The Effects of Urban Density on the Efficiency of Dockless Bike Sharing System

- A Case Study of Beijing, China

by

Wencong Cui

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Urban and Environmental Planning

Approved April 2018 by the
Graduate Supervisory Committee:

Michael Kuby, Chair
Deborah Salon
Calvin Thigpen

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Bicycle sharing systems (BSS) operate on five continents, and they change quickly with technological innovations. The newest “dockless” systems eliminate both docks and stations, and have become popular in China since their launch in 2016. The rapid increase in dockless system use has exposed its drawbacks. Without the order imposed by docks and stations, bike parking has become problematic. In the areas of densest use, the central business districts of large cities, dockless systems have resulted in chaotic piling of bikes and need for frequent rebalancing of bikes to other locations. In low-density zones, on the other hand, it may be difficult for customers to find a bike, and bikes may go unused for long periods. Using big data from the Mobike BSS in Beijing, I analyzed the relationship between building density and the efficiency of dockless BSS. Density is negatively correlated with bicycle idle time, and positively correlated with rebalancing. Understanding the effects of density on BSS efficiency can help BSS operators and municipalities improve the operating efficiency of BSS, increase regional cycling volume, and solve the bicycle rebalancing problem in dockless systems. It can also be useful to cities considering what kind of BSS to adopt.

DEDICATION

"While his parents are living, a son should not go far abroad; if he does, he should let them know where he goes." — *Confucius*

This thesis work is dedicated to both my parents, for their love, endless support and encouragement during the challenges of graduate school and life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Over the past decades, bike sharing systems (BSS) have expanded rapidly all over the world. Attributed to the creation of new technologies, the systems are innovating continuously. The newest type of "dockless" bike sharing system, which eliminates docks and stations, has become popular in China since early 2016.

As of now, the BSS, born in 1965 in Europe, has been developed for three generations. These include the first generation, called white bikes (or free bikes); the second generation of coin-deposit systems; and the third generation, or information technology (IT) based systems. These three generations are collectively referred to as docked systems, which need several fixed stations with docks in each station used to store bicycles and finish rent and return operations. The dockless system, also considered as the fourth-generation system, based on mobile app and GPS, which completely eliminates stations and docks. Passengers can easily pick up and drop off bikes anywhere using their cell phone.

The rapid increase in use of this successful system has exposed its drawbacks. Without the regulation of docks and stations used in regular public bike systems, bike parking has become a major problem, and in some urban hot spots, like the Central Business District (CBD) and tourist sites, it has led to a chaos, especially during rush hour (Figure 1).

Cities and towns that want to improve their transportation systems with bike sharing can learn much from the results of BSS implementation elsewhere. One of the problems they will need to deal with if they choose a dockless systems is chaotic parking. Here I argue that building density is one of the most important factors influencing the efficiency of dockless BSS.

Figure 1. Chaotic parking problem of dockless systems

Source: http://news.hexun.com

China first innovated the dockless system at the end of 2015, and most of users worldwide are still found in China. Dockless trip data are not open access. Literature about these systems are hard to find. Because of the lack of data, the only research about this dockless systems was done by one of the bike sharing operation companies in collaboration with Tsinghua University Urban Planning Institute (Sharing Bicycles and

Urban Development White Papers 2017). This white paper mainly discusses the trip characteristic and riders' behavior based on the data from all over China. It shows that the modal share of dockless bike systems have reached 6.8% of all trips, while traditional bike trips account for only 4.8%. However, if we have a look at dock system, a traditional type system which was first operated in Denmark in 1991 (DeMaio, 2009), with docks and stations, there is a larger literature. Generally, these studies can be divided into three perspectives of actions for improving bicycle-sharing systems: infrastructure planning and utilization; measures for integrating bicycles with public transit; and resolutions for rebalancing problem (Fishman, Washington, & Haworth, 2013).

For dock station infrastructure planning, García-Palomares et al. (2012) applied a GIS-based location-allocation model to optimize bike-sharing stations in central Madrid, and tested two solutions in this model: minimizing impedance (minimizes the distance between supply and demand) and maximizing coverage (the stations are concentrated in the zones with greatest potential demand). The empirical study shows that the maximize-coverage solution is more efficient in Madrid because it maximizes the potential demand covered by the stations and the minimize-impedance solution is more advantageous in terms of spatial equity because it generates a more uniform coverage. Some research shows differences in the usage rate of bike share programs globally. Usage rates vary from around zero to eight trips per bicycle per day (Fishman et al., 2013). In addition to the spatial differences, the usage rate also varies over time and purpose of trip. Randriamanamihaga, Côme, Oukhellou, & Govaert, (2014) applied Poisson mixture models on two months of trips data recorded on the Vélib' Bike-Sharing System of Paris and found that the usage of the system is mostly governed by the weekday/weekend distinction, which clearly differentiates between utilitarian and recreational usage. Faghih-Imani et al. (2014) did a comprehensive study on exploring the factors affecting bicycle-sharing flows and usage based on the data obtained from a major bicycle-sharing system (BIXI) in Montreal. They examined the influence of meteorological data, temporal characteristics, bicycle infrastructure, land use and built environment attributes on arrival and departure flows at the station level using a multilevel approach to statistical modeling. As a result, they found that adding additional stations (either by relocating large stations to smaller stations with lower capacity in multiple locations or adding new bicycle stations) is more beneficial in terms of arrival, departure flows and usage rate compared to adding capacity to existing stations. This research also found that population density of a station's transport analysis zone (TAZ) positively affects the bicycle flows.

Integrating bicycle-sharing with public transit is also a meaningful and practical research area. In most scenarios, cycling is not the only transportation mode for the whole trip: rather it is integrated with other modes, such as metro and bus, and affected by them. BSSs have frequently been cited as a way to solve the "last mile" problem and connect users to public transit networks. Fishman et al. (2013) identify two themes of the relationship between BSS and public transit: modal integration and modal substitution. The first focuses on the location of bikeshare infrastructure near transit stations so that passengers can use bikeshare in conjunction with transit. The other refers to trips that shift from public transit to bikeshare. The modal integration of BSSs and public transit has been shown to strengthen the benefits of both modes in Netherlands (Brons, Givoni, & Rietveld, 2009), North America (Ma, Liu, & Erdoğan, 2015), Australia (Pucher & Buehler, 2012) and China (Pan, Shen, & Xue, 2010). In New York City, BSS stations that are located near subway stations, particularly stations with high monthly boardings, showed higher bikeshare usage (Noland et al., 2016). In China, a study on the different rider group of BSS in five cities shows that commuters using them as a supplementary form of travel for the first/mile of their journey to and from work (Zhang, Zhang, Duan, & Bryde, 2015). Jäppinen et al. (2013) modeled the travel times between the population and 16 important destinations in the central Helsinki by public transportation alone compared with public transportation extended with shared bikes. The result is that a large-scale bicycle sharing system could complement a traditional public transport system and thereby improve accessibility. Regarding modal substitution, the evidence from New York City shows that after separately controlling for bike lane infrastructure, almost 50% of trips now made by bikeshare were previously made by bus. The model without bike lanes suggests that approximately 70% of bikeshare members may be substituting bikeshare for bus use (Campbell & Brakewood, 2017). Coincidentally, another case study of a BSS in Shanghai showed that a large share of BSS users shift from walking and mass transit to use bike-sharing system for commuting and other daily activities (Zhu, Pang, Wang, & Timmermans, 2013). Another study of Washington DC and Minneapolis suggests that the denser the bikesharing network and the denser the urban form, the more bikeshare members substitute biking for public transit (Martin & Shaheen, 2014). However, Singleton and Clifton (2014) mentioned that the relationship between public transit and BSS can vary by time. In the short-run, BSS can be a substitute of public transit, which means passengers will shift to bike-sharing system from public transit. However, they could become complements in the long-term, in other words BSS and public transit can work together, a public transit network and BSS network could impact future travel behavior.

The final group of bike-sharing studies, the bicycle rebalancing problem (BRP) for dock systems, is also a popular research area. Most of these papers are talking about the design of solution algorithms (Erdoğan, Battarra, & Wolfler Calvo, 2015; Erdoğan, Laporte, & Wolfler Calvo, 2014). Some papers use empirical data. For example, Dell'Amico et al. (2014) use Mixed Integer Linear Programming and a Branch-and-Cut algorithm created several formulas to find out the solution of BRP in Reggio Emilia, Italy based on the net flow of bike in docked stations and population density of each stations (Dell'Amico, Hadjicostantinou, Iori, & Novellani, 2014).

