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

Miraya Gupta '25 Data Science Major Capstone

Miraya Gupta '25 Data Science Major Capstone

Research Question

To what extent is the number of bikes in use at a given point correlated with hour, day of week and season? What is the usage of a typical bike? Based on the bikes in use at a given time, is there opportunity for dynamic pricing?

Background

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

Data

Source: Blue bikes rides and station data
Population: rides from April 2022–April 2023
Variables of interest: bikeid, month, season, day, starttime, stoptime, startstation, stopstation, usertype
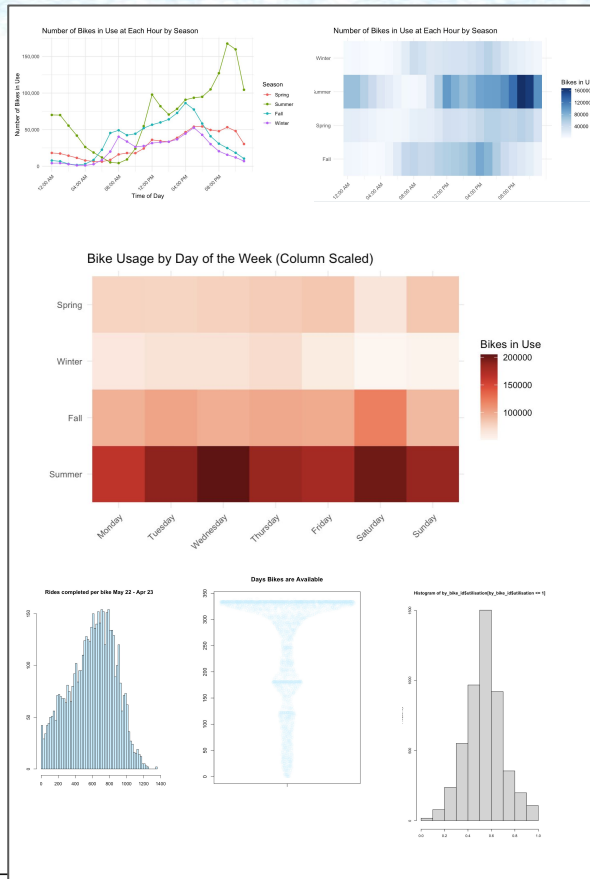**Missingness:** Handled through deletion

Analysis

*Bikes are most used in the summer, then the fall, winter and spring. In winter, most bikes are in use during working hours, and in summer most bikes are in use after 12 noon.
*Weekends have highest usage in summer, winter and fall. In spring, weekdays have highest usage.
*Most bikes are available for use year-round. However, the proportion of days they are used for trips is normally distributed with mean 0.5.

Number of Bikes in Use at Each Hour by Season

Number of Bikes in Use at Each Hour by Season

Bike Usage by Day of the Week (Column Scaled)

Rides completed per bike May 22 - Apr 23

Days Bikes are Available

Histogram of by_bike_id$utilisation[by_bike_id$utilisation <= 1]

Models

*A linear model was constructed using number of bikes in use at a given hour as the response variable
*I made the assumption that time variables like day, week, month and season were uncorrelated
*The linear model will forecast how many bikes are in use at a given time

Discussion and Conclusion

*Based on the results of the linear model, it is possible to interpret regression coefficients to understand what internal factors drive bike usage.
*Given the current pricing scheme:
-$2.95 for up to 30 min, +$0.25 per min after that
- $29 per month or $13 per month (annually) for members
It makes sense for BlueBikes to use a dynamic pricing strategy at peak usage hours, like summer weekend nights or fall Saturdays.

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.