# The Battle of the Neighborhoods in Italian dishes

## 1. Introduction

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. According to the United Nations Development Program, Toronto has the second-highest percentage of constant foreign-born population among world cities, after Miami, Florida.

According to Thrillist, Food is a big deal in Toronto. Not only do we need it to live, but they've come into a foodie destination. New restaurants are constantly opening, and chefs continue to push culinary boundaries to come up with new and innovative ideas. It's a bit dizzying just how MUCH good stuff there is to eat here. But don't be daunted.

In this project we will analyze all the Italian Restaurants currently present in Toronto, and check the top Italian Restaurants to give a full analysis.

### 1.1. Background

- **Business Problem**

In this project we will check all the Italian Restaurants that are present in Toronto. We will analyze the data of Toronto's postal codes and see where are the most Italian Restaurants. the major purpose of this project, is to suggest visualize what does the best Italian restaurants have and where are they located for future Italian restaurants.

- **Target Audience**

The people who will benefit for this project are future chefs and business persons who will like to set up a new Italian restaurant in Toronto, knowing where will be the best point of location in Toronto of where is the best place to set up this new restaurant. Also, to know the highest competitors and what does people expect and like from them.

### 1.2. Foursquare API

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about the restaurants that we need.

### 1.3. Workflow

Using Foursquare API credentials, we will mine the different Italian restaurants around Toronto, and we will search the top 5, and see the different attributes of each one of them.

### 1.4. Libraries Used to Developed the Project

- Pandas: For creating and manipulating data frames.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Geocoder: To retrieve Location Data.
- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.

## 2. Data Description

**Toronto Postal Codes data Link:**
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area. This table consist of "Postal Code", "Borough" and "Neighborhood"

**Latitude and Longitude**: https://cocl.us/Geospatial_data

We will wrangle and clean the data, leaving them useful for the project, erasing any "Not Assigned" in the **Borough** column, also if there is any "Not Assigned" value in the **Neighborhood** column we will copy the same of the **Borough** column. Also, if there is a same postal code and borough for different neighborhood, we will group them in to a single row.

**Foursquare API** We will utilize the Foursquare location data, with the Data Frame cleaned. Venues of all Italian Restaurants in Toronto city and their rating, likes and tips information data. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

**Folium Package** Visualize the data in form of Boundary Map. Folium package (Folium is a python library that can create interactive leaflet map using coordinate data.)

## 3. Methodology

The project started by importing and installing all necessary libraries for the project in this there was **Pandas, Numpy, requests, geopy, matplotlib, folium, yellowbrick and sklearn.**

### 3.1. We extract the Postal Code table from Wikipedia as a pandas Data Frame

[3]:

| | Postal Code | District | 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 |

### 3.2. We cleaned the Data frame for a better use and understanding later in the project.

Cleaned the dataframe, removed Boroughs that are 'Not assigned' and if a neighborhood is not assigned, but it has a borough, then the neighborhood will be the same as the borough.

[5]:

| | Postal Code | 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 |

### 3.3. Importing LAT and LONG data with .csv

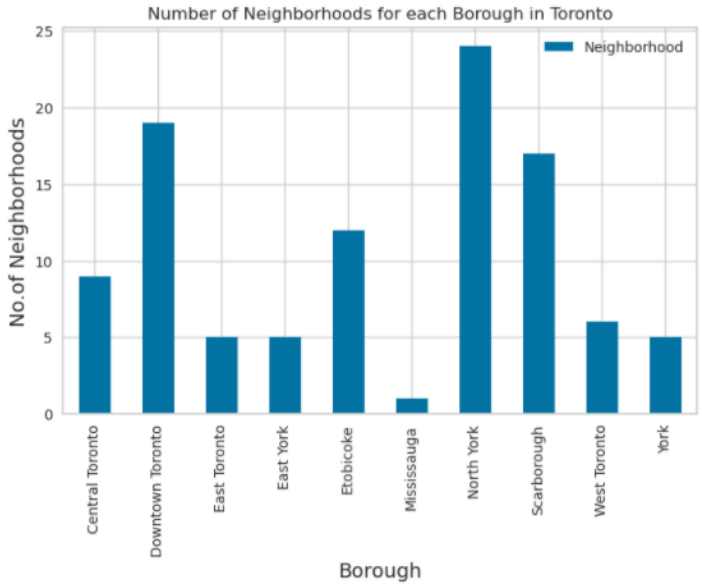
A .csv file was imported with the Latitude and Longitude data for each postal code, later it was merged with the first data frame leaving this data frame.

