Availability and Usage of Blue Bike Stations

Ali Kyrouz ‘25 Data Science Major Capstone

WELLESLEY

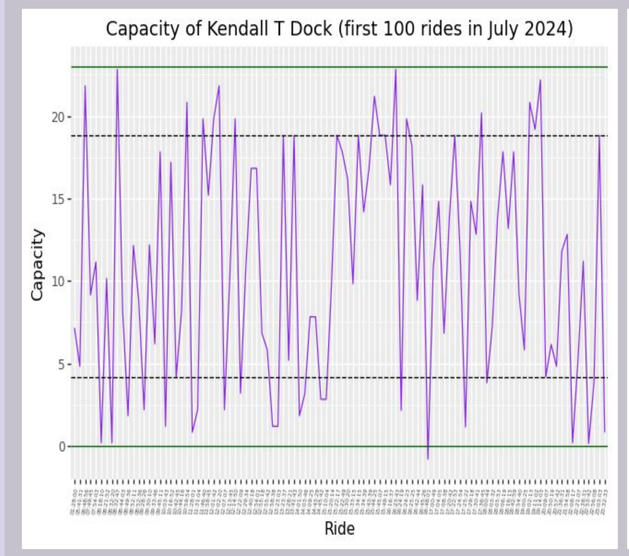
Introduction

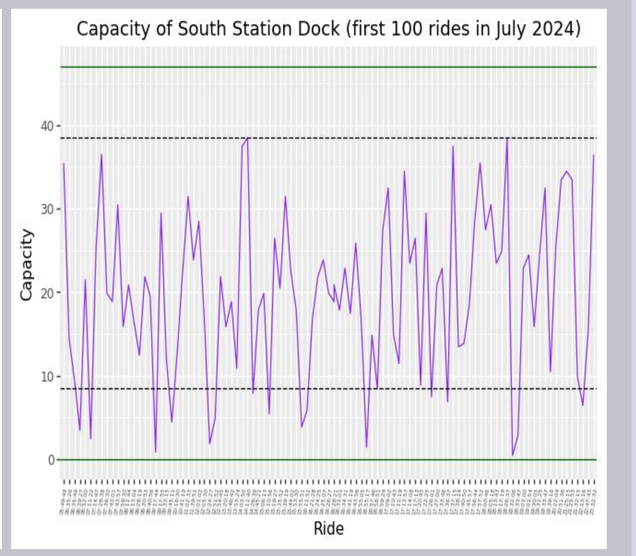
As of 2024, there are almost 500 Blue Bike stations in the greater Boston area, and almost 5000 bikes within the system. With this widespread availability, there are a range of qualities that one station could have that may make it more desirable to a potential customer, or, conversely, that could be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

How strongly correlated are features like availability of bikes, dock space, or electric bikes to the proportion of trips from one station out of the total trips in a month? How are these correlated with the proportion of total trips from the station by members?

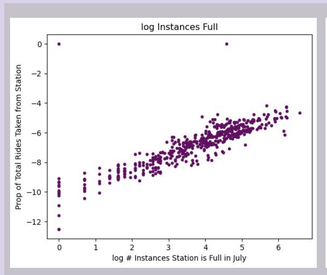
Data Cleaning & Algorithm

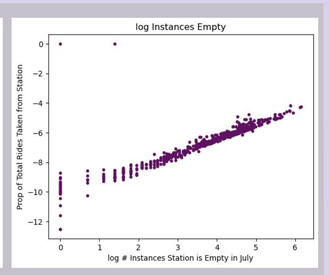
- Looked at data from July 2024, could also be applied to all months after addition of electric bikes, from December 2023 to current.
- Removed information from ‘Current Blue Bike Stations’ file when stations were not present in the trip data.
- Used average dock space for 5 stations missing data
- Calculated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

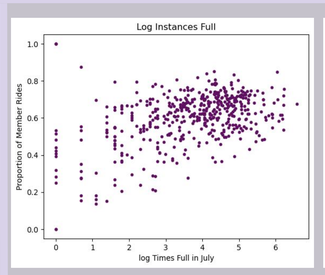


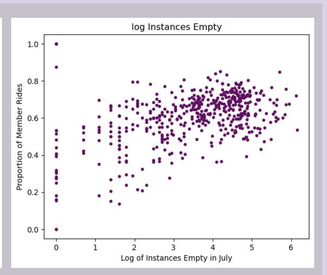


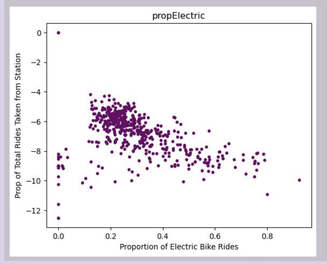
Dark green solid lines: max and min capacity, black dotted line: value adjusted to to keep capacity in a logical range.

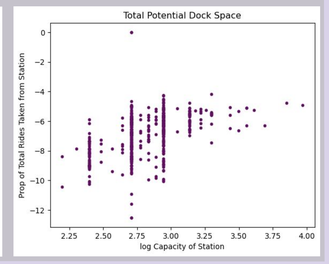


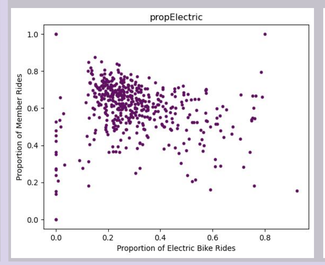


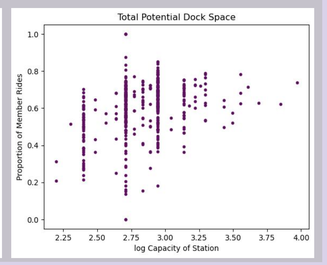












Features: 'Total Capacity', 'Times Empty', 'Times Full', 'propElectric', 'propTrips', 'propMembers'

$R^2 = 0.9343$ $R^2 = 0.3495$

Model

I fit two linear regression models: one predicting the proportion of total trips in the month that left from a given station, and one predicting the proportion of rides taken by members from a station. If the company’s re-stocking of bikes between stations is successful, one would expect there not to be a relationship between how many trips leave from a station and how frequently the station runs out of bikes: they would be refilled proportionally to activity level. This is clearly not the case, as the best predictors for proportion of rides from a station was the number of times it ran out of bikes. However, within rides from a station, there is not a similarly strong predictor for how many of these rides were by members.

Conclusion

Despite redistribution efforts, higher activity levels at a station are still correlated with less availability of bikes, but how many of the rides are from members is not strongly correlated with a difference in availability. Members likely ride bikes more frequently, which may explain the slight correlation between times empty and full and proportion of members.

Sources :)
Ashish Kabra, Elena Belavina, Karan Girotra (2019) Bike-Share Systems: Accessibility and Availability. Management Science 66(9):3803-3824. https://doi.org/10.1287/mnsc.2019.3407
Bluebikes System Data | Blue Bikes Boston. https://bluebikes.com/system-data. Accessed 6 Dec. 2024.