DATA SCIENCE CAPSTONE PROJECT

Bike sharing customer behavior

Long Le

May 13, 2020

1. Introduction

Bike sharing is one promising solution to help reduce carbon dioxide footprint and increase public health. In the EU area, bike sharing has been implemented in many countries. Many bike sharing platforms have been introduced to spread the trend and they have proved to be quite successful.

However, bike sharing is not a new model. In China, the model has once been so popular that many issues emerged as the significant growth in the number bike sharing platforms and number of bikes available. For instance, the infrastructure in many big cities was not ready yet for the sudden, but tremendous increase in the number of bikes. There was not enough parking area, bike maintenance services, etc. Consequently, bikes were parked everywhere on the street and negatively affected the urban experience of citizens who did not use public bikes. Millions of broken bikes were put in bike graveyards without being fixed due to the lack of capability. As such, the effort to promote bike sharing service was a failure in achieving sustainable goals. Also, the above-mentioned issues reflect a tremendous waste in national resources.

I see a need to look at bike sharing activities in order to understand customer behavior. The insights would be useful for the bike sharing platform to better manage their businesses and avoid issues like those that once happened in China.

2. Study goals and research questions

The main goal of this study is to create an initial view on customer behavior in using public bikes via bike sharing platform. This study only focuses on one city in Germany. Before starting to start the analysis, I have some hypotheses:

a. Bikes are used mostly outdoors. Thus, weather conditions would very likely have an impact on the number of trips made.
b. Also, weather conditions could influence the duration of each trip.
c. There are multiple ways to use bikes. For instance, in Finland, bikes could be taken from stations and must be returned to stations.

However, Germany bikes could be offered at stations or as floating bikes. The question is, would users prefer stations or floating?

3. Description of data used

In this study, I will use the bike sharing data of one of the bike sharing platforms in the EU. The data was shared for study purposes and this study only focuses on one city in Germany and in the year 2019. The data considers the duration of the trips, the number of trips made, pick-up and return locations. After cleaning, the data set contains 154360 rows, which represent the trips made in the year, and 22 columns, which represent the relevant features for the study.

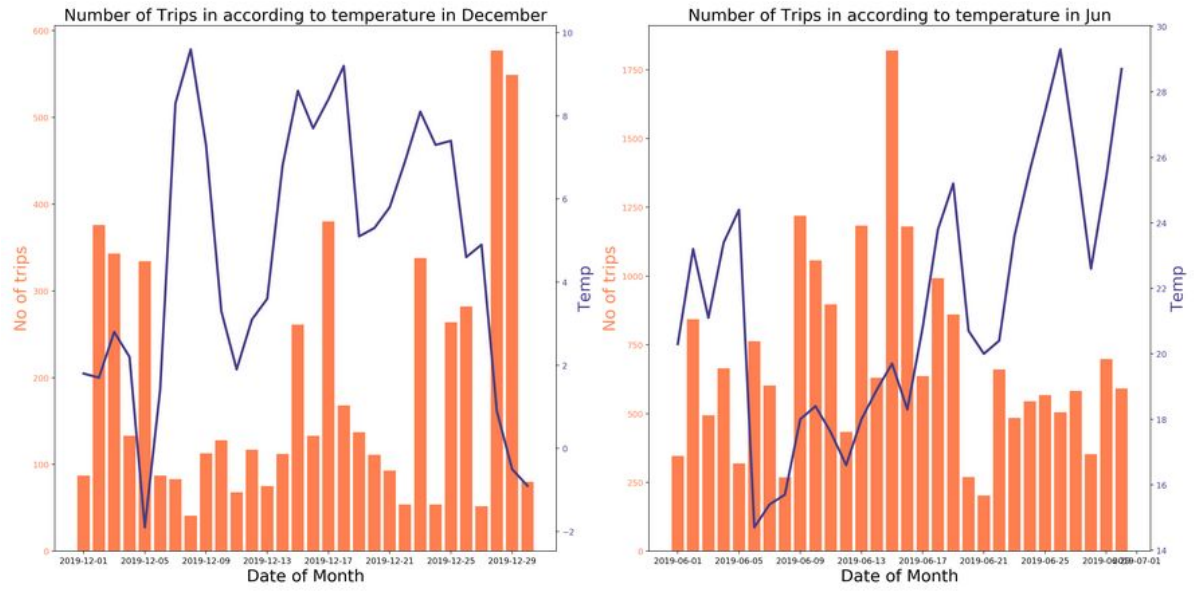
Since one of the research goals is to explore the relationship between weather conditions and bike usage, I collect historical weather data from weatherbit using its free API.

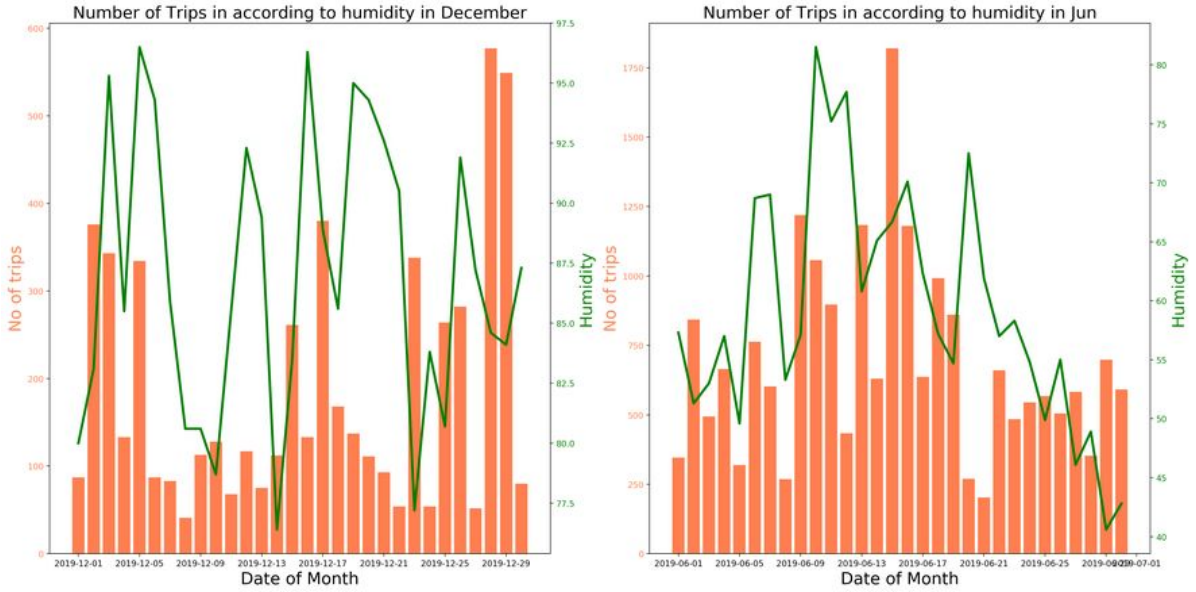
Foursquare API is also employed to collect data on popular locations in the focused city. This information is important to discover the user behavior pattern in regards to locations.

4. Data Analysis

a. Relationship between weather condition and trips made

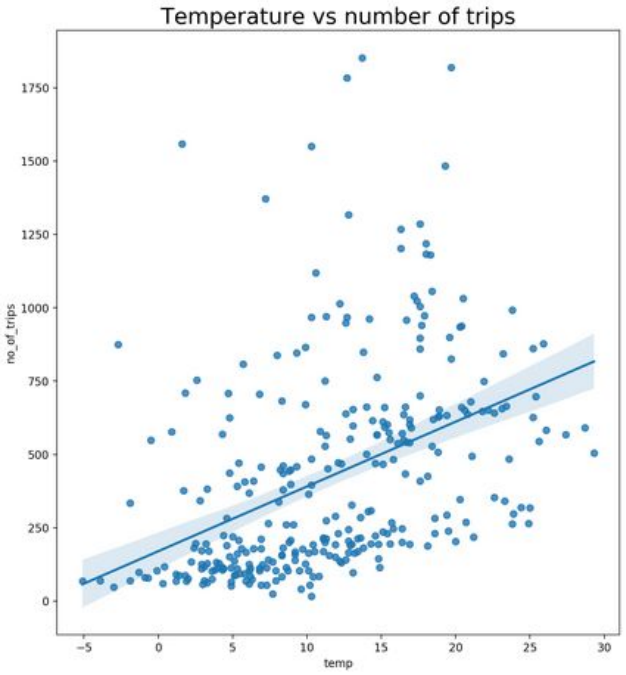
In order to do the analysis for the relationship between weather condition and number of trips made, the study selected temperature and humidity as the variables for weather. To see the relationship, the study started with plotting the data of two months, one in winter and one in summer. Doing this shows not only the relationship the study is aiming to explore but also a direct comparison between two different times of the year.





The charts show there is a negative relationship between number of trips made and temperature. The higher the temperature, the more trips made. Another important factor is humidity. So I use the same method to see the relationship between humidity and the number of trips made. However, it is more challenging to interpret the result for humidity. The humidity level in December fluctuates drastically while the number of trips showed a clear pattern without much fluctuates. Nevertheless, humidity level and number of trips in June show a close positive relationship.

To better explore this relationship, a simple linear regression was performed. The regression results are shown as below



Regression results for Temp vs Number of trips

OLS Regression Results

Dep. Variable:	no_of_trips	R-squared:	0.185
Model:	OLS	Adj. R-squared:	0.182
Method:	Least Squares	F-statistic:	68.63
Date:	Wed, 13 May 2020	Prob (F-statistic):	3.88e-15
Time:	12:37:50	Log-Likelihood:	-2192.1
No. Observations:	305	AIC:	4388.
Df Residuals:	303	BIC:	4396.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	169.9190	35.789	4.748	0.000	99.492	240.346
temp	22.0822	2.666	8.284	0.000	16.837	27.328

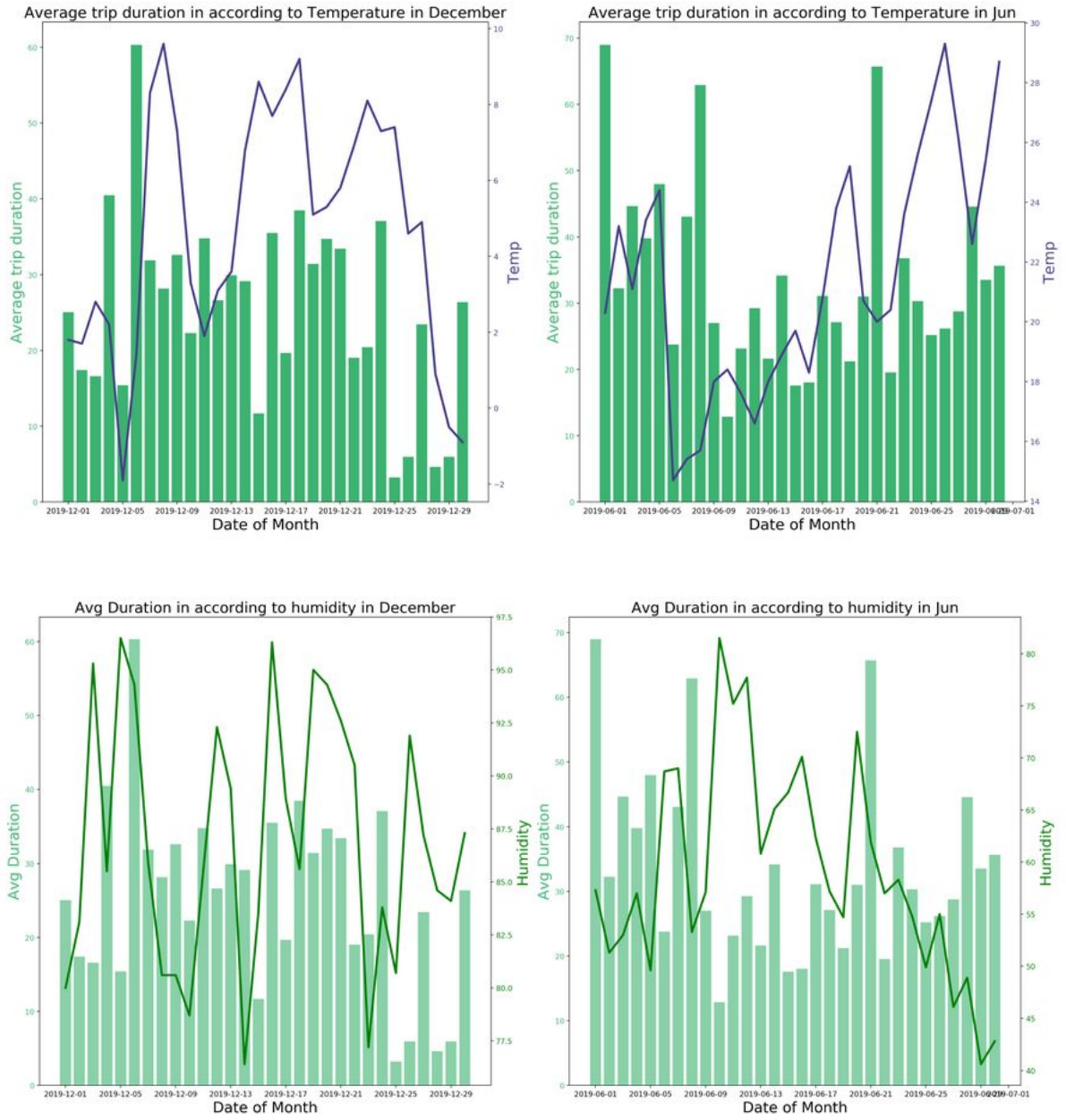
Omnibus:	106.816	Durbin-Watson:	1.664
Prob(Omnibus):	0.000	Jarque-Bera (JB):	286.149
Skew:	1.647	Prob(JB):	7.30e-63
Kurtosis:	6.416	Cond. No.	26.2

Figure: Regression plot and regression results of independent variable Temperature and the dependent variable Number of trips

There is a positive relationship between temperature and the number of trips made. With significant P value and coefficient 22.08, this relationship is strong. As

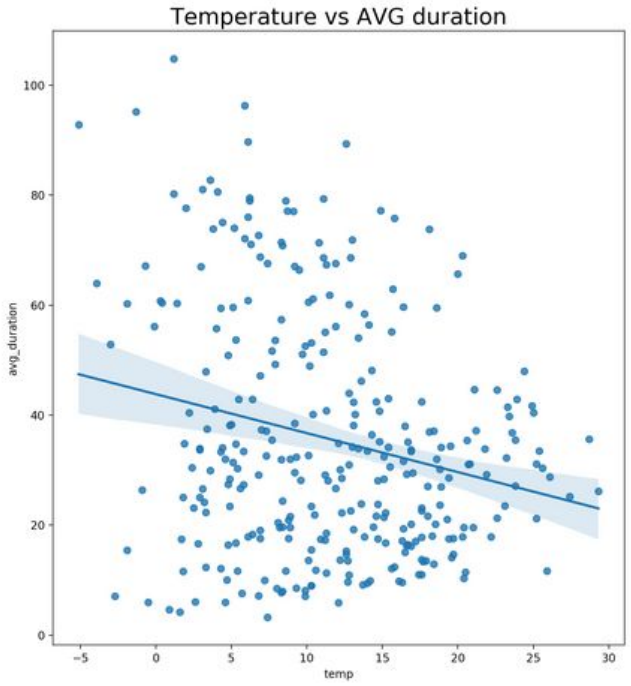
temperature increases, more trips are made. This finding confirms a relatively typical customer behavior that when it is warmer, it is more likely for people to stay outdoors. Also, summer time is a good time for people to stay active after several winter months staying mostly indoor. Considering the slope at more than 22, it is considerably steep trend for temperature and number of trips.

b. Relationship between weather condition and average trip duration



The charts in the first canvas show a positive relationship between temperature and the average duration of each trip. As temperature increases, the longer trips are.

Although the temperature during winter is lower than that in June, the trend holds true for both months.



