

Data Science Project Week 5th

INTRODUCTION

This is IBM Data Science Professional Certificate capstone project. In this project, I am creating a hypothetical scenario for a concept that there may not be enough Restaurants in Toronto Area for a particular cuisine. Therefore it might be a great opportunity for an entrepreneur who is based in Canada. As the particular cuisine that has less number of restaurants is popular among particular community, so the entrepreneur might think of opening its business in areas where the particular community resides. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

BUSINESS PROBLEM:

The objective of this capstone project is to find the most suitable location for the entrepreneur to open Restaurant in that cuisine with less number of restaurants in Toronto, Canada. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question : In Toronto, if an entrepreneur wants to open an a Restaurant, where should they consider opening it?

TARGET AUDIENCE:

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

DATA:

To solve this problem, we will need below data:

List of neighborhoods in Toronto, Canada

Latitude and Longitude of these neighborhoods

Venue data to find which cuisine restaurant is less available. This will help us find the neighborhoods that are more suitable to open the particular restaurant.

EXTRACTING THE DATA:

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

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

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 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 Foursquare developer 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 specific cuisine like “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.

RESULT

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 the less number of Indian restaurants.
- Cluster 1: Neighborhoods with no Indian restaurants.
- Cluster 2: Neighborhoods with a more number of Indian restaurants

The results are visualized in the above map with Cluster 0 in green, Cluster 1 in blue, Cluster 2 in red.

RECOMMENDATIONS

Most of the Indian restaurants are in cluster 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.