A few papers on dock systems discuss the issue of density of stations and of users. Liu, Jia, & Cheng (2012) explored the reason the Public Bicycle System (a docked system operated by city government of Beijing) suffered a decline after the 2008 Olympic Games. One of the main reasons is inefficient distribution of bicycle stations, with too many stations and low utilization in low-density areas and too few stations in high-density areas. García-Palomares et al., (2012) analyzed where dock stations should be located using GIS based on population density and CBD areas. For instance, there should be several docked bike-sharing stations around a subway station in the CBD area, while in residential areas there should be more stations in high-density areas and fewer stations in low-density areas. However, their method assumed that people usually consider CBD areas as the end point of their trips from suburban residential areas. In other words, it focuses on certain kinds of commuting and shopping trips from residential areas to CBDs, which ignores other kinds of bike-share trips and requires the analyst to specify CBD boundaries.

Although there are no stations, the efficiency of dockless systems also vary spatially. Newspapers and other media outlets have highlighted the chaotic parking problem in high density areas (CBS8 News, March 20th). In addition, low population density areas often experience low usage rate per bike. Users may have to walk a long distance to access a bike, affecting service quality. Bikes left in low-density areas may go unused for long periods of time. On the positive side, dockless bicycle users can conveniently bike to their exact destination in low-density areas, rather than park the bike at a station, which may be far from their final destination. As dockless systems start to spread to more sprawling cities and to the suburbs of cities, density will become an increasingly important factor. For these reasons, density should be a key concept for understanding dockless bike-share systems, but there appear to have been no studies published on the effect of density of such systems.

CHAPTER 2

PROBLEM STATEMENT

This research investigated the relationship between building density and efficiency of dockless BSS and in this way to discover the impact of building density on bike use, so as to optimize regional cycling volume.

The figure below is a conceptual model of this research and included all the hypothesis. The +/- sign on the arrow means positive/negative impact. The other (unmeasured) variables are not included in this study due to the lack of data.

The hypothesis of this study also included in the conceptual model:

1. Urban density can have a positive effect on BSS efficiency. As I mentioned before, low density area often has relatively low usage rate per bike while high density area has more daily activities, the efficiency in high density area should be higher.
2. Metro stations can be a positive variable for the efficiency of BSS. One of the objectives of BSS is to integrate with mass transit and to solve last mile problem. The more metro station and rail lines in an area means it should have more flow of people and activities. In other word, this area is more attractive for people using BSS.
3. Idle time and rebalancing are two indicators I use to reflect the BSS efficiency. Idle time is the time that the bike stays somewhere unused, in other word it can be treated as the waste time of BSS. Thus, the higher idle time means a lower efficiency of BSS system.
4. As for the rebalancing, this process is done by operators and government, at least partly to reduce the idle time in this area. That means, the more rebalancing process,

the lower idle time and therefore improve the efficiency of BSS in that area.

5. Also, I assumed the trip can have some impact on efficiency: Morning and evening peak can have positive effect, and midnight and daytime except rush hour can have a negative impact on the efficiency. In addition, the efficiency of BSS on weekends may be lower than weekdays.

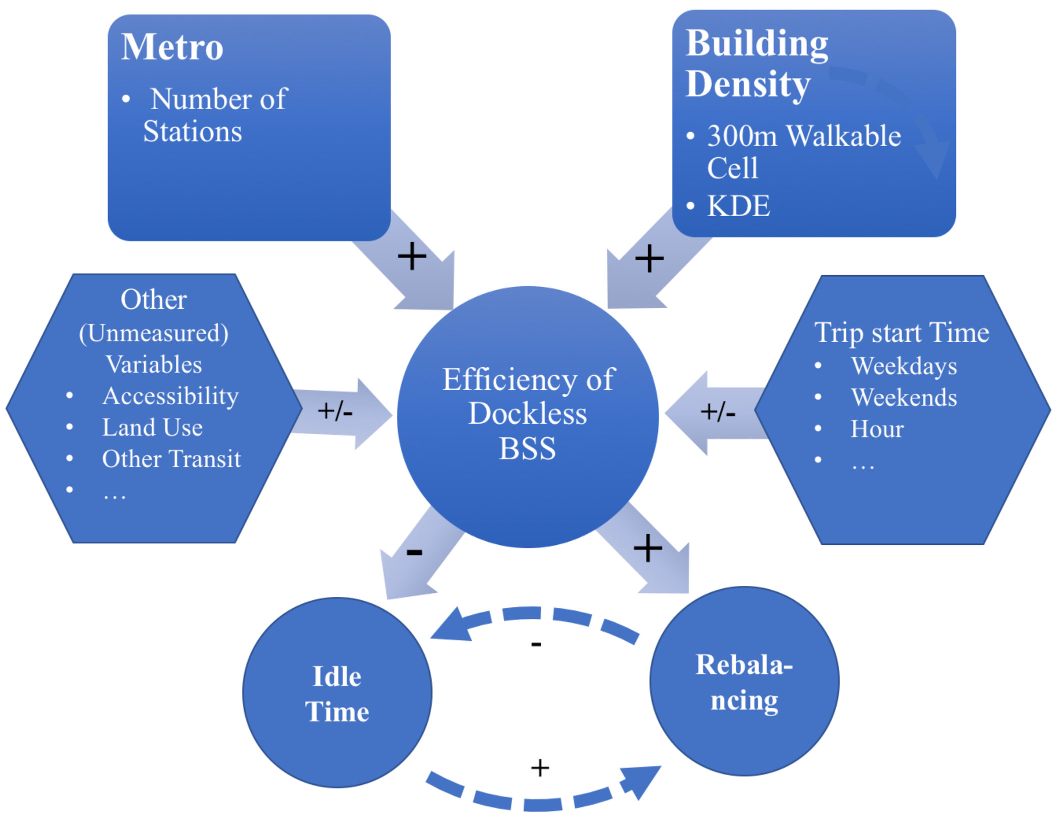

Figure 2. Conceptual model

CHAPTER 3

METHODOLOGY

Case Study: Mobike System in Beijing

Beijing, China's capital and second largest city, had a population of over 21.5 million people in 2015. This study analyzed the bike-sharing trips in Beijing's central area, within the Fifth Ring Road (Figure 3(B)). This area includes all of Xicheng, Dongcheng, Xuanwu, and Chongwen districts, and parts of Chaoyang, Haidian, and Fengtai districts. It covers approximately 775 square kilometers and accounts for 4.6% of the total area of Beijing (Figure 1(A)). The area was home to 10.5 million people in 2015, or 49% of Beijing's total population. This central area, with its very high population density, is a concentration of residences, employment, shopping, entertainment, culture, and political power.

In 2017, after dockless BSSs had been operating in Beijing for a year, the mode share of total bicycle trips had increased significantly, from 5.5% to 11.6%, and the share of car trips had decreased from 29.8% to 26.6% (Sharing Bicycles and Urban Development White Papers 2017).

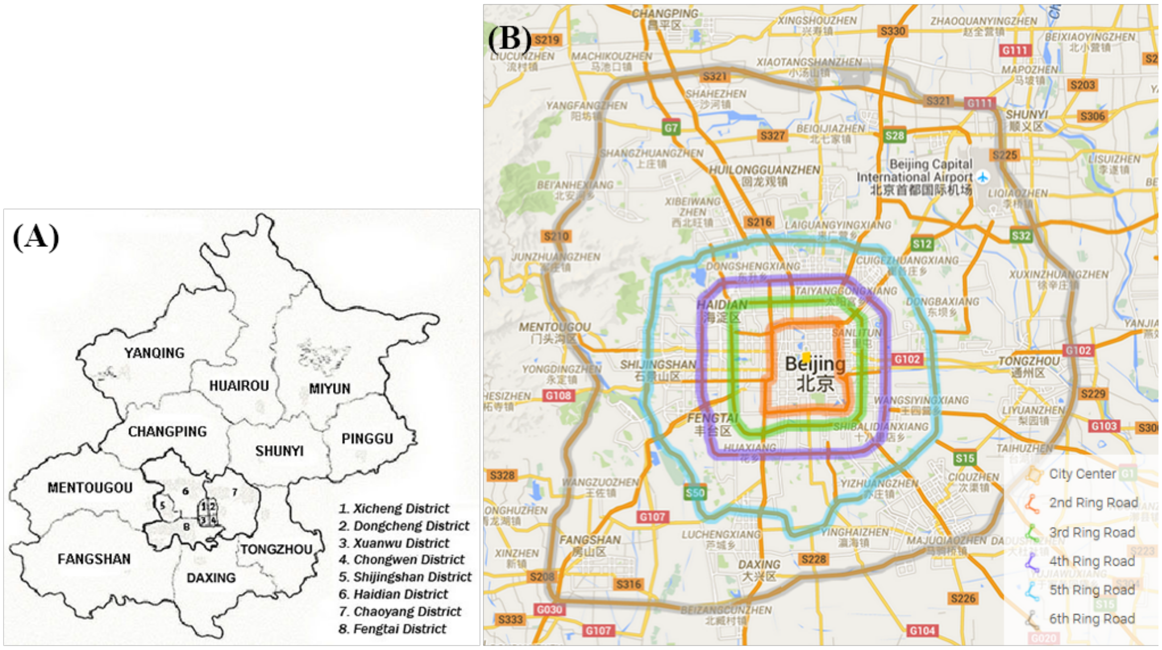

Figure 3(A). Administrate districts in Beijing

Figure 3(B). Ring roads in Beijing

Mobike is one of the biggest companies operating dockless bike-sharing systems in China, with a market share of about 50%. Mobike is the only company to install GPS trackers on their bicycles; the trackers provide useful, accurate trip data. Mobike began operating in Beijing on September 1, 2016, after launching its first operation in Shanghai in April of the same year. As of this writing, Mobike operates more than 700,000 bicycles in Beijing.

