Tweet NLP Analysis

Will Byrd
Flatiron School 2024

Business Overview

| Brand	# Of Tweets
iPad	2346
Google	2289
Apple	1764
iPhone	934
Android	505

- Tweet Analysis
  - **SXSW-2011**
  - Multi-classification
    - **Sentiment** in Tweets
- Objectives
  - Better predict Sentiment of Tweets
  - Determine relative opinion of **brands** at SXSW
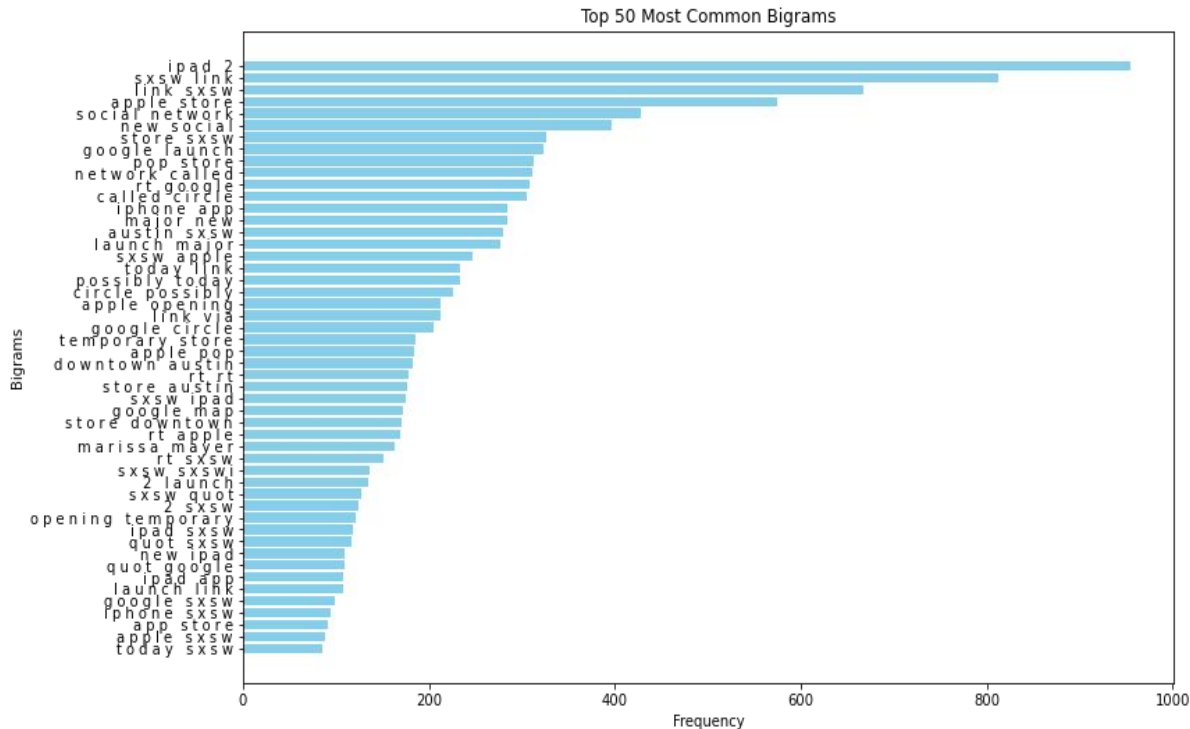
Data Overview

Top 20 Most Frequent Words
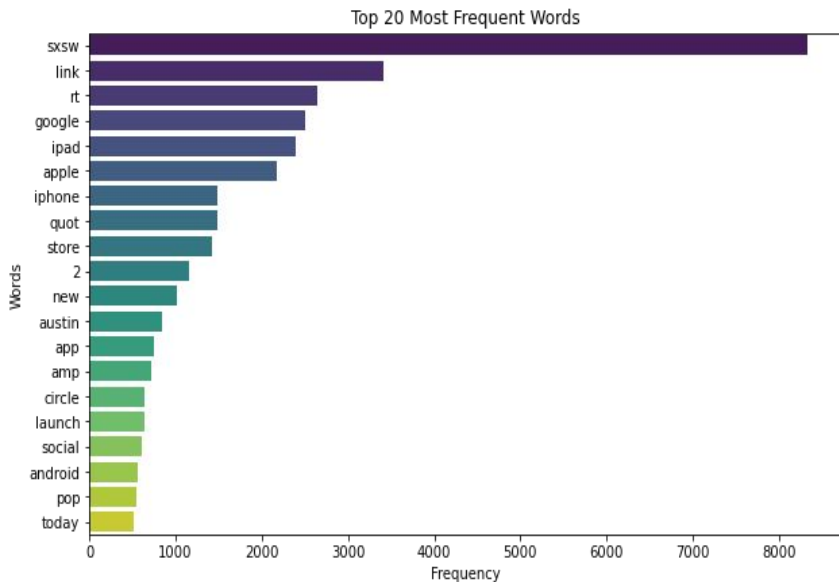
- 8721 SXSW Tweets
  - Sentiment rating
    - Negative, Neutral, Positive
  - Brand relevance
   - iPad, iPhone, Apple, Google, Android
  - nGrams

