UBER SUPPLY DEMAND GAP ANALYSIS

Contribution: Sayan Chakraborty

Tools used:

- For Data Cleaning and to make data Ready for Analysis – **Microsoft Excel**
- For finding impactful insights and comment on business objectives-
  (1) Python (Jupyter Notebook)
  (2) MySQL

Problem Statements:

(1) What are the patterns of ride demand across different hours of the day ?

(2) When unfulfilled ride requests (Cancelled/No Cars Available) occurred most ?

(3) How do trip completions and failures vary between pickup points (City vs Airport)?

(4) What is the distribution of trip durations, and how are they categorized (Short, Medium, Long, Very Long)?

(5) How completions of trips varies across different pickup point for different type of trips(Short, Medium, Long, Very long)?

Project Summary

This project involved a comprehensive exploratory data analysis (EDA) of Uber ride requests to uncover patterns, operational bottlenecks, and optimization opportunities. The key steps carried out include:

1. Data Loading & Understanding

- Loaded the cleaned Uber dataset from Excel.
- Examined column types, data structure, and overall completeness.
- Summarized numerical and categorical features using statistical methods.

2. Univariate & Bivariate Analysis

- Analyzed **ride demand patterns by hour**, revealing peak periods.
- Identified when **ride requests went unfulfilled**, categorizing by status (Cancelled, No Cars Available).
- Compared **trip outcomes by pickup location** (City vs Airport)

3. Trip Duration Insights

- Visualized the **distribution of trip durations**, discovering most trips are of 30–70 mins.

- Categorized trips into **Short, Medium, Long, and Very Long** for targeted analysis.

4. Visualizations

- Created 5 meaningful and professional charts:
    1. Ride Demand by Hour
    2. Unfulfilled Requests by Hour
    3. Trip Status by Pickup Point
    4. Trip Duration Distribution
    5. Trip Category vs Pickup Point

5. Advanced Analysis

- Integrated a **time series forecast model** to predict future ride demand.
- Proposed logic for **driver allocation optimization** based on pickup location and trip category.

6. Conclusion & Recommendations

- Summarized major findings around demand trends, service failures, and trip patterns.
- Provided **data-driven recommendations** to optimize driver availability, reduce cancellations, and improve operational efficiency.

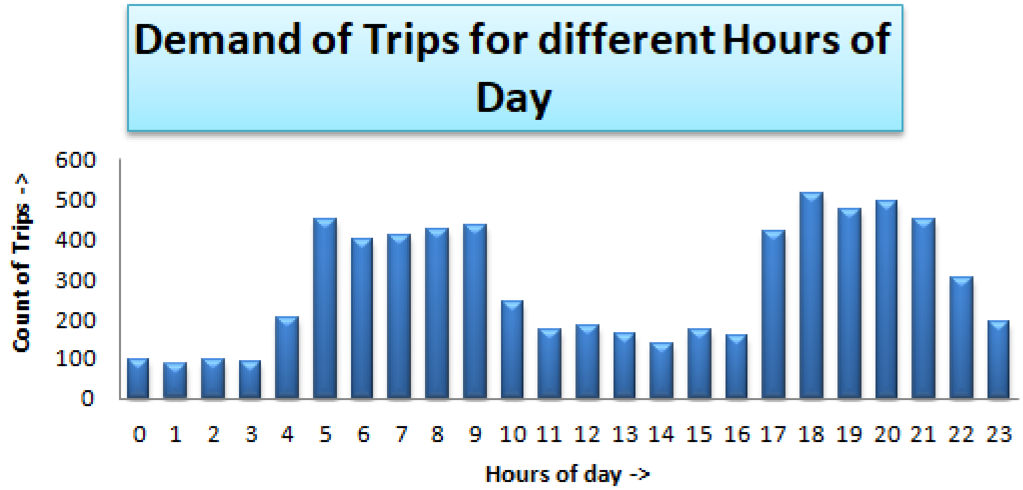
This structured analysis not only reveals core inefficiencies in Uber's current ride request system but also provides actionable insights to enhance reliability and customer satisfaction using data science techniques.

