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.

- 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 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.

IMG- Indian Districts

Data

To tackle the business problem, we need to have the dataset that contains the following features –

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

The Wikipedia page – link 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.

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

Literature Review

There are specific factors within the characteristics of the population which makes the tourism industry lead to an improvement of the socio-economic conditions of the population. This will eventually result in low rates of unemployment and a higher percentage of the working population. The former improves the socioeconomic conditions of the population whereas the latter helps finance, through different tax burdens, public policies aimed at achieving a higher level of economic development. It also demonstrates that countries with regressive population pyramids have greater difficulties for tourism growth to improve their socio-economic conditions.

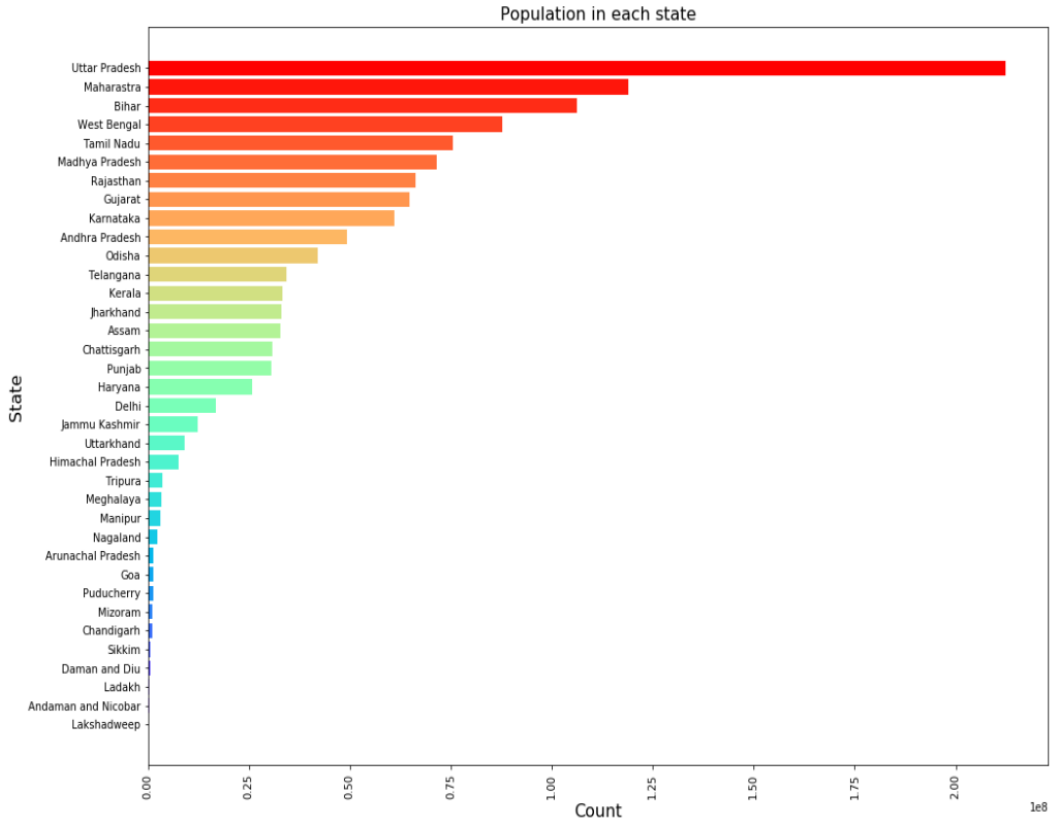
The survey from Annual Report Tourism of India provides us with the following facts –

- Tourism has contributed around 5.06% share in GDP during 2016-17
- There were 1854.93 million domestic tourist visits all over the country during the year 2018.
- Foreign Tourist Arrivals during 2019 were 10.89 million (Provisional) with a growth of 3.2% over the same period of the previous year
- Foreign Exchange Earnings during the period during Jan 2019 – Dec 2019 were Rs.2,10,981 crores (Provisional estimates) with a growth of 8.3% over the same period of the previous year.
- According to Tourist Satellite Account, the tourism industry has provided around 87.50 million people employment opportunities in the year 2018-19 vi.

The above-obtained statistics highlight the importance of the Tourism Industry in the overall development of the country.

