Tingting Huang

Work Sample

Bike-share Spatial-temporal Analysis in Philadelphia

Abstract

Bike-share systems have been applied in many cities around the world. The bike-share system is an important part of public transport, serving as a "last mile" connection to public transit. While affecting the daily travel patterns of urban residents, bike-share is also slowly reshaping cities. The key issue of bike-share operation and management is how to rebalance the bicycles in each station to meet demand over time. This paper shows the spatial and temporal patterns of bike-share use in Philadelphia based on Indego origin-destination data, followed by an analysis of the impact of factors like weather, time and space in the prediction capabilities of the time-series model. The finding shows that more trips are taken in summer, and that travel patterns on weekdays and weekends are different. There is a clear ridership peak during morning and evening rush hours on weekdays. Many people take bike-share trips to travel the first and last miles for their daily commutes by public transit. The regression model suggests that there is a correlation between time, space, weather conditions and bike-share trips.

Keywords: bike-share program, spatial analysis, demand prediction

Introduction

It has been over 50 years since the first bike-share program launched, but in the last decade the prevalence and popularity of bike-share was increased significantly around the world (Shaheen, Guzman, & Zhang, 2010). In the bike-share system, people can borrow bikes via mobile app from a bike-share docking station paying a small fee, ride to their destination and return the bikes to the bike-share docking station near the destination. The evolution of the bike-share system has been categorized into four generations (S. Shaheen, Martin, & Cohen, 2013). The first public bike-share program started in Amsterdam as far back as the late 1960s. Since anyone could use the bikes without any control, bikes were damaged and stolen, finally, the program failed (DeMaio, 2009). In the 1990s, the bike-share system improved customer tracking. People needed to provide identification and a deposit to use a public bike. Since 2005, bike-share programs have combined with information technology to provide a more convenient rental process. In 1998, the first bike-share scheme using smartcard came out in France. After that other systems using new technology began to develop worldwide (Lyon, 2005). Today, many innovative technologies are applied in rebalance and route management in the bike-share system, bike-share is becoming increasingly important in urban transportation (S. Shaheen et al., 2013; S. A. Shaheen, Guzman, & Zhang, 2010).

There is a growing trend in cycling in Philadelphia. As shown in Figure 1, the percentage of trips made by bike experienced significant growth from 2005 to 2017 according to the American Community Survey. The bike commuting mode share increased from 0.9% to 2.6% during the decade. Commuting by bike continued to become popular. The improvement of bike lanes is one of the reasons that cycling has gotten popular. PBC promoted the construction of bicycle lanes on the major arteries of Spruce and Pine Streets in 2009, and we can see a dramatic increase at that time. Additionally, the bike-share system largely contributes to the cycling increase after 2015.

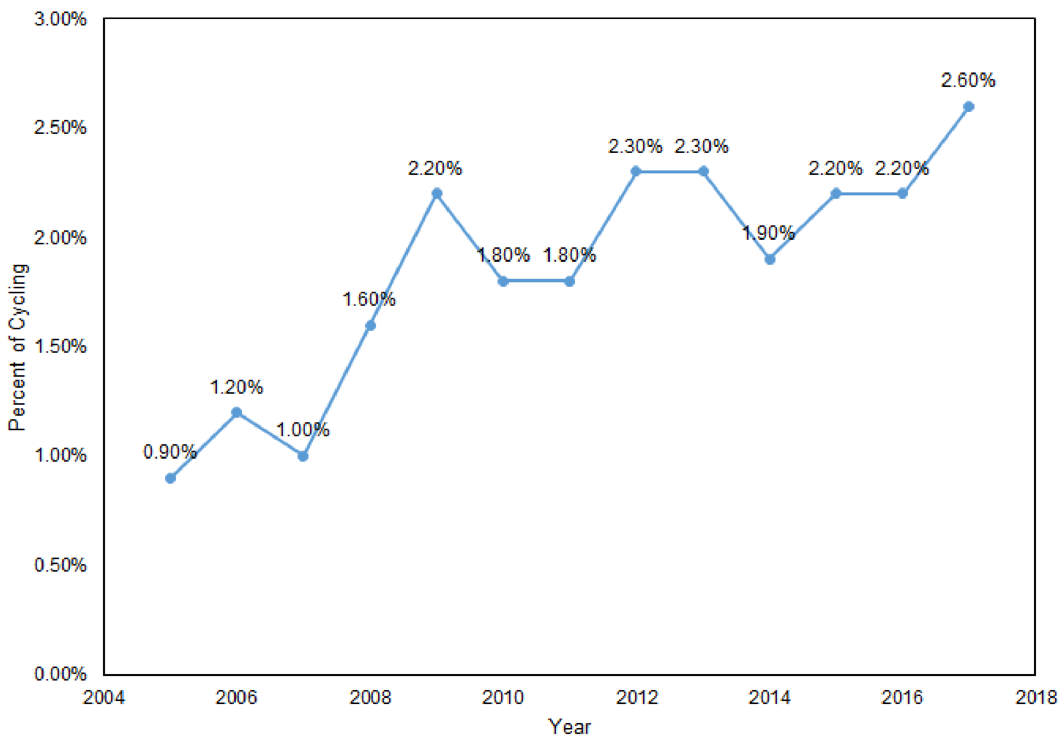

Figure 1: The trend of commuting by bike, Philadelphia, 2005 - 2017

Data Source: American Community Survey, 2005 – 2017

Indego is the first public bike-share program in Philadelphia. It was launched in 2015. Bike stations were selected based on proximity to public transit, workplaces and community resources. The service area of the bike-share system covers the main commercial district, employment center and residential area of Philadelphia. More than one thousand bikes were added into the system in 2019. The bike-share program together with the expansion of the bike lanes will lead to an increase in cycling, and at the same time simulate the transformation from driving to other more sustainable modes. To satisfy the increasing trend of cycling and demand in public bikes, more public bikes and bike stations should be added. Through the analysis of spatial-temporal cycling, we can have a better understanding of high demand location and high demand time, which is helpful for bike share rebalancing. This paper explores the relationship between demand for bike-share trips and time, weather and space. Finally, we will discuss the role Philadelphia's bike-share system plays in daily commute and give advice on bike station allocation.

