The Battle of Neighbourhoods

– **Exploring Toronto neighbourhoods for starting a new restaurant business**

Introduction

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571, it is the most populous city in Canada and the fourth most populous city in North America. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. It is a tourism hub and has a booming restaurant sector too. The food sector is one of Toronto's most important sectors with annual sales of $17.75 billion.

Business Problem

In this project we will be analysing the neighbourhoods of Toronto to see which ones would be the most ideal for opening a new restaurant.

Additionally, we would also study which type of restaurant i.e. restaurants of which cuisine would be more suitable for which neighbourhood as people in different neighbourhoods maybe having different preferences.

Target Audience

The target audience for this project would be investors or entrepreneurs looking to start a new restaurant in Toronto. This study can also be of use to franchise owners of restaurant chains to understand which neighbourhood would be suitable to start a new outlet.

Data

To explore the potential solutions to the above business problem the following data is required:

1. The Postal code, Borough and names of all the Neighbourhoods present in Toronto. This information was scraped from the website :
   https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M' . The postal codes in this data would be useful in merging with the dataset containing geographical coordinates.
2. Geographical coordinates of Toronto neighbourhoods were obtained using the "https://cocl.us/Geospatial_data" csv file. These coordinates will be helpful in getting location information using the Four Square API.

3. Information about venues present in each neighbourhood of Toronto were obtained using Four Square API. This included the names of the Venue, Venue Category, Latitude and Longitude of the venues in each neighbourhood. The required information about restaurants and their types that is needed for this project can be further extracted from this data.

Methodology

For this project, we first need to acquire the necessary data, clean and prepare the data for analysis, perform exploratory analysis and run Machine learning algorithms to provide a solution that can be leveraged by the business stakeholders to make their decision.

First, we need data about Toronto and its suburbs which includes the names of the borough, neighbourhoods along with postal codes of each neighbourhood. We will get this data by webscraping from Wikipedia using BeautifulSoup. We convert this into a pandas data frame and check for any values which are null or not assigned. The first few rows of the resultant dataset looks as fig.1

	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

This dataset consists of 103 rows which means it has details of 103 neighbourhoods of Toronto and its suburbs.

The latitudes and longitudes of these neighbourhoods have been obtained using the "https://cocl.us/Geospatial_data" csv file. These location coordinates were merged with the above dataset to get a new dataset which looks as follows:

	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

This dataset also contains 103 rows.

Next, we need information about restaurants present in each neighbourhood of Toronto. This was obtained using Four Square API. The data obtained includes details of all types of venues in the neighbourhoods. It has the names, latitudes, longitudes, and venue categories of venues present in each neighbourhood of Toronto as we can see in the figure below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant

This dataset has many venue categories present such as farms, metro, banks, parks, gas stations etc. but we only need data regarding restaurants so we extract the required information and create a dataset with only those rows which have restaurants as their venue category. It looks as follows:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Guildwood, Morningside, West Hill	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
3	Woburn	43.770992	-79.216917	Korean Grill House	43.770812	-79.214502	Korean BBQ Restaurant
4	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka Restaurant

This dataset has 295 rows which implies that there are 295 restaurants in Toronto.

Using the Folium package we can visualise these restaurants on the map of Toronto and observe the location of restaurants and how they are scattered across Toronto.
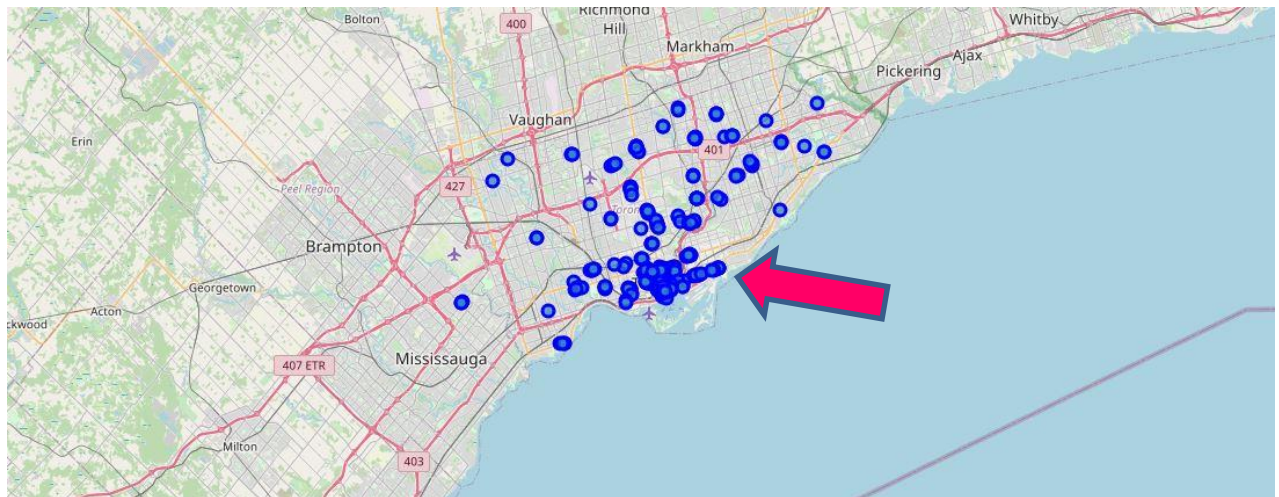
As expected, in the above map we can see that the majority of restaurants are located in the neighbourhoods located near the downtown area of Toronto.