Regression results for Temp vs Avg Duration

OLS Regression Results

Dep. Variable:	avg_duration	R-squared:	0.050
Model:	OLS	Adj. R-squared:	0.047
Method:	Least Squares	F-statistic:	16.00
Date:	Wed, 13 May 2020	Prob (F-statistic):	7.96e-05
Time:	12:37:19	Log-Likelihood:	-1365.5
No. Observations:	305	AIC:	2735.
Df Residuals:	303	BIC:	2742.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	43.8113	2.380	18.406	0.000	39.127	48.495
temp	-0.7092	0.177	-4.000	0.000	-1.058	-0.360

Omnibus:	16.845	Durbin-Watson:	1.275
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17.355
Skew:	0.549	Prob(JB):	0.000170
Kurtosis:	2.598	Cond. No.	26.2

Figure: Regression plot and regression results of independent variable Temperature and the dependent variable Average duration of trips

Using the same method for the other relationship exploration, the regression results show that there is a negative relationship between the temperature and the average duration of the trips. The coefficient -0.7092 (with significant P value <0.05) indicates the statistically significant relationship.

c. Popular areas for bike usage

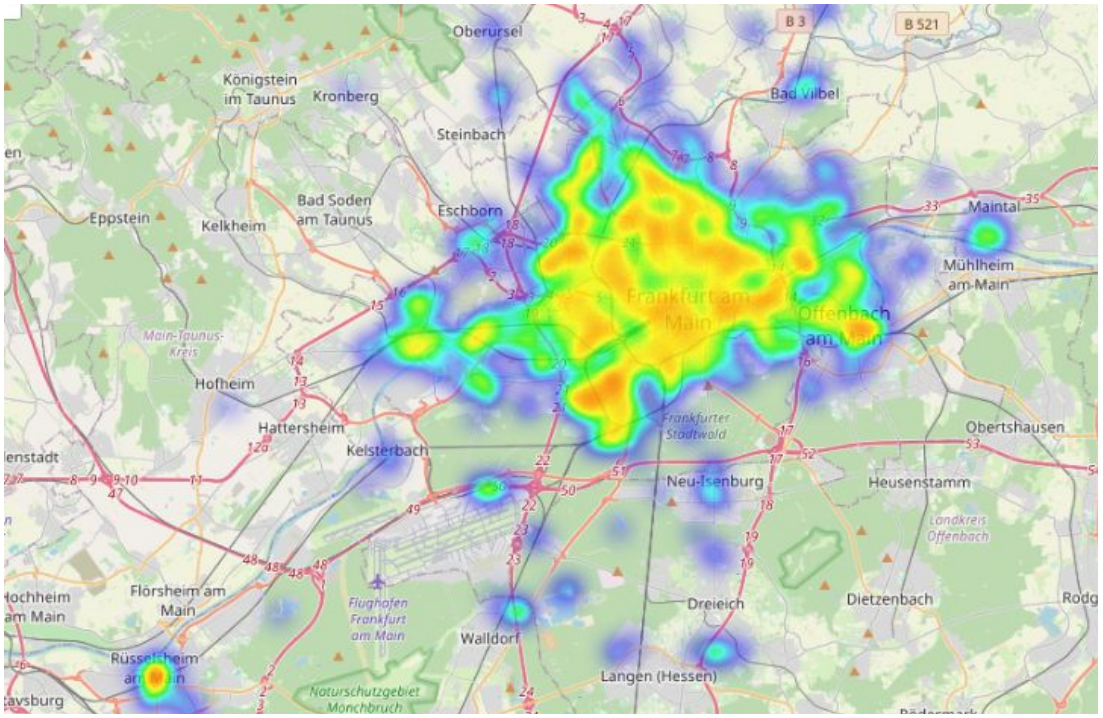


Figure: Heat map of the end position of the trips made throughout the year 2019

From the map, we can see that the city center is the main area where people use the bike. This is not a surprise since the intention of the bike system is for urban mobility.

Because the number of trips in the whole year is quite many, it would be useful to look at the data for a month with less trips to explore where people go. In this study, I take December as the target to study.

