INDUSTRIAL ENGINEERING AND MANAGEMENT

BACHELOR THESIS

Optimizing Truck Routes: A Mathematical Model to Reduce Empty Driven Kilometres

A.S.A AL-ARIKI

March.2024

UNIVERSITY OF TWENTE.

Colophon

University of Twente,
Industrial Engineering and Management
Faculty of Behavior, Management, and Social Sciences

DATE
March.2024

SUPERVISORS
Ing. J.P.S Piest (First supervisor)
Dr. I. Seyran Topan (Second supervisor)
Pascal Esmeijer (Company supervisor)

VERSION
2

STATUS
FINAL

AUTHOR(S)
Amr Al-Ariki

TELEPHONE
+31614411947

EMAIL
A.S.A.AL-ARIKI@STUDENT.UTWENTE.NL

POSTAL ADDRESS
P.O. Box 217
7500 AE Enschede

WEBSITE
www.utwente.nl

FREE HEADER
Optimizing Truck Routes: A Mathematical Model to Reduce Empty Driven Kilometres

FILENAME
Bsc THESIS IEM

REGISTRATION DETAILS
Registration details

Management summary

This Thesis is conducted in Vos transportation group and under the supervision of the University of Twente. Vos Transportation Group is considered to be the one of most sustainable companies in the logistics sector in the Netherlands, therefore the success of this Thesis will aid in realising the company's goals and objectives. Problem identification shows that minimizing the empty driven kilometres can be achieved through providing the planners with a technical tool that calculates the empty driven kilometres using all possible marketplace websites. Therefore, the objective of this research is to develop a mathematical tool to minimize the empty driven kilometres for Vos Trucks using CRISP-DM methodology. The main objective of this research aims to answer the following researcher question.

"How can mathematical models support the planners of Vos transportation group with finding better loading destinations for using the marketplace in Finland and Denmark to decrease the empty driven kilometres from a percentage of 6% in Denmark and 18% in Finland to a percentage of 4,89% and 7,24% respectively."

Literature shows that minimizing empty driven kilometres should be divided into defining the problem formulation and selecting the solution method. The problem formulation revealed that this thesis problem is considered a variant of the vehicle routing problem; hence, after understanding the desired process. The problem is considered to be a single vehicle routing problem with capacity, time-window, and backhaul customers (SVRPCTWB). To solve the SVRPCTWB, the exact approach solution of mixed integer linear programming was chosen based on the knowledge of the researcher to delivering accurate results.

The stakeholders related to the project were identified based on their degree of involvement, then the current and desired processes were identified to construct the problem formulation and solution method. The desired model is considered the case study of the mixed integer linear programming. The desired model states that the company aims to minimize the empty driven kilometres when using the marketplace, constraining (filtering) the vehicle capacity of 13,6 loading meters (LDM), pick-up minimum load of 6 LDM, and pick-up time window of one day.

Afterwards, Vos 2022 data was cleaned to conduct the test with reliable data. Several tests were conducted to analyse the data. The first analysis showed that the actual empty driven kilometres average for Denmark is 8,04%, and the empty driven kilometres while using the marketplace is 4%. The analysis showed that the usage of the marketplace for Denmark and Finland was very low with a usage of 6,7% in Denmark and 1% in Finland. Therefore, only Denmark is considered in the evaluation due to the low amount of data Finland offered.

Evaluating the model with two tests showed positive. To begin, in the first evaluation test the planners' decisions that resulted with the 2022 empty driven kilometres and model performance that results in another 2022 empty driven kilometres were compared. The results showed that the model performance when using the marketplace resulted in 5,36% of empty driven kilometres, while the actual 2022 planners' performance when using the marketplace was 3,95%. The model resulted in higher empty driven kilometres, because the model calculates the empty driven kilometres for the stage of travelling to the pickup location in Denmark and the stage of dropping-off the picked load before returning to the depot, while the 2022 data only provided the empty driven kilometres when the truck is traveling to the pick-up location. Therefore, after calculating the model empty driven kilometres for the stage between the linehaul customer and the pick-up location, the model then resulted with 3% empty driven kilometres, showing in improvement when conducting a compassion of the same conditions.

The second evaluation test was experimental and involved the use of the developed model with live data and the planners. The empty driven kilometres of the model were lower with approximately an average of 53 kilometres than the planners' decisions. However, the computation time of the model was approximately one hour, while the planners searched for an average of 3 minutes. The main reason of having high computation time is due to the computation of the distances and times between every location and all other possible locations as MILP is considered to NP-hard. Though, the computation time of the model is high, several minor improvements can be approached to reduce the computation time, such as using the Euclidean distance method for computing the distances and time or storing the locations for every instance the model is ran and then use the stored locations for future instances.

Preface

Dear reader,

This Thesis endeavours to contribute to the fields of operational research and heuristics, with a special focus on enhancing transportation efficiency for Vos transportation group in Denmark. By delving into the complexities of reducing empty driven kilometres through mathematical modelling, hence addressing a practical industry challenge. This work is submitted in partial fulfilment of the requirements for the bachelor's degree of Industrial Engineering and Management at the University of Twente.

My fascination with operational research and heuristics stems from my past experiences in the study program of Industrial Engineering and Management, where I observed the potential transformative power of optimization in solving real-world problems. This Thesis is not only an academic endeavour but also a personal quest to understand how optimization can be applied at an industry level.

I would like to express my deepest gratitude to my primary supervisor, Ing Sebastian Piest, and secondary supervisor, Dr. Ipek Seyran Topan, for their invaluable guidance, patience, and support throughout this research. My thanks also go to the Vos company supervisor, Pascal Esmeijer, for their insightful comments and constant encouragement which I was lucky to receive from an industry veteran.

Special thanks to my family and friends for their unwavering support and understanding during this challenging yet rewarding journey.

This thesis is organized into nine main chapters. Chapter 1 introduces the research background, objectives, and significance. Chapter 2 provides an overview of the business structure from the stakeholders' point of view. Chapter 3 reviews the literature. Chapters 4 and 5 incorporate an analysis of the data. Chapter 6 highlights the construction, integration, and validation of the Mixed integer linear programming model. Chapter 7 involves an evaluation of the findings. Chapter 8 goes over the deployment plan and framework. The final chapter, chapter 9, discusses the implications of the findings and concludes with suggestions for future research.

The scope of this research is focused on the reduction of empty driven kilometres in Denmark. While every effort was made to cover all relevant aspects, limitations were encountered in data availability and time.

Ethical considerations were meticulously observed throughout the research process ensuring data confidentiality among other considerations.

It is my hope that this Thesis not only contributes to the existing body of knowledge on operational research and heuristics but also inspires further research in this area. The journey encapsulated in the following pages is one of discovery, challenge, and intellectual growth. I warmly invite readers to engage with this research, in hopes that it may inspire further innovation and inquiry in the domain.

Amr Al-Ariki,
March.2024

Table of Contents

List of Figures

List of Tables

List of Equations

MIXED INTEGER LINEAR PROGRAMMING

List of Abbreviations

CRISP-DM: Cross Industry Standard Process for Data Mining

ALNS: Adaptive large neighborhood search

MILP: Mixed integer linear programming

VRP: Vehicle routing problem

SVRPCTWB: Single vehicle routing problem with capacity, time-window, and backhaul customers

ELM: Electronic logistic marketplace

EDK: Empty driven kilometres

LDM: Loading metre

API: Application programming interface

1. Introduction

The first chapter introduces a brief introduction to the thesis. This involves establishing an understanding of the background and solution methodologies that will be utilized to address the action problem identified in this research plan. Starting from **Chapter 1.1**, the background behind the issues is introduced, along with the respective stakeholders in **Chapter 1.2**. This is followed by introducing the problem in **Chapter 1.3** in which the problem context. Hence, the problem is formulated through a why-why analysis and a cluster map illustrating the cause-and-effect relationship leading up to the core problem in **Chapter 1.3.2**. Furthermore, the research objective is stated in **Chapter 1.4** with the research reliability, validity, scope and limitations in **Chapter 1.4.1** and **Chapter 1.4.2.** In **Chapter 1.5** the selected research methodology approach will be employed to support the structure of the Thesis. the research question and sub-research questions are introduced in **Chapter 1.6**. Finally, the thesis structure is visualized in **Chapter 1.7**.

1.1 Background

Empty driven kilometres are considered non-value-added activities because the amount of time, money, and effort spent on transport service does not increase the service's worth to customers (Islam, 2019). Empty driven kilometres differ between the international and national transportation for freight trucks. The Eurostat analysis (Road freight transport by journey characteristics, 2023) shows the empty driven kilometres percentage for the European average and each country individually as shown in **Figure 1.1**. The statistics can be used to evaluate the companies empty driven kilometres in the region they are operating in. At the EU level, empty driven kilometres are considered one-fifth of the total driven kilometres.

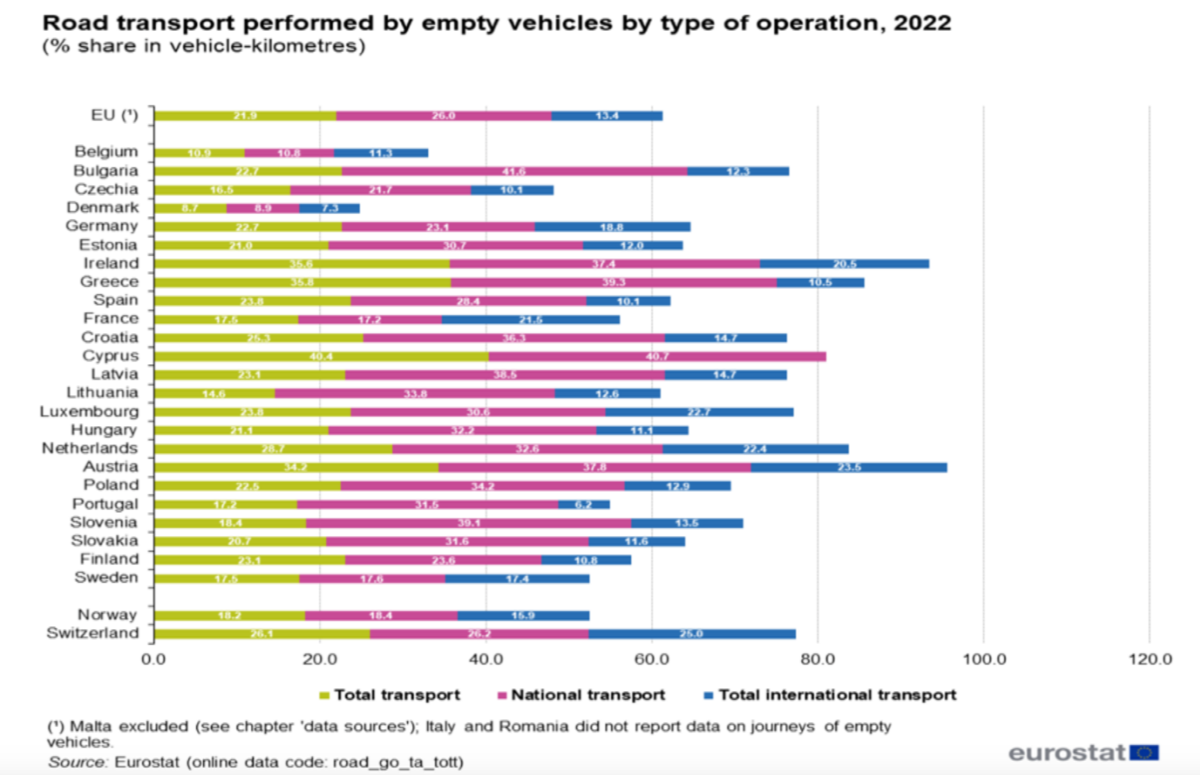

Figure 1.1 – Empty driven kilometres percentage in Europe (Road freight transport by journey characteristics, 2023)

1.2 Company

Vos Transportation Group is an international company in the field of transportation and distribution, owning more than 450 trucks. The company mainly transports shipments to Western and Northern Europe. Vos owns four offices in the Netherlands, with two cross-dock locations in Deventer and Ittervoort, a warehouse in Alblasserdam, and an Expedition office in Groningen. With 70 years of experience, Vos has an objective to lead the sustainability sector in Logistics, where the companys' slogan is "Leading the way in Sustainability" (Leading the Way in Sustainability, 2021).

1.3 Problem identification

Vos Transportation group in collaboration with the IEBIS department at the University of Twente has created a new project to adapt to the current technical generation. The project aids Vos to be directed into the red queen effect in being competitive and running harder and smarter than its' competitors (Voelpel et al., 2005). The red queen effect is a business concept that states that companies need to remain competitive in their respective sectors to avoid being outperformed by rivals. Thus, with recent advancements in the logistics sector, technology has played a vital role, which requires Vos to evolve with the competition.

Vos transports shipments internationally to their customers. To minimize empty driven kilometres and cost, Vos seeks on ensuring that trucks returning to the depot in the Netherlands are not empty. To achieve this goal, planner return the trucks to the Netherlands using two strategies. Firstly, the trucks are returned loaded from standard customers that are arranged by the planners. This strategy is mostly effective and mainly used as the first option. However, if the planners are unable to find a return shipment from the standard customer, the planners use the marketplace to search for a returning shipment.

When Vos planners use the marketplace, Vos planners are confronted with a vast number of shipments on the marketplace that they are required to select from. Therefore, the current process lacks the use of algorithmic or any systematic approach that could assist planners in selecting the best shipment for an empty truck. Hence, a high workload is required from planners. This issue is a consequence of the absence of support from the IT department. To substantiate that this problem is, in fact, the actual problem, an in-depth analysis is required. Thus, a Why-Why analysis is conducted, and a cluster map is constructed in **Chapter 1.3.2**.

1.3.1 Problem context

The key problem context associated to the empty driven kilometres in the logic sector are stated in the following point. specifically, the dynamics of the electronic logistics marketplaces is explored with their relation to the usage of marketplace in reducing the empty driven kilometres. While also stating the safety implication posed by empty driven trucks.

1) Electronic logistics marketplaces (ELM)

ELM is defined as an electronic hub using Web-based systems that link shippers and carriers together for the purpose of collaborating or trading (Wang et al., 2007). There are two types of ELM; open and closed (Wang et al., 2011). Open ELM is the only relevant type for this research, as it is used for spot sourcing, while close ELM is a restricted marketplace to only authorized members. This type of open ELM refers to the ad-hoc procurement of transporting services to only fulfil an urgent and short-term needs between the shipper and carrier (Kale et al., 2016). Open ELM is limited to basic load posting and matching services with the absence of the benefits of efficient searches and coordination of costs (Kale et al., 2016).

2) Empty driven kilometres danger

Empty driven kilometres are the number of kilometres driven by a truck with an empty trailer (Heilmann, 2020). Empty trucks can be dangerous in some situations. From a safety perspective, (Rescot, 2009) states that empty trucks should limit their speed to half the range compared to loaded trucks; this is due to the instability of the empty trucks. Another study that has been conducted in Texas highways indicates that in wet tracks, empty trucks are thrice as likely to be involved in an accident than loaded trucks (CHIRA-CHAVALA, 1968).

1.3.2 Core problem

To reach the core problem of the thesis and to find the root causes and action problem, a why-why analysis and a problem cluster are developed. These techniques aim to visualize the problem and to identify all perspectives and views.

Firstly, the Why-Why analysis is a method of questioning used to link the cause-and-effect relationships of possible problems, leading to the main goal of identifying the root cause of the problem (Theisens, 2021). Since, the initial problem is the high empty driven kilometres for the empty trucks in Denmark and Finland compared to the countries' average. The Why-Why analysis results shown in **Figure 1.2** has helped in finding the root cause of this problem. The root cause is the absence of technical support of the IT department in planning the routes for the empty trucks using the open electronic logistic marketplace (ELM). Finally, performing the why-why analysis substantiated the initially provided problem as the actual cause problem.

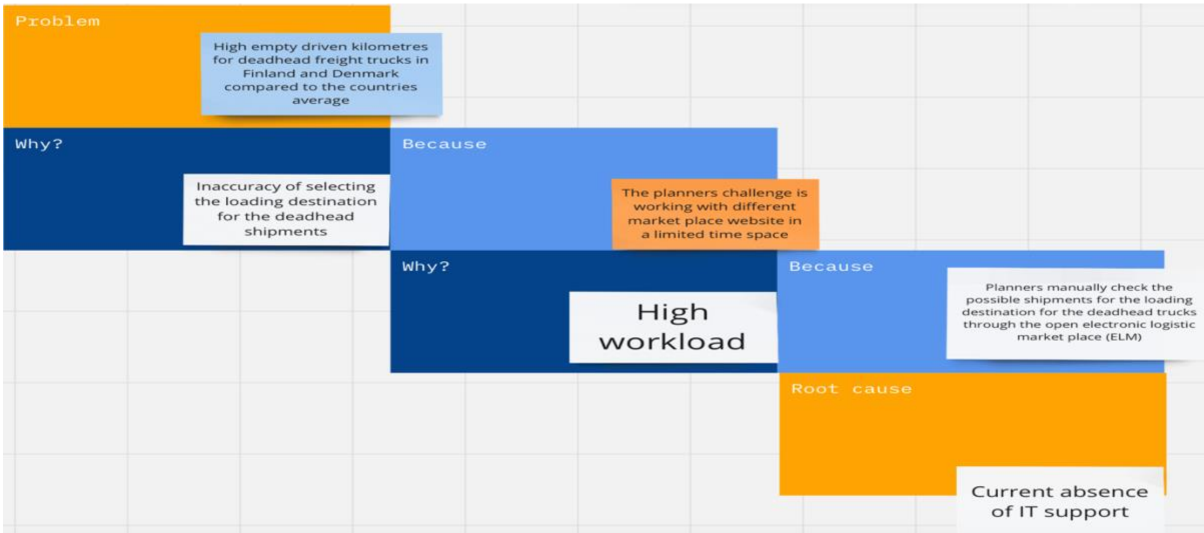

Figure 1.2 – why-why analysis

Secondly, a problem cluster was constructed to find the inventory of problems and to link them to the core problem. The action problem is the gap between the norm and reality, so solving the core problem will have a favourable outcome on the action problem. This implies that the action problem that Vos currently faces is the higher empty driven kilometres in both Denmark and Finland as compared to the European average in both countries in 2022 (Road freight transport by journey characteristics, 2023). In 2022 Vos' empty driven kilometres percentage in Denmark was 7.5%, while the European average in Denmark was 7.3%. In the same year, the percentage in Finland was 18,09%, while the European average in Finland was 10.8%. The norm and reality are visualized in **Table 1.1**.

Table 1.1 – 2022 Vos performance average versus European Average (Action Problem)

	Vos	European average
Finland	18,09%	10.8%
Denmark	7,5%	7.3%

The norm for Vos is to have a percentage equal or lower than the European average empty driven kilometres. The goal of the research is to verify if the absence of IT support in selecting the path for the empty trucks is the main reason for having high empty driven kilometres.

Different problems that cause high empty driven kilometres are visualized in the problem cluster found in **Figure 1.3**. To begin, Vos believes that the planners can contribute to inaccuracy when selecting the loading location from the marketplace, due to the fact that the planners do not search for all possible loading customers, which might result in better options. This results from the high workload on planners and the number of unused marketplace websites. To elaborate on both problems, the high workload originates from the limited time window between the marketplace customers offering the shipment, and the planner process of selecting and contacting the marketplace customers. Consequently, with the limited time window, the planner has insufficient time to search for many possible customers on several marketplace websites to find the optimal returning shipment based on minimizing the empty driven kilometres.

To conclude, the evidence presented thus far indicates that the core problem of this research is the sole dependency on manual work by the planners to accomplish tasks. These tasks include choosing the loading and unloading locations, negotiating the costs with the marketplace, and finally communicating to the drivers the respective loading locations they need to head to. Hence, the core problem is that the planners are manually selecting the loading locations for the empty trucks traveling in Denmark and Finland, and not being assisted by any technical mathematical system while making these decisions.

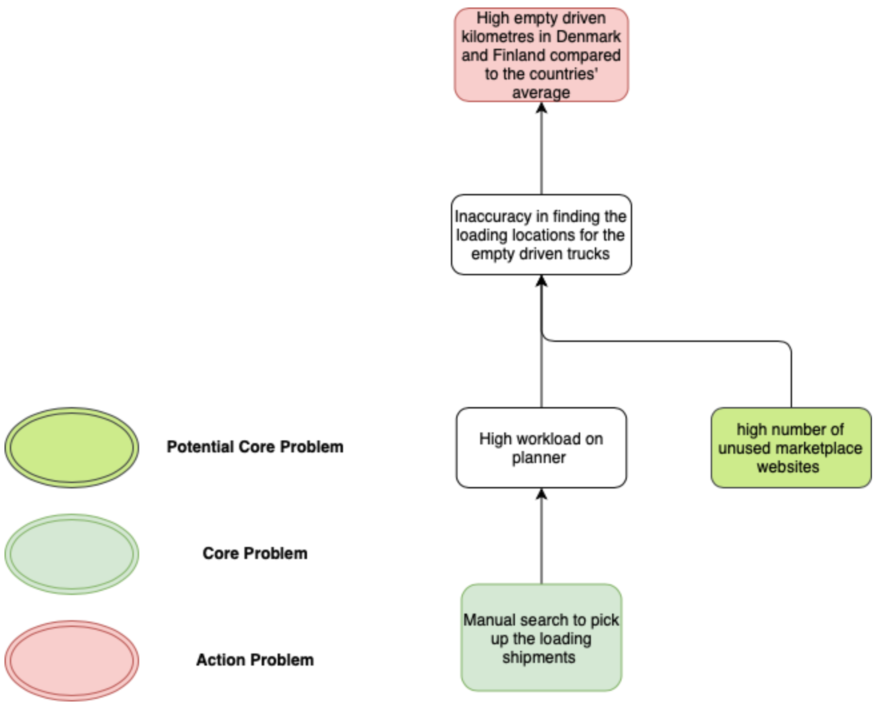

Figure 1.3 – Problem cluster

1.4 Research objectives

The noted problems demonstrate the necessity for constructing a technical tool to minimize the empty driven kilometres and increase efficiency of the planner's time. To develop a technical tool, the current process needs to be studied in a way to create a desired process that will improve the current process according to the company's needs. In this Thesis, the company's current and desired process showed that the problem relays on the vehicle routing problem variants. Using the literature, vehicle routing problem variants are indicated based on the constraints of the process. Therefore, the company process indicates that the company is facing a vehicle routing problem classified as a single vehicle routing problem with vehicle capacity, time constraint, and backhauls (SVRPCTWB). After conducting a systematic literature review, it has been found that the vehicle routing problem are solved using three approaches; exact, approximate, and hybrid approach. Each approach has its compensations as well as limitations, but since the exact approach provides the most accurate solutions it has been chosen to be constructed in this thesis. The research objective is to develop a mathematical model MILP and integrate the model into Python to build a prototype that search and selects the marketplace customers based on minimizing the empty driven kilometres for Vos truck. To accomplish the main research objective, historical data from Vos for the period of 2022 has been used to evaluate the performance of the model through comparing the empty driven kilometres generated by the decisions of the planners in 2022 and model decisions. To evaluate the model, marketplace customers data is scripted from the marketplace websites recent offers to feed the model with marketplace customer locations. However, to have a non-discriminatory comparison, the marketplace customer locations that the planners selected in 2022 will be compared to the scripted data set, using a map tool to check the distance difference between locations.

