Report

Introduction

I would like to help recommend a location in Toronto to open a restaurant. I will leverage Foursquare data on popular restaurants in Toronto in order to visualize the competition. I will also incorporate data on population per square kilometre and average income to help the restaurant owner make the decision.

Data

The competitors' restaurants will be from Foursquare, and the population per square kilometre and average income will be from Statistics Canada Census data from 2016 (https://open.toronto.ca/dataset/neighbourhood-profiles/)

Methodology

First, I wanted to get data on population per square kilometre and average income for every neighbourhood. This is because I think opening a restaurant in an area of many people or high income would be a good idea.

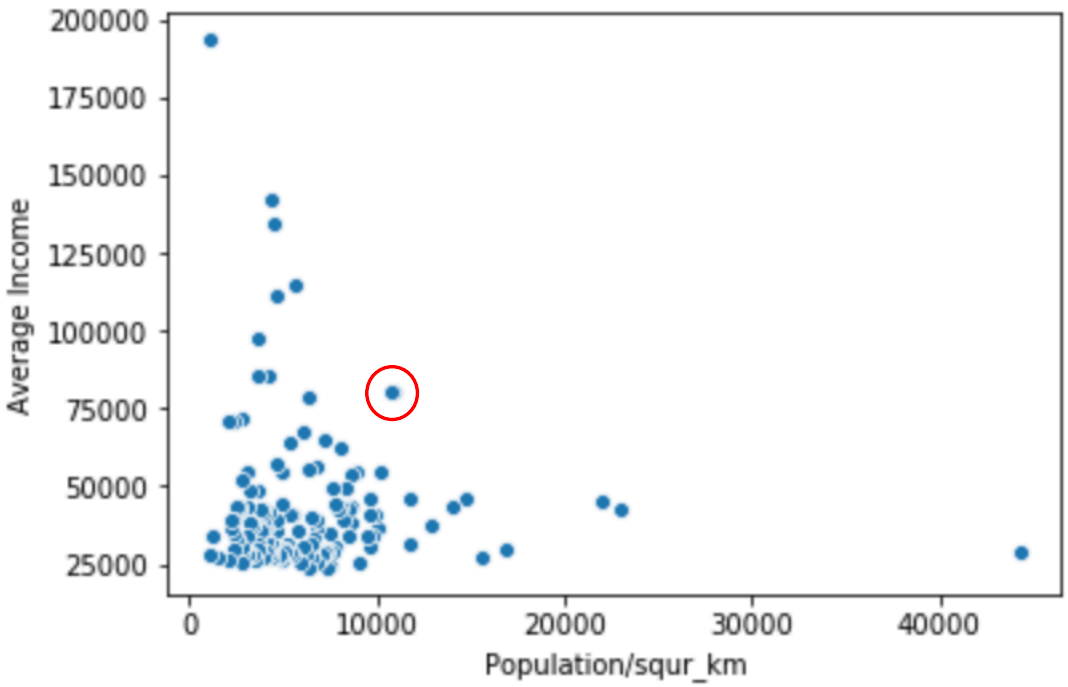
Here are the top 10 neighbourhoods for population density:

	Population/squr_km	Average Income
North St. James Town	44321.0	28541.0
Church-Yonge Corridor	23044.0	43039.0
Mount Pleasant West	21969.0	45441.0
Regent Park	16880.0	29525.0
Taylor-Massey	15528.0	27064.0
Moss Park	14753.0	45877.0
Bay Street Corridor	14097.0	43427.0
Little Portugal	12859.0	37924.0
Kensington-Chinatown	11806.0	31282.0
High Park North	11726.0	45893.0

