IBM DATA SCIENCE APPLIED CAPSTONE PROJECT

Open a New Bar in Bengaluru,Karnataka,India

By :Prashant Kambi

Business problem

To find out best location or neighborhood in capital city of Karnataka, Bengaluru(Bangalore) , India. Bengaluru is becoming a hub of pubs.

With college students and travelers filling the city of Bangalore, the capital of Karnataka, it is increasingly becoming one of the top cities for party goers.

We try to find out the best places for someone who is looking to open a bar in Bengaluru.

Target Audience

The project is useful for investors who are looking to open or invest in opening new bars in city(Bengaluru).

The project would briefly explain with valid reasons to all types of investors whether be small scale or large scale and then recommend a best place according to the needs of investor.

Data

The following data will be used to solve the business problem

1 . List of areas or neighborhoods in Bengaluru.

2 . Location of neighborhoods (latitude and longitude)

3. Venue category (data about nearby centers of particular venue)

Source of Data

Web-scraping:

This is 'http://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore' webpage from which data scrapped using python modules like beautifulsoup and requests.

Only the required data i.e.neighborhood name is extracted from the webpage and then their corresponding location(Latitude and longitude) is retrieved from geocoder package in python.

Venue categories :

We use foursquare API to get venue data for neighborhoods.This API helps us to retrieve categories of venue data out of which we are interested in "BAR" category.This will help us solve the business problem.

Methodology

1.Web-scraping wikipedia page to get neighborhoods list

2.Get latitude and longitude coordinates using geocoder

3. Use foursquare API to get venue data

4.Group data by neighbourhood and take mean of frequency of each category.

5.Filter venue category by “BAR”

6.Perform Clustering using data by K-means

7.Visualise clusters on map using folium

Results

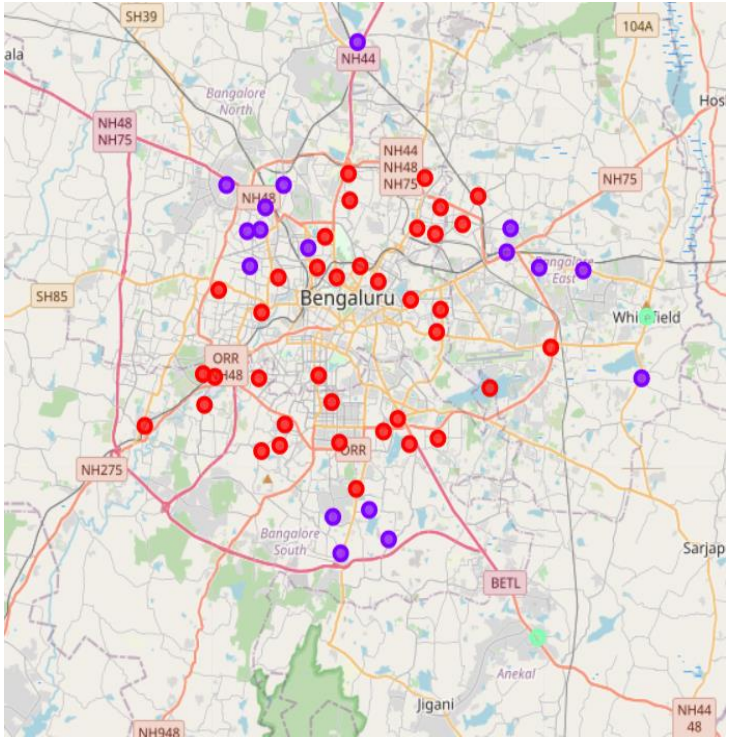
There are 3 clusters formed

Red clusters(cluster 0) are those with least number of bars

Purple clusters(cluster 1) are those with moderate number of bars

Light green(cluster 2): cluster 2 is satisfactory as it agrees with real world Electronic city Whitefield has more bars

Electronic city and Whitefield are neighborhoods where more IT companies more concentrated



Conclusion

Bars in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of bars

On the other hand, cluster 0 has very low number to totally no shopping mall in the neighborhoods

Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 1 with moderate competition

Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhoods in cluster 1 with little competition.