



# Toronto Neighborhood Classification for Business Expansion

By: David Marcus Thierry  
January 7, 2021



# Objective

- The objective of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood in Toronto to open a Italian restaurant.
- Through this project, we will find the most suitable location for an entrepreneur to open a new Italian restaurant in Toronto, Canada.



# Target Audience

- Target Audience Entrepreneurs who are passionate about opening an Italian restaurant in a metropolitan city such as Toronto would be extremely interested in this project.
- The project is also for business owners and stakeholders who want to expand their businesses and wonder how data science could be applied to the questions at hand.



# Data

The data that will be required will be a combination of CSV files that have been prepared for the purposes of the analysis from multiple sources.

- Source 1: Neighborhoods in Toronto (via Wikipedia)
- Source 2: Geographical location of the neighborhoods ([http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data))
- Source 3: Venue data pertaining to Italian restaurants (via Foursquare)

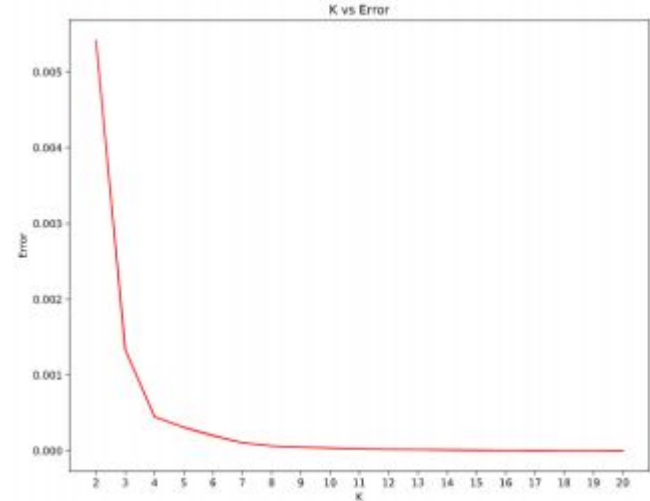
The Venue data will help find which neighborhood is best suitable to open an Italian restaurant.

# Model & Tuning

We will use K-Means clustering as our clustering method.

In order to tune our model, I used my knowledge of Clustering from earlier in the course and found that the optimal number of K in this particular application was  $k = 4$ .

Thus, the optimal number of clusters in our model should be 4.





# Results

After merging our predictions and our main data together, we notice two things.

- Our data is categorized based on our own heuristic
- We have designated intervals assigned to each cluster

	Neighborhood	Italian Restaurant	Cluster Labels
0	Berczy Park	0.000000	0
1	Brockton, Parkdale Village, Exhibition Place	0.045455	3
2	Business reply mail Processing Centre, South C...	0.000000	0
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0
4	Central Bay Street	0.047619	3



# Cluster Analysis

## Cluster 1

- Cluster 1 was mainly in the East Toronto Area. First Canadian Place, Underground city, Richmond, Adelaide, King were among some neighborhoods that were in that cluster. Cluster 1 had 173 unique Venue locations and out of those only 1 were Italian Restaurants. Cluster 1 had the lowest average of Italian Restaurants equating to 0.0.

## Cluster 2

- There was a total of 140 neighborhoods, 63 different venues and 9 Italian Restaurants. Therefore, the average amount of Italian Restaurants that were near the venues in Cluster 2 is the highest being 0.07. In the map and from our numbers, we can see that nodes of Cluster 2 were dispersed mostly evenly throughout Toronto making it one of the most sparsely populated clusters.



## Cluster Analysis (Part 2)

### Cluster 3

- Cluster 3 had the second to lowest average of Italian Restaurants. Cluster 3 was mainly located in the Downtown area but also had some neighborhoods in West Toronto, East Toronto and in North York. Neighborhoods such as Ryerson, Toronto Dominion Center, Garden District, Queen's Park and many more were included in this cluster. There was a total of 162 unique venues and out of those 20 were Italian Restaurants.

### Cluster 4

- Cluster 4 venues were located in the Downtown, West, East Toronto areas. Neighborhoods such as Central Bay Street, St. James Town, Cabbagetown were some of the neighborhoods that made up this cluster. There was a total of 91 unique Venues in Cluster 4 with 10 Italian Restaurants. This made up the second-highest average of Italian Restaurants in that cluster which was approximately 0.047.





# Discussion

After carefully inspecting the details of each cluster, and by also looking at the nearby venues, the optimal place to put a new Italian Restaurant would be in Cluster 1. There are many Neighborhoods in the area and virtually no Italian Restaurants, therefore, eliminating any opening competition.

The second-best Neighborhoods that have a great opportunity would be in areas such as which is in Cluster 3.



# Conclusion

In conclusion, we had an opportunity on a business problem and tackled it in a way that it was similar to how a genuine data scientist would do. Utilizing numerous Python libraries to fetch the information, control content and break down and visualize those datasets. Through the utilization of Foursquare API to investigate the settings in neighborhoods of Toronto and getting a great measure of data from Wikipedia from general scraped. Finally, applying AI strategy to anticipate the error given the information and utilized Folium to picture it on a map.