Tudor Bulf (S5618258) 34%
Joel Boer (S5542332) 33%
Tomas de Rijk (S5596483) 33%

Intro to Machine Learning
Project proposal

Task

Testing the influence of the weather on the prices of ride-hailing services like Uber and Lyft. We aim to provide a machine-learning model that can predict these prices based on weather information.

Dataset:

We will use the publicly available Kaggle dataset titled *Uber and Lyft Cab Prices*. This dataset includes features such as:

- **Categorical:** Cab type (Uber/Lyft), source, destination, ride type (e.g., economy, luxury).
- **Numerical:** Price, distance, time (hour of the day), and weather conditions (temperature, precipitation).

We won't be using every column in this dataset, for example we will not take into consideration the source and destination of the ride since we think that will not be necessary for our purposes.

Models:

We are not completely sure what models we would want to use for this task but we have a few options in mind:

- Linear Regression
- Decision tree regressor
- Baysian linear regression

Short Description of the Project:

1. **Exploratory Data Analysis:**
    - Handle missing values, encode categorical variables.
2. **Model Training and Evaluation:**
    - Train baseline models to establish reference performance.
    - Evaluate the model by splitting the dataset into training, testing and validation sets.
    - Evaluate the model using different metrics such as Mean squared error(MSE) and Root mean squared error (RMSE)
3. **(optional) Additional Evaluation Tasks:**
    - Analyze how performance varies with training set size (sample complexity curves).
    - Introduce label noise to assess model robustness.

We're open to any suggestions or alterations to the project.