

June 2020

Italian Restaurants Analysis

**Opening italian restaurants in
best possible neighbourhood of
Toronto,CA**

Prepared by: Ridham kacchadiya
Student/Aspiring Data Analyst

Coursera Capstone

Battle of neighborhoods

Introduction

Toronto is one of the most densely populated area in Canada. Being the land of Opportunity, it brings in a variety of people from different ethnic backgrounds to the core city of Canada, Toronto. Being the largest city in Canada with an estimated population of over 6 million, there is no doubt about the diversity of the population. Downtown Toronto being the hub of interactions between ethnicities, brings many opportunities for entrepreneurs to start or grow their business. It is a place where people can try the best of each culture, either while they work or just passing through.

Toronto is well known for its great food. The objective of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the best neighborhoods in Toronto to open a Italian restaurant. Pizza and Pasta are one of the most bought dishes in Toronto originating from Italy. Toronto being the fourth largest home to Italians with a population over 500k, there are huge opportunities and favorable target audience to open a new Italian restaurant. Through this project we will find the most suitable location for an entrepreneur to open a new Italian restaurant in Toronto, Canada. It will also people who want to shift to other neighborhoods and prefer Italian food as their daily prioritized needs

Target Audience:

- Entrepreneurs who want to open an Italian Restaurant in Toronto.
- People who want to shift to other neighborhoods of Toronto, it will help them select a preferable neighborhood whose preferences include enjoying Italian cuisines.

Data Overview

The data required to carry out the analysis will be a combination of CSV files containing data information collected from multiple sources. They are mentioned as follows:-

- The list of postal codes with their respective borough and neighborhoods were collected from Wikipedia
- The corresponding geographical locations were obtained from the geocoder package and the given link (https://cocl.us/Geospatial_data).
- At last, the venues present in each neighborhood which consisted of Italian restaurants and pizza places were fetched from the Foursquare API. Their geographical locations were also included in it.

Methodology

First we will need extract data from all the data sources.

1) Source-1

- Wikipedia
- <https://www.zipcodesonline.com/2020/06/postal-code-of-toronto-in-2020.html?m=1>
- Both were used to collect data of neighborhoods and its postal codes using BeautifulSoup and then converted into a pandas dataframe.
- It was cleaned efficiently by removing rows with 'Not assigned' boroughs and respective manipulations like comma, parenthesis removal were also performed.

2) Source-2

- https://cocl.us/Geospatial_data
- This CSV file was used to map geographical locations and its coordinates with their respective postal codes and neighborhoods
- It was then merged to the above mentioned dataframe(fig1)

3) Source-3

- *Foursquare API.*
- *The retrieval of location, venue name and its category(italian restaurant and pizza places) was done using this API by signing up for its developer account and making requests based on neighborhood's locations.*
- *It was stored in a pandas dataframe.(fig2)*

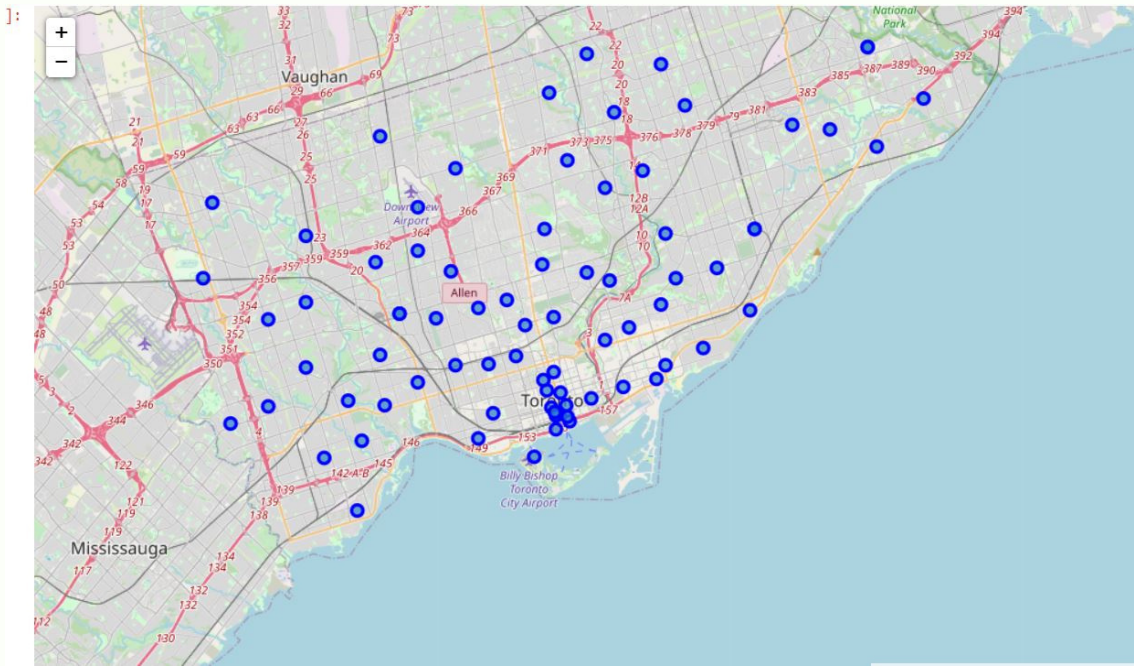
	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M5H	Downtown Toronto	Adelaide,Richmond	43.650571	-79.384568
1	M3H	North York	Bathurst Manor,Downsview North ,Wilson Heights	43.754328	-79.442259
2	M5V	Downtown Toronto	Bathurst Quay ,CN Tower ,Harbourfront West ,Is...	43.628947	-79.394420
3	M2K	North York	Bayview Village	43.786947	-79.385975
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306

