DATA ANALYSIS OF CRIME IN BOSTON

21CSS101J – PROGRAMMING FOR PROBLEM SOLVING

Mini Project Report

Submitted by

C DHIYA [Reg. No.: RA2211026010221]
B.Tech. CSE - AI & ML

DAKSHYA T [Reg. No.: RA2211026010250]
B.Tech. CSE - AI & ML

SCHOOL OF COMPUTING
COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Under Section 3 of UGC Act, 1956)
S.R.M. NAGAR, KATTANKULATHUR – 603 203
KANCHEEPURAM DISTRICT

December 2022

TABLE OF CONTENTS

PROBLEM STATEMENT

Data Analysis of any dataset using NumPy, Pandas and Matplotlib.

METHODOLOGY/PROCEDURE

We decided to analyse the datasheet of crimes in Boston for the mini project.

We can understand which are the most violent neighbourhoods in Boston, the most common crimes in the area and also what time we can walk safely in the beautiful tourist spots of the city.
Here we can extract some relevant information, such as the most violent neighbourhoods, the most common crimes, the year with the highest number of crimes, the most frequent time, among other things.

CODING IN PYTHON

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import folium
from folium.plugins import HeatMap
%matplotlib inline


data = pd.read_csv('/content/crime.csv', encoding='latin-1')
data.head()


data.isnull().sum()



max_street_crime = data['STREET'].value_counts().index[0]
max_year_crime = data['YEAR'].value_counts().index[0]
max_hour_crime = data['HOUR'].value_counts().index[0]
max_month_crime = data['MONTH'].value_counts().index[0]
max_day_crime = data['DAY_OF_WEEK'].value_counts().index[0]

month = ['January','February','March','April','May','June','July',
         'August','September','October','November','December']

print('Street with higher occurrence of crimes:', max_street_crime)
print('Year with highest crime occurrence:', max_year_crime)
print('Hour with highest crime occurrence:', max_hour_crime)
print('Month with highest crime occurrence:', month[max_month_crime-1],
max_month_crime)
print('Day with highest crime occurrence:', max_day_crime)




plt.subplots(figsize=(15,6))
sns.countplot('YEAR',data=data,palette='RdYlGn_r',edgecolor=sns.color_p
alette('dark',7))
plt.xticks(rotation=90)
plt.title('Number Of Crimes Each Year')
plt.show()
plt.subplots(figsize=(15,6))
sns.countplot('HOUR',data=data,palette='RdYlGn_r',edgecolor=sns.color_p
alette('dark',7))
plt.xticks(rotation=90)
plt.title('Number Of Crimes Each Hour')
plt.show()



plt.subplots(figsize=(15,6))
```

```
sns.countplot('OFFENSE_CODE_GROUP',data=data,palette='RdYlGn_r',edgecol
or=sns.color_palette('dark',7))
plt.xticks(rotation=90)
plt.title('Types of serious crimes')
plt.show()
```

```
plt.subplots(figsize=(15,6))
sns.countplot('DISTRICT',data=data,palette='RdYlGn',edgecolor=sns.color
_palette('Paired',20),order=data['DISTRICT'].value_counts().index)
plt.xticks(rotation=90)
plt.title('Number Of Crimes Activities By District')
plt.show()
```

```
pd.crosstab(data.DISTRICT,data.OFFENSE_CODE_GROUP).plot.barh(stacked=Tr
ue,width=1,color=sns.color_palette('RdYlGn',5))
fig=plt.gcf()
fig.set_size_inches(12,8)
plt.show()
```

