



Data Analyst Project

By -- Sushant

1. Project Overview :

This project focuses on analyzing a large-scale dataset consisting of **1 lakh+ OLA ride records** from Bengaluru over a one-month period. The primary goal of this analysis is to gain a deep understanding of customer booking behavior, ride patterns, operational challenges, and overall performance metrics within the ride-hailing ecosystem.

The project involves studying various critical aspects such as booking trends, cancellation patterns, customer and driver ratings, vehicle-type performance, and revenue distribution across multiple categories. By examining these metrics, the analysis aims to uncover hidden patterns, identify inefficiencies, and highlight improvement areas to enhance customer satisfaction and operational efficiency.

To achieve this, the project utilizes **Power BI** for interactive dashboard creation and visualization, **SQL** for querying and extracting actionable insights from the data, **Excel** and **Power Query** for initial data cleaning and transformation, and **DAX** for building dynamic measures and KPIs within dashboards.

Overall, this project demonstrates a complete end-to-end data analysis workflow—from data preparation and exploration to visualization and business insights—providing valuable recommendations that can help OLA optimize ride allocation, reduce cancellations, improve user experience, and boost revenue performance.

2. Project Objective:

The primary objective of this project is to perform a comprehensive analysis of OLA's ride data in order to uncover meaningful patterns and derive insights that can support business decision-making and operational improvements. The specific goals include:

1. Identify Key Ride Trends Over Time

Analyze daily and hourly ride volumes to understand demand patterns, peak hours, weekday vs weekend differences, and seasonal fluctuations. This helps in forecasting future ride demand and optimizing driver allocation.

2. Understand Major Cancellation Reasons

Study customer-initiated and driver-initiated cancellations to identify the top reasons behind failed rides. This includes analyzing operational inefficiencies such as driver availability, navigation issues, customer behavior, and external factors impacting ride fulfillment.

3. Analyze Customer & Driver Ratings

Evaluate customer satisfaction and driver performance through rating patterns. Identify factors influencing low ratings, assess service quality, and compare ratings across vehicle types and ride attributes.

4. Evaluate Vehicle-Type Performance

Compare the performance of different vehicle categories such as Auto, Bike, Mini, Prime Sedan, Prime SUV, etc. Assess them based on ride volume, distance covered, customer preference, revenue contribution, and ratings

5. Study Revenue Distribution & Payment Behavior

Analyze revenue generated across rides, vehicle types, customer segments, and payment modes (Cash, UPI, Card, etc.). Understand which categories drive the most revenue and how payment preferences vary across users.

6. Provide Actionable Insights to Improve OLA Operations

Present data-backed recommendations to reduce cancellations, enhance customer experience, improve driver efficiency, and optimize pricing or service availability. The insights aim to support OLA in making strategic decisions to strengthen overall operational performance.

3. Dataset Description :

The dataset used in this project represents **one month of OLA ride operations in Bengaluru**, consisting of **1,00,000+ ride records**. It captures a wide range of operational, behavioral, and transactional details that allow for deep analysis of booking patterns, customer behavior, and service performance. Each row in the dataset corresponds to a unique ride instance, and includes the following fields:

- Date & Time**

These fields indicate when the ride was booked, allowing analysis of daily/hourly demand patterns, peak periods, weekend vs. weekday differences, and time-based performance metrics.

- Booking ID**

A unique identifier assigned to each ride, used to track individual trips and ensure accuracy in filtering and querying operations.

- Booking Status**

Shows the final state of the ride such as *Successful*, *Cancelled by Customer*, *Cancelled by Driver*, or *Incomplete*. This field is essential for cancellation analysis, operational efficiency studies, and KPI calculations.

- Customer ID**

A unique identifier for each customer, enabling customer-level behavior analysis, top customer identification, and loyalty or usage pattern assessment.

- Vehicle Type**

Specifies the category of vehicle chosen (Auto, Bike, Mini, Prime Sedan, Prime SUV, etc.). This helps compare the performance and popularity of different vehicle types and analyze category-wise revenue or ratings.

- **Pickup & Drop Locations**

These fields capture the geographical start and end points of rides. They help identify high-demand zones, frequently traveled routes, and areas with higher cancellation or delay rates.

- **VTAT (Vehicle Time to Arrive)**

Represents the time taken by the driver to reach the pickup location. Used to analyze operational efficiency, driver availability, and delay patterns.

- **CTAT (Customer Time to Arrive)**

Measures the time taken by the customer to reach the pickup point. Helps in understanding customer delay trends and their impact on cancellations.

