Living Next Door

Get the best neighbourhood

TABLE OF CONTENTS

Project Guide : Dr. Shriram Sir

Our Team :

Introduction

Problem Statement

Our friend is going to be moving in to the city of Bangalore, located at the edge part of the Indian subcontinent. He needs to find suitable housing. However, he frets. He wants to search for a good neighborhood that can provide him easy accessibility to daily use facilities, good pricing of houses and very less crime rate.He has come to us for help, asking us to analyze the different areas in the city of Bangalore and find which neighborhoods would be good for a newbie like him.

So as a team we have decided to help Mr. Nolan using a bit of data and a bit of science.

Objective

This project aims to utilize all the concepts learned through Data Science and find a good neighborhood in Bangalore. We define a domain problem, the data that will be used and using this we are able to analyse it by applying various ML tools. Analysis of data is step by step procedure, involves Data Gathering, Data Cleaning, Exploratory Data analysis, Data Modelling and interpreting (final output). In this project we will go through all these process and provide a conclusion that can help the people to to find a good and safe place for them to live in while moving to a new city..

Block Diagram Of The System

Data Collection

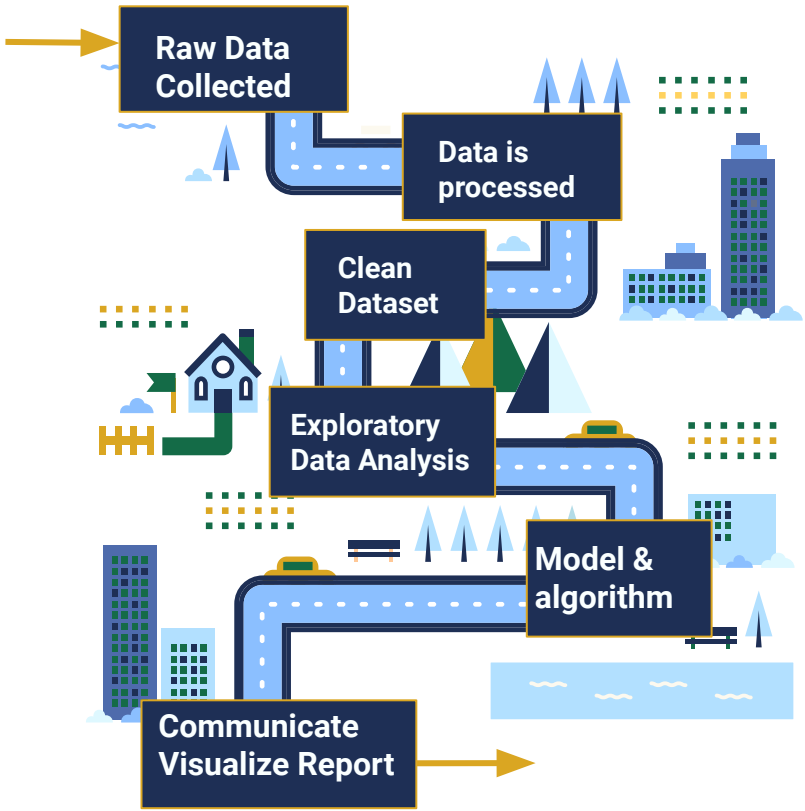
Data Collected are from APIs and different datasets

Data Processed

Data collected is translated into usable information

Data Cleaning

To fix or remove incorrect, corrupted, incorrectly formatted, complete data within a dataset



Exploratory Data Analysis (EDA)

Used to to analyze and investigate data sets and summarize their main characteristics

Model & Algorithm

By using different machine learning algorithms, taking insights from data

Data Visualization

Translating information/data into a visual context and making decisions

Dataset

a combination of CSV files that have been prepared for the purposes of the analysis from multiple sources

1) The list of neighbourhoods in Bangalore (via wikipedia done by web scraping).

