ALDI New Location Analysis

Sarah Chang

A. Introduction

A.1. Description & Discussion of the Background

ALDI is a discount supermarket chain that has over 10,000 stores in 20 countries. Discount supermarkets typically offer a reduced product range to enable themselves to gain economies of scale and offer a lower price than conventional supermarkets. ALDI has also been gaining economies of scale through growth and has been rapidly increasing store count worldwide. In fact, in the U.S., ALDI is predicted to become the third largest grocer by store count by 2022.

However, new store openings are a risky and capital-intensive endeavor. To reduce risk, supermarkets will need to analyze an array of data to predict which locations have the greatest chance of success for their brand. As a former Manhattan resident, I am interested in which areas I can expect ALDI stores to open up next within Manhattan. ALDI management would be the target audience for this report.

When selecting a location for a new ALDI store, we can assume:

- A high population around an ALDI store will increase sales.
- Opening in an area without competitors will make it easier to be profitable.
- As ALDI's main customer base are price-conscious shoppers looking for cheap goods, ALDI will most likely thrive in low income areas.

Taking into account the above assumptions, I decided to look for neighborhoods in Manhattan without competition that had a low income and a high population.

A.2. Data Description

The below data can be used to address the problem:

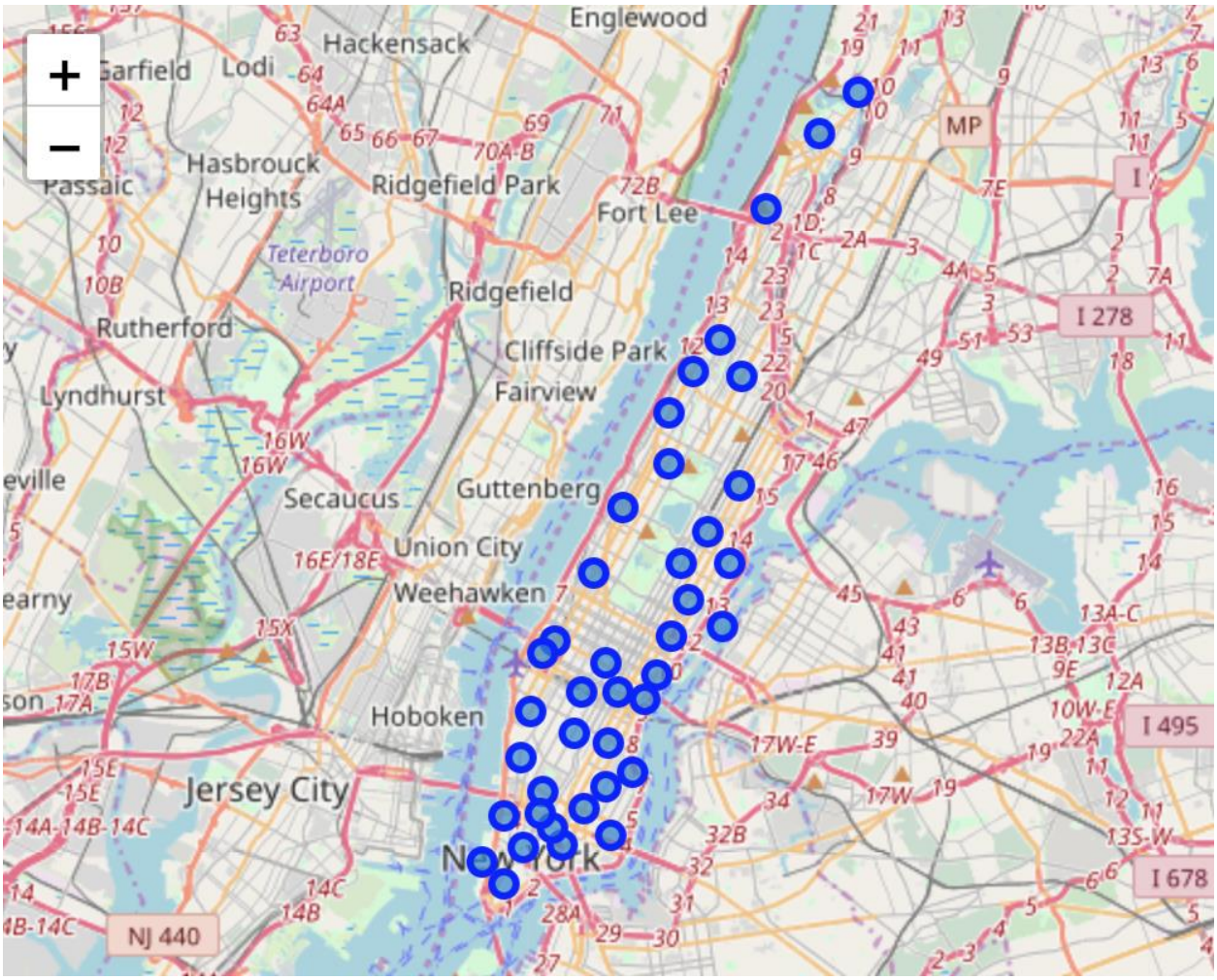
- I will use Foursquare location data to determine areas in Manhattan that have no competition for ALDI. There have to be no discount supermarkets nearby. This data will be used to explore Manhattan. I will use the explore function to determine the most common venue categories in each neighborhood. I will use the Folium library to visualize the NYC neighborhoods.