Data

To examine the relationship between urban density and the efficiency of dockless BSS in Beijing, I used data on building density, Mobike trips, and metro station locations.

For the building density data, I used a CAD building map showing the number of floors in each building in downtown Beijing. The map was purchased from AutoNavi

Software Co., Ltd. (amap.com), an online map company in China (Figure 4). The reason why I use building density rather than population density is that building density represents both population density and the high density commercial and office area (e.g., CBD). In contrast, population density can only reflect the density of residential area. As commuting trips plays an important role in BSS usage, residential density reflects only one end of the trip. But in addition, bike share can be used for shopping, restaurants, errands, school, entertainment, tourism, and other trip purposes, and may not have the riders home at either end of the trip. Therefore, the building density is more reasonable to applied in this study.

Figure 4. Building block CAD map for downtown Beijing

Data on bike-sharing trips in Beijing were provided by Mobike. This data included all 1,830,100 trips for one week, Wednesday May 10 to Tuesday May 16, 2017, made by all of that week's 349,600 Mobike users, on Mobike's 485,500 bicycles. The data included trip start times, starting and end points (latitude and longitude), user ID, bicycle ID, and bike type. These data were missing an important field: trip end time. Therefore, I used the average biking speed in Beijing, as provided by Mobike, of 8.5km/h to estimate end times for the trips, using Manhattan distance.

Some studies of BSS have used only one explanatory variable, (e.g., building density) and have lacked control variables (e.g., metro stations), which can result in omitted variable bias. To avoid this problem, I included metro station data and start day and time for each trip as control variables in my analysis. Metro stations are relevant because one of the main purposes of bike-sharing is to solve the "last mile problem" in public transit (Martens, 2004). Metro-station location data were obtained from a Baidu Map (comparable to Google maps), and were transformed to WGS-1984 Coordinate System.

Methods

This study investigated the relationship between building density and efficiency of dockless BSS, using a big-data approach. It was first necessary to define efficiency for dockless BSS. Based on literature and the characteristics of dockless BSS, idle time and percentage of rebalanced trips were selected as indicators of efficiency. The raw data on the 18.3 million Mobike trips were processed using Pandas in the Python environment. Then, both statistical and spatial analyses were performed to determine the correlation between building density and the efficiency variables. Analyses were carried out using SPSS and ArcGIS.

a. Data Processing

1. Density Data

I used two methods to present building density: Kernel Density Estimation (KDE) and 300m resolution grid cell.

Kernel Density Estimation (KDE) techniques in geospatial analysis may be applied to point datasets with spatially extensive attributes (De Smith, Goodchild, & Longley, 2007). The result of a KDE is usually a raster dataset (Longley et al., 2005) where each cell has a density value that is weighted according to distance from the starting features. Also, each point can have a "population" field to present its weight in the model. In this research, I used building volume (Floor Area * Number of Floor) as "population" field to weight each building point. The bandwidth is an important parameter for KDE model. The higher bandwidth can make the model global and vice versa. In this study, I generated three KDE models with different bandwidths: 0.005, 0.01 and 0.02 decimal degree (Around 500m, 1km and 2km respectively). After comparing the three models, I chose the model with 0.01 bandwidth (Figure 5) as it balanced the bias and variance better than others. For the kernel of KDE model, I use Gaussian kernel, chose 0.0018 decimal degree as cell size and use ArcMap to generate the KDE model.

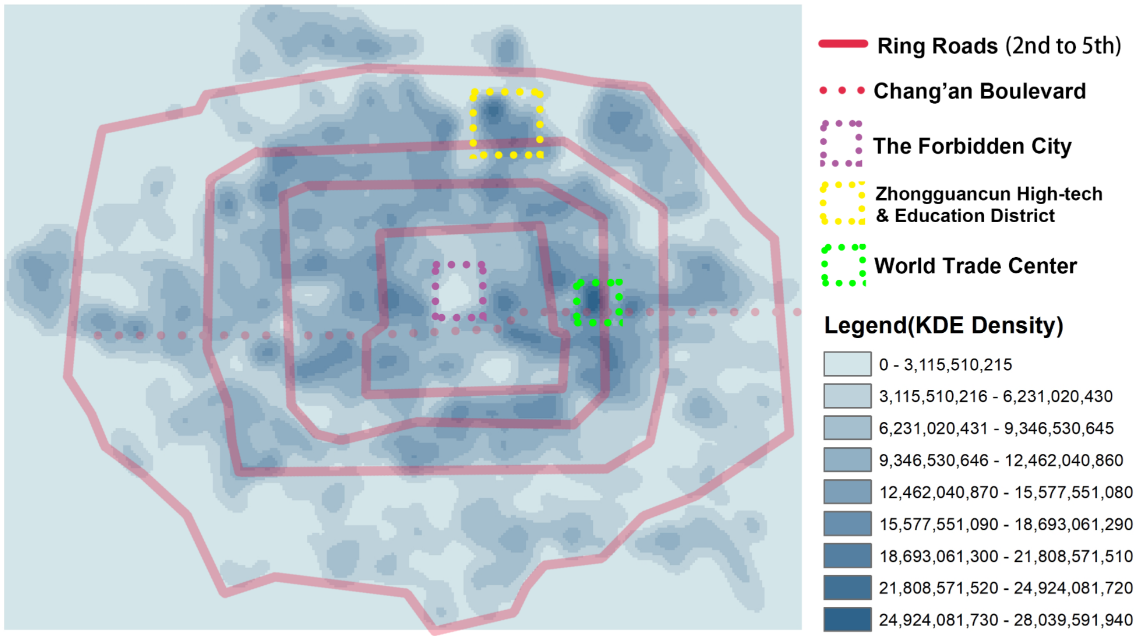

Figure 5. KDE density distribution of urban area in Beijing, with 0.01 decimal degree bandwidth

The KDE model can give each point a more precise density value without arbitrary boundaries, and I used this model to make a scatter map for each trip point and a histogram for net density change.

For the 300m grid cell method, I used Floor Area Ratio (FAR) to present building density (Figure 6). FAR can be expressed as:

FAR = Floor Area * Number of Floor/Area of the Plot.

With this method, I was able to calculate the average value of a spatial region and also reduce the computation time. Most spatial analyses in this research are based on a 300m grid-cell map.

The selection of a resolution for the grid cell is a trade-off process between bias and variance. If the resolution is too low (e.g., 1,000m), it will lead to a large bias in the local building density estimation, which produces an inaccurate result. If the resolution is too high (e.g., 50m), the number of data points for each cell used in estimation may become too low and create a large variance in the local building-density estimates. I chose 300m as the resolution level because it is a walkable distance. Moreover, the density of buildings in a person's immediate area can affect his or her feelings and behavior; for example, some people dislike, and prefer to avoid, crowded areas. The higher the building density, the more crowded an area is likely to be.

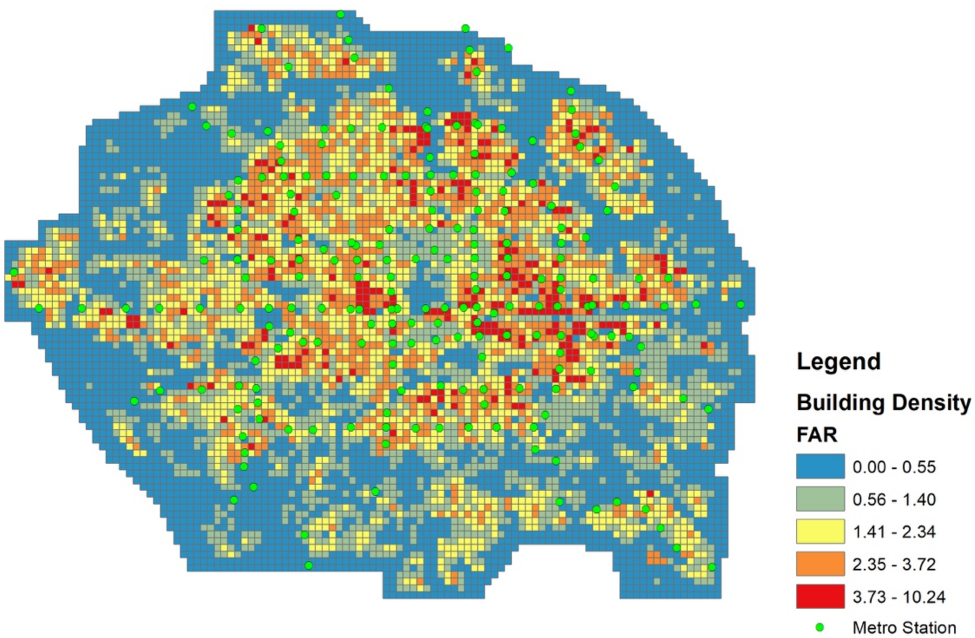

Figure 6. 300m resolution grid-cell floor ratio area (FAR) map of downtown Beijing
Made by author, Data Source: *AutoNavi Software Co., Ltd. (amap.com)*

2. Trip Data

I grouped the bike-trip data by bike id, and then calculated the Manhattan distance for each trip to estimate actual biking distance, since Beijing's street network is grid-like.