2) The Geographical location of the neighbourhoods (via Kaggle & Geocoder package)

3) Venue data pertaining to venues (via Foursquare API)

	Neighborhood	Latitude	Longitude
0	Agram	45.813177	15.977048
1	Amruthahalli	13.066513	77.596624
2	Attur	11.663711	78.533551
3	Banaswadi	13.014162	77.651854
4	Bellandur	58.235358	26.683116
...	...	...	...
347	Virupakshipura	13.024075	76.469658
348	Vishwanathapura	13.273529	77.649099
349	Yadamaranahalli	12.427249	77.379083
350	Yadavanahalli	12.789855	77.751454
351	Yeliyur	12.509896	76.828661

```
address = 'Bangalore, Karnataka'

geolocator = Nominatim(user_agent="bangalore_explorer
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinate of bangalore are {
```

```
The geographical coordinate of bangalore are 12.9767936, 77.590082.
```

	Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0	Air Force Hospital	12.964027	77.627500	GOLDEN ACES	12.962327	77.630772	Casino
1	1	Air Force Hospital	12.964027	77.627500	inox	12.966699	77.624227	Multiplex
2	2	Amruthahalli	13.066513	77.596624	Reliance Fresh	13.066264	77.596961	Convenience Store
3	3	Amruthahalli	13.066513	77.596624	Bata	13.065699	77.599402	Shoe Store
4	4	Amruthahalli	13.066513	77.596624	Print Central	13.062902	77.596335	Print Shop

Dataset Description

This dataset provides us a look on crime in bangalore

The csv file of this dataset has been chosen from kaggle which will provide the list of neighbourhoods in Bangalore with their crimes.

	Police Station	Year	Type	Date	Time	Place	Latitude	Longitude
0	Banashankari	2016.0	Murder	29/07/2016	21:00:00	22nd Cross, Jayanagara 6th Block Bangalore	12.933	77.5838
1	Banashankari	2016.0	Murder	13/11/2016	23:45:00	On the Road of 30th Cross, Opp Vinayaka textti...	12.9304	77.5498
2	Banashankari	2017.0	Murder	2017-10-05 00:00:00	21:00:00	No. 282, 36th Cross Jayanagara 7th Block Banga...	12.9237	77.5818
3	Banashankari	2017.0	Murder	19/07/2017	19:30:00	Near Transfaram, Opp Amejan, Kanakapura Main R...	12.9241	77.5769
4	Banashankari	2018.0	Murder	21/06/2018	23:30:00	No. 53, 2nd Floor, 8th Main 4th Block Jayanaga...	12.9244	77.5827
...	...	...	...	...	...	...	...	...
9236	V V Puram	2019.0	Assault	2019-02-09 00:00:00	21:30:00	No, 107, 3rd cross, Manjunh colony, Bsk 3rd st...	12.9316	77.562
9237	V V Puram	2019.0	Assault	2019-03-10 00:00:00	10:45:00	No, 33, Kaverinagar, Khriguppe,Bsk 3rd stage, ...	12.9279	77.563
9238	V V Puram	2019.0	Assault	15/10/2019	17:30:00	Near Mayura Bar, Bhuvneshwarinagar, Bsk 3rd st...	12.923	77.549
9239	V V Puram	2019.0	Assault	13/12/2019	23:30:00	No, 10/27, 4th Main, 10th cross, Srinivasanaga...	12.9337	77.5569
9240	V V Puram	2019.0	Assault	20/12/2019	18:00:00	No, 119, 3rd cross, 5th main,Bsk 3rd stage, be...	12.9278	77.5439

Dataset Description

This dataset provides us the prices and other details about the type of houses with their basic structural information.

