



**MIRIAM BROWN,
MERON DESTA,
KATIE SALARIS,
ROBERT SIMON,
GARETT TAFFEY**

RESEARCHING THE INTERNET RESEARCH AGENCY

AUGUST 2019

WHY THIS?



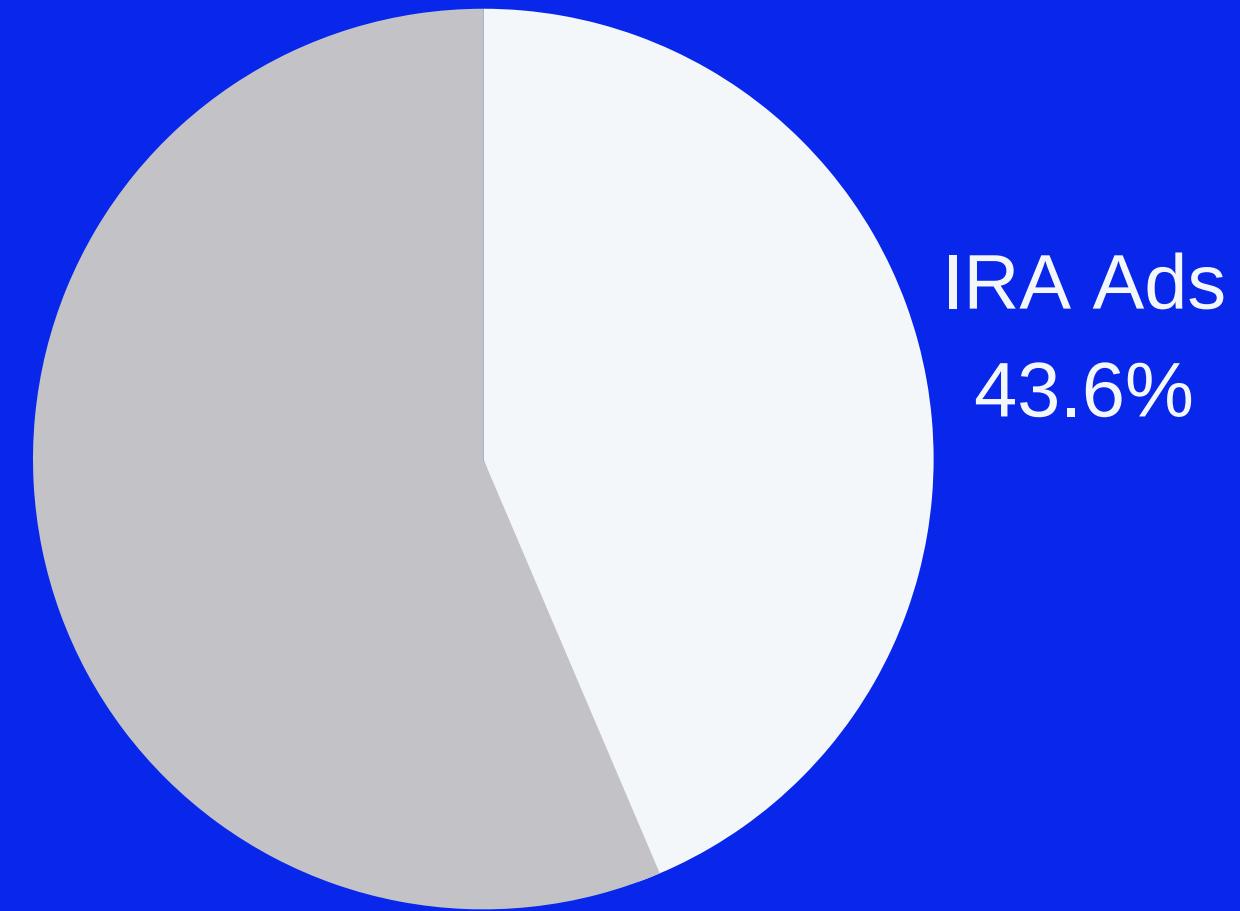
Moscow’s influence campaign followed a Russian messaging strategy that blends covert intelligence operations—such as cyber activity—with overt efforts by Russian Government agencies, state-funded media, third-party intermediaries, and paid social media users or “trolls.”