I then applied average bike speed in Beijing (8.5 km/h) to estimate trip end time for each trip. I calculated the time gap (idle time) between every two adjacent trips. Next, I determined whether the bike has been rebalanced; in other words, each trip was classified by its rebalanced status, in one of three classes: “Rebalanced”, “Non-rebalanced,” and “No Previous Record” (see Figure 8 for class parameters). I defined a trip with a Manhattan distance of less than 100 m between the previous ending point and the next starting point as non-rebalanced. Percentage of rebalanced trips was calculated based on total trips made on the same bicycle in a 300m grid cell.

Trips in the “non-rebalanced” class means these trips didn't experience a rebalance process after the end of the previous trip and before the start of next trip. In other words, the end location of this trip is (roughly) the same as the start point of next trip using the same bike. Therefore, the time gap between the end time of first trip and start time of next trip is exactly the idle time of this bike between these two trips at this location.

In contrast,” trips in the class “rebalanced” means the bike experienced a manual rebalancing process after the end of trip and before the start of next trip. This rebalancing process can be conducted by the BSS operator or government by truck (Figure 7). Under perfect conditions (urban activities are completely evenly distributed), the manual rebalancing is not necessary for BSS system, the bike can be rebalanced automatically be passengers themselves. However, in fact, there is a lot of unevenly distribution in the urban area, such as building density, land use type, and metro stations, that leads to the necessity of rebalancing. For rebalanced trips, it is impossible to calculate the amount of time the bike was idle at that location before being checked out by a customer, because the time gap between the end time of first trip and start time of next trip included the time the bike was idle at a different location plus the time required for rebalancing.

Figure 7. A photograph of rebalancing process
Source: www.theatlantic.com

The remaining class “no previous record” is for the first trip of each bike recorded in the dataset. For these trips, I cannot calculate the idle time because I cannot get the end time from the previous trip (they don't have a previous trip).

Overall, I can use the trip in the first class (“Non-Rebalanced”) to do analysis of idle time, and use second class (“Rebalanced”) to do analysis for rebalancing.

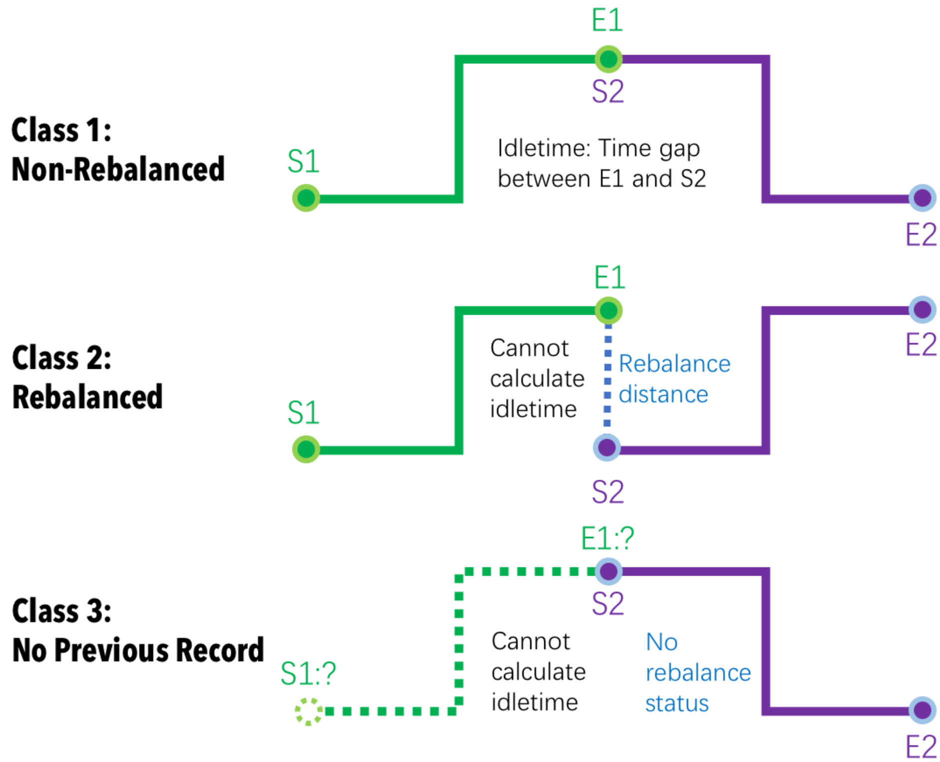

Figure 8. Three classes based on rebalance status

Note: S1= Start Point of Trip 1, E1= End Point of Trip 1.

For the idle time analysis, the density of each trip is based on start points. That's where the bike sat idle before being used. For rebalancing analysis, the density at both start points and end points are potentially of interest, but the density of a location that it is rebalanced *from* is what triggers the need to move the bike. In most cases, when a BSS company or government rebalances a dockless bike, it's more likely to be because it was left in an undesirable location—either "in the way" or where there is an oversupply, rather than because the company needs to find bikes to move to undersupplied areas—although the latter can also be an issue. For this analysis, I use individual trip data and

KDE model to reflect building density. In this way, the analysis can be more accurate with no aggregation.

The percentage of rebalancing should be another indicator for the efficiency of BSS, both docked and dockless. The higher percentage shows higher efficiency, because in general, rebalancing brings bikes from area with low demand to high demand, this allows bicycles to be fully utilized. The percentage of rebalancing can be expressed as:

Percentage = (Number of rebalanced trip / Total trips in each grid cell) * 100%

I use 300m grid cell for the rebalancing analysis because each cell can have multiple trips, and I can do the calculation for percentage of rebalancing for each grid cell. It's easier to quantify the rebalancing than using individual trips with KDE map.

In addition, the time when the trip is recorded is also included in this study. I use 23 dummy variables ("hour1" to "hour23") to reflect the hour and another 6 dummy variables ("Mon" to "Sun") to reflect the day of the trip recorded. "Hour1" means the trip start time falls from 1:00 to 2:00 (in 24-hour), "Mon" means the trip start time recorded in Monday, respectively. For instance, if the trip start time is 3:20am at Monday, the variable "hour 3" and "Monday" of this trip are "1", and other day and time variables are "0". The base case for hours is midnight to 1:00 am; no dummy variable is included for trips starting at this time. Similarly, the base case for day of week is Monday, there is no dummy variable for Wednesday.

For metro station data, which is based on 300m grid cell. for the cell has a metro station in it got a value 1, for those station which have more than one metro line (transfer station) got the value equals to the number of lines it has. I assume the passenger flow at the two-line transfer station is twice that of the station with only one line. The highest value for metro station field is 3.

There were 7,282 300m walkable cells with building density in downtown Beijing. However, bike trips were not distributed evenly in these cells: the densest cell had 2683 trip records while some cells had none. Thus, I needed to cleanse the data before analysis to avoid inaccuracies caused by too few data samples. I used a minimum threshold of 10 trips per cell, and removed all cells with fewer than 10 trips.

b. Statistical and Spatial Analyses

To map the indicators of BSS efficiency, average idle time and percentage of rebalanced trips were both calculated for each 300m grid cell. A higher idle time indicates lower efficiency, and a higher percentage of rebalanced trips indicates lower efficiency because rebalances are assumed to be made from a low-demand area to higher-demand area. I mapped each of the two indicators for each grid cell to put the results of data analysis into a spatial perspective. I also created scatter maps of the relationship between building density and average idle time, and building density and percentage of rebalanced trips.

Next, I did regression analyses, with building density as the independent variable, and idle time and percentage of rebalanced trips as the dependent variables. In addition, metro-station location was used as another independent variable in a regression analysis to understand the effect of metro stations on BSS. Trip start time, which are reflected by day and hour variables, is also included in this study to find out the time effect on BSS usage. The regressions were conducted with OLS (Ordinary Least Squares) to investigate the relationships among variables.

$$Y_i = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{ik} + \varepsilon_i \, , i = 1, \ldots , n$$

where:

Y_i = dependent variables: percentage of rebalanced trips or idle time

X_{ik} = independent variables: building density, metro line count, hour and day dummy variables

α, ε = constant

CHAPTER 4

RESULTS

Descriptive Findings

Table 1 shows some features of riding behavior in the Mobike system in Beijing. To get an idea of how trip behavior in the Mobike system compares to that in a docked system, I looked at data from a docked system in another nation's capital, Capital Bike, in Washington, D.C. Trip distance and trip duration for the Mobike system were relatively lower than those for the Capital Bike, where the mean trip duration is 900 seconds (15 minutes), and mean trip distance is 2.9km. The main purpose of the Mobike system in Beijing is to cover short-distance trips to make connections with public transit.

