**CRIME RATE PREDICTION**

**INT 300 – INTERNSHIP 2**

**PROJECT REPORT**

***Submitted by***

**SUDHARSHINI R – E0322019**

***In partial fulfilment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**(Artificial Intelligence and Data Analytics)**

**Sri Ramachandra Faculty of Engineering and Technology**

**Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai -600116**

**APRIL 2024**

**CRIME RATE PREDICTION**

**INT 300 – INTERNSHIP 2**

**PROJECT REPORT**

***Submitted by***

**SUDHARSHINI R – E0322019**

***In partial fulfilment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**(Artificial Intelligence and Data Analytics)**

**Sri Ramachandra Faculty of Engineering and Technology**

**Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai -600116**

**APRIL, 2024**

**BONAFIDE CERTIFICATE**

Certified that this project report **“CRIME RATE PREDICTION”** is the bonafide record of work done by **“SUDHARSHINI R – E0322019”** who carried out the internship work under my supervision.

**Signature of the Supervisor Signature of Programme Coordinator**

|  |  |
| --- | --- |
| **Dr. A Sathya**  **Associate Professor,**  Department of Artificial Intelligence and Data Analytics  Sri Ramachandra Faculty of Engineering and Technology,  SRIHER, Porur, Chennai-600 116. | **Dr. Uma Satya Ranjan**  **Professor and Head,**  Department of Artificial Intelligence and Data Analytics  Sri Ramachandra Faculty of Engineering and Technology,  SRIHER, Porur, Chennai-600 116. |

**Evaluation Date:**

**Signature of the Supervisor**

**ACKNOWLEDGEMENT**

I express my sincere gratitude to our Dean of Sri Ramachandra Faculty of

Engineering and Technology, **Dr. T. Ragunathan** for providing the required

facilities for this study.

I would also like to extend heartfelt thanks to our Head of the Department

**Dr. Uma Satya Rajan** for helping us with additional support during this study.

I wish to thank my faculty supervisor, **Dr.A. Sathya,** Department of Artificial Intelligence and Data Analytics, Sri Ramachandra faculty of Engineering and Technology for extending help and encouragement throughout the project. Without his/her continuous guidance and persistent help, this project would not have been a success for me.

I am grateful to all the members of Sri Ramachandra Faculty of Engineering and Technology, my beloved parents and friends for extending the support, who helped us to overcome obstacles in the study.

**TABLE OF CONTENTS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CHAPTER NO.** | | **TITLE** | | **PAGE NO.** | |
|  | | **ABSTRACT** | **6** | |
|  | | **LIST OF SYMBOLS** | **7** | |
|  | | **LIST OF FIGURES** | **8** | |
| **1** | | **INTRODUCTION** | **10** | |
| **2** | | **LITERATURE REVIEW** | **11** | |
| **3** | | **PROPOSED METHODOLOGY**  3.1. Techniques involved  3.1.1. Exploratory data analysis  3.1.2. K-means clustering  3.1.3. ARIMA modelling | **12** | |
| **4** | | **IMPLEMENTATION**  4.1. Data Collection  4.2. Data Preprocessing  4.3. Visualisation  4.4. Clustering Analysis  4.5. Forecasting with ARIMA Model  4.6. Prediction for Uttar Pradesh (UP) - Dowry Deaths  4.7. Evaluation and Visualization  4.8. Conclusion and Recommendations | **14** | |
| **5** | | **RESULTS AND DISCUSSIONS** | **17** | |
|  | | **APPENDICES** | **19** | |
|  | | 6.1.Appendix-1: Code – Technical detail |  | |
|  | | 6.2.Appendix-2: Screenshots |  | |
|  | | **REFERENCES** | **35** | |
|  | | **WORKLOG** | **38** | |
|  |  | | |  | | |

**ABSTRACT**

In India with a women population of 1.45 billion urbanisation has created a massive change with respect to women safety. Crime against Women affects the women's safety and their quality of living. Certain action has to be taken to analyse and predict crime activity which is happening in various parts of the country. Using this Analysis, it is able to identify the trends used by the criminals and alert the women. There are numerous databases on criminal activity against women available on the internet for the researchers to analyse and predict the trends. The proposed model acts as an eye opener to women about the different crime activities happening around them and their patterns. Among these algorithms, K-means clustering stands out for its ability to uncover hidden patterns and group similar types of criminal activities together. Additionally, the Autoregressive Integrated Moving Average (ARIMA) model is utilized to forecast future crime trends based on historical data.

**LIST OF SYMBOLS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **ABBREVIATION** | **EXPANSION** |
| **1.** | PANDAS | PYTHON DATA ANALYSIS LIBRARY |
| **2.** | NUMPY | NUMERICAL PYTHON |
| **3.** | SKLEARN | SCIKIT-LEARN |
| **4.** | MATPLOTLIB | MATHEMATICAL PLOTTING LIBRARY |
| **5.** | PLOTY EXPRESS | GRAPHING LIBRARY |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **FIGURE NAME** | **PAGE NUMBER** |
| **1.** | Block Diagram | **12** |
| **2.** | Kmeans head after Scaling | **17** |
| **3.** | Classifying states into safe/unsafe | **17** |
| **4.** | ARIMA model | **18** |
| **5.** | Total cases vs crimes | **27** |
| **6.** | States vs Total cases | **27** |
| **7.** | States vs Rape cases | **28** |
| **8.** | States vs Kidnapping and Abduction | **28** |
| **9.** | States vs Dowry Deaths | **29** |
| **10.** | States vs Assault on women with intent to outrage her modesty | **29** |
| **11.** | States vs Insult to modesty of Women | **30** |
| **12.** | States vs Importation of girls | **30** |
| **13.** | All crimes dataset | **31** |
| **14.** | Kmeans head after Scaling | **31** |
| **15.** | Classifying states into safe/unsafe | **32** |
| **16.** | Year vs Dowry Deaths | **32** |
| **17.** | Partial Autocorrelation | **33** |
| **18.** | acf plot | **34** |
| **19.** | pacf plot | **34** |

**CHAPTER 1**

**INTRODUCTION**

As societal, political, and economic dynamics shifted, women gradually lost their prominence, often being relegated to the sidelines. This decline was accompanied by the emergence of detrimental customs and traditions, which bound women within the confines of their households.

Disturbing trends, such as declining sex ratios, deteriorating health indicators, low literacy rates, and limited participation in various spheres, underscored the challenges faced by women.

Moreover, the proliferation of social maladies like dowry-related deaths, underage marriages, domestic abuse, sexual violence, and exploitation in the workplace further exacerbated the situation. Instances of humiliation, abduction, molestation, dowry-related fatalities, torture, and spousal abuse have surged over time, casting a shadow over women's safety and well-being.

Predicting crime rates against women in India poses a complex challenge. Efforts to combat these issues must involve comprehensive strategies encompassing education, legal reforms, community empowerment, and stringent law enforcement.

Moreover, fostering a culture of respect, equality, and empathy towards women is crucial in addressing the root causes of gender-based violence and ensuring a safer and more equitable society for all.

