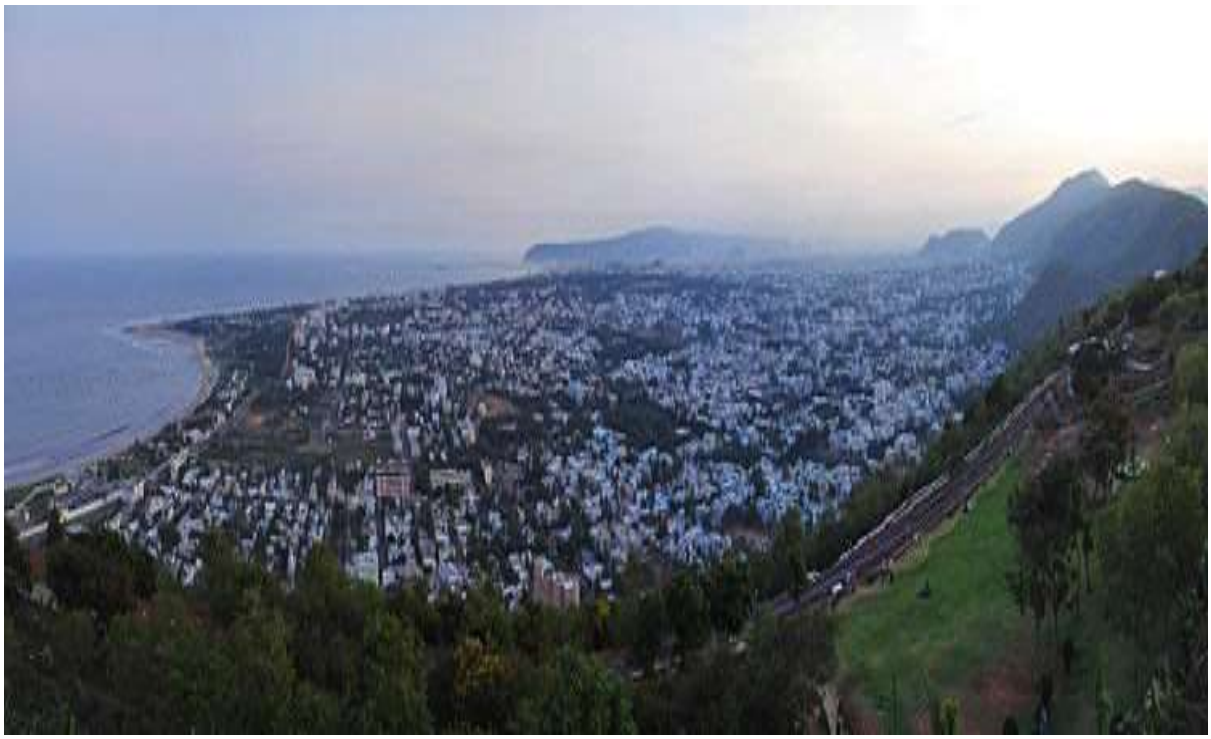
The Battle of Neighbourhoods

IBM Applied Data Science Capstone

Identifying the Better Location to Open a Hotel/Resort in Visakhapatnam, Andhra Pradesh, India

By: Rama Kishore Darlapudi
December 2019

The city of Destination, Vishakhapatnam, Andhra Pradesh, India



Introduction

Vishakhapatnam is one of the fastest growing tire2 cities in Andhra Pradesh and it is named as city of destination for its unique attraction towards the tourism. Since the city is expanding at a faster phase and the tourism has increased a lot in the past 3 to 4 years the city is struggling to provide the accommodation to the increasing tourists at a faster phase. The accommodations are concentrated at a given place and so the commutation is another important aspect to consider while opening a new hotel or resorts. Since the city is on the shore of Bay of Bengal the demand for the beach resorts is also increasing in the recent years. The infrastructure developers also struggling to know the best places to start hotels and resorts which are going to be success in terms of services and getting the better business deals. Also, the city is small in terms of areas and lot of other places in and around this city are also have good places to visit and have a peaceful holiday and vacations.

