Team Alpha

MahilAI Datathon 2021

19 OCTOBER 2021

Team Members

Charvi Bannur and Jaaswin D Kotian

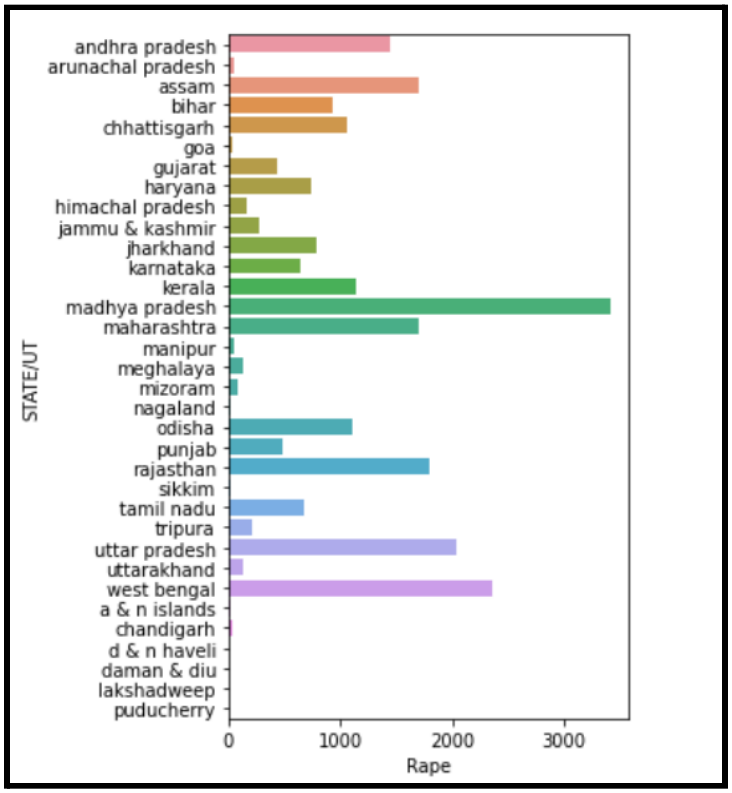
Problem Statement

In India, as per the latest National Family Health Survey (NFHS) 2015-16, 29.5% of women have experienced physical violence since age 15. Moreover, as per the latest National Crime Records Bureau (NCRB) report, an average of 87 rape cases was registered daily in India in 2019 (NCRB, 2020). Violence against women is deep-rooted and has unequal power and gender relations (CARE International, 2018), it transcends social and economic boundaries, and affects women and girls of all socio-economic backgrounds. So, to create awareness about the enormity of the issue, the following dataset has been chosen.

Dataset: https://drive.google.com/file/d/1OEzc9pXfAcOZizJ8UtNWGP5tycEhrGhP/view

Task-1

- **The data given was preprocessed-Converted to lowercase, separated with respect to year of survey, total cases per district element were removed, etc. Graphs were plotted for each category.**



- **Safest State/UT**, as per our analysis we found that the safest state was **Lakshadweep** followed by **Mizoram**.
    - For eg. We found the Average number of rape cases per district in each state for a particular year. We then considered the 5 states with the least number of reported rape cases for each category and for each year. The mode of all those values is our Safest State.

- **Most Unsafe State/UT**, as per our analysis we found that the most unsafe state was **Madhya Pradesh** followed by **Uttar Pradesh**.
    - A similar algorithm was followed

- **The cases with respect to the national average have been found out for each category in each year. The colab link for the same has been attached below.** (The leftmost column shows a state-wise comparison with the national average in percentage for the reported rape cases in 2011)

STATE/UT	DISTRICT	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	Percentage of rape	Rape w.r.t National Average (In %)
andhra pradesh	TOTAL	2011	1442	1612	599	4849	3658	13376	0	18.727273	107.446899
arunachal pradesh	TOTAL	2011	42	60	0	51	0	18	0	0.545455	93.957857
assam	TOTAL	2011	1700	3192	121	1193	8	5246	2	22.077922	144.562918
bihar	TOTAL	2011	934	3050	1413	790	11	2607	10	12.129870	34.365744
chhattisgarh	TOTAL	2011	1053	365	104	1654	174	834	2	13.675325	51.485149

