



Social Media Computing

**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: 3 October 2021

Table of Contents

Assignment 1 – Social Media Campaign & Metrics	1
Introduction.....	1
Huawei Mobile	1
Huawei current Social Media Campaign.....	5
Samsung	6
Samsung current Social Media Campaign.....	10
Dashboard Implementation and Metric used for analysing	12
Assignment 2	15
Introduction.....	15
Q1. What is the opinion of users towards the industry?	15
Q2. Which are the brands that are popular among users within the industry?	18
Q3. What are the brands that are viewed negatively or positively?	19
Samsung.....	19
Apple.....	20
Huawei.....	20
Oppo	21
Xiaomi	21
Q4. Within the industry, what is the users' sentiment towards the different brands?	22
Samsung.....	22
Apple.....	22
Oppo	23
Huawei.....	24
Xiaomi	24
Conclusion	25
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?	25
Q6.What are the important aspects (and its sentiment) that are discussed by the users specifically on the target company.....	32
Assignment 3 – Social Network Analysis	43
Introduction.....	43
Data Extraction.....	43
Centrality Measures	43
HuaweiMobile	44
SAMSUNG	51
References.....	58

In figure 1, we can see some bold words: “huaweiwatch”, “phone”, “huaweimateview”, “worldheartday”, “huaweimatepad”, “camera” which represent the most prominent or used words in tweets. It can be observed that all these words are interrelated. Here, the joined words are used as the hashtags. Few statements are made from this word cloud:

- This cloud is presenting the feature of the phone using sentimental words which are: “camera stunning”, “wireless”, “productive” and “perfect”.
- Some activities are mentioned: “workout”, “traveling”, “swimming”, which are very inspiring to the audiences and thus can help to draw the attention
- The most used hashtags are “huaweimatepad”, “huaweimateview”, “huaweiwatch” which can be a possible campaign that companies are running.

Word cloud provides a quick view but for deeper analysis sometimes word cloud is not sufficient itself to reach the conclusion. For a better analysis, we extract hashtags for the tweets. Nowadays most of the business are using hashtag based campaign since hashtag is very helpful to attract new customers, generate the engagement and get followers to retweet or post about the hashtag (ThriveHive, 2015). This is why we decided to use the hashtag entity comparing with other metrics to keep track of running campaigns.

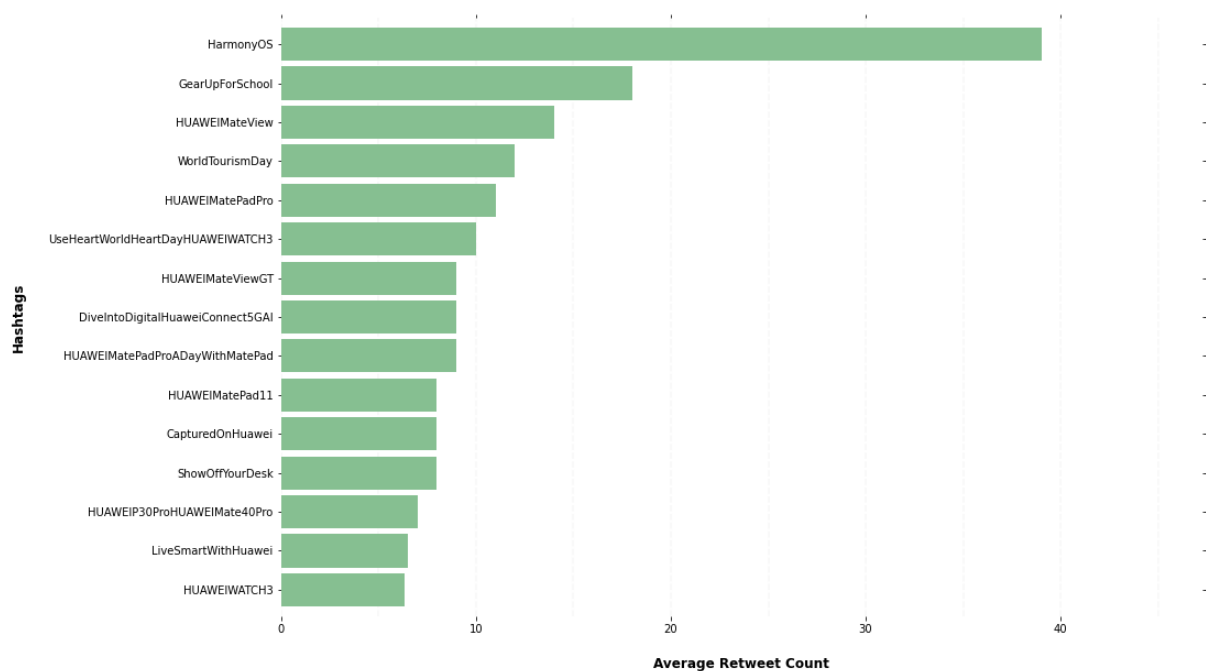


Figure 2 : Average Retweet Count of hashtags of Huawei

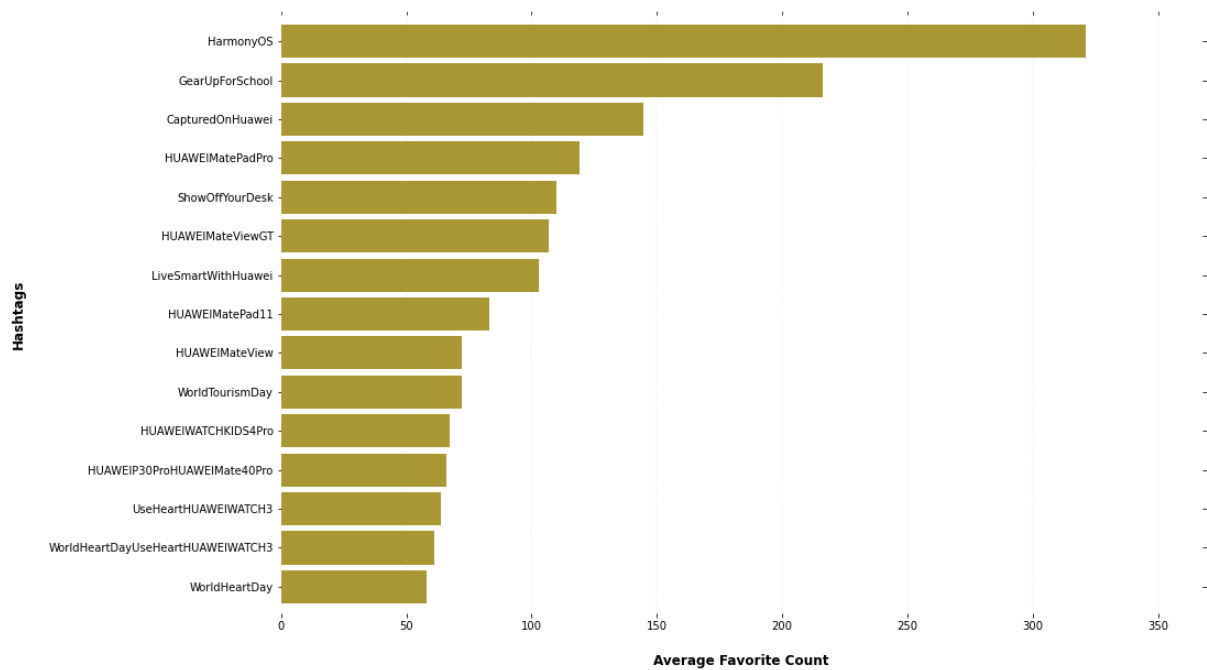


Figure 3 : Average Favorite Count of hashtags of Huawei

Figure 2 and 3 show the most average retweet and favourite count of used tweets by descending order which used hashtags. Both figures show the top hashtags which are “HarmonyOs”, “GearUpForSchool”.

HarmonyOs is Huawei’s proposed next-generation based operating system that brings a lot of new features between smart devices (Wikipedia, 2019). This hashtag shows a product promotion-based campaign that hit the most retweet count.

‘GearUpfornewschool’, ‘ShowoffourDesk’ are activity-based campaigns where it asked the followers especially the young ones to show up their e-learning setup. Giving a task to the follower can be a fun activity that also helps to reach for more engagement.

These analytics show that campaigns which is related to an innovative feature in products or showcase or task given to the followers are more entertained among the followers. Therefore, such a campaign is suitable for Huawei’s social media marketing.

Discovering the best day to tweet a post is an important key point for an effective social media campaign strategy (Patel, 2021). This is why we plot a graph to see the behaviour change of Retweet count and Favourite count over time for hashtags tweets. It shows, there’s a peak in count on 13,20,22,27 September. According to the calendar, these days are mostly between Wednesday and Friday. On the other hand, the count was very low on 2, 9, 16, 23 September

and the interesting fact is all these days fall on weekends. Therefore, we can conclude running a campaign on weekdays may potentially grow more reach.

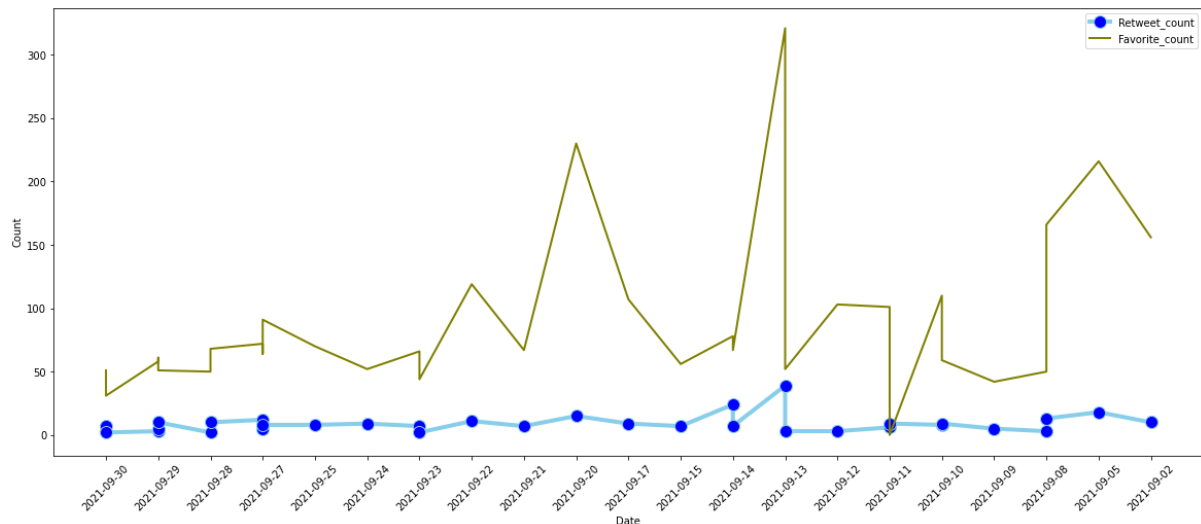


Figure 4 : Huawei Retweet and Favourite count over time

Social Media Marketing agencies or communities play a very important role in giving businesses a chance to follow their consumer's activities or potential buyers. This also helps the business to know their target, likes, interest so that they can create a better marketing strategy to attach such customers. To extract this information we extract the source data of tweets which tells from where or through which platform the tweets are posted.

	Date	Text	Retweet_count	Favorite_count	hashtags	Source
36	2021-09-12	Which is your favourite city to travel to? 🌍\n...	3	103	CapturedOnHuawei	Falcon Social Media Management
37	2021-09-11	Marvel at these beautiful fuchsia tones #Captu...	6	101	CapturedOnHuawei	Falcon Social Media Management
38	2021-09-11	RT @Huawei: #DiveIntoDigital at #HuaweiConnect...	9	0	DiveIntoDigitalHuaweiConnect5GAI	Twitter Web App
40	2021-09-10	Say goodbye to desktop clutter and hello to Wi...	8	110	ShowOffYourDesk	Falcon Social Media Management
41	2021-09-10	Pair your #HUAWEIMatePad11 with the HUAWEI Sma...	9	59	HUAWEIMatePad11	Falcon Social Media Management
43	2021-09-09	Set up your #HUAWEIWATCH3 Series and connect i...	5	42	HUAWEIWATCH3	Falcon Social Media Management
45	2021-09-08	📱\n\nTell us which you'd choose below!\n\n#...	3	50	LiveSmartWithHuawei	Falcon Social Media Management
46	2021-09-08	Busy days spent commuting back and forth? #HUA...	13	166	HUAWEIMatePad11	Falcon Social Media Management
48	2021-09-05	It's that time again... #GearUpForSchool!\n\nTel...	18	216	GearUpForSchool	Falcon Social Media Management
53	2021-09-02	Heading back to school in style 🎒\n\nMake your...	10	156	LiveSmartWithHuawei	Falcon Social Media Management

Table 1 : Huawei Data Frame

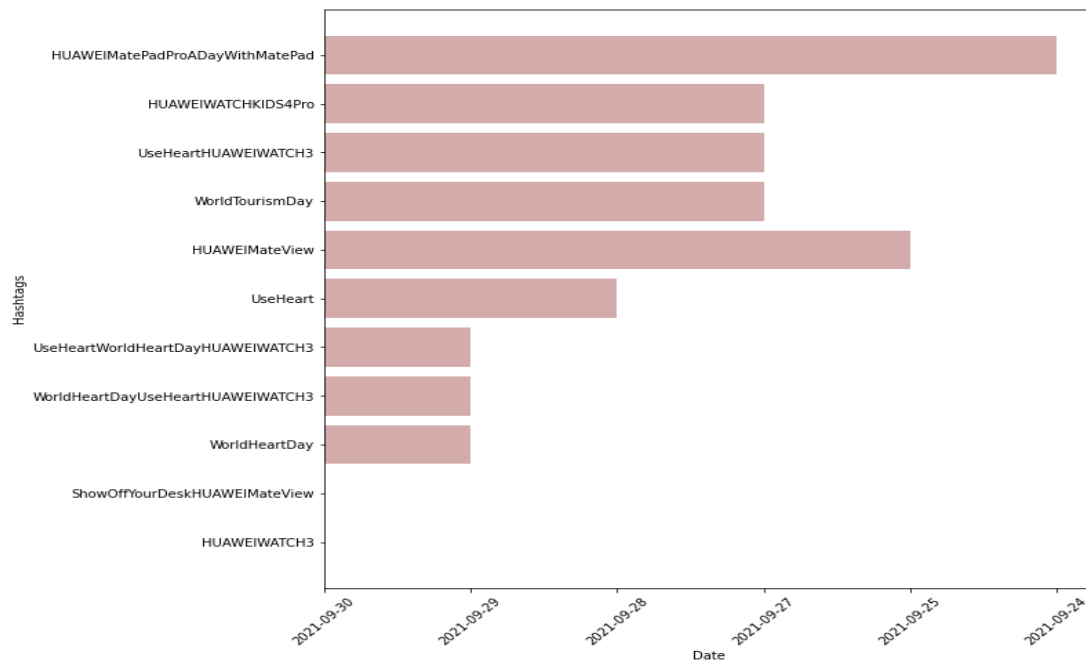


Figure 6 : Huawei recent Date of hashtags

Figure 6, shows the most recently used hashtags are HUAWEIMATCH3, Huaweimateview, and WordheartDay HUAWEIWATCH3, useheartwordday, useheart. In brief, the recent ongoing campaign is targeted at the popular hashtags trend and created hype on their new product Huaweiwatch3. This product can store health information, maintain daily routine and help to manage a heart-healthy life. From the date data, we observed that 29th September is ‘world heart-day’. So this can be a possible statement that Huawei used this heart day trend to promote their new product. Looking for popular trends and hashtags plays an important role in social media campaigns. It illustrates that everyone will be able to see how on-point the product and the brand are while they are showcasing their brand personality. So overall, it is concluded that their latest campaign is related to health awareness by promoting the importance of healthy-life activities.

Samsung

We choose Samsung as a competitor to Huawei. To examine the best campaign, it’s very necessary to analyse the other competitor’s business marketing to know the behaviour and change within the strategy.

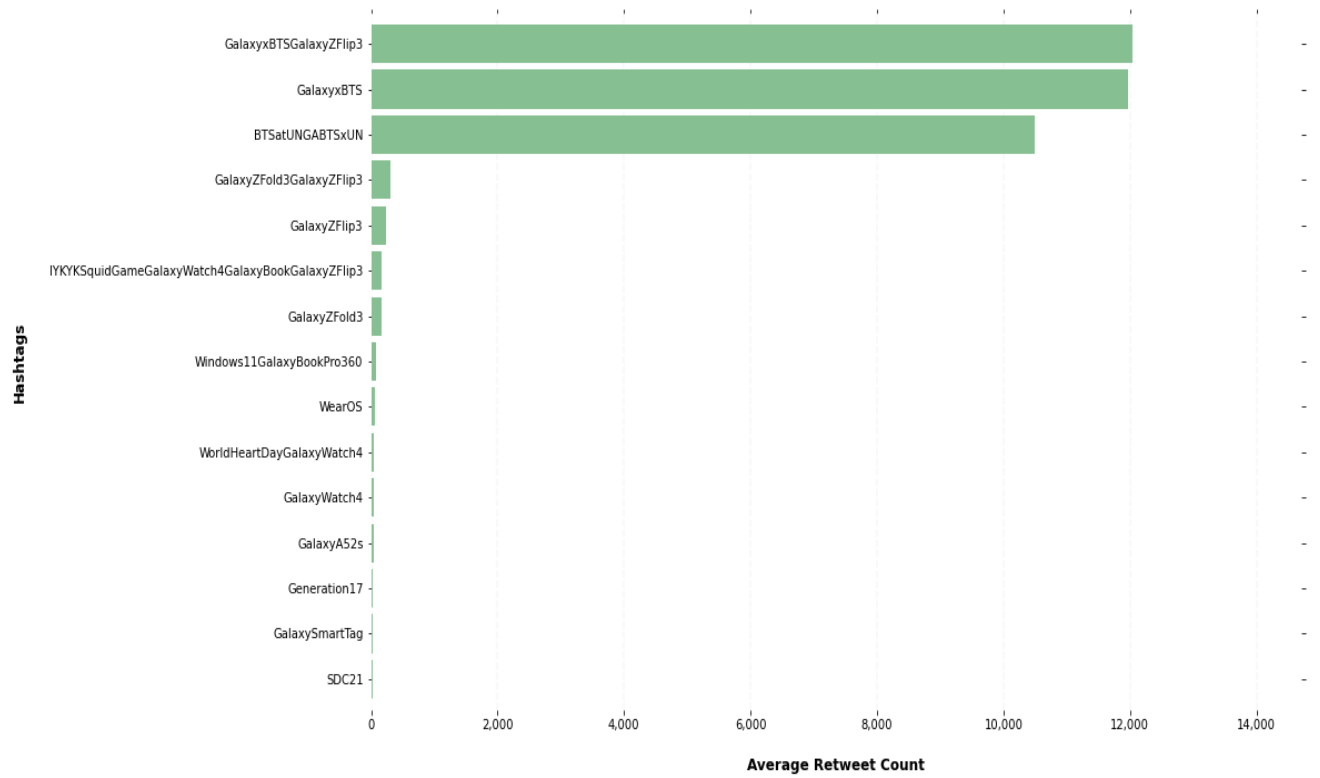


Figure 8 : Average Retweet Count of hashtags of Samsung

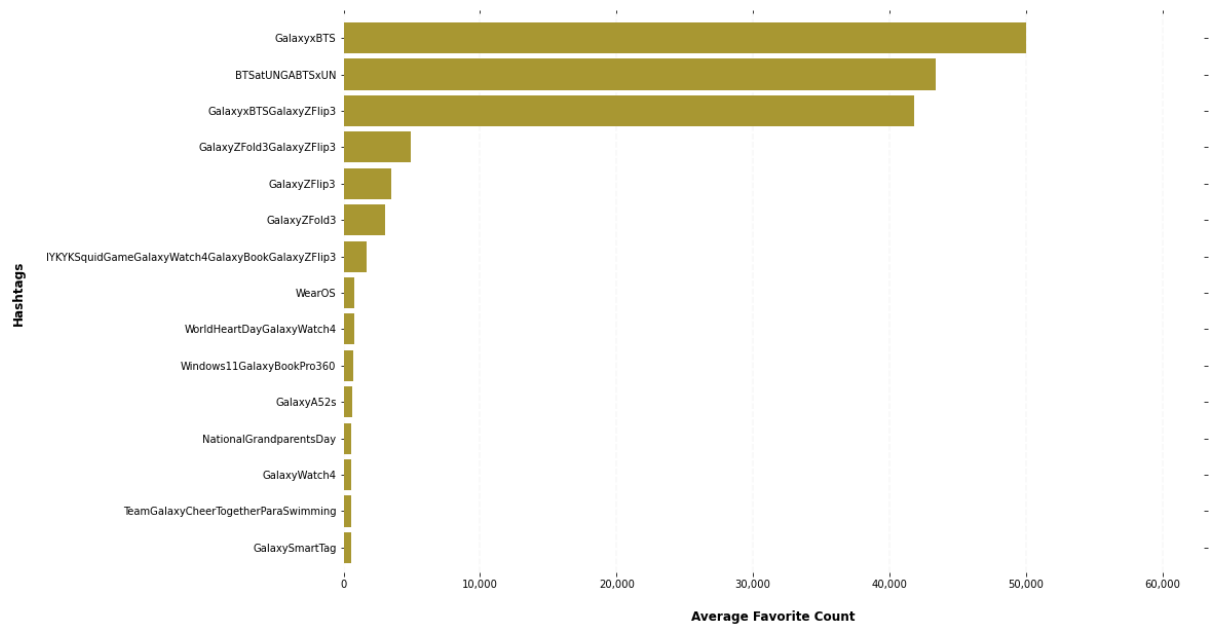


Figure 9 : Average Favourite Count of hashtags of Samsung

Figure 8 and figure 9 shows that the top 3 hashtag trends are 'GalaxyBTSGalaxyZFlip3', 'GalaxyBTS', 'BTSatUNGABTSxUN' with respect to the most retweet count and favourite count. The common fact in between all of these hashtags is, they

are highlighting the hashtags adding the BTS tag. BTS is worldwide famous for its music. They have joined Samsung as global ambassador (Chan, 2021). This is one of those social media campaigns where the brand makes a partnership with an influencer. It is also known as the brand ambassador program. A brand ambassador program plays a partnership relationship between brand and influencer in order to improve brand awareness and drive sales. Indeed, the influencer becomes the main tag for the brand and they use their platform to promote a product or service. Thus, it helps to expand both the influencer and brand's audiences. This type of exposure is what successful partnerships strive for. We see this partnership reflection in our measuring metric. Therefore, we can conclude that 'Partnership with global ambassador' is suitable for Samsung as it helped to grow more reach to the audiences.