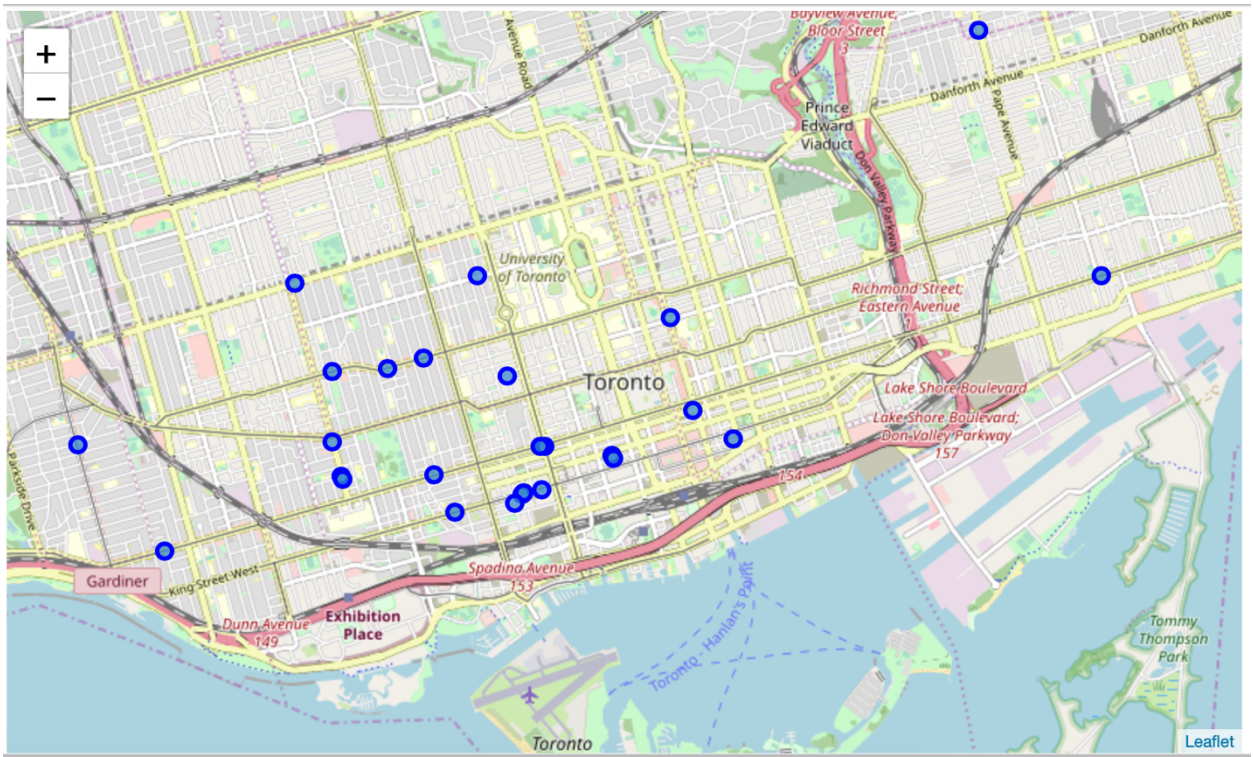
Here are the top 10 neighbourhoods for average income:

	Population/squr_km	Average Income
Bridle Path-Sunnybrook-York Mills	1040.0	193454.0
Forest Hill South	4380.0	142627.0
Rosedale-Moore Park	4500.0	134865.0
Casa Loma	5683.0	115033.0
Lawrence Park South	4685.0	111586.0
Kingsway South	3593.0	97836.0
Bedford Park-Nortown	4209.0	85678.0
Leaside-Bennington	3596.0	85496.0
Yonge-St.Clair	10708.0	80555.0
Annex	10863.0	80138.0

