

IBM Applied Data Science Capstone

Recommending to Start a Business at a Tourist Venue

Sridevi Sribhashyam

Dec 11, 2020

Table of Contents

- [1. Introduction](#)
 - [1.1 Background](#)
 - [1.2 Business Problem](#)
 - [1.3 Target Audience](#)
- [2. Data Collection and Anotomaization](#)
 - [2.1 Data Source](#)
 - [2.2 Description of the Data](#)
- [3. Methodology - Exploratory Data Analysis](#)
- [4. Methodology - Predictive Modeling & Clustering](#)
- [5. Results & Recommendations](#)
- [6. Conclusion](#)
- [7. Future Direction](#)
- [8. References](#)

1. Introduction

1.1 Background

For long, tourism has been prohibitively expensive for many. Today, it has become a way of life, especially for Millenials. Tourism comprises the activities of persons traveling to places and spending time outside their normal daily environment for leisure, business, and other purposes. Travel exposes one to life outside; gives confidence in dealing with change and provides independence and freedom. It boosts our self-awareness and helps us get real-life education. Travel to a foreign country helps us understand foreign culture, their history, meet new people, make new friends, and helps us navigate difficult situations.

Globalization, as well as diplomatic relations among countries, has made traveling increasingly easy. As a philomath and a historian, I acknowledge that tourism plays a prominent role for a globetrotter and, in fact for any traveler.

Tourism in India has significant potential considering the rich cultural and historical heritage, variety in ecology, terrains and places of natural beauty spread across the country.

Constituting a major part of Deccan plateau, which is south of India, Telangana has a pleasing climate, with abundant natural and water resources. Telangana is a highly developed state and society with proud history and grand heritage.

Telangana apart from the historical perspective of various dynastic and regional and cultural influences offers its visitors a plethora of places of pristine natural beauty. With serene lakes, verdant woods, rocky regions, abundant wildlife, flora and fauna, elaborately carved temples, magnificent palaces and forts reflecting architectural blends of Hindu, Pathan and Persian styles. The artistic temper of the local people finds expression in the varied art forms of the different tribes and regions of Telangana in paintings, handicrafts in wood and metal and textile of International repute.

Hyderabad is famous as the former seat of the fabulously wealthy Nizams. Built on the banks of the river Musi, and surrounded by huge prehistoric rock formations like petrified, grey elephants, Hyderabad is a blend of the fairy tale and earthy, a potpourri of old and new. Endearingly called the Pearl City, it offers a variety of tourist attractions ranging from Heritage monuments, Lakes, Parks, Gardens, Resorts, Museums to delectable cuisine and a delightful shopping experience. Hyderabad is an important center of Islamic culture, places which reflect that culture are FALAKNUMA PALACE-Dazzling beauty and sophisticated magnificence that surpasses your wildest imagination, and CHARMINAR-An example of unpatrolled artistic brilliance.

Warangal is the second largest city in Telangana. Originally known as Orugallu which means 'one stone', Warangal was ruled by the Kakatiya dynasty. It is believed that the entire town was carved out on a single rock. The Kakatiyas built many monuments in the state which include iconic fortresses, gateways and temples. Among the most famous monuments of Telangana, the Thousand Pillar Temple in Hanamkonda, was built by the Kakatiya King Rudra Deva in 1163 AD in Chalukyan style of temple architecture. The city of Warangal was also described in detail by Marco Polo during his visit to the country. The city is a cluster of three towns and is popular for many attractions. While visiting Warangal, one should definitely stop by the beautiful Khush Mahal located near the Warangal Fort, the symbol of a glorious past.

Other prominent art works of the state include Nizamabad panels, Nirmal painted furniture, Dokra castings, Silver Filigree, Cheriyal Scroll Paintings, Bidri craft, Pembarthi brassware etc.

Importance of Tourism

- ❖ 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 boosts the revenue of the economy. Total contribution by the travel and tourism sector to India's GDP is expected to increase from Rs 15.24 lakh crore in 2017 to Rs 32.05 lakh crore in 2028. This is a huge earning to India and creates thousands of jobs.
- ❖ Tourism is a significant source of foreign exchange for the country.
- ❖ Tourism develops the infrastructures of a country, and plants a sense of cultural exchange between foreigners and citizens.
- ❖ In FY20, 39 million jobs were created in the tourism sector in India; this accounted for 8.0% of the total employment in the country. The number is expected to rise by two% annually to 52.3 million jobs by 2028.
- ❖ Telangana, a state in India is steadily becoming favourite destinations for foreign tourists. As per statistics shared by the Telangana tourism department, there is a 76% jump in foreign tourist count in 2017.

1.2 Business Problem

The objective of this project is to analyze the tourist places of Telangana state in India and recommend the best location where one can open a Restaurant or Hotel to make the best use of the opportunity.

1.3 Target Audience

The target audience for this project are people who are interested in opening a Restaurant, Lodging/Hotel/Motel, Transport Services or any other businesses which fall within the tourism industry in the state of Telangana, India.

2. Data Collection and Anotomaization

To find solution to the business problem mentioned above, we should have the dataset that contains

- ❖ All the states of a particular country, India in this case, and which has 36 states
- ❖ Latitude and longitudes of all the districts of individual states, a total of 739 districts in India
- ❖ Telangana state is chosen for this problem and the state has 13 districts

2.1 Data Source

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. Also used Google Places API to get latitude, longitude for some locations.

2.2 Description of the Data

The output shows the final dataset. The dataset consists of a single Dataframe 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.

```
import pandas as pd
df = pd.read_excel('Capstone_TelanganaDataSet.xlsx')
df.head(15) # Display first 15 rows
```

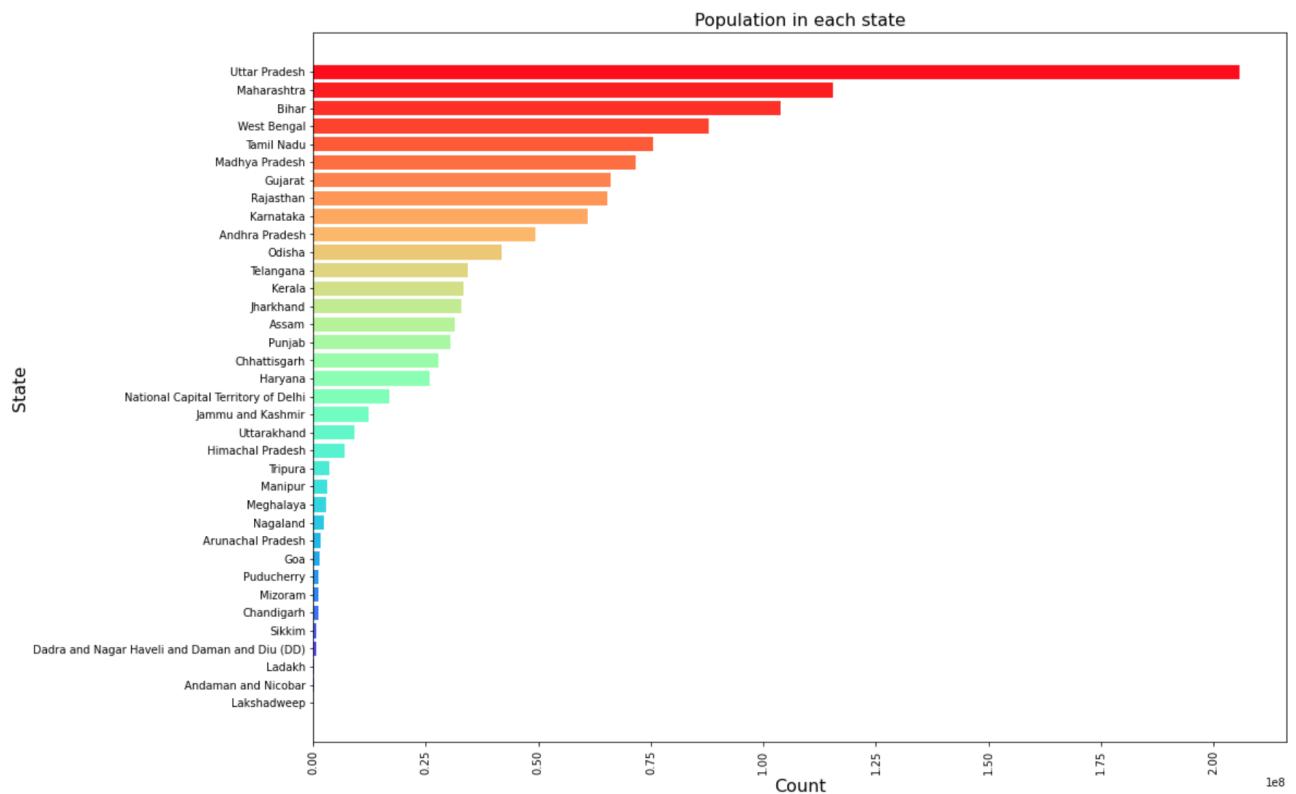
	State	Code	District	Headquarters	Population(2011)	Area(km2)	Density(/km2)	Latitude	Longitude
0	Andaman and Nicobar	NI	Nicobar	Car Nicobar	36842	1841.00	20.0	7	93
1	Andaman and Nicobar	NaN	North and Middle Andaman	Mayabunder	105597	3736.00	28.0	12.6112	92.8317
2	Andaman and Nicobar	SA	South Andaman	Port Blair	238142	2672.00	89.0	10.7057	92.4875
3	Andhra Pradesh	AN	Anantapur	Anantapur	4083315	19130.00	213.0	14.6546	77.5563
4	Andhra Pradesh	CH	Chittoor	Chittoor	4170468	15152.00	275.0	13.1601	79.1556
5	Andhra Pradesh	EG	East Godavari	Kakinada	5151549	10807.00	477.0	17.2335	81.7226
6	Karnataka	BK	Bagalkot	Bagalkot	1890826	6583.00	288.0	16.1853	75.6968
7	Karnataka	BL	Ballari	Ballari	2532383	8439.00	300.0	15.2485	76.8324
8	Karnataka	BG	Belgaum	Belgaum	4778439	13415.00	356.0	15.8573	74.5069
9	Karnataka	BR	Bangalore Rural	Bangalore	987257	2239.00	441.0	13.0011	77.3361
10	Karnataka	BN	Bangalore Urban	Bangalore	9588910	2190.00	4378.0	12.9451	77.5536
11	Telangana	AD	Adilabad	Adilabad	708952	4185.97	171.0	19.6641	78.532
12	Telangana	KB	Komaram Bheem	Asifabad	515835	4300.16	106.0	19.3619	79.293
13	Telangana	BK	Bhadradri Kothagudem	Kothagudem	1304811	8951.00	143.0	17.556	80.617
14	Telangana	HY	Hyderabad	Hyderabad	3441992	4325.29	18172.0	17.385	78.4867

