Capstone Project — the Battle of Neighborhoods

Italian restaurants in Toronto –Final report

By: Aljawharah Alareefi

Introduction:

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 . In 2016, foreign-born persons made up 47 per cent of the population, compared to 49.9 per cent in 2006.

Food is a big deal in Toronto. Not only do we need it to live, but also we have come into our own as a foodie destination. New restaurants are constantly opening, and chefs continue to push culinary boundaries to come up with new and innovative ideas.

In this project, we will try to analyze all the **Italian Restaurant** currently present in Toronto's different neighborhood and will able to provide an final result/analysis at the end of this project. This study can help to decide the preferred location to guarantee the best success for opening new Italian restaurant.

Business problem:

The business problem is that we are looking for a location in the city of Toronto to open an Italian restaurant. The study must include the all the Italian restaurants and how they are distributed in the city and the places where they are concentrated.

We will analyze the data of Toronto's Borough and Neighborhoods and see where the Italian Restaurant are in highest number and in high demand and then we will compare all the Italian Restaurant according to their Likes, Tips and Rating. We will visualize all the great Italian Restaurant in different types of graphs and map. Then try to decide the preferred location to guarantee the best success for the new restaurant.

Target Audience:

- Business Analyst :who are interested to invest and launch new Italian Restaurant in Toronto's different Neighborhood, will get an idea about currently in which neighborhood people are loving Italian food and where new Italian Restaurant needs to be opened to let people have the taste of Italian Food. Also, where there will a tough competition between other great Italian restaurants who are currently running in smooth way and if he sets a plan to open a new Italian restaurant then this comprehensive guide will help him/her to know the prime competitors.

- Foreign-born population can look into our analysis and find out the top place to visit for Italian Cuisine in Toronto and enjoy their meal.
- Tourists :who are eager to taste various cuisines including Italian dishes.

Data Description:

The data needed for this project is required for analyzing the restaurants in Toronto. The data obtained is related to the Toronto Area, its central area, downtown, east and west areas, the postal codes of the neighborhoods and their latitude and longitude. With the help of this data, we get restaurants' data from Foursquare API along with each latitude and longitude.

Data Sources and tools:

1- For the Toronto neighborhood data, this Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M contains Postal Code, Borough, and Neighborhood of Toronto. I will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a panda's data frame using BeautifulSoup package.

2- Then get the latitude and the longitude coordinates of each neighborhood from https://cocl.us/Geospatial_data.

	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	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

3- Then Foursquare APIs will be used to fetch the Venues of all Italian restaurant in Toronto city and their rating, likes and tips information data.

Methodology:

1- **Gathering the Data:** Importing all the python packages and libraries that will be used in this project. I have extracted the data as per the direction mentioned above through Web scraping and geopy client etc.

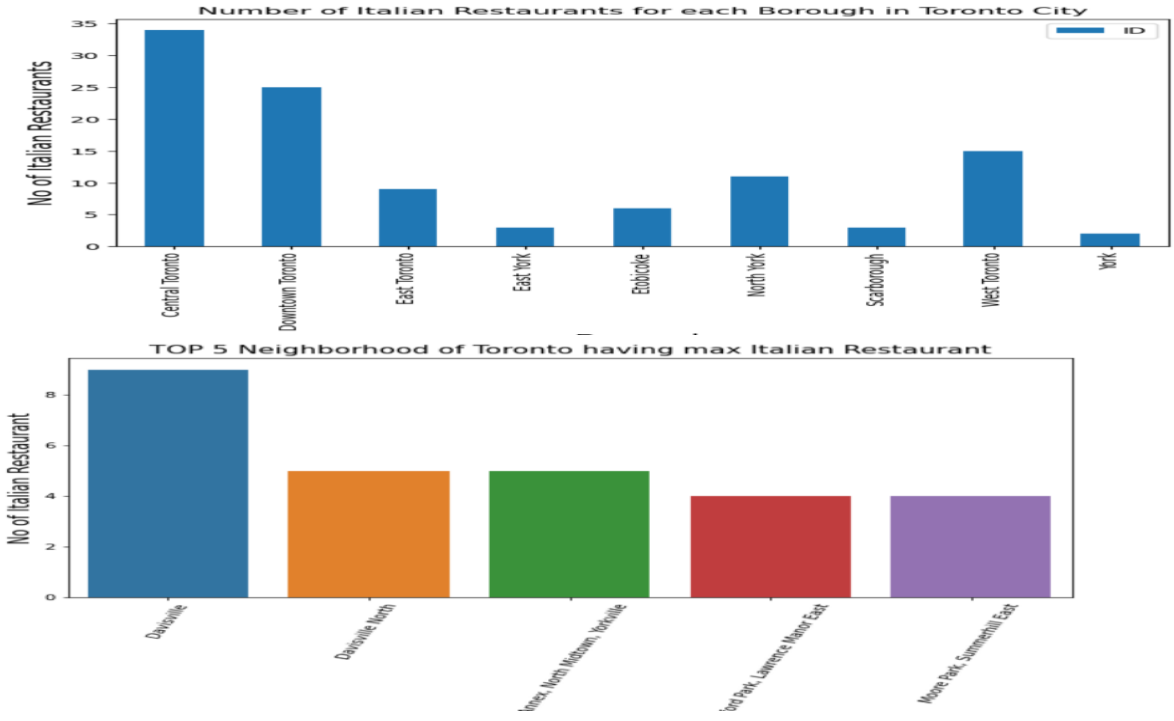
	PostalCode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