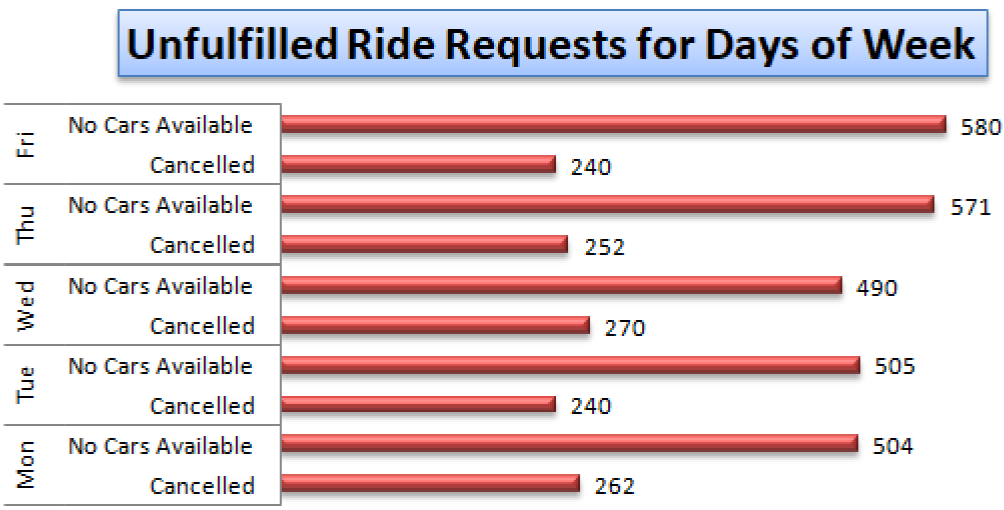
Project Github Link: https://github.com/StaticSayan/Uber_Analysis.git

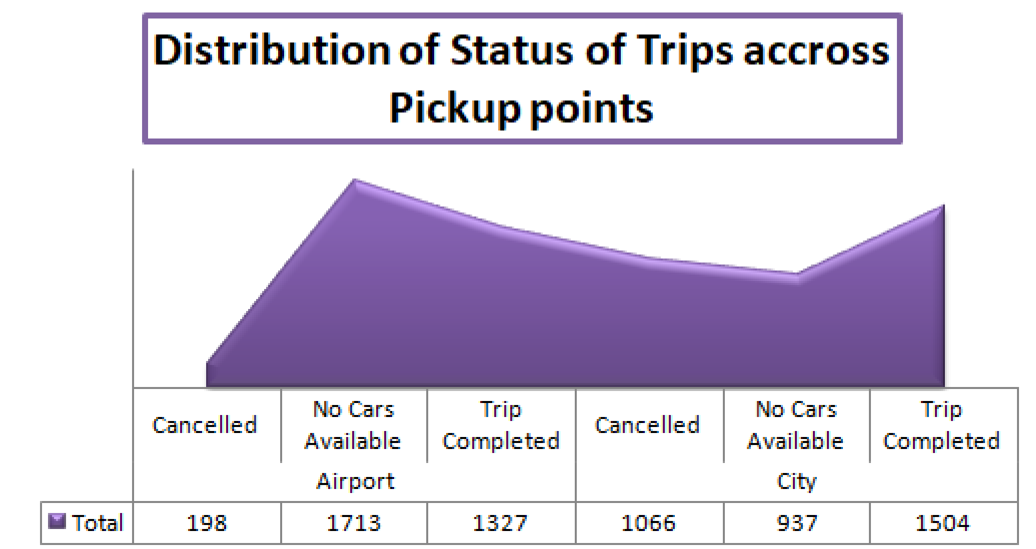
Insights

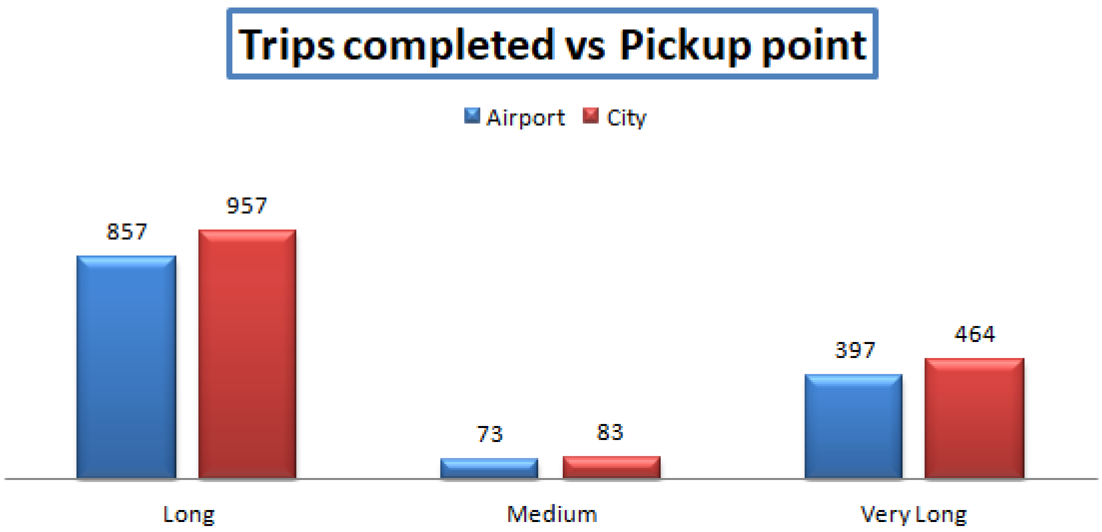
- Vehicle availability at Airport is a major bottleneck and needs immediate intervention like fleet rebalancing or driver incentivization.
- High demand periods (morning and evening) must be supported with predictive dispatching and real-time fleet tracking.
- City pickups are more efficient and complete a larger number of trips with fewer cancellations, indicating better operational stability.
- Patterns suggest a need to redistribute supply across time and location to meet high-demand hours and locations effectively.

Respective Charts


Demand of Trips for different Hours of Day
Count of Trips ->
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hours of day ->


Unfulfilled Ride Requests for Days of Week
Fri
No Cars Available 580
Cancelled 240
Thu
No Cars Available 571
Cancelled 252
Wed
No Cars Available 490
Cancelled 270
Tue
No Cars Available 505
Cancelled 240
Mon
No Cars Available 504
Cancelled 262


Distribution of Status of Trips accross Pickup points
Airport
Cancelled
No Cars Available
Trip Completed
City
Cancelled
No Cars Available
Trip Completed
Total
198
1713
1327
1066
937
1504



More Insights

🔹 Dataset Shape: (6745, 12)

