

Uber Rides Data Analysis and Dashboard Documentation

• Project Proposal:

Objective: Create an engaging and interactive Power BI dashboard that provides comprehensive insights into Uber ride operations, uncovers booking patterns, and supports data-driven decision-making for service optimization.

Scope:

- **Booking Performance:** Analyze ride completion rates, cancellation patterns, and booking status distribution across different vehicle types and locations.
 - **Customer & Driver Insights:** Understand cancellation reasons, ratings distribution, and behavioral patterns for both customers and drivers.
 - **Operational Analysis:** Assess average waiting times (VTAT & CTAT), ride distances, booking values, and payment method preferences.
 - **Geographic Trends:** Compare performance metrics across different pickup and drop-off locations to identify high-demand areas and optimization opportunities.
-

• Project Plan

Phase	Tasks	Duration	Milestone
1-Data Collection	Gather booking data, customer details, driver information, and ride metrics	Week 1	Dataset finalized
2-Data Cleaning & Transformation	Handle missing values, standardize data types, impute nulls using group means/modes, and create calculated columns	Week 2	Cleaned dataset ready
3-Data Analysis & Metrics Creation	Calculate KPIs: completion rate, cancellation rates, average ratings, revenue metrics, and operational efficiency indicators	Week 3	Key insights identified
4-Data Modeling & Visualization	Build Star Schema in Power BI, create relationships, and design interactive visualizations across 4 dashboard pages	Week 4	First version of dashboard
5-Review & Refinement	Improve visuals, verify accuracy of calculations, optimize performance, and finalize report design	Week 5	Final dashboard ready
6-Presentation and Documentation	Prepare findings, document analysis methodology, present results to stakeholders	Week 6	Project completed

- **Resources:** Python (Pandas, NumPy, Matplotlib), Excel, Power BI

- **Task Assignment & Roles:**
 - Data Collection: Ahmed Osama, Fawzy Shams
 - Data Cleaning & Transformation: Anton Samy, Ahmed Refaa
 - Data Analysis and Metrics Creation: Hassan Ashraf, Essam
 - Dashboard Development: Ahmed Osama, Fawzy Shams
 - Review & Refinement: All team members
 - Presentation and Documentation: All team members
-

- **Risk Assessment & Mitigation Plan:**

- **Risk:** Data inconsistency and missing values → **Solution:** Implemented group-based imputation strategies using vehicle type, location, and customer ID groupings.
 - **Risk:** Performance issues with large dataset → **Solution:** Optimized data model with Star Schema, used DAX measures efficiently, and implemented proper indexing.
 - **Risk:** Incomplete ride data affecting analysis → **Solution:** Created proper handling for incomplete rides with reason categorization.
-

- **KPIs:**

Operational Metrics:

- Total Rides
- Completed Rides & Completion Rate
- Customer Cancellation Rate
- Driver Cancellation Rate
- Incomplete Ride Rate

Financial Metrics:

- Total Booking Value
- Average Booking Value
- Revenue by Vehicle Type
- Revenue by Payment Method

Service Quality Metrics:

- Average Driver Rating
- Average Customer Rating
- Average VTAT (Vehicle Turn Around Time)
- Average CTAT (Customer Turn Around Time)
- Average Ride Distance

Geographic Performance:

- Rides by Pickup Location
 - Rides by Drop Location
 - Revenue by Location
-

• Dataset Overview:

Column Name	Description
Date	The date when the booking was made
Time	The time when the booking was made
Booking ID	A unique identifier for each booking
Booking Status	Status of the ride (Completed, Cancelled by Customer, Cancelled by Driver, Incomplete, No Driver Found)
Customer ID	A unique identifier for each customer
Vehicle Type	Type of vehicle booked (Auto, Bike, eBike, Go Sedan, Premier Sedan)
Pickup Location	Location where the ride started
Drop Location	Destination location for the ride
Avg VTAT	Average Vehicle Turn Around Time (in minutes)
Avg CTAT	Average Customer Turn Around Time (in minutes)
Cancelled Rides by Customer	Binary indicator (0/1) if customer cancelled
Reason for cancelling by Customer	Reason provided for customer cancellation
Cancelled Rides by Driver	Binary indicator (0/1) if driver cancelled
Driver Cancellation Reason	Reason provided for driver cancellation
Incomplete Rides	Binary indicator (0/1) for incomplete rides
Incomplete Rides Reason	Reason for ride incompleteness
Booking Value	Total fare amount for the ride
Ride Distance	Distance covered during the ride (in km)
Driver Ratings	Rating given by customer to driver (1-5 scale)
Customer Rating	Rating given by driver to customer (1-5 scale)
Payment Method	Method used for payment (UPI, Debit Card, Credit Card, Cash)

2. Literature Review

- **Feedback & Evaluation:** Gathered insights from ride-sharing industry best practices and stakeholder requirements for operational dashboards.
 - **Suggested Improvements:**
 - Implemented time-series analysis for trend identification
 - Added geographic heatmaps for location-based insights
 - Created cancellation reason analysis for service improvement
 - Designed rating distribution analysis for quality monitoring
 - **Final Grading Criteria:** Based on data accuracy (cleaned 150,000 records), visualization effectiveness (4-page interactive dashboard), user interaction capabilities, and actionable storytelling through insights.
-

