Market Analysis for an Indian Restaurant in Toronto

Table of Contents

[Introduction/Business Problem 2](#_Toc40574889)

[Target Audience 2](#_Toc40574890)

[Data 2](#_Toc40574891)

[1. Collecting Data 2](#_Toc40574892)

[2. Data Cleaning 3](#_Toc40574893)

[3. Data Exploration / Analysis 3](#_Toc40574894)

[4. Clustering 3](#_Toc40574895)

[Methodology 3](#_Toc40574896)

[Visualize Toronto Map 3](#_Toc40574897)

[Venues in Toronto using FourSquare 3](#_Toc40574898)

[Indian Demography from Statcan Portal 4](#_Toc40574899)

[Frequency Distribution of Indian Restaurants / Competitor Landscape 4](#_Toc40574900)

[Frequency Distribution of Indian Population / Demography 5](#_Toc40574901)

[K Means Clustering 5](#_Toc40574902)

[Find best k 5](#_Toc40574903)

[Clustering using k = 6 6](#_Toc40574904)

[Results 8](#_Toc40574905)

[Discussion 8](#_Toc40574906)

[Conclusion 9](#_Toc40574907)

# Introduction/Business Problem

Ajay is a successful restaurateur based in Mumbai, India. Under his brand AJ hospitality group, he has a chain of fine dining restaurants across India. He intends to expand his business and plans to start an Indian fine dining restaurant in Toronto, Canada in 2021. To begin with, AJ hospitality group approaches an analytics company in Toronto and assigns them with the task of conducting a market feasibility study.

The objective for the company is to understand and analyze the neighborhoods in Toronto suitable for AJ hospitality group to start their operations. They have to collect information on Indian population in different neighborhoods in Toronto and the competitors in the area. Finally, they intend to segregate the neighborhoods to different clusters based on similarities of factors mentioned above. This will help AJ hospitality group to make an intelligent decision by January 2021.

## Target Audience

Such a study would be beneficial for anyone planning to setup a restaurant in Toronto. The analytics company can make use of this model for their other restaurant clients who are interested in understanding the ethnicity-competitor landscape before setting up their base in Toronto.

# Data

The different processes involved in this stage are mentioned below

## 1. Collecting Data

We will use the below wikipedia page to gather the details of all the neighbourhoods in Toronto.

https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M

We will use BeautifulSoup python library for web scraping.

Geographical co-ordinates for the neighborhoods are obtained from the below link. https://cocl.us/Geospatial\_data

Foursquare API will be used to get the list of Indian restaurants in each neighborhood. This can be used to establish a competitor landscape

We will use 'Statistics Canada' portal to find out the demography details (South Asian population)

https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3520005&Geo2=PR&Code2=35&Data=Count&SearchType=Begins&SearchPR=01&B1=All

## 2. Data Cleaning

We will bring all the data from step 1 together, into a single dataframe which consists of neighborhoods, postal codes, geo-coordinates and South Asian (or Indian) population details.

## 3. Data Exploration / Analysis

We will use Foursquare API to explore the Indian restaurants in each neighbourhood and then we will group rows by neighborhood and then take the mean of the frequency of occurrence of Indian restaurants in each neighborhood. A similar exercise is performed for population as well

## 4. Clustering

The final stage would be clustering where in k-Means clustering technique to come up with multiple clusters based on competitor landscape and South Asian population.

This final stage will help AJ hospitality to make an intelligent decision as to where the restaurant should be located.

# Methodology

## Visualize Toronto Map

Initially, a map of Toronto was created with neighborhoods superimposed on top. I used Folium library for map visualization.



