Data-based application scenarios for e-scooters

Jan Fauser, Natascha Sigle, Dieter Hertweck

Reutlingen University, Faculty of Computer Science and Reutlingen Research Institute

KEY WORDS: E-sccoter sharing, Data Mining, Clustering, E-Mobility, Free-floating

ABSTRACT:

In various German cities free-floating e-scooter sharing is an upcoming trend in e-mobility. Trends such as climate change, urbanization, demographic change, amongst others are arising and forces the society to develop new mobility solutions. Contrasting the more scientifically explored car sharing, the usage patterns and behaviors of e-scooter sharing customers still need to be analyzed. This presumably enables a better addressing of customers as well as adaptions of the business model to increase scooter utilization and therefore the profit of the e-scooter providers. The customer journey is digitally traceable from registration to scooter reservation and the ride itself. These data enable to identifies customer needs and motivations. We analyzed a dataset from 2017 to 2019 of an e-scooter sharing provider operating in a big German city. Based on the datasets we propose a customer clustering that identifies three different customer segments, enabling to draw multiple conclusions for the business development and improving the problem-solution fit of the e-scooter sharing model.

1. INTRODUCTION

In our cities nowadays, trends such as climate change, urbanization, demographic change, amongst others are arising. According to the United Nations (UN) 68% of the world's population will live in cities by the year 2050 (Nations, 2018). New mobility solutions can be a part to relief that problem. Vehicle free-floating sharing concepts are increasing around the world and in Germany, providing on-demand mobility for the customers while tackling the problem of crowded streets and lacking parking spots in metropolitan areas. The main idea of these business models is the simplicity to share individual vehicles with other customers to save costs and increase flexibility. In addition, fewer vehicles are needed overall and environmentally harmful CO_2 can be avoided through lower production of e-scooters. Free-floating e-scooter sharing is a recent variation of this idea and can already be found in multiple big German cities ((EMMY, no date), (*Stella Stadtwerke Stuttgart*, no date), ('MVV', no date), (Düsseldorf, no date)) It combines the advantages of vehicle sharing with those of e-mobility and is supposed to reduce air and noise pollution in urban environments.

In this paper, we build on our study (Degele *et al.*, 2018) and analyze the data from 2017 - 2019, drawing new conclusions for the e-scooter sharing provider. We analyzed a free-floating e-scooter sharing services which is characterized by providing their service in a mobile app and offering the flexibility to pick up and leave scooters in any publicly accessible location in the respective business area. for business development. For our research we received trip and anonymized customer information from a German e-scooter sharing provider. The main objective was how can the value for customers of the service be increased. Therefore, we investigated the customer profiles and used data mining to cluster to build customer segmentation. This enables the service provider to target a specific customer segment and develop new features to improve the value of the service for the specific customer segment. We followed the cross industry standard process for data mining model (CRISP-DM Model) (Chapman *et al.*, 2000; IBM, 2012) the following research question has been formulated: How can customers of mobility (e-scooter) service be segmented?

2. RELATED WORK

Related literature analyzing the usage of sharing models mainly concentrates on geographic factors with few exceptions. Schmöller et al. investigate correlations between the spatial distribution of car sharing bookings and the respective city structure (Schmöller *et al.*, 2015). Several works examine relocation strategies for bike and car sharing systems aiming to attenuate imbalances in the vehicle distribution (Vogel, Greiser and Mattfeld, 2011), (Weikl and Bogenberger, 2013) ,(Reiss and Bogenberger, 2017).

