Evaluation of Brooklyn and Manhattan Neighborhoods
Recommendations for Moving

Margaret Mayer
June 4, 2020

1. Introduction

1.1 Background

The boroughs of Brooklyn and Manhattan have many neighborhoods, each of which tends to have its own unique atmosphere and culture. Much of this can be characterized by the type of venues in the neighborhood which contribute to the lifestyle for the residents in the area. The rental cost to live in a given neighborhood also varies widely, but the conventional wisdom is that Brooklyn is more affordable than Manhattan.

1.2 The Problem

The overarching question to be addressed by this project is can one find a neighborhood in Manhattan with similar characteristics to one in Brooklyn that is equally affordable. More specifically, the client for this project is my daughter. Currently she works from home, in the Clinton Hill neighborhood of Brooklyn, and has an office she will return to at some point in the Financial District neighborhood of Manhattan. She is currently concerned about using public transportation given the pandemic and so a location closer to the office might be desirable. So the initial questions to be answered by this project are:

1. Are there neighborhoods nearer to her office that are similar in terms of popular venues?

2. Do any of these similar neighborhoods also have apartments in the price range she is targeting?

1.3 The Audience

The most direct stakeholder in this research is my daughter. Since she is a computer engineer, it will be useful to her to have a data analytics approach to add her existing knowledge of NYC. I will also check in as the results emerge to see if there are additional issues that she would like explored.

More generally, the classification scheme I will use could provide anyone considering a move in NYC a background on which to base their decision.

2. Data Sources and Preprocessing of Data

Three data sources were utilized for this analysis.

2.1 New York Neighborhood Location Data

A json file of data identifying the longitude and latitude of each NYC neighborhood was obtained from New York University's Data Services Spatial Data Repository (https://cocl.us/new_york_dataset). To simply extract the data needed to analyze just the boroughs of Brooklyn and Manhattan, I downloaded the json file information for all five boroughs as a csv file and then edited/cleaned that file to include only the two boroughs of interest.

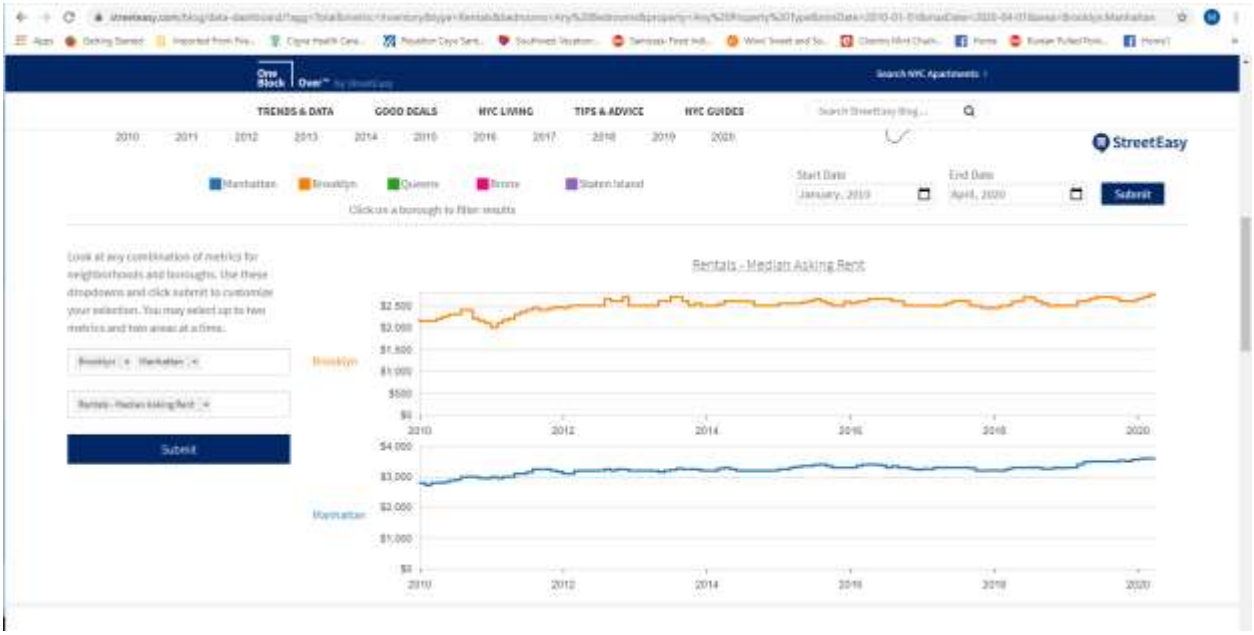
2.2 Popular Venue Data

The Foursquare API (foursquare.com) was queried to extract popular venue information for each neighborhood. In particular I used an “explore” call to identify up to 100 venues per neighborhood by location, name, and category.

