**Battle of the Neighbourhoods**

**Open new restaurant in Toronto**

April 11, 2019

**1. Introduction**

The entrepreneur is planning to open a new restaurant in Toronto, but he is not sure which location would be most appropriate for his new venue. We noticed that the Toronto already has a lot of restaurants in town, but we need to help this entrepreneur to find this location.

**2. Data**

We had to discover the most important factors that contribute to the restaurant’s success. We can expect these factors to be among the following list: neighbourhood wealth, accessibility, crime rates, visibility, competition, etc. We should use the datasets from Toronto Opendata website to address some of these considerations.

From there, we were able to get the city’s average housing prices list. We will be working with Get Wellbeing Toronto - Economics data set that includes average house price by neighbourhood. Also, we were using the “Foursquare” location data to retrieve the food venues. We would use Foursquare location data in conjunction with the average house price by neighbourhood to determine the best possible location for a new restaurant.

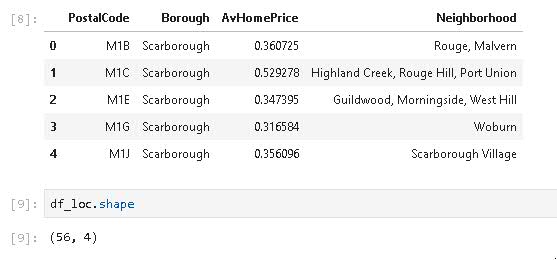
We started with the dataset called “Get Wellbeing Toronto – Economics” which included the average house price by neighbourhood This dataset contains the child case spaces, debt risk score, home prices, local employment and social assistance recipients for 138 neighbourhoods. Due to the scope of this project we were only interested in home prices data.

Next, we’ve got the Postal code data for the Toronto. To create this dataframe we had to parse <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> html page using Beautiful Soup. After removing all rows where we had not assigned values in borough column and replacing the not assigned values in neighbourhood with corresponding values of borough we end up with the dataframe for 211 postal codes.

**3. Methodology**

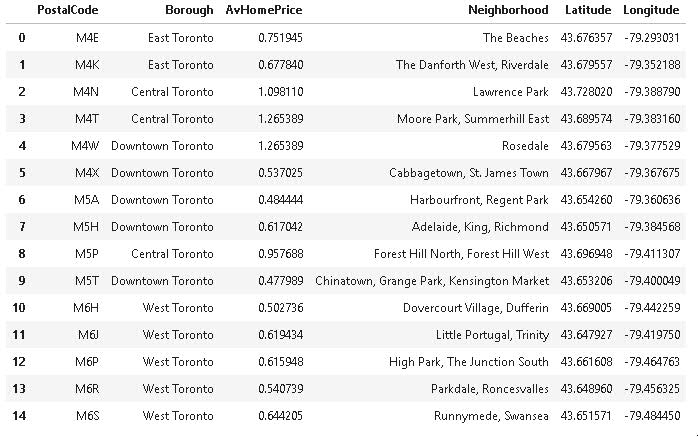
To analyse the acquired data, we would combine the average house price from “Get Wellbeing Toronto – Economics” dataframe with the Neighbourhood postal code dataset to get house prices per postal codes. Then we would get the venues from food category using the “Foursquare” location data. We would cluster the combined data and would try to determine the best possible location for the new restaurant.

Starting from “Get Wellbeing Toronto – Economics” dataframe we added the average house prices column to postal code dataframe, matching the values to the respective neighbourhoods. We converted the average house prices to the units of millions. And removed the postal code rows that had no value for the average house price. This gave us the complete data for 54 postal codes.