Table 1. Descriptive Statistics for Mobike Trip Dataset

Data	count	Min	25%	Medium	75%	Max	Mean	Std. Dev.
Trip Distance (km)	1,830,100	0.15	0.76	1.07	1.53	44.74	1.23	0.95
Trip Duration (Second)	1,830,100	65	324	453	647	18900	522	403
Idle time (Second)	183,484	-6,417*	987	5,076	21309	556,336	17,034	30,825

Rebalance Distance (km)	1,210,673	0.15	1.22	2.9	5.96	8.75	4.38	4.56

* For negative values in "idle time" field, because trip distance is estimated using Manhattan distance and the duration time is estimated using average speed in Beijing (8.5km/h), the estimated idle time may deviate from actual time.

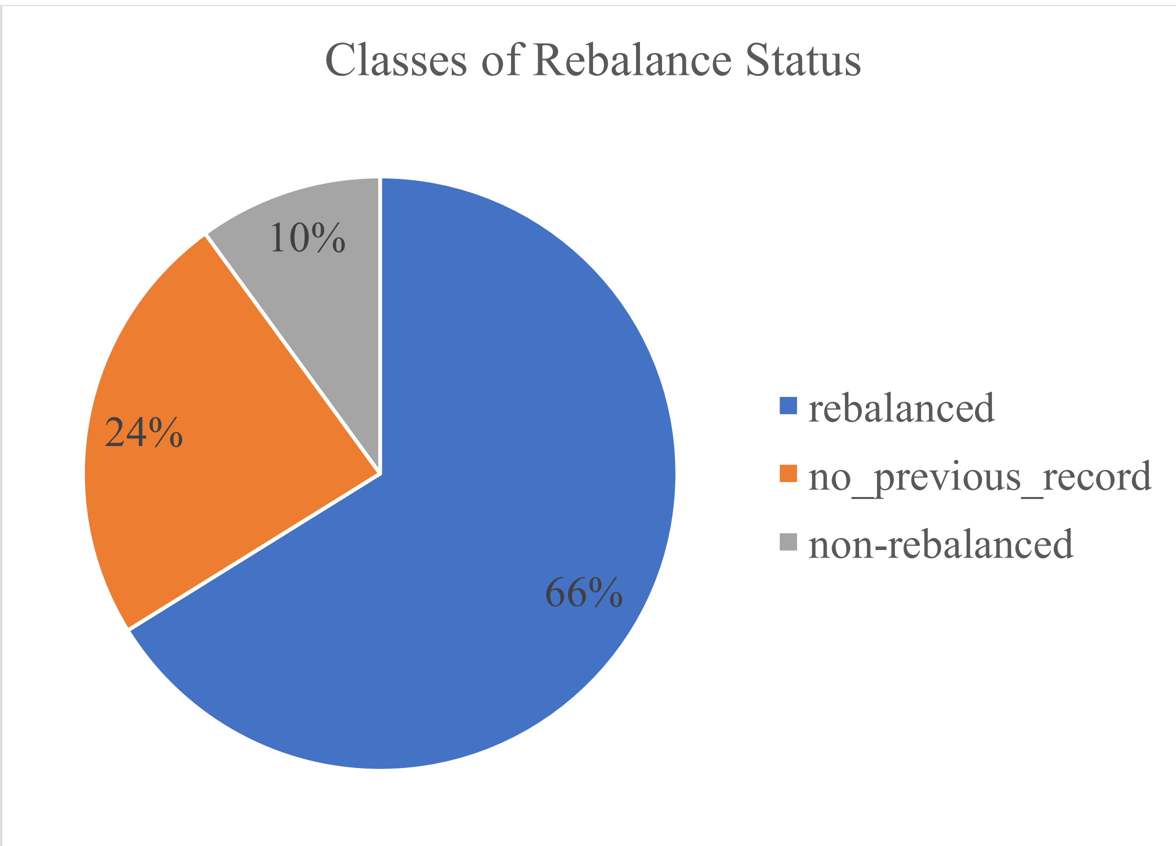

Figure 9. Distribution of each class of rebalance status (n=1,830,100 trips)

The distribution of each class of rebalance status is shown in Figure 9. For total 1,830,100 trips, 1,210,673 trips are rebalanced (66%), 435,943 trips have no previous record (24%) and 183,484 trips are non-rebalanced (10%). This result is somehow unexpected. The rebalanced trips account for a much larger proportion than I expected before. That means over half of trips had experienced a manual rebalance from the destination of last trip (e.g. moved by Mobike company from an area with low demand to another area with high demand). To further study the rebalanced trip, I calculated the net change of density (Δ Density) and made a histogram to present the result (Figure 10):

Δ Density = Density $_{\text{rebalance end}}$ – Density $_{\text{rebalance start}}$

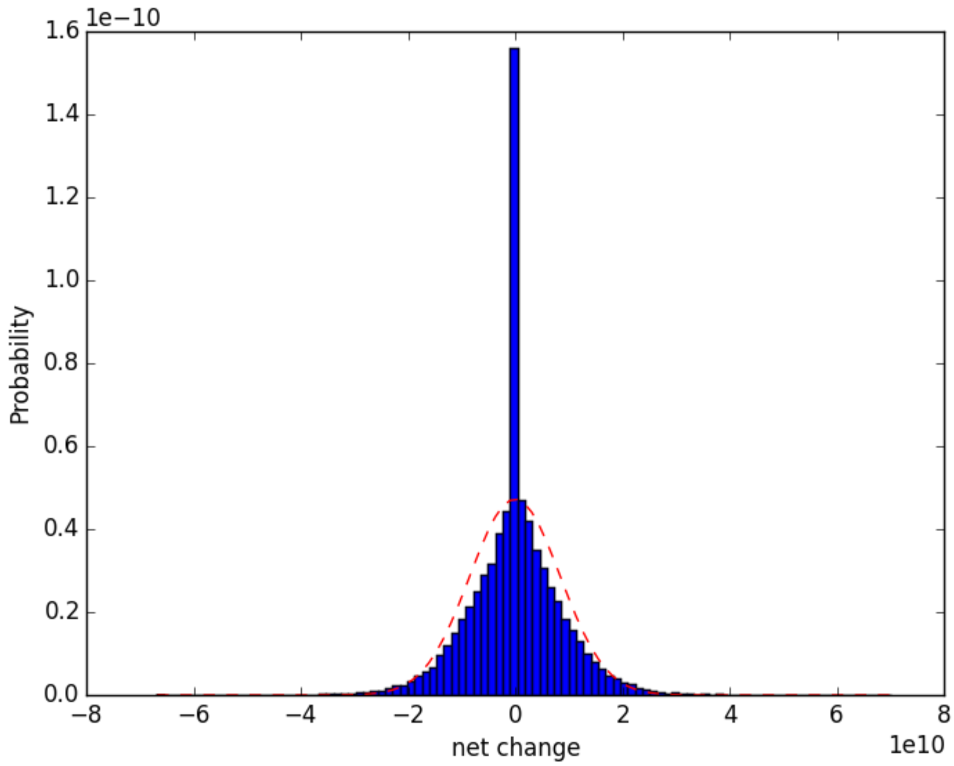

Figure 10. Histogram for the net change of density for rebalanced trips (n=1,210,673 rebalanced trips)

The result shows amazing symmetry and a normal distribution, which means there is no obvious direction of density change for rebalance process. Another finding from the histogram is a large part of trips has a relatively small net change before and after rebalancing. That means the building density is not a significant factor during rebalance process. I sent an enquiry email for more detailed information about rebalance process to Mobike company but did not get any response yet.

Relationship Between Density and Efficiency

a. Relationship Between Building Density and Idle Time

Using the 300m walkable cell to generate a map for spatial distribution of average idle time in downtown Beijing (Figure 11) The map shows a great spatial heterogeneity, the average idle time in south region is obvious higher than north region. Compare with the density map, the spatial distribution of average idle time is in general the opposite of density map: in the center city area, which has a high building density, bikes have a lower average idle time, and vice versa. This map result visual evidence for my hypothesis: high density area shows a higher efficiency of dockless BSS.

