



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 8      Issue: VIII      Month of publication: August 2020**

**DOI: <https://doi.org/10.22214/ijraset.2020.30937>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Malfeasance Factual Anatomization

Anika Sharma<sup>1</sup>, Abhishek Singh<sup>2</sup>, Rythm Seth<sup>3</sup>

<sup>1, 2, 3</sup>Apex Institute of Technology, Chandigarh University

**Abstract:** *Data analysis is one amongst the hard tasks within the coinciding world. the dimensions of knowledge is increasing at a awfully high rate owing to the sexual practice of peripatetic gadgets and sensors connected. to create that information legible is another difficult task. The dataset contains sure columns like State, District, year, Rape, dowry Deaths, snatch and Abduction, Assault on girls, Insult to modesty, Cruelty by husband and importation of ladies. This analysis is extremely vital to grasp within which state or within which District most variety of cases happened and within which Year what proportion specific offense happened and plenty of a lot of. to try to to this analysis we tend to used Hadoop, Pig, Hive, etc. Hadoop is one amongst the most effective and economical tools to figure on an outsized dataset, we are able to conjointly work on live information sets like twitter data by mistreatment Hadoop. There square measure regarding quite one hundred tools in Hadoop and every tool has its specific specification. It works on each style of information whether or not it's Structured, or Unstructured. Hadoop is extremely low cost as a result of it's ASCII text file and may store an outsized quantity of knowledge, it makes analysis terribly straightforward. we tend to simply got to apply queries to search out the vital info from the info. we are able to use it on Ubuntu and on Cloudera that is that the IBM info sphere. Predictive analytics is that the branch of the advanced analytics that is employed to create predictions regarding unknown future events. prophetic analytics uses several techniques from data processing, statistics, modeling, machine learning, and AI to research current information to create predictions regarding future. In our project we tend to square measure mistreatment machine learning to perform prophetic analysis.*

## I. INTRODUCTION

### A. Problem Definition

Increasing crime day by day is that the main issue before of human society. Crime happens once the personnel area or work area of wrongdoer and target intersects at one purpose. Target could also be single person or cluster of individuals or say a territory. Crime may well be accidental or it'd be planned. Accidental crime happens at any places. The cluster folks|of individuals} fight with others for atiny low matter which can harms the people that don't seem to be having any relation thereupon matter. Planned crime is that the crime that is enforced advisedly.

Crime mapping is employed to research, map and visualize crime incidents or crime pattern to possess an inspiration for predicting the crime prevalence. Crime mapping so helps the safety still as police to accommodate their resources consequently for preventing crime. these days each students still as practitioners have the capability to map the crime victimisation out there criminal spatial knowledge and with the assistance of developed advance technology.

### B. Project Overview

Crime analysis will be outlined because the study of daily reports and crime to work out the situation, time of day, special characteristics, and similarities to different crimes still as any vital knowledge could} or may determine the existence of patterns of criminal behavior. this is often done through the gathering, collation, analysis of reports, and analysis of crime knowledge.

Crime analysis could be a important initiative in crucial the necessity for crime/loss interference programs still as characteristic issues on the location. Crime analysis comes into play once more within the analysis of programs. is that the town rate up and yours down? you'll have an occasional rate during a high crime space. thus however does one get such information? Police departments break their crime down into areas or zones. every zone could also be given variety, for instance, one to eight and you're situated in zone six. Effective programming begins with results of relevant and reliable data/information.

## II. METHODOLOGY

This project is to analyse a knowledge set associated with crime records in several States from 2001 until 2012. It includes the knowledge and data regarding the varied crimes committed in several regions, at totally different Year and therefore the total variety of cases; It provides a State wise perspective analysis of the expansion and deviation between varied crimes committed.

The options of our project square measure given below:

- 1) Framing the info sector wise for higher understanding.
- 2) information munging for scouring or removing orthogonal information.  
To accomplish the projected work i.e. prophetic Analytics, we have a tendency to square measure engaged on Hadoop Ecosystems and Machine Learning. There square measure five basic steps in doing Crime Analysis:
  - a) information assortment
  - b) Classification\
  - c) Pattern Identification
  - d) Prediction
  - e) Visualization

### III. LITERATURE SURVEY

#### A. Existing System

In earlier time the info concerning crime area unit largely the police complaints, news paper's report and articles that area unit obtainable in hand written format or written however because the technological development advances the info concerning crime area unit obtainable in textual matter still as soft copy format.

Past situations area unit completely different because the lower rate was there, the info generated concerning criminal activities was additionally low. on it less quantity information{of knowledge{of information} ancient data analysis techniques area unit economical to research and predict the crime.

The past knowledge associated with criminal activities plays an important role in mapping crime and prediction of places wherever crime will occur.

Analyzing that knowledge obtainable in earlier time was terribly tedious and time overwhelming task by ancient data processing techniques despite the fact that the info was terribly less. knowledge generation today is huge thanks to inflated rate that can not be handled by ancient knowledge analysis techniques. This huge generated knowledge is huge knowledge which might be simply treated with the assistance of huge knowledge Analytics.

Previous knowledge is useful to predict the volatile places or say hotspots. once applying some data processing techniques like agglomeration, classification and alternative techniques the places having higher possibilities of crime to be occur were known and police capabilities may be allotted there.

Today the employment of net is increasing apace. the employment of net is additionally accountable to produce communication between criminals for finishing their targeted mission.

That the knowledge generation is in Brobdingnagian quantity that is generally in semi structured or unstructured format and might be analyzed exploitation agglomeration for giant knowledge. to research such Brobdingnagian quantity of knowledge either in semi-structured or unstructured format ancient data processing techniques aren't that a lot of capable. For that purpose knowledge huge knowledge Analytics is employed.

#### B. Proposed System

Using Python Tool and BigData Analytics, we have a tendency to area unit planning to do the crime analysis. Distribution of knowledge geographically is the initial section wherever the obtainable knowledge is distributed over geographical areas. Here the obtainable knowledge is said to crime.

Hadoop platform is employed for cluster analysis purpose that is second section. Clusters created in primary section area unit used as input this section and appropriate algorithms area unit applied up here for the analysis purpose. Hadoop will perform process{multiprocessing{data process} on completely different clusters the processing are going to be quick as compared to ancient processing capabilities. this can end in less time consumption and provides the output prior the conventional data {processing} cluster analysis process.

### C. Workflow

The progress work of the project flows by defining the issues related to the earlier analysis reports. The workflows as follow:



Fig.. I Workflow Diagram

1) The advancement goes as:

Purpose of the analysis and wish to hold the analysis – explained within the introduction section.

- 2) Data requirement: The knowledgeset used for the analysis are crime data (size=1.52 GB).It contains over sixty five, 00000 records. This embody the crime knowledge from 2001 until 2012 and that we have to be compelled to realize vital data from the info. In this, dataset of crime there square measure several columns State, District, Year, Rape, capture and Abduction, dower Deaths, Insult to modesty, cruelty by husband, Importation of ladies.
- 3) In the primary, the info can get loaded into the hive information through making table.
- 4) Furthermore, the analysis are performed on the info exploitation HQL.
- 5) Next, knowledge are foreign to the tablu for image.
- 6) In the analyzed results are accustomed perform prophetic analysis.

### D. Drawback Formulation

In today's world, each institution is facing ever growing challenges which require to be coped up quickly and with efficiency. this trend of analysis on Crime information Analytics, many variants of those algorithms are developed. this Project is to review and develop algorithms that provides a applied math illustration of the information from past Affairs yet as predicts what is going to happen in future from the Past information that is additionally called prognosticative information Analytics.

