Uber Supply Demand Gaps

Project Summary:

The aim of this project is to perform Data Analysis on Uber request data. The data reveals a gap between the demand and supply of Uber services. The goal of this project is to perform data cleaning, analysis and to visualize the data and uncover insights that can help bridge this gap between demand and supply.

About the Dataset, the dataset consists of a single CSV file named uber_request_data.csv, which contains 6,745 records and 10 columns. The column names are: request_id, pickup_point, driver_id, status, request_timestamp, request_time, request_time_status, drop_timestamp, drop_time, and drop_time_status.

Problem Statement:

There is a gap between demand and supply in Uber services. Many requests are cancelled during the early morning and late-night hours. Additionally, cars are often unavailable during these times. This indicates a significant mismatch between demand and supply during odd hours.

Data cleaning, analysis and visualization steps:

Step1: Data cleaning in Excel

Step2: Find insights using SQL queries

Step3: Visualize the data and find insights using python libraries (matplotlib, seaborn)

Step4: Create the dashboard in Excel

Step1: Data Cleaning

- In data cleaning process, create the new columns like ‘Request_time’ , ‘Request_time_category’, ‘Drop_time’, ‘Drop_time_category’. I get these columns from ‘Request_timestamp’ and ‘Drop_timestamp’ columns accordingly.
- Transform the column names into lower case
- Fill the missing values in driver_id and drop_timestamp.
- Check the duplicate value.

Step2: Find Insights using SQL Queries

- **Count Total Request**

SELECT COUNT(request_id) total_request

FROM uber_request_data;

total_request bigint
6745

Insight: Total request from different pickup points and different time of day.

- **Number of Request by Pickup Points**

SELECT pickup_point, COUNT(*) total_request

FROM uber_request_data

GROUP BY pickup_point

ORDER BY total_request DESC;

pickup_point text	total_request bigint
City	3507
Airport	3238

Insights: 3507 requests pickup point is city and 3238 requests pickup point is airport.

- **Number of Drop Request Time of the Day**

SELECT drop_time_status, COUNT(*) drop_request

FROM uber_request_data

GROUP BY drop_time_status

ORDER BY drop_request DESC;

drop_time_status text	drop_request bigint
Not Applicable	3914
Morning	781
Night	605
Evening	518
Afternoon	413
Early Morning	307
Late Night	207

Insight: Maximum drop request time of day is morning that is 1876 request. 1676 drop requests time of day is Night, In Evening number of drop request is 1290, In early morning drop request is 775, in afternoon drop request is 654 and in late night drop request is 474. 3914 drop request not applicable because of no driver availability or no car availability and drive cancelled.

- **Number of Requests by Status**

SELECT status, COUNT(*) request_by_status

FROM uber_request_data

GROUP BY status

ORDER BY request_by_status DESC;

status text	request_by_status bigint
Trip Completed	2831
No Cars Available	2650
Cancelled	1264

Insight: Number of requests completed are 2831, number of requests cancelled are 1264, Car not available for the number of requests are 2650.

- **Number of requests are cancelled in early morning and late night**

SELECT COUNT(request_id) AS cancelled_requests

FROM uber_request_data

WHERE status = 'Cancelled'

AND request_time_status IN ('Early Morning', 'Late Night');

cancelled_requests bigint
389

Insight: Total number of requests are cancelled in late night and early morning are 389.

- **Top 10 drivers with most completed trip**

```
SELECT driver_id, COUNT(*) completed_trip
FROM uber_request_data
WHERE status = 'Trip Completed'
GROUP BY driver_id
ORDER BY completed_trip DESC
LIMIT 10;
```

driver_id text	completed_trip bigint
22	16
233	15
184	15
176	14
126	14
118	14
23	14
134	14
69	14
24	14

Insight: List of top ten drivers with driver id and number of request competition.

