

UNIVERSITY OF TWENTE.

Dashboard as a Decision Support Tool to
Minimize Empty Miles in Shipping
Network: Locating Region for
Improvement



Author

Novianita Isnaynizahra

Research Information

Industrial Engineering and Management Bachelor Thesis

Title

Dashboard as a Decision Support Tool to Minimize Empty Miles in Shipping Network:
Locating Region for Improvement

Author

Novianita Isnaynizahra
novianitaisnaynizahra@student.utwente.nl

Educational Program

Bachelor Industrial Engineering and Management
Department of Industrial Engineering and Business Information Systems

Educational Institution

University of Twente
Industrial Engineering and Management
PO BOX 217, 7500 AE Enschede, tel (053)489111

Host Organization

Emons Group B.V
Rijksweg 4, 6596 AB Milsbeek

Examination Committee

First Supervisor: Sebastian Piest
Second Supervisor: Amin Asadi
External Supervisor: Kevin Lemmens

Date

10 October, 2022

Version

This is the second version of the research report. Due to confidentiality, some pictures, customer names, and location names in this document are being anonymized for publication

Preface

Dear reader,

This document you are about to read is my bachelor thesis “Dashboard as a Decision Support Tool to Minimize Empty Miles in Shipping Network: Locating Region for Improvement”. This research was conducted for approximately six months since the first contact with the company. It has been a great experience implementing knowledges I acquired during my bachelor program into an industry problem. More than that, I also have learned a lot about my capacities as a person when conducting this independent project in an extensive period of time. The journey was often quite challenging. However, I am really grateful that I was able to passed the challenging phase and arrived at this point. This thesis would not be able to finish without the help and support of many people. Therefore, here I would like to express my gratitude for the people who have supported me throughout this process.

Firstly, I would like to thank Sebastian Piest as my first supervisor for his enormous support and guidance since the day I was searching for my thesis assignment until the last stage of my project. Sebastian has helped me to get connected with Emons, the company where I did my thesis assignment. It was not easy to find a company which can give a bachelor assignment to international students like me, however, luckily Sebastian helped me in that process. Moreover, his continuous encouragement, feedback, and advice have really helped me in the progression of my thesis. Further, I also thank my second supervisor Amin Asadi who has dedicated his time to give valuable feedback for my thesis.

I also would like to express appreciation to Emons for accommodating me as a research intern and providing the necessary support during the internship. Enormous gratitude I also give to Kevin Lemmens as my supervisor in Emons for his time, support, and continuous guidance over the past few months. I appreciate his patience and dedication to answer the enormous amounts of questions I gave him. I have gained quite substantial insight about the complexity of the shipping network. Working on the assignment at Emons, specifically at the Service Centre Department has been really enjoyable for me. During this assignment I have developed new technical skills by learning and working on a software I have never used before. Overall it was a rewarding experience working on this final assignment.

Enschede,
October, 2022

Management Summary

As a freight logistic company, Emons has a dynamic shipping network that connects loadings and unloading locations across 12 countries in Europe and the United Kingdom. In the freight logistic network, empty miles is a common problem in which the truck travels empty during backhaul and when driving to the next loading location. The empty miles are certainly expensive. For that reason, Emons tries to minimize the empty miles by loading additional pallets during the shipment trip and searching for the additional customer to increase the network density. However, it is challenging to determine which region needs to be focused on for adding customers or pallets during the trip, more empty miles could consequently appear if the additional pallets or customers' locations are too far. Before deciding on the improvement region, the company needs to understand how different location or lane contributes to the total empty miles of the network. Therefore, this research is conducted to create a decision support tool that would determine how lanes contribute to the empty miles of the total network, this tool will then be used to determine improvement region.

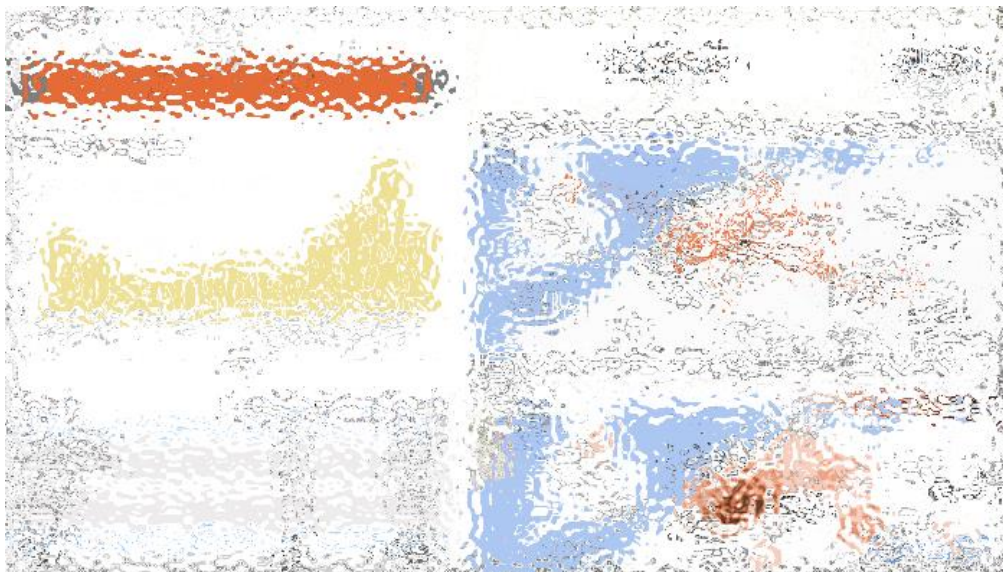
To conduct essential improvement of the approach company does, an evaluation of Emons' current tool was conducted. Currently, the company uses a shipment frequency matrix to understand the network balance for each country. This method provides an overview of the loading and unloading locations distribution for each country. However, it has a limitation on providing interactive and comprehensive insight for dynamic shipping network with high complexity. Therefore, the context analyses lead to the conclusion that there is a need to have a comprehensive network insight as a decision support tool.

The literature review revealed the potential of utilizing machine learning approaches in the decision support tool. It was found that the regression method is an approach that is commonly used for a shipping network problem. In addition, eight different visualization techniques that are used for transport network data visualization were also collected from the literatures. The visualization techniques were evaluated to see the potential contribution of the expected tool. From the eight collected techniques, we select four types of visualization that are most closely fulfilled the tool objective. Through visualization techniques exploration and literature review, self-service machine learning visualization features in Power BI were selected to be the primary feature of the decision support tool, moreover, the four other selected visualization techniques are used as supporting features of the tool.

The decision support tool prototype is created in a form of a dashboard. The dashboard is designed with a scaling-down concept that provides systematic data insight from the broad overall shipping network view into the specific small region of the network. The dashboard is divided into four main sections as follows:

1. Overall Network Performance Page: The first section of the tool displays the customers' performance based on the overall network evaluation.
2. Key Influencers Page: The second section provides general insight of which cities and customers are contributing to high empty miles of the shipping network.
3. Decomposition Tree and Tree Details Pages: The third section enables users to explore the empty miles distribution across different cities and countries using a decomposition tree. Further, it is connected to the tree detail page to see the network visualization empty miles percentage.
4. Lane Analysis Page: The fourth section displays a table that contains empty miles contribution for each lane in the shipping network.

Among the four sections in the dashboard, section three has the most interactive feature. The picture of the decomposition tree page and the tree details page are displayed below to represent the dashboard prototype. To test the dashboard, an analysis is conducted using the two pages below and combined with the lane analysis section. The result shows that high empty miles between R and P, further there is high empty miles percentage around the country Z.



To summarize, the prototype has been able to solve the core problem of determining empty KM contribution to the network. It has been validated that the tool has the potential implementation for the Sales department. However, the lane analysis section has not been sufficient to be implemented in practice. It is mainly due to the incomplete components that need to be added and taken into account for determining lanes' added value to the overall network. Thus, further research should be conducted to optimize the tool's capabilities.

Lists of Figures

Figure 1.1: Stakeholder onion chart.....	2
Figure 1.2: Problem cluster	4
Figure 1.3: DSRM process model	6
Figure 2.1: Network locations distribution	9
Figure 0.2: Shipping network illustration	11
Figure 4.1: Python code to validate the location attributes	19
Figure 4.2: Slicer	22
Figure 4.3: Timeline slicer	23
Figure 4.4: Clustered bar chart	23
Figure 4.5: Ribbon chart	23
Figure 4.6: Column chart	24
Figure 4.7: Table	24
Figure 4.8: Heat map	25
Figure 4.9: Flow map	25
Figure 4.10: Card.....	25
Figure 4.11: Buttons	26
Figure 4.12: Drill through feature	26
Figure 5.1: Page guideline framework	27
Figure 5.2: Overall network performance page	28
Figure 5.3: Key influencers page	28
Figure 5.4: Decomposition tree page (before being operated)	29
Figure 5.5: Decomposition tree page (while being operated)	30
Figure 5.6: Tree detail page	30
Figure 5.7: Lane analysis page	31
Figure 5.8: Analysing customer B performance using lane analysis page	32
Figure 5.9: Lane analysis table after removing several lanes	33
Figure 5.10: Previous empty KM decomposition tree of Customer B	33
Figure 5.11: Previous location heatmap of city S.....	34

<i>Figure 5.12: Next empty KM decomposition tree of Customer B</i>	<i>34</i>
--	-----------

<i>Figure 5.13: Tree detail of next empty KM.....</i>	<i>35</i>
---	-----------

List of Tables

<i>Table 1.1: Research questions</i>	8
<i>Table 2.1: Weekly calculation</i>	10
<i>Table 2.2: Aggregated weekly calculation of shipment frequency</i>	10
<i>Table 2.3: Dataset attributes connections with location and lanes</i>	11
<i>Table 3.1: Visualization Techniques Literature Review Summary</i>	16
<i>Table 4.1: Dashboard requirement framework</i>	18
<i>Table 4.2: Attributes data types</i>	20
<i>Table 4.3: Dashboard sections</i>	21
<i>Table 4.3: Measures for the dashboard</i>	22
<i>Table 6.1: Good data characteristics (Gualo et al., 2021)</i>	36

Table of Contents

Research Information	i
Preface.....	ii
Management Summary	iii
Lists of Figures	v
List of Tables	vii
Table of Contents	viii
Glossary List	x
Chapter 1: Introduction	1
1.1 Company Profile	1
1.2 Empty Miles of Shipment	1
1.2 Empty Miles of Shipment	2
1.3 Stakeholders	2
1.3.1 Emons Sales.....	3
1.3.2 Emons Account Management.....	3
1.3.3 Emons Expedition Department.....	3
1.4 Problem Identification.....	3
1.4.1 Problem Context	3
1.4.2 Problem Cluster	4
1.4.3 Core Problem.....	5
1.5 Research Objective.....	5
1.6 Research Questions	5
1.7 Research Methodology.....	6
1.7.1 Research Design	6
1.7.2 Problem Solving Approach	6
1.7.3 Research Validity and Reliability	8
Chapter 2: Context Analysis	9
2.1 Network Location Distribution	9
2.2 Frequency Matrix	10
2.3 Data Understanding.....	11
2.4 Conclusion: Research Scope	11
Chapter 3: Literatures Review	13
3.1 Decision Support Tool for Logistic Network.....	13
3.2 Machine Learning and Business Intelligence.....	14
3.2.1 Key-Influencer Visual: Self-Service Regression.....	14
3.2.2 Decomposition Tree Visual: Self Service Decision Tree	15
3.3 Data Visualization Techniques.....	16

3.4 Conclusion and Summary	17
Chapter 4: Dashboard Design and Development	18
4.1 Design Requirement Framework	18
4.2 Data Preparation	18
4.3 Dashboard Development	20
4.3.1 Dashboard Sections	20
4.3.2 Dashboard Features	21
Chapter 5: Implementation	27
5.1 Page Guideline	27
5.2 Scenario Analysis	31
5.2.1 Case	31
5.2.2 Problem Approach	32
5.2.3 Analysis	32
Chapter 6: Validation	36
6.1 Data Validation	36
6.2 Content Validation: Stakeholder input	37
6.2.1 KPI Measurement	37
6.2.2 Visualizations and Features	37
6.2.3 Final Prototypes	38
Chapter 7: Conclusion and Recommendation	39
7.1 Conclusion	39
7.2 Limitation	40
7.3 Recommendation	40
Reference	42
Appendix A: SLR summary	44
Appendix B: Stakeholder input	46

Glossary List

AI	: Artificial Intelligence
BI	: Business Intelligence
DAX	: Data Analysis Expression, formula expression language used in Analysis Services, Power BI, and Power Pivot in Excel
FastTree	: An efficient implementation of the MART gradient boosting algorithm
DSMR	: Design Science Methodology Research
Lane	: The shipment route that a truck travel from one location to another different location
L-BFGS	: A method used to solve large scale nonlinear optimization problems matrices are expensive to compute
MART	: MART (Multiple Additive Regression Trees) is an implementation of the gradient tree boosting methods for predictive data mining (regression and classification)
ML	: Machine Learning
ML.Net	: An open source, and cross platform machine learning framework for the .NET developer platform
Regression	: a statistical technique that relates a dependent variable to one or more independent (explanatory) variables
SDCA	: (Stochastic Dual Coordinate Ascent) method, a state-of-the-art optimization technique for convex objective functions.

Chapter 1: Introduction

This introductory chapter explains the problem background and the overview of how the research is conducted. The chapter begins with section 1.1 which introduces the company business, then followed by section 1.2 explaining the problem background. Section 1.3 describes the relevant stakeholders that need the expected solution implementation. Further, the problems are identified and clustered in section 1.4. The research objective and research questions are stated in sections 1.5 and 1.6 respectively. Lastly, the research methodology of how each step is conducted is explained in section 1.7.

1.1 Company Profile

Emons Group is a logistics company that focuses on innovative and sustainable transport solutions, their aim is to reduce overall logistics costs by providing innovative services. Emons 2WIN cargo is one of Emons Group's cargo brands. It is Europe's largest multi-customer double-deck transport network. Compared to a standard trailer, 2WIN cargo can transport over 64% more volume by offering a capacity of up to 54 euro pallets (*Double-Decker International Shipping Service*, 2022)

The Emons 2WIN cargo is operating in 12 country destinations within Europe and United Kingdom. In order to optimize international freight network, Emons connect individual shipment from different customers to create a transport network. Network optimization is done by developing and implementing freight concepts for the best product route in terms of costs, services, sustainability, and security.