## Venues in Toronto using FourSquare

I used FourSquare API to list all the venues in the Toronto neighborhoods. Further, only venues with the keyword ‘Indian’ was filtered out. A percentage of this was calculated and appended to the dataframe



## Indian Demography from Statcan Portal

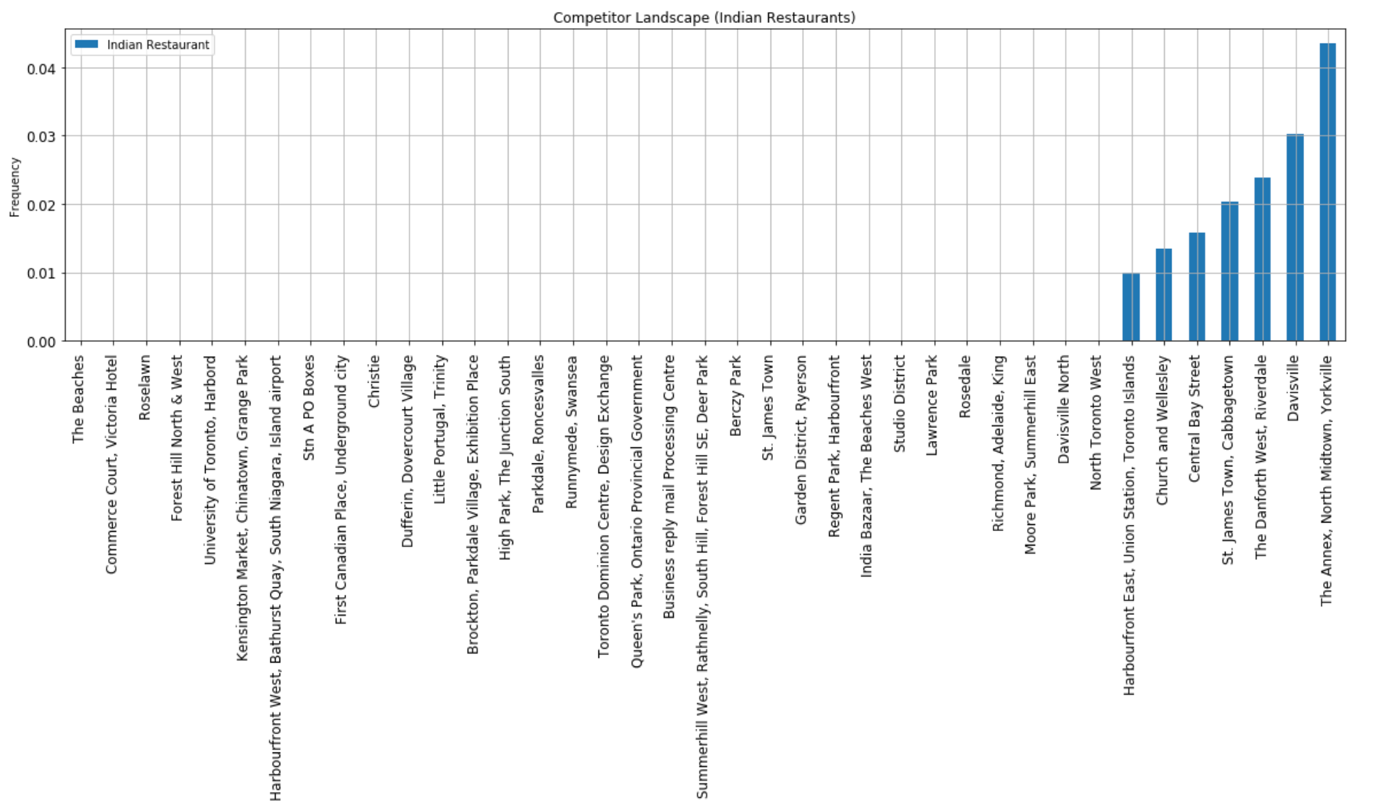
From the statcan portal (<https://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/FullFile.cfm?T=1201&LANG=Eng&OFT=CSV&OFN=98-310-XWE2011002-1201.CSV>), the Indian dwelling information was retrieved as CSV. This is based on the postal codes.

This was merged with the neighborhood dataframe to form the full dataset which includes ‘neighborhood’, ‘Indian Restaurant Percentages’ and ‘Dwellings’ information.



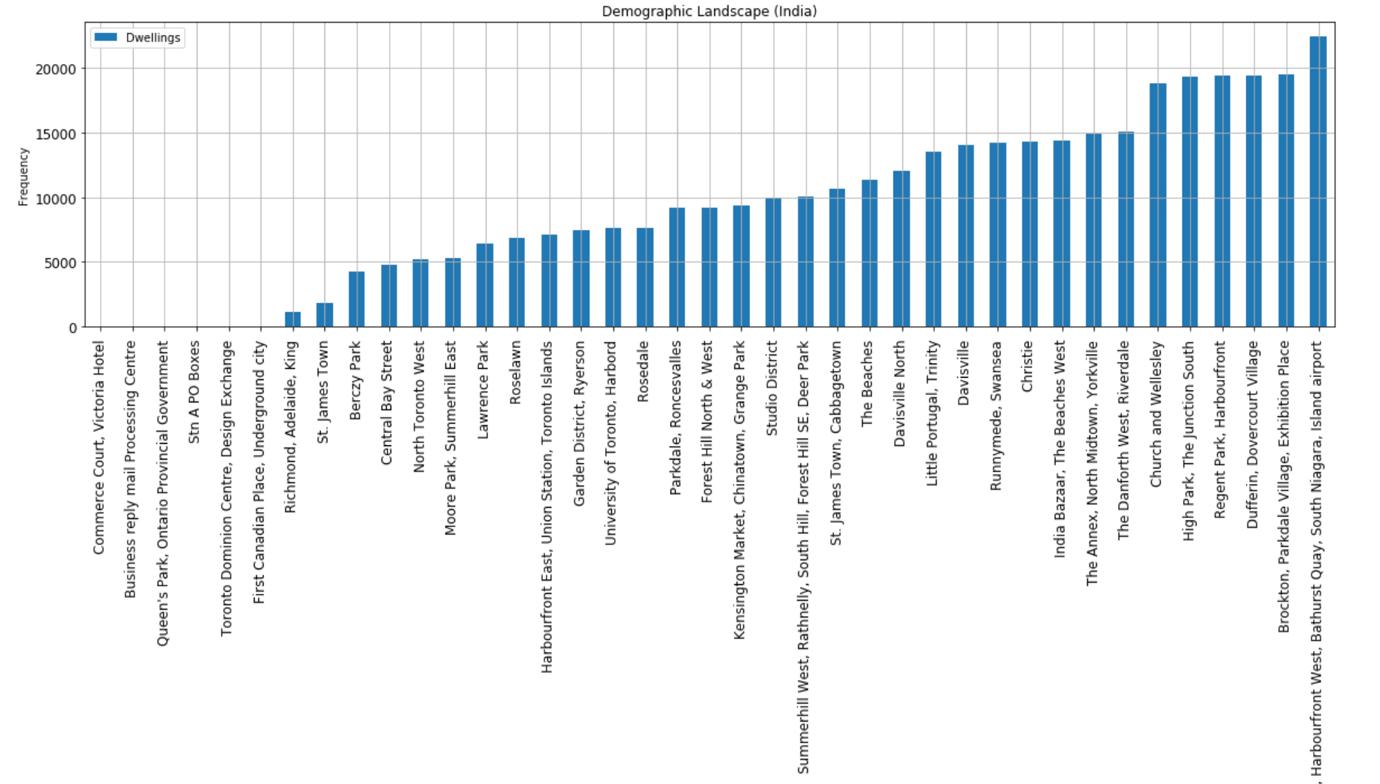
## Frequency Distribution of Indian Restaurants / Competitor Landscape

The data from the dataframe above was used to plot the competitor landscape (as bar plot)



