Coursera Capstone Project

IBM Applied Data Science Capstone

Analyze the restaurants in Ayutthaya province to invite foreign tourists

By: Chayapol Mahatthanachai

7 April 2020

Introduction

For many Foreign Tourist who travel in Ayutthaya province at Thailand. The first thing that they will think about is popular temple, Grocery shopping or absolutely Coffee Shop. But one thing that they cannot forget is the restaurant. Ayutthaya Province not only famous on tourist attraction. But many Foreign Tourist also advert on Ayutthaya restaurant because many restaurant has originating cooking from antiquity period.

From according above. It became the idea about how can we invite foreign tourist who travel around the most popular restaurant nearby that district. So we provide a group of restaurant in Ayutthaya to analyze how distribution of neighborhood would be better. As a result, there are more popular restaurant nearby the central of Ayutthaya Province: Phra Nakhon Si Ayutthaya District. Opening the restaurant in central of Phra Nakhon Si Ayutthaya District could be invite Foreign Tourist and Thai people so much. Particularly, the location of restaurant is one of the most important decisions that will determine the advert of tourist.

Business Problem

The objective of this capstone project is to analyze the most popular restaurant to invite foreign tourist who travel in Ayutthaya, Thailand by using data science methodology and machine learning technique such as loading the Dataset, Scrapping Website or Clustering group of data that is unsupervised learning. This project aims to provide the solution for instructive entrepreneur who give emphasize only Phra Nakhon Si Ayutthaya district that was the most famous district in Ayutthaya. If the foreign tourism need to find which neighborhood has most popular restaurant, how did you invite those tourism to find out the most popular restaurant in that neighborhood?

Target Audience of this project

This project is particularly useful to foreign tourist who looking for the restaurant in Ayutthaya province. As we know that not only Thai people who has trip or visit in Ayutthaya. But another foreign tourist like Europe continent also has vacation too. The data from prachachat local economy website release in 2018 said 'The amount of revenue will increase 1.5 billion baht and the tourist amount also increase around 8 million people. Forecast in 2019 to 2020 the tourism in Thailand especially Ayutthaya province will increase 8-10% too.

Data

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

- List of neighborhoods in Ayutthaya Province. This defines the scopes of project which is confined the restaurant in Ayutthaya, the popular district on tourism of Thailand
- Latitude and Longitude coordinates of neighborhoods. This required in order to plot the map and get the venue data.
- Venue data, we use this data to clustering on the neighborhoods.
- Geospatial map to coordinate the restaurant in Ayutthaya district.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Phra_Nakhon_Si_Ayutthaya_Province) contains a list of district (neighborhoods) in Ayutthaya province, with a total of 70

neighborhoods. 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 neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare API will provide many categories of the venue data, we are particularly interested in the restaurant category in order to invite foreign tourist visited. 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).

Methodology

Firstly, we need to get the list of district (neighborhoods) in Ayutthaya. The district could be available in (https://en.wikipedia.org/wiki/Phra_Nakhon_Si_Ayutthaya_Province). We will do web scraping using Python requests and beautifulsoup package to extract the list of neighborhood data. And then we need to get the geographical coordinates in form of latitude and longitude in order to able by using Foursquare API. We will Geocoder package that will convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into pandas DataFrame and then visualize the neighborhoods in a map using Folium package.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 1000 meters. We need to call Foursquare passing in the geographical coordinates of the neighborhoods in python loop. Foursquare will return venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were return on each neighborhood and examine how many

unique categories can be curates from all the returned venues. Then, we will analyses each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. We are also preparing the data for use in clustering. Since we are analyzing the “Restaurant” data, we will filter the “Restaurant” as venue category for the neighborhoods.