In this, we've got done crime analysis and that we need to notice vital data from the information. In this, dataset of crime there area unit several columns State, District, Year, Rape, capture and Abduction, gift Deaths, Insult to modesty, cruelty by husband, Importation of women. This project is to investigate a knowledge set associated with recorded crimes. It includes the knowledge and information regarding completely different offence committed. It offers AN area/region wise perspective analysis of the expansion and deviation between numerous countries offence. This information contains regarding ten thousand rows approx. So, we've got created a number of the issues or queries, for example:

- 1) However will the speed of crime changes with time?
- 2) That State had the foremost crime?
- 3) What's the count of explicit crime?
- 4) That area unit the highest states wherever crime happened the most?,

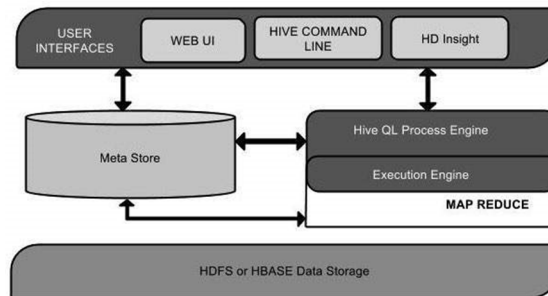
E. HIVE

Hive may be a information warehouse infrastructure tool to method structured information in Hadoop. It resides on prime of Hadoop to summarize massive information, and makes querying and analyzing straightforward.

- 1) Initially Hive was developed by Facebook, later the Apache software system Foundation took it up associate degree developed it more as an open supply underneath the name Apache Hive. it's employed by totally different corporations. as an example, Amazon uses it in It stores schema in a very info and processed information into HDFS.
- 2) It is meant for OLAP.
- 3) It provides SQL kind language for querying known as Hive or HQL.
- 4) It is acquainted, fast, scalable, and protractible. Amazon Elastic Map Reduce. Hive isn't
- 5) A on-line database
- 6) A style for on-line dealings process (OLTP)
- 7) A language for period of time queries and row-level updates Features of Hive

F. Architecture of Hive

The following part diagram depicts the design Hive:



G. HIVE: Marked Blue

```
hive> use crime;
hive> create table crimeyear (int_murder int, int_atkmeat_murder int, int_homicides int, int_cotodial Rape int, int_other Rape int, int_allopinning int, int_auto theft int, other_theft int,
> int_hurt int, int_dowry_details int, int_violence_relatives int, int_sporotation_girls int, int_death_engilence int, other_crimes int, total_crimes int);
Time taken: 2.287 seconds
hive> load data local inpath '/home/training/Desktop/crime.csv' overwrite into table crime;
Copying data from file:/home/training/Desktop/crime.csv
Copying file file:/home/training/Desktop/crime.csv
Loading data to table crime.crime
Deleted http://localhost/user/hive/warehouse/crime.db/crime

Time taken: 0.368 seconds
hive> select * from crime;
hive> use crime;
hive> select year, count(*)
from crime
group by year;
+-----+-----+
| year | count(*) |
+-----+-----+
| 2014 | 4380      |
| 2015 | 4189      |
| 2016 | 4190      |
| 2017 | 4191      |
| 2018 | 4192      |
| 2019 | 4193      |
| 2020 | 4194      |
+-----+-----+
Time taken: 4.866 seconds
hive> select year, count(*)
from crime
group by year;
+-----+-----+
| year | count(*) |
+-----+-----+
| 2014 | 4380      |
| 2015 | 4189      |
| 2016 | 4190      |
| 2017 | 4191      |
| 2018 | 4192      |
| 2019 | 4193      |
| 2020 | 4194      |
+-----+-----+
Time taken: 4.866 seconds
hive> select year, count(*)
from crime
group by year;
+-----+-----+
| year | count(*) |
+-----+-----+
| 2014 | 4380      |
| 2015 | 4189      |
| 2016 | 4190      |
| 2017 | 4191      |
| 2018 | 4192      |
| 2019 | 4193      |
| 2020 | 4194      |
+-----+-----+
Time taken: 4.866 seconds
```

```
hive> use crime;
hive> select year, count(*)
from crime
group by year;
+-----+-----+
| year | count(*) |
+-----+-----+
| 2014 | 4380      |
| 2015 | 4189      |
| 2016 | 4190      |
| 2017 | 4191      |
| 2018 | 4192      |
| 2019 | 4193      |
| 2020 | 4194      |
+-----+-----+
Time taken: 4.866 seconds
```

```

Applications - Hadoop System
[bin] job_201909240148_0005

2019-11-06 07:37:06,473 Stage-1 map = 100%, reduce = 100%
Map job = job_201909240148_0005
OK
2001 7601
2002 6823
2003 3284
2004 6170
2005 5711
2006 5420
2007 5080
2008 4544
2009 4534
2010 4401
2011 4091
2012 4966
2013 5047
2014 5110
Time taken: 13.587 seconds
View show crimes:
FAILED: Parse Error: Line 1:5 Failed to recognize predicate 'crimes'. Failed rule: 'addColumn' in show table grants
View desc crimes:
FAILED: Parse Error: Line 1:8 cannot recognize input near 'desc', 'crimes', '<EOF>'
View desc crimes:
year int
murder int
attemp number int
suicide int
suicide_rate int
other_type int
murder_rate int
other_crimes int
total_crimes int
Time taken: 0.040 seconds
View []

```

```

Applications - Hadoop System
[bin] job_201909240148_0005

1164 2003
1176 2004
1171 2005
1400 2006
1090 2007
4544 2008
4534 2009
4401 2010
4091 2011
4966 2012
5047 2013
5110 2014
Time taken: 13.615 seconds
View select year,max(murder) from crimes group by year:
totalMapReduce jobs = 1
Number of reduce tasks not specified, Estimated from input data size: 1
In order to change the average load per a reducer (in bytes):
  * set hive.exec.reducers.bytes.incr to a larger number
  * set hive.exec.reducers.max to a larger number
  * set mapred.reduce.tasks to a larger number
Start time = job_201909240148_0005, Tracking URL = http://localhost:50030/jobdetails.jsp?jobId=job_201909240148_0005
2019-11-06 07:36:57,394 Stage-1 map = 0%, reduce = 0%
2019-11-06 07:36:58,424 Stage-1 map = 100%, reduce = 0%
2019-11-06 07:37:06,473 Stage-1 map = 100%, reduce = 100%
Map job = job_201909240148_0005
OK
NULL NULL
2001 7601
2002 6823
2003 3284
2004 6170
2005 5711
2006 5420
2007 5080
2008 4544
2009 4534
2010 4401
2011 4091
2012 4966
2013 5047
2014 5110
Time taken: 13.597 seconds
View []

```

### H. Predictive Analytics

Predictive analytics flip the information into valuable, unjust info. prognostic analytics uses information to work out the probable outcome of an occurrence or a probability of a state of affairs occurring. Predictive analytics holds a spread of applied mathematics technique from modeling, machine, learning, data processing and theory of games that analyze current and historical facts to create prediction regarding future event.

There ar 3 basic cornerstones of prognostic analytics-

- 1) Predictive modeling
- 2) Decision Analysis and improvement
- 3) Transaction identification

Why is prognostic analytics important? Organizations ar turning to prognostic analytics to assist solve troublesome issues and uncover new opportunities.