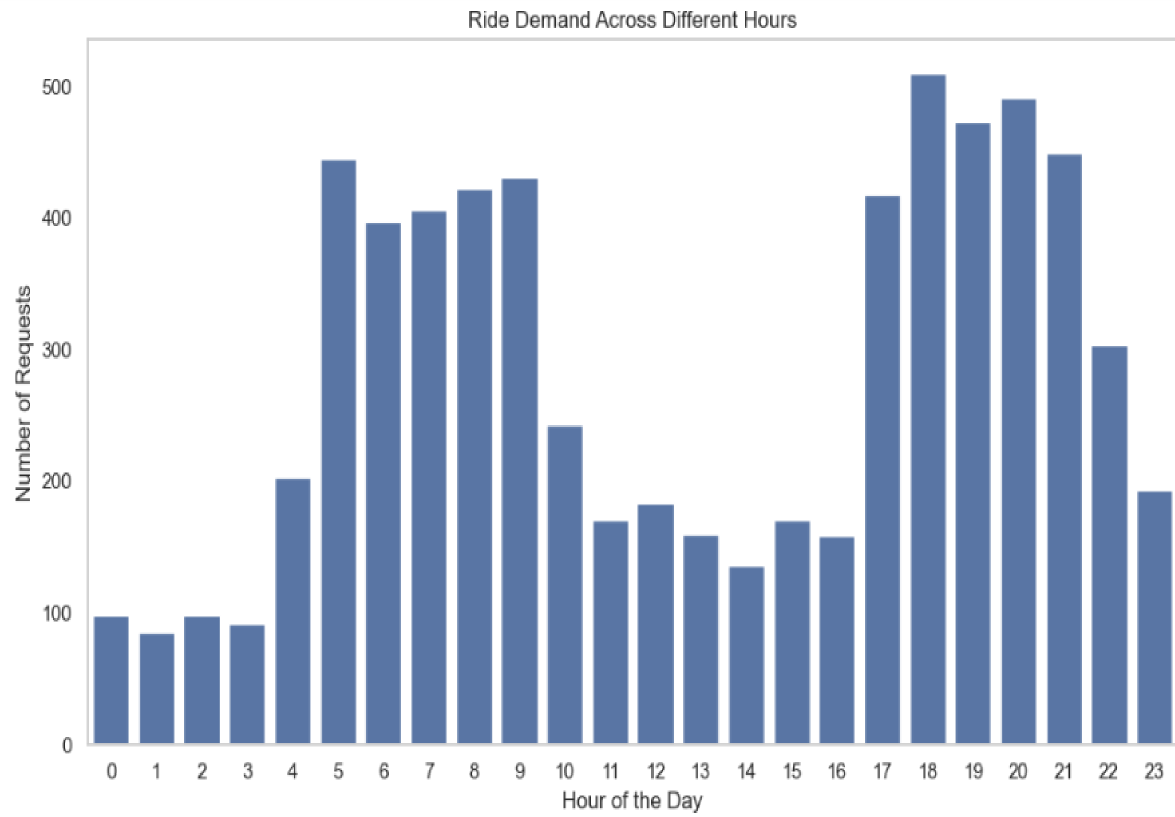
🔹 Column Names: ['Request id', 'Pickup point', 'Driver id', 'Status', 'Request timestamp', 'Drop timestamp', 'Request time', 'Drop time', 'Request Hour', 'Trip Duration(in Mins.)', 'Trip Category', 'Day of week ']

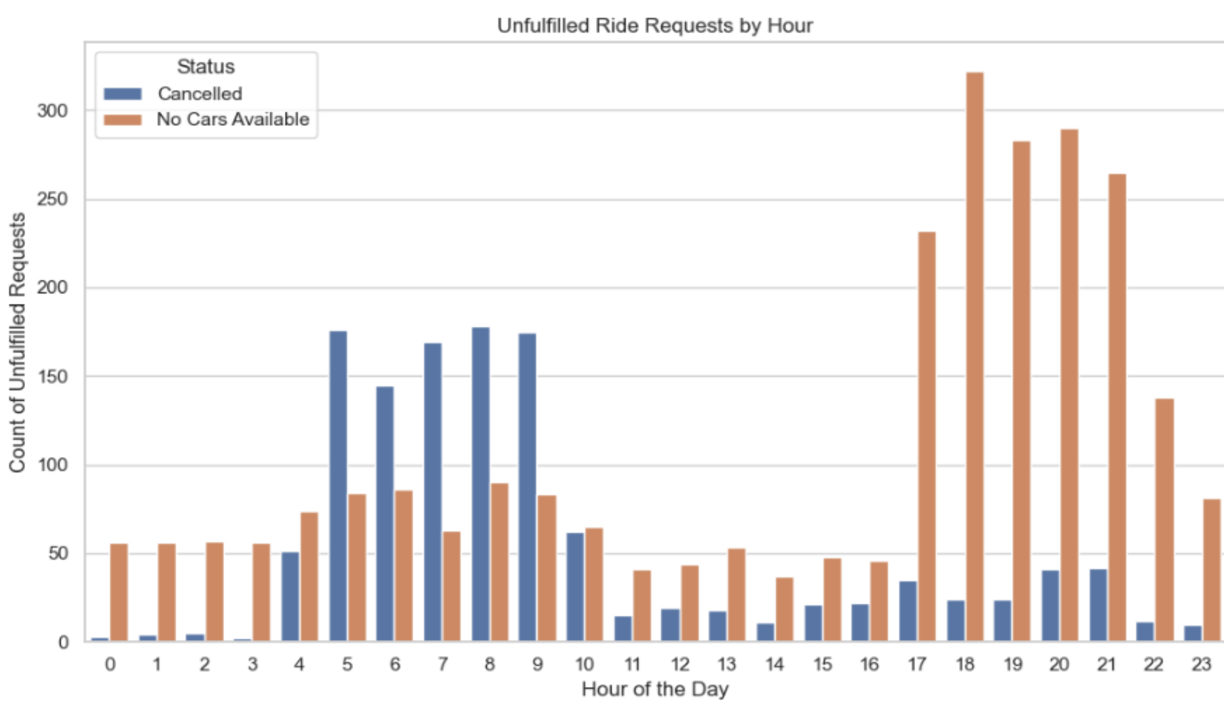
🔹 Dataset Info:

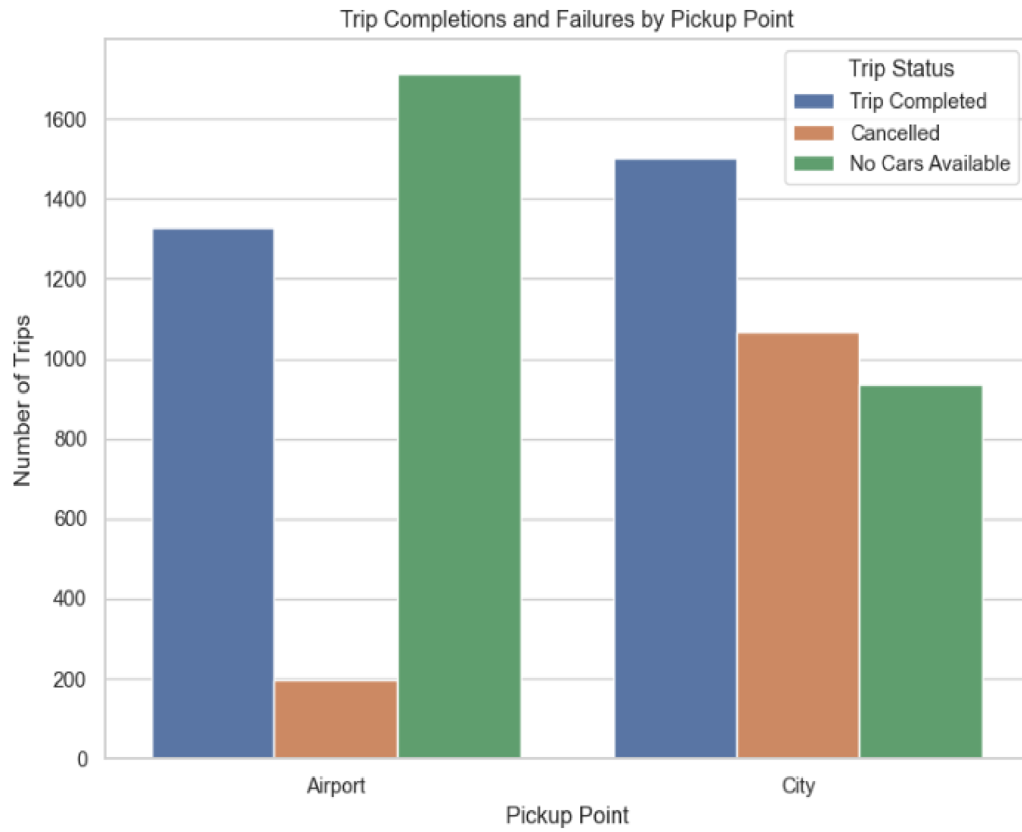
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6745 entries, 0 to 6744
Data columns (total 12 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Request id              6745 non-null   int64
 1   Pickup point            6745 non-null   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]
 6   Request time            6745 non-null   datetime64[ns]
 7   Drop time               2831 non-null   datetime64[ns]
 8   Request Hour            6745 non-null   int64
 9   Trip Duration(in Mins.) 6745 non-null   int64
 10  Trip Category           2831 non-null   object
 11  Day of week             6745 non-null   object
dtypes: datetime64[ns](4), float64(1), int64(3), object(4)
