IBM Applied Data Science Capstone

Recommending to Start a Business at a Tourist Venue

Sridevi Sribhashyam

Dec 11, 2020

Importance of Tourism

❖ Tourism industry creates commercial demand and growth for many more industries

❖ Tourism boosts the revenue of the economy and it cretaes 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

Tourism Growth In India

❖ 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

❖ 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

Business Problem

Analyze the tourist places of Telangana state in India and recommend the best location where one can open a Restaurant or a Hotel to make the best use of the opportunity

Data Source and Description

❖ The Wikipedia page https://en.wikipedia.org/wiki/List_of_districts_in_India scraped to obtain all the districts of India

❖ Geopy API and Google API is used to obtain latitude and longitude of the respective districts

❖ Removed missing values present in the dataset and created a Dataframe with 9 columns

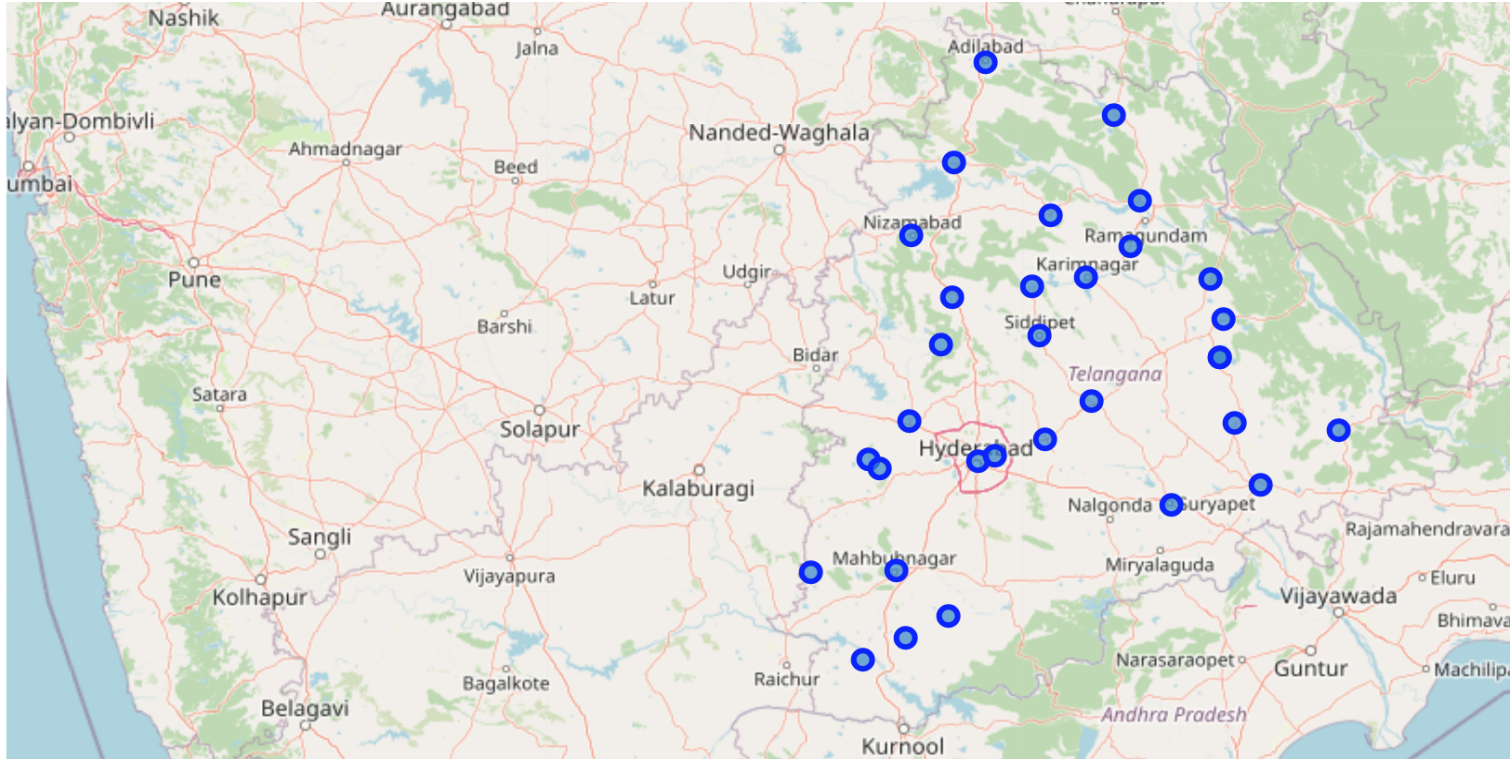❖ The final dataset contains 713 rows which includes 36 states of India

