VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

Final Review Report

Programme: B.Tech

Course: Data Visualization

Slot: D2

Faculty: Dr. Joshan.Athanesious J

Component: J

Title: Crime Analysis in India

Team Member(s):

Dolly Agarwala (20BCE1863)

Dhruvi Ochani (20BCE1882)

Aryan Bhardwaj (20BCE1908)

ABSTRACT

Crime analysis in India involves the collection and analysis of data related to criminal activities to understand patterns and trends of crime. The process of crime analysis involves the use of advanced technology and tools to identify and analyze the factors contributing to crime, including socio-economic factors, environmental factors, and demographic factors.

In recent years, crime analysis has become an essential tool for law enforcement agencies in India to improve public safety and reduce crime rates. The use of advanced technologies such as geographic information systems (GIS), predictive analytics, and data mining has helped law enforcement agencies in identifying crime hotspots, predicting crime patterns, and allocating resources effectively.

However, crime analysis in India faces various challenges, including the lack of standardized data collection and management, inadequate funding, and insufficient training of law enforcement officials. Additionally, the high rates of underreporting of crimes and the lack of trust in law enforcement agencies in some regions of the country pose significant obstacles to effective crime analysis.

This project will indicate a very clear depiction of various different crimes in India happening in all sectors of life and how they have increased over the last few years in all Indian states and Union territories. Tableau and python have been used for all the visualizations and to build a predictive model so as to identify all the 'most' important attributes contributing to the total number of crimes.

Keywords: crime, analysis, tableau, visualizations, python, GIS, management, attributes, criminal, activities, law, agencies

1. INTRODUCTION

Tableau is a powerful data visualization tool that can help crime analysts explore and present crime data in a way that is easily understandable and actionable. In this project, we will use Tableau to analyze crime data and create visualizations that can help law enforcement agencies identify areas of high crime and potential crime hotspots.

Initially, we have gathered crime data from various sources such as police reports, crime databases and other related websites. Next, we have used Tableau to create a range of visualizations that can help us understand the patterns and trends in the crime data. For example, we have created bar graphs to show the concentration of crimes in different areas, line graphs to show how crime rates have changed over time, and all other relevant visualization graphs to identify correlation and similarities between different types of crimes.

We have also used Tableau to create interactive dashboards that allow us to explore the data in more detail. For example, we have created a dashboard that allows us to filter the data by location,

crime type, and time period, or a dashboard that allows users to drill down into specific crime categories or neighborhoods. In addition to creating visualizations and dashboards, we can also use Tableau to conduct advanced statistical analyses on the crime data. For example, we may use regression analysis to identify the factors that are most strongly associated with certain types of crimes, or clustering analysis to identify groups of crimes that share similar characteristics.

After creating an interactive dashboard, we then moved to developing a predictive model using 'Multilinear Regression' by first identifying all the necessary attributes via feature scaling and then finding out how much they contribute to 'total crimes' which is our target variable. This model will help us to identify what type of crime is responsible and how much it impacts the overall crime ratio in India. This has been done with python.

Overall, this project will provide an introduction to crime analysis using Tableau, and will demonstrate how data visualization tools can be used to help law enforcement agencies and public safety officials identify and prevent crimes in their communities along with identifying the most frequently occurring crimes which have a major contribution to the total number of crimes in India.

2. LITERATURE REVIEW

1. **"Crime Mapping and Analysis for Effective Policing in India"** by K.P. Modi and R.K. Singh (2015): This paper highlights the importance of crime mapping and analysis for effective policing in India. The authors argue that crime analysis can help law enforcement agencies identify crime hotspots and develop targeted interventions to reduce crime rates.

2. **"Crime Analysis and Mapping using GIS: A Case Study of Delhi"** by S. Raj and M. Kansal (2014): This paper presents a case study of crime analysis using GIS in Delhi. The authors use spatial analysis techniques to identify crime patterns and hotspots in the city, and suggest that this information can be used to develop targeted interventions to prevent crime.

3. **"Data Mining for Crime Analysis in Indian Cities"** by V. Vijayarajan and M. Baskaran (2016): This paper explores the use of data mining techniques for crime analysis in Indian cities. The authors use clustering and classification algorithms to identify crime patterns and predict future crime rates, and suggest that these techniques can help law enforcement agencies develop effective crime prevention strategies.

4. **"Spatial Analysis of Crime in Mumbai using GIS"** by S. Varghese and S. Sreekumar (2018): This paper uses GIS to analyze crime patterns in Mumbai. The authors use kernel density estimation to identify crime hotspots, and suggest that this information can be used to allocate police resources more effectively.

5. **"Crime Pattern Analysis using Data Mining Techniques: A Case Study of Bangalore"** by R. Poojary and R. Shetty (2017): This paper presents a case study of crime pattern analysis using data mining techniques in Bangalore. The authors use association rule mining and clustering algorithms to identify crime patterns and suggest that this information can be used to develop effective crime prevention strategies.

6. **"Predictive Policing in India: A Study of its Prospects and Challenges"** by Arindam Roy and Biswajit Mohanty. This paper explores the use of predictive policing in India and discusses its potential to enhance crime prevention and detection. The authors suggest that the successful implementation of predictive policing requires the development of accurate and reliable predictive models.

7. **"Community Policing and Crime Prevention: A Study of Delhi Police"** by Ravi Kumar and Rajesh Kumar. This study evaluates the effectiveness of community policing in preventing and reducing crime in Delhi. The researchers found that community policing has led to improved police-community relations and a reduction in crime in some areas.

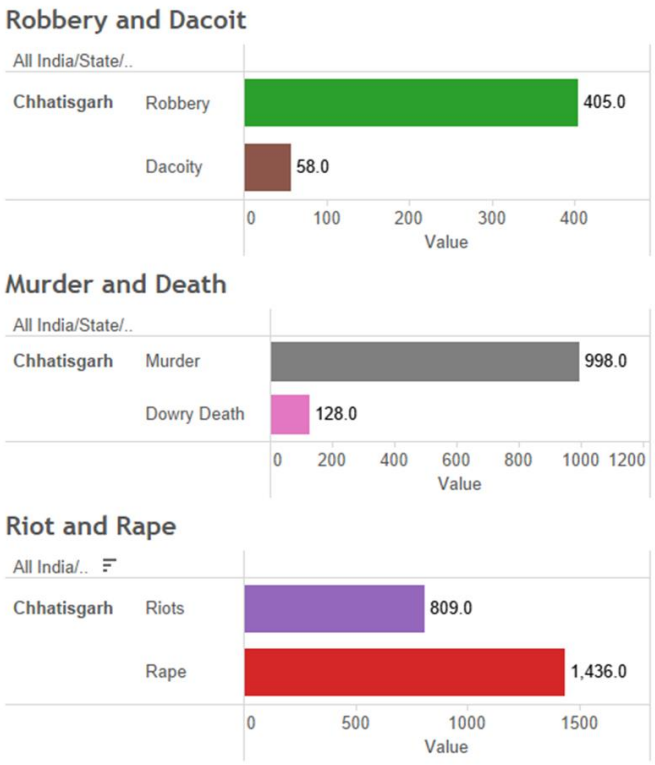
8. **"Crime Analysis and Mapping of Delhi, India"** by Shalini Aggarwal and Ankit Kumar. This study uses crime mapping and spatial analysis techniques to identify hotspots of different types of crime in Delhi. The researchers found that crime is concentrated in certain areas of the city, and that this concentration varies by type of crime.

9. **"Cybercrime in India: A Study of the Types, Victims, and Offenders"** by N. Vignesh and M. Asaithambi. This paper examines the types of cybercrime prevalent in India, as well as the characteristics of the victims and offenders. The authors suggest that the effective prevention and detection of cybercrime requires a multi-disciplinary approach that involves law enforcement agencies, policymakers, and the private sector.

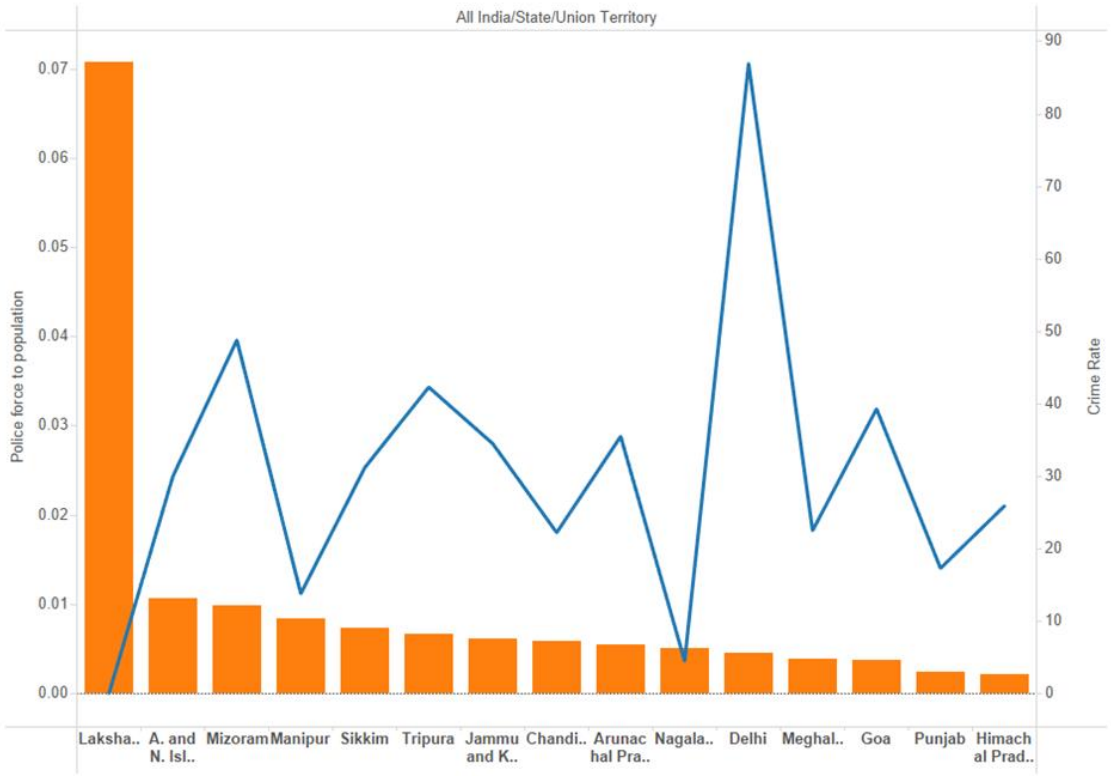
10. **"Intelligence-Led Policing in India: An Analysis of its Application and Effectiveness"** by Ashish Kumar Singh and Arun Kumar Singh (2021): This paper evaluates the effectiveness of intelligence-led policing in India, focusing on its application in several states. The authors argue that intelligence-led policing has the potential to enhance crime prevention and detection efforts, but it also requires significant organizational and technological investments.

