

## Introduction

### Background:

Bluebike is a public Metro Boston bike share program, which placed thousands of bikes and hundreds of stations across the Greater Boston area since launching in 2011. In 2022, Mayer Wu of City of Boston has announced the *Better Bike Lane* 3-year project, in hopes to improve safety and encourage the bike network.

### Research Question:

Can we classify Bluebike stations into high, medium, or low demand categories based on bike lane types, station location, and user trip characteristics?

### Hypothesis:

The more the bike lanes are safe in the morning of central area, the more in demand the nearby Bluebike station will be.

## Results

Table 1. Performance Metric Scores of Classification Models

	Decision Tree	Random Forest	KNN	SVM	Gradient Boosting
Accuracy Score	0.9290	0.9706	0.9529	<b>0.9882</b>	0.9349
Precision Score	0.9417	0.9713	0.9541	<b>0.9919</b>	0.9390
Recall Score	0.9204	0.9648	0.9333	<b>0.9778</b>	0.9222
F1 Score	0.9213	0.9628	0.9329	<b>0.9815</b>	0.9194
Kappa Score	0.8757	0.9496	0.9180	<b>0.9793</b>	0.8863
AUC Score	0.9400	1.000	0.9992	<b>0.9973</b>	0.9904

## Visualizations

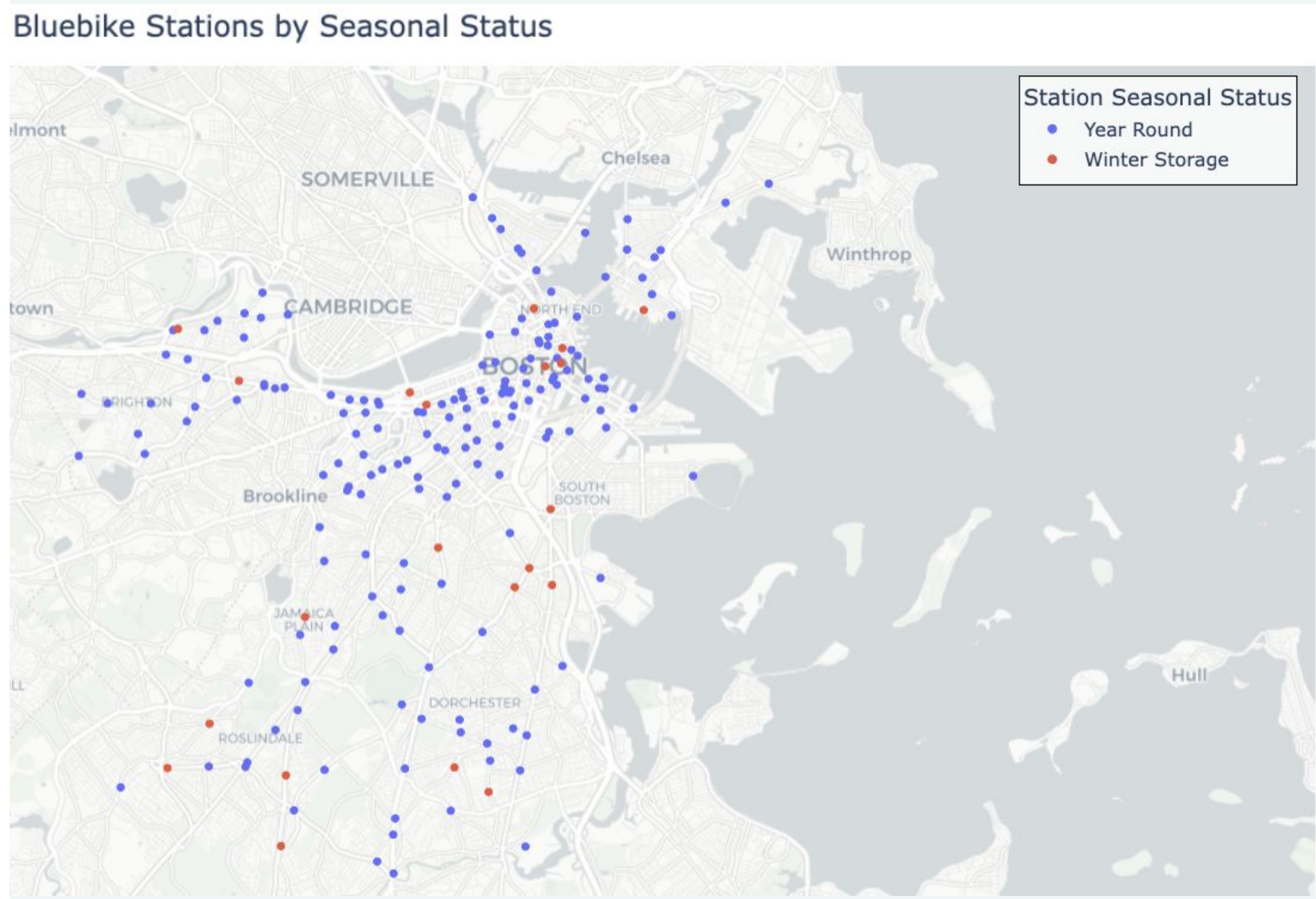
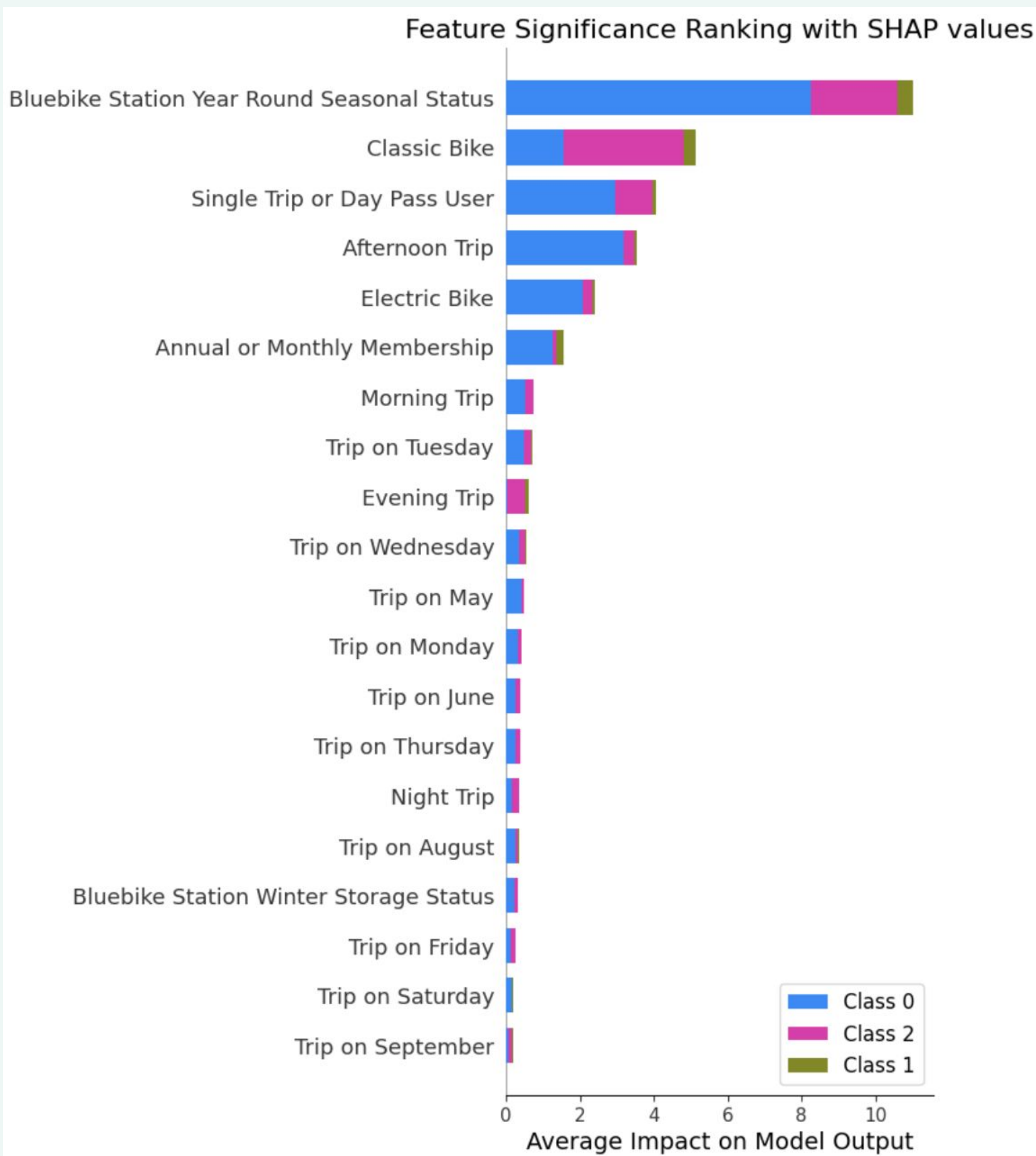


