Using discrete event simulation to quantify performance of a car-sharing network with a heterogeneous shared fleet

Seminar Paper

Author: Luca Elias Fanselau (Student ID: 7369806)
Supervisor: Univ.-Prof. Dr. Wolfgang Ketter
Co-Supervisor: Muhammed Demircan

Department of Information Systems for Sustainable Society
Faculty of Management, Economics and Social Sciences
University of Cologne

January 23, 2022

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Luca Elias Fanselau

Köln, den 23.01.2022

Abstract

In the rapidly growing market of car-sharing mobility, operational decision-making can determine success or failure for car-sharing organizations. Predicting the impact of these decisions analytically is a complex and high dimensional problem. Further complexity comes from the heterogeneity of the vehicle fleet a customer is offered. This paper proposes a discrete event simulation that aims to provide quantitative metrics, indicating the performance of the car-sharing networks and therefore providing a computationally efficient way to analyze the effect of operational decisions. The simulation environment models a non-floating one-way car-sharing system, equipped with a vehicle choice classifier that is trained using existing car-sharing rental data. The classifier receives a customer request as input, that is defined by socio-demographic factors as well as the planned route. The proposed framework is then implemented for a specific car-sharing deployment in Berlin, Germany. The dynamics of the model, in terms of the presented metrics, are then analyzed in regard to the fleet size as well as the substitution effect, that captures the choice uncertainty when selecting a specific vehicle type that is typically present in real-world decision-making.

Contents

Luca Elias Fanselau

List of Figures

List of Tables

List of Symbols

C	Parameter of the simulation model. Represents the maximum capacity of each vehicle class at each station
α	Parameter of the simulation model. Represents the substitution effect.
$\mathbb{C}$	The set of all vehicle class choices
$\mathbb{R}$	The set of all real numbers
$\mathbb{N}$	The set of all natural numbers
$cost$	Function that maps the class and rental time to a price
$\mathbb{X}$	Set of socio-demographic categories that define an area
a	A specific area in regard to the categories $\mathbb{X}$
$\mathbb{S}$	Set of all stations in the simulation
$\mathbb{E}$	Set of all connecting edges of the graph that is formed by $\mathbb{S}$
$L : \mathbb{E} \to \mathbb{R}$	Distance mapping for each edge
Δ_s	The delta function of a station. Indicates the number of vehicles of each type that is rented up until a specific simulation time. Always greater than $-C$
r	A customer request, containing a start station $s_{r,0}$ and an end station $s_{r,1}$
D	A function that returns the set of all feasible vehicle choices for a request, with regard to the substitution effect
D'	Restricted subset of D that checks availability of the vehicle type at $s_{r,0}$
URR	Unsatisfied Request Ratio
π	The total profit of the simulation stage
TD	The total distance driven during the simulation stage

1 Introduction

Car sharing services have recently seen a rapid rise in popularity and valuations of the global car-sharing market expect a growth to a total value of USD 6.5 billion by 2024, from just USD 1.1 billion in 2015 (Jiyeon Jung, 2018, p. 1). A primary reason for that is the high value-proposition a car-sharing service can offer to its customers. This includes positive environmental impact in the form of reduced CO2 emission by up to 312 kg CO2 / year per individual, as well as lower individual mobility costs, compared to owning a vehicle (Jörg Firnkorn, 2011, p. 1525). Together with the broad adoption of smartphones, which, thought the use of custom apps, can support the reservation and operation of shared vehicles, Car Sharing Organizations (CSOs) can provide an appealing alternative to other public or private transportation measures.

These services can be categorized into free-floating, meaning vehicles of the fleet can travel freely in a restricted area, and non-floating systems, where vehicles must travel between discrete stations that typically offer a fixed amount of capacity. Additionally, there is also a distinction between one-way and two-way car-sharing systems, the former describes services where the user can travel from point A to point B, which he can choose freely, while the latter expects the user to return the vehicle to the spot where it was rented.

Operating large scale on-demand car-sharing networks tends to be difficult, since deploying operational decisions is often bound to large expenses and, especially if the system is already in use, cannot be done in a trial and error fashion. One way to test the feasibility of seemingly optimal proposals quickly is to develop a simulation which tries to capture the real world interaction of actors in a car sharing network to a sufficient degree. With this, an operator can quickly assess if the solution holds up to various real world restrictions that are characteristic for car sharing networks, such as demand served or, in the case of an EV car-sharing platform, the vehicle charge levels (Burak Boyaci, 2016, p. 224)

The focus of this seminar thesis is firstly, to analyze existing literature on CSOs and especially the use of simulation environments and the optimization problems that were solved using them. Secondly, to use this information to design and implement a discrete event simulation, that models a non-floating one-way on-demand car-sharing service. The simulation is then equipped with a vehicle choice classifier model, based on socio-demographic and request specific features, to model the real world mobility dynamics of a car-sharing network. The main objective of the proposed framework is to use metrics obtained by the simulation to study the impact of fleet size and substitution effects on overall performance, evaluated on the specific case of SHARE NOW in Berlin, Germany.

2 Literature Review

A lot of pervious research on car-sharing and its impact has been conducted throughout the last two decades. The focus of this section is to provide an overview of papers that convey relevant information to the simulation environment or the topic of vehicle choice. For a broader literature review and further research pointers the reader can refer to dedicated papers on the topic such as Illgen and Höck (2019) or Ferrero et al. (2018).