Literature Review

Although many European countries and America have encouraged bicycle transportation and developed public bicycles program since the late 1960s, research on public bicycle sharing began in the early 2000s, mainly focusing on the management and operation of sharing bicycles, spatial-temporal patterns of bike-share trips, the impacts of bike-share on people's travel behavior as well as the relationship between bike-share and public transit. Also, numerous studies are concentrated on the demographics of bike-share users and their motivations, preferences and riding purposes. However, despite the growing interest in bike-share, research on bike-share system and travel mode changes in big cities is still limited. A limited number of studies have visualized the spatial-temporal characteristics of bike-share and analyzed cycling flow between stations. Few studies try to predict the supply-demand balance of bike-share stations. The key to operating a bike-share system successfully is effectively rebalancing bike based on demand. How to address the rebalance problem and optimize the bike-share system are current important issues.

Many studies explore the issue of mode substitution in bike-share programs and how bike-share systems influence people's travel choices. Some survey reports also indicate that people use more public transit when bike-share is available. Yang et al. found that more than 50% of people use bike-share to link the metro systems with their desired destinations in Beijing and Shanghai, China (Yang et al., 2011). The possible relationships between public transit and bike-share include complementarity, substitution or no relationship (Campbell et al., 2017). Fishman et al. found that the variables that are related with public transit can predict bike-share ridership since bike-share connects users to public transit (Fishman et al., 2014).

Thanks to the application of information technology in bike-share systems, we can collect trip data, including specific time and location data, to study the spatial-temporal pattern of bike-share. Etienne and Latifa used departure and arrival counts in each station to explore bike-share patterns in Paris by Poisson mixtures based model and clustering techniques (Etienne & Latifa, 2014). Zhou visualized flow clusters and use hierarchical clustering to show the spatial spatial-temporal patterns in Chicago (Zhou, 2015). Previous studies have found that weather conditions influence cycling patterns (Heinen et al., 2011; Brandenburg et al., 2007). Corcoran et al. found that weather and calendar events have an impact on the geographic and temporal patterning of bike-share trips in Brisbane and reveal the demand and availability in different times and spaces (Corcoran et al., 2014). In addition to the weather, some studies have confirmed that the time of day and the location of bike stations are related to the demand for shared bikes. We can use time and space data to forecast the demand for shared bike. Froehlich et al. applied the clustering method to investigate user behavior in the Barcelona bike-share system, and conclude that simple prediction models are able to predict station usage and identify station state (Froehlich et al., 2009). Faghih-Imani et al. used data from New York City's bicycle-sharing system to explore the spatial and temporal effects on modeling bicycle demand. The results show that there is spatial and temporal dependency for bike station's departure and arrival rates (Faghih-Imani, 2016). Time series analysis is an efficient way to determine the factors influencing seasonal and short-term bike-share trips (Vogel & Mattfeld, 2011).

Data

The data used in this paper are bike-share trip data, weather data and demographic data in census tract level. This paper uses trip data for the whole year of 2018. The trip data was derived from the Indego website (https://www.rideindego.com/about/data/). Philadelphia government and Bicycle Transit System, a private company that manages the bike-share system in Philadelphia, share anonymized Indego trip data with the public. The dataset provides bike-share trip data from April 2015 and updates the trip data every three months. The data includes origin and destination stations, start and end times of trips, and trip duration. Also, the station location data are provided on the website. The data includes station name, station ID and status etc. For the weather and temperature data, the paper assumes the weather and temperature in the study area are the same as that in Philadelphia airport. There are free and open weather data APIs available in the R community, which provides precipitation, temperature and wind data. The demographic characteristic data were gathered from 2017 ACS estimates.

Methodology

Method for Spatial-temporal Pattern

The first step of the process was to download bike trip data, weather data, and social-demographic data. To ensure the accuracy of the data and the reliability of the analysis results, it was necessary to eliminate interference data before analysis, so the next step was to clean the trip data. I deleted the trip records without specific origin and destination location or time. After data cleaning, 66,371 riding trips were used in the study. Then, the paper shows the overall bike-share temporal pattern in 2018 through summarizing riding records in each day. It gives us an overall understanding of how riding trips change in different seasons. Next, the study compares the change in hourly trips per station at a different time in one day by aggregating the data that are at the same time in 2018. Usage differences between weekdays and weekends are taken into account in the paper. To show the difference of trip patterns in weekdays and weekends as well as the comparison between different days in a week, I aggregated the trips on weekdays and weekends separately, and the trips happened on Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday, respectively, and then show the differences through line chart.

To show the spatial patterns of bike-share trips in Philadelphia, I randomly selected a weekday and a weekend day to analyze. Since we only want to see how time influences the distribution of riding trips in space, we need to control other variables. The two days should have similar weather and temperature condition to compare the influence of time only. The first day is June 16th, 2018. It was a sunny Saturday with moderate temperatures. Another day is June 20th, 2018, which is a sunny Wednesday, also had a similar temperature as June 16th. Density analysis can clearly show the spatial distribution of riding, so I did kernel density analysis for bike-share trips in two days. To see how the riding varied in peak hour and non-peak hour, I selected trip data from three periods on June 20th for analysis. Considering the season of the study days, I assumed the AM peak hour is from 7 to 10 am, and the PM peak hour is from 5 to 8 pm, the non-peak hour is other time of the day. Since the travel activities in peak hour and non-peak hours are largely influenced by work, there may not be an obvious peak on the weekend, I just selected the weekday to analyze the difference of spatial distribution in different time of a day.

