Project Report

on

CRIME RATE PREDICTION USING MACHINE LEARNING

Submitted in partial fulfillment of the
requirement for the award of the degree of

Bachelor of technology

Computer science and engineering

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Mr. Mukesh Kumar Jha

Submitted By

Ayush Verma (19SCSE1180014)
Devansh Saini (19SCSE1180102)

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
DECEMBER, 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **"CRIME RATE PREDICTION USING MACHINE LEARNING"** in partial fulfillment of the requirements for the award of the B.Tech submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the Period of october 2021 to December 2021 , under the supervision of Mr. Mukesh Kr. Jha(Assistant Professor), Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me/us for the award of any other degree of this or any other places.

Ayush Verma
Devansh Saini

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Mr. Mukesh Jha
Assistant
Professor

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of ~~Name: Admission No~~ has been held on _____ and his/her work is recommended for the award of ~~Name of Degree.~~

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: November, 2013

PLACE: GREATER NOIDA
Fall 2021 - 2022

TABLE OF CONTENTS

ABSTRACT

Crime investigation and prediction is a precise methodology for breaking down and recognizing various examples, connections and patterns in crime. The locales with high likelihood of event of crime is predicted by the framework. The framework created will assist with accelerating the most common way of settling crime for the law authorization offices. The utilization of AI and machine learning to identify crime by means of sound or cameras presently exists, is demonstrated to work, and expected to keep on growing. The utilization of AI/ML in anticipating crime or a singular's probability for committing a crime has guarantee yet is even a greater amount of an unknown. The current information from the police is utilized which subsequent to utilizing diverse prediction and clustering algorithms gives an understanding that will assist with foreseeing the probability of incidents, track crimes and help the law authorization specialists to convey assets and furthermore settle crime cases at a quicker rate.

Upgrades in crime prevention innovation will probably prod expanded all out spending on this innovation. We likewise attempt to make our characterization task more significant by consolidating different classes into bigger classes. At last, we report and think about our outcomes with various classifiers, and well on roads for future work.

Introduction

Crimes are the significant threat to the humankind. There are many crimes that happen in regular intervals of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster.

Through many documentation and cases, it came out that machine learning and data science can make the work easier and faster. The inputs to our algorithms are time (hour, day, month, and year), place (latitude and longitude), and class of crime:

- Act 379 - Robbery
- Act 13 - Gambling
- Act 279 - Accident
- Act 323 - Violence
- Act 302 - Murder
- Act 363 - Kidnapping

Literature Survey

The principle issue is that everyday the population will be expanded and by that the wrongdoings are likewise going to be Increased in various regions by this the crime percentage can't be precisely anticipated by the authorities. The authorities as they center around many issues may not foresee the wrongdoings to be occurred later on. The authorities/cops in spite of the fact that they attempt to diminish the crime percentage they may not lessen in undeniable way. The crime percentage forecast in future might be hard for them.

There has been incalculable of work done identified with wrongdoings. Enormous datasets have been looked into, and data, for example, area and the kind of violations have been removed to assist individuals with observing law requirements. Existing strategies have utilized these information bases to recognize wrongdoing areas of interest dependent on areas.

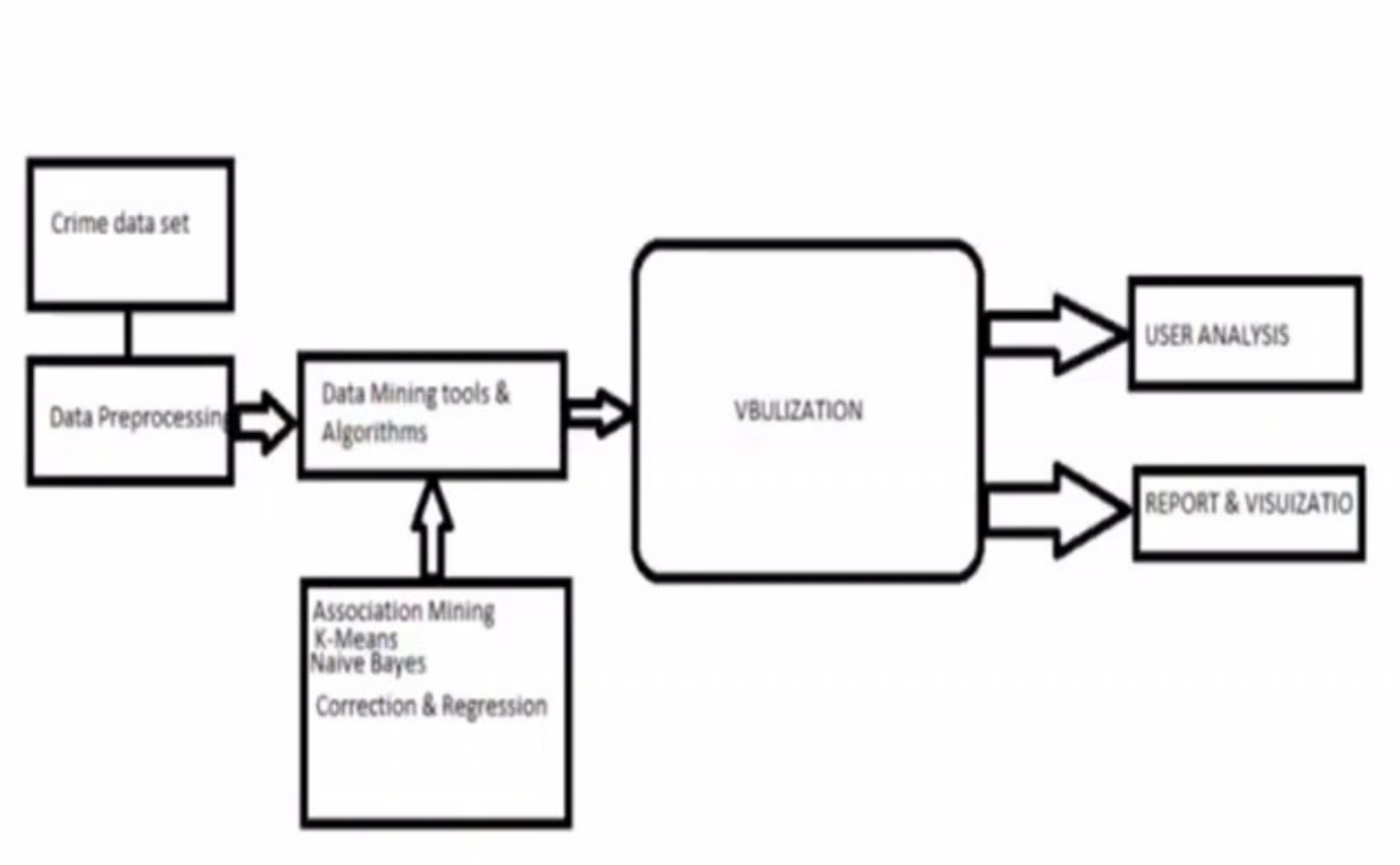
Despite the fact that wrongdoing areas have been distinguished, there is no data accessible that incorporates the wrongdoing event date and time alongside strategies that can precisely foresee what violations will happen later on.

Our review expects to discover spatial and transient criminal areas of interest utilizing a bunch of genuine world datasets of wrongdoings. We will attempt to find the most probable wrongdoing areas and their incessant event time. Moreover, we will foresee what sort of wrongdoing may happen next in a particular area inside a specific time. At long last,