The academic research on the market of e-scooters as well as e-scooter sharing in general is quite rare (Hardt and Bogenberger, 2019). Bogenberger and Hardt (2019) analyze the state-of-the-art literature on e-scooters. They find little research on the topic in general, with some expectations on state regulations and market research. The authors analyzed a pilot project in Munich, where e-scooter usage was studied. Their investigation however focused on the usage of privately owned e-scooters (Hardt & Bogenberger, 2019). When discussing e-scooter sharing, there is even fewer literature to be found. The available literature concentrates mostly on car sharing, with some approach to research bike sharing as well. Aguilera-Garcia et. al (2020) point out, that e-scooters are even less academically addressed than bike sharing (Zagorskas and Burinskienė, 2019; Aguilera-García, Gomez and Sobrino, 2020). If e-scooter
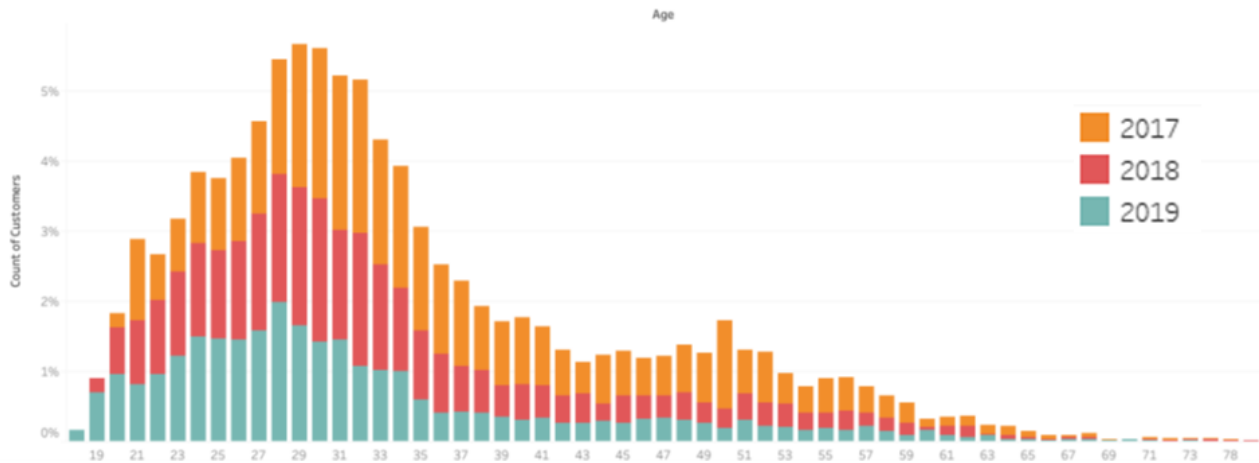
Figure 1. Age structure of customers

sharing is attended to in the academic field, it is mostly placed into a global perspective and the side-mention of the outline of other sharing options such as car or bike sharing. The usage and customer profiling of e-scooters is a rather poorly researched field, with hardly any scientific user description (Aguilera-García et al., 2020; Zagorskas & Burinskienė, 2019). We will focus on the customer itself, rather than on vehicles.

Greiser, Mattfeld and Vogel (Vogel, Greiser and Mattfeld, 2011) who are researching spatiotemporal dependencies of activity patterns at bike sharing stations, acknowledge the importance of a customer-centric analysis and designate the development of customer profiles for future work. Similarly, to our intended approach the paper "Topology of carsharing members" (Morency, Trepanier and Agard, 2011) categorizes car sharing customers. The authors make use of frequency of rides, the most common weekday and the distance travelled for their clustering. Aguilera-Garcia et. al (2020) conducted a customer-centric research who identified key drivers for the adoption of usage in general and the frequency of usage of e-scooters for Spanish citizens. The study suggests that the factors were most significant for customer differentiation and thus segmentation were sociodemographic and travel-related variables (Aguilera-García et al., 2020). This work recognizes the importance of a customer-centric analysis and thus focuses on the development of customer profiles for e-scooter users.

3. METHODOLOGY

For the customer segmentation we followed the cross industry standard process for data mining model (CRISP-DM Model). CRISP-DM is a general-purpose methodology which is domain independent, technology neutral and according to (Kdnuggets, 2014) the most referenced and used DM methodology in practice. The lifecycle of the Model is split into six phases: Phase 1: Business Understanding, Phase 2: Data Understanding, Phase 3: Data Preparation, Phase 4: Modelling, Phase 5: Evaluation and Phase 6: Deployment. The first phase engages with the "Business Understanding" of the problem or question. The generated business objectives and requirements are transformed into a data mining problem that can be undertaken in the subsequent phases. The second phase is concerned with the initial "Data Understanding", which is closely connected to the first phase of business understanding. It also involves the recovery of possible data quality problems, and new interesting leads for additional hypotheses. The third phase of "Data

Preparation" includes the transformation of the raw data into the final dataset. To achieve this data may be cleaned, poor quality data may be removed, and new variables and attributes may be constructed. In the Modelling phase, several modelling techniques are selected and applied, and their parameters optimized. The fifth phase "Evaluation" involves the analyzing of the results and needed to be verified with the business objectives and requirements of the first phase. In the last phase "Deployment", it's about visualization and presentation of the knowledge and obtained it that has been derived from the previous phases.