- **Total Requests, Completed Request, Gap in demand and supply request by Time of Day**

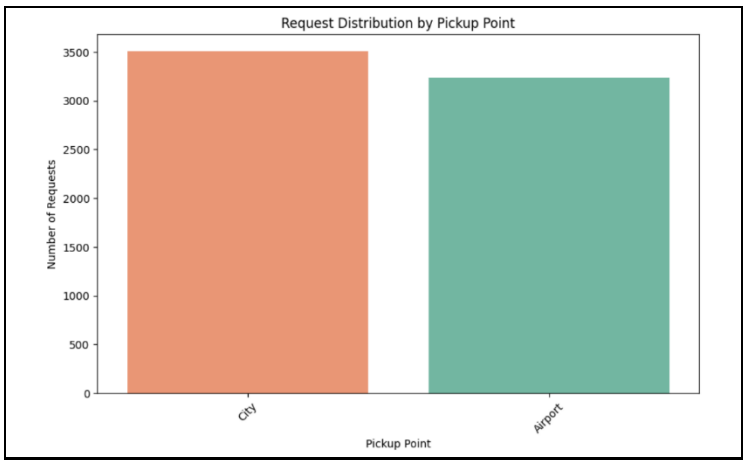
```
SELECT request_time_status, COUNT(*) AS total_requests,
  SUM(CASE WHEN status = 'Trip Completed' THEN 1 ELSE 0 END) AS completed_requests,
  SUM(CASE WHEN status != 'Trip Completed' THEN 1 ELSE 0 END) AS gap
FROM uber_request_data
GROUP BY request_time_status
ORDER BY gap DESC;
```

request_time_status text	total_requests bigint	completed_requests bigint	gap bigint
Evening	1559	571	988
Morning	1671	733	938
Night	1439	560	879
Early Morning	1047	430	617
Afternoon	652	401	251
Late Night	377	136	241

Insight: In morning total requests are 1671, requests complete 571, gap is 988. In early morning total requests are 1047, complete request 430, gap 617. In Evening total requests 1559, request complete 571, gap 988. In night total request 1439, complete request 560, gap 879.

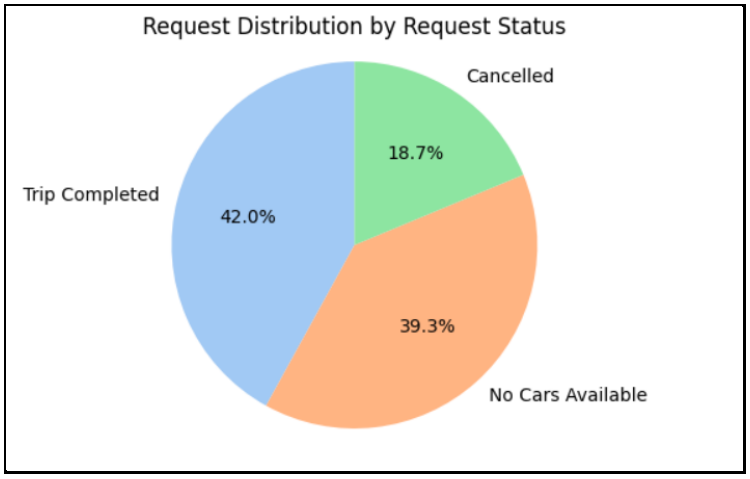
Step3: Visualization and Find Insights Using Python Libraries (pandas, matplotlib, seaborn)

- **Request Distribution by Pickup Point**



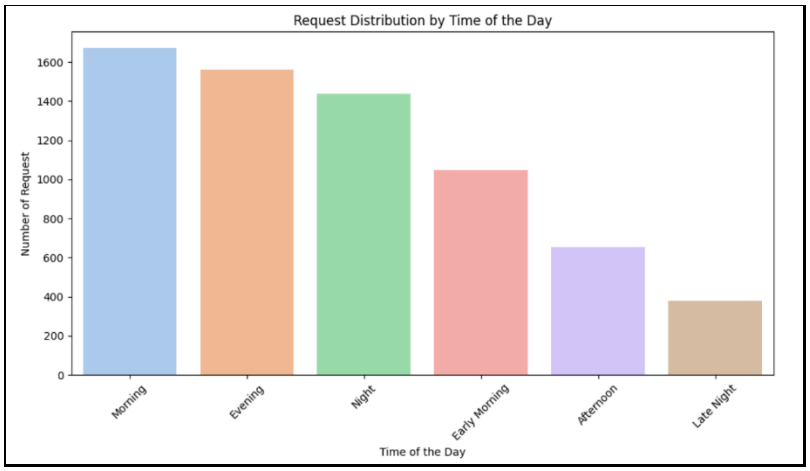
Insight: Maximum number of request pickup point is City.

