IBM Data Science Capstone Project

Week 4

Aman Kumar

“ I have taken my first steps towards acquiring skills related to data science by doing the IBM Data Science Professional Certificate course on Coursera. The last module of this course is a capstone project. This project is about using a data science toolset on a real-life problem and demonstrating the creation of value by applying the learned skills. I present here the summary of my project and the findings. The analysis was performed in Python. “

1. Introduction

Bengaluru, India is the IT hub of India. It is multicultural. It provides a lot of business opportunities and a business-friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, etc.

As it is highly developed, there is a huge problem in waste management in the city. Many organization and individual want to recycle and reduce carbon footprint but in most of the cases, they can’t find proper resources.

2. Business Problem

The problem I’m trying to solve is to establish a recycling center in strategic places inside the city so that the maximum volume of the waste gets recycled also providing revenue to the organization at the same time.

3. Data Used

There are two approaches to find the solution for this problem. The first is to target the residential areas as waste is generated in large volume from there also, but due to the limitations in data available for the Bengaluru city, it is not possible. Second to target the areas where there is a maximum number of businesses.

For this first, I have scrapped the Wikipedia page which has a list of all the neighborhoods in the city, and then I have collected the coordinates of each neighborhood using geocoder and put that in a pandas data frame.

Methodology

1. Web Scraping Wikipedia page to get the list of neighborhoods in Bengaluru city and storing it as pandas Dataframe.
2. Adding coordinates of each neighborhood in a data frame using geocoder.
3. Plotting all the neighborhoods on the folium map.
4. Getting a list of venues present in each neighborhood using Foursquare.
5. Counting and putting it in a data frame with neighborhood present with a maximum number of venues.
6. Plotting the same in a folium map.
7. At last, knowing where should be a recycling collection center should be opened for maximum efficiency.
8.

Resources

Wikipedia page:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

Link to notebook:

https://github.com/aman-kumar-004/Coursera_Capstone/blob/main/W4_IBM_Data_Science_Capstone_.ipynb