Uber Supply-Demand Gap

**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. Let’s hear more about such problems that Uber faces during its operations.

Play Video

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.

**[Uber Request Data](https://cdn.upgrad.com/UpGrad/temp/76b3b6a4-d87d-4e82-b1c3-3f6e10b9c076/Uber%20Request%20Data.csv" \o "Uber Request Data.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/76b3b6a4-d87d-4e82-b1c3-3f6e10b9c076/Uber%20Request%20Data.csv" \o "Uber Request Data.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/76b3b6a4-d87d-4e82-b1c3-3f6e10b9c076/Uber%20Request%20Data.csv" \o "Uber Request Data.csv" \t "_blank)**

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.
   * 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.
   * Find the time slots when the highest gap exists
   * 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**.

Please refer to the rubric provided on the next page to get a clear idea of what parameters will you be evaluated on.

**Evaluation Rubric**

| **Criteria** | **Meets expectations** | **Does not meet expectations** |
| --- | --- | --- |
| **Data Cleaning and Manipulation**  **(25 %)** | All data quality issues are correctly identified and reported.     The data is converted to a clean format suitable for analysis in Python. New metrics are derived wherever required and are used for analysis. | Data quality issues are overlooked or are not identified correctly.    The data is not converted to a clean format which is suitable for analysis or is not cleaned using commands in Python. New metrics are not derived or are not used for analysis. |
| **Data Analysis**  **(45 %)** | The right problem is solved which is coherent with the needs of the business. The analysis has a clear structure and the flow is easy to understand.    Realistic assumptions are made and proper reasons are given for all choices made.    The time slots and problems are identified correctly with valid reasons.    Univariate and segmented analysis are done correctly and successfully identify the problems.    The demand and supply are defined properly and the numbers are correct.    All relevant plots during the analysis are created. The choice of plots is correct, i.e. the plots clearly display the important insights. The reason for choosing certain plots, aesthetics and geometries etc. is mentioned in the comments.    All major issues are correctly pointed out. | The analyses do not address the right problem or deviate from the business objectives.  The analysis lacks a clear structure and is not easy to follow.    Realistic assumptions are not made wherever required or unrealistic ones are made.      The time slots and problems are selected without any sound reasoning.    Univariate and segmented analysis are not done correctly or are not able to identify the problems.    The demand and supply numbers are incorrect/not defined meaningfully.    All relevant plots are not created. The choice of plots is not ideal and the plots are either difficult to interpret or lack clarity or neatness. The reason for choosing certain plots, aesthetics and geometries etc. is not mentioned in the comments.    Not all major issues are pointed out/many minor issues are pointed out. |
| **Presentation and Recommendations**  **(20 %)** | The presentation has a clear structure, is not too long, and explains the most important results concisely.    The recommendations to solve the problems are realistic, actionable and coherent with the analysis.    If any assumptions are made, they are stated clearly. | The presentation lacks structure, is too long or does not put emphasis on the important observations.    Contains unnecessary details or lacks the important ones.    The recommendations to solve the problems are either unrealistic, non-actionable or incoherent with the analysis.    Assumptions made, if any, are not stated clearly. |
| **Conciseness and Readability of the Code**  **(10 %)** | The code is concise and syntactically correct.    Wherever appropriate, built-in functions are used instead of writing long code (if-else statements, for loops).    Custom functions are used to perform repetitive tasks.    The code is readable with variables appropriately named and detailed comments are written wherever necessary. | The code has errors in the syntax.      Long and complex codes used instead of shorter built-in functions.    Custom functions are not used to perform repetitive tasks resulting in the same piece of code being repeated multiple times.    Code readability is poor because of vaguely named variables or lack of comments wherever necessary. |

feedback

**Report an error**