- The csv file of this dataset has been chosen to provide the list of neighbourhoods in Bangalore, the Geographical location of the neighbourhoods (via Kaggle & Geocoder package)
- To make it more user-friendly we tried to represent the results with the help of graphs(via. matplotlib package)

```
[ ] data2 = pd.read_csv('/content/drive/My Drive/DATASETS_CAPSTONE/Bengaluru_House_Data_Price.csv')
    data2.head()
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

Module 1 :

The master data's main components were Neighborhood, Latitude and Longitude as depicted below for the city of Bangalore.

The data for neighborhood and their latitude and longitudes was scraped from wikipedia pages using bs4.

Using Foursquare, individual neighborhoods are searched to find nearby venues and their categories within a 500m radius of a randomly chosen neighborhood, Banaswadi.

One hot encoding was then carried out to group more than 200 venue categories by neighborhood.

To make the analysis more interesting, we wanted to cluster the neighbourhoods based on the neighbourhoods

Module 2 :

- This module is our Price Prediction Module. In this module we will sort and arrange the locations according to the prices .

- The Kaggle dataset for prices of all location available to us provided all the information about the neighbourhood, house details, area of the house and the calculated price per square ft. along with the number of rooms in the various house locations present in Bangalore.

- To reach to the final result we performed a bit of data cleansing to remove duplicate and null values. The neighbourhoods are then grouped by the name of the neighbourhood and type of area, so data clustering is made easier later on.

Module 3

The data contains information on crime rate for the city of Bangalore.

Plotting of data using bar is done to show the crime reported and to show the highest number of cases reported.

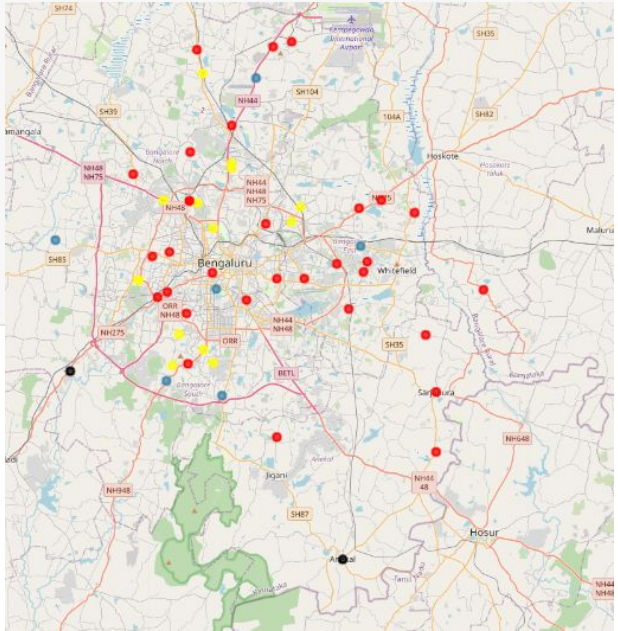
After that line graph is plotted to show the crime rate in different areas. This gives us a clear picture of which areas to avoid.

Module implementation and Results

Result from Module 1:

```
[ ]
# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(bangalore_merged['Latitude'], bangalore_merged['Longitude'], bangalore_merged['Neighborhood'], bangalore_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=colors[int(cluster)],
        fill=True,
        fill_color=colors[int(cluster)],
        fill_opacity=0.7).add_to(map_clusters)


map_clusters
```



Module Implementation and Results :

```
data2i.isnull().sum()
```

```
area_type          0
availability       0
location           1
size              16
society         5502
total_sqft         0
bath              73
balcony          609
price              0
dtype: int64
```

```
[13] data2 = data2i.dropna()
     data2.head()
```

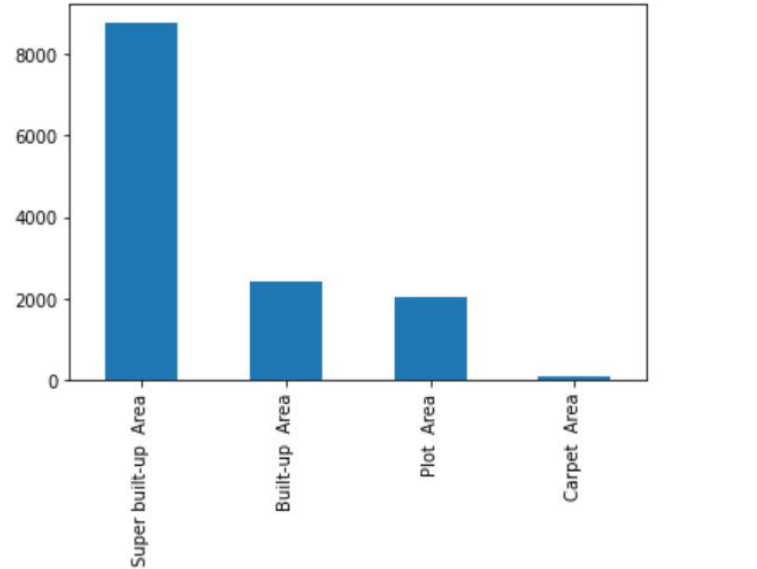
	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
5	Super built-up Area	Ready To Move	Whitefield	2 BHK	DuenaTa	1170	2.0	1.0	38.00

0s completed at 11:04 PM

Type here to search

