Coursera Capstone
Battle of the neighborhoods - building a hotel in Brooklyn, New York City

By: Zachary Kolander

Introduction

In New York City, the borough of Manhattan is an important location for many groups of individuals and businesses. Aside from being a financial hub for thousands of business travelers each year, it is also a prime destination for millions of tourists traveling to New York City annually. This has resulted in the cost of hotels in Manhattan being exponentially higher than that of the neighboring boroughs.

Background and Business Problem

This purpose of this project is to find the best, or optimal, neighborhood in Brooklyn, New York to build a hotel for vacationers and business people alike, who choose not to stay in the more expensive hotels in Manhattan. Specifically, the analysis will determine the closest neighborhood in Brooklyn relative to Manhattan, with the lowest number of existing hotels. The results of this report can be used as a reference by hotel stakeholders who are interested in building a hotel in Brooklyn for Manhattan bound travelers.

One option to alleviate this problem for travelers is to build a hotel in the less expensive boroughs of Bronx, Brooklyn, or Queens. From there, travelers can book a hotel in a neighboring borough and commute to downtown Manhattan. It is important to build a hotel as close to downtown Manhattan as possible, as to reduce travel time and cost.

For a hotel builder, knowing the optimal neighborhood in a neighboring borough to build the hotel is key to solving these problems. By knowing the best neighborhood to build the hotel, Manhattan bound travelers can reduce booking costs, travel time, and overall travel cost. Additionally, knowing the number of existing hotels within the neighborhood in question and where they are located can minimize competition from other hotels in that area.

This project will focus on finding the optimal neighborhood in the Brooklyn borough

Data Acquisition

In order to address this problem, we will need to find the neighborhood that is closest to downtown Manhattan with the minimum number of hotels. We will need to acquire data on the neighborhoods and hotels in New York City. Specifically, we will need location data on all neighborhoods in the borough of Brooklyn and the total number of hotels in each neighborhood within that borough. Data was obtained from two locations.

Neighborhood data was acquired from New York University (NYU)

url: https://geo.nyu.edu/catalog/nyu_2451_34572

The neighborhood data downloaded from the NYU website comes in JSON format and has the coordinates of all the neighborhoods in all five New York City boroughs.

Hotel data was acquired from Four-Square and venue platform.

url: www.foursquare.com

Data Exploration

The neighborhood data downloaded from the NYU website comes in JSON format and has the coordinates of all neighborhoods in all five boroughs in New York city. Since our focus is on the boroughs of Brooklyn and Manhattan, the original dataset was cleaned and reduced to just the neighborhoods in Brooklyn. The following data frame consisting of the Borough, Neighborhood, Latitude and Longitude was created.

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

Data Exploration cont’d

The hotel data for each Brooklyn neighborhood was obtained from the Four-Square location platform using the coordinates of each neighborhood. A data frame consisting of the hotels in each Brooklyn neighborhood was created.

Neighborhood	Hotel
Bay Ridge	1
Borough Park	1
Broadway Junction	2
Downtown	1
Fort Hamilton	1
Fulton Ferry	1
Greenpoint	1
North Side	1
Sheepshead Bay	1

Methodology

As previously mentioned, our goal is to find the neighborhood that's closest to midtown Manhattan with the minimum number of hotels. In the section above we acquired and cleaned the necessary data needed to accomplish this goal. In this section I applied various data science techniques and methods to the data to determine the optimum Brooklyn neighborhood. The following task was performed.

• I used the Python Folium map library to visualize and analyze the layout of the Brooklyn neighborhoods relative to midtown Manhattan.

