PROFESSIONAL TRAINING REPORT

at

Sathyabama Institute of Science and Technology

(**DEEMED TO BE UNIVERSITY**)

Submitted in partial fulfilment of the requirements for the award of

Bachelor of Engineering Degree in

Computer Science and Engineering

By

PUNITH KUMAR M R (Reg. No. 3511449)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,

CHENNAI – 600119, TAMILNADU

APRIL 2018

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

(Established under Section 3 of UGC Act, 1956)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600119

www.sathyabama.ac.in

SCHOOL OF COMPUTING

BONAFIDE CERTIFICATE

This is to certify that this Professional Training Report is the bonafide work of **PUNITH KUMAR M R (Reg. No. 3511449)** who underwent the professional training in "**UDACITY NANO DEGREE DATA ANALYTICS"** under our supervision from January 2018 to April 2018.

Internal Guide

Mr. MURARI DEVAKANNAN KAMALESH, M.E., (Ph.D.),

Head of the Department

Dr. S. MURUGAN, M.E., Ph.D.,

Submitted for Viva voce Examination held on_____

Internal Examiner **External Examiner**

DECLARATION

I, **PUNITH KUMAR M R** (Reg. No. 3511449) hereby declare that the Professional Training Report on **"UBER RIDES DATA ANALYSIS"** done by me under the guidance of **Mr. Murari Devakannan Kamalesh**, **M.E., (Ph.D.),** at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: **SIGNATURE OF THE CANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. M. Lakshmi, M.E., Ph.D.,** Dean, School of Computing and

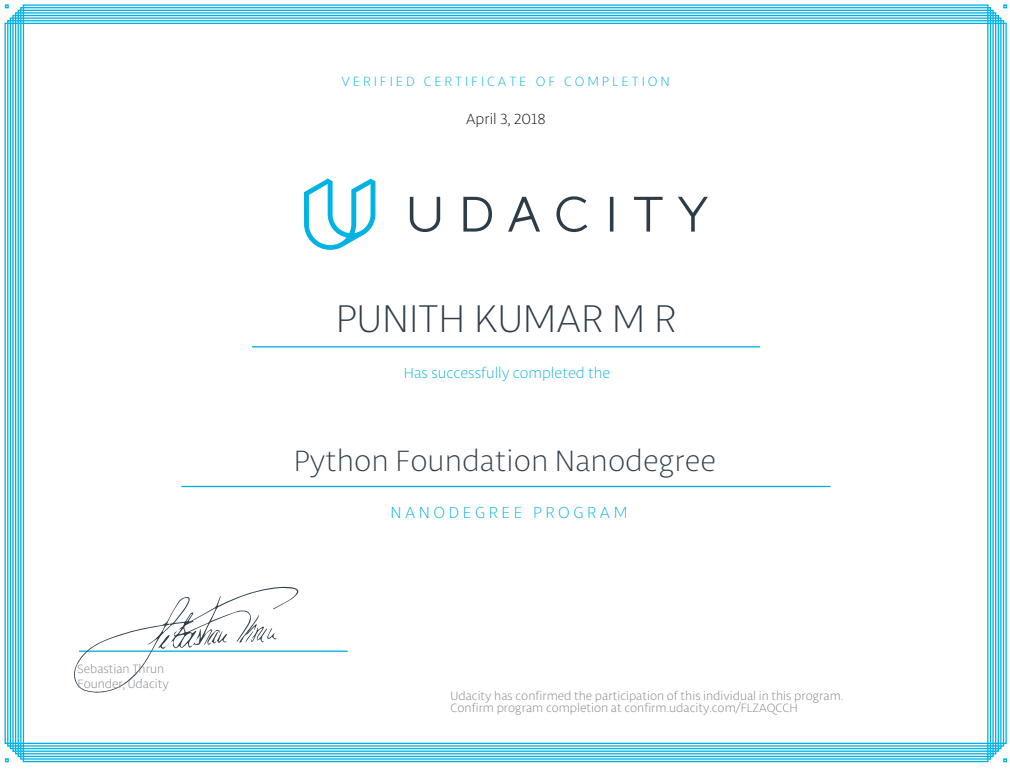
Dr. S. Murugan M.E., Ph.D., Head of the Department, Department of Computer Science and Engineering for providing necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide of **Mr. Murari Devakannan Kamalesh**, **M.E., (Ph.D.),** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of **Computer Science and Engineering** who were helpful in many ways for the completion of the project.

Also, I thank the **Almighty** and my **Parents** for supporting me in the completion of the professional Training.

CERTIFICATE



ABSTRACT

With market competition stiffening, top organizations are turning to data analytics to identify new market opportunities for their services and products. As things stand today, 77% of top organizations consider data analytics a critical component of business performance. Uber Rides Analysis is a data visualization based project where data is analysed. 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 rides data and analysed them for months of April 2014 to September 2014. Compared this data set with other cab service data sets to analyse 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

NYC	-	New York City.
TLC	-	Taxi and Limousine Commission.
EDA	-	Exploratory Data Analysis.
CDA	-	Confirmatory Data Analysis.
CSV	-	Comma Separated Value.
Lat	-	Latitude.
Lon	-	Longitude

CHAPTER 1
INTRODUCTION

This chapter will provide a brief understanding about background of the study, definition of the project justification, risks, project deliverables and project budget and schedule

1.1 BACKGROUND STUDY

Transportation has been proved as the most vital service in large cities. Diverse modes of transportation are accessible. In large cities in the United States and cities around the world, taxi mode of conveyance plays a foremost role and used as the best substitute for the general public use of transportation to get their necessities. For instance, by today in New York there are nearly 50,000 vehicles and 100,000 drivers are existing in NYC Taxi and Limousine Commission. There are many misperceptions in TLC (Taxi and Limousine Commission) of the New York city, that how the taxi services should be disseminated in the city that too based on certain assumptions, like most pickups, time, distance, airport timings. In order to provide very good taxi service and plan for effective integration in city transportation system, it's very important to analyse the demand for taxi transportation. The dataset provides the relating information such as where taxis are used, when taxis are used and factors which tend public to use taxi as divergent to other modes of transportation. In present day transportation dataset contains large quantity of information than the preceding data. For specimen, from the year 2010 TLC using the Global Positioning System(GPS) data type for every taxi trip, including the time and location (latitude and longitude) of the pickup and drop off. In this a complete traffic data which contains nearly 180 million rows of data in the year 2014.

1.2 LITERATURE REVIEW

Anaconda is a free and open source distribution of Python and R programming languages for data science and machine learning. It was released on September 6, 2012. Anaconda was in beta version 0.1 in September 2012, then the stable version 1.0 was released a year later on February 2013. Now the current stable version 5.2.0

was released on May 30 2018. It is available for download on Windows, macOS and Linux distributions. Continuum Analytics, the parent organization responsible for developing Anaconda released a cloud version of their distribution for hardware heavy computations. The data set which we will be using in here is the bike share analysis where we will be operating the US cities which are providing the bike analysis for the further study of the report. In 2012 The cities like Boston, Washington DC, New York, and many more highly developed cities have given study using this model and visualization.

1.2 DATA ANALYSIS

Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

Data mining is a particular data analysis technique that focuses on modelling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing mainly on business information. In statistical applications, data analysis can be divided into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data while CDA focuses on confirming or falsifying existing hypotheses. Predictive analytics focuses on application of statistical models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All of the above are varieties of data analysis.

Data integration is a precursor to data analysis, and data analysis is closely linked to data visualization and data dissemination. The term data analysis is sometimes used as a synonym for data modelling.

CHAPTER 2
AIM AND SCOPE OF PRESENNT INVESTIGATION

2.1 OUTLINE OF THE PROJECT

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 analyse 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.

2.2 SCOPE FOR THE ANALYSIS

Uber Rides Analysis is a data visualization based project where data is analysed. 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 rides data and analysed them for months of April 2014 to September 2014.

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 analyses 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.

