Improving immersive, highly realistic in-lab, cycling experiences for analyzing active travel

Center for Transportation, Environment, and Community Health
Final Report

by

Ricardo A Daziano, So-Yeon Yoon, Cacho Wang

September 30, 2021

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Improving immersive, highly realistic in-lab, cycling experiences for analyzing active travel		5. Report Date September 30, 2021 6. Performing Organization Code
7. Author(s) Ricardo A Daziano (ORCID ID # 0000-0002-5613-429X)		8. Performing Organization Report No.
9. Performing Organization Name and Address School of Civil and Environmental Engineering Cornell University Ithaca, NY 14853		10. Work Unit No. 11. Contract or Grant No. 69A3551747119
12. Sponsoring Agency Name and Address U.S. Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report 07/01/2020-09/30/2021 14. Sponsoring Agency Code US-DOT
15. Supplementary Notes		

16. Abstract

In recent years, the continuous growth of private cars, the tight supply of land resources and the continuous poor air conditions have led policy makers to advocate sustainable public transportation. Bike sharing system has been introduced by many cities and developed worldwide rapidly, due to its advantages in reducing environmental pollution and alleviating traffic congestion (Fishman, Washington et al. 2015). It is recognized as a strategic tool to integrate public transportation and promote sustainable urban transportation (Martin and Shaheen 2014). Cities around the world seek to reshape urban transportation to a greener and healthier way with the help of bike sharing system.
Since bike sharing systems were put into use, the service has gone through several generations. From the initial free bike system (DeMaio 2009) in the 1960s, to the coin-deposit system (Shaheen, Guzman et al. 2010), and then to the information technology-based system (Wang, Zhang et al. 2010), it has now developed into the multimodal system (Mátrai and Tóth 2016). However, with the development of˝ bike sharing system, many challenges related to its operation have not been well solved, including modeling the demand and encouraging users to purchase memberships.

This report focuses on modeling the demand of bike sharing system and exploring the factors that affect users' subscription membership, in the context of the Citi Bike initiative in New York City. It is organized as follows. Chapter 1 is the literature review of related topics. Chapter 2 describes the data as well as the context of the survey. Chapter 3 develops different discrete choice models and discusses estimates of these models, with focuses on both attributes of a bike sharing system pass and socio-demographics of users. Conclusions found through these models are also discussed in this part.

17. Key Words	18. Distribution Statement
Discrete choice model, cycling, subscription, virtual reality	Public Access

19. Security Classif (of this report)	20. Security Classif. (of this page)	21. No of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-69)

Improving immersive, highly realistic in-lab, cycling experiences for analyzing active travel

September 2021

1 Introduction

1.1 Introduction

In recent years, the continuous growth of private cars, the tight supply of land resources and the continuous poor air conditions have led policy makers to advocate sustainable public transportation. Bike sharing system has been introduced by many cities and developed worldwide rapidly, due to its advantages in reducing environmental pollution and alleviating traffic congestion (Fishman, Washington et al. 2015). It is recognized as a strategic tool to integrate public transportation and promote sustainable urban transportation (Martin and Shaheen 2014). Cities around the world seek to reshape urban transportation to a greener and healthier way with the help of bike sharing system.

Since bike sharing systems were put into use, the service has gone through several generations. From the initial free bike system (DeMaio 2009) in the 1960s, to the coin-deposit system (Shaheen, Guzman et al. 2010), and then to the information technology-based system (Wang, Zhang et al. 2010), it has now developed into the multimodal system (Mátrai and Tóth 2016). However, with the development of′ bike sharing system, many challenges related to its operation have not been well solved, including modeling the demand and encouraging users to purchase memberships.

This report focuses on modeling the demand of bike sharing system and exploring the factors that affect users' subscription membership, in the context of the Citi Bike initiative in New York City. It is organized as follows. The subsequent part of Chapter 1 is the literature review of related topics. Chapter 2 describes the data as well as the context of the survey. Chapter 3 develops different discrete choice models and discusses estimates of these models, with focuses on both attributes of a bike sharing system pass and socio-demographics of users. Conclusions found through these models are also discussed in this part.

1.2 Literature Review

In terms of forecasting the demand of bike sharing system, the demand might be affected by various factors. Rixey (2013) examines effects of travelers' demographic characteristics, built environment characteristics near bike sharing system and bike sharing system network characteristic. Statistically significant relationships between demographic characteristics and bike sharing ridership are established. However, Campbell, Cherry et al. (2016) point out through a survey of bike sharing system in Beijing, that detailed factors such as distance, temperature, precipitation, air quality, etc. are major factors affecting the short-term demand under relatively fixed economic conditions, while travelers' demographic characteristics (including income, gender, occupation, etc.) have no significant effect. Both these two studies do not pay enough attention to spatio-temporal factors. However, ignoring the presence of such effects (for example, if a station is full, there will not be any arrivals until someone rent a bicycle from the station) might result in bias when estimating models. Thus, Faghih-Imani and Eluru (2016) estimate comprehensive econometric models to incorporate for the influence of spatiotemporal interactions. Spatio-temporal factors considered in the research include time-of-day and day-of-the-week on bike sharing system usage, population and employment density, the length of bicycle routes and streets, the presence of subway, the number of capacity of bicycle stations, restaurants and area of parks within 250m vicinity of bicycle stations. Vogel, Greiser et al. (2011) apply cluster analysis to the ride data when trying to alleviate imbalance in the distribution. During the process, they also reveal spatio-temporal dependencies of pickup and return activity pattern at a docking station. In conclusion, current studies have examined bike sharing system from different dimensions, mainly including travelers' socio-demographics characteristics, bike sharing system and network infrastructure, and spatio-temporal interactions.

From the perspective of modeling methods, methods from traditional models such as linear OLS models to the latest method of machine learning are applied in different papers. Faghih-Imani and Eluru (2016) adjust a pooled linear regression model for panel data considering spatial specific effects by replacing the random error with spatial specific random error, and prove that temporally or spatially lagged variables significantly improve the fitness of the model. Tang, Pan et al. (2017) implement binary logistic model in their study to analyze what factors would help to increase the probability of travels' using bike sharing system. Campbell, Cherry et al. (2016) build a multinomial logit mode switching model to estimate the likelihood of respondents switching from original choices to using bike sharing system. Kaspi, Raviv et al. (2016) use Bayesian model to continuously keep track and updating of the estimation of the number of unusable bicycles in a station and focus on each bicycle independently. Bacciu, Carta et al. (2017) test the performance of support vector machine model with Gaussian kernel and random forest on predicting OD of a bike sharing ride, and successfully capture the seasonal pattern of the system. Compared to traditional models, the integrated models of machine learning have higher goodness of fit and lower standard error in predicting short-term real-time demand of bike sharing system. Zeng, Yu et al. (2016) extract global features and predict the demand for bike sharing system by using gradient boosting decision tree and neural network technology, which can overcome the limitation of applying a station-centric model to sparse data. Ai, Li et al. (2019) compare a deep learning approach, the convolutional long short-term memory (conv-LSTM) network to long short-term network in their study on short-term spatio-temporal distribution forecasting of dockless bike sharing system. Spatial-temporal variables taken into consideration include time of a day, number of bicycles in area, usage distribution, etc., and the result shows that conv-LSTM performs better. Some studies also make predictions through data mining. Kaltenbrunner, Meza et al. (2010) notice the clear pattern of travelers' behavior and thus use the auto-regressive moving average model to predict the amount of bicycles in the stations based on past data, and thus deepen the understanding of the demand of travelers. Chen, Ma et al. (2017) also find the sparsity and locality properties of the pattern and then formulate the pattern as an ill-posed inverse problem. Further evaluation on bike sharing system data from Washington D.C. and New York City proves this method can recover strong bike flows effectively and will be helpful when modeling travelers' demand. In conclusion, current studies mainly focus on demand modeling based on existing bicycle stations or for new bicycle stations under construction, only a few discuss about dockless bike sharing system. Compared to traditional bike sharing system, dockless bike sharing system eliminates fixed docking station, which brings more convenience to travelers when borrow or return bicycles. However, this will also lead to more obvious imbalances in time and space. For example, during rush hours, it can be hard to borrow bicycles near some bus or subway stations, while some are surrounded by disorderly parked bicycles. In addition, the bicycle damage rate is higher compared to tradition systems since it is more difficult to manage bicycles when they are parked everywhere, which will affect travel experience of travelers.

Current studies also show that there is a significant difference between the behavior of temporary users and annual members. Buck, Buehler et al. (2013) investigate travelers' behaviors through the comparison of short-term users and annual members, which includes trip purpose, helmet use and travel modes. The result indicates that annual members in D.C. are more likely to be women, younger with lower incomes, and be less likely to own vehicles. Faghih-Imani and Eluru (2015) perform quantitative analysis based on distinguishing between short-term users and annual members. Schoner and Levinson (2014) use linear regression to model travelers' choice behavior of the origin station in Minnesota. This modeling process also reveals the difference between the preferences of workers and non-workers. Through the elasticity, the study shows that commuters are more sensitive to changes in network density and college enrollment compared to other travelers. Fishman, Washington et al. (2015) predict bike sharing membership through logistic regression model and a wide range of variables are tested. The model can be used to show how the odds of the bike sharing system memberships change at different scenarios. Final variables included in the model have a big overlap with variables who have significant effect on the demand of bike sharing system, including sociodemographics characteristics (age, income) and spatio-temporal factor (whether work within 250m of docking station or not). In addition, other variables, including impact of mandatory helmet legislation on riding, riding activity in the past month and workplace end-of-trip facilities (showers, lockers, etc.), provide a different perspective when dealing with individual choice behavior. These variables are more micro compared to those also have significant effect on the demand of the bike sharing system. In addition to the differences between temporary users and annual members, some studies also explore the differences within annual members. de Chardon and practice (2019) clearly interpret the contradiction of benefits, purposes and outcomes of bike sharing system from different perspectives of travelers, operators, politicians, and society. The study also points out there is a large variance in bike sharing system frequency of use between members. People may purchase memberships out of environmental awareness or other reasons. Since membership does not equate use, Winters, Hosford et al. (2019)

put forward the concept of “super-users”. Multivariable logistic regression and backward stepwise regression are used to construct a model with the lowest AIC value. The profile of a superuser in the Vancouver case is a young male with relatively low income. Though the super-users account for only 10% of the members, they produce more than half of the trips.

