Supply & Demand Gap of Uber

NGUYEN THACH THAO
NGUYEN THI TUYET TRINH
MENTOR: NGUYEN NGOC MINH – PHONG LUU

Table of Content

Context

Uber business problem

“Uber is facing - driver cancellation and non-availability of cars leading to loss of potential revenue.

The aim of the 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 analysis, we 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.”

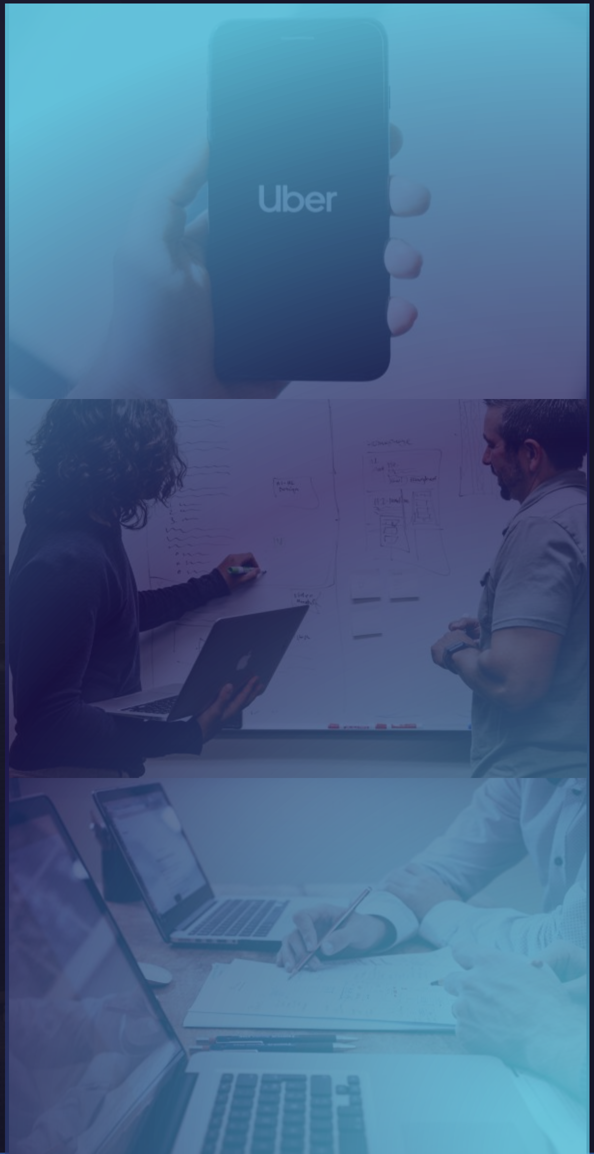
Business Understanding

About Uber

“**Uber** is an American transportation conglomerate that mainly provides taxi services where individuals can hail a taxi (Uber) in an app on their phone.

The company has over 131 million monthly active users and 5.4 million active drivers worldwide. The company has an average of 23 million trips each day through all of their services combined.”

Uber.com



Problem

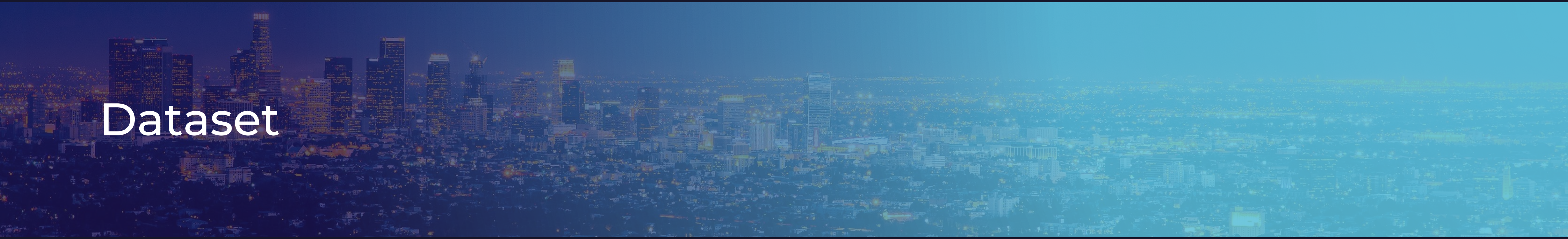
Driver cancellation and non-availability of cars leading to loss of potential revenue

Objective

Finding root cause(s) and possible hypotheses of the problem(s) and recommend.

