OLA Data Analyst Project

ChatGPT Prompt to Create Data Please create a spreadsheet with 20k rows, for Bengaluru city. Give the following columns.

The data will be for 1 month. use the following column –

1. Date

2. Time

3. Booking ID

4. Booking Status

5. Customer ID

6. Vehicle Type - Auto - Prime Plus - Prime Sedan - Mini - Bike - eBike - Prime SUV

7. Pickup Location (Create dummy location points Take any 50 areas from Bangalore)

8. Drop Location (Take from dummy pickup locations)

9. Avg VTAT (Time taken to arrive at the vehicle)

10. Avg CTAT (Time taken to arrive the Customer)

11. Cancelled Rides by Customer

12. Reason for cancelling by Customer - Driver is not moving towards pickup location - Driver asked to cancel - AC is not working (Only for 4-wheelers) - Change of plans - Wrong Address

13. Cancelled Rides by Driver - Personal & Car related issues - Customer related issue - The customer was coughing/sick - More than permitted people in there

14. Incomplete Rides

15. Incomplete Rides Reason - Customer Demand - Vehicle Breakdown - Other Issue

16. Booking Value

17. Ride Distance

18. Driver Ratings

19. Customer Rating

Keep the overall booking status success for this data at 62%. If the booking status is successful, then only fare charge ratings, average VTAT, average CTAT, and other data will be there.

SQL Questions with solution:

1. Retrieve all successful bookings:

```
Create View Successful_Bookings as
SELECT * FROM bookings
WHERE Booking_Status = 'Success';

#1. Retrieve all successful bookings:
Select * From Successful_Bookings;
```

2. Find the average ride distance for each vehicle type:

```
Create View ride_distance_for_each_vehicle As
SELECT Vehicle_Type, AVG(Ride_Distance) as avg_distance
FROM bookings
GROUP BY Vehicle_Type;

#2. Find the average ride distance for each vehicle type:
Select * from ride_distance_for_each_vehicle;
```

3. Get the total number of cancelled rides by customers:

```
Create View cancelled_rides_by_customers As
SELECT COUNT(*) FROM bookings
WHERE Booking_Status = 'cancelled by Customer';

#3. Get the total number of cancelled rides by customers:
Select * from cancelled_rides_by_customers;
```

4. List the top 5 customers who booked the highest number of rides:

```
Create View Top_5_Customers As
SELECT Customer_ID, COUNT(Booking_ID) as total_rides
FROM bookings
GROUP BY Customer_ID
ORDER BY total_rides DESC
LIMIT 5;

#4. List the top 5 customers who booked the highest number of rides:
Select * from Top_5_Customers;
```

5. Get the number of rides cancelled by drivers due to personal and car-related issues:

```
Create View Rides_cancelled_by_Drivers_P&C_Issues As
SELECT COUNT(*)
FROM bookings
WHERE cancelled_Rides_by_Driver = 'Personal & Car related issue';
```

#5. Get the number of rides cancelled by drivers due to personal and car-related issues:
Select * from Rides_cancelled_by_Drivers_P&C_Issues;

6. Find the maximum and minimum driver ratings for Prime Sedan bookings:

```
Create View Max_Min_Driver_Rating As
SELECT MAX(Driver_Ratings) as max_rating, MIN(Driver_Ratings) as min_rating
FROM bookings
WHERE Vehicle_Type = 'Prime Sedan';
```

#6. Find the maximum and minimum driver ratings for Prime Sedan bookings:
Select * from Max_Min_Driver_Rating;

7. Retrieve all rides where payment was made using UPI:

```
Create View UPI_Payment As
SELECT * FROM bookings
WHERE Payment_Method = 'UPI';
```

#7. Retrieve all rides where payment was made using UPI:
Select * from UPI_Payment;

8. Find the average customer rating per vehicle type:

```
Create View AVG_Cust_Rating As
SELECT Vehicle_Type, AVG(Customer_Rating) as avg_customer_rating
FROM bookings GROUP BY Vehicle_Type;
```

#8. Find the average customer rating per vehicle type:
Select * from AVG_Cust_Rating;

