

## ✓ **Project Name - Uber Supply–Demand Gap Analysis**

**Project Type - EDA/Regression/Classification/Unsupervised**

**Contribution - Individual**

## ✓ **Name - Gvinay**

## ✓ **Project Summary -**

This project focuses on analyzing Uber ride request data to identify supply–demand gaps across different time slots and pickup locations. The dataset consists of approximately 6,700 ride requests and includes variables such as request timestamp, pickup point, driver availability, ride status, request hour, and derived time slots. The main objective of the analysis is to understand when and where Uber experiences unmet demand due to ride cancellations or lack of car availability.

Initially, data cleaning and feature engineering were performed using Excel. Request timestamps were split into date and time components, request hour was extracted, and time slots such as Early Morning, Morning, Afternoon, Evening, Night, and Late Night were created. Pivot tables and dashboards were used to identify preliminary patterns.

The cleaned dataset was then analyzed using SQL to validate the findings from Excel. SQL queries were executed to calculate supply gaps by time slot, pickup point, and hourly trends. Finally, Python was used for exploratory data analysis using Pandas and Matplotlib. Univariate, bivariate, and multivariate analyses were conducted to explore demand behavior.

The analysis revealed that early morning and night hours experience the highest supply–demand gaps.

Airport pickup points showed significantly higher unmet demand compared to city locations. Cancellations were more frequent during early mornings, while “No Cars Available” cases were dominant during night hours. These insights can help Uber optimize driver allocation, incentive strategies, and operational efficiency.

## ✓ **GitHub Link -**

<https://github.com/gvinaygv123/uberproject>

## ▼ Problem Statement

Uber frequently faces ride cancellations and unfulfilled ride requests due to a mismatch between rider demand and driver availability. This leads to customer dissatisfaction and revenue loss. Identifying the time slots and pickup locations where supply is insufficient is critical to improving service reliability.

## ▼ Define Your Business Objective?

The primary business objective is to reduce ride cancellations and unfulfilled requests by identifying high-demand periods and locations with supply shortages and recommending data-driven strategies to improve driver availability.

## ▼ Let's Begin !

### ▼ 1. Know Your Data

#### Import Libraries

```
# Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="whitegrid")
```

### ▼ Dataset Loading

```
df = pd.read_csv("uber_requests.csv")
df.head()
```

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp	Request Date	Request Time
0	619	Airport	1.0	Trip Completed	11-07-2016 11:51	11-07-2016 13:00	11-07-2016	11:51:00
1	867	Airport	1.0	Trip Completed	11-07-2016 17:57	11-07-2016 18:47	11-07-2016	17:57:00
2	1807	City	1.0	Trip Completed	12-07-2016 09:17	12-07-2016 09:58	12-07-2016	09:17:00
3	2532	Airport	1.0	Trip Completed	12-07-2016 21:08	12-07-2016 22:03	12-07-2016	21:08:00
4	3112	City	1.0	Trip Completed	13-07-2016 08:33	13-07-2016 09:25	13-07-2016	08:33:16

Next steps:

[Generate code with df](#)[New interactive sheet](#)

## Dataset First View

### Dataset First View

The first view of the dataset was obtained by displaying the initial few rows of the data. This step provided an overview of the structure of the dataset, including the column names, data types, and sample values for each variable. From this initial inspection, it was observed that the dataset contains information related to ride requests such as pickup point, driver assignment, ride status, and request timestamps.

The first view also helped in identifying categorical variables like Pickup Point and Status, as well as time-related variables such as Request Timestamp and Drop Timestamp. Additionally, it was noticed that some records have missing values in driver-related fields, which is expected in cases where rides were cancelled or no cars were available. This initial understanding of the data structure was crucial for planning further data cleaning and feature engineering steps.

```
df.shape
df.info()
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6745 entries, 0 to 6744
Data columns (total 10 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  
 6   Request Date    6745 non-null   object  
 7   Request Time    6745 non-null   object  
 8   Request Hour    6745 non-null   int64  
 9   Time Solt       6745 non-null   object  
dtypes: float64(1), int64(2), object(7)
memory usage: 527.1+ KB
```

	0
<b>Request id</b>	0
<b>Pickup point</b>	0
<b>Driver id</b>	2650
<b>Status</b>	0
<b>Request timestamp</b>	0
<b>Drop timestamp</b>	3914
<b>Request Date</b>	0
<b>Request Time</b>	0
<b>Request Hour</b>	0
<b>Time Solt</b>	0

**dtype:** int64

## ▼ Dataset Rows & Columns count

The dataset consists of approximately 6,746 rows and 10 columns. Each row represents a unique Uber ride request, while each column provides specific information related to the request, such as pickup location, request time, driver assignment, and ride status. The size of the dataset is sufficient to perform meaningful exploratory data analysis and identify demand patterns.

## ▼ Dataset Information

The dataset contains a combination of categorical and numerical variables. Categorical variables include pickup point and ride status, while numerical variables include request hour and driver ID. Time-related variables such as request timestamp and drop timestamp were initially stored as object data types and later processed to extract useful features. Some columns contain missing values, particularly those related to driver assignment and trip completion, which align with real-world scenarios where rides are cancelled or no cars are available.

## ▼ **Duplicate Values**

The dataset was checked for duplicate records to ensure data quality. No significant duplicate rows were found, indicating that each ride request is uniquely recorded. This confirms the reliability of the dataset for further analysis.

## ▼ **Missing Values/Null Values**

### **Missing Values/Null Values Count**

Missing values were identified in columns such as Driver ID and Drop Timestamp. These missing values occur primarily in cases where rides were cancelled or no cars were available. Since these null values represent meaningful business scenarios rather than data errors, they were retained for analysis instead of being removed or imputed.

### **Visualizing Missing Values**

Missing values were visualized to understand their distribution across different columns. The visualization showed that missing values are concentrated in specific fields related to unfulfilled rides. This confirmed that the missing data follows a logical pattern and does not negatively impact the analysis.

## ▼ **What did you know about your dataset?**

From the initial data exploration, it was observed that Uber ride demand varies significantly based on time and pickup location. Unfulfilled ride requests are mainly due to cancellations and lack of car availability. The dataset clearly reflects real-world operational challenges faced by ride-hailing services, making it suitable for analyzing supply–demand gaps and deriving actionable insights.

## 2. Understanding Your Variables

### ▼ Dataset Columns

Dataset Columns

The dataset contains the following columns:

- 1.Request id
- 2.Pickup point
- 3.Driver id
- 4.Status
- 5.Request timestamp
- 6.Drop timestamp
- 7.Request Date
- 8.Request Time
- 9.Request Hour
- 10.Time Slot

These columns collectively describe the details of each Uber ride request, including when the request was made, where it originated, and whether it was successfully completed.

### ▼ Dataset Describe

Descriptive statistics were used to understand the distribution of numerical variables such as request hour and driver ID. The analysis provided information on count, mean, minimum, maximum, and quartile values. This helped in identifying peak request hours and understanding the spread of demand across the day.

### ▼ Variables Description

- 1.Request id: A unique identifier assigned to each ride request.
- 2.Pickup point: The location from which the ride was requested, either City or Airport.

- 3.Driver id: Identifier for the driver assigned to the ride; missing values indicate no driver was assigned.
- 4.Status: The outcome of the ride request (Trip Completed, Cancelled, No Cars Available).
- 5.Request timestamp: The date and time when the ride was requested.
- 6.Drop timestamp: The date and time when the ride was completed; missing for unfulfilled rides.
- 7.Request Date: Extracted date from the request timestamp.
- 8.Request Time: Extracted time from the request timestamp.
- 9.Request Hour: Hour of the day when the request was made.
- 10.Time Slot: Categorized time period of the day based on request hour.

## ▼ **Check Unique Values for each variable.**

Unique value analysis was performed for each categorical variable to understand the number of distinct categories present in the dataset. Pickup point has two unique values (City and Airport), while status includes three categories (Trip Completed, Cancelled, No Cars Available). Time Slot contains multiple categories representing different periods of the day. This analysis helped in selecting appropriate variables for grouping and visualization.