Business Problem

The key objective of this study is to analyze and identify the best locations in Visakhapatnam city to start a new hotel or resort. Provide a solution to this problem by using tools and techniques of data science and advanced machine learning algorithms and provide insights to the local government officials and infrastructure develops to identify the best locations to improve the hotels and resorts to meet the demand for emerging tourism destination.

Benefited Parties

This project is mainly helping the government officials and infrastructure developers to identify the better locations not only in the Visakhapatnam city but also the places which are close by and safe to the tourists in terms of commutation, safety and reachability to various places of interest.

Introduction to data

In this section the data related to hotels and resorts that are in Visakhapatnam city and the surrounding neighbourhoods within the limits of 10 miles are considered for the study. By clustering the hotels and resorts based on the concentration of the location and neighbourhoods.

Data

The following data is considered for this problem to solve:

1. Since the scope of this project is confines to Visakhapatnam city, considered list of neighbourhoods in and around Visakhapatnam city proposed executive capital Andhra Pradesh state in India.
2. In order to building the meaningful visuals using python the latitude and longitudes of the Visakhapatnam city are considered.
3. Also, the data related to the hotels and resorts that are already exist in the city are considered to build a meaningful cluster using machine learning

Data sources

The sources of the above mentioned data is Wikipedia and the data related to Visakhapatnam city is fetched from URL https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Visakhapatnam contains data for 121 neighbourhoods in Visakhapatnam. Web scraping is used to extract the data from the Wikipedia using python and beautiful soup package. For geographical coordinates data Python Geocoder package is used to get the latitude and longitude for all the neighbourhoods.

The Foursquare API the largest locations data base with more than 100 million places used to get the information and details about the places of interest (Hotels and Resorts in Visakhapatnam). Post gathering the data related to places of interest applied K means clustering to create the homogeneous groups of the places of interest and analyse each of these clusters and provide recommendations.

Methodology

The following are the steps followed to address this problem:

