Is education level a factor in domestic violence statistics?

WAIDATATHON 2021

Group 11

Karin Cheong
Dana Smith
Pallavi Kale
Vivian Chan

WAI

Combat domestic violence with Data & AI

Background

Aim: Analyze the data to find contributing factors towards perpetrating or accepting domestic violence

The dataset contains several demographics-based questions regarding attitudes towards domestic violence – marital status, education level, etc.

We wish to discover the significance of education level as a factor.

WAI

Angola	F	Marital status	Widowed, divorced, separated
Angola	F	Education	Primary
Angola	F	Employment	Employed for kind
Angola	F	Education	No education
Angola	F	Residence	Rural
Angola	M	Education	Higher
Angola	M	Age	35-49
Angola	M	Marital status	Married or living together
Angola	M	Residence	Urban
Angola	M	Education	Secondary
Angola	M	Age	25-34

Tableau
R
Python
SPSS

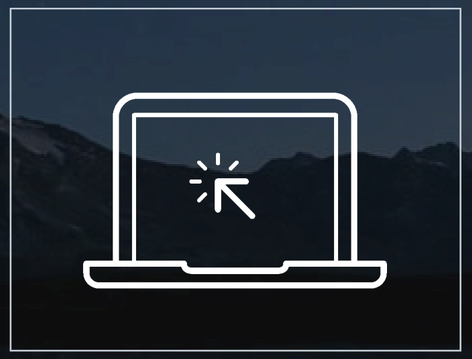
Approach

Education focus

Statistical Analysis

Data Visualization

Twitter Sentiment Analysis



Machine Learning

Visualization

Analysis of Data to Raise Awarness of Education in the Society

Crime and violence against women and girls remains a concern in the 21st Century.

Many countries in Africa,Asia and South America has a history of violence.
The data shows that both men and women justifies violence against female partners.
The crime percentage in no education group is high in comparison to high educational levels.
Women themselves justifies violence and the educating more girl child for awareness is the key..

Demograp..	Demograph..	Gender	
Education	No education	F	28.34%
		M	14.24%
	Primary	F	26.00%
		M	13.55%
	Secondary	F	19.17%
		M	12.20%
	Higher	F	9.28%
		M	6.65%

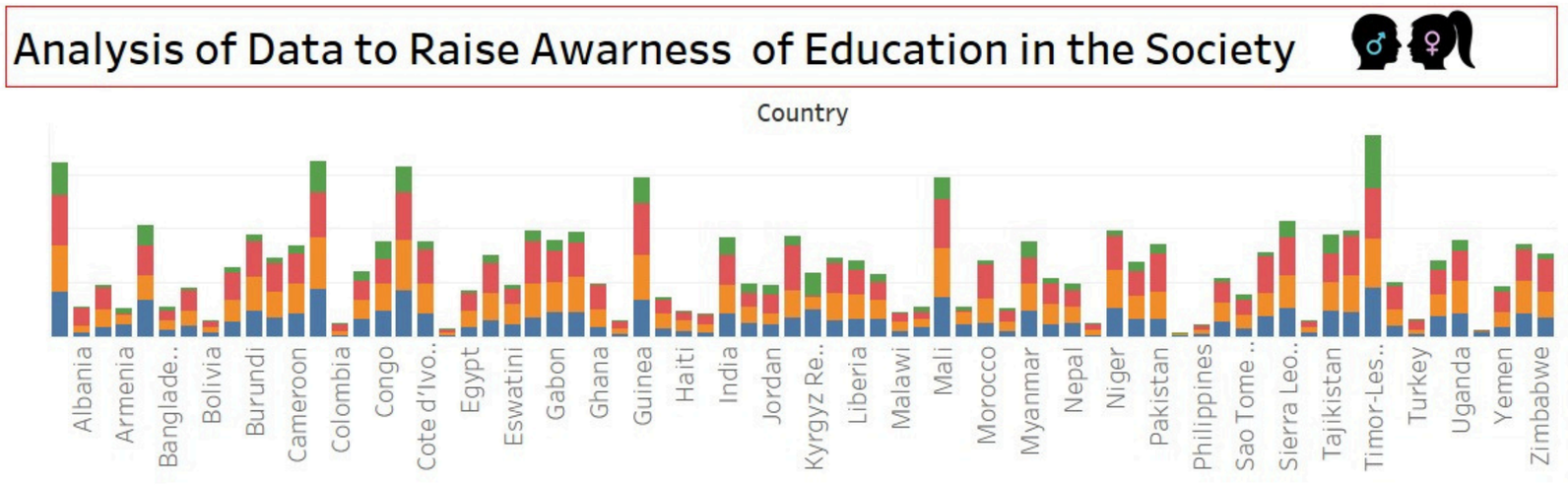
Select an indicator to see other demographics.

