Applied Data Science Capstone

Presentation – Week 5

NEIGBORHOOD RECOMMENDED TO OPEN AN ITALIAN RESTAURANT IN TORONTO

Carlos A. Gallo
October 2019

Introduction & Discussion

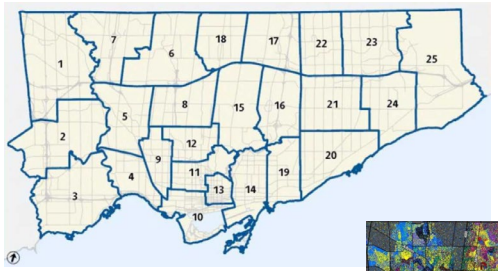
Toronto is Canada’s largest city and is recognized as one of the most multicultural cities in the world. Toronto is the most populous city in Canada and the capital of Ontario province.

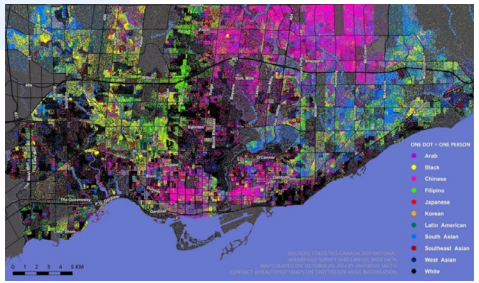
Since Toronto is one of the world’s top cities to invest in and for starting a company and has become an important destination for immigrants to Canada, where more tan 50% of the residents belong to an especific population group, this project will try to recommend which neighborhoods are more suitable to open an Italian restaurant. Besides, sinde acording to the data (2017) Italian is placed in the top 20 of ethnic origins in this city with almost 7% of the total population, it would be a great business idea to start a restaurant in certian neigborhoods that provide a higher profit.

Description of the data

The following information will be used in the project:

- **https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M** - “List of Postal code of Canada: M”: Information related to all the neighborhoods in Toronto.
- **https://cocl.us/Geospatial_data**: A csv file that contain all the geographical coordinates of the neighborhoods.
- **https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversityTo** - “Demographics of Toronto”: Information about the distribution of population by their ethnicity.
- **https://developer.foursquare.com/docs** - Foursquare API: Information about location and venues in Toronto





FOURSQUARE

Methodology

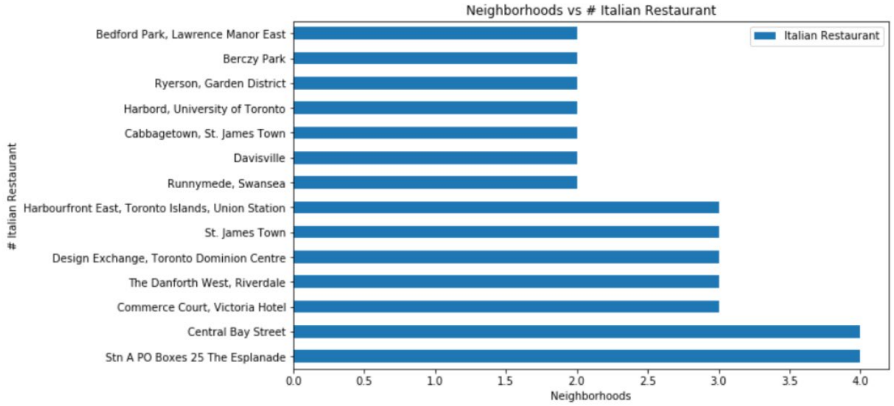
GETTING DATA & PREPROCESSING

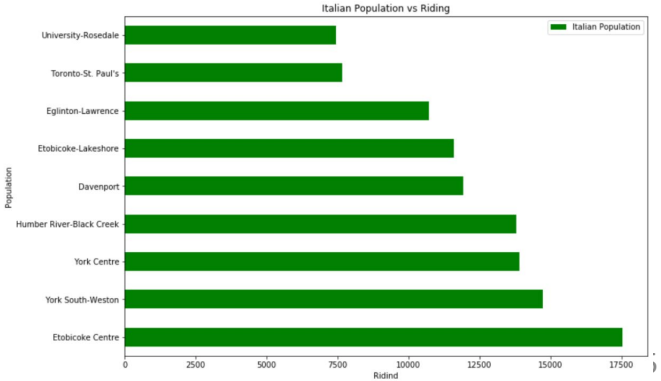
- The data referring to Toronto Neighborhoods was obtained by scraping the Wikipedia webpage: “List of Postal code of Canada: M”.
- Cells just with Borough assigned were processed and if a cell had a borough but a Not assigned neighborhood, then the neighborhood was the same as the borough
- Using the csv file: https://cocl.us/Geospatial_data”, the corresponding coordinates (latitude & longitude) were assigned to each neighborhood
- Data related to ethnic distribution was obtained by scraping the Wikipedia webpage: "Demographics of Toronto". The information belongs to the ridings of:
- Using the Foursquare API, information about location and venues in Toronto was obtained.

