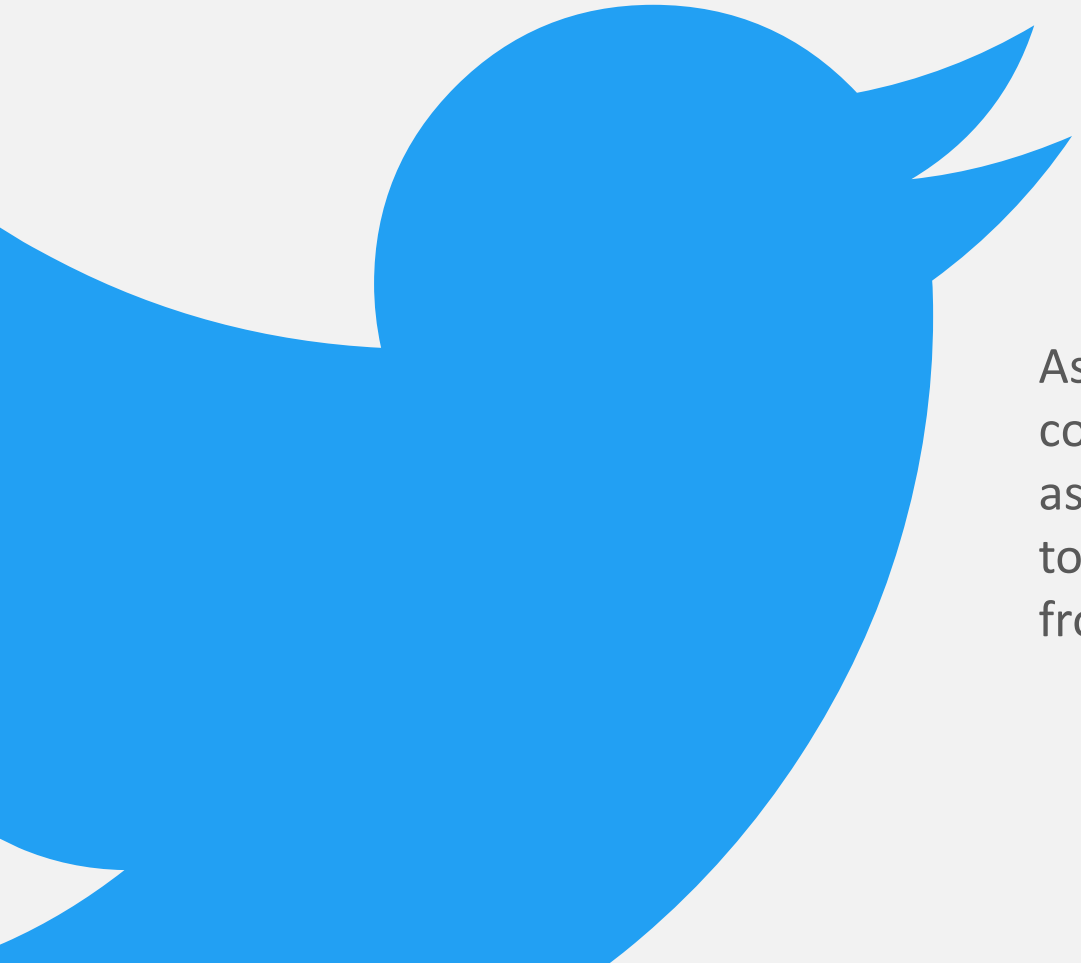
Capstone Project: Comparing Consumer Sentiment of Apple, Google, and Android from the Past and Present on Twitter

By: Sam Lim
Date: 7/28/21



BUSINESS PROBLEM

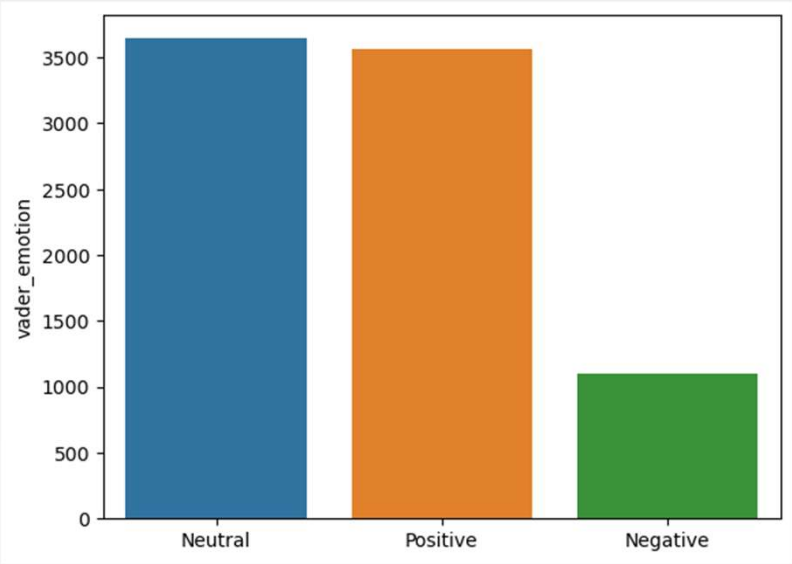
As consumers' perception of these large companies have been changing, Twitter has asked to find how people's sentiments towards these companies have changed from the past using tweets.

