Master's Thesis

Integration of Bicycle Commuting to Public Transit in New York City

Graduate School of Architecture, Planning and Preservation

Columbia University

Zhiyuan Han

May 6th, 2016

ABSTRACT

In this thesis, the author is dedicated to exploring the bicycle commuting trend in New York City and discussing the integration of bike-share system to public transit modes. Rather than looking into the whole population of cyclists, the analysis focuses on the group using Citi Bike, the bike-share system in NYC, as a commuting tool. Determinants of Citi Bike usage is examined through bivariate and multivariate correlation analysis. Specifically, the thesis consists of 6 parts. Chapter 1 goes through an overview on the basic development of trend of bicycle commuting. Chapter 2 looked into a bunch of early studies researching on the determinants of cycling level and statistical analysis methods. Extra attention is paid to the discussion about what has been influencing the usage of bike-share system. Chapter 3 overall introduces the data sources and ideas about data preprocessing. Research question is raised and the basic hypothesis described. More importantly, the principle and techniques of the two major analysis in this study is explained in detail. Implementation process and findings of the major analysis, the temporal and spatial analysis as well as the correlation analysis, are discussed separately in Chapter 4 and Chapter 5. Eventually, Chapter 6 concludes on the findings and arguments the author has proposed through the whole study and raises some of the ideas for further studies.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

Bicycle commuting has drawn much attention in recent years, especially in large cities in developed countries. In the United States, transportation departments have been producing policies that benefit bicyclists and implementing bicycle and active commuting programs all over the states in order to encourage bicycling for both commuting and recreation. It is explicitly listed in a federal document, Executive Order (EO) 13514, "Federal Leadership in Environmental, Energy, and Economic Performance," signed by President Obama on 5 October 2009, to "implement strategies and accommodations for transit, travel, training, and conferencing that actively support lower-carbon commuting and travel by agency staff." Bicycling is definitely an important component of low-carbon commuting transportation, given that if one of 10 commuters switched to a bike, CO_2 emissions would be reduced by 25.4 million tons per year.[1]

Bicycling to work is recognized as a great opportunity to increase residents' physical activity. Studies show that thirty minutes of daily cycling saves $544 per person in annual medical costs.[2] Unlike other sports, bicycle commuters do not have to spend extra time on their busy workdays. As a result, they obtain health benefits on the way to work and suffer much lower risk of obesity and heart diseases. Some companies even offer additional incentive to encourage employees riding their bikes to work. Microsoft and Fairchild Semiconductor, for instance, offer to pay for employee gym memberships if their employees would integrate cycling into their commuting.

---

[1] Implementing a Successful Bicycle and Active Commuting Program in the Washington, DC Metropolitan Area, May 21, 2010.
[2] bikesmakelifebetter.com

Bicycle commuting also contributes to the construction of sustainable transportation systems since it would reduce air pollution and energy consumption. Transportation-related emissions account for approximately 33 percent of U.S. CO_2 emissions.[1] Meanwhile, as automobiles has been the dominant mode of commuting in most developed regions, a large amount of the private vehicles are occupied by drivers commuting alone. In New York City, 21.4 percent of the employees drove solo to work by 2009, (Miller, 2014) resulting in a huge waste of transport space and energy although the percentage has been decreasing in recent years. The low occupancy rate of vehicles during travelling leads to higher vehicle ownership, and thus contribute to more pollutant emission. Bicycle commuting provides a more efficient and healthy transport mode by saving resources and eliminating negative impact on the environment. Promoting cycling would switch some motorized trips to flexible and space-saving bikes, thus reducing traffic congestions and parking problems.

Bicycling in New York City, however, still has much room to grow. Compared to the 5.9 percent of residents commuting by bike in Portland, 4.5 percent in Washington, DC, 3.8 percent in San Francisco, 2.3 percent in Philadelphia, and 1.9 percent in Boston, the percentage of New York City reached only 1.2 in 2013 after years' growth. (Pucher and Buehler, 2011) In terms of the integration of bicycling with public transport, New York City even takes the last place among equivalent large cities in North America in the supply of bike parking at transit stations and bike accessibility to buses and subways. In cities like San Francisco and Washington, DC, thousands of bike parking facilities are provided at transit stations and almost 100 percent of the buses are equipped with bike racks, while in New York City, there are no such facilities at all by 2008. (Miller 2014) So it is of great necessity to talk about integrating bicycling with public transport in New York City.

2 LITERATURE REVIEW

A quantity of literature has examined the determinants of bicycle commuting. Since commuting manner is quite a subjective choice, which to a large extent depends on personal situation, the demand of bicycle commuting varies a lot among the areas with different demographic features, including age, gender, income, car ownership, education attainment, etc. The results show conflicts within different study levels. Besides inherent individual factors, it turns out that the supply of infrastructure also has a large impact on citizens' choice of bicycle as a transit tool. Bike paths and lanes, as well as trip-end facilities such as showers and lockers, usually contribute to more bicycling in the city, given that cyclists care a lot about safety and cleaning issues. Given the same reason, the potential accident risks and more physical effort than required by driving would definitely deter some citizens from the choice of bicycle commuting.

Bivariate correlation tests and multiple regression are conducted to verify the influences that possible factors have on bicycle usage. Chi-square is usually used for testing nominal variables and t-test for interval variables. Some literature converted interval variables to nominal ones by using quartiles so that it would be easier to figure out and elaborate the relationships. Multiple regression and corresponding diagnoses were conducted through different models, integrating Ordinary Least Square(OLS), logarithm and odds ratio, in order to figure out the best estimate of cycling level.

As part of the public transport itself, bike share systems provide a huge convenience to cyclists. Commuting was found to be the most common purpose for annual users. (Fishman, 2014) Bike share is also used as connection between workplace and other public transport, such as train stations and bus stops. As a result, the bike station's proximity to work or home is identified as one of the strongest motivations for citizens' cycling. The bike usage is measured either by trips per day per bike or single trip duration in minutes. It could be influenced by both motivating factors and demographic features.

2.1 DATA AND FINDINGS

A vast number of studies researched on the determinants of bicycle commuting. Most of the determinants were found having significant impact on bicycle usage as a commute tool by using empirical analysis, and fall into three categories: socioeconomic characteristics, infrastructure, and potential costs of bike commuting.

2.1.1 SOCIOECONOMIC CHARACTERISTICS

In terms of the socioeconomic characteristics, most studies got the demographic information from the US census data. Most researches found that both gender and age lay an impact on bicycle usage. There is a significant difference between male and female share of bike trips on the nation level. The percentage of male cyclists has seen an upward trend from 2001 to 2009, while that of female cyclists remains on the same level, resulting in an even lower percentage of bike trips by females. However, compared to the falling female cycling share in the US, there has been an increase in Canada given the safer cycling environment. So what may explain the inverse trend in these two countries is that women have more concern on cycling safety. (Pucher and Buehler, 2011) Variation in share of bike trips is also found among different age groups. According to the American Community Survey report, the percentage of bike trip share generally decreases as the age increases, showing the most share of bike trips falls in the group of 16-24 years old. (Mckenzie, 2014) Another study shows a conflict finding that the group of 16-24 years old comes with the lowest share of bike trips except for the group of 65 and older, taking into consideration a new group of 5-15 years old which reaches the highest share percent. (Pucher and Buehler, 2011)

Income is another important determinant of bicycle usage, which is highly correlated to the affordability of different transportation modes and car ownership. It is reported in the American Community Survey that the percent share of bike trips in all trips generally declines as household income level moving up through the ten income ranges. Several studies categorize the different income levels into four quartiles. Household income level increases from the first to the fourth quartile. In one national study, however, almost no difference is found in bike preference among the four quartiles. In another research conducted in the Washington, DC metropolitan area, the top 25% income is found to have statistically significant difference from the other income levels, sharing 35% of the bike trips. In contrary, the 2009 National Household Travel Survey indicates a higher percentage of bike mode share in the lowest income quartiles. Hence, it is assumed that lower-income people mainly cycle for commuting or other utilitarian purposes while the higher-income group cycles more for entertainment and physical exercise. (Jennifer, 2009)

The impact of automobile ownership and educational attainment on bike usage is also mentioned a lot in the socioeconomic characteristics discussion. According to the National Household Travel Survey, the bike mode share of all trips for households with zero vehicles has the highest percent of 2.4, while those for households with one, two, three or more vehicles reach a much lower level of 0.7, 0.9, 0.8. (NHTS 2001) We can assume that the group of households with no vehicles have more rigid demand of bike either for commuting or for other utilitarian purposes.

In terms of educational attainment, it is found that workers with the highest education level, who obtained a graduate or professional degree, have the highest rate of bicycle commuting (0.9), followed by those with the lowest educational level, who drop out before high school graduation, reaching the rate of 0.7. (Mckenzie, 2014) This interesting result indicates that the least educated workers cycle to commuting mostly for the economic saving concern, while the best educated workers for physical exercise.

However, the automobile ownership and educational attainment factors are generally significantly correlated to income level. As a result, extra attention should be paid when integrate these two determinants as independent variables to a multiple regression model together with income.

2.1.2 INFRASTRUCTURE

In terms of the impact of infrastructure on cycling level, what has been examined most are bike paths and bike lanes. Bike lane is defined as "a portion of a roadway which has been designated by striping, signing and pavement markings for the preferential or exclusive use of bicyclists." Bike path, also called shared use path, is generally referred to as "a bikeway physically separated from motorized vehicular traffic by an open space or barrier and either within the highway right-of-way or within an independent right-of-way." (AASHTO Executive Committee, 1999) Shared use paths are also accessible for walking, jogging, skating, and other non-motorized activities.

Pucher, John R., and Ralph Buehler measured bike lanes and paths by miles per square mile of city area, finding that New York City has the most supply of bike lanes and paths among 100 largest U.S. cities. (2001) However, New York City doesn't even get into the top ten list of bike facility supply per 100,000 population, where Scottsdale ranking the first, indicating that although New York City has invested a lot on bicycle infrastructure, the resources shared by each resident are still insufficient. It would be relatively crowded to cycle in New York and thus discourage workers commuting by bike.

