Capstone Project 1 – Uber Data Analysis

Submitted by:

Nayini Rithwik, PGA46

Table of Contents

Problem Statements:

Uber, as a ride-hailing service, faces challenges in optimizing its operations and improving customer satisfaction. The dataset provided reflects real-world issues such as:

1. High-demand areas – Identifying which pickup locations experience the highest ride demand.
2. Vehicle type efficiency – Understanding which vehicle types perform better in terms of usage and distance.
3. Revenue optimization – Determining which payment methods and routes contribute the most revenue.
4. Customer behavior – Analyzing frequently traveled routes and ride patterns.
5. Operational challenges – Evaluating booking statuses (completed vs. canceled rides) to detect inefficiencies.
6. Time-based trends – Monitoring daily revenue fluctuations to support seasonal demand forecasting.

Project Summary:

This project focuses on analyzing Uber ride data to uncover patterns in demand, revenue, and customer behavior. Using SQL queries, the data was explored to extract ride counts, revenue distribution, and route analysis. A Jupyter Notebook (Python) was used for data cleaning, EDA, and visualization, while Power BI dashboards provided interactive exploration of the dataset.

Key findings highlight that certain pickup locations dominate demand, digital payments drive revenue, and specific routes account for most rides. Operational inefficiencies such as cancellations were also identified.

The project concludes with actionable recommendations: optimizing driver allocation, promoting digital payments, reducing cancellations, and preparing for seasonal peaks. Overall, this integrated approach demonstrates how data-driven insights can help Uber improve efficiency, maximize revenue, and enhance customer satisfaction.

INTRODUCTION:

The objective of this project is to analyze Uber ride data to gain insights into ride patterns, customer behavior, and revenue generation. The analysis integrates SQL queries, Python-based Jupyter Notebook exploration, and Power BI dashboards.

DATA SOURCE:

The dataset `Uber_data` contains records of Uber rides, including Pickup & Drop Locations, Vehicle Type, Ride Distance, Booking Value, Payment Method, Booking Status, and Date. This dataset enabled analysis of ride frequency, revenue, and operational trends.

SQL ANALYSIS:

SQL queries were used for descriptive analytics on the Uber dataset:

1. Total Rides – Counted the total number of rides.

2. Top Pickup Locations – Ranked locations by ride count.

3. Vehicle Type Performance – Analyzed ride distribution and average distances by vehicle type.

4. Payment Methods & Revenue – Summarized booking value across payment modes.

5. Most Frequent Routes – Identified top pickup–drop combinations.

6. Booking Status Distribution – Counted rides by completion/cancellation.

7. Daily Revenue Trends – Aggregated booking values per date for revenue analysis.

PYTHON (JUPYTER NOTEBOOK) ANALYSIS:

The Jupyter Notebook was used for deeper analytics:

- Data Cleaning: Handling missing values in ride distance and booking value.

- Exploratory Data Analysis (EDA): Distribution plots for ride distance, booking value, and ride frequency.

- Visualization: Graphs highlighting top pickup locations, routes, revenue per payment method, and ride trends.

- Insights: Confirmed SQL results and highlighted customer demand fluctuations.

POWER BI DASHBOARD:

The Power BI dashboard provided interactive analytics:

- KPI Cards – Total rides, revenue, and average ride distance.

- Maps – Geographical distribution of pickups and drops.

- Bar/Column Charts – Rides by vehicle type, payment method revenue share.

- Trend Lines – Daily and monthly revenue trends.

- Filters/Slicers – For date, vehicle type, and payment method.

FINDINGS & INSIGHTS:

Key insights include:

- Certain pickup locations dominate demand.

- Vehicle types vary in usage patterns.

- Digital payments contribute the highest revenue.

- A few pickup–drop routes account for a majority of rides.

- Revenue trends fluctuate daily with seasonal patterns.

- Booking status analysis revealed cancellation issues.

RECOMMENDATIONS:

- Optimize driver allocation in high-demand areas.

- Promote digital payment discounts to increase adoption.

- Offer route-based pricing/offers for popular routes.

- Monitor cancellations to reduce inefficiencies.

- Plan resources for seasonal peaks in demand.

CONCLUSION:

This capstone project integrated SQL, Python, and Power BI to perform end-to-end analysis of Uber ride data. The insights derived can directly support operational efficiency, customer experience improvements, and revenue growth strategies.

REFERENCES:

1. Kaggle. *Uber Data Analysis Datasets & Notebooks*. https://www.kaggle.com.
2. Uber. *Company Information and Ride-Hailing Data Insights*. https://www.uber.com.