MUELLER REPORT

Our Data

- 1 HOUSE INTELLIGENCE COMMITTEE
- 2 FACEBOOK AD LIBRARY

Facebook Ad Library
56.4%



Machine Learning

CAN WE DISTINGUISH THE 'FAKE' ADS?



CLEAN DATA

Eliminated records where ads did not have any text and corrected parsing errors

LANGUAGE PROCESSING

Removing stop words and punctuation and lemmatize words (reduce to base root)

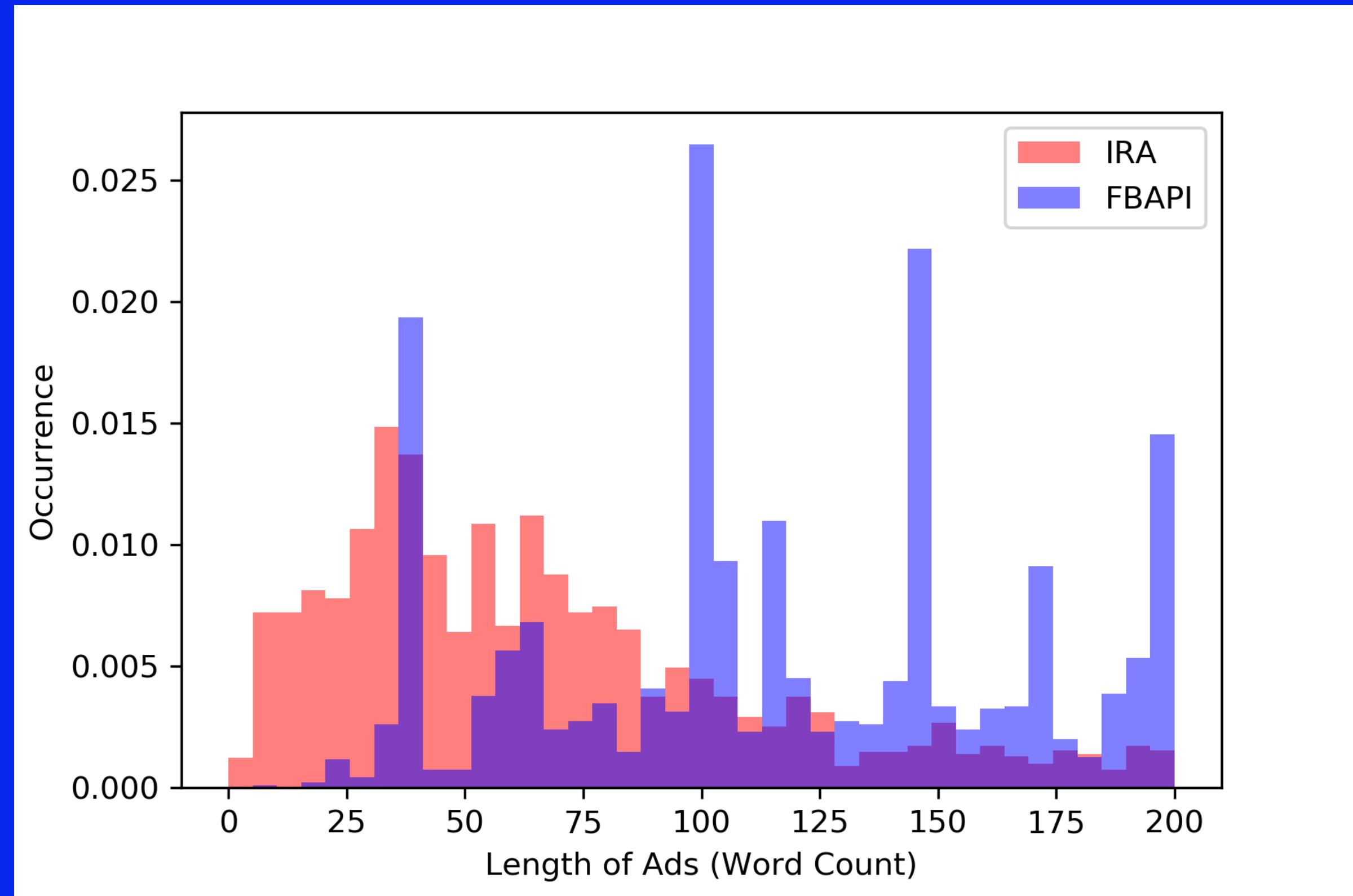
TF*IDF VECTORIZATION

Translate words into numbers (vectors) to prepare for analysis through the machine learning model

