Applied Science Final Assignment

Business Proposal Analysis for opening a Training Centre in India

VAISHNU BALAKRISHNAN SUGUMARAN

Business Problem:

The objective of this project is to analyze and select the best locations in India to open training centers. This project is mainly focused on the analysis of 2011 census data to understand where there is a need to start- up educational institutions. Using data science methodology and machine learning techniques like clustering and data visualization, this project aims to provide solutions to low literacy rates in India

Data:

 To solve the problem, we will need the following data: • 2011 district census data. This dataset contains data like literacy rate, population, sexratio and growth rate. • Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data. • Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods. Sources of data and the methods to extract the data I got the 2011 district census data from the official Indian census website. I have used web scraping techniques to extract the data from the census page, with the help of Python requests and beautiful soup packages. Then we can get the latitude and longitude coordinatesof the neighborhoods using Python Geocoder package. After that, I have used the Foursquare API to get the venue data for those neighborhoods. Foursquare API will provide many categories of the venue data, and we are particularly interested in the residence category in order to help us solve the problem. This is a project that will make use of many data science skills, from web scraping (census data), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).