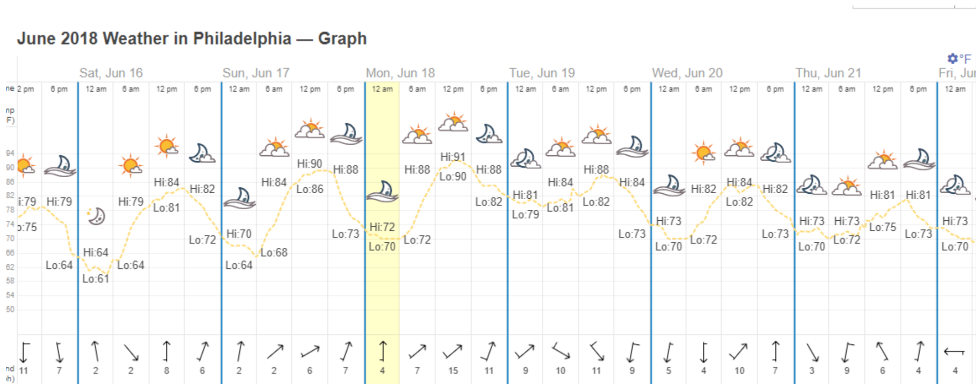

Figure 2: Weather, Temperature and Wind in the Study days

Source: https://www.timeanddate.com/weather/usa/philadelphia/historic?month=1&year=2018

Method for Demand Prediction

I used the time-series model to predict bike-share trips. The prediction was based on the assumption that the bike-share ridership in the future is highly correlated with the demand at the same time in the past. Then, we can use trips that happened in the past to predict the future demand for bike share trips. As an overview of the demand prediction method, the first step is to create space-time panel data. We need to create a panel dataset that includes every unique station and the number of trips each hour every day. Each period is represented by a row, as for the station that didn't have any riding trips starting or ending, I still put zero to describe the demand condition at each period.

Some studies have proved the presence of time-lag in bike-share prediction (Vogel & Mattfeld, 2011). Because of the autocorrelation of time series, the model tends to take the true value of the previous time as the predicted value of the next time, causing the true value lag behind the predicted value. For example, holidays would influence the demand for days before and after it. So I create several time lag variables to compare the prediction results between linear models without time lag and linear models with a time lag. Next, I split the data into two datasets: one is training data and the other is test data. The training set includes data from January to June (data in June are not included). The test set is data on which trips happened in June, and the total number of the test data set is approximately 20% of the training data set. Then I ran three linear models using the training dataset to see the relationship between the bike-share ridership and time, weather and location, and used the regression model to predict trips in June. Finally, I calculated the mean absolute error of the models to assess the prediction accuracy.

Analysis and Discussion

The_Spatial-temporal Properties of Bike-share in Philadelphia

There are seasonal variations in bike-share trips in Philadelphia. As Figure 2 shows, the bike-share trips have seasonal variation: the average numbers of riding trips are the highest during summer and fall. Bike share trips decrease dramatically in winter. Unlike driving, cycling tends to be influenced by weather conditions. We can see the trend of temperature throughout the year coincides with the change in bike share trips: as the temperature becomes lower, less bike-share trips occur. This makes sense since people are not likely to ride outside in cold weather. In winter, extreme weather is more likely to happen, people travel less so the bike-share ridership reduces accordingly.

