# Analysis and comparison of Uber, Taxi and Uber request via Transit 

Jing Guo (Changsha University of Science and Technology)<br>jingguo97@outlook.com<br>Mentors: Dr.Candace Brakewood, Dr. Cheng Liu, Dr. Kwai Wong cbrakewo@utk.edu, cheng.n.liu@gmail.com, kwong@,utk.edu<br>August 3, 2018


#### Abstract

As urban residents who are unable to drive by themselves, their transportation within the city is mainly public transportation including buses, subways and railways, taxis and the Uber which is very popular among people recent year. The above three modes of transportation are important components of a city, and each has its own characteristics. Among them, there are competitive taxis and Uber, the main difference is the way of calling and the way to pay.

New York City is a city with a complex transportation system and a large residential community. Its traffic problems are most obvious and prominent. Focusing on the analysis of the choice of transportation modes of residents in New York City and the analysis of traffic characteristics are very typical.

In the era of rapid development of mobile communication devices (smartphones). People's lives are closely linked to smartphones. Technology has also greatly facilitated people's lives. This paper is about the apps called Transit and Uber. The Transit app can be used to query real-time public transit time information. Uber can be used to call or book a Uber driver by a smartphone.

How different time and different environments affect different passengers to make different choices is the main research issue of this paper. Visualization of data to analyze travel time and spatial characteristics of each user.

Using four data sets to analyze travel characteristics of users with different travel modes and the factors that influence users' choice of transportation mode. The data set includes: Transit (which seeks to easily informs users of transit, bike-share, car-share, and Uber access based on their geographic position), Taxi data (including Yellow and Green)from the New York City Taxi and Limousine Commission ,Uber ride-hail origin data released publicly through the New York City Taxi and Limousine Commission and New York Central Park weather data comes from the National Climatic Data Center.


## 1.BACKGROUND

New York City's transportation is very complex and diverse. Because it is the largest and most crowded city in the United States and the largest metropolitan area in the world, the city's traffic is very large.

As an essential urban transport for travel, it has a great impact on the economic development of the entire city and the lives of the residents. For those who do not use their own vehicles or do not want to drive. The choices for travel are as follows: urban public transport systems including buses ,subways and train, taxis, and fhv(for-hire vehicles) such as Uber.

Taxi is almost a must-have for every city. This type of transportation has been around for a long time and is still popular in urban. However, with the advancement of the technology era and the development of mobile communication equipment, taxis have not been linked to mobile devices. User have to use cash to pay.
Uber is becoming more and more popular. It develops mobile applications to connect passengers and drivers, and provides shared economy services for passenger car rental and real-time sharing. Passengers can reserve these passenger vehicles by sending text messages or using the mobile app, and they can also track the location of the vehicle when using the mobile app. The premise of using uber is that you must have a mobile app to call it and you need to complete the payment online, you can not use cash.

Transit is a very useful application that provides a public transit timetable query service and calculates the waiting time for the user's location based on GPS positioning. It also provides users with the option of sharing bicycles and shared cars like uber. This is an app that brings convenience to users and has always been very popular among users.

## 2. INTRODUCTION

### 2.1 INTRODUCTION OF TRANSIT APP

Figure1: Screenshots of the Transit app main function


Public transportation timetable inquiry


Select the target location directly on the map


Trip Planner

### 2.2 INTRODUCTION OF UBER APP

Uber is a mobile app taxi program. A car owner with a private car can earn money by verifying that it is a driver of Uber. Passengers can call online or reserve an Uber driver. Passengers must pay for the fare online.Over time, Uber has developed into a variety of ride modes including carpooling. Its price is generally cheaper than a taxi. Therefore, the development of Uber is very rapid.

