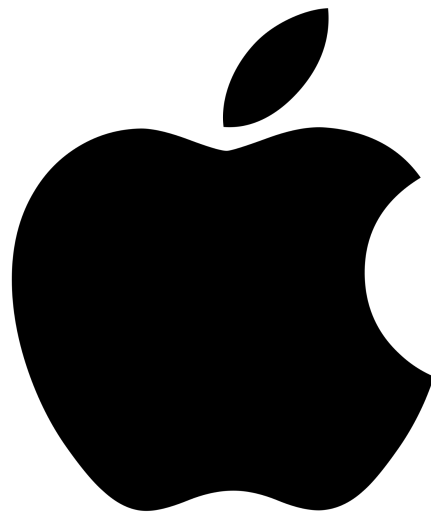


Apple Customer Support Tweets Analysis



Team 5

Mira Daya

Misha Khan

Ankit Jain

Peggy Lee

Yashmini Vuppu

BANA 275

Spring 2021

Table of Contents

1. Overview
2. Dataset
3. Data Preprocessing
4. Exploratory Data Analysis
5. Methodology
 - a. Sentiment Analysis
 - b. Topic Modeling
6. Recommendations

Overview

INTRODUCTION

Started in 1976, Apple is a multinational technology company that designs, develops, and sells consumer electronics, computer software, and online services. After years of successful product launches, such as the iPhone, iPad, Macbook, and more, Apple has built its reputation of being the top competitor in the technology industry.

OBJECTIVE

Because of the company's immense success, our team wanted to observe common pain points among Apple products that their users face. By extracting Apple Customer Support Tweets from @AppleSupport, we were able to generate a list of repetitive complaints that Apple can address to improve their products.

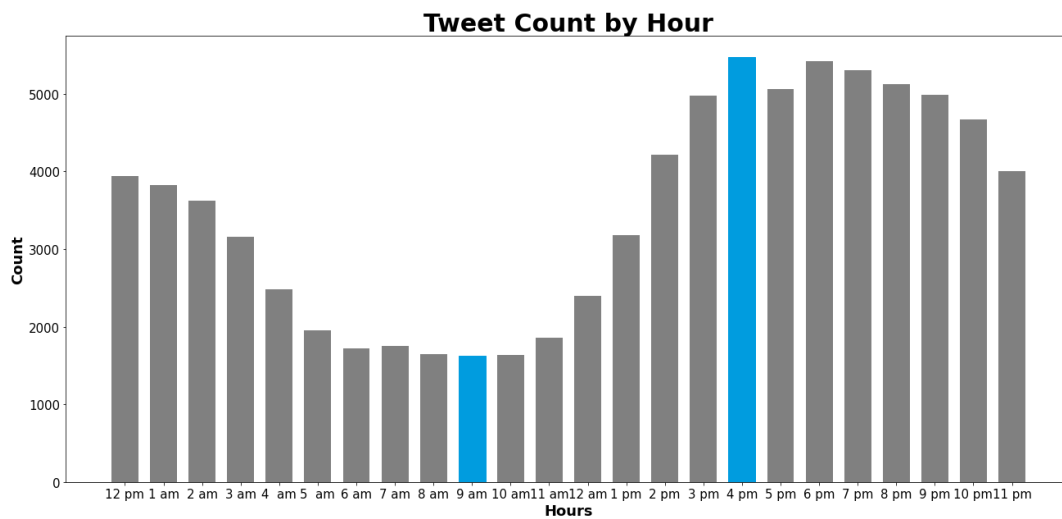
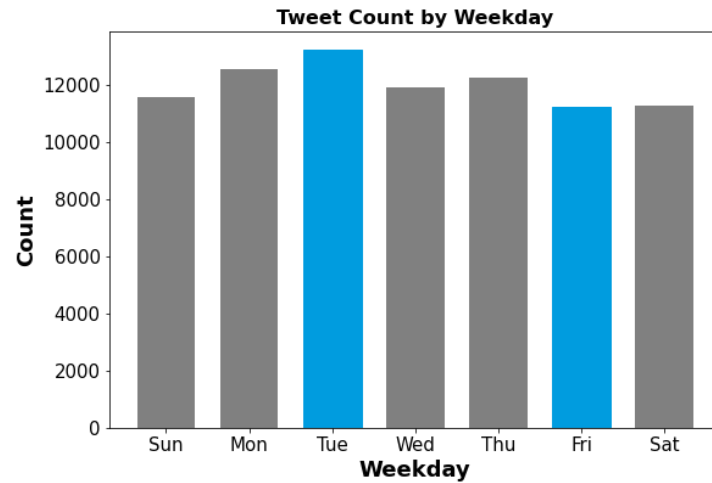
Dataset

We utilized a dataset from Kaggle.com called "Customer Support on Twitter". The dataset's time frame is from March 4th, 2016 to December 3rd, 2017 with a sample size of 95,147 tweets. In order to conduct analysis, we only focused on the following three variables that were meaningful: author_id, date_created, and text.

Data Preprocessing

Before further analysis, we cleaned the text to make our analysis more efficient. First, we converted all date time objects from text data into date data. Next, using the regular expressions and tweet processor package, we removed emojis, mentions (@), URL links, digits, punctuations, and empty tweets. To make it simpler, we converted all text to lowercase.

EDA



We decided to take a look at the distribution of volume of tweets throughout the week by hours of day. We found that the users are reaching out to Apple Support the most on Tuesdays and the least on Fridays and the weekend. Additionally, users are least active during late night and early morning hours. Notice that from 12 PM to 9 AM, there are less frequent tweets to Apple Support, the least amount being at 9 AM. After 10 AM, tweet volume starts to increase, reaching a high at 4 PM.

Methodology

TEXT ANALYSIS

For a better understanding of what issues customers are facing, we used a TF-IDF Vectorizer to investigate commonly occurring words. We set parameters so that a word or phrase must occur at least two times in the corpus of documents ($\text{min_df} = 2$) but cannot occur more than 80% of the entire corpus ($\text{max_df} = 0.8$). The lower your tf-idf score, the more frequent the word or phrase occurs in the entire corpus relative to all other possible words and phrases. We created 4 different types of vectorizers with different values for n-grams, and we got the top 10 commonly occurring words for each case.

	tfidf
iphone	2.707482
ios	2.786113
phone	2.795077
update	3.116456
apple	3.428243
battery	3.637551
new	3.649739
app	3.948400
screen	4.097581
updated	4.142475

Case 1: ($n_gram = (1,1)$)

	tfidf
iphone plus	4.908906
ios update	4.963631
battery life	5.175124
new update	5.360384
apple music	5.419731
new ios	5.500843
iphone ios	5.534560
new iphone	5.624795
updated ios	5.675706
high sierra	5.689818

Case 2: ($n_gram = (2,2)$)

	tfidf
new ios update	6.728261
latest ios update	7.068390
iphone plus ios	7.188098
phone keeps freezing	7.291534
macos high sierra	7.385562
brand new iphone	7.596457
apple watch series	7.650042
ios iphone plus	7.659258
apple pay cash	7.746270
low power mode	7.819601

Case 3: ($n_gram = (3,3)$)

	tfidf
check overall battery health	8.730472
overall battery health iphone	8.815029
check battery health iphone	8.907402
iphone plus running ios	9.296867
updated phone last night	9.396951
new ios update phone	9.396951
update phone keeps freezing	9.396951
battery health iphone plus	9.508176
ios phone keeps freezing	9.568801
latest ios update iphone	9.568801

Case 4: ($n_gram = (4,4)$)

	tfidf
iphone plus	4.908906
ios update	4.963631
battery life	5.175124
new update	5.360384
apple music	5.419731
new ios	5.500843
iphone ios	5.534560
new iphone	5.624795
updated ios	5.675706
high sierra	5.689818

Case 5: ($n_gram = (2,4)$)

From the vectorizer results, we see that most tweets are about the iPhone, specifically the iPhone plus. Additionally, many tweets are related to the battery health of Apple devices and the latest iOS updates. There were also a few tweets that mention macOS High Sierra software.