While only a few existing studies explore choice models about whether travelers are willing to subscribe memberships of bike sharing systems, other fields have more mature research on this issue. Thøgersen (2009) apply motivation, opportunity, ability model in his study and discuss the effect of price promotion (a free month travel card) on car drivers’ travel behavior. The result is quite encouraging, since price promotion not only greatly increases the frequency of car drivers’ using public transportation, but also strengthens their awareness of using public transportation instead of private cars. Schlereth and Skiera (2012) compare the differences between bucket pricing plan and pricing plans without a marginal price (including pay-per-use, two-part, and three-part pricing plans). The hierarchical Bayes method is employed in the study to capture consumers’ decisions and the results show that bucket pricing plan can increase the number of subscription when ensuring profit per subscription. Kim, Nam et al. (2017) explore what do consumers prefer for music streaming services. They develop a multinomial logit model with random utility theory to estimate users’ marginal willingness-to-pay for a new music streaming service. Attributes including price, service duration, music quality, offline access, distribution channel, personalization, mobile application, and community features are tested. Based on the relatively high sum of the marginal willingnessto-pay of the attributes, they conclude that consumers are willing to pay a higher price for services provided by Spotify or Apple Music. Wang, Levin et al. (2020) build a customer’s dynamic choice model to capture strategic customers’ behavior when offed passes and draw the conclusion that passes may encourage consumption of customers that are note fully strategic by offering discount. Besides, customers’ uncertainty about future consumption might enable companies to gain more profit.

Overall, current studies have explored demand modeling of sharing bicycle systems with docking stations, customers’ choices on membership subscriptions in other fields, and pricing strategy that can help companies to gain more profit. There is still big space to discuss about demand modeling of dockless bike sharing systems and travelers’ choice behavior about membership subscriptions.

2 Data

2.1 Data Collection

An online survey was designed and implemented to collect data about commuting behavior and cycling preferences. Respondents were recruited from a representative Qualtrics panel of adult commuters in the metropolitan area of New York City, reaching a sample size of N = 801. One of the core section of the survey, in addition to more standard sections of sociodemographics and travel behavior questions, was a series of choice experiments as described below.

2.2 Data Description

2.2.1 Descriptive Statistics

As mentioned above, all respondents were adult commuters who live in the New York City Metro Area. Table 1 summarizes socio-demographics characteristics of the sample.

Table 1. Sample Demographic Statistics

Characteristics	Respondents(N=801)(%)
Male	39
18-24 years old	22
25-34 years old	31
35-44 years old	22
45-54 years old	13
55-64 years old	8
65+ years old	4
White	46
Black or African American	26
Asian	12
Hispanic/Latino	28
High school diploma or less	17
Some college experience	32
Bachelor's degree	33
Graduate or professional degree	18
Employed as a worker/student	89
Household income less than $25,000	16
Household income $25,000 to $34,999	9
Household income $35,000 to $49,999	11
Household income $50,000 to $74,999	20
Household income $75,000 to $99,999	15
Household income $100,000 or more	23
Household income prefer not to tell	6

Figure 1 shows the distribution of origin and destination county of the sample, considering only boroughs of NYC. The residence of the respondents is relatively evenly distributed in the four main boroughs of NYC, while the distribution of work/study places is quite uneven, which is highly concentrated in Manhattan.

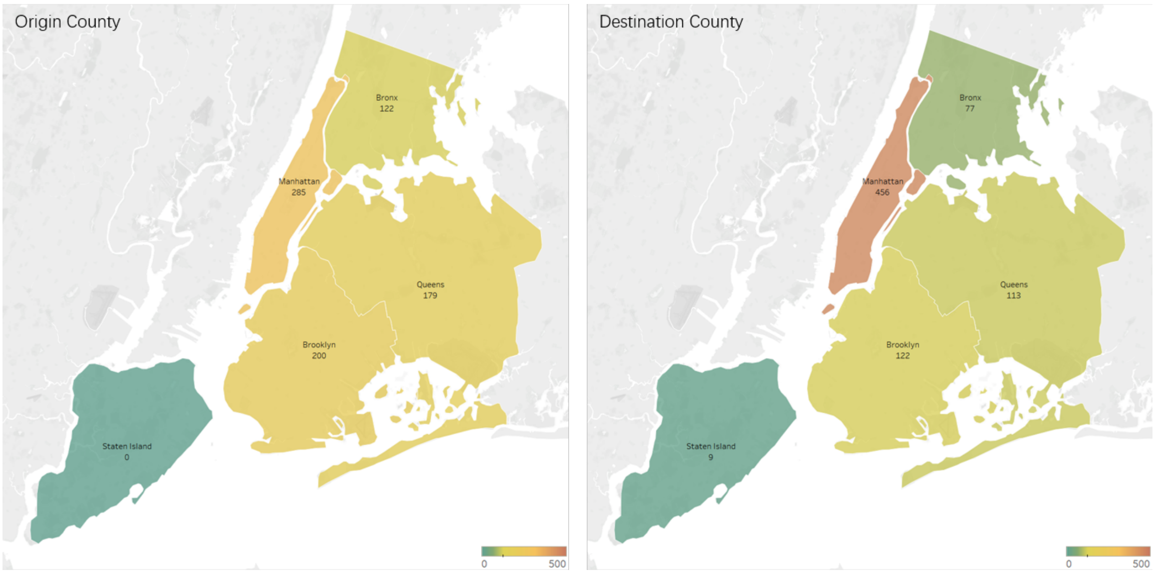


Figure 1. Distribution of Origin and Destination County

Among all the respondents, one third have no access to bikes, 108 are unable to ride a bike, while 242 consider themselves as beginner cyclist, 264 as intermediate cyclist and 187 as advanced cyclist. Figure 2 shows the percent of respondents with different cycling experiences in each group.

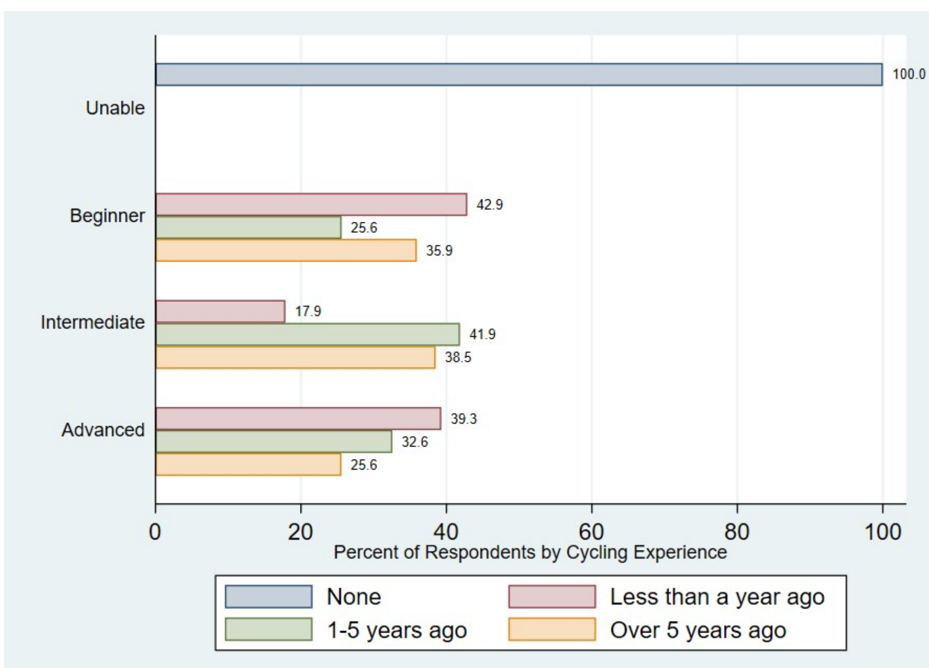
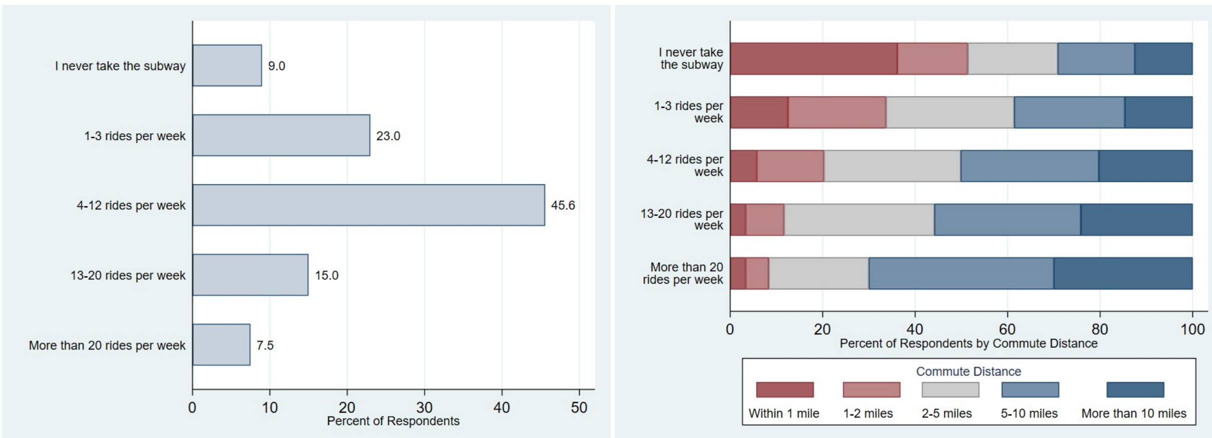


