The Battle of the Neighborhoods : Visakhapatnam City, INDIA

Applied Data Science Capstone Course by IBM/Coursera

—

Shaik Shahid
121722501020@gitam.in

Introduction/Business Problem

Visakhapatnam is the largest city in the Indian state of Andhra Pradesh, serving as headquarters of the Visakhapatnam district. It lies between the Eastern Ghats and the coast of the Bay of Bengal. The city is set to become the executive capital of Andhra Pradesh once a decentralization bill is enacted in the Andhra Pradesh Legislative Assembly. It is one of the four smart cities of Andhra Pradesh selected under Smart Cities Mission. It is the most populous city in the state, with a population of 2,035,922 as of 2011, making it the 11th largest city in the country and the fourth most populous in South India. It is also the ninth most populous metropolitan area in India, with a population of 5,018,000. With an output of $43.5.billion, Visakhapatnam is the ninth largest contributor to India's overall gross domestic product as of 2016.

Whenever people move to any other place, they explore the place and try to fetch as much information as possible about it. It can be the neighborhood, locality, market, price of the place and many more factors including neighborhood analysis. This can be termed as a request for a search algorithm which usually returns the requested features such as population rate, median house price, school ratings, crime rates, weather conditions, recreational facilities etc. It would be beneficial and nice to have an application which could be made easy by considering a comparative analysis between the neighborhoods with provided factors.

This project helps the end user or the stakeholder to achieve the results which will not only be recommended but also saves a lot of time in manual search. This will indeed save the time and money of the user.

This project can be used by the user at the time of rental apartment or buy house in a locality based on the distribution of various facilities available around the neighborhood. As an example, this project would compare 2 randomly picked neighborhoods and analyse the top 10 most common venues in each of those two neighborhoods based on the number of visits by people in each of those places. Also, this project uses K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category such as restaurants, park, coffee shop, gym, clubs etc. This would give a better understanding of the similarities and dissimilarities between the two chosen neighborhoods to retrieve more insights and to conclude with ease which neighborhood wins over others.

Data Description

Foursquare API:

This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. Many organizations are using to geo-tag their photos with detailed info about a destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform location search, location sharing and details about a business. Foursquare users can also use photos, tips and reviews in many productive ways to add value to the results.

Visakhapatnam City Data :

To get data about Boroughs, PinCode, and Neighbourhood we will scrape this data from a webpage (https://www.indiatvnews.com/pincode/andhra-pradesh/visakhapatnam) to create our own dataset.

Workflow:

HTTP requests would be made to this Foursquare API server using pin codes of the Visakhapatnam city neighborhoods to pull the location information (Latitude and Longitude).

Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods.

Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Visakhapatnam city over an interactive leaflet map.

Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to derive the conclusions.

Data Science Methodology

- The data about Boroughs, PinCode, and Neighbourhood was extracted using web scraping from the url (https://www.indiatvnews.com/pincode/andhra-pradesh/visakhapatnam) to create our own dataset.
- The data consisted of 650+ rows.
- Feature Selection and cleaning was carried out on the data.
- Map for all the Neighbourhoods in Visakhapatnam City was plotted.
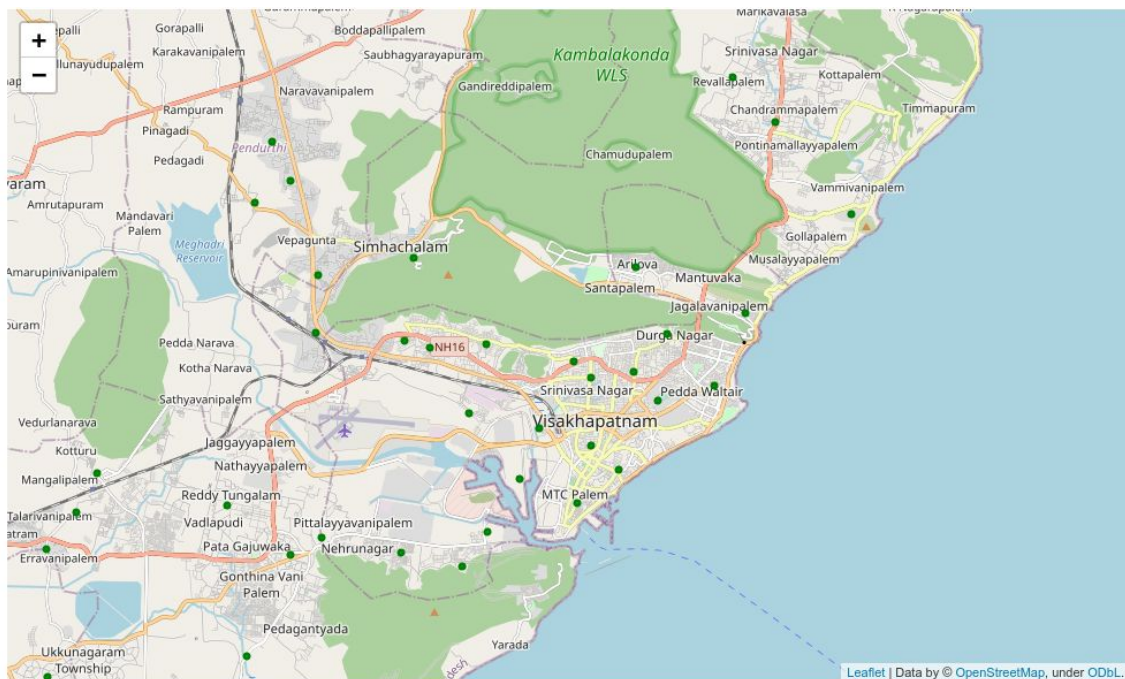
Neighbourhoods in Visakhapatnam City

- Foursquare API was used to find all the venues near each Neighbourhood.
- OneHotEncoding was used to create dummy variables for the VenueCategory as it is a categorical variable.
- The venues were clustered using KMeans ML algorithm.
- The plot of all the Clustered venues and the Neighbourhood was made on the map of Visakhapatnam.
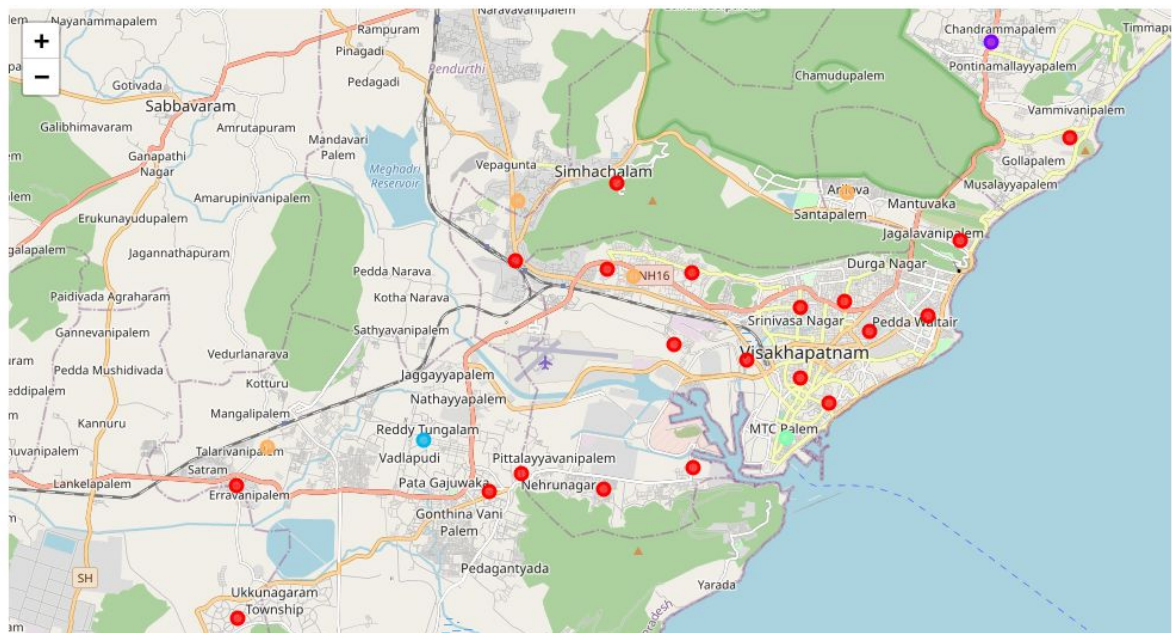
Clustered Neighbourhoods in Visakhapatnam City

Results and Discussions

- Most of the neighborhoods fall into Cluster 1 which are mostly business areas with cafes, restaurants, supermarkets etc.
- Mostly business areas in Cluster 5 are ATMs, Pet services and Pharmacies.
- Most business areas in Cluster 4 are just Harbor / Marina.
- Mostly business areas in Cluster 3 are Business Service and Warehouse Store.
- Mostly business areas in Cluster 2 are just Sculpture Gardens.

Conclusion

With this we can conclude that the people who are looking for new apartment or home for rent or purchase can select the Neighborhoods that belong to Cluster 1 as there are many facilities and business areas.