| | Postal Code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Malvern, Rouge | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Rouge Hill, Port Union, Highland Creek | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood, Morningside, West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |

Finally, we had the principal data frame that was used in the whole project, with 5 columns (Postal Code, Borough, Neighborhood, Latitude and Longitude) with 103 instances in the data frame.

### 3.4. Exploratory Data Analysis

The data analysis began by exploring the number of neighborhoods that each borough had.




**North York** Borough has the highest number of Neighborhoods in Toronto City, mostly 25 number of neighborhoods.

Later, an analysis by a bar plot with the **Foursquare API** was made to discover the number of Italian restaurants in each borough in Toronto City.

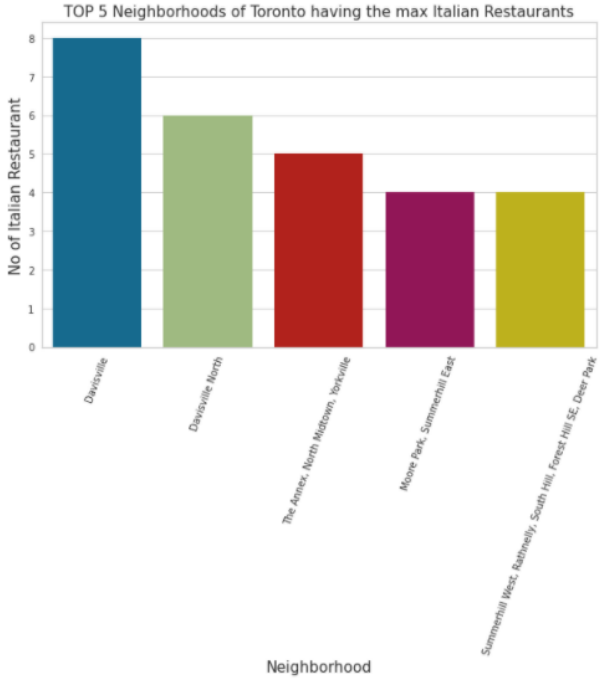
**Central Toronto and Downtown Toronto** have the greatest number of Italian Restaurants in the city according of the bar plot above. So, we will see if there is any difference in the number of Restaurants. For better understanding of the amount of restaurants per Borough a table was created.

```
Borough
Central Toronto      34
Downtown Toronto     34
East Toronto         10
East York             3
Etobicoke             6
North York           10
Scarborough           3
West Toronto         14
York                  2
Name: ID, dtype: int64
```

It can be analyzed that **Central Toronto** Borough has 34 Italian Restaurants as **DownTown Toronto**, followed by **West Toronto** which has 14 Italian Restaurants, as for **North York**, the Borough with the greatest number of neighborhoods, has only 8 Italian Restaurants.

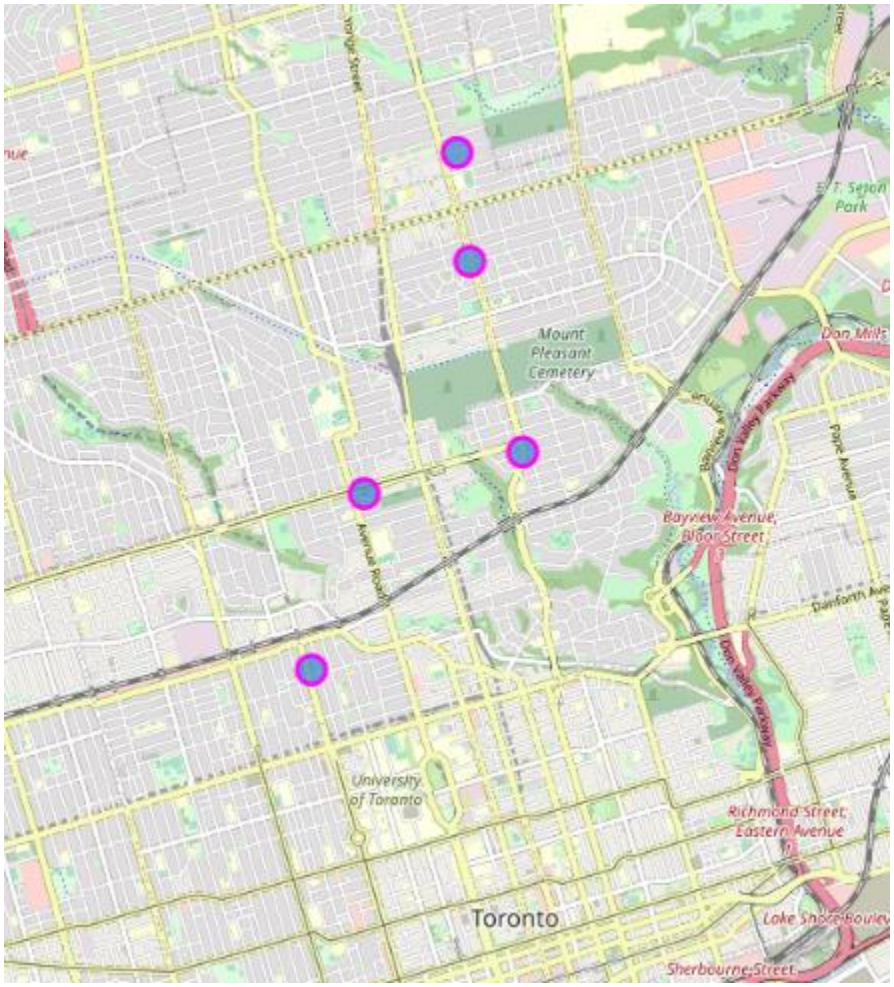
Later the top 5 neighborhoods having the greatest number of Italian restaurants was checked.

| | Neighborhood | Total Italian Restaurants |
|---|---|---|
| 0 | Davisville | 8 |
| 1 | Davisville North | 6 |
| 2 | The Annex, North Midtown, Yorkville | 5 |
| 3 | Moore Park, Summerhill East | 4 |
| 4 | Summerhill West, Rathnelly, South Hill, Forest... | 4 |



Analyzing the chart and the Top 5 table. The **Davisville** neighborhood is the first with 8 restaurants, followed by **Davisville North** with 6 restaurants. This falls in **Central Toronto Borough**, which hast the maximum number of Italian restaurants of the city.

| | Neighborhood | Total Italian Restaurants | Postal Code | Borough | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 0 | Davisville | 8 | M4S | Central Toronto | 43.704324 | -79.388790 |
| 1 | Davisville North | 6 | M4P | Central Toronto | 43.712751 | -79.390197 |
| 2 | The Annex, North Midtown, Yorkville | 5 | M5R | Central Toronto | 43.672710 | -79.405678 |
| 3 | Moore Park, Summerhill East | 4 | M4T | Central Toronto | 43.689574 | -79.383160 |
| 4 | Summerhill West, Rathnelly, South Hill, Forest... | 4 | M4V | Central Toronto | 43.686412 | -79.400049 |




Furthermore, using the **Foursquare API** a 116 instances list was created with each Italian Restaurant with its Bourough, Neighborhood, name, likes, rating and tips.

| | Borough | Neighborhood | ID | Name | Likes | Rating | Tips |
|---|---|---|---|---|---|---|---|
| 0 | Scarborough | Rouge Hill, Port Union, Highland Creek | 4bccf9bf511f9521ce0eb4c7 | Fratelli Village Pizzeria | 18 | 7.6 | 15 |
| 1 | Scarborough | Dorset Park, Wexford Heights, Scarborough Town... | 4bf32a9594af2d7fd4363972 | Nova Ristorante | 17 | 7.1 | 5 |
| 2 | Scarborough | Clarks Corners, Tam O'Shanter, Sullivan | 4b5a2bf2f964a520deb128e3 | Remezzo Italian Bistro | 46 | 7.8 | 19 |
| 3 | North York | Willowdale, Willowdale East | 4b69ad98f964a520f3ac2be3 | Paisano's | 17 | 6.7 | 13 |
| 4 | North York | Willowdale, Willowdale East | 4ad4c05ff964a52031f720e3 | Mezza Notte Trattoria | 12 | 6.1 | 9 |

Later, twith the Fousquare API was checked that **Terroni** got the Max Likes and Tips, and that its rating isn't far from **Noce** rating. Also, it belongs to Downtown Toronto Borough and Garden District, Ryerson Neighborhood. For this we visualized the neighborhoods with the maximum average ratings of Italian Restaurants.