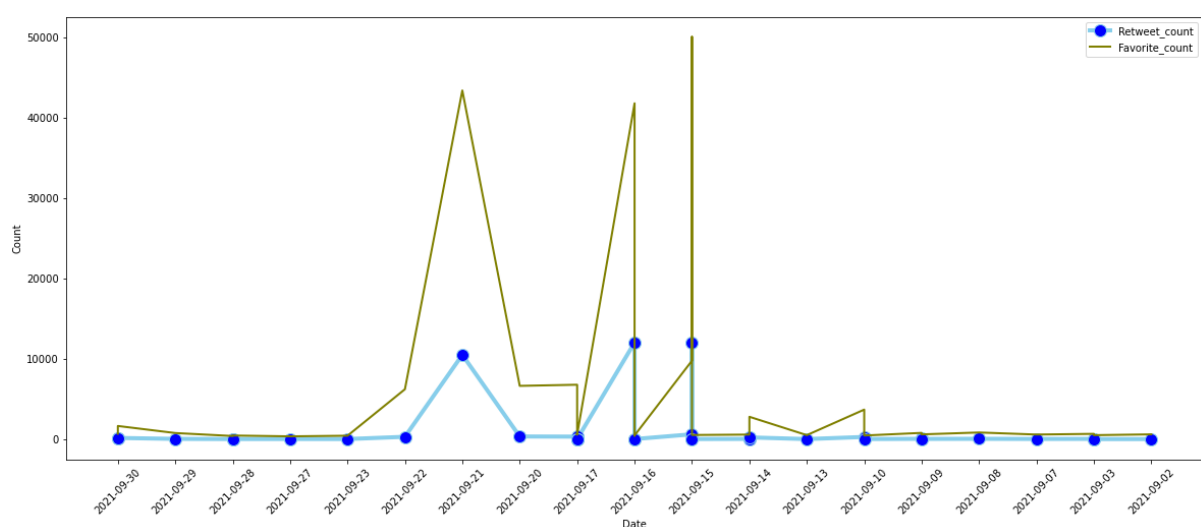


Figure 10 : Samsung Retweet and Favourite count over time

Figure 10 shows the count with respect to retweet count and favourite count over time. We can see some peak value at 21, 15 (weekdays) September and peak down at 16, 17 September which is the weekend day of the month. Lead to the conclusion that, running a campaign on weekdays have much reach, effective and can be considered a good strategy than running a campaign on weekend.

We also wanted to say, in case there any effect of used platform/source on the reach of tweets and therefore, we extract the source information where we try to tally the data with a retweet and favourite count. Table 2 denotes that using the source: "Twitter media studio" very effective effect on the engagement metrics than "Twitter web app". Basically twitter media studio platform is to measure, manage and monetize uploaded contents that tell a story using videos or images or gifs, overall using visual data which also means that content with videos or images helps to expand the reach to the audiences.

Samsung current Social Media Campaign

	Date	Text	Retweet_count	Favorite_count	hashtags	Source
30	2021-09-13	Enjoy life's little moments, magnified – with ...	26	514	GalaxyS21	Twitter Web App
31	2021-09-10	Something sweeter is already here. #GalaxyZFol...	290	3688	GalaxyZFold3GalaxyZFlip3	Twitter Media Studio
32	2021-09-10	Missing Grandpa and Grandma? This #NationalGra...	29	467	NationalGrandparentsDayGalaxyTabS7	Twitter Web App
33	2021-09-09	Carry your maps on your wrist. Know where you'...	56	794	WearOS	Twitter Media Studio
34	2021-09-09	Thanks to its abundance of camera features, #G...	46	608	GalaxyZFlip3	Twitter Media Studio
35	2021-09-08	Two-toned hues that fit any style, and in any ...	58	844	GalaxyZFlip3	Twitter Media Studio
39	2021-09-07	With more screen real estate, #GalaxyZFold3 5G...	40	588	GalaxyZFold3	Twitter Media Studio
42	2021-09-03	Introducing the #GalaxyZFold3 5G and the new S...	49	674	GalaxyZFold3	Twitter Media Studio
45	2021-09-03	Giseong Jo and @JessicaLong, let's get some me...	21	511	TeamGalaxyCheerTogetherParaSwimming	Twitter Web App
47	2021-09-02	We're rooting for Giseong Jo and @JessicaLong!	24	611	TeamGalaxyCheerTogetherParaSwimming	Twitter Web App

Table 2: Samsung Data Frame

Current Social Media campaign: To have a quick impression on recent tweets in Samsung, we plot a word cloud.

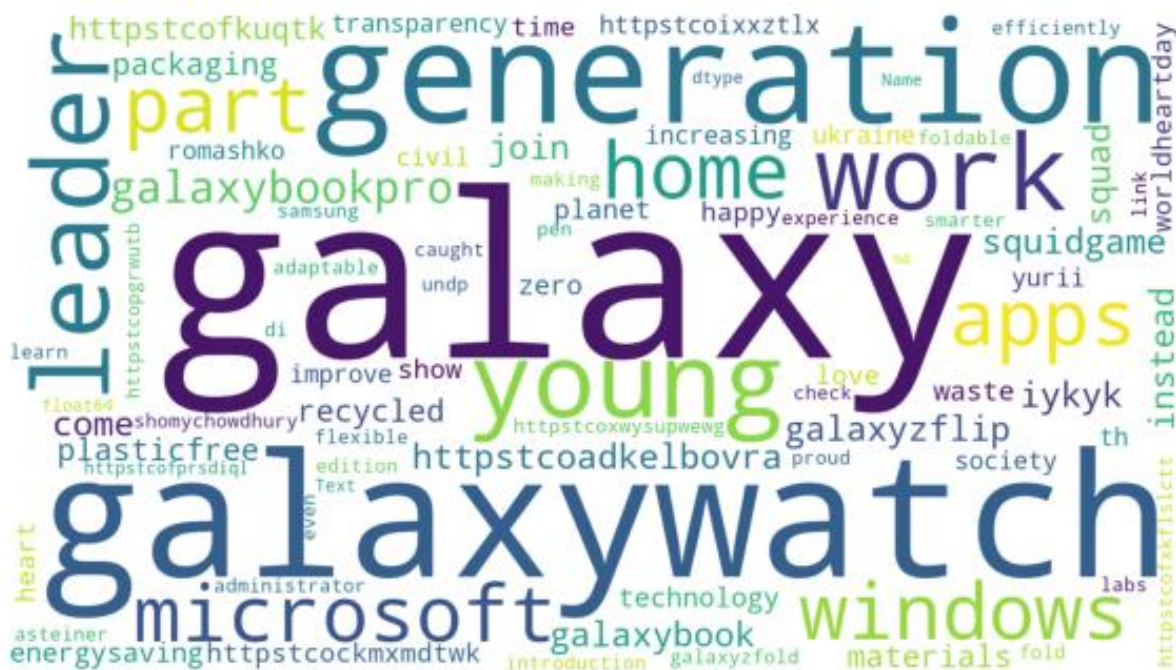


Figure 11 : Word cloud of recent Social media campaign of Samsung

From figure 11, it can be perceived that here the most prominent words are “generation”, “galaxywatch”, “Microsoft”, “leader”. Apparently, it can be assumed that they are running some new campaigns. We also analysed recent hashtags trends to keep track of the latest social media campaign, where the data period was from the last week of September.

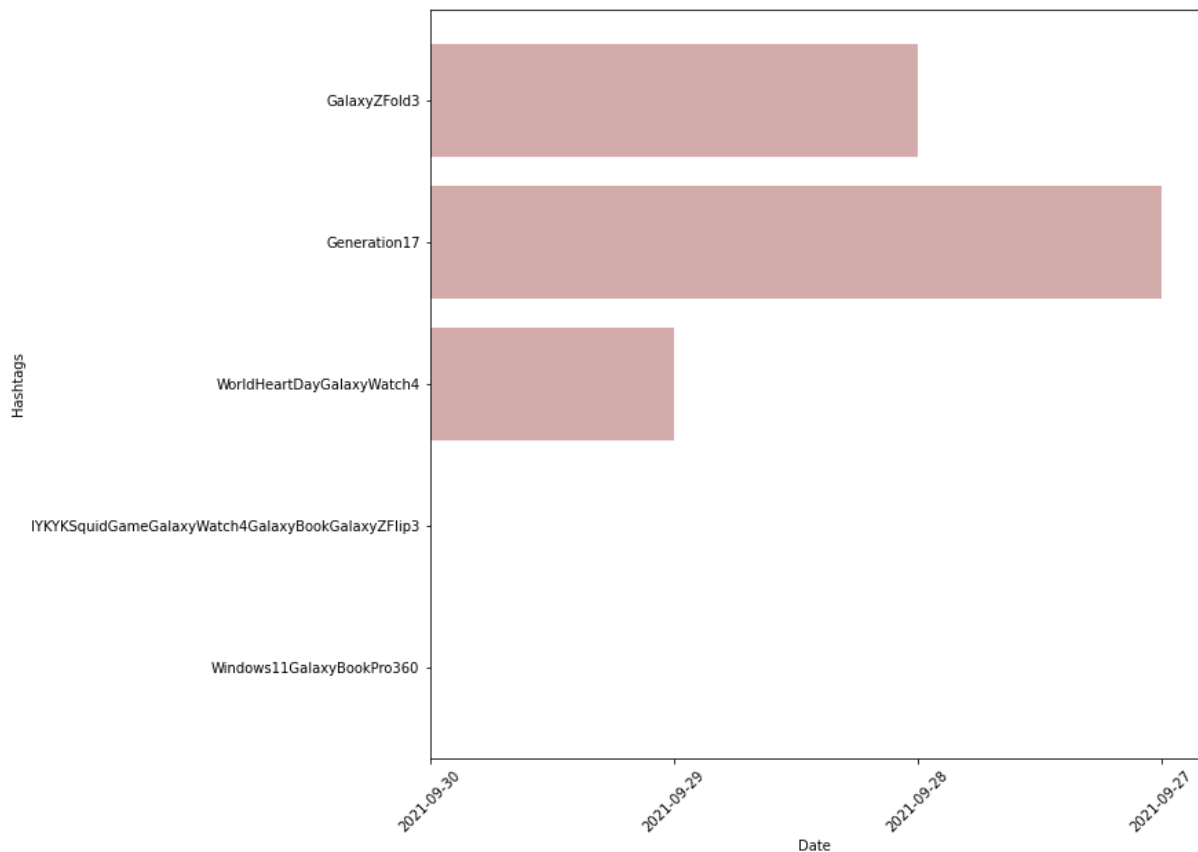


Figure 12 : Huawei recent Date of hashtags

Figure 12 shows the current social media campaign over time. Though our data is collected for the last week of September, some days are missing because no hashtag trends were run on those days. The most recent hashtag trend is “Windows11GalaxyBookPro360” which starts trending on 30th September. The message they are spreading through this campaign is their new product Galaxy Book Pro 360 5g which will not only support connectivity but will also become the first Samsung laptop to ship with windows 11 (Jones, 2021). We can see the hype in this campaign which is very innovative and has never been done before. This hashtag also includes “windows11” which is a branded tag. This hashtag can also attract the audience from the windows fan base and hence more reach to the audiences. Overall, this leads to the conclusion that Samsung is currently implementing a branding hashtag campaign.

Dashboard Implementation and Metric used for analysing

We used 4 types of metrics to monitor and measured both company in term of non-financial and financial.

Daily increase of followers is measured because new followers mean that audiences think your profile is important or interesting enough to watch and regularly get your content in their feeds. This is a useful metric for brand awareness and growing your audience.

Conversely, if the number drops, you might need to investigate further to see what the causes were.

Daily number of tweets is measured because we can see how active is the page or the company that tweets their content and see how much engagement or new followers did they get per tweet.

Daily increase of retweet count is measured because retweets can help companies to promote their tweet more further and expand their marketing to reach different types of customers and attract new customers to their company or pages.

