

CSCE 4205/5215: Machine Learning Project Proposal

1. Project Title:

Analyzing the Impact of Weather Conditions on Uber and Lyft Ride Services.

1.1 Team:

Group Number	12
Team Member	Member ID
Poojitha Gorantla	11597766
Sijo Rejigeorge	11708233
Kishan Kumar Zalavadia	11685261
Valapadasu Uday Bhaskar	11696364

2. Goals and Objectives:

2.1 Motivation:

The weather has a huge impact on people's preference to travel around the city and the availability and reliability of the various transportation systems. In this study we will narrow down the entire transportation system to just ride-sharing apps in and around Boston, MA and how weather could impact the computer's decision-making process on booking such services as well as their availability.

By closely examining the data we hope to understand the influence of weather on various aspects, including ride durations, the surge in commuters booking Uber/Lyft rides, the difference in fair charges, hotspot pickup/drop-off locations and other related issues faced by commuters on ride-sharing apps during adverse weather conditions. We aim to drive insights from the data and the questions raised during our analysis and use this knowledge to predict the availability of rides.

2.2 Significance:

The importance of this project lies in its ability to enhance the ride-sharing user experience in adverse weather conditions. Users can choose services that reduce expenses and promote safer travel by knowing which service is better suited for different weather conditions. This in turn would enhance the accessibility and overall dependability of ride-sharing services.

2.3 Objectives:

2.3.1 Data Collection:

We will gather data from both Uber and Lyft that covers a wider range of weather conditions over a specified timeframe.

Our primary focus will be on extracting critical information, such as temperature and weather reports, which will serve as the foundation for our analysis.

2.3.2 Data analysis:

Data Preprocessing:

- Removing outliers.
- Handling missing values, and duplicate records.
- Classification: Identifying days with poor weather conditions to analyze.

Data Visualization: Using Tableau to generate graphs and visualization reports.

2.3.3 Implementation:

We will apply K-Nearest Neighbors (KNN) algorithm for analysis, aiming to predict the availability of rides during similar weather conditions.

2.3.4 Data Analytics:

We will examine the dataset to get a thorough understanding on the various impacts weather has on commuter's decision-making process on booking through ride sharing apps.

We aim to go through various key aspects that could influence the overall ride experience of a commuter including,

- **Ride duration** - We will analyze how adverse weather conditions may negatively impact the overall ride duration compared to mild weather and how various weather conditions can lead to variations in the ride duration.
- **Surge hours** - We will analyze the data to determine whether there is a surge in commuters seeking ride sharing apps during extreme weather conditions. Additionally, we will compare this data to check the difference in mild weather conditions, with the aim of identifying various patterns of increased demand.
- **Fair charges** - We will also investigate the difference or variances in fair charges during extreme weather conditions.
- **Hotspot pickup and drop-off location** - We will examine whether weather conditions influence the choice of pickup and drop-off locations, identify any hotspot and such.

Apart from the above-mentioned aspects, we also hope to find other correlations between weather and Uber and Lyft rides. We intend to use the gained knowledge to develop predictive models that can forecast the availability of rides in similar weather conditions.

2.4 Features:

- Uber and Lyft datasets containing relevant trip details, pricing, and weather data.
- Implementation of data preprocessing techniques to remove outliers and ensure data quality.
- Utilization of data visualization tools and libraries, such as Matplotlib, Seaborn, or Tableau, for creating visualizations.
- Development of user-friendly report or dashboard for presenting the data visually.

3. Reference:

Dataset: <https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma>