CIS – 602 PROJECT REPORT
ON
UBER PICK UP'S ANALYSIS

BY: VISHNU VARDHAN KUMAR PALLATI
ID: 01468680

ABSTRACT

Uber Pick up Analysis is a data visualization based project where data is analyzed. This is a private cab service where in each individual owning a car can serve and earn as well for quite a low price. In this report I have considered the uber pickups data and analyzed them for months of April 2014 to September 2014. Compared this data set with other cab service data sets to analyze this data set more and also compared its data for a week day and weekend to observe any patterns in these specific times too. Thus, came up to an analysis showing some amazing facts of this Uber pick up service.

INTRODUCTION

UBER is an American based transportation network company, which develops, markets and operates the Uber mobile app, which allows people with smartphones to add/submit a trip request based on which the Uber drivers routes their car. Uber is evolving the way the world moves by seamlessly connecting riders to riders through apps. This new innovation in the field of transportation has made it easier to access any point of a city for less cost. It also opens more possibilities for riders and more business for drivers. Uber has covered almost all the major cities in the United States and is performing good by making it possible for anyone use a car. In this project, we analyze various data frames that has all the information's related to Uber's pick up date, time, latitude and longitude.

Data visualization is a process to learn and present data to an observer in a way that yields insight and understanding. Today's data visualization tools go beyond the standard charts and graphs used in excel spreadsheet, displaying data in more sophisticated ways such as infographics, dials and gauges, geographic maps, spark lines, heat maps and detailed bar, pie and fever charts. It has become a key vital part in day today's business activities. Almost all the business intelligence software vendors embed data visualization tools into their products, either developing the visualization technology themselves or sourcing it from companies that specialize in data visualization.

In this report, three major analyses have been made in order to know to predict the efficiency of the Uber based on their past data and propose new ways to improve and plan their future operations. The key features taken for analyzes are to know he density or the number of pick-ups every month with respect to the New York City map, know the hourly average number of pick-ups within 24 hours for each

month and compare the performance in each month, and compare Uber's pick up with other major cabs operators in the New York City and graph them to know their outreach and performance.

RELATED WORKS

There are numerous study going on around the world about a revolution in the transportation field. People are working hard to create a transportation module that is most cost and fuel economic. We see numerous videos online promoting the usage of share car drives to and fro office, and sometimes long trips as well. Few reasons for this to become vital in current day situation is to reduce the number of cars on the roads, to save the globe by burning less fuel per day, to reduce the usage of the non-renewable crude oil products, and to facilitate better living by sharing.

As we see in the article "The Public transit should be Uber's New Best friend" by Nate Silver and Reuben Fischer, New York City is the biggest market for the public transit in the country, in fact 40 percent of all public transit trips in the United States occur in New York metro area. This article also talks about the popularity of Uber in this area by mentioning that about 93 million Uber and Taxi trips in New York between April and September 2014. That says it should have a huge amount of data in this period frame itself.

The Article " Uber is Serving New York's outer Boroughs More Than taxis Are" by Carl Bialik, Andrew Flowers, Reuben Fischer and Dhrumil Mehta shows the popularity of Uber in New York and its surroundings. Hey also talk about the availability of huge data generated by Uber users in a short period of time including date, time, coordinates of the pick-ups.

All the above mentioned article supports the importance of interpreting the huge amount of data generated by the Uber users over a time period. It would also help to enhance the service provided by this organization by better sharpening their operations and current procedures. It is also important to know the Uber's strategic impact on the taxi service field which made it unique and top in class.

DESCRIPTION

For Uber pick-ups analysis the data set gathered had files in csv with attributes shown in the below table.