- **Cancelled by Customer (Reason)**

If the ride was cancelled by the customer, this field documents the reason (e.g., driver not moving, wrong address, AC not working, change of plans). This helps identify service quality issues and customer pain points.

- **Cancelled by Driver (Reason)**

If the driver cancelled the ride, the reason is recorded (e.g., personal issue, customer behavior issue, sick customer, extra passengers). This assists in diagnosing driver-related issues and operational constraints.

- **Incomplete Rides (Reason)**

Provides insights into rides that started but did not complete due to issues such as vehicle breakdown, customer demand, or other disruptions.

- **Booking Value**

Indicates the revenue earned from each ride. This is useful for revenue analysis, vehicle-type profitability, and payment method trends.

- **Payment Method**

Specifies whether the customer paid via Cash, UPI, Credit/Debit Card, etc., helping analyze customer payment preferences and platform-specific revenue shares.

- **Ride Distance**

Shows the total distance covered in each trip. This helps analyze typical travel distances, long-route rides, and operational efficiency by vehicle type.

- **Driver Rating & Customer Rating**

These fields capture satisfaction levels from both the customer and driver perspectives. They allow correlation studies, service quality evaluation, and identification of factors affecting ratings.

4. Tools & Technologies Used :

- Power BI (Dashboard Development + DAX Measures)**

Power BI was used to build interactive dashboards that visualize key metrics such as ride volume, booking status distribution, cancellation patterns, revenue trends, and rating comparisons.

Using DAX (Data Analysis Expressions), custom measures and calculated fields were created to track KPIs like total bookings, success rate, cancellation percentage, average ride distance, revenue contribution, and customer-driver rating differences. The dashboard allows dynamic filtering and drill-down analysis for deeper insights.

- SQL (Data Extraction, Filtering & Analytical Queries)**

SQL played a crucial role in performing detailed data analysis and retrieving insights from the dataset. Various SQL queries were written to:

- Identify successful and cancelled rides
- Analyze ride distance by vehicle type
- Retrieve top customers
- Calculate total revenue from successful trips
- Analyze cancellation reasons
- Evaluate rating distributions

SQL ensured accurate aggregation and segmentation of the data before visualization.

- Excel / Power Query (Data Cleaning & Preprocessing)**

Excel and Power Query were used for initial data cleaning and transformation. Tasks included:

- Removing duplicates and handling missing values
- Standardizing vehicle types and booking status labels
- Formatting date-time values
- Validating business rules (success rate, cancellation thresholds)

- Preparing the dataset for SQL and Power BI
Power Query automated the cleaning process, making the data ready for analysis.

- **DAX (Calculated Measures & KPI Development)**

DAX formulas were used within Power BI to create meaningful KPIs such as:

- Average customer and driver ratings
- Revenue per day and per vehicle type
- Cancellation %
- Distance-based performance metrics
- Trend calculations and time-intelligence measures

These DAX expressions enabled more accurate, dynamic, and interactive insights in the dashboard.

5. Data Cleaning & Preparation:

The data cleaning and preparation phase was carried out using **Excel** and **Power Query**, ensuring the dataset was accurate, consistent, and fully ready for analysis. Key steps included:

- Removal of NULL Values & Duplicates**

The dataset was inspected for missing values, incomplete entries, and duplicate records. All NULL and duplicate rows were removed or appropriately handled to maintain data integrity and prevent misleading analysis.

- Standardization of Categorical Fields**

Columns such as **Booking Status**, **Vehicle Type**, and **Payment Method** contained inconsistencies in formatting (e.g., misspellings, case variations). These were standardized to uniform labels to ensure accurate grouping and analysis across all tools.

- Extraction of Date & Time Components**

The timestamp field was split into separate **Date** and **Time** columns. This enabled time-series analysis, peak hour identification, and weekday/weekend trend evaluation.

- Verification of Business Constraints**

To ensure realistic and usable data, the dataset was validated against predefined business rules:

- ~62% successful ride completion rate
- <7% cancellations initiated by customers
- <18% cancellations initiated by drivers
- <6% incomplete rides

Records not aligning with these expected ratios were corrected or regenerated to maintain dataset authenticity.

- Validation of Ride Distance and Ratings**

Ride distance values were checked to ensure no negative or unrealistic distances existed. Similarly, customer and driver ratings were validated to fall within the standard rating range (1.0 to 5.0), enabling accurate analysis of satisfaction metrics.

- **Ensuring Accuracy of Numerical Fields**

Metrics such as **Booking Value**, **Vehicle Time to Arrive (VTAT)**, and **Customer Time to Arrive (CTAT)** were reviewed for outliers, incorrect entries, and unrealistic times. Inaccurate values were corrected or removed based on reasonable operational thresholds.