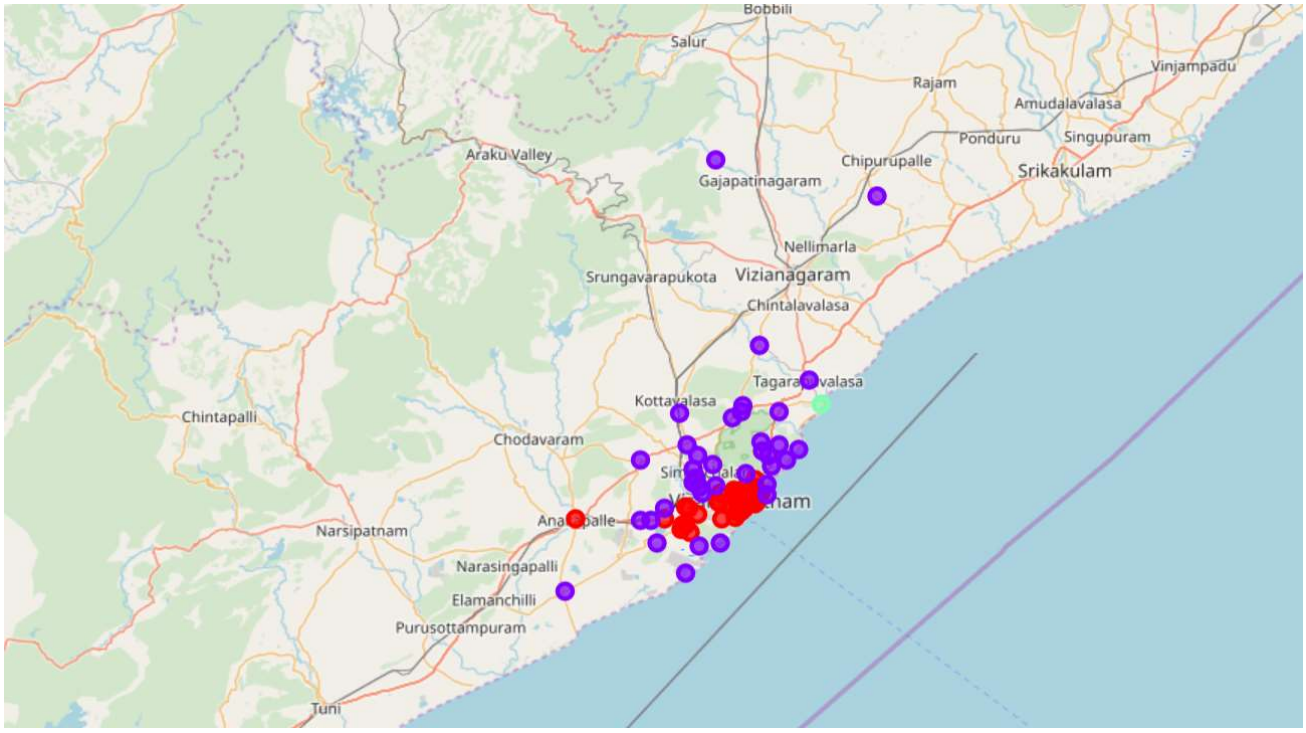
1) Getting Neighbourhood data: First, we need to get the data related to the neighbourhoods of Visakhapatnam city which are available from Wikipedia at URL https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Visakhapatnam.
2) To get this data use web scrapping methodology using the tool request and beautiful soup packages in python. This gives us the list of names of neighbourhoods.
3) Next one has to bring the geo coordinates, latitude and longitude for all these neighbourhoods and this can be achieved by using Geocoder package in python. create a data frame with this data for easy processing in further steps.
4) Plot these coordinates on a location map using folium package to create a visual.
5) Now, by using Foursquare API we would like to bring the 100 venue details around the neighbourhoods withing the radius of 5000 meters. To access Foursquare API, one has to create the client ID and client password beforehand. This provides the information related to venues like, hotels, restaurants, resorts, shopping malls…etc.
6) This also create the JSON file with all the required details about the venues, such as **['Neighborhood', 'Latitude', 'Longitude', 'VenueName', 'VenueLatitude', 'VenueLongitude', 'VenueCategory']**
7) Since our venue of interest for this exercise is Hotels & Resorts, we will filter the data only for Hotels and resorts only.
8) Once we have the relevant data, we can perform the clustering exercised to create the clusters by using hotels and resorts information and assign each of the hotel or resort to a cluster based on the distance.
9) Now analyse each cluster and identify the areas where each cluster are representing and what are the key take away from each cluster.
10) Based on the analysis, provide recommendations to start a new Hotel or Resort what is best suited place or location in and around Visakhapatnam.

Results

The results of k-means clustering are as described below: we have restricted to 3 cluster for easy analysis and provide better recommendations based on the frequency occurrence of hotels and resorts.

The below table gives us the evidence that the concentration of the hotels and resorts in Visakhapatnam are in south and north parts of the city only.

Cluster Number	#Hotels or Resorts	Description
Cluster 0	55	More hotels and resorts are concentrated in North part of the city
Cluster 1	54	More hotels and resorts are concentrated in South part of the city
Cluster 2	2	Very a smaller number of hotels on east part of the city



Discussion

Based on the above results the following are the key discussion points to be considered.

1) The concentration of the hotels and resorts are north and south part of the city.
2) Cluster0: which is in red colour indicates the hotels are very near to the sea shore within main city though the city has a costal belt or sea shore about 30 to 40 kilometres.
3) The second cluster Cluster1 is mainly represents the business areas of the city and so the concentration of high number of hotels and resorts.
4) Concentrating on these two locations is not beneficial for new hotel or resort since these are highly competitive areas.
5) Since the city has a sea shore about 30 kilometres Bhimunipatnam as the second end of the shore one can concentrate on this stretch.

Limitations

1) Only one factor is considered for analysis that is frequency of occurrence of a hotel or resort in the neighbourhood areas.
2) There is a scope to consider other factors such as population, proximity to various places to visit or places of attraction within and outside of the city.
3) Commutation facilities to the tourist places near by to the main city

Conclusion

Based on the study it is very evident that the concentration of hotels and resorts are very near to the business locations or only part of the sea shore. But based on the cluster analysis cluster 2 which is part of Bhimunipatnam and Muralinagar are the best locations to start a new hotel or resort.

References

- Category: Suburbs in Visakhapatnam, India. *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Visakhapatnam
- Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs