

Uber Fare Predictor

Brian Plattus

01

Overview

02

Results

03

Next Steps

Overview

W

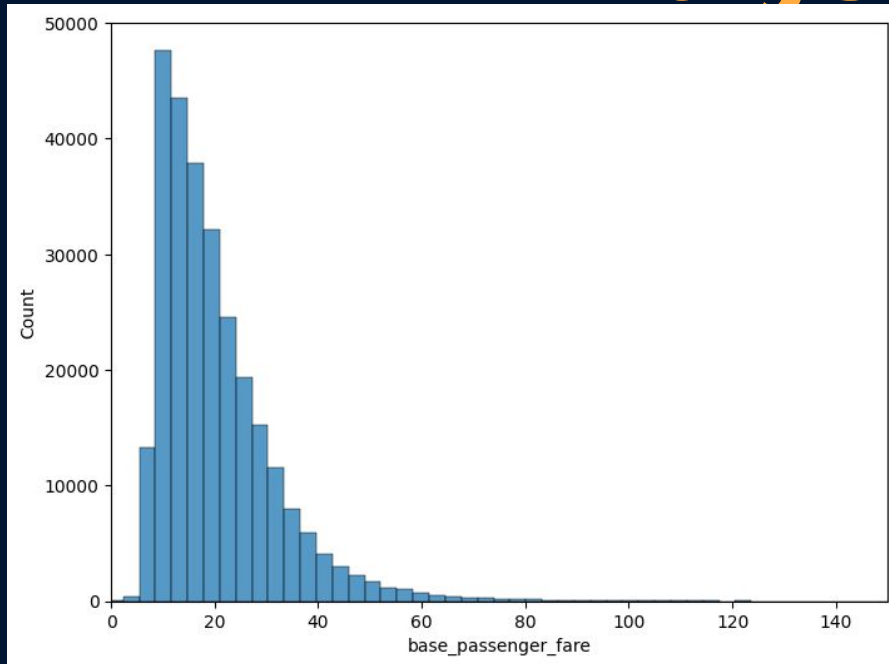
The Uber logo, consisting of the word "Uber" in a bold, black, sans-serif font, is centered within a white rectangular box.

Goal: Use NYC Taxi and Limousine Commission (TLC) data to predict future Uber passenger fares in Manhattan

Main Features:

- Trip Mileage
- Total Trip Time
- Pickup and Dropoff Location
- Pickup and Dropoff Time

