Predicting best location for buying home

Ankit Patel

October,2019

**1. Introduction**

* 1. Background

Consider a scenario, when you want to buy a home in New York City. What you will do? Most of us will reach the internet and property advisor in order to get the answer about property prices and options available to make their final decision. Is this really a wise decision? Most of us will say yes, but it’s not.

* 1. Problem

Does the property advisor or the internet considering the liveability index parameters such as medical facilities, ease of travelling and shopping facilities around the suggested area? The answer surely is not, there naïve calculation is just based on the price factor and size of house needed without considering these important factors.

* 1. Interest

We are particularly interested in finding areas in New York city with good liveability index and pocket friendly property rates. Liveability of place considers following parameters

* Medical Facilities
* Shopping & Services
* Food and Restaurant
* Arts & Entertainment Zones
* Outdoor & Recreational activities

**2. Data needed**

Data needed for our decision making:

* number of Medical Centers in the neighborhood
* number of Arts & Entertainment places in the neighborhood
* number of Shopping stores in the neighborhood
* number of Outdoor & Recreational places in the neighborhood
* number of Food & Restaurants in the neighborhood
* number of Travelling options in the neighborhood
* property price in the neighborhood
* different neighborhood in New York City

We will need geo-coordinates of each neighbor in order to find various liveability index parameters (using foursquare api) described in section 1.3.

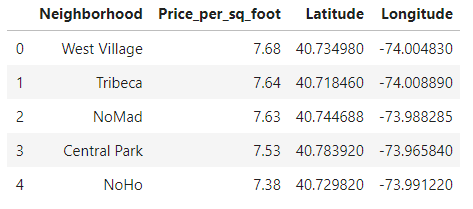
Finally, we can know the liveability of that area using the above data i.e. whether the number of Medical Facilities, Shopping & Services, Outdoor & Recreational places and Travelling options are high or not in that area and whether the property price is low or not. Using the above data, we can also compare various neighbors of chosen city and choose the best pocket friendly highly livable area.

Thus, we can achieve our task of predicting best location for buying home.

**3. Methodology**

3.1 Identifying Neighborhood and property prices

Using this web page <https://www.zumper.com/blog/2019/05/nyc-by-square-foot-see-which-neighborhood-gets-you-the-most-space-for-your-money/> we can get different areas of New York city and their property prices. We will use BeautifulSoup to scrap the different neighbors and their respective property prices from the above web page and store it in pandas dataframe. Then we will use HERE maps api to Geocode the neighbors address in order to get Geo-Coordinates of each location.



