

Sentiment Analysis

by: Adam Marianacci

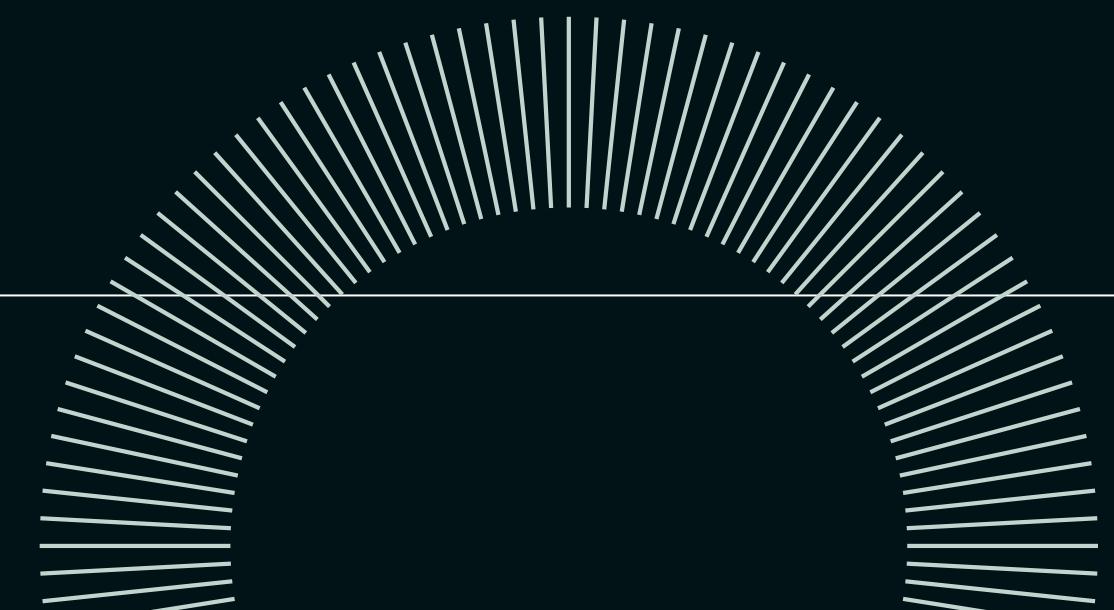
A photograph of a large concert crowd from behind, looking towards a brightly lit stage. The stage is filled with musicians and equipment, and colorful spotlights create a vibrant atmosphere. In the background, the dome of the Texas State Capitol building is visible against a dark sky.