Jian et al. (2017) has proposed a Multiple discrete-continuous extreme value (MDCEV) modeling framework to predict a user vehicle type preference. Additionally, a simulation procedure is employed to evaluate the MDCEV models, these results are then quantified by goodness-of-fit in regard to the data source. The proposed model was developed to provide a useful tool for operators of car-sharing fleets. They could prove a consistency of customer usage patterns regarding travel time, mileage and monetary expenditure. Furthermore, an effect of various socio-demographic features such as income, age, urban structure was discovered, indicating a statistical significance of these features regarding the vehicle choice, that will also be discussed in this thesis. As part of the MDCEV modeling procedure the satiation parameter was determined, that captures the willingness to drive more with the current choice of vehicle type. It indicated that this correlates with the usability, ease of driving and driving experience. Overall this paper laid down a basis for the vehicle choice model in this paper and provided sufficient evidence that a statistical important of the described features exists based on real world data.

Mehdi Nourinejad (2014) research has provided a dynamic optimization-simulation model for one-way car-sharing networks, that acts based on online requests with a reservation time and tries to maximize system profit. This is in contrast to the on-demand framework that is developed throughout this thesis. Their research aims to study the effects of fleet size and factors which impact fleet size and vehicle relocation cost. A benchmark model is proposed that has complete knowledge and calculates an optimal solution, together with the dynamic model, that could be used as a support tool for CSO operations. They also outlined the benefits of a discrete event simulation framework from a computational standpoint. Findings of the paper included a positive correlation of optimal fleet size and demand and that the mean required fleet size per demand decreases with higher demand. They also were able to quantify the operational effects of reservation times of 30 min, in contrast to the on-demand model, can lead to a reduction in fleet size of up to 86%. Additionally, they concluded that the dynamic model, without total knowledge, converges to the results calculated by the benchmark model and therefore verifies the feasibility of a discrete event driven simulation for further studies.

Perboli et al. (2017) have used a simulation approach to study the effects of customized tariff plans from the business model point of view. The proposed tool simulates different tariff plans with regard to different profiles of car-sharing users, according to mobility needs and urban structure. This is verified with the specific case of Turin, Italy. During the study, different car-sharing organizations are characterized in terms of their business model canvas, a structured form of the companies main business aspects. A separation of user profiles was made, into three distinct classes: commuters, professional users and casual users, each differing in time of departure and range of distance. Findings for the specific case of Turin include a favorability of dynamic pricing structures and concluded that private vehicle ownership is beneficial, in regard to the best performing car-sharing service in Turin at that time, only after about 8000 km/year for casual users and 10000 km/year for commuters or professional users. Additionally, a strong emphasis was made on the importance of simulation based models in car-sharing operation management.

Jochem et al. (2020) have conveyed a survey to study the effects of car-sharing membership on individuals, with the specific case of SHARE NOW which is also the basis for the simulation in this thesis. They specifically focused on the effects of such a membership on the car ownership. In an optimistic case one car-sharing car could replace up to 20 regular vehicles.

Burak Boyaci (2016) developed an integrated multi objective mixed integer linear programming optimization framework to optimize operational decision in regard to vehicle and personnel relocation in a car-sharing system of electric vehicles with reservations. The output of the optimization phase was then tested in terms of feasibility by a discrete event simulation stage. The determining factor of feasibility used is if the charge levels of electric vehicles would be enough the satisfied a calculated schedule. The car-sharing system covered in this paper was a non-floating one-way car-sharing system, as also covered in this thesis. Findings of this study includes a near linear dependency of demand to amount of relocation necessary and also verified the positive benefits from an operational standpoint that car-sharing services with reservation times enable, which was also discovered by Mehdi Nourinejad (2014). Additionally, the benefit of controlling demand by incentives/disincentives was discovered, as a higher amount of requests with the same resources could be served. This paper also used a clustering algorithm to group demand patterns into general area, which closely mimics the concept of stations, which will be introduced in section 3.1.3, this enabled the simulation to find results of real-world importance while reducing computational complexity.

3 Method

In a heterogeneous fleet of vehicles that is typically deployed by a car-sharing operator, having insights of the features determining the vehicle choice of a customer is key to operate in an efficient manner. This includes features like availability, travel-time and driving distance, but also socio-demographic features of the area in which the car-sharing system is deployed (Jian et al., 2017). During this study a classifier is developed which captures this relation and uses it to analyze an inter-area connected network of stations during a discrete event driven simulation phase to get insights on performance critical metrics.

3.1 Concepts

Firstly a set of concepts is defined, that will be important in the further discussion. The reader can also refer to the list of symbols for an overview of every symbol and a short description.

3.1.1 Vehicle Classes

Firstly a set of vehicle classes $\mathbb{C}$, to be the set of possible vehicle choices a customer can make, is defined, to represent a heterogeneous fleet. A discrete example of these vehicle classes can be found in the case study later on. These classes are also used to determine the pricing of a trip. An implementation of this framework is expected to define a pricing function:

$$\text{cost} : \mathbb{C} \times \mathbb{R} \rightarrow \mathbb{R}$$

where the input is the rented class and the time traveled in minutes and the output is the cost in local currency.

