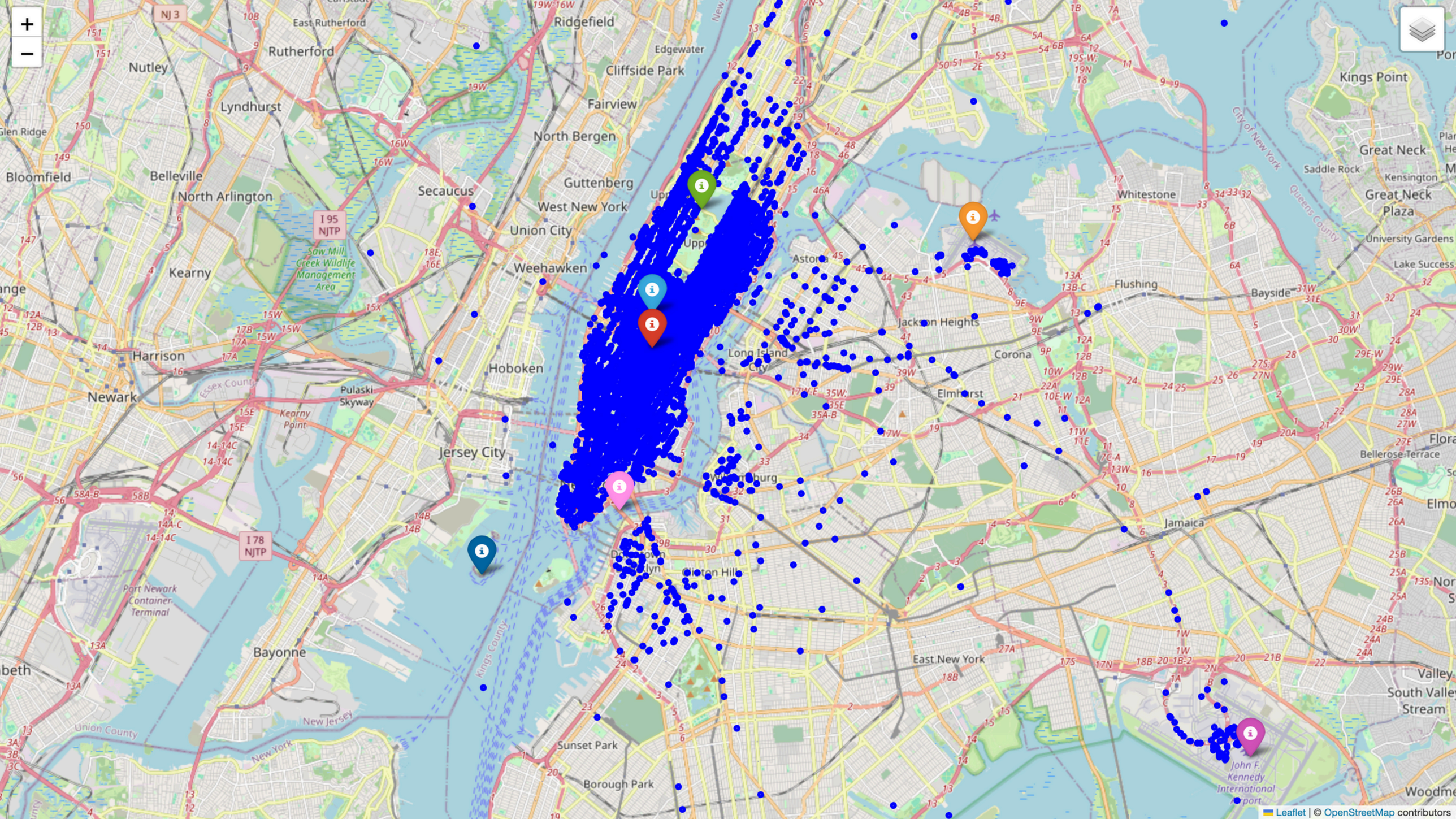

Ridgefield
East Rutherford
Rutherford
Nutley
Lyndhurst
Edgewater
Cliffside Park
Fairview
North Bergen
Guttenberg
West New York
Union City
Weehawken
Secaucus
Bloomfield
Belleville
North Arlington
Kearny
Harrison
Newark
Hoboken
Jersey City
Pulaski Skyway
Kearny Point
Port Newark Container Terminal
Bayonne
Sunset Park
Borough Park
Clinton Hill
Long Island City
Astoria
Jackson Heights
Corona
Elmhurst
Flushing
Whitestone
Bayside
Jamaica
East New York
Kings Point
Great Neck
Great Neck Plaza
Saddle Rock
Kensington
University Gardens
Lake Success
Bellerose Terrace
John F. Kennedy International Airport
Sow Mill Creek Wildlife Management Area
I 95 NJTP
I 78 NJTP
NJ 3
Essex County
Kings County
Union County
New Jersey
New York
City of New York
Queens County
Leaflet | © OpenStreetMap contributors

