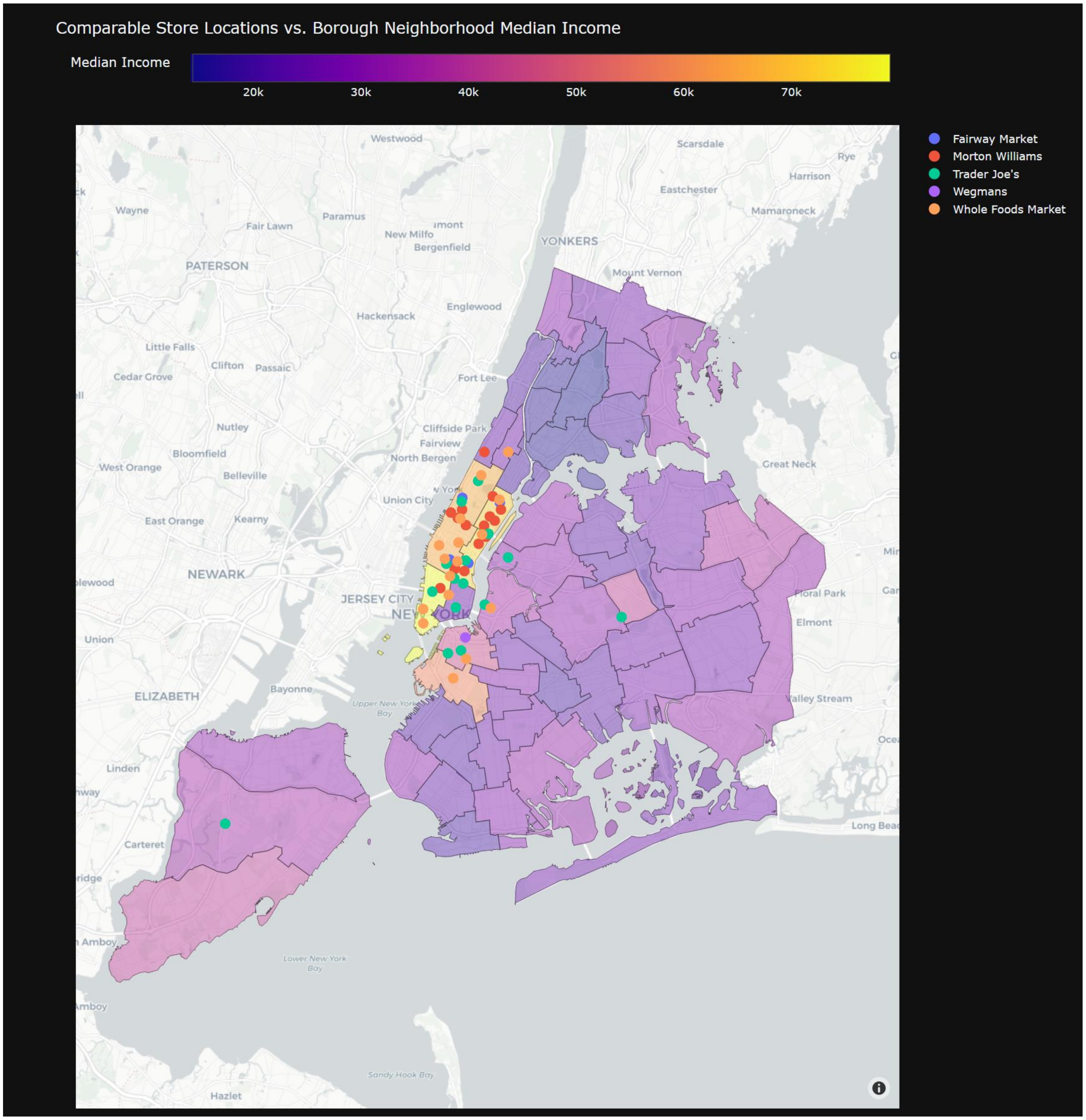
Question: Where in New York City would you advise Trader Joe's to open their next store?



