**Recommendation for Restaurant in Toronto, Canada**

1. **Introduction** 
   1. **Background**

Shifting demographics and changing lifestyles are surge in food-service businesses. Busy customers don’t have time to cook food they want to appetite their thrust without the hustle of baking that increase the demand of restaurants. Though the future look bright for the restaurants industry but there are no guarantees for the success. Many restaurants fail during their first year, due to the lack of proper analysis and planning. Location is the important factor which become the success of restaurant business.

Toronto is the capital of the Ontario and is the most populous city in the Canada. It is the multicultural and the hub of tourist attraction. Proper analysis of the location and the type of food people like in the area are the key to success. Competition is also important factor which need to be consider.

* 1. **Business Problem**

Data that help to identify the location that best for new restaurant business. Need a neighbourhood with steady stream of customer visiting other restaurants, which hopefully not meet the need in the area. The aim of this project is to find the interesting point of interest for restaurant business.

* 1. **Interest**

The target audience of this project are the investors who want to invest in the restaurant industry and the entrepreneur who want to start their restaurant business in the Toronto. One more important character which get benefit from this project are the food lovers who can easily identify in which area he/she can find their choice of restaurants.

1. **Data acquisition and cleaning** 
   1. **Data Source**

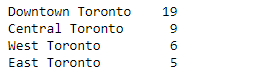
To get location and the other information of the city Toronto the data is collected from the Wikipedia [1] we scrap the postal code, borough and the neighbourhood data from the Wikipedia page. After getting the postal code we found the coordinates of each postal code from the geocoder library provided by the Google. Geocoding is the process of converting the human readable addresses into the geographic coordinates (like latitude 37.233, longitude -122.266) which is used to place marker and position the map.

Foursquare API are used to fetch venue information and analyse different neighbourhood. To fulfil the requirement of our problem we used the longitude and latitude and take the venues information we will look at the specific type of data from the foursquare API that help in our analysis process and help us to identify the most popular category of restaurants in each location which are of our interest.

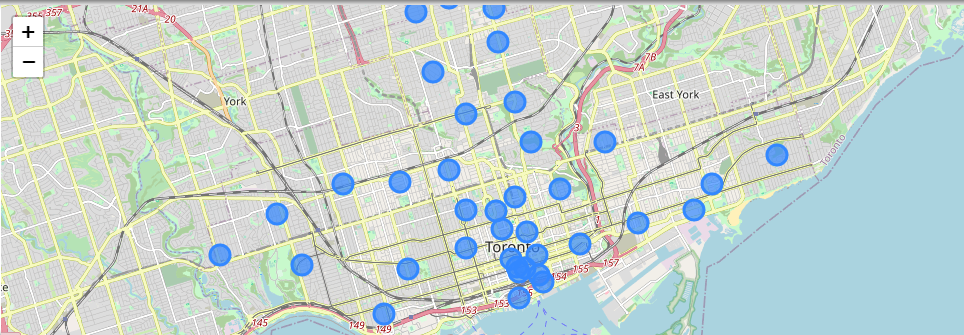
* 1. **Data Cleaning**

Data of postal code and the borough scrapped from the Wikipedia has missing values and some records have ‘not assigned’ value. We have few options to deals with the missing and not assigned data but the quantity of that records is not much so we had decided to remove this data because it will help in our analysis process.

After fixing the above problems, we check for repeated data but there is no repeated data found in our data. In our dataset we have data of the whole city of Toronto, but we are interested into the borough whose name contains Toronto keyword, so we filter the data and take only the neighbourhood within the following borough.



Plot of neighbourhood on the map of Toronto with the help of folium library.



*Fig 1 - Neighbourhood in the downtown, central, west and east Toronto*

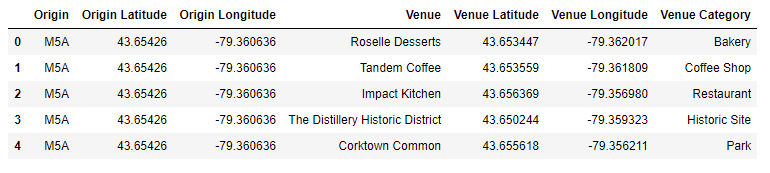
1. **Methodology and Exploratory Data Analysis**

After filtering our neighbourhood from the dataset, we request Foursquare API for the venues at each neighbourhood within the radius of 1 km and the limit of 100 nearby venues by providing the longitude and latitude of our selected neighbourhood. Foursquare API return a lot of information about each venue we filtered the interested information which help in our analysis.