2.3 Average Apartment Rental Price Data

Data concerning the average rental price for each neighborhood was collected from One Block Over™ by StreetEasy, a real estate search site that maintains a dashboard with downloadable data updated monthly.

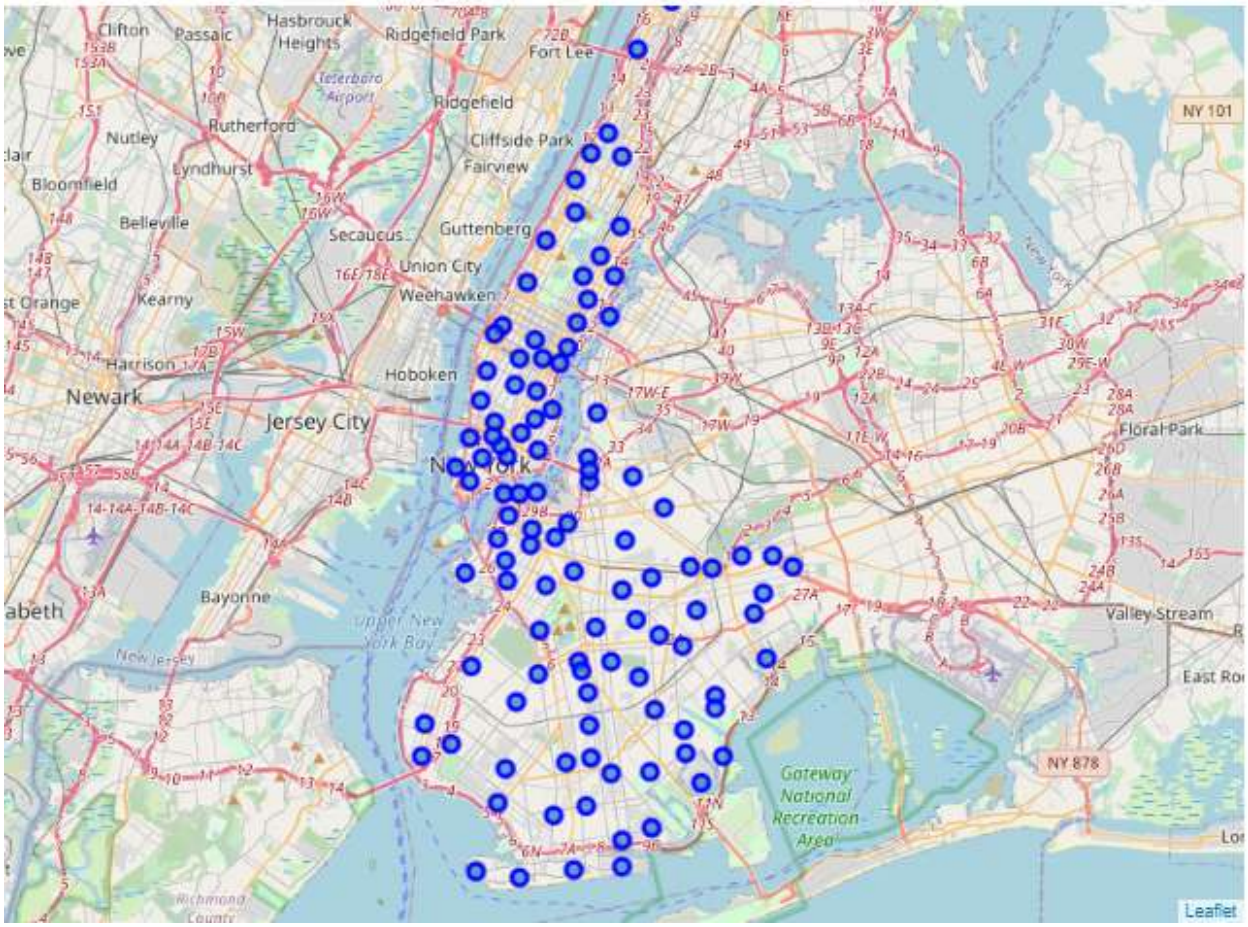
https://streeteasy.com/blog/data-dashboard/?agg=Total&metric=Inventory&type=Rentals&bedrooms=Any%20Bedrooms&property=Any%20Property%20Type&minDate=2010-01-01&maxDate=2020-04-01&area=Brooklyn,Manhattan



