**Finding the most suitable area in Mumbai to open a restaurant**

**Week 5**

**Final Report**

**Introduction**

Mumbai is the financial and cultural capital of India. It is home to Indians of all states and is also a tourist attraction. What is more intriguing than its coastal location is the plethora of cuisines it hosts.

Be it vegetarian or non-vegetarian, the dishes in Mumbai boast of rich taste, fieriness and impressive flavors. The cuisine of Mumbai covers a large assortment of interesting, authentic dishes and zesty seafood dishes. The staple foods consumed by the residents of Mumbai include rice, aromatic fish curries, Indian bread (chapatis and rotis), vegetable curries, pulses and desserts. Coconuts, cashew nuts, peanuts and peanut oil are some of the major ingredients used in many of Mumbai's traditional dishes.

To monopolise on the hunger pangs of its residents, Mumbai is a rather great location for someone looking to enter the restaurant business. But being so populated, the high competition is obvious. Hence, it is sensible to analyse and settle for a location which would be most economically profitable.

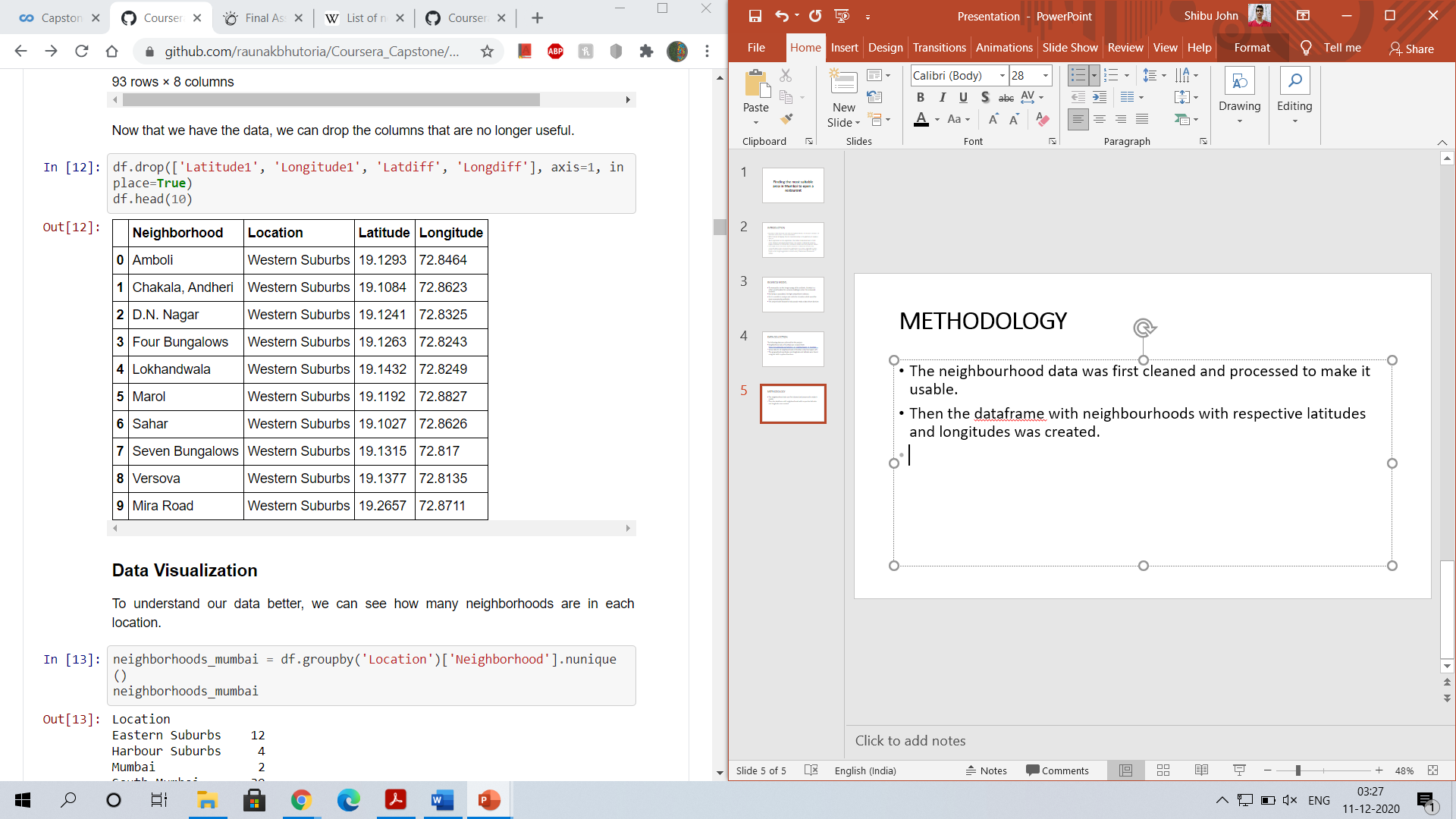
This project looks forward to help people make a data driven decision.

**Data Used**

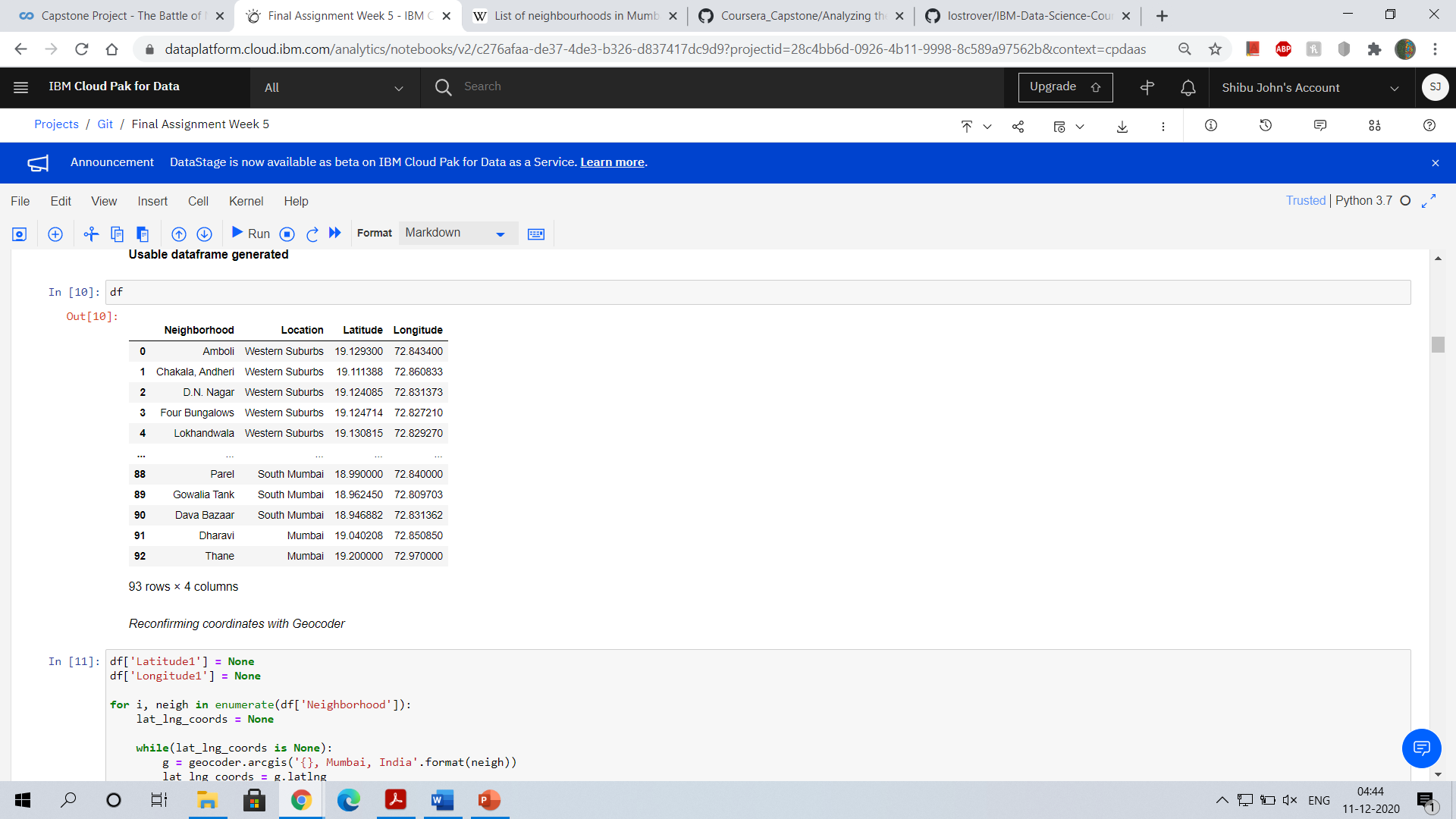
In this project neighbourhood data from Wikipedia (<https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Mumbai>) will be used. The geographical coordinates of these neighbourhoods will be required. Also the longitudes and latitudes of these neighbourhoods will be required as well. This data can be occupied using the built in Python libraries Geopy and Geocoder.

The venue data with recommendations will be collected using Foursquare API.

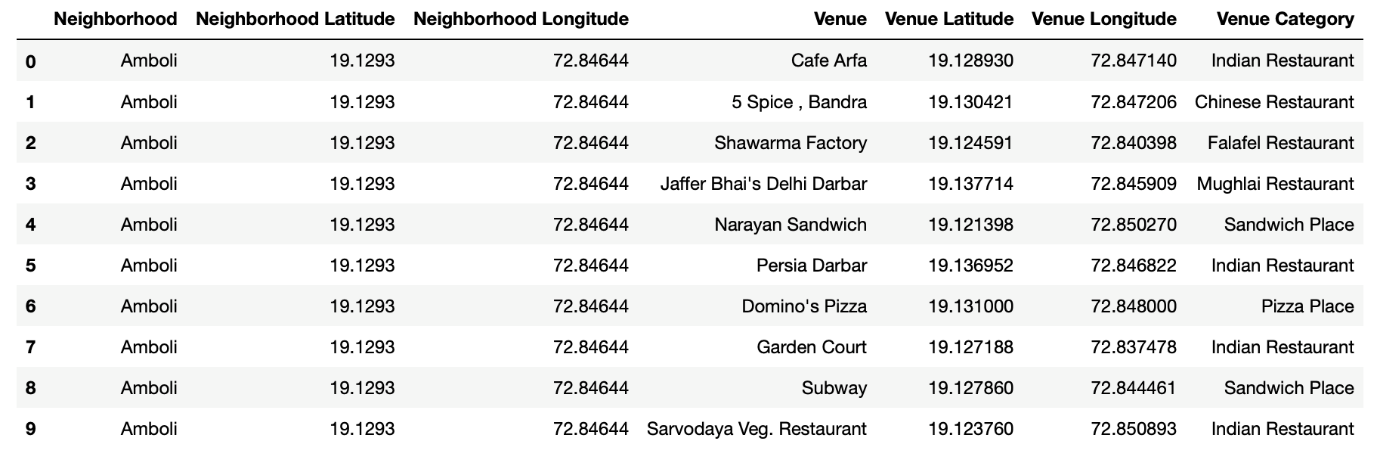
The following dataframe is the created initially after scraping the neighbourhood data from Wikipedia.



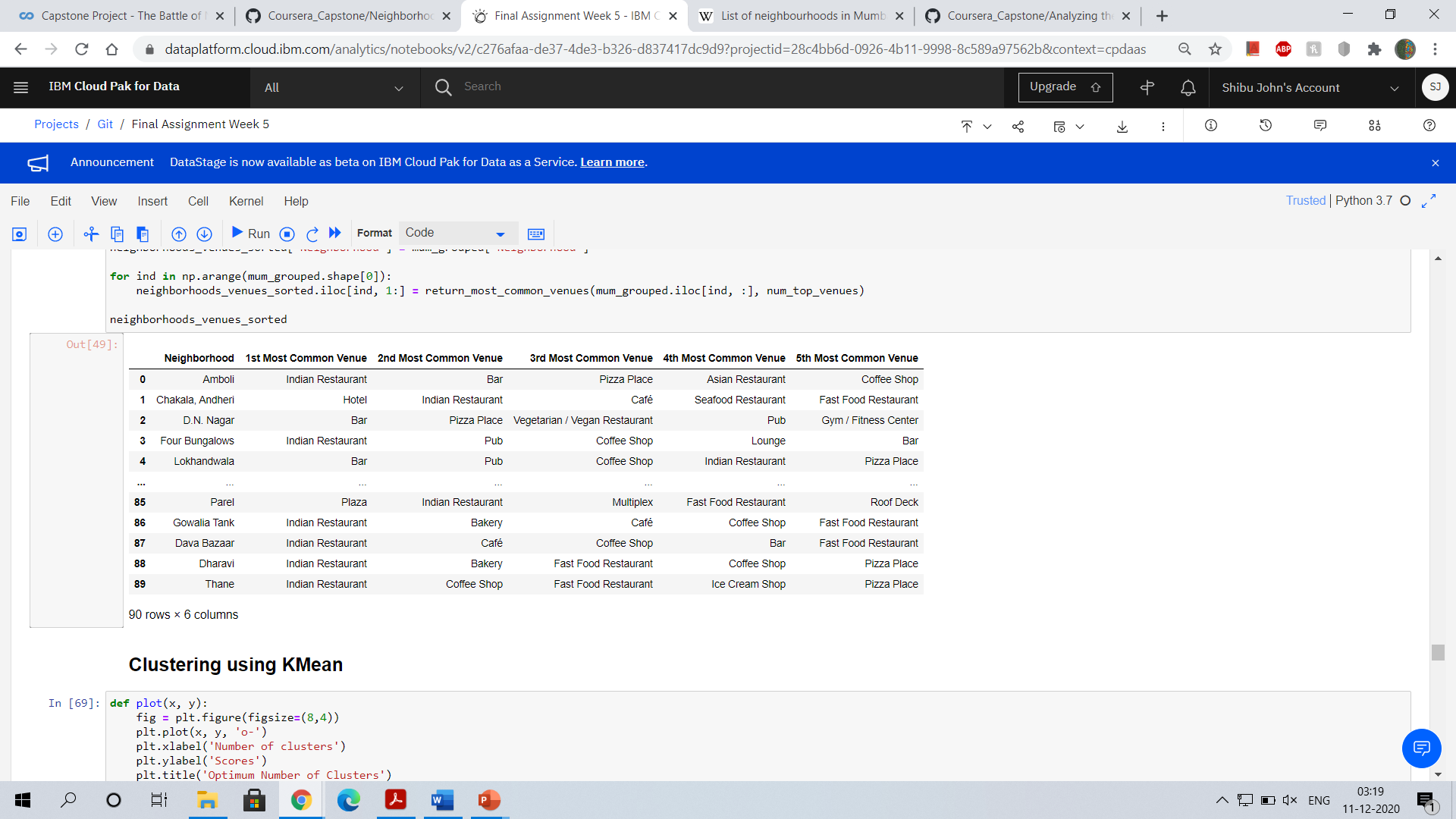
The following data was obtained after processing the above mentioned dataframe, to make it more usable.



