Analysis and Visualization of Cab Rides

Vishnu Shashank(20MCB1009),

Vellore Institute of Technology,

Chennai, India

Akshay Kumar

Yadav(20MCB1015)

Vellore Institute of Technology

Chennai, India

ABSTRACT

This The objective of this project is to understand the traffic in different boroughs of New York City and to try and categorize the various zones within various boroughs.

It can be office destinations, residential destination, popular brunch destination or party destination according to its popularity given the time-range and the day of the week.

Keywords
Cab data, visualization, Analysis.

1. 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 has covered almost all the major cities in the United States and is performing good by making it possible for anyone use a car.
New York City (NYC) is the most populous city in the United States. It is also the most densely populated major city in the United States.
The city is considered as one of the busiest places in the world and also different kind of people live in the city have their own professions and go to work or by their love place.
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.

2. Functional Requirements

- OS: WINDOWS 7,10
- R Studio
- RAM 2GB
- 256 TB HARDISC

3. Methodology

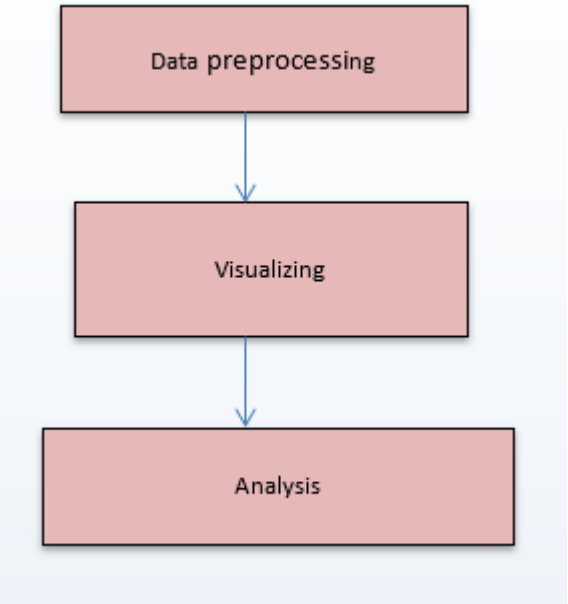


Fig. Workflow of project

4. Data Cleaning and Preprocessing

The raw data downloaded from Kaggle repository. We want to work out when and how temperature is changing at certain times of the day and aggregate this result to daily, monthly, and yearly levels. As such, we use Pandas to add month, year, and date columns. Simple stuff in preparation, and we can then output plots as required. With the data cleansed, we now have non-uniform samples of the weather at a given station throughout the year, at a sub-hour level. To make meaningful plots on this data, we can aggregate over the days and months to gain an overall view and to compare across stations.