- **Request Distribution by Status**



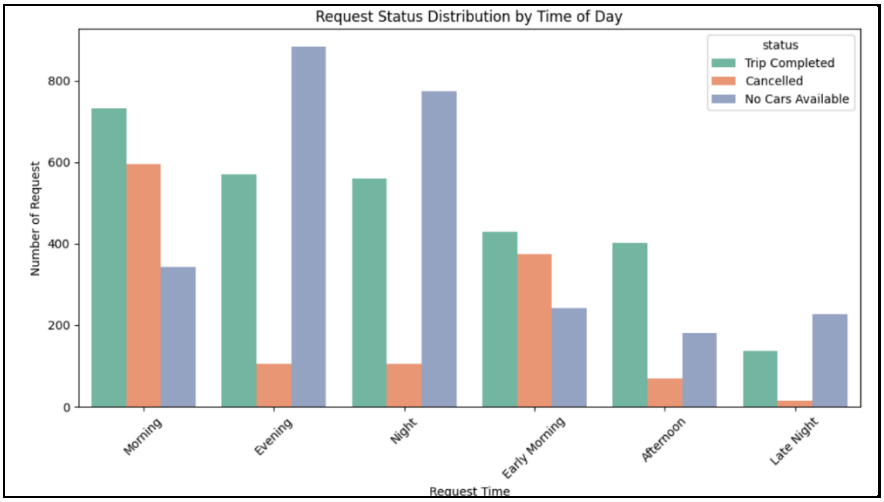
Insight: Trip completed is 42% of total request. Request cancelled is 18.7% of total number of requests. No car available for request is 39.3% of total number of requests.

- **Request by Time of Day**



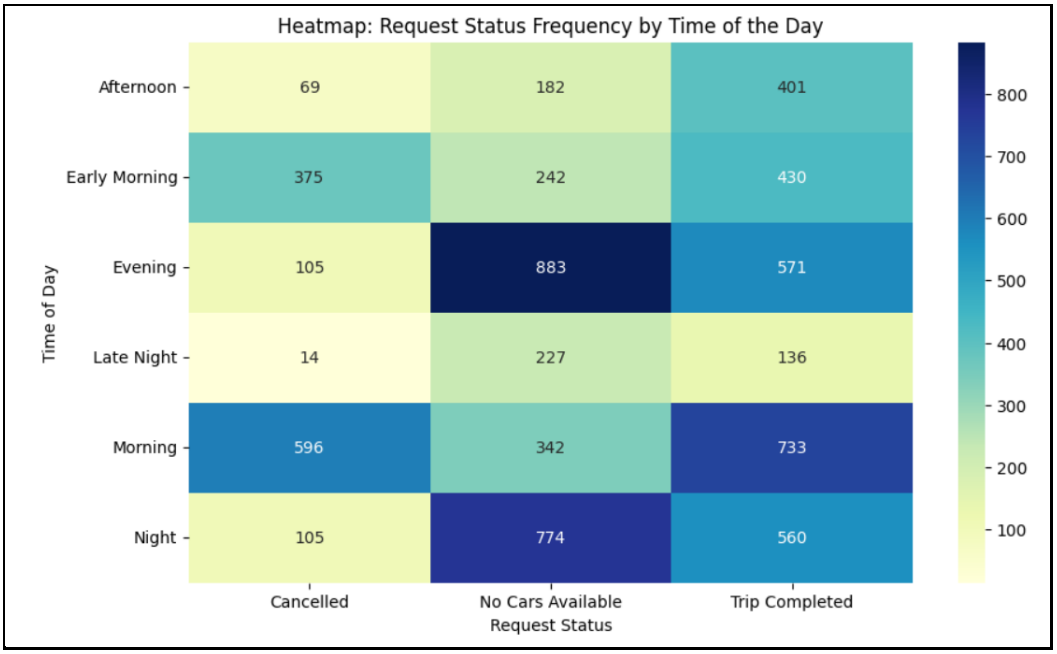
Insight: Request is high in morning and evening and low in late night.