Figure2: Uber app interface screenshot


Uber app interface


Schedule a Ride Wed, Jul 25 at 10:15 AM-10:30 AM


Schedule a Ride

### 2.3 INTRODUCTION OF TAXICABS IN NEW YORK CITY

Taxi is a common public transportation in the city. Unlike public transportation such as buses and subways, taxi lines are not fixed and random. The density of taxis varies greatly from region to region. It is non-commuting and frequent. The main means of transportation, and its travel path, time, and get-off point information are closely related to human activities, which can better reflect the behavior patterns of urban residents [1].
Taxicab is a very important part of New York City's transportation architecture. They come in two varieties: yellow and green. Taxis painted canary yellow are able to pick up passengers anywhere in the five boroughs. Those painted apple green are allowed to pick up passengers in Upper Manhattan, the Bronx, Brooklyn, Queens (excluding LaGuardia Airport and John F. Kennedy International Airport), and Staten Island. Both types have the same fare structure.

## 3. RESARCH QUESTION

Will the living environment of taxis be affected by Uber? Taxi has many similar aspects to Uber, including their passenger capacity, reachable area and ride comfort. But the price will be more expensive than some Uber types. Uber can only be booked by Uber app in the smartphone. Taxis is a typical offline payment method. It is a very obvious competitive relationship between Uber and taxicabs.

Transit is an app that queries the time of real-time traffic information. The purpose of most of its users is to check the timetable and waiting time of the vehicle that they are going
to take. Users can also use this for travel planning. When they query a route, Transit will give them a travel advice, one of which is Uber. When the user clicks on Uber, the Uber request data is generated in the background. What this paper wants to know is why a user who wants to travel by public transportation first will choose Uber?

Weather, especially extreme weather, has always been a key consideration for residents to choose their transportation options. This paper will analyze the impact of weather on Uber users, taxi users, and users who make Uber requests via the Transit app.

## 4. DATA

The first data set is a sample of the origins of Uber travel in May 2015. The second data set includes an Uber request that was summoned via the Transit app from November 2016 to October 2017. The third data set includes the May 2017 yellow and green taxi trips in New York City. And the weather data set. These data sets are detailed in the following paragraphs.

### 4.1 TLC UBER TRIP RECORD DATA

Uber is considered one of for-hire-vehicles (FHVs) under New York law. All for-hirevehicles (FHVs) drivers in New York City are required to be licensed by the Taxi and Limousine Commission (TLC) and associated with a dispatch base. Uber origin data has been reported to the TLC from the dispatch bases since 2014. Since these data are now public now.it became a candidate for a Freedom of Information Law (FOIL) request. The datajournalism blog FiveThirtyEight.com conducted a FOIL request for the Uber origin data and was rewarded with six months (April through September 2014) worth of Uber origin data by latitude/longitude, and six months (January - June 2015) of Uber origin data by 'taxi zone' (a TLC Specific geographic unit a bit smaller than a zip code). They used this data to examine several claims about Uber use and traffic congestion on their popular blog [2][3]. FiveThirtyEight then made this data accessible to the public along with the FOIL documentation on a GitHub repository in 2015 [4]. Although the data set for transportation is a full year from November 2016 to October 2017. Uber's data location information from 2016 to 2017 is not public. The Uber data here is selected from May 2015 because we can't get the latest data.

### 4.2 TLC taxi Trip Record Data

The yellow and green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab \& Livery Passenger Enhancement Programs (TPEP/LPEP)[5].

The CSV file is too large due to too much dataset information. The computer equipment used by the author could not complete the data reading and analysis for a whole year. Therefore, the analysis of the passenger characteristics of the New York City taxis and the travel mode are based on the data for May 2017.

### 4.3 Transit app Uber data

The data on Transit, which was researched in this paper, all applied for their consent. The main function of the Transit application is to provide users with real-time information about public transportation. However, due to the diversification of user needs and the links and transitions between various modes of transportation, Transit has gradually diversified its services. Users can now query most of the "shared" mobile services, including car sharing (Uber and car2go) and bike sharing.

