

Introduction: Business Problem

Toronto is the most important and populous city in Canada. It is very diverse and serves as the financial capital of Canada. The city is a center for trade, finance, governmental services, finance, arts, real estate and media in the Canada. The owners of a very popular restaurant chain in the United States want to expand their chain to Canada and would like to open some delis in Toronto to begin their Canadian operations.

Toronto is the fourth largest in North America with more than 2.9 million inhabitants and a metropolitan area population of 5.93 million. It's made up of more than a hundred neighborhoods with different characteristics. Marketing consultants has determined that we must pay special attention to entertainment services, because they are considered attractive to potential customers.

The project will consist of obtaining information of the neighborhoods and make recommendations about best places to start the deli/bodegas chain expansion.

Data

Required data will be sourced from:

- Toronto information, including boroughs, postal codes and neighborhoods, can be obtained from Toronto
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
 - In order to make recommendations about where to situate the delis, Toronto segmentation will be made based on different neighborhoods.
 - Full list of neighborhoods can be obtained from Wikipedia Toronto information, but only their names will be used. They must be geolocated in order to use Foursquare services for obtaining venues.
- For geolocation of neighborhoods Python geocoder will be used.
 - Every neighborhood obtained will be geolocated using python geocoder package, using neighborhood name plus city and country.
 - Geocoder returns longitude and latitude information for every neighborhood center, then it will be used as main Foursquare input.
- In order to obtain venues and their categories FOURSQUARE (<https://foursquare.com/>) will be used.
 - Using services provided by Foursquare one can obtain venues for every neighborhood. Such services require as input geo-localization, it means the latitude and longitude obtained in previously described step.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	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

	PostalCode	Borough	Neighborhood	Latitude	Longitude
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village,Martin Grove Gardens,Richvie...	43.688905	-79.554724
101	M9V	Etobicoke	Albion Gardens,Beaumont Heights,Humbergate,Jam...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest	43.706748	-79.594054

Methodology

A Jupyter Notebook will be developed in order to process data and segment the neighborhoods. Following steps will be implemented:

1. Build neighborhoods list

A list of districts is obtained from Toronto Wikipedia page. That list contains the names of the neighborhoods, boroughs and postal codes for the city. The data for them is then wrangled as in the Peer Reviewed assignment on Toronto clustering and segmentation.

2. Neighborhoods geolocation

Every element in the neighborhoods dataset is geolocated using *Python Geolocator* and two columns are updated containing latitude and longitude for each district or neighborhood by using the name of boroughs containing the word *Toronto*.

The geolocator service has some problems, many times gives time out error, for this reason in this step the information obtained is saved in a text file in CSV format.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

	PostalCode	Borough	Neighborhood	Latitude	Longitude
34	M6P	West Toronto	High Park,The Junction South	43.661608	-79.464763
35	M6R	West Toronto	Parkdale,Roncesvalles	43.648960	-79.456325
36	M6S	West Toronto	Runnymede,Swansea	43.651571	-79.484450
37	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494
38	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558

3. Venues compilation

As next step Foursquare services are used for obtaining venues for every neighborhood. The output is a new dataset with many records for every neighborhood containing the venues found for each of them.

