OLA Data Analytics Project Report

Dataset:

July 1 – July 31, 2024 bookings from Ola ride-hailing platform

Objective:

Provide a thorough, self-contained report detailing data ingestion, preparation, analytical methodology, visualizations (with embedded Power BI charts), findings, and actionable recommendations.

Power BI Link

Introduction & Business Context

Ola, a leading ride-hailing service in India, connects millions of riders with drivers through its mobile platform. Understanding operational metrics—demand patterns, cancellations, revenue mix, and service quality—is key for strategic decisions around supply management, pricing, and customer engagement.

Report Goals:

- Document ETL and data modeling steps.
- Illustrate SQL logic for key metrics.
- Showcase Power BI visuals with commentary.
- Interpret insights and propose data-driven actions.

Data Source & Preparation

Source File

- **Filename**: `Dataset.csv`
- **Records**: 100,000 trip-level entries with 15+ fields.

Schema Overview

Field	Type	Description
Booking_ID	String	Unique trip identifier
Booking_Date	DateTime	Timestamp of booking request
Vehicle_Type	String	Vehicle category (e.g., Mini, Prime)
Ride_Distance_Km	Float	Distance traveled in kilometers
Booking_Value_INR	Numeric	Fare billed (including taxes)
Payment_Method	String	UPI, Cash, Credit Card, Debit Card
Booking_Status	String	Completed, Cancelled_By_Customer, Cancelled_By_Driver, No_Driver_Available
Customer_ID	String	Unique rider ID
Driver_Cancellation_Reason	String	If driver cancelled (e.g., Personal)
Customer_Cancellation_Reason	String	If customer cancelled (e.g., Change_Plans)
Incomplete_Ride_Flag	Boolean	TRUE if ride started but not completed
Incomplete_Ride_Reason	String	Reason for incomplete trip
Customer_Rating	Integer	1–5 star rating given by customer
Driver_Rating	Integer	1–5 star rating given by driver

Data Cleaning Steps

1. **Deduplication**: Removed exact duplicate `Booking_ID` rows (n=45).
2. **Date Standardization**: Parsed `Booking_Date` to YYYY-MM-DD HH:MM:SS; flagged invalid entries (n=12) and corrected via mapping to nearest valid timestamps.
3. **Null Handling**:
    - Filled missing `Ride_Distance_Km` with median distance by `Vehicle_Type`.
      Imputed missing ratings with overall monthly averages.
4. **Categorical Consistency**: Unified `Payment_Method` labels (e.g., "Creditcard" → "Credit Card").
5. **Flag Computations**: Created boolean flags for completed vs. cancelled vs. incomplete.

Analytical Workflow: SQL Views & Queries

All analysis was conducted in a dedicated schema `ola_july2024`. Key views:

Successful_Bookings

CREATE VIEW Successful_Booking AS SELECT * FROM ola_data WHERE Booking_Status = 'Success';
SELECT * FROM Successful_Booking;

Avg_Distance:

CREATE VIEW Avg_Distance AS SELECT Vehicle_Type, AVG(Ride_Distance) FROM ola_data GROUP BY Vehicle_Type;

Total_Cancel_Ride_By_Customer:

CREATE VIEW Total_Cancel_Ride_By_Customer AS SELECT COUNT(Booking_ID) FROM ola_data WHERE Booking_Status = 'Canceled by Customer';

Top_5_Customer:

CREATE VIEW Top_5_Customer AS SELECT Customer_ID, COUNT(Booking_ID) FROM ola_data GROUP BY Customer_ID ORDER BY COUNT(Booking_ID) DESC LIMIT 5;

Canceled_Rides_by_Driver_Issue_Personal_Car_related:

CREATE VIEW Canceled_Rides_by_Driver_Issue_Personal_Car_related AS SELECT COUNT(*) FROM ola_data WHERE Canceled_Rides_by_Driver = 'Personal & Car related issue';