It is also found that cities with well-supplied bike lanes and paths usually have a higher cycling level. The result of Bivariate Pearson's correlations showed that the cycling level in cities with the most supply of bike lanes is more than three times of that in cities with the least supply, while the multiple relationship is twice in terms of the supply of bike paths. The multiple regression models examined the relationship between cycling levels and each of the independent variables all together, while controlling for some other variables at the same time. It turned out that bike lanes and paths explains a large percentage of variation in cycling level. Meanwhile, cycling safety turned out to have a statistically significant impact, while other control variables, such as climate, population density, vehicles ownership, did not. (Pucher, John R., and Ralph Buehler, 2001)

Studies also showed significant impact of trip-end facilities on bike commuting, such as bike parking facilities, showers and clothes lockers at workplace. In a study conducted in Washington DC, 11.1% of the commuters reported to have comprehensive trip-end facilities at workplace encouraging bike commuting while 38.2% only have bike parking. Adjusted odds ratio(AOR) was examined between bicycle commuting decision and trip-end facilities, together with transit commuter subsidies, race, gender, age, household income, car and bike ownership, trip distance, miles of bike lanes and paths, and climate as control variables. The results indicated that commuters whose workplaces provide bike parking but not showers or lockers are 78% more likely to commute by bicycle, while those whose workplaces provide bike parking, showers, and lockers are 386% more likely to commute by bike than individuals without any trip-end bicycle facilities at work. In addition, it is noteworthy that the AOR of miles of bike lanes and paths per 1000 population is only 1.11, indicating that for each additional mile of bike lanes and paths, workers would be 11% more likely to commute by bike. However, this result is statistically significant at 99% significance level, indicating the supply of bike lanes and paths has strong positive impact on workers' decision of bicycle commuting. (Buehler, 2012)

2.1.3 POTENTIAL COSTS

Except for the positive determinants, there are definitely quite a few factors that deter workers' choice of bicycle commuting. Physical effort, such as slopes and frequent stops, is considered to have the most effect on discouraging workers from commuting by bike. Even if one cycles on even paths, the energy consumption would be 14.7 kJ per minute, generally more than three times that of driving on roads which is only 4.2 kJ per minute. (Kölbl and Helbing, 2003) Frequent stops would consume even more energy from cyclists, since changing the state of an object needs more effort than keeping it.

Long travel time would be another major hindrance. The average travel time for bicycling is 20 minutes. Workers living and working within the city have a notably higher rate of bicycling commuting than those working outside the city of their residence. The former reached a rate of 1.2 percent, about 4 times higher than the latter. (McKenzie, 2014) Sometimes the distance from transit stations to home, or to work, is also considerable, since many workers cycle as their secondary transport mode.

The potential accident risks are also considered to play a large role in non-motorized commuting mode since cyclists and walkers are relatively more vulnerable with a certain travel distance than car drivers. (Rietveld and Daniel, 2004) Another study provided evidence of higher risks that compared with car drivers, bike-share system users experienced an increase of 0.03 in death from traffic accidents and 0.13 from air pollution. (Rojas-Rueda, Nazelle and Tainio, 2011)

2.2 RESEARCH APPROACH

2.2.1 VARIABLES

Many studies examined the determinants of cycling level with all purposes, including commuting and other utilitarian activities, as well as recreation or exercise. Cycling levels were generally measured by either the absolute number of annual bike trips or the share of bike mode of all trips. Pucher and Buehler even looked at the share of all bike trips by more detailed purposes including "commute or work related", "shopping", "personal business", "school, church, doctor", "visit friends", "recreational or vacation", "transit access or egress". (2011) Bicycle share of commute trips is also examined and turned out to be higher than the share for all trip purposes. (Buehler, 2012)

Most of the studies found that infrastructure has significant impact on residents' decision to travel by bike. Except that bike lanes and paths are measured by length per area unit of the city, bike parking facilities as well as lockers and showers at workplace are measured as dummy variables, or nominal variables, to examine whether the supply of such facilities would make significant difference to cycling level.

In estimating the separate impact of infrastructures on cycling levels, quite a few variables are commonly controlled in the literature for the accuracy and effectiveness of the multiple regression model. Pucher and Buehler included population density to estimate the effect of land use. In terms of socioeconomic factors, they included share of students in the population assuming certain group would contribute more to the cycling level, and percent of households without automobile as the measurement of car ownership. They carefully excluded income assuming high correlation between income and car ownership. Safety is examined by looking at cyclist fatalities per 10,000 bicyclists, the data of which is only available on the state level. Another important control variable is climate. Assuming extremely cold and hot temperatures would discourage residents' travelling by bike, they examined the average annual number of days that reach the temperatures of over 90 °F. The annual precipitation level and extremely cold temperatures were excluded since the impact is small and statistically insignificant. (2011)

2.2.2 BIVARIATE CORRELATION

Bivariate correlation tests are conducted to examine the separate impact of all the independent variables on cycling level. Chi-square test is generally used to examine whether the nominal variables have significant impact, and for instance, whether the supply of bike parking facilities would significantly increase cycling level. Buehler even converted some of the interval and ratio variables into nominal variables to which Chi-square tests were also applied, assuming that when such variables reach a certain level, more significant impact would be laid on cycling level. (2012) T-test was mostly applied to ratio variables such as the supply of bike lanes and paths, population density, car and bicycle ownership. All the variables examined turned out to be statistically significant at 95% significance level, which is a good preparation for the further multiple regression analysis.

Pucher and Buehler grouped the 100 cities into quartiles of bike lanes and paths supply, with the fourth quartile containing the most supply and the first quartile containing the least. This method corroborated the bivariate correlation by providing more intuitive results telling how different the fourth quartile is from the first. The comparison between different measurement of cycling levels, share of bike commuters and bike commuters per 10,000 population, also cross validated the results, indicating those with consistent significance levels are more credible.

2.2.3 MULTIPLE REGRESSION

Pucher and Buehler set up two sets of multiple regression models to find out the best estimate of cycling level. The first model was an Ordinary Least Square (OLS) regression with the natural logarithm of bike commuters per 10,000 population as dependent variable. The natural logarithm was applied to both dependent and independent variables to make sure the error terms are normally distributed, so that the assumption of OLS would be met. Three modifications were made to the first model, thus the first model was run four times, to overcome the potential multicollinearity among independent and control variables. To improve the model, a composite variable was included to examine the interaction among possibly multicollinear variables, and a variable was excluded as imperfect proxy in the following tests. Adjusted R^2 and F statistics were compared among four tests with the first model.

A second set of models was estimated using a Binary Logit Proportions Model with the share of bike commuters as dependent variable. The purpose was to verify the robustness of the results returned by the first regression model. Instead of directly using the number of measurement for each variable, the natural logarithm of the odds ratio of each variable was applied. The model became more complicated but much better met the assumption of multiple regression with the linear relationships between x and y from the regression equation and homoscedastic error terms.

Rietveld and Daniel estimated a semi-log linear regression model with the natural logarithm of the share of bicycle use as dependent variable. Quite a few similar determinants, as mentioned above, were used as independent and control variables. The fitness of the model is good since all the parameters are statistically significant at 0.10 significance level and the multicollinearity examination proved the robustness of the results. The adjusted R2 reached 0.726 and the F statistics is 20.30, indicating the model efficiently explains a large proportion of the variation in the share of bicycle use.

2.3 BIKE-SHARE SYSTEM

Bike share usage is commonly measured by the metric trips per day per bike, in order to control for variation in the number of bikes in a system. Studies showed bike share usage patterns by looking at the usage over time during the day and over months during the year. Usage of bike share system is also compared among different cities.

Another thing discussed a lot is motivating factors. In most studies, convenience is found to be the main motivating factor, with docking station proximity to work and home identified as the second strongest motivator. Finally, saving money on transport is found to be another important motivator.

Demographics of bike share users, as a common focus of attention, are also examined. It turned out that bike share users are more likely to be female, younger, and own fewer cars, in comparison to regular bicycle riders.

While the usage and demand of bike-share systems have been discussed a lot in early studies, the analysis in these study on bike-share usage was mostly conducted based on the information of single users or the duration of a single trip, which provide more significance in terms of bike-share customer market expansion or facility adjustment at individual stations. A majority of these studies adopted the approach from the perspective of geography or bike-share system design. Bike-share system's contribution as a green efficient tool to the public transit is seldom studied in depth from the perspective of urban planning. Therefore, based on the intuition on the determinant and major purpose of bike-share trips taken from existing studies on bike-share systems, this thesis will integrate the research approach and influential factors, as well as spatial and statistical methods to the analysis of bike-share system in New York City, focusing on integration of bicycle commuting to public transit.

3 METHODOLOGY

The methodology applied in this thesis includes two important techniques: Geographic Information System (GIS) based spatial analysis, and statistical correlation analysis. GIS is not only powerful in figuring out the spatial patterns of certain factors, but also very efficient in combining information stored in tables with spatial factors based on some common attribute or spatial relationships. Only in this way can the statistical analysis get the appropriate input and present persuasive results.

3.1 DATA

This study would mainly use open-source data and focus on the census tract level. The demographics, including population density, age, gender, household income, and education attainment, would be extracted from the American Community Survey (ACS). The ACS 5-year estimates would also provide access to the commuting characteristics by sex which covers the percent of cycling in all commuting modes in each census tract. NYC Department of Transportation provides open data for bike share system, bike routes, bike parking, ridership and safety statistics. The data of subway stations is available from the Metropolitan Transportation Authority (MTA).

For bike share specifically, the system data of Citi Bike provides a quantity of trip records, including trip duration in seconds, the name and location of start and end stations, the time and date when a trips starts and ends, user type, gender and age. These are very useful information for the analysis of bike share system in New York City.

3.2 RESEARCH QUESTIONS

A majority of studies looking into cycling level and bicycle commuting were conducted on the scope of metropolitan or inter-city comparison. The macro scope analysis has shown valuable findings regarding national cycling trend, the corresponding causation of the variation of cycling level over time and space, as well as the determinants of workers' bicycle commuting preference. This research, however, would look specifically into New York City on a much smaller scale, the census tracts, to conduct analysis on inner-city bicycle commuting issues. Rather than taking all the cyclists into consideration, the author will focus on the usage of Citi Bike, the bike-share system in New York City, which has been recording over 100,000 bike-share trips every month since the system was put in use in 2013.

The author will first try to figure out how Citi Bike has been influencing citizen's commuting manner in New York City by testing whether the bike-share service made a significant difference on bicycle commuting population over time and over space. Then more detailed correlation analysis would be conducted to explore what the determinants are that influence the bike-share usage for commuting purpose.

