**SYNOPSIS**



**Data Analysis Project**

**NAME: Mohan S**

**Reg. No:** **AF0380435**

**Using ChatGPT Prompt to Create Data.**

**Please create a spreadsheet with 1 lakh rows, for Bengaluru city. Give the following columns. The data will be for December 2024. 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

1. Incomplete Rides
2. Incomplete Rides Reason

* Customer Demand
* Vehicle Breakdown
* Other Issue

1. Booking Value
2. Ride Distance
3. Driver Ratings
4. Customer Rating

**CONSTRAINTS:**

* 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%
* Keep order value high on weekends.

Business 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 | 1. C\_TAT 2. cancelled\_Rides\_by\_Customer 3. cancelled\_Rides\_by\_Driver 4. Incomplete\_Rides 5. Incomplete\_Rides\_Reason 6. Booking\_Value 7. Payment\_Method 8. Ride\_Distance 9. Driver\_Ratings 10. Customer\_Rating |

Segregation of the views:

* 1. **Overall**
  + Ride Volume Over Time
  + Booking Status Breakdown
  1. **Vehicle Type**

- Top 5 Vehicle Types by Ride Distance

* 1. **Revenue**
  + Revenue by Payment Method
  + Top 5 Customers by Total Booking Value
  + Ride Distance Distribution Per Day
  1. **Cancellation**
  + Cancelled Rides Reasons (Customer)
  + cancelled Rides Reasons(Drivers)
  1. **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, analyzing correlations.