IBM Applied Data Science Capstone

Understanding Tourism in India for Business

Nithiya Nanmaran

Introduction

Tourism in India is important for the country's economy and is growing rapidly. The Ministry of Tourism designs national policies for the development and promotion of tourism. The Ministry consults and collaborates with other stakeholders in the sector including various central ministries/agencies, state governments, union territories and private sector representatives. Concerted efforts are being made to promote niche tourism products such as rural, cruise, medical and eco-tourism. The Ministry of Tourism maintains the Incredible India campaign focused on promoting the tourism in India.

Business Problem

One of the importance of tourism is the employment opportunity which it offers to the people of that country. The purpose 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.

This project helps the people who are interested in opening a restaurant, lodging, transport services.

Data Source

Districts in India 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.

Here, we are considering one of the states of India, **Tamil Nadu**.

The beautifulsoup4 package, to scrape information from the web pages and convert it into a pandas data frame. Then we use Python geopy package to get latitude and longitude for all the districts. Foursquare API is used to understand the neighbourhoods in the districts to give clear idea to the stakeholders.

The dataset contains: District, Code, Area, Population, Population Density, Latitude, Longitude.

	District	Code	Area	Population	Population_Density	Latitude	Longitude
1	Ariyalur	AR	1949.31	754,894	390	11.076	79.117455
2	Chengalpattu	CGL	2,944.96	2,556,244	868	12.6841	79.983637
3	Chennai	CH	426	4,646,732	26,076	13.0837	80.270186
4	Coimbatore	CO	4,723	3,458,045	732	11.0018	76.962842
5	Cuddalore	CU	3,703	2,605,914	709	11.7564	79.763464

Literature Review

Tourism in India is important for the country's economy and is growing rapidly. The World Travel and Tourism Council calculated that tourism generated ₹16.91 lakh crore (US$240 billion) or 9.2% of India's GDP in 2018 and supported 42.673 million jobs, 8.1% of its total employment.

- The sector is predicted to grow at an annual rate of 6.9% to ₹32.05 lakh crore (US$450 billion) by 2028 (9.9% of GDP).
- In October 2015, India's medical tourism sector was estimated to be worth US$3 billion, and it is projected to grow to US$7–8 billion by 2020.
- In 2014, 184,298 foreign patients travelled to India to seek medical treatment.

The Ministry of Tourism designs national policies for the development and promotion of tourism. In the process, the Ministry consults and collaborates with other stakeholders in the sector including various central ministries/agencies, state governments, union territories and private sector representatives. Concerted efforts are being made to promote niche tourism products such as rural, cruise, medical and eco-tourism. The Ministry of Tourism maintains the Incredible India campaign focused on promoting the tourism in India.

Methodology

The following steps are followed to solve the problem taken into consideration.

- One hot encoding to obtain famous Tourist spots in each district.

- From the venue details, business opportunity can be determined.

- Using k-means algorithm to cluster the business details.

The state we taken into consideration for understanding of Tourism is **Tamil Nadu.** The dataset we have here,

|	District	Code	Area	Population	Population_Density	Latitude	Longitude
1	Ariyalur	AR	1949.31	754,894	390	11.076	79.117455
2	Chengalpattu	CGL	2,944.96	2,556,244	868	12.6841	79.983637
3	Chennai	CH	426	4,646,732	26,076	13.0837	80.270186
4	Coimbatore	CO	4,723	3,458,045	732	11.0018	76.962842
5	Cuddalore	CU	3,703	2,605,914	709	11.7564	79.763464

