

DATA SCIENCE PROJECT
LABORATORY COURSE
TOPIC-Uber Pickup Analysis

MIS:- VIPIN INGLE 111903125

VIPUL GAIKWAD 111903126

VIRAJ NAJAN 111903127

For the assigned data science project, do the following:

1) Do the business requirement for the assigned project

There is a lot of information stored in the traffic flow data of any city. This data when mined over location can provide information about the major attractions of the city, it can help us understand the various zones of the city such as residential areas, office/school zones, highways, etc. This can help governments and other institutes plan the city better and enforce suitable rules and regulations accordingly.

The data when monitored over time can help us identify rush hours, holiday season, impact of weather, etc. This knowledge can be applied for better planning and traffic management. This can at a large, impact the efficiency of the people in the city and can also help avoid disasters, or at least faster redirection of traffic flow after accidents.

2) Identify the project objectives from your perspective

- ◆ To find out the pickup Hotspots from a particular area.
- ◆ Analysing the data of pickups on weekdays and weekend.
- ◆ Hour-wise analysis of Uber Pickups.
- ◆ To find out weekdays which have more pickups.
- ◆ To find out Hours of the day which are mostly busy
- ◆ To find out the impact of the Holidays on Uber Pickups.
- ◆ To analyse the data based on time.

3) Identify the data required

There are six files of raw data on Uber pickups in New York City from April to September 2014 (uber-raw-data-(month)14.csv). The files are separated by month and each has the following columns

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,

<https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city?select=uber-raw-data-aug14.csv>

4) Write a data scrapper, scrap the web for generating your corpus. The scrapped data should have minimum 10,000 instances.