3.1.2 Area

An area is an approximation of different socio-demographic distributions with regard to distinct categories, such as age, income or marital status. Formally a category is defined by the random variable X which comes from an unknown distribution with value range K. For the framework, X is then captured in terms of probability of being part of specific subclasses R_K, where R_K is a partition of K, meaning that $\bigcup_{r \in R_K} r = K$ and $a \cap b = \emptyset \ \forall a, b \in R$. Leading to the specification of the distribution of X for our framework:

$$Z_X = \{(r, P(X \in r)) \mid r \in R\}$$

These values are sourced by empirical studies and are specified in the actual implementation of the proposed framework. Due to the fact that R is a partition of K the following invariant must hold true: $\sum_{z \in Z_X} z = 1$. An area is then defined in terms of the set of categories $\mathbb{X}$ that are analyzed in the implementation and can therefore be described as:

$$a = \{Z_X \mid \forall X \in \mathbb{X}\}$$

3.1.3 Station

The Station is an important concept for the simulation stage of the proposed framework. A station is always inside a particular area, namely there exists a function $\text{area}(s)$ which returns the area that the station is placed in. Additionally, a station has a state that is kept and updated throughout the simulation stage, this includes a function which for every time $t \in \mathbb{R}$ and class $c \in \mathbb{C}$ returns the delta of cars of class c rented up until that time: $\Delta_s : \mathbb{R} \times \mathbb{C} \to \mathbb{N}$. This function is bounded by the parameter C (Capacity), such that $\Delta_s(t, c) \geq -C \; \forall t \in \mathbb{R} \wedge c \in \mathbb{C}$. Representing the inability to satisfy customer requests after a maximum capacity is reached.

Leading to the spatial structure of the simulation environment, that is a fully connected graph where the nodes are the set of all stations in the simulation, called $\mathbb{S}$, and the edges of that graph called $\mathbb{E}$. An implementation is required to provide a distance mapping $L : \mathbb{E} \to \mathbb{R}$, providing weights for the edges of the graph.

3.1.4 Customer Request

Another crucial part is the customer request r, since that is the data that is used for the classification task and the simulation stage. Prior to deciding on the vehicle class a user has knowledge about the starting station $s_{r,0}$ and end station $s_{r,1}$ that he/she wants to travel between. The request is also expected to be sourced at station $s_{r,0}$ and therefore adheres to the socio-demographic factors defined by the $\text{area}(s_{r,0})$. A User request is always created inside a station graph that describes the simulation environment and can therefore be extended by the implicit attribute l_r (distance travelled), that is given by the travelled edge $e_r = (s_{r,0}, s_{r,1}) \in \mathbb{E}$ and the distance measure $l_r = L(e_r)$. The set of all possible customer request will be referred to as R.

3.2 Classification Model

Using those basic concepts a classification model can be defined whose main objective is to find a relation between a customer request, combining the socio-demographic factors of the area and the planned travel distance, and the decision that would be made. Additionally, the probability of each class should be determined, so that the confidence of a decision can be quantified. Therefore, a model is proposed whose objective is to approximate the function $P(c \mid r)$ where $c \in \mathbb{C}$ the vehicle class and $r \in R$ the request. Given this classifier, a prediction can be made which vehicle type is most likely to be chosen based on a particular customer request. The statistical evidence for this relation is given by Jian et al. (2017)

A typical classifier that would be used in this scenario is the Naive Bayes Classifier (NBC) that utilizes Bayes rule to rewrite the above-mentioned probability into $P(c \mid r) = \frac{P(c)P(r \mid c)}{P(r)}$ which can then be computed based on the strong assumption of conditional independence of the features. Although this assumption is often violated, as with our proposed framework, since the feature groups are technically not independent of one another, this type of classifier delivers competitive results. Other benefits of the NBC include a high computational efficiency, fairly low variance and most importantly for our setting a direct predication of posterior probabilities. (Webb, 2010)

3.3 Substitution effect

It is however noteworthy that the vehicle type decision is only partly driven by just the factors, that are provided by the customer request. Especially in free-floating car-sharing services other factors, like distance to the closest vehicle, can dominate the actual decision process even against personal preference. To capture this effect in a simplified form, a parameter $\alpha \in \mathbb{R}$ is introduced, called the substitution effect, that captures the willingness to decide against the most likely decision at random. Formally the set of all class decision D of a customer request $r \in R$ and the substitution effect α can then be defined as follows:

$$D(r, \alpha) = \{c \mid c \in \mathbb{C} \wedge P(c \mid r) \geq \max\{P(d \mid r) \mid d \in \mathbb{C}\} - \alpha\}$$

This set then includes every class that in regard to the value of α a user would consider renting.

3.4 Simulation

Understanding the dynamics of a complex system like a car-sharing system analytically is exceptionally difficult. One can however use a simulation to assess typical dynamics and get a quick insight on important performance metrics like ratio of satisfied customers, total distance driven or total profit. Additionally, the simulation environment should be easily extensible to enable further study on different network effects.

A discrete-event simulation is employed that aims to model a typical day of operation for the car-sharing network. A discrete event simulation is a simulation where the flow of time is not continuous but driven by events, such as time passed. It enables a computationally efficient way to model complex systems and is a good fit for the proposed simulation. A discrete event simulation consists of processes that can interact with other processes, access and modify shared resources and create and listen to events.

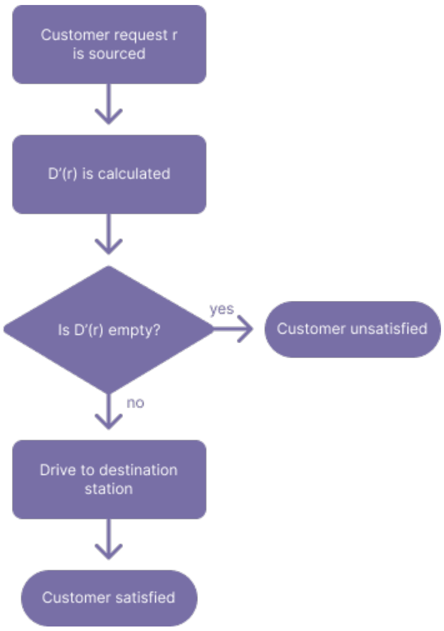

Figure 1: Simulation flow

The process that a typical customer request would flow through is given by Figure 1. The flow is started by the event that a customer request is sourced. This event happens periodically based on the average time period of demand p_h, determined beforehand by existing demand data for each hour of the day $h \in [0, 24) \subset \mathbb{N}$. The timestamp a customer request is registered with the system is called t_0 from now on.

