



Comprehensive Report
on
“Data Analytics and Data Visualization with Tableau”
(Summer Course)

Submitted by:

Name: VIGHNESH M S
SRN: PES2UG19CS452
SEM: 7TH

Faculty Handling the Course:

Dr.Sudeepa Roy Dey
Dr.Prajwala TR
Prof.Ruby Dinakar

4th -9th, July 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
PES UNIVERSITY, EC campus

(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, Ind

Table of Contents:

<u>Content</u>	<u>Page no.</u>
Data Analytics <ul style="list-style-type: none">- Dataset- Identify the nature of dataset whether it is monotonic- Find correlation among the features using Pearson Correlations- Applying random forest	2-4
Data Visualization <ul style="list-style-type: none">- Cyber Criminals- The Trend- India Crime Stats-Crime Ratio	5-6
Insights	7

Data Analytics:

Dataset:

```
[ ] df
```

	City	Personal Revenge	Anger	Fraud	Extortion	Causing Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	...	Sale purchase illegal drugs	Developing own business	Spreading Piracy	Psycho or Pervert	Steal Information	Abetment to Suicide	Others	Total
0	Agra	5	0	19	0	0	0	0	0	0	...	0	0	0	0	0	0	46	70
1	Allahabad	0	0	222	11	8	0	0	0	0	...	0	0	0	0	0	0	0	241
2	Amritsar	2	0	5	0	0	0	2	0	0	...	0	0	0	0	0	0	0	9
3	Asansol	6	1	3	0	0	0	0	0	0	...	0	0	0	0	0	0	11	21
4	Aurangabad	5	2	51	0	0	0	21	0	0	...	0	0	0	0	0	0	0	82
...
63	Vadodara	0	0	9	0	2	0	0	0	0	...	0	0	0	0	0	0	1	12
64	Varanasi	28	29	124	5	27	39	36	2	0	...	0	0	18	0	0	0	127	448
65	Vasai Virar	1	0	16	0	0	0	1	1	0	...	0	0	1	0	0	0	9	29
66	Vijayawada	8	4	205	7	0	1	9	1	0	...	0	2	1	0	2	0	0	240
67	Vishakhapatnam	6	8	359	0	0	0	0	3	0	...	0	0	0	0	0	0	24	400

38 rows x 24 columns

Identify the nature of dataset whether it is monotonic:

```
[ ] df['City'].is_monotonic
```

```
False
```

```
[ ] df['Personal Revenge'].is_monotonic
```

```
False
```

```
[ ] df['Anger'].is_monotonic
```

```
False
```

```
[ ] df['Fraud'].is_monotonic
```

```
False
```

```
[ ] df['Extortion'].is_monotonic
```

```
False
```

```
[ ] df['Causing Disrepute'].is_monotonic
```

```
False
```

```
[ ] df['Political Motives'].is_monotonic
```

```
[ ] False
```

```
[ ] df['Prank'].is_monotonic
```

```
False
```

```
[ ] df['Terrorist Activities'].is_monotonic
```

```
False
```

```
[ ] df['Sale purchase illegal drugs'].is_monotonic
```

```
True
```

```
[ ] df['Developing own business'].is_monotonic
```

```
False
```

```
[ ] df['Spreading Piracy'].is_monotonic
```

```
False
```

```
[ ] df['Psycho or Pervert'].is_monotonic
```

```
False
```

Find correlation among the features using Pearson Correlations:

df.corr(method = "pearson")