Figure 2. Feature Significance Ranking with SHAP values  
Figure 3. Bluebike Stations by Seasonal Status

## Data & Data Preparation

### Data Sources:

- Bluebikes System Data
  - Bluebikes Comprehensive Trip Histories of Year 2024
    - Trip Duration (seconds), Start Time and Date, Start Station Name & ID, Bike ID, User Type\*\*
  - Bluebikes Station Data
    - Number, Name, Latitude/Longitude, Municipality, Total docks
- City of Boston Existing Bike Network (2024)\*\*
  - Street Name, Bike Lane Divided, Existing Facilities, Shapefile

### Data Manipulation:

- By converting both geographic coordinates or shape data into GeoDataFrame, Bluebike Station and Bike Network dataset were combined where the closest bike lane were matched to each Bluebike stations within the 50 meter radius.
- Pre-defined by the City of Boston, each type of bike lanes in the area were assigned a comfort score of 0-2, 0 being least comfortable, and 2 being most comfortable.
- Users' trip distance and duration were calculated using geographic coordinates of the stations and start to end trip time.
- Station and bike lane informations were merged by each trip's start station ID



Figure 1. Types of Bike Lanes Ranked by Comfortability

## Discussion & Conclusion

- The top 5 most significant attributes to determine demands of Bluebike stations are: (1) Bluebike Station Year Round Seasonal Status, (2) Classic Bike, (3) Single Trip or Day Pass User, (4) Afternoon Trip, (5) Electric Bike
- Seasonal Bluebike stations are not necessarily limited to Downtown Boston, well-spread out across Greater Boston
- Casual Bluebike users contribute to more or less demand than subscribed membership users
- Limitation: (1) additional data on Bluebike users and other methods of transportation, (2) the City of Cambridge launches a similar project called *Broadway Safety Improvement Project* over the Winter to Spring 2025
- Future Work: including the Cambridge area for Cambridge Bluebike stations and MBTA subway/bus routes