**Ideal Locations for Restaurant Setup in Toronto**

Amit Pillai

July 17th, 2020

**1. Introduction**

**Business Problem**

* Toronto, which is located on Lake Ontario, is the most populous city in Canada and the provincial capital of Ontario. People would be interested in setting up a business here in Toronto. One of the most important criteria for starting a business is location.
* Through analysis of data, I wish to narrow down the location search for people wishing to start a Multi-cuisine Restaurant here in Toronto.
* The result of the analysis would lead to a list of suggested neighbourhoods for his/her business (with importance given to more population, less crime, less business competition)

**2. Data Section**

**For this project we will be using**

* Venues data from Foursquare API.
* The Toronto neighbourhood crime dataset from Kaggle (Crime data - 2019, population - 2016 census): https://www.kaggle.com/alincijov/toronto-crime-rate-per-neighbourhood/data#

**How these data will help**

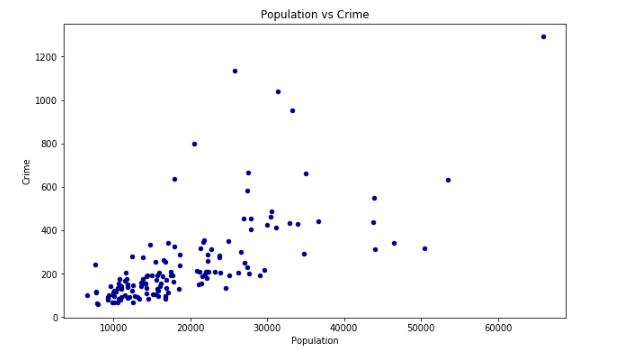
* Venues data from Foursquare.com will help us get popular business categories of a neighbourhood.
* Crime data will help us locate a neighbourhood with comparatively less crime rate.

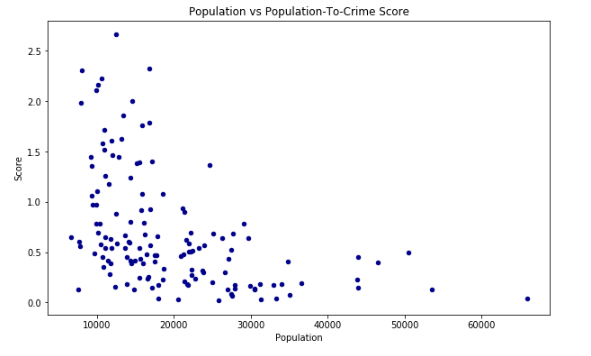
3. Methodology

Procedure

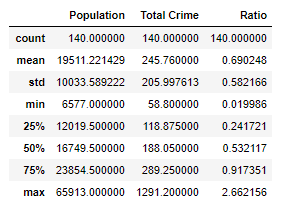
* Finding neighbourhoods with High population and Low crime rate.
* Finding number of restaurants in each neighbourhood.
* Comparing and choosing neighbourhoods

Starting with the Toronto crime dataset

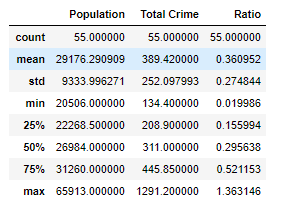
* Creating column which represents Total Crime in a neighbourhood.
* Scatter Plot of Population vs Total Crime - shows that it is not a perfect correlation between Population and Crime. So that means there could be values with High Population and Low Crime, which is what we are looking for.
* So, we need a value which can be used to find Neighbourhoods with low crime compared to its population.
* Creating column which represents the Ratio of Population/Total Crime\*2(higher the ratio, safer the neighbourhood in terms of crime).



* We see that majority of neighbourhood with high ratios are in the 10,000 – 20000 range.
* But population is a big factor too, for a business like a Restaurant, so we take the mean of all population, and consider only neighbourhoods above that mean.

