Data Analytics in Free-Floating Carsharing: Evidence from the City of Berlin

Sebastian Wagner
University of Freiburg
sebastian.wagner@is.uni-freiburg.de

Tobias Brandt
University of Freiburg
tobias.brandt@is.uni-freiburg.de

Dirk Neumann
University of Freiburg
dirk.neumann@is.uni-freiburg.de

Abstract

Carsharing has emerged as an alternative to vehicle ownership and is a rapidly expanding global market. Particularly through the flexibility of free-floating models, carsharing complements public transport since customers do not need to return cars to specific stations. We present a novel data analytics approach that provides decision support to carsharing operators – from local start-ups to global players – in maneuvering this constantly growing and changing market environment. Using a large set of rental data, as well as zero-inflated and geographically weighted regression models, we derive indicators for the attractiveness of certain areas based on points of interest in their vicinity. These indicators are valuable for a variety of operational and strategic decisions. As a demonstration project, we present a case study of Berlin, where the indicators are used to identify promising regions for business area expansion.

1. Introduction

Through air pollution, emissions, and traffic jams, the progressing urbanization witnessed around the globe substantially reduces the quality of life of the very people that are drawn to the cities. While public transportation networks are growing, they continue to face obstacles with respect to public perception and flexibility. To serve the entire area at low cost, bus and train services often use neither the quickest nor the shortest route. Commuters are also bound by fixed departure and arrival times. This lack of flexibility prevents people from renouncing car ownership; yet, once a car is owned, the willingness to use public transportation decreases substantially [23]. Thus, people continue to rely on their own cars, thereby contributing to pollution, congestion, and the impending infarct of urban centers.

Free-floating carsharing has emerged as a possible solution to this dilemma. In contrast to one-way or round-trip carsharing models with fixed parking lots, the free-floating model allows customers to return the car anywhere within the operation area [34]. Driven by this flexibility, it complements public transportation with both components integrated into a hybrid transportation system. The popularity of carsharing has increased tremendously in recent years. In North America, membership has grown from approximately 16,000 in 2002 up to more than one million at the beginning of 2013 [26] – a compound annual growth rate of more than 45 percent. Customers can easily locate, reserve, and pay for the closest vehicle through smartphone applications and an online payment system, while an RFID card provides access to the vehicle. From an ecological and economical point of view, carsharing contributes to saving fuel, reducing accidents, and decreasing the total number of cars on the streets, thereby reducing CO2 emissions [6, 10, 8].

This growth in popularity and demand causes providers to constantly adapt their network, balance their vehicle capacities, and search for new regions to expand into. Currently, this process is often ad-hoc, with operators relying largely on general long-term strategies. Hence, a detailed business analysis can provide a substantial competitive advantage in the permanently changing market environment. Big Data analytics has emerged as an important field of IS research to improve timely business decisions [12]. Hence, the research presented in this paper employs a novel business analytics methodology to support the operation and management decisions of free-floating carsharing providers.

For this purpose, we cooperate with a globally leading carsharing provider and analyze the usage behavior of thousands of carsharing customers. This includes the analyses of millions of rentals from April 2012 through October 2013 in Berlin. Furthermore, we investigate whether specific locations (points of interest, POIs) influence the attractiveness of the surrounding area as destinations for carsharing customers. Thus, our first research question is:

Research Question 1: What is the impact of different points of interest on the driving behavior of carsharing customers?

1530-1605/15 $31.00 © 2015 IEEE
DOI 10.1109/HICSS.2015.112

We investigate this question empirically by relating the demand for a vehicle to surrounding points of interest. In addition to rental data and more than 180,000 POIs, our analysis also relies on census data, which is included as a control variable. We account for spatial variation by dividing the operating area into thousands of finely-granulated sub-areas, relating to the data mining research for geo-spatial Big Data [24]. The influence of each POI type, such as shopping malls or night clubs, on carsharing activity is determined through a zero-inflated Poisson regression. To validate this approach, we compare the results of the zero-inflated model to a geographically weighted regression, which leads to the following research question:

Research Question 2: Are the empirical results substantially changed when allowing coefficients to vary locally?

If the global regression coefficients serve well to explain the spatial behavior of carsharing customers, we are able to calculate the expected vehicle demand not only at locations within the operation area, but also for new areas. Hence, our final research questions is:

Research Question 3: What is the benefit of the presented approach with respect to strategic decisions concerning the expansion of the operation area?

Ultimately, our approach seeks to support carsharing providers in making decisions concerning the adaption and expansion of the operating area, as well as fleet balancing and management. In the next section, we will provide a brief overview on related work, before we address the research questions successively in the remainder of the paper.

2. Related Work

The history of carsharing dates back to the year 1948 when a housing cooperative founded the SEFAGE ("Selbstfahrgemeinschaft") in Zurich [11, 27]. Between the late 1980s and the 1990s, carsharing became more common, particularly in several European countries, such as Germany, Switzerland, and the Netherlands [28], but also in the U.S. [19], [14]. However, the share of costumers in the entire population was still very low (0.52‰).

2.1 Round-trip and one-way systems

Traditional, station-based, carsharing business models include the *round-trip* and the *one-way* concepts. In general, round-trip carsharing is strict and rather inflexible, since customers have to return cars to the station they started from [2]. Round-trip research is no longer actively conducted because the more flexible one-way concept has emerged. This model allows customers to arbitrarily choose a station to return the rented vehicle at. Consequently, researchers have to face a new challenge of temporal and spatial imbalances between vehicle supply and demand.

Barth et al. [3] introduce a user-based relocation mechanism that urges customers to share or split rides depending on the system demand. Kek et al. [18] test a decision support system for vehicle relocation on commercial data from a carsharing company in Singapore. Their results indicate a reduction in staff costs of 50% and the number of relocations by 37.1% to 41.1%. Further case studies and simulations for one-way carsharing are performed to establish models that seeks to optimize the imbalance between supply and demand [6, 4, 15].

