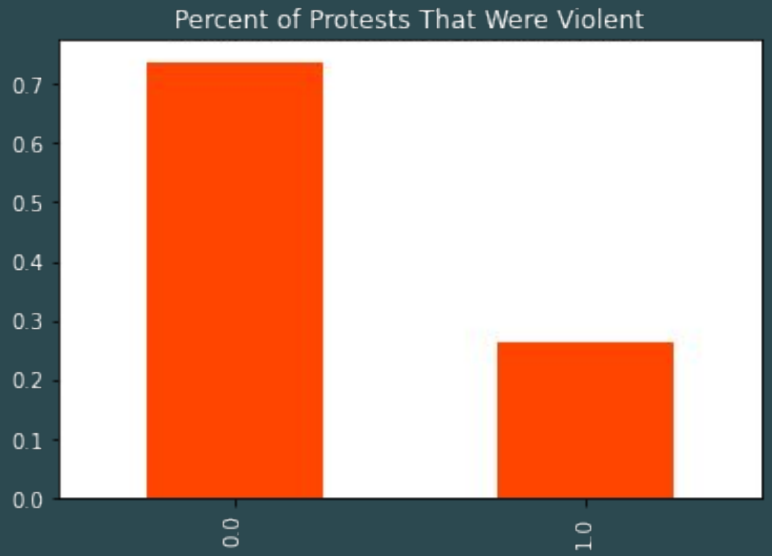
Predicting Protester Violence in Protests

Adrian Chapman | Jennifer Hofer | Sonali Bhatia | Steve Hydzik

Introducing the project

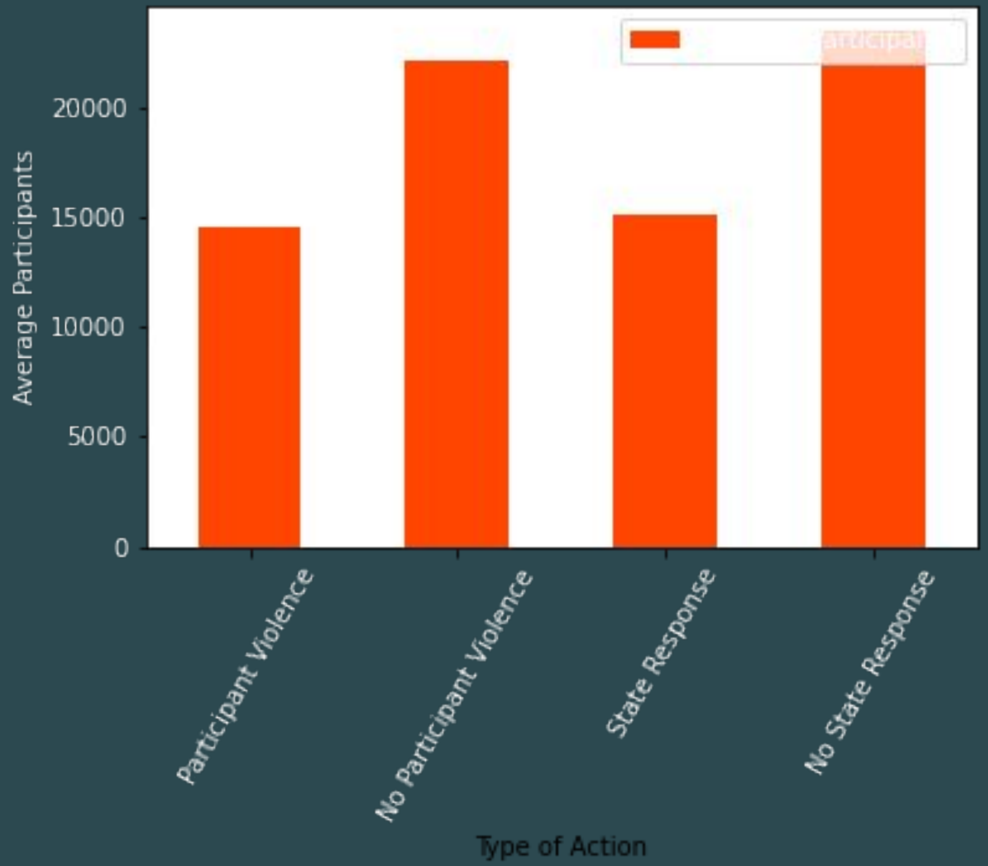
- Observing Data from The Mass Mobilization Project
  - Provides Data from protests all around the World
  - Data we will be focussing on: State Response, Protester Violence, Protester Demand, Notes(articles written about each protest), Countries, Dates, Duration and number of protesters
- Data spans from 1990 to 2019
- Discuss Data Cleaning (numeric and text)
- Discuss interesting findings within specific types of protests
- Different Modeling to predict Protester Violence

Variables



- Target: Protester Violence
  - Non Violent: 73.705%
  - Violent: 26.295%
- Features:
  - Duration
  - Dates
  - Country & Region
  - State Response (Ignore, Crowd Dispersal, Accomodation, Arrests, Beatings, Shootings, Killings)
  - Protester Demands (Social Restrictions, Political Behavior, Removal of Politician, Police Brutality, Land / Farm Issue, Labor Wage Dispute, Tax Policy)
  - Number of Participants in the Protest