The customer then needs to make a vehicle decision based on the request r and the current substitution effect α resulting in $D(r, \alpha)$. The framework then calculates a more restricted set of available possible vehicle choices $D'(r, \alpha) =$

$\{c \mid c \in D(r,\alpha) \wedge \Delta_{s_{r,0}}(t_0, c) > -C\}$ which includes the capacity restriction mentioned in Section 3.1.3. If D' is an empty set, the customer request is labeled "unsatisfied" and r is added to the set of unsatisfied requests $\mathbb{UR}$. Otherwise, a random unbiased choice between the classes in the set D' is made resulting in the chosen class c and the state of the start station is updated such that

$$\Delta_{s_{r,0}}(t_0 + \epsilon, c) = \Delta_{s_{r,0}}(t_0, c) - 1$$

Following this the customer starts its rental and travels to $s_{r,1}$. This takes him an amount of time δt given by $\delta t = l_r/\mathrm{AS}$ where AS (average speed) is the average speed of urban commute. This customer then reaches station $s_{r,1}$ at time $t_1 = t_0 + \delta t$ and increases the delta function of the target station similarly to

$$\Delta_{s_{r,1}}(t_1 + \epsilon, c) = \Delta_{s_{r,1}}(t_1, c) + 1$$

The rental is then counted as complete, added to the set of satisfied requests $\mathbb{SR}$ and the total profit π is increased by $\mathrm{cost}(c, \delta t)$, completing the customer request process. One run of the simulation framework simulates a whole day and includes the station network as a shared resource.

3.5 Metric & Performance

Resulting in performance metrics that can give measurable insights on the dynamics of the environment and allows to analyze the effects of the models parameters such as the substitution effect α and the capacity restriction C. The presented metrics aim to be of importance for operational decision-making and connect the results of the simulation stage to real world performance of a car-sharing network.

Metric	**Defintion**	**Description**
URR	$\frac{\lvert\mathbb{UR}\rvert}{\lvert\mathbb{UR}\rvert+\lvert\mathbb{SR}\rvert}$	The **U**nsatisfied **R**equest **R**atio, eg. the ratio of unsatisfied customers to all customers
π	–	The total profit of the simulation, calculated as part of the simulation stage
TD	$\sum_{r\in\mathbb{SR}} l_r$	The total distance driven during the day

4 Case Study Berlin SHARE NOW

The methods described in Section 3 are now applied to the specific case of SHARE NOW in Berlin during a period between October 2019 and March 2020. Firstly, SHARE NOW as a company will be described and then later the dataset and the insights it can deliver as well as the trained classifier and its application in the simulation environment that has been described above.

4.1 SHARE NOW

SHARE NOW is the worldwide leading free-floating car-sharing service. As a free-floating service the user can pick up an available car anywhere in a defined zone and finish the trip anywhere in that zone. It is currently available in 16 European cities with 11000 vehicles, with nearly 3000 electric vehicles. With about 3.4 million customers it has an unprecedented set of car-sharing users (*ShareNow About Us*, 2021). It was founded as part of a larger a joint venture of the BMW Group and the Daimler AG, which also includes services like PARK NOW, CHARGE NOW, REACH NOW and FREE NOW. Both firms brought in their existing car-sharing solutions, namely Car2Go a subsidiary of the Daimler AG and DriveNow a subsidiary of the BMW Group into the joint venture.

4.2 Data sources

This case study is primarily based on a dataset which includes 1.983.246 data-points, each representing a trip made through SHARE NOWs service. The data source was provided by the university chair and is not publicly available. Along other interesting fields each data point contains the vehicle model of the rental, the location where the rental was started as well as the distance of the trip. This data was then joined with publicly available socio-demographic data from Zensus 2011, a census commissioned by the statistical federal office (*Zensus2011*, 2020).

4.3 Fleet

SHARE NOW uses an extensive fleet of vehicles for its car-sharing service according to their website. The data for Berlin however indicates a slightly smaller set of available vehicles for that region. Throughout the period of data collection, a total of 3946 unique vehicles were tracked in the zone, with the following distribution.

As can be seen in Figure 2, the number of vehicles of each type is far from equally distributed. The dominant model in terms of number of vehicles is by a large margin the smart fortwo with 1007 unique vehicles, offering a high value

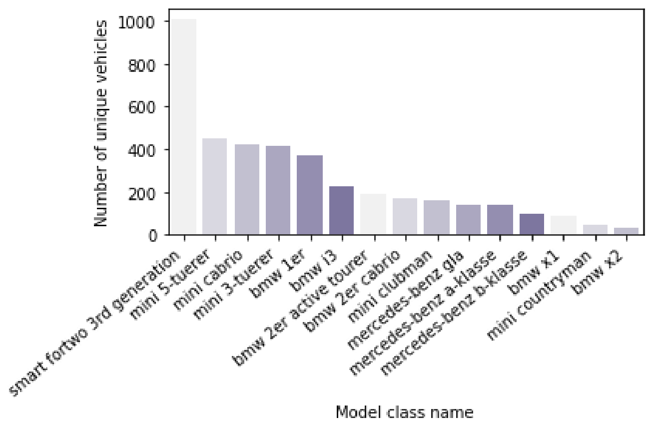

Figure 2: Number of unique vehicles by class

proposition in terms of parking space and ease of use for urban commuting. A clear tendency towards smaller cars, such as smarts or minis is visible in the distribution.