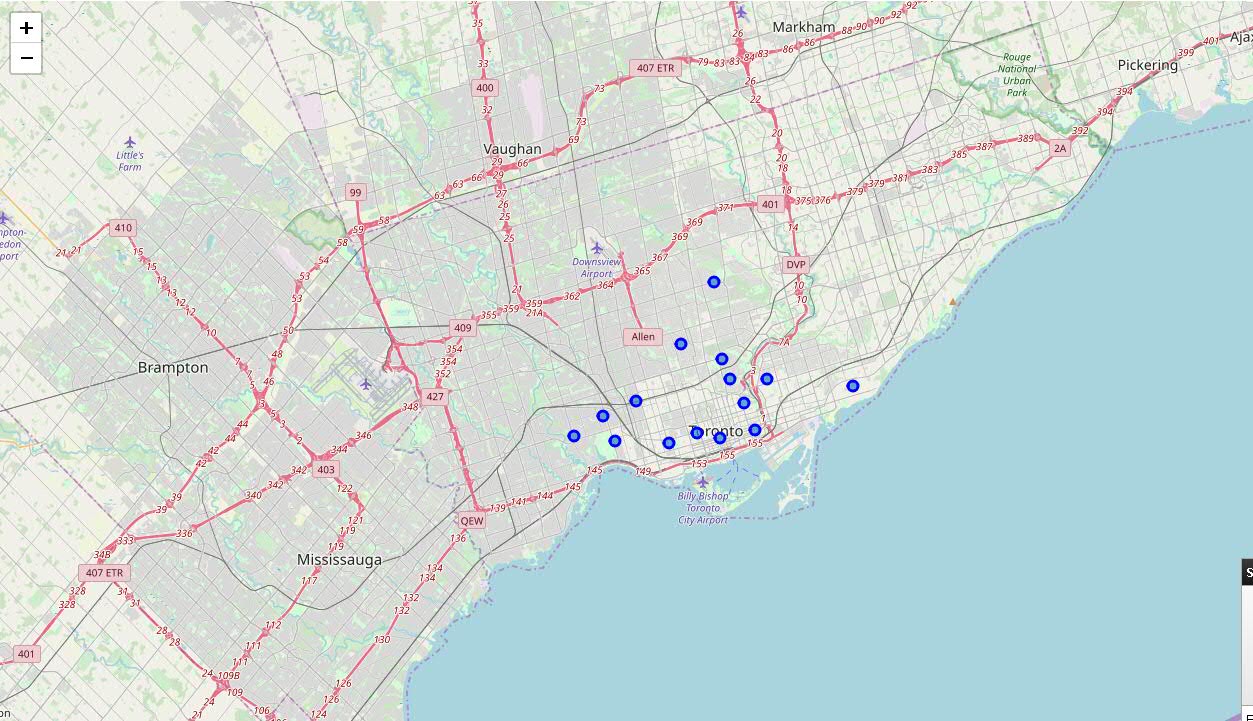


**Figure 1. Postal code dataframe with average home prices**

We importedGeospatial\_data file containing the coordinates by postal code from <http://cocl.us/Geospatial_data> We found the respective coordinates by the Postal code in Geospatial\_data dataframe and add it to Neighbourhood dataframe. To limit the dataframe to Toronto only postal codes we filter out the borough’s that contain “Toronto”. This gave as a 14 row dataframe.