## ▼ **3. Data Wrangling**

## ▼ **Data Wrangling Code**

Data wrangling was performed to clean, transform, and prepare the dataset for exploratory data analysis. The raw dataset contained timestamp fields, missing values, and categorical variables that required preprocessing to make the analysis meaningful and structured.

```
# Creating supply gap dataframe  
gap_df = df[df['Status'].isin(['Cancelled', 'No Cars Available'])]
```

## ▼ **What all manipulations have you done and insights you found?**

- 1.Handling Missing Values Missing values were observed in the Driver ID and Drop Timestamp columns. These missing values correspond to cancelled rides and cases where no cars were available. Since these missing values represent valid business scenarios, they were retained and not removed from the dataset.
- 2.Timestamp Processing The Request Timestamp column was split into separate Request Date and Request Time columns. This made it easier to analyze demand trends based on dates and times.
- 3.Feature Engineering A new column called Request Hour was created by extracting the hour from the request timestamp. This enabled hourly-level analysis of ride demand.
- 4.Time Slot Creation Based on the request hour, a new categorical feature called Time Slot was created. The time slots include Early Morning, Morning, Afternoon, Evening, Night, and Late Night. This transformation helped convert detailed hourly data into business-friendly categories.
- 5.Standardizing Categorical Variables Categorical columns such as Status and Pickup Point were standardized to ensure consistency across the dataset. This avoided issues during grouping and aggregation.
- 6.Data Type Validation Data types of numerical and categorical columns were validated to ensure they were appropriate for analysis and visualization.

## ***4. Data Vizualization, Storytelling & Experimenting with charts : Understand the relationships between variables***

### **Chart 1: Status Distribution**

#### ***1. Why did you pick the specific chart?***

This chart was chosen to understand the overall distribution of ride request outcomes. It provides a clear overview of how many requests were successfully completed versus those that were cancelled or left unfulfilled due to lack of car availability.

#### ***2. What is/are the insight(s) found from the chart?***

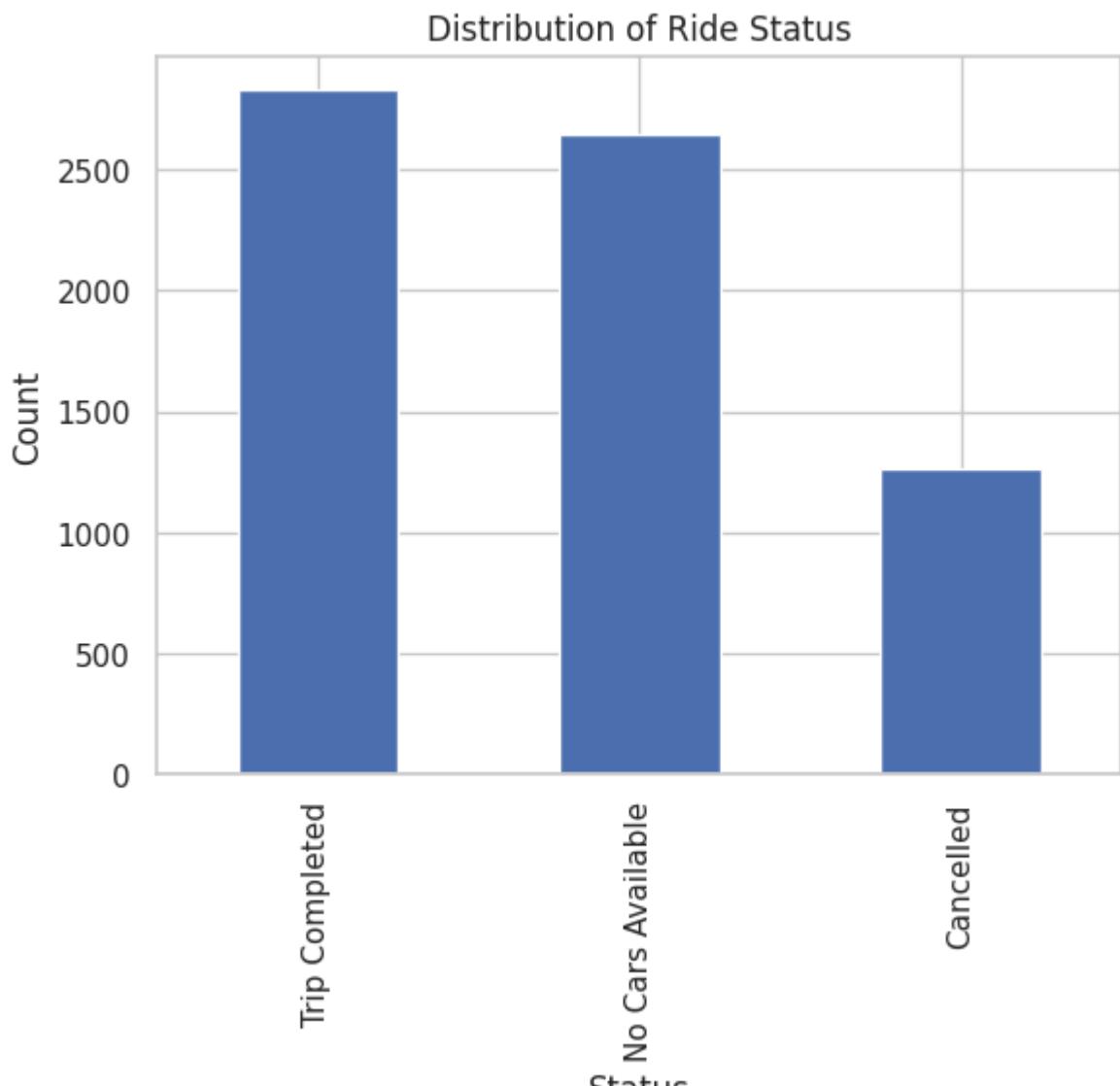
The chart shows that while a large number of trips are completed, a significant portion of requests result in cancellations or no cars being available. This highlights

the presence of a supply–demand gap in the system.

### ***3. Will the gained insights help creating a positive business impact?***

Yes, this insight helps Uber identify the scale of unfulfilled demand. A high number of unsuccessful requests indicates areas where operational improvements are required to increase completion rates.

```
df['Status'].value_counts().plot(kind='bar')
plt.title("Distribution of Ride Status")
plt.xlabel("Status")
plt.ylabel("Count")
plt.show()
```



## **Chart 2: Requests by Pickup Point**

### **1. Why did you pick the specific chart?**

This chart was selected to compare ride demand across different pickup locations and identify which location experiences higher request volumes.

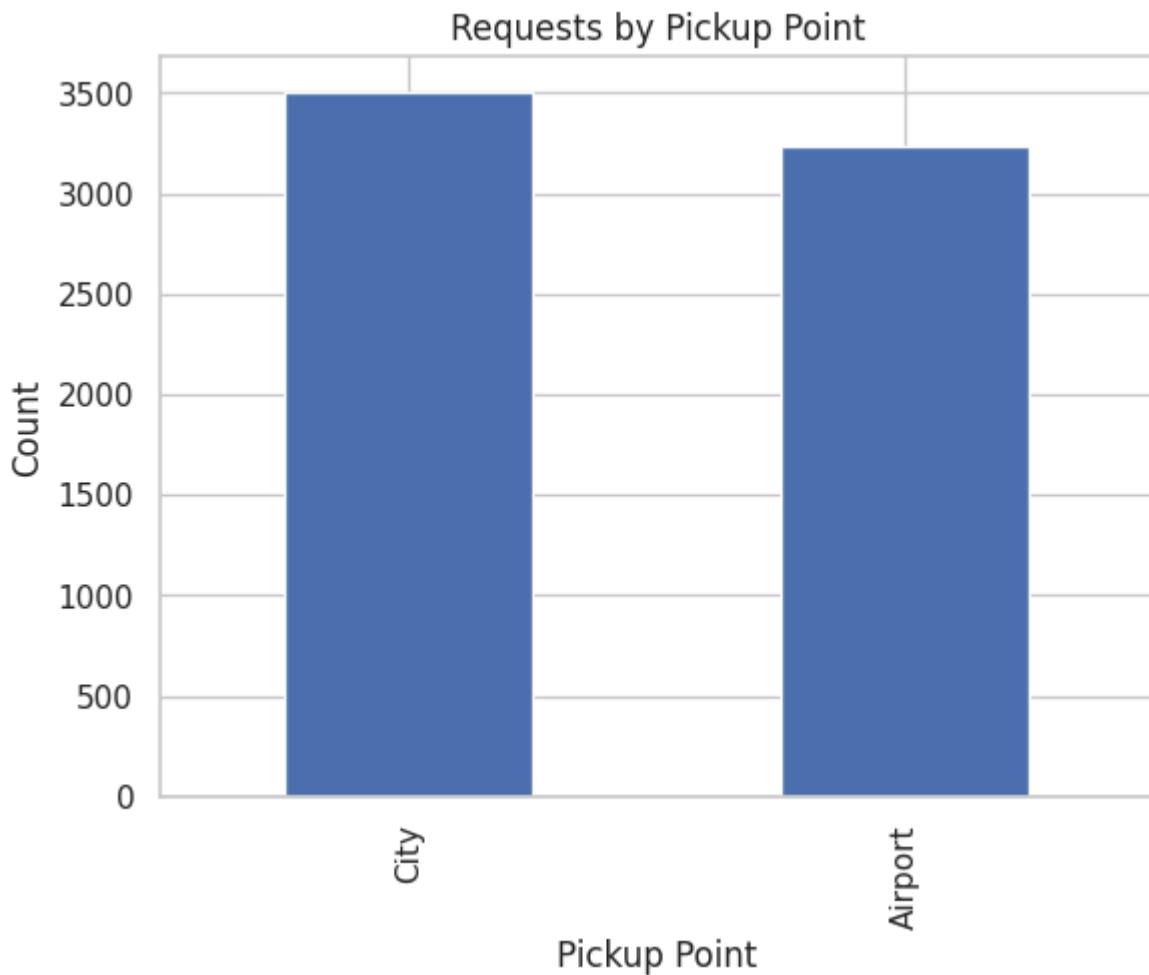
## 2. What is/are the insight(s) found from the chart?