**CHAPTER 2**

**LITERATURE REVIEW**

Researchers compared many data mining algorithms using a variety of real-world applications with some related works. Apart from those one of them is Crime against Women (CAW) Analysis and Prediction in Tamil Nadu Police Using Data Mining Techniques where they used Clustering in WEKA utensils, Euclidean distance calculation by **S. Lavanyaa, D. Akila [1]** which states that Clustering in WEKA utensils, Euclidean distance calculation gives improved exactness in the metropolitan urban areas violations rate to decrease and predict, according to the investigation. Analysis and Prediction of Crimes using Clustering, Classification and General algorithm by **Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi [2]** where the main objective of occurrence frequency during different years. We used a theoretical model focused on data mining techniques including clustering and classification to analyse a real crime dataset collected by the police in England and Wales between 1990 and 2011. In this certain kind of weights are assigned so as to refine the model which is being used and eliminating the lesser values. Using the RapidMiner tool, the Genetic Algorithm (GA) is used to optimise the Outlier Detection operator parameters. Crime Analysis using K-Means Clustering on crime dataset using rapid miner tool by **Jyoti Agarwal ,Renuka Nagpal, Rajni Sehgal [3]** where this project is based on the concept of crime analysis by using clustering algorithm on the obtained data set using a rapid miner tool. The analysis is done by taking in a particular type of crime that is homicide and presenting it in a graphical form, matching it with to the particular year it happened. The outcome derived was, that the rate of homicide is decreasing from year 1990 to 2011. Crimes Against Women in India: Analysis of Trends Using Regression and Visualization by **R. Devakunchari, Bhowmick S, Bhutada S P, Shishodia Y [4]** where Detection technologies improves identifying of the incidents and make use of public safety equipment as soon as possible. With faster identification of incidents, this helps to improve the response time and thus in this the accuracy and reliability of incident response and reporting, as well as the distribution of investigative resources, can all be improved with technology. This can also help in boosting up the clearing rates.

**CHAPTER 3**

**PROPOSED METHODOLOGY**

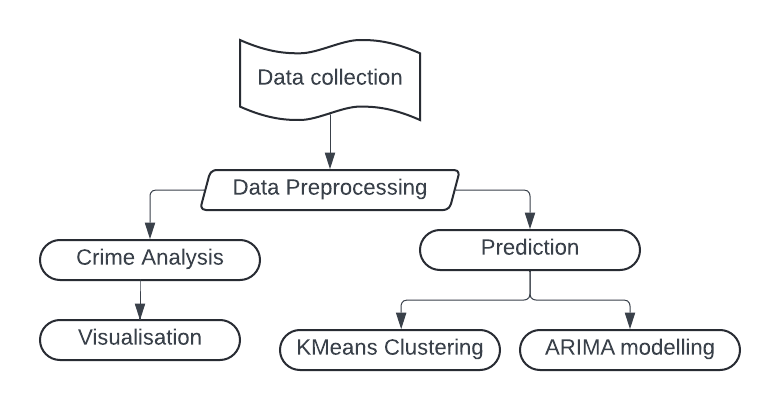


Fig.1Block Diagram

This section elaborates on the methodology employed to analyze and predict crime rates against women in India for the years 2001 to 2014, utilizing data gathered from a Kaggle dataset. The objective is to visualize the analysis and prediction of crime against women across different regions of the country based on historical criminal records.

Before applying machine learning algorithms, rigorous data preprocessing was conducted. This involved the removal of undefined and unnamed features from the Kaggle dataset, as they held no relevance to the analysis. By cleansing the dataset, we ensured that only pertinent information was utilized for modelling.

The system leverages two primary techniques: KMeans clustering and ARIMA (AutoRegressive Integrated Moving Average) modelling.

KMeans clustering is employed to identify distinct clusters or groups within the dataset. By grouping regions with similar crime patterns, this technique enables us to gain insights into geographical variations in crime rates against women over the years.

ARIMA, a time series forecasting method, is utilized to predict future crime rates against women based on historical data trends. By analyzing temporal patterns and incorporating seasonality and trends, ARIMA facilitates accurate predictions for the forthcoming years.

**3.1. TECHNIQUES INVOLVED**

**3.1.1. EXPLORATORY DATA ANALYSIS**

EDA involves examining the data to understand its characteristics, identifying patterns, correlations, and potential outliers that may impact the analysis.

**3.1.2. K-MEANS CLUSTERING:**

Utilizing the k-means clustering algorithm to group states/UTs into two clusters based on selected features, such as crime rates, healthcare access, and socio-economic indicators, to categorize them as "safe" or "unsafe."

**3.1.3. ARIMA MODELLING:**

Implementing the ARIMA (AutoRegressive Integrated Moving Average) model, specifically SARIMAX, to analyse time series data on dowry deaths. This involves fitting the model to historical data, identifying the appropriate parameters (e.g., auto-regressive, differencing, and moving average terms), and assessing the model's performance using various metrics and diagnostic tests.

**CHAPTER 4**

**IMPLEMENTATION**

**4.1. Data Collection:**

The dataset, sourced from Kaggle, spans the years 2001 to 2014 and comprises 10,677 entries in a pandas DataFrame format. The DataFrame consists of 11 columns such as STATE/UT, DISTRICT, Unnamed: 0, Year, Rape, Kidnapping and Abduction, Dowry Deaths, Assault on women with intent to outrage her modesty, Insult to modesty of Women, Cruelty by Husband or his Relatives, Importation of Girls each containing valuable information for analysis and insights generation.

**4.2. Data Preprocessing:**

The dataset was pre-processed using techniques such as data cleaning, handling missing values, and transforming variables as needed. Remove undefined and unnamed features, and ensure consistency in data formats. An unnamed column was removed from the dataset, repeated states were identified and eliminated from the dataset, all entries were converted to capital letters and the DataFrame was filtered to include only columns containing numeric data.

**4.3. Visualisation:**

Visualizing the crime dataset is essential for gaining insights into the patterns and trends of crimes against women across different regions and over time.

Bar plots are utilized to compare the prevalence of different crime types across various states and union territories (UTs).

Each bar represents a state/UT, and the height of the bar corresponds to the count of a specific crime type, such as rape cases, dowry deaths, or importation of young girls.

This visualization enables us to identify regions with higher or lower rates of specific crimes and assess geographical disparities.

**4.4. Clustering Analysis:**

The KMeans clustering algorithm was used to segment the dataset into two clusters based on crime patterns. The mean of crime rates against women as the threshold. Regions with crime rates above the mean are categorized as 'Unsafe', while those below are labelled 'Safe'.

**4.5. Forecasting with ARIMA Model:**

Conduct exploratory data analysis to understand the temporal patterns in crime rates against women using line charts, histograms, and box plots.

Plot AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) graphs was used to identify the parameters (p,d,q) for the ARIMA model.

The ARIMA model was fit to the dataset, specifying the identified parameters (p,d,q).

The future crime rates against women was predicted using the ARIMA model, considering the temporal trends and seasonality observed in the data.

**4.6. Prediction for Uttar Pradesh (UP) - Dowry Deaths:**

The dataset was filtered to include only entries related to Uttar Pradesh and focusing on the 'Dowry Deaths' column for prediction. The optimal parameters (p,d,q) are determined for the ARIMA model specific to the 'Dowry Deaths' column and it was fit in the ARIMA model to the data for Uttar Pradesh, considering the identified parameters. The dowry deaths was forecasted for the next 7 years in Uttar Pradesh using the ARIMA model.

**4.7. Evaluation and Visualization:**

The accuracy of the clustering and forecasting results was implemented using appropriate metrics and techniques and visualization of the clustering results was done using bar graphs and pie charts to illustrate the division of regions into 'Safe' and 'Unsafe' categories.

**4.8. Conclusion and Recommendations:**

The findings of the analysis was summarized, highlighting insights into crime patterns against women and future trends.

Based on the analysis results, policy interventions and preventive measures can be taken to address gender-based violence and ensuring women's safety and well-being.

**CHAPTER 5**

**RESULTS AND DISCUSSIONS**

The project successfully analysed and predicted crime rates against women in India using a combination of KMeans clustering and ARIMA modelling techniques. The results obtained provide valuable insights into the spatial and temporal dynamics of gender-based violence, aiding in evidence-based decision-making and policy formulation.

Using k-means clustering to categorize states/UTs into "safe" and "unsafe" categories provides a data-driven approach to understanding and addressing safety concerns. By identifying patterns and similarities among states/UTs, policymakers and stakeholders can better allocate resources and implement targeted interventions to improve safety and security nationwide.