```
[ ] data2.area_type.value_counts().plot(kind='bar')
```

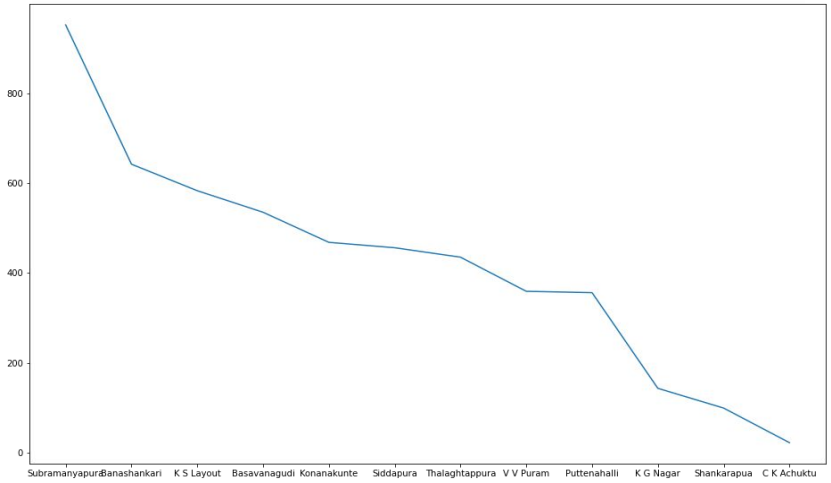
```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff9a3574cd0>
```



Module implementation and Results

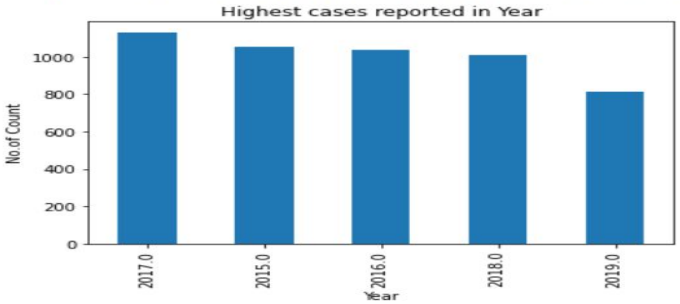
Result from Module 3:

```
plt.xlabel('Police Station')
plt.ylabel('No.of Count')
plt.title('Highest cases reported in police station')
plt.figure(figsize=(16, 10))
plt.plot(cdf2[cdf2.columns[0]].value_counts())
plt.show()
```



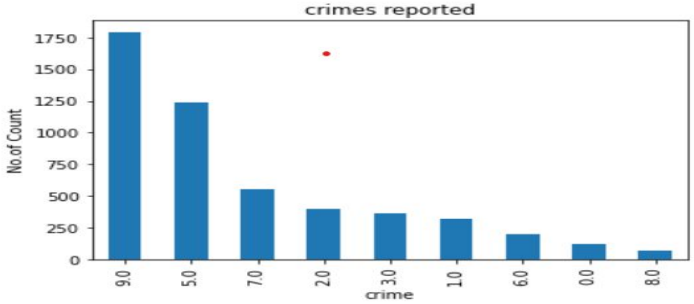

```
plt.xlabel('Year')
plt.ylabel('No.of Count')
plt.title('Highest cases reported in Year')
cdf2.Year.value_counts().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ffb64c9d390>
```




```
plt.xlabel('crime')
plt.ylabel('No.of Count')
plt.title('crimes reported')
cdf2.Type.value_counts().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ffb64a14390>
```



References

https://dphi.tech/courses/introduction-to-exploratory-data-analysis

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

https://www.kaggle.com/rmenon1998/bangalore-neighborhoods

https://www.kaggle.com/vineetsingh26/bangalore-south-crime-details

THANK YOU!