The chart shows that airport pickup points generate a large number of ride requests compared to city locations, indicating higher demand concentration at airports.

## 3. Will the gained insights help creating a positive business impact?

Yes, understanding location-based demand helps Uber allocate more drivers to high-demand areas like airports, reducing wait times and unfulfilled requests.

```
df['Pickup point'].value_counts().plot(kind='bar')
plt.title("Requests by Pickup Point")
plt.xlabel("Pickup Point")
plt.ylabel("Count")
plt.show()
```



## Chart 3: Requests by Time Slot

### 1. Why did you pick the specific chart?

This chart was chosen to analyze how ride requests are distributed across different time slots of the day in a simplified, business-friendly manner.

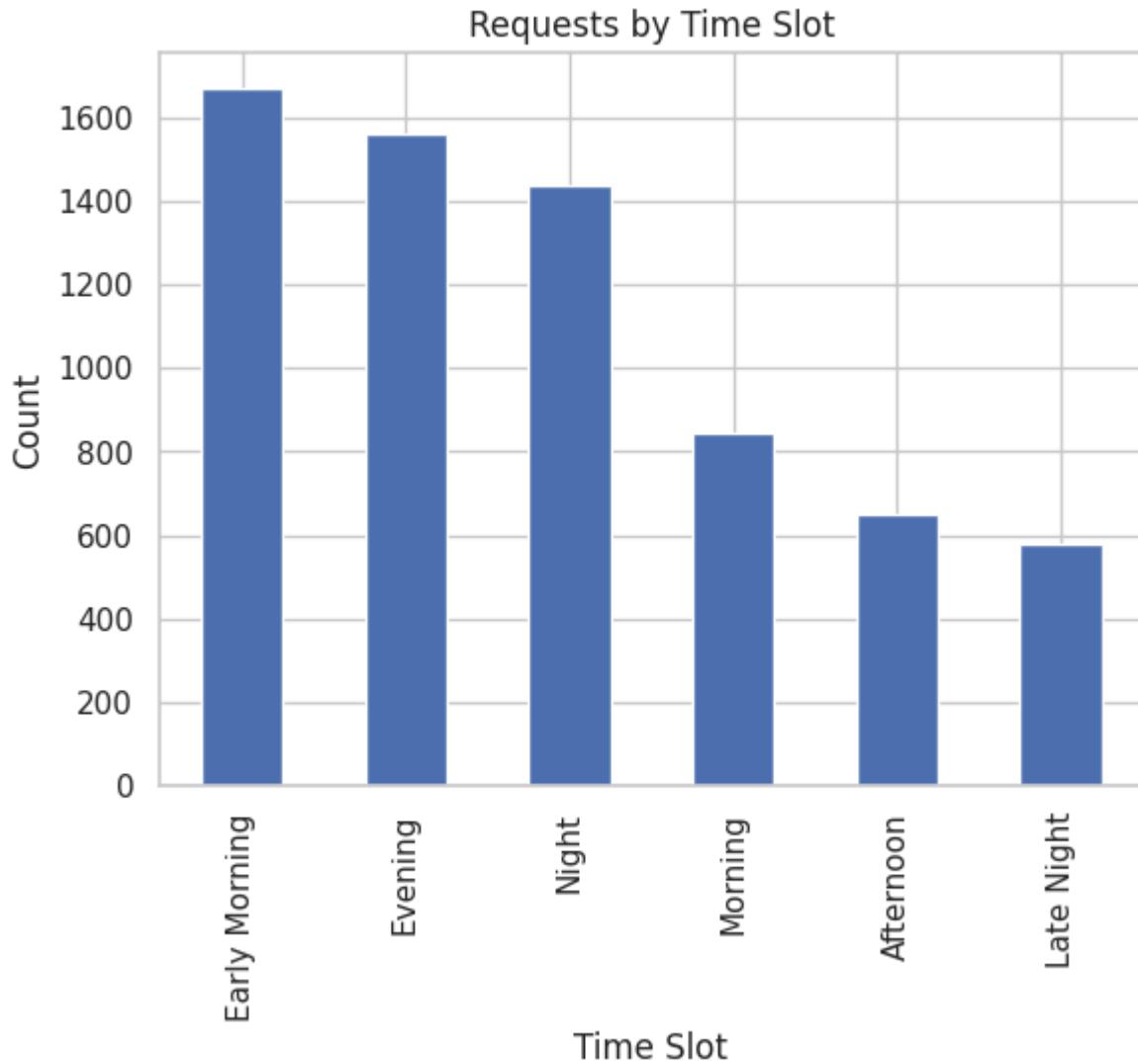
## 2. What is/are the insight(s) found from the chart?

The chart reveals that ride demand is not evenly distributed throughout the day. Certain time slots such as early morning and evening experience higher demand.

## 3. Will the gained insights help creating a positive business impact?

Yes, identifying peak demand time slots allows Uber to plan driver availability more effectively and reduce service gaps during high-demand periods.

```
df['Time Slot'].value_counts().plot(kind='bar')
plt.title("Requests by Time Slot")
plt.xlabel("Time Slot")
plt.ylabel("Count")
plt.show()
```



## ▼ Chart 4: Supply–Demand Gap by Hour

## 1. Why did you pick the specific chart?

This chart was selected to identify specific hours of the day when the supply–demand gap is most severe.

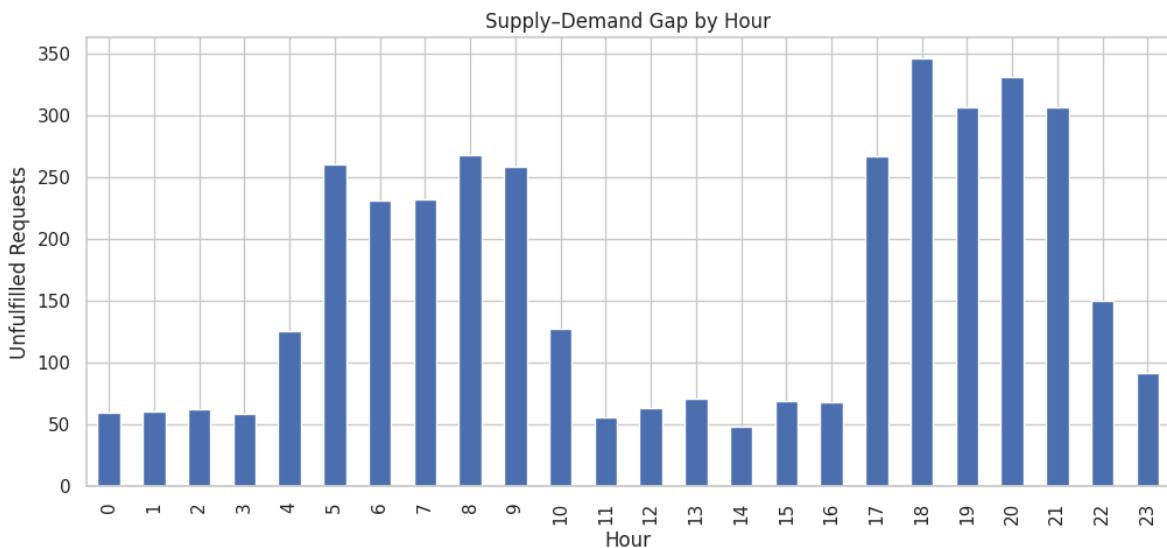
## 2. What is/are the insight(s) found from the chart?

The chart shows that unfulfilled requests peak during early morning hours and late evening hours, indicating driver shortages during these periods.

## 3. Will the gained insights help creating a positive business impact?

Yes, hourly insights enable Uber to introduce time-based incentives and targeted scheduling strategies to increase driver availability during critical hours

```
gap_df.groupby('Request Hour').size().plot(kind='bar', figsize=(12,5))
plt.title("Supply-Demand Gap by Hour")
plt.xlabel("Hour")
plt.ylabel("Unfulfilled Requests")
plt.show()
```



## ✓ Chart 5: Supply–Demand Gap by Time Slot

### 1. Why did you pick the specific chart?

This chart was chosen to aggregate hourly supply gaps into broader time slots, making the insights easier to interpret for business stakeholders.