The districts are plotted on to the map of Tamil Nadu.
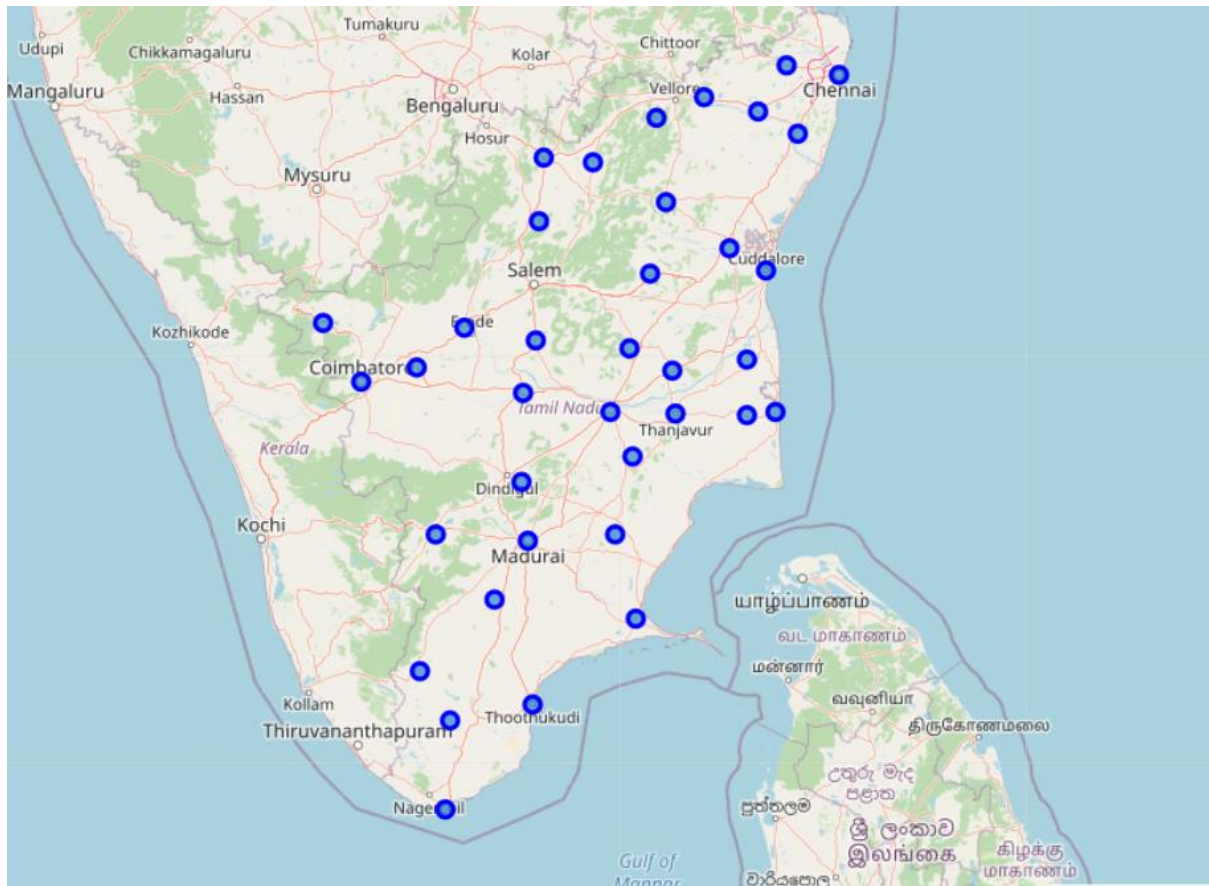
Using FourSquare API, get the venues of the districts in the data frame. There are different categories of the venues. We have removed few of the categories from the list. We are considering Nightlife spot and Outdoors & Recreations for Tourism category. For business category, we are taking into consideration Food, Shop & Service and Travel & Transport.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Ariyalur	11.076036	79.117455	Maha Magam Tank	10.956496	79.381546	Pool
7	Ariyalur	11.076036	79.117455	Gangai Konda Chozhapuram	11.206286	79.449432	Garden
29	Chengalpattu	12.684089	79.983637	Krishna's Butter Ball	12.618988	80.192722	Sculpture Garden
35	Chengalpattu	12.684089	79.983637	Mahabalipuram	12.861927	80.228115	Beach
53	Chennai	13.083694	80.270186	Anna Nagar Tower Park	13.086571	80.213627	Park
...	...	...	...	...	...	...	...
708	Viluppuram	11.913787	79.507893	Paradise Beach	11.875693	79.820578	Beach
718	Viluppuram	11.913787	79.507893	Promenade Beach	11.932650	79.836295	Beach
721	Viluppuram	11.913787	79.507893	Zest Private Beach	11.805728	79.797210	Beach
724	Viluppuram	11.913787	79.507893	Bharathi Park	11.932934	79.834294	Park
728	Viluppuram	11.913787	79.507893	Serenity Beach	11.971024	79.844426	Beach

