Data Analyst Project

By Apoorv Srivastava

ChatGPT Prompt to Create Data

Please create a spreadsheet with 10000 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.

Make sure orders cancelled by customers should not be more than 7%
Make sure orders cancelled drivers should not be more than 18%

Also, increase the number of orders on weekends and match days. Keep match day by using the following dates.
keep incomplete rides less than 6%
Keep order value high on weekends

in Food Category keep around 67 Indian
keep order ID with 10 digits starting with CNR and then digits
keep orders under 500 value 70%
keep orders above 500 value 28%
keep remaining orders above 1000

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

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 & Answers

```
CREATE DATABASE ola;
USE ola;
```

#1. Retrieve all successful bookings:

```
CREATE VIEW successful_bookings AS
SELECT * FROM booking
WHERE Booking_Status = 'Success';
SELECT * FROM successful_bookings;
```

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

```
CREATE VIEW average_ride_distance_for_each_vehicle AS
SELECT Vehicle_Type, AVG(Ride_Distance) FROM booking GROUP BY Vehicle_Type;
SELECT * FROM average_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 booking WHERE Booking_Status="Canceled by Customer";
SELECT * FROM cancelled_rides_by_customers;
```

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

```
CREATE VIEW top_5_customer AS
SELECT Customer_ID, COUNT(Booking_ID) AS total_rides FROM booking
GROUP BY Customer_ID
ORDER BY total_rides DESC LIMIT 5;
SELECT * FROM top_5_customer;
```

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

```
CREATE VIEW cancelle_ddrivers_personal_car_related_issues AS
SELECT COUNT(*) FROM booking WHERE Canceled_Rides_by_Driver="Personal & Car related issue";
SELECT * FROM cancelle_ddrivers_personal_car_related_issues;
```

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

```
CREATE VIEW max_min_driver_ratings AS
SELECT MAX(Driver_Ratings) AS max_rating, MIN(Driver_Ratings) AS min_rating FROM booking WHERE Vehicle_Type = "Prime Sedan";
SELECT * FROM max_min_driver_ratings;
```

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

```
CREATE VIEW payment_upi AS
SELECT * FROM booking WHERE Payment_Method="UPI";
SELECT * FROM payment_upi;
```

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

```
CREATE VIEW avg_customer_rating AS
SELECT Vehicle_Type, AVG(Customer_Rating) FROM booking
GROUP BY Vehicle_Type;
SELECT * FROM avg_customer_rating;
```

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

```
CREATE VIEW Total_booking_value AS
SELECT SUM(Booking_Value) AS total_booking FROM booking WHERE Booking_Status
='Success';
SELECT * FROM Total_booking_value;
```

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

```
CREATE VIEW incomplete_ride_reason AS
SELECT Booking_ID, Incomplete_Rides_Reason FROM booking WHERE
Incomplete_Rides = "Yes";
SELECT * FROM incomplete_ride_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 averagecustomer ratings for different vehicle types.
5. cancelled Rides Reasons: A bar chart that highlights the common reasons for ridecancellations by customers and drivers.
6. Revenue by Payment Method: A stacked bar chart displaying total revenue based onpayment methods (Cash, UPI, Credit Card, etc.).
7. Top 5 Customers by Total Booking Value: A leaderboard visual listing customers who havespent 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.