



The Product Sentiment Project



Twitter and Natural Language
Processing



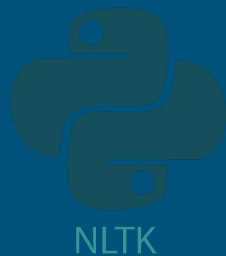
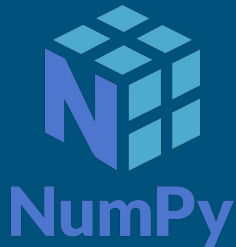
Goal

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

11,244 rows of text data:

- tweet
- brand/product
- sentiment

Data Sources

data.world/crowdflower

brands-and-product-emotions data set

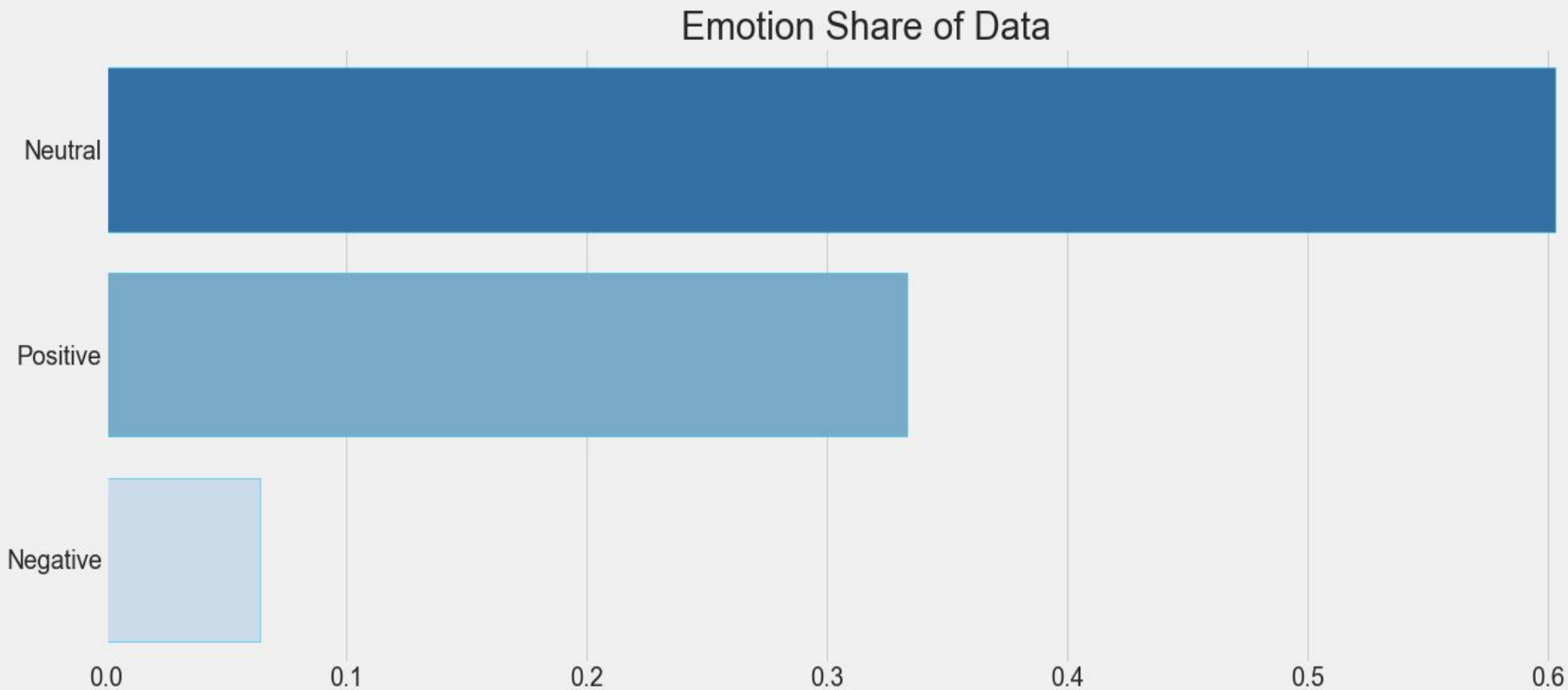
apple-twitter-sentiment data set

Data Sources

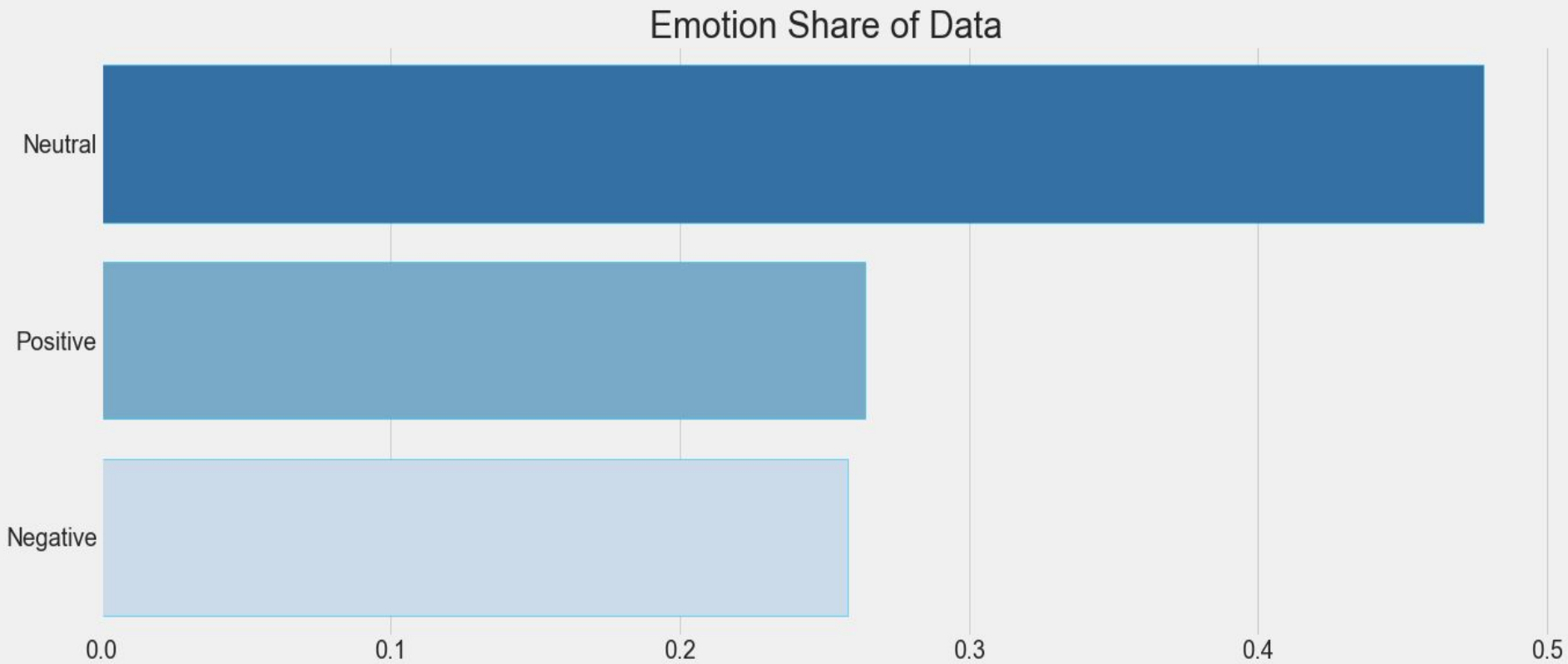
kaggle.com/shashank1558

preprocessed-twitter-tweets data set

Class Balance in Original Data