Figure 2. Cycling Experience of Respondents

2.2.2 Commuting Mode

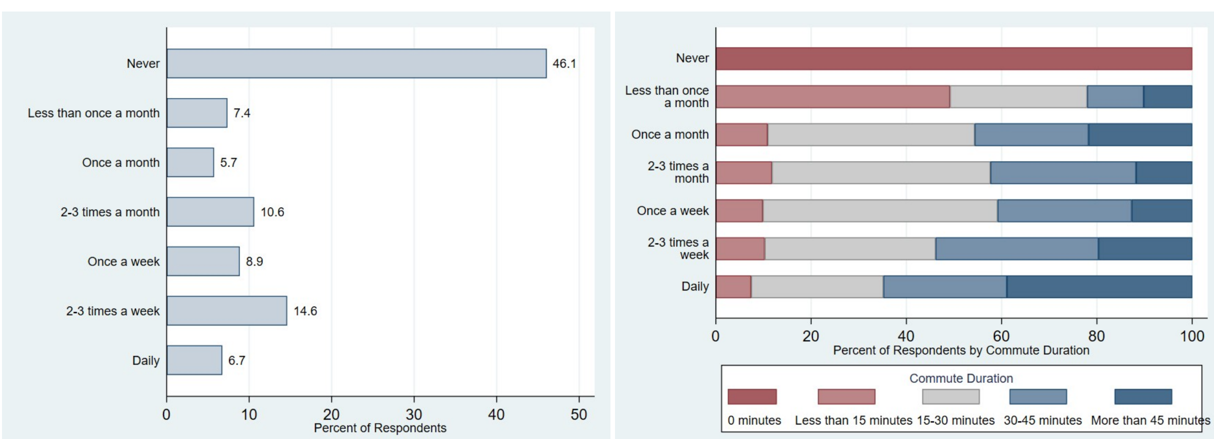
Figure 3a shows that nearly half of the respondents take the subway 4-12 rides per week. Frequency that is too low or too high may not be a good representative of commuters, since NYC has a well developed metro system and many commuters prefer to take the subway for commuting (twice a day, 5 days a week). Figure 3b indicates that respondents with shorter commute distance seem to take the subway less frequently, while those with longer commute distance seem to take the subway more frequently. This phenomenon is particularly obvious in the extreme groups (within 1 mile, 1-2 miles, more than 10 miles).



a) Percent b) Percent by Commute Distance

Figure 3. Frequency of Taking the Subway

Unlike the high frequency of taking the subway, few respondents choose to commute by bike, leading to the low frequency of biking. Figure 4a shows that nearly half of the respondents never commute by bike and only one fifth of the respondents commute by bike rather frequently. If the commute distance is less than 15 minutes by bike, respondents are unlikely to choose biking for commuting as figure 4b shows. Daily users have longer commute duration on average.



a) Percent b) Percent by Commute Distance

Figure 4. Frequency of Commuting by Bike

2.2.3 Safety Score

Respondents are randomly divided into two groups, and each group rates safety of cycling blocks with fast or slow car speed. Figure 5 shows the score distribution of different cycling blocks. It can be indicated from this figure that cycling blocks with protected cycle paths gets higher score than those unprotected ones. Besides, when the car speed is slower, respondents seem to feel safer cycling.

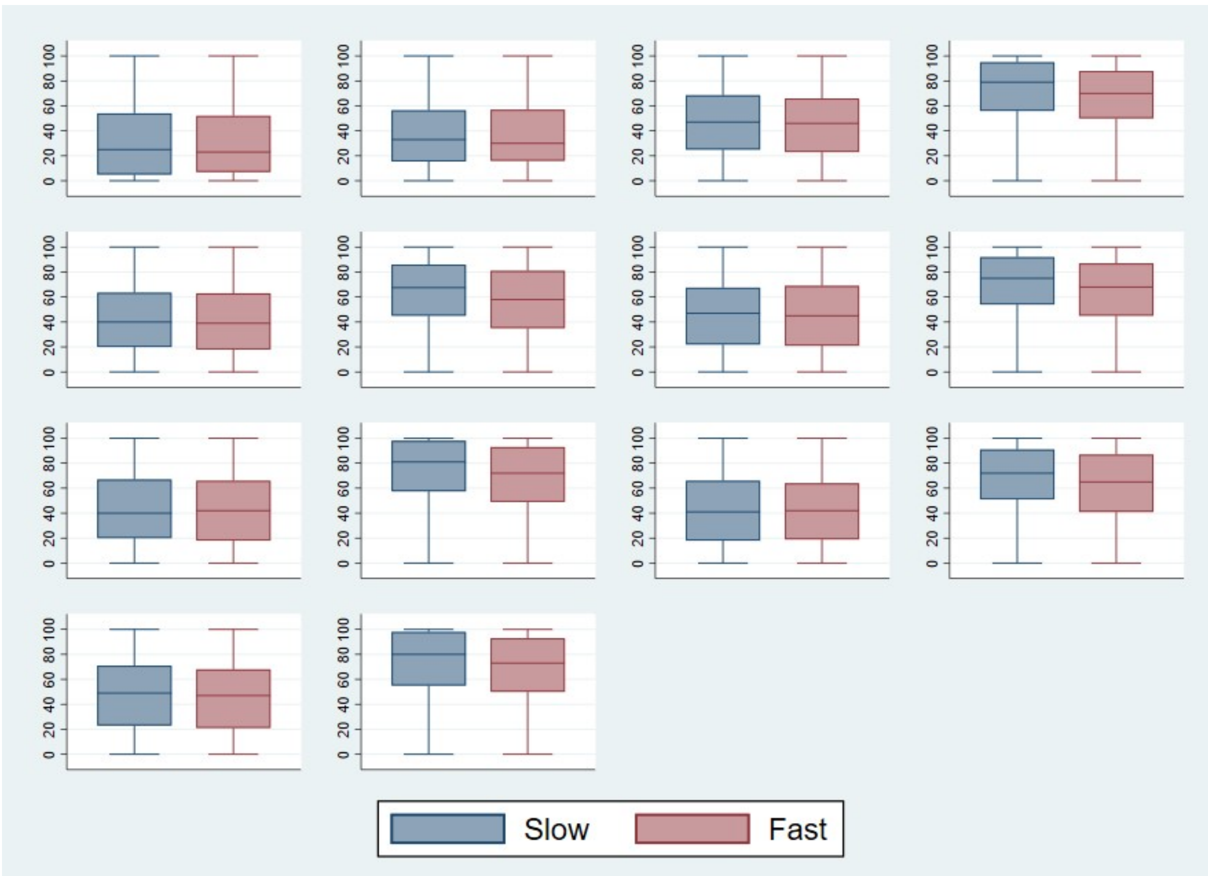
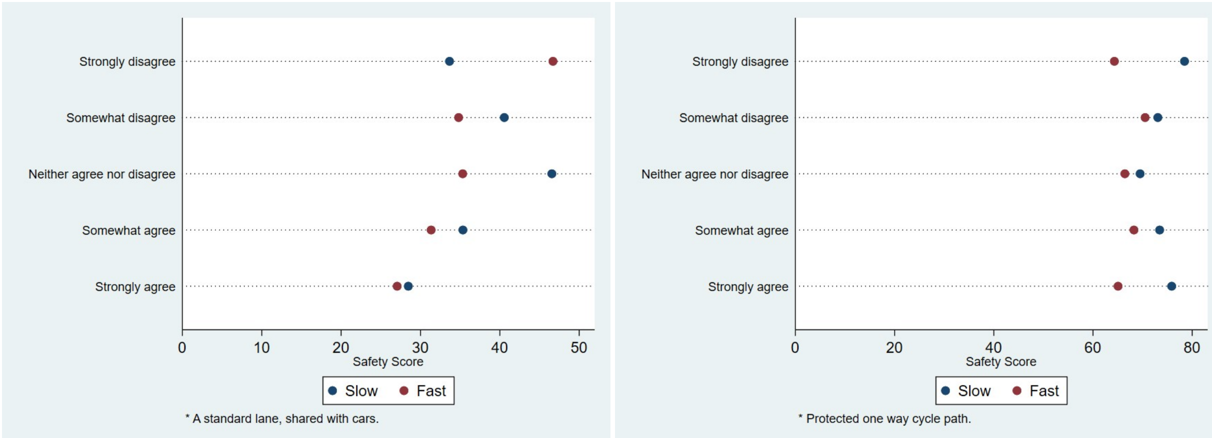


Figure 5. Safety Score of Different Cycling Blocks

If we go further and explore the relationship between safety scores and respondents' concerns about accidents, the score distributions have different patterns in different degrees. Figure 6a indicates that when the cycling path is standard lane sharing with cars, the two groups have big difference. When the car speed is relatively high, respondents who strongly disagree with accidents keep them from biking give the highest scores. This is reasonable since they are not worried about accidents at all. However, when it comes to slow car speed, respondents with neutral views give the highest score. As for respondents' concerns about lack of bike lanes, Figure 6b indicates that when there is a one-way cycle path, respondents' who hold neutral views give the lowest scores at slow speed and the highest scores at fast speed.



a) Worry about Accidents b) Lack of Lanes

Figure 6. Safety Score by Degree

2.3 Discrete Choice Experiments

One of the choice experiments was designed to represent the hypothetical decision of buying a pass for a ride sharing program. The experiment was introduced using the following text: ”Suppose that you are taking a leisure trip to a city in the summer. You are spending 3 full days in that city. You will now see a series of hypothetical passes for bike share. You will have the option of choosing a singleride, 1-day, or 3-day pass.” The experiment thus considered 3 labeled alternatives: a single ride, a day pass, and a 3-day pass. Attributes were: price per pass, maximum minutes included on a ride, extra cost for additional time (15 minute increment), whether the bike was a classic or an electric one, and whether the pickup/return was a docked or dockless system. Table 2 summarizes the attribute levels that were considered to design the choice experiments.

Table 2. Attribute Level of Choice Scenarios

Variable	Levels

Attribute levels were combined in 12 hypothetical scenarios following a D-efficient experimental design (D-error: 0.037624). Figure 7 shows a sample choice card as seen by respondents of the survey.

Single Ride	Day Pass	3-Day Pass
$6/trip	**$16/day**	**$30**
one ride up to **30 minutes** on a **classic bike**	Unlimited **45-minute rides** on a **classic bike**	Unlimited **30-minute rides** on an **electric bike**
Extra **$2** for additional 15 min	Extra **$5** for additional 15 min	Extra **$2** for additional 15 min
Docked	**Dockless**	**Dockless**
Pickup/return at designated racks	Pickup/return anywhere	Pickup/return anywhere

Figure 7. Sample Choice Card

3 Models and Conclusions

3.1 Logistic Regression

To explore the characteristics of respondents who have used Citi Bike, a logistic regression is fit. Table 3 presents results of related questions.

Table 3. Result of Related Questions

Questions	Respondents (N=801)	
	Yes	No
Do you have the app to use Citi Bike?	272	529
Have you actually used Citi Bike?	257	544

Socio-demographic characteristics and other questions related are chosen as the explanatory variables of the model. After moving away some very insignificant variables, the remaining variables and their descriptions are presented in Table 4, and the result of this logistic regression is shown in Table 5.

Table 4. Description of Variables

Variable	Description
app_use	Have actually used Citi Bike, yes (1) vs. no (0)
age	Age
height	Height
white	White, yes (1) vs. no (0)
black	Black or African American, yes (1) vs. no (0)
high_income	Household income higher than average level ($75,000), yes (1) vs. no (0)
high_educ	Education level higher than high school graduates, yes (1) vs. no (0)
bike_access	Have access to bikes through bikeshare service, have (1) vs. not (0)
car_ownership	Ownership of private cars, have (1) vs. not (0)
cycling_proficiency	Cycling proficiency better than beginner, yes (1) vs. no (0)
commute_close	Commute distance less than 2 miles, yes (1) vs. no (0)
accidents	Agree with worries about accidents keep them from biking more, yes (1) vs. no (0)
bike_frequently	Commuting using a bike more than once a week, yes (1) vs. no (0)
active_trans_workday	Walk/Biking more than 10 min on weekday, yes (1) vs. no (0)
spring_transit_bike	Commuting by transit and bike more than 2 days for a typical week in spring, yes (1) vs. no (0)
origin_manhattan	Living in Manhattan, yes (1) vs. no (0)
recreation_long_time	Duration of recreation by bike longer than 30 mins, yes (1) vs. no (0)
moderate_frequently	Doing moderate-intensity sports more than once a week, yes (1) vs. no (0)

This result indicates that respondents with the following characters are more likely having used the app:

- Socio-demographic characteristics: Young respondents with high household income and education degree are more likely having used the app. White and black or African American also present higher probability.

- Cyclist status: Respondents who are experienced cyclists and access bikes through bike sharing system are more likely having used the app. Cyclists who have private cars present higher probability than those who do not.
- Commuting-related attributes: Respondents who live in Manhattan and have commuting distance farther than 2 miles are more likely having used the app. Those who commute by bike more than once a week or commute by transit & bike more than 2 days for a typical week in spring also present higher probability.
- Activity-related attributes: Respondents who walk or ride a bike more than 10 minutes on weekdays, or use a bike for recreation more than 30 minutes each time are more likely having used the app. As for doing moderate-intensity sports, respondents with frequency less than once a week present higher probability.

Table 5. Logistic Regression Result

app_use	Coef.	Odds Ratio	*z*	*P > |z|*
*age******	-.0391102	.9616448	-4.44	0.000
height	-1.58939	.20405	-1.64	0.101
*white***	.8039918	2.234443	3.05	0.002
*black***	.8385914	2.313106	2.89	0.004
*high_income***	.6495001	1.914583	2.89	0.004
*high_educ**	.6438396	1.903777	2.12	0.034
*bike_access****	1.423147	4.150161	5.96	0.000
*car_ownership***	.8225094	2.276205	3.45	0.001
*cycling_proficiency****	1.484237	4.4116	6.10	0.000
commute_close+	-.4081616	.6648714	-1.67	0.095
*accidents**	-.4396678	.6442504	-2.17	0.030
*bike_frequently***	.8350546	2.30494	3.36	0.001
*active_trans_workday****	1.131404	3.100004	3.79	0.000
*spring_transit_bike****	1.320418	3.744986	4.77	0.000
*origin_manhattan***	.6576541	1.930259	3.05	0.002
*recreation_long_time**	.4233655	1.527092	1.99	0.047
*moderate_frequently**	-.4742986	.6223214	-2.20	0.028
_cons	-1.314596	.2685827	-0.82	0.410

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

In addition to the above conclusions, body-related variables (height, weight and health status) and income-related variables deserve further exploration.

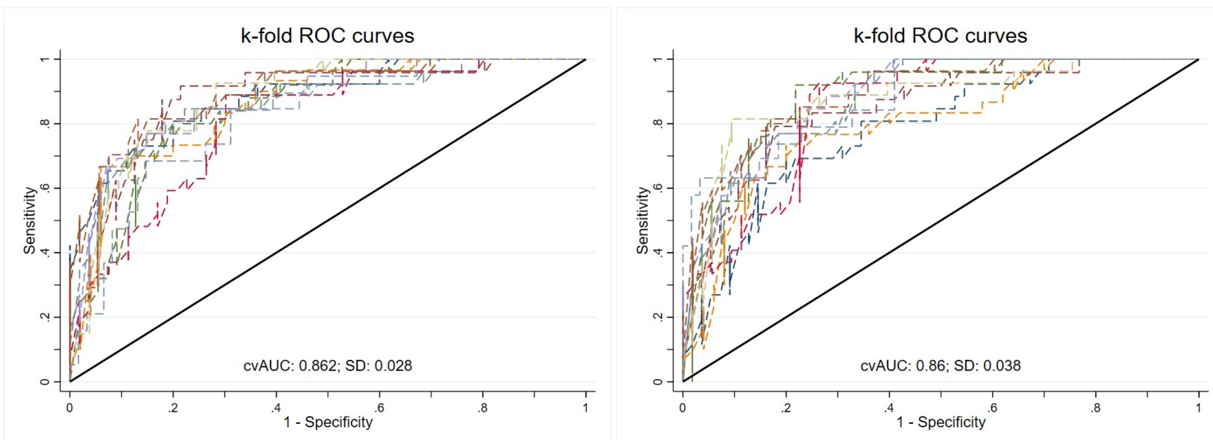
3.1.1 Body-Related Variables

Among all body-related variables obtained from the survey, only height is included in the logistic regression. Taller respondents seem less likely having used the app. Possible reason might be bikes of bike sharing systems are suitable for people of average height. For those who are taller than average level, those bikes might be uncomfortable since they have longer body.

Although reasonable conjectures can be given from this result, the influence of other body-related variables is still worth exploring. Since weight and health status are statistically insignificant and have been excluded from the logistic regression, a newly generated variable BMI is introduced. Body mass index (BMI) is a convenient rule of thumb used to broadly categorize a person as underweight, normal weight, overweight, or obese based on tissue mass (muscle, fat, and bone) and height. Equation 1 shows how to calculate BMI.

$$BMI(kg/m^2) = \frac{Weight}{Height^2} \tag{1}$$

When height is replaced by BMI in the logistic regression, the $p-value$ of BMI is 0.904, which means BMI is statistically insignificant in this logistic regression. In addition, Figure 8 shows the ROC curves (with 10-fold cross validation) of the original logistic regression and the one with BMI instead of height.



a) Original Regression b) Regression with BMI

Figure 8. ROC Curves of Regressions with Different Body-Related Variables

The mean AUC and AUC's standard deviation of the original logistic regression are both smaller than the one with BMI instead of height, thus height will be the only body-related variable in the logistic regression.

3.1.2 Income-Related Variables

In the logistic regression, a binary variable (high income) is generated to denote whether a respondent's household income is higher than average level (i.e., \$68,703 in 2019[1], taking into account options in the survey, \$75,000 is chosen to be the demarcation) or not.

However, there are other ways to deal with the income-related variable. The household income data obtained from the survey can be treated as a continuous variables with missing values (some respondents choose "prefer not to tell"). The descriptions of two newly generated variables along with the one used in the regression are shown in Table 6.

[1] Federal Reserve Bank of St. Louis. "Real Median Household Income in the United States." Accessed Sept. 24, 2020.

Table 6. Descriptions of Income-Related Variables

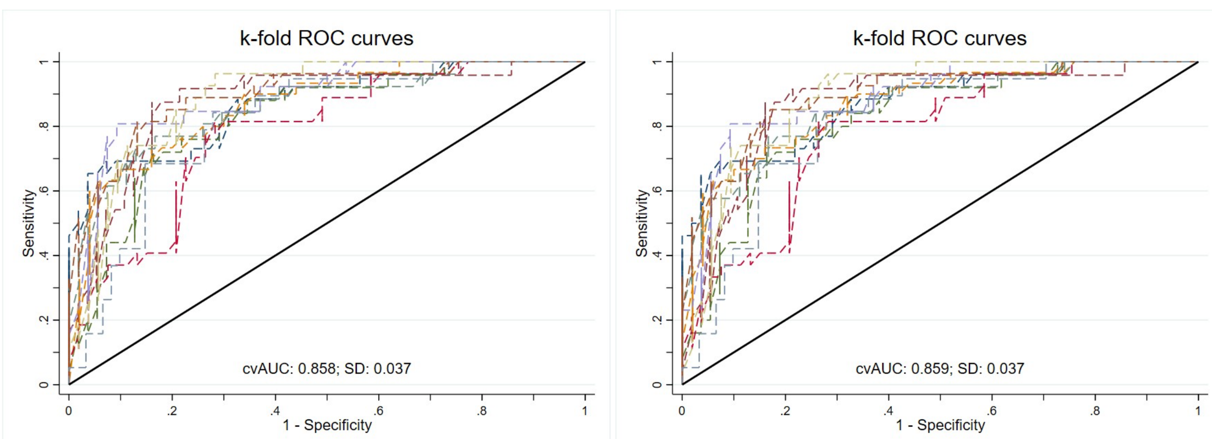
Variable	Type	Description
income	Continuous	Income, missing values are replaced by $0
missing_income	Binary	Income information is missing, yes (1) vs. no (0)
high_income	Binary	Household income higher than average level ($75,000), yes (1) vs. no (0)

When other explanatory variables in the logistic regression remain unchanged, different income related variables are added into the regression. Table 7 shows the detail of three different logistic regressions.