1.4.1 Reliability and Validity

Reliability and validity are key elements of successful research. As (Saunders et al., 2018) express the reliability and validity. Reliability "Extend to which data collection technique(s) will yield consistent findings", while Validity "Extent to which data collection method or methods accurately measure what they were intended to measure". The data used for the project will be acquired from the Vos planning database for the period of 2022, and the marketplace customers to feed the model will be scripted in January 2024. The 2022 period needs to be investigated due to the war in Ukraine and check if that affected the transportation sector in both Finland and Denmark, and if there are any Covid-19 influences in 2022 in both countries effecting either transportation or manufacturing sector. Both data collection will result in big data to ensure reliability and real-time evaluation data to ensure validity. Data cleaning methods and the use of Excel will be conducted to identify missing data and inconsistent from the company's 2022 big data.

1.4.2 Scope and limitations

The scope of the research involves the selection of loading locations for the empty trucks operated by Vos using the marketplace while addressing the issue of high empty driven kilometres for Vos trucks in both Denmark and Finland compared with these countries' empty driven kilometres average as issued by the Eurostat (Road freight transport by journey characteristics, 2023). The research's goal is to minimize the empty driven kilometres distance to assist the planners in selecting the loading locations for the empty trucks, henceforth decreasing the time spent searching for rides on the open electronic logistic marketplace, this will be achieved using the Vos IT-department resources.

To conduct the research in an efficient way, a researcher needs to understand the limitations of the project. The project goal as mentioned is to decrease the empty driven kilometres for Vos trucks by developing technical mathematical models to improve the results. Evaluating the results of the project is part of the limitations of the research, since the project prototype should be compared with decisions of the planners in the same timeline, though, the plan to collect some data to pattern the direction the results, however, the project is timeline is around 10 weeks, so the limited time constraint limit the project from prototyping and evolution. Secondly, Vos currently owns 3 warehouses in the Netherlands, however, to test the functionality of the developed model. The research will only focus on the result of one warehouse to decrease the complexity of data collection from different warehouses.

1.5 Research methodology

After considering different methodology cycles such as MPSM, DMAIC, and CRISP-DM, the CRIPS-DM methodology cycle was chosen in this research. The reason for this choice, is because CRISP-DM is the finest methodology in data mining and data sciences projects. CRISP-DM runs through six phases, which are business understanding, data understanding, data preparation, modelling, evaluation, and deployment. The key is to understand where the research will be conducted in each phase. The deliverables of each phase of the CRISP-DM are shown in **Figure 1.4.**

Cross-Industry Standard Process for Data Mining Deliverables					
Business understanding	Data understanding	Data preparation	Modelling	Evaluation	Deployment "Plan"
• Stakeholders identification • Current process investigation • Construct the desired process and select the KPIs to evaluate it	• Extract the data sets • Determine the quality of the data sets • Explore the data	• Clean the data • Build an entity relation diagram to connect the data sets • conduct the required analysis on the cleaned data	• Construct the selected model • Integrate the model in Python • Validate the model using a scenario-based approach	• Evaluate the selected KPIs from the business identification section • Conduct experimental testing between the constructed model and the planners	• Formulate a deployment plan using Scrum framework and integrate SMART framework in the Scrum sprints

Figure 1.4 – CRISP-DM deliverables

1.6 Research questions

To gain a better understanding of the main core problem and find the appropriate literature, analyse the solution methods, find the set of solutions, and then select the best one. The process starts with defining the research question of this Thesis. In this case, the main research question is based on technical models that are developed to assist in building a unique model to optimize the empty freight shipments for Vos when using the marketplace. This method will support the planners in decision-making, while decreasing Vos empty driven kilometres in Finland and Denmark to be in the range of the countrys' total international empty driven kilometres percentage.

"How can technical models support the planners of Vos transportation group with finding better loading destinations for the empty freight trucks in Finland and Denmark to minimze the empty driven kilometres from a percentage of 7,5% in Denmark and 18,09% in Finland to a percentage of 7,3% and 10,8% respectively."

To support the main research question, several sub-questions and sub-sub-research questions are developed as the following:

1. **What are the current decision-making criteria for selecting the loading location in marketplaces?**
    a. **Which marketplaces are available to script customers for both Finland and Denmark?**
2. **Which KPI's can be used to evaluate the new model?**
    a. **How does the performance of the new model affect the KPI's?**
3. **Which criteria are necessary when selecting new models?**
4. **Which Vehicle routing problem variant in literature contest with the company's process?**
5. **Which solution approaches are in literature that can be used in optimizing routes?**
    a. **What is the most effective solution approach to solve vehicle routing problems and decrease empty driven kilometres?**
6. **What available data will be used in the mathematical model?**
    a. **How can cleaning the data ensure the validity of the data?**
7. **What is the characteristic of the solution model?**
8. **How can a Scrum framework assist in the deployment plan?**

1.7 Report structure

As shown in CRISP-DM deliverables, the report structure will assess all deliverables by assigning the research sub-questions to each deliverable of the CRISP-DM. **Table 1.2** shows how every sub-research question is allocated into the CRISP-DM phases and in which chapters of the Thesis the assessments of the sub-research questions can be found. In chapter 9, the conclusion of the thesis is stated, discussing the results of each sub-research question and stating the limitations, recommendations, and further projects of this thesis.

Table 1.2 – Report structure

Research sub-question	Phase	Chapter
1. What are the current decision-making criteria for selecting the loading location in the marketplaces?	Business understanding	Chapter 2
2. Which KPI's can be used to evaluate the new model?	Business understanding	Chapter 2
3. Which criteria are necessary when selecting new models?	Business understanding	Chapter 2
4. Which Vehicle routing problem variant in literature contest with the company process?	Literature review	Chapter 3
5. Which solution approach that are in literature can be used in optimizing routes?	Literature review	Chapter 3
5a. What is the most effective solution approach to solve vehicle routing problems and decrease empty driven kilometres?	Literature review	Chapter 3
1a. Which marketplaces are available to script customers for both Finland and Denmark?	Data understanding	Chapter 4
6. What available data will be used in the mathematical model	Data understanding	Chapter 4
6a. How can cleaning the data ensure the validity of the data	Data preparation	Chapter 5

7. What is the characteristic of the solution model?	Modelling	Chapter 6
2a. How does the performance of the new model affect the KPI's?	Evaluation	Chapter 7
8. How can a Scrum methodology assist in the deployment plan?	Deployment	Chapter 8

2. Business understanding

According to (Kruse et al., 2021), the business situation is performed to get an overview of the required resources that will be used in the project; determining the goals and objectives of the project is one of the most important aspects in this phase. In the phase of understanding the goals and objectives of the company, it is necessary to discuss with the stakeholders; identified in **Chapter 2.1**, how the current process is performed as shown in **Chapter 2.2**. Moreover, to have a good understanding of the problem, modelling the desired model helps in understating the problem and identifying the bottlenecks in the process in **Chapter 2.3**.

2.1 Stakeholders

The stakeholders related to the project are the operational department representing the planners, IT-department, company management, and financial department. Recently, Vos has been working on the collaboration between the IT-department and the operational department to integrate technology in developing the performance of the company KPIs related to sustainability.

Firstly, to shed light on positive impact on the planners, If the project goals are achieved, the planners' issues as mentioned in the problem cluster will be resolved. Therefore, planners will have an efficient and productive workflow in the selection of loading locations for the company deadhead trucks, subsequently reducing search time spent in this phase. Currently some planners work overtime and on breaktimes searching for loading locations due to the competitive marketplace environment, given that the objectives of the project are attained, work outside working hours will be significantly reduced. The unfavourable outcomes resulting from this project will be experienced in the long term, with the involvement of machine learning and integrating the technology in the planning departments of logistics companies, this may lead to full dependence on machine learning algorithms to fulfil planning needs, which could subsequently cause loss of jobs. As mentioned, this is a problem for the long term, however, at the current time, planners' hold an integral role in terms of negotiations with relevant parties during the selection of loading locations, hence, solidifying their positions within logistics companies. To conclude, the outcome of this research will assist the planners in streamlining the selection process, ultimately saving time, and empowering them during negotiations.

Furthermore, the parties that are affected directly and indirectly are the IT-department, financial department, and company management. Firstly, the IT-department will need to implement and maintain the project. The project has considerable expansion opportunities into different functional operations of the company. Therefore, the IT-department can grow and employee more technicians for more projects. Secondly, from the financial aspect, the project aims to reduce the empty driven kilometres which will directly affect the transportation cost leading to savings in the financial department. As mentioned, extra technicians may need to be hired in the IT-department, which will require financing. Finally, the company management will be able to achieve their sustainability objectives allowing for a stronger alignment with the company's mission statement.

2.2 Current process

The planners' greatest challenge in the current route planning process is mostly caused by the limited time frame between the time a customer posts a request on the marketplace and the time at which the empty driven truck needs to be at the loading posting, since customers usually post last minute requests. This results in an uncertainty when choosing the best routes for the empty driven trucks since fast decision-making has significance. The daily route planning process, which is the same for planners in Denmark and Finland, starts with login, time consuming filtering of the marketplace offers and impromptu decision-making. Once requests fitting the search criteria are found, emails are sent to the respective customers. When a response is received, an offer is negotiated and if both parties agree to the terms, the necessary information are entered into the Vos system. This information can later be viewed in the driver's dashboard. Usually, the planners must work through this process in one to two days, because marketplace requests are posted one to two days before the required day of transportation. In addition to, a quantity of approximately 300 offers in Denmark and 50 offers in Finland in the several marketplace websites. The current process is shown in **Figure 2.1**.

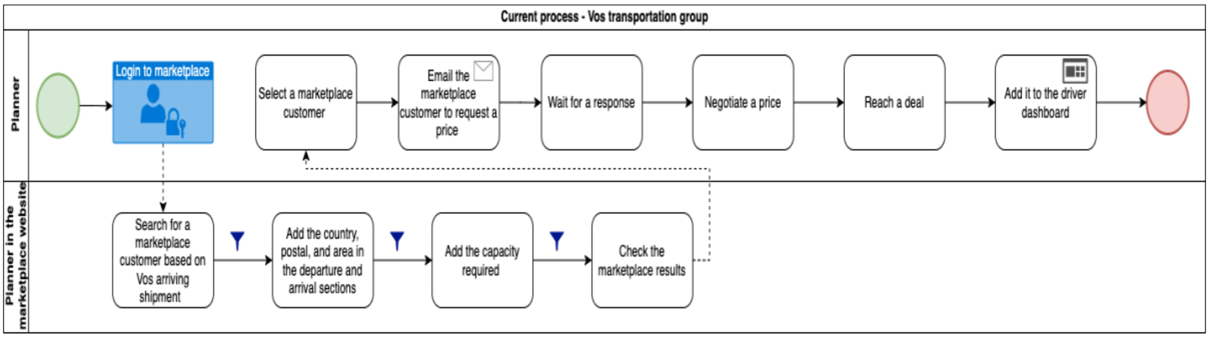

Figure 2.1 – Current process

Vos planners select transportation paths for the empty trucks mainly based on cost. The target of the company is to at least have 30% cost return from the truck heading back to the depot and a 70% cost return from the truck heading to the unloading location (main customer). This target increases the overall opportunity cost when it comes to optimizing the cash return through selecting a better loading location for the empty trucks, as planners can opt for a loading location that would result in a lower return although a better alternative was available.

After several discussions with the operational management, it is believed that the planners struggle with the current process is caused by the highly competitive environment of the marketplace. Therefore, if an offer is not chosen quickly, the planner risks to lose that offer to a rival company. This is why Vos planners mostly work during break-times and sometimes even after normal working hours to search for marketplace customers for the empty driven trucks. The lack of technical support and relying solely on the planner's experience in selecting these routes amplifies the issue. Although, the planners are required to search in all the marketplaces provided by the IT department, only a single marketplace website is usually used when searching for loading locations. Even though, using more marketplace websites would be considered best practice since it usually results in loading locations that serves the interests of the company. Using multiple websites requires a certain level of familiarization that ensures a fluent experience, however, from the planner's perspective, this is a time-consuming process.

2.3 Desired process

The IT department provides no support in the daily decision-making of the route planning process, this is substantiated by the absence of technical approach in selecting routes for the empty driven trucks when using the marketplace. Such support could decrease the workload of the planners and present the planners with a level of certainty that could guarantee an ideal route selection process. Thus, an issue in the current system is highlighted through the planners' challenges caused by the need of impromptu decision-making in the face of numerous marketplaces offers. Consequently, less favourable loading locations are selected, primarily because of the high workload and the aforementioned numerous offers that are found on multiple websites.

The desired process should include a mathematical model that will assist the planners with finding the marketplace customers from the marketplace website and will differ from the current process by excluding the inefficient time the planners use to search for marketplace.

The desired process focuses on building a mathematical model and integrate the model using Python to search and select marketplace customers based on minimizing empty driven kilometres. Thus, the desired process will start with the planner entering the shipment ID that is stored in the system, afterwards, the model will search for all possible marketplace customer based on their pickup and drop-off locations. The model then will select the marketplace customer and will visualize the choice to the planners, so the planner can start contacting the marketplace customer. The desired process is shown in **Figure 2.2**.

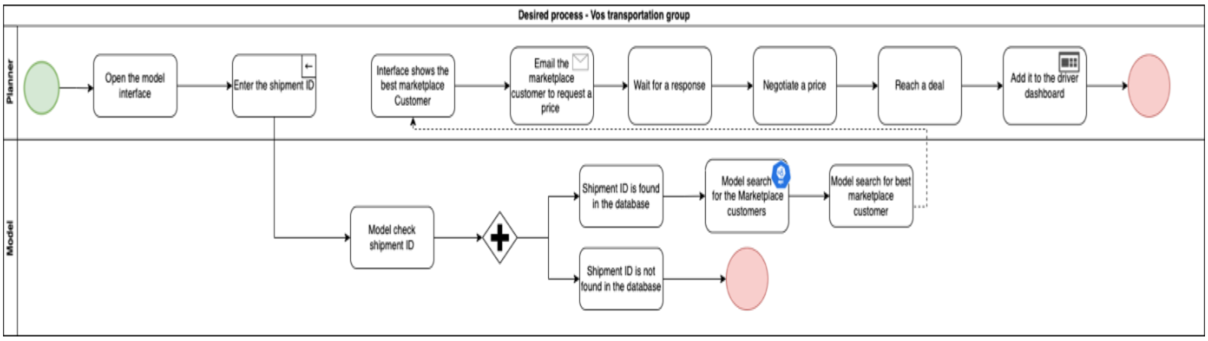

Figure 2.2 – Desired process

The model should have several criterions to function and make decisions in the same way the planners want the system to work. After the discussion with the operational and IT team, the following constraints should be included while constructing the new model.

To begin, Vos travels to Scandinavia using a standard truck with a maximum capacity of 13,6 loading meters (LDM) and a minimum pickup load of six loading meters (LDM) from the marketplace. Furthermore, several routing constraints should be implemented, as the goal is to minimize the empty driven kilometres trading of the total driven kilometres; the company requested to have a radius constraint for when the truck is empty, and that is after delivering to the linehaul standard customer and after dropping off the final shipment before returning to the depot. Therefore, the radius to search for a backhaul location is inside 200 kilometres, and 100 kilometres radius that is between the last drop-off and the depot.

Finally, constraining the time is divided into three steps. Firstly, is to consider the arrival of the truck to the linehaul standard customer, and then check the pickup times of the marketplace customer location. The time travel between the linehaul to the backhaul is then calculated using Google Maps API and if the arrival time to the backhaul is inside the time window, then the shipment will be considered. Overall, all pickup shipments should be accomplished in one to two days, considering the resting time of the drivers. Secondly, when constraining the time, the loading time of one hour and unloading time of half an hour must be considered for all stops. Finally, the resting time should be considered, taking into circumstances the previous driver trips and if there are one or two drivers in the truck.

However, historical data does not include times, and the marketplace customer doesn't often include times with the offers; therefore, for the model prototype, the days of delivery and pickups are considered with the exception of the time. Thus, an assumption of the arriving time at the linehaul customer is at 08:00 am on a particular day and all pickups must picked up at the same day, also assuming that the driver is valid for the whole day and has rested before the delivery. Loading and unloading times will be implemented in the prototype model.

The company defined several KPIs to evaluate the performance of the model. The first KPI is linked to the main research question of this thesis, which is the evaluation between the norm and reality of 2022, and the second KPI is the computation time of the model. Furthermore, an experimental analysis will be conducted between the planners and the model prototype to evaluate the performance of both based on the computation time and the empty driven kilometres. The analysis will start with providing the planner with several shipment IDs and needs to optimize the routes using the scripted marketplace customers. Afterward, the model will search for the same shipment ID's, and the results of both will be evaluated. The goal is to compare the computation time and the empty driven kilometres when the model and the planners have the same marketplace customers. Currently, the IT team wants the computation time to be less than one minute, believing that the planners take more time when searching for the shipment.

2.4 conclusion

In this chapter, the stakeholders related to the thesis were introduced, and then the current process was analysed. Finally, the desired process was developed based on the discussion with the stakeholders. The stakeholders related to this Thesis are the operational department representing the planners, the IT department, company management, and the financial department. Each stakeholder will be directly affected by deploying the desired process. The current process showed the planners' high workload when using the marketplace, which results in uncertainty in their decisions in selecting marketplace customers, due to the high number of marketplace customers. The desired process goal is to minimize the empty driven kilometres when using the marketplace by building a solution approach to search and select the optimal marketplace customer location. The desired process also provided all the necessary information to build a case study that will be used to construct the solution approach.

3. Literature review

In this chapter, a literature review is formulated to gain the knowledge necessary to assess the thesis goals. **Chapter 3.1** search for the different vehicle routing problem variants that can result in the problem formulation to answer the sub-research question “Which Vehicle routing problem variant in literature contest with the company process?”. **Chapter 3.2** search for the solutions approaches to solve the problem formulation identified in Chapter 3.1 and answer the sub-research question “Which solution approach that are in literature can be used in optimizing routes?”. Finally, **Chapter 3.3** connects both the problem formulation and the solution method to assess the sub-research question “What is the most effective solution approach to solve vehicle routing problems and decrease empty driven kilometres?”

3.1 Problem Formulation

The vehicle routing problem is one of the most studied combinatorial optimization problems with many variants and extensions, the goal of VRP is to find a set of routes that starts from the depot and ends in the depots and assign vehicles to route the customers and fulfil their needs, therefore, the objective of VRP is to minimize the objective function of the research (Kabadurmus et al., 2023). An example of the objective functions can be minimizing the route cost, all transportation costs, distance travelled, number of vehicles needed to cover all customers, and the overrun (Sitek et al., 2021).

VRPTW is the Vehicle routing problem with a time window. In VRPTW each customer has a time window in which its demand is to be delivered. The definition of VRPTW defined by (Ursani, 2011) states that a problem of finding a minimum cost route for several homogeneous vehicles stationed at a depot, which have the task of delivering goods to several customers with different rules. The first rule is that the route of the vehicle must start and end at a depot. The second rule is each customer must be served once by only one vehicle satisfying all customer demands. Finally, the truck capacity should be lower or equal to the demands when arriving at the customer inside the time window.

CVRP refers to the capacitated vehicle route problem. CVRP is when a fleet of delivery vehicles with a uniform capacity goal is to fulfil the customer demands from a single depot with a minimum transit cost or total distance travelled (Li et al., 2019). The properties of the CVRP are a distribution centre (depot), a number of geographical random distribution delivery points, and vehicles from the depot to serve the customer’s demand (Ni et al., 2023). To find the best route to travel while reducing the delivery time and cost, three conditions must be met. Firstly, the maximum load of each vehicle must be greater or equal to the maximum demand of the delivery route. Secondly, the maximum distance that each vehicle can travel must not be less than the maximum distance required for each delivery route. Finally, there is only one delivery vehicle at the distribution centre (Ni et al., 2023).

The vehicle routing problem with backhauls involves two types of customers. The first type is the base deliveries (linehauls), and the second type is the pickups (backhauls). Classic additional constraints enhanced in the VRPB are each vehicle must perform all deliveries satisfying the demand before making any pickups, routes with only backhauls are disallowed, however routes with only linehauls are allowed (Wassan et al., 2016). The VRPB is considered to be more valuable than the classical VRP in terms of cost-saving and placing fewer vehicles on the roads. The MT-VRPB is the multi-tour VRPB, therefore, vehicles can perform several trips within a given time constraint. The objective of the MT-VRPB is to minimize the total travel distance while minimizing the total cost with characteristics of a given set of customers divided into two subsets of delivery and pickups, a homogenous fleet of vehicles with a property that a vehicle may perform more than one trip in a single planning period, and finally all delivery customer are served before any pickups (Wassan et al., 2016).