Population Characteristic of India
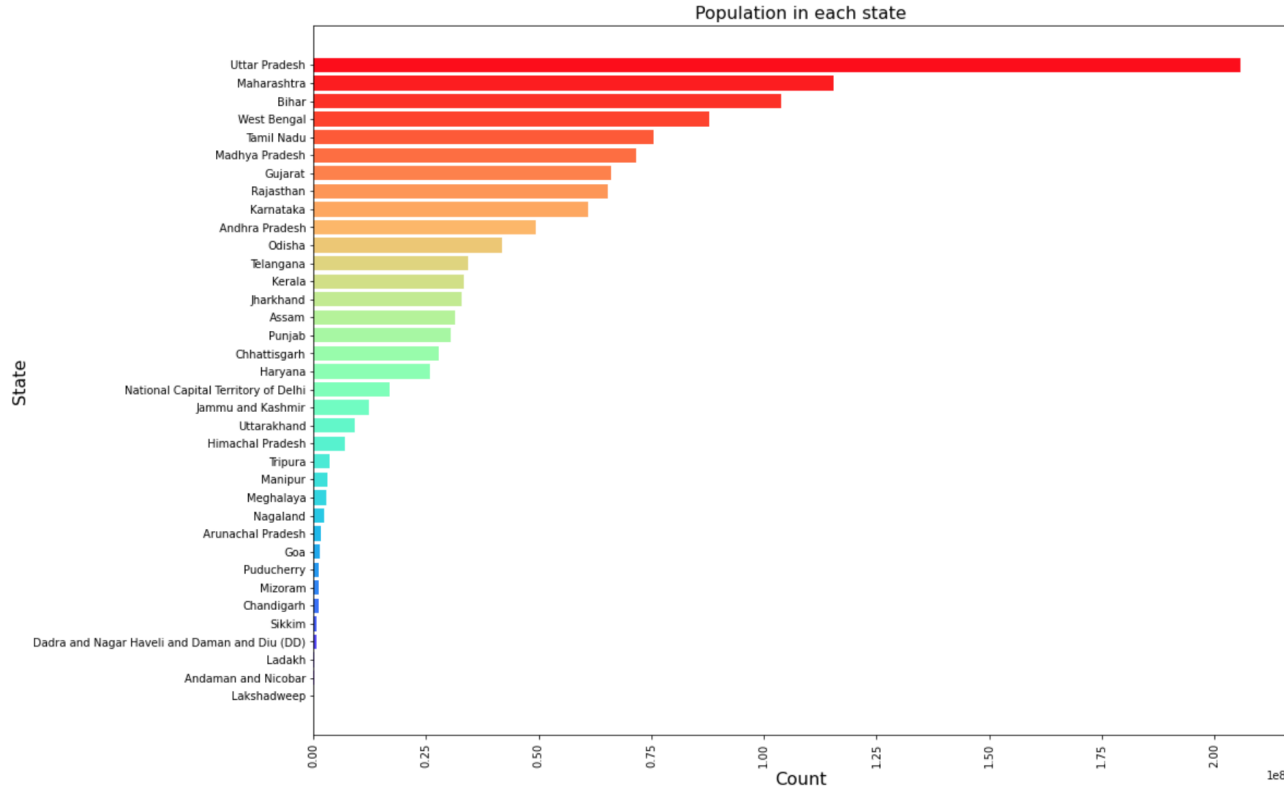
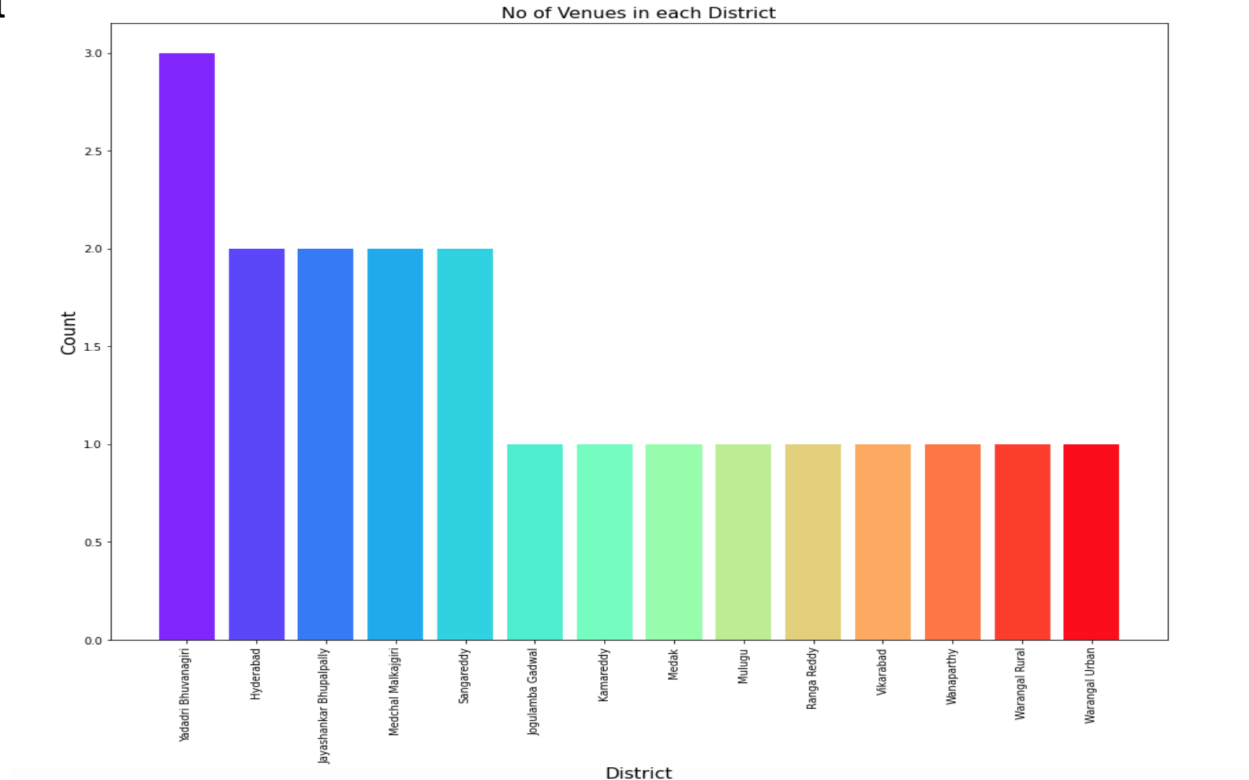★ Population characteristics makes the tourism industry lead to an improvement of the socio-economic conditions