• I used the Python Folium map library to visualize and analyze the layout of the Brooklyn hotels relative to midtown Manhattan

• Use Matplotlib to create bar charts of the hotels in each Brooklyn neighborhoods for more analysis

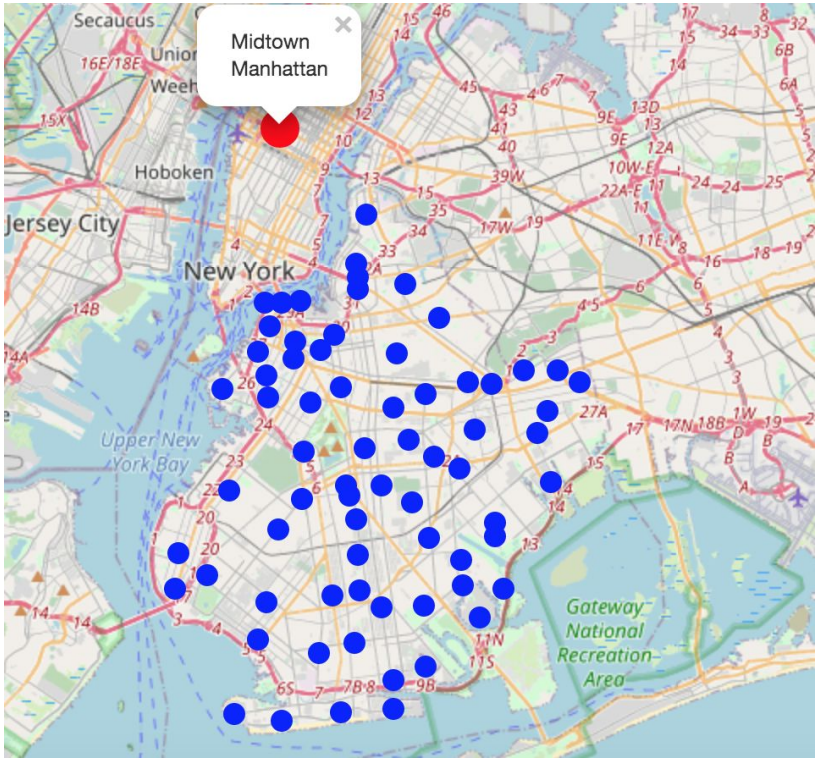
• Use the Haversine formula to calculate the distance from each Brooklyn neighborhood to midtown Manhattan. This distance data was added to the final feature set for clustering

• Combine and normalize all relevant features for clustering

• Apply K-means clustering method to the feature set

• Use the Python Folium map library to the visualize and analyze the clusters

Data Visualization



The map to the left shows the overall layout of the Brooklyn neighborhoods in blue with respect to midtown Manhattan in red. As we can see there are a few neighborhoods along the border of Brooklyn and Manhattan that could be good candidates to build a hotel due to their close distance to Manhattan. However, other factors such as the number of hotels already existing in these neighborhoods needs to be considered. With that said let's create a map of the hotels throughout Brooklyn.

Data Visualization cont’d

The map to the left show the current hotels in the Brooklyn borough. We can see that there are a group of hotels that are fairly close to Manhattan, but have some distance between each other.

Data Visualization cont’d

The bar graph to the left does not seem to help as much as the maps, as it appears that the most hotels that a neighborhood in Brooklyn has is 1 hotel. We are ideally looking for the neighborhood in Brooklyn that is closest to midtown Manhattan with the fewest amount of hotels. Seeing that the closest neighborhoods each have 1 hotel does not help us to narrow down which neighborhoods to look into very much. We can see that further analysis is needed.

Modeling

Thus far, I have used the visualization tools Folium and Matplotlib to gain some useful insight into the data. Now let use some machine learning tools to improve our results. I applied the K-means machine learning algorithm to a feature set formed from both the neighborhood and hotel datasets. Our feature set consists of the following.

