**Uber vs Lyft Fare Comparison and Analysis**

**Group 5:** Georgina Young, Revano Harahap, Dominique Kelsey, Val Sanhueza  
**Dataset:** [Uber and Lyft Dataset – Boston, MA (Kaggle)](https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma)

**Overview**

This project explores ride-sharing behaviors in Boston, MA, by comparing Uber and Lyft services using a public dataset from Kaggle. We aim to understand fare pricing differences, trip duration trends, and temporal demand patterns for both services. Our analysis employs both Tableau for interactive visualizations and predictive modeling techniques in Python.

**Research Questions**

1. **How do Uber and Lyft fares compare across different regions?**

* We used bar charts to display the average fare across Boston neighborhoods, segmented by ride-sharing service.
* The map visualization reveals geographical trends in pricing hotspots for both companies.

1. **What are the average trip durations for Uber vs Lyft, and how do they vary by time of day?**

* Line charts track average trip durations across each hour of the day.

1. **Do Uber or Lyft have more demand during specific hours (e.g., rush hour)?**

* A heat map visualizes the volume of rides per hour.
* We observed noticeable peaks during early-morning commutes and late-night periods. Uber held more demand than Lyft.

**Tableau Dashboards**

We developed two Tableau stories to visualize the dataset and address our core research questions.

**Story 1: Fare Comparison and Regional Analysis**

* Dashboard reflects the bar chart and map interactively. Featuring a dynamic filter to pinpoint what service had higher usage.
* Bar chart comparing average Uber and Lyft fares by region.
* Heat map highlighting demand levels by hour, giving insight into peak and off-peak usage.
* A leaderboard-style table ranks the highest and lowest fare regions.
* Scatterplot to drive home the prices of Lyft were higher than Uber and the direct correlation of pricing vs. distance.

**Story 2: Trip Duration and Time-of-Day Analysis**

* Dashboard reflects the line chart, map, and bar chart of services by style.
* Line chart showing the change in average trip duration throughout the day.
* Interactive map of Boston displaying fare distribution by service.

Both stories utilize a consistent, clean color theme for clarity and professional presentation.

**Predictive Modeling**

 Objective: Predict the price of a ride based on features like distance, surge prices, sources and destinations, and ride type — and determine which service (Uber or Lyft) is cheaper.

 Initial Approach:

* Started with basic regression models (e.g., linear regression, decision tree regressor).
* Results showed low R² scores, indicating poor performance.

 Second Attempt:

* Reframed the problem as a classification task using a Random Forest Classifier.
* Created a new column to label which service was cheaper.
* Achieved decent metrics (accuracy, precision, recall), but the approach didn’t match the real-world use case well.

 Final Model, with Correct Logic:

* Tested additional regressors: Ridge, Lasso, Random Forest Regressor.
* Random Forest Regressor delivered the highest R² scores and best overall performance.

 Pipeline:

* Built a preprocessing + modeling pipeline to handle input features and predictions.
* Exported the final pipeline as a .pkl file using pickle for use in the web app.

**GitHub Repository Structure**

***To be completed as a team***

**Team Roles & Contributions**

| **Team Member** | **Primary Role** | **Secondary Role** |
| --- | --- | --- |
| Revano Harahap | Data Cleaning & Preprocessing | Model Assistance |
| Val Sanhueza | Predictive Modeling | Data Cleaning Support |
| Georgina Young | Tableau Visualizations (Fares & Duration) | Dashboard Integration |
| Dominique Kelsey | Tableau Visualizations (Demand Analysis) | Dashboard Integration |