Clustering Cities in INDIA

Aneel Kumar

May 18,2020

1.Introduction

India is the second largest country in the world after China .There are 4000 cities and towns in India .About 300 cities have population over 1,00,000 Seven cities have population more than 3 million .Greater Mumbai still is the most populated city in its 440sq.Km. area followed by Delhi, Kolkata ,Bangalore and Chennai.

In this project over 150 cites data are collected from different web sources and use it for clustering cities in India. These cities are clustered on basis of the venues in the city. By this we can show same kind of cities on basis venues in it.

2.About the data

I obtained all cities data from a Wikipedia page titled "List of cities in India by population" .I extract this data using read_html function of pandas.

Using this cities list I obtain its longitude and latitude data using Geolocator.

Now using this longitude and latitude data I use get requests to Foursquare api calls and get data of venues in each city.

This data is cleaned and changed to categorical values.

3.Modeling

Here I am clustering cities in India. So for that classification I used K-means clustering algorithm and clusted cities into 5 clusters. These Clustering is on basis of venues in the cities. And I used Folium to visualize the data ie., cities and colour representing the cluster is produced on map.

4.Result

These cities are Clustered into 5 clusters and these cities are pointed in map using Folium. Top venues in each cluster was extracted .

5.Conclusion

In this project we obtained data from various data sources like Foursquare and cleaned it and categorized . This data is clustered using K-means algorithm and data is visualized on maps using Folium.