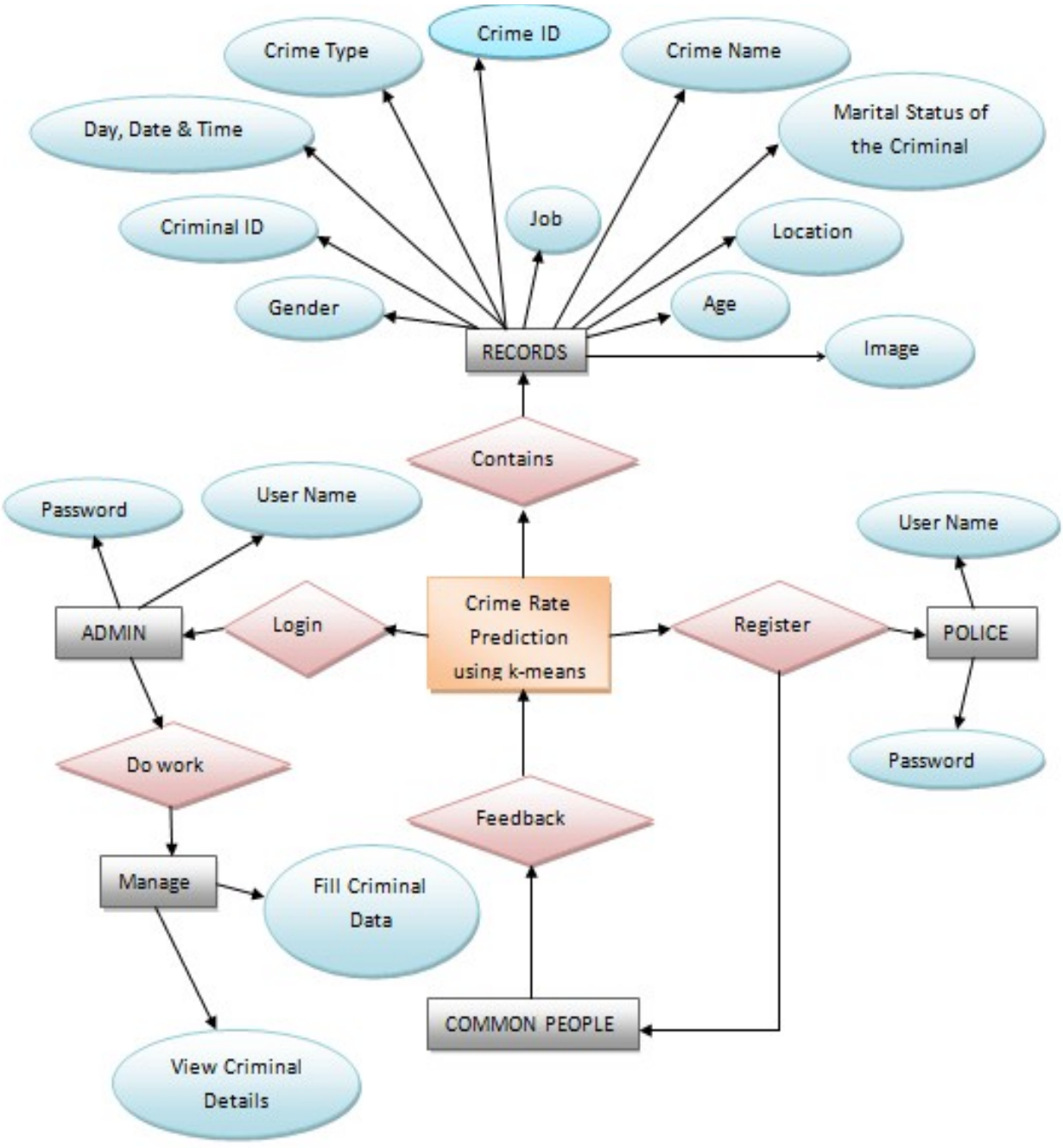
we plan to give an examination study by joining our discoveries of a specific violations datasets with its demographics data.