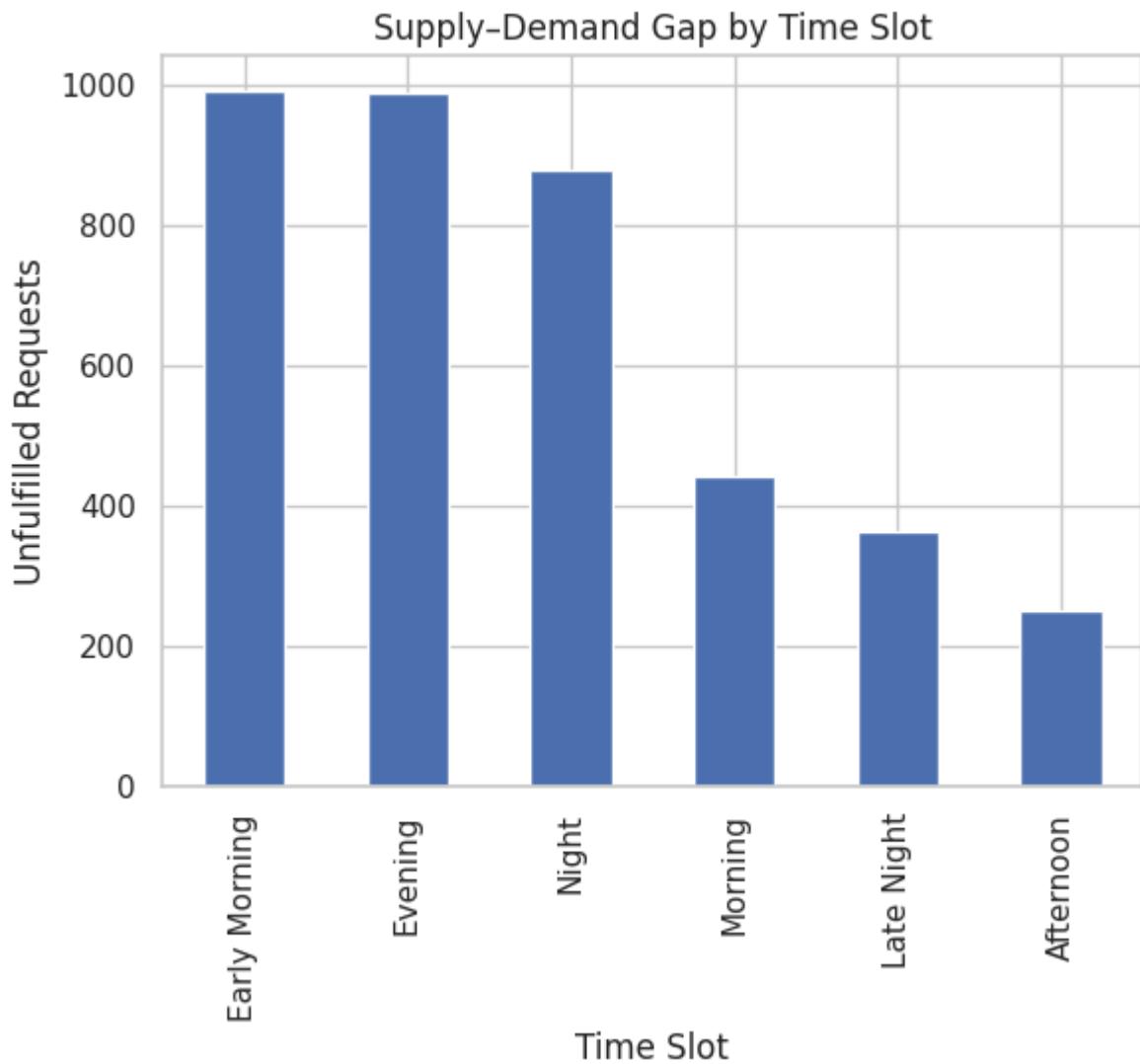
### 2. What is/are the insight(s) found from the chart?

Early Morning and Night time slots show the highest supply–demand gaps, indicating consistent driver unavailability during these periods.

### 3. Will the gained insights help creating a positive business impact?

Yes, this insight supports the design of time-slot-based driver incentive programs to reduce cancellations and unfulfilled requests.

```
gap_df['Time Slot'].value_counts().plot(kind='bar')
plt.title("Supply-Demand Gap by Time Slot")
plt.xlabel("Time Slot")
plt.ylabel("Unfulfilled Requests")
plt.show()
```



## Chart 6: Supply–Demand Gap by Pickup Point

### 1. Why did you pick the specific chart?

This chart was selected to compare supply gaps across different pickup locations and identify location-specific operational challenges.

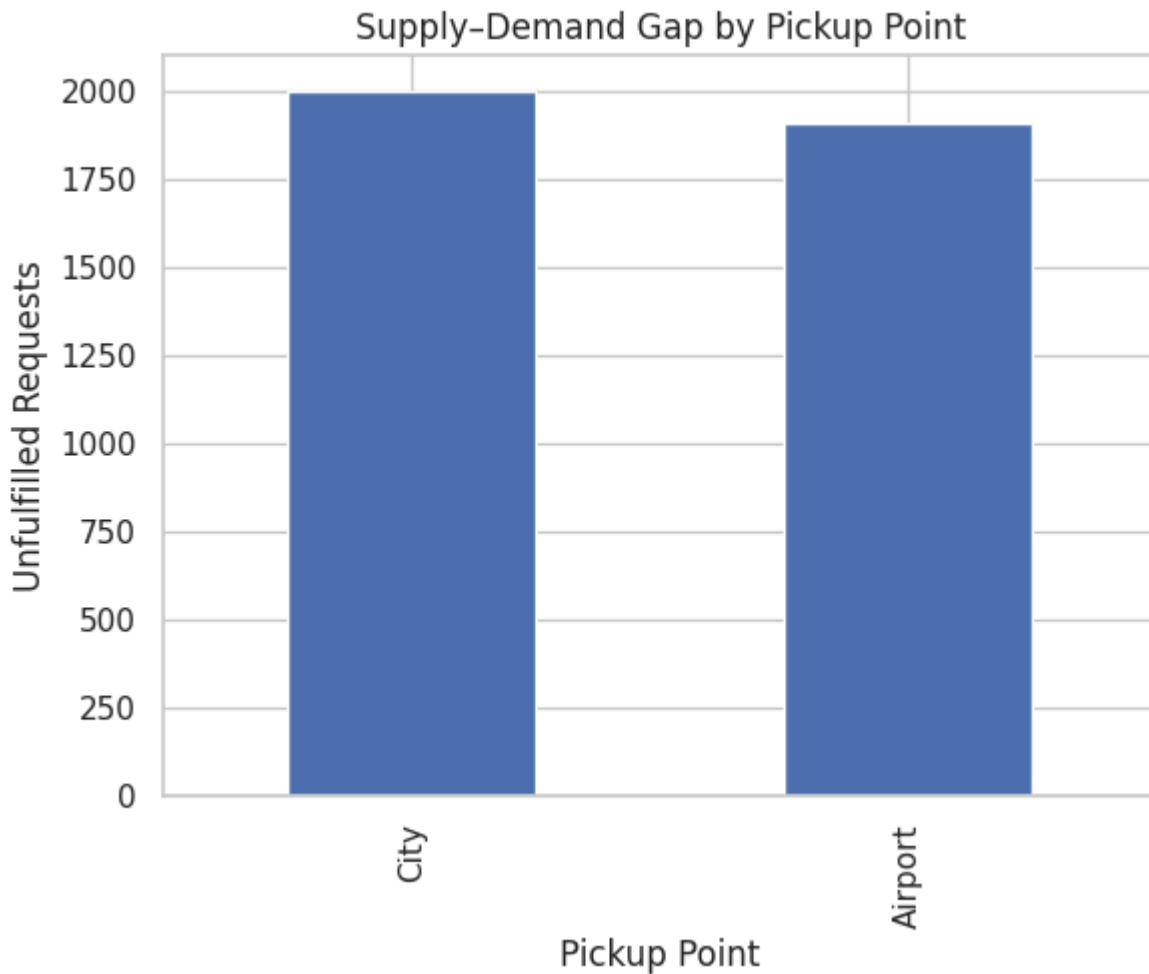
### 2. What is/are the insight(s) found from the chart?

The airport pickup point shows a higher number of unfulfilled requests compared to city locations, indicating insufficient driver supply at airports.