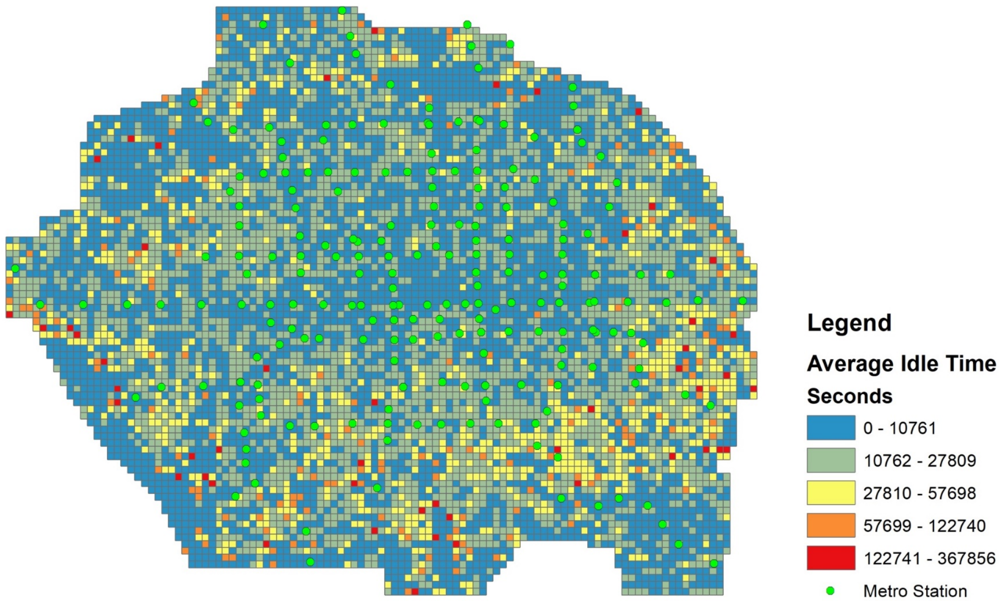

Figure 11. Map of average idle time distribution

To further test my hypothesis and to find more specific relationship between building density and idle time, I did some statistical analysis: made a scatter plot between building density (Figure 9) and applied an OLS model.

An OLS model has been applied to observe the detailed relationship between average idle time (dependent variable) and building density, metro station and start day and time (explanatory variables.)

Table 2. Ordinary Least Squares Regression Model for Idle Time (n = 183,484 trips)

Independent Variable	Coefficients	t-value	VIF
(Constant)	19483.497	14.327	

Metro Station	-1116.487***	-3.602	1.009
Building Density	-7.84E-07***	-50.294	1.013
hour1	-314.846	-0.134	1.481
hour2	4983.558*	1.803	1.308
hour3	9024.481***	2.941	1.236
hour4	18974.668***	7.53	1.394
hour5	24628.007***	15.526	3.495
hour6	22267.77***	15.839	10.994
hour7	16186.336***	11.852	27.276
hour8	8131.439***	5.96	28.595
hour9	4555.157***	3.3	17.556
hour10	5366.81***	3.867	14.824
hour11	3558.237**	2.59	20.736
hour12	361.056	0.264	24.28
hour13	-190.342	-0.138	19.989
hour14	1210.739	0.877	17.222
hour15	1207.533	0.875	17.861
hour16	1774.573	1.293	21.865
hour17	2709.397**	1.987	30.074
hour18	2768.249**	2.031	30.71
hour19	1263.714	0.923	23.674
hour20	850.373	0.617	18.314

hour21	569.385	0.41	15.019
hour22	990.08	0.696	8.931
hour23	1531.823	0.987	3.945
Wed	-8862.494***	-33.108	1.574
Thu	-2803.508***	-11.09	1.671
Fri	-2112.358***	-8.28	1.653
Sat	2372.049***	8.768	1.558
Sun	2239.971***	8.444	1.595
Tue	-658.446***	-2.657	1.711

Note: Adjusted R^2 = 0.074, VIF = variance inflation factor, DF (degrees of freedom) = 31. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. “hour1” means the trip start time during 1:00 to 2:00 (in 24-hour), “Mon” means the trip start time recorded in Monday, respectively.

The OLS result has a relatively low R^2 0.074 but the p value (Sig) of most explanatory variables are very low (0.000), which means that most explanatory variables show high significance. However, the lower R^2 shows that there may be some other factors that can also impact idle time but were not included in this regression.

The parameter estimates show the idle time has a negative correlation with both metro station and building density. Also, for time analysis, idle time can be obviously reduced in morning peak (from 6:00 to 10:00) and evening peak (17:00 to 19:00) is also showing significance (Figure 12). The model uses the midnight hour (trip during 0:00-1:00) as the base time, and the result shows the idle time has an obvious increasing during the time before morning peak (2:00 to 6:00). That is because the time period from midnight to morning peak has fewer urban activities compared to other times of day. In this way, the idle time can experience a significant accumulation and keep increasing until the morning peak. In addition, the time variable is based on trip start time, which means the idle time is measured from the previous trip. For example, a bike finished the first trip and was dropped off at midnight until the next trip of this bike that started in 6 am, this six-hour-long idle time is applied to the second trip which is recorded in 6am (morning peak). This can explain why during the first two hours in morning peak (6 and 7) the coefficient is still very high—because early in the morning, most bikes have sat idle overnight, or longer.

For weekday analysis, day variables show high significance level, and weekends can have a significant positive impact on idle time (Figure 13). That means, the usage rate of BSS is lower in weekends than those in weekdays. Another finding is Wednesday has a very low (strongly negative relative to Monday) coefficient compare to other weekdays. This result is due to the dataset, which recorded the trip during the whole week from May 10th, 2017 to May 16th, 2017. May 10th is Wednesday, so that many trips on this day have no idle time ("no previous record"), and I can only calculate idle time for those trips that had a previous trip recorded on the same day. In other word, the longest idle time in Wednesday's data is entire day (24 hours), they don't have any idle time recorded that longer than one day otherwise it will be recorded to the next day. For example, if the first trip of a bike in this dataset is recorded ending at 1am on Wednesday, the idle time of this trip cannot be calculated, because there is no previous trip record for this bike. Then, the next trip started at 11pm (23:00), that means, the idle time of this bike on Wednesday is

22 hours. If the next trip is started at 1am the next day (Thursday), the idle time should be 24 hours, but this idle time will be recorded on Thursday not Wednesday.

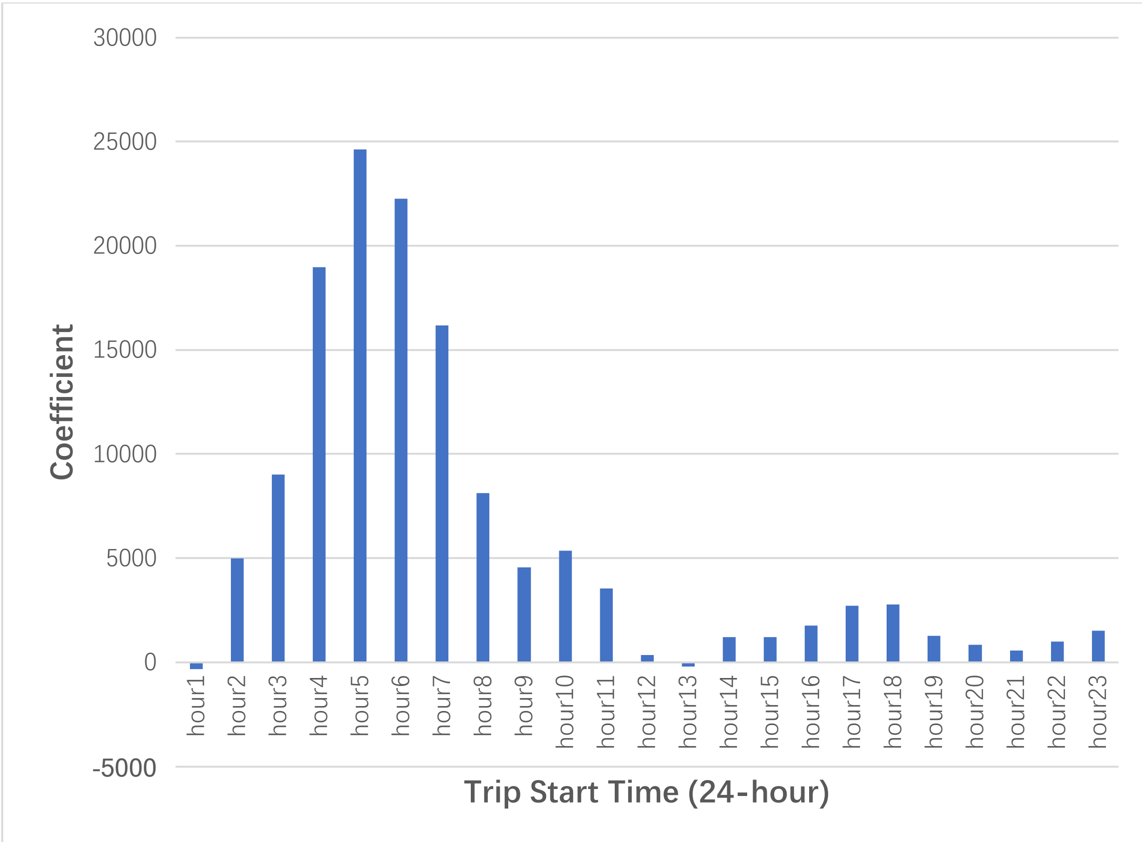

Figure 12. Coefficient of idle time distribution for 24 hours (based on hour 0)

