**Case Study: Analysis of Ride Booking Dataset**

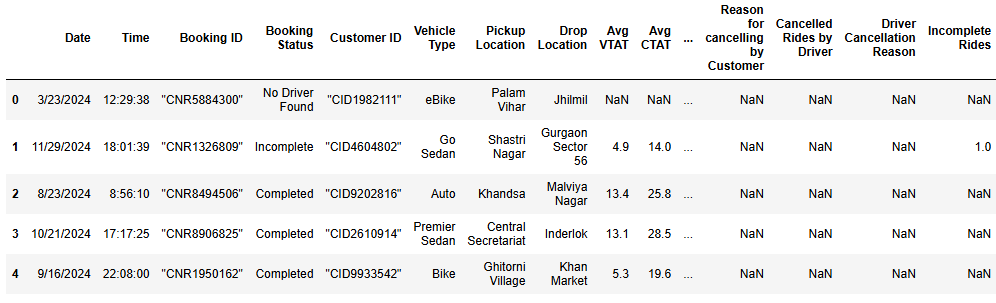
**Understand Problem Statement**

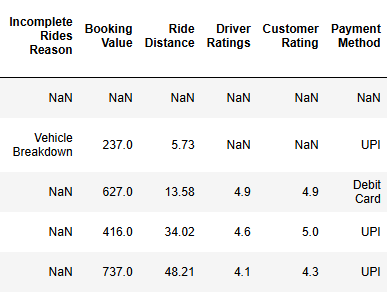
The Problem statement contains the ride booking information including several attributes. the data has missing values outliers and inappropriate formats which may affect the analysis the problem is to clen and preprocess the data and get the required results and insights and obtained desired visuals.

**Data Collection / Sourcing**

1. In this case the dataset is provided in the form of csv file, read the csv file import it to data frame for further operations.

* Demo of the Raw data with attributes Before Cleaning

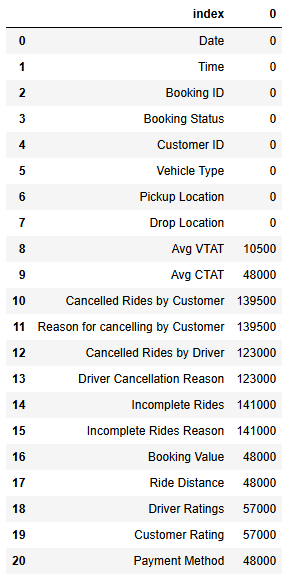
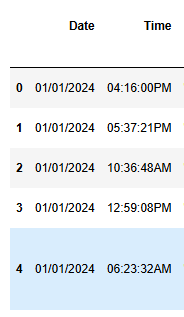


****

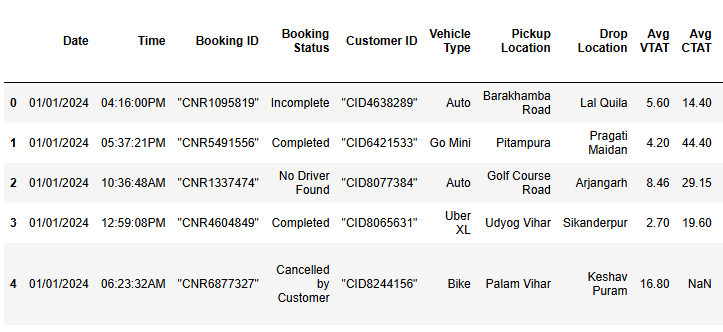
**Methodology**

* **Data Cleaning / Understand your variables**

1. Start by importing all the required libraries-based n the operation needs to be performed: pandas, NumPy, matplotlib, and display.
2. Data Cleaning: Find all the missing values and fill them with appropriate data using the fillna () function.
3. Convert the date and Time column to the appropriate format.



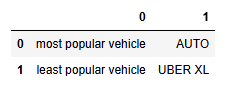
* Demo of the data with attributes After Cleaning

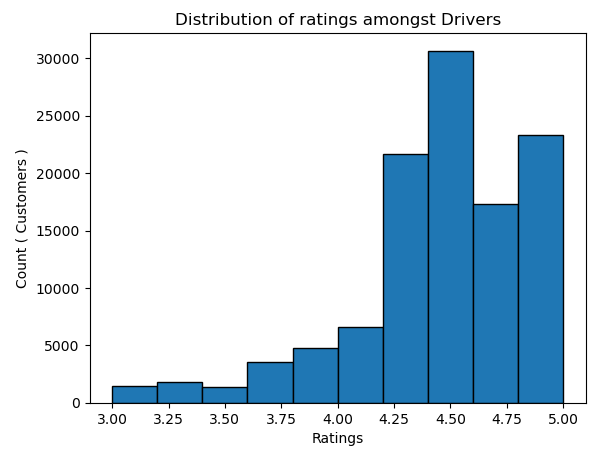
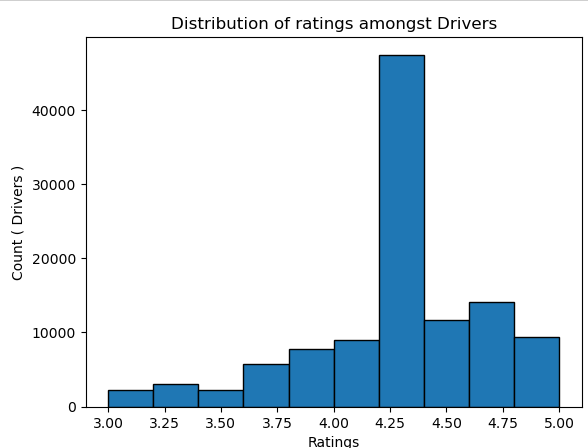


**Results and Insights**

* **Descriptive Analysis:**

1. The most popular vehicle is an **auto**.
2. The least popular vehicle is **Uber XL**.



1. Average ride distance is **24.64.**
2. Average booking value is **499.**
3. Distribution of ratings across drivers and customers is visualized using a histogram.
4. most common Reason for cancelling by Customer **wrong address**
5. most common Reason for cancelling by Driver **customer related issue**

* **Customer Behavior Insights:**

1. **Drivers** are the most frequent cancellers of rides.
2. Total number of rides canceled by drivers are **27,000.**

* **Operational Matrices**

1. **Ashram** is the location with high demand, with a total count of 936 rides.
2. **05:44:57** PM is the time slot with highest demand

**Additional Observations:**

1. Around 13 attributes have null values.
2. The most common and popular payment method is UPI.
3. The lowest rating any driver has gotten is three.
4. Maximum ride distance is 50.
5. Total rides cancelled by drivers are 27,000.
6. Total rides cancelled by customers are 10,500.
7. The total number of rides cancelled by both customers and drivers is 37,500.

**Conclusion**

This project focused on cleaning and analyzing a ride booking dataset, applying data preprocessing, visualization, and analysis techniques to uncover key patterns in customer behavior and operations, while strengthening skills in Python, pandas, and data-driven decision making.

Prajwal S. Chougule

Prajwalchougule41@gmail.com