INTRODUCTION

The project explores the best locations of Indian restaurants in the Toronto neighborhood. Toronto is Canada's largest city with diverse

population with high international immigration. There are restaurants of different cuisines in the city but the project will focus on the

Indian cuisine. According to census Canada in 2016, there were 350,040 South Asian people in the Toronto region. This indicates a significant

amount of peolpe who might be looking to dine at the Indian restaurants.

The target audience for this project would be-

- A person looking to invest in an existing restaurant
- A person looking to open a new restuarant
- People who are looking to move in a different neighborhood

DATA

- Data for the Toronto neighborhoods and zip code would be collected from the webpage 'https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto'
- The longitude and latitude coordinates would be collected from https://cocl.us/Geospatial_data
- Data would be transformed into a dataframe as well as cleaned and refined
- Data related to locations and quality of Indian restaurants would be obtained via the FourSquare API
- FourSquare would be used to locate all venues and then filtered by Indian restaurants. Ratings, tips, and likes by users will be counted and added to the dataframe.
- Data would then be sorted based on rankings
- Finally, the data be will be visually assessed using graphing from various Python libraries.

METHODOLOGY

The first step is to get the list of neighborhoods from the Wikipedia webpage : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

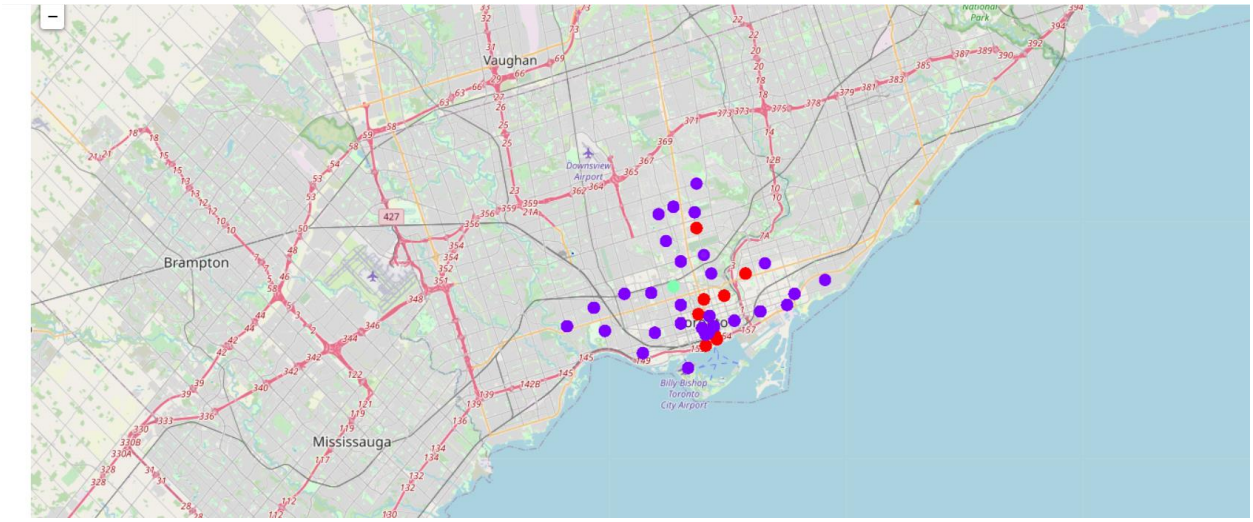
We used the web scraping method by utilizing beautiful soup scraping method as it is easier and more convenient to pull tabular data into a data frame.

We need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinate, the CSV file provided by IBM team to match the coordinates of Toronto neighborhoods was used. After gathering the coordinates, the map of Toronto was visualized using Folium package to verify whether these are correct coordinates. Next, the Foursquare API was used to pull the list of top 100 venues within 500 meters radius. The Foursquare developer account helped pull the data. From Foursquare, data about the names, categories, latitude, and longitude of the venues was obtained. With the help of this data, we could also see how many unique categories we could get from these venues. To prepare ourselves for clustering, each neighborhood was then analyzed by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category.

Lastly, the k-means clustering method was performed. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. The neighborhoods in Toronto were made into three clusters based on their frequency of occurrence for “Indian food”. Based on the results (the concentration of clusters), we would be able to recommend the ideal location to open the restaurant.

RESULT

CLUSTERS



The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Indian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with a smaller number of Indian restaurants.
- Cluster 1: Neighborhoods with no Indian restaurants.
- Cluster 2: Neighborhoods with a greater number of Indian restaurants

The results are visualized in the above map with Cluster 0 in red, Cluster 1 in blue, Cluster 2 in green.

RECOMMENDATIONS

Most of the Indian restaurants are in cluster 0 which is around Central Bay Street, Church and Wellesley, Berczy Park, Union Station, Richmond, lowest in Cluster 1 areas which are in North Toronto West and Parkade areas. Also, there are good opportunities to open near St James Town, Cabbagetown. Looking at nearby venues, it seems cluster 2 might be a good location as there are not a lot of Indian restaurants in these areas.