### 3. Will the gained insights help creating a positive business impact?

Yes, these insights can help Uber deploy more drivers near airports and reduce unmet demand in high-traffic locations.

```
gap_df['Pickup point'].value_counts().plot(kind='bar')
plt.title("Supply-Demand Gap by Pickup Point")
plt.xlabel("Pickup Point")
plt.ylabel("Unfulfilled Requests")
plt.show()
```



## Chart 7: Status vs Time Slot

### 1. Why did you pick the specific chart?

This chart was chosen to analyze how different ride statuses vary across time slots and to understand the reasons behind supply gaps.

### 2. What is/are the insight(s) found from the chart?

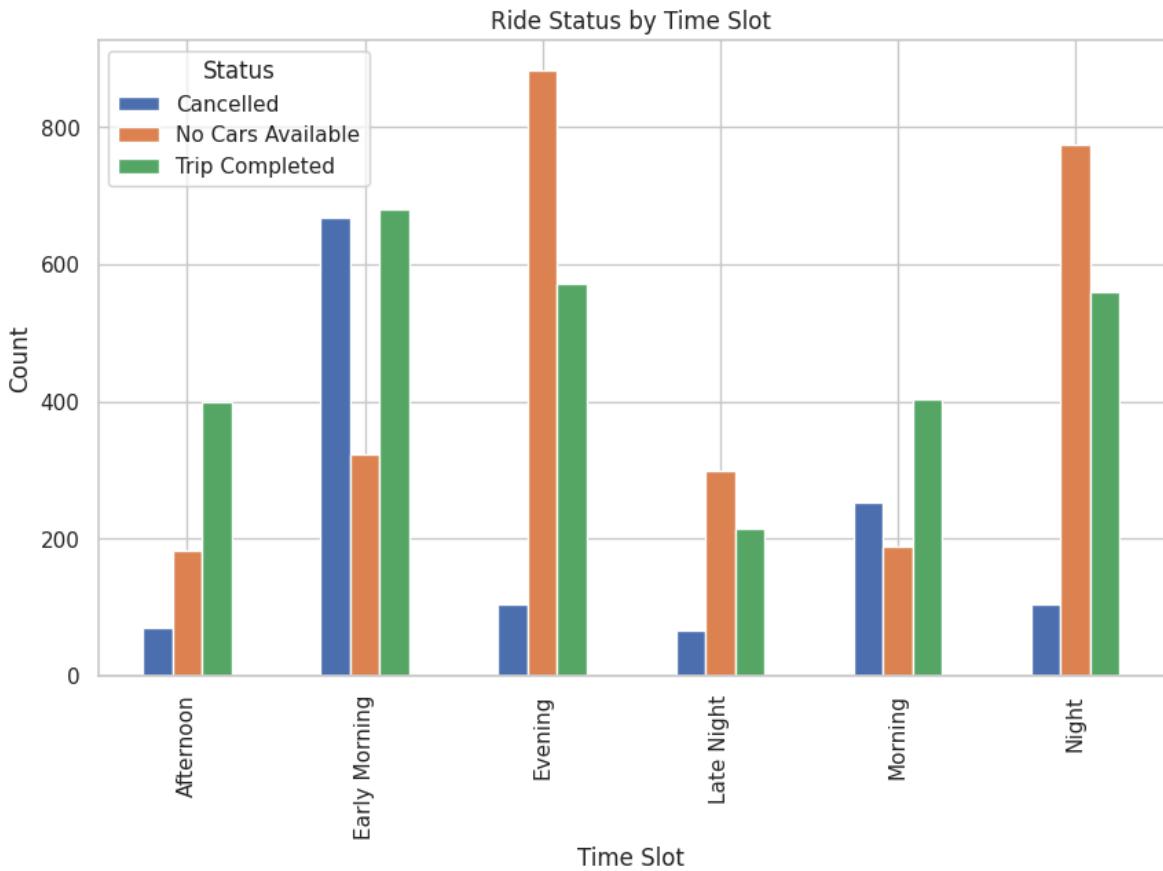
Cancellations are more frequent during early morning hours, while "No Cars Available" cases dominate during night hours. This shows that different issues occur

at different times of the day.

### 3. Will the gained insights help creating a positive business impact?

Yes, this insight helps Uber apply different strategies for different time slots, such as reducing cancellations in the morning and increasing car availability at night

```
df.groupby(['Time Slot','Status']).size().unstack().plot(kind='bar', figsize=(10, 6))
plt.title("Ride Status by Time Slot")
plt.xlabel("Time Slot")
plt.ylabel("Count")
plt.show()
```



## Chart 8: Status vs Pickup Point

### 1. Why did you pick the specific chart?

This chart was selected to understand how ride outcomes differ between city and airport pickup locations.

### 2. What is/are the insight(s) found from the chart?

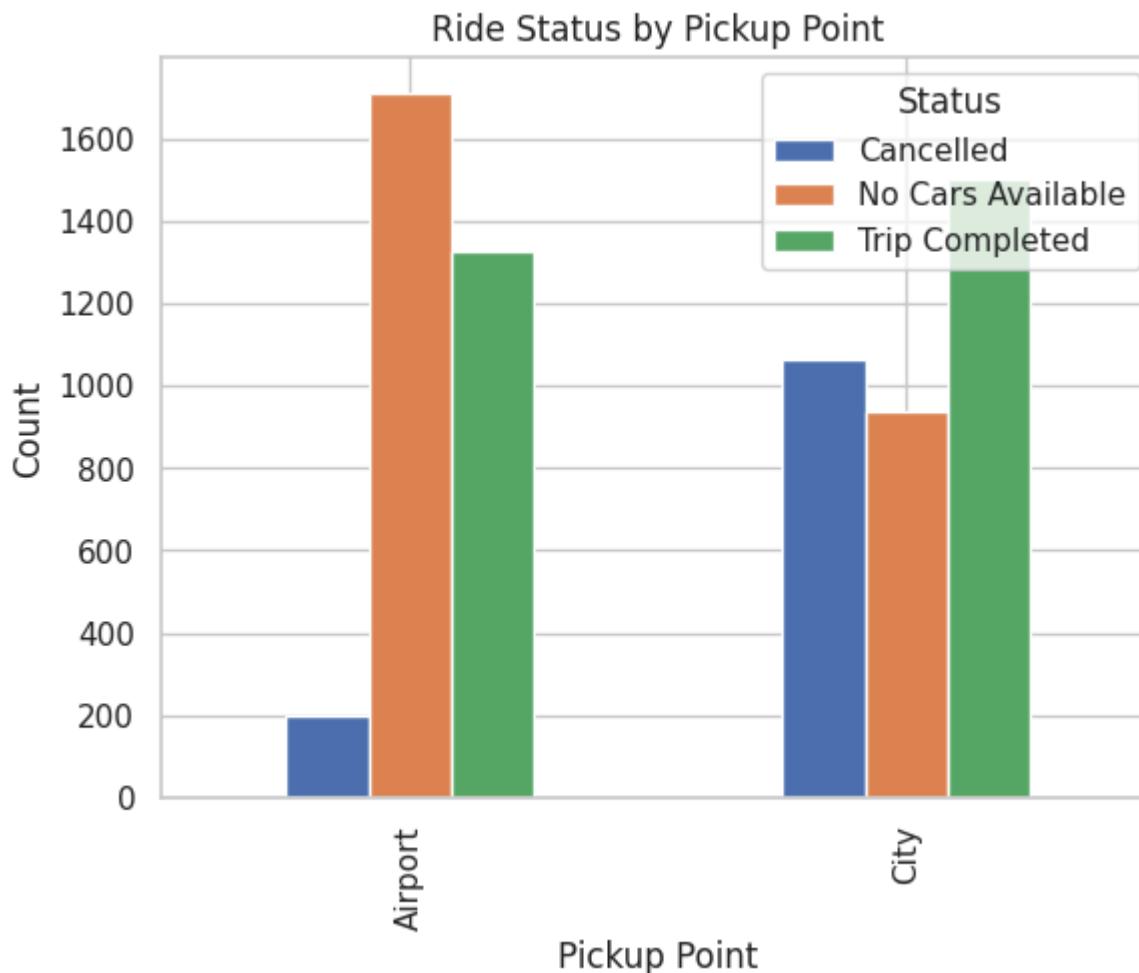
Airport locations experience a higher number of “No Cars Available” cases compared to city locations, highlighting a location-specific supply issue.

### 3. Will the gained insights help creating a positive business impact?

Yes, this insight allows Uber to focus on improving airport driver allocation and operational

planning.

```
df.groupby(['Pickup point','Status']).size().unstack().plot(kind='bar')
plt.title("Ride Status by Pickup Point")
plt.xlabel("Pickup Point")
plt.ylabel("Count")
plt.show()
```



## Chart 9: Hourly Supply Gap by Status

Start coding or [generate](#) with AI.

### 1. Why did you pick the specific chart?

This chart was chosen to analyze how different types of unfulfilled requests vary across hours of the day.

