**Predicting best neighborhoods to open a restaurant in Toronto**

Yakov Gazman

Nov 17 2020

1. Introduction
   1. Background

Toronto Ontario is one of the most diverse and populous cities in Canada. It a vibrant city with communities from all over the world. As divers the city population is the restaurant industry is even more colorful and exciting. Cooking styles can range from simple home made to any country traditional cooking style in addition to infinite fusion system that combine two or more styles. As the city has a large population and it is a great vacation destination as well the restaurant industry is a very lucrative industry to be in

* 1. Problem

Toronto has a huge number of establishment and new restaurant popping up everyday. As we know of the restaurant industry has a very high closer rate and this issue is much more critical and impactful in a large and competitive city as Toronto.

* 1. Interest

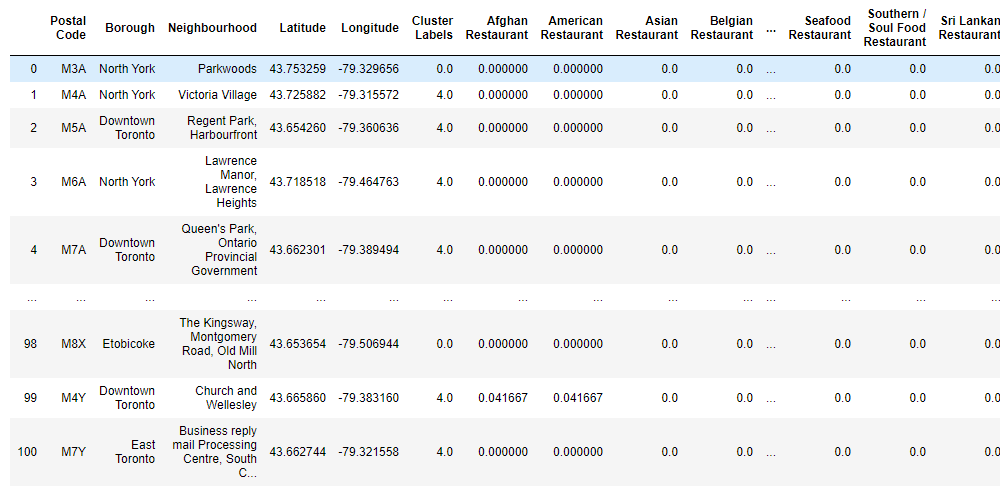
We are proposing an analytical tool that would be able to extract and analyst restaurant location data for the city of Toronto. After the data build, we will able to cluster and identify the neighborhoods were the competition would be the lowest and which cooking style will be preferred in the area. This tool would be a great addition in building a business case for new and experienced entrepreneurs in this industry

1. Data acquisition and cleaning
   1. Data sources

The main data source will be the four squares API and the Wikipedia page for the neighborhoods in Toronto.

* 1. Data cleaning

Data preparation and cleaning will find all the neighborhoods in Toronto and the associated coordinates for each area. The following step will be to connect the venues data from four squares with the neighborhood data. When data is assembled, we will filter and change columns, create a matrix to 1 and 0 from the different restaurant types and if they are present in the neighborhood

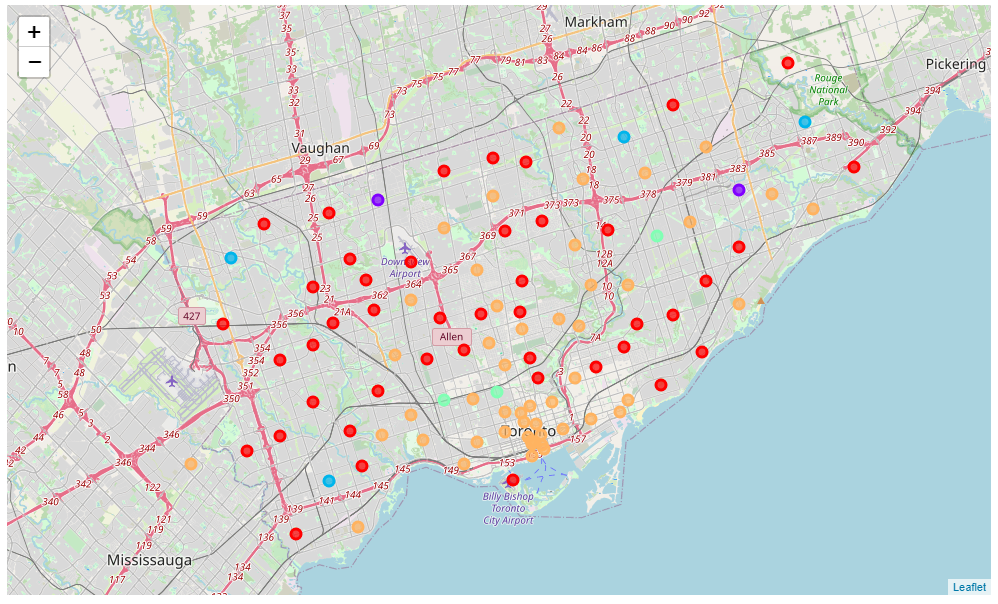


1. Methodology

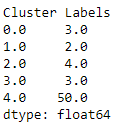
The methodology will include normalization of the restraint frequency depending how often the specific type of establishment appears in the area. We will use k mean clustering method to cluster the different neighborhoods according how many restaurants they have. After clustering we will aggregate the data to find the best clusters where completions in the restaurant business is the lowest and we would assume has the best to potential to open a successful business. The project will go future by analysing which types of restaurant will work best in each specific area

1. Results

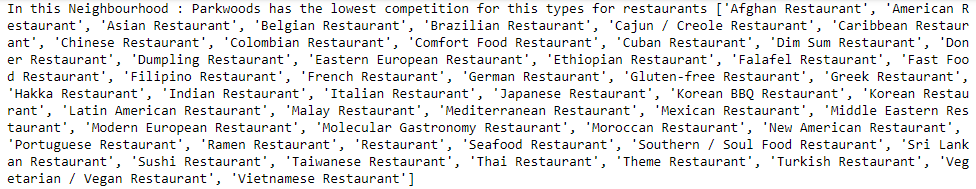
The map below is all Toronto neighborhoods represented in in clusters, each clusters appears in a different color.



The table below shows the overall density for each area with cluster 0- red, 1-purpule, 2-blue, 3-green, orange-4



The list below is an example of the output expected when selecting a specific restaurant type, for each neighborhood we are able to show which area and restaurant types:



1. Discussion

After performing our analysis, we discover that the orange neighborhoods have the highest density of restaurant in the area and are not recommended to new business due to saturation in the market. In addition all other areas don’t have high density in most establishment type and will present a good area to start a new restaurant. This project assumes that population density and economical ability in the city is normally distributed and does not affect the over all results. We are solely looking at restaurant density in the area to draw this result

1. Conclusions

Per the project presented, out model was able to successfully predict the area in Toronto city that are less dense in restaurant location and provide a good overall look to potential restaurant owner to where they should locate their business. In addition to identifying the areas the project was also able to give the owners a comprehensive list of potential establishment types that are in low density in the selected area

1. Future directions

For future project enhancements we would pull additional data and normalize the results to neighborhoods population and income data to give another edge and a full understanding to the potential business owners.