- a) **Common uses Include:** Detecting fraud. Combining multiple analytics strategies will improve pattern detection and forestall criminal behavior. As cyber security becomes a growing concern, superior activity analytics examines all actions on a network in real time to identify abnormalities that will indicate fraud, zero-day vulnerabilities and advanced persistent threats. Optimizing selling campaigns. prognostic analytics ar wont to verify client responses or purchases, similarly as promote cross-sell opportunities. prognostic models facilitate businesses attract, retain and grow their most profitable customers. Improving operations. several corporations use prognostic models to forecast inventory and manage resources. Airlines use prognostic analytics to line price tag costs. Hotels try and predict the quantity of guests for any given night to maximise occupancy and increase revenue. Prognostic analytics allows organizations to operate additional expeditiously. Reducing risk. Credit scores ar wont to assess a buyer's probability of default for purchases and ar a widely known example of prognostic analytics. A credit score may be a variety generated by a prognostic model that includes all information relevant to a person's trustworthiness. different risk-related uses embrace insurance claims and collections. The wide used algorithms in information analysis ar rectilinear regression and neural network.
- b) **Linear Regression:** The straightforward regression model assumes that the linear relationship exists between the input and therefore the output variables.

c) *Neural Network*: A neural network impressed by the human brain, a network of neurons that are interconnected that's it's a collection of procedure units, that takes a collection of inputs and transfer the result to a predefined output. The procedure units are ordered organized in layers so the options of associate degree input vector are often connected with the feature of associate degree output vector.

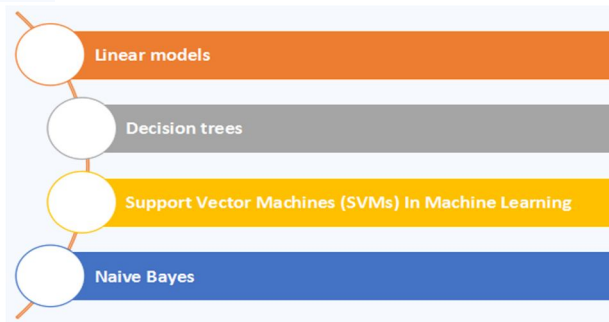
The idea behind usually this can be often to educate neural networks to model the relationships at intervals the provided information.

- Produce a model that is predicated on the principles established by the rule throughout the coaching part.
- Take a look at the model on the verification information set – the information is fed to the model and therefore the foretold values are compared to the particular values. therefore the model is tested for accuracy.
- Use the model on new incoming information and take action supported the output of the model.

### I. Other Vital Algorithms

Predictive models are available varied forms. There are totally different strategies that may be used to produce a model, and most of them are being developed all the time.

The most common prognostic models are:



- 1) *Linear Models*: It's a really wide used applied mathematics rule to make a relationship model between 2 variables. One variable is termed variable quantity whose worth is gathered through experiments, whereas the opposite variable is termed response variable whose worth springs from the variable quantity.
- 2) *Decision trees (also called Classification and Regression Trees or CART)*: It's a graph used to represent potentialities and their outcome within the variety of a tree. The nodes within the graph represent an occurrence or alternative and therefore the edges of the graph represent the choice rules or conditions.
- 3) *Support Vector Machines (SVMs)*: In Machine Learning: The support vector machine searches the nearest points and is thought as “support vectors” the name is as a result of the particular indisputable fact that points are like vectors that the only line “depends on” or is “supported by” the closest points.

Once it detects the nearest points, it attracts a line connecting them by doing vector subtraction (point A – purpose B). The support vector machine then declares the most effective separating line to be the road that bisects — and is perpendicular to — the connecting line

### J. Descriptive Analytics And Predictive Analytics: Marked Yellow

```

In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

In [2]: data = pd.read_csv('crime.csv')
data.head()

Out[2]:
   YEAR  MURDER  ATTEMPT  CULPRIBLE  CURTOSIAL  OTHER  NONAPPING  AUTO  OTHER  ROBTS  CHEATING  HURT/GREIVOUS  DOWNY  CRUELTY  BY  HUSBAND  OTHER  RELATIVES
0  2001    181    60      17        0        50        16    22    177    78    104        1131    16    175
1  2001    181    128        1        0    23        23    67    308    188    88        1643     7    154
2  2001    181    47        2        0    27        25    164    588    194    200        2888    14    186
3  2001     80    53        1        0    20         5    36    137    184    37         795    17    87
4  2001     82    87        1        0    23        23    190    871    79    229        1244    12    247

```

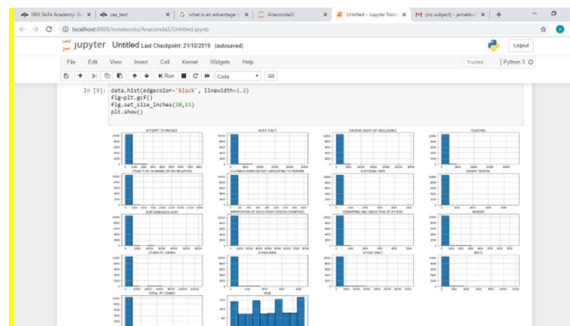
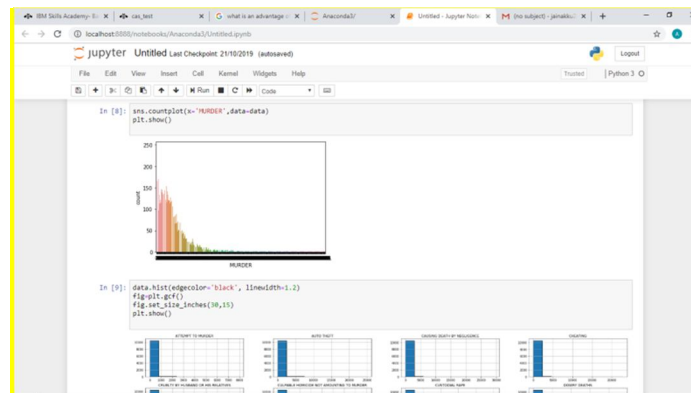
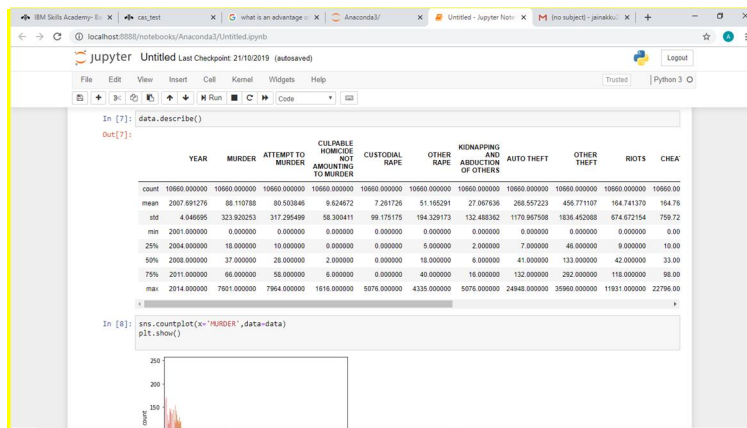
```

In [5]: data.size
Out[5]: 151888

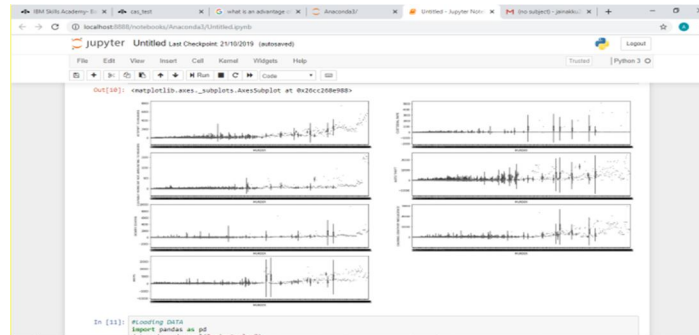
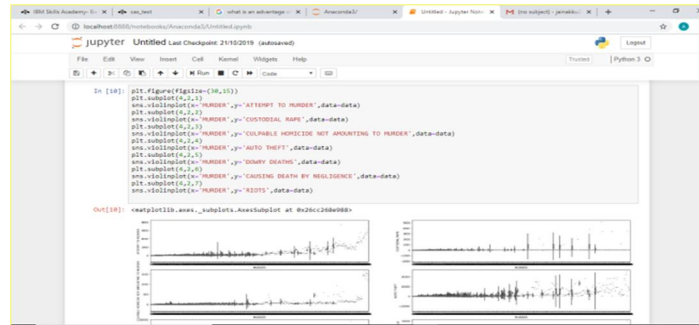
In [6]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10660 entries, 0 to 10659
Data columns (total 18 columns):
YEAR                10660 non-null int64
MURDER              10660 non-null int64
ATTEMPT TO MURDER  10660 non-null int64
CUSTODIAL RAPE     10660 non-null int64
OTHER RAPE         10660 non-null int64
KIDNAPPING AND ABDUCTION OF OTHERS  10660 non-null int64
AUTO THEFT         10660 non-null int64
OTHER THEFT        10660 non-null int64
RIOTS              10660 non-null int64
CHEATING           10660 non-null int64
HURT/FOREVSIUS HURT  10660 non-null int64
DOWRY DEATHS      10660 non-null int64
CRUELTY BY HUSBAND OR HIS RELATIVES  10660 non-null int64
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES  10660 non-null int64
CAUSING DEATH BY NEGLIGENCE          10660 non-null int64
OTHER IPC CRIMES                      10660 non-null int64
TOTAL IPC CRIMES                      10660 non-null int64
dtypes: int64(18)
memory usage: 1.5 MB

In [7]: data.describe()