3.3 DATASET PREPROCESSING

As the major dataset used for core analysis in this study, Citi Bike system data is preprocessed based on the specific focus of bicycle commuting. The system data covering the whole year of 2015 is downloaded for preprocessing. A small preprocessing program has been written in Python making it much easier to read and write a bunch of csv files, as well as to make a series of smart calculations in a short time. Since the whole dataset is too large, containing millions of trip records, for the convenience of further analysis, a random sample with population of 10,000 for each month is selected based on the following three criterions. First, one attribute in the dataset called "start time" contains the date and time of the day when a certain trip starts. The starting time of bike trips should fall into two periods in a day, 7-9am and 4-6pm, assuming that people use Citi Bike in these two time periods to commute on weekdays, which are the peak hours of bike share usage for commuting. (Pfrommer, Warrington, Schildbach, and Morari, 2014) The second selecting criterion would therefore be that the sample only includes trips performed on weekdays. The "datetime" library in Python could automatically calculate day of the week based on the date in the "start time" attribute of each trip. Third, the "user type" attribute in the dataset specifies whether the user is a long-term subscriber or a temporary customer. In order to exclude as many trips of purposes other than commuting as possible, only trips performed by subscribers are selected, provided that a number of customers also tend to be active, with exercise and recreational purposes for instance, in the selected time periods, especially from 4 to 6 in the afternoon.

The preprocessing program goes through every one of the dataset files covering all 12 months in 2015, selects 10,000 random trips fulfilling the screening terms from each month, and finally outputs one compiled file with 120,000 random Citi Bike trips, the riders of which are supposed to travel for the purpose of commuting. While carefully inspecting the compiled dataset, which often turns out to be necessary to make sure the accuracy of data, it is worth noting that there is a little flaw about Citi Bike system data that the order of attributes "gender" and "birth year" might vary among different monthly datasets. As a result, the automatically concatenated trip records end up containing two columns of gender and birth year mixed with each other. In order to fix this unexpected mismatch, a new field is created getting the minimum value between the two mismatched columns as gender, since gender is indexed as “1 for male, 2 for female, and 0 for unknown”, which are definitely smaller than the value of birth year. Since no analysis would be conducted on bike users’ ages, at this moment, the dataset of bike trips is ready to use. This preprocessed dataset would be displayed as points in ArcGIS based on the coordinates information for further analysis.

3.4 TEMPORAL AND SPATIAL ANALYSIS

Commuting characteristics by sex were collected by census tract units for the year 2012 and 2014. A temporal analysis would be conducted to test whether there is significant difference on bicycle commuting preference before and after 2013 when Citi Bike started service. Then a spatial analysis would be used to test whether there are more residents choosing to commute by bike in census tracts with Citi Bike stations than those without bike stations. In order to eliminate the impact of population growth and the unequal spatial distribution of population density, bicycle commuting preference will be measured in the percent of cycling population in all population of each census tract.

In order to get data ready for the test, the Spatial Join tool in ArcGIS was used to join the points feature of Citi Bike stations to census tracts, identifying the census tracts with or without Citi Bike stations and attaching this attribute to the bicycle commuting table. Census tracts with bike station located inside are marked as “1”, the other tracts marked as “0”.

A paired t-test would be conducted on the census tracts with Citi Bike stations located inside. The bicycle commuting percent would be tested for the year 2012 and 2014 as the temporal analysis. The hypothesis tested here is that the bike-share system has contributed to a significant rise in workers’ bicycle commuting percent in total population of each census tract. An unequal-variance unpaired t-test would be conducted between the group of census tracts with bike stations and the group without to explore the spatial variation of bike commuting resulted from the spatially limited Citi Bike service, assuming that the bicycle commuting percent in census tracts with Citi Bike stations would be much higher than that in census tracts without Citi Bike stations.

In addition, spatial statistical analysis would be conducted in ArcGIS to further explore the spatial pattern of Citi Bike usage. Concretely, two spatial statistics tools, Cluster and Outlier Analysis (Anselin Local Moran's I) and Hot Spot Analysis (Getis-Ord Gi^*), would be helpful with the analysis on bike-station-based point level and census tract level. As preparation for the Cluster and Outlier Analysis, given that many selected trips occurred at the same station and all of them have been displayed as points, the "Collect Event" tool in ArcGIS will be used to eliminate duplicated points at the same location and get count of the number of trips for each station, which is used as the input field of the analysis. As a result, the output will include clusters, outliers and insignificant points. In terms of the census tract level analysis, the "Spatial Join" tool will be used to sum up the count of bike trips in each census tract, as the input of Hot Spot Analysis. Similarly, the result will present the spatial distribution of hot spots, cold spots, and insignificant spots. Actually, both of these spatial statistical tools are dedicated to figure out where high and low values are clustered. High-value clusters are called hot spots, and low-value clusters are cold spots.

3.5 DETERMINANT ANALYSIS

3.5.1 VARIABLES

The author will be trying to explore the determinants of Citi Bike usage based on the unit of census tracts. Thus, the dependent variable will be the number of trips occurring in each census tract. Two circumstances need to be analyzed separately. Citi Bike trips that start from a certain census tract and that end in the same tract might probably be determined by difference factors. As mentioned in the data preprocessing section, only trips occurring in certain time periods, 7-9am and 4-6pm on weekdays, and performed by subscribing users are selected because the focus of this thesis is the Citi Bike trips for commuting purpose.

In addition, there are actually several other ways to measure bike usage, such as trip duration measured in seconds, or number of trips standardized with land area or population excluding the impact of the size of census tracts. The reason why trip duration is not adopted is that this determinant analysis is focused on exploring what has been determining workers' choice of Citi Bike as a commuting mode or part of it, and how to encourage more citizens to integrate cycling to their commuting mode. Trip duration is more relevant to the concept of cycling level, usually measured by distance or time duration, which has been talked about in the overview of early studies. In other words, rather than trying to encourage people to cycle more, this thesis is aiming to call for higher proportion of cycling in all commuting modes.

In term of the standardization of the dependent variable, it is true that there is no way to ignore the impact of census tract size and population. Census tracts with higher level of bike usage will probably turn out to have more bike stations. However, in order to present better intuition of bike usage at census tract level, the number of bike trips will be used directly as the dependent variable, and the influential factors will also be included as determinants of bike usage level in the analysis. Thus we are able to expect achieving equivalent effect of standardizing the number of trips.

The determinant analysis will mainly focus on the effect of infrastructures that are considered to play an essential role in encouraging workers to commute by bike, while controlling for the demographics which are also important factors that influencing bike usage. The independent variables will include the length of bike lanes, the number of bike racks, proximity to subway stations and the density of bike stations. Meanwhile, population density, gender, age, income, education attainment will be included as control variables.

One assumption is that people are more willing to travel along roads with bike lanes since it would be much safer and less congested. Although the whole study is conducted on census tract level, it is unreasonable to measure the length of bike lanes completely within the boundary of a certain census tract. A trip might occur at any relative position of a census tract, either near the edge or close to the center, and obviously, the range of this certain trip could not be limited within the scope of a single tract, meaning that the supply of bike lanes in the neighborhood also matters a lot. So a half-mile buffer is created around the centroid of each census tract and total length of bike lanes within this buffer will be used as the measurement representing the influence of bike lanes on bike usage.

The proximity to subway stations is initially measured by the average distance from each bike station within the census tract to the nearest subway station. However, the regression results turn out to be insignificance. Adjustment is made by changing the nearest distance to number of subway stations within a quarter-mile buffer of each bike station. Then the author takes the average of this count of all bike stations in each census tract as the second important independent variable. It is assumed that many subway customers use Citi Bike as the tool to achieve the last mile transit. The hypothesis is that subway stations within the walking distance of each bike station will have significant impact on the bike usage at these stations. So the buffer radius is set to a quarter mile since it is the critical value of comfortable walking distance.

The number of bike racks per square mile in each census tract is used as the third independent variable assuming bike racks would provide more convenience to bike users. The density of bike stations in each census tract, which is the result of number of bike stations divided by the area of each census tract, is another variable as mentioned above to control the impact that more stations result in more trips. The reason the author is using density rather than simply number of stations is that integration of bike station density not only shows concern on the impact of station number, but makes more sense reflecting the mutual influence among bike stations that are close to each other. The concern about density and number will be further explained in Section 4.2.

Several factors are controlled in order to make better prediction through the regression model. Population density, as mentioned in explaining the measurement of the dependent variable, is included to control both the population and the size of census tracts. The percentage of male is used to control for the impact of gender composition. The percentage of students, population aged between 5 and 24, is also included, assuming this age group contributes most to the bike usage. Also, median household income and the percent of education attainment of bachelor or more are integrated as control variables.

3.5.2 CORRELATION ANALYSIS

The bivariate correlation will be examined between each determinant and bike-share usage. All the interval variables are divided into quartiles. T-test will be conducted to examine the difference in bike usage level between the first and the fourth quartile, two groups with the lowest and the highest values, of each determinant. The results are expected to show clear intuition of the impact of each determinant on Citi Bike usage. Then more comprehensive correlation analysis, multiple regression, will be conducted to explore the relationship between Citi Bike usage and all the independent and control variables mentioned in last section above at the same time. The multiple regression model could present the mutual influence among the variable more accurately, and better predict the variation trend of the dependent variable. Multicollinearity and influential outliers will be carefully checked to guarantee the accuracy of the regression model. The model could be continuously improved by dropping multicollinear variables and influential outliers.

4 INFLUENCE OF BIKE-SHARE SYSTEM

Following the boom of bike-share programs all over the world, New York City carried out its own bike-share system, Citi Bike, in July, 2013. During Citi Bike's first month in service, it saw over 800,000 bike trip records reaching a total trip duration of more than 233,837 hours. However, the bike stations were only accessible in lower Manhattan and downtown Brooklyn, which means there could be much more demand for public shared bikes in other areas of New York City. In order to explore the influence of bike-share system on the city's overall cycling level and residents' preference of commuting manner, the author would conduct both temporal and spatial analysis based on census tracts.

4.1 BEFORE AND AFTER CITI BIKE STARTING SERVICE

In this temporal analysis, the author will compare the percent of workers commuting by bike in the total population of each census tract a year before and after 2013 when Citi Bike started service. Census tracts with Citi Bike stations located inside are selected and compared correspondingly regarding the bicycle commuting percent in 2012 and 2014. So a paired t-test is necessary for this kind of examination of two trials from a single group.