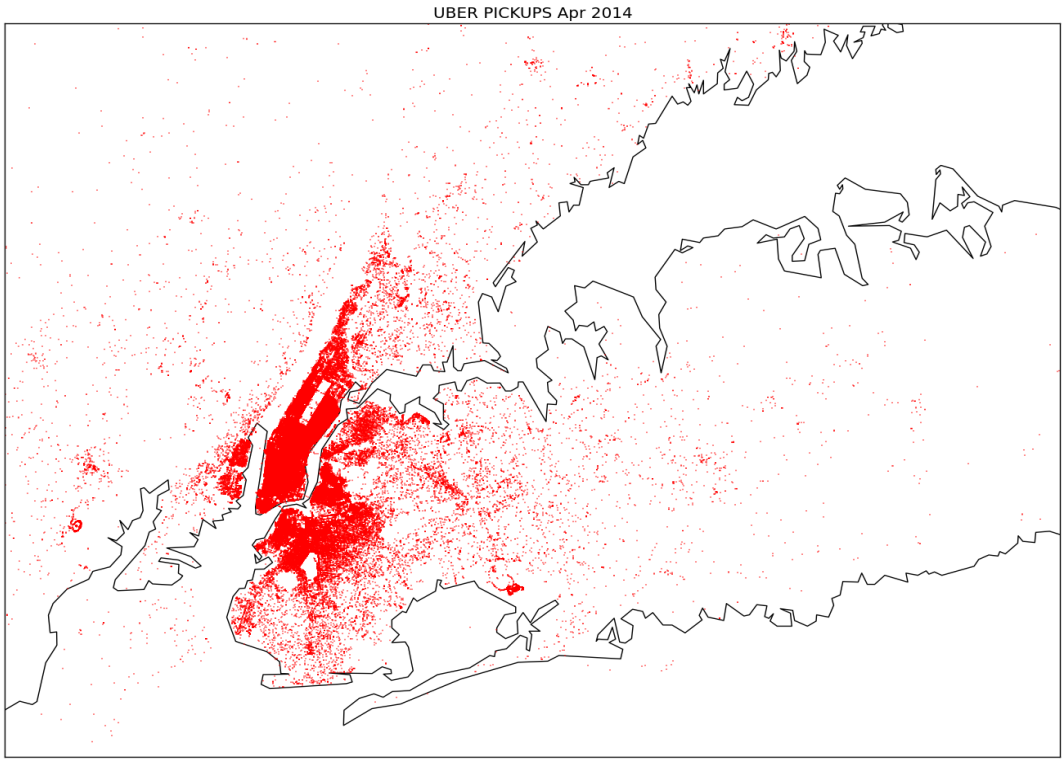
Header	Definition
`Date/Time`	The date and time of the Uber pickup
`Lat`	The latitude of the Uber pickup
`Lon`	The longitude of the Uber pickup
`Base`	The TLC base company code affiliated with the Uber pickup

From the data set after cleaning and grouping the data according to the requirements using pandas library, data frame objects, date time objects I gathered attributes as shown in the figure below.

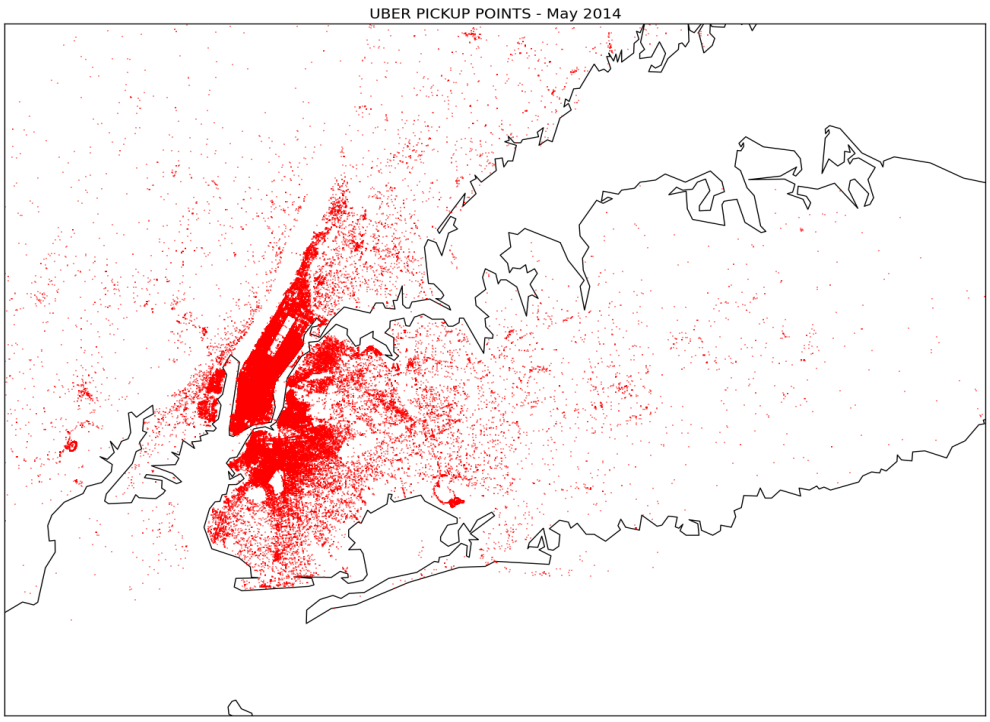
Date Time	Latatitude	Longitude

Through these attribute I could plot the graphs using the Matplotlib library. And the Maps I drew using mpl_toolkit.basemap library. Thus using these libraries and the above attributes I came up with the graphs and maps that I am going to detail about in Results section.

