UBER Request Dataset Analysis Project

Project Name: Uber Request Data Analysis Project

Project Type: Exploratory Data Analysis (EDA)

Contribution: Individual

Project Summary:

Objective: To analyse Uber request data to understand demand-supply dynamics, identify peak hours, and uncover reasons behind unfulfilled requests (cancelled or no cars available).

Workflow: Demonstrates a comprehensive data analysis workflow by integrating Excel, SQL, and Python (Pandas) for different stages of the project.

Problem Statement:

- **Core Problem:** Identify and address the underlying causes of trip cancellations and unfulfilled requests ('No Cars Available') within the Uber service.
- **Goal:** Improve service reliability and customer satisfaction

Data Overview:

Dataset: Uber_Dataset.xlsx (or Uber dataset cleaned.csv)

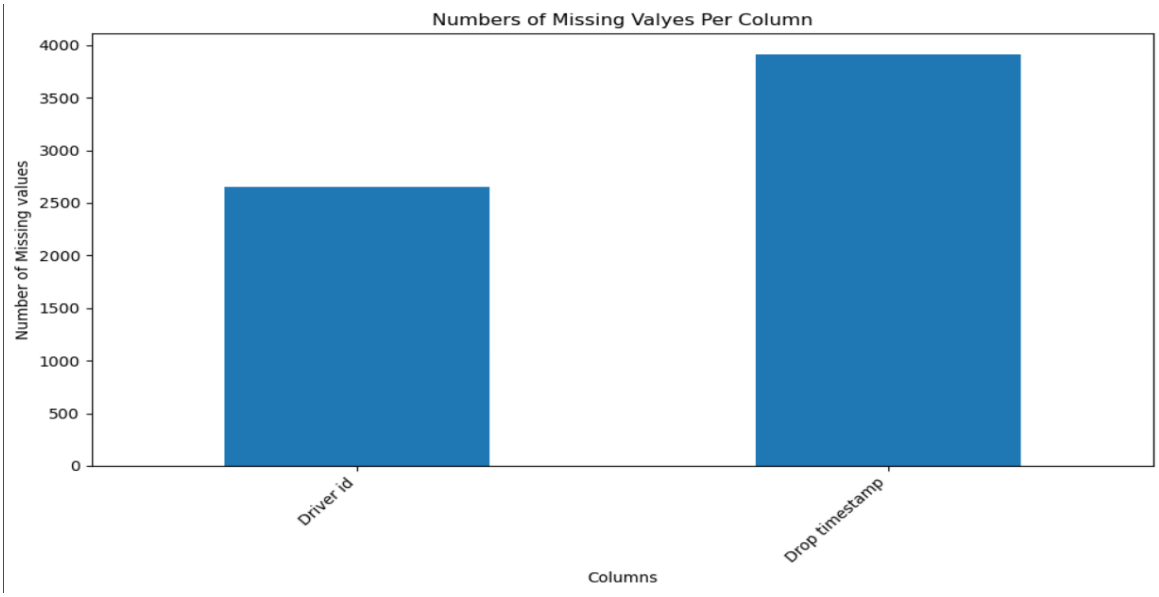
Rows & Columns: 6745 entries (rows) and 6 original columns (before feature engineering).

Key Columns:

- Request id: Unique identifier for each request (Integer).
- Pickup point: Location of the request (Airport/City).
- Driver id: Driver assigned (Float, contains missing values).
- Status: Outcome of the request (Trip Completed, Cancelled, No Cars Available).
- Request timestamp: Date & time of request (Datetime).
- Drop timestamp: Date & time of trip completion (Datetime, contains missing values).

Data Quality - Missing Values

- **Observation:**
    - Driver id had 2650 missing values.
    - Drop timestamp had 3914 missing values.
- **Interpretation:** These missing values are significant and likely correlate with uncompleted requests (e.g., 'Cancelled' or 'No Cars Available' statuses), as a trip that wasn't completed wouldn't have a drop timestamp or an assigned driver.