1. The following above choropleth scatter chart displays current location of Trader Joe's and comparable high end grocery stores with overlayed median income statistics in each public use micro data area (PUMA) boundary for the NYC area. We can see that little to no comparable stores exist in the **Staten Island** borough while, slightly tough to tell, the Borough on average has the second to highest median income across PUMA boundaries. This chart does not fully explain nor pinpoints an exact address for the Trader Joe's placement – it is a simple enough opener to initially explain the reasoning though. All charts were built in plotly and are saved as webpages in the associated zip folder.

2. One can think of this problem as peeling back the layers of an onion. There is only finite NYC land area where a store can be placed. The goal was to start at a macro level and cross off as many bad locations, backed with layers of data, as possible. This is a breakdown of my thought process and steps for the assignment:
    a. Research for the top drivers when opening a new store. The major ones quoted were - Source Link:
        i. Competition of other retailers in the area
        ii. Economies and demographics of the area
        iii. Traffic Patterns
        iv. Recent and planned residential growth
    b. Next step was to leverage Google Map's API to pull in all comparable grocery stores to Trader Joes within each of the 5 Boroughs. The API gave me access to the latitude, longitude and other meta data for each store. Allowing me to see where the least amount of competition would be for a new Trader Joe's without cannibalizing current stores in operation.
    c. Utilizing Data.gov I found (and learned) about PUMA's, or public use micro data areas, which are specked out specifically for the Census Bureau. These are non-overlapping statistical geographic areas that partition each state into areas containing no fewer than 100k people. The NYC Neighborhood Financial Health Index provides a range of scores covering 5 goals that point to different aspects of financial health for each neighborhood. The following goals are scored on a scale of 1-10:
        i. Access to Affordable, High-Quality Financial Services
        ii. Access to Affordable, High-Quality Goods & Services
        iii. Access to Quality Jobs & Income Supports
        iv. Stable Housing & Capacity to Limit Financial Shocks
        v. Opportunities to Build Assets & Plan for the Future

       I used all of the Goal Index Scores, plus the ethnicity breakdown of each Borough to create radar charts. The radar charts allowed me to compare all of the goals and ethnicities to each Borough in a single capture. The Borough with the most area in the NFH Goals chart will be the best off comparatively. The best-looking radar chart here, with the caveat of intense store competition/cannibalization in Manhattan from the prior layer, was **Staten Island**. When adding the ethnicity radar chart, we can also see that Staten Island has the highest percentage of White ethnic makeup. Unfortunately I do not have quantitative empirical evidence to back this up, but have personally visited several Trader Joe's. With confidence I can say their health-conscious product offerings and price levels are designed to target higher educated and wealthier ethnic groups though. Another drilldown radar chart to the PUMA districts in Staten Island would narrow the scope down to the neighborhoods of, New Springville, South Beach, Tottenville, Great Kills and Annadale.

       The radar charts are attached at the end of the document.

d. To scale each Borough for population density I calculated the total area of each Borough and divided by the number of PUMA districts (used as a proxy for population count). This is a strike against Staten Island, because it is the least populous dense out of the 5 boroughs