The following data was obtained after processing the Foursquare API.



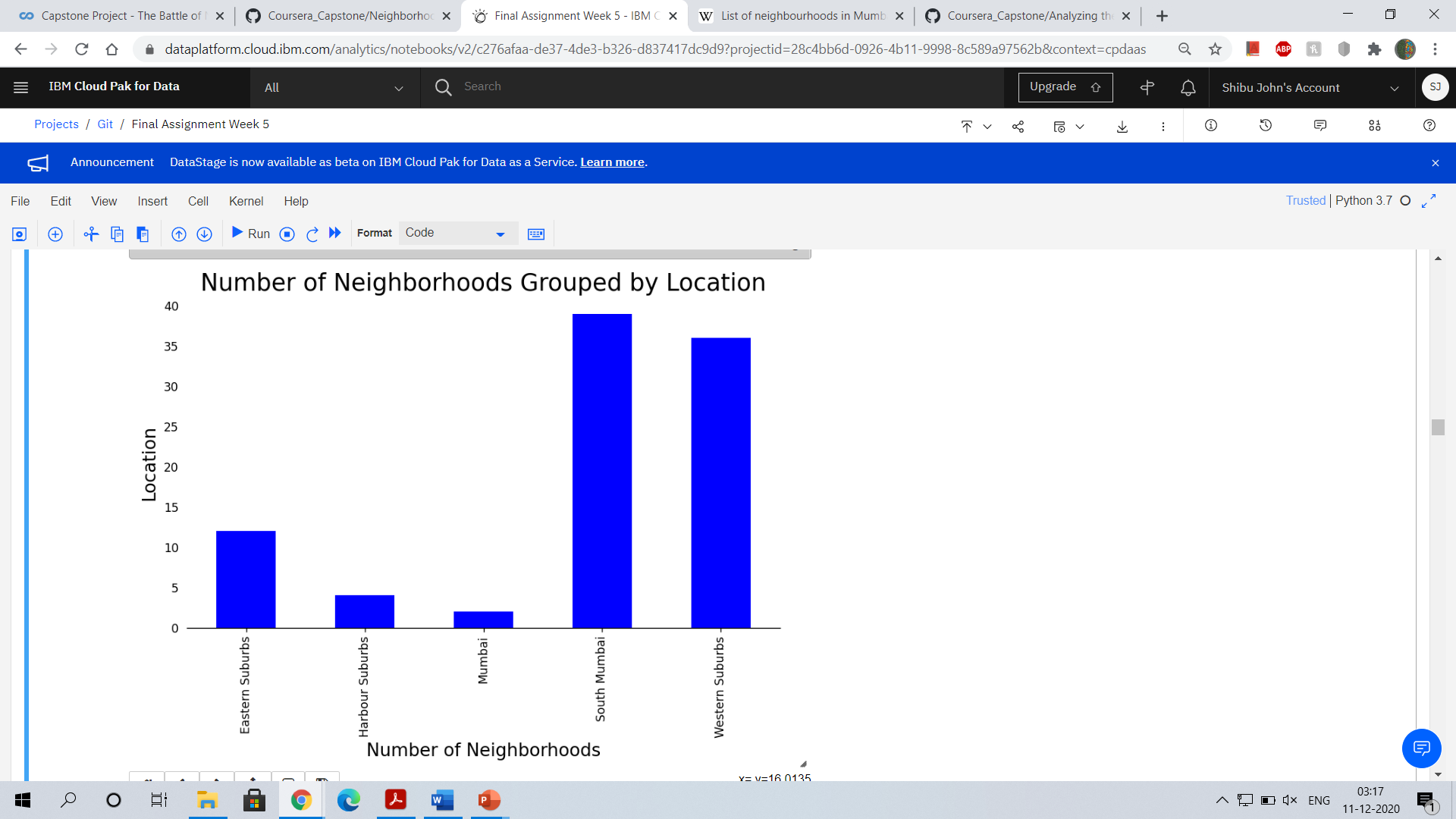
The following dataset was used for clustering