The venues are listed for each district. The count of venues for each district are calculated, they are put in to visualization for better understanding.
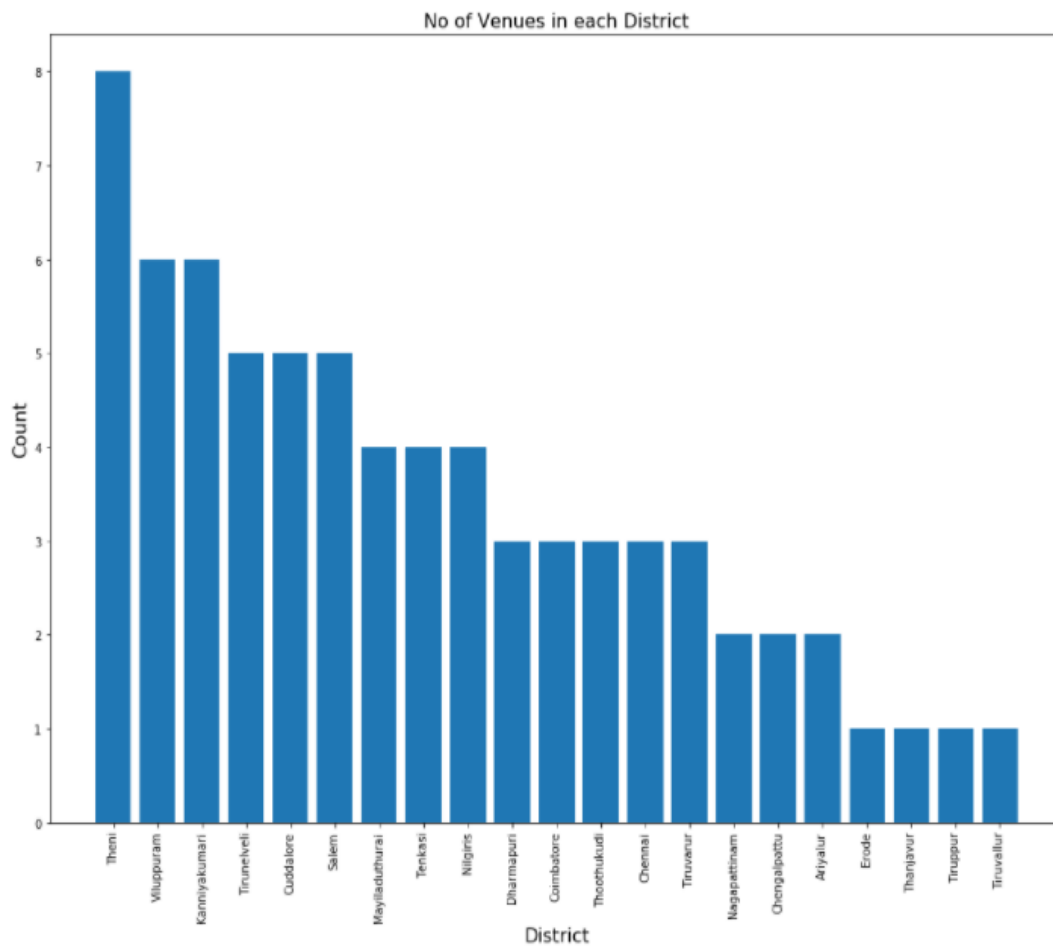
With the venue details, using Foursquare API, we are getting the businesses near the venues.

|	Venue	Business	BLatitude	BLongitude	Business Category
1	Maha Magam Tank	Hotel Raya's	10.957723	79.380810	Hotel
2	Maha Magam Tank	Rayas Grand	10.956523	79.380908	Hotel
3	Maha Magam Tank	Rathimeena Travels - KMU	10.958472	79.381596	Bus Station
8	Krishna's Butter Ball	Mamalla Heritage Hotel Tamil Nadu	12.618603	80.193797	Indian Restaurant
10	Krishna's Butter Ball	Moonrakers	12.619642	80.196394	Seafood Restaurant
...	...	...	...	...	...
583	Bharathi Park	Le Chateau	11.928759	79.833679	French Restaurant
585	Bharathi Park	Kasha Ki Aasha	11.930277	79.831313	Café
586	Bharathi Park	Nilgiris Supermarket	11.934763	79.830949	Department Store
588	Serenity Beach	Bodhi Cafe	11.971132	79.844410	Café
591	Serenity Beach	Theevu Plage	11.970995	79.844329	Seafood Restaurant

343 rows × 5 columns

The count of each businesses near the venues are taken. There are 48 unique businesses. After the one hot encoding the dataset looks like.

|	Venue	ATM	American Restaurant	Asian Restaurant	Bakery	Bistro	Boat or Ferry	Bookstore	Breakfast Spot	Bubble Tea Shop	...	Restaurant	Salon / Barbershop	Sandwich Place	Seafood Restaurant	Shop & Service
1	Maha Magam Tank	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Maha Magam Tank	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Maha Magam Tank	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
8	Krishna's Butter Ball	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
10	Krishna's Butter Ball	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0

The average of the number of each business near the venue are taken. Then top 5 business are listed.

