

# PROTESTS

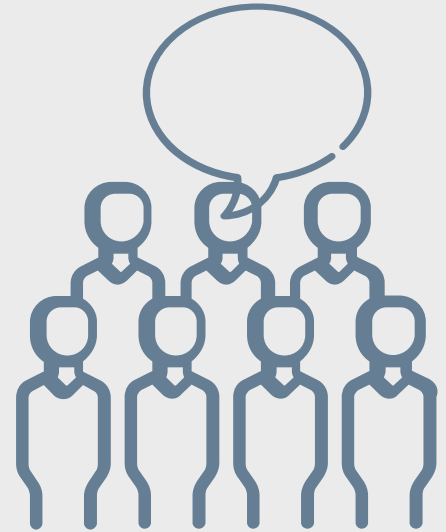
## A Data Science Approach

Group Project 5, DSI Cohort #1019

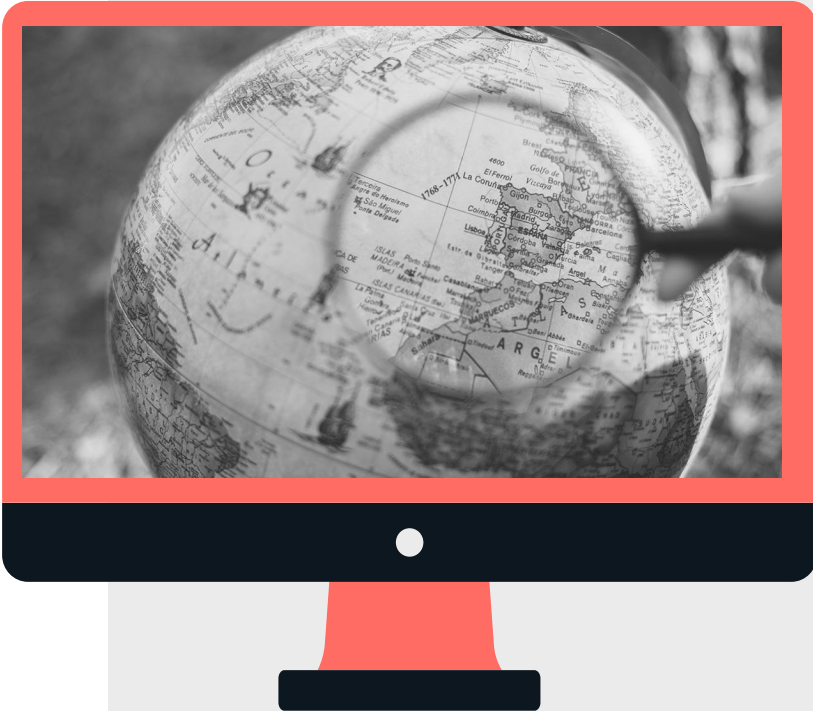
*Group Members:*  
Stephen Strawbridge  
Nicholas Kovacs  
Young Park  
Chris Burger

# Presentation Contents

1. Problem Statement/Background
2. Data Organization/Engineering
3. Exploratory Data Analysis (EDA)
4. Models Used
5. Conclusion/Recommendations



# Problem Statement



Protests groups are *not* reaching full potential

Use prediction models to guess government response with reasonable confidence

# Background

**Data Source** – Mass Mobilization Dataset

**Data Features** – demands, responses, location, brief notes on each protest, etc

# Data Organization & Engineering

## Key points

- Engineer *protest length* feature
- Dummify protester *demands* and government *responses*
- Quantitative vs Qualitative (notes) features





# Data Organization & Engineering (cont.)

Our \*7 target variables – *gov. responses*

1. Ignore
2. Accomodation
3. Arrests
4. Beatings
5. Crowd dispersal
6. Killings
7. Shootings

\*...Or *consolidate*  
target variables

# Data Organization & Engineering (cont.)

Our \*4 consolidated variables

1. Ignore
2. Accomodation
3. Adverse Reaction (arrests, crowd dispersal)
4. State Violence (beatings, shootings, killings)

# Protests from around the world



A protester waves a red flag in front of anti-riot police in Rome in October 2011. Tens of thousands marched as part of a global day of protests inspired by Occupy Wall Street.



Amid fires set around Independence Square, known as the Maidan, in Kyiv, Ukraine, protesters demonstrate against the government of then-President Viktor Yanukovich in February 2014.



