The Best Place to Open an Italian Restaurant in Toronto, Canada

Introduction

Business problem
Italian cuisine is undoubtly one of most famous cuisine. However, from the perspective of an investor or a chef dreaming of opening his/her own italian restaurant, it is a remarkable challenge in getting into the market due to the highly saturated supply of Italian food.
The objective of this project is to find the best neighborhood in Toronto to open an Italian restaurant. In this project, we'll be trying to find ideal places to start a Italian restaurant using Foursquare location data and clustering analysis.

Target Audience
Italian chef's trying to open their own restaurant. Investors looking to establish a business related to Italian food. Tourists who are interested in Italian cuisine.

Data Description

The data that will be used in the project will be as follows-.
Data Source 1 – Neighborhood Data The dataset about neighborhoods of toronto will be extracted from a Wikipedia page. Features such as borough, neighbourhood, postal code will be extracted.
Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Data Source 2 – Geographical Coordinates Geographical coordinates for each neighborhood will be obtained with the aid of GEOPY Library such as latitude and logitude.
Data Source 3 – Venue categories I will use the Foursquare API to retrieve venues, using the coordinates obtained in Data Source 2 above.

Methodology

In this section, we prepare the dataframe to be use for the modelling.
The summary of the flow is:

1. Webscrape the data from city info webpage

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

2. Get and connect the Postal Codes to the neighborhoods

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

3. Check the business venues in the vicinity of the neighborhood using Foursquare

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Central Bay Street	43.657952	-79.387383	KFC	43.660400	-79.383900	Fast Food Restaurant
Berczy Park	43.644771	-79.373306	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant
North Toronto West	43.715383	-79.405678	Roots	43.716194	-79.400661	Clothing Store

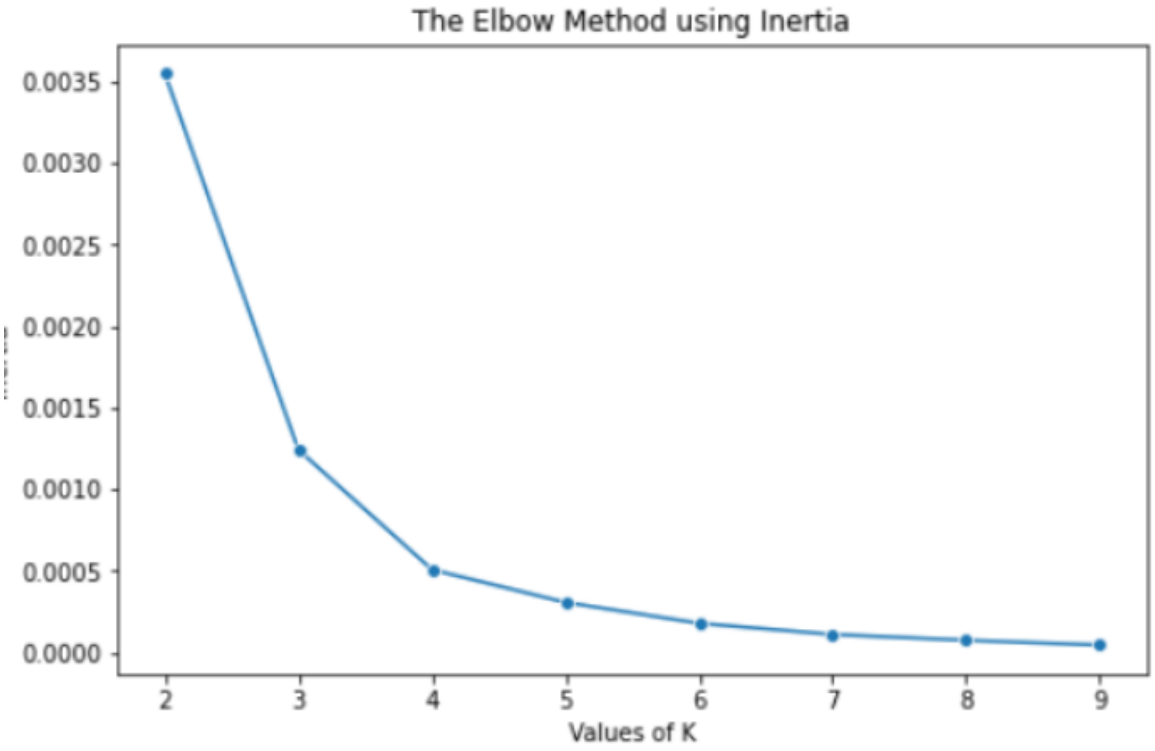
4. Use K-means clustering to group the neighborhoods

Analysis and discussions

Taking account of only Italian restaurants

	Neighborhood	Italian Restaurant
0	Berczy Park	0.017241
1	Brockton, Parkdale Village, Exhibition Place	0.041667
2	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
3	Central Bay Street	0.047619
4	Christie	0.062500
5	Church and Wellesley	0.012658