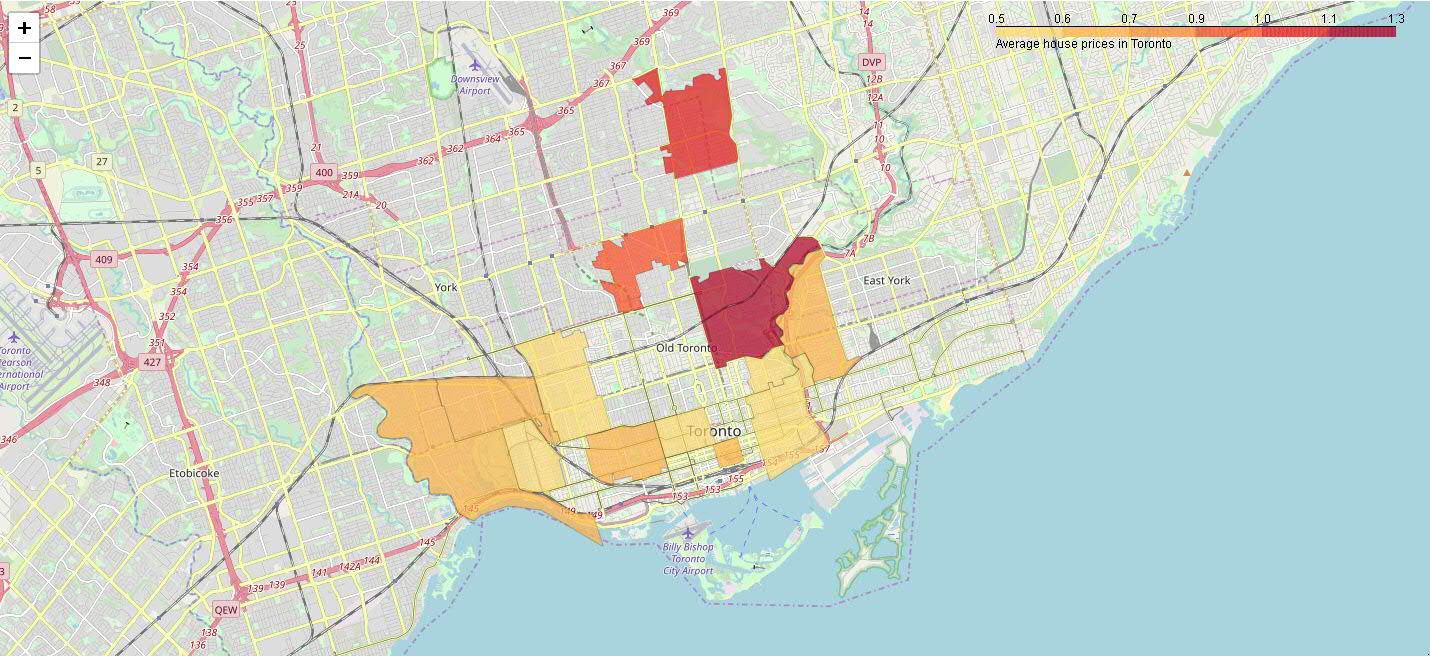
****

**Figure 2. Postal code dataframe with average home prices and geographical coordinates**

****

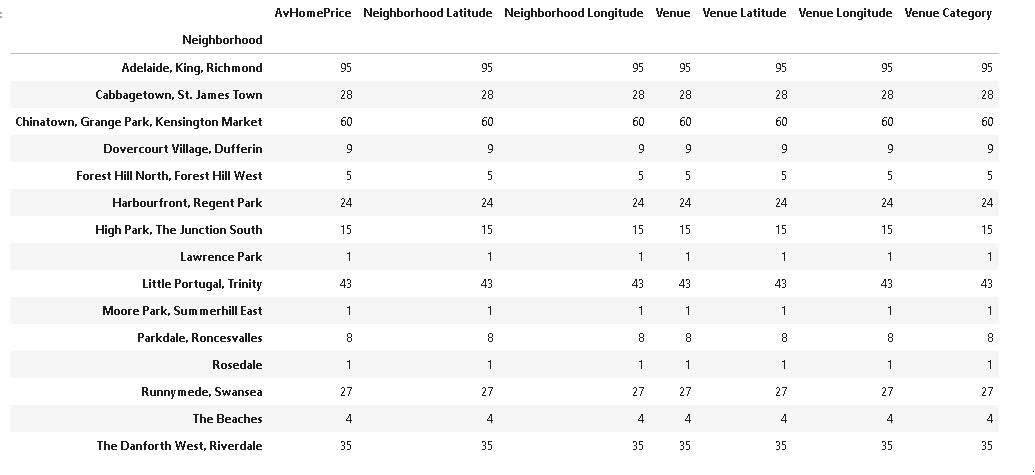
**Figure 3. Toronto map with Neighbourhood markers**

On Figure 4c is the choropleth map showing the Average house prices for Toronto. Unfortunately, we were unable to get the average price data for all Toronto neighbourhoods, but we even though we can see three areas where the average house price is above 1 million - around Rosedale, Summerhill East and Lawrence Park.