	Personal Revenge	Anger	Fraud	Extortion	Causing Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	Terrorist Recruitment	...	Sale purchase illegal drugs	Developing own business	Spreading Piracy	Psycho or Pervert	Steal Information	Abetment to Suicide	Others	Total	Crime Rate	Year
Personal Revenge	1.000000	0.843807	0.334453	0.189108	0.059288	0.612099	0.618562	0.526043	0.126098	NaN	...	NaN	-0.018836	0.732927	-0.060308	0.152727	NaN	0.726222	0.614392	0.554645	0.051287
Anger	0.843807	1.000000	0.390607	0.181303	0.111092	0.815348	0.558858	0.425349	0.005801	NaN	...	NaN	0.040850	0.782035	-0.046792	0.058394	NaN	0.701346	0.670580	0.604937	0.117120
Fraud	0.334453	0.390607	1.000000	0.483221	-0.034409	0.128187	0.077272	0.092752	-0.033169	NaN	...	NaN	0.149114	0.151698	-0.061418	0.237429	NaN	0.290915	0.903391	0.894634	0.079642
Extortion	0.189108	0.181303	0.483221	1.000000	0.050817	0.182828	0.224344	0.069262	-0.057376	NaN	...	NaN	0.320940	0.199807	-0.016735	0.227113	NaN	0.203209	0.508298	0.573967	0.029353
Causing Disrepute	0.059288	0.111092	-0.034409	0.050817	1.000000	0.093221	0.013233	-0.028911	NaN	NaN	-0.046642	0.203473	0.083677	-0.046232	NaN	0.110974	0.163081	0.170817	0.090898
Prank	0.612099	0.815348	0.128187	0.182828	0.165125	1.000000	0.575660	0.203687	-0.012015	NaN	...	NaN	-0.035246	0.851620	-0.030584	-0.012015	NaN	0.661209	0.458199	0.442149	0.058115
Sexual Exploitation	0.618562	0.558858	0.077272	0.224344	0.093221	0.575660	1.000000	0.296765	-0.042289	NaN	...	NaN	0.005664	0.662700	-0.058722	0.072738	NaN	0.563006	0.377425	0.377534	0.041538
Political Motives	0.526043	0.425349	0.092752	0.069262	0.013233	0.203687	0.296765	1.000000	0.246493	NaN	...	NaN	-0.033016	0.200766	-0.040847	0.054933	NaN	0.117976	0.195399	0.148711	0.057647
Terrorist Activities	0.126098	0.005801	-0.033169	-0.057376	-0.028911	-0.012015	-0.042289	0.246493	1.000000	NaN	...	NaN	-0.022718	-0.022777	-0.014925	-0.014925	NaN	-0.027256	-0.029730	-0.041824	-0.122169
Terrorist Recruitment	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Terrorist Funding	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Inciting Hate against Country	0.360897	0.428698	0.017656	0.096788	0.133620	0.463623	0.398650	0.062692	-0.040672	NaN	...	NaN	-0.061906	0.527483	-0.040672	-0.040672	NaN	0.335441	0.218858	0.228531	-0.142678
Disrupt Public Service	0.716552	0.773469	0.137479	0.214998	0.210159	0.865925	0.664334	0.197199	-0.023218	NaN	...	NaN	-0.035329	0.978516	-0.023218	-0.023218	NaN	0.738249	0.492934	0.486001	0.047511
Sale purchase illegal drugs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Developing own business	-0.018836	0.040850	0.149114	0.320940	-0.046642	-0.035246	0.005664	-0.033016	-0.022718	NaN	...	NaN	1.000000	-0.018952	-0.022718	0.286242	NaN	0.138618	0.152193	0.194642	0.185952
Spreading Piracy	0.732927	0.782035	0.151698	0.199807	0.203473	0.851620	0.662700	0.200766	-0.022777	NaN	...	NaN	-0.018952	1.000000	-0.022777	0.028851	NaN	0.759021	0.507273	0.499939	0.087005
Psycho or Pervert	-0.060308	-0.046792	-0.061418	-0.016735	0.083677	-0.030584	-0.058722	-0.040847	-0.014925	NaN	...	NaN	-0.022718	-0.022777	1.000000	-0.014925	NaN	-0.051404	-0.058893	-0.068654	0.122169
Steal Information	0.152727	0.058394	0.237429	0.227113	-0.046232	-0.012015	0.072738	0.054933	-0.014925	NaN	...	NaN	0.286242	0.028851	-0.014925	1.000000	NaN	-0.051404	0.195413	0.181204	0.122169
Abetment to Suicide	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Others	0.726222	0.701346	0.290915	0.203209	0.110974	0.661209	0.563006	0.117976	-0.027256	NaN	...	NaN	0.138618	0.759021	-0.051404	-0.051404	NaN	1.000000	0.617755	0.584880	-0.018168
Total	0.614392	0.670580	0.903391	0.508298	0.163081	0.458199	0.377425	0.195399	-0.029730	NaN	...	NaN	0.152193	0.507273	-0.058893	0.195413	NaN	0.817755	1.000000	0.982679	0.089448
Crime Rate	0.554645	0.604937	0.894634	0.573967	0.079642	0.377534	0.148711	-0.041824	NaN	NaN	0.194642	0.499939	-0.068654	0.181204	NaN	0.584880	0.982679	1.000000	0.095679
Year	0.051287	0.117120	0.079642	0.029353	0.090698	0.058115	0.041538	0.057647	-0.122169	NaN	...	NaN	0.185952	0.087005	0.122169	0.122169	NaN	-0.018168	0.089448	0.095679	1.000000

23 rows x 23 columns

Applying random forest:

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **AdditiveRegression -S 1.0 -I 10 -W weka.classifiers.trees.DecisionStump**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: **10**
☐ Percentage split % **66**
 More options...