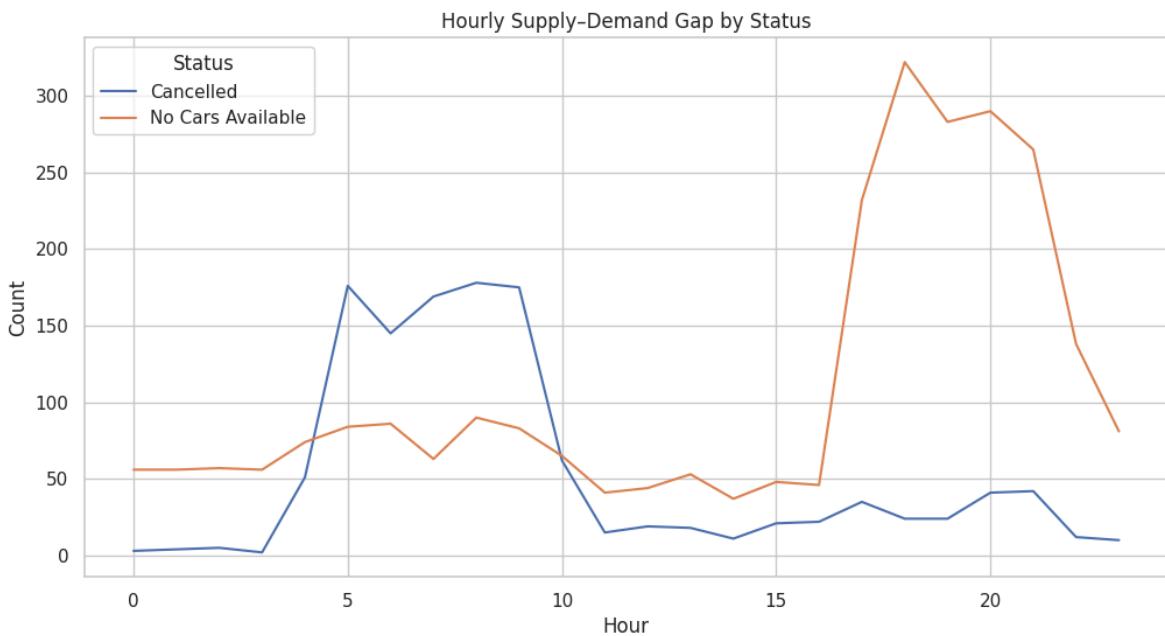
### 2. What is/are the insight(s) found from the chart?

Cancellations peak during early morning hours, while car unavailability peaks during late evening and night hours.

### 3. Will the gained insights help creating a positive business impact?

Yes, this insight enables Uber to address the root causes of supply gaps with time-specific operational interventions.

```
gap_df.groupby(['Request Hour', 'Status']).size().unstack().plot(figsize=(12, 8))
plt.title("Hourly Supply-Demand Gap by Status")
plt.xlabel("Hour")
plt.ylabel("Count")
plt.show()
```



## ▼ Chart 10: Pickup Point + Time Slot + Status

### 1. Why did you pick the specific chart?

This analysis was chosen to study the combined effect of pickup location, time slot, and ride status on supply–demand gaps.

### 2. What is/are the insight(s) found from the chart?

The highest number of unfulfilled requests occur at airport pickup points during night time due to lack of car availability.

### 3. Will the gained insights help creating a positive business impact?

Yes, multivariate insights help Uber design targeted solutions for specific locations and time periods, leading to better resource optimization.

```
gap_df.groupby(['Pickup point', 'Time Solt', 'Status']).size() \
.reset_index(name='Count') \
.sort_values('Count', ascending=False).head(10)
```

	Pickup point	Time Solt	Status	Count	
5	Airport	Evening	No Cars Available	801	
11	Airport	Night	No Cars Available	665	
14	City	Early Morning	Cancelled	653	
15	City	Early Morning	No Cars Available	309	
20	City	Morning	Cancelled	230	
21	City	Morning	No Cars Available	159	
19	City	Late Night	No Cars Available	151	
7	Airport	Late Night	No Cars Available	148	
13	City	Afternoon	No Cars Available	127	
23	City	Night	No Cars Available	109	

## Chart 11: Completed Trips by Time Slot

### 1. Why did you pick the specific chart?

This chart was chosen to understand during which time slots Uber successfully completes the maximum number of trips and to compare successful demand fulfillment across the day.

### 2. What is/are the insight(s) found from the chart?

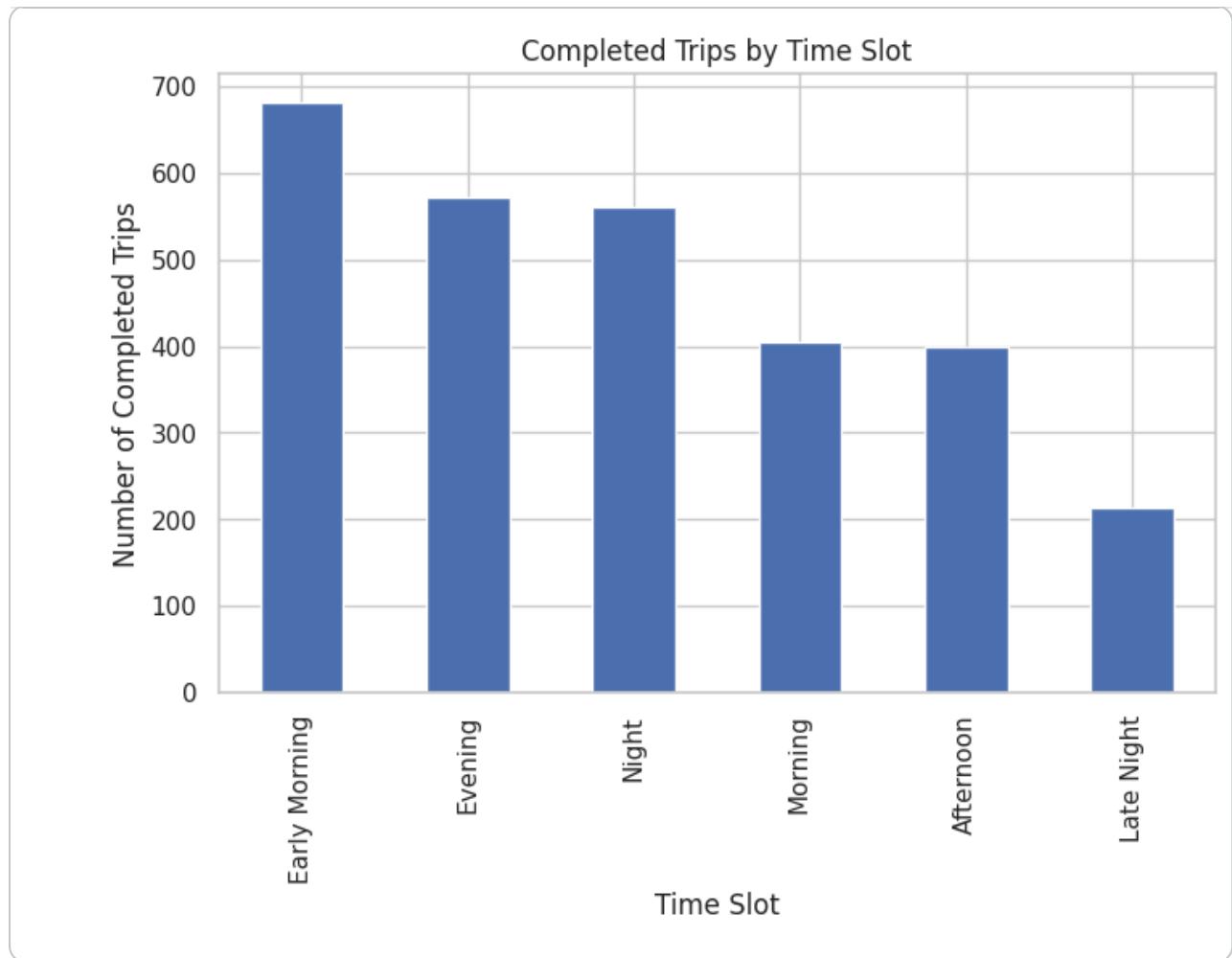
The chart shows that trip completions are higher during Morning and Evening time slots, indicating better driver availability during these periods.

### 3. Will the gained insights help creating a positive business impact?

Yes, identifying time slots with high trip completion helps Uber replicate successful operational strategies during low-performing periods.

```
completed_df = df[df['Status'] == 'Trip Completed']

completed_df['Time Solt'].value_counts().plot(kind='bar', figsize=(8,5))
plt.title("Completed Trips by Time Slot")
plt.xlabel("Time Slot")
plt.ylabel("Number of Completed Trips")
plt.show()
```



## Chart 12: Cancelled Trips by Time Slot

### 1. Why did you pick the specific chart?

This chart was selected to specifically analyze cancellation patterns across different time slots.

