

Applied Data Science Capstone

*Find the best place to open a new Brazilian restaurant in Toronto,
Canada*

Introduction

Assuming that I am a Brazilian immigrant who wants to open a new Brazilian restaurant in Toronto, I need to find out where is the best place to start my enterprise. As I could see when I lived in Brazil, people who like to eat Italian food have a big tendency to like Brazilian food. Therefore, with this assumption, the objective of this project is to define the best place to open my Brazilian restaurant based on the location of existent Italian restaurants.

Business Problem

By using the data analysis and machine learning tools, allied to the clustering machine learning algorithms learned throughout the course I want to answer the question: Where is the best place to consider if I want to open a Brazilian restaurant in Toronto?

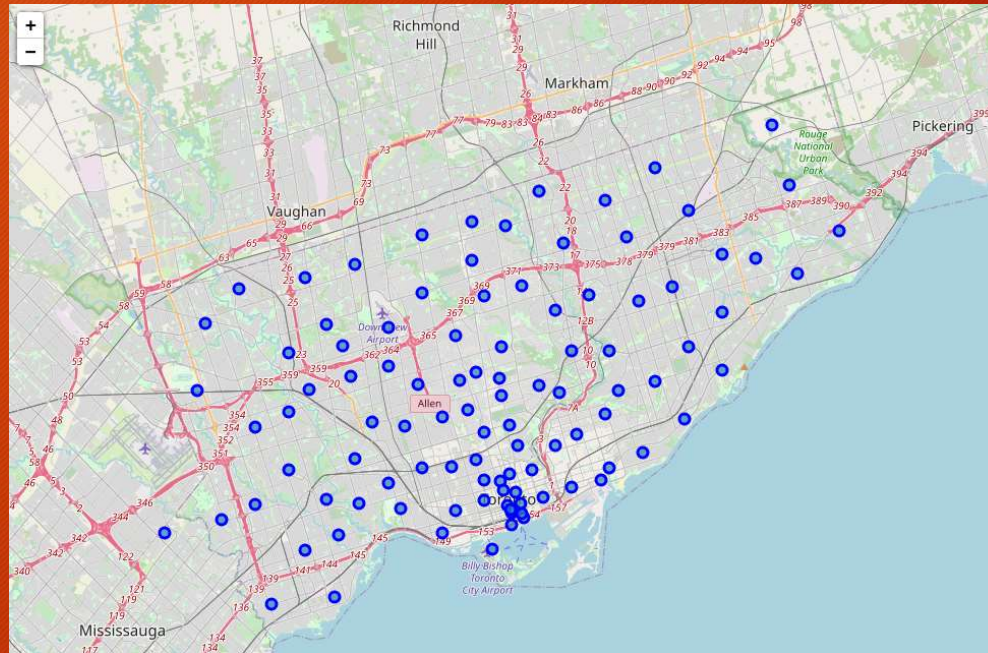
The target audience of this work are entrepreneurs who wants to find the location to open a reliable Brazilian restaurant in Toronto.

Data Sources

- A list of neighborhoods in Toronto, Canada: the information source is the Wikipedia web site https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and the data will be extracted from the table present on this site.
- The geospatial coordinates of these neighborhoods: the coordinates were obtained from the 'Geospatial_Coordinates.csv' file, provided by the course
- Venue data from Foursquare of these neighborhoods: The data was obtained using the Foursquare API to find out where the Italian restaurants are located in Toronto.

Methodology

- Get the list of neighborhoods in Toronto and clean the data
- Merge the result set with the geospatial data
- Plot the results on a map



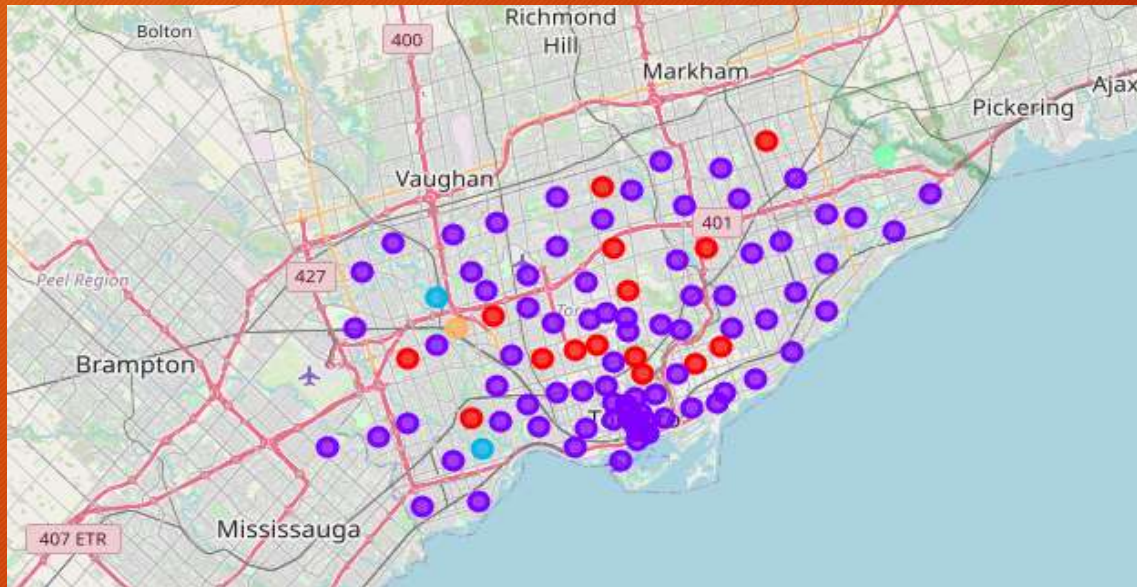
Methodology

- Get venue data from Foursquare
- Group the result set by category
- Prepare the data to the cluster analysis

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Skating Rink	Latin American Restaurant	Breakfast Spot	Clothing Store	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant
1	Alderwood, Long Branch	Pizza Place	Coffee Shop	Sandwich Place	Athletics & Sports	Pub	Pool	Skating Rink	Gym	Concert Hall	Department Store
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Park	Deli / Bodega	Supermarket	Middle Eastern Restaurant	Sushi Restaurant	Ice Cream Shop	Shopping Mall	Mobile Phone Shop
3	Bayview Village	Café	Bank	Chinese Restaurant	Japanese Restaurant	Women's Store	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Thai Restaurant	Sandwich Place	Restaurant	Juice Bar	Butcher	Café	Indian Restaurant	Pub

Methodology

- Run k-means algorithm - 5 clusters



Results and Discussion

- Get the cluster with the most Italian Restaurants: Cluster 2
- Get the borough which the most common venue are Italian Restaurants

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
55	North York	1	Italian Restaurant	Coffee Shop	Thai Restaurant	Sandwich Place	Restaurant	Juice Bar	Butcher	Café	Indian Restaurant	Pub

Conclusion

With the assumption that people who like to eat Italian Food and after the analysis of the clustering data, obtained with a dataset created with Foursquare venue data, we came to the conclusion that the best borough in Toronto, Canada, to open a Brazilian Restaurant is North York.