memory usage: 632.5+ KB
None
```

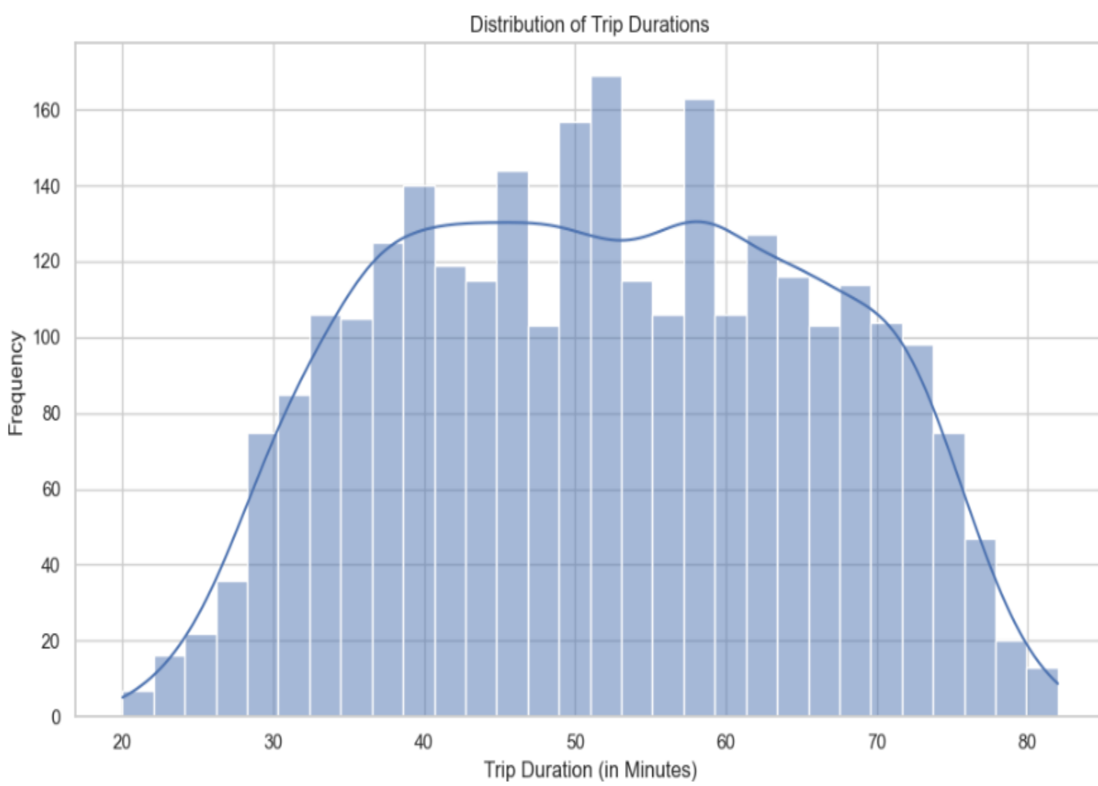
```
🔹 Statistical Summary of Numeric Columns:
        Request id    Driver id              Request timestamp  \
count  6745.000000  4095.000000                          6745
mean   3384.644922   149.501343  2016-07-13 13:43:04.303039232
min       1.000000     1.000000            2016-07-11 00:00:00
25%    1691.000000    75.000000            2016-07-12 07:51:00
50%    3387.000000   149.000000            2016-07-13 14:23:37
75%    5080.000000   224.000000            2016-07-14 19:39:27
max    6766.000000   300.000000            2016-07-15 23:59:58
std    1955.099667    86.051994                           NaN