Users must transfer to the already downloaded Uber app to complete the car booking if they want to choose Uber in Transit app. The Uber request via Transit dataset used in this article is selected from Transit's data from November 1, 2016 to October 31, 2017. At that time, Transit could not directly book Uber. But we can research the Uber request sent by the user. Why would a user who might have chosen to travel on public transport choose Uber from a platform that provides information on public transportation services? This is the question we need to find answer.

Uber request via Transit data includes user ID, device ID, pickup time, pickup latitude, pickup longitude, uber type and whether the user has installed the Uber app. The Transit app works based on the principle of GPS positioning. It accurately locates the location of the smart mobile device and returns nearby bus route information, timetables and real-time information. It also searches for nearby Uber vehicles and returns to their location, the waiting time the user needs to spend and the approximate price required to reach the destination.

The Transit app dataset analyzed for this paper contains records of user interactions with the software interface and is considered a form of data exhaust [6].

The user interactions in the dataset include opening an application, finding a destination or querying a route, show nearby lines, clicking on a tile to get more information, or retaining a shared mobile service. To provide users with real-time traffic information, location information for each device, and related applications. While data sets are very detailed in terms of activity and data queries, it is geared towards the operation of the smartphone application, and does not reveal the user's identity information. Names or demographic variables are not requested nor stored, which protects the anonymity of users [7].
Since the object of this research was unified into New York City, the Transit dataset screening removed cities that were not New York City.

### 4.4 WEATHER DATA

The weather data set comes from National Climatic Data Center. Weather in New York central Park is chosen to represent the weather in New York City. The data set contains information such as precipitation, maximum temperature, minimum temperature, wind speed,etc.

## 5. DATA DESCRIPTIVE STATISTICS \& ANALYSIS

The statistics and analysis of the data set will be divided into two parts, the first is the data set from Transit. The Transit dataset contains data for the entire year from November 1, 2016 to October 31, 2017. Focus on analysis and statistics of Uber requests in transit data. Analyze the travel characteristics of the user. The second part is analysis and comparison of Uber, Taxi and Uber request via transit app. In order to make it comparative, the three sets of data will use one month of data uniformly. Uber raw data is selected from May 2015, New York taxi data is selected from May 2017, and Transit data set is selected from May 2015(compare to Uber data) and May 2017(compare to Taxi). Visualize the three parts of the data, compare the user's usage time and location distribution, and observe the impact of weather changes on the number of users.

### 5.1 UBER REQUEST ANALYSIS (2016-11~2017-10)

### 5.1.1frequency of Uber request via Transit

Figure3 : Frequency statistics of Uber request via Transit


Uber_request frequency(1~10times)


Although some users use Transit to find or book Uber over 50 times in a year, it can be seen from the figure that the proportion of users who have only used it once is very large. This means that most of the Uber request senders will only choose to use Transit to book Uber once.
The reasons for this phenomenon are as follows:
First of all. Back to the Transit app itself, the main function of this app design is to provide public transportation information services. The vast majority of users download or use the app for the purpose of querying routes and public transit real-time information. Only when they can't find a suitable public transportation mode or they don't want to spend time waiting, they may turn to Uber after thinking.
Second. Since the user cannot use the Transit application to directly subscribe to Uber within the time range counted in this dataset. The users who want to use Uber will open the Uber app directly instead of opening the Transit app next time after they are told that they can't book Uber directly from the Transit app.

### 5.1.2 Uber type statistics

figure4:Uber type selected by Transit user

figure5 : Statistics of Uber app installation


Obviously, from figure4, most Transit users have chosen UberX when they choose to use Uber. The possible reasons are as follows: First, UberX is the most widely distributed type of Uber. Second, the UberX has a capacity of 1 to 4 people and the fare is lower than others. For Transit users, most of them are originally planned to travel by public transport, so the user's travel cost budget is low fares such as public transport fares, and the number of people traveling together is generally no more than 4 people.