Fig.1

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M5H	Downtown Toronto	Adelaide,Richmond	43.650571	-79.384568
1	M3H	North York	Bathurst Manor,Downsview North ,Wilson Heights	43.754328	-79.442259
2	M5V	Downtown Toronto	Bathurst Quay ,CN Tower ,Harbourfront West ,Is...	43.628947	-79.394420
3	M2K	North York	Bayview Village	43.786947	-79.385975
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306

Fig.2

Map containing all neighborhoods of toronto

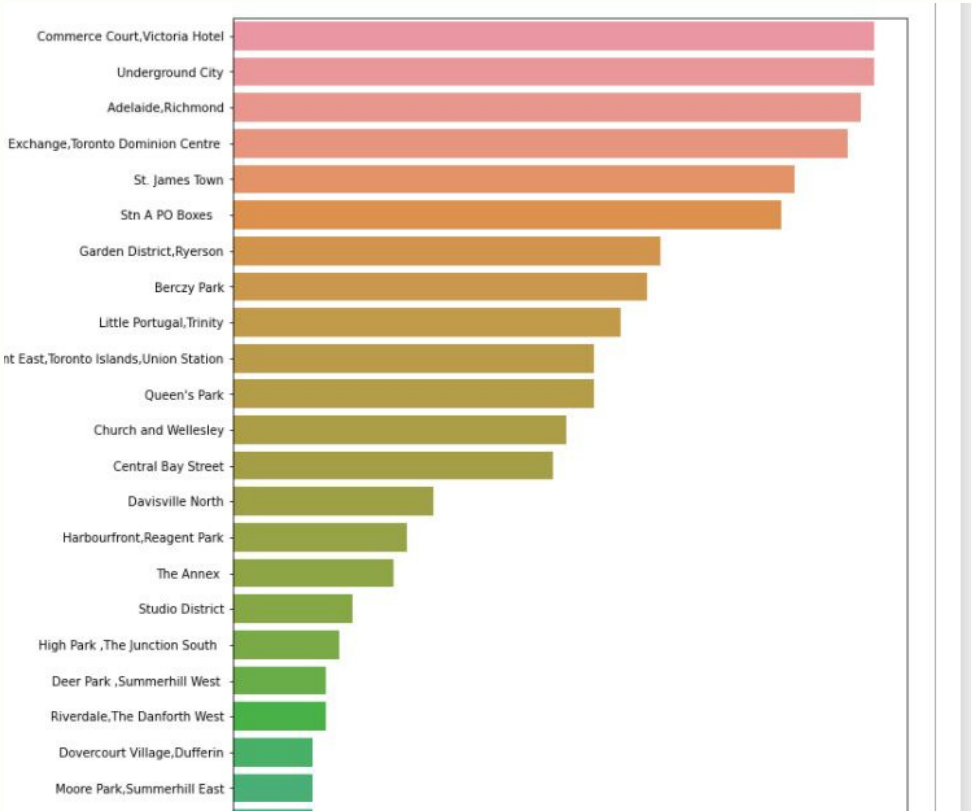


Analysis

- After all the relevant data was collected ,Grouping of data was done on venue dataframe to know the number of italian places present in each neighborhood and stored in a dataframe.(fig3)

