Capstone Project - The Battle of Neighborhoods

Introduction

Problem Background:

The Capital City of the province of Ontario, Toronto is the largest and most populated city in Canada. Toronto is one of the most multicultural cities and serves as a metropolis having significant contribution to the country's economic, financial, and historical development. Diverse and influential the city attracts people from all over that want to experience the city for its skyscrapers, restaurants, college education, and many other reasons.

Its large population and popularity provide opportunity, but this comes at the expense of a saturated market filled with competition. Decisions must me made cautiously, they need to be informed with the current and past information that is available. It needs to answer what, when, and where. This strategic planning helps provide insight, reduce risk, and lets you select a potentially more profitable, prosperous location regardless of the business venture you are pursuing.

Problem Description:

Specifically, I will be investigating the restaurants in the city of Toronto and selecting the best neighborhoods to open a restaurant. Toronto is known for its ethnic and cultural diversity; it's diversity and high population has allowed different food cultures and practices to intertwine and be brought together in the city. Demographically, almost or slightly over half of the population are visible minorities and they are represented in ethnic neighborhoods like Chinatown, Greektown, Koreatown, Little Italy, Little India, Little Portugal, Little Jamaica etc.

Therefore, neighborhood locations are very important in determining where to open a restaurant not just the amount located there and the type. All restaurant types will be included and the number of restaurants in each neighborhood will be investigated.

Data

The dataset I will be using is the Toronto Neighborhood data that we used previously for segmenting and clustering neighborhoods in Toronto in Week 3. The dataset is located on Wikipedia and contains 103 observations and includes 3 variables: Postal Code, Borough, and Neighborhood. It features 103 Neighborhoods that are located across 15 Boroughs.

Additionally, the latitude and longitude coordinates of the neighborhoods are also included using the geospatial data. This dataset will be utilized with the Foursquare API so that venue information can be used and explored within each neighborhood in Toronto. Toronto's diverse culture is evident in its variety of restaurant types like Belgian, Brazilian, Caribbean, Ethiopian, German, Greek, Moroccan, Portuguese, and many more.

Datasets that will be used:

- Toronto Neighborhood Data
  - Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - Description: The dataset contains the borough, neighborhood, and postal code information of the city of Toronto
- Geospatial Data
  - Source: https://cocl.us/Geospatial_data
  - Description: By using this we can get the geospatial data of Toronto such as the neighborhood and venue latitude and longitude coordinates to provide further information on the location
- Restaurant Venues in Toronto
  - Source: Foursquare API
  - Description: Allows us to retrieve the venue types and categories in each neighborhood which we can further filter to restaurant venues