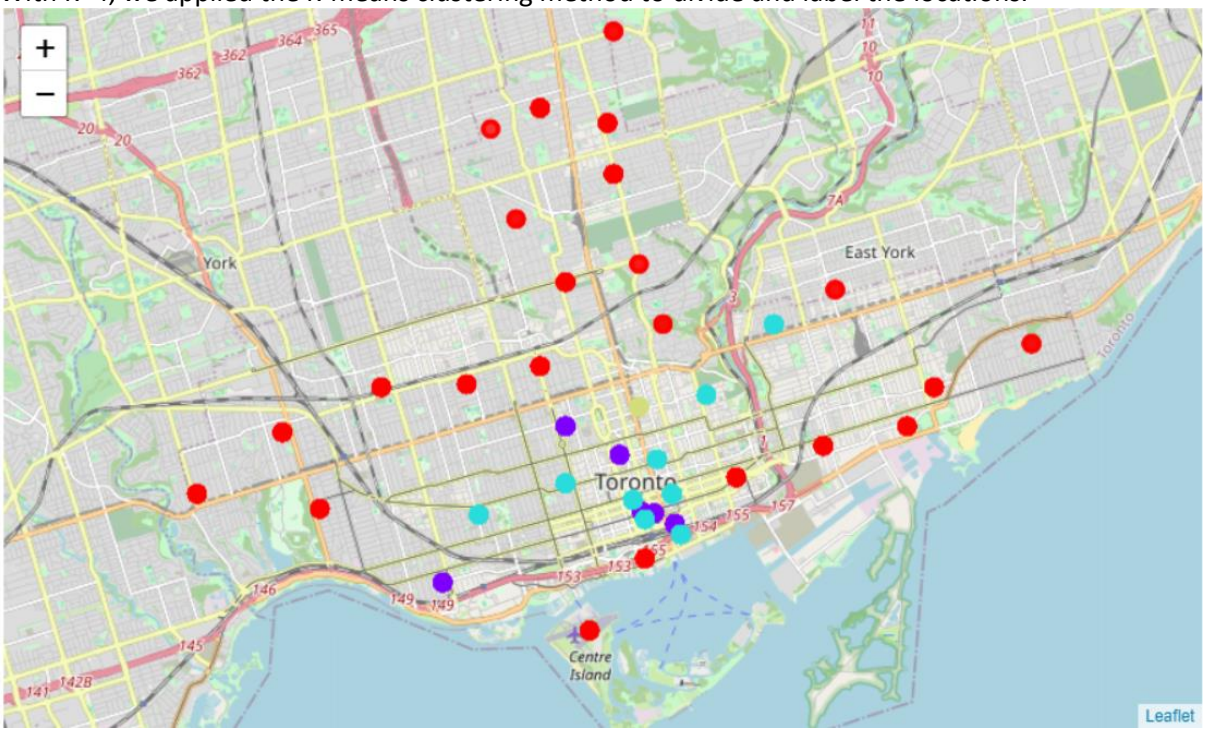
Use K-means clustering to group the neighborhoods



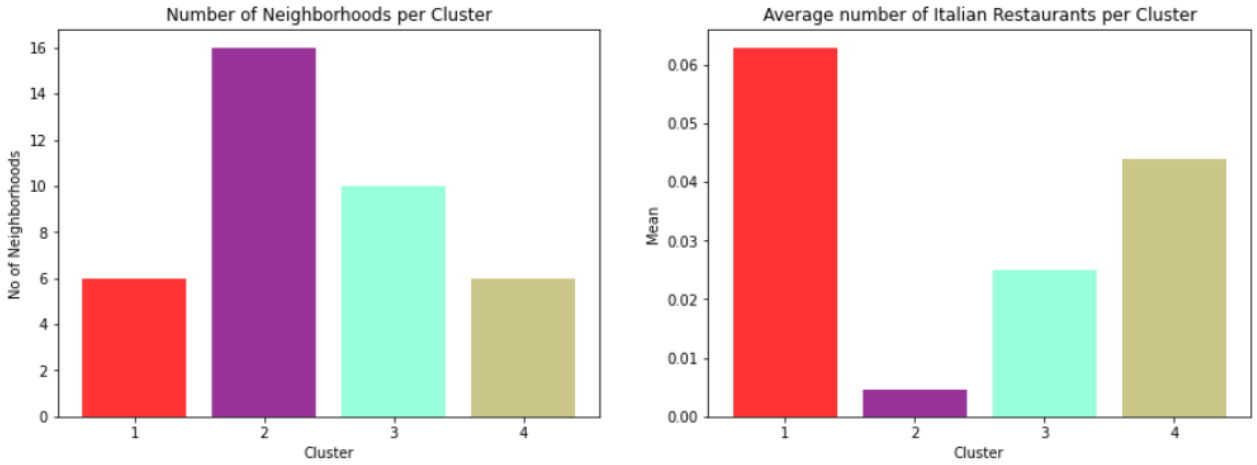
Our results show that the optimal K value is 4. Hence, K=4 will be used to divide our dataset.

Results

Applying the K-mean clustering
With K=4, we applied the K-means clustering method to divide and label the locations.



```
Text(0.5, 1.0, 'Average number of Italian Restaurants per Cluster')
```



Conclusion

From the perspective of a tourist, it may be best to go to the neighborhoods in cluster 1 to eat Italian food. There are a lot of Italian restaurants that will cater to the adventurous pallete. From the perspective of a businessman, the optimal location to open a new Italian restaurant is in cluster 2, that is in neighbourhoods such as Kensington Market, Chinatown, Grange Park, Harbourfront East, Union Station, Toronto Islands,Garden District, Ryerson where there are a lot of potential demands for an Italian restaurant while having the least competition around the vicinity. Of course it might be possible that these neighborhoods are less commercialize as compared to cluster 1.