An Emerging Mode of Transportation: Predictors of Electric Bike Usage in Boston

Nile Rankin-Watson, Wellesley College '25, Data Science Major Capstone

Background

As the world becomes more environmentally conscious, citizens have begun to swap individual cars for more eco-friendly options like public transportation, ride-sharing, and, increasingly, bicycling. In 2017, the Municipalities of Boston, Cambridge, Brookline, Everatt, and Somerville adopted the regional bike sharing system, Blue Bikes. Using the Blue Bikes app, customers can rent bikes for single trips, 24 hour passes, and monthly memberships. Blue Bikes now operates 4000+ bikes and 400+ stations operating at every T line, ferry, commuter rail, and over 60 bus routes. In 2023, Blue Bikes introduced e-bikes as an alternative to the classical mechanical bike. Since then users have increasingly opted into this option.

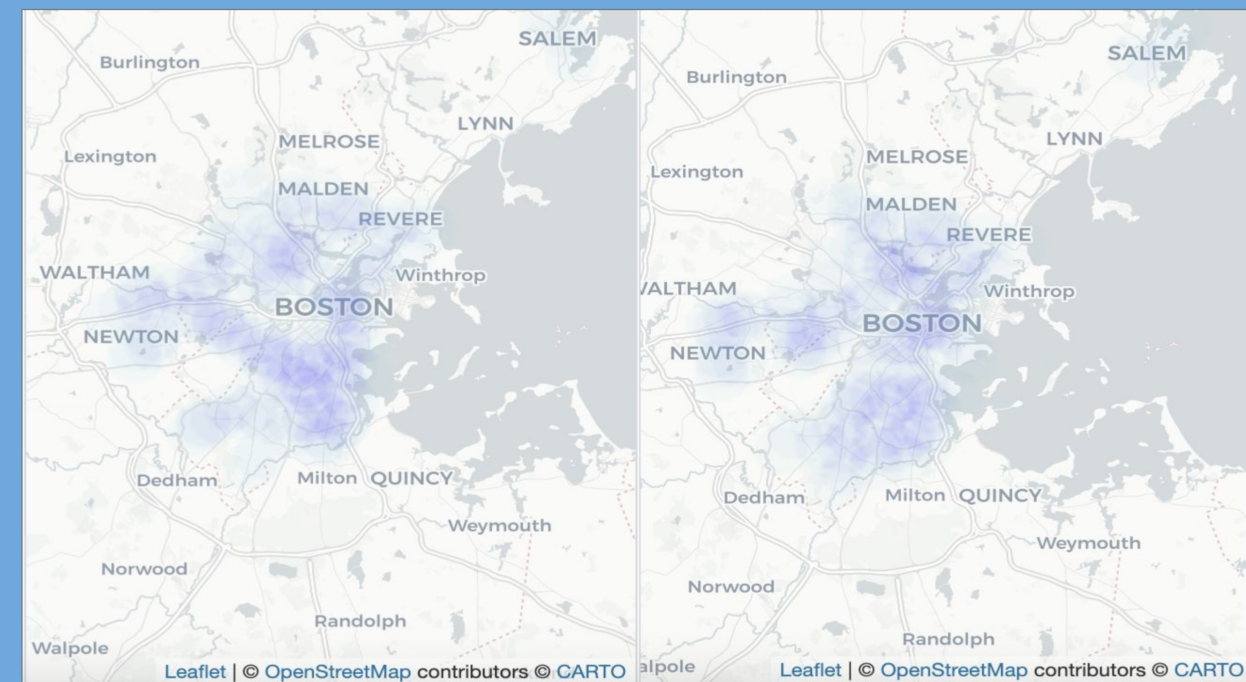
Research Question

What are the predictors of a Blue Bike rider's usage of an e-bike rather than a classic bike in 2024?

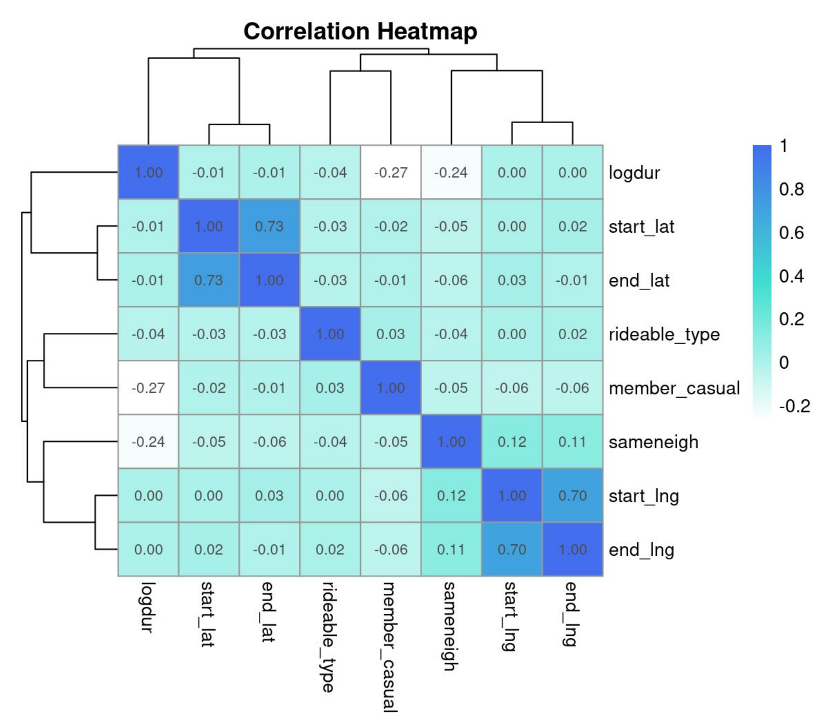
Data

Source: This exploration uses "Blue Bikes Comprehensive Trip Histories", a comprehensive dataset outlining individual trips made by patrons including the time of the trip, the start and ending stations, the bike's ID, the bike type, and the patron's user type.

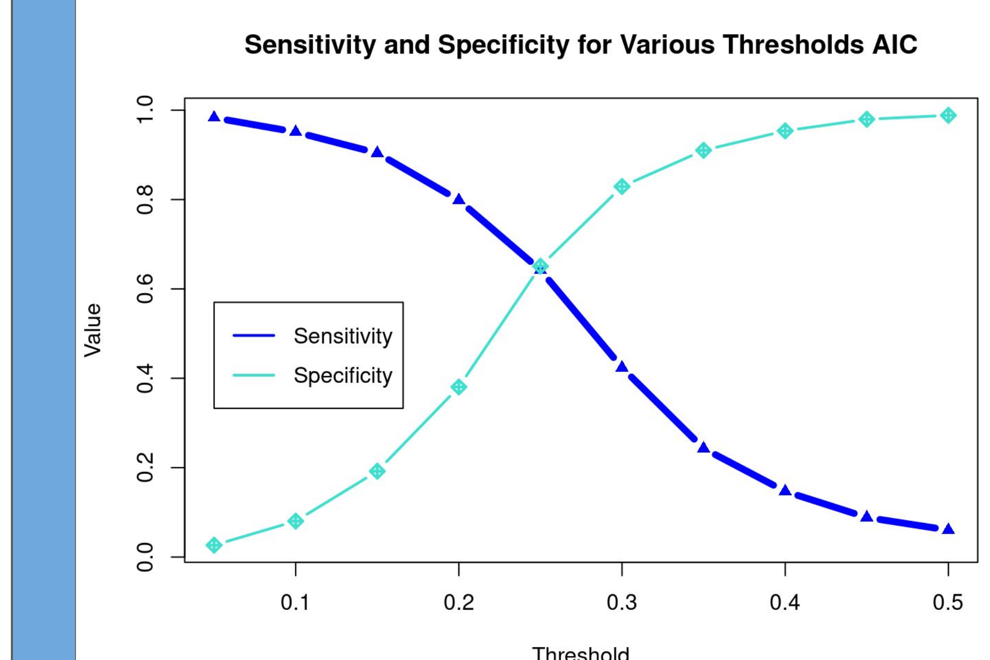
Manipulation: This data was cleaned by combining all of the 2024 trip history data into one dataset and sorting the data by date. I created a separate date variable as well as a time started and time ended variable for each trip. I deleted the rows with NA values, as relative to the total number of rows, the number of rows with NA values was negligible. I created variables representing the neighborhoods in which users started and ended their rides, as well as one that signifies whether or not users started and ended in the same neighborhood. Finally, I created a subset of 10,000 random rows for analysis.