Before accurate statistical analysis, Figure 1 shows the general trend in terms of how bicycle commuter percent has been changing over the past ten years in New York City. An overall upward trend is found in all categories - whole population, male and female. With extra attention drawn to the change between 2012 and 2013, a slightly faster increase could be found in the percent of bicycle commuters in whole population, while male bicycle commuters present a much more obvious growing trend. However, female bicycle commuter percent shows a relatively flatter trend, whose increase even slows down from 2013 to 2014. This intuitive trend indicates that Citi Bike has probably accelerated the increase of bicycle commuter percent overall in New York City. The change of male bicyclists is especially subjected to this new commuting option, which is, however, much less influential to female commuters. This assumption could be verified through t-tests, specifically between areas with and without Citi Bike service, as well as between male and female commuters.

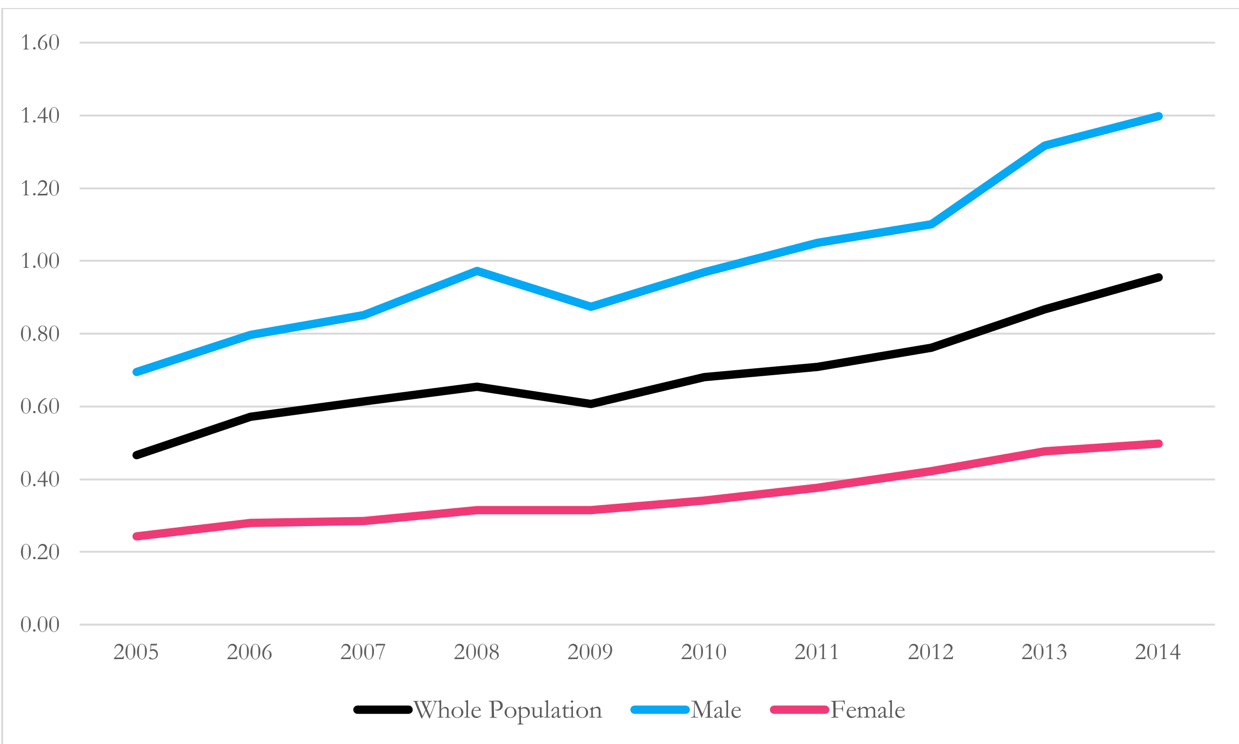

Figure 1 Changing Trend of Bicycle Commuter Percent in NYC

(Source: American Community Survey 2005-2014, Commuting Characteristics by Sex)

The result of the paired t-test between the bicycle commuting percent in 2012 and that in 2014 (See Table 1) reveals that under 97.5% confidence level, there is a significant increase in the percentage of workers who commute by bike. In other words, after Citi Bike started service, bike commuting became more popular among workers in the census tracts that are within the Citi Bike service area. The increase of bike commuting percent, however, could also be a general trend all over the city which means that it has nothing to do with the bike-share program. In order to verify this assumption, a similar paired t-test was conducted between the bicycle commuting percent in 2012 and 2014 in the census tracts without Citi Bike stations located inside. As expected, the result also shows significant increase of bicycle commuting percent from 2012 to 2014. Nevertheless, a slight difference is found in the variation of mean between these two t-tests. For census tracts with Citi Bike stations, the average of bike commuting percent increased in a large range, from 1.73 to 2.36, while in census tracts without bike stations, the average percent increased only 0.12, from 0.56 to 0.68. It is indicated that Citi Bike has probably promoted the increase of bicycle commuting percent.

Regarding the increase range of bicycle commuting percent, another t-test was conducted on the percent change of workers commuting by bike in total population of each census tract. One group is census tracts with Citi Bike stations and the other group is those without Citi Bike stations. The result (Table 2) shows that the average bicycle commuting percent in census tract with bike stations is significantly higher than that in census tracts without bike stations. So it could be concluded that the bike-share system in New York City is accelerating the increase of bicycle commuting percent in total population of each census tract rather than determining the increase.

Table 1 T-test Results on Bicycle Commuting Percent Over Time

	Year	*Obs*	*Mean*	*SD*	*t*	*Pr(T>t)*	*Alpha*
CTB	2014	235	2.36	2.60	5.62	0.00	0.025
	2012	235	1.73	2.05			
CTNB	2014	1883	0.68	1.36	5.47	0.00	0.025
	2012	1883	0.56	1.27			

(CTB = Census tracts with Citi Bike stations; CTNB = Census tracts without Citi Bike stations)

Table 2 T-test Result on Bicycle Commuting Percent Change

	Obs	*Mean*	*SD*	*t*	*Pr(T>t)*	*Alpha*
CTB Change	235	0.63	1.73	4.45	0.00	0.025
CTNB Change	1883	0.12	0.97			

In addition, when examining the bicycle commuting percent change by sex, it is surprising that there is no significant difference in the percent of female bicycle commuters in census tracts with Citi Bike stations from 2012 to 2014, while the group of male bicycle commuters shows significant increase as the whole population does. Comparatively, both male and female in census tracts without Citi Bike stations show significant difference in the year 2012 and 2014. (Table 3) However, there is no reason to interpret that bike-share system hinders more females from choosing bike commuting. Further tests indicate that the possible explanation might be the different sensitivity of males and females towards the bike-share system as a new commuting manner.

Further t-tests were conducted regarding the percent change of male bicycle commuters and female bicycle commuters in census tracts with and without Citi Bike stations. The results indicate that there is significant difference between the percent change of male and female bicycle commuters in census tracts with Citi Bike stations, while in contrast, no statistically significant difference is found from the two groups in census tracts without Citi Bike stations. (Table 4) In the census tracts with

Citi Bike stations, given higher average percent change, the percent of male bicycle commuters is increasing much faster than that of female, while there is no such sign in census tracts without Citi Bike stations. So it is reasonable to interpret that male bicycle commuters are more sensitive to the alternative bike commuting tool - Citi Bike, which lays little impact on females' commuting preference. In other words, due to the operation of bike-share system, more males are motivated to commute by bike than females.

Kaufman has pointed out in her study on Citi Bike and gender that the main reason for the sexual difference in Citi Bike usage is that the safety among car traffic lays more impact on women's preference towards transit tools. (2014) Given that a majority of Citi Bike stations are distributed across the most congested parts of Manhattan and Brooklyn, it makes sense that female cyclists share a much lower percent of Citi Bike trips.

Table 3 T-test Results on Bicycle Commuting Percent Over Time by Sex

	Gender	*Year*	*Obs*	*Mean*	*SD*	*t*	*Pr(|T| > |t|)*	*Alpha*
CTB	Male	2014	235	3.38	3.96	5.57	0.0000	0.05
		2012	235	2.30	2.78			
	Female	2014	235	1.20	1.68	1.11	0.2689	0.05
		2012	235	1.09	1.98			
CTNB	Male					3.74	0.0002	0.05
	Female					4.60	0.0000	0.05

Table 4 T-test Results on Bicycle Commuting Percent Change by Sex

	Gender	*Obs*	*Mean*	*SD*	*t*	*Pr(|T| > |t|)*	*Alpha*
CTB	Male Change	235	1.08	2.97	4.42	0.00	0.05
	Female Change	235	0.11	1.56			
CTNB	Male Change	1880	0.14	1.65	1.16	0.25	0.05
	Female Change	1880	0.09	0.87			

4.2 WITH AND WITHOUT CITI BIKE SERVICE

In this part, the author is conducting a spatial analysis examining the bicycle commuting percent in all the census tracts across New York City. To start with, the percent of bicycle commuters in total population of each census tract is visualized in a map, showing the spatial variation of bicycle commuting proportions across the city. Then similar quantified method will be used as conducted in the temporal analysis to verify the significant difference among areas with and without Citi Bike service, with extra attention drawn on the difference in gender. Furthermore, the author will zoom in to Citi Bike service area to explore the spatial pattern of Citi Bike usage with spatial statistical methods.

Having imported the preprocessed Citi Bike trips data into ArcGIS, the census tracts with Citi Bike station are recognized based on the spatial relationship as mentioned in Section 3.4, and highlighted in the black boundary as shown in Figure 1, which are also called the Citi Bike service area in this thesis. The gradients of grey colors are showing different percent levels of bicycle commuting in each census tract. Based on intuition, a majority of census tracts with high bicycle commuting percent, shown in dark grey, is found within or around the Citi Bike service area, meaning that a larger proportion of workers living in or near the service area tend to commute by bike.

Figure 2 Bicycle Commuting Percent Spatial Distribution (%)

If we look at some statistics, the average bicycle commuting percent is 0.98% for all census tracts in the city. In the Citi Bike service area, 65% of the census tracts have reached or exceeded the average line, while in the rest area, only 23% of the census tracts have above-average percent of bicycle commuters. (Figure 2) Looking specifically into the census tracts above average in terms of bicycle commuting percent, as shown in black, a quarter of these tracts fall in the Citi Bike service area, sharing only 6.5% of New York City's land area.

