Patterns in Blue Bike Usage: analysis at a per-bike level

Miraya Gupta '25 Data Science Major Capstone

WELLESLEY

Research Question

To what extent is the number of bikes in use at a given point correlated with time of day, day of week, season and location? Can these variables be used to predict the number of bikes in use?

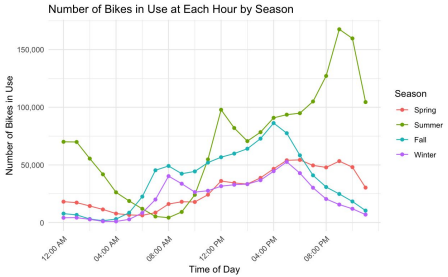
Background

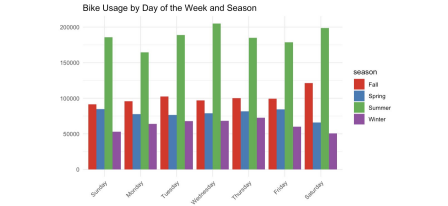
Redistribution of bikes to stations with more demand in New York City became an issue for the company. Want to understand demand patterns in Boston and what the company can do to match them.

Blue bike is a company as well as a public service so it is important that it both stays profitable and also serves customers for the best outcome.

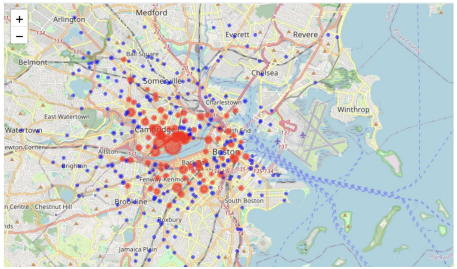
Data

Source: Blue bikes rides and station data
Population: rides from April 2022-April 2023, since these contain the relevant variables
Variables of interest: bikeid, month, season, day, starttime, stoptime, startstation, stopstation, usertype
Missingness: Handled through deletion





Number of bikes in use by start station on March 13, 2023



Models and Analysis

Bikes are utilised most in the Summer and Fall
There are certain stations at which bikes are utilised much more frequently than others
Plan to correlate this with data on number of bike docks
Plan to fit one of NN, linear regression, CART
Explore tripduration's correlation with bikeid

Discussion and Conclusion

Limitations

Predictive value: Since the relevant variable **bikeid** is no longer collected post April 2023, this method may not be useful in making predictions beyond a few years from now.