

# **Exploring the neighborhoods of Brooklyn based on venues and rent**

Utsav Dihingia

January 10, 2021

## **1. Introduction**

### **1.1 Background**

You find yourself in a position that you have to move to a neighborhood in Brooklyn, NY but since you are not a native of New York you don't have a slightest idea of which neighborhood you should to move to. This report tackles with the simple idea of choosing the best neighborhood for you depending on your budget and taking in the factor of availability of resources around the neighborhood be it a neighborhood with delicacies of multiple ethnicities or a one with more accessible to parks and leisure activities.

Here we will make analysis on multiple neighborhoods and cluster them to fit for multiple individuals that have different needs and criteria's for choosing one.

### **1.2 Interested audience**

The target audience for such a project is quite specific as to it is centered around Brooklyn Neighborhood but individuals can take this as an example and build similar cases with region of their choice

## **2. Data**

### **2.1 Data Sources**

The dataset on the New York neighborhoods and their corresponding coordinates will be mainly collected from the dataset that is readily available in the IBMDeveloperSkillsNetwork portal.

We will be collecting data mainly from the Foursquare API through which we will get insights on the venue available in the neighborhoods

We will also be taking in the average rent market trend data from the website Rentcafe which will further help in narrowing down neighborhoods.

From the IBMDeveloperSkillsNetwork portal we collect the neighborhood names and corresponding coordinates

From the Foursquare API we would retrieve nearby venues of the neighborhood

From the website Rentcafe we retrieve the market rent trend on the neighborhood

## 2.2 Data Cleaning

Since data might not to be available to us in csv format everytime. In these scenarios we have to extract data from the url directly. Using the python library beautifulsoup we can easily extract any tables into csv format from the url. Although they may not be in the perfect format so data cleaning is always necessary with it.

We now collect average rent data for neighborhoods in Brooklyn. We parse the html page of the website using beautifulsoup library available in python and store the data we collected in a new csv file.

In[18]:

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
2	Brooklyn	Sunset Park	40.645103	-74.010316
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438
14	Brooklyn	Brownsville	40.663950	-73.910235
17	Brooklyn	Bedford Stuyvesant	40.687232	-73.941785
25	Brooklyn	Cypress Hills	40.682391	-73.876616
26	Brooklyn	East New York	40.669926	-73.880699
27	Brooklyn	Starrett City	40.647589	-73.879370
30	Brooklyn	Mill Island	40.606336	-73.908186
32	Brooklyn	Coney Island	40.574293	-73.988683
35	Brooklyn	Dyker Heights	40.619219	-74.019314

Figure 1: List of the missing data

We take the mean value of the average rent data column and assume it as the rent of the missing neighborhoods and merge it with the rent datasets.

Out[25]:

	Borough	Neighborhood	Latitude	Longitude	Average Rent
0	Brooklyn	Bay Ridge	40.625801	-74.030621	2434.313721
2	Brooklyn	Sunset Park	40.645103	-74.010316	2434.313721
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438	2434.313721
14	Brooklyn	Brownsville	40.663950	-73.910235	2434.313721
17	Brooklyn	Bedford Stuyvesant	40.687232	-73.941785	2434.313721

Figure 2: Missing data with mean rent

## 3. Methodology and Exploratory Data Analysis

This project aims in identifying the best neighborhood suitable for a individual in Brooklyn, NY. It takes in the factor of average rent of each neighborhood and venues available around a radius of 0.5kms. First step includes collecting data, which we do by collecting the NY Borough Dataset from IBMDeveloperSkillsNetwork portal. We select only the required neighborhood i.e. Brooklyn in our case and extract the features. We use this dataset to find all the nearby venues around 0.5kms radius. We also extract average market trends from the website for us to group the neighborhoods in a more informed way. The next step includes cleaning of data by checking the required data type of the values in the dataset and unwanted extra characters and check for missing values. The third step is where we do Analysis on our data. We use K-means clustering to group the neighborhood based on venues near it and its

average rent. Based on the analysis we label the clusters on their pattern and plot it in a map with different colored markers for each cluster.

The last step is to discuss our findings from the analysis and suggest the best neighborhood according to the individual's needs.