1.2 Empty Miles of Shipment

In every shipment, truck travels from one loading location to an unloading location, each of these routes is called a lane. Every shipment has its lane which uniquely links with a customer. When combining lanes from different customers there is a distance gap where the truck's trailer has an empty load. It happens when the truck travels from one unloading location to the next loading location. This empty load distance is commonly known as empty miles of shipment in logistics. The empty miles turn into a problem as most customers do not pay for the trailer return trip. Although in practice, sometimes customers pay for empty load distance, however, more commonly this lane is unpaid.

In order to minimize the empty miles, Emons tries to find the additional shipment(s) near the empty lane which they can load onto the truck to get paid by that particular shipment distance. Thus, the empty load distance is reduced. However, as a consequence of taking additional pallets along the way, the initial shipment network will change due to some lanes being added, while others can be discontinued by a changing customer base, thus some other lanes are removed. Changes in one network shipment might cause some changes in other parts of the same network shipment. Currently, there is a lack of supporting tools to determine how changing lanes impacts the whole network, thus it becomes difficult to assess the network performance and see where in the network the opportunities for improvements are. Further, it also becomes challenging to understand how changing lanes in one network will impact the total network.

Essentially, the shipment route is a dynamic network, it keeps changing depending on demand and season. By that means the shipment route for this week could be completely different from the shipment route for next week, further the route shipment of the exact same week each month

is not necessarily the same. (i.e shipment route in the first week of February is not the same as the route in the first week of January). Thus it depends on the situation of that particular time. Then by having additional customers for minimizing the empty miles, the network dynamic becomes more challenging.

1.2 Empty Miles of Shipment

In every shipment, truck travels from one loading location to an unloading location, each of these routes is called a lane. Every shipment has its lane which uniquely links with a customer. When combining lanes from different customers there is a distance gap where the truck's trailer has an empty load. It happens when the truck travels from one unloading location to the next loading location. This empty load distance is commonly known as empty miles of shipment in logistics. The empty miles turn into a problem as most customers do not pay for the trailer return trip. Although in practice, sometimes customers pay for empty load distance, however, more commonly this lane is unpaid.

In order to minimize the empty miles, Emons tries to find the additional shipment(s) near the empty lane which they can load onto the truck to get paid by that particular shipment distance. Thus, the empty load distance is reduced. However, as a consequence of taking additional pallets along the way, the initial shipment network will change due to some lanes being added, while others can be discontinued by a changing customer base, thus some other lanes are removed. Changes in one network shipment might cause some changes in other parts of the same network shipment. Currently, there is a lack of supporting tools to determine how changing lanes impacts the whole network, thus it becomes difficult to assess the network performance and see where in the network the opportunities for improvements are. Further, it also becomes challenging to understand how changing lanes in one network will impact the total network.

Essentially, the shipment route is a dynamic network, it keeps changing depending on demand and season. By that means the shipment route for this week could be completely different from the shipment route for next week, further the route shipment of the exact same week each month is not necessarily the same. (i.e shipment route in the first week of February is not the same as the route in the first week of January). Thus it depends on the situation of that particular time. Then by having additional customers for minimizing the empty miles, the network dynamic becomes more challenging.

1.3 Stakeholders

In reducing the empty miles, several stakeholders involve in some decision-making regarding network optimization. As shown in Figure 1, the first inner layer outside the core implies Emon's selected options, the second inner layer implies the action that needs to be taken correlated to the layer under it and lastly, the most outer layer implies the stakeholder who is related with the action under it. This section will elaborate more on three stakeholders as indicated in Figure 1. Section 1.3.1 elaborates on the Sales role, section 1.3.2 elaborates Account Management role and section 1.3.3 elaborates expedition role.

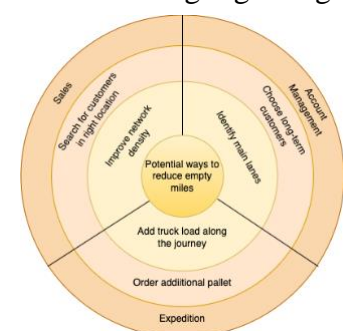


Figure 1. 1: Stakeholders onion chart

1.3.1 Emons Sales

This department is responsible for finding potential customers, refining, and signing contract agreements with customers. When searching for potential customers, the loading and unloading locations are one of the most important aspects sales need to take into consideration. It is due that customer location will impact the network. The better the network-fit, the less empty miles of the network. By having a better understanding of the network and the gaps of the network (the empty miles), sales would have a better insight into which region should be focused on for finding customers.

1.3.2 Emons Account Management

Account management is a role that focuses on maintaining customer relationships in order to retain their relationship with Emons. From all customers who have an agreement with Emons, this department needs to map which one they should have a long-term and short-term relationship with. For potential long-term customers, they aim to build a partnership, meanwhile the non-partnership customers have a contract agreement in which they have a shorter period of agreement. Customer categorization (how important a customer is) depends on various factors, such as the revenue it generates for Emons. In regards to the topic of our network, the customer can have an important contribution to the network. If a customer discontinues their cooperation with Emons, it will consequently impact the total network performance.

1.3.3 Emons Expedition Department

The role of the expedition department is to matters that are happening during the expedition. In regards to minimizing the empty load distance, expedition responsible in ordering additional pallets to be shipped to fill the distance of empty load. They look for the pallet via several marketplaces such as *Spot Market*.

1.4 Problem Identification

This chapter discuss the process of narrowing problems into research objective. It starts by identifying the problems in Section 2.1, mapping the problems into causal relationship diagram in Section 2.2, concluding the core problem in Section 2.3, and lastly defining the research objective in Section 2.4.

1.4.1 Problem Context

In order to understand the current situation, several interviews and discussions with Emons were conducted. The cause and effect of empty miles of shipment were investigated. Then after several meetings we conclude that there are four main problem contexts as describe below:

1. *Lack of insight in network density*

As Emons operates across Europe, there is a big transport network of shipment that connects various shipment locations. Thus, the network has a lot of dependencies across different lanes. Ideally, the network supposed to be so dense in such a way the empty miles is minimum. In practice, the empty miles is not possible to be reduced to zero. Therefore, some insight of network density is needed to see where in the network location that still need improvement of density level. However, at this moment there is no method to give insight of how different lanes contribute to the network density.

2. *Unnecessary cost*

When there are empty miles, the truck will travel unpaid for certain empty load distance. As a result, some unnecessary operational cost has to be incurred to keep the

truck travel to the next customer destination. Among the operational costs there are the truck maintenance cost, truck driver wages and fuel cost.

3. Unnecessary CO2 emission

When truck travel with empty load it implies that the truck journey does not have any significant contribution to logistic shipment. Hence, the truck travel is practically unnecessary in which release unnecessary CO2 emission.

4. Lack of decision support tool to find customers and additional pallet

Considering that each lane indicates shipment route of a customer, therefore lanes have linear correlation with customers from Sales perspective. In minimizing empty miles, selecting lanes become a crucial step to improve network density in location where the gap of empty miles is big. Thereupon, in the context of this project choosing lanes can directly be translated as choosing customers. As we know that currently there is a lack of insight in network density, as consequences properly finding customers in the accurate location to reduce empty miles become quite difficult. Moreover, Expedition department also need to know similar information of where in the network the gap is high and additional lanes are needed. Therefore, Sales and Expedition need a tool that would help them to select customers based on network insight.

1.4.2 Problem Cluster

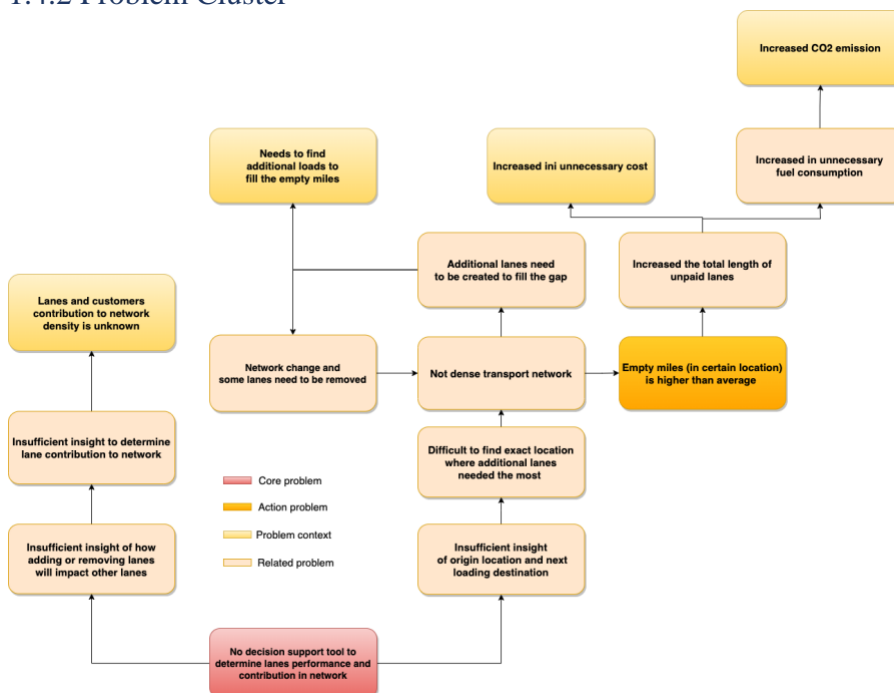


Figure 1. 2: Problem cluster

The identified problem contexts in which colored yellow are mapped into a causal relationship as in Figure 2. Several other causes which link the identified problems are addressed as related problems. The problem context which has been elaborated in Section 2.1 is mainly related to shipment lane, cost, gas emission, and customers. In order to draw a relationship between them, we step back into the process of these problems being connected. After analyzing the causal relationship diagram the core problem and action problem are determined.

The identified action problem is there is the chance of high miles of shipment in certain region compare to the overall region density if the current condition is not improved. As has been

stated in Section 2.1 the empty miles issue is an optimization problem that is not possible to be reduced until zero. In practice, logistic companies try to minimize the empty miles length to reduce unnecessary costs and resource utilization. Therefore, the level of empty miles could sometimes fluctuate, however, the goal is to always keep it minimum. Currently, the company does not specifically calculate the empty miles for different regions, thus the current level of empty miles in different locations are still unknown. For that reason, the norm and reality of the action problem are not quantitatively defined. Based on problem cluster analysis the gap of action problem could be solved by focusing on the core problem that there is no decision support tool to determine the importance of lane.

1.4.3 Core Problem

After mapping the identified problems into a causal relationship diagram in Figure 1.2 and assessing the cause and effect of each relation, we conclude that *no decision support tool to determine the importance of lane as performance and contribution in network* as the core problem. The lack of decision support tools mainly being a challenge for the relevant stakeholder in optimizing the network and finding the most suitable customers. As explained in section 1.3, the Sales department which consists of Sales and Account Management role a challenge to properly select and refine customers to optimize the network. Moreover, the Expedition department also needs some insight on where to order additional pallets in a certain region. Thus the tool will be mainly developed for the Sales department and Expedition department.

1.5 Research Objective

The identified core problem revealed that there is a need of having a decision support tool in which can provide some insight of network density performance for Sales and Expedition departments to use. Thus, it leads to research objective to develop a logistic network decision support tool by making use of shipment historical data. Through this research, we aim to discover a method to build the expected tool while at the same time will also directly apply the method into a product. The result of the discussion with Emons revealed that the expected tool has desired function as follows:

1. Interactive data insight tool with filter feature to slice and dice information
2. Provide lanes origin and destination statistic insight
3. Provide scenario analysis for lanes changing

These criteria become the foundation of how the expected tool should work. At the end of the research, it is expected that an applicable logistic network decision support tool for minimizing empty miles of shipping network and a systematic approach how to the tool is developed will be delivered as the main result.

1.6 Research Questions

In order to achieve the research objective, we formulate the main research question. This research question is then narrowed into more specific sub-questions related to the topic.

How can a decision support tool analyses and provide data insight for helping Emons Sales departments to minimize empty miles of the overall logistic network?

1. What are existing techniques that can be implemented to fulfill the required functions for the decision support tool?
2. What are the data analysis techniques that suit the tool implementation?

- What are the main components of the tool and how will they be implemented using the collected data?

1.7 Research Methodology

This section mainly discusses the methodology on how the research will be conducted, what result as expected, as well as the limitation of this research. Section 1.7.1 explain the selected research design, Section 1.7.2 elaborate the problem approach of the research, and Section 1.7.3 discuss the research limitation from the perspective of validity and reliability.

1.7.1 Research Design

In order to systematically solve the occurring problem, the Design Science Research Methodology (DSRM) (Peppers et al., 2007) problem solving approach is chosen as the research methodology. DSRM framework is a problem-solving approach for doing research that seeks to enhance the creation of innovative artifacts (Brocke et al., 2020). The goal of a Design Science Research aims to extend the limitation of human and organizational capabilities by creating new and innovative artifacts represented by constructs, models, methods, and instantiations (Gregor & Hevner, 2013; Hevner et al., 2004). Further, in this research we aim to develop a tool for the creation of an artifact, thus the DSRM approach is suitable to be used for this research method. The DSRM process model has six phases that include: problem identification and motivation, defining the objectives for a solution, design and development, demonstration, evaluation, and communication.

