Presentation

In this undertaking, I am making a speculative situation for the idea that there may not be sufficient Indian Restaurants in Toronto Area. In this manner, it may be an incredible open door for a business person who is situated in Canada. As Indian food is well known among the Asian people group, so this a business person may consider starting its business in zones where Asian network lives. In view of the reason, finding the area to open such a café is one of the most significant choices for this business visionary and I am structuring this task to assist him with finding the most reasonable area.

BUSINESS PROBLEM

The target of this capstone venture is to locate the most reasonable area for the business person to open another Indian Restaurant in Toronto, Canada. By utilizing information science strategies and instruments alongside AI calculations, for example, bunching, this task intends to give answers to answer the business question: In Toronto, if an a business person needs to open an Indian Restaurant, where should they think about opening it?

To take care of this issue, we will require underneath information:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these areas
- Venue information identified with Indian eateries. This will assist us in finding the neighborhoods that are progressively appropriate to open an Indian Restaurant.

How to get the data

- Scrapping of Toronto neighborhoods through Wikipedia
- Getting Latitude and Longitude information of these areas through Geocoder bundle

● Using Foursquare API to get scene information identified with these neighborhoods

METHODOLOGY

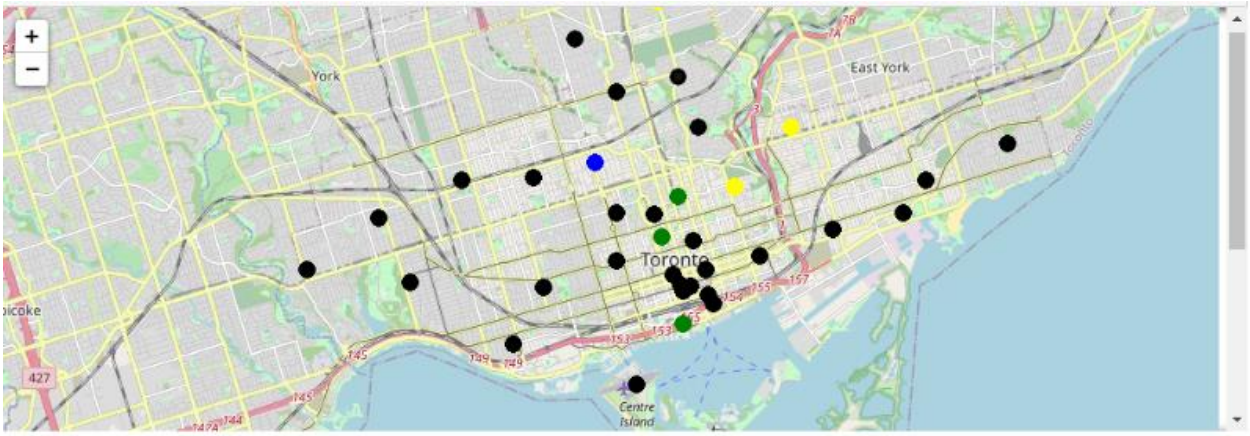
First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M I did the web scraping by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame. However, it is only a list of neighborhood names and postal codes. I need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder Package but it was not working so I used the CSV file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering these coordinates, I visualize the map of Toronto using Folium package to verify whether these are correct coordinates. Next, I use Foursquare API

to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names,categories, latitude, and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Here, I made a justification to specifically look for “Indian restaurants”. Lastly, I performed the clustering method by using k-means clustering. 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. I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Indian food”. Based on the results (the

concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

Results



The results from k-means clustering show that we can categorize Toronto neighborhoods into 4 clusters based on how many Indian restaurants are in each neighborhood.

Cluster black—With very no Indian restaurants

Cluster Green—With less number of restaurants

Cluster Blue—Slightly more than green

Cluster yellow—With maximum number of restaurants

Conclusion

Most of the Indian restaurants are in cluster 3 and 2 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 0 might be a good location as there are not a lot of Indian restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Indian restaurant in these locations.