Data Analysis

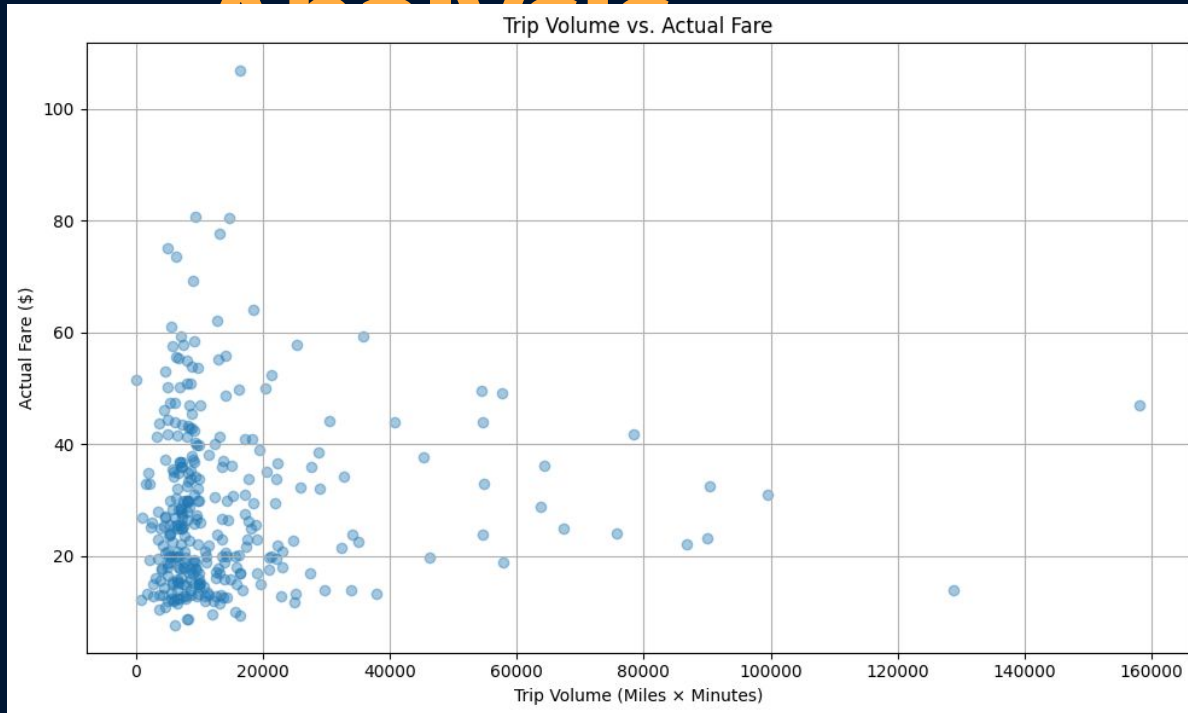


The majority of the data falls between \$10 and \$30

Skewed heavily right as there are fares as high as \$120

Data

Analysis



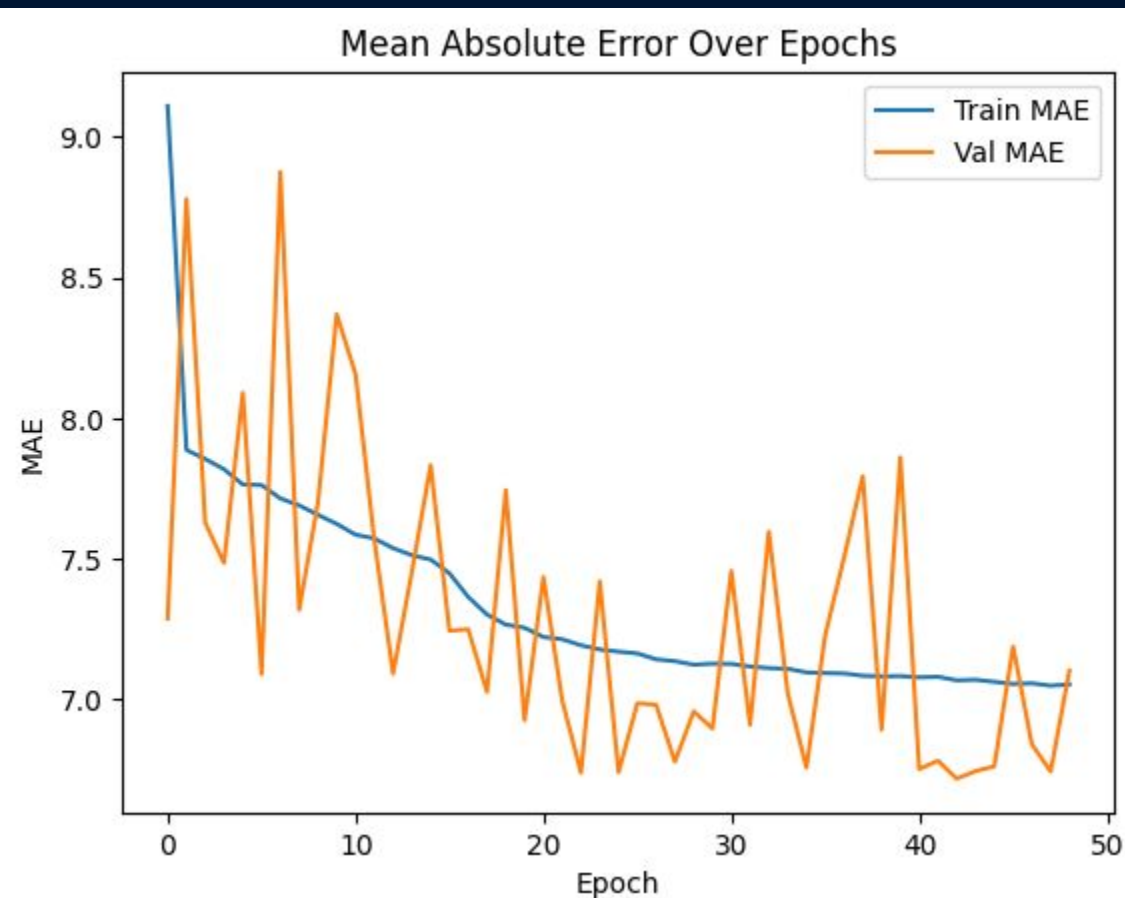
Model Selection

I decided to use a neural network due to extremely large amount of data available