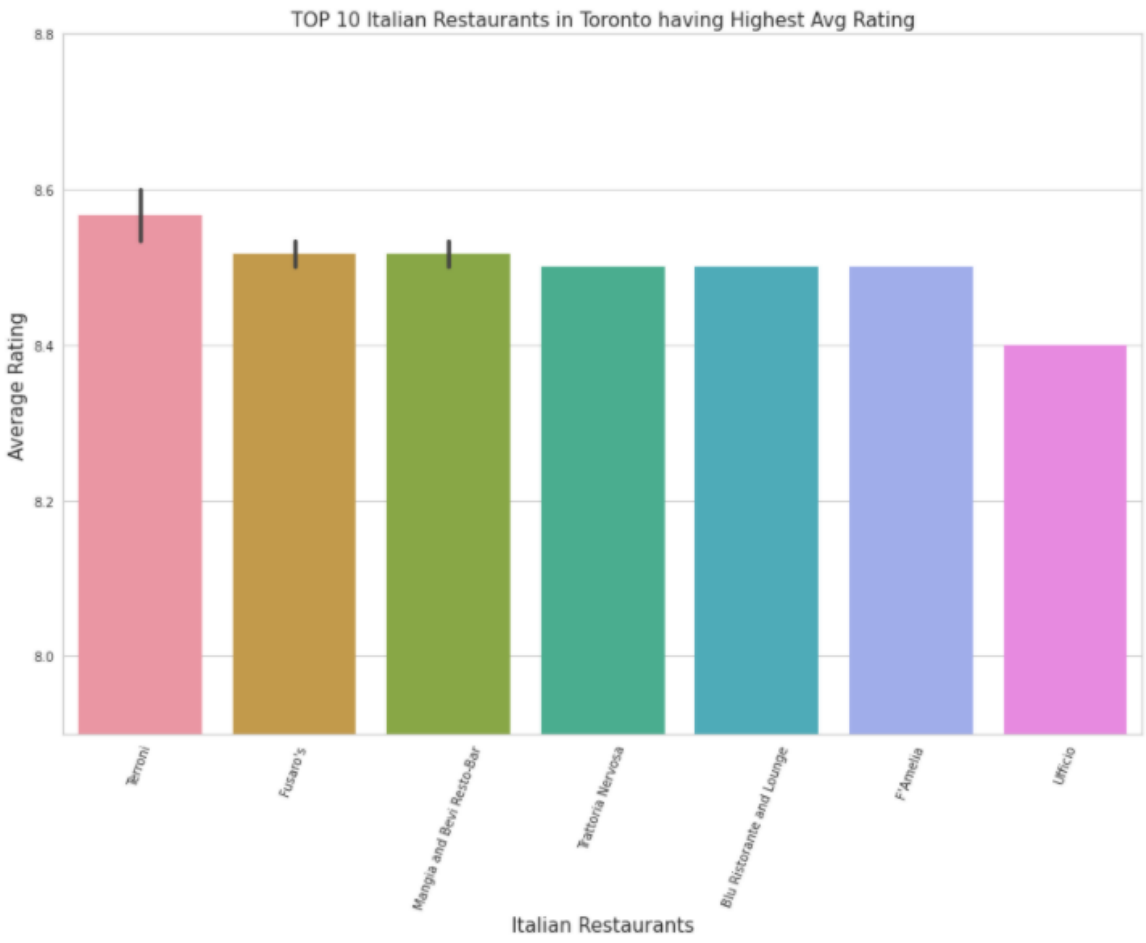
| | Neighborhood | Average Rating |
|---|---|---|
| 8 | Commerce Court, Victoria Hotel | 8.600000 |
| 41 | St. James Town | 8.533333 |
| 42 | St. James Town, Cabbagetown | 8.500000 |
| 36 | Regent Park, Harbourfront | 8.500000 |
| 6 | Church and Wellesley | 8.500000 |

