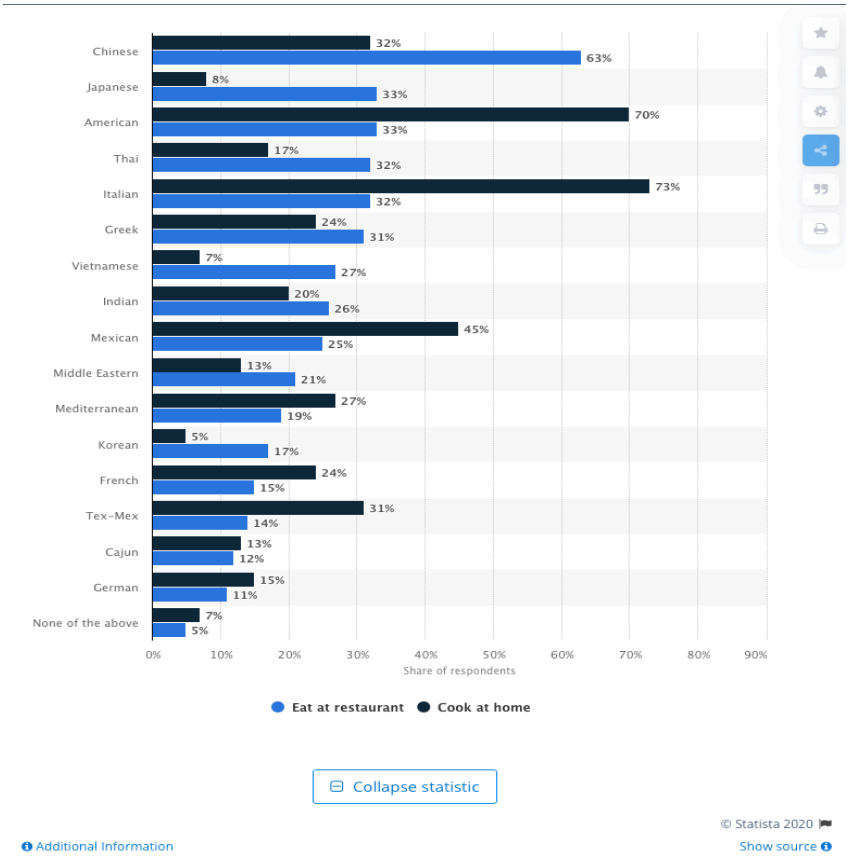
Data Science Project -
The Battle of Neighborhoods- Week 4

INTRODUCTION

This is an IBM Applied Data Science Capstone project. The concept behind this project is to find out areas in which there less or no Italian Restaurants in Toronto. As per the survey, the Italian cuisine is highly liked in Toronto, as shown in the below statistics.



Therefore it might be a great opportunity for an entrepreneur, based in Canada to invest in an Italian restaurant. So, I am laying out an outline of my observation to help the entrepreneur find the best and most suitable location.

BUSINESS PROBLEM

The objective of this project is to find the most suitable location for an entrepreneur/firm to setup a restaurant of Italian cuisine in Toronto, Canada. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question: "If an entrepreneur wants to open an Italian restaurant in Toronto, which is the most suitable location?

TARGET AUDIENCE

Entrepreneur who wants to find a suitable place to open an Italian restaurant.

DATA

In this project, following data has been used:

1. **List of neighborhoods in Toronto, Canada**
2. **Latitude and Longitude of these neighborhoods**
3. **Venue data related to Italian restaurant.**