3. Methodology - Exploratory Data Analysis

Data that is required to solve the business problem is saved in Capstone_TelanganaDataSet.xml. The final dataset has nine columns as shown below

	State	Code	District	Headquarters	Population	Area	Density	Latitude	Longitude
0	Andaman and Nicobar	NI	Nicobar	Car Nicobar	36842	1841.0	20.0	7	93
1	Andaman and Nicobar	NaN	North and Middle Andaman	Mayabunder	105597	3736.0	28.0	12.6112	92.8317
2	Andaman and Nicobar	SA	South Andaman	Port Blair	238142	2672.0	89.0	10.7057	92.4875
3	Andhra Pradesh	AN	Anantapur	Anantapur	4083315	19130.0	213.0	14.6546	77.5563
4	Andhra Pradesh	CH	Chittoor	Chittoor	4170468	15152.0	275.0	13.1601	79.1556

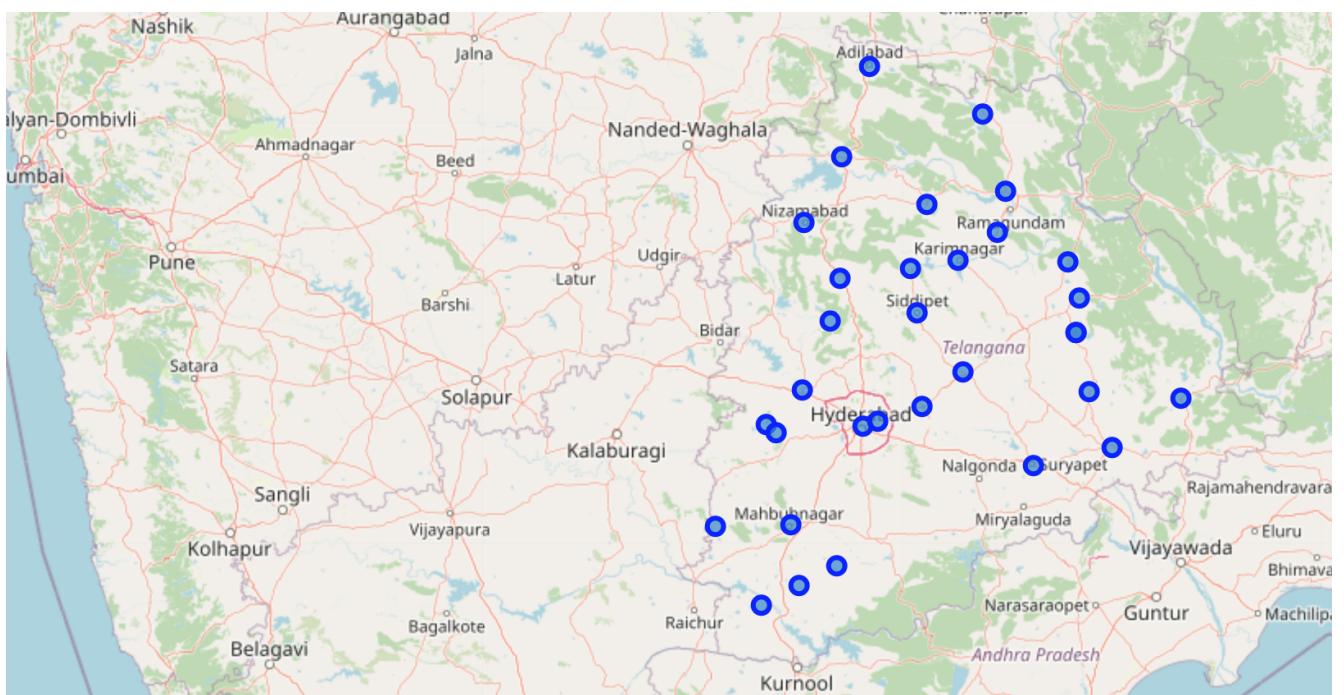
There are 36 states which have been retrieved from the webpage and stored in the dataset. There were some missing values in the dataset and they were removed. The below graph shows the population in each state.



```
print('The data has {} states'.format(len(df['State'].unique()))))
```

The data has 36 states

Telangana state in India, is used in solving the problem. A map with all the districts of Telangana is displayed below

