Analyzing User Sentiment Towards Industries and Brand

Multimedia University

Cyberjaya

Submitted to

Ms. TS. DR. GOH HUI NGO

Lecturer

Prepared by

Name	Student ID	Group
Ong Pei Yong	1171100655	TC1V
Ng Zhen Xiong	1181302692	TC1V
Arpita Bhattacharijee	1171301208	TC1V

Date of Submission: 7 November 2021

Table of Contents

Assignment 2

Introduction

The industry we are investigating is the mobile phone industry and our target brand is Samsung. Along with Samsung, other famous brands in the industry are Apple, Oppo, Huawei and Xiaomi. This question will focus on the sentiment analysis of these brands.

The data we collected are one week before 26th of October 2021 because Twitter only allows to query tweets one week before and only in one week the amount of tweets is huge already and it is sufficient for the analysis.

The collected tweets are then pre-process first before doing the analysis. We perform pre-processing like removing stop words, numbers, mentions, emoji, links, hashtags and frequent words to the data we collected.

Q1. What is the opinion of users towards the industry?

To determine the user's opinion towards industry we collect the data by keyword search which is “phone /smartphone/mobile” in order to keep the context of the data within this industry. After extracting the data we had 50k data. The extract was from 29 October data. As the topic domain for the phone industry is so big, this is why every day there are million tweets generated only towards the phone or mobile industry.

After data extraction we performed data cleaning and data pruning. During data cleaning, we filter out unnecessary syntax, punctuations, emoji’s, links, stop words etc. to prepare the data for further better analysis. After cleaning the data, still the data left with 50k.

The data amount is vast and there are different types of comments towards many entities. It is usually not easy to generalize the comments as an opinion without any context or aspect. Aspect based opinion mining involves extracting aspects of features of an entity and identifying out the opinions about those aspects. Opinions with entities are helpful for analysis but opinion about aspects of those entities are more granular and insightful. This is why we performed aspect based opinion mining to detect the aspect specific sentence of users in other word the actual opinion towards the industry.

In the next step, we applied post tagging method to get the noun phrases from the data. A noun phrase is basically a noun where it pulls all of the stuff that surrounds and modifies the noun like prepositional phrase, adjectives etc. It gives the idea that the text might be "about". Once we detect all the noun phrases of each text data, we apply compactness pruning on data in order to trim those noun phrases where they do not exist in meaningful word definition or word length is less than two. Now, the data is left with 26572 noun-phrases. We remove the duplicates from the noun-phrases features, since having duplicates does not help the analysis and in some cases it gives biased results. Finally the data left with 11878 noun-phrases.

Now we have the noun-phrases which will be used as word features for next analysis. Then we used these noun-phrases to iterate over the whole tweets which were extracted initially and count the amount of repeating of those noun-phrases over the tweets. After that, we generated our word feature with count.

Since our data is still more than 11k, we want to prune our data more to take out the most insightful data. We applied a threshold value, where the value is average of counts of each word feature and applied on the word features. After that, we take out the top 30 count of word features as the final most frequent word feature.

Once we have the final word-features, we apply these word features as filters over the tweets to detect only those tweets which are surrounding or relating to these word features. Now we have the aspect level sentence. (Fig: 1). The figure shows the top aspect level sentences. Regarding phones, the most comments were related to screenshot, watching movies on the phone at night, phone charge, and activities using the phone. To identify more meaningful and relevant context, we decide to draw the graph of word frequency.

```
['daily routine use phone charge phone use phone charging',
 'send calls another user cell phone press button multiple transfer buttons phone know user available phone check website voip
voipservices callhandeling',
 'pull screenshot phone rub ypur phone persons called networking digital age',
 'dumb shit watch movie phone night phone dead',
 'rainbow garden mobile phone photo',
 'day collide chocar inktober inktober pencils paper cellphone smartphone mobile phone street walk girl boy style art illustrat
ion ink blackandwhite biticol staedler fineliner',
 'mini crossbody phone purse mobile sling bag cross body purse linen padded phone sleeve chezvies etsy',
 'dc get lawyers phone lawyers phone looking bag',
 'hi poppy please send us direct message case number looked please include full post code phone number email address reg number
thank mt',
 'sorta keep refreshing roblox login screen phone daughters phone talk stress x']
```

Fig: 1(Aspect Level Sentences)

We want to look at the word frequency to see which words are repeated most often in the comments. Word frequency helps in the sense of giving a quick sight of data such as how people feel or their opinions within the context. We visualize the word frequency [Fig: 2] by bigram words as bigram helps to show the relation between words and get better insight. After plotting the graph on the first ten word frequency “phone” was most repeated which is also obvious since most of the extracted data contain this word. This is why we did not consider the first ten words frequency and plot second ‘top thirty’ words frequency.

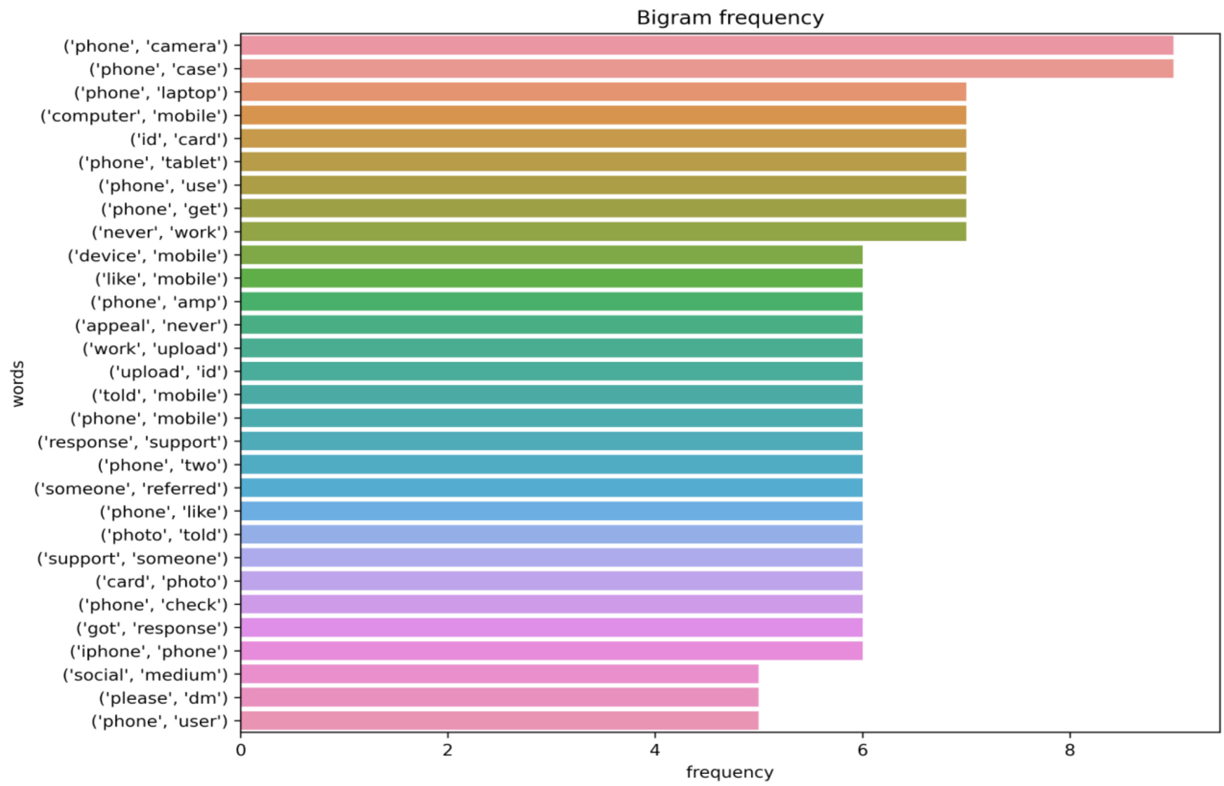


Fig: 2(Word Frequency of Aspect Level Sentences)

From figure 2 we can see the most repeated words are phone, camera, and case which are basically properties of a phone. Beside that we can also know the existence of words user, upload, social, work, dm (direct message) which possibly mean the usage of phone in daily life and sharing the life activities. Another interesting matter is that we found the only brand name ‘IPhone’ which also possibly depicts the most popular topic or expressed opinions towards iPhone. Finally we can conclude that the most opinions towards the mobile industry are mostly related to the properties of phones and sharing the usage of phone activities within the range of data we collected.

Q2. Which are the brands that are popular among users within the industry?