Telangana State Map

Map of Telangana State

Data Exploratory Analysis

❖ Get nearby tourist venues within a radius of 50km and Count the number of venues in each district
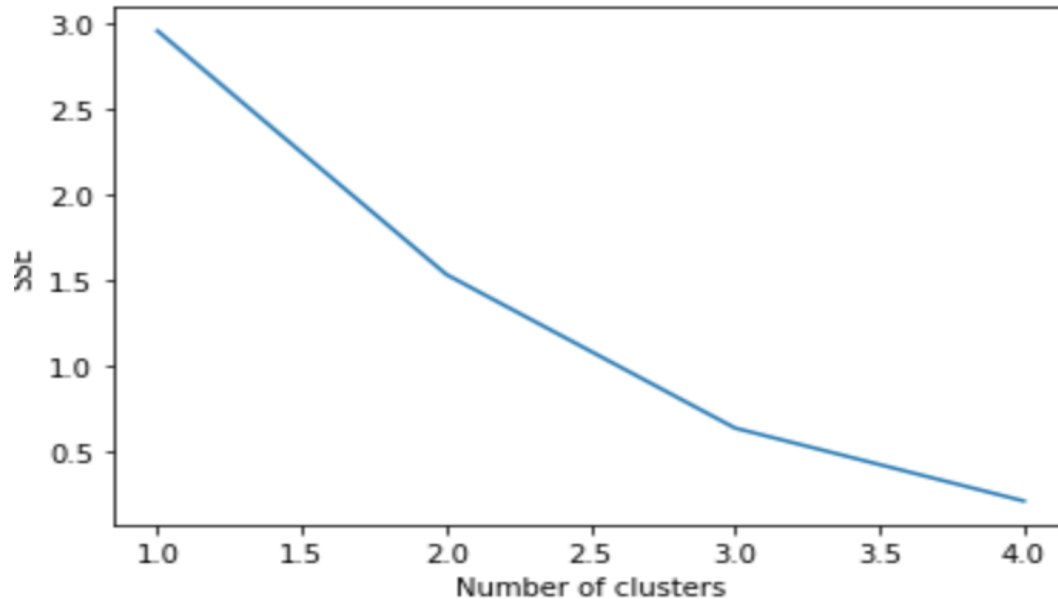
Predictive Modeling and Clustering