SVRPTWCB is a combination of several variants; explained above; of the VRP problems and symbolizes the vehicle routing problem with a single vehicle having a specific capacity, where the problem has time window for each node and include backhaul customers. SVRPTWCB is chosen based on the desired process of the company.

As SVRPTWCB is a variant of the VRP, then SVRPTWCB can be solved using either the exact, approximate, or hybrid approach. Exact approach includes mainly the mixed integer linear programming, approximate approach uses either heuristics or Meta-heuristics, and finally the hybrid approach focus on integrating more than one method to solve the VRP.

3.2 Model solution

The model solutions involve optimization models that will be used to address and optimize the SVRPTWCB. The model solution will consist of three main approaches, which are the exact approach, approximate approach, and hybrid approach. Below each approach, as shown in **Figure 3.1** each model will be evaluated explicitly.

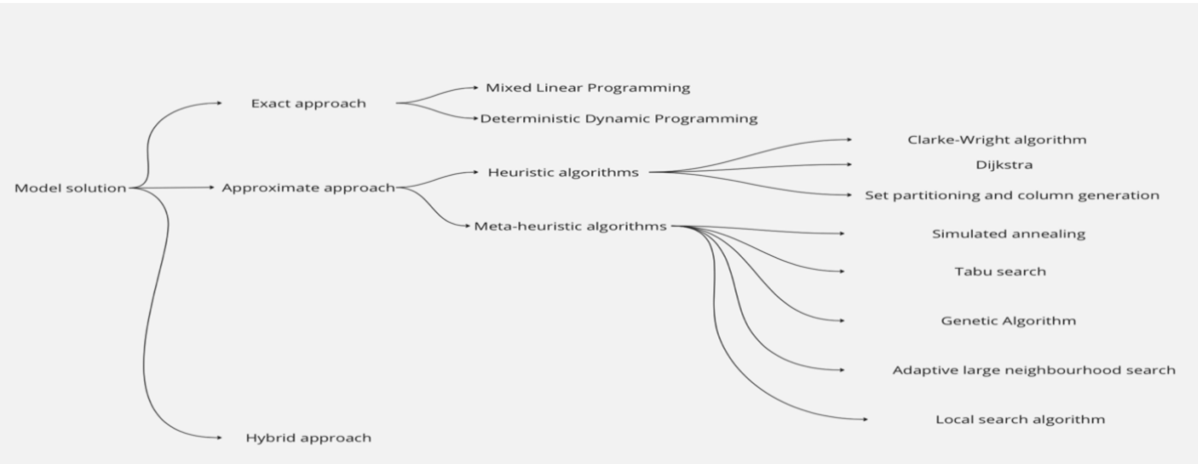

Figure 3.1 – Model solution framework

3.2.1 Exact approach

The exact approach can find the optimal solution for a problem using mathematical laws or data structure search. The main exact approaches used for the VRP variants are Mixed integer programming (MIP), and the Deterministic Dynamic Programming method. For small-size complex problems, the exact algorithm approach has the power to find the optimal solution in a reasonable time. However, when the size of the complex problem is large, finding the optimal solution requires a high amount of time (Ye et al., 2022).

Firstly, the MIP is based on Linear Programming with slight differences. Linear programming (LP) is a tool for optimization problems and has been used to solve optimization problems in several industries which including the logistic industry. To start building Linear programming, a decision variable needs to be defined, the decision variables need to address all decisions that need to be made toward the case study. Secondly, the objective function needs to be formulated, the decision-makers need to either maximize or minimize the function of the decision variables. The function that needs to be maximized or minimized is called the objective function. Finally, constraints need to be built to restrict the model to the case study. the differences between LP, MIP, and integer programming are as the following. To begin, all mathematical models serve the same systematic procedure. IP is an LP in which all or some variables need to be a non-negative integer and an IP in which only some of the variables need to be integer is called a mixed integer programming. Moreover, additional more complex constraints need to be added to the IP (Winston, 2004).

Secondly, dynamic programming is used to solve more than one optimization problem. Most dynamic programming techniques find the solution by breaking the large problem into smaller problems and starting the problem from the end into the beginning. Dynamic programming is mostly used in solving networks, inventory, and resource allocation problems. Through the network problem, dynamic programming is used to find either the shortest or longest path, depending on the case study, that links two points to build a route in a given network. To build a model using dynamic programming, different characteristics need to be identified. The model should consist of the stage, state, decision, principle of optimality, and the recursion formula (Winston, 2004).

3.2.2 Approximate approach

The approximate approach is one of the models' solutions approaches. The approximate approach is divided into the traditional heuristic algorithms and the Meta heuristic algorithms. The traditional heuristics are mainly into constructive heuristics, while meta heuristics differs by accepting non-feasible solutions or non-optimal solutions.

3.2.2.1 Heuristic algorithms

The heuristic algorithms refer to the traditional heuristics which involves the constructive heuristic. The constructive heuristics include the Clarke-Wright, Dijkstra, and finally Set partitioning and column generation (Ye et al., 2022).

Firstly, the Clarke-Wright algorithm objective is to merge tours together based on the expected savings in either distance travelled or cost. Based on the theory from (Wegen et al.,2017), the algorithm used in Clarke-Wright is as follows. Begin with constructing N tours from the Depot to the customer and back to the depot, and in every tour a customer demand must be supplied. Formerly, calculate the possible saving using the succeeding formula shown in equation 3.1 for all routes of i and j. The previous formula works as if s_{ij} is larger than zero without violating any constraints of the case study, then the saving algorithm of Clarke-Wright tries to merge two tour routes and search for the highest saving. After calculating the possible saving, the algorithm picks the largest possible saving based on the saving formula and the constraints. Finally, the algorithm repeats itself until there is no possibility for any savings (Wegen et al., 2017).

$$(\boldsymbol{S_{ij}} = \boldsymbol{C_{Depoti}} + \boldsymbol{C_{jDepot}} - \boldsymbol{C_{ij}}) \qquad \textit{Equation 3.1, (Wegen et al.,2017)}$$

Secondly, Dijkstra's algorithm was invented by the Dutch computer scientist Edsger Dijkstra. The algorithm runs to calculate the shortest route from one node to all other nodes. The key attribute for Dijkstra is the continuous expansion from the starting point following all nodes till they reach the endpoint (Ye, et al., 2022).

Finally, partitioning and column generation are heuristic algorithm, and from the book of (Wegen et al., 2017), performing only partitioning will result in the perfect route for the truck, as the goal of set partitioning is to partition a set of customers into a number of groups, such that every group corresponds with one tour (Wegen et al., 2017), to elaborate, the system goes through the whole possible tour for one truck and choose the best route and preform the same process for every shipment. The disadvantage for this algorithm is that it requires enormous amount of time for the computer to perform the process, thus, column generation is used to reduce the computation time to perform the process. The column generation is making pre-selection of tours that are done under the method of enumeration, algorithmic selection of likable tours, and heuristic constructive algorithms. The goal and advantages of column generation is that it decreases the set of selection, with the restriction that the developer wants to add.

3.2.2.2 Meta-heuristic algorithms

Meta-heuristics as (Ye et al., 2022) state ''setting up the initial solution of the problem, the solution is appropriately disturbed by using disturbance factors so that the problem's solution is continuously optimized, and finally, a satisfactory solution to the problem is obtained''. Meta-heuristics algorithms consist of the simulated annealing, Tabu search, genetic algorithm, Adaptive large neighbourhood search, and Local search.

Firstly, the simulated annealing is a stochastic algorithm aimed at finding the global optimum. The probability of the algorithm changes according to the quality of the move (Uddin et al., 2023). To have a better understanding of the simulated annealing, the book of Planning and Scheduling, (Pinedo, 2009) clarifies the simulated annealing. The algorithm goes through several iterations, in iteration X, the model will have a current solution S_X and the optimal solution found in the whole process till iteration X and called S_0. $G(S_k)$ and $G(S_0)$ refer to the corresponding values of the objective function. Performing a search for several solutions in iteration X through the neighborhood of S_X, and exterminated solutions are called the candidate solutions (S_C). If S_C is selected from the neighborhood, the selection can be done through a random or organized process. The process of selection is as follows, if $G(S_c) < G(S_x)$, then a move is made and $s_{x+1} = S_c$. If $G(S_c) < G(S_0)$, S_0 is adjusted to equal the S_c. Finally, if $G(S_c) \geq G(S_x)$. Then, in this stage probability for acceptance and rejection arrives, worse solutions are allowed in this stage and the reasoning is to give an opportunity to move away from the local minimum and find a better solution in the following iterations. A move to the following S_C is made with a probability of equation 3.2, and the rejection is made with a probability of $(1 - P(S_x, S_c))$.

$$P(S_x, S_c) = exp\left(\frac{G(S_x)-G(S_c)}{\beta_x}\right);$$ *Equation 3.2, (Pinedo, 2009)*

Secondly, tabu search shares the same procedure as the simulated annealing, but with one major difference. As stated before, tabu search differs from the simulated annealing by the selection of the solution from the neighbourhood of the current solution using deterministic algorithms instead of stochastic. The tabu tries to prevent the search from going to previously visited solutions (Pinedo, 2009).

Thirdly, the genetic algorithm is also similar to both simulated annealing (SA) and tabu search, though, SA and tabu search are indicated as special cases of the genetic algorithm with a population size of 1. The genetic algorithm has a special neighbourhood; thus, the neighbourhood is a collection of different solutions. The design of the neighbourhood goes through the construction of combined parts of different solutions within the data set. The advantage of a genetic algorithm is the easy coding; however, the genetic algorithm requires a high amount of computation time (Pinedo, 2009).

Fourthly, the adaptive large neighbourhood search (ALNS) has been used in many solution models for the VRP's due to their simplicity in adapting to new problems (Sacramento et al., 2019). ALNS is an extension of the large neighbourhood search (LNS). So, the LNS model is to continuously improve the initial solution and repeatedly destroy and repair the current solution. The ALNS uses many destroy and repair approaches that are gathered through statistical tests according to the performance of the model. The use of the destroy method is to eliminate part of the current solution in the model, while the repair method, uses stochastic to avoid constructing the same solutions more than once. The algorithm is firstly built using pseudo-code and the elapsed time to stop the algorithm is measured through the CPU time.

Finally, (Luo et al., 2021) discuss the use of local search. Local search algorithm is used for solving combinatorial optimization problems. Local search uses local moves through the candidate solutions to find a better solution. Local search operates using three classical improving operators explicitly 2-opt operator, 2-opt move operator and relocate moves which are used to generate new solutions randomly. The three operators are widely used in VRP problems. The local search stops when either a better solution is used compared to the input solution or when the maximum stated iterations are reached. Moreover, (Luo et al., 2021) define the three operators as the following:

- 2-opt operator: randomly select two shipments in one route and reverse the path between the two shipments while including the two shipments.
- 2-opt move operator: randomly select two paths from two different routes, then swap the two paths between two routes.
- Relocate operator: randomly select one shipment from one route, then insert the shipment to another route.

3.2.3 Hybrid approach

A hybrid approach is the combination of both the exact approach and the approximate approach or the combination of both two heuristic algorithms to gain in improving the solution quality and decrease the computation time by having an efficient model. (Sun et al., 2020) shows the hybrid use of LP and Tabu search and (Hong et al., 2023) shows the hybrid algorithm based on the ALNS heuristics and the use of MILP.

3.3 SVRPTWCB and MILP

After understanding the different types of VRPs with the model formulation, the company's desired process requires vehicle capacity, backhaul shipments, and time window, and the transportation is driven in one homogenous truck. Therefore, the company case is a variant of the vehicle routing problem and can be stated as single vehicle routing problem with capacity, time window, and backhaul (SVRPCTWB). The goal is to minimize the empty driven kilometres using the mixed integer linear programming, as the MILP is a study of operational research, and the researcher has knowledge in this area. However, the most efficient approach according to the literature is the adaptive large neighbourhood search (ALNS); as the ALNS can provide very close solutions to the MILP, while also having a very fast computation time.

Finding a SVRPTWCB problem that is similar to this thesis problem was not possible. Therefore, the goal is to find a problem that aims to minimizing the travelled distance and in this thesis project, the transportation cycle will be divided into different stages and each stage will have its own constraints. Azi et al., 2006 works with a VRPTW problem, VRPTW differs from this thesis problem. However, the variants of the VRP differ in the constraints of the MILP. Hence, the VRP main constraints that are implemented to function the model is integrated in most of the VRP problems. (Azi et al., 2006) problem formulation is a representation of the MILP model, the problem is defined as a directed graph that represents the nodes and arcs. Hence, **Table 3.1** show the parameters of (Azi et al., 2006) problem. As stated, the problem is VRPTW, however, there is a constraint limiting the vehicle to a specific capacity.

Table 3.1 - Literature MILP problem formulation sets, parameters, and variables (Azi et al., 2006)

Sets	
N	Set representing the customers N = {1, 2,...,n}
A	Set representing every arc {i,j}
K	Set of routes, K = {1,2, ..., K}
Parameters	
$\boldsymbol{t_{ij}}$	Travel time for arc i to j
$\boldsymbol{d_{ij}}$	Distance for arc i to j
$\boldsymbol{Q}$	Vehicle capacity
Variables	
$\boldsymbol{x_{ij}^r}$	Binary variable 1 if arc ij is used in route r, 0 otherwise
$\boldsymbol{y_i^r}$	Binary variable: 1 if customer i is in route r, 0 otherwise

Afterward, several constraints and the objective function that are necessary to function the model are shown below, equations 3.4 to 3.8 are considered to be subtour elimination constraints to ensure linear inequality. equation 3.3 is the objective function and aims to minimize the distance travelled between the arc ij when the binary variable x_{ij} equals one. Moreover, equation 3.4 represents the first constraint, indicating that if an arc exists then i must be visited, creating a relation between the node visits and the arcs taken. Equation 3.5 states that every customer should be visited exactly once. Furthermore, equations 3.6 to 3.8 show the flow conservation constraints that describes the vehicle path and ensure infeasibility. Equation 3.9 states that the total demand on a route should not exceed the vehicle capacity. Finally, equation 3.10 ensures that the arrival time to state i plus the service time and the time travel should be smaller or equal to the arrival time to state j.

$$min \sum_{r \in K} \sum_{(i,j) \in A} d_{ij} x_{ij}^r$$ *Equation 3.3, (Azi et al., 2006)*

S.t.

$$\sum_{j \in N^+} x_{ij}^r = y_i^r, \ \ i \in N, \ \ r \in K,$$ *Equation 3.4, (Azi et al., 2006)*

$$\sum_{r \in k} y_i^r = 1, \ \ i \in N$$ *Equation 3.5, (Azi et al., 2006)*

$$\sum_{i \in N^+} x_{ih}^r - \sum_{j \in N^+} x_{hj}^r = 0, \ \ h \in N, \ \ r \in K,$$ *Equation 3.6, (Azi et al., 2006)*

$$\sum_{i \in N^+} x_{0i}^r = 1, \ \ r \in K$$ *Equation 3.7, (Azi et al., 2006)*

$$\sum_{i \in N^+} x_{i(n+1)}^r = 1, \ \ r \in K$$ *Equation 3.8, (Azi et al., 2006)*

$$\sum_{i \in N} q_i y_i^r \leq Q, \ r \in k$$ *Equation 3.9, (Azi et al., 2006)*

$$t_i^r + s_i + t_{ij} - M\left(1 - x_{ij}^r\right) \leq t_j^r, (i,j) \in A^+, r \in K$$ *Equation 3.10, (Azi et al., 2006)*

3.4 Conclusion

In conclusion, after gaining knowledge from the systematic literature review and linking the literature review to the research objective. As the action problem of this Thesis refers to the high empty driven kilometres and the goal of the Thesis is to develop solution approach to support the planners with decision making, furthermore, the modelling phase requires high knowledge of the chosen model that will be constructed in this Thesis. The MILP was chosen as the solution approach for this Thesis project to solve the SVRPCTWB. The exact approach can provide accurate solutions and has high flexibility due to the constraints.

4. Data understanding

In the data understanding, the goal is to collect data from the data sources, explore it, describe it, and check the data quality (Kruse et al., 2021). Therefore, in **Chapter 4.1** addressing the sub-research question of "which available data will be used in the mathematical model?" will be conducted to collect the data from Vos databases for both Denmark and Finland and understand the links and quality in the data will be shown in **Chapter 4.2** and **Chapter 4.3**. Moreover, the sub-research question of "which marketplaces are available to script customers for both Finland and Denmark?" will be shown in **Chapter 4.1** with the attributes and results of the collection.

4.1 Data description

The data is separated into two data sheets stated as A and B as shown in **Figure 4.1**. Data set A contains the historical data for the Vos transportation group, and the Data set B includes the manually scripted marketplace customers gathered from the available marketplace websites which are lkw, trans.eu and Timocom. Data set A contains the shipment number, starting date, ending date, depot, returning depot, unload customer with all specifications, load customer with all specifications, total driven kilometres, empty driven kilometres, and the customer type. Data set A attributes are stated in **Table 4.1**. Data set A contains a set of **6345** shipments delivered to Finland and Denmark.

Furthermore, the data gathered from the marketplaces in data set B contained marketplace pickup location, marketplace drop-off location, and the distance between both locations. Filtering technique was used when acquiring the marketplace locations; for Denmark, the marketplace pickup area was limited to 200 KM around Copenhagen, and the drop-off location is within an area of 100 KM around Deventer, and 100 marketplace locations were acquired, while for Finland, 18 marketplace customers were acquired from a radius of 200 KM around Stockholm. The filtering technique was conducted to limit the amount of data collect, as the data was collected manually not using the marketplace API's due to the limited timeline of the thesis. The attributes for data set B are shown in **Table 4.2**.

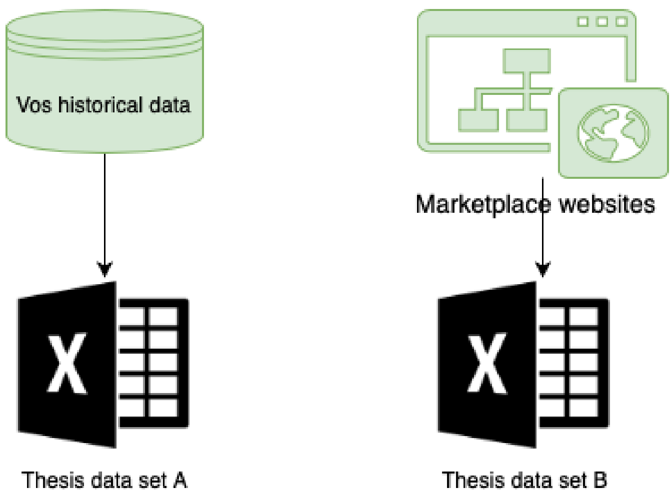

Figure 4.1 – Data sets

Table 4.1 – Data set A (Vos historical Data attributes)

COLUMN	DATA TYPE	FORMAT	DESCRIPTION
SHIPMENT_NUMBER	String	General	Shipment ID used in the company, describing the route of the shipment, and uniquely assigned for every shipment.
STARTINGDATE	Date	Date, DAY/MONTH/YEAR	Departure of the truck from the depot
ENDINGDATE	Date	Date, DAY/MONTH/YEAR	Arrival of the truck back to the depot after completing the journey
DEPOT	String	General	Starting point of the journey
RETURING_DEPOT	String	General	End point of the journey
UNLOAD_CUSTOMER	String	General	Standard customer who orders the order from the Netherlands to Finland or Denmark, the unload customer is defined as the linehaul customer
UNLOAD_CITY	String	General	Standard customer city that the trucks need to travel to deliver the full shipment
UNLOAD_LATITUDE	Float, **String (NULL)**	General	Standard customer coordinates
UNLOAD_LONGITUDE	Float, **String (NULL)**	General	Standard customer coordinates
UNLOAD_CUSTOMER	String	General	Marketplace customer that the planners of VOS decided to assign to pick up the loads and drop-off the loads in the radius of the depot.
LOAD_CITY	String	General	Marketplace customer city location
UNLOAD_LATITUDE	Float, **String (NULL)**	General	Marketplace customer coordinates
UNLOAD_LONGITUDE	Float, **String (NULL)**	General	Marketplace customer coordinates
EMPTY_DRIVEN_KM	Integer, **String (NULL)**	General	Empty driven kilometres for the full route
TOTAL_DRIVEN_KILOMETERS	Integer, **String (NULL)**	General	Total driven kilometres for the full route
CUSTOMER TYPE	Integer, **String (NULL)**	Genera, {1: SC, 2: MPC}	Customer type, {standard customer, Marketplace customer}

Table 4.2 – Data set B (Marketplace Data attributes)

COLUMN	DATA TYPE	FORMAT	DESCRIPTION
MARKETPLACE CUSTOMER NAME	String	General	Marketplace customer name
DATE OF ARRIVAL	Date, string	DAY/MONTH/YEAR, Day, Month/ +#	Date of picking up the shipment
COUNTRY – PU	String	{DE, DK, SE}	Country code of picking up the shipment
Postal code – PU	String, Integer	{####}, {#### AB}	Postal code of picking up the shipment
CITY – PU	String	General	City of picking up the shipment
DATE OF DILVER	Date, string	DAY/MONTH/YEAR, Day, Month {+#}	Date of dropping off the shipment
COUNTRY – DO	String	{DE, NL}	Drop-off country code
POSTAL CODE – DO	String	{####}, {#### AB}	Drop-off postal code
CITY – DO	String	General	Drop-off city
COST	String	EUR {#Cost}	Shipment transportation cost
Kilometres/KM	Integer	General	Kilometres between the pickup location and the drop-off location
POST DAY	Date, string	DAY/MONTH/YEAR	Offer post day
METERS/LDM	Float	General	Loading meters of the shipment.
WEIGHT/T	Float	General	Shipment weight

The scope of the data set A focused on three important criterions. Firstly, the data from Vos will only include the shipments that are delivered to Finland and Denmark, and will be separated into two columns to be analysed individually. Secondly, the time frame is limited to 2022, as the Thesis main research question focuses on comparing the performance of 2022 and the European average in both countries' in 2022. Finally, the project focuses on decreasing the empty driven kilometres when working with the marketplace. Therefore, only marketplace customers will be used in the model.