2- **Cleansing Data**: We will then clean the data according to our need remove unassigned values in borough or unwanted features and merge different neighborhood with same borough etc.

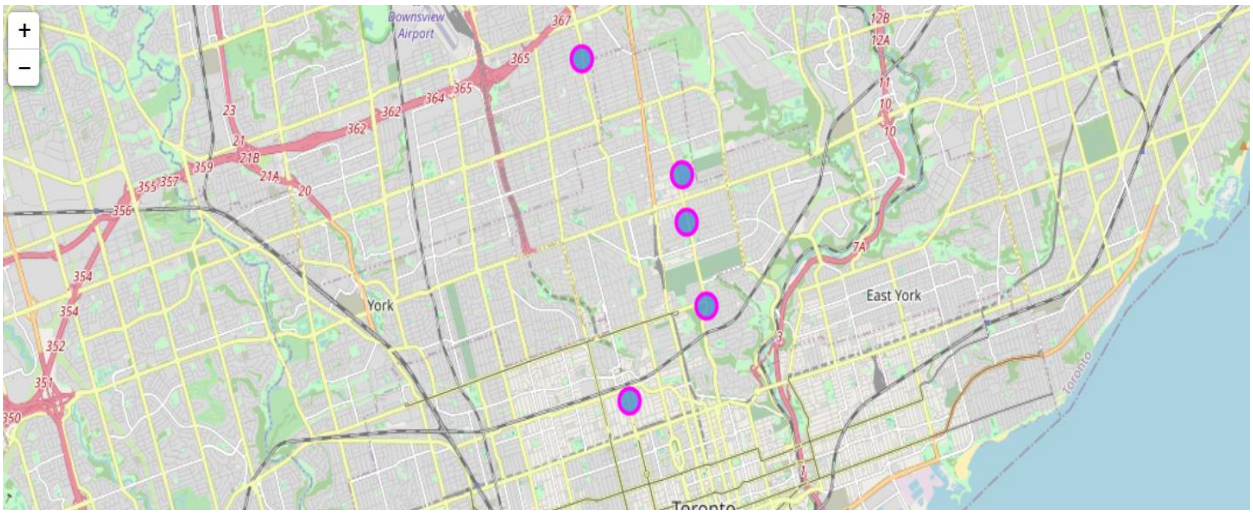
	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

3- **Retrieve the Italian restaurant**: Foursquare APIs used to fetch the Venues of all Italian restaurant

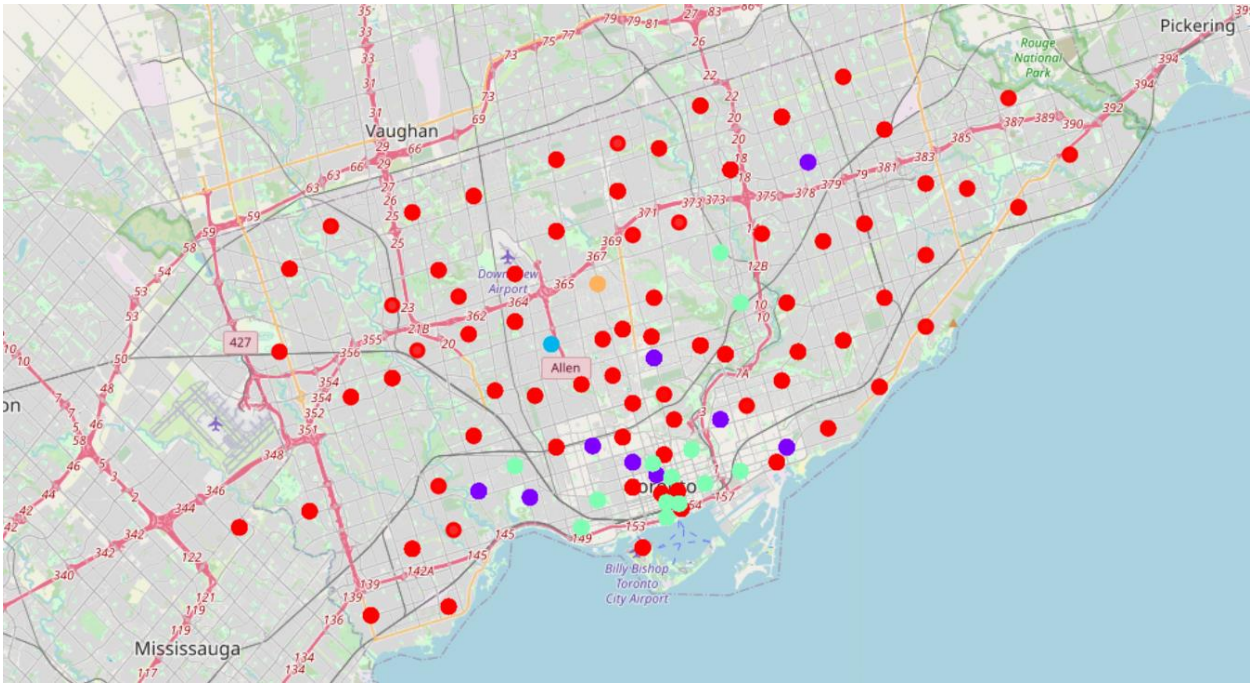
4- **Explore Data:** Now analysis will done on the data that we have prepared and different calculation and decisions will be taken accordingly. We have plot the data in bar chart and bar graph to see the outcome which will make us decide to move further on our analysis for Italian restaurants.