(Num) Disrupt Public Service

Start Stop

Result list (right-click for options):
 08:16:03 - trees.RandomForest
 08:16:16 - functions.LinearRegression
 08:16:27 - functions.MultilayerPerceptron
 08:16:57 - meta.AdditiveRegression

Classifier output

==== Run information ====
 Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
 Relation: Cyber Crime Stats
 Instances: 68
 Attributes: 24
 City
 Personal Revenge
 Anger
 Fraud
 Extortion
 Causing Disrepute
 Prank
 Sexual Exploitation
 Political Motives
 Terrorist Activities
 Terrorist Recruitment
 Terrorist Funding
 Inciting Hate against Country
 Disrupt Public Service
 Sale purchase illegal drugs
 Developing own business
 Spreading Piracy
 Psycho or Pervert
 Steal Information
 Abetment to Suicide
 Others
 Total
 Crime Rate
 Year
 Test mode: 10-fold cross-validation
 ==== Classifier model (full training set) ====
 RandomForest
 Bagging with 100 iterations and base learner
 weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
 Time taken to build model: 0.03 seconds
 ==== Cross-validation ====
 ==== Summary ====
 Correlation coefficient 0.8587
 Mean absolute error 0.261
 Root mean squared error 0.8867
 Relative absolute error 58.4298 %
 Root relative squared error 64.2979 %
 Total Number of Instances 68

Status OK

Log x 0

Applying Linear Regression:

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **AdditiveRegression -S 1.0 -I 10 -W weka.classifiers.functions.LinearRegression**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

More options...

(Num) Disrupt Public Service

Start Stop

Result list (right-click for options)

08:16:03 - trees.RandomForest

08:16:16 - functions.LinearRegression

08:16:27 - functions.MultilayerPerceptron

08:16:57 - meta.AdditiveRegression

Classifier output

Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4

Relation: Cyber Crime Stats

Instances: 68

Attributes: 24

City

Personal Revenge

Anger

Fraud

Extortion

Causing Disrepute

Prank

Sexual Exploitation

Political Motives

Terrorist Activities

Terrorist Recruit-ment

Terrorist Funding

Inciting Hate against Country

Disrupt Public Service

Sale purchase illegal drugs

Developing own business

Spreading Piracy

Psycho or Pervert

Steal Information

Abetment to Suicide

Others

Total

Crime Rate

Year

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Linear Regression Model

Disrupt Public Service =

-0.3581 * City=Faridabad,Gwalior,Bhopal,Raipur,Varanasi +

0.3708 * City=Gwalior,Bhopal,Raipur,Varanasi +

1 * City=Bhopal,Raipur,Varanasi +

-0.5 * City=Raipur,Varanasi +

1.679 * City=Varanasi +

0.3708 * Spreading Piracy +

-1 * Psycho or Pervert +

-0.0128

Time taken to build model: 0.04 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.8848
Mean absolute error	0.3003
Root mean squared error	0.7853
Relative absolute error	67.2191 %
Root relative squared error	62.5934 %
Total Number of Instances	68

Status OK

Log

Applying Additive Regression:

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **AdditiveRegression -S 1.0 -I 10 -W weka.classifiers.functions.LinearRegression**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

More options...

(Num) Disrupt Public Service

Start Stop

Result list (right-click for options)