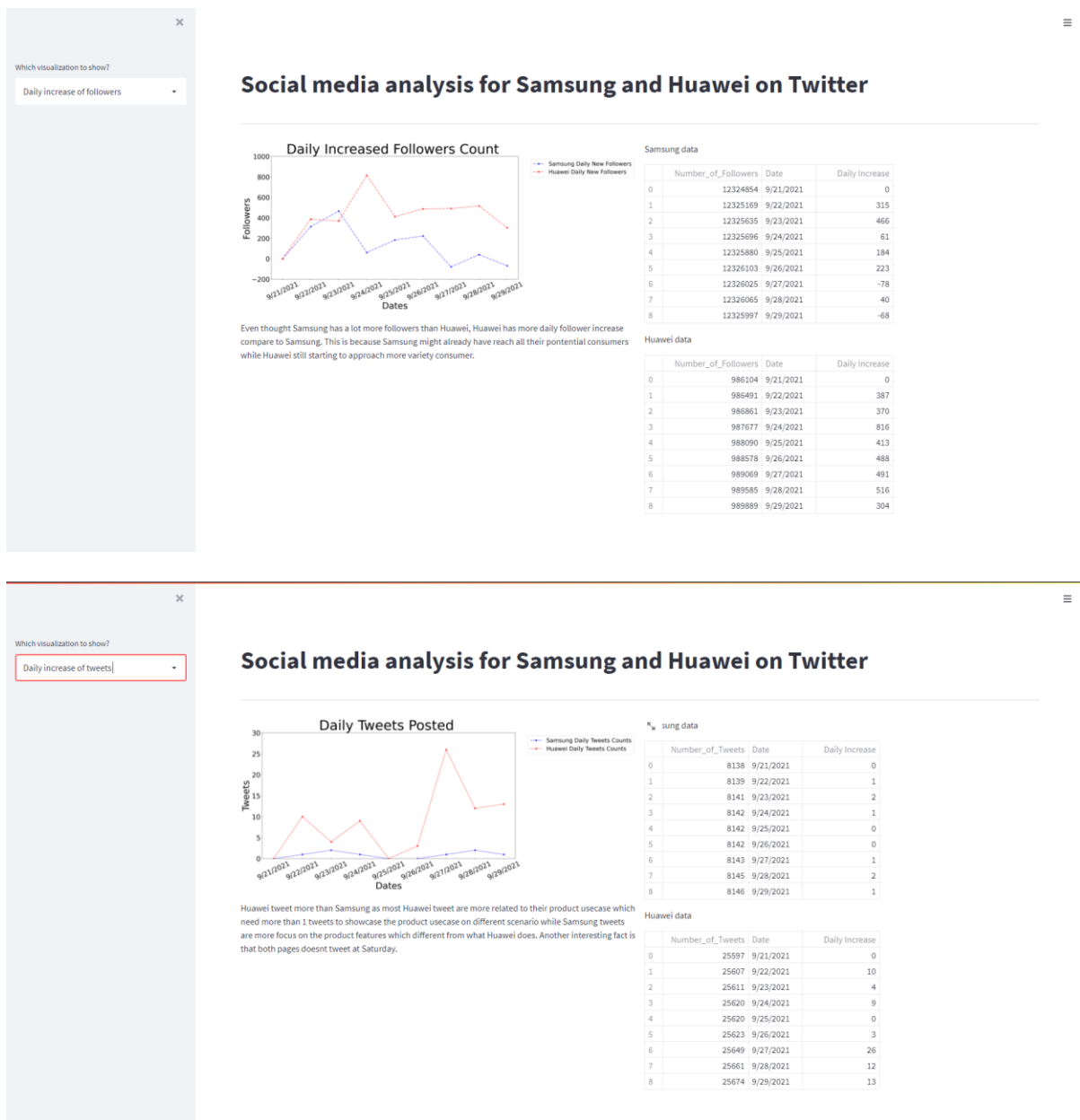
Daily increase of favourite count is measured because it at least can show that the customer interacts with the post which can be an engagement matrix to be monitored.

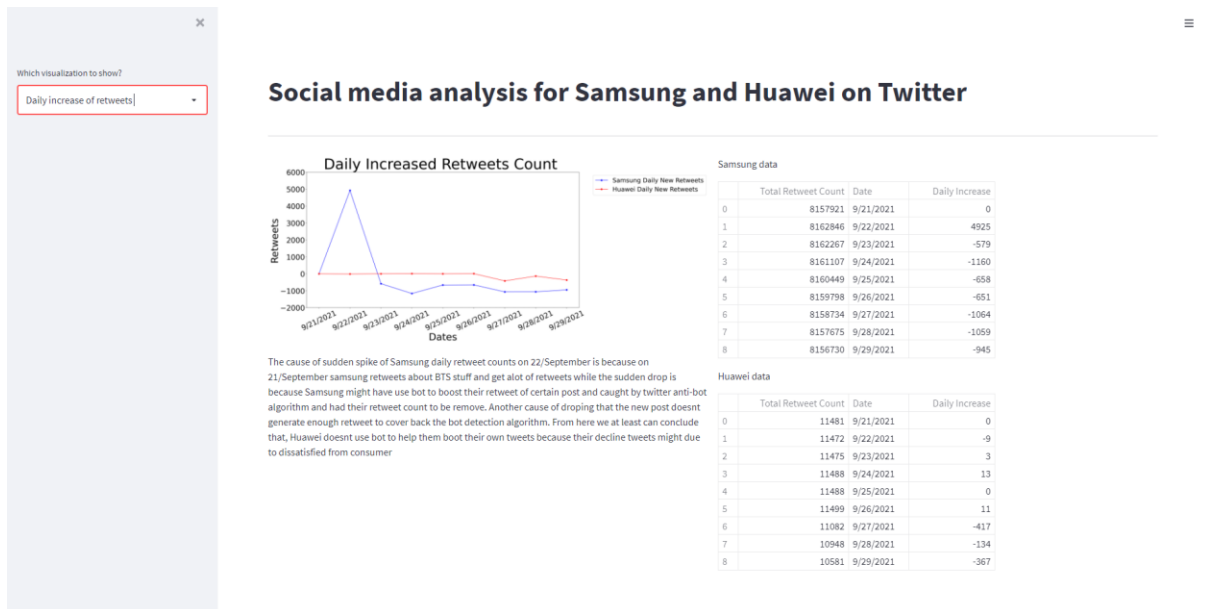
Using the script that we created, we manage to scrape the data of Samsung and Huawei from their Twitter account. Data we collected are Number of Tweets, Number of Followers, Number of Friends, Total Retweet Count, Total Favourite Count, Average Retweet Count, Average Favourite Count and Date. Metrics that we are measuring are the number of daily increases of followers, tweets posted, favourites and retweets for both brands.

So there will be 4 graphs to show the increment of the data for both brands. For the graph, we are using Matplotlib to plot it out and write some story telling based on the line graph. Steps are relatively easy, first we read data we collected from the .csv file and plot accordingly to followers, tweets posted, retweets and favourites. Do note that the data we scraped are the daily total, so to obtain the daily increase we will need to take the number of days after minus

the number of days before. This can be archive in Python by a single line of code: `datalist = [y-x for x,y in zip(datalist,datalist[1:])]`

With the graph obtained, we will then create a dashboard using Streamlit and visualize all of the results for a better overview.





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.

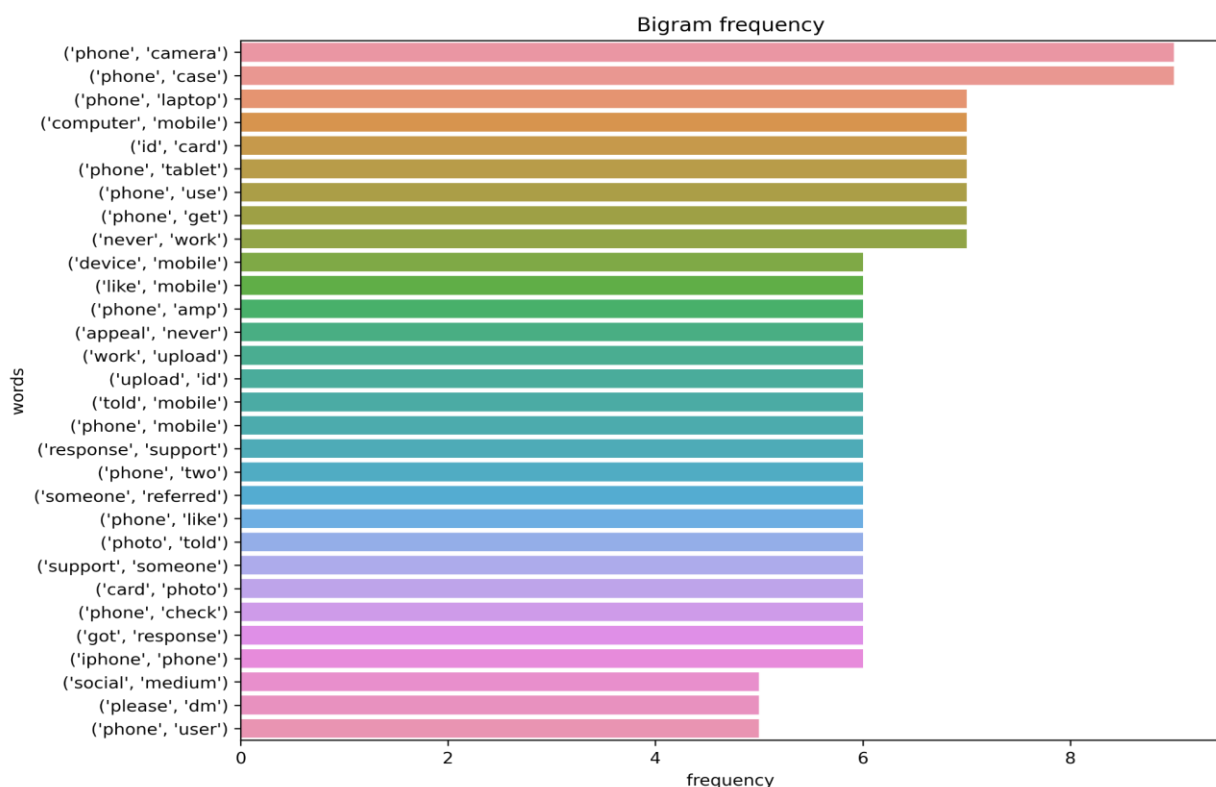


Fig: 2(Word Frequency of Aspect Level Sentences)