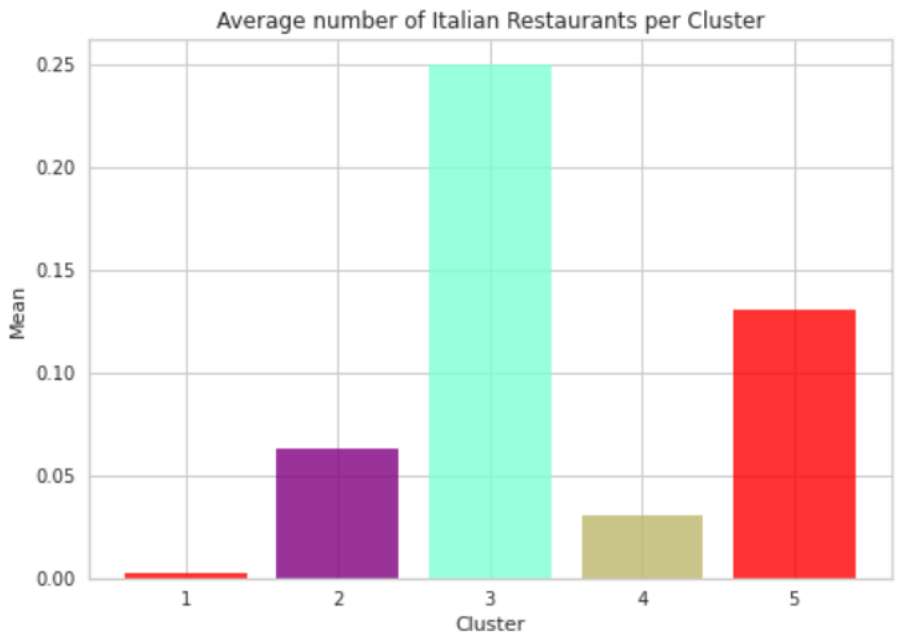


5- **Visualize in map**: Now we will visualize the data in Map using Folium package as below .



6- **Clustering of the neighborhoods**: We will use K-means clustering technique with Elbow visualizer to get the best K.





Results:

After analyzing the data, we observe different results as follows:-

1- **Central Toronto** Borough has 34 Italian restaurants followed by **Downtown Toronto** has 24 and **East Toronto** which has 9 Italian Restaurant whereas **York** Borough has least Italian Restaurant preset its count is only 2 in Toronto City respectively.
2- **Davisville** Neighborhoods has maximum no of Italian Restaurant with a count of 9 followed by **Davisville North** which has 5 Italian Restaurant each in Toronto City respectively and each of them fall under **Central Toronto** Borough
3- Most of Italian cluster under **cluster 3**

Discussion:

It is recommended to start a business of Italian restaurant in **York** Borough which has least no of Italian Restaurant and the demand will be high as expected. The **Davisville** and **Davisville North** Neighborhoods has maximum no of Italian Restaurant which also marked as competition zone.

Conclusion:

Finally, to conclude this capstone project, I have a small glimpse of how real life data-science projects look like. In this project I have imported different types of python libraries and packages such as panda, numpy, matplotlib, sk-learn, seaborn, Folium etc . I have also used BeautifulSoup package to web scrape data. I have also used Foursquare api to get the latitude and longitude data of Toronto City by Geopy Client. I have explored the different Borough, Neighborhood of Toronto city and analyse the data to get different outcome for Italian Restaurants of different parts of the city. I have also used Machine learning technique K-Means clustering to cluster the neighborhoods and predicted a result which may help many business enthusiasts for opening Italian Restaurant in Toronto city where profit will be maximum and the demand is high.

Refrences:

- coursera
- wikipedia.

- https://cocl.us/Geospatial_data

- https://foursquare.com/