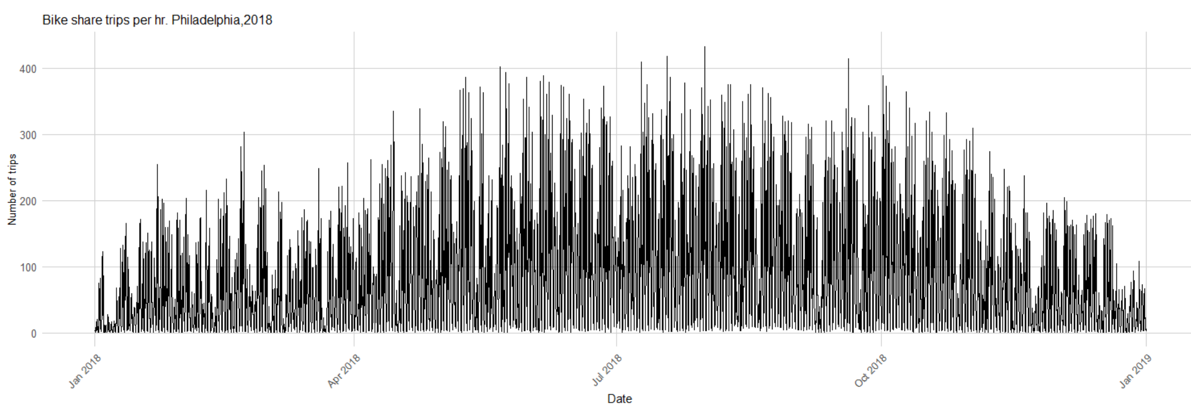

Figure 3: Bike Share Trips per Hour, Philadelphia, 2018

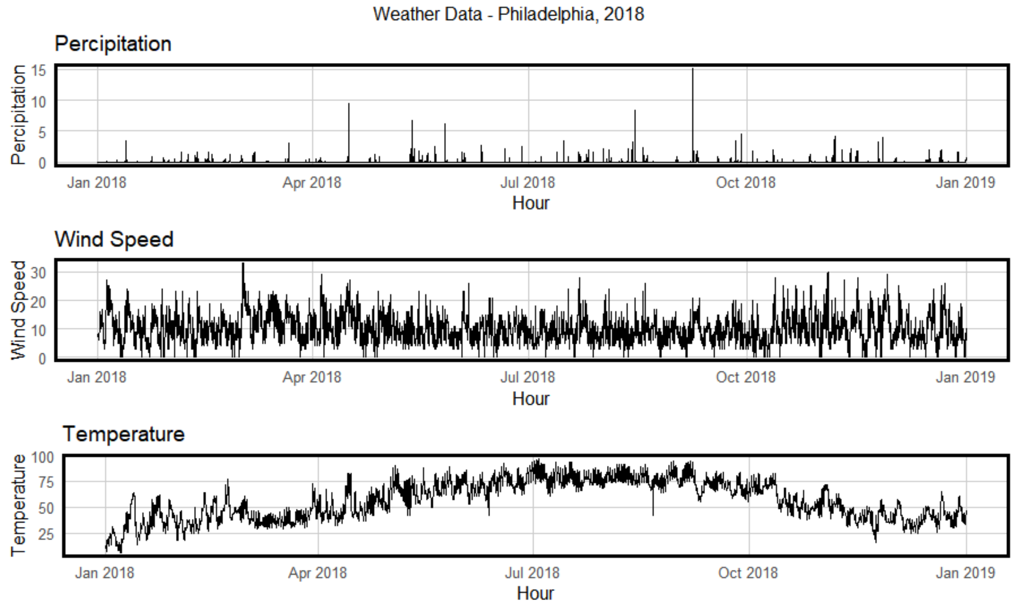

Figure 4: Weather Condition, Philadelphia, 2018

Bike-share system services are an important tool for daily commutes on weekdays, because there is a large demand for shared bikes during rush hour. There is a clear difference in trip trends between weekdays and weekends (Figure 5 & Figure 6). The total number of bike-share trips on weekdays is larger than that on weekends. We can see bike-share trips peak at the morning and evening rush hour periods during the weekday, which indicates that the bike-share system serves as an important part of commuting to work. Besides, the bike-share peak on weekdays differs by morning and evening. The trips increase rapidly in the morning rush and peak at 8:00 am and decline rapidly. At night, however. trips decline faster after evening peak hours than in the morning. There is a larger demand for shared bikes in the evening peak than at any other times of the day. Finally, there is a slight increase in riding volume at noon, which may be due to riders traveling for lunch.

Bike share trips on weekends are mainly for non-work travel, and there are no significant morning or evening peaks. As we can see that the number of bike-share trips keeps steadily increasing from 10:00 am and peaks in the afternoon and evening. This single peak is due to most people traveling for recreational activities on the weekend, and there is no fixed time for starting and ending for these activities.

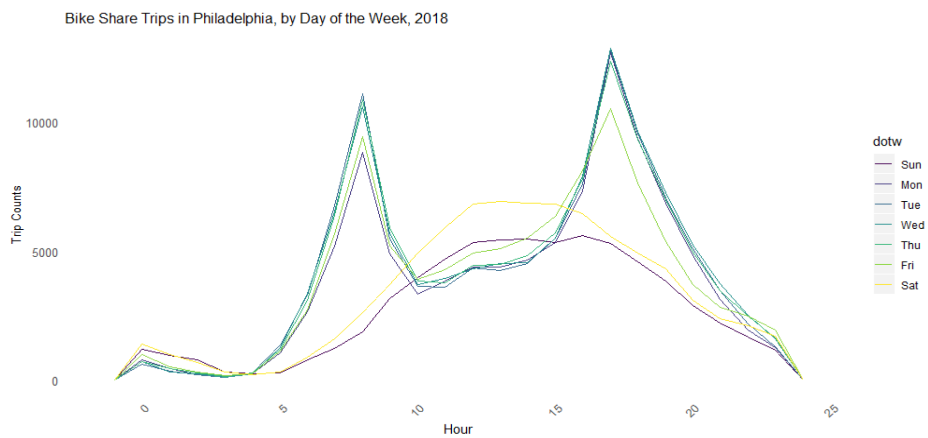

Figure 5: Bike Share Trips by Day of the Week, Philadelphia, 2018

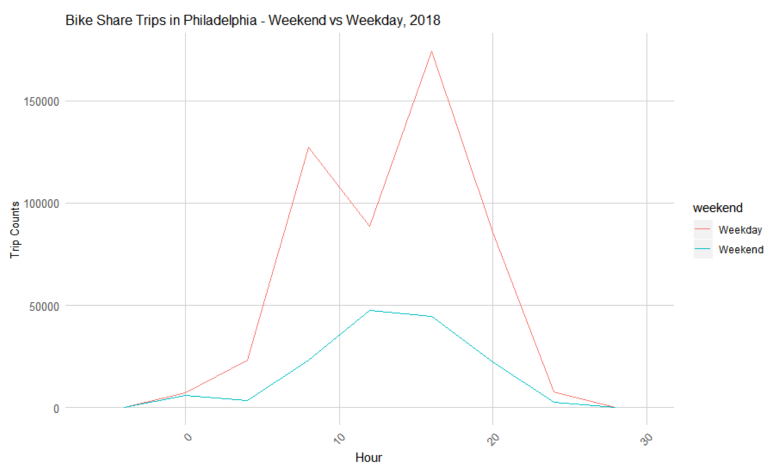

Figure 6: Bike Share Trips on Weekdays and Weekends, Philadelphia, 2018

The majority of the bike share trips are short-haul. As we can see from Figure 6, most bike-share trips take between 5 to 15 minutes. The bike stations are concentrated in center city, so station to station distance not very far. On the other hand, since public bikes usually serve as a connection to public transit, we don't need to spend a lot of time traveling the "last mile" by bike-share in places with good public transportation, such as central Philadelphia. We can conclude that bike-share is an important tool for short-distance travel.