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

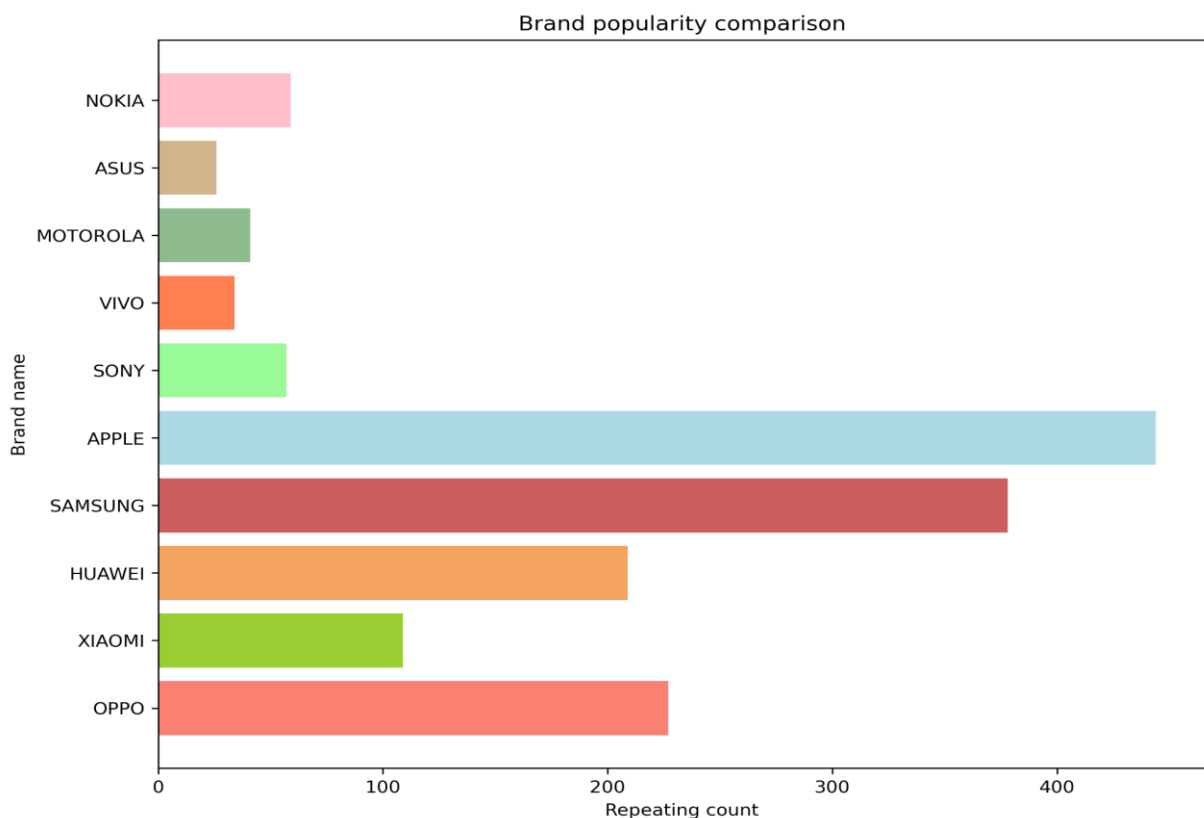


Fig: 3(Brand Popularity Comparison)

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.

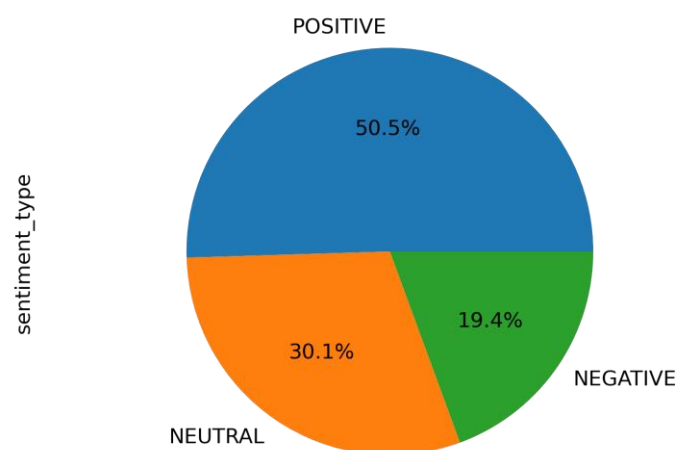


Fig: 4(Sentiments of Samsung)

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.

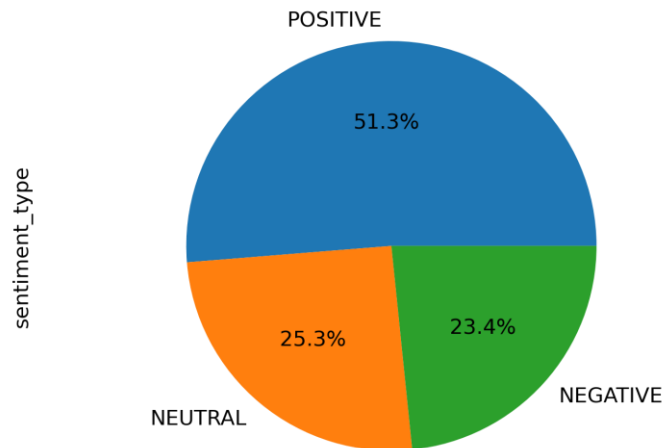


Fig: 5(Sentiments of Apple)

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.

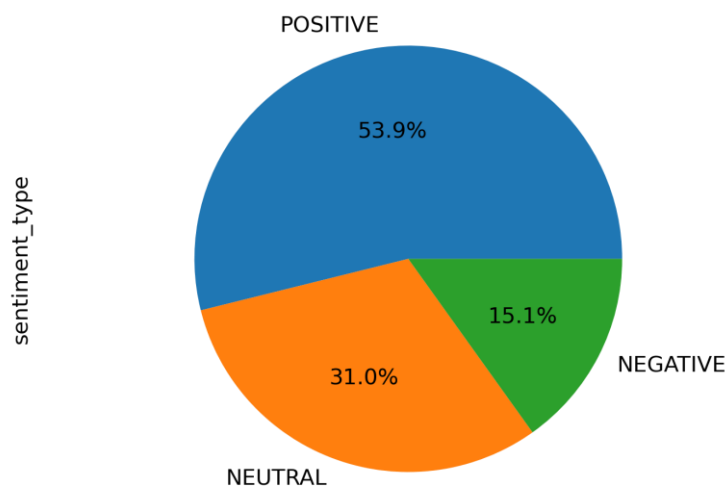


Fig: 6 (sentiment of Huawei)

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.

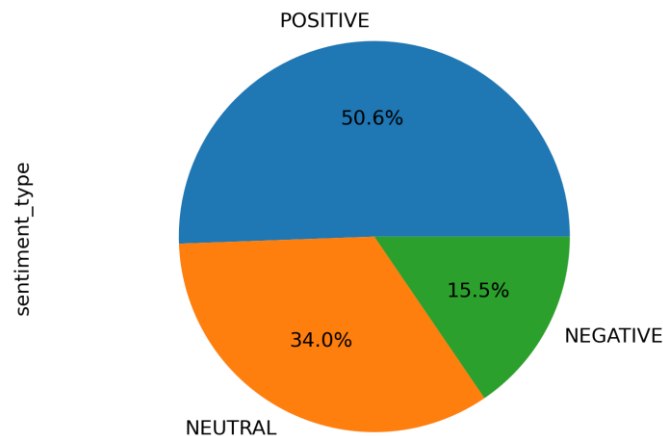


Fig: 7(Sentiments of Oppo)

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.

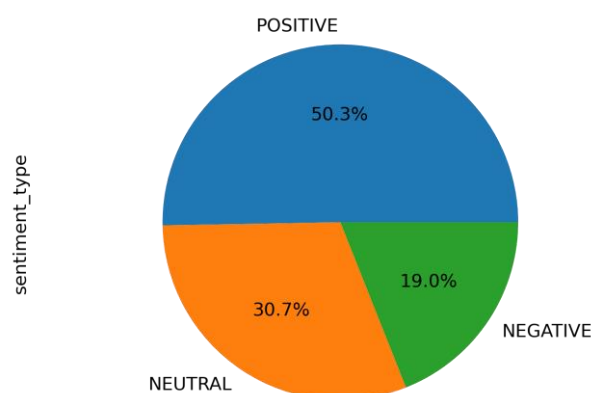


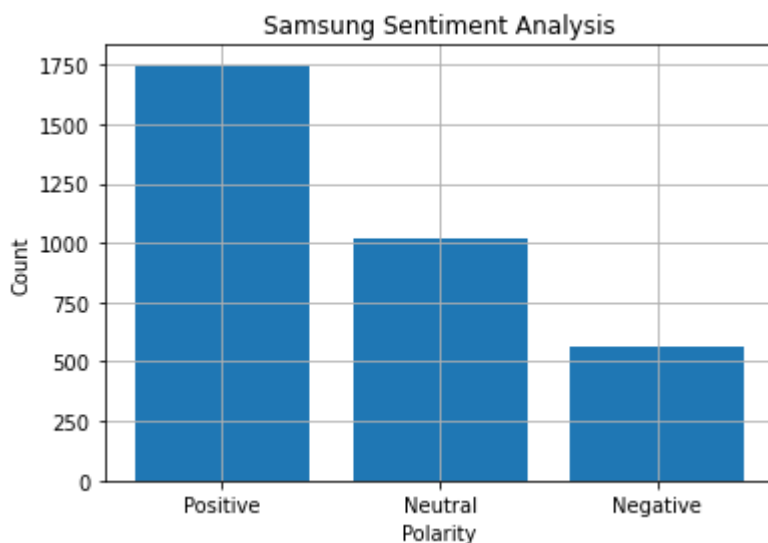
Fig: 8(Sentiments of Xiaomi)

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

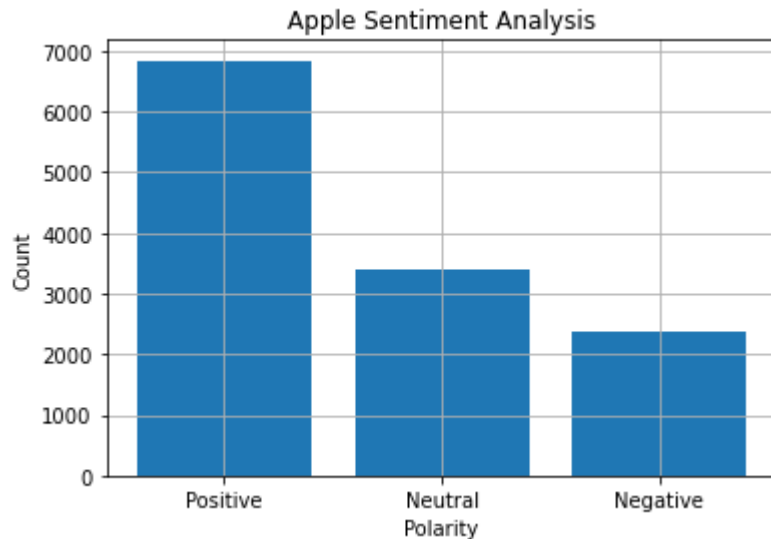
Next, we now acquired the cleaned data and are ready to be analysed. We are using Text Blob to analyse the sentiment of users towards the brands. First is the target brand Samsung:



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

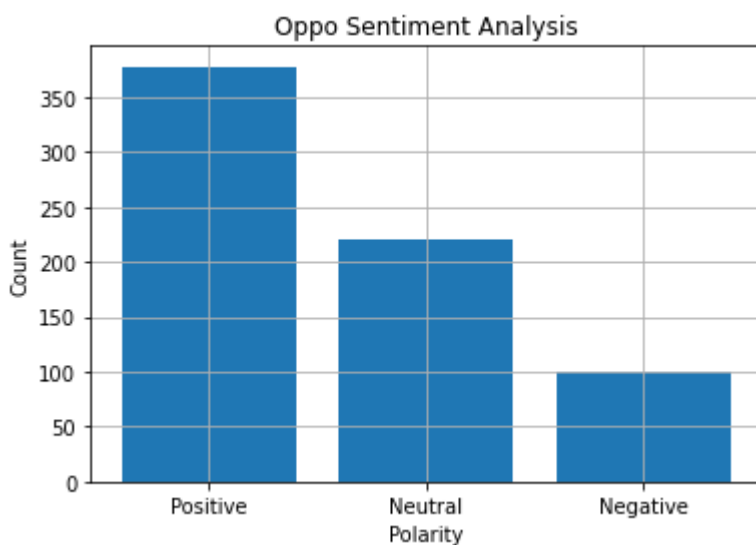
Next, we will be analyzing the sentiment of users towards Apple. Similar to Samsung, Apple is one of the most discussed brands in the mobile phone industry by having 80k tweets collected over one week and the cleaned data of Apple still having around 12k tweets to be analysed.