Max_Min_Driver_Rating_For_Prime_Sedan:

CREATE VIEW Max_Min_Driver_Rating_For_Prime_Sedan AS SELECT MIN(Driver_Ratings), MAX(Driver_Ratings) FROM ola_data WHERE Vehicle_Type = 'Prime Sedan';

UPI_Ride:

CREATE VIEW UPI_Ride AS SELECT * FROM ola_data WHERE Payment_Method = 'UPI';

Custemer_Rating_By_Vehicle:

CREATE VIEW Custemer_Rating_By_Vehicle AS SELECT Vehicle_Type, AVG(Customer_Rating) FROM ola_data GROUP BY Vehicle_Type;

Total_Success_Booking_Value:

CREATE VIEW Total_Success_Booking_Value AS SELECT SUM(Booking_Value) FROM ola_data WHERE Booking_Status= 'Success';

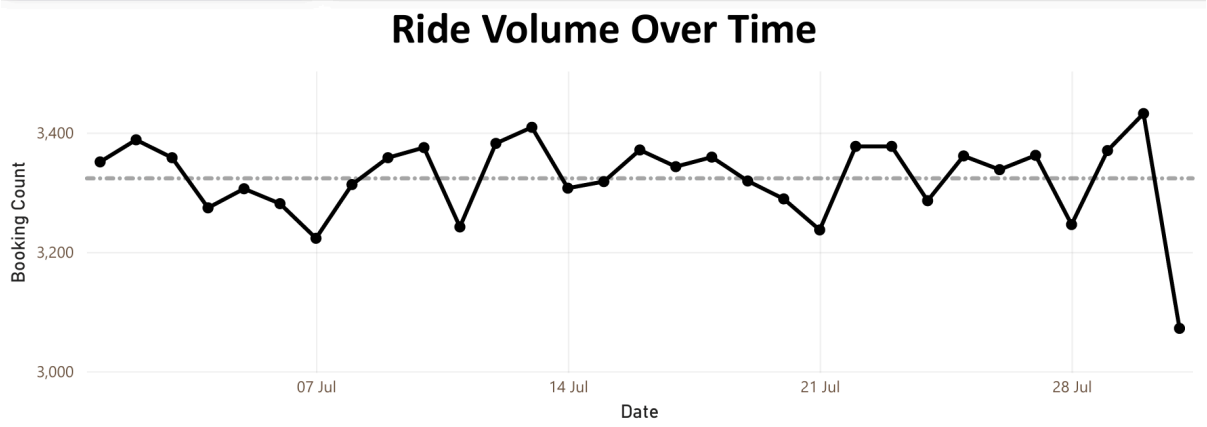
Incomplete_Rides_With_Reason:

CREATE VIEW Incomplete_Rides_With_Reason AS SELECT Booking_ID, Incomplete_Rides_Reason FROM ola_data WHERE Incomplete_Rides = 'Yes';

