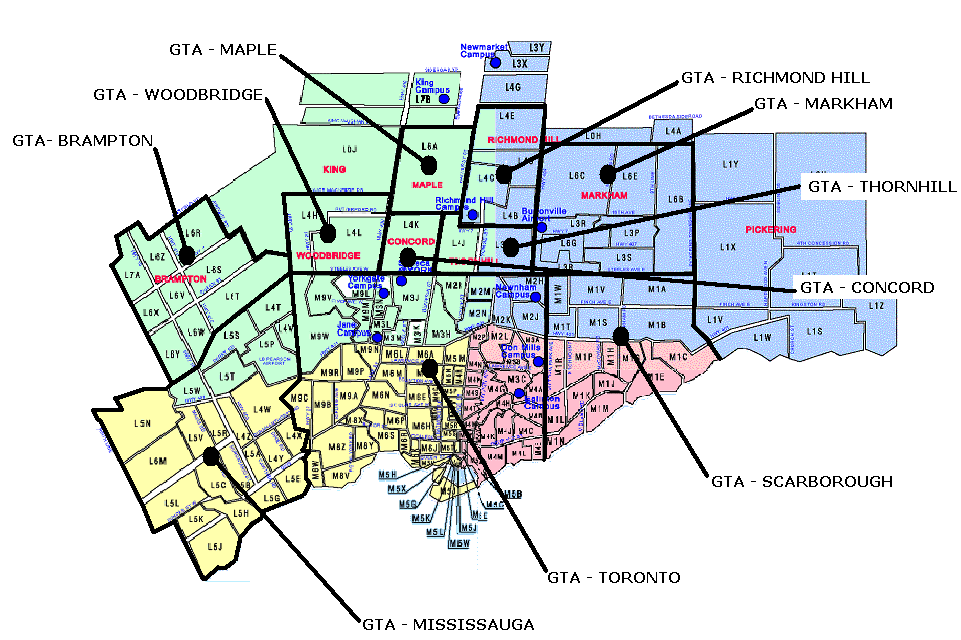
**Capstone-Project---The-Battle-of-Neighborhoods**

Bradley Berg

[NOTE! There are 2 ipynb files of coding for this project:

bwb Capstone Project Battle of the Neighborhoods week 5 pizza.ipnyb

bwb Capstone Project Battle of the Neighborhoods week 5 fastfood.ipnyb ]



**1A) Background Discussion**

Toronto is the capital city of the Canadian province of Ontario. It is the most populous city in Canada. Toronto is an international center for business and finance. It's the financial and industrial center with a high concentration of banks and brokerage firms in the Financial District.   
  
From wikipedia:

*“In 2016 (last census), Toronto's city proper had a population of 2,731,571; the urban area had a population of 5,429,524; the census metropolitan area had a population of 5,928,040; and the Greater Toronto Area metropolitan area had a population of 6,417,516.   
  
The city's foreign-born persons made up 47% of the population. No single nationality or culture dominates Toronto's immigrant population, placing it among the most diverse cities in the world. In 2016, the three most commonly reported ethnic origins overall were Chinese (12.5%), English (12.3%) and Canadian (12.0%).  
  
Toronto historically has a low crime rate making it one of the safest major cities in North America. In 2007, the homicide rate for Toronto was 3.3 per 100,000 people, compared with Atlanta (19.7), Boston (10.3), Los Angeles (10.0), New York City (6.3), Vancouver (3.1), and Montreal (2.6). Toronto's robbery rate also ranks low, with 207.1 robberies per 100,000 people, compared with Los Angeles (348.5), Vancouver (266.2), New York City (265.9), and Montreal (235.3).”*

**>>>>This demographic background makes Toronto an attractive destination for potential business investment.**  
  
**1B) Business Problem to be Analyzed**  
**>>>> This analysis of Toronto is from the perspective of a potential investor who is interested in opening one or more restaurants.**   
  
In particular, this investor is looking at investing and opening fast food restaurants and/or pizza restaurants. In this case 'fast food restaurants' refer to the well-known national brands (eg Arby's, Mcdonalds, Burger King, Subway, Wendy's, etc.) These franchises have international brand recognition as well as established business plans, logistic support, and pricing.