We will be analyzing the repeating count of each brand name over the collected data. We believe that the most popular brand will be in topic for most of the time, in other words they will appear in the topic for most count .Hence, the count of brand name appearance will help to reach the conclusion towards the popularity. The count of appearance for each brand name is shown in Fig: 3

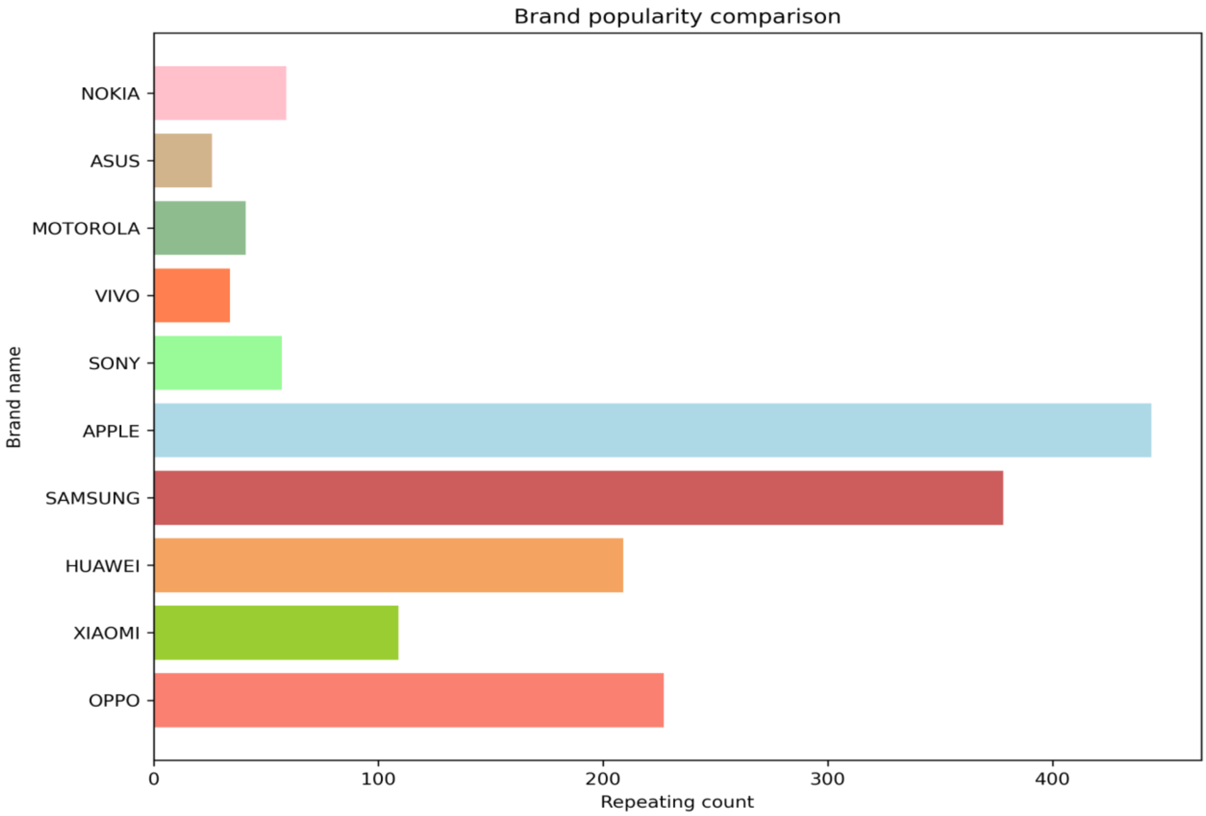


Fig: 3(Brand Popularity Comparison)

From Figure 3 we can see that the apple was counted for the most of the time. After Apple Samsung can be considered their next competitor and followed by Oppo, Huawei. From this point of view it can be said that Apple, Samsung, Oppo, and Huawei are leading in the market or industry. On the other hand , Nokia, Asus , Vivo , Sony , Motorola have very less count and therefore less appearance in the data compared to other existing brands. Overall it can be concluded that Apple, Samsung, Huawei, Oppo are the most popular brands among the users within the industry.

Q3. What are the brands that are viewed negatively or positively?

To figure which brand is viewed positively or negatively, we perform sentiment analysis on each brand. We have chosen Vader sentiment analysis to statistically label the sentiment of each data. Vader is optimized and can yield good results when especially the data from Twitter, Facebook etc. Vader sentiment produces the polarity of the word and their probabilities of positive, negative, neutral and compound. Compound is the average value of probabilities of positive and negative scores. In our analysis we designed the sentiments in a way where polarity > 1, polarity == 1, and polarity < 1 are considered as Positive, Neutral and Negative. Once we have the sentiments for each brand, we plot a pie chart by showing the percentage of each sentiment of each brand. Here, we analyse the sentiments for each brand individually.

Samsung

From Fig: 4 we can see the positive sentiments of Samsung is 50.5% of total sentiments and less negative sentiments with 19.4% which illustrate the sentiments with positivity towards Samsung.

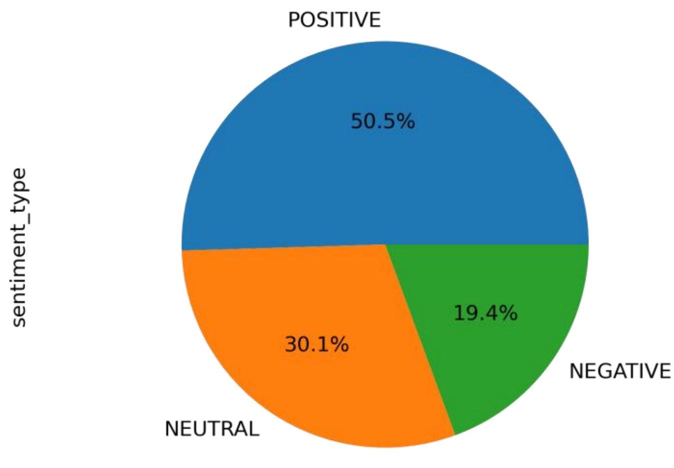


Fig: 4(Sentiments of Samsung)

Apple

Figure 5 also shows the high percentage of positive sentiments with 51.3% and 23.4% negative sentiments. From here, we can conclude that Apple is evaluated more positively among the users.

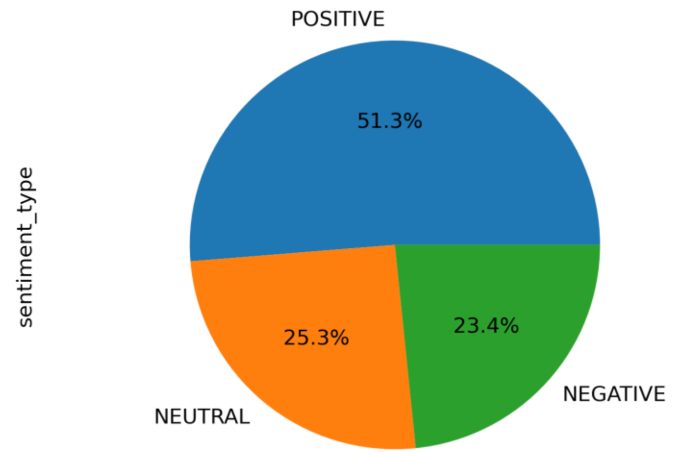


Fig: 5(Sentiments of Apple)

Huawei

We can see from figure 6 that Huawei has more than half of the total sentiments positive and also very less negative sentiments which is 15.1%. This can possibly tell us that Huawei is the favourite of its users.

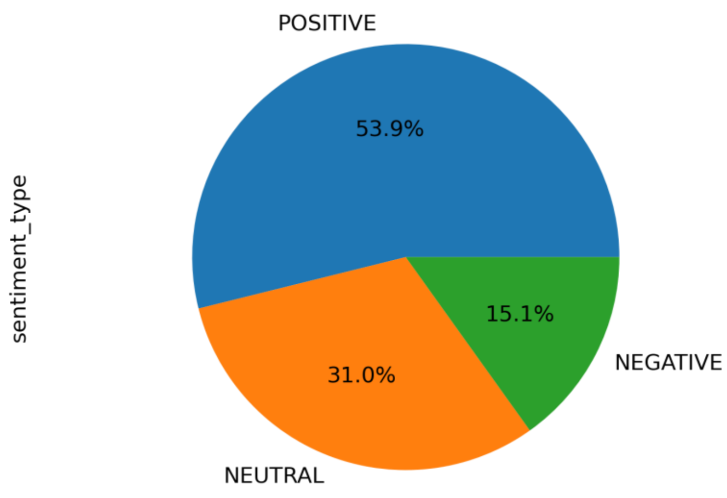


Fig: 6 (sentiment of Huawei)

Oppo

Figure 7 shows the percentage of positive sentiments 50.6%, neutral sentiments 34.0% and 15.5 % negative sentiments. The result is quite similar to Huawei. So it can be concluded that like Huawei, Oppo is also a favourite of its users.