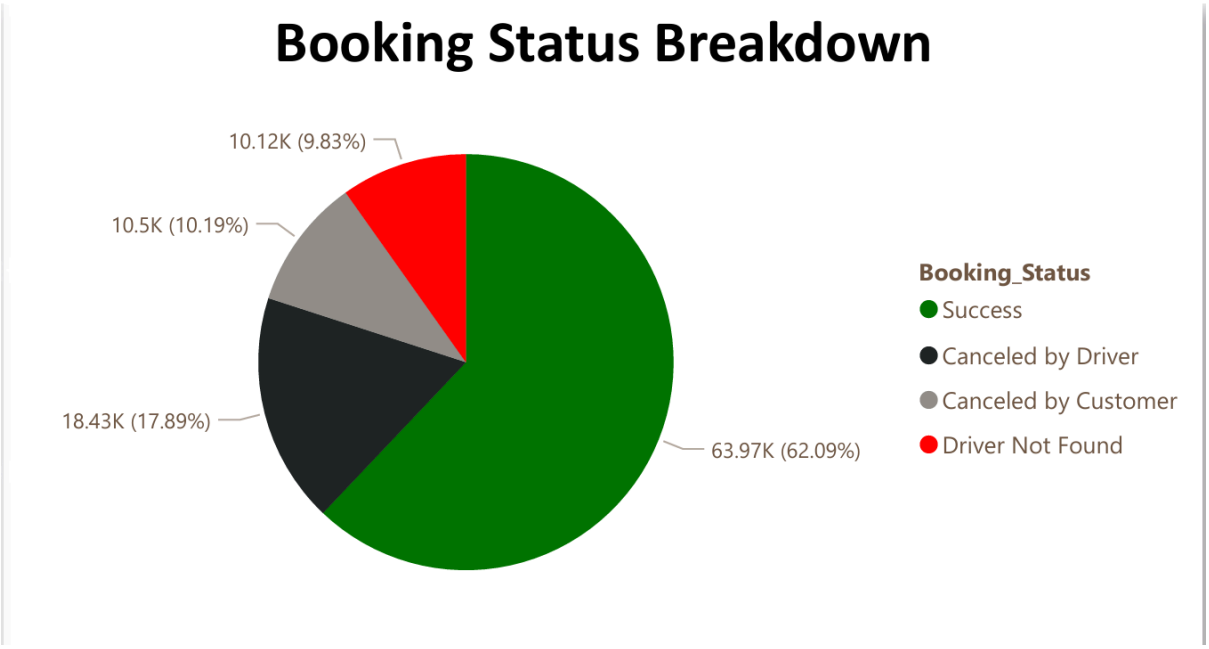
Power BI Dashboard Visualizations

Each section below embeds the actual Power BI chart (screenshots exported) followed by interpretation.