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

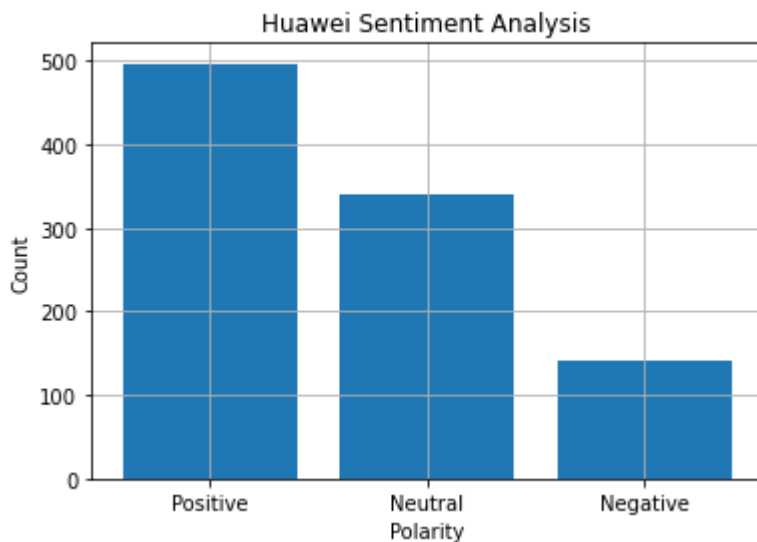
Oppo is also one of the famous mobile phone brands in the industry. Although it is not as popularly discussed as Samsung and Apple, it is also important for us to understand the sentiment of twitter users towards this brand.



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

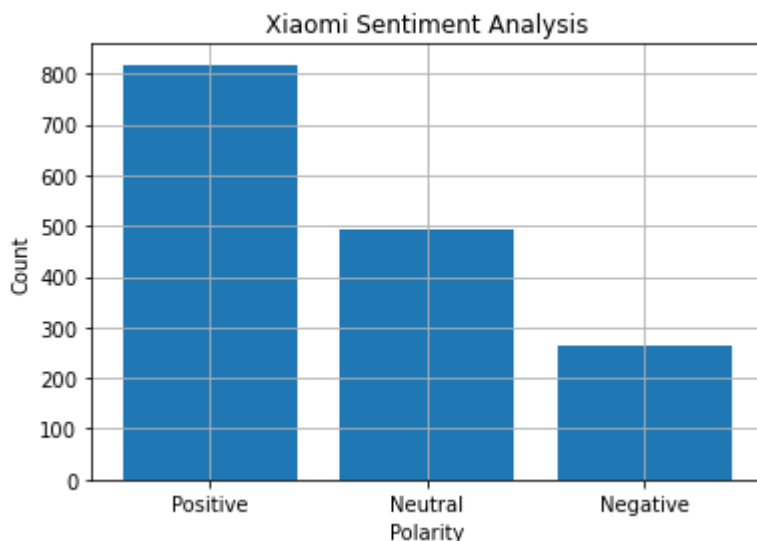
Next brand we will be analyzing is the brand we use as a comparison to Samsung in assignment 1, Huawei. Same as Oppo, Huawei is also a China brand and not so popularly discussed on twitter by having only 8k raw tweets and 1k cleaned tweets.



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.



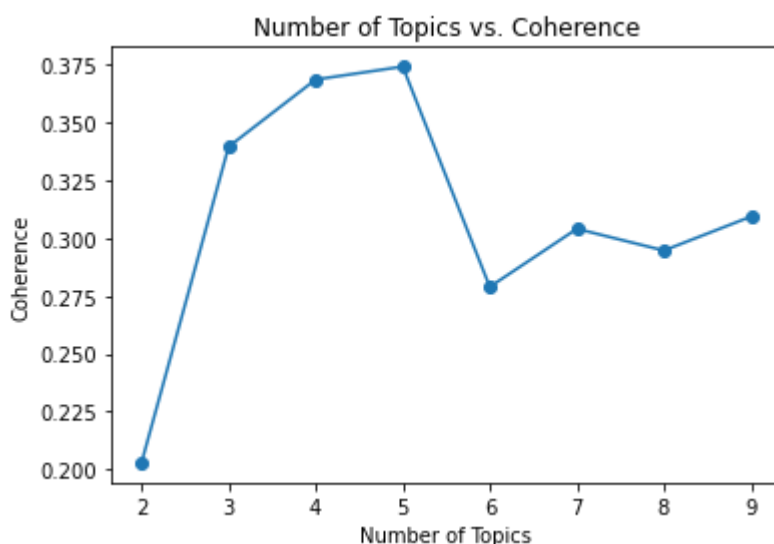
The figure shows the sentiment analysis of Xiaomi on tweets collected and it is positive overall. Xiaomi is getting 818 positive sentiment tweets from users, 492 neutral tweets and 265 negative sentiment tweets from the users.

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...

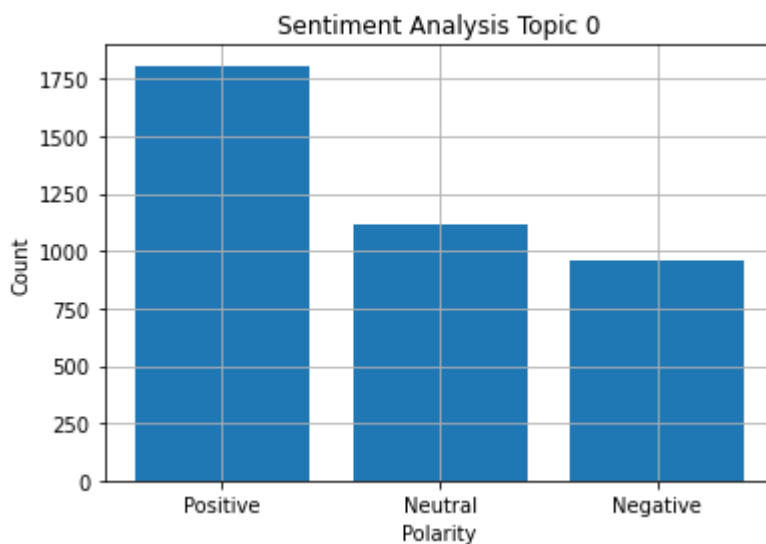
Next we can extract and group the tweets that have the same topic and perform sentiment analysis on each of the topics.

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

Topic 0

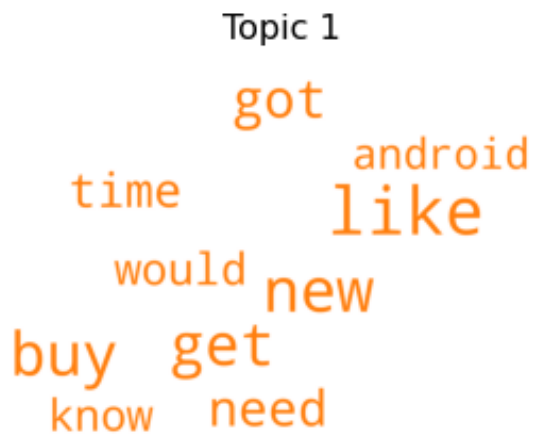
year product
service bought
india issue
month
please
day battery

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

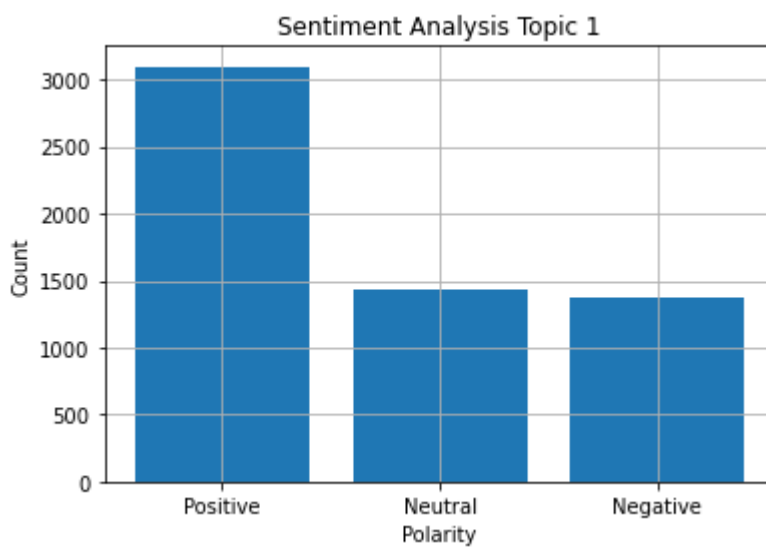


Looking at the sentiment analysis of topic 0, most of the users have positive sentiment when they are discussing topic 0. Positive sentiments are the majority but the number of negative and neutral sentiments is more than the positive sentiment. This can mean most of the users are happy with the topic 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.



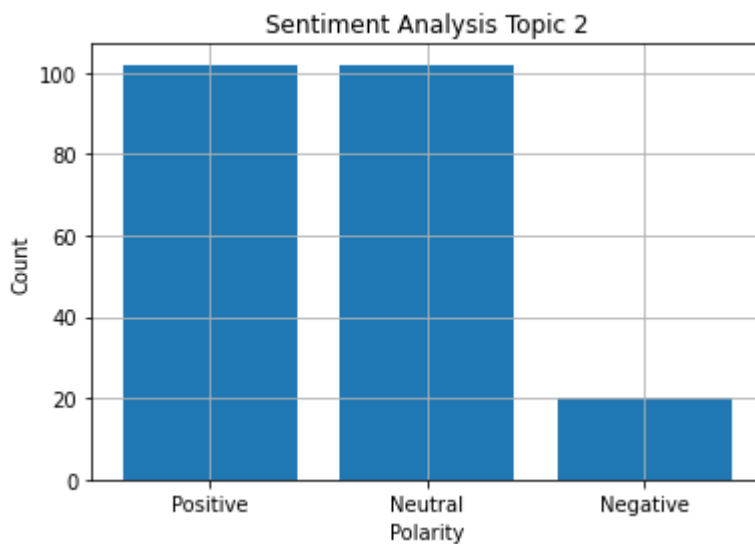
As the figure shows most of the user sentiments towards topic 1 are in a positive way. Neutral and negative sentiments are less than half of the positive sentiments.

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

Topic 2

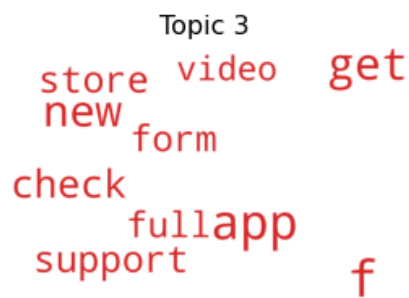
android
gb update
smartphonerealme
camera series
price g amp

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.

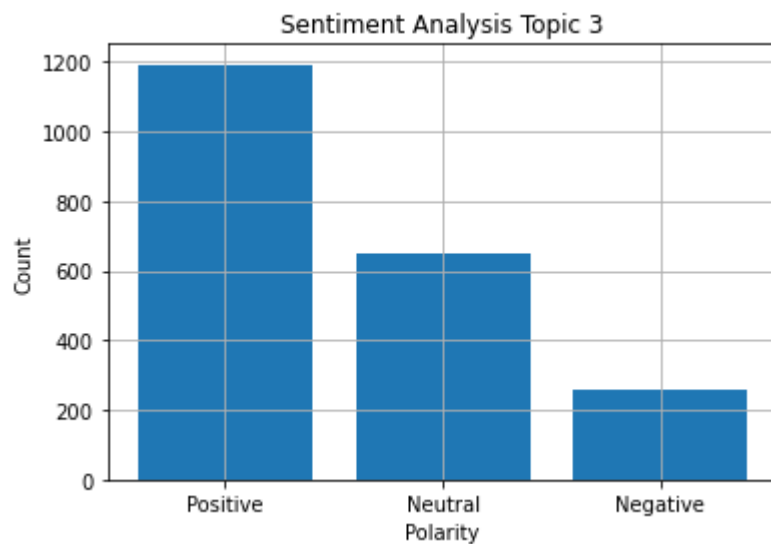


When talking about the specification and price of the mobile phone, most of the users are having positive and neutral sentiments towards the topic. Only a minority of the users are having negative sentiments towards it.