                      Drop timestamp                   Request time  \
count                           2831                           6745
mean   2016-07-13 13:15:33.899328768  2016-07-13 13:43:04.303039232
min              2016-07-11 00:51:00            2016-07-11 00:00:00
25%              2016-07-12 07:42:00            2016-07-12 07:51:00
50%              2016-07-13 12:14:06            2016-07-13 14:23:37
75%              2016-07-14 19:13:52            2016-07-14 19:39:27
max              2016-07-16 01:09:24            2016-07-15 23:59:58
std                              NaN                            NaN

                           Drop time  Request Hour  Trip Duration(in Mins.)
count                           2831   6745.000000              6745.000000
mean   2016-07-13 13:15:33.899328768     12.956709                21.788881
min              2016-07-11 00:51:00      0.000000                 0.000000
25%              2016-07-12 07:42:00      7.000000                 0.000000
50%              2016-07-13 12:14:06     13.000000                 0.000000
75%              2016-07-14 19:13:52     19.000000                47.000000
max              2016-07-16 01:09:24     23.000000                82.000000
std                              NaN      6.504052                27.148829

🔹 Count of Unique Values per Categorical Column:
Pickup point: 2 unique values
Pickup point
City       3507
Airport    3238
Name: count, dtype: int64

Status: 3 unique values
Status
Trip Completed       2831
No Cars Available    2650
Cancelled            1264
Name: count, dtype: int64