- I will also use the following dataset, which contains the 5 NYC boroughs, the neighborhoods that are in each borough, and the latitude and longitude of each neighborhood: https://geo.nyu.edu/catalog/nyu_2451_34572

- I will use demographic data to determine areas with high population density and low median household income: http://www.city-data.com/nbmaps/neigh-New-York-New-York.html.

B. Methodology

I used the Python Folium library to visualize the different neighborhoods of Manhattan by creating a map of Manhattan with the neighborhoods superimposed on top. I used the latitude and longitude values of each neighborhood to generate the below map:



I then used the Foursquare API to explore the Manhattan neighborhoods. I set the limit at 100 venues and the radius at 500 meters for each neighborhood based on each neighborhood's latitude and longitude. Below is the head of the resulting dataframe:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Blink Fitness Riverdale	40.877147	-73.905837	Gym

I then grouped the rows in the dataframe by neighborhood and took the mean of the frequency of occurrence of each category. Below is the head of the dataframe (please note that this is only showing some of the columns since there are a total of 343 columns, venue categories):

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store
0	Battery Park City	0.0	0.0	0.0	0.000000	0.010101	0.0	0.0	0.0	0.0	0.0	0.000000	0.00	0.0
1	Carnegie Hill	0.0	0.0	0.0	0.000000	0.010000	0.0	0.0	0.0	0.0	0.0	0.000000	0.01	0.0
2	Central Harlem	0.0	0.0	0.0	0.065217	0.043478	0.0	0.0	0.0	0.0	0.0	0.021739	0.00	0.0
3	Chelsea	0.0	0.0	0.0	0.000000	0.030000	0.0	0.0	0.0	0.0	0.0	0.010000	0.00	0.0
4	Chinatown	0.0	0.0	0.0	0.000000	0.040000	0.0	0.0	0.0	0.0	0.0	0.000000	0.00	0.0

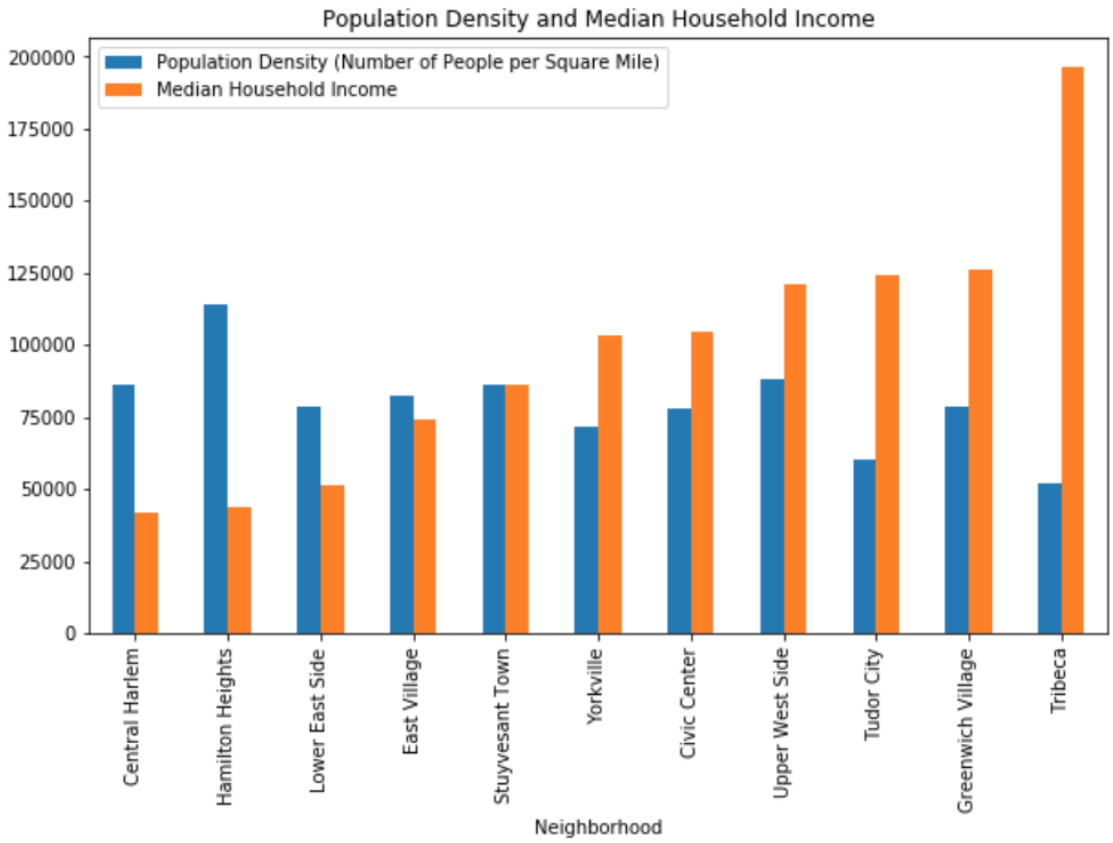
I then dropped neighborhoods from the dataframe that had a mean frequency of occurrence greater than zero for Big Box Store, Discount Store, Grocery Store, and Supermarket because these venue categories are potential competitors for ALDI and I am only looking at neighborhoods which do not have any competition.