```







```
In [11]: #Loading DATA
import pandas as pd
data=pd.read_excel("crimes.xlsx")
#Importing Algorithms
from sklearn.linear_model import LogisticRegression
from sklearn import svm
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import *
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
train,test=train_test_split(data,test_size=0.3)
print(train.shape)
print(test.shape)

(2892, 11)
(948, 11)

In [12]: train_x=train[["CRUELTY BY HUSBAND OR HIS RELATIVES","DOWRY DEATHS","CHEATING","RIOTS","AUTO THEFT","OTHER THEFT","KIDNAPPING AND
train_y=train.HURDER
test_x=test[["CRUELTY BY HUSBAND OR HIS RELATIVES","DOWRY DEATHS","CHEATING","RIOTS","AUTO THEFT","OTHER THEFT","KIDNAPPING AND
test_y=test.HURDER
print(train_x.head())
print(train_y.head())

CRUELTY BY HUSBAND OR HIS RELATIVES  DOWRY DEATHS  CHEATING  RIOTS \
852 2 2 1 1
2014 2 2 1 2
1066 4 2 1 2
1806 4 2 1 2
2812 4 1 1 2
```

```
In [12]: train_x=train[["CRUELTY BY HUSBAND OR HIS RELATIVES","DOWRY DEATHS","CHEATING","RIOTS","AUTO THEFT","OTHER THEFT","KIDNAPPING AND
train_y=train.HURDER
test_x=test[["CRUELTY BY HUSBAND OR HIS RELATIVES","DOWRY DEATHS","CHEATING","RIOTS","AUTO THEFT","OTHER THEFT","KIDNAPPING AND
test_y=test.HURDER
print(train_x.head())
print(train_y.head())

CRUELTY BY HUSBAND OR HIS RELATIVES  DOWRY DEATHS  CHEATING  RIOTS \
852 2 2 1 1
2014 2 2 1 2
1066 4 2 1 2
2812 4 3 1 2
1448 2 1 1 2

AUTO THEFT  OTHER THEFT  KIDNAPPING AND ABDUCTION OF OTHERS  OTHER RAPE \
852 5 3 1 18
2014 4 2 1 12
1066 4 5 1 0
2812 4 3 2 6
1448 4 3 2 4

CULPABLE HOMICIDE NOT AMOUNTING TO HURDER  ATTEMPT TO HURDER
852 7 6
2014 11 9
1066 12 13
2812 9 9
1448 12 16
```

```

Cruelty by Husband or his Relatives  DOWRY DEATHS  CHEATING  RIOTS  \
852 2 2 1 1
2814 2 2 1 2
1066 4 2 1 2
2812 4 3 1 2
1448 2 1 1 2

AUTO THEFT  OTHER THEFT  KIDNAPPING AND ABDUCTION OF OTHERS  OTHER RAPE  \
852 5 3 3 18
2814 4 2 1 12
1066 4 5 1 8
2812 4 3 2 6
1448 4 3 2 4

CULPABLE HOMICIDE NOT AMOUNTING TO MURDER  ATTEMPT TO MURDER
852 7 6
2814 11 9
1066 12 13
2812 9 9
1448 12 16

Name: MURDER, dtype: int64

In [13]: lr= LogisticRegression()
lr.fit(train_x,train_y)
predictions= lr.predict(test_x)
print ("accuracy of logistic regression", accuracy_score(predictions,test_y))
print("Mean Absolute Error:", mean_absolute_error(predictions, test_y))

```

```

In [13]: lr= LogisticRegression()
lr.fit(train_x,train_y)
predictions= lr.predict(test_x)
print ("accuracy of logistic regression", accuracy_score(predictions,test_y))
print("Mean Absolute Error:", mean_absolute_error(predictions, test_y))

C:\Users\jains\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:432: FutureWarning: Default solver will be changed
to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning:
C:\Users\jains\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:469: FutureWarning: Default multi_class will be cha
nged to 'auto' in 0.22. Specify the multi_class option to silence this warning.
"this warning.", FutureWarning)

accuracy of svm 0.3404251510148936
Mean Absolute Error: 1.3925531914893616

In [14]: from sklearn.metrics import *
from sklearn import svm
model= svm.SVC()
model.fit(train_x,train_y)
prediction= model.predict(test_x)
print ("accuracy of svm", accuracy_score(prediction,test_y))
print("Mean Absolute Error:", mean_absolute_error(prediction, test_y))

C:\Users\jains\Anaconda3\lib\site-packages\sklearn\svm\base.py:193: FutureWarning: The default value of gamma will change from
'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid t
his warning.
"avoid this warning.", FutureWarning)

accuracy of svm 0.7574468885186383
Mean Absolute Error: 1.3925531914893616

```

```

In [14]: from sklearn.metrics import *
from sklearn import svm
model= svm.SVC()
model.fit(train_x,train_y)
prediction= model.predict(test_x)
print ("accuracy of svm", accuracy_score(prediction,test_y))
print("Mean Absolute Error:", mean_absolute_error(prediction, test_y))

C:\Users\jains\Anaconda3\lib\site-packages\sklearn\svm\base.py:193: FutureWarning: The default value of gamma will change from
'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid t
his warning.
"avoid this warning.", FutureWarning)

accuracy of svm 0.7574468885186383
Mean Absolute Error: 1.3925531914893616

In [15]: from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=9)
knn.fit(train_x,train_y)
prediction=knn.predict(test_x)
print ("accuracy of knn", accuracy_score(prediction,test_y))
print("Mean Absolute Error:", mean_absolute_error(prediction, test_y))

accuracy of knn 0.4991744688851864
Mean Absolute Error: 1.3925531914893616

In [16]: #Dcts on Tree Classifier
from sklearn import tree
dct=tree.DecisionTreeClassifier(random_state=0)
dct.fit(train_x,train_x)
prediction= dct.predict(test_x)