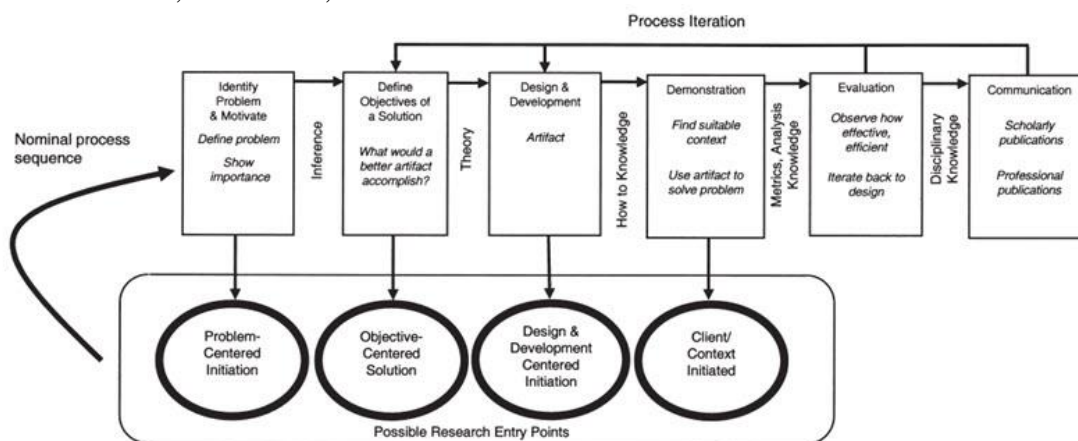


Figure 1. 3: DSRM process model

1.7.2 Problem Solving Approach

In conducting the research, the six methodology phases will be executed to answer the formulated research questions. Further, during the research more knowledge questions will come up along the way as direction on what has to be achieved in each phase. In this section, the problem approach is elaborated and linked to each research question and knowledge questions.

Phase 1: Identify the problem

The goal of the first phase is to identify the core problem. In this process we gather information about the current situation by answering several knowledge questions. The information then later be mapped in a problem cluster diagram to find the core problem. To gain better understanding of the situation the following questions were formulated:

- How Emons currently reduce the empty miles?

2. Who are the relevant stakeholders to the expected tool and what are their anticipated benefit of implementing the solution?

Phase 2: Define Objective solution

After finding the core problem the objective of the research is formulated. In this research, the objective is formulated by creating main research question as research goal and several knowledge questions to fulfill to support the main research question. As the goal of this research is to develop a tool, then in this phase the requirements of the tool are defined. Then the following knowledge questions were formulated:

3. What is the desired situation?
 - a) What are the main data insights that the tool should produce?
 - b) What are the inputs of the tool?
 - c) What are the expected outputs?
 - d) What are the requirements and constraints of the tool?

Phase 3: Design & Development

In this phase the design and development of the tool will be working. There are several steps which should be done in the process of creating the tool. Firstly, we have to choose the platform or environment where the tool will be developed, thus some tool comparison and selectin will be conducted. Second, we need to create a framework in which shows the tool component combined in one platform thus the design of the tool clearly defined before being developed. Lastly, we will develop the tool using shipment network historical data. Thus, during this phase the following knowledge questions are formulated.

- a) What are existing techniques that can be used to develop the required functions for logistic network decision support tool?
- b) What data analysis techniques that is suitable for the tool implementation?
- c) What are the main components of the tool and how will they be implemented using the collected data?

Phase 4: Tool Demonstration

When the tool development has been completed, we will test the performance by demonstrate the tool. One of the ways to demonstrate it is by doing scenario analysis with several case scenario. Thus, the following knowledge questions are formulated.

- a) How the tool will be used by the relevant stakeholder?
- b) What are possible scenarios that are need to be tested?

Phase 5: Evaluation and Recommendation

Before finalizing the research, we will reflect on the whole research process to see whether the main research question has been answered. Moreover, we will see how the implementation of the expected solution solve the identified problem. The evaluation of tool demonstration will be summarized as final recommendation for further research and future improvement.

Phase 6: Communication

The overall result and summary of research are documented in a form of written report. The report and tool will be presented to Emons as problem owner. Further, the whole research process be defended at the university through colloquium. Finally, the content of the report is structured as based on the knowledge question as follow:

Purpose	Knowledge question	
Identify problem	How Emons currently reduce the empty miles?	Chapter 1
	Who are the relevant stakeholders to the expected tool and what are their anticipated benefits of implementing the solution?	Chapter 1
Define Objective	What are the main data insights that the tool should produce?	Chapter 2
	What are the inputs of the tool?	Chapter 2
	What are the expected outputs?	Chapter 2
	What are the requirements and constraints of the tool?	Chapter 2
Systematic Literature Review	What are existing techniques that can be used to develop the required functions for logistic network decision support tool?	Chapter 3
	What data analysis techniques that is suitable for the tool implementation?	Chapter 3
	What are the main components of the tool and how will they be implemented using the collected data?	Chapter 3
Tool Demonstration	How the tool will be used by the relevant stakeholder?	Chapter 5
	What are possible scenarios that need to be tested?	Chapter 5

Table 1. 1: Research question

1.7.3 Research Validity and Reliability

Validity in this research context refers to the extent how which the data insight precisely explains the real situation of empty miles. Shipment network in practice is a complex dynamic network. As this research has time and capability constraints, the data and situation that will be researched have been simplified from the real situation of the network. Among the simplifications is the shipment time unit. In real life, the shipment time is counted in precise hours and minutes, however, to simplify the situation, the period will be counted in the day, week, month, or year depending on the selected data range. Moreover, in this research we assume that the truckload is always full, thus the demand is linear with the number of the truck unit. Apart from that, we need to take into account that the term “empty miles” is used based on the most common term in literature research, however, for the collected data in this research the di distance unit that will be used is KM.

Meanwhile, reliability in this research context refers to the extent of how consistent the process of measuring the data and how good the quality of data is being used. To assess the reliability of this research, a validation phase with the stakeholder who works with the related data in a real situation will be conducted. Further, a data quality assessment will be conducted to preserve the reliability of the research result.

Chapter 2: Context Analysis

This chapter provides a further in-depth explanation of the empty miles problem. In Section 2.1, a detailed explanation of the network is provided first. Following that, Section 2.2 explains the current approach to reducing empty mileage. Finally, it leads to how the present technique may be improved for additional research in Section 2.3

2.1 Network Location Distribution

The shipment loading and unloading location are spread across 14 countries in Europe. As shown in Figure 4, each country has different number of cities. Each of the locations is connected by a shipment lane based on the loading and unloading location. Some regions have high density while some others have low density. Currently, there is no tool to calculate the density of the network, however, the density is approximated by looking at the crowdedness on the map. Since the density level for different regions are different, as consequence driving to a region with low density is more likely to yield high empty miles compare to driving to a high-density region. This is due to far distance trucks need to travel between the unloading location

to the next loading location. Ideally, to have an optimum network with minimum empty miles, the network density should be equally spread out within the region. Further, it would be more ideal if the gap between loading and unloading locations in a region is minimized. Intuitively when the number of loading and unloading location in a region is closely equal, then the truck will have more near options of where to travel next after finishing unloading customer's good, as a result, the empty miles is minimized. This approached is currently used by company to balance the network. However, in reality that condition will not always be applicable in practice. It is due to the time constraints of when customers order the shipment service varies, thus time range needs to be take into consideration for this approach.



Figure 2. 1: Network locations distribution

This network density issue can be approached from different level of plannings. Firstly, in the strategic level company could focus on long term planning of how they should re-design the network shipment flow and construct region limit of where they operate. Furthermore, for the tactical level the problem could be seen from a medium level, this could be perceived by how the company decides which area they want to improve then try to manage the customers balance for that particular region. Finally, the lowest level of planning is operational level. In operational level the problem focus on how to schedule the shipment and find additional customers in daily basis. For this research, company aims to tackle the problem from the tactical point of view on how to find region where additional customers are needed to balance the shipment network.

2.2 Frequency Matrix

In finding additional customers to fill the empty miles the company looks at locations that need more additional customers. In determining the location and the number of additional shipment needed for the network, the company uses a frequency matrix that measures the shipment balance in each country. The shipment balance is defined by an equal number of loading and unloading frequencies in a region, in Table 1 shipment balance is indicated by 0 value.

As shown in Table 2.1, for each country they subtract the number of unloading by the number of loading. The more deviate the number from 0, the worse the shipment balance in that region is, whether the number is positive or negative. When the number is more positive it means more loading frequencies are needed, and meanwhile if the number is negative then more unloading frequencies are needed. However, it is harder to find additional unloading locations as it is dependent on the loading location of where the shipment starts.

Country	Loadings	Unloadings	Unloading - Loading
A	14	14	0
B	0	0	0
C	23	3	-20
D	0	4	4
E	6	7	1
F	111	62	-49
G	11	17	6
H	56	106	50
I	3	4	1
J	3	1	-2
K	5	5	0
L	37	46	9
M	2	2	0
N	0	0	0
Total	271	271	0

Table 2. 1: Weekly calculation

The calculation is gathered and then merged into a matrix with a larger time range as shown in Table 2.2 Each row in the matrix represents the shipment balance for each country, meanwhile, the column indicates the shipment balance in a specific week. As shown in the table that the number fluctuates every week, it could be positive or negative. Thus, the negative and positive numbers from different weeks will balance one another when seeing the shipment in a longer time range when accumulating the total balance in the Total column. However, a high deviation from 0 still happens in 60% of the countries. As shown in table 2.2 that the upper and lower rows are hugely deviating from 0, this indicates that those upper countries have more unloading locations, meanwhile, the rows with negative values indicate the lack of unloading locations in the corresponding countries. The table calculates data based on weekly data collection, therefore the total number would change on a weekly basis.

Country	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	Total
A	48	75	56	60	63	62	35	58	48	48	50	8	4	0	0	0	615
B	5	1	11	2	3	2	2	1	4	9	6	4	-2	-2	0	0	46
C	9	-4	3	-8	-2	-9	8	4	7	4	9	9	0	0	0	0	30
D	3	3	1	4	-2	4	3	2	2	1	1	-1	0	0	0	0	21
E	2	3	0	3	2	2	3	0	-3	1	4	0	0	1	0	0	18
F	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	-4	-1	-1	-4	-3	-2	-3	-2	-6	-2	-2	3	0	0	0	0	-27
K	-10	-3	-5	-3	-1	-4	-4	-2	-4	-3	0	-2	-3	-1	0	0	-45
L	-9	-11	-10	-1	-12	-10	-5	-7	-6	-2	1	-1	0	0	0	0	-73
M	-19	-21	-18	-14	-17	-13	-17	-20	-15	-19	-20	-13	1	2	0	0	-203
N	-25	-42	-37	-39	-31	-32	-23	-34	-27	-37	-49	-7	0	0	0	0	-383
Total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2. 2:Aggregated weekly calculation of shipment frequency

The matrix shows whether there is a balance in the number of loading and unloading locations in each country. However, the insight is still limited in making a decision to optimize the network density. The limitation is listed as follow:

- It does not quantify the empty miles
- It does not consider the shipment distance
- It does not provide the specific location (city)
- The time frame is fixed and hard to be manipulated for further analysis

2.3 Data Understanding

The data set chosen for the dashboard is based on historical shipment data for the period of 2019 to 2021. Each row of the data set contains information about a single shipment, and each shipment has three lanes that are interconnected. In this paper, the lanes are referred to as "Previous Empty KM Lanes," "Full KM Lane," and "Next Empty Lane". Each lane is connecting a loading location as illustrated in Figure 4.1, each row contains two loading locations and two unloading locations, which are connected by three lanes. The lanes in which the trucks are paid are called "Full KM Lane", meanwhile on the other two lanes, the trucks are not paid. Each lane consists of several fields that indicate detailed information about the loading and unloading location. In addition, for each shipment, the shipment date and customer names are also provided on the same rows. In order to have a better overview of the data field in the shipment detail, below we provide a table summary of the data field based on data types and its correlation with location and lanes categories.

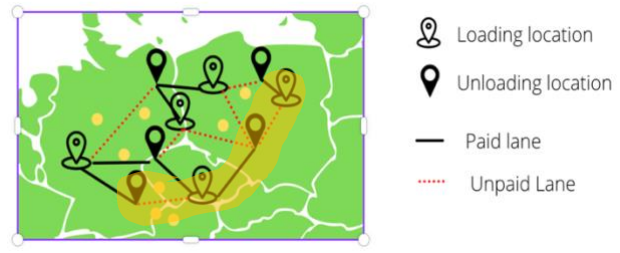


Figure 2. 2:Shipping network illustration; the yellow shaded region indicates data within one row in data set

Field Name	Location Category		Lane Category		
	Loading Location	Unloading Location	Previous Empty KM Lane	Full KM Lane	Next Empty KM Lane
ShipmentID	X	X	X	X	X
CustomerName	X	X	X	X	X
CustomerID	X	X	X	X	X
ShipmentDate	X	X	X	X	X
ShipmentLane	X	X	X	X	X
Distance	X	X		X	
DistanceFromPrevious			X		
DistanceToNext					X
FirstAddressID	X		X	X	
FirstCountry	X		X	X	
FirstZipcode	X		X	X	
FirstCity	X		X	X	
LastAddressID		X		X	X
LastCountry		X		X	X
LastZipcode		X		X	X
LastCity		X		X	X
PreviousAddressID		X	X		
PreviousCountry		X	X		
PreviousZipcode		X	X		
PreviousCity		X	X		
NextAddressID	X				X
NextCountry	X				X
NextZipcode	X				X
NextCity	X				X

Table 2. 3:Dataset attributes connections with location and lanes

2.4 Conclusion: Research Scope

As explained in Section 2.1 the shipment network is complex and dynamic. Therefore an interactive data insight is needed to provide deeper understanding of how the network perform. This research will focus on improving the limitation found in the current method as explained

in Section 2.2 by developing data visualization tool. Further, as adding or removing a lane will directly impact other lanes in which connected to it, then we should know which lane is worth removed or added. Then an analysis method should also be added to the tool.

Chapter 3: Literatures Review

This chapter examines existing research papers to answer several knowledge questions as listed in section 1.7. Firstly, section 3.1 discusses several related research on decision support tool of logistic network in order to find techniques that could potentially be applied for this research. Secondly, section 3.2 elaborate the selected analyses approach. Thirdly, section 3.3 examine various visualization techniques that have been used in transport network visualization. Finally, section 3.4 summarize the chapter by concluding the selected techniques and tool for this research.

3.1 Decision Support Tool for Logistic Network

The examination of the literature reveals several uses for decision support tools for networks and logistical transport. The majority of literature discusses how specific techniques are utilized as a tool to optimize the transport network to increase effectiveness and reduce cost. The network optimization is developed using a variety of approaches, such as improving energy efficiency (Farahani et al., 2018), allocating transportation resources more efficiently (Alam & Habib, 2021), or enhancing agent coordination and interaction inside a given port (Irannezhad et al., 2020). Moreover, the development of decision support tools to assess the risk of accidents in transportation networks is also covered in other publications (Balster & Friedrich, 2019; Loza-Hernandez & Gendreau, 2020). Finally, Gonçalves et al. developed a technique to prioritize close-proximity intervention in railways (2022).