I found the population density and median household income for these remaining 11 neighborhoods from **http://www.city-data.com/nbmaps/neigh-New-York-New-York.html.**

I sorted the remaining 11 neighborhoods in ascending order based on median household income because ALDI's main customer base are price-conscious shoppers looking for cheap goods, so it is important to be able to easily identify the neighborhoods with low median household income:

	Neighborhood	Population Density (Number of People per Square Mile)	Median Household Income
0	Central Harlem	85964	41680
1	Hamilton Heights	114298	43673
2	Lower East Side	78448	51649
3	East Village	82502	74265
4	Stuyvesant Town	85975	86345
5	Yorkville	71597	103234
6	Civic Center	77876	104375
7	Upper West Side	88376	121032
8	Tudor City	60507	124047
9	Greenwich Village	78728	126368
10	Tribeca	52160	196692

I created a bar chart using this sorted data to easily compare between population density and median household income for each neighborhood:



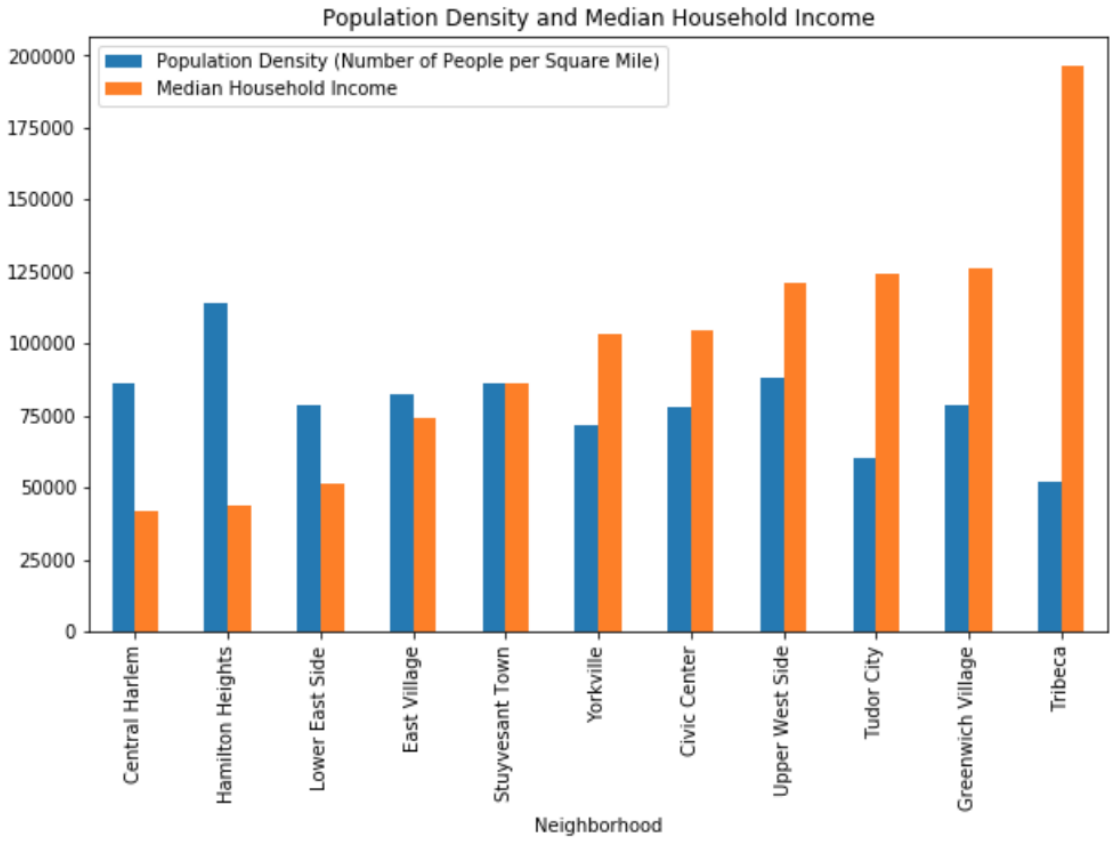
C. Results

Exclusion of all neighborhoods that had a known competitor further narrowed the neighborhood list to 11 number of neighborhoods.

The results from the 11 remaining neighborhoods can be found below, sorted from low median household income to high median household income.

	Neighborhood	Population Density (Number of People per Square Mile)	Median Household Income
0	Central Harlem	85964	41680
1	Hamilton Heights	114298	43673
2	Lower East Side	78448	51649
3	East Village	82502	74265
4	Stuyvesant Town	85975	86345
5	Yorkville	71597	103234
6	Civic Center	77876	104375
7	Upper West Side	88376	121032
8	Tudor City	60507	124047
9	Greenwich Village	78728	126368
10	Tribeca	52160	196692

The graph below makes visualizing the comparison the data much easier, and allows a quick comparison between population density and median household income for each neighborhood.



D. Discussion

From the data, we can see that of the 11 neighborhoods without competition in Manhattan, there is a negative correlation between population density and median household income.

All 5 neighborhoods with above 75000 people per square mile were also the 5 poorest neighborhoods.

Hamilton Heights and Central Harlem are the two most favorable locations for an ALDI. Hamilton Heights has the biggest population density and the second lowest median income, while Central Harlem has the lowest median income and the second highest population density.