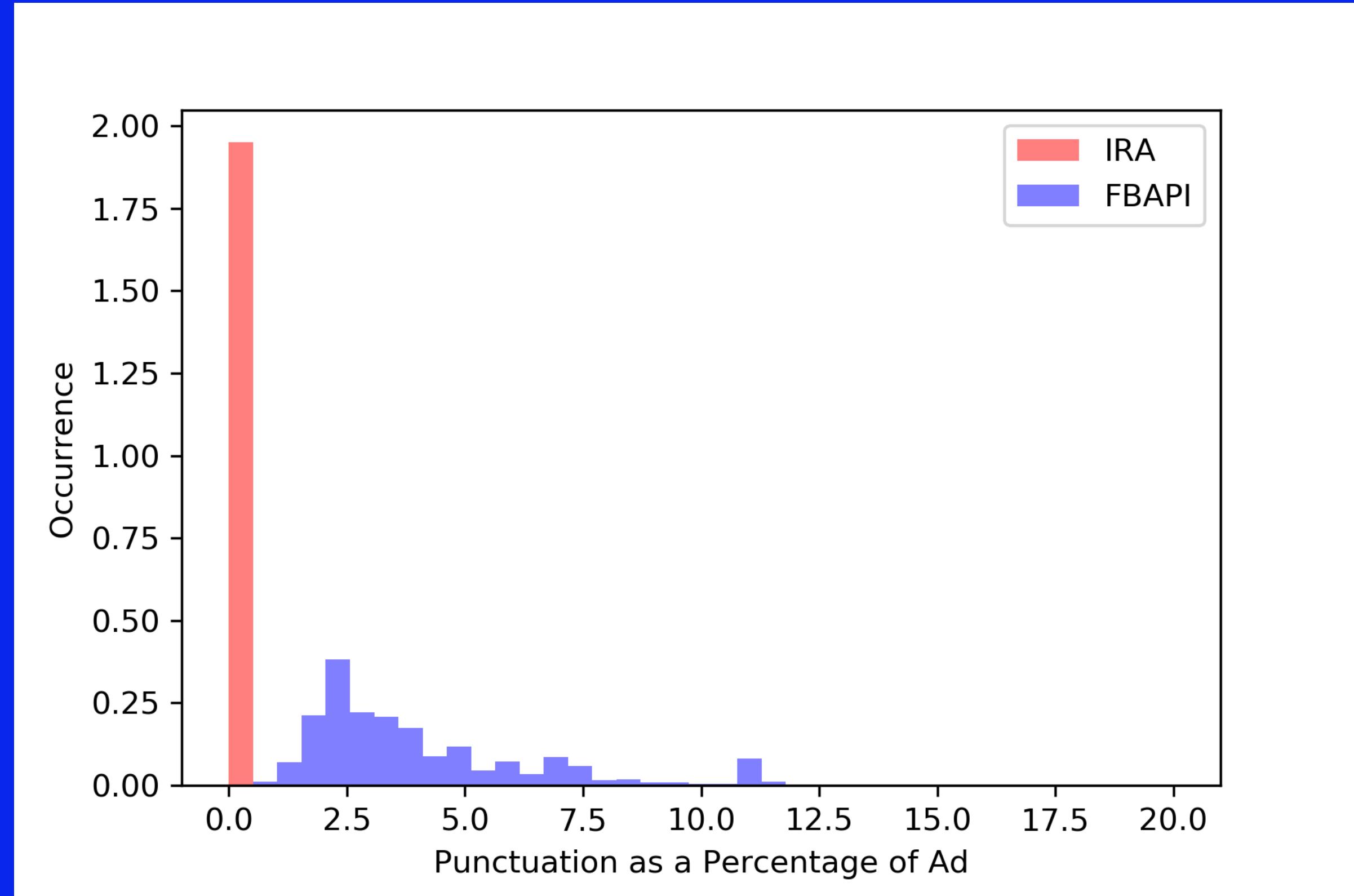
RUNNING IT THROUGH THE MODEL

Used Random Forest Classifier