Feature Engineering - Adding Time-Based Insights

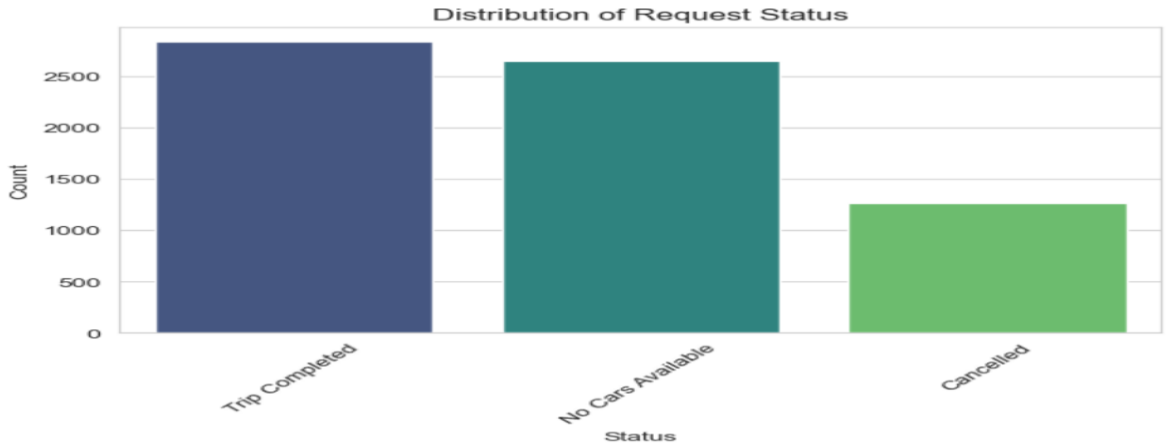
- **Process:** Converted Request timestamp and Drop timestamp to datetime objects.
- **New Features Created:**
    - Request Hour: Hour of the day the request was made.
    - Request Day of Week: Day of the week the request was made (e.g., Monday, Tuesday).
    - Request Month: Month the request was made (e.g., July).
    - Trip Duration: Calculated duration for completed trips in minutes.
- **Purpose:** To analyse patterns and trends based on time.

Missing Value Handling

- **Strategy:**
    1. **Driver id:** Imputed NaN values with 'No cars available' to reflect the reason for unfulfilled requests.
    2. **Trip Duration:** Imputed NaN values with 0 for non-completed trips, as these trips had no duration.
- **Result:** All Driver id and Trip Duration missing values were handled. (Drop timestamp still has NaT, which is expected for uncompleted trips).

Data Visualisation & Understanding with Charts

Chart - 1. Distribution of Trip Status



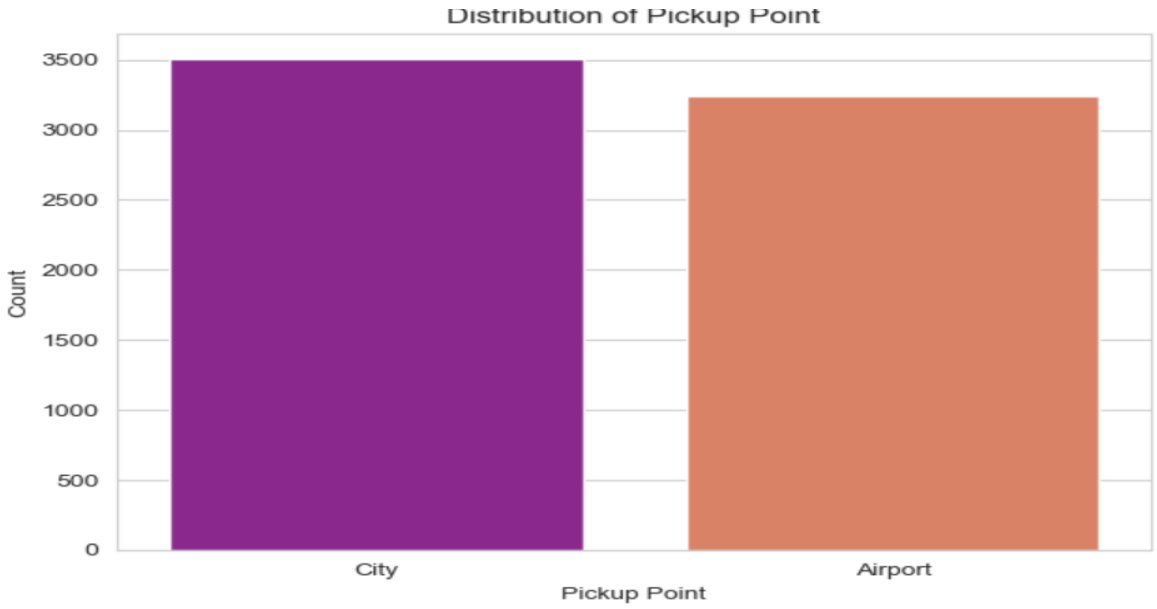
I picked a bar chart as it is excellent for showing the frequency or count of items in different categories

