

Project Proposal

Uber and Lyft Price Prediction (BOSTON)

Group Primrose

Amol Bohora

Ayush Shah

Sumit Pagar

Abstract

We found a dataset which has price data for rideshare apps (Uber & Lyft) which is approximately 5 years old. Our plan with this project is trying to find which model is most accurate for predicting rideshare prices and find most imp factors contributing towards the prices.

Data Source

The dataset we are referring to is hosted on Kaggle, linked [here](#).

About Data

The dataset includes information about both Lyft and Uber services.

It records numerous aspects of each ride, including the distance traveled, the source and destination locations, and the money paid for the journey, the pricing.

Each ride is timestamped, and the dataset tracks surge pricing via the “surge_multiplier” column, which indicates if the ride's price was subject to surge pricing.

Each entry has an id and includes the precise cab product type in both the “product_id” and name fields (for example, "Shared", "Lyft XL").

Methodology

We will start by cleaning the data and creating basic visualizations to uncover initial insights. This will enable us to identify trends and patterns within the dataset. Next, we will apply various algorithms, including Linear Regression, to train the data. By comparing the accuracies of these algorithms, we aim to determine which one is most suitable for this dataset. This methodical approach will deepen our understanding and refine our modeling efforts.

Future Scope

- **Geospatial Analytics:**
 - Utilizing geospatial data (source and destination) to predict ride durations, fares, or traffic hotspots. This could involve clustering certain zones by demand and offering tailored services for high-demand areas.
- **Customer Segmentation:**
 - Grouping customers based on their behavior (e.g., frequent vs. occasional riders, luxury vs. budget users) can help companies personalize marketing efforts, offer loyalty rewards, or target promotions.
- **Ride-Sharing Optimization:**
 - Investigating ride-sharing patterns (for services like Lyft Line or UberPool) to optimize shared ride routes and reduce the total time spent or distance covered while still maximizing profit.