**Methodology**

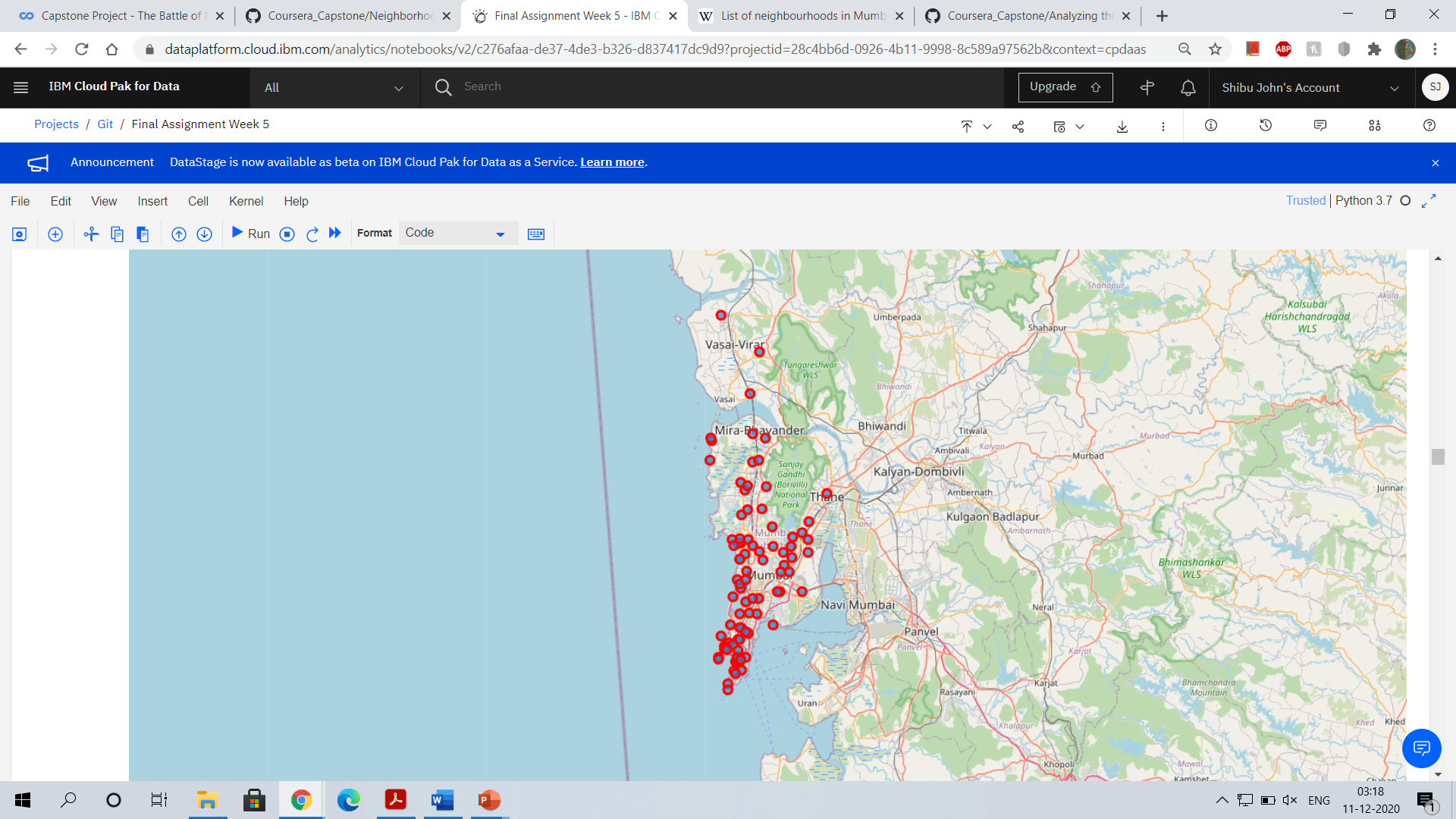
* The neighbourhood data was first cleaned and processed to make it usable.
* Then the dataframe with neighbourhoods with respective latitudes and longitudes was created.
* The area with maximum number of neighbourhoods was determined to be south Mumbai
* The venue recommendations were collected using Foursquare API.
* The dataframe with the top 5 common venues was created.
* Clusters were formed using KMean clustering method.