```
pd.crosstab(data.YEAR,data.OFFENSE_CODE_GROUP).plot.barh(stacked=True,w
idth=1,color=sns.color_palette('RdYlGn',5))
fig=plt.gcf()
fig.set_size_inches(12,8)
plt.show()
```

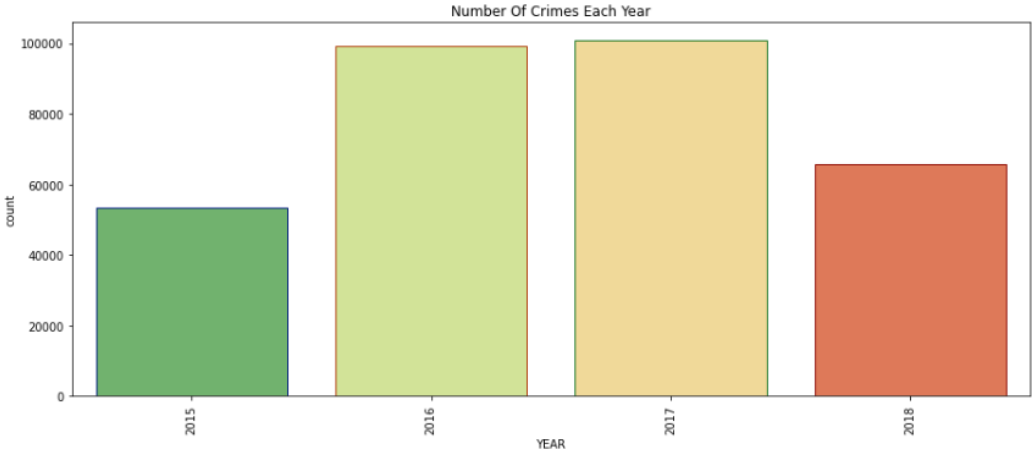
```
G1=data[data['MONTH'].isin(data['MONTH'].value_counts()[1:11].index)]
pd.crosstab(G1.YEAR,G1.DISTRICT).plot(color=sns.color_palette('dark',6)
)
fig=plt.gcf()
fig.set_size_inches(18,6)
plt.show()
```

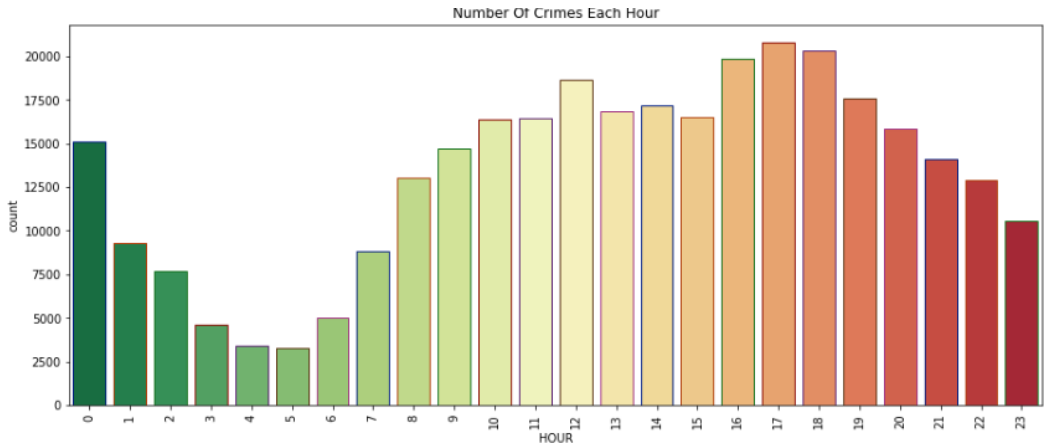
RESULTS

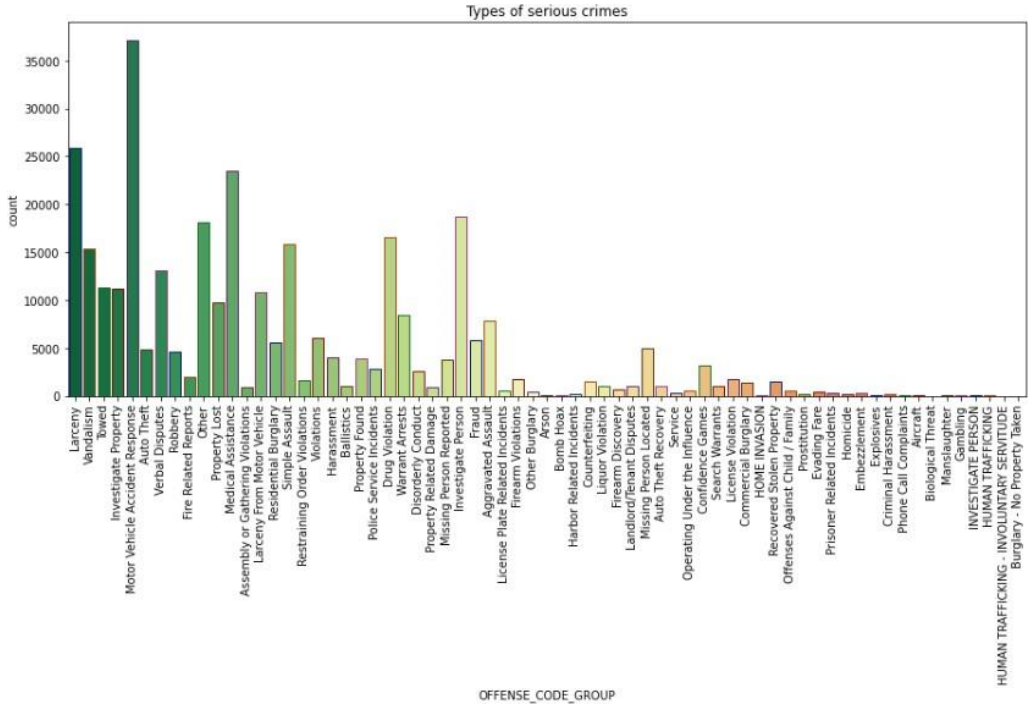
	INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location
0	I182070945	619	Larceny	LARCENY ALL OTHERS	D14	808	NaN	2018-09-02 13:00:00	2018	9	Sunday	13	Part One	LINCOLN ST	42.357791	-71.139371	(42.35779134, -71.13937053)
1	I182070943	1402	Vandalism	VANDALISM	C11	347	NaN	2018-08-21 00:00:00	2018	8	Tuesday	0	Part Two	HECLA ST	42.306821	-71.060300	(42.30682138, -71.06030035)
2	I182070941	3410	Towed	TOWED MOTOR VEHICLE	D4	151	NaN	2018-09-03 19:27:00	2018	9	Monday	19	Part Three	CAZENOVE ST	42.346589	-71.072429	(42.34658879, -71.07242943)
3	I182070940	3114	Investigate Property	INVESTIGATE PROPERTY	D4	272	NaN	2018-09-03 21:16:00	2018	9	Monday	21	Part Three	NEWCOMB ST	42.334182	-71.078664	(42.33418175, -71.07866441)
4	I182070938	3114	Investigate Property	INVESTIGATE PROPERTY	B3	421	NaN	2018-09-03 21:05:00	2018	9	Monday	21	Part Three	DELHI ST	42.275365	-71.090361	(42.27536542, -71.09036101)

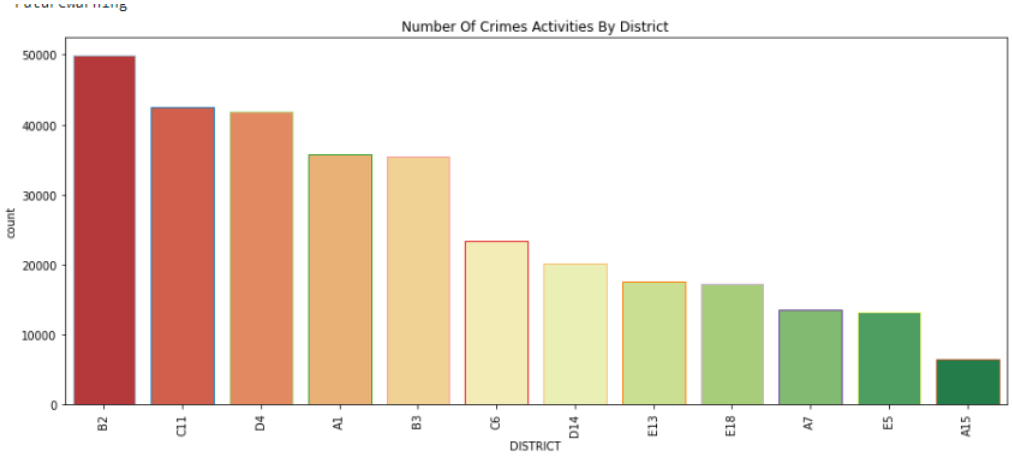