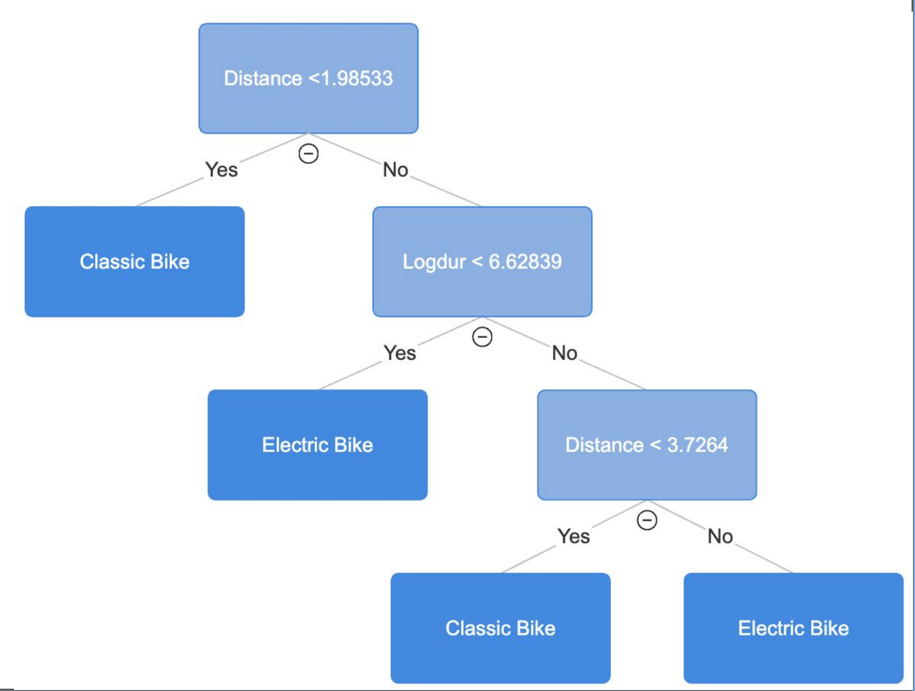
Maps of electric bike usage at starting (left) and ending (right) points. More people end their trips in the heart of Boston than start them.



Correlation heat map matrices. Many of the variables listed have small correlations to rideable_type (response variable). Latitude, trip duration, and starting and ending in the same neighborhood all have negative correlations, while longitude and user type (member_casual) have positive correlations



Overlaid line chart depicting sensitivity and specificity at a range of thresholds. Sensitivity is highest when threshold is low while the opposite is true for specificity.



Classification tree model. The tree created had four terminal nodes and used only two variables: Distance and logged duration. Based on 10-fold cross validation, this is the optimal pruned tree model.

Models & Analysis

I created three models based on my response variable (rideable_type). I coded a tree model and two binary classification logistic regression models: Akaike Information Criterion (AIC), and a Bayesian Information Criterion. After comparing all three models using various measurements, the AIC model proved to be the most optimal.

	Accuracy	Sensitivity	Specificity	AUC
AIC	**0.742**	**0.060**	0.060	**0.674**
BIC	0.740	0.045	**0.991**	0.673
Tree	0.735	0.180	0.935	0.598

Next, I assessed different thresholds to find the optimal threshold for sensitivity and specificity. I found that the optimal threshold for sensitivity was 0.015.

Results & Discussion

Result: The final model chose the variables of logged duration, the distance travelled, the neighborhood and city in which they started, whether or not their trips started and ended in the same neighborhood, and their starting and ending latitudes.

Key Findings: People are less likely to rent e-bikes if their trip starts and ends in the same neighborhood. E-bike rentals are more likely in neighborhoods like Dorchester and East Boston and in cities like Brookline and Revere, though some neighborhood data may be skewed due to limited sample sizes. Conversely, neighborhoods in Cambridge and Riverside have lower electric rental likelihoods. The analysis shows that trips ending further east are more likely to involve e-bikes, while trips starting eastward are less likely. This suggests people may use e-bikes to commute into cities like Boston but use other means for departure. Trip duration and distance have a smaller effect, but faster and farther trips are more likely to involve e-bikes, as they offer a time-saving alternative for longer journeys. This information can guide future decisions on e-bike distribution and station allocation.