Including these features:

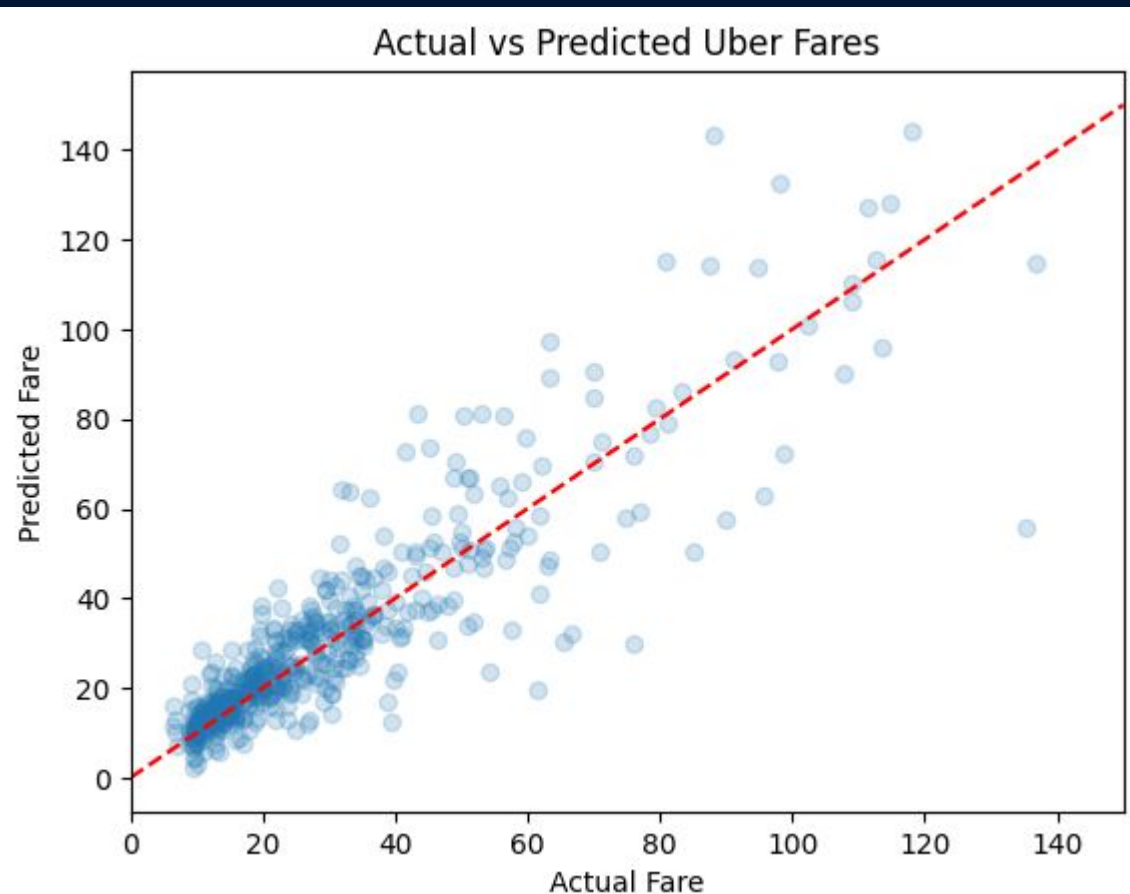
- **1** hidden layer to keep the baseline model simple
- **Linear** Activation since I'm trying to predict a continuous value
- **15%** test size due to the large amount of rows

