

Use data analysis and machine learning to help customers choose the best location to open a Chinese restaurant

Introduction

The history of the country has its strengths and weaknesses, its territory has its strengths and weaknesses, its strengths are strong and its weaknesses, its populations are large and small, and its ethnic composition, religious beliefs, political power, and economic structure are also different. Therefore, the food culture in different countries is different.

From the perspective of history, the Chinese food culture has continued for tens of thousands of years (the discovery of Hunan pottery and rice varieties), divided into four stages of development: raw food, cooked food, natural cooking, and scientific cooking, with more than 60,000 traditional dishes and more than 20,000 industries Food, colorful feasts and colorful flavors.

In terms of connotation, Chinese food culture involves the development and use of food sources, the use and innovation of food utensils, the production and consumption of food, the service and reception of food and beverage, the operation and management of food and beverage and food industry, and food and beverage , The relationship between diet and literature and art, diet and the realm of life, etc., is profound and extensive.

From the perspective of extension, Chinese food culture can be classified from various angles such as time and technology, region and economy, ethnicity and religion, food and utensils, consumption and level, folklore and function, showing different cultural tastes and reflecting different The use value of the product is colorful.

From the perspective of characteristics, the Chinese food culture emphasizes the camp-and-health theory of supporting and benefiting (vegetarian-based, with emphasis on medicinal diet and supplementation), and pays attention to the "color, fragrance and taste". The five-harmonic realm theory (distinct flavor, palatable ones, known as "tongue dish"), bizarre and varied cooking methods (kitchen rules-based, flexible), and a delightful view of cuisine (literate, elegant "Teaching to food" and other four attributes have a natural beauty that is different from the food culture of overseas countries. In addition to paying attention to the bright and picturesque colors of the dishes, the Chinese food culture also needs to match with the dining atmosphere. It is the personality and tradition of the Chinese nation, and it is also a highlight of the traditional etiquette of the Chinese nation.

In terms of impact, Chinese food culture directly affects Japan, Mongolia, North Korea, South Korea, Thailand, Singapore and other countries, and is the axis of the eastern food culture circle; at the same time, it also indirectly affects Europe, America, Africa and Oceania. For example, China' s vegetarian

culture, tea culture, sauce and vinegar, pasta, medicinal dishes, ceramic tableware and soybeans have benefited billions of people around the world.

In short, Chinese food culture is a long-term regional culture with wide vision, deep level, multi-angles, and high taste; it is the people of all ethnic groups in China in more than 5,000 years of production and life practice, in food source development, tableware development, food conditioning , Nutrition and health care and diet aesthetics create, accumulate and influence the material wealth and spiritual wealth of neighboring countries and the world.



Business Problem

The goal of the project is to find the ideal place to open a Chinese restaurant in Toronto, Canada. Although there are many Chinese in Toronto, there are not many Chinese restaurants, but I am looking for a suitable place to open a Chinese restaurant. My purpose is to find the answer to my business question "Where should I open a Chinese restaurant?" In this peak project, with the help of data science technology and machine learning methods (such as clusters), we intend to find ways to solve business problems.

Target audience

Our target audience is all Chinese families or entrepreneurs who want to open or expand Chinese restaurants in good places, because this is not only good for entrepreneurs, but also for full-time parents or partners who are not working but still working Opportunity to find a way to make money.

Data Description

To solve this problem, I would be requiring the following data:

1. List of neighborhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto, Canada.
2. Coordinates of these neighborhoods. This is required in order to plot the map and also to get the venue data.
3. Venue data related to our business. We will use this data to perform clustering on the neighbourhoods.

Sources of Data

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) contains a list of neighbourhoods in Toronto Canada with a total of 103 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page. Then we will get the geographical coordinates of the neighbourhoods to get the latitudes and longitudes of the neighbourhood.

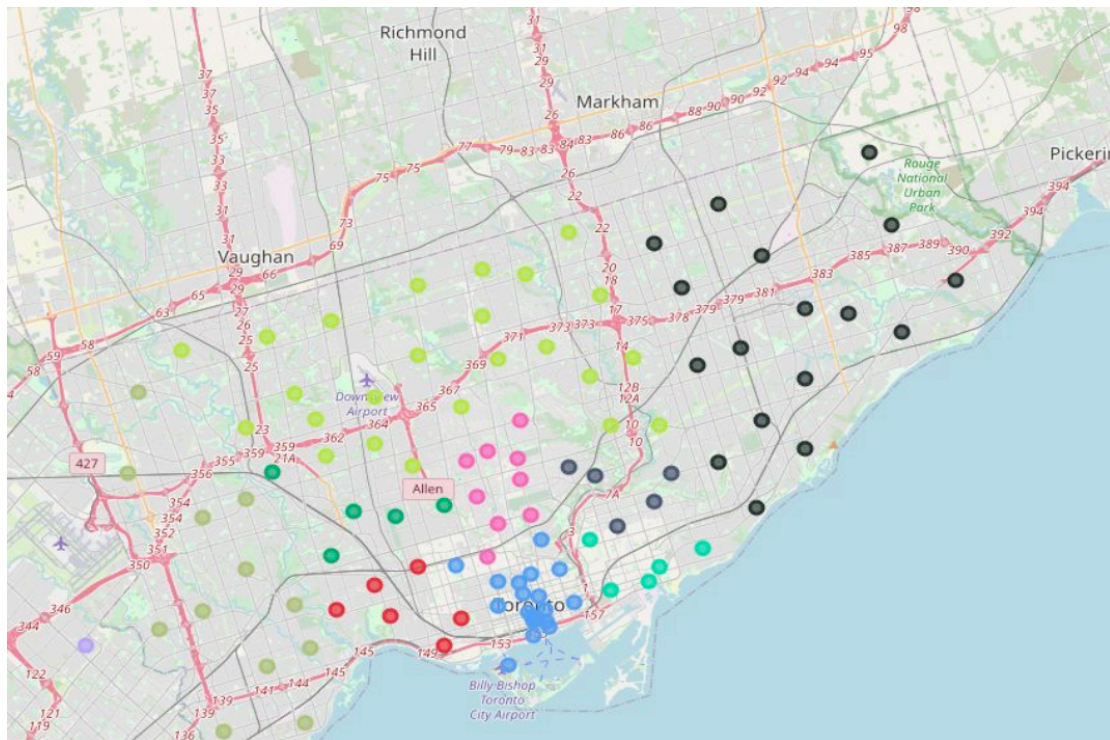
This project makes use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

