Technische Universität München

GENERATION AND DISTRIBUTION OF SHARED BICYCLE TRIPS BASED ON SPATIAL AND TEMPORAL ANALYSIS OF DB CALL-A-BIKE DATA

Master's Thesis

Author:

Katrin Tambunan
03662488

Supervisor:
Dr. Carlos Llorca Garcia

Transportation Systems

Date of submission: 2017-07-30

DECLARATION

I hereby confirm that the presented thesis work has been done independently and using only the sources and resources as are listed. This thesis has not previously been submitted elsewhere for purposes of assessment.

I declare the details state to be true and complete.

Munich, July 30th 2017

(Katrin Tambunan)

ABSTRACT

DB Call-a-Bike is a bike sharing program in Germany. Users can access the locked bicycles by entering a combination of numbers sent to their smartphones. In this way, DB Call-a-Bike can track the usage of each bicycle and produces the database of the bicycles' usage. The database consists of several sets of information such as the start and end time of the usage, origin and destination of each bicycle trip, date of the usage, name of the city, and other related information. This database can be useful to know the current usage pattern of bicycles, predict the bicycles usage in the future, and find factors that affect the usage of a bicycle sharing system.

The purpose of this thesis is to model trip generation and trip distribution based on the DB Call-a-Bike usage database. The first step is to model the trip generation of DB Call-a-Bike data with multilinear regression analysis. Explanatory variables used are population (from Germany census data 2011) and points of interest (from Geofabrik) which were merged to each set of bicycle trip information based on the synthetic bicycle zone of each trip. Several models were produced and the outputs were compared with the observed data to investigate the quality of the model. The second step is to model trip distribution with a gravity model. From DB Call-a-Bike observed data, the total number of trips for 3 years from each origin and destination zone were used as the inputs of the gravity model. Similar to the trip generation model, the results of trip distribution were compared with observed trips.

The results of the trip generation are the relation of each explanatory variable with the number of trips through the regression coefficient. These coefficients were used to build the trip generation model that will be applied to some scenarios and then will be used as an input of the trip distribution model. The changes of the trip after and before the application of the scenarios can be seen and analyzed. However, the results are different for each model, depending on the objective of the study. The model can be further improved by modifying the synthetic zones.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

Bicycle sharing system is a transportation program that allows people to use bicycles in a flexible way. The first bicycle sharing system was introduced in the 1960s in Amsterdam which was known as the “White Bikes”. Bicycles were distributed around the city for public use, but unfortunately this programme was aborted due to repeated cases of stolen bicycles (El-Assi et al. 2017). In the new digitalized era, bicycle sharing system also developed. The most popular bike sharing system is by providing unattended bike stations with locked bicycles where users can drop and pick bicycles whenever they want. They can also unlock the bikes with smartcards (Shaheen et al. 2012) (DeMaio 2009) or with a combination of numbers sent through a bike sharing app or to the users phone (Frade, Ribeiro 2014). In some cities in Germany for example Munich, bicycle sharing systems such as DB Call-a-Bike and MVG Rad provide more flexibility by allowing bike sharing users to pick and drop the bicycles wherever they want without obstacle of searching for a bike stations. Some of these systems charge the users on an annual, monthly, daily, or per-trip basis (Shaheen et al. 2012).

1.1. Background

Nowadays, the bicycle sharing systems use IT based public bicycle sharing program where the users can unlock the bicycles by entering a combination of number codes through an application on their smartphone. This system allows the bike sharing provider to identify the customer, provide costumer with information of nearest bicycle location, and track the usage of bicycle (El-Assi et al. 2017). Through this system DB Call-a-Bike can produce the database of DB Call-a-Bike usage.

DB Call-a-Bike is one of the most popular bike sharing systems by Deutsche Bahn which is one of the biggest transportation providers in Germany and available in 47 cities. In 2016, DB has been provided an open database of Call-a-Bike for free through its open data portal http://data.deutschebahn.com (Deutsche

Bahn 2016). These datasets covered all information from 2014-2016. The datasets consist of several information which are station data, booking data, vehicle data, and tariff data.

The purpose of this thesis is to study the behavior of DB Call-a-Bike users spatial and temporal interaction by analyzing these datasets. The dataset contains a register of anonym users and bookings, start and end time of the usage, start and end location of the bicycle usage, etc. These datasets could be useful to predict number of trips generated and their origin and destination using the trip generation and trip distribution model.

Other spatial data such as the geographical data was also taken in to account to support the analysis of DB Call-a-Bike datasets. Spatial information such as location of points of interest (POIs) consist of education places like school, shopping places like malls and supermarkets, entertainment places like cinemas or theatres, and food and recreational data like restaurants and sport places is also important to know purpose of the trips of DB Call-a-Bike users. Ideally, population and POIs are the best predictors of number of trips. Krykewycz et al. 2010 hypothesize that zones with higher population density will produce more trips and zones with more POIs will attract more trips.

Adding variables that are assumed to influencing trip generation and distribution models is important to determine which temporal and spatial factors influence people to use DB Call-a-Bike. Investigating how the model changes by adding and eliminating the variables for example population, number of bike stations, and POIs will give a sense of which factors influence bike sharing trips. Comparing the observed data and the trips predicted by the model is also important to explore the room for improvements in the model. Models are build only to help people gain insight of what probably would happen if a series of events occur. Therefore, models do not guarantee the validation of the results.

1.2. Research question

The purpose of this research is to predict the number of trips with trip generation and trip distribution models by finding the relation between factors that probably affects trips with DB Call-a-Bike. The purpose would be more explainable by the following research questions.

- How does the relationship between spatial factors (number of bike zones, population density, and diverse POIs) and temporal factors (season, weekend, weekday, peak hour, offpeak hour) affect the number of trips?

Analyzing the observed data will give the overview of the DB Call-a-Bike usage and gain insight into the preferences of people for bike sharing program. Several steps below are established to support the process of answering the main research questions.

a. To perform a temporal and spatial analysis and to understand the behavior of DB Call-a-Bike users. This analysis will be limited to a selected number of cities in Germany.
b. To predict the number of trips and their destinations based on spatial and temporal (daily, monthly, and yearly) characteristics, by estimating trip generation and distribution model.
c. To identify factors that stimulate bike sharing usage (e.g. higher area density leads to a higher bike sharing usage rate). After that, analyze the appropriate city profile in Germany that suitable to continue the DB bike sharing system.
d. To forecast bike sharing trips under different scenario of expanding bike sharing zones.

1.3. Organization of thesis

The outline of this thesis reported as below

- Chapter 2: Literature review

This chapter gives a brief explanation about DB Call-a-Bike sharing system in Germany, other study of bike sharing usage in the world, and models used to predict trip generation and distribution.

- Chapter 3: Methodology

This chapter explains methods implemented to analyze the data and steps to prepare and combine various data to be processed.

- Chapter 4: Data collection and analysis

This chapter describes details of the data and how to process it to produce desired models and compare the results from the model with observation data.

- Chapter 5: Application

This chapter presents the application of the model by expanding bike sharing system zones and shows the changes before and after the application.

- Chapter 6: Conclusion and discussion

This chapter containt the conclusion of the research results, constraints, findings, and recommendation for better research in related area.

CHAPTER 2 LITERATURE REVIEW

A lot of bicycle sharing study explored factors that trigger bicycle sharing usage to investigated factors that attract bicycle sharing systems customers. Not every city in Germany has a good DB Call-a-Bike coverage. Most of the cities only have 1 DB Call-a-Bike stations in the central train station area of the city (Hauptbahnhof). From DB Call-a-Bike open data portal, only 10 cities in Germany that have more than 5 bike sharing stations. Each of the city has different total number of bike sharing trips, different number of bike stations, and different distance from 1 bike stations to other bike stations. This situation leads to a question of what factors that make each city has different bike sharing activity.

2.1. Call-a-Bike in Germany

DB Call-a-Bike has provided bike sharing systems in 47 cities in Germany in 2000 (Deutsche Bahn). The coverage area of DB Call-a-Bike depends on the size of the city. DB Call-a-Bike serves 47 German cities and places the docking stations in the central train station area (hauptbahnhof). Additionally, for larger cities such as Berlin, Hamburg, Kassel, etc. docking stations cover almost all areas in the city. In some cities, DB Call-a-Bike cooperates with the municipality or local transportation authorities. For example, in Kassel, DB Call-a-Bike cooperates with Konrad Kassel and in Hamburg with StadtRAD Hamburg (Deutsche Bahn).

Most of the cities have station based rentals, which means there are dedicated docking stations to rent and drop the bicycles. Other cities for example Munich, Cologne, and some areas in Frankfurt am Main do not have docking stations, so users can rent and drop bicycles wherever they want. There are two types of rental costs, the basis tariff and the comfort tariff. The basis tariff allows users to pay only one time when customers are using the bicycles. For the basis tariff, users are obligated to pay 3€ per year as registration cost, and an additional 1€ for every 30 minutes and a maximum 15€ for 24hour usage. For the comfort tariff, users are obligated to pay 49€ per year and allows users to use the bicycles for free in the first 30 minutes everyday in the year and an additional 1€ for the next

30 minutes. The maximum tariff is 12€ for 24hour usage. DB Call-a-Bike does not serve every city in Germany since 2000. In Damstart, Call-a-Bike started from April 2014 (ASTA TU Darmstadt 2014), in Frankfurt from 2013, and in Marburg and Rüsselsheim started from March 2014.

2.2. Bicycle sharing systems

There are already several studies on temporal and spatial bike sharing usage with the data from bike sharing systems providers. All these research consider similar spatial factors for their bike sharing systems research. Gregerson et al. researched in Seattle about indicators that stimulate bike sharing usage which were population density, non-institutionalized group quarter housing, job density, retail job density, commute trip reduction companies, tourist attractions, parks/recreation areas, topography, regional transit stations, streets with bicycle lanes, and local transit stops. This research divided the entire city to a series of ten meter squares of cells and generated a score for each cell based on the numbers of indicators (Gregerson et al. 2010).

Vogel et al. analysed bike sharing systems in Vienna and clustered the bike stations according to spatial factors similarly to Gregerson et al. research of indicators that stimulate bike sharing usage. This research used data of Citybike Wien, a bike sharing system in Vienna. The data was from bicycle pick-ups and returns from the bike stations for certain time spans. This research purposed was to know the patterns of the bike sharing usage along the day. With the help of Geo Business Intelligence, they clustered the bike stations according to spatial factors (e.g. population, houses, public transport, etc.) to get the type of the costumers and trip purpose. This research comes up with an analysis of five different patterns of temporal pickup and return activity which are *Return Morning Pickups Evening* stations, *Pickups Morning Returns Evening* stations, *Active Night Pickups Morning* stations, and *Active Daytime* stations (Vogel et al. 2011). This research also clustered the study area according to the spatial condition which are stations around working places, stations around tourist attractions, stations around night clubs and bars, and stations around residential buildings. This research proved the hypothesis that bike sharing activity and customers demand depend on bike sharing stations' location. The analysis of this research focusing on the bike stations locations and took the assumption based on stations locations activity for the customer demand. For better results, process of better profiling customers based on location would be support the hypothesis stronger.

Bachand-Marleau et al. researched on BIXI bike sharing in Montreal found another factor that stimulate bike sharing systems usage. Based on the online survey results, the study showed that a higher number of docking stations close to the origins of potential users was highly likely to generate and increase number of bike sharing users. This study also confirmed that on of the reason of using bike sharing system was to avoid the fear of bicycle theft. Limitation of online survey is the distribution of the survey. Because this is a voluntary survey, the data collected have a risk that it cannot represent population. More disperse survey should be conducted to get information from all groups of the population (Bachand-Marleau et al. 2012). Other limitation is this research focusing on customer origin to know factor that trigger bike sharing systems usage and only took consider the distance from home to downtown area. Broader destination factor that trigger the bicycle sharing trip should be take in to account to give more understanding of bicycle sharing trip usage.

The previous bike sharing system usage study in Vienna emphasize on bike station location or trip attraction while bike sharing system usage study in Montreal emphasize on customer profile and origin. In 2012, a study of bike sharing system in Seville and Barcelona, Spain, consider both origin and destination factors of bike sharing usage in sub city district level. Factors that considered in this study were population, business, transport, and leisure recreation and university. The dependant variable of this study was the hourly arrival and the hourly departure rates of bicycle in sub city districts of the city. The population data were from Eurostat Urban Audit while the attraction data were from Tele Atlas Points of Interest. This study conclude that number of bike stations, population, and labor market size are strongly related to number of trips generated by bike sharing systems in Seville and Barcelona (Hampshire, Marla 2012). Each of the sub city district has different area. The wider the area will have higher POIs and population and so does the vice versa. To analyse the bike sharing system user, it would be better to do the analysis in the same wide of the area so the bias of the number of spatial factors can be avoided.

Similarly to Vogel et al. research, Rixey used seven months of station level data to forecast bike sharing ridership of Capital Bikeshare in Washington, D.C., Nice Ride in Minneapolis, and Denver B-cycle in Colorado. This research used regression analysis to identify variables that had significant relations with bike sharing trips. Variables he chose are population density, retail job density, income, education, the presence of bikeways, days of precipitation, and race. Datasets used in this research was the monthly average rentals by stations provided by the bike sharing system providers. Spatial variables that he chose were based on census block level data and census tract level data. They also built environment factors according to the location for parks and recreation facilities. For parks location, if data from government agencies was not available, they used Open Street Map (OSM) data. The data were aggregated to the 400-meter buffer surrounding each station. This study showed that all independent variables in the regression were statistically significant to bike sharing trips at 1% chance of the random event except retail jobs at 5% chance of random event (Rixey 2013). Similarly to Bachand-Marleau et al. researched, this research focusing on demand profile, with additional of transportation network and built in environment factor and not considering trip destination or trip attraction spatial factor of bike sharing users.

2.3. Trip generation

A trip generation model is a model to predict the number of trips generated by a zone. There is already some previous study did research about factors that affect bicycle trips. Some of the factors are population density, job density, location of tourist attractions, proximity to parks and recreation, and infrastructure (Krykewycz et al. 2010). Another factor other than spatial factors that affect bicycle trips is temporal factor. Tin Tin et al. did research about temporal factors that affect cycle volume in Auckland. According to this research, there were significant differences in mean hourly cycle volumes by hour of the day, day type (weekday or weekend), month of the year, and especially weather. The cycle volume increased by 3.2% hourly for 1^oC increase in temperature but decreased by 10.6% hourly for 1 mm increase in rainfall (Tin Tin et al. 2012).

Furthermore to trip generation factors, Ortúzar, Willumsen 2011, ©2011 divided factors that affect trip generation to three main groups which are personal trip production, personal trip attractions, and freight trip production and attractions. There are only two main groups of factors in bike sharing that affect trip generation. First is the personal trip production that has income, car ownership, family size, household infrastructure, the value of land, residential density, and accessibility as trip generation factors. The second group is personal trip attractions which have space available for industrial, commercial and other services as trip generation factors.

One of the trip production factor as Ortúzar, Willumsen mentioned is accessibility. Accessibility can be defined as the degree of easiness or ability to reach mean activities which might require traveling to the place where those opportunities are located (Handy 2005). The accessibility formula is based on gravity model. Equation 2.1 shows the description of the formula.

$$A_i = \sum_{j=1}^{n} D_j \mathrm{e}^{-\beta c_{ij}}$$

Equation 2.1 Accessibility equations (Geurs et al. 2006)

Where Ai is the measure of accessibility in zone i to all opportunities D in zone j, Cij is the generalized cost of travel between i and j, and β represents the cost sensitivity parameter (Geurs et al. 2006). D is the population of each zone, this means to target the total number of people that can reach the destination zone. The generalized cost of travel between i and j can be distance or travel time from origin to destination. In addition to cost sensitivity parameter, Iacono et al. did a research in Twin Cities, Minnesota, about the β value which represents the level of deterrence to travel with various mode of transport by distance. The bicycle has a limitation as a mode of transport because it is using direct human power to travel, therefore there is a limitation of the duration of bicycle usage as a transport mode for commuting. Based on the sample used in this research, for a bicycle, the travel distance reached falling within range 10km. This research found various β value depends on the trip purposes and the time range of the day. The β value will prevent the accessibility and gravity model to overpredict shorter duration and underpredict the longer duration of the bicycle travel. This research found that the cost sensitivity parameter for bicycle was depended on the trip purpose. For bike trips, the ß parameter range from 0.12 for school-related trips to over 0.5 for shopping trips (Iacono et al. 2008).

There are several options to perform trip generation model. In this paper, multiple regression model is used to perform the trip generation model. The dependent variable is the number of observed trips and the independent or explanatory variables are the factors that assumed to affect number trips (Krykewycz et al. 2010). Equation 2.2 is the formula of the linear regression model

$$Y = a + b_1x_1 + b_2x_2 + \ldots\ b_nx_n$$

Equation 2.2 multiple linear regression model (Ortúzar, Willumsen 2011, ©2011)

Where Y is the dependent variable (number of trips), x_1, x_2, are independent variables (populations, attraction factors, season) b_1, b_2,..b_n are regression coefficients to show the relation of independent variables to the dependent variable (Ortúzar, Willumsen 2011, ©2011).

2.4. Trip distribution

Trip distribution is a model to predict the number of trips for each origin and destination zone. There are several methods to predict this number. In this thesis, double constraint gravity model is selected and using travel time as impedance. The predicted numbers then compare with observed trips number to know whether the model produces the reasonable number (Ortúzar, Willumsen 2011, ©2011). Equation 2.3 explains the gravity formula:

$$T_{ij} = \beta i * Pi * \alpha j * Aj * f(cij)$$

Equation 2.3 gravity model (Ortúzar, Willumsen 2011, ©2011)

Where:

T_{ij} = Trips produced at i and attracted at j

β_i = Balancing factor for row i (production constraint)

P_i = Total trips production in zone i

α_j = Balancing factor for column j (attraction constraint)

A_j = Total trip attraction to zone j

$F(c_{ij})$ = Impedance function

Impedance function usually considers in-vehicle and out of vehicle time and multiply it by the cost. The cost can be travel time or distance. There are several studies about impedance function of the bicycle trip. A study in Twin cities Minnesota was aimed to estimate distance decay function to describe travel impedance of various transportation network. The distance decay function for bicycle varies from -0.03 for work trips, -0.182 for shopping trips, -0.199 for school trips, and -0.105 for recreation trips. Figure 2.1 shows the comparison of the travel time with the share of the trips according to the bicycle trip purpose. The results show that people are willing to travel more for work and recreational trips rather than school and shopping (Iacono et al. 2008).