Model Performance



- The smoothness of the Training MAE indicates the model is learning well
- The Validation MAE spiking could indicate that its not generalizing well
- Would expect the training MAE to drop significantly when properly tuned

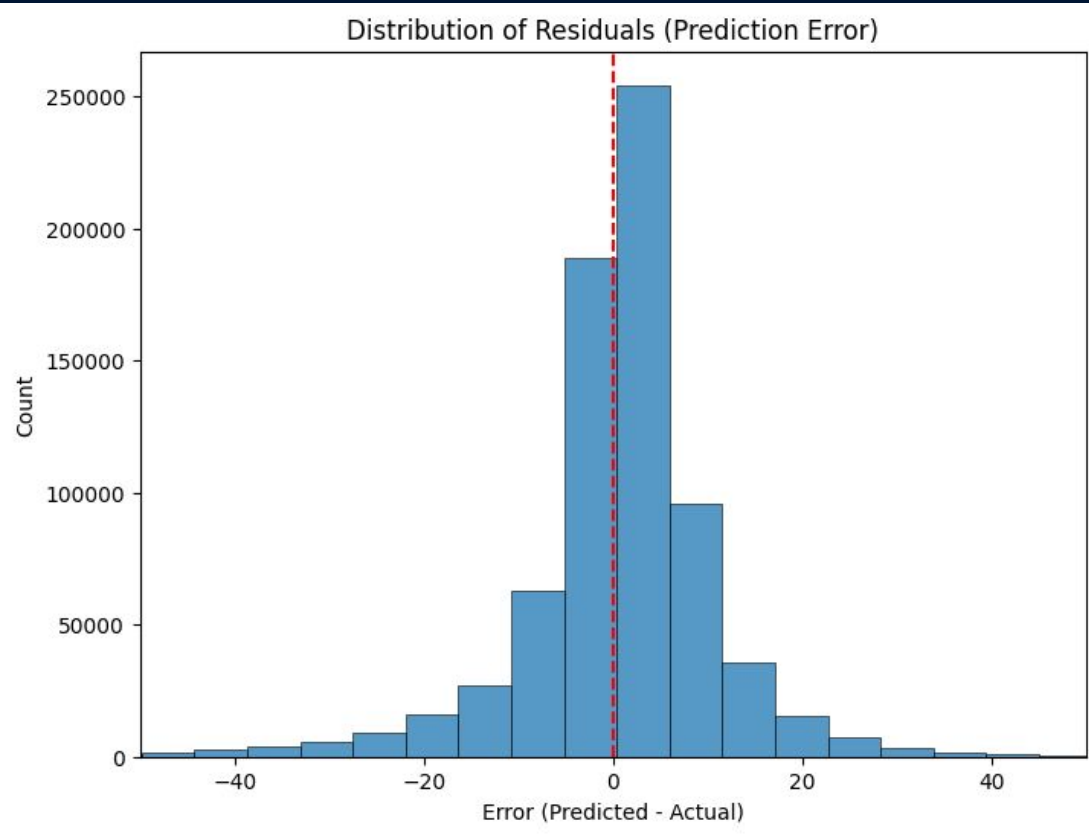
Results



- Easier to predict lower fare trips as accuracy starts to drop after 40 dollar fares
- Same predicted error between minimum fare and 40