Exploratory Analysis

Exploring the data further, we group the data by neighbourhoods to see how many restaurants are located in each neighbourhood. We can view the results in a bar chart below:

Number of restaurants in each neighbourhood of Toronto
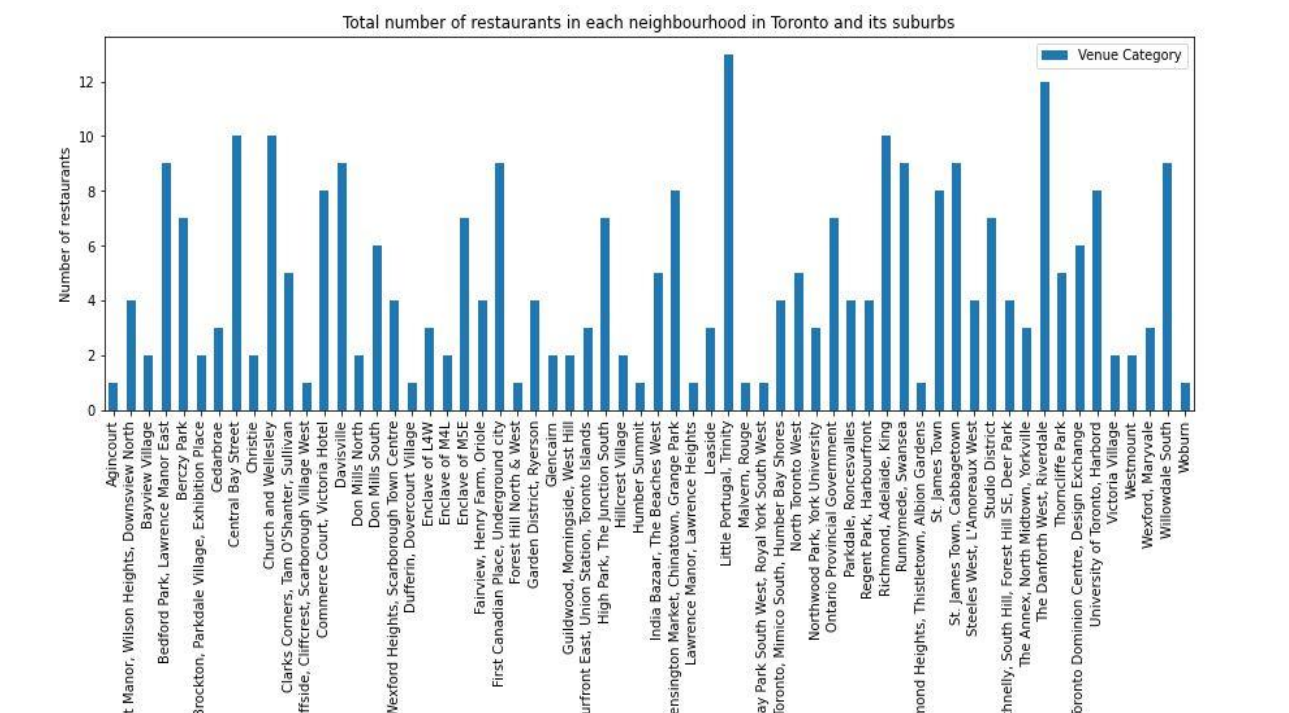
We can clearly see that some neighbourhoods like Little Portugal,Trinity and The Danforth West, Riverdale have a large number of restaurants whereas others like Agincourt, Humber