PREDICTING NYC TAXI FARE

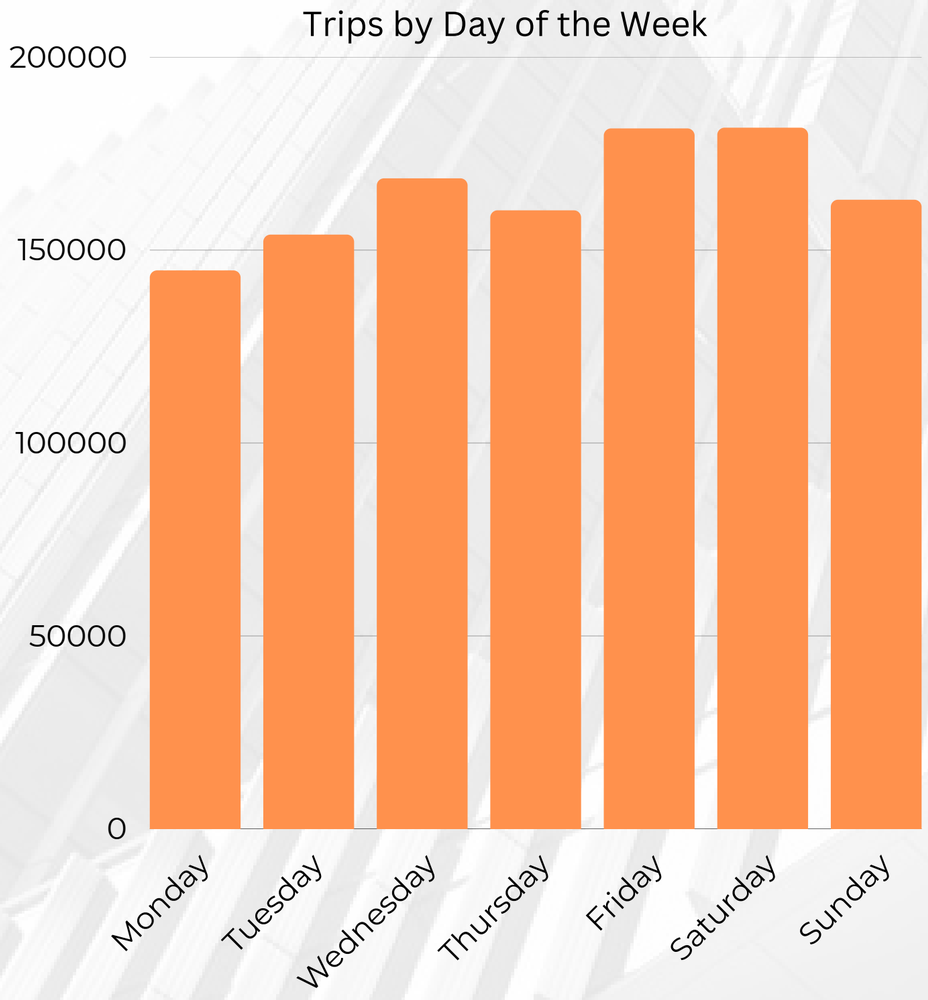
Xinning Zhang

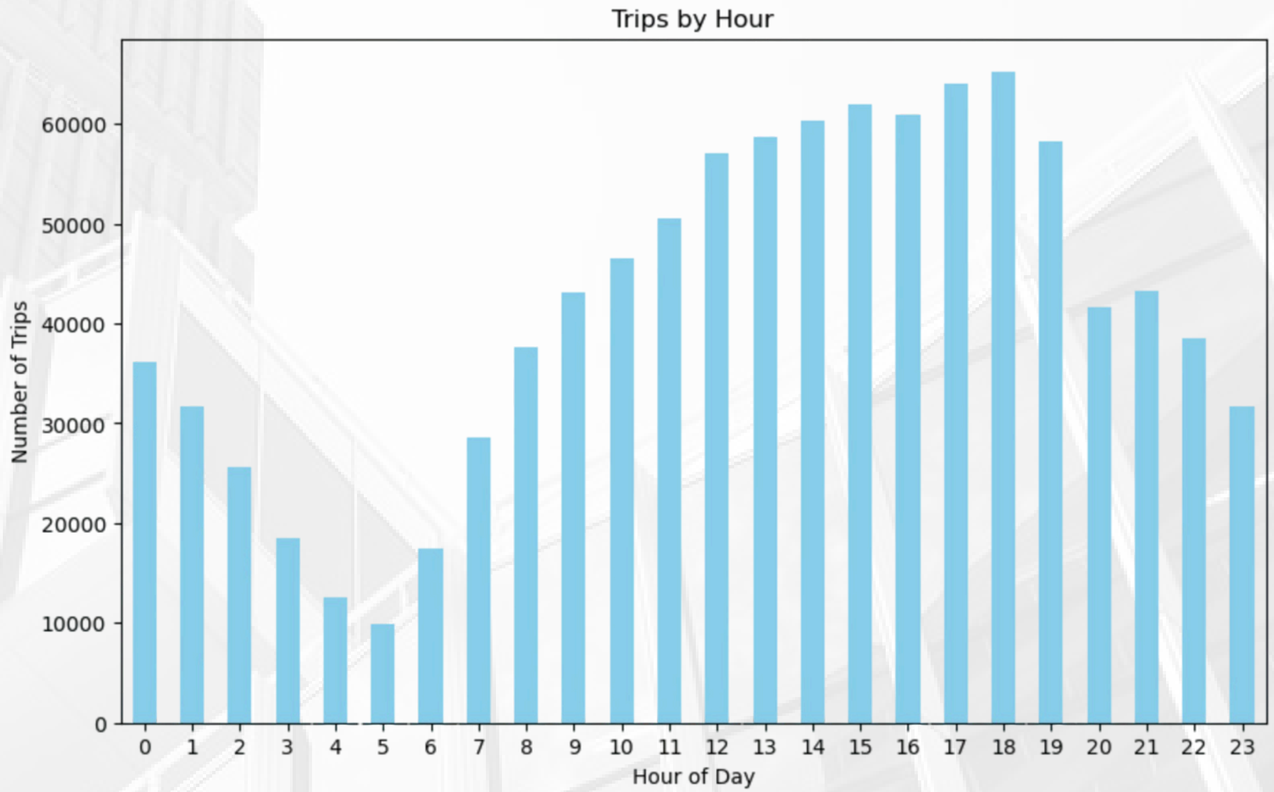
PROJECT OVERVIEW

This project focused on predicting New York City taxi fares. A linear regression and a random forest regression model were developed to improve fare estimation accuracy by analyzing features such as trip distance, number of passengers, and time.

This work highlights key insights into New York City taxi demand and pricing trends, providing practical tools for decision making.