- **Request Status by Time of Day**



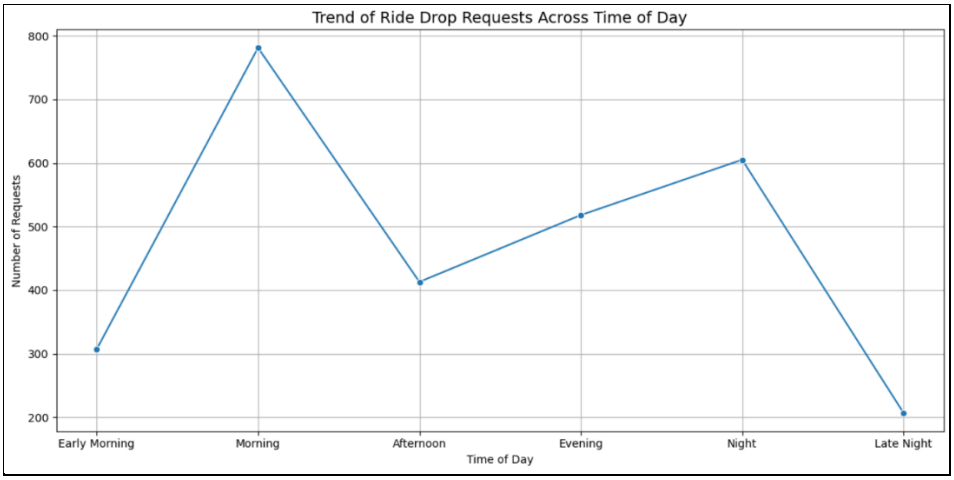
Insight: Trip competition is maximum in morning, car availability is low in evening and night, maximum requests are cancelled in morning and early morning.

- **Heatmap: Request Status Frequency by Time of Day**



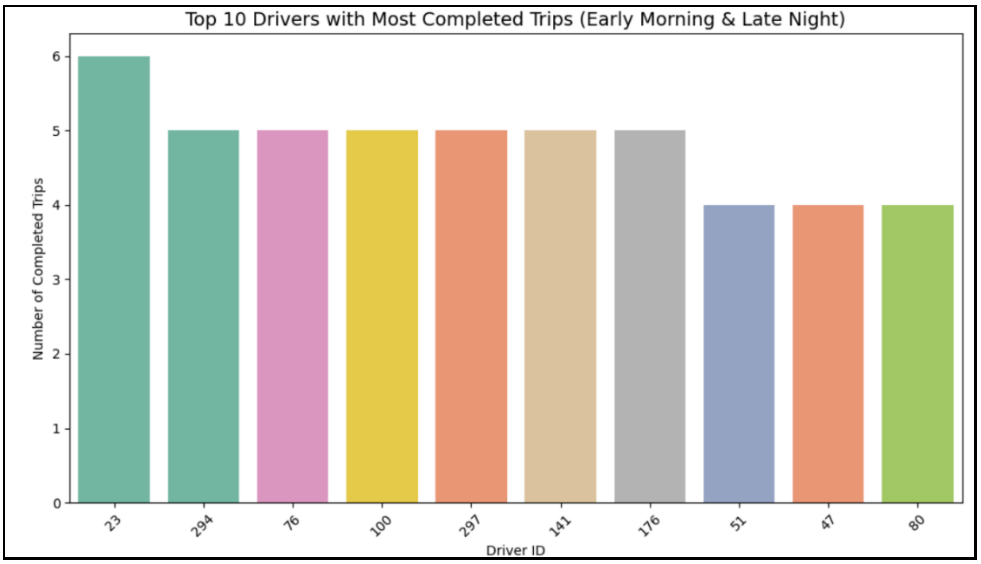
Insight: Maximum Trip competition is 733 in morning, minimum trip competition is 136 in late night. Maximum cancel request is 596 in morning, minimum cancel request is 14 in Late night. No car availability is maximum in evening that is 883 and night 774.