3. MATERIALS & METHODS

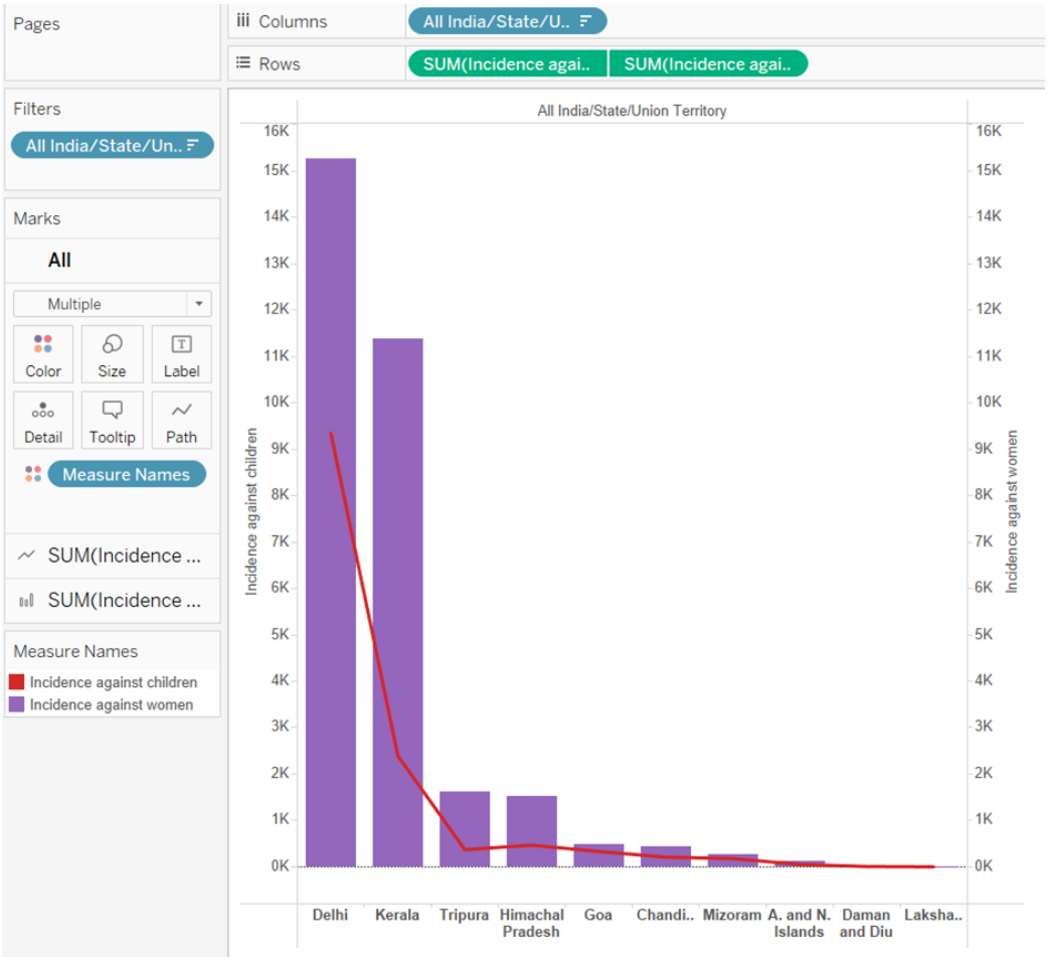
3.1 INFORMATION ABOUT MODELS



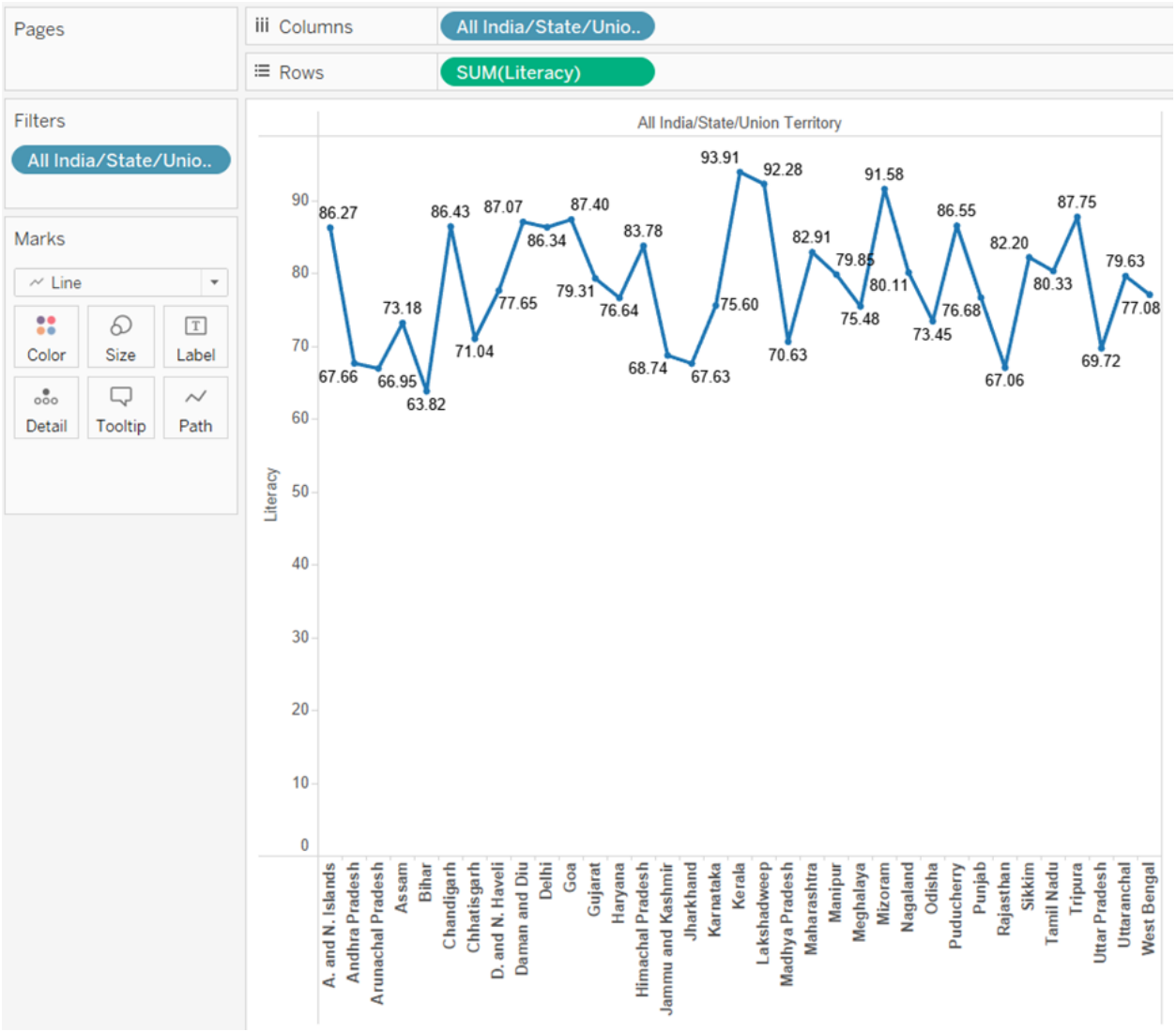
The above graph shows the different crime rates for the state of Chattisgarh. Similarly, graphs can be made for all the states of India.



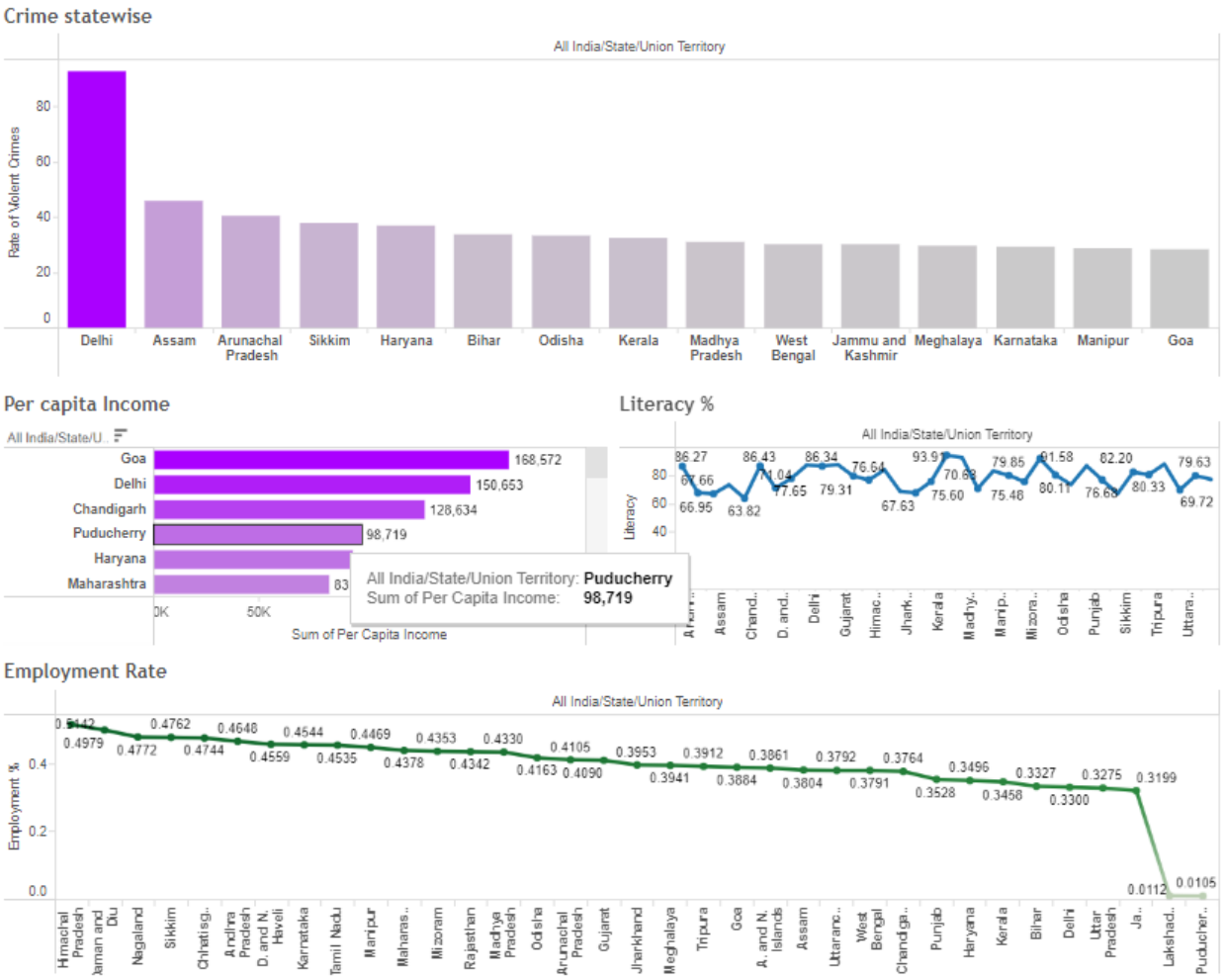
The above graph shows the police force to population ratio and crime rate for all the states.