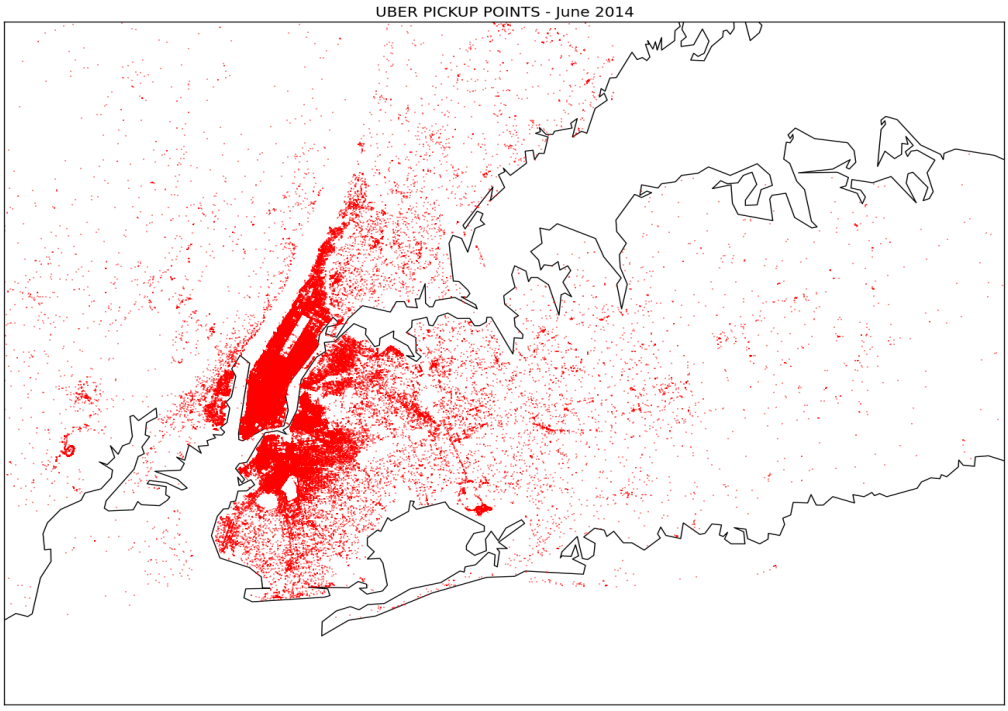
RESULTS



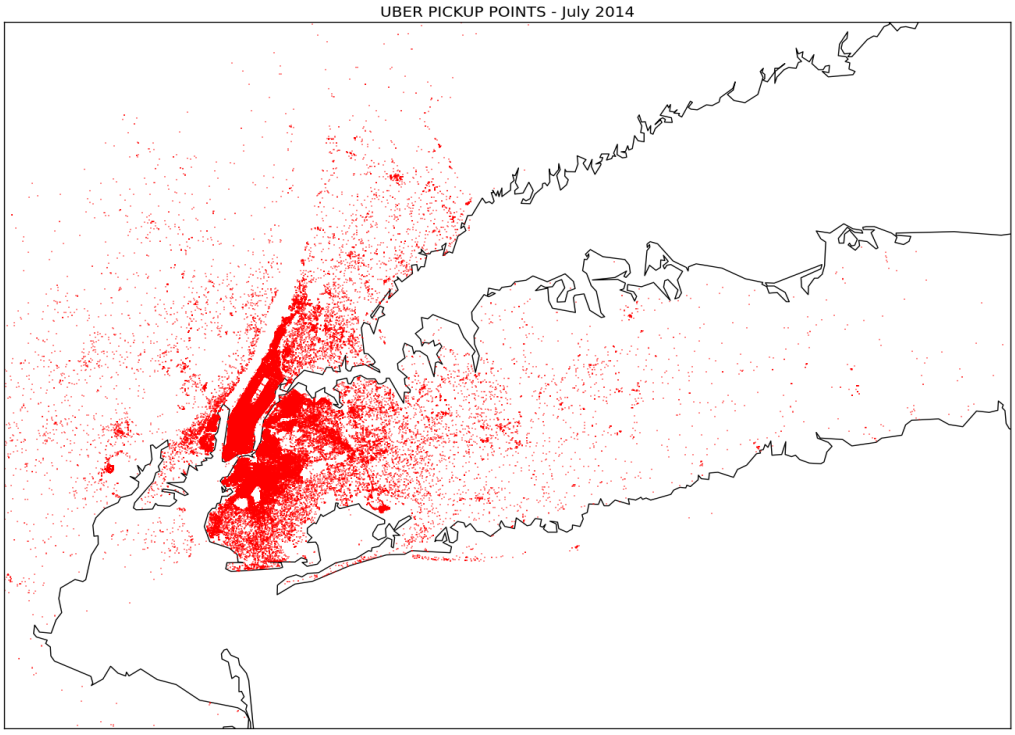
The plot shows pick up points on New York City map. There were around 56 thousand pick up points in April 2014. Each point is plotted on the map. The area with highest density reflects places with more number of Uber pickups. There are no pickup points in the Central park as cars cannot enter the park. The heart of the city is densely populated and the density decreases in outskirts.