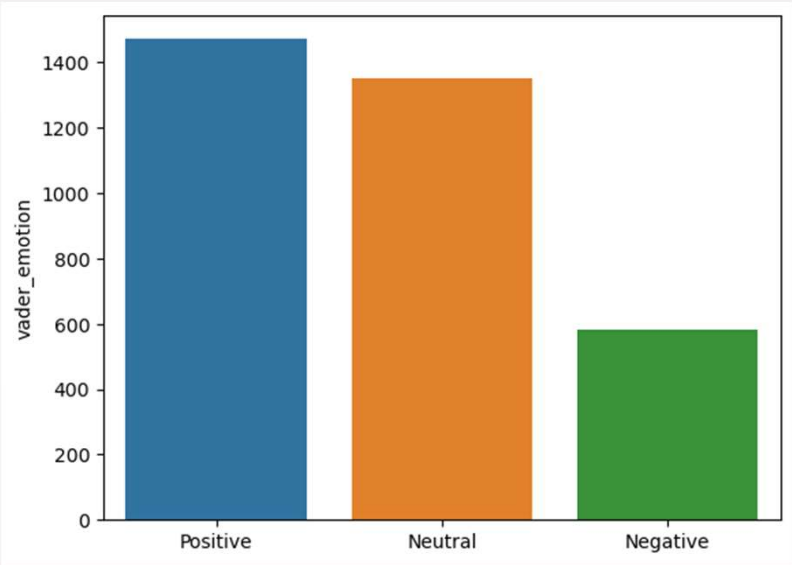
DATA COLLECTION

- Using Tweepy and Twitter's Developer API, I was able to gather around 3,000 current user tweets about these companies.
- For each company, I collected 1,500 tweets, but around 1/3 of the tweets did not contain any information of said companies.
- The data containing older tweets were provided.

Data Imbalance

- For both the old tweets and new tweets, there was an imbalance in the number of the three sentiments.
- The top image represents the number of old tweets with the respective sentiments, and the bottom represents the new tweets with the respective sentiments.
- For both old and new tweets, the number of negative tweets were around a third of the size of the neutral tweets and positive tweets.



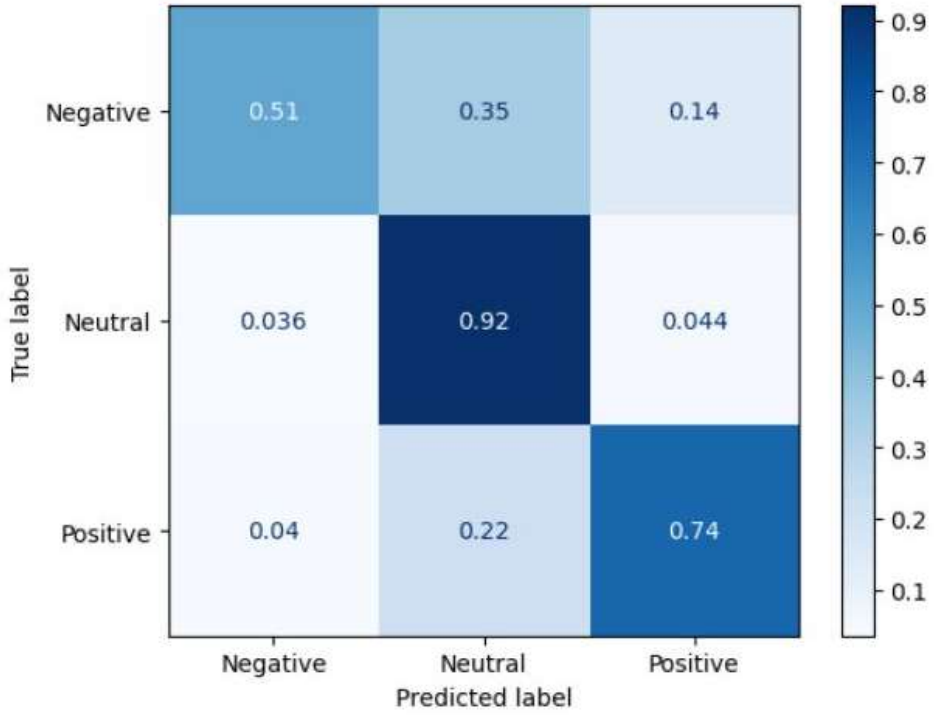


Random Forest
(Old Tweets)

- For each old and new tweets, Random Forest Modeling and Naïve Bayesian Modeling were performed.
- The image on the right shows the confusion matrix of the random forest model after the gridsearch was performed.
- As shown before, the number of negative tweets were far less, which resulted in the low recall score for the negative tweets.

