OLA Ride Analysis Project

Data Analytics Documentation Report

Prepared By

Arpit Dalal

arpitsunildalal@gmail.com

https://www.linkedin.com/in/arpitdalal9/

Note: Both SQL and Power BI analyses are based on the same underlying dataset, ensuring consistency in insights and recommendations across all analytical approaches.

1. Executive Summary

This project presents a comprehensive analysis of ride-hailing operations for Ola, one of India's leading cab aggregator services. The analysis leverages both **SQL** for data extraction and **Power BI with DAX** for interactive visualization and business intelligence.

Project Objective: To analyze booking patterns, customer behavior, revenue trends, cancellation reasons, and operational efficiency in order to generate actionable insights that drive business growth and enhance service quality.

Key Technologies:

- **Database:** MySQL
- **Business Intelligence:** Power BI
- **Languages:** SQL, DAX (Data Analysis Expressions)
- **Dataset Size:** 100,000 booking records
- **Time Period:** One month of operations in Bengaluru, India

2. Project Overview

2.1 Dataset Description

The analysis is based on a comprehensive dataset containing ride-hailing data with the following characteristics:

- **Total Records:** 100,000 bookings
- **Location:** Pune Region
- **Duration:** One month of operational data
- **Booking Success Rate:** ~62%
- **Vehicle Types:** Auto, Bike, eBike, Mini, Prime Plus, Prime Sedan, Prime SUV
- **Payment Methods:** UPI, Cash, Credit Card, Debit Card, Wallet

2.2 Data Columns

The dataset includes 19 columns covering various aspects of ride operations:

No.	Column Name	Description
1	Date	Booking date
2	Time	Booking time
3	Booking_ID	Unique booking identifier
4	Booking_Status	Status (Success, Cancelled by Customer, Cancelled by Driver, Incomplete)
5	Customer_ID	Unique customer identifier
6	Vehicle_Type	Type of vehicle booked
7	Pickup_Location	Starting location
8	Drop_Location	Destination location
9	V_TAT	Vehicle Time to Arrival
10	C_TAT	Customer Time to Arrival
11	Cancelled_Rides_by_Customer	Customer cancellation indicator
12	Cancelled_Rides_by_Driver	Driver cancellation indicator
13	Incomplete_Rides	Incomplete ride indicator
14	Incomplete_Rides_Reason	Reason for incomplete rides
15	Booking_Value	Fare amount
16	Payment_Method	Mode of payment
17	Ride_Distance	Distance traveled in kilometers
18	Driver_Ratings	Customer rating for driver (1–5)
19	Customer_Rating	Driver rating for customer (1–5)

3. SQL Analysis

3.1 Purpose

SQL queries were designed to extract key business metrics and answer critical operational questions. The analysis uses MySQL database with VIEW creation for reusable query structures.

3.2 Key Business Questions Answered

The SQL analysis addresses 10 critical business questions:

1. **Successful Bookings Analysis** - Identifying completed rides for revenue calculation
2. **Vehicle Performance** - Average ride distance by vehicle type
3. **Customer Cancellations** - Quantifying customer-initiated cancellations
4. **Customer Loyalty** - Identifying top 5 most frequent customers
5. **Driver Cancellations** - Analyzing driver-side operational issues
6. **Service Quality** - Rating analysis for Prime Sedan category
7. **Payment Preferences** - UPI payment adoption analysis
8. **Customer Satisfaction** - Average ratings by vehicle type
9. **Revenue Calculation** - Total booking value from successful rides
10. **Incomplete Rides** - Identifying and categorizing incomplete journeys

3.3 Technical Approach

- **Database:** MySQL
- **Views Created:** 10 reusable views for each business question
- **Query Types:** Aggregations (SUM, AVG, COUNT), Filtering (WHERE), Grouping (GROUP BY), Sorting (ORDER BY)
- **Optimization:** Views enable query reusability and simplified reporting

4. Power BI Dashboard Analysis

4.1 Dashboard Architecture

The Power BI dashboard consists of **5 interactive pages**, each focusing on specific business dimensions:

No.	Page Name	Description
1	Overview Page	High-level KPIs and performance metrics
2	Vehicle Page	Vehicle type analysis and contribution
3	Revenue Page	Revenue analysis across multiple dimensions
4	Rider Page	Customer behavior analysis and segmentation
5	Location Page	Geographic and temporal distribution analysis

4.2 DAX Measures and Calculations

DAX (Data Analysis Expressions) is used extensively to create:

Category	Description
KPI Measures	Booking counts, revenue metrics, and distance metrics
Calendar Table	Time intelligence table for temporal analysis
Customer Segmentation	Categorization based on ride frequency
Time Slot Analysis	Identification of peak hours
Calculated Columns	Weekend flags, weekday names, and month names

4.3 Page-by-Page Breakdown

Page 1: Overview

KPIs Displayed:

- Completed Bookings: 67,000
- Lost Bookings: 33,000
- Total Revenue: ₹31.17 Million
- Total Distance: 1.09 Million km
- Average Distance: 16.20 km

Visualizations:

- Booking status breakdown (Success: 67.31%, Driver Cancelled: 19.59%, Customer Cancelled: 7.68%)
- Daily booking trends (line chart)
- Daily revenue trends (bar chart)
- Revenue by vehicle type (bar chart)
- Top locations by booking count
- Average customer and driver ratings (4.40 stars)

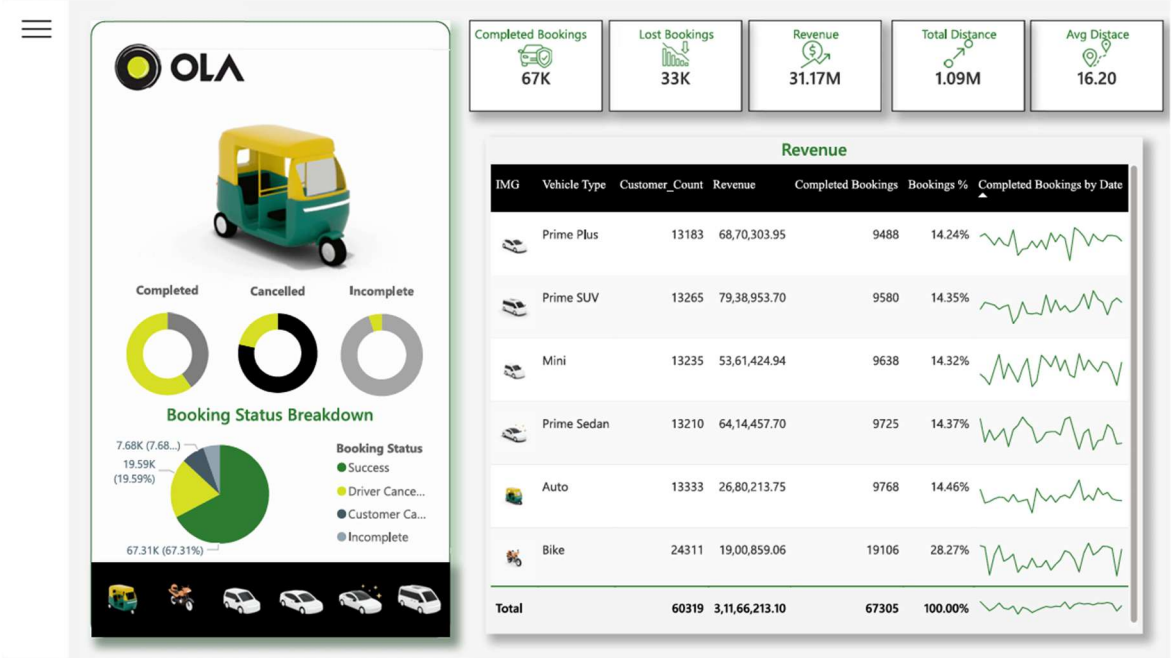
Page 2: Vehicle Analysis

Metrics by Vehicle Type:

- Customer count
- Total revenue
- Completed bookings
- Booking percentage
- Booking trends over time

Key Insights:

- Bike category has highest booking volume (28.27%)
- Prime SUV generates highest revenue (₹7.9M)
- All vehicle types show consistent booking patterns



Page 3: Revenue Analysis

Revenue Breakdown:

- By customer (Top 7 customers identified)
- By payment method (UPI: ₹14.0M, Cash: ₹11.9M)
- By vehicle type
- Monthly and quarterly trends