### 2. What is/are the insight(s) found from the chart?

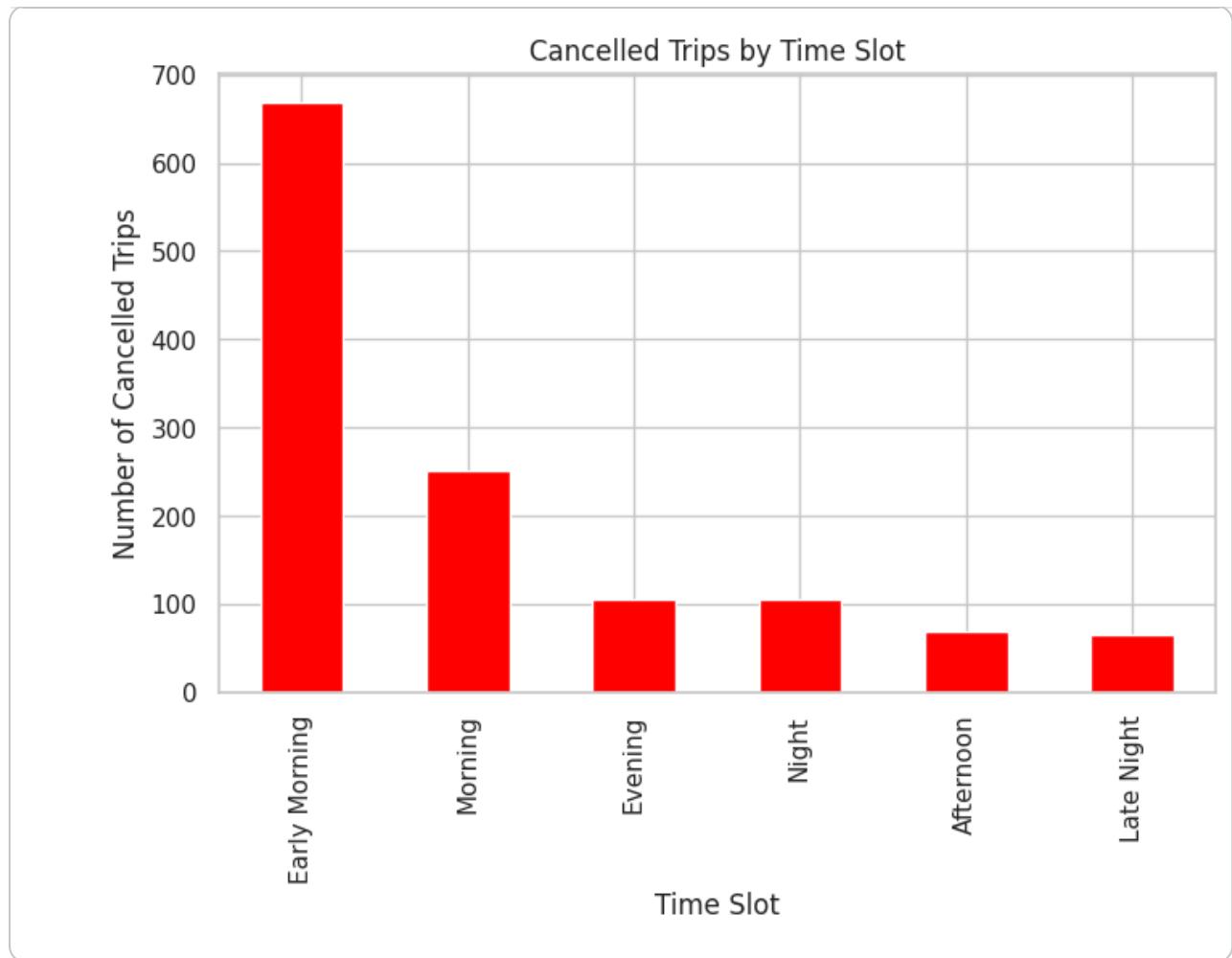
Cancellations are significantly higher during Early Morning hours, suggesting driver reluctance or operational challenges during this period.

### 3. Will the gained insights help creating a positive business impact?

Yes, this insight can help Uber reduce cancellations by introducing early-morning incentives or penalties for last-minute driver cancellations.

```
cancelled_df = df[df['Status'] == 'Cancelled']

cancelled_df['Time Slot'].value_counts().plot(kind='bar', figsize=(8,5), color='blue')
plt.title("Cancelled Trips by Time Slot")
plt.xlabel("Time Slot")
plt.ylabel("Number of Cancelled Trips")
plt.show()
```



## Chart 13: No Cars Available by Time Slot

### 1. Why did you pick the specific chart?

This chart was chosen to isolate and analyze cases where customer demand could not be fulfilled due to lack of available cars.

### 2. What is/are the insight(s) found from the chart?

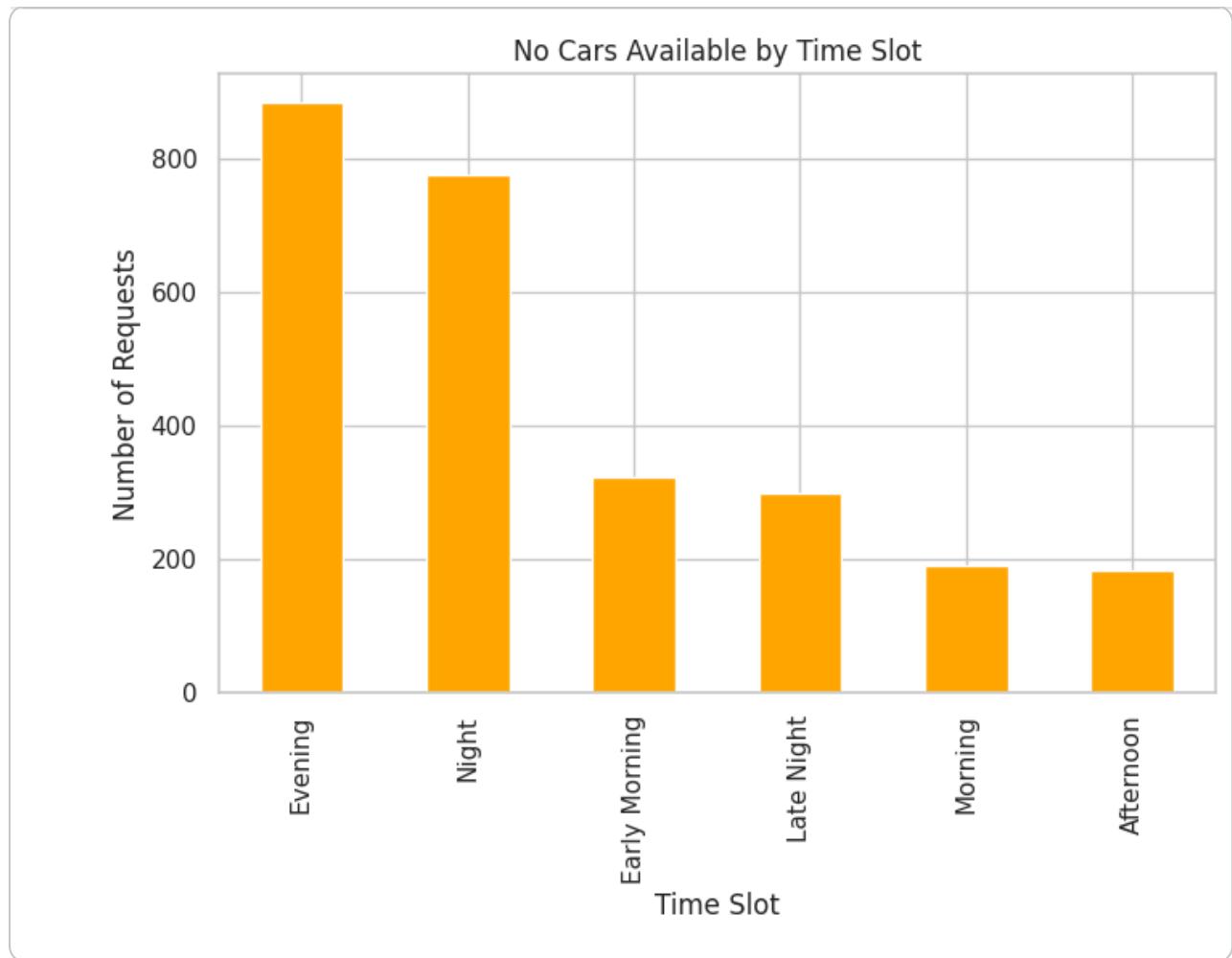
The Night time slot shows the highest number of “No Cars Available” cases, highlighting severe supply shortages during late hours.