TOPIC MODELLING

The following section goes over the different types of modeling methods that were used in topic modeling.

Model 1: Non-Negative Matrix Factorization (NMF)

NMF Models is an unsupervised learning method that discovers underlying patterns in text and identifies words that are representative of a group of documents, or in our case, a group of tweets. Since our Case 5 vectorizer presented the most information about phrases that were commonly occurring in the tweets, we employed that for this model. After trying out different values for n_components, we realized that Case 10 gave us a list that had the most distinctive words. These were the topic results we obtained with the NMF Model:

- Topic 1: iphone plus
- Topic 2: iphone ios
- Topic 3: ios update
- Topic 4: apple music
- Topic 5: ios iphone
- Topic 6: new update
- Topic 7: high sierra
- Topic 8: plus ios
- Topic 9: battery life
- Topic 10: updated ios

The results from the NMF model provided similar results of the vectorizer. Both displayed topics such as iPhone Plus, iOS update, battery life, High Sierra. However, Apple Music is an additional topic mentioned in several tweets.

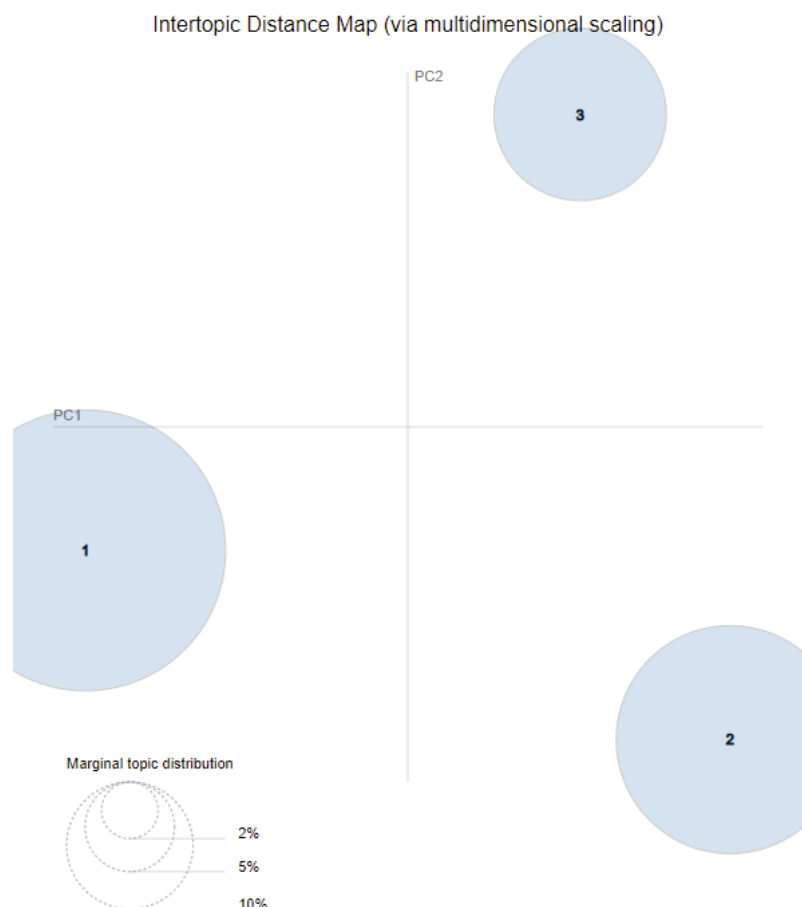
An issue with the NMF model is that some phrases that are representative of the group or topic are being repeated. These phrases are shown in a reversed fashion. For example, Topic 2 and Topic 5 are essentially the same topic repeated again. Since we are not able to make a full distinction between these groups, we proceeded to try out the next model (LDA).

Model 2: Latent Dirichlet Allocation (LDA)

Similar to NMF, LDA is also an unsupervised learning method. The key difference between the two is that it reverse engineers the process of finding topics. Instead of finding ‘n’ topics, it assumes that there are already ‘n’ topics existing and finds the words most commonly associated with those ‘n’ topics. This allows for more intertopic distinction. We tried out different scenarios with the number of topics. Initially, we started with 10 to match the results we had with the NMF model. But we noticed that to get more distinct topics, we needed to reduce the number of topics we filled in. The coherence score is the best parameter to understand which number of topics is most ideal to get the best results. The higher your coherence score, the better the distinction between topics.

Number of Topics = 10	Number of Topics = 5	Number of Topics = 3
0.2541281575790699	0.2405731550131318	0.2768282042923937

The best coherence score was obtained with the number of topics = 3, and the inter-topic distinction can be seen here:

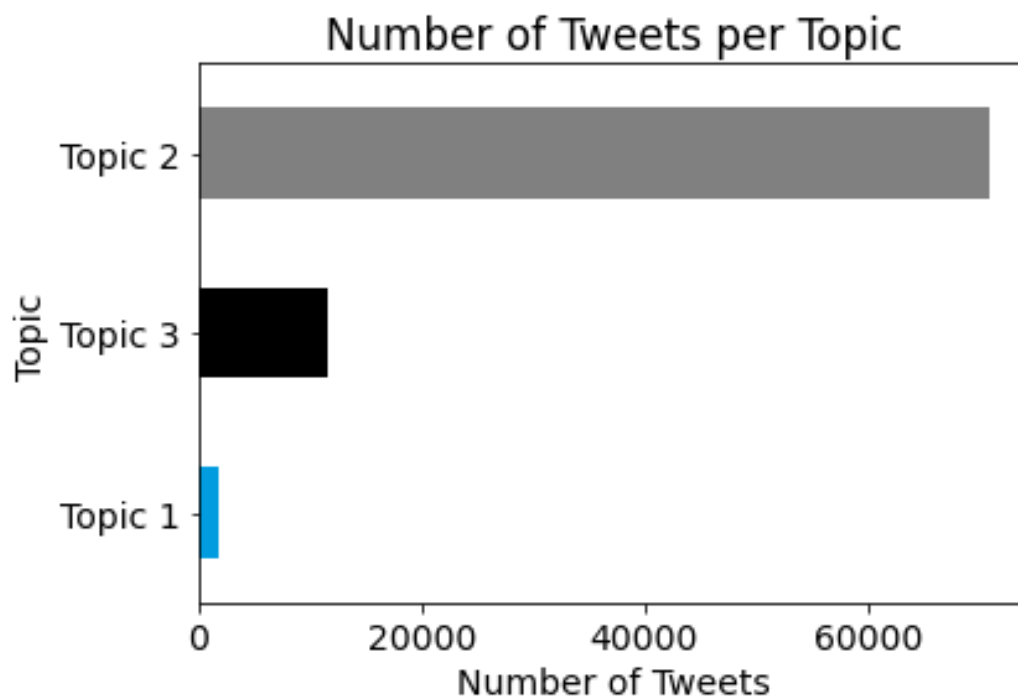


These were the topics which were obtained by the LDA Model:

```
[ (0,
  '0.068*"apple" + 0.024*"music" + 0.022*"amp" + 0.018*"want" + 0.016*"charge" '
  '+ 0.015*"store" + 0.015*"itunes" + 0.014*"buy" + 0.014*"account" + '
  '0.013*"month"'),
  (1,
  '0.055*"iphone" + 0.051*"phone" + 0.044*"update" + 0.031*"io" + 0.026*"app" '
  '+ 0.020*"new" + 0.019*"use" + 0.019*"issue" + 0.016*"battery" + '
  '0.013*"happen"'),
  (2,
  '0.028*"try" + 0.020*"work" + 0.020*"screen" + 0.019*"keep" + 0.017*"back" + '
  '0.016*"turn" + 0.014*"need" + 0.014*"day" + 0.014*"call" + 0.013*"restart"') ]
```

The following are inferences about the three topics above:

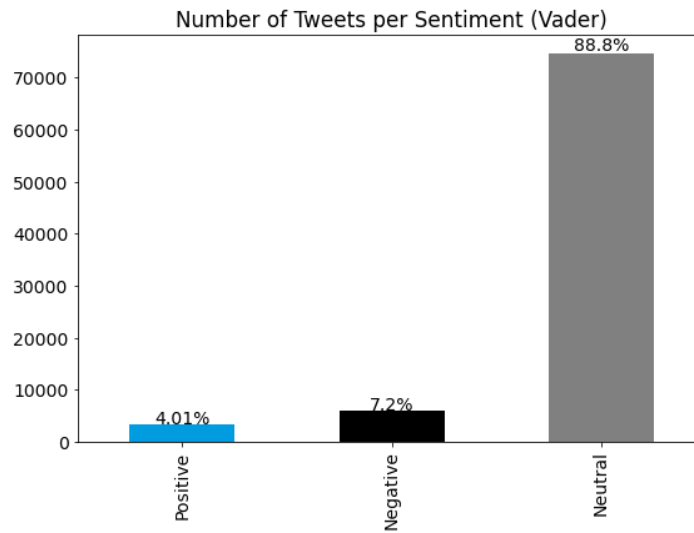
- **Topic 1:** users accounts, and iTunes and App Store charges and purchases issues
- **Topic 2:** iOS software, iPhone software issues
- **Topic 3:** mainly hardware issues



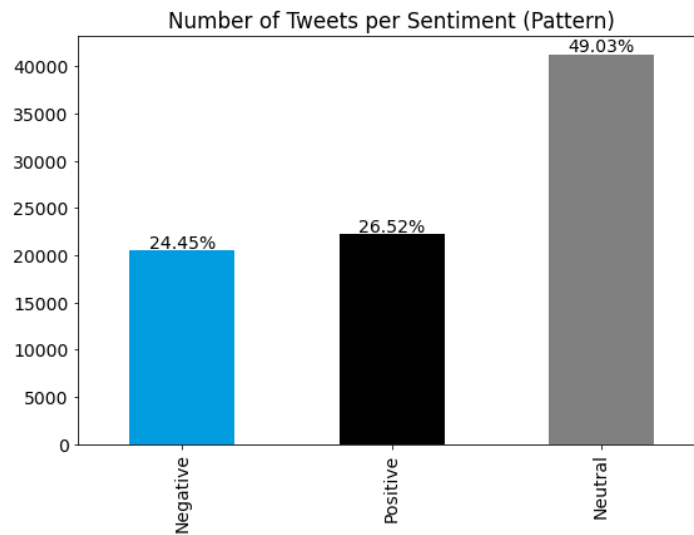
SENTIMENT ANALYSIS

We used three different classifiers to extract information about the sentiment of a tweet. We used the following:

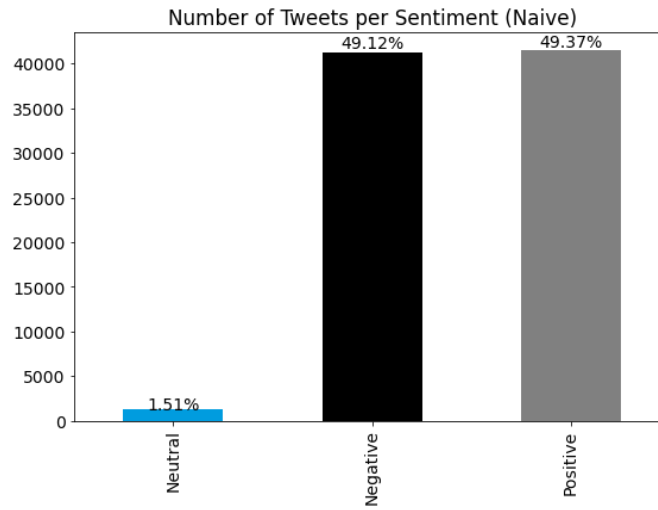
1. Vader Sentiment Intensity Analyzer



2. Textblob Pattern Analyzer



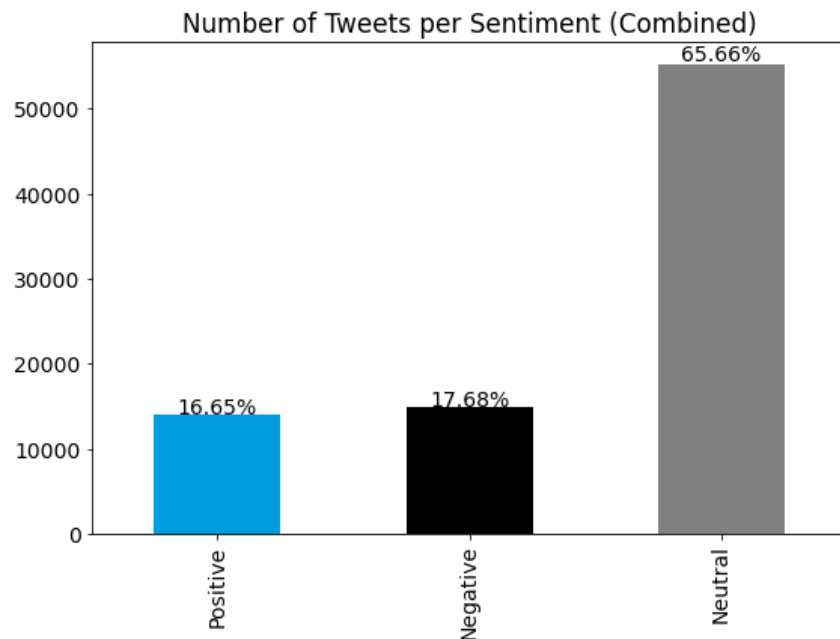
3. Textblob Naive Bayes Analyzer



Overall, the Naive Analyzer performs better than Vader Sentiment and Textblob Pattern Analyzer. It was able to capture both positive and negative sentiment but not the neutral. Instead, Vader and Textblob Pattern Analyzer does a better job of understanding neutral sentiment.