Table 7. Regressions with Different Income-Related Variables

Regression	1	2	Original
Income-related variable	*income*	*income*, *missing_income*	*high_income*

Logistic regression 1 is the one with *income* instead of *high_income*, and logistic regression 2 is the one with *income* and *missing_income* instead of *high_income*. In logistic regression 1, the $p\text{−}value$ of *income* is 0.507; in logistic regression 2, the $p\text{−}value$ of *income* is 0.588 and the $p\text{−}value$ of *missing_income* is 0.412. Neither *income* nor *income* and *missing_income* is statistically significant in the logistic regression. Figure 9 shows the ROC curves (with 10-fold cross validation) of logistic regression 1 and 2.



a) Regression with *income* b) Regression with *income* & *missing_income*

Figure 9. ROC Curves of Regressions with Different Income-Related Variables

The mean AUC and AUC's standard deviation of the original logistic regression are the smallest among all. Logistic regression 2 is slightly better than logistic regression 1, indicating that income data is not missing at random.

3.2 Conditional Logit Model

3.2.1 Basic Conditional Logit Model

As shown in the example choice card (Figure 7), five different attributes are tested in the experiment and they are listed in Table 8.

Table 8. Descritions of Attributes

Attribute	Description
price	Price of the pass
timeinc	Time included in the pass
xtratimefee	Extra fee for extra time
dockless	Bikes that can be pickup/return anywhere or not
ebike	Electric bike (with a small motor to assist pedaling) or not

With all these attributes, a conditional logit model is fitted, and the result is shown in Table 9. Number 1-4 are used to represent single ride, day pass, 3-day pass and opt out, respectively.

Table 9. Conditional Logit Model Result

	Estimate	s.e.	*p-value*
*asc_1****	0.279840	0.085014	$9.9590e^{-4}$
*asc_2****	0.638737	0.115603	$3.290e^{-8}$
*asc_3****	1.002093	0.201287	$6.410e^{-7}$
asc_4	0.000000	NA	NA
*price****	-0.022528	0.005290	$2.055e^{-5}$
timeinc	0.003320	0.003517	0.34519
xtratimefee	0.021569	0.016278	0.18517
*dockless****	0.237442	0.037614	$2.743e^{-10}$
ebike+	0.067542	0.037264	0.06991

AIC: 10867.3; BIC: 10917.66.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

As the result shows, respondents prefer pass with lower price and longer included time. However, they also prefer higher extra time fee, which is hard to explain. As for attributes of the bike sharing system, respondents prefer dockless system and electric bikes rather than docked system and classic bikes. As for different types of passes, respondents' preference increased from single ride to day pass, to 3-day pass.

The background of the experiment is "taking a leisure trip to a city in the summer and spending 3 full days". When traveling in other cities, people are less likely to drive there, which means they will use public transportation more. Since summer is a good season for cycling and people will spend three full days, it is interpretable that 3-day pass is the most popular one. Convenience is an important factor that people will consider when traveling, thus instead of docked system with limited stations, respondents prefer dockless system. Electric bikes are preferred for similar reason, that is electric bikes can make riding easier and let people enjoy leisure better. As for basic attributes (price, time included) of the pass, consumers always make wise choices, and thus they prefer lower price and longer time included.

Since *timeinc*, *xtratimefee* and *ebike* are not very significant in the model, a conditional model including only price and dockless is fitted. The result of the new model is shown in table 10. Lower AIC and BIC indicate that the new model has better performance than the previous one. What respondents care about the most during a leisure trip is the type, price of the pass and whether the system is dockless.

Table 10. Conditional Logit Model Result

	Estimate	s.e.	*p-value*
*asc_1******	0.42202	0.056385	$7.172e^{-14}$
*asc_2******	0.77619	0.075181	0.000
*asc_3******	1.13235	0.136921	$2.220e^{-16}$
asc_4	0.000000	NA	NA
*price******	-0.02226	0.003895	$1.096e^{-8}$
*dockless******	0.22572	0.037102	$1.175e^{-9}$

AIC: 10868.97; BIC: 10900.44.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

3.2.2 Conditional Logit with Latent Classes

In the basic conditional logit model, preferences are homogeneous. To capture preference heterogeneity across individuals, one way is assigning different classes to individuals according to their characteristics. Conditonal logit with two classes is used to judge whether the model has convergence problem. If the model fails to converge when there are only two classes, it will not converge in the case with more castegories. After excluding the combinations that fail to converge, descriptions of newly introduced variables in addition to some variables explained in Table 4 are shown in Table 11.

Table 11. Description of Variables

Variable	Description
female	Female, yes (1) vs. no (0)
hired	Employed as a worker/student, yes (1) vs. no (0)
not interested	Agree with no interests in biking keeps them from biking more, yes (1) vs. no (0)
not accessible	Agree with no access to bikes keeps them from biking more, yes (1) vs. no (0)
active trans weekend	Walk/Biking more than 10 min on weekends, yes (1) vs. no (0)
subway frequently	Commuting by subway more than 12 rides per week, yes (1) vs. no (0)
fall bike	Commuting by bike more than 2 days for a typical week in fall, yes (1) vs. no (0)
bike long time	Duration of commuting using a bike longer than 30 mins, yes (1) vs. no (0)
vigorous frequently	Doing vigorous-intensity sports more than once a week, yes (1) vs. no (0)

With selected variables, the optimal number of latent classes is selected by examining AIC and BIC. The model will not converge when there are more than four classes. AIC and BIC of models with different classes are shown in Table 12. As the result shows, models with latent classes are better than basic models. AIC and BIC are minimized with four classes.

Table 12. AIC and BIC of Tested Models

Number of Classes	2	3	4
AIC	8852.33	8295.67	7959.846
BIC	9052.05	8664.39	8497.558

The result of conditional logit model with four classes is shown in Table 13. For Class 1, class assignment is set as base, whereas for Class 2, 3 and 4, class assignment is a function of socioeconomic covariates. The share of Class 1, 2, 3 and 4 is 38.8%, 24.4%, 21.8% and 15.0%.

Table 13. Conditional Logit Model Result - 4 Classes

	Class 1		Class 2		Class 3		Class 4	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
A: Mean and Standard Deviations								
price	**-0.027**	0.013	**-0.050**	0.012	-0.037	0.060	-0.012	0.013
dockless	**0.408**	0.079	**0.378**	0.118	0.118	0.143	-0.162	0.444
asc_1	**2.084**	0.266	**2.217**	0.472	**2.600**	0.478	**-3.630**	0.415
asc_2	**3.521**	0.332	**3.254**	0.487	1.073	0.959	**-4.114**	0.466
asc_3	**2.821**	0.493	**5.908**	0.634	1.588	2.009	-3.738	.
B: Variable for Class Assignment								
constant			-0.780	0.556	-0.340	0.587	-.0660	0.635
age			0.010	0.009	0.009	0.010	**0.048**	0.009
female			0.056	0.241	0.058	0.252	-0.397	0.285
white			-0.384	0.276	-0.416	0.292	**-0.682**	0.332
black			-0.123	0.305	-0.002	0.319	**-0.934**	0.383
hired			-0.371	0.340	-0.184	0.352	0.542	0.415
app_use			0.113	0.308	-0.148	0.333	**-2.973**	1.267
car_ownership			0.092	0.248	-0.024	0.253	-0.539	0.280
bike_access			0.015	0.284	-0.031	0.308	0.778	0.436
cycling_proficiency			-0.026	0.273	-0.007	0.285	**-0.797**	0.333
not_interested			-0.117	0.256	0.230	0.241	**0.541**	0.269
not_accessible			-0.388	0.245	-0.014	0.244	**-0.621**	0.282
active_trans_weekend			0.247	0.332	-0.143	0.316	**-0.882**	0.301
subway_frequently			0.383	0.254	-0.016	0.279	-0.516	0.359
fall_bike			0.567	0.315	0.220	0.370	1.272	0.784
bike_long_time			0.116	0.251	-0.127	0.284	**-1.724**	0.549
vigorous_frequently			0.143	0.237	-0.431	0.261	**-0.909**	0.339
class share	38.8%		24.4%		21.8%		15.0%	

AIC: 7959.85; BIC: 8497.56.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

As inferred by the negative coefficients for *price*, all classes show preference to passes with lower price, though Class 3 and 4 show heterogeneity within the classes. However, preference for dockless system varies through classes. The willingness-to-pay for dockless system is shown in Table 14.

Respondents of Class 1, 2 and 3 prefer dockless system, while respondents of Class 4 prefer docked system. Notice that estimates of willingness-to-pay for dockless system of Class 3 and Class 4 are insignificant at 5%. Respondents vary widely in their willingness-to-pay for dockless system within the classes. As for pass types, different classes also show different preference. On average, Class 1 respondents prefer day pass to 3-day pass, singe ride and opt out; Class 2 respondents prefer 3-day pass to day pass, single ride and opt out; Class 3 prefer single ride to 3-day pass, day pass and opt out; Class 4 respondents prefer opt out to single ride, 3-day pass and day pass.