Methodology

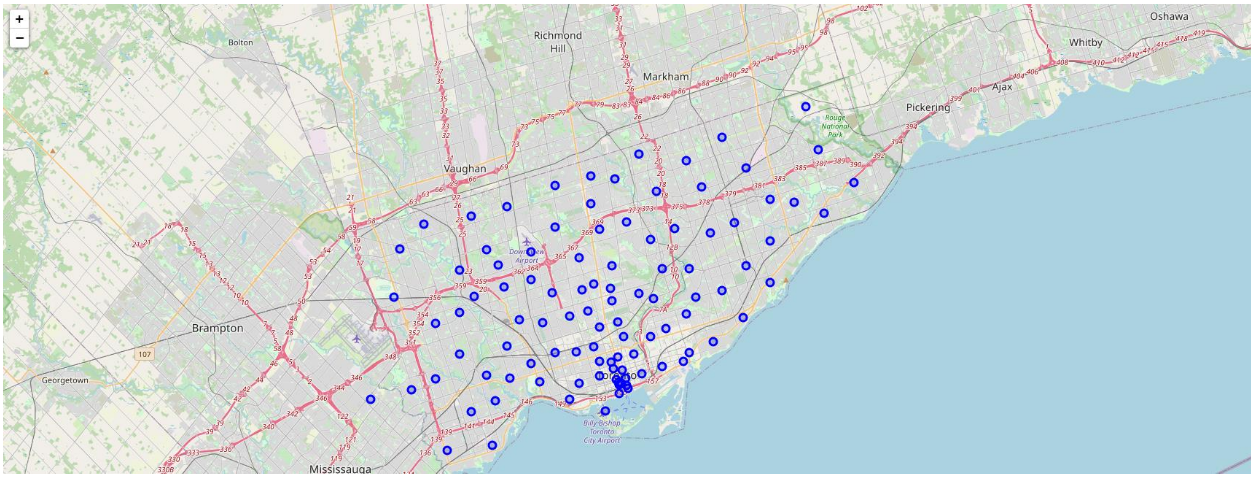
Exploratory Data Analysis

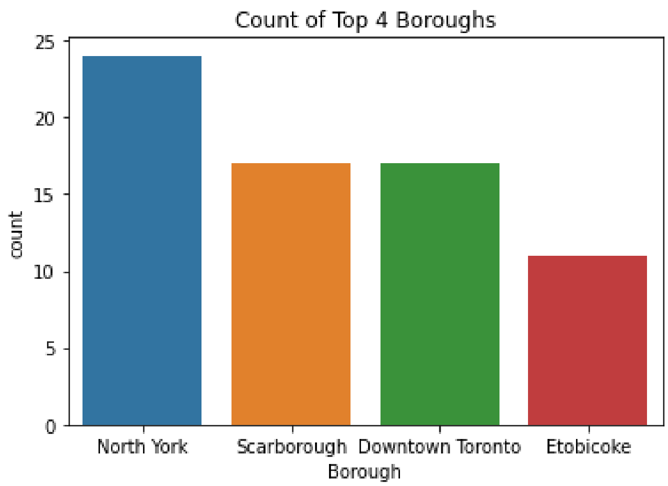
Reading in the dataset BeautifulSoup was used to web scrape and transform it into a pandas data frame. The Toronto Neighborhood dataset includes 103 observations and 3 variables: Postal Code, Borough, and Neighborhood which are categorical and describe the geographical location of the city of Toronto. The objective of this project is to use an algorithm to find the best location, the best neighborhood and borough for restaurants to open and expand to in Toronto. As all the variables are categorical and there is not a model being created there was no need to look at trends or check for assumptions like normality and multicollinearity.

	Postal Code	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
...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto Business	Enclave of M4L
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

103 rows × 3 columns

All the 103 observations record a different neighborhood in Toronto and a map of these Toronto neighborhoods are shown below, many of them are scattered, but there is a small cluster in south Toronto towards the University of Toronto. These 103 neighborhoods encompass 15 different boroughs. A count plot of the top 4 boroughs are shown below indicating that these are North York at 24, Scarborough and Downtown Toronto around 16, and Etobicoke around 11 observations.


Oshawa
Whitby
Ajax
Pickering
Richmond Hill
Markham
Vaughan
Bolton
Brampton
Georgetown
Mississauga


Count of Top 4 Boroughs
count
0
5
10
15
20
25
North York
Scarborough
Downtown Toronto
Etobicoke
Borough

Results

Merging and Hot Encoding

To include the geographical coordinates the Toronto Neighborhood and Geospatial datasets were merged to provide more specific information that enables the use of the Foursquare API. A function and a Foursquare API reference were then incorporated to retrieve the venue information of each neighborhood and use it within the data frame. The image below shows the dataset grouped by neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
...	...	...	...	...	...	...	...
2096	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	South St. Burger	43.631314	-79.518408	Burger Joint
2097	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Wingporium	43.630275	-79.518169	Wings Joint
2098	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Dollarama	43.629883	-79.518627	Discount Store
2099	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Healthy Planet	43.630214	-79.518495	Supplement Shop
2100	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Artisano Bakery Café	43.631006	-79.518172	Bakery

424 rows × 7 columns

The Venue Categories were then hot encoded to illustrate which venues were present in a neighborhood with 0 showing its absence and 1 showing its presence in a neighborhood which we could later use for a machine learning algorithm. Here is a small snippet of the most common venues in each neighborhood ranked by their frequencies.