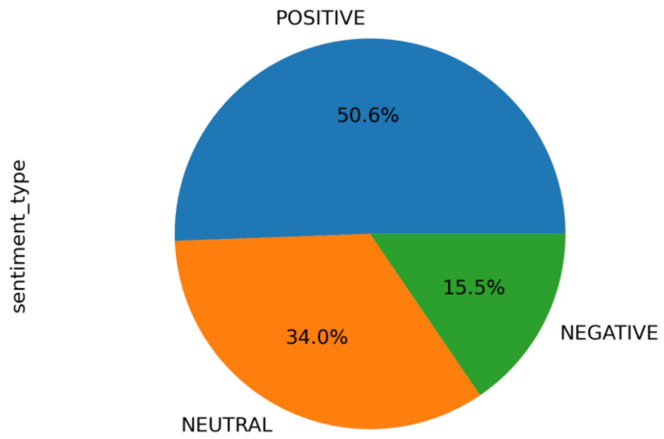


Fig: 7(Sentiments of Oppo)

Xiaomi

Figure 8 illustrates the sentiments of Xiaomi, here it shows the positive sentiments is 50.3%, 30.7% negative assignments and 19.0% negative assignments. Like other brands Xiaomi also shows more positive acceptance by their users.

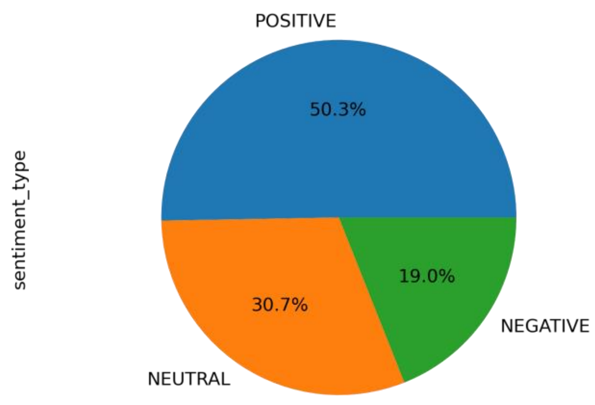


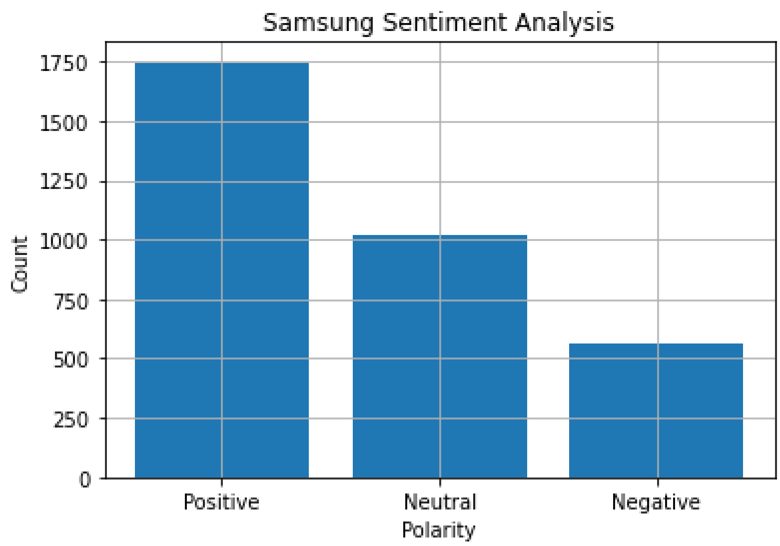
Fig: 8(Sentiments of Xiaomi)

In conclusion, we can say that all the brands are viewed very positively by the users. Compared to all brand sentiments, Apple has more negative sentiments than others and Huawei has less negative sentiments than others. However overall, all the brands that we collected are getting mostly positive sentiment from the users of twitter. This means that most of the users are having positive thoughts about these brands in the industry within the range of data we collected

Q4. Within the industry, what is the users' sentiment towards the different brands?

Samsung

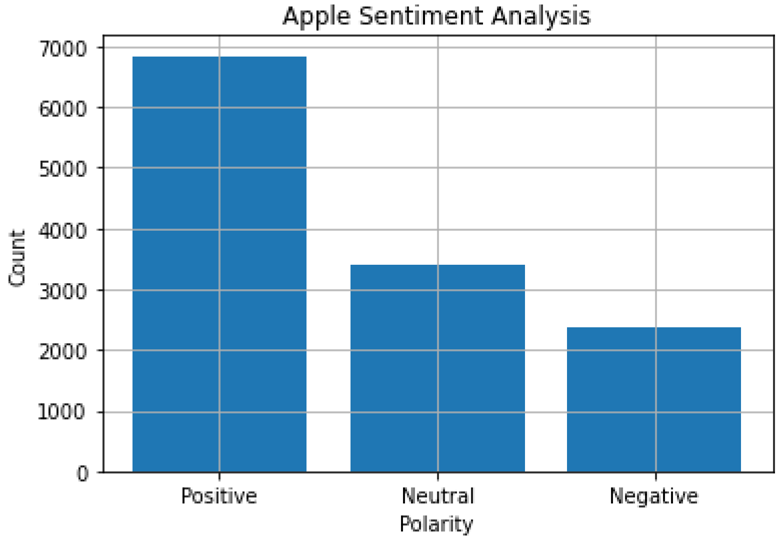
Next, we now acquired the cleaned data and are ready to be analysed. We are using Text Blob to analyse the sentiment of users towards the brands. First is the target brand Samsung:



We collected around 70k tweets for Samsung but most of it are spam and after the pre-process we still have around 3k tweets and the figure above shows the sentiment analysis of users towards Samsung. Most of the users are having positive sentiment towards Samsung with a count of 1746, 1015 of them having neutral sentiment and the count of negative sentiment towards Samsung is 559. Overall the sentiment of the user in the tweets that we collected are positive towards Samsung.

Apple

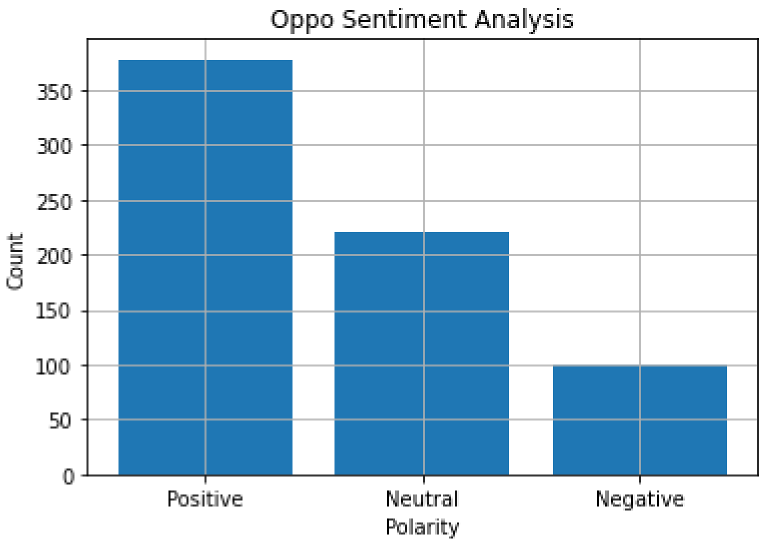
Next, we will be analyzing the sentiment of users towards Apple. Similar to Samsung, Apple is one of the most discussed brands in the mobile phone industry by having 80k tweets collected over one week and the cleaned data of Apple still having around 12k tweets to be analysed.



We can observe that in the figure above the overall sentiments of users towards Apple are positive. Based on the output of Text Blob, Apple acquired 6841 positive sentiment, 3409 neutral sentiment and 2369 negative sentiment. Most of the users are having positive sentiments with the most discussed brands of the data collected.

Oppo

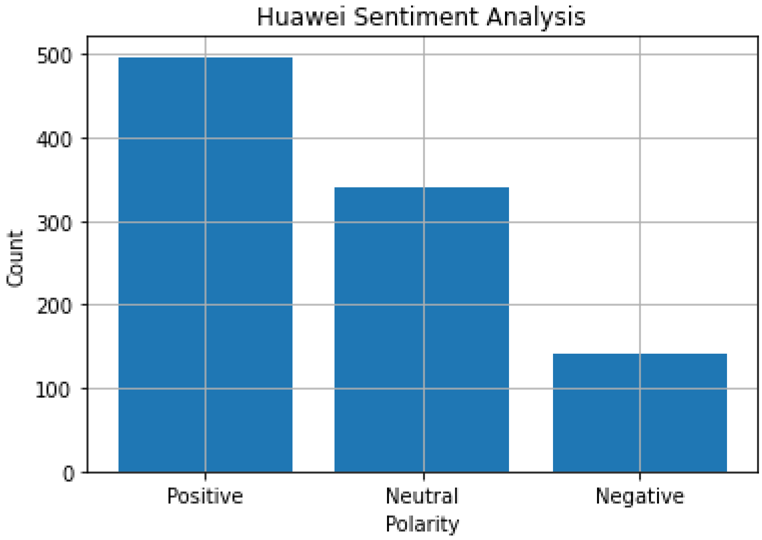
Oppo is also one of the famous mobile phone brands in the industry. Although it is not as popularly discussed as Samsung and Apple, it is also important for us to understand the sentiment of twitter users towards this brand.



With the 700 cleaned data of Oppo dataset, we can conclude that the overall sentiment of users towards the brand Oppo is also positive. Oppo has 378 positive sentiment tweets, 220 neutral sentiment tweets and 100 negative tweets from the user of the cleaned dataset.

