Find the best Neighborhood in Toronto as a recommendation to open a New Restaurant

Moses Jebaraj

11-Jun-2020

1. Introduction

1.1 Background

For a new Restaurant to be launched in an appropriate neighborhood is not the case with all new restaurants being launched daily across the world. Here we are going to help entrepreneur trying to launch a new restaurant in Toronto, Ontario (Canada). We have a fascinating problem on our hands considering Toronto consistently wins awards for being so lovably liveable, which is contributed by the night life and restaurants. So we got to be precise on recommending a neighborhood for launching a new restaurant. Also Toronto has 8000 plus Restaurants across its neighborhoods, making it an interesting city to find the pattern for an Restaurant to be successful. Our major assumption here after analyzing the initial dataset is that Toronto has lot of restaurant, so we got to locate the similar neighborhoods with most number of restaurants and suggest that as the best place.

1.2 Problem

Data we will be analyzing might include Boroughs, their Neighborhoods, and list of Venue categories around each neighborhood. With that we would be able to find the Borough with most number of restaurants in their top 10 most common venues. This will be our recommendation to launch a New Restaurant.

1.3 Interest

Entrepreneurs will be interested in knowing the neighborhoods which has the pattern for having the Restaurants as their top common venues. Also, considering similar Neighborhoods with same neighborhoods where there are less number of Restaurants, this will be the place where a Restaurant is likely to have better reception and less competition.

2. Data Acquisition and Data Cleaning

2.1 Data Sources

Our initial Borough and Neighborhood details for entire Toronto City can we be scraped from this wiki page here. Then we can use the Four Square API service to extract the list of venue categories around each neighborhood. Geocoder package can be used to extract the Latitude and Longitude data of each Neighborhood we scraped from the Wiki page.

