OLA Data Analyst Project

Prompt for dataset:

please create a spreadsheet with 1 lac rows, for pune city. give the following columns.
The data will be for 1 month. Use following columns-

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 pune)
8) Drop location(Take from dummy pickup locations)
9) Avg VTAT(time taken to arrive the vehicle)
10) Avg CTAT(time taken to arrive the customer)
11) Canceled 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 issue
-customer related issue
-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 overall booking status success for this data 62%.if booking status is success then only fare charges ratings,Avg VTAT,Avg CTAT and other data will be there.

Make sure order canceled by customers should not be more than 7%
Make sure order canceled by Drivers should not be more than 18%

Also,make sure to increase the number of orders on weekends and match days.keep match day 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 digit start with CNR and then digits
keep orders under 500 value 70%
keep order above 500 value 28%
keep remaining orders above 1000
Keep overall booking status

SQL Answers:

---1) Retrieve all successful bookings:

```
SELECT * FROM ola1
WHERE booking_status = 'Success';
```

---2) Find the average ride distance for each vehicle type:

```
SELECT vehicle_type,AVG(ride_distance) AS avg_distance
FROM ola1
GROUP BY vehicle_type;
```

---3) Get the total number of canceled rides by the customers:

```
SELECT COUNT(*) FROM ola1
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 ola1
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 ola1
WHERE reason_for_cancelling_by_driver = 'personal & Car related issue'
```

---6) find the maximum and minimum driver rating for prime sedan bookings:

```
SELECT MAX(driver_ratings)AS max_rating,
```

```
MIN(driver_ratings)AS min_rating
FROM ola1
WHERE vehicle_type = 'Prime sedan';
```

---7) find the average customer rating per vehicle type:

```
SELECT vehicle_type,AVG(customer_rating)AS avg_cust_rating
FROM ola1
GROUP BY vehicle_type;
```

---8) calculate the total booking value of rides completed successfully:

```
SELECT SUM(booking_value) AS total_booking
FROM ola1
WHERE booking_status = 'Success';
```

---9) List all incomplete rides along with the reason:

```
SELECT booking_id,incomplete_rides_reason
FROM ola1
WHERE incomplete_rides = 1;
```

Power BI 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 donut chart displaying the proportion of different booking status.

3.Top five 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 customers ratings of different vehicle types.

5.Canceled rides reasons:A bar chart that highlights the common reasons for ride cancellations by customers and drivers.

6.Revenue:A stack bar chart displaying total revenue.

7.Top 5 Customers by total Booking value:A leaderboard visual listing customers who have spend the most on bookings.

8.Ride distance Distribution:A histogram or scatter plot showing the distribution of ride distance for different bookings.

9.Driver Ratings 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.