



# The Product Sentiment Project



Twitter and Natural Language  
Processing



# Goal

---

To develop a natural language processing model to classify tweets as either negative or positive.

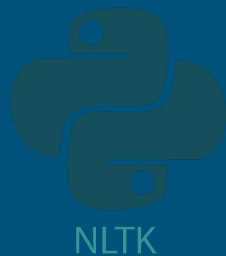
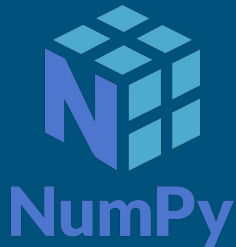
# Business Understanding

---

Companies can benefit from understanding how consumers perceive their brands and products, and sentiment analysis of text data from twitter can help provide this knowledge in a timely manner.

# Methods

---



# Data

---

11,242 rows of text data:

- tweet
- brand/product
- sentiment

# Data Sources

---

**[data.world/crowdflower](https://data.world/crowdflower)**

brands-and-product-emotions data set

apple-twitter-sentiment data set

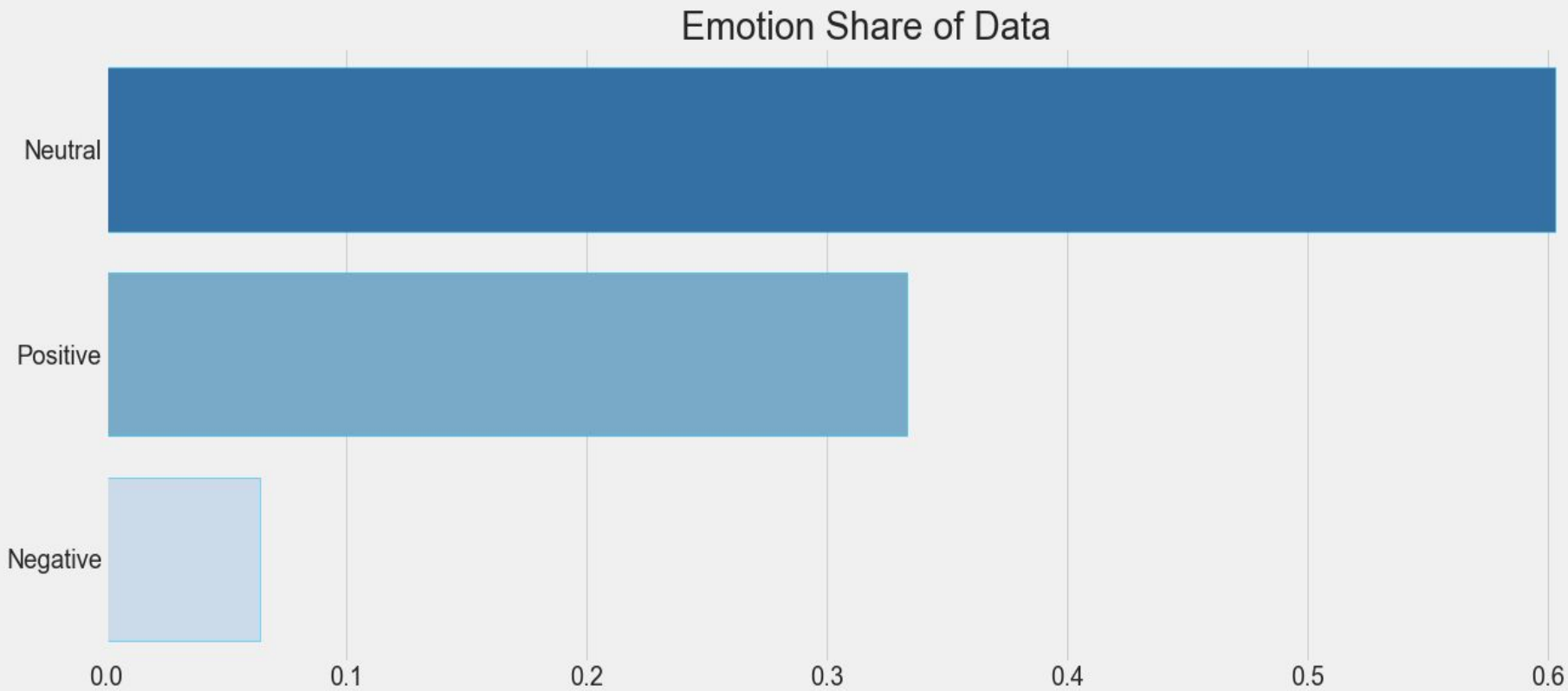
# Data Sources

---

**[kaggle.com/shashank1558](https://kaggle.com/shashank1558)**

preprocessed-twitter-tweets data set

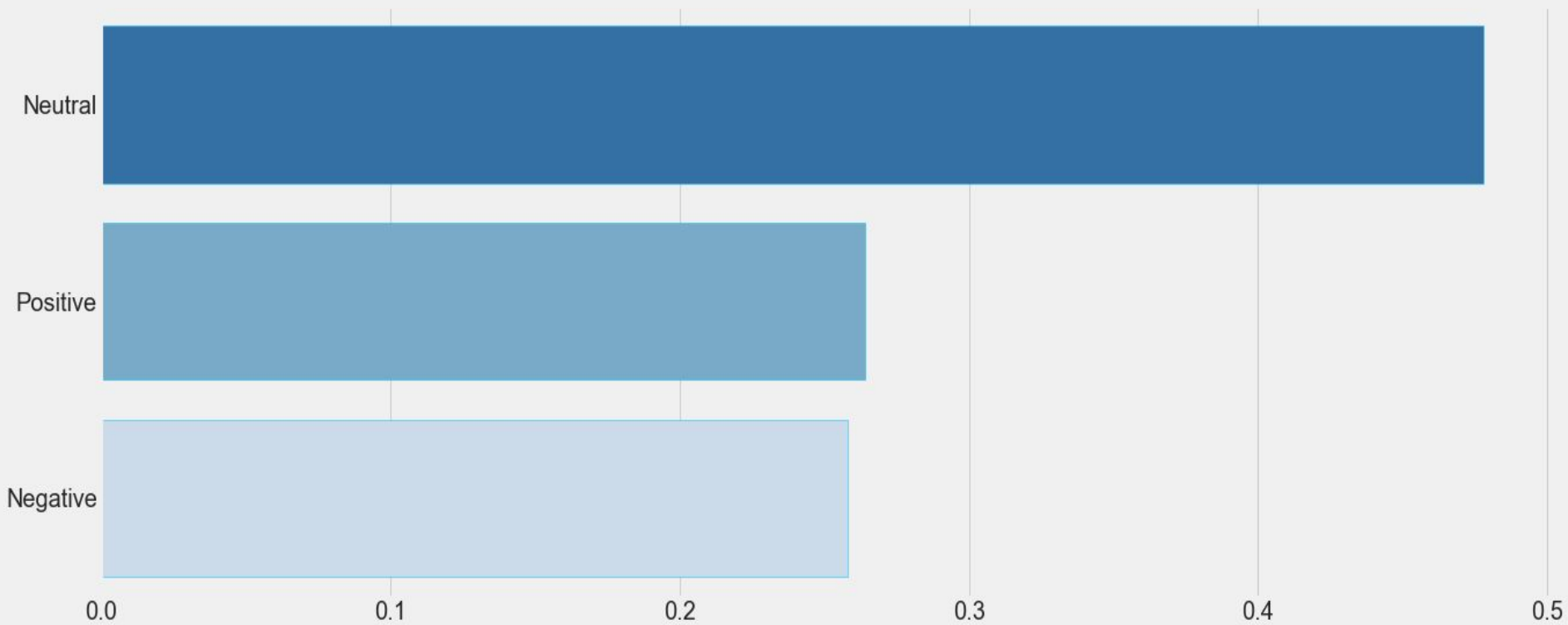
# Original Data Class Balance





# Augmented Data Class Balance

Emotion Share of Data



