Introduction

Book lovers in India are always on the lookout for a nearby library where they could go, relax and read . Libraries provide us with a plethora of opportunities to explore and tickle our creativity. They should be located at places where they can be easily accessible to the public as well as provide a conducive environment for reading.There are a few public libraries in Hyderabad(India), but as a business decision for opening a public library needs a lot of consideration. The location specially needs to be considered so that it is accessible not only to the school and college going students but also to their parents and the elderly.

Business Problem

The business problem that this capstone project will be looking at is to find an appropriate location for a book lover/investor to open a library for the public. Using Data Science methodologies this project aims to provide an answer to the question : What is the recommended place to open a public library in Hyderabad?

Target Audience

This project is useful to investors/book lovers who are interested in opening and a public library in the city of Hyderabad, India.

Data

In order to get this problem solved we will need the following data:

- The list of neighbourhoods of Hyderabad. This can be obtained from scraping the wikipedia page.
- The latitudes and longitudes of those neighbourhoods. This is required to plot the map. This can be done using the geocoder.
- To perform the clustering of neighbourhoods we will also need the data related to libraries. In order to obtain this we will need to use the Foursqaure API(Foursquare Venues & Places Database) and retrieve the library category from it.

Sources of data and how to extract them

This wikipedia page(https://en.wikipedia.org/wiki/Category:Cities_and_towns_in_Hyderabad_district,_India)has the neighbourhoods of Hyderabad. We will be using web scraping to extract all the data with the help of Python and beautifulsoup packages.
Then we will get the latitudes and longitudes of the neighbourhoods using the Python Geocoder package.
After this we will use the Foursqaure API to get the venue data of the neighbourhoods. We are interested in the category library of the venue data.
This project in a way will help us make use of data science skills, web scraping, data wrangling, using Foursqaure API, clustering and creating maps for visualization.