Table 14. WTP for Dockless System

	Class 1	Class 2	Class 3	Class 4
Mean	15.17	7.56	3.18	-13.56
SE	7.47	2.87	7.78	44.67

Class 1, with the largest percentage of respondents, is likely composed of individuals who are younger, white or black, and more professional in cycling. They are unlikely to opt out and prefer day pass the most. Excellent cycling skills together with yound and good body condition make them choose such a environmental friendly travel mode. They may also take the time in the round trip into consideration, and thus, they prefer day pass to 3-day pass. The highest willingness to pay for dockless system among all classes indicates these individuals attach great importance to the convenience of the trip. With dockless system, they can save both time and energy in finding a bike sharing station.

Class 2 is likely composed of individuals who commute by subway more than 12 rides per week, have at least one car and have used the app before. The experience of using the app and the habit of using public transportation in their daily life make these individuals use bike sharing system in the trip. They are interested in biking and they will commute by bike for more than 30 minutes. In addition to their love for biking, they exercise a lot in their daily life. They do vigorous-intensity sports more than once a week and walk or ride bikes more than 10 minutes on weekends. The smallest estimate of price indicates their sensitivity to price. For these budget-conscious sports enthusiasts, the cheaper and longer the pass, the better.

Class 3 is likely composed of individuals who do not have access to bikes in their daily life. Instead of purchasing passes with long time, they prefer to pay for single ride. Perhaps bike sharing system is quite new to them. They are willing to try, however, they are unlikely to purchase day pass or 3-day pass until they have some good using experience.

Class 4, with smallest percentage of respondents, appears to be composed of individuals who are male or other gender, and hired as a worker or student. Though differences exists within the class, these individuals show less interests in using the bike sharing system in general. Some of them may be less aware of the increasingly serious environmental problems and choose to drive cars or take a taxi in the trip, while the others may choose to travel only by public transportation.

Compared to the basic conditional logit model, this model expose heterogeneity through classes. Though some of the estimates are insignificant at 5% level, average preferences of individuals with different characteristics are still useful guidelines for developing bike sharing system.

3.3 Mixed Logit Model

In the basic conditional logit model, preferences are homogeneous. To capture preference heterogeneity across individuals, another way is using random variables instead of fixed ones. A mixed logit model with all attributes of the pass is fitted. All attributes are assumed to have normally distributed parameters and the result is shown in Table 15.

Table 15. Mixed Logit Model Result

	Estimate	s.e.	*p-value*
*asc_1******	2.167913	0.143707	0.00000
*asc_2******	3.826300	0.221293	0.00000
*asc_3******	4.657485	0.364701	0.00000
asc_4	0.000000	NA	NA
*price_mean******	-0.131417	0.012648	0.00000
*price_se******	0.161758	0.009244	0.00000
timeinc_mean	0.005626	0.005490	0.30551
*timeinc_se******	0.042535	0.008433	$4.567e^{-7}$
*xtratimefee_mean****	-0.079278	0.031201	0.01106
*xtratimefee_se******	0.522975	0.034157	0.00000
*dockless_mean******	0.350612	0.063388	$3.181e^{-8}$
*dockless_se******	0.793729	0.103592	$1.821e^{-14}$
ebike_mean	0.042300	0.062183	0.49634
*ebike_se******	0.917856	0.088188	0.00000

AIC: 8855.4; BIC: 8937.24.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

With the mean and standard deviation estimated, the distributions of random parameters are shown in Figure 10. The distributions of *price* and *timeinc* are more concentrated than the other three, indicating less heterogeneity across individuals.

Similar to the result given by the conditional logit model, the result of the mixed logit model shows respondents' preferences to passes with lower price, longer included time, dockless system, electric bikes and longer pass time (i.e., preferences increase from single ride to day pass, to 3-day pass). However, in the conditional logit model, respondents show preferences to higher extra time fee, which is hard to explain. This time, in the mixed logit mode, the mean of xtratimefee indicates that respondents actually prefer lower extra time fee.

Since the mean of *timeinc*, *xtratimefee* and *ebike* are not very significant in the model, which is the same as the result given by the conditional logit model, mixed logit models neglecting these attributes and taking some of the characteristics into consideration are fitted. Characteristics including gender, education level, income level, cycling proficiency and experience of using the app are tested separately.

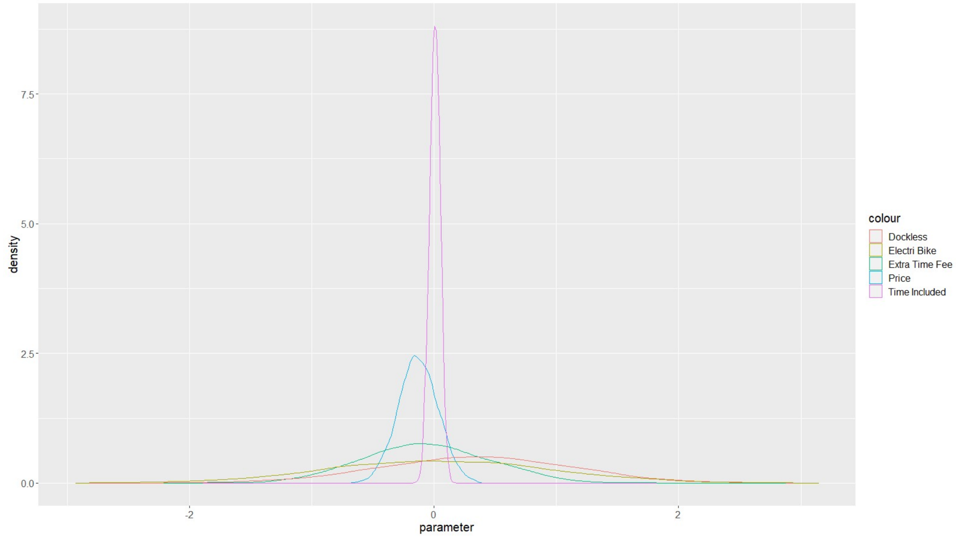


Figure 10. Distribution of Parameters

3.3.1 Interaction with Gender

Among all the respondents, 486 (60.67%) are female, 313 (39.08%) are male and 2 (0.25%) are other gender. A mixed logit model considering the interaction between *dockless* and *gender* is fitted and the result is shown in Table 16.

Table 16. Mixed Logit Model Result - *dockless* × *gender*

	Estimate	s.e.	*p-value*
*asc_1****	1.63756	0.08796	0.00000
*asc_2****	3.17856	0.13824	0.00000
*asc_3****	3.72389	0.22267	0.00000
asc_4	0.00000	NA	NA
*price_mean****	-0.13662	0.01044	0.00000
*price_se****	0.19385	0.01060	0.00000
*dockless_female_mean****	0.34054	0.07581	$7.047e^{-6}$
*dockless_female _se****	0.95270	0.10728	0.00000
dockless_male_mean^{+}	0.17173	0.08895	0.05352
*dockless_male _se****	0.82335	0.13116	$3.437e^{-10}$
dockless_other_mean	0.07951	2.64422	0.97601
dockless_other _se	2.57489	2.99366	0.38973

AIC: 9213.62; BIC: 9282.87.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The distribution of willingness-to-pay is shown in Figure 11. The mean WTP of female respondents is nearly twice the mean WTP of male respondents, indicating female respondents have stronger preference for dockless system. It seems like female respondents are more concerned about the convenience of the journey. The significant standard deviations of the two groups indicate heterogeneity across individuals within each gender group.

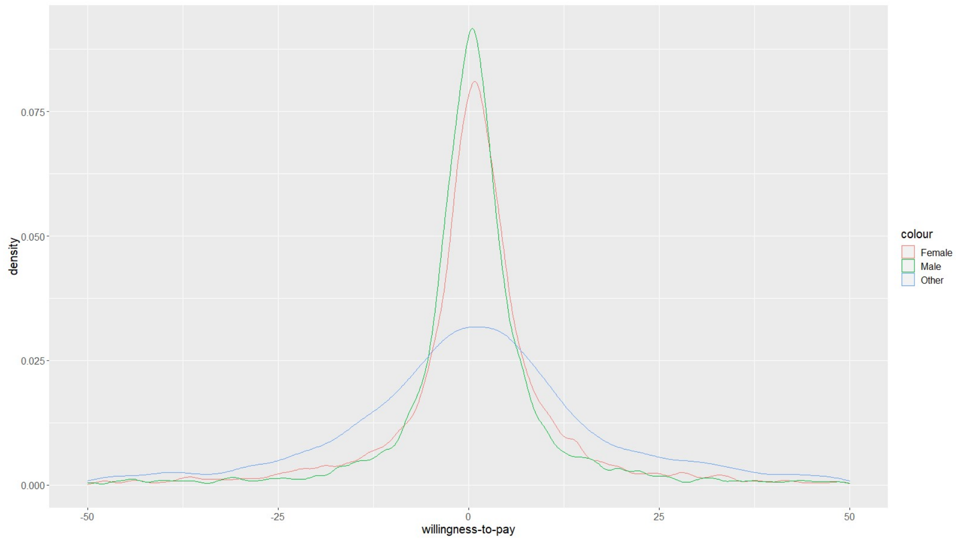


Figure 11. Distribution of Willingness-to-Pay - *dockless* × *gender*

Parameters related to other gender are insignificant in this model. However, since we only have 2 respondents of other gender in the sample, it is inappropriate to make a conclusion on the preference of other gender for dockless system.

3.3.2 Interaction with Education Level

Among all the respondents, 663 (82.77%) have education degree higher than high school graduates and 138 (17.23%) do not. A mixed logit model considering the interaction between *dockless* and *high_educ* is fitted and the result is shown in Table 17.

