Developing and applying a framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

E.P. Kamphuis BSc[a], Supervisor: Dr. ir. W.J.A. van Heeswijk[a], Supervisor: Prof. Dr. M.E. Iacob[a], and Supervisor: ir. H. Stevens[b]

[a]Department of Industrial Engineering & Business Information Systems, University of Twente
[b]Global Supply Chain, Fast Moving Consumer Goods company

November 20, 2020

Foreword

Dear reader,

Thank you for taking the time to read my master thesis.

For almost ten months, I have been working on my master thesis, mostly from my room. I spent my first months working from the research company's office, and even was a part of moving to a new office. I am glad this short period of working with colleagues allowed me to meet a lot of amazing people. Their welcoming attitude made me feel a part of the team for my entire internship period. I would like to thank my team for their support and the experiences we had together. From this team, I would especially like to thank Hadassa, who has always supported me and also pushed me to reach my full potential. Furthermore, I would like to thank Javi. He always took the time to listen to me, and I think he might even have taught me a life lesson or two.

From the University of Twente, I want to thank Wouter for supporting me for this long period. Even though I did not focus on a subject that is close to his field of expertise, he always managed to provide useful feedback. I also want to thank Maria for her valuable contributions during the final stages of writing my thesis. Without feedback from my two supervisors, I would not have managed to write a thesis of this quality. Furthermore, I want to thank all other employees of the University of Twente who have made it possible for me to have an amazing time as a student, which must now come to an end.

Last, but certainly not least, I want to thank my family, friends, and my girlfriend Marloes for supporting me during these trying times. It was not always easy to work from home in solitude, but in the end, you guys helped me get through it all. In special, I also want to thank all my friends who took the time to provide me with feedback on this thesis.

Now we will see what the future will bring. When one door closes, a new one always opens. I'm curious to see where it will lead me.

Eric Kamphuis

Management summary

In this thesis, we successfully created a novel framework for Fast-Moving Consumer Goods (FMCG) companies that describes how to use the output of a Cost-To-Serve (CTS) analysis to find business improvements. A CTS analysis is an approach to determine what the actual logistics costs are of serving a customer by performing cost allocations. By visualizing the output of a CTS analysis in a tool, FMCG companies can find opportunities for business improvements related to topics such as transport optimization, warehouse optimization, and network design.

The research took place in the global organization of a large FMCG company that wished to increase the use of the output from CTS analyses by their operational companies (OpCos). They saw many OpCos that received a CTS implementation achieve significant business improvements and savings in the past, but only 28 of the 42 CTS OpCos still used a CTS analysis a year later. This situation presented a problem for the research company because it means OpCos are missing out on potential benefits. After analyzing the problem context in collaboration with the research company, we decided to increase the usefulness of their CTS analysis by emphasizing diagnostic, predictive, and prescriptive insights rather than mainly focusing on descriptive insights. Eventually, we created a framework for FMCG companies to find business improvements using the output of a CTS analysis to solve the problem for the research company.

The framework consists of four phases, which contain various steps based on reviewed literature and practices of the research company. Figure 1 shows how the Design Science Research Methodology (Peffers et al., 2007), which we followed in this thesis, inspired the phases of the framework. Additionally, we designed a generic algorithm that finds root-causes for a high cost-to-serve of a chosen entity as a potential feature to develop during the Design and Development phase. This algorithm provides users with similar entities, showing for which variable they differ and what the potential savings are, would the difference be resolved. The design of the algorithm originated from a requirement of the research company, but other FMCG companies can consider developing this feature in the Define the Objectives of a Solution phase as well.

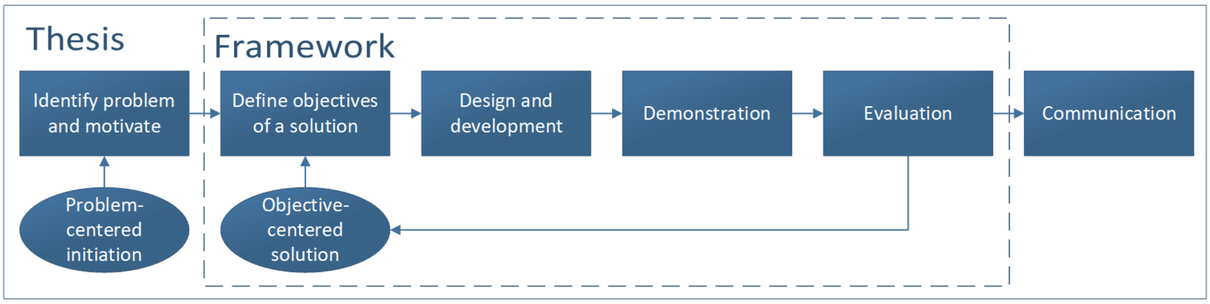


Figure 1: The steps of the Design Science Research Methodology (Peffers et al., 2007) followed in this thesis and included in the framework designed in this thesis

We validated the framework by applying it to the case of the research company, which supplied the research company with an improved CTS tool and two secondary deliverables. We created a new CTS tool in MS Power BI that showed a 38% improvement according to the tool-evaluation method from the Evaluation phase of the framework. Furthermore, we created a separate Power BI report that visualizes past opportunities found by OpCos after a CTS implementation by applying the steps of the Demonstration phase of the framework. Most of those opportunities included a measurement of potential savings, showing an average cost reduction of 3% per OpCo that the research company can use to benchmark future CTS implementations. Finally, we created the root-cause analysis method using R, but could not include it in the CTS tool due to IT restrictions. Nevertheless, the algorithm showed great promise by revealing potential savings up to 20% of costs in scope for different data sets, but the performance of underlying models varied between R^2 values of 0.17 and 0.93, leaving room for improvement.

In conclusion, the case study validated that the framework enables FMCG companies to find business improvements by using descriptive, diagnostic, predictive, and descriptive features in a tool that visualizes the output of a CTS analysis. The case study only revealed four possible improvements we can make to the framework and showed many potential improvements for the research company. The main recommendation for the research company is to start continuously improving its process for visualizing the output of CTS analyses by using the framework. To support the research company, we created a roadmap with recommendations shown in Figure 2.

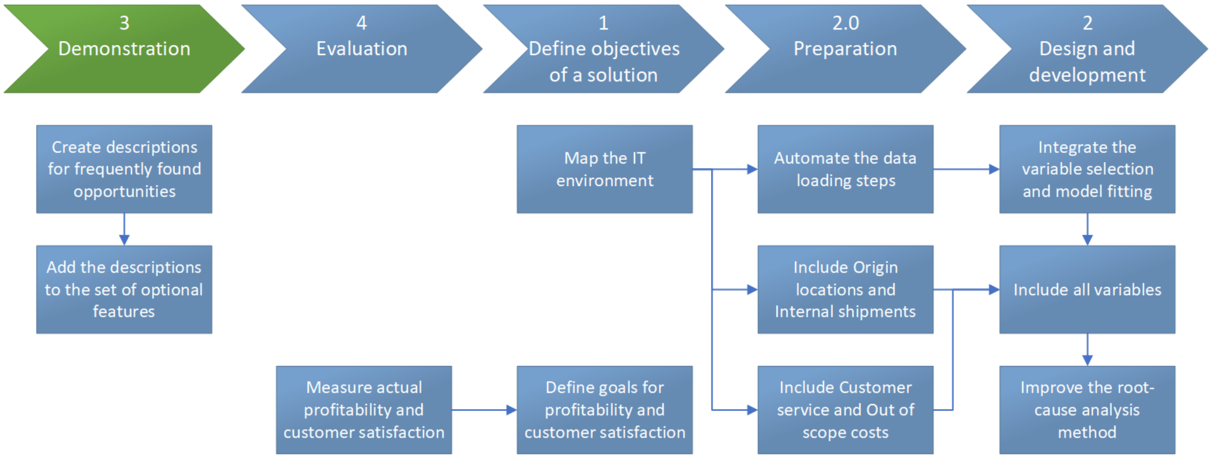


Figure 2: The roadmap with recommendations related to the phases of the framework for fast-moving consumer goods companies to visualize the outcome of a cost-to-serve analysis (the green block indicates the research company is currently in the Demonstration phase)

Contents

Abbreviations

Table 1: Abbreviations used in this report

Abbreviation	Description
AHP	Analytical Hierarchy Process
CTS	Cost-To-Serve
DSRM	Design Science Research Methodology
ERP	Enterprise Resource Planning
FMCG	Fast Moving Consumer Goods
KPI	Key Performance Indicator
MAE	Mean Average Error
MOQ	Minimum Order Quantity
NPS	Net Promoter Score
OpCo	Operational Company
OTC	Order-To-Cash
PW	Production Warehouse
QVD	QlikView Data
RFE	Recursive Feature Elimination
RMSE	Root Mean Squared Error
SKU	Stock Keeping Unit
TPM	Total Productive Maintenance

Chapter 1

Introduction

In this master thesis, in the field of Industrial Engineering and Management, we design a framework that describes how to use the output of a cost-to-serve analysis to find business improvements in fast-moving consumer goods companies. A cost-to-serve analysis is an approach to determine what the actual cost is of serving a customer. "The cost-to-serve analysis provides unique insights into the true profitability of your key customers" (Freeman et al., 2000). A key indicator in a cost-to-serve analysis is the cost-to-serve per volume-unit, which expresses what the costs are of serving a customer one unit of volume. The considered volume-unit depends on the company or the product. The cost-to-serve approach is comparable to the Activity-Based Costing method (Turney, 1992) that allocates resources to an activity and the time-driven Activity-Based Costing method (Kaplan and Anderson, 2003) that involves the time required to perform an activity. However, cost-to-serve takes a more simplistic approach by allocating actual process costs and overheads to orders by using a large amount of data. Ultimately, the cost-to-serve method allocates logistics costs in various cost buckets on an order line level, which means each customer-product combination in an order receives costs related to different activities. We perform the research at a fast-moving consumer goods company, which we refer to as the research company. The research company applies a hybrid variant combining Activity-Based Costing methods with cost-to-serve methods, using the following cost buckets:

- Inter-company Transport
- Delivery to Customer
- Warehousing
- OTC (Order-To-Cash)
- Overheads
- Trade Terms
- Other

Braithwaite and Samakh (1998) introduced the cost-to-serve analysis around the beginning of this millennium. The research company has been performing cost-to-serve analyses for the last five years, but until now, the focus was mainly on individual cost-to-serve implementations, paying less attention to the continuous development of their cost-to-serve analysis as a whole. Figure 1.1 situates the cost-to-serve analysis in the slope of enlightenment, indicating it is fundamental to every business but still under development (Tohamy, 2020). Currently, there is little guidance in the use of the output of a cost-to-serve analysis in literature and practice.

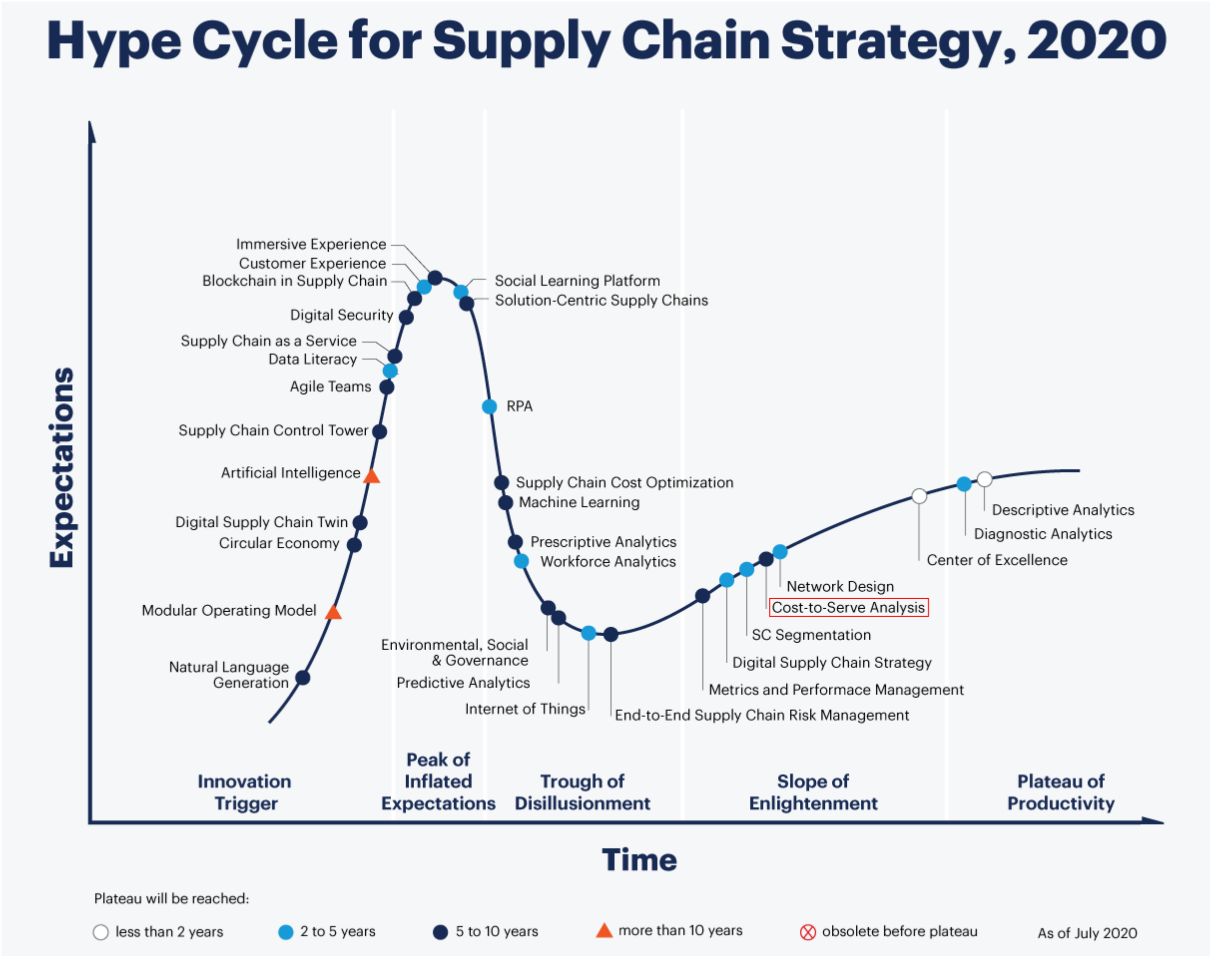


Figure 1.1: The Gartner Hype Cycle for Supply Chain Strategy, 2020 (Tohamy, 2020)

The artifact used to visualize the output of a cost-to-serve analysis is the cost-to-serve tool. Users of this tool can attempt to find business improvements, which can relate to transport optimization, warehouse optimization, network design, or other topics. The framework developed in this thesis, describing how to use cost-to-serve analysis output to find business improvements in fast-moving consumer goods companies, is put to practice by improving the cost-to-serve tool used by the research company.

This chapter describes the process leading up to the research approach. First, Section 1.1 introduces the research of this master thesis by providing background information related to the study and explaining the situation at the start of this research. Based on this situation, Section 1.2 presents the requirement of research, which includes the assignment formulated in collaboration with the research company that served as a starting point for a thorough problem identification process. Section 1.3 presents the problem identification and how we choose a core problem based on the assignment. Finally, Section 1.4 defines the research goal and deliverables. Based on these, we designed steps to solve the core problem, which we linked to research questions. At the end of this section, we present the outline of the report, where we link chapters to the chosen research methodology.

1.1 Background

The purpose of this section is to describe the background of this research. First, Section 1.1.1 introduces the research company. Then, Section 1.1.2 explains how the research company performs cost-to-serve implementations, which start with data collection and end with finding opportunities.

1.1.1 The research company

The research company is a large company active in over 100 countries, employing thousands of people, and selling over 300 different types of fast-moving consumer goods internationally. There is a global organization, and there are multiple Operational Companies (OpCos). An OpCo consists of one or multiple production locations and warehouses within a country. The research company has a decentralized structure. So, each OpCo is an entity responsible for its performance and can make its own decisions to a certain extent. The level of autonomy differs per OpCo as the research company manages some topics globally.

We performed this research from a position in the Global Customer Service team, which is a part of the Global Supply Chain department. At the start of this research, the Customer Service team consisted of 11 people, including a manager, five senior leads, four leads, and the researcher. Every member of the team works on various projects related to capabilities. A capability is a globally developed program that the research company can deploy at an OpCo to improve its performance. The Cost-To-Serve (CTS) capability is the focus area of this research.

At the start of the research, there were three people from the Customer Service team working in the CTS team, enabling OpCos to use their data to determine the cost of serving customers. The research company calculates the CTS on an order line level, determining the CTS for products, origins, vehicles, and shipment types. Then, with the allocated costs and other data, OpCos can improve their business by taking advantage of discovered opportunities through analysis of the output. By making use of these opportunities, the research company improves on their measure of success for CTS implementations, which is the potential savings found. Currently, the research company has a well-working approach that allocates costs on an order line level, which we assumed is valid. However, cost allocations do not automatically provide opportunities for business improvements. Section 1.1.2 explains how the research company performs CTS implementations.

1.1.2 Cost-to-serve implementations

The CTS team has been performing CTS implementations at OpCos since 2015. The CTS capability is important, with more than 40 performed implementations at different OpCos over the past years and more scheduled to come. The steps of a CTS implementation, which did not change much over the years, are as follows:

1. Kick-off, project scoping and tool fit assessment
2. Data collection
3. Data processing and tool calibration
4. User training and Baseline analysis workshop
5. Opportunities assessment
6. The final presentation of results of the opportunities assessment

The kick-off, project scoping, and tool fit assessment do not take much time. The data collection steps take the most time. Then, there is a sequence of steps depending on software solutions to generate insights. Figure 1.2 shows these steps.

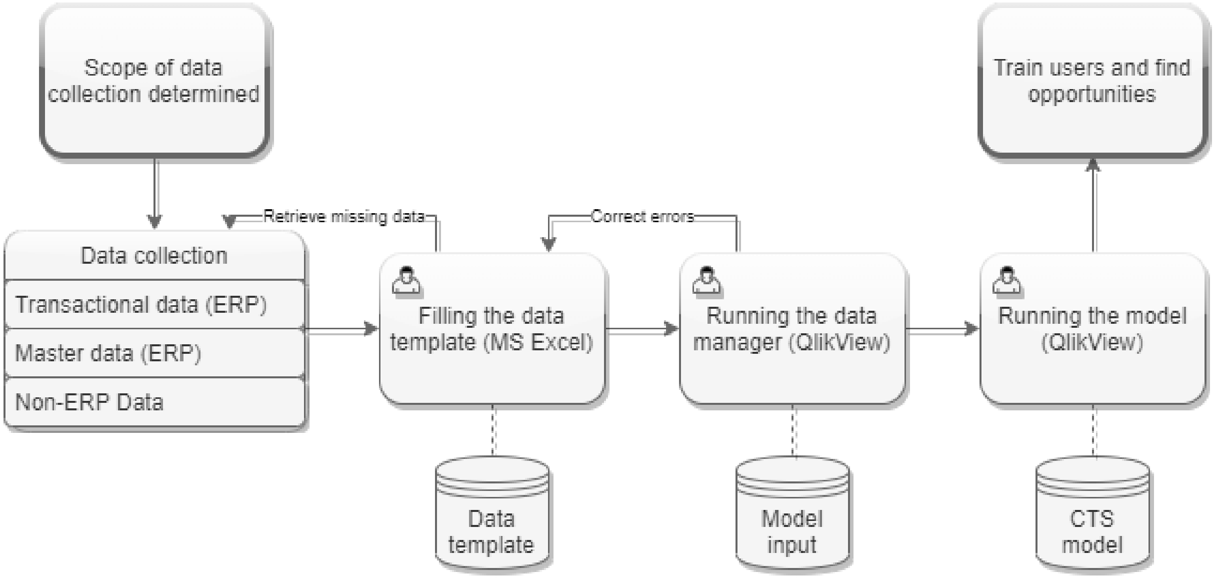


Figure 1.2: The required steps from a scope for data collection to a CTS model that the research company uses to find opportunities

The process from step to step is not linear. After we determine the scope of the data collection, OpCos fill an MS Excel data template. The data template is the location where OpCos combine data from different sources. Data sets usually extracted from the Enterprise Resource Planning (ERP) system are stock transfer orders, sales orders, customer information, product information, and units of measure. Additionally, data sets often collected from other sources are freight, warehousing, warehouse overhead, and order handling costs, as well as product allocation information and logistics discounts, bonuses, and penalties. Depending on the OpCo, some of this data might also come from an ERP system. After the data collection, OpCos process data and calibrate the tool. Then, they load data into a Data manager created with analytics software QlikView. The Data manager performs the cost allocations for each order line. Table 1.1 shows how a cost allocation example concerning the transport costs for a single shipment that goes to one customer. The costs column shows the division of the total costs of 100 per order line based on their weight.

Table 1.1: An example of the cost allocation of a single shipment to a single customer that cost 100

Order line	Product	Quantity	Weight	Costs
1	Small	24	60	5.00
2	Large	12	660	55.00
3	Medium	48	480	40.00

The allocation method of each cost bucket often depends on the route to the customer, shipment type, or product. Sometimes a different method is required based on the preference of an OpCo or available information. Which cost allocation methods the Data manager applies depends on how OpCos fill the data template. So, the QlikView tool incorporates substantive algorithms to handle many situations. As shown in Figure 1.2, a user might have to go back to a previous step due to missing data or errors. The number of times OpCos repeat steps differs. In the end, another QlikView application transforms the model input files created by the data manager into a model. Figure 1.3 shows a simplified version of the model.

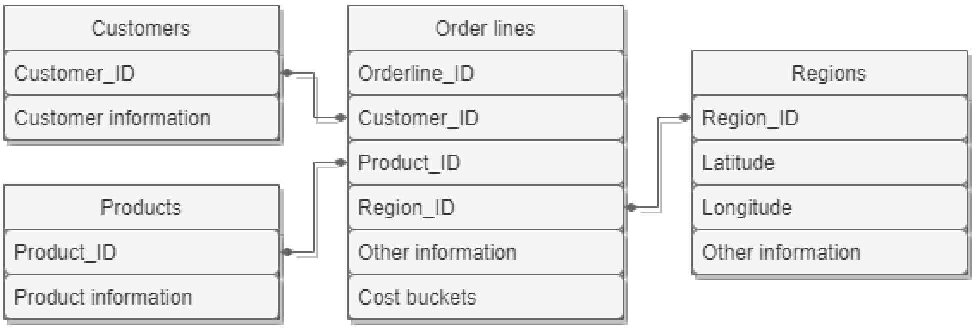


Figure 1.3: A simplified view of the CTS model loaded in QlikView

The actual model is more detailed and contains various tables to support visualizations. The cost bucket table is the most important, as it holds the allocated costs that result from the CTS analysis. In the same application, the model is used to visualize the output of a CTS analysis in the tabs shown in Table 1.2.

Table 1.2: The tabs of the current QlikView tool and their contents

Tab	Contents
Overview	An overview of different costs and where they were made
Key numbers	Key figures for different shipment types and the CTS for different dimensions
Graphs	Nine different graphs with varying functionalities
Validation	Many table views
Reporting	Allows a user to recreate profit and loss reports
Scenario	Allows a user to compare scenarios using various visualizations from other tabs
Maps	Customer locations plotted on different backgrounds
Details	The contents of the three main tables; order lines, products, and customers
Reload	Allows a user to load baseline data, base case data, and run scenarios

Once the OpCos load the model, the CTS team trains users to find opportunities for business improvements using the various visualizations in the tool. Users require training because they must often use different features in combination to find opportunities. Therefore, users require knowledge about the tool, and preferably experience working with the tool, to use the tool to its full potential. Opportunities found in a CTS analysis usually relate to customer collaboration or supply chain optimization. An example of a CTS visualization is shown in Figure 1.4 (Cecere, 2015).

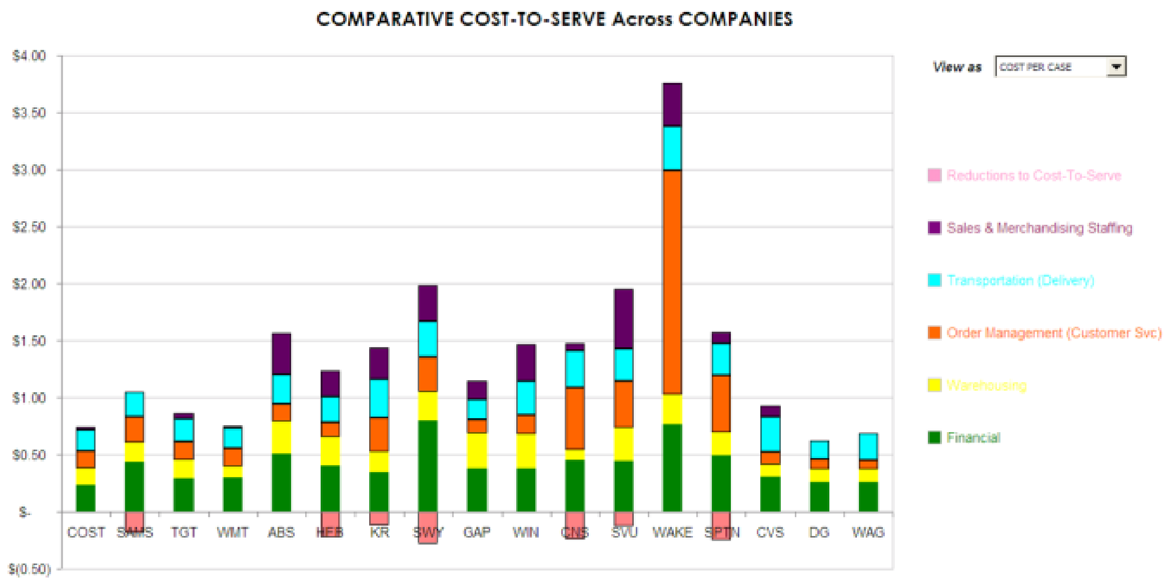


Figure 1.4: CTS visualization example that similarly displays information as the research company (Cecere, 2015)

The visualization shows for which customer the CTS per volume-unit is higher than others, and the impact of each cost bucket. Based on this, OpCos could research how to mitigate the high CTS per volume-unit. The categories shown in Figure 1.4 do not correspond with those of the research company, but it resembles a visualization used in the research company's tool. Finally, the OpCo creates action plans to reap the benefits of the opportunities found, marking the end of the CTS implementation. However, finding these opportunities is a difficult task, according to the CTS team, which leads to the initiation of research in Section 1.2. In some cases, the CTS team follows up on the success of the business changes, but that is not a standard procedure.

1.2 Research initiation

This section presents the starting point of this research. First, Section 1.2.1 presents the motivation for this research based on the context described in Section 1.1. Then, resulting from the research motivation, Section 1.2.2 presents an assignment formulated in collaboration with the research company. The assignment serves as the basis for research into the problem context in Section 1.3. Finally, Section 1.2.3 presents an overview of the stakeholders involved with this research.

1.2.1 Research motivation

Over the past years, many OpCos received CTS implementations that led to significant business improvements and savings, but the research company has not focused on improving the use of the output of the implementations. Many OpCos benefit from CTS implementations and actively support their current operations using the output. Therefore, the research company started initiatives around the start of the research to focus on reviewing the way a CTS analysis can lead to improving the way they deliver products to customers. The CTS team believes their cost allocation methods are strong. However, determining the next steps based on implementations is difficult for OpCos, especially when they already capitalized on the most straightforward opportunities. So, there is a lack of knowledge on how to find opportunities for business improvements using the output of a CTS implementation. Because new technologies emerge, and OpCos experience an increasing difficulty finding new opportunities using the tool, the CTS implementation of the research company is at risk of becoming outdated. Therefore, the requirement of the research company to improve the use of CTS analyses motivates this research.

An underlying reason for the requirement to improve CTS analyses is that a CTS analysis is a part of company frameworks to facilitate continuous improvement. However, not all OpCos continuously use CTS, while the research company aims to develop OpCos with a framework that incorporates this. The research company based their framework on Figure 1.5 but contains other pillars tailored to the company strategy.

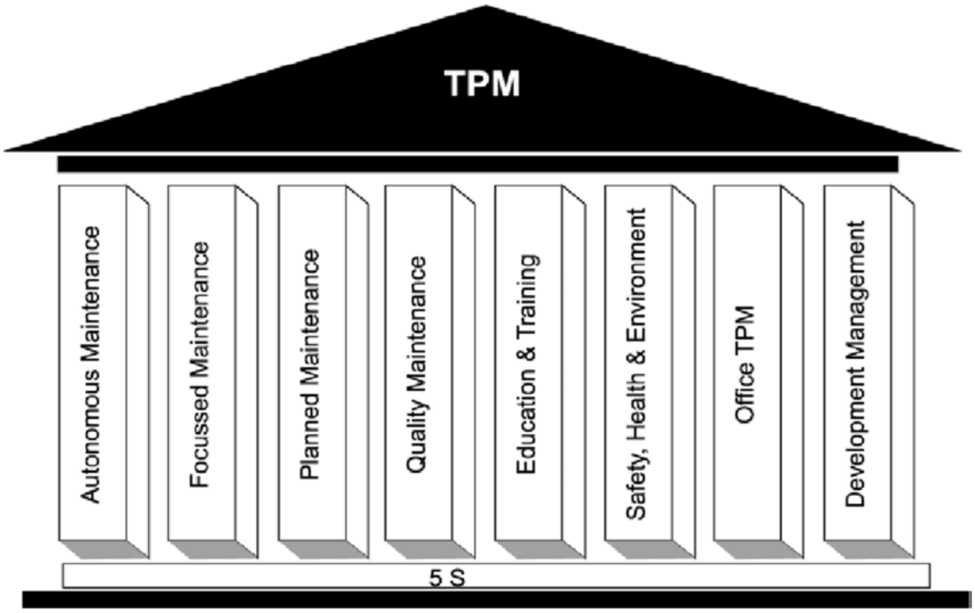


Figure 1.5: Pillars of Total Productive Maintenance as known in literature, which serve as the inspiration for the TPM pillars of the company (Singh et al., 2013)

Total Productive Management or Total Productive Maintenance implies continuous improvement (Nakajima, 1988), which is also the case for the framework adopted by the research company. Therefore, elements of the pillars as the CTS capability should incorporate continuous improvement, which was not the case at the beginning of this research. Thus, we formulated an assignment in Section 1.2.2.

1.2.2 Assignment

Based on Section 1.2.1, the assignment at the research company revolves around the continuous use of CTS. There have been many CTS implementations, but there are not enough OpCos who continue to find opportunities using the CTS tool. It appears that a CTS implementation provides a snapshot of the business at the time of the implementation, but there is no continuous use of the output. It is the wish of the research company to integrate the CTS analysis into the way of working on a strategic/tactical level of OpCos to supply them constantly with opportunities for improvements related to customer collaboration, supply chain optimization, or other areas. In collaboration with the company supervisor, we formulated the assignment as the following problem statement:

> *The current situation is that from the 42 cost-to-serve implementations, only 28 OpCos are still using a cost-to-serve analysis continuously a year later.*

This is a problem for the research company as they plan to perform more implementations and increase the number of OpCos continuously using their CTS analysis. However, looking at this situation, it seems that the CTS capability adds limited value to the process of continuous improvement described in Section 1.2.1 as a limited number of OpCos continues to reap benefits from their CTS analysis. The CTS team desires that all OpCos that have received a CTS implementation should still use their CTS analysis continuously a year later, meaning that the output is regularly updated and reviewed, as is the case for the 28 OpCos. However, 14 OpCos incidentally consult their CTS analysis or have discontinued the use of their CTS analysis. So, that 67% of the OpCos that received a CTS implementation still work continuously with their CTS analysis is too low, serves as the starting point of the problem identification in Section 1.3.

1.2.3 Stakeholders

In this research, we distinguish several stakeholders. Stakeholders can be a person, a group of people, or even an entire company. Table 1.3 shows an overview of the involved stakeholders.

Table 1.3: Stakeholders for the master thesis research

Stakeholder	Description
First university supervisor	Dr. ir. W. J. A. van Heeswijk from the University of Twente
Second university supervisor	Prof. Dr. M. E. Iacob from the University of Twente
Company supervisor	ir. H. Stevens from the Customer Service team, and CTS team
CTS team	A team of three people working actively on CTS implementations
Customer Service team	A team of ten people working on Customer Service capabilities
OpCos	A decentralized branch of the research company

All stakeholders play a different role in this research. We consulted University supervisors to maintain a thesis worthy of graduating from the master's of Industrial Engineering and Management. The company supervisor represents the problem owner of the problem identified in Section 1.3. This stakeholder was consulted, informed often, and played a significant role in verifying outcomes. The CTS team is the problem owner and was involved when requiring more than the single view of the problem owner. The entire Customer Service team should understand the working of the CTS capability. Therefore, we informed them of outcomes to ensure they can understand changes made to the CTS tool or process. Last, OpCos receive CTS implementations and are the final users of the CTS tool. CTS implementations must aim to answer business questions that OpCos have. Therefore, we took the view of OpCos into account during the research.

1.3 Problem identification

This section analyzes the problem context surrounding the assignment presented in Section 1.2.2. First, Section 1.3.1 describes the problem context by creating a problem cluster and selecting potential core problems. Then, Section 1.3.2 evaluates core problems and decides on a focus for this research.

1.3.1 Problem context

The problem context is important in understanding what problems are related to the action problem resulting from the assignment in Section 1.2.2, which is that too few OpCos continuously use a CTS analysis. A way to visualize the problem context is by creating a problem cluster, which is a part of the Managerial Problem-Solving Method described by Heerkens and van Winden (2012). We used this specific part of the method for the problem identification performed in this section. Section 1.4.3 presents the general research methodology used in this research. Conversations with members of the CTS team and materials related to the research company led to insights into the problem context and the creation of the problem cluster. The final problem cluster was verified and agreed upon by the members of the CTS team. Figure 1.6 shows the problem cluster, portraying the problem context for the research.

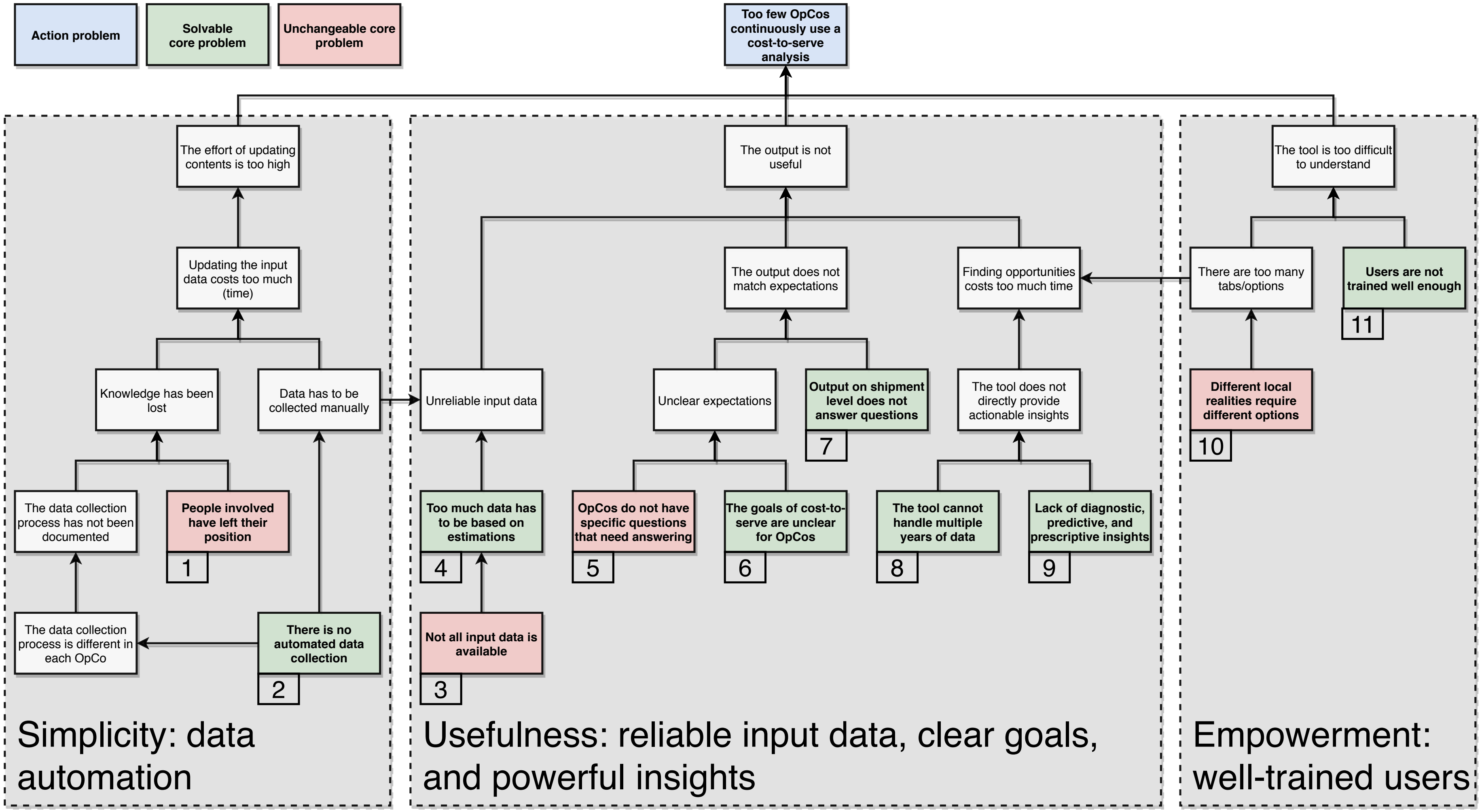


Figure 1.6: Problem cluster showing the problem context at the research company

The problem cluster visualizes causal relations between problems that occur within the problem context. It starts with the action problem at the top of the figure and shows the causes of each problem. When there is no cause for a problem, it is considered a core problem. Core problems are the problems that have the highest effect when solved, as solving these problems has a positive influence on all problems related to this problem. Figure 1.6 shows multiple core problems that influence the action problem which are summarized in Table 1.4. To clarify the problem cluster, we created three main categories. The first category is simplicity, with problems concerning data automation that can severely simplify the CTS implementation process. The second category is usefulness, which concerns reliable input data, clear goals, clear expectations, and handling multiple years of data. So, the tool should provide powerful insights that can drive the business forward. The third category is empowerment, which implies that the tools users are fully empowered to make to get the most out of the tool.

Table 1.4: Summary of core problems (Red = Unchangeable core problem, Green = Solvable core problem)

Number	Problem	Explanation
1	People involved have left their position	People previously involved in CTS have left their position, which results in a loss of tacit knowledge.
2	There is no automated data collection	The absence of an automated data collection leads to an intensive data gathering process.
3	Not all input data is available	Some OpCos cannot provide all the necessary input data. Sometimes, the required input data does not exist at all for an OpCo, which decreases the usefulness of the tool.
4	Too much data has to be based on estimations	Various estimations are made in the CTS implementation. However, it is not sure whether they are all valid.
5	OpCos do not have specific questions that need answering	OpCos do not know which problems they want to solve using CTS. Often they are baffled by the possibilities.
6	The goals of cost-to-serve are unclear for OpCos	OpCos might focus more on obtaining the tool than considering what the general goals are of the implementation.
7	Output on shipment level does not answer questions	The current tool provides data on a shipment level. Including data on different levels (e.g. invoice or order) could provide more insights.
8	The tool cannot handle multiple years of data	Because the tool does not allow for multiple years of data trend analysis or monitoring change is not possible.
9	Lack of diagnostic, predictive, and prescriptive insights	The current tool mostly displays data descriptively. Therefore, it is more difficult to obtain insights quickly.
10	Different local realities require different options	Due to inherent differences between OpCos, requirements can differ heavily.
11	Users are not trained well enough	Using the tool requires training. More training would lead to better use of the tool.

Starting from the action problem determined in Section 1.2.2, stating that too few OpCos continuously use a CTS analysis, we found seven solvable core problems. Section 1.3.2 discusses problems that are potential core problem for this research. In that section, we make a methodological decision regarding which problem(s) to focus on based on impact, strategic importance, and effort.

We cannot solve four of the potential core problems presented in this research. The first problem that we cannot solve is Problem 1, that people involved have left their position, as there is no way to influence the career path of individuals as a part of this research. The second problem that is not possible to solve for two reasons is Problem 3, that not all input data is available. First, when an OpCo does not have to collect specific data, there is no reason to change this situation. Second, when data should be available, the OpCo should facilitate this locally, as they maintain their own IT systems. The global organization can help OpCos to collect specific data, but it is not the focus of this research. The resulting problem is Problem 4, that too much data relies on estimations, which is considered as a potential core problem because the problem connects to Problem 3 by a single relationship. Third, Problem 5, stating that OpCos do not have specific questions that need answering, presents a circumstance that we cannot influence. Last, Problem 10, that different local realities require different options, concerns local factors. It is not within the power of this research to change elements as culture or legislation in countries. So, we do not focus on these issues in this research. Section 1.3.2 presents the problems that we do consider.

1.3.2 Core problem

This section considers the solvable core problems identified in Section 1.3.1, determining which of these problems yields the highest impact when solved and is solvable within the time provided for this research. First, we evaluate the solvable problems and choose a focus for the research. Then, we review the understanding of the current reality and the norm for the selected problem.

Selection of the core problem

Section 1.3.1 showed seven potential core problems. We made a well-funded decision concerning a focus for this research using the Analytical Hierarchy Process (AHP) developed by Saaty (2000). We applied this method to deal with the subjectivity of the respondent as the method includes a consistency check. We performed the AHP with the company supervisor, who was asked questions without knowing how the answers would influence the outcome. Details of the process are shown in Appendix A. Table 1.5 shows the outcome of the AHP.

Table 1.5: The final outcome of the AHP showing the overall priority of each problem based on the determined criteria

Rank	Problem	Impact	Effort	Strategic importance	Overall priority
1	2: There is no automated data collection	0.11	0.00	0.11	0.23
2	9: Lack of diagnostic, predictive, and prescriptive insights	0.10	0.01	0.10	0.20
3	11: Users are not trained well enough	0.08	0.01	0.08	0.17
4	6: The goals of cost-to-serve are unclear for OpCos	0.06	0.01	0.06	0.14
5	8: The tool cannot handle multiple years of data	0.03	0.02	0.05	0.10
6	4: Too much data has to be based on estimations	0.05	0.02	0.03	0.10
7	7: Output on shipment level does not answer questions	0.02	0.02	0.02	0.06

Table 1.5 shows that the problem with the highest priority is problem 2, that there is no automated data collection. This problem is important for the development of the tool but less suitable as a problem to research for an Industrial Engineering & Management master thesis since the scope of this problem merely relates to a data connection and lacks an element related to improving business operations. Therefore, Problem 9 is the focus of this thesis:

There is a lack of diagnostic, predictive, and prescriptive insights.

This problem poses as the starting point for the problem approach presented in Section 1.4. By using the many features of the research company's CTS tool, which mainly incorporates descriptive analytics, users can obtain higher-level insights. However, the extent to which this is possible strongly relies on the expertise of users. So, improving the presence of different types of features can lead to actionable insights, increase the likelihood that OpCos continue to use the output of their CTS analysis, and thereby solve the action problem. Thus, the focus is on which descriptive, diagnostic, predictive, and prescriptive analytics the research company should apply when visualizing the output of a CTS analysis to increase user-value, where analytics are underlying mechanisms used in features. The next section explains the differences between types of analytics.

Norm and reality

To determine a clear focus for this research, we consider the current reality and the desired norm for the problem solved. As an example, the absence of an automated data connection describes a clear reality, where the norm is an established automated data connection. For the lack of diagnostic, predictive, and prescriptive insights, it is not directly clear what the norm and reality entail. The Analytic Ascendancy Model (Elliott, 2013), shown in Figure 1.7, makes a clear distinction between different types of insights.

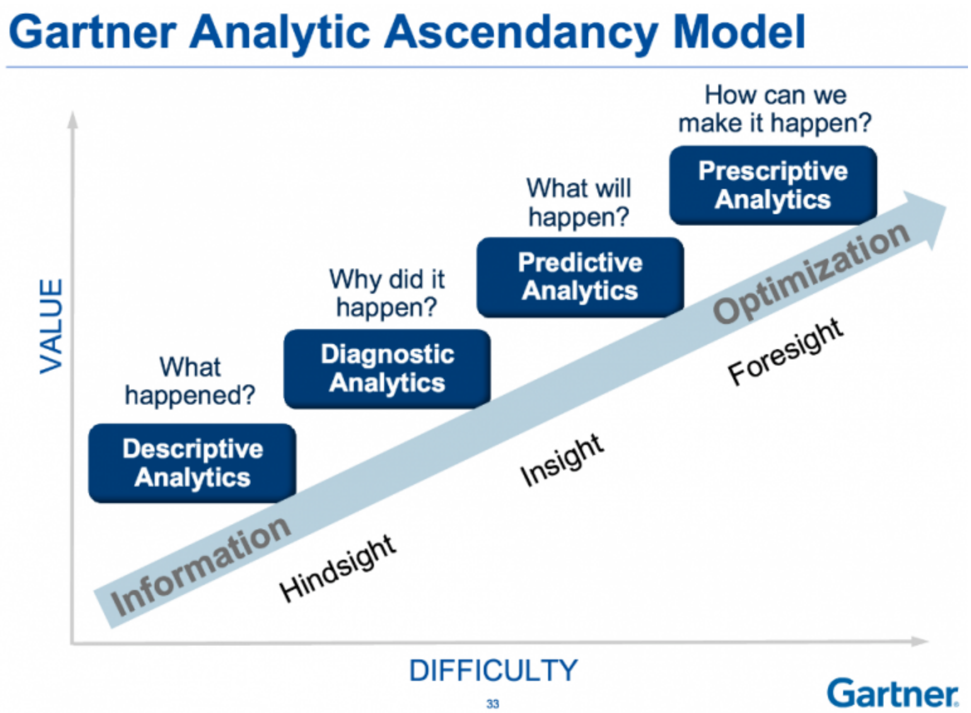


Figure 1.7: The Analytic Ascendancy Model by Gartner showing types of data analytics (Elliott, 2013)

Elliott (2013) states that analytics other than the descriptive kind provide more value, but they are more difficult to provide. In the case of this research, the reality is that there is a certain number of features in the current tool that mostly utilize descriptive analytics. The norm contains an increase in the presence of diagnostic, predictive, and prescriptive analytics compared to the number of descriptive analytics. So, successful research should lead to an increase in features that use higher-level analytics to provide insights without relying on the expertise of users. Realizing this, more OpCos should start using the output of CTS analyses continuously. Section 2.1.1 provides a more detailed description of the model shown in Figure 1.7. Regarding the stakeholders described in Section 1.2.3, the CTS team is the problem owner. The entire CTS team is affected directly by the problem and would benefit when the problem is solved, as solving the problem provides the CTS team with a better tool to support OpCos.

1.4 Problem approach

This section presents the approach to solve the core problem identified in Section 1.3.2. First, in Section 1.4.1, we demarcate the scope and formulate the goal of the research. Then, Section 1.4.2 presents the main research questions and supportive research questions, and Section 1.4.3 presents the Design Science Research Methodology as the methodology for this thesis. Finally, Section 1.4.4 outlines this report, combining the research questions and the research methodology.

1.4.1 Research scope and goal

This research applies to Fast-Moving Consumer Goods (FMCG) companies that use CTS analyses or want to start using CTS analyses. We answer several questions in the context of the research company, but because the outcome of the research applies to many OpCos that are individual FMCG companies with a distinctive local reality, we can generalize findings. Designed solutions fit all OpCos, taking factors such as different regions and sizes into account. Furthermore, we took OpCo specific delivery strategies, vehicle types, and other factors influencing the CTS into account. Finally, designed solutions add value to the Customer Service team. As guidance, features focus on customer-related improvements.

Section 1.3.2 showed that the problem to solve is the lack of diagnostic, predictive, and prescriptive insights. So, the focus of this research is on the analytics used to visualize the output of the CTS implementation. Consequently, we determined the research goal and the main deliverable accordingly. Figure 1.8 shows how, starting from the formulation of the assignment, these were determined.

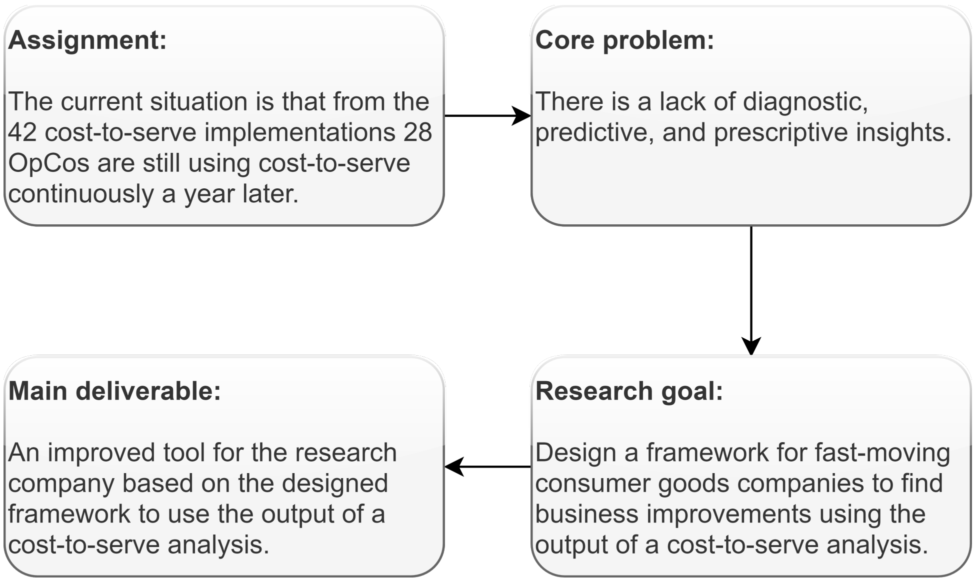


Figure 1.8: The connection between the assignment, the selected core problem, the research goal, and the main deliverable of this thesis

As the research in this thesis applies to FMCG companies that want to find business improvements, a generic framework can outline how to visualize the output of a CTS analysis. Findings apply to any FMCG company regardless of their software or business processes. Thus, the generalized research goal is as follows:

> *The goal of this research is to design a framework for fast-moving consumer goods companies to find business improvements using the output of a cost-to-serve analysis.*

For the research company, the aim is to increase the number of OpCos that continuously use CTS analyses. The CTS tool visualizes the output of the analysis, but the analytics used in the tool's features appear insufficient as a large proportion of the OpCos does not continue to use the tool. The framework created in this thesis includes the development of the CTS tool. By applying the framework to the research company, we created a new CTS tool. The reason to focus on the CTS tool is that it is at the pinnacle of finding opportunities. The main deliverable of the thesis is linked to this and defined as follows:

> *For the research company, we create a new tool based on the designed framework to find business improvements using the output of a cost-to-serve analysis.*

There are two reasons for creating a new tool rather than extending the current CTS tool. Firstly, manufacturers no longer maintain QlikView software. And secondly, the research company selected new tooling software as a company-wide solution.

1.4.2 Research questions

This section presents the main research questions, as well as several supportive questions. Section 1.4.4 details which chapters and sections correspond to the questions. We formulated two main research questions for this thesis based on the assignment, the core problem, the research goal, and the main deliverable shown in Figure 1.8. The first main research question, which focuses on the research goal, is as follows:

> ***1 How can fast-moving consumer goods companies find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a cost-to-serve analysis?***

Answering the first main research question provides a framework to find business improvements using the output of a CTS analysis, which we based on literature and findings related to the research company. Additionally, we formulated a second main research question for the application of the framework in practice. This research question focuses on the main deliverable, a new CTS tool for the research company, created as a part of a case study. The second main research question is as follows:

> ***2 Can the research company improve the use of the output of cost-to-serve analyses by applying the framework designed in this research to create a new tool?***

The answer to the second main research question is of value to the research company, as it includes the creation of a new tool. We answer the research question by applying the framework to the research company in a case study. Besides answering this research question, the case study allows for a validation of the framework. Several research questions support the two main research questions, which we present in the rest of this section.

Framework development

Answering the first main research question creates a framework. The first supportive research question to develop the framework is:

> *1.1 Which descriptive, diagnostic, predictive, and prescriptive features can visualize the output of a cost-to-serve analysis?*

When developing the CTS tool, certain features and analytics add value where others do not. We reviewed literary sources to understand which features and analytics to apply when visualizing the output of a CTS analysis, and we assessed how users appreciate parts of the research company's current tool by surveying users in the CTS team and OpCos. Then, we consolidated the results of this research by creating a set of optional features to develop for a CTS tool as a part of the framework to find business improvements using the output of a CTS analysis.

1.2 How can fast-moving consumer goods companies find a root-cause for a high cost-to-serve per volume-unit based on data used in the cost-to-serve analysis?

Section 4.2.1 argues the development of a feature using diagnostic analytics to find root-causes for a high CTS per volume-unit for the research company. This method allows users to select an entity or a combination of entities and find causes for a high CTS per volume unit. An algorithm finds similar entities with a lower CTS per volume-unit, showing where FMCG companies can potentially improve each entity. The purpose of this research question is to develop a generic method based on literature regarding input variables, variable selection, and predictive models. This method is applicable in multiple settings as we applied general techniques that are not only applicable in the case of the research company.

1.3 How can fast-moving consumer goods companies map business improvements obtained with insights from a cost-to-serve analysis?

The goal of a CTS analysis is to find business improvements. Preferably, FMCG companies can learn from past business improvements to strengthen their CTS analysis. To understand how insights can lead to opportunities for business improvements, we performed a literature study and researched past CTS implementations performed by the research company. Answering this research question provided a structure to map opportunities resulting from obtained insights.

1.4 Which techniques can assess the quality of a tool visualizing the output of a cost-to-serve analysis?

FMCG companies should measure the effect of changes made to a tool to track improvement. For this, we reviewed the literature concerning technology assessment models, developing a method to evaluate the performance of tools visualizing the output of a CTS analysis. The resulting method includes a measurement of success, tool assessment, and feature assessment that FMCG companies should apply before every iteration that changes to a CTS tool.

Case study: Applying the framework in the research company

After the creation of the framework by answering the first main research question, we validate the framework by applying it to the research company. Consequently, we developed a new tool for the research company. The first research question that supports the second main research question is:

2.1 How is the cost-to-serve implementation of the research company performing?

The purpose of this research question is to understand what type of business improvements the research company found in the past and assess the performance of CTS implementations in the research company. First, related to Research Question 1.3, we validated the structure to map opportunities. We mapped and collected opportunities found in the research company in the past years, creating an interactive report. Second, related to Research Question 1.4, we validated the method to evaluate the performance of tools visualizing the output of a CTS analysis. We assessed the performance of the research company's CTS implementations and the current CTS tool by surveying users.

2.2 Which descriptive, diagnostic, predictive, and prescriptive features should the research company use to visualize the output of a cost-to-serve analysis?

The answer to this research question clarifies what to include in the new CTS tool of the research company. Related to Research Question 1.1, we validated the set of options by assessing which features from the current CTS tool should remain and what to develop in the new CTS tool of the research company. We validated decisions made by surveying users of the CTS tool.

2.3 How can the research company incorporate the chosen descriptive, diagnostic, predictive, and prescriptive analytics into a new tool?

The purpose of this question is to clarify how we can create a usable CTS tool by presenting the design and creation of the new CTS tool. We paid attention to the preparation steps required and the development of the new CTS tool that contains descriptive analytics from the current CTS tool and analytics that are new to the research company. Related to Research Question 1.2, we created a modular root-cause analysis approach that the research company can potentially include in the new CTS tool.

1.4.3 Research methodology

As explained in Sections 1.4.1 and 1.4.2, we created a framework for visualizing the output of a CTS analysis and put it to practice by designing a new tool. So, we designed a new artifact based on research. Therefore, the Design Science Research Methodology (Peffers et al., 2007) (DSRM) presents a suitable methodology. The DSRM is a widely accepted framework for Design Science Research in Information Systems. According to Peffers et al. (2007), Design Science is of importance in the creation of successful artifacts. In the case of this research, the artifacts refer to methods in the framework. These methods lead to the development of a CTS tool. The DSRM consists of a nominal process sequence containing six steps that can start in any of the first four steps depending on the research. Figure 1.9 shows the steps and possible research entry points. With the DSRM, one can continuously improve existing solutions.

