CLUSTERING DISTRICTS AND MANDALS IN ANDHRA PRADESH

Sreemukh Mantripragda
B.Tech Second Year, ECE
Natioanl Institute of Technology,
Warangal
Visakhapatnam, AP, India
sreemntr@gmail.com

Abstract— The project focusses on clustering every Mandal in each district in the state of Andhra Pradesh, India and in turn cluster the districts. The factors that are chosen for the clustering are the kind of places that are present in each Mandal based on the data from Foursquare API.

Keywords— Clustering, Folium, Foursquare API, Geocoding, Python, Web Scraping.

I. INTRODUCTION

A. PROBLEM STATEMENT

The idea behind the project is to get a clear picture about each Mandal in a district and about how districts differ in a state like Andhra Pradesh. The clustering of places can be done on various bases, but the one chosen in this project is based on the type and number of establishments within the boundaries of a place. The word establishments, here is used in a broad sense, it includes any sort of a venue for example, ATMs, restaurants, parks, movie theatres, shopping malls and so forth.

B. BACKGROUND

The underlying reason is to find out if the project results have any implications onto the previously calculated demographic measures like literacy rate, M/F ratio, percentage of forest area and so on.

II. METHODOLOGY

The process begins with gathering data required i.e. list of all districts in the state and all mandals in every district. The data used for the project is present in the form of tables in websites listed in the references and data column below. After the data required is scraped from respective websites and cleaned and processed, it is read into pandas Data Frames. The step that follows after names of all districts and mandals in each of them is read into data frames is getting location data for each one of them. For location data, Geopy package is used which retrieves latitude and longitude information about each place. Based on the location data and a pre-defined radius to search for venues, Foursquare API calls are made to retrieve all sorts of venues near a given latitude and longitude within the given radius. Based on the venue data, clusters are identified and labelled. The clustering is done using the KMeans algorithm. We start with number of clusters equal to three. The value of number of clusters can be chosen based on the graph between mean distance of points to the centroid of the cluster (a measure of how similar the points are) vs number of clusters. The number of clusters at which we observe an "elbow" point is to be chosen.

The radius as mentioned above is not a constant for every mandal, but mandals are very similarly distributed within a district. To continue with the process, the mandals are approximated to be uniformly distributed within a district and are considered to be circular with an area equal to area of the district divided by number of mandals. The clusters are colour coded and marked on maps using a plotting library Folium.

III. RESULTS

The clustering process is done for each district. When looking into a district, all the mandals are searched for nearby venues and are clustered on that basis. Then maps are plotted based on clusters identified. An example of two such maps of the 13 districts in Andhra Pradesh is shown in Fig. 1 and 2.
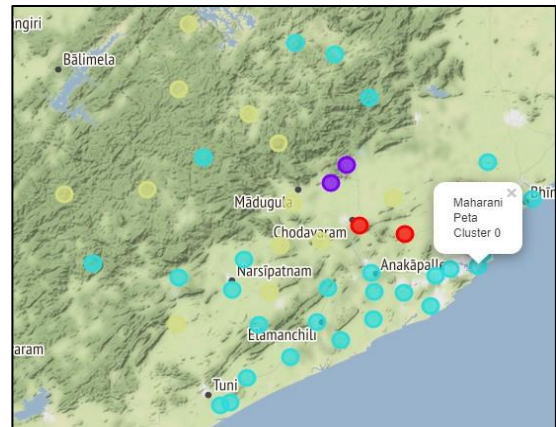

Fig. 1.   Map of Visakhapatnam showing different clusters in colours
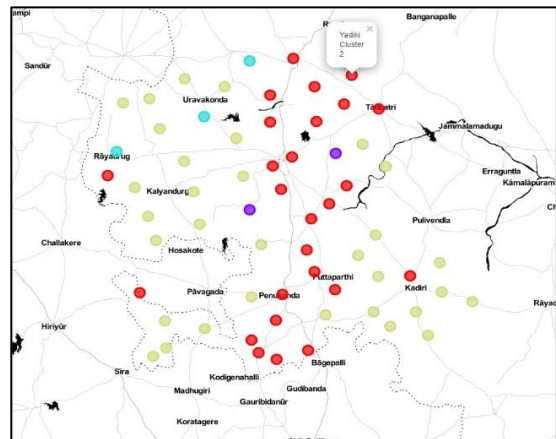