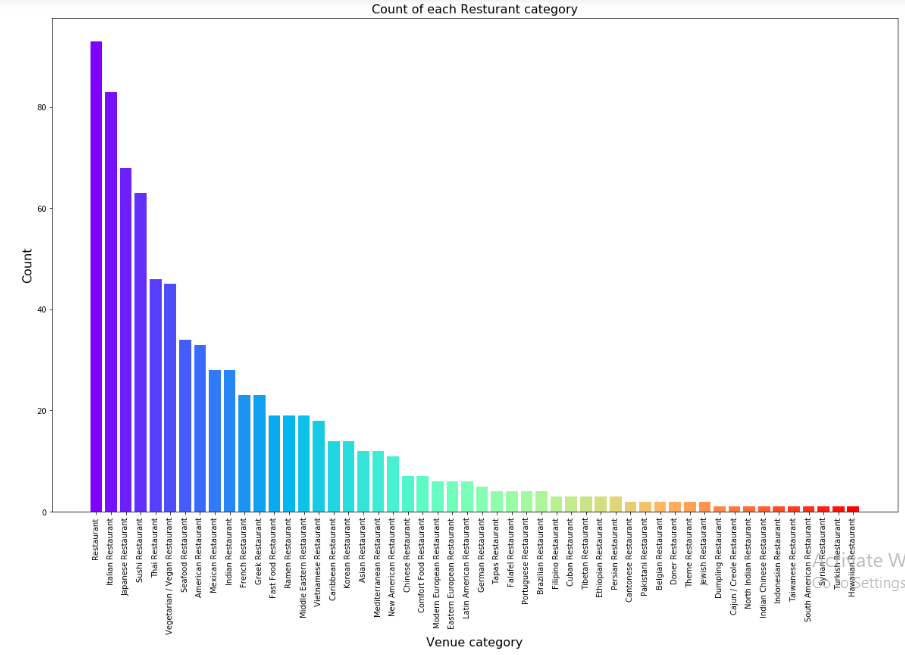
From Foursquare API endpoint [2] we retrieved the following for each venue.

* **Name** = Name of the venue
* **Longitude** = geographical coordinates of the venue
* **Latitude =** geographical coordinates of the venue
* **Category** = Venue category

Wemerge the venues data with the neighbourhood data the shape of our data is now (3188, 7).

*Fig 2 - Origin with venues and their categories*

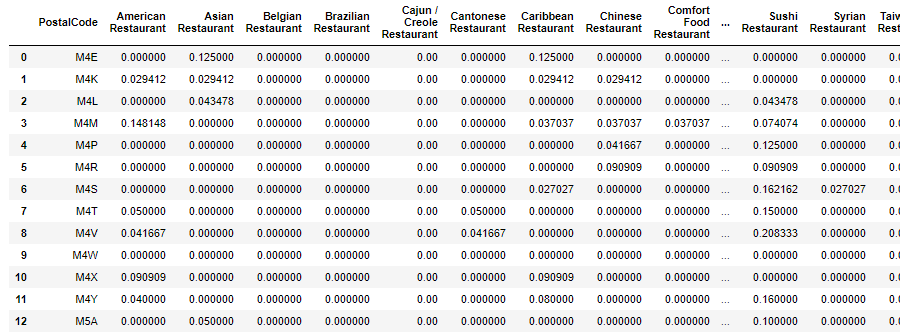
In the venue category we found 277 unique categories of venues in our Dataset. So, we filter the venues categories with contains the keyword restaurant because we are analysing the restaurant category. After filtering we have 52 unique categories of restaurant in the Toronto.



*Fig 3 - Number of Restaurant of each category in the Toronto*

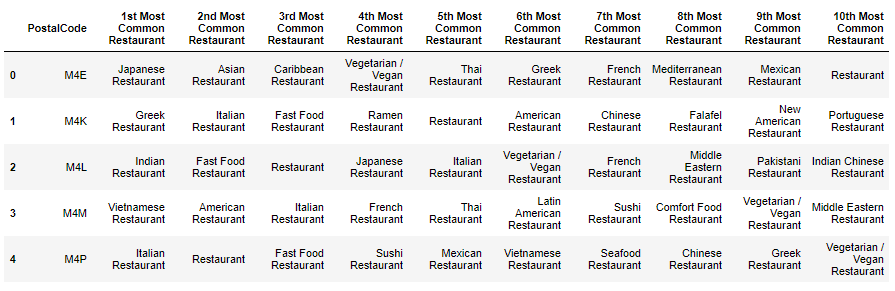
Figure 3 shows that in Toronto the most popular categories of restaurant are general, Italian and the Japanese restaurant. Furthermore, we identify the total number of venues at each neighbourhood which identify the total number of venues and the opportunity for the new venues.