## Frequency Distribution of Indian Population / Demography

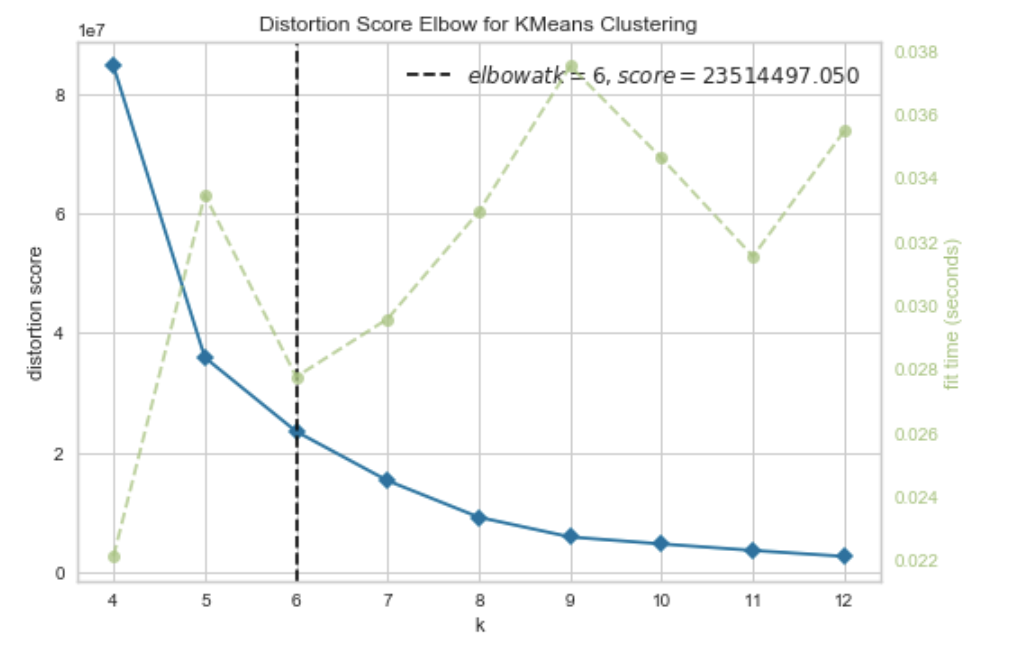
Similarly, a bar plot was plotted to visualize the Indian demography details in each neighborhood



