Vrije Universiteit Amsterdam

Master Thesis

Optimization of a Multi-period Two-stage Transportation Problem for a Subscription-based Company

Author: Y.T. Reurich (2672887)

1st supervisor: Alessandro Zocca
daily supervisor: Melle Minderhoud (Swapfiets)
2nd reader: Gabriele Benedetti

September 28, 2022

Acknowledgements

This thesis is submitted as part of the completion of the Master's degree program in Business Analytics at VU Amsterdam. The research described herein was commissioned by the Data Science department at Swapfiets between March 2022 and September 2022.

In this study, a mixed integer programming approach is proposed to solve the multi-period two-stage transportation problem for Swapfiets. The model focuses on efficiently distributing bikes for a time window of multiple months over the most profitable customers.

I would like to use this opportunity to thank the Data team at Swapfiets for the opportunity to write my thesis in their company and for all the support I have received during the past seven months. In particular, I would like to thank Melle Minderhoud, my supervisor at Swapfiets. Thank you for the dedicated support, the critical feedback and for the challenges to maximize the result of my internship.

Additionally, I would like to thank my supervisor from the Faculty of Science at Vrije Universiteit Amsterdam, Alessandro Zocca. Thank you for the advice, support and feedback throughout the Master Project Business Analytics. I greatly appreciate the bi-weekly progress meetings.

I would also like to thank Gabriele Benedetti from VU Amstersdam for assessing my work.

Yorran Reurich
Amsterdam, September 11, 2022

Management Summary

Since the beginning of Swapfiets, Swapfiets has experienced extreme growth, and this growth is expected to stay. However, if Swapfiets wants to outgrow the scale-up phase to become a profitable organization, there are multiple processes that need to be changed and optimised. One of these processes is the supply of bikes and e-bikes to business units. Swapfiets currently works with a reactive pull system when it comes to supplying stores with bikes and e-bikes. In the current approach a yearly demand forecast is created which serves as the base of the supply plan. Based on this plan bikes are allocated to stores, and whenever stores run out of bikes and e-bikes, store managers inform the head quarter after which they will be supplied with new stock.

A mixed integer programming model has been proposed with the aim to define a more proactive logistical strategy that can be used for multiple periods for all bike types. Another problem that can be tackled by the model is the fact that the total demand of new customers will eventually exceed the supply. This has been tackled by incorporating a component that represents the average profitability of different members based on locations and starting month, namely the lifetime value. Based on this feature it is possible to target the most profitable customers. Furthermore, regarding the transportation, bikes and e-bikes are not transported per piece, but are transported in trucks with a fixed capacity, with either [redacted] e-bikes, [redacted] bikes, or a mixture of bikes and e-bikes. This allows the model to efficiently fill trucks as a way of reducing transportation costs. Additionally, bikes and e-bikes will be returned by customers due to churn that happens during the period that is included in the logistical strategy. These bikes and e-bikes are added to the supply in the model.

All of these components have been included in the mixed integer programming model. As a next step, the model has been validated by applying extreme condition testing, based on a subset of four business units for three periods,

instead of five. The model worked as intended and running the full model, with a runtime of 29 seconds, showed that by following the suggested supply strategy, Swapfiets can generate a total profit of € [redacted] until the end of 2022. In addition, employees from the Operations department believe it is vital to fill trucks efficiently to become profitable. However, the model shows that the transportation costs are so little compared to the leasing costs and the revenue that this should not be the main focus for Swapfiets. As a final remark, it is interesting to note that the model shows that Swapfiets has enough bikes and e-bikes to fulfill sales demand until the end of the year, which was not expected, and thus some stores were not allowing new customers. Based on the results of the model these stores can be opened again to generate more revenue.

Contents

Chapter 1

Introduction

1.1 Background

Three young students from the University in Delft, Richard Burger, Dirk de Bruijn and Martijn Obers, were having a coffee at the TU Delft and were talking about their love for road cycling and their entrepreneurial mindset. As they were discussing, they saw a female student passing by on a bike that was slowly falling apart as she progressed. The girl on the bike was very frustrated by the fact that her bike was not working properly. This was the moment that the idea for Swapfiets was born. The three students realised that the girl was frustrated because she now had to bring her bike to a repairman, and that it would take some days before the bike got repaired, which meant that she would not be mobile for a few days. Besides that, repairing a bike can be expensive. Or she could repair the bike herself, but that can be a hassle as well. This story, in combination with the fact that riding a bike is part of the daily life for most people in The Netherlands, resulted in a great plan.

To put this plan to practice, on 7 November 2014, the three students founded the company Swapfiets N.O.V. At that time they bought 40 bikes from Marktplaats, which is the dutch version of eBay/Craigslist, and tried to repair them. When repairing the bikes they realised that they needed something to make these bikes recognizable. The initial idea was to use a blue front and back tire, blue as a reference to Delft, the color "Delfts blauw". However, throwing away so many good working tires would be a waste. Eventually via Whatsapp they decided that only changing the front tire to a blue one would be better to reduce waste and since it is the easiest tire to change.

They started with a small pilot group and the concept showed to be successful. After they increased their pilot group to 150 customers the growth exceeded the demand, and even after this fact was clear to the aspiring customers, the aspiring customers kept on pre-ordering bikes. The number of pre-orders gave a good indication for the possibilities of growth for Swapfiets. In August 2016 Ponooc came on board after which things really started to take off. Swapfiets has grown rapidly over the last years, and can be considered to be a scale-up with around [redacted] active members in over 50 cities in 9 different countries in Europe at the beginning of 2022. Figure 1.1 shows how the number of active members have changed since the beginning of 2018.

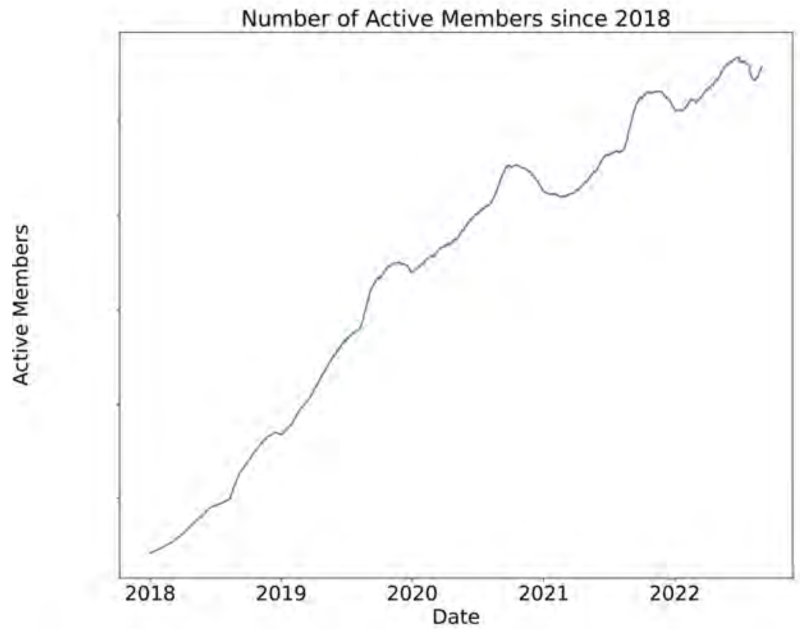

Figure 1.1: Figure showing the number of active Swapfiets members since the beginning of 2018

The figure shows that Swapfiets grew from around [redacted] thousand members at the beginning of 2018 to over [redacted] thousand four years later in 2022. They expect to keep this growing trend going, and strives to a huge milestone of 1 million active members.

1.2 Problem Statement

When a company grows at this pace, processes become more complex at a high rate. Instead of being able to execute tasks (manually) on a small scale, processes need to be well defined and efficient to achieve profitability. The logistics is one of the fields that becomes more complex when a company grows and expands to different locations, transporting products from A to B, in this case three types of bikes and two types e-bikes.

The three normal bikes are the Original 1, Original 1+ and the more luxurious version, the Deluxe 7. These bikes are offered for €18,90 per month for the Original 1 and 1+ or €21,90 for the Deluxe 7 per month. Similarly, for the e-bikes, the basic model is the Power 1 and the more luxurious version is called the Power 7, for €59,90 and €79,90 per month, respectively.

In the current situation the distribution of bikes and e-bikes over all the different stores is based on a planning that consists of two components. In the first place the Data Science team creates a forecast to predict the demand per store. However, this forecast merely functions as a basis and gives an indication of the expected needs in the long run. The second component of the current stock allocation methodology relies on the store employees. In each store the employees keep track of their stock and inform the head office whenever more stock is needed. Based on this, new bikes and e-bikes are transported to the store. Additionally, a tier ranking method is applied to prioritise stores. Based on the profitability of the customers that go to the a particular store, the store gets assigned a different tier, where 1 resembles the best stores. In case of a shortage, first the stores with the highest tier will be provided with new bikes and e-bikes. This process is very reactive and mistakes can easily be made, as the process is prone to human errors.

This strategy was focused on growth, and it basically meant accepting everybody who was interested in getting a bike or e-bike. In the current phase where demand exceeds supply, this is an undesirable strategy. The intention is to transition to a situation where the distribution of stock over all the stores can be managed proactively and where stock is distributed in such a way that the most profitable customers can be prioritized. This means, switching from a pull system, where store employees ask for new stock, to a push system where enough supply is pushed towards the stores.

1.3 Theoretical Background

The problem of efficiently redistributing physical products from a set of locations (often warehouses or factories) to a number of stores which all have a particular number of demand has been around for a long time, and is called the “transportation problem”. It is initially known as the Hitchcock-Koopman problem. The mathematical formulation was first formulated by Frank L. Hitchcock in 1941, in his paper “The Distribution of a Product from Several Sources to Numerous Localities” [1]. Unfortunately his paper did not get much attention at the time. In 1947 Koopman began his research on the potentialities of linear programs for the study of problems in economics, which he reported on in his paper

"Optimum Utilization of the Transportation System". The combination of his research in [2], with the basics being already documented by Hitchcock eventually resulted in the Hitchcock-Koopmans Transportation problem .

The formulation of the problem remained the same and is still being used frequently nowadays to tackle transportation problems. In [3], Armenta, Maldonado-Macias et al. addressed transportation for an independent transporter of fresh fruit and vegetables in Mexico. They used an alternative objective function to consider certain constraints of the transporter in Mexico. The main objective of this research was to determine the optimal configuration of products that must be purchased and how to transport them to the different destinations to achieve maximum profitability. They dealt with the problem by applying the simplex method that maximized the objective function which was a combination of the market price with a deduction of the purchase price and travel expenses. Furthermore, they used a set of constraints that makes sure that the demand is met at all the destinations and a set of constraints that makes sure that the purchased supply does not exceed the number of products that is actually available. Up until this point the transporter experienced monetary losses due to the fact that the transporter did not rely on a plan to determine the number of products that need to be bought and transported. The study showed that a model that includes the demand, the possible supply and the transportation costs can be used to set up a profitable purchase and transportation plan. As shown in [4], Shiang-Tai Liu conducted a study where he concluded that most classical transportation problems work with supply and demand quantities that are precisely specified to single numbers. He explains that sometimes it can be hard to get these precise numbers due to changing economic conditions, and thus stresses the importance to create a model that can compute an upper and lower bound for the transportation costs. In 2011 in [5], Raj and Rajendran focussed on solving a two-stage transportation problem. In a two-stage transportation problem the supply chain consists of plants, that produce products, which are transported to distribution centers after which the products are distributed over customers. They considered two scenarios. In the first scenario transportation costs per unit where used and every route had fixed costs, coupled with unlimited capacity at every distribution center. The second scenario is focused on opening new distribution centers. They tried to represent the first scenario as a single-stage fixed-charge transportation model to solve it with a genetic algorithm. They showed that it is possible to solve the first scenario with a genetic algorithm. However, only small problem instances were tested, with a maximum size of four plants, three distribution centers and five demand locations. In 2021, Zhu, Ji and Shen tackled the transportation problem that included uncertainty parameters in [6]. Due to the incompletion of historical data, parameters such as purchasing prices, supply levels, and demand were missing. The goal was not to minimize transportation costs, like in the classical transportation model, but a maximization of profit. Similarly to Raj and Rajendran, they solved the transportation problem by using a genetic algorithm. Additionally they used Particle Swarm Optimization (PSO) to solve the problem, and found out that PSO outperforms the GA. In [7], in 2017, Mogale, Dolgui et al. researched the possibility of tackling a multi-period inventory transportation model for a food supply chain in India to satisfy the demand for grain for several deficit Indian states. The problem involved the transportation of bulk food by vehicles with limited capacities from surplus states to deficit states. They showed that the multi-period problem can be formulated as a mixed integer non-linear programming (MINLP) model that seeks to minimize the overall costs including transportation costs, storage costs and operational costs.