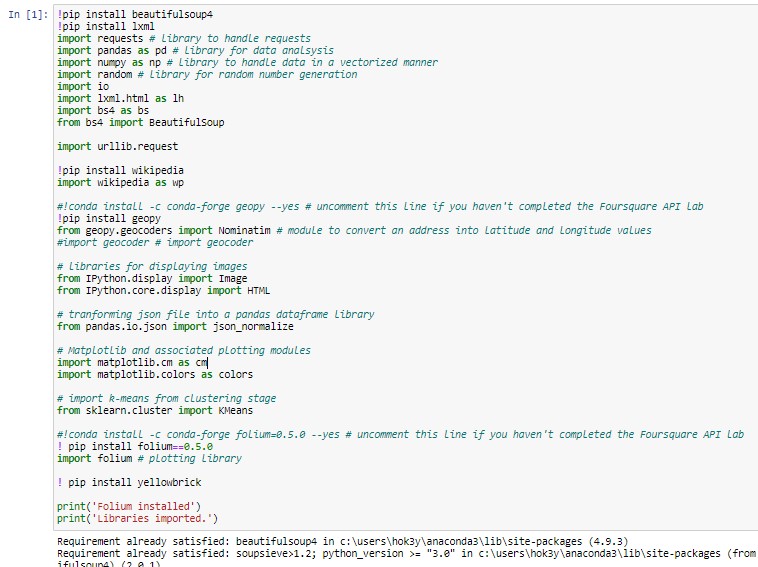
Pizza restaurants are a special case of 'fast food' and require a different level (perhaps lower) of investment and logistics than the national brands (although franchises such as Papa Johns, Pizza Hut, and Dominos could be investment options).  
  
Determining the potential location of a restaurant is one of the most important factors affecting its future success or a failure. So our project will attempt to answer the question: “Where should the investor open a fast food franchise and/or a pizza place?”.   
**1C) Problem Statement**

**What are the best locations for fast food franchises and/or pizza places in Toronto (specifically postal codes starting with 'M')?**

**2A) Data Collection**

In order to answer the above questions, data is collected on Toronto neighborhoods/boroughs including: postal codes, neighborhood name, latitude, longitude, population, population density, income, number of fast food restaurants, and number of pizza restaurants.   
  
All of this data is compiled and sorted by postal code.  
  
Data has been collected from the FourSquare API utilized via the Request library in Python.   
Additional data has been collected from cybo.com, wikipedia.com, yellowpages.ca, and www150.statcan.gc.ca.

i) First, the necessary Python libraries are imported:

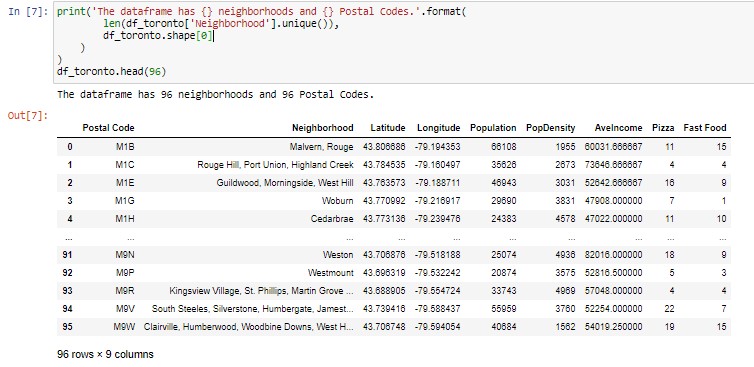


ii) Scrape the Toronto postal code, borough, and neighborhood data from wikipedia. Then clean the dataframe (The result basically removes certain postal codes. For example, M7A is the Legislative Assembly of Ontario and M7Y is the Letter Processing center.)