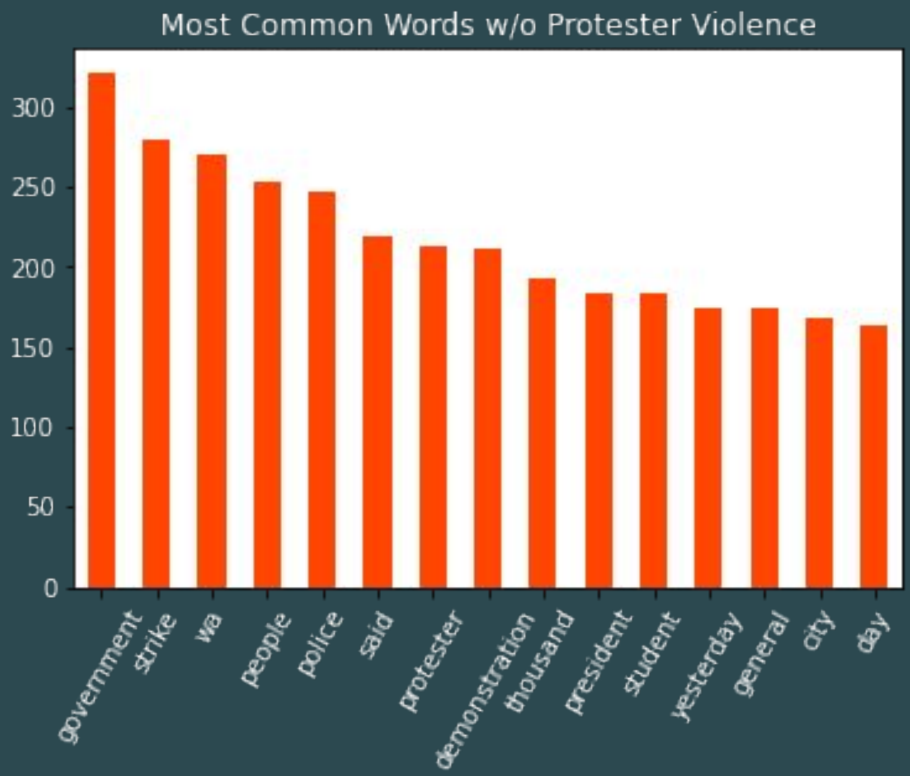
Interesting Finding

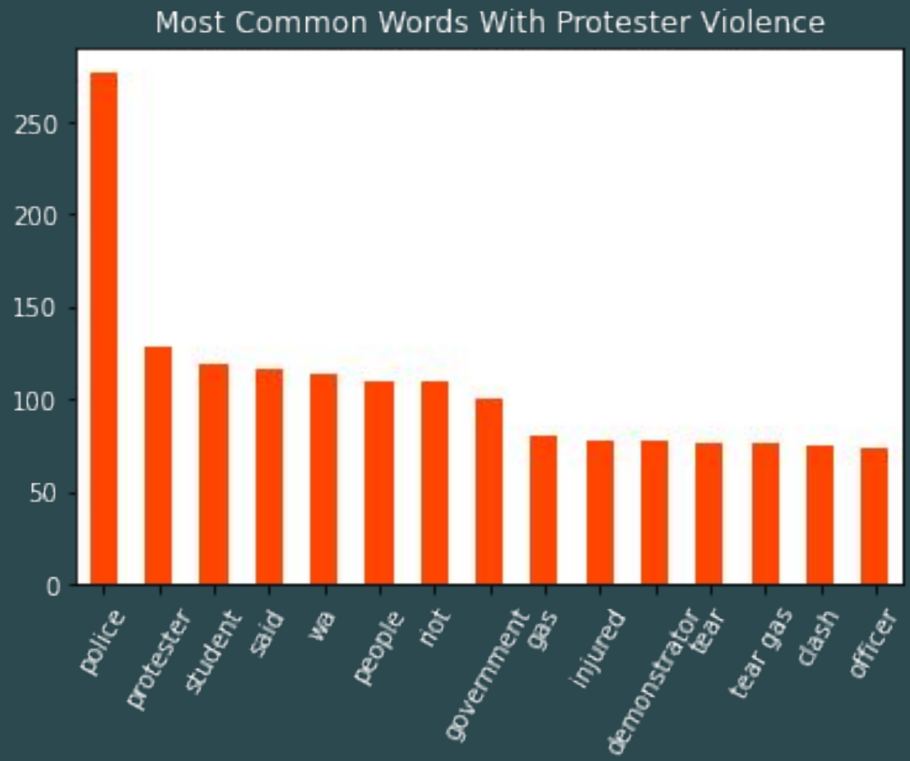


Text Cleaning

- Notes: Text from articles written describing the protest
- Step One: Lemmatize
  - Breaking down each word into its base form.
- Step Two: TFIDF (Word Vectorizer that shows the frequency of each word appearing)
  - Able to use this to get rid of common stop words (all articles were written in English)
  - Able to remove one and two letter words
  - Able to split each word into a seperate column for better summarization