	Neighborhood	Number of italian restaurants
0	Commerce Court,Victoria Hotel	48
1	Underground City	48
2	Adelaide,Richmond	47
3	Design Exchange,Toronto Dominion Centre	46
4	St. James Town	42

Fig.3

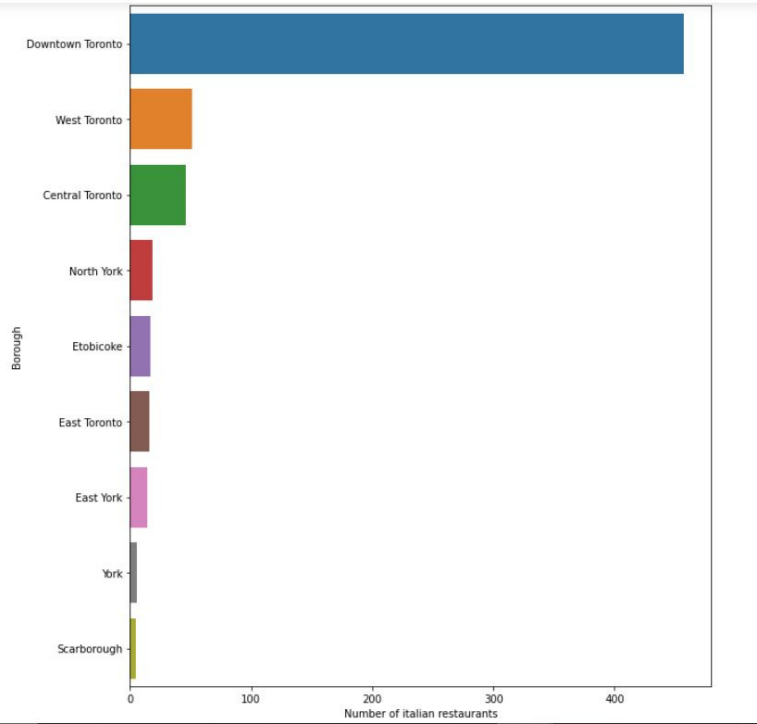


Bar plot :Total number of italian food places in each neighborhood.

- The other analysis performed was grouping venues based on Boroughs in which they were present.(fig4).

	Borough	Number of italian restaurants
0	Downtown Toronto	457
1	West Toronto	51
2	Central Toronto	46
3	North York	19
4	Etobicoke	17

Fig.4



Bar plot :Total number of italian food places in each borough.

- Pie chart representing proportion of italian restaurants in percentage for all boroughs in toronto city.(fig5).