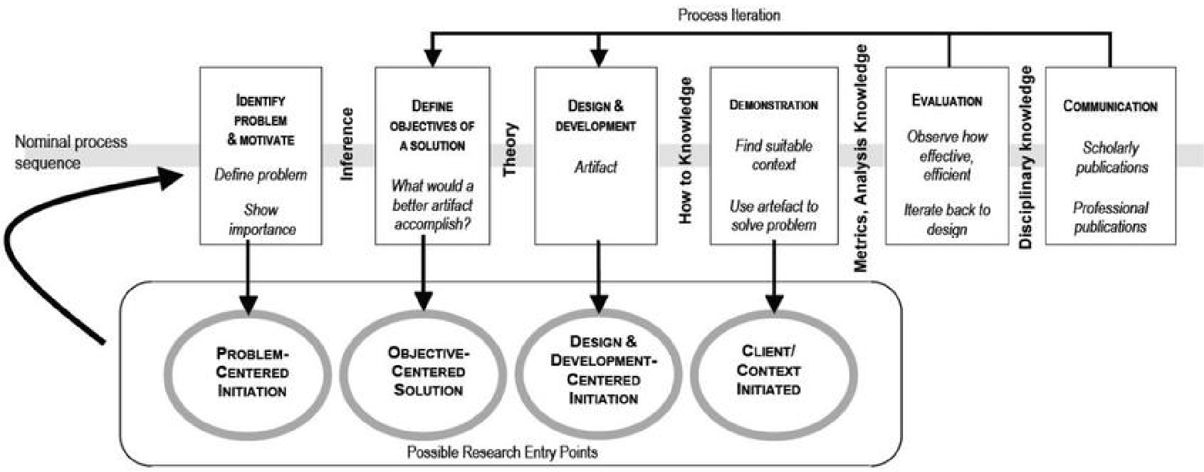


Figure 1.9: The Design Science Research Methodology framework (Peffers et al., 2007)

In this research, there is a problem-centered initiation based on Section 1.2.2, which presented the initial thesis assignment. Therefore, the process starts with identifying and motivating the problem, which we did in Section 1.3. Section 1.4.1 addressed the next step, which is to define the objectives of a solution. After this step, we follow all the steps in this thesis. The last step is to publish the research in the form of the master thesis. Section 1.4.4 explains how we incorporate these steps into the structure of this thesis.

1.4.4 Report outline

In this thesis, we answer research questions using the DSRM framework presented in Section 1.4.3. That section explains this research has a problem-centered initiation formulated in Section 1.2, which served as the entry point for the problem identification stage in Section 1.3 and the definition of objectives in Section 1.4.1. Figure 1.10 shows these sections, along with other chapters in the outline of the report.

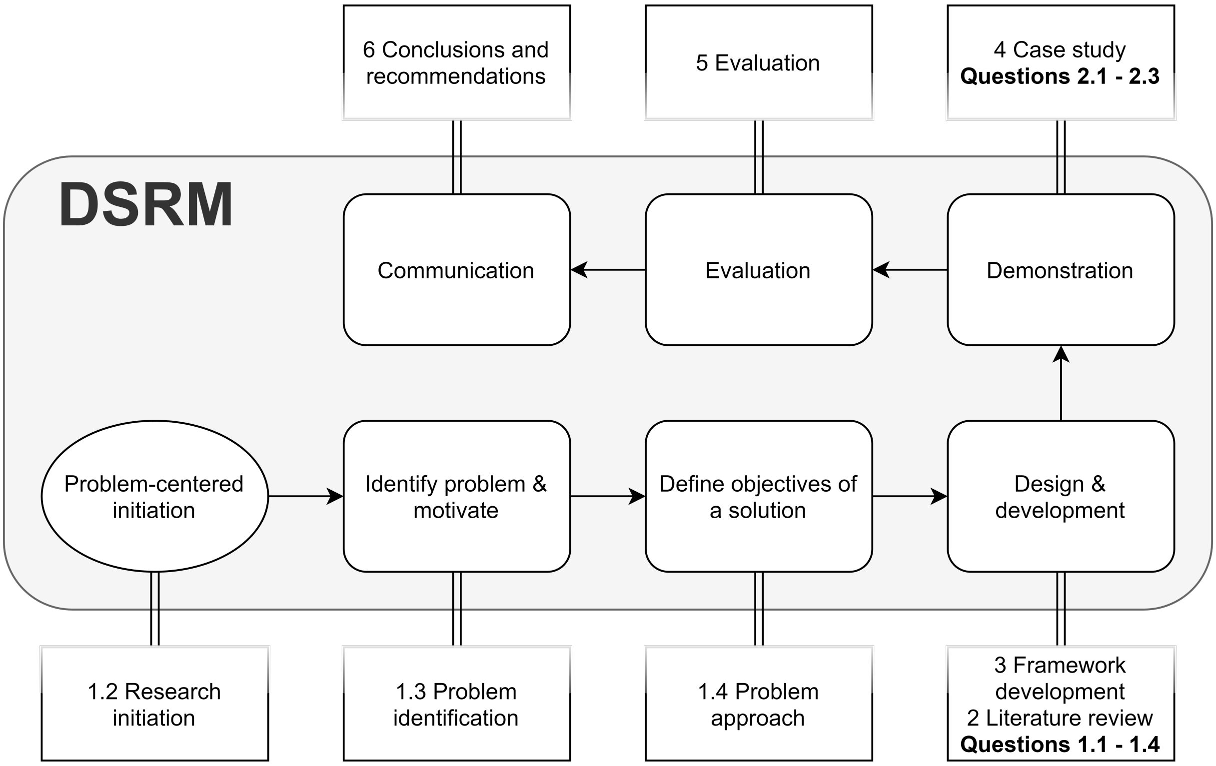


Figure 1.10: The outline of the thesis combined with the Design Science Research Methodology framework adapted from Peffers et al. (2007)

Chapters 3, 4, 5, and 6 represent the steps of the DSRM. Chapter 2 supports the creation of the framework by providing an overview of relevant literature. In Chapter 3, we create the framework for fast-moving consumer goods companies to find business improvements using the output of a CTS analysis. The framework shows how FMCG companies can visualize the results of a CTS analysis in an ideal situation. This chapter includes the development of a root-cause analysis feature that uses diagnostic analytics. Section 4.2.1 argues the development of the root-cause analysis as a part of this research. So, in Chapter 3, we answer the following questions:

> ***1 How can fast-moving consumer goods companies find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a cost-to-serve analysis?***
>
> *1.1 Which descriptive, diagnostic, predictive, and prescriptive features can visualize the output of a cost-to-serve analysis?*
>
> *1.2 How can fast-moving consumer goods companies find a root-cause for a high cost-to-serve per volume-unit based on data used in the cost-to-serve analysis?*
>
> *1.3 How can fast-moving consumer goods companies map business improvements obtained with insights from a cost-to-serve analysis?*

1.4 Which techniques can assess the quality of a tool visualizing the output of a cost-to-serve analysis?

In Chapter 4, we applied the framework to the research company in a case study. We validated the framework by showing how it performs in practice and developed a new tool for the research company. So, in this chapter, we answer the following questions:

2 Can the research company improve the use of the output of cost-to-serve analyses by applying the framework designed in this research to create a new tool?

2.1 How is the cost-to-serve implementation of the research company performing?

2.2 Which descriptive, diagnostic, predictive, and prescriptive features should the research company use to visualize the output of a cost-to-serve analysis?

2.3 How can the research company incorporate the chosen descriptive, diagnostic, predictive, and prescriptive analytics into a new tool?

Chapter 5 evaluates the implementation of the framework in the research company and the new CTS tool. This evaluation is similar to the assessment of the current situation in Section 4.1. Then, Chapter 6 answers all research questions in the form of conclusions and presents recommendations, which include advice on how the company should structure future developments in light of the implementation of the framework. Finally, the discussion section reviews the applicability of the framework, limitations, opportunities for future research, and the contributions to theory and practice.

Chapter 2

Literature review

This chapter reviews literature that supports the answering of the following main research question:

> ***1 How can fast-moving consumer goods companies find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a cost-to-serve analysis?***

First, Section 2.1 defines key constructs by explaining important concepts applied in this thesis. Then, Sections 2.2 and 2.3 present the results of literature studies on the benefits of cost-to-serve and potential analytics from literature. The first review focuses on what the end goals of a cost-to-serve analysis are, and the second on applicable techniques for visualizing the output of a cost-to-serve analysis. The focus is not descriptive analytics, as the research company uses these extensively in the current situation described in Section 4.1. Then, Section 2.4 presents variables that can serve as potential cost drivers for a high cost-to-serve per volume-unit, Section 2.5 presents methods to reduce the number of variables, and Section 2.6 presents literature concerning models to predict a high cost-to-serve per volume-unit. The root cause analysis in Section 3.3.4 uses these variables and models. Finally, Section 2.7 performs research into methods to determine the performance of a tool, which are used in Section 3.5.2.

2.1 Key constructs

Throughout this report, we apply various concepts, structures, and research components. The purpose of this section is to shed light on these topics to ensure a clear understanding of their definition. Concepts explained in this section support the answering of Research questions 1.1 and 1.4. First, Section 2.1.1 presents the different types of analytics referred to throughout this thesis. Then, Section 2.1.2 presents a model that visualizes requirement engineering for feature development. Finally, Section 2.1.3 defines the concepts frameworks, methods, and roadmaps.

2.1.1 Opportunities and analytics

In this thesis, various sections mention opportunities for business improvements. An opportunity is an "exploitable set of circumstances with an uncertain outcome, requiring a commitment of resources and involving exposure to risk" (BusinessDictionairy.com, 2020). In this research, an opportunity is within the scope defined in Section 1.4.1 and found with the CTS tool. A tool refers to an application, which is "a software program that runs on your computer" (Christensson, 2008) and leads to opportunities by providing insights. Section 1.3.2 presented the Gartner Analytic Ascendancy Model (Elliott, 2013), explaining different kinds of analytics supply hindsight, insights, and foresight. These levels of insights answer different types of questions (Jong, 2019). Table 2.1 shows the types of questions the different analytics answer.

Table 2.1: The types of questions (Jong, 2019) answered with descriptive, diagnostic, predictive, and prescriptive analytics (Elliott, 2013)

Analytic	Type of questions answered
Descriptive	What happened?
Diagnostic	Why did something happen?
Predictive	What is likely to happen in the future?
Prescriptive	Which action should be taken to gain a future advantage or mitigate a threat?

We first apply this understanding of different types of analytics in Section 3.5.3, and often in later sections. The answers to these questions provide insights that lead to opportunities. Analytics are mechanisms used in features that provide insights.

2.1.2 Features and requirements

A feature is a part of a functional item, which is a tool in this research. Features are often visualizations of a certain kind. To create a new or improve an existing feature, we must consider requirements. Figure 2.1 shows a model created by Pandey et al. (2010) for the collection of requirements.

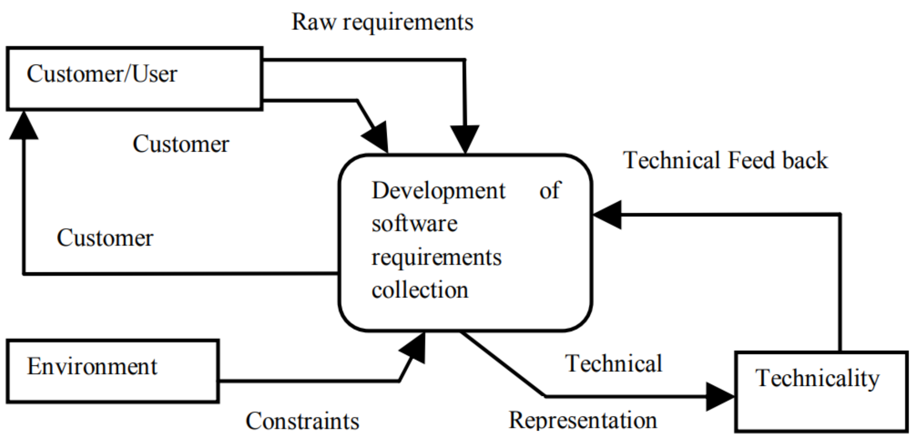


Figure 2.1: Development of requirements (Pandey et al., 2010)

Requirements are dependent on users, the environment, and technical aspects. Furthermore, the progress of requirements development influences the users and the technical aspects. Hence, it is important to formulate clear requirements that are agreeable from these three points of view. We apply this model in Section 3.1.2.

2.1.3 Frameworks, methods, and roadmaps

In this thesis, we create and apply a framework, methods, and a roadmap. We create a framework based on the DSRM (Peffers et al., 2007) that contains several methods in Chapter 3. For example, Section 3.1 requires a prioritization of features. Multi-Criteria Decision Making can support this process as one of the most well-known branches of decision making (Triantaphyllou, 2000). Figure 2.2 (Chen and Hwang, 1992) shows various decision making methods.

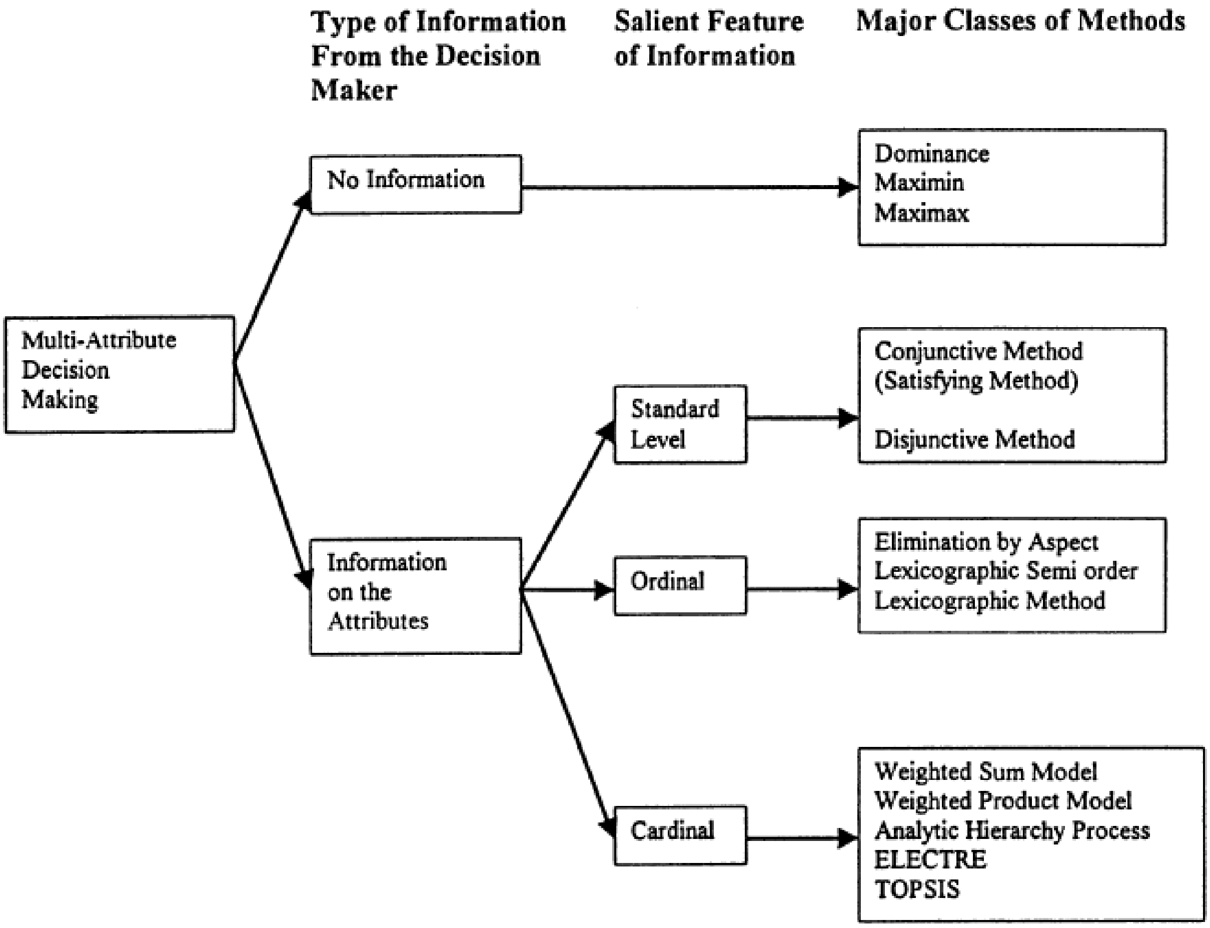


Figure 2.2: A taxonomy of Multi-Criteria Decision Making Methods (Chen and Hwang, 1992)

The AHP method, applied in Section 1.3.2, is a part of this overview. When deciding what to develop in Section 3.1, this technique is applicable. Finally, Section 6.2 includes a roadmap. "A roadmap is a strategic plan that defines a goal or desired outcome and includes the major steps or milestones needed to reach it" (ProductPlan, 2020).

2.2 Benefits of cost-to-serve

This section reviews the obtained benefits from cost-to-serve analyses outside of the research company. We mainly do this to answer the Research question 1.3 How can fast-moving consumer goods companies map business improvements obtained with insights from a cost-to-serve analysis?, by identifying main categories of opportunities found in a cost-to-serve analysis. Section 3.1.1 applies the outcome of this literature study to assess benefits held by different potential features, and Section 3.5.1 to determine measures of success.

Poole (2017), who performed a study in an FMCG company, states that the goal of a CTS analysis is to support three factors:

- It supports customer service optimization by redesigning processes to make them more efficient.
- It should support collaboration strategies, where actions benefit both the company and the customer.
- It helps a company to provide the right service at the optimal costs by providing insights into how they service customers and what the incurred cost was of different activities.

In general, Poole describes that a CTS analysis is about improving customer service. Kolarovszki et al. (2016) take a different view by focusing on the relationship between customer profitability and a CTS analysis through segmenting customers. The focus on profitability often reoccurs. Kone and Karwan (2011) say about a CTS analysis that it "allows companies to monitor their costs and therefore manage their pricing strategy and profitability." Findings of a study by Guerreiro et al. (2008) show that measuring the CTS enables a more comprehensive customer profitability analysis than the compared studies. Holm et al. (2012) support the research of Guerreiro et al. by stating "the measurement of cost-to-serve provides specific customer information that enables a more comprehensive CPA than when only measuring gross profit from products," where CPA is referring to a Customer Profitability Analysis. Most studies do not focus on the customer, contrary to the research company. When reviewing the benefits mentioned in this section, the main goal is usually profit. Poole (2017) explains that the purpose of customer service optimization is to make processes more efficient, which implies a reduction in costs. Furthermore, customer collaboration means the same but focuses on shared benefits, and a profitable pricing strategy also focuses on increasing profits. Thus, companies should measure obtained profits to assess the performance of a CTS analysis.

Besides profits, customers also hold importance concerning the benefits of CTS. As customers are critical to the success of FMCG companies, customer satisfaction is also an important indicator. Furthermore, performing this research from the Global Customer Service team of the research company, there is an increased interest in customers. A customer is the point of sales for the company. In the case of an FMCG company, a consumer usually follows the customer. A general method to measure customer satisfaction is by asking customers how satisfied they are and dividing the positive responses by the total number of respondents (Tripathi, 2020). The Net Promoter Score (NPS) is another method that measures customer satisfaction that the research company uses. Reichheld (2003) developed this method, which relies on customers answering the question: "How likely is it that you would recommend our company to a friend or colleague?" Then, responses returning a 9 or 10 (out of 10) represent promoters, responses returning a 7 or 8 represent passives, and responses returning a lower score represent detractors. Respondents that are passives do not influence the NPS score, which Reichheld calculates as follows:

$$NPS = \frac{Promoters}{Respondents} - \frac{Detractors}{Respondents}$$

According to Reichheld (2003), an NPS above 75% indicates world-class customer loyalty. However, the NPS method gained some criticism. Fisher and Kordupleski (2019) claim that the NPS method performs poorly and propose a superior Customer Value Management method. Furthermore, Poole (2020) proposes a Value Enhancement Core method that outperforms the predictive accuracy of the NPS concerning customer loyalty. So, there are various rivaling methods available. In general, FMCG companies have certain customer satisfaction measures in place. The important factor is that companies should measure customer satisfaction in a way and take it into account when evaluating the performance of CTS analyses.

In conclusion, the main benefits of a CTS analysis relate to profitability and customer satisfaction. Specifically, a CTS analysis provides benefits that relate to customer service optimization, customer collaboration, and a profitable pricing strategy. The measurement of profitability is straightforward, and for customer satisfaction, various measurement techniques are available, which includes the NPS measurement.

2.3 Analytics using cost-to-serve

Section 2.2 clarified that a cost-to-serve analysis focuses on increasing company profitability and customer satisfaction. Here, we present various analytics based on the output of CTS analyses that provide insights that can lead to the mentioned benefits. We mainly do this to answer research question 1.1 Which descriptive, diagnostic, predictive, and prescriptive features can visualize the output of a cost-to-serve analysis?, by supporting the collection of potential analytics in Section 3.1. Furthermore, the contents of this literature review inspired the importance of a cost allocation step in the framework presented in Section 3.2.2, and the inclusion of hierarchies and segmentation in the framework presented in Section 3.3.1.

Kone and Karwan (2011) focus on predicting the CTS of new customers by clustering customers using various classification attributes. Sun et al. (2015) continue on this with an improved attribute selection, and more recently, Wang et al. (2020) also focus on the estimation of the CTS, concerning the routing costs for new customers in particular. Wang et al. state that their model is more accurate than the previous models. Özener et al. (2013) provides another example of estimating the costs of including a new customer by evaluating customers that use a Vendor Managed Inventory. With a Vendor Managed Inventory, a company can control the inventory levels of its customers. Thus, there are various examples where the focus is on estimating costs for new customers. It is of importance to take into account that such estimations often apply to other dimensions, such as products, as well. Another specific example of the use of CTS data is for optimizing third-party logistics service delivery (Ross et al., 2007). That research focuses on identifying the cost-drivers of third-party-logistics providers within the internal supply chains. So, it is not customer-focused research.

Everaert et al. (2008) focus on improving logistics using a time-driven Activity Based Costing strategy, proving this strategy is more accurate than the regular Activity Based Costing methods. Everaert et al. describe various key-drivers of profitability that mainly focus on the difference between targeted costs, which are theoretically estimated costs, and actual costs. Furthermore, Everaert et al. mention initiatives to enhance profitability, where most efforts relate to negotiating and sales, which are not in the scope of this research. However, relevant techniques mentioned are introducing minimum order value policies, maximal discount policies, which indicate the discount a company can offer to a customer in exchange for improvements for the company, optimizing delivery routes, and improving capacity planning. Everaert et al. used simulations to evaluate different scenarios. We could translate these initiatives into analytics to provide actionable insights using the output of a CTS analysis.

In general, there is not much research available on analytics that could provide useful insights based on the output of a CTS analysis. Possibilities for analytics can rely on cost buckets used in a CTS analysis. So, FMCG companies must have a complete set of cost buckets. Often these buckets are valued using Activity-Based Costing (Turney, 1992), which assigns resources to activities (Kaplan and Anderson, 2003). For example, Kolarovszki et al. (2016) determine the CTS based on the number of visits, distance from the customer, and the frequency of contacts. Freeman et al. (2000) take a more comprehensive approach by distinguishing many activities related to sales, marketing, and physical distribution. The key in all these different approaches is to assign costs to activities and thereby determine the CTS of an order (line). The mentioned approaches differ in each case and appear tailored to a specific industry or environment. Of course, enriching the output of a CTS analysis with general data such as customer or product data allows for more possibilities for providing insights through analytics.

To make features more insightful, companies can focus on a group of customers or products that is of importance to them, which requires a type of segmentation. Segmentation might not offer direct insights, but it can support other features. Traditionally, a market is segmented based on attributes such as geography, channel, or demographics (Bolton and Tarasi, 2006). Before this, Bonoma and Shapiro (1984) have already made this concrete by stating five segmentation attributes to use in a nested hierarchy, which are demographics, operating variables, purchasing approach, situational factors, and personal characteristics. Companies can also apply nested hierarchies for products, using different attributes (Davis, 2010). By using nested hierarchies, companies can visualize data on different levels, but this requires higher data availability. Kolarovszki et al. (2016) describe a three-dimensional model based on a customer's relationship value, development potential, and CTS, focusing on customers that are relevant to a company. In the end, segmentation can assign labels to customers or products. Guerreiro et al. (2008) give an example that describes a pyramid model based on the profitability of customers. This model contains the smallest group with the most profitable customers at the "platinum" level. The "gold" level, "iron" level, and "lead" level include larger groups of customers with lower profitability. However, using this segmentation is difficult, as the largest group of customers represents the most potential savings. However, combining such groupings with features can provide a user with valuable insights.

In conclusion, the literature shows analytics related to estimation methods, methods for predicting the CTS, minimum order value policies, maximal discount policies, and various other analytics. Furthermore, it showed the importance of cost allocations and the usefulness of integrating segmentation approaches.

2.4 Cost drivers

Historically, selecting variables is a difficult task. Many people experience difficulties in identifying cost drivers (Pohlen and La Londe, 1994). This section focuses on finding potential variables that could drive a high cost-to-serve per volume-unit because these variables are of importance when creating advanced analytics. The review in this section supports answering the Research question 1.1 Which descriptive, diagnostic, predictive, and prescriptive features can visualize the output of a cost-to-serve analysis?, serving as input for the assessment of the required data in Section 3.1.2. Furthermore, this literature review revealed cost buckets that are supplementary to the cost buckets applied in the research company presented in Chapter 1, which we use in Section 3.2.2. Finally, we also use the variables for the root cause analysis in Section 3.3.4.

Stapleton et al. (2004) provide an overview of cost drivers for logistics and marketing. Focusing on the logistics activities, Stapleton et al. distinguish thee activities:

- Order management activities driven by the number of orders received
- Warehousing and shipping activities driven by the number, size, or weight of units shipped
- Customer service activities driven by the number of returns and complaints.

The last activity is not included in logistics costs in other literature but is very relevant for this research that we perform in the Global Customer Service team of an FMCG company. Varila et al. (2007) show the importance of using multiple cost drivers in combination. Again, mentioning the number and weight of units as well as the number of orders. Guerreiro et al. (2008) include previously mentioned distribution, warehousing, and order management costs. Furthermore, Guerreiro et al. use the time and type of visits to customers by salespeople and promoters as cost drivers, arguing that visits or calls for logistics purposes are potential cost drivers.

Up to this point, we saw individual cost drivers for distribution, warehousing, order management, customer service, and activities of representatives. Baykasoğlu and Kaplanoğlu (2008) introduce a more detailed split for overhead costs that includes many drivers such as distance, number of vehicles, number of drivers, amount of freight, and the area used. Baykasoğlu and Kaplanoğlu show that all these factors drive overhead costs, introducing multiple drivers for a single cost bucket. Focusing on distribution, Bokor (2010) extends the work of Stapleton et al. (2004), Varila et al. (2007), and Guerreiro et al. (2008) by including the distance and the number of distribution-related trips as cost drivers leading to multiple drivers for the customer delivery cost bucket.

More recently, van Niekerk and Bean (2019) used various before mentioned cost drivers and applied them in a CTS framework. A new cost driver and a new category are introduced by van Niekerk and Bean (2019), considering picking activities as relevant cost drivers. When picking requires splitting up a full pallet, this generates more costs due to increased picking times. The new category is the primary distribution, which considers internal distribution between a production location and a final depot. Finally, Kaçan (2019) introduces three new dimensions by stating that the origin of a shipment, the utilization of resources, and the supplier used can also drive transportation costs.

In conclusion, the reviewed literature showed that cost drivers in logistics often explain a specific cost bucket or activity. There are various general uses of drivers, but the exact selection of drivers appears to differ depending on the situation, for example, when having to choose between the number, size, or weight of units shipped (Stapleton et al., 2004).

2.5 Variable selection

In this section, we evaluate methods that can reduce the number of variables to fit a predictive model. Thereby, we support the answering of Research question 1.2 How can fast-moving consumer goods companies find a root-cause for a high cost-to-serve per volume-unit based on data used in the cost-to-serve analysis?, by presenting approaches to reduce the number of variables and ensure variables are suitable for model fitting. We include these methods in Section 3.3.3, where we design a process that selects variables and fits models.

In this research, a solution must apply to many different situations. Therefore, we focused on a general approach to cope with different settings, which may result in a large set of variables, which is not beneficial for fitting models. By selecting a subset of variables, we reduce the complexity to decrease the run-time of model-fitting algorithms and yield better model predictions (Andersen and Bro, 2010). Kuhn et al. (2008) state that it is not beneficial when a variable only has very few distinct values. Such variables are not necessary to consider when fitting a model, as these variables do not influence the predicted value or can cause erratic behavior. Thus, including such variables, named near zero-variance predictors, will only increase run-time. Furthermore, Kuhn et al. explain that multicollinearity negatively influences the performance of many models. Multicollinearity occurs when there is a high correlation between predicting variables. So, it can be beneficial to remove such variables. We do not consider linear relations with the predicted variable, as these relations are what is looked for when fitting a model. The Pearson correlation coefficient can determine a linear relation between variables referred to as X and Y:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

A value below 0.1 indicates a negligible relationship where a value above 0.9 indicates a very strong relationship (Schober et al., 2018). More interpretations of values are shown in Table 2.2.

Table 2.2: Example of a Conventional Approach to Interpreting a Correlation Coefficient (Schober et al., 2018)

Correlation coefficient	Interpretation
0.00-0.10	Negligible correlation
0.10-0.39	Weak correlation
0.40-0.69	Moderate correlation
0.70-0.89	Strong correlation
0.90-1.00	Very strong correlation

Table 2.2 provides an overview of interpretations of the coefficient that FMCG companies can use to determine a threshold. A final set of variables may not contain any correlation coefficients above the threshold. Furthermore, Kuhn (2019) warns to avoid linear dependencies. Linear dependencies occur when two combined variables determine the value of a third variable. For example, when the sum of two variables always equals the value of another variable. So, to reduce the number of variables, FMCG companies must avoid the presence of linear dependencies, highly correlated variables, and near-zero variables in the final set of variables. In this research, this could result in a different set of variables for each data set. Nevertheless, we do not consider variables that are not potential cost drivers or strongly correlated to variables present in the final variable set created for a data set.

Section 2.6 includes an explanation of Random Forests. We can apply specific techniques to reduce the number of variables in a random forest model. Kuhn (2019) describes the use of Recursive Feature Elimination. This technique starts by fitting a model with all predictors. Then, in each iteration, the top predictors remain in the model, the model is fit again, and the algorithm evaluates the performance. Figure 2.3 shows the algorithm.

Algorithm 1: Recursive feature elimination

Tune/train the model on the training set using all predictors
Calculate model performance
Calculate variable importance or rankings
for *Each subset size* S_i, $i = 1 \ldots S$ **do**
  Keep the S_i most important variables
  [Optional] Pre–process the data
  Tune/train the model on the training set using S_i predictors
  Calculate model performance
  [Optional] Recalculate the rankings for each predictor
end
Calculate the performance profile over the S_i
Determine the appropriate number of predictors
Use the model corresponding to the optimal S_i

Figure 2.3: The recursive feature elimination algorithm (Kuhn, 2019)

The optional step of preprocessing the data is not required when using preprocessed input variables. Furthermore, Svetnik et al. (2004) show that the optional step for recalculating the rankings for each predictor decreases the performance when it comes to random forests. Alternatively, we can apply Simulated Annealing (Kirkpatrick et al., 1983) by making small changes to the selected subset of predictors. However, this is only beneficial when outcomes differ strongly due to randomness.

In conclusion, FMCG companies may reduce the selection of variables by omitting linear dependencies, highly correlated variables, and near-zero variance predictors to improve the model-fitting performance of predictive models. The removal of highly correlated variables requires the definition of a suitable threshold. Furthermore, Recursive Feature Elimination or Simulated Annealing can reduce the number of variables in Random Forest models.

2.6 Model selection

Here, we present a literature review to support the model fitting process in Section 3.3.3 by supporting the answering of research question 1.2 How can fast-moving consumer goods companies find a root-cause for a high cost-to-serve per volume-unit based on data used in the cost-to-serve analysis?. In the root cause analysis in Section 3.3.4, the goal is to fit a model that accurately predicts the cost-to-serve per volume-unit.

Designed solutions must handle a large amount of data and ensure the applicability to different settings. Therefore, we consider basic machine learning techniques, including regression models and random forests. In this research, we use known independent and dependent variables. So, as we use labeled samples to train a model, supervised learning is applied (Ayodele, 2010). Supervised learning comprises of two learning types, which are classification and regression (GeeksforGeeks, 2020). In the case of this research, the CTS per volume-unit is a continuous dependent variable. According to Zhou (2018), “in classification, the label indicates the class to which the training example belongs; in regression, the label is a real-value response corresponding to the example”. Brownlee (2017) supports this logic when comparing classification and regression techniques. So, the focus of the root cause analysis is on regression.

Having determined that we must select a suitable regression model, FMCG companies may test different types and choose the best fit. A well-known regression type is a relatively simple Linear Regression with multiple independent variables. Sometimes Linear Regression is referred to as Multiple Linear Regression (Freedman, 2009), where more than one variable (X_i) has a linear effect (a_i) on the outcome of the dependent variable (Y):

$$Y = a_1 * X_1 + a_2 * X_2 + a_3 * X_3 ... a_n * X_n + b$$

There might be no linear relations between independent and dependent variables. In such a case, Polynomial Regression might perform better (Seif, 2018). A Polynomial Regression can handle curvature in a way a simple linear regression cannot.

$$Y = a_1 * X_1 + (a_2)^2 * X_2 + (a_3)^3 * X_3 ... (a_n)^n * X_n + b$$

Seif (2018) also presents Lasso and Ridge Regression. These are two regression model types that deal with collinearity. "When two or more predictors measure the same underlying construct or a facet of such construct, they are said to be collinear (Kock, 2015)". So, in such a case, there is a linear relationship between two of the independent variables. These regression models deal with collinearity by applying penalty functions to ensure collinear variables do not both obtain too high coefficients. The loss function, which regression algorithms minimize, incurs the penalty. Below we see an example of a loss function:

$$L = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 + \lambda \sum_{j=1}^{m} \beta_j^2$$

Besides minimizing the squared error between actual and predicted values $\hat{y}_i$ and y_i, the model coefficients β_j are squared, multiplied by a factor λ, and added to the loss function. So, the loss function penalizes high coefficients. The difference between the Lasso and Ridge Regression is that the Lasso variant takes the absolute value of the coefficient (Chakon, 2017). ElasticNet Regression presents a hybrid of Lasso and Ridge regression, combining both ways of penalization. However, we can also avoid correlation and collinearity in the preprocessing phase when variables are selected, as explained in Section 2.5.

Alternatively, we can fit more complex models. An example of such a model is a Random Forest model (Breiman, 2001). The Random Forest technique creates a regression tree. Figure 2.4 shows an example of a regression tree. In such a tree, samples are split based on the value of a certain attribute. As the name random forest indicates, the technique tests random splits and measures the performance.

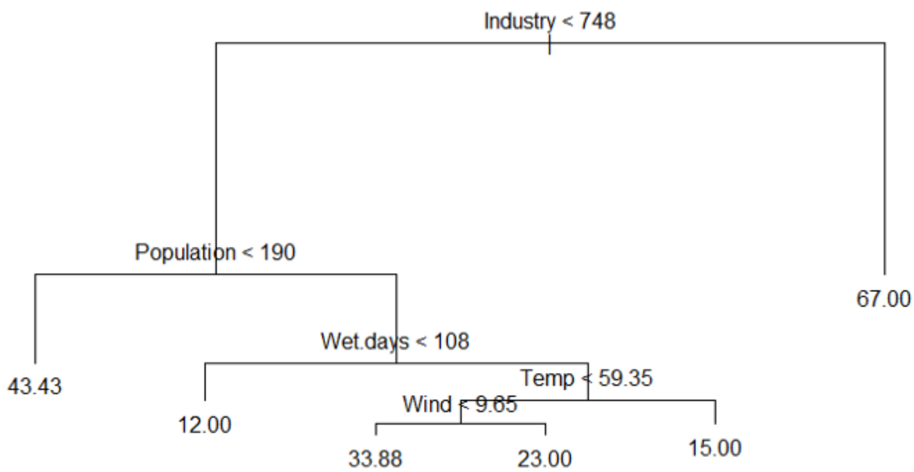


Figure 2.4: An example of a regression tree, or decision tree (Vala, 2019)

An advantage of this model type is that the model is excellent at handling collinearity between variables. A downside of applying a Random Forest technique is that the model fit makes it difficult to understand the variable importance. Liaw et al. (2002) explain that the importance of a variable is how much the prediction error increases when excluding the variable. Due to sometimes complex interactions with other variables, it can be difficult to estimate if a variable has a generally positive or negative influence on the predicted variable.

When fitting more than one model, we must consider how good the model fits are to compare them. The caret package (short for Classification And REgression Training) by Kuhn (2019) includes suitable techniques to perform a regression analysis in R. The caret package uses the Root Mean Squared Error (RMSE), simple R^2 statistic, and the Mean Absolute Error (MAE). The formula for the RMSE is as follows, considering actual values y_i, predicted values $\hat{y}_i$, and sample size n:

$$RMSE = [\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}]^{1/2}$$

The R^2 statistic has several variants. Kvålseth (1985) compares different measures of the R^2 statistic by showing how they differ depending on a model. For this research, we include the squared correlation coefficient between actual and predicted values, as Kvålseth shows that results do not differ much for different models. Section 2.5 explained this correlation coefficient in more detail. Last, the formula for the MAE is as follows:

$$MAE = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n}$$

In conclusion, FMCG companies can use the five basic regression models mentioned in this section and a random forest model to fit a model that predicts a high CTS per volume-unit. To compare models, they can consider several goodness-of-fit measures. For the framework, we include these models and techniques in the model-fitting process in Section 3.3.3.

2.7 Rating tool performance

This section considers the theory supporting the rating of tool performance to answer the Research question 1.4 Which techniques can assess the quality of a tool visualizing the output of a cost-to-serve analysis?. Technology acceptance models evaluate a tool's performance, showing how likely a user is to adopt a certain technology. The adoption of technology is important in this research, as the main goal is to have more continuous users of the cost-to-serve tool after a new implementation. Figure 2.5 shows a recent review of these models by Taherdoost (2018).

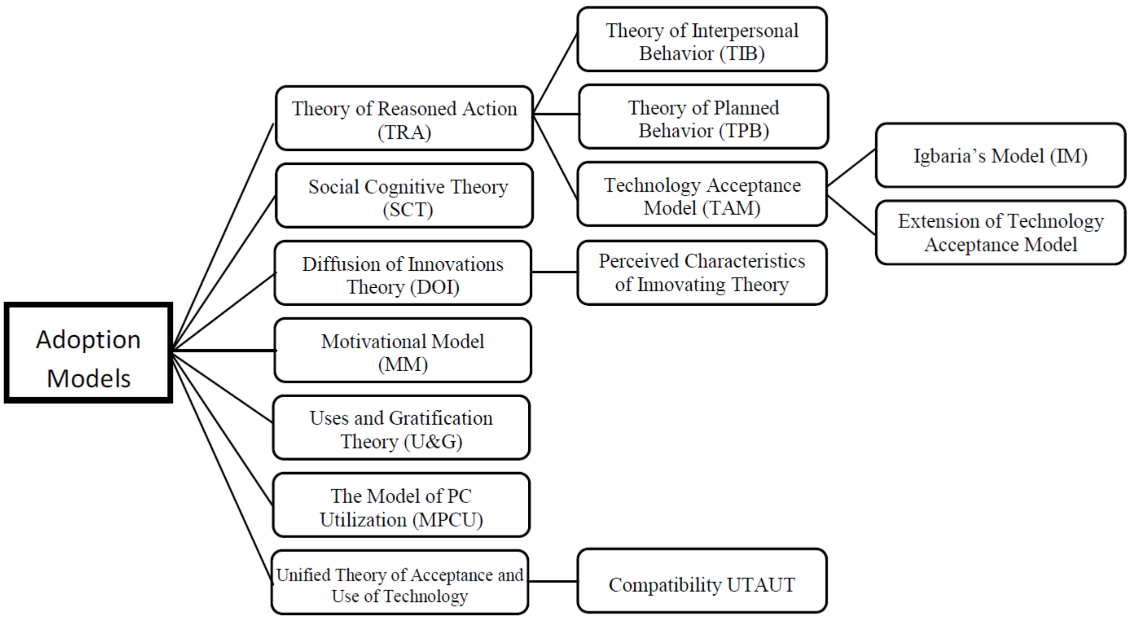


Figure 2.5: An overview of adoption/acceptance models (Taherdoost, 2018)

Four theories are used most commonly in practice. Firstly, the Theory of Reasoned Actions by Fishbein and Ajzen (1977) describes that an Attitude and a Subjective Norm directly influence the behavior of a user. The second theory is the Theory of Planned Behavior by Ajzen et al. (1991), which extends related the Theory of Reasoned Actions by including Perceived Behavioral Control as a variable influencing Behavioral Intention as well as the final Behavior. Madden et al. (1992) state that a positive attitude as a higher perceived behavioral control positively influences the final behavior. Both models are shown in Figure 2.6.

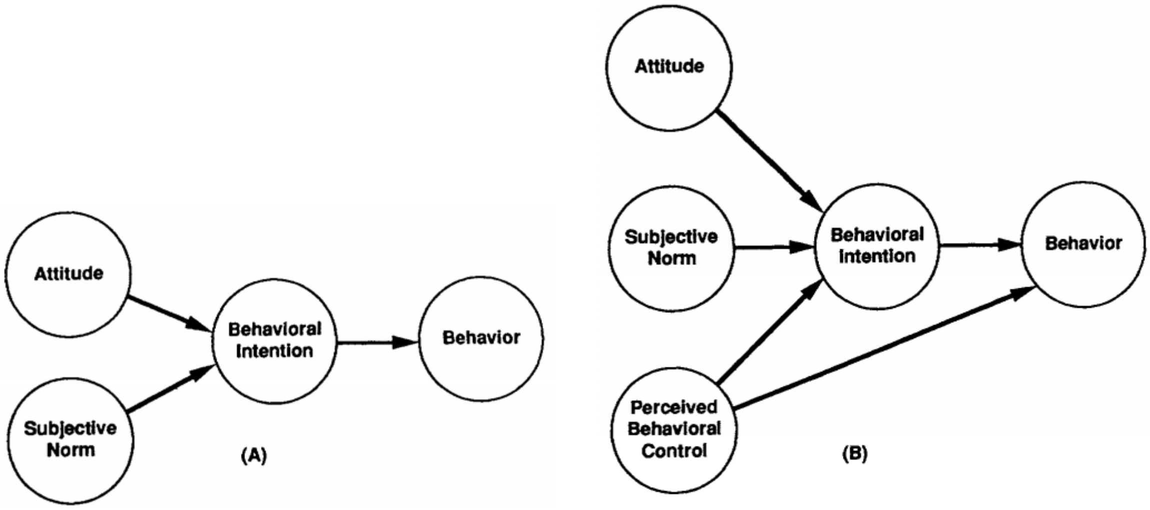


Figure 2.6: The Theory of Reasoned Actions (A) and the Theory of Planned Behavior (B) (Madden et al., 1992)

The third commonly used theory is the Technology Acceptance Model (Davis, 1985), which also extends on the Theory of Reasoned Actions. This theory states that Perceived Usefulness and Perceived Ease of Use influence the Behavioral Response. Figure 2.7 visualizes this model.

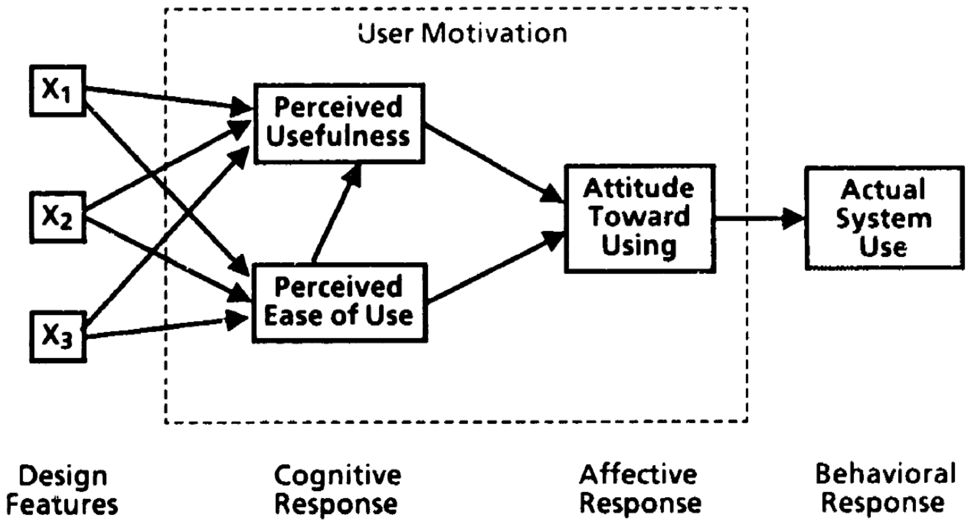


Figure 2.7: The Technology Acceptance Model (Davis, 1985)

This model includes Design Features as external variables and relates strongly to the Usefulness and Simplicity categories found in the problem context in Section 1.3.1. Then, the last widely used model we consider is the Unified Theory of Acceptance and Use of Technology developed by Venkatesh et al. (2003). This theory states that four determinants that are moderated by four variables influence behavior. Figure

2.8 shows the model.

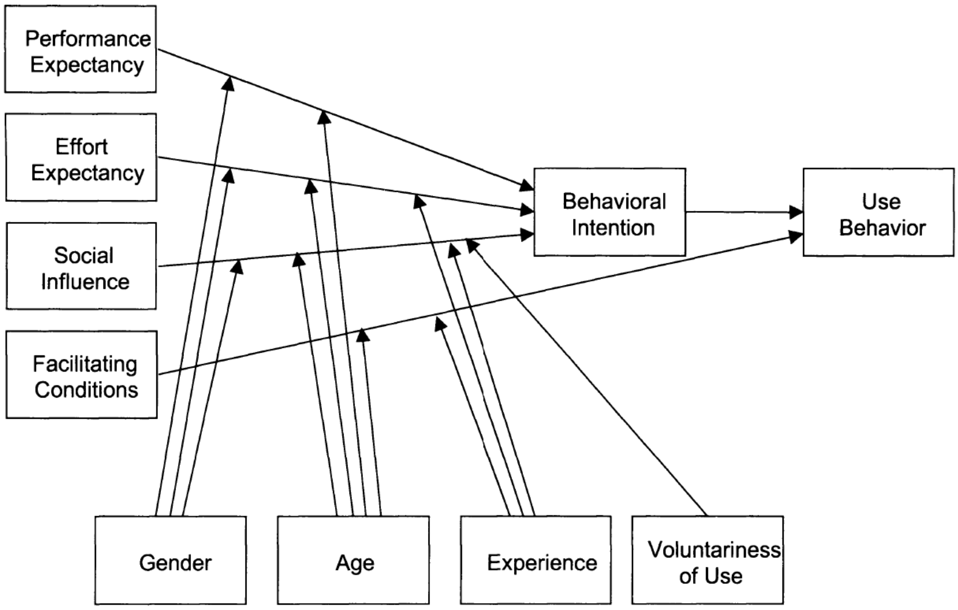


Figure 2.8: The Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003)

This model provides an interesting viewpoint on the importance of moderating variables when it comes to technology acceptance. Furthermore, this model replaces the Perceived Usefulness in the Technology Acceptance Model by the Performance Expectancy, but the terms appear interchangeable. Both concepts relate to the obtainment of cost benefits explained in Section 2.2. Similarly, the Perceived Ease of Use is the Effort Expectancy in the Unified Theory of Acceptance and Use of Technology, implying that technology is easy to use. Additionally, the model includes Social Influence and Facilitating Conditions.

In conclusion, there are several suitable models in the reviewed literature for rating tool performance. Putting this literature review to practice, Section 3.5.2 combines and applies the Technology Acceptance Model and the Unified Theory of Acceptance and Use of Technology.

Chapter 3

Framework development

This chapter presents the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis. Thereby, we answer the following main research question:

> ***1 How can fast-moving consumer goods companies find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a cost-to-serve analysis?***

Figure 3.2 (next page) shows the answer to this question in the form of a framework that combines elements from literature and the research company into a continuous cycle with four different phases that contain several steps. The steps inspired by the research company are generic, as related elements apply to every OpCo regardless of differing characteristics. The framework's intended users are people in FMCG companies responsible for managing and developing a CTS analysis. Depending on the company, this can be from a global or local position related to customer service or logistics. The starting phase also depends on the company. FMCG companies that already perform CTS analyses can commence with any framework phase, while others must start with the Define Objectives of a Solution phase. The DSRM methodology shown in Figure 1.9, which we follow in this thesis, inspired the framework phases. Figure 3.1 visualizes the relationship between the research methodology of this thesis and the framework.

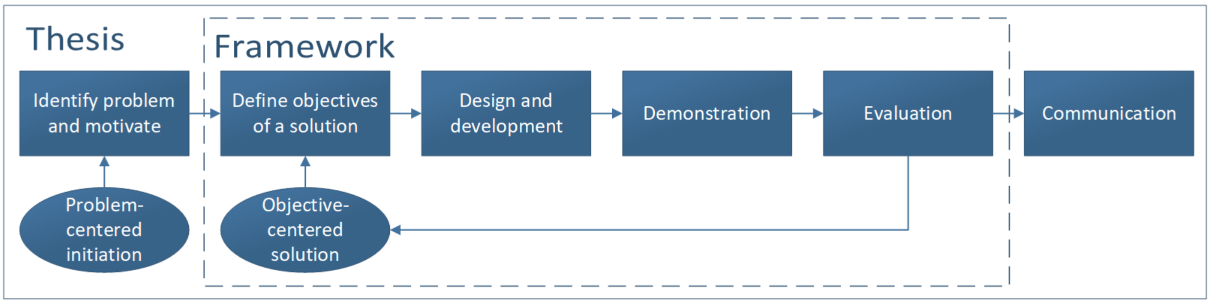


Figure 3.1: The steps of the Design Science Research Methodology (Peffers et al., 2007) followed in this thesis and included in the framework designed in this thesis

This chapter presents each part of the framework in detail. First, Section 3.1 presents the Define Objectives of a Solution phase. Then, Section 3.2 addresses the preparatory steps of the Design and Development phase, which FMCG companies must follow when creating a new tool. Finally, Sections 3.3, 3.4, and 3.5 present the Design and Development, Demonstration, and Evaluation phases. In the Design and Development phase, special attention is paid to descriptive and diagnostic analytics as the research company currently mainly incorporates descriptive analytics, whereas diagnostic analytics concern the level that follows.

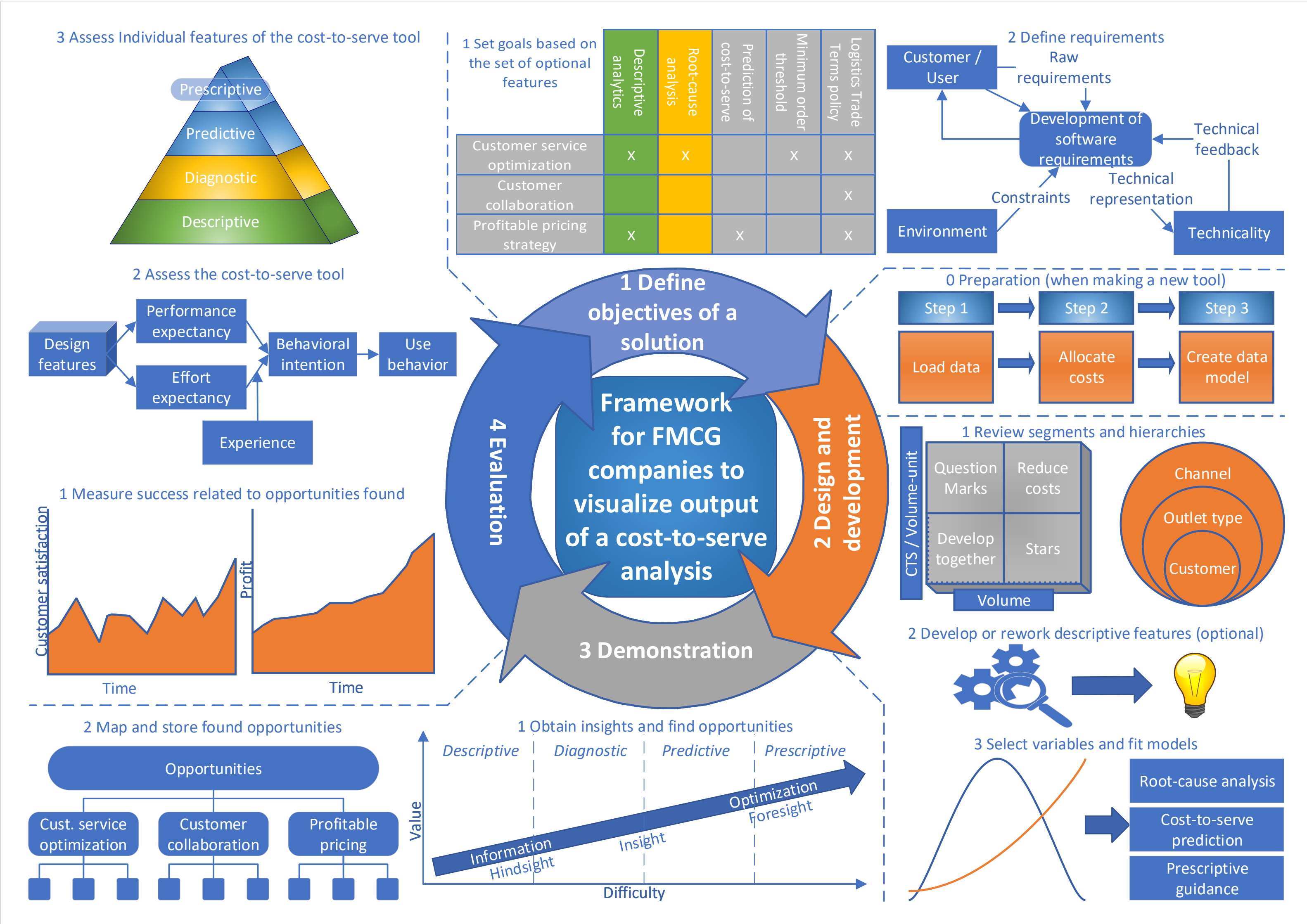


Figure 3.2: The framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

3.1 Phase 1: Define objectives of a solution

Defining the objectives of a solution includes setting goals and defining requirements. Here, we outline how to do this when it comes to visualizing the output of a cost-to-serve analysis, which includes a focus on levels of analytics explained in Section 2.1.1. We answer the following question:

> *1.1 Which descriptive, diagnostic, predictive, and prescriptive features can visualize the output of a cost-to-serve analysis?*

The input for the Define Objectives of a Solution phase comes from the Evaluation phase. FMCG companies can use the success measurements, tool assessment, and feature assessment to make decisions. Figure 3.3 shows the framework's steps in the Define Objectives of a Solution phase. The output of this phase is a set of goals, including features to develop with corresponding requirements. FMCG companies must perform these steps before the Design and Development phase, as shown in the framework in Figure 3.2.

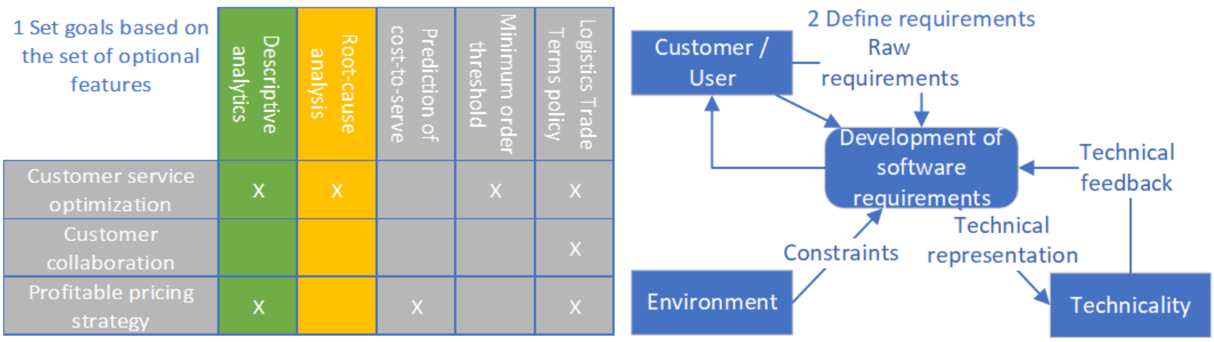


Figure 3.3: The Define objectives of a solution phase of the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

We explain both steps of the Define Objectives of a Solution phase in this section. First, Section 3.1.1 combines previous findings into a set of optional features for FMCG companies to develop as a part of their cost-to-serve tool. FMCG companies can use this set of features when setting a goal. Then, Section 3.1.2 explains how to define requirements for chosen features with a focus on data requirements.

3.1.1 Review options

Here, we create a set of options that we combine with the types of analytics presented in Section 2.1.1 and potential benefits found in Section 2.2. First, we present options from literature. Then, we expand the set of options by including features from the tool of the research company. We assess the perceived value of features from the research company's current tool to determine what to include in the set of options. Then, we discuss which tables to include in the set of options, as a table is the most obvious feature using descriptive analytics. Finally, we combine all potential features into the final set of options that FMCG companies can use to set goals for the Design and Development phase in Section 3.3.