Data visualization



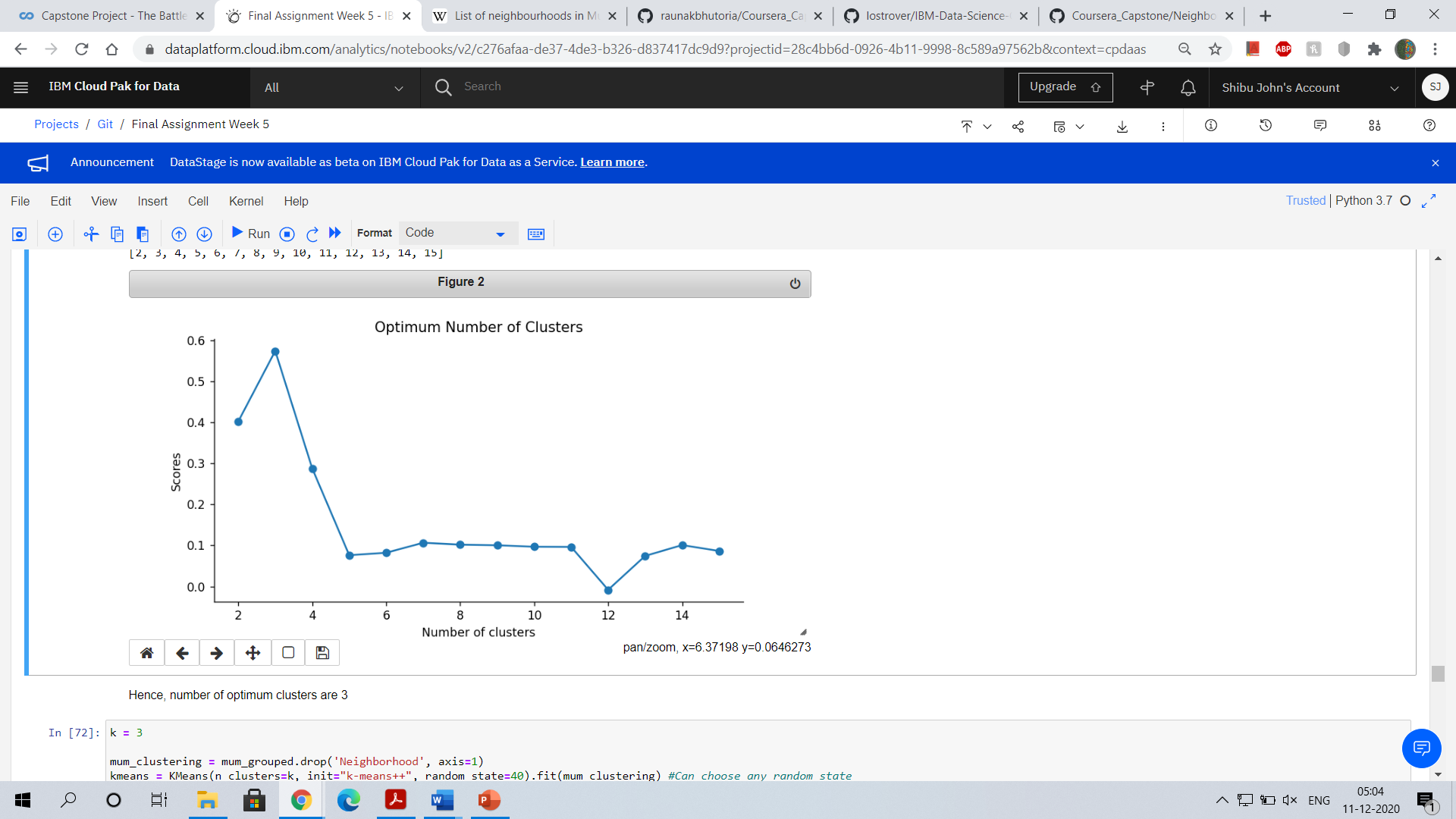
From the following graph it is visible that the south Mumbai area has the maximum number of neighbourhoods.

We can see how the neighbourhoods are distributed across Mumbai in the map below.