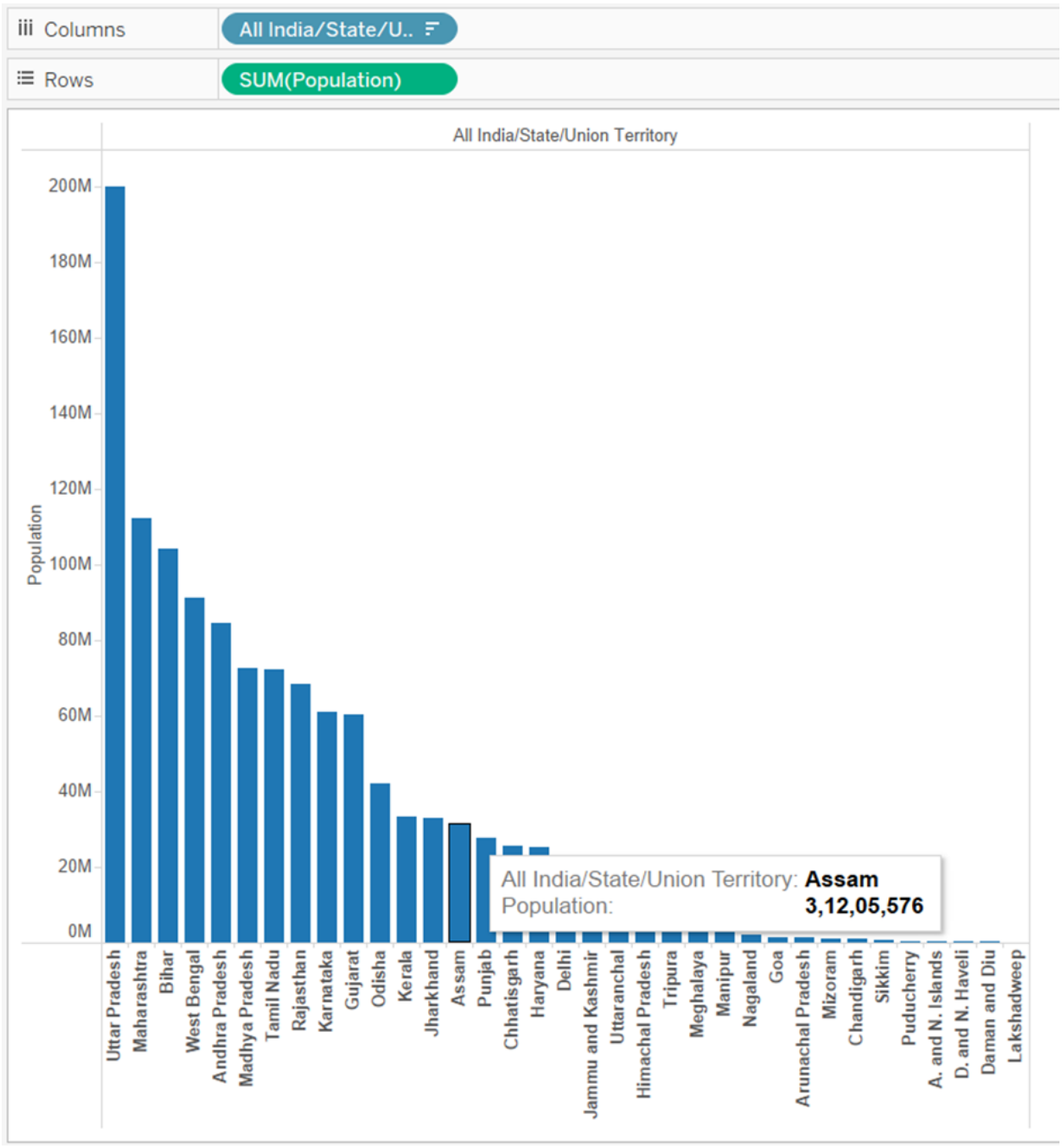
The above graph shows the number of incidents against children and women for all the states.



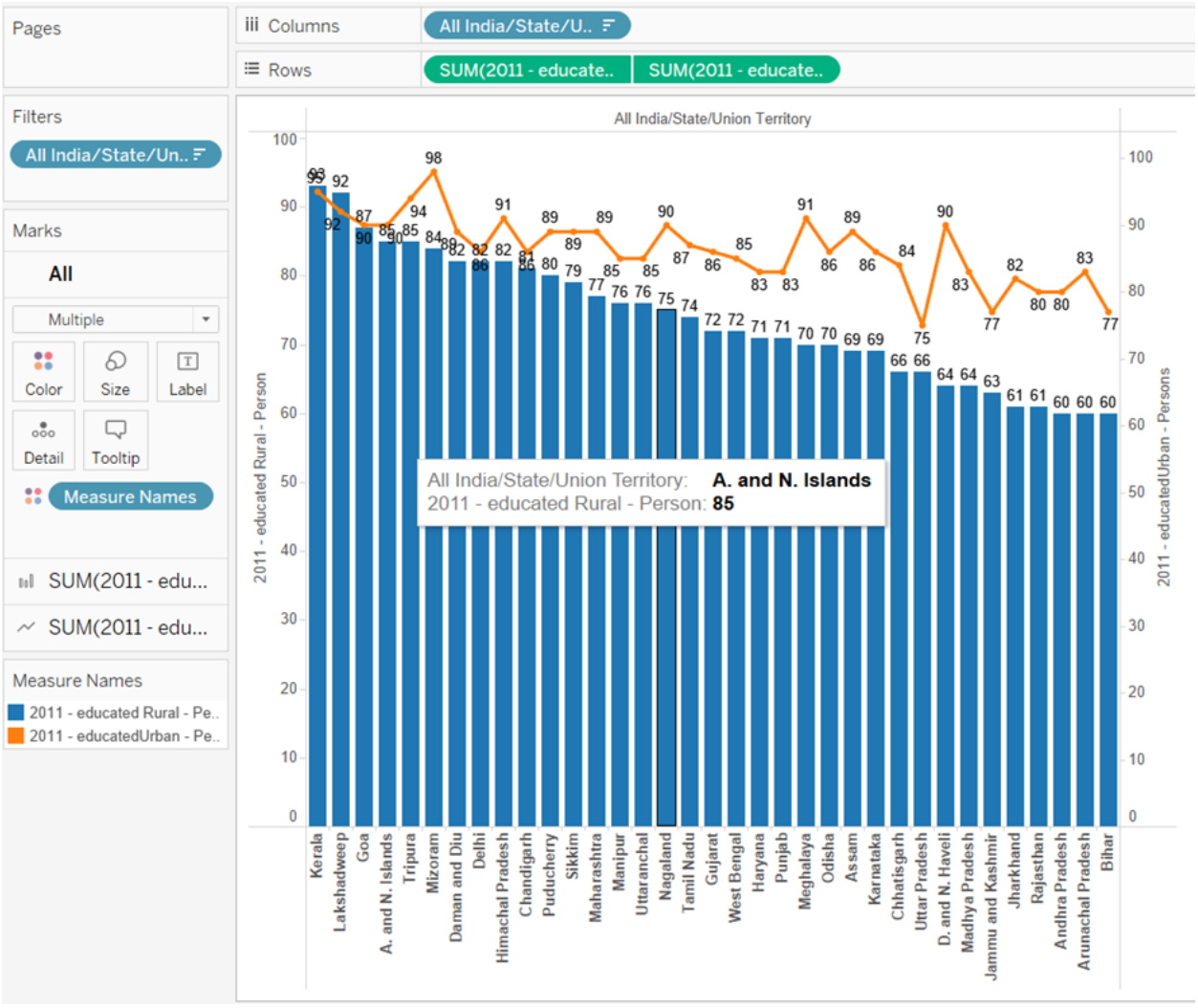
The above graph shows the literacy rate for all the states.



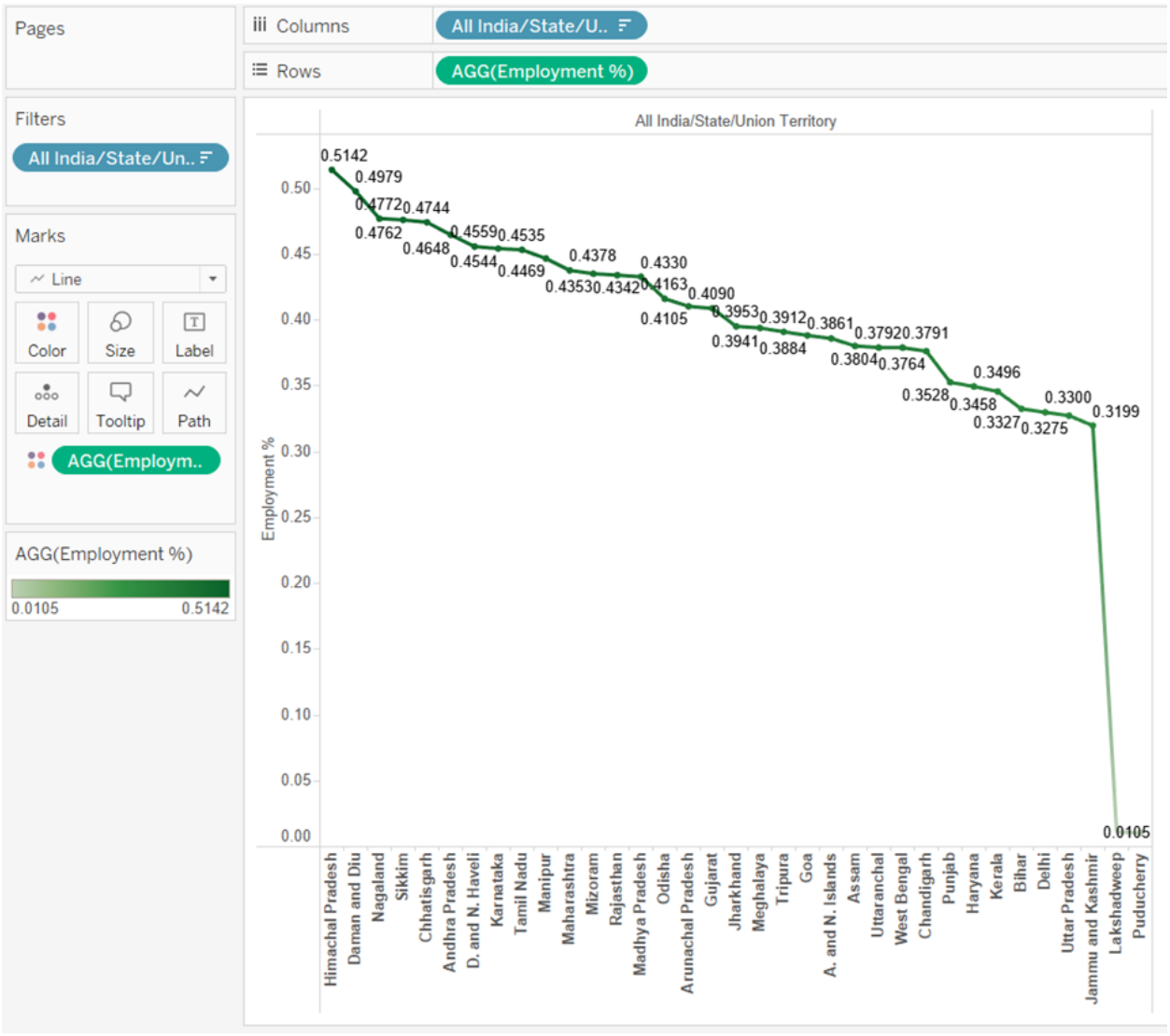
The above is a dashboard showing 4 different visualizations: Crime statewise, Per capita Income, Literacy rate and Employement rate. We can hover over any state to get a detailed numeric value.