	Borough	Neighborhood	Latitude	Longitude	Distance	Hotel
0	Brooklyn	Bay Ridge	40.625801	-74.030621	9.299109	1.0
1	Brooklyn	Bensonhurst	40.611009	-73.995180	10.070021	0.0
2	Brooklyn	Sunset Park	40.645103	-74.010316	7.788030	0.0
3	Brooklyn	Greenpoint	40.730201	-73.954241	2.590410	1.0
4	Brooklyn	Gravesend	40.595260	-73.973471	11.184066	0.0

Modeling cont’d

	Borough	Neighborhood	Latitude	Longitude	Distance	Hotel	Cluster Labels
0	Brooklyn	Bay Ridge	40.625801	-74.030621	9.299109	1.0	0
1	Brooklyn	Bensonhurst	40.611009	-73.995180	10.070021	0.0	0
2	Brooklyn	Sunset Park	40.645103	-74.010316	7.788030	0.0	4
3	Brooklyn	Greenpoint	40.730201	-73.954241	2.590410	1.0	1
4	Brooklyn	Gravesend	40.595260	-73.973471	11.184066	0.0	2

This feature set of only numerical data was normalized and the K-means clustering algorithm applied with K=5. Results of the clusters were merged with the original neighborhood dataset.

The above table doesn’t give us much information in its current form. To make more sense of the clusters I applied some statistical method.

Statistical Methods

I group the clusters from the previous table and took the mean of each cluster to get the following result.

	Latitude	Longitude	Distance	Hotel	Cluster Labels
0	40.625801	-74.030621	9.299109	1.0	0
1	40.611009	-73.995180	10.070021	0.0	0
2	40.645103	-74.010316	7.788030	0.0	4
3	40.730201	-73.954241	2.590410	1.0	1
4	40.595260	-73.973471	11.184066	0.0	2

Statistical Methods cont'd

	Borough	Neighborhood	Latitude	Longitude	Distance	Hotel	Cluster Labels
3	Brooklyn	Greenpoint	40.730201	-73.954241	2.590410	1.0	1
15	Brooklyn	Williamsburg	40.707144	-73.958115	3.791646	0.0	1
18	Brooklyn	Brooklyn Heights	40.695864	-73.993782	4.209000	0.0	1
38	Brooklyn	Clinton Hill	40.693229	-73.967843	4.525600	0.0	1
40	Brooklyn	Downtown	40.690844	-73.983463	4.559319	1.0	1
49	Brooklyn	East Williamsburg	40.708492	-73.938858	4.248871	0.0	1
50	Brooklyn	North Side	40.714823	-73.958809	3.302832	1.0	1
51	Brooklyn	South Side	40.710861	-73.958001	3.564233	0.0	1
61	Brooklyn	Fulton Ferry	40.703281	-73.995508	3.704304	1.0	1
62	Brooklyn	Vinegar Hill	40.703321	-73.981116	3.711685	0.0	1
65	Brooklyn	Dumbo	40.703176	-73.988753	3.697199	0.0	1