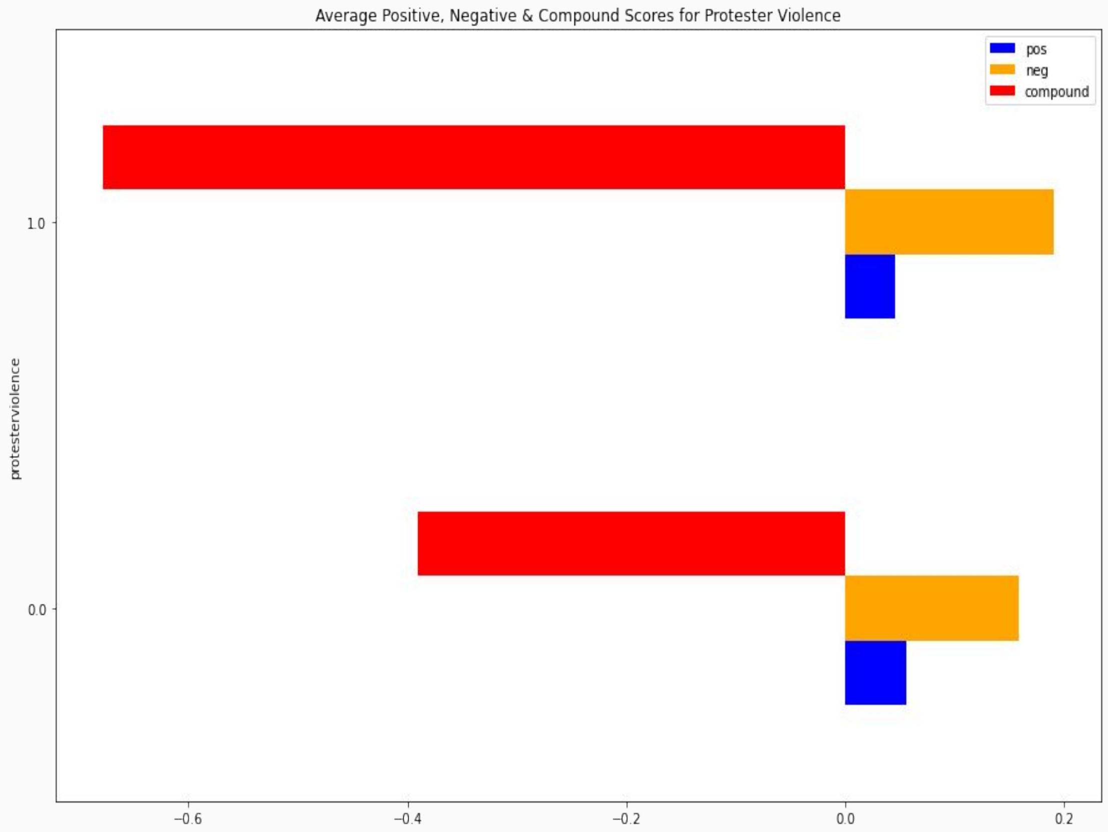
Findings:


Most Common Words w/o Protester Violence
0
50
100
150
200
250
300
government
strike
wa
people
police
said
protester
demonstration
thousand
president
student
yesterday
general
city
day


Most Common Words With Protester Violence
0
50
100
150
200
250
police
protester
student
said
wa
people
riot
government
gas
injured
demonstrator
tear
tear gas
clash
officer

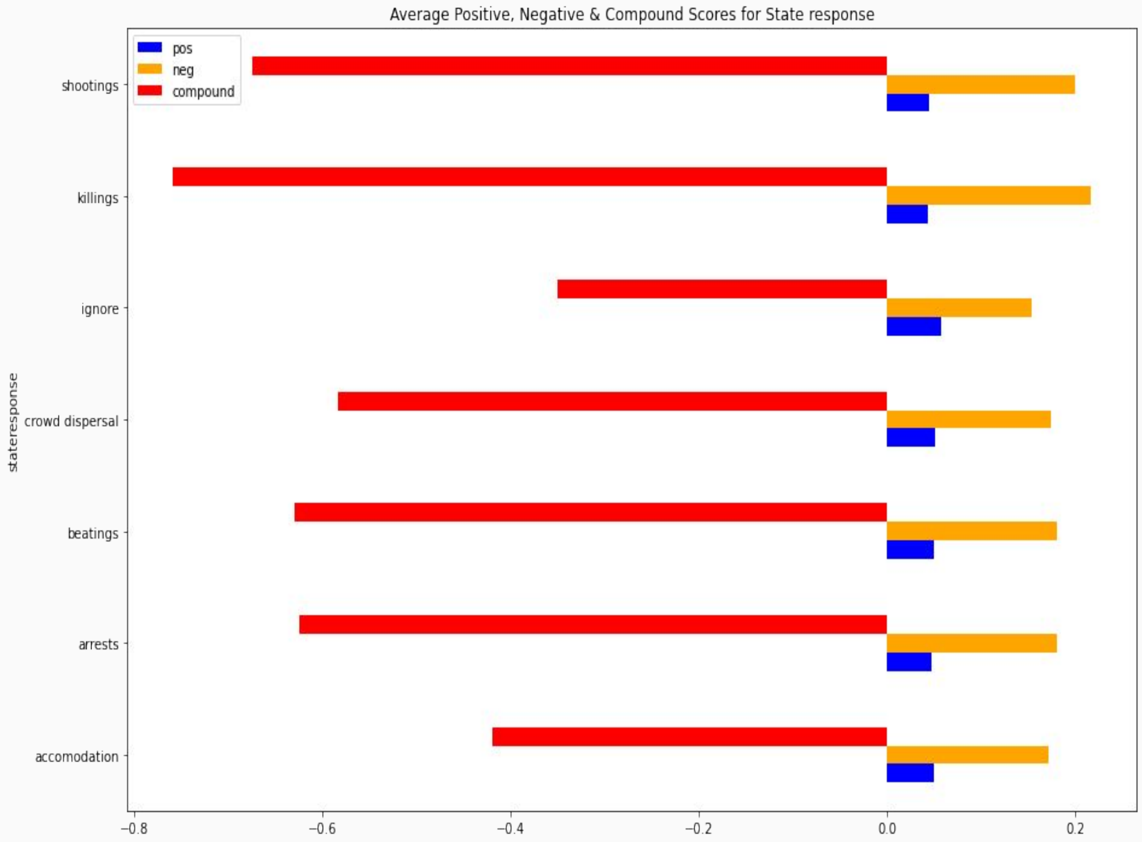
Sentiment Analysis using VADER

Adjacent chart shows sentiment analysis for text against target variable protester violence. For more violent protests it shows most negative sentiment while for less violent protest shows less negative sentiment.



Sentiment Analysis using VADER

Doing same analysis for variable State Response it shows maximum negative sentiment for more violent state response like killing , shootings , beatings while less negative sentiment for less violent state response like ignore or accomodations.

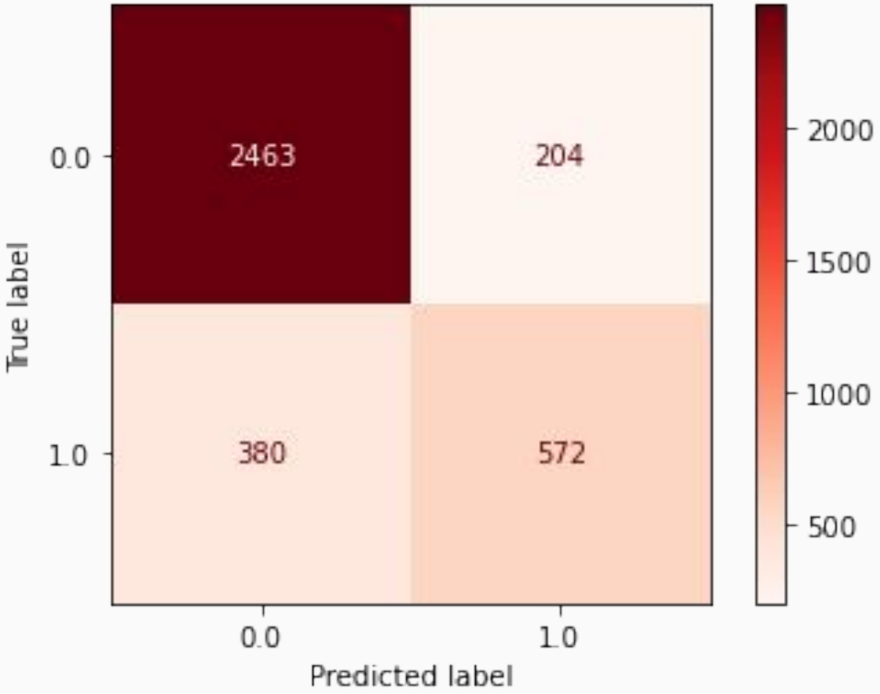


Modeling with text and numeric features using Adaboost

Followed by text EDA we did modeling for text and numeric data using ensemble model Adaboost combined with DecisionTreeClassifier for target variable protesterviolence, which gave below results :

Train Score : 0.86 Test Score : 0.83

Specificity : 0.92 Sensitivity : 0.60



Modeling with text And numeric features using SVM

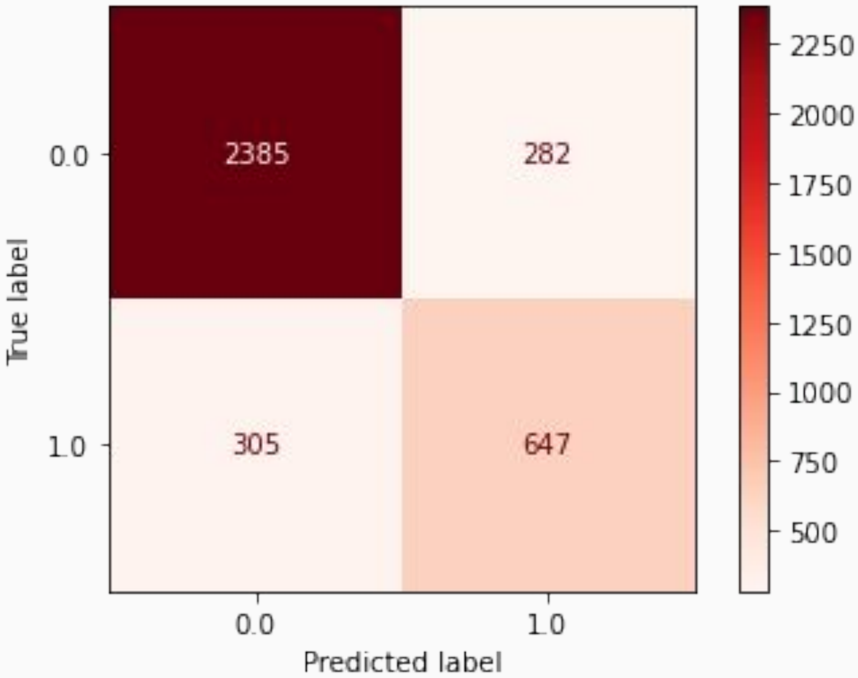
We also decided to model text and numeric data using SVM to see if we get better results.