Methods

Dataset: Kaggle
Python
Power BI
Logic Tree (Mc Kinsey)



Dataset

Request ID

A unique identifier of the request

Pickup Point

The point from which the request was made (City, Airport)

Driver ID

The unique identification number of the driver

Status

The final status of the trip, can be either completed, canceled by the driver or no cars are available

Request Timestamp

The date and time at which the customer made the trip request

Drop Timestamp

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

Cleaning data by Python

```
#Checking null values in column
df.isnull().sum()
#DriverID Null -> reason: no cars available
#Drop timestamp Null -> reason: no cars available, cancel
```

```
Request id             0
Pickup point           0
Driver id           2650
Status                 0
Request timestamp      0
Drop timestamp      3914
dtype: int64
```

Check null

```
df.info()
#Check data type
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6745 entries, 0 to 6744
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Request id         6745 non-null   int64
 1   Pickup point       6745 non-null   object
 2   Driver id          4095 non-null   float64
 3   Status             6745 non-null   object
 4   Request timestamp  6745 non-null   object
 5   Drop timestamp     2831 non-null   object
dtypes: float64(1), int64(1), object(4)
memory usage: 316.3+ KB
```

Check datatype

```
#Check duplicate
duplicate_rows = df.duplicated(keep=False)
print(duplicate_rows)
```

```
0       False
1       False
2       False
3       False
4       False
        ...
6740    False
6741    False
6742    False
6743    False
6744    False
Length: 6745, dtype: bool
```

Check duplicate

```
#Convert data: Request timestamp, Drop timestamp - object -> convert to datetime
df["Request timestamp"] = pd.to_datetime(df["Request timestamp"])
df["Drop timestamp"] = pd.to_datetime(df["Drop timestamp"])
df.head(10)
```

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	2016-11-07 11:51:00	2016-11-07 13:00:00
1	867	Airport	1.0	Trip Completed	2016-11-07 17:57:00	2016-11-07 18:47:00
2	1807	City	1.0	Trip Completed	2016-12-07 09:17:00	2016-12-07 09:58:00
3	2532	Airport	1.0	Trip Completed	2016-12-07 21:08:00	2016-12-07 22:03:00
4	3112	City	1.0	Trip Completed	2016-07-13 08:33:16	2016-07-13 09:25:47
5	3879	Airport	1.0	Trip Completed	2016-07-13 21:57:28	2016-07-13 22:28:59
6	4270	Airport	1.0	Trip Completed	2016-07-14 06:15:32	2016-07-14 07:13:15
7	5510	Airport	1.0	Trip Completed	2016-07-15 05:11:52	2016-07-15 06:07:52
8	6248	City	1.0	Trip Completed	2016-07-15 17:57:27	2016-07-15 18:50:51
9	267	City	2.0	Trip Completed	2016-11-07 06:46:00	2016-11-07 07:25:00

Convert data

```
#Replace "/", "-" in Request timestamp, Drop timestamp
df["Request timestamp"] = df["Request timestamp"].replace("/","-")
df["Drop timestamp"] = df["Drop timestamp"].replace("/","-")
```

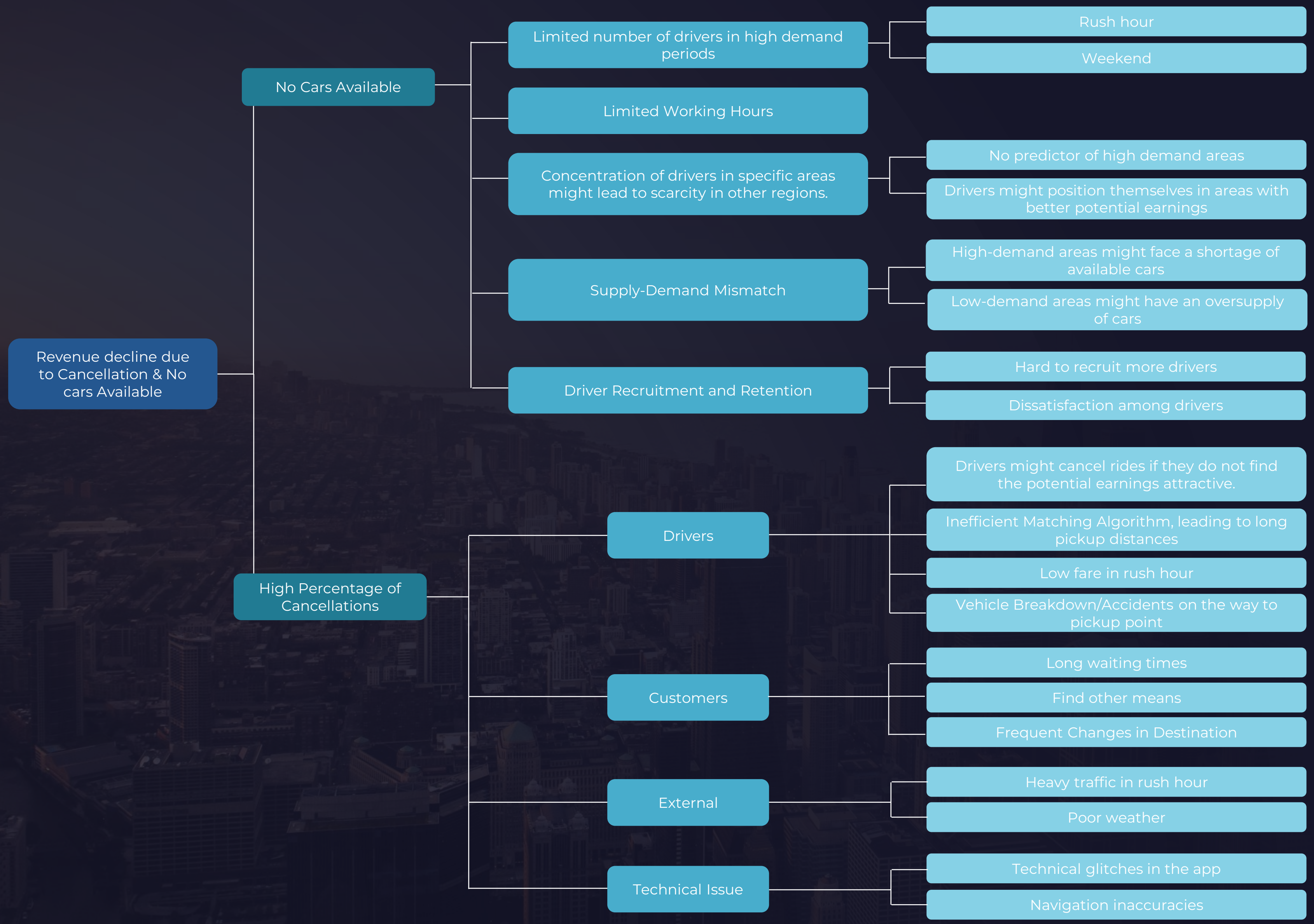
```
#Check datatype again
df.info()
df.head(10)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6745 entries, 0 to 6744
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Request id         6745 non-null   int64
 1   Pickup point       6745 non-null   object
 2   Driver id          4095 non-null   float64
 3   Status             6745 non-null   object
 4   Request timestamp  6745 non-null   datetime64[ns]
 5   Drop timestamp     2831 non-null   datetime64[ns]
dtypes: datetime64[ns](2), float64(1), int64(1), object(2)
memory usage: 316.3+ KB
```

Replace

Finding Hypothesis Problems and Root Causes

Creating Issue Tree by Mc Kinsey to find hypothesis problems



Overview

We can clearly see from this chart that No cars available is **more than** the number of trips cancelled, nearly double.

In order words, Uber is losing quite a lot of revenue mainly due to shortage of available cabs.

Chart 1. Status of request at City and Airport

Request by days

The initial assumption was that ride **numbers would rise on weekends**.

However, after graphing data for each day of the week, two key observations emerged:

- Saturdays and Sundays were not part of the dataset, preventing testing of the hypothesis.
- Regardless of the day of the week, minimal variation was seen in ride statuses. This suggests that the day of the week does not affect the existing supply-demand gap evident in earlier plots.

