**OPTIMIZING THE BEST LOCATION FOR OPENING AN AYURVEDIC AND INDIAN ORGANICS STORE IN MANHATTAN, NYC**

**Introduction :**

NYC is the financial capital and the most populous city of the United States. I am from India and like many other Indians, I’ve used and continued to use many Ayurvedic remedies for various common and uncommon health and beauty issues. Ayurveda is a natural system of medicine which is 3000 years old. I wish to spread the magic of this “God’s gift to Indians” to the entire world, and what better place than NYC to start something like this.

It will be a clinic with highly qualified and experienced ayurvedic doctors. Along with the clinic, there will also be an Indian organics store which will have all sorts of raw, pure and unprocessed products which are used in ayurvedic remedies.

The target customers are the people who are fitness freaks, crazy about yoga and are aware of the cons of using processed and chemical-intensive products for health and beauty.

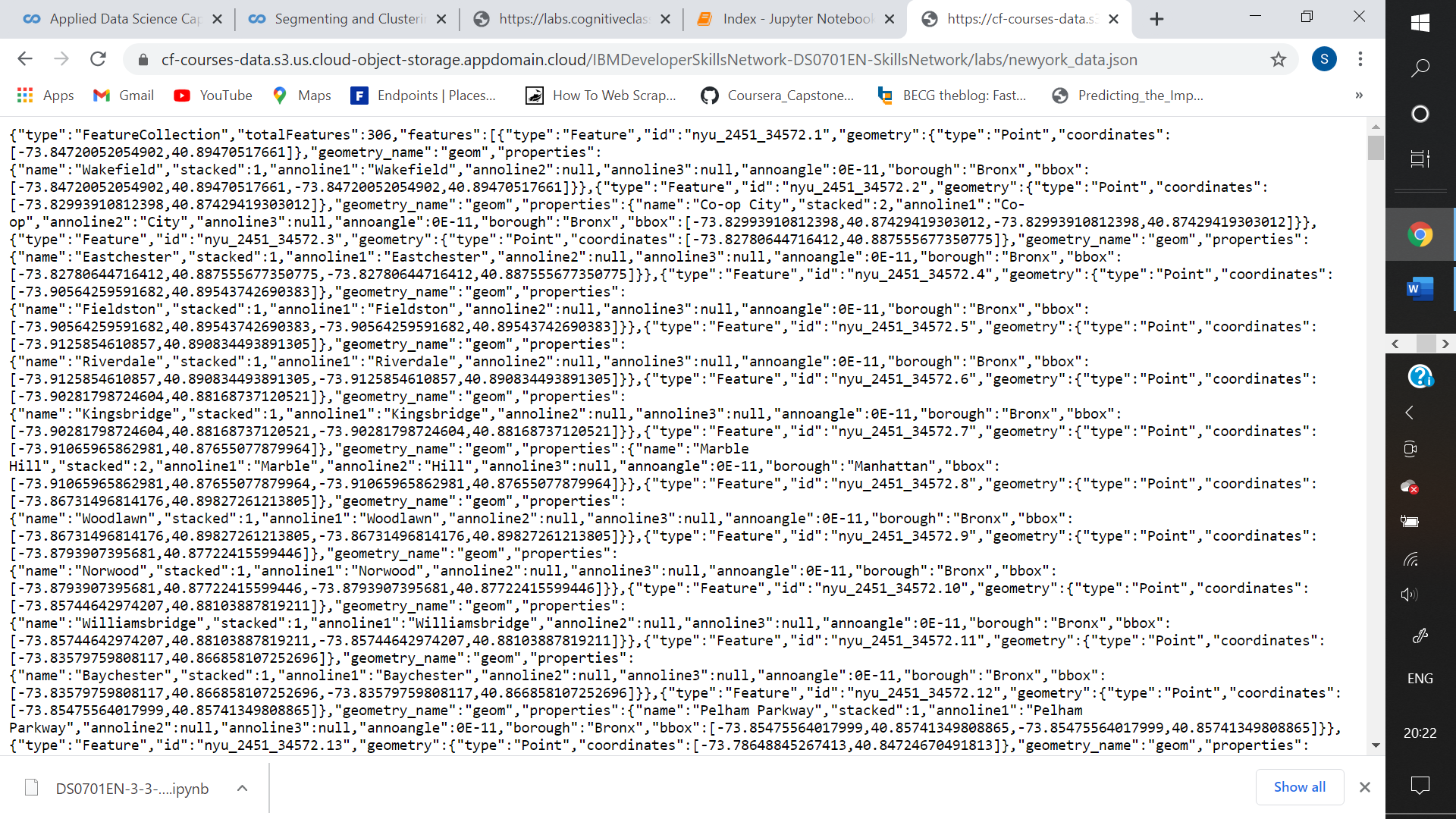
In this project, I will be optimizing the best location in Manhattan, NYC for maximizing the target audience for the clinic and store.

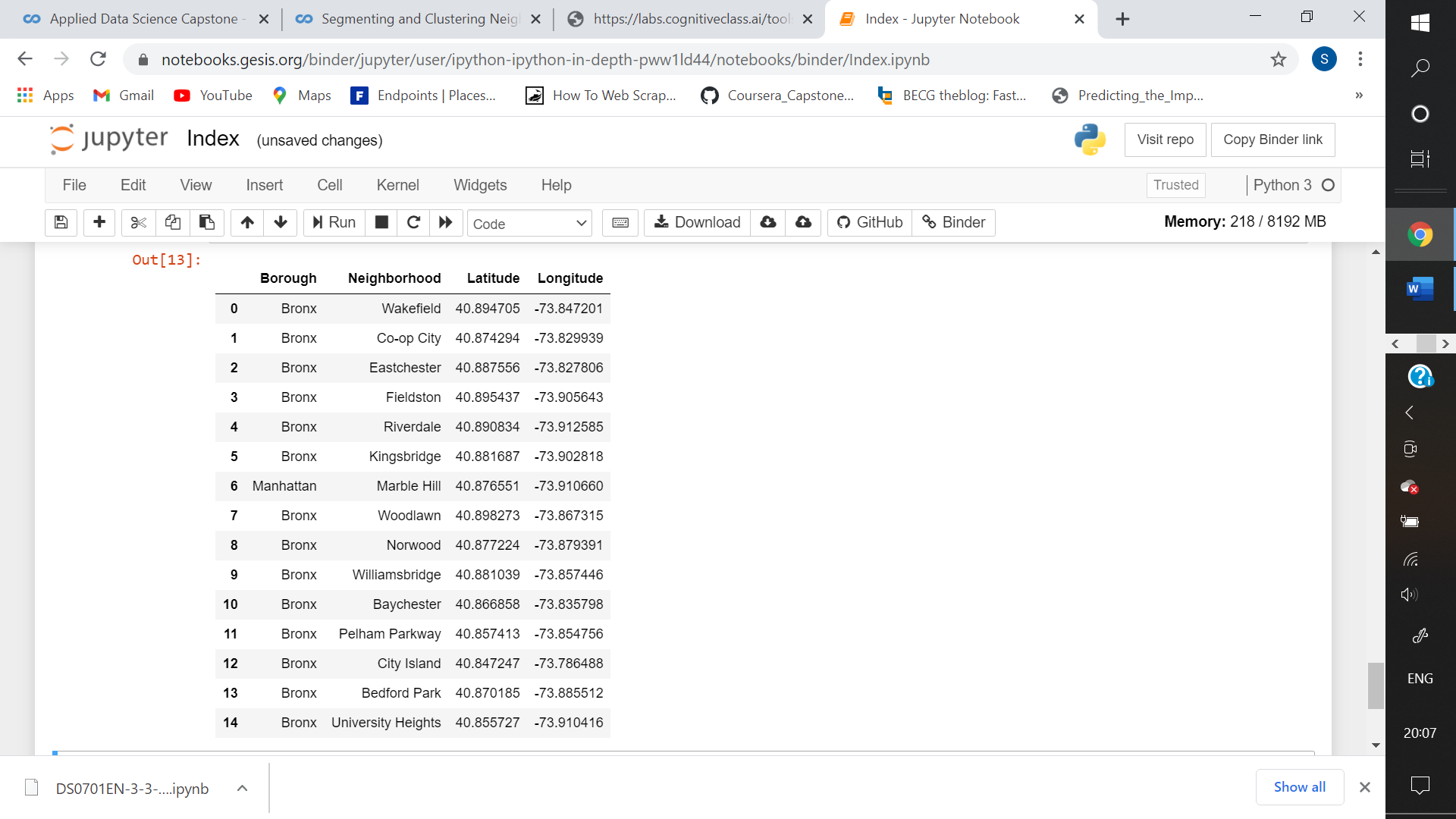
The target audience for this project is a client who wants to open an Ayurvedic clinic in NYC.