4. BUSINESS UNDERSTANDING

Following the CRISP-Methodology the first process activity is Business Understanding. We analyzed the business models for e-scooters operated by a municipal utility in a major city in Germany. We used three data collection methods for case study research to increase construct validity and reliability (Yin, 2009): semi-structured interviews, participating observation and document study. The utility's focus of business is the provision of electricity, gas, and heat. Additionally, they produce regenerative energy. The energy turnaround in Germany and the possibility of easily switching electricity or gas suppliers via internet comparison portals are having a negative impact on the profits of energy suppliers. Therefore, the analyzed utility in our case study expanded their business area and introduced a e-scooters sharing system. For implementation of the business model, they used a white label solution, that is also active in other cities under different names. They started in 2016 with 15 e-scooters around the city, but quickly realized that the size of the fleet had to be increased. In 2020 they operated their e-scooter sharing service with 200 scooters. If the battery of an e-scooter is fully charged, the maximum reach is 100 km. The e-scooters can be picked up and dropped off anywhere within the geographical region and the legal requirements of the city. The pricing model of e-scooter sharing has changed over the years. In 2017 and 2018 the price of the service was calculated either for the price per minute or per kilometer (km), whichever was cheaper. Customers took advantage of this functionality and parked the scooter but did not switch to the parking mode. As a result, parking was free and the price per km would be charged at the end. For the season of 2019 the e-scooter sharing provider switch to a time-based billing. Additionally, in 2019 customers of the utility pay 0,19€/min, others must pay 0,24ct/min. It was intended that this campaign would lead to cross-selling effects
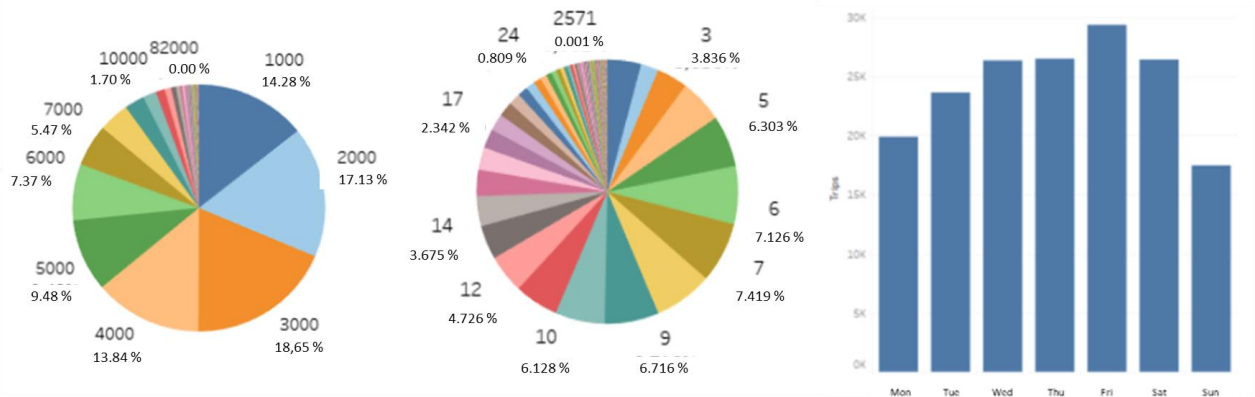
Figure 2: Driven distance in percent, Usage time in percent and Rentals per weekday

and make customers that use the e-scooters service also became customers of the municipal utilities in the segment of electricity, gas or district heating.

5. DATA UNDERSTANDING

In the following section we investigated the second phase of the CRISP-DM Model: data understanding. After getting an understanding of the business, the underlying data needs to be investigated and connected to the business objectives. The business objective, to increase the customers' value by establishing a targeted marketing strategy and business development plan, has underlying data as essentials. Data mining goals (DMG) were implemented to understand and analysis the data logically, with the final aim of identifying and segmenting user groups by clustering. In preparation of the modelling (section 6) these objectives were essential to define. The following data mining goals were identified in collaboration with the e-scooter sharing provider: 1. DMG: Understand the customers profiles of e-scooter-sharing users. 2. DMG: Understand how e-scooters are used. 3. DMG: Identify and segment user groups. The first and second data mining goal were concerned with the underlying customer profiles, the usage of the e-scooters as well as the geospatial distribution. These analyses give a vast overview of the behavior of e-scooter users. This delivers the basis for the clustering and supports the identification of clustering variables, which is a crucial measure for the development of clustering.