We then combined the results of the three classifiers, here are the following results:

From the graph, we see that most tweets seem to be neutral. This makes sense because users are simply trying to resolve their issue, therefore being a combination of positive and negative resulting in neutral sentiment. There is a higher degree of negative sentiment as opposed to positive. We believe that is due to frustration that customers express in their tweets when having concerns about their products.

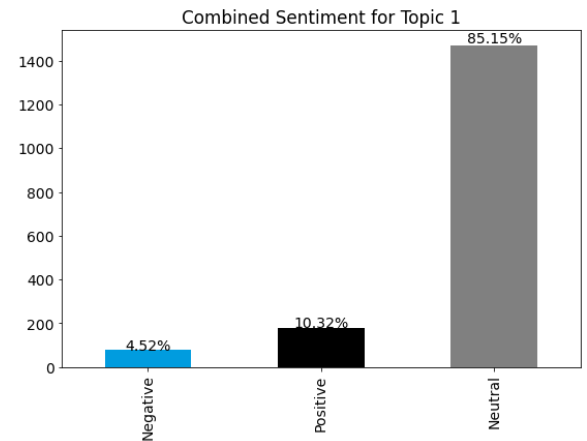


Next, we analyzed each topic and its sentiment:

Topic 1, Account Related Issues:

The following graph shows us that most tweets are neutral (85.15%) for tweets that are account based. We found that it was interesting that more tweets tend to be positive, which is possibly due to the fact that this issue is easier to solve and customers respond in a more positive tone. For example:

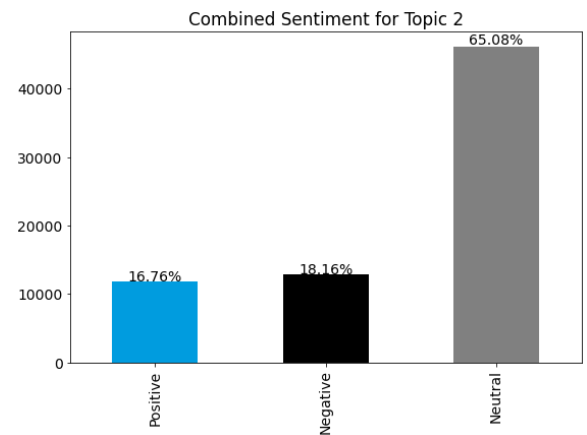
“@AppleSupport thank you for sorting out my phone issue, it took time but I am a happy customer. Also the staff member in the Trafford store that helped was excellent. She needs a promotion 😊. Good job 👍”



Topic 2, Software Related Issues:

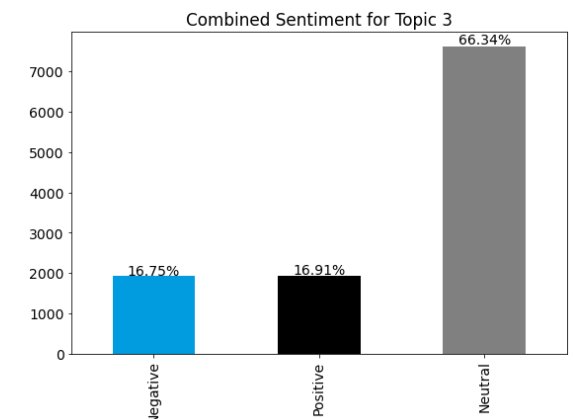
Here, there seems to be an overall neutral sentiment for software related issues (iOS update) but, there are more negative sentiment tweets compared to positive ones. From reading some of the tweets, this could be due to the fact that most users do not receive a response in a timely manner and turn to other resources for help. For example:

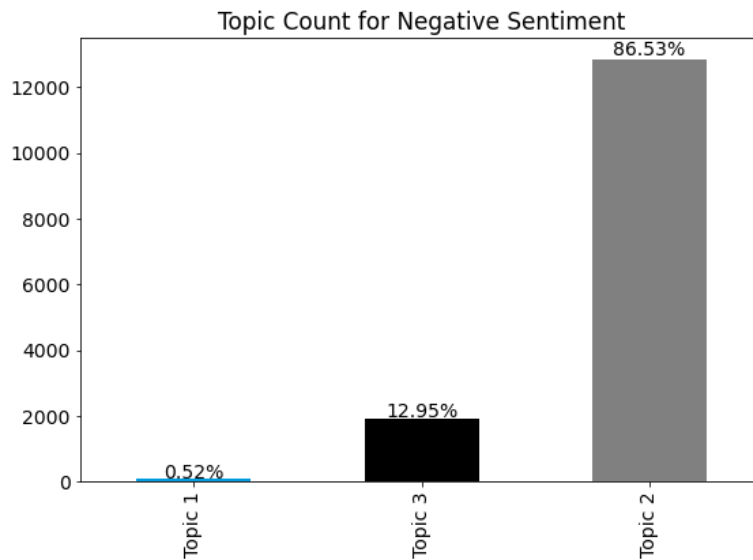
“@AppleSupport I figured it out, without help of customer service. That was a waste of time.”



Topic 3, Hardware Related Issues:

Similar to the previous two cases, the neutral sentiment is most prominent for hardware related issues. Interestingly, the number of positive and negative sentiment tweets are very close to one another.



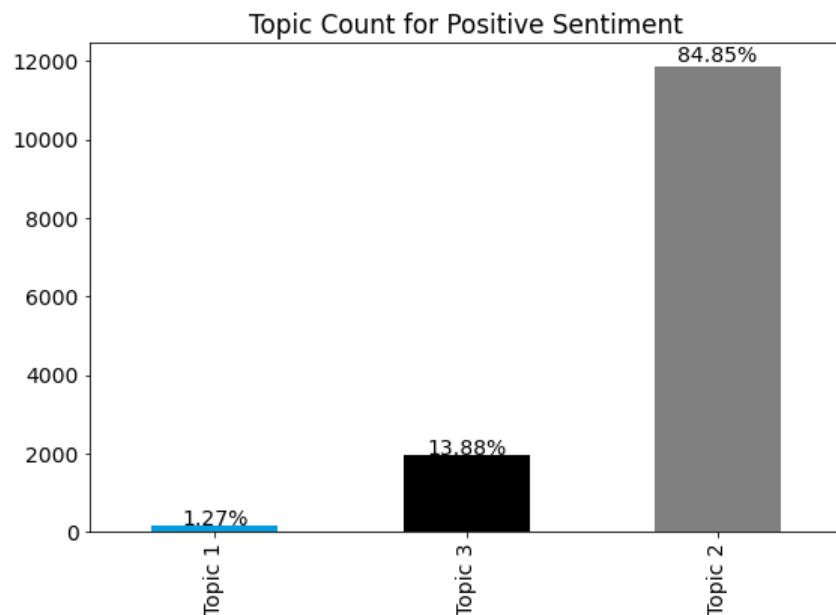


From the graph above, notice that the majority of negative sentiment tweets to Apple Support occurred within Topic 2 (Software). Looking specifically at tweets associated with the negative sentiment, the most commonly occurring words based on a tfidf vectorizer are:

Note that the top occurring issues with negative sentiment include battery life, ios update, and new update which verifies that Topic 2 has the most tweets with negative sentiment.