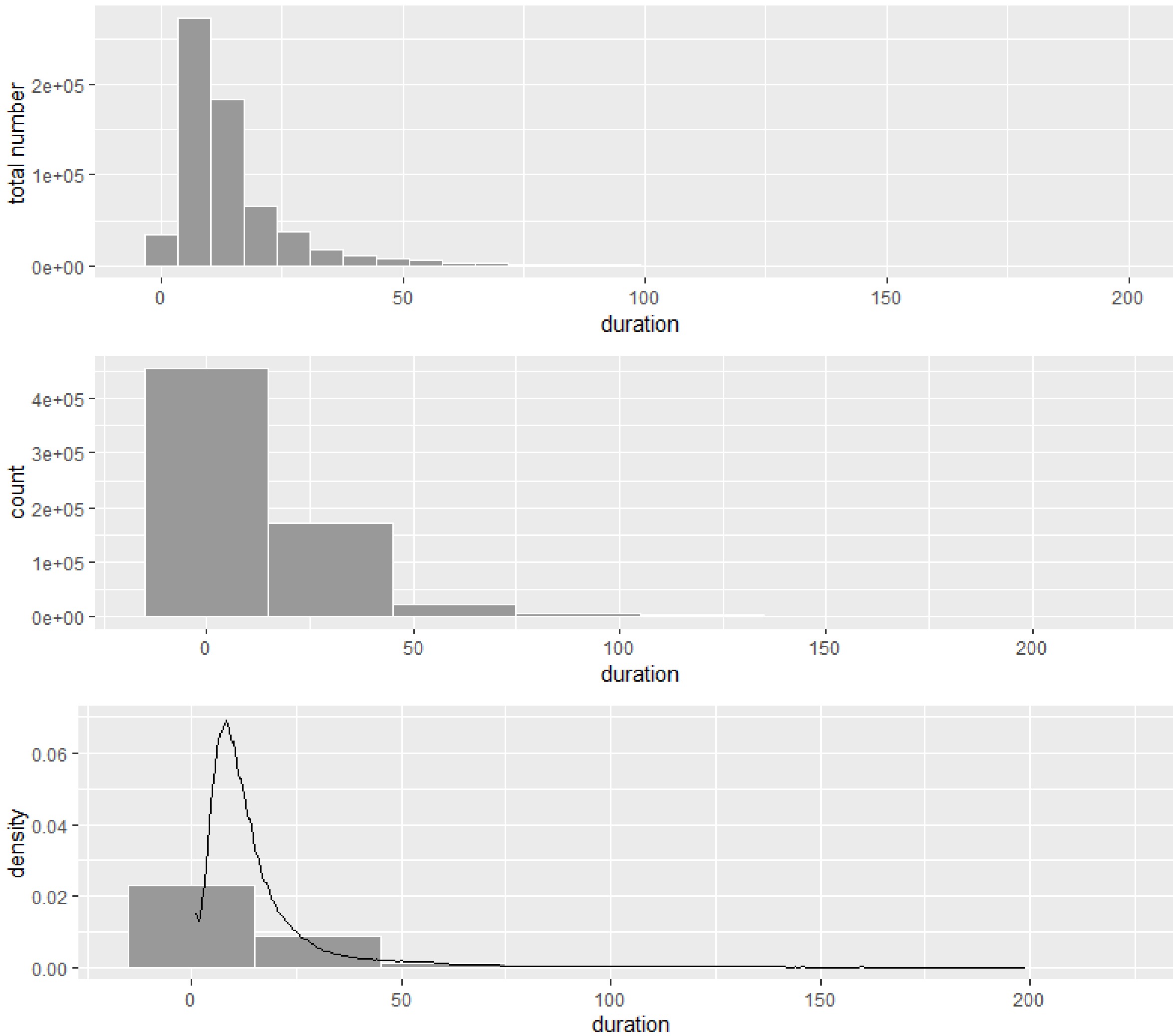

Figure 7: Distribution of Cycling Duration, Philadelphia, 2018

Weekday bike-share hot spots are concentrated in business and employment centers, and bike-share trips in recreational places and tourist attractions increase significantly on weekends. To explore the spatial pattern of bike-share trips, I randomly selected a weekday and a weekend to do a comparison. Since the weather significantly influences cycling, I selected the sample from sunny days. By doing kernel density analysis, we get to know where there is a high demand for public bikes. Whether on weekday or weekend, most bike-share trips happen in Center City which is the center of commercial business and employment, bike-share hot spots formed around public stations. But the spatial distribution of trip origin and destination are different between weekdays and weekends. On the weekday, there are three hot spots for bike-share trips in the city center. One is along Market Street, the main arterial of Philadelphia, and connects with another hot spot in Rittenhouse square. Also, many trips started around the

Kimmel Center, which is a major theater in Philadelphia. These places are in commercial corridors, with a large number of retail stores and, office buildings nearby, which leads to heavy traffic. Bike-share hot spots for the weekend are not as centrally concentrated in the city center as on weekdays. A clear bike-share hotspot was formed around the Philadelphia Museum of Art, and bike-share trips increased significantly in tourist attractions and Parks, such as the Barnes Foundation, the Liberty Bell, and Philadelphia's Magic Gardens. We should consider increasing bike supply in these high-demand locations.