Project Design

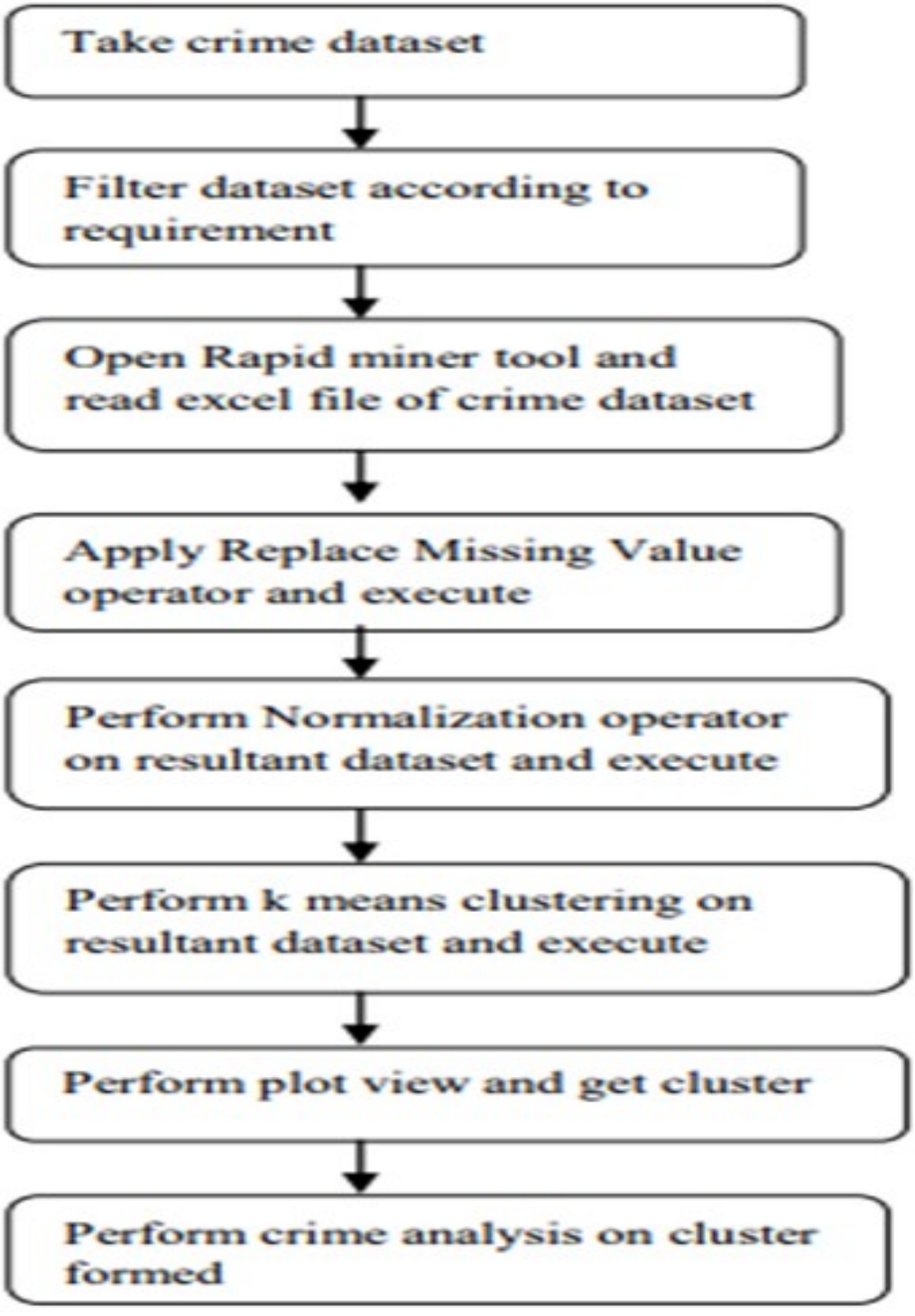
DATA FLOW DIAGRAM:-



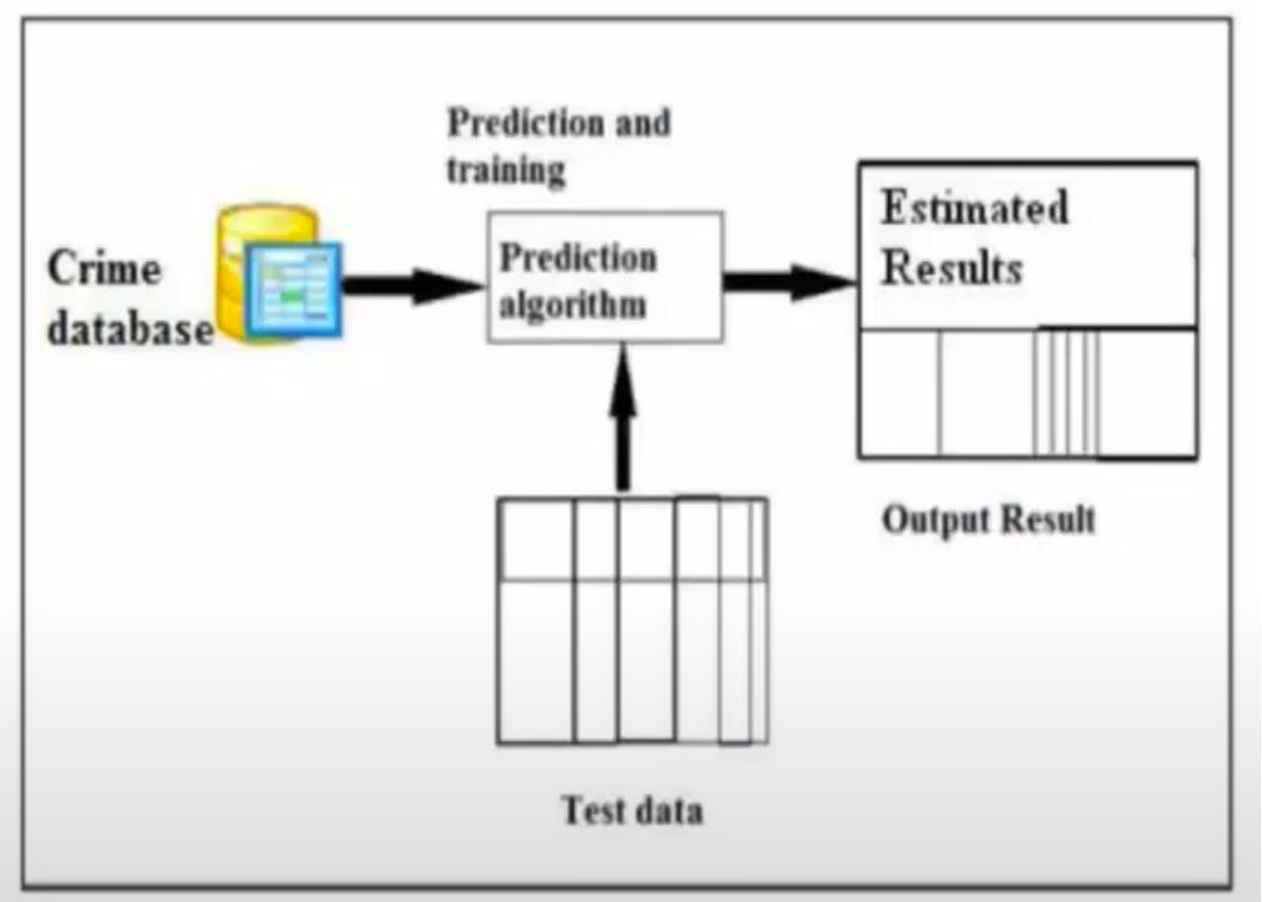
ER DIAGRAM:-


Crime ID
Crime Type
Crime Name
Day, Date & Time
Marital Status of the Criminal
Criminal ID
Job
Location
Gender
Age
RECORDS
Image
Contains
Password
User Name
User Name
ADMIN
Login
Crime Rate Prediction using k-means
Register
POLICE
Do work
Password
Feedback
Manage
Fill Criminal Data
COMMON PEOPLE
View Criminal Details

FLOWCHART:-


Take crime dataset
Filter dataset according to requirement
Open Rapid miner tool and read excel file of crime dataset
Apply Replace Missing Value operator and execute
Perform Normalization operator on resultant dataset and execute
Perform k means clustering on resultant dataset and execute
Perform plot view and get cluster
Perform crime analysis on cluster formed

ARCHITECTURE DIAGRAM:-



Dataset

‘/kaggle/input/up-crime-data-year-2019/MR data - Compiled Data Set.csv'

Requirements of Project

- Linux operating System
- Python 2.7
- Flask Framework
- Flask wtForms
- Flask Mysqldb
- Numpy
- Flask Mail
- SciPy
- Scikit-learn

Implementation

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image:
https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all
files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets
preserved as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved
outside of the current session
```

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

```
/kaggle/input/up-crime-data-year-2019/MR data - Compiled Data Set.csv
```

+ Code + Markdown

[15]:

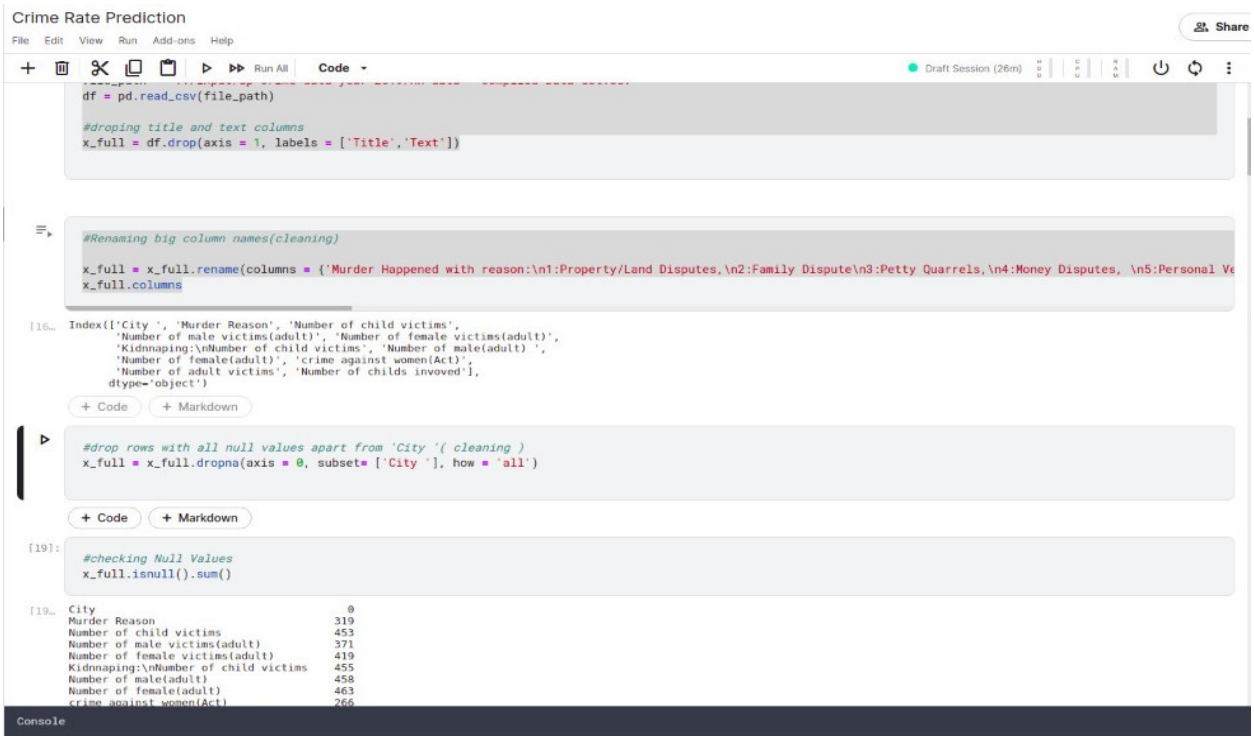
```
file_path = '../input/up-crime-data-year-2019/MR data - Compiled Data Set.csv'
df = pd.read_csv(file_path)

#droping title and text columns
x_full = df.drop(axis = 1, labels = ['Title','Text'])
```

[16]:

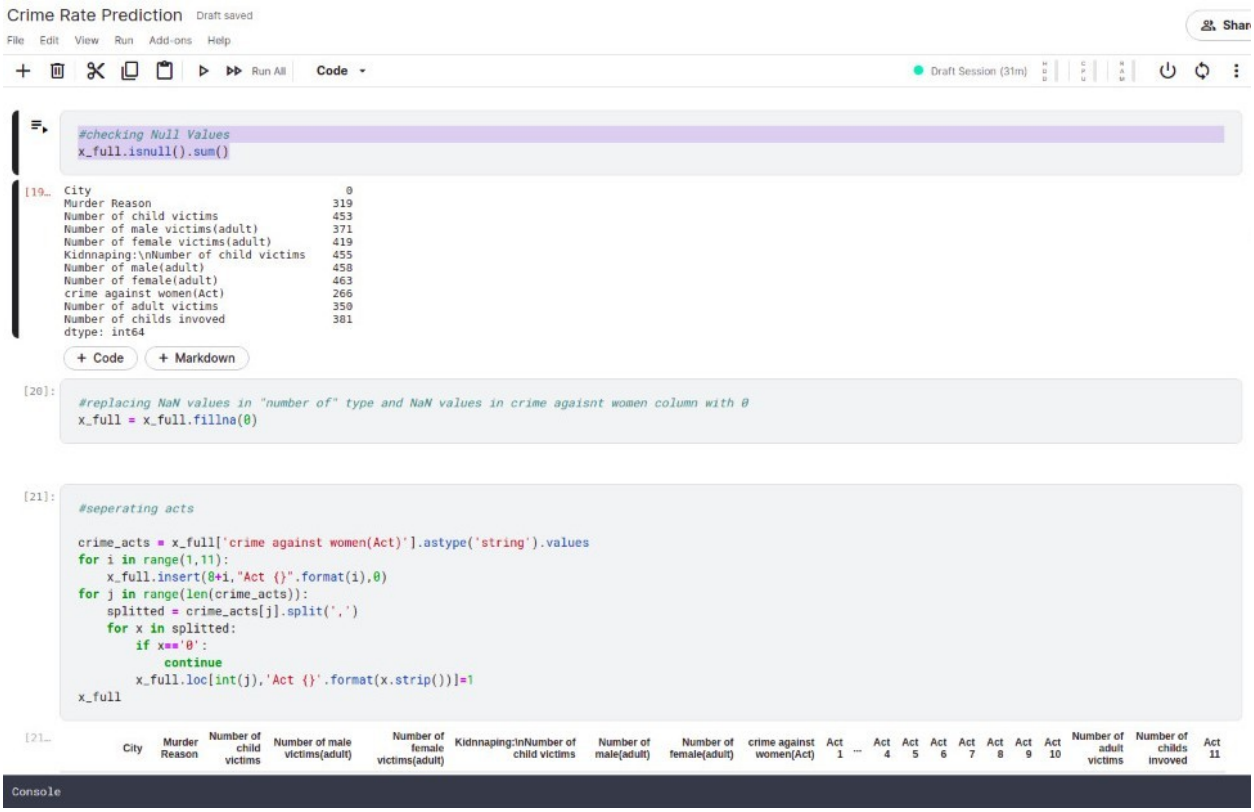
```
#Renaming big column names(cleaning)

x_full = x_full.rename(columns = {'Murder Happened with reason:\n1:Property/Land Disputes,\n2:Family Dispute\n3:Petty Quarrels,\n4:Money Disputes, \n5:Personal Ve
x_full.columns
```

Console