EXPLORATORY DATA ANALYSIS





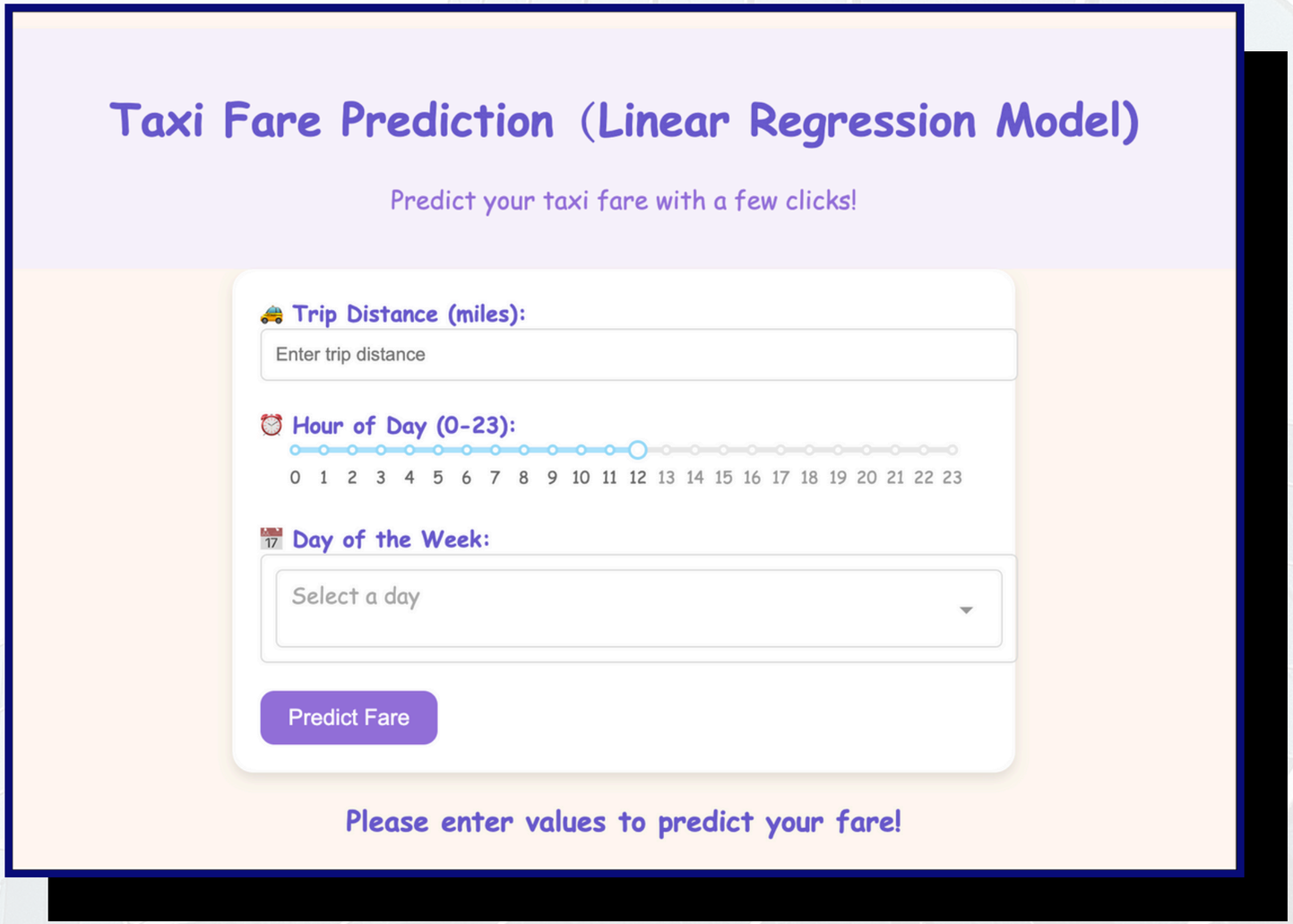
LINEAR REGRESSION MODEL

Model Accuracy:
RMSE: 6.98
R^2 Score: 0.71

Good Fit

Moderate Error

Potential for Improvement

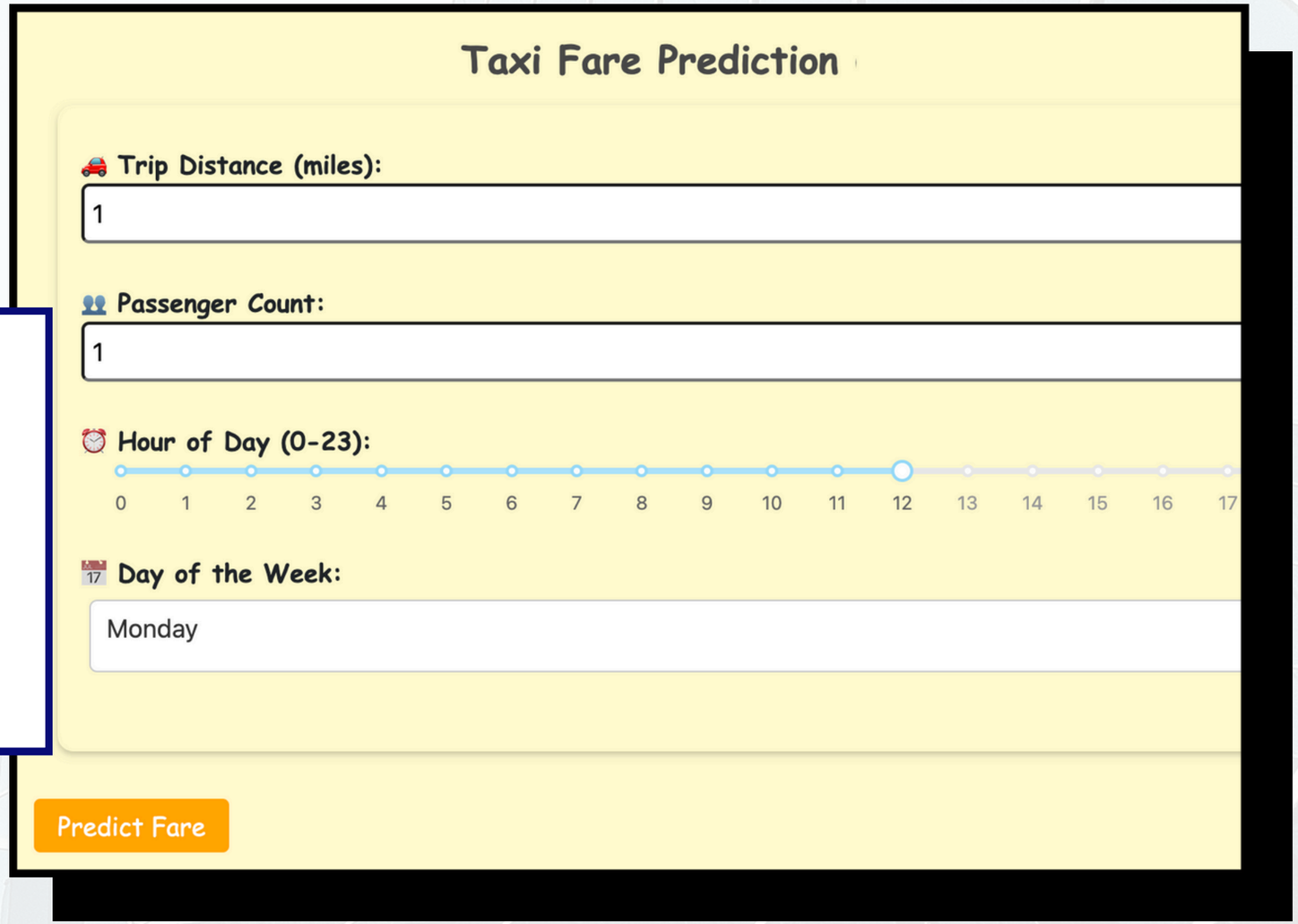


RANDOM FOREST MODEL

Model Accuracy:
RMSE: 7.71
R^2 Score: 0.65

Performance

The model shows **moderate accuracy**, capturing trends well but leaving room for improvement with better features or tuning.



MODEL COMPARISON

The Linear Regression model is more effective for the current dataset due to its better accuracy and simpler structure. However, **the Random Forest model** could improve with further optimization, making it a robust choice for datasets with more complexity or non-linear patterns.