The literature research showed that formulating the transportation problem as a mixed integer programming model offers a good solution. Even when the model becomes a two-stage problem or when it contains multiple periods, the mixed integer programming model can still be used. During this research an algorithm has been created that is able to redistribute bikes and e-bikes over all the destinations, such that the demand is met as good as possible in an efficient manner. This can be translated into the following research question:

"*How can optimization algorithms be used to redistribute products of subscription-based companies over a set of demand locations such that supply and demand can be matched to increase profitability?*"

1.4 Thesis Outline

The remainder of this thesis is structured as follows: in Chapter 2 the description of the proposed method to solve the transportation problem is given. In chapter 3 the data that is used as input for the transportation model is described. Furthermore the data has been analysed, and the results can also be found in chapter 3. Chapter 4 elaborates on how the validity of the model is tested. In chapter 5 the results of the transportation model are shown. Not only the results of the transportation model in terms of profit, but also a stress test to test the computational performance, is included. Chapter 6 discusses the key results of the model, limitations of the model are stated and recommendations are done. In the final chapter, Chapter 7, the research is concluded and the research question is answered.

Chapter 2

Development of The Transportation Model

The previous chapter elaborated on the background of Swapfiets and introduced the problem that they are currently facing. In this chapter the optimization model, a mixed integer programming model, that has been created to tackle this transportation problem is described. The goal of the model is to determine the number of bikes and e-bikes that have to be transported to particular locations, in such a way that the profit is maximized. The development from the basic transportation model to a complex multi-period two-stage model is described in this chapter. The purpose of showing the different phases is to show the process of adding complexity to the model step by step and thus making it easier to follow.

2.1 Single-stage Transportation Problem

The first model that has been created is a variant of the classical transportation model. As stated by Dantzig in [8], in 1997, for the classical transportation model the goal is to determine an optimal schedule of shipments that originate from a source where known quantities of a commodity are available. These commodities are allocated and sent directly to their final destinations, where the total amount received must equal the known quantities required. Instead of using stores as sources and destinations, business units are used. Business units can be defined as the cities where bikes and/or e-bikes are being offered or are required. Business units can have multiple stores. The model that has been created is a mixed integer programming model. A mixed integer programming model, or MIP model for short, can be defined as a linear programming model where at least one of the variables is required to be an integer, as stated in [9]. The model works for one bike type only, to keep it readable. It contains m locations that function as supply locations. These locations function as supply locations if in some way they will provide bikes or e-bikes for this single period model. This means at least one of the following events is happening: either they currently contain bikes, repairs will be executed at the particular location, newly ordered bikes or e-bikes will be arriving at these locations or bikes or e-bikes will be returned due to churn. These locations are collected in a set M. Furthermore, the model contains a set N, with all the n locations that require a particular number of bikes or e-bikes, which similarly to set M, differs in size per bike type. The MIP model is formulated as follows:

$$\max \quad \sum_{i \in M} \sum_{j \in N} u_{ij} x_{ij} \tag{2.1}$$

$$\text{s.t.} \quad \sum_{j \in N} x_{ij} = s_i \qquad \forall i \in M \tag{2.2}$$

$$\sum_{i \in M} x_{ij} = d_j \qquad \forall j \in N \tag{2.3}$$

$$x_{ij} \geq 0 \qquad \forall i \in M, \quad \forall j \in N \tag{2.4}$$

The model uses the same structure as the classical transportation model as stated in [8]. However the difference is that this objective function is being maximized, while the classical model focuses on minimizing the transportation costs. The objective function (2.1) is the function that quantifies the profit that can be achieved by following the strategy that is proposed by the model. In the objective function, u_{ij} is the total utility that is generated by transporting one bike that originates from location i to location j, where it will be sold. The utility in this case is defined as the net profit, which is the difference between the customer lifetime value for a customer from a particular business unit j, including the deduction of the transportation costs. These are the costs of transporting one product from supply location i to demand location j. To clarify, imagine that a bike that is offered to a customer in Utrecht, which will result in a customer lifetime value of l euros. However, the bike was initially stationed in Rotterdam, it means that some transportation expenses of t euro's will be made first. This will result in a utility of $l - t$ euro's. Furthermore, in the objective function, x_{ij} is the decision variable, which indicates the number of bikes that have to be transported from i to j.

The second equation,(2.2), is the supply constraint. This constraint ensures that the number of products that are transported from a particular supply location i cannot exceed the amount of products that are present at that location. The next equation, (2.3), is focused on the other side of the network, namely the demand. This constraint makes sure that the demand in each demand location is met.

2.2 Multi-period Transportation Problem

On top of the basic model as described in the previous section, one may want to include other bike types as well, as Swapfiets does not only offer one type of bike, but five. Two e-bikes, the *Power 7* and *Power 1*, and three bikes, the *Original 1*, *Original 1+* and the *Deluxe 7*. The demand and the supply of these products differ from each other, and thus should be considered separately. This can be achieved by adding another index to the decision variable and constraints. These five bike types are included in the set O, *bike type* $k \in O$. Additionally, Swapfiets intends to look further into the future instead of only looking at a single period. The following set has been introduced to include multiple periods, set *time* $t \in T$. Instead of using x_{ij} to indicate the number of bikes that are transported from i to j, x^t_{ijk} is used. x^t_{ijk} indicates the number of bikes of type k that are transported from location i to location j at time t. This results in the following model:

$$
\begin{aligned}
\max \quad & \sum_{t \in T} \sum_{i \in M} \sum_{j \in N} \sum_{k \in O} u^t_{ijk} x^t_{ijk} && && (2.5)\\
\text{s.t.} \quad & \sum_{j \in N} x^t_{ijk} = s^t_{ik} && \forall i \in M, \forall k \in O, \forall t \in T && (2.6)\\
& \sum_{i \in M} x^t_{ijk} = d^t_{jk} && \forall j \in N, \forall k \in O, \forall t \in T && (2.7)\\
& x^t_{ijk} \geq 0 && \forall i \in M, \forall j \in N, \forall k \in O, \forall t \in T && (2.8)
\end{aligned}
$$

The structure of the MIP model is similar to the previous version in Section 2.1, with the only addition being the two extra indices, and thus does not require further explanation. As a final remark, it should be noted that the different bike types k and the times t are not linked and therefore this model is the same as solving independent instances of the previous model for each value of t and k. To clarify what the network looks like, the flow for one period for one bike type has been visualized in Figure 2.1.

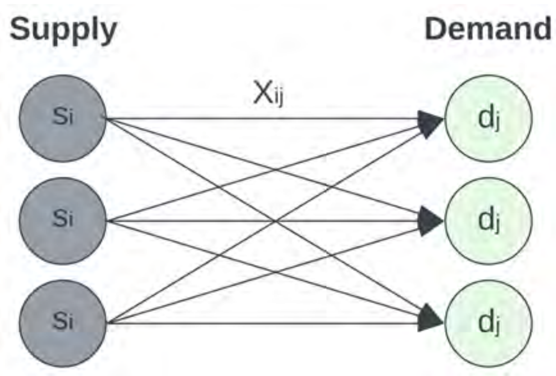

Figure 2.1: Single stage transportation model

2.3 Multi-period Two-stage Transportation Model

In this section the final model will be described. Compared to the previous model, a lot of new components have been added, making the model much more complex than the previous stage. However, this was crucial to make the model as realistic as possible, by incorporating a lot of processes of Swapfiets. In the first place, instead of working with a multi-period single-stage transportation model, as shown in Figure 2.1, now the MIP model has been changed to a multi-period two-stage transportation model. To clarify this principle, see Figure 2.2 in which the flow has been shown for one bike type.

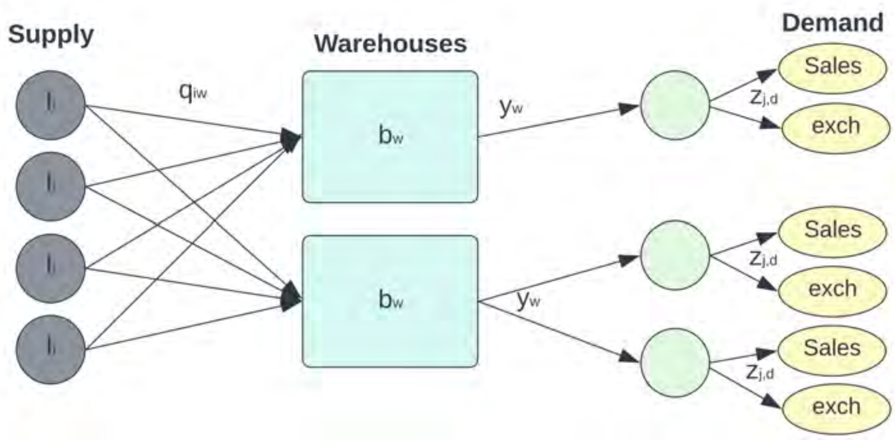

Figure 2.2: Two stage transportation model

The figure shows that first bikes and e-bikes are transported from the business units to a number of warehouses, after which they are distributed over the business units to satisfy demand. Another change is that bikes are now transported in trucks in the first stage, as this is more realistic than transporting them per piece, like in the previous model. This means that batches are transported at a fixed cost instead of a fixed cost per bike. Another change compared to the previous model is the fact that in the second stage it is not possible to transport bikes and e-bikes from a particular warehouse to all business units. Warehouses are only able supply pre-defined business units with bikes and e-bikes. This drastically reduces the number of possibilities in the second stage for the model. After the second stage of the transportation model (indicated by the variable y_w, which resembles the total number of bikes that are transported from warehouse w to any connected business unit), bikes and e-bikes are distributed within the business units to fulfill new demand "sales", and secondly exchanges. Exchanges happen whenever a customer's bike needs to be replaced, which could happen when a bike is broken and needs to be repaired. The repairs can be divided in two different levels, level 1 repairs and level 2 repairs, with level 2 repairs meaning problems with the bike or e-bike that cannot easily be solved. In case of these level 2 repairs, the members' bike is exchanged with a new one. Additionally, whenever a bike goes missing because it gets stolen, the bike will also be exchanged for a new one. This service is part of the subscription, and thus applies to every member. Exchange demand should always be satisfied before giving bikes to new customers, as serving the existing customer base is preferred. This means that even fewer bikes remain to supply the sales demand, and thus stresses the importance of prioritizing the most profitable customers.

Additionally, to make the model more realistic, churn will be included. In the previous models churn based on bikes and e-bikes that were sold before the starting moment of the model was already and were part of the supply figures. But based on the MIP model, bikes will be transported and sold, which creates new churn eventually. This means that every month a percentage of the bikes and e-bikes that have been sold in the preceding months will be returned and can be used as supply again.

Another addition the the model is that leasing costs are included. Whenever a bike or e-bike is not used, and remains in the warehouse or in a store, monthly leasing costs have to be paid. [redacted].

One of the problems of the previous model was the fact that the times t and bike types k were not connected. This is as if the model would be solved eight times separately instead of one time. To deal with this problem an inventory system has been introduced. This enables the model to transport stock in a particular month, then hold it in stock, to use it in a later month. This is especially useful as bikes and e-bikes are now transported in batches, which means that the required number of bikes for several months on one location can be transported in the same truck and kept in stock. This allows the model to efficiently fill trucks, to reduce the transportation costs.

Based on these changes, multiple new sets have been introduced. Swapfiets has two types of demand they have to deal with. The first type being the "exchange" demand and secondly, demand of new customers, sales. These two types of demand have been captured in set D, $d \in D$: {sales, exchange}. The second set that has been introduced is W, this set contains all the warehouses $w \in W$. The third and fourth set are $a \in A$, which contain all the e-bikes {Power 1, Power 7} and $b \in B$ which contain all the normal bikes {Original 1, Original 1+, Deluxe 7}. The sets with the different bike types make the split between bikes and e-bikes as they are of different sizes, which is relevant when trucks have to be filled. The last sets that have to be included in the transportation model, are the sets that state which business units are connected to which warehouse for the second stage of the model. These sets can be defined as following $j \in N_w$, where w presents the warehouse. This results into the following model:

$$\max \quad \sum_{d\in D}\sum_{k\in O}\sum_{j\in N}\sum_{t\in T} u^t_{jkd} z^t_{jkd} - \sum_{t\in T}\sum_{j\in N}\sum_{i\in M} c_{iw} q^t_{iw} - \sum_{t\in T}\sum_{k\in O}\sum_{i\in M} l_k I^t_{ik} - \sum_{t\in T}\sum_{k\in O}\sum_{w\in W} l_k b^t_{wk} \tag{2.9}$$

$$\text{s.t.} \quad \begin{cases} s^0_{ik} = I^0_{ik} & \forall i\in M, \forall k \in O, \quad \text{if } t=0 \\ I^{t-1}_{ik} + s^t_{ik} - \sum_{w\in W} x^{t-1}_{iwk} + r^t_{ik} = I^t_{ik} & \forall i\in M, \forall k\in O, \forall t\in T \end{cases} \tag{2.10}$$

$$\begin{cases} \sum_{i\in M} x^0_{iwk} = b^0_{wk} & \forall w\in W, \forall k\in O, \quad \text{if } t=0 \\ b^{t-1}_{wk} + \sum_{i\in M} x^t_{iwk} - y^{t-1}_{wk} = b^t_{wk} & \forall w\in W, \forall k\in O, \forall t\in T \end{cases} \tag{2.11}$$

$$\theta q^t_{iw} \geq \sum_{a\in A} x^t_{iwa} + \frac{90}{130}\sum_{b\in B} x^t_{iwb} \qquad \forall i\in M, \forall w\in W, \forall t\in T \tag{2.12}$$

$$\begin{cases} r^0_{jk} = 0 & \forall j\in N, \forall k\in O, \quad \text{if } t=0 \\ r^t_{jk} \leq \sum_{b=0}^{t-1} \alpha^b_{j,t-b,k} z^b_{jk,sales} & \forall j\in N, \forall k\in O, \forall t\in T \end{cases} \tag{2.13}$$

$$\begin{cases} r^0_{jk} = 0 & \forall j\in N, \forall k\in O, \quad \text{if } t=0 \\ r^t_{jk} \geq (\sum_{b=0}^{t-1} \alpha^b_{j,t-b,k} z^b_{jk,sales}) - 1 & \forall j\in N, \forall k\in O, \forall t\in T \end{cases} \tag{2.14}$$

$$y^t_{wk} \geq \sum_{j\in N_w} v^t_{jk,exchange} \qquad \forall w\in W, \forall k\in O, \forall t\in T \tag{2.15}$$

$$y^t_{wk} \leq \sum_{d\in D}\sum_{j\in N_w} v^t_{jkd} \qquad \forall w\in W, \forall k\in O, \forall t\in T \tag{2.16}$$

$$y^t_{wk} \leq b^t_{wk} \qquad \forall w\in W, \forall k\in O, \forall t\in T \tag{2.17}$$

$$\sum_{d\in D}\sum_{j\in N_w} z^t_{jkd} \leq y^t_{wk} \qquad \forall w\in W, \forall k\in O, \forall t\in T \tag{2.18}$$

$$\begin{cases} z^t_{jkd} = v^t_{jkd} & \forall j\in N, \forall k\in O, \forall t\in T, \quad \text{if } d = exchange \\ z^t_{jkd} \leq v^t_{jkd} & \forall j\in N, \forall k\in O, \forall t\in T, \quad \text{if } d = sales \end{cases} \tag{2.19}$$

$$\sum_{w\in W} x^t_{iwk} \leq I^t_{ik} \qquad \forall i\in M, \forall k\in O, \forall t\in T \tag{2.20}$$

$$I^t_{ik} \geq 0, x^t_{iwk} \geq 0, y^t_{wk} \geq 0, z^t_{jkd} \geq 0, r^t_{jk} \geq 0, b^t_{wk} \geq 0, q^t_{iw} \geq 0 \tag{2.21}$$

2.3.1 Variables

When it comes to the variables, a lot of changes have been made compared to the previous model. To be able to understand the multi-period two-stage model, these variables need to be explained. In this subsection they will be introduced one by one.

- z^t_{jdk}: has been introduced, which indicates the number of bikes or e-bikes of type k that will be transported to location j to satisfy demand of type d (exchange or sales) at time t.
- q^t_{iw}: represents the number of trucks that will transport bikes and/or e-bikes from supply location i to warehouse w at time t. In contrast to the transport in the second stage, in the first stage all the business units are connected to all the warehouses. This means that bikes and e-bikes from every city can be transported to each warehouse.
- I^t_{ik}: is the variable that represents the number of bikes or e-bikes of type k at business unit i that is in stock at time t before transporting bikes or e-bikes to the warehouses in the same period. This variable allows the model to see how many bikes or e-bikes of type k can be transported from location i at time t to warehouse w.
- b^t_{wk}: represents the number of bikes or e-bikes of type k that is in stock in warehouse w at time t. This variable represents the end stage of the first step in the transportation model and functions as the starting point of the second stage of the transportation model.
- x^t_{iwk}: is the variable that represents the number of bikes or e-bikes of type k that are transported from location i to warehouse w at time t. This variable depends on the stock available at location i, of type k at time t (I^t_{ik}), as it cannot exceed this number.
- y^t_{wk}: is the variable that indicates the number of bikes or e-bikes of type k that are transported from warehouse w to any location j that the warehouse is connected to at time t. This variable is mainly used to determine the remaining stock levels at a particular warehouse for the next moment in time.
- r^t_{jk}: represents the variable with the bikes or e-bikes of type k that return at location j at time t due to churn. This number is caused by the decisions the MIP model makes, as based on the volume of bikes that are used to satisfy sales demand, churn is generated.

2.3.2 Parameters

Besides a lot of variables, the MIP model also contains a lot of static values, that originate from the data which are captured in the parameters. These parameters are described in this subsection.

- $u^t_{jkd} \geq 0$: is the average lifetime value of customers at demand location j for bike type k of demand type d (sales or exchange). When $d = exchange$, the parameter value is 0, since exchanging a bike does not result in new revenue. This parameter has been included since the expected demand exceeds the supply, which means that the problem is an unbalanced transportation problem and not all the demand can be satisfied [10]. Using this measure, the most profitable customers can be prioritized.

- $c_{iw} \geq 0$: is the parameter that is used for the travel expenses. c_{iw} are the costs of sending one truck from location i to warehouse w.

- $l_k \geq 0$: are the leasing costs per bike or e-bike of type k for one time period.

- $s^t_{ik} \geq 0$: is the total supply of bikes or e-bikes of type k at location i at time t. Which consists of multiple components, this will be elaborated on in Section 3.1.

- $\theta \geq 0$: is the parameter that represents the capacity of a truck in number of e-bikes.

- $\alpha^t_{j,u,k} \geq 0$: is the churn rate that has been used to determine the number of bikes and e-bikes that will be returned due to churn, which can be used as supply again. $\alpha^t_{j,u,k}$ is the churn rate at location j, for a bike given out at time t, u time periods back in the past of bike type k .

- $v^t_{jdk} \geq 0$: is the parameter that represents the demand at location j of type d (sales or exchange) of bike type k at time t.

2.3.3 Objective & Constraints

The sets, variables and parameters that have been explained in the previous sections have been used to create the objective function and the constraints. The objective function and the constraints have been explained one by one in this subsection.

- Equation (2.9) is the objective function. This is the function that represents the profit that can be made by following the suggested transportation actions that are advised by the MIP model. The goal of the MIP model is to find the combination of decision variables that maximizes this function. To make it more clear what exactly is computed, the function can easily be split into different parts. In the first part ($u^t_{jkd}z^t_{jkd}$), the sum of the lifetime value of customers is determined for every bike that is given away at any location for any moment in time. This part makes up for the revenue part of the objective function, while the remaining parts are the costs associated with following the strategy as suggested by the MIP model. The second part ($c_{iw}q^t_{iw}$) are the transportation costs generated by driving trucks around between the supply locations and the warehouses. The third part ($l_kI^t_{ik}$) accounts for the leasing costs that are paid to the bank by Swapfiets per bike per time unit that are in stock in the business units. Similarly, the final part ($l_kW^t_{wk}$) accounts for the leasing costs of bikes that are in stock in the warehouses. Bikes and e-bikes that are in stock are not being used, meaning that they are collecting dust. It might seem strange that the leasing costs are only paid for the bikes and e-bikes that are not being used, but the idea is that the leasing costs are included in the lifetime values as well. This means that for the bikes that are given away, the leasing costs are also included, but they are included in the parameter u^t_{jkd}.

- Equation (2.10) is the constraint that computes the stock levels for all business units at any moment in time for every bike type before the bikes or e-bikes are transported to the warehouses. The constraint has been split into two variants. Whenever $t = 0$, the first variant will be used. In this variant the only supply that the model can use is s^0_{ik}, which is the external supply. Since it is the first period, there is no churn that is generated by the model, there is no outflow in the previous month and there is no previous stock value. The second version does include all these components. Whenever $t \geq 1$ the second version is used. In this version, the first part (I^{t-1}_{ik}, the stock in the previous period) is the base of the stock at time t for location i for bike type k. The stock of the previous month is changed by first adding the external incoming stock of the current month (s^t_{ik}), subtracting the stock that has been moved away to warehouses in the previous month (x^{t-1}_{iwk}) and adding the bikes that return in the current month because of churn caused by the model (r^t_{ik}).

- Equation (2.11) is the constraint that is used to determine the stock levels for all the warehouses. The constraint is also split into two versions. The first one is only used in case $t = 0$, because at $t = 0$ there is no previous stock and no satisfied demand in the previous period to consider. For the second variant, the stock in warehouse w depends on the stock level at $t - 1$, (b^{t-1}_{wk}) and the changes in the current period t. These changes are in the first place an addition of stock that is transported from the business units to the warehouses ($\sum_{i \in M} x^t_{iwk}$) in the current period t and a deduction of the satisfied demand in the previous period y^{t-1}_{wk}.

- Equation (2.12) is the constraint that is used to efficiently fill the trucks with bikes and e-bikes. Instead of transporting bikes per piece, bikes and e-bikes are in fact transported in batches using trucks. The truck capacity is captured in the parameter θ, since e-bikes are larger than bikes, smaller batches of e-bikes can be transported, meaning that a truck has a fixed capacity of either [redacted] e-bikes or [redacted] normal bikes. In this case the truck capacity for the e-bikes has been used and is captured in the parameter θ. However, to fill the trucks as good as possible, to reduce the costs of transportation, mixed trucks will be used, which means that both bikes and e-bikes should be able to fit into the same truck. This has been accomplished by making the weight of a bike [redacted] instead of 1 as it is for the e-bikes.

- Equation (2.13) is a constraint that is used to compute the number of bikes that are returned due to churn, that are a result of following the strategy as suggested by the MIP model. Similar to Equation (2.10) there are two variants. The first variant is only used if time $t = 0$, in this case the number of returned bikes caused by the model is 0 as it is the first period of the model. In any other cases, the second version is used. At the right hand side of the inequality the churn rate of bikes that are given away at time t, u periods ago of bike type k for location j ,$\alpha^t_{j,u,k}$, is multiplied with the number of bikes of type k that are given away at time t at location j to fulfill sales demand (z^t_{jdk}, where $d = sales$). This is repeated for every moment before the current time t, and the sum of these values is taken as the number of bikes that return because of churn. As a final remark, it seems strange that the constraint is an inequality with the $\leq$ sign, while a "=" sign might seem more logical. The reason behind the $\leq$ sign is that the churn rates are rates between 0 and 1, which means that multiplying this with an integer (the number of bikes), most likely results in a decimal number. It does not make any sense to use decimal numbers for bikes, so now the model takes the largest integer below the actual result.

- Equation (2.14) is related to the previous constraint in a sense that it is used to create the correct churn figures. Whenever the supply exceeds the demand, the model sets the churning bikes to 0. This is a logical reaction due to the maximization property of the objective function. In the objective function leasing costs are deducted from the objective value whenever bikes are not used. Due to the fact that the churn is included as a decision variable in the model, the model will put the variable to 0 since this lowers the total leasing costs, and thus results in a higher objective value. This is an undesired result, as in fact churn will always happen and thus should always be included. By adding this constraint to the model, churn is always included, even in scenarios where the extra bikes are not needed.