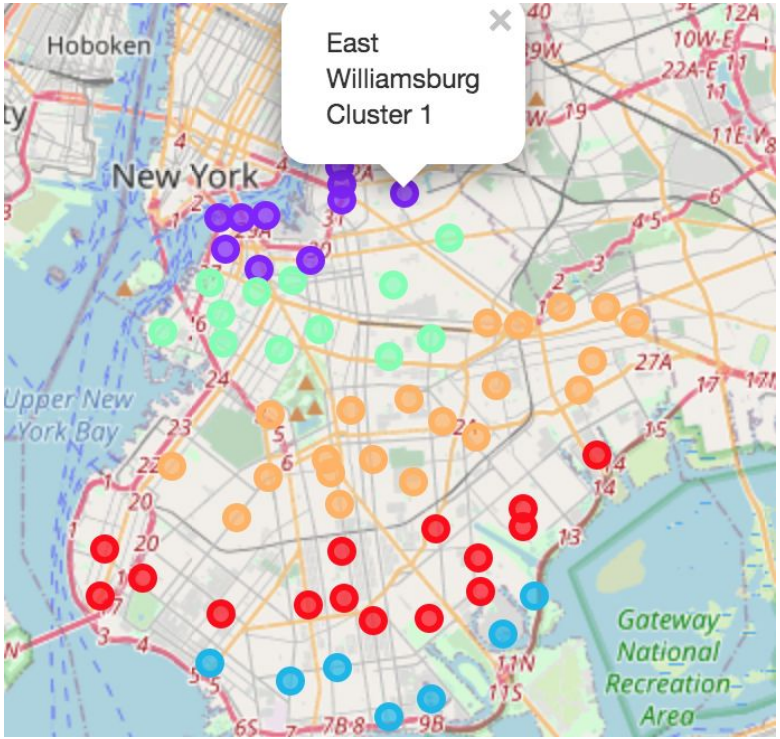
From this table we can now see which cluster of neighborhoods are ideal to build a hotel in Brooklyn. Going through the table I narrowed the search down to neighborhoods in cluster 1 and cluster 3. Cluster 1 has the closest neighborhoods to Manhattan. It does not have the most hotels, but does not have the fewest either. Cluster 3 has the second closest neighborhoods to Manhattan with slightly fewer hotels. With the amount of hotels being relatively similar, cluster 1 fits the requirements better because it is closer to midtown Manhattan.

Results

	Borough	Neighborhood	Distance	Hotel
51	Brooklyn	South Side	3.564233	0.0
65	Brooklyn	Dumbo	3.697199	0.0
62	Brooklyn	Vinegar Hill	3.711685	0.0
15	Brooklyn	Williamsburg	3.791646	0.0
18	Brooklyn	Brooklyn Heights	4.209000	0.0
49	Brooklyn	East Williamsburg	4.248871	0.0
38	Brooklyn	Clinton Hill	4.525600	0.0

In summary we determined the Brooklyn neighborhoods closest to Manhattan with the minimum amount of hotels, by clustering the number of hotels in each Brooklyn neighborhood and the distances from each Brooklyn neighborhood to Manhattan. Our analysis found that there are 7 Brooklyn neighborhoods within 5 miles of Manhattan with no hotels. All 7 neighborhoods are potential candidates to build a hotel and further investigation is needed to determine the ideal neighborhood and a prime building location within that neighborhood. The recommended list of Brooklyn neighborhoods are shown in the table to the left.

Results cont’d



Looking at the map the clusters of the neighborhoods closest to Manhattan cluster 1 is in color purple.

Discussion

The results obtained from my analysis is just the starting point for hotel companies interested in finding the optimum Brooklyn neighborhood to build a hotel. The feature set used in my analysis to determine the optimum neighborhood or neighborhoods was limited to neighborhood location data(Lat, long), the distance from each Brooklyn neighborhood to Manhattan(distance) and the number of existing hotel in each neighborhood(Hotel). Other dataset such as demographic, property value, taxi fares from each neighborhood, mass transit, and crime can be applied to refine the list of neighborhoods. Once a neighborhood has been picked then another dataset also may be needed to determine a suitable location within that neighborhood.

Conclusion

The purpose of this project is to find the Brooklyn neighborhoods or neighborhoods closest to Manhattan with the lowest number of existing hotels. The results of this analysis can be used by hotel companies to build hotels in Brooklyn neighborhoods that are an alternative to the more expensive hotels in Manhattan. Manhattan bound business travelers and vacationers who can't afford the high hotel cost in Manhattan can then reside in Brooklyn hotels and commute to Manhattan. After fetching data from several data sources and processing them into a clean data frame, I then apply the K-Means clustering algorithm and picked the cluster that was closest to Manhattan with the lowest number of existing hotels. The final decision of the optimal hotel location will be made by the stakeholders based on specific characteristic of each neighborhood in the cluster. As I mentioned before, dataset on demographics, property value, traffic, mass transit, and crime can be applied to each neighborhood being considered.

Thank you!