9. Calculate the total booking value of rides completed successfully:

```
Create View total_successful_ride_value As
SELECT SUM(Booking_Value) as total_successful_ride_value
FROM bookings
WHERE Booking_Status = 'Success';
```

#9. Calculate the total booking value of rides completed successfully:
Select * from total_successful_ride_value;

10. List all incomplete rides along with the reason

```
Create View Incomplete_Rides_Reason As
SELECT Booking_ID, Incomplete_Rides_Reason
FROM bookings
WHERE Incomplete_Rides = 'Yes';
```

#10. List all incomplete rides along with the reason: Select * from Incomplete_Rides_Reason;

Power BI Answers:
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

Answers:

1. Ride Volume Over Time: A time-series chart showing the number of rides per day/week.

2. Booking Status Breakdown: A pie or doughnut chart displaying the proportion of different booking statuses (success, cancelled by the customer, cancelled by the driver, etc.).

3. Top 5 Vehicle Types by Ride Distance: A bar chart ranking vehicle types based on the total distance covered.

4. Average Customer Ratings by Vehicle Type: A column chart showing the average customer ratings for different vehicle types.

5. cancelled Rides Reasons: A bar chart that highlights the common reasons for ride cancellations by customers and drivers.

6. Revenue by Payment Method: A stacked bar chart displaying total revenue based on payment methods (Cash, UPI, Credit Card, etc.).

7. Top 5 Customers by Total Booking Value: A leaderboard visual listing customers who have spent the most on bookings.

8. Ride Distance Distribution Per Day: A histogram or scatter plot showing the distribution of ride distances for different Dates.

9. Driver Rating Distribution: A box plot visualizing the spread of driver ratings for different vehicle types.

10. Customer vs. Driver Ratings: A scatter plot comparing customer and driver ratings for each completed ride, analysing correlations.

Key Statistical Insights

1. Booking Overview
    - Total Bookings: 20,407
    - Successful Bookings: 12,652 (~62%) ✅
    - Cancellations by Customers: 2,081 (~10%)
    - Cancellations by Drivers: 3,654 (~18%)
    - Incomplete Rides: 795 (~4%)

This shows success rate is strong (~62%), with driver cancellations higher than customer cancellations — an area for operational improvement.

2. Vehicle Type Demand (Top 5)
    - eBike → 2,997 rides
    - Bike → 2,949 rides
    - Prime Sedan → 2,945 rides
    - Prime SUV → 2,945 rides
    - Auto → 2,916 rides

Demand is well distributed across vehicle types, but eBikes and Bikes are slightly more popular, suggesting a growing shift towards affordable/eco-friendly rides.

3. Ride & Revenue Metrics
    - Average Ride Distance: 14.16 km
    - Average Booking Value: ₹546.32

Short-to-mid distance rides dominate. Higher value rides may be concentrated on weekends or SUVs/Sedans.

4. Customer & Driver Experience
    - Average Driver Rating: 4.0
    - Average Customer Rating: 3.99

Both ratings are close, showing a generally balanced satisfaction level, but there's scope to improve customer experience (slightly below 4).

Learnings from the Project

- **Data Cleaning**: Handling duplicates, missing values, and ensuring realistic distributions sharpened your ability to prepare raw data for analysis.
- **SQL Skills**: Practiced writing aggregation, grouping, filtering, and ranking queries (e.g., top 5 customers, revenue analysis).
- **Power BI**: Built dashboards showing booking trends, cancellation patterns, and customer/driver insights, mirroring what analysts deliver in real companies.
- **Business Understanding**: Learned how operational KPIs (cancellations, incomplete rides, average ratings) impact customer satisfaction and revenue.
- **Analytical Thinking**: Translated numbers into **business insights** like demand trends, cancellation issues, and customer loyalty.

Summary Insight:

The project highlighted that while OLA maintains a healthy booking success rate (~62%), **driver cancellations (18%)** are significantly higher than customer cancellations (10%). Demand is **strong for two-wheeler options (Bike & eBike)**, showing customer preference for affordable and quick transport. Improving **driver reliability and customer ratings** could directly boost revenue and customer retention.