❖ 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

K-means Clustering Algorithm

❖ Elbow method used to find optimum value of k for the given data

Visualization of Clusters
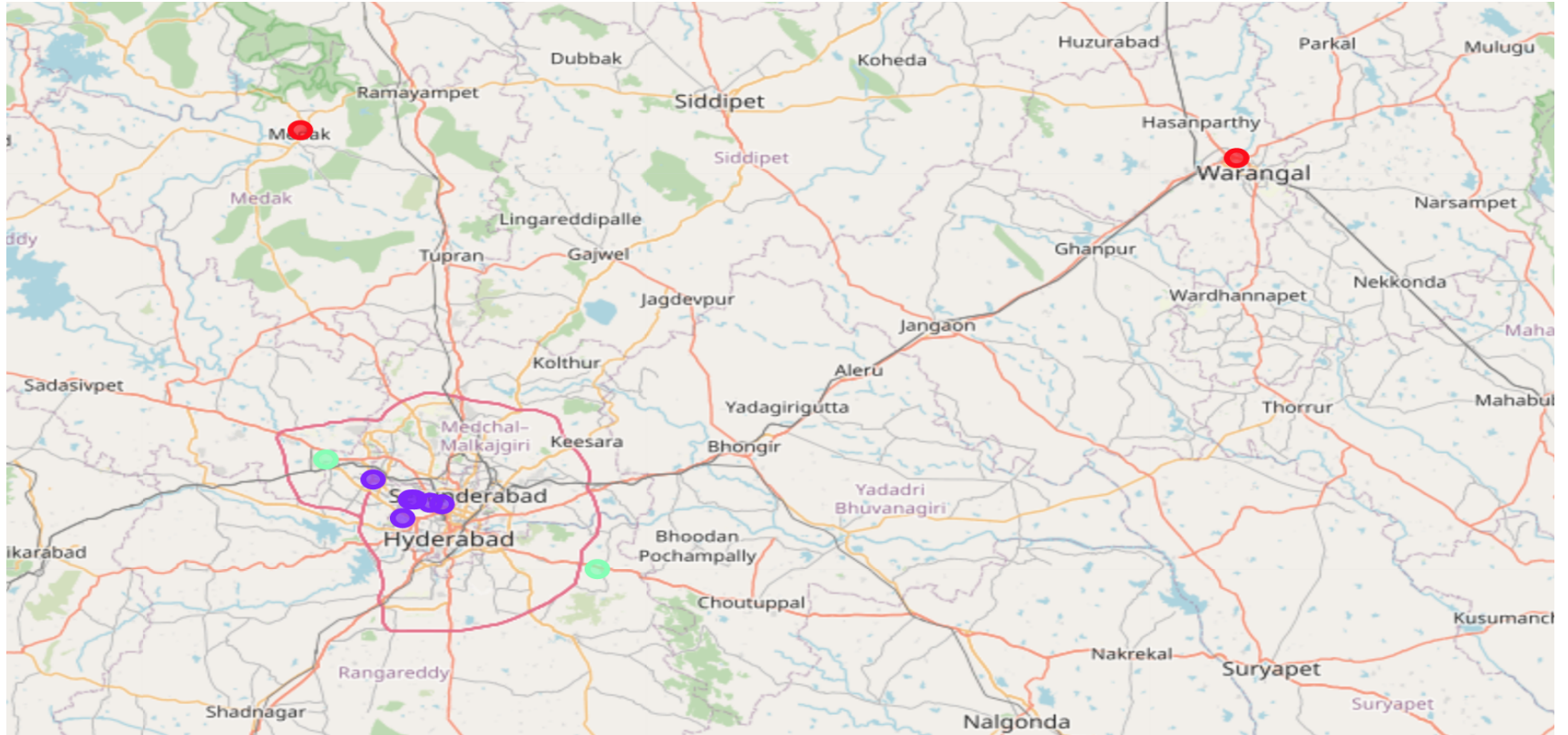
★ Purple - cluster 1; Green - cluster 2; Red - cluster 3;

Results and Recommendations

❖ As per the results the most common business in cluster one are Coffee Shops and Fast Food Restaurants. Opening similar ones can attract many tourists

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

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

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

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