Linear Regression Model

RMSE: 6.98

R^2 Score: 0.71

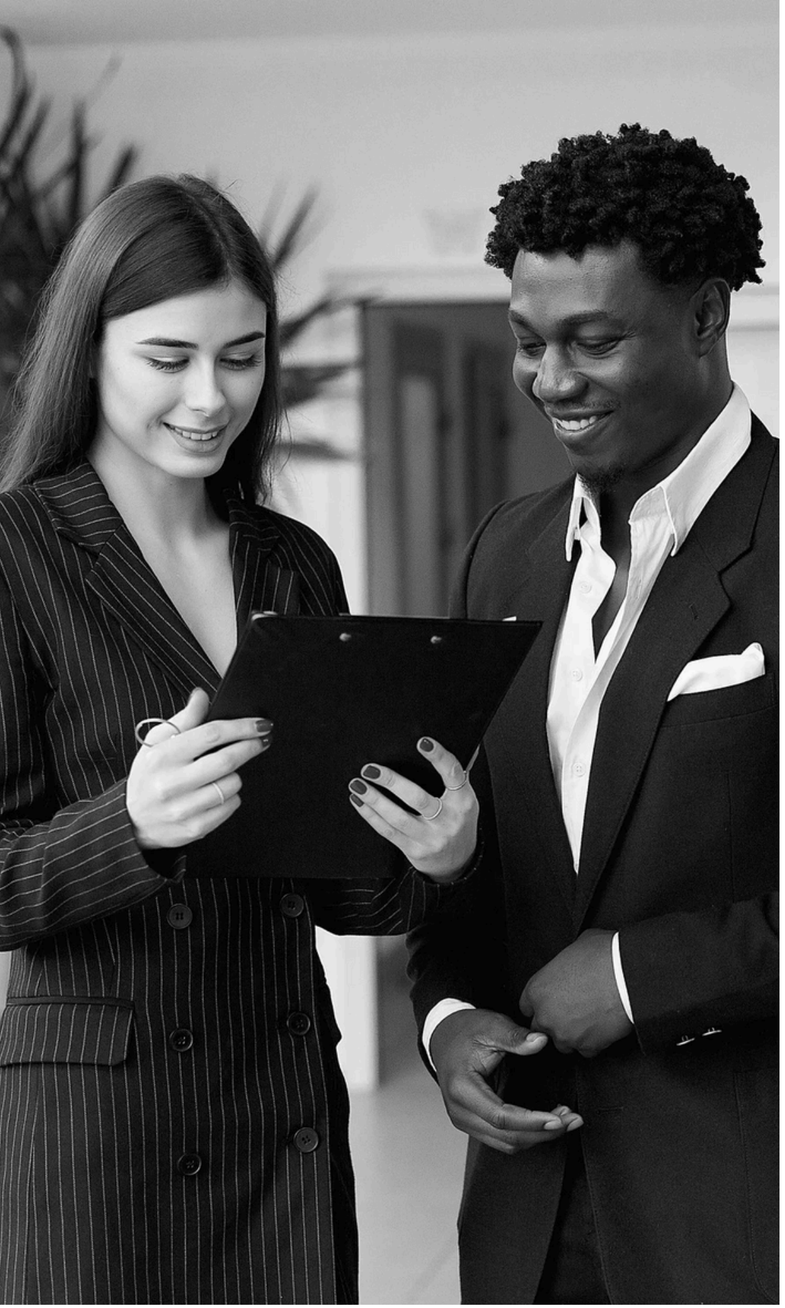
Simple, interpretable, and fits well

Random Forest Model

RMSE: 7.71

R^2 Score: 0.65

Handles non-linear patterns but needs tuning or better features



SUMMARY

Goal

Build predictive models to enhance NYC taxi fare estimation and decision-making.

Key Components

Built and compared Linear Regression and Random Forest models with an interactive map.

Insights

Trip distance is the strongest predictor; maps show high-demand zones for fleet optimization.