

Suitable Location Recommender

By:-J.sasikiran

1. Introduction:-

1.1 Background:-

In present days Finding perfect place to open restaurants is very important because it is directly impacting growth of that restaurant.

If we can find best place to open restaurant then our sales will increase eventually. There are some things which impact growth of restaurant those are place, competition from other restaurants. For example if we open another restaurant which offer Italian food in place where there are many such type of restaurants are present , then our business will not success. But if we can open some other restaurant which offer Indian food then there will be no or little competition. So that our business will be successful.

1.2 Problem:-

The objective of this capstone project is to find suitable regions for entrepreneur to open restaurants with different countries food recipes. By using Data science methods and machine learning techniques like clustering (K-means) and foursquare api , we can find suitable places for opening different types of restaurants.

This project aims to provide answer to business Problem: In Toronto , If an entrepreneur wants to open different types of restaurants where there will be no competition i.e... which places are good/suitable to open new restaurant?

1.3 Target audience:-

The entrepreneur who wants to find the location to open Different types of restaurants. “Entrepreneurs find this very interesting because it will greatly boost their business.”

2. Data Sources

To solve this Business problem, I am going to use following Data:-

- 1) List of Neighborhoods in Toronto, Canada
- 2) Latitude and Longitude of Neighborhoods
- 3) Details related to Restaurants like name , type of food etc..., from Foursquare. This will help me find most suitable region to open restaurants.

2.1 Extracting the Data :-

Scrapping of Toronto neighborhoods via Wikipedia

Getting Latitude and Longitude data of these neighborhoods via Geocoder

Using Foursquare to get venue details in these neighborhoods.

3. Methodology

First, I need to get the list of neighborhoods in Toronto, Canada. This can be achieved by extracting the list of neighborhoods from wikipedia page ([“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M))

I did the web scraping by utilizing pandas read_html table scraping method as it is 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 latitude and longitude values to utilize Foursquare to pull the list of venues in 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 and merged this with neighborhoods dataframe . After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

Next, I use Foursquare API to pull the required venues data within 500 meters radius. I have created a Foursquare developer account in order to obtain account

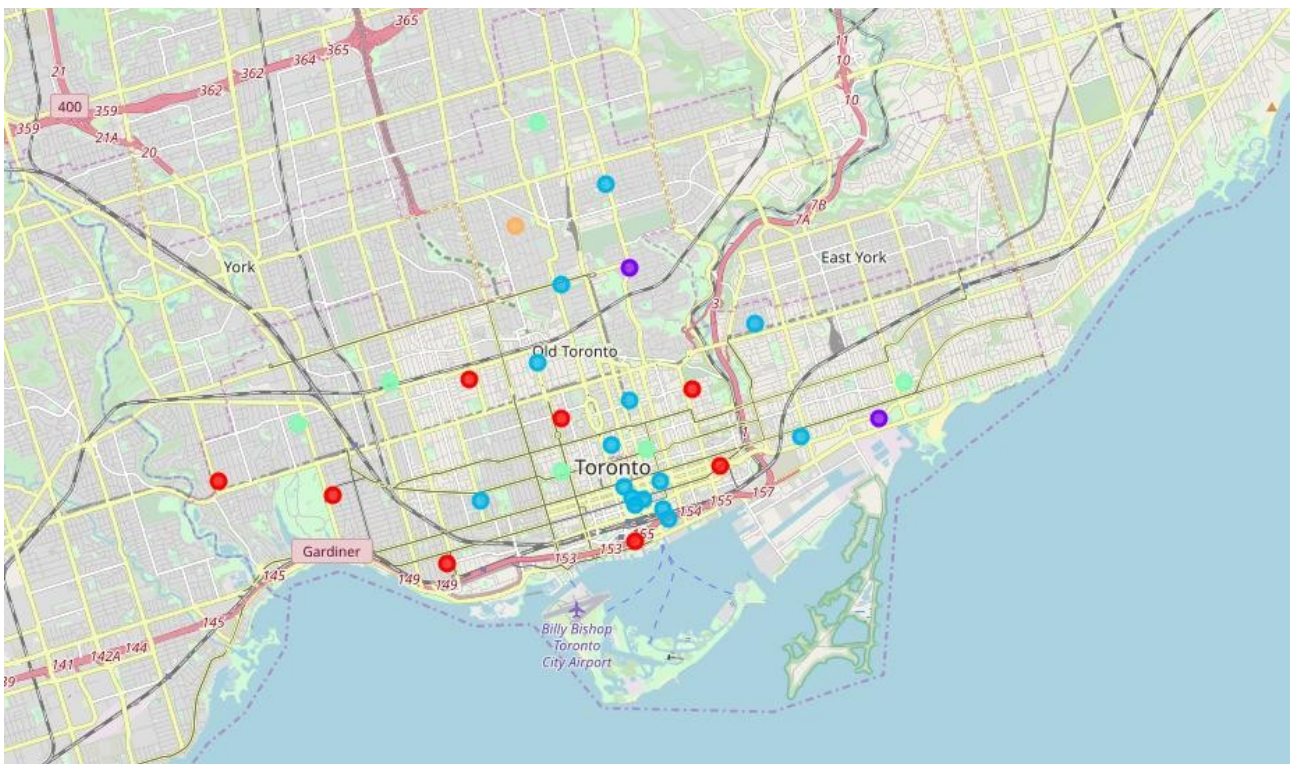
secret ID and secret key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeroids, 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 suitable for this project. I have clustered the neighborhoods in Toronto into 5 clusters. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the Different types of restaurant with no or little competition.

We will find types of restaurants in each cluster region then we will recommend type of restaurant an entrepreneur can open.

4.Results :-

Clusters



The results from k-means clustering show that we can categorize Toronto neighborhoods into 5 clusters based on most common Restaurants:-

- Cluster 0: Neighborhoods in downtown Toronto and west Toronto with most of the restaurants are Italian restaurants
- Cluster 1: Neighborhoods in central , east Toronto with Vietnamese Restaurants.
- Cluster 2: Neighborhoods with most of them are American and Asian restaurants
- Cluster 3: Neighborhoods with high number of Fast food restaurants
- Cluster 4: Neighborhoods with high number of sushi restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color, Cluster 2 in sky blue color, Cluster 3 in light blue color and Cluster 4 in orange color.

5. Recommendations:-

From above cluster 1 we can conclude that most of the restaurants are Italian. So we can suggest Investors or entrepreneur to start Restaurants other than Italian in downtown Toronto or West Toronto ,so that there will be no competition.

From Cluster 2 we can come to know that there are no Italian restaurants in central Toronto ,east Toronto so entrepreneur start Italian restaurants here.

and most importantly it is best opportunity for entrepreneur who are willing to open Indian restaurants , They can establish their restaurants anywhere in Toronto region because there is little or no competition.

6. 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 the stakeholder to open different restaurants.