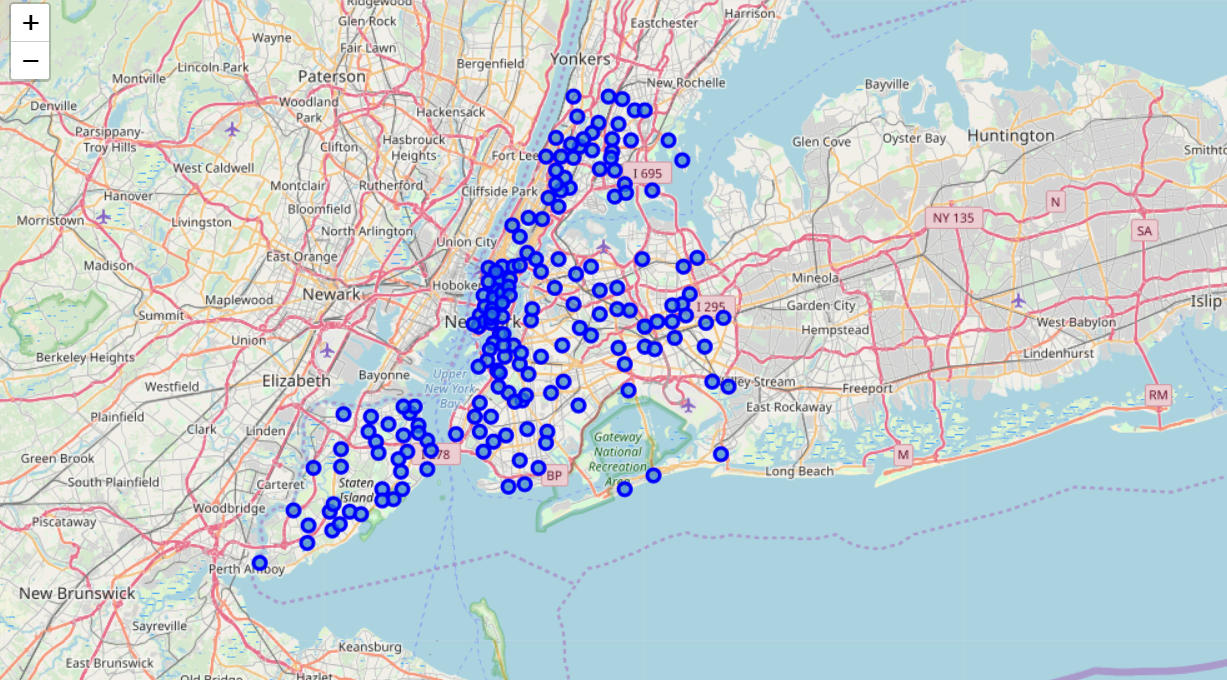
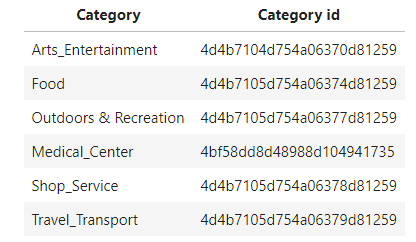


Fig 1: Marker points on different areas of New York City

3.2 Getting liveability index parameters for each neighborhood

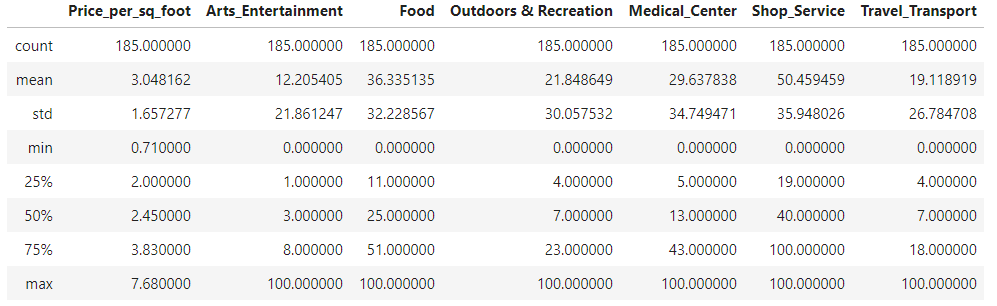
After finding the list of neighborhood, their Geo-Coordinates and property prices, we then connect to the Foursquare API to gather information about number of venues inside each neighbor. We are only interested in following categories:



The Category id's of different categories are obtained from following web page: <https://developer.foursquare.com/docs/resources/categories>

3.3 Statistical view of our data

Clearly from below table we can find that values of column Price\_per\_sq\_foot ranges from 0.71 to 7.68 and that of another lie between 0 to 100. Therefore, there is a need to scale these features before clustering.