As in our research, Stillwater et al. [30] provide a geographic IS based study to explain carsharing demand using built-environment and demographic data. While they argue that recent research is only partially able to find characteristics of neighborhoods that make carsharing successful, we will elaborate on these factors in the course of this paper.

2.2 Free-floating systems

As a result of increased demand for carsharing services, an extension of the one-way trip concept emerged known as "free-floating" carsharing. This concept allows users to leave the rented car anywhere within the provider's operation area. Thus, it increases flexibility on the consumer side but also complexity in terms of balancing supply and demand on the provider side. Previous research on free-floating systems focuses mainly on relocation strategies, for instance through a two-step relocation algorithm consisting of an offline and online module [34]. While the former pre-calculates possible relocation strategies based on historical data, the latter measures the current state and selects the best relocation strategy. Various empirical studies examine the impact of free-floating carsharing on other means of transportation and environmental effects [9, 10]. Results indicate a reduction of CO2-emissions and the overall number of vehicles in a city.

Finally, several studies attempt to investigate the criteria an urban region should have to establish a successful business. Millard-Ball et al. [21] show that most conducted rentals are associated with several points of interest, such as grocery stores or shops, while only 12% of all trips are work related. Celsor and Millard-Ball [5] found that neighborhood and public transport are more important success indicators for carsharing than the demographic characteristics of customers.

Melville [20] notes the important contributions information systems can have on sustainable transportation. However, while Degirmenci and Breitner [25] discuss a DSS to determine optimal locations for prospective stations, carsharing as one component of sustainable transportation has received very little attention from IS research [7].

2.3 Research Gap

Currently, carsharing providers lack sophisticated decision support to determine expansion and operation strategies. The research presented in this paper builds upon neighborhood characteristics that have already been identified as important determinants for a successful carsharing business in the literature. We develop a novel data-driven method to estimate carsharing demand based on points of interest and emphasize the tremendous impact IS research can have on the success of emerging sustainable transportation services.

3. Descriptive Statistics and Visualization

Since customers are allowed to park a rented vehicle at any location within a predefined area under a free-floating concept, fleet control becomes a major challenge. However, to our best knowledge, none of the recent publications investigated the influence of neighborhood features on the vehicle demand of a carsharing provider. As one of the main contributions of this paper, we conduct an extensive case study for the city of Berlin in cooperation with a globally leading carsharing provider. Rental data of 1,200 shared vehicles was collected over a period of 1 year. This includes more than one million trips conducted by a customer base of more than 55,000 members in total. We collected different types of anonymized information, for example, the GPS coordinates of the start and end position of each trip. Thereby, the customers themselves do not have to provide any information regarding the trip, such as the final destination or details on additional passengers.

The operation area encloses a region of almost 300km^2. Making decisions based on data of this volume and incorporating various sources, such as rental data, POIs, and demographic information, fulfills two main classifications of IBM's Big Data definition [1]. To keep on track, managers are consistently faced with changes, such as increasing vehicle fleets, new customers, or changes in the urban built environment. Even small variations like the inclusion or exclusion of a large shopping mall can result in substantial profits or losses. Therefore, one of the most important decisions and a key success factor, especially for free-floating carsharing providers, is the appropriate definition of the operation area.

In the course of this paper, we use Berlin as a reference city, since it will be used later in the case study. However, the approach is generally applicable to any city or urban area. As a first step, we mathematically define the operation area. A spatial area A is defined as a closed polygon of at least 3 different GPS-points, given by the following n-tuple[1]

$$A_n = (a_1^n, \dots, a_{m_n}^n, \dots, a_{\overline{m_n}}^n, a_1^n) \tag{1}$$

$$\text{s.t.} \qquad |A_n| = \overline{m_n} + 1 \geq 4$$

$$a_m^n = (\phi_m^n, \lambda_m^n),$$

with ϕ as latitude and λ as longitude value, which describes the corner point of the polygon.

In general, a free-floating carsharing business operates in at least one spatial region A, in which customer are allowed to end their rentals. In contrast, it is prohibited to end a rental at any location outside that area. Thus, an operation area O of a carsharing provider is defined as a set of areas $O = \{A_1, A_2, \dots, A_{\bar{n}}\}$. A rental itself is defined as the following 6-tuple

$$r_q = (\phi_{start}^q, \phi_{end}^q, \lambda_{start}^q, \lambda_{end}^q, t_{start}^q, t_{end}^q), \tag{2}$$

while ϕ and λ are again the GPS latitude and longitude values, and t the timestamp of the start and end position of the respective rental.

To give a graphical representation, Figure 1 visualizes the operation area, as well as the vehicle demand, of a usual business day for the city of Berlin in April 2012. The red points represent end positions of rentals for one day. The overall number of rentals on that day was more than 2,400. The blue polygons in Figure 1 represent the borders of different areas where customers are allowed to end their rentals. Of course, it is possible to cross these zones and to leave a car temporarily outside them during the course of a trip. However, as soon as the rental ends, the vehicle has to be located within the operation zone. As not all customers observe these rules, we see a few red markers outside the operation area.

[1] For the remainder of this paper we use an overline to indicate the maximum value of an index.

