IBM Applied Data Science Capstone

Recommending a Business at a particular Tourism Site

Quan Kien Minh

Introduction

Tourism has always been a thriving sector across the world. No matter which country you are living in, you can always come across a group of people, big or small, who always like to visit attractions. I am a big fan of adventure, and I acknowledge this fact as to how tourism plays a salient role for a traveler/explorer. Tourism is not only an important aspect of a country's economy but also for its global standing.

Why Tourism is important to any country?

The tourism industry is important for the benefits it brings and due to its role as a commercial activity that creates demand and growth for many more industries. Tourism not only contributes to more economic activities but also generates more employment, revenues, and play a significant role in development.

- Tourism activity creates demand.
- Tourism industry value chain meets & spreads demand across industries & boosts more economic activities.
- Tourism induces more consumption.

Business Problem

All the benefits of tourism tend to reflect on the employment opportunity which it gives to the people of that country. The objective of this project is to analyze the tourist places of a given state in Vietnam, and try to recommend the best location where they can open a restaurant or lodging to make the best use of the opportunity.

The target audience for this project includes people who are interested in opening a restaurant, lodging, transport services, or any other similar businesses which fall within the tourism industry. This also recommends travelers' tourist venues to be visited in a given state of a country.

Data Anatomization

To tackle the above mentioned problem, we need to have the dataset that contains -

- All the provinces of a Vietnam.
- Latitude and longitudes of all the districts.

The major sources of data are derived from:

-
https://vi.wikipedia.org/wiki/Th%C3%A0nh_ph%E1%BB%91_H%E1%BB%93_Ch%C3%AD_Minh

- https://en.wikipedia.org/wiki/Hanoi

- https://en.wikipedia.org/wiki/Da_Nang

Those links obtain all the districts of Ho Chi Minh, Ha Noi, Da Nang municipalities. We then use beautifulsoup4 package, a Python module that helps to scrape information from the web pages to extract all the tables from this Wikipedia page and convert it into a pandas dataframe. Then we use Python's geopy package to obtain the latitude and longitude of all the districts present in the dataframe.

Description of the data

The output shows the final dataset. The dataset consists of a single Dataframe with 7 columns containing Municipalities, District/Municipal City, Area (km^2), Population (person), Wards, Latitude, Longitude.

	Municipalities	District/Municipal City	Area (km2)	Population (person)	Wards	Latitude	Longitude
0	Ho Chi Minh City	Thu Duc City	21156.0	1.013.795	34 wards	10.829830	106.761790
1	Ho Chi Minh City	District 1	772.0	142.625	10 wards	10.774845	106.699350
2	Ho Chi Minh City	District 3	492.0	190.375	12 wards	10.771551	106.698380
3	Ho Chi Minh City	District 4	418.0	175.329	13 wards	10.759243	106.704890
4	Ho Chi Minh City	District 5	427.0	159.073	14 wards	10.756129	106.670375