IBM Data Science Project

Harry

Table of Contents

INTRODUCTION

This project will create a hypothetical scenario for a concept that there may not be enough Chinese Restaurants in Toronto Area. Therefore, it might be a great opportunity for an entrepreneur who is based in Canada. As the Chinese food is popular among Asian community, so this entrepreneur might think of opening its business in areas where Asian 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 the project will help entrepreneurs 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 a new Chinese Restaurant 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 local business owner wants to open an Chinese Restaurant, where should they consider opening it?

TARGET AUDIENCE

Local business owner who wants to find the location to open authentic Chinese 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 related to Chinese restaurants. it will help us find the eighborhoods that are more suitable to open an Chinese Restaurant.

EXTRACTING THE DATA

- Scrapping of Toronto neighborhoods via Wikipedia

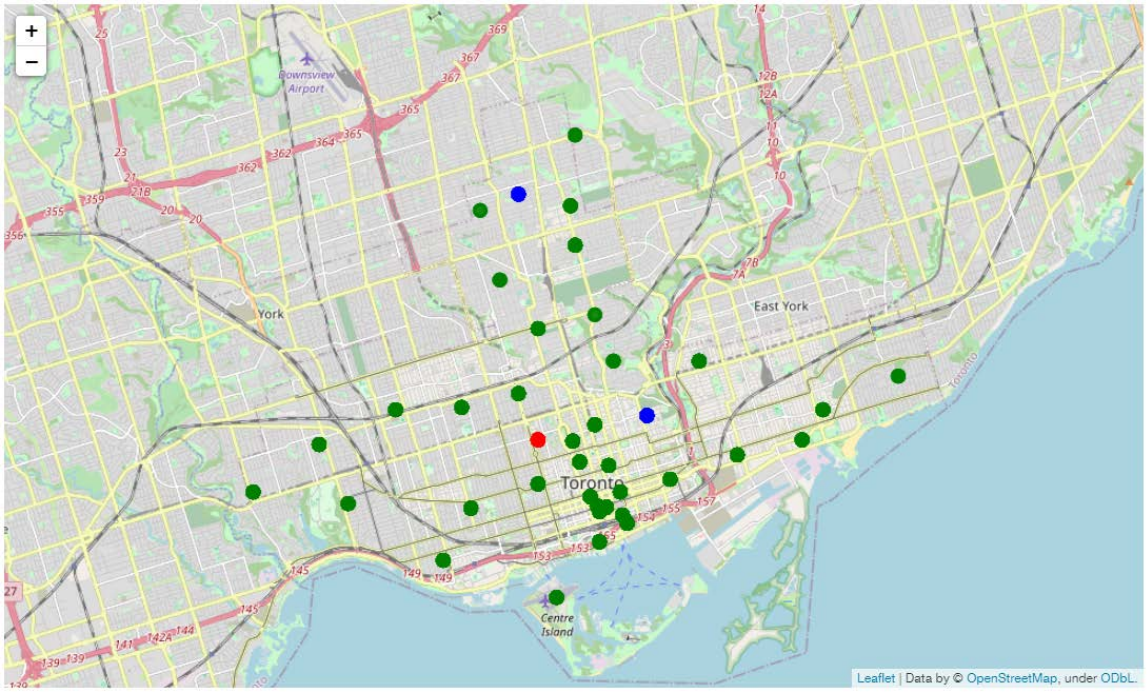
● Getting Latitude and Longitude data of these neighborhoods via Geocoder package
● Using Foursquare API to get venue data related to these neighborhoods

METHODOLOGY

This project requires the whole list of neighborhoods in Toronto and the location of Chinese restaurant. First of all, the team extracted the list of neighborhoods from Wikipedia: List of postal codes of Canada by utilizing Pandas model in Python. It is not only easier to use but also it is convenient to store tabular data from the website into the data frame.

The coordinates of each neighborhood have been generated by using CSV file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering these coordinates, the map of Toronto has been generated using Folium package to verify whether these are correct coordinates. Next, used Foursquare API to pull the list of top 100 venues within 500 meters radius. From Foursquare, the project team able to pull the names, categories, latitude, and longitude of the venues. With this data, total amount of unique categories from these venues has been imported to the python script. By analyzing each neighborhood by grouping the rows by neighborhood, taking the mean on the frequency of occurrence of each venue category, total number of Chinese restaurants in each neighborhood has been visualised in table. 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. Project clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for Chinese restaurant. Based on the the concentration of clusters, the project 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 Chinese restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with no Chinese restaurants.
- Cluster 1: Neighborhoods with the less number of Chinese restaurants.
- Cluster 2: Neighborhoods with a more number of Chinese restaurants

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

RECOMMENDATIONS

Toronto Dominion Centre, Design Exchange, Garden District, Ryerson, Harbourfront East, Union Station, Toronto Islands are the great place to open the new Chinese Restaurant. Looking at nearby venues it seems cluster 0 might be a good location as there are not a lot of Chinese restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Chinese restaurant in these locations.