Data set B was extracted from the marketplace's recent offers in 2024, this raises a question of the reliability of the analysis, as the marketplace customer locations between the extracted data in 2024 differ from the marketplace customer locations in Vos historical data. Although, the IT department believes that the 2022 marketplace customer locations are relativity close in distance with the 2024 marketplace customer location. Hence the IT-department statement should be test. Therefore, to increase the test reliability of having different marketplace customers for each test group, an analysis of comparison between the marketplace customer locations for both data sets should be conducted in the data preparation section.

4.2 Data quality

To have reliable findings at the end of the project, the goal is to start with reliable data. The data quality section will focus on identifying the missing data, errors, and relevance to the research question for data set A. Furthermore, when extracting data set B, all necessary data was extracted except for the pickup times, as many customers only provide the pickup day; therefore, data set B is already identified as a reliable quality.

Starting with the missing data and errors, **Table 4.3** shows a sample of Data set A with several colours shading the cells of the data. The highlighting colours represent different kinds of errors and missing data, and these errors appeared in many instances of the data sheet.

To begin, yellow represents duplicates in the shipment ID, while the shipment ID is supposed to be a unique string representing a unique shipment route. Following, red represents when data is missing from the data sheet. Afterward, it was noticed that sometimes there are empty driven kilometres higher than the total driven kilometres, which indicates an error in the row and is represented as purple coloured.

Furthermore, the data set contains data irrelevant to the project. The orange colour represents when the starting and returning depots are not the same, since the Thesis case problem is considered to be a variant of the vehicle routing problem, therefore the starting and returning depot should be the same for the whole shipment ID route. Afterward, the thesis case study focuses only on shipments related to the use of the marketplace. Therefore, the customer type should equal to two in the customer type column, as two represent the usage of the marketplace when returning to the depot.

Shipment Number	StartingDate	EndingDate	Depot	Returning Depot	unload_customer	Customer type	unload_city	unload_latitude	unload_longitude	load_customer	load_city	load_latitude	load_longitude	Empty driven kilometers	Total driven Kilometers
FI01099	31/12/2021	04/01/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01100	04/01/2022	13/01/2022	DEVENTER	DEVENTER	K14810	1	HAMEENLINNA	60.96458	24.51265	K11255	BUCHHOLZ	53.33993	9.84188		
FI01100	04/01/2022	13/01/2022	DEVENTER	DEVENTER	K14810	1	HAMEENLINNA	60.96458	24.51265	K11255	BUCHHOLZ	53.33993	9.84188		
FI01101	07/01/2022	10/01/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01102	12/01/2022	20/01/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K06489	VANTAA	60.28336	24.98416		
FI01102	12/01/2022	20/01/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K06489	VANTAA	60.28336	24.98416		
FI01103	14/01/2022	15/01/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01104	18/01/2022	28/01/2022	DEVENTER	DEVENTER	K01501	1	TAMPERE	61.48187	23.76026	K06489	VANTAA	60.28336	24.98416		
FI01105	21/01/2022	25/01/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01106	25/01/2022	03/02/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K04108	HAGENOW	53.42325	11.1848		
FI01107	28/01/2022	29/01/2022	DEVENTER	DEVENTER	K05206	3	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01108	01/02/2022	12/02/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K04108	HAGENOW	53.42325	11.1848		
FI01108	01/02/2022	12/02/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K04108	HAGENOW	53.42325	11.1848		
FI01108	01/02/2022	12/02/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K04108	HAGENOW	53.42325	11.1848		
FI01109	04/02/2022	07/02/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01110	09/02/2022	24/02/2022	DEVENTER	EMMEN	K00715	1	SIURO	61.47442	23.35758	K04108	HAGENOW	53.42325	11.1848		
FI01111	09/02/2022	17/02/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K06489	VANTAA	60.28336	24.98416		
FI01112	11/02/2022	14/02/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01113	15/02/2022	04/03/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K04108	HAGENOW	53.42325	11.1848		
FI01114	18/02/2022	21/02/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01115	23/02/2022	10/03/2022	DEVENTER	DEVENTER	K02690	1	KAUHAJOKI	62.45394	22.20583	K06489	VANTAA	60.28336	24.98416		
FI01116	25/02/2022	04/03/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01117	01/03/2022	17/03/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K12611	CARLOW	53.75955	10.94402		
FI01117	01/03/2022	17/03/2022	DEVENTER	DEVENTER	K06489	1	VANTAA	60.28336	24.98416	K12611	CARLOW	53.75955	10.94402		
FI01118	01/03/2022	18/03/2022	DEVENTER	DEVENTER	K14810	1	HEINOLA	61.20761	26.03249	K12123	LEIPZIG	51.41946	12.44357		
FI01118	01/03/2022	18/03/2022	DEVENTER	DEVENTER	K14810	1	HEINOLA	61.20761	26.03249	K12123	LEIPZIG	51.41946	12.44357		
FI01118	01/03/2022	18/03/2022	DEVENTER	DEVENTER	K14810	1	HEINOLA	61.20761	26.03249	K12123	LEIPZIG	51.41946	12.44357		
FI01118	01/03/2022	18/03/2022	DEVENTER	DEVENTER	K14810	1	HEINOLA	61.20761	26.03249	K12123	LEIPZIG	51.41946	12.44357		
FI01118	01/03/2022	18/03/2022	DEVENTER	DEVENTER	K14810	1	HEINOLA	61.20761	26.03249	K12123	LEIPZIG	51.41946	12.44357		
FI01118	01/03/2022	18/03/2022	DEVENTER	DEVENTER	K14810	1	HEINOLA	61.20761	26.03249	K12123	LEIPZIG	51.41946	12.44357		
FI01119	04/03/2022	09/03/2022	DEVENTER	DIEST	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01120	08/03/2022	25/03/2022	DEVENTER	DEVENTER	K12531	1	VANTAA	60.2887	25.07715	K06489	VANTAA	60.28336	24.98416		
FI01121	11/03/2022	14/03/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01122	15/03/2022	25/03/2022	DEVENTER	DEVENTER	K01854	3	HELSINKI	60.17256	24.93407	K06489	VANTAA	60.28336	24.98416	1201	1065
FI01123	18/03/2022	21/03/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01124	23/03/2022	31/03/2022	DEVENTER	COEVORDEN	K05206	1	VANTAA	60.28939	25.02953	K04108	HAGENOW	53.42325	11.1848		
FI01125	23/03/2022	06/04/2022	DEVENTER	DEVENTER	K02690	1	HYVINKAA	60.63587	24.86054	K10729	GOTHA	50.96709	10.72881		
FI01126	25/03/2022	28/03/2022	DEVENTER	DEVENTER	K05206	1	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953		
FI01127	30/03/2022	30/03/2022	DEVENTER	LUBECK	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	440
FI01128	02/04/2022	04/04/2022	DEVENTER	DEVENTER	K05206	3	VANTAA	60.28939	25.02953	K05206	VANTAA	60.28939	25.02953	0	862

Table 4.3 - Data set A defects

4.3 Data exploration

In the data exploration section, the goal is to start forming a data set that can be relevant to the project. Checking the attributes of the historical data in **Table 4.1**. Firstly, checking the data for both Finland and Denmark, which are represented in the same columns differentiated by the first letters of the unique string of the shipment number. Therefore, to have a clear data set, the shipment number column will be separated into two subsets for the shipments related to Denmark and Finland. After separating the shipment IDs, it shows that there were **111** shipments in Finland and **3784** shipments in Denmark; however, these numbers represent uncleaned data.

After having all the necessary tools for the research, the current task is to start exploring the uncleaned company data. The analysis will follow a root of applying the descriptive statistics, and ratio test between the marketplace and standard customers. Before starting the analysis, strings in the numeric columns will be removed to be able to conduct the analysis.

To begin, the descriptive statistics for Denmark and Finland empty driven kilometres were performed in Excel using Excel tools, and the results are shown in **Table 4.4** for Denmark and in **Appendix A** for Finland. The descriptive statistic for Denmark shows that the average empty-driven kilometres are 1249,1 Km, while the average total driven kilometres is 1058,6 Km; hence, the empty driven kilometres percentage in Denmark is **8,04%**; the analysis disproves the starting percentage of empty driven kilometres for Denmark that was provided by the company; which was 6,03%. Furthermore, the standard deviation is very large, with a value of 140,03, demonstrating that there is a wide range between values, which can be proven by the difference between the maximum and minimum. Furthermore, when analysing Finland empty driven kilometres, the average empty driven kilometres percentage resulted with **6,43%**, disproving again the given empty driven kilometres from Vos team; which was stated as **18,09%**.

Table 4.4 - Vos 2022 descriptive analysis for Denmark

DENMARK EMPTY DRIVEN KILOMETRES	
MEAN	124,9178466
STANDARD ERROR	1,904053337
MEDIAN	61,64
MODE	441,53
STANDARD DEVIATION	140,0482681
SAMPLE VARIANCE	19613,51739
KURTOSIS	0,63613646
SKEWNESS	1,426122635
RANGE	792,61
MINIMUM	0
MAXIMUM	792,61
SUM	675805,55
COUNT	5410

Additionally, an analysis is conducted to check the usability of the marketplace by Vos in both Finland and Denmark using the function {=count IF (customer type, Desired type)}. The results showed that for Denmark, the standard customers are used for 2361 shipments, while 237 shipments were returned using the marketplace, demonstrating that the marketplace was used only 6,7% of the time. Remarking that there are 1423 shipments which have an unknown customer type, which the company states that these are standard customers. Likewise, Finland was conducted with the same analysis, and the results state that 89 shipments were returned from standard customers and two returning shipments were returned using the marketplace, demonstrating that the marketplace was only used 1,0% of the time. **Table 4.5** demonstrates the empty driven kilometres results and the usage of the marketplace.

Table 4.5 – Uncleaned empty driven kilometres percentage and customer type usage

UNCLEANED DATA	*Provided empty driven kilometres*	*Actual empty driven kilometres*	*Standard customers usage*	*Marketplace usage*
Denmark	6,03%	8,04%	2361	237 (6,7%)
Finland	18,09%	6,43%	89	1 (1,0%)

Moreover, three tests are conducted to check if the data is considered normally distributed; firstly, the kurtosis should be close to 3, and the skewness should be close to 0; however, the kurtosis for Vos in Denmark is 0.6, which indicates that there are several outliners, while the skewness is 1.4 stating that the distribution is skewed to the right. Both the skewness and kurtosis output and clarification can be verified by checking the histogram in **Figure 4.2**. After the statistical analysis, we can verify that the data does not follow a normal distribution.

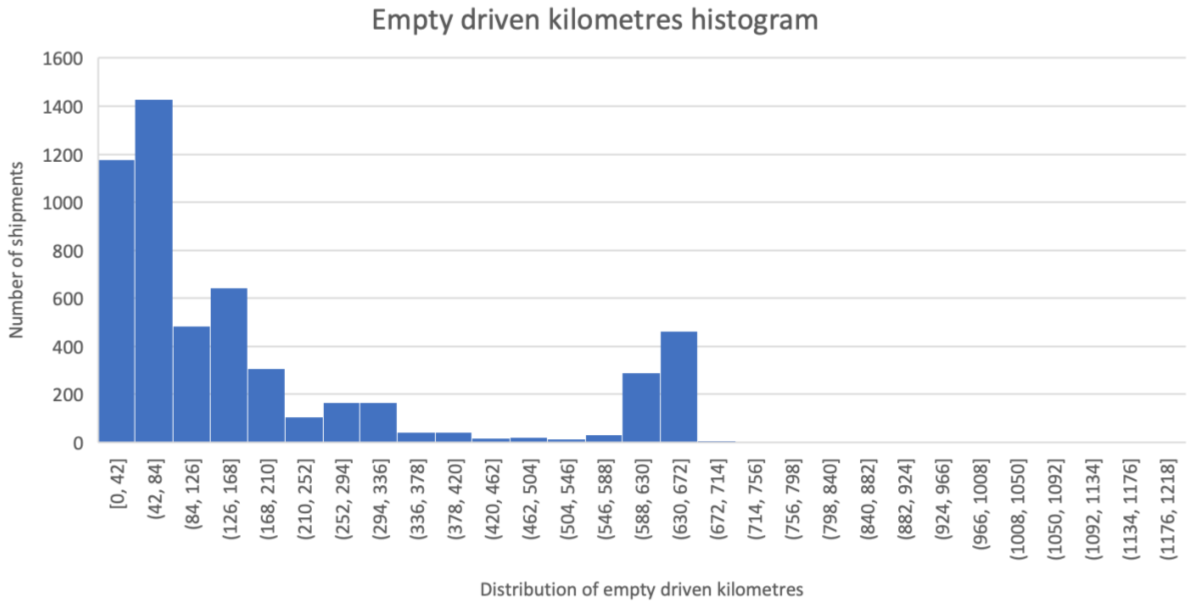

Figure 4.2 – Vos 2022 histogram for Denmark

4.4 Conclusion

In the section, two data sets were extracted; data set A describes the historical data of Vos, while data set B describes the marketplace customers' data. Afterward, historical data quality was examined to find the defects, unnecessary data, and missing data. Finally, data set A was explored to find that the average empty driven kilometres for Vos in Finland and Denmark deferred from the given average empty driven kilometres given by Vos. Moreover, the exploration showed that the marketplace usage was very limited in 2022, given that only 6,7% of the shipments were returned using the marketplace for Denmark, while only 1,0% of the shipments were returned using the marketplace in Finland.

5. Data preparation

In the data preparation phase, the objective in this phase is to have clean and reliable data to address the sub-research question of “how can cleaning the data ensure the validity of the data?”. Therefore, cleaning the data will be performed using Microsoft excel in **Chapter 5.1**. Afterwards, the goal in **Chapter 5.2** is to create a data model through an entity relationship diagram to visualize the connection between data sets. Finally, an analysis in **Chapter 5.3** will be conducted using the cleaned data to check the normality, the seasonality usage for the marketplace against the standard customers, comparison test for the marketplace customer location for data set A and B, and finally a peak test to identify the reasons of the outliners.

5.1 Clean data

In the data understanding section, the three main defects criterions regarding the uncleaned data were identified, and now it should be cleaned to be used in the model. Cleaning the data is performed using the Tables of Microsoft Excel filtering technique and starts with removing the duplicates of the shipment IDs and tackling all NULL, which represents the missing data. Afterward, data that are not relevant for the Thesis case study are removed; hence, all unloading linehaul cities that are not in Finland or Denmark were removed, and due to the VRP requirement, the starting depot and ending depot should be the same.

Through cleaning the data, the findings have showed that there were 1049 duplicates of shipment IDs in Denmark and 35 duplicates in Finland. For Denmark, there were 34 of the linehaul customers that were not in Denmark or in the radius of Denmark. hence, after completing all the cleaning of the data, the results showed that the data decreased for Finland from **111 shipments** to **1 shipment**, while for Denmark, the data decreased from **3784** to **141 shipments**.

The results of the cleaning data demonstrate that the marketplace was not used widely in Vos for both countries and the planners usually collaborate with standard customers, showing that the marketplace was only used for 0,59% in Finland and 4,0% in Denmark as shown in **Table 5.1**.

Table 5.1 – count and marketplace usage percentage for cleaned and uncleaned data set A

	Uncleaned data set	*Cleaned data set*	*Marketplace usage for uncleaned data*	*Marketplace usage for cleaned date*
Denmark	3784	141	6,7%	4%
Finland	111	1	1,0%	0,59%

5.2 Working data

Currently there are two cleaned data sets that were extracted for this thesis; however, the data needs to be structed clearly to work without any interpretations. Therefore, an entity relation diagram is built to visualize the linkage between the data sets. While building the entity relation diagram, Finland shipments were removed from the working data as there is only one shipment, because the amount of data is not enough to conduct the analysis or evaluate the model. Therefore, data set A will only use shipments that travelled to Denmark.

Moreover, to ensure each marketplace customer has a unique identifier, new columns for pick-up ID and drop-off ID were constructed to represent the primary keys for the marketplace locations. Furthermore, to construct the linking of data set A with data set B, new column was added to address the arrival date at the unloading customer location, which matches the starting date of the marketplace pickup location. The entity relation diagram is shown in **Figure 5.1**.

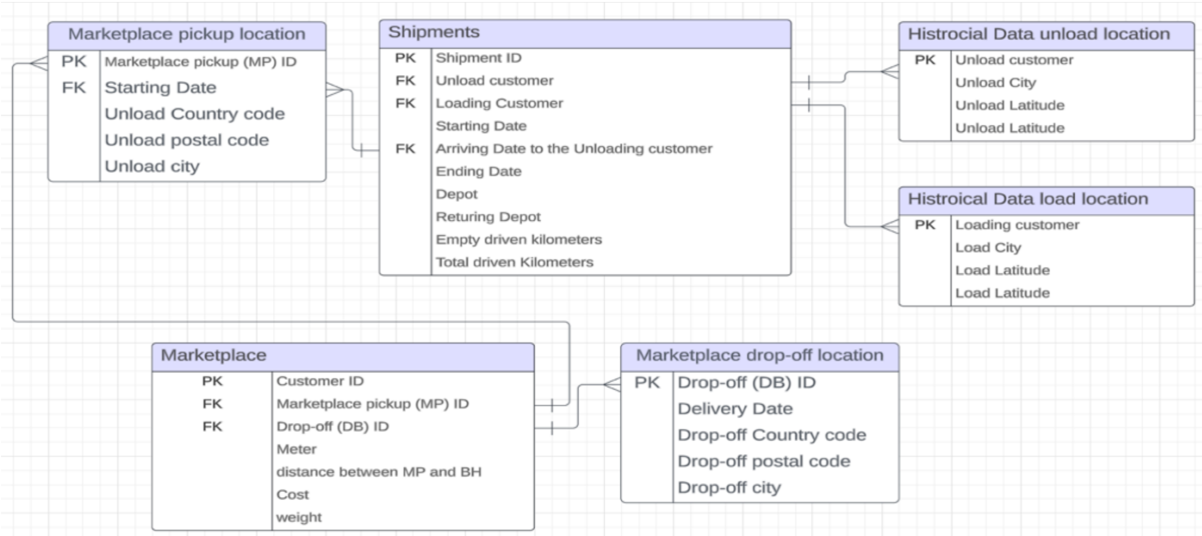

Figure 5.1 – Entity relation diagram

5.3 Analysis

The analysis of the working data goes through several tests. Starting with analysing the normality of Denmark' working data, follows with a seasonality test for the usage of the marketplace and compared to the usage of the standard customers. Afterwards comparison of the marketplace customer locations between the data set A and B is conducted. Finally, peak test for the historical data is performed to analyse the peaks of the empty driven kilometres.

5.3.1 Normality test

Conducting the normality test for Denmark's cleaned data starts with forming a descriptive statistic and performing the bin analysis to construct the histogram chart. The descriptive data was applied using Excel function. The data is shown in **Table 5.2**. The findings show that the kurtosis of 5.9 represents a leptokurtic distribution, which indicates that the data has a sharper peak and several outliners. The skewness of 1.55 shows that the data has a strong right skew, proven by having a higher mean than the median and can be seen in the histogram in **Figure 5.2.** A data is normally distributed when the kurtosis should be close to three, and the skewness should be close to zero; hence, Vos empty driven kilometres for Denmark are not normally distributed.

Table 5.2 – Descriptive analysis

DENMARK EMPTY DRIVEN KILOMETRES	
MEAN	65,34
STANDARD ERROR	3,14
MEDIAN	55,61
MODE	128,64
STANDARD DEVIATION	45,68
SAMPLE VARIANCE	2086,69
KURTOSIS	5,90944
SKEWNESS	1,55
RANGE	343,04
MINIMUM	0
MAXIMUM	343,04
SUM	13785,92
COUNT	211

Afterward, constructing the histogram of the empty driven kilometres started by identifying the bins that present the distribution of small data set samples. The bin size is found using the Excel data analysis function; the process of finding the bin inputs of the histogram starts with a quick heuristic approach of finding the count by square rooting the data count. Afterward, the maximum number of the data is divided by the data count to find approximate bin counts. Furthermore, the bins indicate the distribution using a difference of 35 empty driven kilometres between each bin, and the frequency shows the size of each bin bound, as seen in **Figure 5.2**. The histogram shows that the data is heavily skewed to the right, proving again that the data is not normally distributed. The histogram shows the outliners at the frequency of bins 210, 245, and 350. In **Chapter 5.3.4**, an analysis of the peaks will be conducted, and if the outliners are due to errors, then the outliners will be removed.