3. Methodology

3.1 Exploratory Data Analysis

The first requirement for the exploration was to identify each of the neighborhoods in the two boroughs of interests using the NYU location data and to plot these on a map of New York.



Each circle represents the geographical center of a neighborhood in Manhattan or Brooklyn. The next step was to create a function that would iteratively call the Foursquare API for an exploration of each neighborhood that produced information about up to 100 venues per neighborhood. The specific data required for the clustering analysis was then extracted into a pandas dataframe (table) that identified 5837 venues in 386 unique venue categories.

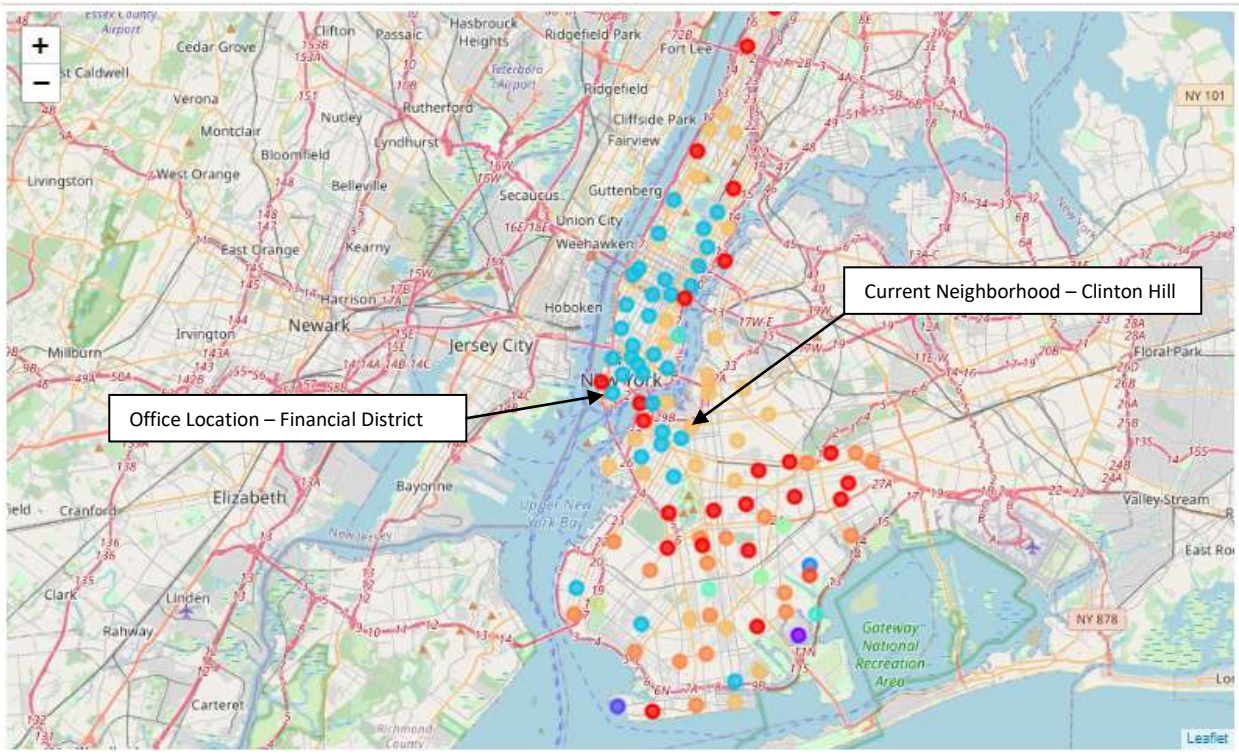
One hot encoding was then used to transform the categorical venue variables in the table to data usable for the k-clustering algorithms by generating a percentage of venue category type per neighborhood for each neighborhood. This data was then grouped into a new dataframe identifying the most common types of venues in each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bath Beach	Pharmacy	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Pizza Place	Bubble Tea Shop	Cantonese Restaurant	Gas Station	Rental Car Location	Coffee Shop
1	Battery Park City	Park	Coffee Shop	Hotel	Gym	Memorial Site	Beer Garden	Shopping Mall	Gourmet Shop	Food Court	Plaza
2	Bay Ridge	Spa	Italian Restaurant	Chinese Restaurant	Pizza Place	Bar	Bagel Shop	American Restaurant	Greek Restaurant	Playground	Grocery Store
3	Bedford Stuyvesant	Bar	Coffee Shop	Pizza Place	Café	Bagel Shop	Boutique	Gift Shop	Gourmet Shop	BBQ Joint	Thrift / Vintage Store
4	Bensonhurst	Pizza Place	Dessert Shop	Sushi Restaurant	Italian Restaurant	Donut Shop	Bakery	Ice Cream Shop	Park	Pet Store	Russian Restaurant

