**The best place for a Brazilian restaurant**

**Wesley Silva**

**October 20, 2019**

**1. Introduction**

**1.1 Background**

The problem consists in selecting the right location where to open a Brazilian restaurant in Toronto, in order to make the right choice in terms of neighborhood.

This is interesting because it requires both location data (through the Foursquare API of course) and other kind of data to evaluate how a restaurant is doing.

It is of extreme importance for a businessman to find the right spot where to open a restaurant that can generate profits for a long time. 70% of restaurants close after six months, so the problem is real and evident.

**2. Data acquisition and cleaning**

**2.1 Data sources**

The data used will be the Canada postal codes list from [Wikipedia](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and to complement this data set we will be using the Foursquare API and a csv file containing the latitude and longitude downloaded from Coursera. After all data merged, we will look at the neighborhood for each of the Toronto postal codes found in the refined list.

**2.2 Data cleaning**

The data scraped from Wikipedia will be merged with the data returned from Foursquare API into a panda dataframe using python, we will refine the dataframe, as you can see in the figure below, we need to remove some unusual data from the Wikipedia and then display some relevant information regarding the data.

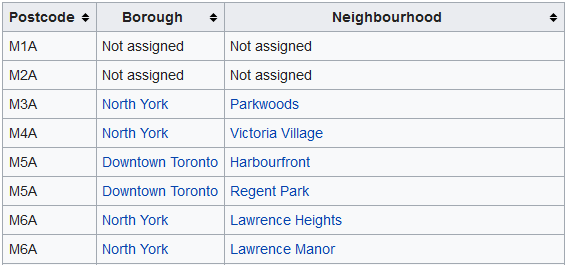


Figure 1. List of Canada postal codes

The following image is the scraped and refined data from Wikipedia loaded into a python dataframe merged with csv file downloaded from Coursera (as the foursquare API has a limitation for free account, we will be analyzing the East Toronto borough only and its neighborhood):

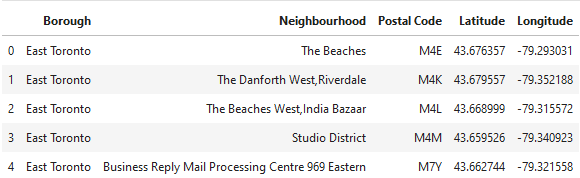


Figure 2. Refined postal code list

**3. Methodology**

With the refined dataframe, I got from foursquare API a list of 10 most common venues from each East Toronto neighborhoods found in the dataframe as showed in the figure below:

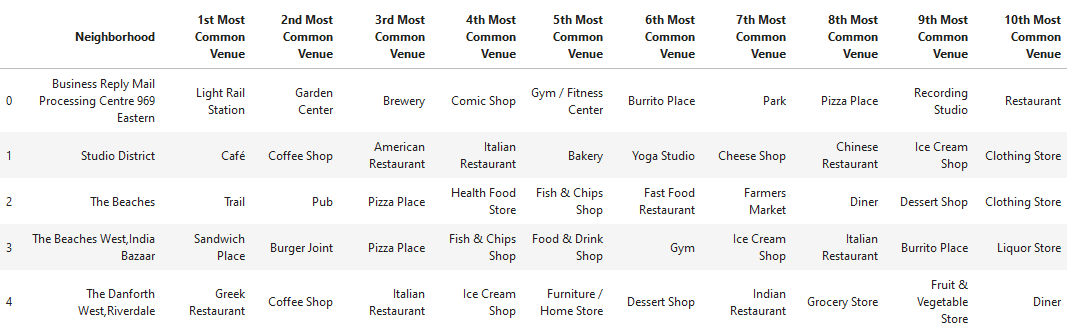


Figure 3. The 10 most common venues in Toronto neighborhoods

Once I discovered the 10 most common venues in East Toronto neighborhoods, I applied a k-means clustering algorithm and the number of clusters choose was 5 and we can see the results in the next section.

**4. Result**

The cluster map looks like this:

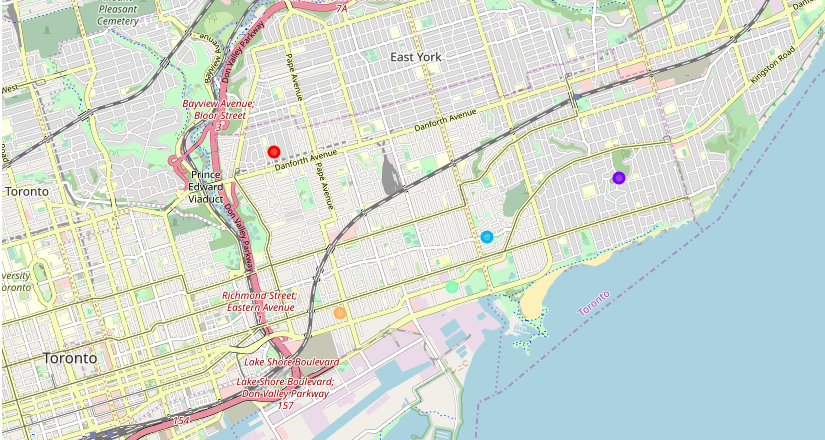


Figure 4. Cluster map from East Toronto

By examining the cluster individually, we will have the following results:

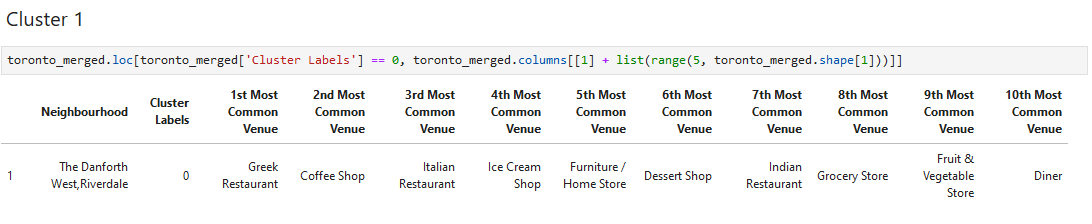


Figure 5. The Danforth West, Riverdale Cluster

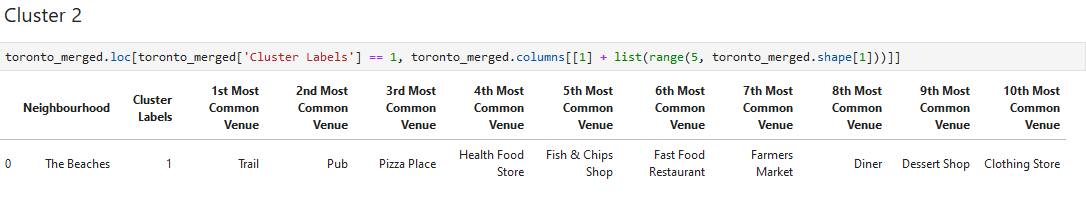


Figure 6. The Beaches Cluster

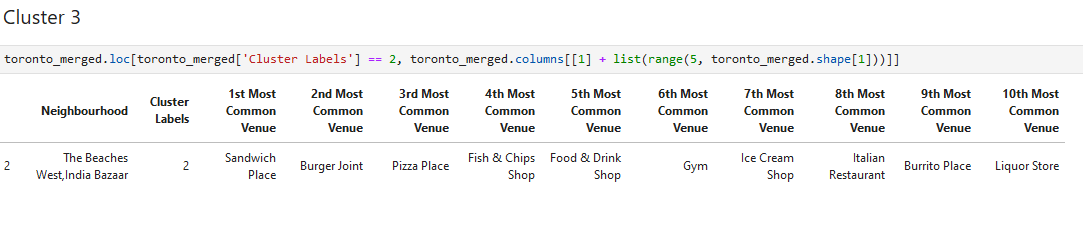


Figure 7. The Beaches West, India Bazaar Cluster



Figure 8. Studio District Cluster

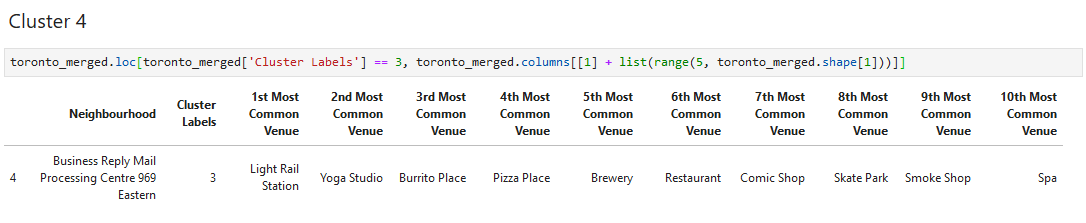


Figure 9. Business Reply Mail Processing Centre 969 Eastern

**5. Discussion**

It is clear in the map analyzes that the East Toronto is a good place to have a restaurant, but the question is, in which neighborhood? By analyzing the clusters individually, we will have an idea of which kind of venues are already there so we can decide based on that analysis the best place for the Brazilian restaurant. For example, the Beaches West, India Bazaar Cluster show us a list of food services, is the intention to have competition with these already stablished restaurants? If the answer is yes, so we should consider this Borough, if not, what about the Business Reply Mail Processing Centre 969 Eastern neighborhood? It doesn’t have many restaurants, is well located and must be take in consideration as a potential neighborhood to have the Brazilian restaurant.

**6. Conclusion**

To conclude this report, I can say that the analysis gave a satisfactory result, of course that it could have been done using different approaches and different machine learning algorithms, but for the propose of to know potential venue to open a Brazilian restaurant this report has a good utility and can be used to have some insights.