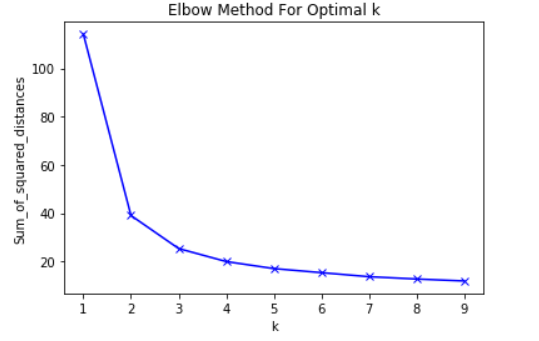
3.4 Applying one of Machine Learning Techniques (K-Means Clustering)

A. We have selected following features for clustering purpose and scaled them:

* Price\_per\_sq\_foot
* Arts\_Entertainment
* Food
* Outdoors & Recreation
* Medical\_Center
* Shop\_Service
* Travel\_Transport

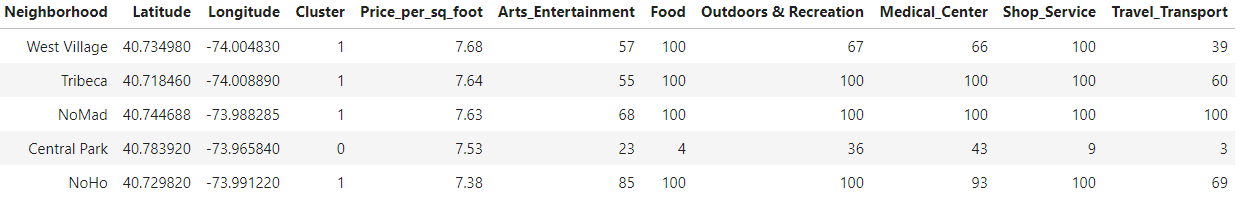
B. Finding Optimal k

Using the elbow method optimal value of k is found to be 3



C. K-Means Clustering:

Now with k=3 we can use k means clustering to agglomerate data based on above selected transformed features.

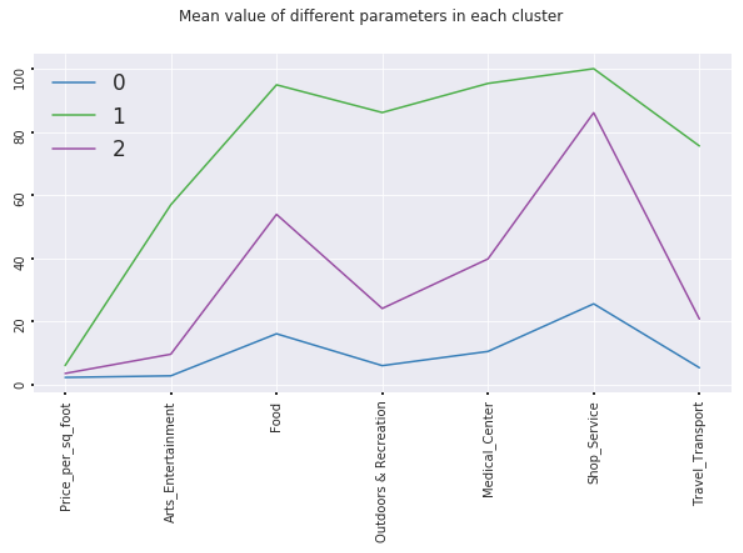


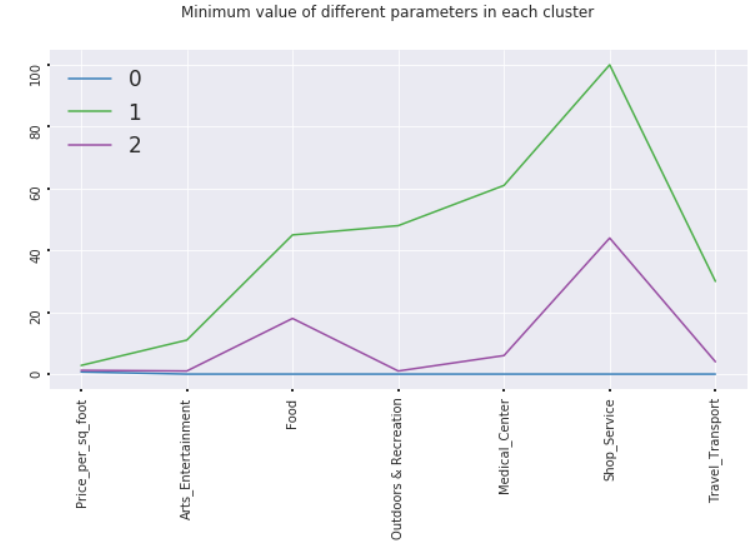
**4. Results**

4.1 Quick Analysis

From below two graph we can see liveability of area in each cluster as follows:

* For Cluster 1 liveability is high i.e. number of restaurants, hospitals, travelling, shopping and entertainment options are great but the starting property price in this region is high.
* For Cluster 2 liveability is medium i.e. compromise in either number of restaurants, hospitals, travelling, shopping and entertainment options or less property price, but budget friendly options might also be present in this cluster.
* For Cluster 0 liveability is low i.e. number of restaurants, hospitals, travelling, shopping and entertainment options are less.





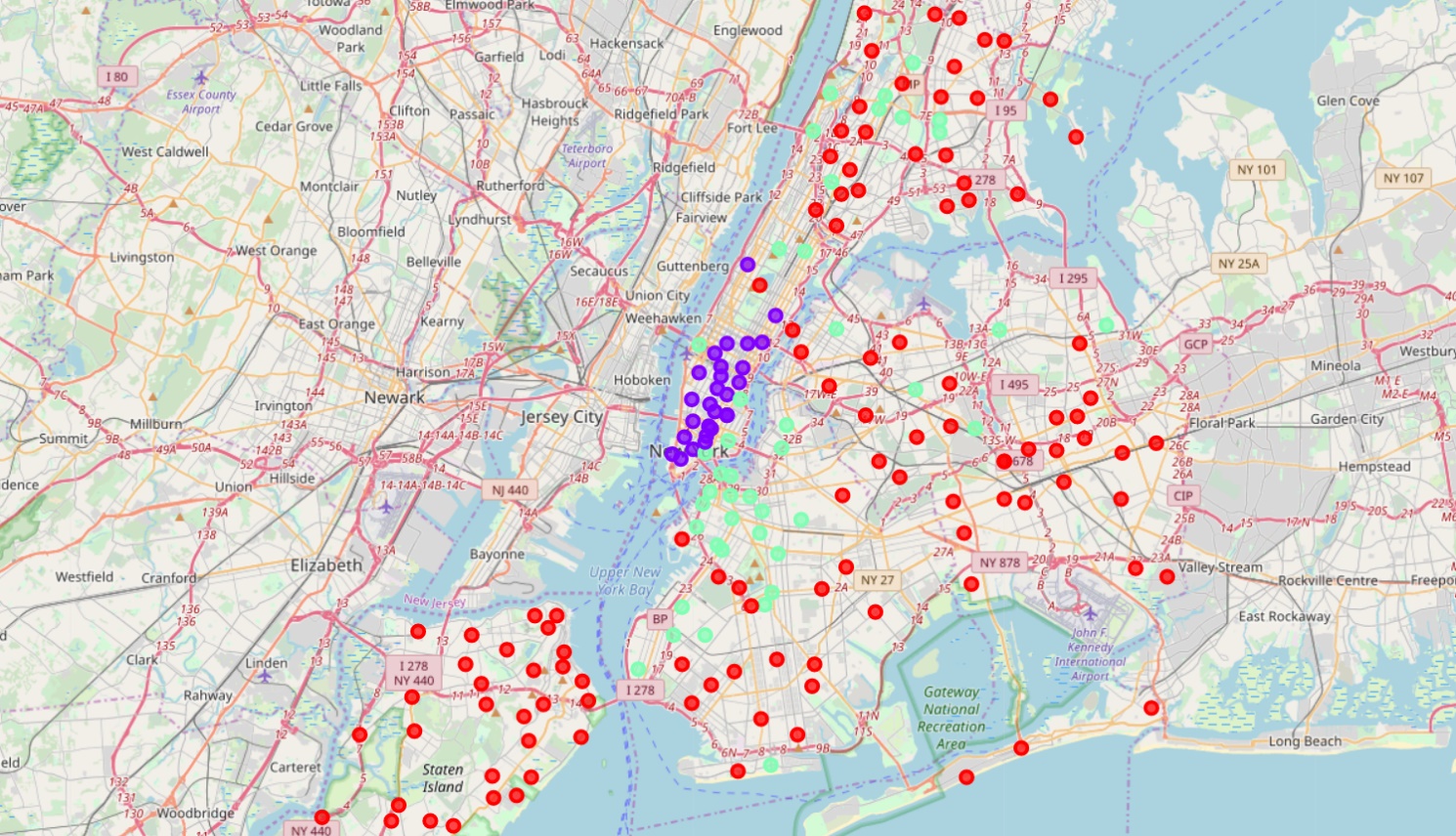


Fig 2: Cluster Marker on New York City Map

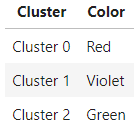


Table: Marker color and their corresponding clusters

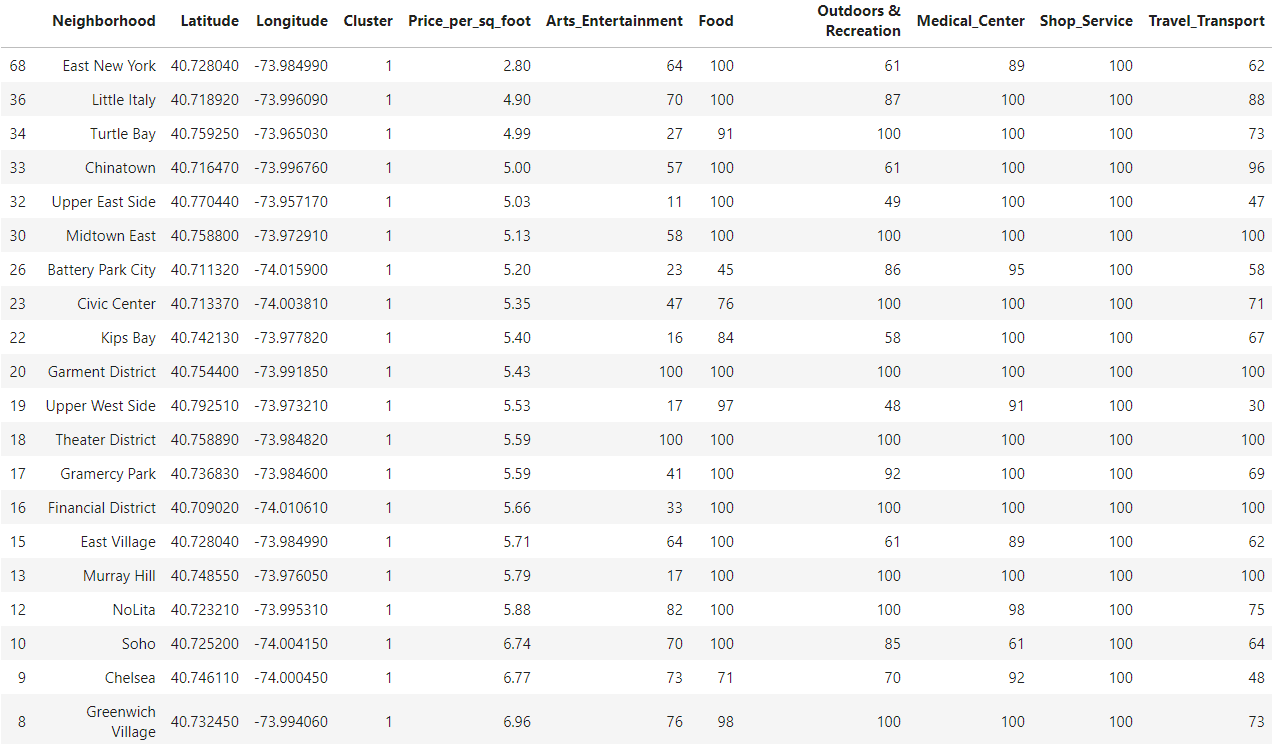
4.2 Deep Analysis

1. Cluster 1

The property price in this cluster ranges from 2.80 to 7.68 dollar per sq foot which is quite high because the liveability is very high in these places.