	precision	recall	f1-score	support
Negative	0.68	0.51	0.58	338
Neutral	0.74	0.92	0.82	1091
Positive	0.89	0.74	0.81	1063
accuracy			0.79	2492
macro avg	0.77	0.72	0.74	2492
weighted avg	0.80	0.79	0.78	2492

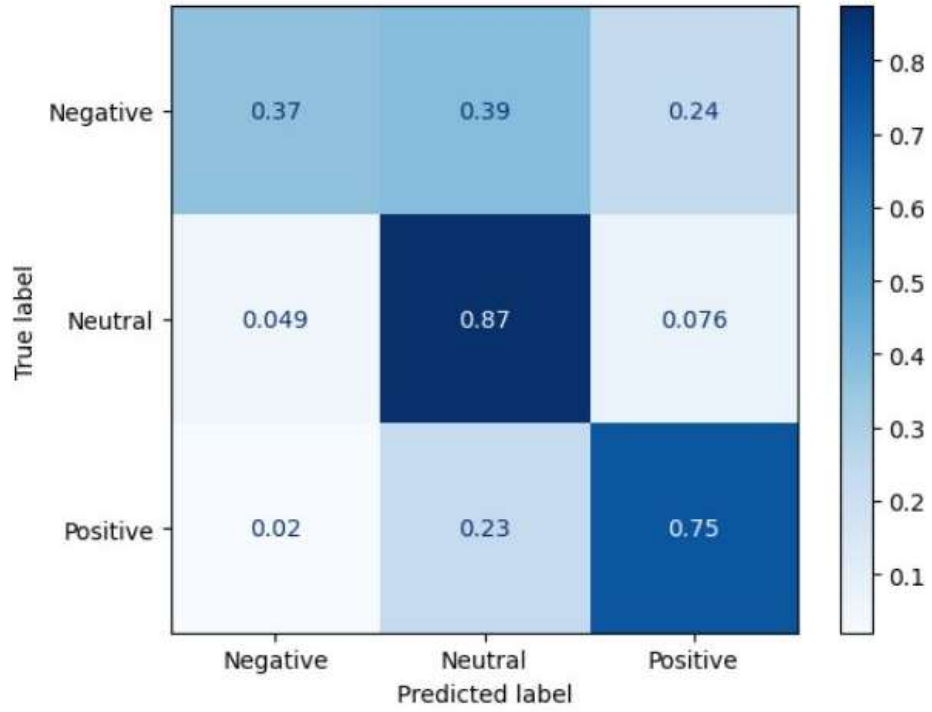
Training Score = 0.95
Test Score = 0.79



Random Forest
(New Tweets)

	precision	recall	f1-score	support
Negative	0.69	0.37	0.48	175
Neutral	0.68	0.87	0.76	406
Positive	0.82	0.75	0.78	441
accuracy			0.73	1022
macro avg	0.73	0.66	0.68	1022
weighted avg	0.74	0.73	0.72	1022

Training Score = 0.99
Test Score = 0.73



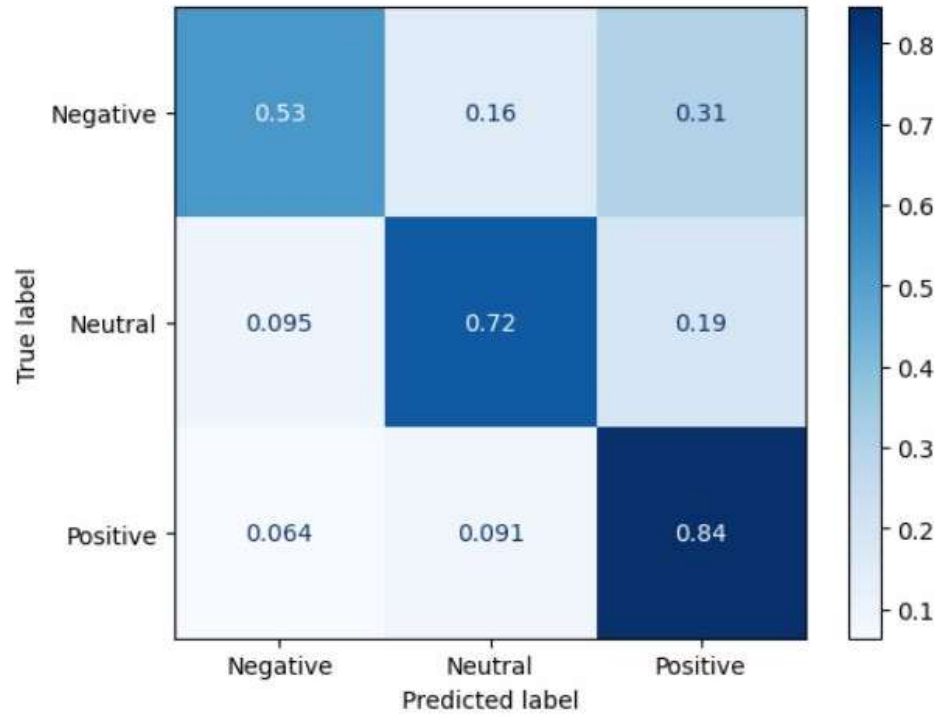
- The recall score of the random forest model with grid search for the new tweets also show the same problem.
- However, I would still recommend this model (random forest with gridsearch) above others because it was able to maintain a relatively high recalls core for the positive and neutral tweets while improving the recall score for the negative tweets.

