TRADER JOE'S

NYC Location Analysis

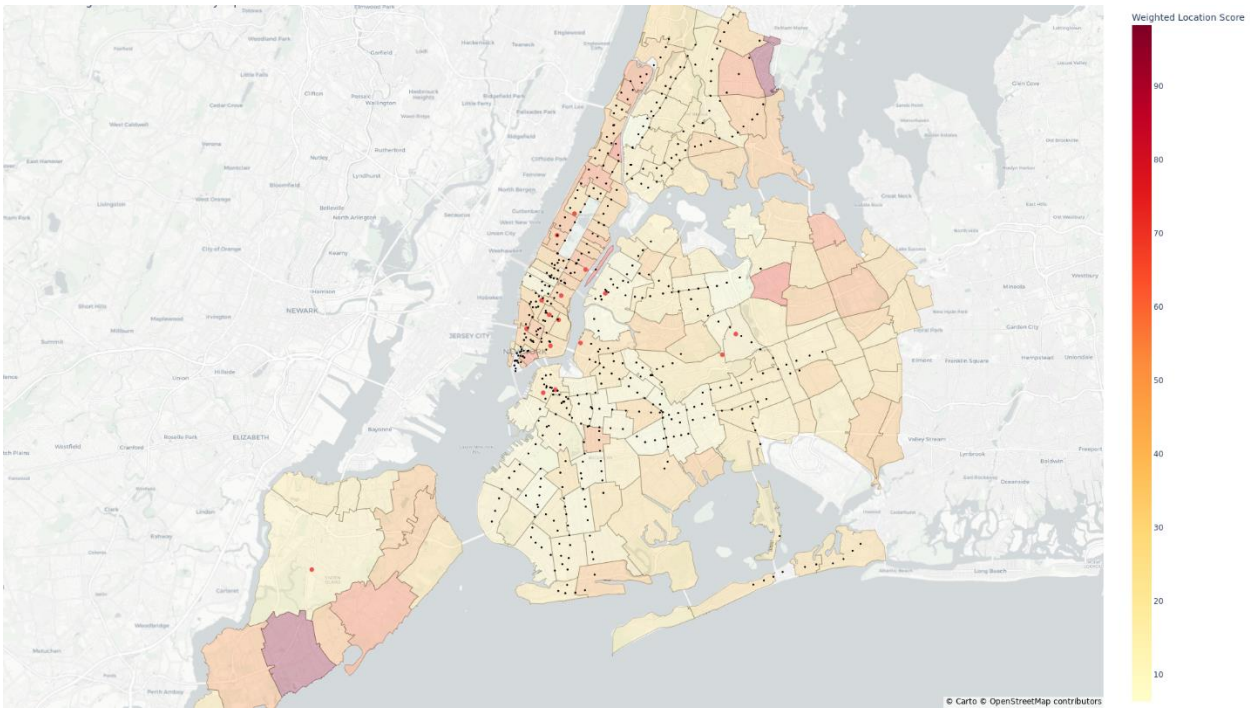
Logan Chalifour

Linked content – including all code, interactive visuals, and data sources files:

- GitHub repository
- Interactive web notebook
- Source notebook

I. Overall Recommendation

The map below highlights optimal zip codes for a new Trader Joe's store in dark red, with less optimal locations in pale yellow. This is based on a single weighted location score which is derived from several other data points (explained in Section 2). The sixteen existing Trader Joe's stores are highlighted as red points with subway stops as smaller black points to understand where existing stores are already located as well as their proximity to public transportation. You can see all existing stores, including competitors, in Appendix – Exhibit A.



With the data provided above, I return the top-scoring zip code for each borough. I manually filtered out Manhattan's highest-scoring zip code in Roosevelt's Island (10044) since it's unlikely residents from other zip codes would shop there due to its isolated nature. I also strip out a zip code in Washington Heights (10039) to keep the suggestion for Manhattan further downtown. This is acceptable since the location scores of the top three zip codes in Manhattan were all relatively close in the 60-70 range.