08:16:03 - trees.RandomForest

08:16:16 - functions.LinearRegression

08:16:27 - functions.MultilayerPerceptron

08:16:57 - meta.AdditiveRegression

Classifier output

Model number 6

Decision Stump

Classifications

City = Raipur : 0.48393021120293844

City != Raipur : -0.014664551854634496

City is missing : 1.193897921701685E-17

Model number 7

Decision Stump

Classifications

Others <= 99.5 : -0.007672892591162319

Others > 99.5 : 0.5140838036078746

Others is missing : -1.1020596200323244E-17

Model number 8

Decision Stump

Classifications

Psycho or Pervert <= 0.5 : 0.007129291873943334

Psycho or Pervert > 0.5 : -0.477662555542032

Psycho or Pervert is missing : 8.163404592832033E-19

Model number 9

Decision Stump

Classifications

City = Bhopal : 0.25760407628592674

City != Bhopal : -0.00780618412987656

City is missing : 2.2449362630288092E-17

Time taken to build model: 0.03 seconds

=== Cross-validation ===

=== Summary ===

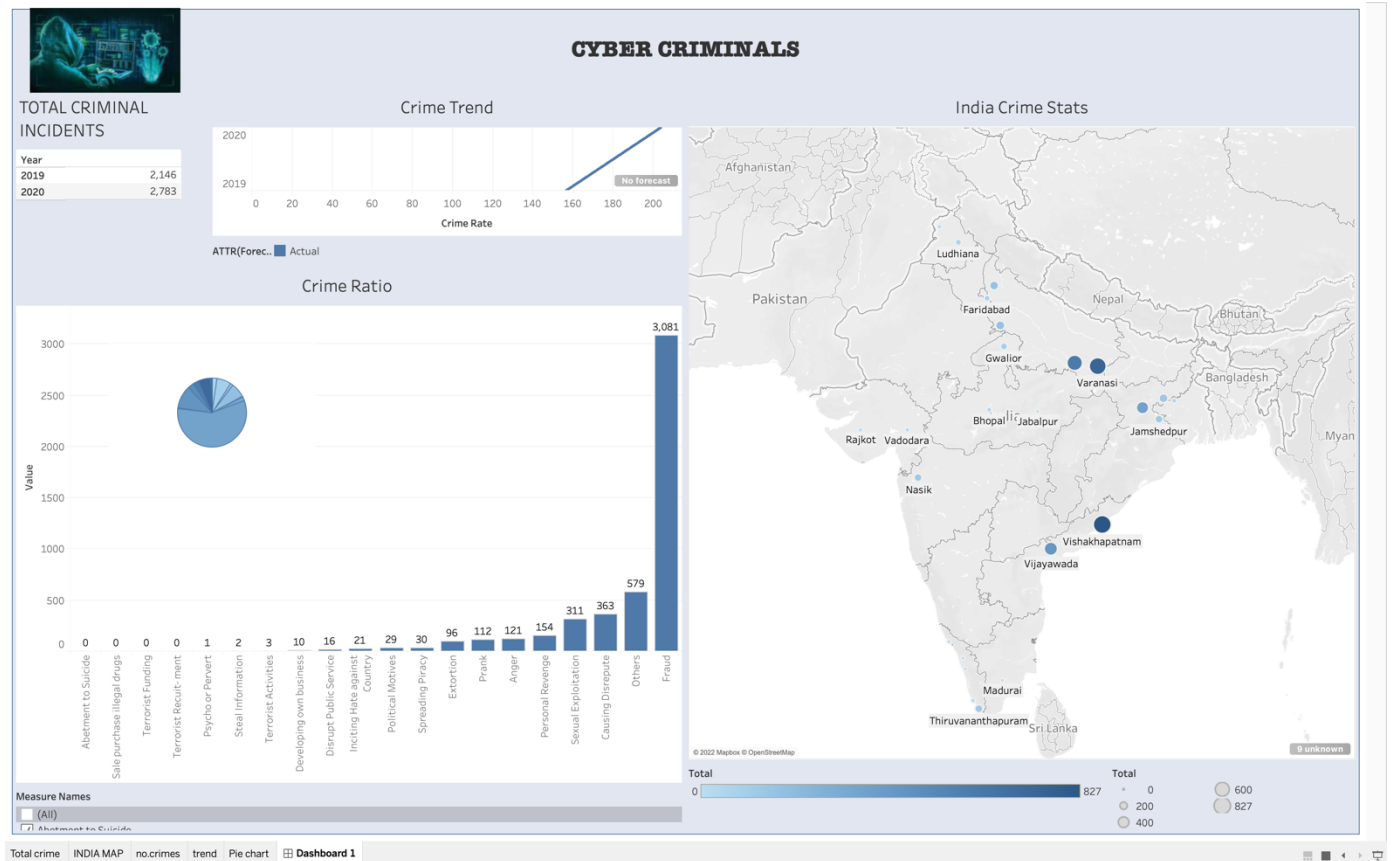
Correlation coefficient	0.6783
Mean absolute error	0.2932
Root mean squared error	1.115
Relative absolute error	65.6317 %
Root relative squared error	88.8709 %
Total Number of Instances	68

Status OK

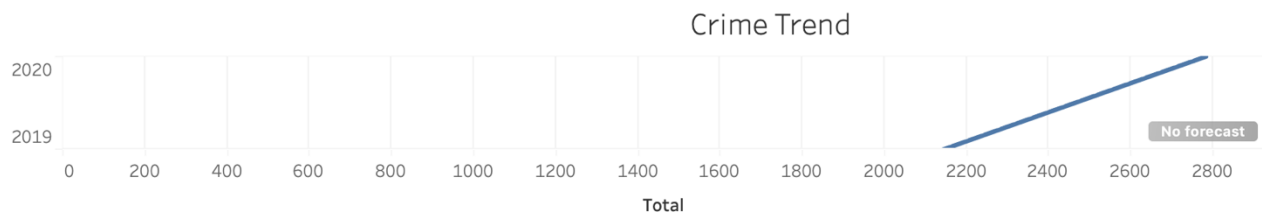
Log

Data Visualization:

Cyber Criminals:



The Trend and Forecast:

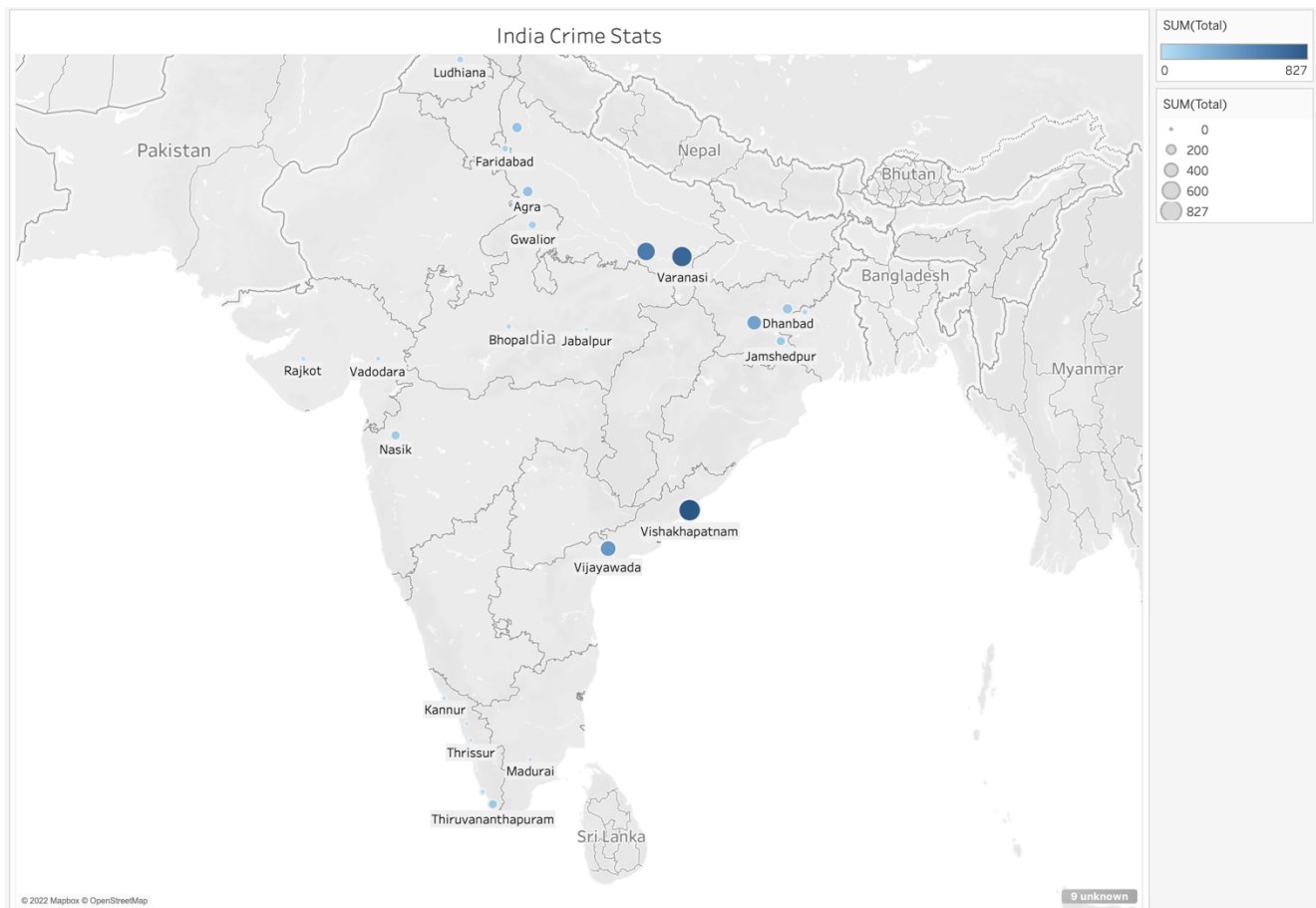


TOTAL CRIMINAL INCIDENTS:

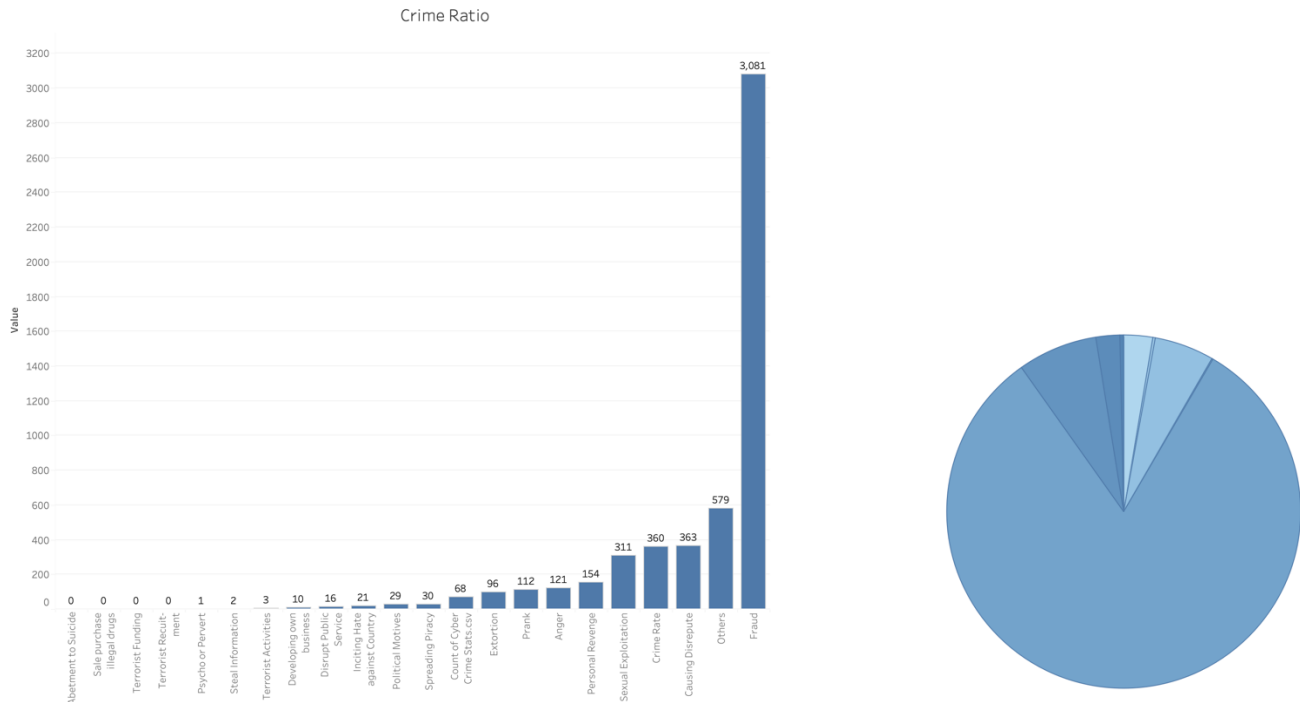
TOTAL CRIMINAL INCIDENTS

Year	
2019	2,146
2020	2,783

India Crime Stats:



Crime Ratio:



Insights:

Here we are trying to analyze the cyber crime data of the past couple of years.

- We can see that the cyber crime has increased from the year 2019 to 2020 in most of the cities except for those cities which have incorporated cyber security measures and have spread the awareness.
- We can see that the fraud is the highest committed cyber crime.
- People need to be made aware about the phonsi schemes and not to enter any phishing sites.
- The political crimes have been increased increasing as the year goes by, this is due to ability to not to trace back to the political parties for the scam.
- We need to increase women awareness as we can see the sexual exploitation has been increasing.