- Equation (2.15) ensures that from any warehouse at least enough bikes or e-bikes are transported to the stores in the second stage to meet the sum of the required number of exchange bikes for all the business units that it is connected to. In this case the $\geq$ is used because the number of bikes transported to the business units should at least be enough to meet the exchange demand, but it can be higher to also satisfy some or all of the sales demand for these business units. The constraint basically sets a lower bound for the number of bikes that should be transported from a particular warehouse to the business units.

- Equation (2.16) is comparable to (2.15), however (2.15) was seen as a lower bound which should always be satisfied, (2.16) sets the upper bound which should never be exceeded. This constraint makes sure that the number of bikes of type k that are transported to all the business units that are connected to the particular warehouse can never exceed the sum of the demand of exchanges and sales.

- Equation (2.17) ensures that the number of bikes that can be transported from any warehouse to any of the connected business units does not exceed the number of bikes of type k that are present in the warehouse.

- Equation (2.18) is used to make sure that the sum of all the bikes that are given away to satisfy sales or exchange demand to all business units that are served by the same warehouse does not exceed the total amount of bikes that are transported from that particular warehouse in the second stage. To clarify, the total number of bikes that are transported from a particular warehouse to any business unit that is connected to the warehouse is captured in the variable y^t_{wk}. This number is vital to the model to get a complete overview of the total number of bikes leaving the warehouse. This information is used to compute the warehouse stock level for time $t+1$ compared to t in (2.11).

- Equation (2.19) is split into two versions. The first version makes sure that for each business unit enough bikes and e-bikes are transported from the warehouse to the business unit to satisfy the exchange demand. The second version makes sure that the number of bikes transported to a particular business unit j does not exceed the actual forecasted demand of sales. This prevents that too many bikes are transported to business units.

- Equation (2.20) prevents the MIP model to transport more bikes and e-bikes from a particular location i to any warehouse w than they currently have in stock (I^t_{ik}).

The full model has been used to generate a transportation strategy that can be used by Swapfiets for each bike type for the coming period. In this plan a lot of constraints are used to resemble the processes of Swapfiets as accurate as possible. The strategy ensures that exchange demand is always met, and that with the remaining products it focuses on the most profitable customer while keeping the costs of transportation and the leasing cost as low as possible.

Chapter 3

Model Input & Data Analysis

The mixed integer programming model that has been used to tackle the transportation problem requires a lot of data as input to fill all the parameters. In this chapter, the transportation process has been split into different components, from the beginning of the supply chain, the supply, to the final stage, the demand. The data that is required for these stages and all the in-between steps are described and analysed in this chapter.

3.1 Supply

The first component that is required, are the supply figures, the number of available bikes and e-bikes. The total supply of bikes and e-bikes is a combination of several different flows. In the first place, the bikes that Swapfiets currently has available are considered as supply. These bikes and e-bikes are currently stored in warehouses and the different business units. Secondly, there is also a stream of external inbound bikes and e-bikes. Those are bikes that are currently not yet available to Swapfiets, but they are being manufactured and will be delivered at some point in batches. Furthermore, since Swapfiets offers bikes and e-bikes in a subscription form, it means that members have the opportunity to cancel their subscription, which results in a certain percentage of churn each month. Finally, a lot of repairs are performed in warehouses and stores, these bikes and e-bikes are also included. These figures only contain the supply of the next eight months, starting from August 2022 as the first month. The supply figures are only collected for the coming eight months due to a limitation of the demand forecast. The reason will be explained in more detail in Subsection 3.3.

3.1.1 Current Stock

Currently Swapfiets has bikes available in several warehouses and business units that are not being used. These bikes can immediately be used to satisfy new demand, or they can be used to exchange broken or missing bikes. These locations are a combination of 66 stores that keep a small amount of stock in The Netherlands, Belgium, Germany, France, Italy, Austria, Spain, England and Denmark. Furthermore some stock is also stored in the 26 warehouses. In total, Swapfiets currently has the following number of bikes available in stores and warehouses:

Table 3.1: Table with the number of currently available bikes in the stores and warehouses

Original 1	**Original 1+**	**Deluxe 7**	**Power 1**	**Power 7**
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

As these bikes are immediately available, they are included in the transportation model in the first month. This results in a big difference in supply levels in the first time period compared to the remaining periods.

3.1.2 External Inbound

The second part of the total supply originates from the bikes and e-bikes that have been ordered by Swapfiets. These bikes and e-bikes are being produced and will arrive in the coming months, and thus can be seen as a continuous stream of supply to include in the model. The external inbound stream of bikes and e-bikes is added to the main warehouse, which is located in Barneveld.

3.1.3 Churn

Another component of supply that should be incorporated, are bikes that are returned by customers that churned. Every month a percentage of all the active members cancel their subscription and return their bike or e-bike. To get an idea of the churn rate, the churn rates for Amsterdam and Delft have been analysed and compared visually, see Figure 3.1.

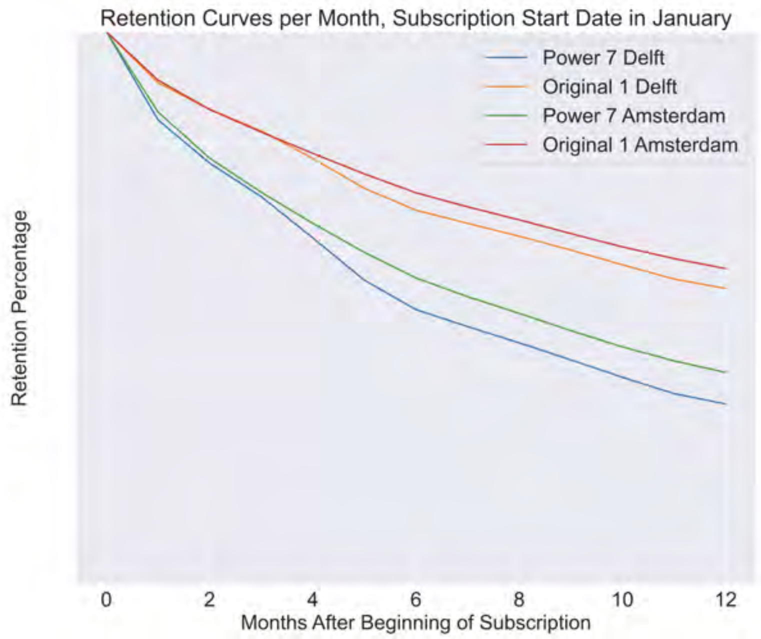

Figure 3.1: Plot showing the retention curve for Amsterdam and Delft for the Power 7 and Original 1 for the first 12 months after start of subscription

It can be observed that in Delft members churn much faster than in Amsterdam. Especially when the end point of the plot is compared at the far right in the figure. For Delft only between [redacted] and [redacted]% of the members still have their Power 7 or Original 1 after 12 months, while in Amsterdam these percentages range between [redacted] and [redacted]%. Furthermore, based on the figure it can be concluded that the churn rate differs per bike type. Between the Power 7 and Original 1 it shows that members with a Power 7, the e-bike, churn faster than members with an Original 1. Finally, the churn rates have also been analysed by the starting month of the subscription. The starting month of a subscription is one of the properties that can be used to segment different customer types. For example, customers that start their subscription around August or September are most likely students, it could be the case that these different customer types might also churn at different rates, See Figure 3.2.

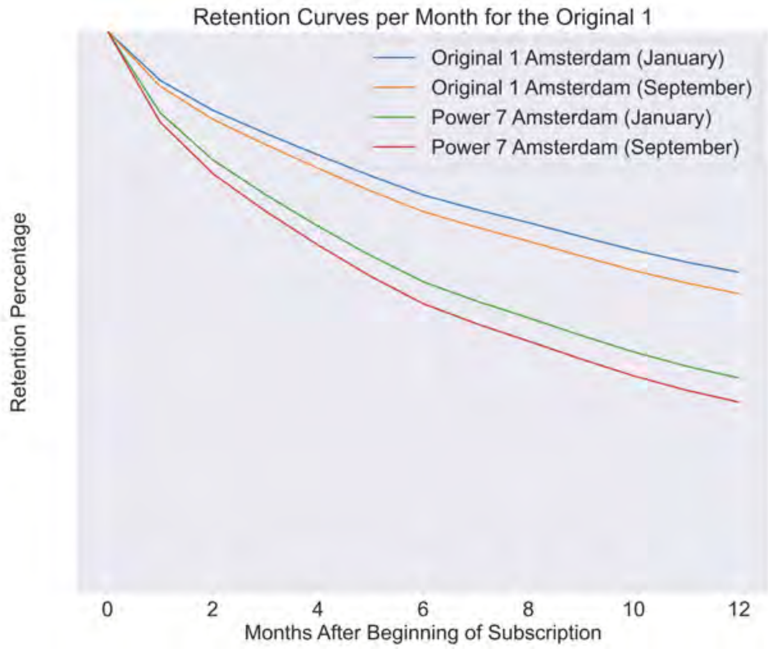

Figure 3.2: Plot showing the retention curve of Amsterdam for the Power 7 and Original 1 for the first 12 months after start of subscription in January and September

Figure 3.2 shows that also different retention curves can be observed when the member starts his or her subscription in a different month. The difference is smaller than it is for different bike types, but still present and thus will be considered in the transportation model.

These churn figures have been used in two different ways. In the first place, the bikes and e-bikes of customers that started their subscription before the eight months that are used in the transportation model have been included in the supply figures. The second part is regarding the churn percentages itself. These percentages are used to compute the churn that is caused by the model, by churn happening based on the bikes and e-bikes that are given to new members during these eight months. These numbers can be added to the supply of the transportation model again, as described more elaborately with an example in the previous chapter in Equation 2.13. Considering the observations made based Figure 3.1 and Figure 3.2, it can be concluded that using a churn rate that makes a distinction between the business units, different bike types and the different starting months is the most desirable method as this is as close to the reality as possible, which eventually increases the value of the results.

3.1.4 Repairs

Another source of supply originates from the repairs of broken bikes. These repairs are both done in the warehouses and stores. The stores mainly focus on the smaller and easy repairs. In total the following number repairs will be performed in the coming eight months:

Table 3.2: Table with the number of repairs that will be performed

Original 1	Original 1+	Deluxe 7	Power 1	Power 7
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Summing up all the supply figures, this results in the a total of [redacted] Original 1 bikes, [redacted] Original 1+ bikes, [redacted] Deluxe 7 bikes, [redacted] Power 1 e-bikes and [redacted] Power 7 e-bikes. To get a better understanding of the volume between the different sources, the supply has been visualized in Figure 3.3.

[redacted]

Figure 3.3: Figure showing the supply of the coming eight months for the Power 7 E-bike

The figure shows how the supply is build up for the Power 7 e-bike. It can be observed that only in the first month (at $t = 0$), logically, the current stock is considered. For the remainder of the plot it can be observed that the repairs account for the largest part of the supply. This pattern can also be found for the remaining bike types. As a final remark, after $t = 4$, the number of repairs show diverging values compared to the first 5 months, this is caused by the fact that the repair plan is only made for the current year. Earlier it is mentioned that due to limitations of the demand forecast it is only possible to include the coming eight months, but due to the fact that the repairs account for a large part of the supply and the plan is not made for next year, it makes sense that the model will only consider the first five months, starting from August.

3.2 Transportation Expenses

The stock that is available at the 66 locations needs to be transported to business units to fulfill the demand. Transportation will be done in trucks with a fixed capacity. One truck has a capacity of either [redacted] normal bikes, [redacted] e-bikes or a mixture of bikes and e-bikes. The transportation costs per truck from the warehouse in Barneveld to 22 different business units are known. However, these are only the costs from a warehouse to several business units, but the problem is more complex and requires more information, information about the transportation expenses for the 66 business units to all 26 warehouses.

Even though these 22 data points are very limited, this information can be used to come up with reasonable transportation figures for all the other business units as well. Using a Python package called "Geopy", the coordinates of these locations can be retrieved. With these coordinates it is possible to determine the distance between the warehouse in Barneveld and the 22 business units. Combining the distance between these locations and the costs that are given, the relationship between them can be analysed. These two measures have been visualised in Figure 3.4.