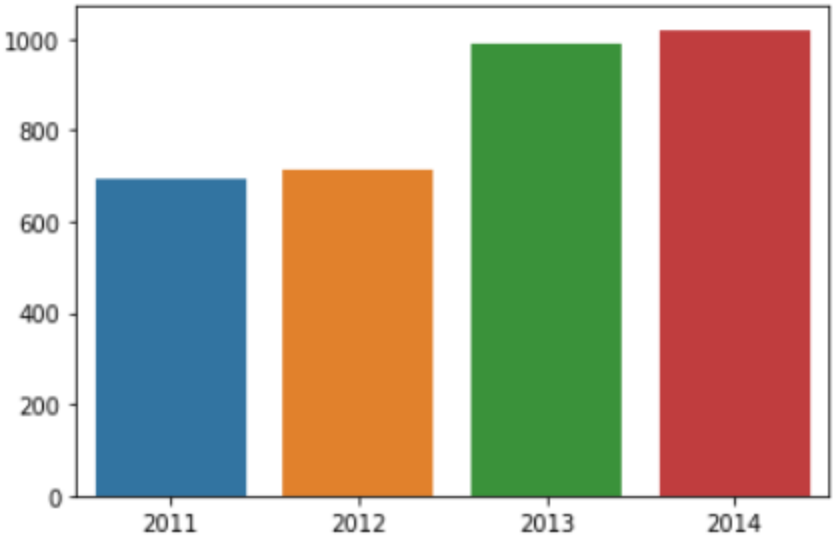
- **A similar approach was used for the rest of the categories.**
- **The percentage increase in the number of cases is also calculated**

```
[430] ((narape2014-narape2011)/narape2011)*100

      46.79769258977179
```

Rape cases increased by a rate of 46.79% all over India from 2011 to 2014

- **A Bar graph was plotted to show the increase in the number of crimes for each year.** (increasing number of rape cases from 2011-2014)



Conclusions from Task-1

* Rape cases increased by a rate of 46.79% all over India from 2011 to 2014

* Kidnapping and Abduction cases increased by a rate of 61.66% all over India from 2011 to 2014

* Dowry deaths decreased by a rate of 5.78% all over India from 2011 to 2014

* Assault on women with intent to outrage her modesty increased by a rate of 83.56% all over India from 2011 to 2014

* Insult to modesty of Women increased by a rate of 9.35% all over India from 2011 to 2014

* Cruelty by Husband or his Relatives increased by a rate of 18.95% all over India from 2011 to 2014

* Importation of girls decreased by a rate of 84.65% all over India from 2011 to 2014

Task-2

Safety index is calculated using the formula $1/log(total_{cases})$ * 100

STATE/UT	DISTRICT	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	Totalcases	safety_index
andhra pradesh	adilabad	2011	84	70	20	112	142	429	0	857	34.095006
andhra pradesh	anantapur	2011	16	83	36	156	16	198	0	505	36.991943
andhra pradesh	chittoor	2011	17	47	25	124	168	421	0	802	34.433194
andhra pradesh	cuddapah	2011	41	65	15	144	17	122	0	404	38.367371
andhra pradesh	cyberabad	2011	89	90	51	158	234	1411	0	2033	30.228491
...	...	...	...	...	...	...	...	...	...	...	...
lakshadweep	lakshadweep	2014	1	0	0	1	2	0	0	4	166.096405
lakshadweep	total	2014	1	0	0	1	2	0	0	4	166.096405
puducherry	karaikal	2014	3	1	0	12	1	1	0	18	79.663977
puducherry	puducherry	2014	7	6	1	20	7	3	0	44	60.847508
puducherry	total	2014	10	7	1	32	8	4	0	62	55.791377

Task-3

Model to predict the safety index was created using Random forest classifier and label encoder.

- Our Model-

```
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor(n_estimators = 1000, random_state = 42)
rf.fit(X_train,y_train)
```

```
RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
                      max_depth=None, max_features='auto', max_leaf_nodes=None,
                      max_samples=None, min_impurity_decrease=0.0,
                      min_impurity_split=None, min_samples_leaf=1,
                      min_samples_split=2, min_weight_fraction_leaf=0.0,
                      n_estimators=1000, n_jobs=None, oob_score=False,
                      random_state=42, verbose=0, warm_start=False)
```

- Accuracy-76.79%

```
mape = 100 * (errors / y_test)
accuracy = 100 - np.mean(mape)
print('Accuracy:', round(accuracy, 2), '%.')
```