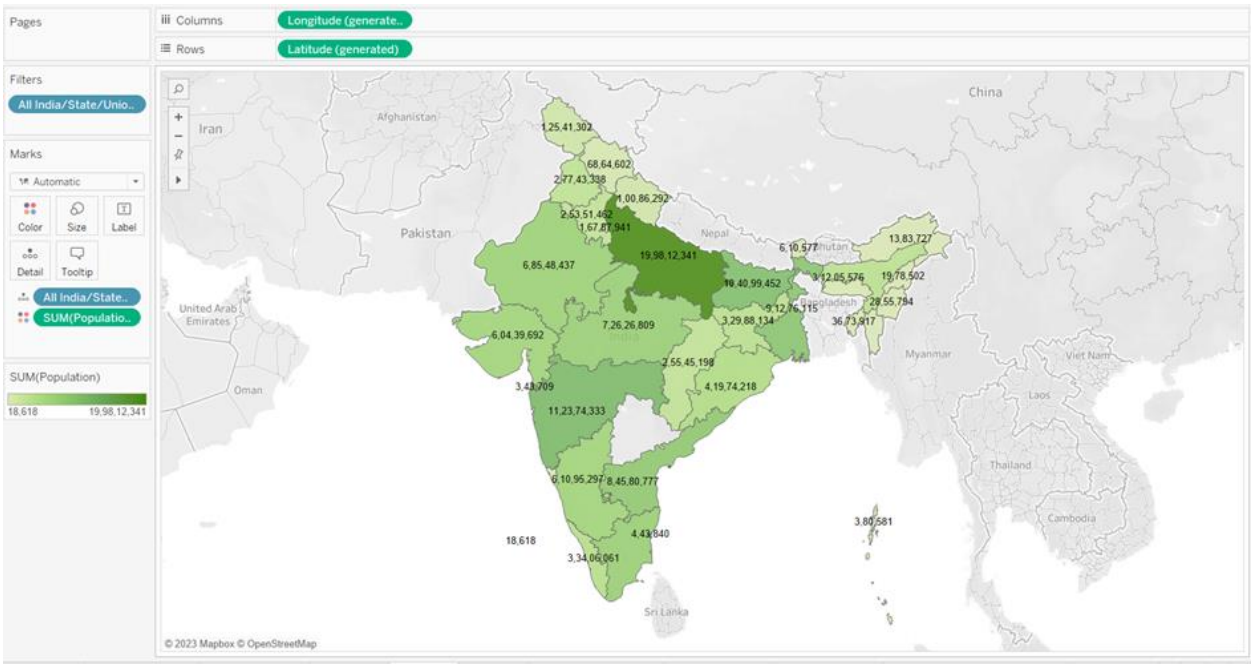
The above graph shows the population for all the states. We can hover over any state bar to get a detailed numerical value. This can be done for all graphs in the project.



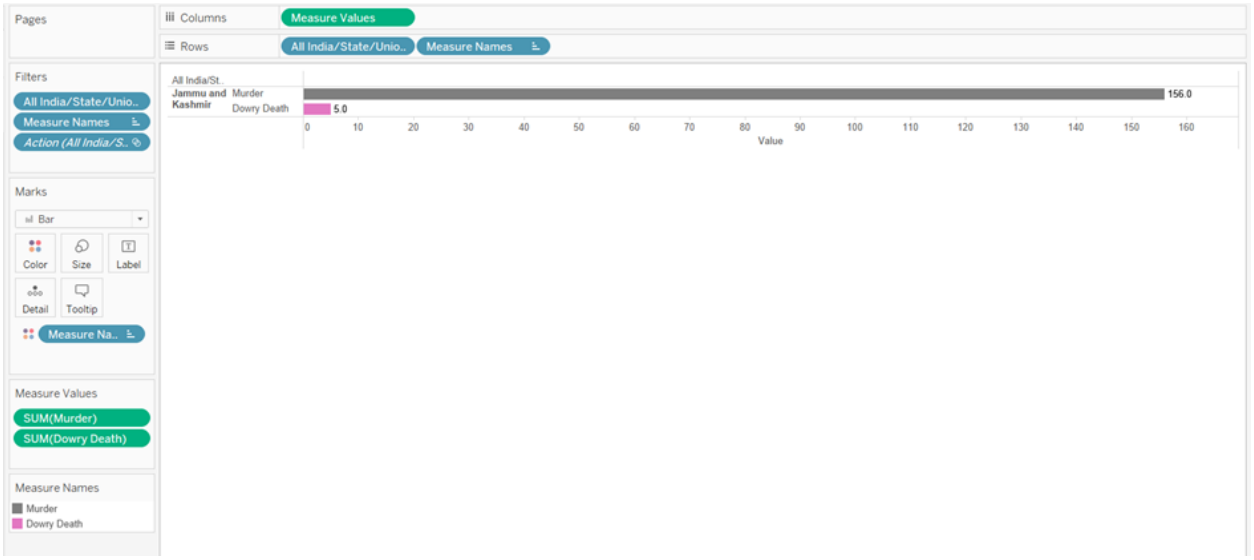
The above graph shows the educated rural and urban population for all the states. We can hover over any state bar to get a detailed numerical value. This can be done for all graphs in the project.

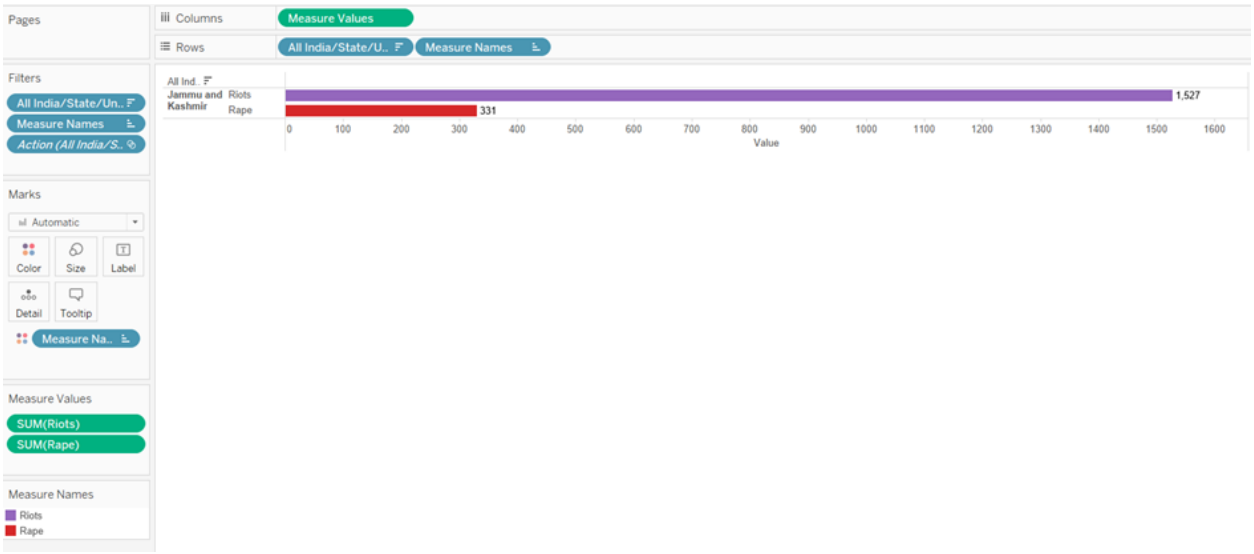


The above graph shows the employment rate for all the states.



The above graph also shows the population for all the states.





The above two graphs show the horizontal bar graphs for different crimes for the state of Jammu and Kashmir. We can select the number of different crimes we want in each graph (here, 2 different crimes are shown in each graph). Similarly, we can find for all the different crimes and for all the other states.

ALGORITHM/PSEUDOCODE FOR MULTILINEAR REGRESSION

```
// X is a matrix of input features (m samples x n features)
// y is a vector of target values (m samples)

function multilinear_regression(X, y):
    // Add a bias column to X
    X = add_bias(X)

    // Initialize w to all zeros
    w = zeros(n + 1)

    // Define the learning rate and number of iterations
    alpha = 0.01
    num_iterations = 1000

    // Perform gradient descent
    for i in range(num_iterations):
        // Compute the predicted values
        y_pred = dot(X, w)

        // Compute the error
        error = y_pred - y
```

```
    // Compute the gradient of the cost function
    gradient = dot(X.T, error)

    // Update w using the gradient descent rule
    w = w - alpha * gradient

// Return the final model parameters
return w
```

3.2 DATASET

The dataset we have used in our project has been obtained from **https://data.gov.in/**. It is the official open data portal of the Government of India. The portal provides access to a wide range of datasets, documents, and applications released by various government departments and agencies.

In addition to providing datasets, data.gov.in also offers a range of tools and applications to help users visualize and analyze data. These include data visualization tools, data analysis software, and APIs that allow users to access data programmatically.

The data available on data.gov.in is provided under the Open Government Data (OGD) Platform India initiative, which aims to promote transparency, accountability, and innovation in government by making data more accessible to citizens, researchers, and businesses.