Huawei

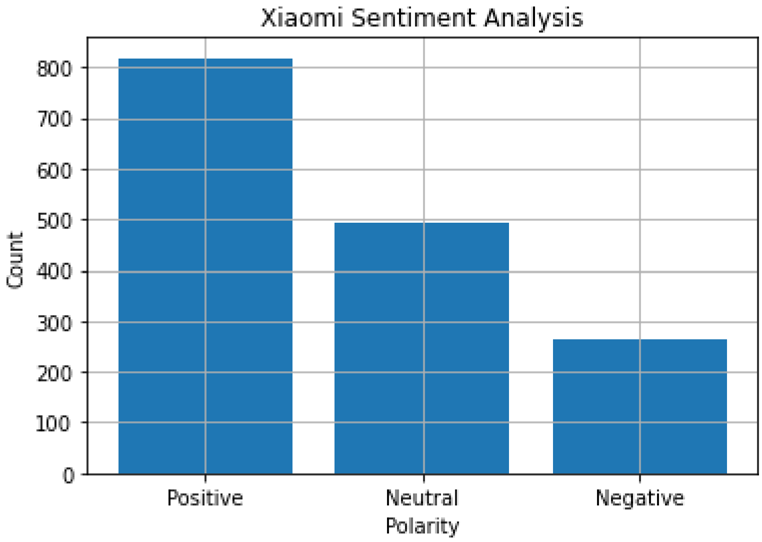
Next brand we will be analyzing is the brand we use as a comparison to Samsung in assignment 1, Huawei. Same as Oppo, Huawei is also a China brand and not so popularly discussed on twitter by having only 8k raw tweets and 1k cleaned tweets.



Same as the previous brand, the sentiment of users towards Huawei is also positive most of the time by having 497 positive sentiment tweets from users. 340 of them are neutral sentiment tweets and only 141 are negative.

Xiaomi

The next brand we are going to analyse is also one of the biggest brands in the industry which is Xiaomi. Raw data of Xiaomi are around 7.5k tweets and cleaned data having 1.2k tweets.



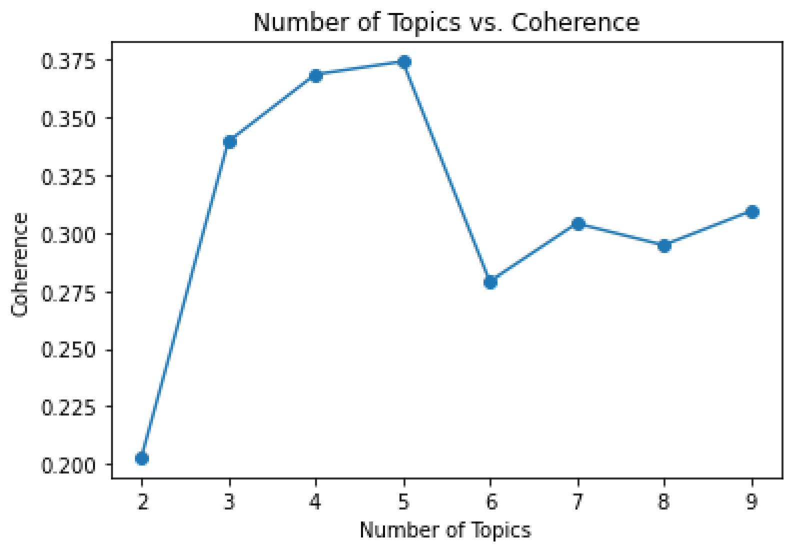
The figure shows the sentiment analysis of Xiaomi on tweets collected and it is positive overall. Xiaomi is getting 818 positive sentiment tweets from users, 492 neutral tweets and 265 negative sentiment tweets from the users.

Conclusion

In conclusion, all the brands that we collected are getting mostly positive sentiment from the users of twitter. This means that most of the users are having positive thoughts about these brands in the industry within the range of data we collected.

Q5. What are the frequent topics that users mention when they interact with the different brands over social media? What is their sentiment towards those topics?

To answer this question, we concat all the cleaned data of the collected brands into the same set of data and perform topic modelling. Next is to perform some other pre-process such as lemmatization and turn the dataset into numerical representations using a bag of words. After all this we can use the coherence score to find out what is the best number of topics for our dataset.



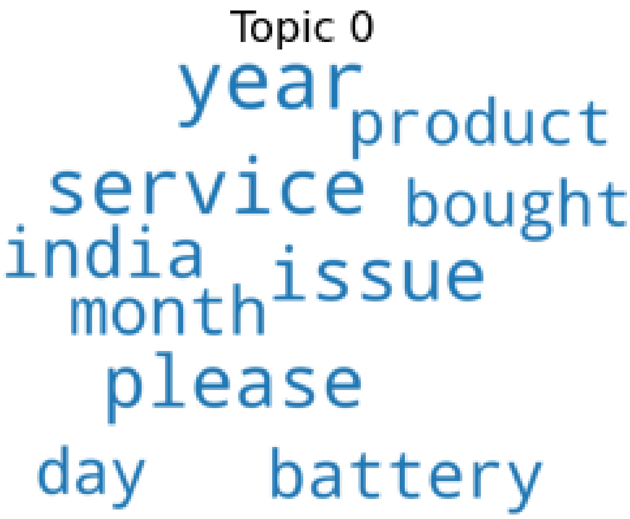
After calculating the coherence score of different numbers of topics, we decided to use 5 as the k for the LDA model to perform topic modelling. The frequent word of each topic are shown in the word cloud below:

With the LDA model we obtained, we can then create a data frame to show what is the dominant topic for each tweet and the topic percent contribution.

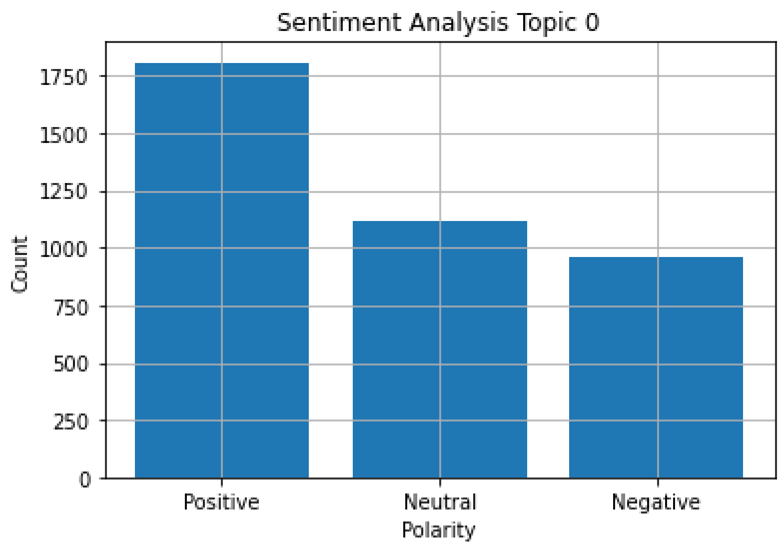
	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	4.0	0.5999	note, pro, case, x, cover, charging, max, powe...	[fuck]
1	1	4.0	0.6025	note, pro, case, x, cover, charging, max, powe...	[chale, much, galaxy, touch, plug]
2	2	2.0	0.6854	gb, camera, android, g, update, smartphone, se...	[true, copy, paste, device, window, laptop, us...
3	3	2.0	0.4954	gb, camera, android, g, update, smartphone, se...	[markzen, daily, task, click, naver, daum, wiki]
4	4	1.0	0.5313	like, new, buy, get, got, need, time, would, k...	[jungkook, behind, scene, galaxy, posted, inst...
5	5	0.0	0.2000	year, service, issue, please, india, month, ba...	[]
6	6	0.0	0.3767	year, service, issue, please, india, month, ba...	[party, popper, free, galaxy, purchase, anothe...
7	7	2.0	0.3276	gb, camera, android, g, update, smartphone, se...	[yes, lily, mobiletrans, support, text, messag...
8	8	2.0	0.8895	gb, camera, android, g, update, smartphone, se...	[brand, new, galaxy, way, like, customize, gal...
9	9	1.0	0.6588	like, new, buy, get, got, need, time, would, k...	[yea, look, nice, better, processor, thats, mo...

Next we can extract and group the tweets that have the same topic and perform sentiment analysis on each of the topics.

First we will be observing the Topic 0 of the LDA model given us from the data.

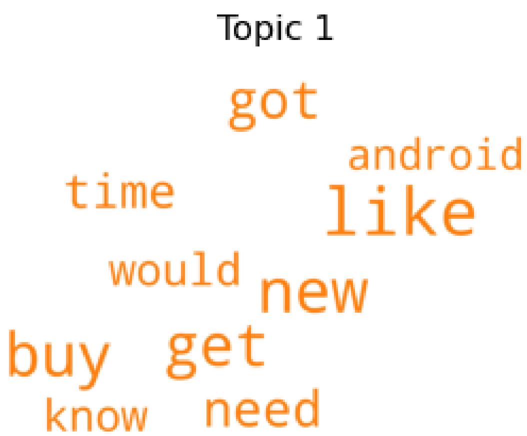


With the keyword of topic 0, we can infer that topic 0 could be related to the mobile product after service and warranty issues.