Count	*Bin*	*Frequency*
1	35	52
2	70	73
3	105	45
4	140	34
5	175	3
6	210	2
7	245	1
8	280	0
9	315	0
10	350	1
11	385	0
12	420	0
13	455	0
14	490	0
15	525	0

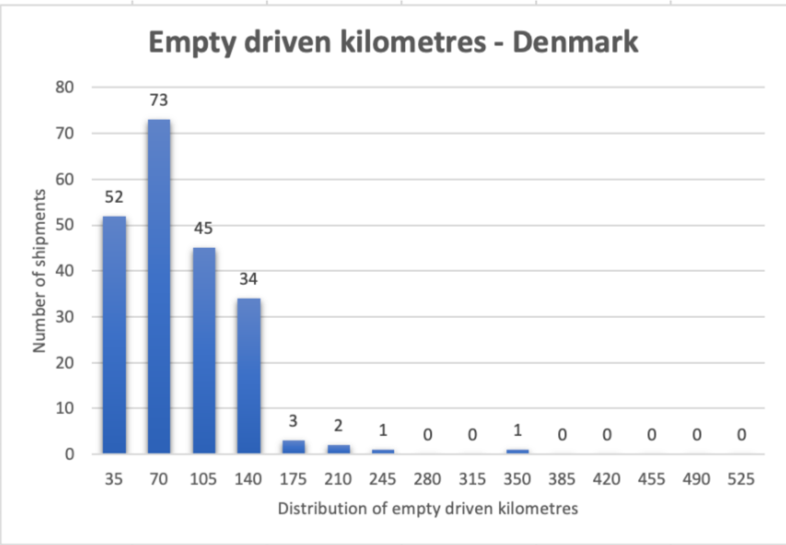

Figure 5.2 – Histogram Vos Denmark

5.3.2 Marketplace usage timeline

Analysing the marketplace usage timeline is performed through seasonal analysis and comparison of usage test. To begin, the seasonal chart in **Figure 5.3** was constructed using the pivot table in Excel, and the marketplace usage was computed using the count of the repeatability of the dates in the Vos data. The analysis shows peaks in marketplace usage in July, August, October, and November. July and August are holiday periods; therefore, Denmark's factories are closed, with fewer standard customers return shipments. The high usage of the marketplace in October and November may resulted from the rupture of the Nord Stream gas pipelines (Reuters, 2022) which influenced in increasing the gas cost in the Netherlands.

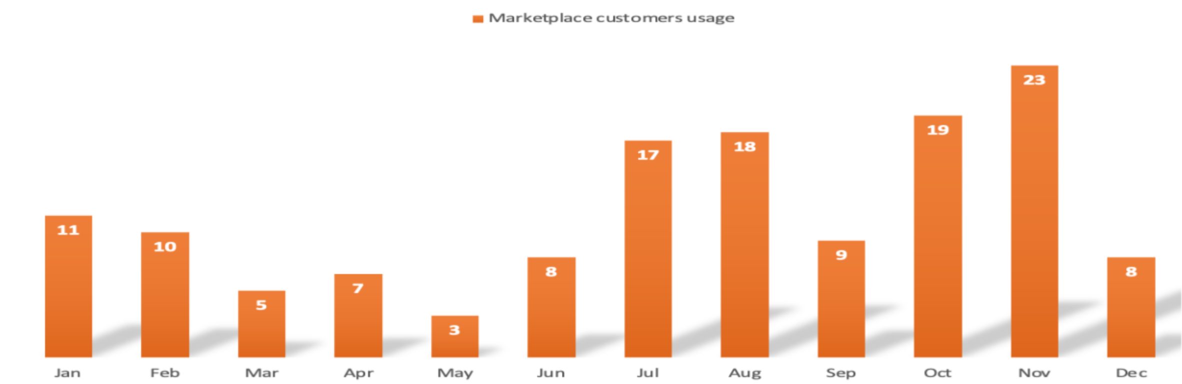

Figure 5.3 – Seasonal test

Furthermore, **Figure 5.4** compares the usage between the standard and marketplace customers. The comparison shows the huge difference in usage between the standard and marketplace customers. The usage of standard customers ranges between 171 and 217, with a close shipment range of usage each month. Since the peaks of the marketplace are July, August, October, and November, an individual analysis is conducted in these months. Firstly, July shows there was a shortage of the standard customers, as July represents the third lowest month of using the standard customers. Secondly, August shows that the standard customer returned to work with high usage; therefore, this month shared a high usage of both the marketplace and standard customers. Thirdly, October and November shared a high usage of both the marketplace and standard customers. In October, the standard customers were higher than those in November. However, the Marketplace customers in November are higher than the marketplace customers in October by a close proportion.

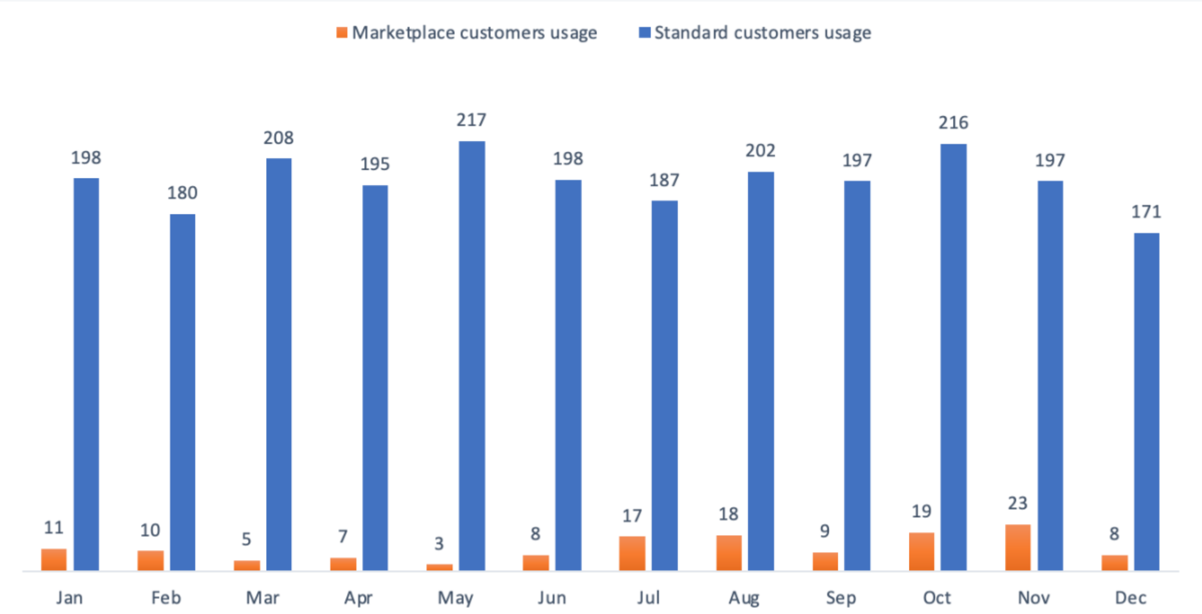

Figure 5.4 – Returning customers' usage comparison

5.3.3 Marketplace customer locations

In the evaluation section, a test will be conducted to analyse the empty driven kilometres resulted from the decisions of the planners in 2022 versus the empty driven kilometres resulted from the decisions of the developed model. One of the main challenges is the marketplace offers, as the marketplace offers in 2022 differ from the marketplace offers scripted for this project in 2024; however, the IT-department in Vos believes that there is a similarity between the marketplace locations in 2022 and 2024; therefore, a test needs to be conducted to compare the marketplace customer locations in 2022 and 2024.

To check if there is a similarity between the marketplace customer locations in 2022 and 2024, using Tableau, the marketplace locations of 2022 and 2024 were used as the testing variables. **Figure 5.5** shows the distribution of the marketplace locations in Denmark for both 2022 and 2024. 2022 marketplace locations are represented with blue colour, and 2024 marketplace locations are represented with orange colour.

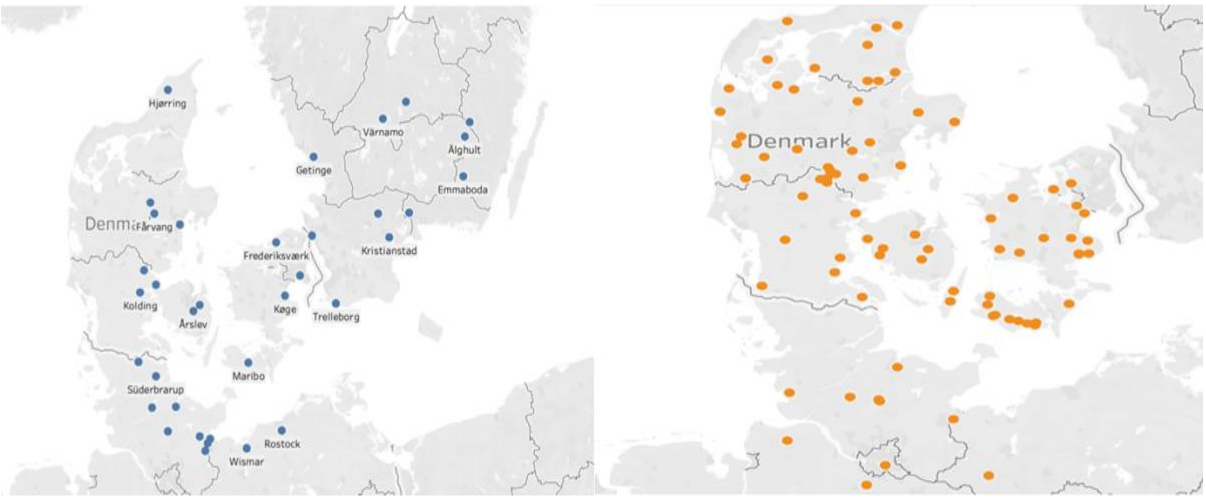

Figure 5.5 – Marketplace customer locations in 2022 and 2024

Afterward, the longitude and latitude of the 2022 and 2024 marketplace customer locations were used to construct the scatter plot, as shown in **Figure 5.6**. The scatter plot shows that the distribution of 2022 and 2024 marketplace customer locations are very close. Therefore, this results in accepting the usage of the 2024 marketplace customer locations to test the model and then comparing the planners results in 2022.

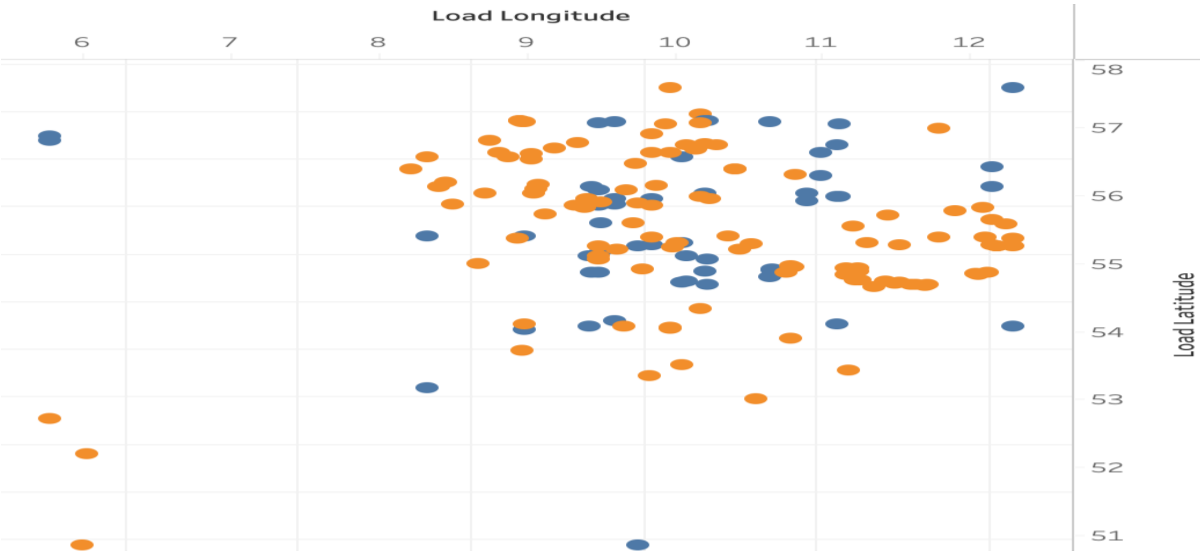

Figure 5.6 – Scatter plot for the 2022 and 2024 Marketplace customer locations

5.3.4 Peak test

A peak test is conducted to check the outliners of the empty driven kilometres and check the reasoning behind the peaks. The peak line chart is shown in **Figure 5.7**. The peaks are identified as any empty driven kilometres higher than the average empty driven kilometres. When checking the top five empty driven kilometres, it shows that the loading locations from the marketplace were picked from locations very far away from Denmark; for example, the highest peak of approximately 335 empty driven kilometres had a trip from Denmark, and the pickup location was in Apeldoorn in the Netherlands. The main reasoning behind the high empty driven kilometres is that the planners were planning based on reducing the total driven kilometres through searching for shipment on the way back from Denmark, not taking into consideration the empty driven kilometres. In conclusion, as stated in **Chapter 5.3.1**, the peak analysis is conducted to check if the peaks resulted from errors as found in the Data understanding section. However, after performing the peak analysis, the analysis shows that there are no errors with the data. Hence, the decisions of the planners are the reasoning behind the empty driven kilometre peaks.

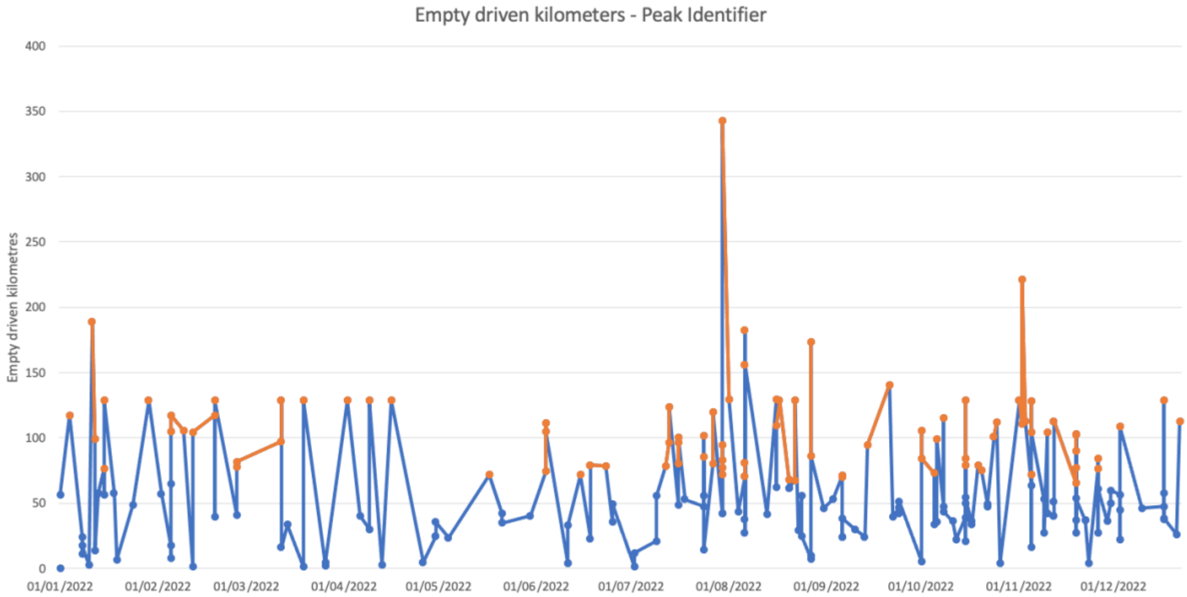

Figure 5.7 – Peak analysis

5.4 Conclusion

In this chapter, the data was cleaned and resulted in decreasing the data in Denmark from **3784** to **141**, while in Finland the data decreased from **111** to only **one** shipment. The first decision in this chapter is that **Finland will not be included** in the data preparation for the model, because no data is available from Finland shipments. The marketplace usage for Denmark cleaned data showed that only 3,95% of the shipments were returned from Denmark using the marketplace. Afterward, an entity relation diagram was constructed to build a relation between Data set A and Data set B.

Furthermore, an analysis is conducted against the cleaned data. Firstly, normality test on the cleaned Data set A showed that the data is not normally distributed. Secondly, the marketplace usage periods were analysed, and the results showed that the marketplace was used mostly in the holidays and during the end of the year might be because of the rupture of the Nord Stream gas pipelines which influenced in increasing the gas cost in the Netherlands. Thirdly, the marketplace customer locations of Data set A and B where compare based on their longitude and latitude. The findings showed that Data set A and B had close marketplace customer locations, therefore, an assumption can be made to use Data set B marketplace customer locations to evaluate the performance of the model and compare it with the results of the Data set A. Finally, peak analysis was conducted to check if the peaks were resulted from any defect, as in the data understanding. However, the peaks were a result of the planner’s strategy in decreasing the total driven kilometres when selecting marketplace customer location on the way of the truck, not considering the distance between the truck drop-off and the pickup from the marketplace.

6. Solution modelling, implementation, and Validation

Modelling phase requires building the test case (prototype) and the model and selecting the parameters and clarifying the model are essential elements (Kruse et al., 2021). In this Thesis, the modelling phase will consist of modelling the mathematical integer linear programming in **Chapter 6.1** to approach the sub-research question of What is the characteristic of the solution model?, then implementing the model in python to create the test case prototype in **Chapter 6.2**. Afterwards, in **Chapter 6.3** the built model should be verified using test data and conducting all possible scenarios.

6.1 Mixed integer linear programming

As demonstrated in the literature and business understanding phases, the desired process the company desires to achieve shows that this thesis vehicle routing problem is considered as single vehicle routing problem with capacity constraint, time window, and backhaul shipments (SVRPCTWB).

This section will provide the model solution for the SVRPCTWB. The transportation process is defined as a directed graph of G = (N, A), where N is the node set representing the locations of the depot, linehauls, and backhauls. De, L, PU, and DO ⊆ N, where De represents the depot, L is a Set of linehaul customer that travels from the depot to the customer standardized by the company planners, PU is a Set of Backhaul customers that are selected from the marketplace website, finally DO is a set representing marketplace returning drop-offs locations, and they are considered to be linehaul for the market place customers.

The phases of all elements in N are visualized clearly in **Figure 6.1**, and the following sets 1,2,3 and 4 describe both the sets of N with the subsets De, L, and M and describe the stages of each element in the transportation cycle. {A} represent the set of arcs {i,j} considering the stages S, when {i,j}∈N. S is a set that describes the stage path of the arcs; hence, it shows where the arc path follows.

As the problem contains different stages with different constraints, a set of stages is constructed to represent each terminal. Six stages were constructed, as shown in set 5, representing the flow between the N subsets. **Table 6.1** represents the notations of sets.

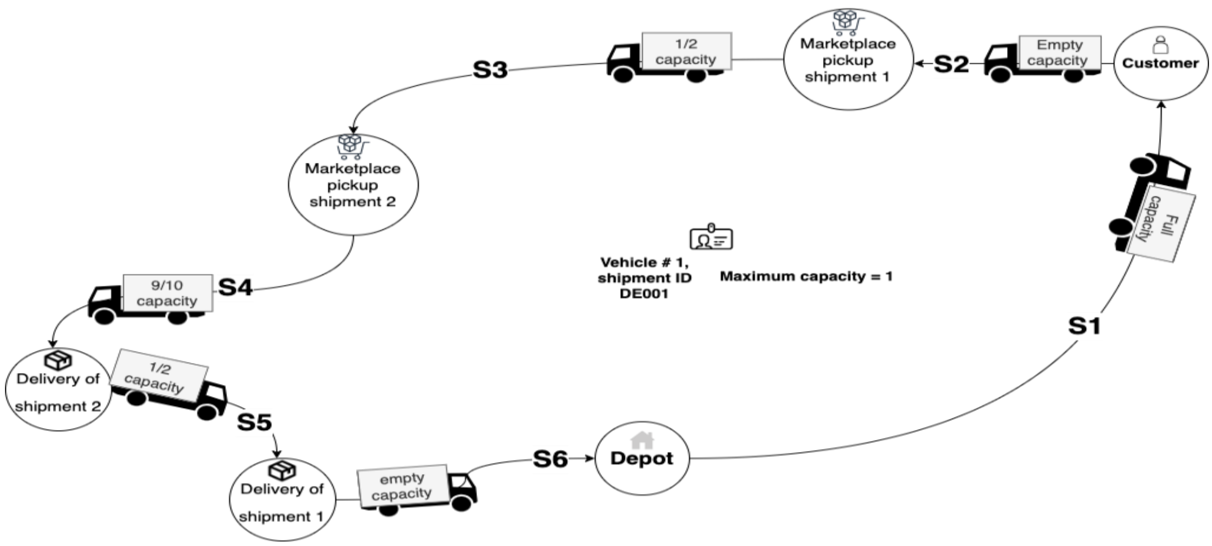

Figure 6.1 – Problem solution process visualization

To assess the model, several parameters need to be added based on the company criteria. **Table 6.2** shows the model parameters. Firstly, the model should calculate the distance between all node N, therefore, a library API that is called using the model programming tool Python will calculate the distance d_{ij} between all nodes N with the travelling time t_{ij}. Afterwards, as we have one linehaul customer set, the linehaul customer needs to be assigned by the planners manually, therefore, a parameter l which select the linehaul customer which will be changed according to the planners linehaul customer. Moreover, to assess the constraints of the model, several parameters of the process need to be constructed. Hence, truck constraints are the capacity of 13,6 LDM and minimum pickup load of 6 LDM. The model will be considering a loading time at the marketplace of 1 hour and an unloading time at the linehaul customers of half an hour. An assumption was made for the time windows of the marketplace pickups, as the marketplace customer usually do not include the pickup times and the company wants to pick up the shipments in one day, therefore, the assumption will state that the vehicle will arrive to the linehaul customer at 8 am and needs to finish the pickups from the marketplace till 8 pm. Therefore, two pick-up locations are feasible, if the maximum travel time between the linehaul and the farthest marketplace location of 200 kilometres will result in approximately two hours, in additional to this, there is an unloading time of half an hour in the linehaul location and two hours in the pick-up marketplace locations. Therefore, the working hours of 10 hours minus the used hours will result into 330 minutes. The 330 minutes are the maximum travel time between the first marketplace location and the second marketplace location.