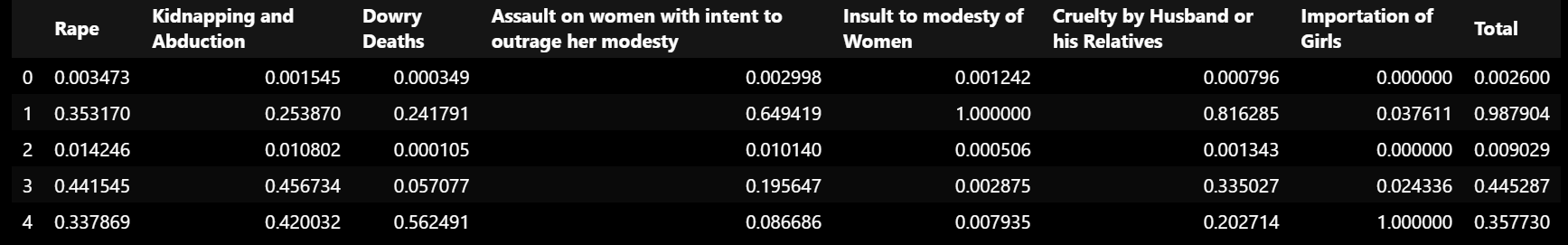
Fig.2Kmeans head after Scaling



Fig.3Classifying states into safe/unsafe

Additionally, employing the ARIMA model, specifically the SARIMAX with parameters (1, 1, 1), facilitates the prediction of "Dowry Deaths" based on historical time series data. The model's coefficients provide insights into the influence of auto-regressive and moving average terms on the dependent variable, while significance levels assess the reliability of these effects. ARIMA modelling offers a robust framework for forecasting future trends, enabling proactive measures and informed decision-making in addressing societal issues like dowry deaths.

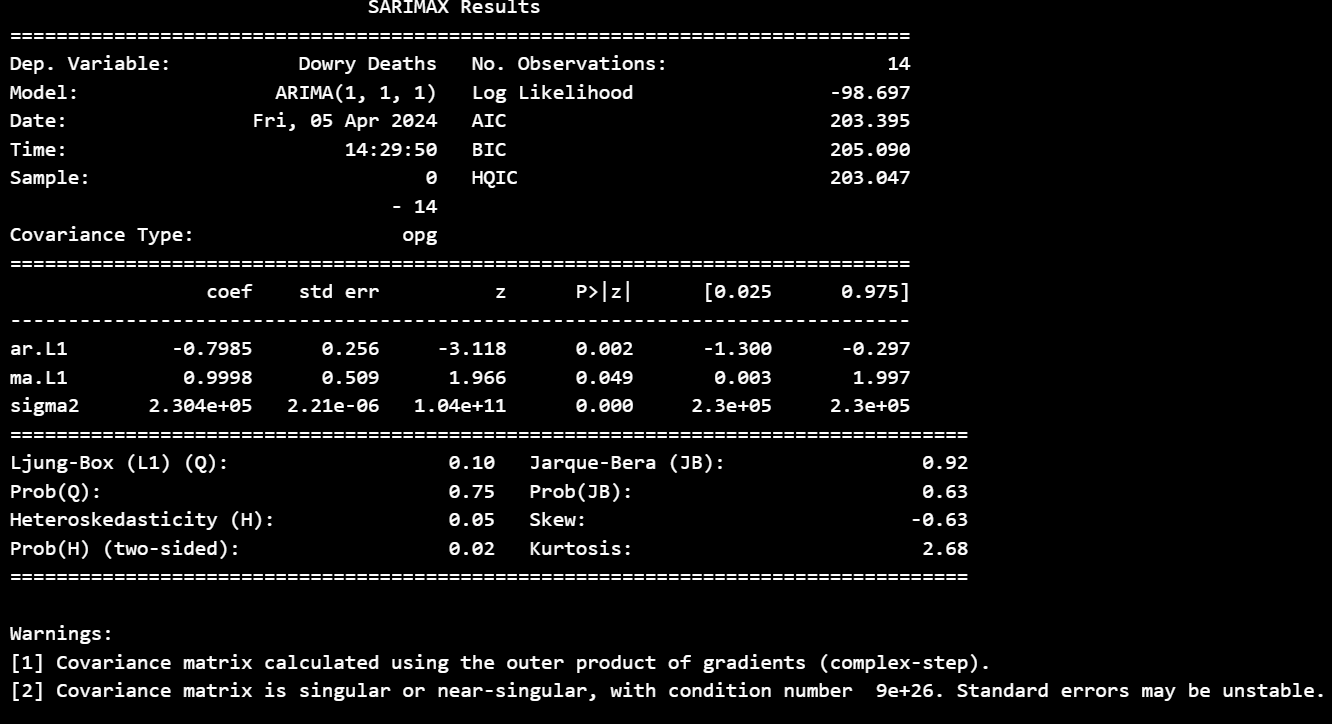


Fig.4ARIMA model

The statistical analysis and visualization techniques further enhanced the understanding of underlying patterns and trends in crime rates against women, facilitating informed decision-making and strategic interventions.

Overall, the insights derived from this project can contribute to efforts aimed at combating gender-based violence, promoting women's safety and empowerment, and fostering a safer and more inclusive society for all.

**APPENDICES**

**APPENDIX-1: CODE COMPILER**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib

import matplotlib.pyplot as plt

import plotly.express as px

%matplotlib inline

import warnings

warnings.filterwarnings('ignore')

**PRE PROCESSING:**

df=pd.read\_csv('/content/crimes\_against\_women\_2001-2014.csv')

df.head()

df.drop('Unnamed: 0',axis=1,inplace=True)

df

df.loc[df['STATE/UT'] == 'A&N Islands', 'STATE/UT'] = 'A & N ISLANDS'

df.loc[df['STATE/UT'] == 'D&N Haveli', 'STATE/UT'] = 'D & N HAVELI'

df.loc[df['STATE/UT'] == 'Delhi UT', 'STATE/UT'] = 'DELHI'

df['STATE/UT'] = pd.Series(str.upper(i) for i in df['STATE/UT'])

df['DISTRICT'] = pd.Series(str.upper(i) for i in df['DISTRICT'])

state\_all\_crimes = df.groupby('STATE/UT').sum()

state\_all\_crimes = state\_all\_crimes.drop(columns=['DISTRICT'])

state\_all\_crimes.drop('Year',axis=1,inplace=True)

numeric\_columns = state\_all\_crimes.select\_dtypes(include=['int64', 'float64'])

state\_all\_crimes['Total'] = numeric\_columns.sum(axis=1)

print(state\_all\_crimes)

all\_crimes = state\_all\_crimes

all\_crimes

**ANALYSIS:**

state\_all\_crimes.sort\_values('Total',ascending=False)

state\_all\_crimes

total\_df=state\_all\_crimes.sum(axis=0).reset\_index()

tf=pd.DataFrame(total\_df)

tf

sorted\_df = state\_all\_crimes.sort\_values('Total',ascending=False)

fig = px.bar( x=tf["index"],y=tf[0], color=tf[0],

labels={'x': "Crimes", 'y': "Count"}, title="Total Cases",

color\_continuous\_scale='burg')

fig.show()

fig = px.bar( x=state\_all\_crimes['STATE/UT'],y=state\_all\_crimes["Rape"], color=state\_all\_crimes["Rape"],

labels={'x': "States", 'y': "Count"}, title="Rape Cases",

color\_continuous\_scale='burg')

fig.show()

fig = px.bar( x=state\_all\_crimes['STATE/UT'],y=state\_all\_crimes["Kidnapping and Abduction"], color=state\_all\_crimes["Kidnapping and Abduction"],

labels={'x': "States", 'y': "Count"}, title="Kidnapping and Abduction Cases",

color\_continuous\_scale='burg')