Figure 2.1 Distance decay curves for bicycle trips (travel time) (Iacono et al. 2008)

Another travel impedance study used data from National Household Travel Survey of United States to investigate the influence of trip distance on non-motorized transport. This study divided the classification of the bike users according to gender, age, education, occupation, income, purpose, and travel day. The travel impedance values are according to the user classification and varies from -0.178 for people who obtained bachelor degree until -0.371 for shopping purpose (Mondal et al. 2015).

2.5. Current study in context

The previous bike sharing literature mentioned focusing on demand profile or bike stations location, or combine both of origin and destination factors. The previous study mentioned also focus in 1 until 3 cities in the same country. There are still more factors of the trip attraction factors that can be explored. The bike sharing usage data provided by DB Call-a-Bike has origin and destination information for each trip. This data can be useful to achieve the focus of this study which is exploring the trip production and the trip attraction factors that influence bike sharing in Germany based on its spatial condition. Ten cities from Germany from different states will be analyzed. Each city has different demographic therefore it will be interesting to know the different of bicycle sharing system usage depends on the city.

Both of trip production and trip attraction factors will be taken into account in this study. Points of interest as trip attraction and population as trip production are expected to explain the reason behind different DB Call-a-Bike usage situations in each city. Using the OSM data to locate POIs will be very useful to analyze trip attraction factors comprehensively. From previously mentioned literature, only one study used OSM data for built environment of parks and recreation, while in fact, OSM database covers the POIs information in detail (for Germany).

This study focusing on factors that generate and attract bike sharing system usage according to temporal and spatial condition, therefore dividing each city in the same size is expected to give more detail of spatial factors that influence bike sharing in a certain area rather than analyzing according to the buffer of each station or according to the sub-city district level to avoid bias in the spatial level. Factors that have relation with DB Call-a-Bike usage and model applied will be a useful consideration to give a decision of extensive application of bike sharing system in Germany.

CHAPTER 3 METHODOLOGY

Analysis of DB Call-a-Bike data consists of several steps. The first step is to collect all required data according to trip generation and distribution model. After that, data from DB Call-a-Bike will be sorted and tidied up based on required information according to the trip generation and distribution factors and variables. This includes to categorized registered trips to several temporal factors such as dates, hours, months, peak and off-peak hour, and type of the day. After all the sets of the data are complete, the trip generation and distribution analysis will be processed according to the workflow shown in Figure 3.1.

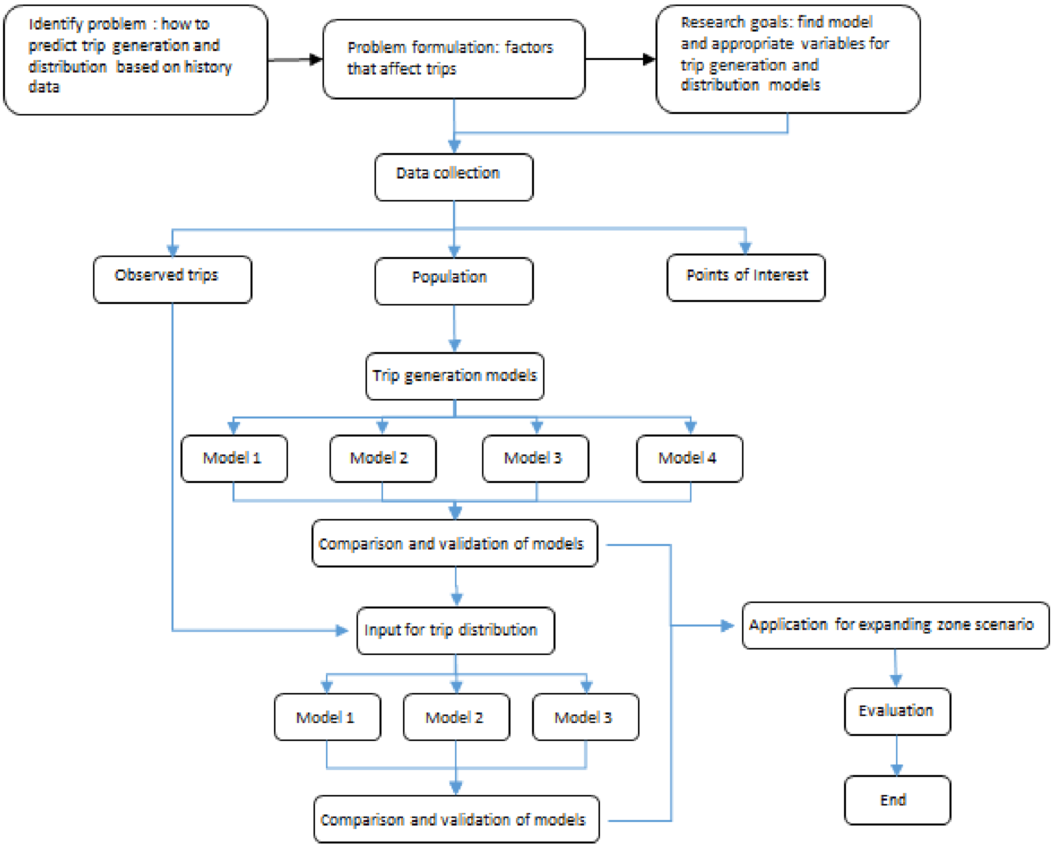


Figure 3.1 Workflow of the research

3.1. Data collection

All data collected are freely accessible from related websites which are DB Call-a-Bike usage data, the population from Germany census data and POIs from Geofabrik websites. The DB Call-a-Bike usage data contains registered trip information from 2014-2016 in every city in Germany. The information of each registered trip is booking id, detail of time for each trip, origin and destination zone for each trip, trip length in minutes, vehicle ID, and rental zone name with its coordinates (latitude and longitude). The latest census data in Germany is freely accessible from 2011 census report from www.zensus2011.de (Statistische Ämter des Bundes und der Länder 2011). This website provides various population data based on gender, age class, marital status, and citizenship. The information is available in Excel file and contains the number of population for each 100m^2 and 1km^2. Each Excel cell has population, latitudinal, and longitudinal information. Therefore, the population information according to the location can be processed through Arc-GIS software.

Gitter_ID_1km	x_mp_1km	y_mp_1km	Einwohner
1kmN2689E4337	4337500	2689500	8
1kmN2689E4341	4341500	2689500	7
1kmN2690E4341	4341500	2690500	3
1kmN2691E4340	4340500	2691500	3
1kmN2691E4341	4341500	2691500	22
1kmN2692E4341	4341500	2692500	20
1kmN2692E4344	4344500	2692500	9
1kmN2693E4340	4340500	2693500	28
1kmN2693E4341	4341500	2693500	8
1kmN2693E4343	4343500	2693500	3
1kmN2694E4340	4340500	2694500	12
1kmN2694E4343	4343500	2694500	12
1kmN2695E4340	4340500	2695500	15
1kmN2695E4343	4343500	2695500	7
1kmN2695E4344	4344500	2695500	3
1kmN2696E4337	4337500	2696500	3
1kmN2696E4339	4339500	2696500	3
1kmN2696E4341	4341500	2696500	4
1kmN2696E4342	4342500	2696500	13

Figure 3.2 Example of census data (Statistische Ämter des Bundes und der Länder 2011)

Geofabrik is a website that extracts, filter, and provides a free geographical database by OpenStreetMap project (Geofabrik, OpenStreetMap). OpenStreetMap is a community project founded in UK where everyone can contribute to provide a geographical database. Most of the database were collected by project members using their GPS devices. Information provided are transport infrastructure (streets, paths, railways), rivers, points of interest, buildings, natural features, land use information, coastlines, and administrative boundaries. Geofabrik converts this data to shapefile or OSM raw data so it can be processed by GIS software such as Arc-GIS. Other data from literature review will also be used, such as the cost sensitivity which is a parameter to obtain accessibility and impedance factors as inputs in trip distribution.

All data mentioned previously have location information through latitudinal and longitudinal coordinates. With GIS software (Arc-GIS), these datasets can be compiled and processed. Therefore, these datasets can be used to analyze the trip generation and the trip distribution of DB Call-a-Bike according to the spatial factors. In this way, each registered trip information can have more detail spatial information.

3.2. Data analysis

3.2.1. Trip generation model

Explanatory variables that will be used in the trip generation model are population, total POIs per zones in several categories, the number of bike stations, temporal factors, and accessibility. These variables then will be added to each registered trip. This thesis will test several multilinear regression models with different explanatory variables. The trip generation will be tested from the simplest model with one explanatory variable until the complex model which includes all the variables that available in the data. The model will produce regression coefficients for each explanatory or independent variable. These coefficients will give information about the relation of each variable to the number of observed trips. These coefficients will also be used to create the trip generation model. The output of each model, which is the number of trips, will be compared to the observed data to know which model can produces the closest number to the observed trips. With regression model, the goodness of fit test to prediction and observed trips will be performed through the number of R-square value.

3.2.2. Trip distribution model

A simple trip distribution model which only consider travel time as impedance factor will be used. Some modification for trip distribution impedance factor will be performed to reach the validity of the model. The iteration for the gravity model will be processed manually using Microsoft Excel. The iteration will be limited to 10 iterations. The 10^{th} iteration is expected to produce the output with the scale near to one. The result of trip distribution model is an Origin-Destination (OD) matrix that contains the number of trips arise for each OD pair. This result will be compared to the observed OD matrix to know which trip distribution model produces the closest number to the real data.

3.2.3. Application of the model

Tested trip generation and trip distribution model with the closest number to observed trips will be chosen as inputs for several scenarios. The scenarios that were performed are expanding bike sharing zones, add more bike stations to current existing bike zones, and increase the number of POIs in existing bike zones. The results are the number of trips arise from the new expansion zones and existing zones.

CHAPTER 4 DATA COLLECTION AND ANALYSIS

4.1. Data collection and description

Three data sources will be used to model trip generation and trip distribution: the trip information and zones location from the DB Call-a-Bike portal (http://data.deutschebahn.com/dataset/data-call-a-bike), the POIs from Geofabrik (http://www.geofabrik.de), and the population data, which can be accessed from the Germany census data of 2011 (https://www.zensus2011.de). The trip information covers the travel time of each trip in minutes, but instead of the actual travel time, it is the full duration when a bicycle is activated (from unlocking to locking) rather than travel time from origin to destination point. Therefore, a valid travel time is important for the trip generation and the distribution model. The valid bicycle travel time can be obtained from google maps via google API, which is a set of application programming interfaces (APIs) developed by Google. It allows communication with Google Services and their integration into other services, including Google maps. The script, written in JAVA, pulls the travel time and distance from Google maps, from point to point, based on the latitudinal and longitudinal coordinates. The distance from the origin to destination are obtained from the shortest bicycle route from google maps.

The data provided from the DB Call-a-Bike portal consist of several sets of CSV data. For this thesis, there are two datasets that will be used, which are hackathon booking and hackathon rental zone. These two datasets contain detailed information as listed in Table 4.1

Table 4.1 List of information from DB Call-a-Bike data

Hackathon booking Call-a-Bike	Hackathon rental zone Call-a-Bike
Booking ID	Station coordinates
Time frame of each trip	Station rental station name
Origin and destination station name	Company
Trip length in minutes	
Vehicle ID	

The trip information is contained in a 6GB CSV archive, therefore it cannot be accessed using Microsoft Excel. With R, a statistical program, this dataset is split according to cities, to reduce the size and simplify the analysis. The temporal information is not well-organized in the original datasets. With R, this temporal information is tidied up. For the spatial information, Arc-GIS, displays the location of every bike stations on the map for further analysis by zones. Figure 4.1 shows the example of the original data from the DB Call-a-Bike portal.

DATE_BOOKING	DATE_FROM	DATE_UNTIL	START_RENTAL_ZONE	START_RENTAL_ZONE_GROUP	END_RENTAL_ZONE	END_RENTAL_ZONE_GROUP	CITY_RENTAL_ZONE
2016-05-05 13:23:49.0000000	2016-05-05 13:23:49.0000000	2016-05-05 14:09:06.0000000	Busparkplatz	Busparkplatz	Kino	Kino	Baden-Baden
2016-05-03 09:27:03.0000000	2016-05-03 09:27:03.0000000	2016-05-03 09:57:38.0000000	Hauptbahnhof	Hauptbahnhof	Kurhaus	Kurhaus	Baden-Baden
2016-05-04 09:33:26.0000000	2016-05-04 09:33:26.0000000	2016-05-04 09:55:45.0000000	Hauptbahnhof	Hauptbahnhof	Kurhaus	Kurhaus	Baden-Baden
2016-05-04 17:59:57.0000000	2016-05-04 17:59:57.0000000	2016-05-04 18:00:33.0000000	Kino	Kino	Kino	Kino	Baden-Baden
2016-05-03 08:29:11.0000000	2016-05-03 08:29:11.0000000	2016-05-03 08:29:32.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-05-03 08:30:14.0000000	2016-05-03 08:30:14.0000000	2016-05-03 08:30:33.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-05-03 08:29:44.0000000	2016-05-03 08:29:44.0000000	2016-05-03 08:29:56.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-05-03 08:30:49.0000000	2016-05-03 08:30:49.0000000	2016-05-03 08:32:29.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-05-03 08:34:00.0000000	2016-05-03 08:34:00.0000000	2016-05-03 09:01:32.0000000	Hauptbahnhof	Hauptbahnhof	Augustaplatz	Augustaplatz	Baden-Baden
2016-05-03 15:24:21.0000000	2016-05-03 15:24:21.0000000	2016-05-03 15:47:21.0000000	Kurhaus	Kurhaus	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-05-04 19:15:21.0000000	2016-05-04 19:15:21.0000000	2016-05-04 19:27:00.0000000	Busparkplatz	Busparkplatz	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-05-04 08:32:20.0000000	2016-05-04 08:32:20.0000000	2016-05-04 08:45:57.0000000	Hauptbahnhof	Hauptbahnhof	Kino	Kino	Baden-Baden
2016-06-30 19:07:37.0000000	2016-06-30 19:07:37.0000000	2016-06-30 19:25:27.0000000	Hauptbahnhof	Hauptbahnhof	Busparkplatz	Busparkplatz	Baden-Baden
2016-05-03 18:09:01.0000000	2016-05-03 18:09:01.0000000	2016-05-03 18:23:16.0000000	Hauptbahnhof	Hauptbahnhof	Busparkplatz	Busparkplatz	Baden-Baden
2016-05-04 15:36:30.0000000	2016-05-04 15:36:30.0000000	2016-05-04 15:56:14.0000000	Kurhaus	Kurhaus	Wohnmobilparkplatz	Wohnmobilparkplatz	Baden-Baden
2016-05-04 16:00:43.0000000	2016-05-04 16:00:43.0000000	2016-05-04 16:18:24.0000000	Kurhaus	Kurhaus	Kino	Kino	Baden-Baden
2016-03-03 15:37:56.0000000	2016-03-03 15:37:56.0000000	2016-03-03 15:42:39.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-04-28 14:18:43.0000000	2016-04-28 14:18:43.0000000	2016-04-28 14:19:31.0000000	Kino	Kino	Kino	Kino	Baden-Baden
2015-06-03 16:14:44.0000000	2015-06-03 16:14:44.0000000	2015-06-03 17:58:10.0000000	Kurhaus	Kurhaus	Kurhaus	Kurhaus	Baden-Baden
2015-07-27 16:33:11.0000000	2015-07-27 16:33:11.0000000	2015-07-27 16:48:00.0000000	Busparkplatz	Busparkplatz	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2015-06-03 16:39:53.0000000	2015-06-03 16:39:53.0000000	2015-06-03 18:17:07.0000000	Wohnmobilparkplatz	Wohnmobilparkplatz	Kurhaus	Kurhaus	Baden-Baden
2015-09-10 18:17:18.0000000	2015-09-10 18:17:18.0000000	2015-09-10 18:52:22.0000000	Wohnmobilparkplatz	Wohnmobilparkplatz	Klosterplatz	Klosterplatz	Baden-Baden
2015-09-10 18:16:40.0000000	2015-09-10 18:16:40.0000000	2015-10-07 15:37:54.0000000	Wohnmobilparkplatz	Wohnmobilparkplatz	Wohnmobilparkplatz	Wohnmobilparkplatz	Baden-Baden
2016-04-14 15:51:43.0000000	2016-04-14 15:51:43.0000000	2016-04-14 16:55:16.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-04-10 08:08:35.0000000	2016-04-10 08:08:35.0000000	2016-04-10 08:23:45.0000000	Hauptbahnhof	Hauptbahnhof	Busparkplatz	Busparkplatz	Baden-Baden
2015-09-10 19:03:01.0000000	2015-09-10 19:03:01.0000000	2015-09-10 19:17:01.0000000	Busparkplatz	Busparkplatz	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-01-12 14:27:35.0000000	2016-01-12 14:27:35.0000000	2016-01-12 17:00:43.0000000	Kurhaus	Kurhaus	Kurhaus	Kurhaus	Baden-Baden
2016-03-24 06:56:09.0000000	2016-03-24 06:56:09.0000000	2016-03-24 17:26:21.0000000	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-03-24 04:27:35.0000000	2016-03-24 04:27:35.0000000	2016-03-24 04:45:40.0000000	Kurhaus	Kurhaus	Hauptbahnhof	Hauptbahnhof	Baden-Baden
2016-04-18 17:00:58.0000000	2016-04-18 17:00:58.0000000	2016-04-18 17:17:02.0000000	Hauptbahnhof	Hauptbahnhof	Busparkplatz	Busparkplatz	Baden-Baden
2015-10-10 13:05:25.0000000	2015-10-10 13:05:25.0000000	2015-10-10 13:13:38.0000000	Kurhaus	Kurhaus	Busparkplatz	Busparkplatz	Baden-Baden
2015-09-10 19:18:32.0000000	2015-09-10 19:18:32.0000000	2015-09-10 19:22:12.0000000	Augustaplatz	Augustaplatz	Augustaplatz	Augustaplatz	Baden-Baden
2016-04-29 17:03:08.0000000	2016-04-29 17:03:08.0000000	2016-04-29 19:31:12.0000000	Busparkplatz	Busparkplatz	Augustaplatz	Augustaplatz	Baden-Baden
2015-07-27 20:29:26.0000000	2015-07-27 20:29:26.0000000	2015-07-27 21:24:40.0000000	Kurhaus	Kurhaus	Kurhaus	Kurhaus	Baden-Baden
2015-07-27 20:33:31.0000000	2015-07-27 20:33:31.0000000	2015-07-27 21:27:47.0000000	Kurhaus	Kurhaus	Kurhaus	Kurhaus	Baden-Baden
2015-07-27 20:34:21.0000000	2015-07-27 20:34:21.0000000	2015-07-27 21:26:58.0000000	Kurhaus	Kurhaus	Kurhaus	Kurhaus	Baden-Baden
2016-01-12 17:00:59.0000000	2016-01-12 17:00:59.0000000	2016-01-12 17:11:35.0000000	Kurhaus	Kurhaus	Klosterplatz	Klosterplatz	Baden-Baden
2016-05-01 15:05:01.0000000	2016-05-01 15:05:01.0000000	2016-05-01 16:44:53.0000000	Wohnmobilparkplatz	Wohnmobilparkplatz	Wohnmobilparkplatz	Wohnmobilparkplatz	Baden-Baden

Figure 4.1 Original trip information file from DB Call-a-Bike

Not all rental point coordinates are available in the hackathon rental zone dataset. Therefore, some of the rental points that do not have coordinates information are omitted. Only ten cities with more than five bike stations or rental points are chosen for this research. For cities without station-based systems such as Munich, some areas of Frankfurt and Cologne, the hackathon rental zone dataset provides information on the coordinates of the bike sharing central activity zone. The website does not give further explanation of the zone radius for each set coordinates. Therefore, three cities without station-based systems can still be analyzed, as well as city with station-based systems. The example of DB Call-a-Bike “central activity zone location” that found in Frauenlobstrasse Munich shown as in Figure 4.2. There are no bike stations, only bicycles in the sidewalk of Frauenlobstrasse.