Figure 5: More than half of the users did not install Uber app. Which means that there is no or very little Uber travel in the travel history of these users. Users who have installed the Uber app do not choose to open Uber directly, but use Transit first. This shows that the user originally wanted to choose public transportation. But they finally chose Uber after comparing the waiting time, arrival time, ride comfort and other factors.

### 5.1.3 User request time trend

Figure6 : pickup time trend by month uber request by month



This is the time distribution heat map of the user who issued the Uber request from November 1, 2016 to October 31, 2017.According to the figure, the number of users in November is much higher than other times. In December, as the temperature dropped and the holiday arrived, the number of users also decreased.

Figure7 : pickup time trend by hour

## uber request by hour



As can be seen from the figure7, after 6 am, the user distribution is more uniform. However, it can be seen that the number during the morning peak period ( 7 am to 9 am ) and the evening peak period ( 4 pm to 6 pm ) is higher than other times.

### 5.1.4 UBER REQUEST \& WEATHER

Figure8 : Uber request \& Weather(precipitation)


The pink histogram above is the number of Uber requests per day, and the blue polyline is the amount of precipitation per day. Show the two in a picture and analyze the trend of the Uber request as the weather changes. The data for this picture was selected from the first five months of 2017. Because there is too much data for a whole year. It is difficult to see all the inputs into the picture. Therefore, after analysis and comparison, a representative five months was selected.

As can be seen from the above picture, on the day of the rainy weather, the number of Uber requests is significantly higher than the two days before and after. However, if the rainfall is too large, such as the 127th day in the picture, Uber request will not be as big as expected, probably because the user will give up the trip because of the bad weather. As a result, the overall number of users will decrease.

### 5.2 Uber \& Taxi \& Uber request via Transit Analysis and Comparison

The data in this part is all in one month. New York City taxis and Uber request via Transit in May 2015 and May 2017, and Uber original travel data in May 2015.The data will be analyzed and compared from the following sections :

- Proportion analysis of different methods
- Trip time trend (line chart)
- weather \& trip
- Origins location distribution (Density map)


### 5.2.1 Proportion analysis of different methods

Figure9: The Proportion of usage in various ways within one month


This is the pie chart of the number of users per mode in a month. Users who use Transit to book Uber are very few in comparison of all the data. It is also possible to verify previous assumptions: the most primitive purpose of most users of Transit is to query public transit real-time information or use public transportation. Yellow taxi account for the largest proportion. They are more widely distributed than green taxis. Yellow taxis are still a very important part of transportation choices.

### 5.2.2 Trip time trend by hour

Figure10 : Line chart of different travel mode start time distribution



Uber Weekday Averages for May 2015


Green Taxi Weekday Averages for May 2017



The above is Line chart of start time distribution in May in four different ways. From the above five figures, the following conclusions can be drawn:

- Uber's usage rate is higher than other methods in the middle of the night, because Uber's driver's working time is determined by himself and is more flexible.
- Users who use Transit late at night have noticeably reduced. And its general trend is completely different from Uber and taxi. This also proves the particularity of the Transit user community.
- Uber and taxi are similar traffic patterns to some extent. It can also be seen from the graph that their time distribution trends are roughly similar, especially between yellow taxis and green taxis.
- Uber and taxi travel data shows that Uber and taxi usage rates at night peaks are higher than early peaks, probably because some users are more tired after work.
- The Uber request in the Transit app does not change regularly over time. This shows that the user has a randomness in using Transit to call Uber, Uber and taxis have a very regular time distribution as well-known travel modes. .
- The taxi pick-up time distribution has a very obvious morning and evening peak period, while the Uber's early peak is not obvious. This is because most of the Uber drivers do not work on this full-time job, and the working hours are more flexible. Most drivers May not be willing to get up early.