**Commerce Court, Victoria Hotel** Neighborhood has the highest average rating Italian restaurants in Toronto City with 8.6 rating

Top 10 Italian Restaurants in Toronto City with its average rating, name, neighborhood and borough.

| | Name | Borough | Neighborhood | Average Rating |
|---|---|---|---|---|
| 8 | Terroni | Downtown Toronto | Commerce Court, Victoria Hotel | 8.600000 |
| 41 | Terroni | Downtown Toronto | St. James Town | 8.533333 |
| 41 | Fusaro's | Downtown Toronto | St. James Town | 8.533333 |
| 41 | Mangia and Bevi Resto-Bar | Downtown Toronto | St. James Town | 8.533333 |
| 6 | Trattoria Nervosa | Downtown Toronto | Church and Wellesley | 8.500000 |
| 6 | Blu Ristorante and Lounge | Downtown Toronto | Church and Wellesley | 8.500000 |
| 42 | F'Amelia | Downtown Toronto | St. James Town, Cabbagetown | 8.500000 |
| 36 | Mangia and Bevi Resto-Bar | Downtown Toronto | Regent Park, Harbourfront | 8.500000 |
| 36 | Fusaro's | Downtown Toronto | Regent Park, Harbourfront | 8.500000 |
| 28 | Ufficio | West Toronto | Little Portugal, Trinity | 8.400000 |

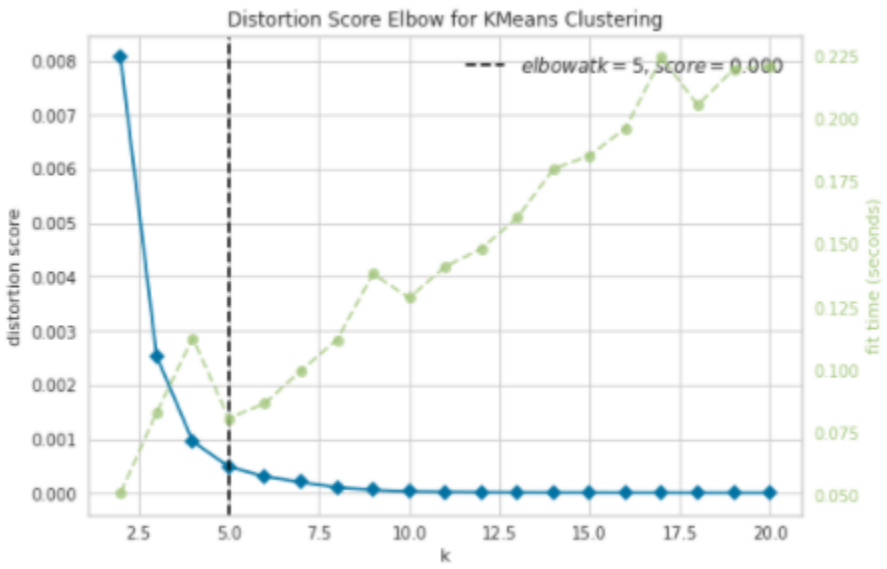
We can analyze that Terroni has 2 restaurants in the same Borough, being those two with the highest Rating in their neighborhoods. Also, Mangia and Bevi Resto-Bar, and Fusaro's hast 2 restaurants both in Downtown Toronto, also in 2 different Neighborhoods with the same rating.




### 3.5. Clustering

In the clustering part of the project, K-means was used as the clusterin method, using the Elbow point to get the best K number of clusters.