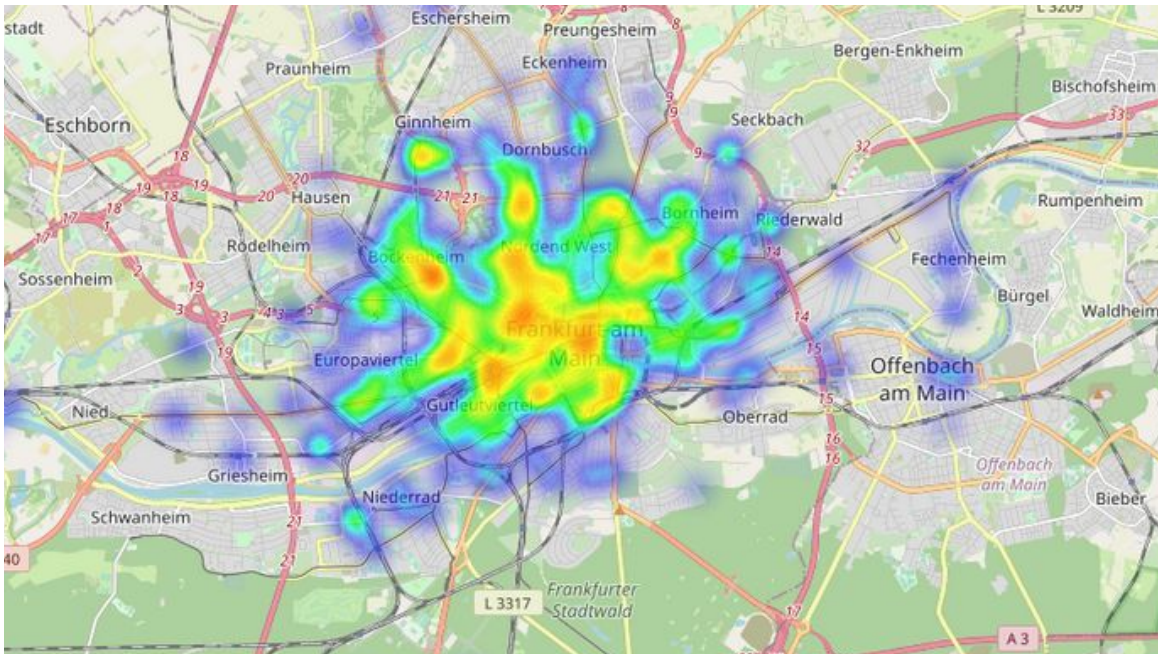


Figure: Heat map of the end position of the trips made in December, 2019

The heat map also shows that people use bikes mainly in the city center. This is again the main purpose of urban bikes as they are designed for urban mobility. The next step is to see the most popular are where people go. To do that, I start with a hypothesis:

> In winter, as the temperature drops, people will be more likely to go outdoors only if necessary. Therefore, locations such as markets, schools and offices would be the places to go.

I use Foursquare to get the locations of the above-mentioned location and plot them on the map to see if they match the heatmap.