Fig. 2.   Map of Anantapur showing major boundaries and markers

Colour vs Cluster number:

🟥 **2**   🟦 **1**   🟪 **0**   🟫 **-1**

The cluster numbers show two main properties, one is the similarity of all the mandals that are in the same cluster and dissimilarity with those not in its cluster, while the other is that areas with cluster numbers 0>1>2>-1, here the greater than signifies that mandals marked 0 are the best places with maximum possible facilities within a district, similarly cluster1 places are better than the ones in cluster2. The ones in cluster (-1) show the ones that are deprived of resources like ATMs, schools, hospitals etc. Based on venues present in each mandal, the total types of venues and their counts are summed up to get required data for each district, then the districts are clustered based on the available data. Fig. 3 shows the districts clustered and marked on a map.
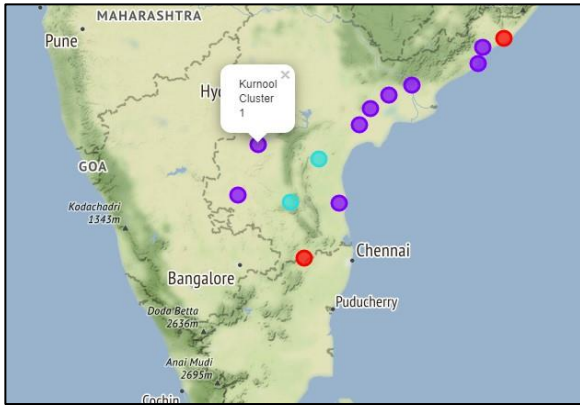

Fig. 3. Map of south India showing clusters of districts in AP

IV. DISCUSSION

The results of the project are summarised in the form a table shown in Table. 1.

Name of the District	Cluster number
Anantapur	1
Chittoor	0
East Godavari	1
Guntur	1
YSR Kadapa	2
Krishna	1
Kurnool	1
Nellore	1
Prakasam	2
Srikakulam	0
Visakhapatnam	1
Vizianagaram	1
West Godavari	1

Table. 1. Cluster number assigned to each district in AP

The results here can say nothing about how the lifestyles would be across different districts, but they show us how similar two places in a same cluster are. The structures of the places within the districts in the same clusters are very similar.

The key takeaway point from the project is that if a certain kind of a proposal or a project worked out fine in any one of the districts, then it has a fair chance that the same response or outcome occurs in other districts within the same cluster. The same applies to different mandals within a district. For the clusters of mandals in each district, the clusters show a clear relationship with the geographic conditions also, for example, places that are on high altitudes (seen as darker green in the Fig. 1) are scarce of facilities that available on the planer regions and hence are in cluster -1.

V. DATA

The data used in the projects is taken from various sources in the internet.

[1] https://simple.wikipedia.org/wiki/List_of_districts_in_Andhra_Pradesh
[2] https://en.wikipedia.org/wiki/Anantapur_district
[3] https://en.wikipedia.org/wiki/Chittoor_district
[4] https://en.wikipedia.org/wiki/East_Godavari
[5] https://en.wikipedia.org/wiki/Guntur_district
[6] https://en.wikipedia.org/wiki/Kadapa_district
[7] https://en.wikipedia.org/wiki/Krishna_district
[8] https://en.wikipedia.org/wiki/Kurnool_district
[9] https://en.wikipedia.org/wiki/Nellore_district
[10] https://en.wikipedia.org/wiki/Prakasam_district
[11] https://en.wikipedia.org/wiki/Srikakulam_district
[12] https://en.wikipedia.org/wiki/Vizianagaram_district
[13] https://en.wikipedia.org/wiki/Visakhapatnam_district
[14] https://en.wikipedia.org/wiki/West_Godavari_district

VI. CONCLUSION

The project can be further extended by taking into account of more factors for deciding clusters. Factors such as temperature, humidity, land to forest ratio, population, density etc. can be added to improve the clusters' meaning. The work currently done gives an estimate of how similar the places within the same cluster are.