O'KOSH
SXSW 2023

PROJECT GOAL

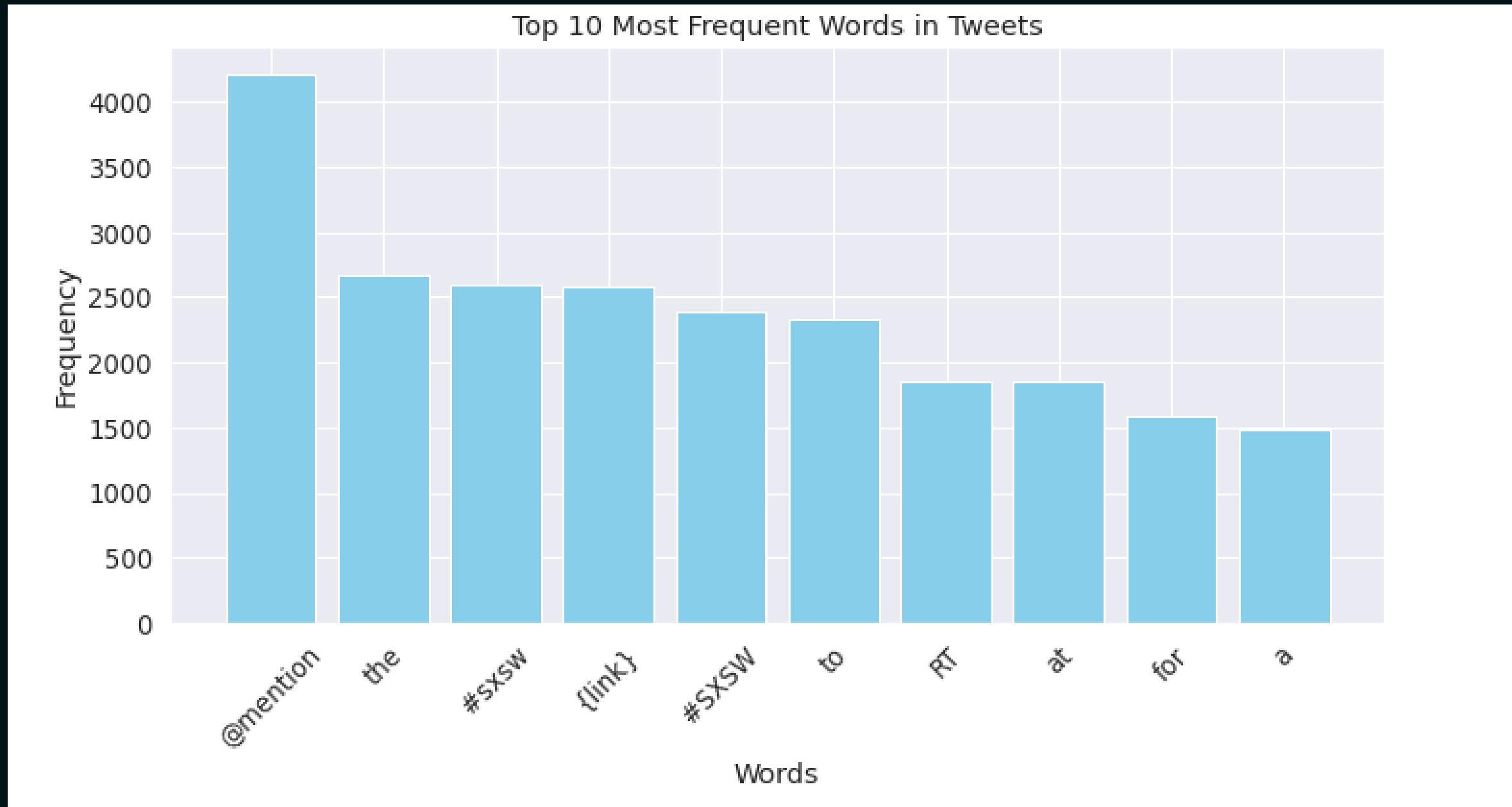


Help SXSW detect positive sentiment in tweets about their event.



