Using Sentiment Analysis to Improve Google Products

Raymond Leong

Flatiron Phase 4 Project

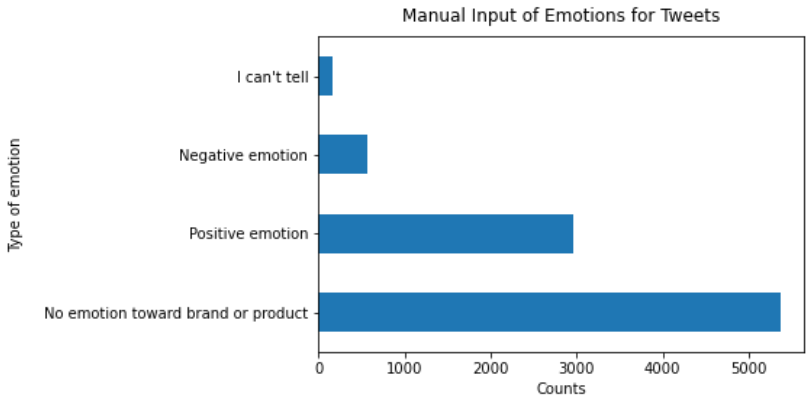
Introduction

- What is Twitter?
  - An independent platform where opinions are shared about almost any topic
    - The main page of twitter usually shows what is trending
    - Past data about tweets
- This project can help accentuate the results of Google Duplex based on the Sentiment of a tweet

Business Problems

- Business problem: Are tweets a good resource to help enhance Google products?
  - What recommendations can we make based on user's raw opinions?
  - Positive tweets helpful in identifying trends?
  - Negative tweets representative of problems that Google can solve?

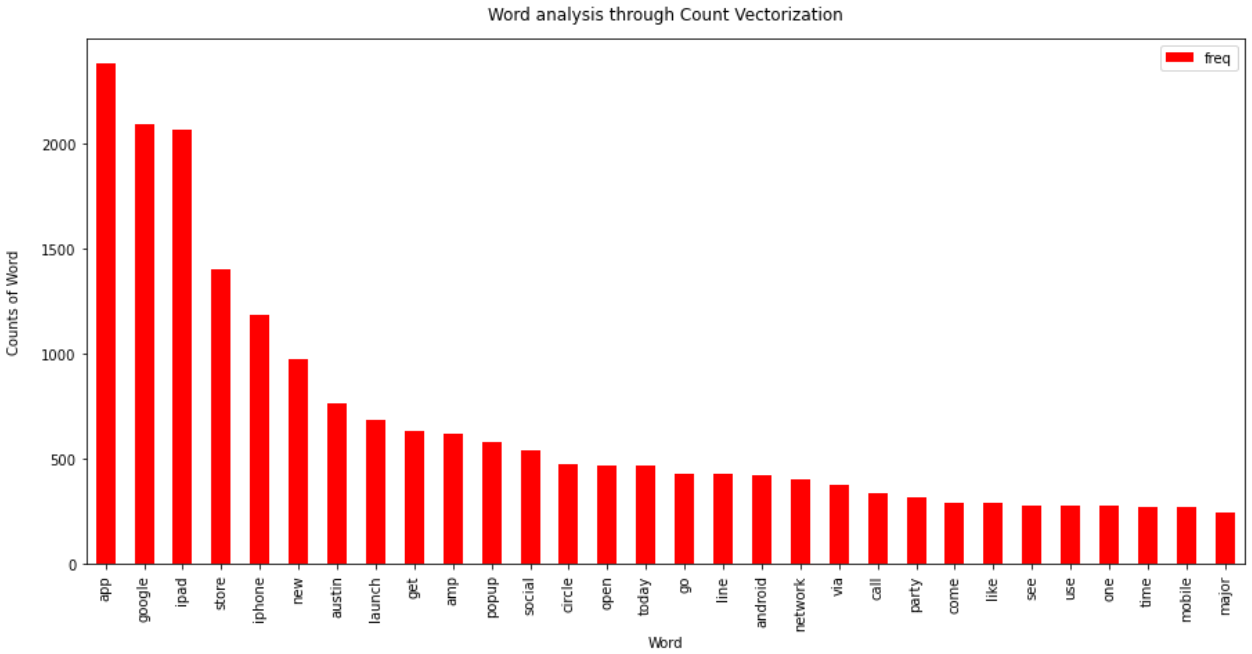
The Data



- Data
  - ~9000 tweets with 24 duplicated tweets
  - ~4000 tweets that tell you what object the tweet is referencing
  - ~4392 “neutral tweets”
- Data is imbalanced
  - Emotions through Amazon Turk were used and this is also biased
  - Some tweets are unidentifiable

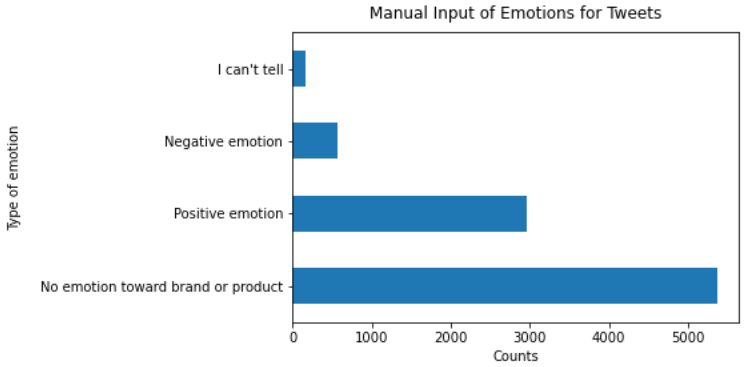
A Glance at the Tweet Data

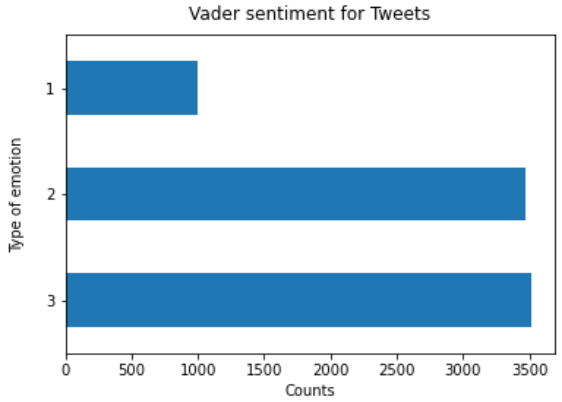
- What is mostly discussed in both Google and Apple products
  - Apps, physical stores, location of a new launch for a phone, people attending said launches
    - Popup problems, which probably an android issue



Using Vader to improve Sentiment

- Why?
  - The data column of sentiment may be biased
- Under vader:
  - 1 means there's a negative emotion
  - 2 means the emotion is neutral
  - 3 means the emotion is positive



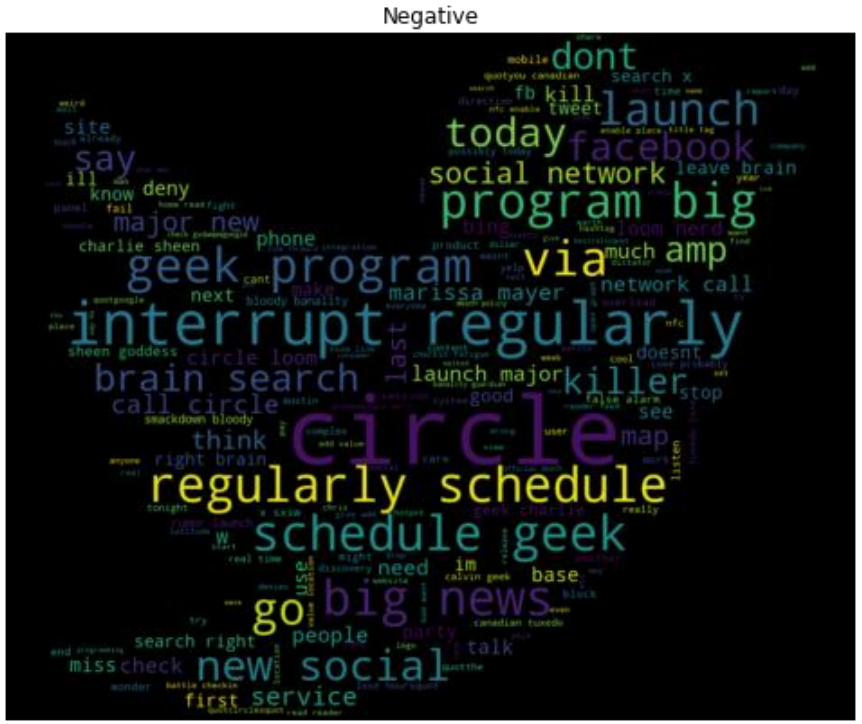


Positive words about Google

- Most words are just talking about their experience with Google
- App, Map, TV may be areas that Google are doing well but more information is needed
- **Positive words are not representative of information that can help Google products**

Negative words

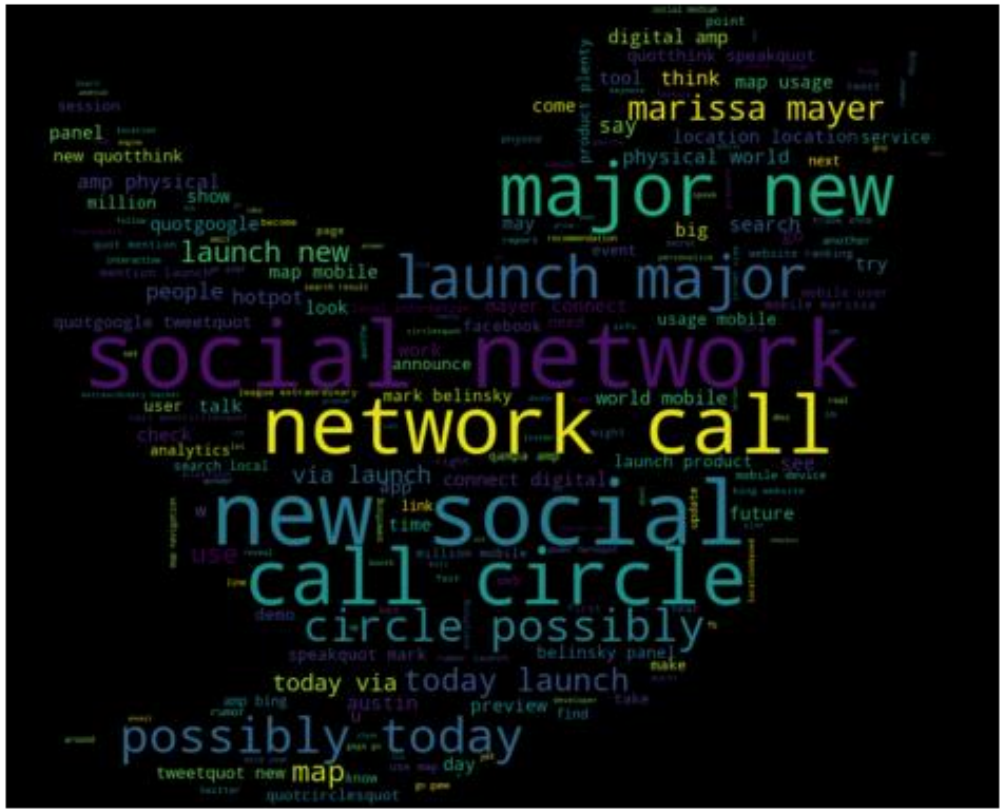
- Negative words that surround twitter need improvement.
  - PR may need to address how the population perceives it as a “geek program”
  - Interruptions appear to be regular, but there is no correlation between the two words
  - Circle seems to be a prominent problem but no attributes about it are clearly identifiable
- **There are no immediate action needed to Negative tweets- more insight is necessary**