Methodology

EXPLORATORY DATA ANALYSIS

- It was extracted the Neighborhood and Italian Restaurant column from dataframe obtained in the previous analysis
- Population tables with the ethnic percentage by neighborhood were merged
- The neighborhoods with highest Italian population percentage were picked out
- A dataframe with Italian ethnic percentage & riding was created.
- Since there are few ridings, I created a dataframe and fill with its respective neighborhood
- The dataframe that contains the Italian populations with the riding was merged with the dataframe that contains information related to neighborhoods and ridings.
- Finally, it was obtained a relationship between Italian population and Italian restaurants

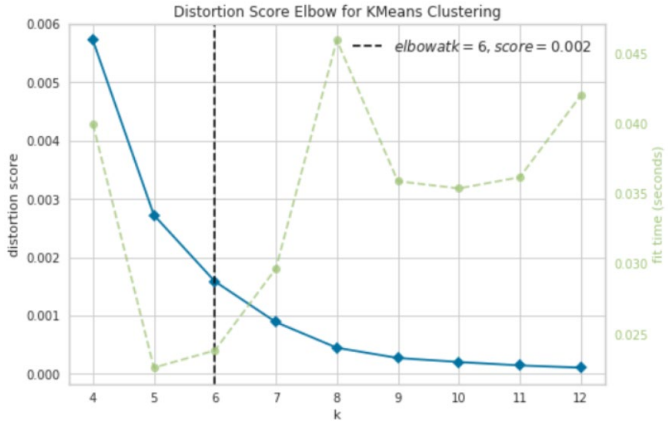




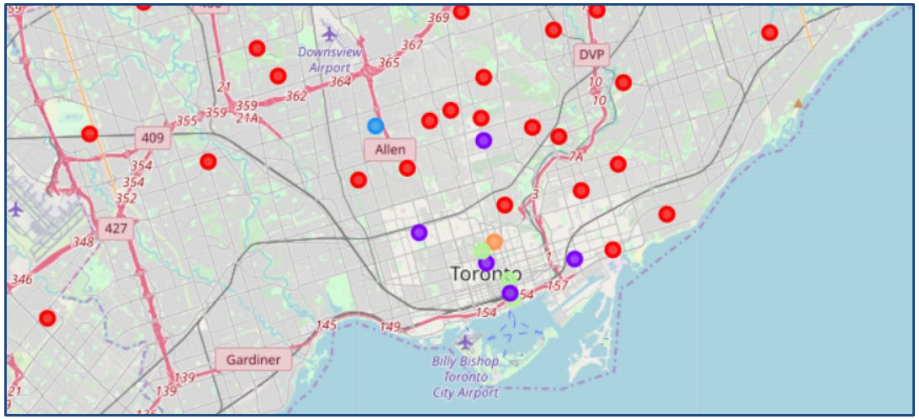
Methodology

MODELING K-MEANS

- After using Distortion Score Elbow for KMeans Clustering for each K value, k = 6 was determined as the best value



- It was clustered the Toronto Neighborhood using K-Means (k = 6) and each cluster was evaluated..



Results

- Regarding to the clusters it is possible to see:

CLUSTER	OBSERVATION
0	The neighborhoods have the smallest number of Italian restaurants.
1	The neighborhoods have a medium amount of Italian restaurants
2	The neighborhood where more Italian restaurants exist.
3	No one neighborhood is close to this centroid
4	The neighborhoods have a few amount of Italian restaurants.
5	The restaurants are scattered throughout that neighborhood

- Humber River-Black Creek, York Centre, York South-Weston and Etobicoke Centre are the ridings where more Italian population is present.

- Among others, Central Bay Street is the neighborhood where more Italian restaurants exist so it will not be an option as a place to open an Italian restaurant.

- The riding of Humber River—Black Creek include the neighbourhood of **Humber Summit** which has a considerable amount of Italian population is the ideal neighborhood to open a restaurant of the characteristics proposed since according to the cluster analysis performed, it has a very low concentration of Italian restaurants.

Dicussion & Conclusion

Discussion

After the analysis carried out, it can be conclusively concluded that Humber Summit is the neighborhood that would provide the greatest benefits when opening an Italian restaurant since it not only has a relatively high Italian population but also the concentration of Italian restaurants is significantly less than the other neighborhoods.

Conclusion

During this analysis, there was an opportunity not only to put into practice what was learned, but additional research was also necessary. I had the chance to analyze information from different data sources and use the Foursquare API to meet the proposed objective.

THANK YOU!