IBM Applied Data Science Specialization Final Project

Recommending location for opening pub in the city of Bangalore.

Introduction

This project is intended to recommend places where pub’s can be started in the city of Banglore, India.
Banglore is place inviting several startup’s in India. There are sevral corporate sectors moving to Banglore. Pub could be a very beneficial business, if opened in Banglore.

Business Problem

Opening pub’s in the city of Banglore can make a good scope of business. It is very attractive place for lot of multi national companies and startups. Millions of people move to Banglore to enhance their career and reside in Banglore, thus providing evergreen scope for the business of pub culture.In project we will figure out which cities in Banglore have more pubs and which location don’t have many, where it is recommended for a pub opening.

As the target place attracts lot of people, mostly the target audience would be the people who want to open a pub in the city of Banglore. This project suggest the places where one can start their pub with least competition and hindrance in the market.

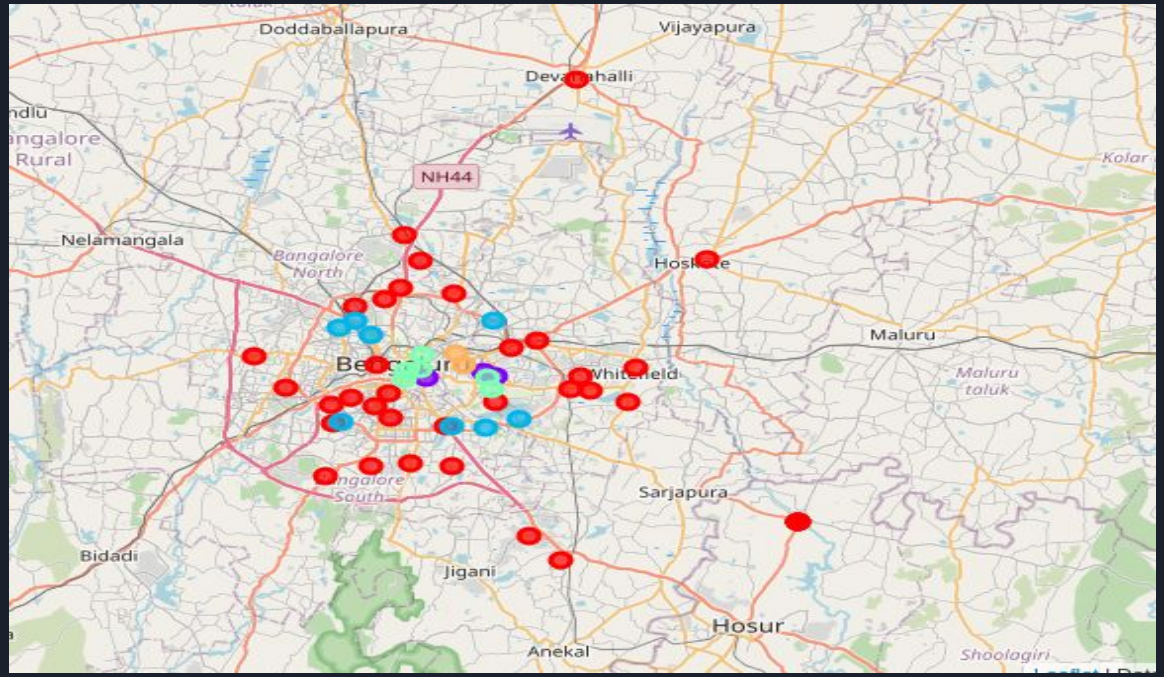
Data Source and Extraction Method

The list of cities has been extracted from Wikipedia using following link. (https://commons.wikimedia.org/wiki/Category:Suburbs_of_Bangalore). Cities data has been scraped from the above link of wikipedia using Beautiful soup.
Location coordinates for the above obtained cities is gathered from geocoder package.
The venues information for this cities has been gathered from foursquare developer api, which provides all the information regarding the neighborhood of a city

Methodology

First we are concerned on finding the list of cities in Banglore and
then we find all theneighborhoods and venues using FoursquareAPI. We are using Folium for data visualization.
Then we find all the coordinates of the cities using Geocoder.
Then we filter the data having VenueCategory as Pub.
Then grouping the data based on the VenueCategory Then we merge cities data with coordinates data and then use the data for Clustering.
Kmeans Clustering algorithm is used for clustering the banglore
cities data. We cluster data into 5 clusters.

Result


Doddaballapura
Vijayapura
Devanahalli
Rural
NH44
Nelamangala
Bangalore North
Maluru
Whitefield
Maluru taluk
Sarjapura
Bidadi
Jigani
Hosur
Anekal
Shoolagiri

Results

Based on the clustering results.Complete data is seperated into 5 clusters and based on the clusters, it is clear that cluster 3 is best for opening the Pub as it has only one pub.

Cluster 0 : 37 datapoints
Cluster 1 : 3 datapoints
Cluster 2 : 8 datapoints
Cluster 3 : 6 datapoints
Cluster 4 : 2 datapoints

Cluster 0 : Red Colour
Cluster 1 : Blue Colour
Cluster 2 : Purple Colour
Cluster 3 : Green Colour
Cluster 4 : Orange Colour

Clearly from data cluster 4 is good to open the pub.