Analyzing Tweets: Positive or Negative

Joseph Denney

Purpose

A client is looking to design and manufacture a new smart phone and will invariably compete with Apple and Google products. They have provided us with a data set of Tweets and would like more detail regarding negatively and positively charged Tweets directed at both iPhone OS and Android OS phones.

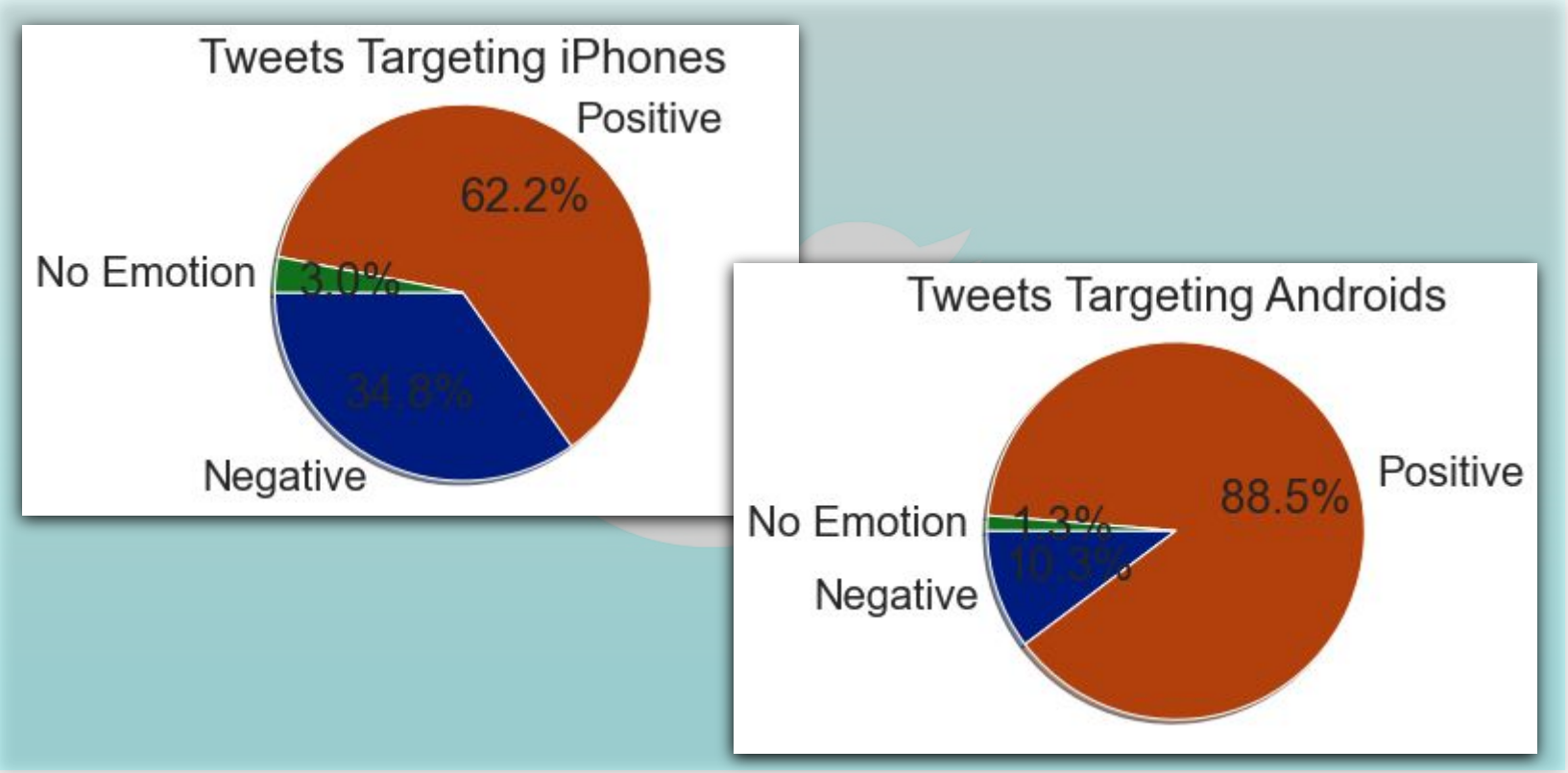
Our challenges are -

1. *To highlight any negative features of iPhones and Androids so that they can reduce them in their new product and*
2. *To highlight positive features of iPhones and Androids so that they can implement or improve them in their own product*
3. *To provide recommendations that will improve their future product*