The chart reveals the overall success rate and failure modes of ride requests.

This insight can drive strategies to:

- Improve driver availability.
- Reduce cancellations.
- Focus on converting 'No Cars Available' requests into completed trips.

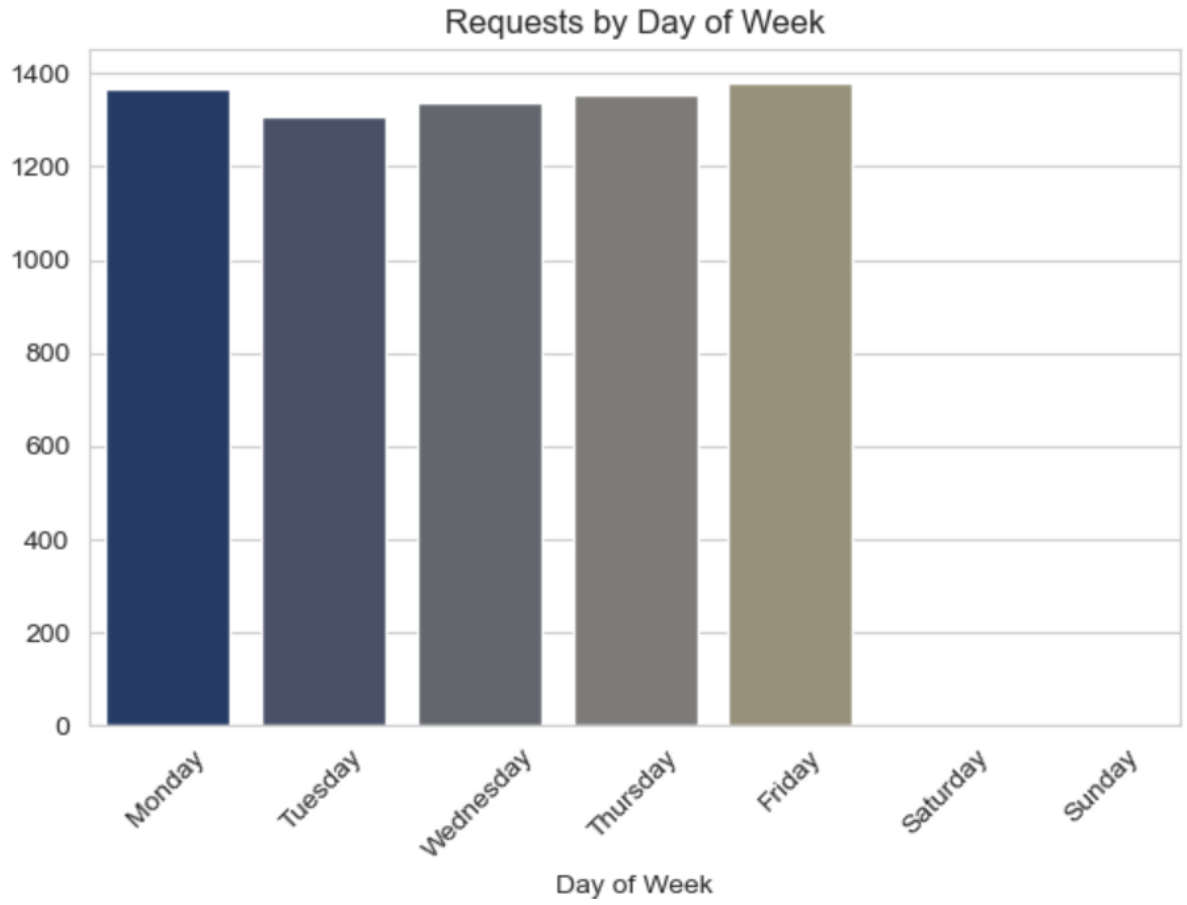
Chart - 2. Distribution of Pickup point



The chart shows the demand split between 'Airport' and 'City' pickup points. It highlights which location generates more ride requests.