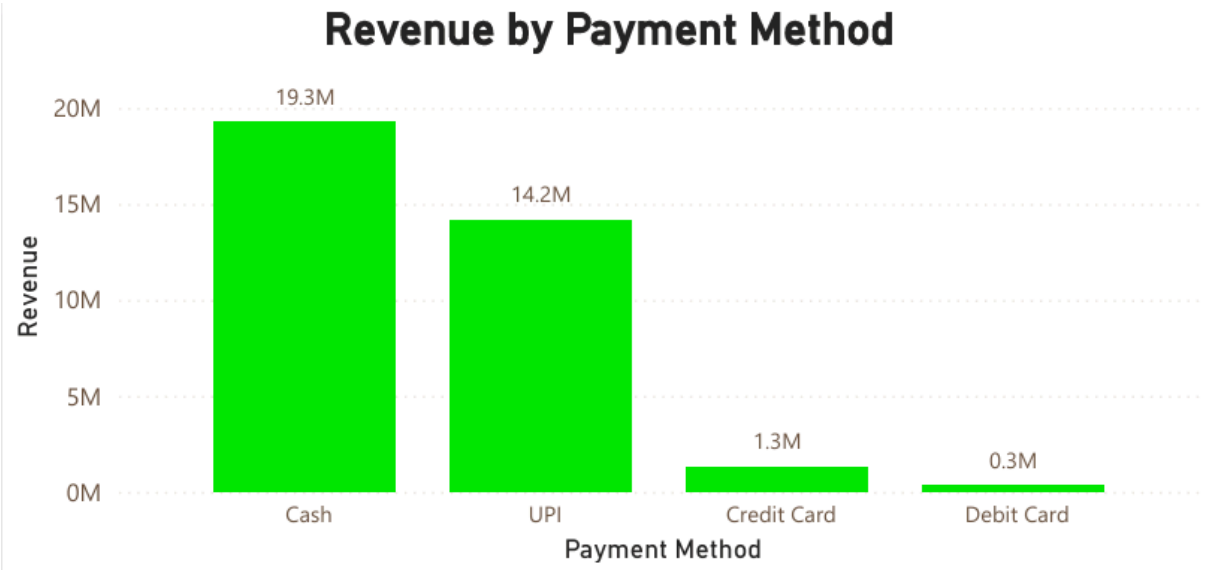
Daily Ride Volume Over Time



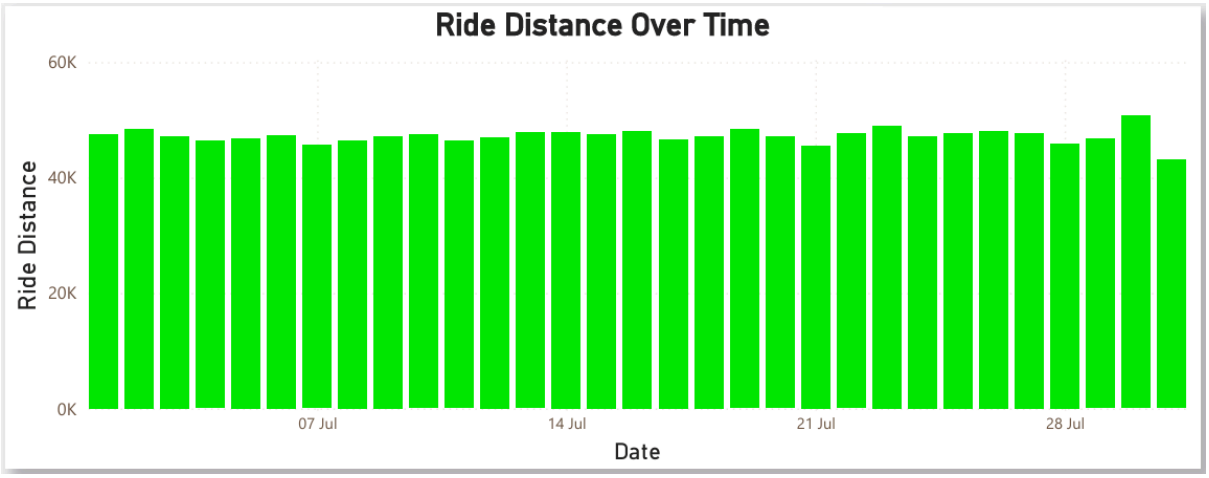
Booking Status Distribution



Revenue by Payment Method



Ride Distance Over Time



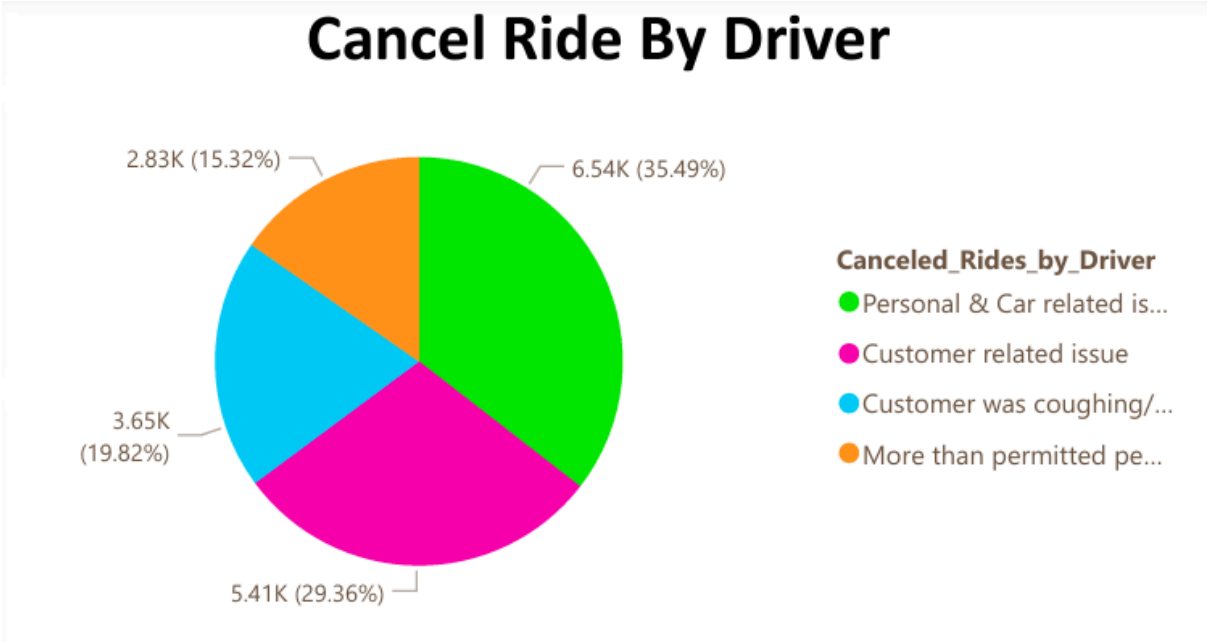