```

```

In [16]: #DecisionTreeClassifier
from sklearn import tree
dct=tree.DecisionTreeClassifier(random_state=0)
dct.fit(train_x,train_y)
prediction= dct.predict(test_x)
print("accuracy of Decision Tree Classifier", accuracy_score(test_y, prediction))
print("Mean Absolute Error:", mean_absolute_error(predictions, test_y))

accuracy of Decision Tree Classifier 0.8425331914893617
Mean Absolute Error: 1.3925531914893616

In [17]: #RandomForest
rf=RandomForest()
rf.fit(train_x,train_y)
predictions= rf.predict(test_x)
print("accuracy of RandomForest", accuracy_score(test_y, predictions))
print("Mean Absolute Error:", mean_absolute_error(predictions, test_y))

accuracy of RandomForest 0.4780227699774467
Mean Absolute Error: 2.08810818697972

In [19]: #for function is attached to 'Show Result' button
def fun():
    i1=i1+1
    a1=CHEATING,MURDER,HISRELATIVES,GET()
    a2=DOWRYDEATHS,GET()
    a3=CHEATING,GET()
    a4=RAPES,GET()
    a5=AUTOTHEFT,GET()

```

```

In [29]: #fun function is attached to 'Show Result' button

def fun():
    list1=[]
    a1=CRUELTYBYHUSBANDORHISRELATIVES.get()
    a2=DOINGDEATHS.get()
    a3=CHEATING.get()
    a4=RIDTS.get()
    a5=AUTOHEFT.get()
    a6=OTHERHEFT.get()
    a7=KIDNAPPINGANDABDUCTIONOFOTHERS.get()
    a8=OTHERRAPE.get()
    a9=CULPABLEHOMICIDENOTAMOUNTINGTOMURDER.get()
    a10=ATTEMPTTOMURDER.get()

    list1.append(a1)
    list1.append(a2)
    list1.append(a3)
    list1.append(a4)
    list1.append(a5)
    list1.append(a6)
    list1.append(a7)
    list1.append(a8)
    list1.append(a9)
    list1.append(a10)

    ind=[CRUELTYBYHUSBANDORHISRELATIVES, 'DOINGDEATHS', 'CHEATING', 'RIDTS', 'AUTOHEFT', 'OTHERHEFT', 'KIDNAPPINGANDABDUCTIONOFOTHERS', 'CULPABLEHOMICIDENOTAMOUNTINGTOMURDER', 'ATTEMPTTOMURDER']

    text_xi=pd.DataFrame(list1, index=ind)

    model=DecisionTreeClassifier
    model.fit(train_x, train_y)
    
```

```

prediction=model.predict(test_xi.T)
if prediction==0:
    lab=Label(f1,text="RESULT- CHANCES OF MURDER",font = ('arial', 14, 'bold'),bd = 14, anchor = "w",bg="lavender")
else:
    lab=Label(f1,text="RESULT- SAFE",font = ('arial', 14, 'bold'),bd = 14, anchor = "w",bg="red")
lab.grid(row=9,column=1)

In [30]: #Reset Function will reset the entry widgets
def reset():
    CRUELTYBYHUSBANDORHISRELATIVES.set("")
    DOINGDEATHS.set("")
    CHEATING.set("")
    RIDTS.set("")
    AUTOHEFT.set("")
    OTHERHEFT.set("")
    KIDNAPPINGANDABDUCTIONOFOTHERS.set("")
    OTHERRAPE.set("")
    CULPABLEHOMICIDENOTAMOUNTINGTOMURDER.set("")
    ATTEMPTTOMURDER.set("")

    lab=Label(f1,text="")
    lab.grid(row=9,column=1)

In [31]: # Import tkinter module
from tkinter import *

# Import other necessary modules
import time
    
```

```

In [32]: # Import tkinter module
from tkinter import *

# Import other necessary modules
import time
import datetime

root=Tk()
root.geometry("1400x700")
root.title("Crime Prediction")

#Top Frame for Title and Time
tops = Frame(root, width = 1600, relief = SUNKEN)
tops.pack(side = TOP)

#Left Frame for Label, Entries and Button
f1 = Frame(root, width = 800, height = 700, relief = SUNKEN)
f1.pack(side = LEFT)

localtime = time.asctime(time.localtime(time.time())) #for time

#Title Label
labelInfo = Label(tops, font = ('helvetica', 33, 'bold'), text = " Crime \n Analysis and Prediction",fg = "black", bd = 30, anchor = "center", grid = (row = 0, column = 0))
#Time Label
labelInfo = Label(tops, font = ('arial', 14, 'bold'), text = localtime, fg = "Steel Blue", bd = 18, anchor = "w")
labelInfo.grid(row = 1, column = 0)

#Creating the Menu
menu=Menu(root)
    
```