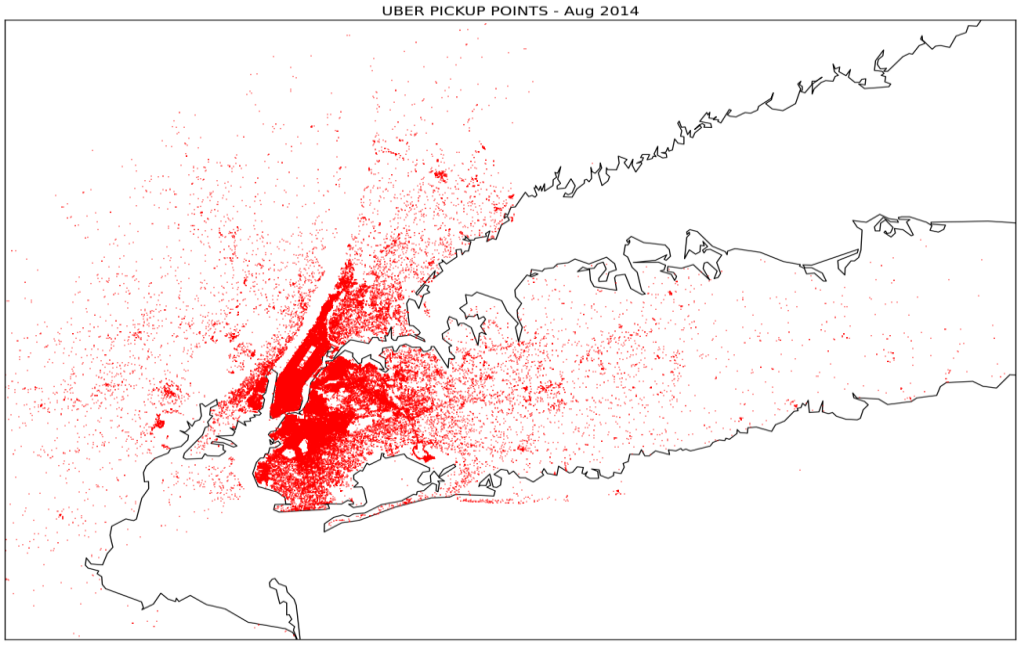
The above graph shows Uber pick-ups for the month May 2014. There were around 65 thousand pick-ups for this month. And these 65 thousand points are plotted on the map. The highly populated areas have many point showing majority of pickups from that place.