DATA

'Brands and Product Emotions" dataset from
CrowdFlower via data.world



Low semantic value

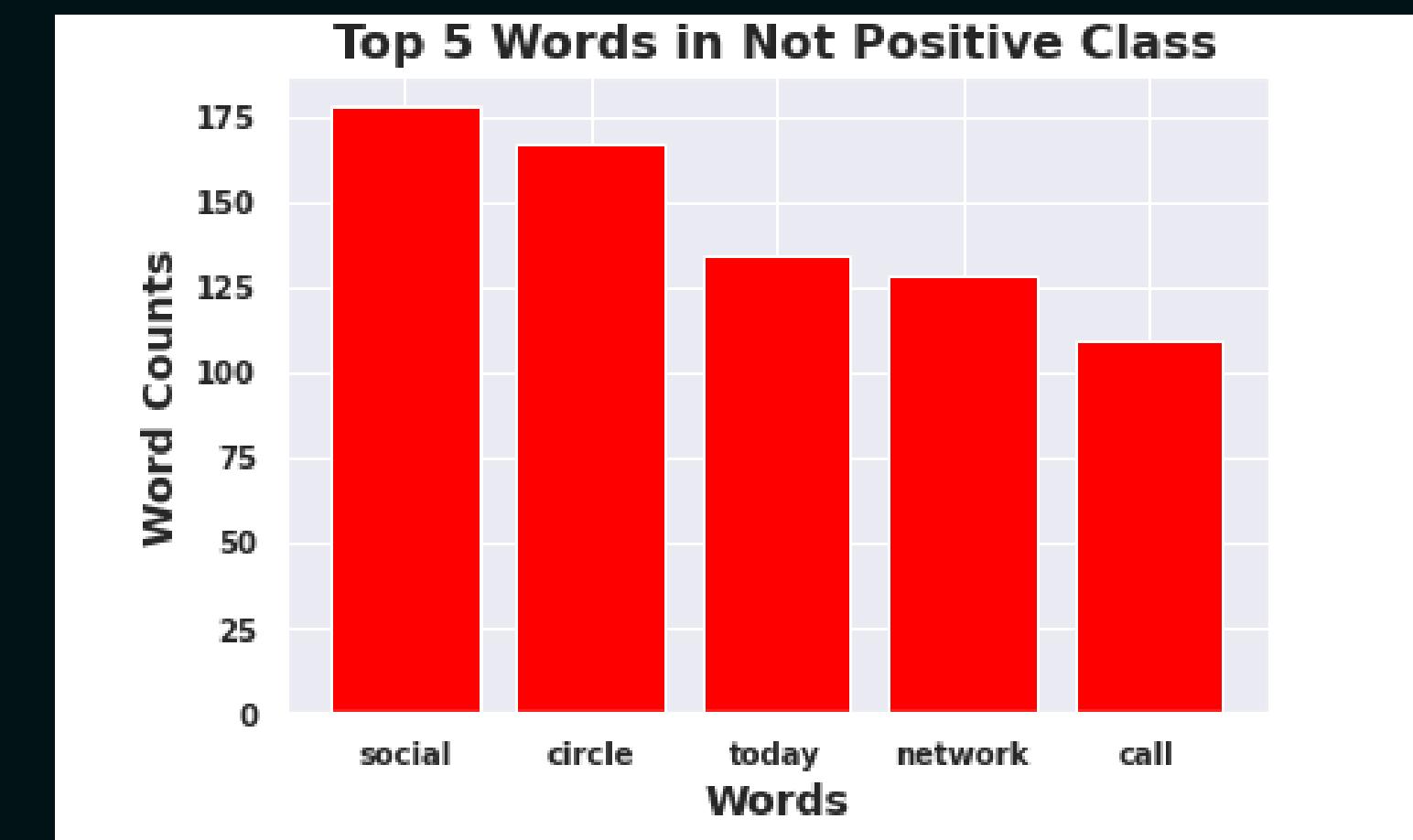
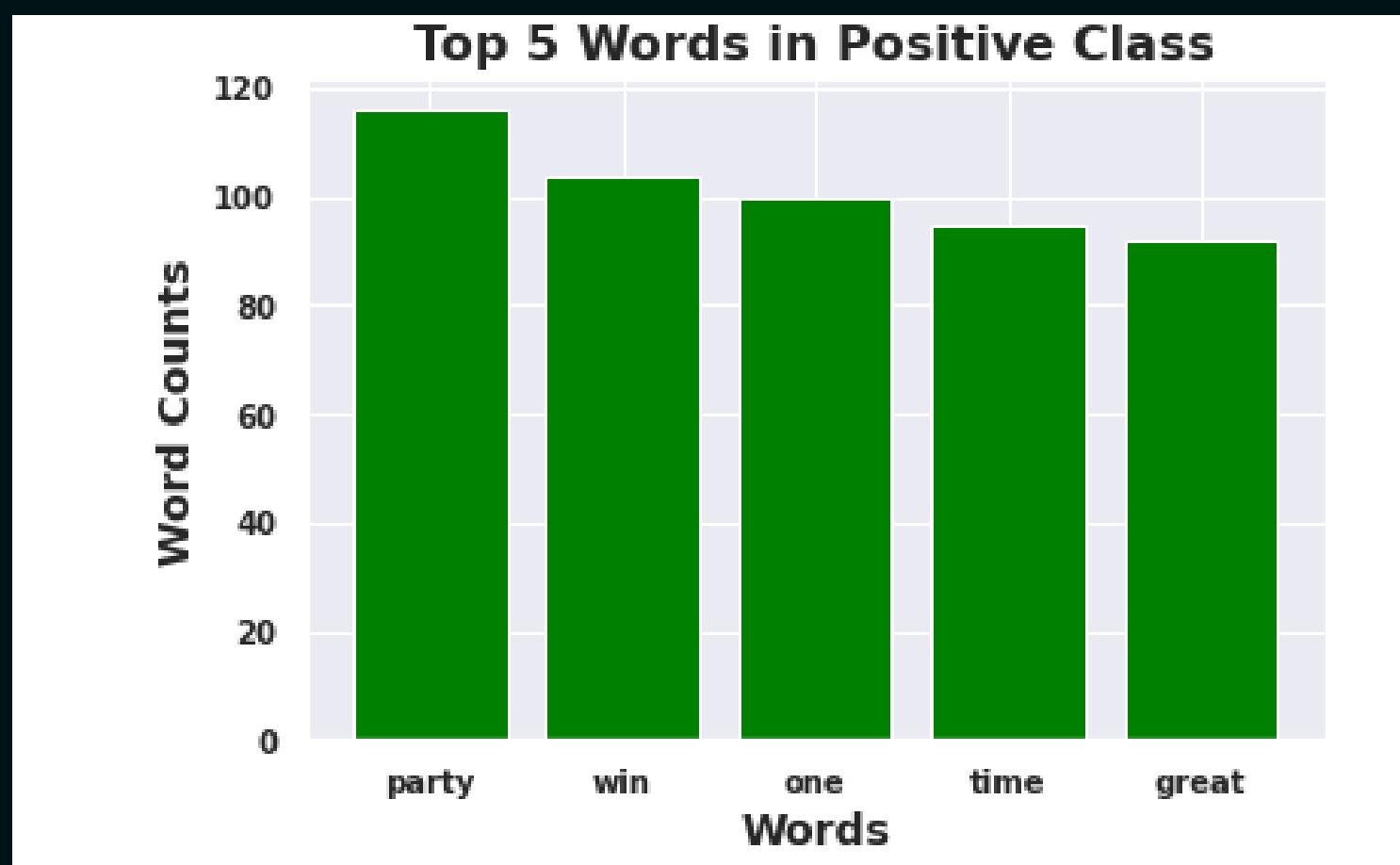
METHODS

