Coursera Capstone Project

Opening New Indian Restaurant in Toronto City, Canada

By

Santosh Sawant

06.08.2020

Introduction

- Capital city of the Canadian province of Ontario
- Most Populous and diverse culture city
- Most favored city by immigrants
- Major economic activities are executed in Toronto
- Higher population from India after China in immigration count

Business Problem

“Using data science methodology and machine learning tools, help investors to select the best locations in the city of Toronto, Canada to open a new Indian Restaurant”

Data and Source of Data

- List of neighborhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto in Canada.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to Indian restaurants. We will use this data to perform clustering on the neighborhoods.

- ***Source: Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)***

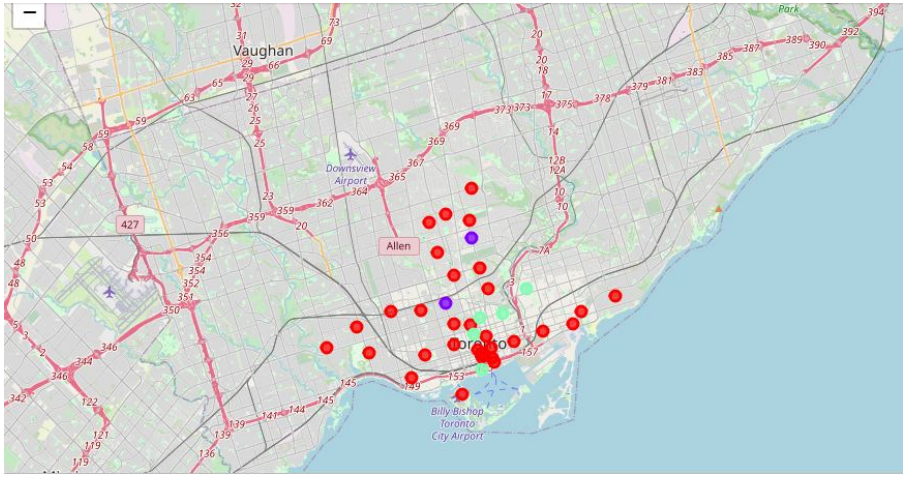
Methodology

- Using web scraping using Python requests and beautifulsoup packages to extract the list of neighborhoods data.
- Geographical coordinates in the form of latitude and longitude using Geocoder in order to be able to use Foursquare API.
- Populate the data into a pandas Data Frame and then visualize the neighborhoods in a map using Folium package.
- Using Foursquare API to get the top 100 venues that are within a radius of 2000 meters.
- Analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. (Focusing Indian restaurant category)
- K-means clustering algorithm to identify cluster based on occurrence of Indian restaurants at neighborhoods.

Results

Before Clustering

After Clustering



- Cluster 0: Neighborhoods with Zero number of Indian restaurants
- Cluster 1: Neighborhoods with moderate number of Indian restaurants
- Cluster 2: Neighborhoods with high concentration of Indian restaurants

Conclusion

The neighborhoods in cluster 0 are the most preferred locations to open a new Indian Restaurant. Further decision to identify location in cluster 0 neighborhoods may be based on other factors such as population density of Asian community / income levels of population.

The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding higher competition as in cluster 2.

Thank You