Availability and Usage of Blue Bike Stations

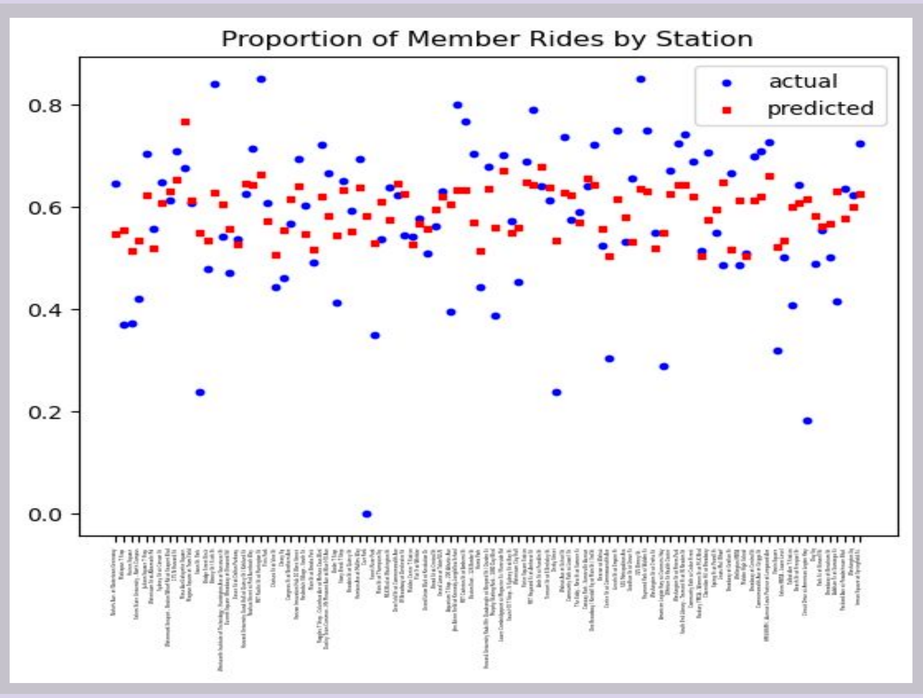
Ali Kyrouz ‘25 Data Science Major Capstone

WELLESLEY

intro

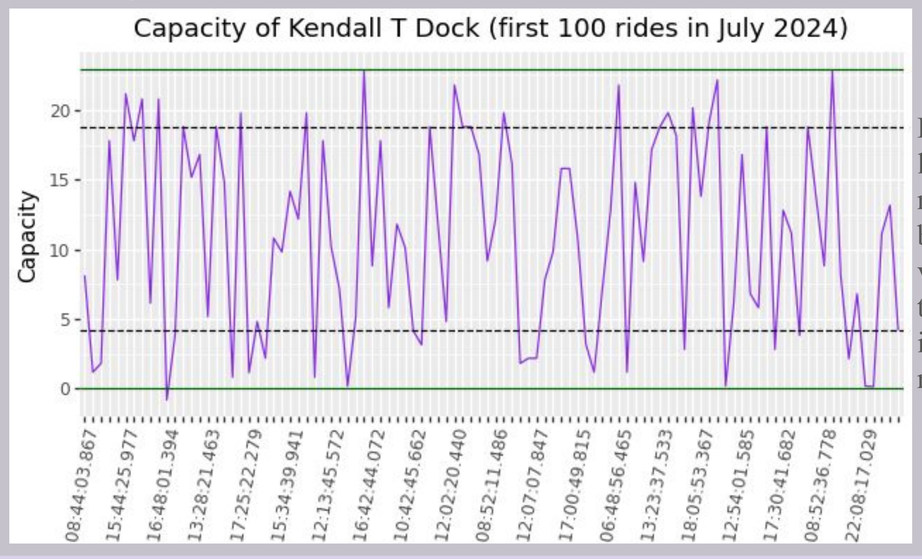
As of 2024, there are almost 500 Blue Bike stations in the greater Boston area, and almost 5000 bikes within the system. With this widespread availability, there are a range of qualities that one station could have that may make it more desirable to a potential customer, or, conversely, that could be put in place in response to a dedicated local population of customer.

How strongly correlated are features like frequent availability of bikes and dock space, or availability of electric bikes to the frequency at which a station is used? How is this correlated with the frequency of use of a station by members vs. non-members?



Data cleaning/algorithm

- Looked at data from July 2024, could also be applied to all months after addition of electric bikes, from December 2023 to current.
- Removed information from ‘Current Blue Bike Stations’ file when stations were not present in the trip data.
- Used average dock space for 5 stations missing data
- Calculated statistics for each station, such as the proportion of rides that are taken by members, or that use electric bikes.



Dark green solid lines: max and min capacity, black dotted line: value adjusted to to keep capacity in a logical range.

Model

Attempting to predict proportion of rides taken by members out of total rides for a given station, I fit a linear regression model with features including the total number of docks at a station, the amount of times the station was empty or full over the month, the total rides from the station and the proportion of electric bikes out of the total rides from the station. these are not very good predictors :(

Conclusion

There is not a very strong correlation between these current features, although the proportion of electric bikes had the strongest correlation out of all of them.

Sources :)

blue bike and something else :)