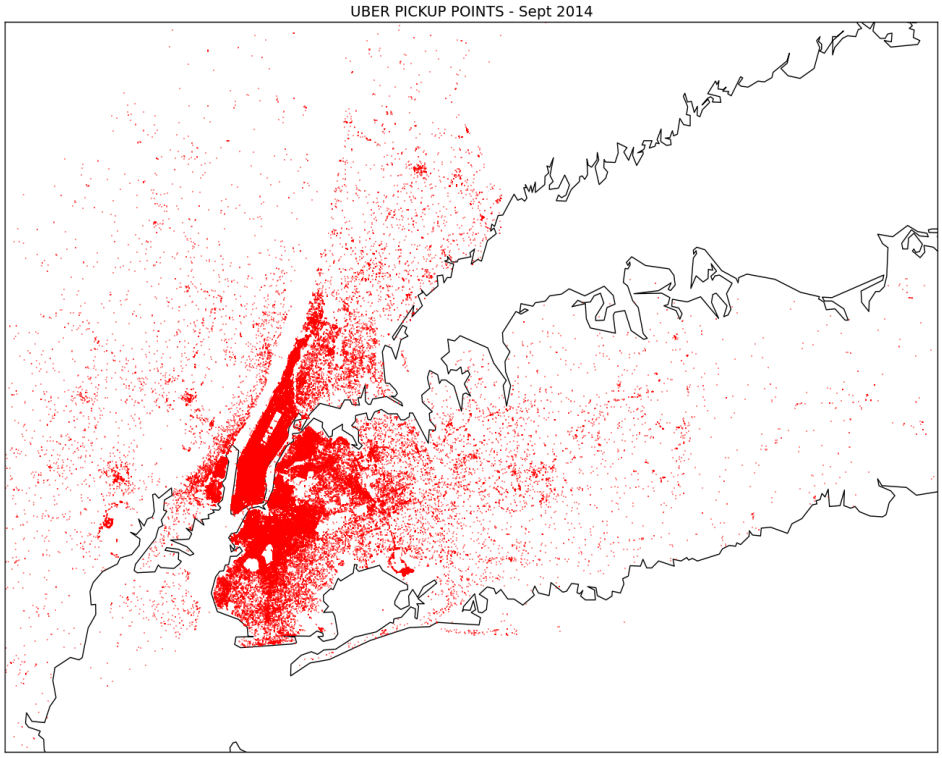
The above graph shows Uber pick-ups for the month June 2014. There were around 66 thousand pick-ups for this month. And these 66 thousand points are plotted on the map.



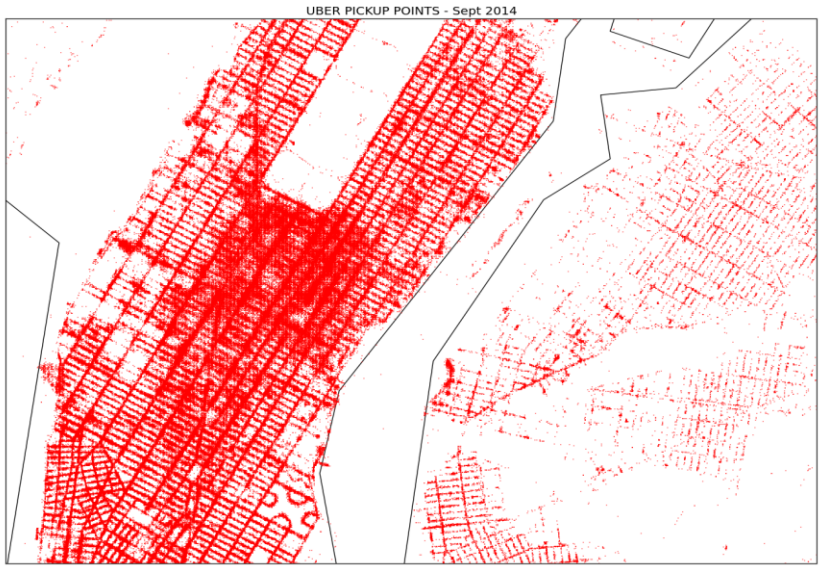
The above graph shows Uber pick-ups for the month July 2014. There were around 79 thousand pick-ups for this month. And these 79 thousand points are plotted on the map.