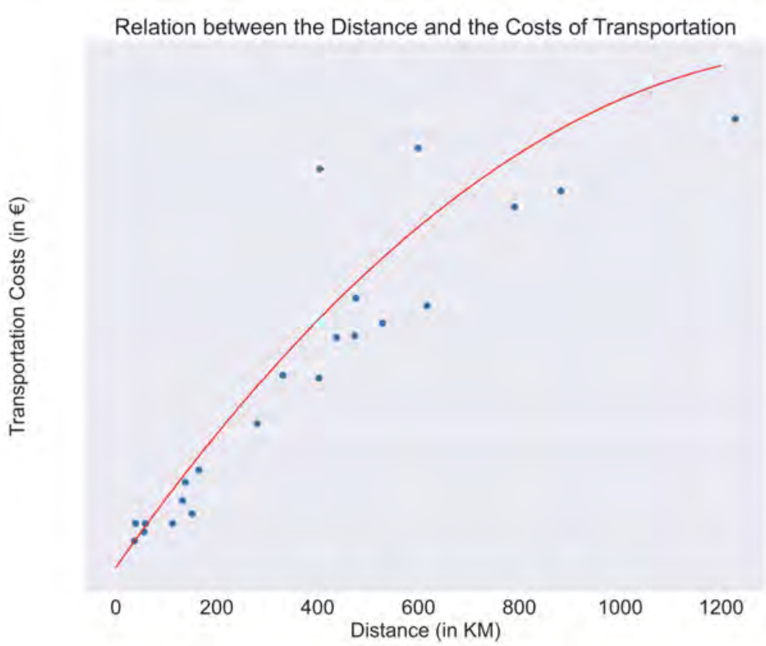

Figure 3.4: Figure of the relationship between the distance between locations (in KM) and the travel expenses (in €)

The plot shows a clear relation between the distance between the locations and the costs made for transport. The plot almost shows a straight line between the costs and distance, indicating a linear relation. Based on the figure, it can be concluded that a regression model could be a suitable approach to predict the missing transportation expenses for the other business units as well. Since the relationship cannot be completely explained by a perfect straight line, a polynomial regression analysis has been executed, as it could be the case that a concave downward function might offer a better solution to explain the relationship between the variables in the best manner. As a final remark, the two data points that have a distance between 400 and 600 kilometers and with costs over €[redacted] are the routes from Barneveld to London and Copenhagen. Logically, it makes sense that transporting products to the UK, who are not part of the EU anymore is expensive. Additionally, the distances are the distances in a straight line, and thus to cross the North Sea more kilometers have to be made. This also applies for the route between Barneveld and Copenhagen.

3.2.1 Polynomial Regression

A polynomial regression analysis with different degrees of polynomials have been used to model the relationship. In a lot of situation a linear relation between a dependent variable Y and any independent variable X is not sufficient and realistic, it could be that the rate of change is not stable for all possible values of X. In [11], Rawlings states that in these cases where the relationship is not linear, but contains more complexity, a polynomial regression model can offer a good solution.

There are different types of polynomial regressors that can be used, where the simplest non-linear form is the second-degree polynomial, better known as the quadratic model. The polynomial is of the following form:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{3.1}$$

In this form, X is also used in a quadratic form, X^2. This model can be extended by using a higher degree polynomial. In case the degree is higher than 2, not a quadratic model but a polynomial model is used. The model is given by the following form:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdots + \beta_p X^p \tag{3.2}$$

Where the polynomial is now of the degree p, and thus is called a pth degree polynomial. During this research different degrees of polynomials have been tested to find which degree results in the best fit. One of the advantages of using different degrees of polynomials is that also the linear regression model can be tested. The linear regression is used when the degree of the polynomial is equal to 1. This can be of use as Figure 3.4 only showed a slight bend in the relation which could indicate that a linear model should not be excluded from the possibilities.

To compare these different models, to see which degree is the best fit to describe the relationship between X (the distance in KM) and Y (the travel expenses in €), first the data has been split into a training set and a test set. A 70/30% split has been used. The next step is to determine a performance metric that can be used to compare the performance on the training set and test set for the different models. For this task the performance metric "Root Mean Squared Error" (RMSE) has been used, which is commonly used when it comes to regression analyses. As stated in [12], the RMSE is commonly used since it has a few advantages. In the first place taking the root of the Mean Squared Error makes the metric smaller and thus easier to grasp. Furthermore, an advantage of the RMSE is that squaring the errors assigns a relatively high weight to the larger errors, and thus penalizing larger errors more heavily. As stated in [13], the RMSE can be computed with the following formula:

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(p_n - \hat{p}_n)^2}{N}} \tag{3.3}$$

With p_n the real value, $\hat{p}_n$ the predicted value, and N the number of observations. The results of the analysis have been visualized in Figure 3.5.

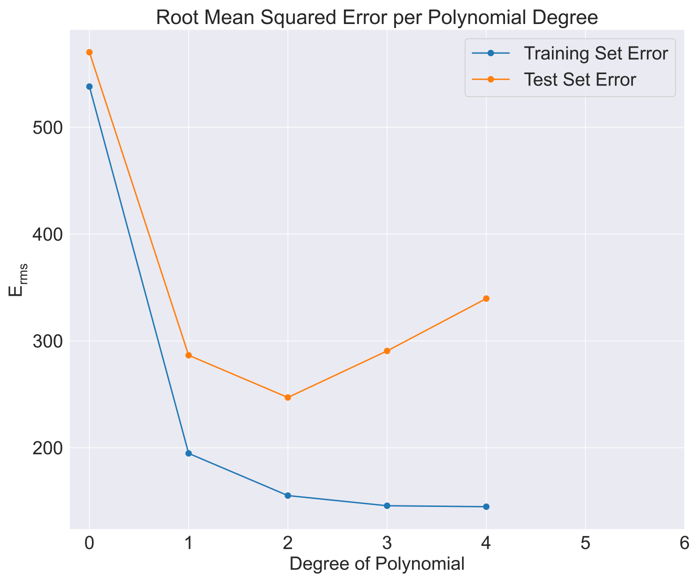

Figure 3.5: Figure that shows the Root Mean Squared Error of the polynomial regression models for different degrees

The figure shows that testing more degrees is not required, as the error on the test set increases when a higher degree than two is used. The main takeaway from the figure is that a second degree polynomial regression results in the best solution. This can be concluded by the fact that after 2 degrees the test set error increases again while the error of the training set decreases, this is an indication of an overfit of the model on the training set.

This means that the generalizability of the model becomes worse, which is undesirable. The polynomial has the following form:

$$Y = \text{[redacted]} + \text{[redacted]}X - \text{[redacted]}X^2 \tag{3.4}$$

What is interesting to see is the coefficient used at the seconds degree ([redacted]), this indicates that indeed the relationship between the distance and costs is almost linear but only has a very small bend, as observed in Figure 3.4. The function has been visualized in Figure 3.4. As a final remark, the model shows that the intercept of the model is at €[redacted], those are the costs that can be linked to filling and emptying the truck. However this means that even when the distance between two locations is zero, €[redacted] of transportation costs are used. Logically, this does not make any sense, and thus the costs of transportation are set to zero when the locations are the same.

3.3 Demand

The next component is the demand per business unit per bike type. The demand is required to know how many products should be transported to which locations, to both deal with the new sales and the exchanges. Currently Swapfiets has an on-going project where the data science team predicts the demand, which is the aggregation of new sales and exchanges, per business unit for each bike type. These predictions are made on a daily, weekly and monthly level with a forecasting tool from Amazon, called Amazon Forecast. The daily forecast predicts two weeks ahead, while the weekly figures are available for the coming two months and the monthly forecast for the next eight months. This forecast is on-going in a sense that continuous improvements are made to make it more realistic and reliable. The forecasted demand is automatically updated on a weekly basis to include the most recent figures of the number of new members. Before the data can be used, a decision needs to be made regarding the different levels of granularity in the forecast. The most desirable way to work with this data is to use the demand on a monthly level, as it is most realistic to transport bikes to fulfill a monthly demand instead of having to transport bikes on a daily level. To get a feeling of what these numbers look like, they have been analysed and visualised. In the first place, the number of active business units where Swapfiets currently offers bikes and e-bikes has been determined. Active business units can be defined as cities where Swapfiets currently has more than 50 active subscriptions. Currently there are 51 active business units in eight different countries. All these countries need to be included in the same model. The total demand per month has been visualised in Figures 3.6 and 3.7.

[redacted]

Figure 3.6: Plot showing the demand of the coming eight months for each bike type

[redacted]

Figure 3.7: Plot showing the demand of the coming eight months for each e-bike type

It can be observed that the number of Original 1 and Deluxe 7 bikes are much more volatile than the forecast for the Original 1+. In the left figure a peak for the number of Originals and Deluxe 7's in August and September ($t = 0$ and $t = 1$), at the beginning of the left plot can be observed. The explanation behind this peak is the fact high schools and universities start in September, and a lot of students travel to their universities with a bike, which results in a lot of potential customers. This pattern cannot be found for the e-bikes, which makes sense as kids and students are more likely to use normal bikes. For the e-bikes, only the Power 7 is visualized, since the Power 1 has a constant value for the predicted demand. This is due to the fact that the data for the Power 1 is not sufficient for the Amazon forecast algorithm to predict the demand, alternatively the algorithm just takes the average and uses it as a static value. In contrast to the normal bikes, the e-bikes show a completely different pattern. The number of e-bikes increase around February and March. These findings have been compared to the patterns in the previous year to confirm the validity of the forecast. In the previous year the same patterns were found for the bikes and e-bike.

3.4 Prioritization of Profitable Customers

It is not possible for Swapfiets to completely fulfill the total forecasted demand with the supply for the coming eight months. This means that bikes have to be given away with care to the right customers. To get an idea of the situation, see Figure 3.8 and Figure 3.9.

[redacted]

Figure 3.8: Plot showing the cumulative demand versus the cumulative supply for the coming 8 months for all bikes

[redacted]

Figure 3.9: Plot showing the cumulative demand versus the cumulative supply for the coming 8 months for all e-bikes

The left Figure shows the cumulative number of supply and demand for all the bikes for the coming eight months, while the right figure shows the same but for the e-bikes. Per bike type the same color has been used, but the supply is a solid line while the demand is a dashed line. For the Original 1 and the Power 7, the total demand already exceeds the supply that Swapfiets is able to deliver. For the remaining bike types this is not yet the case, but they all show the same pattern, which is that the rate of change of the demand is higher than the supply. This indicates that most likely eventually this will also become a problem and a prioritization of customers is desirable.

One way to tackle this problem is to define a measure which can be used to prioritize the most profitable customers. Luckily, between the different business units and customer types, there is a form of heterogeneity when it comes to profitability, which makes it possible to target the right customers. However, there are many different measures that can be used to quantify the profitability of a customer. One approach would be to use the lifetime value (LTV). The lifetime value can be defined as "*a calculation of how much profit a business could make from one customer over the whole period that they remain a customer*". The LTV would suffice as a good measure as it includes the whole process, from the customer paying a monthly subscription fee with a deduction of the expected costs that will be made over the whole period the customer remains a Swapfiets customer. To break it down into details, the customer lifetime value for a Swapfiets customer is computed in the following way:

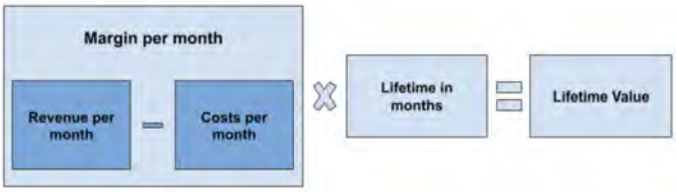

Figure 3.10: Lifetime value logic

The total monthly revenue of subscribers are computed, with a deduction of the percentage of bad debt, for customer who do not pay their monthly fee. The remaining number is increased by the monthly add-on revenue, which is revenue that originates from customers paying an extra fee to get add-on's such as baskets on their bike or e-bike. This number gets reduced by the average costs for a customer. To get an idea of what these numbers look like, they have been visualized, see Figure 3.11 and Figure 3.12.