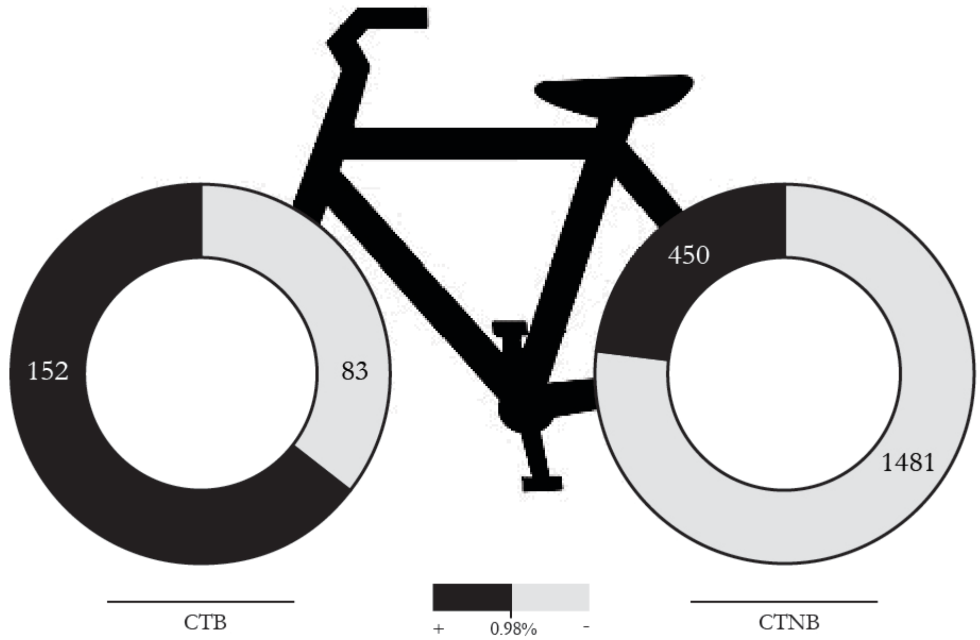

Figure 3 Number of Census Tracts Above or Below Average Percent of Bike Commuting

(CTB for census tracts with Citi Bike stations; CTNB for census tracts without bike stations)

Both the spatial distribution pattern and the proportions comparison are showing a sign that Citi Bike has laid considerable impact on worker's commuting modes. Census tracts within Citi Bike service area are presenting remarkably higher proportions of bicycle commuters. In order to statistically test such difference over space, which is actually done in part in the temporal analysis for comparison purpose, t-test would be conducted between CTB and CTNB. In Section 4.1, the author has already examined the significant difference in the percent change of bicycle commuters between CTB and CTNB. (Table 2) More specifically in this part, the author would examine the percent change by census tracts type for both male and female commuters respectively. The results show even stronger contrast than that in the temporal analysis, in the significance of different genders regarding the bicycle commuting percent change over space. (Table 5)

Table 5 T-test Results on Bicycle Commuting Percent Change by Census Tracts Type

		Obs	*Mean*	*SD*	*t*	*Pr(|T| > |t|)*	*Alpha*
Male Change	CTB	235	1.08	2.97	4.75	0	0.05
	CTNB	1880	0.14	1.65			
Female Change	CTB	234	0.11	1.56	0.20	0.84	0.05
	CTNB	1880	0.09	0.87			

Regardless of the variation in gender, more significant spatial pattern of Citi Bike usage could be presented through the "Spatial Statistics" tool in ArcGIS. Both the Cluster and Outlier Analysis and the Hot Spot Analysis are basically taking the values, which is the number of bike trips in this case, at certain location and the neighborhood into consideration, and aiming to find out the statistically significant cluster of high values (HH) and cluster of low values (LL). On contrary, high values surrounded by low values or low values surrounded by high values are considered as outliers.

The Cluster and Outlier Analysis result (Figure 3) turns out to be surprising. Almost all the high-value clusters, which could be considered as where the most popular Citi Bike stations aggregate, are found in Manhattan, while the majority of low-value clusters, where the bikes stations bearing low bike usage, is located in Brooklyn. The Hot Spot Analysis result (Figure 4) is to some extent showing a consistent trend in terms of the spatial distribution of hot spots and cold spots. The color gradient corresponds to different significance levels from 0.1 to 0.01 (90% confidence to 99% confidence), with darker color showing higher significance level. If we only look at the most significant clusters (those in dark red and dark blue), it could be noticed that there are far more hot spots than cold spots. The great number of the hot spots with high significance level indicates that the bike stations are probably having considerable positive impact on each other regarding bike usage. The possible reason for such distinctive regional distribution might be that Citi Bike stations located in Manhattan exhibit higher density on average, resulting in better connection among stations and more convenience provided to users when transiting from or towards another commuting mode.

This argument, however, might be doubted since higher density also indicates more stations in each census tract in Manhattan, which would therefore contribute to higher values of bike usage for each tract. Actually, evidence could be easily found that either number or density contributes to similar results, if the Hot Spot Analysis is conducted based on the value of average bike trips per census tract by making trip count divided by the number of bike stations in each tract. The Cluster and Outlier Analysis also proves the fact that higher value of bike usage per station occurs mostly in Manhattan. It is clear that such consistency of number and density in the spatial pattern is derived from the high correlation between number and density themselves, and thus no further clarification is necessary. What the author is trying to address here is that higher density probably accounts for larger influence that bike stations are laying on each other.

In order to verify the assumption of density difference, the average Citi Bike station density in each census tract will be compared between the service area in Manhattan and that in Brooklyn through a t-test. The result of one-tailed t-test proves that the average bike station density in Manhattan is statistically significantly higher than that in Brooklyn with the significance level of 0.025. It also indicates that the mutual influence among bike stations should be taken into account as a considerable factor in the analysis of bike usage determinants.

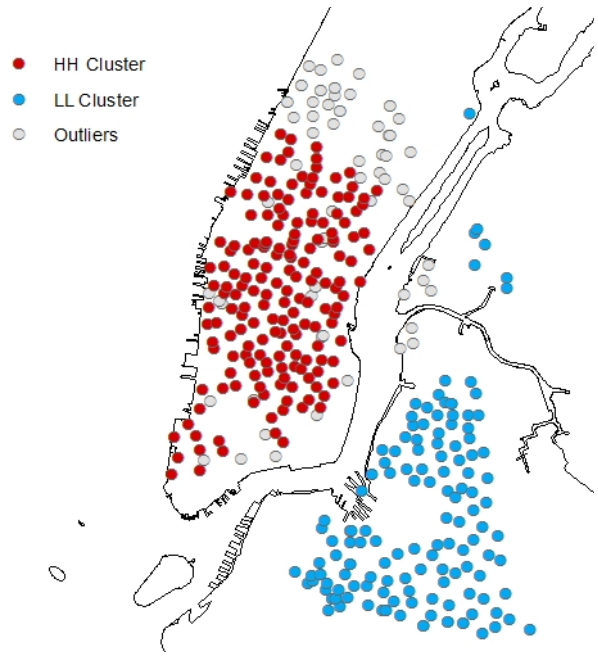

Figure 4 Cluster and Outlier Analysis

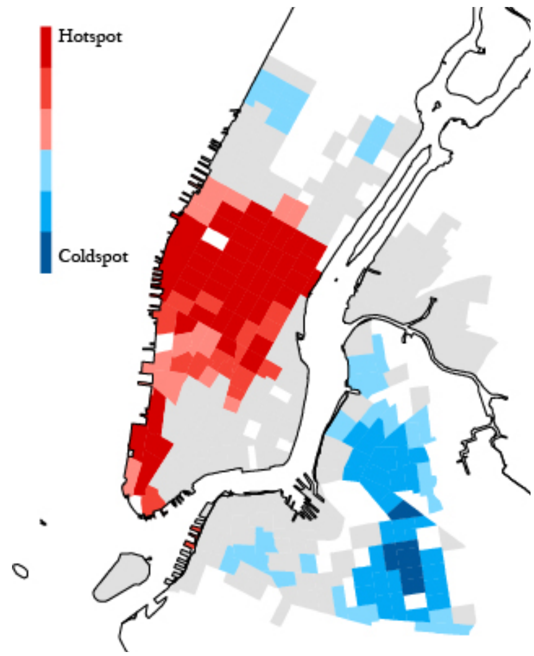

Figure 5 Hotspot Analysis

Table 6 T-test Result on Citi Bike Station Density

	Obs	*Mean*	*SD*	*t*	*Pr(T>t)*	*Alpha*
Manhattan	145	32.8	17.1	2.097	0.019	0.025
Brooklyn	90	27.7	18.3			

5 BIKE-SHARE USAGE ANALYSIS

Based on the temporal and spatial analysis on the influence of the bike-share system in New York City, we can tell that Citi Bike has made a significant difference to workers' choice of commuting mode, specifically increasing the bicycle commuting proportion. In this part, a deeper analysis of the determinants of bike-share usage for commuting purpose will be conducted through bivariate correlation analysis and multiple regression analysis.

5.1 DESCRIPTIVE STATISTICS

Before the analysis, Table 7 is showing a general description for the variables to be included in the correlation analysis with some basic statistics. All the analysis will be conducted based on the analysis unit of census tracts. So the number of observations is actually indicating the number of census tracts covered by the Citi Bike service, meaning that there are 241 census tracts with Citi Bike stations located inside based on the 2015 Citi Bike system data. The table is showing number of bike trips performed at both start station (origin) and end station (destination) because as the dependent variables, trips at start station and end station will be examined separately in the correlation analysis. Statistics of the other variables remain the same in either circumstance.