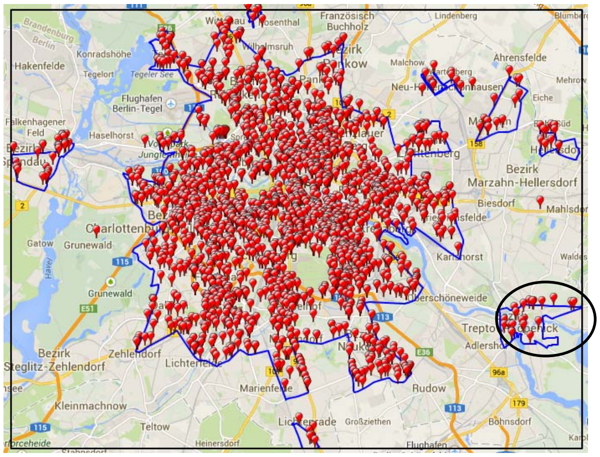


Figure 1. Ended rentals in Berlin on a usual weekday

Each location in Figure 1, whether inside or outside the permitted zones, is characterized by various landmarks. Obviously, some locations are more attractive for carsharing customers than others. The question arises as to which features cause this appeal. To answer this question, we look at regions with a high and dense amount of red markers. Thereby, we assume that the amount of ended rentals at a locations may be a proxy for the attractiveness of that respective region. Hence, we assume that locations with a high number of completed rentals are promising destinations for carsharing users, while locations where no rental was completed are considered to be unattractive. Due to the extensive data set, we are able to omit the settling-in period directly after the launch of the carsharing business. Thereby, we highly increase the probability that trips are carried out for a specific purpose and not just for trying out carsharing, making the above assumption applicable.

By using the geographical information of each rental, we are also able to visualize the driving behavior of customers. In fact, Figure 1 already demonstrates the substantial level of carsharing activity. However, Figure 2 provides a closer look at the circled area in Figure 1. While a cross marker represents the starting point of a rental, a square marks the destination. Since destinations are likewise the locations of new start points, each square should be overlaid by a cross in theory. Nonetheless, Figure 2 exhibits more squares than crosses, which is often caused by commuters who continue travel by public transport, and is one of the main reasons for the necessity of relocation algorithms.

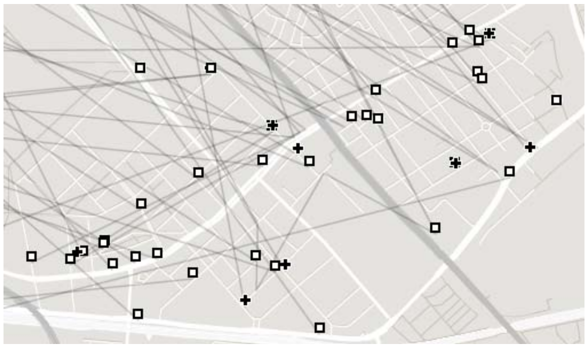


Figure 2. Snapshot of regional start and end positions of carsharing rentals

Based on our data set, we found that more than 80% of all rentals are completed within 0-30 minutes, while the duration exceeds two hours only for two percent. Evidently, customers use carsharing as a service for short trips rather than for longer periods of travel. This aligns with the perspective of using vehicles as a means of urban transportation in addition to buses or subways. Furthermore, the traveled distance is less than 10 kilometers for 76% of all trips. On average, the trip length is approximately 8 kilometers and, thus, rather short. Again, less than two percent of all rentals are longer than 30 kilometers, confirming the above statement of carsharing as an urban means of transportation. Concerning different times of day, almost 30% of all rentals take place between 4 and 8 p.m., while only about 5% are performed during the morning hours from 4 to 8 a.m. Further, there is almost no difference between days of the week. However, as the weekend draws near, the number of rentals slightly increases. This is associated with a shift of average end times to night and midday. We assume that this behavior is caused by societal activities, like having a drink in a bar, going out for dinner or a movie, or just visiting friends. Excluding the last activity, a special point of interest can be identified as the destination for each respective trip. Since the driving behavior changes at the weekend, we decide to focus only on weekdays in this research.

4. Empirical Investigation

In April 2012, the city of Berlin launched the largest carsharing project in the world [31]. Since that date the customer base, the number of vehicles, and size of the operation area has steadily increased, while managers are faced with new challenges. In order to make the right decisions, the ongoing business has to be continually analyzed by incorporating real-world developments. Therefore, we subsequently introduce a novel business analytics approach based on urban features (points of interest) to determine the expected vehicle demand at a certain location. To our best knowledge, no previous research has been able to derive precise performance indicators to establish a successful carsharing business.

To investigate the driving behavior of carsharing customers, we analyze more than 180,000 points of interests and their impact on driving behavior in the respective operation area (cf. Figure 1). To distinguish the POIs, they are tagged with specific categories, such as bar, gym, or restaurant, with means of public transport like bus, train, or subway stations also included. We assume that these locations are potential rental destinations for carsharing users, since research shows that more than 80% of all trips are related to personal or private purposes, like shopping or leisure activities [21]. All points of interest p_i are ordered according to Equation 3.

$$P = \{p_i\}_{1 \le i \le \bar{\imath}} \tag{3}$$

s.t.

$$p_i \mapsto (\phi_i, \lambda_i, \gamma^i)$$
$$\gamma = (accounting, \dots, zoo)$$
$$|\gamma| = 92$$
$$\gamma^i \subseteq \gamma$$

Each point is defined as a tuple $p_i = (\phi_i, \lambda_i, \gamma^i)$, with ϕ_i, λ_i as the respective GPS latitude and longitude values, and γ^i the categories the corresponding POI is tagged with. γ defines the tuple of all categories, with 92 different POI types in total.