SVM model resulted with :

Train Score : 0.94 Test Score : 0.83

Which is a overfit model with

Specificity : 0.89 Sensitivity : 0.67



Text Model Comparison

- Both the model show imbalance in specificity and sensitivity due to imbalanced data.
- Model with Adaboost shows better results as the model is not highly overfit.
- SVM model is overfit but shows better sensitivity score.
- Future recommendations for model with text would be to try to balance specificity and sensitivity.

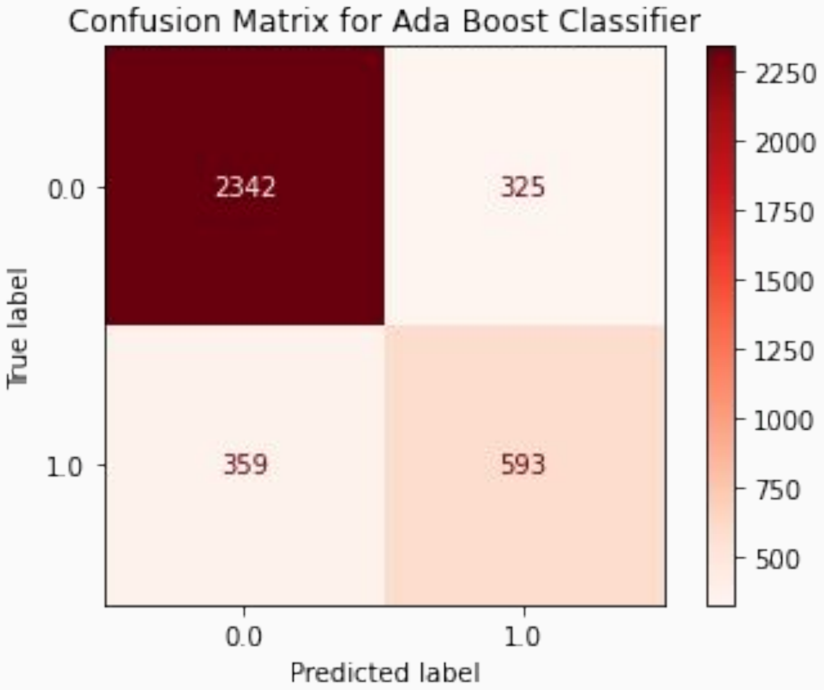
Modeling without text features

Though the text descriptions were useful in describing violent and non-violent protests, we were interested to see what numeric and categorical features were most predictive of protester violence without a natural language description of the event.

Ada boost classifier was chosen for its predictive and explanatory power.

Our best model after grid search produced:

- Train score: 85.6%
- Test score: 81.1%



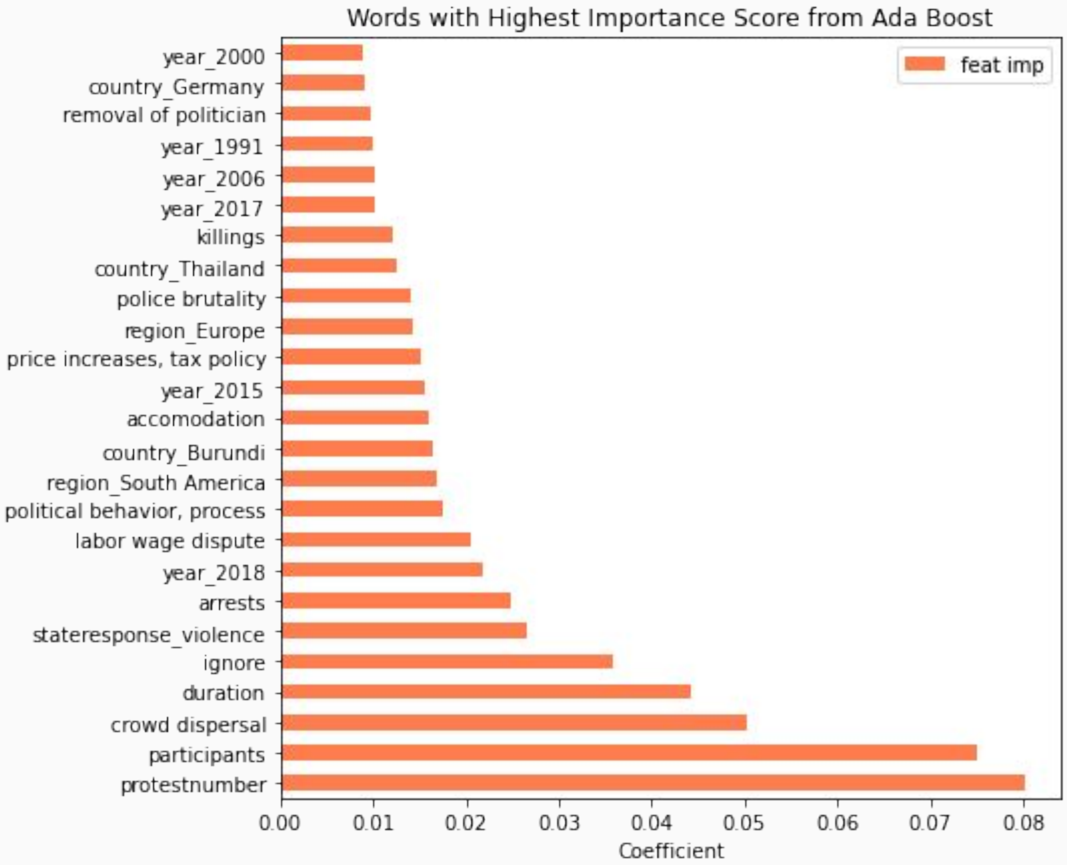
Important Features

Examining the Importance Scores from the Ada Boost model gives insights into the features most useful for modeling.