- **Drop Request Across Time of Day**



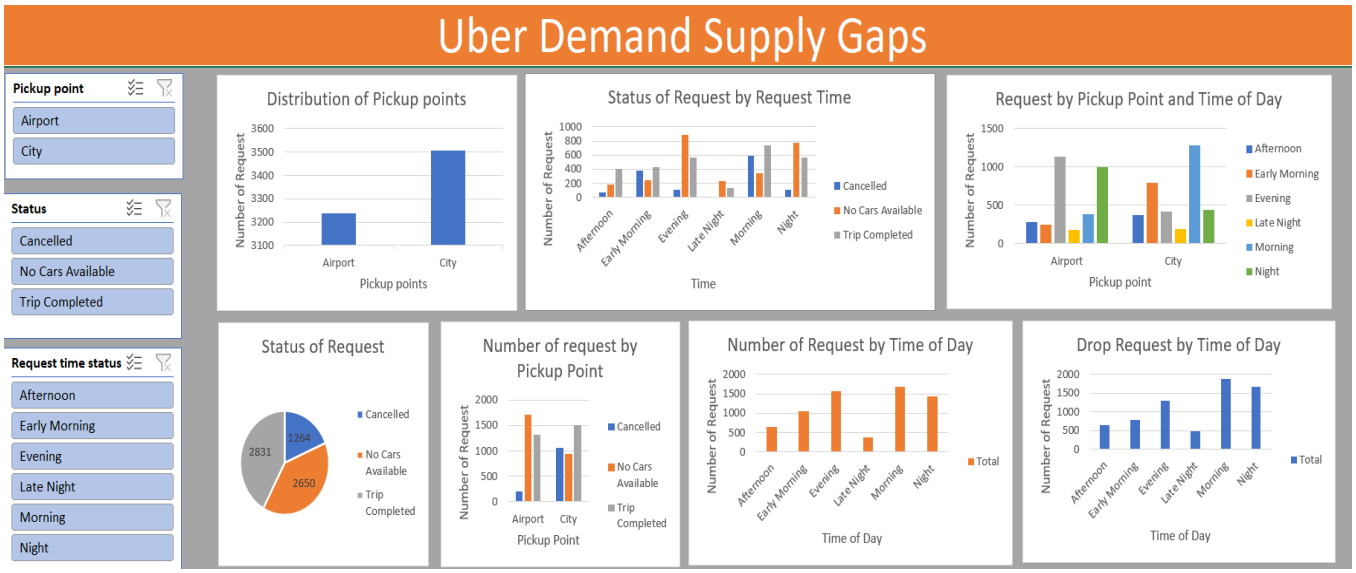
Insight: Drop request is high in morning and night. Low in early morning, late night and afternoon.

- **Top 10 Drivers Who Completed the Request in Early Morning and Late Night**



Insight: List of driver's id who completed the request in early morning and late night.

Step4: Create the Dashboard in Excel



Insights:

- **Distribution of Pickup Point:** Maximum request pickup point is City.
- **Status of Request-by-Request Time:** Trip completion is high in morning, Car availability is low in evening and night, cancelled request is high in morning and early morning.
- **Request by Time of Day and Pickup Point:** Number of requests is high in morning and early morning from pickup point city. Number of requests is high in evening and night from pickup point Airport.
- **Status of Request:** Completed request is 2831, 42% of total request. Request cancelled is 1264, which is 19% of total request. Request nit completed because No car available is 2650 which is 39% of total request.
- **Status of Request by Pickup Point:** No car availability is high with pickup point airport. Trip competition is high with pickup point city. Cancelled request is high with pickup point is city.
- **Number of Request by Time of Day:** Maximum request in morning and evening. Minimum in late night.
- **Drop Request by Time of Day:** Maximum drop request in morning and night. Minimum in late night.

Recommendations:

- Deploy more drivers and vehicles during early morning and late-night hours.
- Offer additional incentives or bonuses to drivers who complete rides during these off-peak times to reduce cancellations.
- Encourage users to book rides during low-demand hours by offering price discounts and enhanced safety features.