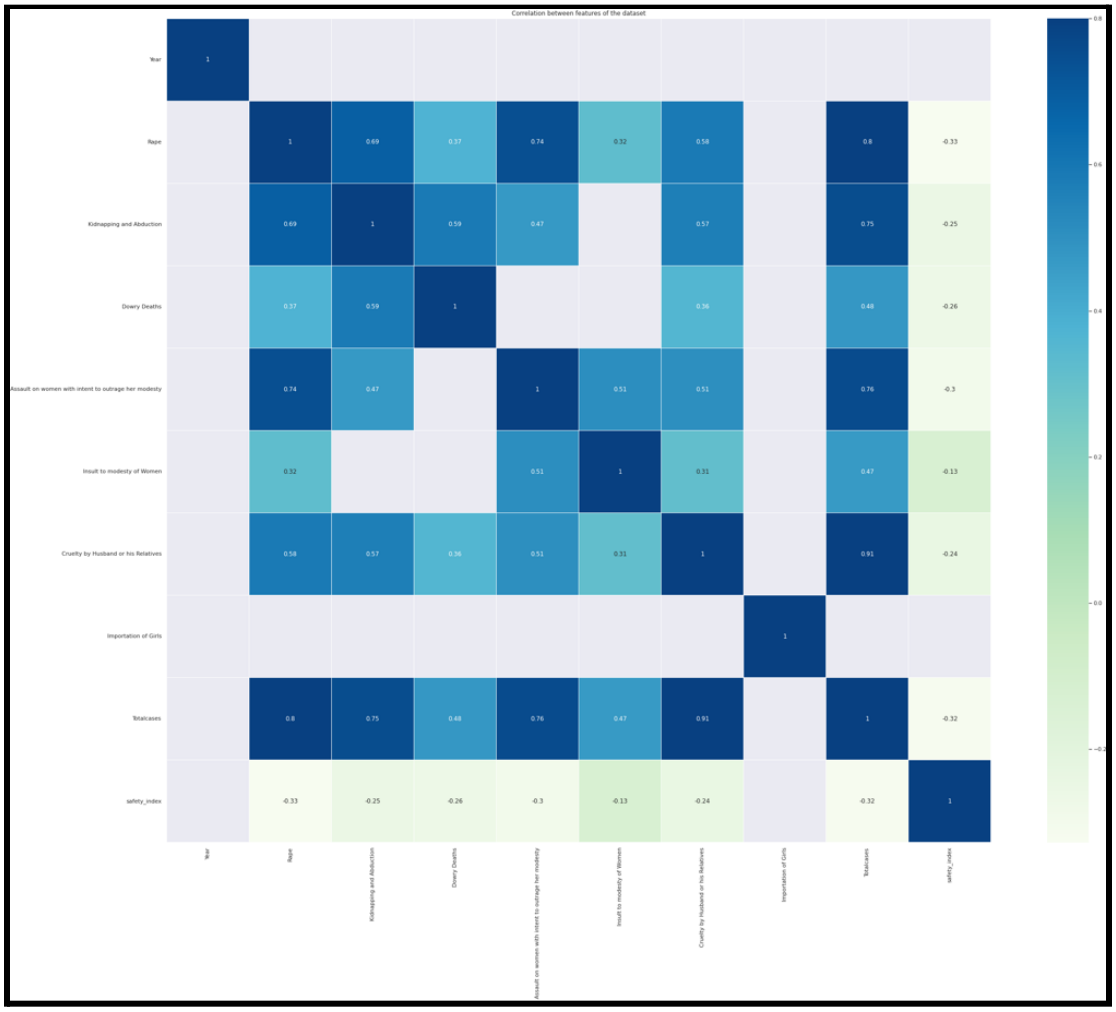
```
Accuracy: 76.79 %.
```

- Predicted Output for BANGALORE COMMR. is 37.76 (Indicates the fact that crimes are reducing)

```
predict('bangalore commr.',2021)
```

```
[37.76083638]
```

- Correlation graph:



Colab file links:

- Task-1: https://colab.research.google.com/drive/1LaQfbhzhnSOWJ4YMNxayEE6BfuxW_YAO?usp=sharing
- Task-2 and Task-3: https://colab.research.google.com/drive/1xtNNDobEI3oqfmKnMc84Tdl2reKiUi54?usp=sharing