

Final Report

1. Introduction

This project is aimed at helping people finding better neighbourhoods in New York. People from different places around the world migrate to USA in search of work and settle down in different states. New York is one such state which attracts a large number of crowd owing to its lively atmosphere and being the heart of United States.

The project helps in making a comparative analysis for the people looking to settle in New York City to search for better neighbourhoods in terms of good schools, vicinity to the supermarkets, ease of commute etc.,

It will help people to become aware of the area and neighbourhood before moving to a city, state, country or place for their work.

2. Data Collection

Foursquare API

We will need data of the places in the vicinity of the houses in a particular borough. For this, we will have to make a call to Foursquare API with the radius kept to 500 metres. This will fetch the details of all the nearby places which will make it easy for comparative analysis amongst the places under consideration.

The data for the analysis is available from the link mentioned in the lab for the boroughs in New York City. The link to the file is given below: -

URL: https://geo.nyu.edu/catalog/nyu_2451_34572

For average housing price, the following link is used: -

URL: <https://www.rentcafe.com>

For School Rating, the following link is used: -

URL: <https://www.greatschools.org>

The call to the Foursquare API will return following data as seen in the clustering assignment: -

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood 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

With the above data at hand, it will be easy to compare the boroughs and select the ideal borough to settle.

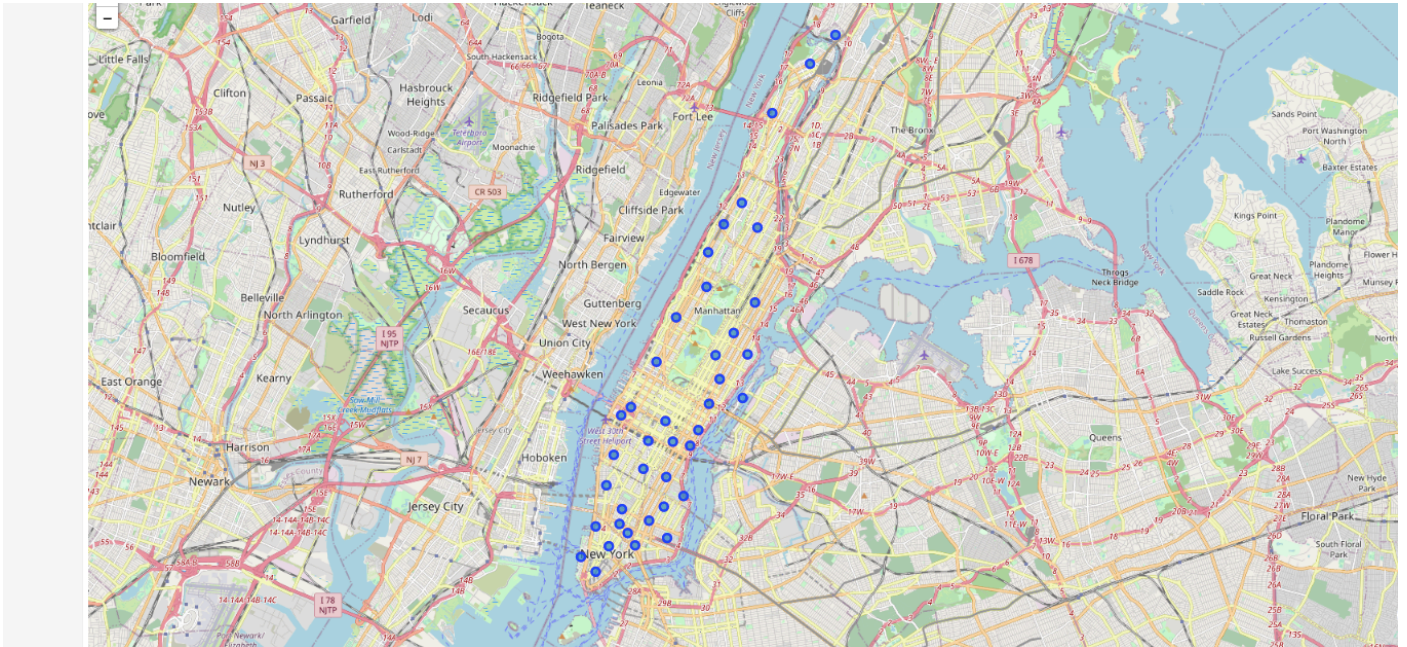


Figure 1: Map of Manhattan

3. Methodology

Clustering Approach

We will do a comparative analysis of the cities based on the neighbourhoods where we will segment them, and group them into clusters. We will then find similar neighbourhoods in a New York. We will use K-means clustering algorithms which is an unsupervised learning algorithm to generate clusters.

