

Group 5: Proposal for Uber vs Lyft Fare Comparison and Analysis

By: Georgina Young, Revano Harahap, Dominique Kelsey, and Val Sanhuesa

1. Dataset and Why We Chose This Topic:

We will be working with the *Uber and Lyft Dataset from Boston, MA*, available on Kaggle: [Uber VS Lyft - Kaggle](#).

This dataset contains detailed information about Uber and Lyft rides, including fare amounts, trip durations, and geographical data (latitude and longitude) for the Boston area. We have chosen this dataset because it offers valuable insights into two of the most popular ride-sharing services in a central metropolitan area. We all use these services regularly and are interested in how the data can reveal patterns or trends in ride fares, demand, and other factors. We also want to explore how ride-sharing behaviors change by time of day, region, and service.

2. Research Questions:

Here are the 3 guided research questions we intend to explore:

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

- **Goal:** Understand whether there is a significant difference in fares between Uber and Lyft across different areas of Boston.
- **Visualizations:** A bar chart to show average fare comparisons per region, possibly broken down by service (Uber vs Lyft).

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

- **Goal:** Compare the average trip durations for Uber and Lyft and see how they vary by different times of day (e.g., morning, afternoon, evening).
- **Visualizations:** A line chart showing the change in average trip duration across different hours of the day for each service.

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

- **Goal:** Examine whether there is a peak demand for Uber or Lyft during rush hour or other specific times.
- **Visualizations:** A heat map or line chart illustrating the demand (number of rides) for both Uber and Lyft over time.

1. **MAP** – geographical location between Uber and Lyft

3. Data Exploration and Inspiration:

- **Inspiration and References:**

- We will draw inspiration from existing visualizations on Kaggle and Tableau Public. We will review other relevant visualizations highlighting fare comparisons and time-of-day trends in transportation data.
- Specifically, we will look at the [*“Uber/Lyft Fare Prediction Model”*](#) on Kaggle for insights into fare modeling.
- Tableau Public will inspire designing clean and interactive dashboards.

- **Tableau Visualizations:**

- As part of our analysis, we will create multiple visualizations in Tableau to address our research questions, such as:
 - Bar charts, line charts, heat maps, and a map to visualize demand by geographical location (using latitude and longitude).
 - Dashboards with a clean, color-coordinated theme using a palette from [Palette Selection](#).

4. Tableau Dashboards/Stories:

We will create two Tableau dashboards to address the following:

1. Dashboard 1:

- Title: **Fare Comparison and Regional Analysis**
- **Visualizations:**
 - A bar chart comparing average fares for Uber and Lyft by region.
 - A map showing the geographic distribution of Uber vs Lyft rides in different neighborhoods.
 - Possibly a table or leaderboard showing the top regions with the highest/lowest fares.

2. Dashboard 2:

- Title: **Trip Duration and Time-of-Day Analysis**
- **Visualizations:**
 - A line chart displaying average trip durations for Uber and Lyft across different times of day.

- A heat map showing the number of trips at different times of the day (peak vs off-peak hours).

5. Predictive Models:

Based on the data available, we will implement a regression model to predict the following:

- **Prediction Task:**
We will try to predict the **price** based on features such as trip duration, region, and time of day.
- **Model Type:**
We will evaluate regression models (e.g., linear regression and decision trees) to predict the continuous price.
- **Target Column:**
Our target column will likely be **price** (continuous).
- **Future Work:**
We could also explore building a K-Nearest Neighbors (KNN) model for recommendation purposes, predicting which ride-sharing service a user might prefer based on historical data.

6. Roles & Responsibilities:

Team Member 1:

- **Primary Role: Data Cleaning and Preprocessing**
- **Secondary Role: Assist with Predictive Modeling**

Team Member 2:

- **Primary Role: Predictive Modeling**
- **Secondary Role: Assist with Data Cleaning**

Team Member 3:

- **Primary Role: Tableau Visualizations (Research Questions 1 & 2)**
 - Lead the development of visualizations for the first two research questions (Uber vs Lyft fares and trip durations).
 - Create bar charts, line charts, and heat maps for the visual analysis.
- **Secondary Role: Assist with Dashboard Integration**

Team Member 4:

- **Primary Role: Tableau Visualizations (Research Question 3)**
 - Lead the development of visualizations for the third research question (demand patterns by hour).
 - Focus on heat maps and demand-related visualizations.
- **Secondary Role: Assist with Dashboard Integration**

7. Timeline:

Day 1: Today (Monday, April 1st)

- **Primary Focus: Data Cleaning and Preprocessing**
 - **Team Member 1:** Clean the dataset, handle missing values and format data.
 - **Team Member 2:** Support data cleaning and begin defining key features for the analysis.
 - **All Team Members:** Start familiarizing yourselves with the research questions.
- **Tableau Visualizations (Research Question 1):**
 - **Team Members 3 & 4:** Start building basic visualizations for Research Question 1 (Uber vs Lyft fares across different regions).

Day 2: Thursday, April 4th

- **Primary Focus: Continue Building Tableau Visualizations**
 - **Team Members 3 & 4:** Continue refining the visualization for Research Question 1 and start visualization for Research Question 2 (Trip durations for Uber vs Lyft).
- **Predictive Modeling:**
 - **Team Members 1 & 2:** Begin the predictive modeling work (regression models for fare prediction).
 - **Data Integration:** Start working on integrating the predictive models into the Tableau dashboard.

Day 3: Monday, April 8th

- **Complete Tableau Visualizations:**
 - **Team Members 3 & 4:** Finalize visualizations for Research Question 1 and Research Question 2. Begin visualizations for Research Question 3 (Demand patterns for Uber vs Lyft).
- **Predictive Modeling:**
 - **Team Members 1 & 2:** Continue working on the predictive modeling and refining.
- **PowerPoint Presentation Start**
 - **Team Member 3:** Starts working on the presentation and gathers input from all team members for their sections.
- **README file and Written Report**
 - **Team Member 2:** Begins draft of both.

Day 4: Tuesday, April 9th

- **Final Refinements:**
 - **All Team Members:** Continue working to refine the Tableau dashboard and integrate all visualizations for the 3 research questions.
- **Final Model Testing:**
 - **Team Members 1 & 2:** Finalize the regression models for fare prediction, ensuring they work correctly within Tableau.
- **Dashboard Integration:**
 - **Team Members 3 & 4:** Visualizations are well-integrated and cohesive in the final Tableau dashboard.
- **PowerPoint Presentation**
 - Finalize the presentation and make any adjustments based on team feedback.
 - Do a run-through of the presentation, practicing the timing and flow of each section.

Day 5: Wednesday, April 10th (*We can try to do this Tuesday too*)

- **Final Review & Presentation Prep:**
 - **All Team Members:** Review the entire project—visualizations, model integration, data accuracy, and overall coherence.

- Prepare for the final presentation.
- Make any last-minute refinements to ensure everything runs smoothly.

8. Deliverables:

- **PDF Report:** Comprehensive write-up, including research questions, visualizations, and insights.
- **Two Tableau Dashboards:** One focused on fare comparison by region, and another on trip duration and time-of-day analysis.
- **GitHub Link:** We will share our code and dataset on GitHub for reproducibility.

Link to GitHub Repository:

[Group 5 GitHub](#)