	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	...	Train Station	Truck Stop	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant	Warehouse Store	Wine Bar	Wings Joint	Women's Store
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
93	Willowdale South	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.030303	0.0	0.0	0.0	0.0
94	Willowdale West	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
95	Woburn	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
96	Woodbine Heights	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
97	York Mills West	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0

98 rows × 269 columns

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Latin American Restaurant	Skating Rink	Clothing Store	Breakfast Spot	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run
1	Alderwood, Long Branch	Pizza Place	Coffee Shop	Skating Rink	Pub	Sandwich Place	Gym	Airport Gate	Event Space	Escape Room	Electronics Store
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Pizza Place	Supermarket	Deli / Bodega	Sushi Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Shopping Mall	Mobile Phone Shop
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Women's Store	Dim Sum Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant
4	Bedford Park, Lawrence Manor East	Sandwich Place	Coffee Shop	Italian Restaurant	Restaurant	Liquor Store	Pharmacy	Pub	Butcher	Café	Sushi Restaurant

Filtering Restaurant Venues

As the objective is to find the best neighborhood and borough for restaurants to locate to, the dataset was then filtered with a list being created for all the restaurant venues which were then counted within each neighborhood.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	5	5	5	5	5	5
Alderwood, Long Branch	7	7	7	7	7	7
Bathurst Manor, Wilson Heights, Downsview North	21	21	21	21	21	21
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	25	25	25	25	25	25
...	...	...	...	...	...	...
Willowdale South	33	33	33	33	33	33
Willowdale West	5	5	5	5	5	5
Woburn	5	5	5	5	5	5
Woodbine Heights	7	7	7	7	7	7
York Mills West	2	2	2	2	2	2

98 rows × 6 columns

Being hot encoded a machine learning algorithm, K-Means Clustering was implemented to find the optimal number of k's to split the dataset into using silhouette coefficients as our scoring method. N=2 had the highest silhouette score at 0.724 meaning 2 clusters did the best at quantifying the quality of clusters and separating them into neighborhoods with and with no or little restaurants. K-Means Clustering with 2 clusters was then ran on the data.

```
For n_clusters=2, The Silhouette Coefficient is 0.7244763858438001
For n_clusters=3, The Silhouette Coefficient is 0.6099089526000726
For n_clusters=4, The Silhouette Coefficient is 0.5247198774942157
For n_clusters=5, The Silhouette Coefficient is 0.5269345881108605
For n_clusters=6, The Silhouette Coefficient is 0.5167511100569303
For n_clusters=7, The Silhouette Coefficient is 0.4491962015721038
For n_clusters=8, The Silhouette Coefficient is 0.4435289262238617
For n_clusters=9, The Silhouette Coefficient is 0.4366609466012446
```

N=2 has the highest silhoutte coefficient at 0.7244763858438001 so we will use 2 clusters

After remerging to include the postal code and borough information, a new category of total and total sum was created to indicate the number of restaurants in that area. Cluster 1 had the higher total and total sum at 27.4 and 54.8 indicating that it is more saturated. This information is useful as we can now explicitly state that cluster0 has less saturation and more market potential for restaurants.