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
0	Anna Nagar Tower Park	Indian Restaurant	Vegetarian / Vegan Restaurant	Burger Joint	Juice Bar	Market	Coffee Shop	Clothing Store	Fast Food Restaurant	Electronics Store	Asian Restaurant
1	Bharathi Park	Coffee Shop	Hotel	Pizza Place	Bakery	Café	Indian Restaurant	Vegetarian / Vegan Restaurant	Department Store	Italian Restaurant	French Restaurant
2	Blend - The High Energy Bar	Clothing Store	Hotel	Bookstore	Juice Bar	Steakhouse	Shopping Mall	Sandwich Place	Indian Restaurant	Breakfast Spot	Coffee Shop
3	Bryant Park	Hotel	Bakery	Indian Restaurant	Pizza Place	Bus Station	Fast Food Restaurant	Modern European Restaurant	Café	Vegetarian / Vegan Restaurant	Clothing Store
4	Bush's Pasture Park	American Restaurant	Vegetarian / Vegan Restaurant	Café	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop

With the data collected and pre-processed, we then use k-means clustering to divide the observations into clusters. We use elbow method to get the optimal value of k. Below is the elbow method implemented.
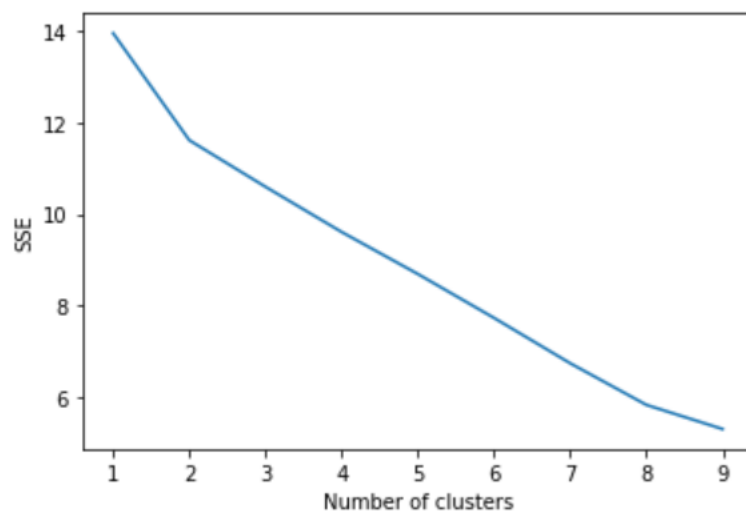
With the value of k taken as 3, below is the clusters of the districts.
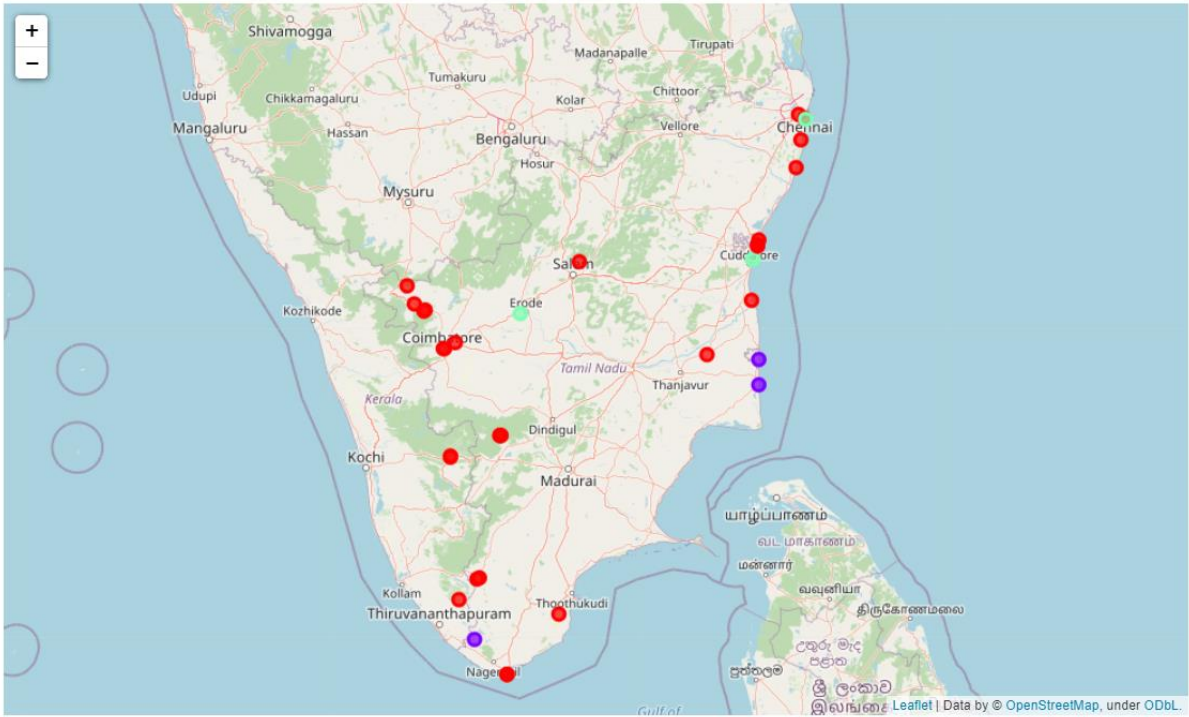
The Red, Green and Purple dots represent observations of 3 different clusters. Few observations in each cluster are taken screenshot and mentioned below.