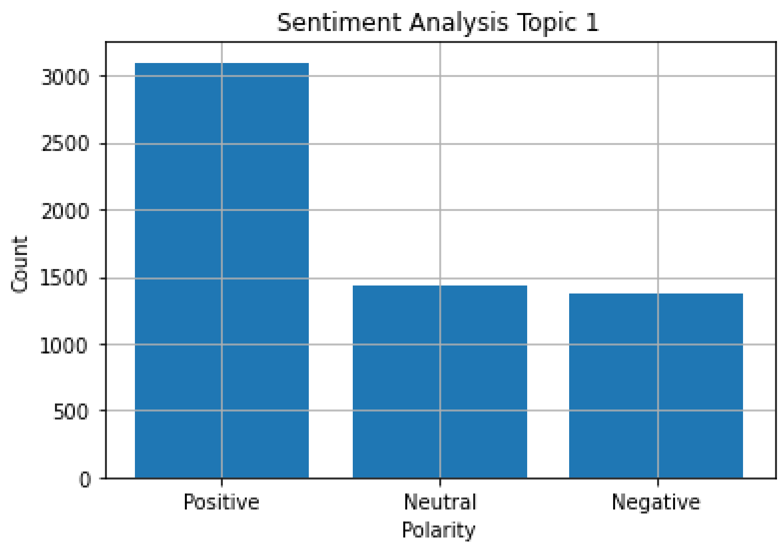


Looking at the sentiment analysis of topic 0, most of the users have positive sentiment when they are discussing topic 0. Positive sentiments are the majority but the number of negative and neutral sentiments is more than the positive sentiment. This can mean most of the users are happy with the topic discussed in topic 0 but some of them are having unpleasant experiences and issues when dealing with the thing discussed in topic 0.

Next, we will be analyzing topic 1 given by the LDA model.



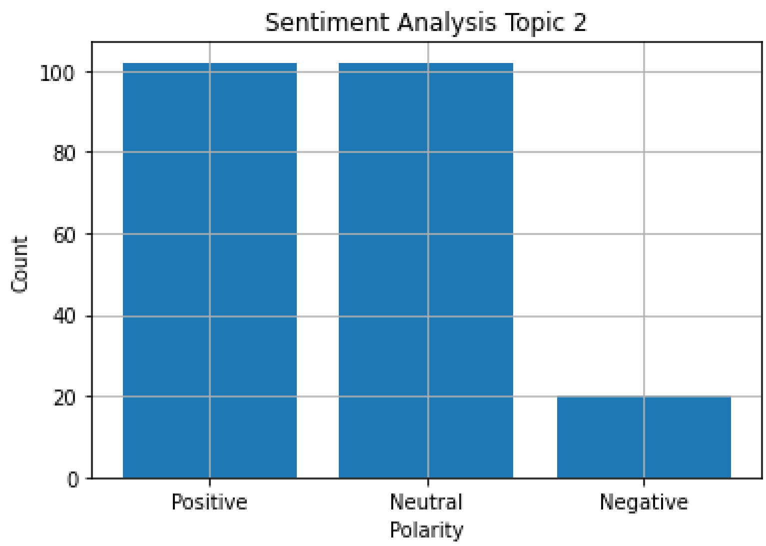
From the frequent words of topic 1, we can observe that there are words like get, new and buy. We can infer that this topic is where the user discusses getting a new phone of one of the brands we analyse on.



As the figure shows most of the user sentiments towards topic 1 are in a positive way. Neutral and negative sentiments are less than half of the positive sentiments.

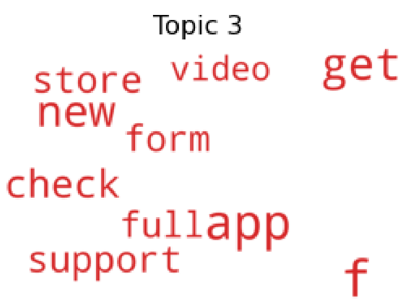
After that, we will be analyzing topic 2. Figure below shows the keyword in topic 2.

From the keyword of topic 2, we can infer that the topic might be talking about how the specification of a mobile phone affects the price because of the words like: camera, GB and series appear.

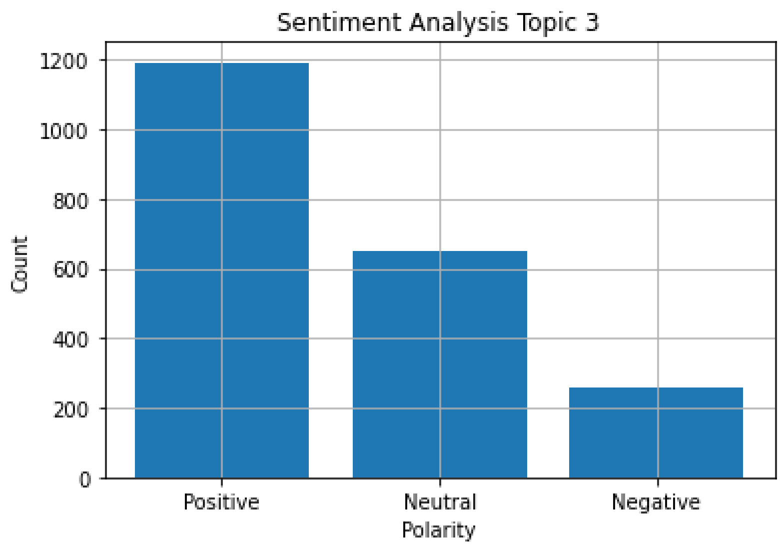


When talking about the specification and price of the mobile phone, most of the users are having positive and neutral sentiments towards the topic. Only a minority of the users are having negative sentiments towards it.

Next, we will analyze the keyword from topic 3 and see what we can infer.

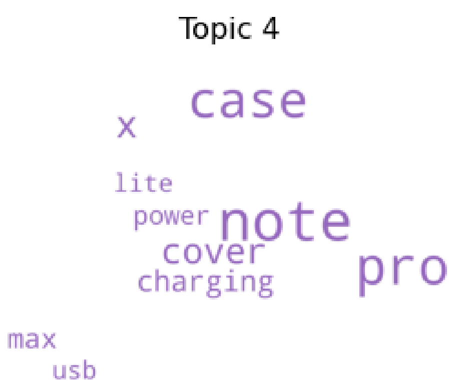


As the figure above shows, the keyword from topic 3 is kind of hard for us to infer what is the main focus of the topic. The only idea we can infer is maybe users are talking about mobile phone applications that need to have support for new features or something.

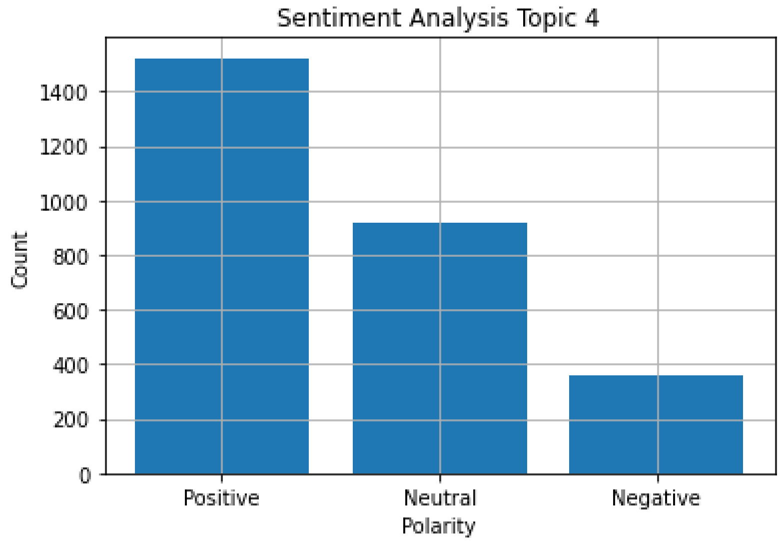


When it comes to topic 3, most of the user sentiment towards it are positive and some of them are neutral. Only a minority is negative sentiments.

Last but not least, we will infer the keyword of topic 4 given by our LDA model.