to evaluate vectorized data

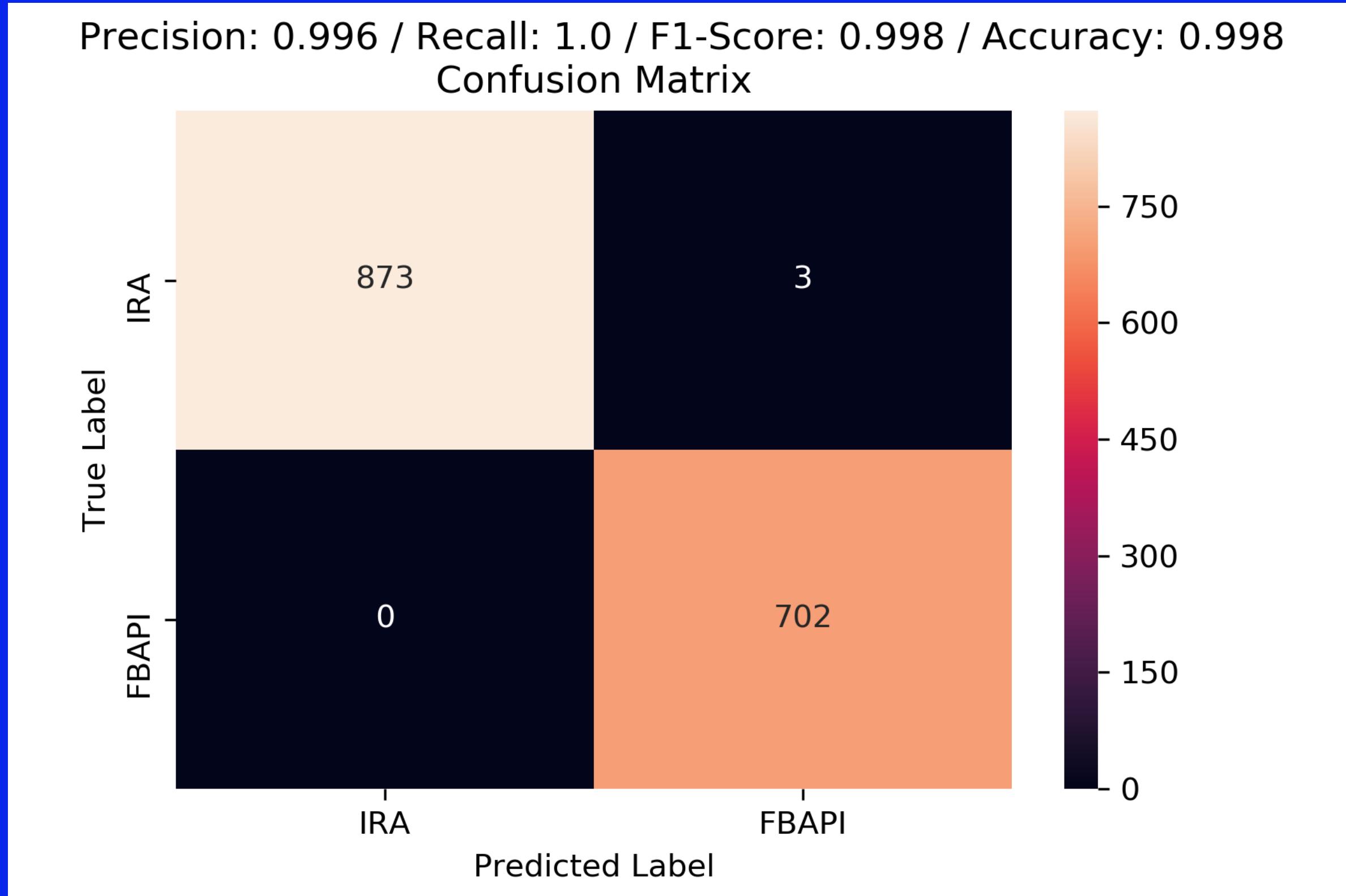
What do our data
sets tell us?