- Removed low semantic value words
- Created two classes of sentiment
 - Positive and *Not* Positive
- Converted text to numerical data
- Built models



RESULTS

Words with high semantic value



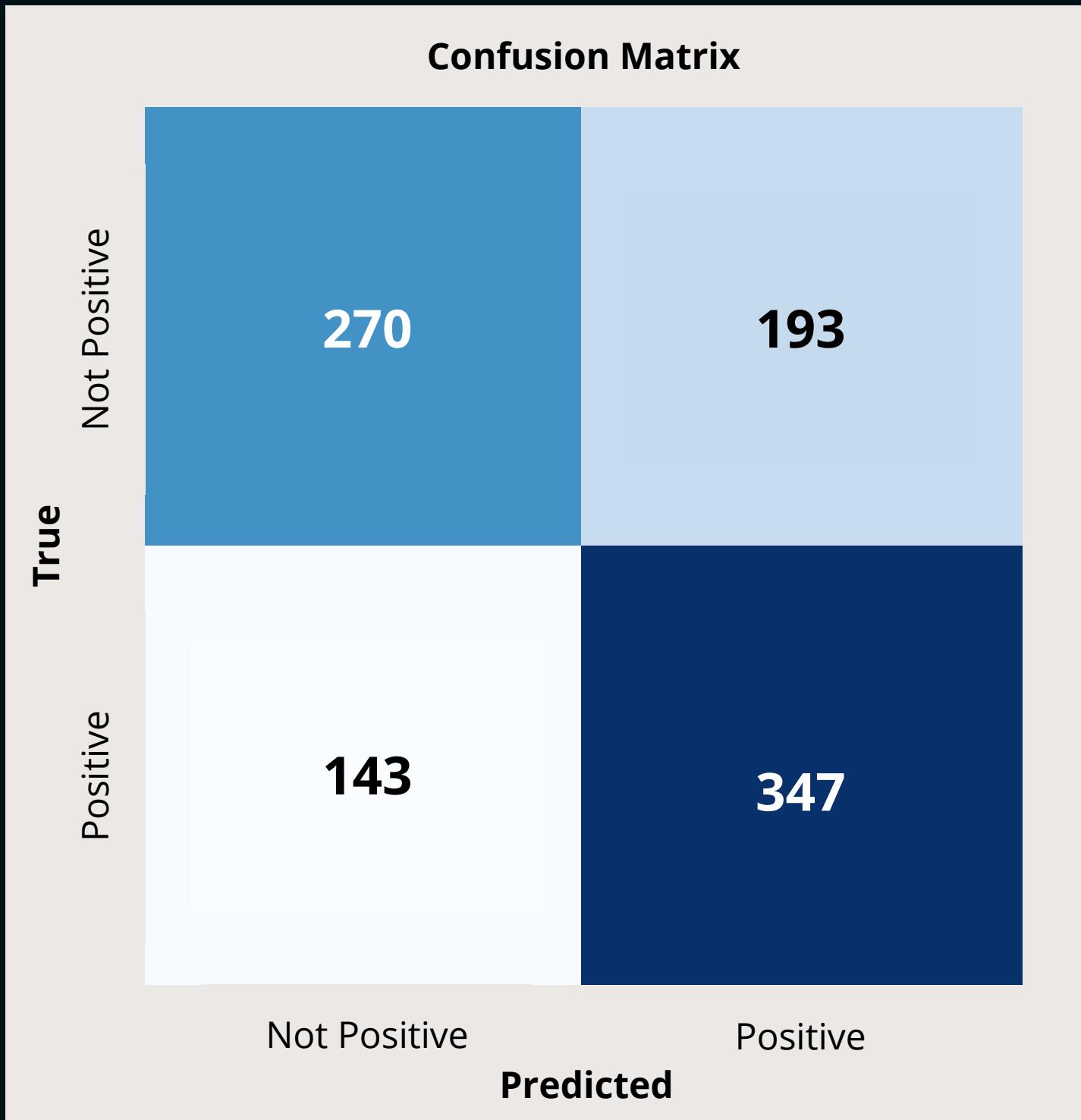
MODEL EVALUATION

- Low feature importance scores
- 65% accuracy score
- Accuracy - How often a model correctly predicts the outcome

$$\left(\text{true positives} + \text{true negatives} \right) \div \text{total number of instances} = \text{accuracy}$$



ACCURACY



- Sample Size - 953
- The model correctly predicted a word as positive or not positive 65% of the time

CONCLUSIONS



Top 5 words in
each class



Need More Data to
Improve the Model



Need to Eliminate
No Emotion
Sentiment



LIMITATIONS

- Not enough Data
 - Negative sentiment
- Features showed to lack importance in models
- Missing values
 - sentiment towards brands



