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.

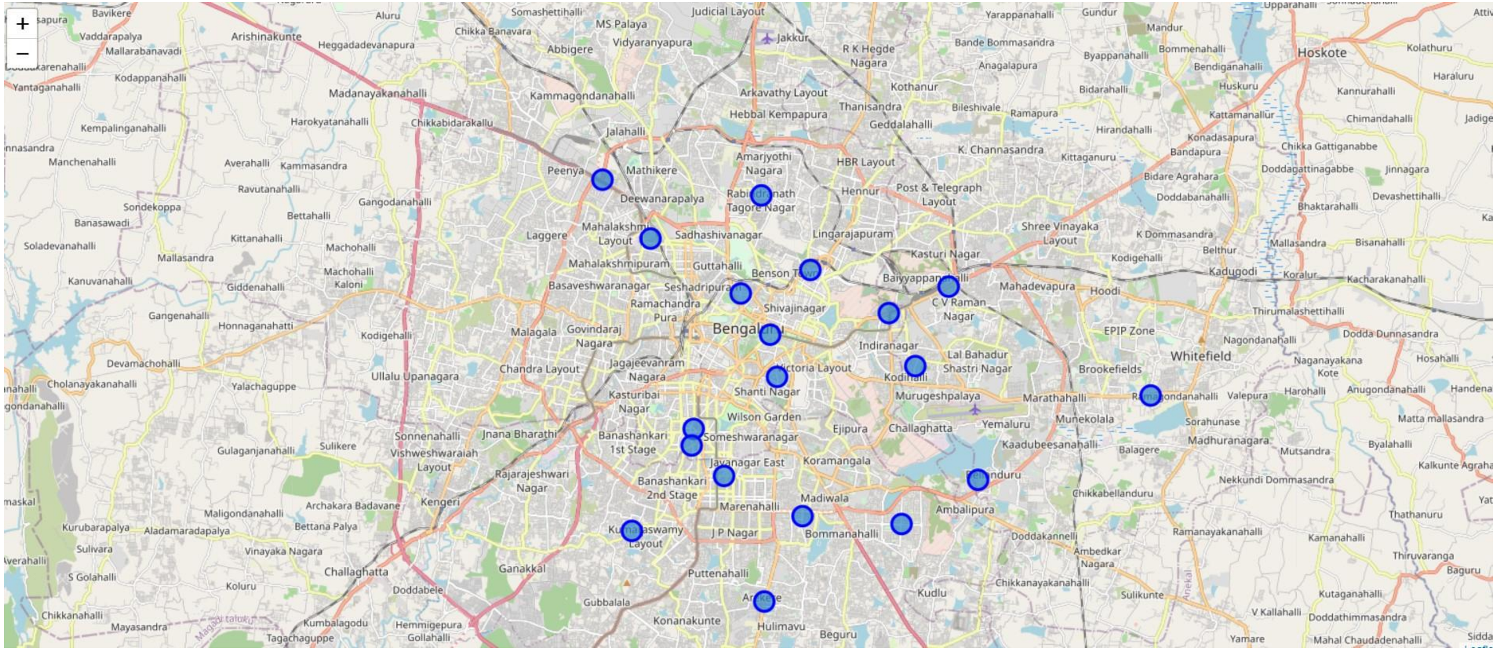
Result

List of places at which recycling center that should be opened.

Neighborhood	Number of Venues	Latitude	Longitude
Milk Colony	50	13.00966	77.55683
Jayanagar, Bangalor	41	12.92872	77.58281
Fraser Town, Bangal	33	12.99894	77.61276
BTM Layout	29	12.91495	77.61001
Brookefield	28	12.99333	77.66123
Statue of Queen Vic	19	12.97675	77.59879
Gandhi Bazaar	18	12.94478	77.57213
HSR Layout	17	12.91222	77.6447
Bellandur	16	12.92734	77.67169
Richmond Town	15	12.96233	77.60123
Vasanth Nagar	15	12.99073	77.58861
Binnamangala, Bang	15	12.98403173	77.64047977
Basavanagudi	14	12.93898	77.57137
Arekere	12	12.88568	77.59668
R. T. Nagar	12	13.02445	77.5959
Yeswanthpur	12	13.02955	77.54022
Siddapura, Bangalor	11	12.95613	77.73196
Chikkalasandra	11	12.90995	77.55051
Kodihalli, Bangalore	11	12.9662	77.64982

Result

Map of those places



Discussion

This project meets all the requirements of the assignment but still, there is room for improvement. I went with the “venue available in each neighborhood” approach but one can also approach it from the “residential density” approach where they can analyze the high-density areas and open recycling plants.

Also due to the limitations in API calls from foursquare I have to limit the radius of each neighborhood. It can be improved with premium API calls.

Conclusion

In this Capstone project, I started with collecting data from different sources, I learnt how to perform scrapping. With data analysis, we came to know the strategic places ideal for opening recycling centres and finally plotting it on a map for visual understanding.

I enjoyed and learnt so many new skills during this project.

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