# Unsupervised Learning

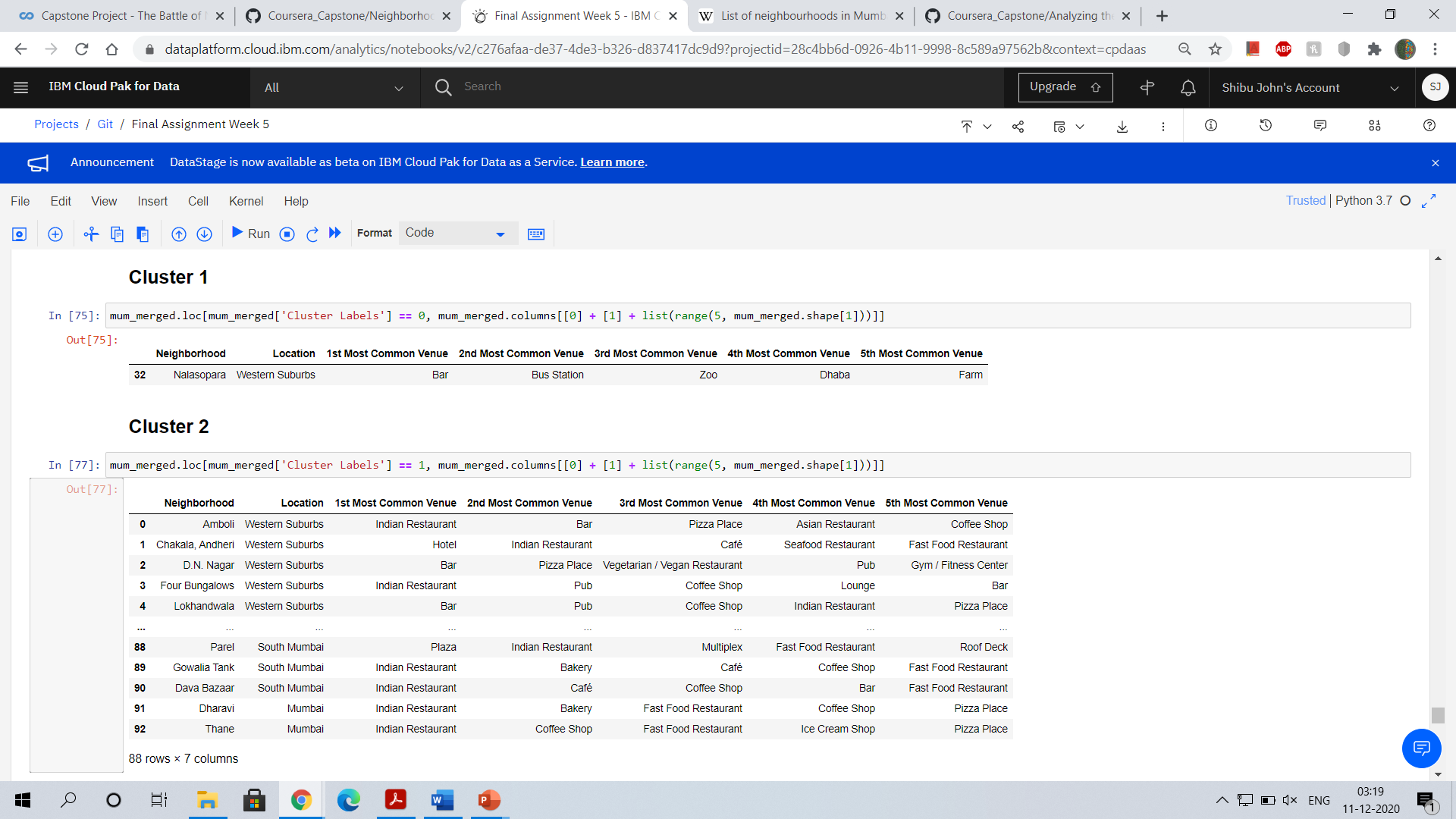
K-means unsupervised learning technique was used to cluster the neighborhoods based on the category of venues near the neighborhoods. One important aspect of the k-means model is to determine the number of clusters to use in model development.