Finally, we analyse each neighbourhood separately to further explore our point of interests. We group row with postal code and take the mean of the frequency of occurrence of each restaurant category.



*Fig 4 - Mean frequency of occurrence category at each neighbourhood*

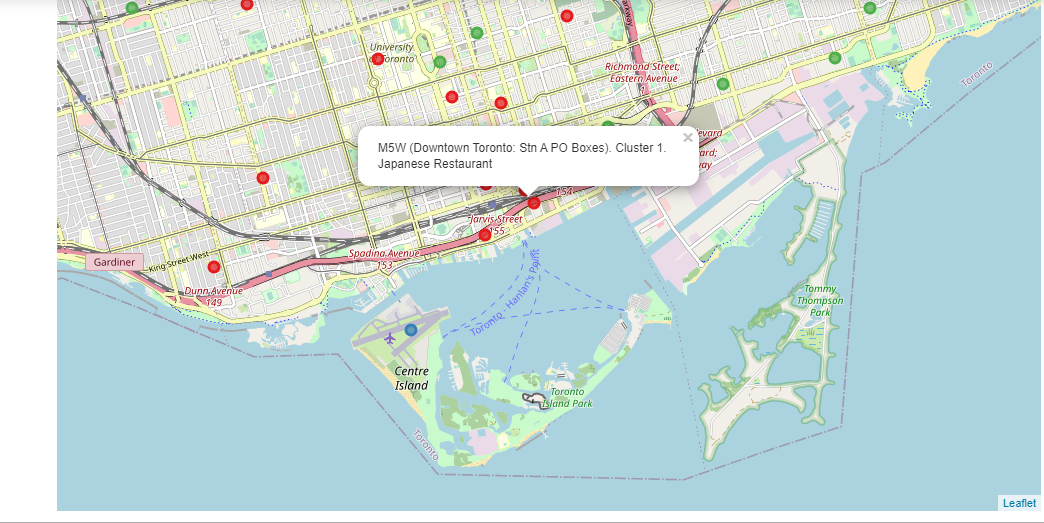
Sort the restaurant category mean frequency in the descending order help us to identify the most famous category. Finally, we found the top 10 most common restaurant category in the neighbourhood.



*Fig 5 - Top 10 most common restaurant*

* 1. **Clustering**

Finally, we used the cluster analysis in the unsupervised learning clustering is used to identify the interesting patterns in the data by grouping them based on the common features and the behaviours. Clustering has many versions in our analysis we are using k-mean clustering which is the most common algorithm it is the fastest and minimize the variance within each cluster. We choose to divide our data into 5 clusters.



*Fig 6 - Clustering of restaurants venues*

In Figure 6 we see the 5 clusters the clusters are identify with the colours each neighbourhood are coloured with in which cluster it is belong. On clicking the neighbourhood, we can see the postal code, name of the neighbourhood, cluster in which it belongs and the top restaurant category in that neighbourhood.

1. **Result**

Based on the above analysis performed, the following results can be finalized

* Restaurant, Italian and Japanese restaurant are the most common restaurant in the central Toronto City.
* Indian, Mexican and American restaurant have less competition and there is opportunity for these categories.
* Syrian, Turkish and Hawaiian restaurants are the least common.
* Clustering group the categories into group to further analyse the location. For example, if you are interested in opening Indian restaurant then cluster 1 is best for you because its on the 4th position on the most common list there is a margin for the new restaurant.

1. **Conclusion**

The purpose of this project is to explore the best appropriate location to open a new restaurant in the city of Toronto by analysing the existing restaurants in the city their location. The city postal code borough and the longitude and latitude of the neighbourhood are found from the Wikipedia and the google geocoder. We focus on the central Toronto which has the higher population and the point of interest for the visitors so here the opportunity for new entering in this business. Foursquare provided us the venues at each location by analysing the frequency restaurant category it helps us to identify the customer likeness and the taste of restaurant they want. There are various others factors which need to be consider before opening a restaurant such as cost of the property, menus, parking places, customer need and many more but this analysis help the investor and the entrepreneur to analyse the city locations where other restaurants are located their neighbours and the most popular category of restaurant for each neighbourhood.

1. **References**

[1]<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

[2] https://api.foursquare.com/v2/venues/search