Table 6.1 – Model sets

SETS	
N	*Set representing the nodes of the case study. where De, L, MP, and DO $\subseteq N$.*
A	*Set representing the arcs $\{i, j\}$ considering the stage s. where*
S	*Set representing the stage of the vehicle v in the process.*

$N = \{0, l, l+1, \dots, l+m, l+m+1, \dots, n\}$ *Set 1*

$L = \{l\}$ *Set 2*

$MP = \{l+1, l+2, \dots, l+m\}$ *Set 3*

$DO = \{l+m+1, l+m+2, \dots, l+m+d\}$ *Set 4*

$S = \{1 = De \rightarrow l\,, 2 = l \rightarrow PU, 3 = PU \rightarrow PU, 4 = PU \rightarrow DO, 5 = DO \rightarrow DO,$
$6 = DO \rightarrow De\}$ *Set 5*

Table 6.2 – Model parameters

PARAMETERS	
t_{ij}	*Travel time for* $arc\{i,j\}$
d_{ij}	*Distance travelled on* $arc\{i,j\}$
$w_{i,j}$	*Vehicle weight in state i or j*
CA	*Capacity of the vehicle,* $CA \leq 13{,}6\ LDM$
MPCA	*Minimum pickup load of the vehicles,* $MPCA \geq 6LDM$
$MTT_{S=3}$	*Maximum travel time in stage 3,* $MTT \leq 330min$
l	*The linehaul city selected for the route by the planner*

Decision Variables:

$$x_{ijs} = \begin{cases} 1\ arc(i,j)\ is\ used\ \ in\ stage\ s \\ 0\ otherwise \end{cases}$$

$$y_{is} = \begin{cases} 1\ if\ node\ i\ is\ visited\ in\ stage\ s \\ 0\ otherwise \end{cases}$$

$$u_{is} = continous\ variable\ representing\ Load\ of\ the\ truck\ after\ leaving\ the\ node\ i$$

Objective function:

$$Z_{EDK} = min \sum_{s\in\{2,6\}} \left(\sum_{j\in MP} d_{lj} \times x_{lj2} + \sum_{i\in BH} d_{i0} \times x_{i06} \right) \qquad OF\ 1$$

The objective function shown in equation OF1 computes the minimum distance between the linehaul and the first marketplace and the minimum distance between the marketplace drop-off location and the depot. The objective function checks every marketplace customer with their pick-up and drop-off location, and if all constraints fulfil the conditions, then the customer with the minimum distance will be chosen as the optimal location based on the empty driven kilometres.

Constraints:

As discussed in the case study of the MILP, several stages are constructed to assess the planning cycle. Constraints are formulated starting from stage two till stage six, as stage one will be applied by the planners manually. The constraints sections will individually outline and describe the constraints of each stage of the case study to enhance clarity and comprehension. Constraints 1 and 2 are implemented for the whole stages. Therefore, both constraints are denoted as general constraints. Constraint 1 checks the cumulative weight of the visited node y and stores the capacity in the cumulative decision variable U. Constraint 2 then ensures that the cumulative weight should not exceed the vehicle capacity of 13,6 LDM. Moreover, constraints 3, 4, and 5 are sign restriction constraints for the continuous variable of the cumulative load.

$U_{is} = \sum w_j \times y_j \qquad i \in MP,\ j \in MP, i \geq j,\ s \in [0,6]$ *C 1*

$U_{is} \leq CA \qquad i \in MP, s \in [0,6]$ *C 2*

$u_{is} \geq 0$ C 3

$x_{ijs} = \{0,1\}$ C 4

$y_{is} = \{0,1\}$ C 5

Stage 2:

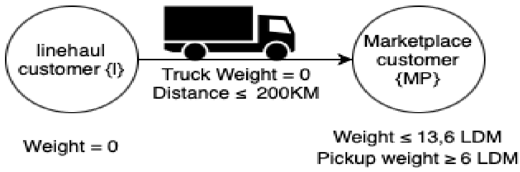

Figure 6.2 – Stage two

Stage 2 shows the truck's travel from the linehaul customer to the marketplace customer. At the beginning of stage 2, the truck departs from the linehaul customer empty-weighted and travels to one of the marketplace customers within a radius of 200 kilometres and has a loading quantity of not more than 13,6 LDM and not less than 6 LDM. **Figure 6.2** visualizes the path of stage 2.

Constraint 6 ensures that a vehicle will always visit a location in stage two if all other constraints in this stage also satisfy their constants. Constraint in stage 2 control x_{lj2} , which represents the usage of arc ij, will be equal to one or zero. Hence, using the constraints of distance limiter, minimum pickup quantity, and maximum load with the objective function that aims to minimize the travelled distances will result in the optimal x_{lj2} that can proceed to the next stage. However, if the optimal x_{lj2} does not satisfy the constraints in the next stages, then another arc of x_{lj2} will be selected. Constraint 7 states that if an arc between the linehaul and a marketplace exists in stage 2, then node i must be visited in stage 2, which, in other words, $y_i = 1$. This constraint creates a relation between the node visited and the arc taken. Constraint 8 is built to limit the empty driven kilometres in stage two to an area of 200 kilometres radius; therefore, any locations outside the 200 kilometres of the visited linehaul city will be neglected. Constraint 9 denotes that the minimum pick-up weight for the vehicle from the marketplace location should not exceed the minimum pick-up weight, which is equivalent to 6 LDM.

$\sum x_{lj2} = 1 \qquad j \in MP$ *C 6*

$\sum x_{lj2} = \sum y_j \qquad j \in MP$ *C 7*

$\sum d_{lj} \times \sum x_{lj2} \leq 200 \qquad j \in MP$ *C 8*

$\sum x_{lj2} \times w_j \geq MPCA \qquad j \in MP$ *C 9*

Stage 3:

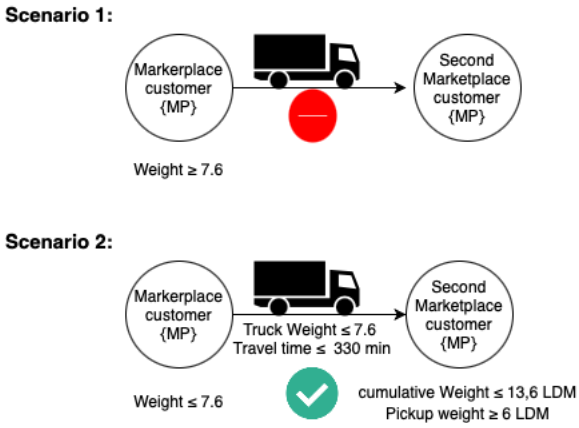

Figure 6.3 – Stage 3

In stage three, there are two scenarios. The first scenario states that if the truck already loaded more than 7,6 LDM in stage two, then the truck will not travel to a second marketplace customer. While, In the second scenario, if the truck loaded a load less than 7,6 LDM, then there is an opportunity to visit a second marketplace location; taking into consideration, that a cumulative load of both marketplace locations is less than 13,6 LDM, the pick-up load greater than 6 LDM, and a travel time between the first marketplace customer and the second marketplace customer less than 330 minutes. **Figure 6.3** visualizes the path of stage 3.

Constraint 10 denotes that if there is an arc from the linehaul customer to the first marketplace, then an arc from the same marketplace to other marketplaces can exist. The constraint ensures continuity in the network process from one stage to the other. Constraint 11 ensures that the pickup load will be more than 6 LDM from the second marketplace location. However, if the truck has already loaded 7,6 LDM in stage two, then using constraints 1 and 2 which control the cumulative pickup load will already convert the arc x_{ij3} into zero, therefore, it will neglect the use of constraint 11. Constraint 12 shows the time constraint in stage 3, the constraint checks all possible arcs ij and if the travel time between the arcs is larger than the maximum travel time of 330 minutes, then the arc x_{ij3} will be converted to zero.

$$\sum x_{ij3} \leq \sum x_{li2} \qquad i \in MP, j \in MP, i \neq j \qquad \text{C 10}$$

$$\sum x_{ij3} \times w_j \geq MPCA \qquad i \in MP, j \in MP, i \neq j \qquad \text{C 11}$$

$$\sum t_{ij} \times x_{ij3} \leq MTT \qquad i, j \in MP, i \neq j \qquad \text{C 12}$$

Stage 4:

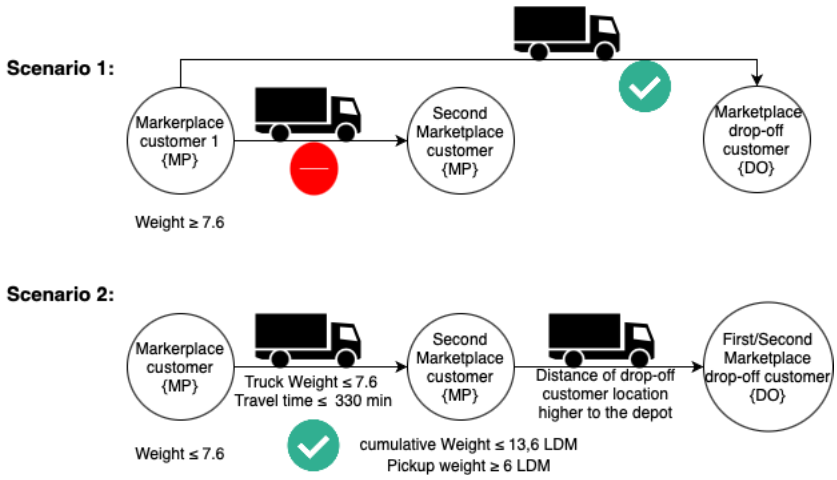

Figure 6.4 – Stage 4

Stage 4 is the flow from a marketplace location into the drop-off of the marketplace location as shown in **Figure 6.4**. Stage four will always be feasible. As known, the travel to the second marketplace location in stage 3 is not mandatory due to the weight, distance, and time constraints. Therefore, stage 4 has two scenarios, if the truck has not travelled through stage 3 then the truck will travel from the first marketplace to its drop-off location. However, if the truck travelled to stage 3 then the truck will deliver the load to one of the marketplace drop-off locations based on the objective function, which is minimizing the distance between the final marketplace drop-off location and depot. For example, if the distance between the first marketplace drop-off location and the depot is 70 kilometres and the second marketplace location and the depot is 50 kilometres then the truck will travel in stage four to the first marketplace drop-off customer. Constraint 13 ensures that if there is an arc in stage two then there should be an arc in stage 4.

$$\sum_{\substack{i \in MP, \\ j \in DO}} x_{ij4} \leq \sum_{\substack{i \in MP, \\ j \in MP \\ i \neq j}} x_{ij2} \qquad \text{C 13}$$

Stage 5:

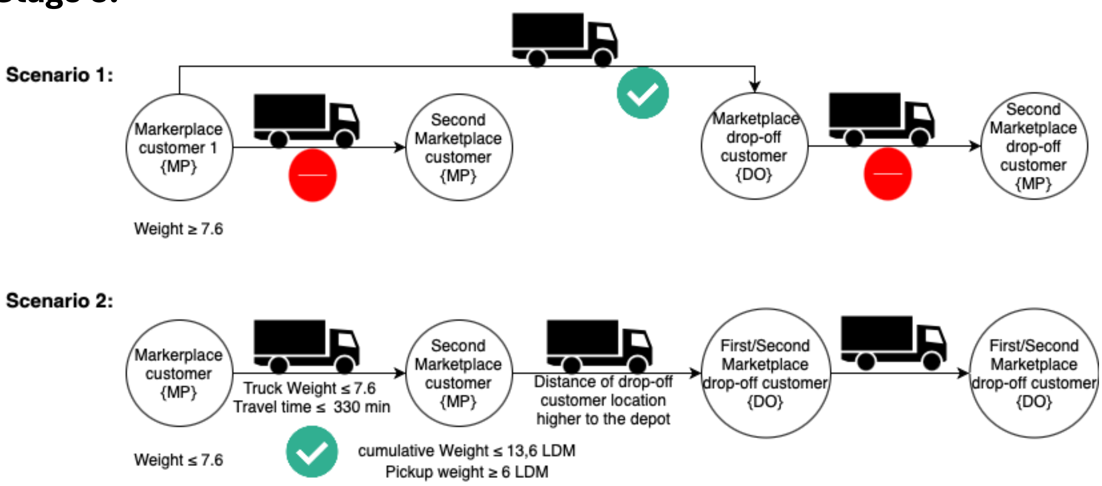

Figure 6.5 – Stage 5

Stage 5 show the flow of the truck from the first marketplace drop-off location to the second marketplace drop-off location as shown in **Figure 6.5**. Constraint 14 states if there is an arc in stage three, then there should be an arc in stage five. However, if only one marketplace was visited then the arc x_{ij3} will be zero that will also enforce x_{ij5} to be zero.

$\sum x_{ij5} \leq \sum x_{ij3} \qquad i \in MP, j \in MP, i \neq j$ C 14

Stage 6:

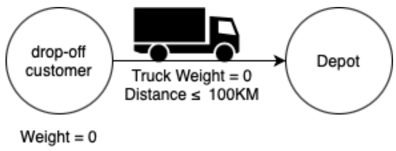

Figure 6.6 – Stage 6

Stage six is the final stage. In stage six as shown in **Figure 6.6**, the truck travels from the final drop-off location to the depot back in Deventer. The distance between the drop-off and depot is limited into 100 kilometres and the truck should return empty.

Constraint 15 operates as constraint 6; however, in this case, the goal of constraint 15 is to ensure that the vehicles will use a route back to the depot. Constraint 16 checks if there was a visited route in stage four which states that there is a vehicle traveling from a marketplace to a drop-off location, then a route will be formed to travel back to the depot. Constraint 16 is part of the connectivity constraints. Similarly, in constraint 17, if there is an arc between a drop-off location denoted by DO to the depot, then node i must be visited in stage 6. Constraint 18 is a distance limiter constraint that limits the distance between the drop-off locations and the depot to a radius of 100 kilometres. If a customer has a radius of less than 200 kilometres in stage 2 and then has a radius of more than 100 kilometres in stage 6, then the customer will be neglected.

$\sum x_{i06} = 1 \qquad i \in DO$ *C 15*

$\sum_{i \in DO} x_{i06} \leq \sum_{\substack{i \in MP, \\ j \in DO}} x_{ij4}$ C 16

$\sum x_{i06} = y_i \qquad i \in DO$ *C 17*

$\sum d_{i0} \times \sum x_{i06} \leq 100 \qquad i \in DO$ *C 18*

6.2 Model implementation

The model was implemented using python, and all experiments were run using Macbook Air with a processer of 1,1 GHz Quad-Core Intel Core i5 and Memory of 8 GB of RAM. The model was developed in python using several libraries of Pulp, Pandas, geopy.geocoders, and googlemaps. Pulp is well known library that is used to solve MILP models and was used to solve the objective function of the model. Moreover, Pandas was used to collect the Excel sheet data set and the geopy.geocoders converted the postal and country codes of the marketplace into latitude and longitude. Finally, the googlemaps library was used to collect the google maps API to compute all the distances and travel times. A sample of the code can be shown in **Appendix B**.

6.3 Validation of models

To ensure the validity of the models, construction of a sample data that can be solved manually using the MILP model and then testing the programs with the same data set to evaluate the results and validate the results in term of similar outputs between the manual and program results.

The sample data set as shown in **Table 6.3** consists of depot, one linehaul customers represented by the shipment ID, while **Table 6.4** represents the multiple marketplace locations and their drop-offs produced manually for this evaluation. Each marketplace customer has different specifications in term of weighting meters**,** distance between the first linehaul customer marketplace customer (Stage 2 – MILP), distance between the final drop-off location and the returning depot (stage 6 – MILP), all dates are as the same day of the arrival to the first linehaul customer as we assumed that the driver will always arrive at the beginning of the working day (08:00) and needs to accomplish all pickups from the marketplace at the same working day till (20:00). The evaluation will be conducted through building different scenarios to cover all possible outcome and the distances will be changed according to each scenario's specification.

Table 6.3 – Sample Vos data

Shipment_Number	StartingDate	Depot	Returning_Depot	Unload_Customer	Unload_City	Unload_Latitude	Unload_Longitude	delivery day
JUT22563	01/01/2022	DEVENTER	DEVENTER	K05206	MIDDELFART	55.50409	9.75568	02/01/2022

Table 6.4 – Sample marketplace data

MP CUSTOMER ID	Country code	postal code pu	city pu	Date of pickup	meters/m	BH CUSTOMER ID	Country code	postal code	city df
MP1	DE	18439	Stralsund	02/01/2022	6	BH1	NL	7811AP	Emmen
MP2	DE	24103	Kilonia	02/01/2022	6	BH2	NL	8442 CH	Heerenveen
MP3	DE	24937	Flensburg	02/01/2022	13.6	BH3	NL	7811AP	Emmen

6.3.1 First scenario – Distance

The first scenario is distance-based validation, therefore, the goal in this scenario is to test constraints 9 and 10, and the objective function in the model section. Constraint 8 ensures that any distance between the linehaul customer and marketplace customer should not exceeds 200 kilometres, while constraint 18 ensures that the drop-off of the marketplace customer should be in a radius of 100 kilometres with the depot. The objective function goal is to minimize the driven kilometres in stage 2 and 6, through choosing the marketplace customer with the least empty driven kilometres. Hence, **Table 6.5** shows five semi-scenarios of the distance-based validation. In the test column of **Table 6.5**, the matrix test column, represent the distance between the linehaul and the marketplace customer, and the distance between the marketplace drop-off and the depot. The Matrix text column includes the matrix tables. The matrix table column and row abbreviations are as the following:

- MP – Marketplace location
- DO – Marketplace drop-off location
- D.L – Distance between the linehaul and the marketplace
- D.D – Distance between the depot and the marketplace drop-off location

Matrix test	Goal	Expected outcome	Model outcome
$\begin{array}{lcc} & D.L & D.D \\ MP1 & 51 & \\ MP2 & 50 & \\ DO1 & & 50 \\ DO2 & & 50 \end{array}$	Choose the minimum distance between the linehaul and the marketplace pick-up.	MP2	MP2
$\begin{array}{lcc} & D.L & D.D \\ MP1 & 50 & \\ MP2 & 50 & \\ DO1 & & 51 \\ DO2 & & 50 \end{array}$	Choose the minimum distance between the Depot and the marketplace drop-off.	MP2	MP2
$\begin{array}{lcc} & D.L & D.D \\ MP1 & 210 & \\ MP2 & 190 & \\ DO1 & & 10 \\ DO2 & & 90 \end{array}$	The goal is to exclude MP1, since it is exceeding the 200 KM constraint.	MP2	MP2
$\begin{array}{lcc} & D.L & D.D \\ MP1 & 100 & \\ MP2 & 30 & \\ DO1 & & 50 \\ DO2 & & 101 \end{array}$	The goal is to exclude MP2, since it is exceeding the 100 KM constraint between the backhaul and the depot	MP1	MP1
$\begin{array}{lcc} & D.L & D.D \\ MP1 & 300 & \\ MP2 & 50 & \\ DO1 & & 50 \\ DO2 & & 200 \end{array}$	In this case, no customer will be selected, and the model will return a statement of “No shipment was found”	-	-

Table 6.5 – First scenario

6.3.2 Second Scenario - Weight and time

The second scenario, the weight constraints and the time constraints will be tested. The weight constraint 1, 2 and 9 differs between the minimum pickup weight and the maximum load of the vehicle. Hence, the vehicle minimum pick-up load is 4 LDM and maximum capacity of 13,6 LDM. Afterwards, the weight constraint will be test for cases that include the objective function to compute the minimum empty driven kilometres based on the weights of the customer. Thereafter, the time constraint 12 will be tested, the time constraint tests the time difference between the first marketplace and the second marketplace location and shown in the matrix test column in **Table 6.6**. This testing scenario will include all customers for the sample data. The matrix table column and row abbreviations are as the following:

- MP – Marketplace location
- DO – Marketplace drop-off location
- D.L – Distance between the linehaul and the marketplace
- D.D – Distance between the depot and the marketplace drop-off location
- W – Weight of the shipment
- T – Time to reach the second marketplace

Matrix test	Goal	Expected outcome	Model outcome
$\begin{array}{lccc} & W & D.L & D.D \\ MP1 & 13{,}6 & 51 & \\ MP2 & 6 & 50 & \\ DO1 & & & 50 \\ DO2 & & & 50 \end{array}$	Choose the minimum distance between the linehaul and the marketplace pick-up. Even if the maximum load is not reached	MP2	MP2
$\begin{array}{lccc} & W & D.L & D.D \\ MP1 & 13{,}6 & 51 & \\ MP2 & 13{,}6 & 50 & \\ MP3 & 13{,}6 & 100 & \\ DO1 & & & 50 \\ DO2 & & & 50 \\ DO3 & & & 50 \end{array}$	Choose the minimum distance between when all have the same weight.	MP2	MP2
$\begin{array}{lccc} & W & D.L & D.D \\ MP1 & 13{,}6 & 200 & \\ MP2 & 6 & 50 & \\ MP3 & 6 & 100 & \\ DO1 & & & 100 \\ DO2 & & & 60 \\ DO3 & & & 50 \end{array}$	The goal in this test is to test the path of the stages. In the test the model should firstly either chose MP1 or the combination of MP2 and MP3 based on the distance, and in this case, it is shown that the combination of MP2 and MP3 will result in less empty driven kilometres. Next, the model should choose the minimum distance between the linehaul and the marketplace location, similarly with depot.	L->MP2 MP2->MP3	L->MP2 MP2->MP3
$\begin{array}{lccc} & W & D.L & D.D \\ MP1 & 13{,}6 & 200 & \\ MP2 & 4 & 50 & \\ MP3 & 6 & 100 & \\ DO1 & & & 100 \\ DO2 & & & 60 \\ DO3 & & & 50 \end{array}$	The goal is to remove MP2 from the decision list since the minimum weight of MP2 is less than 6. Then, the comparison is between the MP1 and MP3.	MP3	MP3

T W $D.L$ $D.D$ $MP1$ 60 13,6 200 $MP2$ 20 4 50 $MP3$ 370 6 100 $DO1$ 100 $DO2$ 60 $DO3$ 50	Since MP3 has a time larger than 330 minutes, then MP3 is excluded, and the comparison is between MP1 and MP2	MP1	MP1

Table 6.6 – Second scenario

6.4 Conclusion

In this section, the mathematical model was constructed and integrated in python and finally validated using scenario-based approach. The MILP was constructed using 6 stages describing different stages of the transportation cycle. Afterward, the model was integrated in python and the program provided results that was test in the several scenarios to validate the accuracy of the model.

7. Evaluation

In the evaluation phase, the built model will be evaluated based on the selected KPI's derived in the business understanding phase to answer the sub-research question of "how the performance of the new model affects the KPI's?" in **Chapter 7.2**. Furthermore, the planners will be evaluated with and without the visualization tool for the same period to check the effectiveness of the models. **Chapter 7.1** visualize the built graphical user interface that will be used in the experimental testing with the planners in **Chapter 7.3**.