```
INCIDENT_NUMBER            0
OFFENSE_CODE               0
OFFENSE_CODE_GROUP         0
OFFENSE_DESCRIPTION        0
DISTRICT                1765
REPORTING_AREA             0
SHOOTING              318054
OCCURRED_ON_DATE           0
YEAR                       0
MONTH                      0
DAY_OF_WEEK                0
HOUR                       0
UCR_PART                  90
STREET                 10871
Lat                    19999
Long                   19999
Location                   0
dtype: int64
```

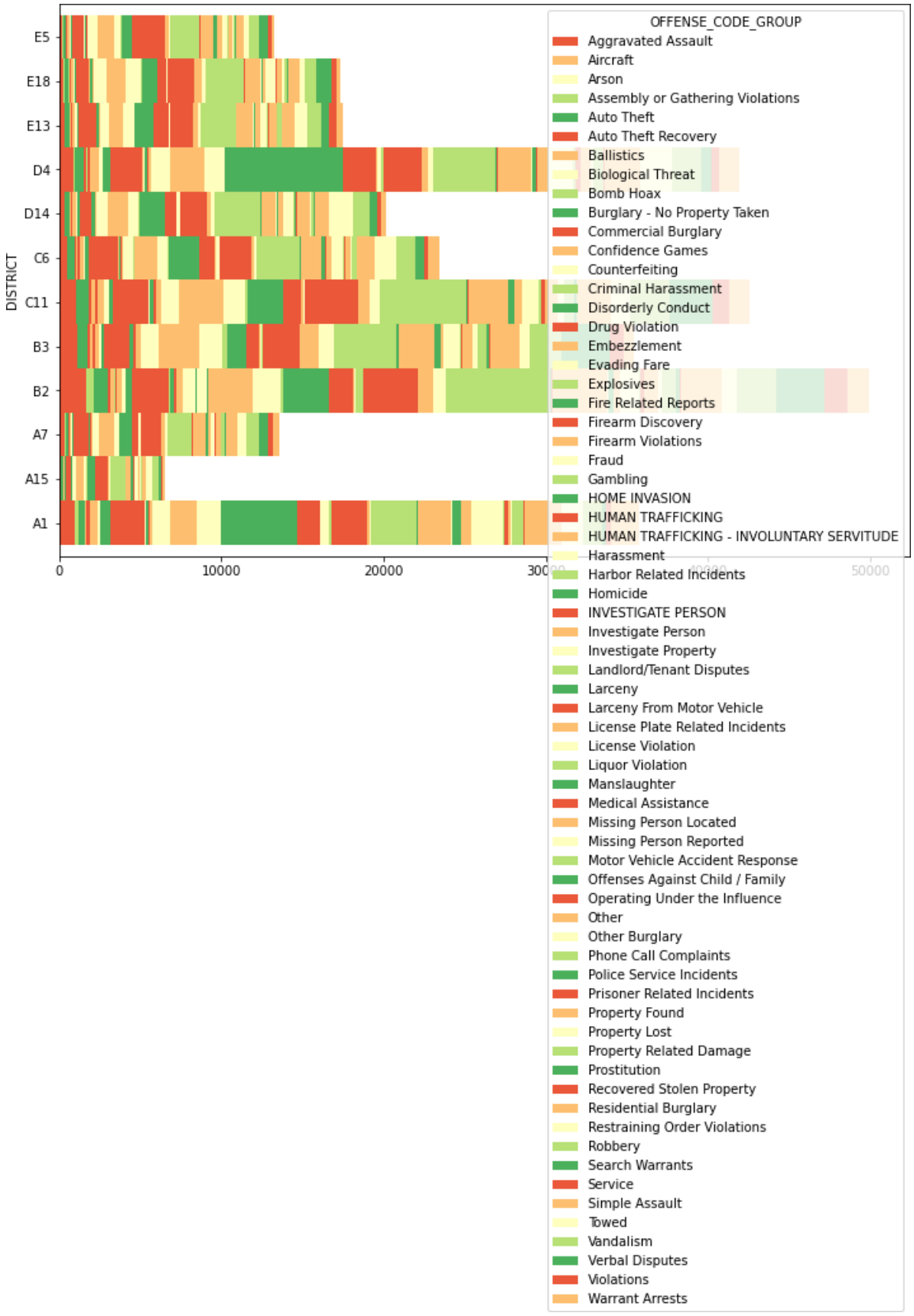
```
Street with higher occurrence of crimes: WASHINGTON ST
Year with highest crime occurrence: 2017
Hour with highest crime occurrence: 17
Month with highest crime occurrence: August 8
Day with highest crime occurrence: Friday
```



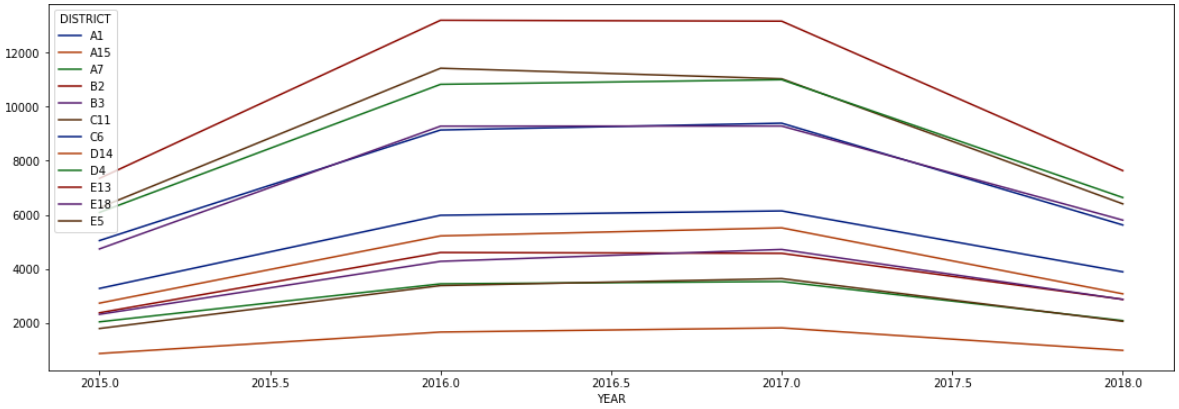

Number Of Crimes Each Hour
count
HOUR


Types of serious crimes
count
OFFENSE_CODE_GROUP


Number Of Crimes Activities By District
count
50000
40000
30000
20000
10000
0
B2
C11
D4
A1
B3
C6
D14
E13
E18
A7
E5
A15
DISTRICT


DISTRICT
E5
E18
E13
D4
D14
C6
C11
B3
B2
A7
A15
A1
0
10000
20000
30000
50000
OFFENSE_CODE_GROUP
Aggravated Assault
Aircraft
Arson
Assembly or Gathering Violations
Auto Theft
Auto Theft Recovery
Ballistics
Biological Threat
Bomb Hoax
Burglary - No Property Taken
Commercial Burglary
Confidence Games
Counterfeiting
Criminal Harassment
Disorderly Conduct
Drug Violation