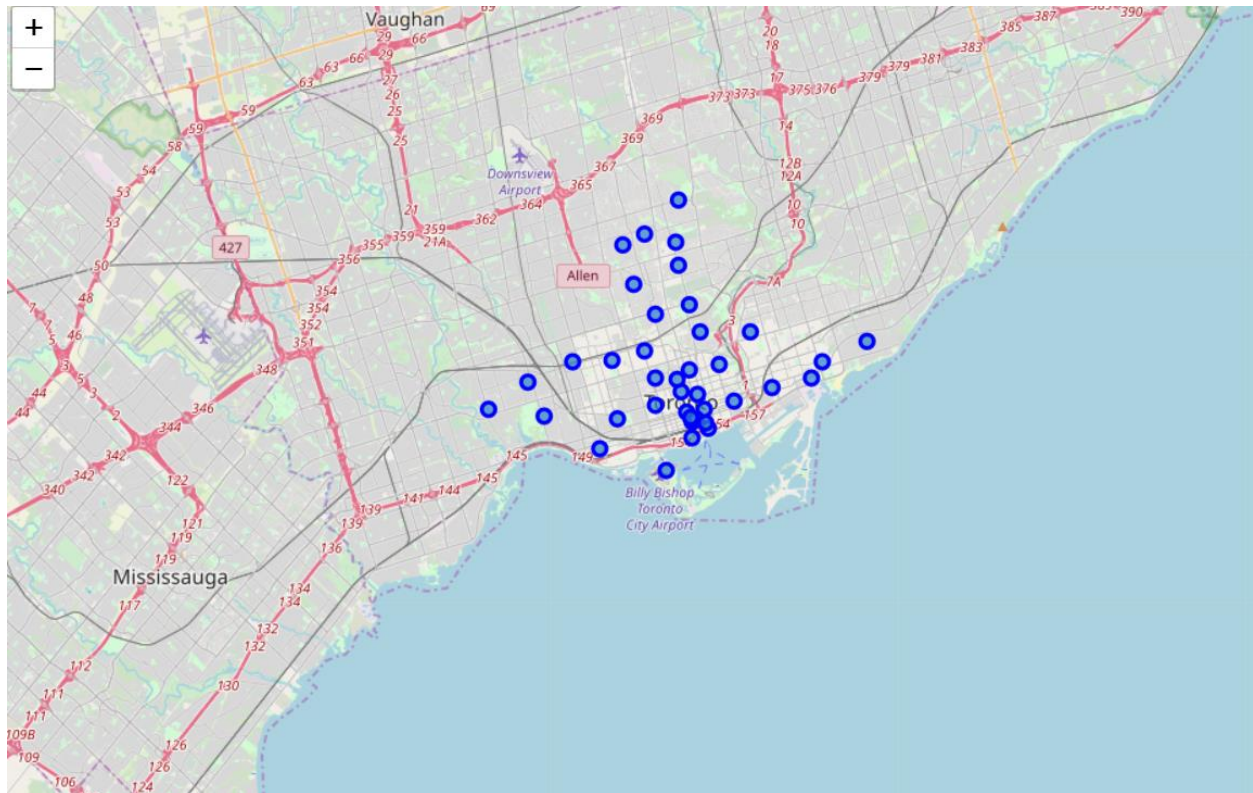
A free Foursquare service with limited count of calls is used. In order to minimize the usage of Foursquare, the information is saved in a text (CSV) file. This was also done using Deli/Bodega category Id `4bf58dd8d48988d146941735` to retrieve data from Foursquare API. After this the explore function is used to get deli categories in the neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	Tori's Bakeshop	43.672114	-79.290331	Vegetarian / Vegan Restaurant
2	The Beaches	43.676357	-79.293031	The Fox Theatre	43.672801	-79.287272	Indie Movie Theater
3	The Beaches	43.676357	-79.293031	Beaches Bake Shop	43.680363	-79.289692	Bakery
4	The Beaches	43.676357	-79.293031	The Beech Tree	43.680493	-79.288846	Gastropub

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3186	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Tim Hortons	43.662644	-79.309945	Coffee Shop
3187	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Measurement Park	43.666916	-79.312631	Park
3188	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Breakfast Club	43.662811	-79.310174	Breakfast Spot
3189	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Carters Landing	43.662414	-79.309898	Bistro
3190	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	Ashbridges Bay Yacht Club (ABYC)	43.658210	-79.311004	Harbor / Marina

`toronto_venues_deli.shape`

(3191, 7)



