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 Wikipedia page https://vi.wikipedia.org/wiki/T%E1%BB%89nh_th%C3%A0nh_Vi%E1%BB%87t_Nam is the major source of data that is being used to obtain all the districts of India. 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. Besides, I will use FourSquare API to discover which businesses are doing in a particular places.

Description of the data

The output shows the final dataset. The dataset consists of a single Dataframe with 10 columns containing Province/City, Number on map, Population of the particular district etc.

Province/City	Number on map	Administrative center	Area (in km^2) [4]	Population [3]	Density (/km^2) [3][note 1]	% Urban [3]	HDI (2012) [5]	GDP per capita (2011 PPP US$)[5]	Region
Bắc Giang Province	20	Bắc Giang	3,895.59	1,803,950 [6]	463	11.4	0.711	1,909.44	Northeast
Bắc Kạn Province	14	Bắc Kạn	4,859.96	313,905 [7]	65	20.7	0.685	1,766.68	Northeast
Cao Bằng Province	10	Cao Bằng	6,700.26	530,341 [8]	79	23.3	0.653	1,564.27	Northeast
Hà Giang Province	9	Hà Giang	7,929.48	854,679 [9]	108	15.9	0.586	1,083.72	Northeast
Lạng Sơn Province	11	Lạng Sơn	8,310.09	781,655 [10]	94	20.4	0.707	2,201.98	Northeast
Phú Thọ Province	17	Việt Trì	3,534.56	1,463,726 [11]	414	18.1	0.715	1,916.23	Northeast
Quảng Ninh Province	21	Hạ Long	6,178.21	1,320,324 [12]	214	64.1	0.784	7,834.55	Northeast
Thái Nguyên Province	15	Thái Nguyên	3,526.64	1,286,751 [13]	365	31.9	0.741	2,547.11	Northeast
Tuyên Quang Province	13	Tuyên Quang	5,867.90	784,811[14]	134	13.8	0.699	1,948.18	Northeast
Lào Cai Province	8	Lào Cai	6,364.03	730,420 [15]	115	23.5	0.670	2,767.55	Northeast
Yên Bái Province	12	Yên Bái	6,887.46	821,030 [16]	119	19.8	0.657	1,845.51	Northeast