Table 17. Mixed Logit Model Result - *dockless* × *education level*

	Estimate	s.e.	*p-value*
*asc_1****	1.6406	0.08798	0.00000
*asc_2****	3.1822	0.13827	0.00000
*asc_3****	3.7307	0.22273	0.00000
asc_4	0.0000	NA	NA
*price_mean****	-0.1369	0.01046	0.00000
*price_se****	0.1942	0.01061	0.00000
*dockless_high_educ_mean****	0.2814	0.06412	$1.142e^{-5}$
*dockless_high_educ _se****	0.8889	0.09268	0.00000
dockless_low_educ _mean+	0.2506	0.14154	0.07659
*dockless_low_educ _se****	1.0022	0.18861	$1.074e^{-7}$

AIC: 9212.87; BIC: 9269.53.

Note: + *p < 0.1*; * *p < 0.05*; ** *p < 0.01*; *** *p < 0.001*.

The distribution of willingness-to-pay is shown in Figure 12. The mean WTP of respondents who have education degree higher than high school graduates is slightly higher than the mean WTP of respondents who do not. Respondents with higher education level show more preference to dockless system, though the heterogeneity across education level is not very obvious. The significant standard deviation of the two groups indicate heterogeneity across individuals within each education level group.

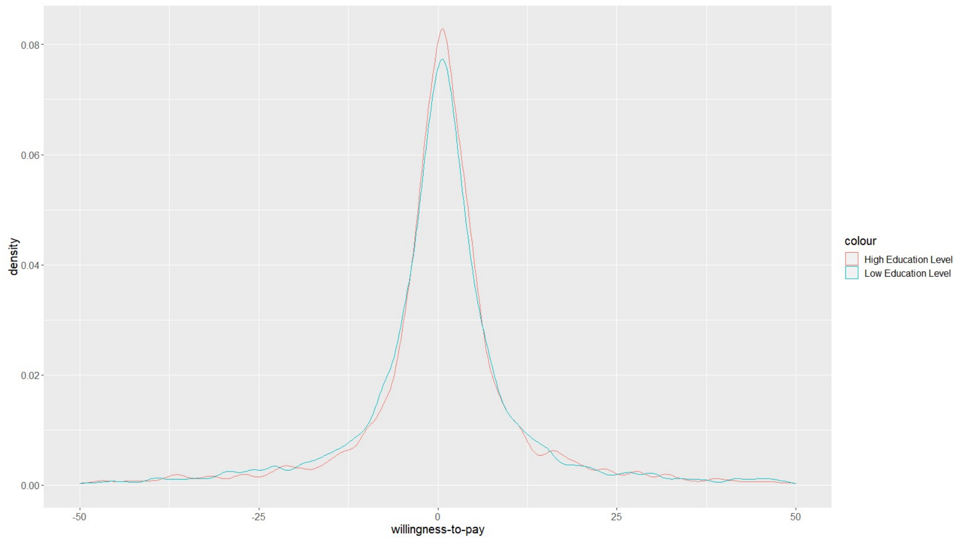


Figure 12. Distribution of Willingess-to-Pay - *dockless* × *education level*

3.3.3 Interaction with Income Level

Among all the respondents, 310 (38.70%) have income level higher than the average level ($75,000) and 491 (61.30%) do not. A mixed logit model considering the interaction between *dockless* and *high_income* is fitted and the result is shown in Table 18.

Table 18. Mixed Logit Model Result - *dockless* × *income level*

	Estimate	s.e.	*p-value*
*asc_1****	1.6424	0.08803	0.00000
*asc_2****	3.1872	0.13833	0.00000
*asc_3****	3.7395	0.22279	0.00000
asc_4	0.0000	NA	NA
*price_mean****	-0.1373	0.01046	0.00000
*price_se****	0.1944	0.01063	0.00000
*dockless_high_income_mean***	0.2480	0.08756	0.004628
*dockless_high_income_se****	-0.7730	0.13860	$2.445e^{-8}$
*dockless_low_income_mean****	0.2947	0.07655	$1.1789e^{-4}$
*dockless_low_income_se****	-0.9880	0.10462	0.00000

AIC: 9211.33; BIC: 9267.99.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The distribution of willingness-to-pay is shown in Figure 13. The mean WTP of respondents whose income level is higher than the average level is slightly lower than the mean WTP of respondents whose income level is lower than the average level. Respondents with higher income level show less preference to dockless system, though the heterogeneity across income level is not very obvious. The significant standard deviation of the two groups indicate heterogeneity across individuals within each income level group.

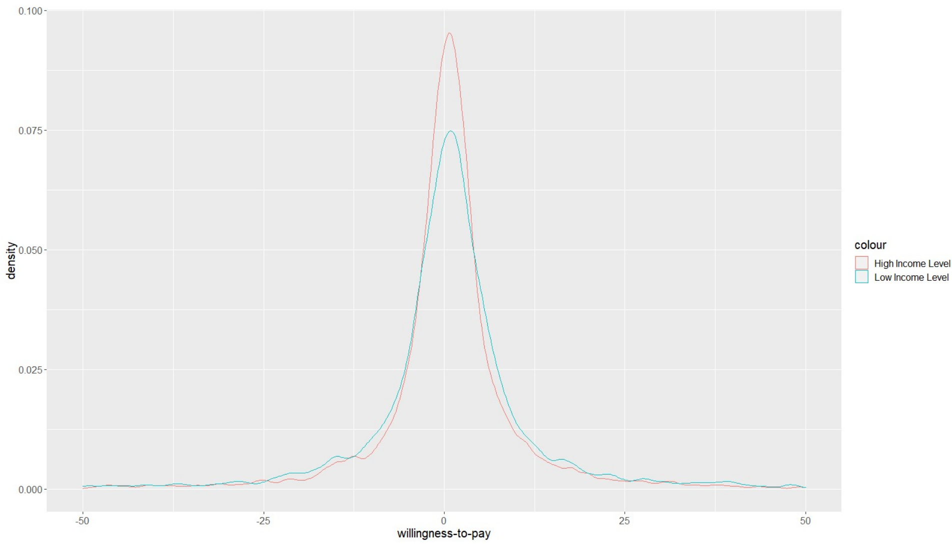


Figure 13. Distribution of Willingess-to-Pay - *dockless* × *income level*

3.3.4 Interaction with Cycling Proficiency

Among all the respondents, 451 (56.30%) have cycling skill better than beginners and 350 (43.70%) do not. A mixed logit model considering the interaction between *dockless* and *high_pro* is fitted and the result is shown in Table 19.

Table 19. Mixed Logit Model Result - *dockless* × *cycling proficiency*

	Estimate	s.e.	*p-value*
*asc_1****	1.6396	0.08777	0.00000
*asc_2****	3.1717	0.13821	0.00000
*asc_3****	3.7311	0.22235	0.00000
asc_4	0.0000	NA	NA
*price_mean****	-0.1352	0.01030	0.00000
*price_se****	0.1894	0.01038	0.00000
*dockless_high_pro_mean****	0.3248	0.07039	$3.940e^{-6}$
*dockless_high_pro_se****	0.7387	0.10815	$1.601e^{-11}$
dockless_low_pro_mean+	0.1673	0.10108	0.09786
*dockless_low_pro_se****	-1.1939	0.13590	0.00000

AIC: 9204.29; BIC: 9260.95.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The distribution of willingness-to-pay is shown in Figure 14. The mean WTP of respondents whose cycling skill is better than beginners is nearly twice the mean WTP of respondents who have less cycling experience, indicating these respondents have stronger preference for dockless system.

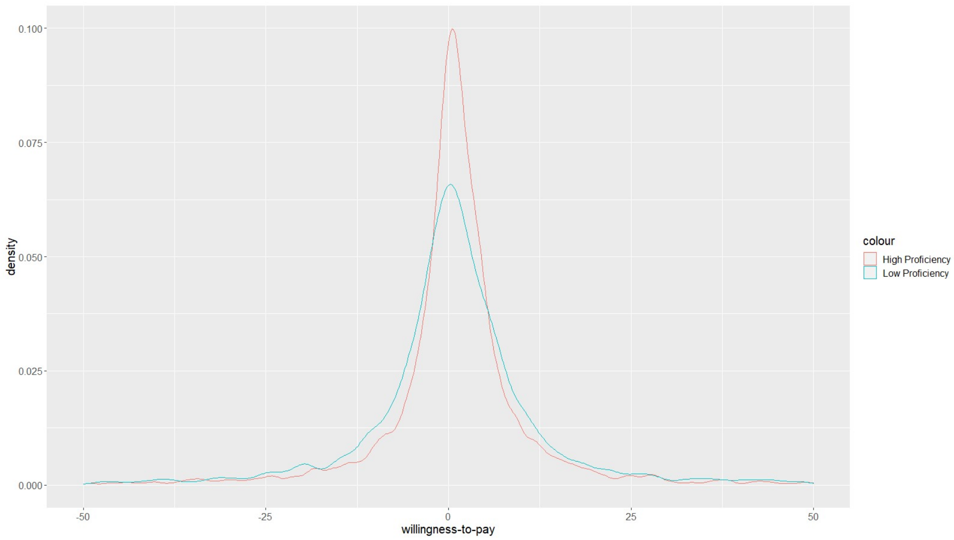


Figure 14. Distribution of Willingness-to-Pay - *dockless* × *cycling proficiency*

It seems like respondents with more cycling experience are more concerned about the convenience of the journey. The significant standard deviations of the two groups indicate heterogeneity across individuals within each proficiency group.

3.3.5 Interaction with App Use

Among all the respondents, 257 (32.08%) have used the app and 544 (67.92%) have not. A mixed logit model considering the interaction between *dockless* and *app_use* is fitted and the result is shown in Table 20.

Table 20. Mixed Logit Model Result - *dockless* × *app use*

	Estimate	s.e.	*p-value*
*asc_1****	1.6465	0.08786	0.00000
*asc_2****	3.1680	0.13826	0.00000
*asc_3****	3.7243	0.22220	0.00000
asc_4	0.0000	NA	NA
*price_mean****	-0.1341	0.01025	0.00000
*price_se****	0.1877	0.01026	0.00000
*dockless_use_mean****	0.5190	0.08605	$1.623e^{-9}$
*dockless_use_se****	-0.5254	0.16116	0.001113
dockless_not_mean^{+}	0.1177	0.07800	0.131396
*dockless_not_se****	-1.0979	0.10430	0.000000

AIC: 9191.72; BIC: 9248.37.