Summit and Woburn have only one restaurant. Further sorting this data and plotting it gives a clearer picture of neighbourhoods and their restaurants as seen in the figure below:

Total number of restaurants in each neighbourhood in Toronto and its suburbs
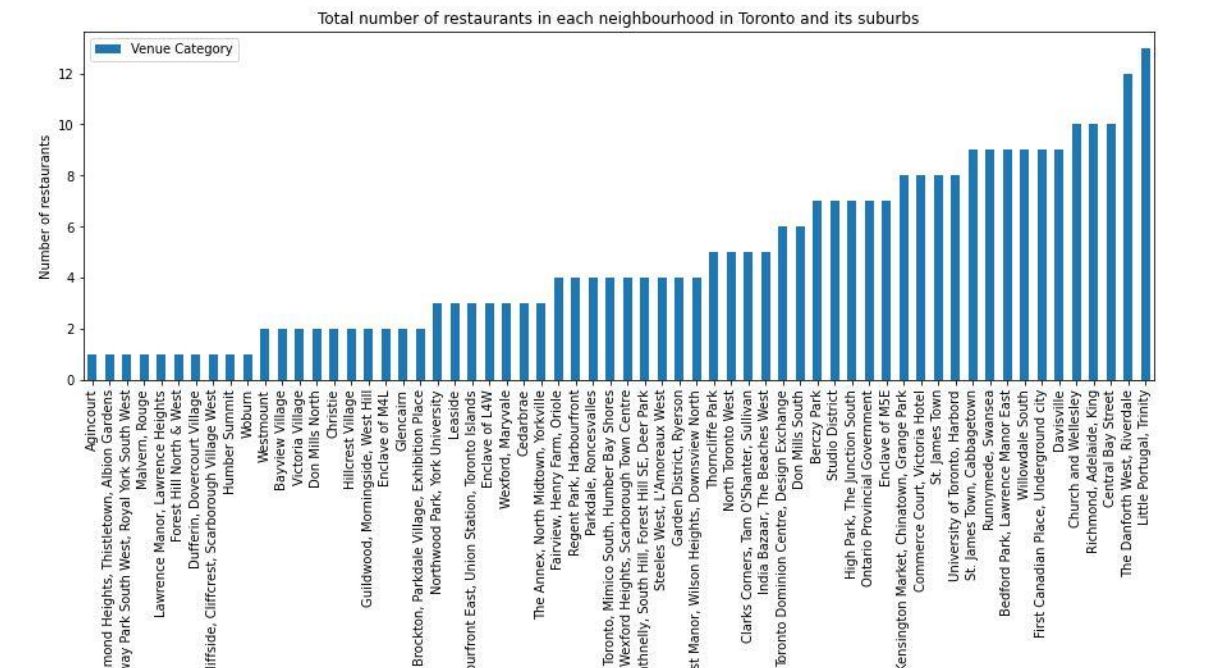
We can also explore the different types of restaurants present in Toronto. There are 40 unique categories of restaurants present in Toronto and a frequency table of their count is given below:

Restaurant	40		New American Restaurant	4
Italian Restaurant	27		Latin American Restaurant	3
Fast Food Restaurant	19		Mediterranean Restaurant	2
Sushi Restaurant	18		Cuban Restaurant	2
Japanese Restaurant	18		Gluten-free Restaurant	2
Thai Restaurant	15		Falafel Restaurant	2
American Restaurant	14		Portuguese Restaurant	2
Seafood Restaurant	13		Dim Sum Restaurant	1
Greek Restaurant	12		Hakka Restaurant	1
Middle Eastern Restaurant	12		Belgian Restaurant	1
Mexican Restaurant	11		Korean BBQ Restaurant	1
Vietnamese Restaurant	10		Theme Restaurant	1
Indian Restaurant	9		Cajun / Creole Restaurant	1
Vegetarian / Vegan Restaurant	9		Modern European Restaurant	1
Chinese Restaurant	8		Tibetan Restaurant	1
French Restaurant	7		Eastern European Restaurant	1
Comfort Food Restaurant	6		Colombian Restaurant	1
Ramen Restaurant	6		Korean Restaurant	1
Caribbean Restaurant	6		Ethiopian Restaurant	1
Asian Restaurant	5		Taiwanese Restaurant	1

This is data can be plotted in the graph below:

We observe that Toronto has a large number of Italian and Fast Food restaurants. Asian cuisines such as Sushi, Japanese and Thai restaurants follow close behind. The maximum observations go to the first category Restaurants. These maybe assumed to be multi cuisine restaurants or data about their type is not available.