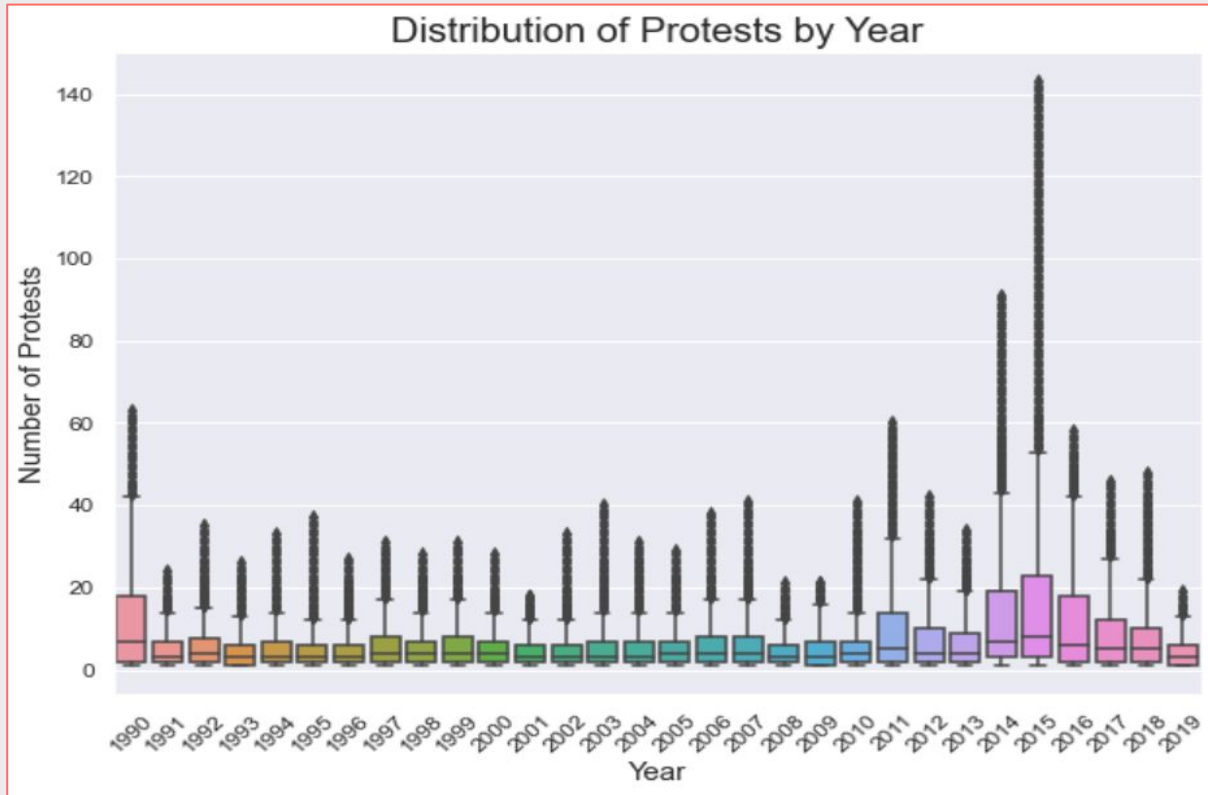
In November 2015, a man walks amid hundreds of shoes in Paris, left to protest climate inaction after marches were banned following terrorist attacks.



Iraqi protesters join hands after taking part in prayers during anti-government demonstrations in the Shiite holy city of Najaf, in November 2019.

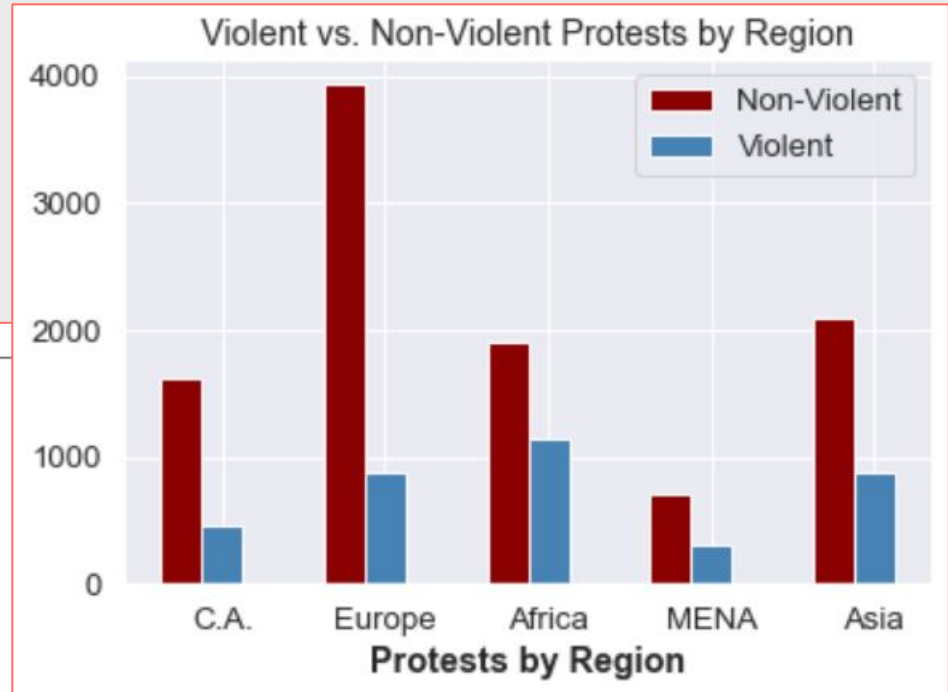
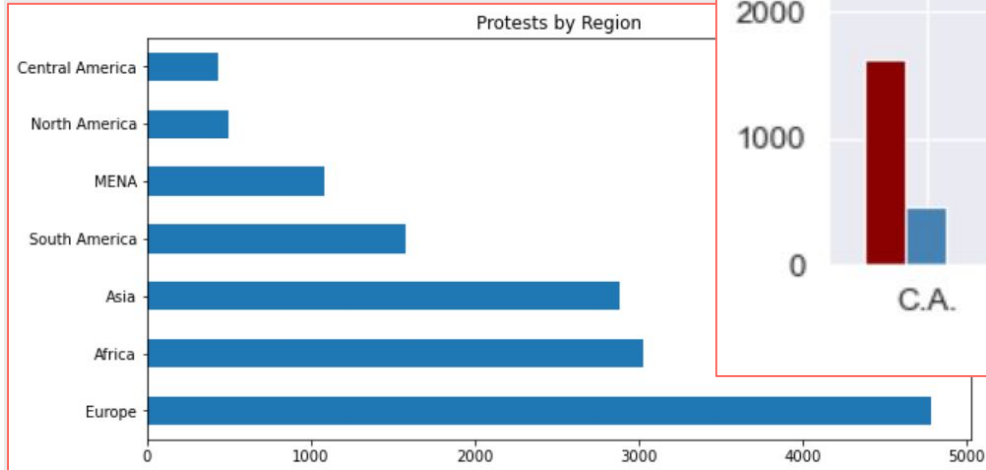