****

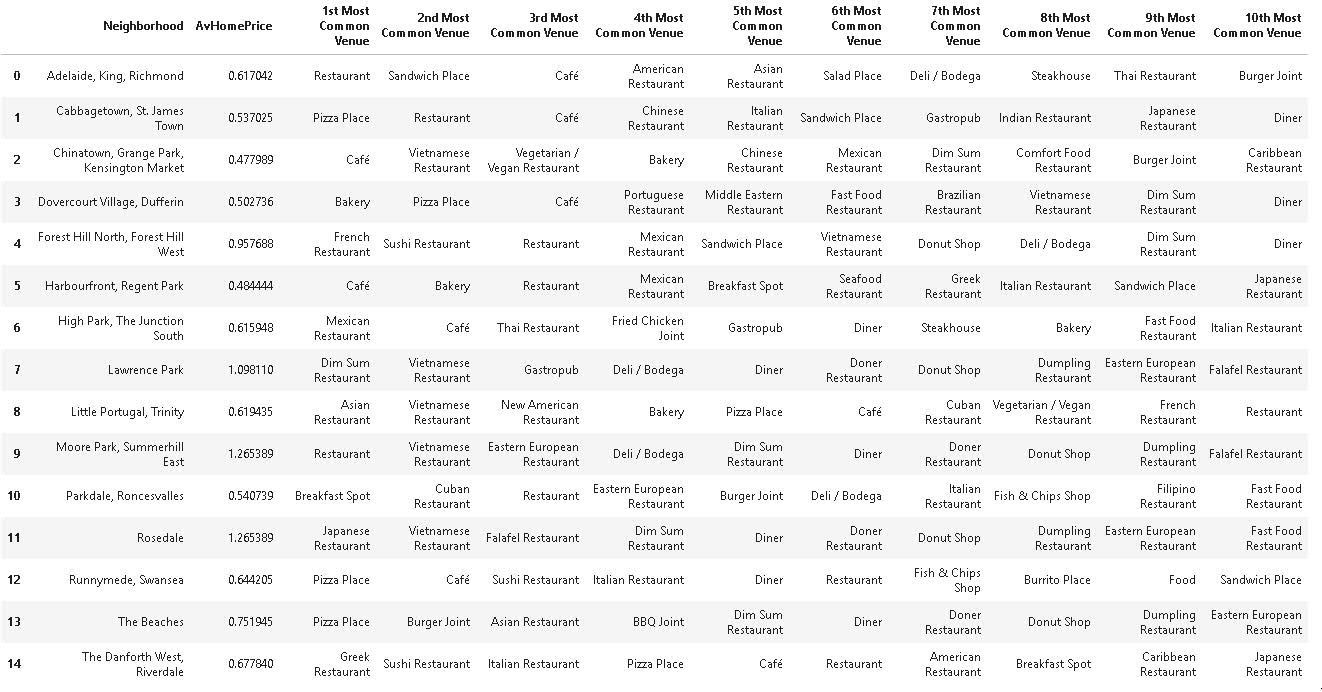
**Figure 4. Average house price in Toronto**

Using “Foursquare” API we got the get nearby venues from Food category for all Toronto neighbourhoods. This gave us all restaurants in the selected area. We created a new dataframe that had all the venues for each of the neighbourhoods. The resulting dataframe, had 73 unique categories.