One Hot Encoding

One hot encoding is used to create dummy variables for converting categorical variables to numeric variables required for analysis. In our project we create dummy variables for the different restaurant types and calculate the mean of restaurant types grouped by neighbourhood. The resultant dataset looks as follows:

	Neighbourhood	American Restaurant	Asian Restaurant	Belgian Restaurant	Cajun / Creole Restaurant	Caribbean Restaurant	Chinese Restaurant	Colombian Restaurant	Comfort Food Restaurant	Cuban Restaurant	...	Ramen Restaurant
0	Agincourt	0.000000	0.0	0.0	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.000000
1	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0.0	0.0	0.0	0.0	0.25	0.0	0.000000	0.0	...	0.000000
2	Bayview Village	0.000000	0.0	0.0	0.0	0.0	0.50	0.0	0.000000	0.0	...	0.000000
3	Bedford Park, Lawrence Manor East	0.111111	0.0	0.0	0.0	0.0	0.00	0.0	0.111111	0.0	...	0.000000
4	Berczy Park	0.000000	0.0	0.0	0.0	0.0	0.00	0.0	0.142857	0.0	...	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...
57	Victoria Village	0.000000	0.0	0.0	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.000000
58	Westmount	0.000000	0.0	0.0	0.0	0.0	0.50	0.0	0.000000	0.0	...	0.000000
59	Wexford, Maryvale	0.000000	0.0	0.0	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.000000
60	Willowdale South	0.000000	0.0	0.0	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.333333
61	Woburn	0.000000	0.0	0.0	0.0	0.0	0.00	0.0	0.000000	0.0	...	0.000000

We then define a function to sort restaurant types in descending order in each neighbourhood and create a data frame showing the 5 most common type of restaurants in each neighbourhood. The first few rows of this data frame is shown below:

Top 5 Restaurant types for each neighbourhood in Toronto

|	Neighbourhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
0	Agincourt	Latin American Restaurant	Vietnamese Restaurant	Indian Restaurant	Greek Restaurant	Gluten-free Restaurant
1	Bathurst Manor, Wilson Heights, Downsview North	Sushi Restaurant	Chinese Restaurant	Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant
2	Bayview Village	Japanese Restaurant	Chinese Restaurant	Eastern European Restaurant	Hakka Restaurant	Greek Restaurant
3	Bedford Park, Lawrence Manor East	Italian Restaurant	Sushi Restaurant	Comfort Food Restaurant	Greek Restaurant	Indian Restaurant
4	Berczy Park	Seafood Restaurant	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	French Restaurant

Clustering neighbourhoods based on the similarity of restaurant types
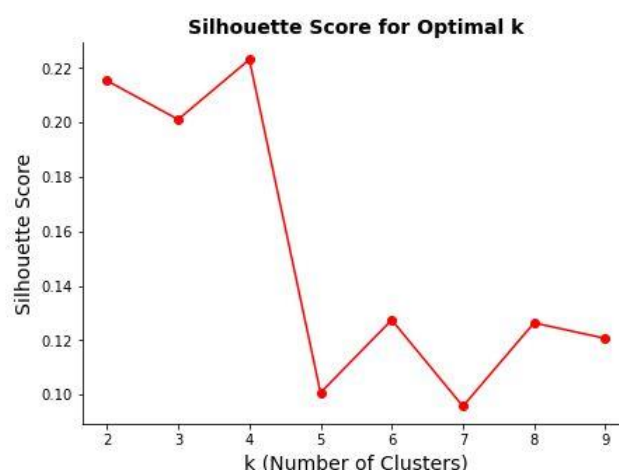
Clustering using K Means

The neighbourhoods will be clustered based on a set of similar characteristics, in this case, based on similarity in types of restaurants present. K-Means clustering, which is used in this part of the analysis, is a machine learning algorithm that creates homogeneous

subgroups/clusters from unlabeled data such that data points in each cluster are as similar as possible to each other according to a similarity measure (e.g., Euclidian distance).