Table 7 Descriptive Statistics for Variables in the Analysis

Variable	Obs	Mean	SD	Min	Max
#Trips (Origin)	241	495.99	621.41	1	3,847
#Trips (Destination)	241	496.27	656.39	0	3,792
Bike Lane (mile)	241	5.09	2.55	0	10.69
Bike Rack (# per sq.mile)	241	493	488.74	0	3,083
#Subway Station (average)	241	1.75	1.33	0	5
Station Density (# per sq.mile)	241	30.35	17.85	1.64	136.29
Population Density	241	68,542	39,195	559.52	182,800
%Male	241	0.49	0.06	0.38	0.75
%Students	241	0.19	0.09	0	0.54
Median Income	241	87,427	43,074	12,479	232,266
%Bachelor or more	241	0.48	0.21	0.02	0.77

5.2 BIVARIATE CORRELATION

All variables other than the bike usage are involved in the bivariate correlation analysis so that the influence of each of them on bike usage is tested respectively. Each variable is divided into quartiles and the critical values to cut off the first and the fourth quartile are shown in the second and third columns in Table 8. Thus two groups are extracted based on each of the variables. One group contains the highest values of the certain variable, which is the fourth quartile. The other group contains the lowest values, which is the first quartile. Then for the two quartile groups of each variable, the means of the corresponding bike usage level are compared through a t-test. Take the variable bike lane for instance, the first quartile will be census tracts containing the least miles of bike lanes in the half-mile buffer, less than 3.0393 miles specifically, and the fourth quartile will consist of census tract with more than 7.1966 miles of bike lanes. Then the author examines the two means of trip counts in these two quartile groups. The difference of means in number of trips between the two quartiles is 677.2 and t-test shows the difference is statistically significant, indicating that the number of bike trips in the quartile with the most bike lanes is statistically greater than that in the lowest quartile. So there is a positive bivariate correlation between bike lane and Citi Bike usage which is measure by number of trips in each census tract, meaning that greater total miles of bike lanes in length in the buffer of census tracts will contribute to more bike trips.

The variables also showing positive bivariate correlation with bike usage level include bike rack, subway station, station density, median income, education attainment, and age. For the quartile of census tracts with the most bike racks, the number of Citi Bike trips is statistically greater than that in the quartile with the fewest racks. In census tracts where the quarter-mile buffer of every bike station covers the most subway stations, more bike trips are found than in those where bike stations have poor access to subway stations, which indicates many users are using Citi Bike to connect subway stations with their destination. In the quartile of census tracts with the highest bike station density, the bike usage level is also statistically higher than that in the lowest quartile. The gap is 567.7 trip records more on average, next to the highest difference range in the quartiles of bike lane. Similarly, median household income, percent of bachelor attainment or more, percent of students aged 5-24 are all showing positive bivariate correlation with Citi Bike usage, meaning that the higher value of these variables, the more bike trips will be found.

Population density, however, does not show significant impact on Citi Bike usage, getting the least difference between the first and the fourth quartile. The impact of the percent of male in total population is hardly significant with a p-value of 0.051 under 0.05 significance level. What needs to be noticed is that the percent of male examined here is not the same thing as what the author talked about in Chapter 4. In the temporal and spatial analysis, the author looked into the percent of male bicycle commuters in all commuting modes, which shows statistically significant difference between census tracts with Citi Bike service covered and those without. However, in the correlation analysis, it is only the percent of male in total population that is examined as a control variable, which seems to have little influence.

The author is only testing the bivariate correlation between determinants and number of bike trips performed at starting stations. Test conducted on bike trips counted at destinations could possibly result in different correlations with the same determinants. However, in this single test, the core determinants emphasized by this thesis, including bike lane, bike rack, subway station, and station density, are all showing strong correlation with bike usage. The purpose of this bivariate correlation is thus already achieved. In the multiple regression, more comprehensive analysis will be conducted on Citi Bike trips counted at both start stations and end stations.

Table 8 Bivariate Correlation with Citi Bike Usage

Variable	25% Percentile	75% Percentile	Difference in #Trip (4th quartile – 1st quartile)	Significance (p-value)
Bike Lane (mile)	3.0393	7.1966	677.2	0.000
Bike Rack (#/sq.m)	118.31	706.37	342.4	0.002
#Subway Station	1	3	399.0	0.000
Station Density (#/sq.m)	15.9954	37.8004	567.7	0.000
Population Density	39,838	95,445	95.0	0.411
Median Income	54,653	110,793	425.4	0.000
%Bachelor or more	0.3226	0.6508	464.7	0.000
%Male	0.451	0.517	245.7	0.051
%Student	0.122	0.232	214.3	0.034

(Alpha=0.05, two-tailed t-test with unequal variance)

5.3 MULTIPLE REGRESSION

The bivariate correlation analysis presented above only examined the relationship between Citi Bike usage and each independent variable, one at a time. The multiple regression model to be estimated below examines the comprehensive impact of all the independent variables, including availability of bike lanes and bike racks, proximity to subway stations, and bike station density, while controlling for population density, percent of male population, percent of students, median household income and percent of bachelor attainment or more. It is assumed that Citi Bike trips counted at the start stations and the end stations would result in different relationship with the independent variables, so the multiple regression model will be estimated separately regarding origin bike trips and destination bike trips. Then comparison will be made based on the difference in the two regression results.

5.3.1 ORIGIN BIKE TRIPS

The basic analysis units are census tracts in New York City. All the attributes of variables have been attached to census tracts units based on carefully matching and transformation from normal table data to spatial attributes as explained in the methodology. The first multiple regression model is estimated in terms of the impact of each independent variable on Citi Bike usage, measured by total number of bike trips counted at start stations per census tract. (Table 9) The model gets the Adjusted R-squared of 0.39, indicating that the regression model explains 39% of the variation in the number of bike trips. All the independent variables have joint significance at the 95% level (F=18.02, Pr>F=0), indicating the model overall is statistically significant. The 241 observations are referring to the 241 census tracts with Citi Bike station located inside, or in other words, within the Citi Bike service area.

Table 9 Multiple Regression of Origin Bike Trips (#Trip per CT)

Variable	COE	SE	t-value	p-value
Bike Lane (mile)	81.02	15.61	5.19	0.000*
Bike Rack (#/sq.m)	-0.19	0.08	-2.50	0.013*
#Subway Station	10.15	30.14	0.34	0.737
Station Density (#/sq.m)	11.66	1.85	6.30	0.000*
Population Density	-0.0001	0.0009	-0.15	0.878
%Male	1431	610.3	2.35	0.020*
%Student	1192	525.6	2.27	0.024*
Median Income	0.0005	0.0014	0.36	0.722
%Bachelor or more	1037	382.2	2.71	0.007*

(Obs = 241; Adj R^2 = 0.39; F = 18.02, Pr>F = 0.00; Alpha = 0.05)

Most of the variables are showing the expected signs, not all of them are statistically significant though. The average number of subway stations within quarter-mile of each bike station does not show statistically significant impact on Citi Bike usage. Neither do two of the control variables, population density and median household income. Among the significant variables, however, bike racks per square mile is showing negative impact on Citi Bike usage, although the impact is slight since the coefficient (COE) is -0.19, meaning that every five-unit increase in the number of bike racks per square mile will result in one less bike trip. Some of the signs indicate that the accuracy of the multiple regression model might be doubted, so the error terms, the multicollinearity problem, and outliers should be further checked.

The test of Variance Inflation Factors for independent variables (VIF) is conducted to examine the multicollinearity among all the independent variables and the control variables. The result confirms the multicollinearity problem especially in the control variable, percent of bachelor attainment or more with the highest VIF value of 6.47. In order to improve the accuracy of the regression model, such variable with fairly high VIF value should be removed from the model.

The predicted bike trips and error terms are calculated for each trip record. Results of the skewness and kurtosis tests showed that there is a statistically significant difference between the distribution of error terms in this model and a normal distribution. In other words, the error terms in this model are not normally distributed, which is inconsistent with the assumption of OLS regression. In order to fix this problem, outliers are carefully examined and finally 6 outliers with the absolute value of rstandard greater than 2.58 are removed from the trip records.

After the adjustment of the model and the dataset, the multiple regression model is reestimated. (Table 10) In comparison, the Adjusted R-squared has been raised from 0.39 to 0.44, indicating the model is now explaining 5% more of the variation in Citi Bike usage. All the independent variables account for 44% of the variability in the number of Citi Bike trips per census tract.

Table 10 Adjusted Multiple Regression Model of Origin Bike Trips

Variable	COE	SE	t-value	p-value
Bike Lane (mile)	76.93	12.20	6.30	0.000*
Bike Rack (#/sq.m)	-0.08	0.06	-1.33	0.184
#Subway Station	38.86	23.94	1.62	0.106
Station Density (#/sq.m)	10.10	1.65	6.13	0.000*
Population Density	0.0006	0.0007	0.79	0.428
%Male	732.70	494.69	1.48	0.140
%Student	-1.22	292.57	-0.00	0.997
Median Income	0.0028	0.0007	4.06	0.000*

(Obs = 235; Adj R^2 = 0.44; F = 23.73, Pr>F = 0.00; Alpha = 0.05)

Bike lane and station density has been showing high statistical significance level before and after the model is adjusted. With the coefficient of 76.93, every one-mile increase in the total bike lane length within the half-mile buffer around each census tract will contribute to approximately 77 Citi Bike trips increase in number in the corresponding census tract. There is no doubt that bike lanes have the most significant impact on Citi Bike usage, playing one of the most important role as infrastructure in promoting bicycle commuting. The impact of station density in each census tract remains on the same level in the two regression models. For every one-unit increase in station density in a census tract, about 10 more Citi Bike trips could be found in the same tract. The significance of station density also indicates that Citi Bike stations are laying considerable mutual influence on each other. Higher station density, which contributes to more bike trips, probably means more flexible route choice of bike trips and better connection between the origins and destinations. As a result, it would turn out to be more convenient for workers to pick up Citi Bike as a commuting tool for the whole trip or just part of it as a connection to the other public transit modes.

However, in the adjusted model, many variables turn out to be insignificant, including bike rack and subway station. Comparing to the negative correlation with Citi Bike usage shown in the former model, the insignificant result of bike rack seems to make more sense. Unlike ordinary cyclists, Citi Bike users do not have rigid demand towards bike racks since the network of Citi Bike stations is providing sufficient convenience for users to pick up and drop off bikes. Subway station access is assumed to have significant influence on Citi Bike usage because the flexibility of Citi Bike trips makes it a possible alternative transit tool in many areas with poor access to subway service. However, the average number of subway stations in quarter-mile buffer of each bike station does not show a statistically significant impact on Citi Bike usage. Nevertheless, the assumption that Citi Bike plays the role of alternative transit tool could not be denied. Many workers might prefer to use Citi Bike as a last-mile transportation solution from train stations or subway station to work, as opposed to taking a Citi Bike trip when beginning a commuting trip to a transit station. As a result, subway station turns out to be insignificant in the multivariate correlation test on Citi Bike trips counted at start stations. That is also the reason why separate models are estimated regarding origin bike trips and destination bike trips, as bicycle commuting behaviors could probably be determined by different factors in these two circumstances.

