Coursera Capstone

Opening a New Asian Restaurant in Ho Chi Minh City, VietNam

By: Tran Hong Vinh

Jun 2020

Introduction

Today, Vietnamese often eat at the restaurant. Life is so busy so it makes cooking at home and cleaning becomes a burden for many people, especially in big cities. So breakfast which is easy to buy at convenience stores or restaurants such as bread, sticky rice, noodles, wet bread ... is preferred. The Officers usually have more time so that they choose to have breakfast with friends at the cafe. Lunch time is short, only about 1 hour, so most people choose to order office lunch or invite each other to the restaurants. Only dinner or weekends left to spend time cooking family meals.

The trend of eating out has led to the increase of a variety of dining types such as restaurants and bars from luxurious to popular to serve different audiences. Korean and Japanese restaurants have poured into Vietnam market and despite the rise for nearly ten years, until now, it has attracted a lot of customers. Well-known brands are always in the right place at all times.

Opening asian restaurant allows to earn income. Of course, as with any business decision, opening a new asian restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the asian restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the Ho Chí Minh city, VietNam to open a new asian reataurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the Ho Chí Minh city, VietNam, if the businessman is looking to open a new asian restaurant, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to the businessman and investors looking to open or invest in new asian restaurant in the Ho Chí Minh city, VietNam. This project is timely as the city is currently suffering from oversupply of asian restaurant. According to the General Statistics Office just published, in the first month of 2020, accommodation and catering services revenue is estimated at over 45,000 billion VND, up 0.5% from the previous month and up 14.7%. over the same period in 2019.

Data

To solve the problem, we will need the following data:

• List of neighbourhoods in Ho Chi Minh city. This defines the scope of this project which is confined to the Ho Chi Minh city.

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to asian restaurant. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_districts_of_Vietnam#H%E1%BB%93_Ch%C3%AD_Minh_City)contains a list of neighbourhoods in Ho Chi Minh city, with a total of 24 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make 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). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.