Deli restaurants in Toronto

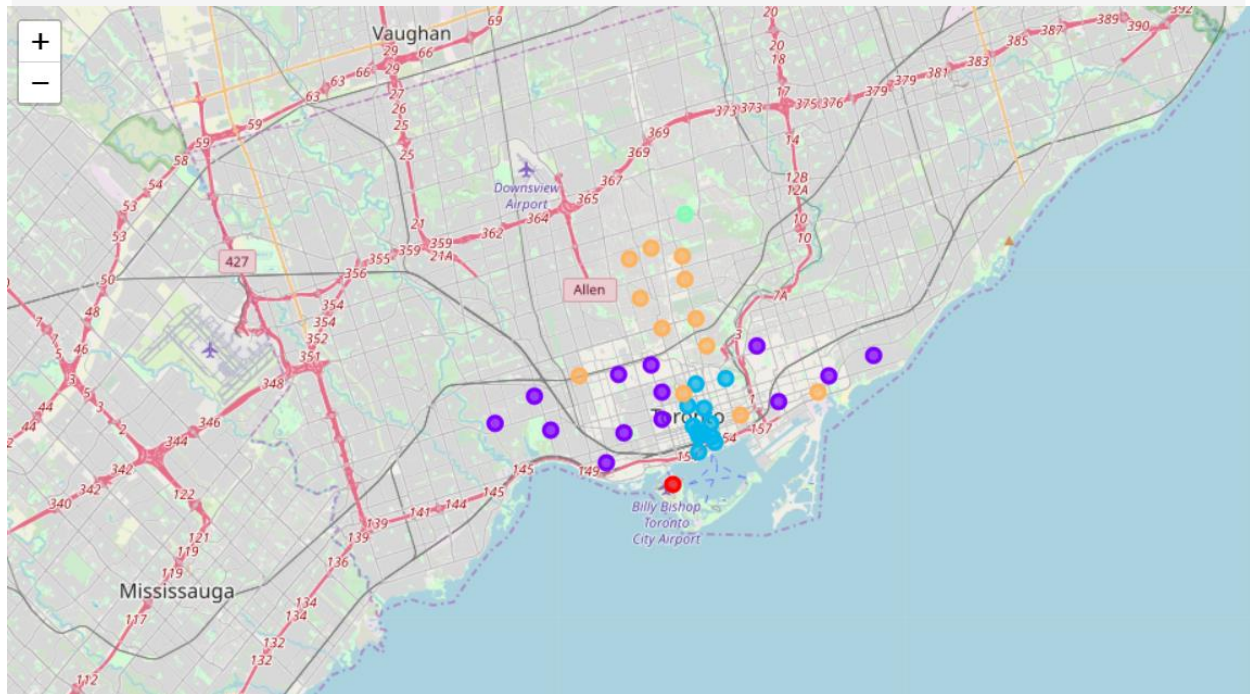
4. Neighborhoods segmentation

The problem on hand is a case of unsupervised segmentation and, from the possible machine learning algorithms, K-means was chosen.

- Taking in account that the venues information obtained from Foursquare is categorical, it must be previously processed in order to be handled by K-means algorithm. For this "*pandas.get_dummies*" is used for dummies variables.
- The list of dummy variables obtained are then grouped as features of every neighborhood.
- K-means algorithm is executed with $K = 5$ chosen.
- Next step is to build the segmentation dataframe, composed of the top venues for every neighborhood plus a segment label determined by K-means.

	Neighborhood	Zoo	Accessories Store	Airport	Airport Lounge	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Deal
0	Adelaide,King,Richmond	0.00	0.00	0.000000	0.000000	0.020000	0.00	0.00	0.00	0.00	0.010000	0.010000	0.020000	0.000000	
1	Berczy Park	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.020000	0.000000	0.000000	0.000000	
2	Brockton,Exhibition Place,Parkdale Village	0.00	0.01	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.010000	0.020000	0.000000	0.010000	
3	Business Reply Mail Processing Centre 969 Eastern	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.020833	0.000000	0.000000	
4	CN Tower,Bathurst Quay,Island airport,Harbourt...	0.00	0.00	0.066667	0.066667	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	

	Neighborhood	Zoo	Accessories Store	Airport	Airport Lounge	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Deal
0	Adelaide,King,Richmond	0.00	0.00	0.000000	0.000000	0.020000	0.00	0.00	0.00	0.00	0.010000	0.010000	0.020000	0.000000	
1	Berczy Park	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.020000	0.000000	0.000000	0.000000	
2	Brockton,Exhibition Place,Parkdale Village	0.00	0.01	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.010000	0.020000	0.000000	0.010000	
3	Business Reply Mail Processing Centre 969 Eastern	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.020833	0.000000	0.000000	
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	0.00	0.00	0.066667	0.066667	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
5	Cabbagetown,St. James Town	0.00	0.00	0.000000	0.000000	0.027027	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
6	Central Bay Street	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.020000	0.020000	0.000000	0.000000	
7	Chinatown,Grange Park,Kensington Market	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.00	0.030000	0.010000	0.000000	0.000000	
8	Christie	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.010000	0.000000	0.000000	0.000000	
9	Church and Wellesley	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.00	0.00	0.00	0.000000	0.010000	0.000000	0.000000	
10	Commerce Court,Victoria Hotel	0.00	0.00	0.000000	0.000000	0.020000	0.00	0.00	0.00	0.00	0.020000	0.000000	0.020000	0.000000	
11	Davisville	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	



The feature above is used to group the neighborhoods into clusters using K-means clustering algorithm. The Folium library helps to visualize the neighborhoods in Toronto and its emerging clusters.