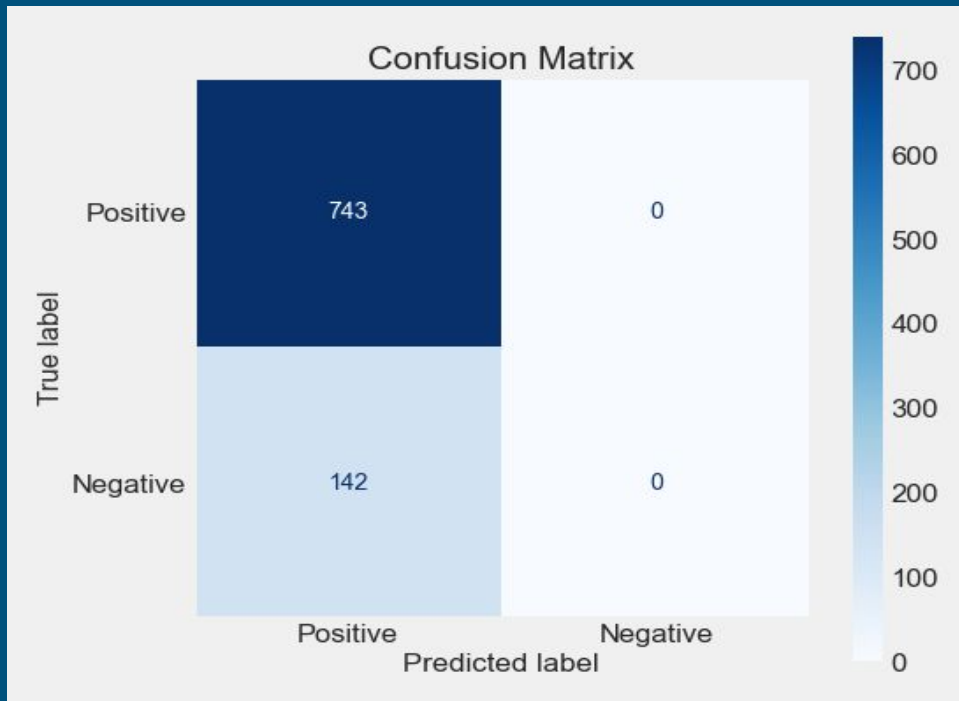
# Positive Word Cloud



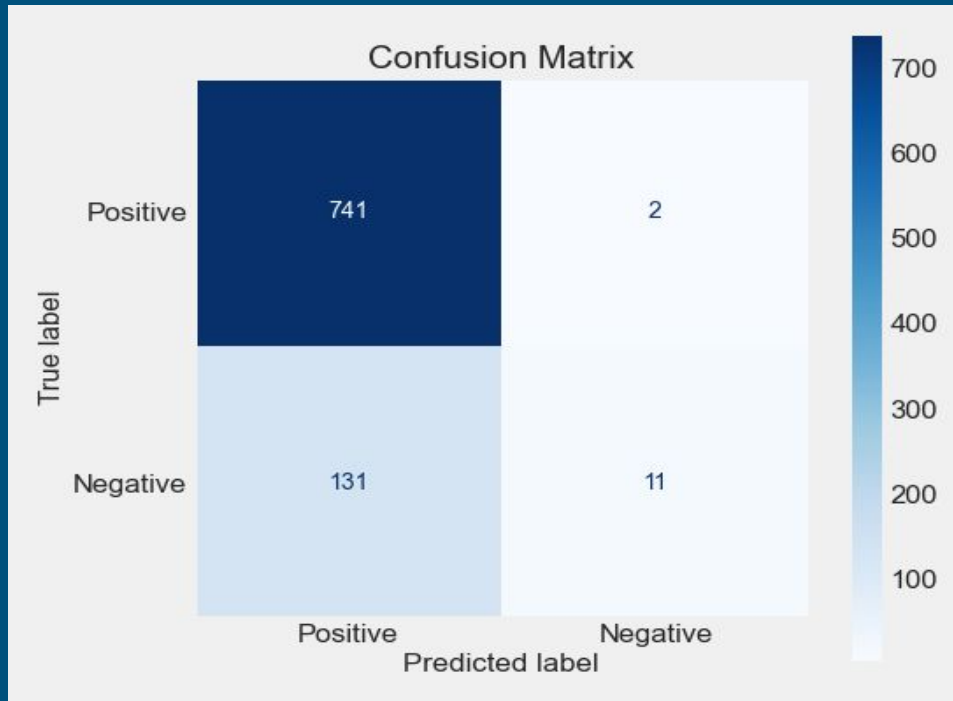
# Negative Word Cloud



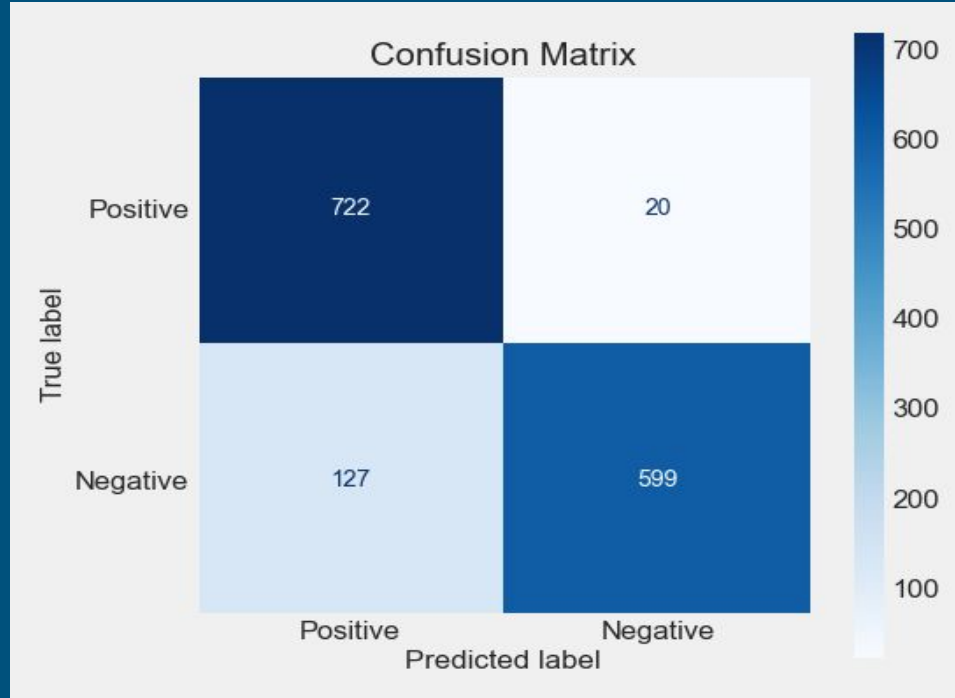
# Baseline: $F1=0$ (Recall=0, Precision=0)



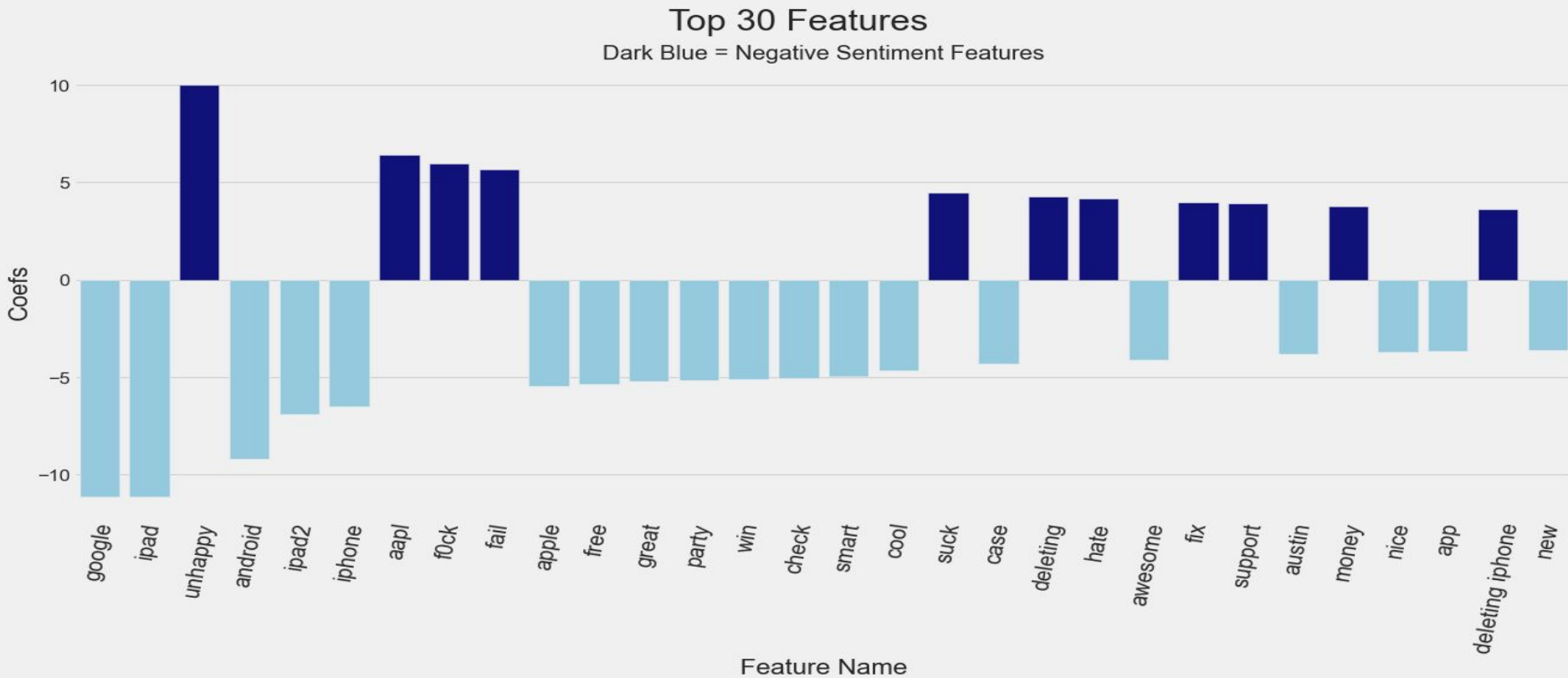
# Simple: $F1=.14$ (Recall=.08, Precision=.85)