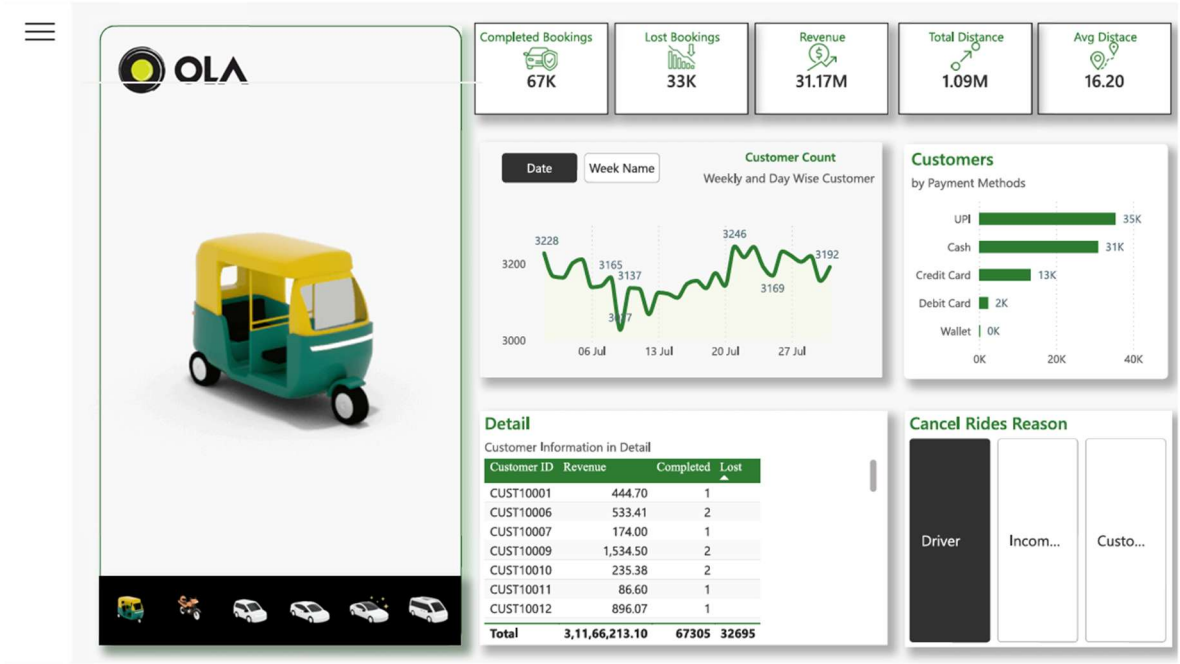
Page 4: Rider Page

Customer Segmentation:

- New Customer: 1 ride
- Occasional: 2-3 rides
- Frequent: 4-6 rides
- Loyal: 6+ rides

Analysis:

- Cancellation reasons breakdown
- Payment method distribution (UPI: 35K, Cash: 31K)
- Weekly customer count trends
- Detailed customer information table



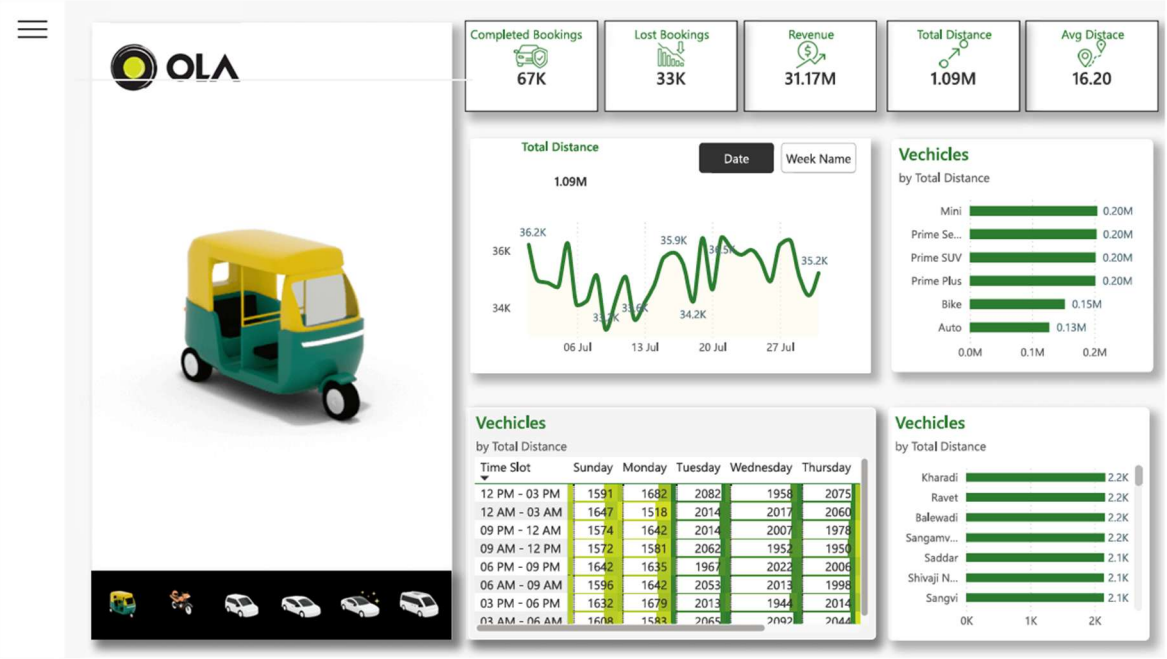
Page 5: Location Page

Spatial and Temporal Analysis:

- Monthly total distance trends
- Distance by vehicle type
- Busy time slots (8 time slots: 3-hour intervals)
- Busy areas (Top 7 locations)
- Weekday vs weekend patterns

Time Slot Analysis:

Peak booking times identified across different weekdays using heat map visualization.