5.3.2 DESTINATION BIKE TRIPS

The Citi Bike trips counted at the end stations, also called destinations, should also be taken into consideration because the impact of the independent variables could be quite different at trip origins and destinations. The second multiple regression model below is estimated in terms of the impact of each independent variable on Citi Bike usage, measured by total number of bike trips counted at end stations per census tract. (Table 11) The model gets the Adjusted R-squared of 0.41, indicating that the regression model explains 41% of the variation in the number of bike trips. All the independent variables have joint significance at the 95% level (F=19.47, Pr>F=0), indicating the model overall is statistically significant. The 241 observations are referring to the 241 census tracts with Citi Bike station located inside, or in other words, within the Citi Bike service area.

Table 11 Multiple Regression of Destination Bike Trips (#Trips per CT)

Variable	COE	SE	t-value	p-value
Bike Lane (mile)	77.61	16.22	4.79	0.000*
Bike Rack (#/sq.m)	-0.16	0.08	-2.02	0.045*
#Subway Station	42.94	31.29	1.37	0.171
Station Density (#/sq.m)	11.65	1.88	6.20	0.000*
Population Density	-0.0006	0.0010	-0.66	0.511
%Male	1518	634.21	2.39	0.017*
%Student	1087	546.16	1.99	0.048*
Median Income	0.0019	0.0015	1.28	0.201
%Bachelor or more	778.33	397.81	1.96	0.052

(Obs = 241; Adj R^2 = 0.41; F = 19.47, Pr>F = 0.00; Alpha = 0.05)

The results are showing very similar signs as the multiple regression of origin bike trips is before adjustment. Bike lane and station density are still at very high significance level with p-value of 0.000. Bike rack is again statistically significant in the initial model. The p-value of 0.045 is pretty close to the critical value of 0.05. So the further adjustment might have large impact on the significance level of bike rack. Subway station is not showing the sign as expected to have significant impact on Citi Bike usage counted at end stations. In terms of the control variable, the percent of male and percent of students are showing significant impact on Citi Bike usage, while none of population density, median household income, percent of bachelor attainment or more turns out to be statistically significant. Getting no perfect signs as expected, the author will once more improve the accuracy of the multiple regression model by checking the error terms, the multicollinearity problem, as well as outliers.

The VIF test is conducted to examine the multicollinearity among all the independent variables and the control variables. The result also confirms the multicollinearity problem especially in the control variable, percent of bachelor attainment or more with the highest VIF value of 6.49. In order to improve the accuracy of the regression model, such variable with fairly high VIF value should be removed from the model.

In order to check the error terms and outliers, the predicted bike trips and error terms are predicted for each trip record. With the results of the skewness and kurtosis tests showing statistically significant difference between the distribution of error terms in this model and a normal distribution, the error terms in this model are not normally distributed, which is inconsistent with the assumption of OLS regression. In order to fix this problem, outliers are carefully examined and finally 8 outliers with the absolute value of rstandard greater than 2.58 are removed from the trip records.

As a result of the adjustment in the model and dataset, specifically dropping a multicollinear variable, the percent of bachelor attainment or more and 8 influential outliers, the multiple regression model of destination bike trips is reestimated, getting an Adjusted R-squared of 0.46, raised by 5% from 0.41 in the former model. (Table 12) All the independent variables now account for 46% of the variability in the number of Citi Bike trips per census tract. The F statistics is still showing that the adjusted model is overall statistically significant at 0.05 confidence level. The 233 observations are the number of census tracts included in the analysis after dropping 8 outliers.

Table 12 Adjusted Multiple Regression of Destination Bike Trips

Variable	COE	SE	t-value	p-value
Bike Lane (mile)	76.76	12.19	6.30	0.000*
Bike Rack (#/sq.m)	-0.03	0.06	-0.46	0.648
#Subway Station	56.59	23.99	2.36	0.019*
Station Density (#/sq.m)	9.15	1.67	5.48	0.000*
Population Density	-0.0002	0.0007	-0.32	0.748
%Male	320.92	493.24	0.65	0.516
%Student	216.18	290.10	0.75	0.457
Median Income	0.0032	0.0007	4.81	0.000*

(Obs = 233; Adj R^2 = 0.46; F = 26.12, Pr>F = 0.00; Alpha = 0.05)

The core independent variables are all showing the expected signs. Bike lane is still statistically significant, indicating that for every one-mile increase of total bike lane length within the half-mile buffer of each census tract, about 77 more Citi Bike trips would be found. Although mile is a large-scale unit, with the coefficient of 76.76, bike lane is surely one of the most important factors that encourage workers to integrate Citi Bike to their commuting mode. Higher value of total bike lane length in the buffer actually means better access to bike lanes in the surrounding area of the destination. Thus it would be safer and less congested for users to access their workplace located in such census tracts.

Also, station density has as significant impact on Citi Bike usage as bike lane does. Every one-unit increase in the number of bike stations per square mile in each census tract will approximately contribute to 9 more bike trips ending in this census tract. In other words, high station density in a certain census tract will attract more commuters who work in the corresponding census tract to use

Citi Bike to cover part of or their whole commuting route from home to work, since it would be more convenient to drop off the bikes as closed to their workplace as possible.

Subway station finally shows statistically significant impact on Citi Bike usage after the adjustment. For every one-unit increase in the average subway station amount within a quarter-mile buffer of each bike station in each census tract, Citi bike usage ending in this tract will be raised by 57 trips. The significance of this variable provides strong evidence for the assumption that many commuters would prefer to use Citi Bike as their last-mile transit tool to connect work place with public transit stations. Citi Bike service plays an important role in the areas with poor access to subway service. This is not only about the distance to the nearest subway station, but about the access to alternative subway services, because one single subway line, covering very limited area of destinations, can hardly satisfy citizens' commuting demand with high uncertainty of flowing direction, regardless of the difference in the convenience of each subway line to transfer to another. That also explains why this variable is measured by the number of subway stations within a buffer of each bike station instead of the distance to the nearest subway station, and why the buffer radius is set to a quarter mile, the critical value of comfortable walking distance which is assumed to be shorter for commuting trips than in normal conditions.

Bike rack density turns to be insignificant in the adjusted model as in the multiple regression of origin bike trips. Even though it showed significant impact on Citi Bike usage in the initial model, the coefficient was -0.16, indicating very slight negative impact on the number of Citi Bike trips counted at end stations. (Table 11) Bike racks do not really account for much variability in Citi Bike usage since the bike-share system is providing sufficient facilities for bike parking through widely distributed bike stations according to the predicted bike flow at each station.

As is the case with the multiple regression of origin bike trips, median household income becomes statistically significant after percent of bachelor attainment or more is removed. It is showing the sign that more trips flow to census tracts with higher income level. The possible reason could be that there are more jobs in such census tracts, which is under verified in this thesis and needs further exploration. In terms of other controls variables, none of population density, percent of male population or percent of students is presenting statistically significant impact on Citi Bike usage.

5.4 SUMMARY

The multiple regression mainly examined the impact of infrastructure on Citi Bike usage on census tract level while controlling for some major related demographic characteristics. Basically, two regression models are estimated on Citi Bike usage measured by the number of bike trips counted at start stations and that counted at end stations respectively. Both models have been adjusted based on the examination of the multicollinearity problem, the distribution of error terms, and influential outliers. The two final models present different signs regarding the same independent variables, but provides the evidence to support each other.

Bike lane has always been statistically significant at the 95% confidence level. It is the most important piece of bike infrastructure encouraging more Citi Bike trips for the purpose of commuting no matter near the origins or the destinations. The variable is measured by the total length of bike lanes within a buffer of the same scale, half a mile, around the centroid of each census tract so as to include the impact of bike lane supply not only within the census tract but also in the broader neighborhood around, because what matters most in terms of bike lanes are the continuity in a certain range as well as the connectivity of the bike lane network.

Station density shows the same significance as bike lanes do in both models. There is no doubt that denser distribution of bike stations in a census tract will contribute to more bike commuting trips go across this region by providing better connected Citi Bike network. More convenience is thus created, also meaning better access to bike stations, so that commuters could pick up or drop off bikes as close to home, or work places, or transit stations as possible.

Subway stations only show statistically significant impact on Citi Bike usage in the last model when destination bike trips are examined. This difference presented in the two models looking into origin bike trips and destination bike trips providing persuasive argument that commuters are more possibly using Citi Bike as a solution for the last-mile transit problem in areas with poor access to subway service, than transferring from Citi Bike to subways as a home-subway connection. Actually, the measurement of distance to the nearest subway station was adopted at the first place. The multiple regression was run several times looking into either the average distance to the nearest subway station of each bike station or the distance from the centroid of each census tract to the nearest subway station. None of the measurement with regard to the minimum distance accessing the subways presented statistically significant result. The reason is probably that the Citi Bike service area, mostly covering lower Manhattan and downtown Brooklyn, is the best connected region in New York City in terms of public transport. The possibility of getting easy access to subways varies little across this region. So it is no surprise to get insignificant results by simply testing the distance to the nearest subway station.

Bike racks show unstable significance in both models before and after adjustment and eventually become insignificant in the final models. The uncertain impact reminds us that bike racks, showing considerable influence on cycling level in many early studies, is not a demand for Citi Bike users since the bike-share system supplies parking facilities itself which are users' prior choice. Therefore, the importance of station density stands out further in terms of providing more flexible trips nodes incorporating both origin and destination function.

Population density never shows significant impact on Citi Bike usage which is consistent with the result of bivariate correlation analysis. Nor does the percent of male population, although it shows significant impact in the under-adjusted models. It is the same case with the percent of students. Neither the percent of male nor the percent of students shows constant significance, and therefore, they are considered to have no significant impact on Citi Bike usage.

Census tracts with higher income level or more population of higher educational attainment tend to output more bike trips as the origins, and attract more trips as the destinations. However, these two variables are found to have high probability of multicollinearity, as a result of which, one of them need to be removed from the multiple regression model. The percent of bachelor attainment or more is actually removed due to the highest VIF score. Thus, the median household income turns out to be statistically significant in both models. The opposite way would result in similar effect – the percent of bachelor or more would result in high significance level if median income were removed. Both ways were tested but the latter was not presented in the regression results. Given such possible inter relation between educational attainment and income level, it is reasonable to interpret that there would probably be more Citi Bike trips in census tracts with either high income level or more population who have attained a bachelor's degree or more.