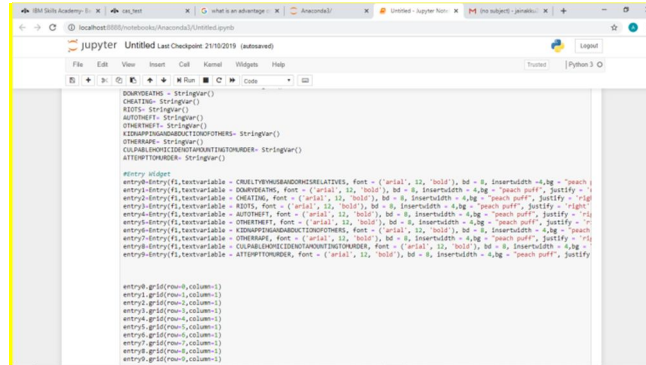
```

root.config(menu=menu)
subMenu=Menu(menu)

#Title Menu
menu.add_cascade(label="File", menu=subMenu) #main menu
subMenu.add_command(label="Reset", command=reset) #submenus
subMenu.add_command(label="Exit", command=root.destroy)

#Labels
label0=Label(f1,font = ('arial', 13, 'bold'),text="CRUELTY BY HUSBAND OR HIS RELATIVES",bd = 14, anchor = "w")
label0.grid(row=0,column=0,sticky=E)
label1=Label(f1,text="DOING DEATHS",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label1.grid(row=1,column=0,sticky=E)
label2=Label(f1,text="CHEATING",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label2.grid(row=2,column=0,sticky=E)
label3=Label(f1,text="RIDTS",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label3.grid(row=3,column=0,sticky=E)
label4=Label(f1,text="AUTO HEFT",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label4.grid(row=4,column=0,sticky=E)
label5=Label(f1,text="OTHER HEFT",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label5.grid(row=5,column=0,sticky=E)
label6=Label(f1,text="KIDNAPPING AND ABDUCTION OF OTHERS",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label6.grid(row=6,column=0,sticky=E)
label7=Label(f1,text="OTHER RAPE",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label7.grid(row=7,column=0,sticky=E)
label8=Label(f1,text="CULPABLE HOMICIDE NOT AMOUNTING TO MURDER",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label8.grid(row=8,column=0,sticky=E)
label9=Label(f1,text="ATTEMPT TO MURDER",font = ('arial', 13, 'bold'),bd = 14, anchor = "w")
label9.grid(row=9,column=0,sticky=E)

#Text Variables For Entry
CRUELTYBYHUSBANDORHISRELATIVES = StringVar()
    
```



```

DOORWEIGHT = StringVar()
ONERATING = StringVar()
RSDTS = StringVar()
AUTOFRETT = StringVar()
ONERMEFF = StringVar()
CONFERREDBASCONSTRONERS = StringVar()
ONERNAME = StringVar()
COLPARELONCECONSTRONETONER = StringVar()
ATTEMPTONER = StringVar()

entryWidget
entry0=Entry(f1,textvariable = ONERATING,font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry1=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry2=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry3=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry4=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry5=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry6=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry7=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry8=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")
entry9=Entry(f1,textvariable = ONERATING, font = ('arial', 12, 'bold'), bd = 8, insertwidth = 4, bg = "search puff", justify = "left")

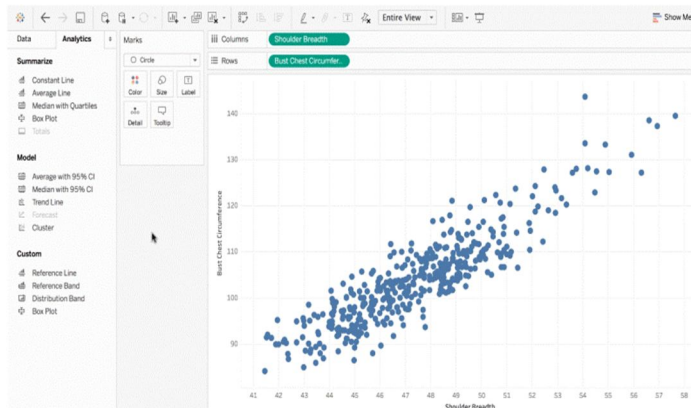
entry0.grid(row=0,column=1)
entry1.grid(row=1,column=1)
entry2.grid(row=2,column=1)
entry3.grid(row=3,column=1)
entry4.grid(row=4,column=1)
entry5.grid(row=5,column=1)
entry6.grid(row=6,column=1)
entry7.grid(row=7,column=1)
entry8.grid(row=8,column=1)
entry9.grid(row=9,column=1)
    
```

**K. Tableau Visualization**

Tableau may be a knowledge visual image tool that's wide used for Business Intelligence however isn't restricted to that. It helps produce interactive graphs and charts within the kind of dashboards and worksheets to achieve business insights. and every one of this is often created potential with gestures as straightforward as drag and drop!

	Key Features	Other Features	Operating System	License
<b>Tableau Desktop</b>	Creating Dashboards and Stories locally	Tableau Personal - limited data sources, non connectivity to Tableau Server  Tableau Professional - Full enterprise capabilities	Windows, Mac	Personal - \$999 Professional - \$1999
<b>Tableau Public</b>	A Massive, public, non commercial Tableau Server	All data published in public	-	Free
<b>Tableau Online</b>	Creating Dashboards and Stories on the Cloud	Live Connections	-	\$500 per year per
<b>Tableau Reader</b>	View Dashboards and Sheets locally	Cannot modify workbooks or connect to the server	Windows, Mac	Free
<b>Tableau Server</b>	Connect to Data sources and share Dashboards	Users can directly interact with Dashboards via browser	Windows	Core Licensing

Data visualisation is another kind of visual art that grabs our interest and keeps our eyes on the message. once we see a chart, we tend to quickly see trends and outliers. If we are able to see one thing, we tend to internalise it quickly. It's storytelling with a purpose. If you've ever stared at a vast computer program of knowledge and couldn't see a trend, you recognize what proportion more practical a visualisation are often.



When you think about knowledge visualisation, your initial thought most likely forthwith goes to straightforward bar graphs or pie charts. whereas these could also be associate integral a part of visualizing knowledge and a standard baseline for several knowledge graphics, the correct visualisation should be paired with the correct set of data. straightforward graphs ar solely the tip of the iceberg. There's a full choice of visualisation strategies to gift knowledge in effective and attention-grabbing ways in which.

1) Common General forms of Knowledge Visualization

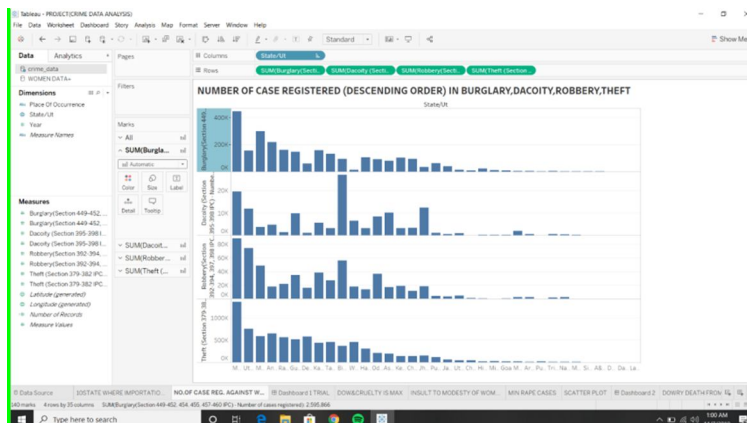
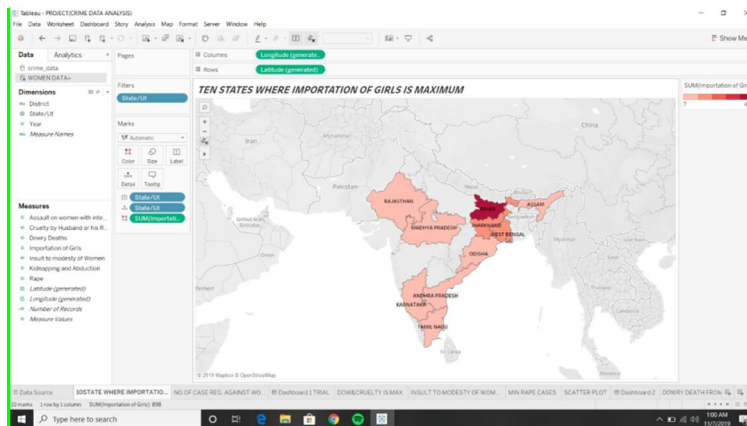
- Charts
- Tables
- Graphs
- Maps
- Infographics
- Dashboards

