

Problem Statement

Introduction

This data set is a masked data set which is similar to what data analysts at Uber handle. Solving this assignment will give you an idea about how problems are systematically solved using EDA and data visualisation.

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.

As an analyst, 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 to the client the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

Data Understanding

Download the dataset from below.

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

1. Request id: A unique identifier of the request
2. Time of request: The date and time at which the customer made the trip request
3. Drop-off time: The drop-off date and time, in case the trip was completed
4. Pick-up point: The point from which the request was made
5. Driver id: The unique identification number of the driver
6. 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 assignment, only the trips to and from the airport are being considered.

Data Cleaning and Preparation - Hints

1. Identify the data quality issues and clean the data so that you can use it for analysis.
2. Ensure that the dates and time are in the proper format. Derive new variables which will be useful for analysis.

Results Expected

1. Visually identify the most pressing problems for Uber.
 - a. Hint: Create plots to visualise the frequency of requests that get cancelled or show 'no cars available'; identify the most problematic types of requests (city to airport / airport to city etc.) and the time slots (early mornings, late evenings etc.) using plots
2. Find out the gap between supply and demand and show the same using plots.
 - a. Find the time slots when the highest gap exists
 - b. Find the types of requests (city-airport or airport-city) for which the gap is the most severe in the identified time slots
3. What do you think is the reason for this issue for the supply-demand gap? Write the answer in less than 100 words. You may accompany the write-up with plot(s).
4. Recommend some ways to resolve the supply-demand gap.

Present the problem, the analyses and the recommendations using plots to the Chief Data Scientist in a well-formatted presentation (make sure to submit a PDF version of the PPT). Also, include a commented jupyter Notebook in your submission. Please note that the assignment has to be done completely in Python including the plots. However, you may recreate the plots in Tableau to include in the presentation.