Data Analyst Project

ABOUT THE PROJECT:

This project demonstrates a comprehensive analysis of a dataset containing **20,000 rows** of Ola data, utilizing SQL, Excel, and Power BI. The objective is to extract valuable insights from raw data and present them effectively through interactive visualizations. The project covers the complete data analytics workflow, from data extraction to creating impactful dashboards.

SQL for Data Extraction & Preparation:

- Queried a large dataset of 20,000 rows to extract relevant information.
- Cleaned and prepared the data for analysis.
- Managed complex data structures efficiently to address real-world problems.

Excel for Data Analysis:

- Conducted in-depth data analysis using tools like sorting, pivot tables, and advanced formulas.
- Summarized and interpreted data to extract actionable insights.
- Showcased Excel's versatility in handling analytical tasks.

Power BI for Dashboard Creation:

- Created an interactive and visually engaging dashboard to present findings.
- Transformed data insights into compelling visual stories for stakeholders.
- Highlighted trends and key metrics effectively using dynamic visualizations.

End-to-End Workflow Execution:

- Executed a complete data analytics workflow, from raw data to insightful dashboards.
- Developed critical data analytics skills and gained a holistic understanding of the process.

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 Rating

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 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;
select * from bookings;
```

#1. Retrieve all successful bookings:

```
Create View Successful_Bookings As
SELECT * FROM bookings
WHERE Booking_Status = 'Success';
```

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

```
Create View ride_distance_for_each_vehicle As
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:

```
Create View cancelled_rides_by_customers As
SELECT COUNT(*) FROM bookings
WHERE Booking_Status = "Canceled by Customer";
```

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

```
Create View Top_5_Customers As
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:

```
Create View Rides_cancelled_by_Drivers_P_C_Issues As
SELECT COUNT(*) FROM bookings
WHEREcancelled_Rides_by_Driver = 'Personal & Car related issue';
```

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

```
Create View Max_Min_Driver_Rating As
SELECT MAX(Driver_Ratings) as max_rating,
MIN(Driver_Ratings) as min_rating
FROM bookings WHERE Vehicle_Type = 'Prime Sedan';
```

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

```
Create View UPI_Payment As
SELECT * FROM bookings
WHERE Payment_Method = 'UPI';
```

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

```
Create View AVG_Cust_Rating As
SELECT Vehicle_Type, AVG(Customer_Rating) as avg_customer_rating
FROM bookings
GROUP BY Vehicle_Type;
```

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

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

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

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

Retrieve All Answers

#1. Retrieve all successful bookings:

```
Select * From Successful_Bookings;
```

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

```
Select * from ride_distance_for_each_vehicle;
```

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

```
Select * from cancelled_rides_by_customers;
```

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

```
Select * from Top_5_Customers;
```

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

```
Select * from Rides_cancelled_by_Drivers_P_C_Issues;
```

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

```
Select * from Max_Min_Driver_Rating;
```

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

```
Select * from UPI_Payment;
```

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

```
Select * from AVG_Cust_Rating;
```

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

```
Select * from total_successful_ride_value;
```

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

```
Select * from Incomplete_Rides_Reason;
```