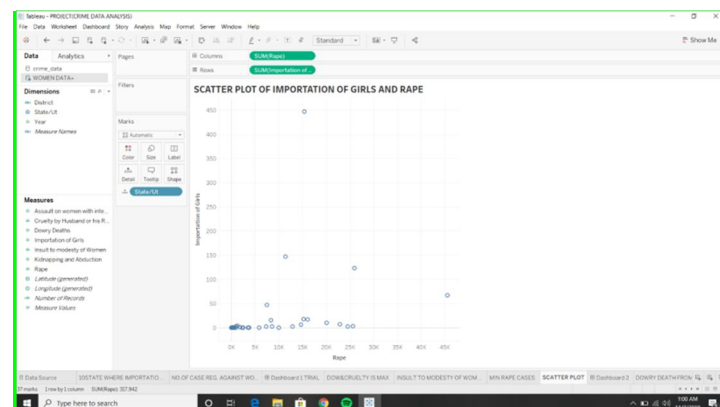
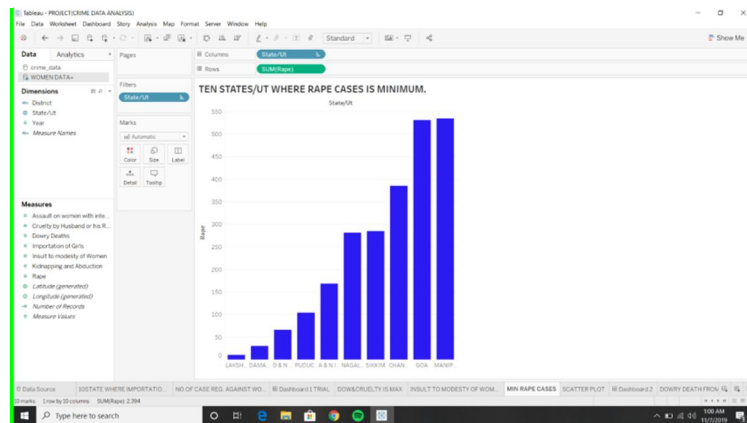
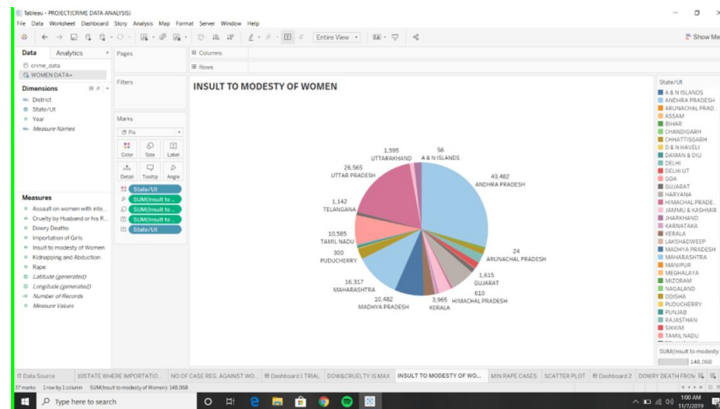
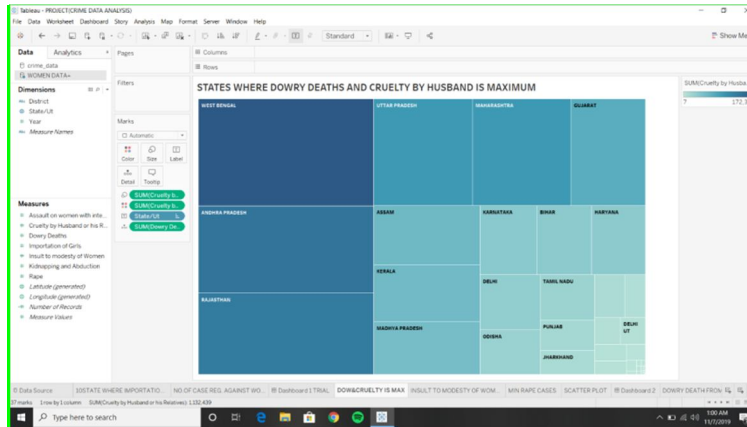
2) More specific Samples of Strategies to check Data

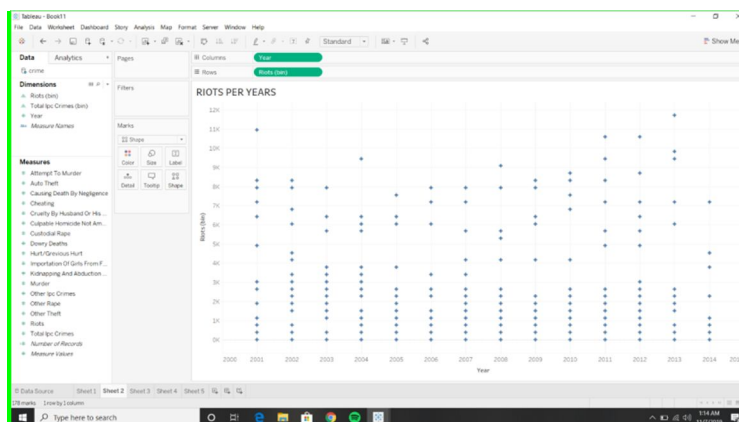
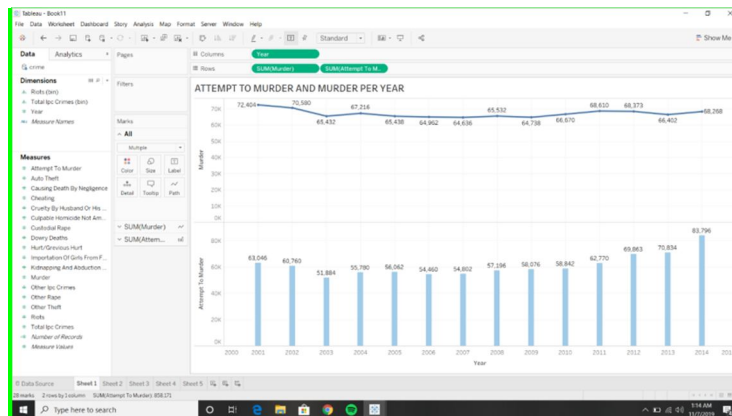
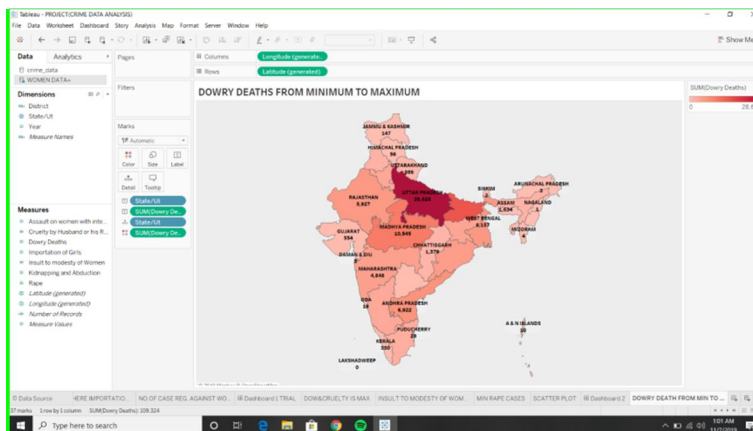
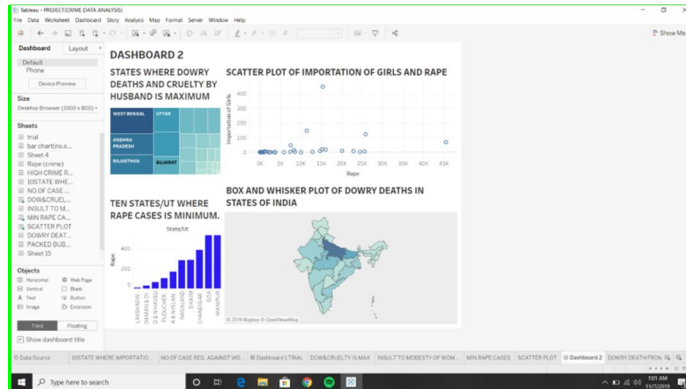
- Area Chart
- Bar Chart
- Box-and-whisker Plots
- Bubble Cloud
- Bullet Graph