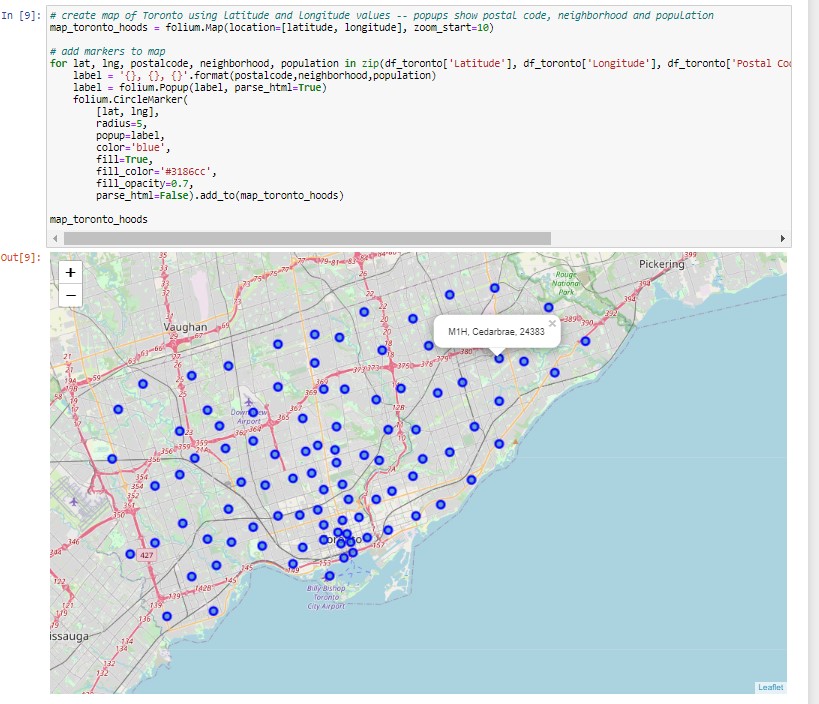
iii) The latitude and longitude of each postal code are read from 'C:\Geospatial\_Coordinates (1).csv'.

Data collected manually from cybo.com, wikipedia.com, yellowpages.ca, and www150.statcan.gc.ca were read from '[c:/toronto\_pop\_inc\_postalcode\_pizza\_ff2.csv](file:///c:/toronto_pop_inc_postalcode_pizza_ff2.csv)'.

The resulting dataframe:



iv) A map of Toronto was created. Each popup shows postal code, neighborhood name, and population.



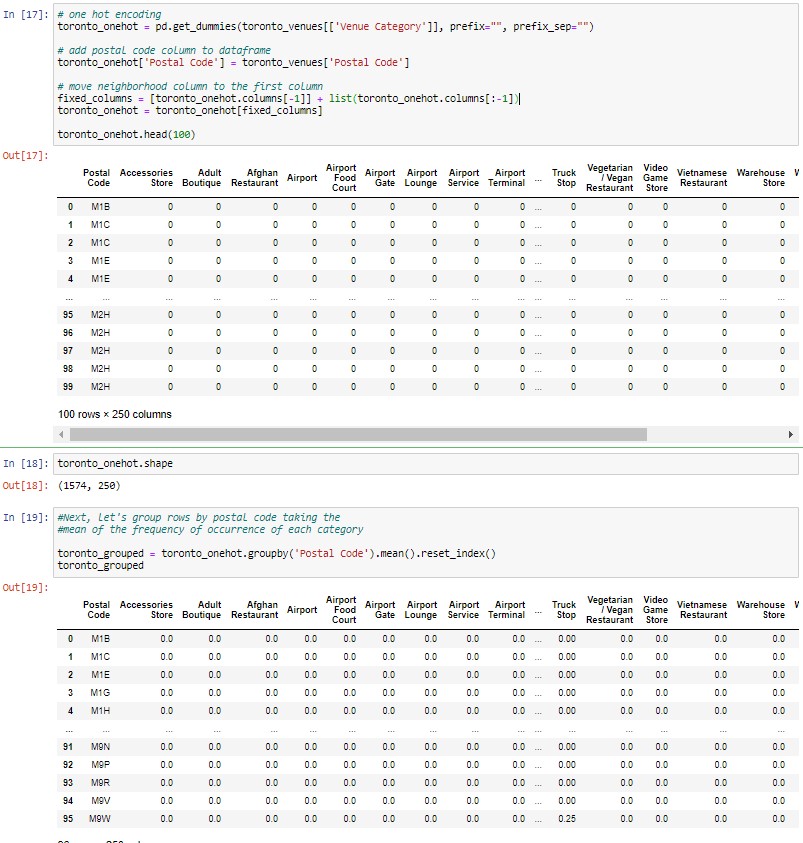
v) Another dataframe was created using FourSquare data that included the venues and venue categories below:



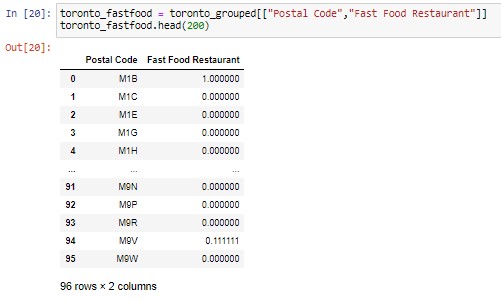
**2B) Methodology**

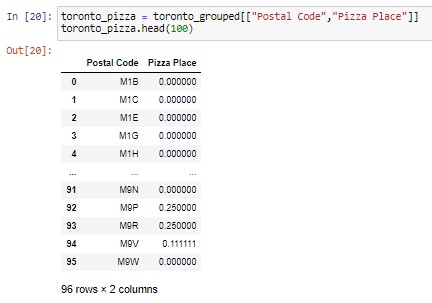
i) To help potential investors decide which locations might be best for opening a fast food restaurant and/or a pizza place I will be clustering similar neighborhoods using K-means clustering. This is a type of unsupervised machine learning algorithm that clusters groups of data based on a predefined cluster size.

ii) First, the data is prepared using one-hot encoding. One-hot encoding is used when the variables being used are ‘categorical’ as is the case here:



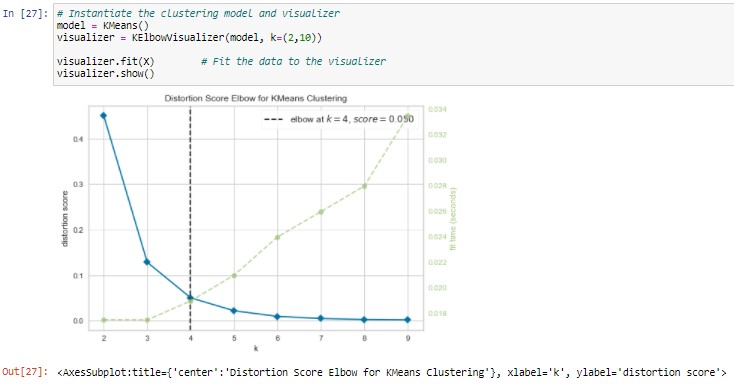
iii) Then Fast Food Restaurants and Pizza Places are grouped by Postal Code:



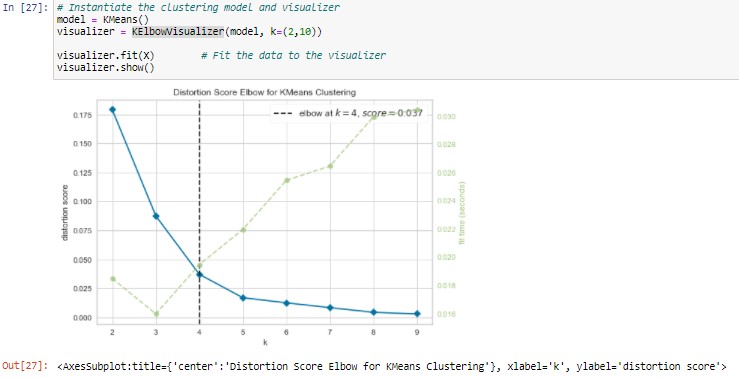


iv) Next, the number of clusters needs to be determined for Fast Food and Pizza Places. Here, I’m using the KElbowVisualizer from yellowbrick.

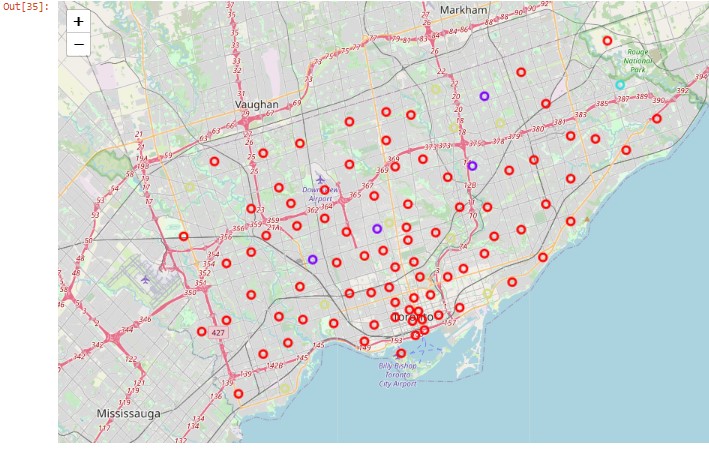
For Fast Food k = 4:



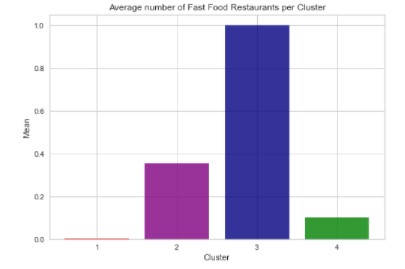
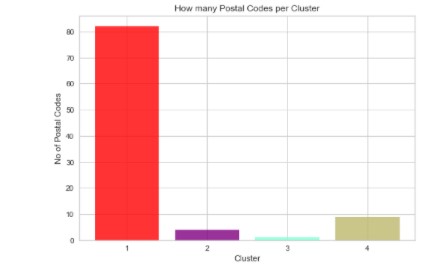
For Pizza k = 4:



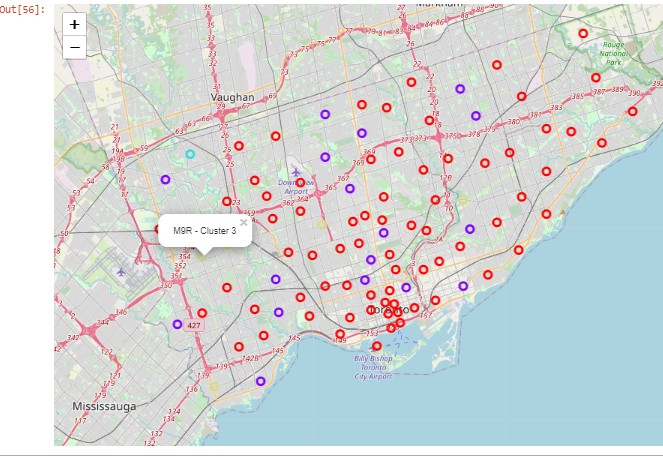
v) Map of the 4 clusters for Fast Food:



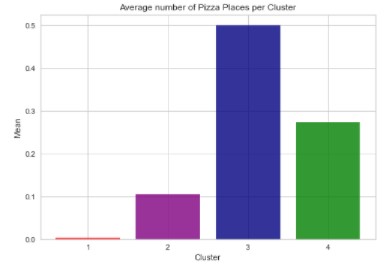
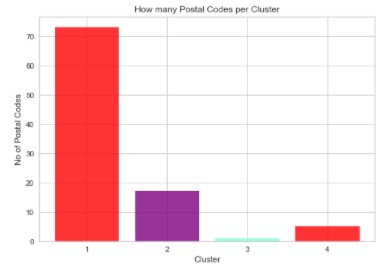
Fast Food bar charts (according to the FourSquare data):



vi) Map of the 4 clusters for Pizza Places:



Pizza Place bar charts (according to the FourSquare data):



vii) Now it should be noted that the data from '[c:/toronto\_pop\_inc\_postalcode\_pizza\_ff2.csv](file:///c:/toronto_pop_inc_postalcode_pizza_ff2.csv)' which was manually collected from cybo.com, wikipedia.com, yellowpages.ca, and www150.statcan.gc.ca DOES NOT show the same number of fast food or pizza place venues seen in the FourSquare data.

So my approach here is to analyze the postal codes sorted by 1) Cluster Labels and then

2) by Population. I also added two columns to the dataframes: PizzaPer1000 and FastfoodPer1000.

These are the number of Pizza Places per 1000 residents and Fast Food Restaurants per 1000 residents, according to cybo.com

Again, I will look at the postal code Cluster Labels according to FourSquare data **as well as** the postal code Population data in relation to the cybo.com data.

Income is ignored here because of the lower price points in the international fast food franchises as well as the relatively low price of getting a slice or pie.