Class Balance in the Augmented Data



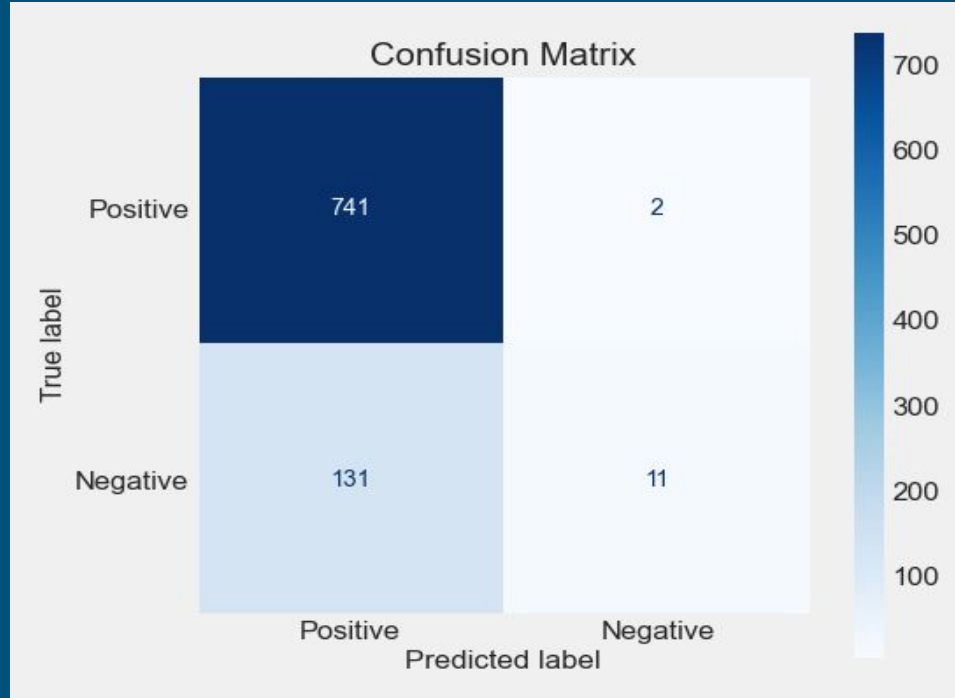
Positive Word Cloud



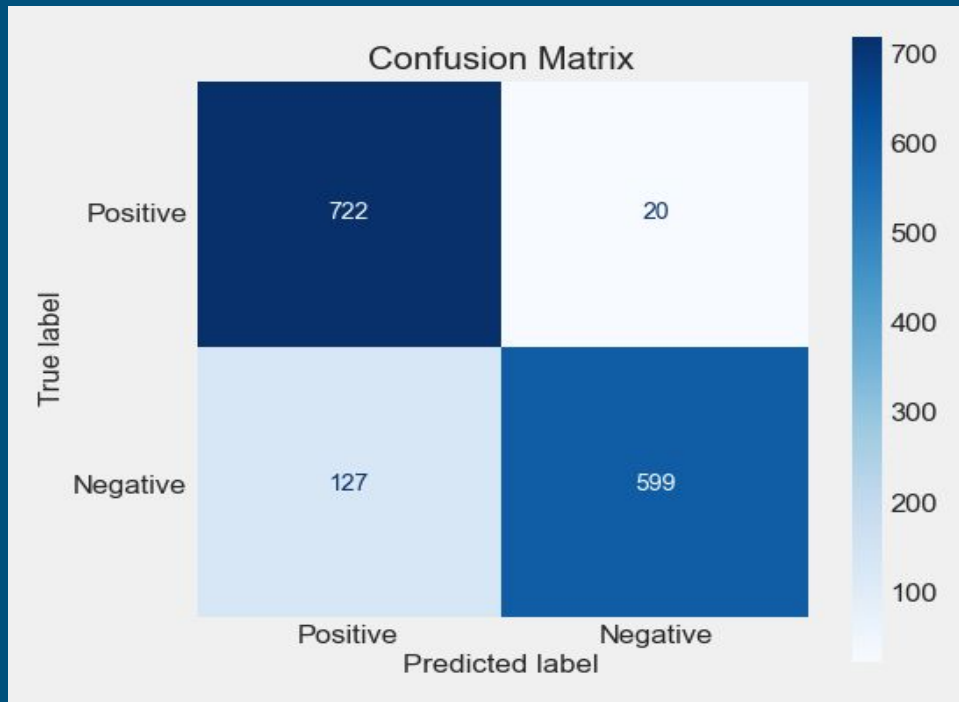
Negative Word Cloud



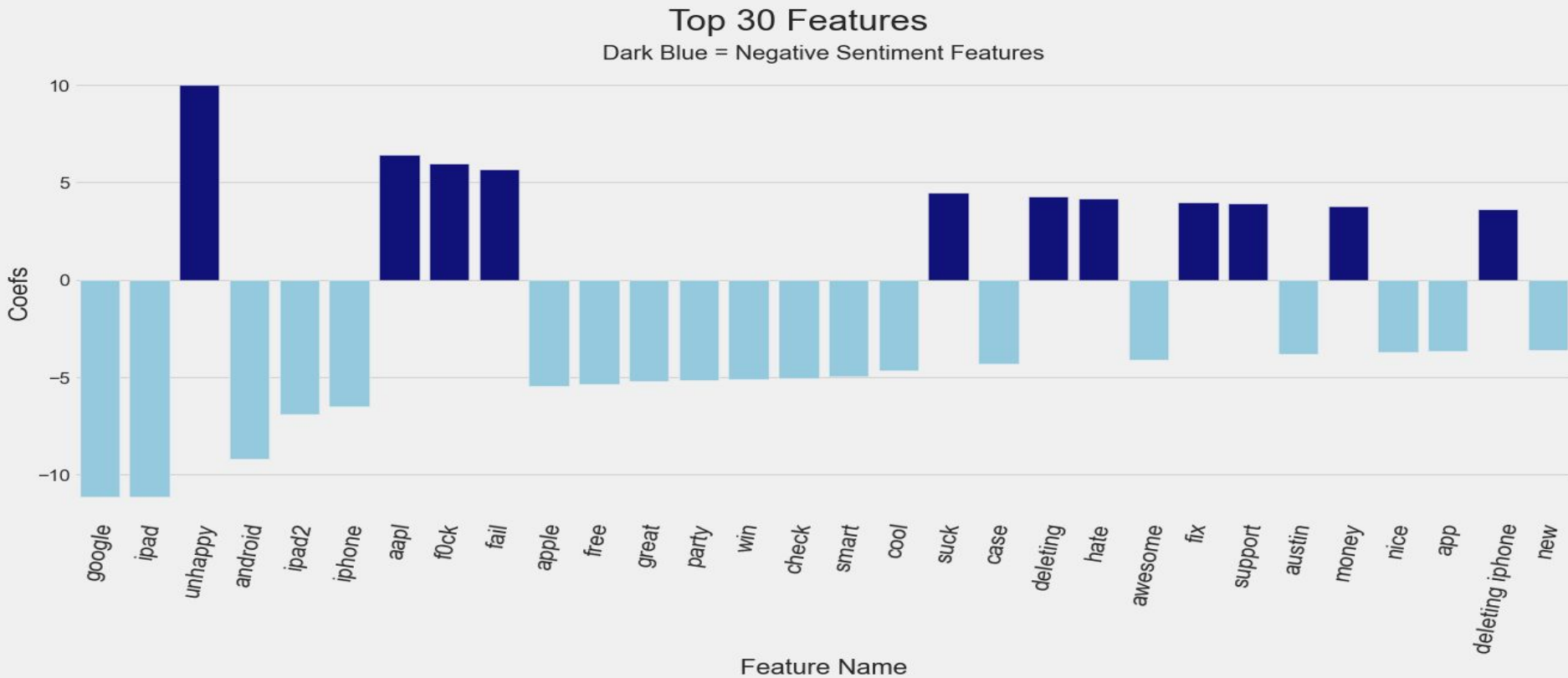
FSM: F1=0.14 (Recall=.08, Precision=.85)



Final: F1=0.89 (Recall=0.83, Precision=0.97)