### 5.2.3 weather and trip

Figure 11 : Trip \& Weather


The above is the combination of the quantitative statistics bar graph of the four methods in one month and the rainfall trend line graph of this month. As can be seen from the above picture: On most rainy days, there will be obvious or not very obvious transition trends. This trend is an increase compared to the previous two days. This shows that the general bad weather will increase the chances of citizens taking Uber and taxi. Among them, the number of yellow taxis is the most evenly distributed, which also proves that it occupies an important part of the transportation structure of New York City.

### 5.2.4. Origins location distribution

## Ask a question:

As the city's main transportation tool, there is a competitive relationship between them. What are the characteristics of different modes of transportation in space? Which way is the user in the airport area more inclined? Is the spatial distribution of vehicles in the morning and evening peaks the same? What is the relationship between the starting point of users in various ways and the distribution of subway lines in New York City? Will the Uber
request from Transit (the timetable that most users use to query the bus or subway) be closer to the subway line distribution than other methods?

## Solution:

The location of all user starting points is reflected on the map of New York City, and the spatial distribution of the different modes is analyzed by using the color depth of the density map.

## Work preparation:

Due to the need to protect user privacy issues, TLC's data does not have accurate latitude and longitude positioning, but instead divides New York City into blocks. The user's location is displayed as a block number. Therefore, the author uses QGIS to take the central position of each block as the latitude and longitude coordinates of the block, and counts the number of points in each block, and uses this as the density of the position.

## Tools used:

QGIS, python。

## Origins distribution density map: :

The location distribution density map has the following categories:

- Uber Request from November 2016 to October 2017, Uber Request in May 2015, Uber Request in May 2017, Uber Raw Data in May 2015, Lease in May 2017 Car raw data
- The comparison of working days and weekends within one month
- The comparison of working days' morning and evening peak
- The comparison of weekends' morning and evening peak

This article will analyze the maps with representative features, and all the other maps are attached.

Figure 12: Density of Uber request origins from the Transit for 2016-11 to 2017-10


图 13: Density of Uber origins from The Transit app for 2015 [8]
Density of Uber Origins
Figure 3. from the Transit App for 2015


Figure 12 is the Uber request density map from Transit from November 2016 to October 2017. Compare it to the picture from the paper--- Interactive Travel Modes: Uber, Transit and Mobility in New York City to find out: The location of users in 2015 was mainly distributed in Lower Manhattan, and in 2017, the location of users in Lower Manhattan was significantly reduced, but concentrated around the subway lines in other areas. The possible reasons are as follows:

- The economy in Lower Manhattan has developed rapidly, and the number of people using Transit in the region has decreased.
- This part of the user has used the Transit app before 2017 and was told that they can't call Uber directly from the software. If they are used to Uber, they will download the Uber app from the app store, so when they When you want to call Uber, you will open Uber directly instead of Transit.
- For specific reasons, we have sent the relevant pictures to Transit and are awaiting their response.

Figure 14: Density of Green taxi origins for May 2017


Due to New York City regulations, the green taxi pick-up area does not include lower Manhattan, and the part circled in the picture belongs to Lower Manhattan. This part still has most of the green taxi pick-up, thus indicating the passenger flow in the area. Very big.

Figure15:Density of Yellow taxi origins for May 2017


Figure16:Density of Uber origins for May 2015


The circled part of the map is the two airports in New York City. The yellow taxi and Uber are the most densely distributed in this area. It also shows that the users of the airport are more inclined to call the yellow taxi and Uber. And the location distribution of Uber and the yellow taxi is similar.

Figure17: Comparison of density of green taxis' am and pm peak for May 2017

> Density of Green Taxi for May 2017 workingday(6am-9am) vs workingday(16pm-20pm)


As shown in Figure 17, the location distribution of all travel modes in the morning and
evening peak hours (see Appendix for the rest) is almost the same, which also shows that the travel distance of users by taxi or Uber or public transportation is not very far.