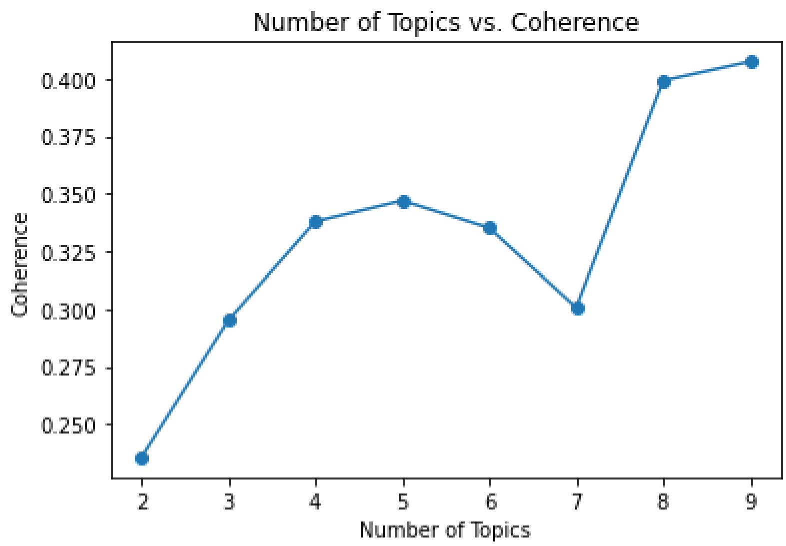
Based on the keyword in topic 4, we can conclude that there are two topics that are discussed and they might be related. First, the user might be discussing the series and the model of mobile phones of different brands because terms like: x, lite, note and pro appear. Next the users are discussing the charging and the power of the battery. This two topic should be related to each other as the charging ability of different phone model can be different.



As the figure shows, just like previous topics the topic 4 also getting positive sentiments overall. Few of them are having neutral sentiments and a small group of users are having negative sentiments towards topic 4.

Q6.What are the important aspects (and its sentiment) that are discussed by the users specifically on the target company.

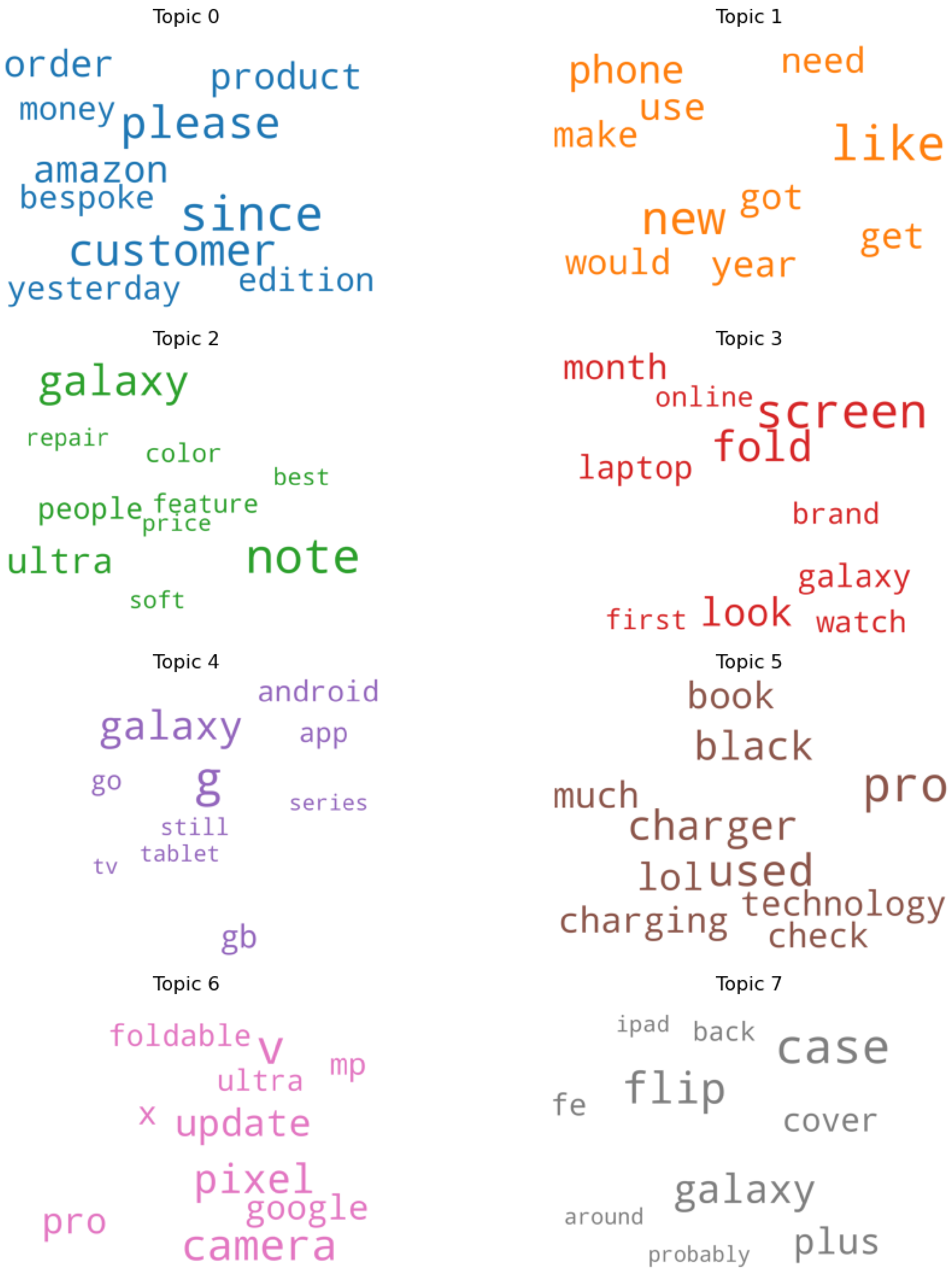
Similar to the previous question, we will run topic modelling on the target company which in our case is Samsung. We will first run a function to get the coherence score for each number of topics to get the optimal k.



We will be using k = 8 in the LDA model. Below figure shows the data frame of the Samsung LDA model. It contains the dominant topic and keywords of each tweet.

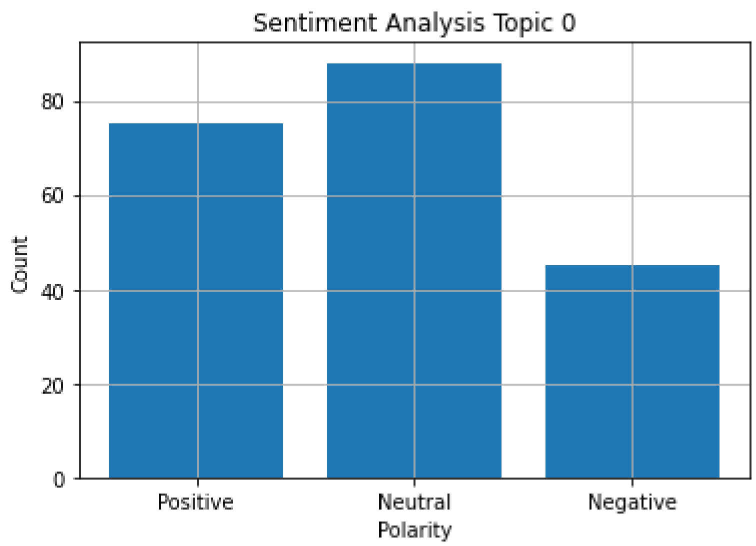
	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	7.0	0.5625	case, flip, galaxy, plus, cover, fe, back, ipa...	[fuck]
1	1	5.0	0.6387	pro, used, charger, black, lol, book, much, ch...	[chale, much, galaxy, touch, plug]
2	2	3.0	0.5516	screen, fold, look, month, laptop, galaxy, wat...	[true, copy, paste, device, window, laptop, us...
3	3	3.0	0.7595	screen, fold, look, month, laptop, galaxy, wat...	[markzen, daily, task, click, naver, daum, wiki]
4	4	6.0	0.8904	v, camera, pixel, update, pro, google, x, mp, ...	[jungkook, behind, scene, galaxy, posted, inst...
5	5	0.0	0.1250	since, please, customer, amazon, order, produc...	[]
6	6	4.0	0.8253	g, galaxy, gb, android, app, go, still, tablet...	[party, popper, free, galaxy, purchase, anothe...
7	7	7.0	0.6182	case, flip, galaxy, plus, cover, fe, back, ipa...	[yes, lily, mobiletrans, support, text, messag...
8	8	1.0	0.4043	like, new, phone, use, year, get, got, make, w...	[brand, new, galaxy, way, like, customize, gal...
9	9	3.0	0.4464	screen, fold, look, month, laptop, galaxy, wat...	[yea, look, nice, better, processor, thats, mo...