Naïve Bayesian Model
(Old Tweets)

- The Naïve Bayesian model after the gridsearch also performed quite well.
- However, this model's recall score for the neutral and positive tweets were lower than that of the random forest model, hence why this model is not recommended.

	precision	recall	f1-score	support
Negative	0.51	0.53	0.52	338
Neutral	0.84	0.72	0.77	1091
Positive	0.74	0.84	0.79	1063
accuracy			0.75	2492
macro avg	0.70	0.70	0.70	2492
weighted avg	0.75	0.75	0.75	2492

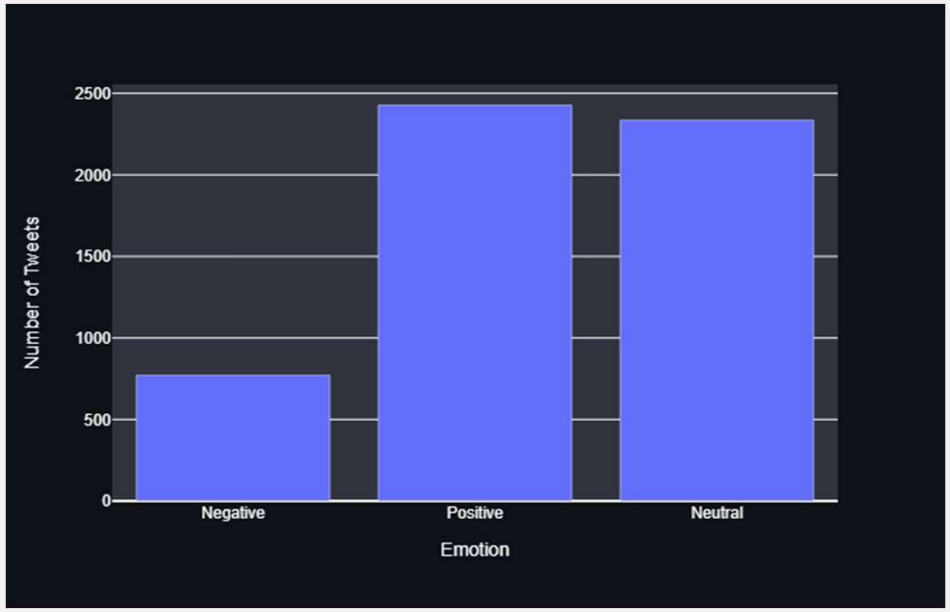
Training Score = 0.95
Test Score = 0.75



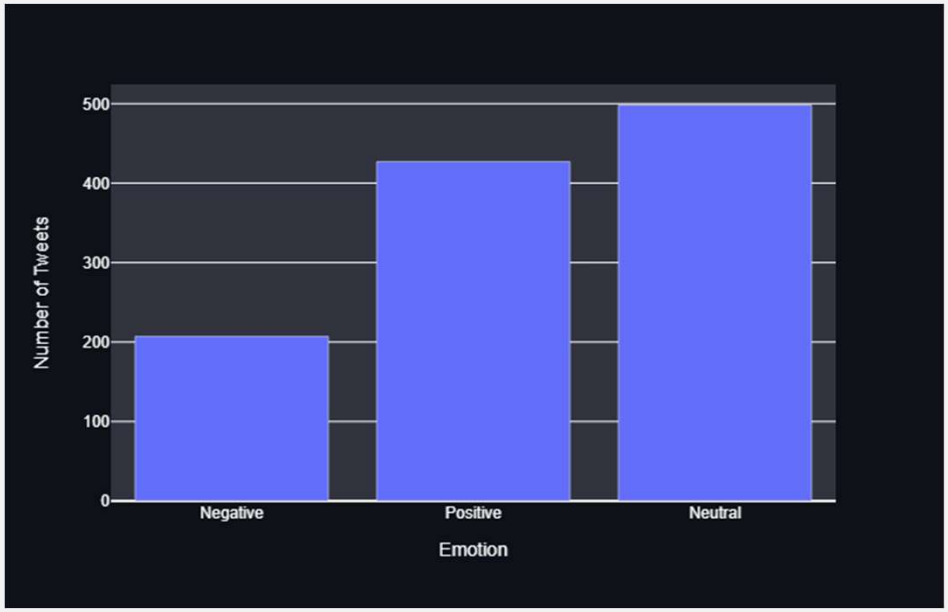
APPLE

- When comparing the user sentiments generated from the old tweets to the newer tweets, we can see an increase in the ratio of negative tweets.