In order to determine neighborhood features (POIs) at a given location and to prepare this information for upcoming regression analyses, we divide the operation area O in thousands of tiles as follows.

$$G(\Delta\lambda, \Delta\phi, \bar{x}, \bar{y}) = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,\bar{y}} \\ g_{2,1} & g_{22} & \cdots & g_{2,\bar{y}} \\ \vdots & \vdots & \ddots & \vdots \\ g_{\bar{x},1} & g_{\bar{x},2} & \cdots & g_{\bar{x},\bar{y}} \end{pmatrix} \tag{4}$$

s.t.

$$\bar{x}, \bar{y} \in \mathbb{N}$$
$$g_{x,y} \mapsto (\phi_{x,y}, \lambda_{x,y}),$$

with $\Delta\phi$ and $\Delta\lambda$ as the changes in latitude and longitude of the edge length of each tile. The resulting grid G consists of different tiles $g_{x,y}$, (sub area), while the geographical coordinates represent the center of the respective location. As a next step, we need to define the edge length of each tile to specify the size of each sub area. This value also determines the granularity of the business analytics procedure introduced in this paper. The smaller the edge length of the tiles, the greater the number of locations considered by the approach. However, note that as soon as the edge length drops below a certain threshold, neighboring cells barely differ, while complexity and the required computational power greatly increases. This is also the reason why we decided to use a latitude delta of $\Delta\phi$ = 0.0009 and a longitude delta of $\Delta\lambda$ = 0.001485, which correspond to an edge length of 100 meters. The resulting final grid consists of 80,925 individual tiles. However, considering Figure 1, the black rectangle that includes the whole visible area also includes many tiles that are not part of the operation area and, thereby, irrelevant for an initial empirical investigation. Hence, only a subset $G' \subset G$, which contains tiles exclusively inside the operation area, needs to be taken into consideration at this stage. To decide whether a certain location is inside or outside the operation area, we make use of the geographical information of each tile. The point in polygon algorithm [29] is used to face the above problem in linear time.

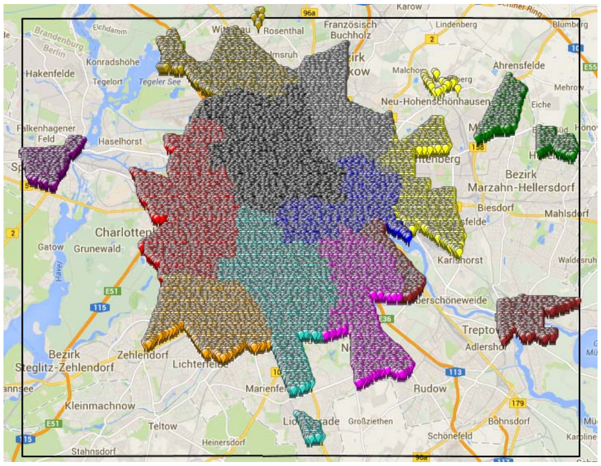


Figure 3. Operation Area Divided Into 24,280 Tiles

To give a graphical representation of G', Figure 3 shows a 3D visualization of the operation area divided into thousands of sub locations. The surrounding black rectangle marks the borders of the grid G. The colors in Figure 3 represent the 12 different districts of Berlin. Thereby, we are able to add demographic, educational, and economical information, such as population density, number of people with high or low education, unemployment and foreigner rates, and income per person, to each subarea. Since literature shows that such information drives carsharing success, we likewise incorporate this kind of data into our model. Another main reason for incorporating demographic data is to make the regression results more robust. It is possible that the sign of a coefficient changes because of a lack of demographic, educational or economical information.

Even rather small districts will be separated into many sub areas to emphasize any kind of minor change in the urban environment. In total, we have a number of tiles $|G'| = 24.280$, representing 30% of the whole area represented by G.

In order to use the initially mentioned POIs as independent variables to explain the number of completed rentals at a given location, the vicinities of each of the above areas need to be defined. In the model of van der Goot [33], this vicinity is restricted to an upper bound of a 40 minute walk to the target location. Within this timeframe, the incentive of reaching the desired destination decreases linearly, implying the "willingness to walk" of the respective user. As 40 minutes appears to reflect a very optimistic assumption of people's willingness to walk, we limit the vicinity to a radius of one kilometer. Thus, we assume that all POIs with a distance of more than one kilometer to a given location have no relevance, while a POI at the exact location has an impact of one. This also agrees with Tobler's Law "Everything is related to everything else, but near things are more related than distant things" [32]. Further, we deviate from van der Groot's model concerning the functional form. Instead of a linear relationship, we use a segment of the cosine wave. The reason for this is that the coordinates of POI locations do not necessarily align with, for instance, the entrance to the respective building. This concern is especially valid for large buildings, like museums. The cosine wave segment reflects this uncertainty by only weakly discounting the first couple of hundred meters, followed by a basically linear slope as in [33]. Hence, the impact $\iota_{i,x,y}$ of a certain POI p_i at a given location $g_{x,y}$ is calculated as

$$\iota_{\,i,x,y} = \begin{cases} \cos\left[\frac{\pi}{2} \cdot d(p_i, g_{x,y})\right], & if\ d(p_i, g_{x,y}) \leq 1 \\ 0, & otherwise \end{cases} \tag{5}$$

$$\text{s.t.} \qquad \iota_{i,x,y} \mapsto \mathbb{R} \in [0,1],$$

while the distance $d(p_i, g_{x,y})$ between two points p_i and $g_{x,y}$ is given as the great-circle distance on a sphere, calculated by the haversine formula[2].

We specify the neighborhood features of each tile by calculating the individual impact $\iota_{i,x,y}$ of all 180,000 POIs on the respective tile. Thereafter, we derive specific POI densities $\delta_{x,y,k}$ at each location by summing up all impact values of a certain category k.

$$\delta_{x,y,k} = \sum_{p_i \in \{P \mid k \in \gamma^i\}} \iota_{x,y,i} \tag{7}$$

Essentially, instead of saying that there are 10 bars within 1,000 meters of the center of tile, we consider the distance between each bar and the tile. This is of great importance if, for instance, all bars are within a distance of 999 meters. The resulting "bar-density" $\delta_{x,y,bar}$ is only 0.016. Assuming now a different tile $g_{x',y'}$ that is only 10 meters from all bars, it would have a $\delta_{x',y',bar}$ value of 9.998. Despite the fact that both tiles enclose the same 10 bars within a range of 1km, the bar-density and, thus, the respective impact of bars on tile $g_{x',y'}$ is more than 600 times higher.