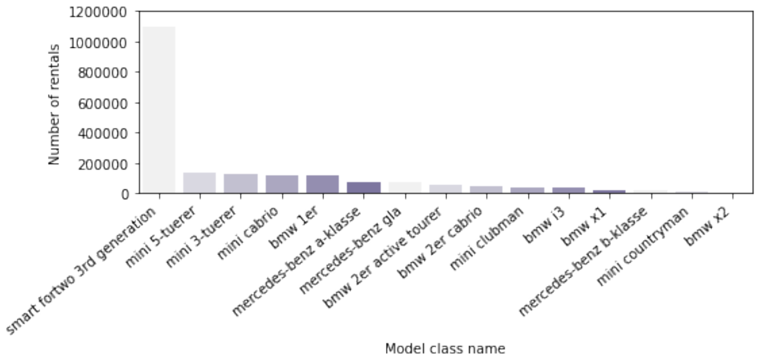

Figure 3: Number of rentals by class

The actual amount of rentals, as shown in Figure 3, has a very similar order to the fleet size with only a few exceptions. Notably though, the smart has seen even more demand than indicated by the number of vehicles.

The missing component however is the set of decidable classes. Similarly to the real world fleet deployment the set of models is structured in terms of vehicle classes, namely $\mathbb{C} = \{\mathrm{XS}, \mathrm{S}, \mathrm{M}, \mathrm{L}\}$. Then each model was assigned to a class, the mapping can be found in Table 1. These classes play an important role in the operational management of the car-sharing fleet since they differ in terms of costs and therefore profit. Additionally, the cost function used in the simulation stage of the framework is defined, based on the actual minute based pricing model of SHARE NOW. Although SHARE NOW also offers other tariffs, a focus was made on the on-demand minute based charge system for this implementation (*ShareNow Pricing*, 2021). The resulting function can be found in Table 2.

4.4 Data preparation

Before the actual classifier could be trained, the data had to be prepared to adhere to the form required by the framework. Using the census data an area description for each request was build where the set of categories to be analyzed includes the age distribution as well as the martial status. $\mathbb{X} = \{\text{Age}, \text{Marital Status}\}$. Then ranges for each were defined as can be seen in Table 3. Subsequently, rows with missing of malformed data were removed.

In early training iterations another important imbalance in the dataset was discovered. Classification models trained on the data had a tendency to favor the "XS" class since there were far more rentals done with vehicles from that class, since the set of features was missing an indicator for availability of vehicles. To counteract that the data for each class was shuffled and truncated in such a way that $\frac{nv(c_a)}{nv(c_b)} = \frac{nd(c_a)}{nd(c_b)}\ \forall c_a, c_b \in \mathbb{C}$, where $nv : \mathbb{C} \to \mathbb{N}$ is a function that returns the number of vehicles of a specific category and $nd : \mathbb{C} \to \mathbb{N}$ is a function that returns the amount of datapoints in the training set for each category. Reducing the impact of the availability of vehicles for the classification task.

4.5 Classifier

Resulting in a dataset that mapped the framework definition of the customer request onto the class that the customer choose. That data was then normalized into a zero to one range and split into a training and a testing set and a Complement Naive Bayes classifier (CNB) was fit using the training data. A Complement Naive Bayes Classifier is an adaption of a Multinomial Naive Bayes classifier, most prominently known from text classification tasks (Rennie, Shih, Teevan, & Karger, 2003). Additionally, the CNB typically has an advantage when dealing with imbalanced dataset such as ours.

4.6 Simulation

Before the actual implementation of the simulation stage, some data had to be prepared. Firstly the set of stations was to be defined, setting up the real world context of the simulation framework. Since the data was sourced in the city of Berlin, the simulation will also take place in that setting. Firstly the area data regarding the age and marital status where sourced for various districts in Berlin, as can be seen in Table 4. Then a subset of these were selected to contain stations of our proposed simulation framework. This then forms our simulation graph as can be seen in Figure 4.

Afterwards the distance function $L : \mathbb{E} \to \mathbb{R}$ was determined for each edge in

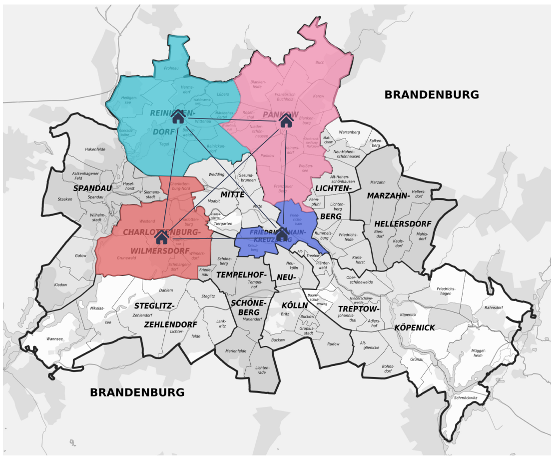

Figure 4: Simulation Graph

the proposed graph and can be found in Table 5. Using this the set of stations $\mathbb{S} = \{\text{Pankow}, \text{Reinickendorf}, \text{Kreuzberg}, \text{Charlottenburg}\}$ can be defined, where the area function is the identity function.