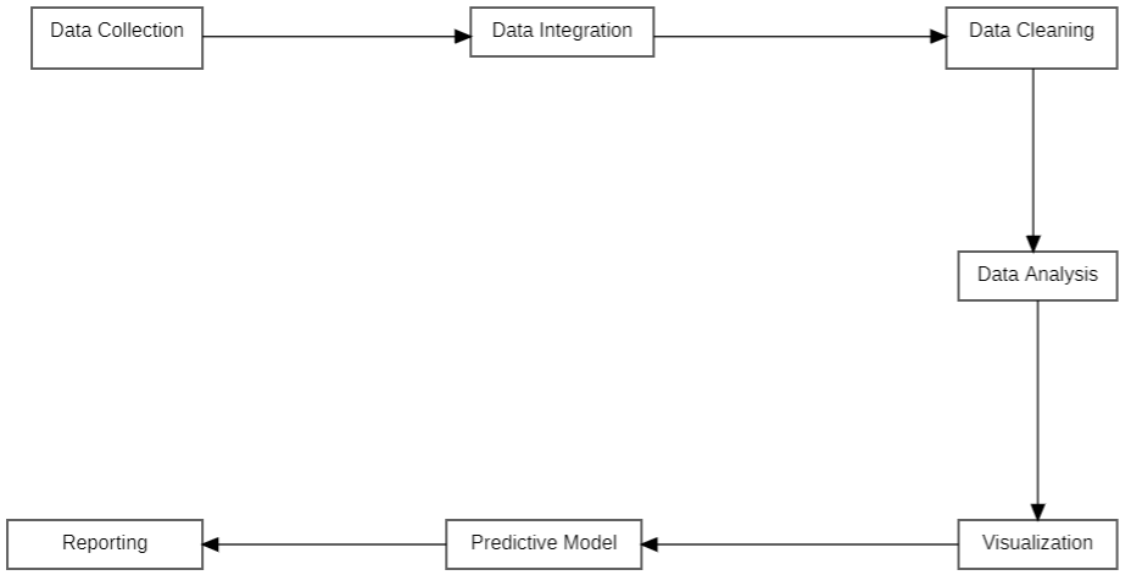
The dataset is in the form of a csv file which was imported in Tableau for making the desired dashboard. It contains 35 rows and 42 columns. Each row depicts a state/union territory whereas each column depicts a certain different type of crime.

Here is a peak to our dataset -

	All India/State/Union Territory	2011 - educated Rural - Person	2011 - educatedUrban - Persons	Actual Strength of Police force - Total	Arson	Assault On Women	Attempt to commit Culpable Homicide	Attempt To Commit Murder	Attempt To Commit Rape	Crime Rate	...	Rape	Rape 1	Rate of Total Cognizable Crimes	Rate of Total Cognizable Crimes 1	Rate of Violent Crimes	Riots	Robbery	Total Number of IPC cases for Investigation including pending Cases from previous year	Total Sexual Offences Under IPC
0	A. and N. Islands	85	90	4043.0	5	41	2	11	0	30.00	...	32	32	36.8	44.2	18.4	10	2	1466	78
1	Andhra Pradesh	60	80	89404.0	450	4547	1	1540	165	32.80	...	961	961	13.1	65.1	14.0	1013	433	241502	8322
2	Arunachal Pradesh	60	83	7455.0	39	121	0	48	10	35.50	...	83	83	28.6	57.4	40.6	33	61	3808	217
3	Assam	69	89	54069.0	743	3099	14	1142	227	34.20	...	1980	1980	11.6	123.4	45.9	2640	1038	142823	5307
4	Bihar	60	77	67546.0	517	572	463	4379	484	4.60	...	1127	1127	5.0	31.3	33.6	13566	1600	245409	2252
5	Chandigarh	81	86	6156.0	13	87	4	33	6	22.22	...	59	59	52.0	62.3	26.3	33	65	6105	154
6	Chhatisgarh	66	84	44107.0	297	2122	0	716	21	29.00	...	1436	1436	43.4	49.6	27.3	809	405	64304	3658
7	D. and N. Haveli	64	90	306.0	5	2	0	1	0	4.23	...	5	5	8.7	11.1	9.6	6	2	761	8
8	Daman and Diu	82	89	351.0	0	3	3	6	0	4.85	...	2	2	7.7	14.6	12.0	4	5	379	5
9	Delhi	82	86	75117.0	190	4322	988	770	70	86.96	...	2096	2096	166.9	169.1	92.5	160	6464	100416	7849
10	Goa	87	90	5399.0	41	202	0	42	2	39.30	...	95	95	63.5	53.9	28.5	133	32	5885	356

3.3 ARCHITECTURE AND EXPLANATION

The entire idea of depicting various different crime visualizations has been deployed on Tableau. The following is the architecture of the way this project has been carried out -



1. Data Collection: The first step in crime analysis is to collect data related to crime. This can be done by collecting data from police stations, courts, and other sources. The data should include details about the crime, such as the type of crime, the location, the time of day, and any other relevant information.

2. Data Integration: Once the data has been collected, it needs to be integrated into a central database. This can be done by using data integration tools that can merge data from different sources into a single database.

3. Data Cleaning: The data collected may contain errors, duplications, or missing values. Therefore, the data needs to be cleaned and transformed into a standard format to ensure consistency and accuracy.

4. Data Analysis: The next step is to analyze the data using statistical and machine learning techniques. This can help identify patterns and trends in crime, as well as predict future crimes.

5. Visualization: The results of the analysis need to be visualized in a way that is easy to understand. This can be done by creating graphs, charts, and maps to represent the data.

6. Developing a predictive model - Multilinear regression has been used to make predictions based on the most significant attributes contributing towards total crimes in India by using feature scaling according to the correlation matrix.

7. Reporting: The final step is to create reports based on the analysis and visualization. These reports can be used by law enforcement agencies, policymakers, and other stakeholders to make informed decisions.

Overall, the architecture of crime analysis in India project should focus on data collection, integration, cleaning, analysis, visualization, and reporting. It might also use advanced technologies such as machine learning, artificial intelligence, and data visualization tools to provide insights into crime patterns and trends.

4. PROPOSED WORKS

4.1 NOVELTY

Developing a multilinear regression model on our dataset to find all the 'most important' attributes and how significantly they are contributing towards the target variable 'total violent crimes' is something we have added on in this project apart from the visualization and creating a dashboard on tableau. More information regarding the model will be provided in the appendix.

4.2 PROJECT CONTRIBUTIONS

4.2.1

Dhruvi Ochani (20BCE1882) found the dataset on **https://data.gov.in/** and had major contribution in making the dashboard on tableau with all the necessary visualizations and making a part of the report ensuring all the submissions were made on time.

4.2.2

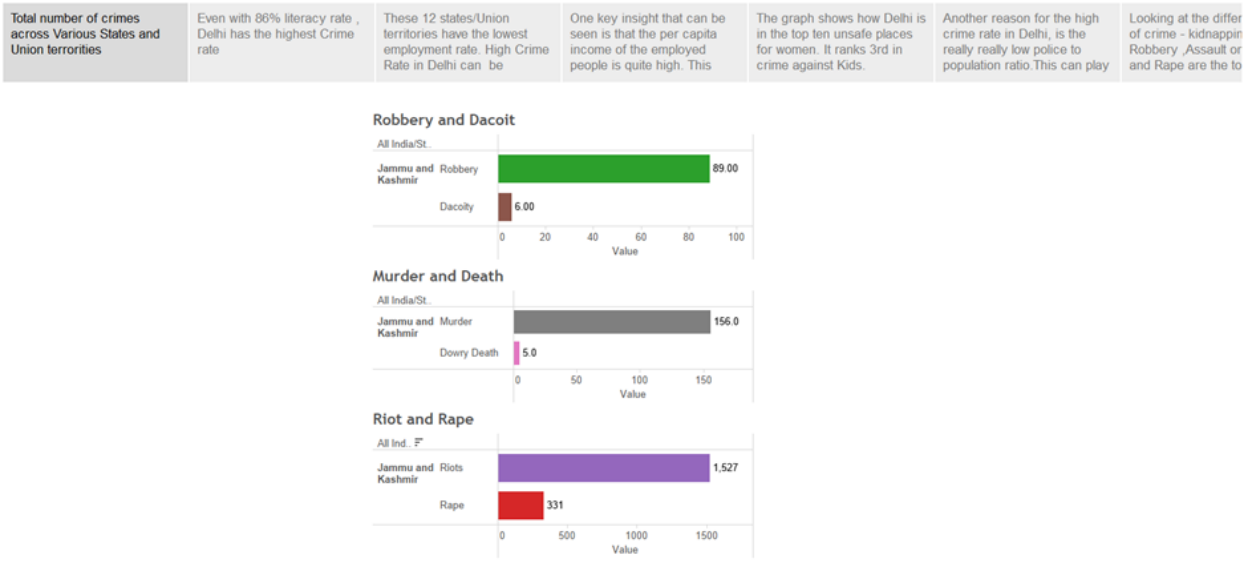
Dolly Agarwala (20BCE1863) did her part in making the tableau dashboard, finding most of the research papers required for literature survey and making a significant part of the report along with finding resources for the project.

4.2.3

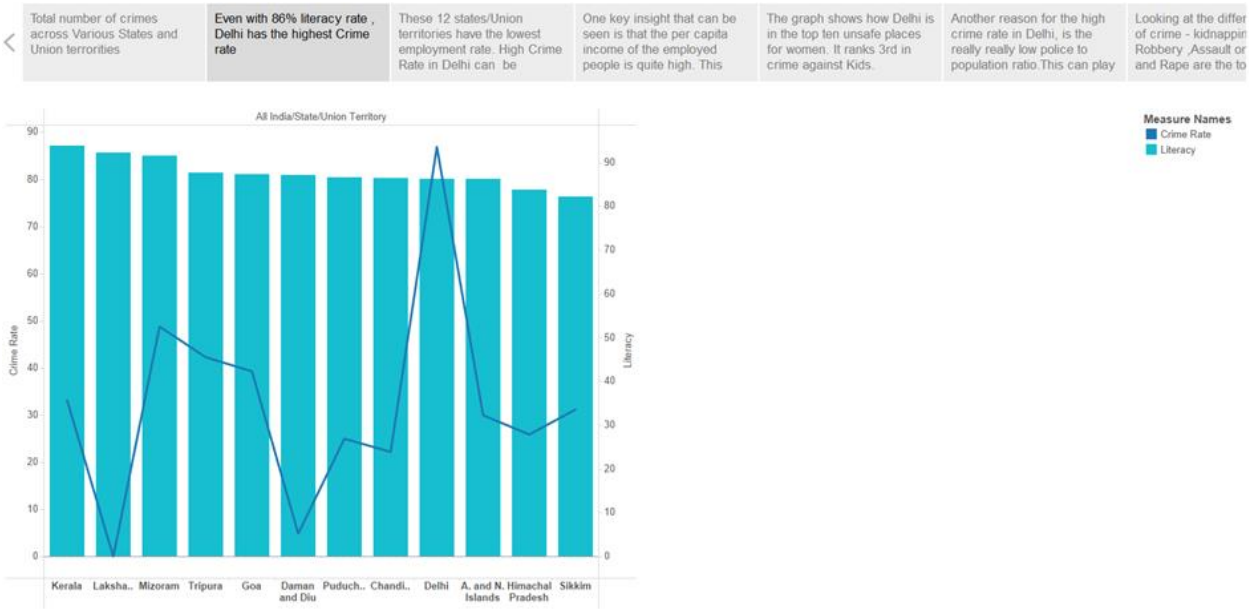
Aryan Bhardwaj (20BCE1908) did his part in making a few visualizations for the dashboard, finding research papers for literature review and developing the multilinear regression model for making predictions along with making a part of the report.