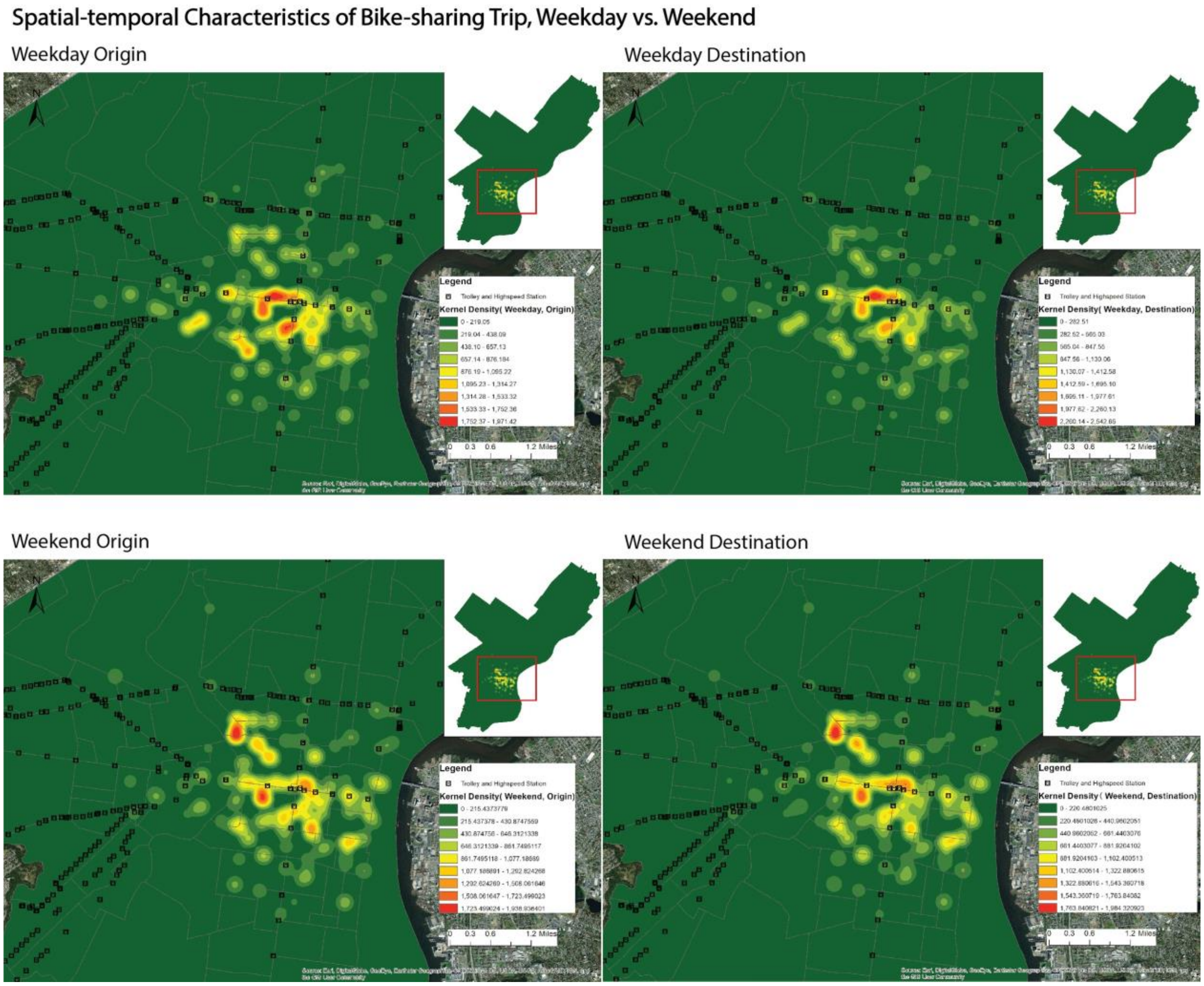

Figure 8: Spatial-temporal Characteristics of Bike-share Trips, Weekday vs. Weekend

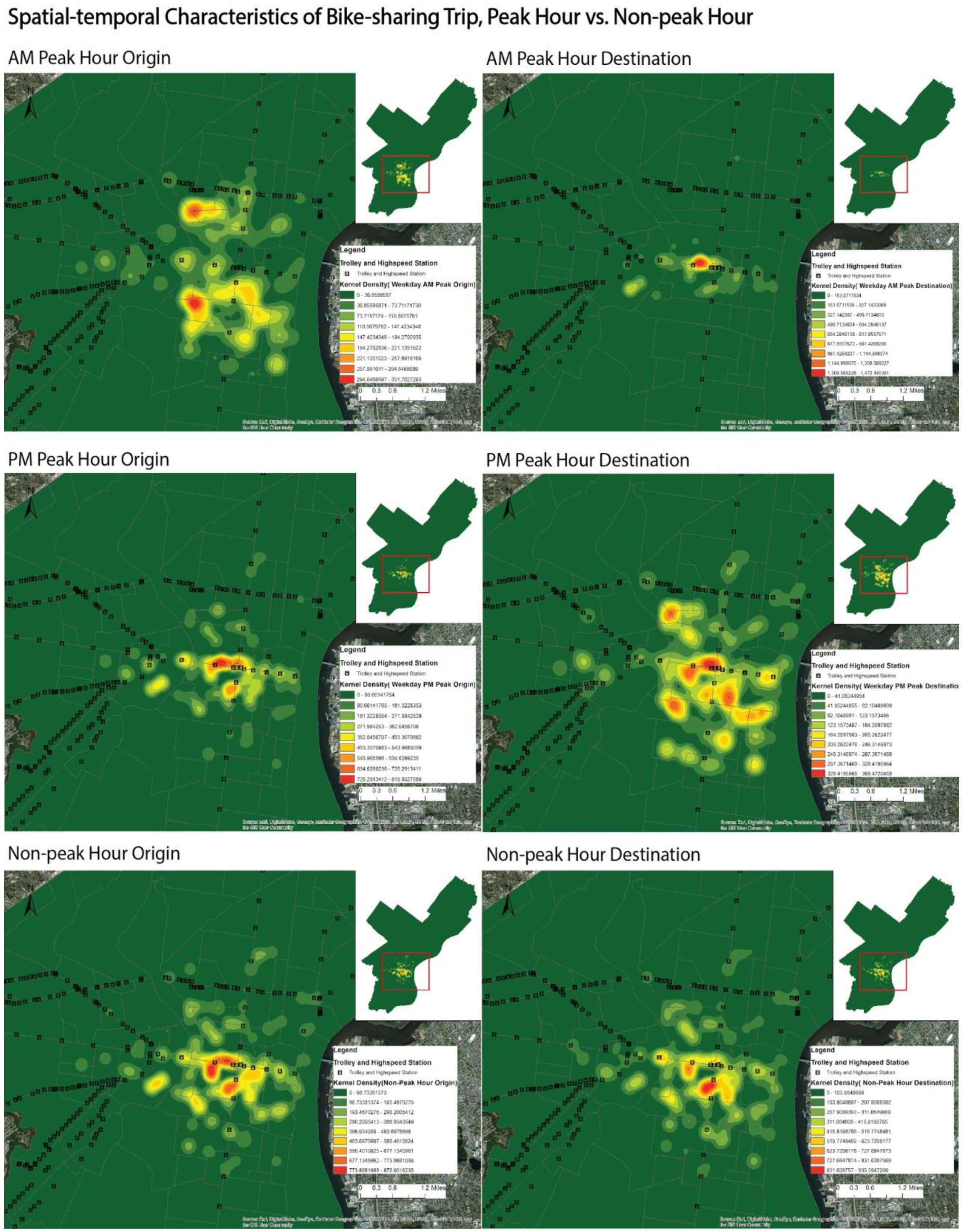

Figure 9: Spatial-temporal Characteristics of Bike-share Trips, Peak Hour vs.Non-peak Hour

The bike-share system connects users to public transit for the first mile and last mile of commuting. There are obvious morning and evening bike-share peaks on the weekday, so I did kernel density analysis for rush hour and non-rush hour to see the distribution of the trip's origin and destination at a different time in a day. As shown in Figure 9, the hot spots of the bike-share origin are concentrated in the residential areas outside the Center City in AM rush hour. Most of the trips started from the residential areas in Spring Garden and Fitler square. There are no trolley or subway stations near the Origin hot spots. The cycling destinations, however, are concentrated near City Hall. There are many public transit stations here, and it is also the employment and commercial center of Philadelphia. We can conclude that most people take bike-share trips to get to public transportation stations faster, and the bike-share system serves as the connection between home and public transit, and many people take bike-share trips for daily commuting from home to workplace at AM rush hour.