However, literature that focuses on the issue of minimizing empty truck and empty miles in logistics are still limited, especially in the context of locating specific region for network improvement. Some papers discuss the issue of the empty truck from the perspective of truck capacity and load (Epstein et al., 2012; Kuzmicz & Pesch, 2019). Meanwhile, some others discuss empty trucks' problems with truck collaboration approach. There are two alternative ways to collaborate. The first is to use mathematics to develop a truck appointment system that would plan truck activities and enable trucks to cooperate for transportation between terminals and clients. (Schulte et al., 2017) and secondly by using simulation to configure regional driving fleets to minimize empty miles (Don Taylor et al., 2006). The two discussed approaches of changing truck capacity and creating truck collaboration are outside this research scope, thus they are not able to be applied to solve the action problem of this research.

The closest literature found on the topic of finding a specific location in transport networks are presented in the context of locating empty container storage to reduce the mileage (Lei & Church, 2011) and the selection of ramp service area to minimize the circuitous miles (Taylor et al., 2002). Despite the potential of the two studies, neither of them will fit to this research. First, it is anticipated that the techniques would produce more proposed strategies than decision support tools. Additionally, the literatures discussed and the Emons "empty miles" issue are different. The literatures attempt to identify a single unique location to construct a store using a mathematical model and simulation method; as a consequence, there is a single conclusion. The strategies mentioned in the literature reviews are not appropriate for our problem since the goal of our research is to examine and comprehend potential alternative locations for enhancements. As a result, the mathematical model method is inappropriate for solving our problem. In order to bridge the gap in the logistical network decision support tool to locate empty miles region, this research was conducted. For summary, the method classification of the literatures mentioned in this section are attached in the Appendix A.

3.2 Machine Learning and Business Intelligence

Data-driven research has been revitalized by the growth of big data analytics for freight transportation and the developments in machine learning. Therefore, for creating a decision support tool with numerous amounts of data, machine learning method is one of possible approach to be used. Tsolaki et. al (2022) grouped data mining and machine learning methods according to the application areas. One of the application areas is the problem of predicting traffic flow and location, it is the transportation research branch that is most closely related to network performance analysis. According to the literature, this problem can be solved using a variety of machine learning techniques, including regression trees, decision trees, patterns, artificial neural networks, and other techniques (Tsolaki et al., 2022).

Implementing machine learning techniques frequently requires a thorough understanding of mathematics and computer science; nevertheless, because of the high degree of expertise required, ML is difficult to use. Modern research has focused on creating ML frameworks that enable self-service adoption of ML components rather than trying to provide a unified ML solution for every user and requirement. A self-service ML framework can therefore be utilized by analysts with less technical knowledge (Patriarca et al., 2022). Several research have provided self-service machine learning deployment for business intelligence using Microsoft Power BI (Patriarca et al., 2022; Sufi, 2021, 2022).

Microsoft Power BI is an interactive data analytics software that enables users to explore business intelligence through advanced analytics and interactive data visualization. It provides variety of features and connectivity with several databases systems and other Microsoft products, thus it enables users to innovate and customize analytics dashboard based on their needs. AI-Visulaization is one of Power BI's distinctive features. Key-Influencer visualization, Decomposition Tree, and Anomaly Detection are the three different types of AI-Visualization in Power BI. Each visualization is created by an ML.NET-based machine learning algorithm. For the expected solution of this research, we will use Key Influencer and Decomposition Trees among the three AI-Visualizations. The next paragraphs provide a brief explanation of their working concept.

3.2.1 Key-Influencer Visual: Self-Service Regression

The purpose of key influencers visualization is to make it easier to identify certain elements that affect a metric. The user is given the opportunity to dive further by finding correlations within the dataset. Prior to displaying them as the most crucial elements in the visualization, it first analyses the data, classified the key variables, and then crucial variables based on its importance or contribution. This visualization is produced by ML.NET using two different types of regression techniques depending on the variable type, which are namely linear and logistic regressions.

Sufi (2021) described that linear regression was carried out for data of the numerical type using Microsoft's ML.Net's SDCA regression solution. In linear regression, numerical data are plotted using a mathematical model, and then the data points are compared to a straight line. The algorithm finds points with the best-fit line and uses it to construct an equation that depicts the dependent and independent variables as being in a linear relationship. The mathematical model use equation as follow:

$$y = b_0 + b_1x_1 + \varepsilon \quad \dots (1)$$

y = dependent variable; called response x_1 = independent variable; called predictor ε = random error

b_1 = slope indicating the change of y b_0 = intercept

Further, categorical data's logistic regression was carried out using ML.Net's L-BFGS logistic regression. The kind of data that can be grouped together is known as categorical data. Using supervised learning methods, logistic regression seeks patterns and assesses correlation between groups of data (Sufi, 2021). The following statistical equation is used to model this regression:

$$\text{Log} \left[\frac{P}{1-P} \right] = b_0 + b_1x_1 + b_2x_2 + \dots b_nx_n \quad \dots (2)$$

P = probability of event Y occur

x_1, x_2, \dots, x_n = independent variable; called predictor

b_1, b_2, \dots, b_n = slope indicating the change of y b_0 = intercept

In both logistics and linear regression, the strong relationship between variables are determined by coefficient correlation. This coefficient lies in range $[-1,1]$, in other word the value will not exceed 1 or be less then -1. When the correlation is greater than 0 then it is called positive relationship in which indicate strong relationship. In contrast, if the coefficient is less than 0 then it is called negative relationship in which indicate weak relationship. The more positive coefficient, the better the quality of the relationship. The downside of the key influencer visual feature is that the coefficient correlation of the data are not displayed, Therefore the feature does not explain what benchmark of coefficient correlation they use to consider that variables have enough strong relation. However, according to Microsoft (2022), they employ the Wald test to decide if a variable is an influencer. The threshold is established by the visual using a 0.05 p-value.

3.2.2 Decomposition Tree Visual: Self Service Decision Tree

The decision tree concept is used by the decomposition tree visual to show data along several dimensions. This tool is helpful for data exploration and doing root cause analyses. In the visualization, data is gathered in the graphic, which also allows users to dig down into data dimensions in any sequence. In order to deliver the most representative segmentation, it looks for data that is relatively high in a certain statistic of interest.

This decision tree procedure employs FastTree, it is an efficient implementation of the MART gradient boosting. It learns step-wise an ensemble of multiple regression trees (Patriarca et al., 2022). The MART gradient boosting algorithm uses statistical regression and classification as its primary methods for predictive data mining. The idea behind the gradient booting procedure is to train a system using a mathematical model to maximize a differentiable loss function from an array of random "input" and "output." (Friedman, 2001). By generating a regression tree that roughly approximates the gradient of the loss function in each step, the ensemble is produced. It then incorporates it into the tree that was previously computed using the factors designed to reduce the loss of the new tree. The decision tree algorithm, where it has x as input, can be written as follows:

$$h_m(x) = \sum_{j=1}^{I_m} a_{jm} 1_{R_{jm}}(x) \quad \dots (3)$$

$h_m(x)$ = the decision tree

I_m = the number of tree's leaves in decision trees;

R_{jm} = a region of the input space partitioned in I_m distinct regions ($R_{1m} \dots R_{I_m m}$)

a_{jm} = predicted value in a certain region R_{jm} .

The way the equation works is that each predicted value (a_{jm}) in the equation is multiplied by a certain number of coefficients in order to reduce the loss function and update the model as necessary. In reality, the algorithm groups data segments that share a numerosity that cannot be attributed to a splitting anomaly after each split (Patriarca et al., 2022)

3.3 Data Visualization Techniques

There are a number of research papers that examine various transport network visualization techniques; each paper presents different visualization forms depending on the context of the data and the accessibility of feature options on the platform being employed. Table 3.1 provides a summary of the visualization identified in the literature review.

No .	Author	Visualization Techniques							
		Route Map	Bar Chart	Line Graph	Bubble Map	Network Navigator	Flow Map	Heat Map	Simulation
1.	(Bachechi et al., 2022)	x	x	x					
2.	(Pan et al., 2019)				x	x			
3.	(Pais Montes et al., 2012)					x	x		
4.	(Kalamaras et al., 2018)	x						x	
5.	(Čerić, 1997)								x

Table 3. 1: Visualization techniques literature review summary

Bachechi et al. (2022) give visualization for traffic monitoring in a variety of methods, mainly to provide visualization of the location and data insight summary. For location visualization in particular, they begin by using a route map to highlight a particular traffic condition in a given area. In addition, bar charts presented to display the quantity of vehicles and line graphs to display time interval as part of data insight summary. Further, the other literatures more focus on showing the visualization of the network by emphasizing the location point of view. Kalamars et al.(2018) created interactive visual analytics platform for transport system. In addition, using route maps as visualization tools, the platform also features colored dots scattered across the route, where the color spectrum demonstrates how the vehicles behave in terms of traffic flow. A heat map is also included to display the number of accidents on the map and is linked to the route map's transportation behavior. Light blue to red are the colors of the heat map, with red spots denoting places with a high accident rate. For the light blue tinted road, it is the opposite.

Moreover, network navigator is used in by Pan at al. (2019) to display hub connections between cities in China. The connections presented as lines and the cities presented as circle. The size of the circle indicates the number of connections associated to the city, thus the bigger the

circle, the more connection the city has. However, the network navigator only shows connection of cities without presenting the real location of each city on map. Thus, to compensate the downside of network navigator, they also use bubble map to show the city connectivity density on the map. Bubble map shows the location of cities on map while at the same time also shows the city connectivity density with circle at each location. However, the different between bubble map and network navigator is that bubble map does not show the line of connection between cities. Similar, Pain Montes et al. (2012) also displayed the containership route network emerging using network navigator, however in this instance they utilize Flow Map to show a better picture of the location map. Although the notion of a flow map and a network navigator is similar, in flow map the line thickness places greater emphasis on the number of connections. Additionally, the place is depicted on the map as being in its actual position.

The final literature featured visual interactive modelling and simulation in a network of railroads for logistical transit (Čerić, 1997). They describe how to create simulations with ServiceModel software. The simulation demonstrates interactive visualization by fusing many graphical user interface types. Although the simulation concept is interactive, a new visualization must be created. In contrast to map visualization, which is typically easily used in BI software like Power BI, visualization in simulation software is typically not easily developed in the software. Thus, various network components such as sections, sidings, stations, and various train kinds were constructed via the graphical user interface in order to simulate the railway network.

After examining the visualization techniques that were described, we eliminated a number of methods that were deemed to be excessively sophisticated and less relevant for our study. The route map firstly displays a detailed image of the location and road; however, as Emons' logistical shipments are typically made between cities, having a detailed representation of the road is not always necessary. Therefore, the solution implementation does not require a route map. Second, only the regions with a high number of connections would pop out on the bubble map; this method would conceal the regions with a low number of connections. Thirdly, the precise location on the map was not displayed in the network navigator. If we want to identify possible areas for improvement, then having access to the precise depiction on the map is a key requirement. Last but not least, building an interactive visualization using simulation is a challenging process. Using a BI tool is a more effective option because Power BI currently offers a variety of interactive data visualizations.

3.4 Conclusion and Summary

From the examination of literature review, we did not find an existing logistics network decision support tool methods that can be further developed and used for the empty miles issue of Emons company. It is primarily due to the way network decision support tool in the literatures designed are not suitable with the current need of the company. Finally, self-service machine learning for business intelligence visualization is selected as the primary feature for the decision support tool. In addition, several visualization techniques has been reviewed and selected to be deployed to the tool. The selected techniques are bar chart, line graph, flow map, and heat map. Combining the selected techniques, it is concluded that the decision support tool will be created as dashboard using Microsoft Power BI.

Chapter 4: Dashboard Design and Development

The aim of this chapter is to display the process of creating data insight tool. The structure of this chapter follows the data transformation framework (Bascur & O'Rourke, 2020). First section 4.1 present the Design Framework of the dashboard. Afterward, data preparation process is elaborated in section 4.2. Finally, section 4.3 show the classification process of dividing selected insight into different section.

4.1 Design Requirement Framework

Analysis has been done to create the design requirement framework of the decision support tool in Table 4.2. This framework was created based on the objective in Chapter 1, evaluation of the current used method in Chapter 2, and examination of existing studies in Chapter 3 to select suitable techniques for the expected dashboard implementation.

Purpose	User	Functional and Feature	Content	Support in decision making
Overall insight of network	Sales Department	Heat map visualize network density	Loading and unloading location data field	Understanding which regions that already has big amount of customers and which region which need more
Empty Miles Measurement		Empty miles measurement displays in a card and calculate in measure feature using DAX	Calculation of previous and next empty miles. Calculated by full and empty distance.	Knowing how good the performance of overall network
Frequency trend over certain period of time		Bar chart in which can be adjusted based on desired period of time	Shipment frequency calculated by counting shipment number during certain period of time	Analyse which period of time has low or high shipment frequency, so they can track when is the best time to add more shipments
Rank the best and worst customers		Bar chart in which shows the gap of certain parameter for each customer	Calculate the empty miles per customer	Knowing which customers has high contribution to empty mile of network indicates that they are not customers that is good to be considered for long-term contract
Root Cause Analysis		Decomposition trees for root cause analysis	Empty miles measure	Understanding how each region contributes to empty KM
City Connection Analysis		Drill through to a page of lane analysis	Visualize the network lane, location density, lane frequency and distance, shipment frequency chart over certain period of time	Finding and determining which location and lane need to be prioritize for improvements
Time Slicer		Timeline slicer that the date based on month, quartile and years	Shipment date data field	Slice the data based on desired time period
Location Slicer		Normal slicer to select specific country and lanes	First Country, Last Country, First City and Last Cty data fields	Slice the data based on desired region
Customer		Normal slicer to select specific customer	CustomerName	Slice the data based on specific customer

Table 4. 1:Dashboard requirement framework

4.2 Data Preparation

Before implementing the data into the dashboard, it is important to prepare the data and make sure the dataset is valid and has good quality. To validate the data we need to make sure the dataset that will be implemented in the dashboard is sufficiently accurate (Beck, 2008). Furthermore, it also has to be justified that the dataset is consistent and complete both of which

indicate good quality of the data (Gualo et al., 2021). The data preparation steps are elaborated in the following paragraphs.