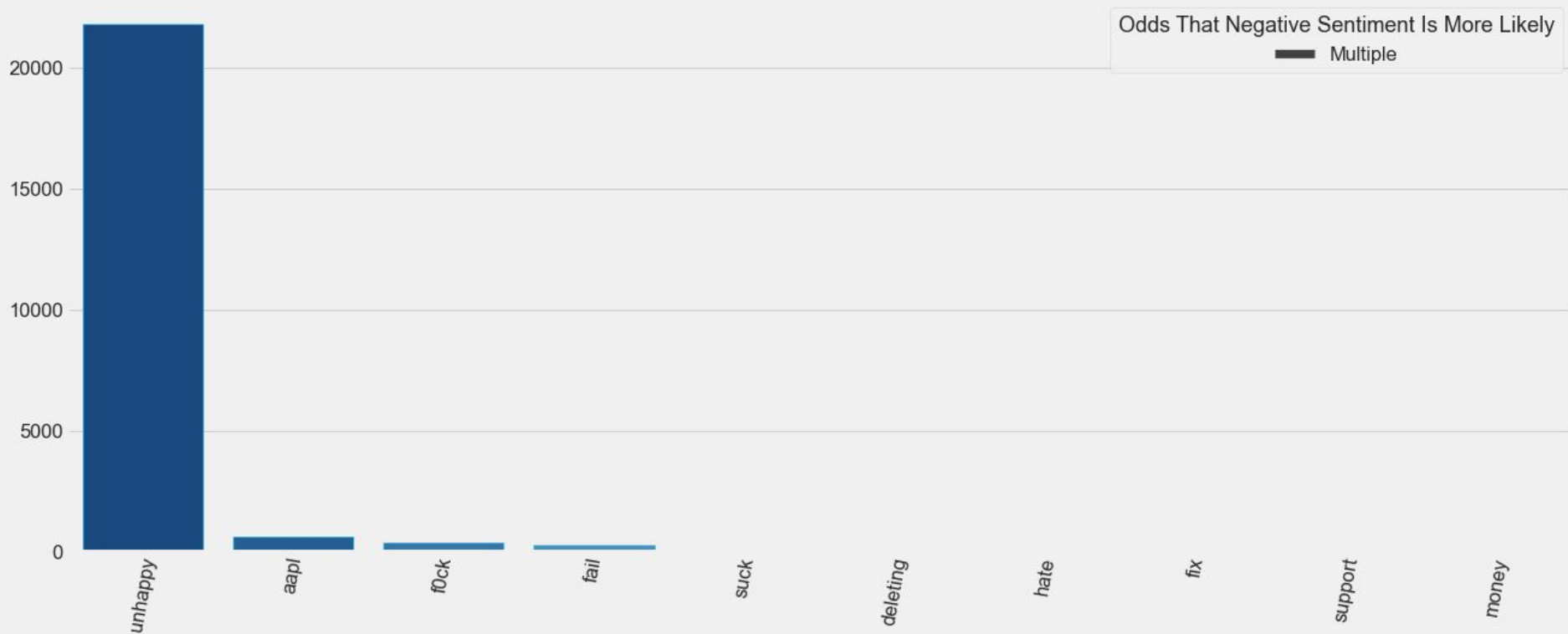


Coefficients of Top Features

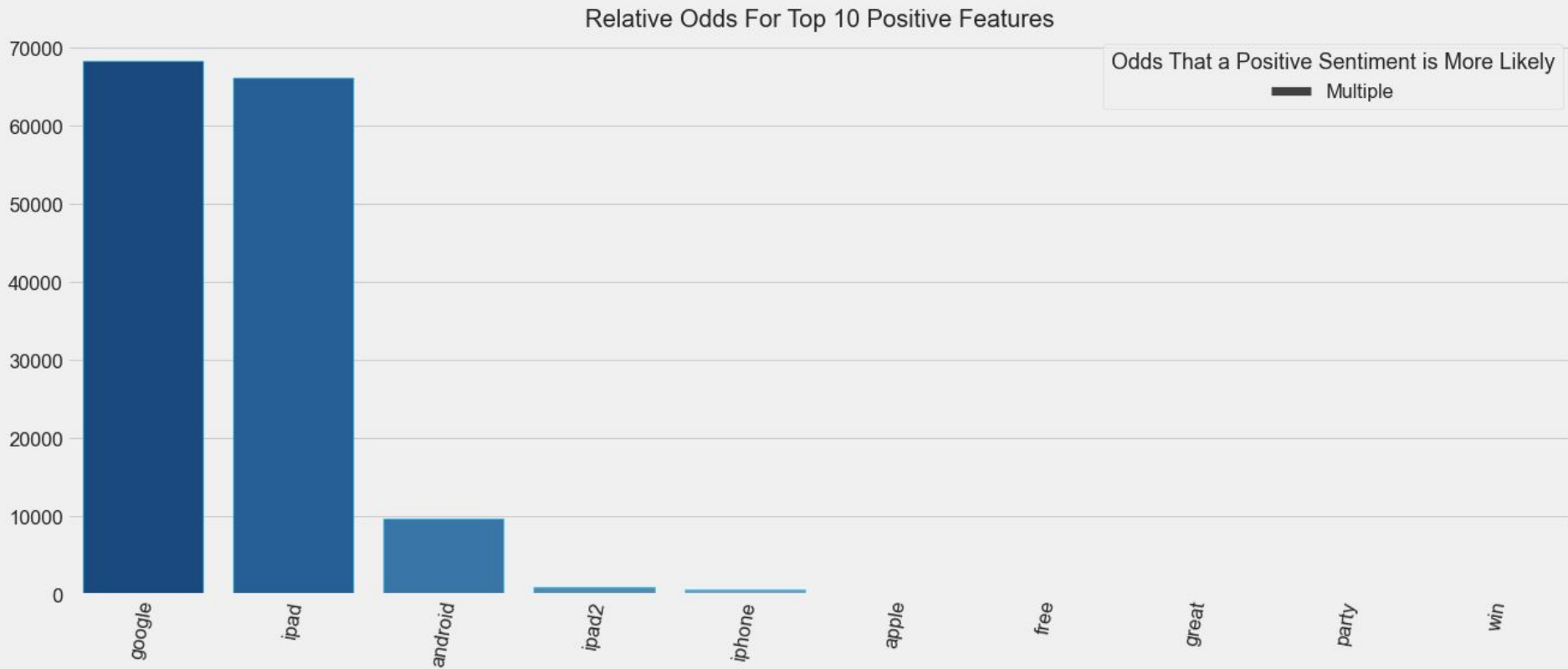


Negative Relative Odds

Relative Odds For The Top 10 Negative Features



Positive Relative Odds



Next Steps

Next steps for the project include:

- Using an advanced word embedding method and tuning an RNN classifier.
- Implementing a multiclass classifier and adding neutral tweets to the model.
- Further investigating the final model's adherence to the underlying assumptions of logistic regression.

Thank You!

Email: jeffrey.h.watson@protonmail.com

GitHub: @jeffreyhwatson