At PM rush hour, most of the bike-share trips started from public transit in the Center City, such as the 19th St trolley station and the 13th St subway station. The destinations are diversified and scattered, including the residential area outside downtown, commercial districts, and recreational places. Additionally, many trips ended at transit stations that were far from the Center City. We can conclude that some people take a bike-share trip for daily commuting from workplace to home as well as recreational activities after work. Also, the bike-share system helps users to reach their final destination from a public transportation station. Finally, the distributions of bike-share origin and destination are similar at the non-peak hour. Most of the bike-share trips started and ended in the Center City.

Bike-share Demand Prediction Based on Time-series Model

Through the previous analysis, we get to know that the bike-share ridership is highly fluctuant with the hour of a day, the day of the week and the weather condition. Therefore, we selected the hour of a day, the day of the week, the temperature and precipitation to predict the demand for bike-share demand. The first models include spatial-temporal and weather controls, the other two regression models contain all of the lag information. The regression models are shown below:

Regression model 1:

$$Trips = a_1\text{Start station} + a_2\text{The hour of a day} + a_3\text{ The day of a week} + a_4\text{Temperature} + a_5\text{Precipitation} + a_0$$

Regression model 2:

$$Trips = a_1\text{Start station} + a_2\text{The hour of a day} + a_3\text{ The day of a week} + a_4\text{Temperature} + a_5\text{Precipitation} + a_6\text{LagHour} + a_7\text{Lag2Hours} + a_8\text{Lag3Hours} + a_9\text{Lag12Hours} + a_{10}\text{Lag1day} + a_0$$

Regression model 3:

$$Trips = a_1\text{Start station} + a_2\text{The hour of a day} + a_3\text{ The day of a week} + a_4\text{Temperature} + a_5\text{Precipitation} + a_6\text{LagHour} + a_7\text{Lag2Hours} + a_8\text{Lag3Hours} + a_9\text{Lag12Hours} + a_{10}\text{Lag1day} + a_{11}\text{LagHoliday} + a_{12}\text{Holiday} + a_0$$

The regression results show that time, space, weather conditions, and time lag have a significant relationship with the demand of bike-share trips (see in the appendix). The MAE of regression models that contain time lag is better than the regression model 1. However, the overfitting of the three regression models is not good. The adjusted R square of the three regression models is 0.072, 0.313, 0.313, respectively.

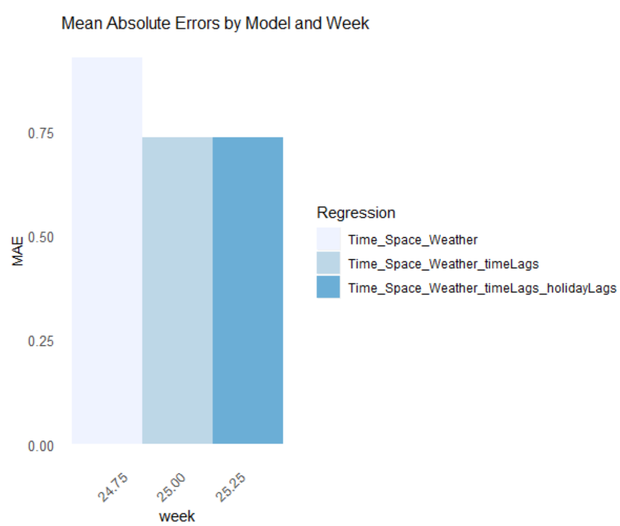

Figure 10: The MAE by Regression Model and Week

As shown in Figure 11, the time-series regression model can predict the bike-share demand in normal conditions well but unable to predict the peaks very well. The trips in high demand periods are under predicted, so the regression model still needs improvement. However, the regression models are not worthless. Adding the time-lag variables significantly increases the accuracy of the prediction model. This means that there is a time-lag effect on bike-share demand. To improve the prediction model, we can consider adding more space variables, social and demographic variables. The high bike-share ridership often related to public transit, so some variables that reflect the positional relationship between the bike docks and public transit.

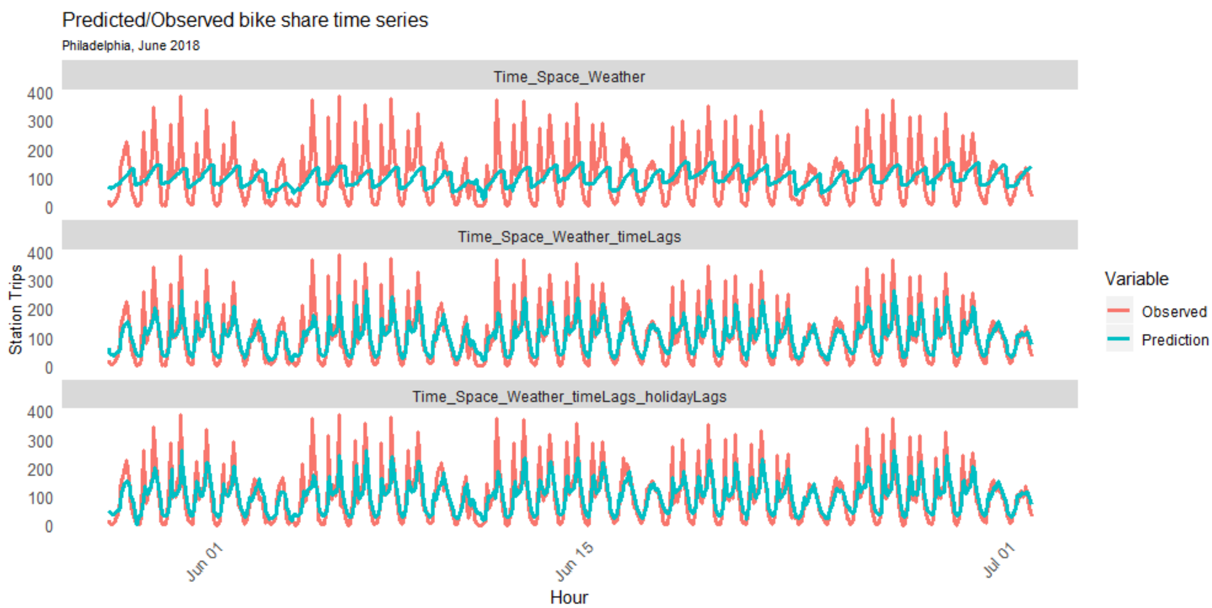

Figure 11: The Prediction Value and Observed Value of Bike-share Trips

Conclusion

This paper analyzed the spatial-temporal patterns of bike-share trips in Philadelphia based on Indego origin-destination data. The findings show that there are obvious seasonal variations in bike-share trips in Philadelphia. People are more likely to take bike-share trips in summer. There is a large demand for shared bike during rush hour. The demand is higher during PM rush hour than AM rush hour. We should improve rebalancing efficiency in rush hour on the weekday to satisfy the high demand for the shared bike at that time.