## 6. Acknowledgements

This project is sponsored by Oak Ridge National Laboratory, Joint Institute for Computational Sciences, University of Tennessee, Knoxville and Changsha University of Science and
Technology.
Thank you very much for the careful guidance and help from Dr. Candace Brakewood, Dr. Cheng Liu and Dr. Kwai Wong.

## Appendix

New York City Zoning Map


Density map:



## Density of Uber Request <br> Origins from the Transit App by Zone for May-2015


$\begin{array}{llllllllllll}0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & k m\end{array}$

## Density of Uber Request Origins from the Transit App

 by Zonefor 2016-11-01~2017-10-31

| Legend |
| :--- |
| $\quad$ |
| Subway Stations |
| subway line |

Count
$\square$ 0-390

$\square$| $390-780$ |
| :--- |
| $\square$ |
| $780-1170$ |
| $1170-1560$ |
| 1560-1950 |
| 1950-2340 |
| $2340-2730$ |
| $2730-3119$ |
| $3119-3509$ |
| $3509-3899$ |
| $3899-4289$ |
| $4289-4679$ |
| $4679-5069$ |
| $5069-5459$ |


$\begin{array}{lllllllllllllllllll}0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & k m\end{array}$

## Density of Green Taxi Origins reported by the TLC by Zone for May-2017

Legend


## Density of Yellow Taxi Origins for May 2017 Weekend(6am-9am) VS Weekend(16pm-20pm)



Density of Uber Origins for May 2015 Weekend(6am-9am) VS Weekend(16pm-20pm)


Density of Uber Request for May 2017 Weekend(6am-9am) VS Weekend(16pm-20pm)


Density of Uber Request for May 2015
Weekend(6am-9am) VS Weekend(16pm-20pm)


Density of Green Taxi Origins for May 2017
Weekend(6am-9am) VS Weekend(16pm-20pm)


Density of Yellow Taxi Origins for May 2017 Workingday VS Weekend


Density of Uber Origins for May 2015 Workingday VS Weekend


Density of Uber Request Origins from Transit for May 2017 Workingday VS Weekend


Density of Uber Request Origins from Transit for May 2015 Workingday VS Weekend


Green Taxi Density of Origins:


Density of Yellow Taxi Origins for May 2017 Workingday(6am-9am) VS Workingday(16pm-20pm)


Density of Uber Origins for May 2015 Workingday(6am-9am) VS Workingday(16pm-20pm)


## Density of Uber Request for May 2015 Workingday(6am-9am) VS Workingday(16pm-20pm)



Density of Uber Request Origins for May 2017 Workingday (6am-9am) vs workingday(16pm-20pm)


Density of Green Taxi for May 2017 workingday(6am-9am) vs workingday(16pm-20pm)


## References

1.CHENG Jing, LIU Jiajun, GAO Yong. Analyzing the Spatio-Temporal Characteristics of Beijing's OD Trip Volume Based on Time Series Clustering Method[J]. Journal of Geoinformation Science , 2016, 18 (9): 1227-1239.
2. Bialik, C., Fischer-Baum, R., \& Mehta, D. (2015, December 9). Is Uber Making NYC Rush- Hour Traffic Worse? Retrieved July 2016, from FiveThirtyEight.com:
3. Fischer-Baum, R. (2015, October 13). Uber Is Taking Millions Of Manhattan Rides Away From Taxis. Retrieved July 2016, from FiveThirtyEight.com:
http://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-360 from-taxis/
4. FiveThirtyEight.com. (2015, September 22). uber-tlc-foil-response. Retrieved July 2016, from GitHub: https://github.com/fivethirtyeight/uber-tlc-foil-response
5.Taxi and Limousine Commission from :http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
6. Transit app. (2015). 'uber-request' table.
7.Davidson, Peters \& Brakewood (November 15, 2016 )Interactive Travel Modes: Uber, Transit and Mobility in New York City.
8. Davidson, Peters and Brakewood (2017). Interactive Travel Modes: Uber, Transit and Mobility in New York City. Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington, DC.

