# Introduction

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of more than 2.7 Million. It 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.

Toronto is remarkable in diversity in its food and restaurants. Pakistani, Persian, Portuguese; aboriginal and new fusion; Japanese pancakes and Korean barbecue; fresh pasta in Little Italy, shawarmas in Greektown and the best damn dumplings in Chinatown. Torontonians love to eat out, whether it's sitting at sidewalk bistros on a warm summer night or getting all bundled up for some hot Vietnamese pho.

So, I would love to explore a new investment and business in food. If we think of both city residents and travelers, they may want to choose different types of food. At the same time, they may want to choose local, district or signature Toronto's dishes.

In our business problem, we will perform analysis to understand the current market and decide which market and types of food that we will invest, and create a map with clusters to reveal which areas we should consider to start our business.

# Data

The data that I used include

1. Table of postal codes in Toronto will be used to identify the Postal Code, Borough and Neighborhood

Data source: Wikipedia

1. Geospatial Data of Canada will be used in order to get the latitude and longitude coordinates of a given postal code in Canada as well as Toronto.

Data source: Geospatial Data of Canada from Coursera

1. Foursquare API data will be used to explore venues around Toronto, the top 5 venues around these areas, explore users' tips, and finally perform clustering.

Data source: Foursquare

# Methodology

1. **Data Preprocessing and Visualisation – Borough and Neighborhood in Toronto**

In order for us to understand the data and explore different types of data analysis, the first exercise that I need to do is to extract the data, clean the data and understand the data.

In the first dataset, I establish a dataframe which consist of three columns: Postal Code, Borough and Neighborhood. The boroughs that has unassigned borough are removed; The neighborhoods that has unassigned values are assigned with the name of Borough; Rows with different neighborhoods that under the same postal codes are combined. The dataframe is shown below.

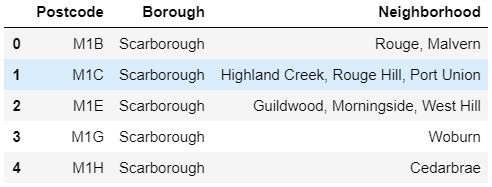


Figure 1. Dataframe with Postcode, Borough and Neighborhood

Once I build a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name, I also need to get the latitude and the longitude coordinates of each neighborhood, which will be later used in Foursquare API for location data. Therefore, I extracted the data Geospatial Data of Canada from Coursera and combined with the Dataframe shown above. The result is shown below.

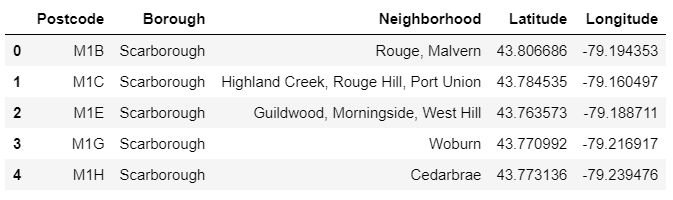


Figure 2. Dataframe with Postcode, Borough, Neighborhood, Latitude and Longitude

Since only Toronto is considered, I filter the Borough that contains ‘Toronto’ only. As a result, the dataframe now only contains Borough with Central Toronto, Downtown Toronto, East Toronto and West Toronto.

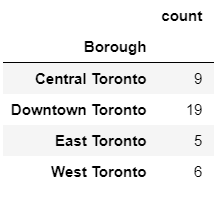


Figure 3. Groupby function to show the number of rows in each of Borough

The dataframe is prepared and with the use of Geopy and Folium, a local map is generated to visualize the neighborhoods that within each Borough in Toronto, as shown in figure 4.