Protest number, participant number, and duration all ranked highly, indicating the length and size of the protest can affect the chance of violence in the outcome.

Several state controlled responses ranked highly as well: crowd dispersal, choosing to ignore the protest, issuing arrests, and providing accommodation.

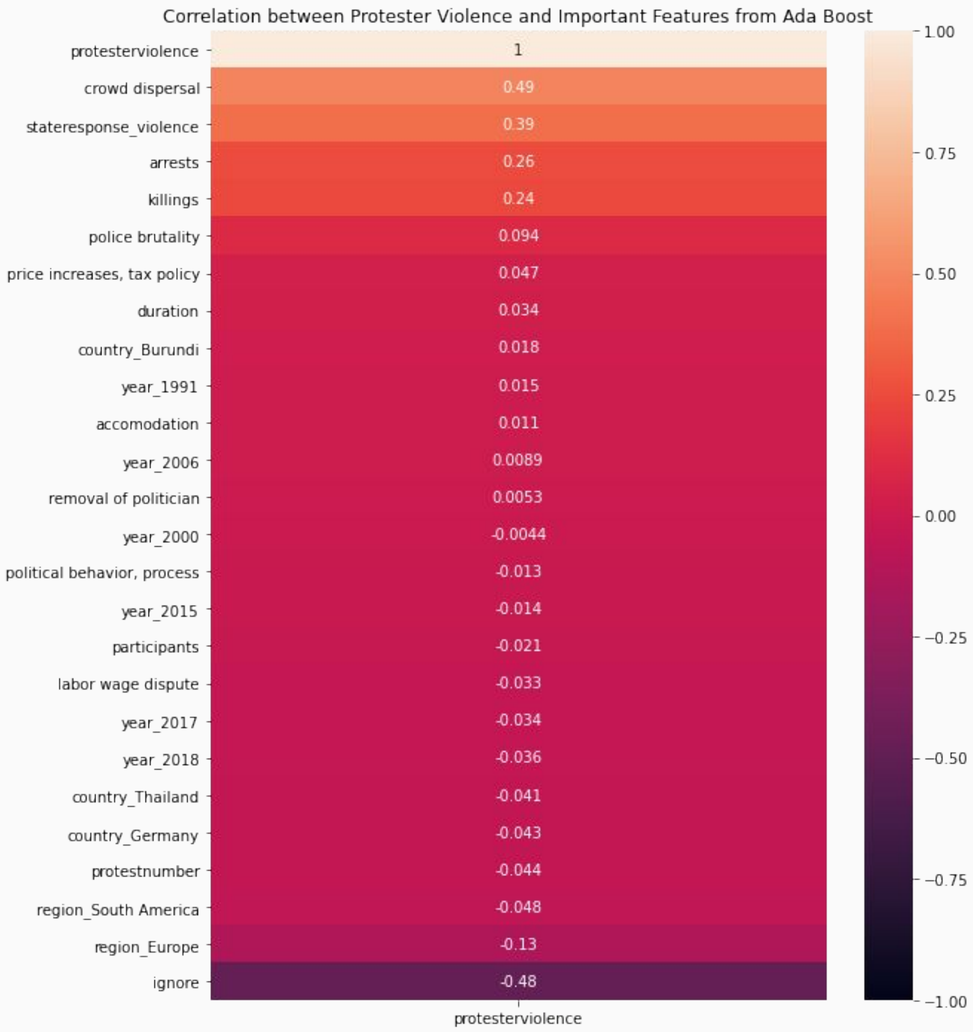
Many protester demands, specific years, and regions showed up as well.