5. Visualization for Year-data analysis

5.1 Trips by Day and Month

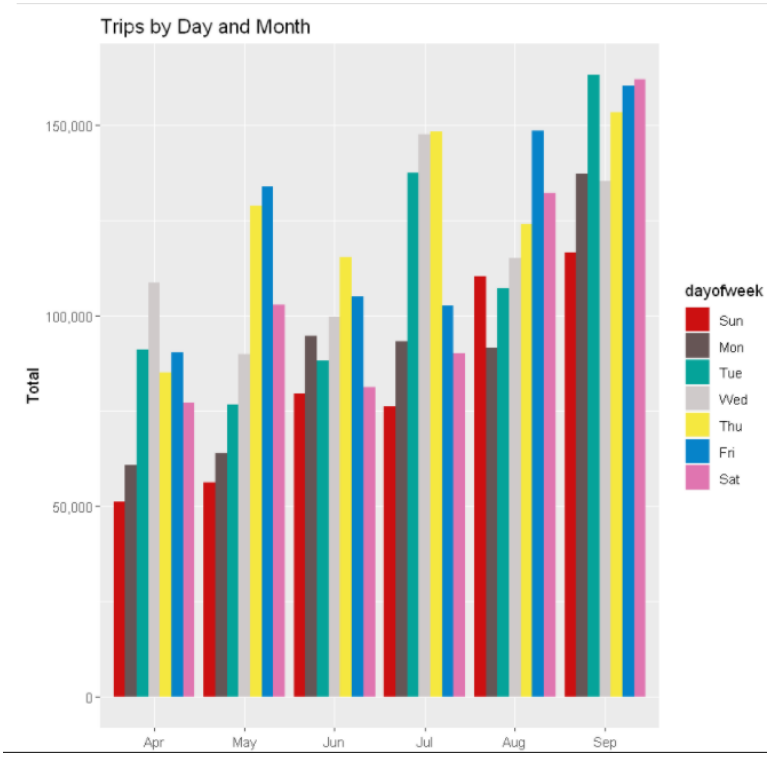
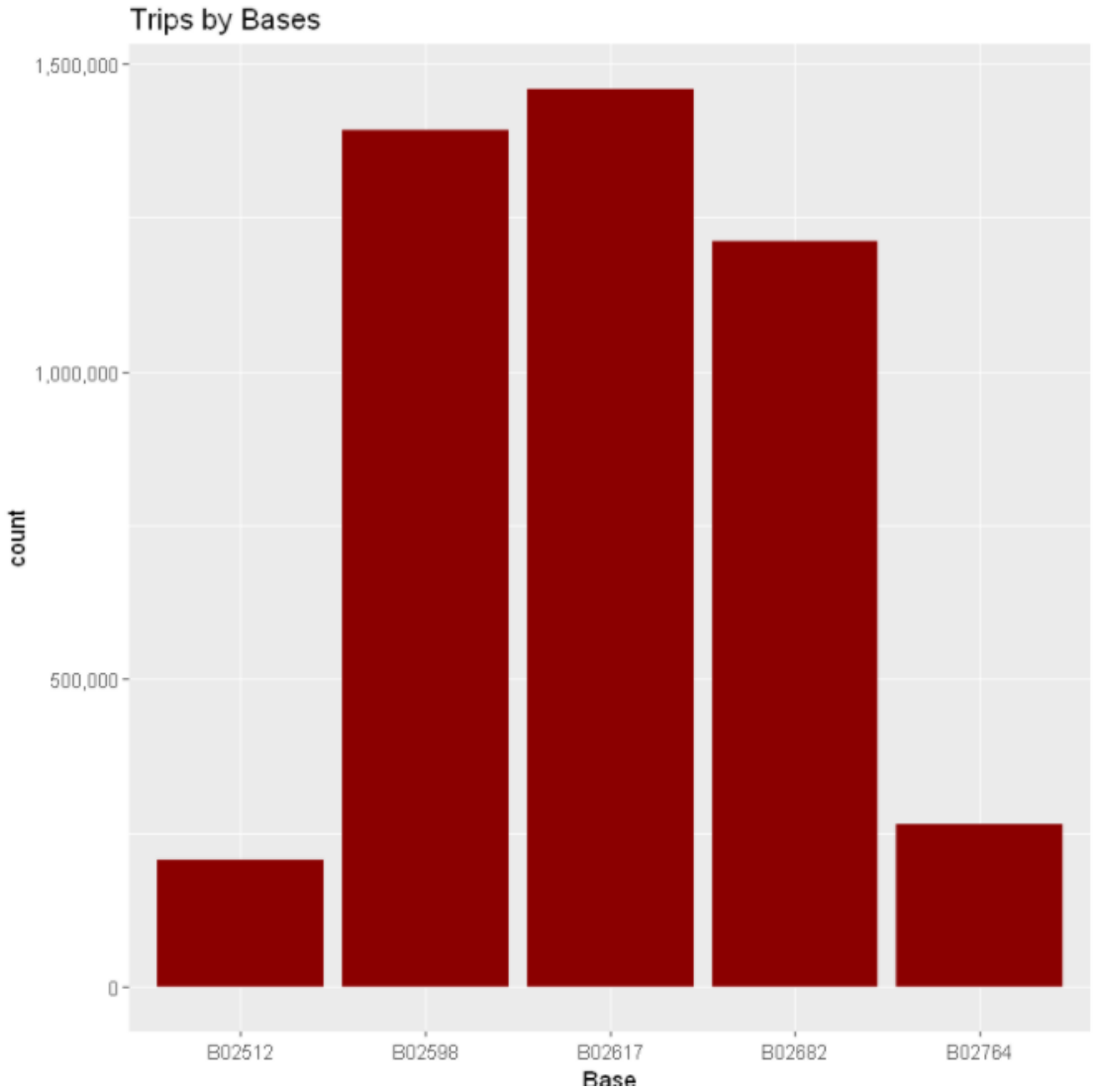