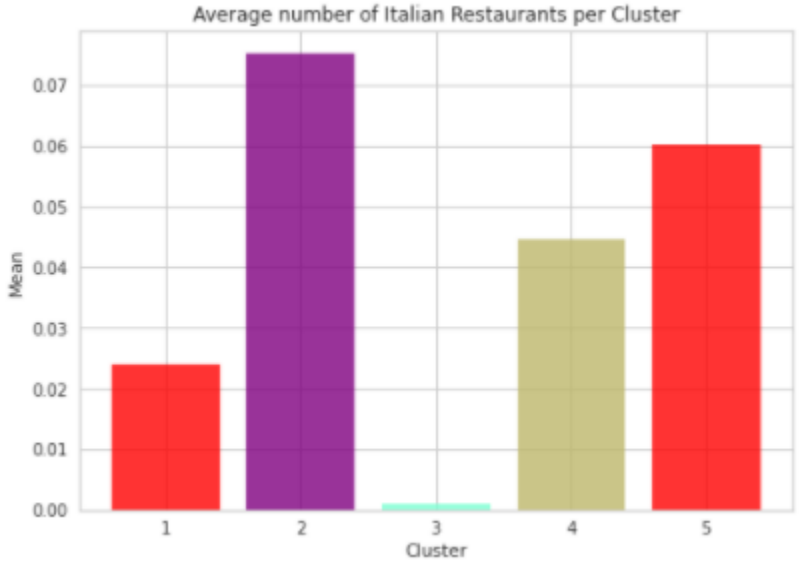
Importing the KElbowVisualizer from the yellowbrick.cluster library we could visualized that the perfect number for K clusters was **5.**




We labeled clusters with neighborhoods according of the amount of the italian restaurants.

| | Neighborhood | Italian Restaurant | Cluster Labels |
|---|---|---|---|
| 0 | Agincourt | 0.00 | 2 |
| 1 | Alderwood, Long Branch | 0.00 | 2 |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | 0.00 | 2 |
| 3 | Bayview Village | 0.00 | 2 |
| 4 | Bedford Park, Lawrence Manor East | 0.08 | 1 |

Making the 5 different clusters of Toronto city, we analyzed that the cluster 2 hast the greatest number of Italian restaurants followed by cluster 5 and 4.



## 4. Results

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

- Central Toronto and Downtown Toronto have the greatest number of Italian Restaurants in the city according of the bar plot above. So, we will see if there is any difference in the number of Restaurants.
- We saw that Central Toronto Borough has 34 Italian Restaurants as Downtown Toronto, Followed by West Toronto which has 14 Italian Restaurants, as for North York, the Borough with the greatest number of neighborhoods, has only 8 Italian Restaurants.
- The Davisville neighborhood has the greatest number of restaurants with 8 restaurants, followed by Davisville North with 6 restaurants
- We saw that Terroni got the Max Likes and Tips, and that its rating isn't far from Noce rating. Also, it belongs to Dowtown Toronto Borough and Garden District, Ryerson Neighborhood.
- Commerce Court, Victoria Hotel Neighborhood has the to highest average rating Italian restaurants in Toronto City with 8.6 rating.
- We analyzed that Terroni has 2 restaurants in the same Borough, being those two with the highest Rating in their respected neighborhoods. Also, Mangia and Bevi Resto-Bar and Fusaro's hast 2 restaurants both in Downtown Toronto, also in 2 different Neighborhoods with the same rating.
- Cluster 2 hast most of the Italian restaurants in the city, followed by Cluster 5 and Cluster 4.

## 5. Discussion

According to the analysis we conclude that to taste the best Italian Food and visit the best Italian Restaurant in the Toronto city we have to visit Terroni as currently it has the highest Average Rating of 8.6 in Commerce Court, Victoria Hotel Neighborhood in Downtown Toronto. Also, for Business purposes it is recommended to start a business of Italian food in York Borough which has the least number of Italian Restaurants and the demand will be high as expected, but if you want to start a Italian Restaurant in a middle demand area it can be at the East Toronto Borough or East York Borough. But, for better engagement with client you could set up a restaurant near Little Italy, that is near Christie Neighborhood in Downtown Toronto.

Now the competition here will be in between Central Toronto and Downtown Toronto Borough as these are the top two areas where Italian restaurants are found.

## 6. Conclusion

Finally, this is a small glimpse of how real-life data-science projects look like. In this project I have imported different types of python libraries such as panda, numpy,

matplotlib I have also used sckitlearn for cluster modeling. 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.

This project gave me the knowledge and strenghts to encourage me with no fear of involving me in future Data Science projects as a junior data scientist. I had the opportunity to engage in data wrangling, data cleansing, data analytics, data graphics and merge this with awesome results and conclusions that we could continue furthermore.