Old



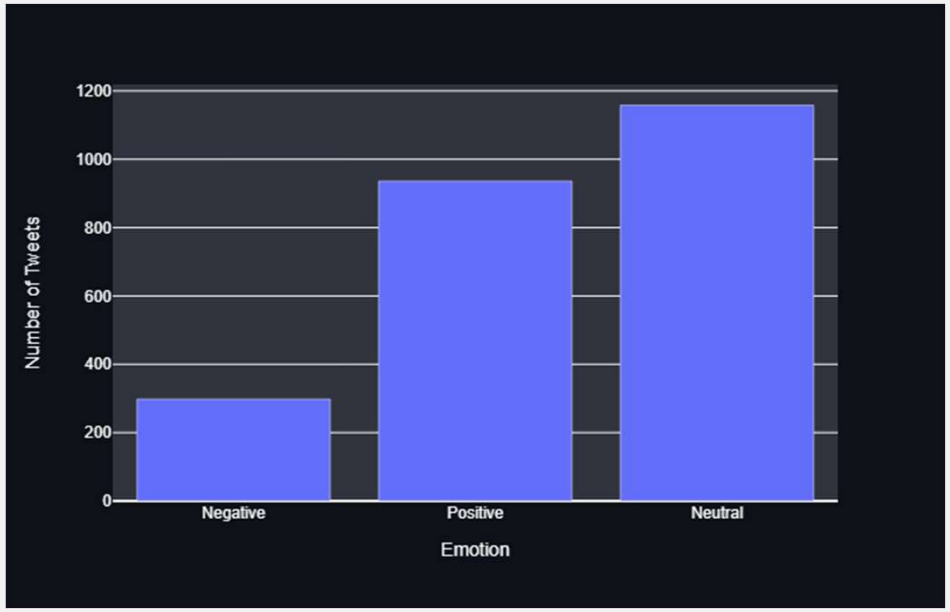
New



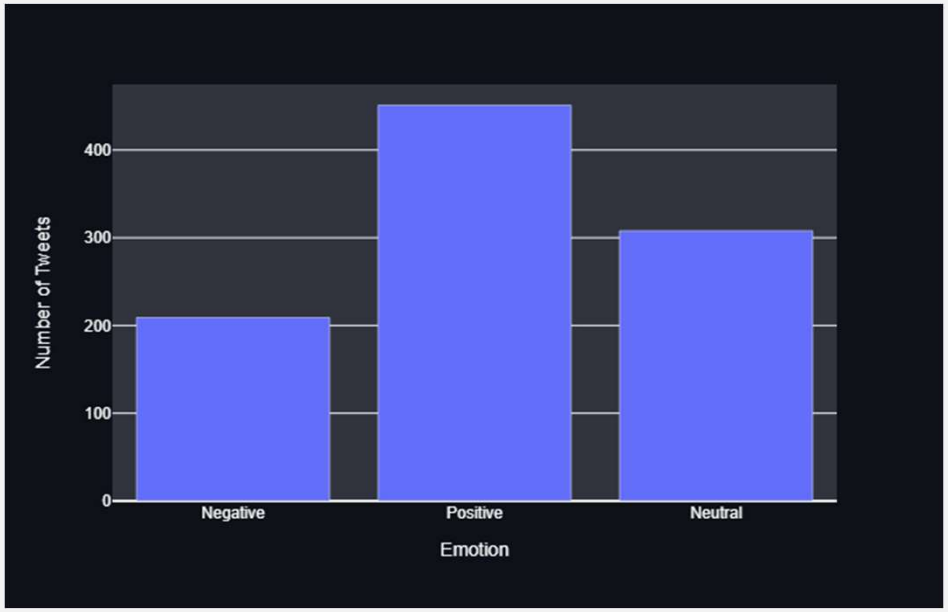
Google

- Sentiment towards Google also became slightly worse.
- While the graphs may seem to represent a positive change, the ratio of negative to positive comments have increased by nearly two-fold.

Old



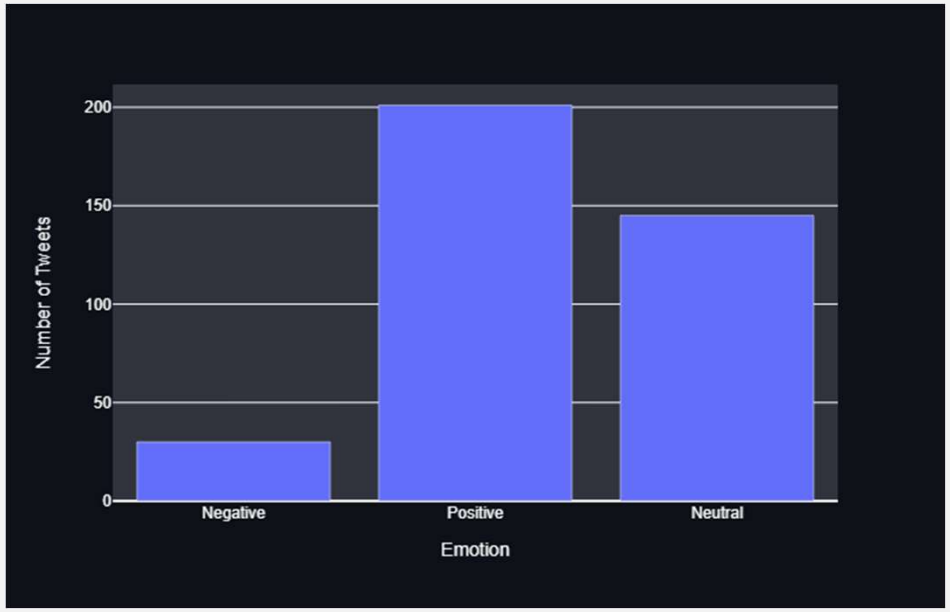
New



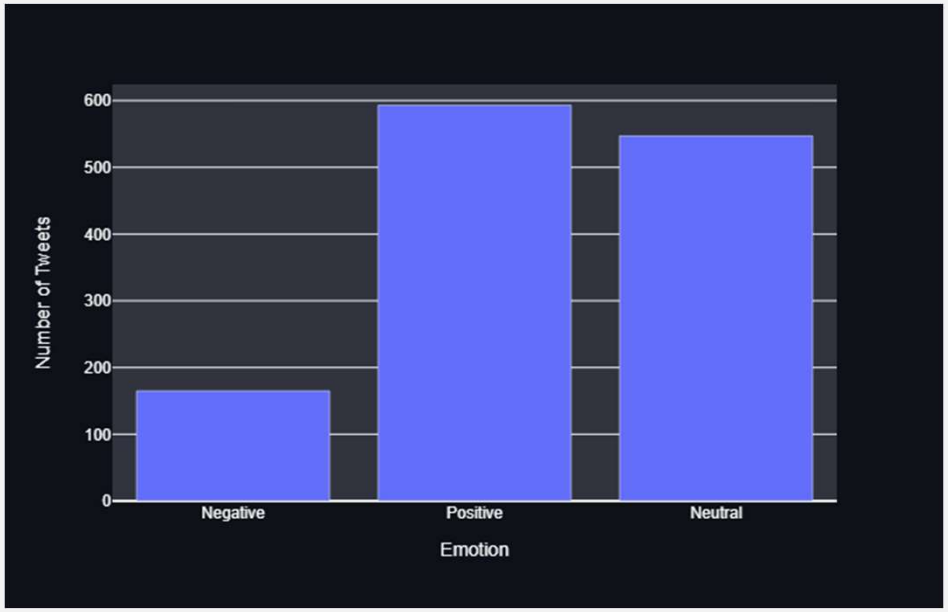
Android

- User's sentiment towards Android has improved significantly.
- The graph may seem to show similar sentiment, the ratio of negative to positive sentiments are not too different, but the total number of positive sentiments has increased.