This became the basis for grouping neighborhoods based on the similarity of venue categories (most common venue types) in the neighborhood.

3.2 K-Means Clustering

k-Means Clustering is a type of machine learning that allows segmentation of data into groups with similar characteristics. In this case we used similarity of most common venue category to classify Brooklyn and Manahattan neighborhoods into similar segments. Using the KMeans function from the sklearn.cluster library we identified 12 similar “clusters” each cluster identified by a different color on our updated neighborhood map.



Clinton Hill is in Cluster 9, the Financial District is in Cluster 4.

The question we want to answer is if there are other neighborhoods in Cluster 9 that are closer than Clinton Hill to the Financial District.

3.3 Location and Price Analysis

By visual analysis, we identified four neighborhoods that meet the criteria of being in Cluster 9 and being physically closer to our client's office location in the Financial District. These are:

- East Village (Manhattan)
- Gramercy (Manhattan)
- Cobble Hill (Brooklyn)
- Vinegar Hill (Brooklyn)

Accordingly, we bring in the median rental price in for each of these neighborhoods and for Clinton Hill.

Neighborhood	Latitude	Longitude	Cluster	Average Rental	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
East Village	40.727847	-73.982226	9	$3150	Bar	Mexican Restaurant	Cocktail Bar	Coffee Shop	Pizza Place	Ice Cream Shop	Ramen Restaurant
Gramercy	40.737210	-73.981376	9	$3900	Bagel Shop	Bar	Coffee Shop	Pizza Place	Grocery Store	Mexican Restaurant	Playground
Cobble Hill	40.687920	-73.998561	9	$3500	Coffee Shop	Playground	Pizza Place	Bar	Yoga Studio	Italian Restaurant	Ice Cream Shop
Vinegar Hill	40.703322	-73.981116	9	$3900	Food Truck	Coffee Shop	Art Gallery	Café	Wine Shop	Ice Cream Shop	Factory
Clinton Hill	40.693229	-73.967843	9	$2995	Italian Restaurant	Pizza Place	Mexican Restaurant	Wine Shop	Thai Restaurant	Restaurant	Japanese Restaurant

4. Answers

4.1 Results

Among the neighborhoods under consideration, based on the kMeans Clustering model, the East Village in Manhattan is the best answer to the questions posed in the problem statement of this report.

1. Are there neighborhoods nearer to her office that are similar in terms of popular venues?

2. Do any of these similar neighborhoods also have apartments in the price range she is targeting?

4.2 Discussion

The median rental price in the East Village is closest to the median rental price in Clinton Hill among the neighborhoods that are similar to Clinton Hill and nearer (potentially in walking distance) to my daughter's office in the Financial District. An additional requirement for a new apartment revealed by our client is having an elevator. With this information, a further examination of the real estate data base used suggested that one downside could be that many of the apartment buildings in the East Village are older and may not have elevators.

4.3 Conclusions and Recommendations

The recommendation of the East Village as a target for the new apartment search was positively received based on two parameters:

- Liking the neighborhood for its prevalence of good/interesting restaurants.
- Being surprised that the average rent is not that much higher than the current neighborhood, Clinton Hill.

As the move will probably not take place until at least six months from now, the recommendation would be to revisit the analysis closer to the target date and to begin to examine the apartment inventory in the East Village to find a building that fills all the specific requirements at the right price.