```
file_path = '../input/up-crime-data-year-2019/MR data - Compiled Data Set.csv'
df = pd.read_csv(file_path)

#droping title and text columns
x_full = df.drop(axis = 1, labels = ['Title','Text'])

#Renaming big column names(cleaning)

x_full = x_full.rename(columns = {'Murder Happened with reason:\n1:Property/Land Disputes,\n2:Family Dispute\n3:Petty Quarrels,\n4:Money Disputes, \n5:Personal Vendetta, \n6:Love Affairs, \n7:Casteism\n8:Unknown/other':'Murder Reason', 'Crime Against Women ( Combined ):\n1. Murder with Rape \n2. Dowry Deaths(Sec. 3048)\n3. Suicide(sec 305/306)\n4. Kidnapping(All)\n5. Acid Attack(Sec. 326A IPC)\n6. Cruelty by Husband/in-laws((Sec.498 A IPC)\n7. Rape only(Sec. 376 or 511 IPC)\n8. Assault on Women with Intent to Outrage her Modesty (Sec. 354 IPC)\n9. Cyber Crimes against Women\n10. Protection of Children from Sexual Offences Act':'crime against women(Act)'})
x_full.columns
#drop rows with all null values apart from 'City '( cleaning )
x_full = x_full.dropna(axis = 0, subset= ['City '], how = 'all')
```



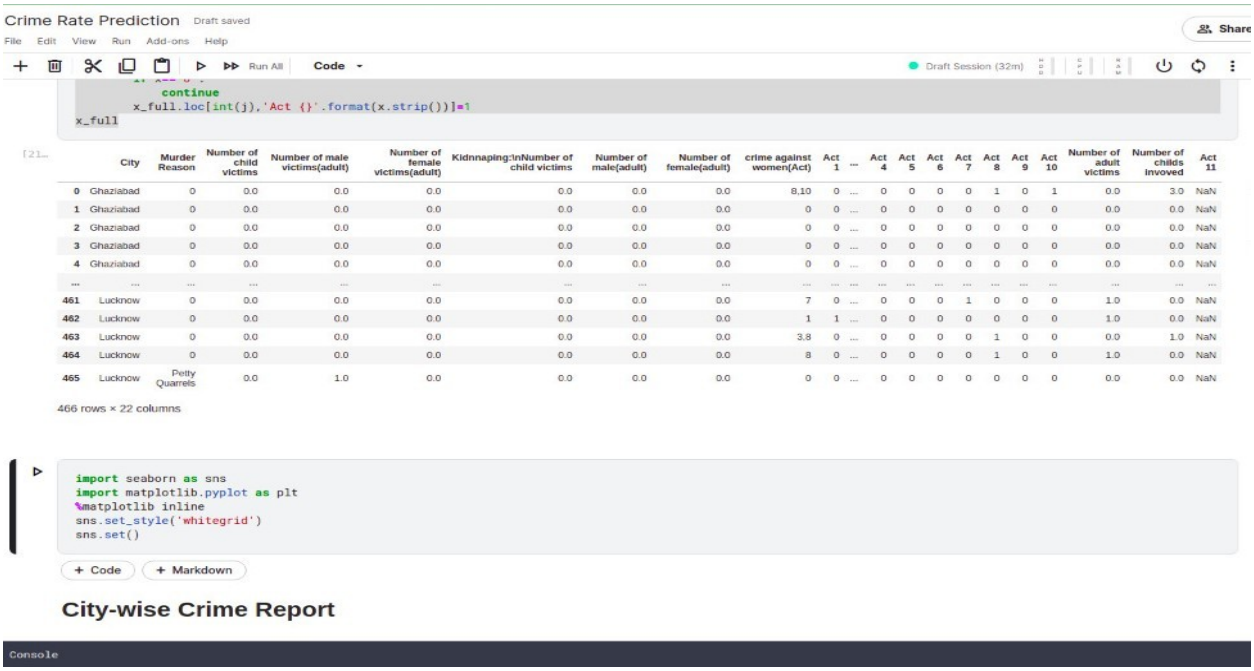
#checking Null Values
x_full.isnull().sum()