It seems that users are reaching out to Apple Support with mainly three concerns under Topic 2: battery life, iOS update, and new update. Previously, we saw that Topic 2 has the most tweets with negative sentiment. This can be a signal for Apple to consider improving battery life for their products as well as creating a smoother experience for iOS updates.

	tfidf
battery life	4.262548
ios update	4.556695
new update	4.834987
iphone plus	4.975217
new ios	5.181478
update phone	5.525024
hard reset	5.609216
new iphone	5.708589
updated ios	5.731231
iphone battery	5.746616



From the graph above, notice that the majority of positive sentiment tweets to Apple Support also occurred within Topic 2. Looking specifically at tweets associated with the positive sentiment, the most commonly occurring words based on a tfidf vectorizer are:

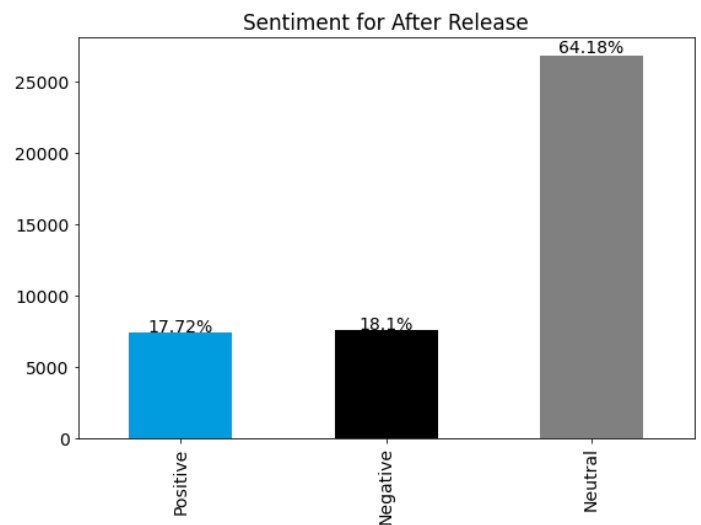
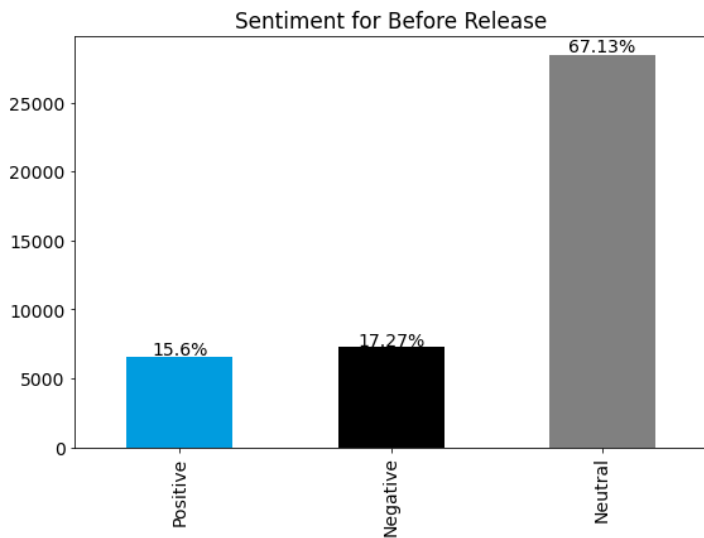
Note that the top occurring issues with positive sentiment include high sierra, new iPhone, and new ios which verifies that Topic 2 has the most tweets with positive sentiment.

A handful of these positive tweets seem to be associated with MacBook, such as MacOS High Sierra, and new products which users might have purchased.

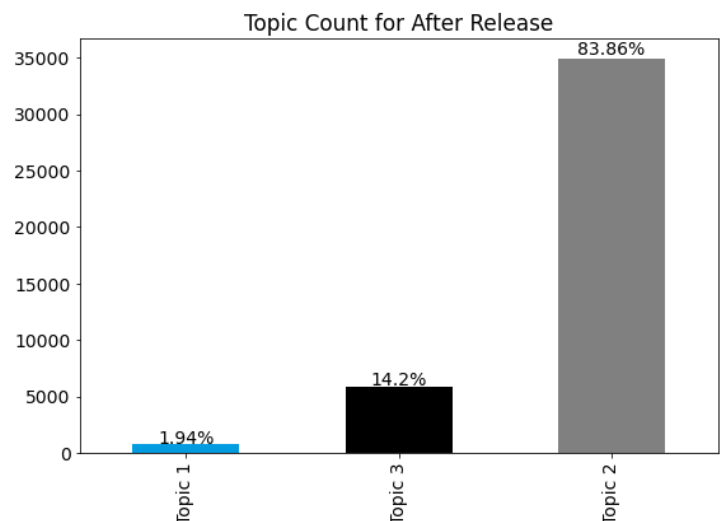
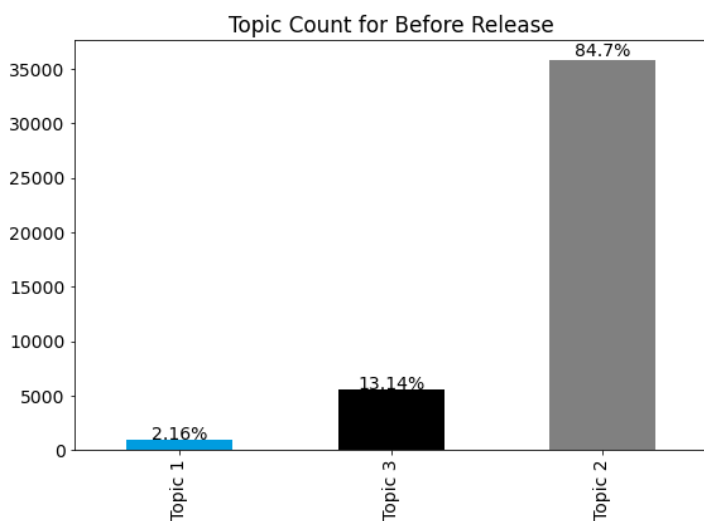
	tfidf
high sierra	4.243694
new iphone	4.611418
new ios	4.922295
new phone	5.099575
new update	5.171034
latest ios	5.253008
iphone plus	5.263109
apple music	5.315204
macbook pro	5.458716
ios update	5.483717

Analysis of Tweets with Product Releases

The iPhone X release date was on November 3rd, 2017. We analyzed the change in tweets before and after the release of the product to see what issues were being addressed by customers upon the release of a prominent new product. The main takeaways are summarized below.



The most noteworthy change in sentiment after the release of the iPhoneX is that there was a slight increase in positive sentiment tweets to Apple Support.



There were more noticeable changes in topic count before and after the release of the iPhone X versus sentiment changes. There was an increase in tweets related to Topic 3, hardware issues, and slight decreases to Topic 2, software issues, and Topic 1, account issues.

Before Release:

	tfidf
ios update	4.722875
iphone plus	4.769907
battery life	4.945285
new update	5.173325
new ios	5.182570
apple music	5.395702
iphone ios	5.458964
updated ios	5.546347
ios iphone	5.642102
apple watch	5.713198

After Release:

	tfidf
iphone plus	5.071392
ios update	5.281957
new iphone	5.425458
apple music	5.447749
high sierra	5.464278
battery life	5.474750
new update	5.590692
iphone ios	5.617041
updated phone	5.823934
updated ios	5.826924

Looking closely at what was being mentioned in tweets, we can see that issues related to the iOS update and the iPhone plus were both prominent before and after. One big takeaway from this analysis is that after the release of the new phone, battery life related issues became less occurring while topics like Apple music and new iPhone started showing up more frequently.