Old

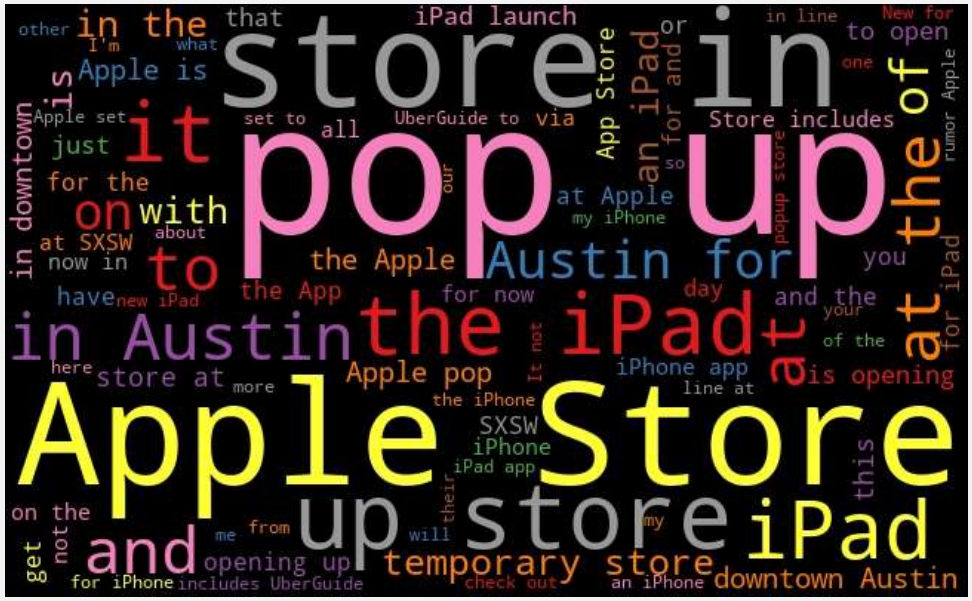


New



Word Cloud Comparison Apple: Negative

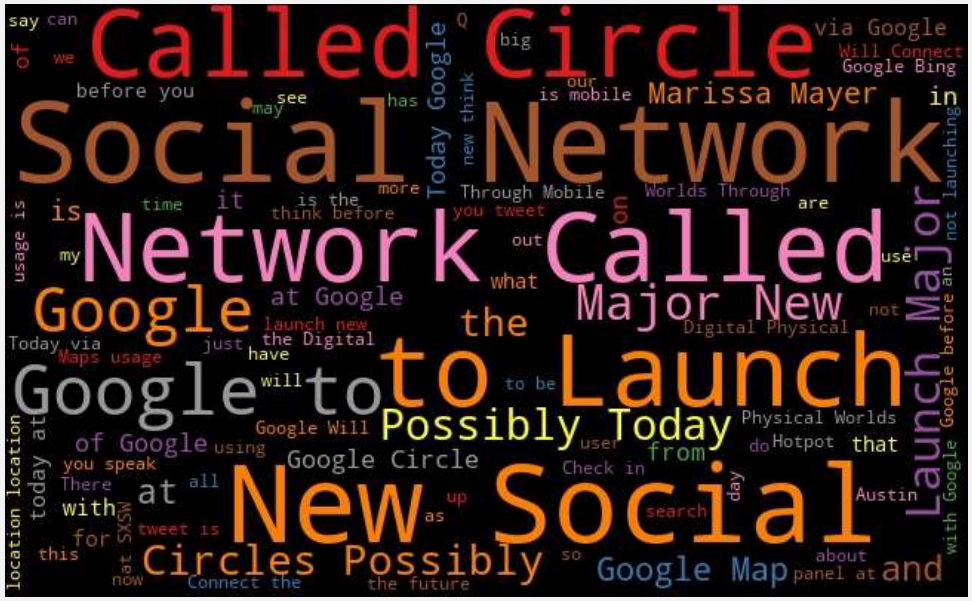
Old



New

Word Cloud Comparison Google: Neutral

Old



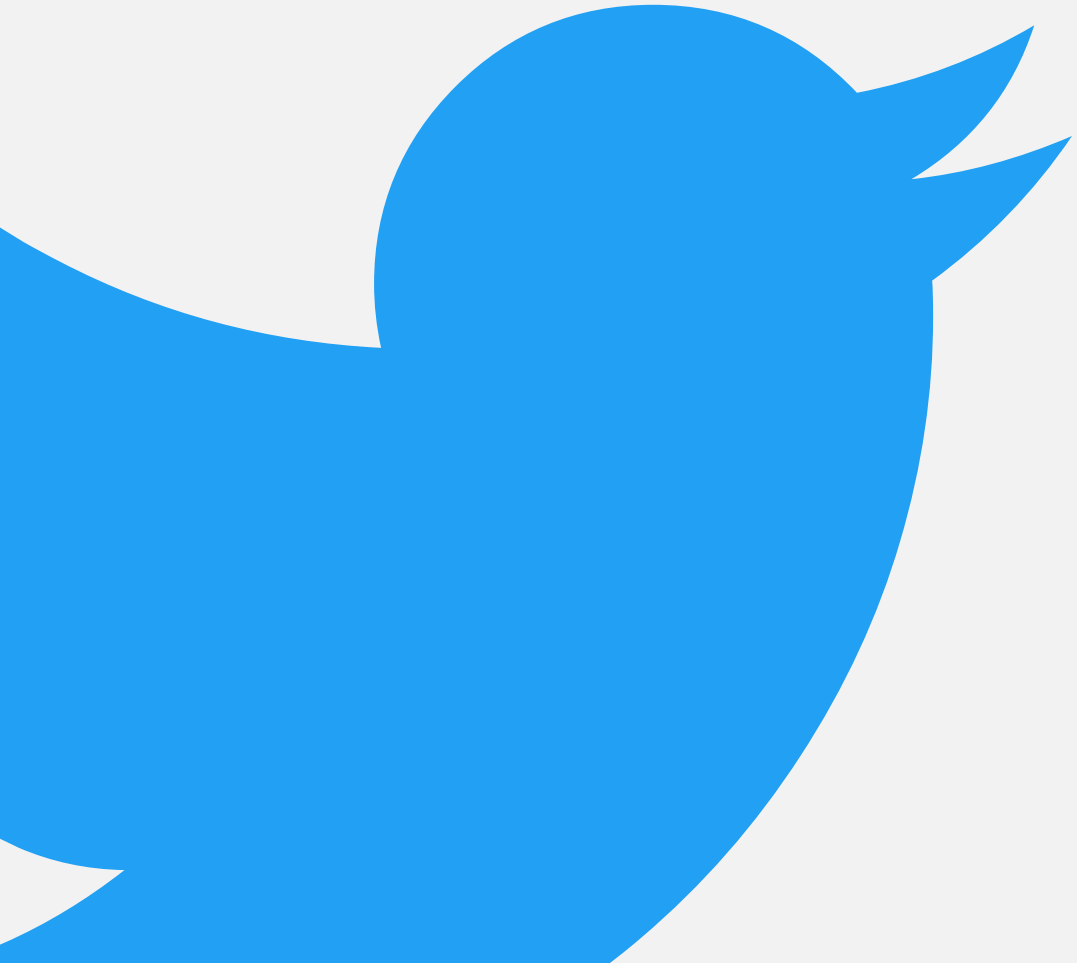
New

Word Cloud Comparison Android: Positive

Old

New

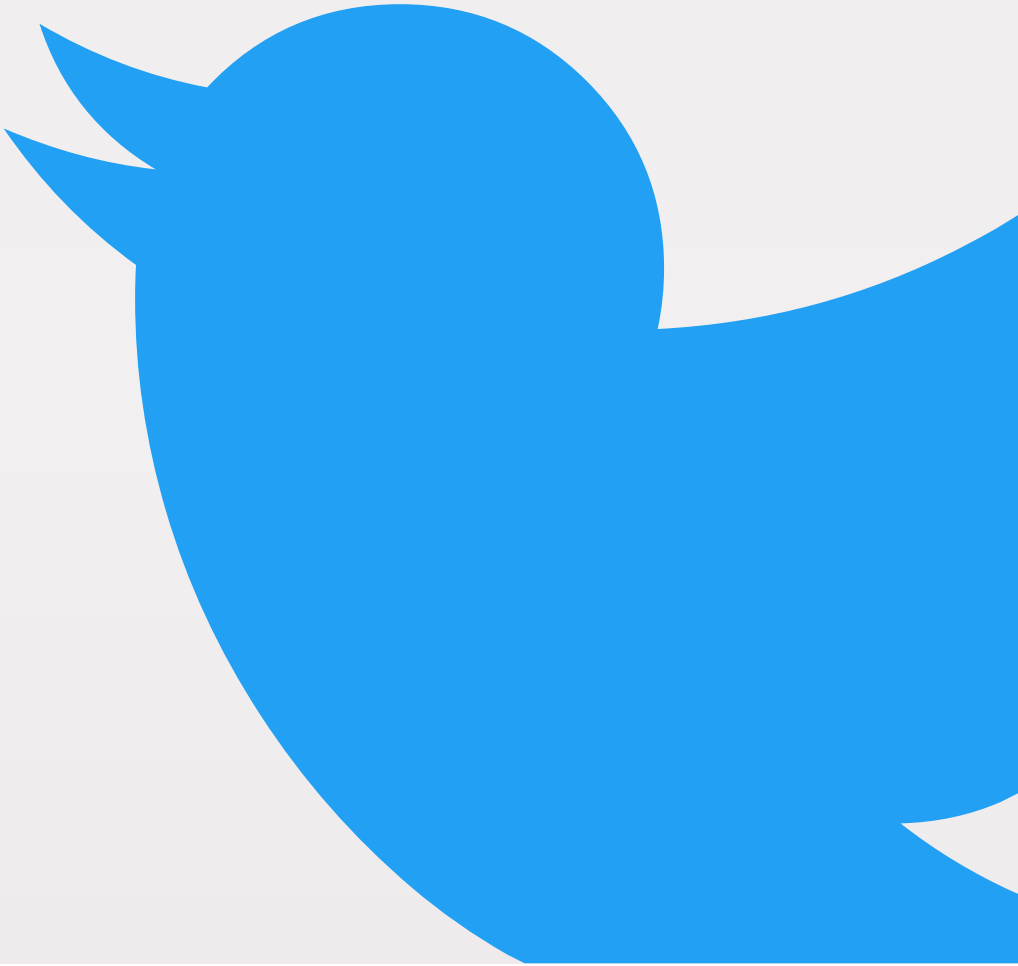
Recommendation



- I recommend using the Random Forest model with gridsearch.
- This model was able to improve on the low recall score for tweets with negative sentiments while maintaining the relatively high recall scores for other tweets.
- Given that the consumer's sentiment towards some of these companies have turned more negative, it could be recommended that these companies start paying more attention to their consumers to retain/improve their relationships with their customers.

Future Works

- Create a timeline that displays continuous changes in public's sentiment towards these companies.
- Add important events to the timelines to have a better understanding of what end consumers' wants and needs.
- Add monthly/quarterly/annual earnings to show the relationship between the rate of increase in earnings to public sentiment.
- Add companies that have continuously improved its relations with the end consumers and report on the changes in their earnings.



Thank You

Q&A