If your budget is above 2.80 dollar, some of the best places to buy house in this cluster

* **East New York** with lowest price of 2.80 dollar per sq foot
* **Little Italy** with price of 4.90 dollar per sq foot
* **Turtle Bay** with price of 4.99 dollar per sq foot and so, on



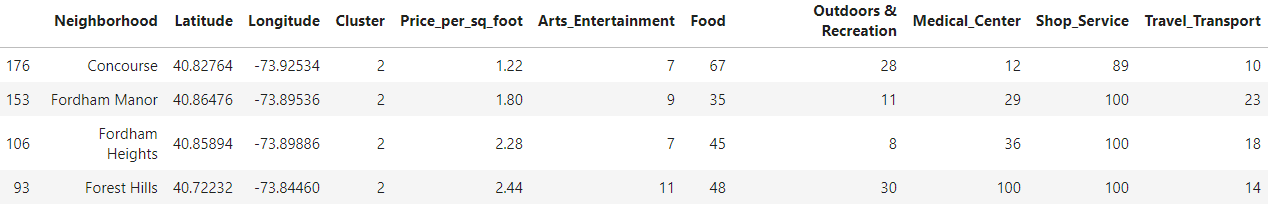


1. Cluster 2

The property prices in this cluster ranges from 1.22 to 6.37 dollar per sq foot.

If you budget ranges from 1.22 to 2.80 dollar the best places in this cluster with medium liveability are

* **Concourse** with a price of 1.22
* **Fordham Manor** with a price of 1.80
* **Fordham heights** with a price of 2.28 and so on



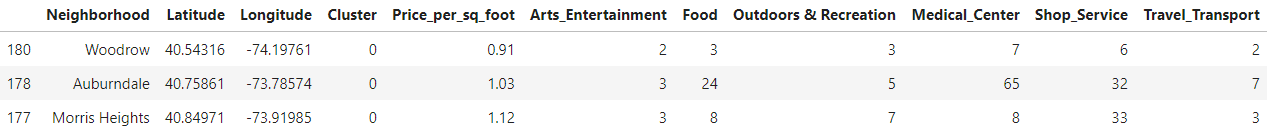
1. Cluster 0

The property prices in this cluster ranges from 0.71 to 7.53 dollar per sq foot.

The lowest property price is in Highbridge, Oakwood, Prince's Bay, Clifton and Woodrock below 1 dollar per sq foot

The best places in this cluster with fair liveability if you budget is below 1.22 dollar are

* **Woodrow** with a price of 0.91
* **Auburndale** with a price of 1.03
* **Morris Heights** with a price of 1.12



**5. Discussion**

The aim of this project is to find the best places where we can buy home. We found out that

* For Cluster 1 liveability is high i.e. number of restaurants, hospitals, travelling, shopping and entertainment options are great but the starting property price in this region is high.
* For Cluster 2 liveability is medium i.e. compromise in either number of restaurants, hospitals, travelling, shopping and entertainment options or less property price, but budget friendly options might also be present in this cluster.
* For Cluster 0 liveability is low i.e. number of restaurants, hospitals, travelling, shopping and entertainment options are less.

**6. Conclusion**

This project helps a person in buying home by considering most important liveability parameters apart from property price and suggesting them better pocket friendly alternatives.

Thank You!!!