#replacing NaN values in "number of" type and NaN values in crime agaisnt women column with 0
x_full = x_full.fillna(0)

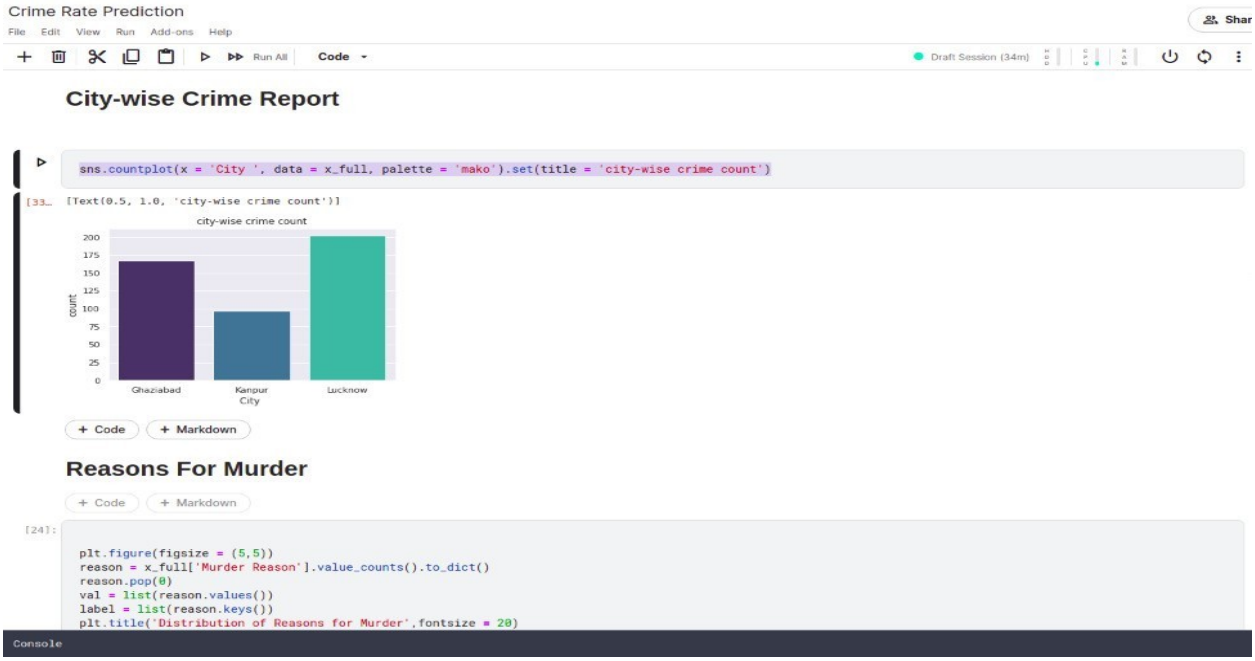
#seperating acts

crime_acts = x_full['crime against women(Act)'].astype('string').values
for i in range(1,11):
    x_full.insert(8+i,"Act {}".format(i),0)
for j in range(len(crime_acts)):
    splitted = crime_acts[j].split(',')
    for x in splitted:
        if x=='0':
            continue
        x_full.loc[int(j),'Act {}'.format(x.strip())]=1
x_full

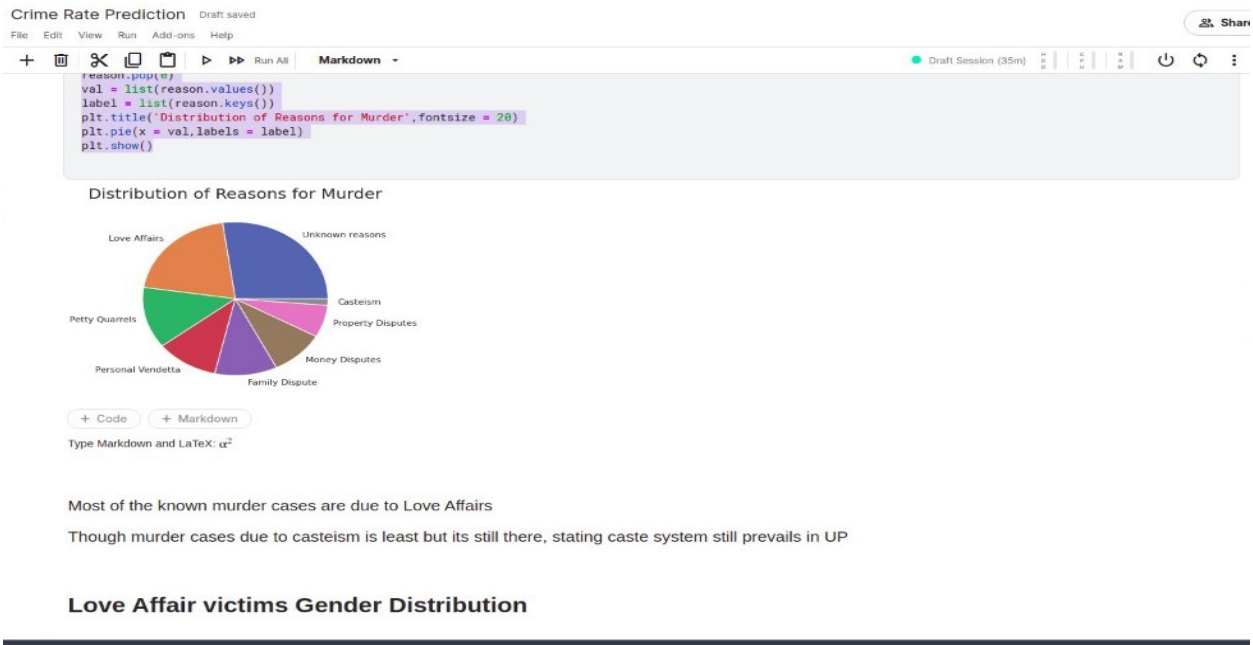


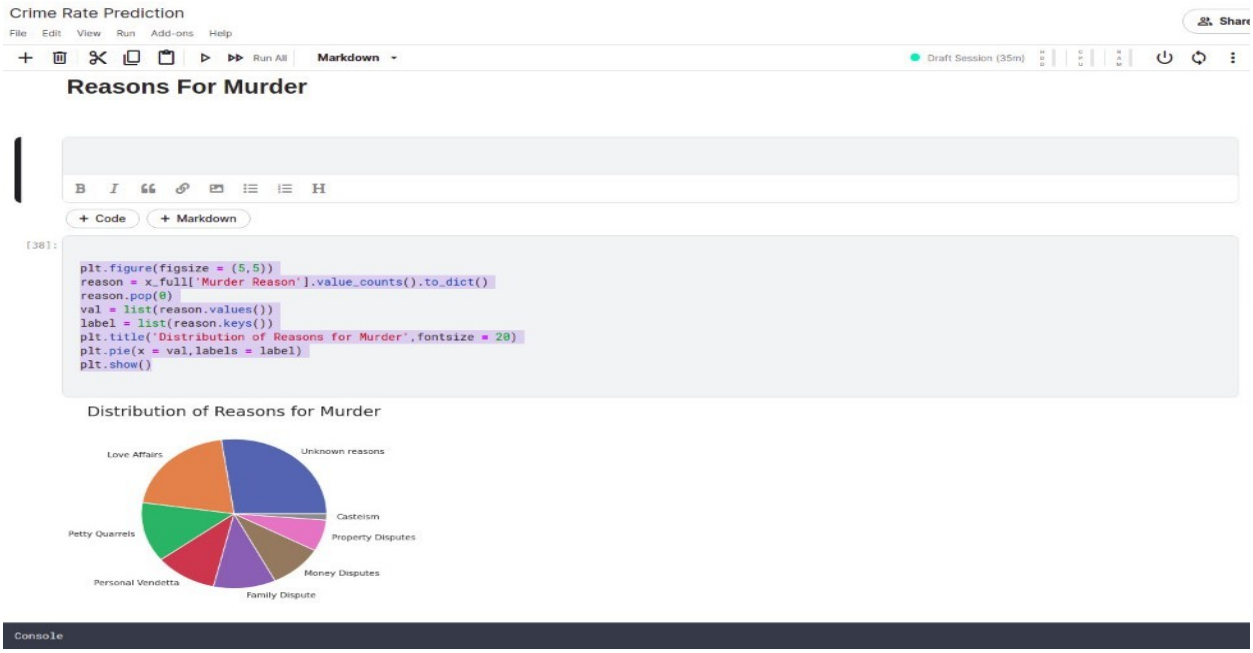
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set_style('whitegrid')
sns.set()

sns.countplot(x = 'City ', data = x_full, palette = 'mako').set(title = 'city-wise crime count')

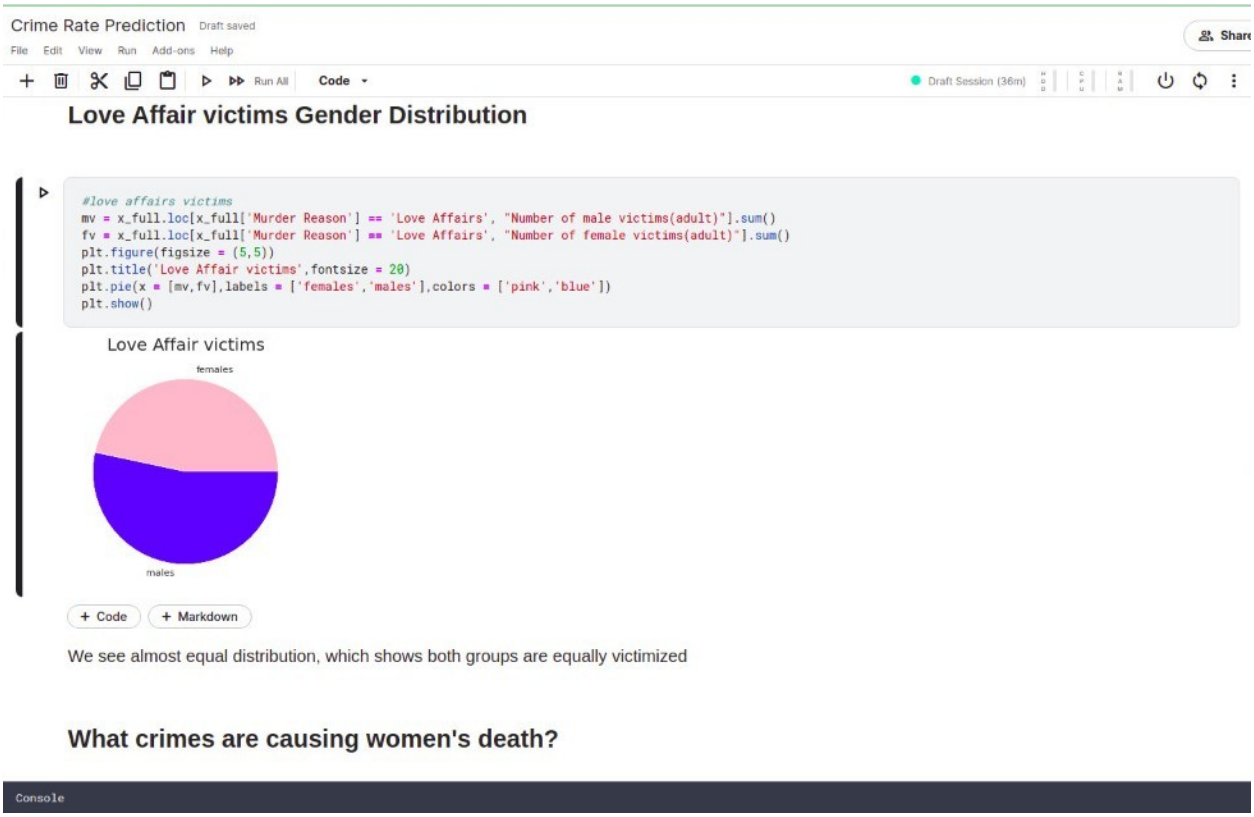


plt.figure(figsize = (5,5))
reason = x_full['Murder Reason'].value_counts().to_dict()
reason.pop(0)
val = list(reason.values())
label = list(reason.keys())
plt.title('Distribution of Reasons for Murder',fontsize = 20)
plt.pie(x = val,labels = label)
plt.show()

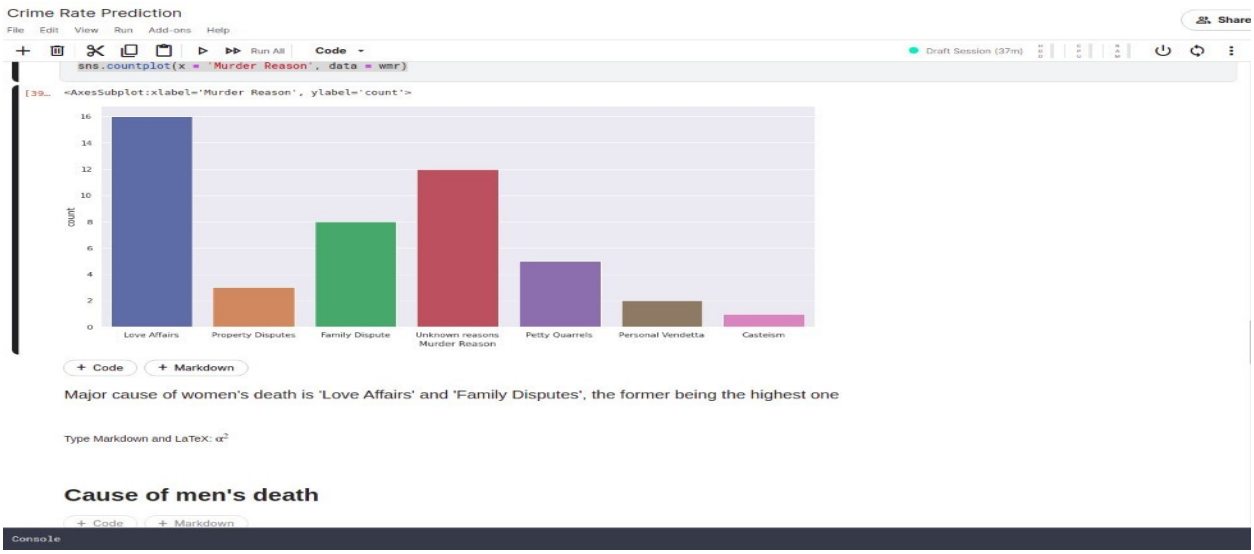




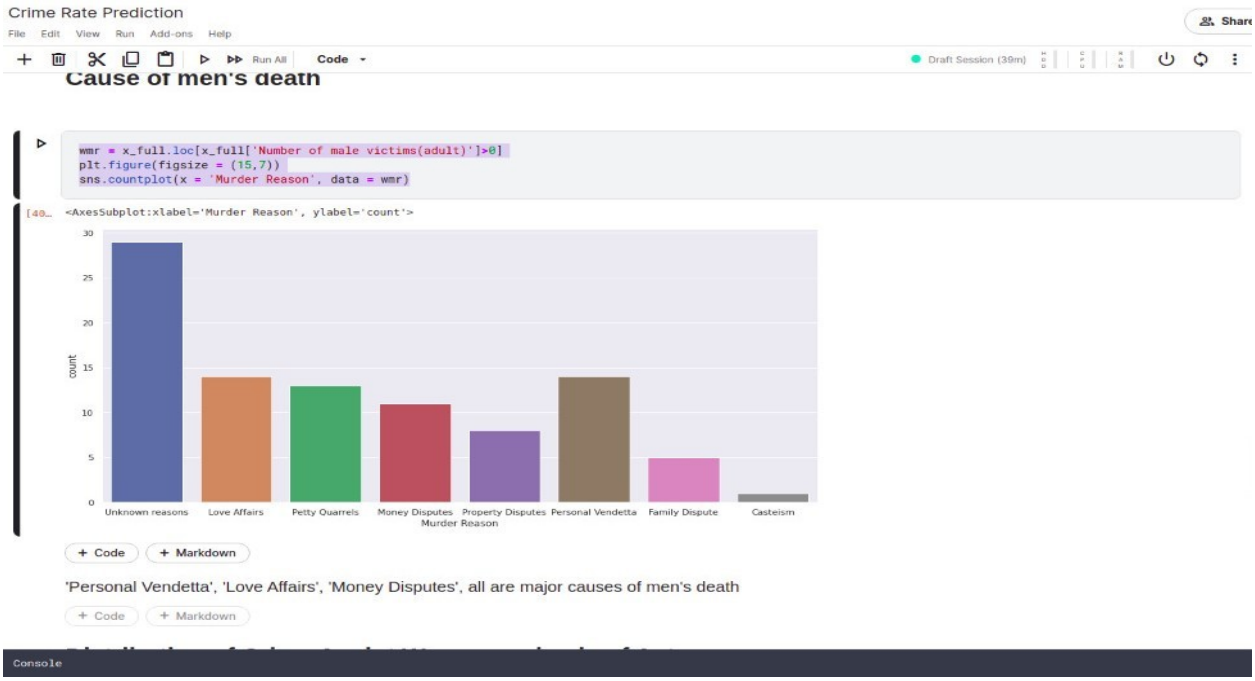

