Image credits: https://worldtraveland.wordpress.com/2012/12/19/secrets-to-enjoying-your-holiday/holidays-beach/

IBM Data Science Capstone Project:
An ideal place to spend a holiday in Visakhapatnam

Jahnavi Mantripragada

Background:

Visakhapatnam is one of the nine districts in the Coastal Andhra region of Indian state of Andhra Pradesh. I am a resident of Visakhapatnam, the city of destiny (as it is well known as).
There are 13 mandals in Visakhapatnam division. 11 each in Narsipatnam, Anakapalle and Paderu divisions. (Data from: https://en.wikipedia.org/wiki/Visakhapatnam_district)

Problem Statement:

To answer the question: **What are the places which are good to spend a holiday?**

Scenario:
Let's consider Ram,when he gets a holiday, he may want to spend it out with his family. So, on an average people would like to visit places where they can find shopping, food, and some entertainment. So, In this project, I am trying to find out all such places. So that people like Ram can plan where to go on a holiday.

Data:

As the data was not available in a csv or excel form I used Wikipedia resources to make a dataset for this project.

Then, I further used geopy and foursquare API:

Geopy: To get the latitude and longitude of the neighbourhoods in Visakhapatnam.

Foursquare API: To get the venues in the neighbourhoods in Visakhapatnam

Data Set:

The Data set consists of 87 Rows and 3 Columns.

The 3 Columns: Neighbourhood, Latitude, Longitude.

Process and Methodology:

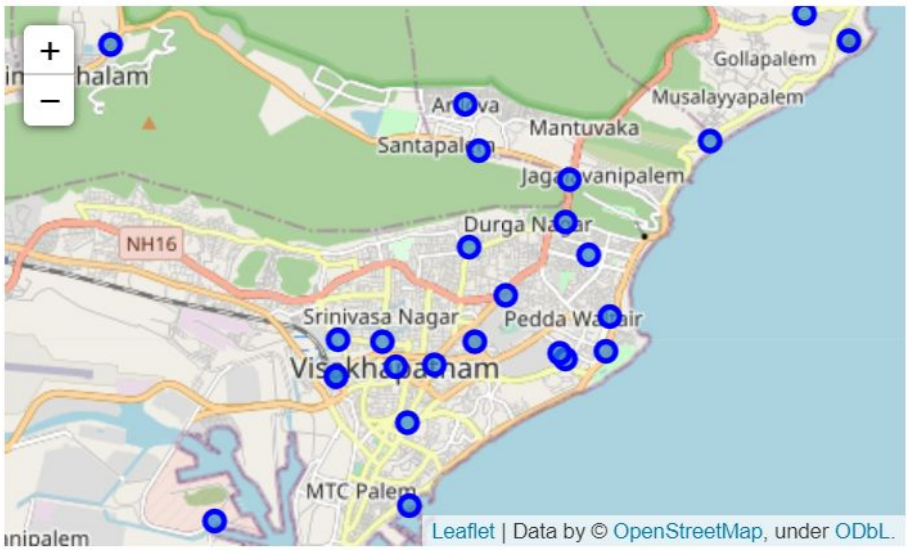
Procedure:

I collected data from different resources. (Web Scraping of Wikipedia, Using Geopy package).

This is how data looks like:

	Neighbourhood	Latitude	Longitude
0	Abidnagar	17.558513	79.134943
1	Adarsh Nagar	30.104309	78.288169
2	Adavivaram	17.777825	83.250659
3	Aganampudi	17.686012	83.135476
4	Akkayyapalem	17.727487	83.299088

And how we use folium to plot the map along with markers (marking the neighbourhoods):



Next, Foursquare API is used to get the nearby venues for each neighbourhood, then we store the venue data based on its category, longitude, latitude.

	name	categories	lat	lng
0	Sangam Sarat Theatre	Indie Movie Theater	17.725508	83.302463
1	Sai Ram Parlour	Indian Restaurant	17.726339	83.303465
2	Pizza Hut	Pizza Place	17.726650	83.305531
3	Pages	Bookstore	17.719736	83.306753
4	Green Park Hotel Visakhapatnam	Hotel	17.715709	83.306315

Next, Merging of the neighbourhoods and their respective venues list is done is make a dataset.

A glimpse:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adarsh Nagar	30.104309	78.288169	Dezloper	30.107285	78.287823	IT Services
1	Adarsh Nagar	30.104309	78.288169	Sitting Elephant Restaurant	30.101748	78.291772	Restaurant
2	Adarsh Nagar	30.104309	78.288169	Ellbee Ganga View	30.101624	78.291794	Hotel
3	Adarsh Nagar	30.104309	78.288169	Rishikesh Market	30.106958	78.292262	Plaza
4	Aganampudi	17.686012	83.135476	Domino's Pizza	17.686542	83.139156	Pizza Place

Then, One hot encoding is done to get to know what all categories are present in each neighbourhood.

On understanding that not all the categories are required as the parameters for the project, some of categories are removed and a new data set is made.

Here in this project, K-Means Clustering is used to cluster the neighbourhoods based on the categories selected. Elbow method is used to get the best K for K-Means Clustering.