### 3.1 Plotting of coordinates

We now plot the coordinates of the different neighborhoods in Brooklyn with circle markers at the corresponding locations. To do this we make use of the folium library in python.

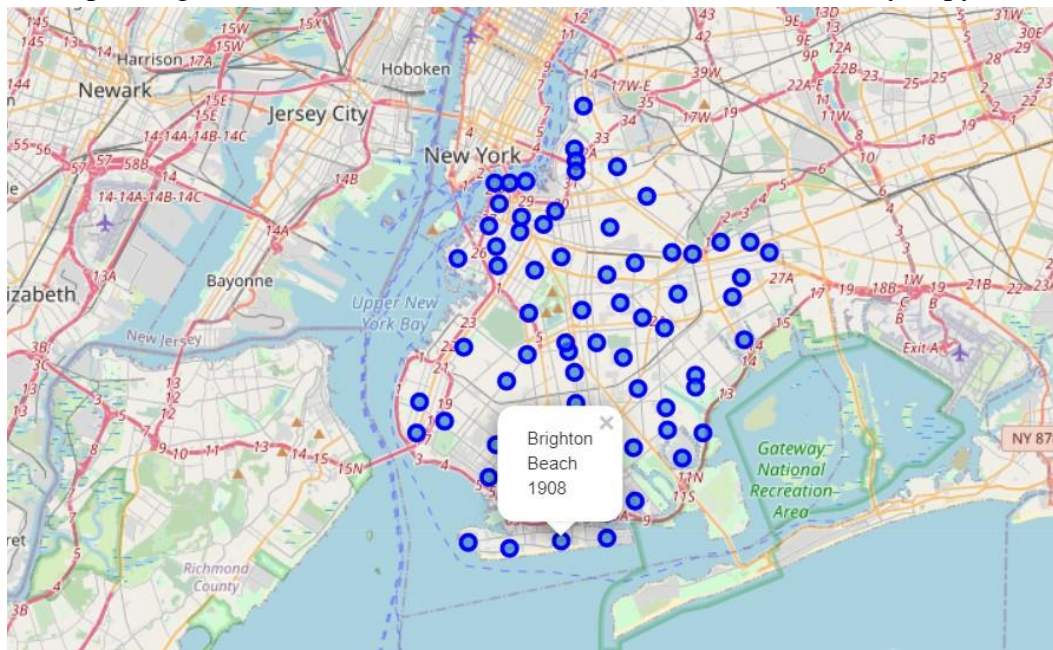


Figure 3: plotting of different neighborhood

### 3.2 Bar plot for rent trend

Below is the pictorial depiction (a bar plot) of the average rent dataset. We can observe that Carroll Gardens has the highest amount of average rent which is something to consider for

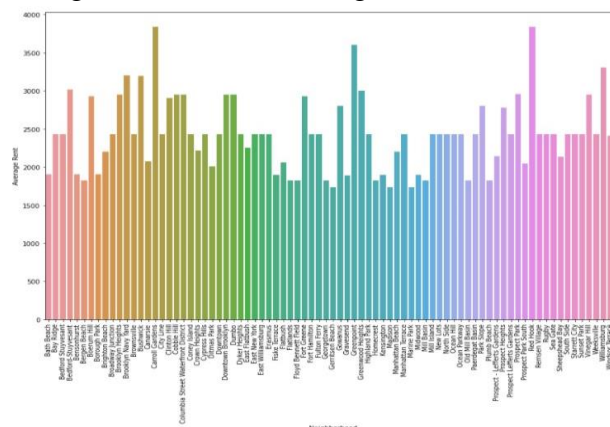


Figure 3: Bar plot for rent trend

### 3.3 Clustering

Before proceeding with k-means clustering we need to decide an optimal value for K. We do this by implementing the Elbow Method. To determine the optimal 'K' we determine the squared error for a range of values for k (range for us is 1 to 11) and plot it on the y axis. On the x axis we plot the values of 'k'. From the graph we choose the elbow point for the graph i.e. the point after which the mean squared errors start decreasing in a linear fashion. For us it seems  $K=4$  seems to be optimal.