5. Cluster Neighbourhoods

```
In [41]: kclusters = 5

manhattan_grouped_clustering = manhattan_grouped.drop('Neighborhood', 1)

kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(manhattan_grouped_clustering)

kmeans.labels_[0:10]
```

Figure 2: Grouping Neighbourhoods into 5 Clusters

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 Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site	Shopping Mall	Food Court	Clothing Store	Gourmet Shop	Boat or Ferry
1 Carnegie Hill	Coffee Shop	Café	Yoga Studio	Wine Shop	Italian Restaurant	Gym / Fitness Center	Gym	French Restaurant	Pizza Place	Bookstore
2 Central Harlem	African Restaurant	Cosmetics Shop	French Restaurant	American Restaurant	Bar	Seafood Restaurant	Chinese Restaurant	Salon / Barbershop	Coffee Shop	Cocktail Bar
3 Chelsea	Coffee Shop	Art Gallery	Bakery	American Restaurant	Italian Restaurant	Ice Cream Shop	Cycle Studio	Cupcake Shop	Café	Market
4 Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	American Restaurant	Dessert Shop	Noodle House	Salon / Barbershop	Optical Shop	Spa	Hotpot Restaurant

Figure 3 Identifying Common Venues near each Neighbourhood

Work Flow:

Using credentials of Foursquare API, features of near-by places of the neighbourhoods are determined. Number of requests per day are set to 100 and the radius of venue search is set to 500 meters.

