MICRO PROJECT - UBER DATA ANALYSIS

Abhijeet J Kartha (CS2 02)
Dilshan Dileep Shyni (CS1 21)
Shawn Siby George (CS2 52)
Reuben Roy Vargis (CS1 51)

Abstract:
This micro project presents a comprehensive analysis of Uber data in Boston, MA, USA during the months of November and December 2018. The dataset used for this analysis consists of a large collection of over 630,000 data points, including records from both Uber and Lyft, two prominent ride-sharing companies.

The project aimed to gain insights into the Uber data through various data analysis techniques. To ensure the accuracy and reliability of the analysis, several preprocessing steps were performed, including data cleaning, transformation, and handling missing values. Additionally, label encoding was utilized to convert categorical variables into numerical format, enabling the subsequent regression models to effectively process the data.

Python programming language and the Jupyter Notebook environment were employed for the analysis. The project made use of the rich resources available in the Python ecosystem for data analysis and visualization. The dataset utilized in this project was obtained from Kaggle.

After preprocessing the data, two regression models, namely Linear Regression and K-Nearest Neighbors (KNN), were implemented to predict prices based on the available features. The performance of each model was evaluated using appropriate scoring metrics to assess their accuracy and effectiveness in price prediction.

The findings and visualizations obtained from the analysis offer a deeper understanding of Uber data patterns and trends in Boston, MA. This project serves as a valuable example of using Python, Jupyter Notebook, and popular regression models for analyzing and predicting prices in the ride-sharing industry.