Features from literature

In general, there are not many specific features mentioned in the reviewed literature that appear to provide more than descriptive insights, but Section 2.3 presented a few potential features. Firstly, users should only focus on customers of products that are of importance to them. A segmentation feature provides this functionality (Bonoma and Shapiro (1984), Bolton and Tarasi (2006), Guerreiro et al. (2008), Davis (2010), Kolarovszki et al. (2016)). On itself, it is not clear how segmentation leads to insights. However, a CTS tool can combine segmentations with other features. We consider a basic segmentation as descriptive because it fully depends on user inputs. An advanced segmentation method is diagnostic as it uses techniques that do not rely on users. Secondly, predicting the CTS of a new customer could prove valuable when arranging the pricing for such a customer (Kone and Karwan (2011), Özener et al. (2013), Sun et al. (2015), and Wang et al. (2020)). Furthermore, methods are presented that focus on optimizing third-party-logistics service delivery (Ross et al. (2007) and Everaert et al. (2008)) and improve capacity planning (Everaert et al., 2008), but optimizing delivery routes and improving capacity planning fall outside of the scope defined in Section 1.4.1. The CTS tool might be the tool to highlight opportunities in this area, but solving these problems is the area of expertise of other teams than the Customer Service team. Finally, Everaert et al. (2008) introduce minimum order value and maximum discount policies. Maximum discount policies relate to Logistic Trade Terms, which are discounts given to the customers when they agree to terms, resulting in more efficiency for the supplying company. In the end, the set of options to develop is as follows:

- Descriptive analytics
- A segmentation of customers and products (descriptive or diagnostic)
- Root-cause analysis for a high CTS per volume-unit (diagnostic)
- Estimating the CTS of a new customer or product (predictive)
- Estimating the CTS of a customer or product in the future (predictive)
- Guide in defining a minimum order threshold (prescriptive)
- Guide in defining a discount policy when implementing Logistic Trade Terms (prescriptive)

We included an additional idea for a feature in the set of options inspired by the definition of diagnostic analytics from Section 2.1.1 that states diagnostic analytics answer the question "Why did something happen?". A root-cause analysis for a high CTS per volume-unit can answer this question. Furthermore, estimating the CTS of a new customer or product inspired the idea for predicting the CTS of an existing customer or existing product to show negative trends. The next two sections detail which descriptive features we included based on the research company's current tool.

Features from the research company's tool

We reviewed features from the research company's current tool, intending to include them in the set of optional features because they are generic features that apply to all OpCos. We sent a survey out to assess how often features are used in the current CTS tool of the research company to include these in the set of options for the framework. Thereby we apply an Evaluation step presented in Section 3.5.3 to the current situation of the research company. The survey, filled by sixteen respondents from OpCos and the Global CTS team, assessed the use of features with a linear scale that awards zero points when respondents do not use a feature and three points when they always use a feature. Figure 3.4 shows the outcome of this survey. Appendix B shows all survey responses referred to in this section.

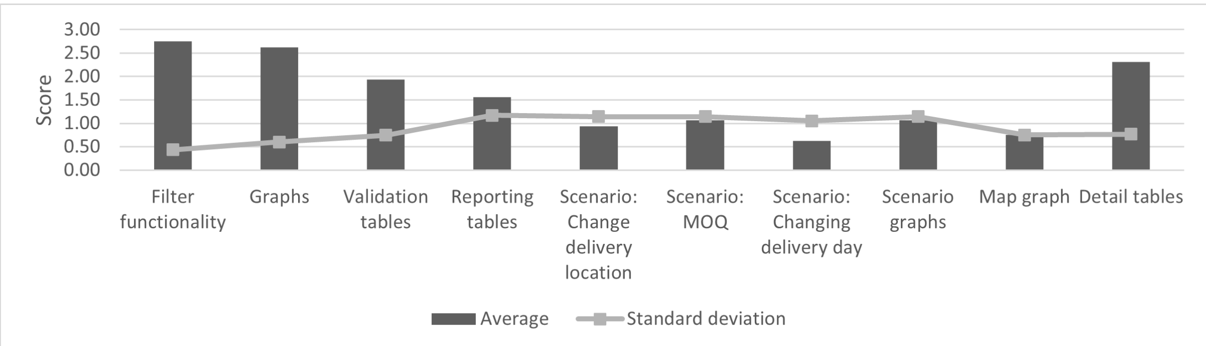


Figure 3.4: The average score and standard deviation for how often a feature is used by the survey respondents (Never 0 - 3 Always)

The filter functionality is always used by most respondents, as indicated by the high score and relatively low standard deviation. The filter functionality is the most appreciated feature but is usually inherent to tooling software. However, as the survey results indicate users value a filter functionality, we included it in the set of options. Users do not use the scenario functionalities often. The high standard deviation means that some respondents do use the functionalities, but the current scenario functionalities appear insufficient. The CTS team indicates that the low usage of scenarios is because they are difficult to understand. So, we did not include them in the set of options in their current state. Users rarely use the map graph, but some research into this showed several users experienced technical issues when viewing the map. Perhaps users would value a normally functioning map graph. So, we included the map graph in the set of options. The Validation, Reporting, and Detail tables are tabs that contain various overlapping tables that users often use. We address which tables we included in the set of options in the next section.

To determine which graphs to include in the set of options, we asked respondents to rate the perceived value concerning different graph visualizations by indicating a rating between one and four stars. Figure 3.5 shows the average and standard deviation of the number of stars awarded.

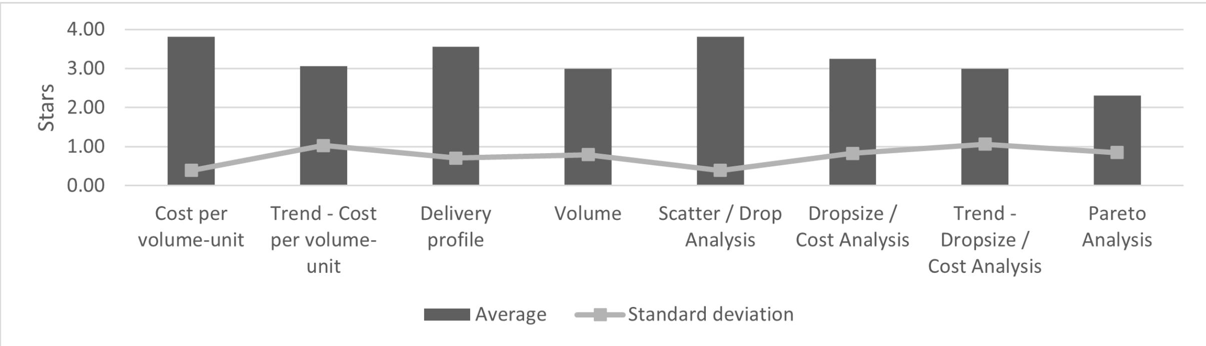


Figure 3.5: The average score and standard deviation of the number of stars awarded to each graph (1 - 4 stars)

Three graphs stand out with a rating above 3.5 and a relatively low standard deviation. We included the Cost per volume-unit, Delivery profile, and Scatter/drop analysis graphs in the set of options. Only the Pareto analysis graph has a relatively low score. So, we did not include that graph, but we did include all others in the set of options as optional to avoid overloading users.

Tables

The research company's current tool has many tables. However, there is much overlap in the descriptive analytics used in them. The goal here is to maximize the availability of descriptive information while minimizing the number of tables required in the tool. All in all, we included the tables in Table 3.1 in the set of options.

Table 3.1: Tables included in the set of optional features for visualizing the output of a cost-to-serve analysis, including what they measure and which dimensions they show

Feature	Measures	Time	Shipments	Customers	Products	Sources	Vehicles
Key numbers	CTS and volumes	x	x	x	x	x	x
Combinations	CTS and volumes	x	x	x	x	x	x
Shipments	None		x				
Customers	None			x			
Products	None				x		
Sources	None					x	
Vehicles	None						x

The research company's tool inspired the tables shown. A key numbers table shows the CTS and volumes along every dimension. Key numbers refer to different volume measurements like weight sold, shipments performed, or customers delivered. The combinations table provides the option to combine two of the listed dimensions. Finally, there are several tables showing attributes and no specific measure. With these seven tables, we extend the information available from tables in the research company's CTS tool, with fewer tables.

Set of options

The previous sections presented the features included in the set of options. We combined these options with the types of analytics presented in Section 2.1.1 and potential benefits found in Section 2.2. Table 3.2 (next page) shows all features and benefits where three factors stand out.

- Many features do not necessarily lead to improvements but merely provide descriptive information or are supplementary to other features.
- Only one feature focuses on Customer collaboration.
- The sole focus of features applying predictive analytics is improvements related to profitable pricing.

Features focused on Customer service optimization can concern Customer collaboration when yielding joint improvements, but this is not their function. Furthermore, as explained in Section 3.5.3, predictive features can serve as stepping stones towards more advanced features that apply prescriptive analytics.

With the set of features in Table 3.2 as options, FMCG companies must decide which features to develop by determining what is feasible considering available resources and what is necessary considering the state of their CTS analysis as evaluated during the three steps of the Evaluation phase. For example, the evaluation from Section 3.5.1 can show an underperforming performance metric, Section 3.5.2 can show that the CTS tool has low expected performance or a high effort, and Section 3.5.3 can show poorly evaluated features. When necessary, Multi-Criteria Decision Making methods presented in Section 2.1.3 can aid in this process. So, taking into account the outcome of the evaluation, FMCG companies can set goals concerning features to rework and new features to develop based on the set of options.

Table 3.2: The set of features that FMCG companies can develop to visualize the output of a CTS analysis showing the types of analytics and benefits

Analytics type	Feature	Customer service optimization	Customer collaboration	Profitable pricing strategy
Descriptive	Filter functionality			
	Graph - Cost per volume-unit (CTS per volume-unit bucket and the volume)	x		x
	Graph - Scatter / Drop analysis (drops per week and the average dropsize)	x		
	Graph - Delivery profile (number of shipments containing a number of pallets)	x		
	Optional: Graph - Dropsize / Cost analysis (dropsize and the CTS per volume-unit)	x		
	Optional: Graph - Trend - Cost per volume-unit	x		x
	Optional: Graph - Volume (volumes of different categories)			
	Optional: Graph - Trend - Dropsize / Cost analysis	x		
	Table - Key numbers	x		x
	Table - Combinations	x		x
	Table - Shipment details			
	Table - Customer details			
	Table - Product details			
	Table - Source details			
	Table - Vehicle details			
	Map (showing volumes and the CTS per volume-unit)	x		
	Customer and product segmentation (basic)			
Diagnostic	Customer and product segmentation (advanced)			
	Root-cause analysis for a high CTS per volume-unit	x		
Predictive	Estimating the CTS of a new customer or product			x
	Estimating the CTS of a customer or product in the future			x
Prescriptive	Guide in defining a minimum order threshold	x		
	Guide in defining a discount policy for Logistic Trade Terms	x	x	x

3.1.2 Define requirements

After FMCG companies define a goal, and it is clear what to develop, they must define requirements. Section 2.1.2 presented a model by Pandey et al., which shows requirements are dependent on users, the environment, and technical aspects. So, the development of features requires clear requirements from a user perspective, and technical aspects play a role in multiple ways, including data availability.

Before-mentioned factors can differ between FMCG companies, but relevant variables might be similar. As applying advanced analytics requires the presence of variables, and we decide to create a feature using advanced analytics for the research company in Section 4.2.1, we address data requirements for developing advanced analytics based on the output of a CTS analysis here. Section 2.4 identified general activities and cost drivers as potential variables. However, it also argued that cost drivers are different in each situation. In this research, solutions apply to FMCG companies. So, we designed a wide range of variables. Table 3.3 shows cost buckets based on a combination of literature from Section 2.4 and the main cost buckets used by the research company as presented in Chapter 1, combined with the activities related to the cost drivers.

Table 3.3: Combination of activities and cost drivers from literature with cost buckets used currently with CTS in the research company

Activity	Cost drivers			
Inter-company Transport	Quantity shipped	Origin	Utilization	Supplier
Delivery to Customer	Quantity shipped	Volume shipped	Weight shipped	Distance
	Number of trips	Origin	Utilization	Supplier
Warehousing	Quantity shipped	Volume shipped	Weight shipped	Picking time
Order Management	Orders received			
Overheads	Distance	Vehicles used	Quantity shipped	Area used
Trade Terms				
Customer service	Returns	Complaints	Visits	Calls

The cost drivers show some ambiguity that we addressed because the variable selection and model fitting process described in Section 3.3.3 requires clear variables. For example, we can evaluate the quantity shipped or the number of trips taken in different ways. Therefore, we designed more specific variables that remain generic and that any FMCG company can use. Appendix C shows a complete overview of all the variables designed to reduce the ambiguity. For Trade Terms, we found no potential cost drivers. This overview of data that FMCG companies should include in their CTS analysis concludes the framework's phase for defining the objectives of a solution. FMCG companies can set goals using the set of options created and define requirements considering users, technical aspects, and their environment.

3.2 Step 2.0: Preparation

This section outlines the preparation steps of the Design and Development phase for visualizing the output of a cost-to-serve analysis. We determined a generalized approach, as steps depend on an FMCG company's view on cost-to-serve and IT infrastructure. FMCG companies must follow these steps when creating a new cost-to-serve tool. When the goal is to improve an existing cost-to-serve tool, they can skip these steps, but because we develop a new tool for the research company in Section 4.4 we address these steps here.

The inputs for the preparation steps of the Design and Development phase are goals and requirements defined in the Define Objectives of a Solution phase. Figure 3.6 shows the preparation steps. The output that follows is a data model containing the output of a cost-to-serve analysis and other relevant information. FMCG companies must complete these steps before they can continue with the other steps of the Design and Development phase, as shown in Figure 3.2.

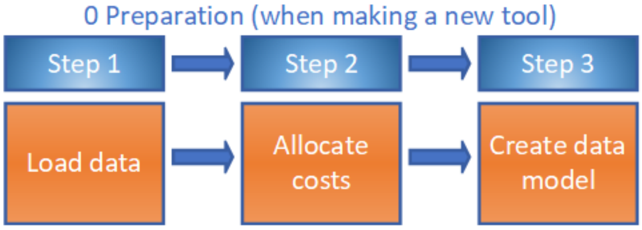


Figure 3.6: The preparation steps of the Design and Development phase from the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

This section outlines each Preparation step. First, Section 3.2.1 explains how FMCG companies can load data. Then, Section 3.2.2 presents different methods for cost allocations and a definition of cost buckets. Finally, Section 3.2.3 shows a data model design for the visualization of output from a cost-to-serve analysis.

3.2.1 Load data

The first preparation step is to load the data required to perform a CTS analysis. We based this process on the process to load data into the research company's CTS tool, as explained in Section 1.1.2. The data includes shipments on an order line level and master data regarding various dimensions. Ultimately, OpCos fill a template with a standard layout, which always has the same fields and data types. Therefore, it makes sense to have an automated process, so the data collection only has to occur once. Preferably modules extract, transform, and load data. Figure 3.7 shows a generic model concerning data loading.

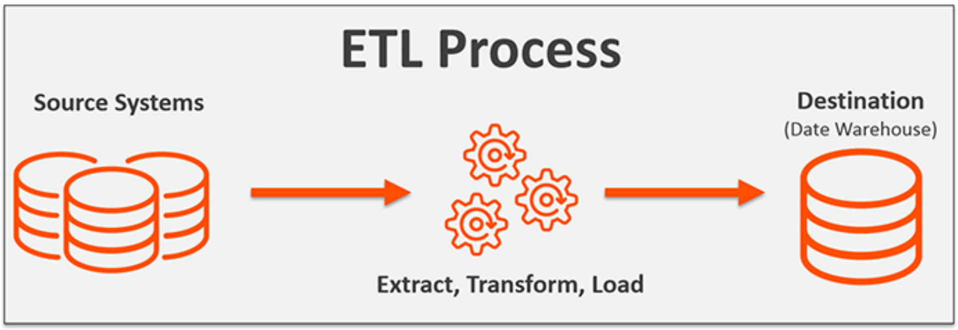


Figure 3.7: The Extract, Transform, and Load process Watts (2017)

This process shows a method to handle the data loading for a CTS analysis in FMCG companies. After extracting data, a module transforms data into the desired format, which is ideally the format of the tables required for the data model presented in Section 3.2.3. Finally, the transformed tables are stored in a data warehouse and loaded into visualization software. Ideally, FMCG companies automate these steps when loading data for a CTS analysis.

3.2.2 Allocate costs

After loading the data for a CTS analysis, FMCG companies must perform cost allocations. Section 2.3 explained that there are different approaches concerning the allocation of cost buckets. However, all approaches allocate costs on an order line level. FMCG companies might prefer CTS methods, which allocate actual costs based on drivers, Activity-Based Costing methods (Freeman et al. (2000), Kaplan and Anderson (2003), Kolarovszki et al. (2016)) or a time-driven variant (Everaert et al., 2008). For example, Section 2.4 mentions a suitable case for a time-driven approach, stating that picking time is a driver of warehouse costs. So, a time-driven activity-based costing approach involving picking time seems applicable to the allocation of warehouse costs. The research company applies a combination of methods, which shows that an FMCG company can have a specific view of the allocation of costs. Table 1.1 showed an example of an allocation as performed by the research company. Nevertheless, we defined generic cost buckets that apply to any FMCG company based on a combination of literature from Section 2.4 and the main cost buckets used by the research company as presented in Chapter 1:

- Inter-company Transport
- Delivery to Customer
- Warehousing
- Order Management
- Overheads
- Trade Terms
- Customer Service
- Out of Scope

These main cost buckets can contain nested cost buckets. For example, the Warehousing cost bucket can consist of sub-buckets for fixed and variable costs, but the definition of sub-buckets is company-specific. The last cost bucket should allocate Out of Scope costs to determine the profitability. In conclusion, FMCG companies fill these buckets in the Transformation step in Figure 3.7 on an order line level.

3.2.3 Create data model

Assuming all required data is available in a data warehouse, the last Preparation step is to create a data model. FMCG companies can use a wide range of software solutions to create a data model. Figure 3.8 shows a generalized data model that applies to FMCG companies. The model used by the research company shown in Figure 1.3 inspired this model.

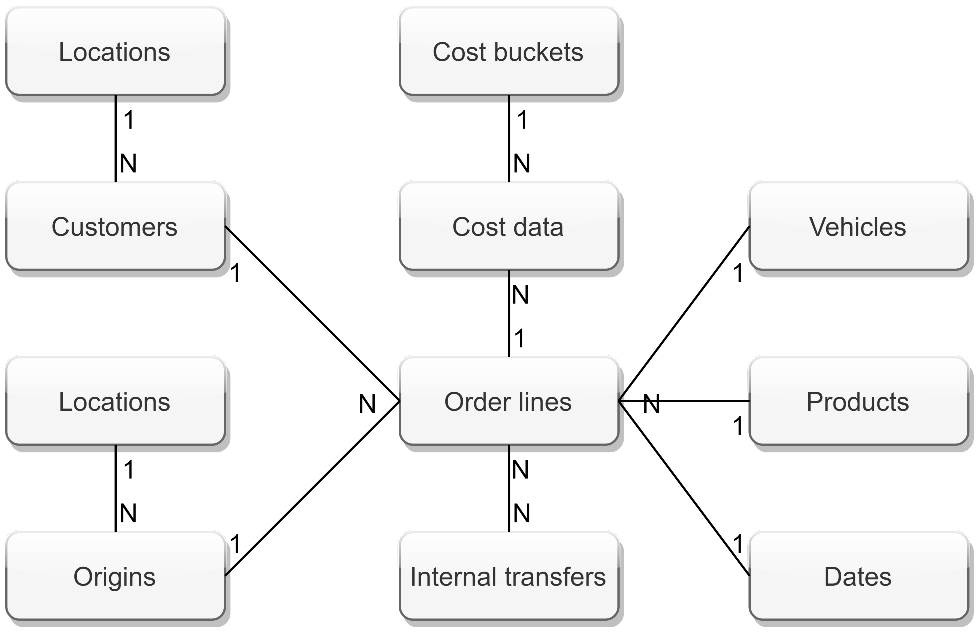


Figure 3.8: The data base scheme of the model for visualizing CTS data

The data model has a snowflake scheme which aims to avoid redundancy (Jensen et al., 2010). Less redundancy means the data requires less storage space. In a snowflake scheme, each dimension has an individual table and contains keys to underlying dimensions. For example, the model shows tables for customer, product, origin, vehicle, and date information. The order lines might be at the center of the data model. However, the cost data has the lowest level of detail, as each order line has multiple allocated cost buckets. As explained in Section 3.2.2, cost buckets can incorporate a hierarchy with main buckets and sub-buckets. Therefore, the storage of the cost bucket names in a separate table is beneficial. Finally, the model includes geographic data concerning customers and sources to determine distances and internal transfers linked to order lines to enable the visualization of routes to customers. FMCG companies can create this data model to visualize the output of a CTS analysis. However, an actual data model likely includes additional supportive tables. The creation of a data model concludes the preparation steps of the Design and Development phase of the framework, providing a basis for the development of a CTS tool.

3.3 Phase 2: Design and development

In this phase, we outline how to rework existing features and develop new features to visualize the output of a cost-to-serve analysis. The focus is mainly on descriptive and diagnostic analytics, as the research company mostly applies descriptive analytics and diagnostic analytics are the next level. All features and analytics mentioned here are generic and applicable to any FMCG company. Consequently, we answer the following question in this section:

> *1.2 How can fast-moving consumer goods companies find a root-cause for a high cost-to-serve per volume-unit based on data used in the cost-to-serve analysis?*

The inputs for the Design and Development phase are goals and requirements defined in the Define Objectives of a Solution phase and the data model from the preparation steps of this phase. Figure 3.9 shows the Design and Development phase, which leads to the development of new features. The features an FMCG company decides to develop influence the necessary steps taken during this phase. The output of this phase is a new or revised cost-to-serve tool. The Demonstration phase follows this phase, as shown in the framework in Figure 3.2.

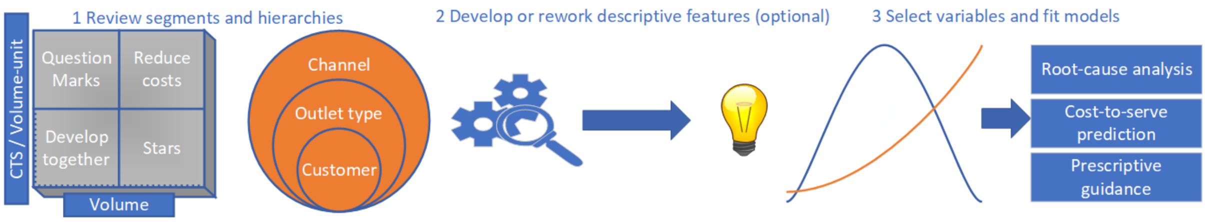


Figure 3.9: The Design an development phase of the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis that focus on developing the tool that is used

First, Section 3.3.1 explains how to incorporate segments and hierarchies. Periodically checking these is important as circumstances change. Then, Section 3.3.2 describes how to review existing features, when this is necessary based on the defined objectives from Section 3.1. Then, Section 3.3.3 explains how to select variables and fit models for the application of diagnostic, predictive, and prescriptive analytics. Finally, Section 3.3.4 develops a generic method find root-causes for a high cost-to-serve per volume-unit that relies on Section 3.3.3. The root-cause analysis is not a mandatory part of the framework, but Section 4.2.1 argues the requirement for a root-cause analysis for the research company as this lies at the foundation of more advanced analytics.

3.3.1 Review segments and hierarchies

Applying hierarchies has a positive effect on the usability of the output of a CTS analysis as mentioned in Section 2.3 (Bonoma and Shapiro (1984), Bolton and Tarasi (2006), Guerreiro et al. (2008), Davis (2010)). A hierarchy means that an entity is a part of a group with multiple entities, which might be a part of another group. FMCG companies can incorporate hierarchies for each dimension. For example, shipping locations belong to customers, that belong to outlet types, that belong to channels. To have a nested hierarchy, as in this example, is beneficial when there are many-to-one relations between groups in different levels of a hierarchy to avoid complicated hierarchies.

Segments imply a grouping of dimensions or groups. For example, segments can help to focus on types of customers that require a similar approach or have other similar attributes. Hierarchies are basic segmentations defined by users. A more advanced segmentation might not rely on user input. An advanced three-dimensional technique is described by Kolarovszki et al. (2016) that attempts to make less tangible factors tangible by expressing customer relationship value and growth potential. If such methods are too advanced for an FMCG company, they can implement a more simple segmentation. We designed a basic segmentation using two factors of the CTS analysis, which are the volume and the CTS per volume-unit. Figure 3.10 shows how to interpret a segmentation using these two variables. In Section 4.4.2, we created this segmentation for the research company.

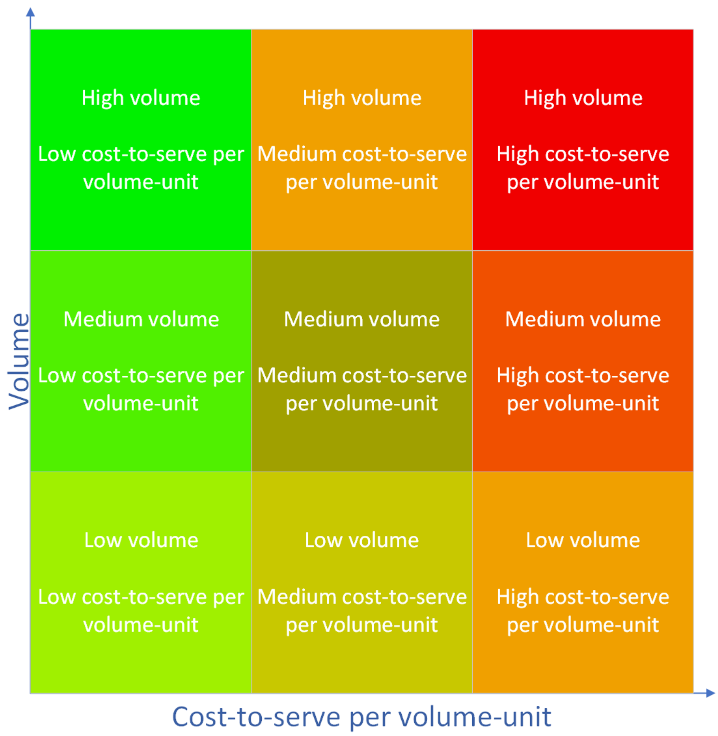


Figure 3.10: A simple segmentation that is proposed based on the total volume and the cost-to-serve per volume-unit (green = a desirable situation, red = an undesirable situation)

The top left corner is the ideal position as it represents a high volume and a low cost per volume-unit. FMCG companies do not want entities in the top right corner due to a high volume and high costs per volume-unit. It makes sense to move an entity from its current position to the left or up, but when an entity moves up one segment, the priority also becomes higher to focus on the CTS per volume unit because entities with more volume can drive up costs more than low-volume entities. In conclusion, FMCG companies should include hierarchies and a form of segmentation in their CTS tool.

3.3.2 Develop or rework descriptive features

Cost-to-serve tools will usually include features using descriptive analytics, as these are the most simplistic features to develop. However, FMCG companies might want to rework these features. Section 1.2 mentioned that the research company wished to review its current solution at the start of the research, which shows an example of an FMCG company wishing to review their current tool. Therefore, we included a step to review existing features based on the Define Objectives of a Solution phase presented in Section 3.1, which FMCG companies can base on the assessment of the state of features from Section 3.5.3. Ideally, users appreciate all features of a CTS tool. However, it is important to reconsider the state of features when this is not the case, which can even result in the removal of certain features. When an FMCG company does not yet use descriptive analytics, the FMCG company likely developed goals related to this in Section 3.1.

3.3.3 Select variables and fit models

More recent studies mentioned in Section 2.3 that apply advanced analytics often use predictive models to make decisions (Kone and Karwan (2011), Sun et al. (2015), Wang et al. (2020)) rather than mathematical models (Ross et al. (2007), Özener et al. (2013)). Predictive models focus on understanding a situation, where mathematical models aim to optimize a certain situation. A wide range of available variables and attributes is beneficial for advanced analytics. Ideally, FMCG companies include all variables presented in Section 3.1.2. Then, they can preprocess variables as presented in Section 2.5 and fit models presented in Section 2.6. Figure 3.11 shows steps we designed from creating the variables to fitting a model. We applied these steps for the research company in Section 4.4.5.

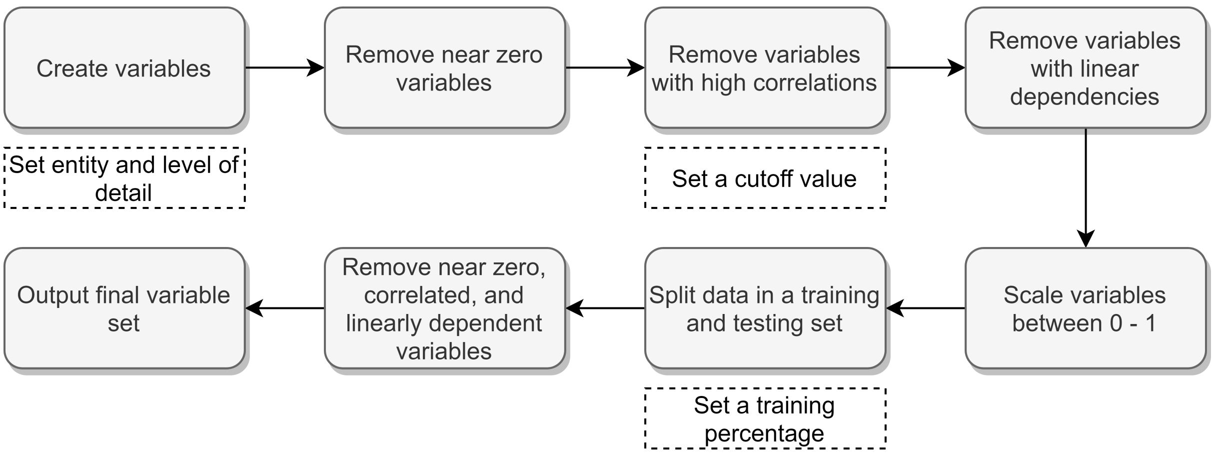


Figure 3.11: The process from variable creation to having a final set of variables including values that users must define

The process starts with the creation of the variables, where users must choose an entity and level of detail. We assume that the variables created do not contain outliers or false inputs. Then, as explained in Section 2.5, FMCG companies should remove near-zero variables, variables that cause high correlations, or variables they can predict with other variables. For avoiding the correlations, users must determine a cutoff value based on Section 2.5. The next step is to scale the variables between zero and one to avoid difficulties when fitting models due to different magnitudes. Also, scaling makes it easier to derive the variable importance from weights in many models. Then, we split the data set into a training and testing set for the validation of models. In this step, users determine a training percentage for the data split. Then, FMCG companies must repeat the second, third, and fourth steps because there is a chance the data splitting led to undesirable values or relations in the training data set. Finally, there is a set of variables for model fitting.

The variable set contains independent variables that predict a dependent variable. For many analytics, this is likely the CTS per volume-unit. Section 2.6 presented models that FMCG companies could fit and methods to evaluate them. For the training of models, they can apply k-fold cross-validation (Allen, 1974). Figure 3.12 shows how k-fold cross-validation works, which is useful for experimenting with different parameter settings.

Figure 3.12: Diagram of k-fold cross-validation with k = 10 (Ashfaque and Iqbal, 2019)

The example shown is for 10-fold validation, which splits the training data into ten partitions. Then, the method trains the model using nine parts and tests on the one singled out for all ten possible combinations. Finally, the error is the average of the ten iterations. Finally, the testing set validates the model-fit resulting from the k-fold validation method, determining the final model performance. In conclusion, we created a process for FMCG companies to select variables and fit models that serves as a basis for many features using advanced analytics. Model fitting concludes the framework steps of the Design and Development phase. Section 3.3.4 presents a feature using diagnostic analytics that is not a mandatory part of the framework.

3.3.4 Root-cause analysis

The application of the Define Objectives of a Solution phase in Section 4.2.1 argues the development of a root-cause analysis approach to show why some entities have a higher CTS per volume-unit than others. Here, we present a generic approach to do this that FMCG companies can decide to apply. We apply the approach presented here in the research company in Section 4.4.5. It is not a mandatory part of the framework but an optional feature that relies on the variable selection and model-fitting process. Section 3.1.2 explained the importance of requirements. Therefore, this approach takes the following user-perspective requirements into account:

- The solution answers the question: “Why does this entity have a higher CTS per volume-unit than another entity?”
- The solution works for any (combination of) attribute(s) used as a key
- The solution shows potential monetary savings
- It is possible to integrate the solution into a CTS tool

Furthermore, we take the following technical requirements into account:

- Data is available on an order line level
- The variables presented in Section 3.1.2 are available on an order line level
- Software is available that facilitates all computations and the fitting of models

Assuming users take all requirements and constraints into account. The next step is to apply the selection of variables and fitting of models as explained in Sections 2.5 and 2.4. The output of this process is a model with the best fit based on the RMSE, R^2, and MAE, as explained in Section 2.6. For example, for all attempted model fits, we scale criteria between zero and 1, inversing the R^2 as a higher value indicates a better model fit. Then, the model with the lowest value for the sum of all criteria has a superior fit. Including the fitted model, the root-cause analysis presented here requires the following input data:

- The model
- The data set with variables used by the model
- The unscaled variables, with the volume included

Figure 3.13 (next page) shows the algorithm designed to compare each entity to all other entities based on this input data. Each entity obtains at most three comparable entities based on three thresholds:

- We exclusively compare entities to other entities with a lower CTS per volume-unit because the end-goal is to find improvements.
- We compare entities to other entities that have an absolute difference concerning the sold volume of no more than ten percent to safeguard the comparison of similar entities and to speed up the algorithm.
- We consider entities as similar concerning a variable if the absolute difference between the value for the two is less than a threshold based on the variable importance in the fitted model.

So, more significant attribute values must be closer to each other to classify as comparable. When the algorithm finishes, each entity has at most three similar entities, so there are alternatives when it is not clear what business changes to make based on the most similar entity. The output shows actual values of both entities, a description on which values the entities differ, and the potential savings corrected for the MAE of the fitted model to avoid an overestimation of potential savings.

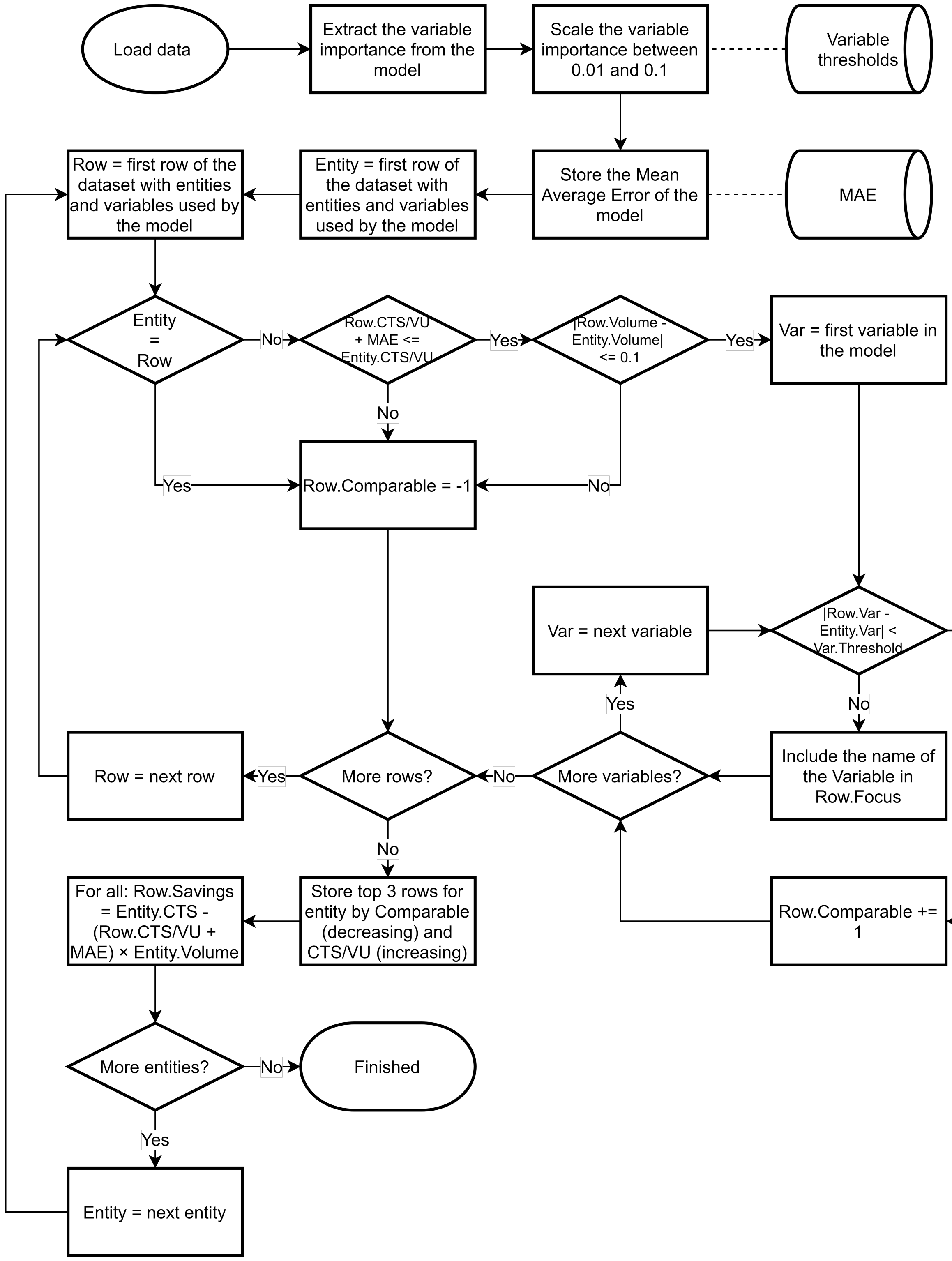


Figure 3.13: A flowchart showing the root-cause analysis (round = start/stop, square = process, triangle = decision)

3.4 Phase 3: Demonstration

The demonstration of the cost-to-serve tool created during development occurs by obtaining insights that lead to opportunities. This section outlines how to obtain insights using the visualized output of a cost-to-serve analysis. Furthermore, it explains how insights can lead to opportunities, which FMCG companies can map for future use. We answer the following question:

> *1.3 How can fast-moving consumer goods companies map business improvements obtained with insights from a cost-to-serve analysis?*

The input for the Demonstration phase is the new or revised cost-to-serve tool created during the Design and Development phase. Figure 3.14 shows the framework's steps in the Demonstration phase. The output of this phase is an overview of opportunities found with the new, or revised, cost-to-serve tool of an FMCG company. These steps are followed by the Evaluation phase, as shown in the framework in Figure 3.2.

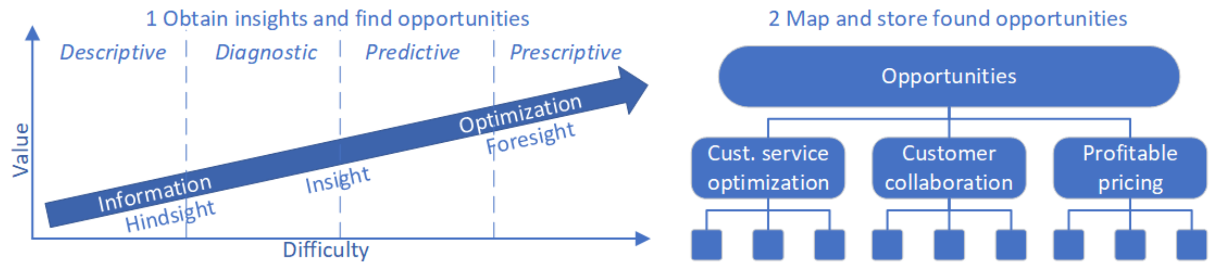


Figure 3.14: The Demonstration phase of the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

First, Section 3.4.1 reflects on how the research company obtains insights. Then, Section 3.4.2 outlines how to map opportunities and learn from implementations by leveraging the past. We created a categorization of opportunities based on opportunities found in the research company.

3.4.1 Obtain insights and find opportunities

After the development process described in Section 3.3 is completed, FMCG companies will distribute the tool among users to obtain insights. Figure 3.15 shows that making a business decision requires human input, except when using prescriptive analytics.

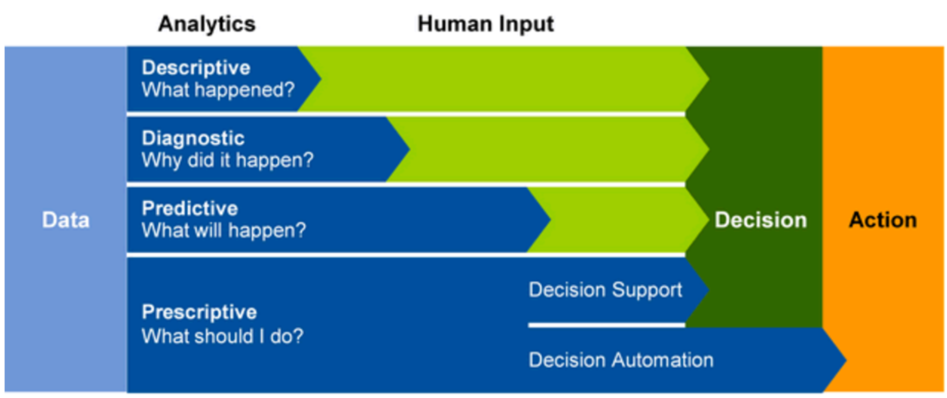


Figure 3.15: The extent to which the decision-making process is left to human judgement as opposed to fully automated decisions (Bosshard, 2017)

When a tool that visualizes the output of a CTS analysis only exists of features using prescriptive analytics, it always provides clear-cut opportunities, but this is not likely. Therefore, users must learn to combine a CTS tool's features to find opportunities. Therefore, FMCG companies should consider training users on how to obtain insights using a CTS tool. For example, Section 1.1.2 explained that the research company trains users of the CTS tool through workshops and supports them with setting up action plans.

3.4.2 Map opportunities

Insights lead to opportunities, which accompany certain benefits. Section 2.2 determined three main types of benefits that are obtainable from a CTS analysis. The main benefits relate to customer service optimization, customer collaboration, and a profitable pricing strategy. To be able to judge the effect of releasing a new CTS tool, FMCG companies require a method to map the resulting opportunities found by users. By combining these benefits with research into CTS implementations in the past of the research company, we created a generic opportunity typology for mapping opportunities. We analyzed historical data of implementations performed by the research company to discover different opportunity types related to the main benefits. Table 3.4 shows the discovered categories and types with their definitions.

Table 3.4: The definitions of the categories and types used during the collections of potential opportunities found in past CTS implementations

Category	Type	Definition
Customer service optimization	General	Focus on optimizing the supply chain
	Data quality	Improve the data quality
	Network design	Change the transportation network
	RPM	Improve Returnable Products Management
	Transport optimization	Optimize transportation activities
	Warehouse optimization	Optimize warehousing activities
Customer collaboration	General	Focus on collaboration with customers
	Logistic trade terms	Logistics-focused customer discounts
	OTC	Improve the Order To Cash process
	SLAs	Review Service Level Agreements
Profitable pricing strategy	General	Change prices or discard products
	Commercial trade terms	Commerce-focused customer discounts
	Product portfolio	Change the product portfolio

The research into opportunity types confirmed the three main benefits as all types were easily nested under one of the benefits. Appendix D shows an overview of the complete typology, including sub-type definitions. The typology was verified with the CTS team and embedded in a simple process for continuously evaluating potential opportunities. In other FMCG companies, some types or sub-types might not apply or be missing. Then, they can easily remove or add types to improve the typology for their situation, but we assume that we found most opportunity types for FMCG companies due to the different nature of the OpCos involved. Figure 3.16 shows the process for mapping opportunities.

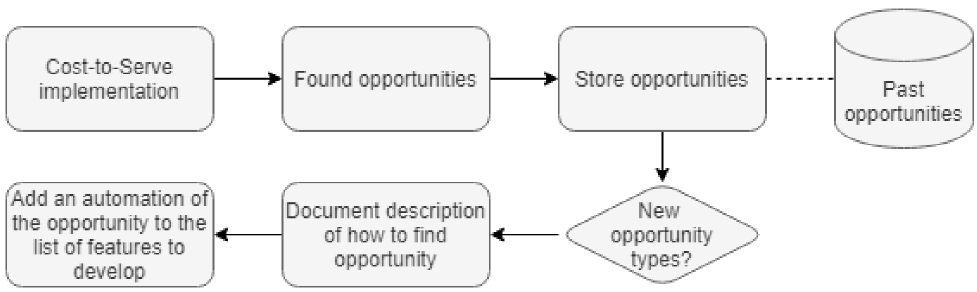


Figure 3.16: The steps to map opportunities that have been found

In the first row, the process shows how to map opportunities using the typology, which FMCG companies can extend if necessary. After mapping opportunities, FMCG companies can assess whether they found new opportunities. If so, documenting how they found them can potentially pave the way for repeating the process in another CTS implementation. Additionally, there is the potential to automate the opportunity and add it to the list of optional features to develop shown in Section 3.1.1. In conclusion, FMCG companies must train users based on the user input required to make business decisions and apply the process to map opportunities in the Demonstration phase.

3.5 Phase 4: Evaluation

This section outlines how FMCG companies can evaluate the performance of their process and tool to visualize the output of a cost-to-serve analysis, which involves performance measurements as well as tool assessments. In this section, we answer the following question:

> *1.4 Which techniques can assess the quality of a tool visualizing the output of a cost-to-serve analysis?*

The inputs for the Evaluation phase rely on the Design and Development phase and Demonstration phase. FMCG companies require the created overview of found opportunities from the Demonstration phase to measure success, and they require the new or revised cost-to-serve tool created in the Design and Development phase to assess it. Figure 3.17 shows the Evaluation phase of the framework. The outputs of this phase are a success measurement, tool evaluation, and feature evaluation. This phase is the last of the framework, but the Define Objectives of a Solution phase follows this phase, as shown in Figure 3.2. The output of the Evaluation phase serves as the input for the Define Objectives of a Solution phase.

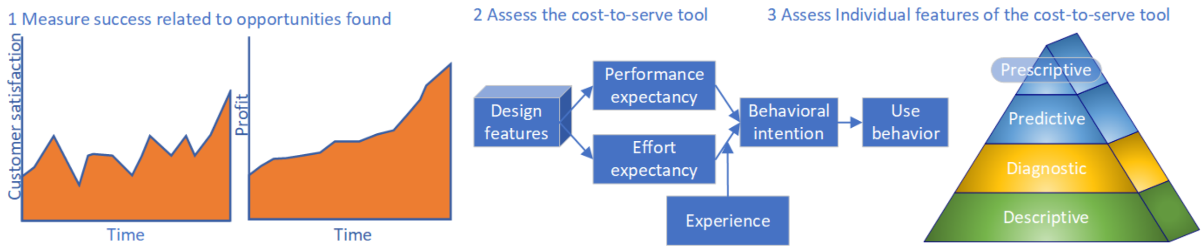


Figure 3.17: The Evaluation phase of the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

This section explains the Evaluation steps. First, Section 3.5.1 outlines how to measure the success of the cost-to-serve analysis as a whole. The execution of this step can depend on the company or team KPIs. Then, Section 3.5.2 presents performance variables for a tool assessment and explains how these variables can measure the performance of a cost-to-serve tool. Finally, Section 3.5.3 presents a method to assess the state of features and analytics applied in a cost-to-serve tool.

3.5.1 Measure success

"What gets measured gets done" might be the most famous saying of performance measurement (Behn, 2003). Hence, it is important to measure success. Here, we present KPIs that FMCG companies can measure to determine the success of CTS implementations. The benefits of CTS, related to potential measurements of success, found in Section 2.2 are Customer service optimization, Customer collaboration, and a Profitable pricing strategy.

Section 2.2 showed that profitability and customer satisfaction measure success related to these benefits. When measuring these KPIs, FMCG companies must carefully scope the area of measurement. A

CTS analysis focuses on logistics costs. So, FMCG companies should consider these costs when measuring profitability or savings. Furthermore, estimating the effect of business changes is complicated as many factors related to a CTS analysis influence customer behavior. Therefore, putting down an exact number when it comes to achieved savings is hard. A way to deal with this is by taking the estimated savings of an opportunity and evaluate savings by indicating if the actual savings were less, about the same, or more than the initial estimate. Comparably, FMCG companies should measure changes in customer satisfaction of customers influenced by changes.

3.5.2 Assess tool performance

Besides measuring success through KPIs, it is important to know how users perceive a CTS tool. Section 2.7 presented several models to measure the performance of a tool. Performance and Effort Expectancy, which are stated as independent variables by Venkatesh et al. (2003), are of critical importance for the acceptance and use of a CTS tool at any FMCG company. Figure 3.18 shows a model to assess a CTS tool's performance, which includes the effect of Design Features mentioned by Davis (1985), and a relevant moderating variable.

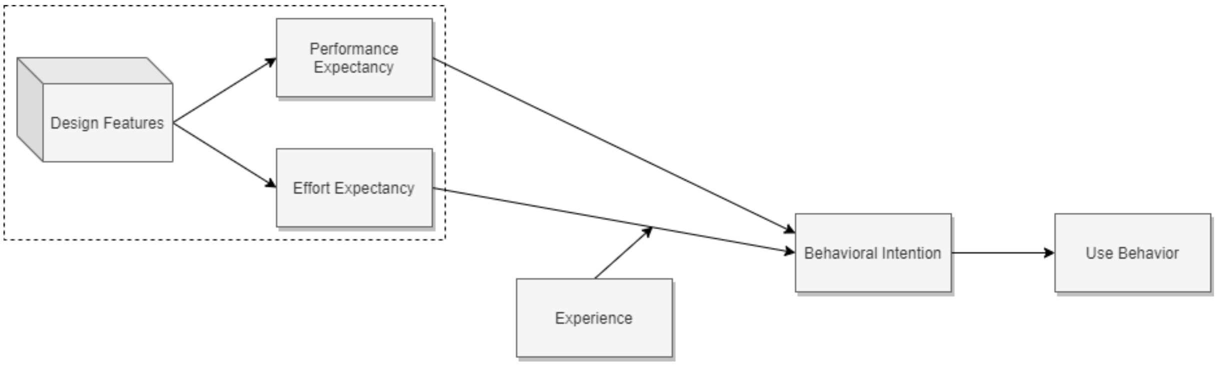


Figure 3.18: The adapted Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003) that is used to measure tool performance in this research

Design features influence Performance Expectancy and Effort Expectancy. The other independent variables shown in Figure 2.8 do not rely on tool design. Furthermore, Experience influences the relation between Effort Expectancy and Behavioral Intention as a moderating variable. We do not consider Gender and Age, assuming employees in a specified position have similar competencies. Gender and Age might influence these competencies, but it is not necessary to measure their influence in this research. Furthermore, the Voluntariness of Use variable exclusively impacts the relationship between Social Influence and Behavioral Intention. Therefore, this variable is also out of consideration.

FMCG companies can use the model in Figure 3.18 to assess their CTS tool's performance. Venkatesh et al. (2003) describe many statements from literary sources to propose to subjects when assessing the expected acceptance and use of new technology. For a CTS tool, the following statements can assess Performance Expectancy:

1. Using the system improves my job performance.
2. Using the system makes it easier to do my job.
3. Using the system significantly increases the quality of output on my job.

The following statements can assess the Effort Expectancy:

1. Learning to operate the system was easy for me.
2. Using the system involves too much time doing mechanical operations (e.g., data input).
3. I find it easy to get the system to do what I want it to do.

Scoring these statements to the extent a subject agrees with them provides a performance measure of the tool. FMCG companies can use a scale from 0 (completely disagree) to 5 (completely agree) to measure performance by assessing answers to the questions concerning the Performance Expectancy, the Effort Expectancy, and the view on the tool as a whole. Furthermore, FMCG companies can apply the method to a specific part of the CTS tool, such as recently added features.

3.5.3 Assess features

Besides measuring the general performance of the CTS analysis and tool, it is important to be aware of the insights currently provided for users. Section 2.1.1 explains the concept of descriptive, diagnostic, predictive, and prescriptive analytics. FMCG companies can assess if they apply enough analytics of a certain type and whether the focus should be on the next level of analytics. Figure 3.17 indicates this by the green color of descriptive insights and yellow color of diagnostic insights as an example, which resembles a case where features applying descriptive analytics are covered, and there is a requirement for diagnostic analytics. Besides creating a mapping of features used in a CTS tool, FMCG companies can involve users to validate assumptions. Then, users must answer two questions:

1. How much value do features in the current tool hold?
2. How much value will the addition of new features add?

When posing these questions, FMCG companies must know what types of analytics features use. Visualizing the output of a CTS analysis starts with features applying descriptive analytics that plot values against various dimensions. The next level, diagnostic analytics, relies on a complete set of descriptive analytics to find causes for events or trends, leading users to areas that require attention. Then, predictive analytics can focus on these areas to estimate the severity of what might be a trend. Finally, prescriptive analytics can drive business decisions directly. Section 3.4.1 explained the extent to which different analytics can drive decisions. Higher-level analytics require less human input in the decision-making process, but features applying higher-level insights require more comprehensive analytics and depend on lower levels of analytics. Therefore, FMCG companies should start by making sure lower-level analytics are complete and valid before moving to more complicated analytics.

The assessment of features concludes the Evaluation phase of the framework, but the framework presented in this chapter incorporates a continuous cycle. Combining the feature assessment with the performance of the CTS analysis and tool, FMCG companies can define the objectives of a solution correctly in the next framework phase. So, after the feature assessment, the process restarts with another Define Objectives of a Solution phase, as presented in Section 3.1.

Chapter 4

Case study

This chapter presents a case study that applies the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis to the research company. Consequently, we answer the following main research question:

> ***2 Can the research company improve the use of the output of cost-to-serve analyses by applying the framework designed in this research to create a new tool?***

The purpose of this chapter is to present how elements of the framework work in practice and outline learnings related to the different phases. Figure 4.1 shows that one section focuses on the current tool, and the others are devoted to the development of a new tool to visualize the output of a cost-to-serve analysis.

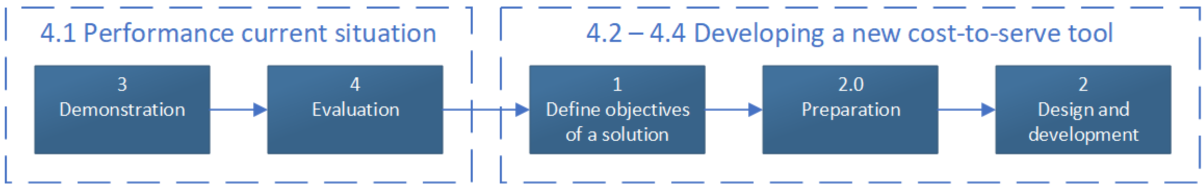


Figure 4.1: The relation between the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis and the sections of the case study in this chapter

To ensure we could apply each framework phase to the research company, we applied some phases to the current situation and others for the development of a new cost-to-serve tool. Each section relates to one or more phases of the framework presented in Chapter 3 and concludes with a comparison between the case study and the intended steps presented in the framework. First, Section 4.1 focuses on the current tool by creating a structure to map opportunities based on framework Phase 3 Demonstration, and evaluating the current performance based on phase 4 Evaluation. Then, Sections 4.2, 4.3, and 4.4 present the application of Phases 1 Define objectives of a solution, 2.0 Preparation (when making a new tool), and 2 Design and development to develop a new cost-to-serve tool. In the end, we applied every step of each framework phase to the research company.

4.1 Performance current situation

This section focuses on the current tool used by the research company. We evaluated the performance of the research company by applying framework Phase 3 Demonstration and 4 Evaluation in the context of the company. Thereby we answer the following question:

First, Section 4.1.1 presents an application of the proposed process for mapping opportunities from Section 3.4.2. We created a MS Power BI report to track opportunities found in cost-to-serve implementations. Then, Section 4.1.2 relies on the Power BI report for measuring success and evaluates the performance as well as the contents of the current tool. Finally, Section 4.1.3 presents differences between the case study and the framework's Demonstration and Evaluation phases.

4.1.1 Phase 3: Demonstration

This section applies the Demonstration phase presented in Section 3.4. The input for this phase is the current CTS tool of the research company, which was in use for several years. The output of this phase is an overview of opportunities found with that tool.

We combined the typology from Section 3.4.2 with data regarding CTS implementations of the research company. This data contained general information, the costs in scope, and data of mapped opportunities that include an OpCo name and a description. Furthermore, many opportunities included an estimation of the effort and a potential annual impact. We assigned a category, type, and sub-type to each opportunity, allowing for the calculation of the relative impact on the costs in the scope of each opportunity type. Finally, we created a report and dashboard in Power BI, providing a good means for analysis of the opportunities and serving as a useful database for future implementations. This report is not a CTS tool, but it is a supplementary solution for the CTS team. We chose Power BI because the research company selected this software as their company-wide reporting solution, which we explain in Section 4.2.1. Figure 4.2 visualizes this process. Appendix E shows screenshots of the report and an example of insights regarding the types of opportunities provided by the report.