# EDA - Protests over Time(worldwide)

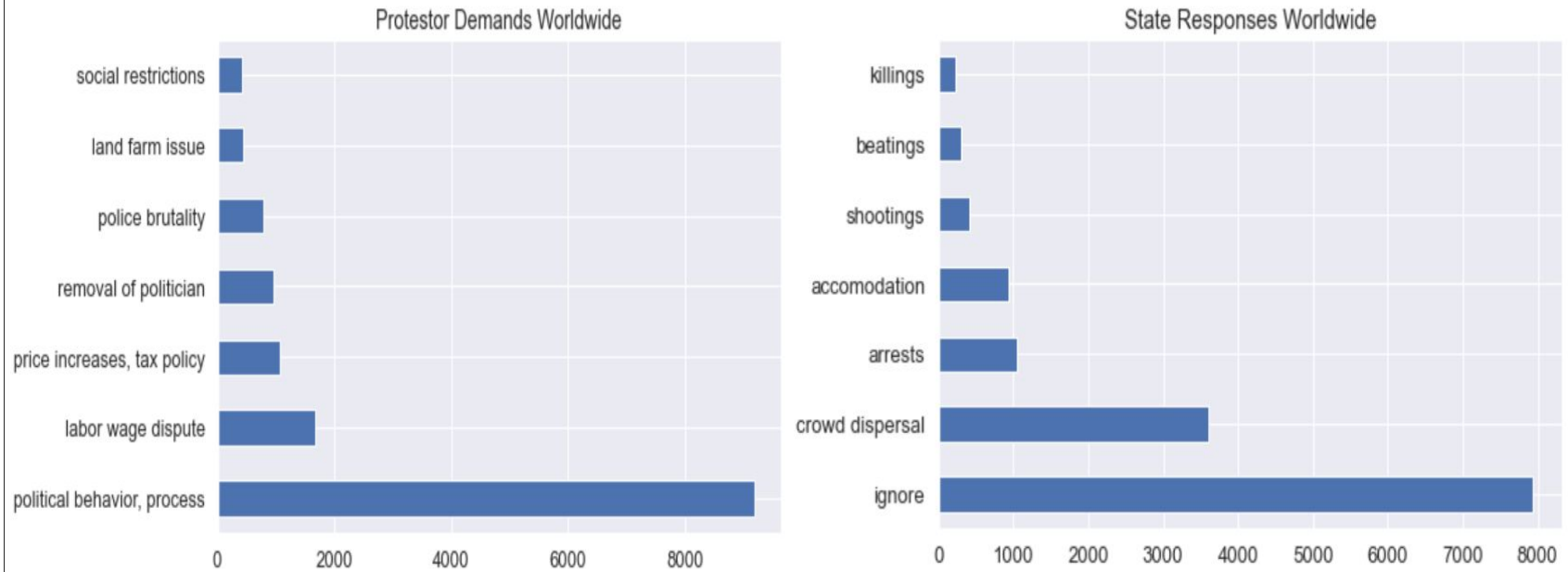


# EDA (cont.)

Protests by Region and the amount of protests by year (Worldwide).

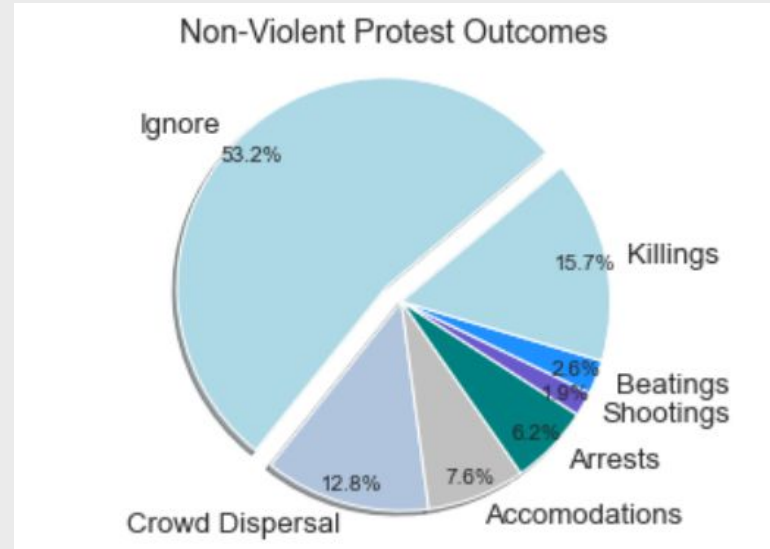
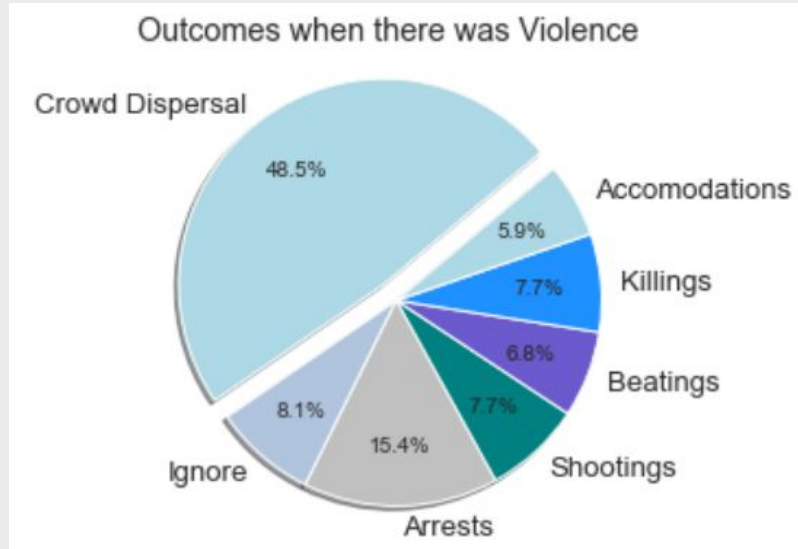


# EDA (cont.)



State responses, worldwide, vs. the protesters' demands.

# EDA (cont.)



The differences in state responses/outcomes to violent vs. non-violent protests can be seen here. It's quite interesting to note the enormous difference in the protest being ignored when it is peaceful vs. not (along with other interesting comparisons).