Figure 4.2 DB Call-a-Bike in Munich

The DB Call-a-Bike dataset provides all information from January 2014 to July 2016. Some cities that joined Call-a-Bike later than January 2014, for example, Darmstadt, have trip information since April 2014.

The travel length minutes from DB Call-a-Bike data was expected to be the travel time in minutes from origin to destination, but as mentioned previously, this number is the total time when the bicycle was activated. Some data shows odd numbers. For example, one OD pair with 1.2km distance, shows a travel time of 60 minutes or higher when the actual normal travel time is 8-10 minutes. This occurs very often, presumably because users do not always take direct route or they leave the bicycles unlocked when they visit a location. Figure 4.3 is an example comparison of Google maps travel time with DB Call-a-Bike travel time in Frankfurt.

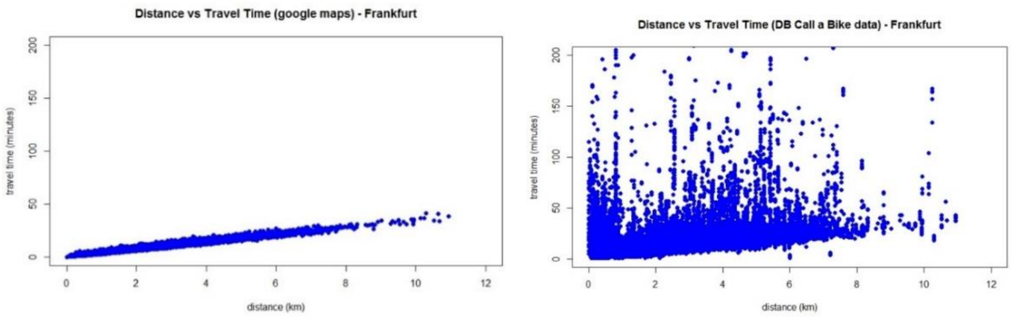


Figure 4.3 Comparison of distance vs travel time from Google maps and Call-a-Bike data

Figure 4.3 shows a lot of trips from DB Call-a-Bike data base that deviate from the normal travel time for a bicycle trip. The information necessary for a trip generation and a trip distribution model is the normal travel time from origin to destination for every trip information. To avoid the odd trip length minutes data, the travel time data was filtered. To filtered the travel time, the data is sorted based on each OD pair, after that, the travel time means, Q1 (first quartile), Q3 (third quartile), maximum and minimum values were calculated. The travel time that falls between Q1 and Q3 range is kept, but the travel time value which is less than Q1 will be substituted with the minimum travel time while the one that greater than Q3 will be substituted with Q3 value. Table 4.2 shows the original number of observed trips from DB Call-a-Bike portal and the number of trips after elimination. This elimination is done because some of the trips do not have detail useful information for the analysis (travel time, date, or OD pair).

Table 4.2 List of time range of observation

City	Observation time range	Number of observed trips for analysis	Number of all trips from DB Call-a-Bike data
Berlin	30 months	791,596	985,253
Darmstadt	27 months	156,615	156,615
Frankfurt	30 months	686,041	910,195
Hamburg	30 months	5,113,524	6,348,447
Kassel	30 months	460,038	460,038
Cologne	26 months	216,946	585,856
Marburg	27 months	97,227	97,227
Munich	26 months	203,441	203,441
Rüsselsheim	27 months	36,506	36,506

The spatial information from Geofabrik is available in a shapefile that consist of POIs information. The POIs from Geofabrik is divided into 27 categories which are school, college, kindergarten, library, bakery, bar, beverage, night club, theatre, mall, park, department store, mall, food& recreation, sports, beer garden, clothes, butcher, café, restaurant, zoo, cinema, fast food, museum, playground, sport, supermarket, and picnic area. To simplify the trip generation model, these categories will be aggregated into four categories which are education, shopping, entertainment, and food and recreation.

Germany census 2011 data provided the population data for each 1km^2 for all Germany area. In some area, there is minus population according to the original census data. For the analysis, these values are substituted to zero, otherwise, it will affect the linear regression analysis. The population data contains the coordinates and inhabitants for every $1km^2$.

4.2. Temporal Analysis

The bike sharing trip frequency for each city was aggregated per month to see the travel pattern along the year. Every city has the same pattern. April until June are the months with the highest frequency of DB Call-a-Bike usage due to the summer season. The lowest is from November until February, this shows that winter season affects the number of bike trips. The number of trips falls from 14,3% in June to 9,7% in July. Figure 4.5 and 4.6 show the fluctuation of the trips number in percentage and the absolute number along the year from 2014-2016.

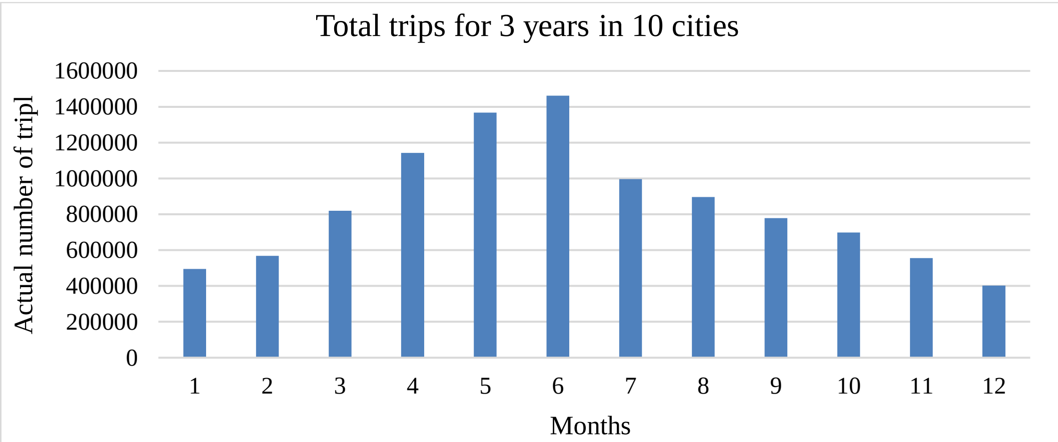


Figure 4.4 Total trips for 3 years in 10 cities in absolute number

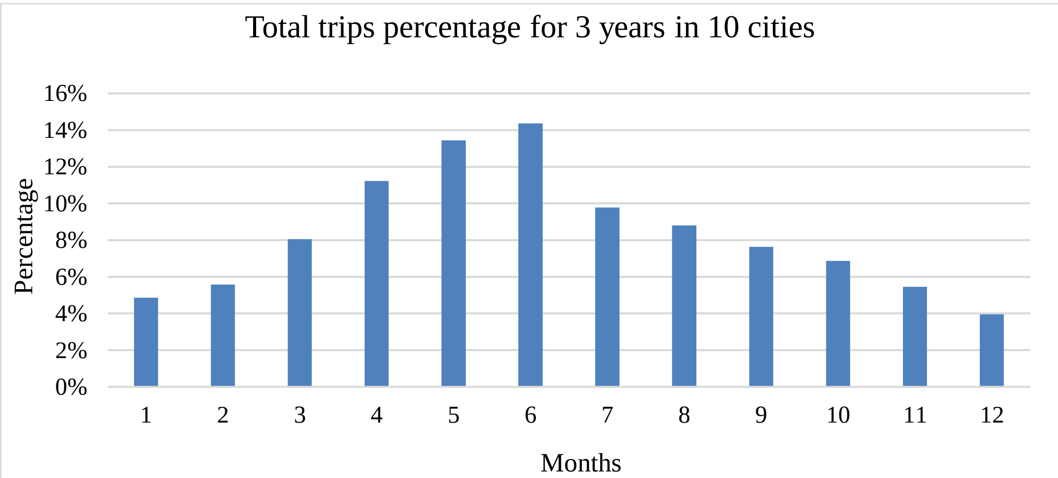


Figure 4.5 Total trips in percentage for 3 years in 10 cities

Table 4.3 shows the relation of the season to the number of trips. This analysis was performed through a regression linear analysis. Each of registered trip has a temporal detail of season that inform to which season each trip belongs to. Season 1 is from January-April, Season 2 from May-August and Season 3 from September-December. The equation of the regression linear is shown as below:

Equation 4.1 Regression linear equation for bike sharing trips season analysis

$$Y = \text{season } 1 + \text{season } 2 + \text{season } 3$$

Equation 4.1 shows the regression equation for bike sharing trips season analysis where Y is the total number of trips in 3 years for each city. The regression linear analysis was performed with R. The season was read as a factor in R to know the influence of season to the number of trips generated. All cities have the same pattern. Season 2 which is from May-August has the highest relation to the number of trips followed by Season 1 (Jan-April) and then Season 3 (Sept-Dec). Table 4.3 shows the relation of each season to number of trips and it is relevant to Figure 4.5 and 4.6.

Table 4.3 Relation of season to the number of trips generated

Variable	Estimate	P value	Significant codes
Season 1	9.66553	<2E-16	***
Season 2	11.37905	<2E-16	***
Season 3	8.81211	<2E-16	***

`Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1`

The number of trips along the day was also analyzed. Figure 4.7 and 4.8 show the comparison of the average usage for 3 years. The charts are divided into weekday and weekend to see the trip pattern difference.

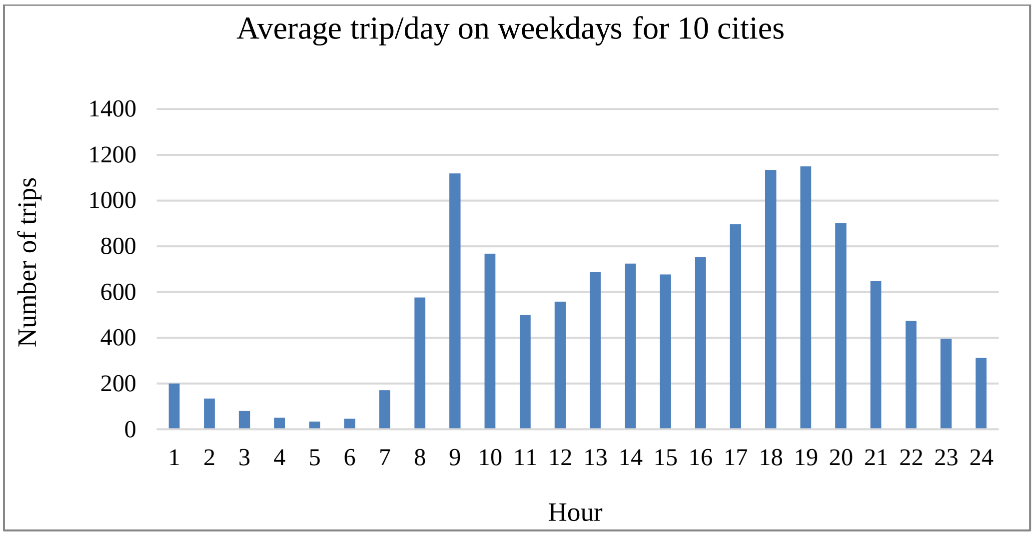


Figure 4.6 average trip/day on weekdays

Figure 4.7 shows that on weekdays, the highest bike usage is from 9.00 - 10.00 AM. Based on this figure, it is assumed that users primarily use DB Call-a-Bike in rush hour to commute to work. After 9.00 AM, the number keeps decreasing until 12.00 AM and then slightly increasing until 14.00 PM. This probably caused by people who also used bicycles in their break time and probably get a quick lunch nearby their job or education location. The number then increasing again at 18.00-19.00 PM. Based on this figure, it is assumed that users primarily use DB Call-a-Bike to commute from work.

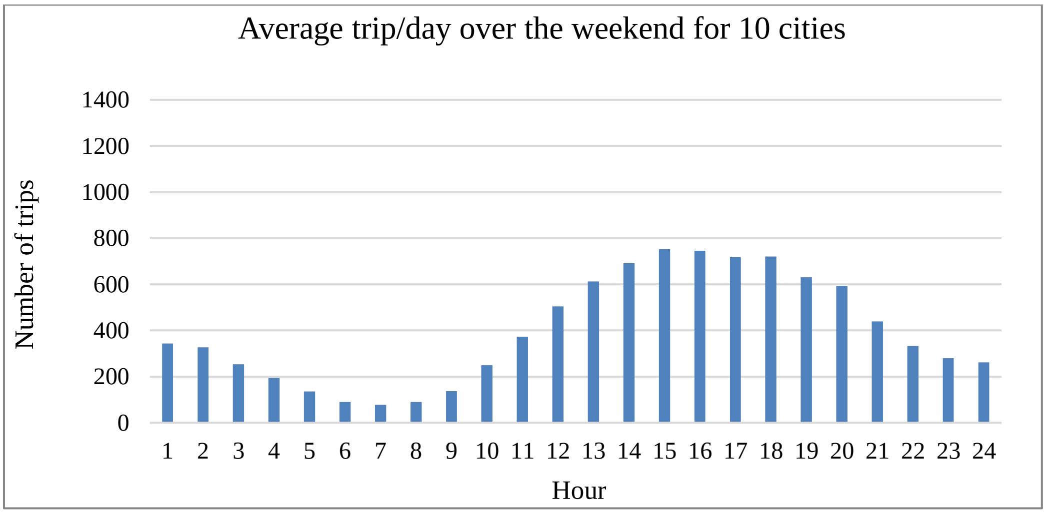


Figure 4.7 Average trip/day over the weekend for 10 cities.

On the other hand, the bike sharing pattern usage over the weekend is different from weekdays. Bike sharing usage number keeps increasing from 8.00AM until 16.00PM, this happened probably because on weekend users prefer a leisurely time when using bike sharing rather than the normal rush hour during weekdays. The bike sharing usage activity number keeps increasing for 7-8 hours until 16.00 PM, decreases until 24.00PM, and then increasing again until 02.00AM. This could be happened presumably by people who use bike sharing to go back from their leisure activity at night. Probably because the public transportation service on weekend served only until limited time. The trip number along the week can also be obtained. During weekdays, the usage is quite constant, around 14-15% from Monday to Friday, but then decreasing to11-13% over the weekend.

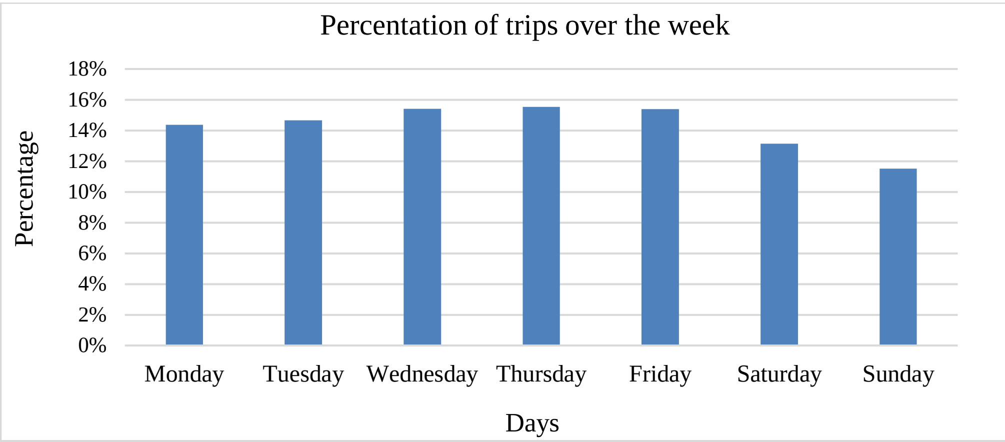


Figure 4.8 Trips fluctuation through the week

Figure 4.7 until 4.9 show that temporal variable is significantly influences the number of trips generated. Another way to measure temporal relation is by comparing the coefficient of multilinear regression. Each of registered trip has a temporal detail of season as mentioned previously, day type (weekday or weekend) and hours' type that divided to 5 types which are listed as in Table 4.4.

Table 4.4 Hours' type categories

Hours' type	Time span
On 1	06:00 – 09:00
Off 2	09:00 – 16:00
On 2	16:00 – 19:00
Off 3	19:00 – 06:00

The regression linear equation for testing the temporal factors with number of trips is shown in equation 4.2.