6. SQL Analysis:

SQL played a crucial role in extracting deeper insights from the OLA ride dataset. A series of analytical queries were written to answer key business questions related to performance, cancellations, ratings, and customer activity. Below are the major SQL analyses performed:

- Retrieve All Successful Bookings**

This query filtered the dataset to show only rides that were successfully completed. It helped calculate success rate, revenue contribution, and overall service reliability.

- Calculate Average Ride Distance by Vehicle Type**

By grouping rides based on vehicle category and averaging the distance traveled, this query provided insights into the typical usage pattern of each vehicle type (e.g., Bike vs. Prime Sedan).

- Find Total Cancelled Rides (Customer & Driver)**

Separate queries were written to determine how many rides were cancelled by customers and how many were cancelled by drivers. This helped understand operational bottlenecks and pain points from both sides.

- Identify Top 5 Customers by Number of Bookings**

This query ranked customers based on their total ride count, helping identify high-value users or frequent riders who contribute significantly to the platform.

- **Analyze Reasons for Driver Cancellations**

This analysis grouped driver cancellations by reason (e.g., personal issues, customer behavior, passenger count, sick customer). It highlighted operational challenges affecting driver participation.

- **Find Maximum and Minimum Driver Ratings for Prime Sedan**

Helpful for understanding service quality within a specific vehicle category. Prime Sedan performance was evaluated by checking the highest and lowest ratings received from customers.

- **Retrieve Rides with UPI Payments**

This query filtered rides where customers paid via UPI. Payment method analysis helped understand user payment preferences and digital adoption.

- **Calculate Total Booking Value of Successful Rides**

Summing the booking value of completed rides provided insights into total revenue generated from fulfilled bookings, excluding cancellations and incomplete rides.

- **List All Incomplete Rides with Their Reasons**

This query identified rides that began but did not complete and grouped them by reason (e.g., vehicle breakdown, customer request, other issues). It helped analyze operational disruptions affecting trip completion.

These SQL analyses provided the foundational insights used in the dashboard and formed the basis for business recommendations aimed at improving OLA's operational efficiency and customer satisfaction.

7. Power BI Dashboard Insights Dashboard Visuals:

Dashboard Visuals

The Power BI dashboard was designed to present a clear, interactive, and comprehensive view of OLA's ride operations. The main visuals included:

- Ride Volume Over Time**

A line chart displaying daily and hourly ride counts. This visualization helped identify demand fluctuations, peak traffic periods, weekday/weekend patterns, and special spikes during match days or events.

- Booking Status Breakdown**

A pie or donut chart showing the proportion of successful rides, customer cancellations, driver cancellations, and incomplete rides. This gave a quick overview of operational reliability and service efficiency.

- Cancellation Reasons (Customer & Driver)**

Two bar charts visualizing the top reasons for cancellations.

- Customer cancellations (e.g., driver not moving, wrong address, AC not working)
- Driver cancellations (e.g., personal issues, customer behavior, sick customer)

These visuals highlighted critical service issues affecting ride fulfillment.

- Average Ratings by Vehicle Type**

A cluster bar chart comparing customer and driver ratings across categories such as Auto, Bike, Mini, Prime Sedan, Prime SUV, etc. This revealed service quality differences among vehicle types.

- Revenue by Payment Method**

A stacked bar chart showing revenue distribution by payment modes like Cash, UPI, and Card. This helped understand customer payment preferences and digital adoption trends.

- **Ride Distance Distribution Per Day**

A histogram showcasing the spread of daily travel distances. It helped identify common trip lengths and segment rides into short, medium, and long-distance categories.

- **Top Customers by Booking Value**

A leaderboard or bar chart ranking customers based on their total spending. This helped identify high-value or loyal customers contributing significant revenue.

- **Driver vs Customer Rating Comparison**

A scatter plot comparing driver and customer ratings for each ride. This helped analyze rating patterns, mutual satisfaction levels, and potential service mismatches.

💡 Key Insights Discovered

1. Higher Demand and Revenue on Weekends & Match Days

Ride bookings and revenue saw significant increases during weekends and cricket match days, indicating leisure and event-driven travel trends.

2. Driver Cancellations Were More Frequent Than Customer Cancellations

Drivers canceled more rides compared to customers, primarily due to personal issues or customer behavior. This highlights an area where operational improvements and driver incentive programs could be beneficial.