Zip Code	Borough	Neighborhood	Weighted Location Score
10475	Bronx	N/A	98.33
10312	Staten Island	Great Kills	95.65
11355	Queens	Queensboro Hill	66.48
10038	Manhattan	Two Bridges	62.10
11225	Brooklyn	Crown Heights	50.80

My recommendation is to open a new Trader Joe's store in one of the zip codes above. As expected, there is no right or wrong answer here necessarily, but rather a relative strength of the recommendation. In this case, I would recommend placing the new store somewhere in Manhattan as opposed to one of the other boroughs. This is because nine of Trader Joe's sixteen existing stores in New York City are located within Manhattan, suggesting it fits their customer base the best.

With a population of roughly 24,000 in 2023 and a two-year population growth of roughly .5%, Two Bridges of downtown Manhattan is an optimal location for a new Trader Joe's store when compared to other zip codes. It has eleven existing grocery stores, only one of which is a large chain (CTown Supermarkets). Thus, the normalized scores for population per store and chain come in at ~34 and ~89 respectively, with a population growth rate scoring at ~64 – meaning the 10038-zip code represents a high potential opportunity with less competition from existing grocery stores and modest population growth. After applying the weights, we arrive at the weighted location score of ~62, the highest in Manhattan (after filtering out two zip codes) and the overall best recommendation for a new Trader Joe's location. As you can see from the plot above, the closest Trader Joe's is in the Lower East Side, so being too close to an existing store is not a concern. Additionally, there are plenty of subway stops that provide easy transportation to the Two Bridges area.

Lastly, Two Bridges is eligible for local tax incentives for new grocery stores within the FRESH zoning guidelines. According to the FRESH boundary specifications, "tax incentives are available for the development, expansion and renovation of full line grocery stores and supermarkets" within the regions specified (see Appendix – Exhibit B). Even though FRESH zoning boundaries were not included in the weighted location score, it is an important factor that can help us decide between multiple options. Because Two Bridges is eligible for the FRESH tax advantage, it ends up as an even better location for a new Trader Joe's store.

II. Thought Process, Assumptions, and Data Sources

The recommendation above is derived from a weighted average location score ("weighted location score" or just "location scores") for each zip code based on borough-normalized 2023 population per store, 2023 population per chain, and 2-year population growth rate from 2021 to 2023.

New York grocery store data was sourced from data.ny.gov, filtered by keywords and popular store names, aggregated by zip code, and then combined with population data by zip code. The choice to normalize the scores by borough as opposed to altogether was made because expected population per store/chain and overall population growth looks very different in densely populated urban areas such as Manhattan when compared to other more suburban boroughs such as Staten Island. For example, the data shows – and we would expect – a higher population per store/chain in those suburban areas since people are more likely to get in their car and drive to stores further away, compared to most individuals in Manhattan without a car. As a result, we normalize the values by each borough between zero and one, accounting for variance.

A higher population per store, higher population per chain, and higher population growth rate will all contribute to a higher location score, which can range from zero to one hundred. Although no zip code received a score of one hundred, high location scores represent the "best" potential zip codes for a new Trader Joe's store, while low location scores represent the "worst" zip codes for a new Trader Joe's store. The reason we use these three variables for identifying a potentially good location for a Trader Joe's store is based on some common assumptions.

First off, we expect there to be less competition among existing grocery stores in zip codes where the population to number of stores ratio is high. This means more residents are available per store, which leads us to believe a new Trader Joe's store will have more shoppers, a higher chance of success, and

ultimately higher revenues. On the flip side, a zip code with a low population per store has more competition, making it less ideal for a new Trader Joe's store.

We also use population per chain, to account for the fact that not all grocery stores are created equal. For example, a small corner bodega commands a much smaller number of customers than a large multi-level Whole Foods. There's no perfect way to filter for this, but we use manual keyword filtering and a dictionary lookup to derive major chain grocery stores. With this in mind, we also compute the population per chain in each zip code to better understand existing competition. By equally weighing the population per store and population per chain in the location score, we account for different competition dynamics in each zip code.