DATA SOURCE

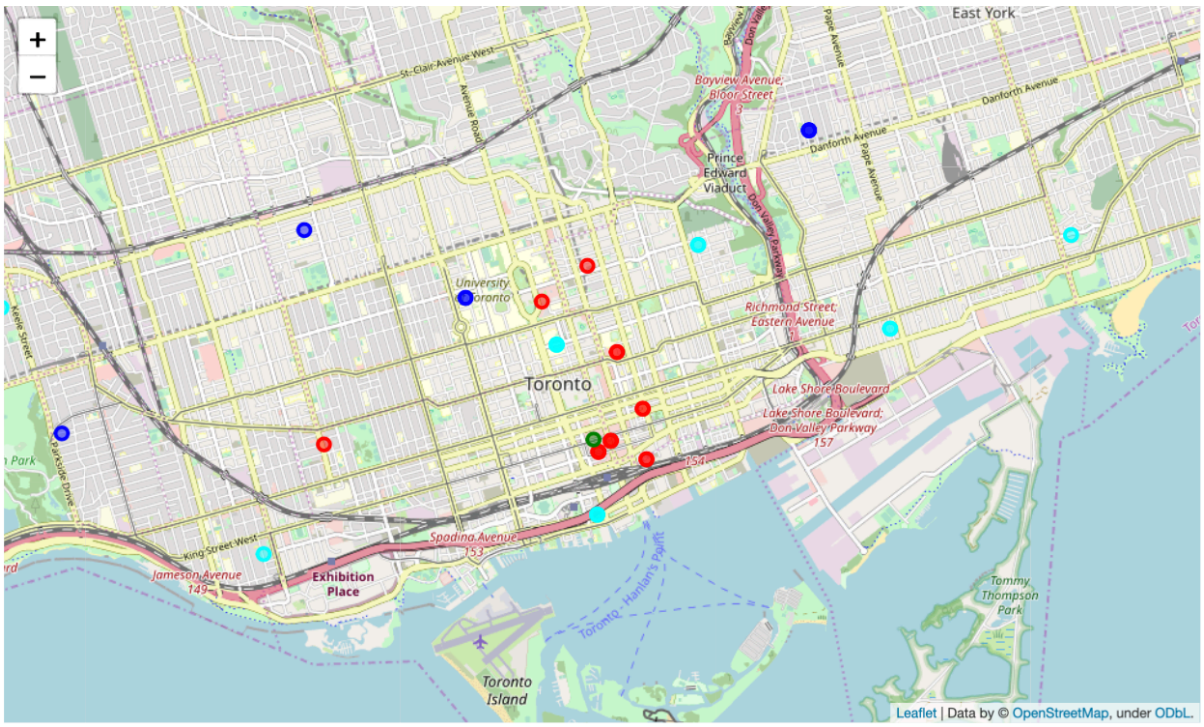
1. List of neighborhoods in Toronto – Scrapping from Wikipedia
2. Latitude and Longitude – Geocoder library
3. Venue data (related to Italian restaurants) – Foursquare API

METHODOLOGY

The first step was to get the list list of neighborhoods in Toronto, Canada. This data is available on Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Web scrapping was done with the help of Pandas HTML library table scraping method as this helps in easy conversion of data on web into a data frame.
The second step was to get coordinates of the retrieved data so as to get popular venues near these neighborhoods. So, to get the coordinates, Geocoder package was used.
The third step was to visualize a map of the listed neighborhoods with the help of Folium package.
The fourth step was to get venues from Foursquare. For this project, a Foursquare developer account and API key was used to retrieve top 100 venues within the limit of 500 meters radius.
Since this project focused on Italian cuisine so specifically that data was retrieved. Lastly, the clustering method was performed using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster.

RESULT



The results from k-means clustering show that we can categorize Toronto neighborhoods into 4 clusters based on the number of Italian Restaurants in each neighborhood:

- Cluster 0: Neighborhoods in this cluster has 15 Italian restaurants altogether.
- Cluster 1: Neighborhoods in this cluster has 9 Italian restaurants altogether.
- Cluster 2: Neighborhoods in this cluster has just 1 Italian restaurant.
- Cluster 3: Neighborhoods in this cluster has 17 Italian restaurants altogether.

RECOMMENDATIONS

Cluster 3 has the most number of Italian Restaurants which includes **Habourfront East, Toronto Islands, Union Station, Brockton, Exhibition Place, Parkdale Village, The Beaches West, India Bazaar, St. James Town.**
Cluster 0 and 1 has 15 and 9 Italian restaurants respectively which includes **Harbord, Christie, The Danforth, Davisville, Church and Wellesley, Queen's Park, Commerce Court, Ryerson and Design Exchange.**
Cluster 2 has only 1 Italian Restaurant in **First Canadian Place.**

So, as per the findings cluster 2 might be a good location as it has only 1 Italian restaurant. Hence, it is recommended for an entrepreneur to open an Italian restaurant in this area.