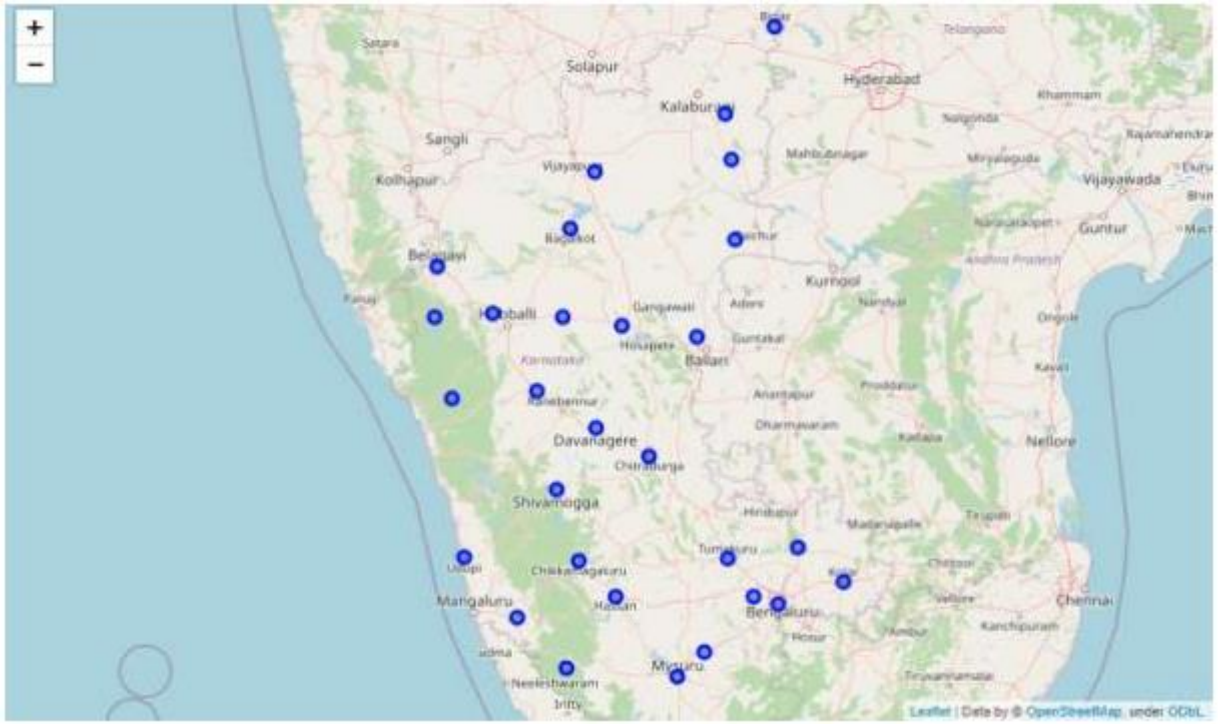
Methodology

- There are 36 states (including Union Territories) which have been retrieved from the webpage and stored in the dataset.
- As mentioned in the literature review, there can be some impacts of the population of a state on tourism. The below graph shows the population in each state.



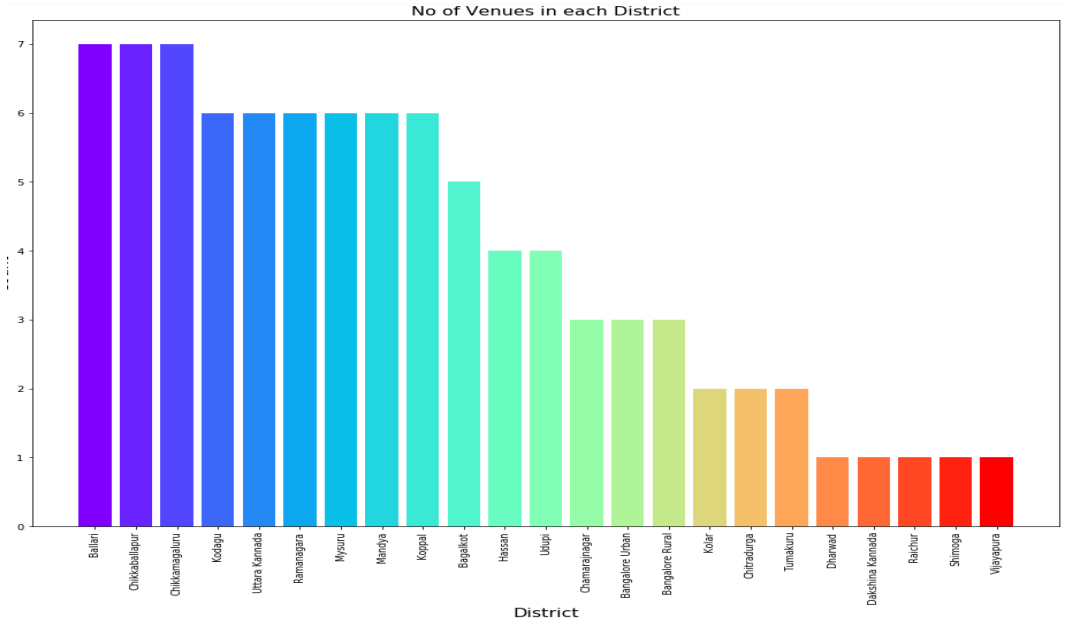
Methodology

The user can enter the state of his choice among the given states. Here Karnataka is taken as a choice. A visualization with all the districts of the given state will be displayed as shown below –



Methodology-2

The next step is to obtain the nearby tourist venues within a radius of 50km. This gives us multiple tourist spots if there are in a district. We visualize a bar graph by plotting District v/s count to obtain the number of venues in each district. The visualization can be shown below –

