

Opening a new Indian Restaurant in Toronto, Canada



Report Submitted By:
Sunny Kumar Sharma

INTRODUCTION

For this project, I am creating scenario for a Indian Restaurateur who wants to explore opening an Indian restaurant in Toronto area.

The idea is behind this project is that there may not be enough Indian restaurants in Toronto and it might present a good opportunity for entrepreneur who is based in Canada. As Indian food is similar to others Asian Cuisines.

So this entrepreneur want to open this restaurants in locations where Asian Food is famous.

With this purpose I am creating this project to help the entrepreneur to locate a correct location.

Business Issues

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Indian restaurant in Toronto, Canada. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to the business question: In Toronto, if an entrepreneur wants to open a Indian restaurant, where should they consider opening it?

Audience

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

Data Extraction

- Scrapping of Toronto neighborhoods via Wikipedia
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

Data

To solve this problem, I will need below data:

- List of neighborhoods in Toronto, Canada.
- Latitude and Longitude of these neighborhoods.
- Venue data related to Asian restaurants.

This will help us find the neighborhoods that are most suitable to open an Indian restaurant.

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")

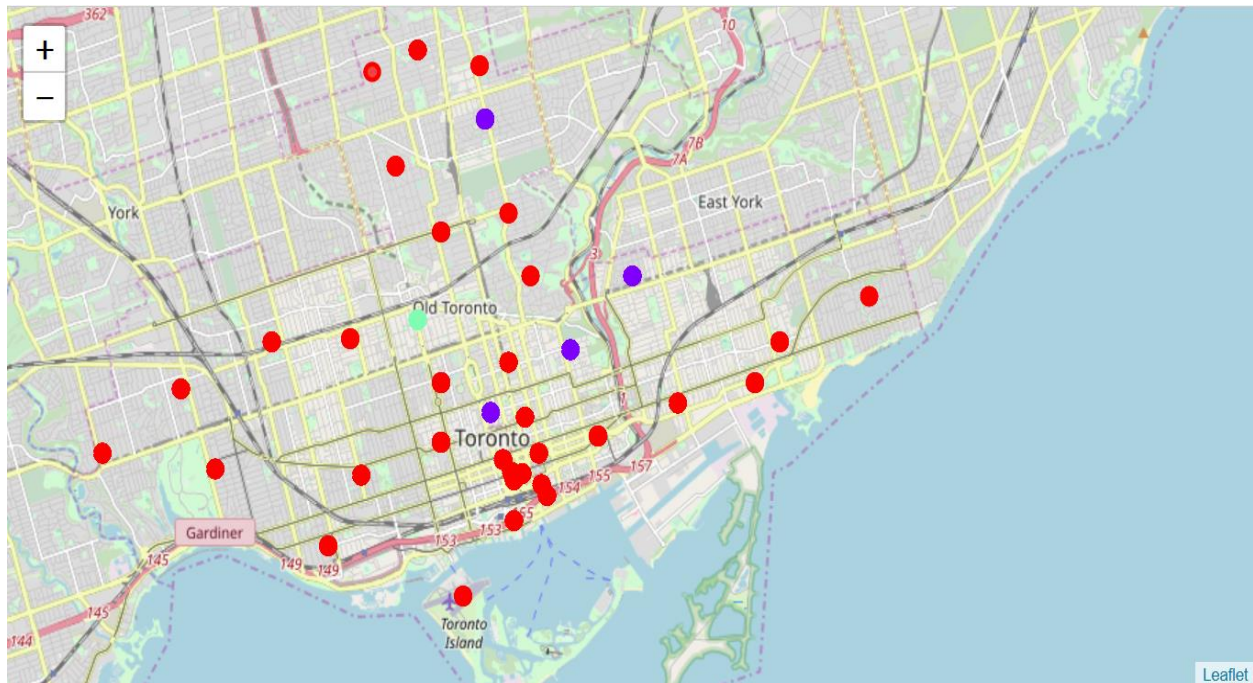
I did the web scraping and save it in csv file. However, it is only a list of neighborhood names and postal codes. I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder package but it was not working so I used the csv file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering 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 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". Previously, when I run the model, I was looking for "Asian restaurants" but there are very few results (maybe due to Foursquare categorization) so I looked for the restaurants closest to Indian cuisine taste

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 3 clusters based on how many Indian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with no Indian restaurants
- Cluster 1: Neighborhoods with little or 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 the Indian Restaurants are in cluster 2 which is around The Annex, North Midtown, Yorkville and in cluster 1 around The Danforth West, Riverdale

And lowest (close to Zero) in Cluster 0 areas which are around Commerce Court, Victoria Hotel and also there are good opportunities to open near Harbourfront, Regent Park, Adelaide, King, Richmond.

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 a 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.

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.

References

List of neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Foursquare Developer Documentation: <https://developer.foursquare.com/docs>