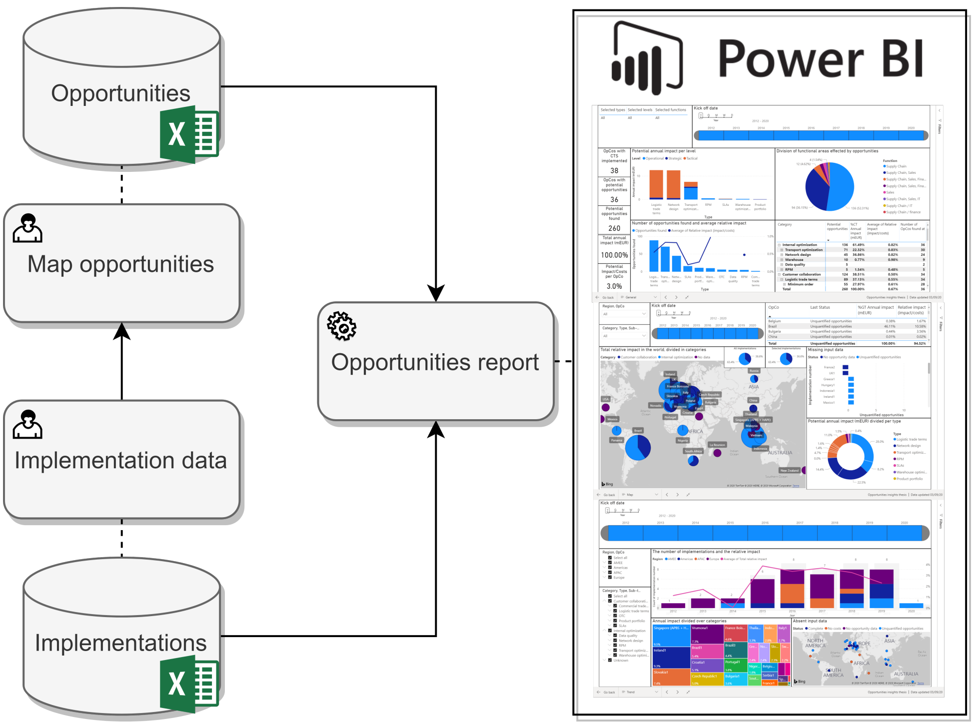


Figure 4.2: The process of for mapping opportunities and implementation data and visualizing this in a Power BI report

The report shows which types of opportunities were most effective in the past. The research company found Logistics Trade Terms opportunities the most over the years. Furthermore, they often found Customer service optimization related opportunities involving Transport Optimization and Network Design.

Drilling down shows that opportunity types under the Logistics Trade Terms type mostly concern implementing a minimum order threshold. The report also includes a page showing a worldwide view of where past CTS implementations took place and the types of opportunities found in those implementations. Furthermore, it clarifies which CTS implementations have missing input data.

4.1.2 Phase 4: Evaluation

This section applies the Evaluation phase presented in Section 3.5. The inputs for this phase are the report with past opportunities from the Demonstration phase and the current CTS tool of the research company. This section starts with measuring the success of cost-to-serve implementations of the research company, which we connected to the report with past opportunities from Section 4.1.1. Then, we assess the research company's current tool by applying the model designed in Section 3.5.2, and we assess the contents of the tool. So, the outputs of this phase are a success measurement for the research company's past CTS analyses and an evaluation of the research company's current CTS tool and its features.

Measure success

The first step in the Evaluation phase is to measure success based on Section 3.5.1. Section 1.1.1 explained that the research company measures success through potential savings found, which we measured in the Power BI report created to map opportunities in Section 4.1.1. This main KPI of the report showed an average cost reduction of 3% per OpCo resulting from CTS implementations. However, a large amount of input data includes unquantified opportunities in the past. 114 of the 259 opportunities do not have a potential annual impact related to them. Furthermore, a page visualizes the success of implementations over the years, showing a minor decline in the average impact made per OpCo. The decline supports the research motivation in Section 1.2.1, which stated in an increase in the effort to find opportunities. However, the decline might be inaccurate due to unquantified opportunities.

The combination of the success measurement with the mapping of opportunities allowed a view of the success of different opportunity types. An analysis of opportunity types led to two insights. First, the most impactful opportunity sub-type is the Minimum Order Threshold sub-type, accounting for 18.5% of all the found opportunities. Second, the Customer Service Optimization category has a high average impact. On average, opportunities yielded a reduction of 0.87% of costs in scope, but they include many subtypes with varying performance. So, it is hard to determine specific actions based on this insight. Appendix E shows all report pages and the analysis of opportunity types.

Assess tool performance

The next step in the framework is the assessment of the current tool's performance. We shared a survey with a group of people that are currently working with the CTS tool in QlikView containing the statements formulated in Section 3.5.2. The group contained people from OpCos and the Global CTS team. Besides the mentioned questions, we assessed the experience of respondents by asking how many years of experience they had with the CTS tool. Furthermore, we asked respondents to state their functional position. Figure 4.3 summarizes the experience of the sixteen respondents.

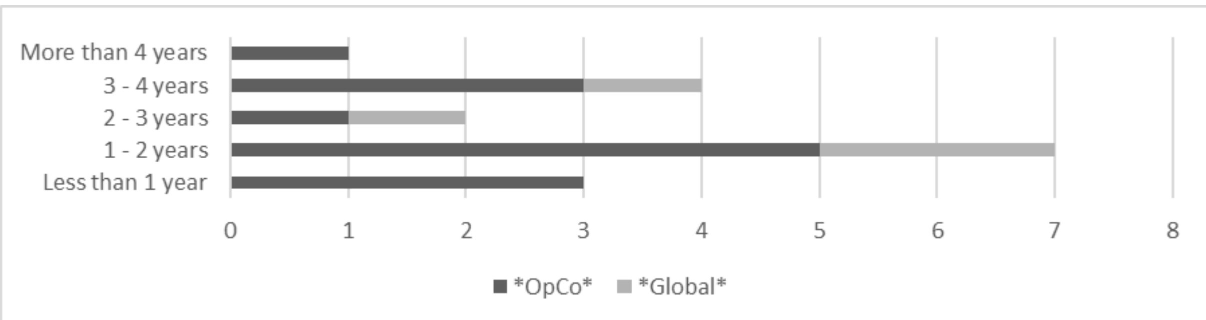


Figure 4.3: A summary of the experience working with the CTS tool and functional position of survey respondents

The Figure shows a total of seventeen respondents. The responses of one respondent count for both groups because that respondent was a part of the global CTS team but currently works with the CTS tool at an OpCo. In general, most respondents have at least a year of experience with the CTS tool, indicating the respondents are quite familiar with it. For the assessment of the current CTS tool, we transformed answers into points depending on the agreeance of respondents. The points range between zero points, given when a respondent completely disagrees, to five points for a respondent complete agreeing. Figure 4.4 shows the average and standard deviation of the scores for each statement. Appendix F shows an overview of all answers.

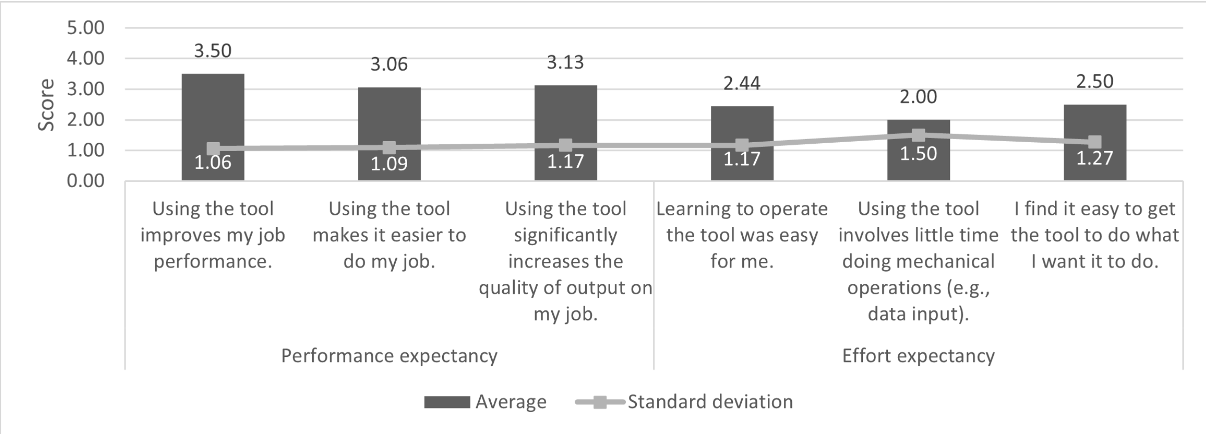


Figure 4.4: The average score and standard deviation for each statement in the assessment of the performance of the current tool (Completely disagree 0 - 5 Completely agree)

On average, the current tool received a score of 2.77 out of 5. The standard deviation is low for answers to each question, which means the agreeance between respondents is high. The CTS tool scored higher concerning the expected performance than the expected effort. In general, respondents somewhat agree with the CTS tool having a positive influence on their performance, and they somewhat disagree that using the CTS tool requires little effort. Figure 4.5 shows the different scores of functional groups and respondents with varying years of experience.

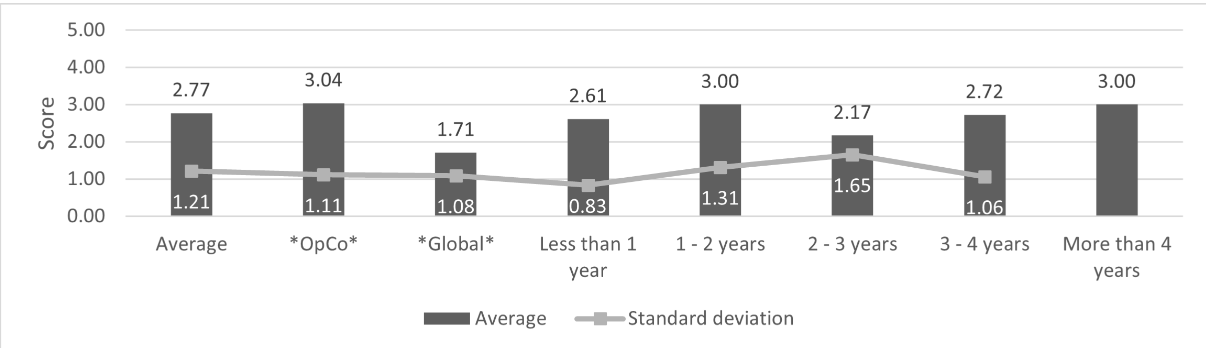


Figure 4.5: The average score and standard deviation for each group of respondents regarding the assessment of the performance of the current tool (Completely disagree 0 - 5 Completely agree)

The difference is high between respondents working with the tool at an OpCo and from a global position. On average, respondents from OpCos give the current CTS tool a score of 3.04, while the global respondents give a 1.71. It seems the Global CTS team is more critical of the CTS tool than people in OpCos. We observed no trends related to respondents with a different number of years of experience. Only one employee with more than four years of experience responded to the survey, which is why the standard deviation for that group equals zero. In general, the assessment method provides understandable results, showing the current tool obtained a mediocre score in the assessment.

Assess features

After we measured the success and assessed the performance of the CTS tool, we assessed the features in the tool based on the framework step presented in Section 3.5.3. To get a view of the insights provided by the current tool, we mapped the tool's features. The mapping of the tool includes all features, their type, what they measure, and along which dimensions they can provide insights. Appendix G shows the complete mapping of the tool. This framework step includes a survey of how users perceive different features, which we already presented in Section 3.1.1 to determine optional features to include in the framework's set of options.

Most features in the CTS tool apply descriptive analytics, showing users what happened, as stated in Section 2.1.1. Users can obtain diagnostic insights, which show why something has happened, but that requires them to conduct a one-time root-cause analysis. The CTS tool also contains three different scenarios that users can run, which use predictive analytics, but the tool uses descriptive analytics to show scenario outcomes. Furthermore, Section 3.1.1 showed that users do not use the scenarios often because they are difficult to understand. In the rest of this section, we focus on the descriptive analytics of the CTS tool. Figure 4.6 visualizes what these features measure.

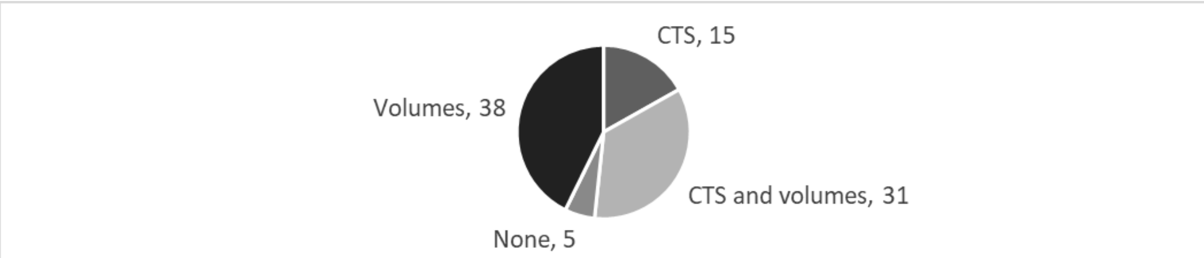


Figure 4.6: The division of what features in the current tool are measuring

Out of the 89 features using descriptive analytics, most features measure volumes, and many features measure a combination of the CTS and volume KPIs. So, the overall focus is on showing volume KPIs, which is contrary to the savings-focused measurement of success used by the research company in Section 4.1.2. Figure 4.7 takes a look at the dimensions used in different features of the tool.

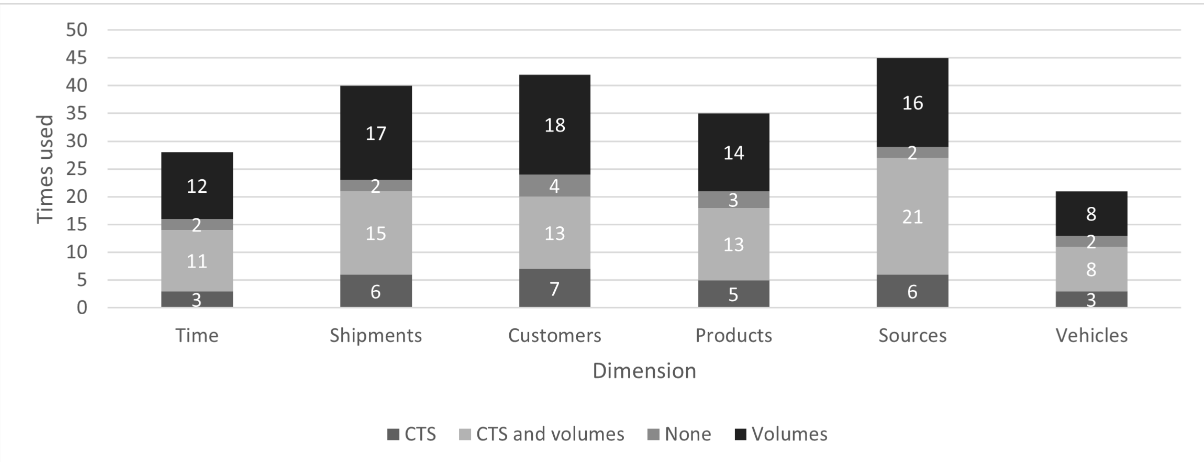


Figure 4.7: The division of dimensions the current tool uses in features

Sources are used as a dimension most frequently, while one would expect to see a focus on customers in the CTS tool. Additionally, many features in the current CTS tool use overlapping dimensions. In many cases, the CTS tool visualizes data differently, but in some cases, the same table occurs more than once. Table 4.1 shows some features with overlapping dimensions.

Table 4.1: Features with similar dimensions that occur more than 5 times in the tool and their measurements

Dimensions used	CTS	CTS and volumes	No measures	Volume	Total
All except CTS buckets and KPIs	1	4	1	2	**8**
Sources		5		3	**8**
Products			1	4	**5**
Shipments	2	1		2	**5**

The most striking example of overlap is that five features show the CTS and volumes along a source dimension, indicating the degree of overlap in the tool. In conclusion, this assessment shows the focus on descriptive analytics and the overlap between features. Additionally, a new CTS tool could pay more attention to showing customer dimensions and measuring the CTS.

4.1.3 Differences with the framework

Compared to the typology and process of the Demonstration phase presented in Section 3.4, there are three differences. Firstly, Internal Optimization refers to Customer Service Optimization, and Profitable Pricing Strategy opportunities are nested under Customer Collaboration because a structure agreed on with the research company was applied when we created the report. Secondly, we did not consider the automation of opportunities for the research company because it was the first time the research company had an overview of past opportunities, which showed a lot of different types. Perhaps the research company could start automating impactful opportunities using predictive or prescriptive analytics. Thirdly, we made an addition to the framework by creating a Power BI report to visualize past opportunities. This report enables the CTS team to view past opportunities in a user-friendly environment and to obtain insights regarding their past performance.

There are three differences compared to the framework step on how to measure success from the Evaluation phase. Section 3.5.1 presented profit and customer satisfaction as two indicators of success, but the research company considers savings rather than profit. It was not possible to include profitability due to unavailable data and company policies. However, the savings compared to scoped costs provided a suitable measure of financial success. Regarding customer satisfaction, the research company does measure this, but it was also not possible to include this data. Secondly, we did not include actual values because the research company only had data concerning estimated savings. Lastly, we measured success on a level of opportunity types due to the connection with the opportunities mapping, allowing the CTS team to determine which types require more or less attention. The last difference again presents an extension of the Evaluation phase, presented in Section 3.5.

We assessed the tool as intended in the framework step from Section 3.5.2, and we performed most of the assessment of features used as intended, considering we assessed the value of different the CTS tool's features in Section 3.1.1 to create the set of options for the framework. However, we determined the potential value of future analytics in Section 4.2.1 after we set goals concerning what to develop in the new tool.

4.2 Phase 1: Define objectives of a solution

Here, we focus on the goals and requirements for a new tool for the research company. We evaluate the performance by applying framework Phase 1 Define objectives of a solution in the context of the research company. Thereby we answer the following question:

> *2.2 Which descriptive, diagnostic, predictive, and prescriptive features should the research company use to visualize the output of a cost-to-serve analysis?*

The inputs for this phase are the success measurement, tool assessment, and feature assessment from the Evaluation phase. This section starts with Section 4.2.1 presenting a review of the options presented in Section 3.1.1. We set a goal stating what to develop as a part of this research. Then, Section 4.2.2 outlines requirements and restrictions concerning the selected features. So, the output of this phase is a set of goals, including features to develop with corresponding requirements. Finally, Section 4.2.3 presents differences compared to the framework steps as presented in Section 3.1.

4.2.1 Define goals

As stated in Section 1.2.1, the research company was considering new tooling software around the start of this research. In the end, they chose Power BI as their standard company-wide reporting solution. Therefore, creating a new tool was in line with company developments.

We made decisions on what descriptive analytics to develop in collaboration with the CTS team. The selected features included all descriptive analytics from the set of options shown in Table 3.2, except for the features marked as optional. Furthermore, we could improve the current descriptive analytics by focusing more on showing relevant data, which we can do by highlighting outliers, showing thresholds, or in other ways. Additionally, we could develop new features relying on descriptive analytics when an opportunity presented itself. So, the first goals defined are as follows:

- Create a new tool in MS Power BI
- Include all non-optional descriptive analytics from Section 3.1.1
- Improve descriptive analytics where possible
- Test new ideas for descriptive analytics that come up during the development

In the next part of this section, we define a goal for a feature using more advanced analytics, based on options from Section 3.1.1. Then, we validate this goal by surveying users.

Review options

The assessment in Section 4.1.2 showed that the focus of the old tool was on descriptive analytics. Therefore, the research company should start exploring higher-level analytics. The CTS team wished to explore the following features from Section 3.1.1:

- Root-cause analysis for a high CTS per volume-unit (diagnostic)
- Estimating the CTS of a new customer or product (predictive)
- Guide in defining a minimum order threshold (prescriptive)
- Guide in defining a discount policy when implementing Logistic Trade Terms (prescriptive)

We did not consider the development of an advanced segmentation method because there was another project in the research company focusing on this. Furthermore, we did not consider an estimation method for the CTS of a customer or product in the future, as the CTS team did not require such a feature.

There was a good fit between potential analytics and requirements of the research company. Firstly, a root-cause analysis for a high CTS per volume-unit is beneficial when understanding the output of a CTS analysis. We could automate the process where users find the reason for a high CTS per volume-unit. Furthermore, Section 4.1.2 showed types of opportunities found differ per implementation. A root-cause analysis would reduce the time spent on finding opportunities. Secondly, estimating the CTS of a new customer is beneficial when acquiring new customers. Also, estimating the CTS of new products can ensure a profitable product launch. However, the research company often relies on single-use solutions for such processes. Thirdly, a feature to guide in defining a minimum order threshold would present users with options for implementing a threshold for a customer or group of customers. Section 4.1.2 showed that many OpCos found opportunities related to minimum order thresholds in the past, which shows that they would appreciate the guidance. However, there is little knowledge within the research company to provide such guidance. Lastly, the feature to guide in a maximum discount strategy relates to Logistic Trade Terms, which are a part of the opportunity typology shown in Section 3.4.2. Such a feature guides users in finding logistic optimizations where the customer shares in a part of the savings as an incentive to agree to certain terms. A minimum order threshold is an example of a Logistic Trade Term that the research company applied often, but the research company has other Logistic Trade Terms as well. In conclusion, all four potential features could benefit the research company.

Considering the four potential features to develop, previously mentioned arguments dictate that the research company should prioritize the development of an analysis that shows root-causes of a high CTS per volume-unit. Section 1.3.2 defined that the current tool contains mostly descriptive analytics, and Section 3.5.3 argued it makes sense to start developing features that take the CTS tool to the next step concerning more advanced analytics because developing higher-level analytics requires the previous level to be solid. For example, you only want to fix something broken. Therefore, you must first know what broke. In the case of the research company, this means focusing on diagnostic analytics. So, the best feature to focus on is the root-cause analysis for a high CTS per volume-unit, leaving the last goal is as follows:

- Create a feature that finds root-causes for a high CTS per volume-unit

Validate decision

We surveyed users to test the assumption that a root-cause analysis should be the next addition to the current tool of the research company. The survey was conducted among the same group of respondents, as shown in Section 4.1.2, and adhering to the model presented in Section 3.5.2, we assessed the performance and effort expectancy. Respondents ranked the four analytics based on their expected performance and effort expectancy. We posed the following statements to respondents:

1. Please rank the following features according to how much it would improve your job performance.
2. Please rank the following features according to how much it would reduce the time spent in your job.

We awarded the feature with the highest rank four points, and the lowest-ranked one received zero points. Figure 4.8 shows the average and the standard deviation of the assessments for each potential feature concerning improved job performance and time reduction. Appendix H presents an overview of all answers given.

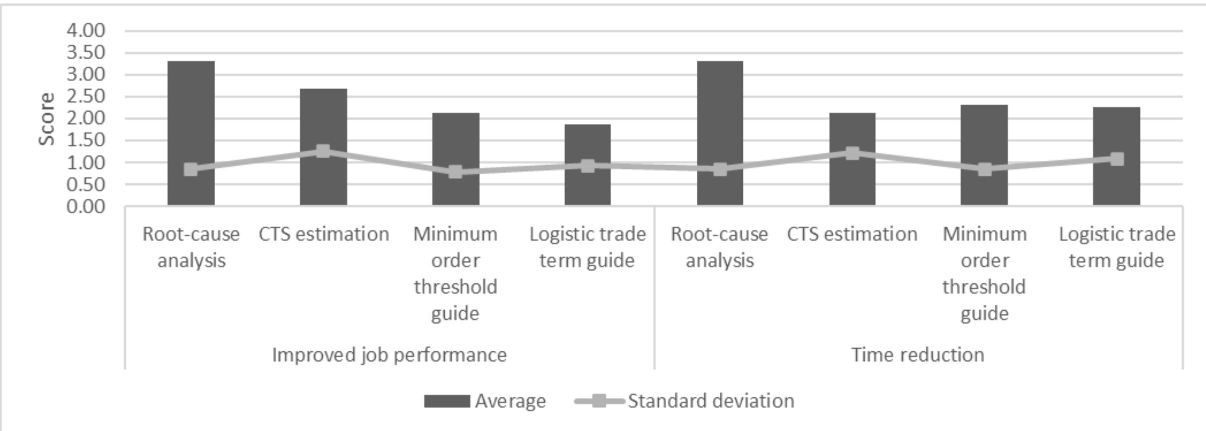


Figure 4.8: The average score and standard deviation for each analytic in the assessment of the expected performance of future analytics (0 - 4)

The standard deviation is low for the average answers meaning the agreeance between respondents is high. Furthermore, the results appear to confirm the requirement of the development of a module that finds root-causes of a high CTS per volume-unit. The root-cause analysis scores the highest in both categories. When it comes to predictive and prescriptive analytics, the results are less clear. Concerning improved job performance, respondents prefer predictive analytics rather than prescriptive analytics, but concerning the expected time reduction, there is no clear preference. We also moderated the experience and functional groups of respondents. To obtain a more detailed view, Figure 4.9 shows the results for different groups of respondents.

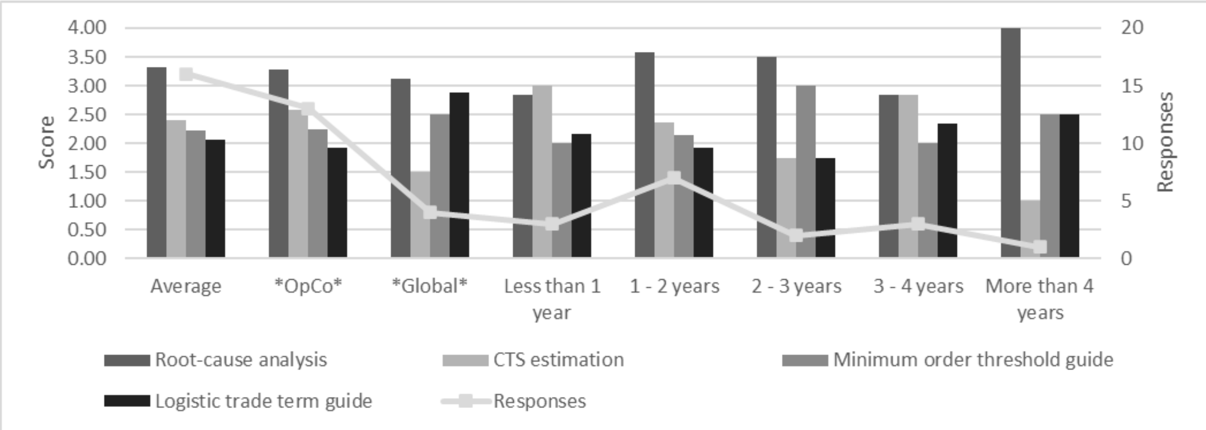


Figure 4.9: The average score and the number of responses for each group of respondents regarding the assessment of future analytics that we could develop (0 - 4)

This again confirms the requirement for the root-cause analysis as only the three respondents with less than a year of experience with the tool expect more value from estimating the CTS of a new customer or product over a root-cause analysis. The average response and the responses from OpCos confirm the assumption that the research company should develop analytics in the specified order. However, the preference is mixed in groups with four respondents or less, but larger groups support the logic to move from descriptive to diagnostic, to predictive, and finally to prescriptive analytics. So, the research company should develop the features in the order shown in the previous section.

4.2.2 Requirements

Having decided what to develop, we defined requirements based on the framework step in Section 3.1.2. This includes, requirements for users, data requirements, and technical constraints that we must take into account. Regarding the latter, Section 4.2.1 mentioned that Power BI is the preferred software and we incorporated the requirements for a root-cause analysis stated in Section 3.13 into the technical constraints, which are as follows:

- Create the tool in Power BI
- The tool should work for all OpCos that received a CTS implementation
- The following input files are available:
  - The data template (.xlsx)
  - Output files from the QlikView tool (.xlsx)

We defined user requirements related to the goals stated in Section 4.2.1. From those goals, we already included the creation of a Power BI tool in the constraints. So, the other user requirements are as shown in Table 4.2.

Table 4.2: User requirements for the research company's new cost-to-serve tool and its features

Scope	User requirement
Power BI tool	Users can load data into the tool The tool is understandable for users from OpCos The tool has an improved user experience compared to the tool in Qlikview
Existing features	Features migrated from the QlikView tool have an equal or improved functionality
New descriptive features	Features developed apply to all OpCos
Root-cause analysis (diagnostic feature)	The solution finds root causes for a high CTS per volume-unit considering any combination of dimensions The solution works for all OpCos The solution shows a value representing potential savings

Having clarified the technical constraints and user requirements, we evaluated the data requirements. Section 3.1.2 presented a set of variables required to fit models that predict a CTS per volume-unit, and Section 3.3.4 mentioned that a proper root-cause analysis requires all those variables. We performed a gap analysis between the required data and the data available in the provided input files. "Gap analyzing is employed to identify the differences between baseline and target architecture based on architectural views" (Rouhani et al., 2015). Table 4.3 shows which potential cost drivers are problematic to create in orange and which are unavailable in red.

Table 4.3: Combination of activities and cost drivers from literature with cost buckets used currently with CTS in the research company (orange = problematic, red = not available)

Activity	Cost drivers			
Inter-company Transport	Quantity shipped	Origin	Utilization	Supplier
Delivery to Customer	Quantity shipped	Volume shipped	Weight shipped	Distance
	Number of trips	Origin	Utilization	Supplier
Warehousing	Quantity shipped	Volume shipped	Weight shipped	Picking time
Order Management	Orders received			
Overheads	Distance	Vehicles used	Quantity shipped	Area used
Trade Terms				
Customer service	Returns	Complaints	Visits	Calls

The utilization of transports is problematic because we can only calculate the truck utilization per shipment. We cannot calculate the truck utilization for a single customer when a truck visits multiple customers. Determining a supplier and vehicle is also problematic because inter-company transports and customer deliveries are done with combinations of suppliers and vehicles, making it hard to assign a single supplier or vehicle type. Furthermore, the distance, the area used, and all customer service cost drivers were unavailable. So, we could not include quite some variables in the root-cause analysis developed in Section 4.4.5, which may negatively influence the quality of results.

4.2.3 Differences with the framework

There were three differences compared to the Define Objectives of a Solution phase presented in Section 3.1. The first difference is that we included the validation of the decision to develop a root-cause analysis, which was a useful addition to the framework steps. Secondly, there was no requirement for Multi-Criteria Decision Making, as logical arguments led to a prioritization of the potential features to develop. Such a situation could also occur in other FMCG companies. The third difference was that we did not map the environment. At the time of the research, the CTS team was exploring Power BI. Therefore, this research also served as an exploration of the possibilities of using Power BI to create the CTS tool. In hindsight, this led to complications regarding the use of advanced analytics in Power BI.

4.3 Phase 2.0: Preparation

This section focuses on preparations for visualizing the output of CTS analysis in a new tool for the research company by following framework step 2.0 Preparation (when making a new tool) in the context of the research company. We followed this step to prepare for the development presented in Section 4.4. Thereby the following question is partially answered:

> *2.3 How can the research company incorporate the chosen descriptive, diagnostic, predictive, and prescriptive analytics into a new tool?*

The inputs for the steps addressed here are the goals and requirements defined in the Define Objectives of a Solution phase. This section starts with Section 4.3.1 addressing the first two preparation steps, which are the loading of data and the allocation of costs. We combined these steps because the cost allocations rely on the old tool of the research company. Then, Section 4.3.2 presents the final data model created in Power BI, which is the output of the preparation steps. Finally, Section 4.3.3 presents differences compared to the framework steps presented in Section 3.2.

4.3.1 Load data and allocate costs

Section 4.2.1 mentioned that the research company was considering new tooling software at the start of this research for two reasons. Firstly, the developers of the QlikView software used for the current tool no longer maintain it. So, the QlikView solution will become outdated. Secondly, the research company strives to have a single worldwide reporting solution. Consequently, they selected Power BI as the preferred software. We designed a way to prepare input data and load this into Power BI based on this decision. Figure 4.10 shows the steps leading up to the creation of a data model in Power BI.

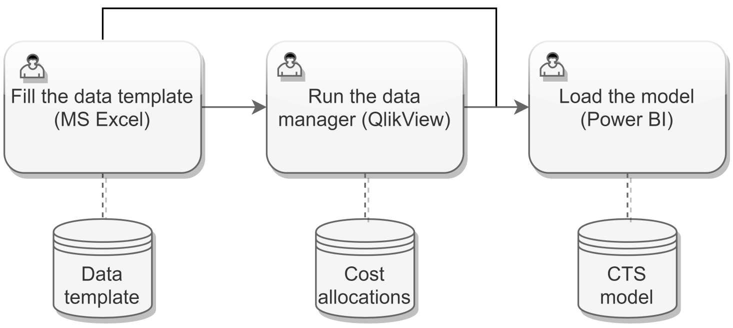


Figure 4.10: The required steps for using the data template and the QlikView data manger to create the CTS model in Power BI

We use the Excel data template where OpCos collect input data as input for the cost allocations in QlikView. Then, we load the original data template and the cost allocations into Power BI. Just as in the current process, these steps still require many manual activities that might cause users to deviate from defined standards. Therefore, Power BI can handle different file types. Compared to the cost buckets presented in Section 3.2.2, the research company does not include customer service costs and out of scope costs.

4.3.2 Create the data model

This section describes the creation of the data model in Power BI, which forms the foundation of the new tool. We created most tables in the model by simply loading the Excel sheet into Power BI. However, some tables required performing transformations or combining sheets. Figure 4.11 shows the resulting model.

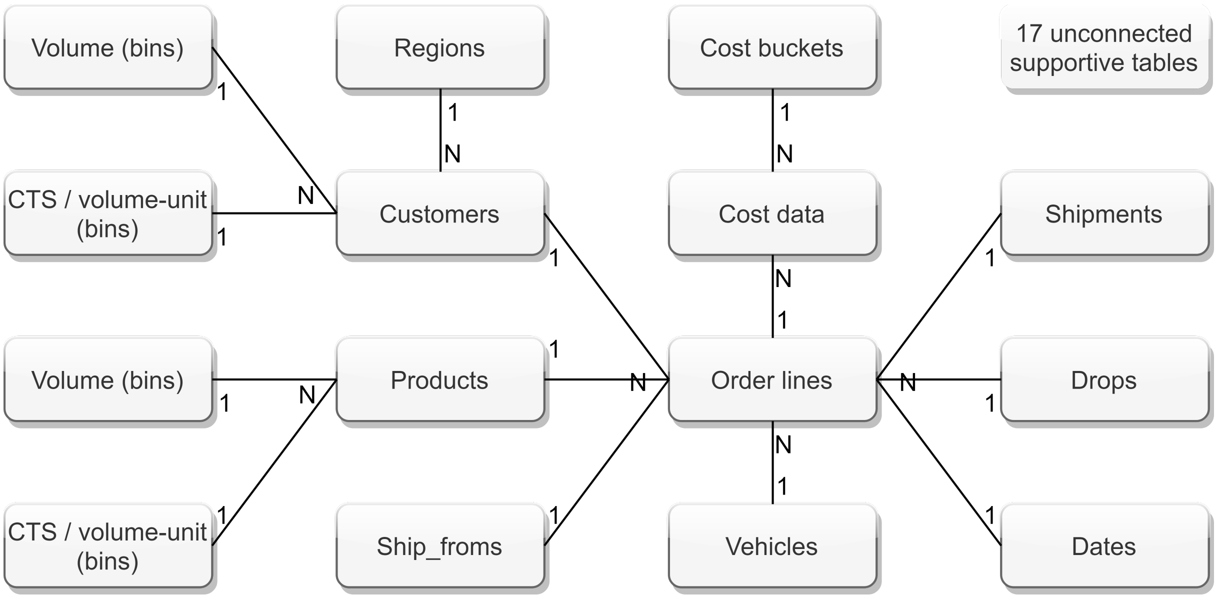


Figure 4.11: The data schema of the data model of the new CTS tool created in Power BI

The scheme includes the relations between tables and forms a snowflake schema, as explained in Section 3.2.3. The table with the lowest level of detail is the Cost data table. This table contains an ID, cost bucket, and allocated costs. Furthermore, most tables are connected to the Order lines table as it contains many keys to other tables. Overall, the model structure provides a setting where it is easy to plot dimensions against each other. We also created many supportive tables, columns, and measures to create features selected in Section 4.2.1. The difference between them is that Power BI creates additional tables and columns when loading the data and calculates measures at the moment a visualization uses them. So, measures are more flexible in use. Some measures are parameters that users define within a given interval. Appendix I shows pseudo-code for the creation of additional tables, columns, and measures.

4.3.3 Differences with the framework

As explained in Sections 3.2.1 and 3.2.2, the process for loading data and cost allocations differ per FMCG company, but we advised a solution that does not require user-interaction and is automated. The solution created for the research company is not a single process, nor is it automated. Compared to the database model presented in Section 3.2.3, two tables are absent. These are the locations table connected to the origins, here referred to as Ship_froms, and the Internal Transfers table. The absence of these tables is due to data unavailability and can restrict the development of features. We also included many tables that support visualizations. Some tables contain bins used in histograms, tables with shipments, and tables with drops. Additionally, seventeen supportive tables are not related to other tables. We required all these tables for specific features, but in general, the data model strongly resembles the data model designed in Section 3.2.3, proving the design is feasible and effective.

4.4 Phase 2: Design and development

This section presents the development of a new tool for the research company. The development applies framework Phase 2 Design and development in the context of the research company. Thereby we answer the following question:

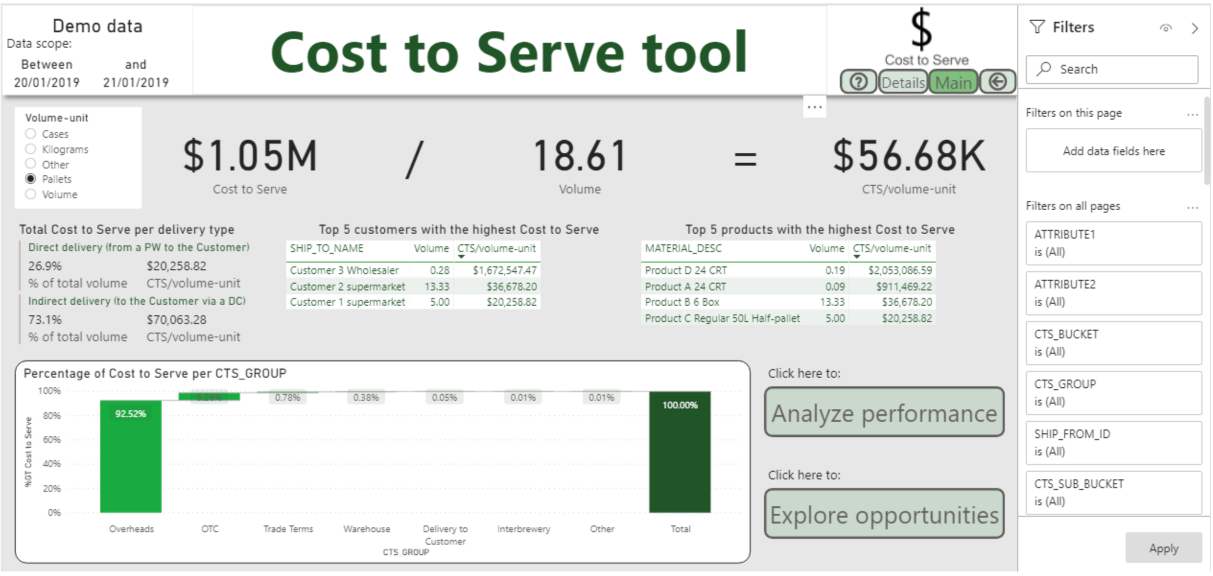
2.3 How can the research company incorporate the chosen descriptive, diagnostic, predictive, and prescriptive analytics into a new tool?

The inputs for this phase are the goals and requirements defined in the Define Objectives of a Solution phase and the data model resulting from the preparation steps of this phase. This section starts with Section 4.4.1, which shows how we designed the report to incorporate a desirable user experience. Then, Section 4.4.2 presents the implementation of hierarchies and a basic segmentation method, Section 4.4.3 presents the recreation, and sometimes enrichment, of features that existed in the old cost-to-serve tool, Section 4.4.4 introduces new features that we included during the development, and Section 4.4.5 shows the implementation of the root-cause analysis approach, which we created in Section 3.3.4. The root-cause analysis is optional, but the variable selection and model fitting processes are a mandatory part of the framework. In the end, the outputs of this phase are a new cost-to-serve tool and the root-cause analysis, which we could not include in the tool. Finally, Section 4.4.6 presents differences compared to the framework steps as presented in Section 3.3.

In this section, we give an impression of the tool created in Power BI and the root-cause analysis solution, but we do not address all features developed in the new cost-to-serve tool. Appendix J presents many features that we do not address in this section.

4.4.1 Tool design

Section 4.2.2 stated two requirements that call for a good design of the new tool. The first requirement was that the tool should be understandable for users from OpCos. Secondly, the tool should have an improved user experience compared to the tool in Qlikview. This section presents elements that support the fulfillment of these requirements. Figure 4.12 shows the main page of the tool.



SHIP_TO_NAME	Volume	CTS/volume-unit
Customer 3 Wholesaler	0.28	$1,672,547.47
Customer 2 supermarket	13.33	$36,678.20
Customer 1 supermarket	5.00	$20,258.82

MATERIAL_DESC	Volume	CTS/volume-unit
Product D 24 CRT	0.19	$2,053,086.59
Product A 24 CRT	0.09	$911,469.22
Product B 6 Box	13.33	$36,678.20
Product C Regular 50L Half-pallet	5.00	$20,258.82

Figure 4.12: The Main page created in the new Power BI tool using demo data

In the top-left, a selection box shows the volume-units users can choose. When a user changes the volume-unit, the tool shows all values accordingly. Next to the chart, there are two buttons related to different sections of the CTS tool. Finally, in the top-right, four buttons are visible. The button on the left leads to the Help page shown in Figure 4.13.

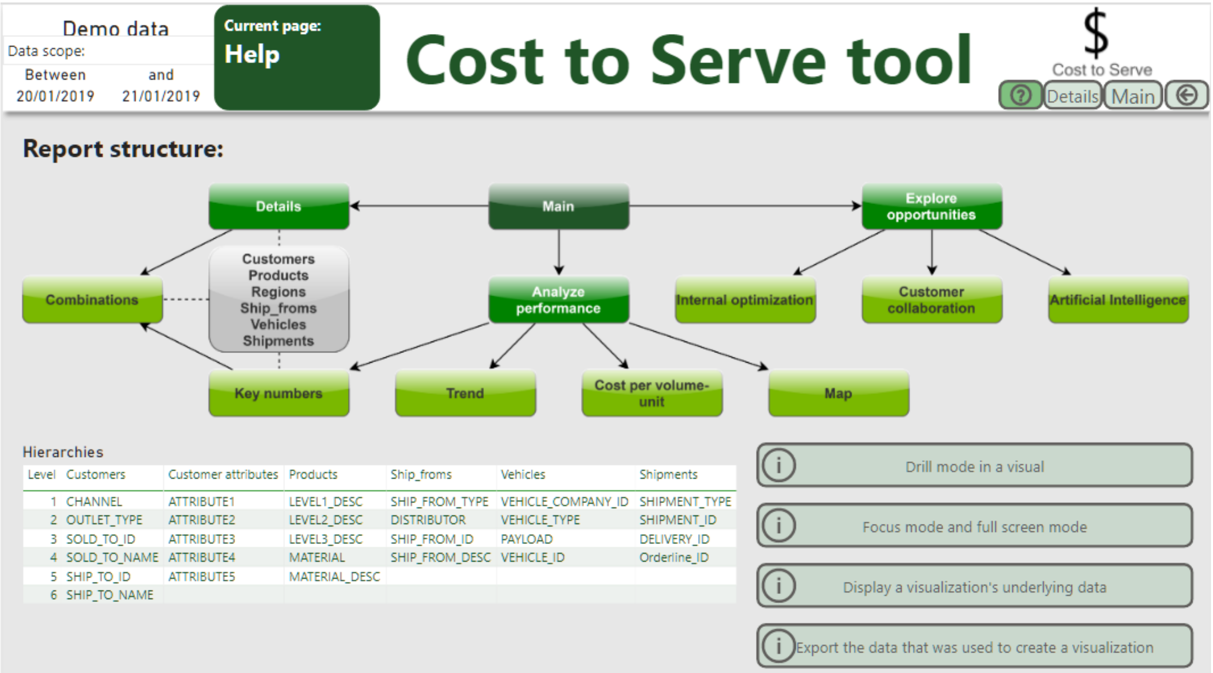


Figure 4.13: The Help page created in the new Power BI tool using demo data

The main element of the help page is a figure showing the report structure. After the main page, there are sections with pages to view details, analyze performance, and explore opportunities. The Details section focuses on different dimensions, showing their attributes, key numbers, and allowing users to combine them, the Analyze Performance section focuses on the current situation of an OpCo, and the Explore Opportunities section focuses on finding potential business improvements. The help page also shows the hierarchies incorporated throughout the report and several buttons with links to web-pages. Every report page has a similar structure to enhance usability. Figure 4.14 shows the page visited when clicking the "Analyze performance" button on the main page.

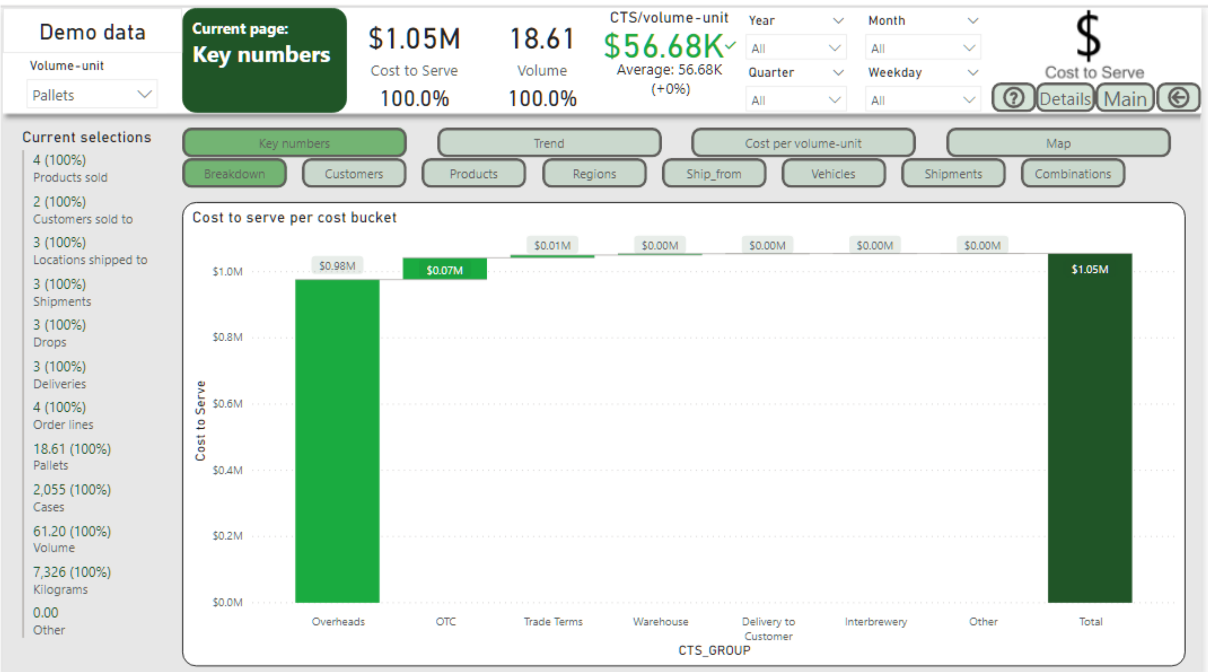


Figure 4.14: The Page structure created in the new Power BI tool using demo data

The top left always shows a name entered by the user when connecting to the data, for example, the name of the OpCo. Below this, users can select a different volume-unit at any time to change how values are displayed. Below the drop-down menu for the volume-unit, the tool shows current selections concerning relevant volume-related KPIs. The percentage behind the number is the current selection divided by the total. When selecting a subset of the data by interacting with visuals or using filters, values change accordingly. The name of the current page is on the top, and to the right of it, three KPIs show the selected CTS, volume, and CTS per volume-unit. The CTS and volume KPIs also show the percentage currently selected, and the CTS per volume-unit shows the deviation from the average based on the selected subset of data. Finally, each page includes several data filters and buttons for navigational purposes.

4.4.2 Hierarchies and segments

In Section 4.2.2 we decided to develop all non-optional descriptive analytics presented in Section 3.1.1, including nested hierarchies and a basic segmentation. Furthermore, Section 3.3.1 argues every CTS tool should incorporate such functionalities. Figure 4.13, which shows the Help page, showed the nested hierarchies included for the research company. The first level indicates the top level in the hierarchy, and a higher level indicates a nested level. We included descriptions in the hierarchies because IDs are often numbers, which might not be familiar to users. Ultimately, we incorporated nested hierarchies in nearly every feature in the tool, allowing users to view data on the aggregation level that suits their needs.

We created the basic segmentation as defined in Section 3.3.1. Figure 4.15 shows the segmentation as visualized in the tool with demo data.

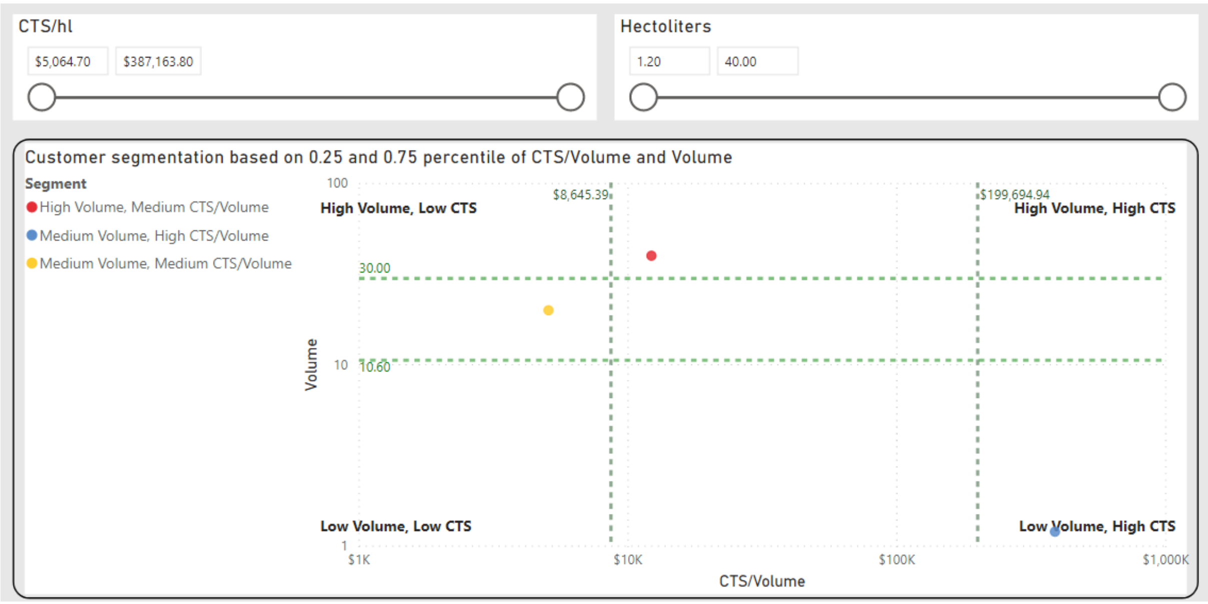


Figure 4.15: The segmentation in the Power BI tool with demo data

As the title of the visualization shows, we applied 0.25 and 0.75 percentiles to create the segmentation. So, segmentation criteria change depending on the input data. By including the segmentation as a filter for all pages, we allowed users to focus on specific groups of customers in the CTS tool. Figure 4.16 shows a segmentation with actual OpCo data.

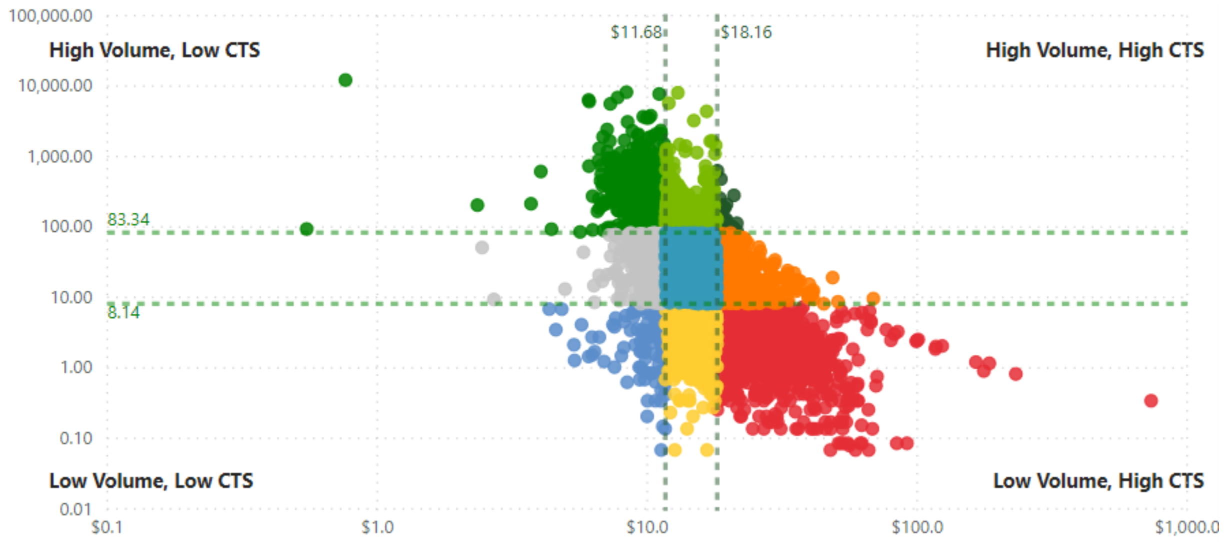


Figure 4.16: The segmentation in the Power BI tool with actual data

The segmentation with actual OpCo data shows approximately 6,000 customers. When a user would choose to focus on a relatively small group with a high volume and CTS per volume-unit, the user would select the corresponding group in a filter and visit other pages to find opportunities for business improvements.

4.4.3 Existing features

In Section 4.2.2 we decided to develop all non-optional descriptive analytics presented in Section 3.1.1. The first feature was the filter functionality. In Power BI, this feature is a part of the user interface with low customizability. So, the filter functionality included in the tool differs from the filter functionality in the QlikView tool. Then, we created the three graphs that received high scores in the evaluation of the features of the old tool, in Section 3.1.1. Figure 4.17 shows the Cost per volume-unit graph. Additionally, we created the Scatter/drop analysis, Delivery profile, and map graphs. Appendix J shows those graphs.

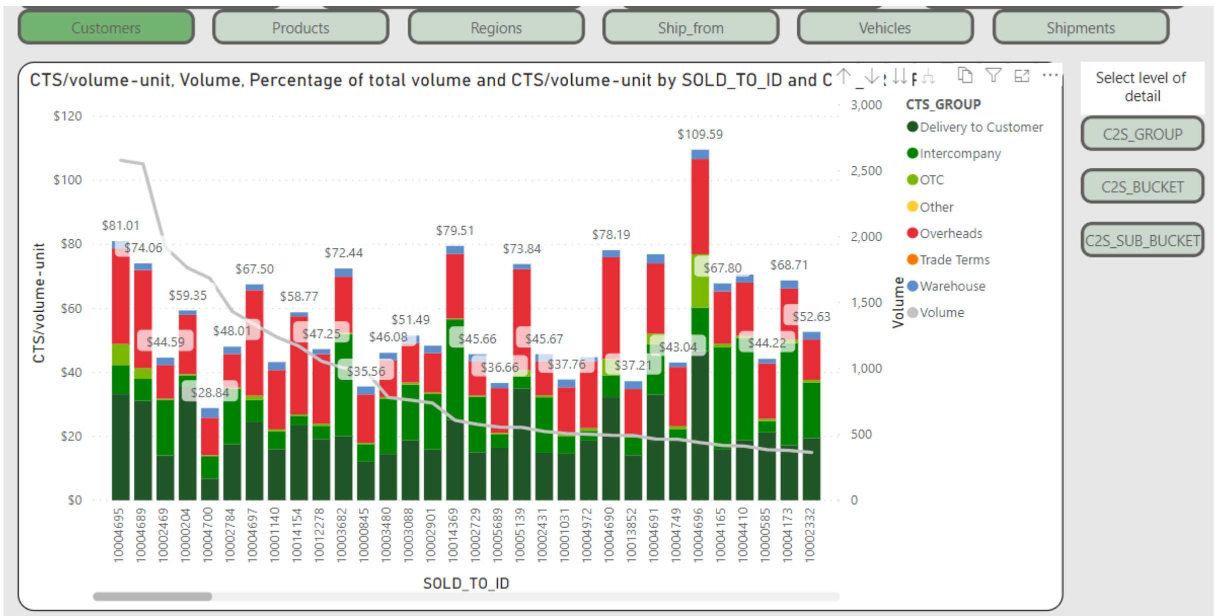


Figure 4.17: The Cost per volume-unit graph created in the new Power BI tool

In the old CTS tool, users could switch between a view showing the line or showing the stacked column. Here, we combined these functionalities and added labels showing the total CTS/volume-unit, and we included buttons above and to the right of the visualization to change the dimension or level of cost buckets. Next, we created the tables suggested by the framework in Section 3.1.1, which are the key numbers, detail, and combination tables. Figure 4.18 shows the Customer's key numbers page, and Appendix J shows the others.