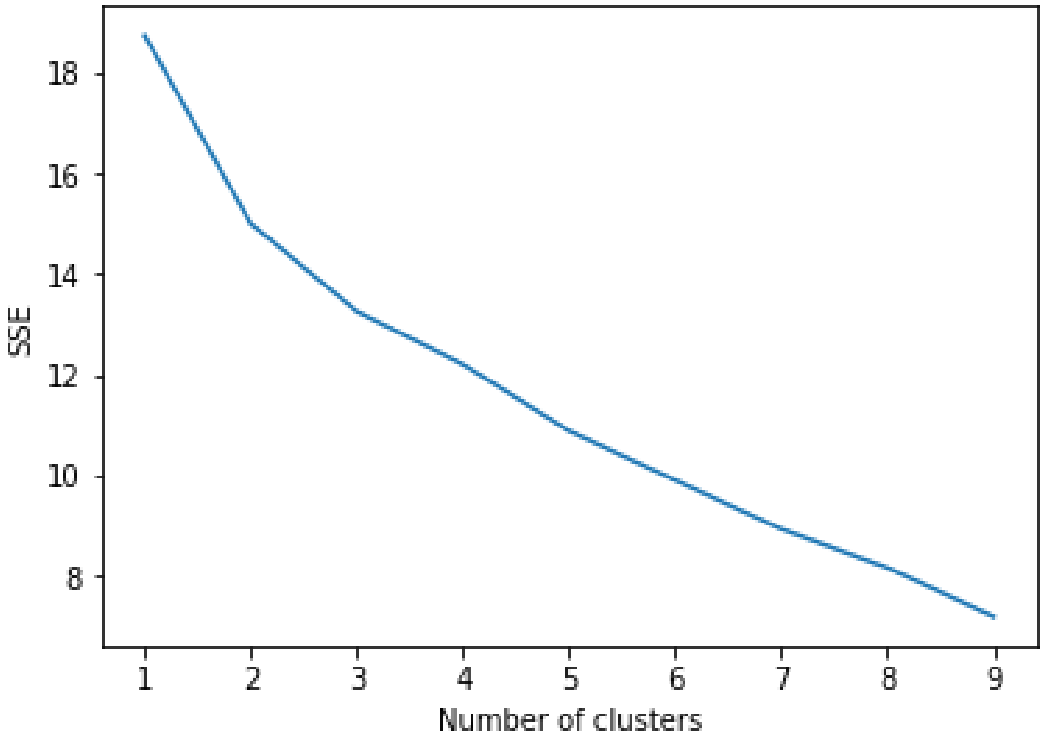


Methodology-3

- Perform one-hot encoding to obtain famous tourist spots in each district.
- Use the venues obtained to find the best business that could be established in the locality.
- Use K-means algorithm to cluster the business categories obtained.

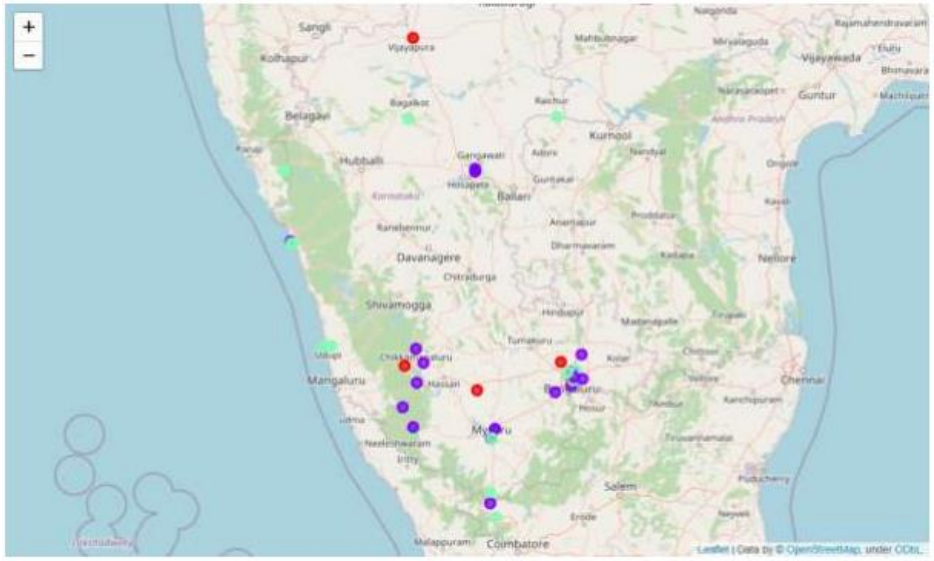
Cluster Validation

Here elbow method is used to determine the optimum value of k to perform K-means clustering. The graph obtained is –



Conclusion

- The results show that the most common business in cluster one at the respective venues are Indian Restaurants.
- So Indian Restaurants are popular in these tourist venues and opening up a similar one can attract many tourists.



Note - The colors purple, green, and red represents cluster 1, 2, and 3 respectively.

Conclusion

- Whereas in cluster two the most sought business is the Hotel, Seafood Restaurants, and Cafeterias.
- The green clusters at the seaside clearly indicate that opening a seafood restaurant would help a person make the best use of the opportunity.
- Finally, in cluster three Fast Food/Vegetarian Restaurants have been given a top priority.