Fig. Bar plot of Total cab rides vs Month

5.2 Trip by Bases



5.3 Heatmap for hour and day

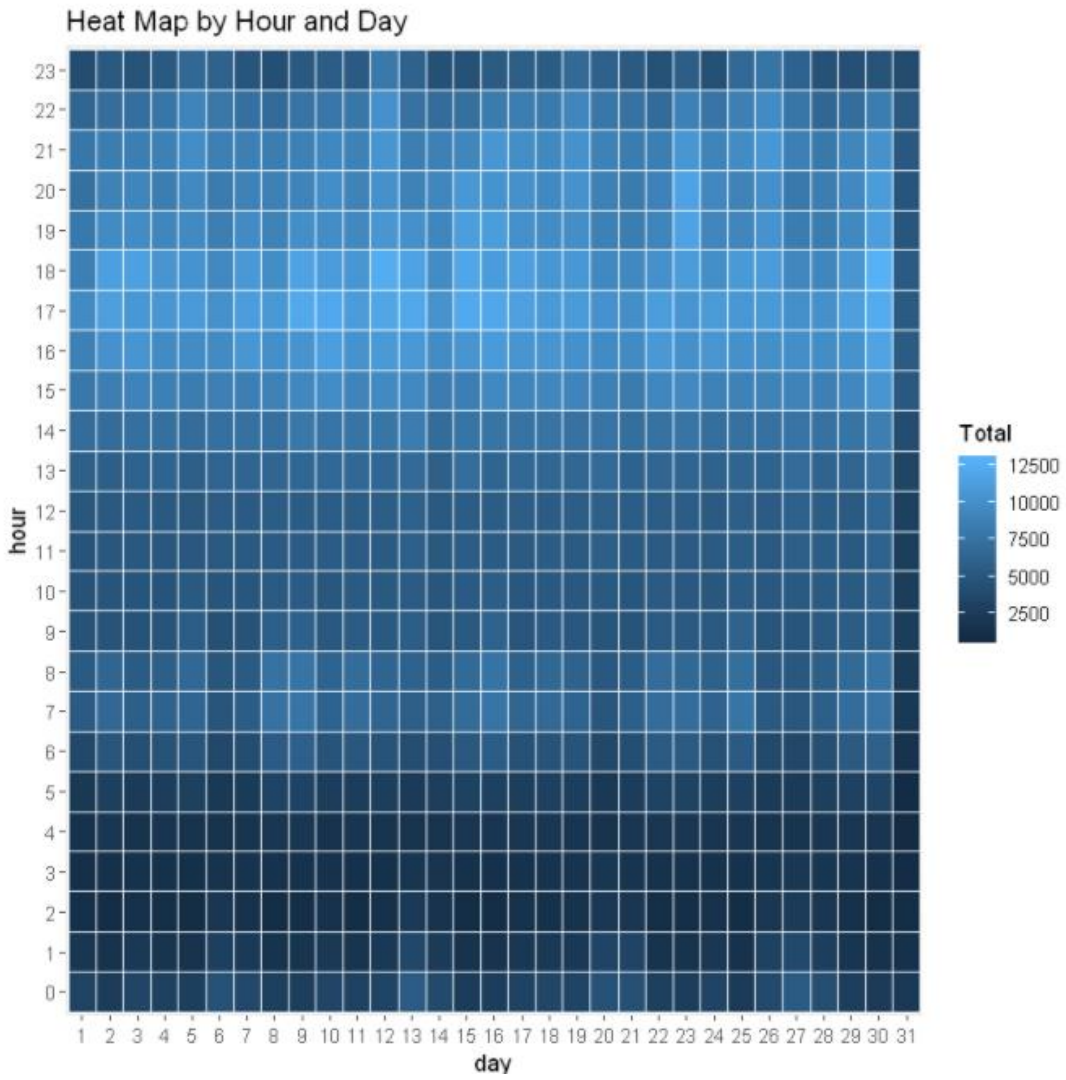


Fig. heatmap of attributes

5.4 Heatmap for month and bases