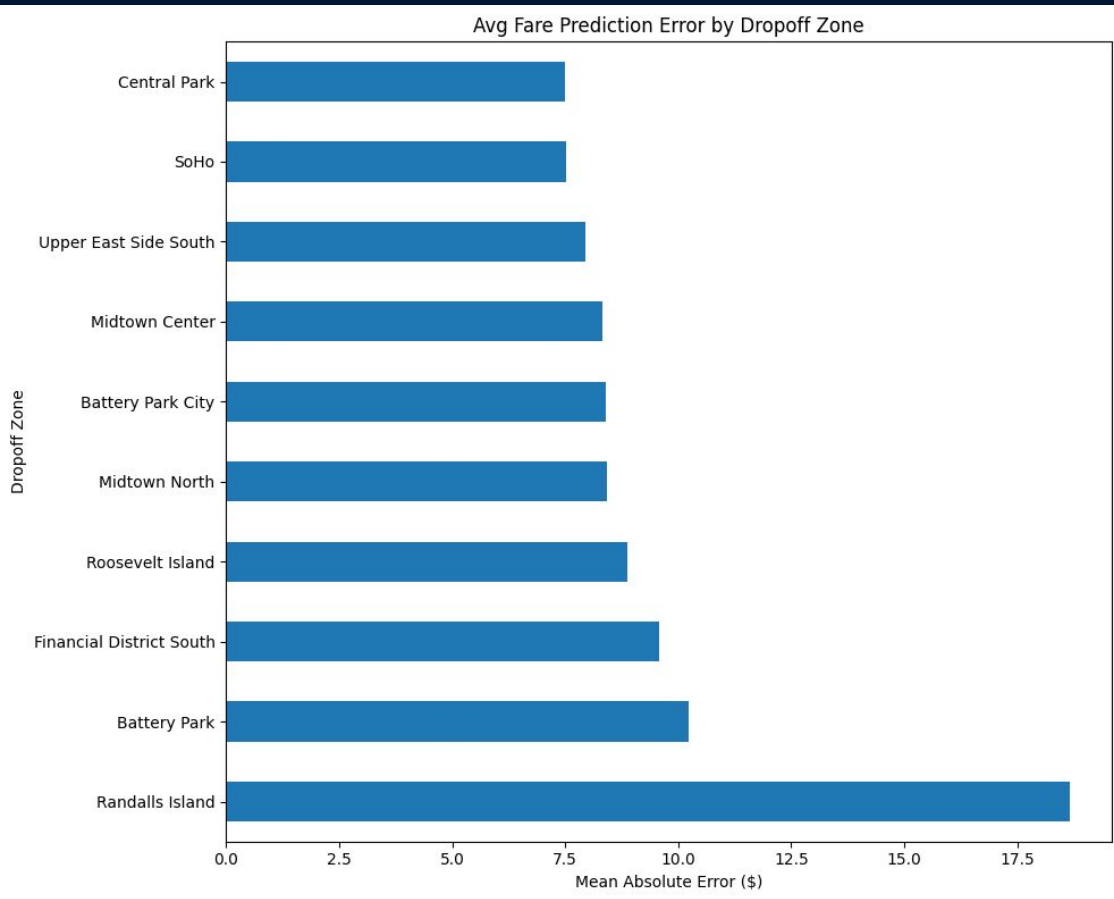
Results

graph



- Overall the model is over predicting the fare for an average trip
- The relatively high number of high error residuals may be due to data input errors

Results



- Randalls Island is almost three times the average MAE of \$6.8
- Parts of Midtown and the Financial District are the hardest to predict

Next Steps

- 1) Continue to clean data to exclude non-realistic trips
- 2) Improve model by tuning hyperparameters (learning rate and number of layers
- 3) Replace Pickup and Dropoff Zones with geographical coordinates
- 4) Use older Uber dataset that include trip coordinates to estimate trip paths for this data

Thanks
