Using Sentiment Analysis to Improve Google Products

Raymond Leong

Flatiron Phase 4 Project

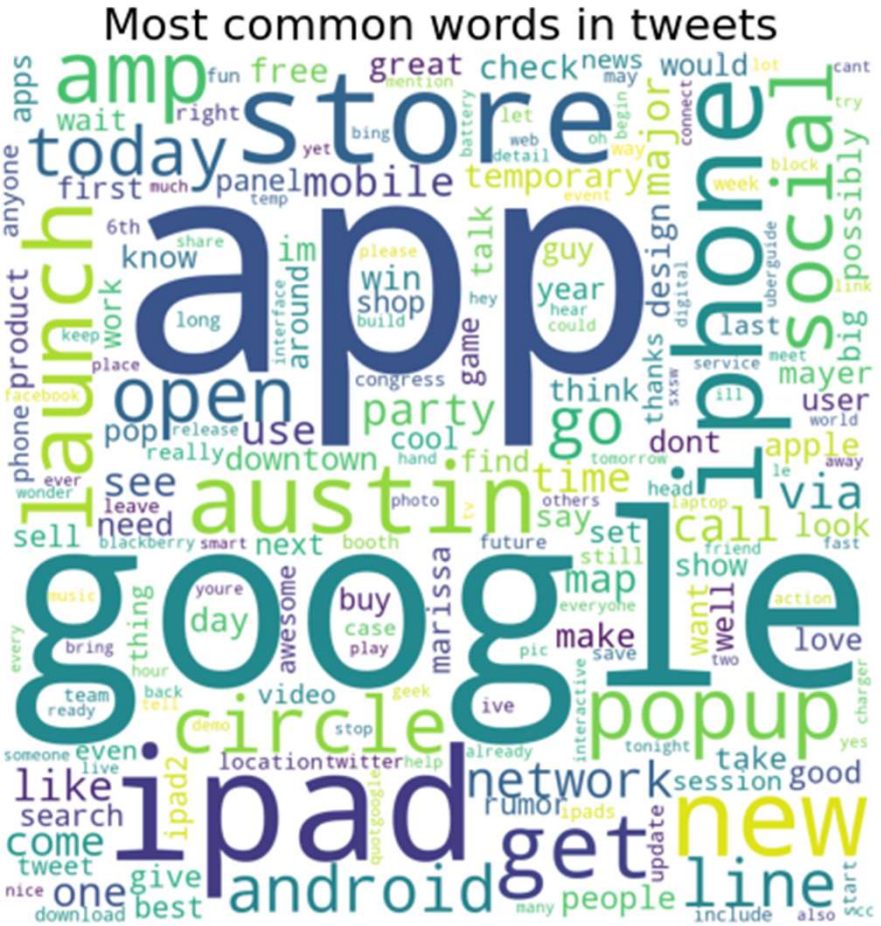
Introduction

- 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
- After Sentiment analysis, Tweets generated from individuals can be of either
    - Complaint
    - Praise
    - Neutral

Business problems and Data

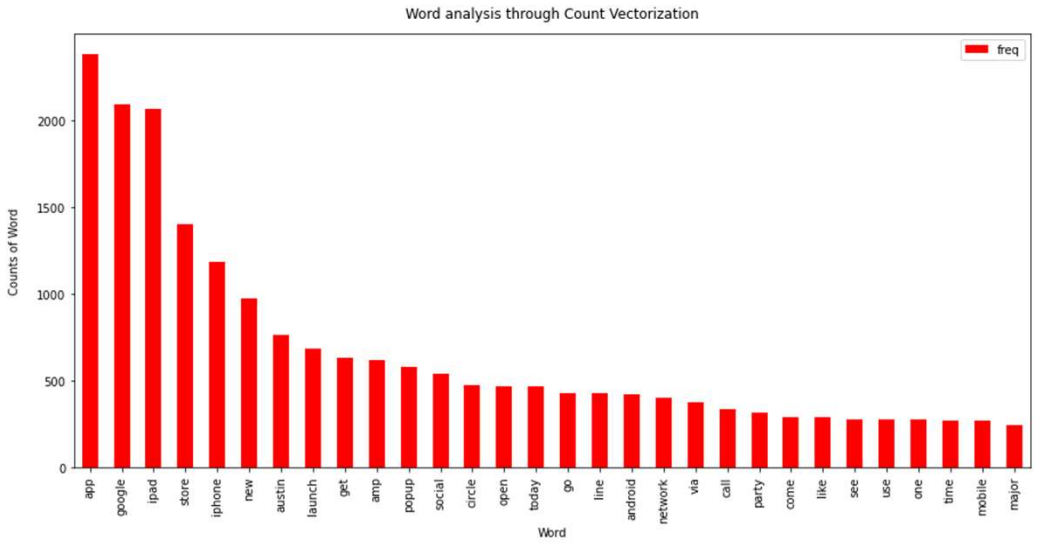
- How can Google use twitter to better their customer experience based on the tweets observed?
  - Solution: Create a Sentiment analyzer model for each tweet then reexamine the data to look for areas of improvement with NLP
  - Validate with a probability model (MNB)
- Data
  - ~9000 tweets with 24 duplicated tweets
  - ~4000 tweets that tell you what object the tweet is

  referencing
  - ~4392 “neutral tweets”



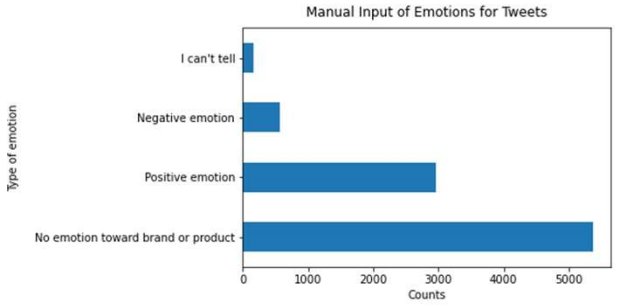
A Glance at the Tweet Data

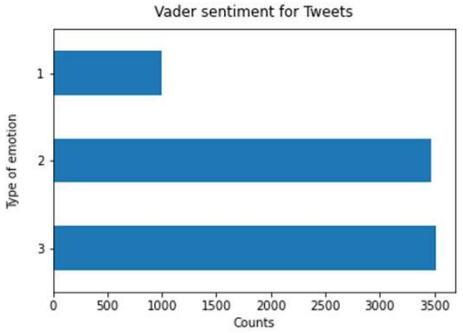
- What is mostly discussed in both Google and Apple products
  - Apps, physical stores, location of a new launch for a phone, people attending said launches
    - Popup problems, which probably an android issue



Using Vader to improve Sentiment

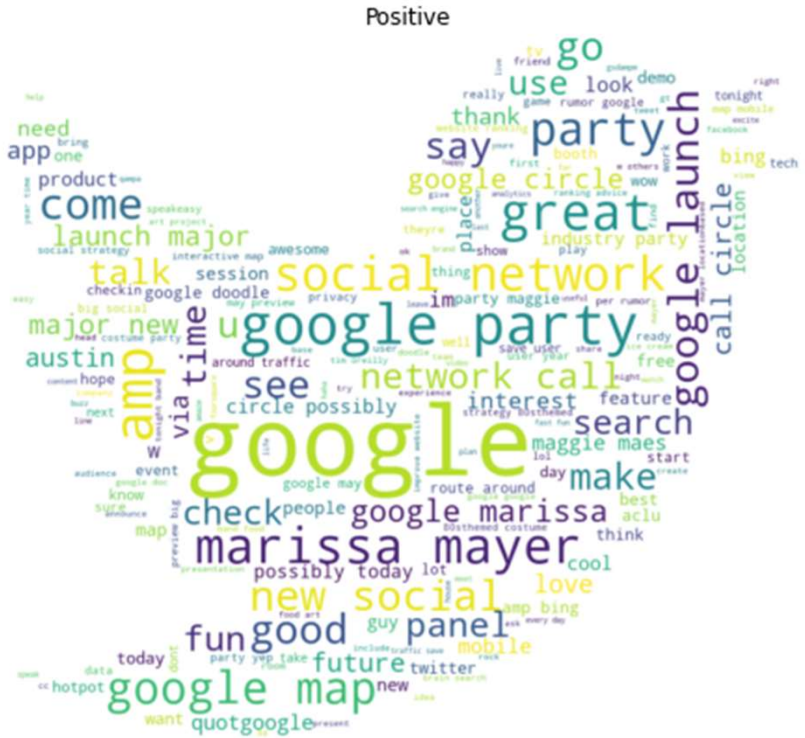
- Why?
  - The data column of sentiment may be biased
- 1 means there's a negative emotion
- 2 means the emotion is neutral
- 3 means the emotion is positive





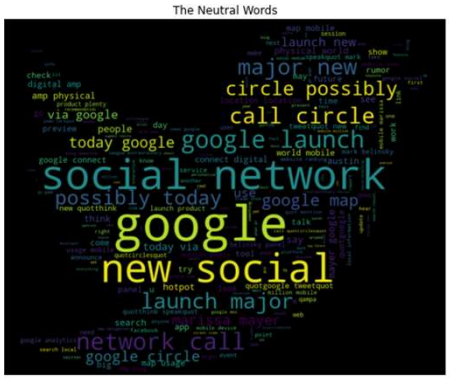
Positive words about Google

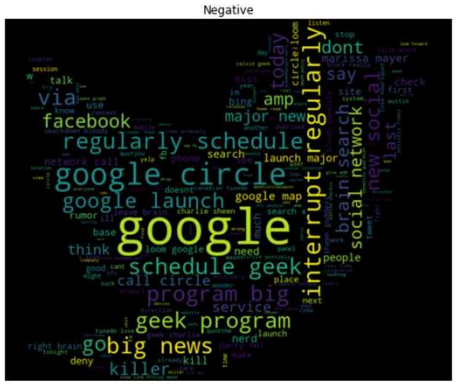
- Positive phrases that surround google suggests it is free, cool and apparently a team feature may be a contributing factor.
- The team feature is distinguishable from other web conference platforms. Thus, money can be invested into it for it to grow but more insight is needed.



Negative and Neutral Words

- Negative words that surround twitter need improvement. For instance, there is probably the perception that it is a “geek program”. Thus, PR may be requested to help shape the popular opinion.
- There is also a problem where it is “interrupted regularly” This may be a service problem and not directly related to google, as T-mobile and Verizon were mentioned.
- Tabs should be kept on how Google reacts with social networks using their Circle app.





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

- NLP requires several iterations of EDA -> Model building, I only did one.
- Acquiring more tweets needs to be more readily accessible in order to reduce class imbalance issues
- Twitter is now going under Elon's guidance- it may not be the same platform in a few months

Thank you!

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