Equation 4.2 Regression linear equation for temporal factors analysis towards number of trips

$$Y = weekday + weekend + on\ 1 + on2 + off\ 2 + off3$$

Table 4.5 as result of equation 4.2. shows a significant difference between weekday and weekend usage. Weekday attracts way higher trips generated rather than the weekend. On-peak hour 2, which is evening peak hour, attracts more trips rather than morning peak hour. This probably happened because people have limited time to reach their commuting destination (education and working place). Therefore, users prefer to use another alternative transportation that faster than a bicycle. But in the evening peak hour, users do not have limited time to reach their destination. Therefore, the number of people uses bike sharing is higher than the morning peak hour. Among all the hour-type categories, the off-peak hour attracts higher trips generated by DB Call-a-Bike. Probably it caused by the number of people that do not have a time limit to reach their activity destination is higher than the number of people that use DB Call-a-Bike in the morning peak hour. Besides that, the time span of the off-peak hour is longer than on-peak hour, this situation leads to the higher number of trips on off-peak hour rather than on-peak hour.

Table 4.5 Temporal factors relation with number of trips according to the city

City	Weekday	Weekend	On 1	On 2	Off 2	Off 3	R square
Frankfurt	776.2	-185.8	314.6	835.4	1041.9	304.9	0.1748
Hamburg	3693.2	-687.8	1186.0	5671.1	6131.1	1691.4	0.2487
Kassel	965.76	41.83	17.83	277.53	1296.30	153.04	0.2219
Marburg	647.75	-91.16	33.57	405.11	1118.71	289.01	0.327
Stuttgart	878.7	-163.2	310.5	839.7	1189.6	211.9	0.1779
Darmstadt	777.2	-212.7	263.8	393.3	1471.7	186.3	0.4162
Munich	857.3	-166.3	195.5	776.1	1608.7	81.3	0.46
Koln	1516.1	-153.4	371.5	1492.9	2565.4	837.2	0.7644
Berlin	834.9	-375.4	520.4	1467.9	2125.2	480.2	0.3309
Rüsselsheim	1122.6	-838.2	918.5	818.7	3355.7	280.7	0.7353

Table 4.6 Overall temporal factors relation with number of trips

Variables	Estimate	P value	Significant codes
Weekday	1591.9	2.46E-13	***
Weekend	-297.2	1.71E-01	
On 1	519.3	6.39E-02	.
On 2	2113.0	2.38E-14	***
Off 2	2677.0	< 2E-16	***
Off 3	657.8	1.78E-02	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The range of the travel time and distance of DB Call-a-Bike users can also be obtained based on the start and end of the trip zones coordinates for each trip. Figure 4.10 and 4.11 show details of DB Call-a-Bike usage frequency according to the distance and travel time. Around 28% people use DB Call-a-Bike to travel in the range 5-10 minutes and more than 45% travel in the range 0-2km. Only 2% of people use it to travel for 30-35 minutes and around 13% use it to travel from 35-1000 minutes. The distance used in this analysis was calculated based on Google maps via Google API. The bicycle speed is set to 16km/h, which is the normal bicycle speed that already set originally by Google maps, therefore 8 minutes travelling by bike could reach 2km distance. This proved that the travel time in Figure 4.10 and distance charts in Figure 4.11 are strongly related.

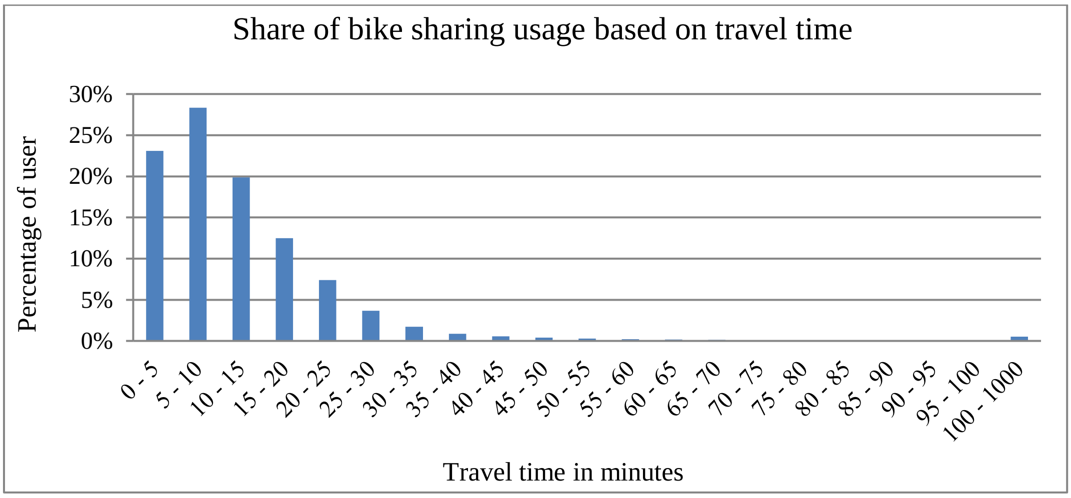


Figure 4.9 Share of bike sharing usage based on travel time

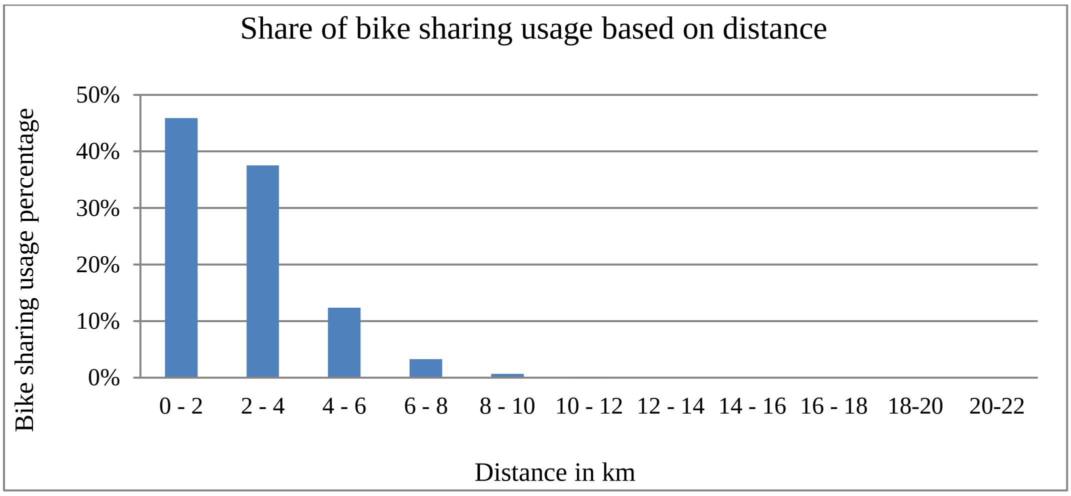


Figure 4.10 Share of bike sharing usage based on distance

Figure 4.10 and 4.11 show that most users tend to do short trips for less than 20 minutes. Most of the people use DB Call-a-Bike for a trip with travel time within 5-10 minutes range. The number of the percentage of bike sharing usage by distance drop significantly within 4km distance from 37% to only 12% for 4-6km trips.

4.3. Trip generation

The multilinear regression for the trip generation model consider some explanatory variables that assumed to have strong relationships with the number of trips generated as mentioned in Chapter 2. Explanatory variables that will be added to each trip information are temporal information (off- and on-peak hour, season and day type), population, accessibility, bike station and POIs. The population and bike station will be added as an attribute to each registered trip according to the origin zone while POIs and accessibility will be added according to the destination zone of each trip.

4.3.1. Bike sharing system zones

The synthetic bike zones were created by dividing each city into several square zones of $4km^2$ area ($2km^2$x$2km^2$). Every city has a different distance from one bike station to other bike station. After some investigation through Arc-GIS, seven cities (all cities for this study except Frankfurt, Berlin, and Hamburg) have 1

- 2 bike stations for every 2km. Therefore, it was decided to have synthetic zones of 4km^2 for each city. Dividing the city into several zones was done with Arc-GIS by using fishnet tool. The study area only focusing in zones that have bike stations. Coordinates of bike zones were used as an input to determine the boundaries of the zones. The fishnet tool will automatically create zones in a rectangular shape. Zones that do not have any bike stations will be removed from the analysis. Figure 4.12 is the example of bike zones in Frankfurt produced by Arc-GIS.

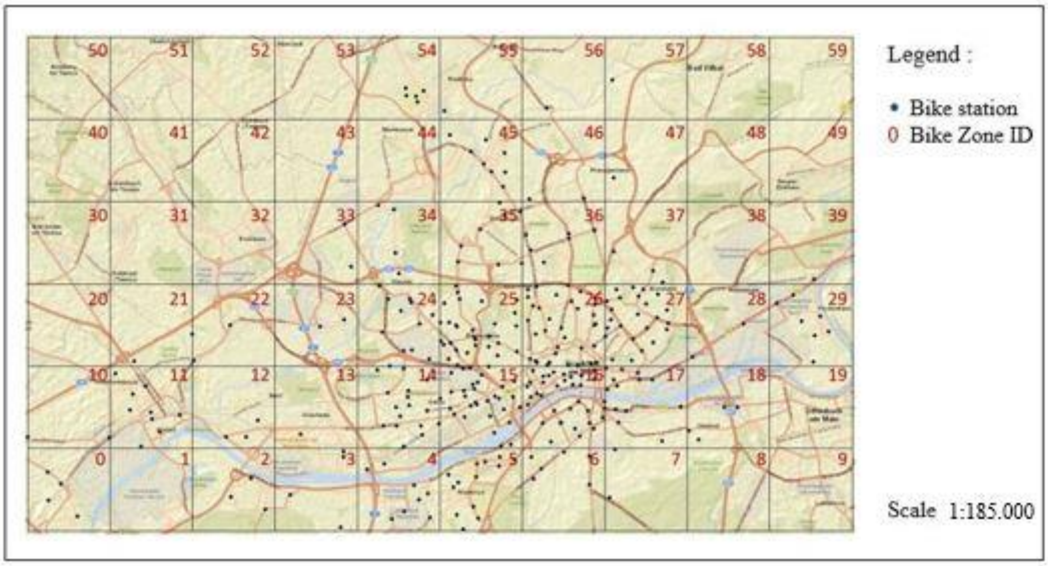


Figure 4.11 Frankfurt bike zones

Each city will have different number of bike zones depends on the coverage area of DB Call-a-Bike. Table 4.7 shows the list of the total number of bike zones that will be analyzed in each city.

Table 4.7 List of total number of bike zone for analysis

City	Number of zones	Number of bike stations
Berlin	25	114
Darmstadt	7	25
Frankfurt	38	252
Hamburg	52	102
Kassel	18	34
Köln	5	80
Marburg	5	15
Munich	9	85
Rüsselsheim	1	8
Stuttgart	17	29

4.3.2. Temporal infromation

According to the season, the data was divided into 3 categories. They are Season 1 from January-April, Season 2 from May-August and Season 3 from September-December. According to the hour, the trips are divided into four categories as mentioned in section 4.2, which are on-peak 1 from 06:00-09:00, off-peak 2 from 09:00-16:00, afternoon peak-hour from 16:00-19:00 and off-peak 3 from 19:00-06:00. The other type of the temporal information is the day type that divided into 2 categories which are weekday and weekend trips.

4.3.3. Points of Interest (POIs)

From the Geofabrik, POIs are divided into 27 categories which are kindergarten, school, college, university, library, bakery, beverages, bar, food court, night club, pub, theater, biergarten, butcher, café, mall, restaurant, zoo, and cinema. Each of this point has latitudinal and longitudinal coordinates. To simplify the model, the variables were aggregated to four categories which are education, shopping, entertainment, and food and recreation.

After zones were created in Arc-GIS, other variables which are population, POIs and bike stations as explained in section 4.1 were compiled with each trip information and the total number of each variable was calculated according to the synthetic zones. The data processed from Arc-GIS will produces the information of zones attribute which consists of the total number of bike stations and the total number of POIs according to the categories and population. Table 4.8 shows the example of zones attributed that can be produced.

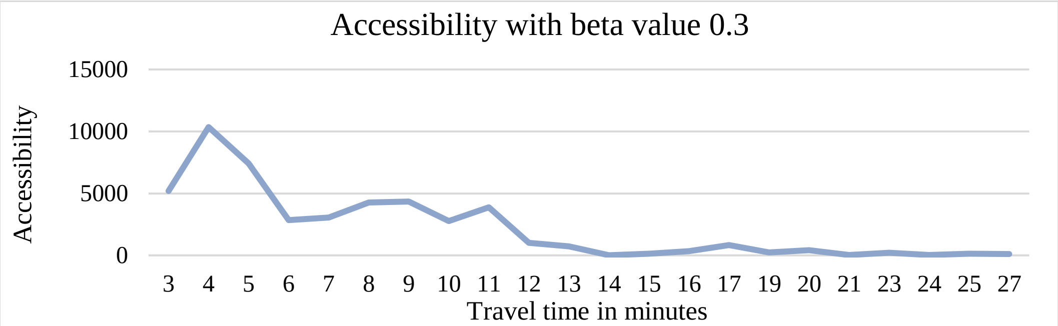
Table 4.8 Example of zone attribute in Kassel

Bike zone number	Bike stations	Education	Shopping	Entertainment	Food & recreation	Population
0	2	0	5	0	5	5027
1	3	0	14	2	15	14437
2	1	0	34	0	16	1702
3	2	1	9	3	5	494
4	2	2	10	3	13	7832
5	2	2	2	0	6	5587
7	3	2	9	4	21	8360
17	2	0	4	2	13	11212
19	1	1	3	1	7	5782

4.3.4. Accessibility

Other zones attribute is the accessibility. Accessibility was calculated according to the formula that already explained in Chapter 2. The accessibility value for the zones was calculated after the zone attribute was produced because the necessary information, which is population according to the zone, is important to calculate the accessibility. The travel time used to calculate accessibility is from each centroid to each other centroid of the zone with the shortest bicycle route from Google maps.

According to Iacono et al. (2008) cost sensitivity parameter (ß) for the bicycle depends on the trip purpose. For bike trips ß parameter, it ranges from 0.12 for school-related trips to over 0.5 for shopping trips. It was decided to take the mid value of 0.3 for the calculation considering that the analysis is for all trip purposes. Aside from that, after comparing the accessibility value with the changes in beta value from 0.1 until 0.5, Figure 4.13 shows that the higher the beta value will lead to giving the trips with higher travel time less weight. While in fact, around 17% of the total trips from the database are trips with more than 20 minutes of travel time, therefore beta with 0.3 value was chosen to avoid underprediction of trips with more than 20 minutes travel time.



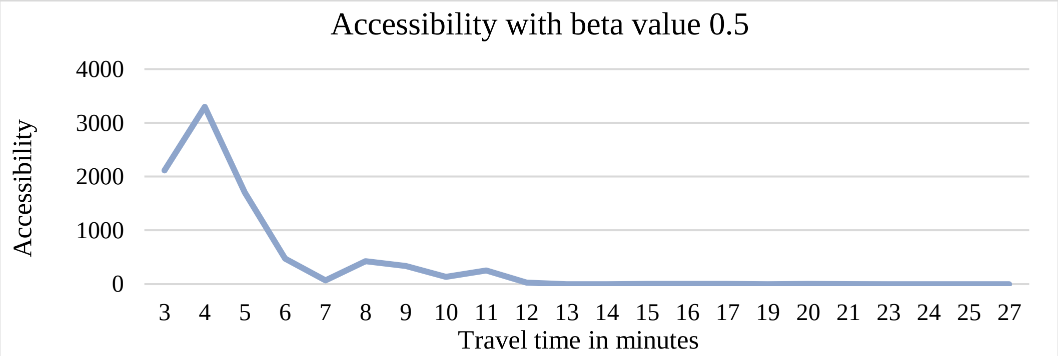


Figure 4.12 comparison of ß 0.3 and 0.5

Each registered trip has the origin-destination information and the zones attribute information from Arc-GIS. The zone attribute created by Arc-GIS can show which bike station belongs to which zone. In this way, the origin and destination bicycle zone for each trip can be added. After that with R, the dataset from DB Call-a-Bike will be merged with the zone attribute data from Arc-GIS according to the origin-destination zones. Therefore, each trip information will have the spatial attribute on it. After that, this dataset will be processed using multilinear regression to find the trip generation model.

4.3.5. Trip generation model results

The trip generation model regression analysis was tested from the simplest model using only one variable (population) until the complex model, which includes temporal and spatial factors as independent variables. Table 4.9 shows the equation of each model that were tested in this study

Table 4.9 Trip generation models and equation

Model	Equation
1. Population	Number of trips = a + b1*Population
2. Population and bike stations	Number of trips = a + b1*Population + b2*Bikezone
3. Population, bikezone, POIs, accessibility	Number of trips = a + b1*Population + b2*Bikestations+ b3*Education + b4*Shopping + b5*Entertainment + b6*Food and recreation + b7*Accessibility
4. POIs	Number of trips = a + b1*Education + b2*Shopping + b3*Entertainment + b4*Food and recreation
5. Population, bike stations, POIS, accessibility, temporal factors	Number of trips = a + b1*Population + b2*Bikestations +b3*Education + b4*Shopping + b5*Entertainment + b6*Food and recreation + b7*Accessibility + b8*Season + b9*Hours'type + b10*Daytype

In the first model, the population in this equation is the population of bike zone origin for each registered trip. For every analysis concerning zone attribute, Rüsselsheim cannot be analyzed because it has only one synthetic zone, but for overall results, Rüsselsheim was included. Table 4.10 shows the relation of population to the number of trips.

Table 4.10 Relationship of population with number of trips

City	Intercept	Population estimate	T value population	P value population	R square	Population
Frankfurt	-651	0.0785	11.400	< 2E-16	0.1391	580,816
Hamburg	541.0214	0.1848	7.588	6.81E-14	0.0488	853,503
Kassel	-597.7282	0.1467	9.602	< 2E-16	0.1470	178,169
Marburg	-191.8798	0.0719	4.327	2.78E-05	0.1130	58,607
Stuttgart	-294.4362	0.0637	5.471	7.66E-08	0.0657	276,168
Darmstadt	94.1018	0.0523	4.690	4.99E-06	0.0968	87,886
Munich	1709.4304	-0.0170	-1.340	17.972E-02	0.0097	328,956
Cologne	2590	-0.0105	0.0489	1.56E-02	0.0489	375,054
Berlin	-1290	0.0760	13.010	< 2E-16	0.1976	685,150
Total 10 cities	1150	0.0345	6.847	8.61E-12	0.0108	3,433,981

Almost all cities except Munich and Cologne have positive population coefficients, indicating that higher population will increase the number of trips generated. The R-square value indicates whether the model explains the variable with the number of trips. The higher the R-square value, the more that this model explains the variable on the equation around its mean. The R-square value for the first model is quite low which is in the range 1-19% for almost all the cities. This shows that the model is unlikely to explain the relationship between population and number of trips.