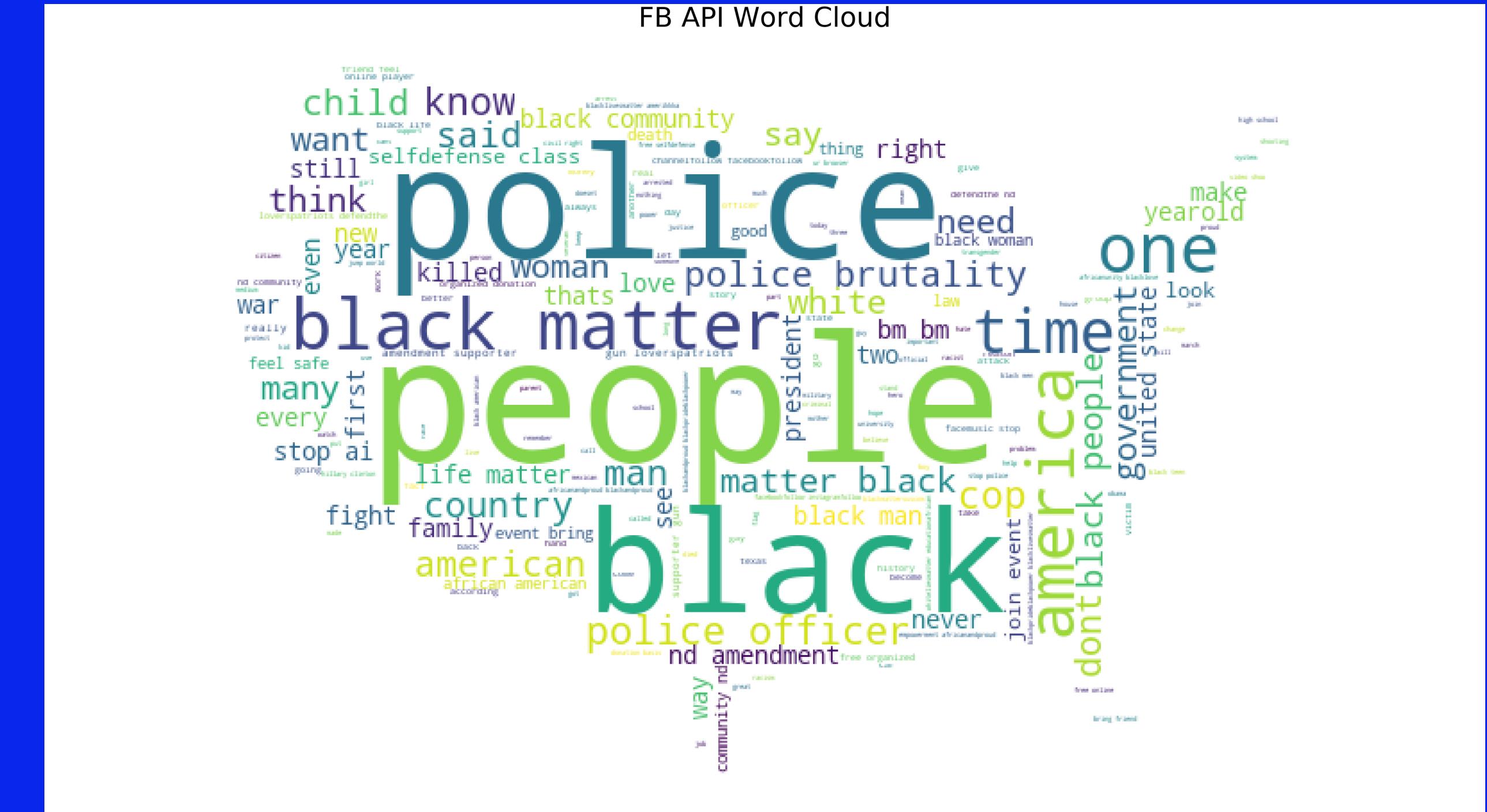
What do our data
sets tell us?



