

Coursera IBM Data Science Capstone Project

Opening a New Indian Restaurant in Toronto Canada

Introduction:

For this capstone Project I am creating a scenario in which an Indian who settled in Canada wants to explore opening an Indian restaurant in Toronto Area. The idea behind this project was because there are less Indian restaurant's in Toronto Area and it might present a great opportunity for the entrepreneur who is settled in Canada as the Number of Indians migrating from India are increasing every year, this entrepreneur is thinking of opening a Indian restaurant in locations where Indian food is popular (where Indian restaurant's and Indians in the neighborhood) with the purpose in mind, of finding the location to open this type of restaurant is big and crucial task, Hence I'm trying to help this entrepreneur to find the most suitable location to open the restaurant.

Business Solution:

As you can see the immigration list to Canada in https://en.wikipedia.org/wiki/Immigration_to_Canada .The percentage of Indians migration to Canada is increasing. Hence The Objective of this Project is to find the most suitable location to open an Indian restaurant in Toronto, Canada. By Using Data Science Methods and Machine Learning Such as Clustering. This Project aims to provide solutions to the Business Question.

Target Audience:

The entrepreneur who wants to find the location to open authentic an Indian restaurant.

Data:

To solve this Problem I will need the following Data

1. List of Neighborhoods in Toronto, Canada.
2. Co-ordinates of these Neighborhoods.
3. Venue Data related to Indian Restaurants
4. This will help us find neighborhoods that are most suitable to open a Indian Restaurants.

Extraction of Data:

1. Scrapping of Data from Wikipedia about Neighborhoods.
2. Getting Latitude of Longitude data of these Neighborhoods using **Geocoded Package**.
3. Using **Foursquare API** to get venue data related to these Neighborhoods.

Methodology:

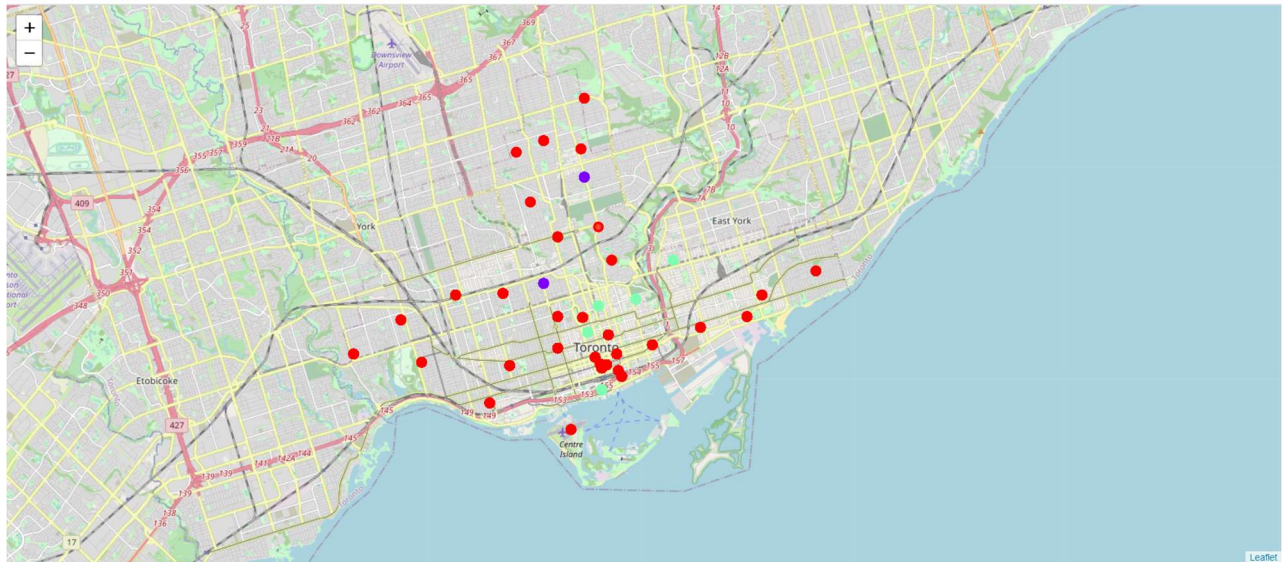
First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia page.

(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:M) I did the web scraping method as it is easier and more convenient to pull tabular data directly from a web page into data frame. However, it is only a list of neighborhood names and postal codes. I will need to get their co-ordinates, I tried using Geocoder package but it was not working so I used the CSV file provided by IBM team to match the co-ordinates of Toronto neighborhoods. After gathering all these co-ordinates, I visualized the map of Toronto using Folium Package to verify whether these are correct co-ordinates.

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. 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 looked for the “Indian Restaurants”. 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 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 3 clusters based on many Indian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with little or no Indian restaurants
- Cluster 1: Neighborhoods with no Indian restaurants
- Cluster 2: Neighborhoods with high number of Indian restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color and Cluster 2 in light green color.

Recommendations:

Most of Indian restaurants are in Cluster 2 which is around **Berczy Park, Harbord** areas and lowest (zero) in Cluster 1 areas which are **Davis Ville, Central Toronto (The Annex, North Midtown, Yorkville)**. Also, there are good opportunities to open near Berczy Park, Harbord as the competition seems to be low. Looking at nearby venues, it seems **Cluster 0** might be a good location as there are not a lot of Asian restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Indian restaurant in these locations with little to no competition. Nonetheless, if the food is authentic, affordable and good taste, I am confident that it will have great following everywhere.

Future Extension:

In this project, I only take into consideration of one factors:

The Occurrence of Indian Restaurants in each neighborhood.

There are many other factors that can be taken into consideration such as

1. Population density.

2. Income of residents

3. Cost of setting up of New restaurant like (Rent of the restaurant, Labor charges, Capital invest, etc.)

However to pull these data into this project is not possible to do alone that too from out of that Country. Future research can take into Consideration of these factors.

Conclusion:

In this Project we have gone through the process of identifying the Business Problem. Specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to stakeholder.

References:

1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:M
2. https://cocl.us/Geospatial_data
3. <https://developer.foursquare.com/docs/>