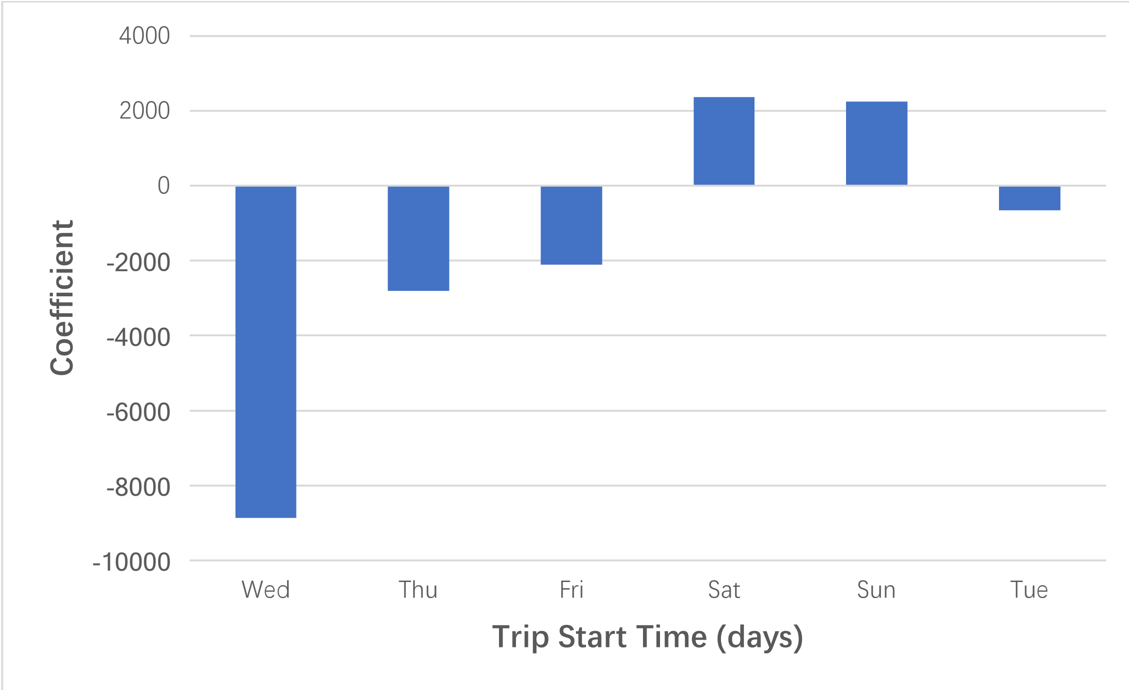

Figure 13. Coefficient of idle time distribution for seven days (based on Monday)

b. Relationship Between Building Density and Percentage of Rebalancing

Using 300m walkable grid cell to generate a map for spatial distribution of percentage of rebalancing in downtown Beijing (Figure 14). The average percentage for 5715 grid cells is 65.1%. Compared with the density map, the spatial distribution of percentage of rebalancing is approximately following the density map, presenting a descending distribution from the center. This result shows a positive correlation between percentage of rebalancing and building density.

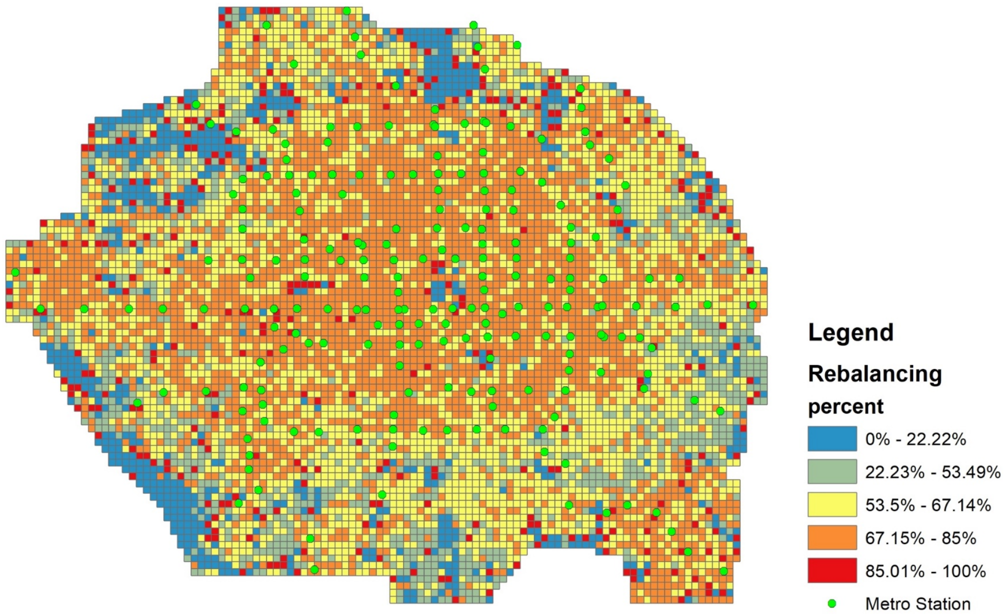

Figure 14. Map of spatial distribution of percentage of rebalancing

An OLS model has been applied to observe the detailed relationship between percentage of Rebalancing (dependent variable) and building density, metro station (explanatory variables.) After eliminated the cells count less than 10 trip data, 5,715 cells remain. This time, the R^2 improves a bit yet is still low: 0.069. However, similarly to the idle time regression, the explanatory variables are highly significant. The parameter estimates show rebalancing has a positive correlation with both metro station and building density. Also, low VIF (1.007) means there is little multicollinearity between variables. (Table 3)

Table 3. Ordinary Least Squares Regression Model for Percentage of Rebalancing (n = 5,715 300m grid cells with > 10 trips per cell)

Independent Variable	Coefficients (t-value)	VIF	Adjust R^2	DF
Constant	0.623*** (335.723)	-	0.069	2
Building Density (FAR)	0.020*** (19.809)	1.007		
Number of Metro Station	0.022*** (4.398)			

Note: VIF = variance inflation factor, DF = degrees of freedom. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

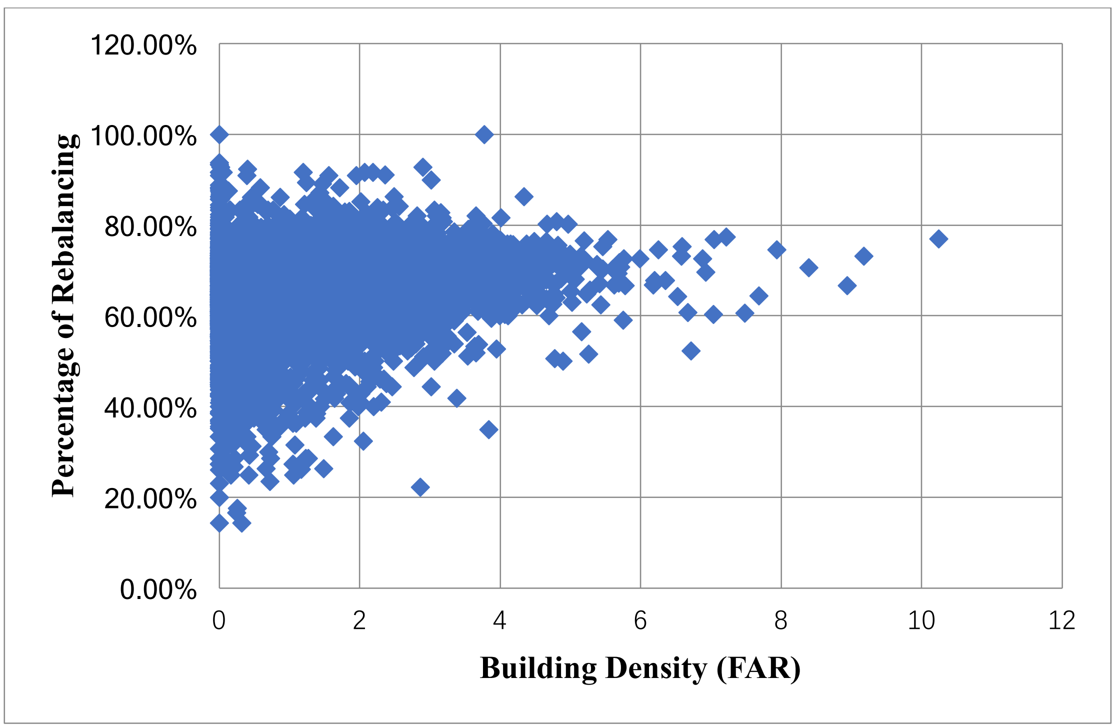

Figure 15. Scatter plot for building density and percentage of rebalancing (n = 5,715 300m grid cells with > 10 trips per cell)

The scatter plot between building density and percentage of rebalancing is shown in Figure 15. The scatter plot uses data in 5715 grid cells. A positive correlation can be found in the scatter plot, but there is also strong evidence of heteroscedasticity.