Chart 2. Number of Request at Airport and City by Day

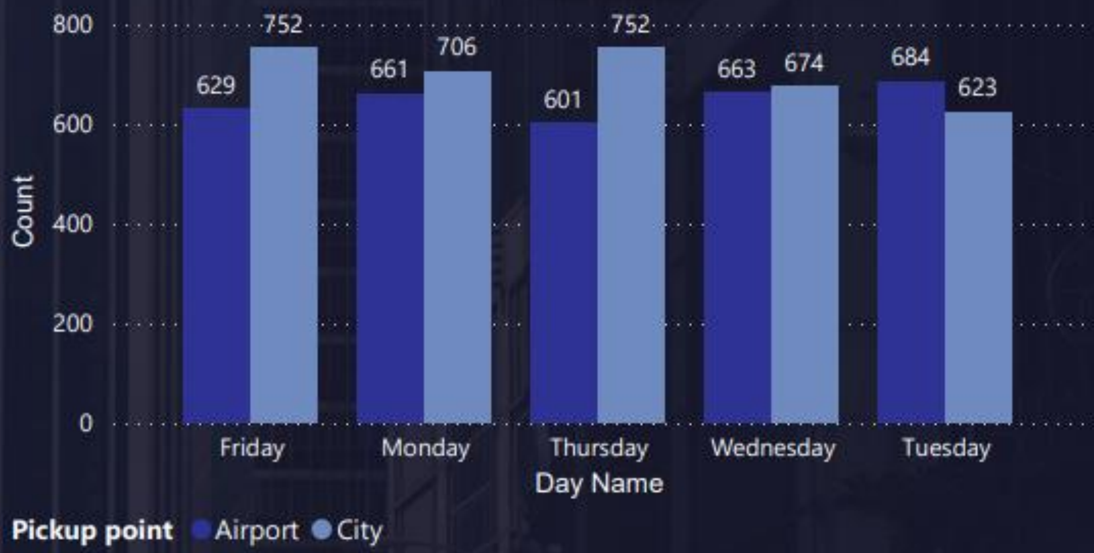
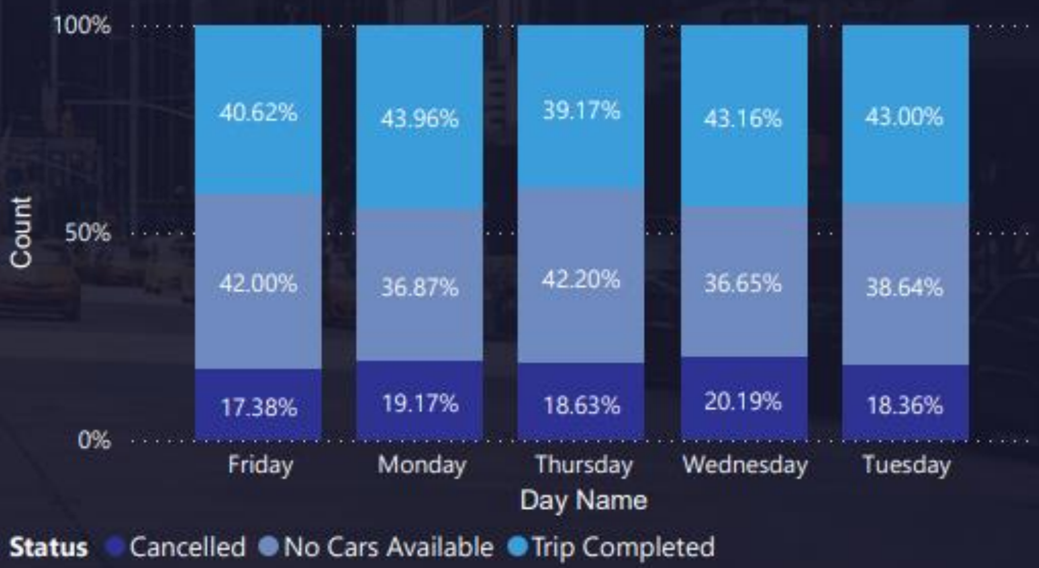


Chart 3. Status rate at City and Airport by Day



Hour peak

Peak cab request times: **5 AM - 9 AM and 5 PM - 9 PM**

Supply-demand gap exists during these hours, particularly in the evening. Evening hours **(5 PM - 9 PM)** indicate a potential car shortage.

