# COURSERA CAPSTONE
# IBM Applied Data Science Capstone

## Opening Restaurant in Toronto, Canada

By : Bernardo Andrey Panggabean

July 2020

# BUSINESS PROBLEM

- Many business decision, opening a new restaurants requires serious consideration and a lot more complicated than it seems
- Location of the restaurants is one of the most important decisions that will determine whether the restaurants will be a success or a failure
- The objective of to analyze and select the best location in Toronto, Canada to open a Restaurants

# DATA

- List of districts in Toronto. This defines the scope of this project which is to confined to the city of Toronto, the capital city of Canada.
- Latitude and longitude coordinates of those districts. This is required in order to plot the map and get the venue data from foursquare.
- Venue data, particularly data related to restaurants. We will use this data to perform clustering on the districts.

# SOURCE OF DATA

- Wikipedia page for districts in Toronto
- Geocoder package for latitude and longitude coordinates
- Foursquare API for venue data

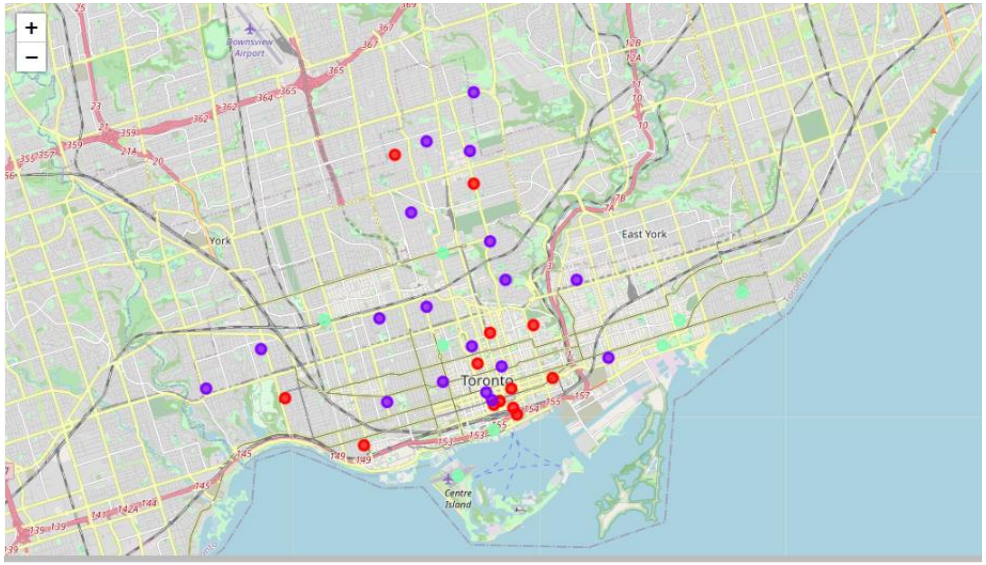
# METHODOLOGY

- Web Scrapping wikipedia page for district list
- Use Geocoder package to get latitude and longitude
- Use Foursquare API to get venue data
- Group data by districts and taking the mean of the frequency of occurrence of each venue category
- Filter venue category by restaurants
- Perform clustering on the data using k-means clustering
- Visualization the clusters using folium package

# RESULT

- The results from the k-means clustering show that we can categorize the districts into 3 clusters based on the frequency of occurrence for restaurants:
  - Cluster 0: Districts with high number to no existence of restaurants
  - Cluster 1: Districts with moderate number of restaurants
  - Cluster 2: Districts with low concentration of restaurants

  The results of the clustering are visualized in the map below with cluster 0 in red, cluster 1 in purple, and cluster 2 in mint green



# CONCLUSION

- Most of the restaurants are concentrated of the South area of Toronto
- Highest number in cluster 0
- Cluster 2 has very low number to no coffee shops
- To open new restaurants, better to open in area cluster 2
- Avoid cluster 0, because there is intense competition.

# DISCUSSION

- Although location is important to place restaurants but many factor success in open new restaurants.
- Can try Hierarchical Clustering to perform this project and that 2 method that have high accurate.
- And this project still using free Foursquare API that have limitations.

# THANK YOU