fig.show()

fig = px.bar( x=state\_all\_crimes['STATE/UT'],y=state\_all\_crimes["Dowry Deaths"], color=state\_all\_crimes["Dowry Deaths"],

labels={'x': "States", 'y': "Count"}, title="Dowry Deaths",

color\_continuous\_scale='burg')

fig.show()

fig = px.bar( x=state\_all\_crimes['STATE/UT'],y=state\_all\_crimes["Assault on women with intent to outrage her modesty"], color=state\_all\_crimes["Assault on women with intent to outrage her modesty"],

labels={'x': "States", 'y': "Count"}, title="Assault on women with intent to outrage her modesty",

color\_continuous\_scale='burg')

fig.show()

fig = px.bar( x=state\_all\_crimes['STATE/UT'],y=state\_all\_crimes["Insult to modesty of Women"], color=state\_all\_crimes["Insult to modesty of Women"],

labels={'x': "States", 'y': "Count"}, title="Insult to modesty of Women",

color\_continuous\_scale='burg')

fig.show()

fig = px.bar( x=state\_all\_crimes['STATE/UT'],y=state\_all\_crimes["Cruelty by Husband or his Relatives"], color=state\_all\_crimes["Cruelty by Husband or his Relatives"],

labels={'x': "States", 'y': "Count"}, title="Cruelty by Husband or his Relatives",

color\_continuous\_scale='burg')

fig.show()

importation\_df = state\_all\_crimes.copy()

importation\_df.loc[importation\_df['Importation of Girls'] <= 50, 'STATE/UT'] = 'Others'

fig = px.pie(importation\_df, values='Importation of Girls', names='STATE/UT', title="Importation of Girls",

color\_discrete\_sequence=px.colors.sequential.Teal\_r)

fig.update\_traces(textposition='inside', textinfo='label+value',

marker=dict(line=dict(color='#000000', width=2)))

fig.show()

**CLUSTERING:**

all\_crimes = all\_crimes.reset\_index()

all\_crimes

m=all\_crimes['Total'].mean()

print('mean=',m)

q = np.quantile(all\_crimes['Total'],[0.25,0.75])

print(q)

l=q[0]

u=q[1]

df\_kmeans = all\_crimes.loc[:,all\_crimes.columns!="STATE/UT"]

output=[]

for i in df\_kmeans['Total']:

if i >= m:

output.append(1)#redzone

elif m > i:

output.append(0)#safe

all\_crimes['output']=output

df\_kmeans\_y=all\_crimes['output']

from sklearn.preprocessing import MinMaxScaler

cols = df\_kmeans.columns

cols

ms=MinMaxScaler()

df\_kmeans = ms.fit\_transform(df\_kmeans)

df\_kmeans = pd.DataFrame(df\_kmeans,columns=[cols])

df\_kmeans.head()

**K-MEANS MODELLING WITH 2 CLUSTERS:**

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=2, random\_state=0)

kmeans.fit(df\_kmeans)

kmeans.inertia\_

labels = kmeans.labels\_

correct\_labels = sum(df\_kmeans\_y == labels)

print('labels:',labels)

print('df\_kmeans output:',df\_kmeans\_y)

print("Result: %d out of %d samples were correctly labeled." % (correct\_labels, df\_kmeans\_y.size))

final=[]

for i in range(len(labels)):

state=all\_crimes['STATE/UT'][i]

label = labels[i]

if label == 1:

final.append([state,'unsafe'])

else:

final.append([state,'safe'])

final\_df = pd.DataFrame(final, columns=['STATES/UT', 'SAFE/UNSAFE'])

final\_df

**FORECASTING THE CRIME FOR FUTURE YEARS USING ARIMA MODEL:**

df\_UP = df[df['STATE/UT'] == 'UTTAR PRADESH']

df\_UP = df\_UP[['Year', 'Dowry Deaths']]

X = df\_UP.groupby(['Year'], sort=True).sum()

df\_UP = pd.DataFrame(X)

df\_UP=df\_UP.reset\_index()

df\_UP

print(df\_UP)

df\_UP.plot(kind = 'line',

x = 'Year',

y = 'Dowry Deaths')

from statsmodels.graphics.tsaplots import plot\_acf, plot\_pacf

from statsmodels.tsa.arima.model import ARIMA

from statsmodels.tsa.stattools import adfuller

def check\_stationarity(series):

result = adfuller(series.values)

print('ADF Statistic: %f' % result[0])

print('p-value: %f' % result[1])

print('Critical Values:')

for key, value in result[4].items():

print('\t%s: %.3f' % (key, value))

if (result[1] <= 0.05) & (result[4]['5%'] > result[0]):

print("\u001b[32mStationary\u001b[0m")

else:

print("\x1b[31mNon-stationary\x1b[0m")

check\_stationarity(df\_UP['Dowry Deaths'])

plot\_acf(df\_UP['Dowry Deaths'],lags=10)

plot\_pacf(df\_UP['Dowry Deaths'],lags=6)

model = ARIMA(df\_UP['Dowry Deaths'], order=(1,1,1))

model\_fit = model.fit()

print(model\_fit.summary())

fc = model\_fit.forecast(7, alpha=0.05)

for i in fc:

print(round(i,0))

**APPENDIX-2: SCREENSHOTS**

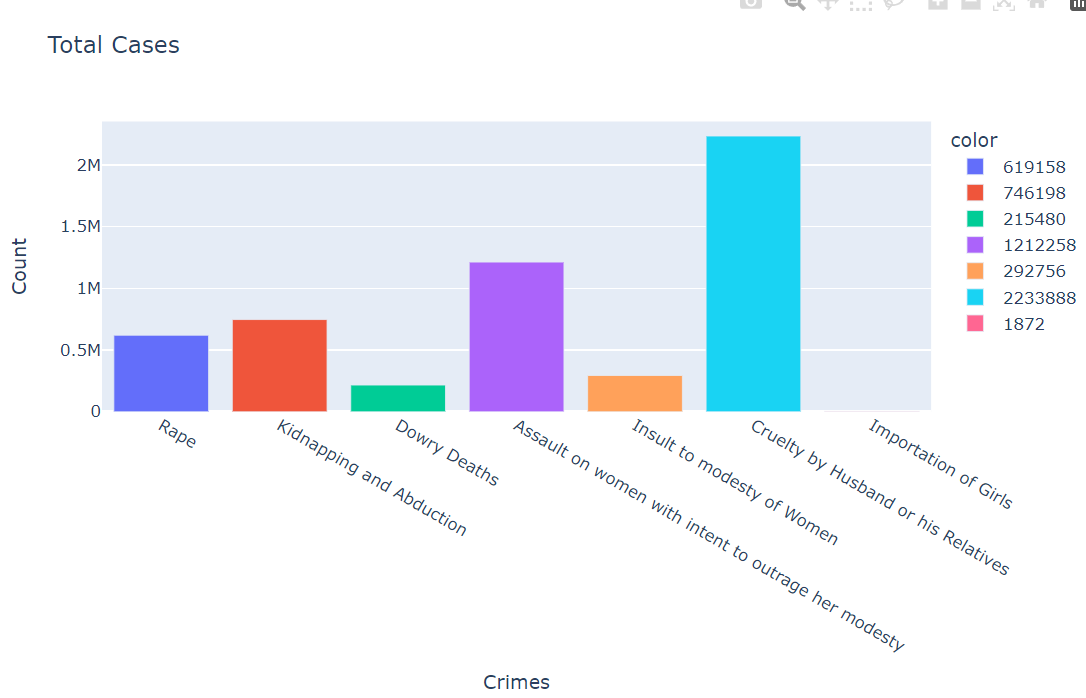
****

fig.5Total cases vs crimes

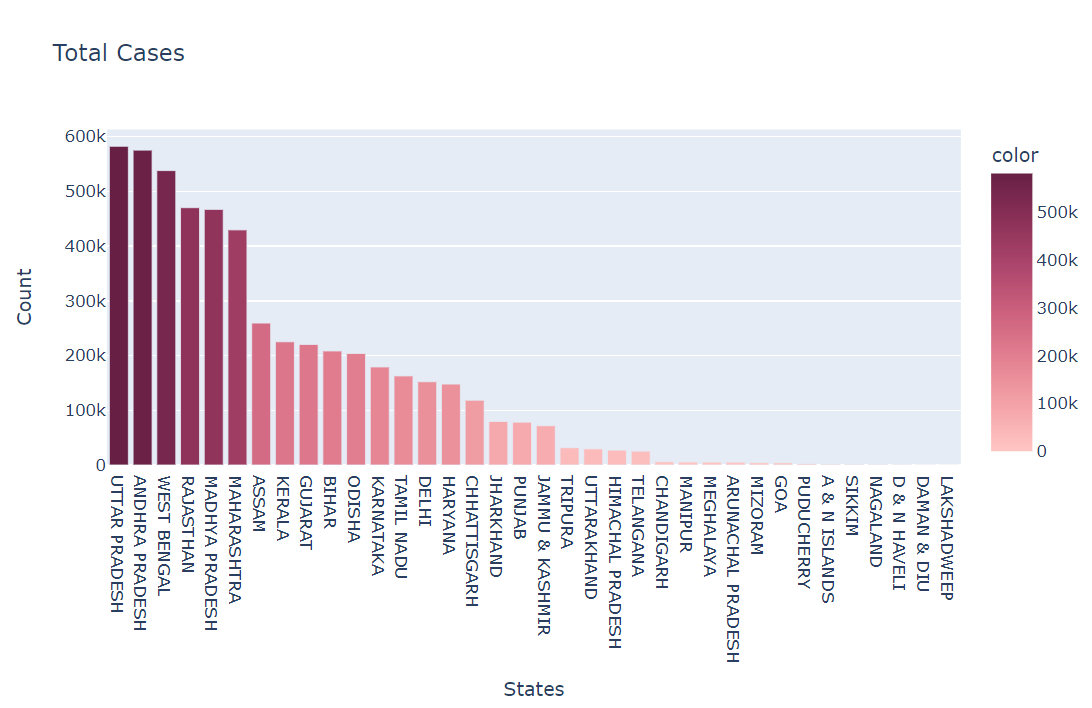
****

fig.6States vs Total cases

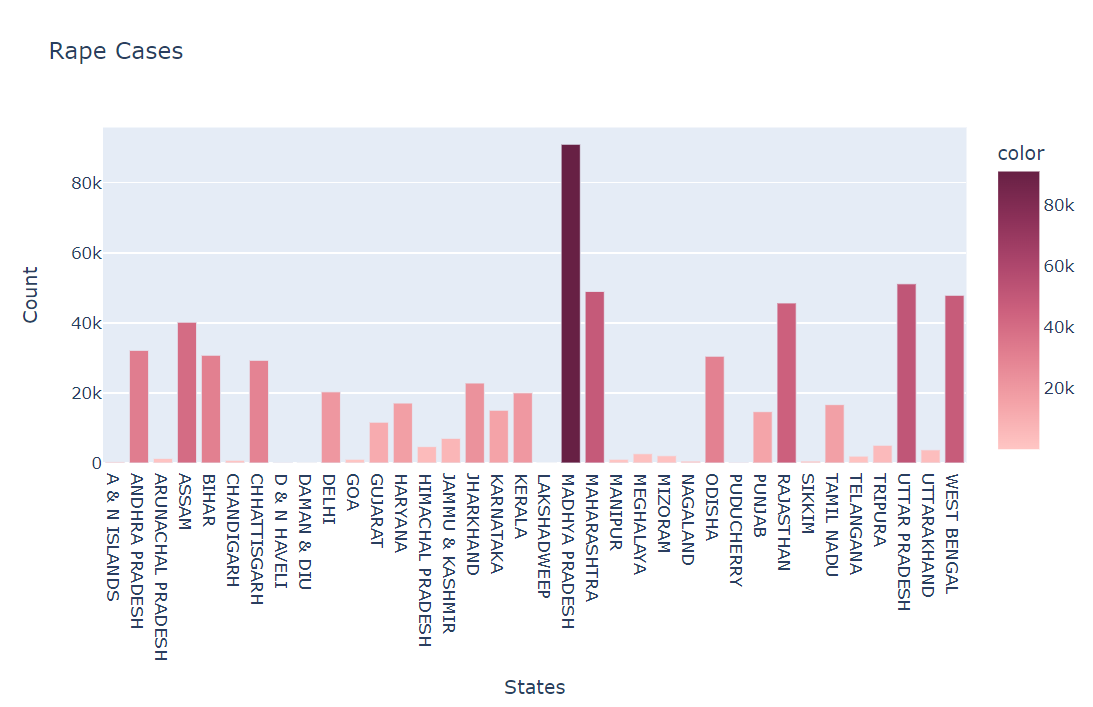
****

fig.7States vs Rape cases

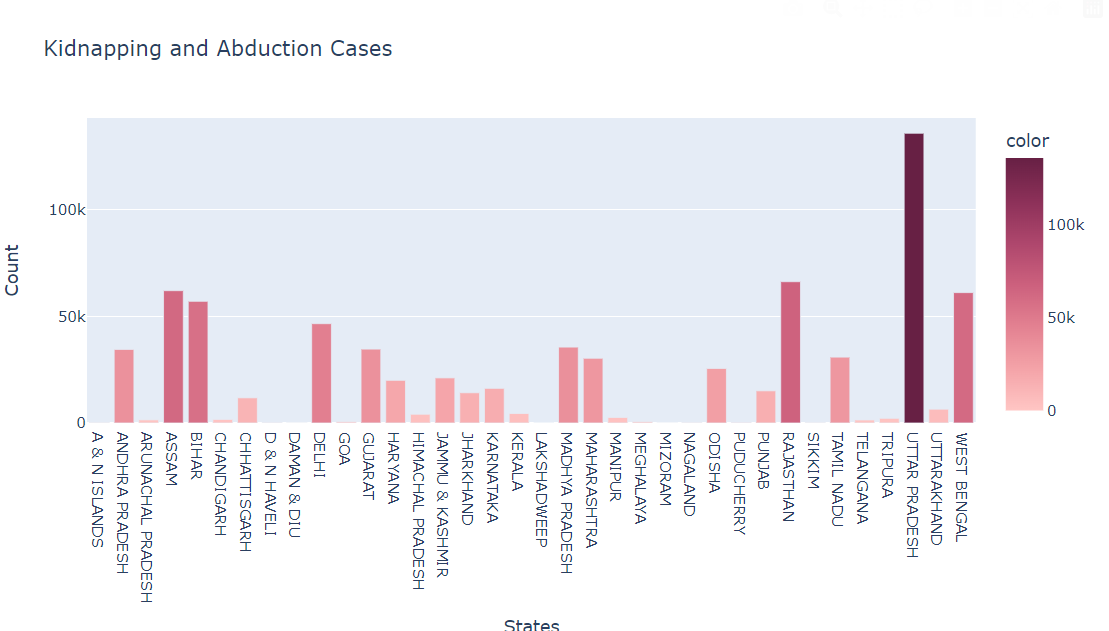
****

fig.8States vs Kidnapping and Abduction

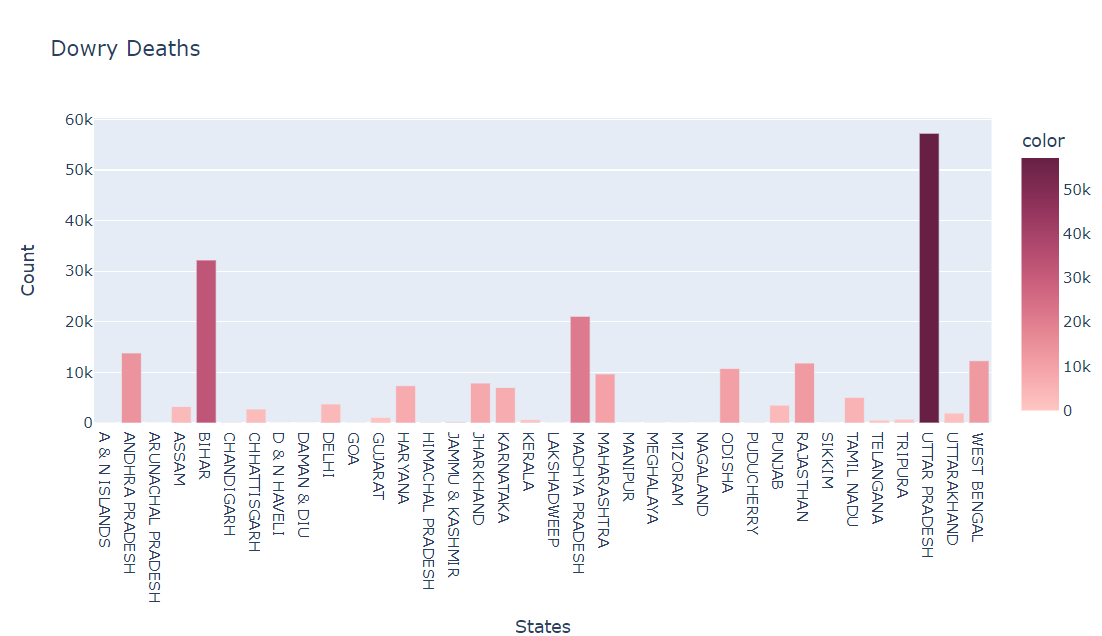
****

fig.9States vs Dowry Deaths

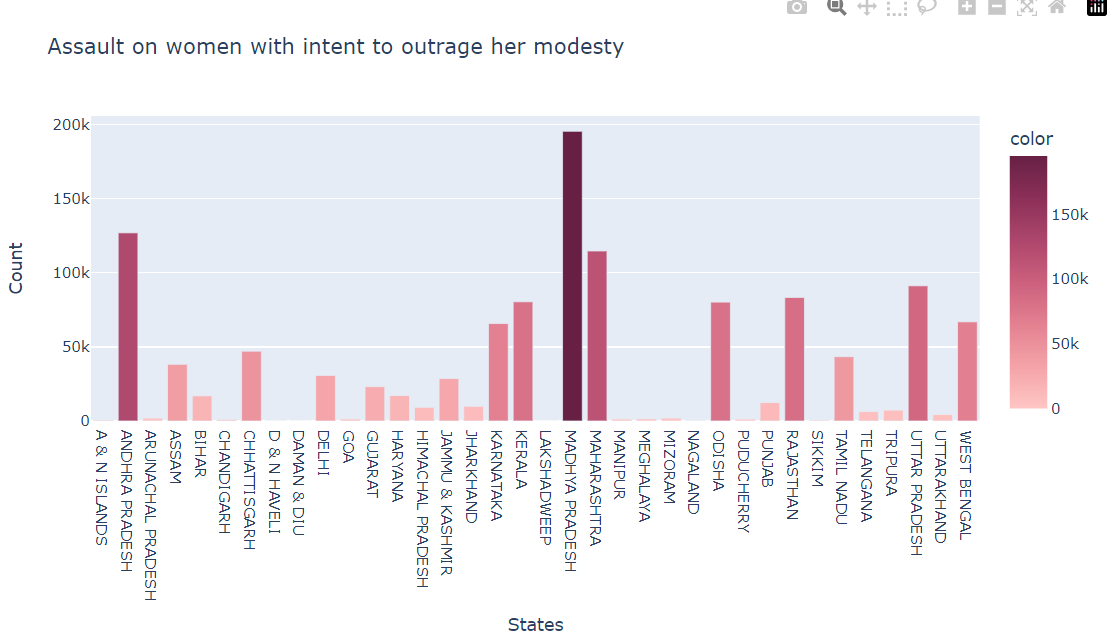
****

fig.10 States vs Assault on women with intent to outrage her modesty

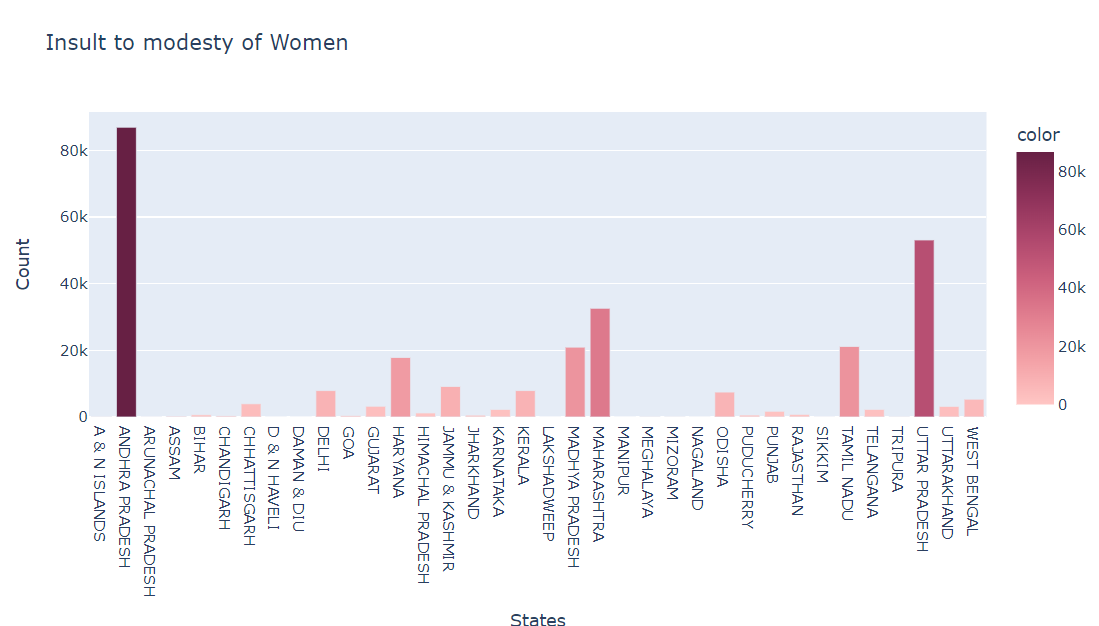
****

fig.11States vs Insult to modesty of Women

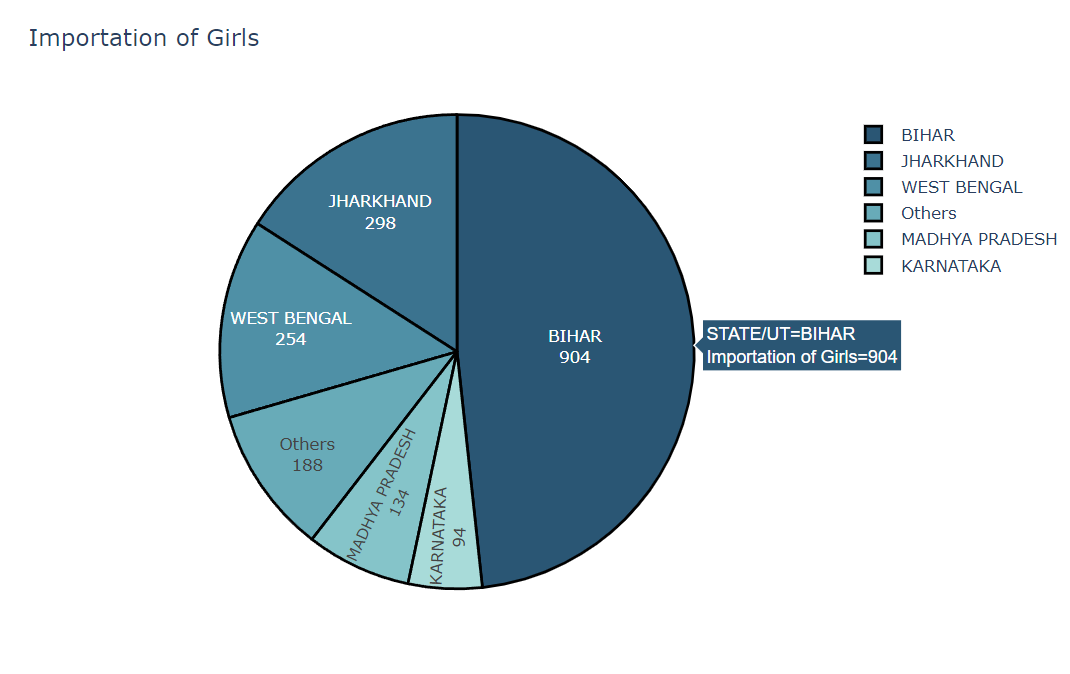
****

fig.12States vs Importation of girls

****

fig.13All crimes dataset

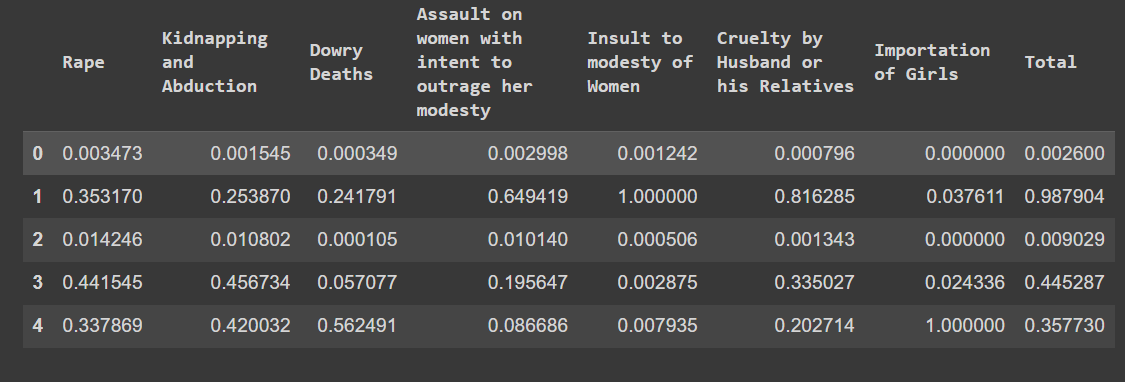
****

fig.14Kmeans head after Scaling

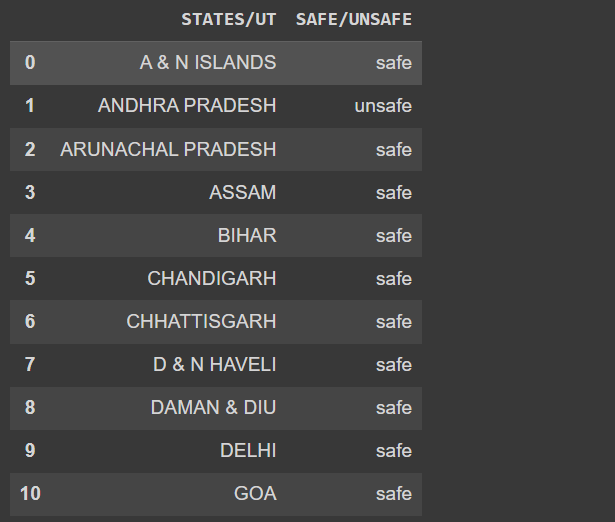
****

fig.15Classifying states into safe/unsafe

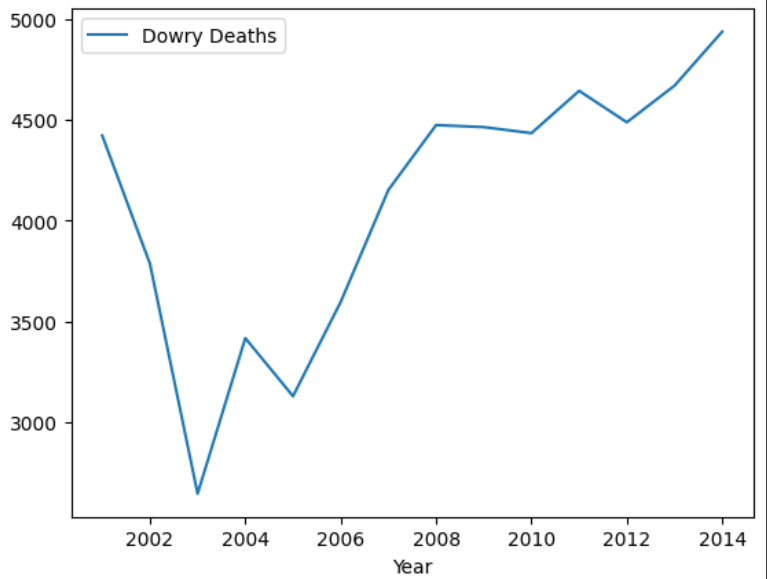
****

fig.16Year vs Dowry Deaths

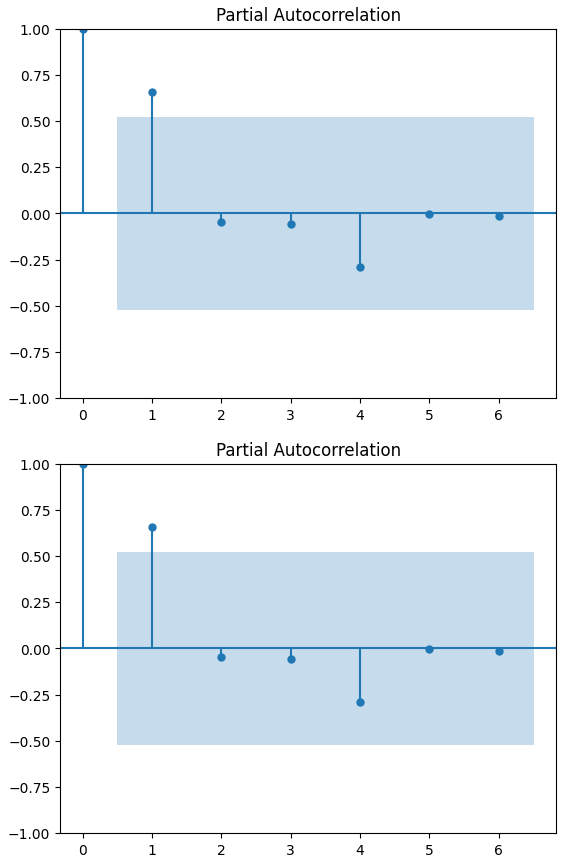
****

fig.17Partial Autocorrelation

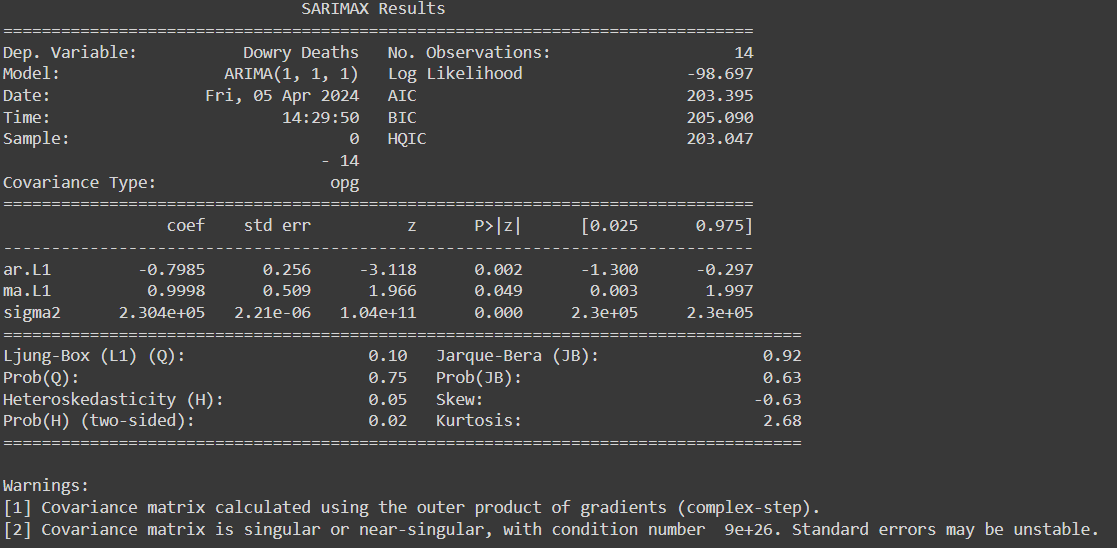
****

fig.18acf plot

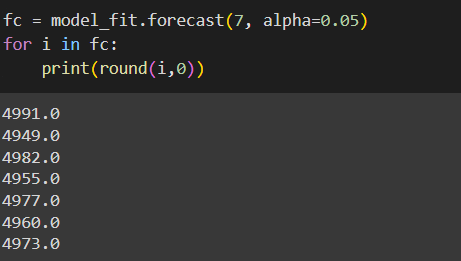
****

fig.19pacf plot

**REFERENCES**

**Journal References:**

[1] S. Lavanyaa, D. Akila Crime against Women (CAW) Analysis and Prediction in Tamilnadu Police Using Data Mining Techniques International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5C, February 2019

[2] Tahani Almanie, Rsha Mirza and Elizabeth Lor Crime Prediction Based on Criminal Types And Using Spatial and Temporal Criminal Hotspots International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015

[3] Mehmet Sait Vural, Mustafa Gok Criminal prediction using Naive Bayes theory September 2017

