



# Data Analyst Project

By Saurabh Singh

## ChatGPT Prompt to Create Data

**Please create a spreadsheet with 20,000 rows, for Gurugram 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. Find the average customer rating per vehicle type.
8. Calculate the total booking value of rides completed successfully.
9. 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 Answers:

### 1. Retrieve all successful bookings:

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

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

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

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

```
SELECT COUNT(*) FROM bookings WHERE Booking_Status = 'cancelled 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 bookings GROUP BY Customer_ID ORDER BY total_rides DESC LIMIT 5;
```

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

```
SELECT COUNT(*) FROM bookings WHERE cancelled_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 bookings WHERE Vehicle_Type = 'Prime Sedan';
```

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

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

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

```
SELECT SUM(Booking_Value) as total_successful_value FROM bookings WHERE Booking_Status = 'Success';
```

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

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

## 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