## K Means Clustering

### Find best k

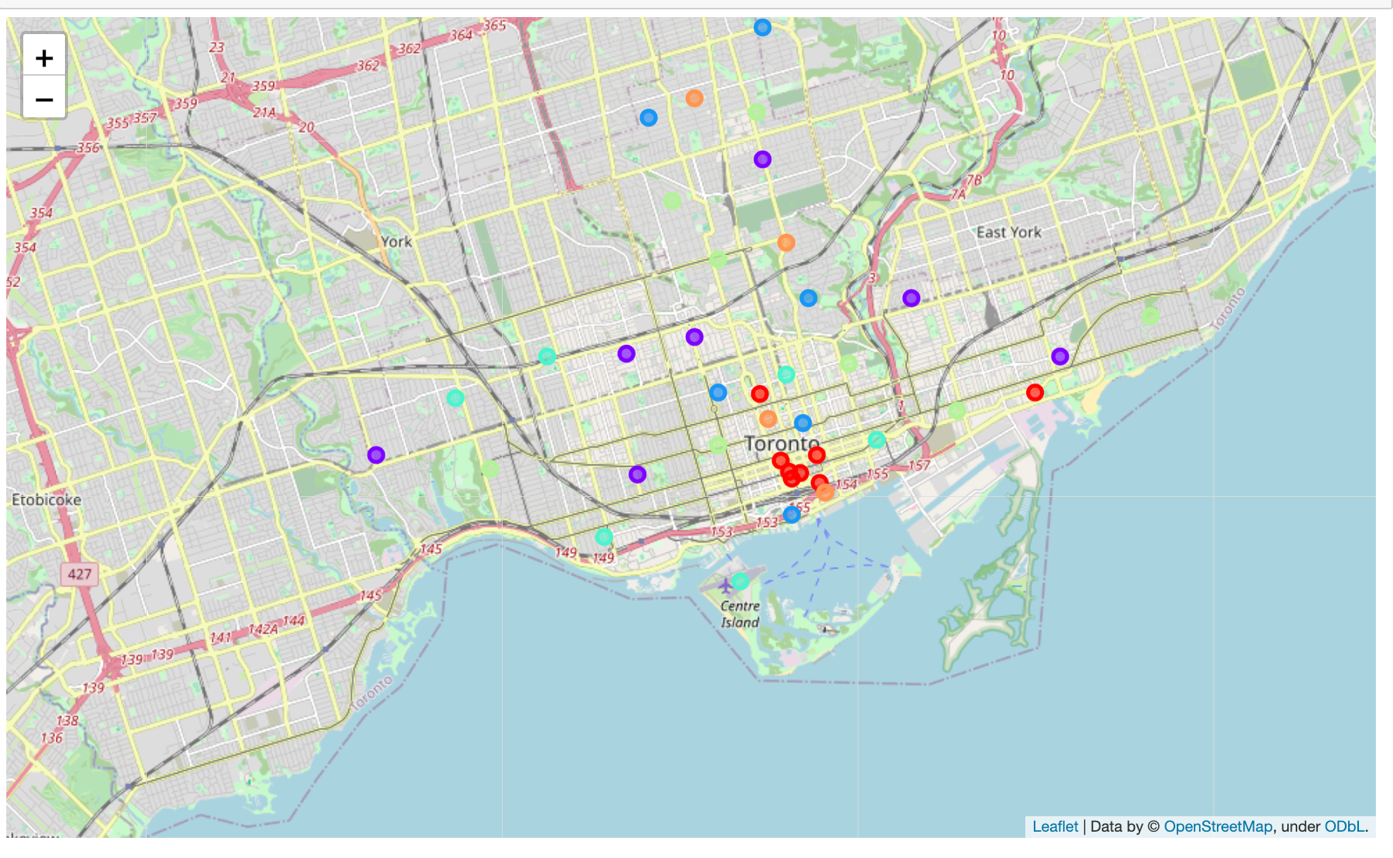
The best k was found out using the KElbowVisualizer from yellowbrick library