Cost to Serve

CHANNEL	Delivery to Customer	Interbrewery	OTC	Other	Overheads	Trade Terms	Warehouse	Total
1 (on trade)	$65.00	$0.00	$20,937.50	$36.00	$71,731.34	$8,200.00	$324.25	**$101,294.09**
2 (off trade)	$126.67	$86.67	$21,937.50	$40.00	$463,283.58	$0.00	$3,568.22	**$489,042.64**
3 (mixed)	$380.00	$3.16	$23,125.00	$2.16	$440,985.07	$0.00	$101.13	**$464,596.52**
Total	**$571.67**	**$89.82**	**$66,000.00**	**$78.16**	**$976,000.00**	**$8,200.00**	**$3,993.60**	**$1,054,933.25**

Volumes

OUTLET_TYPE	Products sold	Customers sold to	Locations shipped to	Shipments	Drops	Deliveries	Order lines	Volume	CTS/volume-unit
Retail	2	1	2	2	2	2	2	18.33	$32,200.19
Wholesaler	2	1	1	1	1	1	2	0.28	$1,672,547.47
Total	**4**	**2**	**3**	**3**	**3**	**3**	**4**	**18.61**	**$56,682.98**

CTS/volume-unit

CTS_GROUP	Delivery to Customer						Interbrewe
SOLD_TO_NAME	Distribution Fixed Charge	Distribution Variable Charge	Primary Customer Delivery	Primary Customer Delivery - Return	Selfcollector	Total	Transfer fr
Customer 3	$0.00	$0.00	$720.00	$648.00	$0.00	**$1,368.00**	
Customer 2	$0.00	$0.00	$0.00	$0.00	$9.50	**$9.50**	
Customer 1	$8.00	$5.00	$0.00	$0.00	$0.00	**$13.00**	
Total	**$2.15**	**$1.34**	**$10.75**	**$9.67**	**$6.81**	**$30.72**	

Figure 4.18: The Key numbers Customers page created in the new Power BI tool showing demo data

This page shows three tables, which we drilled down to different levels. The first table shows costs in each cost group for channels, the second table shows volumes and the CTS per volume-unit for outlet types, and the third table shows the CTS per volume-unit for customers. In the last table, we expanded the cost groups to reveal the underlying cost-buckets, which only shows the costs for customer deliveries, but other groups are visible when using the horizontal scroll bar. There are also buttons on the right to switch between the general customer hierarchy or a hierarchy based on attributes specified by users. Furthermore, we derived a feature from the old tool that was not an option in Section 3.1.1. Figure 4.19 shows it visualizing volume flows from and to customers based on the old tool's overview page.

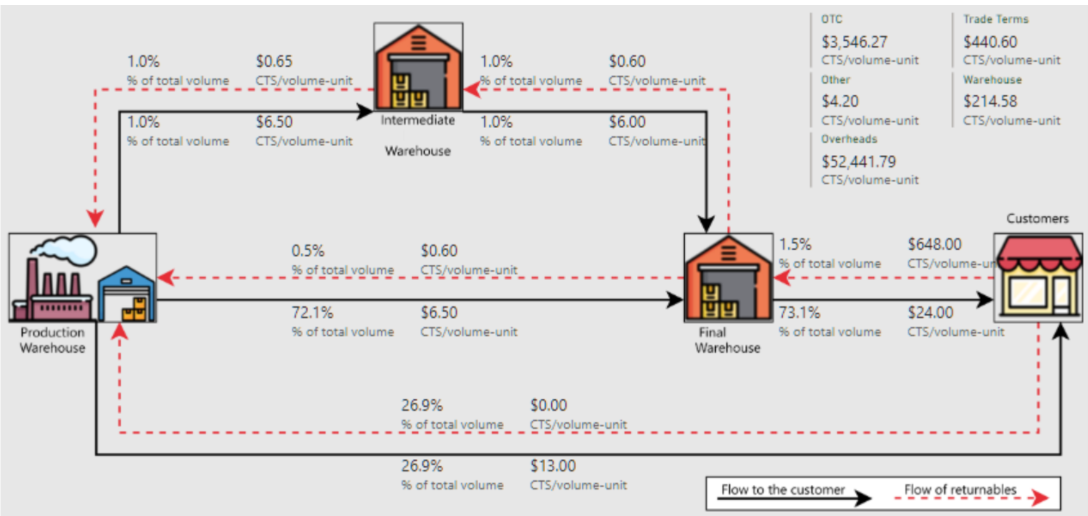


Figure 4.19: The Volume flow feature created in the new Power BI tool showing demo data

The feature shows the percentage of volume with a particular route to the customer. We included the transportation cost for each lane between two locations for finished goods going to the customer and returnable goods returning from the customer. Furthermore, we show the average CTS per volume-unit for cost buckets unrelated to transfers in the top right.

4.4.4 New features

During the development of the Power BI tool, several ideas for features came up. We decided to include them into the new tool in collaboration with the CTS team, and Section 5.1.2 evaluates them along with other features to determine if they should remain in the tool. In the end, we developed ten new pages based on ideas generated during the development process. For example, Figure 4.20 shows a trend page inspired by the Cost per volume-unit - trend graph listed as optional in Section 3.1.1.

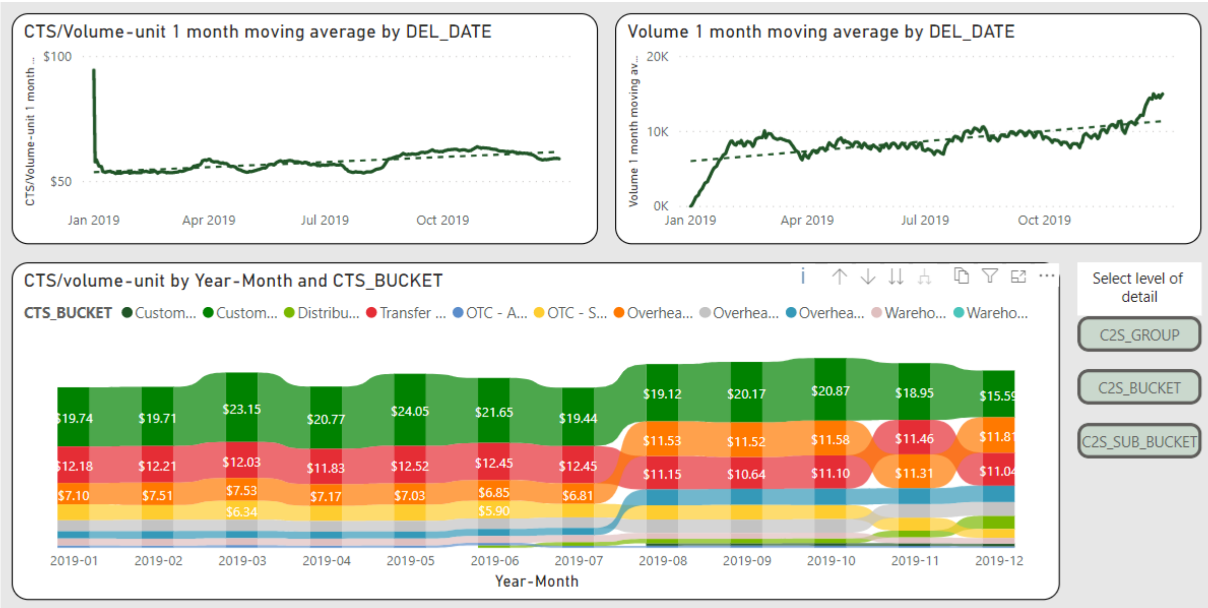


Figure 4.20: The Trend page created in the new Power BI tool

The two graphs on the top indicate the trend of the CTS per volume-unit and the volume, shown by a one-month moving average and a trend line. Ideally, it would show a one-year moving average to include the effects of seasonality. However, only one year of data was available. The bottom graph shows which cost bucket has the highest impact over time for different levels of buckets. Appendix J presents other new features that we developed.

Besides creating new features based on ideas, we also leveraged on newer Power BI visualizations. We created three features that apply artificial intelligence, intending to motivate the research company to explore more advanced solutions that can help OpCos to find opportunities. We created a Q&A visual, a Key influencers visual, and a Decomposition tree visual. The outcomes of these features were often difficult to interpret. Therefore, we did not spend time configuring settings for the features. The research company can look further into how these features can support OpCos in finding opportunities.

4.4.5 Root-cause analysis

This section presents the application of the variable selection and model fitting process from Section 3.3.3, and the root-cause analysis method presented in Section 3.3.4. Besides the gap between the required data and the data available in the research company shown in Section 4.2.2, there is another difference regarding the intended application of the method. Due to technical constraints, we could not incorporate the processes described in 3.3.3 and 3.3.4 into the CTS tool. However, we did follow those processes as intended using Excel, R, and RStudio. We chose R because the research company can, theoretically, embed R scripts in a Power BI report. So, after this research, the research company should be able to integrate the solutions. Figure 4.21 shows the modular solution that we created. We designed a modular solution to enhance the potential useability of the root-cause analysis for the research company.

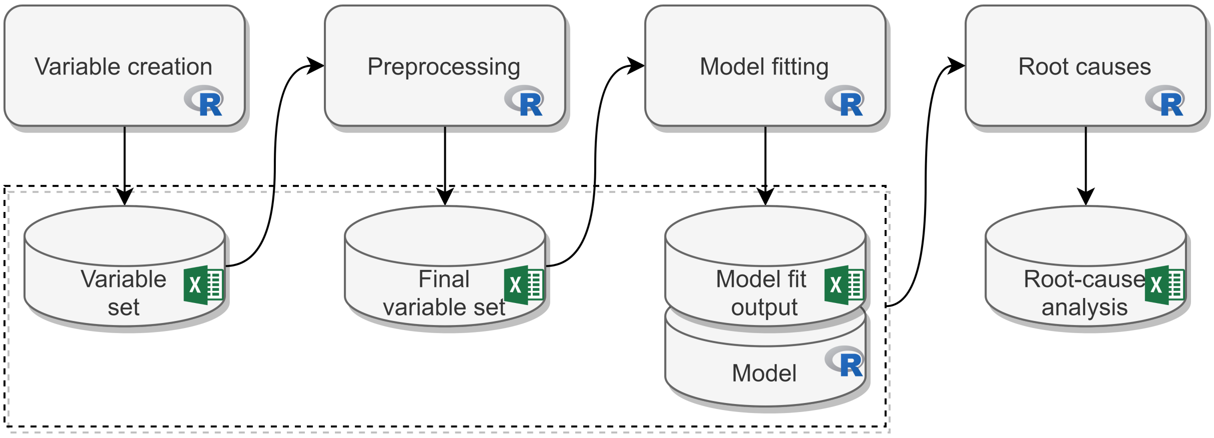


Figure 4.21: The process with subsequently, variable creation, preprocessing, model fitting, and the root-cause analysis showing intermediate output files

The process requires two input files, the shipment and cost details, that users can extract from the QlikView tool. Then, they set the level of detail in the first file and run the R scripts in the sequence shown. An advantage of the modular solution is that users can alter the Excel sheets, which serve as input and output files. Furthermore, users can use the Excel sheets for different reasons than described in this process. Appendix K shows the code of the R scripts.

The variable selection and model-fitting process designed in Section 3.3.3 mentioned several parameters, which are the cutoff value for correlations and the number of repeats and folds for cross-validation. To test the sensitivity of these parameters, we performed two experiments. First, experiments with cutoff values for a data set showed the best performance with a value of 0.9, which avoids "very strong" correlations based on the interpretation of cutoff values from Section 2.5. Especially the Recursive Feature Elimination random forest model (rfe) showed an improved performance with higher cutoff values, which is understandable as Section 2.6 explained that this model handles collinearity well. A higher cutoff value means the script removes less correlating variables, which leads to more variables in the final set. Then, we tested the sensitivity of the parameter values for the number of folds and repeats when performing the cross-validation method presented in Section 3.3.3. In this experiment, we did not observe an improved performance for higher values, but we observed a high increase in runtime. Therefore, a single run with 2-fold cross-validation is the most effective option based on this experiment. Appendix L shows detailed results and more analysis concerning these experiments.

After determining the values for parameters, we ran the four R scripts for two different OpCos to ensure that the solution works for different data sets. For both OpCos, we performed runs for customers, products, and combinations of the two. Table 4.4 summarizes the data sets.

Table 4.4: The data sets for which a root-cause analysis was performed

OpCo	Set	Customers	Products
	Customers	6183	
1	Products		11
	Customers - Products (5118 out of 36103)	727	11
	Customers	269	
2	Products		313
	Customers - Products	269	313

For the combination of customers and products from OpCo 1, we created variables and fitted models based on the entire data set. However, we only found root-causes for a part of the data set due to a very long runtime. The variable creation and preprocessing scripts ran as expected, properly creating all variables and preprocessing them into a final set for the fitting of models. Table 4.5 shows all the results for each data set from the model fitting step of the process.

Table 4.5: The best model fits for the root-cause analysis and their performance for each data set (with a cutoff value of 0.9 and performed a single repeat of 2-fold cross-validation)

OpCo	Set	type	variables	time	RMSE	R^2	MAE
	Customers	rfe	9	1510.83	10.03	0.30	2.53
1	Products	ridge	7	0.45	1.45	0.72	1.36
	Customers - Products	rfe	23	11004.02	13.38	0.17	3.34
	Customers	rfe	3	45.14	6.96	0.93	2.48
2	Products	rfe	20	17.64	1.40	0.82	0.84
	Customers - Products	rfe	33	2420.00	26.11	0.55	3.29

In all cases, except for one, the best fit was a random forest model fitted by recursive feature elimination. Only for the products of OpCo 1, a ridge regression model showed the best performance. Overall, the model performance differs heavily concerning the indicators. The average CTS per volume-unit is 11.18 for OpCo 1 and 3.38. So, the best MAE deviates 12.2%, and the worst MAE differs 97.3% from its corresponding average CTS per volume-unit, which means the model fit is reasonable in some cases, but in others, the script could not fit a good model given the final set of variables. In conclusion, the varying results do not indicate that the variable selection and model fitting scripts created for the research company will provide well-fitted models for all OpCos.

After fitting models, we ran the script with the root-cause analysis algorithm from Figure 3.13, which shows savings for each entity based on a maximum of three peer entities, where the first peer has the most comparable variables to the entity. If there are multiple peers with the same number of similar variables, the first peer is the entity with the lowest CTS per volume-unit. Savings found with the root-cause analysis are corrected based on the MAE. So, a low MAE value leads to more conservative potential savings. Table 4.6 shows an overview of savings found based on the first peers for each set.

Table 4.6: The savings found for the most comparable peers in the root-cause analysis specified in total savings found, and savings when there are 0, 1, and 2 or more focus variables (in thousands)

OpCo	Set	Savings peer 1	With 0 focus variables	With 1 focus variable	With 2 or more focus variables
	Customers	$ 637.74 K	$ 7.42 K	$ 212.72 K	$ 417.60 K
1	Products	$ 4.47 K	$ -	$ -	$ 4.47 K
	Customers - Products	$ 122.36 K	$ 1.10 K	$ 8.16 K	$ 113.10
	Customers	$ 229.96 K	$ 31.79 K	$ 66.31 K	$ 131.86 K
2	Products	$ 1,979.38 K	$ -	$ 14.13 K	$ 1,965.25 K
	Customers - Products	$ 681.34 K	$ 1.73 K	$ 59.62 K	$ 619.99 K

The height of potential savings varies between the data sets, which can have multiple causes:

- The constraint that the volume cannot differ more than 10% restricts the number of possible peers, making it more difficult to find a comparable entity.
- The correction of the potential savings for the MAE can cause savings to be lower.
- Savings might be higher because there is more room for improvement.

In general, the outcomes show high potential savings. For OpCo 1, the customer's closest peers show a potential 10% reduction of the costs in scope, and for OpCo 2, this is 20% for the products. To indicate where savings originate from, we present an actual example of savings found. Customer 1 from OpCo 2 displays Customer 2 as a peer, indicating the focus variable is the Average volume per delivery and the potential savings are $ 3,995.26. Table 4.7 shows how these customers relate to each other.

Table 4.7: A comparison of two customers as shown in the output of the root-cause analysis for products from OpCo 2 (with the focus variable in red)

Key	Volume	CTS	Average volume per delivery	Maximum volume per delivery	Average volume per order line	CTS per volume-unit
C1	3,609.42	$ 25,253.74	25.07	206.77	3.03	$ 7.00
C2	3,261.20	$ 11,128.07	75.84	237.12	6.45	$ 3.41

Both customers have comparable volumes, but the costs and CTS per volume-unit differ. The algorithm determined the difference is caused by the Average volume per delivery, as the Maximum volume per delivery and the Average volume per order line both do not show a significant difference. Assuming the fitted model reflects reality, increasing the Average volume per delivery for Customer 1 leads to the CTS per volume unit of Customer 2. Hence, the calculation of the savings:

$$Savings = CTS_{Customer1} - Volume_{Customer1} \times (CTS\,per\,volume - unit_{Customer2} + MAE)$$

$$Savings = 25,253.74 - 3,609.42 \times (3.41 + 2.48)$$

$$Savings = \$\,3,994.26$$

Table 4.6 also showed the savings found based on different numbers of focus variables. When there are zero focus variables, all variables of the first peer compare to the variables of the entity, which means the model cannot explain the difference in the CTS per volume-unit. Ideally, peers that differ on a single variable show the most savings, as then FMCG companies know what to change. That is the case for customer peers in OpCo 1, showing a potential cost reduction of around 3%. Furthermore, $ 72,621.27 of those potential savings relate to 433 customers that have a Loose case picking percentage (orders > 1 pallet) that is too high. So, according to the analysis, the OpCo could save that amount by reducing the loose case picking in orders of more than one pallet for those customers. On the other hand, Table 4.6 showed many savings related to peers with two or more focus variables. For those cases, it is more difficult to make business changes that impact all variables. In conclusion, the output of the root-cause analysis provides understandable savings on which OpCos determine actions. However, the validity of the savings depends on the quality of the model fit to the data set, but the analysis still highlights problems as expected from a feature applying diagnostic analytics.

4.4.6 Differences with the framework

The development in this section was comparable to the framework steps of the Design and Development phase presented in Section 3.3. First, we developed the hierarchies and basic segmentation as intended. Then, we re-developed features that exist in the QlikView tool in Power BI successfully. For example, we improved the key numbers feature by including the CTS per volume-unit, products sold, and distinguishing between customers sold to and locations shipped to, as a single customer can have multiple shipping locations. Furthermore, Figure 4.19 showed the volume flow feature, which we based on the overview page of the QlikView tool, and included as an extra existing feature. We also included several new features, which shows that new ideas can generate during the development process. Finally, we implemented the process to select variables and fit models in preparation for the root-cause analysis with varying results. Looking at differences with the framework, we could not include all variables in the tool due to data unavailability, which possibly caused the poor model fits shown in Section 4.4.5. Another difference compared to the framework is that we were not able to integrate the variable selection, model fitting, and root-cause analysis into the new CTS tool. We evaluate these differences and differences related to other phases in Section 5.2.

Chapter 5

Evaluation

In this chapter, we evaluate the framework for fast-moving consumer goods companies to find business improvements using the output of a cost-to-serve analysis and the new tool we created for the research company, which resulted from applying the framework in a case study. First, Section 5.1 focuses on the assessment of the new tool, which is similar to the assessment done for the old tool in Section 4.1.2. Then, Section 5.2 evaluates the framework in light of the case study by highlighting differences to determine the applicability of the framework.

5.1 Power BI tool

This section evaluates the new tool that we created based on the designed framework to find business improvements using the outcome of a cost-to-serve analysis. Section 5.1.1 evaluates whether the new tool meets the requirements set before development. Then, Section 5.1.2 shows the outcomes of the evaluation of the new tool and compares the new and old tools. Finally, Section 5.1.3 evaluates whether we achieved set goals in the creation of the new tool. The first two sections include the results of a survey sent out to assess the new tool. Appendix M shows all answers to this survey.

5.1.1 Requirements

This section focuses on requirements formulated for the development of the new tool by evaluating technical requirements and user requirements. We evaluate the requirements listed in Section 4.2.2.

Technical requirements

Concerning requirements listed in Section 4.2.2, we did not meet the data requirement and the requirement for a single solution. In general, requirements depend on the company, but Section 3.1.2 presented various potential cost drivers of a high CTS per volume-unit, which FMCG companies should include in their CTS analysis. The framework requires the availability of these variables for features using descriptive analytics and the root-cause analysis presented in Section 4.4.5. However, Section 4.2.2 showed that it was not possible to include all variables, but we did include all variables from the old tool to avoid a loss of information. To resolve the issue, the research company must expand the data collection during a CTS implementation.

Section 4.2.2 also mentioned three other requirements. The first requirement was to develop the tool in Power BI. Features that apply descriptive analytics met this requirement, but the root-cause analysis did not, due to constraints imposed by the IT environment of the research company. IT experts from the research company can most likely resolve this issue. The second requirement was that the tool should work for all OpCos that received a CTS implementation, and the third requirement specified the use of the data template and output files from the old tool as input files. We tested the new tool with data from five OpCos and the root-cause analysis with two different OpCos. As both solutions used the specified input data and worked for different OpCos, we met the second and third requirements.

User requirements

Here, we subsequently review requirements listed in Section 4.2.2 regarding the Power BI tool, existing features, new descriptive features, and the root-cause analysis. We evaluated several requirements via a survey presented to the CTS team and a CTS tool user from an OpCo.

We formulated three requirements for the Power BI tool. Firstly, we enabled users to load data into the CTS tool. Users can place the input files in the research company's file storage environment, open a template version of the new CTS tool, specify the location of the input files, and load the files. Secondly, the evaluation of the expected effort experienced when using the new CTS tool in Section 5.1.2 shows that the new CTS tool is understandable for users. Thirdly, we assessed whether the new tool improves the user experience compared to the old CTStool by asking respondents how they would rate the user experience in both CTS tools. Figure 5.1 shows the results from the survey, with optional ratings between one and four stars.

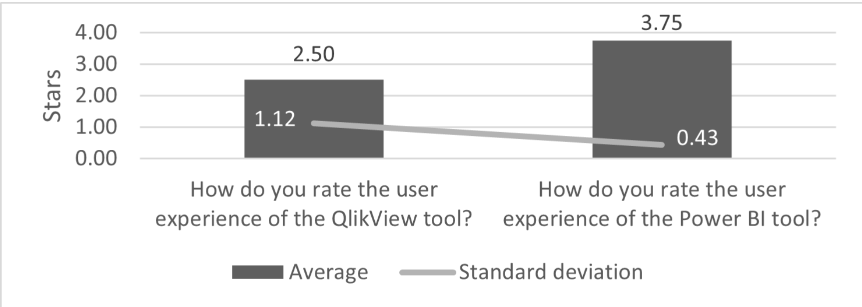


Figure 5.1: The results of the survey questions evaluating user experience of the old and new tool showing the average and standard deviation of responses (1 - 4 stars)

The user experience in the new CTS tool is superior compared to the old CTS tool. Furthermore, the standard deviation of the responses for the new CTS tool is lower, which means there is more agreeance between respondents. So, we conclude that we met the third requirement.

Features that existed in the old tool required equal or improved functionality in the new CTS tool, which we assessed by asking survey respondents to what extent the new tool shows an improvement concerning these features. Possible answers ranged between much worse (-2 points), unchanged (0 points), and much improved (2 points). Figure 5.2 shows the averages and the standard deviations of the existing feature's scores.

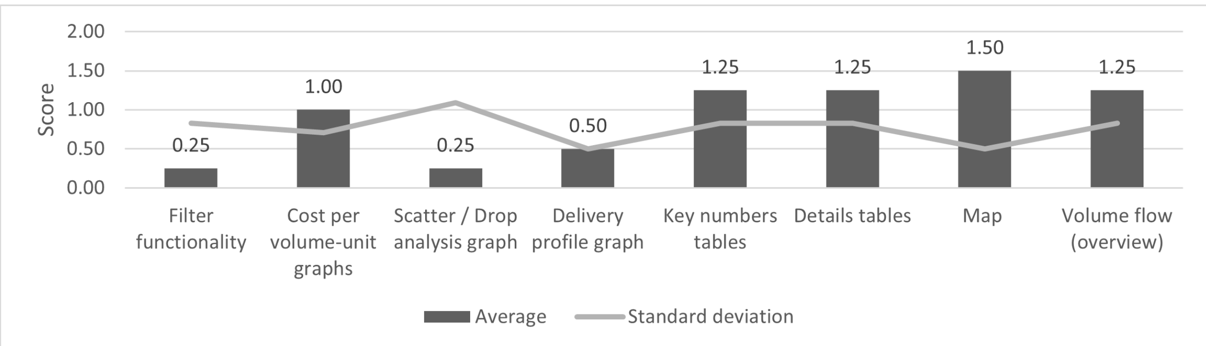


Figure 5.2: The results of the survey questions evaluating the features that existed in the old tool that we recreated in the new tool showing the average and standard deviation of scores (In the new tool compared to the old: Much worse -2 - 2 Much improved)

Most features have an average value between one and two, indicating slight or much improvement. As those features all have a standard deviation lower than the average, we can argue the respondents experienced the functionality of these features is either unchanged or improved. The Filter functionality, the Scatter/drop analysis graph, and the Delivery profile graph have scores between zero and one with a higher or equal standard deviation, which shows perceptions of respondents differ concerning the improvement of these features. However, with average values above zero, it is safe to say the functionality is not worse. Therefore, the existing features met the requirement for improved functionality in the new tool. Furthermore, we stated that new descriptive features should work for all OpCos, which is the case when the input data is complete. So, we met this requirement as well.

We formulated three requirements for the root-cause analysis. Firstly, the solution finds root-causes for a high CTS per volume-unit of any combination of dimensions. In the script where we create variables, users define a key for each order line based on columns available in the shipment details. This key, for which the method determines root-causes for a high CTS, can be a single dimension or a combination. Secondly, the solution works for all OpCos. We extracted input data in a standard format from the old tool for two OpCos. Therefore, the solution works for OpCos that extract data similarly. Thirdly, the solution shows a value representing potential savings, which we achieved by displaying savings related to peers of an entity. So, we met all requirements concerning the root-cause analysis method and all other user requirements formulated for the development of a new tool in this research.

5.1.2 Performance new tool

This section evaluates the performance of the new tool its features. The assessment is similar to the evaluations performed in Section 4.1.2. First, we assess the performance of the new tool. Then, we compare the number of features used in the new and old tools.

Tool assessment

The CTS team and a CTS tool user from an OpCo filled the survey presented in this section. Ideally, there would be more respondents from OpCos, as 13 respondents from OpCos assessed the old CTS tool, but as the research company did not yet distribute the new CTS tool among OpCos, we could only involve a single OpCo. Figure 5.3 shows the results of the assessment of the old and new tools. When a respondent completely disagreed with a statement, we awarded zero points, and when a respondent completely agreed, we awarded five points.

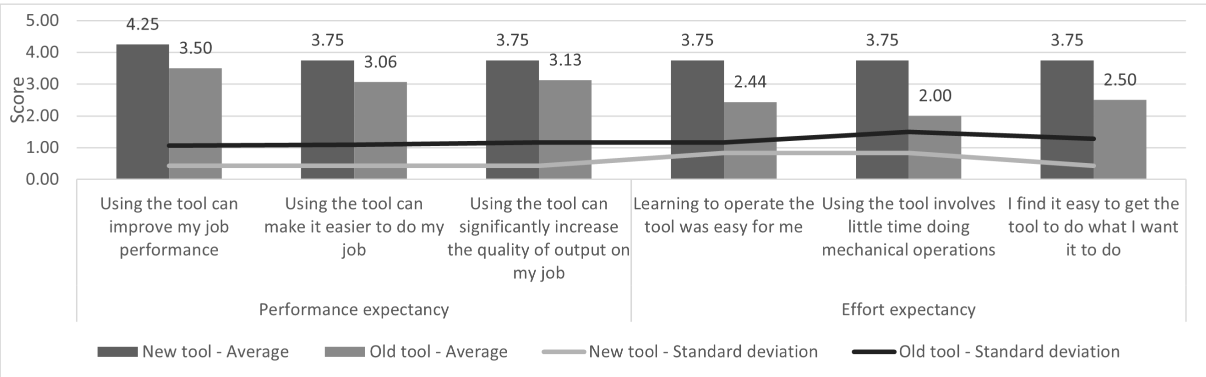


Figure 5.3: The results of the survey questions evaluating the old and new tool showing the average and standard deviation of responses (Completely disagree 0 - 5 Completely agree)

The new tool outperformed the old version regarding all statements shown. The new tool received an average score of 3.83, where the old tool received an average score of 2.77 out of 5. So, in general, the new tool yields an improvement for the research company. The low standard deviation indicates that many responses were close to the high average. We also assessed the new features by asking respondents how they value them by awarding between one and four stars. Figure 5.4 shows the result of this assessment.

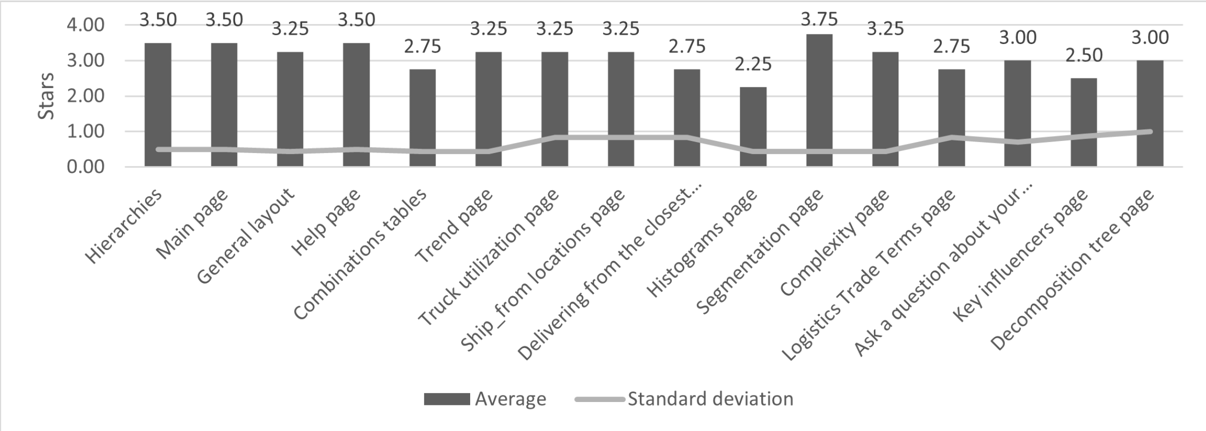


Figure 5.4: The results of the survey questions evaluating new features showing the average and standard deviation of responses (1 - 4 stars)

Section 3.1.1 included features with an average assessment of 3.5 stars or higher in the framework. In this assessment, hierarchies and the segmentation page received 3.5 stars or higher, validating the use of nested hierarchies and segmentation in the framework, and the main and help page also received 3.5 stars, validating the improved user experience. The general layout and six pages received average scores between 3 and 3.5. We can include these pages as optional in the framework, and the research company could attempt to improve them before deciding to exclude them from their CTS tool. Furthermore, they can collect more specific feedback to find improvements for the general layout, and they should improve the Combinations tables because the framework requires them. Delivering from the closest DC, Histograms, and Logistic trade terms pages, based on ideas generated during development, also have low scores. The research company should not continue developing those features, as they did not show potential. Finally, the Key influencers page, containing a standard Power BI visualization, has a low score. The research company could improve this page, as it does not require much effort.

We evaluated the root-cause analysis method similar to the old and new tools to determine whether the research company should continue to develop the designed method and include it in the tool. It was only possible to evaluate the method with the CTS team, as we did not manage to include the method in the new tool. Figure 5.5 shows the results of this assessment.

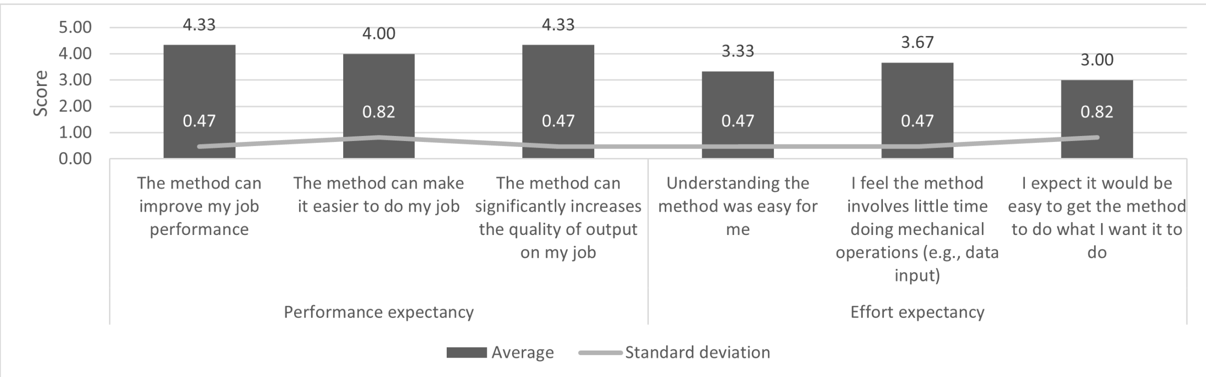


Figure 5.5: The results of the survey questions evaluating the root-cause analysis showing the average and standard deviation of responses (Completely disagree 0 - 5 Completely agree)

The expected performance and effort both show above-average scores. Respondents agree with all the proposed statements, but the effort expectancy scores lower, which might be due to the method not being integrated into the tool. In general, the high scores indicate that the research company should add the root-cause analysis method to the CTS tool.

Feature assessment

In Section 4.1.2, we observed that the old CTS tool incorporated a large number of features. We performed a similar mapping of all features in the new tool that we created in Power BI. Figure 5.6 shows how the new CTS tool compares to the old tool.

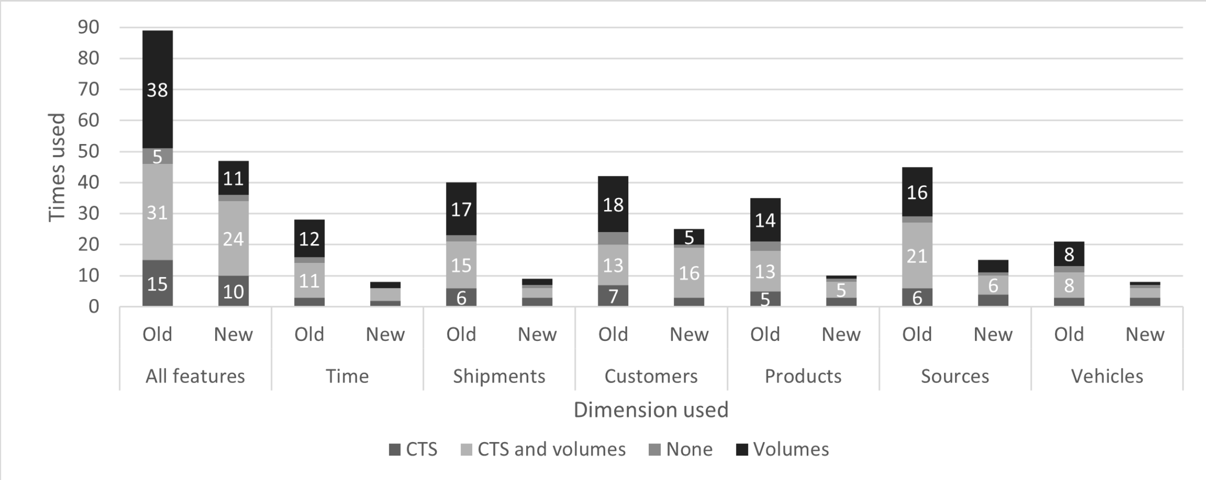


Figure 5.6: A comparison between the features used in the old CTS tool and the new tool showing what features are measuring in both tools and how often dimensions are used (labels for values above 4)

In general, the new CTS tool reduces the number of features used from 89 to 47. So, there are way fewer features users have to explore, even though we added additional features during the Design and Development phase. Furthermore, the new tool focuses more on the CTS compared to the old CTS tool, but volumes remain dominantly present. Regarding dimensions used in features, the new CTS tool focuses mainly on customers, as they are the main focus of a CTS analysis. We also considered sources in more than ten features, and other dimensions are still present, but the new tool includes them way less often. In the end, the features mapped here mainly use descriptive analytics, as was the case with the old CTS tool, but way fewer features are required to provide the same and additional information.

5.1.3 Goals

This section evaluates whether we achieved the goals for the development of the new cost-to-serve tool. Section 4.2.1 presented five goals. We managed to include all non-optional descriptive analytics from Section 3.1.1, improve descriptive analytics where possible, test new ideas for descriptive analytics that came up during the development, and create a feature that finds root-causes for a high CTS per volume-unit. However, we did not integrate the feature that finds root-causes for a high CTS per volume-unit into the tool in Power BI due to constraints in the IT environment. With a thorough mapping of the IT environment, the research company can deal with comparable IT constraints during the next Define Objectives of a Solution phase. The framework advises to map the IT environment, but the CTS team was exploring it themselves at the time of the research, which made it challenging to assess IT possibilities. However, we achieved almost all goals set by applying the framework. In conclusion, realizing four out of five goals shows a successful development of the CTS tool. The new tool outperforms the old CTS tool with fewer features, we validated the inclusion of hierarchies and segmentation in a CTS tool, 13 out of the 16 new features show potential, and the evaluation argues the research company should continue developing the root-cause analysis.

5.2 Framework evaluation

This section evaluates the framework for FMCG companies to find business improvements using the output of a cost-to-serve analysis presented in Chapter 3 by outlining differences between the framework and the application of the framework in the case study performed at the research company presented in Chapter 4. The purpose of this section is to connect the differences observed between the framework and the case study and find potential opportunities for improving the framework. Table 5.1 shows the differences between the framework and the case study.

Table 5.1: The differences between the framework for fast-moving goods companies to find business improvements using the output of a cost-to-serve analysis and the application of it in a case study

Phase	Step	Difference between the framework and the case study
3: Demonstration	Obtain insights and find opportunities	None
	Map opportunities	Different naming in opportunity typology Did not consider the potential automation of opportunities Created a Power BI report to visualize past opportunities
4: Evaluation	Measure success	Measured potential savings rather than actual profit and customer satisfaction Measured success of opportunity types
	Assess tool performance	None
	Assess features	Assessed potential features after setting a goal
1: Define objectives of a solution	Review options	Validated goal including new analytics with users Did not require Multi-Criteria Decision Making methods
	Define requirements	Did not map environment
2.0: Preparation	Load data	Segmented and unautomated solution
	Allocate costs	No Customer service and Out of scope costs
	Create data model	Two absent tables due to unavailable data
2: Design and development	Review hierarchies and segments	None
	Rework existing features	None
	Select variables and fit models	Not all variables included Not integrated into the CTS tool
	Extra step	Included features based on ideas generated during development

In the Demonstration phase, there are two occurrences where we did not perform something described in the framework as intended in the case study. Firstly, some names in the typology created differ from those in the framework, but the meaning remained the same, so it did not matter. In general, FMCG companies can change names in the typology, as long as they create and apply it. Secondly, we did not take the automation of opportunities into account, as the opportunity mapping performed was the first for the research company, and we found many different types. Therefore, we did not focus on which opportunities have the potential for automatizing due to time constraints. The Demonstration phase also presented an element to add to the framework. The report created in Power BI to visualize past opportunities found was of great value for the analysis of the past and measuring success. Therefore, we could add the creation of a report to the process for mapping opportunities.

In the Evaluation phase of the case study, there is one difference compared to the framework that the research company can resolve by making a change in this phase. Ideally, measuring success is done with actual KPIs as these are more valuable in the Define of objectives for a solution phase, but we measured potential KPIs that are less precise. As an addition to the framework, we showed that measuring success is possible for opportunity types, which could be useful to do for other FMCG companies as well.

Three differences between the framework and the case study originated in the Define objectives of a solution phase. Firstly, we assessed potential features after we defined a goal instead of during the Evaluation phase, which allowed for a more concise assessment and avoided asking users about features the research company might not develop. Thus, the step where FMCG companies assess potential features fits better in the Define objectives of a solution phase. Secondly, we did not use Multi-Criteria Decision-Making methods when formulating a goal. We did not require such methods as we could make a decision based on logic, but when other FMCG companies could not do the same, such methods could still be useful. Thirdly, we did not map the IT environment during the definition of requirements. Later, in the Design and Development phase, we were not able to integrate the steps for variable selection and model-fitting into the CTS tool, emphasizing the importance of mapping the IT environment.

In the preparation steps of the Design and Development phase, there are three occurrences where we did not perform something in the case study as described in the framework. Firstly, the data loading process is segmented and not automated, which leads to users experiencing a high effort when working with the CTS tool. However, this does not negatively affect any other phases of the framework. Secondly, the research company does not include Customer Service and Out of Scope costs in their CTS analysis. The absence of customer service elements resulted in a lack of customer satisfaction measurements in the Evaluation phase. Due to the absence of Out of Scope costs, we measured savings rather than profit, but this did not have negative consequences. Savings can even serve as a suitable alternative for profit in the framework, but including both cost-buckets would strengthen the Evaluation phase of the CTS analysis. Thirdly, two tables from the proposed data model were absent, which led to unincluded variables in the Design and Development phase. The absence of variables possibly caused the model fitting performance to vary and restrains the options for the use of analytics.

In the other steps of the Design and Development phase, we decided to include features based on ideas generated during development, which presented an extension to the framework. An advantage of that step was that we developed more optional features, which we directly evaluated, but Section 5.1.2 showed that three of the seven new features lack potential. Therefore, we advised to discard three of them and include the others as optional descriptive features. Furthermore, one of the three standard Power BI artificial intelligence features scored relatively low, but including these features required little effort. So, we advised continuing the development of that feature. In general, FMCG companies should decide beforehand how much time to spend on randomly introduced features. The advantage of creating extra features during development is that there are additional features to consider, but the time investment presents a disadvantage.

In conclusion, many differences require action from the research company, and five differences provide a potential improvement for the framework. These are validating features after defining goals, creating additional features during development, creating a report to visualize past opportunities, measuring the success of opportunity types, and the possibility to include savings rather than profit. The number of differences that require the research company to take action, and the improvement made by the new tool shown in Section 5.1, indicate how following the framework can improve the process of visualizing the output of a CTS analysis for an FMCG company. Taking the differences presented in this section into account, Section 6.2 shows recommendations for the research company in a roadmap.

Chapter 6

Conclusions and recommendations

This chapter presents the conclusions and recommendations of this research. First, Section 6.1 answers the main research questions.

> ***1 How can fast-moving consumer goods companies find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a cost-to-serve analysis?***

> ***2 Can the research company improve the use of the output of cost-to-serve analyses by applying the framework designed in this research to create a new tool?***

So, we reflect on the creation of the framework for fast-moving consumer goods companies to find business improvements using the output of a cost-to-serve analysis and the case study applying this framework in the research company. Then, Section 6.2 presents recommendations for the research company based on the case study. Finally, Section 6.3 discusses the findings in this research in the light of validation, the applicability of the framework, limitations, future work, and the contribution of this research to theory and practice.

6.1 Conclusions

In this section, we present the conclusions of this research in which we developed and applied a framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis. Chapter 3 presented the framework's four phases containing various steps, which we summarize here. The framework describes how FMCG companies can find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a CTS analysis.

Section 3.1 presented the Define Objectives of a Solution phase. We created a set of optional features for a CTS tool by researching which descriptive, diagnostic, predictive, and prescriptive analytics FMCG companies can use and which features of the research company's old CTS tool users value most. Based on the set of options, FMCG companies can set a goal for development and define requirements based on users, the environment, and technical aspects. For the data availability, which is a technical requirement, we created a set of variables based on potential cost-drivers that FMCG companies should include in their CTS analysis.

Section 3.2 presented preparation steps for the Design and Development phase in Section 3.2 that FMCG companies must follow when creating a new tool. These steps are to load data, allocate costs, and create a data model, which is necessary when creating a new CTS tool. We based those steps on processes of the research company, which we supported with reviewed literature. After the preparation steps, the mandatory part of the Design and Development phase in Section 3.3 commences, where we firstly included the use and review of hierarchies and segmentation techniques in a CTS tool. Secondly, we focused on reworking existing features when the Evaluation phase shows this is necessary. And finally, we explained how to select variables and fit models for features using more advanced analytics, which we applied in an algorithm that finds root-causes for a high CTS per volume-unit.

Section 3.4 presented the Demonstration phase, where FMCG companies empower users of the CTS tool to find potential business improvements. They must decide how much training to provide for users based on the amount of human input required to find opportunities, which depends on applied analytics. In this phase, we also researched how to map business improvements obtained with insights from a CTS analysis, resulting in a process that includes a typology to map found opportunities.

Section 3.5 focused on evaluating the CTS analysis. The Evaluation phase starts by measuring the influence CTS analyses have on profitability, and customer satisfaction, which we both derived from the benefits of CTS found in the reviewed literature. Then, an evaluation of the tool takes place. We researched which techniques can assess the quality of a tool visualizing the output of a CTS analysis, creating a model that applies to CTS tools of FMCG companies. Finally, FMCG companies measure the performance of different types of analytics used in the features of the CTS tool. After this phase ends, a new cycle begins with the evaluation phase. By going through the mentioned phases and following the steps, FMCG companies can get the most out of the output from a performed CTS analysis.

In Chapter 4, we applied the framework to find business improvements using the output of a cost-to-serve analysis to the research company to improve the use of results of a CTS analysis and validate the framework. First, Section 4.1 researched how the CTS implementation of the research company was performing by applying the Demonstration and Evaluation phases of the framework. We created an opportunity mapping system, including success measurements that we visualized in a Power BI report. The report includes the research company's success measurement, showing the average potential savings per OpCo are 3% of scoped costs. Then, an assessment of the old tool showed that it performs better on increasing performance than reducing experienced effort, and the feature assessment showed that the research company's CTS tool has many similar descriptive features.

After we assessed the current situation, Section 4.2 presented the Define Objectives of a Solution phase of the framework where we set goals and define requirements, which included one goal to develop a new tool in Power BI. We researched which descriptive, diagnostic, predictive, and prescriptive analytics the research company should use to visualize the output of a CTS analysis by applying the steps to define objectives of a solution of the framework. As a result, we decided to develop all the non-optional features using descriptive analytics from the framework's set of optional features, and we argued the development of a root-cause analysis method, based on a prioritization of diagnostic analytics before predictive and prescriptive analytics, which we confirmed via a survey among users of the CTS tool.

Following the Define Objectives of a Solution phase, Section 4.3 addressed the preparation steps of the framework's Design and Development phase, which was required because we set a goal to develop a new CTS tool. Then, Section 4.4 researched how to incorporate the chosen features into a new tool for the research company by following the other steps of the Design and Development phase of the framework. The result was a fully functional Power BI tool with nested hierarchies, a customer segmentation method, a minimized presence of descriptive tables, highly appreciated graphs from the old CTS tool, and features using descriptive analytics designed during the development process. We evaluated the Power BI tool in Section 5.1, showing a 50% more desirable user experience, that all features that were present in the old CTS tool improved, and that the likelihood to increase performance and reduce effort improved by 38% despite having fewer features. The hierarchies and segmentation page also received a positive evaluation, validating the relevance of the corresponding framework step, and six new features based on ideas generated during development also showed promise. So, FMCG companies could include them in the framework's optional set of features. Finally, we performed the last steps of the Design and Development phase, concerning the variable selection and fitting of models, in the root-cause analysis. It was not possible to integrate them into the CTS tool. It should be possible, but restrictions related to the IT environment of the research company hindered this. Nevertheless, the root-cause analysis method showed great promise revealing potential savings up to 20% of costs in scope for different data sets and a possible cost reduction of more than 3% by improving a single focus variable for customers in one OpCo. The evaluation of the root-cause analysis in Section 5.1 showed the CTS team believes the method can significantly improve performance and reduce the effort of finding root-causes for a high CTS per volume-unit, but a varying model performance and the exclusion from the new CTS tool provide room for improvement.

The case study showed an example of an application of the framework in an FMCG company, revealing a few potential improvements for the framework, which are the possibility to measure the success of opportunity types, include savings rather than profit, validate features after defining goals, create additional features during development, and create a report to visualize past opportunities. Furthermore, we found many possible improvements for the research company, which we address in Section 6.2. Besides validating the framework, the case study delivered an improved CTS tool that the research company will use in practice to find business improvements in many different OpCos. Based on the advancements made in the research company that consists of OpCos with varying characteristics, we proved the framework can support a wide range of FMCG companies. Therefore, we conclude this research successfully produced a framework that enables fast-moving consumer goods companies to find business improvements by using descriptive, diagnostic, predictive, and descriptive analytics in their tool that visualizes the output of a cost-to-serve analysis. In the months after this research, the research company will continue developing this tool, which will show if the improved balance of analytics leads to more OpCos continuously using cost-to-serve.

6.2 Recommendations

The main recommendation for the research company is to start continuously improving the output of cost-to-serve analyses by using the framework. Currently, we completed the framework's Design and Development phase and performed a premature evaluation to show the improvement made with the new tool. So, after this thesis, the research company may distribute the new CTS tool among users so they can find opportunities, which means their CTS tool is in the Demonstration phase. Starting from that phase, Figure 6.1 shows a roadmap for the research company, which was mainly inspired by differences between the case study and the framework presented in Section 5.2. The roadmap shows recommendations for the research company related to parts of the framework.

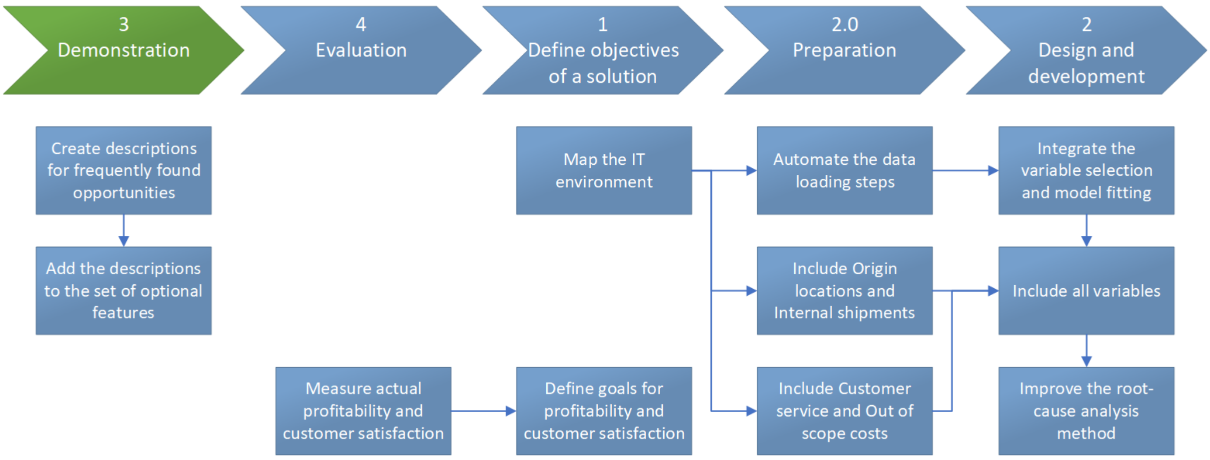


Figure 6.1: The roadmap with recommendations related to the phases of the framework for fast-moving consumer goods companies to visualize the output of a cost-to-serve analysis

The recommendations are supplementary to the advice to incorporate the framework into the research company's way of working. During the next Demonstration phase, the research company can start creating descriptions of frequently found opportunities and add these descriptions to the set of optional features, so they can potentially automate them during a future Design and Development phase. Then, in the next Evaluation phase, they should measure actual profitability and customer satisfaction to have more solid measurements on which they can base future decisions. For measuring customer satisfaction, we advise considering the method used carefully, as Section 2.2 showed that there is criticism towards the NPS method. Then, the research company can also set goals related to these measurements during the next Define Objectives of a Solution phase. Another recommendation related to that phase is to map the IT environment preceding a reconsideration of the preparation steps of the Design and Development phase. Doing so leads to three other recommendations, which relate to another execution of the preparation steps of the Design and Development phase:

1. The research company should automate the data loading process.

We advise prioritizing this, as the problem context in Section 1.3.1 showed this was the most relevant problem. Figure 6.2 illustrates a proposed structure for automating the data loading process.

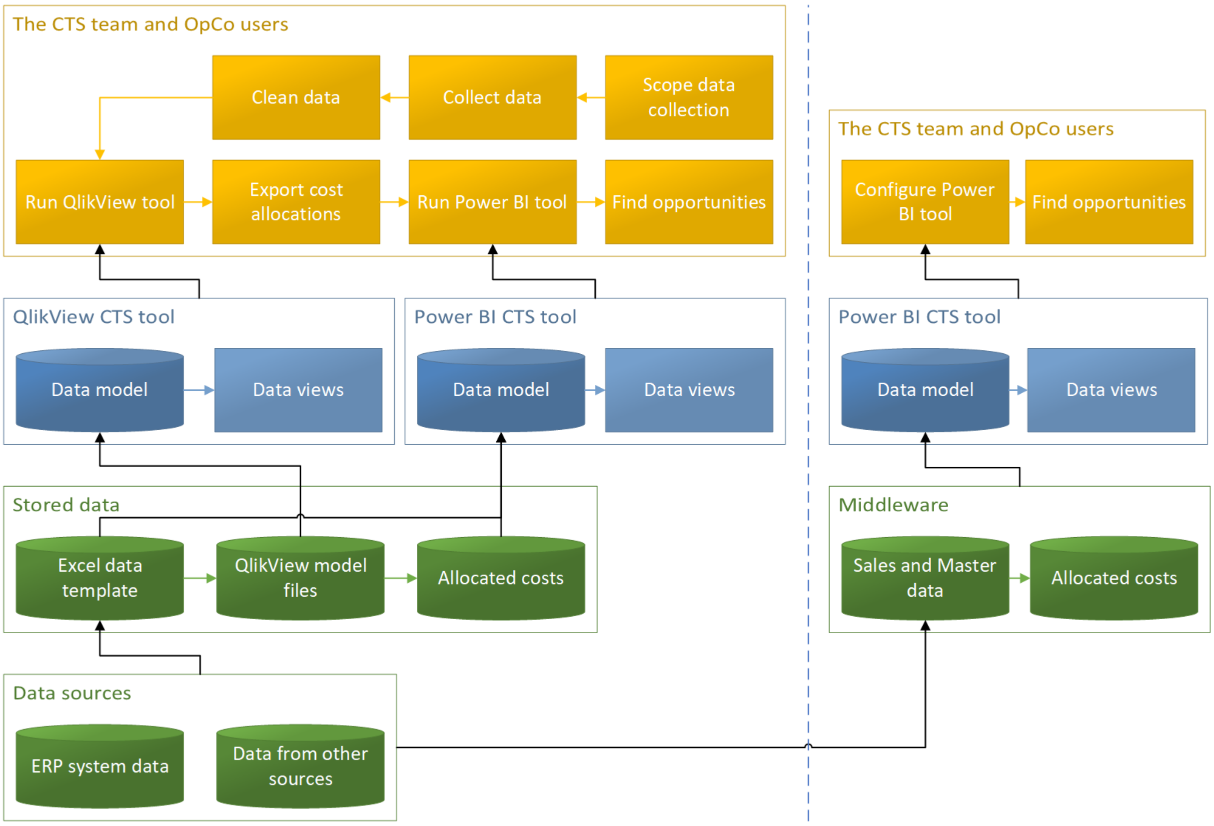


Figure 6.2: The current data connection process (left) and the proposed data collection process (right) for the research company's CTS analyses

The structure relies on the development of middleware that supplies the CTS tool in Power BI with the required source data. When OpCos have comparable data sources, it might even be possible to standardize the middleware created. Then, they can integrate the variable selection and model-fitting scripts into the tool during the Design and Development phase. By doing this, the research company will have all features incorporated in a single CTS tool. Additionally, the research company should continue to explore the possibilities in Power BI. As the software continues to develop, it becomes possible to do more with it over time. Perhaps, it becomes possible to perform a variable selection and fit models within Power BI as well. Then they should focus on the two other recommendations:

2. The research company should include tables with location-data of origins and link a table with internal shipments to order lines.
3. The research company should include Customer service and Out of Scope costs.

With these preparation steps included, the research company should aim to create all variables mentioned in this thesis during the next Design and Development phase. Then, they can review and improve the root-cause analysis method to ensure it is user-friendly and have everything included in a single CTS tool. A final recommendation is to continue developing and reviewing the framework, taking into account the evaluation of the framework presented in Section 5.2, which showed potential improvements for the framework.

6.3 Discussion

This section discusses the outcomes of this research. First, Section 6.3.1 outlines how we took into account the validity and reliability of results in this thesis. Then, Section 6.3.2 reflects on the applicability of the created framework for the research company, FMCG companies, and other types of companies. Section 6.3.3 assesses limitations and presents opportunities for future work. Finally, Section 6.3.4 presents the contributions of this research to theory and practice.

6.3.1 Validation

Cooper and Schindler (2014) describe validity as the extent to which a test measures what it aims to, whereas reliability refers to the accuracy of measurements. We often validated matters with the CTS team, and in some cases, we did this with OpCos. For example, the CTS team validated the problem cluster in Section 1.3.1, showing that it correctly displayed the problem context and was therefore valid. In another example, we ensured a reliable core problem selection using the AHP method with the company supervisor. The AHP method checks that preferences are consistent, which provides a reliable problem selection. Furthermore, the CTS team verified the opportunity typology created in Section 3.4.2, which provided an expert view on potential opportunities, validating the typology for use in practice.