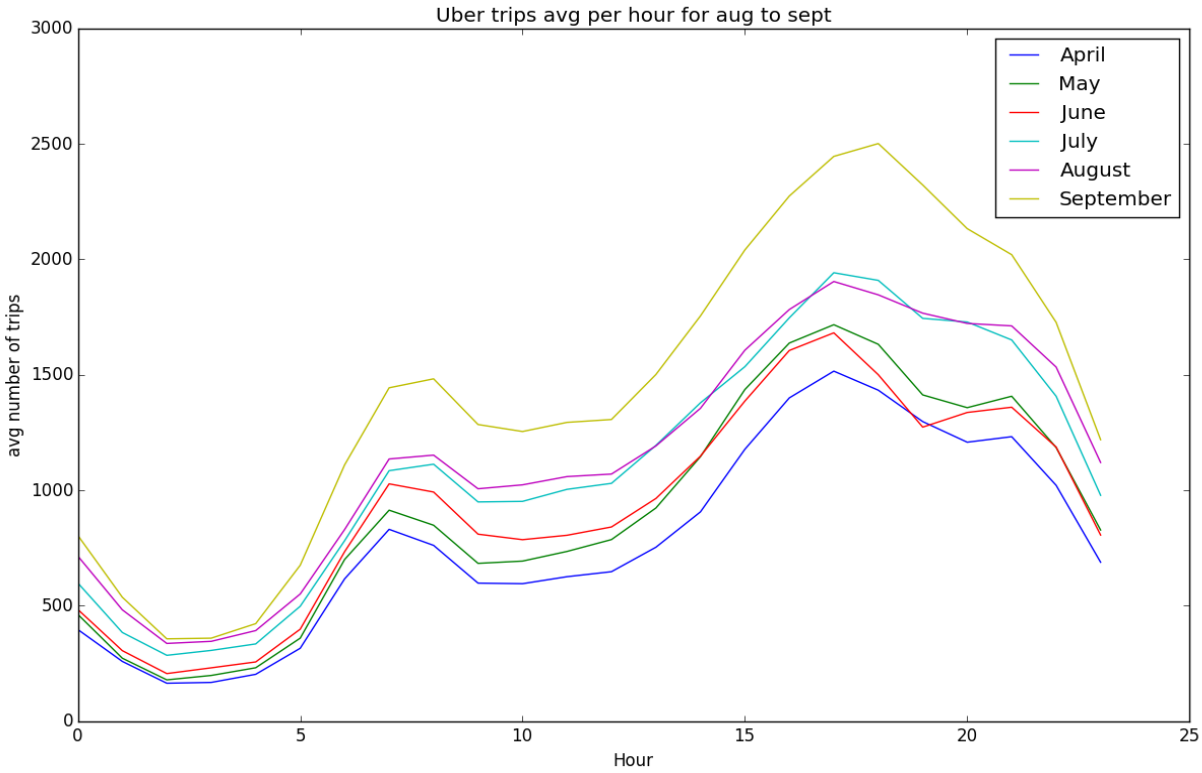
The above graph shows Uber pick-ups for the month August 2014. There were around 82 thousand pick-ups for this month. And these 82 thousand points are plotted on the map.



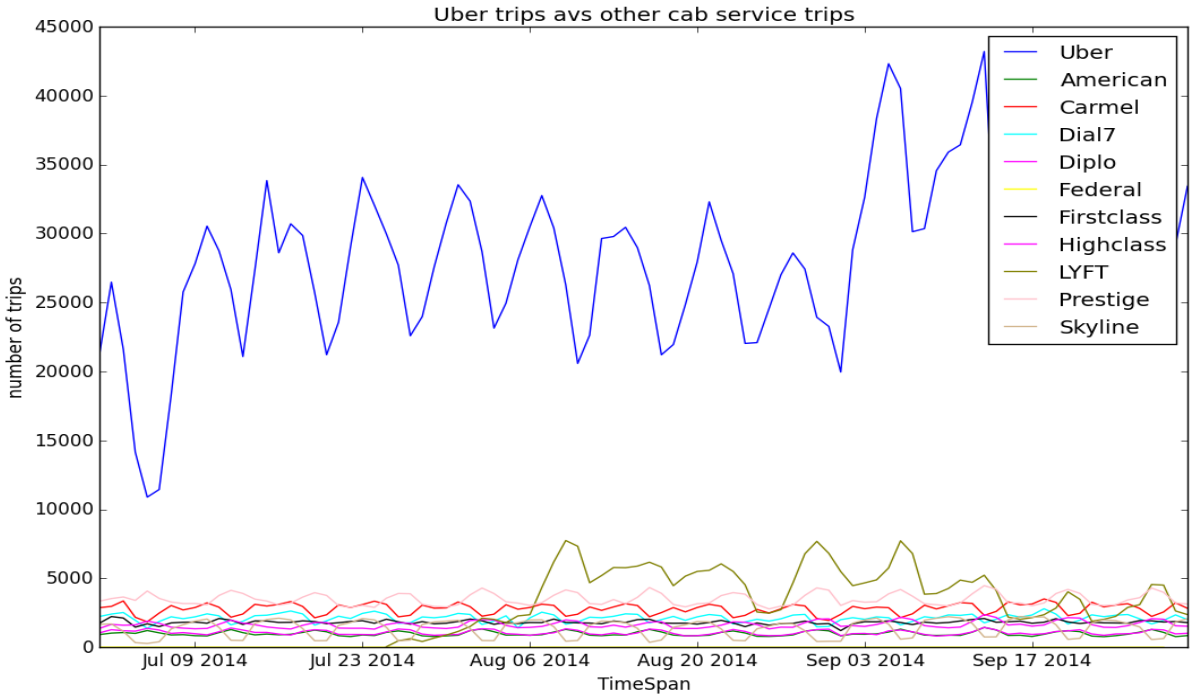
The above graph shows Uber pick-ups for the month September 2014. There were around 102 thousand pick-ups for this month. And these 102 thousand points are plotted on the map. We can observe that for each month the pickups were increasing gradually.