**Data used in the project :**

**Source for NYC data -**

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork\_data.json

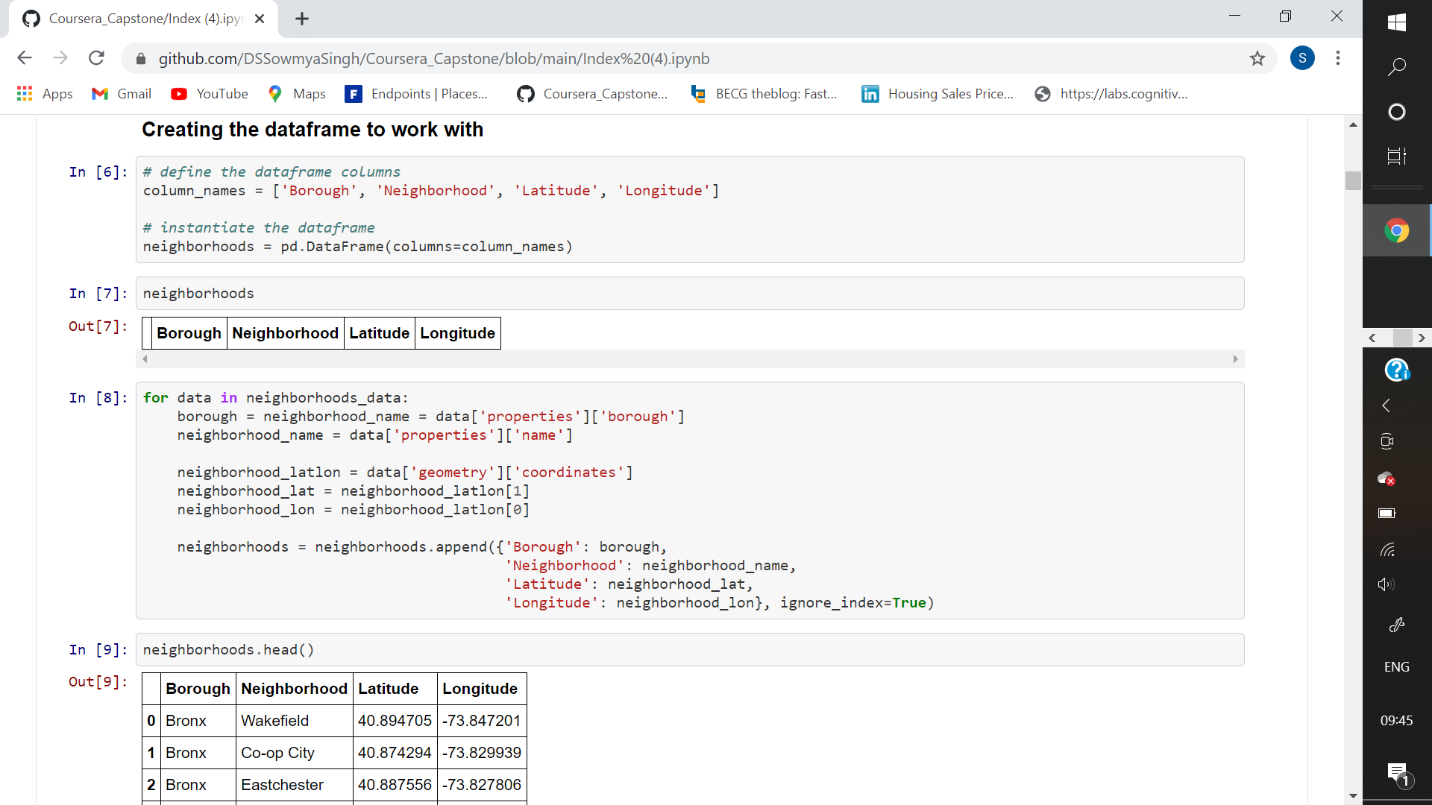




I used k-means clustering to form clusters based on the type/category of venues present in the locations. Analysing the clusters helped me in choosing a suitable location for opening an Ayurvedic clinic in NYC.