4. Results

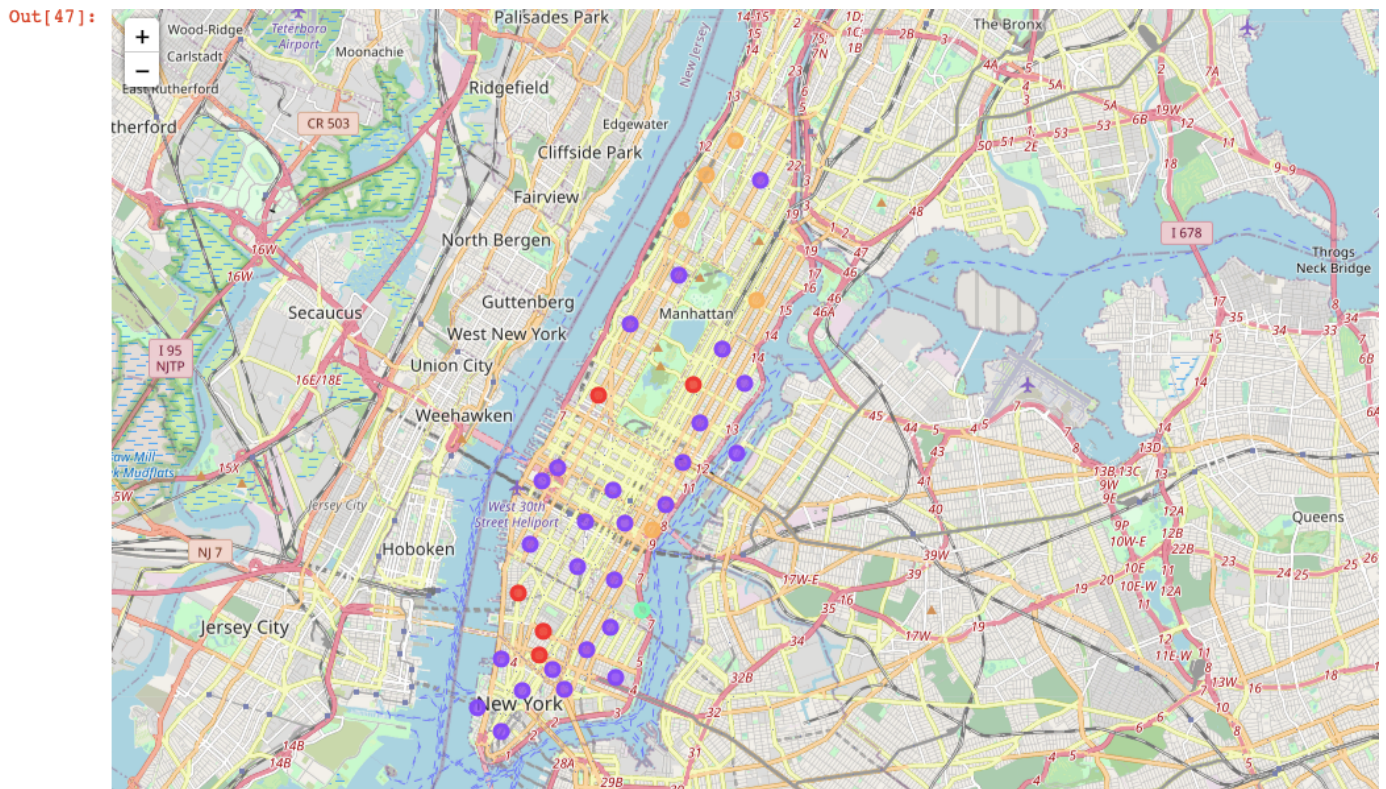


Figure 4 Map of Clusters in Manhattan

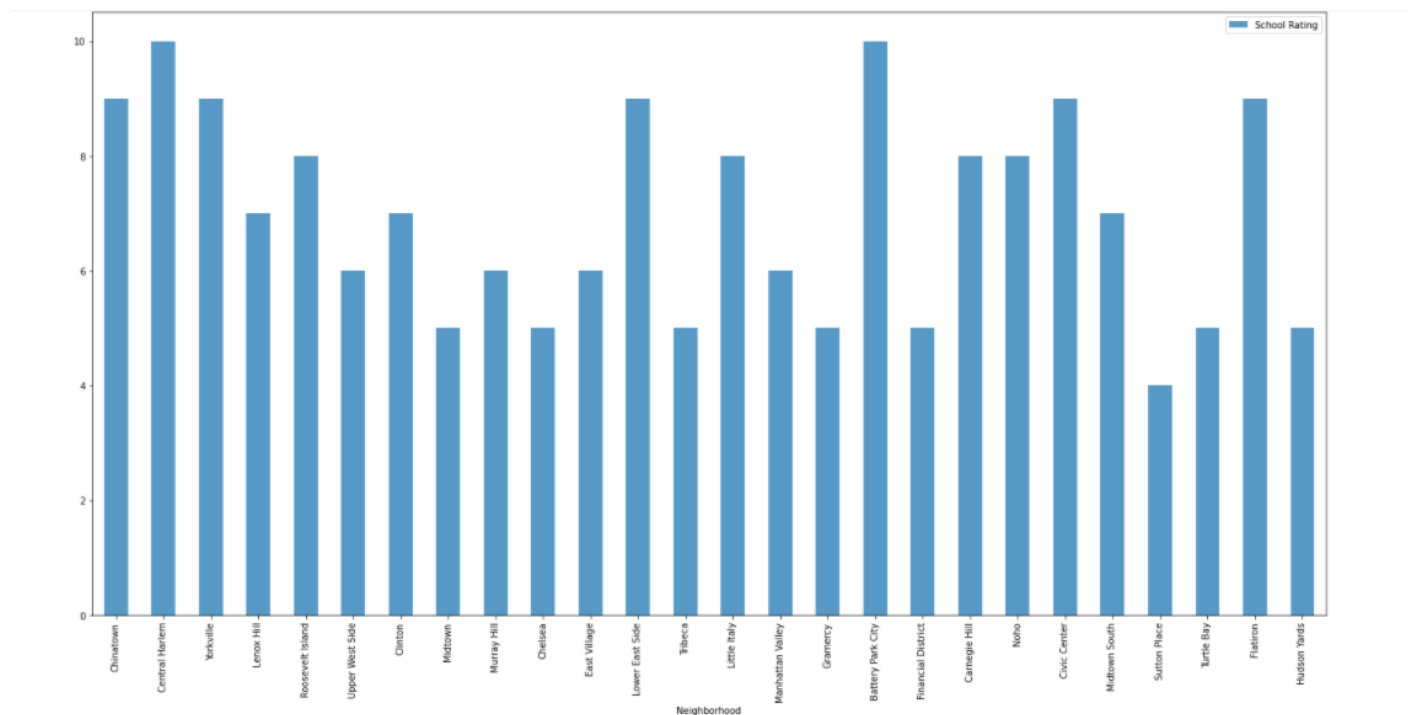


Figure 5 Ratings of schools present in the Neighbourhoods in a cluster

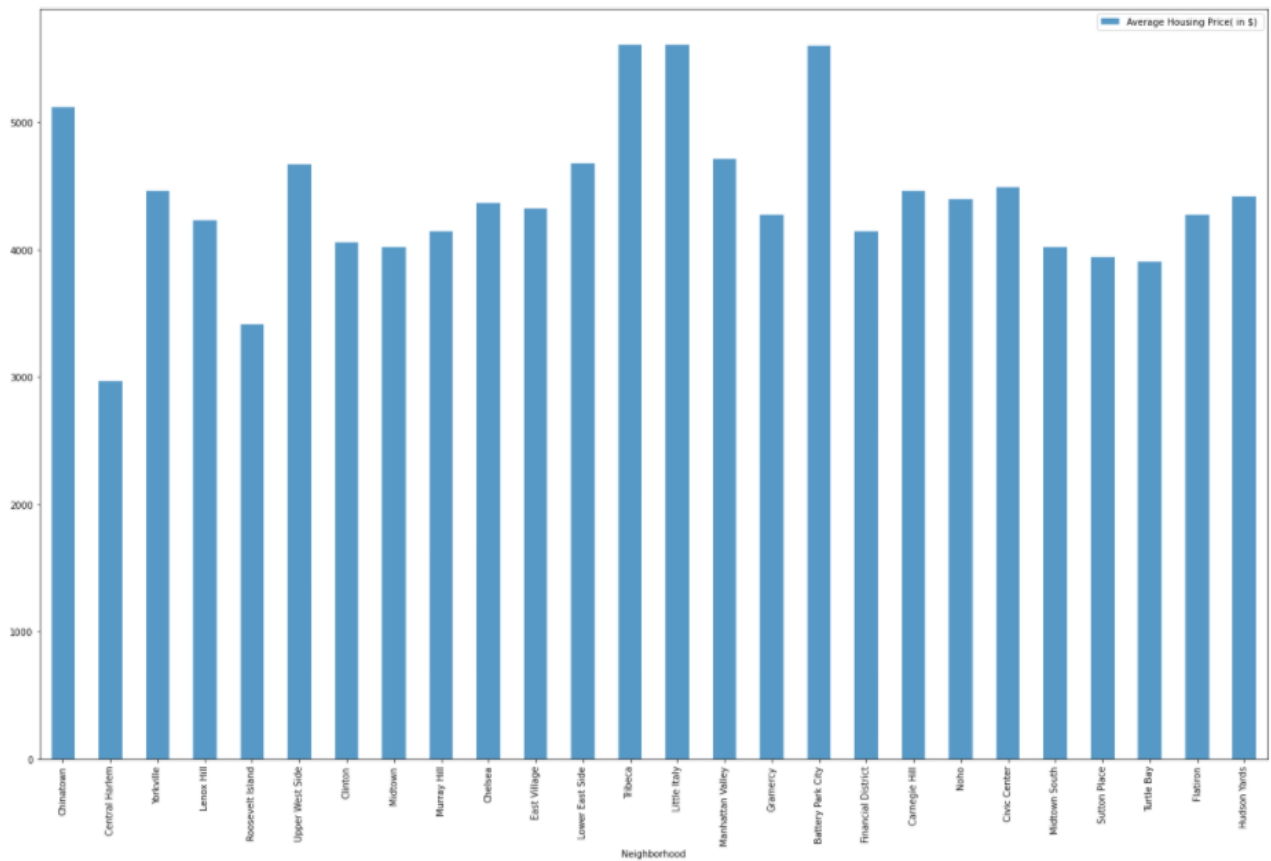


Figure 6 Average Housing Price of Neighbourhoods in a cluster

Location

Manhattan is one of the popular places for the people willing to settle in New York City. It is one of the densely populated areas in New York and has a rich diversity in terms of social culture and living standards.

Foursquare API

The Four Square API is particularly useful in this project as it will help in finding the places in the vicinity of the houses the new immigrants are considering as their probable home. This will help the people is narrowing down to the ideal place that they have in their mind to settle down with their families.

5. Discussion

Problem

The primary goal of the project is to help the person who is a new immigrant in selecting a neighbourhood in the New York City. The selection is made based on variety of factors like: -

1. Vicinity to the super market
2. Ease of commute
3. Presence of reputed schools nearby
4. Less crime rate

The project is likely guide a person by providing him: -

1. List of houses available along with their cost/rent.
2. List of schools along with their ratings, fees, reviews etc.,

6. Conclusion

In this project, I used k-means clustering to divide the neighbourhood into 5 clusters and then analysing each clusters for the top places in the vicinity of those neighbourhoods. The results presented in the above section are backed by

charts which provide a clear picture of the comparative analysis of the neighbourhoods in terms of housing prices and the school ratings.

This course has provided me with an overwhelming experience and I learned a lot throughout this course and I believe that this effort will take me a long way in my career.

Future Works

This model can be further refined to provide the immigrants with the sure shot and a one go solution to selecting the neighbourhood which match their needs and desires and help them in finding an ideal neighbourhood to settle.

Libraries Required:

Pandas: For creating and manipulating Dataframes.

Folium: Python visualization library.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

Geocoder: To retrieve Location Data.

BeautifulSoup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.