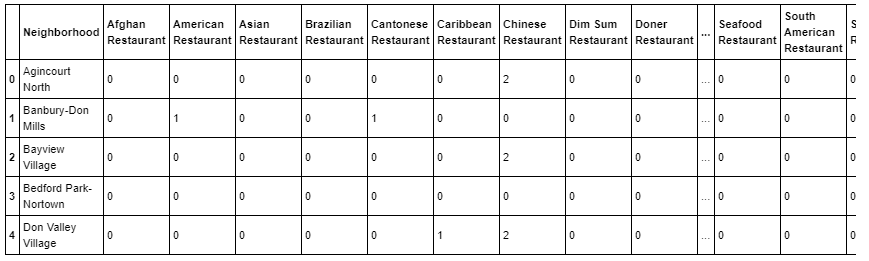


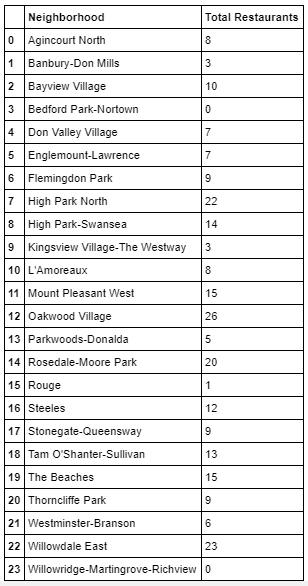
* Filter data to only > 20,000, then we look at mean ratio of this new set of data.



* Now, filtering the data with ratio > 0.4 (mean–0.36) will filter in better neighbourhoods.

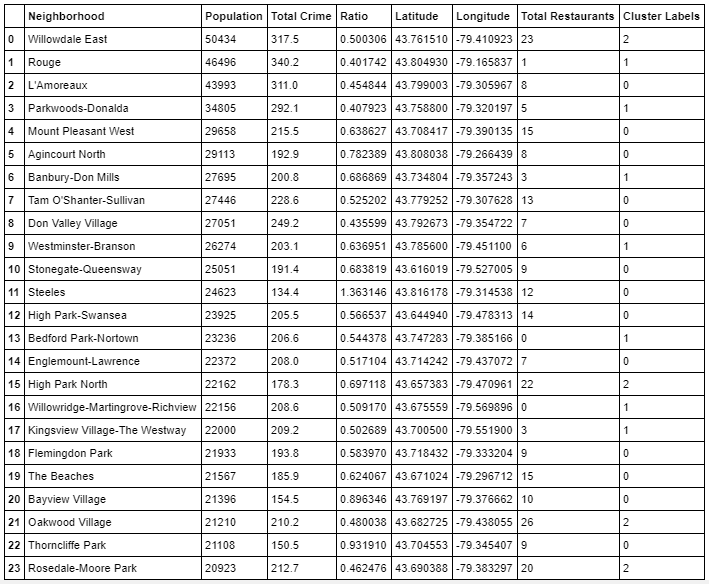
Finding Venues

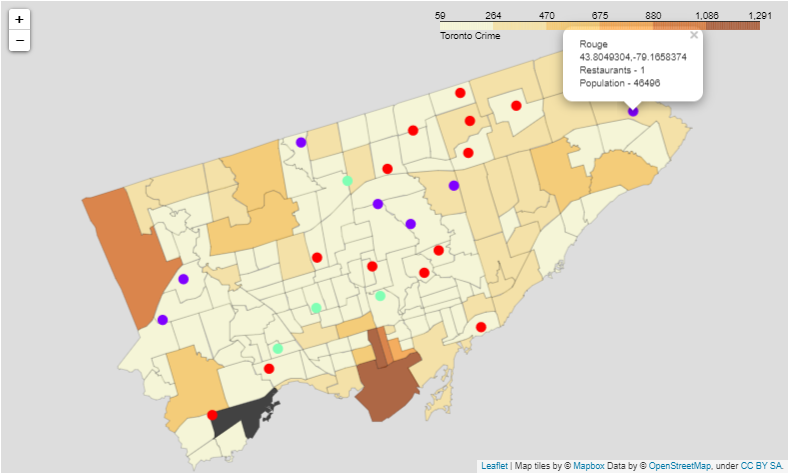
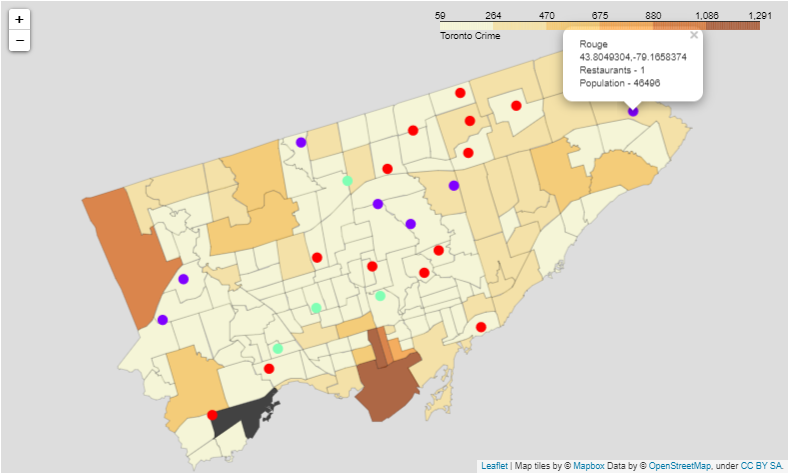
* Using Geocoder, we find the respective latitude and longitude of the neighbourhoods.
* Using foursquare API, we find the venues within 1km radius of the neighbourhoods and convert the json response to a Data frame.
* Using one-hot encoding on venue categories we get the numerical representation of number of venues per neighbourhood.
* Then the Data frame is grouped by neighbourhood and total sum of each venue per neighbourhood is created. From this, Restaurants are filtered out.
* Restaurant counts are summed by column. We now have our total restaurant per neighbourhood Data frame.



* Merging the above table with Crime table on the common column Neighbourhood, will lead us to the final Data frame.
* This ‘Total Restaurant’ column is subjected to K-means algorithm with k=3, for Low, Medium, and High restaurant count.

**Final Dataframe**

****

******Crime Density Map with Neighbourhoods**

Purple = Low Restaurant Count Neighbourhoods (Label 1)

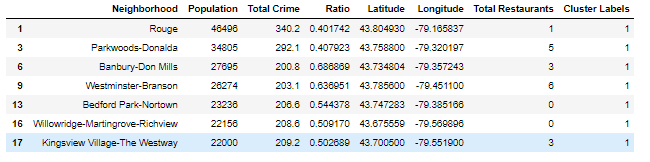
Red = Medium Restaurant Count Neighbourhoods (Label 0)

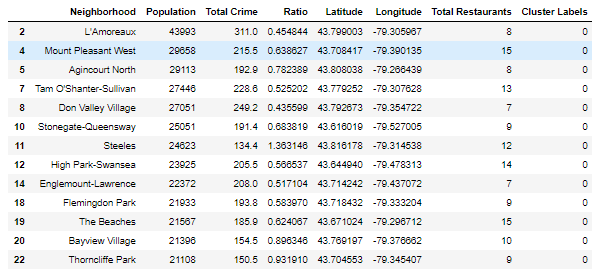
Cyan = High Restaurant Count Neighbourhoods (Label 2)

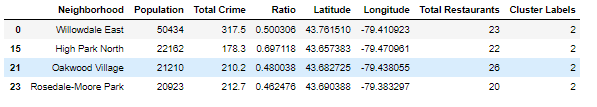
**4. Result**

* In the result Dataframe, we have 3 clusters of neighbourhoods with relatively less crime with respect to its population.
* In the 3 clusters, Cluster 1 (Low Restaurant Count) and Cluster 0 (Medium Restaurant Count) are recommend because these neighbourhoods are less in competition compared to Cluster 2 (High Restaurant Count).
* Cluster 2 could be considered for setting up restaurant chains for an already well established and renowned restaurant.
* These recommendations are based on the only 2 parameters I have considered (population and crime). In reality, there are much more factors to consider, before choosing a neighbourhood for setting up a restaurant.

Low Restaurant Count – Cluster 1



Medium Restaurant Count – Cluster 0

High Restaurant Count – Cluster 2

**5. Conclusion**

In this analysis, I was able to narrow down neighbourhoods with decent population and relatively low crime rate. Then from our results, we chose neighbourhoods with less restaurant count to avoid competition with the veterans. The output of the analysis is a list of neighbourhoods that could be considered for a new Restaurant Setup. (This is just considering Population and Crime rate, in real life more factors would come into play, but this analysis narrows down the search).