2.3 PROBLEM STATEMENT

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 rides data and analysed them for months of April 2014 to September 2014.

The crucial part of the analysis lies beyond the data that, to manage such a huge service like this Uber appoints the data analysist to make investigation of existing data and to suggest the further steps to grow the business in feature. Offering a big deal with their customers and also subscribers.

2.4 OBJECTIVE

Compared this data set with other cab service data sets to analyse 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.

CHAPTER 3
EXPERIMENTAL OR METERIAL METHODS

3.1 METHODOLOGY

Data mining is a particular data analysis technique that focuses on modelling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing mainly on business information. In statistical applications, data analysis can be divided into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data while CDA focuses on confirming or falsifying existing hypotheses. Predictive analytics focuses on application of statistical models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All of the above are varieties of data analysis.

Data integration is a precursor to data analysis, and data analysis is closely linked to data visualization and data dissemination. The term data analysis is sometimes used as a synonym for data modelling.

However, in this project the method that I have used is, **The Exploratory Data Analysis (EDA).** EDA focuses on discovering new features in the data in the present investigation.

3.2 COLLECTION AND DATA WRANGLING

Now it's time to collect and explore our data. In this project, we will focus on the record of Uber rides data from months of April 2014 to September 2014. 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

Table 1: Definition of the data in csv file

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

Table 2: Representation of the data in csv file

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.

3.3 EXPLORATORY DATA ANALYSIS

Now that you have the data collected and wrangled, you're ready to start exploring the data. In this section you will write some code to compute descriptive statistics from the data. You will also be introduced to the matplotlib library to create some basic histograms of the data.

CHAPTER 4
RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS

4.1 RESULTS

While making the analysis we collected the data in geographical format which were plotted below according to Uber pickups in format of .csv extension.

4.1.1 UBER PICKUPS APRIL

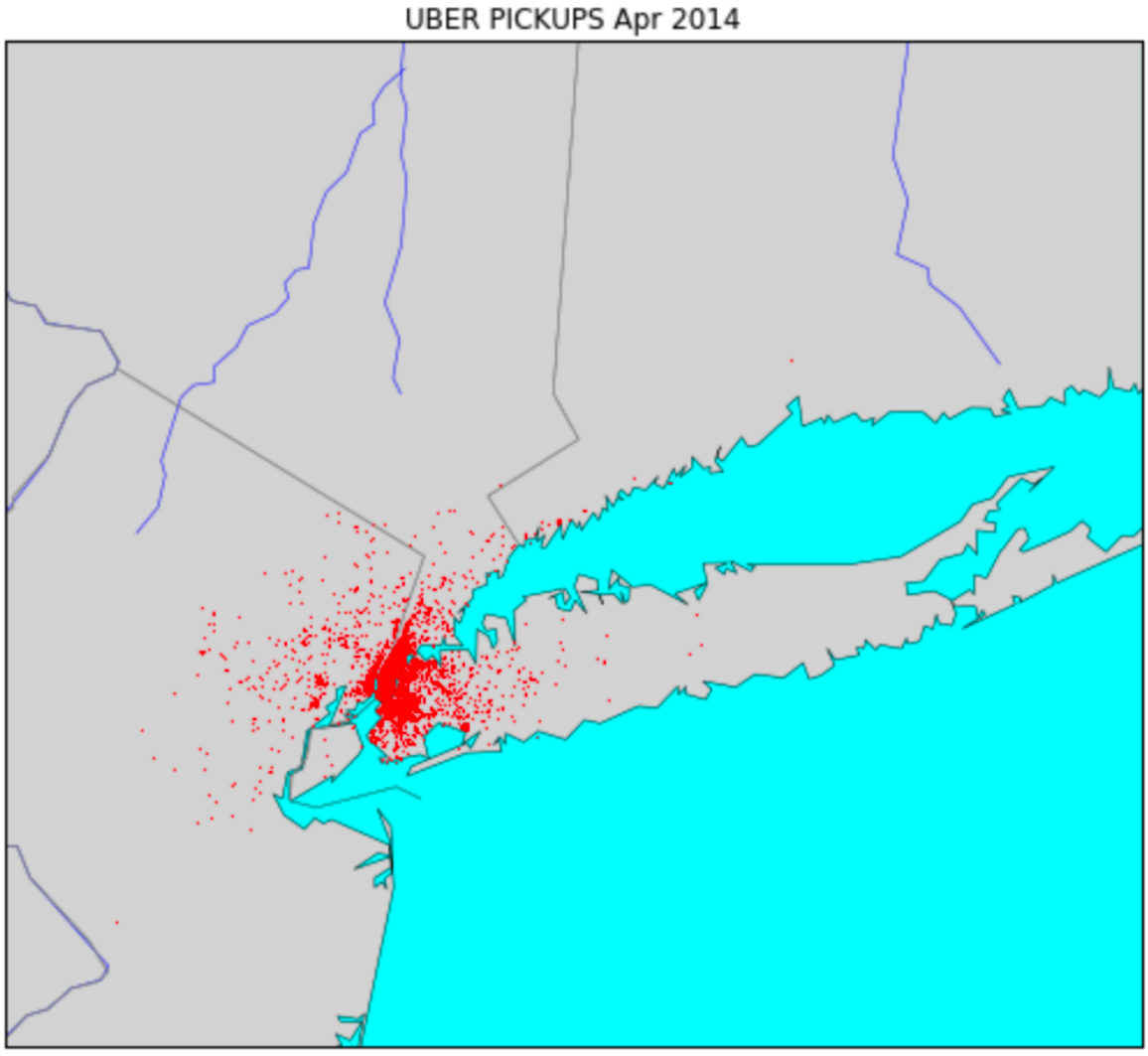


Figure 1: Visualization of the Uber Pickups data of April 2014

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.

4.1.2 UBER PICKUPS MAY

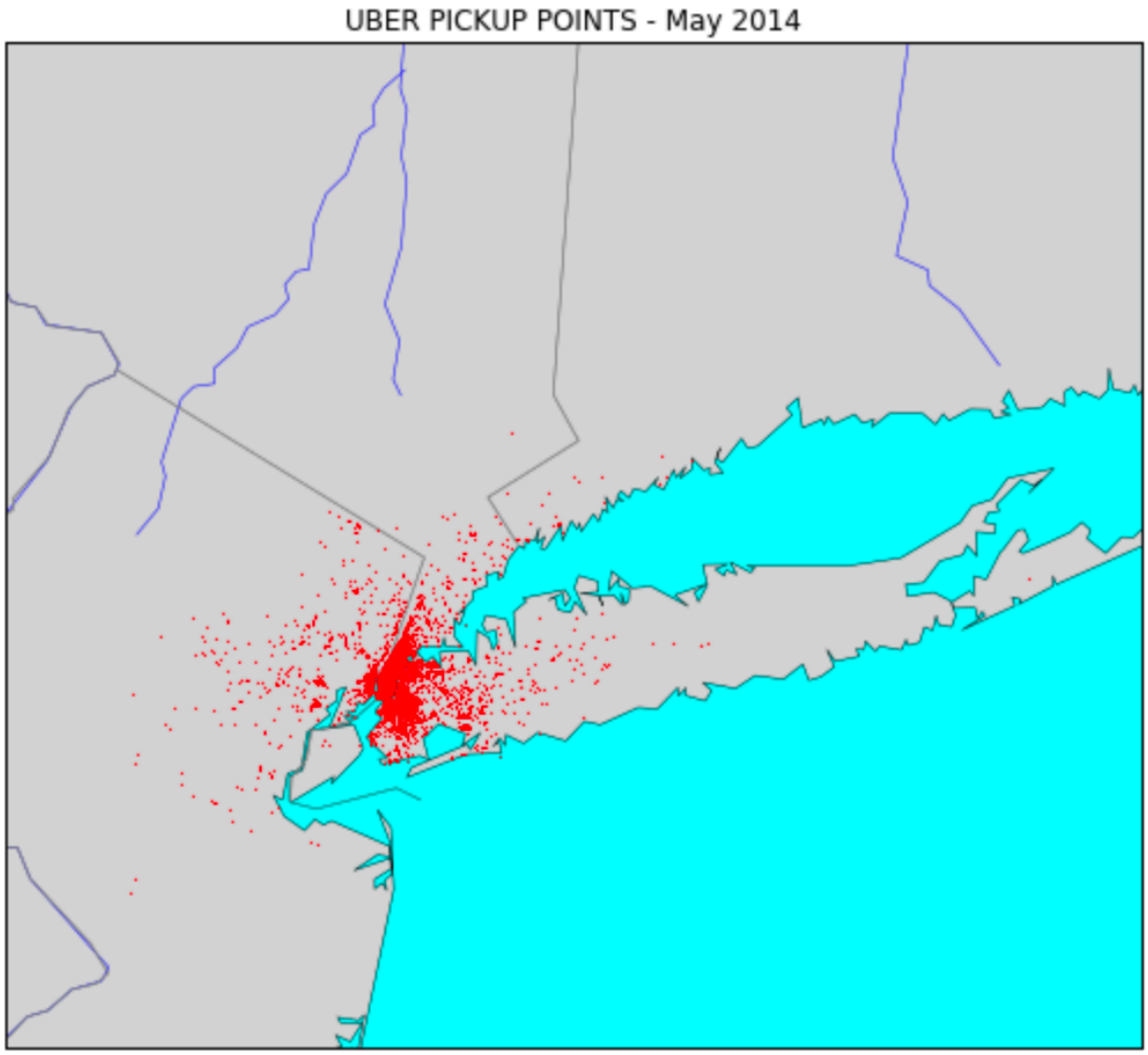


Figure 2: Visualization of the Uber Pickups data of May 2014

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.

4.1.3 UBER PICKUPS JUNE

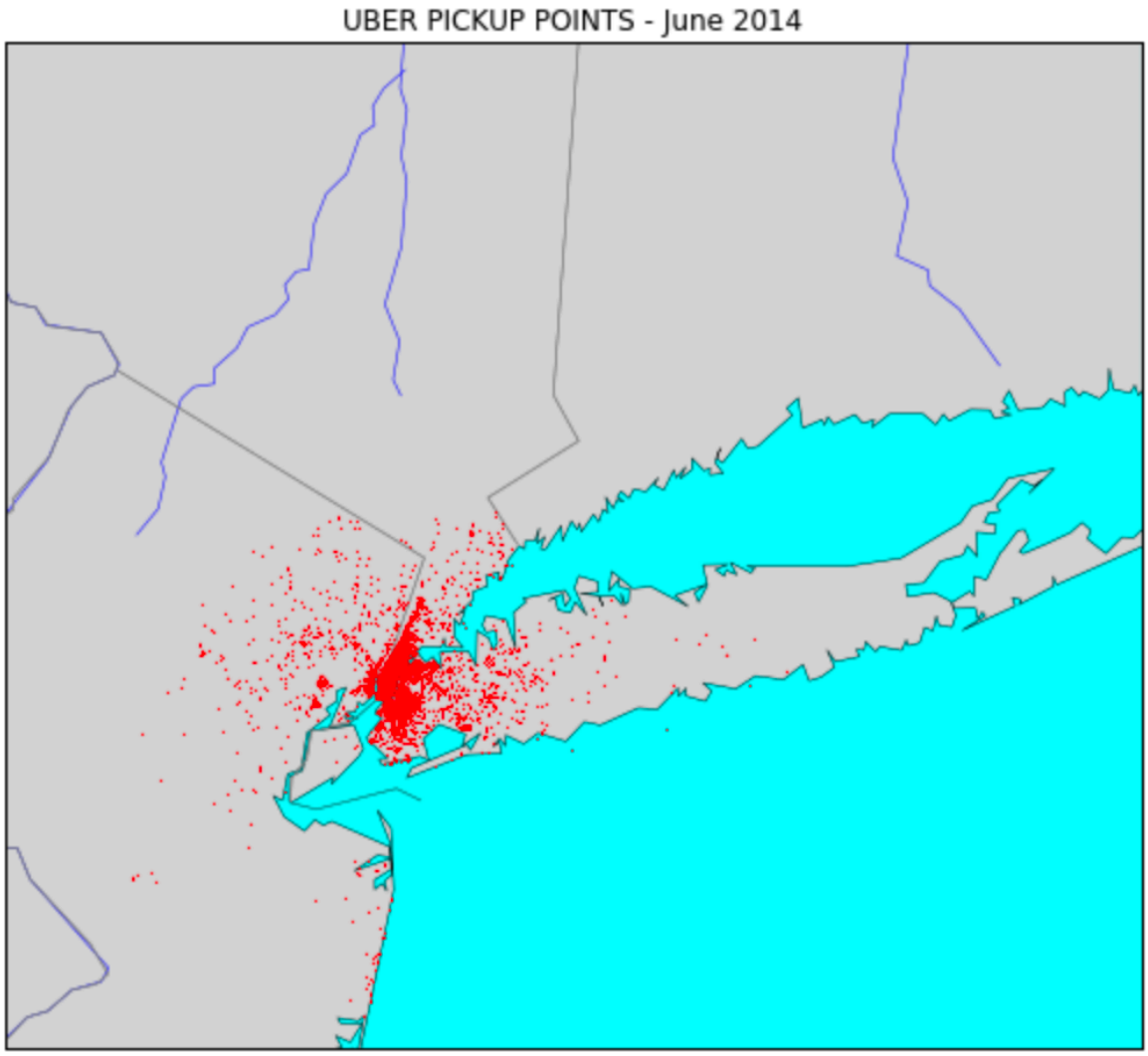


Figure 3: Visualization of the Uber Pickups data of June 2014

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.