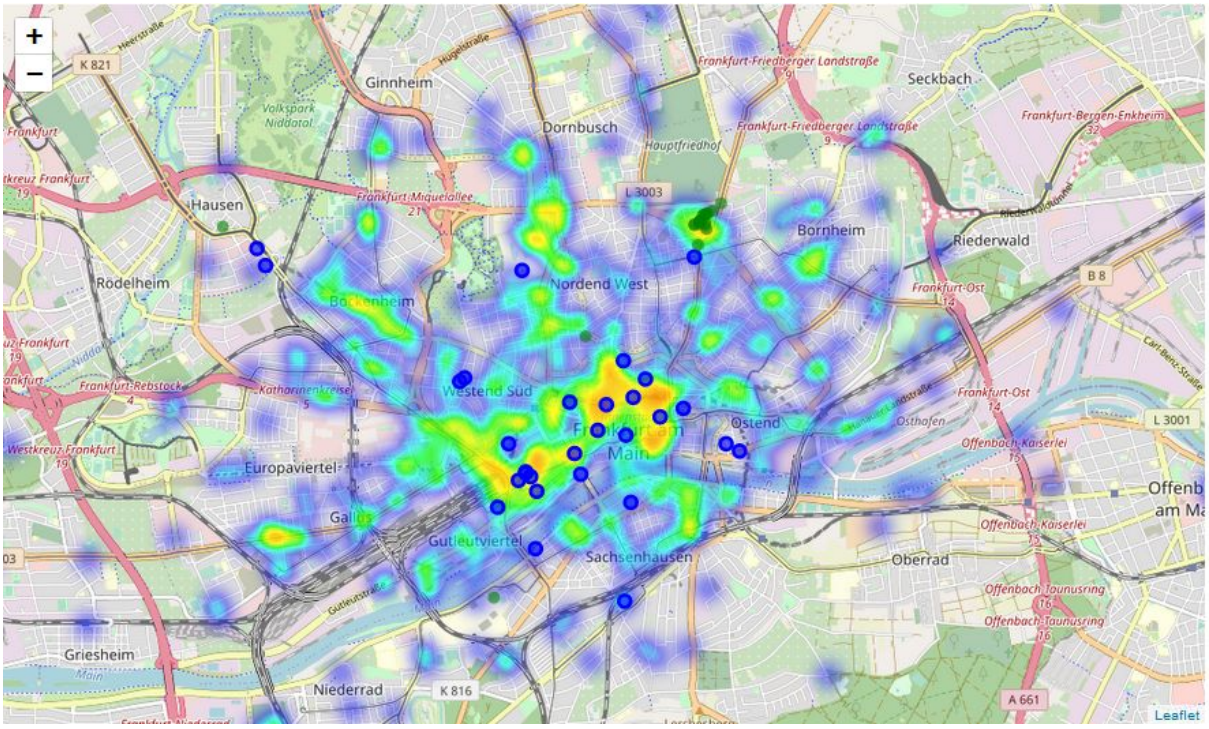


Figure: Heat map of the end position of the trips made in December, 2019.

The blue markers on the map show the markets and some marketing companies as the data took the term 'market' and gave the results with those companies. However, considering that marketing companies are also offices where people go to work, it is acceptable to keep them on the list.

The map shows that the majority of the trips ended at the marketplace. It is worth mentioning that these markets are in the city center, there would be no doubt that there are many other popular places around these markets. But the map someshow demonstrates that users tend to go to these locations.

The green markers are for Universities and colleges location from Foursquare. The map shows that university is another popular location for people who use bikes. As a university can have a contract with the operator to have one of the bike stations on the university campus, it is reasonable that many trips end near the university area. Also, students have special benefits for using bikes. Therefore, it is reasonable to see that university is one of the popular places for bike users.

5. Discussions

The study results reveal interesting behavior patterns of bike users. The regression results show statically significant relationship between the number of trips made and the average duration of the trips against weather conditions.

- The results revealed that there is a statistically significant positive relationship between the number of trips and temperature. As temperature increases, more trips are made. This is consistent with the hypothesis that during warmer periods of time, people can go outside more often and thus bike usage increases. Also, during winter, other means of transport such as buses, personal cars, etc. are more prefered. Therefore, it could be that after an interactive period of the year, people can take bikes as a means to stay active. One suggestion can be proposed for the bike operator is to withdraw bikes from the station during the cold period. Doing this shows less negative impact on the operation of the firm. Withdrawing the bike could help the operator to maintain the bikes. Also, withdrawing bikes helps to prevent bikes from being affected by the harsh weather.

- On the other hand, the results show a negative relationship between temperature and duration of the trips. As the temperature increases, the duration of the trips decreases. This could be explained by two reasons. First, during warmer seasons, people who do not frequently use bikes as their main means of transportation which to use bike. These users might have used the bike for short trips. Second, as the temperature increases, it might be less comfortable to ride bike outdoor. Thus, duration of the trips dropped. Moreover, the plot chart shows a range of temperatures from above 0 to less than 25 degrees that the duration of trips reaches the highest time. This finding is consistent with the hypothesis mentioned at the beginning of the study.

Another finding of the study is the area of bikes used. The map shows that the city center is the main area for bike users. This is consistent with the purpose of the bike sharing system that its main purpose is for urban mobility. Markets and schools are some popular areas where people go to use bikes. One suggestion for the operator is to introduce a new price package for those who commute from home to market or

from home to school using a fixed station. This could help to increase the bikes available in the most popular location.

6. Conclusion

The main goal of this study is to create an initial view on customer behavior in using public bikes via bike sharing platform. The results show that weather condition is a factor affecting customer demand to use bike. By analyzing the correlation between bike trips and weather conditions, the study suggests a way for the bike operator to better perform bike maintenance tasks so that it could avoid the mistake from Chinese bike operators. Moreover, the decision to put more bikes into the market could be decided based on the seasonal study of bike trips. Finally, This study set the first step to study the area of bike usage. More study can be done to explore more popular locations. The results of the future study would help the operator to run relevant marketing plan to leverage its business.