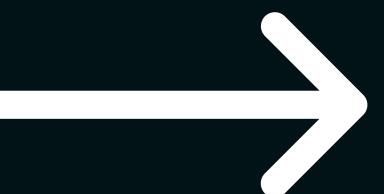
NEXT STEPS



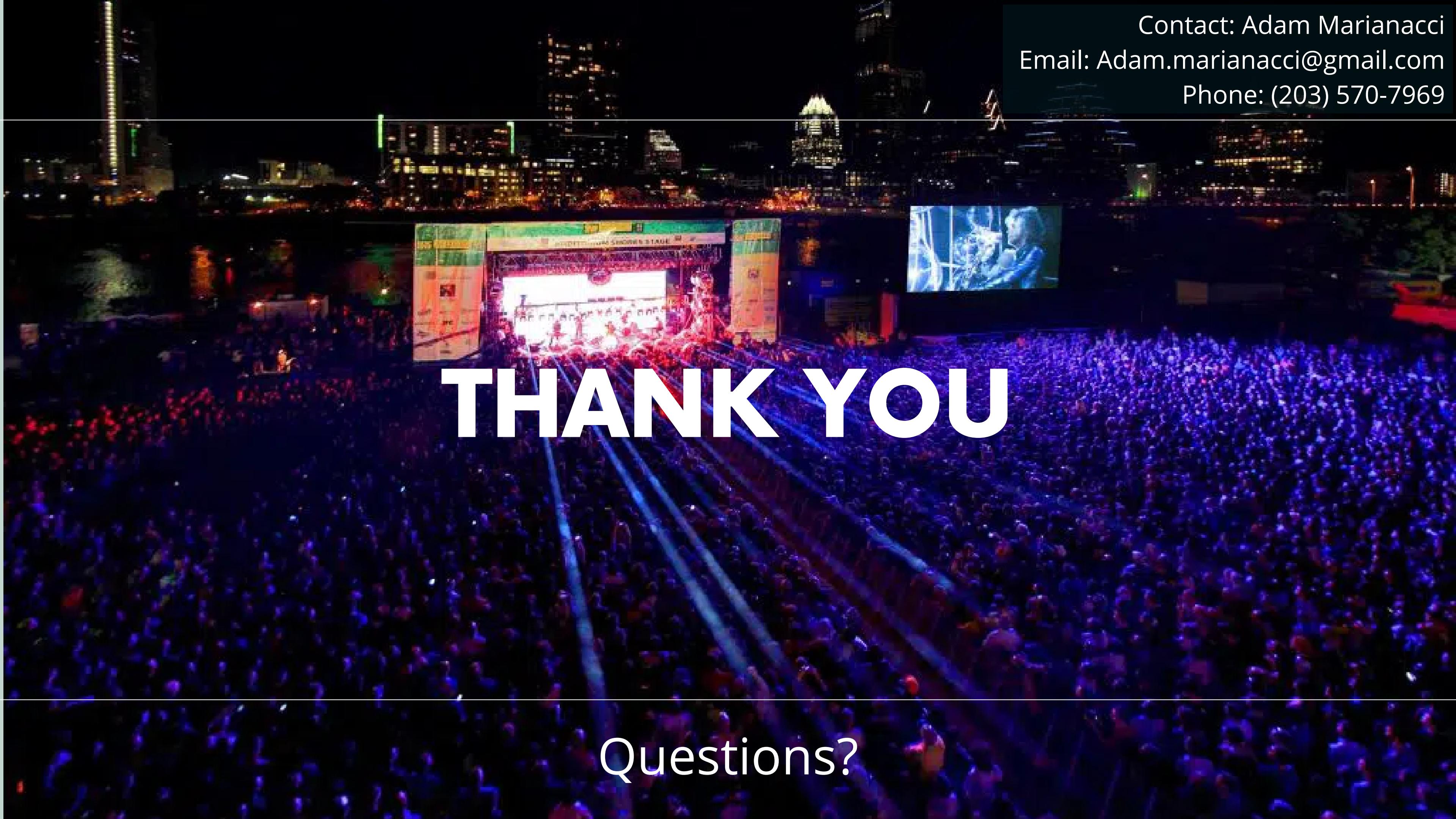
10x more data



Positive and Negative
Sentiment



Brands, film, music



Contact: Adam Marianacci

Email: Adam.marianacci@gmail.com

Phone: (203) 570-7969



THANK YOU

Questions?