Overall, it seems that with the release of the iPhone X, there were people reaching out to Apple Support for issues related to this new iPhone as well as the new iOS update that comes after the release of new iPhones. Apple music moved up in frequency while battery life moved down.

Analysis for Products

The following section dives deeper into Apple Support tweets for specific Apple Products.

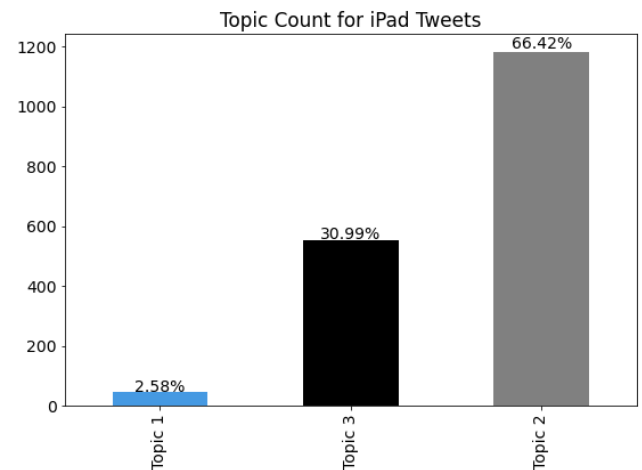
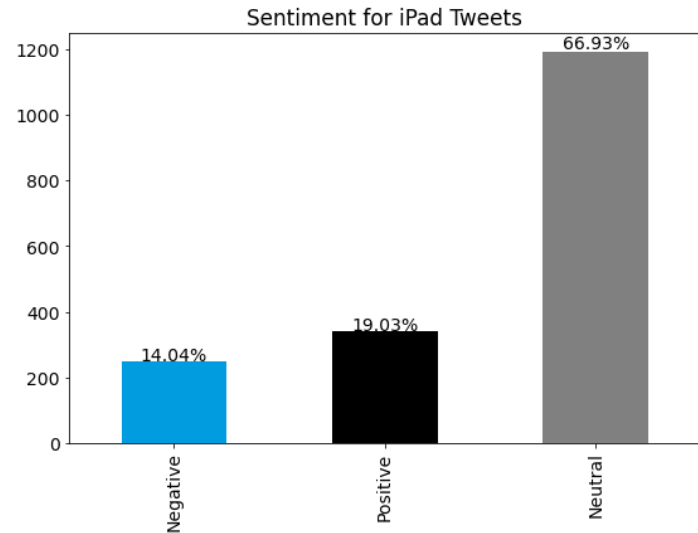
Product: iPad

From this graph displaying sentiment for iPad tweets, the sentiment was more neutral and positive rather than being negative. We found that the most prominent issue is that after the latest iOS update, the iPad's battery life significantly decreased. Many tweets were related to this issue, for example:

"@AppleSupport Upgrading to 11.0.2 has killed my iPad battery life. Didn't happen in iOS 10. Is there a fix coming soon?"

For topic count, the highest number of tweets were related to software (Topic 2), then hardware (Topic 3), and lastly account (Topic 1). Again, this is confirmed by the above analysis where we found the most prominent issue to be battery life and the iOS update.

To further determine top iPad issues, the TFIDF vectorizer confirms that battery life and latest updates are top occurring issues within the iPad related tweets. Notice "app store" also appears in iPad related issues which is actually part of Topic 1.

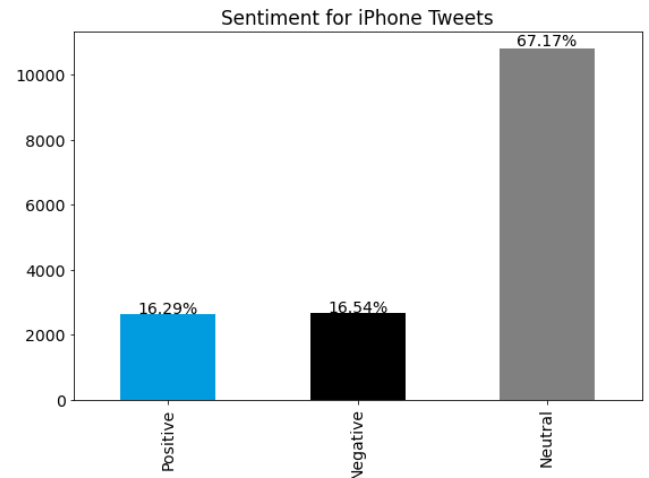


tfidf	
battery life	5.227395
app store	5.307438
latest update	5.394449
last update	5.712903
smart keyboard	5.777441
works fine	5.846434
stopped working	5.920542
software update	6.000585
photo library	6.087596
connect itunes	6.087596

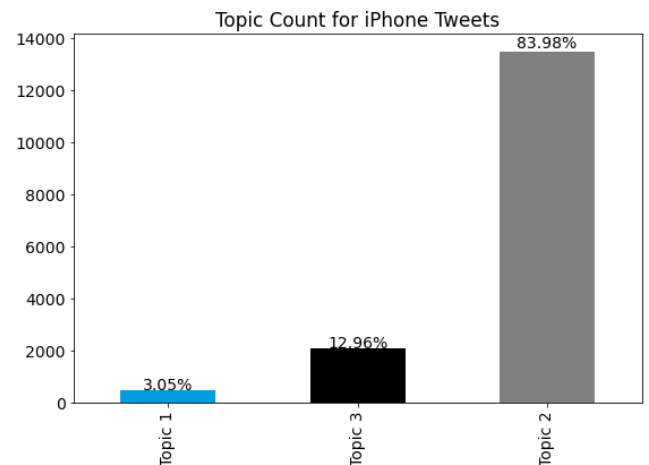
Product: iPhone

From this graph displaying sentiment for iPhone tweets, overall the sentiment is neutral with positive and negative almost being equal. During our analysis, we noticed that the iOS update was a frequently occurring issue. This seems to be common for both iPad and iPhone as we noticed in the above iPad analysis. For example:

“@AppleSupport Battery life on my iPhone 7 taken a hit after 11.0.2 update. It was much better on iOS 10, New iOS 11.0.2 on iPhone7+. Status bar no longer shows if Bluetooth connected & does not show connected device battery life!”



For topic count, the highest number of tweets were related to software (Topic 2), then hardware (Topic 3), and lastly account (Topic 1).