Below figure shows the topic we get for Samsung topic modelling and its keyword:



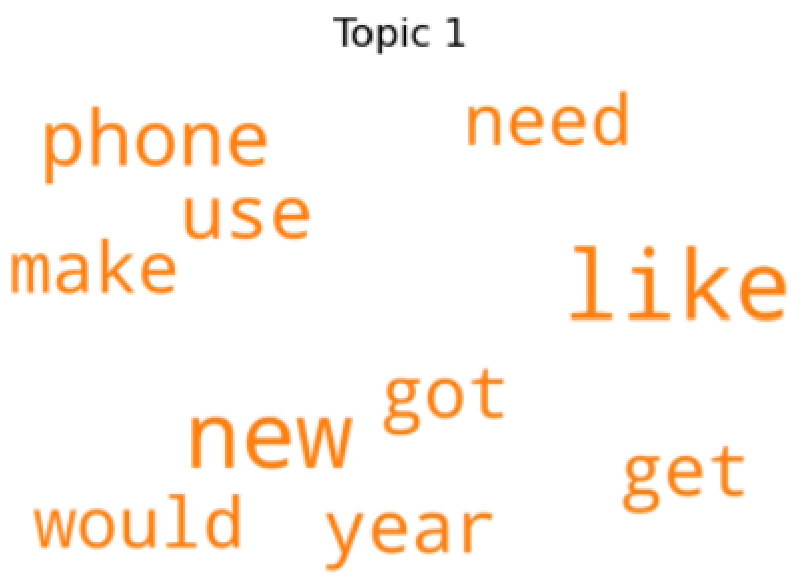
First, we will first infer the topic 0 for Samsung LDA model.

Based on the keyword above, topic 0 might be related to customer service and product after service on the Samsung mobile phone.

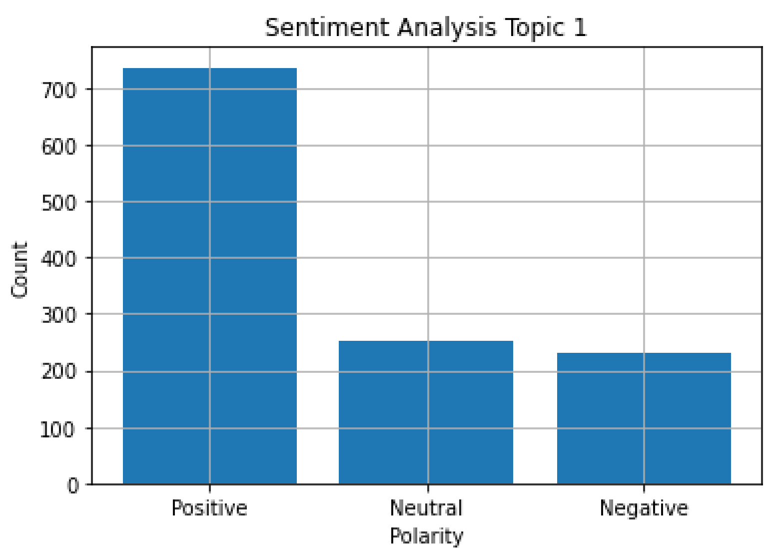


Unlike most of the sentiment analysis results we get along this assignment, sentiment of users towards topic 0 are more neutral instead of positive and negative. With that said, the number of positive sentiments still overpower the negative sentiment in this topic.

Next, we will infer from the keywords of topic 1 to analyse what the user is discussing about Samsung in topic 1.

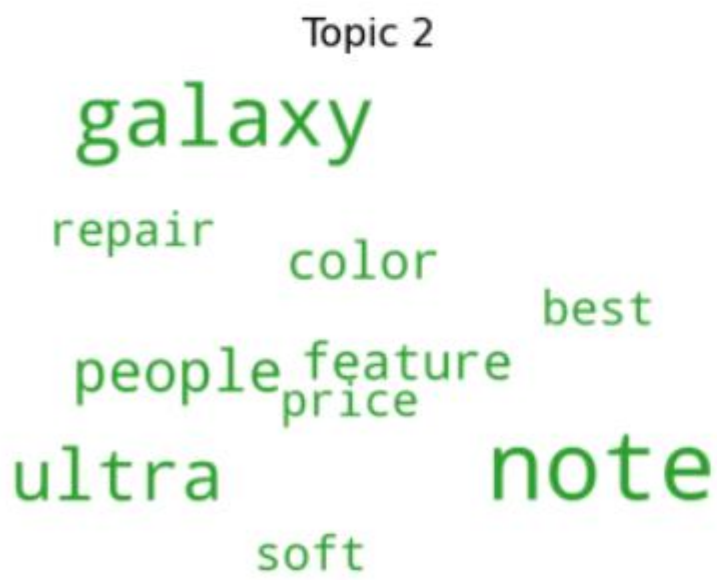


From the frequent words of topic 1, we can observe that there are words like get, new and need. We can infer that this topic is where the user discusses getting a new Samsung phone.

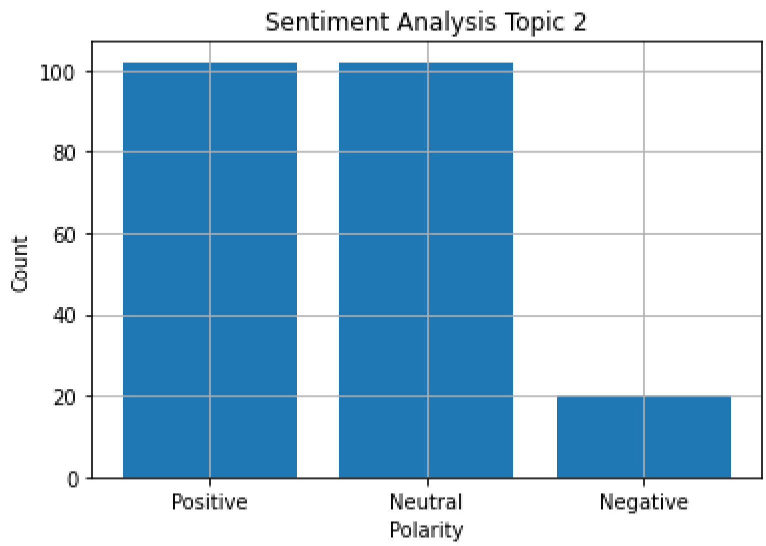


From the figure of sentiment analysis above, we can see that the majority of the tweets get positive sentiment from the users. Only a minority have negative sentiment and neutral sentiment.

Next, we will be analyzing the keyword from topic 2 of Samsung LDA model given below.



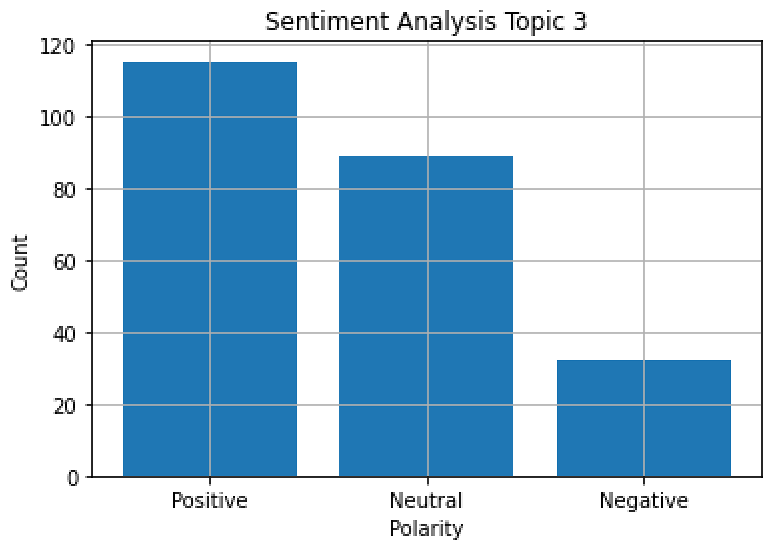
From what we can observe, users might be discussing the specifications of different Samsung phone models. Words like note, galaxy and ultra are the words that Samsung uses to name their phone model series. Users might me discuss these phone models about their specifications such as color and features.



The positive sentiments and neutral sentiments are almost even based on the figure of sentiment analysis of topic 2 shown above. Only a minority of users give negative sentiments towards topic 2.

Next, we will infer the topic of topic 3 from the Samsung LDA model.

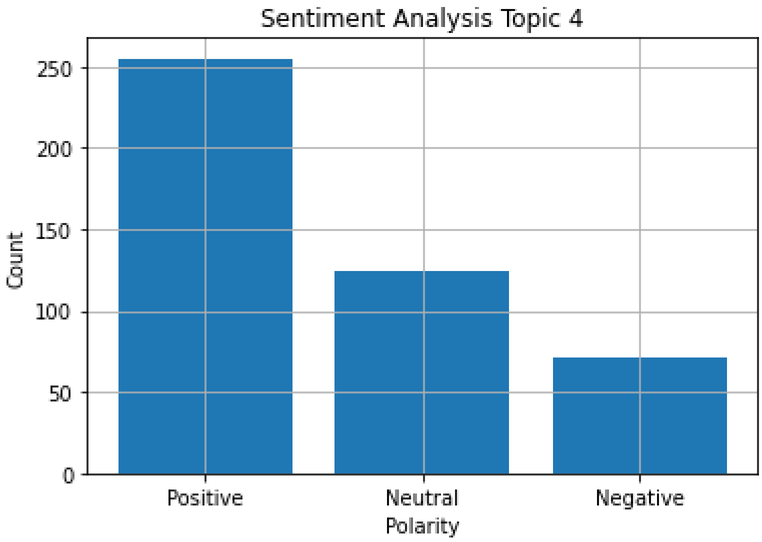
Based on the keyword above, the user might be discussing Samsung products other than mobile phones. The user might be discussing how Samsung is doing with the product other than mobile phones.