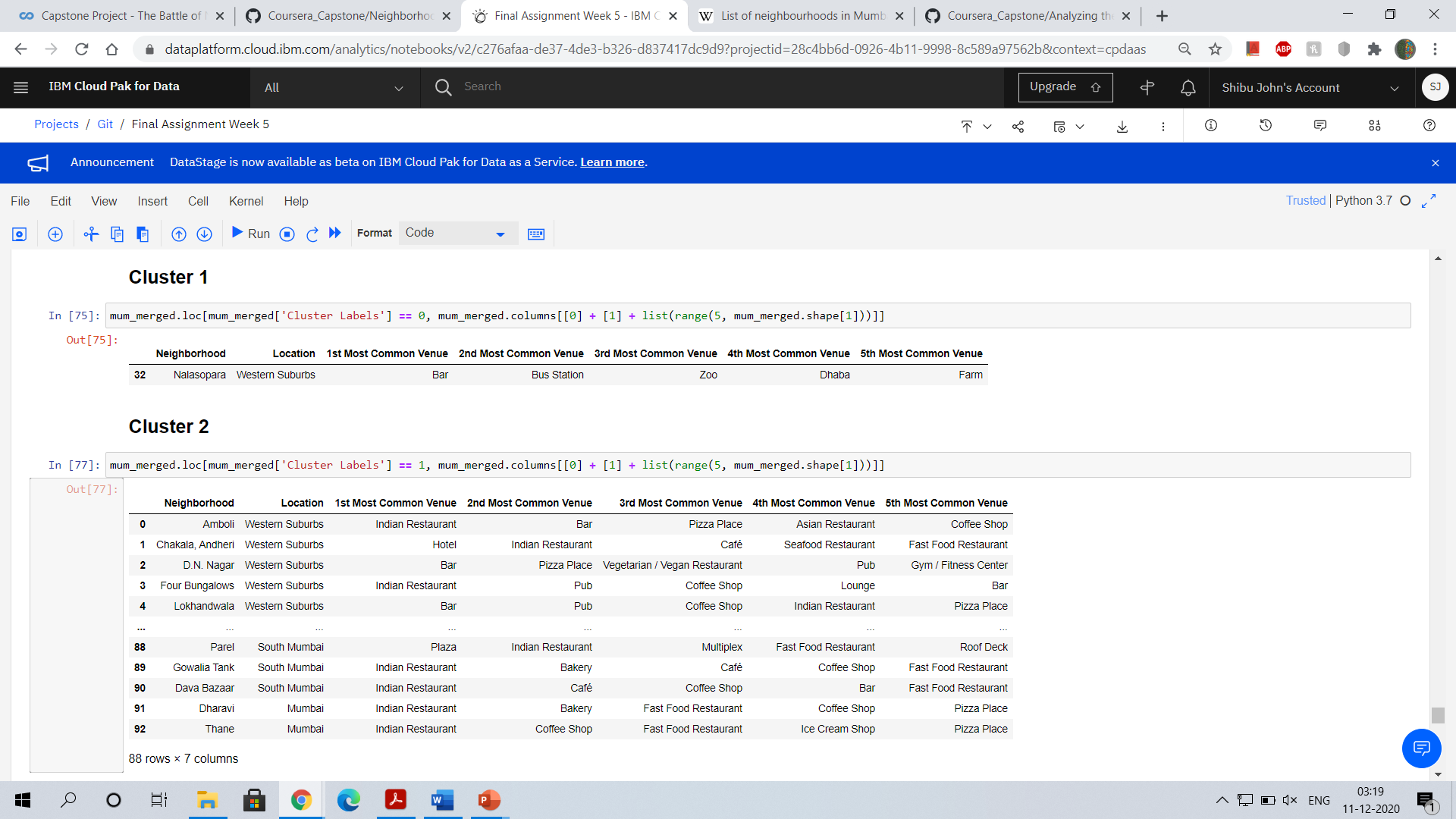


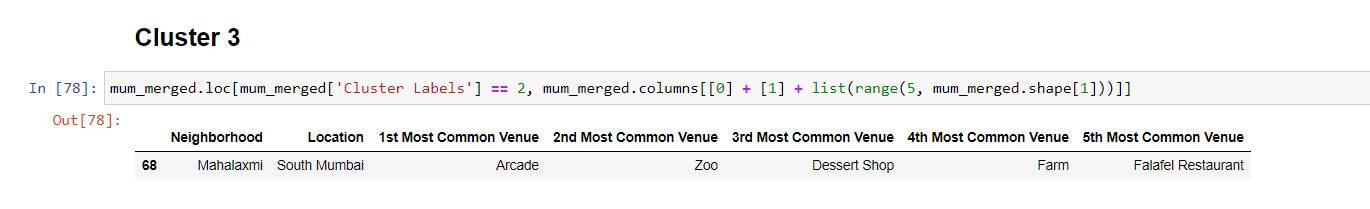
From the above plot it is evident that the optimum number of clusters are 3.

**Results**

The following three clusters are generated:

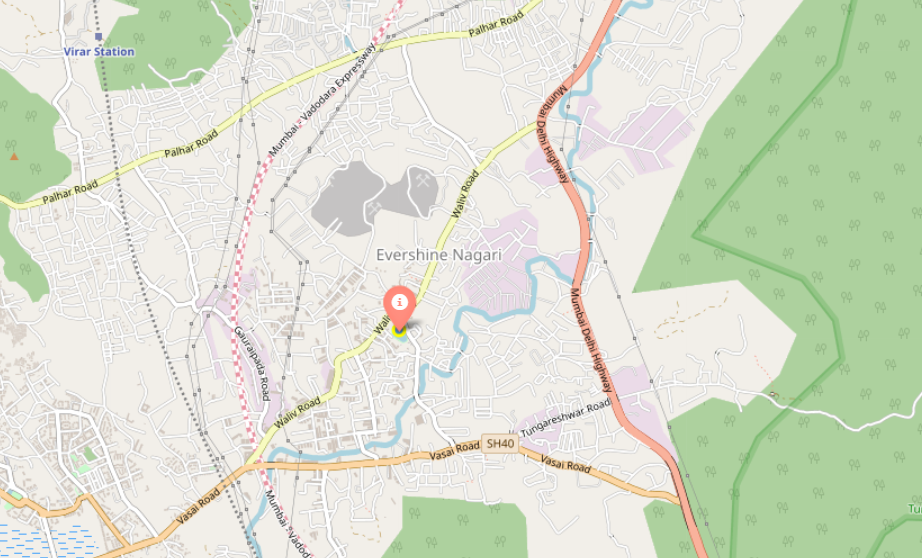






**Discussion**

From the clusters we can plot the map for the region having yielding maximum profits. As it is visible from the map, the region belonging to **cluster 1** is our required target.



Neighbourhoods in this cluster will be most suitable for opening a restaurant.

**Conclusion**

We have successfully analysed the data for the neighbourhoods in Mumbai. We have found the area in Mumbai where it is most suitable to open a new restaurant.