4.1.4 UBER PICKUPS JULY

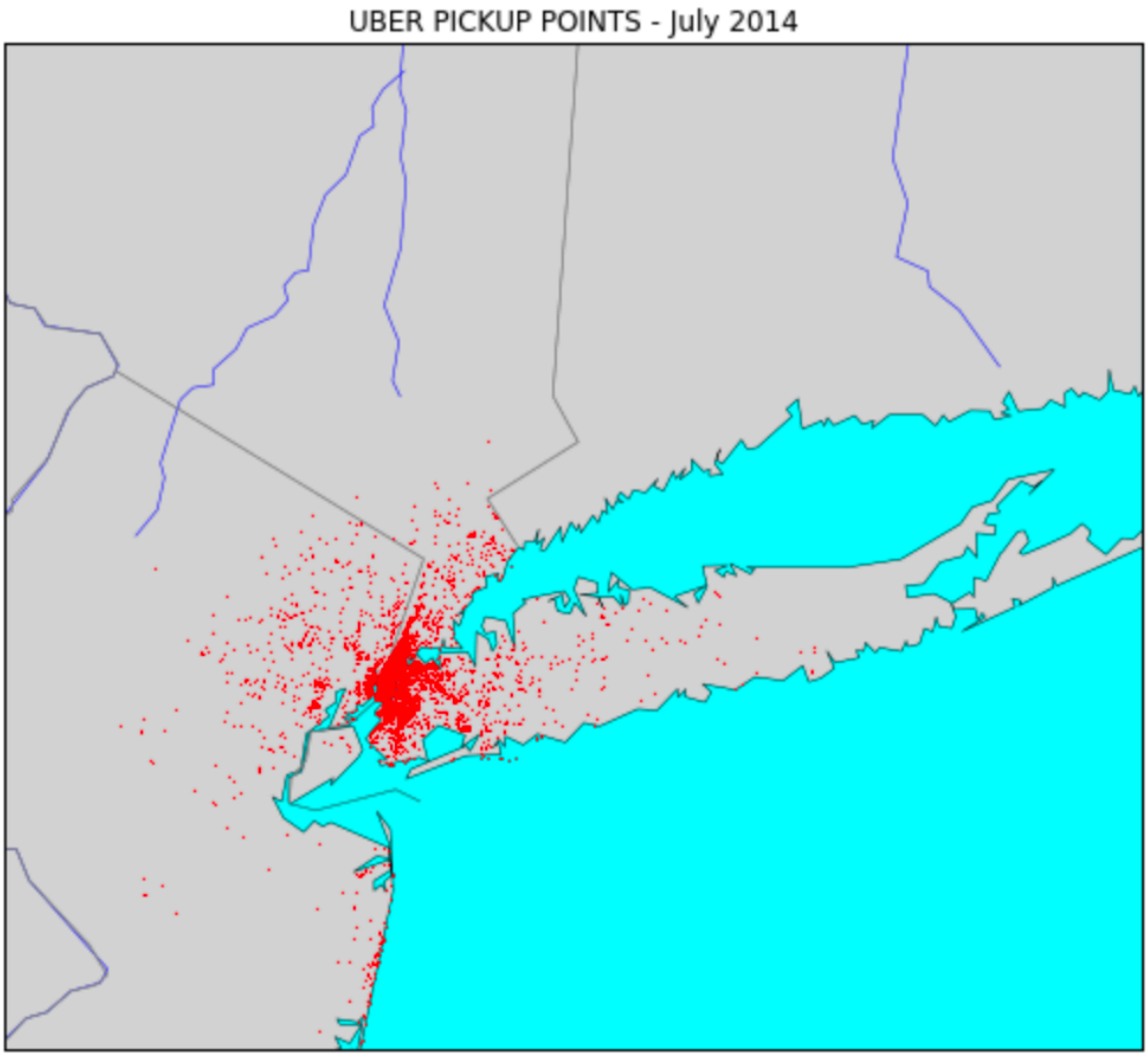


Figure 4: Visualization of the Uber Pickups data of July 2014

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.

4.1.5 UBER PICKUPS AUGUST

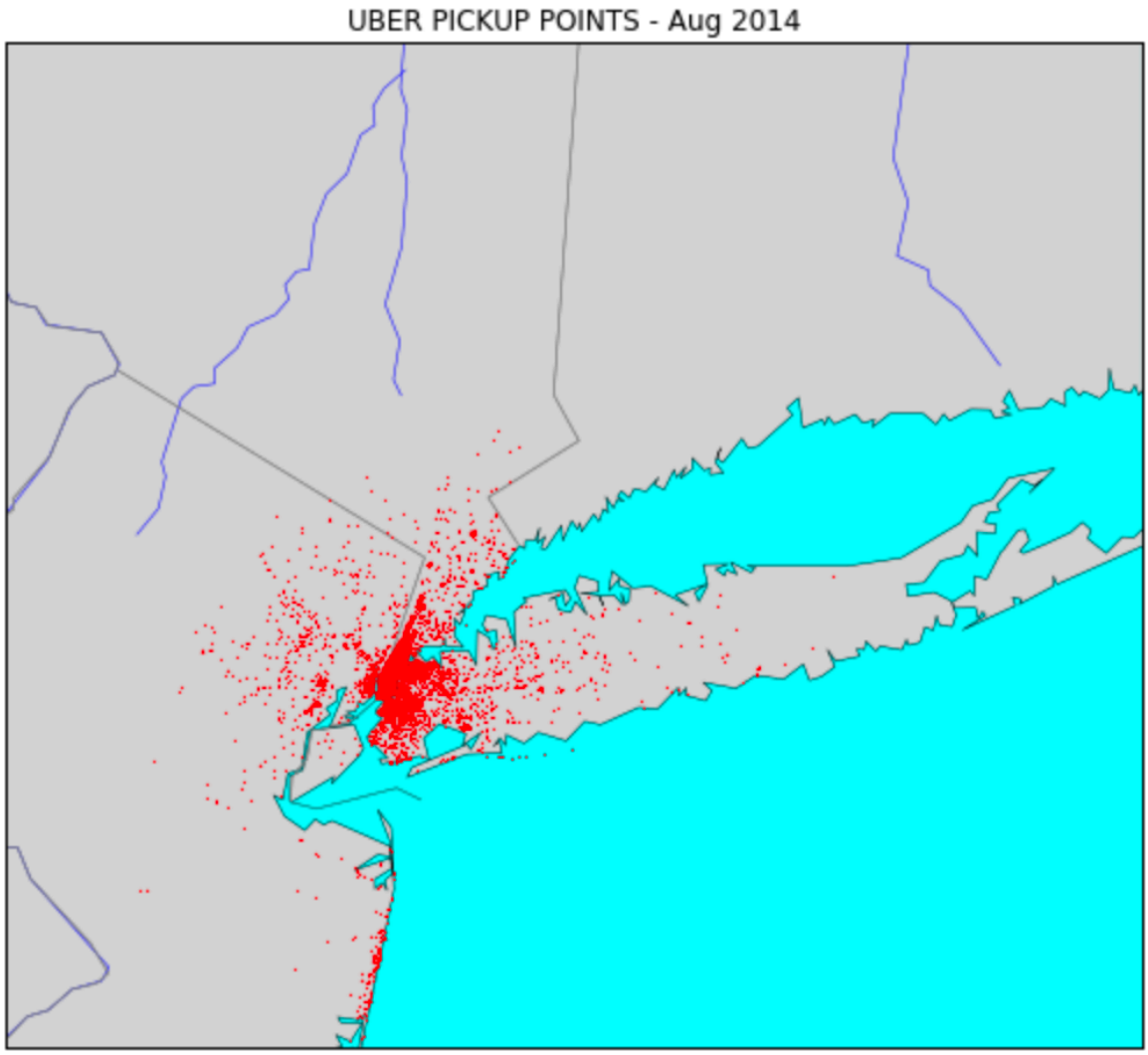


Figure 5: Visualization of the Uber Pickups data of August 2014

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.