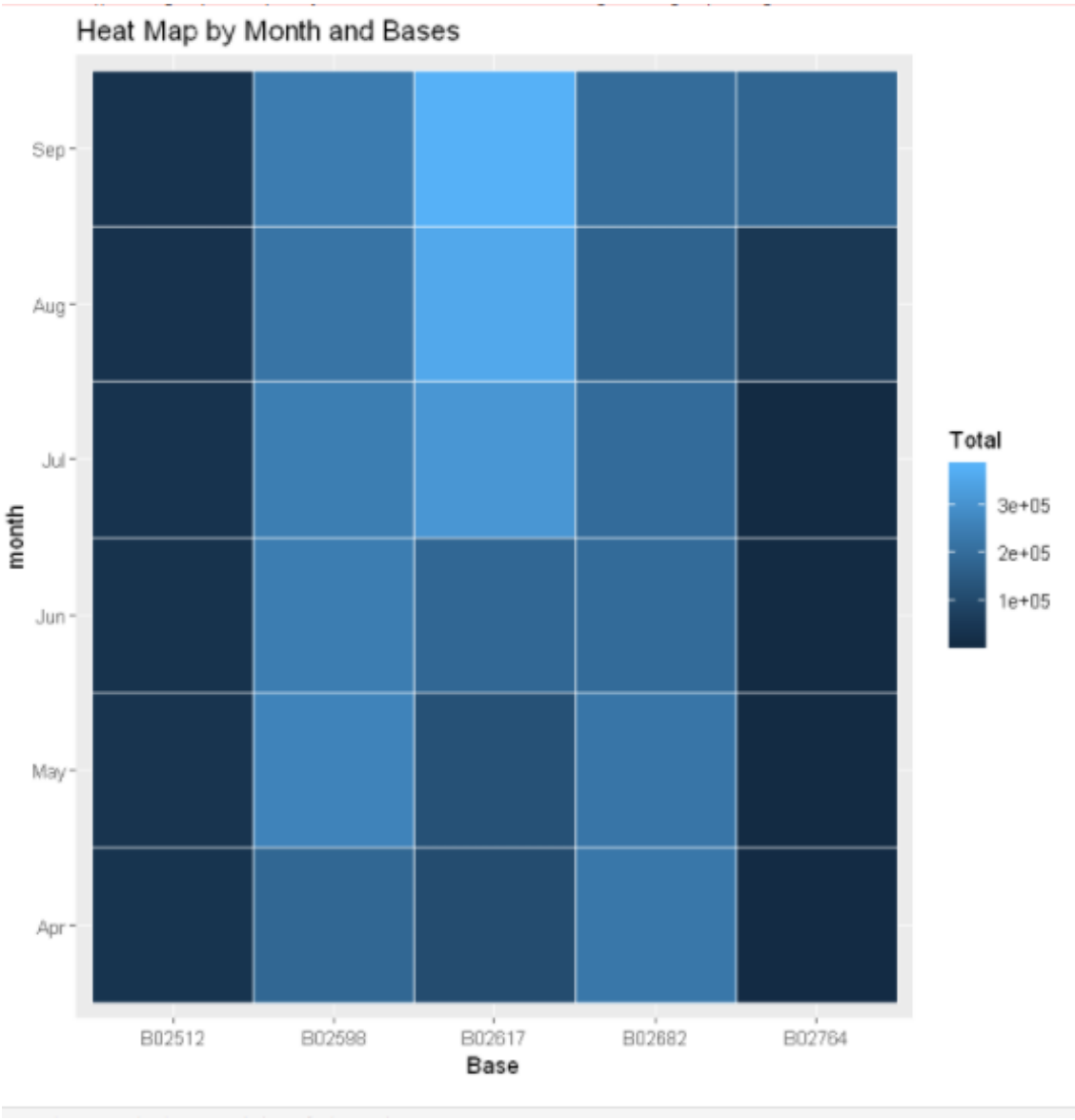


Fig. Heatmap

5.5 Heatmap for bases and day of week