7.1 Historical Results

To analyse the 141 shipments in Denmark that were returned using the marketplace in 2022, the analysation was initiated with evaluating the difference between the average empty driven kilometres. The results have showed that the average empty driven kilometres for the 2022 planner decisions resulted in **65** kilometres with an average of **4,0%**, while the model resulted in **89,6** empty driven kilometres with an average of **5,36%**. Fragment of the results are shown in **Table 7.1**. After reviewing the empty driven kilometres per trip, the review showed that the data with the planner decisions only included the empty driven kilometres for the stage between the linehaul customer and the first marketplace location. The planners' results excluded the stage between the final drop-off shipment and the depot. This is caused because the planners return the trucks to the depot, before delivering to the drop-offs and checks other trucks that will travel in the same route to the drop-off location and drop the shipment. The planners with this strategy converted the case-study from a single vehicle routing problem into a multi trip vehicle routing problem. Therefore, to have a fair comparison between the planners and the model, the model only calculated the empty driven kilometres for the stage between the linehaul customer and the marketplace; the results shows that the model resulted with 49,5 average empty driven kilometres and a percentage of 3% which reduced the planners' average empty driven kilometres with 16 EDK per trip. When evaluating the empty driven kilometres for the whole trip of the model, the average difference is sensible to state that the model optimized the routes for the 2022 historical data.

Table 7.1 – Empty driven kilometres fragment results

	Model / 2022	Planners / 2022	Model/2022
Shipment ID	Empty driven kilometres	Empty driven kilometres	Empty driven kilometres/Stage 2
JUT22566	163.949	56.28	93.13
JUT22617	96.882	24.12	63.717
JUT22621	108.004	11.39	44.153
JUT22626	149.678	17.42	77.72
JUT22629	96.882	2.68	63.717
JUT22646	114.704	188.94	64.32
JUT22666	62.511	57.62	10.653
JUT22683	62.511	57.62	10.653
JUT22689	107.937	76.38	44.019
JUT22764	80.869	48.91	7.303
JUT22840	79.864	56.95	10.117
JUT22858	112.895	17.42	61.037
Average	89.63349761	65.78822967	49.5935273

Furthermore, **Figure 7.1** shows a line chart comparing the model results and the planners' decisions. The chart shows that the model results have no peaks or outliners, and all empty driven kilometres are in a range between 50 and 250, while the planners' decisions show several peaks and outliners, and the empty driven kilometres are between 2 and 512 kilometres. Therefore, the planners' decisions resulted in high level of volatility, this is due to the strategy in selecting the marketplace locations. The planners used to select a marketplace customer who's location is in the route between the linehaul customer and the depot, not considering the empty driven kilometres. While the model results show consistently in the empty driven kilometres, which will decrease the empty driven kilometres in the long term.

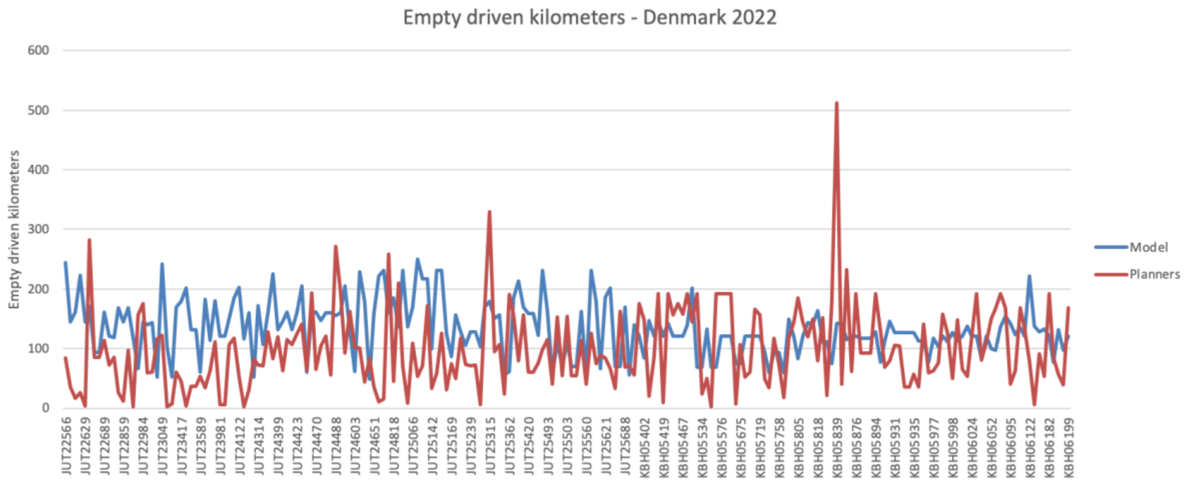

Figure 7.1 – Vos 2022 planners Vs model

7.2 Graphical user interface

The graphical user interface built in Python to visualize the results of the model is divided into three pages. The first page shown in **Appendix C** provides the place where the user enters the shipment ID. While, **Figure 7.2** represents the second page which shows the stages taken and the routes of each stage with their distances and travel time. Moreover, it shows the empty driven kilometres for the shipment ID. Finally, the third page in the **Appendix C** provide all of the customer information, starting with the marketplace customer ID, marketplace customer, number, email, latitude, longitude, and the shipment loading meter.

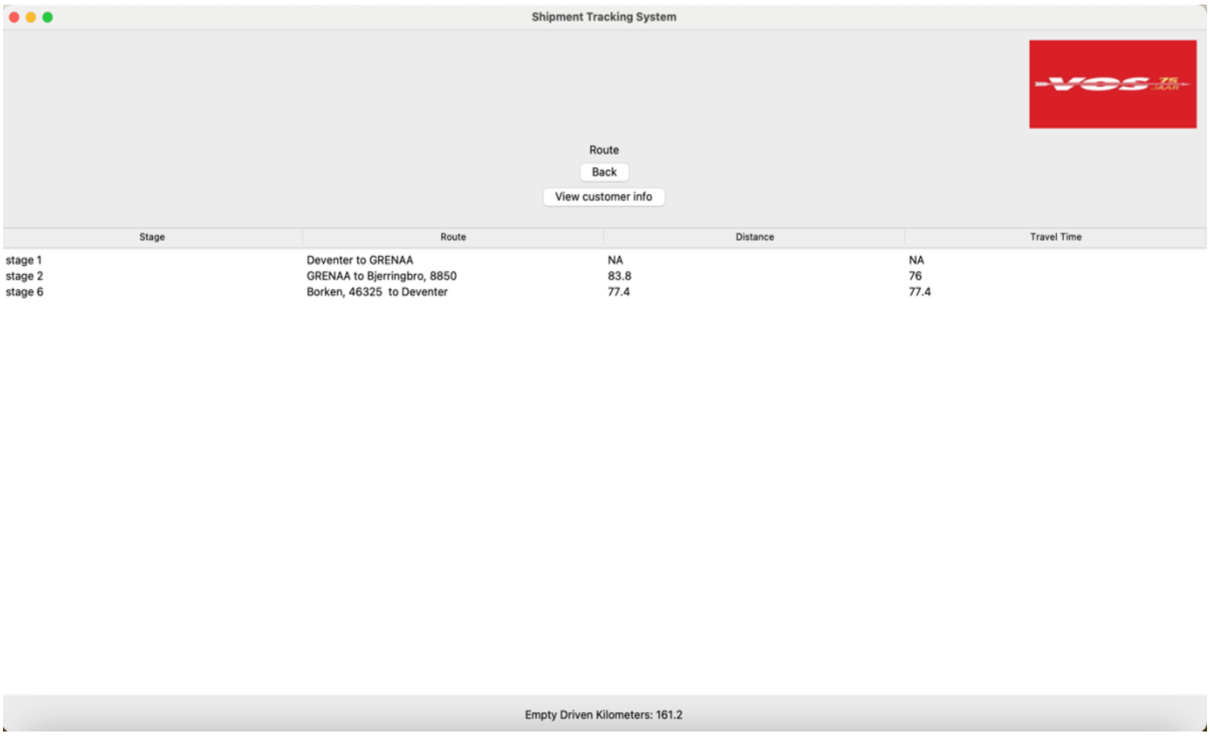

Figure 7.2 – Vos Visualization tool page 2

7.3 Experimental testing results

Experimental testing was conducted with the planner in Vos, the planner searched for the shipments through the scripted marketplace data gathered in the data preparation. The results show the marketplace customer selection, empty driven kilometres, and the computation/searching time of the model and planner. To evaluate both results, the model shows an optimization in the empty driven kilometres and reduces the empty driven kilometres with approximately 50 empty driven kilometres than the planner decisions. The searching time of the model seems to be low, however, this computation time is after calculating all the distances between all arcs using google maps API's and then stored as a dictionary in Python. Therefore, the actual computation time will be one hour of computing the distance plus the model computation time. Hence, the model computation time is a strong disadvantage to the model. The experimental testing results are shown in **Table 7.2**.

Table 7.2 – Experimental testing results (model vs planner)

Live testing	Model results			Planner results		
Shipment_Number	Marketplace customer	Empty Driven Kilometers	computation time	Marketplace customer	Empty Driven Kilometers	searching time
JUT22617	MP56	144.6	00:03:00	MP20	199.1	00:01:27
JUT22621	MP92	161.2	00:00:15	MP24	225	00:04:49
JUT22646	MP92	171.2	00:00:12	MP21	189	00:01:20
JUT22666	MP92	93.3	00:00:13	MP89	145.6	00:04:02
JUT22764	MP92	120.7	00:00:14	MP22	194	00:02:01
JUT22840	MP92	119.2	00:00:13	MP24	162	00:00:58
JUT22858	MP92	168.5	00:00:12	MP14	211	00:05:42
JUT22924	MP92	117.6	00:00:14	MP79	202.6	00:01:26
JUT22985	MP92	140.2	00:00:13	MP24	187	00:03:51
AVERAGE		137.3888889	**00:00:32**		**190.5888889**	**00:02:51**

7.4 Conclusion

In this section, the results of the model show the resulting graphical user interface, that is used in the experimental testing with the planners. The experimental testing showed that the model could produce results lower than the planners' decisions with approximately **52** fewer average empty driven kilometres. Though, the empty driven kilometres are less when using the model, the computation time of the model is higher than the planners' searching time. Afterward, the main research question was evaluated through testing the empty driven kilometres of 2022 with the model and compared with the planners. The results showed that the model produced higher empty driven kilometres. However, the model calculated the empty driven kilometres for two stages of when the empty trucks are driving to the pickup location and driving emptily back to the depot. While the 2022 empty driven kilometres only revealed the first stage, hence, Vos trucks returned to the depot before delivering the pickups. Furthermore, with a comparison of the same conditions, the model produced fewer empty driven kilometres with approximately 16 empty driven kilometres per trip.

8. Deployment plan and Framework

In this chapter, the deployment plan will be formulated to assess the company with deploying the built model of this thesis and a framework is constructed to assess the logistics sector with an initial plan when working with empty driven kilometres. The deployment plan will address the sub-research question of How can a Scrum framework assist in the deployment plan? **Chapter 8.1** will introduce the SMART framework that will be integrated in the Scrump framework. **Chapter 8.2** shows the formulation of the scrum framework with the integration of the SMART framework in the Scrum sprints. Finally, **Chapter 8.3** will show the Framework aiming to reduce empty driven kilometres when using the marketplace.

8.1 SMART

To formulate the goals of the deployment, the SMART framework is used to guide the deployment plan. SMART was developed by George Doran, Arthur Miller, and James Cunningham in 1981; the goal of the SMART framework in this thesis is to create the goals of each sprint of the scrum framework.

Starting from the S for specific, the deployment scope is defined to have goals that can be divided into the scrum sprints. The goal of the deployment plan is to prepare the necessary equipment (such as the APIs), integrate the APIs into the model, test the model with the APIs as initial testing, deploy the model and APIs into the company cloud servers, and finally test the model with live data with evaluating the performance of the model. Secondly, the M for Measurable, each sprint in the scrum will have a criterion to evaluate the progress of the sprint. Thirdly, the A is Achievable, ensuring that all the resources and time are available for the deployment. The company has already invested in this project with the model and APIs; however, due to the limitation of the time frame in this thesis, the deployment was not conducted. Thus, the company already has most of the resources and time to conduct the project. Furthermore, the R for Relevant, the sprints, as mentioned in the specific, are aligned with the overall goal of the deployment; each sprint contributes to assessing an objective to reach the final goal of deploying the model into the company cloud servers. Finally, the t for Time-based, each sprint in the scrum will have a timeline of 4 weeks, except the test of the model with live data, this phase should be evaluated for at least six months to evaluate all aspects of the results.

8.2 Scrum framework

Scrum according to (Theisens, 2022) is a product development approach used in software departments. A product development project that is constructed in applying the Scrum and the progress is measured through several iterations called the “Sprints”. The main advantage of scrum is the flexibility in handling changes within the sprint and aiming on delivering an operational version of the required goal at the end of each Sprint. Six sprints are planned to execute the deployment. The sprints are as the following:

1. Sprint Preparation:
    a. Specific:
        i. Research for the necessary APIs, such as the marketplaces APIs and distance calculator APIs.
        ii. Contact the marketplace websites to purchase the API codes.
    b. Measurable:
        i. The research should state why the marketplace website is important based on its popularity and usage in the targeted countries.

    ii. The research should plan which distance calculator API should be used and a cost plan.
  c. Achievable:
    i. The cost of the marketplace APIs should be discussed in this phase, and how many APIs should be acquired should be planned. Currently, the company already has one marketplace website API. Therefore, the company already has one resource.
    ii. As seen in this Thesis, the cost of computing all distances using Google Maps API is high. Therefore, the company should consider checking the inaccuracy of finding the distances using Euclidean distance method. The Euclidean method will not cost the company any financial resources.
    iii. Time for this research is necessary.
  d. Relevant
    i. The research should answer all questions before the deployment and can be used as a guide for the following sprints.
  e. Time-based
    i. This phase should be conducted in eight weeks. Time is necessary, especially when contacting the API suppliers.

2. Sprint Integration of the API
  a. Specific:
    i. Integrate the APIs into the Thesis model.
    ii. Test the APIs with different scenarios and check if all model constraints fulfil the tasks as shown in **Chapter 6.3**.
    iii. Either Integrate an API to calculate the distances or integrate the Euclidean distance method.
  b. Measurable:
    i. The model should be functioning with the APIs.
    ii. Testing of the model result should be equal to the expected results.
  c. Achievable:
    i. Time is the only necessary resource for this sprint.
  d. Relevant:
    i. The Sprint should result in a working model that can be deployed in the company servers.
  e. Time-based:
    i. This phase should be conducted in eight weeks. Time is necessary when integrating the APIs into the model.

3. Sprint Deployment of the model into the company cloud servers:
  a. Specific:
    i. Deploy the model into the company cloud servers.
  b. Measurable:
    i. The Company should have a working model integrated into the company server.
  c. Achievable:
    i. Time is the only necessary resource for this sprint.
  d. Relevant:
    i. This sprint should result in the main objective of the deployment, which is the deployment of the model into the company server.
  e. Time-based:
    i. This phase should be conducted in eight weeks.

4. Sprint Test the model with live data:
    a. Specific:
        i. The scope is to test the model using live data and resources.
    b. Measurable:
        i. Test the model similar to **Chapter 7.3**, however, for a longer period to increase the reliability of the results.
    c. Achievable:
        i. Time is necessary from the IT-department and the operational department.
    d. Relevant:
        i. This final sprint should result in a decision on whether the company should use the model in the operational department based on the live testing results.
    e. Time-based:
        i. This phase should be conducted in 6 months to gain accurate and reliable results before fully functioning on the model.

8.3 Framework to reduce empty driven kilometres

A framework to reduce empty driven kilometres when using the marketplace is constructed to support the logistic sector in starting new projects with a built framework. The framework shown in **Figure 8.1** shows 3 steps that should be considered when planning to reduce empty driven kilometres through technical support. The framework starts with understanding the company current process and plan a desired process to formulate a case study similar to the Business understanding phase in the CRISP-DM methodology. Next, using literature the company needs to search for possible solution approach that suits the case study. Afterward, the company should search for the active marketplace websites in the desired region to contact and purchase the marketplace API. Resulting in the collection of both the company historical data and the marketplace required data, the data needs to be cleaned from errors and irrelevant data.

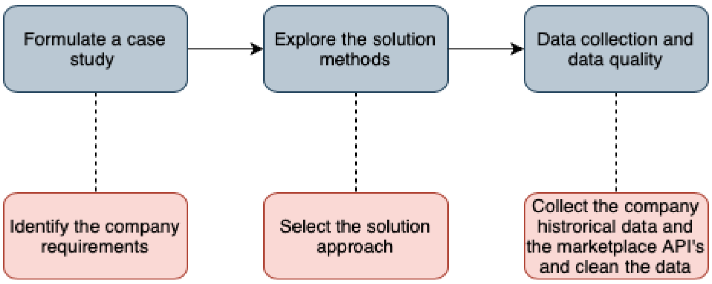

Figure 8.1 – Framework to reduce empty driven kilometres

Formulate a case study:
In the first phase, the company needs to understand the issue they need to solve. Therefore, a case study of the current process needs to be investigated and the improvement areas will be determined. The case study will visualize how the flow of the current process and where are the empty driven kilometres distributed. As the problem is related to empty driven kilometres, then the company can use the vehicle routing problem as the problem identification. The vehicle routing problem (VRP) has several variants that can be sorted depending on the company case study. As stated in chapter three, the variants of the VRP can be the time-constraint, capacity, backhauls, multi-vehicle, single vehicle, and multi-trips. Each variant represents a feature of the

process and depending on the company process, the VRP problem identification can be deduced.

Explore the solution methods:
After introducing the problem identification, the goal of the second phase is to search for a solution method. The solution method should be chosen according to the company requirements. Several solutions are available to solve the vehicle routing problem, such as the exact, approximate, and hybrid approaches. If the company is searching for an approach that will result in a highly accurate solution with flexibility in the modelling, then the exact approach is right solution. However, the major disadvantage of the exact approach is the computation time, as the exact approach is considered to be NP-hard approach, especially when using the mixed integer linear programming. Moreover, approximate approach is widely used in solving vehicle routing problem. Approximate approach includes the use of heuristics or meta-heuristics as the solution algorithm, the best-known approximate approach algorithm is the adaptive large neighbourhood search because of its accuracy in solutions and fast computation time. Finally, hybrid approach is the use of two algorithm solutions to acquire the advantages of both solutions.

Data collection and data quality:
The third phase is to search for and collect available data. To reduce empty driven kilometres using the marketplace, at least two data sets are required. Firstly, the first data set is the available marketplace websites and how active they are in the targeted country. For example, Denmark has around seven marketplace websites; however, only three websites are active. While In Finland, only one website was active with marketplace offers. The marketplace offers is necessary to search for the best route, and the more routes, the better chance of finding the optimal route. To conclude, the company should check the targeted country and begin in searching for the active marketplace websites in that country. Secondly, the company needs to analyse the historical data of the company's performance in the specific country. As in this Thesis, a major issue was detected in the analysis of the historical data. The company wanted to model a MILP for the trucks traveling to Finland. However, the analysis of the historical data showed that the company only used the marketplace in Finland once in one year. Therefore, the analysis resulted in neglecting the modelling for Finland because of the low usage of the marketplace. In conclusion, the analysis of historical data will determine how necessary you need to spend resources in building a solution for an existing issue, and it can be used to support the formulation of the case study by supporting the statements in that phase.

8.4 Conclusion

The deployment plan assists the company with deploying the thesis-built model. The deployment plan used Scrum methodology and integrated the SMART framework in the Scrum sprints. Four sprints were formulated assessing different stages of the deployment. The first sprint is the preparation, and it lets the company start implementing the recommendations of this thesis and searching for the effective API's. In stage 2, the company will implement the APIs in the model and validate the model with the scenario-based approach. Afterward, in stage 3 the company will then deploy the model in the cloud servers to finally test the model in stage 4.

9. Conclusion

The following section will discuss the conclusion, limitations, recommendations, and further projects of this thesis. Starting in **Chapter 9.1**, the research and sub-research questions are discussed, and a conclusion was deduced for each question. Next, the limitations that have restricted this research is discussed in **Chapter 9.2** and recommendations are stated on how to improve the built mathematical model in **Chapter 9.3**. Lastly, further projects that can continue in contributing to the thesis in indicated in **Chapter 9.4**.

9.1 Conclusions

The main research question formulated in the introduction of the thesis aims to minimize the empty driven kilometres for Vos truck when using the marketplace in Finland and Denmark to reach the norm of the company. Twelve sub-research questions were expressed to assist answering the main research question.

Sub-research question 1: What are the current decision-making criteria for selecting the loading location in marketplaces?

The first sub-research question was discussed in the business understanding section. To understand the current process of the planners', it was visualized in a process map. The process map showed that the planners spend excessive time and workload when searching for the pickup customer using the marketplace website for the empty truck in Denmark or Finland. The current process showed that the planners have a target of at least 30% cost return from the truck heading back to the depot and a 70% cost return from the truck heading to the unloading location (main customer). This target provides the planners with freedom when searching for a loading location, therefore, there is a large opportunity of selecting deficient loading locations.

Sub-research questions 2 and 3: Which KPI's can be used to evaluate the new model? Which criteria are necessary when selecting new models?

Both sub-research questions 2 and 3 contributed to the same operation. To answer both questions, firstly, the stakeholders were identified, and a desired process map was constructed based on the stakeholders' desires. The desired process shows that Vos's objective is to construct a technical tool that will search for the loading locations on the marketplace website, aiming to minimize the empty driven kilometres and decrease the planners' workload. The specification of the desired technical tool should involve in selecting the minimum empty driven kilometres when searching for marketplace customers, maximum truck load of 13,6 LDM, minimum pick-up load of 6 LDM, and loading time window of the same day of delivering to the linehaul main customer. Afterward, the criteria of the technical model should evaluate both the results of the empty driven kilometres and the computation time of the model. The company KPI's objective is to reach the norm of the main research question based on the empty driven kilometres, and to have a computation time that is less the one minutes.