4.1.6 UBER PICKUPS SEPTEMBER

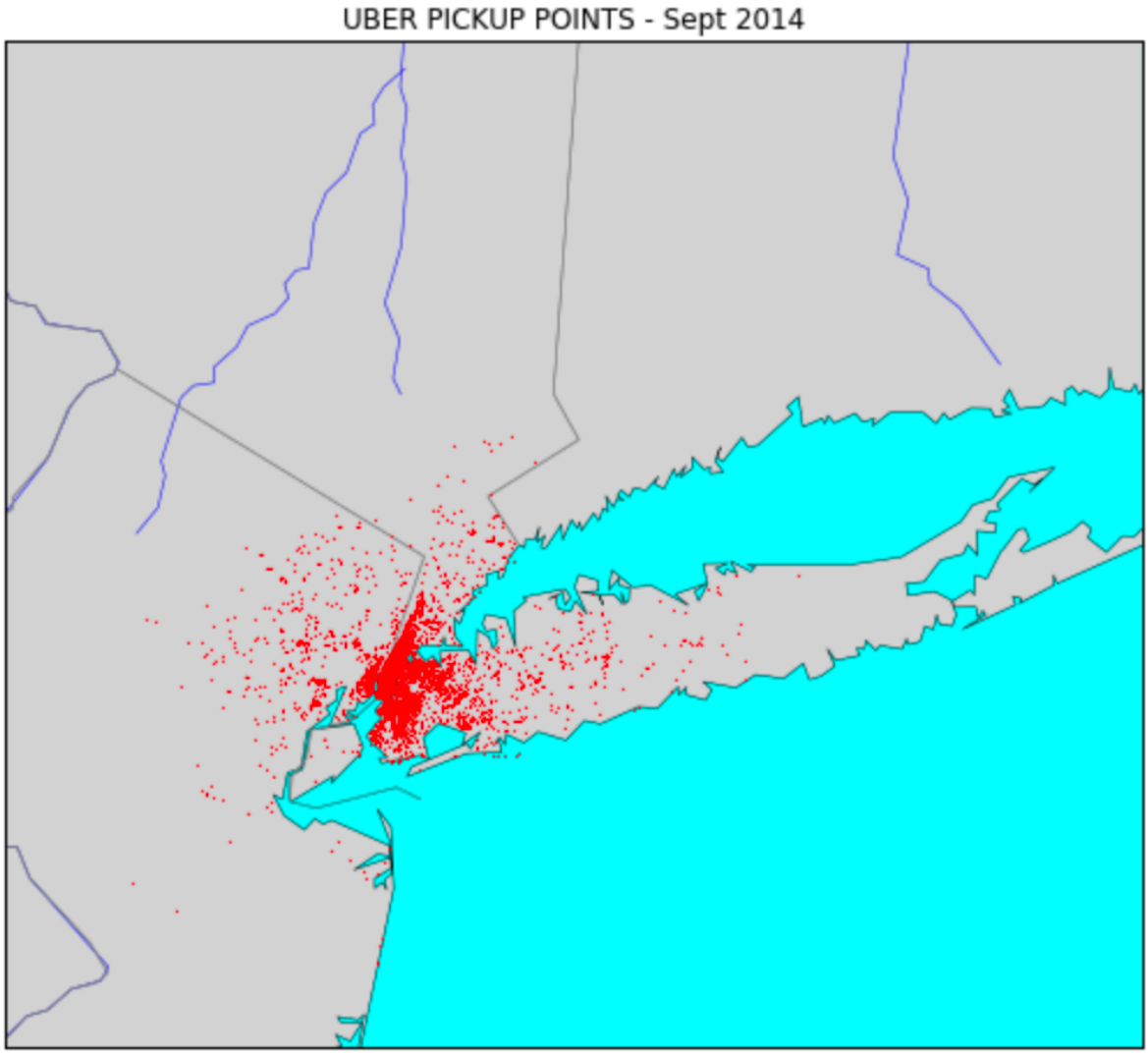


Figure 6: Visualization of the Uber Pickups data of September 2014

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.

4.1.7 UBER PICKUPS NYC STREET VIEW

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.