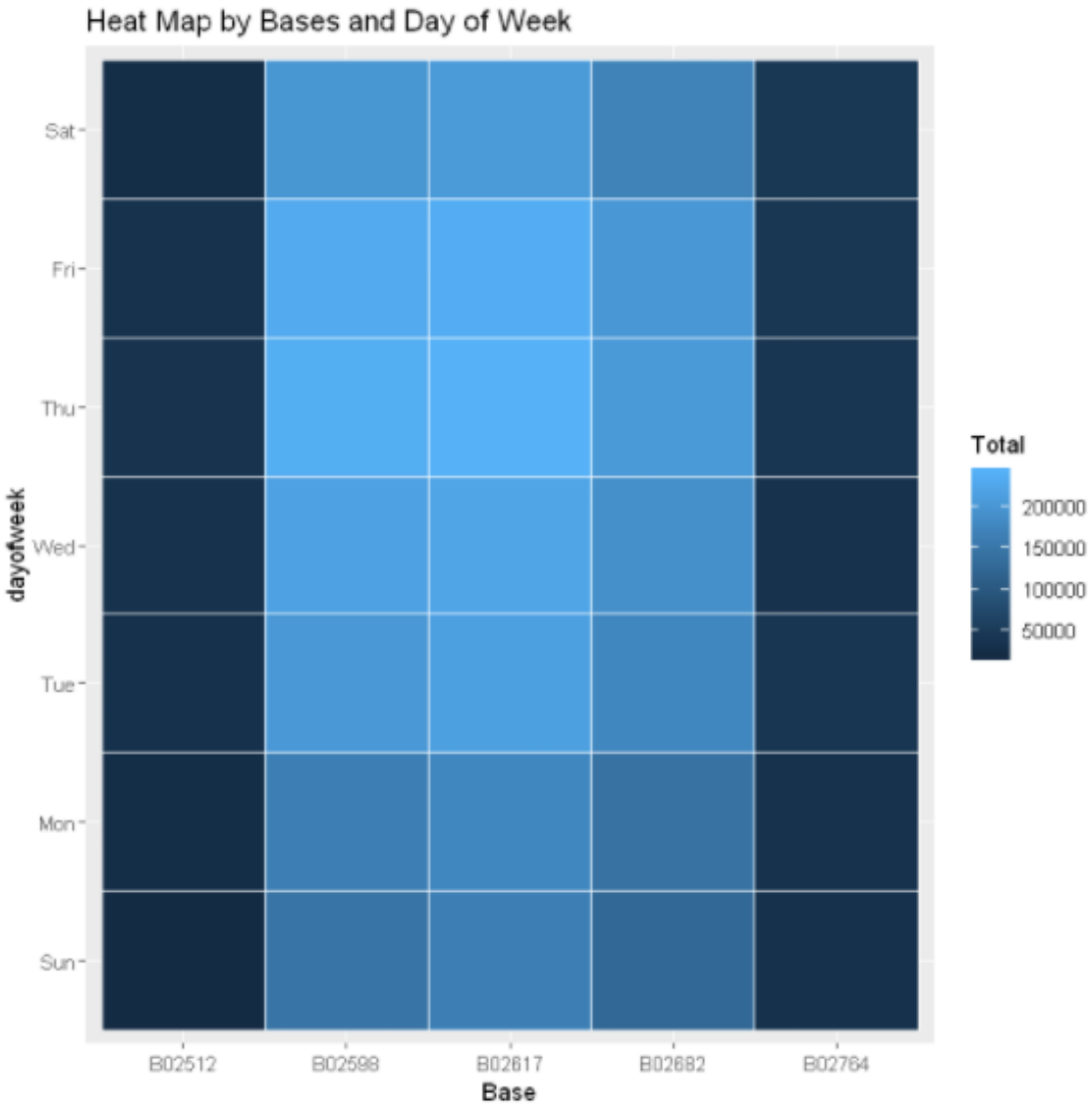


Fig. Heatmap

6. Geo map for cab rides according to different bases