The original dataset comprises 32 columns and 51561 rows which consist of two categories of shipments, which are single and group shipments. Since we only need to analyse the behavior of individual lanes, we excluded the group shipment data for this research in which leaves 48587 rows in total. In addition, the data column was reduced to 25 columns after other redundant columns were also deleted.

The entities of each lane are highly dependable to the entities of other lanes on the same row. Therefore, when there is an error in one entity of a lane, it will directly impact the inaccuracy of other lanes that are connected to it. As elaborated in section 2.3 the lanes are connecting several different locations. Therefore, when there is one missing distance in a lane, we are not able to remove the entire lanes, as in practice some existing lane only has either the previous empty lane or the next empty lane thus some missing distance is considered valid. Further, as a truck could travel on the same lanes more than once, duplicate lanes are also considered normal thus, there is no need to remove lane or date duplicates as long as the shipment ID is unique.

When the data are being collected, there are possibilities of misspelling of location. In order to validate the data accuracy, further data check was done using Python. Python has a library called geopy that enables to translate a location into latitude, longitude, and zip code. We used this library to collect the longitude, latitude, and zip code of each location using the code shown in Figure 4.1. During the data gathering process, it was found that several locations were not able to retrieve latitude and longitude due to misplaced location. From 1924 distinct locations it was found that there were 143 misspelled locations. The dataset then being cleaned by manually adjusting the wrong location name with the right spelling.

```
#importing goopy library
from geopy.geocoders import Nominatim

# calling the Nominatin tool
GeoLocator = Nominatim(user_agent="Locator")

def latitude(Location):
    Loc=GeoLocator.geocode(Location)
    if Loc == "None":
        return "None"
    else:
        return Loc.latitude
def longitude(Location):
    Loc=GeoLocator.geocode(Location)
    if Loc == "None":
        return "None"
    else:
        return Loc.longitude
def zipcode(Location):
    Loc=GeoLocator.geocode(Location)
    address=Loc.address.split(",")
    zipcode=address
    return zipcode

dataset['Distinct Location Zipcode']=" "
dataset['Distinct Location Latitude']=" "
dataset['Distinct Location Longitude']=" "
for i in range(0,1923):
    dataset['Distinct Location Latitude'][i]=latitude(dataset['Distinct Location'][i])
    dataset['Distinct Location Longitude'][i]=longitude(dataset['Distinct Location'][i])
    dataset['Distinct Location Zipcode'][i]= zipcode(dataset['Distinct Location'][i])
    print(i)
```

Figure 4. 1: Python code to validate the location attributes

Furthermore, often time same names of cities could be located in several different countries. Therefore, in order to display location on the map location, the city, zip code, and country attributes are merged to provide accurate locations. Thus, new attributes called “First Location”, “Last Location”, “Previous Location” and “Next Location” are created. The dataset now has been prepared and loaded to Power BI. Before inputting the data into visualization, the data are classified based on its data type as shown is table 4.2.

Field Name	Data Type			
	Number	String	Date	Location
ShipmentID	X			
CustomerName		X		
CustomerID	X			
ShipmentDate			X	
ShipmentLane	X			
Distance	X			
DistanceFromPrevious	X			
DistanceToNext	X			
FirstAddressID	X			
FirstCountry				X
FirstZipcode				X
FirstCity				X
First Location				X
LastAddressID	X			
LastCountry				X
LastZipcode				X
LastCity				X
Last Location				X
PreviousAddressID	X			
PreviousCountry				X
PreviousZipcode				X
PreviousCity				X
Previous Location				X
NextAddressID	X			
NextCountry				X
NextZipcode				X
NextCity				X
Next Location				X

Table 4. 2: Attributes data type

4.3 Dashboard Development

Shipment network has numerous amounts of data insight that can be explored. In order to understand the network thoroughly, there is a process to dive the network from a wide point of view into a narrow point of view. Power BI enables the user to create multiple data visualization pages which can be connected, thus using the requirement framework in section 4.1 the dashboard is designed into several different sections for different purposes. The order of the section follows an order of providing general insight into more detail insight. The dashboard sections and features are explained in the following sub-sections.

4.3.1 Dashboard Sections

1. Network Overview

The first section of the dashboard provides an insight of how the network perform as a whole. In analyzing empty miles for general network performance there are several information the user needs to discover, mainly to see what possible action that can be taken to improve the network. Thus, this section should be able to answer the following questions:

1. How dense does the network look?
2. How good does the network perform?
3. What is the trend of empty miles over a period of time?
4. Which customers that highly contribute to empty miles?

2. Key Influencer

Empty miles are caused by various factors. In the shipping network, there are many components involved, thus it is quite complex to find which components that highly influence the empty miles. After understanding the general performance in section one, we would like to know the key factors that cause empty miles using Key Influencer visualization. In this visualization, we will use empty miles as the main parameter that is being analysed.

3. Root and Cause Analysis

This section aims to investigate why certain regions have high empty miles. This analysis is conducted in two steps. First, the decomposition tree is used to see the empty miles distribution based on location. Decomposition tree has the capability to sort and drilling down data based on category order, then this feature is used to drill down empty miles of the overall network to the empty miles of each country, and lastly it will drill down to empty miles of each city. After understanding how the empty miles are distributed, then another page connected to decomposition, tree is created to understand the network visualization on which location has the potential of improvement for empty miles.

4. Lane Analysis

This section analyses the network based on individual lane. It provides insight of the empty miles percentage of each lanes and how it contributes to the total empty lanes. Therefore, the empty miles insight gives the capability to rank lanes based on their performance. In addition, a feature to remove specific lane from the data collection is added in order to analyse the impact of removing certain lane from the network.

4.3.2 Dashboard Features

Based on Dashboard Sections elaboration, several different types of features are selected to develop the dashboard. The features are gathered to create sections based on their goal. At the end of the development, there are 8 different pages created. Further, the following paragraphs elaborates the functionality of dashboard features during the development phase.

Section	Page Name
Network Overview	Overall Network Performance
Key Influencer	Previous Key Influencer
	Next Key Influencer
Root and Cause Analysis	Decomposition Tree: Previous Empty KM
	Decomposition Tree: Next Empty KM
	Tree Details: Previous
	Tree Details: Next
Lane Analysis	Lane Analysis

Table 4.3: Dashboard sections

1. Measure

In data visualization some additional KPI needs to be created. In power BI measure feature aims to create and calculate collection of data into a new field. The calculated data is then able to be used for visualization. The measure feature use DAX (Data Analysis Expression) query to analyse and aggregate data into a value. Moreover, measure could also be used to manipulate visualization by putting it in the visualization's filters. Measure that are created and their mathematical reasoning behind it are listed in Table 4.3.

Measure Name	Formula
NextEmptyKM	Total sum of DistanceToNext field
PreviousEmptyKM	Total sum of DistanceToNext field
EmptyKM	(NextEmptyKM+PreviousEmptyKM)/2
FullKM	Total sum of Distance field
NextEmptyKM%	(NextEmptyKM measure)/(EmptyKM measure +FullKM measure)
PreviousEmptyKM%	(NextEmptyKM measure)/(EmptyKM measure +FullKM measure)
EmptyKM%	(NextEmptyKM% measure +PreviousEmptyKM% measure)/2
FullKM%	(FullKM measure)/(EmptyKM measure + FullKM measure)
Average Empty KM	(EmptyKM measure)/(Shipemnt Frequency)
ShipmentFrequency	Count the number of ShipmentID field
Exclude	An algorithm that is used to reverse slicer from include to exclude

Table 4.4: Measures for the dashboard

2. Slicer

When exploring the data insight, filtering and selecting certain range of data is the main advantage of using interactive visualization. This data manipulation ability is supported by a feature called slicer. Slicer allows users to choose specific values that they want to display in visualization. The feature can be used in two ways. It can either select one specific value or multiple values. The two functionalities can be adjusted during the development of the dashboard. However, the user can only use either one of them once the dashboard is published. For our dashboard we use two types of slicers in four different uses.

a. (Basic) Slicer

For the proposed dashboard, this slicer is used to (multi) select countries in the “Overall Network Performance” and customers in “Decomposition Tree” page. Furthermore, it is also used to exclude or include lanes from lanes table in the “Lane Analysis” page. Excluding selected value is a reverse functionality of the actual purpose of the feature. In order to exclude the selected value an algorithm measure called ”Exclude” is created. The “Exclude” measure is placed as the Slicer's filter input, meanwhile the input of the slicer is the lane names.

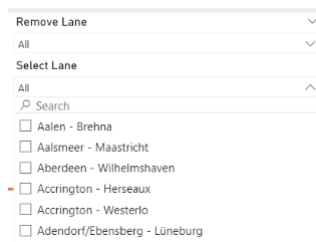


Figure 4. 2: Slicer

b. Timeline Slicer

This slicer is built to create date data type slicer. The specialty of this feature is the ability to categorize date into years, quartiles, months, and days. Therefore when using this slicer we do not need to separate year, month, and day of the date into separated entities. In the propose dashboard this slicers is used in two different ways. First in the “Overall Network Performance” page as month slicer. Secondly, in “Tree Details” pages as year slicer.

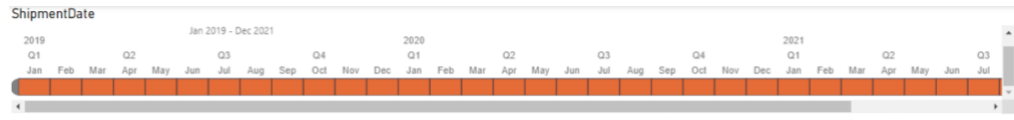


Figure 4. 3: Timeline slicer

3. Data Metrics and Visualization

The data insights are presented in two primary ways. For category comparisons the data are presented in charts. Further, to provide more extended information of how several data connected to one another tables are selected as the presentation feature.

a. Clustered bar chart

This bar chart is used to present the customers data based on empty miles and shipment frequency. This type of bar chart could be used to present different data categories in one visualization, however it is currently not needed for our dashboard. Both visualization is sorted decreasingly so it displays the data based on rank.

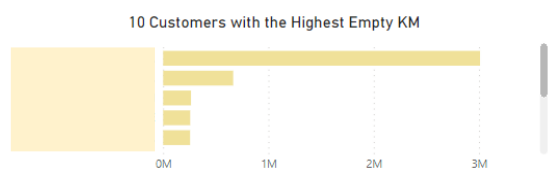


Figure 4. 4: Clustered bar chart

b. Ribbon Chart

This chart is used to compare the the Full KM and Empty KM over times. The visualization combines bar chart and legend, it allows the user to see the compare whether there is significant fluctuation of Full KM and Empty KM over selected period of time.

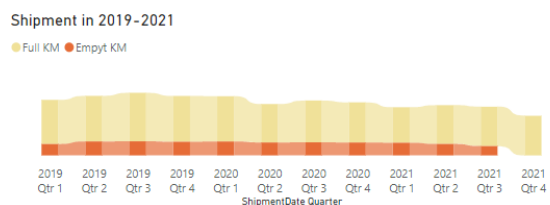


Figure 4. 5: Ribbon chart

c. Column Chart

This bar chart is used to compare monthly shipment numbers in “Tree Details” Pages.

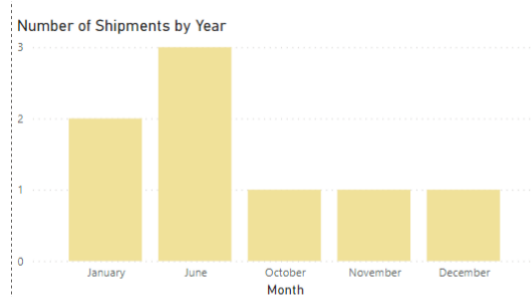


Figure 4. 6: Column chart

d. Table

When analysing the shipment network we want to know how different attributes of certain lanes or region contribute to the total network. Therefore, table is selected as a metric to provide broad data insight. In the proposed dashboard, tables are used in two different ways. The first table is displayed in the “Decomposition Tree” pages to show shipment frequency to a specific city. Further a more comprehensive version is displayed in “Tree Details” pages to show how different cities in a selected lane region contribute to the total empty miles. Since the empty miles data are presented within selected period of time, thus empty lane frequency is also presented in the table. Lastly, the table is used to provide extensive information how each lanes contribute to empty miles. In this table empty miles are presented in several different perspective, which are empty miles percentage per lane, empty miles percentage to the overall network, empty miles in KM, frequency and some other similar information regarding full KM.

PreviousCity	Frequency
Aalst	1
Aartselaar	1
Aberdeen	1
Accrington	2
Adendorf/Ebensberg	1
Agnadello	9
Airdrie	1
Alkmaar	1
Almere	8
Alphen aan den Rijn	25
Total	42633

Figure 4. 7: Table

4. Location Visualization

e. Heat map

Heat map display location visualization with color gradient. The area with higher gradient indicate high level of density, meanwhile low gradient indicate low level of density. In the proposed dashboard the heat map is used to show the density region of loading and unloading location. The loading location is referred as first location, meanwhile the unloading location is referred as last location.

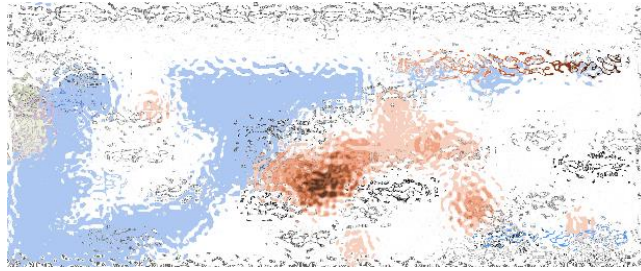


Figure 4. 8: Heat map

f. Flow Map

Flow map display the connection of two location using line. In the proposed dashboard it is used to display the empty lane connection from a selected loading or unloading location.