To further determine top iPhone issues, the TFIDF vectorizer displayed that battery life and latest updates are top occurring issues within the iPhone related tweets.

tfidf	
battery life	4.693805
latest update	5.221438
battery drain	5.833239
last update	5.864988
software update	5.873085
keeps freezing	5.881248
lock screen	5.923095
battery draining	5.931679
battery health	6.031309
battery drains	6.185460

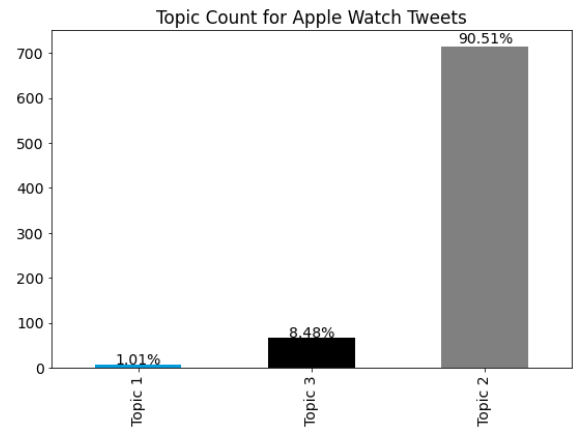
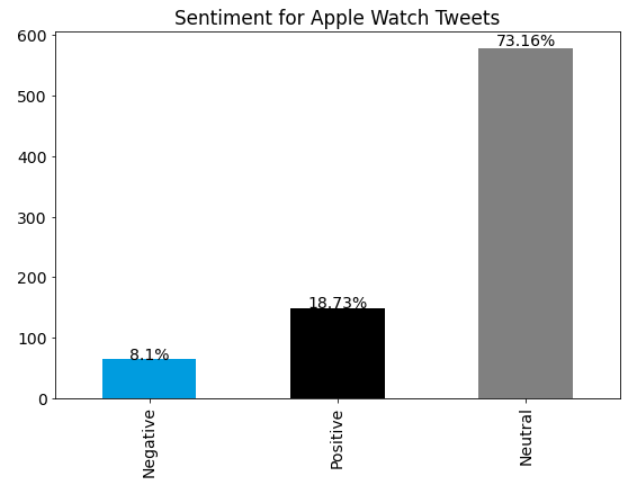
Product: Apple Watch

For the Apple Watch though a lot of people had issues with icons not showing up on the screen. This is the main issue with the Apple Watch and an example of it is:

“@AppleSupport how come the app icons on my Apple Watch come up as blank white squares instead of the actual app icon?”

For topic count, the highest number of tweets were related to software (Topic 2), then hardware (Topic 3), and lastly account (Topic 1). Again, this is confirmed by the above analysis where we found the most prominent issue to be app icons and activity apps.

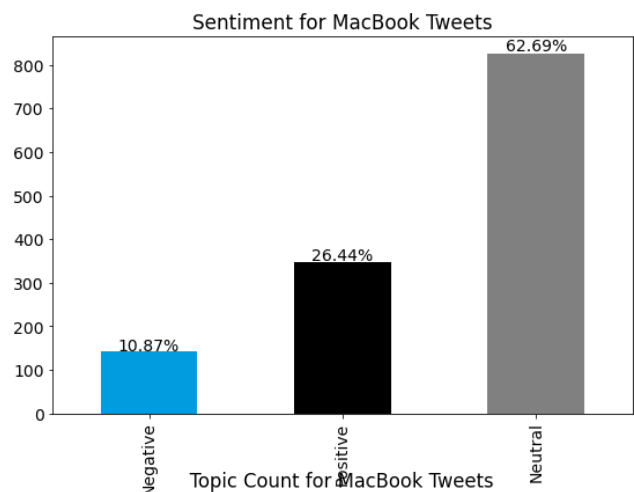
To further determine top Apple Watch issues, the TFIDF vectorizer displayed that app icons and activity apps are top occurring issues within the Apple Watch related tweets.



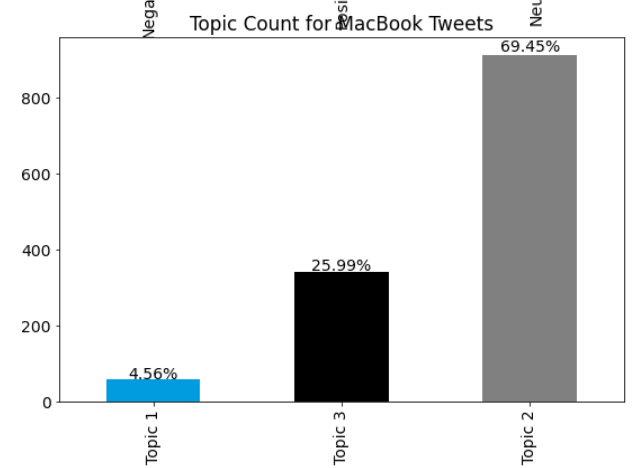
tfidf	
app icons	5.275403
activity app	5.370713
battery life	5.476073
last update	5.476073
workout app	5.593856
latest update	5.727388
music library	5.727388
music app	5.881538
activity data	5.881538
heart rate	5.881538

Product: MacBook

From this graph displaying sentiment of Macbook tweets, we can see that the majority of tweets have a neutral sentiment, following with a significantly high positive sentiment than negative.



For topic count, the highest number of tweets were related to software (Topic 2), then hardware (Topic 3), and lastly account (Topic 1).

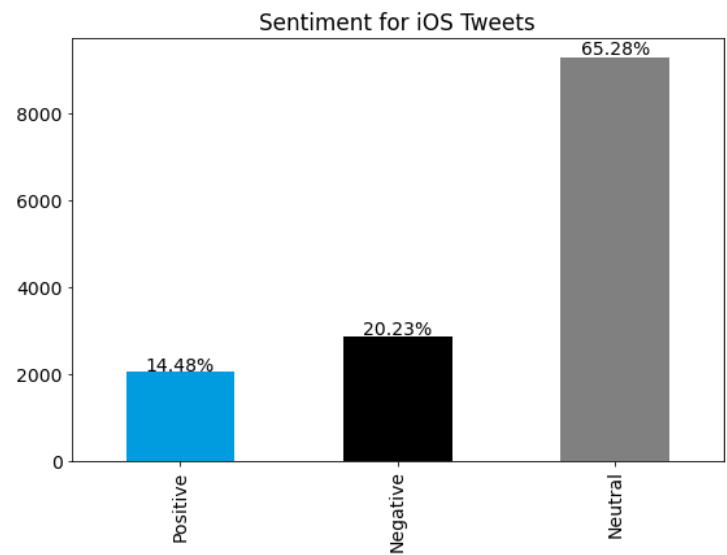


To further determine top Macbook issues, the TFIDF vectorizer resulted: touch bar, hard drive, and stopped working as the top occurring issues within the Macbook related tweets.

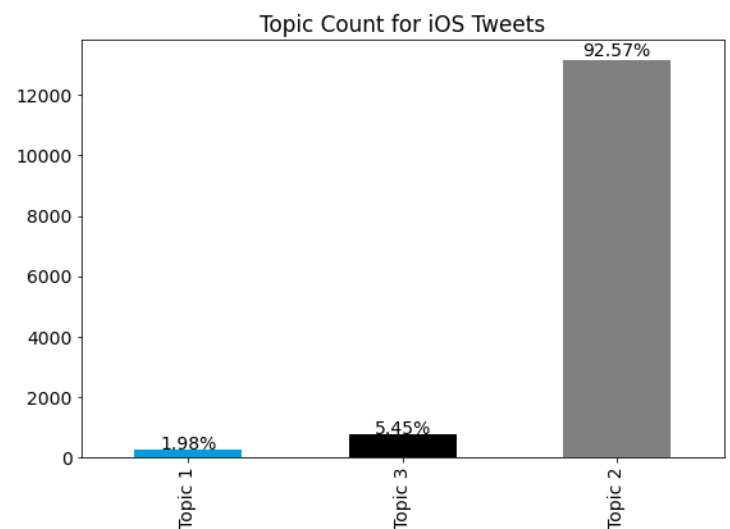
tfidf	
touch bar	5.238673
hard drive	5.238673
stopped working	5.544054
upgraded high	5.698205
updating high	5.698205
weeks ago	5.880527
genius bar	5.880527
month old	5.880527
keep getting	5.985887
logic board	5.985887

Product: iOS

From this graph displaying sentiment of iOS tweets, we can see that the majority of tweets have a neutral sentiment, following with a higher count of negative sentiment than positive.