In this thesis, we sent a survey to the CTS team and various tool users from OpCos to assess the value of features of the QlikView tool in Section 3.1.1, the performance of the QlikView tool in Section 4.1.2, and the desired features using diagnostic, predictive, or prescriptive analytics in Section 4.2.1. This survey had sixteen respondents, including the entire CTS team, which leads to a reliable assessment when taking average values. We applied scales with an even amount of options to force respondents to express either a negative or positive sentiment, rather than allowing respondents to choose a neutral option. Hence, decisions made based on the survey are valid as users always expressed their opinion clearly. We also shared another questionnaire with the CTS team and a single OpCo user to assess the Power BI tool in Section 5.1. Due to the low number of respondents, results from this survey are less reliable, but when we only consider the view of the CTS team, the results are reliable, as the entire team filled in both surveys. When we only take the CTS team into account, we see that the tool improved significantly, as the CTS team had a more critical view of the QlikView tool, and the OpCo respondent also evaluated the new CTS tool positively with a 50% higher score than the old CTS tool. Therefore, we assumed the new tool presented an improvement, but to make future decisions regarding the CTS tool, the research company must have more respondents from OpCos assess the CTS tool to increase reliability. Without good reliability, they could make invalid decisions regarding how to improve the CTS tool further. Additionally, the comparison of the old and new tools also showed an improvement regarding the time involved in doing mechanical operations, which does not make sense, as the Power BI tool relies on the QlikView tool to perform the cost allocations in addition to its data loading process. Perhaps, respondents did not consider loading data into the old CTS tool when giving their opinion regarding that statement.

The Power BI tool must be reliable to ensure users make business decisions validly. To ensure reliability, we compared the values in the Power BI tool with the QlikView tool, assuming values shown in the QlikView tool are correct. We loaded data from five OpCos into the Power BI tool, which showed similar values in overlapping features. So, the Power BI tool is reliable concerning the data and values it shows. We also performed the root-cause analysis for two OpCos to ensure it works in different settings and compared the variables created to similar variables shown in the QlikView tool to validate all descriptive information. However, we do not know whether the potential savings resulting from the root-cause analysis, shown in Section 4.4.5, are reliable. Therefore, the research company should attempt to make business changes based on the root-cause analysis and measure actual savings. A better model fit would also increase the reliability of the savings found. Currently, a difference between entities might result from a cost driver that is not a part of the analysis, but in general, the method is reliable as it works for different OpCos and different dimensions. However, the shown savings are questionable.

6.3.2 Applicability of the framework

Chapter 1 argued the design of a framework for FMCG companies to find business improvements using the output of a CTS analysis. We developed this framework to reduce the lack of diagnostic, predictive, and prescriptive insights in the research company's CTS tool, which we selected as the focus problem together with the CTS team. However, the problem context in Section 1.3.1 also showed other problems. In hindsight, the framework relates to all solvable core problems found. For example, in the case study, we saw again that the absence of an automated data collection is a problem, and we solved that the tool cannot handle multiple years of data. Furthermore, the research company could also solve other problems found during a phase of the framework. For example, they could focus on improving user-training during the Demonstration phase or improve the goal-setting during the Define Objectives of a Solution phase. It appears that the nature of the problem selected for this research caused the focus on the Design and Development phase. When another problem would have resulted from the problem identification, the framework would still provide a good solution, but we might have focused more on another phase.

Chapter 4 showed a successful application of the framework in the research company, which has many OpCos, and because of the decentralized nature of OpCos, we concluded the framework works for other FMCG companies as well. The intended users of the framework presented in Chapter 3 are people in FMCG companies that are responsible for managing and developing a CTS analysis. Depending on the company, this can be from a global or local position related to customer service or logistics. For example, the research company manages CTS analyses globally. If the management in the research company would move to a local level, the framework versions of the tool might diverge, and knowledge sharing would inhibit, but the framework would still apply to the research company assuming local management implies that there are multiple local users of the tool. Then, those users take the place that OpCos have in this thesis. Another difference between FMCG companies can be the scope of their CTS analysis. The framework includes cost buckets for all logistics costs, but an FMCG company could also focus on a segment of the logistics costs. For example, the focus could be on warehousing costs and include all relevant cost drivers for those costs. However, a disadvantage of decreasing the scope of the CTS analysis is that it might not measure all effects of changes. Perhaps improvements that positively influence warehouse costs negatively affect other cost buckets. In general, the scope must not move away from customer-related activities, as these are closely related to the benefits of a CTS analysis, but the framework is still applicable when we do not include one or more of the cost buckets because those buckets do not apply to the FMCG company at hand. In conclusion, the framework applies to FMCG companies regardless of their organizational structure, but to be sure of this assumption, we should test its applicability in more FMCG companies.

Besides being applicable in FMCG companies, the framework might also apply to other types of companies that want to find opportunities related to customer service optimization, customer collaboration, and a profitable pricing strategy. If companies cannot obtain any of these benefits, a CTS implementation becomes useless, so companies performing a CTS analysis should be able to make changes to the way customers are supplied, collaborate with customers, and determine a pricing strategy. Other relevant characteristics of an FMCG company are high volumes and a large degree of Make-to-Stock manufacturing. The combination of those two characteristics can complicate supply chains and make it hard to determine costs on a customer-level, hence the requirement for a CTS analysis. These characteristics can also apply to other company types than the FMCG company, such as fashion companies, original equipment manufacturers, companies selling food, or companies selling flowers. Therefore, the framework appears to apply to such companies as well, which we could also test in practice.

The framework requires the availability of reliable input data. Especially when moving to higher-level analytics, this becomes a more critical requirement due to a higher reliance on input data. The case study showed that the model-fitting process performed poorly in some cases, which might be due to data unavailability. So, companies applying a CTS analysis should focus on data availability and reliability increasingly when incorporating higher-level analytics. For example, the data handling in the research company could improve as it often involves manual work, which might lead to incorrect data. However, if the research company considered it easy to automate data connections and ensure reliable data input at all times, they would do it. So, change management is required to enable the research company to apply the framework as intended, and this might be the case for other companies as well. Alternatively, companies can strive for solutions that are as good as possible given available data, but given the importance of data in a CTS analysis, we could enrich the framework with a data management architecture.

6.3.3 Limitations and future work

The research in this thesis has several limitations, which present opportunities for future work. The previous section showed three limitations and opportunities for future work related to the applicability of the created framework. Firstly, we applied the framework to a single FMCG company in a case study. Due to the nature of the subject of the case study, we argued that the framework applies to other FMCG companies. However, future work could research whether this assumption is correct by applying the framework in several FMCG companies while measuring achieved savings and customer satisfaction. Secondly, the framework is limited to FMCG companies, but we explained that other types of companies could apply the framework as well. So, when after testing the framework in multiple FMCG companies, future work could research whether the framework applies to other types of companies similar to the case study in this thesis. Thirdly, the framework assumes that all input data is available and reliable, while the case study showed otherwise. Thus, future work can focus on the IT architecture and possible change management required to incorporate the framework by creating an ideal IT architecture and performing a gap analysis with the architectures of several companies.

In Chapter 4, we applied the framework to an FMCG company that has been performing CTS implementations for several years, but we would ideally apply it to an FMCG company that does not have CTS-related processes in place. The successful introduction of CTS analyses in an FMCG company using the framework developed in this thesis would provide a strong validation of the framework. In FMCG companies that currently perform CTS analyses, it might be unavoidable to deviate from set standards due to previously mentioned change management. Therefore, future work could set-up a CTS analysis in an FMCG company based on the framework and evaluate differences between the framework and the case study as done in this thesis in Section 5.2.

Section 3.1.1 shows a set of optional features to develop based on reviewed literature concerning potential analytics for visualizing the output of a CTS analysis and features from the QlikView tool. However, Section 4.4.4 presented several additional promising features that emerged during development, showing there are more potential features than we found in the literature. In future work, the assessment of optional features and analytics to develop could focus on other companies that perform CTS analyses besides referring to the literature to find more features that are valuable for end-users. Furthermore, future work could focus on the development of new analytics for a CTS analysis, which does not necessarily relate to a limitation in this thesis. However, Section 3.1.1 showed that only a few features focus on customer collaboration but that we can derive these from customer service optimization features. Furthermore, that section showed that there are no predictive features that focus on customer collaboration or customer service optimization, which presents more possible directions for future work.

In the root-cause analysis in Section 4.4.5, not all data was available, which possibly led to a varying model fitting performance, as explained in Section 6.3.1. Furthermore, we did not integrate the feature into the CTS tool. Future work can leverage the method designed in this thesis and test it in a case where all input data is available to obtain a more stable model fit. Additionally, future work can extend the range of models experimented with, include monthly variables to ensure seasonality is captured, and include experiments with subsets of variables to predict individual cost buckets, which might allow for better inclusion of variables showing utilization, supplier attributes, or vehicle attributes. In such future work, validating savings by assessing how many advised changes are feasible in practice would also present a significant improvement compared to the work in this thesis. Another limitation of the root-cause analysis is that it estimates potential savings for an entity by assuming that if it is comparable to another entity, the CTS per volume-unit will be about the same. However, we could improve this estimation when there is a good model fit. Future work could build on the root-cause analysis in this thesis but explore the possibility of estimating savings by predicting a new CTS per volume-unit with the fitted model. All in all, we can potentially improve the root-cause analysis in many ways, which might present a suitable topic for a new graduate intern at the research company.

6.3.4 Contribution to theory and practice

Chapter 1 started by showing that the expectations for CTS are likely to increase in the coming years. However, there is a gap in the literature explaining how FMCG companies should approach a CTS analysis. In this thesis, we combined many theories related to this subject in a framework. The connection of all elements in the framework provides a new basis for FMCG companies to develop their CTS analyses, creating a holistic overview of what is involved with visualizing the output of a CTS analysis. We linked reviewed literature from Chapter 2 related to CTS analyses with other literature and synthesized it in a framework in Chapter 3, including literature on rating tool performance, potential cost drivers in logistics, and model fitting. The framework also includes the concepts of descriptive, diagnostic, predictive, and prescriptive analytics that can support features that visualize the output of a CTS analysis. Furthermore, Section 3.4.2 showed potential benefits that result from a CTS analysis, which we supported by creating a typology containing types and sub-types, and Section 4.2.1 validated that users prefer diagnostic features rather than predictive or prescriptive features when the current CTS tool focuses on features using descriptive analytics. So, FMCG companies should develop lower-level analytics in full before focusing on higher-level analytics. We also developed a novel algorithm to find root-causes using a fitted model in Section 3.13. All in all, we did not find an approach focus on CTS analyses that compares to the framework developed in this thesis. Therefore, this is the first framework for FMCG companies to support them when visualizing the output of a cost-to-serve analysis.

The future must show whether the improvements made on the use of output a CTS analysis with the framework will lead to more OpCos continuously using CTS in the research company, but regardless of the result, we made a high contribution to their work. We implemented a system using the opportunity typology that visualizes past opportunities found in a CTS implementation in a Power BI report in Section 4.1.1, allowing the CTS team to view and leverage the past. Then, Section 4.2 clarified that the research company should focus on diagnostic analytics before moving to more advanced analytics. And finally, we used this and other knowledge to create a new tool in Power BI that visualizes the output of a CTS analysis in Sections 4.3 and 4.4. The Power BI tool is going to replace the QlikView tool used by the research company, and we created the first version of a root-cause analysis method for a high CTS per volume-unit for the research company in Section 4.4.5, which the research company can continue to develop. The framework in itself is also a contribution to the research company. We proved that the research company could apply the framework, which has led to several improvements and recommendations for the future. In conclusion, this thesis provides the research company with a structure to follow, supporting processes, and an improved tool to visualize the output of a CTS analysis.

Bibliography

Ajzen, I. et al. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16(1):125–127.

Andersen, C. M. and Bro, R. (2010). Variable selection in regression — A tutorial. *Journal of chemometrics*, 24(11-12):728–737.

Ashfaque, J. M. and Iqbal, A. (2019). Introduction to support vector machines and kernel methods. *publication at https://www.researchgate.net/publication/332370436*.

Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*, 3:19–48.

Baykasoğlu, A. and Kaplanoğlu, V. (2008). Application of activity-based costing to a land transportation company: A case study. *International journal of production economics*, 116(2):308–324.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public administration review*, 63(5):586–606.

Bokor, Z. (2010). Cost drivers in transport and logistics. *Periodica polytechnica transportation engineering*, 38(1):13–17.

Bolton, R. N. and Tarasi, C. O. (2006). Managing customer relationships. *Review of marketing research*, 3(3).

Bonoma, T. V. and Shapiro, B. P. (1984). Evaluating market segmentation approaches. *Industrial marketing management*, 13(4):257–268.

Bosshard, T. (2017). Analytics - Does your business need descriptive, diagnostic, predictive, or prescriptive capabilities? https://adviscent.com/blog/analytics-does-your-business-need-descriptive-diagnostic-predictive-or-prescriptive-capabilities.html. Accessed: 2020-10-05.

Braithwaite, A. and Samakh, E. (1998). The cost-to-serve method. *The international journal of logistics management*, 9(1):69–84.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brownlee, J. (2017). Difference between classification and regression in machine learning. https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/. Accessed: 2020-08-22.

BusinessDictionairy.com (2020). Opportunity. http://www.businessdictionary.com/definition/opportunity.html. Accessed: 2020-05-15.

Cecere, L. (2015). Seven thoughts to build a customer-centric supply chain organization. http://www.supplychainshaman.com/demand/demanddriven/seven-thoughts-to-build-a-customer-centric-supply-chain-organization/. Accessed: 2020-04-02.

Chakon, O. (2017). Practical machine learning: Ridge regression vs. lasso. https://hackernoon.com/practical-machine-learning-ridge-regression-vs-lasso-a00326371ece. Accessed: 2020-09-24.

Chen, S.-J. and Hwang, C.-L. (1992). *Fuzzy multiple attribute decision making methods*. Springer.

Christensson, P. (2008). Application definition. https://techterms.com/definition/application. Accessed: 2020-05-15.

Cooper, D. and Schindler, P. (2014). *Business research methods*. The McGraw- Hill Companies.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. PhD thesis, Massachusetts institute of technology.

Davis, M. (2010). Case study for supply chain leaders: Dell's transformative journey through supply chain segmentation. *Gartner research. ID number G*, 208603.

Elliott, T. (2013). Gartner BI: Analytics moves to the core. *Business analytics (blog), February*, 14.

Everaert, P., Bruggeman, W., Sarens, G., Anderson, S. R., and Levant, Y. (2008). Cost modeling in logistics using time-driven ABC. *International journal of physical distribution & logistics management*.

Fishbein, M. and Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research. *None*.

Fisher, N. I. and Kordupleski, R. E. (2019). Good and bad market research: A critical review of net promoter score. *Applied stochastic models in business and industry*, 35(1):138–151.

Freedman, D. A. (2009). *Statistical models: Theory and practice*. Cambridge university press.

Freeman, B., Haasz, S., Lizzola, S., and Seiersen, N. (2000). Managing your cost-to-serve. *Supply chain forum: An international journal*, 1:18–28.

GeeksforGeeks (2020). Ml — Types of learning – Supervised learning. https://www.geeksforgeeks.org/ml-types-learning-supervised-learning/. Accessed: 2020-08-22.

Guerreiro, R., Bio, S. R., and Merschmann, E. V. V. (2008). Cost-to-serve measurement and customer profitability analysis. *The international journal of logistics management*.

Heerkens, J. and van Winden, A. (2012). *Geen probleem, een aanpak voor alle bedrijfskundige vragen en mysteries*. Business school Nederland. Boekredactie.

Holm, M., Kumar, V., and Rohde, C. (2012). Measuring customer profitability in complex environments: an interdisciplinary contingency framework. *Journal of the academy of marketing science*, 40(3):387–401.

Jensen, C. S., Pedersen, T. B., and Thomsen, C. (2010). Multidimensional databases and data warehousing. *Synthesis lectures on data management*, 2(1):1–111.

Jong, Y. d. (2019). Levels of data analytics. http://www.ithappens.nu/levels-of-data-analytics/. Accessed: 2020-05-15.

Kaçan, M. (2019). 5 main drivers of supply chain costs. https://icrontech.com/blog_item/5-main-drivers-of-supply-chain-costs/. Accessed: 2020-09-05.

Kaplan, R. S. and Anderson, S. R. (2003). Time-driven activity-based costing. *Available at SSRN 485443*.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International journal of e-collaboration (ijec)*, 11(4):1–10.

Kolarovszki, P., Tengler, J., and Majerčáková, M. (2016). The new model of customer segmentation in postal enterprises. *Procedia-social and behavioral sciences*, 230:121–127.

Kone, E. R. and Karwan, M. H. (2011). Combining a new data classification technique and regression analysis to predict the cost-to-serve new customers. *Computers & industrial engineering*, 61(1):184–197.

Kuhn, M. (2019). The caret package. https://topepo.github.io/caret/. Accessed: 2020-08-23.

Kuhn, M. et al. (2008). Building predictive models in R using the caret package. *Journal of statistical software*, 28(5):1–26.

Kvålseth, T. O. (1985). Cautionary note about R 2. *The American statistician*, 39(4):279–285.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, 2(3):18–22.

Madden, T. J., Ellen, P. S., and Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and social psychology bulletin*, 18(1):3–9.

Nakajima, S. (1988). Total productive maintenance. Introduction to TPM. *Productivity press 1988.*

Özener, O. Ö., Ergun, Ö., and Savelsbergh, M. (2013). Allocating cost of service to customers in inventory routing. *Operations research*, 61(1):112–125.

Pandey, D., Suman, U., and Ramani, A. (2010). An effective requirement engineering process model for software development and requirements management. *2010 International conference on advances in recent technologies in communication and computing*, pages 287–291.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.

Pohlen, T. L. and La Londe, B. J. (1994). Implementing activity-based costing (ABC) in logistics. *Journal of business logistics*, 15(2):1.

Poole, C. (2017). Supply chain brief: Heinekens journey to improved customer service and collaboration.

Poole, D. (2020). A better way for service to predict future customer loyalty. https://www.gartner.com/smarterwithgartner/a-better-way-for-service-to-predict-future-customer-loyalty/.

ProductPlan (2020). Roadmap basics. https://www.productplan.com/roadmap-basics/. Accessed: 2020-05-15.

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, 81(12):46–55.

Ross, A., Jayaraman, V., and Robinson, P. (2007). Optimizing 3pl service delivery using a cost-to-serve and action research framework. *International journal of production research*, 45(1):83–101.

Rouhani, B. D., Mahrin, M. N., Nikpay, F., Ahmad, R. B., and Nikfard, P. (2015). A systematic literature review on enterprise architecture implementation methodologies. *Information and software technology*, 62:1–20.

Saaty, T. L. (2000). *Fundamentals of decision making and priority theory with the analytic hierarchy process*, volume 6. RWS publications.

Schober, P., Boer, C., and Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768.

Seif, G. (2018). 5 types of regression and their properties. https://towardsdatascience.com/5-types-of-regression-and-their-properties-c5e1fa12d55e. Accessed: 2020-08-22.

Singh, R., Gohil, A. M., Shah, D. B., and Desai, S. (2013). Total productive maintenance (TPM) implementation in a machine shop: A case study. *Procedia engineering*, 51:592–599.

Stapleton, D., Pati, S., Beach, E., and Julmanichoti, P. (2004). Activity-based costing for logistics and marketing. *Business process management journal.*

Sun, L., Karwan, M. H., Gemici-Ozkan, B., and Pinto, J. M. (2015). Estimating the long-term cost to serve new customers in joint distribution. *Computers & industrial engineering*, 80:1–11.

Svetnik, V., Liaw, A., Tong, C., and Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. *International workshop on multiple classifier systems*, pages 334–343.

Taherdoost, H. (2018). A review of technology acceptance and adoption models and theories. *Procedia manufacturing*, 22:960–967.

Tohamy, N. (2020). Hype cycle for supply chain strategy, 2020.

Triantaphyllou, E. (2000). *Multi-criteria decision making methods: A comparative study*. Springer.

Tripathi, N. (2020). What is customer satisfaction score? https://www.surveysensum.com/blog/what-is-customer-satisfaction-score-definition-calculation-applications-and-advantages/. Accessed: 2020-10-05.

Turney, P. B. (1992). Activity based costing. *Management accounting handbook(4th Edition), edited by C. Drury, Butterworth-Heinemann and CIMA.*

Vala, K. (2019). Tree-based methods: Regression trees. https://towardsdatascience.com/tree-based-methods-regression-trees-4ee5d8db9fe9. Accessed: 2020-09-27.

van Niekerk, H. J. and Bean, W. (2019). Evaluation of the cost drivers and allocation framework in outbound logistics of the fast-moving consumer goods industry. *South African journal of industrial engineering*, 30(2):115–130.

Varila, M., Seppänen, M., and Suomala, P. (2007). Detailed cost modelling: A case study in warehouse logistics. *International journal of physical distribution & logistics management.*

Veness, C. (2020). Calculate distance, bearing and more between latitude/longitude points. https://www.movable-type.co.uk/scripts/latlong.html. Accessed: 2020-10-09.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478.

Wang, A., Arbogast, J. E., Bonnier, G., Wilson, Z., and Gounaris, C. E. (2020). Estimation of marginal cost to serve individual customers.

Watts, S. (2017). Etl basics: Extract, Transform & Load. https://www.bmc.com/blogs/what-is-etl-extract-transform-load-etl-explained/. Accessed: 2020-10-02.

Wind, Y. and Saaty, T. L. (1980). Marketing applications of the analytic hierarchy process. *Management science*, 26(7):641–658.

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National science review*, 5(1):44–53.

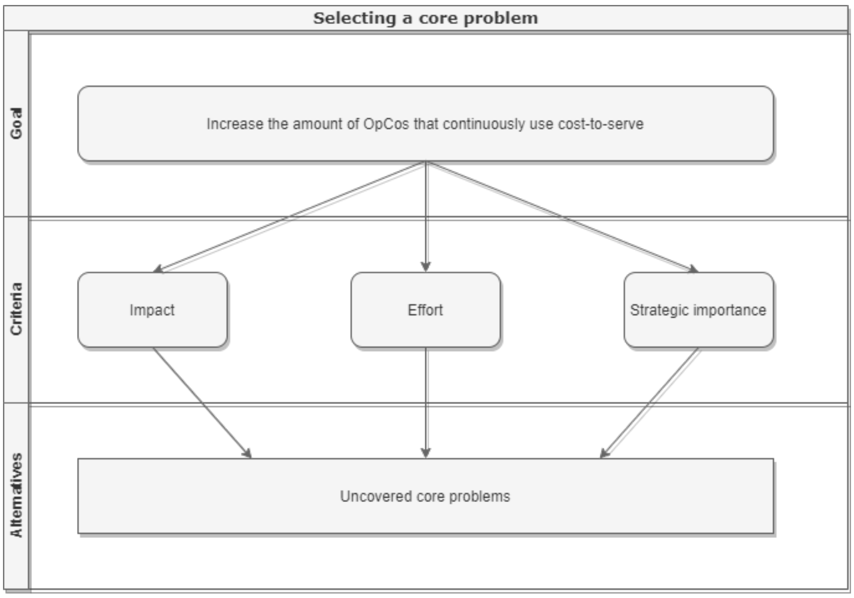
Appendices

A Selection of the core problem

Based on the AHP, the following steps have been applied:

1. Build a hierarchy for a decision (goal, criteria, alternatives)
2. Establish a value scale
3. Determine consistent weights for the criterion
4. Score the alternatives
5. Select a problem

We applied the steps in correspondence with the problem owner from the research company, which is the company supervisor of the researcher. The first step is creating the decision hierarchy that consists of a goal, criteria, and alternatives are determined based on the problem context and shown in the figure below.



The criteria chosen are impact, effort, and strategic importance. These criteria are defined as shown in the table below.

Criterion	Description
Impact	How much impact can solving this problem have on the problem; "Too few OpCos continuously use cost-to-serve"
Effort	How much effort is involved with solving this problem
Strategic importance	What is the strategic importance of solving this problem? Taking into account the research company, the CS team, and the CTS team.

Second, we determined a value scale to compare criteria. For this, we used the value scale shown in the table below.

Judgment (preference)	Rating
Extremely	9
Very-Extremely	8
Very strongly	7
Strongly-Very	6
Strongly	5
Moderately-Strongly	4
Moderately	3
Equally-Moderately	2
Equally	1

We compare each criterion, indicating which one holds a preference over the other. The rating is used in the next step to determine whether the problem owner is consistent. The outcome is that impact and strategic importance are equally important and are both strongly preferred over effort. The third step is to determine consistent weights for the criteria.

Next, the logic behind checking consistency and determining the weights is shown. All pairs combined, the criteria preferred strongly over the other scores a 5:

Problems	Impact	Effort	Strategic importance
Impact	1	5	1
Effort	0.2	1	0.2
Strategic importance	1	5	1
Sum	**2.2**	**11**	**2.2**

Normalize the table and calculate the priority (or weight) as the average of the row:

Problems NORM	Impact	Effort	Strategic importance	**Priority**
Impact	0.4545	0.4545	0.4545	**0.4545**
Effort	0.0909	0.0909	0.0909	**0.0909**
Strategic importance	0.4545	0.4545	0.4545	**0.4545**

The α is calculated by dividing the sum of the row by the average:

Problems	Impact	Effort	Strategic importance	Sum	α
Impact	0.4545	0.4545	0.4545	1.3636	**3**
Effort	0.0909	0.0909	0.0909	0.2727	**3**
Strategic importance	0.4545	0.4545	0.4545	1.3636	**3**

$criteria = 1, 2, 3,$ (Impact, Effort, Strategic importance)

$\lambda_{max} = \frac{\sum \alpha_{criteria}}{criteria} = 3$

Consistency Index $C.I. = \frac{\lambda_{max} - criteria}{criteria - 1} = 0$

The Random consistency Index (R.I.) from Wind and Saaty (1980) is taken:

Matrix size	Random consistency index (RI)
1	0.00
2	0.00
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49

Here we find $R.I. = 0.58$

Consistency Ratio $C.R. = \frac{c.I.}{R.I.} = 0$

As $C.R. < 0.10$, the criteria have been rated consistently.

As we see consistency, we determine the importance of each criterion by its weight. The table below shows these.

Criterion	Weight
Impact	0.4545
Effort	0.0910
Strategic importance	0.4545

The fourth step is to score the alternatives. We performed an interview with the problem owner and ranked problems relative to each other. In all cases, the highest score indicates the problem scores the best for the criterion. So, this means the highest impact, strategic importance, and the lower effort required. The table below shows the results of the interview with the numbers of core problems. We added a short description of the impact criterion for a better understanding.

Rank	Impact	Effort	Strategic importance
1	Problem 2 : Automated data connection	7	2
2	Problem 9 : Mostly descriptive	8	9
3	Problem 11: User training	4	11
4	Problem 6 : Goals unclear	6	6
5	Problem 4 : Data estimations	11	8
6	Problem 8 : Multiple year data	9	4
7	Problem 7 : Output shipment level	2	7

In the table, for example, Problem 2 has the highest impact. Whereas Problem 7 shows the lowest amount of effort. The problem with rank one scores 7 points for a given criterion, and the problem with rank seven receives 1 point. All others receive a score between 1 and 7. Combining all scores relative to each other in a matrix, which is then normalized, we calculated the preferred problems based on the different criteria. The table below shows the process.

Problem	Impact	Effort	Strategic importance
2	0.25	0.04	0.25
4	0.11	0.18	0.07
6	0.14	0.14	0.14
7	0.04	0.25	0.04
8	0.07	0.21	0.11
9	0.21	0.07	0.21
11	0.18	0.11	0.18

Then, we multiply the normalized scores above by previously calculated weights. Finally, the sum for each problem shows an overall priority.

B Survey used features and graphs

Used features

Response	Score
Never used	0
Rarely used	1
Often used	2
Always used	3

ID	Position	Experience	Filter functionality (being able to select specific dimensions)	Graphs (the tab with all the graphs)	Validation tables (the tab with tables related to cost allocations)	Reporting tables (the tab with P&L input data)	Scenario: Change delivery location	Scenario: MOQ	Scenario: Changing delivery day	Scenario graphs (the tab visualizing scenario outcomes)	Map graph (the tab where the map can be viewed)	Detail tables (the tab with tables on shipments, costs, customers, and products)
1	OpCo	2 - 3 years	3	3	2	1	0	0	0	0	1	2
2	OpCo	3 - 4 years	3	3	3	3	1	1	1	2	1	3
3	OpCo	More than 4 years	3	3	3	0	1	1	0	1	1	3
4	OpCo	3 - 4 years	2	2	1	1	2	1	0	2	0	1
5	OpCo	Less than 1 year	2	1	2	2	0	0	0	0	0	2
6	Global	1 - 2 years	3	3	1	0	1	2	1	2	1	3
7	Global	2 - 3 years	2	2	2	2	0	0	0	0	1	2
8	OpCo	1 - 2 years	3	3	2	2	3	3	2	2	1	3

ID	Position	Experience	Filter functionality (being able to select specific dimensions)	Graphs (the tab with all the graphs)	Validation tables (the tab with tables related to cost allocations)	Reporting tables (the tab with P&L input data)	Scenario: Change delivery location	Scenario: MOQ	Scenario: Changing delivery day	Scenario graphs (the tab visualizing scenario outcomes)	Map graph (the tab where the map can be viewed)	Detail tables (the tab with tables on shipments, costs, customers, and products)
9	OpCo	1 - 2 years	3	3	3	3	3	3	3	3	1	3
10	OpCo	1 - 2 years	3	3	3	3	3	3	3	3	1	3
11	Global	1 - 2 years	3	3	1	0	0	0	0	0	0	2
12	OpCo	Less than 1 year	2	3	1	1	1	2	0	0	0	1
13	OpCo	1 - 2 years	3	3	2	3	0	0	0	0	3	2
14	OpCo, Global	3 - 4 years	3	2	2	0	0	0	0	2	1	1
15	OpCo	1 - 2 years	3	3	1	3	0	1	0	0	0	3
16	OpCo	Less than 1 year	3	2	2	1	0	0	0	0	0	3

Used graphs

The results show how many stars each respondent awarded.

ID	Cost per volume-unit	Trend - Cost per volume-unit	Delivery profile	Volume	Scatter / Drop Analysis	Dropsize / Cost Analysis	Trend - Dropsize / Cost Analysis	Pareto Analysis
1	4	2	4	4	4	3	4	2
2	4	4	4	3	4	4	4	3
3	4	4	4	3	4	4	3	3
4	3	3	4	3	4	4	4	3
5	3	1	3	3	3	3	3	3
6	4	4	4	2	4	3	4	1
7	4	3	4	3	4	3	2	2
8	4	4	4	3	3	4	4	2
9	4	4	4	4	4	4	4	3
10	4	4	4	4	4	4	4	4
11	4	3	4	1	4	1	1	2
12	4	2	2	3	4	3	2	3
13	4	3	2	4	4	2	2	1
14	3	3	3	3	4	3	3	2
15	4	1	4	3	3	4	1	1
16	4	4	3	2	4	3	3	2

C Variables root-cause analysis

Ordering behavior	Average/Min/Max/SD/CV
Weeks of demand	Drops/week
Weeks with drops	Deliveries/week
Weeks without drops	Deliveries/drop
	Products/week
	Products/drop
	Products/delivery
	Order lines/week
	Order lines/drop
	Order lines/delivery
	Order lines/product
	Volume/week
	Volume/drop
	Volume/delivery
	Volume/product
	Volume/order line

Route to market and distance	Average/Min/Max/SD/CV
Percentage of drops per shipping location	Shipping locations/week
Percentage of deliveries per shipping location	Shipment types/week
Percentage of products per shipping location	PWs/week
Percentage of order lines per shipping location	PWs/drop
Percentage of volume per shipping location	PWs/delivery
Percentage of drops per shipment type	PWs/product
Percentage of deliveries per shipment type	Paths/week
Percentage of products per shipment type	Paths/drop
Percentage of order lines per shipment type	Paths/delivery
Percentage of volume per shipment type	Paths/product
Percentage of order lines per PW	Distance/week
Percentage of volume per PW	Distance/drop
Percentage of order lines PATH	Distance/delivery
Percentage of volume PATH	Distance/product
	Distance/order line
	Distance/volume

Picking
Percentage of weeks with loose picking
Percentage of drops with loose picking
Percentage of deliveries with loose picking
Percentage of products with loose picking
Percentage of order lines with loose picking
Percentage of volume with loose picking
Loose case picking percentage in pallets
Loose case picking percentage (>1 pallet)
Loose case picking percentage (>2 pallet)
Loose case picking percentage (>3 pallet)
Loose case picking percentage (>... pallet)

Drop days	Average/Min/Max/SD/CV
Percentage of drops per drop day	Drops per day Time between drops

Other drivers
Origin
Origin - attributes that can differ per order line
Supplier
Supplier - attributes that can differ per order line
Vehicle
Vehicle - attributes that can differ per order line
Utilization
Area used

D Opportunity typology

Category	Type	Sub-type	Definition
Customer service optimization	General	General	Focus on optimizing the supply chain
Customer service optimization	Data quality	General	Improve the quality of data which is available
Customer service optimization	Network design	General	Changes the transportation network
Customer service optimization	Network design	Distribution footprint (Customer-DC-Plant allocations)	Change sourcing locations for customers
Customer service optimization	Network design	Route-to-Market (distribution)	Change the distribution strategy
Customer service optimization	RPM	General	Improve the Returnable Products Management
Customer service optimization	Transport optimization	General	Optimize transportation activities
Customer service optimization	Transport optimization	Backhauling	Decreasing the amount empty trucks on return trips
Customer service optimization	Transport optimization	Direct deliveries	Deliver to customer directly from the Production Warehouse
Customer service optimization	Transport optimization	Drop efficiency	Simplifying drops at customers
Customer service optimization	Transport optimization	LSP management	Changing the use of Logistic Service Providers
Customer service optimization	Transport optimization	Pallet configuration	Changing the configuration of pallets
Customer service optimization	Transport optimization	Stock balancing	Optimize stock via inter-unit replenishments
Customer service optimization	Transport optimization	Truck allocation	Changing the use of the truck fleet
Customer service optimization	Transport optimization	Truck efficiency	Improve the vehicle routing
Customer service optimization	Warehouse optimization	General	Optimize warehousing activities
Customer collaboration	General	General	Focus on optimizing collaboration with customers

Category	Type	Sub-type	Definition
Customer collaboration	Logistic trade terms	General	Discounts for customers focusing on improving logistics
Customer collaboration	Logistic trade terms	Customer pick-up	Customers collect their own orders
Customer collaboration	Logistic trade terms	Delivery frequency	Changing the delivery frequency
Customer collaboration	Logistic trade terms	Minimum order threshold	Establishing or changing the minimum order threshold
Customer collaboration	Logistic trade terms	Trade term portfolio	Changing the Logistic Trade Terms portfolio
Customer collaboration	OTC	General	Improving the Order To Cash process
Customer collaboration	SLAs	General	Service Level Agreements made with customers
Customer collaboration	SLAs	Customer collaboration	Change customer collaboration strategy
Customer collaboration	SLAs	Delivery days	Change the amount of delivery days in a period
Customer collaboration	SLAs	Penalty reduction	Change the amount of penalty charged to customers
Profitable pricing strategy	General	General	Changing prices or discarding products
Profitable pricing strategy	Commercial trade terms	General	Discounts for customers focusing on commerce
Profitable pricing strategy	Commercial trade terms	Peak shaving	Stabilizing customer demand (avoiding peaks)
Profitable pricing strategy	Product portfolio	General	Changing the product portfolio

E Opportunities Power BI report

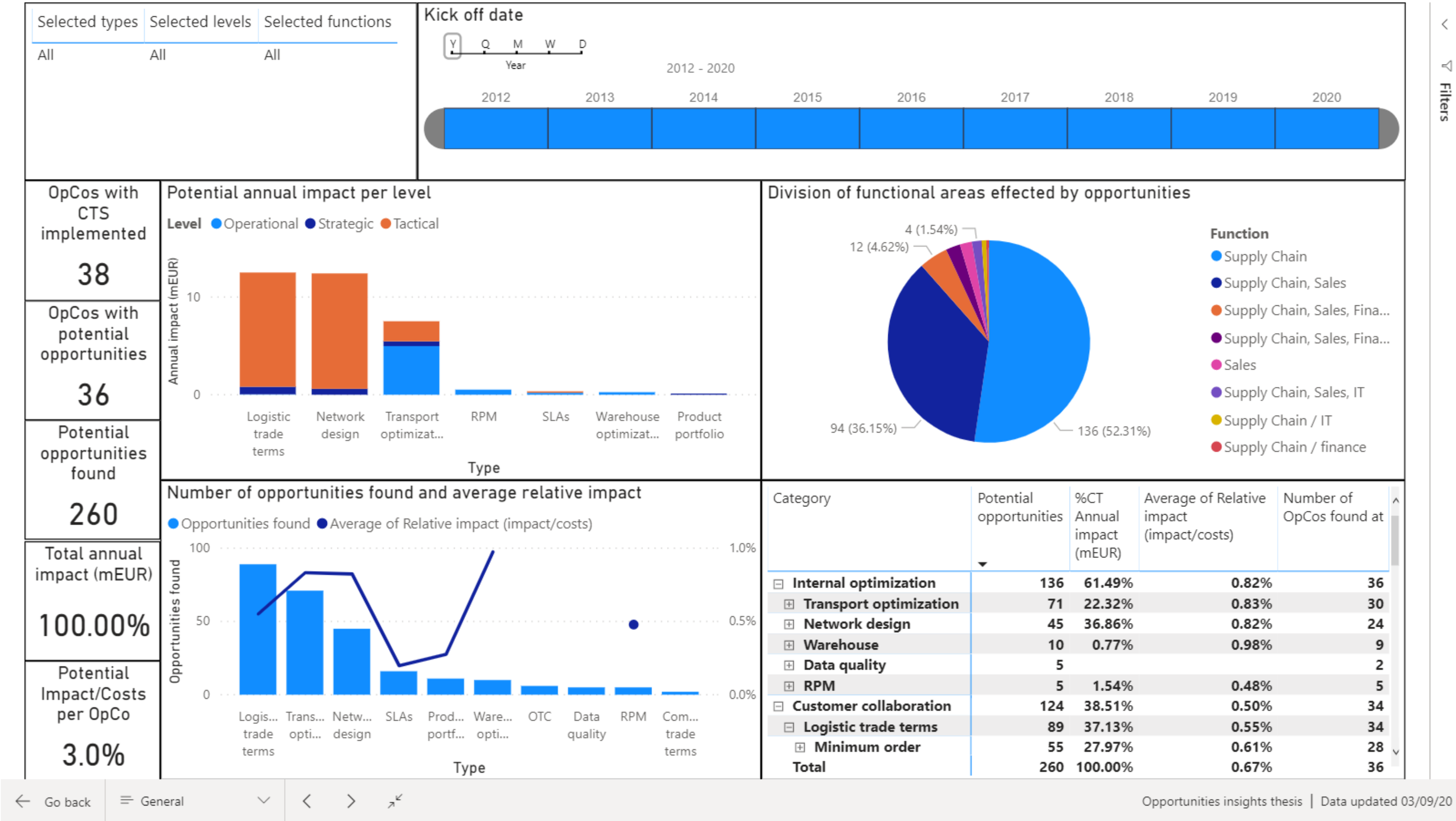


Figure 3: A screenshot of the general tab of the opportunities dashboard with marginalized annual impact

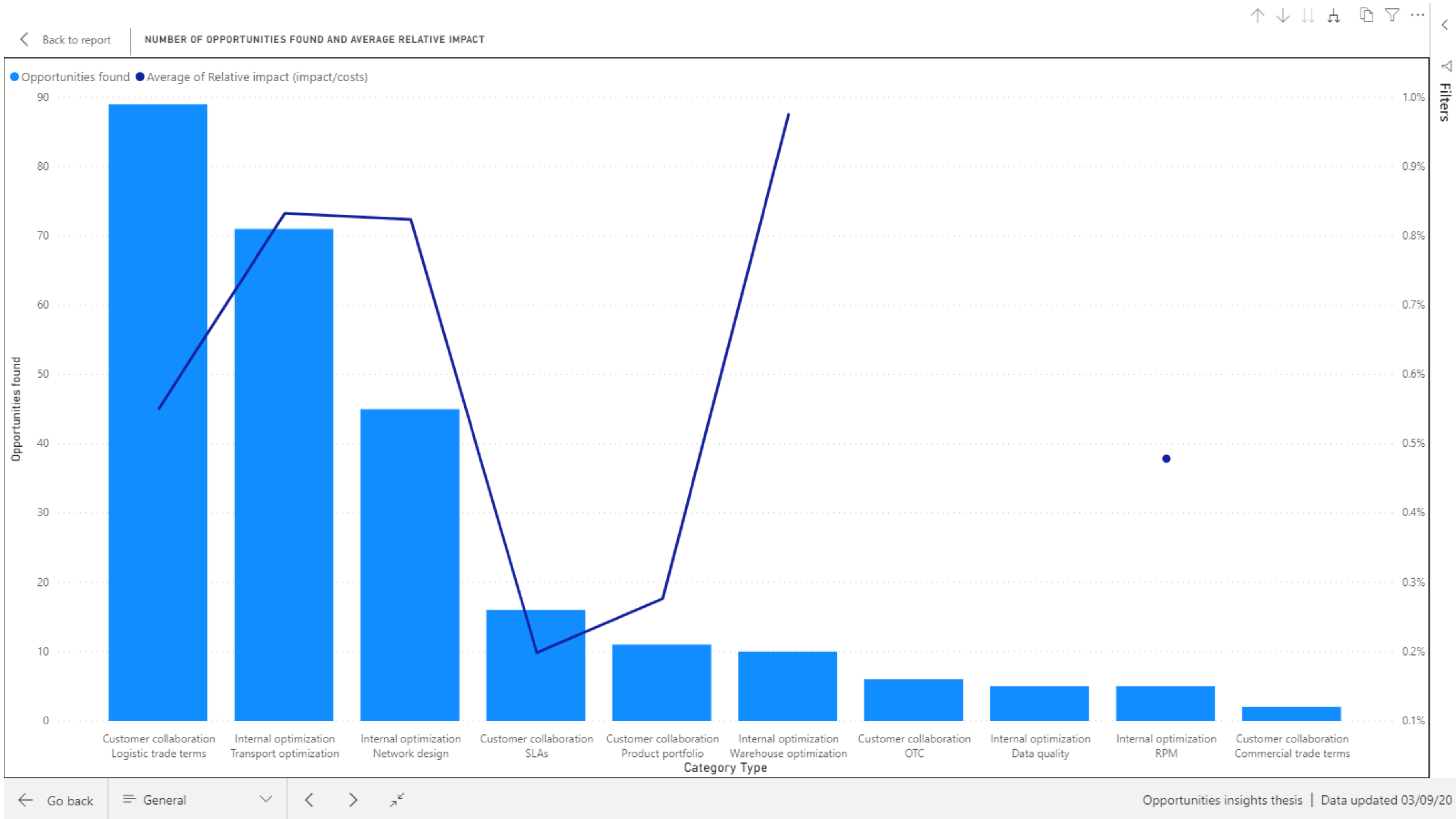


Figure 4: A screenshot showing the number of opportunities of each type that have been found the most in past years including their average relative impact (right y-axis)

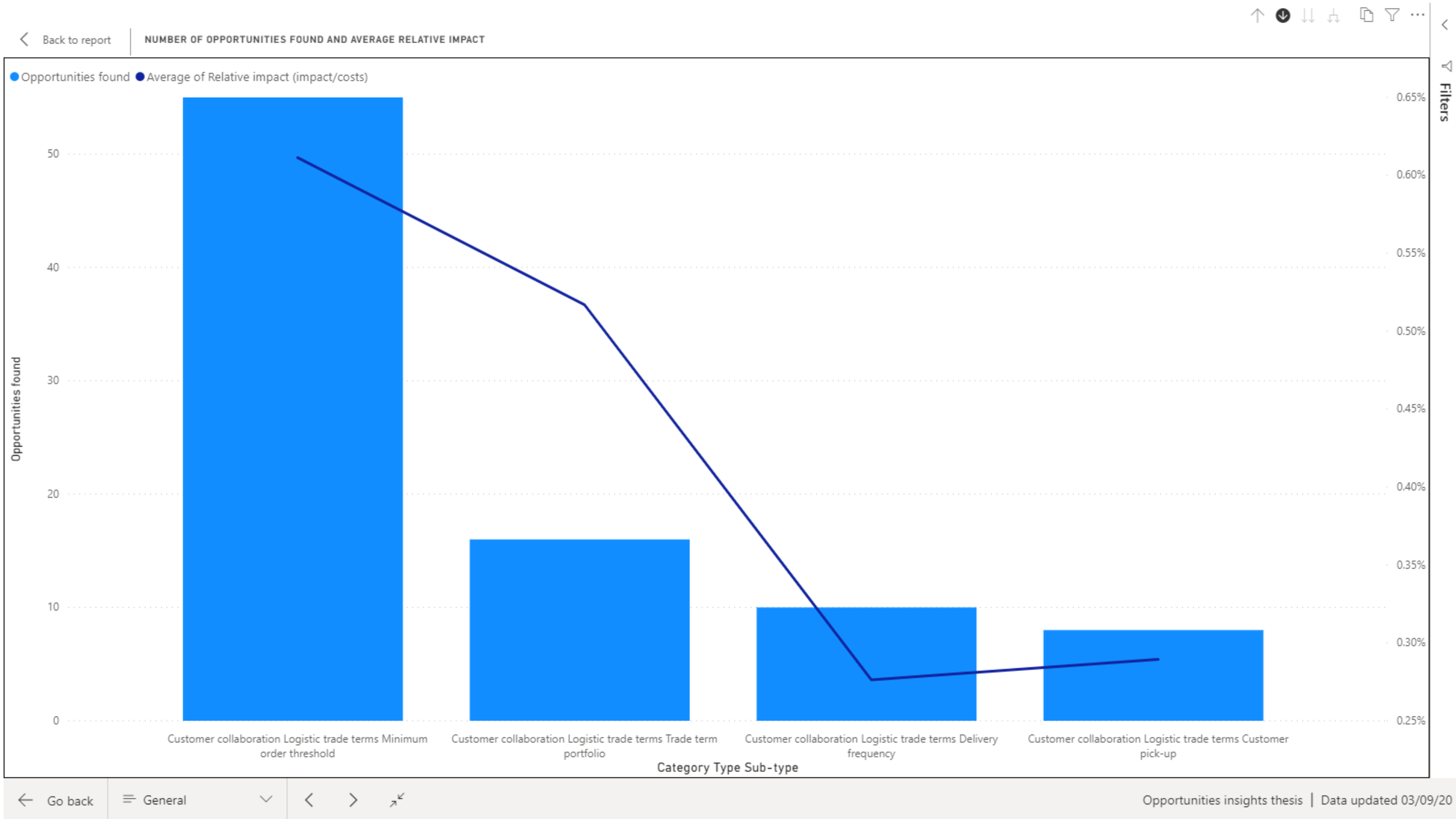


Figure 5: A screenshot showing the number of opportunities of each sub-type that have been found under the Logistics Trade Terms Types, including their average relative impact (right y-axis)

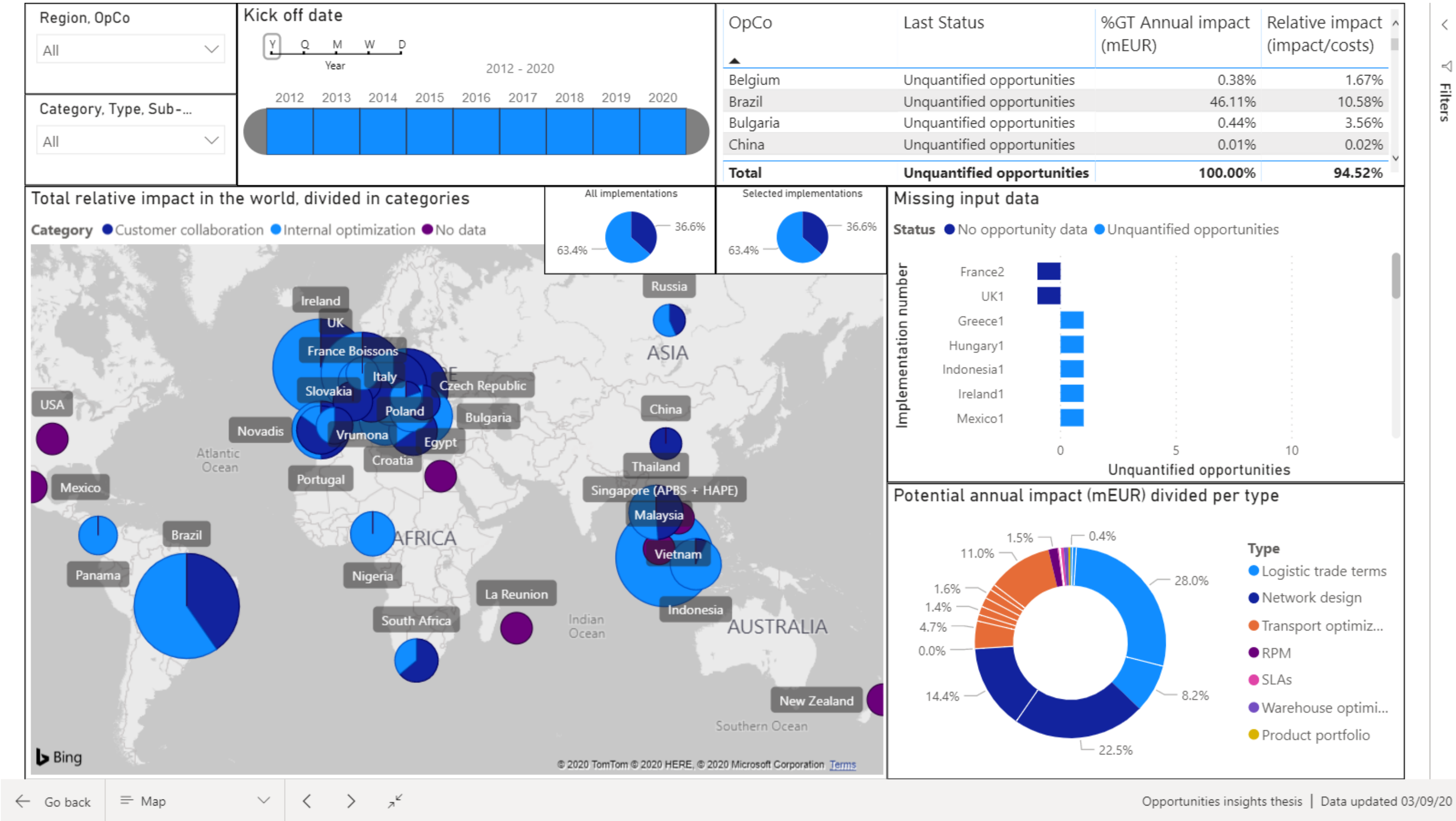


Figure 6: A screenshot of the map tab of the opportunities dashboard with marginalized annual impact

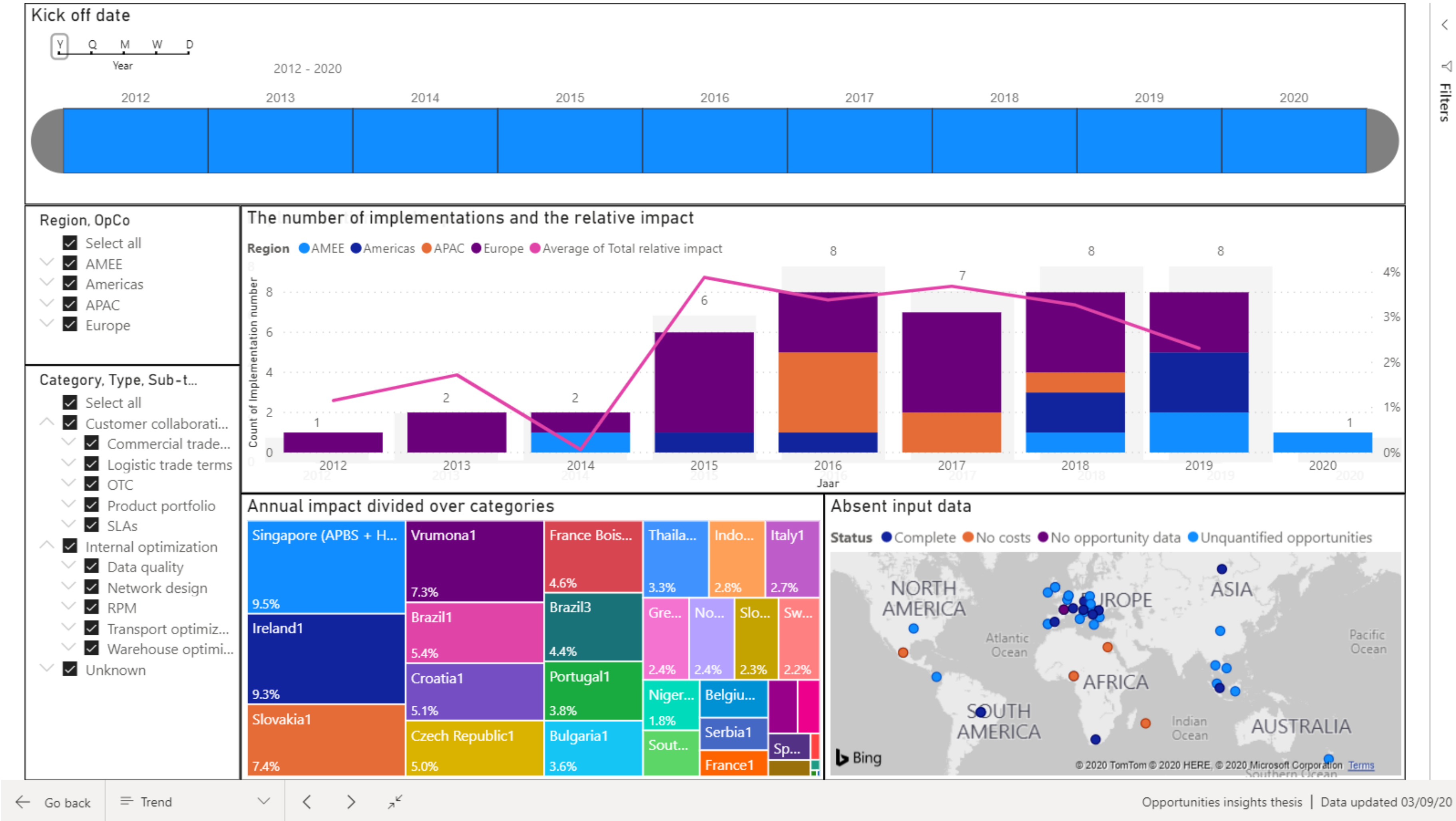


Figure 7: A screenshot of the trend tab of the opportunities dashboard with marginalized annual impact

Two opportunity types stand out. Firstly, the research company often found Minimum Order Threshold opportunities. Secondly, they found a wide range of Customer Service Optimization opportunities with a high average impact. The figure below shows a comparison between the two opportunity types. The Minimum Order Threshold is a sub-type where Customer Service Optimization is a category, but in both cases a relatively large portion of the opportunities found are being considered.

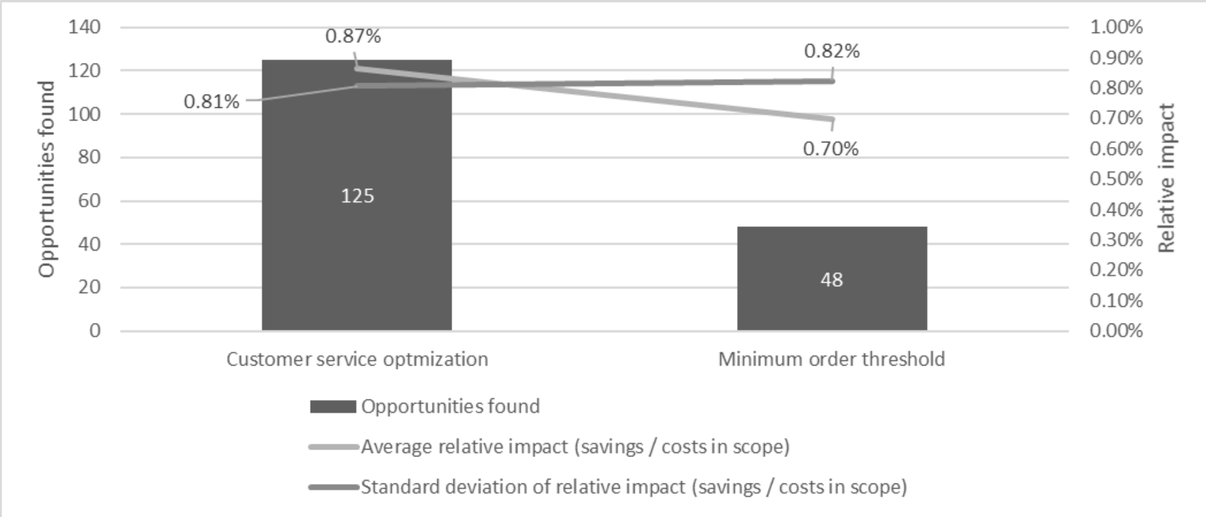


Figure 8: A comparison between Minimum order threshold and Internal optimization opportunities found in the research company

Minimum order thresholds opportunities are more specific than Customer service optimization opportunities, but the standard deviation of Minimum order threshold opportunities is higher than the mean, indicating that the relative savings vary heavily between opportunities found. Customer Service Optimization opportunities made the most impact. Therefore, the figure below takes a close look at the types and sub-types of the Customer Service Optimization category.