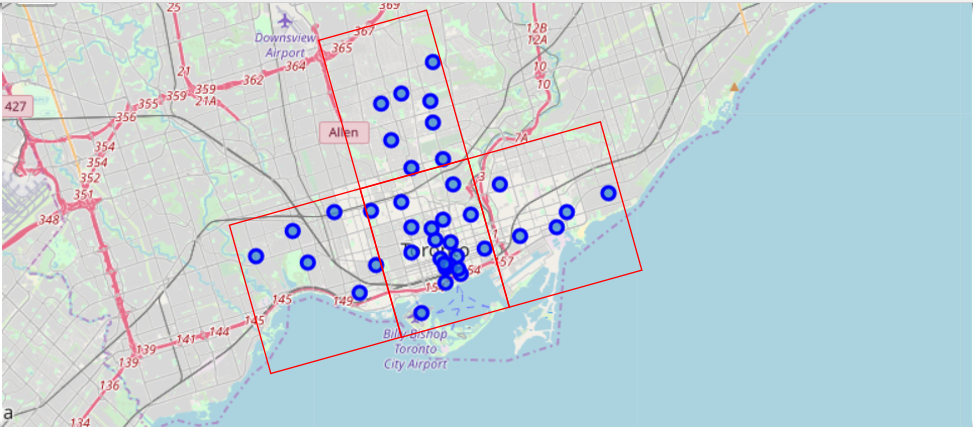


Figure 4. Neighborhoods in the four Boroughs in Toronto

1. **Data Preprocessing and Visualisation – Venues in Toronto**

To explore the venues around Toronto area, a URL is constructed to make a request to the Foursquare API. Nearby venues of each neighborhood are extracted and their associated categories, latitude and longitude are also summarized in a dataframe.

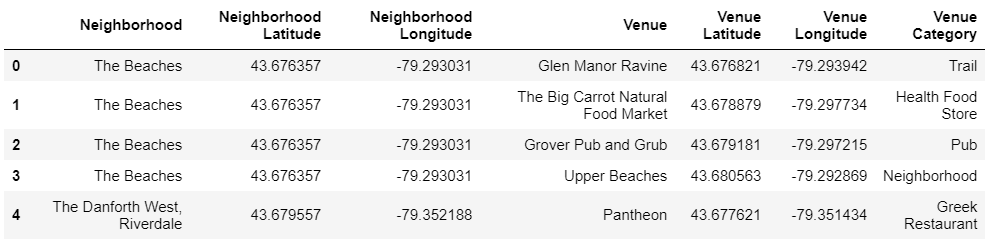


Figure 5. Nearby venues in Toronto

However, in this exercise, my main concern is about the restaurant. That means many other types of venue category are out of the scope, such as “airport”, “airport gate”, etc. So, I add a filtering condition in this dataframe and only venue category contains “Restaurant” will be needed for further analysis. A full table is summarized below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Venue Category** | **Count** | **Venue Category** | **Count** |
| **Afghan Restaurant** | 1 | **Greek Restaurant** | 15 |
| **American Restaurant** | 22 | **Indian Restaurant** | 8 |
| **Asian Restaurant** | 13 | **Italian Restaurant** | 44 |
| **Belgian Restaurant** | 1 | **Japanese Restaurant** | 29 |
| **Brazilian Restaurant** | 2 | **Korean Restaurant** | 2 |
| **Cajun / Creole Restaurant** | 1 | **Latin American Restaurant** | 5 |
| **Caribbean Restaurant** | 4 | **Mediterranean Restaurant** | 4 |
| **Chinese Restaurant** | 13 | **Mexican Restaurant** | 14 |
| **Colombian Restaurant** | 2 | **Middle Eastern Restaurant** | 6 |
| **Comfort Food Restaurant** | 6 | **Modern European Restaurant** | 2 |
| **Cuban Restaurant** | 2 | **Molecular Gastronomy Restaurant** | 1 |
| **Dim Sum Restaurant** | 1 | **New American Restaurant** | 7 |
| **Doner Restaurant** | 1 | **Portuguese Restaurant** | 3 |
| **Dumpling Restaurant** | 4 | **Ramen Restaurant** | 5 |
| **Eastern European Restaurant** | 3 | **Seafood Restaurant** | 24 |
| **Ethiopian Restaurant** | 2 | **Southern / Soul Food Restaurant** | 1 |
| **Falafel Restaurant** | 2 | **Sushi Restaurant** | 19 |
| **Fast Food Restaurant** | 10 | **Taiwanese Restaurant** | 1 |
| **Filipino Restaurant** | 1 | **Thai Restaurant** | 19 |
| **French Restaurant** | 10 | **Theme Restaurant** | 1 |
| **German Restaurant** | 1 | **Vegetarian / Vegan Restaurant** | 18 |
| **Gluten-free Restaurant** | 3 | **Vietnamese Restaurant** | 9 |

Figure 6. Nearby restaurants in Toronto

Once I establish a list of the nearby restaurants in dataframe, the top 10 restaurant types in each of the neighborhood are then calculated and summarized below.



Figure 7. Top 10 nearby restaurants in each of the neighborhood in Toronto

Based on how frequent that the restaurants appearing in the neighborhood, I sum up all the frequency values and found the top 3 types restaurants in each neighborhood, which are Italian, Sushi and Fast Food Restaurant. And these three types will be our final target investment in restaurant.

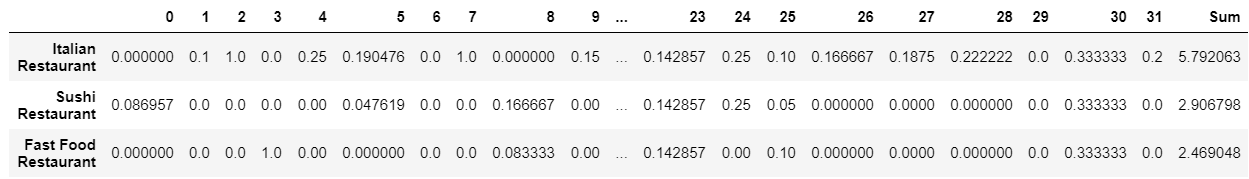


Figure 8. Top 3 types of restaurants in each of the neighborhood in Toronto

1. **Restaurants Clusters in Toronto**

To find out the commonality of the types of restaurants in Toronto, clustering technique which is an unsupervised approach, is used. In total, five clusters are applied and the associated top ten common restaurants are shown.

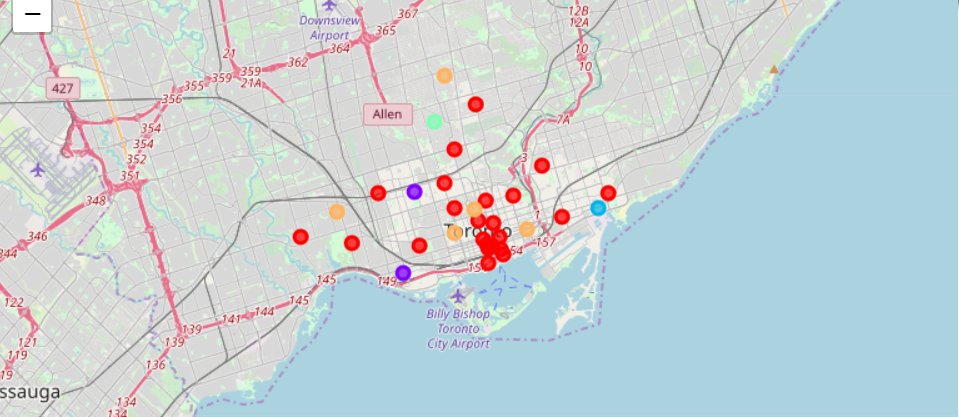
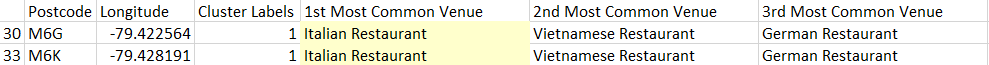


Figure 9. The five clusters in Toronto

**Cluster One**



**Cluster Two**



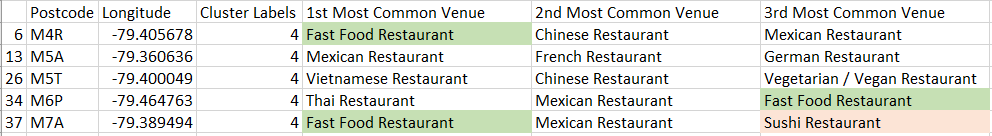
**Cluster Three**



**Cluster Four**



**Cluster Five**



# Results and Discussion

1. **What to invest?**

Referring to figure 8, based on how frequent that the restaurants appearing in the neighborhood, I sum up all the frequency values and found the top 3 types restaurants in each neighborhood, which are

1. Italian Restaurant,
2. Sushi Restaurant; and
3. Fast Food Restaurant.

With limited information, I base on the fact that the more the restaurant appearing, the easier the restaurant that will be accepted by the public. As a result, I decide to invest in these three types of restaurant.

1. **Where to invest?**

For each of the clusters shown above, again with limited information and basing on the fact the more the restaurant appearing, the easier the restaurant that will be accepted by the public, I would recommend considering to invest

1. Italian Restaurant in Cluster One and Two, targeting at Downtown Toronto
2. Sushi Restaurant in Cluster Four, targeting at Central Toronto
3. Fast Food Restaurant in Cluster Three and Five, targeting at Downtown Toronto and East Toronto

# Conclusion

After performing several analyses, the current market share and trend of restaurant in each neighorhood are revealed. Considering the frequency and trend, three types of restaurant are targeted. And by performing a clustering analysis, the preferred locations are identified for each of the targeted restaurant. Further analysis will need to be conducted, such as financial and legal situation, customers’ preferences, investment types and etc, to make a more comprehensive and robust investment plan.