Demographics Question

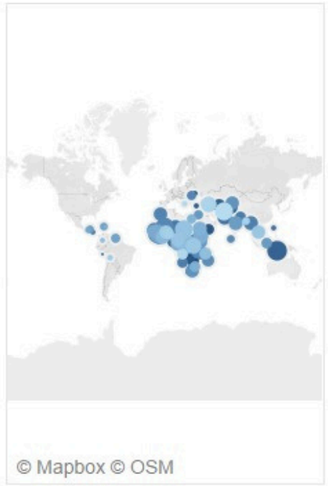
- [ ] Age
- [x] Education
- [ ] Employment
- [ ] Marital status
- [ ] Residence

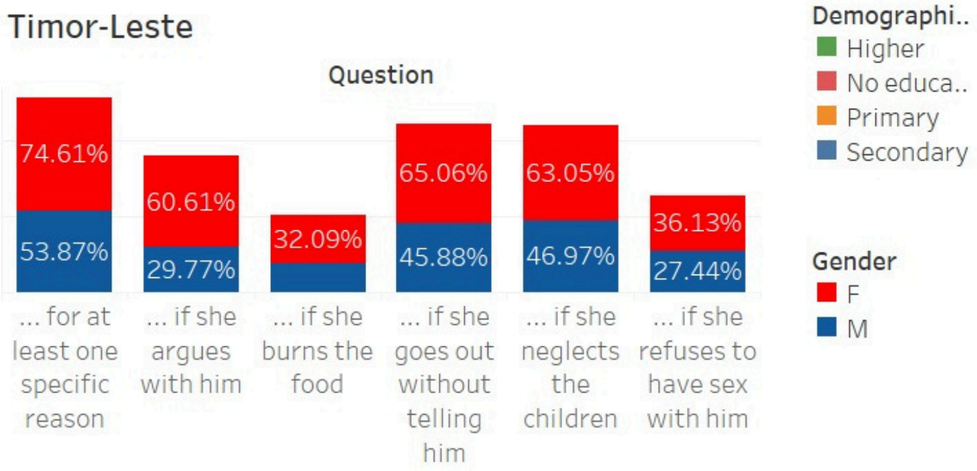
WAI

Visualization



Countries like Timor-Leste, Afghanistan, Chad,Guinea are highest ranking for crime against Women and Girls.





WAI

Visualization

Analysis of Data to Raise Awarness of Education in the Society

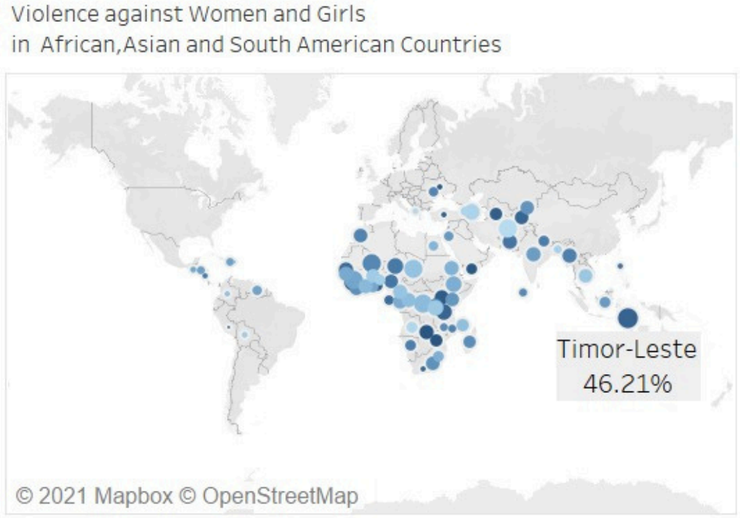
Violence among 15-24 age group is high with an mean of 32.80%.

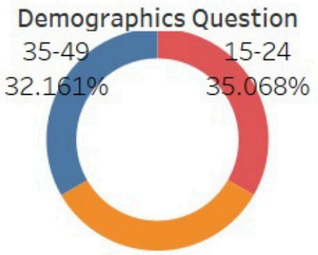
Violence is highest among No education group with a mean of 33.85%.

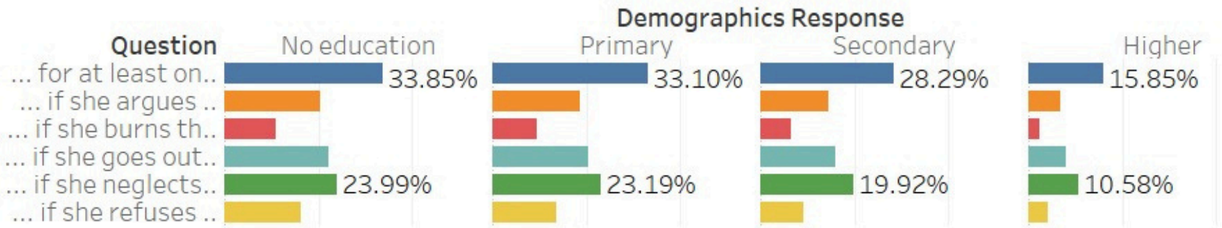
Higher education level has low percent of violence with a mean of 15.85%.

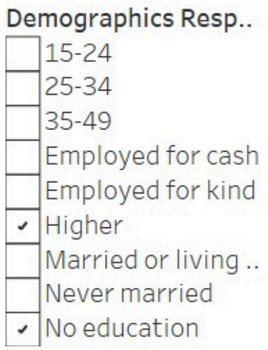
Employed for kind has highest level of violence at 35.66%.

..









WAI

Statistical Analysis

Compared between 3 educational levels

Number of samples length are equal (n=840)

Data are not normally distributed

Hypothesis testing *Kruskal-Wallis Test* on 2 or more samples in non-parametric data

WAI

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	DemographicsResponse	N	Percent	N	Percent	N	Percent
Value	Higher	840	100.0%	0	0.0%	840	100.0%
	Primary	840	100.0%	0	0.0%	840	100.0%
	Secondar	840	100.0%	0	0.0%	840	100.0%

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	DemographicsResponse	N	Percent	N	Percent	N	Percent
Value	Higher	840	100.0%	0	0.0%	840	100.0%
	Primary	840	100.0%	0	0.0%	840	100.0%
	Secondar	840	100.0%	0	0.0%	840	100.0%

Ho :The level of Education does not influences domestic violence

H1: The level of Education does influence domestic violence

Findings : There is statistically significant differences between three education levels. Reject from null hypothesis.

The distribution of Value is the same across categories of DemographicsResponse.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Correlation

Demographic Response vs Value

Nonparametric Correlations

[DataSet2]

Correlations

			DemographicsResponse	Value
Spearman's rho	DemographicsResponse	Correlation Coefficient	1.000	-.282**
		Sig. (2-tailed)	.	.000
		N	3360	3360
	Value	Correlation Coefficient	-.282**	1.000
		Sig. (2-tailed)	.000	.
		N	3360	3360

**. Correlation is significant at the 0.01 level (2-tailed).

Double-click to activate

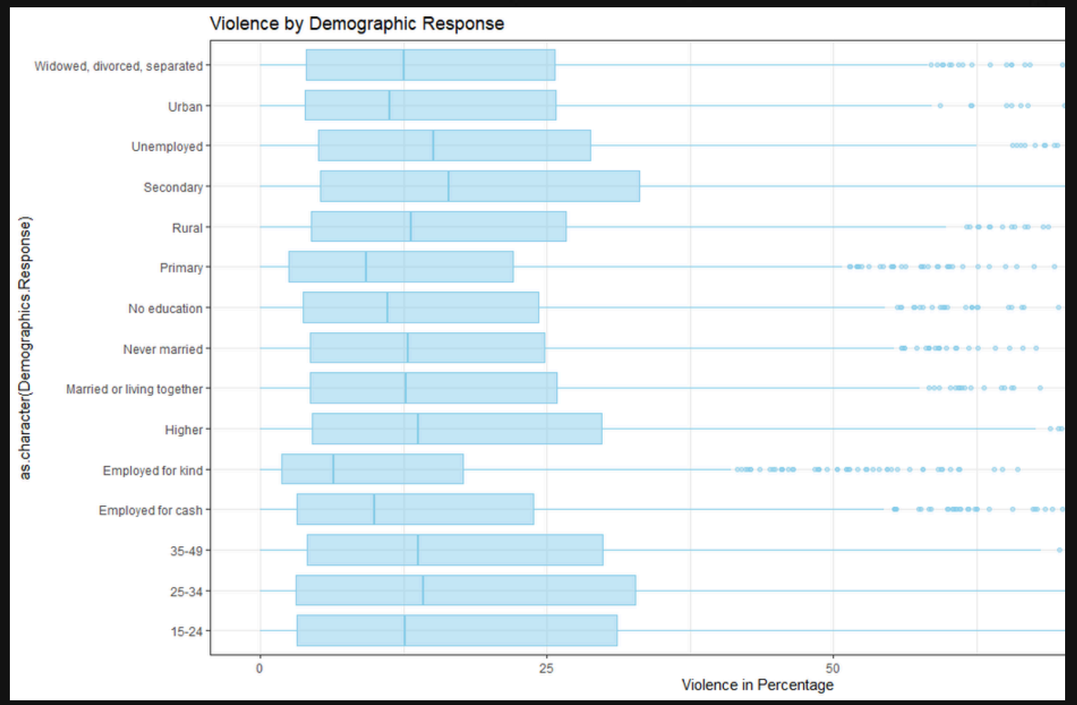
WAI

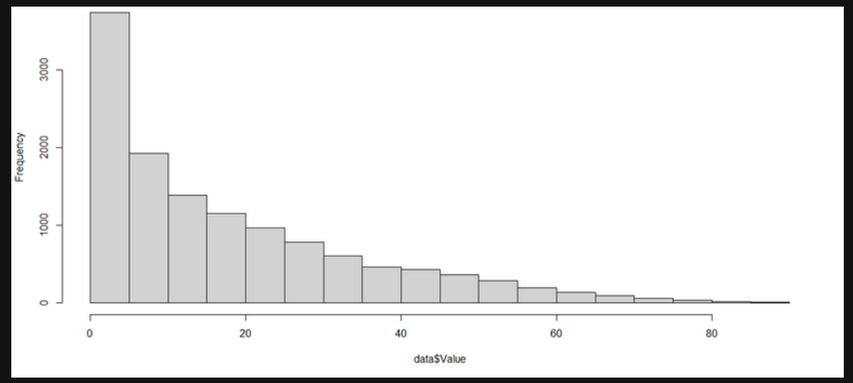
Final outcome:
Correlation statistically significant found

Machine Learning

Regression Model

```
>   mean(data$value)
[1] 17.54631
>   median(data$value)
[1] 12.1
>   max(data$value)
[1] 86.9
>   min(data$value)
[1] 0
>   range(data$value)
[1]  0.0 86.9
>   diff(range(data$value))
[1] 86.9
>   var(data$value)
[1] 295.0686
>   sd(data$value)
[1] 17.17756
>   kurtosis(data$value)
[1] 3.719582
>   skewness(data$value)
[1] 1.144544
>   summary(data$value)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    3.70   12.10   17.55   26.80   86.90
```





WAI

Regression Model

```
Call:
lm(formula = sqrt(Value) ~ Gender + Demographics.Response, data = train)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0907 -1.6227 -0.0581  1.6002  5.8151

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                     3.664479   0.093232  39.305  < 2e-16 ***
GenderM                                        -0.011120   0.046984  -0.237 0.812918
Demographics.Response25-34                      0.121260   0.128322   0.945 0.344702
Demographics.Response35-49                      0.076898   0.128321   0.599 0.549012
Demographics.ResponseEmployed for cash         -0.271818   0.128487  -2.116 0.034413 *
Demographics.ResponseEmployed for kind         -0.845519   0.126781  -6.669 2.73e-11 ***
Demographics.ResponseHigher                     0.107680   0.128432   0.838 0.401818
Demographics.ResponseMarried or living together -0.019534   0.129170  -0.151 0.879799
Demographics.ResponseNever married              -0.111414   0.127883  -0.871 0.383659
Demographics.ResponseNo education               -0.206378   0.127775  -1.615 0.106311
Demographics.ResponsePrimary                    -0.413115   0.127621  -3.237 0.001212 **
Demographics.ResponseRural                      0.018997   0.128601   0.148 0.882565
Demographics.ResponseSecondary                  0.426188   0.127937   3.331 0.000868 ***
Demographics.ResponseUnemployed                 0.007589   0.128266   0.059 0.952821
Demographics.ResponseUrban                     -0.143532   0.128265  -1.119 0.263158
Demographics.ResponseWidowed, divorced, separated -0.090191   0.128940  -0.699 0.484271
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.206 on 8804 degrees of freedom
Multiple R-squared:  0.01576,   Adjusted R-squared:  0.01408
F-statistic: 9.396 on 15 and 8804 DF,  p-value: < 2.2e-16
```

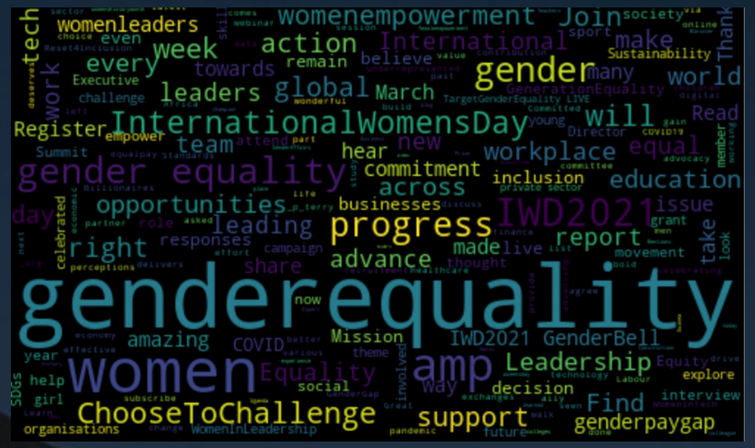
WAI

Linear Regression model was developed with independent variables gender and demographic response to predict violence percent.

Primary and secondary education were seen to be have statistical significance on the dependent variable.

Results
RMSE 21.71
MAPE : 4.18
MAE :14.8

Twitter Sentiment Analysis

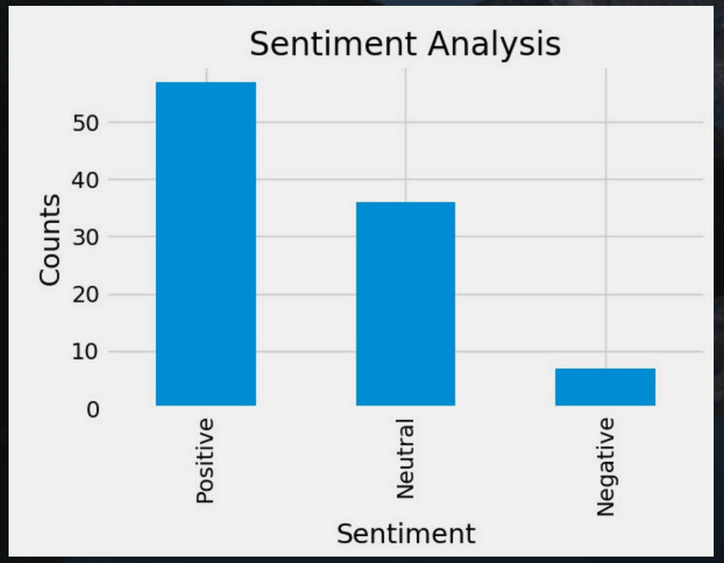


Performed a Twitter Sentiment Analysis in Python using a variety of relevant hashtags

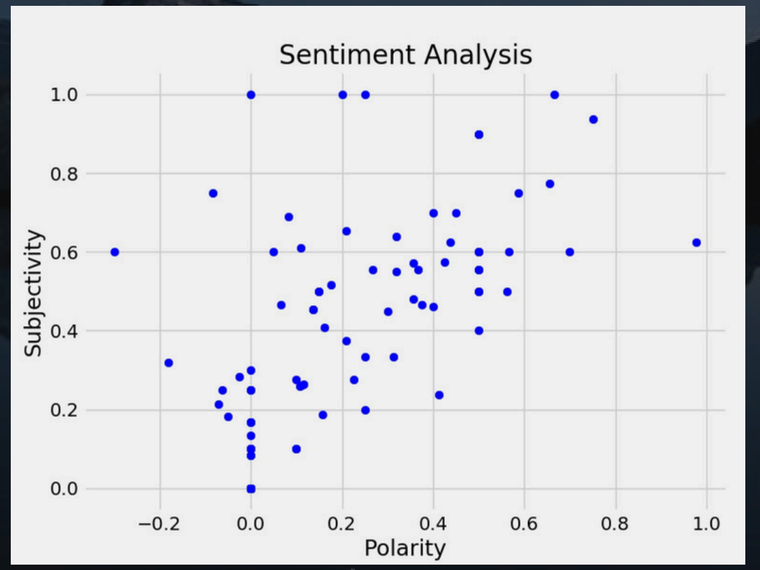
- #FemaleEmpowerment
- #DomesticViolence
- #genderinequality

Positive	57
Neutral	36
Negative	7

WAI



Analyzing these Tweets showed that Twitter users will mostly use these hashtags when they're posting a positive message.
Users tended not to be publicly open with negative comments.



Findings

There is statistical difference in different level of education. Women with less education feel more justified with their husband in hitting or beating them.

Countries with high percentage of higher educated population make up the smallest perpetrators of the abuse.

The majority of Twitter users show a positive attitude. Therefore true sentiment may not be accurately conveyed, but this does not reflect other related statistics.

WAI

Recommendations

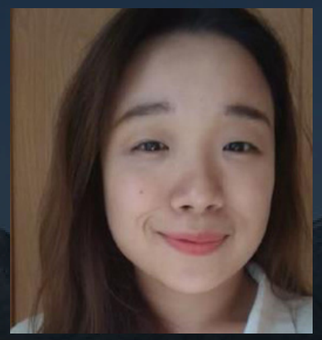
- Data distributed unevenly, some countries are having more survey samples than anothers
- Comparative analysis between countries with high proportion of higher educated population and lower educated population - in terms of how it have a significance on domestic violence.
- Further refining of regression model can bring down the error rate.
- The overall tweet message samples are generally toward positive sentimental scores, users tended not to be publicly open with negative comments.

WAI

Team Members

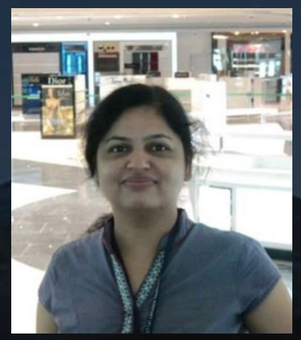
Karin Cheong

Linkedin:
haoiocheong



Vivian Chan

Linkedin:
vivian-jqchen



Pallavi Kale

Linkedin:
pallavi--kale

Dana Smith

Linkedin:
dána-smith-426aa362