### 3. Will the gained insights help creating a positive business impact?

Yes, this insight highlights the need for night-shift driver incentives and better demand forecasting to reduce unmet demand.

```
no_cars_df = df[df['Status'] == 'No Cars Available']

no_cars_df['Time Slot'].value_counts().plot(kind='bar', figsize=(8,5), color
plt.title("No Cars Available by Time Slot")
plt.xlabel("Time Slot")
plt.ylabel("Number of Requests")
plt.show()
```



## Chart 14: Supply–Demand Gap by Pickup Point and Time Slot

### 1. Why did you pick the specific chart?

This chart was selected to analyze how supply gaps vary across pickup locations and time slots simultaneously.

**\*\*2. What is/are the insight(s) found from the chart? \*\***

Airport pickup points during Night and Early Morning experience the highest supply–demand gaps compared to city locations.

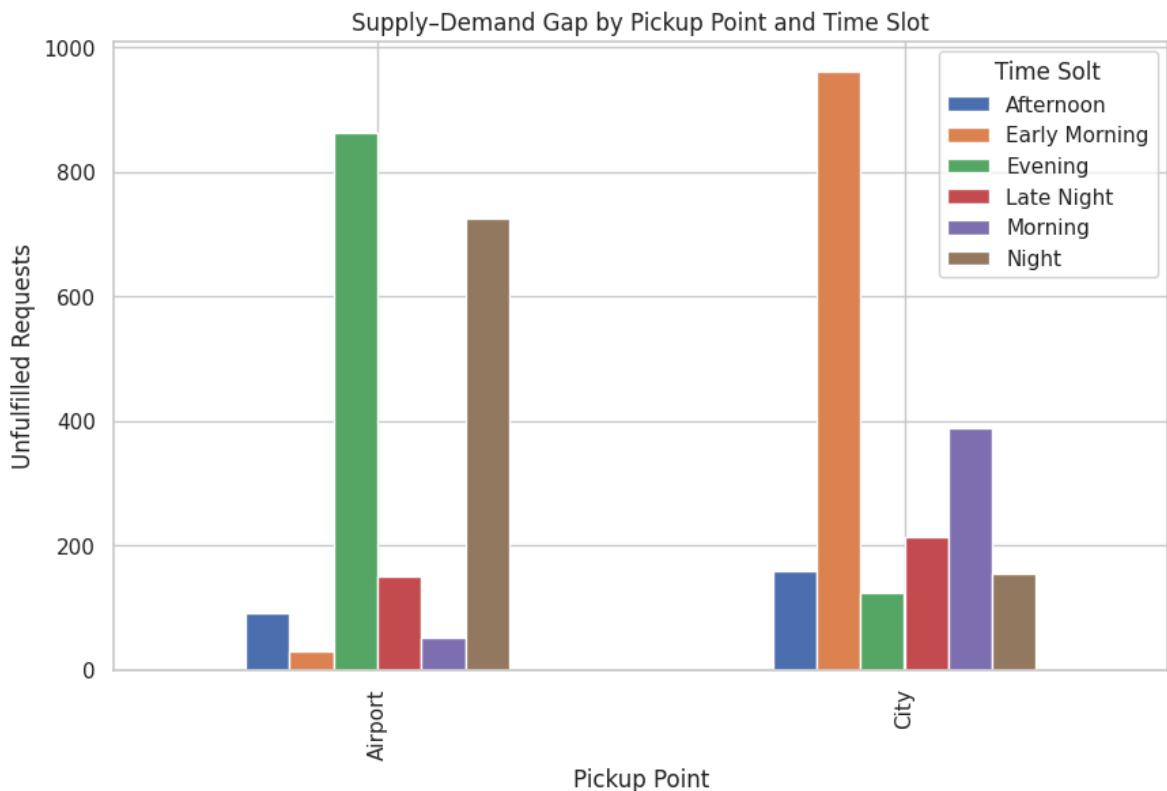
### 3. Will the gained insights help creating a positive business impact?

Yes, this insight allows Uber to deploy location-specific and time-specific operational strategies to reduce unmet demand.

```
gap_df.groupby(['Pickup point','Time Solt']).size().unstack().plot(
    kind='bar', figsize=(10,6)
)

plt.title("Supply-Demand Gap by Pickup Point and Time Slot")
plt.xlabel("Pickup Point")
```

```
plt.ylabel("Unfulfilled Requests")
plt.show()
```



## Chart 15: Proportion of Unfulfilled Requests (Cancelled vs No Cars Available)

### 1. Why did you pick the specific chart?

This chart was chosen to compare the relative contribution of cancellations and car unavailability to overall unfulfilled demand.

### 2. What is/are the insight(s) found from the chart?

The chart shows that “No Cars Available” contributes more to unfulfilled requests than cancellations, especially during night hours.

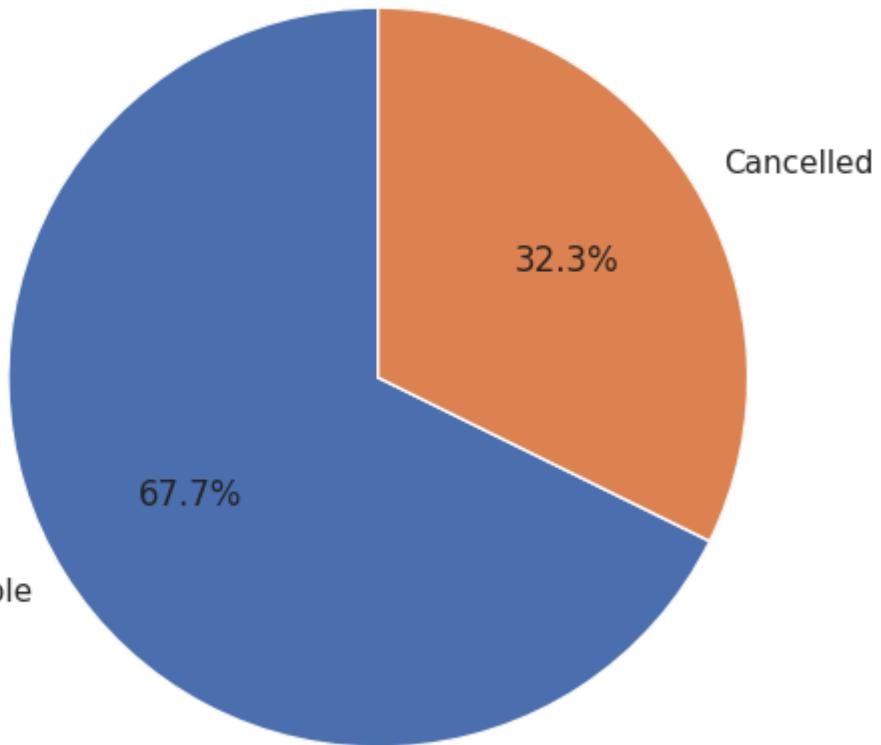
### 3. Will the gained insights help creating a positive business impact?

Yes, this insight helps Uber prioritize increasing car availability over only focusing on reducing cancellations, leading to more effective problem-solving.

```
gap_df['Status'].value_counts().plot(
    kind='pie',
    autopct='%1.1f%%',
    figsize=(6,6),
    startangle=90
)
```

```
plt.title("Proportion of Unfulfilled Requests by Status")
plt.ylabel("")
plt.show()
```

Proportion of Unfulfilled Requests by Status



## ▼ 5. Solution to Business Objective

Uber should increase driver availability during Early Morning and Night time slots, particularly at Airport pickup locations, through targeted incentives and dynamic pricing. Predictive demand forecasting can help position drivers in advance and reduce cancellations. These strategies will minimize unfulfilled requests and improve customer satisfaction.

**What do you suggest the client to achieve Business Objective? Explain Briefly.**

Based on the exploratory data analysis, Uber should focus on improving driver availability during time slots and locations with high supply–demand gaps. The analysis shows that Early Morning and Night time slots experience the highest number of unfulfilled requests, particularly at Airport pickup points. To address this,

Uber should introduce targeted driver incentives such as higher fares, bonuses, or guaranteed earnings during these peak-demand periods.

Additionally, predictive demand forecasting can be implemented using historical data to anticipate high-demand hours and proactively position drivers in critical locations. Dynamic pricing strategies can also be applied during peak hours to balance rider demand and encourage more drivers to be active. Reducing cancellations during early morning hours through driver accountability measures and better communication can further improve trip completion rates. Overall, these data-driven strategies can help Uber reduce unfulfilled requests, enhance customer satisfaction, and improve operational efficiency.

## Conclusion

This exploratory data analysis identified key supply–demand gaps in Uber ride requests across different time slots and pickup locations. The analysis revealed that unfulfilled requests are highest during Early Morning and Night hours, with Airport