To understand more about the data, the 10 most common venues were added.

In this project, wherever there is no data (available), those rows were dropped with a thought(assumption) that they won't be affecting the model.

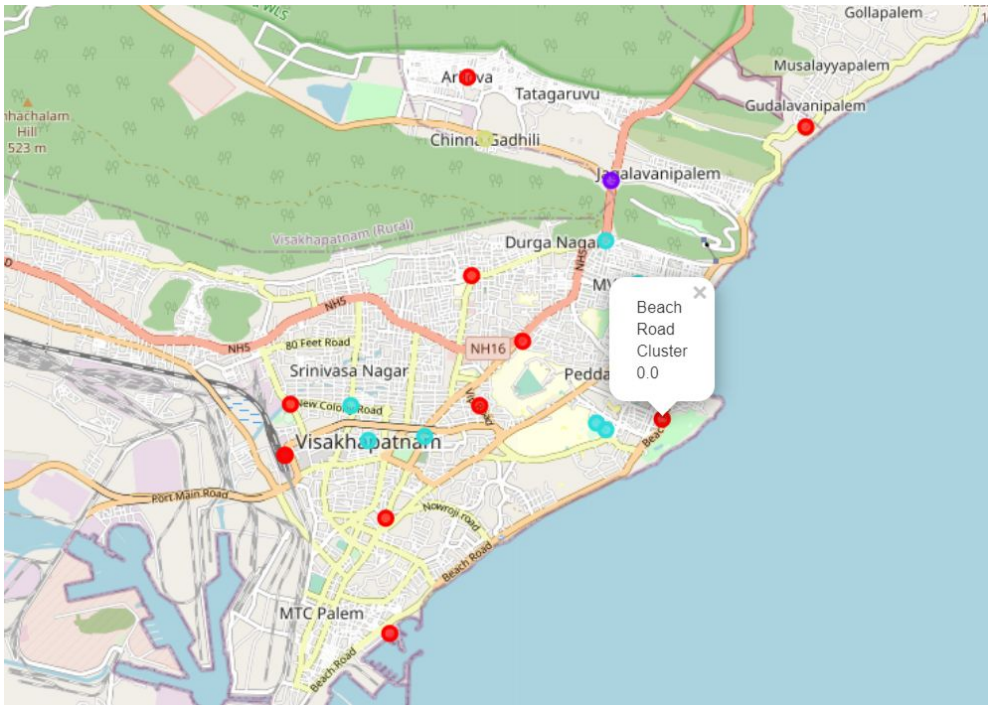
Results

From the elbow method, we found that 4 clusters is the best choice.

Cluster Labels	Neighbourhood
0.0	30
1.0	1
2.0	9
3.0	1

Thus, it can be said that 30 places in Vizag have a cluster of restaurants, coffee shops, movie theatres. [On comparison with the neighbourhoods within Vizag].

On plotting the data from the clusters obtained



The list of neighbourhoods with cluster label 0:

	Neighbourhood	Latitude	Longitude	Cluster Labels
0	Adarsh Nagar	30.104309	78.288169	0.0
1	Aganampudi	17.686012	83.135476	0.0
3	Anakapalle	17.688970	83.003476	0.0
4	Arilova	17.767525	83.313898	0.0
6	Beach Road	17.725861	83.338842	0.0
7	Bheemunipatnam	17.891381	83.451218	0.0
10	Daspalla Hills	17.437015	78.398429	0.0
11	Devipuram	8.557120	76.929593	0.0
12	Duvvada	17.703598	83.151376	0.0
13	Gambhiram	17.879537	83.370890	0.0
15	Jagadamba Centre	17.367304	78.523169	0.0
16	Jalari Peta	17.699563	83.303931	0.0
17	Kailasapuram	17.721431	83.290622	0.0
18	Kakani Nagar	17.738242	83.228431	0.0
19	Lankelapalem	17.687132	83.096176	0.0
20	Lawsons Bay Colony	17.731714	83.339601	0.0
21	Maddilapalem	17.735225	83.320937	0.0
22	Mindi	47.430324	19.455250	0.0
24	Prakashraopeta	17.713747	83.303546	0.0
28	Sagar Nagar	17.761496	83.357269	0.0
29	Seethammadhara	17.743318	83.314492	0.0
30	Sheela Nagar	12.965050	80.202075	0.0
31	Sriharipuram	13.232031	79.766879	0.0
33	Tagarapuvalasa	17.930125	83.425659	0.0
34	Thatichetlapalem	17.727622	83.291263	0.0
36	Vepagunta	17.774427	83.215368	0.0
37	VIP Road	17.727354	83.315422	0.0
38	Visalakshi Nagar	13.023190	80.202751	0.0
39	Waltair Main Road	17.721431	83.290622	0.0
40	Yendada	17.782904	83.374186	0.0

And these according to the data and the model, are the places where one can visit and spend their holiday.

Conclusion/Future Work:

In this, only limited resources were available and used, further by making use of other data resources which may contain more features like the ratings of the places, the more people visit those places.

From this project, One can understand which places have more places to spend a holiday.