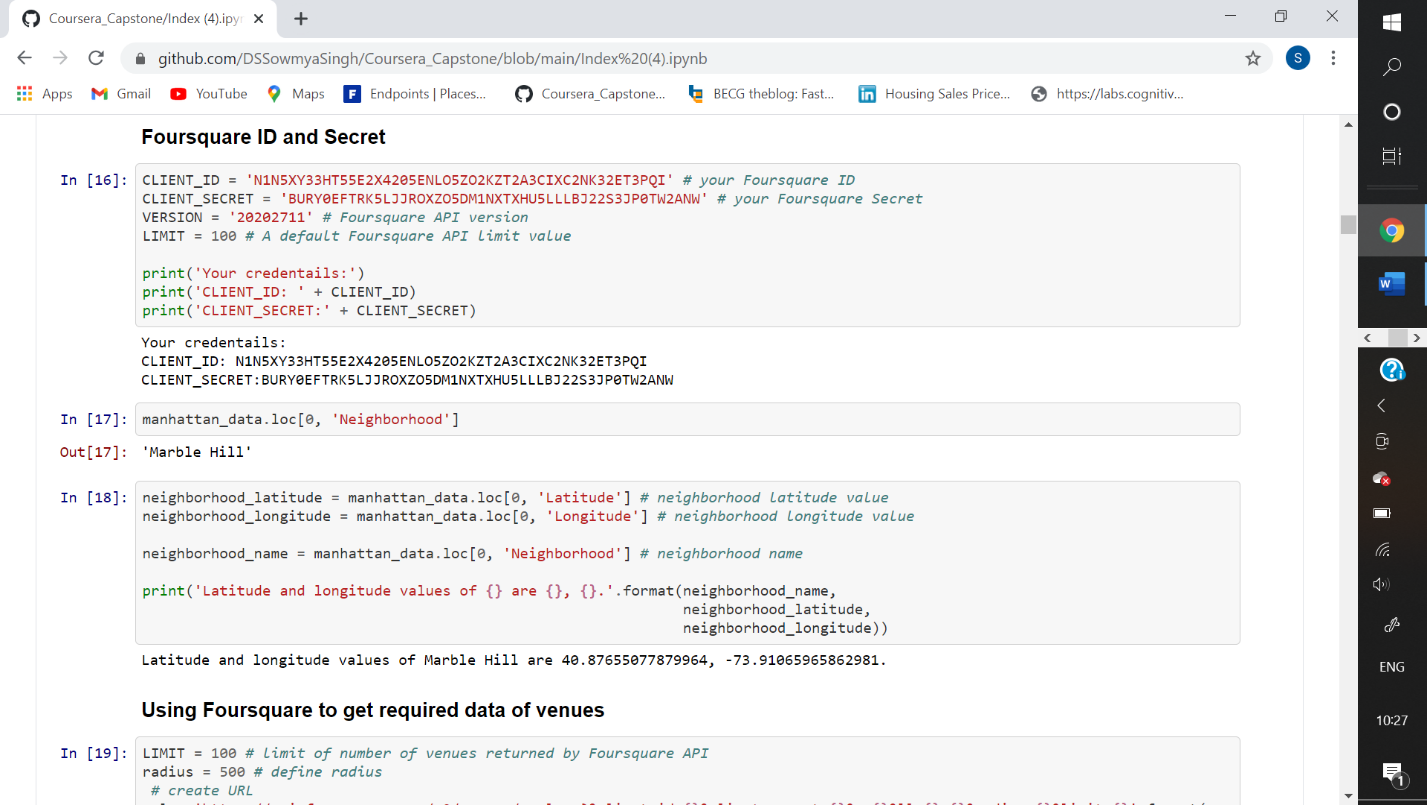
**Methodology :**

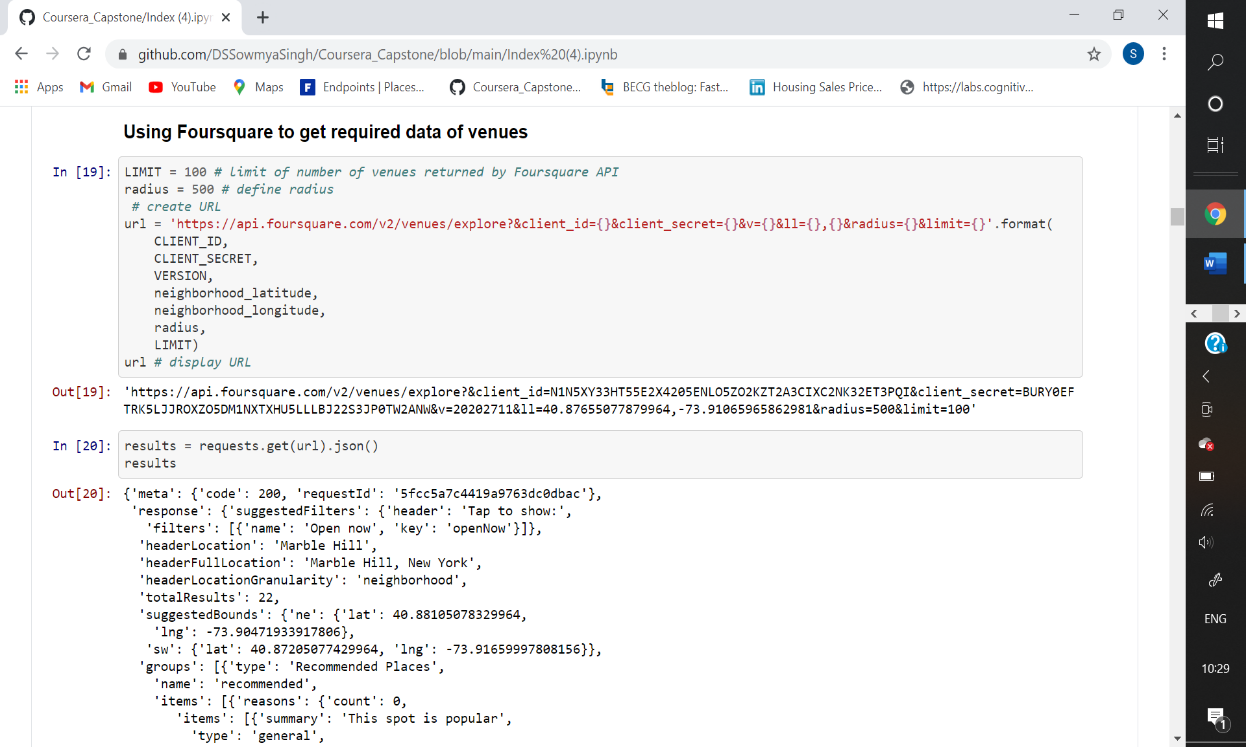
I used the data from the source to create a dataframe with the columns – Borough, Neighbourhood, Latitude and Longitude.



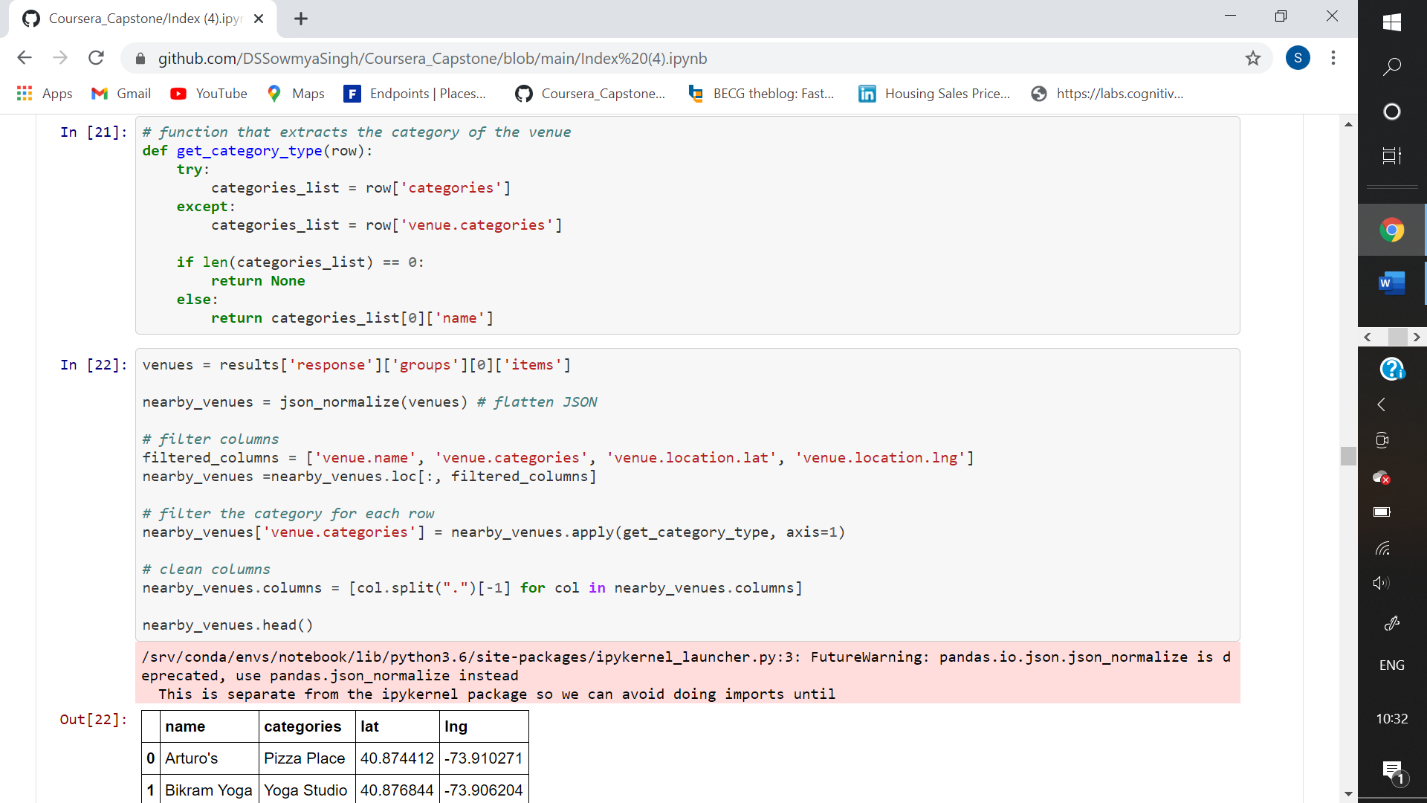
I used this dataframe (‘neighborhoods’) to create another dataframe (‘Manhattan data’) consisting of all neighborhoods of ‘Manhattan’ borough. I found the coordinates of Manhattan using geolocator to create its map with labels showing all its neighborhoods



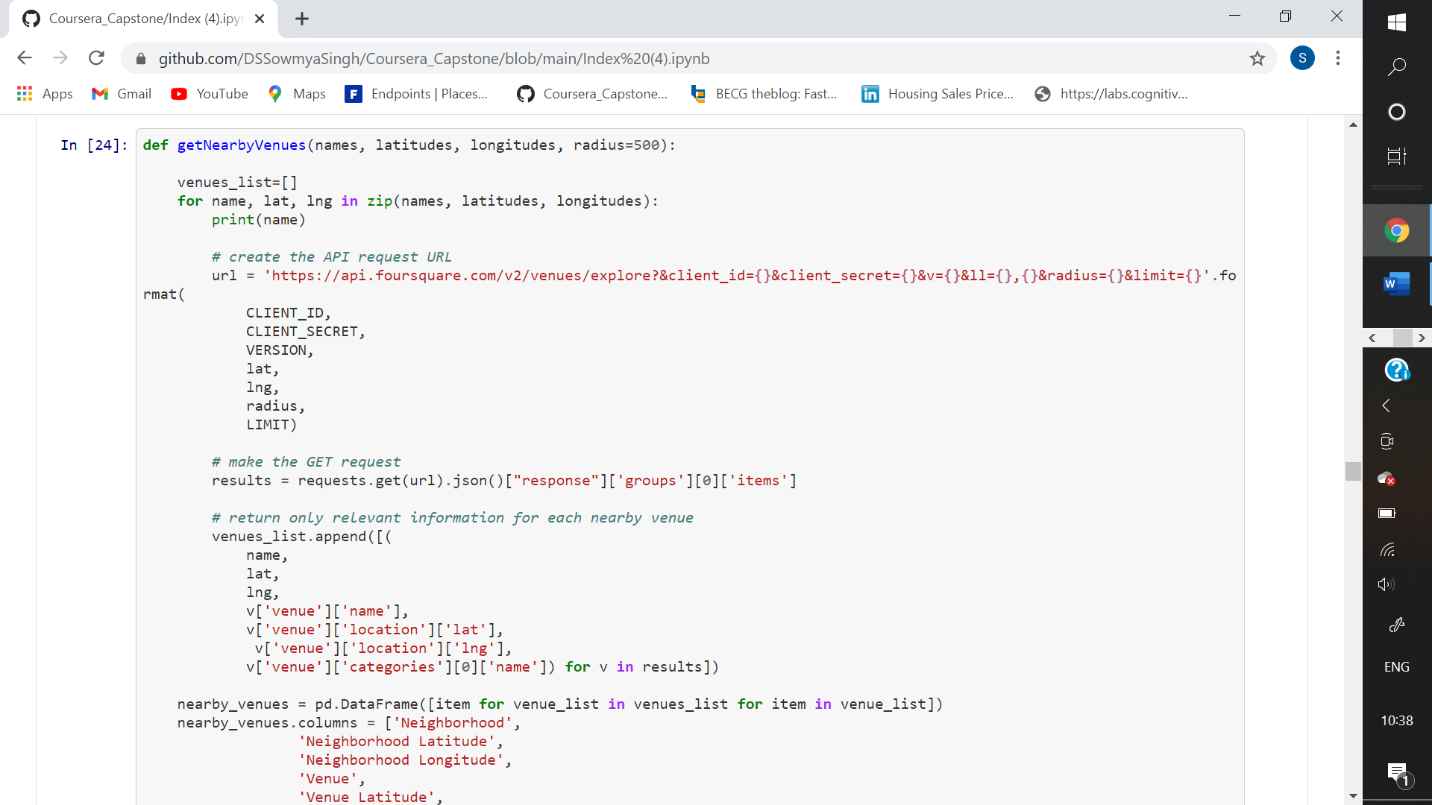
I used Foursquare to get the required data of venues in each neighborhood.



