# A Time-Based Look at Bike Share Station Demand

DS 340H Capstone Project

Jennifer Ruffin

## Introduction

This study analyzes bike share demand at the five most frequently used Blue Bike stations across different months and times of day during the peak summer period from May 2024 to August 2024.

To achieve these goals, the study applies machine learning regression models, including Random Forest and Gradient Boosting Regressors, to predict hourly departure counts based on factors such as station location, month, day of the week, and time of day.

## Research Question

How do month, day of the week, and hour of day influence station-level trip demand for bike share services?

## Data

Source: Blue Bike Ride Data
Population: May 2024-August 2024
Manipulation: Month and start station variables encoded numerically; variables related to arrival stations removed

## Methods

**Random Forest Regression** used for predicting trip counts
**Gradient Boosting** for tuning the model

## Initial Exploratory Data Analysis

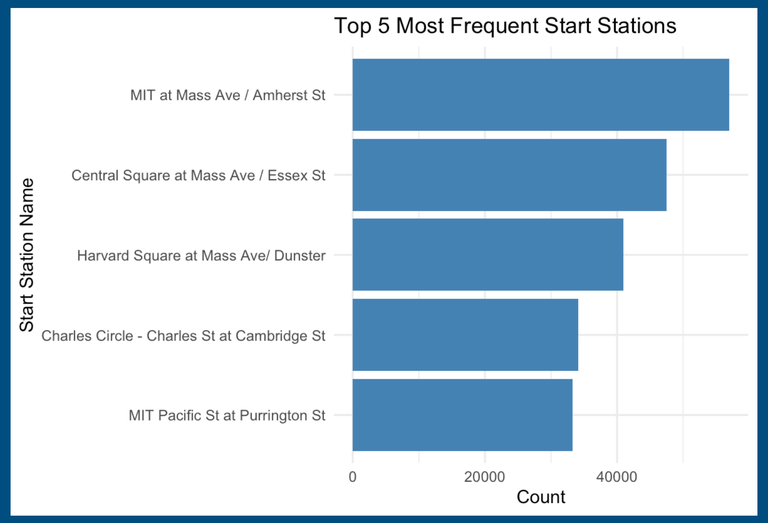


**Figure 1:** Bar plot showing top 5 most frequently used origination stations

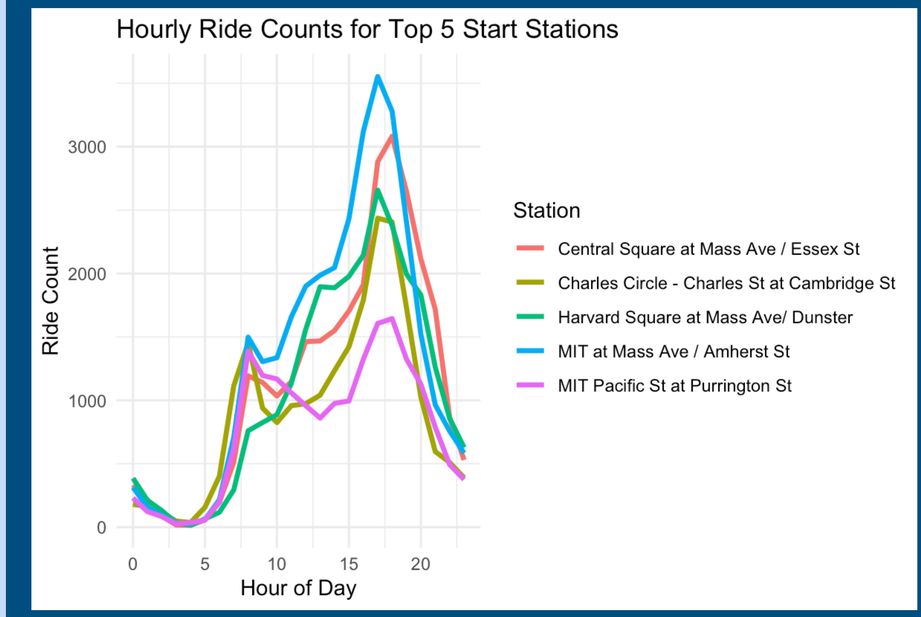


**Figure 2:** Line plot of the top 5 stations ride count by hour of the day

## Results

| | Random Forest Regressor | Gradient Boosting Regressor |
|---|---|---|
| Mean Squared Error | 11.95 | 14.25 |
| R^2 | 0.865 = 81% | 0.089 = 89% |

**Table 1:** Model Output

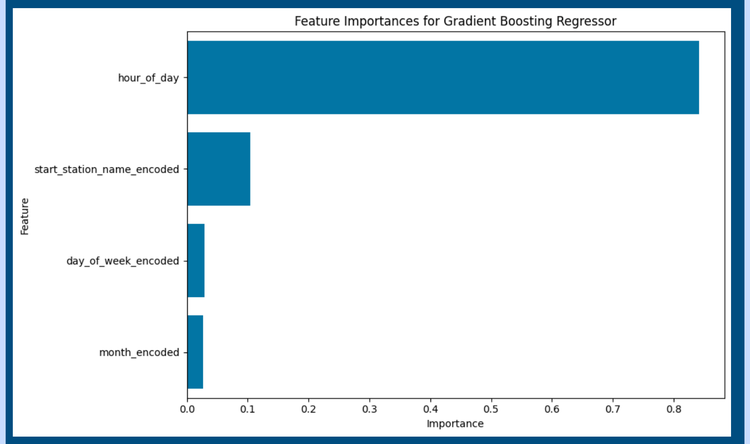


**Figure 3:** Feature Importance for Gradient Boosting Regressor

## Conclusion

- Gradient Boosting model explains 81% of the variability in bike trip counts
- Prediction error is around 14 trips
- Most important feature in predicting ride count is hour of the day
- Least important feature in predicting ride count is month of the year
- Future tests can analyze more specific temporal patterns such as time of the day (morning, afternoon, evening) for more concentrated analysis