Before proceeding, we need to determine the optimum value of k (number of clusters) using the Elbow Method. It calculates the sum of squared distances of data points to their closest centroid (cluster centre) for different values of k. The optimal value of k is the one after which there is a plateau (no significant decrease in sum of squared distances).
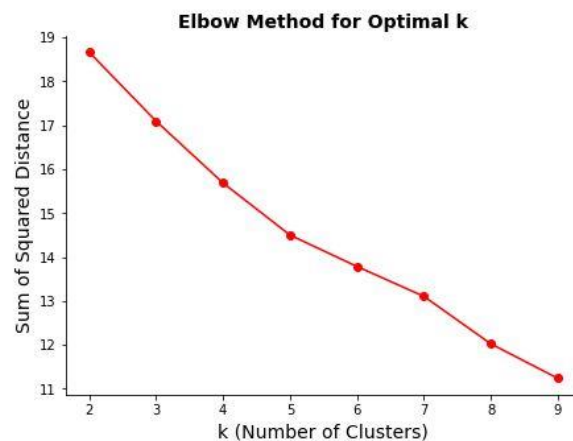
As there is no conspicuous elbow from the plot above, another measure will be used which is known as the Silhouette Score. Silhouette score varies from -1 to 1. A score value of 1 means the cluster is dense and well-separated from other clusters. A value nearing 0 represents overlapping clusters, data points are close to the decision boundary of neighbouring clusters. A negative score indicates that the samples might have been assigned into the wrong clusters.


From the plot above, we can observe that there is a peak at k = 4. Hence, we will proceed with that value as the number of optimal clusters.

Once the clustering is done using K Means, we got the cluster labels for each neighbourhood and merged it with our dataset. The merged data set with cluster labels for each neighbourhood is given below:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
0	M4A	North York	Victoria Village	43.725882	-79.315572	1	Portuguese Restaurant	French Restaurant	Indian Restaurant	Greek Restaurant	Gluten-free Restaurant
1	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1	Greek Restaurant	French Restaurant	Mexican Restaurant	Restaurant	Vietnamese Restaurant
2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant	Gluten-free Restaurant
3	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494	1	Sushi Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Restaurant
4	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	0	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant

There are four clusters of neighbourhoods based on restaurant type. These clusters were colour coded and visualised on the map of Toronto to understand how they are distributed across the region.

Visualising the clusters on a map of Toronto
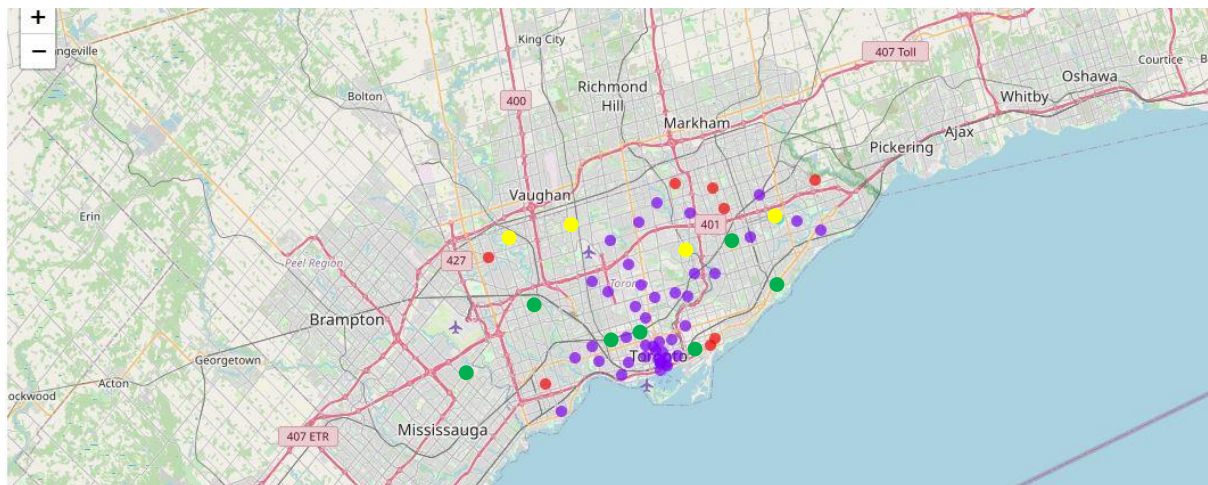
Examination of Clusters

The neighbourhoods belonging to each cluster were filtered out from the dataset previously created. Each cluster was separately analysed to gain an understanding of the distinguishing restaurant types that characterize each of them.