[redacted]

Figure 3.11: Plot showing the average Lifetime Value of the Power 7 in general, for Amsterdam and Rotterdam

[redacted]

Figure 3.12: Histogram showing how the lifetime value is distributed for the Power 7 E-bike

In Figure 3.11, for Rotterdam and Amsterdam the LTV for the Power 7 has been compared to the total average LTV of the Power 7 per month. The LTVs show to be very volatile. The difference between the business units can be caused by multiple reasons, one of which could be the case that customers in one city need more repairs than in the other cities, or store & field costs could be higher for example. This shows that using different LTV values per month, per business unit is required to produce meaningful results. Furthermore, based on Figure 3.12 it can be observed that it is also possible to have negative lifetime values, which is caused when the costs are higher than the total revenue generated by a customer on average. Logically, bikes and e-bikes will not be given away to those new members, especially with the shortage of supply. These large difference between business units and different months are also present for the other bike types, and thus all bikes and e-bikes will have a different LTV per month per business unit. Moreover, a high variance can be observed in the histogram in Figure 3.12 even for the same bike type. This shows how diverse the values of different customer types are and proofs that it is important to target the right customers. As a final remark, it should be noted that also lifetime values for exchange demand are included, however, as they do not result in new customers, these values are set to zero.

3.5 Prioritization of Existing Customers

Swapfiets values satisfied customers more than attracting new customers. This can be translated to the fact that there should always be enough bikes to deal with all the incoming "bike exchange" requests. To get an idea about the number of exchanges that are done, compared to the number of new bikes that are given to new customers the following figure has been made:

[redacted]

Figure 3.13: The percentage of exchanges versus the number of new bikes given out per month for Amsterdam in 2021

The figure shows the number of completed exchanges for previous year, 2021, compared to the number of bikes that have been given away to new customers, the "First Deliveries" for Amsterdam in 2021. What can be observed is that the number of exchanges is in most cases higher than the number of bikes that are used for new customers. This makes it even more difficult to fulfill all the predicted demand with the total supply. Additionally, the figure also seems so show some seasonal pattern where the number of exchanges is lower during the summer months, just before students start at schools and universities again. To account for the exchanges, the forecasted demand has been split according to the exchange/sales percentages of 2021 per business unit per month per bike type into two types of demand (1) “sales”, meant for the new sales and (2) “exchanges”, meant to be used to exchange broken and missing bikes.

3.6 Leasing Costs

Another component to consider, that has been included in the transportation model, are the leasing costs for the bikes and e-bikes. Whenever a bike or e-bike is not used, and remains in the warehouse, monthly leasing costs are still being made. Logically, Swapfiets also pays leasing costs over the bikes that are being used, however those are already included in the lifetime values, and thus do not need to be incorporated again in the objective function. These leasing costs are €[redacted], €[redacted], €[redacted] and €[redacted] euros per month, for the Original 1, Deluxe 7, Power 1 and Power 7 respectively.

Chapter 4

Validation

The MIP model that has been described in the previous chapters results in a model of a size that is difficult to analyse, considering the fact that for five bike types only the network on its own in the first stage exists of 66 supply locations and 26 warehouses. This means that analysing all the results that come out of the MIP model to validate whether it works as intended becomes a difficult task. However, validating whether the model works as intended is vital to the reliability of the results. In this chapter the method that has been used to validate the model has been explained and the results are discussed.

4.1 Extreme Condition Testing

In [14], Sargent states that one of the approaches to validate the workings of a model is by applying a method called "Extreme Condition Testing". Normally extreme condition testing is used for simulation models, however for this project it can also be used. In extreme condition testing it is tested whether the structure and outputs of a model are plausible for any extreme and unlikely combination of levels of factors in a system. For example, if in a production process the inventory that is needed for production is zero, then logically the output should also be zero.

In the transportation problem, the driving factor that is mainly accountable for where bikes and e-bikes end up is the lifetime value of customers. As this value is much larger than the transportation costs per bike, most of the decisions are made based on these values. One way to apply extreme condition testing is by changing the lifetime values for a particular month for some business units before the model is executed. By doing this, and increasing the lifetime values by an unrealistic large amount, when working with a small subset of business units and months, it can be analysed whether the model prioritizes these locations. This also creates the possibility to check all the intermediate steps.

The following values have been used (1) business units: Amsterdam, Delft, Den Hoorn and Rotterdam, (2) the first three months ($t = 0, 1, 2$) for the Power 7 e-bike. These four specific locations have been selected as they are served by two different warehouses. Delft and Rotterdam are both served by the warehouse in Den Hoorn, while Amsterdam is served by the warehouse in Amsterdam. This is mainly done to test whether the business units are correctly linked to the warehouses. Furthermore only three months are used to be able to still track the flow of the products, as this becomes too complicated with a lot of periods. The part where extreme condition testing comes in, is in the manual changes in the lifetime value for Delft. For the lifetime value for sales demand for month 0 the value has been increased by €[redacted]. Which is compared to the values that are observed for the Power 7 in Figure 3.12 unrealistically high, [redacted]. For the same location for month 2 ($t = 1$) the lifetime value has been increased by €[redacted]. For these two months Delft showed to have low demand figures, so they have been increased to [redacted] and [redacted] for month 1 and 2, respectively. With these changes the total supply is [redacted], while the the total demand is [redacted]. The demand is split into [redacted] exchange demand, which results in a lifetime value of 0, but has to be satisfied, and [redacted] new demand. This means that only [redacted] bikes remain and can be used to satisfy the sales demand of [redacted]. For the first period ($t = 0$), the results are visualized in the following figure:

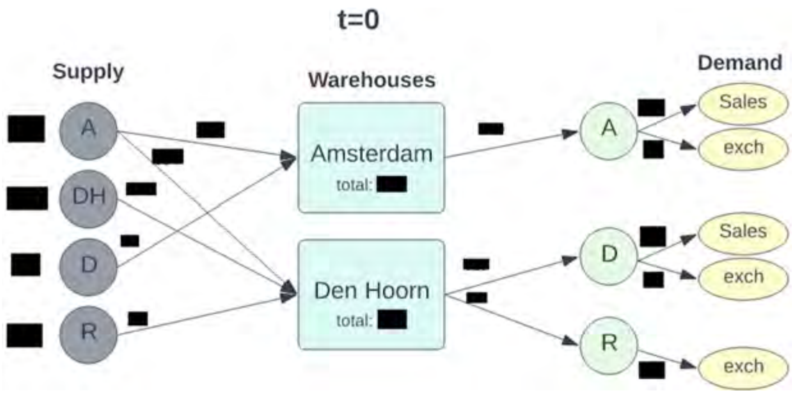

Figure 4.1: Flow of transported Power 7 e-bikes for a subset of the total data for month 1 (t=0)

The figure shows that in the first stage all the e-bikes are transported to the warehouses and no stock remains in the business units, even from Amsterdam to Amsterdam. This is not a problem as transporting bikes from for instance Amsterdam to Amsterdam results in zero travel expenses and transport in the second stage does not increase the costs either. These bikes are then transported to the demand locations which are connected to the warehouse then to be split over sales and exchange demand. As a final remark, Den Hoorn is not present in the last part of the figure as it only serves as a warehouse and thus has no demand. From the figure it can be concluded that the complete sales demand in Delft has been fulfilled. After this first month, the warehouses do not hold any stock anymore as it is completely given away to fulfill exchange and sales demand. For the second month, see Figure 4.2.

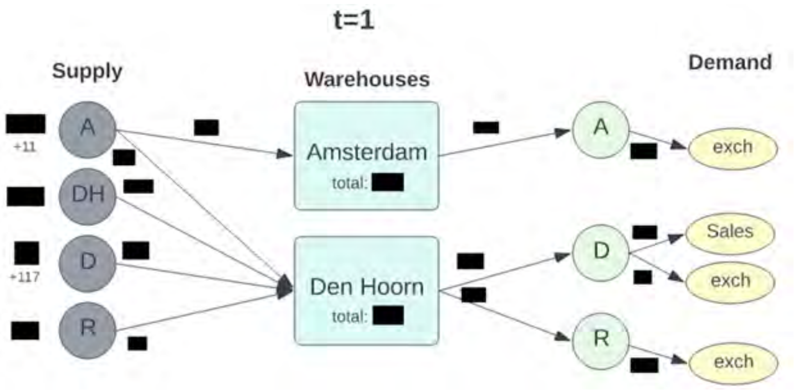

Figure 4.2: Flow of transported Power 7 e-bikes for a subset of the total data for month 2 (t=1)

The first difference with the previous figure is that now extra numbers are included for Supply (11 and 117), these numbers are a result of churn caused by satisfying sales demand in the previous month. Furthermore in this figure it can be observed that once again the sales demand in Rotterdam will not be satisfied because the lifetime value is not high enough. The difference, compared to the previous month, is that for Amsterdam now only the exchange will be fulfilled. As a final remark, it can also be observed that a part of the stock that is available in Amsterdam is not transported and is kept in Amsterdam for the next month to use. For the final month, see Figure 4.3.

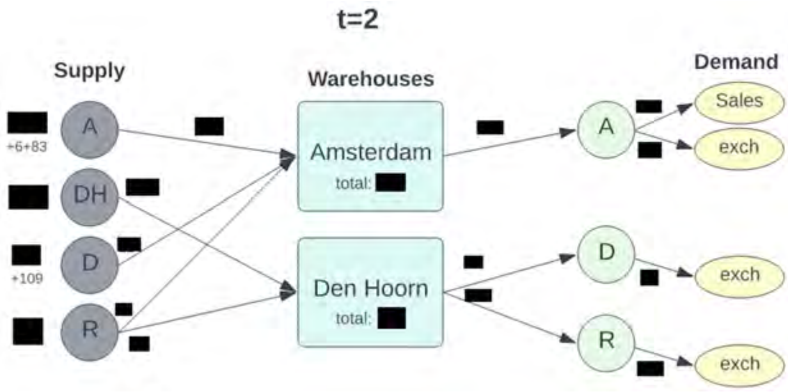

Figure 4.3: Flow of transported Power 7 e-bikes for a subset of the total data for month 3 (t=2)

Besides the external inflow of e-bikes, now for Amsterdam the total supply is a combination of churn (from the e-bikes given away at $t = 0$ and $t = 1$) and stock from the previous month (83). This shows the ability for business units to hold stock and use it in the next month. Furthermore now the demand for new sales in Delft is not satisfied, while in Amsterdam the new sales demand is (partially) satisfied once again. To get a complete overview of the logic behind how the demand is satisfied, see Table 4.1.

Business Unit	t	Demand (Exchange)	Satisfied (Exchange)	Demand (Sales)	Satisfied (Sales)	LTV	Rank	Percentage sales satisfied
Delft	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	1	100%
Delft	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	2	100%
Amsterdam	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	3	89%
Rotterdam	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	4	0%
Amsterdam	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	5	17%
Amsterdam	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	6	0%
Delft	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	7	0%
Rotterdam	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	8	0%
Rotterdam	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	9	0%

Table 4.1: Results of running the transportation model for a subset of business unit and periods

The table is sorted based on the “Rank” column. This column indicates the rank of the lifetime value for the particular location for a particular month. Based on the figure a few conclusions can be drawn. In the first place, the third and fourth column show that the exchange demand is always met. However, the most interesting conclusion is that by applying extreme condition testing, by increasing the lifetime values for Delft, it shows that the demand of these two most profitable business units are fully satisfied and the remainder is divided over the other most profitable cities. Interestingly enough Amsterdam at $t = 2$ is for 89% satisfied while the business unit that is two ranks lower also still receives bikes. Logically Amsterdam at t=2 should have been 100% as well, but the table shows that this is not the case. Multiple reasons can be the cause of this observation. For instance by efficiently filling a truck, the reduction in costs can out-weights the increase in lifetime value in case the bikes would have been used in Amsterdam at $t = 2$. Another, more probable reason in this case, is that giving out bikes at t=0, results in returning bikes due to churn for $t = 1$ and $t = 2$, these bikes can be used again to satisfy even more demand. This indicates that the model tries to find a balance between increasing the objective value by targeting demand with high lifetime value, giving out bikes early to create churn to give them away a second time and reducing transportation costs by efficiently filling trucks. By increasing the lifetime value of Amsterdam at t=2 by an unrealistically large number, €5000 in this case, it showed that the demand for new sales in Amsterdam at $t = 2$ is completely satisfied and the bikes will not be used in Amsterdam at $t = 0$ anymore.

The analysis shows that the model works as intended, by always satisfying exchange demand and then splitting the remainder over the most profitable cities. Furthermore, it shows that months are connected, by being able to carry over stock over time and the fact that churn in months is created based on the actions in previous months.