[4] Shiju Sathyadevan, Devan M. S, Surya S Gangadharan Crime Analysis and Prediction Using Data Mining (ICNSC), 2014

[5] Yamuna.S, N.Sudha Bhauvaneswari D, Data mining Techniques to Analyze and Predict Crimes, International Journal of Engineering And Science(IJES) Vol-1,Issue-2,PP 243-247

[6] Sylvia Walby, Improving the statistics on violence against women, Statistical Journal of the united nation ECE 22(2005)193-216.

[7] S.Bewley, J.Friend and G.Mezey, Eds, Violence against Women, London: Royal College of Obestricians and Gynaecologists, 1997.

[8] Malathi.A and Dr.S.Santhosh Baboo. Article: an enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications,21(1):1-6, May2011.Published by foundation of Computer Science.

[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] Web Crime Data using Data Mining, International Journal of Engineering And Innovative Technology(Ijeit)2(3)

[11] Devendra Kumar Tayal et al., Crime detection and criminal identification in India using data mining techniques,AI & Soc(2015) 30, pp.117-127.

[12] Roslin V.Hausk and Hsinchun Chen., Coplink: A Case of Intelligent Analysis and Knowledge Management, Proceedings of International Conference on Information Systems, 1999, pp.15-28.

[13] Prajakta Yerpude , Vaishnavi Gudur Predictive Modelling of Crime Dataset using Data Mining (IJDKP) Vol.7, No.4, July 2017.

[14] Deepika K.K, Smitha Vinod Crime analysis in India using data mining techniques 2018.

[15] Rasoul kiani, Siamak mahdavi, Amin Keshavarzi, Analysis and Prediction of Crimes by clustering and Classification.(IJARAI) International Journal of Advanced Research in Artificial Intelligence,Vol-4,No:8,2015

[16] Amarnathan, L.C.(2003) Technological Advancement: Implications for the Crime, the Indian Police Journal, AprilJune.

