Data Analyst Project

By Kiran Gupta

OLA Data Analyst Project

Data Columns:

1. Date
2. Time
3. Booking_ID
4. Booking_Status
5. Customer_ID
6. Vehicle_Type
7. Pickup_Location
8. Drop_Location
9. V_TAT
10. C_TAT
11. cancelled_Rides_by_Customer
12. cancelled_Rides_by_Driver
13. Incomplete_Rides
14. Incomplete_Rides_Reason
15. Booking_Value
16. Payment_Method
17. Ride_Distance
18. Driver_Ratings
19. Customer_Rating

SQL Questions:

#1. Identify all bookings that were successfully completed.

```
Select * From Successful_booking;
```

#2. Calculate the average travel distance for each vehicle type.

```
SELECT * FROM avg_Distance_for_each_Vehicle;
```

#3. Count the total number of rides cancelled by the customers.

```
Select * From  rides_cancellation_by_customers;
```

#4. Determine the top 5 customers with the highest number of rides.

```
SELECT * FROM Top_5_Customers;
```

#5. Find out how many rides were cancelled by drivers due to vehicle malfunctions and personal circumstances.

```
SELECT Driver_Cancellation_Count FROM Driver_Cancellations;
```

#6. Find the highest and lowest driver ratings for "Prime Sedan" trips.

```
SELECT * FROM Prime_Sedan_Ratings;
```

#7. List all trips where the payment method used was UPI
SELECT * FROM UPI_Bookings;

#8. Calculate the average customer satisfaction score for each type of vehicle.
SELECT * FROM Avg_Customer_Rating_Vehicle;

9. Calculate the total booking value of rides completed successfully
SELECT Total_Successful_Ride_value FROM Total_Successful_Ride;

#10. Retrieve a list of all incomplete rides and the specific reasons for their incompletion.
SELECT * FROM Incomplete_Rides_with_Reason;

#11. Calculate the percentage of rides that were cancelled by drivers for each vehicle type.
SELECT * FROM Driver_Cancellation_Rate_by_Vehicle;

#12. Find the vehicle type with the lowest average driver rating.
SELECT * FROM Lowest_Avg_Driver_Rating_Vehicle;

#13. Calculate the average ride distance for rides paid with 'Card' versus 'Cash'.
SELECT * FROM Average_Distance_Payment_Method;

#14. For each vehicle type, calculate the percentage of successful bookings.
SELECT * FROM Success_Rate_by_Vehicle;

Power BI Questions:

1. **How ride volumes change over time**
2. **Share of bookings across different statuses**
3. **Five leading vehicle types ranked by total ride distance**
4. **Comparison of customer ratings by vehicle type**
5. **Most common reasons behind ride cancellations**
6. **Revenue distribution across payment methods**
7. **Top customers ranked by overall booking value**
8. **Variation in ride distances on a daily basis**
9. **Pattern of ratings given to drivers**
10. **Side-by-side comparison of customer and driver ratings**

Segregation of the views:

1. Overall
    - Ride Volume Over Time
    - Booking Status Breakdown
2. Vehicle Type
    - Top 5 Vehicle Types by Ride Distance
3. Revenue
    - Revenue by Payment Method
    - Top 5 Customers by Total Booking Value
    - Ride Distance Distribution Per Day
4. Cancellation
    - Cancelled Rides Reasons (Customer)
    - Cancelled Rides Reasons(Drivers)
5. Ratings
    - Driver Ratings
    - Customer Ratings

Data-Driven Recommendations:

Based on the analysis of our project (Ola) data, here are key recommendations to improve operational performance and customer satisfaction:

1. Reduce Cancellations to Improve Reliability

- **Customer Cancellations:** Currently, customer cancellations stand at 13.62%. Implement a strategy to reduce this to a target of **less than 7%**. This will require identifying and addressing the root causes, such as long wait times or mismatched vehicle types.
- **Driver Cancellations:** The data shows that 12.33% of rides are cancelled by drivers. By investigating the reasons (e.g., "Personal & Car related issue" from my SQL script), we can set a realistic goal of keeping this rate **below 10%**.

2. Enhance Operational Efficiency

- **Minimize Incomplete Rides:** With 10.37% of rides being "incomplete," there's a significant opportunity for improvement. A clear goal is to reduce this number to **less than 6%** by analyzing the reasons for non-completion and streamlining the ride process.
- **Maintain Data Integrity:** Our analysis confirmed that all Booking_IDs in the dataset follow the CNR format followed by a series of digits. This indicates a strong data standard that should be maintained to ensure consistent and reliable tracking.

3. Strategic Insights for Growth

- **Optimize for High-Demand Periods:** The Date column in our data can be used to identify weekends and analyze booking trends on those days. We can use this data to develop a

strategy to increase both the number and value of orders during these periods, leveraging data to maximize revenue.

- **Analyze Order Value Distribution:** The Booking_Value column in the Bookings.csv file can be used to analyze the distribution of order values. Based on this data, we can set targets for the proportion of low-value, mid-value, and high-value orders to optimize revenue.