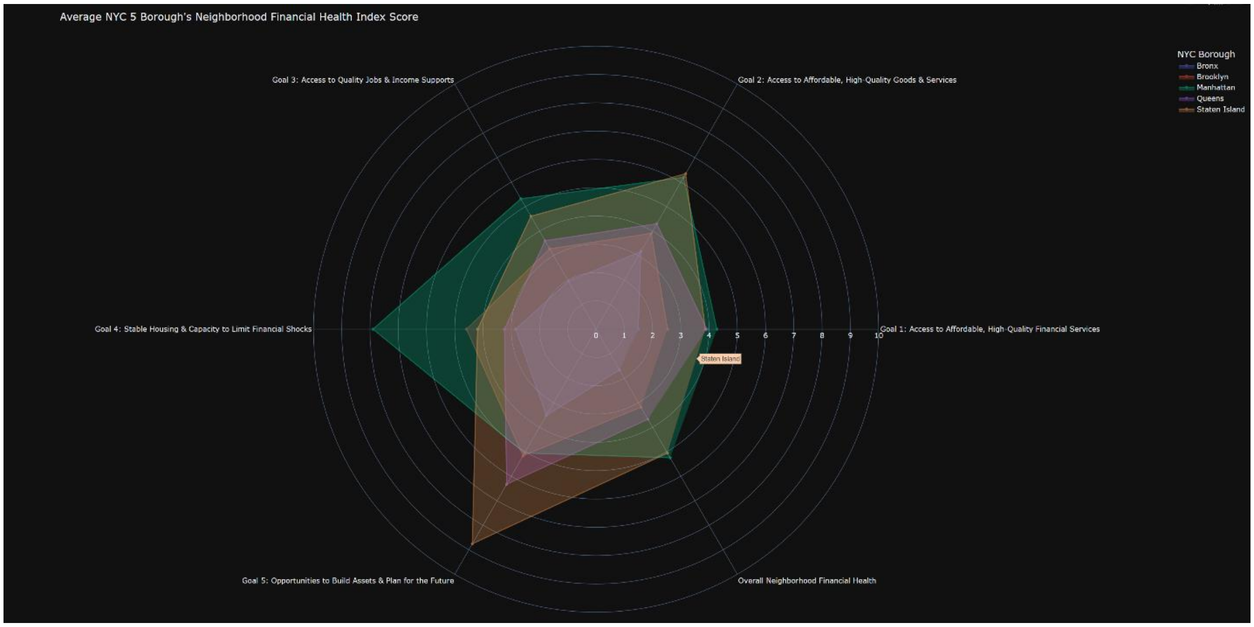
borough	area	puma_count	puma_count_density	X_times_Staten_Island
Bronx	1,187,174,785.83720	10	0.00842	4.55317
Brooklyn	1,934,143,375.00760	18	0.00931	5.03051
Manhattan	636,520,830.59270	10	0.01571	8.49211
Queens	3,042,054,210.09220	14	0.00460	2.48765
Staten Island	1,622,985,023.17000	3	0.00185	0.99916

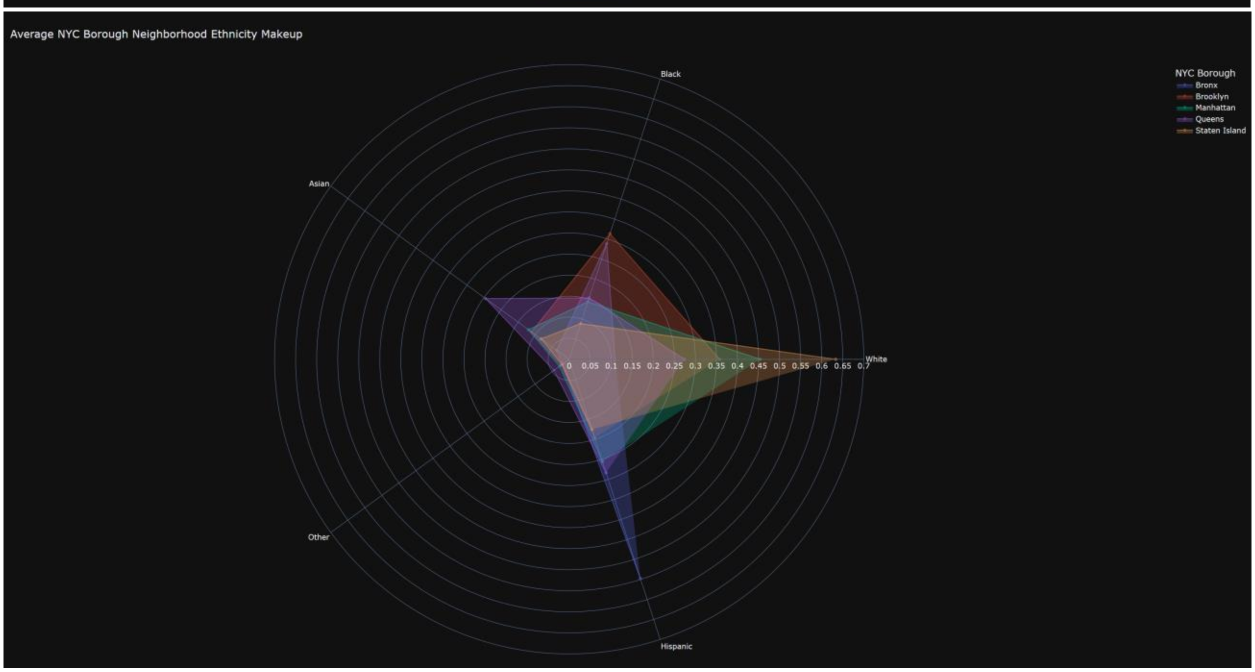
3. If given more time I would have added other data layers in this order:
    a. Foot and vehicular traffic heat maps with public transport routes across Brooklyn, Queens and Staten Island: this data could help assess if there were pockets of high population density and traffic paired with favorable NFH Scores
    b. Commercial vs. residential zoning geospatial data: This kind of data would be used to further target the feasible areas of a commercial storefront development.
    c. Comparable store credit card data – paired with attached demographic metadata: this would help assess discern a “winning” store location vs. a failing store location. Based on the relative winning stores, one could then target specific or idiosyncratic address characteristics which could help explain boosted performance.
        i. I felt I was unable to gather the data necessary to truly model a good vs. bad store and thought it would be better to drill down with visuals and simple statistics to smaller land area where a regression type model could be developed to further analyze