Chapter 5

Results

Previous chapter elaborated on validating that the model works as intended by using a subset of business units and periods. The next step is to analyse the results that the MIP model produces for the full period for all the locations. As stated in Chapter 3 the model has been executed for five months for 66 supply locations and 51 demand locations. The results of the model are shown in this chapter. In the first place a stress will be described, which shows how the computational time of the model develops by adding more locations and periods to the model. Secondly, the key results that are a results of the MIP model will be shown.

5.1 Stress Test

Different solvers have been used to tackle the transportation problem. The purpose of using different solvers is mainly to test if a commercial solver is required or whether a free solver would suffice. The free solver "GLPK" has been compared to the commercial solver "Gurobi". The intermediate model, that is described in Section 2.2 has been solved with GLPK with a subset of business units. When 15 business units were used, the GLPK solver already required 697 seconds to get to a solution. Considering that the final model is more complex and contains over 50 business units, the GLPK solver cannot be used. As a next step, the development of the computational times for the full transportation model have been analysed by working with subsets of business units and periods using the Gurobi solver. These subsets range from a small number of business units and a small subset of the periods to the full data set of 68 business units and 5 months. According to the documentation as stated in [15], the Gurobi solver works iteratively and computes a lower and upper bound for the objective value after each iteration. The absolute gap between these bounds is called the MIPGap (short for Mixed Integer Programming gap). Whenever the absolute difference between these bounds is lower than the threshold, the solver stops and the model is considered to be converged. The threshold of the MIPGap has been changed to 1%. The results are shown in Figure 5.1.

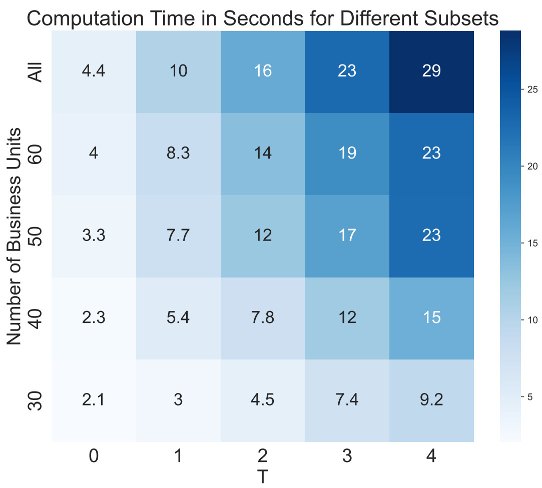

Figure 5.1: Heatmap of computational times for different subsets of the data using the Gurobi solver

The figure shows that the highest computational time that is experienced with the full data set is only 29 seconds. Furthermore it shows that the largest increases in computational time are caused by adding more periods to the model. Additionally the computational times when the default value of 0.001% is used for the MIPGap have been analysed. When the default value for the full multi-period model is used, the computational time is undesirably high, see Figure 5.2.

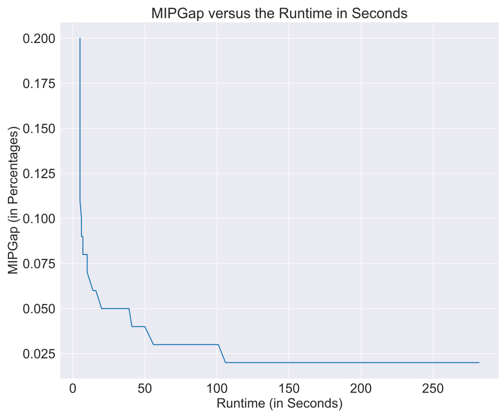

Figure 5.2: Plot showing the runtime of the Gurobi solver and the MIPGap

The figure shows the MIPGap on the y-axis and the runtime of the Gurobi solver on the x-axis. The figure starts after the first iteration for readibility purposes, as the gap after one iteration is 2.69%. Furthermore only the first five minutes of the run are shown. The figure shows how quickly the gap drops to a small value. After the second iteration the gap is already reduced from 2.69% to 0.20%. Furthermore it shows that after 100 seconds the MIPGap is reduced to 0.02% after which it changes to a horizontal line, and no improvements or only very small improvements are found. Even though only the first five minutes of the run are shown, the solver has ran for one hour after which it has been manually stopped since the MIPGap remained at 0.02%. Even though a MIPGap of 1% has been used, as the figure shows the first time the model finds a result with a lower gap, it is already at 0.20%, which is very small.

5.2 Results of the Mixed Integer Programming Model

Considering the size of the model, finding a visual solution to represent the results is a difficult task. For example, only the table that contains the number of bikes that have to be transported contains 42,900 rows of which 558 rows contain non-zero values, and thus are hard to visualize. Alternatively, the results, the values of the variables, are stored in tables. Per variable, one table has been used. To illustrate, a subset of the results for x are shown in the following table.

[redacted]

Table 5.1: Table showing the partial results for the variable x

The names of the columns are self-explanatory except for "value". Value indicates the number of bikes that needs to be transported. The first row of the table means that 130 Original 1 bikes should be transported from the store in Enschede to the warehouse in Den Hoorn in the month $t = 4$, which is November. The table with the number of trucks follows the same pattern.

Regarding the remaining variables, y explains how bikes are transported from the 26 warehouses to the 51 business units, after which the bikes are distributed within the business units over sales and exchange demand, which can be found in the table with the results for the variable z. The results of z have been analysed visually, see Figure 5.3.

[redacted]

Figure 5.3: Figure showing the split between satisfying sales demand and exchange demand for the full duration of the model

The figure shows how the fulfilment of the total demand is split between satisfying sales demand and exchange demand for the five periods that are considered in the model for all bike types. What is interesting to see is that the sales demand that is satisfied decreases, while the exchange demand slightly increases, and thus making the share of exchanges large compared to sales. This pattern where sales fulfilment decreases and exchange fulfilment increases can be found for each bike type separately.

The tables that contain the results for the variables I and b, show the number of bikes and e-bikes that are in stock at any moment in time per business unit and per warehouse. These figures have been visualised in histograms, see Figure 5.4 and 5.5.

[redacted]

Figure 5.4: Histogram showing the frequency of the number of bikes and e-bikes that are in stock in the business units

[redacted]

Figure 5.5: Histogram showing the frequency of the number of bikes and e-bikes that are in stock in the warehouses

The left figure shows a histogram with the sum of all bike types that are in stock per month per business unit, similarly the right figure shows the same information but for the warehouses. What is interesting to see is that the for the number of bikes and e-bikes that are in stock in the business units, three strange values can be observed. These are the three bars on the right side of the figure, which means that there are business units that have a lot of bikes in stock. Upon further investigation these figures are the total stock levels in Heidelberg in Germany. Heidelberg has a supply of around [redacted] Deluxe

7 bikes per month. However, in the time window that is used in the model, a period of five months, the supply of Deluxe 7 bikes exceeds the demand. This means that there is fraction of the bikes that are not needed and are just being stored in either stores or warehouses. That is what is happening in Heidelberg, the supply comes in, piles up and is not being used. This means that these three bars show the development of the stock levels of Heidelberg. As a final remark regarding the stock levels, logically, the right figure shows that stock levels of the warehouses are much higher than in the business units.

Also the churn variable r, that represents the bikes that return due to churning customers, is stored in a table. In addition to the original supply plan, which contains a total of just over [redacted] bikes, [redacted] bikes and e-bikes are added to the total supply due to churning customers that churn in the five months that are used in the model. Comparing this to the total amount of bikes and e-bikes that are used to fulfill sales demand, which is [redacted], it means that within five months already [redacted]% of the customers churn.

When looking at the variables in more details, some more interesting comparisons can be done. For example, x and *trucks* have been compared and visualized. In total [redacted] trucks have been used to transport a combined total of [redacted] bikes and e-bikes. Of this total number the majority consists of bikes. In total [redacted] bikes are transported. A bike is of size [redacted]/[redacted] compared to an e-bike. This means that these [redacted] take [redacted] slots in the trucks. Adding this to the remaining e-bikes, in total [redacted] slots are needed for the complete transportation. In total the [redacted] trucks offer a capacity of [redacted] bike slots, while only [redacted] were required. The efficiency of how the trucks that are not completely filled is visualized, see Figure 5.6.

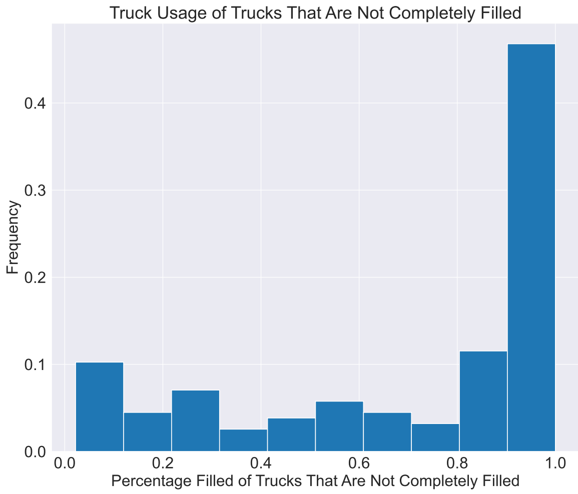

Figure 5.6: Histogram showing the relative frequency of the trucks that are not completely filled with bikes and e-bikes

Each truck has a capacity of either [redacted] bikes or [redacted] e-bikes. However these trucks have been filled in a mixed way, where both e-bikes and bikes fit in the same truck, by adjusting weights in the model. To clarify, for example for Aarhus to Copenhagen at t=0 a total of 1.18 trucks was required to make up for [redacted] bike slots. Logically, the model suggests to transport these bikes and e-bikes in 2 trucks, which means that the second truck was only filled for 18%, and thus 82% was wasted. The percentage that is filled of this last truck, the 18%, is visualized in the figure on the x-axis. The y-axis shows the frequency. It can be observed that almost half of the trucks are completely filled.

Instead of analysing variables separately, the total results of the model have been analysed as well. In the first place, the most prominent measure to analyse is the profit, which is the objective value. Following the strategy as suggested by the model, a profit of €[redacted] can be achieved. However, this profit is build out of different components. In the first place the sum of all the lifetime values, secondly a deduction of transportation expenses and finally a deduction of leasing costs. By combining the results table of the variable z with the lifetime values it is possible to compute what revenue can be generated. This resulted in a total of €[redacted]. When the results of the table containing the values for the transport in terms of trucks is combined with the transportation expenses, it can be concluded that for transport the costs will be € [redacted]. The second type of costs are the leasing costs, which makes up for the largest part of the total costs. Combining the warehouse stock and store stock with the lease costs resulted in a total €[redacted] leasing costs.

5.3 Multi-period versus Single-period Model

In Chapter 2 it was explained how the model was developed, from a single-period to a multi-period model. Another option that could have been executed is to work with a single-period model that can be used five times to tackle all the moments in time ($t = 0$ to $t = 4$) separately. It is interesting to analyse the added value by creating the multi-period model instead of running a single-period model multiple times. To show the difference, the following visualisation has been created:

[redacted]

Figure 5.7: Plot showing the different objective values of the single-period models, cumulative single-period runs and multi-period models

The blue line represent the objective values that are generated by running the model as a single-period model, by including one period per run. For the last period ($t = 4$) no value is shown as the model could not get to a feasible result. The orange line shows the cumulative sum of the objective values of the separate single-period runs, this line results in a total profit of €[redacted] and is required to make a fair comparison to the multi-period model. The third line, the green one, shows the profit that can be achieved by using a multi-period model. The model has been created for different time periods, where for instance at $t = 2$, the model has been executed for 3 periods at once ($t = 0$ - $t = 2$). This means that the final result of the complete model can be found at $t = 4$. The model results in a total profit of €[redacted], which is an increase in profit of €[redacted].

Chapter 6

Discussion

During this research the aim was to tackle the extended version of a multi-period two-stage transportation problem. In this study, Linear Programming was used to formulate a logistics strategy in which the transportation costs are minimized and the most profitable customers are prioritized for the complete network of business units and warehouses. This model has been used for a period of five months, starting in August 2022 and ending in December 2022. In this chapter, the key results are interpreted, after which the limitations are discussed. In additions, a number of recommendations are given for future research.

6.1 Interpretation of Results

As mentioned, during this research a MIP model has been used to create a logistics strategy. In the previous chapter, specifically in Section 5.1 it has been discussed that multiple solvers have been tried to solve the mathematical model. The idea of using multiple solvers is to test whether a free solver, in this case "GLPK", would be sufficient to solve the model, or that a paid version like "Gurobi" is required. It showed that even when the intermediate model, see Section 2.2, is solved with the free solver, it is not possible to work with the complete Swapfiets network of business units. The model resulted in a computational time of 697 seconds for a subset of 15 business units, while the full set contains over 50 business units. This clearly shows that it is required to work with a commercial solver like Gurobi. Additionally, the scalability of the MIP model has been tested. The scalability has been tested by analyzing how the computational times increase by adding more complexity to the model, by extending the single-period model gradually to a multi-period model with eventually five periods and by increasing the number of business units that are used. Eventually Swapfiets would like to run the model for more than just five months and thus it is interesting to analyse how the model develops. Figure 5.1 shows that by adding more periods to the model, the computational time increases more than whenever more business units are added. This makes sense since adding business units only adds a few more transportation routes to the network, while adding a month adds the complete network as a whole to the model again. Even though adding more months increases the computational time, the total time in seconds is at a level where adding more months should not result in undesirably large computational time and thus it is easily achievable. Furthermore, when it comes to the model performance in terms of computational time, the MIPGap has been adjusted. It showed that after the first iteration of the Gurobi solver, the gap is at 2.69%. While after the second iteration the gap already decreased to 0.20%. Since it is desirable to have a model that is able to produce quick results, especially if the model is going to be used in an interactive environment, the MIPGap after which the model can stop iterating has been changed to 1%. Based on the pattern of how the MIPGap decreases quickly during the first few iterations, as observed in Figure 5.2, it is expected that in most cases this will result in an model that is much closer to the optimum than it is in case at the threshold of 1%, with values around 0.20%, which can increase a bit once the complexity of the model is increased by adding more periods. Additionally, the figure shows that after 100 seconds of runtime almost no improvements are found and thus the trade-off of a slightly more accurate model, which is closer to the optimal, and the huge increase in computational time is not worth it. Especially considering the fact that even after running the solver for one hour the MIPGap remained at 0.02%.

Besides the performance of the model in terms of runtime and scalability the key results that are produced by the model also show some interesting insights. In the first place the revenue that can be generated in the coming five months by following the strategy that results from the model is €[redacted], which results in a profit of €[redacted], after the deduction of the transportation costs of €[redacted] and the leasing costs of €[redacted]. Unfortunately it is difficult to compare this to a baseline, as there is no data that shows the expected profit or revenue for the coming months based on the unused bikes, as done by the MIP model.

What is remarkable is the cost of the transportation. The model initially started as a basic transportation model, that is focused on reducing these costs. One of the reasons for the transportation model was that the general idea within Swapfiets is that transportation should be efficient in a sense that trucks should be filled as much as possible to become a profitable company. Figure 5.6 shows that for the trucks that are not completely filled, most of the trucks are still filled quite efficiently, as this is indeed a way to increase the profitability. However, considering the fact that the total costs of transportation are really low compared to the lifetime value and the leasing costs, it can be concluded that minimal improvements can be made here and thus should not be the main priority of Swapfiets.

Furthermore, the cumulative results of the single-period model have been compared to the results of the multi-period model in Figure 5.7. From the figure it can be concluded that the multi-period model creates a strategy that results in a profit that is almost 2 millions higher than the single-period approach, which is a significant amount for Swapfiets. Additionally, by using a multi-period approach it is possible to fulfill all the exchange demand, which should always be the case. This is not the case for the single-period approach, which is caused by the fact that the exchange demand for bikes and e-bike exceeds the supply at $t = 4$. The multi-period approach is able to deal with the problem by saving supply from earlier months.

As a final remark, knowing that Swapfiets is able to account for all the demand until the end of the year is already a valuable insight. Currently Swapfiets is under the impression that even until the end of the year they do not have sufficient supply to account for all the demand. As a result of this, some business units are "turned off". This means that for some business units some bike types are currently not offered on the website, which means that potential customers are not able to subscribe for a new bike, and thus Swapfiets is missing out on revenue. With this information they can undo this and offer bikes again.

6.2 Limitations

Next to the main results of the MIP model, it is important to understand the limitations. One of the limitations that have a huge impact on the results are possible inaccuracies of the lifetime value and demand forecast figures. The entire model depends on these figures and it could be the case that these numbers are incorrect for many reasons. For instance, the data showed that for some business units for some bike types a constant value for the demand results from the forecast, which is caused by the fact that the prediction algorithm did not have enough data to make an accurate forecast. In this case, the forecast is replaced by the average sales for the business unit. Another limitation is that in the first place it seemed like the model can only be used for the coming eight months, as the forecast does not predict any further into the future, but unfortunately the model can only be used for the next five months, due to the repair plan. The repair plan is currently only made for 2022, which means that using the model for more months results in a shortage of supply as these repairs have a value of zero. A final limitation is the lack of transportation costs.

The only transportation costs that are available are the transportation costs from the warehouse in Barneveld to 22 business units, which is only a fraction of what is required for the model. This means that now the model works with predicted transportation costs based on the small data set with 22 observations. Luckily, the transportation costs are small compared to the revenue generated and the leasing costs.

6.3 Recommendations

There are numerous ways in which the MIP model can be improved to be more valuable. In the first place, the model consists of separate components that can be updated and improved one by one, which makes it easy to continuously improve the model. For example, the amount of data that is available keeps on growing as Swapfiets matures. This expanding data set can be used in the Amazon Forecast to create more accurate predictions, and for the business units where the average is used, an actual prediction can be created. Similarly, regarding data availability, once the repair plan is created for 2023 it is possible to use the model for the coming eight months instead of the coming five months.

Furthermore, in the current set-up the demand forecast that is used is treated as the truth, and does not include a stochasticity component. As concluded in [4] by Shiang-Tai Liu in 2003, knowing the exact values of all parameters is not realistic possibly due to changing economic conditions. These changing conditions could result in under- or over-performance compared to the forecast. To deal with this problem it can be interesting to work with different scenarios. For example one scenario where the demand exceeds the prediction and one where the forecast is higher than the actual demand. This would make it possible to create a model that works best on average. Alternatively, a more intelligent approach would be to change the MIP model to a Stochastic LP model, where probability distributions can be used for the random parameters.

Once more data is available and the model can be used for more than five months, the gap between the demand and supply will grow, as the demand increases faster than the supply. This means that it could be the case that for some business units in consecutive months the demand is not fulfilled. Visualizing this can create valuable insights for the marketing department for example. In an earlier state when not all the supply figures were collected, it was the case that the demand exceeded the supply by a huge amount. This was visualized in the following way, see Figure 6.1

business_unit	0	1	2	3	4	5	6	7
Aarhus	100%	100%	0%	100%	100%	100%	100%	100%
Delft	0%	100%	100%	100%	100%	100%	100%	0%
Den Bosch	0%	0%	100%	0%	0%	100%	0%	0%
London	61%	100%	100%	100%	30%	100%	100%	36%
Milan	0%	0%	0%	0%	0%	0%	0%	0%
Vienna	0%	0%	0%	0%	100%	100%	0%	100%

Figure 6.1: Percentage of satisfied demand per business unit for consecutive months

The figure shows the percentage of new sales demand that is satisfied per business unit for the full period the MIP model is used for. From this figure it could have been concluded that in Milan for the coming 8 months no demand will be satisfied. This can be a reason for the marketing team to slim down their marketing campaigns in Milan. A similar conclusion can be drawn for Vienna.

A final interesting addition to the model, would be to include the possibility to change the revenue measure in the objective value. Currently the model uses the lifetime value of a customer. However it could also be interesting to change the measure to not look at the full lifetime, but only a subset of the lifetime value. For example, this makes it possible to create a logistical strategy for 12 months to determine the maximum profit that can be achieved within a year.

Chapter 7

Conclusion

This research followed the development of a mixed integer programming model that is able to solve a complex multi-period two-stage transportation model for a company that works with multiple subscription-based products, such as Swapfiets. Using this approach, the following research question has been addressed:

"*How can optimization algorithms be used to redistribute products of subscription-based companies over a set of demand locations such that supply and demand can be matched to increase profitability?*"

The fact that the model deals with a subscription-based product adds a lot of complexity to the model. It means that customers have the option to cancel their subscription. These products will come back to Swapfiets and can be added to the supply plan. Considering the fact that the model works with multiple periods, it means that based on the decisions that are made in the earlier periods in the model, the number of bikes that can be used for the supply differs later on. For the second component of the transportation process, the transport itself, the complexity is in the fact that products are not transported per piece, but they have to be transported in trucks to the warehouses. These trucks can contain a mixture of different product types, where different products can differ in size. This means that the model has to decide how trucks can efficiently be filled to reduce the transportation expenses. Regarding the final stage of the transportation model, supply is used to satisfy exchange demand and new sales demand. Swapfiets does not have enough supply to satisfy the exchange and sales demand in the long-run. As exchange demand always has to be satisfied, it means that the remainder of the supply has to be used carefully. Based on the profitability of different customers stock has to be used wisely, by focusing on the most profitable customers.

All these components have been included in the model and considering the results presented in Chapter 5, to answer the research question, it can be concluded that a mixed integer programming model is highly suitable to solve a multi-period two-stage transportation problem for a subscription-based company while keeping the computational times acceptable. Additionally, one of the main advantages is the flexibility of the model. Processes that are very company specific can accurately be incorporated in the model since each component of the model is programmed from scratch. Another aspect that is a huge advantage of working with a MIP model is the traceability of the model. In contrast to some machine learning algorithms, where the model can be a black box, in a mixed integer programming model all the steps can accurately be observed and analysed.

As a concluding remark, the model meets, and in some cases, exceeds all the requirements that have been stated at the beginning of the project and is going to be used for multiple different purposes. The transportation model is a good step towards outgrowing the scale-up phase and maturing in a stable profitable company.

References

[1] Frank L Hitchcock. **The distribution of a product from several sources to numerous localities**. *Journal of mathematics and physics*, **20**(1-4):224–230, 1941. 3

[2] George Dantzig. *14. The Classical Transportation Problem*, pages 299–315. Princeton University Press, 2016. 4

[3] Oziely Armenta[1], Aide Maldonado-Macías, Liliana Avelar Sosa, Guillermo Cortés, and Jorge Limón Robles. **Use of Transportation Methodology to Maximize Profits of a Private Transporter**. *Intelligent Decision Support Systems*, page 81, 2016. 4

[4] Shiang-Tai Liu. **The total cost bounds of the transportation problem with varying demand and supply**. *Omega*, **31**(4):247–251, 2003. 4, 48

[5] K Antony Arokia Durai Raj and Chandrasekharan Rajendran. **A genetic algorithm for solving the fixed-charge transportation model: two-stage problem**. *Computers & Operations Research*, **39**(9):2016–2032, 2012. 4

[6] Kai Zhu, Kaiyuan Ji, and Jiayu Shen. **A fixed charge transportation problem with damageable items under uncertain environment**. *Physica a: statistical mechanics and its applications*, **581**:126234, 2021. 5

[7] DG Mogale, Alexandre Dolgui, Rishabh Kandhway, Sri Krishna Kumar, and Manoj Kumar Tiwari. **A multi-period inventory transportation model for tactical planning of food grain supply chain**. *Computers & Industrial Engineering*, **110**:379–394, 2017. 5

[8] G.B. Dantzig and M.N. Thapa. *Linear Programming 1: Introduction.* Springer Series in Operations Research and Financial Engineering. Springer New York, 1997. 7, 8

[9] Wayne L Winston and Jeffrey B Goldberg. *Operations research: applications and algorithms*, **3**. Thomson Brooks/Cole Belmont, 2004. 8

[10] Horst A Eiselt and Carl-Louis Sandblom. *Operations research: A model-based approach*. Springer Science & Business Media, 2012. 15

[11] John O Rawlings, Sastry G Pantula, and David A Dickey. *Applied regression analysis: a research tool*. Springer, 1998. 26

[12] Tianfeng Chai and Roland R Draxler. **Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature**. *Geoscientific model development*, **7**(3):1247–1250, 2014. 27

[13] Miriam Steurer, Robert J Hill, and Norbert Pfeifer. **Metrics for evaluating the performance of machine learning based automated valuation models**. *Journal of Property Research*, **38**(2):99–129, 2021. 27

[14] Robert G Sargent. **Verification and validation of simulation models**. pages 166–183, 2010. 33

[15] Gurobi Optimization, LLC. **Gurobi Optimizer Reference Manual**, 2022. 38