

Indian Sweets Shops in Toronto

Data Science Capstone Project **Final Report**

Introduction

- A lot of Indians live in Canada specially in Toronto. Sweets are favorites of Indians, they can eat sweets anytime and throughout the year. We have occasions when we celebrate by exchanging sweets, we eat sweets after meals to elevate our mood and we serve sweets to guests.
- It's something which has demand throughout the year and is a multi-million dollar business.
- Having an Indian sweets shop nearby an Indian population / restaurant's cluster can be beneficial for owner as well as approachable for customers. Customers passing by can always buy sweets to satisfy their craving for sweets.
- To succeed with the business, one must make sure that the shop is not too far from customer's reach, where one can quickly reach and buy sweets.

Business Problem

- The main objective is to find ideal spots in the city of Toronto where Indian sweets shop chain can be put up aiming at maximum profits.

Data

To solve the problem, we need the following data:

- List of Neighborhoods in [Toronto, Canada](#).
- Latitude and Longitude of this Neighborhood.
- Venue data related to Indian restaurants, which will be obtained using FourSquare API. This will help us find the neighborhoods that are most suitable to open an Indian sweets shop.

Target Audience

The entrepreneur who wants to find the location to open a profitable Indian Sweets chain in Toronto.

Methodology

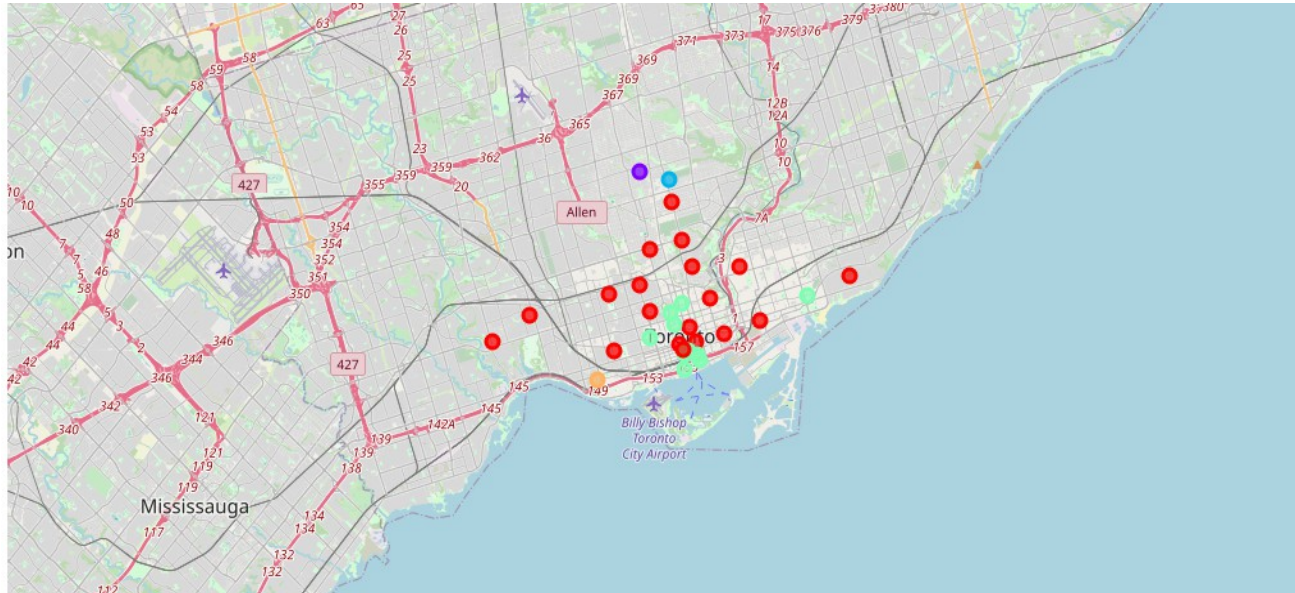
First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from [Wikipedia page](#). I did the web scraping by utilizing the BeautifulSoup library as it is easier and more convenient to pull tabular data directly from a web page into a data frame. 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, either I can use the Geocoder or simply use the CSV file provided by Coursera itself. After gathering all these coordinates, I visualized the map of Toronto using the Folium package to verify whether these are correct coordinates. Next, I use Foursquare API to pull the list of top 100 venues within 900 meters radius. I have created a Foursquare developer account earlier in order to obtain an 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” because I am looking for clusters of restaurants serving Indian cuisine primarily. 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 5 clusters based on their frequency of occurrence for “Indian food”. Based on the results (the concentration of clusters), I can able to recommend the ideal location to open the restaurant.

Results

Map of Toronto



Clusters



The results from k-means clustering show that we can categorize Toronto neighborhoods into 5 clusters based on how many Indian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with little Indian restaurants
- Cluster 1: Neighborhoods with no Indian restaurants.
- Cluster 2: Neighborhoods with little or no Indian restaurants.
- Cluster 3: Neighborhoods with high Indian restaurants.
- Cluster 4: Neighborhoods with little Indian 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 green color and Cluster 4 in lemon yellow color.

Observations And Recommendation

Most of the Indian restaurants are in Cluster 3 and Cluster 0 areas which are Downtown Toronto, East, West. Also, there are good opportunities to open an Indian sweets shop near these areas because of the presence of prospective customers going to restaurants in these areas. Looking at nearby venues, it seems Cluster 0 and cluster 3 might be a good location as there are a lot of Asian restaurants in these areas. Therefore, this project recommends the entrepreneur to open a reliable Indian sweets shops in these locations. Nonetheless, if the sweets is Delicious, reliable, and affordable, I am confident that it will have great demand 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 a recommendation to the stakeholders.