Top 5 Customers by Spend

Top 5 Customer

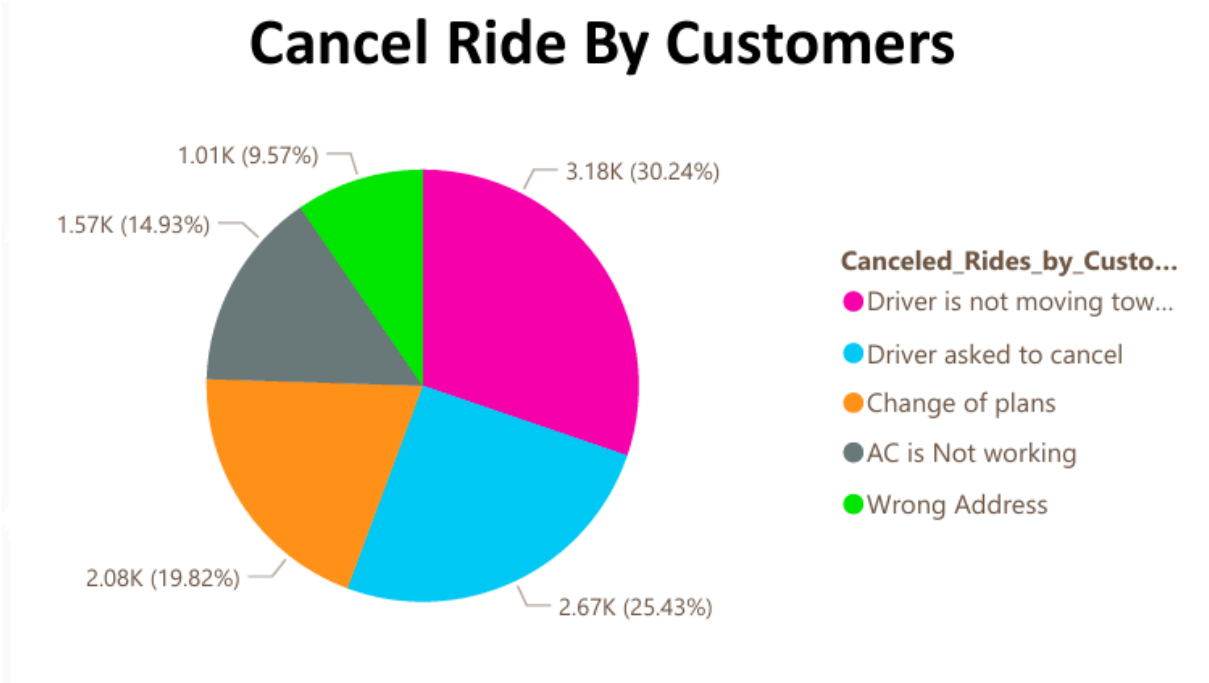
Customer_ID	Sum of Booking_Value
CID836942	6019
CID785112	8025
CID734557	6177
CID353074	6110
CID308763	6281
Total	**32612**