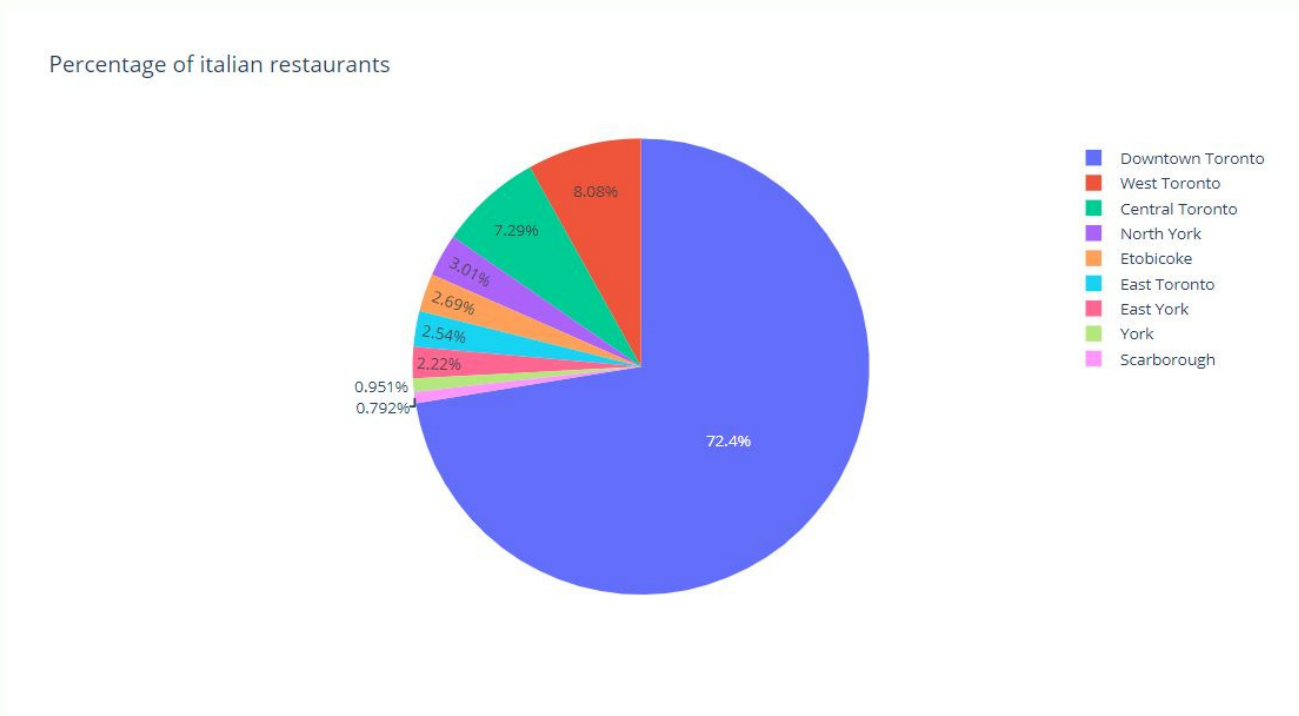


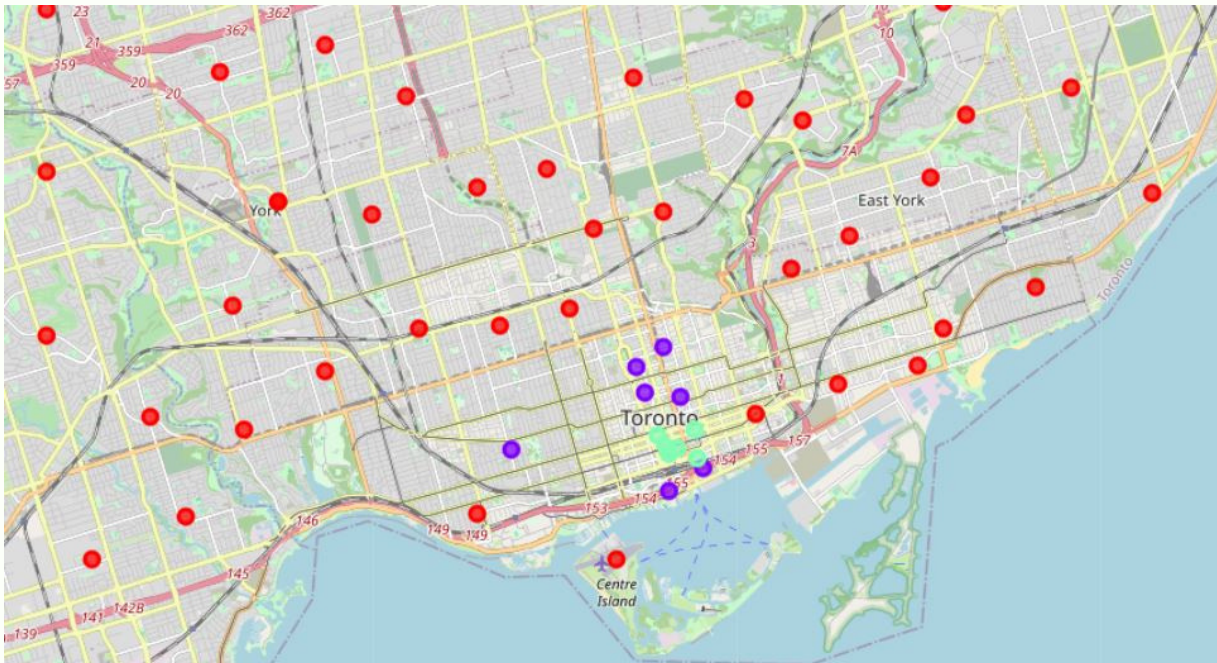
Fig.5

- Clustering of neighborhoods was performed after analysis by grouping neighborhoods in 3 defined clusters based on their similarity.(fig6)

