IBM Data Science Professional Certificate

Final Report

Author
Iraklis Kalamas

Title of the project

Suitable places for a Greek Restaurant in Toronto
(a fictional scenario)

Introduction

The union of restaurants in Greece contact my company (Data Analytics SA - a data analysis company) to produce a report in order to find if there are opportunities for opening a restaurant with Greek cuisine, seventh in popularity according to the CNN travel (https://edition.cnn.com/travel/article/world-best-food-cultures/index.html), in Toronto. The union wants to help new entrepreneurs (or established ones) for having the best possible facts before the decide to take that step.

This is a part of a multi country analysis for the suitable places of opening a restaurant and the union decided to use Toronto as the start point based on the strong Greek community that is established in various cities of Canada.

Business Problem

The objective of the project is to find the most suitable place(s) in order for someone to open a Greek cuisine restaurant in Toronto, Canada. To answer this question our company used data science methods and tools (such as machine learning algorithms). Although the question asked from the union is very broad and accepts many interpretations as to what characteristics has to be met for accept if a place is suitable, our company, at this stage, will produce a fairly objective and round report used as a starting point for a detailed report, and deliverable, at a second stage.

Needed Data for the Analysis

In order to produce the report we need data that will help us to answer the main question and at this stage they will be very basic such as a list of neighborhoods in Toronto, Canada along with their latitude and longitude and some venue data related to Greek restaurants. This will help us find the neighborhoods that are more suitable to open Greek one.

EXTRACTING THE DATA

The methods used for acquiring the need data included:

- 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

Since in the report have to be included areas of Toronto that one can open a restaurant we need to get the list of neighborhoods. This is possible by extracting the list of neighborhoods from Wikipedia:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The web scraping is done by utilizing pandas html table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame.

Having done that we end up with a list of neighborhood names and postal codes which we joined thme with their corresponding coordinates using the http://cocl.us/Geospatial_data service (a .csv file essentially). After that we used Foursquare to pull the list of venues near these neighborhoods. Visualizing the map of Toronto using Folium package was the next step in order to verify whether these are correct coordinates.

In order to extract information of the area of interest, and in order to have some constraints, we used Foursquare API to pull the list of top 100 venues within 500 meters radius. We have created a Foursquare developer account in order to obtain account ID and API key to pull the data. After that we could view several aspects of the data such as how many unique categories there are in the venues or how many of those are restaurants.

We analyzed 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.

All the work is done under the consideration of searching for "Greek Restaurants" in the area.

We used the K-means clustering algorithm in order to separate areas with Greek Restaurants.

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.

We have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for "Greek food". Based on the results (the concentration of clusters), we were able to recommend the ideal location to open a restaurant.

Results

Most of the Greek restaurants are in cluster 1 which is found in the areas of The Danforth West and Riverdale.

The clusters 0 and 2 seems to have the same amount of Greek Restaurants and are located in the areas of Berczy Park, Little Portugal, Trinity, Davisville, Toronto Dominion Centre, Design Exchange, First Canadian Place and Underground city.

Although more detail is needed, such as distance of location from other venues and transportation means or GDP of residents per neighborhood, the recommendation to the union is that places around clusters 0 and 2 are more suitable to open an authentic Greek restaurant.