[17] Mayank Motwani, Pratha Purwar, Rachit Mathur, Aatif J Amshed An Efficient Approach towards crime gainst women using time series algorithm International Journal of Computer Applications (0975 – 8887) Volume 179 – No.34, April 2018.

[18] Dr.D.Akila, Ms.Vidya and Mr.S.Rajesh , "Optimization Based Information Retrieval with the Enhancement of Annotator in WordNet Application", Journal of Advanced Research in Dynamical & Control systems Volume 10 , Issue 2, pp. 318- 323,2018.

[19] D.Akila,S.Sathya, G.Suseendran, “Survey on Query Expansion Techniques in Word Net Application”, Journal of Advanced Research in Dynamical and Control Systems, Vol.10(4), pp.119- 124, 2018.

**Web References:**

1. <https://github.com/ZJUDataIntelligence/HDM-GNN>
2. <https://github.com/vikram-bhati/PAASBAAN-crime-prediction>
3. <https://github.com/sharmaroshan/Sanfrancisco-Crime-Dataset>

**WORKLOG**

|  |  |  |
| --- | --- | --- |
| **Day** | **Date** | **Task Done** |
| Day 1 | 16/02/2024 | Finding a suitable dataset |
| Day 2 | 21/02/2024 | Started data preprocessing using pandas library.  Identified and handled missing values, outliers, and inconsistencies in the dataset. |
| Day 3 | 23/02/2024 | Continued data preprocessing by removing undefined and unnamed features.  Standardized state names to ensure consistency in the dataset. |
| Day 4 | 28/02/2024 | Completed data preprocessing tasks, including data cleaning and standardization.  Conducted quality assurance checks to ensure the dataset's integrity and reliability. |
| Day 5 | 06/03/2024 | Applied KMeans clustering algorithm to segment the dataset into two clusters based on crime patterns.  Determine the optimal number of clusters using the Elbow Method. |