```
#love affairs victims
mv = x_full.loc[x_full['Murder Reason'] == 'Love Affairs', "Number of male victims(adult)"].sum()
fv = x_full.loc[x_full['Murder Reason'] == 'Love Affairs', "Number of female victims(adult)"].sum()
plt.figure(figsize = (5,5))
plt.title('Love Affair victims',fontsize = 20)
plt.pie(x = [mv,fv],labels = ['females','males'],colors = ['pink','blue'])
plt.show()
```



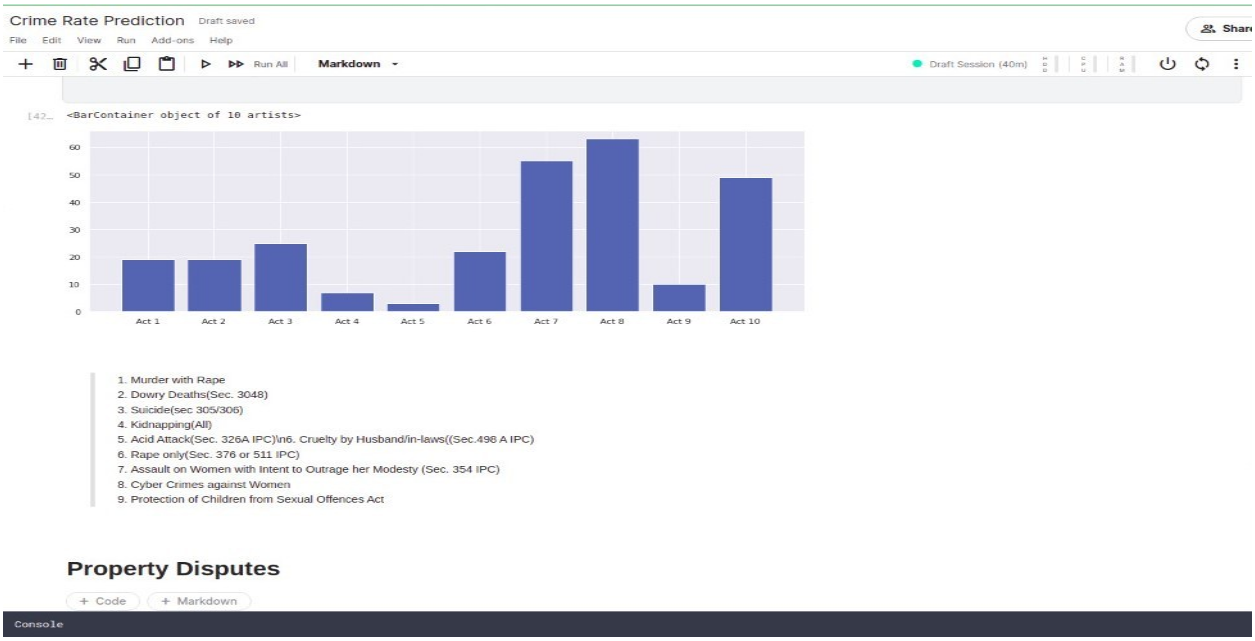

```
#what is the biggest cause of women's death
wmr = x_full.loc[x_full['Number of female victims(adult)']>0]
plt.figure(figsize = (15,7))
sns.countplot(x = 'Murder Reason', data = wmr)
```



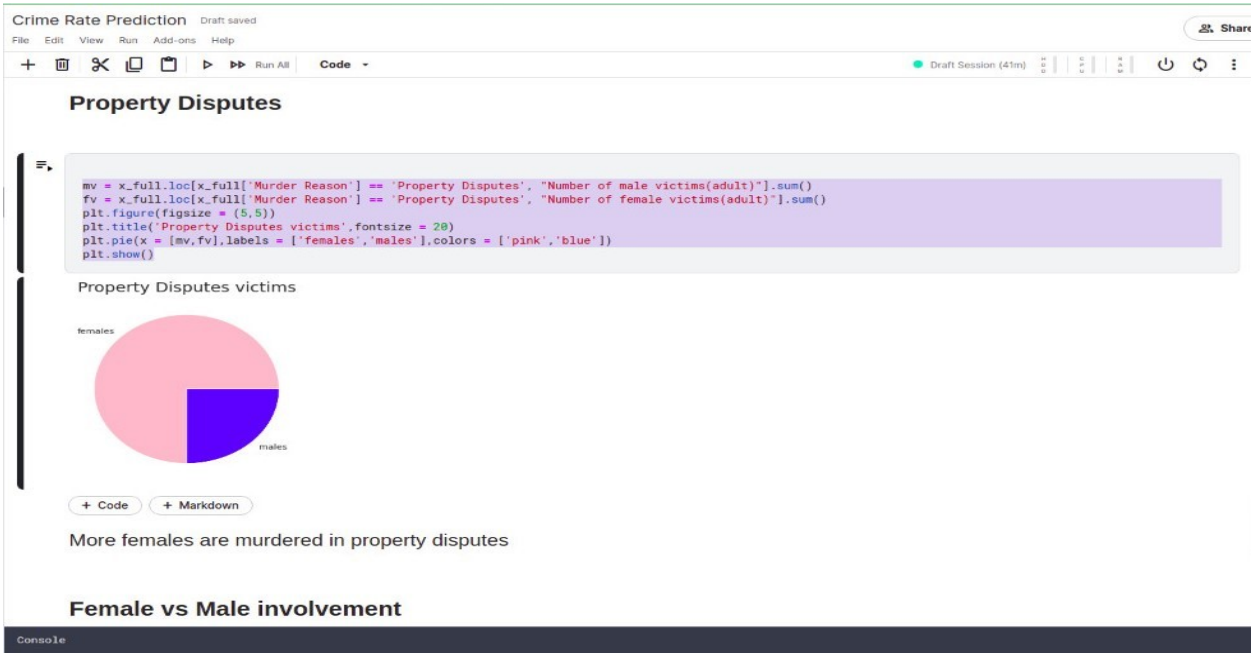

