Uber Data Analysis Report

1. Introduction

This report analyzes Uber trip request data to uncover patterns related to ride requests, cancellations, car availability, and operational performance. The project used Excel for data cleaning, SQL for insight extraction, and Python for Exploratory Data Analysis (EDA).

2. Data Cleaning (Excel)

1. The original dataset contained columns such as request_id, pickup_point, driver_id, status, request_timestamp, and drop_timestamp.
2. Handled Missing Values
    - Identified missing values in Driver_ID and Drop_Timestamp.
    - Created a missing value summary sheet.
3. Formatted Date and Time Columns
    - Standardized the format of Request_Timestamp and Drop_Timestamp
    - Ensured uniform datetime format for consistency across entries.
4. Extracted Time-Based Features
    - Created new columns:
        - Request Date (extracted date)
        - Request Hour (extracted time)
5. Checked and Removed Duplicates
    - Used Excel's Remove Duplicates tool to check for duplicate Request_ID entries. (No duplicates found)
6. Saved Cleaned Dataset
    - Saved the cleaned sheet in .xlsx format to preserve formatting and formulas.

3. SQL Insights

Key insights were derived using SQL queries on the cleaned data:

• Distribution of Trip Status

	status character varying (50)	total_requests bigint
1	Trip Completed	2831
2	Cancelled	1264
3	No Cars Available	2650

- Requests by Pickup Point and Status

	pickup_point character varying (50)	status character varying (50)	total_requests bigint
1	Airport	No Cars Available	1713
2	Airport	Trip Completed	1327
3	Airport	Cancelled	198
4	City	Trip Completed	1504
5	City	Cancelled	1066
6	City	No Cars Available	937

- Peak Request Hours

	request_hour character varying (10)	total_requests bigint
1	18:10	17
2	20:33	15
3	19:02	15
4	18:22	15
5	17:34	14
6	18:29	14
7	19:31	14
8	21:38	14

- Cancellation Rates by Hour

	request_hour character varying (10)	ride_requested bigint	cancellation_rate numeric
1	09:30	2	100.0000000000000000
2	01:07	1	100.0000000000000000
3	05:57	2	100.0000000000000000
4	07:11	1	100.0000000000000000
5	04:57	2	100.0000000000000000
6	16:28	1	100.0000000000000000
7	13:00	2	100.0000000000000000
8	11:19	1	100.0000000000000000

Total rows: 1351 Query complete 00:00:00.227

- Pickup Points with Most 'No Cars Available'

	pickup_point character varying (50)	total_unavailable bigint
1	Airport	1713
2	City	937

• Completion Rates by Pickup Point

	pickup_point character varying (50)	completion_rate numeric
1	City	42.8856572569147419
2	Airport	40.9820877084620136

• Status Breakdown by Day of Week

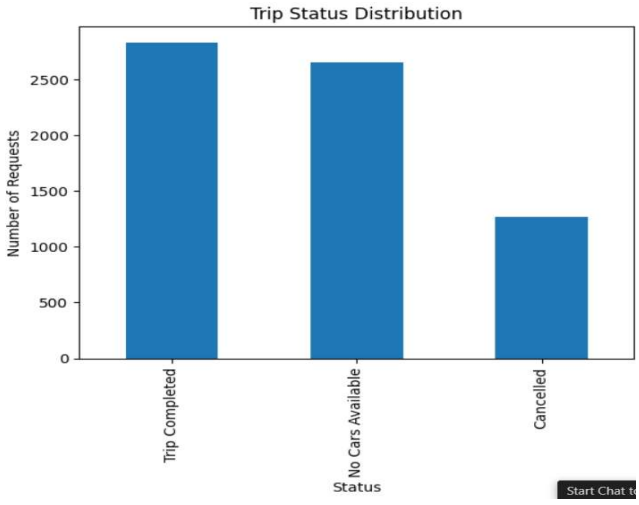
	day_of_week text	status character varying (50)	count bigint
1	Friday	Trip Completed	561
2	Friday	No Cars Available	580
3	Friday	Cancelled	240
4	Monday	Trip Completed	601
5	Monday	No Cars Available	504
6	Monday	Cancelled	262
7	Thursday	Trip Completed	530
8	Thursday	No Cars Available	571

