UBER DATA ANALYSIS

In this project, you have to analyze the Uber Pickups in New York City dataset. This is more of a data visualization project that will guide you towards using the visualisation libraries for understanding the data and for developing an intuition for understanding the customers who avail the trips. So, before you start, take a quick revision to data visualization concepts.

Data Science Project – Uber Data Analysis:

Talking about our Uber data analysis project, data storytelling is an important component of Data Science through which companies are able to understand the background of various operations. With the help of visualization, companies can avail the benefit of understanding the complex data and gain insights that would help them to craft decisions. You will learn how to implement the visualization libraries on the Uber Pickups dataset and at the end, master the art of data visualization in python.

Dataset Description:

This directory contains data on over 4.5 million Uber pickups in New York City from April to September 2014, and 14.3 million more Uber pickups from January to June 2015. All the files are as they were received on August 3, Sept. 15 and Sept. 22, 2015.
FiveThirtyEight obtained the data from the NYC Taxi & Limousine Commission (TLC) by submitting a Freedom of Information Law request on July 20, 2015.

The Data

The dataset contains, roughly, three groups of files:

- Uber trip data from 2014 (April - September), separated by month, with detailed location information
- Uber trip data from 2015 (January - June), with less fine-grained location information
- Aggregate ride and vehicle statistics for all FHV companies (and, occasionally, for taxi companies)

Uber trip data from 2014

There are six files of raw data on Uber pickups in New York City from April to September 2014. 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

These files are named:

- uber-raw-data-apr14.csv
- uber-raw-data-aug14.csv
- uber-raw-data-jul14.csv
- uber-raw-data-jun14.csv
- uber-raw-data-may14.csv
- uber-raw-data-sep14.csv

Uber trip data from 2015

Also included is the file uber-raw-data-janjune-15.csv. This file has the following columns:

- Dispatching_base_num : The TLC (Taxi & Limousine Commission) base company code of the base that dispatched the Uber
- Pickup_date : The date and time of the Uber pickup
- Affiliated_base_num : The TLC base company code affiliated with the Uber pickup
- locationID : The pickup location ID affiliated with the Uber pickup

The Base codes are for the following Uber bases:

B02512 : Unter B02598 : Hinter B02617 : Weiter B02682 : Schmecken B02764 : Danach-NY B02765 : Grun B02835 : Dreist B02836 : Drinnen

For coarse-grained location information from these pickups, the file taxi-zone-lookup.csv shows the taxi Zone (essentially, neighborhood) and Borough for each locationID.

Aggregate Statistics

The file Uber-Jan-Feb-FOIL.csv contains aggregated daily Uber trip statistics in January and February 2015.

CONCLUSION:

Overall you have to submit a detailed visualization report about the Uber pickups in New York city over the period. The trends and seasonality has to be defined in different periods.

ALL THE BEST