The P value used to determine whether the null hypothesis can be rejected is usually set at 5% or lower. Significance level could be made more stringent at 1% to show a highly statistically significant effect. This means less than one in a hundred chance of being wrong. This demonstrates that there is a relationship between dependent and independent or explanatory variable (Cramer, Howitt 2004). Each of the cities except Munich had a P value less than 5% for population. This means the population has strong relationships with the number of trips generated via DB Call-a-Bike, supporting the hypothesis that higher population density leads to a higher number of trips generated.

The second trip generation model tested used two variables, population and number of bike stations. This includes the number of bike stations as a variable in the model which leads to negative coefficient values for the population of some cities, indicating that the higher population reduces the trips generated while the higher number of bike stations increases the trips generated. This case contradicts the normal hypothesis of the trip generation where higher density population generates more trips. For the second model, the R-square value is higher than the first model. For 5 cities, the R-square values are higher than 30% and the rest is in the range of 5-12%. This means that the second model plausibly explains the number of trips generated rather than the first model. Table 4.11 shows the result of the second model for each city

Table 4.11 Regression value of number of trips towards population and bikezone

City	Intercept	Population	Bike stations	R square
Frankfurt	-244.91	-0.0356	167	0.3717
P value	6.54E-02	6.34E-05	< 2E-16	
Hamburg	-3060	0.0619	1250	0.3594
P value	6.35E-09	2.84E-03	< 2E-16	
Kassel	-333.40	0.0131	323	0.3207
P value	3.42E-02	4.59E-01	<2E-16	
Marburg	246.09	-0.0828	342.95	0.3008
P value	2.409E-01	4.65E-03	4.01E-09	
Stuttgart	-4.53	0.0096	403	0.4531
P value	1.75E-02	3.07E-01	<2E-16	
Darmstadt	-97.62	0.0135	134.85	0.1271
P value	5.93E-01	4.574E-01	8.52E-03	
Munich	1525.79	-0.0322	64.87	0.0593
P value	1.39E-03	1.605E-02	2.13E-03	
Cologne	1.63	-0.0080	18.2	0.0799
P value	7.17E-03	7.03E-02	5.058E-02	
Berlin	-5.25	0.0031	229	0.3106
P value	9.78E-03	7.174E-01	< 2E-16	

Some of the P-value of population for the second model is higher than 5% showing that the relation of the population to the number of bike trips is low. In contrast, all P-values for bicycle zones are less than 5% which clearly demonstrates that more bike stations will attract more bike sharing trips. Table 4.12 shows the regression result for all ten cities. The P-value for the population is higher than 5%

and P-value for bike station is less than 5%, indicating that only bike stations have a strong relationship with the number of trips generated.

Table 4.12 Comparison of population and bike stations

Variables	Estimate	P value	Significant codes
Population	1.70E-03	0.759	
Bike stations	1.34E+02	< 2E-16	***