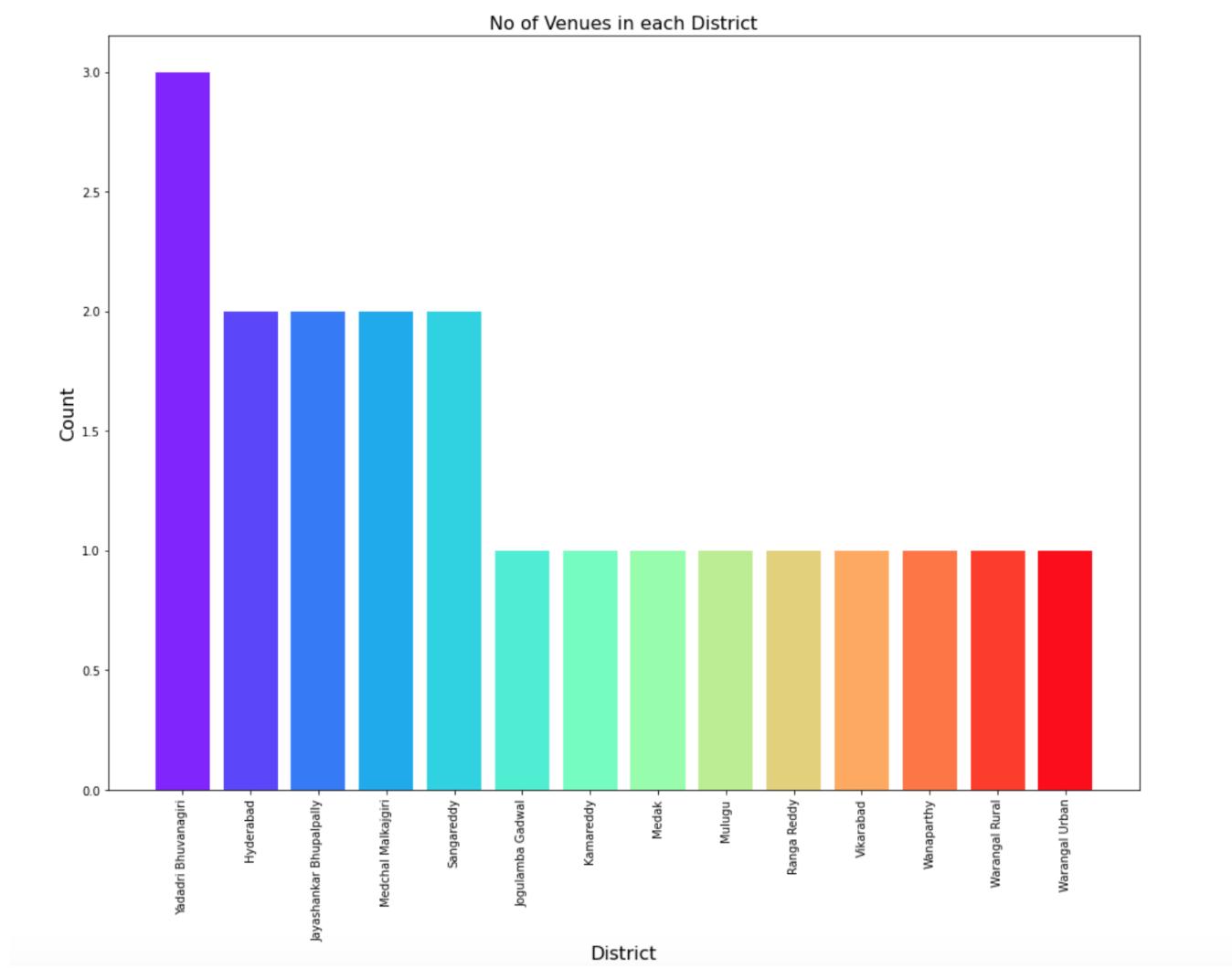


Using Foursquare API, retrieve main categories and subcategories present from venues. This list contains Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service, Travel & Transport. Categories related to tourists are listed below

Category 1 = Arts & Entertainment, Nightlife Spot, Outdoors & Recreation

Category 2 = Food, Shop & Service, Travel & Transport

Now get the nearby tourist venues within a radius of 50km., and this gives multiple tourist spots in a particular district. A bar graph by plotting District v/s Count is shown below to get the number of venues in each district.



Organize unique venue categories and use one hot encoding to analyze each district. Identify top seven most common venue categories.

	District	1st Most Common Venue Category	2nd Most Common Venue Category	3rd Most Common Venue Category	4th Most Common Venue Category	5th Most Common Venue Category	6th Most Common Venue Category	7th Most Common Venue Category
0	Hyderabad	Nightclub	Scenic Lookout	Performing Arts Venue	Park	Historic Site	River	Outdoors & Recreation
1	Jayashankar Bhupalpally	Lake	Scenic Lookout	River	Performing Arts Venue	Park	Outdoors & Recreation	Nightclub
2	Jogulamba Gadwal	River	Scenic Lookout	Performing Arts Venue	Park	Outdoors & Recreation	Nightclub	Lake
3	Kamareddy	Historic Site	Scenic Lookout	River	Performing Arts Venue	Park	Outdoors & Recreation	Nightclub
4	Medak	Historic Site	Scenic Lookout	River	Performing Arts Venue	Park	Outdoors & Recreation	Nightclub

Aggregate all the venues based on the venue category for a given district.

	District	Venue Category	Venue
0	Hyderabad	Historic Site	Qutub Shahi Tombs
1	Hyderabad	Nightclub	Rain Lounge, Heart Cup Coffee
2	Hyderabad	Park	KBR Park
3	Hyderabad	Performing Arts Venue	Lamakaan
4	Hyderabad	Scenic Lookout	Necklace Road

Get most common venues according to the venue category in each District to get an idea where one can start business in a particular district, however this cannot provide a clarity on what type of business can it be.