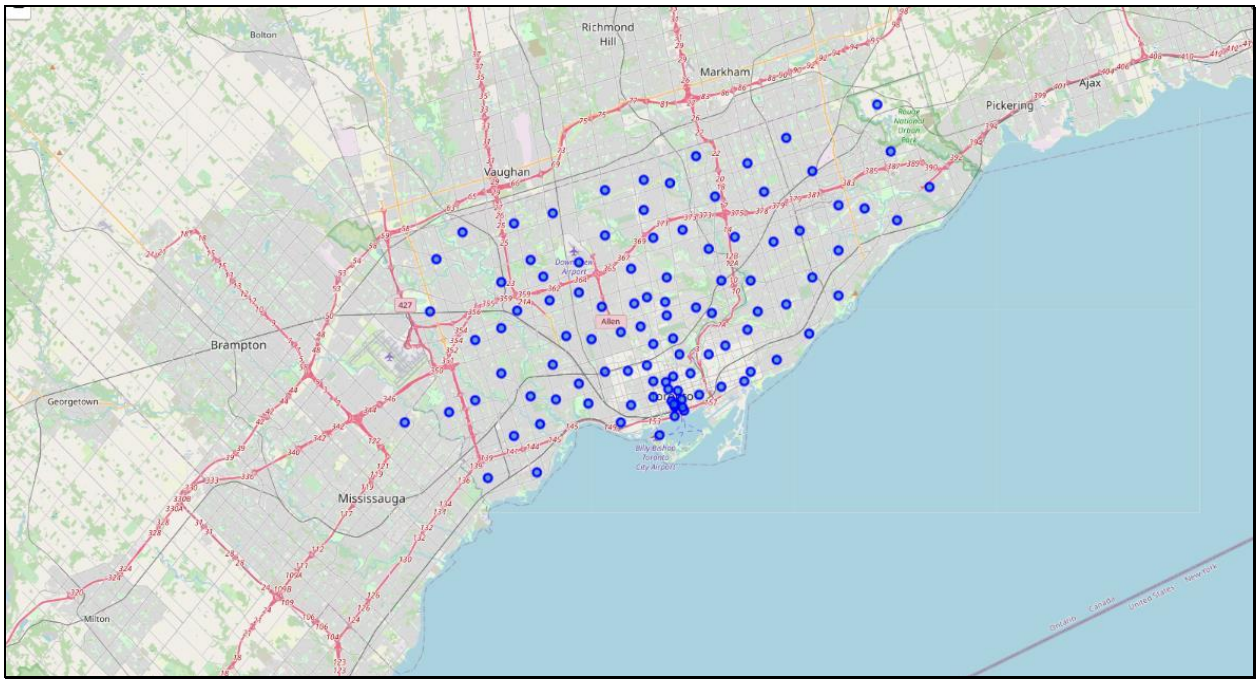
2.2 Data Cleaning

Data I scraped from the wiki page has lot of missing/Non-assigned data, so I created a DataFrame of Borough and Neighborhood removing the Non-assigned data. I am focusing only on the Toronto Boroughs, so I filtered out only the Boroughs containing the string Toronto into a new DataFrame for me to work with.

To identify pattern between these neighborhoods we need to split them into Clusters, once we find the clusters between these neighborhoods, we need to extract each clusters in to different individual DataFrame for us to analyze the Restaurants accumulation in similar neighborhoods.

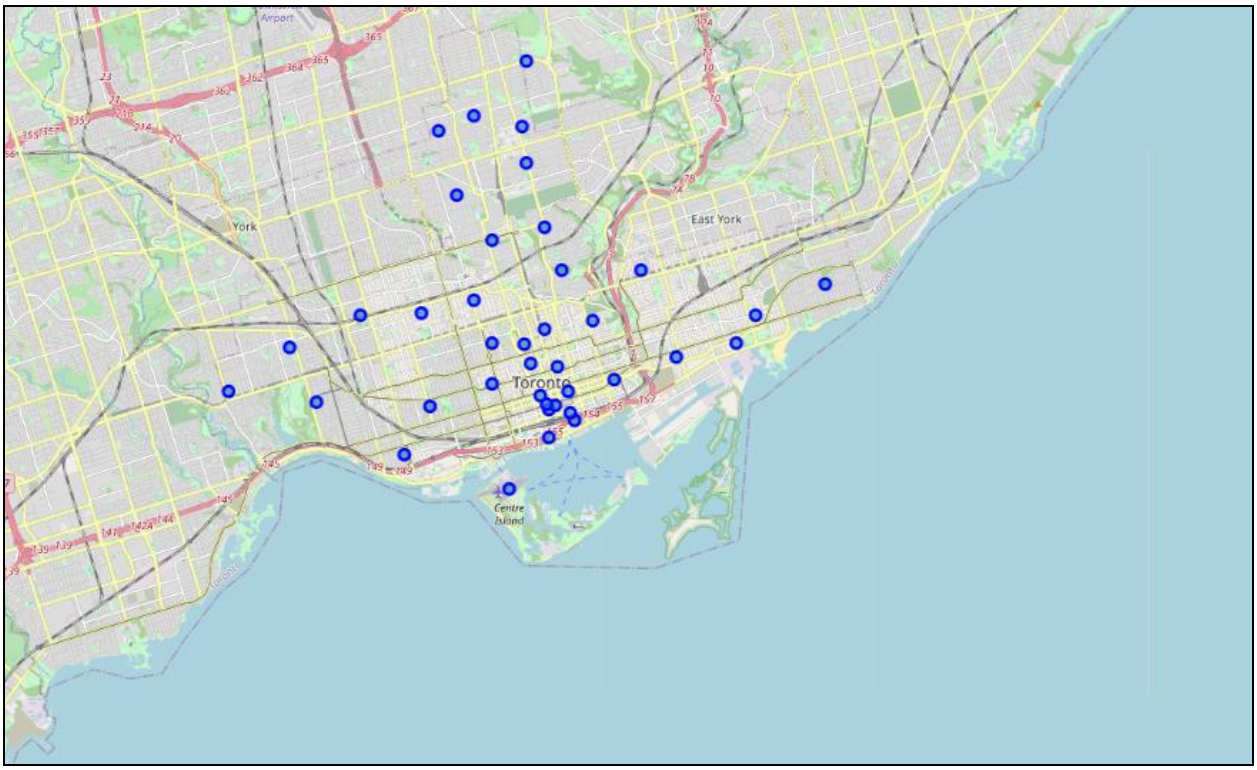
3. Exploratory Data Analysis

Explore the scattering of the Neighborhoods in Toronto city; to view is there are visible clusters. But this is just doesn't mean they are similar neighborhoods. We further need to get the Venues situated in each of Neighborhoods and see if these neighborhoods are similar with respect to the venues situated.