**CONCLUSION**

**To Recap:**

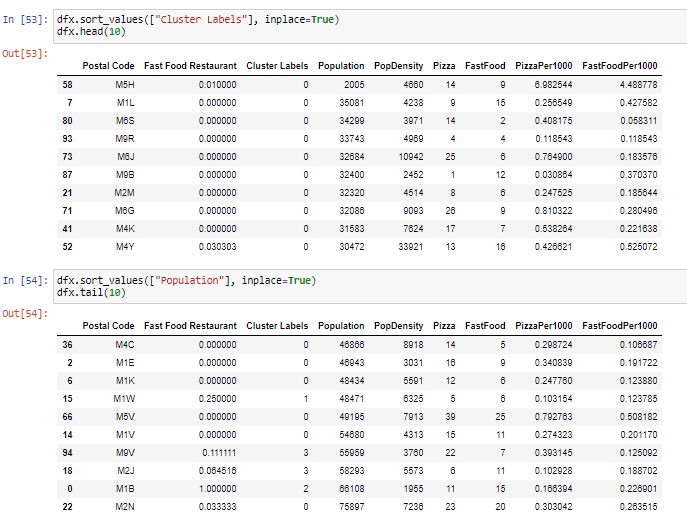
**This analysis of Toronto is from the perspective of a potential investor who is interested in opening one or more restaurants.**

**Question to be answered:**

**What are the best locations for fast food franchises and/or pizza places in Toronto (specifically postal codes starting with 'M')?**

**SEE BELOW:**

**Let’s look at the Fast Food dataframes:**



From the fast food barcharts on page 11, the first cluster had the fewest fast food restaurants. This represents a possible opportunity in ‘underserved’ postal code areas.

The first dataframe above is sorted by cluster and shows the first 10 rows of postal codes. (Cluster 0 has the lowest number of fast food restaurants according to Foursquare.)

The second dataframe above is sorted by population showing the postal codes with highest 10 populations.

None of the postal codes overlap in the two dataframes. (This true even if I go to 20 rows.)

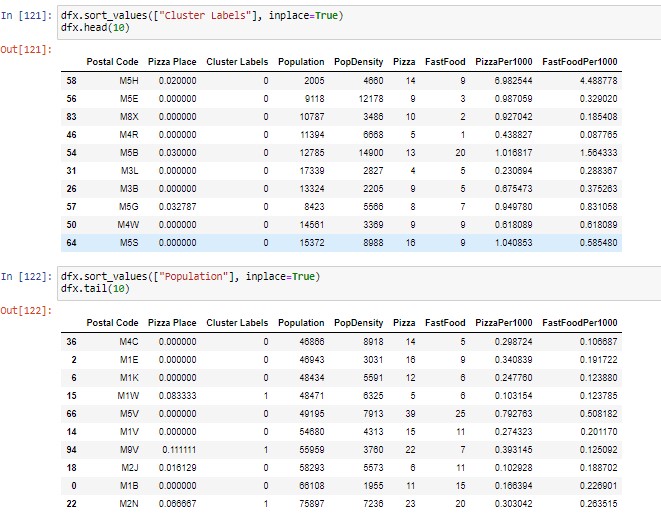
My recommend is take the lowest 5 FastFoodPer1000 (most underserved) from each dataframe and start looking for available properties/locations. (The missing data here are commercial rental rates. That data is not free.)

**Prospective postal codes:**

**From Cluster0: M6S, M9R, M6J, M2M, M4K**

**From highest Population: M4C, M1W, M1K, M9V, M2J**

**Let’s look at the Pizza Place dataframes:**



From the pizza place barcharts on page 12, the first cluster had the fewest pizza places. This represents a possible opportunity in ‘underserved’ postal code areas.

The first dataframe above is sorted by cluster and shows the first 10 rows of postal codes. (Cluster 0 has the lowest number of pizza places according to Foursquare.)

The second dataframe above is sorted by population showing the postal codes with highest 10 populations.

Again, none of the postal codes overlap in the two dataframes.

My recommend is take the lowest 5 PizzaPer1000 (most underserved) from each dataframe and start looking for available properties/locations. (The missing data here are commercial rental rates. That data is not free.)

**Prospective postal codes:**

**From Cluster0: M3L, M4R, M4W, M3B, M8X**

**From highest Population: M2J, M1W, M1B, M1K, M1V**

***Noteworthy are the overlapping postal codes from the Fast Food and Pizza Place recommendations above:***

***M2J, M1W, and M1K***