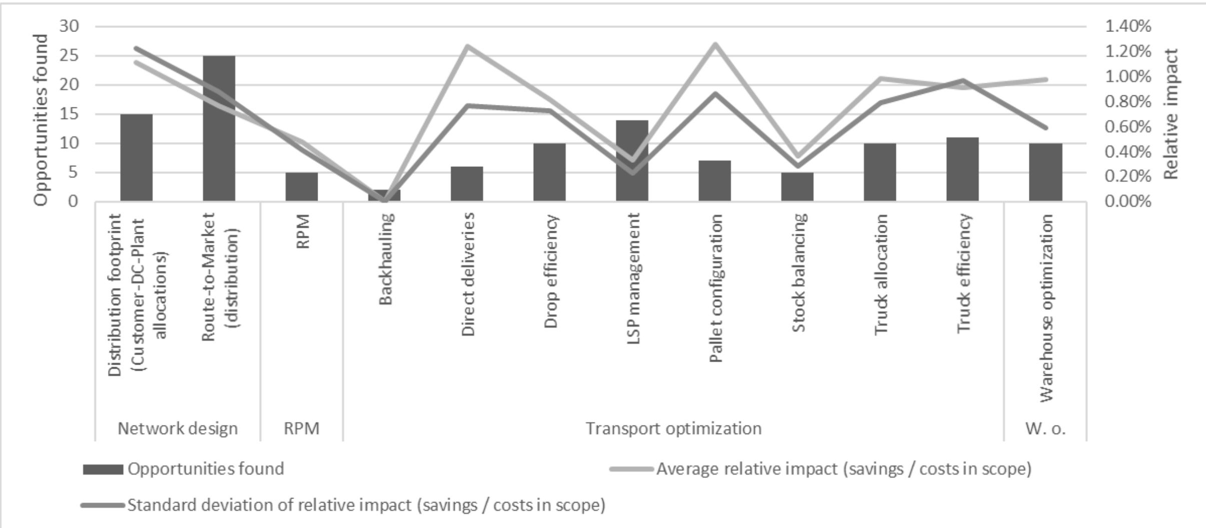


Figure 9: A comparison between the types and sub-types of Customer service optimization opportunities found in the research company

The impact of Customer Service Optimization opportunities varies over different sub-types. Furthermore, the average and standard deviation values for the sub-types differ as well. All in all, it seems that there is a wide range of opportunities in this category. In conclusion, the measurement of success of the research company provided various insights into potential savings, emphasizing the impact differs between types of opportunities. However, due to the missing input data, it is difficult to determine specific actions.

F Survey performance current tool

Response	Score
Completely disagree	0
Mostly disagree	1
Somewhat disagree	2
Somewhat agree	3
Mostly agree	4
Completely agree	5

ID	Position	Experience	Using the tool improves my job performance.	Using the tool makes it easier to do my job.	Using the tool significantly increases the quality of output on my job.	Learning to operate the tool was easy for me.	Using the tool involves little time doing mechanical operations (e.g., data input).	I find it easy to get the tool to do what I want it to do.
1	OpCo	2 - 3 years	5	4	4	2	4	1
2	OpCo	3 - 4 years	4	4	4	4	2	2
3	OpCo	More than 4 years	4	4	4	2	2	2
4	OpCo	3 - 4 years	4	4	2	2	1	4
5	OpCo	Less than 1 year	4	2	4	2	0	2
6	Global	1 - 2 years	4	4	4	2	1	2
7	Global	2 - 3 years	1	1	1	1	1	1
8	OpCo	1 - 2 years	2	2	4	4	2	4
9	OpCo	1 - 2 years	4	4	4	4	4	4
10	OpCo	1 - 2 years	4	4	4	4	4	4
11	Global	1 - 2 years	2	2	2	0	0	0
12	OpCo	Less than 1 year	4	2	2	2	0	4
13	OpCo	1 - 2 years	4	4	2	2	4	4
14	OpCo, Global	3 - 4 years	2	2	2	2	2	2
15	OpCo	1 - 2 years	4	4	2	4	1	2
16	OpCo	Less than 1 year	4	2	5	2	4	2

G Contents of the current tool

To categorize the dimensions, we used the categories shown in the table below. These dimensions often serve as either a row, column, axis, or attribute. Usually, a combination can offer a specific view of the data, but in some cases the CTS of a single dimension is shown.

Table 5: The categories used to map dimensions of analytics in the current tool

Category	Description
Time	Shows time
Shipments	Shows information related to shipments
Customers	Shows customer attributes
Products	Shows product attributes
Sources	Shows sourcing characteristics
Vehicles	Shows vehicle attributes
CTS buckets	Shows the CTS buckets
KPIs	Shows KPIs (e.g. # drops, # shipments)

Often, features are adjustable and can, therefore, show a different axis based on a user preference. Furthermore, features usually measure CTS or volumes. The unit often differs for volumes-measurements.

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
All	Filters	Filter	None	1	1	1	1	1	1	0	0
All	Current selections	Table	None	1	1	1	1	1	1	1	0
All	Key numbers summary - 1	Table	Volumes	0	0	0	0	0	0	0	1
All	Key numbers summary - 2	Table	CTS	0	0	0	0	0	0	1	0
Overview	Volume flow	Diagram	CTS and volumes	0	0	0	0	0	0	0	0
Overview	Cost group	Table	CTS	0	0	0	0	0	0	1	0

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Key numbers	Key numbers - Overview	Table	CTS and volumes	0	1	0	0	0	0	1	1
Key numbers	Key numbers - CTS per dimension	Table	CTS and volumes	1	1	1	1	1	1	1	1
Graphs	Cost per volume-unit	Bar/line chart	CTS and volumes	1	1	1	1	1	1	1	0
Graphs	Trend - cost per volume-unit	Bar/line chart	CTS and volumes	1	0	0	0	0	0	1	0
Graphs	Delivery profile	Bar/line chart	CTS and volumes	1	1	1	1	1	1	1	1
Graphs	Volume - 1	Pie chart	Volumes	0	0	1	0	0	0	0	0
Graphs	Volume - 2	Pie chart	Volumes	0	0	0	0	1	0	0	0
Graphs	Volume - 3	Pie chart	Volumes	0	0	1	0	0	0	0	0
Graphs	Volume - 4	Pie chart	Volumes	0	0	0	1	0	0	0	0
Graphs	Volume - 5	Pie chart	Volumes	0	0	0	1	0	0	0	0
Graphs	Volume - 6	Pie chart	Volumes	0	0	0	1	0	0	0	0
Graphs	Volume - Cases	Stacked bar/line chart	Volumes	1	1	1	1	1	1	0	1
Graphs	Cost of goods sold	Stacked bar/line chart	CTS and volumes	1	1	1	1	1	1	0	0

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Graphs	Scatter / drop analysis	Scatterplot	Volumes	1	1	1	1	1	1	0	1
Graphs	Dropsize / cost analysis	Stacked bar/line chart	CTS and volumes	1	1	1	1	1	1	0	0
Graphs	Trend - dropsize / cost analysis	Stacked bar/line chart	CTS and volumes	1	0	0	0	0	0	0	0
Graphs	Pareto analysis	Stacked bar/line chart	CTS and volumes	0	0	0	1	1	0	0	0
Graphs	SKU profitability	Scatterplot	CTS and volumes	0	0	0	1	0	0	1	0
Graphs	Waterfall	Waterfall chart	CTS	0	0	0	0	0	0	1	0
Validation	Volume & orders - Volume - 1	Table	Volumes	1	1	1	1	1	1	0	0
Validation	Volume & orders - Volume - 2	Table	Volumes	0	0	0	1	0	0	0	0
Validation	Volume & orders - Orders	Table	Volumes	0	0	1	1	1	0	0	0
Validation	Volume & orders - Weight ranges	Stacked bar chart	Volumes	0	1	1	0	0	0	0	0

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Validation	Volume & orders - WH handling	Table	Volumes	0	0	0	0	1	0	0	0
Validation	Cost - inter unit - 1	Table	CTS and volumes	0	0	0	0	1	0	0	0
Validation	Cost - inter unit - 2	Table	CTS and volumes	0	0	0	0	1	0	0	0
Validation	Cost - inter unit - 3	Table	CTS and volumes	0	0	0	0	1	0	0	0
Validation	WH handling cost - Summary	Table	CTS	1	1	1	1	1	1	0	0
Validation	WH handling cost - Storage - 1	Table	CTS and volumes	0	0	0	1	1	0	0	0
Validation	WH handling cost - Storage - 2	Table	CTS and volumes	0	1	0	1	1	0	0	0
Validation	WH handling cost - Transfer	Table	CTS	0	1	0	0	1	0	0	0
Validation	WH handling cost - Customer	Table	CTS and volumes	0	1	0	0	0	0	0	0
Validation	WH handling cost - Return	Table	CTS	0	1	0	0	0	0	0	0

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Validation	WH handling cost - Conditions - 1	Table	CTS	0	0	0	0	0	0	1	0
Validation	WH handling cost - Conditions - 2	Table	CTS	0	0	1	1	1	0	1	0
Validation	WH handling cost - Plant handling	Table	CTS and volumes	0	0	0	0	1	0	0	0
Validation	Outbound freight - Primary delivery - 1	Table	CTS and volumes	0	1	0	0	1	0	0	0
Validation	Outbound freight - Primary delivery - 2	Table	CTS and volumes	0	1	0	0	1	0	0	0
Validation	Outbound freight - Primary delivery - 3	Table	Volumes	0	1	0	0	0	0	0	0
Validation	Outbound freight - F&V delivery	Table	CTS and volumes	0	1	1	0	1	0	0	0

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Validation	Outbound freight - Primary backhaul -1	Table	CTS and volumes	0	1	1	0	1	0	0	0
Validation	Outbound freight - Primary backhaul -2	Table	CTS and volumes	1	1	1	1	1	1	0	0
Validation	Outbound freight - F&V backhaul	Table	CTS and volumes	0	1	1	0	1	0	0	0
Validation	Top down alloc. - OTC	Table	CTS	0	0	1	0	0	0	0	0
Validation	Top down alloc. - Trade terms	Table	CTS and volumes	1	1	1	1	1	1	0	0
Validation	Top down alloc. - WH overhead - 1	Table	CTS and volumes	0	0	0	0	1	0	0	0
Validation	Top down alloc. - WH overhead - 2	Table	Volumes	0	0	0	0	1	0	0	0
Validation	CTS totals	Table	CTS	1	1	1	1	1	1	1	0
Validation	P&L input - 1	Table	CTS and volumes	1	0	1	1	0	0	0	0
Validation	P&L input - 2	Table	CTS and volumes	1	0	1	1	0	0	0	0

Tab	Feature	Type	Measures	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Reporting	P&L report	Table	CTS	0	0	1	1	0	0	0	0
Maps	Map	Map	CTS and volumes	0	0	1	0	1	0	0	0
Maps	Hectoliters	Table	Volumes	0	0	1	0	1	0	0	0
Maps	Drops	Table	Volumes	0	1	0	0	0	0	0	0
Details	Shipment details	Table	Volumes	1	1	1	1	1	1	0	0
Details	Cost details	Table	CTS	0	1	0	0	0	0	0	0
Details	Products	Table	None	0	0	0	1	0	0	0	0
Details	Customers	Table	None	0	0	1	0	0	0	0	0

Tab	Feature	Type	Measures	Scenario	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
All	Filters	Filter	None	1	1	1	1	1	1	1	0	0
All	Key numbers summary - 1	KPIs	Volumes	0	0	0	0	0	0	0	0	1
All	Key numbers summary - 2	KPIs	CTS	0	0	0	0	0	0	0	1	0
Scenario	Key numbers comparison	Bar chart	CTS and volumes	1	0	0	0	0	0	0	0	1
Scenario	Shipment building analysis - Delivery profile	Table	CTS and volumes	1	0	0	0	0	0	1	0	0
Scenario	Shipment building analysis - Tables - Shipments	Table	Volumes	1	0	0	0	0	0	0	0	0
Scenario	Shipment building analysis - Tables - Shipments 2	Table	Volumes	1	1	1	1	1	1	1	0	0

Tab	Feature	Type	Measures	Scenario	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Scenario	Shipment building analysis - Tables - Deliveries	Table	Volumes	0	0	1	0	0	0	0	0	0
Scenario	Shipment building analysis - Tables - Drops	Table	Volumes	0	0	1	0	0	0	0	0	0
Scenario	Shipment building analysis - Per shipment/day - 1	Bar chart	Volumes	1	0	1	1	0	0	0	0	0
Scenario	Shipment building analysis - Per shipment/day - 2	Table	Volumes	1	1	0	0	0	0	0	0	0
Scenario	Shipment building analysis - Nearest neighbour - Map	Map	Volumes	0	0	0	1	0	0	0	0	0

Tab	Feature	Type	Measures	Scenario	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Scenario	Shipment building analysis - Nearest neighbour - Shipments	Table	Volumes	1	1	0	0	0	0	0	0	0
Scenario	Shipment building analysis - Nearest neighbour - Time windows	Table	None	0	0	0	1	0	0	0	0	0
Scenario	Shipment building analysis - Nearest neighbour - Nearest neighbour	Table	Volumes	0	0	1	0	0	0	0	0	0
Scenario	Shipment building analysis - Warehouse handling	Table	CTS	1	1	1	1	1	1	1	0	0
Scenario	Shipment building analysis - Volume	Table	Volumes	1	1	1	1	1	1	1	0	0

Tab	Feature	Type	Measures	Scenario	Time	Shipments	Customers	Products	Sources	Vehicles	CTS buckets	KPIs
Scenario	Shipment building analysis - Details - 1	Table	Volumes	1	1	1	1	1	1	0	0	0
Scenario	Scatterplot - 1	Scatterplot	Volumes	1	1	1	1	1	1	1	0	0
Scenario	Scatterplot - 2	Difference chart	CTS	1	0	0	0	0	0	0	1	0
Scenario	Changing delivery location - 1	Map	CTS	1	0	0	1	0	1	0	0	0
Scenario	Changing delivery location - Changed Ship_from	Table	Volumes	1	0	0	0	0	1	0	0	0
Scenario	Changing delivery location - Changed Ship_to	Table	Volumes	1	0	0	1	0	1	0	0	0
Scenario	MOQ - 1	Bar chart	Volumes	1	0	0	0	0	0	0	0	0
Scenario	MOQ - 2	Table	Volumes	1	0	1	0	0	0	0	0	0
Scenario	MOQ - 3	Table	Volumes	1	1	1	1	1	1	1	0	0
Scenario	Changing delivery day	Bar chart	Volumes	1	1	0	0	0	0	0	0	0
Scenario	Vehicle threshold analysis	Table	Volumes	0	0	0	1	0	0	0	0	0

H Survey future analytics

Rank	Score
1	4
2	3
3	2
4	1

ID	Position	Experience	Improved job performance: Root-cause analysis	Improved job performance: Estimating the CTS of a new customer or product	Improved job performance: Guide in defining a minimum order threshold	Improved job performance: Guide in defining a discount policy when implementing Logistic Trade Terms	Time reduction: Root-cause analysis	Time reduction: Estimating the CTS of a new customer or product	Time reduction: Guide in defining a minimum order threshold	Time reduction: Guide in defining a discount policy when implementing Logistic Trade Terms
1	OpCo	2 - 3 years	4	1	3	2	4	1	3	2
2	OpCo	3 - 4 years	3	4	2	1	4	3	1	2
3	OpCo	More than 4 years	4	1	3	2	4	1	2	3
4	OpCo	3 - 4 years	3	4	1	2	3	4	1	2
5	OpCo	Less than 1 year	4	3	1	2	4	3	2	1
6	Global	1 - 2 years	4	2	1	3	3	1	2	4
7	Global	2 - 3 years	3	4	2	1	3	1	4	2
8	OpCo	1 - 2 years	4	3	2	1	4	3	2	1
9	OpCo	1 - 2 years	3	4	2	1	4	3	2	1
10	OpCo	1 - 2 years	4	1	3	2	3	1	4	2
11	Global	1 - 2 years	4	1	2	3	4	1	2	3
12	OpCo	Less than 1 year	1	3	2	4	1	2	3	4
13	OpCo	1 - 2 years	3	4	2	1	3	4	2	1
14	OpCo, Global	3 - 4 years	2	1	4	3	2	1	3	4
15	OpCo	1 - 2 years	3	4	2	1	4	1	2	3
16	OpCo	Less than 1 year	4	3	2	1	3	4	2	1

I Calculated tables, columns, and measures

Measures are indicated by [Measure] and columns are indicated by Table[Column].

Parameters

Name	Value range	Step size
Average # SKUs per drop - Weight	0 - 1	0.1
CV of cases per week - weight	0 - 1	0.1
Loose picking percentage - weight	0 - 1	0.1
Truck utilization goal	0 - 1	0.05
Weeks with demand - weight	0 - 1	0.1
Weeks with no demand in the demand cycle - weight	0 - 1	0.1

CTS / volume-unit bins - customers

```
distinct values Customers[CTS/volume unit (bins)]
```

Calculated columns

```
Midpoint (CTS / volume-unit) =
if bin = maximum bin
    (bin value + maximum of Customers[CTS/volume unit]) / 2
else
    (lower value + upper value) / 2
end if
```

CTS / volume-unit bins - products

```
distinct values Products[CTS/unit (bins)]
```

Calculated columns

```
Midpoint (CTS / volume-unit) =
if bin = maximum bin
    (bin value + maximum of Products[CTS/volume unit]) / 2
else
    (lower value + upper value) / 2
end if
```

Cost data

Measures

```
Average Cost to Serve KPI goal =
1.00001 [Cost to Serve] / [Volume], ignoring filters

Cost to Serve =
SUM('Cost_data'[Cost])

CTS/volume-unit =
[Cost to Serve] / [Volume]

CTS/volume-unit 1 month moving average =
[CTS/volume-unit], selecting the last month
```

```
Percentage of total Cost to Serve =
[Cost to Serve] /
    [Cost to Serve], ignoring filters

Transport costs =
[Cost to Serve], where CTS Cost buckets[CTS_GROUP] is Customer delivery or
   Intercompany transport
```

Customers

Calculated columns

```
Closest DC =
var Lat1 = Regions[Latitude]
var Lon1 = Regions[Longitude]
return the Ship_from[SHIP_FROM_ID] with the lowest value for:
    var Lat2 = Ship_from[Latitude]
    var Lon2 = Ship_from[Longitude]
    VAR A = 0.5 - COS((Lat2 - Lat1) * PI / 180) / 2 + COS(Lat1 * PI / 180)
        * COS(lat2 * PI / 180) * (1 - COS((Lng2 - Lng1) * PI / 180)) / 2
    12742 * ASIN((SQRT( A )))

CTS/volume-unit =
[CTS/volume-unit]

CTS/volume-unit (bins) =
var binsize = 0.95th percentile of Customers[CTS/volume-unit] / 20
if Customers[CTS/volume-unit] < ROUNDUP(0.95th percentile of Customers[CTS/
   volume-unit] / binsize) * binsize
    var lower = ROUNDDOWN(Customers[CTS/volume-unit] / binsize) * binsize
    lower " - " lower + binsize
else
    ROUNDUP(0.95th percentile of Customers[CTS/volume-unit] / binsize) *
        binsize " +"

Customer =
"Customer"

Volume =
[Volume]

Volume (bins) =
var binsize = 0.95th percentile of Customers[Volume] / 20
if Customers[Volume] < ROUNDUP(0.95th percentile of Customers[Volume] /
   binsize) * binsize
    var lower = ROUNDDOWN(Customers[Volume] / binsize) * binsize
    lower " - " lower + binsize
else
    ROUNDUP(0.95th percentile of Customers[Volume] / binsize) * binsize " +
       "

Savings moving to closest DC =
MAX(SUM(Drops[Estimated savings of moving drop]), 0), for drops at the
   customer

Segment =
```

```
var Volume =
if Customers[Volume] < 0.25th percentile of Customers[Volume]
     "Low"
else if Customers[Volume] < 0.75th percentile of Customers[Volume]
     "Medium"
else
     "High"
var CTS_volume-unit =
if Customers[CTS/volume-unit] < 0.25th percentile of Customers[CTS/volume-
    unit]
     "Low"
else if Customers[CTS/volume-unit] < 0.75th percentile of Customers[CTS/
    volume-unit]
     "Medium"
else
     "High"
Volume "␣Volume,␣" CTS_volume-unit "␣CTS/volume-unit"

Weeks with demand =
[Weeks with demand], for order lines going to the customer
```

Dates

```
all dates between the minimum and maximum of 'Order␣lines'[DEL_DATE]
```

Calculated columns

```
Day =
Dates[Date].Day

Month =
Dates[Date].Month

Month-Year =
Dates[Month] "␣-␣" Dates[Year]

MonthNumber =
Dates[Date].MonthNo

Quarter =
Dates[Date].Quarter

Quarter-Year =
Dates[Quarter] "␣-␣" Dates[Year]

QuarterNumber =
Dates[Date].QuarterNo

Week =
Dates[Date].Week

Week-Year =
Dates[Week] "␣-␣" Dates[Year]

Weekday =
switch WeekdayNumber
      1 = "Monday",
```

```
    2 = "Tuesday",
    3 = "Wednesday",
    4 = "Thursday",
    5 = "Friday",
    6 = "Saturday",
    7 = "Sunday"

WeekdayNumber =
weekday (starting on Monday)

Year =
Dates[Date].Year
```

Drops

```
distinct values 'Order_lines'[Drop_ID]
```

Calculated columns

```
CTS/volume-unit/km =
Drops[Transport costs] / Drops[Volume] / Drops[Haversine distance (km)]

Customers[Closest DC], for a customer in drop

Estimated costs from closest DC =
if Drops[SHIP_FROM_ID] <> Drops[Closest DC]
    var tariff = AVERAGE(Drops[CTS/volume-unit/km], where the ship_from
        location is the closest dc for the customer in this drop and the
        absolute difference in volume is less than 10%
    tariff * Drops[Haversine distance (km)] * Drops[Volume]

Estimated savings of moving drop =
Drops[Transport costs] - Drops[Estimated costs from closest DC]

Haversine distance (km) =
var Lat1 = Regions[Latitude]
var Lon1 = Regions[Longitude]
var Lat2 = Ship_from[Latitude]
var Lon2 = Ship_from[Longitude]
VAR A = 0.5 - COS((Lat2 - Lat1) * PI / 180) / 2 + COS(Lat1 * PI / 180) *
   COS(lat2 * PI / 180) * (1 - COS((Lng2 - Lng1) * PI / 180)) / 2
12742 * ASIN((SQRT( A )))

Volume =
[Volume]

SHIP_FROM_ID =
'Order_lines'[SHIP_FROM_ID], for the drop

SHIP_TO_ID =
'Order_lines'[SHIP_TO_ID], for the drop

Transport costs =
[Transport costs]
```

Volume bins - customers

```
distinct values Customers[Volume (bins)]
```

Calculated columns

```
Average CTS/volume-unit =
[CTS/volume-unit]

Delta Average CTS/volume-unit =
var previous = 'Volume bins - customers'[Average CTS/volume-unit], for the
    previous row sorted descending on 'Volume bins - customers'[Midpoint (
    Volume)]
'Volume bins - customers'[Average CTS/volume-unit] - previous

Midpoint (Volume) =
if bin = maximum bin
    (bin value + maximum of Customers[Volume]) / 2
else
    (lower value + upper value) / 2
end if
```

Volume bins - products

```
distinct values Products[Volume (bins)]
```

Calculated columns

```
Average CTS/volume-unit =
[CTS/volume-unit]

Delta Average CTS/volume-unit =
var previous = 'Volume bins - products'[Average CTS/volume-unit], for the
    previous row sorted descending on 'Volume bins - products'[Midpoint (
    Volume)]
'Volume bins - products'[Average CTS/volume-unit] - previous

Midpoint (Volume) =
if bin = maximum bin
    (bin value + maximum of Products[Volume]) / 2
else
    (lower value + upper value) / 2
end if
```

Order lines

Calculated columns

```
Delivery type =
if 'Order lines'[Route to customer] = "PW-CUST" OR "PW-SELF"
    "Direct delivery (from a PW to the Customer)"
else
    "Indirect delivery (to the Customer via a DC)"

Drop_ID =
'Order lines'[SHIPMENT_ID], 'Order lines'[SHIP_TO_ID]

Extra cases =
'Order lines'[DEL_QTY] mod Products[CASES_PER_PALLET]
```

```
Volume =
'Order_lines'[DEL_QTY] * Products[VOLUME_PER_CASE]

Kilograms =
'Order_lines'[DEL_QTY] * Products[KILOGRAMS_PER_CASE]

Other =
'Order_lines'[DEL_QTY] * Products[OTHER_PER_CASE]

Pallets =
'Order_lines'[DEL_QTY] / Products[CASES_PER_PALLET]

Route to Customer =
var first_part =
if Costs between intermediate and final warehouse > 0
    "PW-DC-DC-"
else if Costs between production and intermediate warehouse > 0
    "PW-DC-"
else
    "PW-"
var second_part =
if Shipment type indicates self collection
    "SELF"
else
    "CUST"
first part second part

Whole pallets =
ROUNDDOWN('Order_lines'[Pallets])

YearWeek =
YEAR('Order_lines'[DEL_DATE]) WEEKNUM('Order_lines'[DEL_DATE])
```

Products

Calculated columns

```
CTS/volume-unit =
[CTS/volume-unit]

CTS/volume-unit (bins) =
var binsize = 0.95th percentile of Products[CTS/volume-unit] / 20
if Products[CTS/volume-unit] < ROUNDUP(0.95th percentile of Products[CTS/
   volume-unit] / binsize) * binsize
    var lower = ROUNDDOWN(Products[CTS/volume-unit] / binsize) * binsize
    lower " - " lower + binsize
else
    ROUNDUP(0.95th percentile of Products[CTS/volume-unit] / binsize) *
        binsize " +"

Volume =
[Volume]

Volume (bins) =
var binsize = 0.95th percentile of Products[Volume] / 20
if Products[Volume] < ROUNDUP(0.95th percentile of Products[Volume] /
   binsize) * binsize
```

```
    var lower = ROUNDDOWN(Products[Volume] / binsize) * binsize
    lower " - " lower + binsize
else
    ROUNDUP(0.95th percentile of Products[Volume] / binsize) * binsize " +"
```

Ship_from

Calculated columns

```
Latitude =
Regions[LATITUDE], where Regions[SHIP_TO_REGION] ID = Ship_from[SHIP_FROM_
    ID]

Longitude =
Regions[LONGITUDE], where Regions[SHIP_TO_REGION] ID = Ship_from[SHIP_FROM_
    ID]
```

Shipments

```
distinct values 'Order_lines'[SHIPMENT_ID]
```

Calculated columns

```
Maximum load =
if Shipment has a vehicle
    Vehicles[PAYLOAD],
else
    Shipments[Pallets]

Pallets =
ROUNDUP(SUM('Order_lines'[Pallets])), for the shipment

SKUs in shipment =
DISTINCTCOUNT('Order_lines'[MATERIAL]), for the shipment

Tons =
ROUNDUP(SUM('Order_lines'[Kilograms] / 1000)), for each order line in the
    shipment
```

Measures

```
Truck utilization =
if shipment being considered > 1
    var max_load = SUM(MAX(Shipments[Maximum load], Shipments[Pallets])),
        for each shipment being considered
    SUM('Order_lines'[Pallets]) / max_load
else
    SUM('Order_lines'[Pallets]) / SUM(Shipments[Maximum load])
```

Volumes

```
{
Volume-unit    Volume
Volume-unit    Kilograms
Volume-unit    Pallets
Volume-unit    Cases
Volume-unit    Other}
```

Measures

```
Average # SKUs per drop =
AVERAGE(DISTINCTCOUNT('Order_lines'[MATERIAL])), for each drop

Complexity =
var weeks_demand_normalized = ([Weeks with demand] - MIN([Weeks with demand
    ])) / (MAX([Weeks with demand]) - MIN([Weeks with demand])), ignoring
    filters
var weeks_no_demand_normalized = ([Weeks with no demand in the demand cycle
    ] - MIN([Weeks with no demand in the demand cycle])) / (MAX([Weeks with
    no demand in the demand cycle]) - MIN([Weeks with no demand in the
    demand cycle])), ignoring filters
var cv_cases_week_normalized = ([CV of cases per week] - MIN([CV of cases
    per week])) / (MAX([CV of cases per week]) - MIN([CV of cases per week])
    ), ignoring filters
var SKUs_drop_normalized = ([Average # SKUs per drop] - MIN([Average # SKUs
     per drop])) / (MAX([Average # SKUs per drop]) - MIN([Average # SKUs per
     drop])), ignoring filters
var loose_picking_normalized = ([Loose picking percentage] - MIN([Loose
    picking percentage])) / (MAX([Loose picking percentage]) - MIN([Loose
    picking percentage])), ignoring filters
([Weeks with demand - Weight Value] * (1 - weeks_demand_normalized) +
     [Weeks with no demand in the demand cycle - Weight Value] * weeks_no_
        demand_normalized +
     [CV of cases per week - Weight Value] * cv_cases_week_normalized +
     [Average # SKUs per drop - Weight Value] * SKUs_drop_normalized +
     [Loose picking percentage - Weight Value] * loose_picking_normalized) /
     ([Weeks with demand - Weight Value] +
     [Weeks with no demand in the demand cycle - Weight Value] +
     [CV of cases per week - Weight Value] +
     [Average # SKUs per drop - Weight Value] +
     [Loose picking percentage - Weight Value])

Customers sold to =
DISTINCTCOUNT('Order_lines'[SOLD_TO_ID])

CV of cases per week =
STDEV.P('Order_lines'[DEL_QTY]) / AVERAGE('Order_lines'[DEL_QTY]), for each
     week in Dates[Week]

Deliveries =
DISTINCTCOUNT('Order_lines'[DELIVERY_ID])

Drops =
DISTINCTCOUNT('Order_lines'[Drop_ID])

Drops per week =
[Drops] / [Weeks between first and last demand]

Volume =
SUM('Order_lines'[Volume])

Locations shipped to =
DISTINCTCOUNT('Order_lines'[SHIP_TO_ID])

Loose picking percentage =
1 - SUM('Order_lines'[Whole pallets]) / SUM('Order_lines'[Pallets])
```

```
Order lines =
DISTINCTCOUNT('Order_lines'[Orderline_ID])

Pallets =
SUM('Order_lines'[Pallets])

Percentage of total volume =
[Volume] /
    [Volume], ignoring filters

Products sold =
DISTINCTCOUNT('Order_lines'[MATERIAL])

Shipments =
DISTINCTCOUNT('Order_lines'[SHIPMENT_ID])

Volume =
switch Selected volume-unit
    "Volume" = SUM('Order_lines'[Volume]),
    "Kilograms" = SUM('Order_lines'[Kilograms]),
    "Pallets" = SUM('Order_lines'[Pallets]),
    "Cases" = SUM('Order_lines'[Cases]),
    "Other" = SUM('Order_lines'[Other])

Volume 1 month moving average =
[Volume], selecting the last month

Volume-unit/drop =
[Volume] / [Drops]

Volume-unit/week =
[Volume] / [Weeks between first and last demand]

Weeks between first and last demand =
DATEDIFF(MIN('Order_lines'[DEL_DATE]), MAX('Order_lines'[DEL_DATE])) + 1

Weeks with demand =
MIN(DISTINCTCOUNT('Order_lines'[YearWeek]), [Weeks between first and last
   demand])

Weeks with no demand in the demand cycle =
[Weeks between first and last demand] - [Weeks with demand]
```

J Power BI tool

Existing features

The figure below shows the scatter/drop analysis graph that was recreated based on the QlikView tool of the research company.

Figure 10: The Scatter / drop analysis graph created in the new Power BI tool showing demo data

An addition to the scatter plot is the inclusion of the dotted lines that show average values. Furthermore, we incorporated no significant improvements compared to the old tool. Of course, hierarchies allow users to drill down to view data on a level. The figure below shows an example where Customer 3 is selected, which was poorly visible previously.

Figure 11: The Scatter / drop analysis graph created in the new Power BI tool showing demo data viewing one particular customer

The figure below shows the loose picking percentage view, which users can access through the button on the top-right.

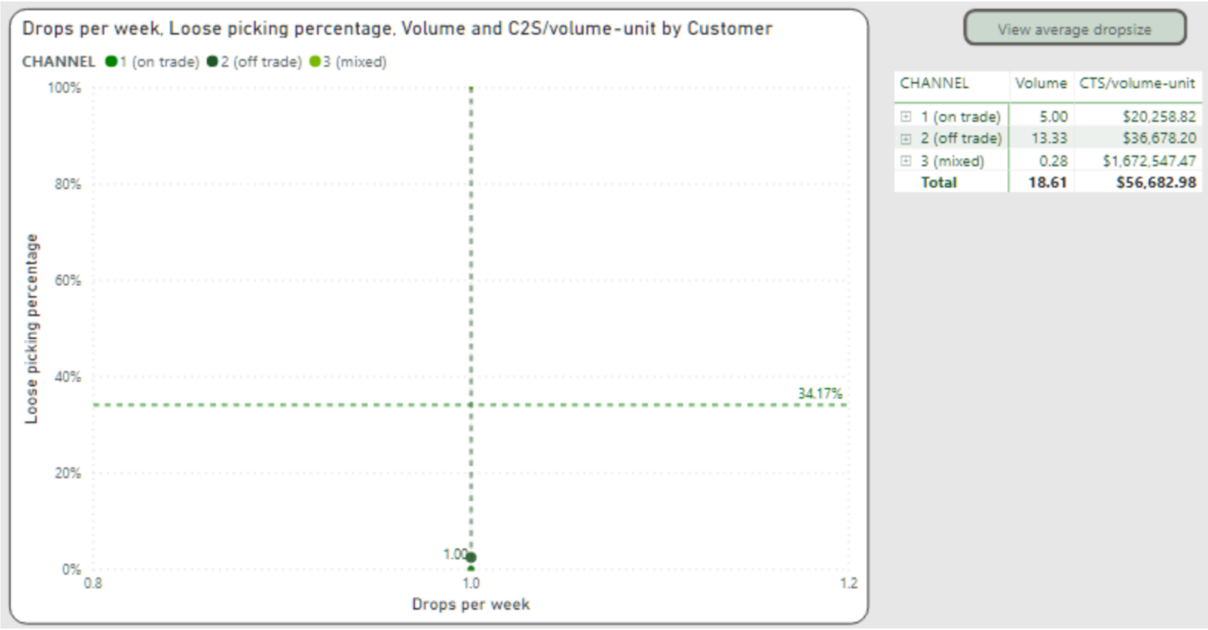


Figure 12: The Scatter / drop analysis graph created in the new Power BI tool showing the loose picking percentage based on demo data

The loose picking percentage is the percentage of pallets that the OpCo did not ship as a full pallet. For Customer 3, it is 100%, as they did not sell any full pallets to this customer, and Customer 1 has a loose picking percentage of 0%, as the OpCo only sold full pallets to this customer. This graph also does not incorporate any significant changes compared to the previous tool besides the dotted lines indicating averages. The last graph feature migrated from the QlikView tool to Power BI is the Delivery profile graph shown in the figure below.

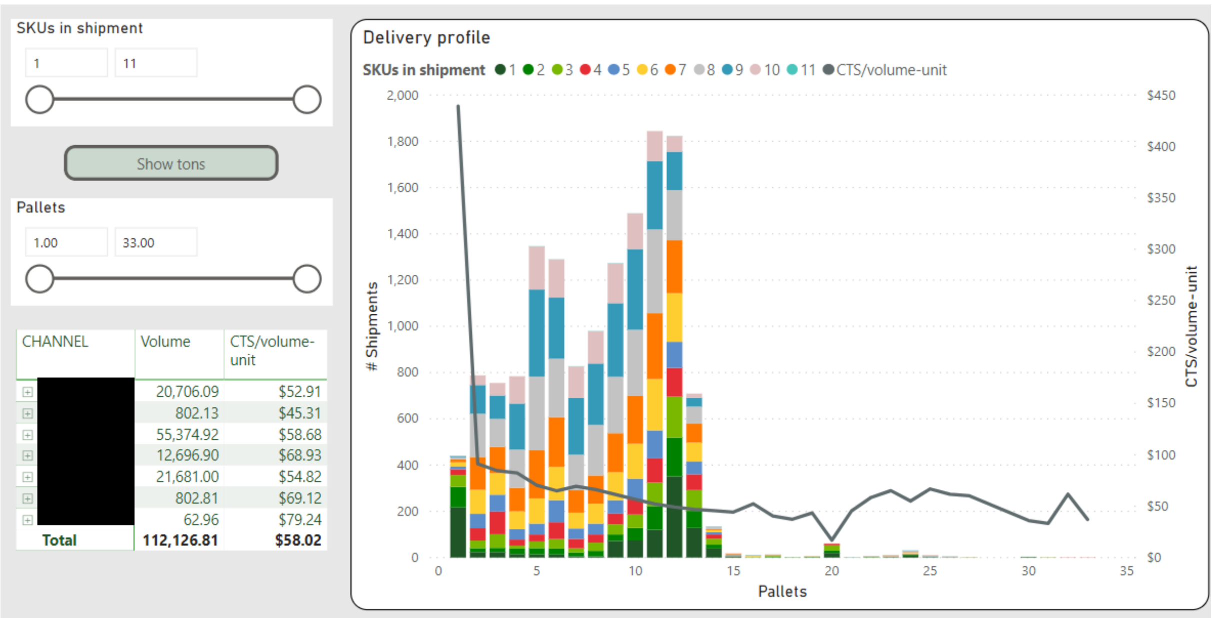


Figure 13: The Delivery profile graph created in the new Power BI tool

Compared to the old tool, the feature shown above combines the stacked columns with the line indicating the CTS per volume-unit as was done with the Cost per volume-unit graph. Besides this, there are no significant improvements compared to the feature in the QlikView tool. Users also have the option to switch to tons rather than pallets by clicking the button on the left.

The figure below shows the landing page of the Key numbers feature, which is the landing page of an environment with various tables.

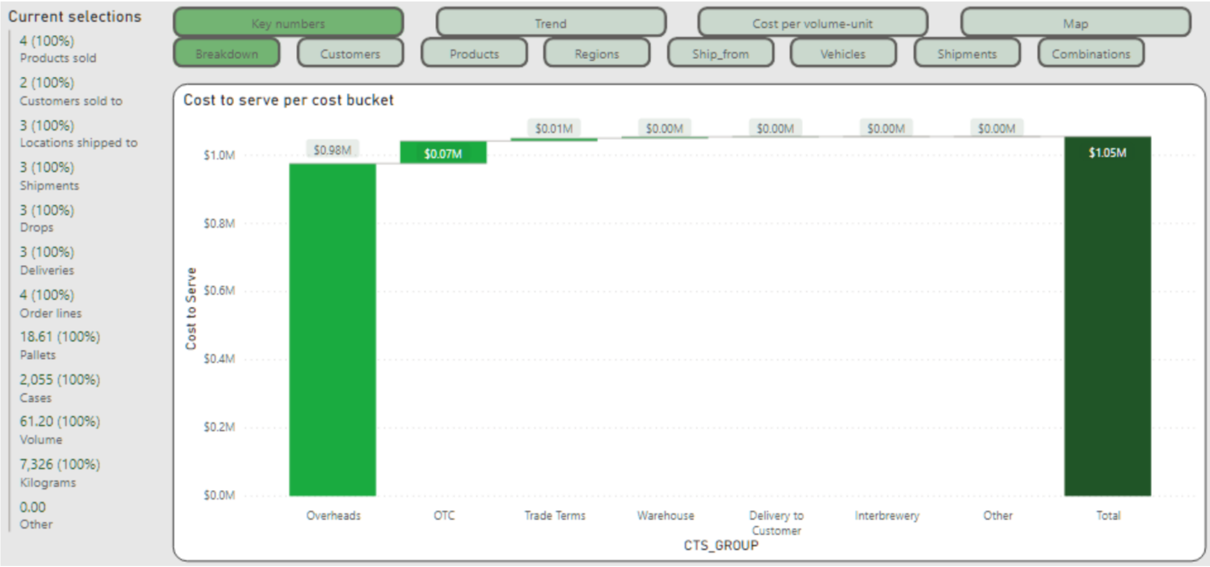


Figure 14: The Key numbers feature created in the new Power BI tool showing demo data

In the previous tool, the Key numbers were all shown in a table form. In an attempt to make the Key numbers more appealing, we made two significant changes. First, we always visualize the Key numbers on the left side of a page in the new tool, including the percentage of the total currently selected. Second, we created a waterfall chart to show the division among cost buckets. The waterfall chart incorporates the cost bucket hierarchy. The figure below shows a drill-down into the Overheads cost group.

Figure 15: The Key numbers feature created in the new Power BI tool showing demo data drilled down on the Overheads CTS_GROUP

Here, the division of costs among the buckets below the groups is shown. Users can drill down further to a sub-bucket level. As is visible on the top-right of the figure above, we created a feature to combine dimensions. The figure below shows the combinations that users can make in the new tool.

Combine	Products	Regions	Ship_from	Vehicles	Shipments
Customers	Show CTS and Volume	Show CTS and Volume	Show CTS and Volume	Show CTS and Volume	Show CTS and Volume
Products		Show CTS and Volume	Show CTS and Volume	Show CTS and Volume	Show CTS and Volume
Regions			Show CTS and Volume	Show CTS and Volume	Show CTS and Volume
Ship_from				Show CTS and Volume	Show CTS and Volume
Vehicles					Show CTS and Volume

Figure 16: The Combinations feature created in the new Power BI tool

The tool in QlikView has the option to add a pivot table that allows users to change dimensions dynamically, which was not possible in Power BI. Therefore, we created the structure above. The figure below shows the combination of customers and products.

Volume

LEVEL1_DESC	Product				Total
SOLD_TO_NAME	Box	Half-pallet	Layer	Total	
Customer 1		5.00		5.00	5.00
Customer 2	13.33			13.33	13.33
Customer 3			0.28	0.28	0.28
Total	13.33	5.00	0.28	18.61	18.61

CTS/volume-unit

LEVEL1_DESC	Product			
SOLD_TO_NAME	Box	Half-pallet	Layer	Total
Customer 1		$20,258.82		$20,258.82
Customer 2	$36,678.20			$36,678.20
Customer 3			$1,672,547.47	$1,672,547.47
Total	$36,678.20	$20,258.82	$1,672,547.47	$56,682.98

Figure 17: The Customers and products page of the combinations feature created in the new Power BI tool showing demo data

The table showing volume incorporates a feature that shows bars indicating high volume. The table showing the CTS per volume-unit indicates low, medium, and high values based on the 33rd and 66th percentile. Compared to the old tool, there are no significant changes besides that data is represented in a slightly different way, and it is more clear that users can view data along multiple dimensions. The last set of tables included in the tool are tables with the dimensions' details. The figure below shows the Ship_from and vehicle details page.

SHIP_FROM_ID	SHIP_FROM_DESC	DISTRIBUTOR	SHIP_FROM_TYPE	DUAL_LOCATED_PLANT_ID	SHIPMENT_TYPE	INT_OUTBOUND_COST_ACTIVE
DC1	Pallet town distribution center	A	Distribution center		0	1
DC2	Central village distribution center	A	Distribution center		0	1
PW1	Green village plant	A	Plant warehouse		0	1
PW2	Wholesalerley village plant	A	Plant warehouse		0	1

VEHICLE_TYPE	PAYLOAD	AVG_UTILIZATION	AVG_VOLUME_LOWER	AVG_VOLUME_UPPER	VEHICLE_ID	VEHICLE_COMPANY_ID	VEHICLE_COMPANY_NAME
26PLT	26.00	100	26	26	200038	B	Bvehicles
26PLT	26.00	100	26	26	200039	B	Bvehicles

Figure 18: The Ship_from and vehicle details page created in the new Power BI tool showing demo data

The pages with details show the master data as entered in the data template. It can be useful for users to be able to export this data without having to go back to the input data. Furthermore, when the research company realizes an automated data connection, they must have this feature as the input data might be unreachable for users. In the old tool, users could only view a customer and product detail page. Then, we included a map feature in the new tool. The figure below shows the map feature.

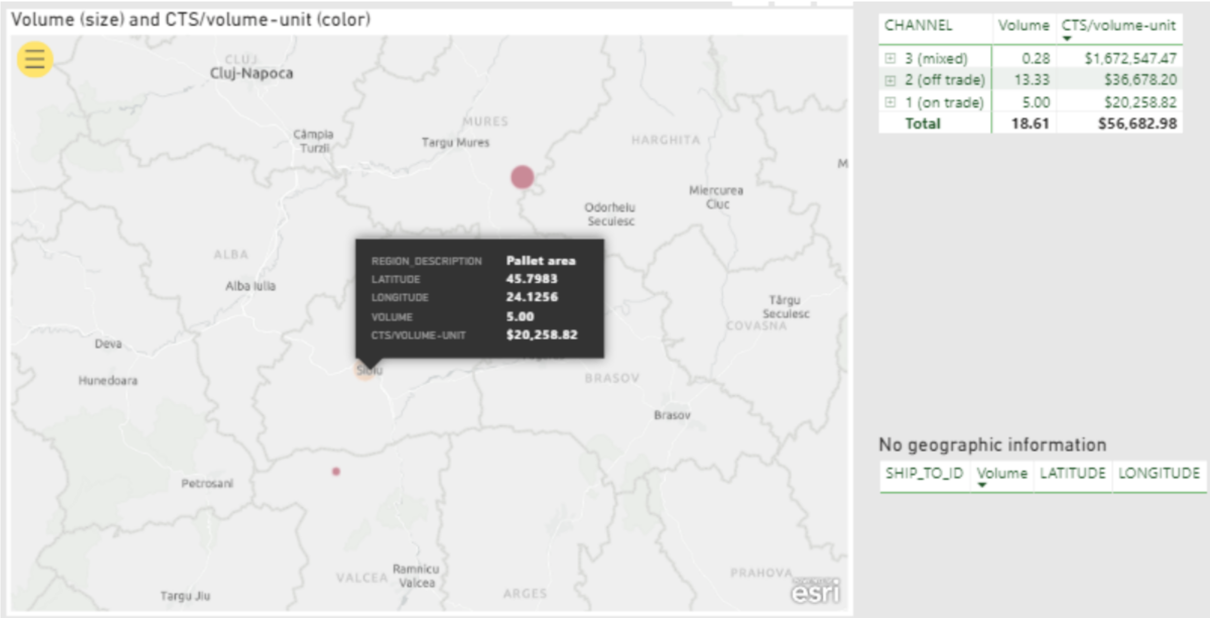


Figure 19: The Map feature created in the new Power BI tool showing demo data

The map indicates volumes and the CTS per volume-unit. Furthermore, a table shows which customers do not have geographic information in the input data. In QlikView, the map was no longer working.

New features

We presented the truck utilization as a potential cost driver for a high CTS per volume-unit. We created a measure that calculates the truck utilization along with a page to visualize the truck utilization related to the origins of shipments, which is presented in Appendix I. The figure below shows the page.

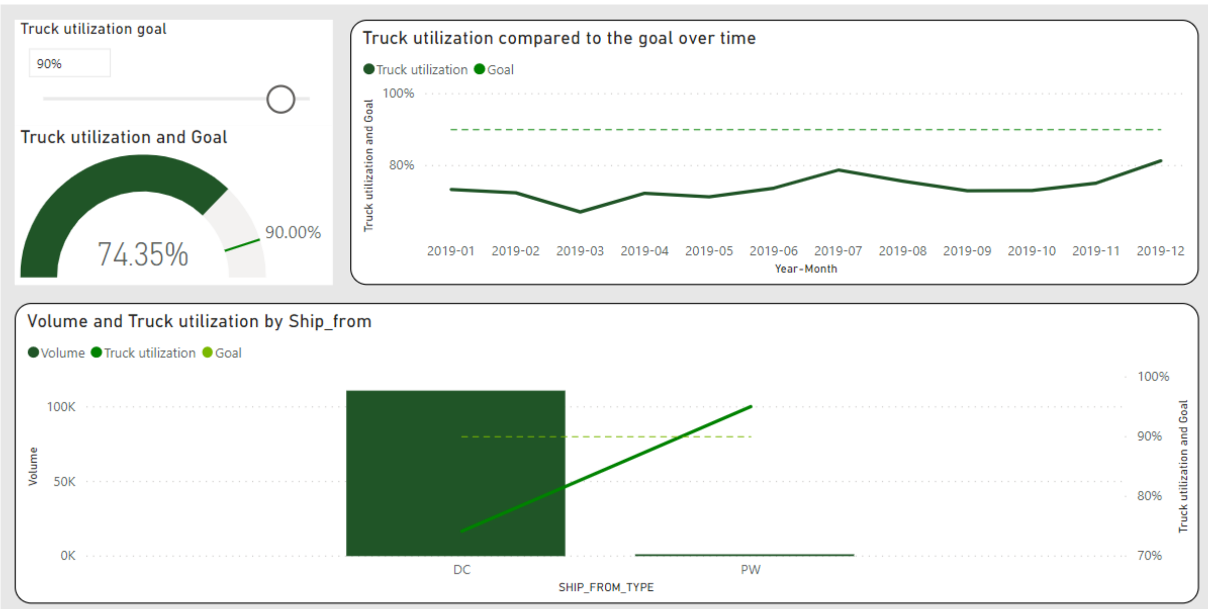


Figure 20: The Truck utilization page created in the new Power BI tool

On this page, users can set a goal on the top left, which shows in all graphs that can be useful when making print screens to share with colleagues. Below the graph feature, a gauge shows the truck utilization of the data selected. By clicking on other graphs, the page selection changes along with this graph. On the top left, truck utilization shows over time. Finally, at the bottom of the page, truck utilization is shown for ship_from locations. In this graph, users can to drill down in the hierarchy to the level of individual shipments. The next page also focuses on Ship_from locations by visualizing the origins of shipments to customers and supplying additional information. The figure below shows the Ship_from locations page.

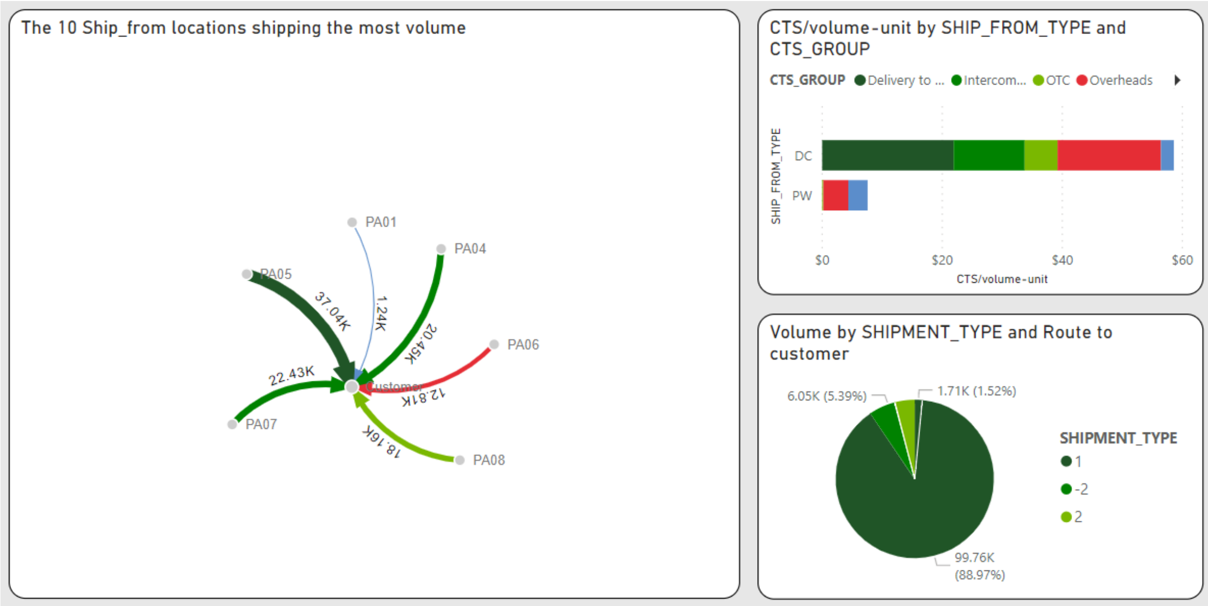


Figure 21: The Ship_from locations page created in the new Power BI tool

On the left ship_from locations with arrows pointing to a customer centroid indicate the flow of goods. A large arrow indicates that more volume is flowing through that route. Furthermore, hovering over arrows shows the CTS per volume-unit for that route to the customer. The two graphs on the right side provide supportive information. By clicking on these graphs, users could, for example, view shipments originating from a PW or of a specific type. Then, we made a page that simplistically incorporates prescriptive analytics by determining whether customers' deliveries originated from the DC closest to the customer. We used the Haversine formula (Veness, 2020) to calculate the distance between two locations. For deliveries not fulfilled from the DC that is the closest to the customer, we determined a tariff to calculate potential savings based on the average costs per volume-unit per kilometer of deliveries with a maximum absolute deviation of no more than ten percent originating from the DC that is the closest. We only considered savings when it is profitable for a customer to move all drops to the DC that is the closest to the customer. The figure below shows the page created.

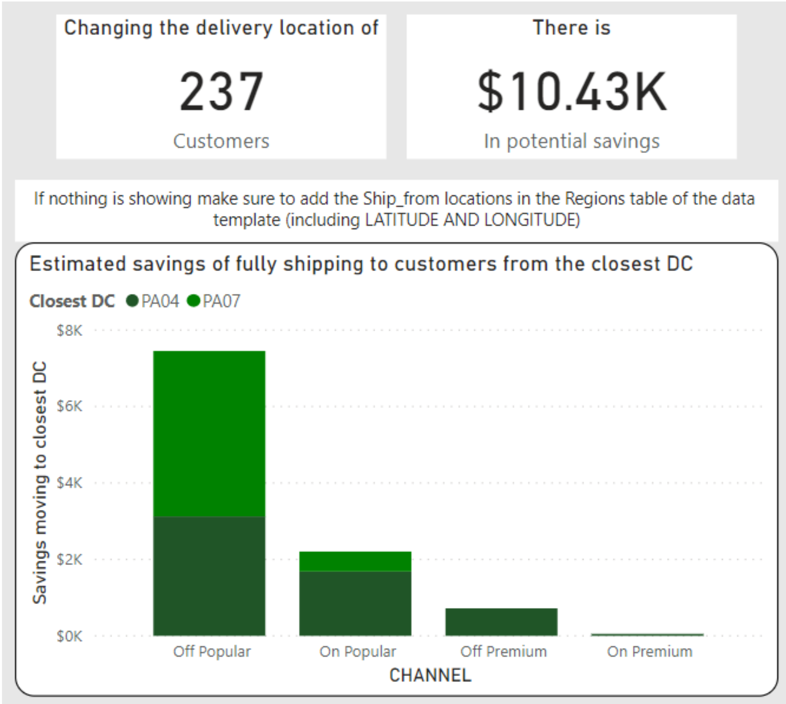


Figure 22: The page considering customers being delivered from the closest DC created in the new Power BI tool

The graph shows customers of whom different DCs can fulfill drops to obtain savings. Furthermore, the page includes a table that shows drops. When a user makes a selection using the graph, the table shows the effected drops. This new feature serves as an example of how the research company could automate opportunities. However, we could improve the feature above by incorporating actual distances and having a validated method for calculating new tariffs. It was not possible to validate the outcomes of this feature during the research. Then, a page was created that focuses on the relation between volume and the CTS per volume-unit because of the suspicion that customers with a higher volume often have a lower CTS per volume-unit due to larger ordering sizes. The figure below shows the page created to research the usefulness of visualizing this relationship using histograms.