Neighborhood plots of Toronto City

Now, plot the neighborhoods of Boroughs with name as Toronto, and see for any visual cluster.



Now I extract the venue data from each neighborhood using the Four Square API, merge them with the original DataFrame with Boroughs and Neighborhoods. After extracting all the venues, I mainly focus on the venue categories as I need to identify how many restaurants are there in a neighborhood and is it the most common venue in these neighborhoods.

We get the data of venues as a JSON, so I cleaned them up and extracted the venue details and created a DataFrame with them. We can extract the venues of all the neighborhoods from the JSON using a function and looping it through the list of Latitude and Longitude data of neighborhoods in Boroughs with name Toronto.

For each Neighborhood there may be many repeated venue categories, for example: “Christie” might have 5 Café. So to find the most common venues and to have the simplified version of data, I found the frequencies of the venues within a neighborhood.

	Neighborhood	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Baby Store	B
0	Berczy Park	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.017857	0.000000	0.000000	0.000000	0.000000	0.000000	0.017857	0.000000	0.01
1	Brockton, Parkdale Village, Exhibition Place	0.041667	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
2	Business reply mail Processing Centre, South C...	0.055556	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000	0.00
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.0000	0.055556	0.055556	0.055556	0.111111	0.166667	0.111111	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
4	Central Bay Street	0.015625	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.015625	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
5	Christie	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.000000	0.058824	0.00
6	Church and Wellesley	0.025000	0.0125	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.012500	0.000000	0.00	0.000000	0.000000	0.012500	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
7	Commerce Court, Victoria Hotel	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.000000	0.00	0.010000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.00
8	Davisville	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
9	Davisville North	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
10	Dufferin, Dovercourt Village	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
11	First Canadian Place, Underground city	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.00	0.010000	0.000000	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.00

Sample of the table I created for Venue Frequencies

Now it's time to see what are all the top 10 venues for each neighborhood, so I sorted the table in descending order of the frequency mean of venues occurrence in each neighborhood. Created a DataFrame with 10 level s of columns to display top 10 common venues in neighborhoods.