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Hyderabad	Rain Lounge, Heart Cup Coffee	Necklace Road	Lamakaan	KBR Park	Qutub Shahi Tombs		
1	Jayashankar Bhupalpally	Laknavaram Lake, laknavaram lake						
2	Jogulamba Gadwal	beechpalli						
3	Kamareddy	Medak Church						
4	Medak	Medak Church						
5	Medchal Malkajgiri	Rain Lounge, Heart Cup Coffee	Necklace Road	Lamakaan	KBR Park	Qutub Shahi Tombs		
6	Mulugu	Thousand Pillar Temple						
7	Ranga Reddy	Palm Exotica Resort & Spa						
8	Sangareddy	Heart Cup Coffee, Rain Lounge	Necklace Road	Lamakaan	KBR Park	Bhel Township	Qutub Shahi Tombs	
9	Vikarabad	Bhel Township						
10	Wanaparthys	beechpalli						
11	Warangal Rural	Thousand Pillar Temple						
12	Warangal Urban	Thousand Pillar Temple						
13	Yadadri Bhuvanagiri	Necklace Road	Lamakaan	KBR Park	Rain Lounge	Ramoji Film City		

Now find the top most desired businesses which can be established at the tourist venue within 1km range and found 34 unique categories.

1	Asian Restaurant	26
2	Auto Garage	3
3	BBQ Joint	22
4	Bakery	16
5	Boat or Ferry	4
6	Bookstore	12
7	Breakfast Spot	25
8	Café	97
9	Clothing Store	16
10	Coffee Shop	36
11	Deli / Bodega	4
12	Department Store	10
13	Dessert Shop	20
14	Diner	14
15	Donut Shop	11
16	Dumpling Restaurant	3
17	Electronics Store	12
18	Fast Food Restaurant	57
19	Food Court	22
20	Food Truck	5

21	Fried Chicken Joint	8
22	Furniture / Home Store	8
23	Hotel	47
24	Indian Restaurant	114
25	Italian Restaurant	20
26	Motorcycle Shop	4
27	Pizza Place	26
28	Restaurant	37
29	Sandwich Place	35
30	Shopping Mall	23
31	Smoke Shop	3
32	Spa	8
33	Vegetarian / Vegan Restaurant	4

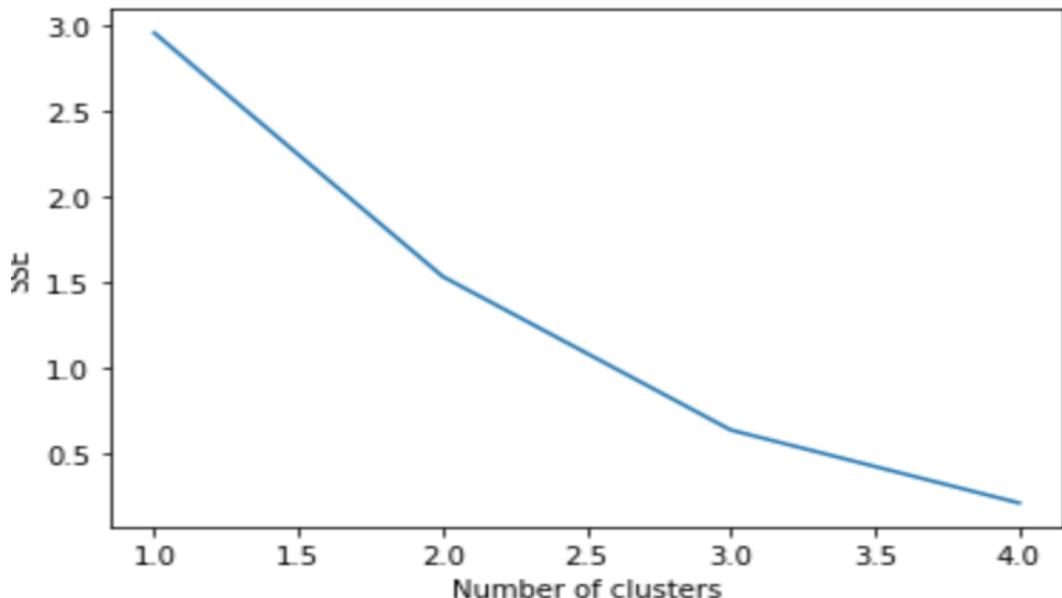
```
: print('There are {} uniques categories.'.format(len(nearby_business['Business Category'].unique())))
There are 34 uniques categories.
```