Most of the bike-share trips happen in the commercial business and employment center. People take bike-share trips for commuting on the weekday. The Bike-share system is an important tool for short-distance travel. It solves the last mile problem by providing links between public transit and the desired destination. Therefore, the government should take public transit into account when setting a bike dock location to create a more efficient multi-modal transportation system.

Bibliography

References:

DeMaio, P. (2009). Bike-share: History, impacts, models of provision, and future. Journal of public transportation, 12(4), 3.

Shaheen, S., Martin, E., & Cohen, A. (2013). Public bikesharing and modal shift behavior: a comparative study of early bikesharing systems in North America.

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia. Transportation Research Record: Journal of the Transportation Research Board, 2143, 159-167. doi: 10.3141/2143-20

Vincent Fraley. Bicycles, 2017 https://philadelphiaencyclopedia.org/archive/bicycles/

Tang, Y.; Pan, H.; Shen, Q. Bike-share systems in Beijing, Shanghai, and Hangzhou and their impact on travel behavior. In Proceedings of the Transportation Research Board 90th Annual Meeting, Washington, DC, USA, 23–27 January 2011; No. 11-3862.

Campbell, K.B.; Brakewood, C. Sharing riders: How bike-share impacts bus ridership in New York City. Transp. Res. Part A Policy Pract. 2017, 100, 264–282.

Fishman, E.; Washington, S.; Haworth, N.; Mazzei, A. Barriers to bikesharing: An analysis from Melbourne and Brisbane. J. Transp. Geogr. 2014, 41, 325–337.

Heinen, E., Maat, K., & van Wee, B. (2011). Day-to-Day Choice to Commute or Not by Bicycle. Transportation Research Record, 2230(1), 9–18. https://doi.org/10.3141/2230-02

Brandenburg, C., Matzarakis, A., & Arnberger, A. (2007). Weather and cycling—a first approach to the effects of weather conditions on cycling. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 14(1), 61-67.

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. Journal of Transport Geography, 41, 292-305.

Etienne, C., & Latifa, O. (2014). Model-based count series clustering for bike sharing system usage mining: a case study with the Vélib'system of Paris. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 39.

Froehlich, J. E., Neumann, J., & Oliver, N. (2009, June). Sensing and predicting the pulse of the city through shared bicycling. In Twenty-First International Joint Conference on Artificial Intelligence.

Faghih-Imani, A., & Eluru, N. (2016). Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. Journal of Transport Geography, 54, 218-227.

Zhou, X. (2015). Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago. PloS one, 10(10), e0137922.

Vogel, P., & Mattfeld, D. C. (2011, September). Strategic and operational planning of bike-share systems by data mining–a case study. In International Conference on Computational Logistics (pp. 127-141). Springer, Berlin, Heidelberg.

Appendix

Table 1: Regression Result

	Trip Count		
	(1)	(2)	(3)
start_station	-0.003***	-0.001***	-0.001***
	(0.00003)	(0.00002)	(0.00002)
hour(interval60)	0.019***	0.002***	0.002***
	(0.0002)	(0.0002)	(0.0002)
dotw.L	0.018***	-0.023***	-0.024***
	(0.004)	(0.003)	(0.004)
dotw.Q	-0.197***	-0.099***	-0.099***
	(0.004)	(0.004)	(0.004)
dotw.C	0.059***	0.047***	0.047***
	(0.004)	(0.004)	(0.004)
dotw4	-0.063***	-0.061***	-0.061***
	(0.004)	(0.004)	(0.004)
dotw5	0.019***	0.034***	0.033***
	(0.004)	(0.004)	(0.004)
dotw6	0.017***	0.008**	0.009**
	(0.004)	(0.004)	(0.004)
Temperature	0.011***	0.003***	0.003***
	(0.0001)	(0.0001)	(0.0001)
Percipitation	-0.110***	-0.049***	-0.049***
	(0.005)	(0.005)	(0.005)
lagHour		0.280***	0.280***
		(0.001)	(0.001)
lag2Hours		0.090***	0.090***
		(0.001)	(0.001)
lag3Hours		0.280***	0.280***
		(0.001)	(0.001)
lag12Hours		-0.001	-0.001

		(0.001)	(0.001)
lag1day		0.273***	0.273***
		(0.001)	(0.001)
holidayLagMinusOneDay			-0.220***
			(0.069)
holidayLagMinusThreeDays			-0.169**
			(0.085)
holidayLagMinusTwoDays			-0.120
			(0.085)
holidayLagPlusOneDay			-0.062
			(0.069)
holidayLagPlustThreeDays			-0.269***
			(0.085)
holidayLagPlustTwoDays			-0.295***
			(0.085)
holiday			0.073
			(0.070)
Constant	9.928***	3.352***	3.353***
	(0.086)	(0.077)	(0.077)
Observations	544,502	544,478	544,478
R^2	0.072	0.313	0.313
Adjusted R^2	0.072	0.313	0.313
Residual Std. Error	1.133 (df = 544491)	0.975 (df = 544462)	0.975 (df = 544455)
F Statistic	4,242.643*** (df = 10; 544491)	16,543.380*** (df = 15; 544462)	11,282.000*** (df = 22; 544455)

Note: *p<0.1; **p<0.05; ***p<0.01