The about map is the street view of NYC near central park. This appears as an art made by thousands of uber users plotting the street view of New York City by their pick-ups.



The above graph show the constant increase in Uber pickups from April 2014 to Sept 2014. Majority of the users use uber at Peak Office hours.

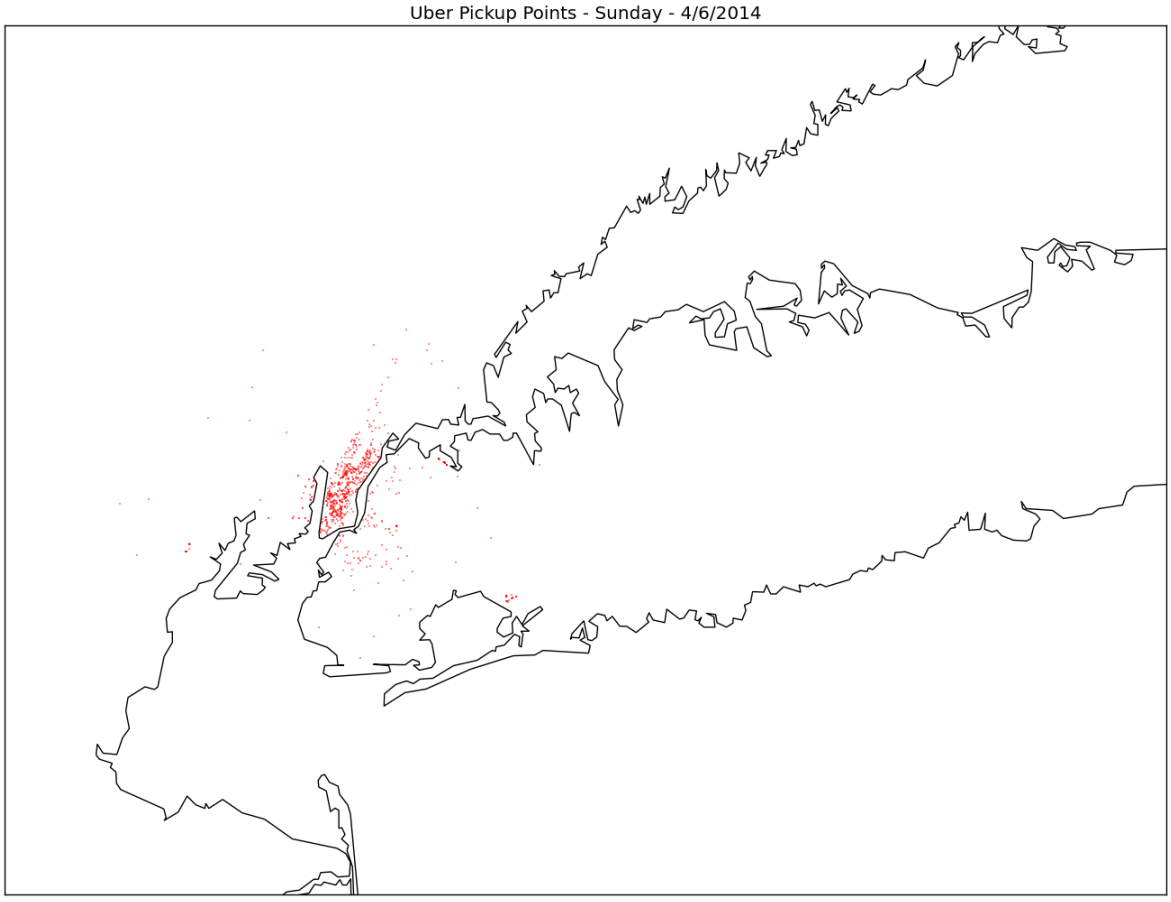


The graph above shows details of Uber cab service in comparison to other cab services. From this graph we can see that uber trips were around 25000 for a period of six months while other cab services where only about 5000. Thus we can