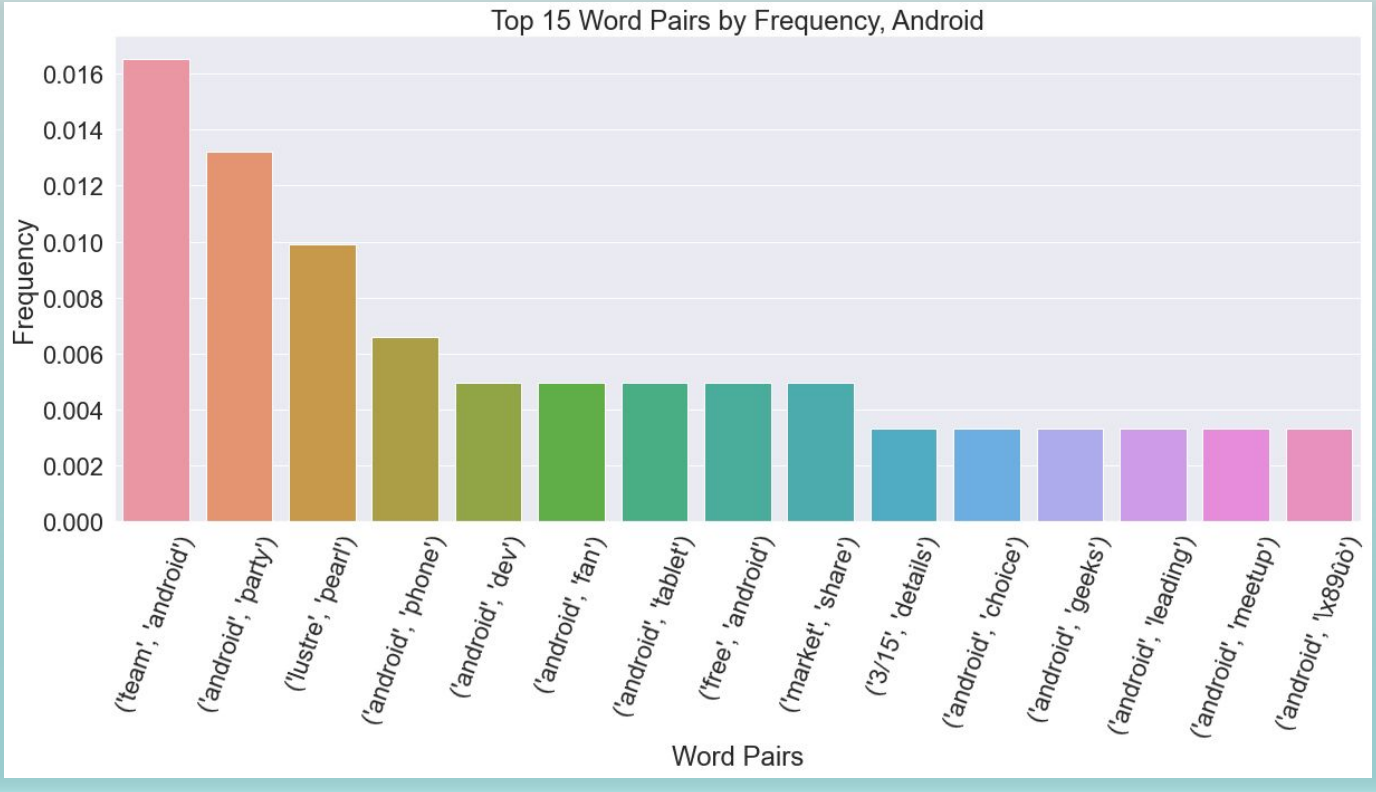
Questions

1. *In tweets targeting either the iPhone or Android phones, which product is more often the subject of negatively charged emotions?*
2. *What words are most common in negative tweets about iPhones and Android phones?*
3. *What are some of the positive features commented about for both iPhones and Android phones?*

