Data Analyst Project

ChatGPT Prompt to Create Data

You can create a spreadsheet with 1 lac rows, for any 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. Canceled Rides by Driver

- Personal & Car related issues
- Customer related issue
- The customer was coughing/sick
- More than permitted people in there

OLA Data Analyst Project

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

SQL Questions:

1. Retrieve all successful bookings:
2. Find the average ride distance for each vehicle type:
3. Get the total number of cancelled rides by customers:
4. List the top 5 customers who booked the highest number of rides:
5. Get the number of rides cancelled by drivers due to personal and car-related issues:
6. Find the maximum and minimum driver ratings for Prime Sedan bookings:
7. Retrieve all rides where payment was made using UPI:
8. Find the average customer rating per vehicle type:
9. Calculate the total booking value of rides completed successfully:
10. List all incomplete rides along with the reason:

Power BI Questions:

1. Ride Volume Over Time
2. Booking Status Breakdown
3. Top 5 Vehicle Types by Ride Distance
4. Average Customer Ratings by Vehicle Type
5. cancelled Rides Reasons
6. Revenue by Payment Method
7. Top 5 Customers by Total Booking Value
8. Ride Distance Distribution Per Day
9. Driver Ratings Distribution
10. Customer vs. Driver Ratings

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

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**

- Canceled Rides Reasons (Customer)
  Canceled Rides Reasons(Drivers)

5. **Ratings**

- Driver Ratings
- Customer Ratings

OLA Data Analyst Project

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, analyzing correlations.

SQL Questions & Answers

```
Create Database Ola;
Use Ola;
```

#1. Retrieve all successful bookings:

```
SELECT * FROM olabookings
WHERE Booking_Status = 'Success';
```

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

```
SELECT Vehicle_Type, AVG(Ride_Distance) as
avg_distance FROM olabookings
GROUP BY Vehicle_Type;
```

OLA Data Analyst Project

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

```
SELECT COUNT(*) FROM olabookings
WHERE Booking_Status = 'canceled by Customer';
```

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

```
SELECT Customer_ID, COUNT(Booking_ID) as total_rides
FROM olabookings
GROUP BY Customer_ID
ORDER BY total_rides DESC LIMIT 5;
```

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

```
SELECT COUNT(*)
FROM olabookings
WHERE canceled_Rides_by_Driver = 'Personal & Car related issue';
```

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

```
SELECT MAX(Driver_Ratings) as max_rating,
MIN(Driver_Ratings) as min_rating
FROM olabookings
 WHERE Vehicle_Type = 'Prime Sedan';
```

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

```
SELECT * FROM olabookings
WHERE Payment_Method = 'UPI';
```

#8. Find the average customer rating per vehicle type:

```
SELECT Vehicle_Type, AVG(Customer_Rating) as avg_customer_rating
FROM olabookings
GROUP BY Vehicle_Type;
```

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

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

OLA Data Analyst Project

#10. List all incomplete rides along with the reason:

```
SELECT Booking_ID, Incomplete_Rides_Reason
FROM olabookings
  WHERE Incomplete_Rides = 'Yes';
```