```
wmr = x_full.loc[x_full['Number of male victims(adult)']>0]
plt.figure(figsize = (15,7))
sns.countplot(x = 'Murder Reason', data = wmr)
```



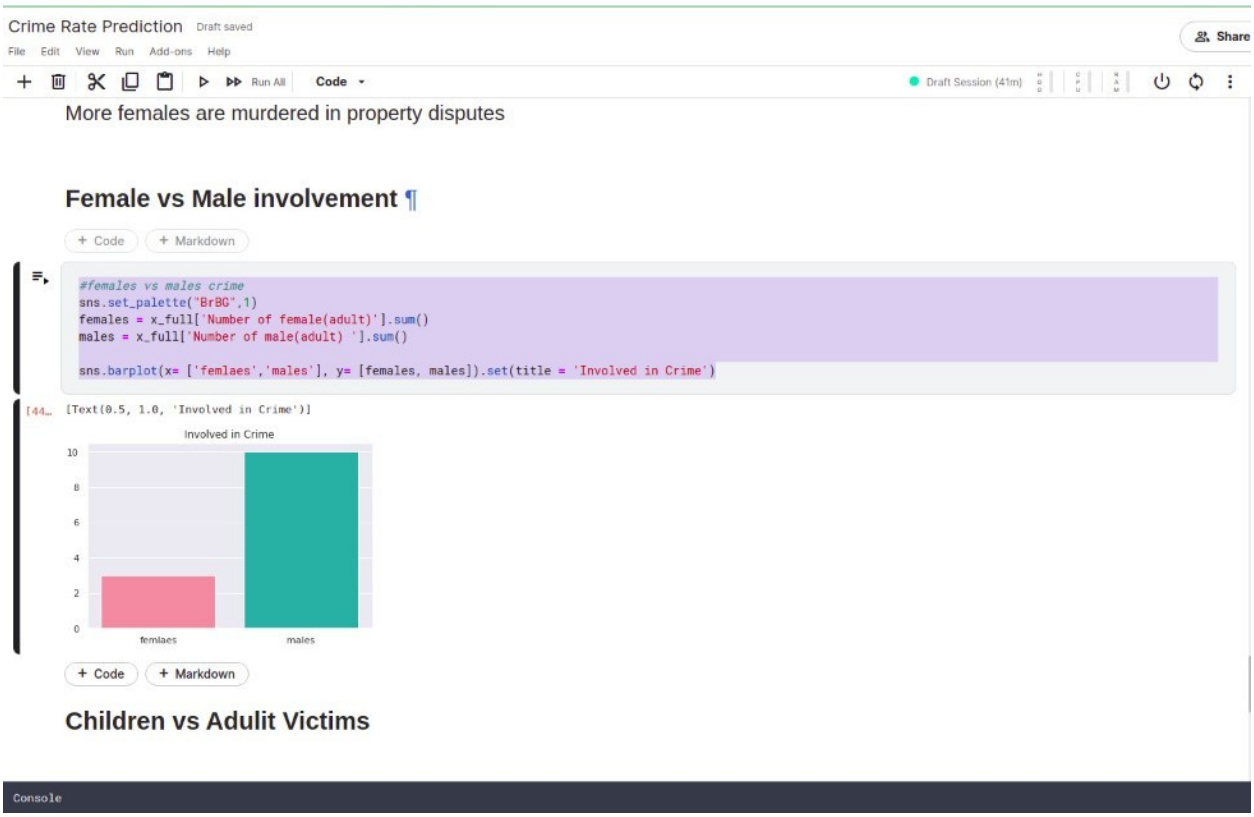

```
s = x_full.iloc[:,9:(9+10)].sum()
plt.figure(figsize = (15,5))
plt.bar(s.index,s.values)
```



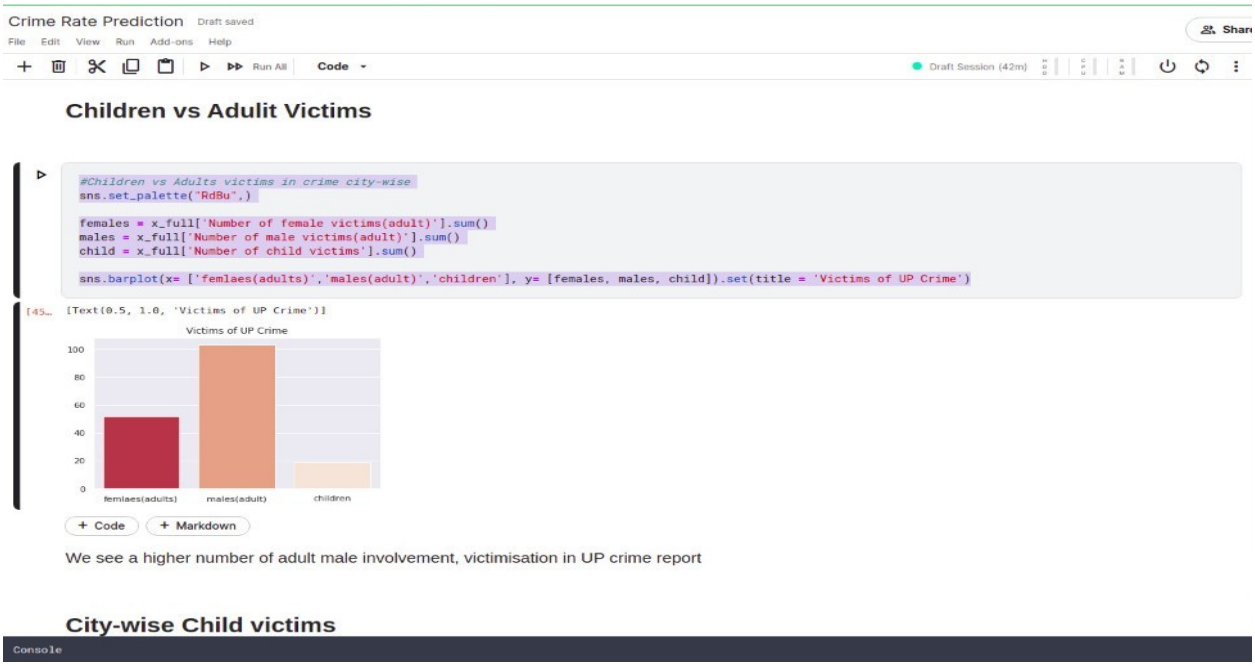

```
mv = x_full.loc[x_full['Murder Reason'] == 'Property Disputes', "Number of male victims(adult)"].sum()
fv = x_full.loc[x_full['Murder Reason'] == 'Property Disputes', "Number of female victims(adult)"].sum()
plt.figure(figsize = (5,5))
plt.title('Property Disputes victims',fontsize = 20)
plt.pie(x = [mv,fv],labels = ['females','males'],colors = ['pink','blue'])
plt.show()
```



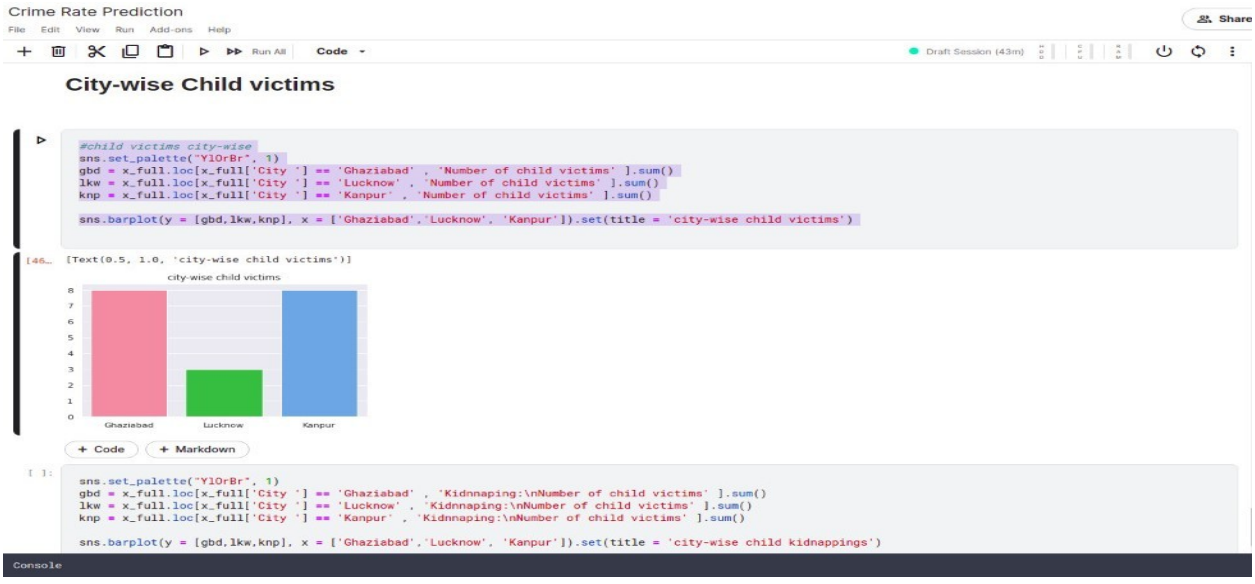
```
#females vs males crime
sns.set_palette("BrBG",1)
females = x_full['Number of female(adult)'].sum()
males = x_full['Number of male(adult) '].sum()