Building upon the above density calculations, the objective of our subsequent analysis is to assess if POIs in general and which POI categories in particular increase or decrease the attractiveness of a location for carsharing. As a measurement for this attractiveness, we use the number of ended rentals $d_{x,y}$ in each tile. The resulting vector of dependent variables is $\vec{d} = \left(d_{1,1}, d_{2,1}, \dots, d_{\bar{x},1}, d_{1,2}, \dots, d_{\bar{x},\bar{y}}\right)^T$.

As might be expected, most of the cells do not include any rentals due to the following reasons. First, it may be forbidden to drive in a region, such as a pedestrian zone, or not possible, as in parks or lakes. Second, it can be assumed that some locations are generally uninteresting points to end a rental, for instance at highways. Third, the amount of zeros is also caused by the high granularity. Eventually, the total share of observed zeros is approximately 42%.

As a next step, we define the covariate matrix C in Equation 8. In addition to the first column for the intercept, this matrix contains all POI densities, as well as all control variables. The numerical index of the densities represents the position of category k in set γ given an alphabetical ordering. Hence, $\delta_{2,3,1}$ is the density of POIs tagged with "accounting" in tile $g_{2,3}$, because "accounting" is the first element in an alphabetical ordering of all elements of set γ. The variables $z_{x,y,1}$ to $z_{x,y,\bar{h}}$ are control variables, such as population density, education, or income per person.

$$C = \begin{pmatrix} 1 & \delta_{1,1,1} & \cdots & \delta_{1,1,|\gamma|} & z_{1,1,1} & \cdots & z_{1,1,\bar{h}} \\ 1 & \delta_{2,1,1} & \cdots & \delta_{2,1,|\gamma|} & z_{2,1,1} & \cdots & z_{2,1,\bar{h}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \delta_{\bar{x},1,1} & \cdots & \delta_{\bar{x},1,|\gamma|} & z_{\bar{x},1,1} & \cdots & z_{\bar{x},1,\bar{h}} \\ 1 & \delta_{1,2,1} & \cdots & \delta_{1,2,|\gamma|} & z_{2,1,1} & \cdots & z_{2,1,\bar{h}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \delta_{\bar{x},\bar{y},1} & \cdots & \delta_{\bar{x},\bar{y},|\gamma|} & z_{\bar{x},\bar{y},1} & \cdots & z_{\bar{x},\bar{y},\bar{h}} \end{pmatrix} \tag{8}$$

Consequently, the regression coefficients are given by the vector $\vec{\beta} = \left(\beta_0, \beta_1, \dots, \beta_{|\gamma|}, \beta_{|\gamma|+1}, \dots, \beta_{|\gamma|+\bar{h}}\right)^T$, where β_0 is the coefficient for the intercept, β_1 to $\beta_{|\gamma|}$ for the POI categories, and $\beta_{|\gamma|+1}$ to $\beta_{|\gamma|+\bar{h}}$ for the control variables.

In order to analyze the influence of the various POI types on $\vec{d}$ and to handle the high amount of

[2] We consider a mean earth radius of 6,371 kilometer.

zeros, we subsequently introduce a zero-inflated regression.

4.1. Zero-Inflated Regression

The zero-inflated model that we use to explain $\vec{d}$ assumes two processes: a Poisson process and a zero-generating process responsible for the excess zeroes [13]. Employing the pscl package in R on our data, the zero-inflated model is fitted using maxim-likelihood. As the output we receive two regressions. The first regression is the count model and returns the logarithm of the expected number of rentals for each cell given the covariates. The second regression returns the logit of the probability that the number of rentals in a cell is zero and caused by the zerogenerating process for each cell given the covariates.

By applying the regression model, we derive location-based success indicators for free-floating carsharing. Naturally, some of the 92 POI categories frequently emerge together, such as different stores, shops, and malls, or even built-in combinations, like ATMs and banks. Therefore, multicollinearity between different categories inevitably exists. To alleviate this problem, we calculate the variance inflation factor (VIF) for each POI category. By using this index, we stepwise delete categories until none of the remaining variables exceeds the generally used cut-off value of $VIF = 5$. Several general classes, like food or restaurant were thus removed, since they are composed of more specific categories. In total, 35 variables were deleted because of multicollinearity.

Further, several additional categories were manually excluded, due to an insufficient number of observations. We conducted the regression analysis using the resulting appropriate data set of independent POI categories and covariates. Table 1 displays the estimates, z values and significance levels of the most interesting regression output coefficients. With respect to research question 1, we can see that the airport density value is significant. This is also one of the most active regions for carsharing customers in Berlin. Interestingly, means of short distance transportation, like buses, are significantly positive, whereas train stations are significantly negative. This result is in accord with the findings of [16]. Furthermore, entertainment facilities, like movie theaters or night clubs, are potential destinations of carsharing trips. The districts where people leave their car are characterized by a high population density and foreigner rate. This could possibly reflect the fact that foreigners and expats are often only in the city for a limited time, making car ownership more unappealing.

Table 1 Zero-Inflated Poisson Regression Results

Dependent variable: Number of end rentals per tile; 24,280 observations	
POI type	Coefficient (t value)
(Intercept)	0.7630 (5.939)***
Airport	0.1298 (6.074)***
ATM	0.0139 (3.172)**
Bus station	0.0240 (12.827)***
Car rental	0.0219 (10.361)***
Car wash	-0.021 (-4.601)***
Gas station	-0.0027 (-0.568)
Meal delivery	0.0058. (1.705)
Meal takeaway	0.0483 (16.561)***
Movie rental	-0.0345 (-5.285)***
Movie theater	0.0170 (3.642)***
Night club	0.0210 (13.805)***
Post office	-0.0408 (-5.285)***
Shopping mall	0.0304 (5.483)***
Train station	-0.0129 (-5.194)***
Distance to center	0.0122 (5.401)***
Foreigners	0.9654 (6.377)***
Age 15 – 45	-0.1943 (-1.281)
High education	-0.9742 (-5.089)***
Income <500	2.4242 (4.868)***
Log. of population density	0.2603 (7.151)***
Significance levels: '***' 0.001 '**' 0.01 '*' 0.05	

4.2. Geographical Weighted Regression

In the following section, we analyze the results by using another method specifically for spatial analysis [17], namely the geographically weighted regression (GWR). This methodological concept is designed to explore nonstationarity in geographic parameters. Depending on a specific bandwidth, or in our case, the amount of nearest grid tiles, the estimates are set into spatial relationship. However, the approach requires caution since research has exposed flaws due to an increased amount of false-positives and faulty recognition of spatial nonstationary [22]. Therefore, an already fitted model – our zero-inflated Poisson model – is generally required before the GWR can be performed. Hence, our preliminary steps allow us to conduct a GWR to investigate the impact of various POI types on a local base instead of only on a global scope and, thereby, clarify research question 2. This is of particular importance if it happens that the coefficients vary from positive to negative values for a certain POI category in different areas. One reason for this could be that a shopping mall A is substantially larger and provides more popular shops in contrast to another shopping mall B. This entails the possibility that customers always prefer mall A instead of B resulting in a substantially lower vehicle demand at B. If such a case occurs, we need to make use of local estimates in order to explain the attractiveness of a certain region instead of using the global values of Table 1. Therefore, we conduct a geographically weighted regression for the whole operation area in Berlin. Since we initially use a zero-inflated Poisson regression in order to handle the high amount of zero-observations in our data set, we need to take this into consideration for the GWR as well. The quasi-Poisson distribution is a remedy to this problem. It estimates a scale parameter and provides the best fit for our model.

As a result of conducting a GWR based on our dataset, we achieve 24,280 local coefficients – one for each tile in G' – for each POI category. To show the spatial variation of estimates, Figure 4 provides a visualization for the POI type "bus station". The coefficient estimates from the GWR confirm the results of the zero-inflated Poisson regression (cf. Table 1). All estimates of the different POI types vary around the global values in Table 1, while at the same time no coefficient turns from positive to negative or vice versa. Hence, the results of the GWR model unveil geographic variation throughout the operation area without substantial effects on the global scope.

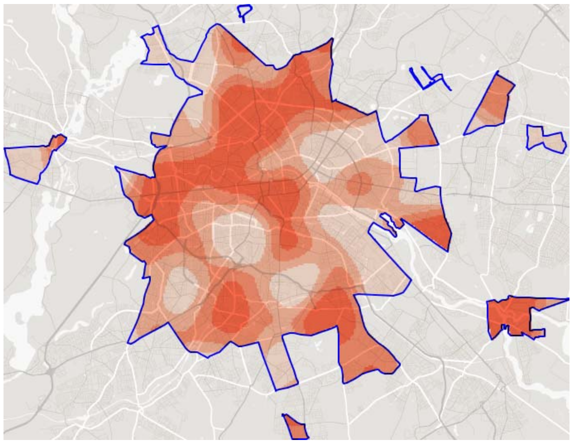


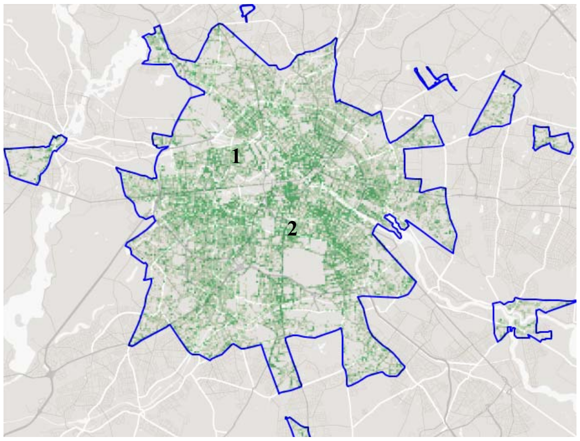
Figure 4. Local Estimates of POI Type "Bus station"

5. Validation & Managerial Implications

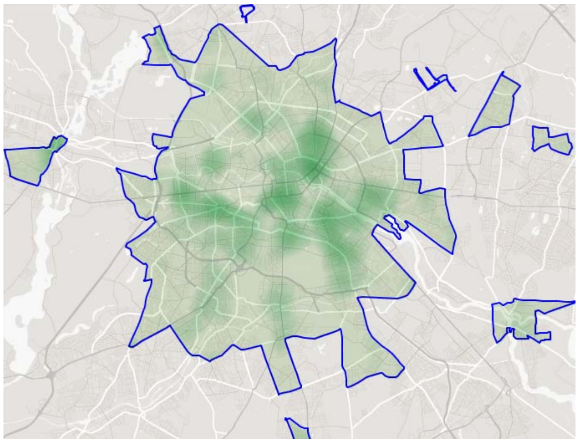
Since we now know the impact of various POIs on the driving behavior of customers (research question 1), we will further provide managerial implications for carsharing providers. The expected vehicle demand of a certain location is given by multiplying each covariate value with the respective coefficient. Further, we multiply this expected demand with the probability that the demand in the tile is equal to zero.

Thereby, we are able to calculate the expected demand for each tile within the initially defined grid G' and clarify research question 2. Figure 5a visualizes the actual vehicle demand based on three months of data (including more than 150,000 rentals) as a heat map. The dark green regions indicate a high number of ended trips, while white regions contain no rentals at all. As can be seen, there are two almost white areas right below the two numbers. Both regions are parks – 1 is the "Tiergarten" and 2 is the "Tempelhofer Park". Naturally, no rentals end at these locations, since it is not possible to park. If we now have a look at Figure 5b, we clearly see the same patterns. Not only the dark green regions match the ones in Figure 5a, but the aforementioned parks can be recognized, as well.

In order to investigate promising regions and to provide decision support for future expansion, we calculate the expected vehicle demand for the whole grid G (cf. Equation 4). Figure 6 illustrates attractive regions again as a heat map, while the model identifies three highly promising areas (black enclosed). In particular, the regions in the south and west seem to be a good choice for an extension, since a) the operation area adaption is rather small, which results in low expenditure, b) the area is rich in public transport, which is highly positively significant in both regression models (cf. Table 1), and c) the overall rental and driving behavior of the business model is assumed to remain the same, since the overall chances are only minor.



(a) Actual Vehicle Demand



(b) Expected Vehicle Demand

Figure 5. Actual vs. Expected Vehicle Demand

As a further managerial implication, the regression shows that the smaller southern areas (red enclosed) are unattractive for customers. Moreover, almost the whole south and east regions seem to provide no added value for carsharing businesses and, thus, are poor options for future expansion. If we look at Figure 6 as a whole, we also recognize that the carsharing potential of Berlin is almost exhausted and covered by the current operation area. Only small adaptions can be conducted in order to improve business. However, even small changes can have a major impact on the overall system.

In order to provide decision support for current carsharing providers, our methodology can be used in multiple ways. On the one hand, under-performing areas can be identified in advance, which results in an enormous cost saving potential. On the other hand, the identification of high-performance areas can serve as a basis for new or incorporated into existing relocation strategies. Furthermore, information about demand at certain areas can be used to adapt the amount of vehicles needed at different times of the day, thereby increasing the overall profit of the carsharing provider.

Naturally, the explanatory power of our model has its limitations. With a close look at the borders, we observe that people frequently leave their rented cars at the edges of the operation area. This is well-known customer behavior since, oftentimes, customers do not live in the operating area and so try to get as close to home as possible. The same phenomena appears if people use carsharing for long trips beyond the borders of the business area. In this case, they naturally try to return the rented car to the closest (permitted) location, which is similarly next to one of the borders.

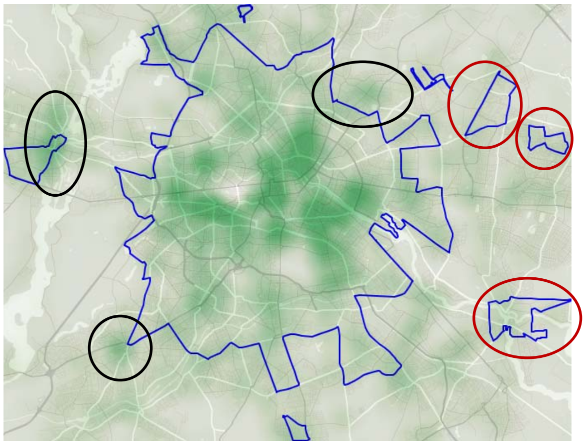


Figure 6. Expected vehicle demand for entire grid

6. Conclusion

During the past decades, the popularity of carsharing as an alternative transportation service has continuously increased, turning it into one of the most promising businesses for sustainable transportation. Companies in this quickly expanding market are constantly required to reassess business strategies, expand operation areas, and react to shifts in demand.

In this paper, we develop a novel method to support this decision-making process. Building upon a large dataset provided by a globally leading carsharing service, we investigate the influence of points of interest on the attractiveness of their surrounding vicinity for carsharing customers. With respect to our initial research questions, we find substantial evidence for a statistically significant relationship, employing both a zero-inflated regression model, as well as a geographically weighted regression. We show that our method is able to approximate the demand for vehicles very accurately and identify key success factors for a carsharing operation area, such as shopping venues or movie theaters. Furthermore, we can empirically support suggestions from related publications that short-distance transportation complements car-sharing activity, while long-distance trains appear to be a substitute. While coefficients naturally vary when enabled to do so in the geographically weighted regression, this variation occurs around the global value and does not have substantial effects, such as changing the sign of the coefficient. Furthermore, we demonstrate how the presented method can be used to predict future demand in locations that are under consideration for an expansion of the operation area. The information provided is valuable to carsharing enterprises of any size.

Our results also emphasize the potential of using information provided by services such as *Google Maps* and *OpenStreetMap* to explain spatial variation in human behavior. In our future research, we will analyze how the accuracy of the model at the edges of the operation area can be improved and further investigate the link between explaining demand and predicting demand in our model [29]. To improve our current model, we will include border-dummies to cover the edge phenomena of free-floating carsharing. Since we use a zero-inflated model and the share of observed zeros is relatively high, future work will also focus on testing and handling overdispersion. Our approach also provides insights for day-to-day operations. Hence, we will analyze how the information gained can be used to derive optimal relocation for carsharing vehicles.

7. References

[1] http://dashburst.com/infographic/big-data-volume-variety-velocity/, accessed 11-5-2013.

[2] Barth, M. and S. Shaheen, "Shared-Use Vehicle Systems: Framework for Classifying Carsharing, Station Cars, and Combined Approaches", Transportation Research Record, 1791(1), 2002, pp. 105–112.

[3] Barth, M., M. Todd, and L. Xue, "User-Based Vehicle Relocation Techniques for Multiple-Station Shared-Use Vehicle Systems", in Proceedings of the 83rd Annual Meeting of the Transportation Research Board. 2004.

[4] Boyacı, B., N. Geroliminis, and K. Zografos, "An optimization framework for the development of efficient one-way car-sharing systems", 13th Swiss Transport Research Conference, 2013.

[5] Celsor, C. and A. Millard-Ball, "Where Does Carsharing Work?: Using Geographic Information Systems to Assess Market Potential", Transportation Research Record, 1992(1), 2007, pp. 61–69.