	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	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Cheese Shop	Beer Bar	Bakery	Café	Restaurant	Eastern European Restaurant	Department Store
1	Brockton, Parkdale Village, Exhibition Place	Café	Bakery	Breakfast Spot	Coffee Shop	Gym	Grocery Store	Pet Store	Performing Arts Venue	Nightclub	Italian Restaurant
2	Business reply mail Processing Centre, South C...	Light Rail Station	Yoga Studio	Auto Workshop	Gym / Fitness Center	Garden Center	Garden	Fast Food Restaurant	Farmers Market	Comic Shop	Pizza Place
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Bar	Harbor / Marina	Coffee Shop	Boat or Ferry	Rental Car Location	Boutique	Plane
4	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Burger Joint	Japanese Restaurant	Department Store	Salad Place	Bubble Tea Shop	Yoga Studio
5	Christie	Grocery Store	Café	Park	Restaurant	Diner	Baby Store	Candy Store	Nightclub	Coffee Shop	Athletics & Sports
6	Church and Wellesley	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant	Gay Bar	Yoga Studio	Bubble Tea Shop	Pub	Burger Joint	Café
7	Commerce Court, Victoria Hotel	Coffee Shop	Café	Restaurant	Hotel	American Restaurant	Gym	Seafood Restaurant	Deli / Bodega	Italian Restaurant	Japanese Restaurant
8	Davisville	Sandwich Place	Dessert Shop	Gym	Café	Italian Restaurant	Sushi Restaurant	Coffee Shop	Pizza Place	Farmers Market	Seafood Restaurant
9	Davisville North	Park	Sandwich Place	Breakfast Spot	Gym / Fitness Center	Hotel	Pizza Place	Department Store	Food & Drink Shop	Dim Sum Restaurant	Diner
10	Dufferin, Dovercourt Village	Bakery	Pharmacy	Bank	Bar	Middle Eastern Restaurant	Café	Supermarket	Pizza Place	Park	Pet Store
11	First Canadian Place, Underground city	Coffee Shop	Café	Gym	Japanese Restaurant	Hotel	Restaurant	Deli / Bodega	Salad Place	Seafood Restaurant	Steakhouse
12	Forest Hill North & West, Forest Hill Road Park	Park	Jewelry Store	Trail	Sushi Restaurant	Bus Line	Department Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant
13	Garden District, Ryerson	Clothing Store	Coffee Shop	Bubble Tea Shop	Café	Middle Eastern Restaurant	Japanese Restaurant	Italian Restaurant	Cosmetics Shop	Tea Room	Ramen Restaurant
14	Harbourfront East, Union Station, Toronto Islands	Coffee Shop	Aquarium	Hotel	Café	Sporting Goods Shop	Restaurant	Brewery	Scenic Lookout	Italian Restaurant	Fried Chicken Joint
15	High Park, The Junction South	Thai Restaurant	Mexican Restaurant	Café	Diner	Bakery	Flea Market	Italian Restaurant	Fried Chicken Joint	Speakeasy	Cajun / Creole Restaurant
16	India Bazaar, The Beaches West	Park	Fast Food Restaurant	Gym	Ice Cream Shop	Fish & Chips Shop	Steakhouse	Restaurant	Pub	Italian Restaurant	Burrito Place
17	Kensington Market, Chinatown, Grange Park	Café	Coffee Shop	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Mexican Restaurant	Pizza Place	Dessert Shop	Bar	Gaming Cafe	Grocery Store
18	Lawrence Park	Park	Swim School	Bus Line	Dessert Shop	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant
19	Little Portugal, Trinity	Bar	Asian Restaurant	Restaurant	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Men's Store	Yoga Studio	Cuban Restaurant	Brewery

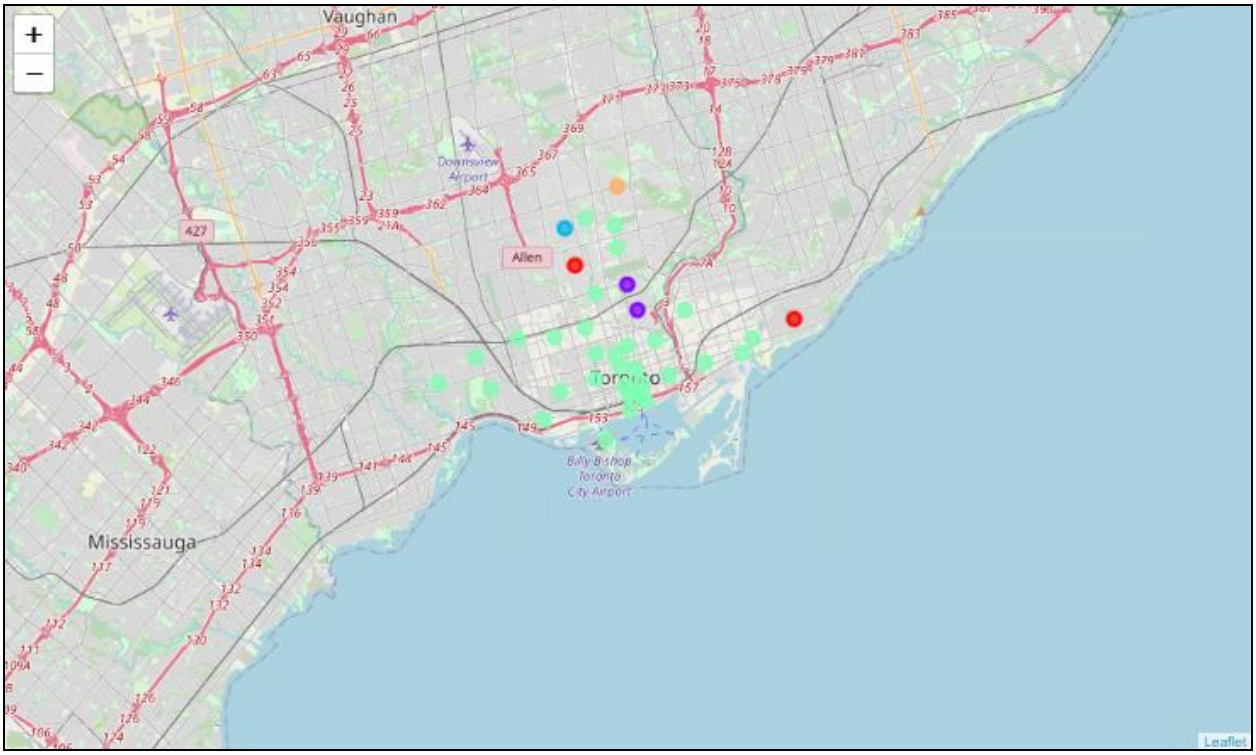
Top 10 common venues in Neighborhood of Toronto Boroughs