Morning hours (5 AM - 9 AM) show relatively balanced supply and demand.

Despite balance, a noticeable gap persists during morning rush. What is the exact status of requests during 5 AM – 9 AM?

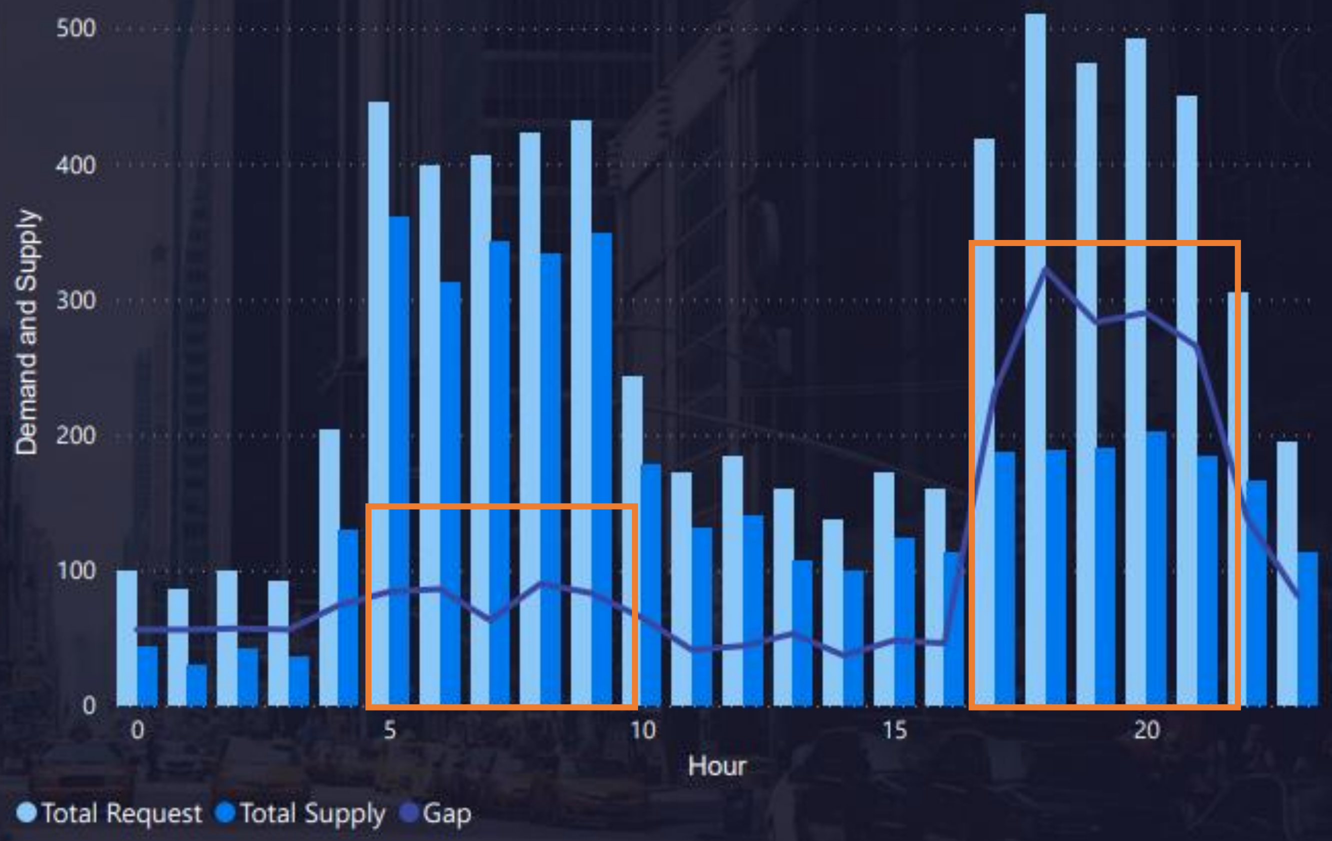


Chart 4. Demand and Supply Gap at Airport and City by Hour

Hour peak by pickup point

There is a high demand of cabs at **Airport** in the **5PM – 9PM**