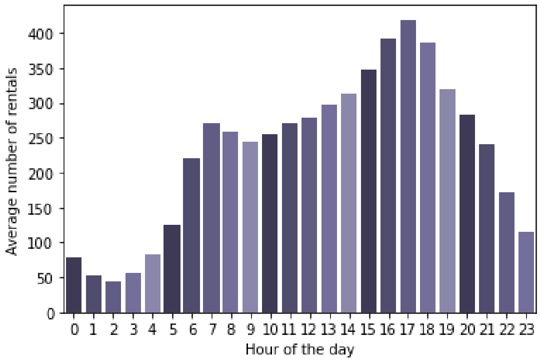

Figure 5: Average hourly demand of selected districts

Secondly, the simulation stage also requires the average time-period between customer arrivals for each working hour. This was computed based on the dataset, by extracting datapoints which happened in the simulation areas $\mathbb{S}$ and grouping into days and then taking the average for each hour. The resulting demand distribution can be seen in Figure 5. The time period $p_h \ \forall h \in [0, 34] \subset \mathbb{N}$ can then be determined by $p_h = \frac{1}{\text{demand}_h}$ where demand_h are the values from the figure. Additionally, the value of AS the average speed of urban commute was determined from the dataset to be $AS = 126.39 \frac{\text{m}}{\text{min}}$.

Using this information and the classifier described in Section 4.5, the simulation environment was build in Python 3.10 using the SimPy library. The simulation receives the parameters α, C (Substitution Effect, Capacity), the set of stations to be simulated, the hourly demand as well as the predictor function. That way different kind of classifiers could be trained to quickly assess the performance of different classification models.

5 Results & Discussion

Succeeding the implementation phase the results of the simulation framework had to be analyzed. Due to the complexity and high dimensionality of the problem, a focus was made on the effects of the parameters C and α on the metrics defined in section 3.5.

5.1 Simulation Result

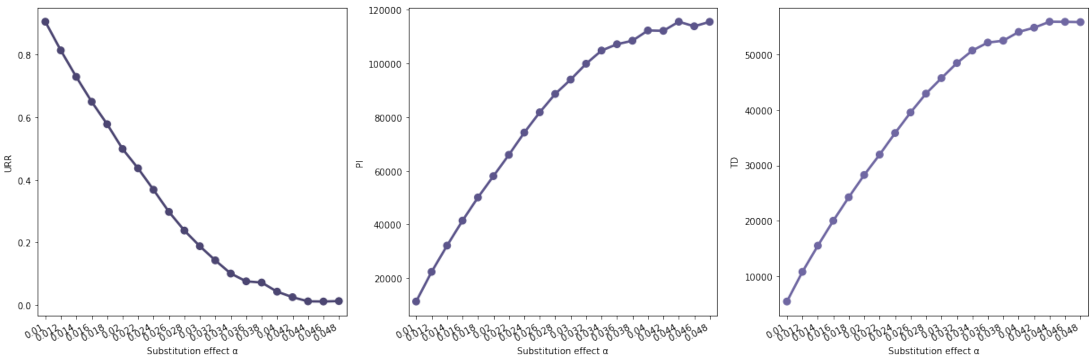

Figure 6: Effects of alpha for performance metrics

Starting with the effect of the substitution effect α. The Figure 6 shows the metrics defined in Section 3.5 in order. For this evaluation the capacity was set to $C = 5$, leading to a higher importance of the substitution effect. A clear dependency of the value of alpha to a higher performing car-sharing network can be seen. However, the curves also indicate that the impact of the parameter decreases with larger values and the random signal, that is due to the random nature of the simulation environment, increases.

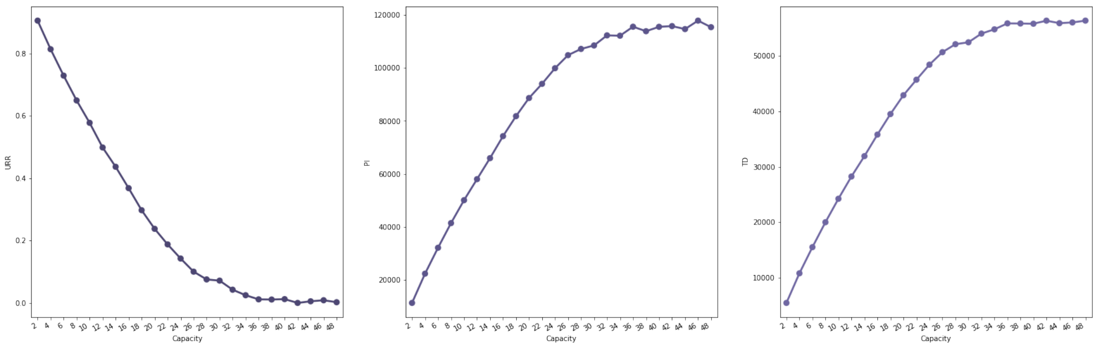

Figure 7: Effects of capacity for performance metrics

A very similar effect can be seen in Figure 7 with the parameter C, the capacity at each station. For these the alpha value has been set to $\alpha = 0.05$, the optimal value based on previous results. From an operational standpoint an optimal fleet capacity could be determined by averaging the metrics over multiple runs and define a threshold below which the increase in overall performance with regard to the increase in fleet size provides no added value to the network. This effect was also discovered by Mehdi Nourinejad (2014).