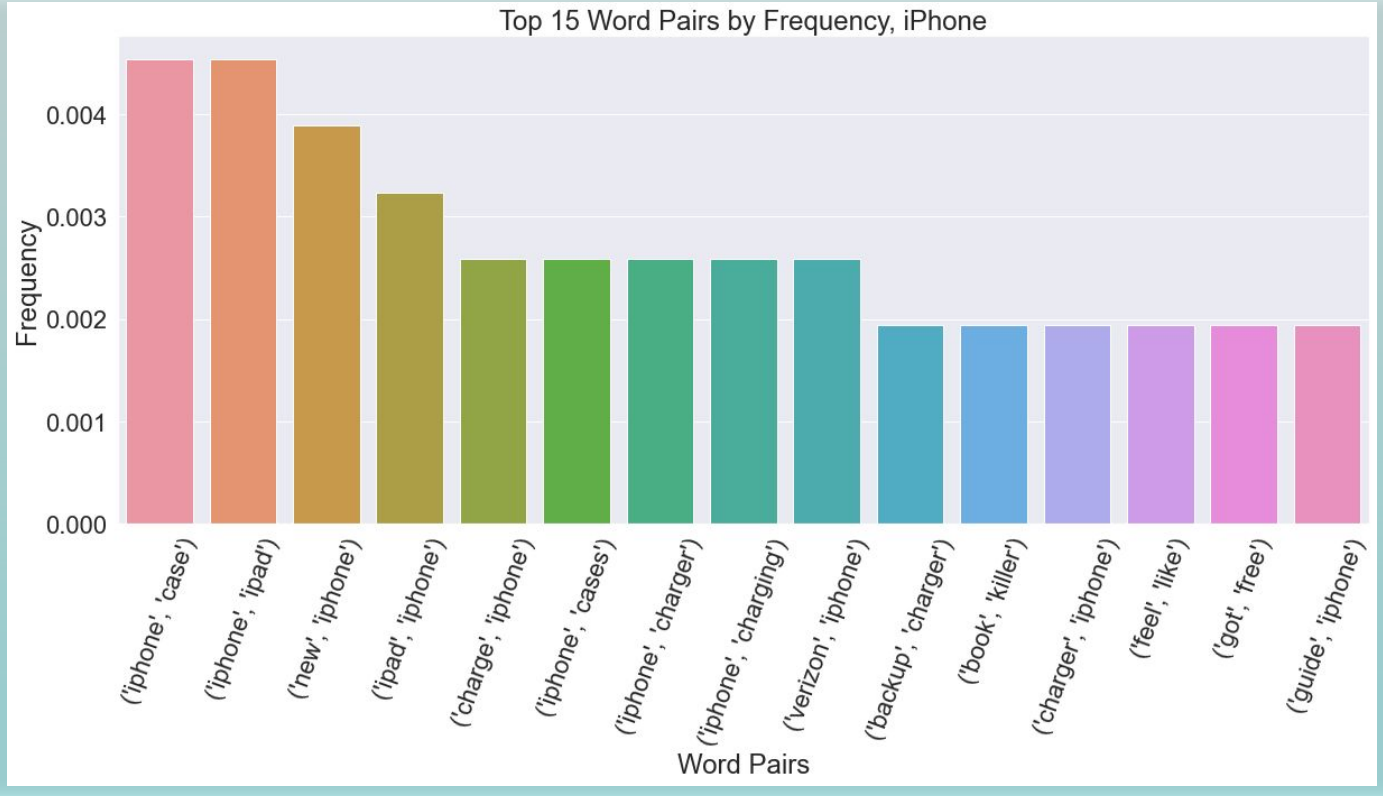
Avoid a Short Battery Life



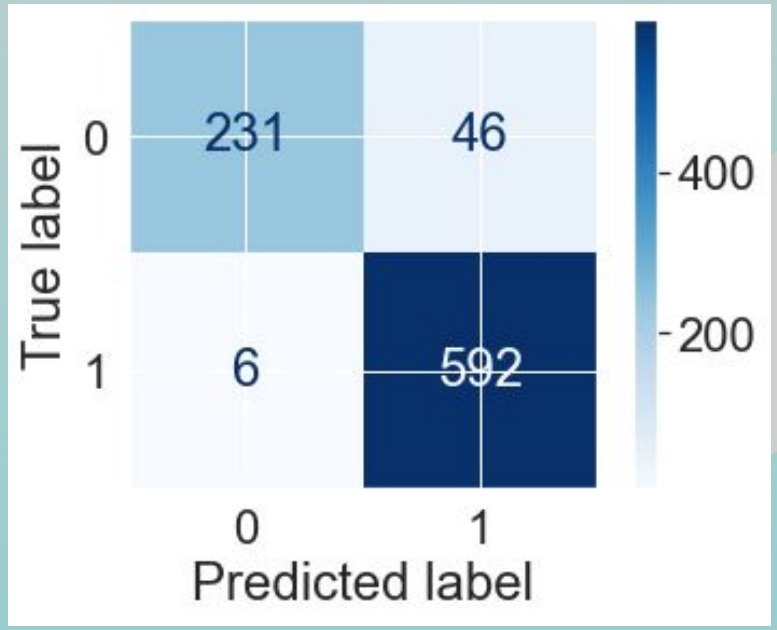
Build a Sleek Phone



Provide Extra Battery Life

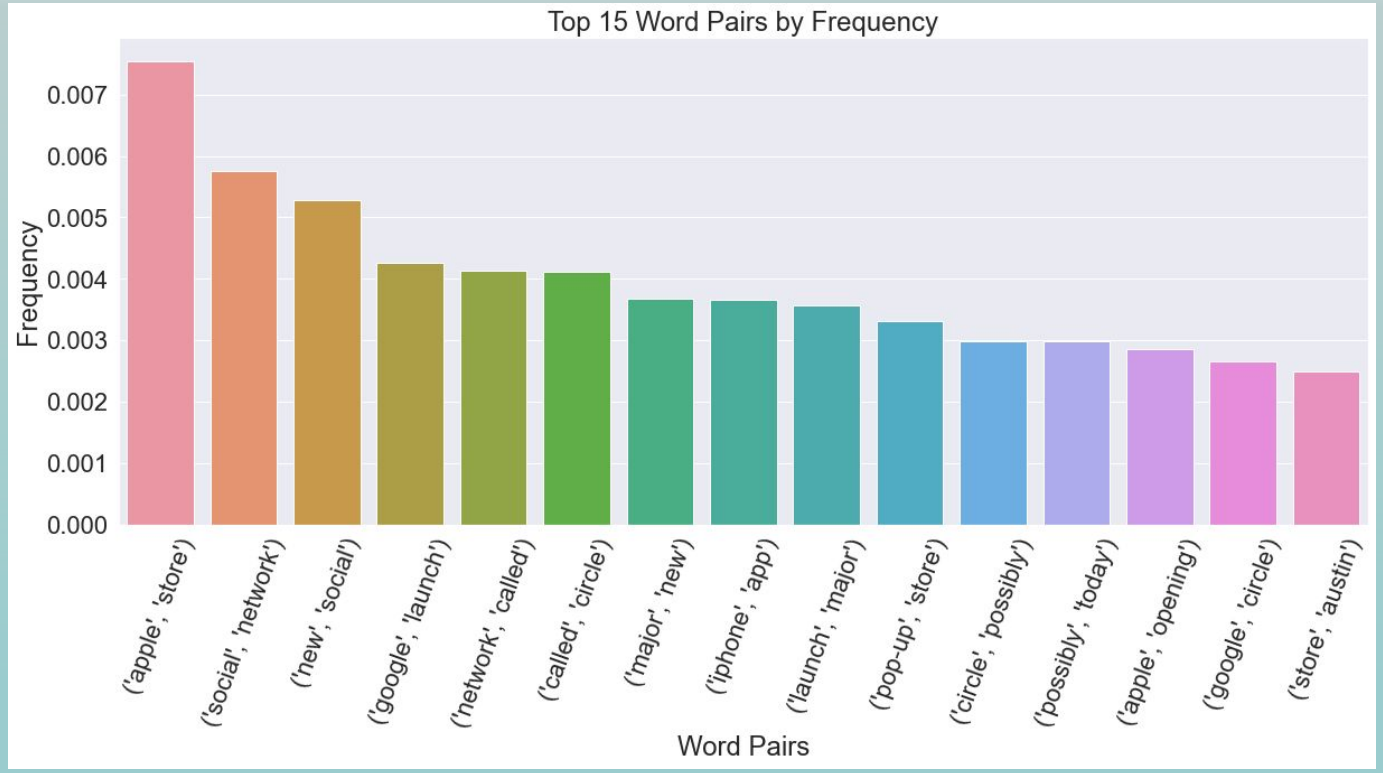


Overall Insights - Supervised Learning

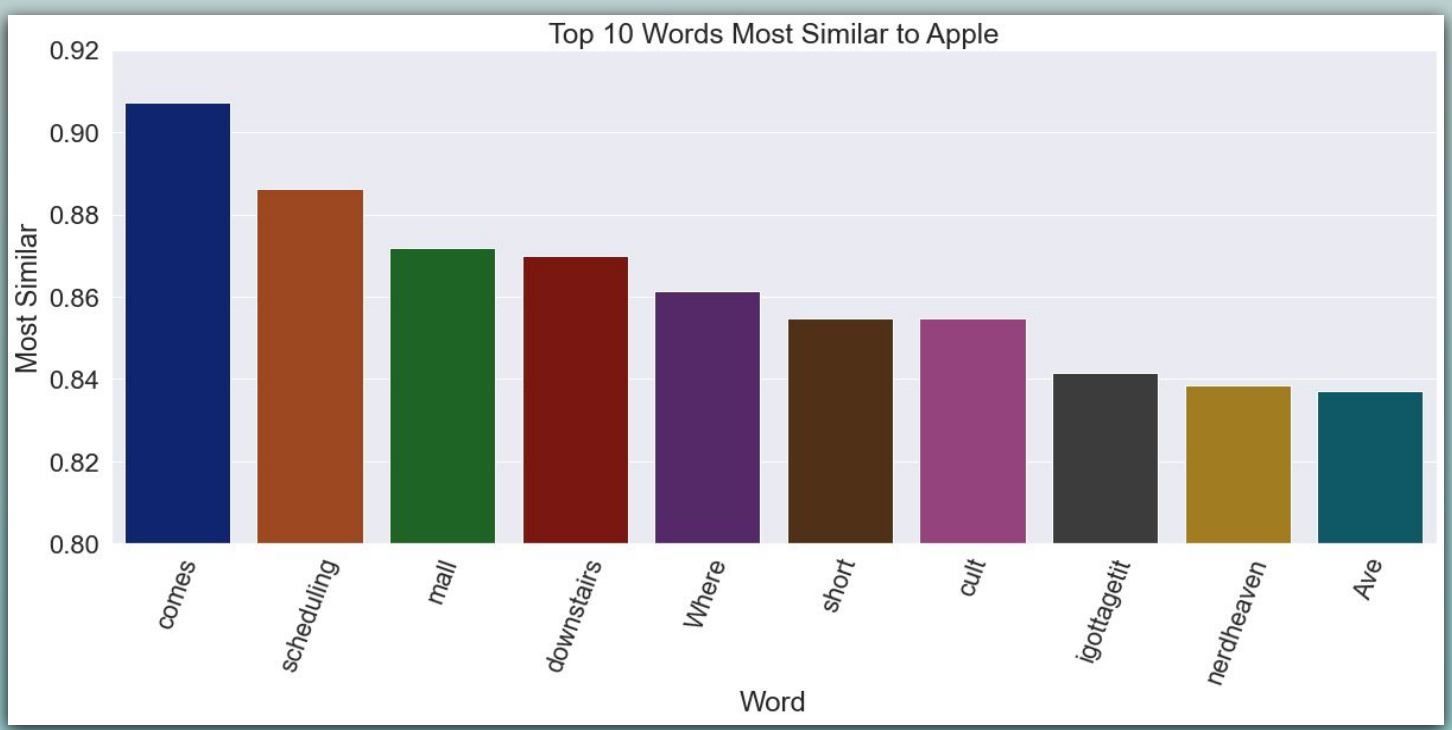


- Random Forest Classifier
- F1-Score of 93% with 94% Overall Accuracy
- 46 out of 648 true positives are incorrectly identified as negative by the model
- Low false positive identification

Insights from Tokenization and Bigrams



Insights using Word2Vec



Conclusions

- Design a power frugal smartphone
- Create a sleek exterior design and a customizable user interface without bugs
- Provide auxiliary power options including charger locations and phone cases with built in battery

Future Work

- Continue to separate iPhone OS and Android OS tweets to look at different behaviors and word usage of those two groups
- Use Spacy NLP to look at Parts of Speech tagging of the bag of words
- Explore which words have the *most* positive and *most* negative connotations in tweets
- Improve multiclass model and create graphics to show the model performance.


Thank
You!

Appendix - Supervised Learning

- Using a multiclass classification sequential Deep Neural Network model, achieved 81.5% (99.8% training accuracy) accuracy on the test data for 4 classes of Tweets -
- Ambiguous
- Negative
- No emotion toward brand or product
- Positive