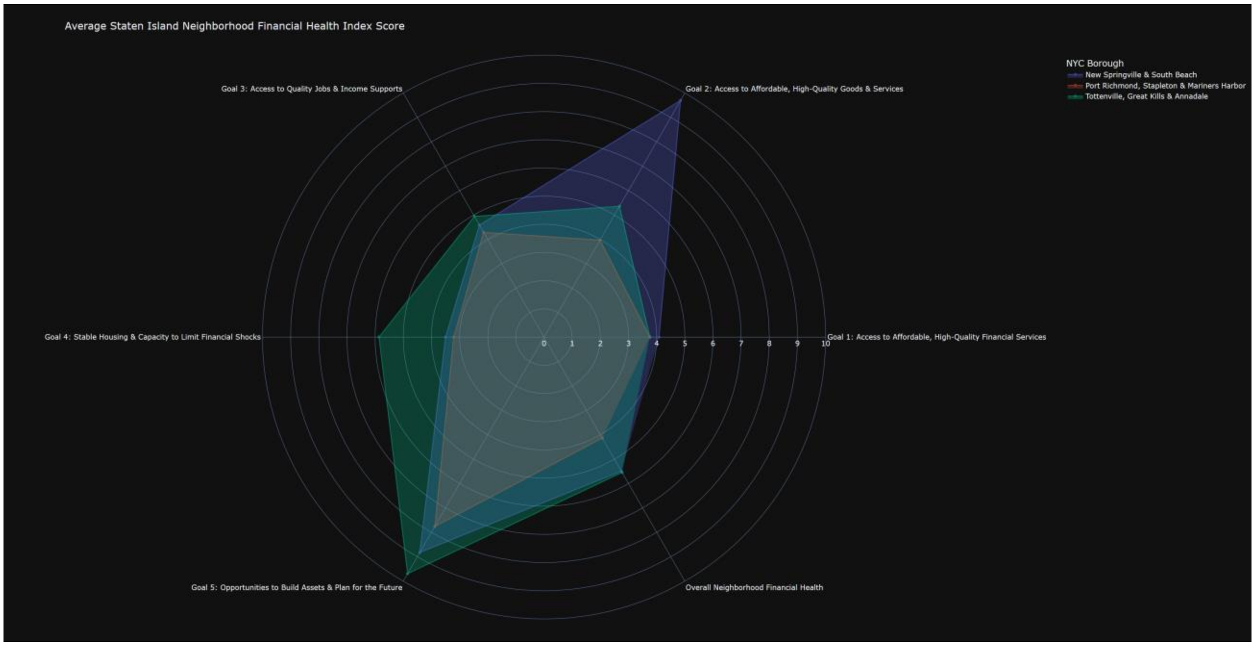
Appendix:

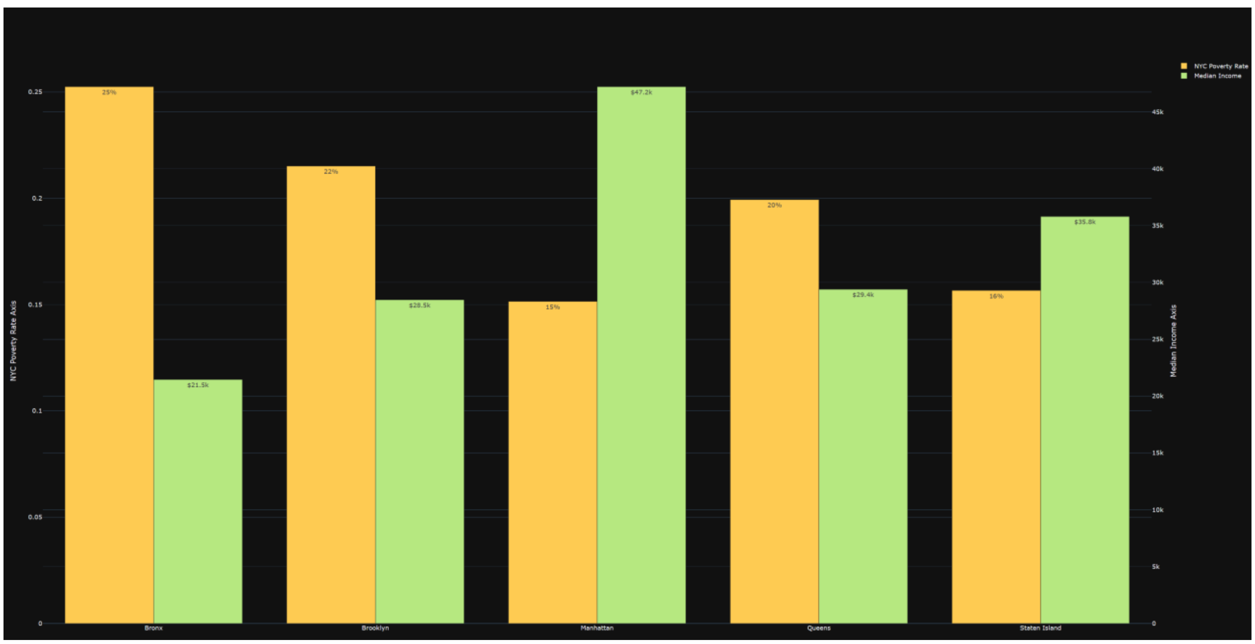
Data Cleaning and Introspection Techniques + Other Commentary:

- Put together a data introspection object to quickly filter and inspect different data sets
- This allowed me to home in on a slightly richer and cleaner dataset faster
- General scrubbing was done in pandas using string manipulation and a single instance of Fuzzy string matching
- Initial code structure that was designed (in the traditional .py files) was too rigid and taking too long, so I reverted to notebooks for the analysis. In the real world, I would harvest and functionalize the good parts of each notebook into a more structured and scalable set of objects


Average NYC 5 Borough's Neighborhood Financial Health Index Score
NYC Borough
Bronx
Brooklyn
Manhattan
Queens
Staten Island
Goal 1: Access to Affordable, High-Quality Financial Services
Goal 2: Access to Affordable, High-Quality Goods & Services
Goal 3: Access to Quality Jobs & Income Supports
Goal 4: Stable Housing & Capacity to Limit Financial Shocks
Goal 5: Opportunities to Build Assets & Plan for the Future
Overall Neighborhood Financial Health
0 1 2 3 4 5 6 7 8 9 10
Staten Island


Average NYC Borough Neighborhood Ethnicity Makeup
NYC Borough
Bronx
Brooklyn
Manhattan
Queens
Staten Island
Black
Asian
White
Other
Hispanic
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7


Average Staten Island Neighborhood Financial Health Index Score
NYC Borough
New Springville & South Beach
Port Richmond, Stapleton & Mariners Harbor
Tottenville, Great Kills & Annadale
Goal 1: Access to Affordable, High-Quality Financial Services
Goal 2: Access to Affordable, High-Quality Goods & Services
Goal 3: Access to Quality Jobs & Income Supports
Goal 4: Stable Housing & Capacity to Limit Financial Shocks
Goal 5: Opportunities to Build Assets & Plan for the Future
Overall Neighborhood Financial Health
0
1
2
3
4
5
6
7
8
9
10


NYC Poverty Rate
Median Income
NYC Poverty Rate Axis
Median Income Axis
25%
$21.5k
22%
$28.5k
15%
$47.2k
20%
$29.4k
16%
$35.8k
Bronx
Brooklyn
Manhattan
Queens
Staten Island