As of now, the results are in line with my assumptions, the density has a positive impact on the efficiency of dockless BSS. However, the R^2 are unexpectedly low.

Relationship Between Idle Time and Percentage of Rebalancing

Finally, I tried to find the connection between the idle time and percentage of rebalancing: applied a bivariate OLS model using average idle time as dependent variable and percentage of rebalancing as explanatory variable (Table 4.) I assuming that average idle time can experience a decrease during the increase of rebalancing. The results proved my hypothesis, the parameter estimates shows there is a significant negative correlation between idle time and percentage of rebalancing. However, the R^2 is still low but better than before (0.137).

Table 4. Bivariate OLS model for Percentage of Rebalancing and Average Idle Time (n = 5,715 300m grid cells with > 10 trips per cell)

Variable	Coefficients (t-value)	VIF	Adjust R^2	DF
Constant	62358.851*** (41.532)	-	0.137	1
Percentage of Rebalancing	-68880.440*** (30.195)	-		

Note: VIF = variance inflation factor, DF = degrees of freedom. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

CHAPTER 5

DISCUSSION AND LIMITATIONS

The relationship between efficiency and building density has been demonstrated from both statistical and spatial analysis: building density can have a positive effect on efficiency of dockless BSS. However, the low R^2 means there should be many other factors can impact efficiency of dockless BSS system beyond building density, metro station and trip start time.

Urban development differences can be an important factor. Compare distribution map of idle time and density can find a common phenomenon: North and South Beijing are very different. Actually, the urban development of Beijing does present a huge difference between North and South. All along, universities, financial and high-tech companies have been gathering in the north. This invigorated the northern part of the city. Zhongguancun, located between the North Fourth Ring and the North Fourth Ring Road, is the most concentrated area of intelligence in China. It has gathered many headquarters of high-tech companies and is known as China's Silicon Valley. The neighboring Wudaokou is China's financial center, which has gathered the headquarters of major banks in China. In contrast, most of the southern area is old city. It has remained the status quo for a long time and has developed slowly. Most areas are urban and rural areas and urban villages, with low-income residential area. Beijing South Railway Station is a large-scale railway hub. Its existence has also greatly limited the development of the southern urban area of Beijing. This shows that urban development level and land use can both have a great impact on the efficiency of dockless BSS. However, this study did not include these variables into the analysis. This is most likely the main reason for the lower R^2, which is also a limitation of this research.

As for the relationship between idle time and rebalancing, the result means a negative correlation between them. This result explains why there is a high portion of rebalanced bikes. From the operator's point of view, one of the Mobike's purposes of rebalancing the bicycle is to reduce idle time. If idle time is high, the company would be expected to rebalance those bikes to get them back into circulation. On the contrary, if rebalancing is high, it can reduce the idle time observed for the remaining bikes. There is circular reasoning involved and the direction of causation goes both directions.

CHAPTER 6

CONCLUSIONS AND FUTURE WORKS

This research demonstrated the relationship between efficiency of dockless BSS and building density using a big data approach. Through this study, I found that the performance of dockless BSS in low-density areas is not ideal, most of them experienced a high idle time. Governments and operators need to develop special plans and policies for using dockless BSS in low-density areas, and operators also need to increase the frequency of rebalancing in low-density areas.

Some specific suggestions based on the density:

1. For high density area: provide more parking space for dockless bikes in the area with high demand (e.g., commercial center, metro station with high passenger). More parking space can help solve the chaotic parking problem, and reduce the idle time as well as the rebalancing percentage, in this way, reduce the cost of operation BSS.
2. For low density area: It may help to create a different rate plan in low density area. The new plan should charge people more in low density area because under the same rebalance percentage, the rebalancing process cost more in low density area than high density area per bike. However, it is complicated because higher pricing will also discourage use of BSSs in low density area, and will increase the idle time.
3. Launch incentives (e.g. reward program) to encourage people to park their bikes in designated locations or in clusters with other bikes. This can improve the idle time in low density area (there will be no bike scattered in remote areas, unused

for long periods of time).

Using 300m walkable cells to do the spatial analysis provide a foundation for future research. Also using FAR as the index of building density make research results easier for the government to formulate policy services.

Understanding the relationship between density and efficiency of dockless BSS can significantly improve the operating efficiency of BSSs, and increase regional cycling volume. The finding on the negative relationship between idle time and rebalancing also contribute valuable information to solve the BRP in dockless system in the future. In addition, finding the nexus between the system performance and building density can also help this new type system promote to other countries.

For future work, I think one valuable research is to include land use in the analysis as mentioned above. More Points of Interest (POI, e.g., restaurant, hotel and bus station) other than metro station can be used to quantify land use data. Some POI with very low building density may have a high attraction for BSS usage (e.g. tourist sites, parks, zoos). Also, the traffic congestion and distribution of bicycle facilities (bike lane and parking space) can be considered in this research.

REFERENCES

Brons, M., Givoni, M., & Rietveld, P. (2009). Access to railway stations and its potential in increasing rail use. *Transportation Research Part A: Policy and Practice, 43*(2), 136-149. doi:10.1016/j.tra.2008.08.002

Campbell, K. B., & Brakewood, C. (2017). Sharing riders: How bikesharing impacts bus ridership in New York City. *Transportation Research Part A: Policy and Practice, 100*, 264-282. doi:10.1016/j.tra.2017.04.017

De Smith, M. J., Goodchild, M. F., & Longley, P. (2007). *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*: Troubador Publishing Ltd.

Dell'Amico, M., Hadjicostantinou, E., Iori, M., & Novellani, S. (2014). The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega, 45*, 7-19. doi:10.1016/j.omega.2013.12.001

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

Erdoğan, G., Battarra, M., & Wolfler Calvo, R. (2015). An exact algorithm for the static rebalancing problem arising in bicycle sharing systems. *European Journal of Operational Research, 245*(3), 667-679. doi:10.1016/j.ejor.2015.03.043

Erdoğan, G., Laporte, G., & Wolfler Calvo, R. (2014). The static bicycle relocation problem with demand intervals. *European Journal of Operational Research, 238*(2), 451-457. doi:10.1016/j.ejor.2014.04.013

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography, 41*, 306-314. doi:10.1016/j.jtrangeo.2014.01.013

Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: a synthesis of the literature. *Transport reviews, 33*(2), 148-165.

Fricker, C., & Gast, N. (2016). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *Euro journal on transportation and logistics, 5*(3), 261-291.

García-Palomares, J. C., Gutiérrez, J., & Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography, 35*(1-2), 235-246. doi:10.1016/j.apgeog.2012.07.002

Jäppinen, S., Toivonen, T., & Salonen, M. (2013). Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography, 43*, 13-24. doi:10.1016/j.apgeog.2013.05.010

Liu, Z., Jia, X., & Cheng, W. (2012). Solving the Last Mile Problem: Ensure the Success of Public Bicycle System in Beijing. *Procedia - Social and Behavioral Sciences, 43*, 73-78. doi:10.1016/j.sbspro.2012.04.079

Ma, B. T., Liu, C., & Erdoğan, S. (2015). *Bicycle Sharing and Transit: Does Capital Bike-share Affect Metrorail Ridership in Washington, DC?* Paper presented at the 83rd Annual Meeting of the Transportation Research Board.

Martens, K. (2004). The bicycle as a feedering mode: experiences from three European countries. *Transportation Research Part D: Transport and Environment, 9*(4), 281-294.

Martin, E. W., & Shaheen, S. A. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. *Journal of Transport Geography, 41*, 315-324. doi:10.1016/j.jtrangeo.2014.06.026

Pan, H., Shen, Q., & Xue, S. (2010). Intermodal Transfer Between Bicycles and Rail Transit in Shanghai, China. *Transportation Research Record: Journal of the Transportation Research Board, 2144*, 181-188. doi:10.3141/2144-20

Pucher, J., & Buehler, R. (2012). *City cycling*: MIT Press.

Randriamanamihaga, A. N., Côme, E., Oukhellou, L., & Govaert, G. (2014). Clustering the Vélib′ dynamic Origin/Destination flows using a family of Poisson mixture models. *Neurocomputing, 141*, 124-138. doi:10.1016/j.neucom.2014.01.050

Singleton, P., & Clifton, K. (2014). Exploring synergy in bicycle and transit use: Empirical evidence at two scales. *Transportation Research Record: Journal of the Transportation Research Board*(2417), 92-102.

Zhang, L., Zhang, J., Duan, Z.-y., & Bryde, D. (2015). Sustainable bike-sharing systems: characteristics and commonalities across cases in urban China. *Journal of Cleaner Production, 97*, 124-133.

Zhu, W., Pang, Y., Wang, D., & Timmermans, H. (2013). *Travel behavior change after the introduction of public bicycle systems: Case study in Minhang District, Shanghai.* Paper presented at the Transportation Research Board 92nd Annual Meeting.