Positive Impact:- Uber can ensure a sufficient supply of drivers in high-demand areas to maximize completed trips and minimize unmet demand.

Chart - 3. Requests by Day of Week



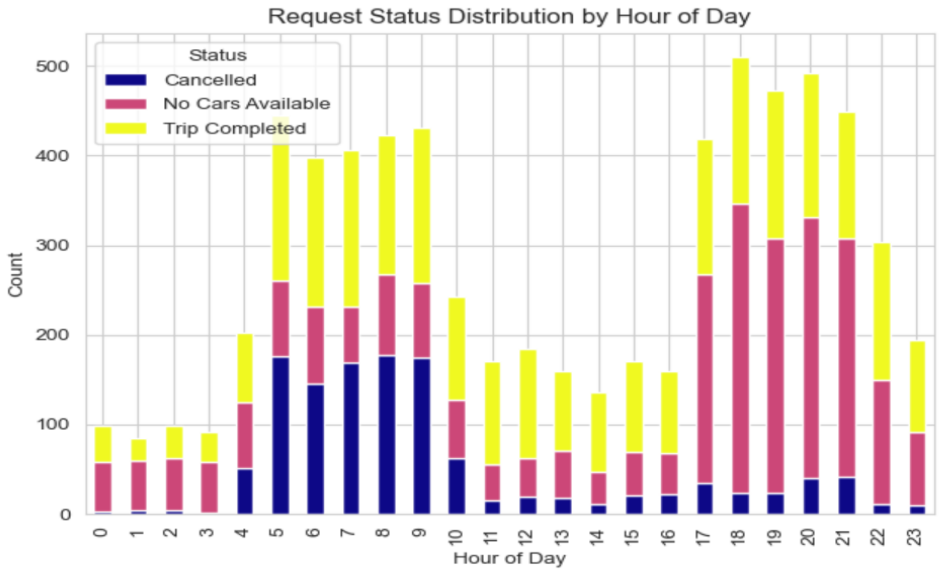
Similar to hourly requests, a bar chart is ideal for showing the distribution of requests across categorical 'days of the week'. It helps in identifying daily patterns in overall demand.

The chart shows how demand varies throughout the week. Certain days might have higher or lower request volumes.

Daily demand patterns inform:

- Weekly driver scheduling.
- Anticipating higher demand on specific days to prepare driver supply.

Chart - 4. Status distribution by Hour of Day (Stacked Bar Chart)



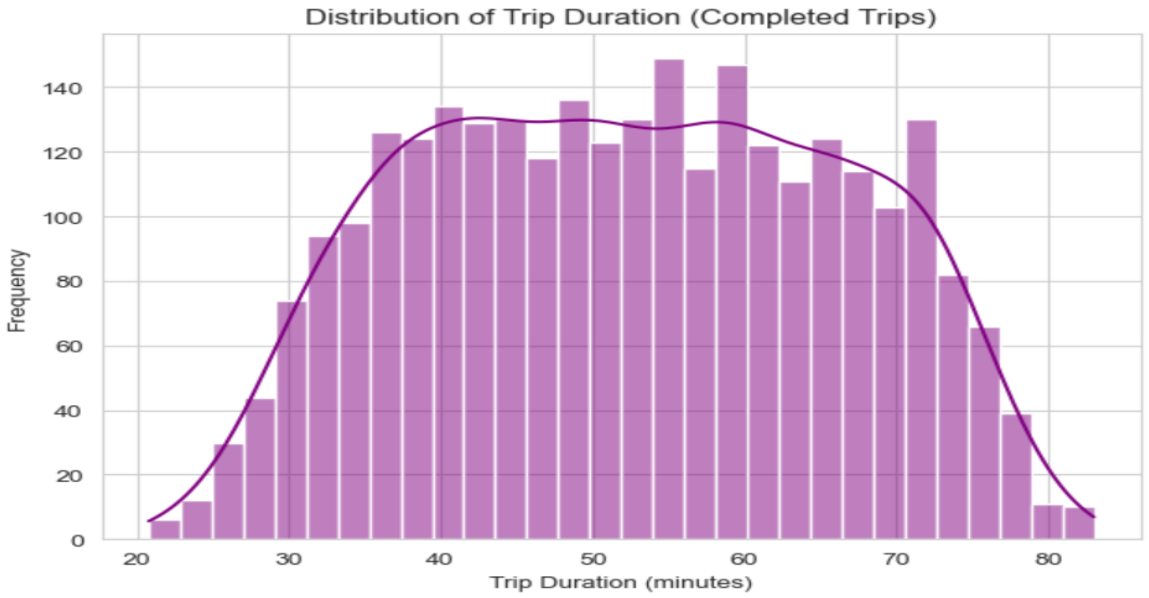
A stacked bar chart is ideal for showing how the distribution of request statuses changes throughout the day.