Lastly, we use the 2-year population growth rate as a factor in our location score since we ideally want to place a new store in a growing neighborhood. For example, even if a neighborhood has a high population per store and chain, it may be a poor decision to place a store in an area that has a shrinking population. An increasing population may indicate economic growth and opportunity, whereas a shrinking population may represent a dying neighborhood that people are leaving. It was decided not to weigh this as heavily in the model as population per store and per chain, but it still carries an important weight. The base year used for this calculation is 2021 and the end year is 2023, thus providing us with a two-year growth rate.

The weights for each variable are as follows and can be adjusted in the notebook itself:

- Population per store = 40% weight
- Population per chain = 40% weight
- 2-year population growth rate = 20% weight

In the end, we get a single location score between zero and one hundred for each zip code, representing the strength of the zip code for a new Trader Joe's store (visualized by the map on page one).

Data Sources:

1. NYC Retail Food Stores CSV (retail food stores licensed by the Department of Agriculture and Markets)
2. 2021 NYC Population by Zip Code
3. 2023 NYC Population by Zip Code
4. NYC Subway Stations
5. GeoJSON File for NY Zip Codes
6. GeoJSON File for FRESH Food Stores Zoning Boundaries

III. Further Analysis

If given access to proprietary and/or third-party data sources obtained for a fee, I would have investigated the success of existing Trader Joe's locations by sales and customer volume to build predictive models, leveraged customer demographic/income data of existing stores and zip codes across New York City, and searched for a specific street address using real estate availability across the city for retail stores.

In the current analysis, we understand where existing Trader Joe's stores and competitors are but have no insight into the actual success or lack thereof for those stores. We don't know how many customers are visiting each location, how much they are spending, and ultimately what the store revenues are – all

of which are important metrics for a store's success. We will likely need to get creative with this data since Trader Joe's is not a public company and is also not required to publicly report sales on a store-by-store basis. Therefore, we could leverage alternative data from credit card providers, online reviews about customer experience, or satellite imagery to get an idea of parking lot fullness – but even that may not work for largely pedestrian-based stores in New York City. Either way, this data would allow us to build sophisticated machine learning models using metrics such as the number of customers, revenue, or average spend as the target variable with signals about the store and zip code as inputs (i.e. population per store, population per chain, population growth, proximity to subway stops, etc.). I imagine we could build a neural network, random forest model, or even a simple linear regression to predict the success of a potential store location. As a result, we would have a better method to compare potential store locations on a predicted number of customers or sales volume, as opposed to an arbitrary location score from zero to one hundred that we currently use.