4. Classification Modeling

4.1 k-Means Clustering

For our classification of the Neighborhoods which share the similar venues and similar lifestyle, I chose k-Means classification. k-Means clustering is a method of vector quantization, originally from signal processing that aims to partition n observations into k number clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Which will create number of clusters, each having list of similar neighborhoods.

I plan to specify the number of clusters to be 5 as initial setup, which seems to be optimum or the Toronto model. My random state for the k-Means would be 0. This would give me a picture of which neighborhoods share the same pattern of lifestyles.



Clusters of neighborhoods with different Lifestyle

This plot of the Clusters gives us a clear picture that when we travel from the west its mostly of same pattern of venues that is same set of lifestyles, as we move across the town and reach North, it tends to have change in list of venues as there might be different lifestyle. And we can also see an outlier at the far end of the town on the eastern side.

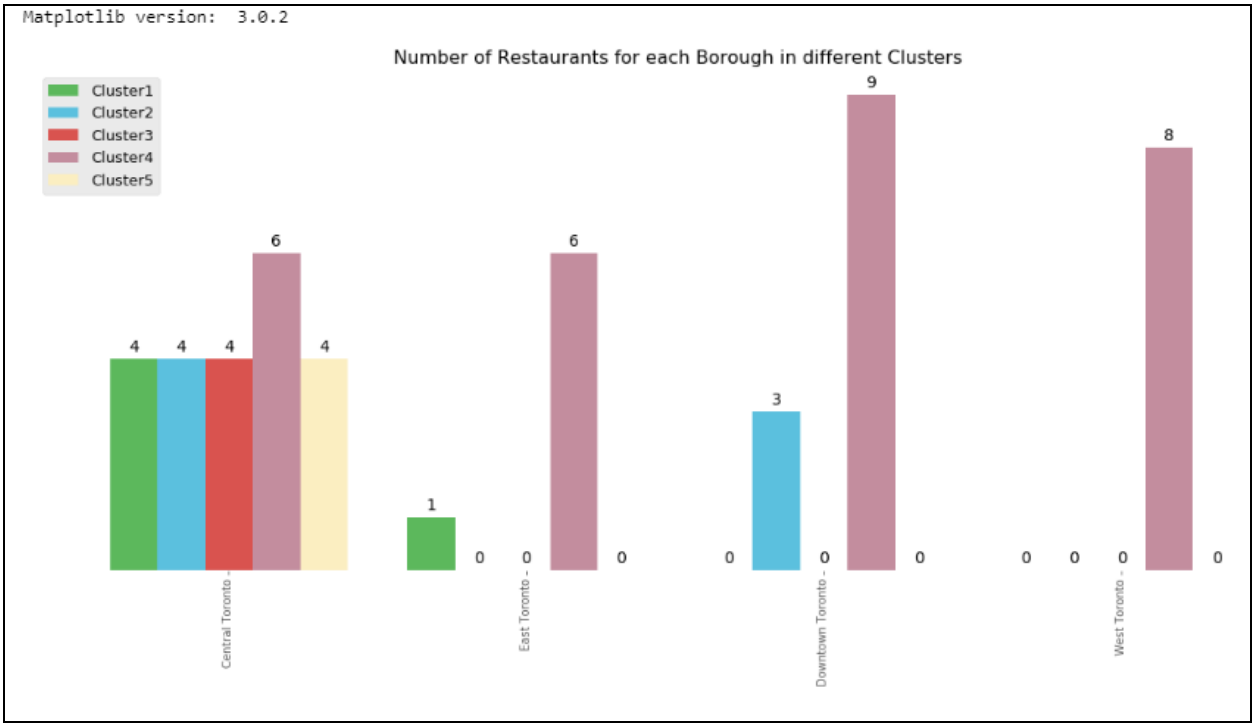
Once these clusters are determined, I placed them in different DataFrames i.e. 5 DataFrames each has different clusters.