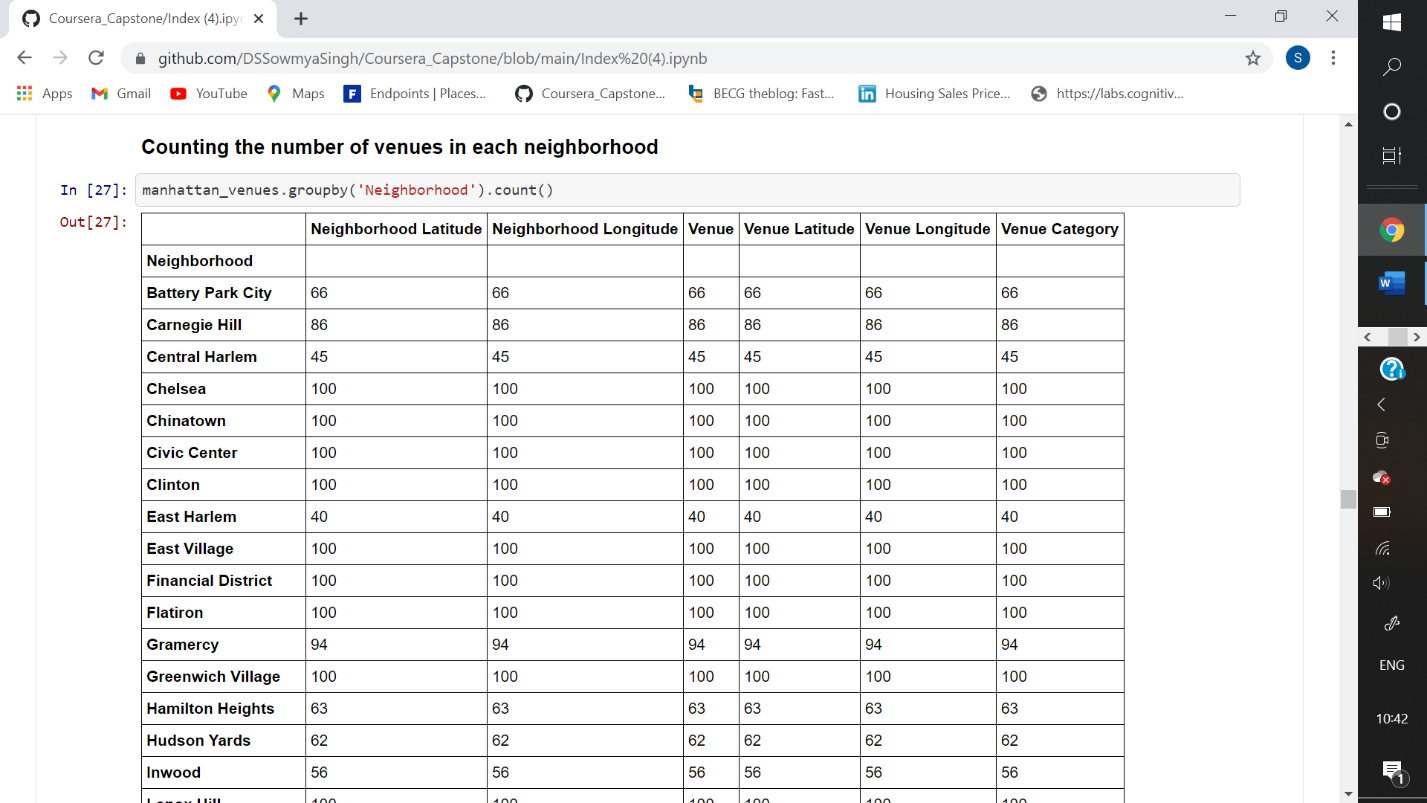
I defined a function to extract the categories of the venues, which will help me in clustering them.



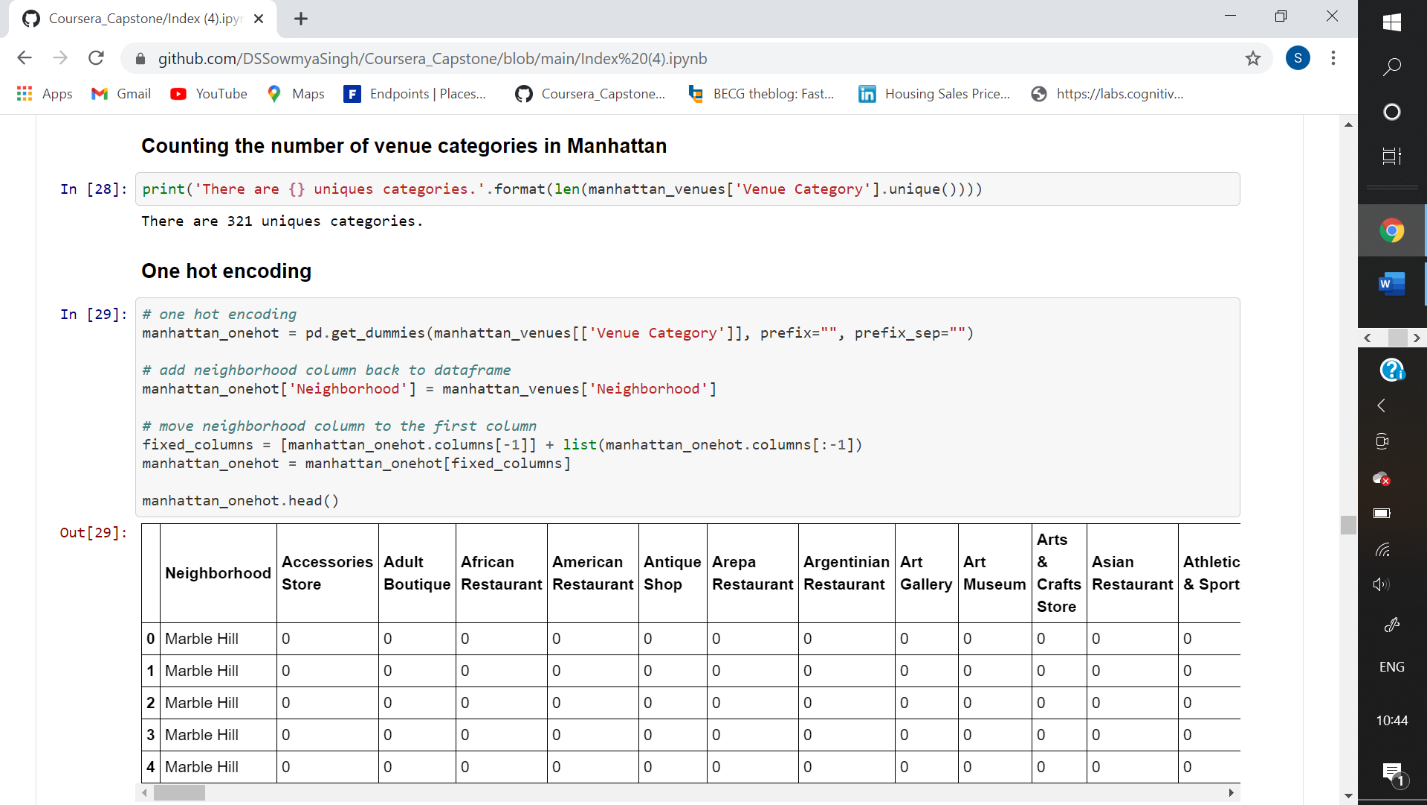
Then I defined a function to get nearby venues for each neighborhood.