3. Requirements Gathering

- **Stakeholder Analysis:** Uber operations managers, fleet managers, customer service teams, and business analysts.
 - **User Stories:**
 - "As an operations manager, I want to see completion rates by vehicle type to optimize fleet allocation."
 - "As a customer service manager, I want to understand cancellation patterns to improve service quality."
 - "As a business analyst, I want to track revenue trends by location to identify growth opportunities."
 - **Functional Requirements:**
 - Real-time operational metrics with filters for date, vehicle type, and location
 - Cancellation analysis with reason categorization
 - Rating distribution visualization for quality monitoring
 - Geographic performance analysis with interactive maps
 - Payment method preference tracking
 - Support for drill-down into detailed ride-level data
 - **Non-functional Requirements:**
 - Dashboard must load within 3 seconds
 - Support for 150,000+ records with smooth interactivity
 - Mobile-responsive design for on-the-go access
 - Clear, intuitive visualizations with consistent color schemes
-

4. System Analysis & Design

• Problem Statement:

The goal is to create an automated Power BI dashboard to analyze Uber ride data at scale. This will replace manual Excel analysis, save significant time, provide real-time insights, and enable data-driven decision-making for operational improvements and strategic planning.

• Use Case Diagram:

- **Operations Manager:** Monitor KPIs, analyze trends, make fleet allocation decisions
- **Customer Service Team:** Review cancellation reasons, identify service gaps, improve customer satisfaction
- **Business Analysts:** Perform deep-dive analysis, create reports, identify revenue opportunities
- **Data Administrators:** Manage data refresh, update dashboards, maintain data quality

• Software Architecture:

Data Layer:

- Source: Cleaned CSV file (150,000 records)
- Storage: Power BI Desktop file (.pbix)
- Refresh: Scheduled refresh through Power BI Service

Modeling Layer:

- Star Schema implementation
- Fact Table: Ride Transactions
- Dimension Tables: Date, Time, Location, Vehicle, Customer, Driver
- Calculated columns and measures using DAX

Presentation Layer:

- 4-page interactive dashboard:
 1. **Overview Page:** High-level KPIs and trends
 2. **Booking Analysis Page:** Completion rates, cancellation patterns, status distribution
 3. **Performance Metrics Page:** Ratings, waiting times, ride distances, revenue analysis
 4. **Geographic Insights Page:** Location-based performance and heatmaps

• Data Flow:

1. **Extraction:** Raw data collected from Uber booking system
2. **Transformation (Python - Colab):**
 - Handled missing values using group-based imputation
 - Standardized data types

- Created calculated fields (e.g., completion indicators)
- Removed duplicates and outliers
- 3. **Loading:** Cleaned data exported to CSV and Excel
- 4. **Modeling (Power BI):**
 - Imported cleaned data
 - Created Star Schema relationships
 - Built DAX measures for KPIs
- 5. **Visualization:** Designed 4-page interactive dashboard with filters and drill-through capabilities

• ER Diagram & Data Modeling:

Star Schema Design:

Fact Table: Rides_Fact

- Booking ID (PK)
- Date Key (FK)
- Time Key (FK)
- Customer Key (FK)
- Vehicle Key (FK)
- Pickup Location Key (FK)
- Drop Location Key (FK)
- Booking Status
- Booking Value
- Ride Distance
- Driver Rating
- Customer Rating
- VTAT
- CTAT
- Payment Method

Dimension Tables:

1. **Dim_Date**
 - Date Key (PK)
 - Date
 - Day
 - Month
 - Year
 - Quarter
 - Day of Week
2. **Dim_Time**
 - Time Key (PK)
 - Time
 - Hour

- Minute
- Time Period (Morning/Afternoon/Evening/Night)
- 3. **Dim_Customer**
 - Customer Key (PK)
 - Customer ID
 - Customer Segment
- 4. **Dim_Vehicle**
 - Vehicle Key (PK)
 - Vehicle Type
 - Vehicle Category
- 5. **Dim_Location**
 - Location Key (PK)
 - Location Name
 - City
 - Region
- 6. **Dim_Booking_Status**
 - Status Key (PK)
 - Status Name
 - Status Category

• UI/UX Design & Prototyping

Dashboard Pages:

Page 1: Executive Overview

- KPI Cards: Total Rides, Completion Rate, Avg Rating, Total Revenue
- Line Chart: Daily ride trends
- Donut Chart: Booking status distribution
- Bar Chart: Top 5 pickup locations

Page 2: Booking Analysis

- Stacked Bar Chart: Booking status by vehicle type
- Pie Charts: Customer cancellation reasons, Driver cancellation reasons
- Table: Incomplete ride details with reasons
- Waterfall Chart: Status breakdown flow

Page 3: Performance Metrics

- Gauge Charts: Driver & Customer ratings
- Line & Column Chart: VTAT & CTAT trends
- Scatter Plot: Ride distance vs booking value
- Clustered Bar Chart: Revenue by payment method

Page 4: Geographic Insights

- Map Visual: Pickup/drop locations with ride volume
 - Heat Map: Revenue by location
 - Matrix: Location-to-location ride frequency
 - Top N analysis: Busiest routes
-

• 5. Deployment System & Integration

- Technology Stack: Power BI for dashboard creation. CSV for data storage. Power BI Service for sharing.
 - Deployment Diagram: Shows how data moves from Excel/CSV to Power BI, and how users access the dashboard.
 - Component Diagram: Shows how Power BI and the data source work together.
-

• 6. Additional Deliverables

- API Documentation: Explains how to connect data if APIs are used.
- Testing & Validation: Ensures the data is correct and the dashboard works.
- Deployment Strategy: How the dashboard will be published and accessed by users.