To find which Borough has the most number of Restaurants in these clusters, we need to find the venues with the category string as “Restaurant” and find the total count for each borough in each cluster. Which will in-turn give us a DataFrame with Boroughs and Clusters mapped with the number of Restaurants. Most importantly we are referring to the DataFrame we created with top 10 common venues.

While doing this analysis, I found that cluster 4 has repeated set of Borough i.e. it has multiple neighborhoods for same Borough sitting in the same cluster. This will have a difficulty for us to find the restaurant count for each Borough, so I merged the venue data of duplicate Boroughs and have it stored in a single column per Borough.

	Borough	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
0	Central Toronto	4	4	4	6	4
1	East Toronto	1	0	0	6	0
2	Downtown Toronto	0	3	0	9	0
3	West Toronto	0	0	0	8	0

After determining number of restaurants for all the Toronto Boroughs across all clusters, we can generate a DataFrame as above.



4.2 Insights

To get a better view of the story in this DataFrame we extracted, I plotted them using bar plot by grouping them by Boroughs and Cluster

- Insight 1 - Bar plot gives us the clear indication that Cluster 4 has more number of Restaurants, however while looking much closer Cluster 4 seems to have more restaurants and performing well across all Boroughs.
- Insight 2 - Also, Cluster 4 in Central Toronto and East Toronto seems to have similar lifestyle but has less number of Restaurants.
- Insight 3 - Also in all Borough, across clusters 1,2,3 and 5 we can see that Central Toronto has good number of Restaurants and performing well. But in same clusters around other Boroughs does seems to have a Restaurant.

5. Conclusion

In this study, I analyzed the Borough, Neighborhoods across Toronto City with respect to their Venues pattern to identify similar lifestyles. We found varies patterns across the city and tried to determine restaurants count for digging the hidden insights that would help us to recommend a neighborhood for launching a new restaurant. So considering the Insights we derived, I have derived some recommendations, these would be functioning across the business strategy of the entrepreneur.

Strategy 1 – Wanting to Launch a Restaurant in a Neighborhood where majority of venues are Restaurants, considering they would catch the same growth as other venues.

For this strategy, Insight 1 would help to decide as it explains the neighborhoods with most number of restaurants across clusters. So my recommendation would be to launch a new restaurant in Cluster 4 more across all boroughs, but specifically on Downtown Toronto and West Toronto, as they have the most number of restaurants as the top 10 common venues.

Strategy 2 – Wanting to Launch a Restaurant in a Neighborhood where people likely to visit Restaurant but they don't have enough number of Restaurants.

For this strategy, Insight 2 and 3 would help to decide as it explains the neighborhood with opportunity to have more number of restaurants but have very less or no restaurants. So my recommendation would be to launch a new restaurant in Cluster 4 in Central Toronto and East Toronto can be a good place to start a New Restaurant. Also, Central Toronto across cluster 1,2,3 and 5 would be a good choice of Neighborhood to launch a New Restaurant.

6. Future directions

I have included the Venue Category as my most trusted data feature to determine the recommendation of the neighborhood for launching new restaurant. In future, we can also avail other data that Four Square API provides, that is we can also include User reviews and how frequent users visit these restaurants, to derive a rating of these restaurants or identifying most suitable neighborhood which has more number of visitors to the restaurants.