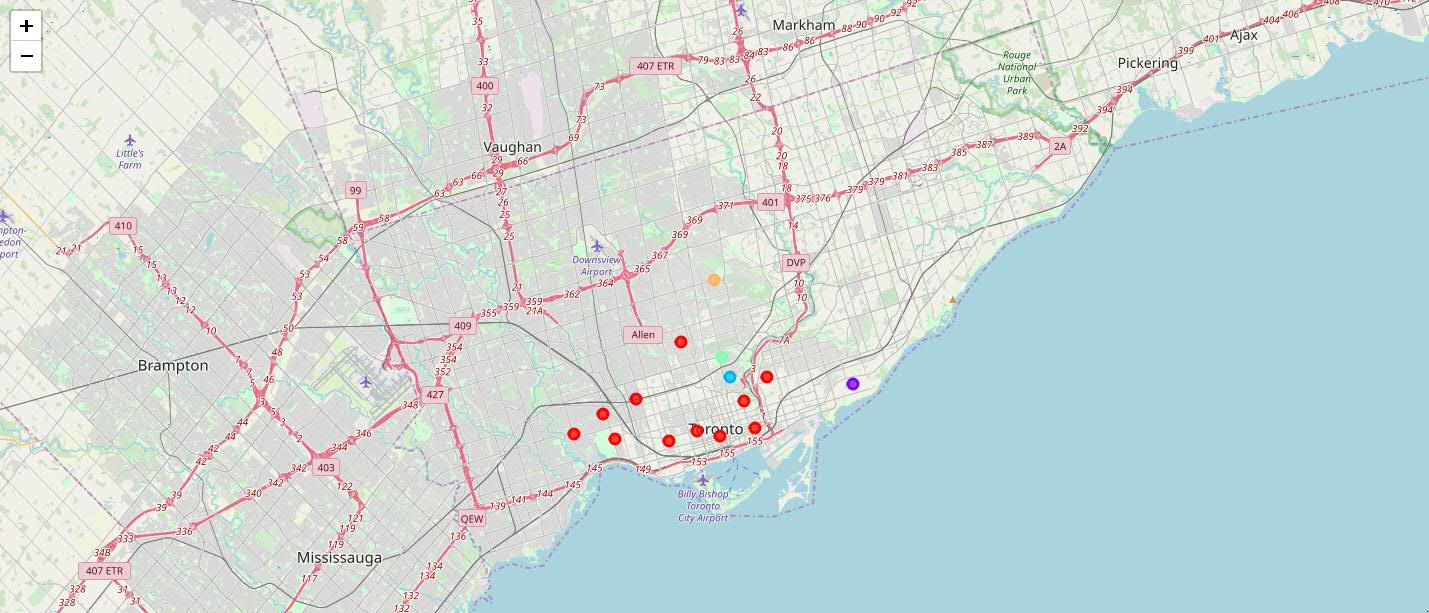
**Figure 5.  Number of venues in Food category were returned for each neighbourhood**

We grouped the rows by neighbourhood and by taking the mean of the frequency of occurrence of each category. Then we sorted the venues in descending order and create the new dataframe that contain the top 10 venues for each neighbourhood.