5.5 LIMITATION

This study is limited in many aspects. The author has adopted a bold methodology to measure and analyze Citi Bike usage and its determinants by integrating all the attributes of variables to the basic spatial analysis units – census tracts. Although the multiple regression models have shown optimistic results regarding the hypothesis and the expected findings of this study, the validity of such method is still under discussion. In addition, in terms of variable measurement, it is kind of arbitrary to decide the radius of the buffers, for instance, the half-mile buffer for bike lane measurement and the quarter-mile buffer for accessible subway station counting.

The dataset selected for this study could be biased since target bike trips are filtered based on several assumptions. For instance, bike trips performed during 7-9am and 4-6pm on weekdays by subscribing users are assumed to be commuting trips. In fact, bike trips for other purposes, such as exercising or amusement, could possibly be mixed in this sample. The time period is also set based on the working hours of conventional jobs.

In the final multiple regression models, most of the control variables are insignificant, and the Adjusted R-squared below 0.5 is showing the models fail to explain half of the variation in the dependent variable, both indicating that more valid control variables need to be integrated to improve the robustness of the models. Besides, both models are estimated under the significance level of 0.05. It would make more sense to conduct analysis under multiple significance levels. In that case, we will be able to obtain significant determinants with different hierarchy of impact.

This thesis is trying to explore the influential factors of bicycle commuting by looking into the usage of Citi Bike, aiming to provide suggestions that help integrate bicycle to public transit modes.

However, another major cycling group, the conventional bike riders, is not covered by the analysis. These two dominant cycling modes probably have potential mutual influence on each other, which should be taken into serious consideration in further exploration. Also, the analysis in this study is based on bike commuting and infrastructure in New York City, where people have access to the best connected public transit network and best-supplied infrastructures in the United States. The method and findings are hardly applicable to other regions, even large cities, around the world.

6 DISCUSSION AND CONCLUSION

This study has integrated spatial analysis and statistical methods to explore the influence of the bike-share system and the determinants of bike-share usage in the background of New York City. Compared to existing studies researching cycling level on the national or metropolitan scales, the analysis conducted in this thesis looked at a much smaller scale based on the analyzing units of census tracts. However, the analysis is not limited to the boundary of census tracts, taking the impact of neighborhood into consideration when necessary. Nor is the method limited to tracking individual users or single bike trip. Taking the spatial units, census tracts, as a whole makes it possible to conduct more comprehensive analysis on bike usage, which contributes to reasonable findings that make good sense from the perspective of urban planning.

Based on the temporal analysis on the influence of Citi Bike, the percent of bicycle commuters in all transit modes is found to have a significant increase after Citi Bike started service in certain areas. The percent variation of male bicycle commuter has widened due to Citi Bike service, while female percent varies little since female commuters are potentially showing more concern about safety issues. Comparatively, the spatial analysis indicates that the areas with Citi Bike service show more significant increase in the percent of bike commuting than those areas with no Citi Bike service. In the Citi Bike service area specifically, there is an eminent spatial pattern that bike trip hotspots are concentrated in lower Manhattan, meaning that the most popular Citi Bike stations cluster at certain areas in Manhattan. That also indicates the possible mutual influence of bike stations which is verified in the further analysis.

The core analysis in this study is the correlation analysis on the determinants of bike-share usage. Bike lane, subway station accessibility and bike station density are showing statistically significant impact on Citi Bike usage on the census tract level, while bike racks present little impact due to the built-in bike parking function of bike-share systems. Nonetheless, there are still several factors not covered by the regression models. The junction density of bike lanes would probably influence Citi Bike usage over space since it is a good measurement of the connectivity of bike lane network. Distance to the nearest subway station, which is found to have little impact on bike usage, might turn out to be influential in areas with very low density of subway stations, which deserves further exploration. In addition, as the control variables, not many of demographic characteristics present significant results. Some of them are showing uncertain impact with significance level varying in a large range. Inclusion of more control variables and improvement of measurement could be considered to raise the robustness of the multiple regression model.

These findings indicate that Citi Bike has been used as an efficient commuting tool and has become an essential part of public transport. Several ideas in this thesis could be implemented to encourage more commuters travelling by bike by improving infrastructures based on the determinant analysis of bike-share usage. For one thing, considering bike users' safety concern, more dedicated bike lanes need to be constructed especially in crowded areas. Given that female commuters share much lower percent of bicycle commuting than male, and females show more concern about safety, better access to bike lanes, which also means safer cycling environment, will particularly attract more potential female commuters to cycle. For another, a good implementation in Citi Bike's expansion plan is to build continuous and well-connected bike-share network, which as a result, will provide huge convenience for users to pick up and drop off bikes. Other than encouraging bicycle commuting, the functionality of Citi Bike as the last-mile-transit solution could also be taken into consideration in transportation planning since this thesis has proved that Citi Bike does a good job connecting workplace with transit stations. For instance, better location bike stations could possibly improve Citi Bike's connection with transit stations.

Furthermore, the potential demand outside Citi Bike service area could be predicted based on the findings of this study. The determinants could be given different weights according to the significance level, expecting to see different intensity of demand over space. Another idea is referring to the cab trip data with certain features, as the way selecting bike commuting trips. The reason that cab trip data would work is that bike-share has been considered as an alternative mode to replace some of the taxi trips, especially those playing a connection role between transit station and workplaces or those short-range commuting trips. However, as the rising of Uber, the contribution of bike-share becomes hard to discern. The potential mutual influence between bike-share and Uber could also interfere the judgment towards the role of bike-share system. Such new rising transport mode should draw extra attention when further analyzing the impact and predicting the demand of bike-share system. The predicted demand of Citi Bike will be able to help policy makers carry out plans calling for more green traffic and healthy travelling manner.

7 BIBLIOGRAPHY

[1] Implementing a Successful Bicycle and Active Commuting Program in the Washington, DC Metropolitan Area, May 21, 2010.

[2] bikesmakelifebetter.com

[3] Pucher, John R., and Ralph Buehler. Analysis of bicycling trends and policies in large North American cities: Lessons for New York. University Transportation Research Center Region 2, 2011.

[4] McKenzie, Brian. "Modes less traveled—bicycling and walking to work in the United States: 2008–2012." US Census Bureau, New York (2014).

[5] Buehler, Ralph. "Determinants of bicycle commuting in the Washington, DC region: The role of bicycle parking, cyclist showers, and free car parking at work." Transportation research part D: transport and environment 17, no. 7 (2012): 525-531.

[6] Kölbl, Robert, and Dirk Helbing. "Energy laws in human travel behaviour." New Journal of Physics 5, no. 1 (2003): 48.

[7] Rietveld, Piet, and Vanessa Daniel. "Determinants of bicycle use: do municipal policies matter?" Transportation Research Part A: Policy and Practice 38, no. 7 (2004): 531-550.

[8] Rojas-Rueda, David, Audrey de Nazelle, Marko Tainio, and Mark J. Nieuwenhuijsen. "The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study." Bmj 343 (2011): d4521.

[9] AASHTO Executive Committee. "Guide for the Development of Bicycle Facilities." Washington, DC: American Association of State Highway and Transportation Officials (1999).

[10] Singhvi, Divya, Somya Singhvi, Peter I. Frazier, Shane G. Henderson, Eoin O'Mahony, David B. Shmoys, and Dawn B. Woodard. "Predicting Bike Usage for New York City's Bike Sharing System." In Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.

[11] Fishman, Elliot. "Bikeshare: A Review of Recent Literature." Transport Reviews ahead-of-print (2015): 1-22.

[12] McKenzie, Brian, and Melanie Rapino. Commuting in the united states: 2009. US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 2011.

[13] Dill, Jennifer. "Bicycling for transportation and health: the role of infrastructure." Journal of public health policy (2009): S95-S110.

[14] Kaufman, Sarah. "Citi Bike and Gender." NYU Rudin Center for Transportation, 2014.

[15] Pfrommer, Julius, Joseph Warrington, Georg Schildbach, and Manfred Morari. "Dynamic vehicle redistribution and online price incentives in shared mobility systems." Intelligent Transportation Systems, IEEE Transactions on 15, no. 4 (2014): 1567-1578.

[16] Buck, Darren, and Ralph Buehler. "Bike lanes and other determinants of capital bikeshare trips." In 91st Transportation Research Board Annual Meeting. 2012.

[17] Chen, Lijun, Adam Pel, Xuewu Chen, Daniel Sparing, and Ingo Hansen. "Determinants of Bicycle Transfer Demand at Metro Stations: Analysis of Stations in Nanjing, China." Transportation Research Record: Journal of the Transportation Research Board 2276 (2012): 131-137.

[18] Martin, Elliot W., and Susan A. Shaheen. "Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two US cities." Journal of Transport Geography 41 (2014): 315-324.

[19] Faghih-Imani, Ahmadreza, Naveen Eluru, Ahmed M. El-Geneidy, Michael Rabbat, and Usama Haq. "How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal." Journal of Transport Geography 41 (2014): 306-314.

[20] Daddio, David William. "Maximizing Bicycle Sharing: an empirical analysis of capital bikeshare usage." PhD diss., University of North Carolina at Chapel Hill, 2012.

[21] Passafaro, Paola, Alessandra Rimano, Maria Paola Piccini, Renata Metastasio, Valeria Gambardella, Gisella Gullace, and Consolata Lettieri. "The bicycle and the city: Desires and emotions versus attitudes, habits and norms." Journal of environmental psychology 38 (2014): 76-83.

[22] Jurdak, Raja. "The impact of cost and network topology on urban mobility: A study of public bicycle usage in 2 US cities." PloS one 8, no. 11 (2013): e79396.

[23] García-Palomares, Juan Carlos, Javier Gutiérrez, and Marta Latorre. "Optimizing the location of stations in bike-sharing programs: a GIS approach." Applied Geography 35, no. 1 (2012): 235-246.

[24] Kaltenbrunner, Andreas, Rodrigo Meza, Jens Grivolla, Joan Codina, and Rafael Banchs. "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system." Pervasive and Mobile Computing 6, no. 4 (2010): 455-466.

[25] Wang, Xize, Greg Lindsey, Jessica E. Schoner, and Andrew Harrison. "Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and from Stations." Journal of Urban Planning and Development (2015): 04015001.