**Introduction/Business Problem**

Toronto is one of the largest and busiest cities in the world. The city is filled with various businesses that competes and flourishes with each other. The amount of opportunism the city provide is immense and due to the recent Covid-19 pandemic, many businesses have shut the door permanently. Although this is very unfortunate for these owners in this tough time, we will overcome the difficulties and that means new businesses will be injected into the city as soon as the economy opens up to revitalize itself.

Once the economy reopens many entrepreneurs will venture out to start their own businesses and open many businesses. The topic or problem that I'm trying to solve for is to provide recommendation for new business owners wanting to open up a business in Toronto based on foursquare data in conjunction with income and population data.

**Data Preparation:**

Scrub data from Wikipedia of Toronto borough and neighbourhood data

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Drop not assigned values:

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Using Geospatial data, determine the latitude and longitude of each neighbourhood

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Now let's incorporate Four Square API data,

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Define a function to push API calls to Foursquare to explore neighbourhoods in Toronto

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

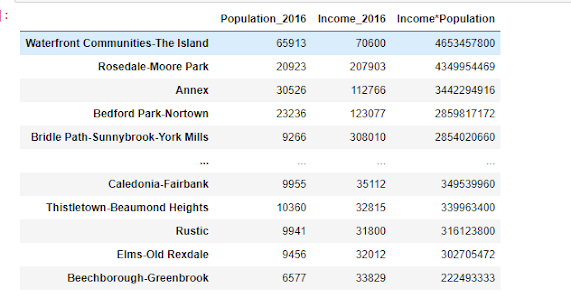
Create a new df combining four square API venue data and neighbourhood data

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Identify top 10 most common venues for each neighbourhood:

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Scrub data from Toronto website containing population and income:

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

**Methodology**

The idea is that due to a recovering economy, although there will be many potentials for new businesses due to shutdown of COVID-19 however the recovery process will be generally slow. The approach is to cluster the neighborhoods into 5 clusters and identify the most common venues and types of the neighbourhoods. This will provide a good suggestion to existing businesses and the type of businesses the population of that area prefers. Then the population and income data will be used so neighbourhoods with the highest potential such as high income and high population should be considered first.

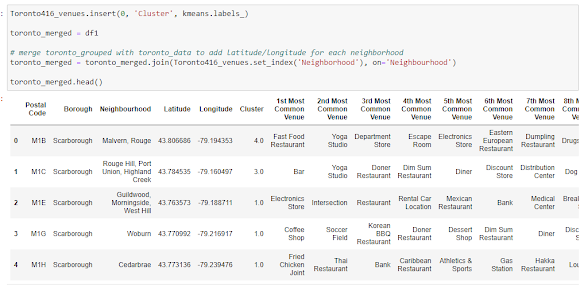
        Four square API will be used to retrieve relevant data, especially the relevant venue category for each neighbourhood. Then that data will be used with K-means to cluster and group them. K-means clustering is used as there is no existing label and we want to identify the relevant type of neighbourhood and thus can target the type of business to open

**Results**

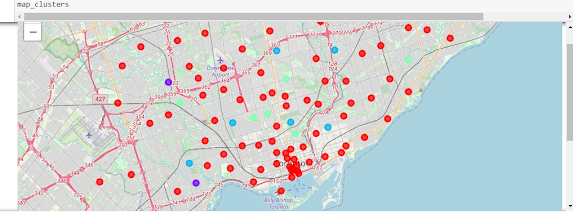
The data is to be separated in to 5 clusters:

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Create a new data frame containing the cluster labels for the original data frame of top 10 venues of each neighbourhood

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

Plotting the clusters:

[](https://www.blogger.com/blog/post/edit/7103840762630661075/5482888539754381128)

**Cluster 0 - Having parks as the most common venue, focused on outdoor recreation with restaurants and snack shops.**

This cluster seem to be largely inhabited by families that are not in a very population-dense area due to park being the most common venue. They are also very health conscious as represented by the abundance of yoga studios and other physical recreation venues. There is also quite a bit of Doner and Dim Sum restaurant which may point to the fact that there are a significant Turkish and Chinese folks living there. With this information, here are some suggestions for types of businesses to consider opening:

1. Group fitness gym
2. Recreational sports for kids (For example, kids basketball league)
3. Grocery/restaurants focused on Turkish/Chinese ethnic foods

**Cluster 1 - Shopping oriented with access to good transportation and restaurants. Lots of amenities**

This cluster contain neighbourhoods that focus on venues for everyday living. They are close to everything they need inters of food, everyday supplies and fast-food restaurants. The noticeable difference between the restaurants in Cluster 1 and Cluster 2 is that Cluster 2 contain fast food restaurants and appear to be less costly than restaurants in Cluster 1. The family income in Cluster 1 should be higher than Cluster 2. Opening business in Cluster 2 should focus on:

1. Instead of prestige items and services, the business should be cost-effective, no-frills, and the best bang for the buck
2. Fast food restaurants or takeout’s that are not overly expensive

**Cluster 2 - Seems to be focused for younger family with kids**

Only one neighbourhood in this category. While the most common venue is playground, it seems that young families live here that are also multi-cultural due to Downer and Dim Sum Restaurant. Opening a venue should focus on:

1. Services for young families such as day care, dog walk, restaurants with take-outs

**Cluster 3 - Focused towards young adults who are into busy city life** Should focus on services for young single adults who wants a busy life such as:

1. Bars and entertainment
2. Fitness class, kickboxing class, etc.

**Cluster 4** - For younger generations and a blend between young families and single adults:

1. Escape rooms, entertainment services
2. Virtual gaming centers

**Based on Toronto income/population data, these are some of the top income\*population neighbourhoods and thus having the best potential:**

**Rosedale having a population\*income of 4,349,954,469:**

This is clusters 0 which belongs to a high-income neighbourhood who enjoys outdoor recreation. It would be beneficial to consider opening a high-end group fitness gym here, or a nice restaurant allowing takeout. Other considerations including extra curriculum activities for after work and for kids such as hockey league.

**Harbourfront， Waterfront communities/Island having income\*population of 4,653,457,800:**

This belongs to Cluster 1 which is shopping oriented with access to good transportation and restaurants and lots of amenities. Would recommend opening:

1. Hybrid Convenience/Grocery store
2. Fast food or restaurant that is not overly expensive

**Bedford park having income\*population of 2,859,817,172 and Annex/Yorkville region having income\*population of 3,442,294,916**: Same recommendation as Rosedale

**Discussion**

Based on my analysis, it is desirable to focus on either Rosedale, Annex, Bedford Park or Harbourfront as the top 4 desirable neighbourhoods to consider building a new business once COVID-19 pandemic is settled. All of the neighbourhoods belong to Cluster 1 except Rosedale (belonging to Cluster 0).

    1. The observation for Cluster 0 (Rosedale) is that since it is in a relatively high-income area (even though population is lower), the households living there are more established and they enjoy celebrating nature due to close proximity of parks. If any business that will be opened there it will need to be reflective of the neighbourhood's state of mind. They likely will have less concerns for cost and more attention will be focused on health and wellness. Thus, opening businesses that have the business aptitude to help residence towards these mentioned qualities will have higher chances of succeeding.

    2. The observation for Cluster 1 (Waterfront, Annex, Bedford) is that even though the income\*population is high and thus giving higher potentials, the income is considerably lower than Cluster 0 with a much higher population. With this group of neighbourhoods, it is important to know that people will be looking for best value, they may not be always focusing on services or goods that are outside of their comfort spending zone. They need services that can help them with the best bang of buck and thus opening businesses towards that will have higher chances of succeeding.

**Conclusion**

This project has utilized various sources of external data and it was effective in determining what businesses are preferable for neighbourhood. Without using this, it would be difficult to physically explore and understand the neighbourhood in details. Having the data available have really helped with determining the flavour of businesses existing and thus the neighbourhood preferences. Although I have used Income\*Population as a standard to determine the high potential neighbourhoods, it is also important to point out that High-income/low-population have very different expectations of businesses to open as compared to lower-income/high-population groups even though the product of the two seem alike. Thus, having the data of the clusters and venues from four square API was very useful in identifying our desired outcomes.