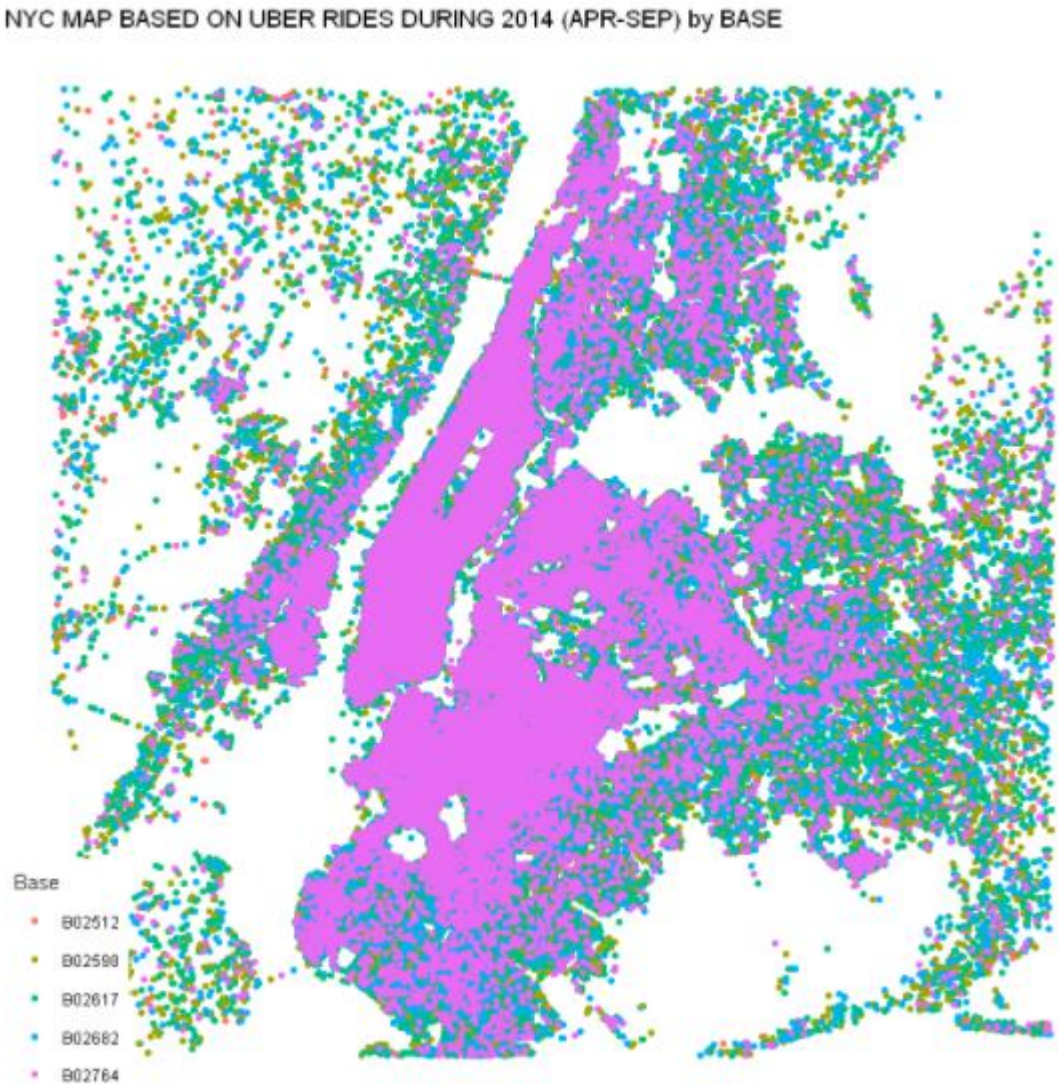
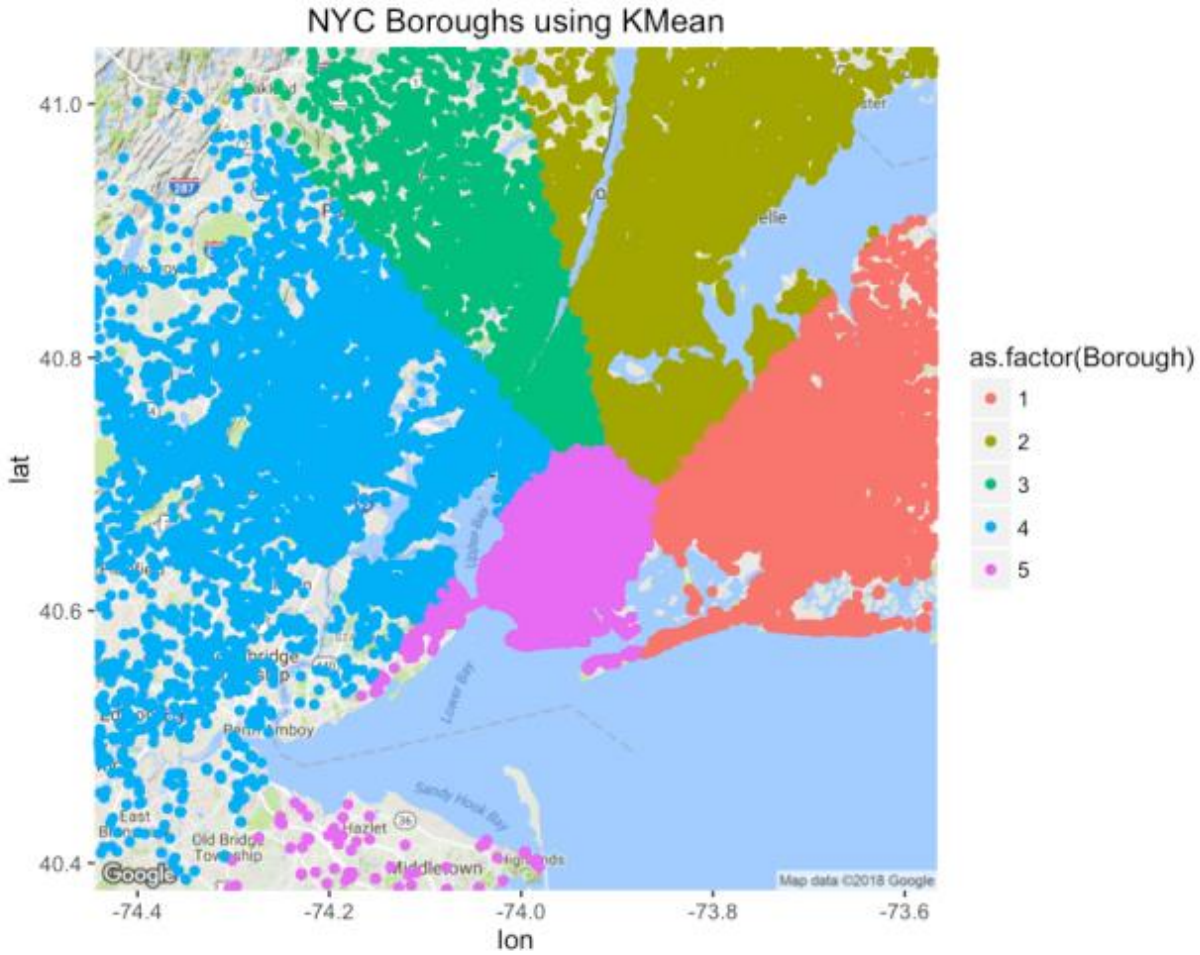