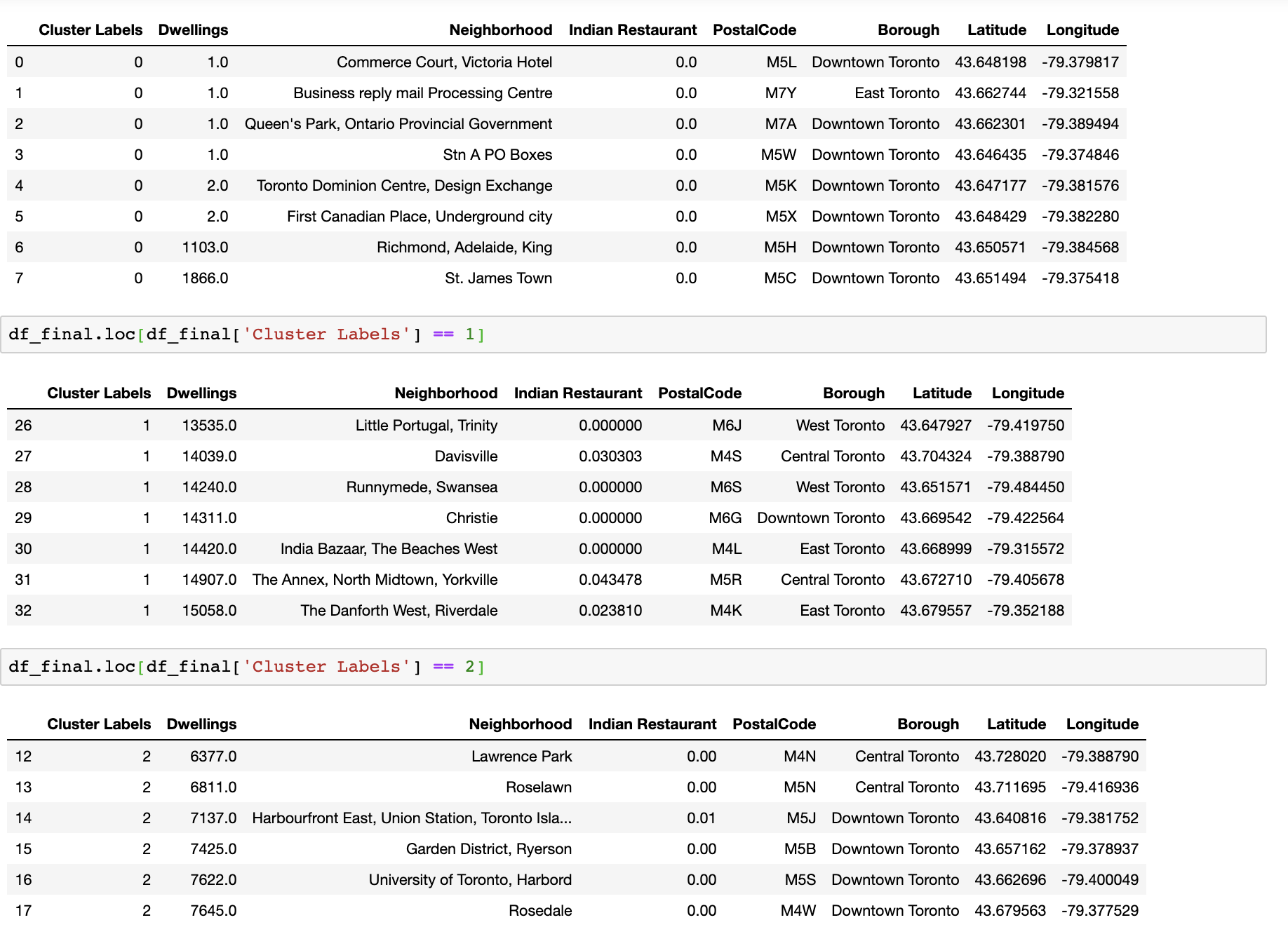
From here, we can see that k = 6

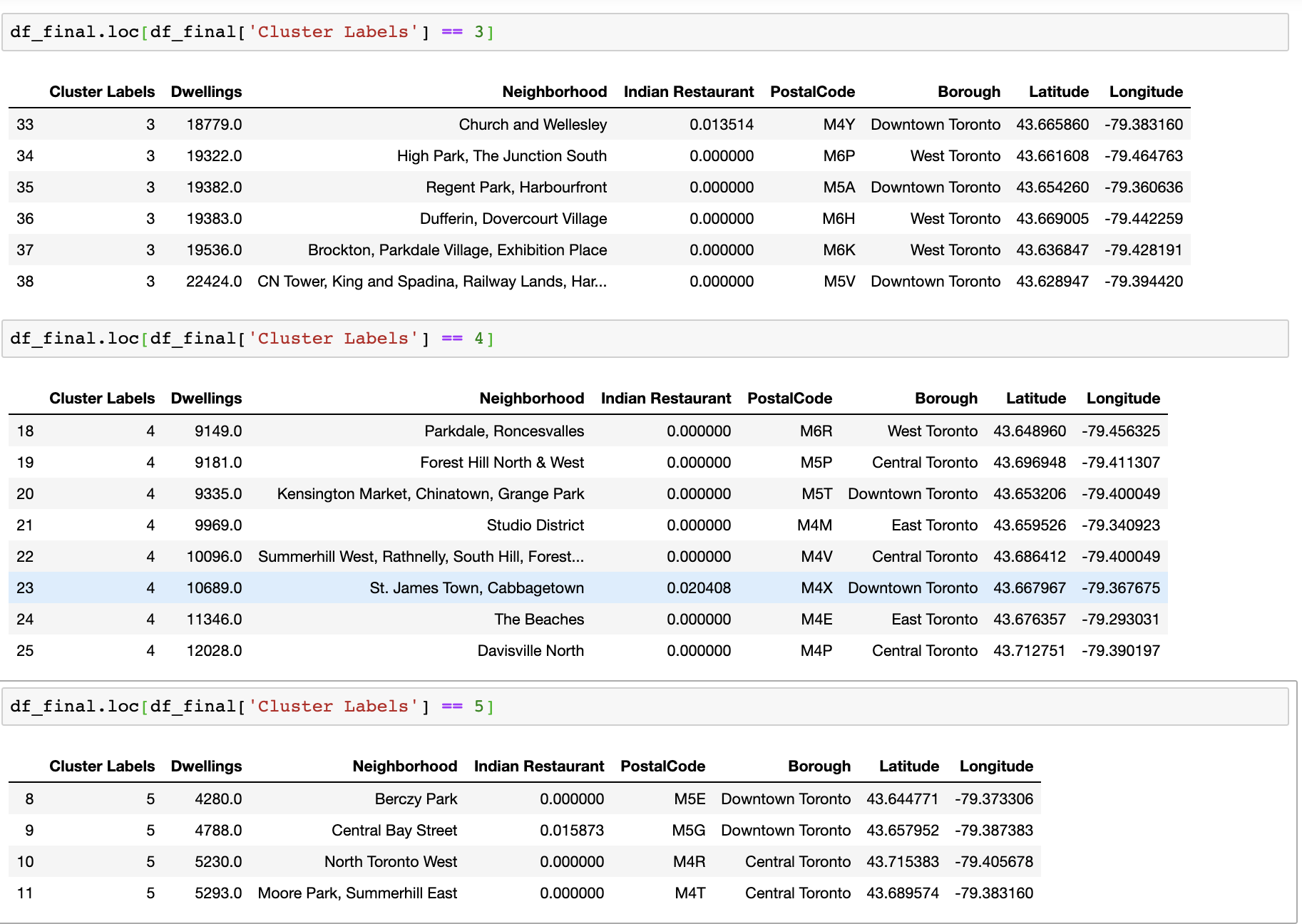
### Clustering using k = 6

Now that the k is defined, I used k Means clustering to come up with 6 clusters. This was then merged to the intial dataframe to get latitude and longitude information which could then be plotted on a map. The different clusters formed are shown below:



Finally, we can examine the 6 different clusters.





# Results

6 different clusters were formed

* Cluster 1 had low number of Indian dwellings and low number of Indian restaurants as well.
* Cluster 2 had relatively high Indian dwellings and a moderate number of competitors.
* Cluster 3 had relatively low to medium dwellings and low number of competitors.
* Cluster 4 had high number of dwellings and low number of competitors.
* Cluster 5 had medium dwellings and low competitors.
* Finally, Cluster 6 had low number of dwellings and low competitors.

# Discussion

* Cluster 1 had low number of Indian dwellings and low number of Indian restaurants as well. Therefore, this cluster would be a **poor choice and shouldn’t be considered.**
* Cluster 2 had relatively high Indian dwellings and a moderate number of competitors. So, this cluster could be a **possible choice.**
* Cluster 3 had relatively low to medium dwellings and low number of competitors. So, this cluster could be **considered; but not a good option though.**
* Cluster 4 had high number of dwellings and low number of competitors. So, this could be a **good option.**
* Cluster 5 had medium dwellings and low competitors. So, this cluster could be **considered.**
* Finally, Cluster 6 had low number of dwellings and low competitors. So, this cluster **shouldn’t be considered.**

Additional parameters like income of the groups and population of other South Asian groups (Sri Lankan, Bangladeshi etc.) could increase the accuracy of the model. So, this can be considered as a future improvement to the model.

# Conclusion

Based on the results from the k-Means clustering model, clusters 4,2 and 5 should be possible options for the client – AJ hospitality group. But, because of the fact that there are relatively low number of competitors in cluster 4, that could be the best option for AJ hospitality group to setup their operations in Toronto (Location shown below).