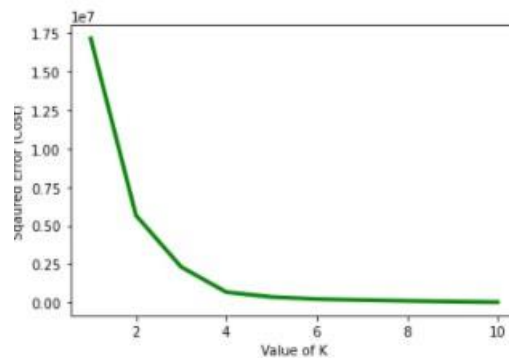


Figure 4: Elbow point

After finding the optimal K i.e. 4 for us we use kmeans clustering which will divide/group the dataset into 4 clusters based on similarity. We can use the cluster labels and merge it with one hot encoded dataset.

Below are the clustered neighbourhoods that are grouped into 4 plotted in the map of **Brooklyn**. We use different colours for each cluster to make it more distinguishable

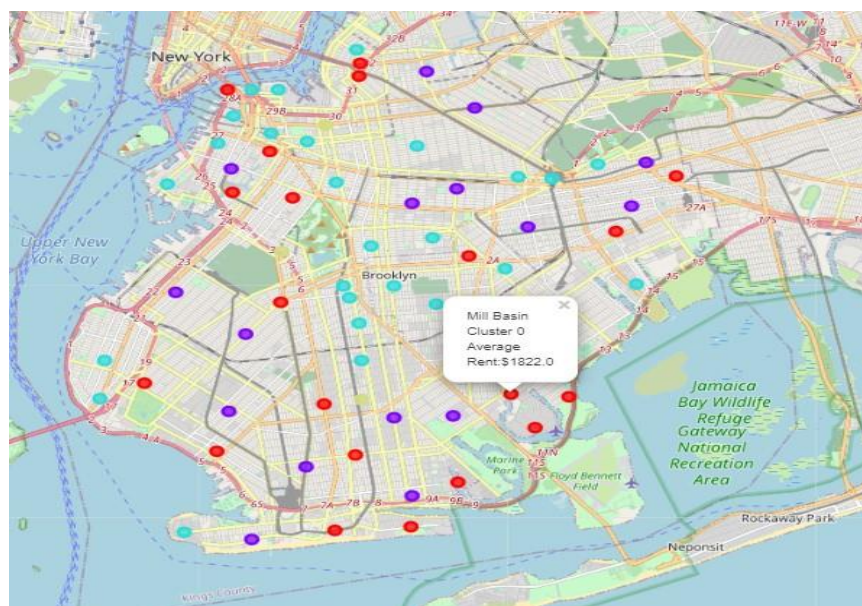


Figure 4: clustered coordinates

## **4. Results and Discussion**

With our clustered neighborhoods we can now determine the best suited neighborhood for the individual moving to Brooklyn, NY. We have gathered data on nearby venues and average market rent trend in the neighborhood and using that we have clustered the neighborhood into 4 clusters. We can now narrow down on the clusters depending on the individual's conditions. We can label the clusters as the following

### **Cluster 0:**

The neighborhoods in this cluster are having the cheapest rent and mostly have all kinds of restaurants and shop for recreational activities.

### **Cluster 1:**

The neighborhoods in this cluster have rents on the higher end and are filled with diverse and multicultural restaurants. There also seem to be a lot of pubs and bars in most of the neighborhoods of this cluster. So there a vivid nightlife in this area.

### **Cluster 2:**

The neighborhoods in this cluster have mostly average rent and are almost similar to cluster 0 in terms of venues. The only differentiator factor here is the rent

### **Cluster 3:**

The neighborhoods in this cluster have highest rent and seem like a posh area. The area has multiple jazz clubs and wine shops

An individual can look up this information and make an informed decision on which area is better suited for him/her.

## **5. Conclusion**

We have finally clustered and completed analysis on the different neighborhoods of Brooklyn, NY based on nearby venues and rent trends. We have used the clustered coordinates and plotted it over the map of Brooklyn. The purpose of this case was to help the individual in making an informed decision on which neighborhood to choose