This project will also use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Methodology

First, I need to get the list of neighbourhoods in Toronto, Canada. This is possible by extracting the list of neighbourhoods from Wikipedia page that is mentioned in the source of data section of this report. I performed the web scraping by utilizing pandas html table scraping technique as it is easier and more convenient to pull tabular data directly from a web page into dataframe.

However, it is only a list of neighbourhood names and postal codes. I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighbourhoods. To get the coordinates, I used the csv file provided by IBM team to match the coordinates of Toronto neighbourhoods. After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

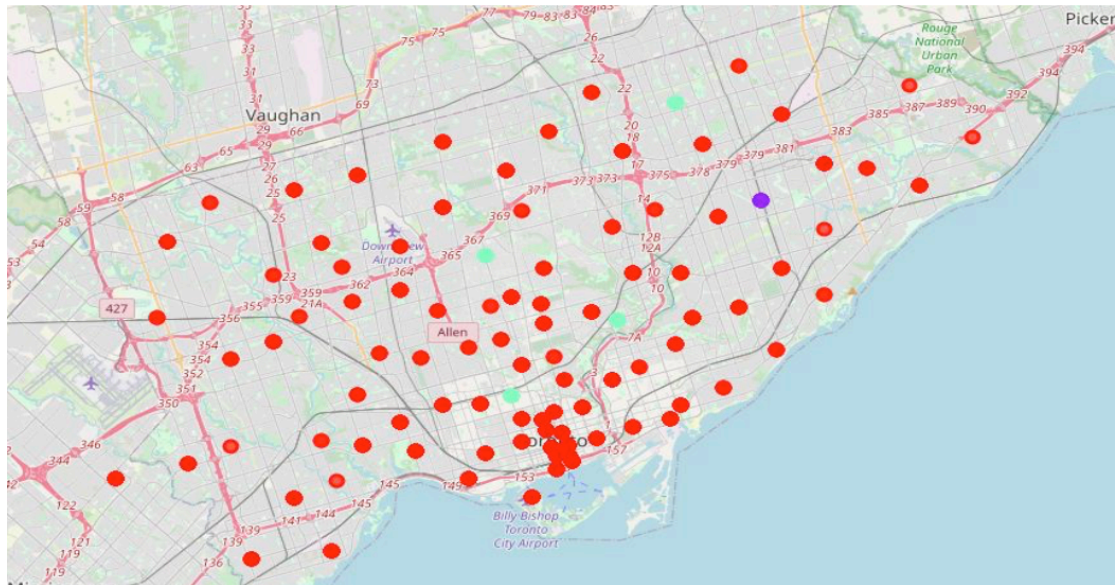


After that, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analysed each neighbourhood by grouping the rows by neighbourhood 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 “Chinese 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 neighbourhoods in Toronto into 3 clusters based on their frequency of occurrence for “Chinese food” .

With the help of the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.



Results

The results from k-means clustering show that we can categorize Toronto neighbourhoods into 3 clusters based on how many Chinese restaurants are in each neighbourhood:

- Cluster 0: Neighbourhoods with high number of Chinese restaurants
- Cluster 1: Neighbourhoods with little to no Chinese restaurants
- Cluster 2: Neighbourhoods with moderate number of Chinese restaurants

The results are visualized in the above map with Cluster 0 in red colour, Cluster 1 in purple colour and Cluster 2 in mintgreen colour.

Recommendations

Most Indian restaurants are in Group 0, located in the Adelaide, King, and Richmond areas, while the lowest (nearly zero) area in Group 1 is Dorset Park in Wexford Heights . In addition, since competition seems to be low, there seems to be a good opportunity to place Chinese restaurants near East Bedford Park in Lawrence Manor in North York, Indiana or near North Yorkshire. Nice location because there are not many Chinese restaurants in these areas. Therefore, the project recommends that entrepreneurs set up an authentic Chinese restaurant in these places with little competition. However, if the food is authentic, affordable and delicious, then the restaurant will become famous anywhere.

Conclusion

In this project, we experienced identifying business problems, specifying required data, extracting and preparing data, performing machine learning by clustering the data into 3 clusters based on similarity, and finally providing relevant stakeholders with Answering the business questions raised in the introduction, the project' s answer is: the community in merger 1 is the best place to set up a new Chinese restaurant, because there are some results of the project that will help entrepreneurs take advantage of opportunities to potentially high Location, while avoiding overcrowding in their decision to set up a Chinese restaurant.