Abstract

The goal of this project was to assist Specialized Bicycle Company* with finding the ideal location for their flagship store in the Central Park area of NYC. I worked with data from MTA stations to explore rider counts between stations across different time frames (across seasons, weekdays/weekends, 2019/2021, and throughout the day). Based on my analysis, locations alongside Central Park have less pedestrian foot traffic compared to other areas further from the park and my recommendation is to build the store near the 96th St location of the 1•2•3 subway line.

Design

Specialized Bicycles is an established bicycle manufacturer and is looking to expand its presence in the cycling community. To do so, they are going to open a flagship store in the Central Park area of New York City. This project provides insight on potential store locations that will maximize the number of individuals renting bicycles to ride in Central Park as well as maximize brand awareness.

Data

MTA turnstile data was used for this analysis. Having consulted Google Maps to determine cycling-friendly streets (so customers can safely get from MTA stations to Central Park), the analysis focused on the following stations: 96th St (lines B•C, 1•2•3), Cathedral Pkwy/110th St (lines B•C, 1•2•3, 4•5•6), 72nd St (lines B•C, 1•2•3), 5th Av/59th St (line N•Q•R•W), and 77th St (line 6). The hours of interest included 8:00-20:00 (inclusive) as these represent typical business hours for bicycle rentals. The months included were April (Spring), July (Summer and primary tourist season), and September (Fall). Lastly, considering the disruptions caused by the COVID pandemic, data for the above months were compared between 2019 and 2021.

Algorithms
- Filtered data to the stations of interest and times of interest (8:00-20:00).
- Added variables of interest (e.g., total counts [entries + exits], seasons, weekdays/weekends).
- Dropped missing data and dropped outliers (total counts that had a z-score > 3 for that station [when grouped by year, season, and weekday/weekend]).
- Created bar plots to visualize the data and determine the ideal location for the store.

Tools

Data was ingested into a SQL database. Using SQLAlchemy, the MTA stations of interest were then queried into Python. Further data cleaning and exploring occurred in Python with data visualizations created using Matplotlib.

Communication

Slides and visuals were presented and posted on my personal GitHub page.

* Fictionalized project with Specialized