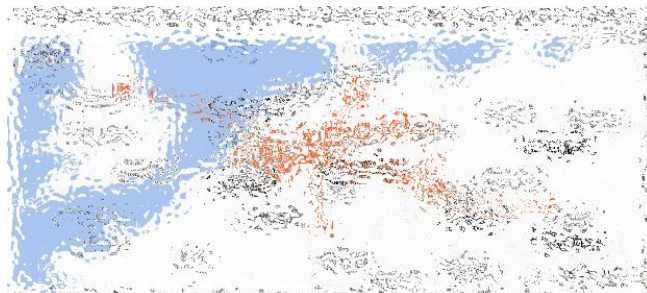


Figure 4. 9: Flow map

5. Supporting Feature

a. Card

It is a visualization that shows a number or metric. In the dashboard it is used to show the performance of the network by displaying full KM and empty KM percentage. This card is used in “Overall Network Performance” page, “Tree Details” pages and “Lane Analysis” page.

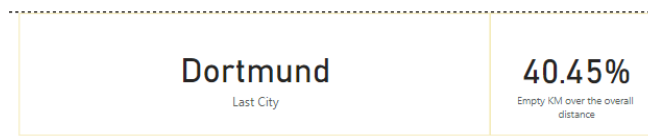


Figure 4. 10: Card

b. Button

Power BI enables to trigger certain actions by using a customized button. In the dashboard some buttons are created to connect several pages. Further, there is also an automatic button called “Page Navigation” that will bring user back to previous page.



Figure 4.11: Buttons

c. Drill through

This feature allows user to dig deeper into the details of specific data. This feature is created by inputting the selected field that will be drill through at the page where the more detail information is displayed. In the dashboard drill through is used to connect “Decomposition Trees” pages with the “Details Tree” pages to allow user to analyse certain location further. The way to use the feature is by clicking right, then choose drill through to specific pages.

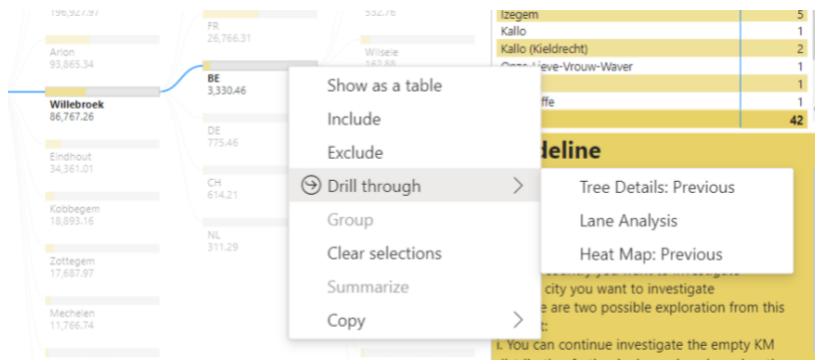


Figure 4.12: Drill through feature

Chapter 5: Implementation

The selected and developed features are gathered to build the dashboard. In this chapter, the implementation of the dashboard is elaborated. It is begin by explaining the guideline on operating the dashboard in section 5.1. Then, a scenario analysis and a demonstration of how the dashboard could be implemented for specific goal is explained in section 5.2.

5.1 Page Guideline

The dashboard consists of 8 different pages in which several of them has the same feature, however presenting different each present different data field. Figure 5.1 shows the workflow of the dashboard. As shown in the diagram, the “Key Influencer”, “Decomposition Tree” and “Detail Tree” pages are created in two different versions, the Previous and Next empty lane. The reason of the separation is to allow the user to investigate lane from two-point of views, which are the Previous trip and the Next trip of shipment. Thus, the user will be able to find individual improvement for each lane. Further guideline to use the dashboard is elaborated in the next paragraphs.

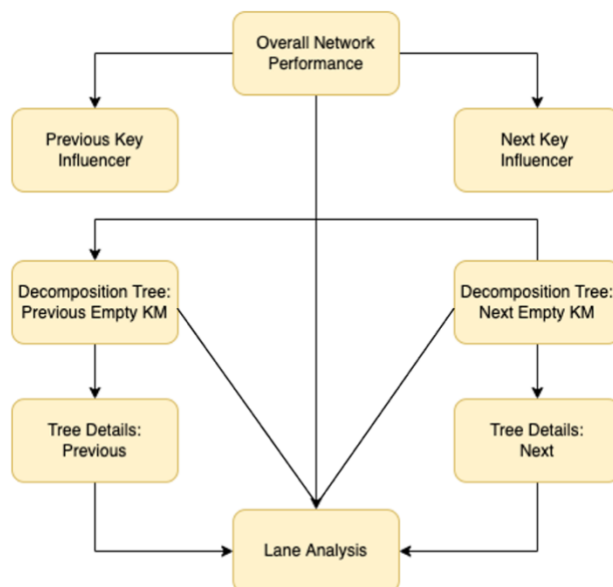


Figure 5. 1: Page guideline framework

1. Page 1: Evaluating Overall Network

Figure 5.2 displays the main page of the dashboard. This page shows the overall insight of the shipping network. This page is divided into four sub sections. The first section is located on the top, it displays shipment date slicer to filter the network insight based on user desired time period. Afterward there is another navigation section on the left. On the left top corner there is a slicer to multi select countries, the aim is to enable the user to evaluate network performance of certain countries. Below the filter, a card is displayed on the left side of the page to show the network performance. In addition, there are five buttons in which will direct the user to other pages. Further, the third section of this page is the two maps in the middle. There are two heat maps that can be zoomed in and zoomed out. The red heatmap shows the density of the loading locations, meanwhile the yellow heatmap shows the density of unloading locations. Finally, on the right side of the page several visualizations are displayed to show the customer ranks and trend of empty and full KM over times.

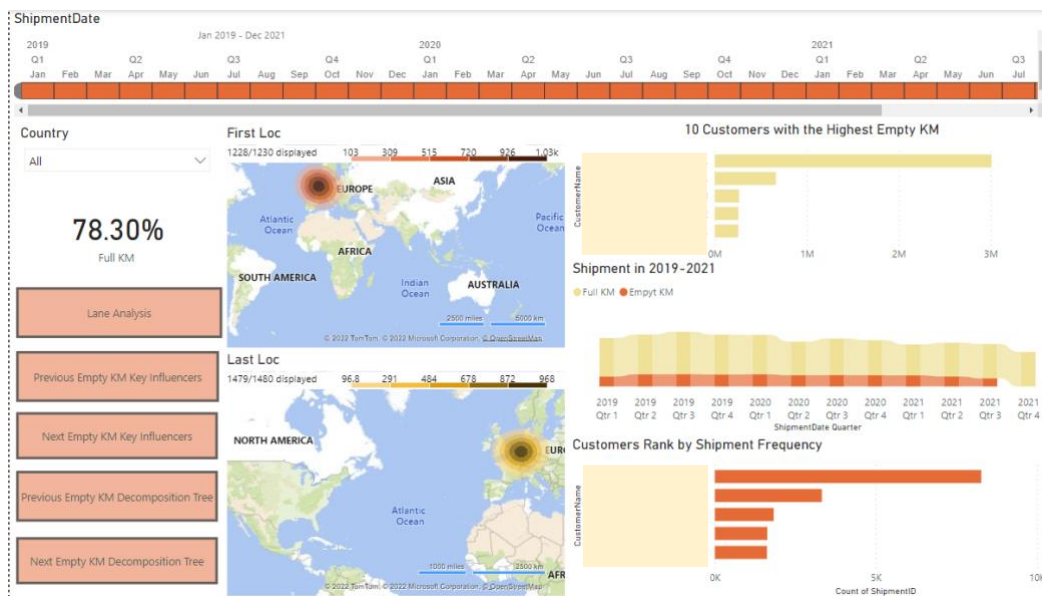


Figure 5.2: Overall network performance page

II. Page 2 and Page 3: Key Influencer

After understanding the overall performance of the network, the user will be able to see which customer or location contributes to high or low empty KM with the key influencer. This page has three main sections. As shown in Figure 5.3, on the top page there is a filter in which will decide what the user wants to predict from the Previous Empty KM, whether the users want the parameter to increase or decrease. Secondly, on the left side of the page there is a list of key factors that influence empty KM. Finally, the right side shows the statistics that explains a selected key influencer on the left. In order to provide better understanding figure 5.3 is will be explained further in the next paragraph.



Figure 5.3: Key influencers page

The key influencer displays several key factors that has major impact on the Previous Empty KM. The result shows that if a lane has previous unloading location from city P or city Q, then it means the Previous Empty KM is more likely to increase. Then, if we select the top right circle that belongs to city P, a statistic visualization is displayed. This statistics bar chart shows that the average of Previous Empty KM per lane is 541.59 KM, meanwhile the empty KM of lane with previous city P is more than 10,000 KM in which far above the average. Thus, lane with previous city P is more likely to increase Previous Empty KM.

III. Page 4 and Page 5: Decomposition Tree

This page consist of two parts, a decomposition tree as the main feature on the left and supporting additional information on the right side. The decomposition tree shows how the empty KM spreads in different locations. As shipment network is connecting multiple locations, this feature allows user to explore the empty KM distribution based on the lane. It is ordered in such a way that user will choose a country they want to explore, then the list of cities in the corresponding country is shown. Further, for additional information there is slicer that selects customer name so user can analyse empty KM of specific customer. In addition, a table showing the shipment frequency to a corresponding city is displayed.

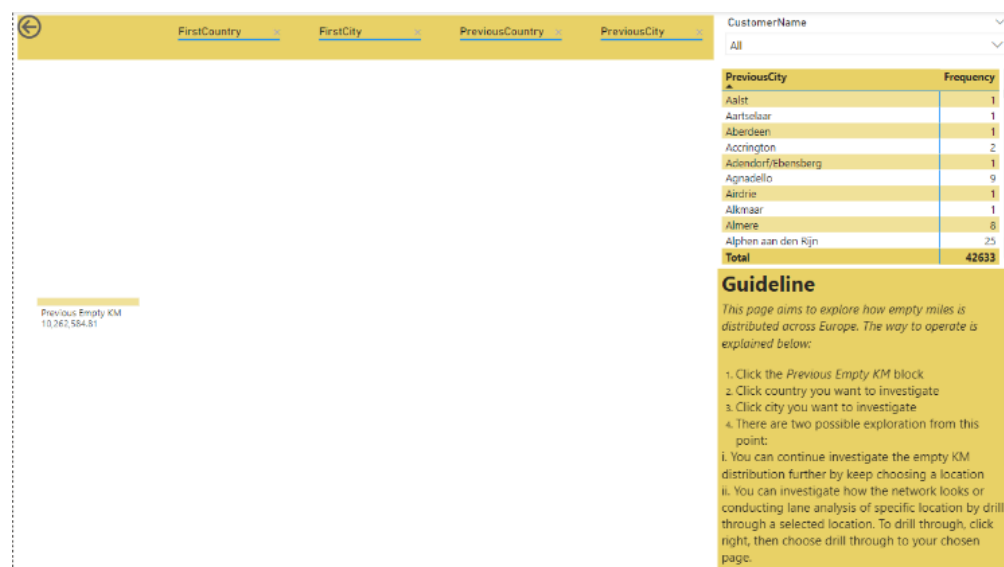


Figure 5.4 Decomposition tree page (before being operated)

Before operating the decomposition tree, the feature only display the total Previous Empty KM block as shown in Figure 5.4. In addition, Figure 5.5 shows how the decomposition visualization after it is being operated. It is also connected with drill through feature in order to analyse the empty miles in selected location more in depth. For further operating information a guideline is provided on the page.



Figure 5.5: Decomposition tree page (while being operated)

IV. Page 6 and Page 7: Tree Detail

This page can only be used after selecting a specific location in the decomposition tree. In the previous section R has been selected to be drilled through, then this page shows details information about the empty KM if the Last City is R.

As shown in Figure 5.6 the top left side of the page display a Shipment Date slicer to select a specific year from the network historical data, meanwhile, the top right side of the page provides information about the name of the selected location and the empty KM percentage. Further, the left side of the page display the statistic of shipment and how the empty miles spreads in different Next Location. The table shows the total empty KM distance, this means that the distance has been accumulated with all shipment frequencies. Therefore, in order to find the empty distance per lane, the user needs to divide the total empty KM distance with the lane shipment frequency. Finally, on the right side of the page network and density is shown. The flow map aims to provide inside potential location in which can be focused on to search for additional customers. Meanwhile the heat map shows the density of shipment frequency. Therefore, by knowing both information user will have an idea which location need to be prioritize based on network and shipment frequency.



Figure 5.6: Tree detail page

V. Page 8: Lane Analysis

The last page as shown in Figure 5.7 is created to analyse the performance of each lane. It consists of four different features. On the right side of the pages several performance measures are displayed. The upper bar chart ranks the lanes with highest empty KM percentage in the overall network. Meanwhile, the lower bar chart ranks lanes in which has the least value offered for the company from distance perspective. In the middle three different slicers are played. Then two bar charts emphasize the rank performance of lanes based on empty KM percentage and lowest added value of lane. The reason that the visualization displays the worst lanes is to enable user to find room for improvement in those lanes. Lastly, an extensive table is displayed at the bottom to show lane performance from several point of views. The table can be sorted based on specific column by double clicking the header of selected column.



Figure 5.7: Lane analysis page

5.2 Scenario Analysis

The process of scenario analysis development framework (Mahmoud et al., 2009) is divided into five stages: scenario definition, scenario construction, scenario analysis, scenario assessment, and risk management. In this section we adapt the framework to make the simplified test of scenario analysis. First, the scenario is defined by choosing a case. Furthermore, the process of constructing, analysing, and assessing the scenario (lane) are combined as a problem approach on solving the case. Meanwhile the management risk stage is out of scope of dashboard implementation, therefore it is not included in this section.

5.2.1 Case

The overall network performance page indicates customer B as the customer with the second highest empty miles. In addition, it is also customer with the second shipment frequency. These two information shows that although this customer contributes a high amount of empty KM, at the same time it also provides high demand for the company. This customer is considered important based on the the number of shipment history which therefore should not be removed. Thus, this problem leads to a case question as follow.

Case: How to Improve Empty Miles from Customer B?

5.2.2 Problem Approach

In order to improve the empty miles, we need to find location in which potentially could be focused on to find additional customers. This approach is conducted for the purpose of testing the dashboard, for the real implementation in the company stakeholder may use their own approach of decision making.