5. RESULTS & DISCUSSION

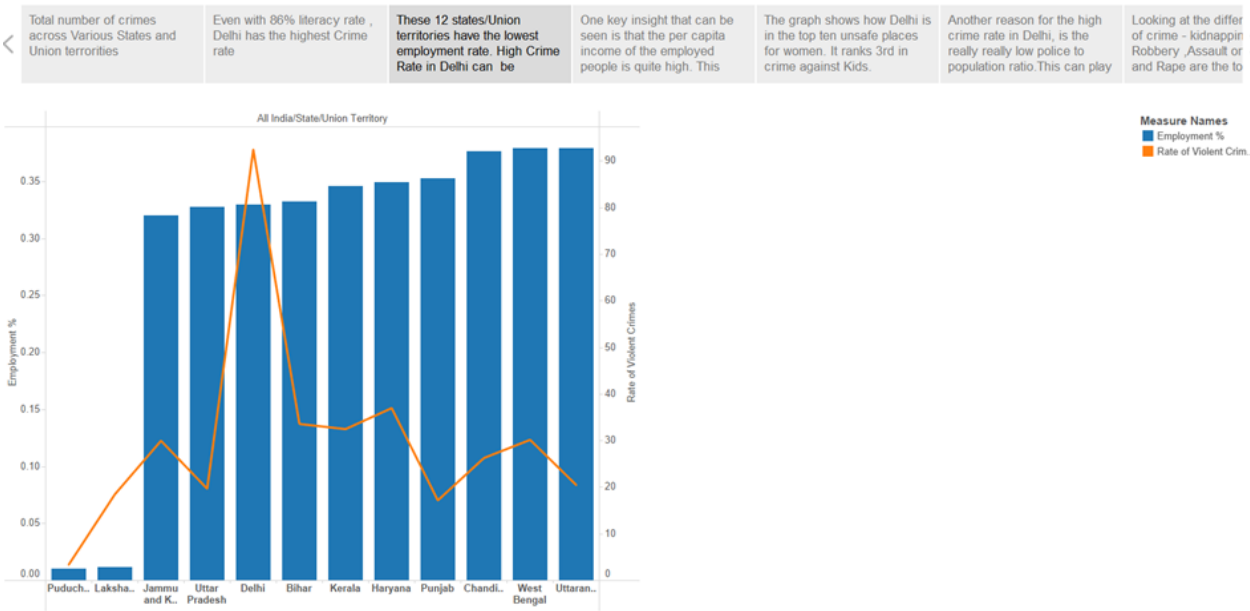
5.1 RESULTS



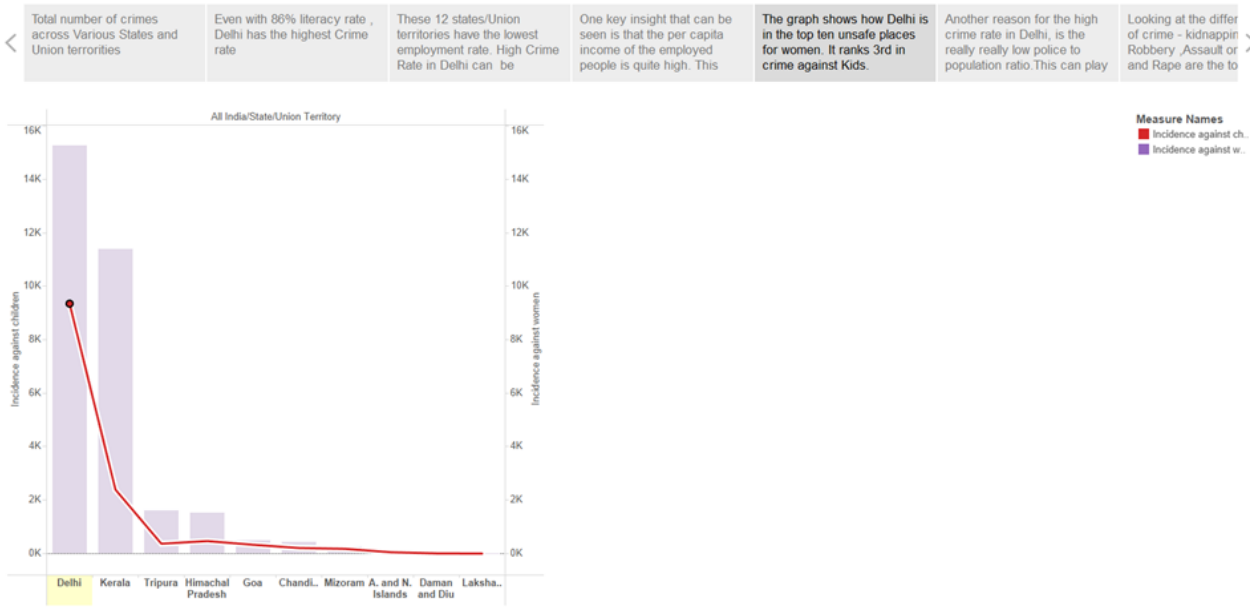
The above graph shows different crime rates for the state of Jammu and Kashmir. Similarly, we can find for all other states.



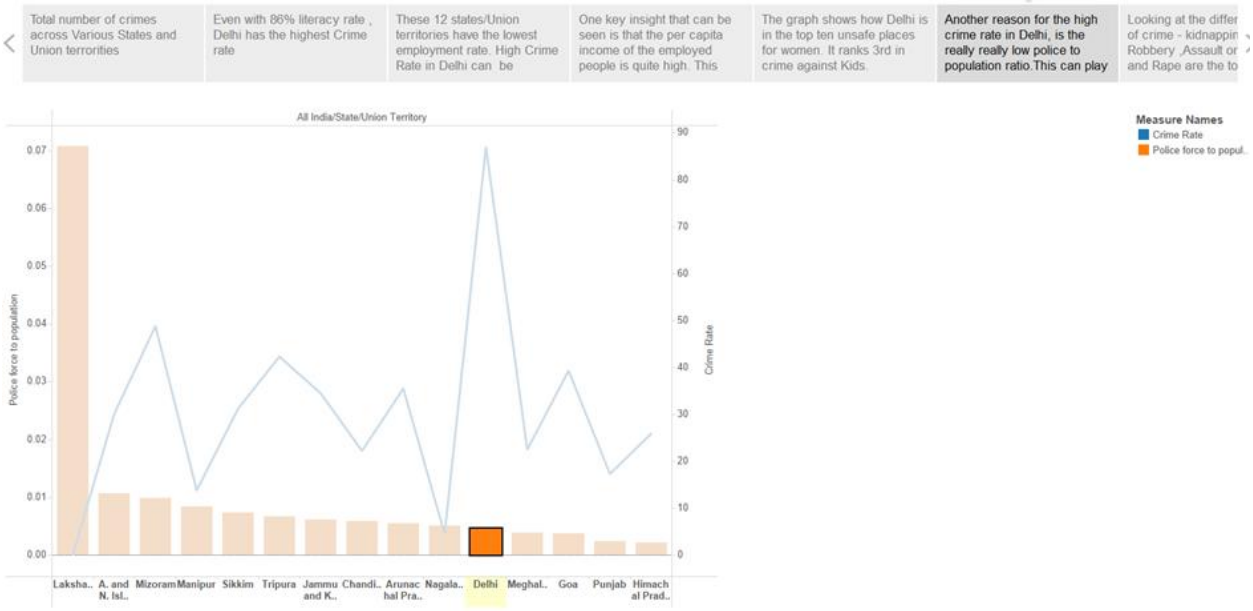
The above graph shows the literacy and crime rates for all the states.



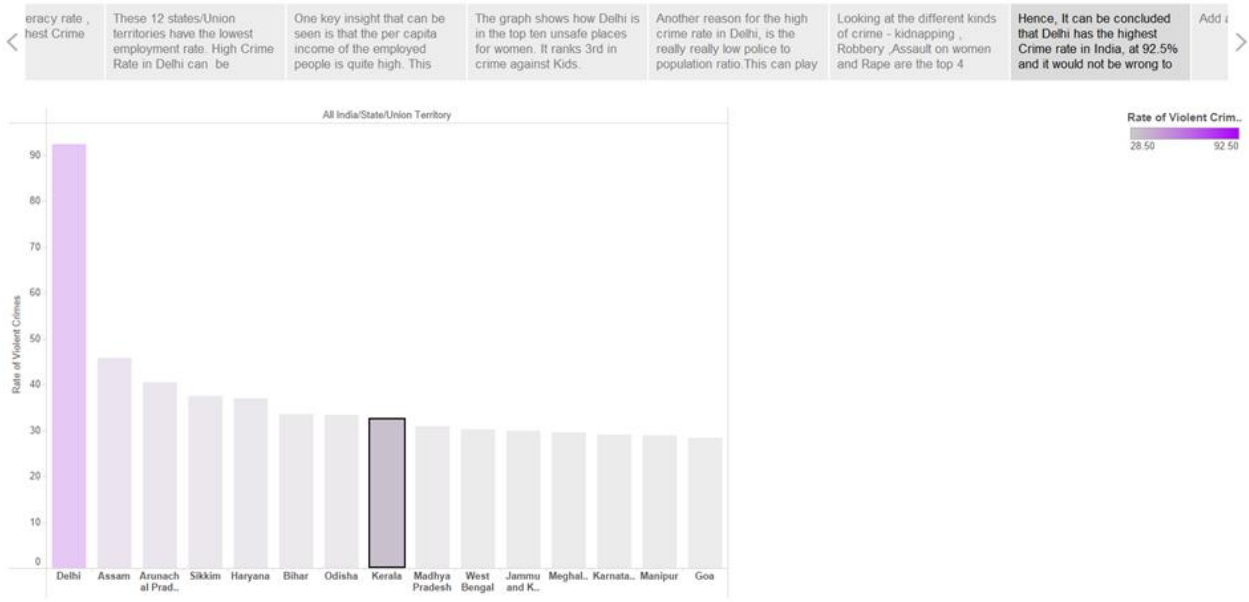
The above graph shows the employment and crime rates for all the states.



The above graph shows the number of incidents against children and women for all the states.

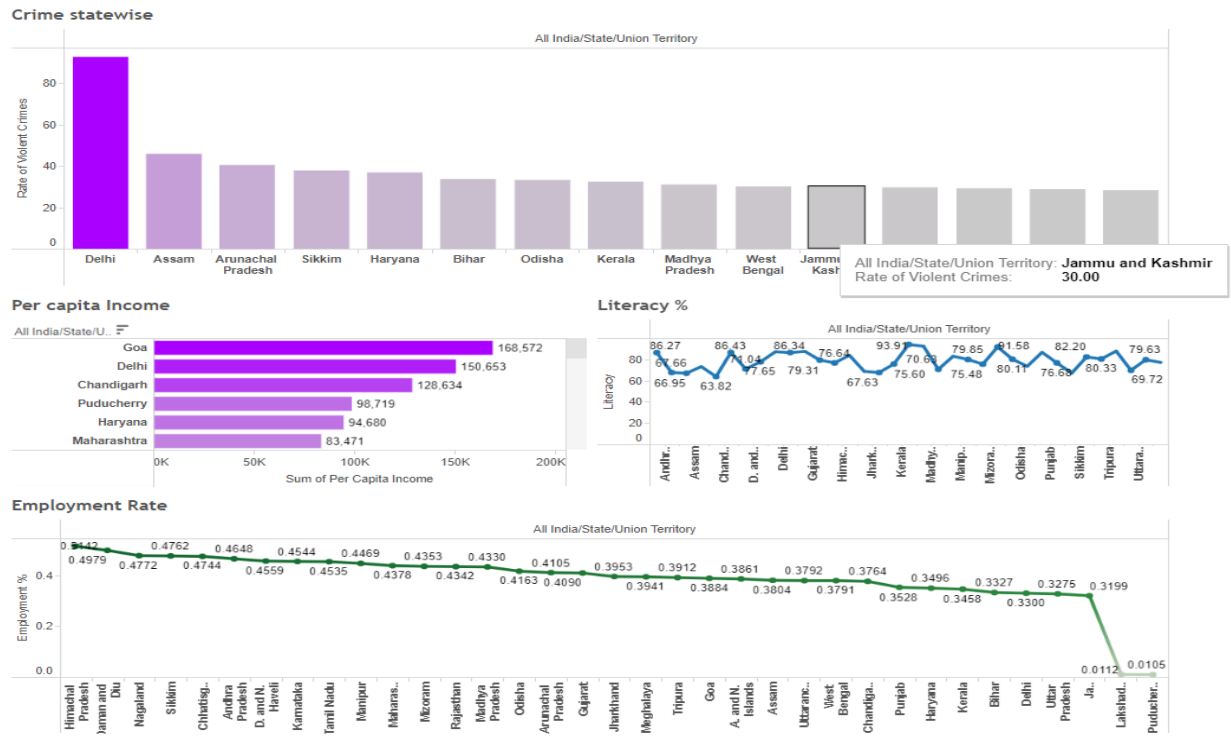


The above graph shows the police force to population ratio and crime rate for all the states.

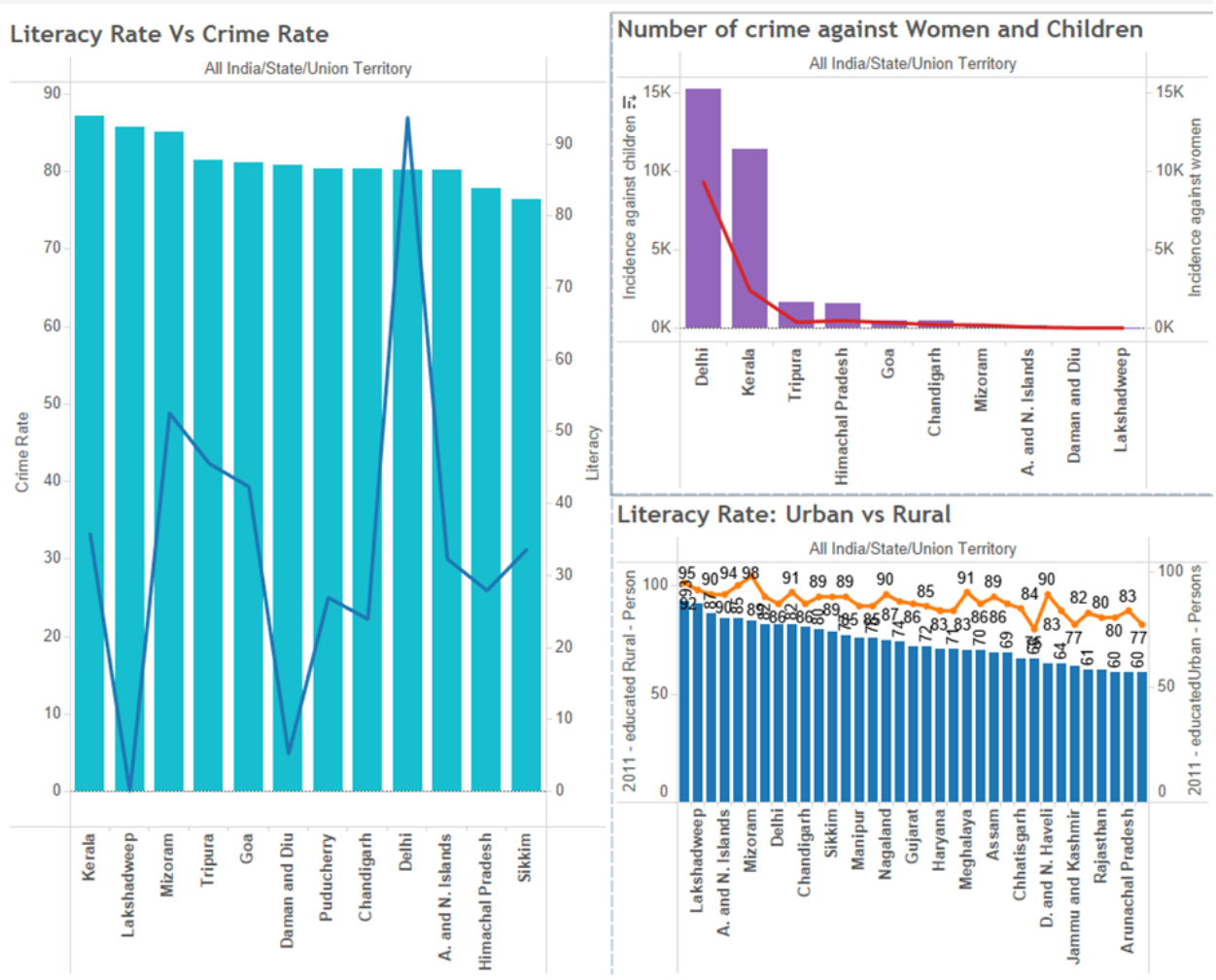


The above graph shows the rate of violent crimes for all the states.

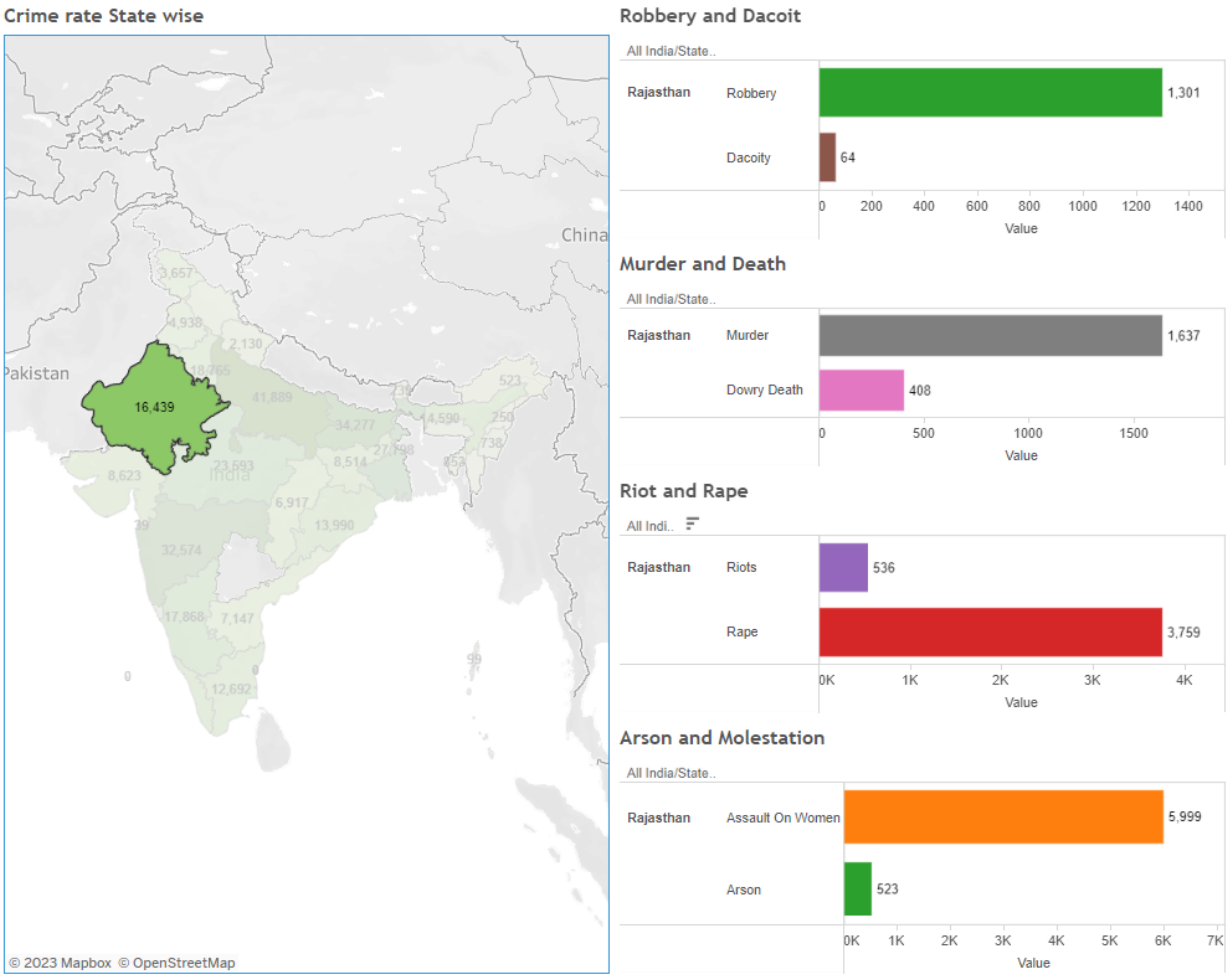
5.2 FIGURES, COMPARISON TABLES



The above is a dashboard showing 4 different visualizations: Crime statewise, Per capita Income, Literacy rate and Employement rate.



The above is a dashboard showing 3 different visualizations: Literacy rate vs Crime rate, Number of crime against Women and Children, Urban Literacy rate vs Rural Literacy rate.



The above shows the various different types of crimes for all the states of India. Clicking on the state on the map gives the graphs for the particular state as on the right.

5.3 EXPLANATION

Crime incidents tend to cluster in certain areas, such as urban slums, industrial estates, and tourist destinations. Spatial analysis techniques, such as hotspot analysis and density mapping, could identify these areas and help prioritize resource allocation and intervention strategies.

They tend to follow seasonal and diurnal patterns. Time-series analysis techniques, such as autoregression, forecasting & multilinear regression in this case, could predict future crime events and inform resource planning and deployment.

Data visualization techniques, such as pie charts , line graphs and bar graphs, could illustrate the frequency and trend of different types of crimes and inform predictions. Crime analysis can

provide valuable insights and evidence-based strategies to prevent and reduce crime and improve public safety in India.

6. CONCLUSION

The Tableau data visualization project on India's crime rates has shed important light on the patterns and trends in the nation's criminal activity.

The initiative has assisted in highlighting the high crime rate locations and the types of crimes that are common in various parts of India through the use of interactive dashboards. It has also been useful in determining the age and gender demographics most at risk for crime.

The project has been successful in conveying difficult information in an aesthetically pleasing and understandable way. This has made it simpler for stakeholders and policymakers to decide on crime prevention and law enforcement tactics in light of relevant information.

Overall, the study has shown the value of using data visualization to spot patterns and trends that would have been hard to spot from just looking at the raw data. It has also underlined the significance of utilizing technological tools like Tableau in order to make data-driven decisions and effectively address social challenges like crime.

7. REFERENCES

1. "Crime Analysis and Crime Mapping in India: A Review of Current Practices and Future Directions" by R. K. Mohanty and S. Mishra.
2. "Spatial Analysis of Crime in Delhi: A Geographical Information System (GIS) Based Approach" by A. K. Singh, S. K. Sharma, and P. Kumar.
3. "Crime Hotspots and Mapping in India: A Review" by S. S. Nair, P. M. Ravi, and P. N. Krishnan.
4. "Exploring Crime Hotspots in India: A Spatial Analysis Approach" by S. Kumari and S. Bhattacharya.
5. "Predicting Crime Hotspots in Bangalore City: A Geospatial Analysis" by S. G. Bhat and S. Mukherjee.
6. C. Chauhan and S. Sehgal, "A review: Crime analysis using data mining techniques and algorithms," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, pp. 21-25, doi: 10.1109/CCAA.2017.8229823.

7. H. Kaur, T. Choudhury, T. P. Singh and M. Shamoon, "Crime Analysis using Text Mining," 2019 International Conference on Contemporary Computing and Informatics (IC3I), Singapore, 2019, pp. 283-288, doi: 10.1109/IC3I46837.2019.9055606.
8. Geospatial crime analysis and forecasting with machine learning techniques by Rajiv Pandey, Sunil Kumar Khatri, Neeraj kumar Singh, Parul Verma.

APPENDIX

```
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
%matplotlib inline
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.feature_selection import RFE
from sklearn.linear_model import RidgeCV, LassoCV, Ridge, Lasso
```

```
df_crime = pd.read_csv("crime_dv.csv")
df_crime.head(10)
```

	All India/State/Union Territory	2011 - educated Rural - Person	2011 - educatedUrban - Persons	Actual Strength of Police force - Total	Arson	Assault On Women	Attempt to commit Culpable Homicide	Attempt To Commit Murder	Attempt To Commit Rape	Crime Rate	...	Rape	Rape 1	Rate of Total Cognizable Crimes	Rate of Total Cognizable Crimes 1	Rate of Violent Crimes	Riots	Robbery	Total Number of IPC cases for Investigation including pending Cases from previous year	Total Sexual Offences Under IPC	Total Violent Crimes
0	A. and N. Islands	85	90	4043.0	5	41	2	11	0	30.00	...	32	32	36.8	44.2	18.4	10	2	1466	78	99
1	Andhra Pradesh	60	80	89404.0	450	4547	1	1540	165	32.80	...	961	961	13.1	65.1	14.0	1013	433	241502	8322	7147
2	Arunachal Pradesh	60	83	7455.0	39	121	0	48	10	35.50	...	83	83	28.6	57.4	40.6	33	61	3808	217	523
3	Assam	69	89	54069.0	743	3099	14	1142	227	34.20	...	1980	1980	11.6	123.4	45.9	2640	1038	142823	5307	14590
4	Bihar	60	77	67546.0	517	572	463	4379	484	4.60	...	1127	1127	5.0	31.3	33.6	13566	1600	245409	2252	34277
5	Chandigarh	81	86	6156.0	13	87	4	33	6	22.22	...	59	59	52.0	62.3	26.3	33	65	6105	154	440
6	Chhatisgarh	66	84	44107.0	297	2122	0	716	21	29.00	...	1436	1436	43.4	49.6	27.3	809	405	64304	3658	6917
7	D. and N. Haveli	64	90	306.0	5	2	0	1	0	4.23	...	5	5	8.7	11.1	9.6	6	2	761	8	39
8	Daman and Diu	82	89	351.0	0	3	3	6	0	4.85	...	2	2	7.7	14.6	12.0	4	5	379	5	37
9	Delhi	82	86	75117.0	190	4322	988	770	70	86.96	...	2096	2096	166.9	169.1	92.5	160	6464	100416	7849	18765

10 rows × 42 columns

```
[6] #finding correlation b/w all the coefficients sing Pearson Correlation
    plt.figure(figsize=(12,10))
    cor = df_crime.corr()
    sns.heatmap(cor, annot=True, cmap=plt.cm.Reds)
    plt.show()
```


2011 - educated Rural - Person
2011 - educatedUrban - Persons
Actual Strength of Police force - Total
Arson
Assault On Women
Attempt to commit Culpable Homicide
Attempt To Commit Murder
Attempt To Commit Rape
Crime Rate
Crime to police %
Dacoity
Dowry Death
Employement
Employment %
Estimated Children Population (In Lakhs)#
Incidence against children
Incidence against women
Kidnapping & Abduction
Literacy
Mid-year Projected Female Population + (in Lakhs)
Murder
Number of policemen per 100 square Kilometres of Area
Number of Records
Per Capita Income
Percentage Contribution To All-India Total
Percentage Contribution To All-India Total 1
Percentage Share
Percentage Share to All India Crimes
Percentage to Total Cognizable Crimes Under IPC
Police force to population
Population
Rape
Rape 1
Rate of Total Cognizable Crimes
Rate of Total Cognizable Crimes 1
Rate of Violent Crimes
Riots
Robbery
Total Number of IPC cases for Investigation including pending Cases from previous year
Total Sexual Offences Under IPC
Total Violent Crimes
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

```
[ ]  #since total violent crimes is our target varibale
     y = df_crime["Total Violent Crimes"]
     cor_target = abs(cor["Total Violent Crimes"])
     #Selecting highly correlated features
     relevant_features = cor_target[cor_target>0.7] #if correlation > 0.7 then hihgly correlated
     relevant_features
```

```
Actual Strength of Police force - Total                                  0.855055
Arson                                                                    0.721331
Assault On Women                                                         0.796802
Attempt To Commit Murder                                                 0.862410
Crime to police %                                                        0.799667
Dacoity                                                                  0.708462
Dowry Death                                                              0.812338
Employement                                                              0.920559
Estimated Children Population (In Lakhs)#                                0.908702
Incidence against children                                               0.776559
Incidence against women                                                  0.892921
Kidnapping & Abduction                                                   0.890547
Mid-year Projected Female Population + (in Lakhs)                        0.929703
Murder                                                                   0.938102
Percentage Contribution To All-India Total                               0.775849
Percentage Contribution To All-India Total 1                             0.894315
Percentage Share                                                         0.818430
Percentage Share to All India Crimes                                     0.999965
Population                                                               0.930050
Rape                                                                     0.761950
Rape 1                                                                   0.761950
Riots                                                                    0.802024
Total Number of IPC cases for Investigation including pending Cases from previous year    0.874550
Total Sexual Offences Under IPC                                          0.818780
Total Violent Crimes                                                     1.000000
Name: Total Violent Crimes, dtype: float64
```

MULTILINEAR REGRESSION MODEL TO PREDICT TOTAL VIOLENT CRIMES

```
[ ] x = df_crime[['Actual Strength of Police force - Total', 'Arson', 'Assault On Women','Attempt To Commit Murder','Crime to police %',
                  'Dacoity','Dowry Death','Employement','Estimated Children Population (In Lakhs)#','Incidence against children','Incidence against women',
                  'Kidnapping & Abduction','Mid-year Projected Female Population + (in Lakhs)','Murder','Percentage Contribution To All-India Total',
                  'Percentage Contribution To All-India Total 1','Population','Rape','Rape 1','Riots','Robbery','Total Number of IPC cases for Investigation including pending Cases from previous year',
                  'Total Sexual Offences Under IPC']]
    y = df_crime['Total Violent Crimes']
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.30)
```

```
[ ] from sklearn.linear_model import LinearRegression
    #Fitting the Multiple Linear Regression model
    mlr = LinearRegression()
    mlr.fit(x_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
print("Intercept: ", mlr.intercept_)
print("Coefficients:")
list(zip(x, mlr.coef_))
```

```
Intercept:  -230.09785448301955
Coefficients:
[('Actual Strength of Police force - Total', 0.015569832749020001),
 ('Arson', -0.10411495747294619),
 ('Assault On Women', 0.2422149854630805),
 ('Attempt To Commit Murder', 2.1132653775664787),
 ('Crime to police %', 979.7630751517521),
 ('Dacoity', 3.2921227083187943),
 ('Dowry Death', 5.192894920319697),
 ('Employement', 0.00013978640462620324),
 ('Estimated Children Population (In Lakhs)#', -37.145989297350766),
 ('Incidence against children', -1.936988527553028),
 ('Incidence against women', -0.1938467684240539),
 ('Kidnapping & Abduction', 0.6592914341496193),
 ('Mid-year Projected Female Population + (in Lakhs)', 116.32827731844698),
 ('Murder', -1.7486572244285794),
 ('Percentage Contribution To All-India Total', 1992.0547816035323),
 ('Percentage Contribution To All-India Total 1', 1258.5433810199675),
 ('Population', -0.00048641986710785956),
 ('Rape', 2.1533515853968113),
 ('Rape 1', 2.1533515854410354),
 ('Riots', 1.2704935621145745),
 ('Robbery', 1.2505359943718202),
 ('Total Number of IPC cases for Investigation including pending Cases from previous year',
  -0.00954464188486289),
 ('Total Sexual Offences Under IPC', -1.6206529317679592)]
```

```
[ ] #Prediction of test set
    y_pred_mlr= mlr.predict(x_test)
    #Predicted values
    print("Prediction for test set: {}".format(y_pred_mlr))
```

```
Prediction for test set: [ 1196.58733722   196.33892036    87.27397105  2769.91986071
   328.94671576   587.20888048   100.1429212  13547.46051936
   252.58061619]
```

```
#Actual value and the predicted value
mlr_diff = pd.DataFrame({'Actual value': y_test, 'Predicted value': y_pred_mlr})
mlr_diff.head(20)
```

	Actual value	Predicted value
13	1390	1196.587337
23	235	196.338920
2	523	87.273971
14	3657	2769.919861
22	809	328.946716
10	550	587.208880
29	239	100.142921
3	14590	13547.460519
24	250	252.580616

```python
#Model Evaluation
from sklearn import metrics
meanAbErr = metrics.mean_absolute_error(y_test, y_pred_mlr)
meanSqErr = metrics.mean_squared_error(y_test, y_pred_mlr)
rootMeanSqErr = np.sqrt(metrics.mean_squared_error(y_test, y_pred_mlr))
print('R squared: {:.2f}'.format(mlr.score(x,y)*100))
print('Mean Absolute Error:', meanAbErr)
print('Mean Square Error:', meanSqErr)
print('Root Mean Square Error:', rootMeanSqErr)
```

```
R squared: 99.94
Mean Absolute Error: 361.79102789007703
Mean Square Error: 261520.4062373008
Root Mean Square Error: 511.3906591220657
```