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

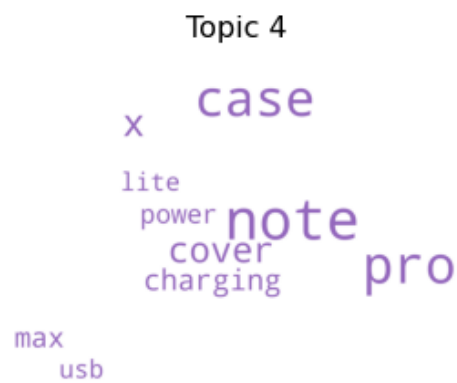


As the figure above shows, the keyword from topic 3 is kind of hard for us to infer what is the main focus of the topic. The only idea we can infer is maybe users are talking about mobile phone applications that need to have support for new features or something.

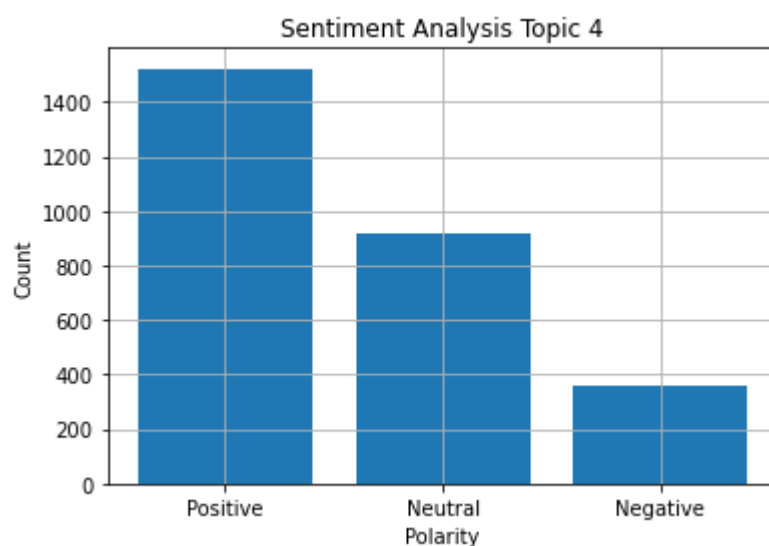


When it comes to topic 3, most of the user sentiment towards it are positive and some of them are neutral. Only a minority is negative sentiments.

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



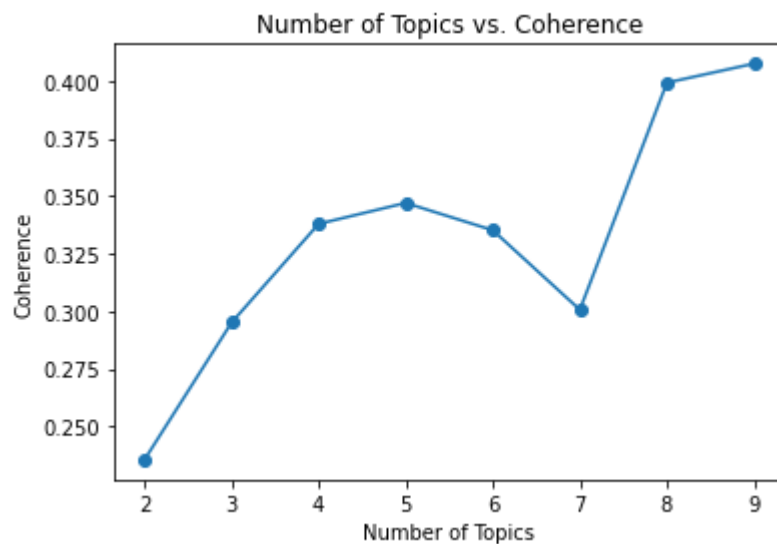
Based on the keyword in topic 4, we can conclude that there are two topics that are discussed and they might be related. First, the user might be discussing the series and the model of mobile phones of different brands because terms like: x, lite, note and pro appear. Next the users are discussing the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

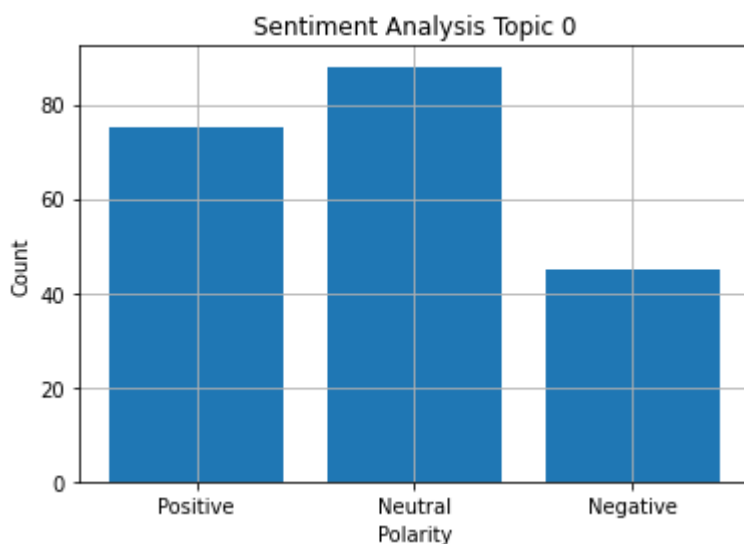


First, we will first infer the topic 0 for Samsung LDA model.

Topic 0

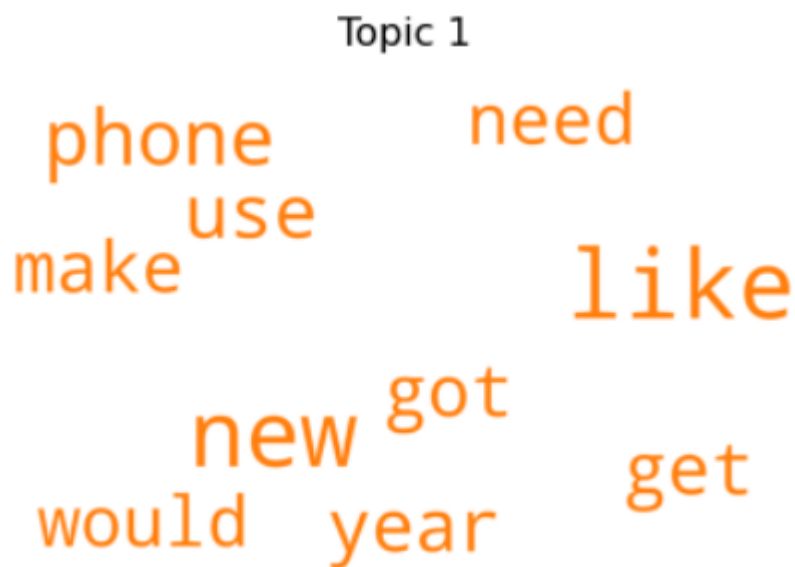
order product
money please
amazon
bespoke since
customer
yesterday edition

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

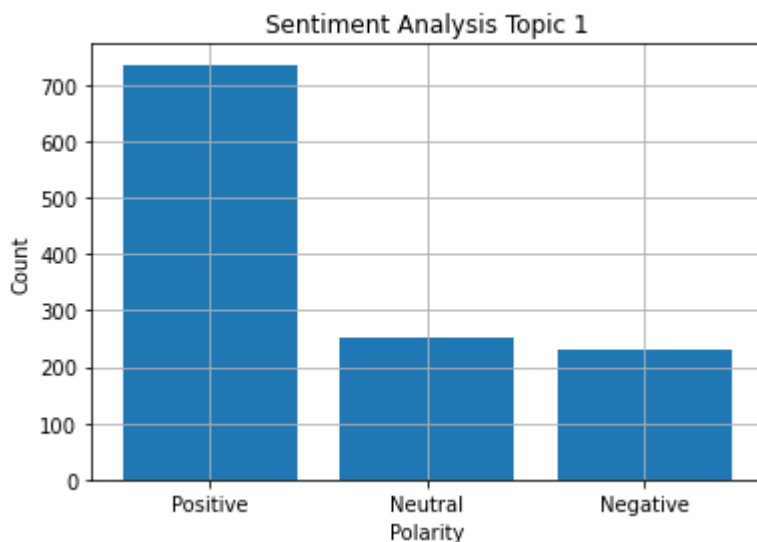


Unlike most of the sentiment analysis results we get along this assignment, sentiment of users towards topic 0 are more neutral instead of positive and negative. With that said, the number of positive sentiments still overpower the negative sentiment in this topic.

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.

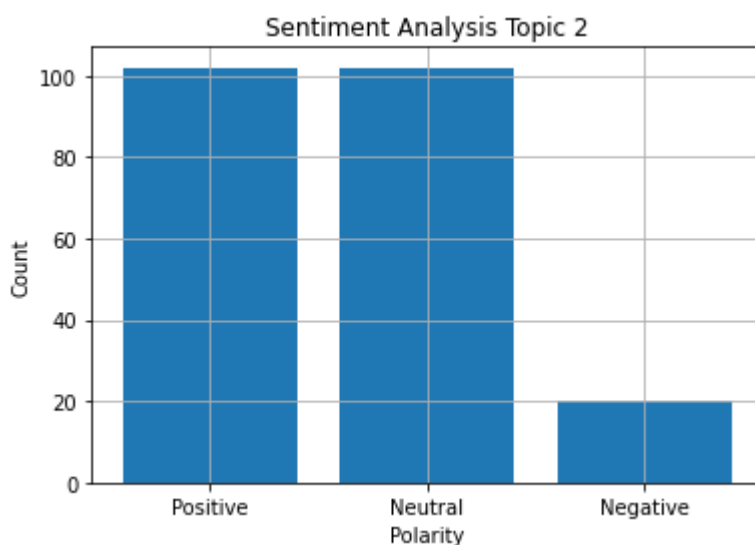


From the figure of sentiment analysis above, we can see that the majority of the tweets get positive sentiment from the users. Only a minority have negative sentiment and neutral sentiment.

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.

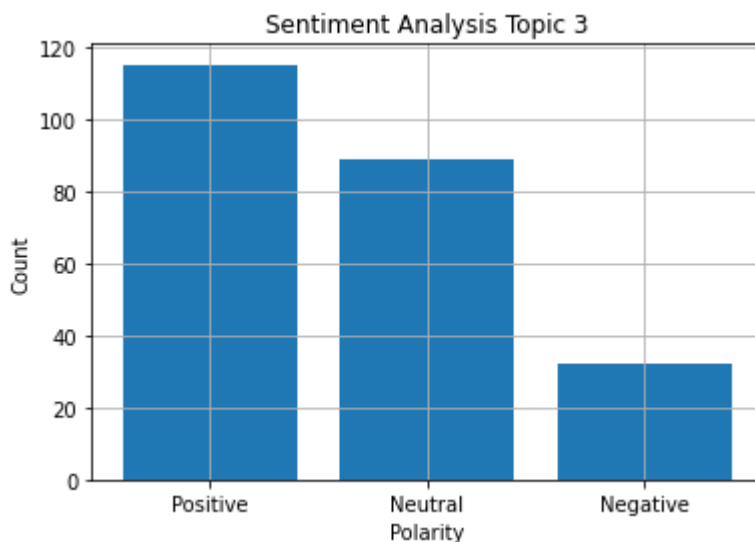


The positive sentiments and neutral sentiments are almost even based on the figure of sentiment analysis of topic 2 shown above. Only a minority of users give negative sentiments towards topic 2.