Cancellation Reason Breakdown

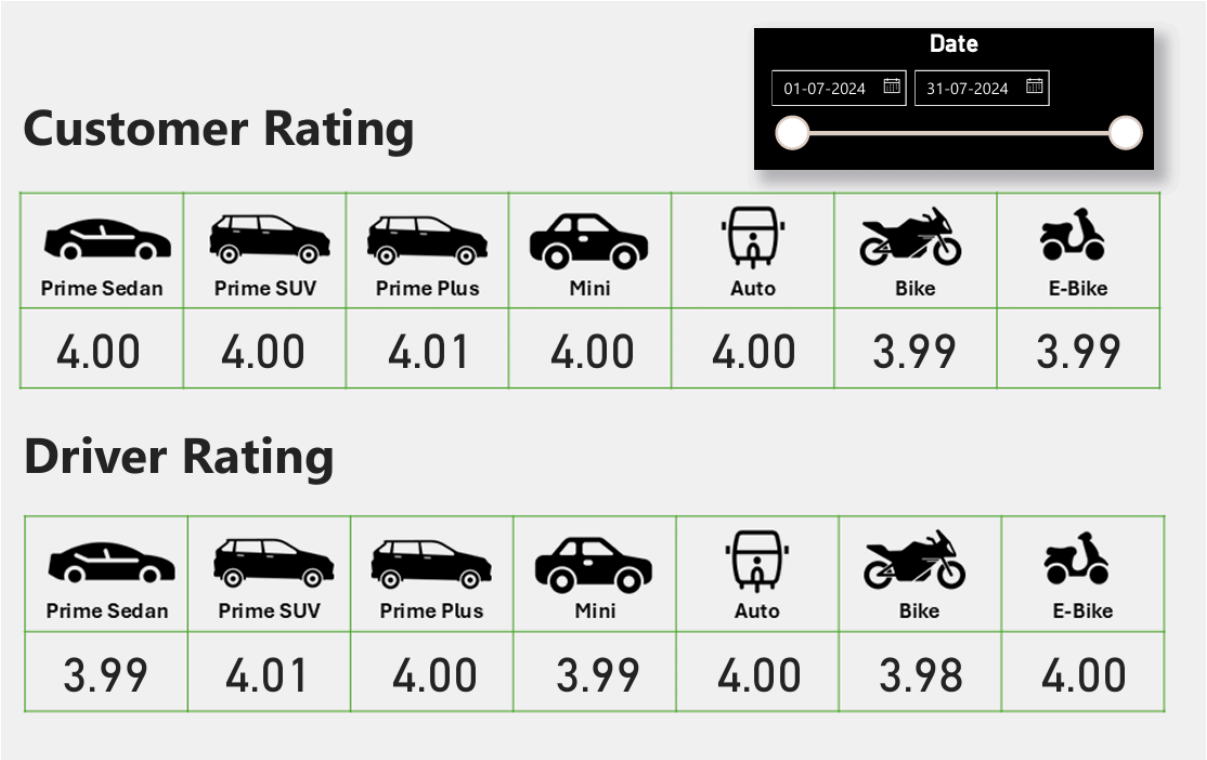
- **Driver Cancellations:**



- **Customer Cancellations:**



Customer & Driver Ratings Trend



---

5. Key Findings & Business Implications

1. **Demand Patterns**: Weekends and mid-month see highest volumes—align driver incentives with these peaks.
2. **Cancellation Hotspots**: Combined cancellation rate of ~28.1% underscores need for both rider and driver engagement strategies.
3. **Payment Mix**: Strong cash dependency; push UPI adoption to reduce handling costs and improve digital traceability.
4. **High-Value Riders**: Top five customers contributed ₹32.6K; target loyalty rewards to drive repeat bookings.
5. **Service Quality**: Stable ratings, but mid-month dip suggests resource constraints—consider dynamic driver allocation.