	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	Adelaide,King,Richmond	Coffee Shop	Café	Theater	Hotel	Restaurant	Pizza Place	Pub	Steakhouse	Clothing Store	Beer Bar
1	Berczy Park	Coffee Shop	Café	Hotel	Japanese Restaurant	Restaurant	Beer Bar	Park	Bakery	Seafood Restaurant	Cocktail Bar
2	Brockton,Exhibition Place,Parkdale Village	Café	Coffee Shop	Bar	Bakery	Restaurant	Furniture / Home Store	Tibetan Restaurant	Arts & Crafts Store	Hotel	Lounge
3	Business Reply Mail Processing Centre 969 Eastern	Park	Coffee Shop	Brewery	Pizza Place	Pet Store	Italian Restaurant	Sushi Restaurant	Burrito Place	Bistro	French Restaurant
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	Café	Harbor / Marina	Coffee Shop	Park	Sculpture Garden	Airport	Airport Lounge	Dog Run	Scenic Lookout	Dance Studio

5. Segments analysis

Every segment is printed individually, were different characteristics can be observed for each group.

Next section describes the results.

Results

As a result of segmenting Toronto, five clusters where defined.

Following their characteristics are shown:

Cluster 1

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	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
27	Downtown Toronto	0	Café	Harbor / Marina	Coffee Shop	Park	Sculpture Garden	Airport	Airport Lounge	Dog Run	Scenic Lookout	Dance Studio

Cluster 2

Cluster 2

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	Cluster Labels	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	East Toronto	1	Pub	Coffee Shop	Beach	Pizza Place	Japanese Restaurant	Breakfast Spot	Bar	Café	Sandwich Place	Bakery
1	East Toronto	1	Greek Restaurant	Coffee Shop	Café	Pub	Spa	Ice Cream Shop	Italian Restaurant	Pizza Place	Fast Food Restaurant	Discount Store
2	East Toronto	1	Indian Restaurant	Coffee Shop	Café	Beach	Park	Gym	Fast Food Restaurant	Brewery	Sandwich Place	Restaurant
3	East Toronto	1	Coffee Shop	Bar	Brewery	Café	American Restaurant	Bakery	Vietnamese Restaurant	Italian Restaurant	French Restaurant	Diner
24	Central Toronto	1	Vegetarian / Vegan Restaurant	Café	Coffee Shop	Italian Restaurant	Museum	Gym	Grocery Store	Bakery	Restaurant	Pub
25	Downtown Toronto	1	Café	Bakery	Bar	Vegetarian / Vegan Restaurant	Coffee Shop	Restaurant	Mexican Restaurant	Bookstore	Park	Vietnamese Restaurant
26	Downtown Toronto	1	Café	Vegetarian / Vegan Restaurant	Bar	Dessert Shop	Tea Room	Art Gallery	Bakery	Coffee Shop	Vietnamese Restaurant	Mexican Restaurant
30	Downtown Toronto	1	Korean Restaurant	Café	Coffee Shop	Grocery Store	Ice Cream Shop	Cocktail Bar	Karaoke Bar	Comedy Club	Bar	Ethiopian Restaurant
32	West Toronto	1	Café	Bar	Restaurant	Bakery	Vegetarian / Vegan	Pizza Place	Italian Restaurant	Cocktail Bar	Coffee Shop	Asian Restaurant