Shows a high 'No Cars Available' rate during early morning hours from the City or high cancellation rates during late-night hours.

Insights:

- Offer surge pricing or driver incentives during specific times to balance supply and demand.

Chart - 5. Distribution of Trip Duration (for completed trips)



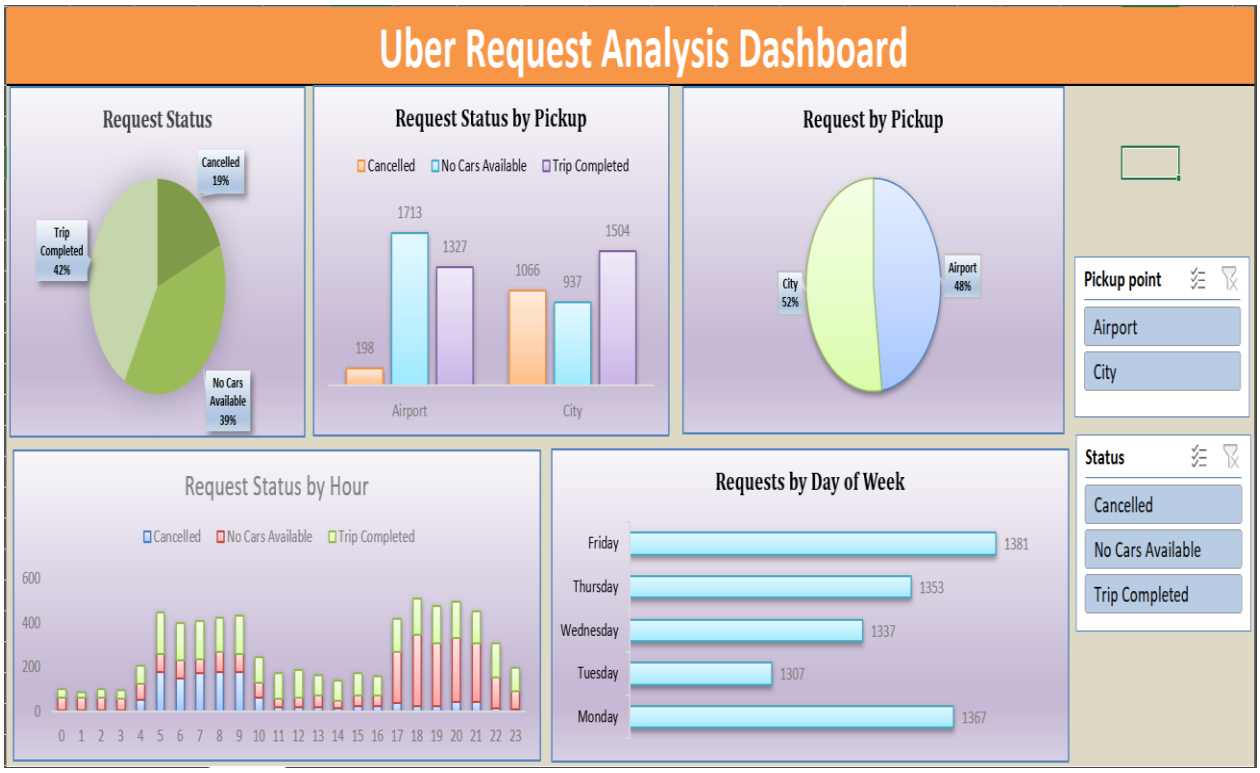
A histogram (with KDE for smoothness) is ideal for visualizing the distribution of a continuous numerical variable like 'Trip Duration'. It shows the most common trip lengths and the spread of durations.

