Case Study 1: Uber supply-demand gap

Introduction

This data set is a masked data set which is similar to what data analysts at Uber handle. Solving this assignment will allow you to demonstrate how problems are systematically solved using EDA (Exploratory data analytics) and data visualization.

Business Understanding

You may have some experience of travelling to and from the airport. Have you ever used Uber or any other cab service for this travel? Did you at any time face the problem of cancellation by the driver or non-availability of cars?

Well, if these are the problems faced by customers, these very issues also impact the business of Uber. If drivers cancel the request of riders or if cars are unavailable, Uber loses out on its revenue. You decide to address the problem Uber is facing - driver cancellation and non-availability of cars leading to loss of potential revenue.

Business Objectives

The aim of analysis is to identify the root cause of the problem (i.e. cancellation and non-availability of cars) and recommend ways to improve the situation. As a result of your analysis, you should be able to present the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

Data Understanding

The data is: Uber Request Data.csv

There are six attributes associated with each request made by a customer:

Request id: A unique identifier of the request

Time of request: The date and time at which the customer made the trip request

Drop-off time: The drop-off date and time, in case the trip was completed

Pick-up point: The point from which the request was made

Driver id: The unique identification number of the driver

Status of the request: The final status of the trip, that can be either completed, cancelled by the driver or no cars available

Note: For this assessment, only the trips to and from the airport are being considered.

Data Cleaning and Preparation - Hints

Identify the data quality issues and clean the data so that you can use it for analysis.

Ensure that the dates and time are in the proper format. Derive new variables which will be useful for analysis.

Results Expected

A. Visually identify the most pressing problems for Uber.

B. Find out the gap between supply and demand and show the same using plots.

C. Find the time slots when the highest gap exists

D. Find the types of requests (city-airport or airport-city) for which the gap is the most severe in the identified time slots

Assessment criteria

You code will not only be assessed on the basis of solution you will provide, but also on code hygiene and coding practices used while writing the code.

Close Ended questions –

    a. Driver ID who has completed the maximum number of trips and number of trips

    b. Percentage cancellation of trips at airport

    c. WeekDay on which maximum number of requests gets a 'Not available' status in the city. Also provide the number of such requests on that day.