Total rows: 15 Query complete 00:00:00.184

4. Exploratory Data Analysis (Python)

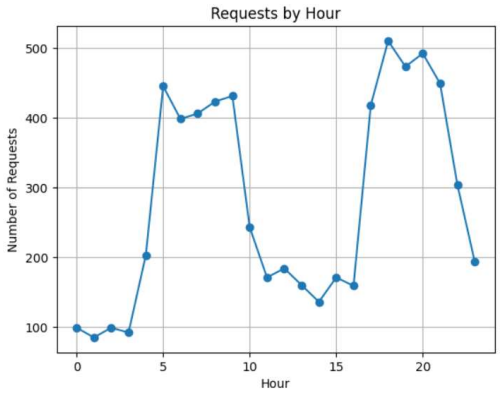
- Trip Status Distribution

  - Analyzed ride outcomes using the Status column (Trip Completed, Cancelled, No Cars Available).
  - Visualized total count of each status using a bar chart.



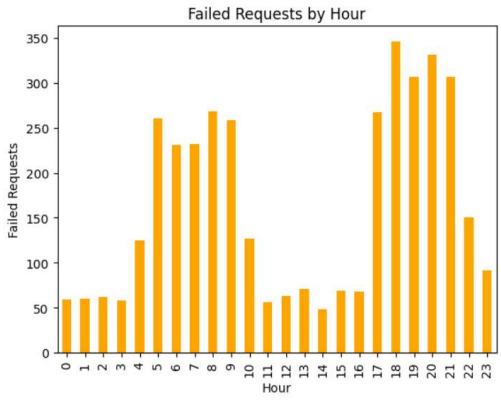
- Request by Hour

  - Created a line plot of number of requests by hour.
  - Peak demand observed during early morning and evening rush hours.



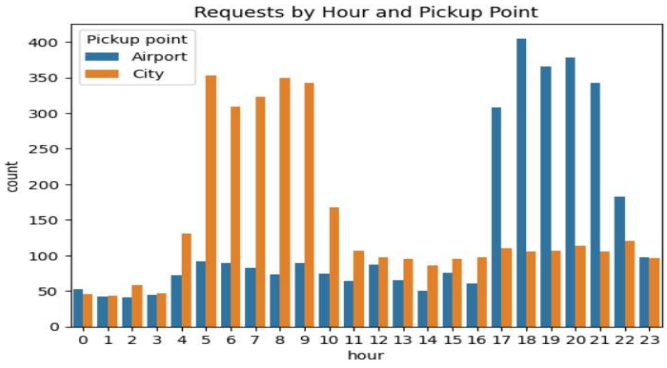
- Failed Requests by Hour

  - Filtered failed requests (Status ≠ "Trip Completed").
  - Bar chart showed failure peaks during high-demand hours.



- Pickup Point Analysis

  - Grouped data by Pickup point and Status.
  - Visualized request volume by pickup point and hour.
  - Stacked bar chart compared status distribution at Airport and City.



- Trip Duration Analysis

  - Trip duration calculated in minutes.
  - Descriptive stats showed variation in trip length.

```
count    2831.000000
mean       52.413753
std        13.850693
min        20.783333
25%        41.000000
50%        52.083333
75%        64.000000
max        83.000000
Name: Trip Duration (min), dtype: float64
```

5.Dashboard (Excel)

Conclusion

From the analysis, we found that most ride failures happened during busy hours in the morning and evening. This was mainly because there were not enough cars available to meet the demand. The problem was especially noticeable at the airport, where many ride requests could not be completed.

While many rides were completed successfully, a large number of cancellations or no-car-available situations affected the overall service quality. Trip durations were mostly normal, but peak-hour delays and unfulfilled requests highlight areas that need improvement.

Recommendations

- Add more cars during peak hours, especially in the early morning and evening.
- Improve car availability at the airport, as this location had the highest number of unfulfilled requests.
- Plan driver shifts better so that more drivers are available when and where they are needed the most.
- Use historical data to predict demand and manage car supply more effectively in high-demand areas.