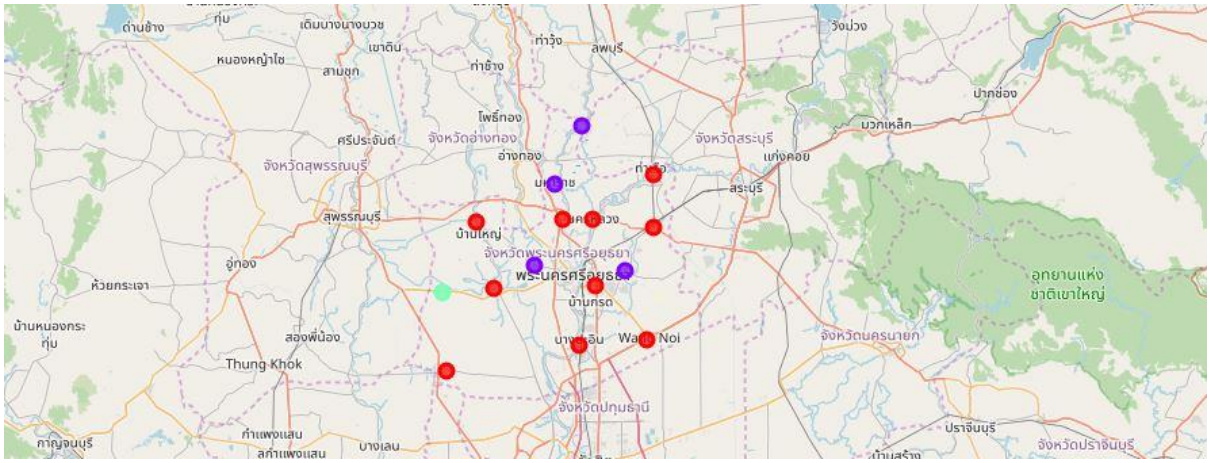
Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroid. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 2 clusters based on their frequency of occurrence for “Restaurant”. The results will allow us to identify which neighborhoods have higher frequency of restaurant while which neighborhoods have fewer number of restaurants. Based on the occurrence of restaurant in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to invite foreign tourist visit.

Results

The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Restaurant”:

- Cluster 0: Neighborhoods with high concentration of Restaurant
- Cluster 1: Neighborhoods with moderate number of Restaurant
- Cluster 2: Neighborhoods with low number of Restaurant

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in mint green color.



Visualized the map that clustering restaurant in Ayutthaya province

Discussion

As observations noted from the map in the Results section, most of the popular restaurant are concentrated in the central area of Ayutthaya district, with the highest number in cluster 0 and moderate number in cluster 1. On the other hand, cluster 2 has very low number with no restaurant in the neighborhoods (Bang Sai (1404), Bang Sai (1413)). This represents a great opportunity and high potential areas to invite foreign tourist to visit and enjoy by themselves. Meanwhile, shopping malls in cluster 1 are likely suffering from intense competition due to high amount percentage of tourism that travel in central area and high concentration of any kind restaurant. From another perspective, the results also show that the popular restaurant mostly happened in the central area of the city. Therefore, this project recommends property the foreign tourism to capitalize on these finding to visited popular restaurant in cluster 0. The foreign tourist can also visit in neighborhoods in cluster 1 with moderate number of restaurant too. Lastly, the foreign tourist should be avoid neighborhoods

in cluster 2 which already have low concentration and number of restaurant which the tourist may not pay attention at this neighborhoods (Bang Sai (1404), Bang Sai (1413)).

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the foreign tourist i.e. the tourist who has trip in Ayutthaya and need to visit on some restaurant. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 0 are the most preferred neighborhoods to invite the foreign tourist visited. The findings of this project will help the foreign tourist to find the popular restaurant around the Tourist attraction which consider in high concentration of clustering neighborhoods.

References

- Phra Nakhon Si Ayutthaya District in Ayutthaya province. Retrieved from

https://en.wikipedia.org/wiki/Phra_Nakhon_Si_Ayutthaya_Province

-Foursquare Developers Documentation . Retrieved from

https://developer.foursquare.com/docs/

-Scrapping Website using BeautifulSoup package. Retrieved from

https://www.crummy.com/software/BeautifulSoup/bs4/doc/

-Prachachat local economy. (2018, March 15). Ayutthaya travel grows over 15 billion per nights. Retrieved from https://www.prachachat.net/local-economy/news-130852