Out of the two, Hamilton Heights seems to be the best choice for an ALDI. While the difference in median income is marginal, the increase in population density is significant.

Hamilton Heights would be an ideal store location for an ALDI. There is a high amount of people that the ALDI can serve, with no competitors nearby, and a potential customer base that will react well to ALDI's focus on price.

While the data points to Hamilton Heights as the best choice, Lower East Side, East Village, Stuyvesant Town and the aforementioned Central Harlem all seem to match the assumptions that we set out with. If the locations are sufficiently spread apart, ALDI may wish to plan stores in each of these area.

One limitation of this study was that it didn't calculate the ratios between population and competitors to see if the higher population in an area outweighs the disadvantage of having a competitor. It also didn't account for total population in an area, using population density instead. If the area of neighborhoods varied widely, this could skew results. Furthermore, all competitors were treated equally. In reality, a Whole Foods, for example, would not target most ALDI customers and if the data was more detailed, these less direct competitors could be excluded from the results to highlight more areas where ALDI may thrive.

There are many future avenues for additional study also. Secondary factors for choosing a store location could also be included in the analysis. What is the availability of store staff in each neighborhood? What pay rates are required to attract staff in these neighborhoods? What are the crime rates for theft and vandalism in each neighborhood? What are land prices and land availability in these neighborhoods? How far do people have to travel out of their neighborhood to get to the nearest supermarket? All of these questions would help ALDI pick the perfect new location for a store. Central Harlem or Hamilton Heights may seem like great places for ALDI, but if high crime rates cut into margins, then other neighborhoods could generate better returns.

F. Conclusion

In conclusion, discount supermarkets like ALDI can use data to find ideal store locations. Discounters look for locations with low-income households, low-competition, and many potential customers living nearby. Out of the 11 Manhattan neighborhoods without competition, Hamilton Heights was found to be the best potential location for a new ALDI, followed by Central Harlem. Future research with more detailed data and more variables analyzed could provide a more complete view of the best locations for ALDI, to aid decision making.