**4 Models  
for  
7 Regions**

# Models Used

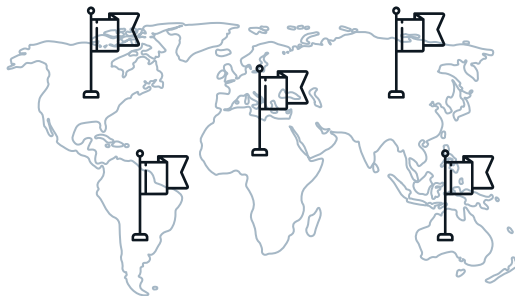


# 1. Random Forests

Baseline Score = 83.56%

Training Score (mean) = 98.23%

Testing Score (mean) = 88.75%

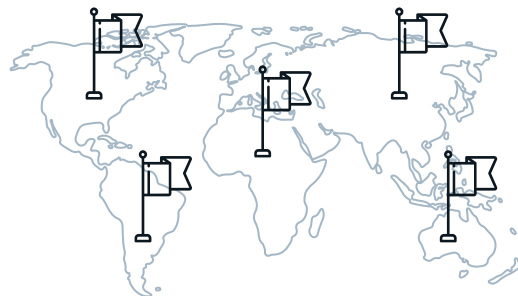


## 2. Support Vector Classifier (SVC)

Baseline Score = 83.56%

Training Score (mean) = 98.36%

Testing Score (mean) = 88.79%



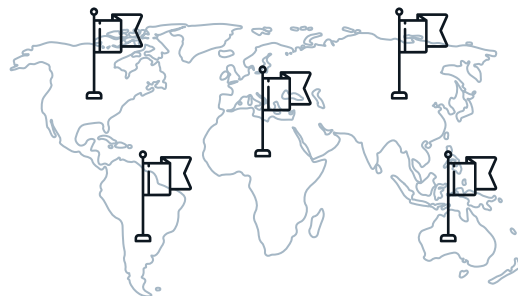
# 3. Logistic Regression

Baseline Score = 83.56%

Training Score (mean) = 97.84%

Testing Score (mean) = 88.28%

Lasso and Ridge both effective



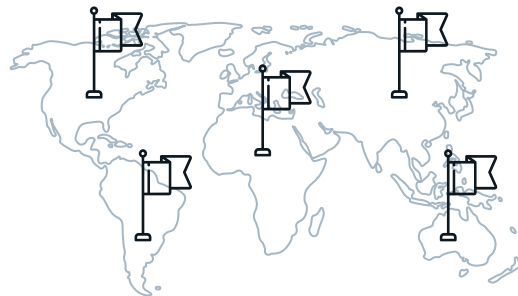
## 4. XGBoost

Baseline Score (7 reponses) = 83.56%

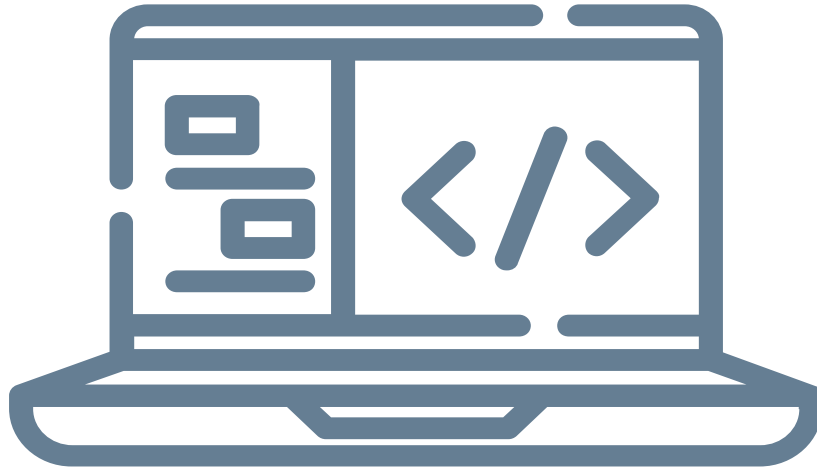
Baseline (4 target responses) = 74.5%

Training Score (mean) = 97.02%

Testing Score (mean) = 89.15%



# Model Demonstration!





# Conclusions/Recommendations

Baseline scores can be hard to beat - specifically the *violent* responses

...But we can predict with confidence (e.g. beat the baseline) for most other responses!

**Successful Models**

Logistic Regression, XGBoost

# Thank you

Questions?



# Sources

- Mass Mobilization Dataset:
- <https://massmobilization.github.io/visualization.html>
- Slides template: <https://slidesgo.com/>
- <https://www.npr.org/sections/pictureshow/2019/12/31/790256816/the-2010s-a-decade-of-protests-around-the-world>