Data Analysis
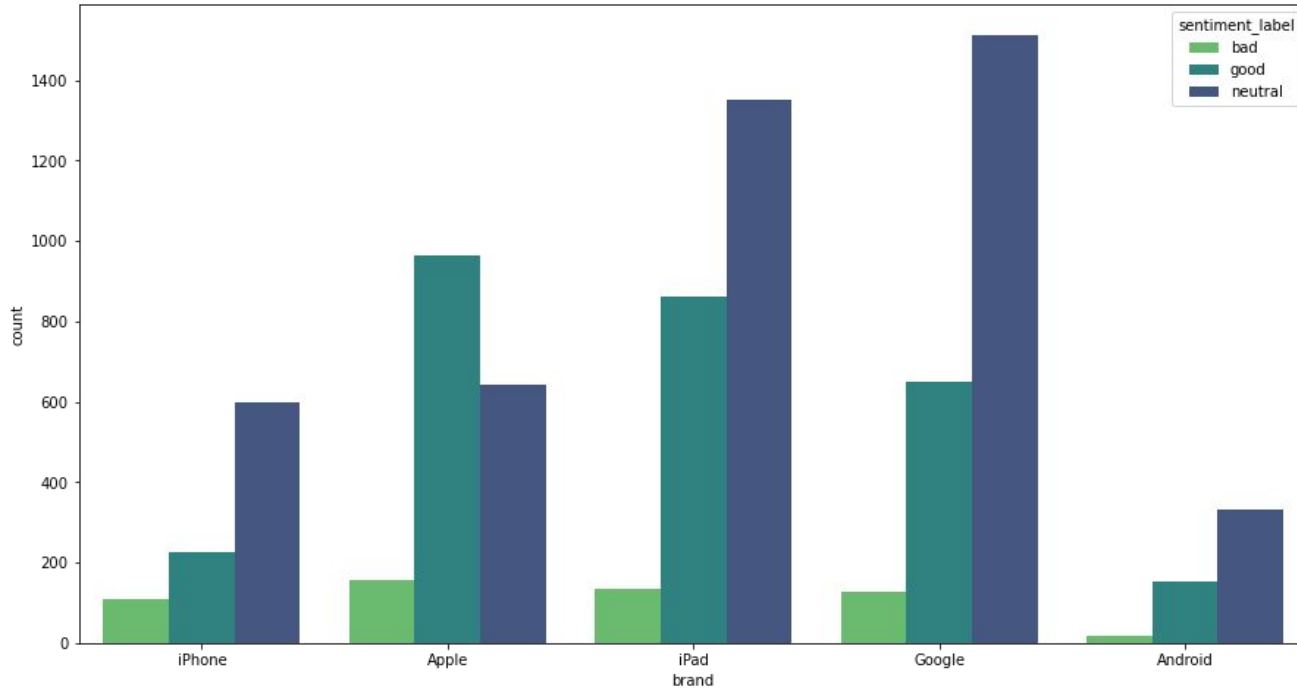
**Sentiment for each brand. Notice Neutral is the most common rating for each brand except Apple

Data Analysis cont.

Top 50 Most Common Bigrams

**Bigrams-pairs of commonly occurring words. Notice ipad 2, apple store, google launch, rt google, iphone app, sxsw apple, and apple opening appearing more commonly. *Android is not even in our list!*

Histogram of TF-IDF Values for the 50 Most Common Words
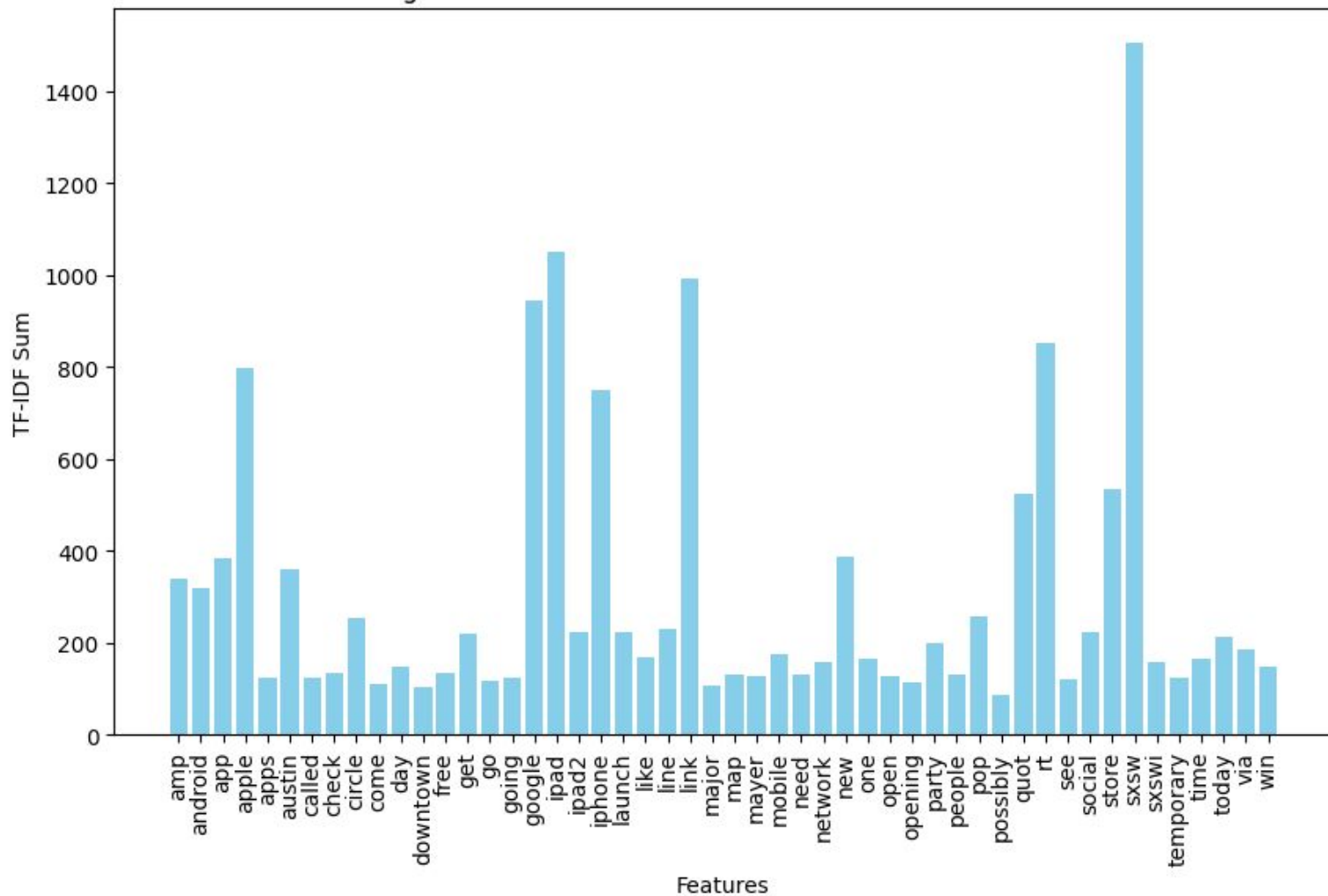
**TF IDF-Term Frequency, Inverse Document Frequency. A measure of importance of a specific word in the corpus, adjusted for the fact that the word occurs commonly in that corpus.

Classification Modeling

	Baseline	Improved
Accuracy	59%	69%
Precision	58%	68%
Recall	59%	69%

- Great improvements to our baseline model via Ensemble Methods
- 3 class, classification guessing=33%
- Precision-tells us how many of the predicted positive instances were actually correct
  - Higher precision = less false positives
- Recall-Tells us how many of the actual positive instances were correctly identified
  - Higher Recall = fewer false negatives

Importance of Metrics

	Baseline	Improved
Accuracy	59%	69%
Precision	58%	68%
Recall	59%	69%

- Accuracy and Precision are most important
  - False positive means our model incorrectly identifies a positive sentiment.
  - Knowing which brands are viewed positively is main goal
  - Minority class was negative tweets, so false negatives are not as detrimental to analysis.
- Positive class-positive sentiment
- Negative class-negative sentiment

Brand	Count	Frequency
iPad	2346	30%
Google	2289	29%
Apple	1764	22%
iPhone	934	12%
Android	505	6%

Recommendations

- *Apple* was the only brand with more positive tweets than any other sentiment
- *iPad* and *Google* were the most talked about brands
- *Android* did not draw much attention
- *Apple and Google should be given most resources for next year's conference and their booths should be in high traffic areas.*
- *Marketing material should promote Apple and Google as they draw more attention than other brands.*
- **Android should not have as many resources next year. If anything, their booth should not be in a high traffic area.**

Additional Questions

Thank you!

By: Will Byrd

Byrdwcrawford@gmail.com

The Flatiron School