Introduction

Client A has been living in Flushing, Queens for most of his life. Client A has recently decided that he wants to move to Brooklyn and be closer to his work. Client A however, wants to make sure that the neighborhood in his new home is like that of his current one in Flushing, Queens. Client A does not care about the demographics, but solely the vast number of restaurants/bars he was accustomed to in Flushing.

We will be using location data from Foursquare to find the store types in his current neighborhood and the neighborhoods in Brooklyn. We will find the stores around the many neighborhoods in Brooklyn and compare it to that of Client A's current neighborhood (Flushing). We will cluster the neighborhoods in Brooklyn according to the store types. The Brooklyn neighborhoods that has the greatest number of Restaurants & Bars will influence Client A's new address.

Data

The data that will be used in this notebook will come from FourSquare's location data, a json file containing New York's location data containing geographical coordinates, the boroughs, and neighborhoods. FourSquare location data will be used to get a list of venue categories. New York's location data will be used to find out the neighborhood names in Brooklyn.

Methods

Data was collected from the New York JSON file inorder to get neighborhood data. First the neighborhood, Flushing was found. Using the Flushing data, a FourSquare request was done inorder to get venue information near a .5 mile radius of Flushing. The neighborhood data on Brooklyn was also collected, this time every neighborhood of brooklyn was found. Using FourSquare, a .5 radius search was found on the neighborhoods in brooklyn. The data from the brooklyn neighborhood was turned into a dataframe. The KMeans algorithm was used on the mean amount of categories in each unique brooklyn neighborhood. The clustering was set to 4, and thus 4 clusters of brooklyn neighborhoods were created. The neighborhoods that were divided into 4 clusters were plotted on a map to visualize the distribution of neighborhoods. The 4 clusters were further analyzed and the number of restaurants and bars in each cluster was found. The 4 clusters were then put into a bar graph to further analyze the data.

Number of Restaurants & Bars
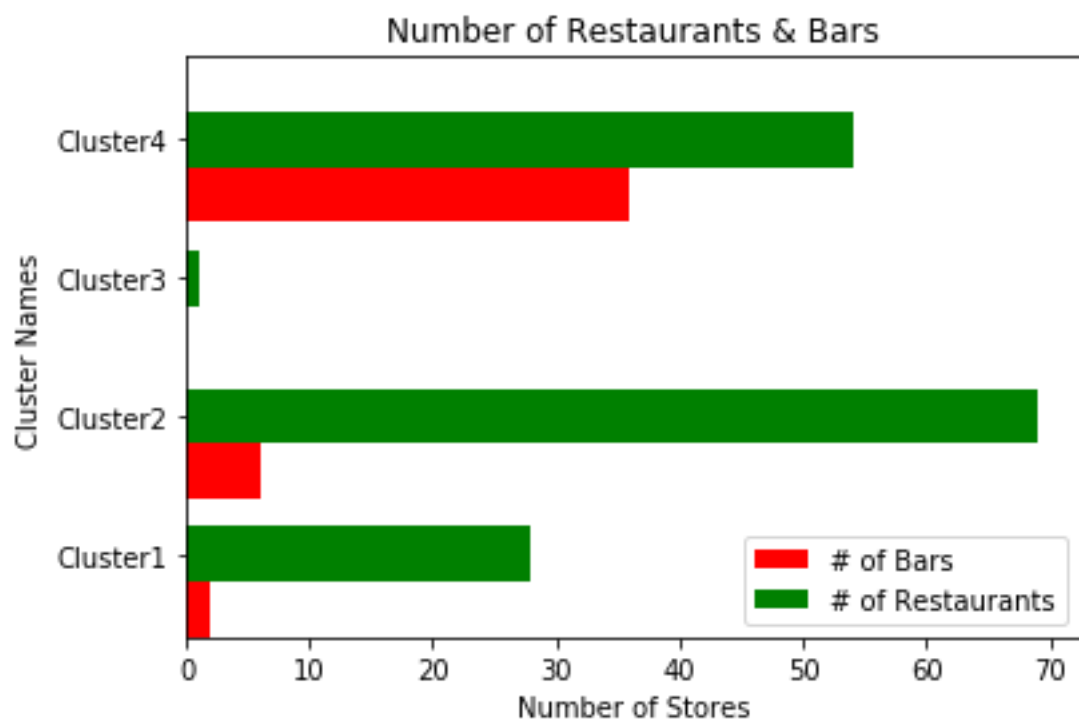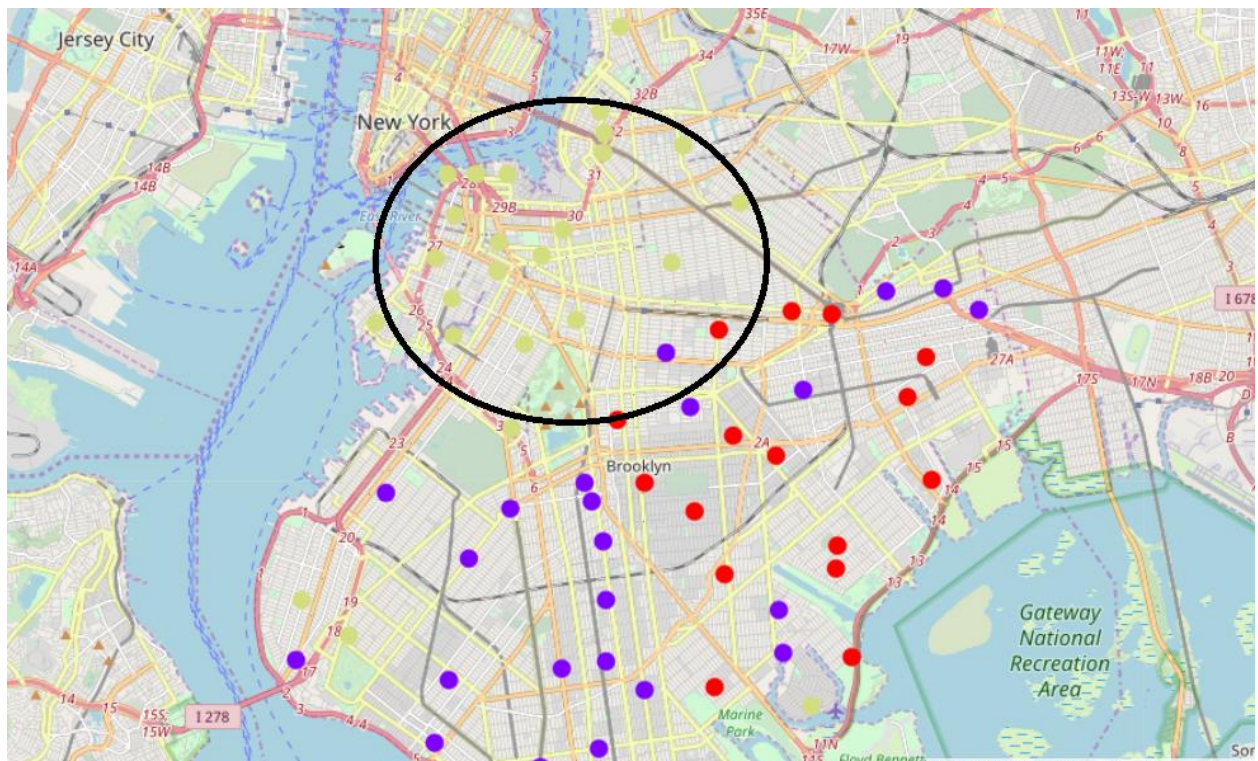
Cluster	# of Bars	# of Restaurants	Total
1	2	28	30
2	6	69	75
3	0	1	1
4	36	54	90