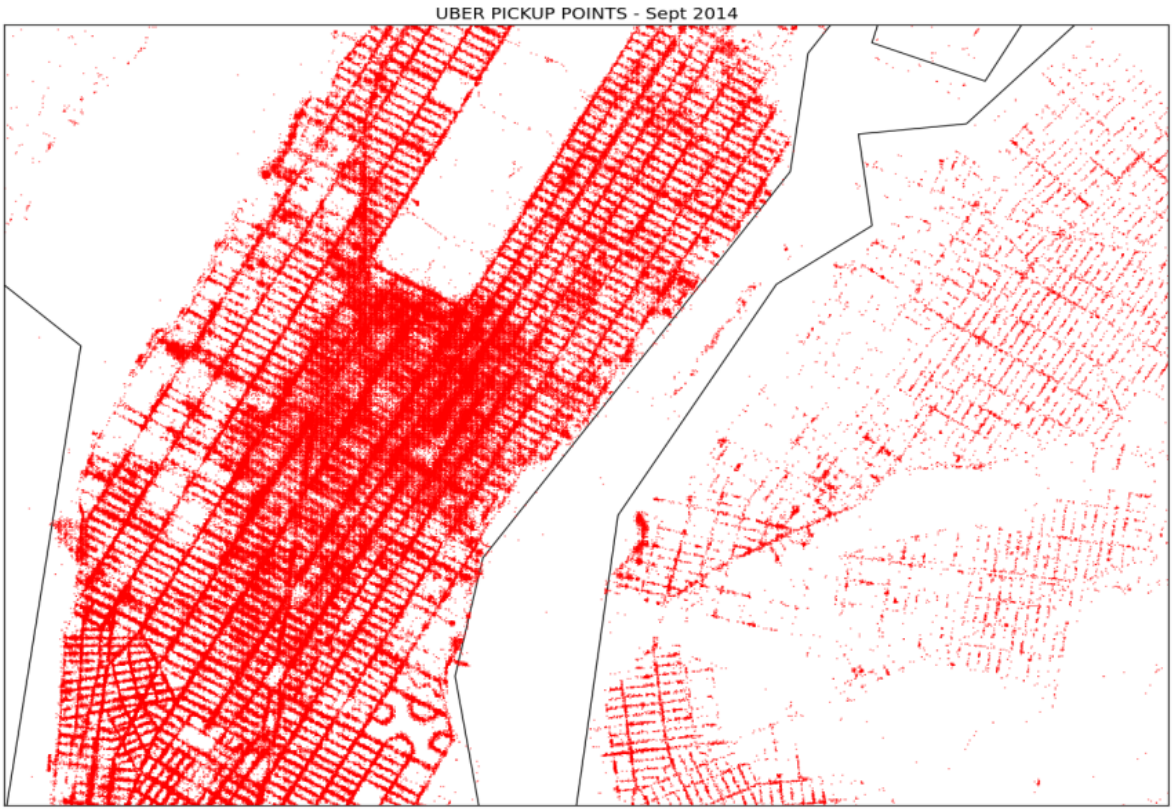


Figure 7: Visualization of the Uber Pickups data in 2014 street view of NYC

4.2 SELF-DIRECTED ANALYSIS

4.2.1 PEAK OFFICE HOURS

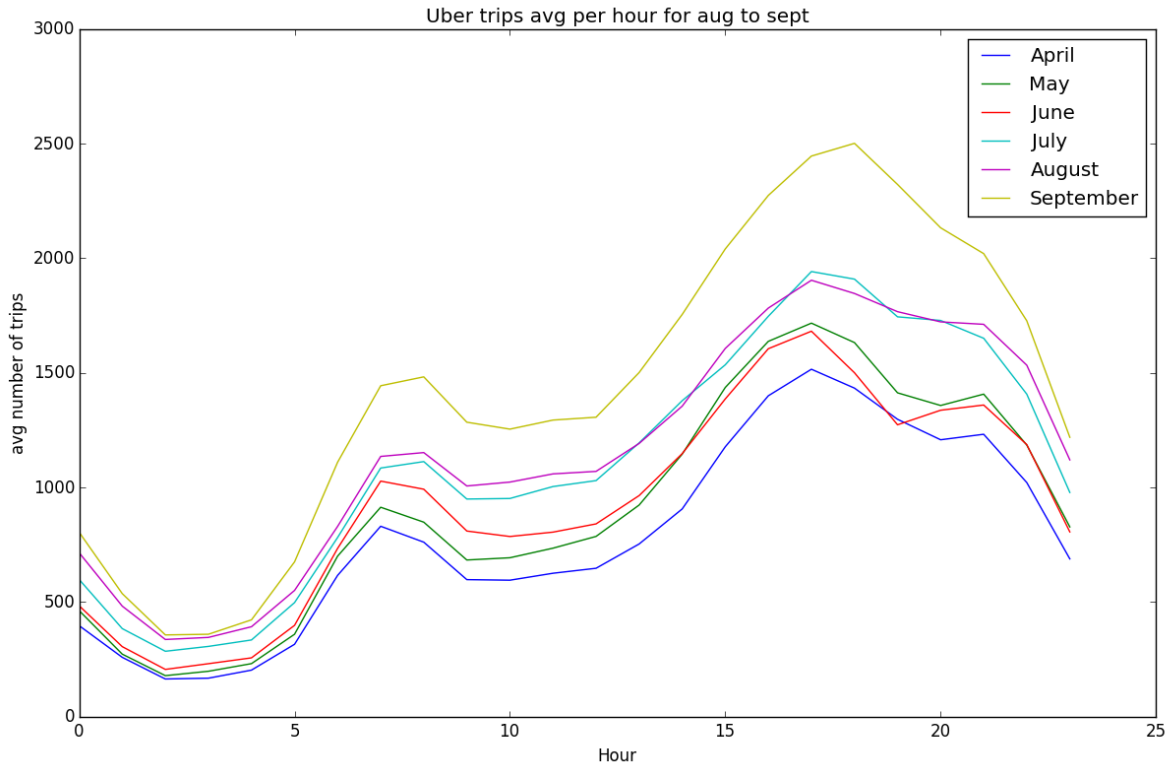


Figure 8: Visualization of the Uber Pickups data at peak hours

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

4.2.2 UBER CAB SERVICE IN COMPARISON TO OTHER CAB SERVICES

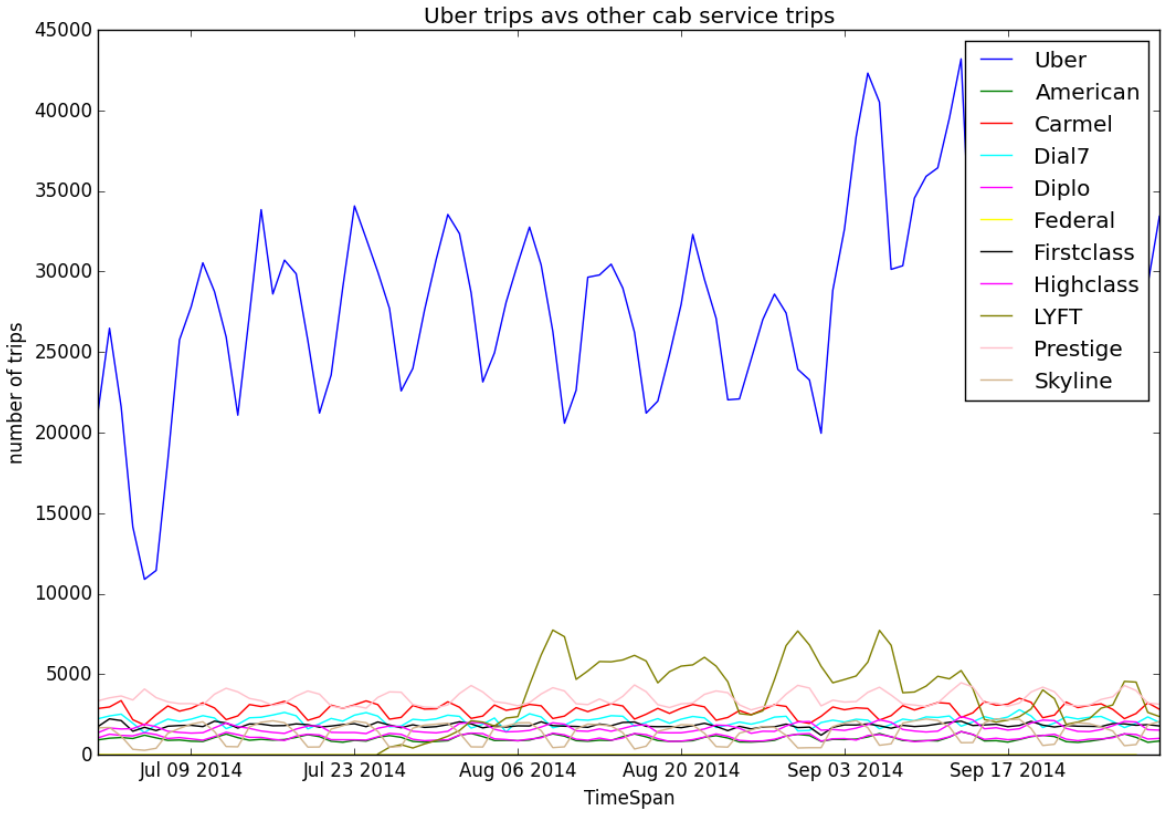


Figure 9: Visualization of comparison with other cab services

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 analysed 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

