CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE CAPSTONE

INTRODUCTION:

Bangalore, officially known as Bengaluru is the capital of the Indian state of Karnataka. It has a population of over 10 million, making it a megacity and the third most populous urban agglomeration in India. It is located in Southern India, on the Deccan Plateau at an elevation of over 900m (3000ft) above sea level, which is the highest among India's major cities. It is multi-ethnic, multi-religious and cosmopolitan in nature.

BUSINESS PROBLEM:

I am currently pursuing my engineering from RNS Institute of Technology, which is a college located in Bangalore. Lots of student studying in this college stay away from their homes during their 4 years of tenure of study. Being a student, I would like to reside in the neighborhood of the college. The main aim of this Capstone Project is to explore the venues of each neighborhood of this college and look for the neighborhoods having ample hostels, malls, restaurants, departmental stores, book stores and transportation etc. This analysis may help students who wish to reside nearby the college campus.

DATA:

I downloaded the required data from a website published by Government of India. Here is a link to the website: https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude

This csv file contains the list of pin codes of all cities of India, officenames (post office), the head office (HO) names of each officename etc. The Latitude and Longitude columns are incomplete for most areas. So, I will try to fill it with the help of Geopy. I am using Foursquare API to get the venues of each neighborhood of the institution.

I have downloaded the required csv file and loaded this data into the project to be used as a data asset. I have then converted the csv file to pandas dataframe and displayed the head of the dataframe below. Further data formating steps are given below under the heading of "Formating and Cleaning Data".

FORMATING AND CLEANING DATA:

1. The data pertaining to Districtname='Bangalore' is needed, so I will filter only the rows having Districtname = 'Bangalore'.
2. Since, RNSIT's Related HeadOffice is "Basavanagudi", only the data pertaining to this condition is kept.
3. Next, we will keep only the columns officename, pincode, longitude, latitude.
4. Multiple occurrences of same pincode are checked.
5. Rename the column 'officename' as 'neighborhood' and combine the neighborhoods having the same pincode with a comma (,).