There is a high demand of cabs at **City** in the **5AM – 9AM**

Chart 5. Demand and Supply Gap at Airport by Hour

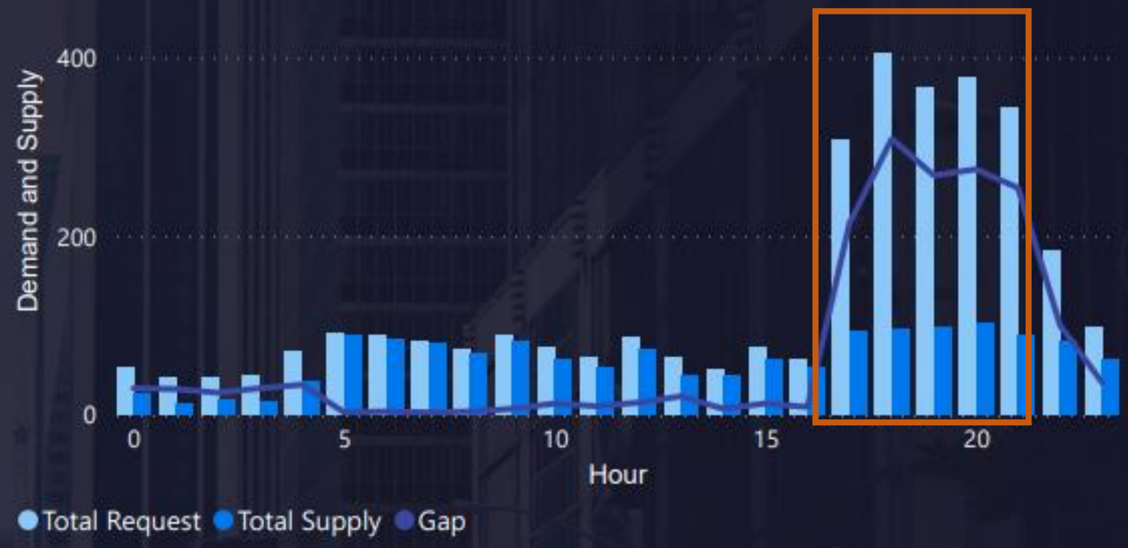
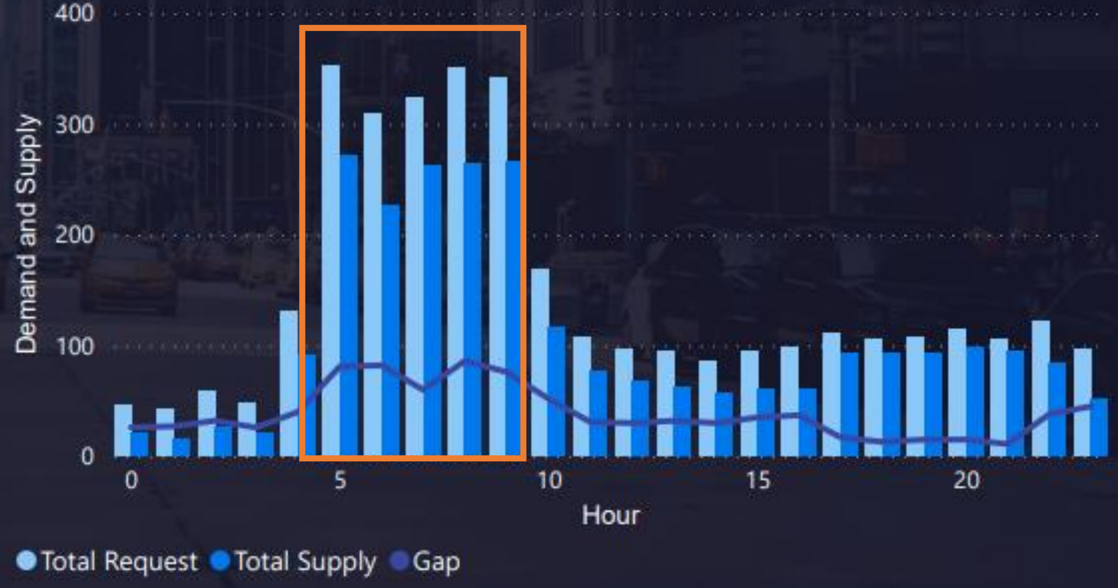


Chart 6. Demand and Supply Gap at City by Hour



Binning Time To Categories

Based on the timing of the request, requests are categorized into 5 homogeneous categories.