3. Major Cancellation Reasons Identified

- **Customer:**
The most common customer cancellation reason was “*Driver not moving toward pickup*”, suggesting coordination issues or delays in reaching pickup points.
- **Driver:**
The top driver cancellation reason was “*Personal / car-related issues*”, reflecting difficulties related to driver availability or vehicle readiness.

4. High Revenue Contribution from Prime Sedan & Prime SUV

Premium vehicle categories generated a larger share of revenue despite lower ride counts, indicating strong profitability and customer preference for comfort during peak times.

5. Better Customer Ratings for Bike & Mini Rides

Short-distance categories like Bike and Mini received consistently higher customer ratings, likely due to faster pickups, affordability, and convenience for quick trips.

6. VTAT & CTAT Increased During Peak Hours

Both driver arrival time (VTAT) and customer arrival time (CTAT) increased significantly during peak traffic hours.

This contributed to:

- Slower ride fulfillment
- Higher cancellation rate
- Lower customer satisfaction

Highlighting the need for better driver allocation and real-time routing optimization.

8. Business Recommendations:

Based on the insights derived from the SQL analysis and Power BI dashboard, the following actionable recommendations can help OLA enhance its operational efficiency, customer satisfaction, and overall business performance:

1. Improve Driver Allocation During Peak Hours & Weekends

Ride volume was significantly higher during peak traffic times and weekends.

To meet demand effectively:

- Deploy more drivers in high-demand zones
- Use predictive analytics to forecast surge periods
- Implement dynamic driver distribution so nearby drivers can be automatically drawn to hotspots

This ensures faster pickups, reduced cancellations, and higher ride completion rates.

2. Address Driver Cancellation Issues Through Training & Incentives

Driver cancellations were more frequent than customer cancellations and were often related to personal or car-related issues. OLA can:

- Provide targeted training to improve driver professionalism
- Offer incentives or bonuses to drivers who maintain low cancellation rates
- Implement quick vehicle inspection programs to manage car-related issues

Reducing driver cancellations will directly improve ride availability and customer trust.

3. Promote High-Performing Vehicle Types During High-Demand Hours

Vehicle categories such as **Prime Sedan** and **Prime SUV** generated the highest revenue. OLA can leverage this by:

- Offering promotional visibility for premium vehicles during peak demand
 - Providing dynamic pricing or exclusive offers to increase bookings for these categories
 - Encouraging drivers of premium vehicles to stay active during high-demand windows
- This helps increase profitability while meeting customer preferences.
-

4. Introduce Customer Communication Features to Reduce Location Confusion

A major reason for customer-side cancellations was "*Driver not moving toward pickup location.*" This leads to frustration and booking failure.

To reduce miscommunication:

- Implement enhanced live location sharing for both driver and customer
 - Add in-app quick chat and voice call options
 - Improve map accuracy with better address recognition
 - Notify customers of driver movement delays
- Clearer communication reduces confusion and improves ride reliability.
-

5. Enhance Real-Time Tracking to Reduce VTAT & CTAT Delays

Both VTAT (Vehicle Time to Arrive) and CTAT (Customer Time to Arrive) were higher during peak hours, impacting cancellations and satisfaction. OLA can:

- Use AI-based routing to identify the fastest, least congested routes
 - Offer incentives for drivers who maintain low VTAT
 - Send timely reminders to customers to reach pickup points on time
 - Highlight nearby pickup points with lower traffic to reduce CTAT
- These improvements will lead to smoother pickups and more successful ride completions.

9. Project Conclusion:

The OLA Data Analyst Project delivered a comprehensive understanding of the platform's operational performance, customer behavior, and financial outcomes. By analyzing over one lakh ride records through SQL, Excel, and Power BI, this study highlighted key patterns in booking trends, cancellation drivers, service quality, and revenue distribution.

The insights derived from the dashboard revealed clear correlations between demand peaks, driver availability, customer satisfaction, and operational bottlenecks. Patterns such as increased weekend demand, higher driver-initiated cancellations, premium vehicle revenue dominance, and elevated VTAT/CTAT during peak hours provided actionable clarity on the overall ride experience.

These findings equip OLA with data-backed strategies to:

- Reduce cancellations on both customer and driver sides
- Optimize driver allocation during high-demand periods
- Enhance communication and tracking systems
- Improve satisfaction scores across vehicle categories
- Maximize revenue through targeted promotion of high-performing services

Overall, this project demonstrates how data-driven decision-making can significantly improve the efficiency, reliability, and profitability of ride-hailing operations. With the right interventions, OLA can strengthen customer trust, streamline operations, and boost long-term growth.