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Number of italian restaurants	clusters
0	M5H	Downtown Toronto	Adelaide,Richmond	43.650571	-79.384568	47.0	2.0
1	M3H	North York	Bathurst Manor,Downsview North ,Wilson Heights	43.754328	-79.442259	0.0	0.0
2	M5V	Downtown Toronto	Bathurst Quay ,CN Tower ,Harbourfront West ,Is...	43.628947	-79.394420	0.0	0.0
3	M2K	North York	Bayview Village	43.786947	-79.385975	0.0	0.0
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	31.0	1.0
5	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848	1.0	0.0
6	M9C	Etobicoke	Bloordale Gardens,Eringate,Markland Wood,Old B...	43.643515	-79.677201	0.0	0.0
7	M4J	East York	Broadview North,East Toronto	43.685347	-79.338106	3.0	0.0

Fig.6

Cluster Map consisting each neighborhood in its respective cluster



Cluster-1 (Red-0)

Cluster 1 consists of neighborhoods with 0-15 italian restaurants in it. It is considered as relatively low frequency of restaurants compared to others and most of the neighborhood lie in this cluster

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Number of italian restaurants	clusters
1	M3H	North York	Bathurst Manor,Downsview North ,Wilson Heights	43.754328	-79.442259	0.0	0.0
2	M5V	Downtown Toronto	Bathurst Quay ,CN Tower ,Harbourfront West ,Is...	43.628947	-79.394420	0.0	0.0
3	M2K	North York	Bayview Village	43.786947	-79.385975	0.0	0.0
5	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848	1.0	0.0
6	M9C	Etobicoke	Bloordale Gardens,Eringate,Markland Wood,Old B...	43.643515	-79.577201	0.0	0.0
7	M4J	East York	Broadview North,East Toronto	43.685347	-79.338106	3.0	0.0

Cluster-2(purple-1)

Cluster 2 consists of neighborhoods with 20-35 italian restaurants in it. It is considered as a moderate frequency of restaurants and few neighborhoods lie in this cluster.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Number of italian restaurants	clusters
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	31.0	1.0
11	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	24.0	1.0
13	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	25.0	1.0
28	M5B	Downtown Toronto	Garden District,Ryerson	43.657162	-79.378937	32.0	1.0
31	M5J	Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752	27.0	1.0
45	M6J	West Toronto	Little Portugal,Trinity	43.647927	-79.419750	29.0	1.0
55	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494	27.0	1.0

Cluster-3(green-2)

Cluster 3 consists of neighborhoods with 40-50 italian restaurants in it. It is considered as a highfrequency of restaurants and few neighborhoods lie in this cluster.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Number of italian restaurants	clusters
0	M5H	Downtown Toronto	Adelaide,Richmond	43.650571	-79.384568	47.0	2.0
16	M5L	Downtown Toronto	Commerce Court,Victoria Hotel	43.648198	-79.379817	48.0	2.0
19	M5K	Downtown Toronto	Design Exchange,Toronto Dominion Centre	43.647177	-79.381576	46.0	2.0
59	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	42.0	2.0
62	M5W	Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846	41.0	2.0
72	M5X	Downtown Toronto	Underground City	43.648429	-79.382280	48.0	2.0

Discussion

Most of italian restaurants are present in cluster 3(green) as those neighborhoods have a high number of variety for italian cuisines and an average of 47 different places to go per neighborhood. So we can suggest neighborhoods like Adelaide, Richmond, underground city etc present in Downtown to people who prefer italian food on a regular basis and are fond of pizzas.

As per business perspective, business officials and entrepreneurs can select a wide variety of neighborhoods from north york, Scarborough, Etobichoke belonging to cluster 1(red) to open an italian restaurant as these have very few number of places. So there is a high chance for restaurant to be a success and satisfying people needs in that area as they have very few options to enjoy a italian delicacy.

Cluster 2(purple) contains a moderate amount of italian places at an average of 27 per neighborhood. It is a moderately competitive region and one can think of starting a new italian restaurant if they are confident enough about their dishes and menu.

Conclusion

In conclusion, to end off this project, we had an opportunity on a business problem, and it was tackled in way that it was similar to how a genuine data scientist would do. We utilized numerous Python libraries to fetch the information , to control the content and to break down and visualize those datasets. We have utilized Foursquare API to investigate the settings in neighborhoods of Toronto, get great measure of data from Wikipedia which we scraped with the BeautifulSoup Web scraping Library. We also visualized utilizing different plots present in seaborn ,matplotlib libraries and created maps tagged with neighborhoods of city Toronto.

It also contains a reasonable amount of drawbacks and can be improved efficiently with increase in accessibility to different types of data. Averaging method can also be used to tackle the problem by considering venues with different categories and then comparing each of them. But ideally, the task performed in this capstone project acts as an initial direction to tackle more complex real-life problems using data-science.