Analyse each venue to get the top businesses at a venue using one hot encoding

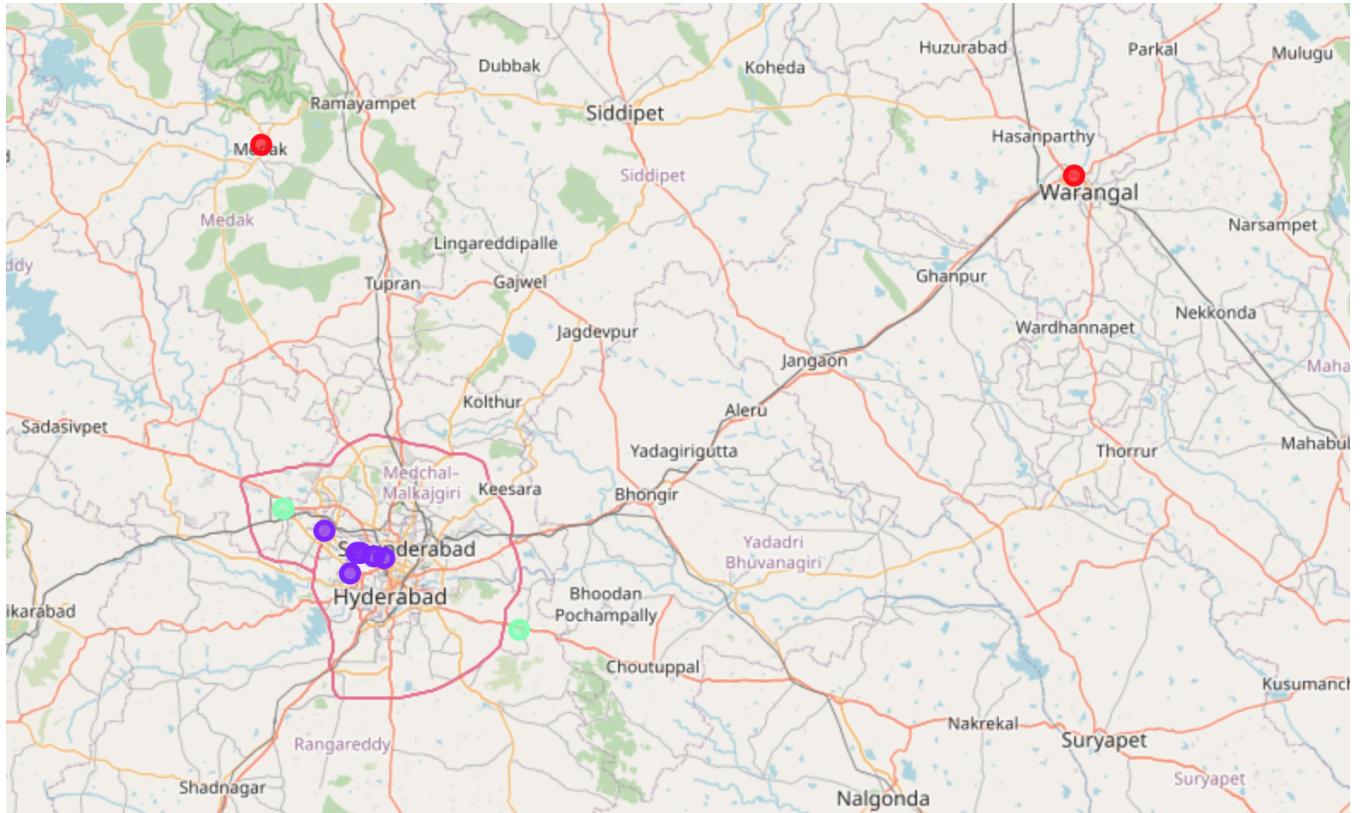
	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
0	Bhel Township	Breakfast Spot	Café	Diner	Dessert Shop	Department Store	Deli / Bodega	Coffee Shop	Clothing Store	Vegetarian / Vegan Restaurant	Spa
1	Heart Cup Coffee	Indian Restaurant	Café	Restaurant	BBQ Joint	Food Court	Department Store	Fast Food Restaurant	Diner	Donut Shop	Breakfast Spot
2	KBR Park	Café	Breakfast Spot	Clothing Store	Asian Restaurant	Dessert Shop	Italian Restaurant	Fast Food Restaurant	Diner	Spa	Electronics Store
3	Lamakaan	Indian Restaurant	Coffee Shop	Hotel	Fast Food Restaurant	Café	BBQ Joint	Shopping Mall	Sandwich Place	Bookstore	Restaurant
4	Medak Church	Indian Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop	Department Store	Deli / Bodega	Coffee Shop	Clothing Store	Café	Breakfast Spot	Donut Shop

4. Methodology - Predictive Modeling & Clustering

Using the K-means clustering algorithm, group the businesses into clusters that aim to partition 'n' observations into k clusters in which each observation belongs to the cluster. Elbow method runs k-means clustering on the data set to determine the optimum value of k.