Neutral words

- Circle appears here and in the negative Word cloud
  - These phrases make separating neutral from Negative words complicated due to the **ambiguity**
- Pairs of words appear because they have the same frequency- not because are associated
- **The lack of association between words makes it difficult for Google to improve on their products based on tweets**

Neutral words

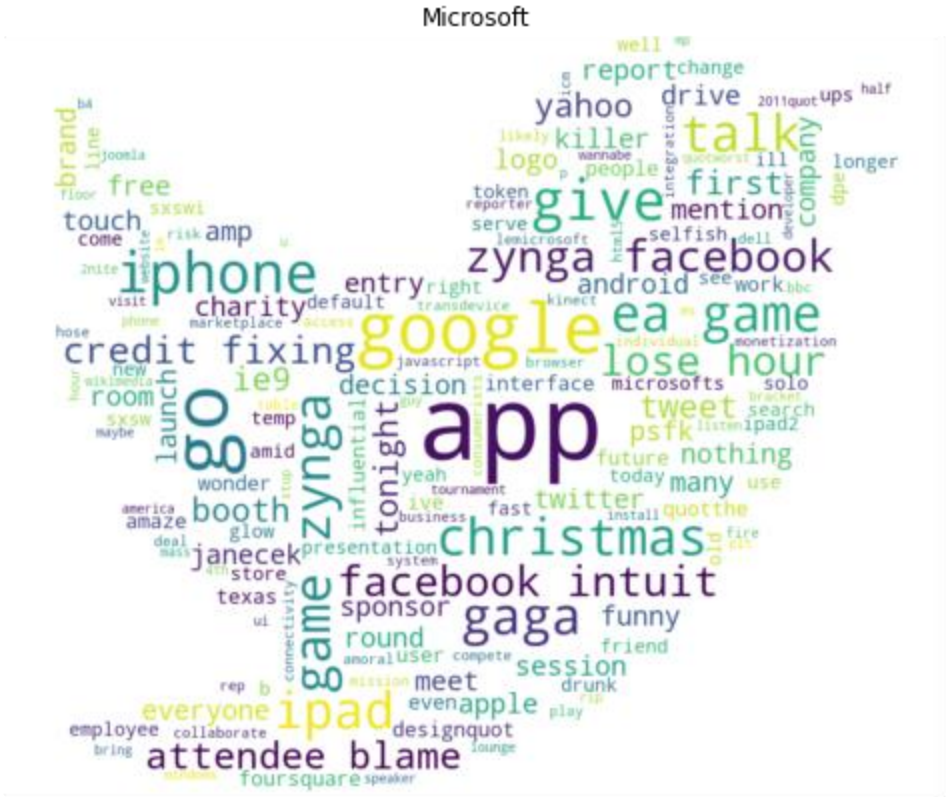


Model Results

- Vader engineered our new sentiments to be 3521 positive tweets, 3477 neutral tweets and 998 negative tweets.
- This came from 4469 Neutral tweets, 2838 Positive tweets, 543 negative tweets and 146 undistinguishable tweets.
- The probability of a tweet being positive, when its actually neutral is 1200/5200 + 1200: which means that 18.75% of all positive tweets will be predicted as neutral unintentionally.
- Accuracy and precision hover in the 75% range for the final model.

Model Limitations

- The sentiment towards the OS doesn't seem to be significant at a glance. Thus not much interpretation can be made
- A pipeline would have been appropriate for this study but the problem resides in the EDA and the encoding problems
- Different stopwords/lemmatization/stemming should be investigated ,



Final Thoughts

- **The analysis doesn’t confirm the need for twitter to be incorporated into improving Google’s product and services**
- Acquiring more tweets needs to be more readily accessible in order to reduce class imbalance issues and improve model vocabulary

Thank you!

- Contact information
- Email : lraymond94sbu@gmail.com
- Linkedin: https://www.linkedin.com/in/raymond-leong-b75b11231
- Github: https://github.com/RaymondLeong94