How does our
model perform?

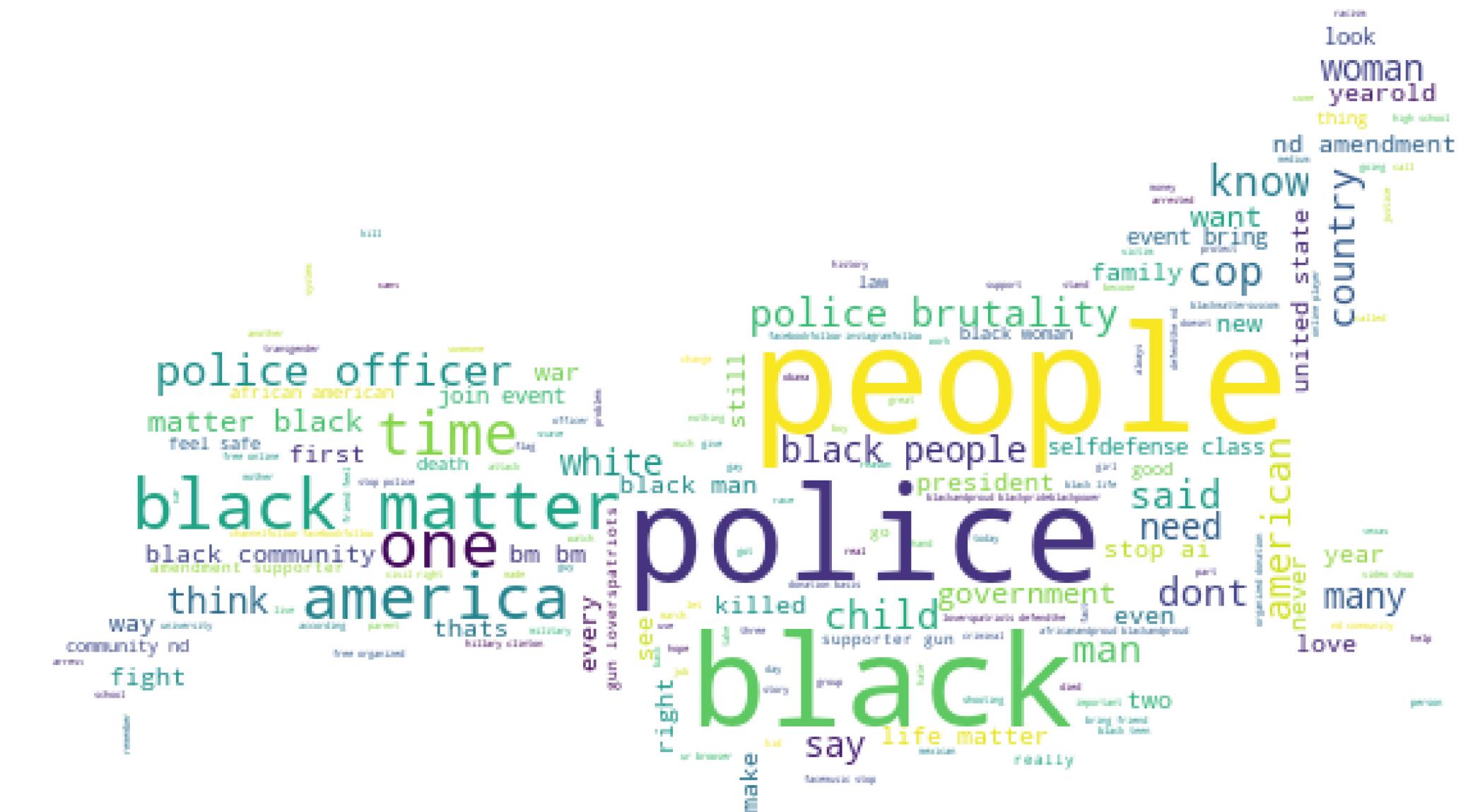


What's in the
ads?