5. Results & Recommendations



Purple - cluster 1; Green - cluster 2; Red - cluster 3;

As per the results, the most common business in cluster 1 are Coffee Shops and Fast Food Restaurants. Opening similar ones can attract many tourists. Coffee and Chai, milk based are very popular in Telangana. Fast foods like Samosa, Chat, Cakes and grab & go foods are very attractive to people across the globe.

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
4	Necklace Road	Hotel	Indian Restaurant	Sandwich Place	Restaurant	Café	Fast Food Restaurant	Boat or Ferry	Coffee Shop	Donut Shop	Food Court
35	Lamakaan	Indian Restaurant	Coffee Shop	Hotel	Fast Food Restaurant	Café	BBQ Joint	Shopping Mall	Sandwich Place	Bookstore	Restaurant
126	Rain Lounge	Café	Indian Restaurant	Asian Restaurant	Sandwich Place	Restaurant	Dessert Shop	Bakery	Fast Food Restaurant	Food Truck	Breakfast Spot
175	KBR Park	Café	Breakfast Spot	Clothing Store	Asian Restaurant	Dessert Shop	Italian Restaurant	Fast Food Restaurant	Diner	Spa	Electronics Store
221	Qutub Shahi Tombs	Café	Asian Restaurant	Fast Food Restaurant	Restaurant	Pizza Place	Diner	Vegetarian / Vegan Restaurant	Department Store	Deli / Bodega	Coffee Shop
230	Heart Cup Coffee	Indian Restaurant	Café	Restaurant	BBQ Joint	Food Court	Department Store	Fast Food Restaurant	Diner	Donut Shop	Breakfast Spot

In cluster 2 the most common businesses are Breakfast Spot, Vegan Restaurants and Shopping Stores. Opening similar ones can attract many tourists.

Cluster 2

nearby_business_merged.loc[nearby_business_merged['Cluster Labels'] == 1, nearby_business_merged.columns[[0] + list(range(4,											
	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
632	Bhel Township	Breakfast Spot	Café	Diner	Dessert Shop	Department Store	Deli / Bodega	Coffee Shop	Clothing Store	Vegetarian / Vegan Restaurant	Spa
877	Ramoji Film City	Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Café	Dessert Shop	Department Store	Deli / Bodega	Coffee Shop	Clothing Store	Breakfast Spot

Finally, in cluster 3, Indian Restaurants, Dessert Shops and Deli Bodega are common businesses. Opening similar ones can attract many tourists.

Cluster 3

nearby_business_merged.loc[nearby_business_merged['Cluster Labels'] == 2, nearby_business_merged.columns[[0] + list(range(4,											
	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
286	Medak Church	Indian Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop	Department Store	Deli / Bodega	Coffee Shop	Clothing Store	Café	Breakfast Spot	Donut Shop
572	Thousand Pillar Temple	Indian Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop	Department Store	Deli / Bodega	Coffee Shop	Clothing Store	Café	Breakfast Spot	Donut Shop

6. Conclusion

In this project, a list of popular tourist locations of Telangana State of India has been identified using Foursquare API. Using k-means clustering algorithm, these tourist locations were clustered based on frequency of the businesses that are present. This indicated a list of business opportunities that can be established in that locality so that the business can attract many tourists.

7. Future Direction

Factors that can affect Tourism are Tourist Count, Fast Reporting, Social Media and Blog Reviews and Physical safety. Research can make use of these factors, and the number of similar businesses that could impact the new business being established. This might help the prediction make the analysis more accurate.

8. References

- [1] [International Journal of Development Research - Impact of tourism in Indian economy](#)
- [2] [Tourism & Hospitality Industry in India](#)
- [3] [Telangana - Favourite destinations for foreign tourists](#)