The chart reveals typical trip lengths. Average trip durations are about 40-60 mins

Understanding trip duration patterns can help with:

- Pricing strategies : Develop more granular pricing models based on typical trip lengths for different zones.
- Driver earnings estimation : Drivers can better estimate their earnings per trip based on common durations.
- Route optimisation : Identify unusually long trips that might indicate inefficient routing or traffic issues, leading to map/routing improvements.

Excel Dashboard



SQL Queries

1.Count the number of requests by pickup point:

```
SELECT "Pickup point", COUNT(*) AS TotalRequests FROM uber_request

GROUP BY "Pickup point";
```

Answer:

"City" 3507

"Airport" 3238

This query groups the requests by Pickup_point and counts the number of requests from each location.

2. Cancellation Rate by Pickup Point

```
SELECT

"Pickup point",

COUNT(*) AS TotalRequests,

SUM(CASE WHEN "Status" = 'Cancelled' THEN 1 ELSE 0 END) AS TotalCancellations,

CAST(SUM(CASE WHEN "Status" = 'Cancelled' THEN 1 ELSE 0 END) AS REAL) * 100 / COUNT(*) AS CancellationRatePercentage

FROM uber_request

GROUP BY "Pickup point"

ORDER BY CancellationRatePercentage DESC;
```

Answer: "City" 3507 1066 30.3963501568292

"Airport" 3238 198 6.114885731933292

Highlights locations with high cancellation rates, which might indicate driver unavailability or other issues in those areas.

3. Demand-Supply Gap by Pickup Point (No Cars Available)

```
SELECT
   "Pickup point",
   COUNT(*) AS TotalNoCarsAvailable
FROM uber_request
WHERE "Status" = 'No Cars Available'
GROUP BY "Pickup point"
ORDER BY TotalNoCarsAvailable DESC;
```

Answer: "Airport" 1713

"City" 937

Reveals specific locations where demand is not being met due to a lack of available cars.

4. Average Trip Duration by Pickup Point for Completed Trips

```
SELECT
   "Pickup point",
   AVG("Trip Duration") AS AverageTripDurationMinutes
FROM uber_request
WHERE "Status" = 'Trip Completed'
GROUP BY "Pickup point"
ORDER BY AverageTripDurationMinutes DESC;
```

Answer: "City" 52.56

"Airport" 52.23

Shows if trips from certain pickup points generally take longer or shorter, which can impact driver efficiency or fare calculations.

5. Breakdown of Trip Statuses by Pickup Point

```
SELECT
   "Pickup point",
   AVG("Trip Duration") AS AverageTripDurationMinutes
FROM uber_request
WHERE "Status" = 'Trip Completed'
GROUP BY "Pickup point"
ORDER BY AverageTripDurationMinutes DESC;
```

Answer: "City" 52.56

"Airport" 52.23

Shows if trips from certain pickup points generally take longer or shorter, which can impact driver efficiency or fare calculations.

Recommendations

- **Based on Peak Hours/Days:** Implement dynamic pricing or incentives for drivers during peak hours (Request Hour, Request Day of Week) to increase supply.
- **Based on Pickup Point Issues:**
    - **High Cancellations (e.g., Airport):** Investigate specific reasons for cancellations. Is it long wait times? Driver preference? Consider dedicated airport queues or bonuses.
    - **'No Cars Available' (e.g., City):** Deploy more drivers to these high-demand areas during critical periods.
- **Driver Management:** Analyse driver behaviour (e.g., cancellation reasons if available) and provide targeted training or support.

- **Communication:** Improve communication with riders during high-demand periods about potential wait times or limited availability.

Conclusion

The project analysed Uber request data, revealing a significant problem of unmet demand and cancellations driven by supply-demand imbalances, particularly during peak hours and at specific locations. Actionable insights highlight the need for targeted driver incentives and operational optimizations to increase trip completions and improve customer satisfaction.