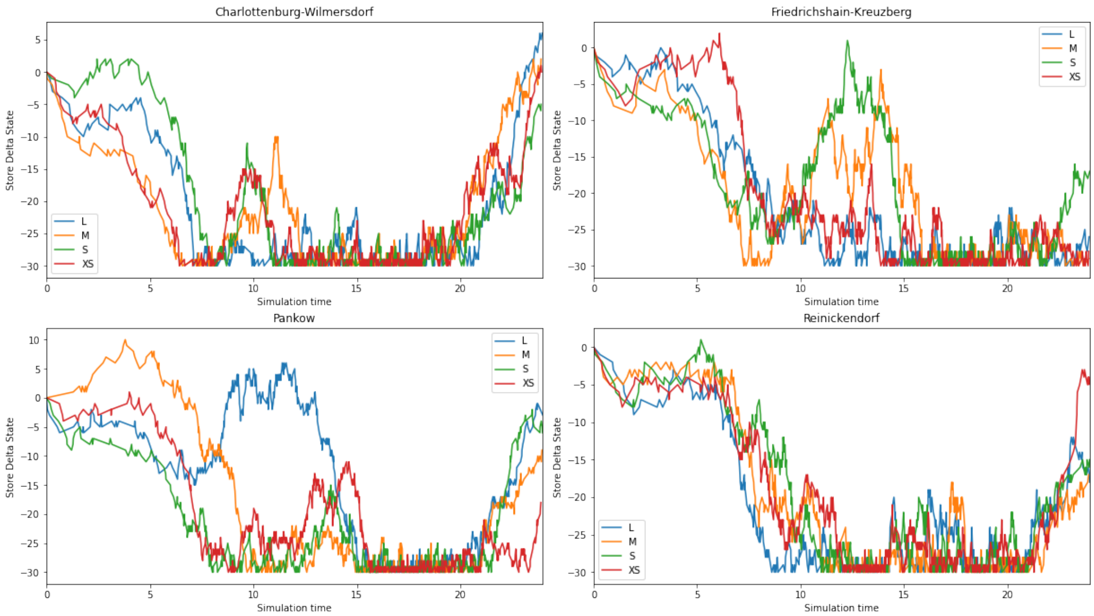

Figure 8: $\Delta_s(t, c)$ for a simulation with $\alpha = 0.003, C = 30$

Lastly the state function of the stations $\Delta_s(t, c)$ has been plotted against the simulation time, to get insights of the state at each station throughout the working day. As can be seen in Figure 8, the impact of the demand described in Figure 5 is clearly visible. Especially during the rush hour in the late afternoon all stations reach capacity limits, but just about manage to keep up with an unsatisfied customer rate (URR) of just 6.50%.

5.2 Discussion & Further Research

As mentioned previously the topic of simulation in a large scale network, like a car-sharing network, is a problem with high-dimensionality, therefore providing a lot of potential further research areas. The focus of this thesis also included to make the simulation environment as general as possible, such that additional research could be conducted on the same code base.

One aspect that could be of interest, is the selection of start and end stations for each customer request. In the real world this is typically not equally distributed but often follows traffic flows, which are among others time and location dependent. Another potential extension, could include an increased set of stations $\mathbb{S}$ to study even larger car-sharing networks.

While training the classifier, it became apparent that the signal of the socio-demographic data of the rental area only correlated loosely with the decision made, and feature importance was quite low. In further research a more sophisticated classifier could be trained that sources a dataset which directly correlates one rental to for example the age of the customer. Due to the modularity of the simulation that model could then easily be exchanged with the classifier trained in this thesis and used in the simulation environment.

Another aspect that would be of particular interest for the simulation stage is the parameter α. This parameter is not controllable by the operator but inherent for a particular customer group. One could convey a study to try to estimate the real world α for different deployments of a car-sharing networks empirically.

6 Conclusion

This research aimed to develop a discrete event simulation and integrate a classification model to get measurable insights into the mobility dynamics of a car-sharing system. By introducing a simulation framework and defining operationally important performance metrics, a clear positive dependency of overall performance to the value of the substitution effect and the fleet capacity has been identified. This has proven the ability to use simulation environments, sourced with real world data, to gain insights applicable to real world scenarios and guide operational decision-making. Therefore, an answer to the main research question proposed in Section 1, about the impact of the parameters α, the substitution effect, and C, the capacity, has been found, and a simulation model has been designed that can be used as a foundation to perform additional research in the field of large scale car-sharing systems.

As outlined in Section 5.2 the proposed framework has multiple aspects that could be improved upon in further research but these enhancements would exceed the scope of this thesis. Overall the method of discrete event simulation has shown to be a good fit to model this complex subject and provide a computationally efficient way of quantitative analysis of the environment. To increase the applicability of the model for real-world usage some assumptions, such as the equal spacial distribution of customer request, should be revisited.

The implementation of the study as well as the source for this thesis can be found on GitHub, to simplify further research on the simulation framework (Fanselau, 2022).

A Appendix

Class	**Vehicles**
XS	smart fortwo 3rd generation
S	Mini 3 door, Mini 5 door, Mini Clubman, Mini Convertible, Mini Countryman
M	BMW 2er Active Tourer, BMW 1er, BMW I3, Mercedes-Benz A-Class, Mercedes-Benz B-Class, Mercedes-Benz GLA
L	BMW 2er Cabrio, BMW X1, BMW X2

Table 1: Vehicle Classes in Case Study

Vehicle class $\mathbf{c}$	$\text{cost}(\mathbf{c}, t) =$
XS	$t * 0.09$
S	$t * 0.28$
M	$t * 0.31 + 0.99$
L	$t * 0.34 + 0.99$

Table 2: Cost function for vehicle classes

Range	**Value**
Z_{Age}	$\{[0, 18), [18, 30), [30, 50), [50, 65), [65, \infty)\}$
$Z_{\text{Martial status}}$	{ "Single", "Pairs", "Single Parents", "Parents with children", "Multiperson households" }

Table 3: Socio demographic category ranges

Start / **Destination**	Pankow	Reinickendorf	Kreuzberg	Charlottenburg
Pankow	0	9.08	9.38	14.18
Reinickendorf	9.08	0	12.39	8.46
Kreuzberg	9.38	12.39	0	10.05
Charlottenburg	14.18	8.46	10.05	0

Table 5: L function: Distance [in km] in Simulation

District	Age [Years]					Marital Status				
	0-17	18-29	30-49	50-64	Over 65	Single	Pairs	Single Parents	Parents with children	Multi person household
Charlottenburg-Wilmersdorf	0.12	0.14	0.30	0.23	0.20	0.56	0.20	0.07	0.14	0.03
Friedrichshain-Kreuzberg	0.14	0.23	0.40	0.14	0.09	0.54	0.15	0.07	0.13	0.11
Lichtenberg	0.13	0.23	0.32	0.16	0.17	0.49	0.24	0.09	0.14	0.04
Marzahn-Hellersdorf	0.05	0.07	0.34	0.27	0.27	0.25	0.64	0.02	0.09	0.00
Mitte	0.14	0.20	0.36	0.17	0.12	0.56	0.17	0.07	0.13	0.06
Neukölln	0.15	0.23	0.35	0.16	0.11	0.54	0.16	0.08	0.13	0.09
Pankow	0.15	0.18	0.46	0.11	0.10	0.55	0.17	0.08	0.14	0.06
Reinickendorf	0.15	0.15	0.29	0.19	0.22	0.49	0.21	0.13	0.12	0.05
Spandau	0.12	0.17	0.28	0.24	0.19	0.56	0.21	0.08	0.13	0.02
Steglitz-Zehlendorf	0.15	0.13	0.28	0.23	0.21	0.47	0.23	0.09	0.18	0.03
Tempelhof-Schöneberg	0.14	0.15	0.33	0.21	0.17	0.54	0.19	0.08	0.14	0.04
Treptow-Köpenick	0.16	0.13	0.31	0.12	0.28	0.46	0.21	0.10	0.20	0.03

Table 4: Districts in the Simulation

References

Burak Boyaci, N. G., Konstantinos G. Zografos. (2016). An integrated optimization-simulation framework for vehicle and personnel relocations of electric carsharing systems with reservations. *Elsevier Transportation Research Part B: Methodological*(95), 214-237. doi: 10.1016/j.trb.2016.10.007

Fanselau, L. (2022). *Github code repository.* `https://github.com/lucafanselau/bachelor-seminar-thesis`.

Ferrero, F., Perboli, G., Rosano, M., & Vesco, A. (2018). Car-sharing services: An annotated review. *Sustainable Cities and Society*, *37*, 501-518. Retrieved from `https://www.sciencedirect.com/science/article/pii/S221067071630573X` doi: https://doi.org/10.1016/j.scs.2017.09.020

Illgen, S., & Höck, M. (2019). Literature review of the vehicle relocation problem in one-way car sharing networks. *Transportation Research Part B: Methodological*, *120*, 193-204. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0191261518300547` doi: https://doi.org/10.1016/j.trb.2018.12.006

Jian, S., Rashidi, T. H., & Dixit, V. (2017). An analysis of carsharing vehicle choice and utilization patterns using multiple discrete-continuous extreme value (mdcev) models. *Transportation Research Part A: Policy and Practice*, *103*, 362-376. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0965856416309028` doi: https://doi.org/10.1016/j.tra.2017.06.012

Jiyeon Jung, Y. K. (2018). Analyzing the effects of car sharing services on the reduction of greenhouse gas (ghg) emissions. *Sustainability 2018*, *2*(10), 539;. doi: 10.3390/su10020539

Jochem, P., Frankenhauser, D., Ewald, L., Ensslen, A., & Fromm, H. (2020). Does free-floating carsharing reduce private vehicle ownership? the case of share now in european cities. *Transportation Research Part A: Policy and Practice*, *141*, 373-395. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0965856420307291` doi: https://doi.org/10.1016/j.tra.2020.09.016

Jörg Firnkorn, M. M. (2011). What will be the environmental effects of new free-floating car-sharing systems? the case of car2go in ulm. *Elsevier Ecological Economics*(70), 1519-1528.

Mehdi Nourinejad, M. J. R. (2014). A dynamic carsharing decision support system. *Elsevier Transportation Research Part E*, *66*(66), 36-50. doi: 10.1016/j.tre.2014.03.003

Perboli, G., Ferrero, F., Musso, S., & Vesco, A. (2017, 09). Business models and tariff simulation in car-sharing services. *Transportation Research Part A: Policy and Practice*, *115*. doi: 10.1016/j.tra.2017.09.011

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th international conference on machine learning (icml-03)* (pp. 616–623).

Sharenow about us. (2021). `https://www.share-now.com/de/de/about-us/`.

Sharenow pricing. (2021). `https://www.share-now.com/de/de/pricing/`.

Webb, G. I. (2010). Naïve bayes. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 713–714). Boston, MA: Springer US. Retrieved from `https://doi.org/10.1007/978-0-387-30164-8_576` doi: 10.1007/978-0-387-30164-8_576

Zensus2011. (2020). `https://www.zensus2011.de/DE/Home/home_node.html`.