Cluster 1

	Borough	Neighbourhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
4	Scarborough	Malvern, Rouge	0.0	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant
17	North York	Hillcrest Village	0.0	Mediterranean Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	Greek Restaurant
30	East Toronto	India Bazaar, The Beaches West	0.0	Fast Food Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Dim Sum Restaurant
50	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	0.0	Fast Food Restaurant	Italian Restaurant	Thai Restaurant	Chinese Restaurant	Dim Sum Restaurant
54	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	0.0	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant
55	Scarborough	Steeles West, L'Amoreaux West	0.0	Fast Food Restaurant	Chinese Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Hakka Restaurant
60	East Toronto Business	Enclave of M4L	0.0	Fast Food Restaurant	Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Greek Restaurant
61	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	0.0	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant

Neighbourhoods in the first cluster seem to have ***Fast Food Restaurants and Asian cuisine restaurants (such as Vietnamese, Dim Sum, Chinese, Sushi)*** as the most common type present.

Cluster 2

	Borough	Neighbourhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
0	North York	Victoria Village	1.0	Portuguese Restaurant	French Restaurant	Indian Restaurant	Greek Restaurant	Gluten-free Restaurant
1	Downtown Toronto	Regent Park, Harbourfront	1.0	Greek Restaurant	French Restaurant	Mexican Restaurant	Restaurant	Vietnamese Restaurant
2	North York	Lawrence Manor, Lawrence Heights	1.0	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant	Gluten-free Restaurant
3	Queen's Park	Ontario Provincial Government	1.0	Sushi Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Restaurant
6	Downtown Toronto	Garden District, Ryerson	1.0	Thai Restaurant	Fast Food Restaurant	Mexican Restaurant	Ramen Restaurant	Vietnamese Restaurant
7	North York	Glencairn	1.0	Japanese Restaurant	Sushi Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Greek Restaurant
8	North York	Don Mills South	1.0	Restaurant	Italian Restaurant	Asian Restaurant	Chinese Restaurant	Dim Sum Restaurant
9	Downtown Toronto	St. James Town	1.0	Italian Restaurant	Restaurant	Japanese Restaurant	Latin American Restaurant	Middle Eastern Restaurant
10	Scarborough	Guildwood, Morningside, West Hill	1.0	Mexican Restaurant	Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Greek Restaurant
11	Downtown Toronto	Berczy Park	1.0	Seafood Restaurant	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	French Restaurant

Above is a snapshot of a few rows of Cluster 2 as it is a very large cluster. ***This cluster seems to have a heady mix of all types of restaurants -*** Most of the neighbourhoods in this cluster are located in or near the downtown area of Toronto, which being a central commercial area,

has many restaurants. Hence, we can see that many different categories of restaurants can be found here.

Cluster 3

	Borough	Neighbourhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
5	North York	Don Mills North	2.0	Japanese Restaurant	Caribbean Restaurant	Eastern European Restaurant	Hakka Restaurant	Greek Restaurant
16	Scarborough	Cedarbrae	2.0	Hakka Restaurant	Thai Restaurant	Caribbean Restaurant	Dim Sum Restaurant	Greek Restaurant
23	North York	Northwood Park, York University	2.0	Vietnamese Restaurant	Caribbean Restaurant	Falafel Restaurant	Eastern European Restaurant	Hakka Restaurant
32	North York	Humber Summit	2.0	Caribbean Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Hakka Restaurant	Greek Restaurant

The third cluster has **Caribbean and Asian (Japanese, Vietnamese, Hakka etc) cuisine** restaurants as the most common type.

