

IBM Applied Data Science

Capstone Project

Recommending a Business at a particular Tourist Venue

by César Fernandes

1. Introduction

Tourism has always been a booming sector 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. I being an adventurer myself 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.

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

2. 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

3. Data Anatomization

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

- All the districts of a particular country.
- 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 data frame. Then we use Python's geopy package to obtain the latitude and longitude of all the districts present in the data frame.

Description of the data

The output shows the final dataset. The dataset consists of a single data frame with 9 columns containing state, districts, latitude and longitudes of the particular 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 | 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 | 2872.0 | 89 | 10.705890 | 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 |