[6] Cepolina, E.M. and A. Farina, "A new shared vehicle system for urban areas", Transportation Research Part C: Emerging Technologies, 21(1), 2012, pp. 230–243.

[7] Degirmenci, K. and M.H. Breitner, "Carsharing: A Literature Review and a Perspective for Information Systems Research", in Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI 14). 2014: Paderborn, Germany.

[8] Fellows, N.T. and D.E. Pitfield, "An economic and operational evaluation of urban car-sharing", Transportation Research Part D: Transport and Environment, 5(1), 2000, pp. 1–10.

[9] Firnkorn, J., "Triangulation of two methods measuring the impacts of a free-floating carsharing system in Germany", Transportation Research Part A: Policy and Practice, 46(10), 2012, pp. 1654–1672.

[10] Firnkorn, J. and M. Müller, "What will be the environmental effects of new free-floating car-sharing systems? The case of car2go in Ulm", Ecological Economics, 70(8), 2011, pp. 1519–1528.

[11] Harms, S. and B. Truffer, "The Emergence of a Nation-wide Carsharing Co-operative in Switzerland", A case-study for the EC-supported rsearch project “Strategic Niche Management as a tool for transition to a sustainable transport system”, EAWAG: Zürich, 1998.

[12] Hsinchun, C., Roger H. L. Chiang, and C.S. Veda, "Business intelligence and analytics: from big data to big impact", MIS Q, 36(4), 2012, pp. 1165–1188.

[13] Johnson, N.L., A.W. Kemp, and S. Kotz, Univariate discrete distributions, 3rd edn., Wiley, Hoboken, N.J, 2005.

[14] Jorge, D. and G. Correia, "Carsharing systems demand estimation and defined operations: a literature review", European Journal of Transport and Infrastructure Research, 13(3), 2013, pp. 201–220.

[15] Jorge, D., Correia, Goncalo H. A., and C. Barnhart, "Comparing Optimal Relocation Operations With Simulated Relocation Policies in One-Way Carsharing Systems", IEEE Transactions on Intelligent Transportation Systems, 2014, pp. 1–9.

[16] Katzev, R., "Car Sharing: A New Approach to Urban Transportation Problems", Analyses of Social Issues and Public Policy, 3(1), 2003, pp. 65–86.

[17] Kauffman, R.J., A.A. Techatassanasoontorn, and B. Wang, "Event history, spatial analysis and count data methods for empirical research in information systems", Information Technology and Management, 13(3), 2012, pp. 115–147.

[18] Kek, Alvina G. H., R.L. Cheu, Q. Meng, and C.H. Fung, "A decision support system for vehicle relocation operations in carsharing systems", Select Papers from the 19th International Symposium on Transportation and Traffic Theory, 45(1), 2009, pp. 149–158.

[19] Loose, W., M. Mohr, and C. Nobis, "Assessment of the Future Development of Car Sharing in Germany and Related Opportunities", Transport Reviews, 26(3), 2006, pp. 365–382.

[20] Melville, N.P., "Information systems innovation for environmental sustainability", MIS Q, 34(1), 2010, pp. 1–21.

[21] Millard-Ball, A., J. ter Schure, C. Fox, J. Burkhardt, and G. Murray, Car-Sharing: Where and How It Succeeds, The National Academies Press, TCRP Report 108, 2005.

[22] Páez, A., S. Farber, and D. Wheeler, "A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships", Environment and Planning A, 43(12), 2011, pp. 2992–3010.

[23] Paulley, N., R. Balcombe, R. Mackett, H. Titheridge, J. Preston, M. Wardman, J. Shires, and P. White, "The demand for public transport: The effects of fares, quality of

service, income and car ownership", Innovation and Integration in Urban Transport Policy, 13(4), 2006, pp. 295–306.

[24] Petry, F.E., "Data Mining Approaches for Geo-Spatial Big Data", International Journal of Organizational and Collective Intelligence, 3(1), 2012, pp. 52–71.

[25] Rickenberg, Tim A. A., A. Gebhardt, and M.H. Breitner, "A Decision Support System For The Optimization Of Car Sharing Stations", in Proceedings of the 21st European Conference on Information Systems. June 5-8, 2013: Utrecht, Netherlands.

[26] Shaheen, S. and A. Cohen, "Innovative Mobility Carsharing Outlook: Carsharing Market Overview, Analysis, and Trends", Transportation Sustainability Research Center, Richmond, California(06.03.2014), 2013.

[27] Shaheen, S., D. Sperling, and C. Wagner, Carsharing in Europe and North American: Past, Present, and Future, University of California Transportation Center, 1998.

[28] Shaheen, S.A., D. Sperling, and C. Wagner, A Short History of Carsharing in the 90's, Institute of Transportation Studies, UC Davis, 01.01.1999.

[29] Shimrat, M., "Algorithm 112: Position of point relative to polygon", Communications of the ACM, 5(8), 1962, p. 434.

[30] Stillwater, T., P.L. Mokhtarian, and S.A. Shaheen, "Carsharing and the built environment: Geographic information system-based study of one U.S", Transportation Research Record, 2009, pp. 27–34.

[31] The New York Times, "Car Sharing Grows With Fewer Strings Attached", 2013.

[32] Tobler, W.R., "A Computer Movie Simulating Urban Growth in the Detroit Region", Economic Geography, 46, 1970, p. 234.

[33] van der Goot, D., "A model to describe the choice of parking places", Transportation Research Part A: General, 16(2), 1982, pp. 109–115.

[34] Weikl, S. and K. Bogenberger, "Relocation strategies and algorithms for free-floating Car Sharing Systems", in 15th International IEEE Conference on: Intelligent Transportation Systems (ITSC), 2012. 2012.