Applied Data Science Capstone

Capstone Project

Opening a Japanese restaurant in Bangkok

Introduction

Thailand is now a very popular tourist spot for the travellers around the globe because of food, shopping malls, markets and outdoor activities. Most people will go to Bangkok in their first Thai experience due to its convenience. Also, ASEAN countries are growing quickly due to external environments. Because of the underlying economic growth and cultural exchange, Thai would start to experience and explore other countries' culture as well. Moreover, more Japanese nationals are moving to Thailand according to Embassy of Japan in Thailand. Opening a restaurant providing Japanese cuisine is a great chance to start or expand the catering business.

Business Problem

The aim of this project is to help businessmen find the best place to open a Japanese restaurant in Bangkok in which popularity and competitiveness are considered concurrently. Locating in a very popular district faces higher competitiveness. In contrast, opening it far away from popular district faces lower return. Therefore, clustering analysis, one of the most popular machine learning methodologies, will be adopted to let them know where has higher business opportunity and lower risk.

Targeted audience

The targeted audience is for whom looking forward to opening a Japanese restaurant. As it involves massive fixed cost and investment, the project can help them start or expand their Japanese catering business in order to seek for the Thai and Japanese customers to minimize their risk and maximize their return.

Data

To solve the problem, the following data is required

- Bangkok's districts
- Districts' latitude and longitude
- Venue data around each district

The sources of data are from Wikipedia and Foursquare. List of districts of Bangkok and districts' latitude and longitude can be found in Wikipedia (https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok). It contains 50 districts with corresponding latitude and longitude. Scraping is used to extract the data from Wikipedia via Python making use of "request" and "Beautiful Soup". However, not all the districts are equipped with longitude and longitude data. Those districts will be removed in this analysis during data cleansing. Then, Foursquare location API is used to explore the venue data near to each district. After clustering analysis, the data should provide us with what are popular venues in each cluster. Finally the result is visualized and analyzed to solve the business problem.

To sum up, the analysis will be done from data scraping from Wikipedia, calling Foursquare location APIs, data cleansing, data wrangling, k-means clustering analysis and data visualization. Finally, the result should be capable to provide an insight for whom they want to open a Japanese restaurant in Bangkok.