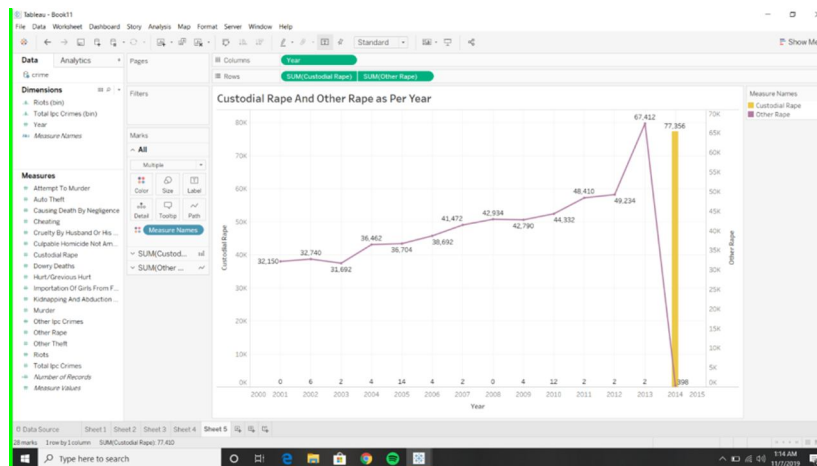
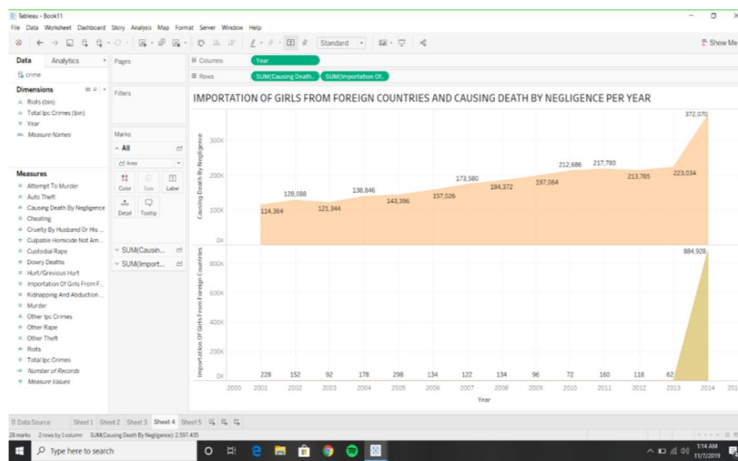
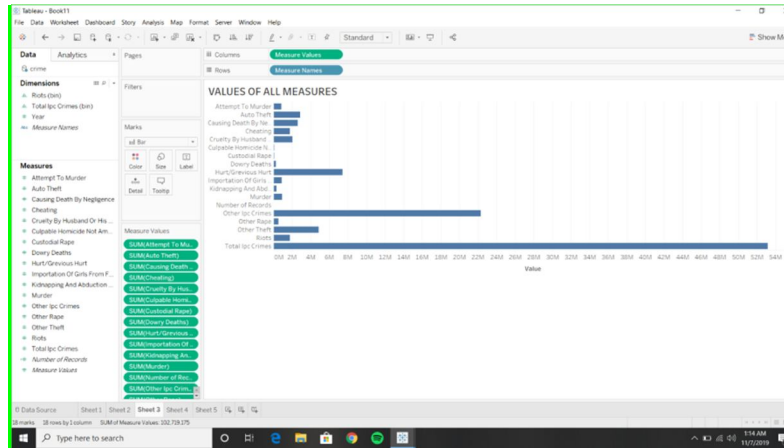
- Cartogram
- Circle read
- Dot Distribution Map
- Gantt Chart
- Heat Map
- Highlight Table
- Histogram

L. Visualization Using Tableau: Marked Green









#### IV. CONCLUSION

From this we are able to conclude that we are able to simply notice something from the dataset whether or not it'll be tiny or giant. We are able to run any variety of queries in Hadoop and dealing on cludera is incredibly simple than alternative software package, all the packages, Hadoop tools, java everything is pre-installed in it. we tend to don't have to be compelled to install something in it. BIn Future huge knowledge has additional scope than the rest and in Big-data Hadoop is should. firms area unit getting Hadoop additional and additional and might get smart package if you master in Hadoop. Hadoop is affordable thus associate one will acquire it as a result of it's an open supply. And we can even predict rate and kind of crime which will be happening in several states of our country.



## V. FUTURE WORK

Our purpose was to review the ways utilized for prognostic crime mapping, from basic approaches presently utilized by crime analysts, to classic models developed by researchers. Our technique was to assess each prediction approach on the premise of accuracy, knowledge needs, hardware and software system needs, and simple use. What we learned is that in several respects, this review is premature. The more refined approaches delineated during this chapter area unit still terribly much within the development stages and might best be thought-about "alpha versions" that have however to be tested by the tip users. Our review conjointly suffers from the actual fact that there area unit few printed, refereed works on these ways. what is more, the variability of the ways themselves precludes a head-on comparison of the accuracy of 1 versus another. nevertheless, this can be associate acceptable purpose within which to require -48 -Forecasting the long run of prognostic Crime Mapping stock of each current observe and future development, as this review may inform mid-course corrections. Perhaps the foremost necessary finding from this review is that, while technology has improved our ability to form, maintain and manipulate knowledge, there's still abundant work to be done before we are able to effectively forecast crime trends. GIS has enabled the creation of geographic knowledge (both crime and crime-related) and also the integration of data from a spread of sources. However, the foremost of times used methods for evaluating the info area unit an equivalent ones that are in use for concerning thirty years. In fact, this state of data looks to indicate that a minimum of 2 of the present and comparatively easy ways (those supported exponential smoothing) area unit as effective — if {not additional|less|no more} thus — than more advanced ones. This is to not say that additional examination of latest ways ought to be abandoned. Recent innovations have however to achieve widespread acceptance and that we won't take care of their accuracy till additional analysis is conducted. Specifically, so as to raised measure the methods that are created, the sphere wants additional head-to-head comparisons of a number of the newer, additional refined ways versus the additional ancient, univariate ones. These comparisons should also take into account the question of what qualifies as associate "accurate" prediction, each in relevancy the dimensions of the anticipated space and also the quality of the prediction itself. could be a 500-foot grid cell necessary, or will a 4,000-foot space (such because the one utilized by Gorr et al.) suffice? However accurate will the spatial prediction have to be compelled to be so as to inspire confidence by police officers? If we are able to determine that there'll be associate increase in crime, however correct will the prediction really want to be for effective intervention? Finally, however tiny will the unit of study need to be so as to be much useful? These area unit necessary questions for additional investigation; till they're answered, individual researchers and analysts can still experiment with their own ways — probably reinventing the wheel — instead of learning from one another. One distressful side of the many of the ways reviewed here is that the lack of guiding theory, that not solely will facilitate develop higher models but conjointly helps US interpret a model's performance in order that interference efforts area unit improved. The accuracy of the variable ways depends upon the appropriateness of the variables enclosed within the model. Since the identification of acceptable variables is grounded in theory, one while not the opposite can solely have restricted utility. Many of these ways also are terribly sophisticated, requiring a high level of specialized applied mathematics and modeling experience still as an oversized volume of knowledge that don't seem to be usually simply out there on a small level. Ultimately, our goal is to seek out comparatively a straightforward technique that's t spots area unit likely to emerge in sure locations, thus on inform interference efforts.

## REFERENCES

- [1] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225- 230). IEEE.
- [2] Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.
- [3] <https://www.kaggle.com/c/sf-crime>.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)