To begin with, we need to know which lanes Customer B shipment go to, this analysis will be conducted in “Lane Analysis” page. Afterward, the analysis will go further by analysing the lane empty miles distribution using both “Decomposition Tree: Previous Empty KM” and “Decomposition Tree: Previous Empty KM”. Finally, we will check the “Details Tree” pages to find the exact lane location for improvement.

5.2.3 Analysis

a. Analysing Customer B Lanes

Figure 5.8 shows the result of lane analysis page after filtering lanes in which only belong to customer B. In overall the customer has 32% of empty KM in which spread in 21 lanes, in addition the average empty KM per lanes is 200.1 KM.



Figure 5.8: Analysing customer B performance using lane analysis page

The Added Value bar chart does not show any visualization since the aim is to only highlight lanes in which has negative added value or in other words lanes in which are not profitable enough. This implies that all lanes in which Customer B are considered good lanes as all lanes have positive values. In order to reduce empty miles of customer B there are two possible approach that can be done:

1. Removing lane with high empty KM
2. Finding additional customer in the empty miles region

The dashboard is developed to assist the two approaches of decision making listed above. The application of both approaches is demonstrated in the next sections.

b. Removing Lanes

The table has an ability to sort data based on selected column. It is done by clicking header of the columns user wants to sort. By sorting the empty KM per lane column we found that there are 9 lanes with high empty KM above average. From these 9 lanes, 4 of them have empty KM

percentage above 40% in which very likely to be reduced. Before removing the four lanes we could see the added value each lane contributes to the network. It is found that three of the lanes are placed as the three lowest added value in the network. The three lanes are:

- S – T
- S – U
- S – V

These lanes are removed from the network by selecting them in the “Remove Lane” filter. Finally, as shown in the figure 5.9 the Empty KM of customer B has reduced by 0.03% compared to empty KM in figure 5.8.

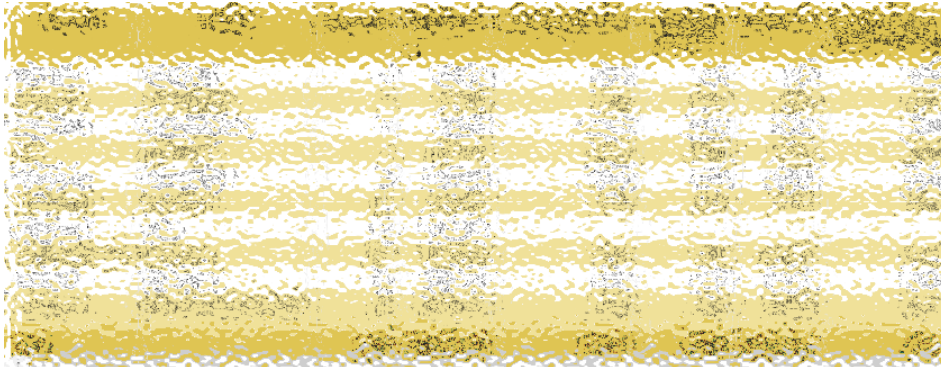


Figure 5.9: Lane analysis table after removing several lanes

c. Locate Region for Improvement

From figure 5.8 lane S – R has high empty KM of 45.86%, however it also has high shipment frequency of 640 times of shipment. This implies that the lane has high importance thus it should not be removed. Therefore, this lane could be improved by trying to find additional customers to fill the empty miles of that line. The investigation is conducted using Decomposition Tree and some related page. The analysis results are explained below:

1. Previous Empty KM

After exploring the decomposition tree, is found that the highest Previous empty KM is in empty lane W – S where it has 308 shipment frequency history.



Figure 5.10: Previous empty KM decomposition tree of customer B

However, as shown in the Figure 5.8 the highest total empty KM to city S is located in country X with total of 542,808.09 KM. In order to understand how the location, spread, a heat map analysis using drill through is conducted. The result displays in Figure 5.9 shows that the empty KM is centralised in three different locations.



Figure 5.11: Previous location heatmap of S

2. Next Empty KM Analysis

The next Empty KM decomposition tree shows that the empty KM is centralised in country Y where the highest came from empty lane to city X as shown in figure 5.10.



Figure 5.12: Next empty KM decomposition tree of customer B

In order to see more improvement possibilities, an analysis to the second country is conducted using drill through. The “Tree Details” shows that empty KM in country Z reach 88.81%. The empty KM percentage implies that the distance of empty KM is higher than the full KM. (Note: Figure 5.13 and Figure 5.6 are both showing the same page and last city, however Figure 5.13 displays higher empty KM percentage by reason that it only calculates the empty KM of customer B with the last city R. Meanwhile, Figure 5.13 evaluate the empty KM of the whole network.)

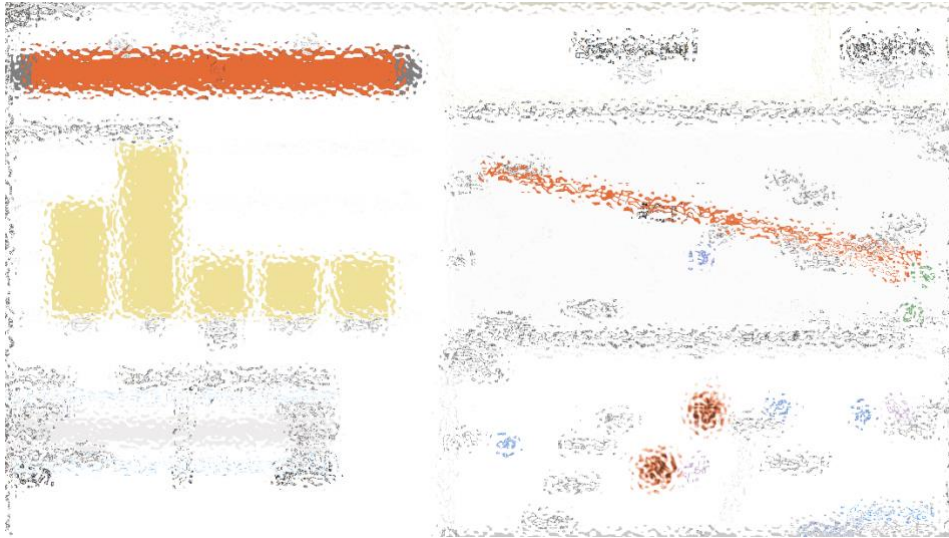


Figure 5.13: Tree detail of next empty KM.

d. Conclusion and Recommendation

The lane examination found two lanes in which can be improved, the lanes are W – R and R – S. Furthermore, high percentage of empty KM is found in country Z. Thus, the company needs to reconsider the network connection from city R to country Z.

Chapter 6: Validation

This chapter elaborates the process of validating the source, content, and tool design with reality in order to create reliable deliverables. It is mainly into two different processes of internal and external validation. The internal validation is validating the data source before implementing it for tool creation, it is elaborated in section 6.1. Meanwhile, section 6.2 elaborates the external validation in which validating the tool design with the stakeholder to adjust the prototype with reality and company needs.

6.1 Data Validation

In this validation process we refer to good quality data characteristics (Gualo et al., 2021) to assess the quality of the dataset we use. Using characteristic listed in table 6.1 several steps of data validation is performed. Some of the steps of data validation has been partly discussed as part of data preparation phase in section 4.2. This section is more focus on assessing the data criteria one by one.

Characteristic	Definition
“Accuracy”	The degree to which the data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.
“Completeness”	The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use.
“Consistency”	The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use.
“Credibility”	The degree to which data has attributes that are regarded as true and believable by users in a specific context of use.
“Currentness”	The degree to which data has attributes that are of the right age in a specific context of use.

Table 6. 1: Good data characteristics (Gualo et al., 2021)

1. Data Accuracy

In assessing the accuracy of data an initial data exploration was conducted. In this process data were visualized in graph and map based on their data types. In assessing the date data type and shipment frequency line trend and bar chart were used to check the correctness of data range and to see if there is an outlier. Furthermore, in order to check the accuracy of location data map is used to visualised the location. It was found that several location are displayed inaccurately. After doing investigation it was found that there are certain techniques in power BI that is used to displayed location accurately. Applying the visualization technique by combining some location entities eventually solved the issue of inaccurate location.

2. Data Completeness

In assessing the data completeness we conducted data sorting to search for missing data. There were found lanes with missing empty lane distance and full lane distance. As explained in section 2.3, lanes connects several different location that are connected by both empty and full lanes. In practice it is possible if certain existing lanes does not have empty lane, therefore missing empty lane distance is not an issue. However, for lanes that are missing the full lane distance the lane attributes considered incomplete, thus the lanes were removed.

3. Data Consistency

To assess the data consistency, data matching and cleaning were conducted using Python as elaborated in section 4.2. In this step, the misspelled locations were found in which consequently double the data entries. The error data then corrected to yield a consistent data set.

4. Data Credibility and Currentness

The data for this research is provided by the company that does the logistic shipment for the research case. Therefore, the credibility and currentness of the data are guaranteed.

6.2 Content Validation: Stakeholder input

The validation phase with the stakeholder is divided into three different main aspects. Firstly, we make sure whether the measurement inside the dashboard is valid. Further, the features and visualizations are evaluated to assess whether they could precisely display the information to the users. Finally, the prototype of the whole tool is presented to check their correctness in reality.

6.2.1 KPI Measurement

The main KPI of the dashboard is the “Empty KM” percentage. In validating the definition of empty miles and how it is mathematically formulated, we had a discussion with company supervisor. The first measurement was not valid as the previous and next empty miles are relatively similar. Thus, to make a better calculation we sum the previous and next empty miles and divided them by two. Furthermore, some more KPI emerged during the development process. In certain page, for instance “Overall Network Performance” the stakeholder addressed that “Full KM” miles would be a better choice to show the network performance. Moreover, several other KPI measurements emerged during the “Lane Analysis” page development phase to complete the lane analysis table overview. Among the KPI measurement are “added value” KPI, and several empty KM and full KM percentages. At the end of the validation phase we noticed that the measurement of “added value” KPI is not sufficiently mathematically elaborated. We realized that to decide a value in which a lane contributes to a network, a more complex analysis and calculation should be performed. However, the research time constraint does not allow us to perform added value analysis, thus the “added value” KPI is not going to be implemented by the company.

6.2.2 Visualizations and Features

In validating the features and visualizations a number of meetings were done to discuss the usability of selected visualizations and features. In the first phase a meeting with the company supervisor to assess the selected network visualization choices for the dashboard. The initial network visualization was the network navigator, however, due to its incapability to display location connection in a map the visualization was removed. Afterward, in the second phase of validation a flow map visualization, key influencers, and decomposition tree were presented to the company supervisor. The flow map has proven to clearly showcase the shipping network in a way company hopes and how it actually looks in reality. Further, the decomposition tree and key influencer presented new insight that the company has never seen. Thus, the second phase of validation was successful. During this phase, we were given feedback to add slicers and bar charts to allow data exploration flexibility and provide more insights.

6.2.3 Final Prototypes

In the final prototype validation phase a meeting was arranged with the company supervisor who has a role as Data Engineer and another company stakeholder from Sales department as user target for the dashboard. At this phase the dashboard has been developed as three different sections of Overall Network Insight, Key Influencers, and Roost Cause Analysis. During the meeting the dashboard was demonstrated by following the dashboard workflow. Afterward, the stakeholder give insight and feedback of how valid the dashboard with the real situation. Furthermore, a questionnaire was given to the stakeholders to provide feedback for the dashboard and asses their satisfaction of the proposed deliverable. The feedback is attached in the Appendix B. The meeting concludes that stakeholders satisfied with the proposed tool with an “excellent” remark.

After the first validation of the final prototype, the dashboard was still further improved to add one page of lane analysis. Afterward, a final validation meeting was done with the company supervisor. During the final validation phase, we found that the tool enables to calculate the impact of removing lanes to the total network performance. The “Lane Analysis” page enables user to explore the lane's performance from a number of different KPIs. Moreover, this section also provides a visualization that compares the value each lane contributes to the network. However, the “added value” feature is not sufficiently reliable to be implemented by the company. In conclusion, excluding the “added value” feature on the “Lane Analysis” page, the dashboard is qualified to be implemented in the company.

Chapter 7: Conclusion and Recommendation

The last chapter covers the summary of the research findings. In section 7.1 the research aim is restated and the research questions is answered into three main points. Further, limitation of the research and tool is explained in section 7.2. Lastly, recommendation for future work is elaborated in section 7.3.

7.1 Conclusion

The research was begin to investigate a method to solve four problem contexts related to empty miles issue which is explained in section 1.4.1. After conducting further problem investigation and research feasibility the aim of the research is narrowed to create a decision support tool to determine how lanes empty miles contribute to the total freight shipping network. The aim of finding the method is to provide insight for Emons' Sales Department to determine how they should order their customers in terms of the added value they offer to the company from the perspective of minimizing the empty miles of shipment. Thus, the problem identification phase lead to the main research question: *"How can a decision support tool analyses and provide data insight for helping Emons Sales departments to minimize empty miles of overall logistic network"*. Further, it developed into three sub-research questions. The research findings based on literature reviews and tool developments are summarized in the following paragraphs.

1. *Limited literatures regarding network shipping decision support tool in tactical level:*
From literature reviews, shipping network and empty miles issues are most often being approached from an operation research point of view. The existing literatures shows that decision support tool for shipping network primarily attempt to design optimum network or to identify specific location for network planning using a mathematical model and simulation method. Thus, the result of the tools normally will be applied in strategic level of planning. It was not found an existing shipping network decision support tool in which aligns with our research goal. Our research aims to create decision support tool to minimize empty miles in the tactical level. Therefore, we found a gap in literature for this particular topic. In order to solve the action problem, we conducted visualization approach exploration in which combined with searching for machine learning potential for empty miles issue.
2. *Self-service Machine Learning feature for shipping network analysis:*
Shipping network has enormous data size with dynamic and complex relation between location. As a result, optimization algorithm in machine learning plays important role on translating data-driven decision making. After conducting literature reviews and visualization exploration in Microsoft Power BI, we found self-service machine learning features in Microsoft Power BI that is able to be incorporated in the decision support tool. The feature enables user to find empty miles region by implementing decision tree and regression algorithms. The features are self-serviced thus it enables users with no programming and mathematical background to implement the feature.
3. *Network analysis with scale down concept:*
The research results in a tool design with main concept of scaling down the shipping network. The decision supports tool allows users to analyses the network from several different point of views. It provides systematic insight from general network view that is able to be scaled down into specific parts of the network. Moreover, the tool uses variety of data insight techniques that are spread over four main sections of the tool. The sections are ordered to follow a systematic approach of scaling down the network,

however once the users understand the functionality of each section, they will have flexibility to explore the network from any section of the tool.

To summarize, this research could contribute to the practical research on how to create empty miles minimization decision support tool for tactical planning. The reason for that is that the majority of research focus on the strategical optimization on designing a network. Meanwhile, based on literature review the number of tool research for tactical approaches are relatively less compare to research for the strategical approaches.

Furthermore, using the proposed tool the company would be able to implement their strategy on finding location to search for additional customers, mainly for the Sales and Expedition of Emons. However, since the problem context is narrowed down from the initial four identified problems, the proposed tool is not capable enough to be used by the account management to determine lane or customer added value to the network for creating customer relationship contract plan.

7.2 Limitation

The prototype was developed based on the simplified situation of the real truck shipping network. This research used a dataset that excluded the cost aspects of the shipping network. As consequence, it was difficult to accurately measure how lanes contribute to network profitability. Although the lanes are able to be evaluated based on their travel distance, the cost aspect still plays an important role in determining lanes' priority order.

In practice, when a lane is removed from a network then it will directly affect the next connecting lane. It will either cause shipment delay due to losing the previous connection, or the other possibility is the need to also remove the connected lanes. Thus, removing lanes in practice could possibly increase the empty miles. Therefore, the aim of reducing empty miles by removing certain lanes could cause a chain of problems. This scenario has not been included in the lane analysis section. In addition, an in-depth analysis should also be conducted to measure whether the removal of certain lanes will yield more profit or loss for the company. Due to the explained reasons, cost plays a crucial role in evaluating lanes' added value to the total network. Therefore, the lane analysis section in the prototype will mainly be used as rough insight on understanding relations between lanes in terms of the empty miles measurement.

7.3 Recommendation

The current decision support tool prototype provides extensive insight into finding which regions contribute to high empty miles. Using the prototype tools the sales department would be able to focus on searching customer in region with high empty KM based on the dashboard data insight. Compared to the current tool of matrix shipment balance, the proposed dashboard has higher reliability on selecting precise improvement region. The reason for that is that the dashboard shows location to the city level and provide the empty miles gap distance. By knowing the empty miles distance, the Sales would be able to have approximate the distance of additional lane that they need to add to the network.

The research has solved one of the four problem contexts which stated in section 1.4.1. Further research should be conducted to solve the other three problems and how they would be integrated with the current prototype. However, a few parts of the tool could also still be improved for further development and research. For instance, by expanding the dataset with cost and several related aspects such as number of trucks as well as the amount of gas emission

per trip, the Key Influencer section could be improved for more comprehensive insight. Furthermore, the current “Lane Analysis” section is still lacking in mathematical reasoning to weight which lane is worth removing. At this stage, the “Lane Analysis” page has shown the empty miles contribution per lane, however high or low number of empty miles is not enough to be used to measure overall network efficiency as there are more aspects that should be taken into consideration.

In order to provide a more valuable lane analysis, travel cost per mile should be included into the model to weight the added value of each lane. Furthermore, the number of lane connection is also crucial to be taken into consideration in deciding lane and network change. Based on literature review, mathematical model and machine learning are two potential approaches that could be implemented for further research. Thus, the current “Lane Analysis” section could be used as a base on analysing the deeper relations between lanes to create algorithm that evaluate lane connection and network optimization in shipping network.

Reference

- Alam, M. D. J., & Habib, M. A. (2021). A dynamic programming optimization for traffic microsimulation modelling of a mass evacuation. *Transportation Research Part D: Transport and Environment*, 97, 102946. <https://doi.org/https://doi.org/10.1016/j.trd.2021.102946>
- Bachechi, C., Po, L., & Rollo, F. (2022). Big Data Analytics and Visualization in Traffic Monitoring. *Big Data Research*, 27, 100292. <https://doi.org/https://doi.org/10.1016/j.bdr.2021.100292>
- Balster, A., & Friedrich, H. (2019). Dynamic freight flow modelling for risk evaluation in food supply. *Transportation Research Part E: Logistics and Transportation Review*, 121, 4-22. <https://doi.org/https://doi.org/10.1016/j.tre.2018.03.002>
- Bascur, O. A., & O'Rourke, J. (2020). Chapter 6 - Measuring, managing, and transforming data for operational insights. In M. Soroush, M. Baldea, & T. F. Edgar (Eds.), *Smart Manufacturing* (pp. 129-165). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-820027-8.00006-X>
- Brocke, J. v., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. In (pp. 1-13). https://doi.org/10.1007/978-3-030-46781-4_1
- Čerić, V. (1997). Visual interactive modeling and simulation as a decision support in railway transport logistic operations. *Mathematics and Computers in Simulation*, 44(3), 251-261. [https://doi.org/https://doi.org/10.1016/S0378-4754\(97\)00055-4](https://doi.org/https://doi.org/10.1016/S0378-4754(97)00055-4)
- Don Taylor, G., DuCote, W. G., & Whicker, G. L. (2006). Regional fleet design in truckload trucking. *Transportation Research Part E: Logistics and Transportation Review*, 42(3), 167-190. <https://doi.org/https://doi.org/10.1016/j.tre.2004.11.001>
- Double-Decker International Shipping Service. (2022, 3 march 2022). Emons Cargo 2WIN. <https://www.2win.eu/2win-double-decker/>.
- Epstein, R., Neely, A., Weintraub, A., Valenzuela, F., Hurtado, S., Gonzalez, G., Beiza, A., Naveas, M., Infante, F., Alarcon, F., Angulo, G., Berner, C., Catalán, J., Gonzalez, C., & Yung, D. (2012). A Strategic Empty Container Logistics Optimization in a Major Shipping Company. *Interfaces*, 42, 5-16. <https://doi.org/10.2307/41472725>
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189-1232. <http://www.jstor.org/stable/2699986>
- Gonçalves, J., Santos, B., & Oliveira, A. (2022). Prioritization of railway proximity interventions: The case of the Portuguese railway network. *Case Studies on Transport Policy*, 10(1), 559-571. <https://doi.org/https://doi.org/10.1016/j.cstp.2022.01.016>
- Gregor, S., & Hevner, A. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37, 337-356. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Gualo, F., Rodriguez, M., Verdugo, J., Caballero, I., & Piattini, M. (2021). Data quality certification using ISO/IEC 25012: Industrial experiences. *Journal of Systems and Software*, 176, 110938. <https://doi.org/https://doi.org/10.1016/j.jss.2021.110938>
- Hevner, A., R, A., March, S., T, S., Park, Park, J., Ram, & Sudha. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28, 75.
- Irannezhad, E., Prato, C. G., & Hickman, M. (2020). An intelligent decision support system prototype for hinterland port logistics. *Decision Support Systems*, 130, 113227. <https://doi.org/https://doi.org/10.1016/j.dss.2019.113227>
- Kalamaras, I., Zamichos, A., Salamanis, A., Drosou, A., Kehagias, D. D., Margaritis, G., Papadopoulos, S., & Tzovaras, D. (2018). An Interactive Visual Analytics Platform for Smart Intelligent Transportation Systems Management. *IEEE Transactions on*

- Intelligent Transportation Systems*, 19(2), 487-496.
<https://doi.org/10.1109/TITS.2017.2727143>
- Kuzmicz, K. A., & Pesch, E. (2019). Approaches to empty container repositioning problems in the context of Eurasian intermodal transportation. *Omega*, 85, 194-213.
<https://doi.org/https://doi.org/10.1016/j.omega.2018.06.004>
- Lei, T. L., & Church, R. L. (2011). Locating short-term empty-container storage facilities to support port operations: A user optimal approach. *Transportation Research Part E: Logistics and Transportation Review*, 47(5), 738-754.
<https://doi.org/https://doi.org/10.1016/j.tre.2011.01.004>
- Loza-Hernandez, L., & Gendreau, M. (2020). A framework for assessing hazmat risk at nodes of transport networks. *International Journal of Disaster Risk Reduction*, 50, 101854. <https://doi.org/https://doi.org/10.1016/j.ijdrr.2020.101854>
- Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., & White, D. (2009). A formal framework for scenario development in support of environmental decision-making. *Environmental Modelling and Software*, 24, 798-808.
<https://doi.org/10.1016/j.envsoft.2008.11.010>
- Pais Montes, C., Freire Seoane, M. J., & González Laxe, F. (2012). General cargo and containership emergent routes: A complex networks description. *Transport Policy*, 24, 126-140. <https://doi.org/https://doi.org/10.1016/j.tranpol.2012.06.022>
- Pan, X., Ning, L., & Shi, L. (2019). Visualisation and determinations of hub locations: Evidence from China's interregional trade network. *Research in Transportation Economics*, 75, 36-44. <https://doi.org/https://doi.org/10.1016/j.retrec.2019.01.001>
- Patriarca, R., Di Gravio, G., Cioponea, R., & Licu, A. (2022). Democratizing business intelligence and machine learning for air traffic management safety. *Safety Science*, 146, 105530. <https://doi.org/https://doi.org/10.1016/j.ssci.2021.105530>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77. <https://doi.org/10.2753/MIS0742-1222240302>
- Schulte, F., Lalla-Ruiz, E., González-Ramírez, R. G., & Voß, S. (2017). Reducing port-related empty truck emissions: A mathematical approach for truck appointments with collaboration. *Transportation Research Part E: Logistics and Transportation Review*, 105, 195-212. <https://doi.org/https://doi.org/10.1016/j.tre.2017.03.008>
- Taylor, G. D., Broadstreet, F., Meinert, T. S., & Usher, J. S. (2002). An analysis of intermodal ramp selection methods. *Transportation Research Part E: Logistics and Transportation Review*, 38(2), 117-134. [https://doi.org/https://doi.org/10.1016/S1366-5545\(01\)00018-7](https://doi.org/https://doi.org/10.1016/S1366-5545(01)00018-7)
- Tsolaki, K., Vafeiadis, T., Nizamis, A., Ioannidis, D., & Tzovaras, D. (2022). Utilizing machine learning on freight transportation and logistics applications: A review. *ICT Express*. <https://doi.org/https://doi.org/10.1016/j.icte.2022.02.001>

Appendix A: SLR summary

a. Decision support tool literature: purposes

No.	Author	Tool Purpose		
		Optimization	Priority evaluation	Risk evaluation
1.	(Farahani et al., 2018)	x		
2.	(Gonçalves et al. (2022)		x	
3.	(Alam & Habib, 2021)	x		
4.	(Irannezhad et al., 2020)	x		
5.	(Loza-Hernandez and Gendreau (2020)			x
6.	(Balster & Friedrich, 2019)			x

b. Decision support tool literature: problem solving approach

No.	Author	Method						
		MDCA	Simulation	Mathematical Model	Reinforcement Learning	Statistics	Event Analysis	Criteria Scoring System
1.	(Farahani et al., 2018)			x				
2.	(Gonçalves et al. (2022)	x				x		
3.	(Alam & Habib, 2021)		x	x			x	x
4.	(Irannezhad et al., 2020)		x		x	x		
5.	Gendreau (2020)						x	
6.	(Balster & Friedrich, 2019)			x				

c. Empty miles/truck minimization literatures: problem solving approach

No.	Author	Solution Approach			
		Changing Truck Capacity	Truck Collaboration	Locating Optimal Location	Minimizing Circuitous Mile
1.	(Epstein et al., 2012)	x			
2.	(Kuzmicz & Pesch, 2019)	x			
3.	(Schulte et al., 2017)		x		
4.	(Don Taylor et al., 2006)		x		
5.	(Lei & Church, 2011)			x	
6.	(Taylor at al., 2002)				x

d. Empty miles/truck minimization solution method

No.	Author	Solution Method		
		Mathematical Modelling/ Operation Research	Simulation	Practical Approaches
1.	(Epstein et al., 2012)	x		
2.	(Kuzmicz & Pesch, 2019)	x		x
3.	(Schulte et al., 2017)	x		
4.	(Don Taylor et al., 2006)		x	
5.	(Lei & Church, 2011)	x		
6.	(Taylor at al., 2002)	x		

Appendix B: Stakeholder input

Shipment Network Analytics Dashboard					
No.		Insufficient ($< 60\%$)	Sufficient ($60\%-80\%$)	Excellent ($> 75\%$)	Comment
1.	How much insight from the dashboard that is helpful to find locations for searching additional customer?			X	gives immediate insight in network gaps
2.	How effective will the dashboard be as decision support tool for Sales?			X	effectiveness is high, this directly follows out of #1
3.	How easy is it to operate the dashboard?			X	pre-reqs are: is introduction
4.	Are the possibilities to slice and dice the information sufficient? (the possibilities that filer provide)			X	yes, more might be required if more data is present
5.	Is there a balance between providing quantitative and visualization insight?			X	Drill through options after visualisation
6.	Are there any other insights or feature you think could be added? Are there anything that you think could be improved?	scenario analysis → as discussed. high impact level of impact of customers if they leave or increase			

KL

Shipment Network Analytics Dashboard

No.		Insufficient ($< 60\%$)	Sufficient ($60\%-80\%$)	Excellent ($> 75\%$)	Comment
1.	How much insight from the dashboard that is helpful to find locations for searching additional customer?			x	gives immediate insight in network gaps
2.	How effective will the dashboard be as decision support tool for Sales?			x	effectiveness is high, this directly follows out of #1
3.	How easy is it to operate the dashboard?			x	pre-requisite is introduction
4.	Are the possibilities to slice and dice the information sufficient? (the possibilities that filter provide)			x	yes, more might be required if more data is present
5.	Is there a balance between providing quantitative and visualization insight?			x	Drill through options after visualisation
6.	Are there any other insights or feature you think could be added? Are there anything that you think could be improved?	scenario analysis → as discussed high impact level of impact of customers if they leave or increase			

KL