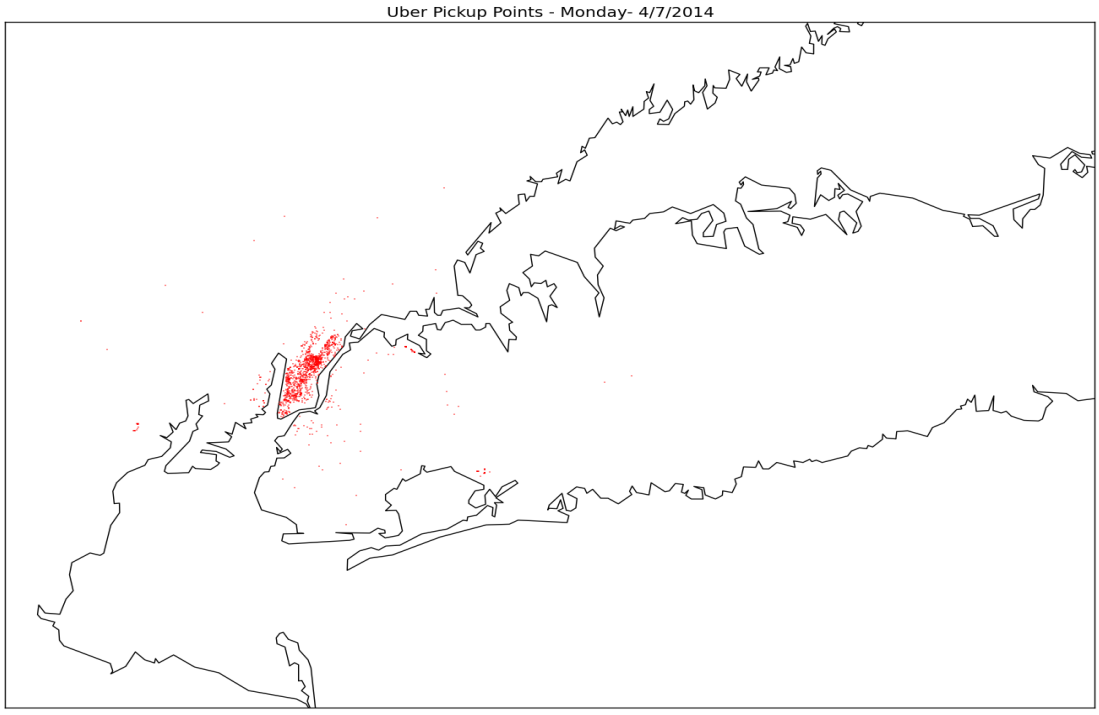
say that Uber has largely conquered the market and its way ahead than its competitors.

In this graph there is also another thing to be noticed that is the graph of UBER has quite a lot of constant ups and down. To know the cause of this I have analyzed it further and realized that this ups and down are due to week days and weekends of every month and thus we can say that Uber is used a lot on week days compared to weekends.

Considering this aspect I also plotted uber pick picks for a random Weekday that is Monday and a week end that is Sunday to show the difference clearly. Below are the two graphs plotted for a random Monday and Sunday from which we can infer that Uber pick ups where more on a weekday than on a weekend.



This is the plot showing a random Sunday being less dense than compared to Monday whose plot is shown below.



DISCUSSION

Let us now discuss about the analysis in this section, the maps I drew for the months of April to September I could analyze the areas where Uber service has been used frequently and the areas where is hasn't been used so much. Uber service is frequently used in comparison to other services. A majority part of the working population of the NYC uses Uber service, this can be inferred as the number of trips during office hours is maximum.

CONCLUSION

Uber Pick up Analysis is a private cab service where in each individual owning a car can serve and earn as well for quite a low price. From the Uber pickups data of months April 2014 to September 2014 while trying to compared the data set with other cab service data sets to analyzing them; from comparing its data for a week day and weekend I could observe few interesting points about this service. Each month the usage of Uber services has been increasing gradually. The map after plots made by thousands of Uber pickups seems to be like an art showing the street view of New York City. Uber has largely conquered the market and its way ahead than its competitors. Uber is used a lot on week days compared to weekends.

REFERENCE

Uber_Proposal.pdf

- http://introtopython.org/visualization_earthquakes.html
- http://pandas.pydata.org/pandas-docs/stable/merging.html
- http://stackoverflow.com/questions/10017876/matplotlib-markers-disappear-when-edgecolor-none
- http://matplotlib.org/1.3.0/examples/pylab_examples/filledmarker_demo.html
- http://pandas.pydata.org/pandas-docs/stable/generated/pandas.to_datetime.html
- http://stackoverflow.com/questions/11391969/how-to-group-pandas-dataframe-entries-by-date-in-a-non-unique-column