Results

The results showed that Cluster 1 had a total of 16 different neighborhoods. Cluster 2 had a total of 27 different neighborhoods. Cluster 3 had a total of 1 neighborhood. Cluster 4 had a total of 26 neighborhoods. Cluster 1 is located mostly on the west of Brooklyn. Cluster 2 is located mostly in the south and eastern parts of Brooklyn. Cluster 3 is in the south by itself in the neighborhood called Sea Gate. Cluster 4 is located mostly on the north of Brooklyn.

From the function Category_Occurrences (to calculate the number of a occurrence of a category in a cluster) it was found out that Cluster 1 had a total of 30 bars & restaurants. 2 coming from bars and 28 coming from restaurants. Cluster 2 had a total of 75 bars & restaurants. 6 coming from bars and 69 coming from restaurants. Cluster 3 had a total of 1 bars & restaurants. 0 coming from bars and 1 coming from restaurants. Cluster 4 had a total of 90 bars & restaurants. 36 coming from bars and 54 coming from restaurants.

Discussion

From the results we can see that the the neighborhoods in Brooklyn were segmented into different clusters inorder to give Client A a variety of locations. Clustering the neighborhoods using KMeans may not have been the best way to solve the problems. However, clustering may have been beneficial to Client A, so that he could choose from more than one neighborhood. The data shows that the neighborhoods Client A will enjoy is mostly in the northern parts of Brooklyn, near the tip of Manhattan. It was also noted that the FourSquare location data did not give the venue names of every single venue in Brooklyn. There was a limit of just 100 venues per neighborhood.The limit in the number of venues was a critical part in determining the categories of stores. It was noted that there were no asian restaurants in Brooklyn, but this is due to the limit in searches from the free FourSquare account. Although the areas with many bars and restaurants were found they were not the same types of restaurants and bars as in Flushing. The data should be further analyzed into giving Client A a more specific area to choose, this can be done by limiting the location radius and finding more specific types of venues.

Conclusion

The analysis of the neighborhoods in Brooklyn was conducted so that Client A can be closer to his work. The data from this analysis showed that Client A should move towards the northern part of Brooklyn. The data also showed that Brooklyn is filled with many restaurants and bars, and Client A will have a fun time transitioning to his new neighborhood. Because of the limit in the amount of searches, unfortunately there was a problem in locating every venue. The categories of stores in Flushing and Brooklyn are very different. Client A can also refine his housing research by looking into the housing costs. In conclusion, if money is not an issue, the northern parts of Brooklyn can be a good place for Client A.