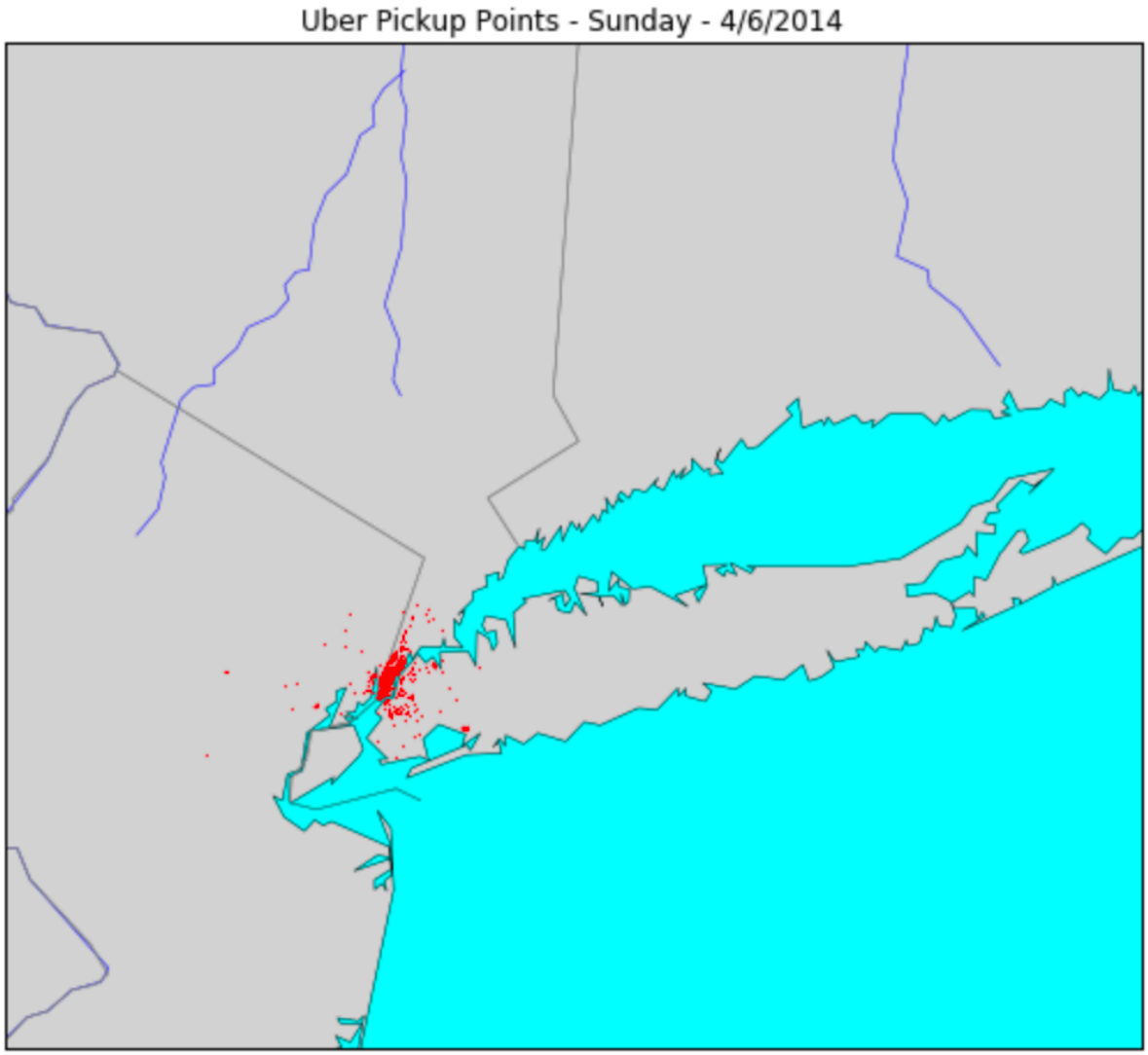


Figure 10: Visualization of the Uber Pickups data at Sundays

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

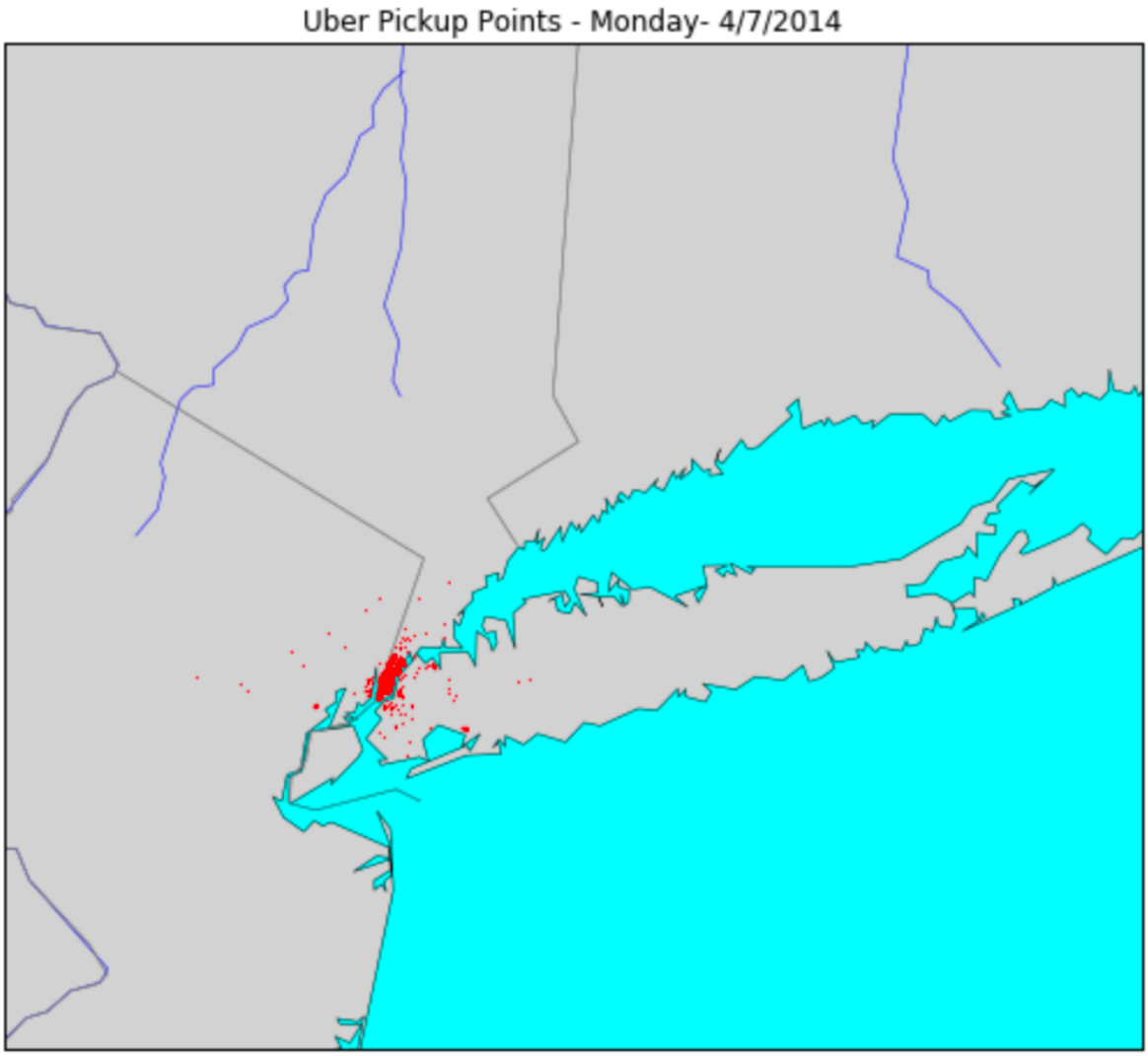


Figure 11: Visualization of the Uber Pickups data at Mondays

4.3 DISCUSSION

Let us now discuss about the analysis in this section, the maps I drew for the months of April to September I could analyse 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.

CHAPTER 5
SUMMARY AND CONCLUSION

5.1 SUMMARY

Let us now discuss about the analysis in this section, the maps I drew for the months of April to September I could analyse the areas where Uber service has been used frequently and the areas where is hasn't been used so much.

The following points are listed below according to the present investigation:

- 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.

5.2 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 analysing 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. The techniques of the data science have become the robust in many vital fields. The techniques are of more precisely business logics oriented. In case one can make many perceptions out of it base on data which you might wanted to make analyse. Whatever the perception can be made it might be small or large we can me use out of it.

REFERENCES

[1] https://www.google.com/

[2] https://www.slideshare.net/

[3] https://www.wikipedia.org/

[4] https://github.com/

[5] https://www.kaggle.com/

[6] https://www.udacity.com/

[7] https://stackoverflow.com/

[8] www.python.org/

[9] https://anaconda.org/anaconda/basemap

[10] Vidyanand Choudhary. Software as a service: Implications for investment in software development. In HICSS ’07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, page 209a, Washington, DC, USA, 2007. IEEE Computer Society.

[11] Rita Dangwal. Public Computing, Computer Literacy and Educational Out- come: Children and Computers in Rural India, pages 59–66. IOS Press, Ams- terdam, The Netherlands, 2005.

[12] I. Foster, Yong Zhao, I. Raicu, and S. Lu. Cloud computing and grid com- puting 360-degree compared. Technical report. Grid Computing Environments Workshop, 2008.

[13] Simson L. Garfinkel. An evaluation of amazon’s grid computing services: Ec2, s3 and sqs. Technical report, 2007.

[14] S.L. Garfinkel. Commodity grid computing with amazon’s S3 and EC2. https://www.usenix.org/publications/login/2007-02/openpdfs/ garfinkel.pdf, 2007.

[15] K.I. Juster. Cloud Computing Can Close the Development Gap. http://www.

salesforce.com/assets/pdf/misc/IT-development-paper.pdf.

[16] Leonard Kleinrock. An internet vision: the invisible global infrastructure. Ad Hoc Networks, 1(1):3 – 11, 2003.

[17] P. Kokil. SAP-LAP Analysis: Gyan Ganga, E-Gram and Communication Information Centers (CIC). http://www.iceg.net/2007/books/3/29_290_3.pdf.

[18] Ralf La¨mmel. Google's mapreduce programming model – revisited. Sci. Comput. Program., 70(1):1–30, 2008.

[19] R. Patra, S. Nedevschi, S. Surana, A. Sheth, L. Subramanian, and E. Brewer. WiLDNet: Design and implementation of high performance wifi based long distance networks. In USENIX NSDI, pages 87–100, 2007.

[20] Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb. A taxonomy and survey of cloud computing systems. Networked Computing and Advanced Information Management, International Conference on, 0:44–51, 2009.

[21] J.E. Smith and R. Nair. An overview of virtual machine architectures. pages 1–20, October 2001. http://www.ece.wisc.edu/∼jes/902/papers/intro.pdf.

