

Recommending a Business at a particular Tourist Venue

IBM Applied Data Science Capstone Project

By César Fernandes

WHY TOURISM IS IMPORTANT TO ANY COUNTRY?

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

BUSINESS PROBLEM

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.

DATA ACQUISITION & CLEANING

- ❖ The Wikipedia page https://en.wikipedia.org/wiki/List_of_districts_in_India scraped to obtain all the districts of India.
- ❖ Geopy API used to obtain latitude and longitude of the respective districts.
- ❖ Removed any missing values present in the dataset
- ❖ The dataset consists of a single Dataframe with 9
- ❖ columns.
- ❖ The final dataset contains 719 rows which includes
- ❖ 36 states.

DATA ACQUISITION & CLEANING

- ❖ The Wikipedia page https://en.wikipedia.org/wiki/List_of_districts_in_India scraped to obtain all the districts of India.
- ❖ Geopy API used to obtain latitude and longitude of the respective districts.
- ❖ Removed any missing values present in the dataset
- ❖ The dataset consists of a single Dataframe with 9
- ❖ columns.
- ❖ The final dataset contains 719 rows which includes
- ❖ 36 states.

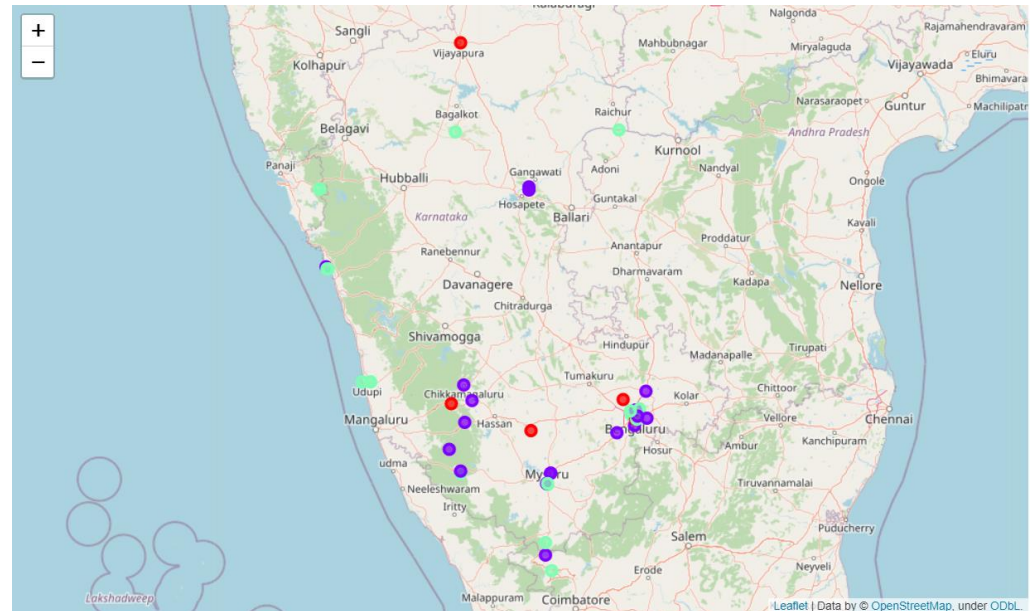
METHODOLOGY

- ❖ 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.

RESULTS & DISCUSSION

The colors purple, green, and red represents cluster 0, 1, and 2 respectively.

- ❖ 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.
- ❖ 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.



FINAL THOUGHTS

Used K-means clustering algorithm to cluster the tourist spots based on exploring the frequency of the businesses that are present which could help indicate a business opportunity that could be established in the locality.

Future possible research could make use of other significant factors which includes the foot traffic,, accessibility, and average business rates. These above-mentioned factors could help the system make the analysis more accurate.