| Day 6 | 08/03/2024 | Calculated the mean of crime rates against women to establish a threshold for categorizing regions as 'Unsafe' or 'Safe'.  Assigned labels to each cluster based on the calculated threshold. |
| Day 7 | 13/03/2024 | Explored ARIMA modeling for forecasting future crime rates against women.  Identified parameters (p,d,q) for the ARIMA model using ACF and PACF analysis. |
| Day 8 | 15/03/2024 | Split the dataset into training and testing sets for model validation.  Fitted the ARIMA model to the training data, specifying the identified parameters. |
| Day 9 | 20/03/2024 | Validated the ARIMA model's performance on the testing data using evaluation metrics such as MAE and RMSE.  Assessed the stability and consistency of the clusters generated by KMeans clustering. |
| Day 10 | 22/03/2024 | Visualized the clustering results and ARIMA forecasting results using matplotlib library.  Presented initial findings and insights to project stakeholders for feedback. |
| Day 11 | 27/03/2024 | Fine-tuned the ARIMA model parameters to improve forecasting accuracy.  Generated forecasts for future crime rates against women using the trained ARIMA model. |
| Day 12 | 03/04/2024 | Conducted a detailed analysis of clustering results and ARIMA forecasts.  Prepared the final report summarizing project findings, methodology, and recommendations. |
| Day 13 | 05/04/2024 | Reviewed and refined the report, ensuring clarity, coherence, and accuracy of information.  Created visual aids such as charts, graphs, and tables to enhance presentation quality. |
| Day 14 | 10/04/2024 | Presented the final project report to stakeholders and key decision-makers. |
| Day 15 | 12/04/2024 | Discussed insights, recommendations, and potential next steps based on the analysis results. |