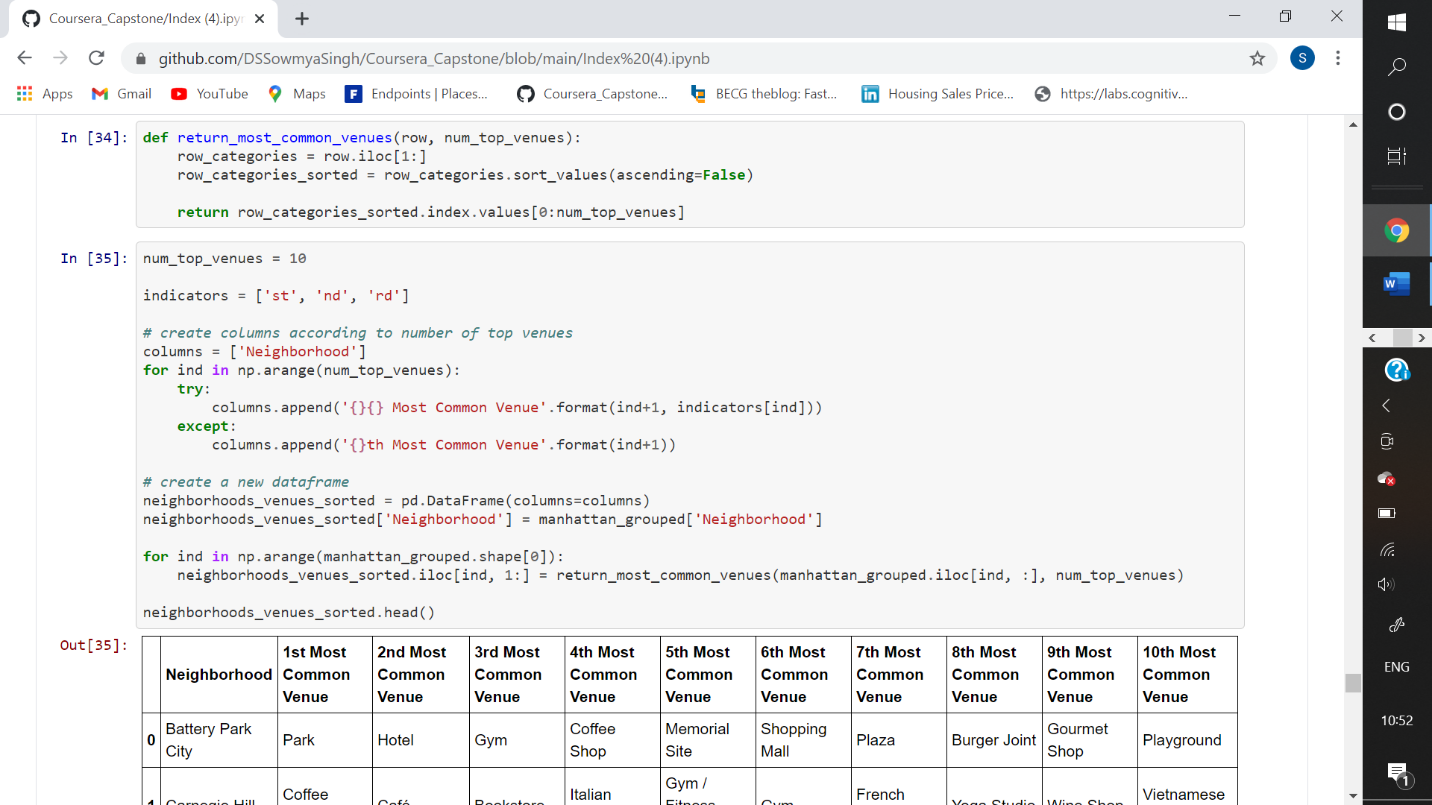
I then created a dataframe of all venues in Manhattan and counted the number of venues in each neighborhood.



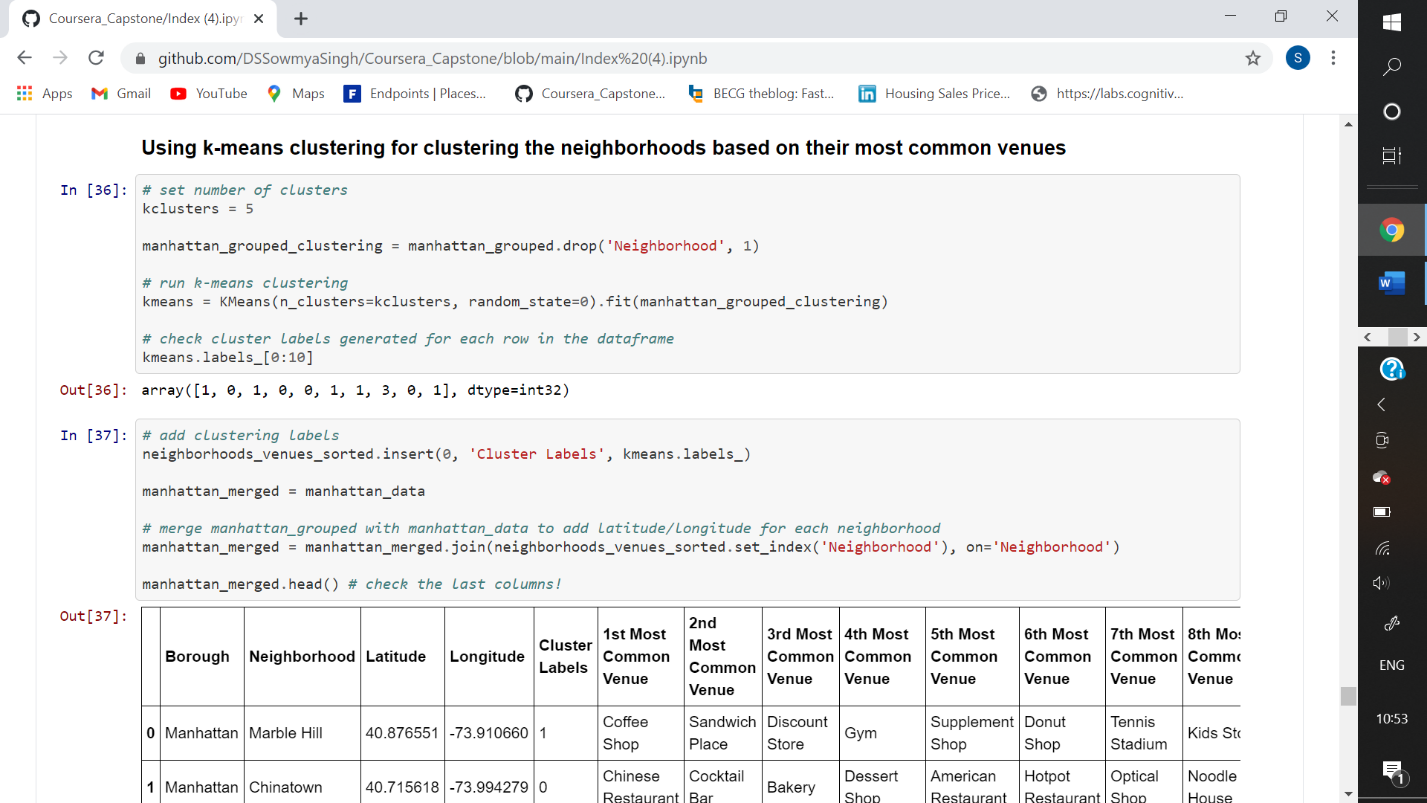
Then I counted the total number of unique categories of the venues in Manhattan and used one hot encoding to find the most common venues in each neighborhood.