Sub-research question 4: Which Vehicle routing problem variant in literature contest with the company's process?

As answered in the sub-research questions 2 and 3, the company's desired process includes vehicle capacity and time-window. The literature did not provide an exact problem formulation for the specific company's case; however, each variant of the vehicle routing problem was explored to find the vehicle routing problem variant of the company. The company's vehicle routing problem is stated as a single vehicle routing problem with vehicle capacity, time window, and backhaul customers (SVRPCTWB).

Sub-research question 5: Which solution approaches are in the literature which can be used in optimizing routes?

To answer sub-research question 5, three main Solution approaches were found. The first solution approach is the exact approach, which aims to find the exact optimal solutions. The exact approach consists of the linear programming and deterministic dynamic programming. As the exact approach aims to find the optimal solutions, it searches for every possible arc to check the distance difference, resulting in high computation time. Secondly, the approximate approach consists of the heuristic traditional algorithms and the meta-heuristic algorithms. The approximate approach aims to find approximately optimal solution with having a low computation time. However, several variants in the approximate approach does not provide sufficient outputs. The finest heuristic algorithm based on several research papers is the adaptive large neighbourhood search, because it provides approximately exact outputs and has a very low computation time. Finally, the hybrid approach is the combination of two solution methods to manage the advantages and disadvantages of the solution methods.

Sub-research question 6: What is the most effective solution approach to solve vehicle routing problems and decrease empty driven kilometres?

According to several researchers, the best approach to minimize a cost of a variable is by using the adaptive large neighbourhood search (ALNS), however, due to the limited knowledge the researcher of this thesis have, the ALNS was not chosen as the solution method for this thesis. On the other hand, the mixed integer linear programming (MILP) is commonly used in several papers that aim to minimize the distances. Therefore, the MILP was chosen as the solution method and the main constraints to function the model is gathered from the literature.

Sub-research questions 7 and 8: Which marketplaces are available to script customers for both Finland and Denmark? What available data will be used in the mathematical model?

The marketplaces available in Denmark and Finland are lkw, trans.eu and Timocom. Two data sets were extracted to assess the mathematical model. Data set A contains the historical data from Vos data bases, while Data set B contains a manually scripted marketplace customer locations from the aforementioned marketplace websites. Data set A showed several defects, irrelevant data and missing data. While, Data set B were scripted according to the necessity of the mathematical model, therefore the cleaning of Data set B was not required.

Sub-research question 9: How can cleaning the data ensure the validity of the data?

Cleaning the data was conducted on Data set A. Data set A contained 3784 shipments in Denmark and 111 shipments in Finland. After cleaning Data set A, the shipment numbers drastically decreased from 3784 shipment to 141 shipment in Denmark, while in Finland the shipments decreased from 111 to 1 shipment. Therefore, due to the limited shipments in Finland and lack of data, it will not be included in the mathematical model.

Sub-research question 10: What are the characteristics of the solution model?

The solution model chosen in the literature section is the mathematical integer linear programming and the goal is to assess the single vehicle routing problem with capacity, time-window and backhaul customers. The MILP started with constructing 6 stages and applying different constraints to each stage. Stage one presents when the truck is traveling from the Depot in the Netherlands to the linehaul standard customer in Denmark, there is no constraint in this stage; however, there an assumption that all trucks arrive in the linehaul standard customer at 8 am of the arriving date. Stage two is the main purpose of the MILP, where the MILP search for the optimal marketplace customer location based on the distance from the standard customer to the pickup location and the drop-off location to the depot, also applying the capacity maximum and minimum constraints of 13,6 and 6 LDM. In stage three, if the truck has loaded a capacity of less than 6,8 and more than 6 LDM, then the truck can also pick another shipment. However, the second pickup location should have a pickup load that does not exceed the maximum cumulative capacity, and have a shipment load bigger than 6 LDM. Moreover, in stage three, 330 minutes are the maximum travel time between the first marketplace location and the second marketplace location. If all constraints are acceptable, then the model will select the option with less empty driven kilometres. In stage four, the truck travel from the selected marketplace customer pickup location to the drop-off location. In stage five, if there is a second pick up customer, then the delivery of either the second or the first pickup location will be made based on the distance from the depot location. In this case, selecting stage six is based on the marketplace customer drop-off that has less empty driven kilometres to the depot. In general, the model was validated using scenario-based approach and showed exact result compared to the expected results of each scenario.

Sub-research question 11: How does the performance of the new model affect the KPI's?

The KPI's selected in the business understanding section are the empty driven kilometres and the computation time of the model. The first performance of the KPI showed that the model resulted in a higher empty driven kilometres than the planners' decisions in 2022. However, the model calculated the empty driven kilometres for two stages that consider the trucks traveling to the marketplace pickup location and traveling back to the depot from the marketplace drop-off location. While the planners' decisions included only one stage which considered travelling only to the marketplace pickup location and then traveling back to the depot before delivering the marketplace shipment. However, a comparison with the same conditions was conducted and resulted in reducing the empty driven kilometres with approximately 16 EDK per trip. Secondly, an experimental testing was conducted between the planners' and the model using 9 shipment ID's and the same marketplace customer location. The experimental testing showed that the model resulted in having 53 fewer average empty driven kilometres than the planners' decisions. The second KPI showed that the computation time of the model is higher than the desired computation time. As the desired computation time is less than one minutes, the model computation time is almost one hour. The high computation time is resulted from computing every possible distance and time between all arcs; though the computation time is high, there is several approaches to decrease the computation time.

Sub-research question 12: How can the Scrum framework assist in the deployment plan?

A scrum framework was formulated to assist the deployment plan. Scrum was integrated with SMART framework to evaluate every sprint in the deployment plan. Four sprints were formulated, while every sprint is a continuous of the previous sprint. Overall, the deployment will require at least a year to understand the effectiveness of the model and deploy the model with confidence of making the model able to select marketplace customers.

9.2 Limitations

In this sub section, the limitations encountered while performing the thesis will be listed. The limitation effected the end results of the thesis, therefore, will be discussed.

1. Limited data in Finland

As aforementioned, after cleaning data set A, Vos only had one shipment travelled back from Finland using the marketplace, therefore, building a model that will apply for the specification of Finland was not applicable. Vos uses ferry's when travelling to Finland and several extra stages were going to be constructed for the use of the ferry's and to evaluate the results and test the validity of the model, data was required.

2. Time limitation

Time was required to deploy the thesis model into the company cloud servers, apply the marketplace API's and conduct live testing of the model. due to the time limitation only experimental testing was conducted using scripted marketplace data and old shipment ID's. The experimental testing provided insight to the performance of the model; however, more time is required to test the model and evaluate the results. Deployment plan due to the time limitation using scrum methodology and SMART framework to assess the company with deploying the built model.

3. Lack of knowledge

Knowledge was required to construct the most effective solution approach based on the literature review. However, as the most effective solution approach which is the adaptive large neighbourhood search require high knowledge, it was not valid to construct it. Therefore, the MILP was constructed as there is no lack of knowledge, flexibility to assess the company requirements, and providing exact solution.

9.3 Recommendations

To begin, the Thesis results were successful in terms of reducing empty driven kilometres, however, the model computation time was very large due to acquiring the distances for every possible arc from the google maps API. To reduce the computation time, Euclidean formula can be used to measure the distances between each arc, the Euclidean formula will reduce the computation time, though the results will not be accurate.

Moreover, the thesis was conduct toward single VRPTWCB, while the company currently returns all shipments to the depot and deliver the marketplace drop-off using different vehicles, converting the problem into multi VRPTWCB, and if the single vehicle resulted in a very large computation time using the MILP models, then multi vehicle will result in a much higher computation time. The company may explore the approximate approach of meta-heuristics especially the adaptive large neighbourhood search. Approximate approach is a well know approach in reducing severally the computation time, however, without resulting in an accurate result. though, the average difference between the MILP model results and the planners' results is large, in my opinion, approximate results will also result in lower empty driven kilometres.

Moreover, the historical analysis cannot provide accurate results due to two factors. The first factor is the marketplace customers are different between the model testing and the planners' decision, though the marketplace customer locations were compared and were in the same range. The results of the historical analysis can be seen as a starting result to either believe or not in the effectiveness of the model. the second factor is that the trucks in 2022 returned to the depot before delivering the drop-off, which contradict with the model approach.

Furthermore, the experimental testing also shows that the model provides an effective output in term of empty driven kilometres when comparing with the planners' decision with the same inputs and circumstances. However, the testing was only conducted for one day and only nine shipments, therefore more testing needs to be conducted before the deploying of the model into the cloud servers of the company as stated in the deployment plan.

9.4 Further projects

Currently, the project goal is to reduce empty driven kilometres without considering the total driven kilometres, therefore, the model currently will result in the minimum empty driven kilometres. However, it will increase the total driven kilometres or will choose a worse solution to reduce the empty driven kilometres. My advice is to create a variable that represent a ratio between the total performance and reducing the empty driven kilometre, and if the total driven kilometres exceed the ratio, then the company should increase the empty driven kilometres to reach a slightly balanced ratio.

This project is a start for several projects that can be integrated in Vos, to begin the MILP models are uniquely build for the specification of the company at Denmark, therefore, integrating the model to the whole system of Vos will result in complications. Hence, the problematic phase of the MILP is the constructive phase and currently Vos has the model and the logic behind the model, therefore, making changes to the model to adapt it to other counties will be easily performed.

If the company decided to implement and integrate the MILP model in their software, a new project can aim to increase the profitability of the company using dynamic programming. As this project is only concerned with minimizing the empty driven kilometres. Dynamic programming can be introduced and aim to minimize the total cost. So, the research question can be how to maximize the profit while also minimize the empty driven kilometres. To answer the research question, while having two decision variables with contradiction factors between them, a weight attractive variable should be introduced as a normalization factor. The dynamic programming will aim to predict the offer cost based on the historical data and after accomplishing the expected offer cost. Two attributes will be available to find the profit. The first attribute is the minimum empty driven kilometres and the second is the expected offer cost. Using the attractiveness weight-sum model, the project can result in the maximum attractiveness based on the two attributes and result in the optimal solution.

Bibliography

Azi, N., Gendreau, M., & Potvin, J.-Y. (2006). An exact algorithm for a single-vehicle routing problem with time windows and multiple routes. *ELSEVIER*, 755-766.

Çelik, M., Ergun, Ö., Johnson, B., Keskinocak, P., Lorca, Á., Pekgün, P., & Swann, J. (2012). Humanitarian Logistics. 2012 TutORials in Operations Research, 18–49. https://doi.org/10.1287/educ.1120.0100

CHIRA-CHAVALA, T. (1968). Problems of Combination Trucks on Wet Pavements: An Accident Analysis T. CHIRA-CHA VALA (pp. 70–75). Committee on Methodology for Evaluating Highway Improvements.

Heilmann, K. (2020). Information Frictions, Load Matching, and Route Efficiency in the Trucking Industry. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3545019

Theisens, H. C. (2021). *Lean Six Sigma Green Belt*. Van Haren Publishing.

Guangying et al. (2023). Selecting risk response strategies to minimize human errors in a design. *Expert Systems With Applications*.

What Defines Effective Delivery Route Apps? (2021, 09 15). From DispatchTrack : https://www.dispatchtrack.com/blog/delivery-route-apps

What Is Route Planning In Logistics: A Detailed Guide. (n.d.). From Shipsy: https://shipsy.io/blogs/what-is-route-planning-in-logistics-a-detailed-guide/#:~:text=Route%20planning%20requires%20considering%20multiple,and%20prone%20to%20human%20errors.

Kruse, et al. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Elsevier*, 526-534.

Leading the Way in Sustainability. (2021, 2 1). From Vostransportationgroup: https://vostransportgroup.com/leading-the-way-in-sustainability/

Voelpel, S., & etc. (2005). Escaping the Red Queen Effect in Competitive Strategy:: Sense-testing Business Models. *European Management Journal*, 37-49.

Islam, S. e. (2019). Minimization of empty container truck trips: insights into truck-sharing constraints. *Emerald Publishing Limited*, 641-662.

Road freight transport by journey characteristics. (2023, 8). From eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_journey_characteristics#Empty_runnings_of_road_freight_vehicles

Road freight transport by journey characteristics. (2023, August). From eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_journey_characteristics&stable=1#Empty_runnings_of_road_freight_vehicles

Rescot, R. e. (2009). Predicting and Mitigating Wind-Induced Truck Crashes on I-70 in Kansas. *ResearchGate*.

Manley, E. (2015). A heuristic model of bounded route choice in urban areas. *ELSEVIER*, 195-209.

Kabadurmus, O., & Erdogan, M. (2023). A green vehicle routing problem with multi-depot, multi-tour, hetergeneous fleet and split deliveries: a mathematical model and heuristic approach. *Jounral of Combinatorial Optimization*, 1-29.

Sitek, P., & et al. . (2021). Optimization of capacitated vehicle routing problem with alternative delivery, pick-up and time windows: A modified hybrid approach. *ELSEVIER*, 670-678.

Montoya-Torres, J., & al., e. (2015). A literature review on the vehicle routing problem with multiple depots. *ElSEVIER*, 115-129.

Montoya-Torres, J. (2015). A literature review on the vehicle routing problem with multiple depots. *ELSEVIER*, 115-129.

Ursani, Z. (2011). Localized genetic algorithm for vehicle routing problem with time windows. *ELSEVIER*, 5375-5390.

Wassan, N., & etal. (2016). The Multiple Trip Vehicle Routing Problem with Backhauls: Formulation and a Two-Level Variable Neighbourhood Search. *ELSEVIER*, 454-467.

Wegen , L., & Heijden , M. (2017). *Vehicle Routing* . Enschede.

Uddin, F., & al., e. (2023). An Improvement to the 2-Opt Heuristic Algorithm for Approximation of Optimal TSP Tour. *Applied Science*.

Pinedo, M. (2009). *Planning and Scheduling in Manufacturing and Services*. New York: Springer.

Sacramento, D., & al, e. (2019). An adaptive large neighborhood search metaheuristic for the vehicle routing problem with drones. *Elsevier* , 289-315.

Tanner, K. (2018). Chapter 14 - Experimental research. In K. Tanner, *Research Methods: Information, Systems, and Contexts (Second Edition)* (pp. 337-356). Chandos Publishing.

Reuters. (2022, October 28). From Sweden to further investigate Nord Stream pipeline damage: https://www.reuters.com/world/europe/sweden-further-investigate-nord-stream-pipeline-damage-2022-10-28/

Vermaas, P. (2011). *A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems.* Cham: Springer.

Kottow, M. (2008). The history of ethics in research with human subjects. *RECIIS*, 7-18.

Puylaert, R. (2023, Novemeber 25). *Research Integrity.* From Universiteiten Van Nederland: https://www.universiteitenvannederland.nl/en/research-integrity

(2018). *Netherlands Code of Conduct for Research Integrity.*

Poel, I. (2011). *Ethics, Technology, and Engineering : An Introduction.* Chicester : Wiley-Blackwell.

Li , Y., & et al. (2019). An improved ant colony optimization algorithm for the multi-depot green vehicle routing problem with multiple objectives. *elsevier*, 1161-1172.

Ni, Q., & Tang, Y. (2023). A Bibliometric Visualized Analysis and Classification of Vehicle Routing Problem Research. *Sustainability* , 1-37.

Montoya-Torres, J., & et al. (2015). A literature review on the vehicle routing problem with multiple depots . *Elsevier* , 115-129.

Ye, C., & et al. (2022). Electric vehicle routing models and solution algorithms in logistics distribution: A systematic review. *Environmental Science and Pollution Research*, 57067–57090.

Luo, H., & et al. (2021). An ACO-based heuristic approach for a route and speed optimization problem in home health care with synchronized visits and carbon emissions. *Soft Computing*, 14673–14696.

Sun, O., & et al. (2020). A Review on Optimization Methods for Biomass Supply Chain: Models and Algorithms, Sustainable Issues, and Challenges and Opportunities. *Process Integration and Optimization for Sustainability*, 203–226.

Hong, Y., & et al. . (2023). Designing sustainable logistics networks for classified municipal solid wastes collection and transferring with multi-compartment vehicles. *Elsevier*.

Ferreirinha, L., & et al. (2022). A VRP Model to Support Last Mile Maritime Containers. *Innovations in Industrial Engineering*, 361-370.

Timothy , C., & et al. (2017). Towards Collaborative Optimisation in a Shared-logistics Environment for Pickup and Delivery Operations. *ICORES*, 477-482.

Saunders, B., & et al. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *SpringerLink*, 1893–1907.

Martínez, R. (2023). Information literacy. 28.

Winston, W. (2004). *Operations Research APPLICATIONS AND ALGORITHMS.* Belmont: Thomson Learning.

Ye, C., He, W., & Chen, H. (2022). Electric vehicle routing models and solution algorithms in logistics distribution: A systematic review. *Environmental Science and Pollution Research*, 57067–57090.

Zhu, M. (2021). The relationship among motivation, self-monitoring, self-management, and learning strategies of MOOC learners. *Journal of Computing in Higher Education* , 321-342.

Appendices

Appendix A

Table 1 – Vos 2022 uncleaned data descriptive analysis (Finland)

FINLAND EMPTY DRIVEN KILOMETRES	
MEAN	68,1759184
STANDARD ERROR	11,5662837
MEDIAN	2,68
MODE	0
STANDARD DEVIATION	140,233737
KURTOSIS	13,9875562
SKEWNESS	3,4328287
RANGE	804,67
MINIMUM	0
MAXIMUM	804,67
SUM	10021,86
COUNT	147

Appendix B

```python
# possible arcs constriants that the arcs needs to follow:
def setup_routing_constraints(model,all_customers,d,x,y):
    # stage 2 - X
    model += lpSum(x[('linehaul', j, 2)] for j in all_customers if j.startswith('MP')) == 1
    # Stage 6 - X
    model += lpSum(x[(i, 'depot', 6)] for i in all_customers if i.startswith('BH')) == 1
    for (i, j, s) in x:
        if s == 2:
            model += d[i][j] * x[(i, j, 2)] <= 200, f"distance_constraint_stage_2_arc_{i}_{j}"
        if s == 6:
            model += d[i][j] * x[(i, j, 6)] <= 100, f"distance_constraint_stage_6_arc_{i}_{j}"
        if i.startswith("MP") and s == 3:
            # Stage 2 to Stage 3 link
            routes_stage2 = lpSum(x[('linehaul', i, 2)])
            model += lpSum(x[(i, j, 3)] for j in all_customers if j.startswith('MP') and i != j) == routes_stage2
    for i in all_customers:
        if i.startswith('MP'):
            # Linking visit decision with route decision
            model += lpSum(x[('linehaul', i, 2)]) == y[i]
            model += lpSum(x[(i.replace('MP', 'BH'), 'depot', 6)]) == y[i]

# the function below sets the constrint that each pickup load should be greater or equal to than the minimum pickup lo
# and that the picked up loads should not exceed the vehicle capacity of 13.6M
def setup_weight_constraints(model,all_customers,x,y,mp_weights,cumulative_weight,min_weight,vehicle_capacity):
    for j in all_customers:
        if j != 'depot' and j != 'linehaul' and j.startswith('MP'):  # Exclude depot
            model += lpSum(x[('linehaul', j, 2)] * mp_weights[j]) >= min_weight * lpSum(
                x[('linehaul', j, 2)]), f"min_weight_constraint_stage_2_customer{j}"  # if the path x is 1, then we ch
            model += lpSum(x[(j.replace('MP', 'BH'), 'depot', 6)] * mp_weights[j]) >= min_weight * lpSum(
                x[(j.replace('MP', 'BH'), 'depot', 6)]), f"min_weight_constraint_stage_6_customer{j}"
```

Figure 1 – Python model sample

Appendix C

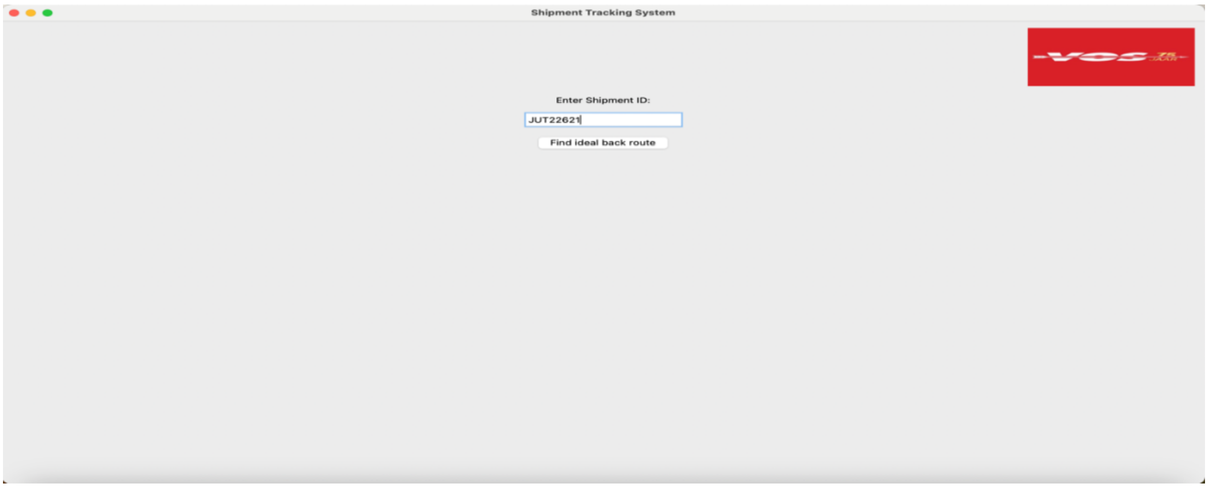

Figure 2 – Vos 2022 Visualization tool page 1

Figure 3 – Vos 2022 Visualization tool page 3