	Fast Food Restaurant	Food & Drink Shop	Portuguese Restaurant	Pizza Place	Restaurant	Breakfast Spot	French Restaurant	Vietnamese Restaurant	Italian Restaurant	Sushi Restaurant	...	Cajun / Creole Restaurant	Dumpling Restaurant	Doner Restaurant	Filipino Restaurant	Bed & Breakfast	Taiwanese Restaurant	Theme Restaurant	Wings Joint	Total	Total Sum
cluster0	0.253012	0.024096	0.024096	0.397590	0.325301	0.132530	0.036145	0.048193	0.192771	0.144578	...	1.204819e-02	-1.040834e-17	-1.040834e-17	-1.040834e-17	-1.040834e-17	1.204819e-02	-1.040834e-17	1.204819e-02	3.349398	6.698795
cluster1	0.600000	0.133333	0.133333	1.266667	2.400000	0.666667	0.466667	0.400000	1.933333	1.000000	...	-1.734723e-18	6.666667e-02	6.666667e-02	6.666667e-02	6.666667e-02	-1.734723e-18	6.666667e-02	-1.734723e-18	27.400000	54.800000

Cluster0: Unsaturated Market

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724	0.0	0.0
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0.0	0.0
2	M4C	East York	Woodbine Heights	43.695344	-79.318389	0.0	0.0
3	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	0.0	0.0
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	0.0	0.0
5	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512	0.0	0.0
6	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0.0	0.0
7	M4J	East York/East Toronto	The Danforth East	43.685347	-79.338106	0.0	0.0
8	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	0.0	0.0
9	M3K	North York	Downsview East	43.737473	-79.464763	0.0	0.0
10	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	0.0	0.0
11	M3L	North York	Downsview West	43.739015	-79.506944	0.0	0.0
12	M6L	North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-79.490074	0.0	0.0
13	M3M	North York	Downsview Central	43.728496	-79.495697	0.0	0.0
14	M9M	North York	Humberlea, Emery	43.724766	-79.532242	0.0	0.0
15	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848	0.0	0.0
16	M3N	North York	Downsview Northwest	43.761631	-79.520999	0.0	0.0
17	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0.0	0.0
18	M5N	Central Toronto	Roselawn	43.711695	-79.416936	0.0	0.0
19	M2P	North York	York Mills West	43.752758	-79.400049	0.0	0.0
20	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577	0.0	0.0
21	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420	0.0	0.0
22	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	0.0	0.0
23	M9W	Etobicoke Northwest	Clairville, Humberwood, Woodbine Downs, West H...	43.706748	-79.594054	0.0	0.0
24	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	0.0	0.0
25	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	0.0	0.0

Cluster1: Saturated Market

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	30.0	1.0
1	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	27.0	1.0
2	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	16.0	1.0
3	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	26.0	1.0
4	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	38.0	1.0
5	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752	24.0	1.0
6	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750	18.0	1.0
7	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	17.0	1.0
8	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576	38.0	1.0
9	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	37.0	1.0
10	M4S	Central Toronto	Davisville	43.704324	-79.388790	17.0	1.0
11	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	21.0	1.0
12	M5W	Downtown Toronto Stn A	Enclave of M5E	43.646435	-79.374846	27.0	1.0
13	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	42.0	1.0
14	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	33.0	1.0

Discussion

As Cluster0 has the lower total and total sum at 3.35 and 6.7 the neighborhood and boroughs within that cluster are more suited and potentially profitable for restaurants to expand to. To name a few this includes Etobicoke, York, East York, Scarborough, and North York within their respective neighborhoods listed. In Cluster1 many of the Neighborhoods listed as being more saturated with restaurants are all within Toronto and more specifically Downtown Toronto. It is also worth noting that 6 out of 25 of the observations in Cluster0 also includes boroughs with Toronto in their name showing that there are places restaurants can move to within it, the neighborhood just also needs to be accounted for. This information should be kept in mind when opening and expanding restaurants within the city of Toronto.

Conclusion

The data included in this project is limited both in size and time. The number of people in each neighborhood as well as the type and number of restaurants are always changing. To provide better results more data could be collected and reference across a longer time span and can also include more information like population or restaurant income. With the data given and analysis conducted I would recommend that restaurants expand to neighborhoods within the first cluster, cluster0 which lists areas like Etobicoke, York, East York, Scarborough, and North York and their respective neighborhoods.