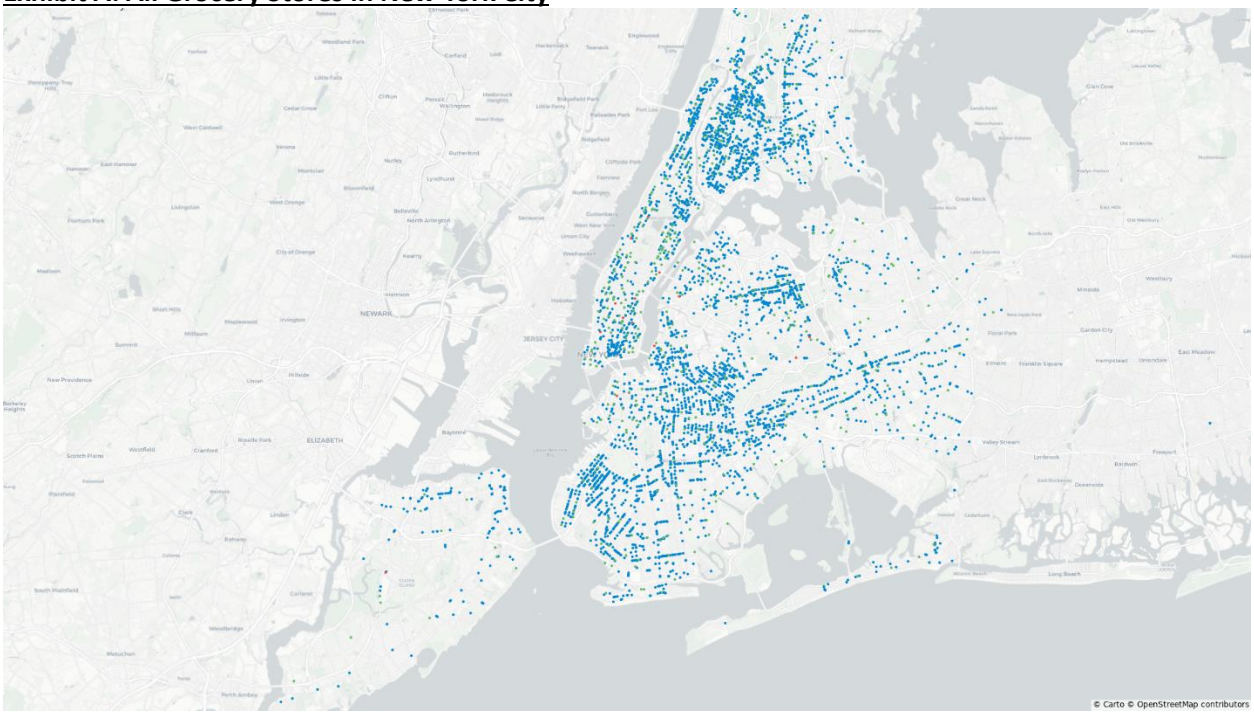
Regardless of whether we get our hands on customer/sales volume and can construct a predictive model, it would be helpful to have further insight into the demographic/income distribution of Trader Joe's customers at each location. While some of this data exists on the web for Trader Joe's as a whole (breakdown of gender, age, race, ethnicity, etc.), it is not complete, nor is it specific to states, zip codes, or stores. Therefore, it wouldn't be that accurate or helpful for our analysis since Trader Joe's is a national brand with nearly 600 locations. By gathering data about the demographic/income distribution of each store in New York City, we would get an idea of the typical customer breakdown for a Trader Joe's store. Afterward, we can cross-reference those findings with each zip code to help select a zip code that has a similar population breakdown. To accomplish this analysis, we may rely on third-party data providers who specialize in customer insights and research. The good news is Trader Joe's business model is highly flexible which is why it's so successful. It appeals to a broad demographic, including low-income households with its affordable prices compared to other chains, while also appealing to higher-income households with its emphasis on quality and freshness. Therefore, the lack of this data currently should not void our analysis.

While we're able to select a store location on a more macro level using zip code, we have limited insight into specific streets or neighborhoods to place a store on a micro level. For example, we have a hard time understanding zoning for residential and commercial areas, and more specifically, what locations are even available for us to place a store. To help with this, we could pull data on available retail properties for sale or rent to determine what specific location to place our store. I attempted using LoopNet to scrape all available ground-level retail storefronts in the city, but many challenges arose since real estate sites like LoopNet do a lot to protect their data from competitive scraping. As we know, however, it's more important to pick a store location based on dynamics at a zip code level, rather than specific streets or available property. On the flip side, using available properties could help Trader Joe's management understand the relative costs of opening a store in different neighborhoods.

None of the options above are meant to suggest the current analysis is unhelpful, but that it's open to improvements with alternative data streams. I searched the web for related data sources to enable such analysis, but the results were limited or too complex to incorporate in the timeframe given. Thus, more time and resources could push this research to the next level by incorporating these alternative data streams. A third-party data provider may have such data available but would need to be explored further.

IV. Appendix

Exhibit A: All Grocery Stores in New York City



Red = Trader Joe's | Green = Large Chain | Blue = Non-Chain

Exhibit B: FRESH Tax-Advantaged Boundaries