Medical Tourism: Clustering hospitals based on the neighborhood characteristics

Data

In order to solve the above problem of clustering the hospitals based on their neighboorhood characteristics. The following data is needed:

1. Name of the Hospital
2. Address of the Hospital
3. Popular venues in the hospital locality

This study focused only on Indian hospitals. The data regarding the name and address of the hospital is obtained from the following website:

- https://www.treatmenttraveller.com/top-hospitals-in-India-2018

This website is selected because it maintains the list of top hospitals in India that are suitable for medical tourism. The study extracts the following information namely name of hospital and address of the hospital from the website using the web scraping facility provided by the *Beautifulsoup* package in Python. The data is processed and formatted to provide a sample dataframe as shown below:

Hospital_Name	Address
Medanta Hospital	Sector 38, Gurugram, Haryana
Apollo Hospital	Jubilee Hills, Hyderabad, Telangana
Rabindranath Tagore International Institute of Cardiac Sciences	Mukundapur, Kolkata, West Bengal
Narayana Health	Hsr Layout, Bengaluru, Karnataka
Continental Hospital	Gachibowli, Hyderabad, Telangana

The abovementioned dataframe containing hospitals is used for extracting the data regarding the popular venues. The major characteristics extracted regarding popular venues are namely name of venue, latitude and longitude of venue and category of venue. The Foursquare is one of the websites that maintains such data regarding the venues or facilities present in an area. Their API could extract the popular facilities in the area with its characteristics. Hence

in study, the data regarding popular venues in an area with its desired characteristics is obtained from following website:

- https://www.foursquare.com

The foursquare API needs the latitude and longitude of the point around which the venues are searched. Hence, the latitude and longitude of the area in which the hospitals were present was extracted using the *geopy* package available in the Python. The resulting sample dataframe is as shown below:

	Hospital_Name	Address	Latitude	Longitude
0	Medanta Hospital	Sector 38, Gurugram, Haryana	28.4351	77.0404
1	Apollo Hospital	Jubilee Hills, Hyderabad, Telangana	17.4308	78.4103
2	Rabindranath Tagore International Institute of Cardiac Sciences	Mukundapur, Kolkata, West Bengal	22.4956	88.3984
3	Narayana Health	Hsr Layout, Bengaluru, Karnataka	12.9127	77.638
4	Continental Hospital	Gachibowli, Hyderabad, Telangana	17.4436	78.352

The extracted data of venues in the hospital locality is used to segment and cluster the hospitals in India. This will help in identifying the similarities and dissimilarities between different localities. These clusters will help any potential patient to select hospital based on its preferred locality characteristics.