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Multiple R-squared:  0.04832,  Adjusted R-squared:  0.04788
```

The third model includes population, accessibility, bike stations, education, shopping, entertainment, and food and recreation as explanatory variables. The population and bike stations are according to the origin bike zone of each registered trip while the rest of the variables are according to the destination bike zone. These POIs (education, shopping, entertainment, food and recreation) are personal trips attraction factors. Table 4.13 shows the relation of each variable to the number of trips.

Table 4.13 Relationship of the third model variables to the number of trips

City	Population	Accessibility	Bike stations	Education	Shopping	Entertainment	Food and recreation	R square
Frankfurt	0.02735	-0.00224	119.4	-59.51	-21.39	-22.42	20.98	0.469
P value	2.16E-01	4.79E-01	2.52E-06	4.2E-03	3.88E-01	0.00029	1.3E-04	
Hamburg	0.00009	0.00004	0.2282	0.04715	0.00906	0.02091	-0.00599	0.1034
P value	< 2E-16	2.9E-10	< 2E-16	2.07E-09	2.9E-10	9.61E-11	6.48E-07	
Kassel	4.75E-03	0.0006639	-0.05615	0.9089	-0.0125	12.119	0.03477	0.2545
P value	4.45E-01	< 2E-16	< 2E-16	< 2E-16	2.42E-05	0.58794	2.95E-13	
Marburg	-0.00954	0.00198	1.621.207	-6.62033	-591.377	3.71656	325.705	0.1516
P value	< 2E-16	< 2E-16	< 2E-16	3.18E-15	2.57E-12	< 2E-16	2.72E-08	
Stuttgart	0.00004	0.00002	0.05805	-0.89	-0.227	-0.2979	0.3154	0.3079
P value	<2E-16	<2E-16	<2E-16	<2E-16	<2E-16	<2E-16	<2E-16	
Darmstadt	0.00007	0.00012	-0.01439	-0.179	0.06598	-0.06286	0.04535	0.13
P value	< 2E-16	3.33E-01	8.38E-01	8.91E-01	7.87E-01	0.00076	4.52E-01	
Munich	0.00384	-0.00543	-2.992	-0.65618	-554.169	852.037	15.123	0.1753
P value	2.66E-06	2.08E-02	7.12E-02	7.31E-01	9.89E-09	0.0001	3.0E-05	
Koln	-0.00029	3.022e-03	-0.8612	-19.6	-30.9	-2.903e+01	27.92	0.2283
P value	8.19E-02	6.41E-02	1.63E-02	1.73E-02	6.9E-09	6.79E-06	2.23E-07	
Berlin	1,57E-04	0.00002	0.05628	-0.0241	-0.011	0.00281	0.00747	0.1338
P value	9.40E-01	< 2E-16	<2E-16	< 2E-16	1.04E-08	0.0265	2.02E-13	

The third model shows different relationships between variables and the number of observed trips, depending on the city. Not every variable had a strong relationship with the number of trips. Most of the cities had a strong and positive relationship of population with the number of observed trips except for Marburg. Six out of ten cities had positive relationship for population with P-values less than 5% and varies from < 2.2E-16 to 1.358E-02, indicating strong relationship between the number of observed trips and population except for Frankfurt, Kassel, and Berlin. Seven out of ten cities show that accessibility had a strong relationship and increases the number of trips generated. The P-values for accessibility varies from <2E-16 to 2.0815E-02 except for Frankfurt and Darmstadt. For Kassel, Darmstadt, Munich and Cologne, the number of bike stations had a negative relationship, but the P-value for this variable was also high from 7.12603E-02 to 8.3E-01, indicating that the number of bike stations does not have high relationship, except for Kassel and Cologne.

The POIs relationship to the number of trips depends on the city. Some of the POIs decrease the number of trips, for example entertainment variable in Frankfurt, Darmstadt, and Cologne. But for all cities except Hamburg, the higher number of food and recreation increases the number of trips. The P-values for food and recreation in every city were also less than 5% except for Darmstadt. This demonstrates that food and recreation can be assumed as a strong factor to generates the number of bike sharing trips in most of the city. The R-square value for this model falls within 10-46%, this shows that adding more variables will make the trip generation model more plausible to explain the number of trips. Table 4.14 shows the relation of each variable to the number of trips of all ten cities.

Table 4.14 Relation of variables with number of trips

Variable	Estimate	P-value	Significant codes
Population	1.709E-04	<2E-16	***
Accessibility	1.0195E-02	<2E-16	***
Bike station	3.385E-03	<2E-16	***
Education	-4.512E-05	<2E-16	***
Shopping	1.472E-04	<2E-16	***
Entertainment	-2.758E-02	<2E-16	***
Food and recreation	6.716E-02	<2E-16	***

The P-values' variables of the model for ten cities is less than 5% which shows a high relationship between variables and the number of trips. This can be happened because not every city has the same number of observed trips. Hamburg has the highest number of observed trip which is around 5 million observed trips while other cities only have around 100,000-600,000 observed trips. This situation leads to all P-values' variables become less than 5% according to Hamburg data in Table 4.13. Only education and entertainment had negative coefficients, shows that these variables reduce the number of trips generated. The rest of the variables increase the number of trips, and according to the P-value, these factors had a strong relationship to the number of trips generated. It can be concluded that shopping, food and recreation, population, accessibility, and number of bike stations are factors that attract bike sharing trips.

The fourth trip generation model tested is to investigate the relation of various POIs as trip attractions with the number of observed trips. The purpose of this model is to find out which trip attraction factors that highly attract the users. Table 4.15 shows the relationship of each POIs.

Table 4.15 Relationship of POIs to the number of trips

City	Education	Shopping	Entertainment	Food and recreation	R square
Frankurt	0.0884	0.0000258	0.0215	-0.0620	0.0835
P value	< 2E-16	1.90E-11	< 2E-16	< 2E-16	
Hamburg	0.1437	0.0509	-0.0032	-0.0007	0.0897
P value	<2E-16	<2E-16	3.11E-01	5.43E-01	
Kassel	1.3070	-0.0492	-0.1567	0.0202	0.2183
P value	<2E-16	<2E-16	<2E-16	<2E-16	
Marburg	2.909	2.429	0.5585	-2.1998	0.1188
P value	< 2E-16	1.08E-07	4.2E-04	4.76E-10	
Stuttgart	-0.7948	-0.3298	-0.4627	0.4532	0.2982
P value	<2E-16	<2E-16	<2E-16	<2E-16	
Darmstadt	12.983	10.669	-15.168	9.162	0.3017
P value	< 2E-16	< 2E-16	1.5E-09	< 2E-16	
Munich	-3.9401	-1.9168	1.8065	3.5827	0.1584
P value	1.29E-05	6.94E-02	1.56E-11	5.98E-02	
Cologne	-8.056	-21.854	-18.274	18.927	0.2265
P value	2.07E-02	3.52E-13	1.85E-08	6.83E-11	
Berlin	0.0000765	-0.0113	0.0238	-0.0430	0.1241
P value	<2E-16	<2E-16	<2E-16	<2E-16	

Used only personal trips attractions as predictors gave different results depending on the city. Some city had food and recreation as an attraction factor (Kassel, Stuttgart, Darmstadt, and Cologne) while other city had the entertainment to attract bike sharing users (Frankfurt, Marburg, Munich, and Berlin). Table 4.16 shows the comparison of POIs that attract bike sharing trips.

Table 4.16 Comparison of factors that attract DB Call-a-Bike users.

Variable	Estimate	P-value	Significant codes
Education	8.943E-06	1.13E-02	*
Shopping	7.361E-05	<2E-16	***
Entertainment	-1.485E-02	<2E-16	***
Food & recreation	7.550E-02	<2E-16	***

Significant codes 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Not like the third model, the fourth model had education as one of the factor that increases the number of bike sharing trips. The third model underestimates education as POI probably because the number of this POI is lower than other factors. The P-value for education was also below 5%, indicating that education can be accepted as factor that increases the number of trips generated.

The fifth model tested used population, bike stations, POIs, accessibility, and temporal factors as predictors. Temporal factors included in the model were Season 1 (January-April), Season 2 (May-August), Season 3 (September-December), weekday, weekend, peak hour morning (On 1), peak hour evening (On 2), off-peak hour day (off 2) and off-peak hour evening (off 3). To test this model, the temporal variables were read as a factor in R to get the relation of temporal factors with the number of observed trips. Table 4.17 shows the results of the fifth model.

Table 4.17 Relation of factors to number of trips generated

Variable	Frankfurt	Hamburg	Kassel	Stuttgart	Darmstadt	Munich	Cologne	Berlin	Marburg
Population	-0.000034	0.000110	0.00000980	0.0000467	-0.000184	-0.00120	-0.000317	-0.00000038	-0.01044
p value	< 2E-16	< 2E-16	1.476E-01	< 2E-16	1.098E-01	1.59E-01	7.806E-02	8.5632E-01	< 2E-16
Bike station	0.08954	0.2216	-0.04573	0.06496	-0.3273	-0.00157	-1.002	0.04976	17.65
p value	<2E-16	< 2E-16	4.4E-12	< 2E-16	3.275E-01	1.42E-02	9.16E-03	< 2E-16	< 2E-16
Education	-0.07693	0.06361	0.5999	-0.8903	-0.3234	-9.898		0.00000985	
p value	<2E-16	2.64E-15	< 2E-16	< 2E-16	49.96E-02	< 2E-16		1.4E-07	
Entertainment	0.01728	0.03082	-0.09971	-0.301	-0.1724	-0.7896	-8.044	0.006481	2.404
p value	<2E-16	< 2E-16	< 2E-16	< 2E-16	6.801E-02	4.42E-02	6.712E-01	< 2E-16	< 2E-16
Accessibility	0.0000182	0.0000437	0.000600	0.0000105	0.000755	-0.00320	-0.001375	-0.0167	-0.000446
p value	<2E-16	< 2E-16	< 2E-16	5.1E-03	3.051E-01	1.55E-01	1.722E-01	< 2E-16	4.718E-03
Food and recreation	-0.0319	-0.01124	0.0003764	0.3224	0.0331	24.25	9.564	-0.0116	-1.764
p value	<2E-16	< 2E-16	5.023E-01	< 2E-16	9.136E-01	< 2E-16	6.205E-01	9.58E-14	4.72E-11
Shopping	-0.0000605	0.008967	-0.02069	-0.2353	0.309	-11.73	-12.02	-0.00308	1.362
p value	<2E-16	5.07E-10	8.6E-10	< 2E-16	3.55E-01	9.05E-09	5.885E-01	9.58E-03	1.23E-04
Season 1	0.9904	-4.85	-1.658	-1.487	1.637	525.6	136.6	2.32	35.58
p value	4.6E-06	< 2E-16	< 2E-16	2.48E-06	5.88E-01	< 2E-16	7.03E-01	< 2E-16	3.62E-14
Season 2	0.458	-0.6686	-0.1539	-0.2414	7.102	573.6	152.6	2.32	45.46
p value	6.87E-13	1.67E-01	3.705E-01	4.39E-01	1.24E-12	< 2E-16	6.70E-01	< 2E-16	< 2E-16
Season 3	-0.2266	-7.243	-1.716	-1.821	2.814	539	123.3	2.32	36.94
p value	2.418E-03	< 2E-16	< 2E-16	8.93E-09	9.58E-03	< 2E-16	7.313E-01	< 2E-16	3.38E-15
Weekend	-2.461	-11.12	-1.074	-1.392	-11.7	-52	39.73	-2.874	-14.6
p value	<2E-16	< 2E-16	< 2E-16	< 2E-16	< 2E-16	3.34E-03	3.96E-03	< 2E-16	< 2E-16
Off 2	2.547	14.26	3.521	3.78	16.1	41.61	26	4.372	18.06
p value	<2E-16	< 2E-16	< 2E-16	< 2E-16	5.63E-08	2.06E-01	2.288E-01	< 2E-16	< 2E-16
Off 3	0.3614	3.991	0.1996	0.2928	2.522	-18.11	2.372	0.7045	4.085
p value	3.54E-04	< 2E-16	2.754E-02	2.46E-02	3.997E-01	6.25E-01	9.206E-01	3.62E-09	5.839E-03
On 1	2.164	4.113	-0.2951	2.567	2.61	2.463	-21.6	3.14	-4.05
p value	<2E-16	< 2E-16	6.96E-03	< 2E-16	4.032E-01	9.52E-01	3.988E-01	< 2E-16	1.917E-02
On 2	1.648	12.67	1.208	2.574	7.213	35.94	1.836	2.657	10.11
p value	<2E-16	< 2E-16	< 2E-16	< 2E-16	1.472E-02	2.84E-01	9.331E-01	< 2E-16	3.36E-12
R square	0.1089	0.1295	0.2841	0.333	0.09497	0.2624	0.2398	0.1569	0.2186

For all cities, off-peak hour day and evening on-peak hour increase the number of trip generated. This can be possible because people prefer to use bike sharing for leisure trips rather than working or education trips where people have limited time (time constraints) to do their trips. The same situation happened for peak hour evening. For all cities, all coefficients of evening peak hour factors are positive. This is different from morning peak hour where three cities had negative coefficients. This can be possible because of the same reason as leisure trips as well as for evening peak hour where people do not have time constraints to do their trips. The weekend was also a strong factor to reduces the number of trips. This supported by the negative coefficients of this factor for all cities except for Cologne. Season was also not a strong factor to the number of trips generated. For Hamburg, Kassel, and Stuttgart, every season decrease the number of trips generated.

Adding temporal factors gave a different relation of POIs, population, and bike stations with the number of observed trips. Most of the factors, except food and recreation, decrease the number of trips of some cities. For Frankfurt, only entertainment as an attraction factor increases the bike sharing trips. The P-value of all factors were less than 5% for almost every city except for Darmstadt, Munich, and Cologne. For most of the city, every factor had strong relationships with the number of observed trips except for populations. Only four cities had P-value less than 5% for populations. The weekend was the only factor with P-value less than 5% in every city with negative coefficient. It can be assumed that weekend was the strongest factor that reduces the number of trips generated. Table 4.18 shows the relation of each factor of 10 cities.

Table 4.18 Relation of factors to the number of trips generated by DB Call-a-Bike

Variable	Estimate	P value	Significant codes
Population	1.38E-04	<2E-16	***
Bike station	3.22E-03	<2E-16	***
Education	-6.58E-05	<2E-16	***
Entertainment	-8.29E-03	<2E-16	***
Accessibility	3.92E-05	<2E-16	***
Food and recreation	2.63E-02	<2E-16	***
Shopping	-2.44E-04	<2E-16	***
Season 1	-2.206E+00	<2E-16	***
Season 2	2.47E+00	<2E-16	***
Season 3	-1.03E+00	<2E-16	***
Weekend	-6.04E+00	<2E-16	***
Off 2	8.20E+00	<2E-16	***
Off 3	2.01E+00	<2E-16	***
On 1	2.68E+00	<2E-16	***
On 2	6.57E+00	<2E-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

For ten cities, the P values of all factors were less than 5%, indicating all factors were related to the number of trips generated. The population, number of bike stations, and accessibility remain positive for all five tested models, showed that trips production factors were more significant to generate trips by bike sharing, while only one of the trip attraction factor, which was food and recreation, increases the number of DB Call-a-Bike trips. The population and the number of bike stations clearly are major factors that increase the number of bike sharing trips because the high availability of the bicycle surrounded by the high number of population will trigger the bicycle sharing usage.

Entertainment had negative relationship with the number of observed trips probably because most of the entertainment POIs (bar, pub, nightclub, cinema, and theatre) are more active in the evening (after work or school) which is colder and windier rather than day time while trip with bicycle is more convenient to do in the daylight environment.

Education had a negative relationship with the number of observed trips plausibly because this trip required limited travel time, so people prefer to use another faster transportation mode. The other reason could possibly because the number of education places is way lower compare to other POIs, while in fact, one education place attracts higher people rather than, for example, one restaurant. For the better analysis of multilinear regression model of the trip generation, the weighting of POIs can be done according to the number of people that can be accommodated for every $1km^2$ of the POIs area. This step probably would give better results of the trip generation model (more explainable).

Shopping also had a negative relationship with the number of observed trips. This probably happened because it is hard for people to carry their heavy groceries or other goods with a bicycle. Therefore, people probably will prefer to use other transport modes such as private vehicle or public transport for shopping trips.

Food and recreation had a positive relationship with the number of observed trips, shows that the high number of this POI leads to higher number of bike sharing trips. People prefer to use the bicycle for leisurely trips because it has no time constraints and the bicycle trip itself is the part of the leisure activity.

The way to calculate accessibility is to consider the population and the distance of the trips from zone to zone. This factor also had a positive relationship with the number of observed trips, because the distance and the travel time of bicycle depends on human power. Therefore, the travel time and the distance strongly influence the decision of people to use the bicycle.

Season 1, which includes winter and spring (January – April), and season 3, which includes fall and winter (September-December), showed negative coefficients, indicating that these factors decrease the number of trips generated. This is clearly explainable because people avoid bicycle as a transport mode due to the temperature. The weekend also reduces the number of bike sharing trips due to the less number of people is active over the weekend.

Evening peak hour had higher coefficient rather than morning peak hour, plausibly, as mentioned before, due to the time constraints of the trip (people has no time constraint to reach their home after work or school). The day off-peak hour (off 2) had higher coefficient among all the hour type variables due to the number of trips. The off-peak 2 is from 09.00AM – 16.00PM or 7hour while the on-peak hour only last for 3hour. For better result, the time range of the off-peak hour should be last the same way with on-peak hour.

Factors that were included in the regression model should be according to the objective of the study. For example, if only wants to see the relation of the population to the number of bike sharing trips, other factors rather than population should not be included. Including all factors to the model will cover up the relation of other factors that should be in the objective of the study. Therefore, choosing trip generation regression model is according to the objective of factors that want to be studied. This can be seen by comparing the results of five trip generation models before. Each result shows different factor that increases the number of bike sharing trips.

4.3.6. Validation of the model

After testing the models, the number of trips generated from each model was compared to the number of observation trips. Comparing R-square value and the pattern of residual value to measure the fitness of the trip generation model and observation data was done to investigate the validity of the model. Only four models were compared with the number of observed trips which are model 2,3,4 and 5 as mentioned in Table 4.9.

To simplify the calculation, the season only divided into 2 categories (summer and winter) and the hours' type also only divided into 2 categories (peak and off-peak hour). Unlike POIs, population or number of bike stations, the temporal factors do not have values on it. Temporal factors only have code to indicate in which time-period the registered trips belong to. Dummy factors were used to mark the temporal factors. Winter, off-peak hour, and weekend have 0 as a dummy factor, and the rest of the temporal factors have 1 as a dummy factor. Table 4.19 shows the R-square results for each model compared.

Table 4.19 Goodness of fit of model with observed trips in percent

City	Model 2	Model 3	Model 4	Model 5
Frankfurt	38.39	40.54	38.8	39.73
Hamburg	36.41	39.83	38.18	47.7
Kassel	32.1	40.03	39.4	50.15
Marburg	30	31	31	27.06
Stuttgart	45	46	42.19	49.79
Darmstadt	12.7	19.97	19.02	40.41
Munich	6	16	10	13
Cologne	8	9	10	37.39
Berlin	30.78	31.98	25.34	40.15
Overall R-square	26.6	30.5	28.2	38.3

Table 4.19 shows that the fifth model, which includes all variables, had higher R-square value rather than other models. This indicates that more factors consider to predicts the trip generation model leads to the higher good of fitness with the observation data. The residual plot also used to investigate the validity of the model. Figure 4.14 shows the difference of residual plots of each model for Hamburg.

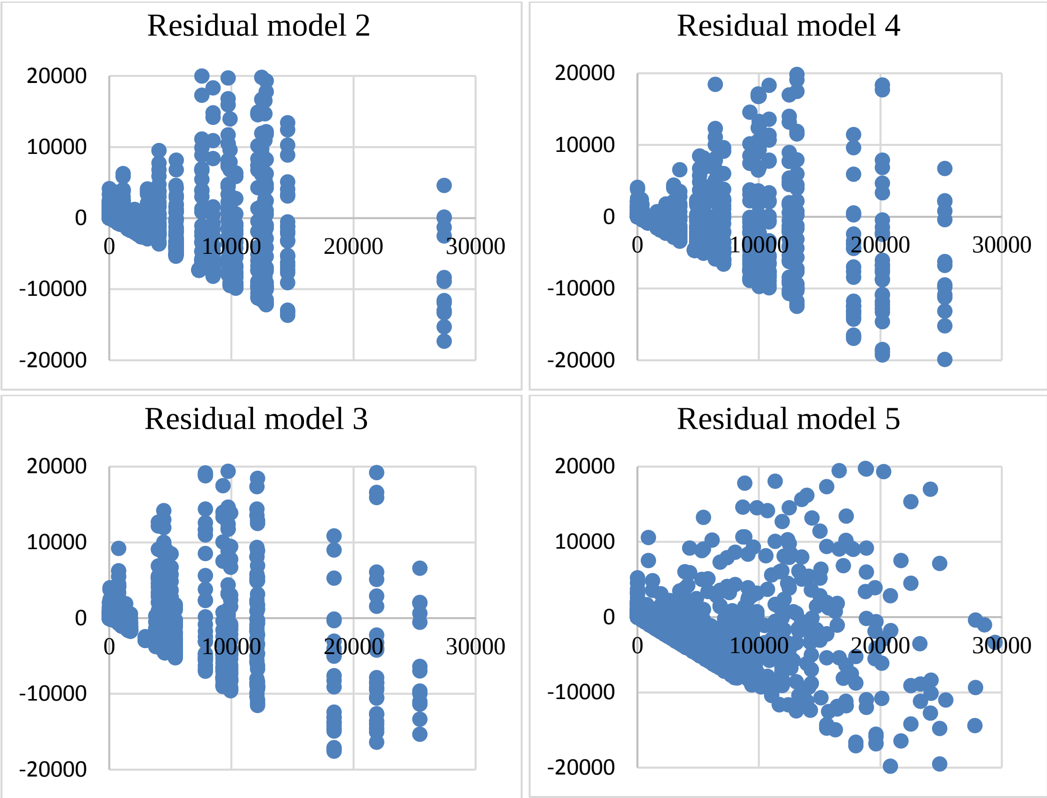


Figure 4.13 Residual plots comparison of trip generation model for hamburg

The x-axis in the figure above is the value of the prediction while the y-axis is the deviation between observed and prediction value. The residual plots above show points that fall under 0 of the y-axis (negative value) were overpredicted while points that fall above 0 of the y-axis (positive value) were underpredicted. The closer the points to the 0 of the y-axis show that the prediction values were closer to the observed data value. According to Figure 4.14, model 5 has more points closer to 0 of the y-axis rather than other model. This means model 5 can predict the trip generation better than the other models. Model 2, 3 and 4 have some gaps in x-axis because the model only considers variables according to the bike zones. Therefore, each line represents one bike zone. This situation shows that trip generation model with only spatial information as an explanatory variable is not significant enough to predict the trip generation. But model 5, which considers temporal factor, shows values that more evenly distributed because the model combined temporal and spatial factors that affect trip generation. Therefore, the trip prediction values were closer to the observed data in model 5. According to the R-square value and residual plot, it can be concluded that the 5th model, which includes spatial and temporal factors, can produces the trip generation model closer to the observed data rather than just consider spatial factors.

4.4. Trip distribution

Required information for trip distribution with gravity model is the number of trip production and attraction for each synthetic zone and travel time from and to each zone. In this case, the number of trips from the real observed data were used instead of the trip generation model output. The real observed data preferred to be used because the trip generation models, that previously tested, produced the number of trips generated with low R-square value as mentioned in Table 4.19. This demonstrates that the quality of the trip generation model is not as accurate as the observed value. Therefore, it was decided to use the observed trips values for the trip distribution inputs. The travel time used for trip distribution was the one obtained via Google maps rather than the travel time from the DB Call-a-Bike database. It was decided to use the travel time from Google maps because some of the travel time on DB Call-a-Bike data are not valid compared to the distance. The real valid travel time from each zone is required for the trip distribution model. As mentioned in Chapter 2, the double constraint of gravity model was used to predict the trip distribution. There are three models for trip distribution that will be tested. The difference is only in the impedance equation. Table 4.20 shows the explanation of impedance models.

Table 4.20 Trip distribution impedance equation

Model	Impedance equation
1. Travel time	$=e^{(\alpha*TravelTimeinMinutes)}$
2.Travel time, POIs, population	$=e^{((\alpha*TravelTimeinMinutes)+(\beta*LN(Population))+(\gamma*TotalPOIs))}$
3. Travel time, POIs, population, observation data	$=e^{((\alpha*TravelTimeinMinutes)+(\beta*LN(Population))+(\gamma*TotalPOIs)+(\omega*ObservedTrips))}$

The first model only used travel time for impedance function input. The higher the travel time the lower the number of trips, therefore the impedance factor for travel time is minus. The impedance factor of -0.15 was chosen to avoid underprediction the trips with travel time higher than 20 minutes. As mentioned in Chapter 2, 18% of total trips have travel time was higher than 20 minutes. Figure 4.15 shows the comparison of the impedance factor.

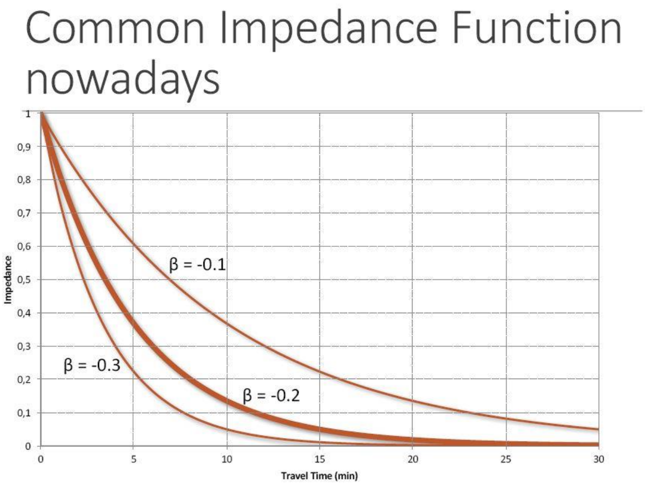


Figure 4.14 comparison of impedance function (Rolf Moeckel 2017)

The results of the trip distribution prediction model were compared to the observed data. The linear regression was used to know the quality of the model. The higher the R-square value the closer the predicted value with the observation data. For the first model, the average result shows the relation (R-square) between trip prediction and observed data is around 80%. Table 4.21 shows the relation of trip prediction and observed data of each city for each model.

To increase the precision between the observed data and the prediction, other spatial factors, which are population, POIs, and observed value, were also considered in the impedance function. The second model included spatial factors in the impedance function. This means to counterbalance the trip impedance because it reduces the trips caused by high travel time, while population and POIs increase the number of trips. Population and POIs were also included in the trip impedance function because the reason people travel with bicycle influenced by attraction factors. Probably, some people willing to travel longer to reach their trip attraction destination despite the distance. In contrary to the travel time, other factors (population, POIs, and observed value) are not the trips resistance. Theoretically, the higher the population and POIs the higher the number of trips. Therefore, the coefficient for these factors are positive. The coefficients for population and POIs are quite small because the R-square value, which interprets the quality of the first gravity model, was already good with the average result of 80%. The impedance coefficients for spatial factors were based on trial and error. After some testing, the optimal impedance coefficients for POIs and population were 0.005. If spatial factors were included in the impedance function, the quality of the model will only slightly improve from 80% in average to 85%. Table 4.21 shows the results of all gravity model for each city.

Table 4.21 Comparison of the trip distribution model goodness of fit

City	Model 1 (%)	Model 2 (%)	Model 3 (%)
Berlin	90.04	94.51	99.8
Kassel	86.25	86.13	53.14
Stuttgart	91.15	91.4	91.9
Munich	40.3	38.69	51.04
Darmstadt	78.9	78.29	53.91
Marburg	97.4	97.45	66.69
Frankfurt	96.29	96.19	53.24
Hamburg	90.26	96.29	12.36
Cologne	87.79	87.79	87.44
Russelsheim	42.4	43	67.89
Overall R-square	84.2	85.2	63.3

The third model included spatial factors and observed OD pair data to the model. Similar to the second model, the factor for observed trips was based on trial and error. After several tests, the most optimum factor obtained was also 0.005. Table 4.21 shows for four cities (Berlin, Stuttgart, Munich, and Rüsselsheim) adding observed data increases the fitness of the observation and prediction. For the rest of the cities, it reduces the fitness. The reason of the low quality of the third model for some cities is because the OD matrix of observation data has a lot of trips that are not exist in certain OD pair. But in the trip distribution model, all OD pairs will be calculated according to the travel time of each OD pair. Another reason is because there are some OD pairs with high trip numbers around 100,000 observation data in 3 years, while another OD pair only have a few trips with less than 10 trips for 3 years. Considering these situations, the third impedance model produces unreliable results for some cities because it will over- or underpredicted the OD trips based on the observed data.

Table 4.21 shows the average R-square results for the third model was lower than the first two model which was around 63%. The third model cannot be used for other application, for example expanding the bike sharing zones, because the number of observed trips in new zones are not exist. Inversely, the number of POIs and population can be known from the census and Geofabrik data.

The residual plot was also used to investigate the quality of the trip distribution model. Figure 4.16 shows the comparison of the residual plot for each model for Berlin. Model 2 had more value closer to zero compared to model 1, indicating that model 2 is better to predict the trip distribution. While for Berlin, model 3 will made the trip generation model output had closer value to the observed data.

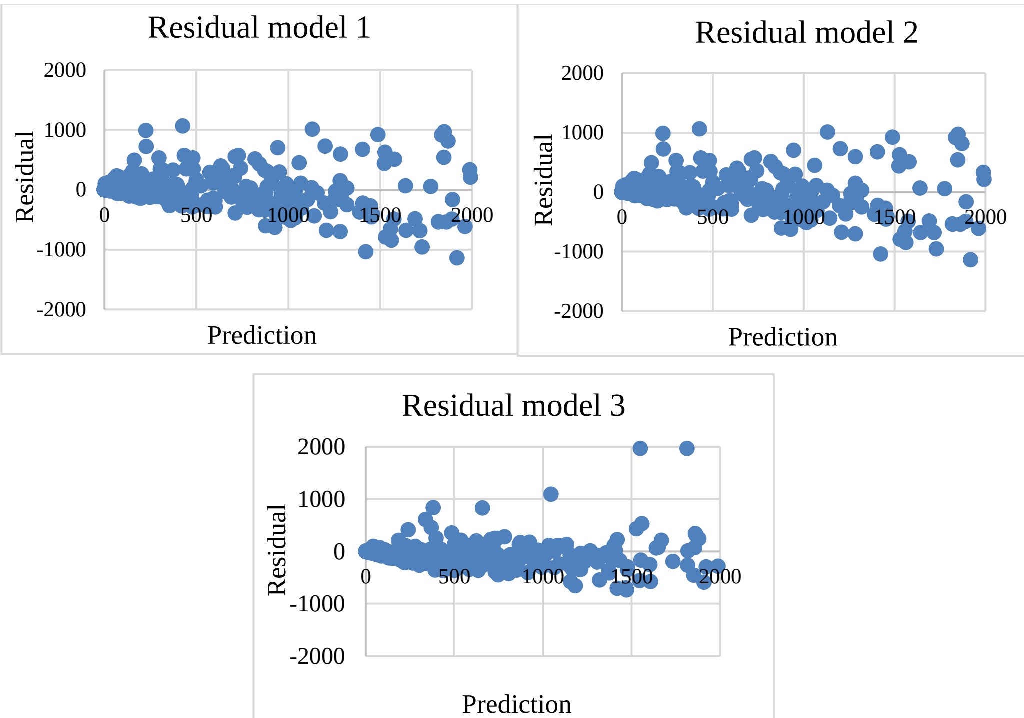


Figure 4.15 Residual plots of trip distribution prediction to observation data of Berlin

Figure 4.17 shows the comparison of the residual plot for each model for Hamburg. As mentioned before, for some cities, model 3 reduces the quality of the trip distribution model. Figure 4.17 shows that model 2 has more closer value to 0 compared to model 3 and 1.

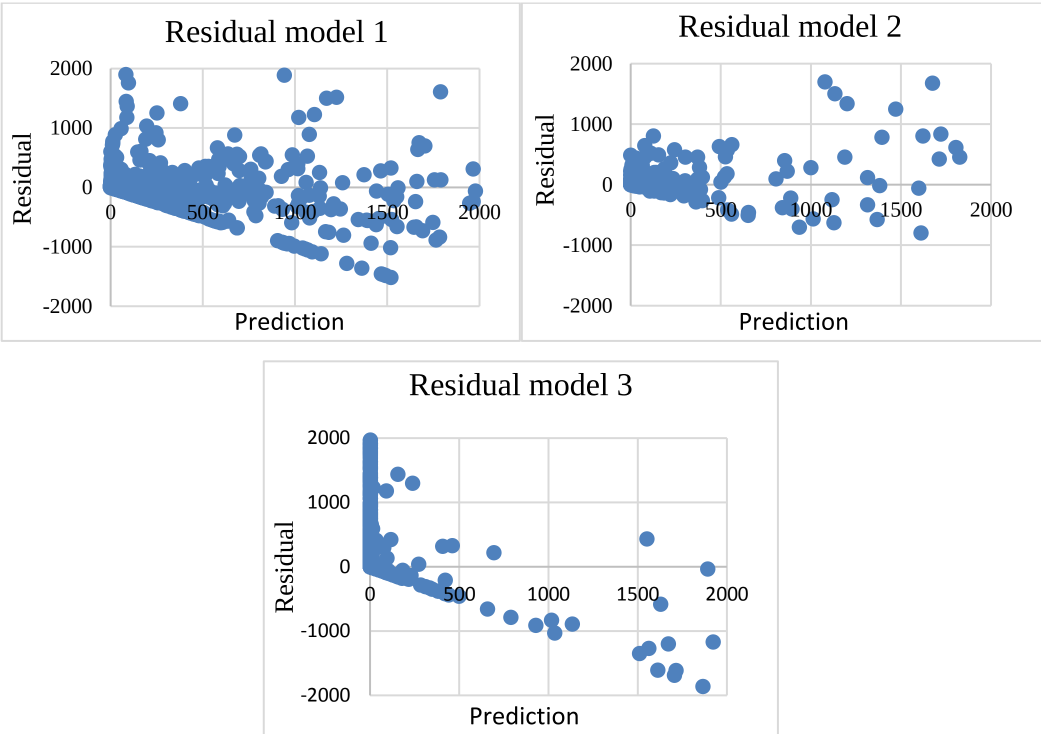


Figure 4.16 Residual plots of trip distribution prediction to observation data of Hamburg

It can be concluded that model 1, which is only using travel time as a travel impedance factor, is already enough to make the prediction of trip distribution model. The reason behind this probably because the bicycle trips rely on the physical condition. It needs direct human power and longer travel time to reach destination compare to another mode of transport. Bicycle users do not need to consider fuel cost or parking as complicated as another mode of transport. Therefore, bicycle users consider travel time way higher as an impedance in choosing their mode of transport. This leads to a decision that only considers travel time as impedance for bicycle sharing trip is already enough to model OD matrix of trip distribution.

Figure 4.16 shows the 3D histogram of Marburg OD matrix and the difference before and after applying the second model to give a better understanding of the difference between OD matrix of observation data and OD matrix of the second model.

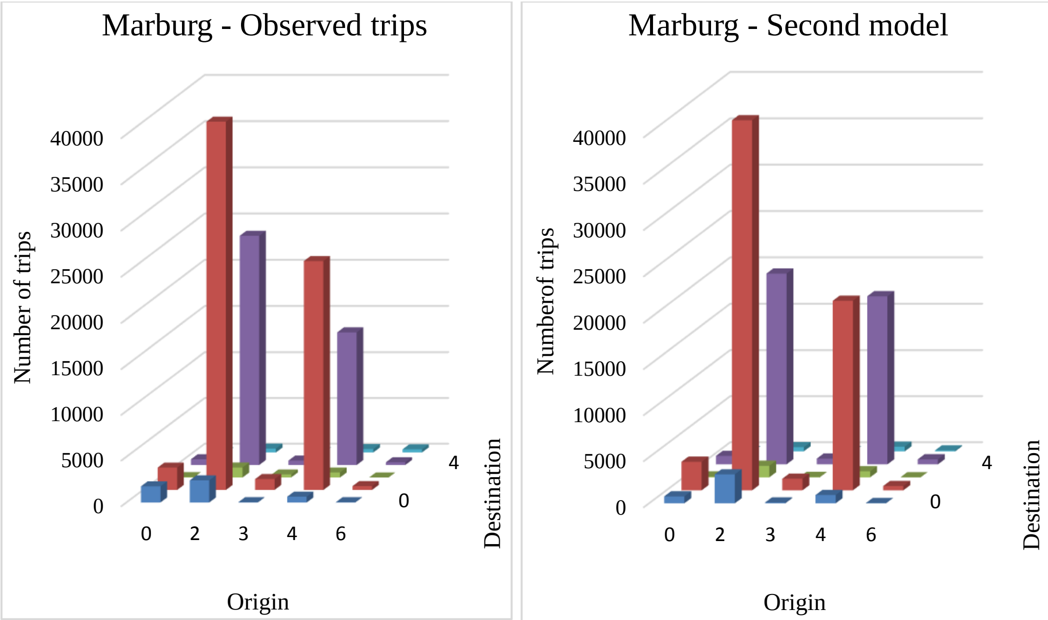


Figure 4.17 Marburg OD matrix comparison of observed and second model trips

Figure 4.16 shows that the number of trips of OD pair 4-2 and 2-4 decreased in the second model, while for zone 4-4, the number of trips were higher. This might possible because the gravity model only considers travel time as an impedance. The inner trips of zone 4 had shorter travel time compared to travel time between zones 2 to 4, therefore the number of trips of inner zone 4 were higher in the second gravity model rather than the observation data. But in fact, in the observation data, the number of trips between zone 2 and 4 were higher because the number of POIs in zone 2 is higher compared to zone 4. This supports that the higher POIs attract more trips in real life. Table 4.22 shows the comparison of zones attribute in Marburg.

Table 4.22 Zones attribute in Marburg

Zone	Bike stations	Population	POIs
0	3	15460	45
2	2	21790	226
3	1	4433	20
4	8	7008	76
6	9	9916	41

4.5. Bike sharing activites and city profile

The number of trips generated in a city depends on different spatial factors based on the trip generation analysis. Table 4.23 shows the comparison of the density of the number of bike stations, population and the number of POIs for each kilometer square in zones serves by DB Call-a-Bike.

Table 4.23 Density comparison of each city

City	Number of bike stations	Population/km	POIs/km	Trip average per day
Hamburg	102	4103	51	7053
Berlin	114	6852	108	1094
Kassel	34	2475	13	511
Stuttgart	29	4061	26	504
Darmstadt	25	3138	24	193
Marburg	15	2930	20	120
Rüsselsheim	8	2418	5	44
Frankfurt	252	3821	20	1011
Cologne	80	3753	43	751
Munich	85	4138	54	260

Overall, the higher the density of population the higher the number of trips generated, except for Kassel. Kassel had higher average trips per day rather than Stuttgart, even though Kassel has lower population covered by DB Call-a-Bike. But Hamburg is an exception, because it had the highest average trip number per day even though it has a lower number of bike stations and POIs compare to Berlin and Frankfurt. The same thing also happens for the number of bike stations. The higher the number of bike stations leads to the higher number of trips. As mentioned before, some area of Frankfurt does not have bike stations. For Cologne and Munich, the DB Call-a-Bike does not provide bike stations at all. Therefore, the number of bike stations cannot be compared with other cities.

Table 4.23 shows that dense population or POIs will not always lead to higher bike sharing trips. Because of this reason, there is a possibility that the inhabitants probably have their own bicycle and not use a bicycle sharing systems.

CHAPTER 5 APPLICATION

Tested trip generation and distribution model will be used to know the changes of the trips if they are applied to a scenario. In this scenario, the output from trip generation model will be used as an input of trip distribution model.

5.1. Trip generation application with expanding bike sharing zones area

The purpose of assign new bike stations to expanded zones is to know the rise of bike sharing usage in a city. The new zones are according to the zones that does not have bike stations before. These zones are previously made on Arc-GIS in the trip generation analysis. The third model of the trip generation, which is using population, bike stations, POIs, and accessibility as explanatory variables, was chosen for this application. Even though the fourth model of the trip generation produced better trip prediction, it is not used because this model needs temporal factor which cannot be known. The number of zones added is different for each city because of the synthetic zones that were created produced a different number of zones depends on the bike sharing coverage of the area. For a city that has a high coverage area of DB Call-a-Bike, the number of zones produced will be higher compared to a city with low coverage area. The difference will be shown according to the synthetic zones created in Cologne and Frankfurt in Figure 5.1 and Figure 5.2

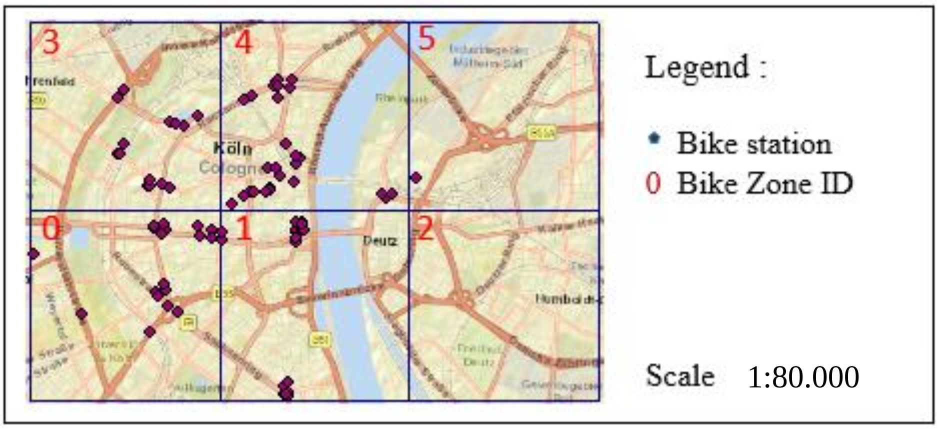


Figure 5.1 Koln DB Call-a-Bike stations coverage area

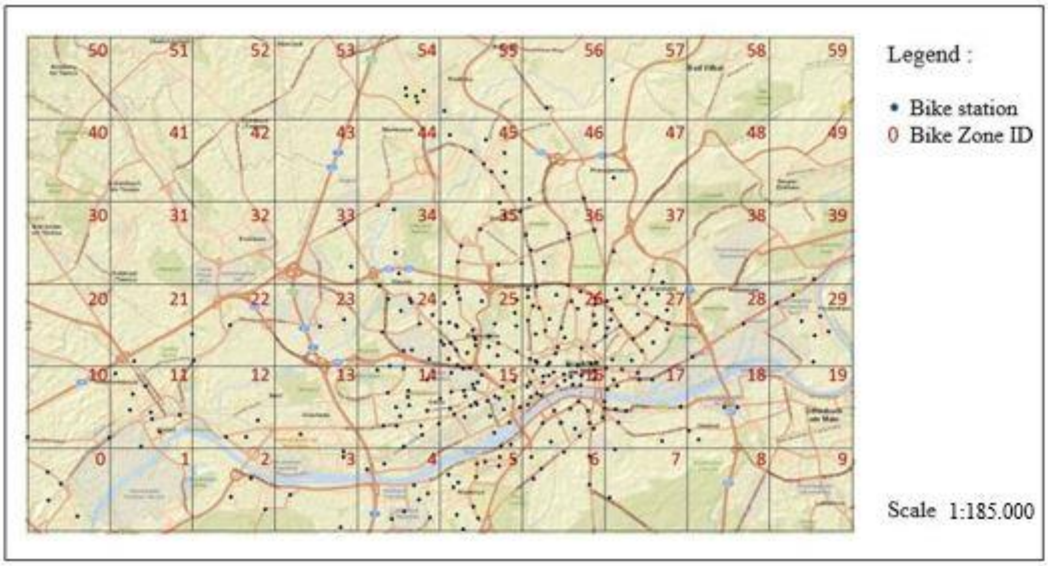


Figure 5.2 Frankfurt DB Call-a-Bike stations coverage area

Cologne has low coverage area of DB Call-a-Bike. The center of bicycle sharing activity only concentrates in the city center. This situation leads to only one zone left without bike station as shown in Figure 5.1. In the other way, Frankfurt with higher coverage area of DB Call-a-Bike and ubiquitous bike stations coverage which is not concentrate only in the city center, will have wider synthetic bike zones with farthest bike stations as synthetic zones boundary rather than Cologne. This caused to more synthetic bike zones without bike stations as shown in Figure 5.2. Each zone which does not have bike stations will have two new bike stations. The trip number will be calculated according to the trip generation model based on each city. Table 5.1 shows the comparison of before and after expanding bike sharing zones.

The number of trips produced from the new bike zones are not high because new zones are not located in the city center and the zones attribute (POIs and bike stations) have significant difference compare to existing zones. Most of the zones have lower population and POIs compared to the city center. Therefore, the number of trips rising only between 0-1% for less than 5 new expansion zones, and 1-9% for more than 19 expansion zones.

Table 5.1 Comparison of trips rising after bike zones expansion

City	Trips observation	Number of trips in expansion zone	Rise of bike sharing usage	Number of zones added
Munich	203441	5564	2.7%	21
Marburg	97227	5616	5.7%	3
Cologne	586007	1672	0.3%	1
Hamburg	5113524	88144	1.7%	55
Frankfurt	686041	387	0.5%	22
Darmstadt	156615	253	0.16%	1
Kassel	460038	53	0.11%	2
Stuttgart	382842	2140	9.18%	19
Berlin	791596	13806	1.74%	30

5.2. Trip generation application with adding points of interest and bike stations in existing zones

Based on the tested trip generation models, factors that highly influence bike sharing trips are population, bike stations, and food and recreation. The purpose of increasing the number of bike stations and food and recreation points in existing bike zones in each city is to know how the number of trips generated changes. Each of the city has a different level of degree of factors that influence bike sharing trips. Therefore, each of the city will have different changes towards increasing bike stations or increasing number of food and recreation. The third model of trip generation, which considers population, bike stations, POIs, and accessibility, will be used for this application. For Munich and Cologne, more central bike sharing activity zones will be added. Four scenarios will be applied to the existing zones. The first scenario is to add two more bike stations, the second scenario is to add four more bike stations, the third scenario is to add 20% more bike stations and lastly the fourth scenario is to add 20% more food and recreation points. The additional bike stations or food and recreation only applied in each existing zone. Table 5.2 shows the changes of number of trips when increasing bike stations and food and recreation points.

Table 5.2 Changes of number of trips according to increasing number of independent variables

City	Add 2 bike stations (%)	Add 4 bike stations (%)	20% more bike stations (%)	Add 20% food and recreation (%)
Berlin	42	44	29	8
Darmstadt	16	32	16	106
Frankfurt	63	90	42	70
Hamburg	26	51	18	10
Kassel	29	57	13	35
Cologne	26	46	1	52
Marburg	156	247	42	170
Munich	13	14	9	63
Stuttgart	112	224	32	15
Average	54	89	22	59

For most of the cities except Berlin, Hamburg, and Stuttgart, additional food and recreation points will generate more trips rather than adding 20% more of bike stations. This happened according to Table 4.15 of regression linear results from the third model of trip generation, where the coefficient of food and recreation factor is positive and it will increase more trips rather than increasing the number of bike stations. For Hamburg, the coefficient factors for food and recreation is negative and it leads to increase only 10% of number of trips generated according to the model.

For most of the cities except Berlin and Munich, doubling the additional number of bike stations will increase the number of trips generated drastically. This is according to the regression result of the third model in Table 4.15 where the number of bike stations strongly influence the increasing number of trips generated.

5.3. Trip distribution with expansion zones

Information needed to produce OD matrix of trip distribution with expansion zones is the number of trips production and attraction in each zone (from trip generation third model output in section 4.3.5), the travel time between zones and the zones attribute. With Arc-GIS, the coordinates of each existing and new zones centroid can be obtained. This coordinates then will be used as an input to obtain the travel time with a bicycle to each zone from Google maps via Google API Javascript. With the similar steps as in section 4.4 of trip distribution, an OD matrix of each city can be produced. The result show that all cities have the same pattern of trip number changes due to the expansion of bike zones area. Figure 5.3 shows the difference of number of trips between each OD pair before and after bike zones expansion for Marburg.

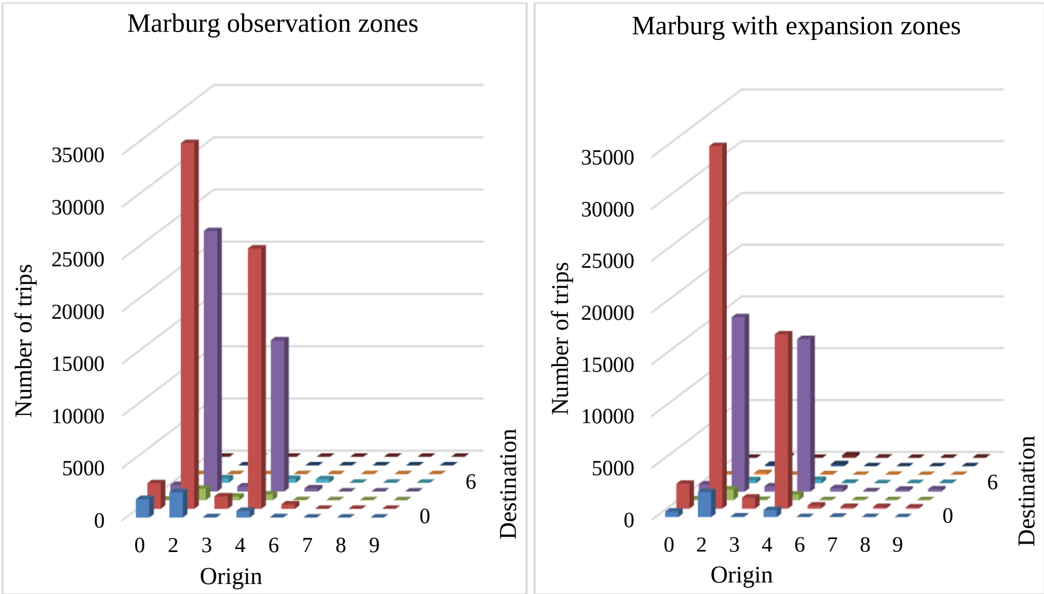


Figure 5.3 Comparison of number of trips between zones in Marburg before and after zones expansion

Figure 5.3 has x-axis as the name of origin zones and y-axis as the name of destination zones in the same order. Marburg only has three more expansion zones which are zone 7,8, and 9. Most of the zones will have lower trip numbers because, as mentioned in section 4.5.6, gravity model only considers travel time as travel impedance. When in real life, POIs also affect number of trips in each OD. New zones will produce only a few number of trips compared to existing zones with the same reason mentioned in section 5.1. This happened according to the lower population in new zones compared to existing zones.

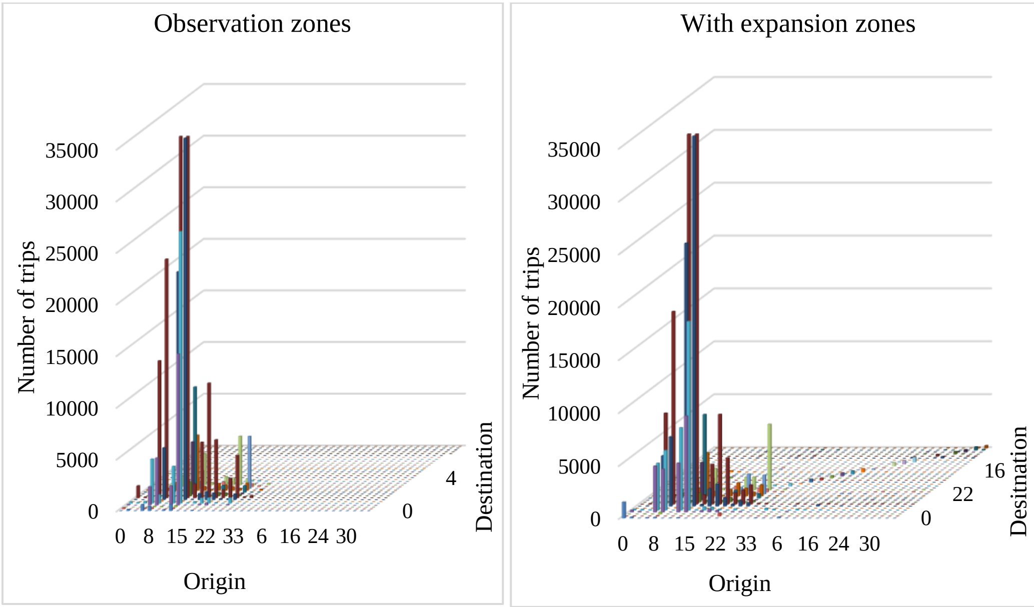


Figure 5.4 Comparison of number of trips between zones in Stuttgart before and after zones expansion

Figure 5.4 shows the example the city of Stuttgart with more than five expansion zones. Stuttgart has 19 new expansion zones. In the new zones, most of the high number of trips produced inside the same zones (inner zones trips) because of lower travel time compared to the travel time between different zones while the other expansion zones arise the lower number of trips. As mentioned before, this happened due to the lower number of population in the new expanded zones.

5.4. Trip distribution with adding points of interest and bike stations in existing zones

The result from the trip generation application in section 5.2, with the same scenarios as mentioned in section 5.2, will be used as an input for the trip distribution. The purpose of testing the trip distribution model is to know how the number of trips changes between each zone if more bike stations or more food and recreation points are assigned to the existing synthetic bike zones. All cities will show the same pattern of changes. Figure 5.5 shows the example of the OD matrix comparison in Marburg after applying the scenarios of adding POIs and bike stations in the existing zones.

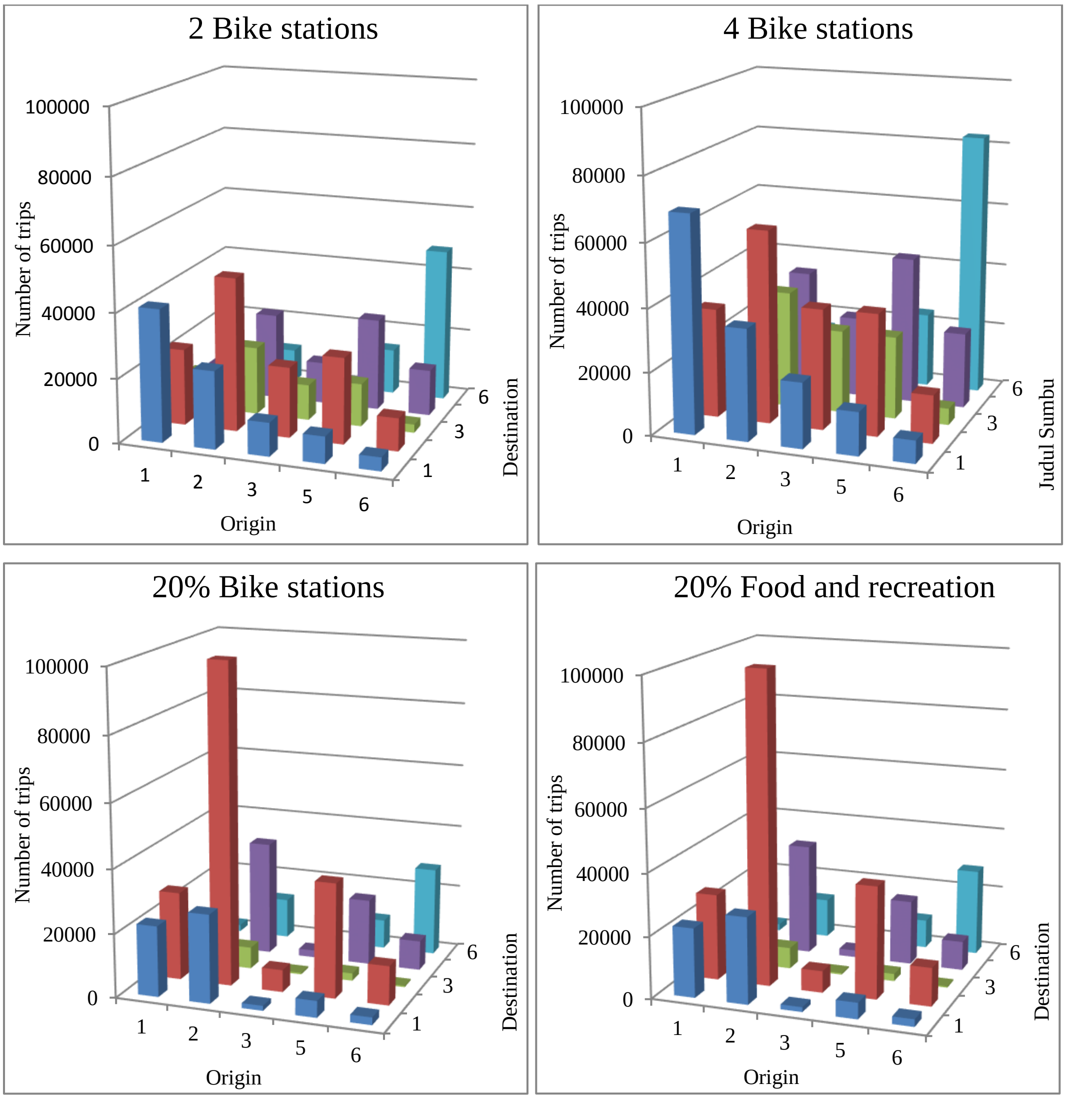


Figure 5.5 Comparison of number of trips between zones in Marburg for adding POIs and bike stations scenario

Placing two and four bike stations in the existing zones arise the number of trips more evenly in each OD pair rather than placing 20% more bike stations or 20% more food and recreation in the existing zones. For scenario 3 and 4 (add 20% more bike stations and food and recreation) the drastic changes of the increasing trips only arise in the zone with the high number of bike stations and food and recreation in the real condition. For other zones, with lower number of bike stations and food and recreation, the changes of the trips will not be significant.

CHAPTER 6 DISCUSSIONS AND CONCLUSIONS

6.1. Discussions

For Frankfurt and Marburg, the first model of trip generation, which consider population and bike stations, produces negative coefficient for population variables. Normally the higher the population and bike stations will lead to higher number of trips. After some investigations, this can be happened as an outcome when assigning the synthetic zones. A lot of trips generated in the border of the zone that has higher population and bike stations. Therefore, when tested the multilinear regression, the results will show that population does not have direct relation with number of trips generated because the trip should be assign to the neighbourhood zones. This will be explained in the Figure 6.1 of Frankfurt city.

Figure 6.1 Bike stations in the border of the zones

There are five bike stations located in the border of zones 25 and 15. The population in zone 25 is higher rather than in zone 15, but more trips were generated in zone 15 rather than zone 25 as shown in Table 6.1

Tabel 6.1 Example of trips in Frankfurt

Hours' type	Season	Bikezone_number	Sum_bike stat	Population	Number of trips
on1	2	15	31	17396	19222
on1	2	15	31	17396	17005
off2	2	15	31	17396	16801
off2	1	25	29	28889	14999
off2	1	25	29	28889	14911
on2	3	26	34	54004	13849
on2	3	15	31	17396	13356
on2	3	15	31	17396	13276
on2	3	15	31	17396	12338
off2	3	25	29	28889	11885
off2	2	26	34	54004	11819
off1	2	16	33	37427	11801
off1	2	16	33	37427	11568
off1	1	26	34	54004	11390
off1	1	15	31	17396	10840

The better methods to assign the bike stations to the right zones in the same size is needed to produce more reliable trip generation regression model. Another thing that could be improved is the trip distribution analysis. The impedance factor of gravity model, which include spatial factors to counterbalance the number of trips, will reduces the impact of travel time as impedance and can improved the quality of the model. This shows that the number of trip attractions is needed in trip distribution process. Destination choice model, which has travel time included in utility function, could be tested to know whether this model will produce closer OD matrix value to the observed data.

6.2. Conclusions

The objective of this research is to develop and validate the trip generation and distribution model based on historical data of DB Call-a-Bike. Results and findings have been presented from Chapter 1-5. The purpose of this chapter is to summarize important conclusions that can be drawn from this research.

DB Call-a-Bike is an effective bike sharing systems for short distance travel. Around 70% of the trips were done in the range 0-15 minutes of travel time.

Temporal factors play important roles in bike sharing systems activity. The highest bike sharing activity was in summer season from May-August. Weekdays and weekend bike sharing usage pattern have significant difference. During weekdays, the majority of the trips generated were in the morning and afternoon peak hour while over the weekend, people are prone to use bike sharing in more leisurely hour.

Spatial factors also play important roles in bike sharing trips. There are two important spatial factors that significantly attract more bike sharing trips which are number of population and bike stations. Based on some tested models of trip generation in this research, not every type of POIs will attract bike sharing activity. But for overall results, food and recreation activity always attracts bike sharing usage in every city.

From all four tested trip generation models, the fourth model, which includes spatial and temporal factors, produces number of trips which is closer to the observation data. Measuring the quality of the fourth model is based on the R-square value, which is 38% to the observation data. The more variables added the higher the quality of the model (R-square value). The R-square value of the first model, which only considers bike stations and population, is 26%. The second model, which considers POIs, is 28%. Lastly, the third model, which considers population, bike stations, and POIs, is 30%.

From three impedance models of trip distribution gravity model, the second model, which includes population and number of POIs to counterbalance the effect of travel time as impedance, produces trip number closer to the observation data. The output of the models was compared with the observation data and the quality of the models are measured according to the R-square value. The second model has overall R-square value of 85.2%. While the first model, which only consider travel time as impedance, has slightly difference results of 84.2%. and lastly the third model, which considers travel time, spatial factors, and observation trips, has lower R-square value of 63.2%.

Using the gravity model to predict trip distribution shows that the lower the travel time the higher the number of trips generated. When comparing the result from the model with the observed data, for some OD pairs, the result is different. From the observed trip data, the zone that has higher POIs will attract more trips even though the travel time between this OD pair is high. But when looking to the same OD pairs from the trip distribution result, the number of trips generated is lower compared to other OD pair from zones with lower travel time.

For application of the tested trip generation and distribution model to expanding bike zones scenario, the third trip generation model was chosen rather than the fourth model because the temporal factors are unknown. Expanding bike zones were slightly increase the number of trips, around 1-9%, due to the spatial factors of the new expansion zones. New expansion zones have lower population and POIs compared to the existing zones. Therefore, the number of trips generated in new expansion zones was not that much. For trip distribution application, the second model, which includes spatial factor in the impedance formula, was chosen. The highest trip activity in the new zones arise in innerzones trips due to lower travel time.

Adding more bike stations and food and recreation points increase the number of trips generated significantly rather than expanding the bike zones. The number of trips increase in the average range of 22-89% if more bike stations are added to each existing zone (add 20%, 2 or 4 bike stations). The number of trips generated also increase to 59% in average if more food and recreation points (20%) are added to each existing zone.

According to the regression linear of trip generation results, the bike sharing system will be ideal to an area of a city that has denser population with higher leisurely POIs. If comparing all ten cities population and POIs density, the denser the population will not always generate the higher number of bike sharing trips. The reason behind this probably because there is a possibility if an area has denser population and POIs, the inhabitant would prefer to have their own bicycle rather than using bike sharing system.

PUBLICATION BIBLIOGRAPHY

ASTA TU Darmstadt (2014): Leihfahrradsystem Call a Bike, checked on 2/15/2017.

Bachand-Marleau, Julie; Lee, Brian; El-Geneidy, Ahmed (2012): Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. In *Transportation Research Record: Journal of the Transportation Research Board* (2314), pp. 66–71.

Cramer, Duncan; Howitt, Dennis (2004): The SAGE dictionary of statistics. A practical resource for students in the social sciences / Duncan Cramer and Dennis Howitt. London: SAGE.

DeMaio, Paul (2009): Bike-sharing: History, impacts, models of provision, and future. In *Journal of Public Transportation* 12 (4), p. 3.

Deutsche Bahn: Call a Bike Unternehmensprofil. Deutsche Bahn. Available online at https://www.callabike-interaktiv.de/de/unternehmen, checked on 2/2/2017.

Deutsche Bahn (2016): Call a Bike Hackathon Data. Deutsche Bahn. Available online at http://data.deutschebahn.com/dataset/data-call-a-bike, checked on 1/2/2017.

El-Assi, Wafic; Salah Mahmoud, Mohamed; Nurul Habib, Khandker (2017): Effects of built environment and weather on bike sharing demand. A station level analysis of commercial bike sharing in Toronto. In *Transportation* 44 (3), pp. 589–613. DOI: 10.1007/s11116-015-9669-z.

Frade, Inês; Ribeiro, Anabela (2014): Bicycle sharing systems demand. In *Procedia-Social and Behavioral Sciences* 111, pp. 518–527.

Geofabrik; OpenStreetMap: Geodata shapefile - Sub Region Germany. Available online at http://download.geofabrik.de/europe/germany.html, checked on 3/2/2017.

Geurs, Karst T.; van Wee, Bert; Rietveld, Piet (2006): Accessibility Appraisal of Integrated Land-Use ? Transport Strategies: Methodology and Case Study for the Netherlands Randstad Area. In *Environ Plann B Plann Des* 33 (5), pp. 639–660. DOI: 10.1068/b31099.

Gregerson, Jennifer; Hepp-Buchanan, Max; Rowe, Daniel; Vander Sluis, John; Wygonik, Erica; Xenakis, Micheal; McCormack, Edward (2010): Seattle bicycle share feasibility study. In *Seattle: University of Washington*.

Hampshire, Robert C.; Marla, Lavanya (2012): An analysis of bike sharing usage: Explaining trip generation and attraction from observed demand. In : 91st Annual meeting of the transportation research board, Washington, DC, pp. 12–2099.

Handy, Susan (2005): Planning for accessibility: In theory and in practice. In : Access to destinations: Emerald Group Publishing Limited, pp. 131–147.

Iacono, Michael; Krizek, Kevin; El-Geneidy, Ahmen (2008): Access to destinations: How close is close enough? Estimating accurate distance decay functions for multiple modes and different purposes. In *Minnesota Department of Transportation Research Services Section* 11.

Krykewycz, Gregory; Puchalsky, Christopher; Rocks, Joshua; Bonnette, Brittany; Jaskiewicz, Frank (2010): Defining a primary market and estimating demand for major bicycle-sharing program in Philadelphia, Pennsylvania. In *Transportation Research Record: Journal of the Transportation Research Board* (2143), pp. 117–124.

Mondal, Swarup; Mitra, Sudeshna; Majumdar, Bandhan Bandhu (Eds.) (2015): Investigating the influence of Trip Distance on Travel Behaviour of Non-Motorized Transport (NMT) users. Conference: 3rd Conference of Transportation Research Group of India. Kolkata, India, 17-20 December. Joy, Kolkata - India. Available online at https://www.researchgate.net/publication/288786922_Investigating_the_influence_of_Trip_Distance_on_Travel_Behaviour_of_Non-Motorized_Transport_NMT_users, checked on 5/27/2017.

Ortúzar, Juan de Dios; Willumsen, Luis G. (2011, ©2011): Modeling Transport. Fourth edition. Chichester, West Sussex, United Kingdom: John Wiley & Sons.

Rixey, R. (2013): Station-level forecasting of bikesharing ridership: Station Network Effects in Three US Systems. In *Transportation Research Record: Journal of the Transportation Research Board* (2387), pp. 46–55.

Rolf Moeckel (2017): Mathematical formulation of trip Impedance. Travel Forecasting Resource. Available online at http://tfresource.org/Impedance, checked on 4/30/2017.

Shaheen, Susan; Martin, Elliot; Cohen, Adam; Firson, Rachel (2012): Public Bikesharing in North America: Early Operator and User Understanding. In *Mineta Transport Institute* 11.

Statistische Ämter des Bundes und der Länder (2011): Bevölkerung im 1 Kilometer-Gitter. Available online at https://www.zensus2011.de/SharedDocs/Aktuelles/Ergebnisse/DemografischeGrunddaten.html?nn=3065474, checked on 3/2/2017.

Tin Tin, Sandar; Woodward, Alistair; Robinson, Elizabeth; Ameratunga, Shanthi (2012): Temporal, seasonal and weather effects on cycle volume: an ecological study. In *Environmental health : a global access science source* 11, p. 12. DOI: 10.1186/1476-069X-11-12.

Vogel, Patrick; Greiser, Torsten; Mattfeld, Dirk Christian (2011): Understanding Bike-Sharing Systems using Data Mining. Exploring Activity Patterns. In *Procedia - Social and Behavioral Sciences* 20, pp. 514–523. DOI: 10.1016/j.sbspro.2011.08.058.