Fig. Geo map

K-Means Clustering with R:

K-means clustering is the most commonly used unsupervised machine learning algorithm for dividing a given dataset into k clusters.
Here, `k` represents the number of clusters and must be provided by the user. You already know `k` in case of the Uber dataset, which is 5 or the number of boroughs. k-means is a good algorithm choice for the Uber 2014 dataset since you do not know the target labels making the problem unsupervised and there is a pre-specified `k` value.

By Applying K-Means as well as by using the GGMAP Library , the following Map Visualization is obtained:



Conclusion and Future Work

- 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.
-
- Through the analysis the Information obtained like:
    - from comparing its data for a weekday and weekend, we observe few interesting points about this service. Each month the usage of Uber services has been increasing gradually.
    - Uber has largely conquered the market and its way ahead than its competitors. Uber is used a lot on weekdays compared to weekends.

References

1) Poornima Selvaraj, Pushpalatha Marudappa SRM Institute of Science and Technology "Analysis of Weather Data Using Forecasting Algorithms", ICCI-2017.
2) Patrik Lundblad , Oskar Eurenius , Tobias Heldring "Interactive Visualization of Weather and Ship Data" ,Conference: 13th International Conference on Information Visualization, IV 2009, 15-17 July 2009, Barcelona, Spain
3) T V Rajini kanth, V V S S S Balaram Sreenidhi Institute of Science & Technology, "Analysis of Indian Weather Data Sets Using Data Mining Techniques" , Fourth International Conference on Advances in Computing and Information Technology, May 2014
4) Alexandra Diehl - University of Zurich , L. Pelorosso- Claudio Delrieux Universidad Nacional del Sur, A. Celeste Saulo- Servicio Meteorológico Nacional, Argentina, "Visual Analysis of Spatio-Temporal Data: Applications in Weather Forecasting" , June 2015
5) Bo Ma- University of Florida, Alireza Entezari , "An Interactive Framework for Visualization of Weather Forecast Ensembles" , IEEE Transactions on Visualization and Computer Graphics , August 2018.
6) ATHUL C R1, Dr. Laksmi Mahesh, Assistant Professor, SCMS, Cochin, Kerala, India, "A Comparative Analysis of Classification Algorithms on Weather Dataset using Data Mining Tool", International Research Journal of Engineering and Technology (IRJET) May 2020.
7) Benjamin Obi Tayo Ph.D. Physicist, Data Science Educator, Writer. Interests: Data Science, Machine Learning, AI, Python & R, Personal Finance Analytics, Materials Sciences, Biophysics, "The Art of Data Visualization — Weather Data Visualization Using Matplotlib and Ggplot2" Nov 2019.
8) Aastha Sharma, Vijayakumar V, "Predictive Analytics In Weather Forecasting Using Machine Learning Algorithms" , EAI Endorsed Transactions on Cloud Systems July 2018.

9) Madhukar Rao G ,Dharavath Ramesh -Indian Institute of Technology (ISM) , “DSSAE-BBOA: deep learning-based weather big data analysis and visualization” , Springer Multimedia Tools and Applications, May 2021.
10) Olanrewaju V.Johnsona, Olayinka T.Jinadub, Olomi I.Aladesotea, “On experimenting large dataset for visualization using distributed learning and tree plotting techniques”, Scientific African , July 2020.