For topic count, the highest number of tweets were related to software (Topic 2), then hardware (Topic 3), and lastly account (Topic 1).

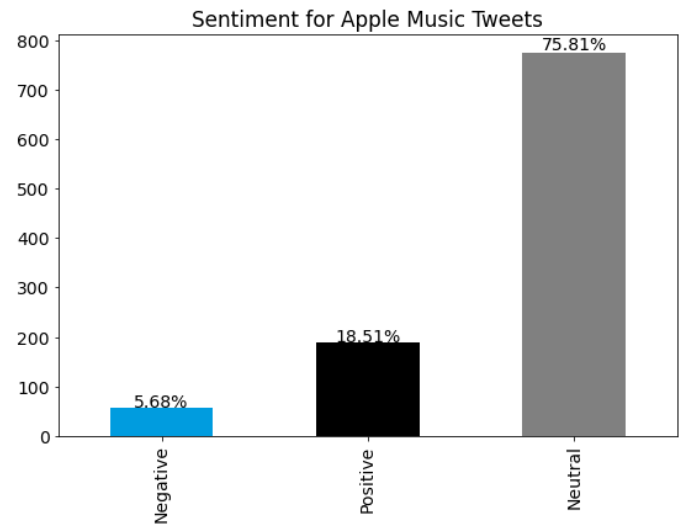


To further determine top iOS issues, the TFIDF vectorizer resulted: battery life, battery drain, and lock screen as the top occurring issues within the iOS related tweets.

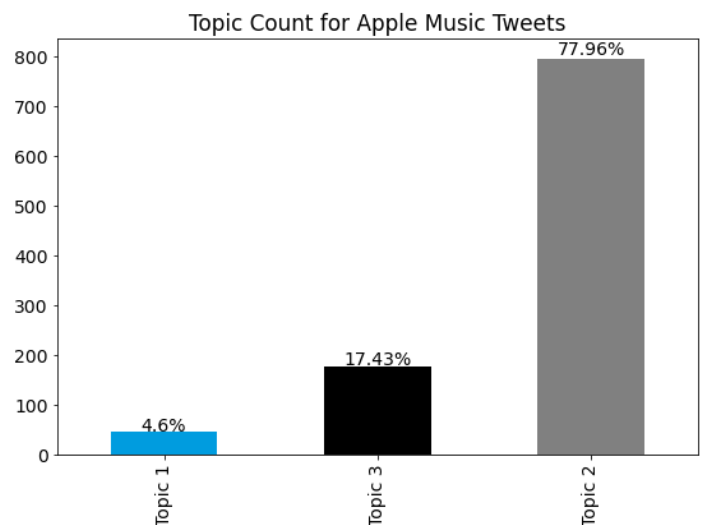
tfidf	
battery life	4.164794
battery drain	5.120971
battery draining	5.404333
lock screen	5.628915
draining battery	5.809799
phone keeps	5.809799
keeps freezing	5.818457
battery drains	5.818457
app store	6.097481
phone freezes	6.314894

Product: Apple Music

From this graph displaying sentiment of Apple Music tweets, we can see that the majority of tweets have a neutral sentiment, following with a high positive sentiment compared to negative sentiment.



For Apple Music, it seems most issues are related to Topic 2, which is software issues. Topic 3 is the second most related issue, hardware issues, then Topic 1 with account issues.

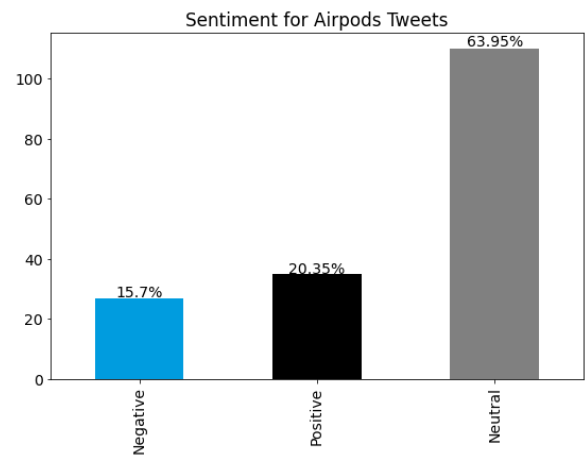


In terms of word frequencies, it seems that “songs downloaded”, “keeps crashing”, “updated iOS” all show up which are indeed related to Topic 2, software issues. Interestingly enough, the most occurring area has to do with membership to Apple Music which is an accounts related topic.

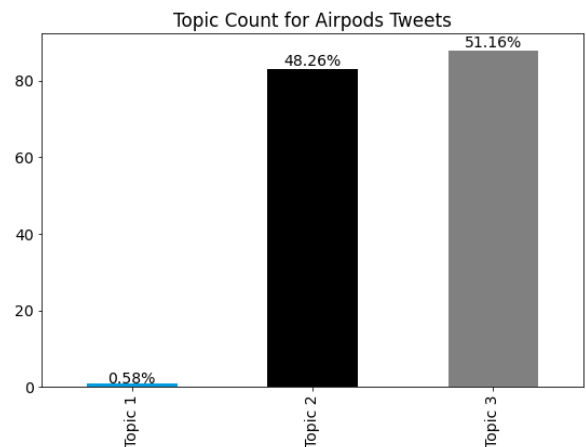
	tfidf
already member	5.221467
songs downloaded	5.290459
keeps crashing	5.364567
even though	5.364567
icloud library	5.444610
updated phone	5.444610
updated ios	5.531621
iphone plus	5.626932
free trial	5.626932
download songs	5.626932

Product: AirPods

From this graph displaying sentiment of AirPods tweets, we can see that the majority of tweets have a neutral sentiment, following with a high positive sentiment compared to negative sentiment.



For AirPods, it seems most issues are related to Topic 3, which is hardware issues. Topic 3 is the second most related issue, software issues, then Topic 1 with account issues.



To further determine top AirPods issues, the TFIDF vectorizer displayed that iPhone Plus and left one are top occurring issues within the AirPods related tweets.

	tfidf
iphone plus	4.361532
left one	4.361532
apple watch	4.361532
charging case	4.543854
left airpod	4.766997
control center	4.766997
playing music	4.766997
anyone issues	4.766997
play audio	4.766997
apple products	4.766997

Recommendations for Apple:

- Apple Support currently operates from 5AM - 8PM Pacific Standard Time, whereas the latter half of the day has more volume of tweets so it's important to have more staff support during times when users are most active. We propose after 9AM to be the opening time since the most active time for tweets were at 4PM.
- Based on the text analysis, we also recommend that Apple looks into improving the post update experience because that was a pending issue in most tweets. This could be done with more intuitive experiences and explanations with possible cognitive walkthroughs and run-throughs of the changes made with the update, for easier usability for customers.
- Lastly issues such as battery life and iOS experience still have not been improved which can be seen in the two examples of the tweets provided. We went through a lot of tweets in the past few days and after the release of the new iOS update, and most of the tweets had almost exactly the same issues which were being addressed back in 2017. If Apple works to fix these issues, it will possibly increase their customer satisfaction and sales as well.