12AM – 5AM
Early Morning

5AM – 10AM
Morning Rush

10AM – 5PM
Day Time

5PM – 10PM
Evening Rush

10PM – 12AM
Late Night

Rush Hours and Unavailability Trends

There is a high demand of cabs at Airport in the Evening Rush (chart 5)

Number of no cars available is high during evening rush (5PM – 10PM)

Most of **No cars available** are from **Airport** pickup point in the **Evening Rush** and a smaller part from **City** in the **Morning Rush**.

Meanwhile,

There is a high demand of cabs at City in the Morning Rush (chart 6)

Number of cancellation is high during morning rush (5AM – 10AM)

Most of cabs **Cancelled** are from **city** pickup point in the **Morning Rush** .

Chart 7. Status of Request at Airport by Time Slot

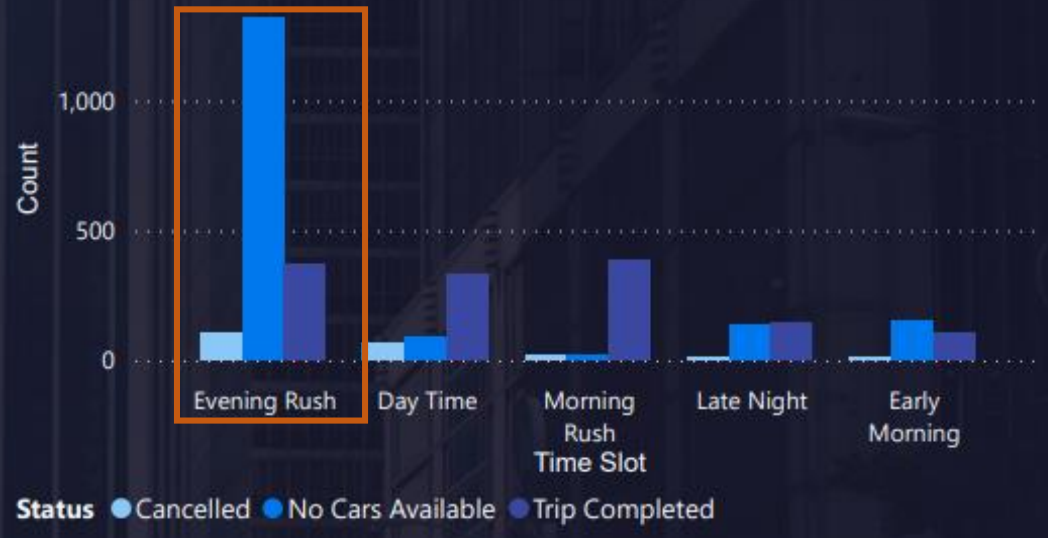
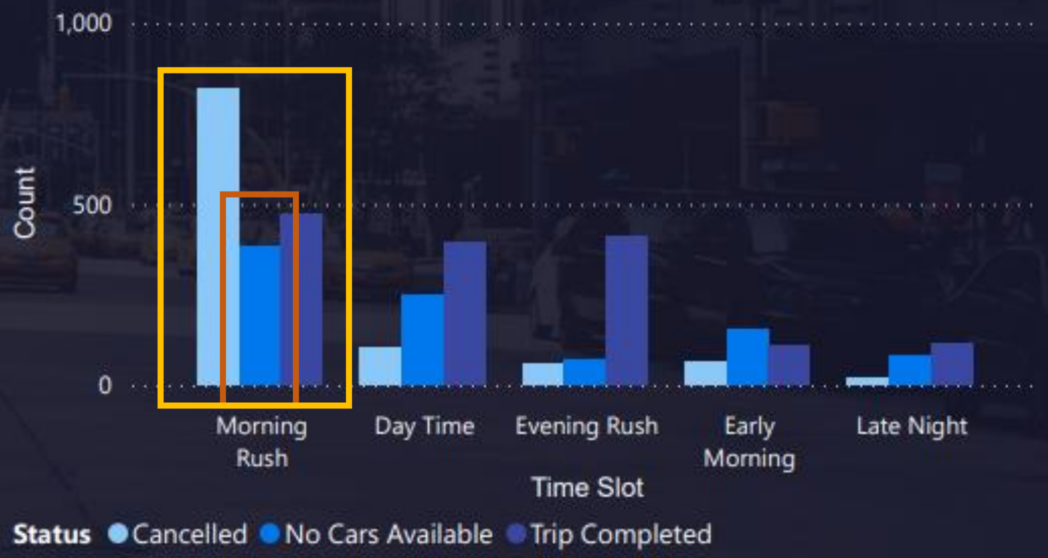


Chart 8. Status of Request at City by Time Slot



Conclusion

In general, total request is mostly higher than the total capacity of supply. Which lead to the results that Uber can not afford the total demand of booking rate in the market. The main reason for loss of revenue comes from lack of supply.

Our analysis identifies two critical areas requiring immediate attention:

- **City Morning Rush:** Substantial Cancellations Notable cancellations during the city's morning rush (5AM - 10AM) impact revenue. Rectifying this issue is vital to optimize service during peak hours.
- **Airport Evening Rush:** Severe Unavailability A substantial lack of available cars during the airport evening rush (5PM - 10PM) is evident, leading to potential revenue loss. Addressing this scarcity is crucial for meeting demand and maximizing earnings.

Focusing on these challenges promises to significantly mitigate revenue loss, enhancing operational efficiency and rider satisfaction.

Recommendation

City Morning Rush: Substantial Cancellations Root Cause: High cancellation rates during the city's morning rush impede service and hinder revenue generation.

Recommendations:

- Driver Incentives: Offer attractive bonuses to drivers for completing a certain number of rides during morning rush hours, reducing cancellations.
- Improved Matching Algorithm: Enhance the algorithm's efficiency to better match drivers and riders, reducing the likelihood of cancellations due to long wait times.
- Clear Fare Communication: Ensure transparent fare estimations for riders, reducing cancellations resulting from fare-related discrepancies.

Airport Evening Rush: Severe Unavailability Root Cause: The scarcity of available cars during the airport evening rush contributes to unmet demand and potential revenue loss.

Recommendations:

- Dynamic Fleet Allocation: Implement an algorithm that dynamically redistributes cars to high-demand areas during peak hours, ensuring better coverage.
- Incentivized Shifts: Introduce targeted incentives for drivers to extend their working hours during peak periods, augmenting vehicle availability.
- Surge Pricing Management: Employ surge pricing to balance supply and demand, encouraging drivers to cater to the heightened airport rush.
- Bonus Policies: Drivers can be given a bonus to complete a trip from the airport in the evening. This will ensure that the supply increases at the airport.

Recommendation (cont)

There are some recommendations expanding on the insights of Cancellation presented in the Issue Tree (Slide 7):

Driver Behavior:

- Driver Satisfaction Enhancement: Implement driver feedback mechanisms to gauge and improve their overall satisfaction, reducing cancellations and attrition.
- Real-time Earnings Display: Provide drivers with real-time earnings estimations for potential trips, addressing cancellations due to unattractive earnings.

Technical Issues:

- App Optimization: Continuously refine the app's performance to minimize technical glitches and navigation inaccuracies.
- Enhanced GPS Integration: Ensure accurate GPS data for seamless matching and routing, reducing cancellations arising from navigation issues.

External Factors:

- Weather-Related Strategies: During inclement weather, consider implementing temporary incentives for drivers to operate, counteracting cancellations due to adverse conditions.
- Traffic Management: Integrate real-time traffic data into the algorithm to provide more accurate ETAs and mitigate cancellations caused by heavy traffic.

Customer Behavior:

- Transparent Wait Time Information: Display estimated wait times clearly to manage rider expectations and reduce cancellations due to perceived long waits.

Thank you

For Your Listening

NGUYEN THACH THAO
NGUYEN THI TUYET TRINH