APPENDEX

A. SOURCE CODE

```
# Plottin the uber pick up points from Apr 2014 to Sept 2014
# author: Punith Kumar M R
# The required CSV files have to be in the same directory of the python file
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
# http://introtopython.org/visualization_earthquakes.html
#map for apr
data = pd.read_csv("uber-raw-data-apr14.csv", parse_dates=['Date/Time'])
print(data['Lat'])
my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)
my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()
x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('UBER PICKUPS Apr 2014')
plt.show()
#map for may
data = pd.read_csv("uber-raw-data-may14.csv", parse_dates=['Date/Time'])
print(data['Lat'])

my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)
my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()
x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('UBER PICKUP POINTS - May 2014')
plt.show()
#map for June
data = pd.read_csv("uber-raw-data-jun14.csv", parse_dates=['Date/Time'])
print(data['Lat'])
my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)
my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()
x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('UBER PICKUP POINTS - June 2014')
plt.show()
```

```
#map for july
data = pd.read_csv("uber-raw-data-jul14.csv", parse_dates=['Date/Time'])
print(data['Lat'])
my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)
my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()
x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('UBER PICKUP POINTS - July 2014')
plt.show()
#map for aug
data = pd.read_csv("uber-raw-data-aug14.csv", parse_dates=['Date/Time'])
print(data['Lat'])
my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)
my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()
x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('UBER PICKUP POINTS - Aug 2014')
plt.show()
#map for Sep
data = pd.read_csv("uber-raw-data-sep14.csv", parse_dates=['Date/Time'])
print(data['Lat'])
my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)
my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()
x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('UBER PICKUP POINTS - Sept 2014')
plt.show()
```

```
# Plottin the uber pick up points from Apr 2014 to Sept 2014
# author: Punith Kumar M R
# The required CSV files have to be in the same directory of the python file

from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd


#map for Sunday - 4/6/2014
data = pd.read_csv("uber-raw-data-apr14.csv", parse_dates=['Date/Time'])
data = (data[6966:7787])


my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)

my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()


x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('Uber Pickup Points - Sunday - 4/6/2014')
plt.show()



#map for Monday - 4/7/2014
data = pd.read_csv("uber-raw-data-apr14.csv", parse_dates=['Date/Time'])
data = (data[7787:9161])

print(data)
my_map =  Basemap(projection='merc', lat_0 = 57, lon_0 = -130,
    resolution = 'h', area_thresh = 1000, llcrnrlon=-77,llcrnrlat=38,urcrnrlon=-70,urcrnrlat=42,)

my_map.drawcoastlines()
my_map.drawcountries()
my_map.fillcontinents(color = 'white')
my_map.drawmapboundary()


x, y = my_map(data['Lon'].values, data['Lat'].values)
# draw a red dot at cities coordinates
plt.plot(x, y, 'ro', markersize=1, marker='.')
plt.title('Uber Pickup Points - Monday- 4/7/2014')
plt.show()
```

```
# Plottin the uber pick up points from Apr 2014 to Sept 2014
# author: Punith Kumar M R
# The required CSV files have to be in the same directory of the python file
import pandas as pd
import matplotlib.pyplot as plt

result = pd .DataFrame()
#april
data = pd.read_csv("uber-raw-data-apr14.csv", parse_dates=['Date/Time'])
grp = data.groupby(by=[data['Date/Time'].map(lambda x : x.hour)], as_index=False).size()
x = pd.DataFrame(grp)
x.columns = ['count']
x['count'] = x['count'] / 30
#print (x)
result['april'] = x['count']
#may
data = pd.read_csv("uber-raw-data-may14.csv", parse_dates=['Date/Time'])
grp = data.groupby(by=[data['Date/Time'].map(lambda x : x.hour)], as_index=False).size()
x = pd.DataFrame(grp)
x.columns = ['count']
x['count'] = x['count'] / 30
#print (x)
result['may'] = x['count']
#june
data = pd.read_csv("uber-raw-data-jun14.csv", parse_dates=['Date/Time'])
grp = data.groupby(by=[data['Date/Time'].map(lambda x : x.hour)], as_index=False).size()
x = pd.DataFrame(grp)
x.columns = ['count']
x['count'] = x['count'] / 30
#print (x)
result['june'] = x['count']
#july
data = pd.read_csv("uber-raw-data-jul14.csv", parse_dates=['Date/Time'])
grp = data.groupby(by=[data['Date/Time'].map(lambda x : x.hour)], as_index=False).size()
x = pd.DataFrame(grp)
x.columns = ['count']
x['count'] = x['count'] / 30
#print (x)
result['july'] = x['count']
#august
data = pd.read_csv("uber-raw-data-aug14.csv", parse_dates=['Date/Time'])
grp = data.groupby(by=[data['Date/Time'].map(lambda x : x.hour)], as_index=False).size()
x = pd.DataFrame(grp)
x.columns = ['count']
x['count'] = x['count'] / 30
#print (x)
result['august'] = x['count']
#september
data = pd.read_csv("uber-raw-data-sep14.csv", parse_dates=['Date/Time'])
grp = data.groupby(by=[data['Date/Time'].map(lambda x : x.hour)], as_index=False).size()
x = pd.DataFrame(grp)
x.columns = ['count']
x['count'] = x['count'] / 30
#print (x)
result['september'] = x['count']
print(result)
plt.plot(result.index, result['april'], label='April')
plt.plot(result.index, result['may'], label='May')
plt.plot(result.index, result['june'], label='June')
plt.plot(result.index, result['july'], label='July')
plt.plot(result.index, result['august'], label='August')
plt.plot(result.index, result['september'], label='September')
plt.xlabel('Hour')
plt.ylabel('avg number of trips')
plt.title('Uber trips avg per hour for apr to sept')
plt.legend()
plt.show()
```

```
# Plottin the uber pick up points from Apr 2014 to Sept 2014
# author: Punith Kumar M R
# The required CSV files have to be in the same directory of the python file

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#uber jul to sep
df = pd.read_csv("uber-raw-data-jul14.csv", parse_dates=['Date/Time'])
df['Date/Time'] = df['Date/Time'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date/Time'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
df1=df
df = pd.read_csv("uber-raw-data-aug14.csv", parse_dates=['Date/Time'])
df['Date/Time'] = df['Date/Time'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date/Time'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
df2=df
df = pd.read_csv("uber-raw-data-sep14.csv", parse_dates=['Date/Time'])
df['Date/Time'] = df['Date/Time'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date/Time'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
df3=df
df = [df1, df2, df3]
result = pd.concat(df)
print(result)
plt.plot_date(result['Date'], result['num of trips'], fmt='b-', color="blue", label='Uber')

#aMERICAN CAB
df = pd.read_csv("American_B01362.csv", parse_dates=['DATE'])
df['DATE'] = df['DATE'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['DATE'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color="green", fmt='b-', label='American')

#Carmel CAB
df = pd.read_csv("Carmel_B00256.csv", encoding='latin-1')
df['Date'] = pd.to_datetime(df["Date"])
df['Date'] = df['Date'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color="red", fmt='b-', label='Carmel')

#Dial7 CAB
df = pd.read_csv("Dial7_B00887.csv", parse_dates=['Date'])
df['Date'] = df['Date'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'], fmt='b-', color="cyan", label='Dial7')
```

```
#Diplo CAB
df = pd.read_csv("Diplo_B01196.csv", encoding='latin-1')
df['Date'] = pd.to_datetime(df["Date"])
df['Date'] = df['Date'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color="magenta", fmt='b-', label='Diplo')


#federal
df = pd.read_csv("Federal_02216.csv", parse_dates=['Date'])
df['Date'] = df['Date'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'], fmt='b-', color="yellow", label='Federal')


#Firstclass CAB
df = pd.read_csv("Firstclass_B01536.csv",encoding='latin-1' , parse_dates=['DATE'])
df['DATE'] = df['DATE'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['DATE'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color="black", fmt='b-', label='Firstclass')



#Highclass CAB
df = pd.read_csv("Highclass_B01717.csv", encoding='latin-1', parse_dates=['DATE'])
df['DATE'] = df['DATE'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['DATE'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color='magenta', fmt='b-', label='Highclass')


#LYFT CAB
df = pd.read_csv("Lyft_B02510.csv", parse_dates=['time_of_trip'])
df['time_of_trip'] = df['time_of_trip'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['time_of_trip'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color='olive', fmt='b-', label='LYFT')



#PrestigeCAB
df = pd.read_csv("Prestige_B01338.csv",encoding='latin-1', parse_dates=['DATE'])
df['DATE'] = df['DATE'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['DATE'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'],color='pink', fmt='b-', label='Prestige')
```

```
#skyline
df = pd.read_csv("Skyline_B00111.csv", parse_dates=['Date'])
df['Date'] = df['Date'].apply(lambda x: x.strftime('%m-%d-%y'))
counts = df['Date'].value_counts()
df = pd.DataFrame(counts)
df.columns = ['num of trips']
df['Date'] = pd.to_datetime(df.index)
df = df.reset_index(drop=True)
df = df.sort('Date')
plt.plot_date(df['Date'], df['num of trips'], fmt='b-', color='tan', label='Skyline')


plt.xlabel('TimeSpan')
plt.ylabel('number of trips')
plt.title('Uber trips avs other cab service trips')
plt.legend()
plt.show()
```