6. DATA PREPARATION

In the Data Preparation phase, we adapted the data workflow described by (Sarstedt and Mooi, 2014). This ensures the data quality by reducing mistakes and furthermore enabling replicability. Our implemented data workflow covers entering, cleaning, describing and transforming the data. The data was provided by e-scooter sharing provider in form of two Excel spreadsheets. The first data set includes information about the customers for example gender and date of birth. The second dataset contains information about reservations or trips. The software KNIME[1] proved to be suitable for the intended preparation and analysis steps. We analyzed the time periods of

[1]https://www.knime.com/

2017, 2018 and 2019. Data generated in 2016 and 2020 was removed, as no complete data set for those years existed. Each year represents a completed season that can be compared with each other. In the preparation process new variables were created and respecified. Two types of transformation of data are distinguished: variable respecification and scale transformation (changing variables values for comparability with other variables). Variable respecification includes the computation of additional fields that are needed for the analyses like 'Age' was created as a new variable based on the birthdates of the customers, 'Time between rides', meaning the time passing until a customer uses an e-scooter again and the exact week-day the trip was taken. Scale transformation in form of normalization is necessary for the clustering and is described later.

7. MODELLING

In this phase, we applied several modelling techniques to answer the different data mining goals. This work follows a customer centric approach. For the first data mining goal we analyzed the customer profiles and specially the age structure and distribution by gender. Most of the customers that are registered are in their late 20s/early 30s (see Figure 1). There is an additional smaller peak with customers in their late 40s/ beginning of 50s. There has been a constant increase of new customers over the years. The distribution of new customers registration has almost remained identical for all years. 76.01 % of the customers were male and 23.99 % were female. There were 3117 customers registered for the service but never rented an e- scooter. Whereas the gender structure is quite similar, there is slight difference in the age structure between the two groups. The customers who have never used an e-scooter were in their late 40s / early 50s.

The second data mining goal examines how sharing e-scooters are used. The number of trips performed increased each year and is shown in Table 1. This correlate with the increased number of customers, e-scooters and the awareness level of citizens. From 2017 to 2018, both the number of trips and the number of customers increased. The ratio increases accordingly from 6.9 to 7.3. Interestingly, the trips from 2018 to 2019 only increased very little in relation to the number of customers, which increased by half. This means that the ratio of 5.9 is even lower than in 2017. The number of trips increases in the warmer months. In 2019, for the first time a year-round service without winter break was implemented. Another aspect we investigated was the distance travelled per trip and the usage time per customer. In figure 2 we illustrated the driven distance and

usage time. In 50% of the trips, distances less than 3 km were driven. 73.02 % of all trips taken were shorter than 5 km. There is a trend over the years, with the tendency that users drive slightly shorter distances. The usage time, which represents the actual time driven (excluding parking time), almost 50 percent were under 9 minutes, 75 percent under 14 minutes and 99 percent of trips have been under 60 minutes. Over 80 percent of the trips are taken by male users. So, the question arises, how does gender and age influence the usage behavior of rentals of e-scooters? We did not detect any significant difference between male and female users in average distance driven and usage time. Users above the age of 55 showed a great variance in the average driven distance. In a next step we investigated the distribution of trips during the day. Most e-scooters were rented between 6 and 7 pm in 2017. In 2018 and 2019 the peak was during 4 and 5 pm. Since the highest peak in the daytime were clearly in the afternoon for all years, we assume that e-scooter sharing is mainly used for leisure time. In figure 2 we aggregated data for all years indicate that most trips are done during or close to the weekend. Over the periods of 2017-2019, the most frequented weekday was Friday. If we look at the time periods separately, the results demonstrate, that the most frequented period of the week shifted even more to the weekend over the years. In 2017, Wednesday and Friday were the peak days. Compared to 2018 were the trips on Wednesday reduced. An even more significant shift we identified in 2019 towards the end of the week from Thursday to Saturday, with Friday being the peak. We also analyzed differences in rental behavior caused by geographic and infrastructural features. The highest frequented areas were in the city center. A high usage could be identified around points of interests., public transportation or other important infrastructure points. This supports the assumption that the e-scooters are used for leisure time activities.