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The distribution of willingness-to-pay is shown in Figure 15. The mean of WTP of respondents who have used the app is more than four times the mean WTP of respondents who have not, indicating these respondents have stronger preference for dockless system. It seems like respondents who have used the app are more concerned about the convenience of the journey. The significant standard deviations of the two groups indicate heterogeneity across individuals within each group.

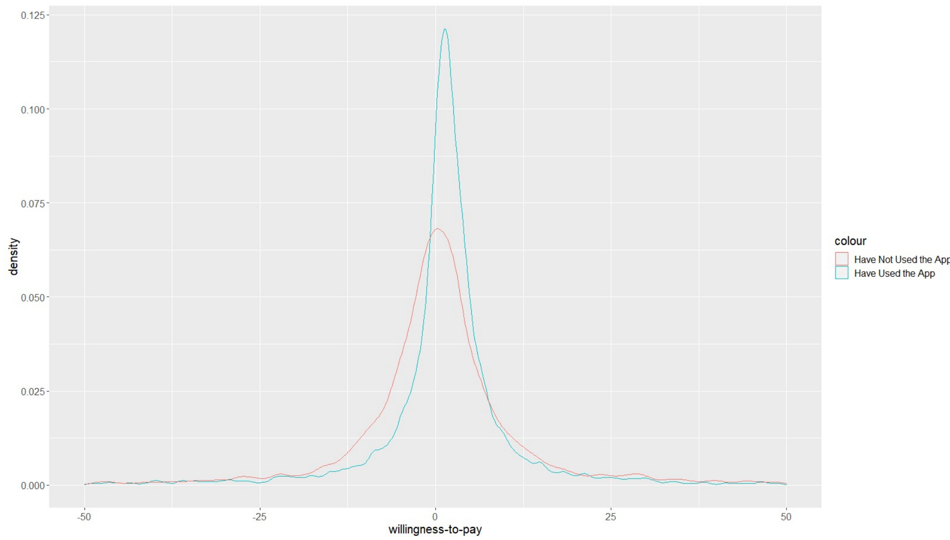


Figure 15. Distribution of Willingness-to-Pay - *dockless* × *app use*

Conclusions obtained from these models that interact with different characteristics are consistent with those obtained from conditional logit with latent class. These models give a more specific explanation of how these characteristics will affect user’s choices.

3.4 Conclusions

Bike sharing system has developed rapidly due to increasingly serious traffic jams and environmental problems. In order to promote shared bikes, bike sharing companies will be curious about what has an important impact on whether users use bike sharing system and subscribing to memberships. In this report, survey data collected from NYC metro area is used.

In terms of the use of Citi Bike app, young and experienced cyclists with high household income and education degree present higher probability. Unlike what might have been assumed, people who own private cars are more likely to have used bike sharing system. In addition, commuting mode and exercise habits also have impacts on the probability.

Conditional logit model is proposed to model the stated choices of respondents. Significant negative estimate of price proves respondents’ preferences for low price. It has been recognized that among all the pass attributes tested, the highest valued feature is being dockless. Respondents’ willingness-to-pay for dockless bike sharing system is $10.53 on average, which is very high even compared with the price of passes. On the contrary, other features seem to be less attractive.

Respondents show some interest in e-bikes, while do not care about time included and extra time fee at all. For day pass and 3-day pass, since they provide unlimited rides in a limited time, users can reborrow the bike at the end of each ride. Though being a bit troublesome, it makes the time included in a ride and extra time fee unimportant. Besides, for a leisure trip, people generally will not choose biking when the distance between two locations is too long.

As a feature most valued by respondents, dockless is further explored. The conditional logit with latent class and mixed logit estimates reveal heterogeneity in the valuation of dockless bike sharing system. 63.2% of respondents exhibit positive willingness-to-pay for dockless bike sharing system and 85% exhibit positive willingness-to for single ride. 15% respondent, mainly consist of non-female workers or students, are going to opt out. It is surprising to find that hired respondents show little interests in biking during a leisure trip. Respondents with higher household income and education level have higher probability of having used the Citi Bike app, but the heterogeneity of willingness-to-pay for dockless bike sharing system across household income level and education level is not obvious (less than $1).

What do affect heterogeneity of the willingness-to-pay for dockless bike sharing system are the gender, cycling skills and experience of using the Citi Bike app. There is no doubt that people with poor cycling skills will not choose to travel by bike, and thus the feature of dockless does not matter to them. The estimates prove this by showing that respondents with better cycling skills are also willing to pay nearly twice for dockless bike sharing system compared with those who are bad at cycling. As for the impact of gender, it is often recognized that female will pay more attention to the experience of trips, which means that they may pay more attention to convenience. In addition to being less likely to opt out, female respondents are willing to pay nearly twice for dockless bike sharing system compared with respondents of other genders, which is consistent with the usual assumptions. Last but not the least, previous experience of using the Citi Bike app matters. Respondents who have acutally used the Citi Bike app before are willing to pay more than four times for dockless sharing system than those who have not. This result shows that people can truly appreciate the convenience brought by dockless bike sharing system only after they have experienced both docked and dockless bike sharing system.

In general, bike sharing companies should first consider encouraging more people to use the bike sharing system. Trying the first ride free may be a good way to attract people who have lower household income or do not present a strong willingness to try. A more easy-to-use app makes it easier for older people to give a try. Then, to attract existing users purchasing passes, passes need a better pricing strategy. For example, since the time included and extra time fee are meaningless when unlimited rides are provided, bike sharing companies could reduce the time included and raise the extra time fee for a single ride, and do the opposite to passes with longer time period. Also, they should considering making dockless bike sharing systems only available to users with passes. In this case, single-ride users can only access docked bike sharing system. Since everyone is willing to pay for dockless bike sharing system, this may attract single-ride users to purchase passes.

There are some limitations in this report. The data set is too small, and thus the population of certain characteristics is too small to make the result representative enough. Also, the data set could be explored in more different ways. For example, studying who is opting in and out may provide valuable advice to the operation of bike sharing companies. Future research could consider discussing further topics with a larger data set.

References

Ai, Y., et al. (2019). "A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system." **31**(5): 1665-1677.

Bacciu, D., et al. (2017). "An experience in using machine learning for short-term predictions in smart transportation systems." **87**: 52-66.

Buck, D., et al. (2013). "Are bikeshare users different from regular cyclists? A first look at short-term users, annual members, and area cyclists in the Washington, DC, region." **2387**(1): 112-119.

Campbell, A. A., et al. (2016). "Factors influencing the choice of shared bicycles and shared electric bikes in Beijing." **67**: 399-414.

Chen, L., et al. (2017). "Understanding bike trip patterns leveraging bike sharing system open data." **11**(1): 38-48.

de Chardon, C. M. J. T. r. p. A. p. and practice (2019). "The contradictions of bike-share benefits, purposes and outcomes." **121**: 401-419.

DeMaio, P. J. J. o. p. t. (2009). "Bike-sharing: History, impacts, models of provision, and future." **12**(4): 3.

Faghih-Imani, A. and N. J. J. o. t. g. Eluru (2015). "Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system." **44**: 53-64.

Faghih-Imani, A. and N. J. J. o. T. G. Eluru (2016). "Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system." **54**: 218-227.

Fishman, E., et al. (2015). "Bikeshare's impact on active travel: Evidence from the United States, Great Britain, and Australia." <u>Journal of Transport & Health</u> **2**(2): 135-142.

Fishman, E., et al. (2015). "Factors influencing bike share membership: An analysis of Melbourne and Brisbane." <u>Transportation research part A: policy and practice</u> **71**: 17-30.

Kaltenbrunner, A., et al. (2010). "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system." **6**(4): 455-466.

Kaspi, M., et al. (2016). "Detection of unusable bicycles in bike-sharing systems." **65**: 10-16.

Kim, J., et al. (2017). "What do consumers prefer for music streaming services?: A comparative study between Korea and US." **41**(4): 263-272.

Martin, E. W. and S. A. Shaheen (2014). "Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two US cities." <u>Journal of Transport Geography</u> **41**: 315-324.

Mátrai, T. and J. J. T. r. p. Tóth (2016). "Comparative assessment of public bike sharing systems." **14**: 2344-2351.

Rixey, R. A. J. T. r. r. (2013). "Station-level forecasting of bikesharing ridership: Station Network Effects in Three US Systems." **2387**(1): 46-55.

Schlereth, C. and B. J. I. J. o. R. i. M. Skiera (2012). "Measurement of consumer preferences for bucket pricing plans with different service attributes." **29**(2): 167-180.

Schoner, J. E. and D. M. J. T. Levinson (2014). "The missing link: Bicycle infrastructure networks and ridership in 74 US cities." **41**(6): 1187-1204.

Shaheen, S. A., et al. (2010). "Bikesharing in Europe, the Americas, and Asia: past, present, and future." **2143**(1): 159-167.

Tang, Y., et al. (2017). "Research on users' frequency of ride in Shanghai Minhang bike-sharing system." **25**: 4979-4987.

Thøgersen, J. J. T. P. (2009). "Promoting public transport as a subscription service: Effects of a free month travel card." **16**(6): 335-343.

Vogel, P., et al. (2011). "Understanding bike-sharing systems using data mining: Exploring activity patterns." **20**: 514-523.

Wang, J., et al. (2020). "Selling passes to strategic customers." **68**(4): 1095-1115.

Wang, S., et al. (2010). Bike-Sharing-A new public transportation mode: State of the practice & prospects. 2010 IEEE International Conference on Emergency Management and Management Sciences, IEEE.

Winters, M., et al. (2019). "Who are the 'super-users' of public bike share? An analysis of public bike share members in Vancouver, BC." **15**: 100946.

Zeng, M., et al. (2016). "Improving demand prediction in bike sharing system by learning global features."