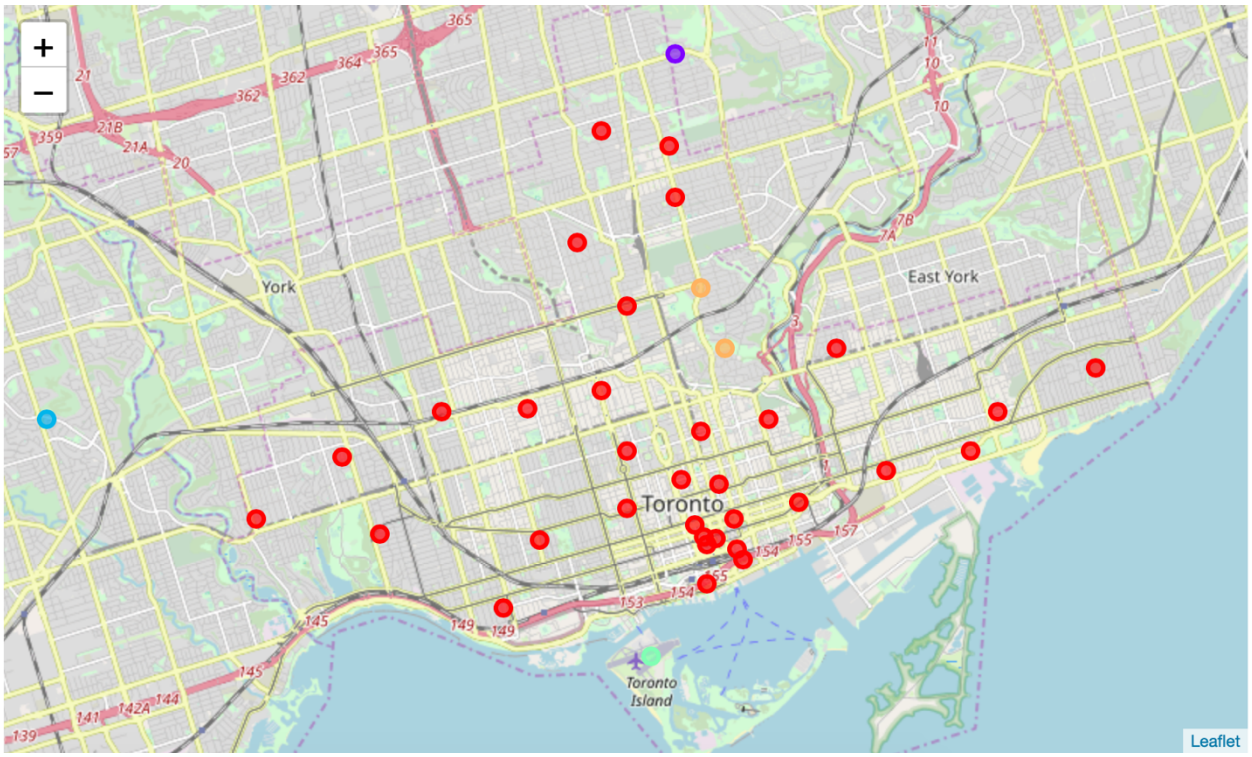
Next, I made a scatterplot to see the relationship between population density and income:



The next group of data I wanted to use was the top 50 restaurants in Toronto. This is useful because it wouldn't be a good idea to open a restaurant in a location that already has many top restaurants. Here are their locations:



I wanted to go even further in depth into the other restaurants in Toronto, so I did cluster analysis based on the types of existing restaurants and their location. I used FourSquare to pull the top 100 restaurants in each Borough of Toronto (some containing multiple neighbourhoods). I then used one hot encoding to transform the data into a matrix containing each Neighbourhood and type of restaurant. The machine learning method of kmeans was used to segment each of the restaurants into one of the 5 clusters. Here is a map of the clusters:



Results

The results of the first inquiry weren't apparent at first, but upon further inspection there were 2 neighbourhoods that stood out, Younge-St. Clair and Annex. I drew a circle around them on the scatterplot. They stand out because they have respectively high population densities of around 11,000 people/km^2 and average incomes of around 80,000.

The results of the top 50 restaurants is less conclusive because they seem to be mostly spread out, except the few in the downtown core.

For the cluster analysis the results were quite interesting. As we can see from the map above, the majority of Neighbourhoods fall within the red cluster. When I looked closer at the most popular types of restaurants within each cluster, I noticed that the red cluster is mostly cafés, pizza and American food. The other 4 clusters seem to favour foreign food types more.

Discussion

One observation I noted in the first analysis of population density and income is that it seems to follow a 1/x distribution. This means that very poor neighbourhoods have high population density and very rich neighbourhoods have low population density. This phenomenon is not especially good for restaurant owners, but the two outliers mentioned above (Younge-St. Clair and Annex) would be my suggestion of a good neighbourhood.

Another observation is that apart from distinguishing national from foreign food preferences, the cluster analysis seems too concentrated in the red cluster. In future trials of this experiment, maybe a different number of clusters or different criteria would prove beneficial.

Conclusion

In this report I used various techniques to help a potential restaurant owner know the best neighbourhood to open a restaurant. I identified that based on population density and average income that Younge-St. Clair and Annex have the best combination and those are the neighbourhoods I would recommend. Based on machine learning cluster analysis I found that in central locations there is a preference for national foods and in surrounding areas the preference is foreign foods. This should help a restaurant owner plan accordingly.