IBM Applied Data Science Capstone

OPENING A NEW AMERICAN RESTAURANT IN TORONTO

Abstract

Neighborhood in Toronto have been analyzed to find a potential location for opening of a new American restaurant in Toronto.

By: Arpan Dutta

Contents

Introduction

The project is for IBM Data Science Professional Certificate. The project assumes that there is a demand for American restaurants in Toronto. An entrepreneur wants to leverage the above referred demand for American food in Toronto and want to open an American restaurant there. The project is about finding potential location for opening an American restaurant in an area where currently less or no American restaurant is there.

Business Problem

The objective of the project is to analyze and find most suitable location for opening of an American restaurant in Toronto. We need to take help of data science knowledge in analyzing the dataset using machine learning algorithm such as clustering. In our analysis we would try to find currently in which areas less numbers of American restaurants are there and where an entrepreneur should consider opening of an American restaurant.

Target Audience

The project is being carried out for helping an entrepreneur in his problem to find the best location to open an American restaurant in Toronto.

Data

To solve the above described problem, we need the following data:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of those neighborhoods
- Venue data of American restaurants in those neighborhoods. We will cluster the neighborhoods on the basis of American restaurants in those areas.

Data Source and Extraction methods

- Wikipedia page containing list of neighborhoods of Toronto – Web scrapping technique to extract data from Wikipedia page, Pyhton requests and beautifulsoup packages were used.
- Geographical Coordinates of these neighborhoods – Python Geocoder package can be used for getting Latitudes and Longitudes but I used CSV file provided by Coursera team (Geospatial_Coordinates.csv)
- Venues data for these neighborhoods – Foursquare API was used to get this data.
- Map of Toronto – Python Folium package was used for this.

Methodology

We get the list of neighborhoods in Toronto from Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). We perform web scrapping to extract the list of neighborhoods from the Wikipedia page. We take help of Python requests and beautifulsoup packages in performing this task. We preview the imported dataset and confirm that It contains list of neighborhood.

Now, we need to get geographical coordinates for these locations. We read the csv file provided by Coursera (Geospatial_Coordinates.csv) for the coordinates. We could have used Python Geocoder package for this but we used already available data for simplicity. We merge the above two datasets.

Next, we need to get venues data for these neighborhoods. So, we take help of Foursquare API to get this data. We have to register for a developer account with Foursquare to use this API service. We can register for free account for this project. Post registration, we need to generate Foursquare ID and secret key (create an app and then generate on their website). We use these credentials to call the API in our Python code. Foursquare returns the dataset in JSON file. We need to normalize the JSON file and extract the venue data from it into dataframe.

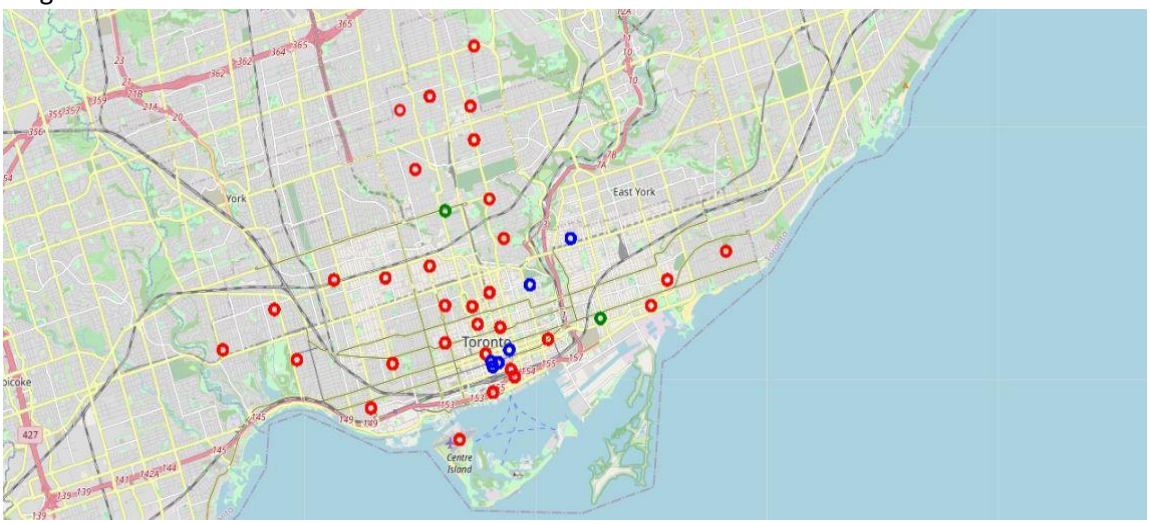
Then, we group rows of dataset by taking mean of frequency of each venue category. We prepare the dataset for clustering by keeping only the required data in the dataframe.

Lastly, we need to perform clustering on the dataset and analyze the thus obtained datasets all neighborhood.

We analyze the clusters and evaluate which is the best location to open an American restaurant in it on the basis of number of already running American restaurants in those areas.

Results and Discussion

- As we can see that maximum number of American restaurants are in Cluster 1.
  Mostly, in Downtown Toronto maximum number of American restaurants are there.
- In Commerce Court, Victoria Hotel, St. James Town, First Canadian Place, Underground city neighborhoods maximum numbers are there.



- If we assume that Downtown Toronto is suitable for American restaurants in view of customer's eating habits, then certainly we can look in cluster 1 wherein the number of American restaurants are not that much.

Conclusion

- So, we have clear potential to open American restaurant in
  - Downtown Toronto in these locations– Cluster 0:
    - Richmond,

- Adelaide,
- King,
- Stn A PO Boxes
- Church and Wellesley.

❖ In East Toronto and Central Toronto in these locations– Cluster 2:
- Studio District,
- Summerhill West,
- Rathnelly,
- South Hill,
- Forest also there is potential to open American restaurant

Appendix:

Notebook Link: https://github.com/ARPAN2405/Coursera_Capstone/blob/master/Week_5.ipynb