Cluster 1

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
252	Thirparappu Falls	Bus Station	Vegetarian / Vegan Restaurant	Café	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop
258	Karaikal Beach	Indian Restaurant	Vegetarian / Vegan Restaurant	Flower Shop	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop	Coffee Shop
268	Velankanni Beach	Vegetarian / Vegan Restaurant	Indian Restaurant	Flower Shop	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop	Coffee Shop
327	Thompson Brewery & Public House	Other Repair Shop	Salon / Barbershop	Bus Station	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop	Coffee Shop

Cluster 2

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
73	Marina Beach | மெரினா கடற்கரை (Marina Beach)	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Train Station	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop	Coffee Shop
124	Zest Private Beach	Coffee Shop	Vegetarian / Vegan Restaurant	Train Station	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop
209	Thindal	Motorcycle Shop	Vegetarian / Vegan Restaurant	Flower Shop	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop	Coffee Shop

Cluster 3

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
1	Maha Magam Tank	Hotel	Bus Station	Vegetarian / Vegan Restaurant	Café	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop
8	Krishna's Butter Ball	Burger Joint	Indian Restaurant	Seafood Restaurant	Restaurant	Bus Stop	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop
15	Mahabalipuram	Restaurant	Fast Food Restaurant	Breakfast Spot	Café	Bus Stop	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store
23	Anna Nagar Tower Park	Indian Restaurant	Vegetarian / Vegan Restaurant	Burger Joint	Juice Bar	Market	Coffee Shop	Clothing Store	Fast Food Restaurant	Electronics Store	Asian Restaurant
49	Blend - The High Energy Bar	Clothing Store	Hotel	Bookstore	Juice Bar	Steakhouse	Shopping Mall	Sandwich Place	Indian Restaurant	Breakfast Spot	Coffee Shop
79	Decathlon	ATM	Bakery	Italian Restaurant	Café	Electronics Store	Donut Shop	Diner	Dessert Shop	Department Store	Cosmetics Shop
83	Race Course Road	Department Store	Vegetarian / Vegan Restaurant	Asian Restaurant	Hotel	Bistro	Indian Restaurant	Dessert Shop	Restaurant	Fast Food Restaurant	Coffee Shop

Discussion

- From the results, the most common business in cluster one is Food business. There are many Indian and Vegetarian cuisine restaurants, opening up other cuisines attract people.
- In cluster two, the most sought business is the Hotel. Opening up lodging service could be best option.
- Finally, in cluster three, Cafes are in business.

Conclusion and Scope of Improvement

With the given information of the tourist spots, finding best options for business help people who are willing to start the business.
Given more information of the tourist spots, we can tune the results to get accurate options.
Using k++, might help in better clusters.

References

https://en.wikipedia.org/wiki/Tourism_in_India
https://en.wikipedia.org/wiki/List_of_districts_in_India