Cluster 3

Cluster 3

```
[27]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

[27]:	Borough	Cluster Labels	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
11	Downtown Toronto	2	Japanese Restaurant	Diner	Gastropub	Café	Park	Coffee Shop	Rock Club	Bistro	Steakhouse	Italian Restaurant
12	Downtown Toronto	2	Coffee Shop	Japanese Restaurant	Park	Sushi Restaurant	Café	Men's Store	Bookstore	Bubble Tea Shop	Burger Joint	Restaurant
14	Downtown Toronto	2	Coffee Shop	Cosmetics Shop	Middle Eastern Restaurant	Restaurant	Clothing Store	Diner	Italian Restaurant	Japanese Restaurant	Plaza	Tea Room
15	Downtown Toronto	2	Coffee Shop	Café	Restaurant	Hotel	Bakery	Cosmetics Shop	Italian Restaurant	Breakfast Spot	Seafood Restaurant	American Restaurant
16	Downtown Toronto	2	Coffee Shop	Café	Hotel	Japanese Restaurant	Restaurant	Beer Bar	Park	Bakery	Seafood Restaurant	Cocktail Bar
17	Downtown Toronto	2	Coffee Shop	Café	Japanese Restaurant	Italian Restaurant	Cosmetics Shop	Park	Tea Room	Thai Restaurant	Bar	Burger Joint
18	Downtown Toronto	2	Coffee Shop	Café	Theater	Hotel	Restaurant	Pizza Place	Pub	Steakhouse	Clothing Store	Beer Bar
19	Downtown Toronto	2	Hotel	Coffee Shop	Aquarium	Café	Gym	Park	Scenic Lookout	Japanese Restaurant	Restaurant	Brewery
20	Downtown Toronto	2	Coffee Shop	Café	Hotel	Steakhouse	Restaurant	Italian Restaurant	Thai Restaurant	Concert Hall	Park	Monument / Landmark
21	Downtown Toronto	2	Coffee Shop	Café	Hotel	Restaurant	Gastropub	Steakhouse	Japanese Restaurant	Beer Bar	Concert Hall	Farmers Market

Cluster 4

Cluster 4

```
[28]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

[28]:	Borough	Cluster Labels	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
4	Central Toronto	3	Pool	College Quad	Bookstore	Café	Trail	Coffee Shop	College Gym	Gym / Fitness Center	Park	Design Studio

Cluster 5

Cluster 5

```
[29]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

```
[29]:
```

	Borough	Cluster Labels	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
5	Central Toronto	4	Coffee Shop	Italian Restaurant	Café	Pharmacy	Gym	Sushi Restaurant	Dessert Shop	Pizza Place	Fast Food Restaurant	Movie Theater
6	Central Toronto	4	Park	Coffee Shop	Italian Restaurant	Sporting Goods Shop	Café	Mexican Restaurant	Diner	Skating Rink	Sushi Restaurant	Food & Drink Shop
7	Central Toronto	4	Coffee Shop	Italian Restaurant	Sushi Restaurant	Middle Eastern Restaurant	Pizza Place	Café	Dessert Shop	Indian Restaurant	Gym	Pub
8	Central Toronto	4	Coffee Shop	Italian Restaurant	Gym	Grocery Store	Park	Bagel Shop	Sandwich Place	Thai Restaurant	Restaurant	Café
9	Central Toronto	4	Coffee Shop	Sushi Restaurant	Park	Italian Restaurant	Thai Restaurant	Gym	Liquor Store	Grocery Store	Pizza Place	Pub
10	Downtown Toronto	4	Coffee Shop	Park	Grocery Store	Filipino Restaurant	Sandwich Place	Bistro	Bank	Japanese Restaurant	BBQ Joint	Trail
13	Downtown Toronto	4	Coffee Shop	Pub	Diner	Café	Park	Theater	Bakery	Restaurant	Breakfast Spot	Italian Restaurant
22	Central Toronto	4	Sushi Restaurant	Italian Restaurant	Café	Bank	Pharmacy	Coffee Shop	Japanese Restaurant	Dry Cleaner	Gastropub	Bakery
23	Central Toronto	4	Park	Café	Coffee Shop	Skating Rink	Italian Restaurant	Gym / Fitness Center	Burger Joint	Sushi Restaurant	Trail	Bank
31	West Toronto	4	Café	Coffee Shop	Park	Sushi Restaurant	Portuguese Restaurant	Bar	Italian Restaurant	Supermarket	Bakery	Art Gallery

Discussion

The aim of this project was to help a deli business trying to open new restaurants in Toronto find the best locations to set up in order to maximize their return on investments. It was observed that both **Clusters 1** (Downtown Toronto) and **4** (Central Toronto) did not have a lot of deli places and hence it would be worthwhile on the part of the company to set up their restaurants in those places. **Clusters 2, 3** and **5** seem to be places of extremely high competition among deli businesses and hence would be risky to open the business from there.

Conclusion

Data was gathered from trusted sources and strong methodology was used for processing. The goals of the project were met and with some additional work can be developed to support the opening of a deli restaurant in an unknown location in a city. The future of this project includes taking other factors such as the cost of living in that cluster into consideration in order to recommend an area for opening the deli business.