****

**Figure 6. Top 10 venues for each neighbourhood.**

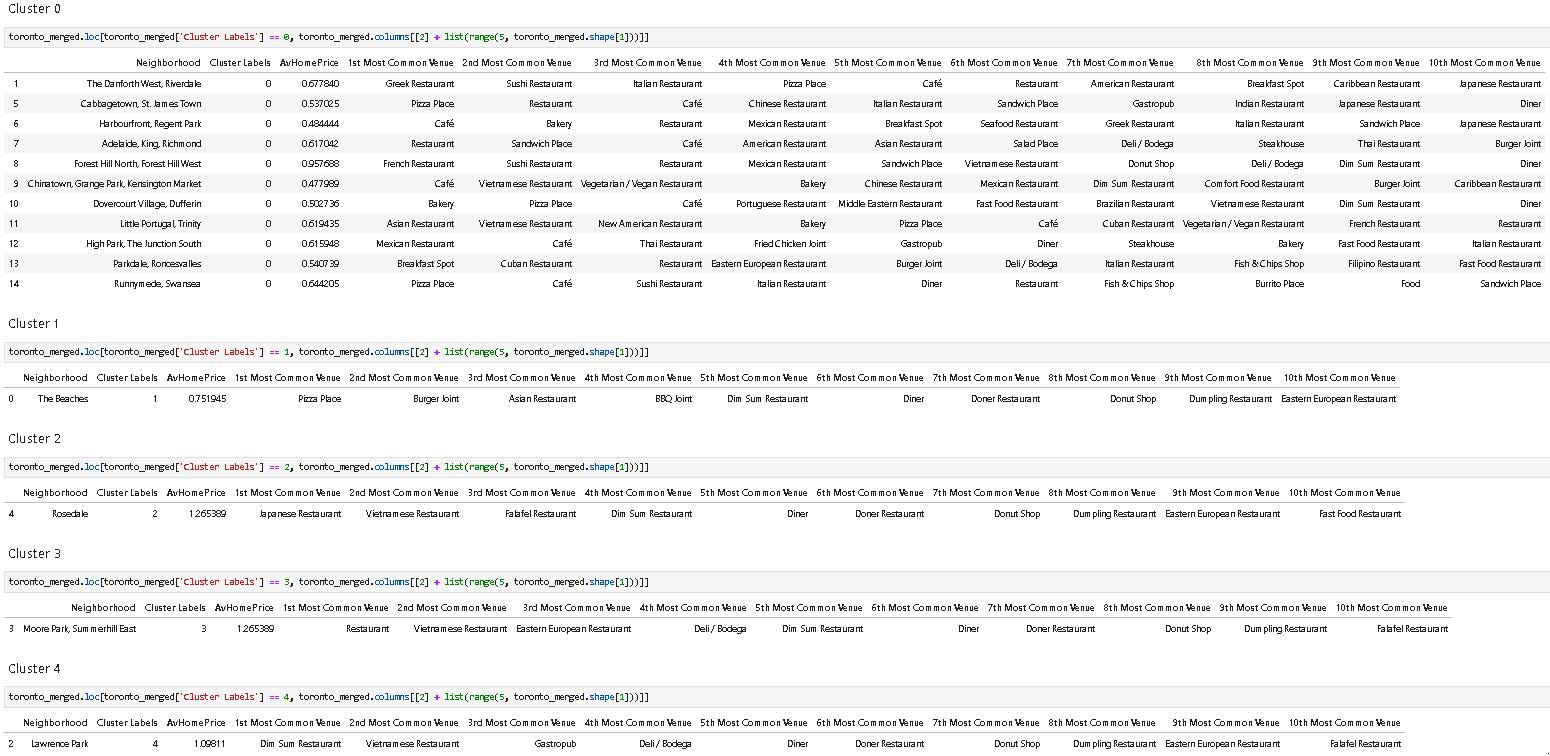
We cluster the Neighborhoods into 5 clusters by runing k-means to cluster the neighborhood into 10 clusters. We created a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.



**Figure 7. Map of resulting clusters.**

**4. Results**

It appears that most of the neighbourhoods are located in the 1st cluster. In the 1st cluster, the most common venues in the neighbourhoods are cafes and pizza places. Also, it appears that three clusters are around the neighbourhoods with the highest average house prices.

****

**Figure 8. Clusters.**

**5. Discussion**

Where should we open a new restaurant? Housing price maps show that the Lawrence Park cluster (4) neighbourhood might be a good candidate. This area looks like a quite densely populated area, so we expect the region to have a lot of foot and car traffic, so good visibility. This neighbourhood has also reasonable average house prices.

**6. Conclusion**

This is only a first-order solution to the question 'Where to open a new restaurant in Toronto?' Using public datasets, we were able to partially address one of the factors that we have mentioned at the beginning - average house prices. There certainly is lot of room for improvement. For example, we have to factor in crime rates, competition etc. Toronto Opendata website should have other datasets that we might use to further improve the results.