IBM Applied Data Science Capstone

Recommending a Business at a Tourist Venue

- *Rishabh Bhatt*

Introduction

Tourism has always been a booming industry across the globe. No matter which country you live in, you can always come across a group of people, big or small, who always like to visit places. Being an adventurer myself, I can 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.

i. Tourism activity creates demand.
ii. Tourism industry value chain meets & spreads demand across industries & boosts more economic activities.
iii. 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 India 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 –

i. All the districts of a particular country.
ii. Latitude and longitudes of all the districts.

The Wikipedia page https://en.wikipedia.org/wiki/List_of_districts_in_India 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.

Description of the data

The output shows the transition of data scraped to the final dataset. The final dataset consists of a single Dataframe with 9 columns containing state, districts, latitude and longitudes of the district etc. Other columns like code, headquarters, population, area and density have also been scraped from the website which can be used for further analysis.

	State	Total Districts	State Population	Population/District
0	Andhra Pradesh	13	49386799	3798985
1	Arunachal Pradesh	25	1383727	57656
2	Assam	34	31169272	944523
3	Bihar	38	104099452	2739459
4	Chhattisgarh	28	25545198	946118

(37, 4)

	State	District #	Code	District	HQ	District Population	Area	Density
0	Andhra Pradesh	1	AN	Anantapur	Anantapur	4083315	19130	213
1	Andhra Pradesh	2	CH	Chittoor	Chittoor	4170468	15152	275
2	Andhra Pradesh	3	EG	East Godavari	Kakinada	5151549	10807	477
3	Andhra Pradesh	4	GU	Guntur	Guntur	4889230	11391	429
4	Andhra Pradesh	5	CU	Kadapa	Kadapa	2884524	15359	188

(741, 8)

	State	District #	Code	District	HQ	District Population	Area	Density	Total Districts	State Population	Population/District
0	Andhra Pradesh	1	AN	Anantapur	Anantapur	4083315	19130	213	13	49386799	3798985
1	Andhra Pradesh	2	CH	Chittoor	Chittoor	4170468	15152	275	13	49386799	3798985
2	Andhra Pradesh	3	EG	East Godavari	Kakinada	5151549	10807	477	13	49386799	3798985
3	Andhra Pradesh	4	GU	Guntur	Guntur	4889230	11391	429	13	49386799	3798985
4	Andhra Pradesh	5	CU	Kadapa	Kadapa	2884524	15359	188	13	49386799	3798985

(741, 11)

Final Cleaned Dataset:

	State	Code	District	Headquarters	Population(2011)	Area(km2)	Density(/km2)	Latitude	Longitude
0	Andaman and Nicobar	NI	Nicobar	Car Nicobar	36842	1841.0	20	7.000000	93.000000
1	Andaman and Nicobar	NaN	North and Middle Andaman	Mayabunder	105597	3736.0	28	12.611239	92.831654
2	Andaman and Nicobar	SA	South Andaman	Port Blair	238142	2672.0	89	10.705690	92.487468
3	Andhra Pradesh	AN	Anantapur	Anantapur	4083315	19130.0	213	14.654623	77.556260
4	Andhra Pradesh	CH	Chittoor	Chittoor	4170468	15152.0	275	13.160105	79.155551