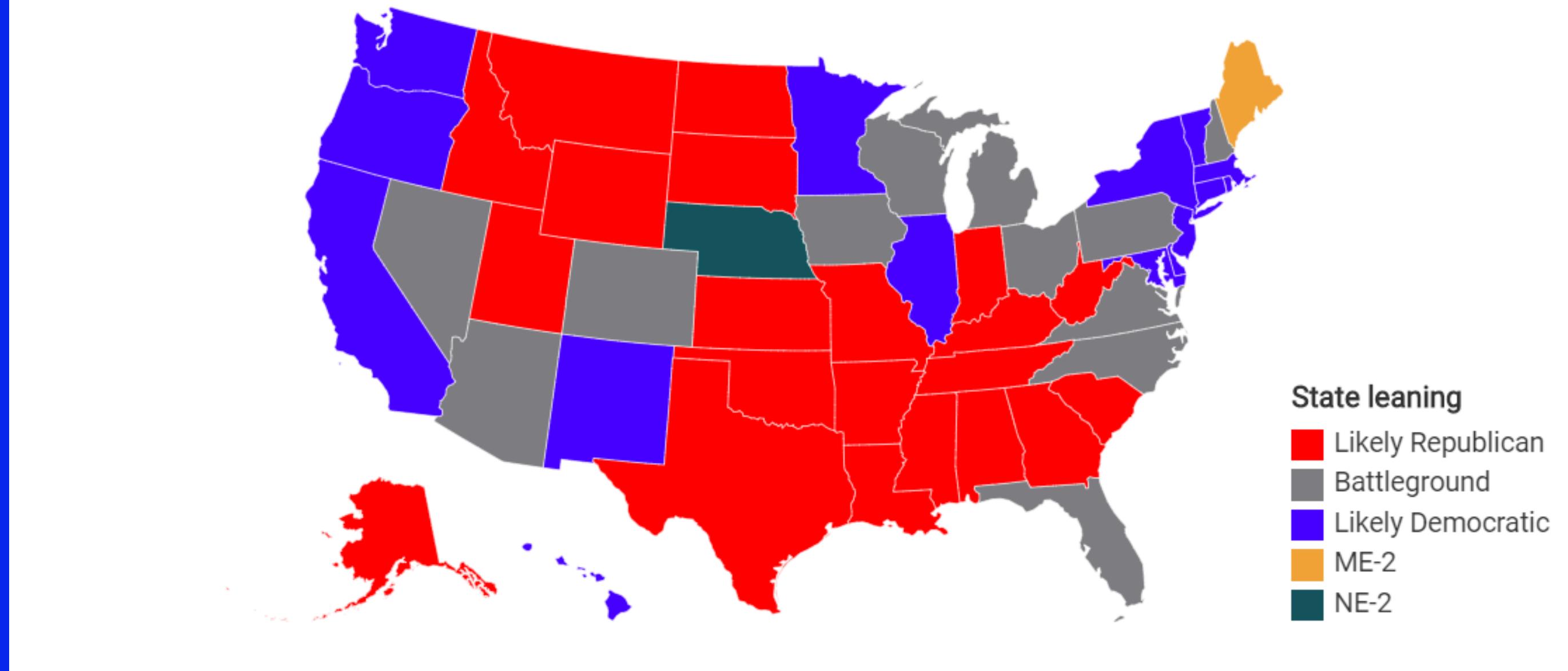
What's in the ad?

Russian FB Ads Word Cloud



Was there a
correlation?

2016 Presidential Battleground States



Was there a correlation?



LANGUAGES

- Python
 - Pandas
 - Numpy
 - Glob
 - NLTK
 - SKLearn
- javascript

VISUALIZATIONS

- Tableau
- Matplotlib
- Wordcloud

WEB DEVELOPMENT

- HTML
- CSS
- Bootstrap
- JQuery
- Flask
- Heroku

VERSION CONTROL/MISC

- Git
- Github
- JSON

Tools Used