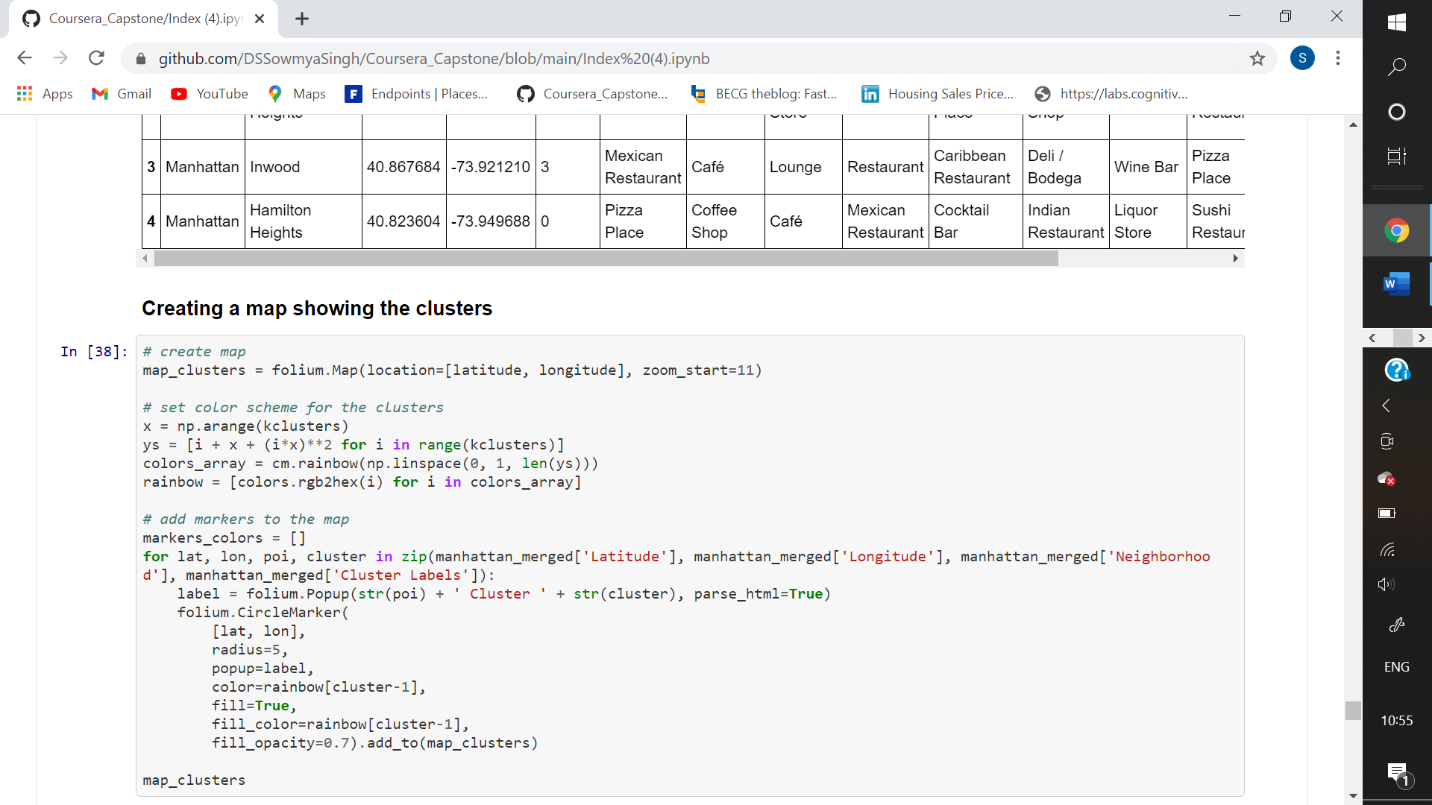


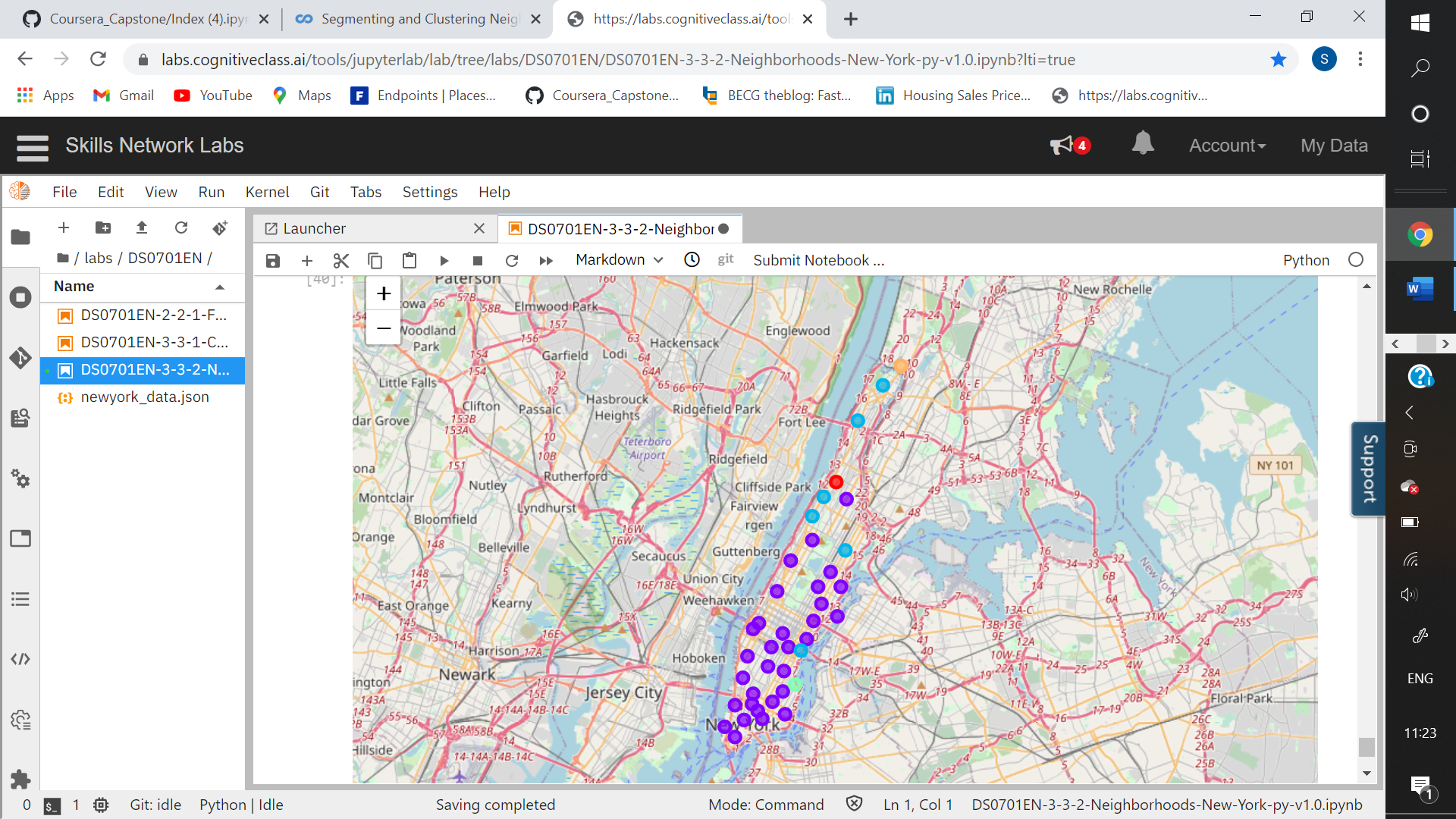
Then I used k-means clustering to cluster the neighborhoods based on their most common venues.



**Results :**

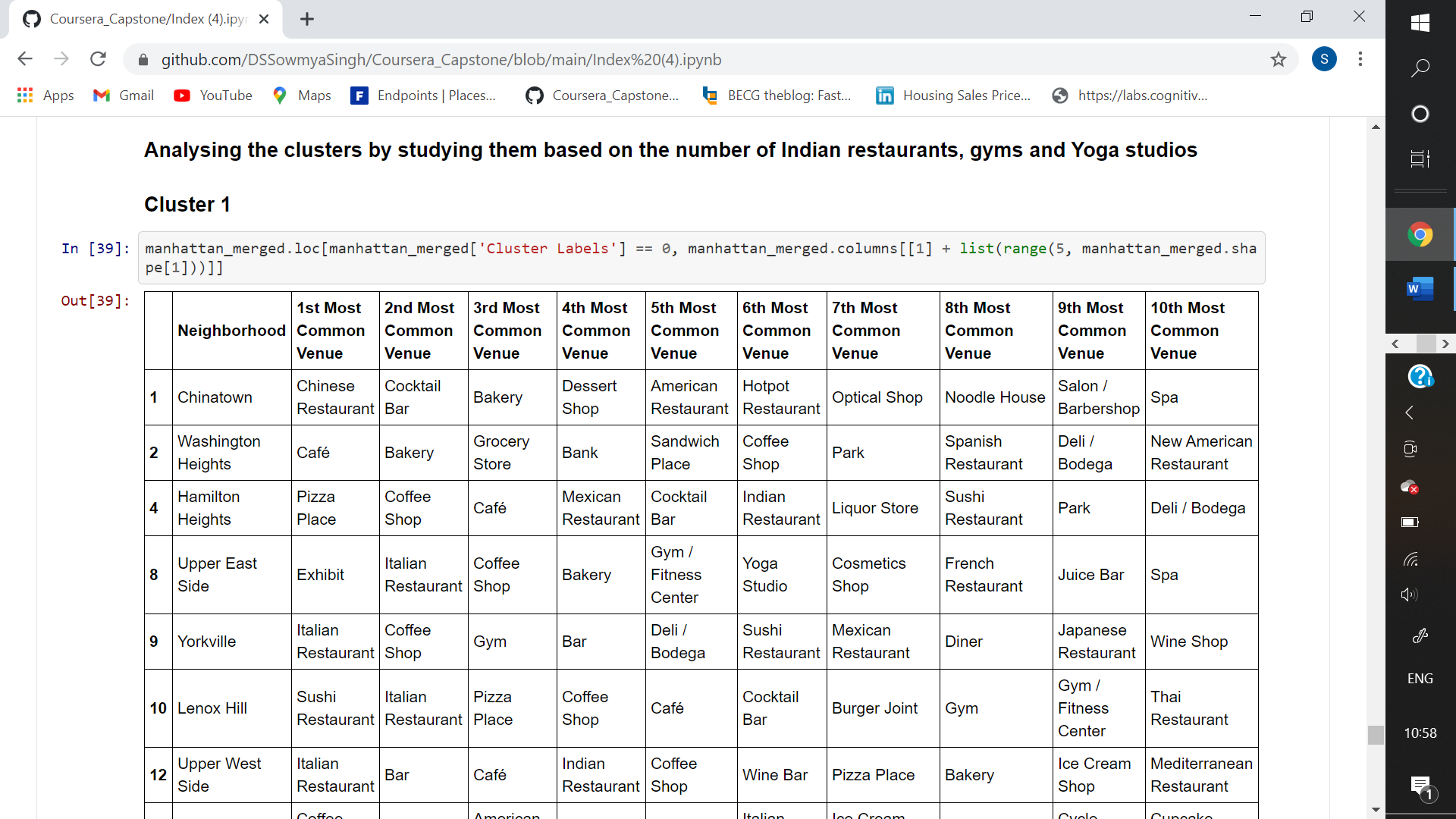
I used a map to show the clusters formed.





**Discussion :**

Then I analysed the clusters based on the number of Indian restaurants, Gyms and Yoga Studios in each cluster.



**Conclusion :**

By analysing the resulting clusters, it can be said that locations in **cluster 1** will be suitable for opening the Ayurvedic clinic, since it has 3 Indian restaurants, 4 Yoga studios and 12 Gyms/Fitness studios under the 10 most common venues. Cluster 2 has 2 Yoga studios and 1 gym, cluster 4 has 1 Indian restaurant and 1 gym whereas clusters 3 and 5 have only 1 gym under the 10 most common venues of their neighborhoods.