The sentiment analysis of topic 3 only has a minority in negative sentiments and most of the users are having positive and neutral sentiments towards the topic discussed in topic 3.

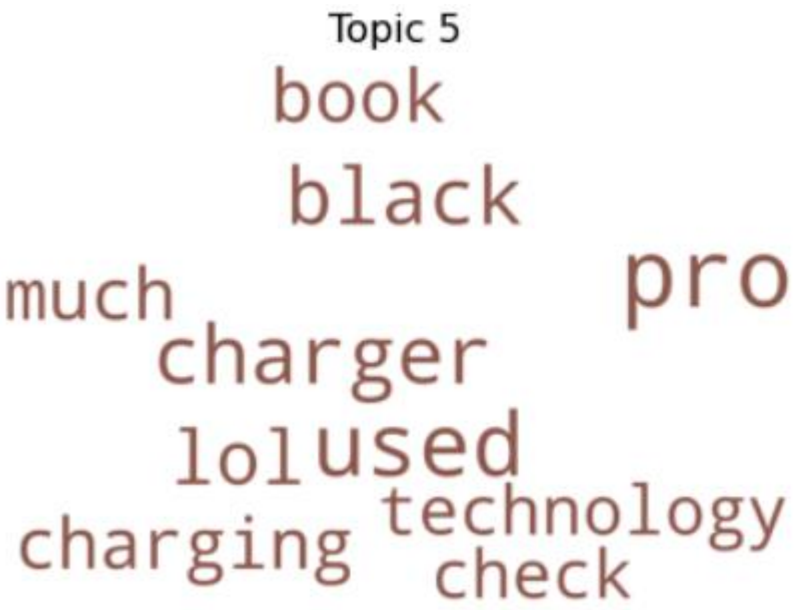
The following topic we are going to infer is topic 4. Below is the keyword of topic 4.

From what the keyword shows, we can infer that topic 4 might be talking about the other products of Samsung like TV and tablet and how they work with the Samsung mobile phones.

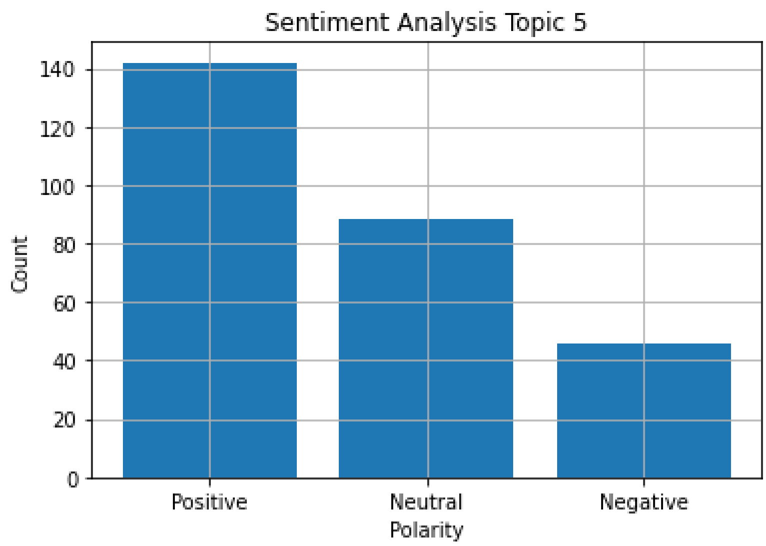


Majority of the user sentiment are positive towards topic 4 while few of them have neutral sentiment and small percentage of it are having negative sentiments.

The upcoming topic we are going to discuss is topic 5 and the figure below shows the keyword of it.



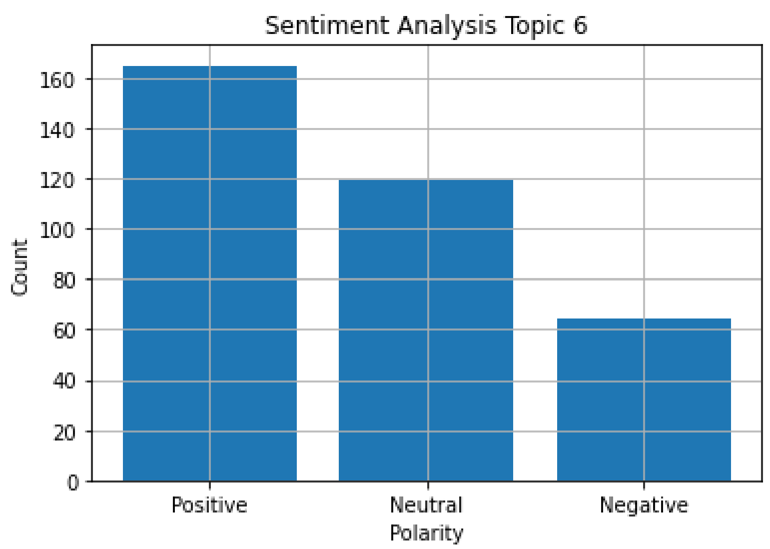
The keywords of topic 5 barely make any sense. The only idea that can be inferred from is the charging issues of Samsung and charger. This topic might be discussing something related with the technology of phone charging of Samsung.



Similar to the previous topics, most of the users are having positive sentiments towards topic 5 and some of them having neutral sentiments. Only a small part of them are negative sentiments.

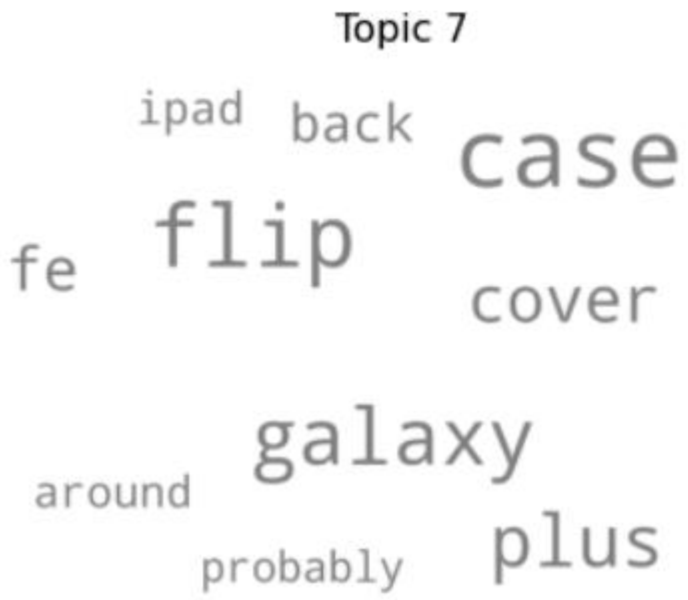
Next, we will be analyzing topic 6 given by the LDA model.

From the keyword we can infer that Samsung and Google are choosing V to represent their brands. V is a member from the Korean group BTS. So the text that belongs to this topic is mostly talking about V promoting their products.

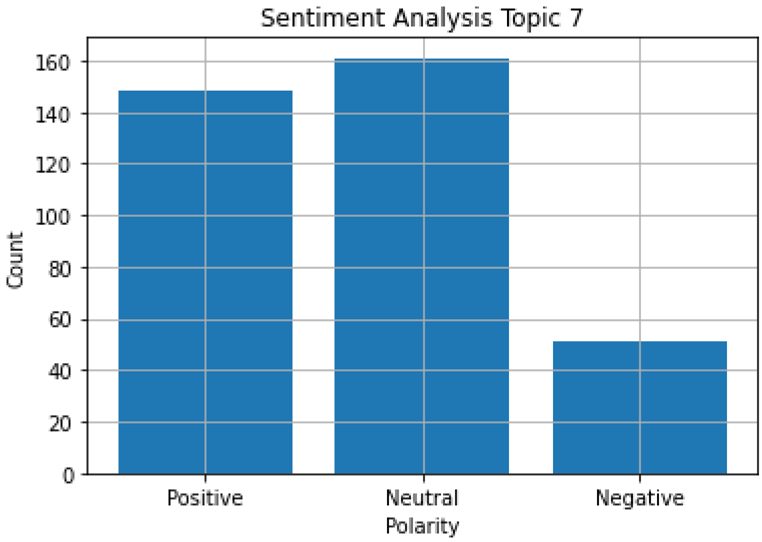


Based on the sentiment analysis figure of topic 6, most of the users are having positive sentiments towards this topic. Only a minority of users having negative sentiments and a large group of users also having neutral sentiments.

Last but not least, we will be analyzing the last topic given by the Samsung LDA model which is topic 7.



Last topic is the users talking about the latest flappable phone by Samsung that is going to launch. The model name is Samsung Galaxy S21 FE so the keyword Fe and galaxy keep appearing.



We can observe that although there are a lot of positive sentiments in this topic, the majority of tweets are neutral. Only a minority sentiment is negative.