Cluster 4

	Borough	Neighbourhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
21	West Toronto	Dufferin, Dovercourt Village	3.0	Middle Eastern Restaurant	Vietnamese Restaurant	Indian Restaurant	Greek Restaurant	Gluten-free Restaurant
33	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	3.0	American Restaurant	Vegetarian / Vegan Restaurant	Hakka Restaurant	Greek Restaurant	Gluten-free Restaurant
34	East Toronto	Studio District	3.0	American Restaurant	Seafood Restaurant	Comfort Food Restaurant	Middle Eastern Restaurant	Italian Restaurant
40	Etobicoke	Westmount	3.0	Chinese Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Greek Restaurant
41	Scarborough	Wexford, Maryvale	3.0	Middle Eastern Restaurant	Vietnamese Restaurant	Indian Restaurant	Greek Restaurant	Gluten-free Restaurant
43	Central Toronto	The Annex, North Midtown, Yorkville	3.0	Asian Restaurant	Middle Eastern Restaurant	Indian Restaurant	Vietnamese Restaurant	Eastern European Restaurant
45	Mississauga	Enclave of L4W	3.0	American Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Eastern European Restaurant	Greek Restaurant

The fourth cluster has **American and Middle Eastern** restaurants as the most common type. This cluster also has many Mediterranean and Greek Restaurants.

Results and Discussion

In this section, we will document all the findings from the above clustering and visualization of the dataset. Exploratory data analysis as well as machine learning and visualization techniques have provided us with some insights into the data about the problem at hand.

There are a total of 295 restaurants in Toronto. Among all the neighbourhoods there are some which have a large number of restaurants such as

1. 1. Little Portugal, Trinity
2. 2. The Danforth West, Riverdale
3. 3. Central Bay Street
4. 4. Richmond, Adelaide, King,
5. 5. Church and Wellesly

These neighbourhoods have a high concentration of restaurants and thus competition would also be much more. Setting up a new restaurant here does not seem feasible.

On the other hand, neighbourhoods such as,

1. Agincourt
2. Malvern, Rouge
3. Lawrence Manor,Lawrence Heights
4. Forest Hill North and West
5. Dufferin Dovercourt Village
6. Cliffside, Scarborough Village West
7. Westmount
8. Woburn
9. Bayview Village

And many others which have a low concentration of restaurants would be more ideal for starting a new restaurant as these neighbourhoods only have one or two restaurants at most. The complete list can be found in the above section.

We also see that Italian, Fast food and Asian cuisine restaurants are most popularly found in Toronto. Greek, Middle Eastern, Mexican, Indian have medium popularity.

On clustering the neighbourhoods, based on the categories of restaurants present, we found that the first cluster has 8 neighbourhoods and has Fast Food and Asian cuisine Restaurants as the most common type. The second cluster has 43 neighbourhoods and seems to have a mix of all types of restaurants as many of the neighbourhoods in this cluster are in or near the Downtown area of Toronto. The third cluster has 4 neighbourhoods and Caribbean and Asian restaurants are the most common type present.

The fourth cluster has 7 neighbourhoods with American and Middle Eastern as the popular restaurant types.

A business investor or restaurant owner should choose the relevant neighbourhood after considering the preference of cuisine of the people living there. So, for example if a stakeholder would like to open a restaurant in a neighbourhood belonging to the first cluster such as Malvern,Rouge or Hill Crest then a Fast food or Asian Food joint would be a better option.

Alternatively, if a restaurant chain serving a particular cuisine say, American would like to see which neighbourhood to open a franchise outlet, then they should probably consider neighbourhoods in the fourth cluster.

Ideally, a restaurant owner should choose a neighbourhood with low concentration of restaurants and see which cluster it falls in, to determine the type of restaurant that is popular in that area.

However, there could also be a flipside to this argument, which is that some restaurant owners might like to explore new markets and open a restaurant in a neighbourhood which does not have restaurant types of that particular category. This would give them the first mover advantage as there would be no competition from similar restaurants.

Conclusion

The above project can be used as a study to make informed decisions by the relevant stakeholders and restaurant owners while deciding about the best neighbourhood to start a new restaurant and the type of restaurant to start there. However, the optimal solution will vary for different entrepreneurs depending on the business strategy that they adopt. Some might be interested in opening restaurants in the downtown areas as these would have a higher footfall even though there would be fierce competition while others might prefer neighbourhoods which are not as central to reduce competition/rent etc. at the cost of relatively less footfall.

For risk averse investors, choosing restaurant types which are popular in a neighbourhood would seem like a better option even though there would be competition. On the other hand, risk lovers might want to start a restaurant type that does not exist in that neighbourhood at all, in order to tap into new markets and face less competition. Whatever the business strategy chosen, the above project would help the relevant stakeholders to make correct decisions and arrive at the optimal solution.