Embezzlement
Evading Fare
Explosives
Fire Related Reports
Firearm Discovery
Firearm Violations
Fraud
Gambling
HOME INVASION
HUMAN TRAFFICKING
HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
Harassment
Harbor Related Incidents
Homicide
INVESTIGATE PERSON
Investigate Person
Investigate Property
Landlord/Tenant Disputes
Larceny
Larceny From Motor Vehicle
License Plate Related Incidents
License Violation
Liquor Violation
Manslaughter
Medical Assistance
Missing Person Located
Missing Person Reported
Motor Vehicle Accident Response
Offenses Against Child / Family
Operating Under the Influence
Other
Other Burglary
Phone Call Complaints
Police Service Incidents
Prisoner Related Incidents
Property Found
Property Lost
Property Related Damage
Prostitution
Recovered Stolen Property
Residential Burglary
Restraining Order Violations
Robbery
Search Warrants
Service
Simple Assault
Towed
Vandalism
Verbal Disputes
Violations
Warrant Arrests


YEAR
2018
2017
2016
2015
0
20000
40000
60000
100000
OFFENSE_CODE_GROUP
Aggravated Assault
Aircraft
Arson
Assembly or Gathering Violations
Auto Theft
Auto Theft Recovery
Ballistics
Biological Threat
Bomb Hoax
Burglary - No Property Taken
Commercial Burglary
Confidence Games
Counterfeiting
Criminal Harassment
Disorderly Conduct
Drug Violation
Embezzlement
Evading Fare
Explosives
Fire Related Reports
Firearm Discovery
Firearm Violations
Fraud
Gambling
HOME INVASION
HUMAN TRAFFICKING
HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
Harassment
Harbor Related Incidents
Homicide
INVESTIGATE PERSON
Investigate Person
Investigate Property
Landlord/Tenant Disputes
Larceny
Larceny From Motor Vehicle
License Plate Related Incidents
License Violation
Liquor Violation
Manslaughter
Medical Assistance
Missing Person Located
Missing Person Reported
Motor Vehicle Accident Response
Offenses Against Child / Family
Operating Under the Influence
Other
Other Burglary
Phone Call Complaints
Police Service Incidents
Prisoner Related Incidents
Property Found
Property Lost
Property Related Damage
Prostitution
Recovered Stolen Property
Residential Burglary
Restraining Order Violations
Robbery
Search Warrants
Service
Simple Assault
Towed
Vandalism
Verbal Disputes
Violations
Warrant Arrests


DISTRICT
A1
A15
A7
B2
B3
C11
C6
D14
D4
E13
E18
E5
12000
10000
8000
6000
4000
2000
2015.0
2015.5
2016.0
2016.5
2017.0
2017.5
2018.0
YEAR

CONCLUSION

EDA is the most important part of any analysis. We will get to know many things about our data. We have tried to show most of the python functions used for exploring the data with visualizations.

From the graph we can see that 2016 and 2017 had the highest incidence of crime in the region, the occurrence of crimes begins to rise from 7am onwards peaking between 5 and 6pm and that the neighbourhood with the highest incidents of crime is B12.

We were able to successfully analyse the data.