Year	Trips	Customers	Trips/Customer
2017	34.812	5.043	6.9
2018	61.479	8.475	7.3
2019	73.071	12.344	5.9

Table 1. Ratio trips to customers in years 2017-2019

In our third data mining goal, we used a clustering approach for the segmentation of user groups. Both customer segmentation and clustering aim to create separated groups from a set of objects, such that the in-group variance is minimal and the variance between groups is maximal (Sarin, 2010). The clustering-based customer segmentation approach based on (Sarstedt and Mooi, 2014) requires selecting a suitable clustering algorithm, preparing the data and visually investigating the data in order to determine the clustering parameters. To define differentiated customer clusters, the choice of clustering variables is of utmost importance. For the identification of cluster variables, we followed the approach of (Kotler, P., Wong, V., Saunders, J., & Armstrong, 2011). According to (Kotler, P., Wong, V., Saunders, J., & Armstrong, 2011), there are four major segmentation variables for consumer markets: geographic, demographic, psychographic and behavioral. Based on the available data, we analyzed which variables are best suited for a customer segmentation in the e-scooter sharing business using clustering algorithms. The variables listed in Table 2 are ready for use after the data preparation and were considered in the cluster analysis. Psychological data could not be derived from the given dataset and is therefore not available for this analysis. We followed the guidelines of (Sarstedt and Mooi, 2014) for validation of the variables.

Segmentation categories	Variables dataset customer	Variables ride dataset	Constructed variables
Demographic	Age (in years)	N/A	N/A
Geographic	N/A	Start longitude; Start latitude; Start zip code	End longitude, End latitude, End zip code
Behavioural	N/A	Usage time (in min), Distance driven (in m), Park time (in min), Cost per ride (in Euro), Free minutes used, Cost per ride (in Euro)	Time between rides (in h), Day of week, Total revenue per customer (in Euro)
Psycho-graphic	N/A	N/A	N/A

Table 2. Variables after data preparation

The first two questions that need to be answered were: Are the variables sufficient to differentiate between the segments? Are the variables adequate to form unique clusters? Our main goal is to segment user groups, understand their background and tasks and improve their value for the e-scooter sharing provider. To identify differences between the customers only a few characteristics are useful. To segment the customers by their value, it is less important to know when they are driving, but rather how frequent they are driving. Therefore, the time between rides is used as a clustering variable. The way of addressing a customer may vary depending on his age, hence the age was included as a variable for clustering. The driven distance and usage time helped us to understand the customer backgrounds and their tasks. Additionally, as e-scooter sharing provider is a profit-oriented company, so revenue is an important segmentation value. With the five clustering variables time between rides, age, driven distance, usage time and revenue per customer, we cover both behavioral and demographic segmentation categories. In the next step, we investigated the question: Is there a strong correlation between the clustering variables? In case two or more clustering variables have a strong correlation, they distort the clustering as they have a higher emphasis than other factors. A correlation is considered strong if the correlation coefficient is larger than 0.9. This work uses Spearman's rho P (rho) correlation for the calculation of the correlation between the variables. This correlation coefficient for variables is measured on an ordinal scale and reflects the intensity of the monotone relationship between two variables. As the correlation matrix in Table 3 shows a relative high correlation between distance driven and usage time, so we exclude usage time form our analysis. In the third step we analyzed the question: Is the ratio of clustering variables to sample size appropriate? A higher number of clustering variables, results in less meaningful clusters and they become difficult to interpret. The formula $n \geq 2^m$, where n equals the sample size and m equals the number of variables, limits the number of cluster variables depending on the size of the dataset. For the chosen four variables, a dataset of 16 records would be sufficient, which is far exceeded by the residual dataset. In the

	Age	Gender	Distance driven	Usage Time	Start zip code	Free minutes used	Park time	Time between rides	Day of week	Cost
Age	1	0.13	0.03	0.04	-0.03	0.34	0.08	0.09	0.01	0
Gender	0.13	1	-0.02	-0.04	-0.01	-0.02	-0.07	0	0	-0.01
Distance driven	0.03	-0.02	1	0.88	0.15	0.05	0.30	0.07	0.03	0.53
Usage time	0.04	-0.04	0.88	1	0.11	0.08	0.32	0.10	0.03	0.56
Start zip code	-0.03	-0.01	0.15	0.11	1	0.02	0.05	0.01	0.01	0.06
Free minutes used	0.03	-0.02	0.05	0.08	0.02	1	0.16	0.23	0.02	-0.43
Park time	0.08	-0.07	0.30	0.32	0.05	0.16	1	0.17	0.04	0.19
Time between rides	0.09	0	0.07	0.10	0.01	0.23	0.17	1	0.01	-0.07
Day of week	0.01	0	0.03	0.03	0.01	0.03	0.04	0.01	1	0.01
Cost	0	-0.01	0.53	0.56	0.06	-0.43	0.19	-0.07	0.01	1

Table3. Spearman's rho correlation matrix

last step we answered the question: Is the underlying data basis of a high quality? The CRISP-DM model and its phases that were executed with the quality of data in mind and throw the data preparation process introduced in section 5.