# Final: F1=.89 (Recall=.83, Precision=.97)



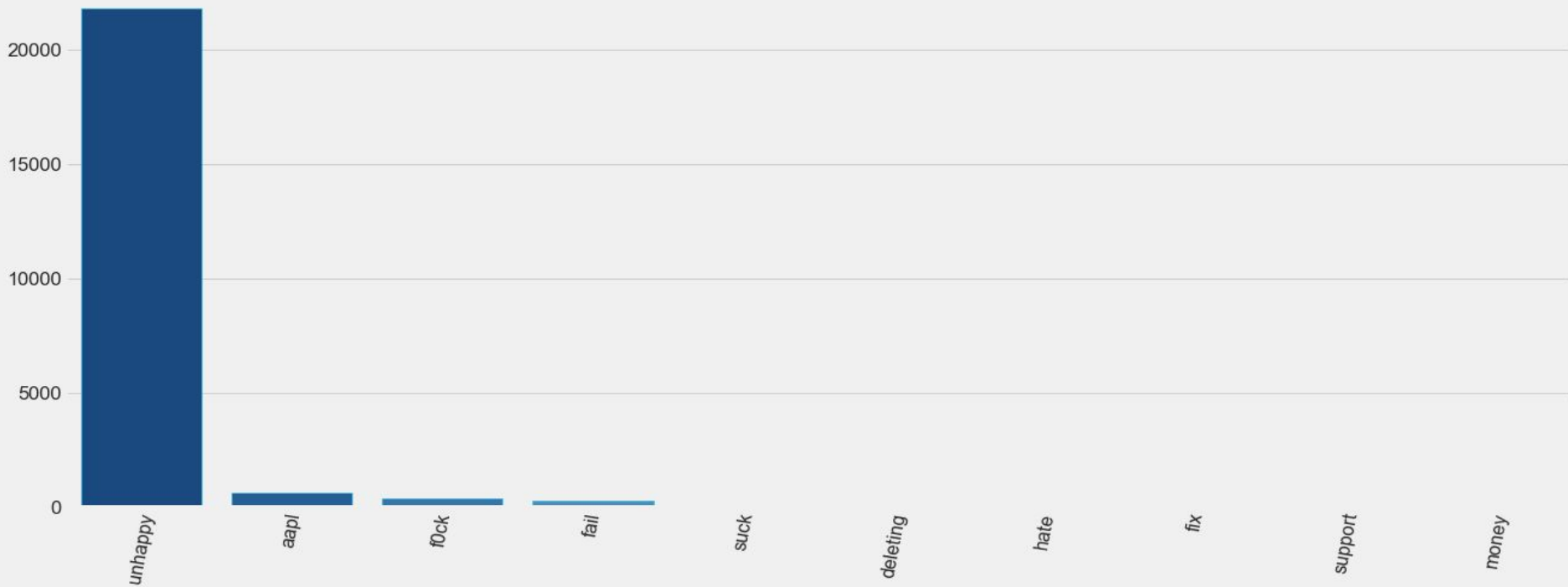
# Coefficients of Top Features



# Negative Relative Odds

## Relative Odds For The Top 10 Negative Features

Higher Bars Mean Higher Odds of a Negative Tweet

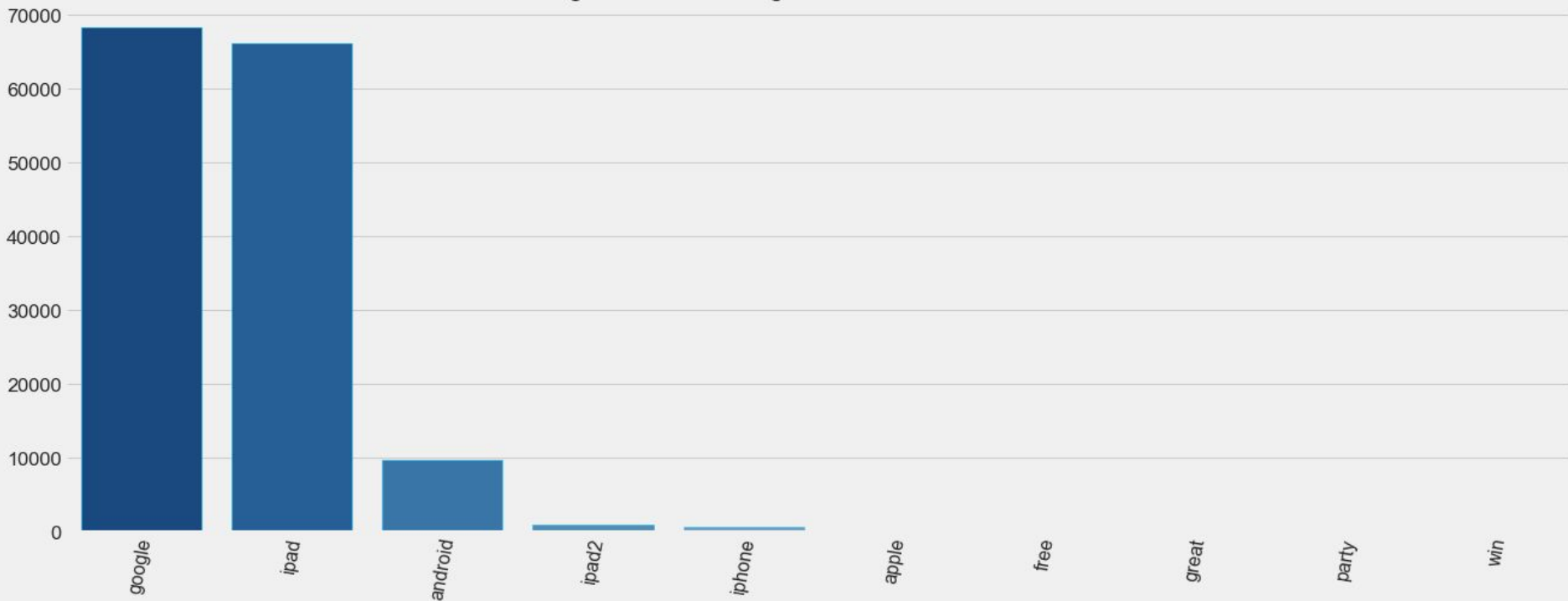




# Positive Relative Odds

## Relative Odds For Top 10 Positive Features

Higher Bars Mean Higher Odds of a Positive Tweet



# Next Steps

---

Next steps for the project include:

- Adding an advanced word embedding method.
- Tuning an RNN classifier.
- Adding neutral tweets with a multiclass classifier.
- Investigating the final model's adherence to the assumptions of logistic regression.

# Thank You!

**Email:** [jeffrey.h.watson@protonmail.com](mailto:jeffrey.h.watson@protonmail.com)

**GitHub:** @jeffreyhwatson