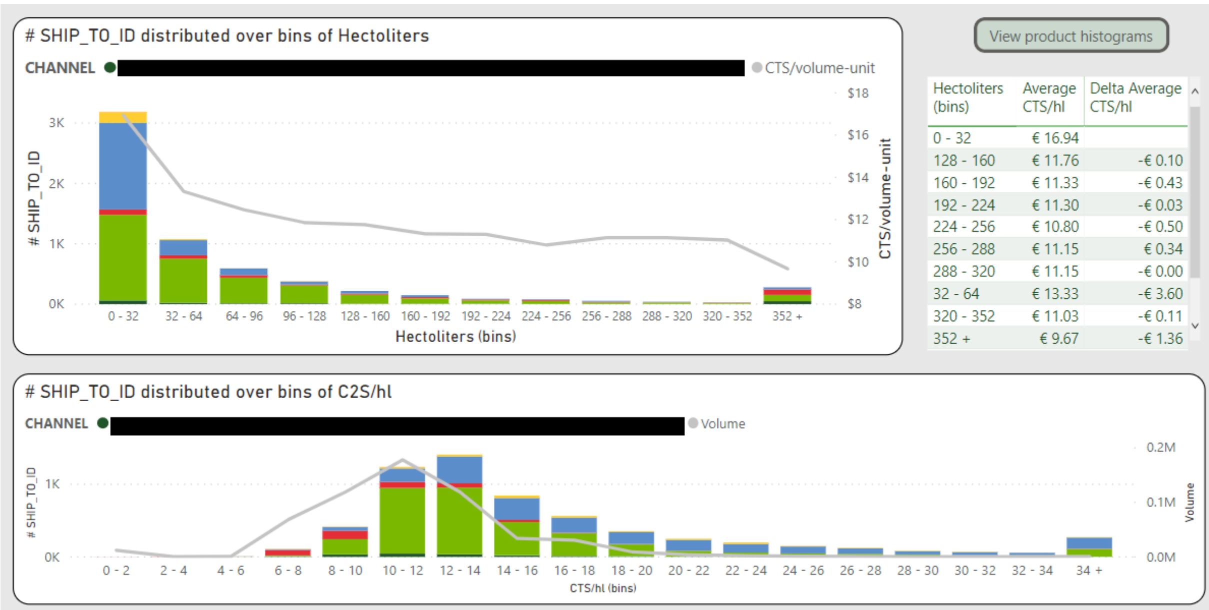


Figure 23: The Histograms page created in the new Power BI tool

The histograms have two applications. First, the top chart shows the change in the CTS per volume-unit between buckets. So, users can estimate savings related to an increase in volume. We also visualized these differences in the table on the top-right, which can be useful when making decisions for incentives to increase sales. Second, when a new customer is acquired, users can make a rough estimation regarding expected logistics costs. We created similar histograms on a product level. The next page focuses on customer complexity. In collaboration with the CTS believes factors that increase complexity have been determined and plotted for customers. These factors are as follows:

- Weeks with demand (inversed)
- Weeks with no demand in the demand cycle
- Coefficient of Variation of cases per week
- Average # SKUs per drop
- Loose picking percentage

We normalized the values of these variables between zero and one. Then, we multiply each variable by a corresponding weight, which results in a variable score. For each customer, we divide the sum of the variables' scores by the sum of the variables' weights. Finally, resulting in scores between zero and 1, or 0% and 100%. The figure below shows the created page.

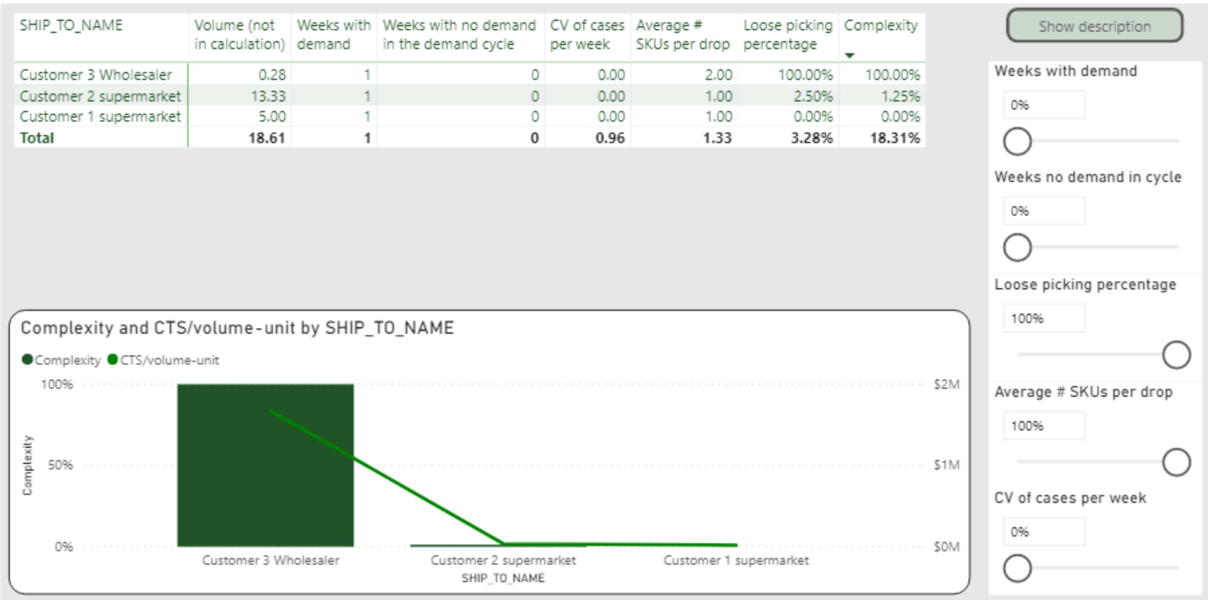
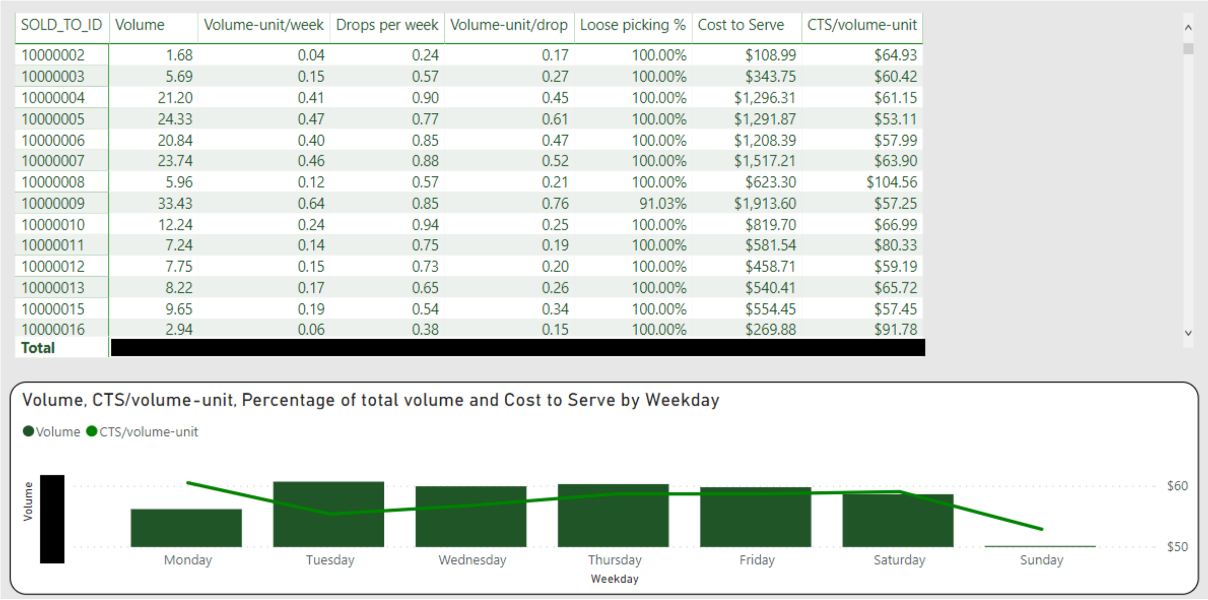


Figure 24: The Complexity page created in the new Power BI tool using demo data

The example using demo data shows the concept. The table shows that the customers have the same value for the first three variables. For simplicity, we set the weight for these values to 0%. Based on the set of customers, Customer 3 has the worst values for the two remaining variables and therefore has a complexity of 100%. Customer 1 has the best values for the two remaining variables resulting in a complexity of 0%. The purpose of the bottom graph is to visualize the relation between complexity and CTS per volume-unit. We created the next page to visualize indicators to find opportunities for Logistics Trade Terms, in collaboration with a Logistics Trade Terms expert. The figure below shows the page that we created.



SOLD_TO_ID	Volume	Volume-unit/week	Drops per week	Volume-unit/drop	Loose picking %	Cost to Serve	CTS/volume-unit
10000002	1.68	0.04	0.24	0.17	100.00%	$108.99	$64.93
10000003	5.69	0.15	0.57	0.27	100.00%	$343.75	$60.42
10000004	21.20	0.41	0.90	0.45	100.00%	$1,296.31	$61.15
10000005	24.33	0.47	0.77	0.61	100.00%	$1,291.87	$53.11
10000006	20.84	0.40	0.85	0.47	100.00%	$1,208.39	$57.99
10000007	23.74	0.46	0.88	0.52	100.00%	$1,517.21	$63.90
10000008	5.96	0.12	0.57	0.21	100.00%	$623.30	$104.56
10000009	33.43	0.64	0.85	0.76	91.03%	$1,913.60	$57.25
10000010	12.24	0.24	0.94	0.25	100.00%	$819.70	$66.99
10000011	7.24	0.14	0.75	0.19	100.00%	$581.54	$80.33
10000012	7.75	0.15	0.73	0.20	100.00%	$458.71	$59.19
10000013	8.22	0.17	0.65	0.26	100.00%	$540.41	$65.72
10000015	9.65	0.19	0.54	0.34	100.00%	$554.45	$57.45
10000016	2.94	0.06	0.38	0.15	100.00%	$269.88	$91.78
Total	[redacted]						

Figure 25: The Logistics Trade Terms page created in the new Power BI tool using demo data

The table shows KPIs that can aid in finding opportunities. For example, when the loose picking percentage is high, the research company can offer a discount to reduce this. The idea is that the savings are higher than the discount given in such a case. Furthermore, the graph on the bottom of the page shows the distribution of volume on different drop days. Using the functionalities of Power BI, users can view this for a single customer, or a group of customers, by making selections. For example, this could lead to a Logistic Trade Term that offers a discount when an OpCo does not have to deliver orders on a possibly more expensive day.

K R scripts root-cause analysis

Variable creation.R

```
#Set working directory——
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

#Libraries——
library(readxl)
library(tidyverse)
library(openxlsx)
library(reshape2)
library(lubridate)

#Global constant variables——
zero <- 0.00001
set.seed(1)

#Functions——
##Determine the average, min, max, SD and CV of a variable
stats <- function(attribute, group, type = c("count", "sum")) {

  df <- shipments %>%
    group_by(Key,
             !! sym(group)) %>%
    summarize(value = ifelse(type == "count",
                             length(unique(!! sym(attribute))),
                             sum(!! sym(attribute))
    )
    ) %>%
    group_by(Key) %>%
    summarize('Average' = mean(value),
              'Min' = min(value),
              'Max' = max(value),
              'SD' = sd(value),
              'CV' = sd(value) / mean(value)
    )
  colnames(df) <- c("Key",
                    paste(colnames(df[-1]),
                          " ",
                          type,
                          " of ",
                          attribute,
                          "/",
                          group,
                          sep = "")
  )

  output <<- left_join(output, df, by = "Key")

}

##Calculate the percentage of an attribute over a group
percentage_per_attribute <- function(attribute_vector, group, type_vector,
    divisor_vector) {

  for (i in 1:length(attribute_vector)) {
```

```r
      df <- shipments %>%
        group_by(Key,
                 !! sym(group)
        ) %>%
        summarize(Values = ifelse(type_vector[i] == "count",
                                  length(unique(!! sym(attribute_vector[i])))
                                    ,
                                  sum(!! sym(attribute_vector[i]))
        )
        ) %>%
        dcast(paste("Key",
                    "~",
                    group),
              value.var = "Values")
      colnames(df) <- c("Key",
                        paste("Percentage of",
                              divisor_vector[i],
                              group,
                              colnames(df[-1])
                        )
      )
      ###Calculate percentages
      output <<- left_join(output, df, by = "Key")
      output[(ncol(output) - ncol(df) + 2):ncol(output)] <<-
        sapply(output[(ncol(output) - ncol(df) + 2):ncol(output)], function(x
           ) x / output[, divisor_vector[i]])
    }
  }

##Calculate the loose picking percentage
percentage_loose_picking <- function(group_vector, type_vector, divisor_
    vector) {

    for (i in 1:length(group_vector)) {
      df <- shipments %>%
        group_by(Key,
                 Picking
        ) %>%
        summarize(Values = ifelse(type_vector[i] == "count",
                                  length(unique(!! sym(group_vector[i]))),
                                  sum(!! sym(group_vector[i]))
        )
        ) %>%
        dcast(Key ~ Picking,
              value.var = "Values"
        ) %>%
        select(Key, `Loose picking`)

      colnames(df) <- c("Key",
                        paste("Percentage of",
                              divisor_vector[i],
                              "involved loose case picking")
      )
      ###Calculate percentage
      output <<- left_join(output, df, by = "Key")
      output[ncol(output)] <<- output[ncol(output)] / output[divisor_vector[i
         ]]
```

```r
    }
}

#Data input——
##Shipments
shipments <- read_excel(paste(getwd(),
                              "/Shipment_details.xlsx",
                              sep = "")
)
###Check that input is distinct and create additional variables
shipments <- distinct(shipments) %>%
  mutate(Drop_ID = paste(SHIPMENT_ID,
                         SHIP_TO_ID),
         Week = as.numeric(strftime(DEL_DATE,
                                    format = "%V")
         ),
         plt_loose = plt_orderline - floor(plt_orderline),
         Picking = ifelse(plt_loose > 0,
                          "Loose_picking",
                          "Full_pallets")
  )

##Costs
costs     <- read_excel(paste(getwd(),
                              "/Cost_details.xlsx",
                              sep = "")
)
###Check for an empty CTS_SUB_BUCKET variable and remove if present
if("CTS_SUB_BUCKET" %in% colnames(costs)) {
  costs <- costs %>% select(-CTS_SUB_BUCKET)
}
###Make sure all values are numeric
costs[]   <- lapply(costs,
                    function(x) as.numeric(as.character(x))
)
###Check that input is distinct and calculate total costs
costs     <- distinct(costs) %>%
  mutate(CTS = apply(costs[,-1], 1, sum))
###Add the total costs to the shipment table
df <- costs %>%
  select(Orderline_ID,
         CTS)
shipments <- left_join(shipments,
                       df,
                       by = "Orderline_ID")

##Include the path of an order line
df <- shipments %>%
  select(Orderline_ID, SHIPMENT_TYPE)

###First part (number of DCs)
if ("Primary_shuttle_from_INT_to_Final" %in% colnames(df)) {
  df <- left_join(df, costs, by = "Orderline_ID") %>%
    mutate(Path1 = ifelse(`Primary shuttle from INT to Final` > zero,
                          "PW-DC-DC-",
                          ifelse(`Primary shuttle from PW to Final` > zero,
                                 "PW-DC-",
```

```
                                   "PW-")))
} else if ("Primary_shuttle_from_PW_to_Final" %in% colnames(df)) {
  df <- left_join(df, costs, by = "Orderline_ID") %>%
    mutate(Path1 = ifelse(`Primary shuttle from PW to Final` > zero,
                          "PW-DC-",
                          "PW-"))
} else {
  df <- left_join(df, costs, by = "Orderline_ID") %>%
    mutate(Path1 = "PW-")
}

###Second part (CUST or SELF)
df <- df %>%
  mutate(Path = ifelse(SHIPMENT_TYPE > zero,
                       paste(Path1, "CUST", sep = ""),
                       paste(Path1, "SELF", sep = "")))

df <- df %>% select(Orderline_ID, Path)
shipments <- left_join(shipments, df, by = "Orderline_ID")

#Choose the level of detail----
##Paste together a unique Key value
shipments$Key <- paste("P", shipments$MATERIAL,
                       sep = "")

#Create output structure----
##Add values that you want to have in the output
output <- shipments %>% select(Key, MATERIAL) %>%
  arrange(Key) %>%
  distinct()
##Add volume and CTS
df <- shipments %>%
  group_by(Key) %>%
  summarize('volume' = sum(volume_orderline),
            'CTS' = sum(CTS)
  )
output <- left_join(output, df, by = "Key")
##Initialize counters
output_intro_len <- length(output)

#Calculate attributes----
##Ordering behavior----
df <- shipments %>%
  group_by(Key) %>%
  summarize('Weeks_with_demand' = max(Week) - min(Week) + 1,
            'Weeks_ordered' = length(unique(Week)),
            'Week_not_ordered' = max(Week) - min(Week) + 1 - length(unique(
              Week)),
            'Drops' = length(unique(Drop_ID)),
            'Deliveries' = length(unique(DELIVERY_ID)),
            'Products' = length(unique(MATERIAL)),
            'Order_lines' = length(unique(Orderline_ID)),
            'Volume' = sum(volume_orderline))

###Join df
output <- left_join(output, df, by = "Key")
```

```r
##Ordering stats----
###Create a data frame with rows that hold an attribute, group, and type
df <- data.frame(attribute = c("Drop_ID",
                               "DELIVERY_ID", "DELIVERY_ID",
                               "MATERIAL", "MATERIAL", "MATERIAL",
                               "Orderline_ID", "Orderline_ID", "Orderline_
                                   ID", "Orderline_ID",
                               "volume_orderline", "volume_orderline", "
                                   volume_orderline", "volume_orderline", "
                                   volume_orderline"),
                 group = c("Week",
                           "Week", "Drop_ID",
                           "Week", "Drop_ID", "DELIVERY_ID",
                           "Week", "Drop_ID", "DELIVERY_ID", "MATERIAL",
                           "Week", "Drop_ID", "DELIVERY_ID", "MATERIAL", "
                               Orderline_ID"),
                 type = c("count",
                          "count", "count",
                          "count", "count", "count",
                          "count", "count", "count", "count",
                          "sum", "sum", "sum", "sum", "sum"),
                 stringsAsFactors = F
)
###Run the stats function for each row
for (i in 1:nrow(df)) {
  stats(df$attribute[i], df$group[i], df$type[i])
}

##Transfers----
###Shipping lcoations
attribute_vector <- c("Drop_ID", "DELIVERY_ID", "MATERIAL", "Orderline_ID",
     "volume_orderline")
group <- "SHIP_FROM_ID"
type_vector <-c("count", "count", "count", "count", "sum")
divisor_vector = c("Drops", "Deliveries", "Products", "Order_lines", "
    Volume")
percentage_per_attribute(attribute_vector, group, type_vector, divisor_
    vector)

###Shipment types
group <- "SHIPMENT_TYPE"
percentage_per_attribute(attribute_vector, group, type_vector, divisor_
    vector)

###Production locations
attribute_vector <- c("Orderline_ID", "volume_orderline")
group <- "PW_ID"
type_vector <-c("count", "sum")
divisor_vector = c("Order_lines", "Volume")
percentage_per_attribute(attribute_vector, group, type_vector, divisor_
    vector)

###Delivery types
group <- "Path"
percentage_per_attribute(attribute_vector, group, type_vector, divisor_
    vector)
rm(attribute_vector, group, type_vector, divisor_vector)
```

```r
##Transfer stats-----
df <- data.frame("attribute" = c("SHIP_FROM_ID", "SHIPMENT_TYPE", "PW_ID",
    "PW_ID", "PW_ID", "PW_ID",
                                  "Path", "Path", "Path", "Path"),
                 "group" = c("Week", "Week", "Week", "Drop_ID", "DELIVERY_
                     ID", "MATERIAL",
                              "Week", "Drop_ID", "DELIVERY_ID", "MATERIAL"),
                 "type" = c("count", "count", "count", "count", "count", "
                     count",
                             "count", "count", "count", "count"),
                 stringsAsFactors = F
)

for (i in 1:nrow(df)) {
  stats(df$attribute[i], df$group[i], df$type[i])
}

##Picking-----
###Percentages are determined along these dimensions
percentage_loose_picking(c("Week", "Drop_ID", "DELIVERY_ID", "MATERIAL", "
    Orderline_ID", "volume_orderline"),
                         c("count", "count", "count", "count", "count", "
                             sum"),
                         c("Weeks ordered", "Drops", "Deliveries", "
                             Products", "Order lines", "Volume")
)
###Initialize temperature and threshold
temp <- 1
threshold <- 0
###Continue checking the loose picking percentage when > threshold pallets
    are ordered
while(temp > 0.05) {

  df <- shipments %>%
    mutate(loose = if_else(plt_orderline > threshold, plt_orderline - floor
        (plt_orderline), 0)) %>%
    group_by(Key) %>%
    summarize('Loose picks' = sum(loose),
              Pallets = sum(plt_orderline)) %>%
    mutate(`Loose picks` = `Loose picks` / Pallets) %>%
    select(-Pallets)
  colnames(df) <- c("Key",
                    paste("Loose case picking percentage (orders > ",
                          threshold,
                          " pallets)")
  )
  ###Join df
  output <- left_join(output, df, by = "Key")
  ###Temperature is the maximum percentage
  temp <- max(output[[ncol(output)]], na.rm = T)
  ###Increment the threshold
  threshold <- threshold + 1
}
rm(temp, threshold)

##Drop days-----
```

```
df <- shipments %>%
  group_by(Key,
           Weekday = c("Sunday", "Monday", "Tuesday", "Wednesday", "
               Thursday", "Friday", "Saturday")[as.POSIXlt(shipments$DEL_
               DATE)$wday + 1]) %>%
  summarize(Value = length(unique(Drop_ID))) %>%
  dcast(Key ~ Weekday,
        value.var = "Value")
colnames(df) <- c("Key",
                  paste("Percentage drop day",
                        colnames(df[-1])))
df <- df[,c(1, 3, 7, 8, 6, 2, 4, 5)]
###Join df
output <- left_join(output, df, by = "Key")
output[(ncol(output) - 6):ncol(output)] <- lapply(output[(ncol(output) - 6)
   :ncol(output)], function(x) x / output$Drops)

##Days between drops-----
df <- shipments %>%
  group_by(Key,
           DEL_DATE) %>%
  summarize('Drops per day' = length(unique(Drop_ID)))

df <- do.call(rbind,
              lapply(split(df, df$Key),
                     function(d) {
                       d$`Days since last drop` <- c(NA, diff(d$DEL_DATE));
                           d
                     }))
###Temporarily disable warnings (due to customers with only one drop going
   to infinity)
defaultW <- getOption("warn")
options(warn = -1)

df <- df %>%
  group_by(Key) %>%
  summarize('Average count of Drops per day' = mean(`Drops per day`),
            'Min count of Drops per day' = min(`Drops per day`),
            'Max count of Drops per day' = max(`Drops per day`),
            'SD count of Drops per day' = sd(`Drops per day`),
            'CV count of Drops per day' = sd(`Drops per day`) / mean(`Drops
                per day`),
            'Average count of Days since last drop' = mean(`Days since last
                drop`, na.rm = T),
            'Min count of Days since last drop' = ifelse(is.infinite(min(`
               Days since last drop`, na.rm = T)),
                                                          0,
                                                          min(`Days since
                                                              last drop`, na.
                                                              rm = T)),
            'Max count of Days since last drop' = ifelse(is.infinite(max(`
               Days since last drop`, na.rm = T)),
                                                          0,
                                                          max(`Days since
                                                              last drop`, na.
                                                              rm = T)),
            'SD count of Days since last drop' = sd(`Days since last drop`,
```

```
                na.rm = T),
            'CV_count_of_Days_since_last_drop' = sd(`Days since last drop`)
                / mean(`Days since last drop`)
  )
###Re-enable warnings
options(warn = defaultW)
rm(defaultW)
###Join df
output <- left_join(output, df, by = "Key")

# ##Cost division----
# df <- shipments %>% select(Orderline_ID,
#                            Key)
# df <- left_join(df,
#                 costs,
#                 by = "Orderline_ID") %>%
#   select(-Orderline_ID) %>%
#   group_by(Key) %>%
#   summarize_all(sum)
#
# df[,c(-1,-ncol(df))] <- lapply(df[,c(-1,-ncol(df))], function(x) x / df$
    CTS)
# df <- df %>% select(-CTS)
# colnames(df) <- c("Key", paste(colnames(df[-1]), "- percentage"))
# ###Join df
# output <- left_join(output, df, by = "Key")

##Include CTS/volume----
output <- output %>%
  mutate(`CTS/volume` = CTS / Volume)

##Output to Excel----
output <- output %>%
  replace(is.na(.), 0)

wb <- createWorkbook(title = "Variable_set.xlsx")
addWorksheet(wb, "Output")
writeData(wb, sheet = "Output", output, rowNames = F, colNames = T)
saveWorkbook(wb, "Variable_set.xlsx", overwrite = T)
rm(list=ls())
```

Preprocessing.R

```
#Set working directory----
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

#Libraries----
library(readxl)
library(tidyverse)
library(openxlsx)
library(caret)

#Global constant variables----
zero <- 0.00001
set.seed(1)
cutoff_value <- 0.9
training_percentage <- 0.8
```

```r
#Data input----
data <- read_excel(paste(getwd(),
                          "/Variable_set.xlsx",
                              sep = "")
)

#Find intro----
intro_length <- 0
found <- F
while (found == F) {

  if (colnames(data[, intro_length + 1]) == "Weeks_with_demand") {
    found <- T
    intro <- data[,c(1:intro_length)]
    data <- data[,-c(1:intro_length)]
  }
  else {
    intro_length <- intro_length + 1
  }

}
rm(intro_length, found)

#Near zero variables in full data set----
data <- data %>%
  select(-nearZeroVar(data[,-ncol(data)]))

#Correlations full data set----
correlations <- cor(data[,-ncol(data)]) %>%
  replace(is.na(.), 0)
remove <- findCorrelation(correlations, cutoff = cutoff_value)
data <- data %>%
  select(-remove)

#Linear dependencies full data set----
combinations <- findLinearCombos(data[,-ncol(data)])
remove <- as.numeric(combinations$remove)
data <- data %>%
  select(-remove)

#Scale data----
data[,-ncol(data)] <- apply(data[,-ncol(data)], MARGIN = 2, FUN = function(
    X) (X - min(X))/diff(range(X)))
data <- data %>%
  replace(is.na(.), 0)

#Data splitting----
trainIndex <- createDataPartition(data$`CTS/volume`,
                                  p = training_percentage,
                                  list = FALSE,
                                  times = 1)

dataTrain <- data[ trainIndex,]
dataTest<- data[-trainIndex,]

#If all samples are in the training set change the percentage to 50%
```

```
if (nrow(dataTest) == 0) {
  trainIndex <- createDataPartition(data$`CTS/volume`,
                                    p = 0.5,
                                    list = FALSE,
                                    times = 1)

  dataTrain <- data[ trainIndex,]
  dataTest<- data[-trainIndex,]
}
rm(trainIndex)

#Near zero variables in training data set----
remove <- nearZeroVar(dataTrain[,-ncol(dataTrain)])
dataTrain <- dataTrain %>%
  select(-remove)
dataTest <- dataTest %>%
  select(-remove)
data <- data %>%
  select(-remove)

#Correlations training data set----
correlations <- cor(dataTrain[,-ncol(dataTrain)]) %>%
  replace(is.na(.), 0)
remove <- findCorrelation(correlations, cutoff = cutoff_value)
dataTrain <- dataTrain %>%
  select(-remove)
dataTest <- dataTest %>%
  select(-remove)
data <- data %>%
  select(-remove)

#Linear dependencies training data set----
combinations <- findLinearCombos(dataTrain[,-ncol(dataTrain)])
remove <- as.numeric(combinations$remove)
dataTrain <- dataTrain %>%
  select(-remove)
dataTest <- dataTest %>%
  select(-remove)
data <- data %>%
  select(-remove)
rm(combinations, correlations, remove)

#Data to Excel----
dataTrain <- dataTrain %>%
  replace(is.na(.), 0)
dataTest <- dataTest %>%
  replace(is.na(.), 0)
data <- data %>%
  replace(is.na(.), 0)
intro <- intro %>%
  replace(is.na(.), 0)

wb <- createWorkbook(title = "Final variable set.xlsx")
addWorksheet(wb, "Intro")
writeData(wb, sheet = "Intro", intro, rowNames = F, colNames = T)
addWorksheet(wb, "Data")
writeData(wb, sheet = "Data", data, rowNames = F, colNames = T)
```

```
addWorksheet(wb, "Training")
writeData(wb, sheet = "Training", dataTrain, rowNames = F, colNames = T)
addWorksheet(wb, "Testing")
writeData(wb, sheet = "Testing", dataTest, rowNames = F, colNames = T)
saveWorkbook(wb, "Final_variable_set.xlsx", overwrite = T)
rm(list=ls())
```

Model fitting.R

```
#Set working directory----
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

#Libraries----
library(readxl)
library(tidyverse)
library(openxlsx)
library(caret)
library(lubridate)

#Global constant variables----
zero <- 0.00001

#Data input----
dataTraining <- read_excel(paste(getwd(),
                                 "/Final_variable_set.xlsx",
                                 sep = ""),
                           sheet = "Training"
)

dataTesting <- read_excel(paste(getwd(),
                                "/Final_variable_set.xlsx",
                                sep = ""),
                          sheet = "Testing"
)

#Model fitting----
fit_models <- function(cv_number, cv_repeat, sa_iters, ga_iters) {

  set.seed(1)

  variables <- c("cv_number" = cv_number,
                 "cv_repeat" = cv_repeat)

  ##Linear model
  start_time <- proc.time()

  lm_control <- trainControl(method = "repeatedcv",
                             number = cv_number,
                             repeats = cv_repeat,
                             verboseIter = T)

  lm_grid <- expand.grid(intercept = c(T, F))

  options(warn = -1)
  lm_model <- train(`CTS/volume`~.,
                    data = as.data.frame(dataTraining),
                    method = "lm",
```

```
                    trControl = lm_control,
                    tuneGrid = lm_grid)
options(warn = 1) #Suppress warning for rank defficiency

#Create and store statistics
run_time <- proc.time() - start_time
lm_variables <- c("variables" = length(lm_model$coefnames)) #Differs per
    model, check this!
lm_time <- c("time" = run_time[3])
lm_predictions <- predict(lm_model, dataTesting)
lm_performance <- postResample(lm_predictions, dataTesting$`CTS/volume`)
lm <- list(c("type" = "lm", variables, lm_variables, lm_time, lm_
    performance))

output <- data.frame(matrix(unlist(lm), nrow=length(lm), byrow=T)) #First
     model goes directly to output
colnames(output) <- c("type", "cv_number", "cv_repeat", "variables", "
    time", "RMSE", "R^2", "MAE")

##Polynomial model
start_time <- proc.time()

gaussprPoly_control <- trainControl(method = "repeatedcv",
                                    number = cv_number,
                                    repeats = cv_repeat,
                                    verboseIter = T)

#gaussprPoly_grid <- expand.grid()

gaussprPoly_model <- train(`CTS/volume`~.,
                            data = as.data.frame(dataTraining),
                            method = "gaussprPoly",
                            trControl = gaussprPoly_control)
#tuneGrid = gaussprPoly_grid)

#Create and store statistics
run_time <- proc.time() - start_time
gaussprPoly_variables <- c("variables" = length(gaussprPoly_model$
    coefnames)) #Differs per model, check this!
gaussprPoly_time <- c("time" = run_time[3])
gaussprPoly_predictions <- predict(gaussprPoly_model, dataTesting)
gaussprPoly_performance <- postResample(gaussprPoly_predictions,
    dataTesting$`CTS/volume`)
gaussprPoly <- list(c("type" = "gaussprPoly", variables, gaussprPoly_
    variables, gaussprPoly_time, gaussprPoly_performance))

gaussprPoly <- data.frame(matrix(unlist(gaussprPoly), nrow=length(
    gaussprPoly), byrow=T))
colnames(gaussprPoly) <- c("type", "cv_number", "cv_repeat", "variables",
     "time", "RMSE", "R^2", "MAE")
output <- rbind(output, gaussprPoly)

##Elasticnet model
start_time <- proc.time()

enet_control <- trainControl(method = "repeatedcv",
                             number = cv_number,
```

```
                                        repeats = cv_repeat,
                                        verboseIter = T)

#enet_grid <- expand.grid()

enet_model <- train(`CTS/volume`~.,
                          data = as.data.frame(dataTraining),
                          method = "enet",
                          trControl = enet_control)
#tuneGrid = enet_grid)

#Create and store statistics
run_time <- proc.time() - start_time
enet_variables <- c("variables" = length(enet_model$coefnames)) #Differs
    per model, check this!
enet_time <- c("time" = run_time[3])
enet_predictions <- predict(enet_model, dataTesting)
enet_performance <- postResample(enet_predictions, dataTesting$`CTS/
    volume`)
enet <- list(c("type" = "enet", variables, enet_variables, enet_time,
    enet_performance))

enet <- data.frame(matrix(unlist(enet), nrow=length(enet), byrow=T))
colnames(enet) <- c("type", "cv_number", "cv_repeat", "variables", "time"
    , "RMSE", "R^2", "MAE")
output <- rbind(output, enet)

##Ridge model
start_time <- proc.time()

ridge_control <- trainControl(method = "repeatedcv",
                                        number = cv_number,
                                        repeats = cv_repeat,
                                        verboseIter = T)

#ridge_grid <- expand.grid()

ridge_model <- train(`CTS/volume`~.,
                          data = as.data.frame(dataTraining),
                          method = "ridge",
                          trControl = ridge_control)
#tuneGrid = ridge_grid)

#Create and store statistics
run_time <- proc.time() - start_time
ridge_variables <- c("variables" = length(ridge_model$coefnames)) #
    Differs per model, check this!
ridge_time <- c("time" = run_time[3])
ridge_predictions <- predict(ridge_model, dataTesting)
ridge_performance <- postResample(ridge_predictions, dataTesting$`CTS/
    volume`)
ridge <- list(c("type" = "ridge", variables, ridge_variables, ridge_time,
     ridge_performance))

ridge <- data.frame(matrix(unlist(ridge), nrow=length(ridge), byrow=T))
colnames(ridge) <- c("type", "cv_number", "cv_repeat", "variables", "time
    ", "RMSE", "R^2", "MAE")
```

```
output <- rbind(output, ridge)

##Lasso model
start_time <- proc.time()

lasso_control <- trainControl(method = "repeatedcv",
                              number = cv_number,
                              repeats = cv_repeat,
                              verboseIter = T)

#lasso_grid <- expand.grid()

lasso_model <- train(`CTS/volume`~.,
                     data = as.data.frame(dataTraining),
                     method = "lasso",
                     trControl = lasso_control)
#tuneGrid = lasso_grid)

#Create and store statistics
run_time <- proc.time() - start_time
lasso_variables <- c("variables" = length(lasso_model$coefnames)) #
    Differs per model, check this!
lasso_time <- c("time" = run_time[3])
lasso_predictions <- predict(lasso_model, dataTesting)
lasso_performance <- postResample(lasso_predictions, dataTesting$`CTS/
    volume`)
lasso <- list(c("type" = "lasso", variables, lasso_variables, lasso_time,
     lasso_performance))

lasso <- data.frame(matrix(unlist(lasso), nrow=length(lasso), byrow=T))
colnames(lasso) <- c("type", "cv_number", "cv_repeat", "variables", "time
    ", "RMSE", "R^2", "MAE")
output <- rbind(output, lasso)

#Feature selection RFE----
start_time <- proc.time()

rfe_control <- rfeControl(functions = rfFuncs,
                          method = "repeatedcv",
                          number = cv_number,
                          repeats = cv_repeat,
                          verbose = T)

rfe_model <- rfe(as.matrix(dataTraining[-ncol(dataTraining)]),
                 as.vector(dataTraining$`CTS/volume`),
                 sizes = c(3:(ncol(dataTraining) - 2)),
                 rfeControl = rfe_control)

#Create and store statistics
run_time <- proc.time() - start_time
rfe_variables <- c("variables" = rfe_model$optsize)
rfe_time <- c("time" = run_time[3])
rfe_predictions <- predict(rfe_model, dataTesting)
rfe_performance <- postResample(rfe_predictions, dataTesting$`CTS/volume
    `)
rfe <- list(c("type" = "rfe", variables, rfe_variables, rfe_time, rfe_
    performance))
```

```
  rfe <- data.frame(matrix(unlist(rfe), nrow=length(rfe), byrow=T))
  colnames(rfe) <- c("type", "cv_number", "cv_repeat", "variables", "time",
      "RMSE", "R^2", "MAE")
  output <- rbind(output, rfe)

  #Save the best model fit
  output[,-1] <- lapply(output[,-1], function(x) as.numeric(as.character(x)
    ))
  output$best <- output$RMSE / output$`R^2` #For a low RMSE and R^2
  model <<- eval(parse(text = paste(output$type[output$best == min(output$
     best)],
                                    "_model",
                                    sep = "")))

  return(output)

}

#Experiments with models
cv_repeat <- 1
cv_number <- 2

if (exists("output")) {
  output <- rbind(output, fit_models(cv_number, cv_repeat))
} else {
  output <- fit_models(cv_number, cv_repeat)
}

save(model, file="model.Rdata")

##Output to Excel----
output <- output %>%
  replace(is.na(.), 0)

wb <- createWorkbook(title = "Model_fit_output.xlsx")
addWorksheet(wb, "Output")
writeData(wb, sheet = "Output", output, rowNames = F, colNames = T)
saveWorkbook(wb, "Model_fit_output.xlsx", overwrite = T)
rm(list=ls())
```

Root causes

```
#Set working directory----
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

#Libraries----
library(readxl)
library(tidyverse)
library(openxlsx)
library(caret)
library(lubridate)

#Global constant variables----
zero <- 0.00001

#Load the best performing model and data file
```

```
load("model.Rdata")

models <- read_excel(paste(getwd(),
                           "/Model_fit_output.xlsx",
                           sep = ""),
                     sheet = "Output"
)

intro <- read_excel(paste(getwd(),
                          "/Final_variable_set.xlsx",
                          sep = ""),
                    sheet = "Intro"
)

data <- read_excel(paste(getwd(),
                         "/Final_variable_set.xlsx",
                         sep = ""),
                   sheet = "Data"
)

output <- read_excel(paste(getwd(),
                           "/Variable_set.xlsx",
                           sep = ""),
                     sheet = "Output"
)

if (!is.null(model[[1]])) {
  output <- select(output, c(gsub("`", "", model$coefnames), `CTS/volume`))
  data <- select(data, c(gsub("`", "", model$coefnames), `CTS/volume`))
} else { #A different method is required for rfe
  output <- select(output, c(model$optVariables, `CTS/volume`))
  data <- select(data, c(model$optVariables, `CTS/volume`))
}

output <- cbind(intro, output)
data <- cbind(intro, data)
intro_length <- ncol(intro)
rm(intro)

#Get the variable importance
if (!is.null(model[[1]])) {
  var_importance <- t(varImp(model)[[1]])
} else {
  var_importance <- t(as.data.frame(model$fit$importanceSD))
}

#Check if we don't have incorrect variables
var_importance <- t(var_importance[, colnames(data[(intro_length + 1):(ncol
    (data) - 1)])])
rownames(var_importance) <- c("Importance")
variables <- ncol(var_importance)

#Set variable threshold for when a variable is comparable with a scale from
     0.01 (most important) to 0.1 (least important)
var_threshold <- 0.1 - (var_importance - min(var_importance)) * 0.09 / (max
    (var_importance) - min(var_importance))
```

```
#Get the Mean Absolute Error
MAE <- models$MAE[models$best == min(models$best)]

#Create output matrix
savings <- select(data, c(1:intro_length, ncol(data)))
savings$peer1 <- NA
savings$focus1 <- NA
savings$savings1 <- 0
savings$peer2 <- NA
savings$focus2 <- NA
savings$savings2 <- 0
savings$peer3 <- NA
savings$focus3 <- NA
savings$savings3 <- 0

#Find comparable customers and see potential savings
start_time <- proc.time()
for (current in 1:nrow(data)) {

  if (current %% 100 == 0) {
    cat(current, "/", nrow(data), "\n")
    run_time <- proc.time() - start_time
    cat(run_time[3], "\n")
  }

  #Compute scores
  for (row in 1:nrow(data)) {

    if (row == current) {

      data[row, "comparable_variables"] <- -1

    } else {
      #Get results Keep rows with a lower CTS/volume with a correction for the MAE, and a comparable volume (no more than 10% deviation)
      if (data[[row, "CTS/volume"]] + MAE < data[[current, "CTS/volume"]] & abs(data[[row, "volume"]] - data[[current, "volume"]]) <= 0.1 * data[[current, "volume"]]) {

        comparable_variables <- 0
        for (var in 1:variables) {
          #Check if the variable is within 0.1 of the current (variables are still scaled)
          if (abs(data[[row, intro_length + var]] - data[[current, intro_length + var]]) < var_threshold[var]) {

            comparable_variables <- comparable_variables + 1
          #If this is the first differing variable put it down as the focus
          } else if (var - comparable_variables == 1) {

            data[row, "focus"] <- colnames(data)[intro_length + var]
          #Else note the number of focus variables
          } else {

            data[row, "focus"] <- paste(var - comparable_variables, "focus variables")
          }
```

```
        }

        data[row, "comparable_variables"] <- comparable_variables
      } else {
        data[row, "comparable_variables"] <- -1
      }
    }
  }

  #And with comparable variables
  result <- subset(data, comparable_variables > 0)
  #Sort results by comparable variables and CTS/volume
  result <- result[order(-result$comparable_variables, result$`CTS/volume`)
      ,]

  #Store three most promising peers
  if (nrow(result) > 0) {

    for (i in 1:min(3, nrow(result))) {
      #Take the first peer
      savings[current, paste("peer", as.character(i), sep ="")] <- result[i
          , "Key"]
      #Read the focus for this peer
      savings[current, paste("focus", as.character(i), sep ="")] <- result[
          i, "focus"]
      #Calculate the savings correcting for the MAE
      savings[current, paste("savings", as.character(i), sep ="")] <- data[
          current, "CTS"] -
        (result[i, "CTS/volume"] + MAE) * data[current, "volume"]
    }
  }
}


wb <- createWorkbook(title = "Root-cause analysis.xlsx")
addWorksheet(wb, "Key drivers")
writeData(wb, sheet = "Key drivers", cbind(var_importance, MAE), rowNames =
     T, colNames = T)
addWorksheet(wb, "Entities")
writeData(wb, sheet = "Entities", output, rowNames = F, colNames = T)
addWorksheet(wb, "Potential savings")
writeData(wb, sheet = "Potential savings", savings, rowNames = F, colNames
    = T)
saveWorkbook(wb, "Root-cause analysis.xlsx", overwrite = T)
rm(list=ls())
```

L Model fitting experiments

The figure below shows the MAE for different models at different cutoff values.

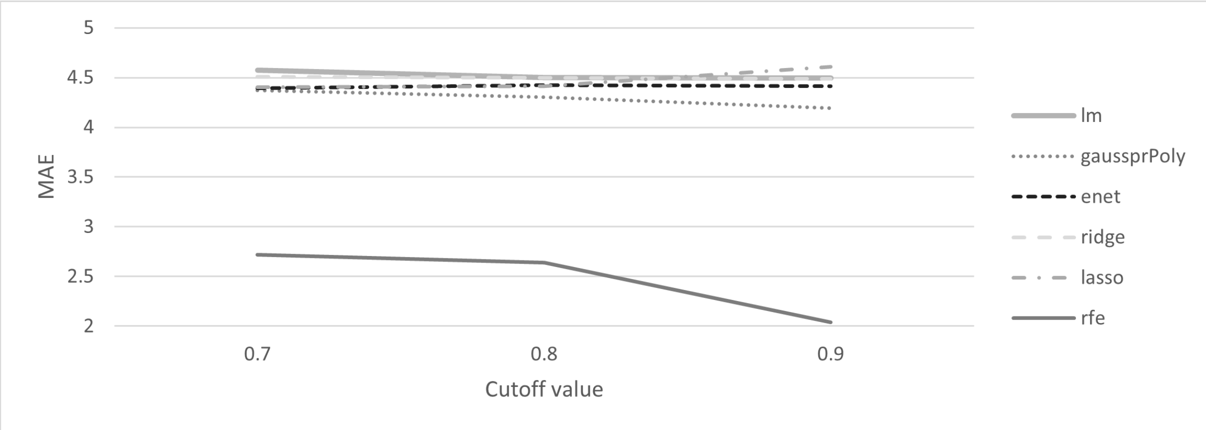


Figure 26: The Mean Average Error for six different model fits considering cutoff values of 0.7, 0.8 and 0.9

For most models, the MAE stayed between 4 and 5, showing a slight improvement when the cutoff value increased. The Recursive Feature Elimination random forest model (rfe) showed a significant performance increase with a higher cutoff value, which is understandable as the model is better with handling collinearity. A higher cutoff value means the script removes less correlating variables, which leads to more variables in the final set. We also experimented with cutoff values above 0.9, but this led to problems with fitting models due to high collinearity. For the RMSE and R^2, we observed similar trends as with the MAE. So, we chose a cutoff value of 0.9, which avoids "very strong correlations". The last part of this appendix shows the results of all runs for this experiment.

The figure below shows the result of tests up to five-fold cross-validation. The last part of this appendix shows the results of all runs for this experiment, including runs with op to ten folds.

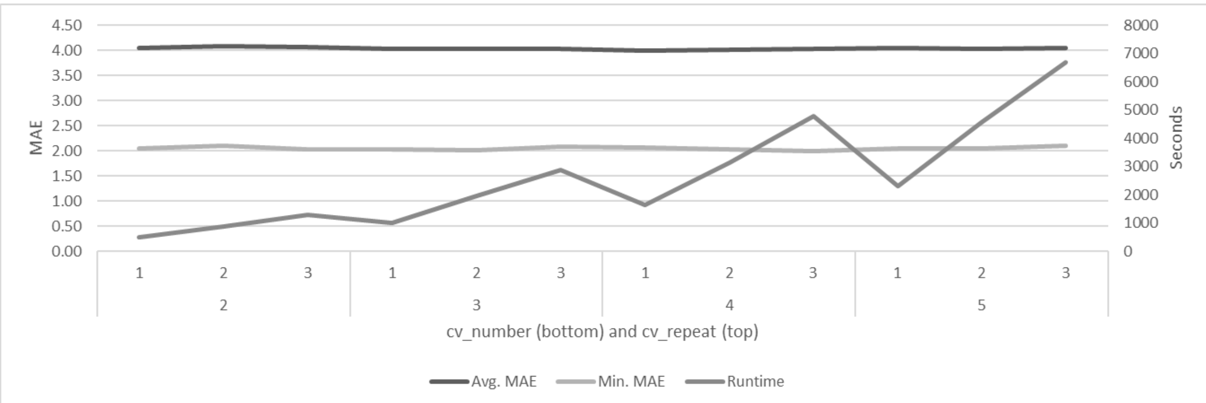


Figure 27: The average and minimum Mean Average Error, and the runtime for six different model fits considering every combination of repeats from one to thee and folds from two to five

Neither the average MAE nor the minimum MAE of the six models changed much when increasing the number of repeats or folds. However, the runtime did increase heavily. Again, we saw no different behavior concerning the RMSE or R^2. Therefore, a single run with 2-fold cross-validation is the most effective option in this experiment.

type	cutoff	cv_number	cv_repeat	variables	time	RMSE	R^2	MAE
lm	0.7	2	1	34	0.86	9.93	0.24	4.57
gaussprPoly	0.7	2	1	34	71.53	9.87	0.25	4.37
enet	0.7	2	1	34	1.53	10.03	0.23	4.39
ridge	0.7	2	1	34	1.4	9.91	0.24	4.51
lasso	0.7	2	1	34	0.94	10.05	0.23	4.41
rfe	0.7	2	1	34	628.25	9.99	0.30	2.72
lm	0.8	2	1	41	0.98	9.91	0.24	4.50
gaussprPoly	0.8	2	1	41	72.45	9.80	0.26	4.30
enet	0.8	2	1	41	1.86	9.89	0.25	4.43
ridge	0.8	2	1	41	1.89	9.91	0.24	4.50
lasso	0.8	2	1	41	1.12	9.89	0.25	4.41
rfe	0.8	2	1	41	1151.36	10.25	0.29	2.64
lm	0.9	2	1	59	0.91	9.81	0.26	4.50
gaussprPoly	0.9	2	1	59	106.03	9.66	0.28	4.19
enet	0.9	2	1	59	2.52	9.83	0.25	4.41
ridge	0.9	2	1	59	2.45	9.79	0.26	4.49
lasso	0.9	2	1	59	1.47	10.31	0.20	4.61
rfe	0.9	2	1	14	387.3	8.00	0.51	2.04

Cutoff value = 0.9:

Model	cv_number	cv_repeat	Variables	Time	RMSE	R^2	MAE
enet	2	1	59	2.52	9.83	0.25	4.41
gaussprPoly	2	1	59	106.03	9.66	0.28	4.19
lasso	2	1	59	1.47	10.31	0.20	4.61
lm	2	1	59	0.91	9.81	0.26	4.50
rfe	2	1	14	387.30	8.00	0.51	2.04
ridge	2	1	59	2.45	9.79	0.26	4.49
enet	2	2	59	3.75	9.83	0.25	4.41
gaussprPoly	2	2	59	157.14	9.80	0.26	4.36
lasso	2	2	59	1.83	10.31	0.20	4.61
lm	2	2	59	1.42	9.81	0.26	4.50
rfe	2	2	17	726.15	8.00	0.51	2.10
ridge	2	2	59	3.04	9.79	0.26	4.49
enet	2	3	59	4.43	9.83	0.25	4.41
gaussprPoly	2	3	59	236.89	9.80	0.26	4.36
lasso	2	3	59	1.90	10.31	0.20	4.61
lm	2	3	59	3.04	9.81	0.26	4.50
rfe	2	3	20	1048.50	7.95	0.51	2.04
ridge	2	3	59	3.79	9.79	0.26	4.49
enet	3	1	59	2.83	9.83	0.25	4.41
gaussprPoly	3	1	59	228.46	9.66	0.28	4.19
lasso	3	1	59	1.49	10.31	0.20	4.61
lm	3	1	59	0.99	9.81	0.26	4.41
rfe	3	1	18	762.73	7.97	0.51	2.03
ridge	3	1	59	2.72	9.79	0.26	4.49
enet	3	2	59	4.65	9.83	0.25	4.41
gaussprPoly	3	2	59	421.56	9.80	0.26	4.36
lasso	3	2	59	2.05	9.75	0.27	4.39
lm	3	2	59	1.37	9.81	0.26	4.50
rfe	3	2	24	1537.98	7.91	0.52	2.02
ridge	3	2	59	4.63	9.79	0.26	4.49
enet	3	3	59	6.56	9.83	0.25	4.41
gaussprPoly	3	3	59	592.64	9.80	0.26	4.36

Model	cv_number	cv_repeat	Variables	Time	RMSE	R^2	MAE
lasso	3	3	59	2.63	9.75	0.27	4.39
lm	3	3	59	1.68	9.81	0.26	4.50
rfe	3	3	25	2274.92	8.07	0.50	2.07
ridge	3	3	59	6.22	9.79	0.26	4.49
enet	4	1	59	3.67	9.83	0.25	4.41
gaussprPoly	4	1	59	403.10	9.66	0.28	4.19
lasso	4	1	59	1.78	9.75	0.27	4.39
lm	4	1	59	1.01	9.81	0.26	4.41
rfe	4	1	24	1223.68	8.01	0.51	2.06
ridge	4	1	59	3.57	9.79	0.26	4.49
enet	4	2	59	6.95	9.83	0.25	4.41
gaussprPoly	4	2	59	726.67	9.80	0.26	4.36
lasso	4	2	59	2.74	9.75	0.27	4.39
lm	4	2	59	1.58	9.81	0.26	4.41
rfe	4	2	23	2385.81	8.06	0.50	2.04
ridge	4	2	59	6.41	9.79	0.26	4.49
enet	4	3	59	9.52	9.83	0.25	4.41
gaussprPoly	4	3	59	1151.30	9.66	0.28	4.19
lasso	4	3	59	3.80	10.31	0.20	4.61
lm	4	3	59	2.00	9.81	0.26	4.50
rfe	4	3	22	3598.00	7.84	0.53	2.00
ridge	4	3	59	8.78	9.79	0.26	4.49
enet	5	1	59	5.84	9.83	0.25	4.41
gaussprPoly	5	1	59	614.99	9.66	0.28	4.19
lasso	5	1	59	2.27	10.31	0.20	4.61
lm	5	1	59	1.14	9.81	0.26	4.50
rfe	5	1	24	1671.08	8.17	0.49	2.06
ridge	5	1	59	5.12	9.79	0.26	4.49
enet	5	2	59	8.71	9.83	0.25	4.41
gaussprPoly	5	2	59	1153.39	9.80	0.26	4.36
lasso	5	2	59	3.34	9.75	0.27	4.39
lm	5	2	59	1.89	9.81	0.26	4.50
rfe	5	2	23	3382.52	7.95	0.51	2.04
ridge	5	2	59	8.09	9.79	0.26	4.49
enet	5	3	59	11.92	9.83	0.25	4.41
gaussprPoly	5	3	59	1649.50	9.80	0.26	4.36
lasso	5	3	59	4.63	9.75	0.27	4.39
lm	5	3	59	2.11	9.81	0.26	4.50
rfe	5	3	24	5006.28	8.06	0.50	2.10
ridge	5	3	59	11.08	9.79	0.26	4.49
enet	6	1	59	5.19	9.83	0.25	4.41
gaussprPoly	6	1	59	566.90	9.80	0.26	4.36
lasso	6	1	59	2.37	9.75	0.27	4.39
lm	6	1	59	1.16	9.81	0.26	4.41
ridge	6	1	59	4.98	9.79	0.26	4.49
enet	7	1	59	5.96	9.83	0.25	4.41
gaussprPoly	7	1	59	747.51	9.66	0.28	4.19
lasso	7	1	59	2.41	9.75	0.27	4.39
lm	7	1	59	1.04	9.81	0.26	4.41
ridge	7	1	59	6.90	9.79	0.26	4.49
enet	8	1	59	5.64	9.83	0.25	4.41
gaussprPoly	8	1	59	946.19	9.80	0.26	4.36
lasso	8	1	59	2.28	9.75	0.27	4.39
lm	8	1	59	1.12	9.81	0.26	4.41
ridge	8	1	59	6.23	9.79	0.26	4.49

Model	cv_number	cv_repeat	Variables	Time	RMSE	R^2	MAE
enet	9	1	59	7.29	9.83	0.25	4.41
gaussprPoly	9	1	59	1040.75	9.66	0.28	4.19
lasso	9	1	59	2.75	9.75	0.27	4.39
lm	9	1	59	1.11	9.81	0.26	4.41
ridge	9	1	59	6.72	9.79	0.26	4.49
enet	10	1	59	6.80	9.83	0.25	4.41
gaussprPoly	10	1	59	1252.03	9.80	0.26	4.36
lasso	10	1	59	2.85	9.75	0.27	4.39
lm	10	1	59	1.30	9.81	0.26	4.41
ridge	10	1	59	6.73	9.79	0.26	4.49

M Survey performance new tool

User experience

The results show how many stars each respondent awarded.

ID	Position	How do you rate the user experience of the QlikView tool?	How do you rate the user experience of the Power BI tool?
1	OpCo	3	4
2	Global	4	4
3	Global	1	3
4	Global	2	4

Features old tool

Response	Score
Much worse	-2
Slightly worse	-1
Unchanged	0
Slightly improved	1
Much improved	2

ID	Position	Filter functionality	Cost per volume-unit graphs	Scatter/drop analysis graph	Delivery profile graph	Key numbers tables	Details tables	Map	Volume flow (overview)
1	OpCo	1	1	-1	1	2	2	1	1
2	Global	0	2	0	0	2	2	2	2
3	Global	-1	0	0	0	0	0	1	0
4	Global	1	1	2	1	1	1	2	2

Tool assessment

Response	Score
Completely disagree	0
Mostly disagree	1
Somewhat disagree	2
Somewhat agree	3
Mostly agree	4
Completely agree	5

ID	Position	Using the tool can improve my job performance	Using the tool can make it easier to do my job	Using the tool can significantly increase the quality of output on my job	Learning to operate the tool was easy for me	Using the tool involves little time doing mechanical operations	I find it easy to get the tool to do what I want it to do
1	OpCo	4	4	4	3	3	3
2	Global	5	4	4	4	4	4
3	Global	4	3	4	5	5	4
4	Global	4	4	3	3	3	4

New features

The results show how many stars each respondent awarded.

ID	Position	Hierarchies	Main page	General layout	Help page	Combinations tables	Trend page	Truck utilization page	Ship_from locations page	Delivering from the closest DC page	Histograms page	Segmentation page
1	OpCo	3	3	3	4	3	3	2	3	2	2	4
2	Global	4	4	4	3	3	3	4	2	4	3	4
3	Global	4	3	3	4	2	4	3	4	2	2	4
4	Global	3	4	3	3	3	3	4	4	3	2	3

ID	Position	Complexity page	Logistics Trade Terms page	Ask a question about your data page	Key influencers page	Decomposition tree page
1	OpCo	3	4	2	2	4
2	Global	4	2	4	4	4
3	Global	3	3	3	2	2
4	Global	3	2	3	2	2

Root-cause analysis

Response	Score
Completely disagree	0
Mostly disagree	1
Somewhat disagree	2
Somewhat agree	3
Mostly agree	4
Completely agree	5

ID	Position	The method can improve my job performance	The method can make it easier to do my job	The method can significantly increase the quality of output on my job	Understanding the method was easy for me	I feel the method involves little time doing mechanical operations	I expect it would be easy to get the method to do what I want it to do
1	OpCo						
2	Global	5	5	5	4	4	4
3	Global	4	3	4	3	4	2
4	Global	4	4	4	3	3	3