sns.barplot(x= ['femlaes','males'], y= [females, males]).set(title = 'Involved in Crime')
```



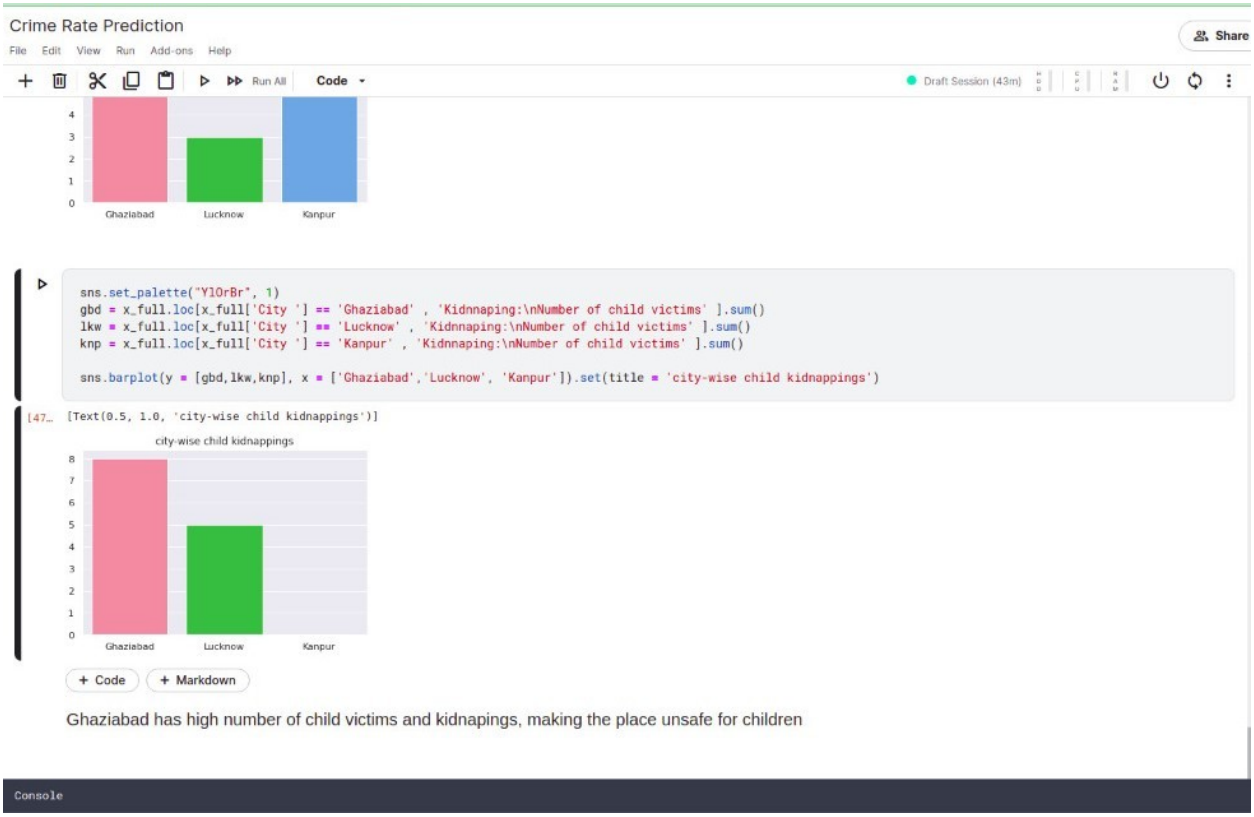

```
#Children vs Adults victims in crime city-wise
sns.set_palette("RdBu",)

females = x_full['Number of female victims(adult)'].sum()
males = x_full['Number of male victims(adult)'].sum()
child = x_full['Number of child victims'].sum()

sns.barplot(x= ['femlaes(adults)','males(adult)','children'], y= [females, males,
child]).set(title = 'Victims of UP Crime')
```




```
#child victims city-wise
sns.set_palette("YlOrBr", 1)
gbd = x_full.loc[x_full['City '] == 'Ghaziabad' , 'Number of child victims' ].sum()
lkw = x_full.loc[x_full['City '] == 'Lucknow' , 'Number of child victims' ].sum()
knp = x_full.loc[x_full['City '] == 'Kanpur' , 'Number of child victims' ].sum()

sns.barplot(y = [gbd,lkw,knp], x = ['Ghaziabad','Lucknow', 'Kanpur']).set(title =
'city-wise child victims')
```



```
sns.set_palette("YlOrBr", 1)
gbd = x_full.loc[x_full['City '] == 'Ghaziabad' , 'Kidnnaping:\nNumber of child victims' ].sum()
lkw = x_full.loc[x_full['City '] == 'Lucknow' , 'Kidnnaping:\nNumber of child victims' ].sum()
knp = x_full.loc[x_full['City '] == 'Kanpur' , 'Kidnnaping:\nNumber of child victims' ].sum()

sns.barplot(y = [gbd,lkw,knp], x = ['Ghaziabad','Lucknow', 'Kanpur']).set(title = 'city-wise child kidnappings')
```



Result

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

We also perform multiple classification tasks – we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

Conclusion

Crime Patterns can't be static since designs change over time, so unique bunching procedures like K-Means; Fuzzy C and so forth are utilized to deal with the changing crime patterns.

Much of the current work is focused in two major directions:

• Predicting surges and hotspots of crime, and

• Understanding patterns of criminal behavior that could help in solving criminal investigations.

References

[1] Lin, Y., Chen,T., Yu, L. Using Machine Learning to assist crime prevention. In: 2017 sixth IIAI-AAI

[2] Kerr, J.: Vancouver police go high tech to predict and prevent crime before it happens. Vancouver Courier, july 23, 2017.

[3] Marchant, R., Haan, S., clancey, G., Cripps, S.: Applying machine learning to criminology: semi parametric spatial demographic Bayesian regression. Security inform. (2018).

[4] M. J. H. B. T. A. M. K. T. Baig, M.Q., “Artificial intelligence, modelling and simulation (aims), 2014 2nd international conference on,” pp. 109–114, November 2014.

[5] Anitha A, Paul G and Kumari S2016 A cyber defence using Artificial Intelligence International Journal of Pharmacy and Technology 8 2532-57

[6] Zeroday. “a lua based firmware for wifi-soc esp8266”,Github. Retrieved APR 2015

[7] A. Bogomoloy, B. Lepri, J. Staiano, N. Oliver, F.Pianesi and A. Pentland, ‘once upon a crime: towards Crime Prediction from Demographics and Mobile Data’, CoRR, vol. 14092983, 2014.

[8] R. Arulanandam, B. Savarimuthu and M. Purvis, 'Extracting Crime Information from Online Newspaper Articles', in Proceedings of the Second Australasian Web Conference - Volume 155, Auckland, New Zealand, 2014, pp. 31-38.

[9] A. Buczak and C. Gifford, 'Fuzzy association rule mining for community crime pattern discovery', in ACM SIGKDD Workshop on Intelligence and Security

Informatics, Washington, D.C., 2010, pp. 1-10.

[10] M. Tayebi, F. Richard and G. Uwe, 'Understanding the Link Between Social and Spatial Distance in the Crime World', in Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12), Redondo Beach, California, 2012, pp. 550-553.

[11] S. Nath, 'Crime Pattern Detection Using Data Mining', in Web Intelligence and Intelligent Agent Technology Workshops, 2006. WI-IAT 2006 Workshops. 2006 IEEE/WIC/ACM International Conference on, 2006, pp. 41,44.

[12] Crimereports.com, 2015.

[13] S. Chainey, L. Tompson and S. Uhlig, 'The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime', Security Journal, vol. 21, no. 1-2, pp. 4-28, 2008.

[14] Data.denvergov.org, 'Denver Open Data Catalog: Crime', 2015.

[15] Imgh.us, 2015. [Accessed: 20- May- 2015].

[16] O. Knowledge, 'Crime — Datasets - US City Open Data Census', Us-city.census.okfn.org, 2015.