Ola Rides Analytics Dashboard – Project Report

1. Introduction

This project focuses on analyzing **Ola ride data** (103K+ records) to generate insights into customer behavior, ride performance, cancellations, vehicle usage, and revenue distribution. The dataset was cleaned, transformed, stored in MySQL, and connected to Power BI & Streamlit for visualization.

---

2. Objectives

- Understand ride volume and patterns over time.
- Analyze booking statuses and cancellation reasons.
- Identify top-performing vehicle types and customers.
- Compare customer vs. driver ratings.
- Track revenue distribution across payment methods.
- Build an interactive dashboard for decision-making.

---

3. Data Source & Preprocessing

- **Source:** `ola_rides.xlsx` (103,024 rows, 20 columns).
- **Tools Used:** Python (Pandas, SQLAlchemy), MySQL Workbench, Power BI, Streamlit.
- **Steps Taken:**
    1. Loaded Excel dataset into Python.

2. Cleaned missing values:
    - Replaced nulls in `V_TAT`, `C_TAT` with `0`.
    - Replaced nulls in categorical columns with `"NA"`.
    - Dropped `Vehicle Images` column (all null).
3. Inserted into MySQL database.
4. Built SQL queries for transformations (COALESCE, UPDATE, DROP, etc.).

---

4. Data Model

- **Fact Table:** Rides (Booking_ID, Date, Ride_Distance, Booking_Value).
- **Dimensions:**
    - Customer (Customer_ID, Customer_Rating).
    - Vehicle (Vehicle_Type).
    - Payment (Payment_Method).
    - Status (Booking_Status, Cancellation_Reason).

---

5. Analysis & Key Metrics

5.1 Ride Volume Over Time

- Line chart of daily/monthly rides.
- Shows peak ride demand in evenings & weekends.

5.2 Booking Status Breakdown

- Pie chart: Success vs. Cancelled (by customer/driver).
- ~38% cancellations due to customer/driver reasons.

5.3 Top 5 Vehicle Types by Ride Distance

- Bar chart: Prime Sedan, Mini, Auto lead ride distance.

5.4 Average Customer Ratings by Vehicle Type

- Heatmap: SUVs & Sedans score higher than Autos/Bikes.

5.5 Cancellation Reasons

- Tree map of reasons:
    - Customer → “Driver not moving to pickup”
    - Driver → “Personal/Car-related issues”

5.6 Revenue by Payment Method

- Donut chart: UPI & Cash dominate.
- Credit Cards & Wallets less used.

5.7 Top 5 Customers by Total Booking Value

- Bar chart ranking customers with highest spend.

5.8 Ride Distance Distribution Per Day

- Histogram of ride distances (most between 5–20 km).

5.9 Driver Ratings Distribution

- Bell curve shows majority between 3.5–4.5 stars.

5.10 Customer vs. Driver Ratings

- Scatter plot comparing fairness of ratings.

---

6. Dashboard Design

- **One-page summary in Power BI** with slicers for:
    - Date, Vehicle Type, Payment Method.
- **Thematic Pages:**
    - Overall Trends
    - Vehicle Insights
    - Revenue Insights
    - Cancellation Analysis
    - Ratings Comparison

---

7. Streamlit App Integration

- Sidebar navigation (Overall, Vehicle Type, Revenue, Cancellation, Ratings).
- Embedded **Power BI dashboard** inside Streamlit using `iframe`.
- Python-driven queries for additional insights (e.g., SQL to fetch top customers).

---

8. Tools & Tech Stack

- **Data Cleaning & ETL:** Python (Pandas, SQLAlchemy)
- **Database:** MySQL
- **Visualization:** Power BI, Streamlit
- **Languages:** Python, SQL, DAX

---

9. Business Insights & Recommendations

- **Improve Driver Training** → reduce driver-based cancellations.
- **Promote UPI Discounts** → strengthen top payment method.
- **Encourage SUVs/Prime Vehicles** → higher ratings & revenue.
- **Loyalty Program for Top Customers** → retain high spenders.
- **Dynamic Pricing on Peak Hours** → optimize ride volume.

---

10. Conclusion

This project demonstrates end-to-end analytics — from **data ingestion → cleaning → SQL queries → visualization → interactive app deployment**. The insights can help Ola improve **customer experience, reduce cancellations, and boost revenue strategies**.