5. Key Findings and Insights

5.1 Operational Performance

1. **Success Rate:** 67.31% booking success rate indicates room for improvement
2. **Cancellation Impact:** 19.59% driver cancellations and 7.68% customer cancellations affect revenue
3. **Average Trip:** 16.20 km average distance suggests medium-distance urban mobility

5.2 Vehicle Performance

1. **Volume Leader:** Bike category dominates with 28.27% of bookings
2. **Revenue Leader:** Prime SUV generates highest revenue despite lower volume
3. **Balanced Distribution:** Other vehicle types show relatively equal distribution

5.3 Customer Behavior

1. **Payment Preference:** UPI is the dominant payment method (₹14M revenue)
2. **Customer Loyalty:** Top customers generate ₹4-5K individual revenue
3. **Ratings:** Consistent 4.40 average rating indicates good service quality

5.4 Geographic and Temporal Patterns

1. **Top Locations:** Kharadi and Ravet lead in booking counts
2. **Peak Hours:** Multiple peak periods throughout the day
3. **Distance Distribution:** Consistent across vehicle types with slight variations

6. Business Recommendations

6.1 Reduce Cancellation Rates

Issue: 27.27% combined cancellation rate (driver + customer)

Recommendations:

- Implement driver incentives for ride acceptance
- Improve driver-customer matching algorithms
- Address top cancellation reasons through targeted interventions

6.2 Optimize Vehicle Deployment

Insight: Bike category has high volume but lower revenue

Recommendations:

- Balance fleet composition to maximize revenue per vehicle
- Deploy Prime SUVs during peak demand periods
- Offer promotions on underutilized vehicle types

6.3 Enhance Digital Payment Adoption

Insight: UPI and Cash dominate; other methods underutilized

Recommendations:

- Incentivize credit/debit card usage with cashback
- Promote wallet adoption through loyalty programs
- Ensure seamless payment experience across all methods

6.4 Customer Retention Strategies

Insight: Customer segmentation reveals loyalty patterns

Recommendations:

- Develop loyalty programs for frequent and loyal customers
- Re-engagement campaigns for occasional customers
- First-ride incentives for new customers

7.Technical Implementation

7.1 SQL Implementation

Database Setup:

```
  -- Create Database
CREATE DATABASE Ola;
  -- Select / Use Database
USE Ola;
```

View Creation Strategy:

- 10 views created for different analytical purposes
- Views enable easy data access for Power BI integration
- Optimized queries for performance

7.2 Power BI Implementation

Data Model:

- Calendar table created using DAX for time intelligence
- Relationships established between bookings and calendar tables
- Calculated columns and measures defined

Dashboard Features:

- Interactive filters for dynamic analysis
- Hide/show filter panel for better space utilization
- Drill-through capabilities for detailed exploration
- Consistent formatting and corporate theme

8. Project Deliverables

1. **SQL Queries Document** - Complete set of 10 SQL queries with explanations
2. **DAX Queries Document** - All DAX measures and calculated columns
3. **Power BI Dashboard** - 5-page interactive dashboard with visualizations
4. **Documentation Report** - This comprehensive project document

9. Conclusion

The **OLA Ride Analysis Project** successfully demonstrates end-to-end data analytics capabilities, from data extraction using SQL to interactive visualization using Power BI. The analysis provides actionable insights across multiple business dimensions:

- **Operational efficiency** can be improved by reducing cancellation rates
- **Revenue optimization** opportunities exist in vehicle type and payment method strategies
- **Customer retention** programs can be designed based on segmentation insights
- **Geographic and temporal patterns** inform optimal resource deployment

This project serves as a robust foundation for data-driven decision-making in the ride-hailing industry and showcases proficiency in SQL, DAX, and Power BI for business analytics.

10. About This Project

Category	Details
Project Type	**Data Analytics Portfolio Project**
Domain	**Transportation & Mobility**
Tools Used	**MySQL, Power BI, DAX**
Dataset	**Synthetic data (100,000 records)**
Duration	**One month operational data**
Location	**Pune, Maharashtra, India**

Skills Demonstrated:

- SQL query writing and optimization
- DAX for business intelligence
- Power BI dashboard development
- Data modeling and relationships
- Business analysis and insight generation
- Data visualization and storytelling