**The Battle of Neighborhoods**

**Introduction**

A businessman wants to open a new restaurant, but he does not know where is the best place. He has information that New York City (EEUU) and Toronto (Canada) are good options to start, but he only has a budget to open only one.

This businessman wants to get better profits reducing cost and receiving many clients.

**Problem**

Use data to identify the best place where a new restaurant get better profits.

**Interest**

The businessman who wants to open a new restaurant.

**Data**

**Data sources**

Information of Toronto and New York is on Internet. In this case, I collected data from FourSquare, Wikipedia and Coursera.

* <https://cocl.us/new_york_dataset>
* <https://geo.nyu.edu/catalog/nyu_2451_34572>
* <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

**Data cleaning**

Steps:

* Data was store in dataframes using pandas library.
* Missing values were dropped.
* To collect information about localization I used FourSquare API.
* Categorical features were transformed to numerical, and then were transformed to One-hot encoding.

**Methodology**

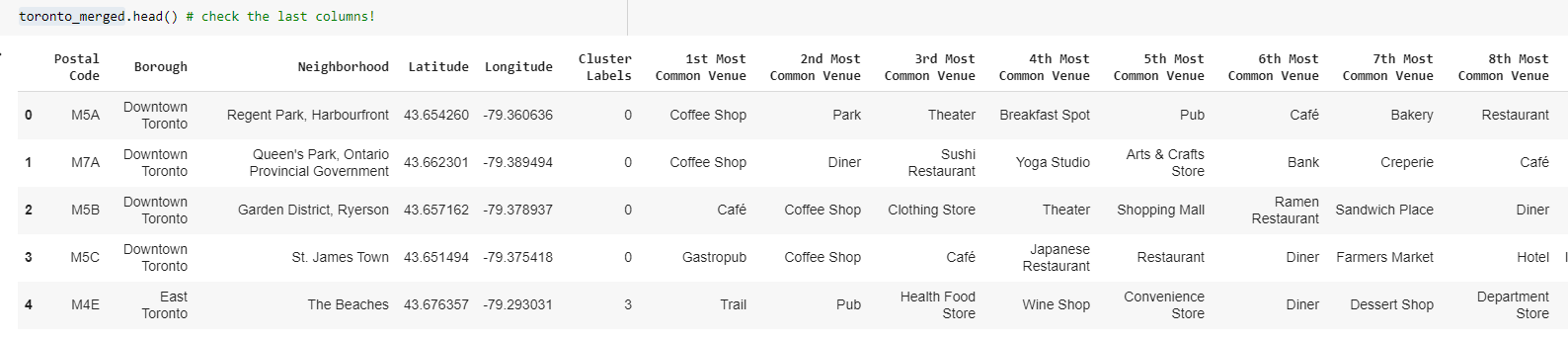
In this work, I decided to use unsupervised methodology like K-means clustering because I wanted to look for common behaviors in the data.

To work with K-means clustering I use categorical transformation, because it is easier to work with clustering.

New York data:



Toronto data:



**Results**

The results show that most common venues in both cities are Coffee Shops. In case of restaurants, there are few of them, but they are more common in New York.

|  |  |
| --- | --- |
| Toronto | New York |
|  |  |

**Discussion**

There is no any observation.

**Conclusion**

Showing the results, a better place to open a restaurant is in Toronto, because there less competitors.