Final Report

Capstone Project – The Battle of Neighborhoods

Introduction

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart choice in selecting their neighborhood in Delhi.

This Project aim to create an analysis of features for a people migrating to Delhi to search a best neighborhood as a comparative analysis between neighborhoods.

The features include median housing price. It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.

Data Selection

Data Link :https://www.kaggle.com/shaswatd673/delhi-neighborhood-data

we will use delhi Dataset consisting of neighborhoods,latitude and longitude.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

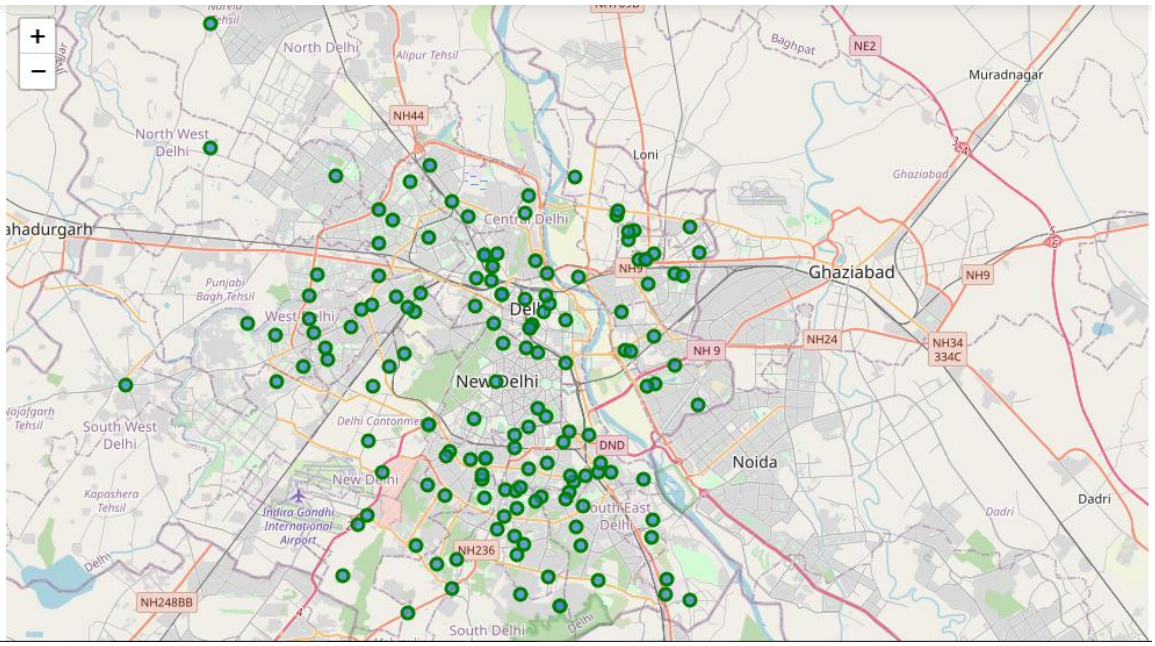
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude . The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant

6. Venue Latitude
7. Venue Longitude
8. Venue Category

Delhi Map :

Out[75]:



Using K-Means Clustering Approach:

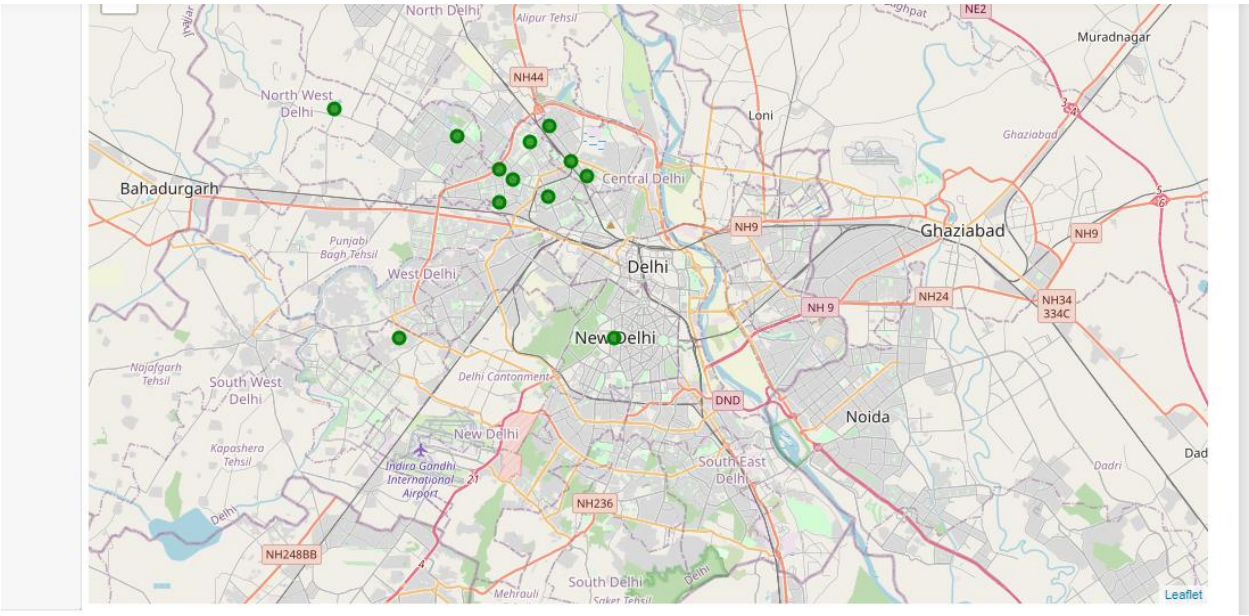
k-Means Clustering Approach

In [46]:
```
# Using K-Means to cluster neighborhood into 3 clusters
delhi_grouped_clustering = delhi_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(delhi_grouped_clustering)
kmeans.labels_
```

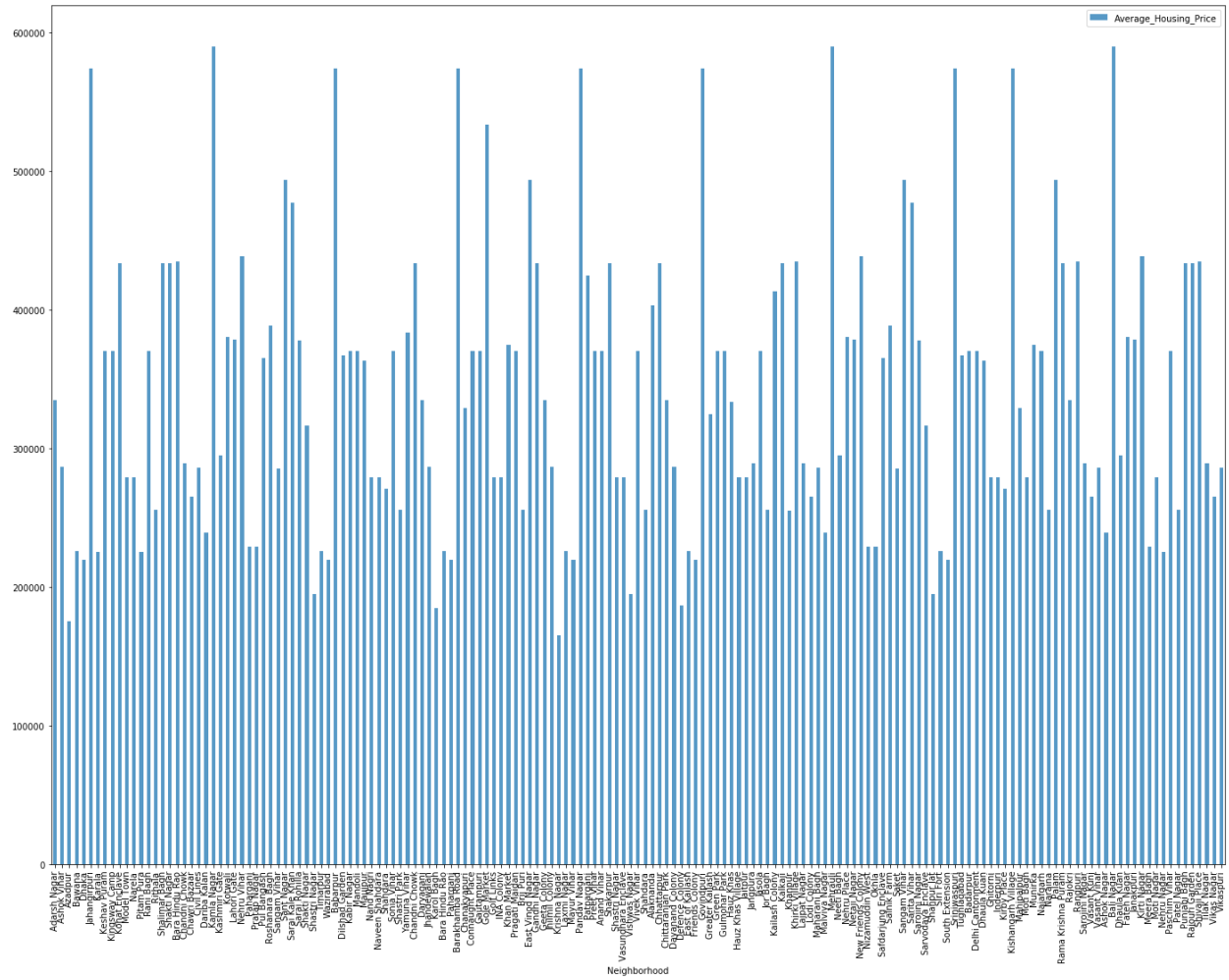
Out[46]:
```
array([1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
       1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       2, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 2])
```

Results Section

Map of Clusters in Delhi



Average Housing Price by Clusters in Delhi:



Discussion Section

Problem Which Tried to Solve:

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shiffting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

1. Sorted list of house in terms of housing prices .

Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 163 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.