Important Feature Correlation with Violence

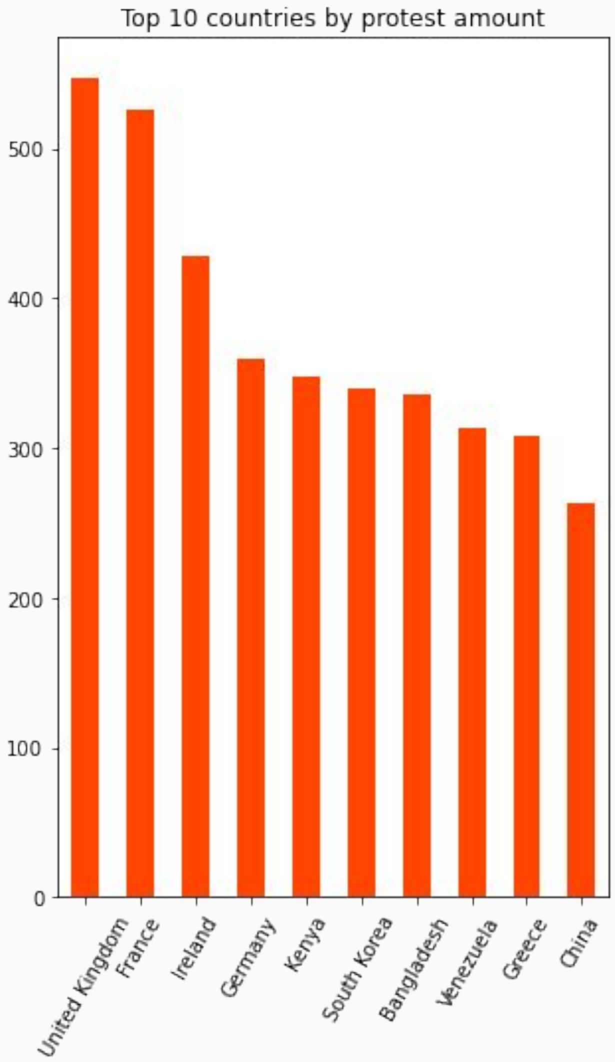
A correlation matrix revealed the directionality of the relationship between the most predictive variables and the target. Crowd dispersal, arrests, and state sponsored killings had the highest positive correlation with protester violence.

While these analyses do not prove causality, it is important to note that choosing to ignore the protest was the feature with the most negative correlation with protest violence. Europe may also be a model for strategies to ensure peaceful demonstration.



Protests by the number

Which countries and regions have the largest number of protests?



UK

1990: Poll tax riots across the UK, largely in Trafalagar Square

2016: Decision to withdraw from the EU. AKA: Brexit

France

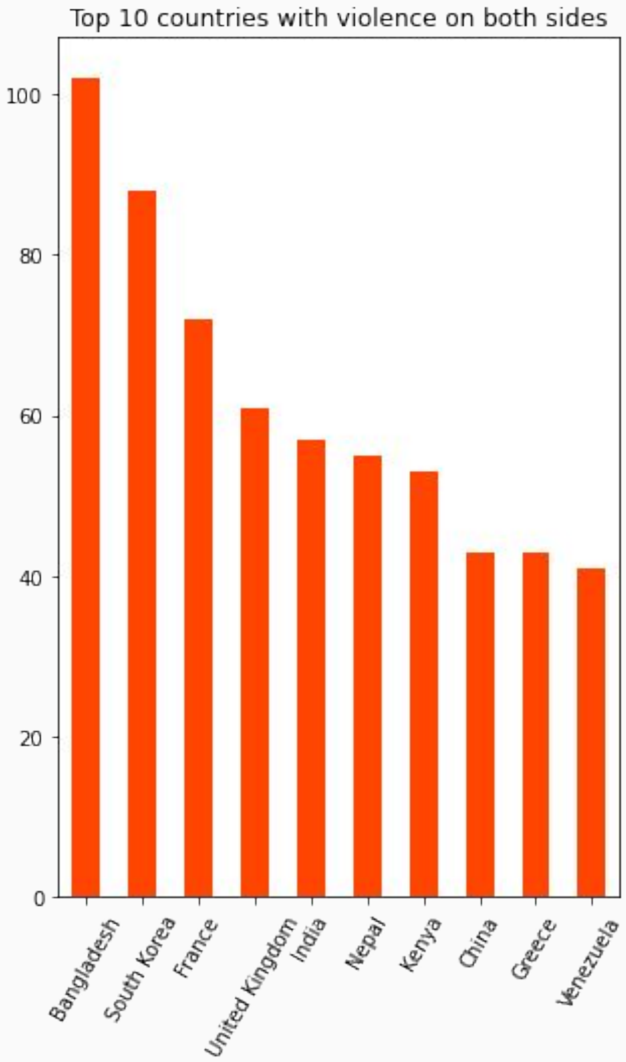
2016: Controversial labor laws to a 48 hour work week

2018: The yellow protests against a fuel increase

Protests by the number

Which countries have the most violence on both sides?

40% correlation between state violence and protester violence.



Bangladesh:

42% State violence

50% Protester Violence

30% Violence on both sides

Top regions:

Africa

Asia

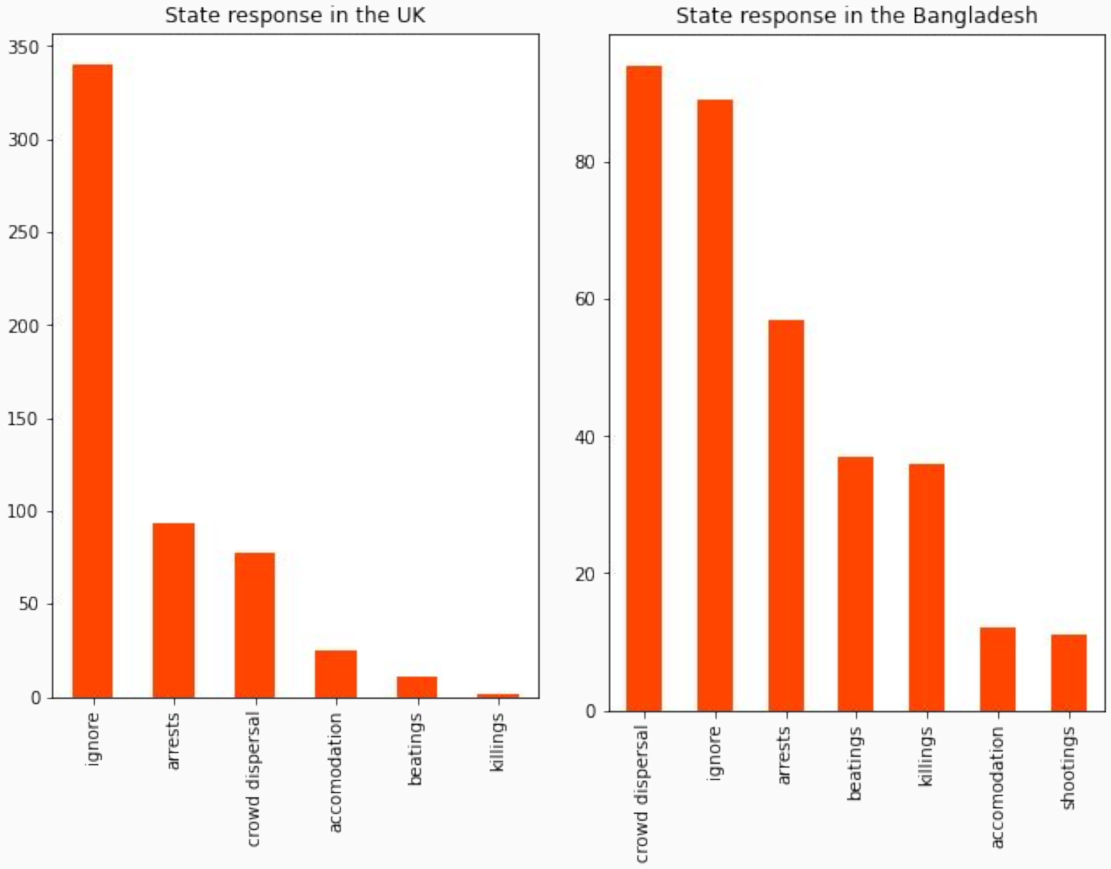
Europe

State response by region

How do different countries law enforcement respond to protests?

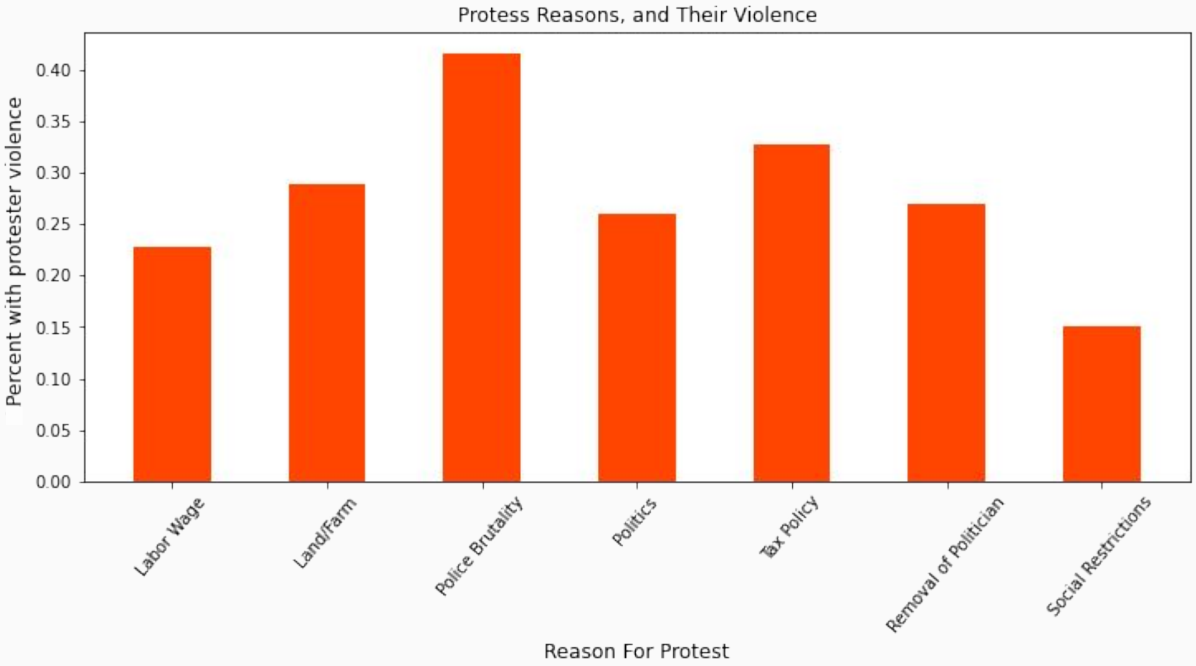
“Crowd dispersal mechanisms: any attempt to move the protesters from their location and break up the protest… Examples might include the use of tear gas, issuing warnings, moving troops into positions and pushing protesters off their positions.”

-Mass Mobilization Project



The Cause and its effect.

Which social causes bring the most aggressive action?



Recommendations

- The strong correlation tells us that violence brings violence.

- Pay attention to the type of protest, in order to minimize violence.
  - Larger protests are less violent
  - Protests regarding police brutality are the most volatile

- Prepare in advance.
  - The majority of protests happen on the 1st or the 15th of the month
  - Countries with the most protester violence: France, Bangladesh, South Korea
  - Countries with the least: Switzerland, Serbia, Czechoslovakia