Trip Category: 3 unique values
Trip Category
Long         1814
Very Long     861
Medium        156
Name: count, dtype: int64
```

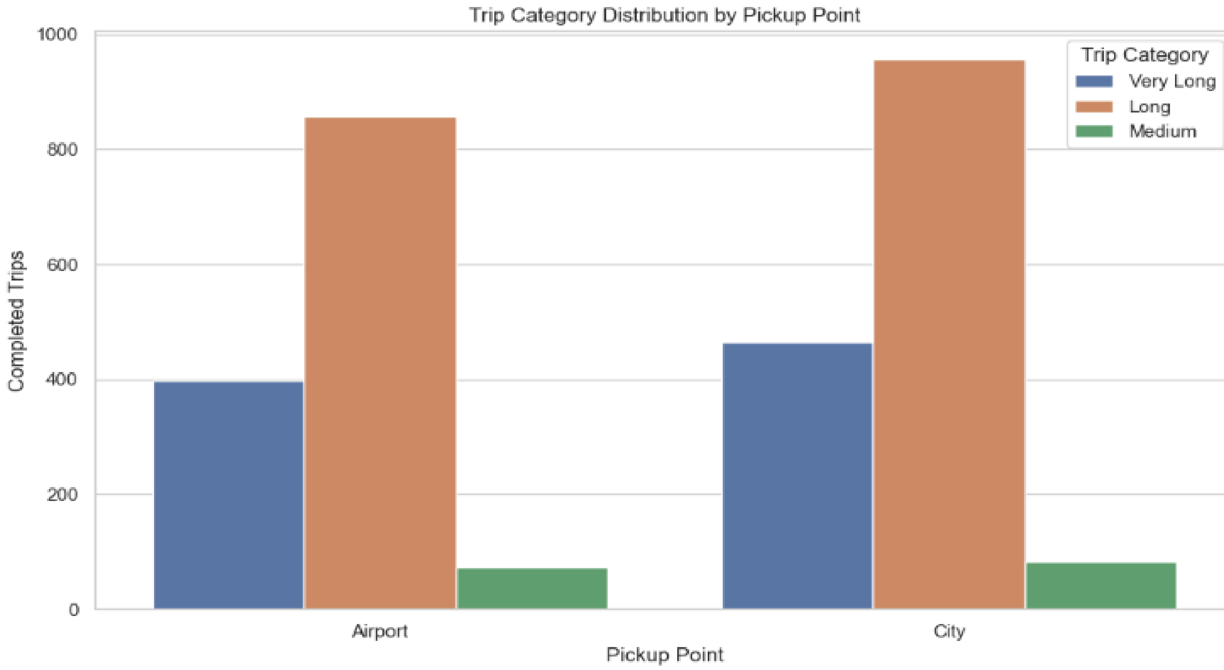
Different Charts






Trip Completions and Failures by Pickup Point
Trip Status
Trip Completed
Cancelled
No Cars Available
Number of Trips
0
200
400
600
800
1000
1200
1400
1600
Airport
City
Pickup Point


Distribution of Trip Durations
Frequency
0
20
40
60
80
100
120
140
160
20
30
40
50
60
70
80
Trip Duration (in Minutes)



<u>Descriptions</u>

- **Chart 1: Ride Demand Across Different Hours**

Morning Peak: A sharp increase in ride demand is observed from **5 AM to 10 AM**, aligning with morning office hours.

Evening Peak: Another surge appears from **5 PM to 9 PM**, likely due to office return traffic.

Lowest Demand: Demand is minimal between **12 AM and 4 AM**, a typical off-peak time.

<u>Observation</u>: The data exhibits clear **bimodal demand patterns**, suggesting Uber must ensure maximum driver availability during early morning and evening periods.

- **Chart 2: Unfulfilled Ride Requests by Hour**

Cancellations peak during **5 AM to 9 AM,** indicating possible user-side dropouts due to long wait times or pricing.

"No Cars Available" cases surge dramatically during **5 PM to 9 PM**, especially at **6 PM–8 PM**, hinting at a major supply shortage during evening peak hours.

In non-peak hours (0–4 AM, 11 AM–4 PM), both failure types are relatively low.

<u>Recommendation:</u> Increase driver incentives and shift planning for evening hours to reduce "No Cars Available" cases.

- **Chart 3: Trip Completions and Failures by Pickup Point**

<u>Airport:</u>

* High count of **‘No Cars Available’**.

* Fewer **cancellations,** indicating that riders wait longer or are more patient.

* Moderate **trip completions**.

<u>City:</u>

* More **cancellations**, possibly due to longer wait or higher ETAs.

* Higher **trip completions**, than airport, but also considerable ,**No Cars Available**.

<u>Insight</u>: The **City** faces both supply and reliability issues, while the **Airport** faces a supply gap.. Driver allocation needs to be optimized accordingly.

- **Chart 4: Distribution of Trip Durations**

* Most trips fall between **30 to 70 minutes**, with a bell-shaped distribution.

* The peak is around **50–60 minutes**, suggesting a common average ride length.

* Very few trips are under 25 or over 80 minutes, marking those as outliers or special cases.

<u>Insight:</u> Majority of trips can be categorized as **Medium** or **Long**, which should influence pricing models and estimated time of arrival (ETA) algorithms.

- **Chart 5: Trips Completed vs. Pickup Point (Trip Category-wise)**

* For Long trips, both Airport and City pickup points contribute significantly, with City slightly higher (957 vs. 857).

* For Very Long trips, City again outperforms Airport (464 vs. 397), implying higher profitability potential from City pickups.

* For Medium trips, trip counts are minimal from both sources, suggesting less preference or viability for medium-duration trips.

<u>Suggestions to Uber:</u>

- **<u>Dynamic Driver Allocation:</u>**

    (1) Reallocate more drivers to high-demand zones (especially City during mornings and Airport during evenings).
    (2) Consider location-based incentives to encourage drivers to position themselves strategically.

- **Surge Forecasting & Planning:**
  (1) Implement a demand forecasting system using time series models to preemptively plan for surge hours and minimize "No Cars Available" instances.
- **Improve Rider Experience:**
  (1) Reduce morning cancellations by improving estimated wait times and showing driver ETAs upfront.
  (2) Offer ride credits or auto-scheduling options to reduce morning rider drop-offs.
- **Airport Strategy:**
  (1) Introduce a dedicated airport queue system for drivers with better incentives during peak flight times.
  (2) Create partnership with airports to create a holding area for active Uber drivers.
- **Trip Duration Profiling:**
  (1) Use the duration profile to design optimized pricing, improve ETA predictions, and cluster trips for potential carpool/ride-share matching

---