To define the number of clusters, we used hierarchical clustering approach. Hierarchical clustering has the advantage of not having to determine a number of clusters in advance. It rather provides guidance for choosing a reasonable number of clusters for other clustering algorithms. However, hierarchical clustering is not suitable for large datasets due to performance constraints. This leads to the decision to perform a hierarchical clustering with a small randomly picked subset of the data in order to decide on the number of clusters used as input for the following partitioning clustering with the whole dataset. We used a dendrogram to determine the quantity of clusters applying a hierarchical clustering algorithm with Euclidian distance measures and average linkage. Due to set size restrictions of the expensive hierarchical clustering, we randomly drew a set of 100 customers from the filtered dataset. We identified three clusters and selected these for further analysis.

Before conducting the clustering, we grouped the ride data by the customers' unique identifier. The chosen aggregation methods display both the riding behavior on average and in total as well as customer demographics. These include age, most frequent day of use, total created revenue, total number of rides, average driven distance and the average time between rides. Another essential step for the preparation of clustering is the normalization of the cluster variables to not distort the results. Therefore, we used a normalization, which converts all values into double values between 0 and 1. After the clustering process, the values were denormalized. The input data for the clustering process were normalized customer data and the clustering variables age, time between rides, average ride distance and revenue per customer. We used the k-means algorithm for the clustering. Table 4 describes the characteristics for each cluster in more detail. Cluster 0 and cluster 2 consist of smaller share of customers, compared to cluster 1. In cluster 0, the average age is 50, whereas in clusters 1 and 2, the younger generation is represented. Furthermore, customers in cluster 0 were at least 38 years old. In compression with cluster 1 the customers were not older than 40 years old. In cluster 1 the average driven distance was shorter compared to the other clusters. Cluster 1 generates 77,90% of the revenue, even in relation with the customer share. Cluster 2 has significantly high times between trips and an in average two rents of an e-scooter. Additionally, in cluster 2 there were more female users compared to clusters 1 and 2. The most common usage day in each cluster was Saturday. All clusters show the highest use of scooters around 4 and 5 pm. This indicates that the e-scooters were mainly used for leisure time activities.

There were no mentionable differences between the geographical distribution of usage of the e-scooters between the cluster. The concentration of the trips is located in the city center. A more detailed analysis showed that the high usage areas were close to point of interests as well as infrastructure such as public transportation.

Cluster characteristics	Cluster 0		Cluster 1		Cluster 2	
Share of customers	21.9 %		62 %		16.1 %	
Gender structure	F 19.28 %	M 80.72%	F 23.24 %	M 76.76 %	F 30.97 %	M 69.03 %
Rentals per customer	16		23		2	
Share of trips	19.40 %		79.23 %		1.37 %	
Average age	50		29		35	
Average time between trips	1046 h (43.6 days)		713 h (29.7 days)		3484 h (145.2 days)	
Average distance	6.0 km		4.9 km		7.4 km	
Share of revenue	21.00%		77.90%		1.10 %	
Most common usage day	Saturday		Saturday		Saturday	

Table 4. Cluster characteristics (F=Female; M=Male)

The service of the e-scooter sharing provider has evolved over the years for instance in strategy, in fleet size and in number of customers. Table 5 represent the results of the different clusters for each period. The number of customers in cluster 2 were increasing over the years. In cluster 0 the users were decreasing. Rentals per customers and the total share of trips increased in cluster 1. Time between trips increase in 2019 in all clusters. This might be due to the fact, that in 2019 there was no winter break. The average distance driven is decreasing for all clusters. The most common usage day change for cluster 0 and 2 form Saturday to Friday in 2019.

8. EVALUATION

There are a few general suggestions that can improve business development for the e-scooters sharing provider. Firstly, the availability of e-scooters in hot spot areas must be ensured. This raises awareness and offers the convenience for the usage of the e-scooters. To increase the awareness of the brand, further cooperation with private and public partners should be aspired. We identified that in areas with a high concentration of leisure time activities offers, the e-scooters were used the most. A high availability of e-scooters in these areas could increase the revenue. To identify more point of interests it could be tracked when and where customers open their app to search for an e-scooter. A fun aspect for customers to get more involved could

Year	2017			2018			2019		
Cluster characteristics	Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2
Share of customers	25.10 %	61.20 %	13.70 %	20.80 %	62.33 %	16.87 %	20.14 %	61.07 %	18.79 %
Rentals per customer	9	12	1	10	12	1	10	15	1
Share of trips	23.48 %	74.88 %	1.65 %	18.45 %	79.68 %	1.87 %	17.26 %	80.73 %	2.01 %
Average age	51	31	36	49	30	36	49	29	34
Average time between trips	835 h (34.8 days)	517 h (23.8 days)	3086 h (128.7 days)	878 h (36.6 days)	646 h (26.9 days)	3557 h (148.2 days)	1178 h (49.1 days)	804 h (33.5 days)	4099 h (170.8 days)
Average distance	6.0 km	5.1 km	7.4 km	6.1 km	5.2 km	7.8 km	5.4 km	4.5 km	5.9 km
Share of revenue	28.13 %	71.03 %	0.84 %	19.54 %	78.77 %	1.69 %	18.72 %	79.02 %	2.26 %
Most common usage day	Saturday	Saturday	Saturday	Saturday	Saturday	Saturday	Friday	Friday	Friday

Table 5. Cluster compression

be a gamification approach to encourage higher usage. Additionally, the pricing model should focus on higher usage of rentals rather than a high registration fee of 19 €, which might be an obstacle for interesting actors. Moreover, as the customers on average drive shorter distances, the pricing model should also reflect these to increase revenue. A cheaper registration fee could be more convenient for tourists to rent an e-scooter.

After building the cluster and customer segmentation each segment of customers can now be individually targeted to optimize the value for customers. developed a 4 P Marketing Mix. The four Ps include Price, Product, Place, and Promotion Through these four Ps each individual customer segment can be addressed according to their needs and interests. The product is e-scooter rental, which can be offered with different interesting features and approaches, varying pricing models, as well as unique promotions. The place for the free-floating system is not suitable for the targeting strategy since the area of sharing service is fixed and accessible for all customers. These areas can be matched to the specific characteristics of the different clusters and a targeted strategy can be derived from it. The characteristics of the clusters were summarized as follows:

Cluster 0 is 50 years old on average and the share of revenue is almost similar to the share of customers. The pricing model might not have the triggering affect if it is only reduced. Although the distances are generally becoming shorter, the desire to travel longer distances is still likely to exist based on the data available. A flat rate might be more interesting for this customer segment including the usage of public transportation. As a fun gadget could be interesting to include information of the weather status and traffic. The best marketing channel would be e-mail or a push notification on the smartphone app.

Cluster 1 includes customers with an average age of 29 years old. This younger generation use the e-scooter mostly for leisure time activities. The share of revenue is higher than the share of customers. To be more involved with the brand and to increase the usage of e-scooters, a gamification approach might trigger this customer's segment. They will get more influenced by special price offering, since they are younger and either students or rather freshly employed. These pricing models could offer different prices on weekends since the leisure time activities rise during the weekends. The channel for reaching this customer segment is mainly through social media, such as twitter, Instagram, or even Facebook. But also, via push notification in the smartphone app.

The third customer segment, cluster 2 is probably the most difficult to reach. A survey especially for those customers could assist with this issue. The average age were 35 years. What is the obstacle to use e-scooters, and how could they be encouraged to increase their usage? It might be important to offer this customer segment a clear process of renting an e-scooter.

Overall, for all customers groups the higher goal should always be to keep the active customers or increase the activities of existing customers in general. For this reason, it is crucial to conduct a survey to ask the customers about their customers jobs to understand how and why they use the rental service. Only then it can be certainly answered how the usage of e-scooters can be increased. Additionally, it is important to review the clusters and the clustering regularly to detect changes in the constellation of clusters as well as the characteristics of the customers of each segment. It is possible, that the number of the clusters needs to be adapted. It is, however, essential to use the same cleaning process of the data to be able to compare the results and build on previous knowledge gained from the data.

9. CONCLUSION

Answering the research question of how customers of mobility (e-scooter) services can be segmented, this work has provided a clustering process with data of a sharing provider in a large German city from 2017 to 2019. We identified three customer segments via clustering and developed individual marketing strategies for each segment. For the clustering we defined the variables age, time between rides, driven distance and revenue per customer. The customers in cluster 1 were responsible for 78% of the revenue and the average age were 29. Following the categories of the 4P Marketing Mix model, multiple ways to address these customers and foster customer loyalty were identified. There are certain aspects that can improve this model. In order to understand the tasks of the customers perfectly, and to include the psychographic information of the customers, surveys could be effective. The knowledge gathered from the surveys could include information about wealth and well-being, occupation and education, pains, gains and customer jobs. This evidence could give a better insight of how the value for customers can be even further increased, as well as how the loyalty of customers can be intensified. A good starting point

could be the low use of scooters by women. What are the reasons why women don't use the e-scooters? Perhaps due to insufficient storage space, too much fear of an accident, problems with the helmets due to styling or hygiene? Additionally, a geographical evaluation could assist to analyze the demand of the rentals. If the demand of the rentals is tracked in real time, e-scooter-sharing could also react in real time to support the demand with the sufficient supply and adapt the relocation strategy of the e-scooters to those numbers.

ACKNOWLEDGEMENTS

REFERENCES

Aguilera-García, Á., Gomez, J. and Sobrino, N. (2020) 'Exploring the adoption of moped scooter-sharing systems in Spanish urban areas', *Cities*, 96, p. 102424. doi: 10.1016/j.cities.2019.102424.

Chapman, P. *et al.* (2000) 'CRISP-DM 1.0'. Available at: http://ieeexplore.ieee.org/document/4579988/.

Degele, J. *et al.* (2018) 'Identifying E-Scooter Sharing Customer Segments Using Clustering', in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, pp. 1–8. doi: 10.1109/ICE.2018.8436288.

Düsseldorf, S. (no date) *Stadtwerke Düsseldorf AG: Eddy E-Roller Sharing*. Available at: https://www.swd-ag.de/mobilitaet/elektromobilitaet/e-roller-mieten/.

EMMY (no date) *Emmy (Berlin, Hamburg)*. Available at: https://emmy-sharing.de/.

Hardt, C. and Bogenberger, K. (2019) 'Usage of e-Scooters in Urban Environments', *Transportation Research Procedia*, 37, pp. 155–162. doi: 10.1016/j.trpro.2018.12.178.

IBM (2012) 'CRISP-DM-Handbuch', p. 59. Available at: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/de/CRISP-DM.pdf.

Kdnuggets (2014) *What main methodology are you using for your analytics, data mining, or data science projects?* Available at: https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html.

Kotler, P., Wong, V., Saunders, J., & Armstrong, G. (2011) *Principles of Marketing*. 14th edn. Pearson Education.

Morency, C., Trepanier, M. and Agard, B. (2011) 'Typology of carsharing members', *In proceeding of: TRB-Transportation Research Board Annual Meeting*.

'MVV' (no date). Available at: https://www.mvv.de/energie/kundenservice/E-Roller/index.jsp.

Nations, U. (2018) *2018 Revision of World Urbanization Prospects*. Available at: https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html.

Reiss, S. and Bogenberger, K. (2017) 'A Relocation Strategy for Munich's Bike Sharing System: Combining an operator-based and a user-based Scheme', *Transportation Research Procedia*. Elsevier B.V., 22, pp. 105–114. doi: 10.1016/j.trpro.2017.03.016.

Sarin, S. (2010) 'Market Segmentation and Targeting', in *Wiley International Encyclopedia of Marketing*. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9781444316568.wiem01033.

Sarstedt, M. and Mooi, E. (2014) *A Concise Guide to Market Research*. Berlin, Heidelberg: Springer Berlin Heidelberg (Springer Texts in Business and Economics). doi: 10.1007/978-3-642-53965-7.

Schmöller, S. *et al.* (2015) 'Empirical analysis of free-floating carsharing usage: The munich and berlin case', *Transportation Research Part C: Emerging Technologies*, 56, pp. 34–51. doi: 10.1016/j.trc.2015.03.008.

Stella Stadtwerke Stuttgart (no date). Available at: https://www.stella-sharing.de/.

Vogel, P., Greiser, T. and Mattfeld, D. C. (2011) 'Understanding bike-sharing systems using Data Mining: Exploring activity patterns', *Procedia - Social and Behavioral Sciences*, 20, pp. 514–523. doi: 10.1016/j.sbspro.2011.08.058.

Weikl, S. and Bogenberger, K. (2013) 'Relocation strategies and algorithms for free-floating car sharing systems', *IEEE Intelligent Transportation Systems Magazine*, 5(4), pp. 100–111. doi: 10.1109/MITS.2013.2267810.

Yin, R. (2009) *Case study research: Design and methods*. 4th edn. Thousand Oaks, Ca.: Sage.

Zagorskas, J. and Burinskienė, M. (2019) 'Challenges Caused by Increased Use of E-Powered Personal Mobility Vehicles in European Cities', *Sustainability*, 12(1), p. 273. doi: 10.3390/su12010273.