

Uber- For Hire vehicles in NYC

Introduction:

In a busy city like New York, need for frequent travel arrangement is essential for daily activities carried out by people living within the city. Taxi and Limousine Commission(TLC) are responsible for managing all taxi movement to ensure that they operate effectively. Uber is a major stakeholder in such an arrangement by ensuring that over a day's time this kind of transport handles more than ten trips(CITE). If the quality of services is to improve then managers must comprehend operational characteristics of Uber as well as customer behavior or service efficiency such are some areas that need deeper understanding. The project is for the in-depth investigation of Uber's operational data. The information was collected from Kaggle and the data comprises 17.45 crore entries, spread out over 24 columns that cut across pickup time and drop time, trip miles' and durations. In our dataset there are 12 files in parquet format with each file corresponding to data from a specific month. Each file is exactly 300 MB in size, resulting in a total file size of around 3.6 GB.

Approach:

Initially, we imported the data from each individual file. Then, we merged them into a single DataFrame using the 'union' operation. The main focus of this project will be doing data processing using PySpark with the help of Databricks software, throughout which these steps are done:

Data Cleaning: The dataset was rigorously cleaned using PySpark. This involved filtering out incomplete records, correcting data entry errors, and dealing with anomalies in trip times and distances that could skew analysis results.

Data Visualization: Created insightful visual representations of data to identify patterns and anomalies.

Exploratory Data Analysis (EDA):

The EDA focused on several key aspects:

Ride Request Patterns: Analysis of the frequency and distribution of ride requests throughout the day to determine peak hours.

Trip Characteristics: Average trip distances and times were calculated to understand typical journey profiles.

High-demand Zones: Areas with the most frequent pickup requests were identified, suggesting where Uber might need to increase vehicle availability.

Wheelchair Accessibility: The demand for wheelchair-accessible vehicles was specifically analyzed to assess how well Uber is serving customers with mobility challenges.

Technical Environment:

This analysis was conducted in Databricks, utilizing the PySpark framework to handle and process large-scale data efficiently. Databricks provided a robust, scalable cloud platform ideal for performing intensive data manipulations and visualizations required by this project.

For a thorough execution of the analysis, it is required to access the Databricks community edition platform by logging into the provided account.

Databricks credentials :

ID: khanderaosharanya033@gmail.com

Password: Shar@2303

Link to project access:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2228397202891690/93080512763204/6938215192770554/latest.html>

Conclusions:

The project delivered comprehensive insights into Uber's operational challenges and opportunities in NYC.

Key findings include:

Strategic Vehicle Deployment: Data-driven strategies were developed for positioning vehicles in high-demand areas during peak times to reduce customer wait times and optimize trip efficiency.

Service Improvement Opportunities: Identification of frequent delays in certain zones led to recommendations for operational adjustments to enhance service reliability.

Enhanced Customer Service: Analysis of service delays and accessibility options provided a foundation for improving customer satisfaction, particularly for disabled passengers.

Outcomes:

This analysis provides Uber with important outcomes to help it plan strategically, giving specific suggestions on how to improve operations, make customers happy and meet all regulations. Through the project's findings, it is hoped that Uber will tweak its services and be able to satisfy varied customers in NYC.