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.

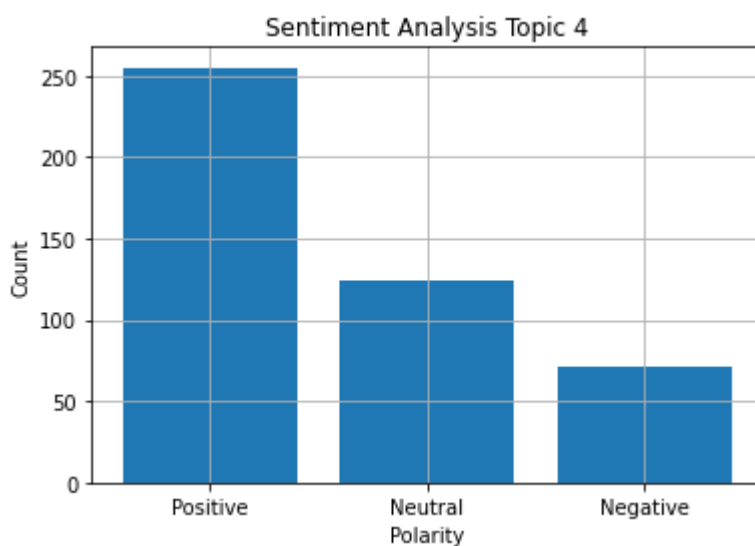


The sentiment analysis of topic 3 only has a minority in negative sentiments and most of the users are having positive and neutral sentiments towards the topic discussed in topic 3.

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

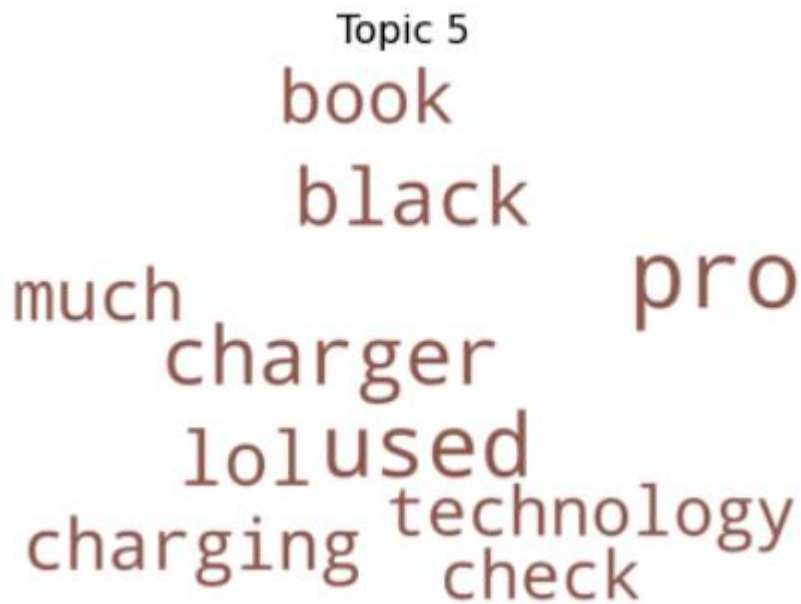


From what the keyword shows, we can infer that topic 4 might be talking about the other products of Samsung like TV and tablet and how they work with the Samsung mobile phones.

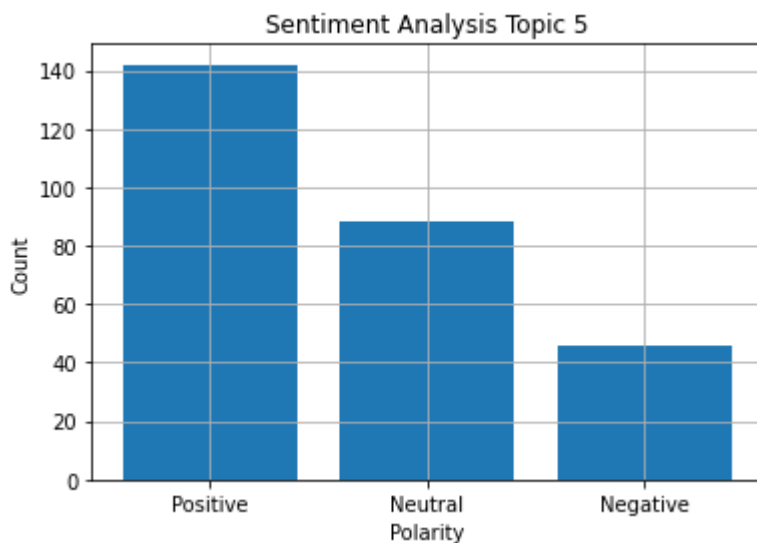


Majority of the user sentiment are positive towards topic 4 while few of them have neutral sentiment and small percentage of it are having negative sentiments.

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.

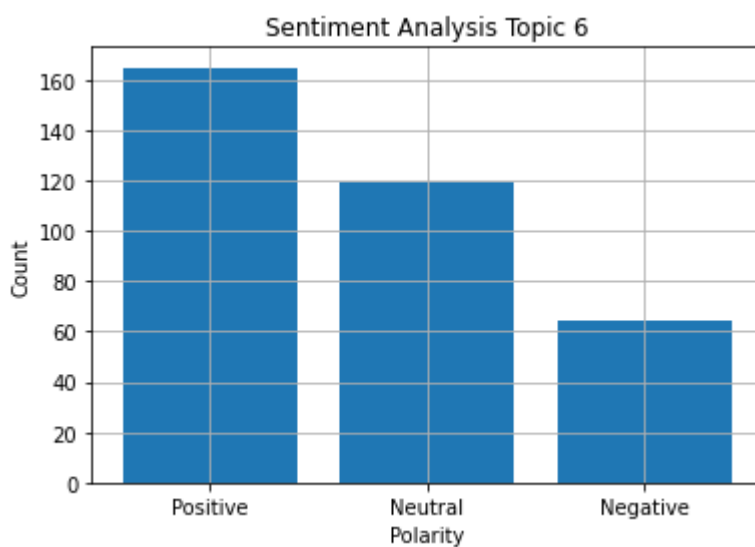


Similar to the previous topics, most of the users are having positive sentiments towards topic 5 and some of them having neutral sentiments. Only a small part of them are negative sentiments.

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

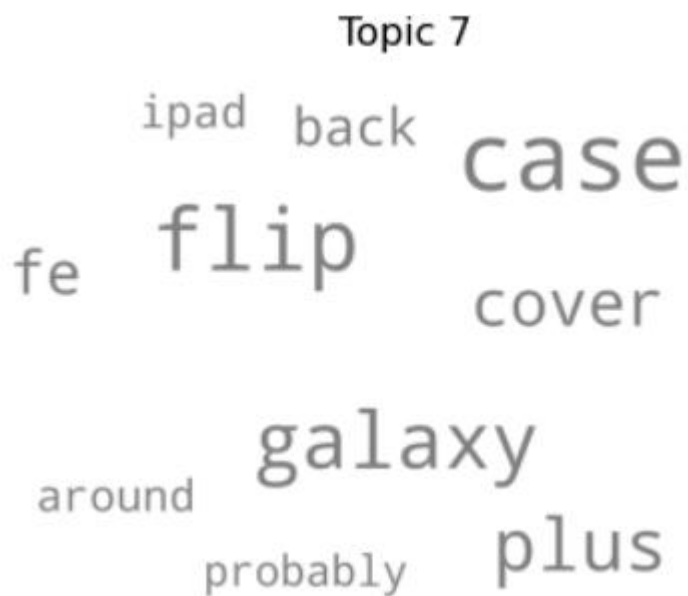


From the keyword we can infer that Samsung and Google are choosing V to represent their brands. V is a member from the Korean group BTS. So the text that belongs to this topic is mostly talking about V promoting their products.

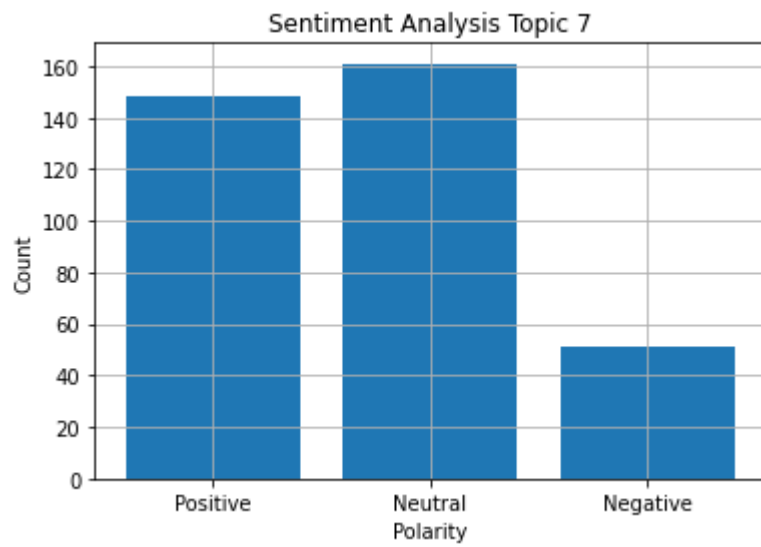


Based on the sentiment analysis figure of topic 6, most of the users are having positive sentiments towards this topic. Only a minority of users having negative sentiments and a large group of users also having neutral sentiments.

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.

Assignment 3 – Social Network Analysis

Introduction

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. In this assignment, we will explain the results we have been found on Samsung and Huawei network.

Data Extraction

The data is extract using tweepy which is an API for twitter. First we scrape only 100 followers for Samsung and Huawei. Then we visualize it normally. After that, we scrape another first 5000 followers from the 100 followers that we scrape later on. For Samsung, we got total of 5940 nodes and 5941 edges in the network while Huawei, we got total of 10556 nodes and 10579 edges in the network.

Centrality Measures

We only use 3 type of centrality measures which is degree centrality, eigenvector centrality and page rank centrality. Degree Centrality is to measure how many nodes are connected to the brand. The degree centrality for a node is simply its degree. For degree centrality, higher values mean that the node is more central and vice-versa.

Eigenvector centrality is to measure the influence of a node in a network. A high eigenvector score means that a node is connected to many nodes who themselves have high scores and vice-versa.

Page rank is used to evaluate the quality and quantity of the links to a page. In our cases, it is to see how popular the brand is by calculating the value of the edges that link to the brand.

Brands	Degree Centrality	Eigenvector Centrality	Page Rank
Huawei	0.009474182851729039	0.010217521076635488	0.0036287085778269686
Samsung	0.016837851490149856	0.011752899619136217	0.006565344263411281

For degree centrality, Samsung brand achieved the highest score of the centrality which indicated there are many connections (edges) for the followers or friends towards the brand and vice-versa. In contrast, Huawei has the lowest score of the centrality.

For eigenvector centrality, the Samsung brands showed many influences of a node in a network. A high eigenvector score means that a node is connected to many nodes who themselves have high scores and vice-versa. Meanwhile, the Huawei brand has the lowest eigenvector score which indicates the least influences of a node in a network.

For page rank, it can see that the Samsung brand has the highest page rank score among the others. Huawei have the lowest page rank score compared to the Samsung brand. Therefore we can see that Samsung is more popular than the Huawei brands.

HuaweiMobile

HuaweiMobile has a total 10556 nodes, 10579 edges and average degree of 2.0044 in the network. From there we found about 23 communities has been form over the network.

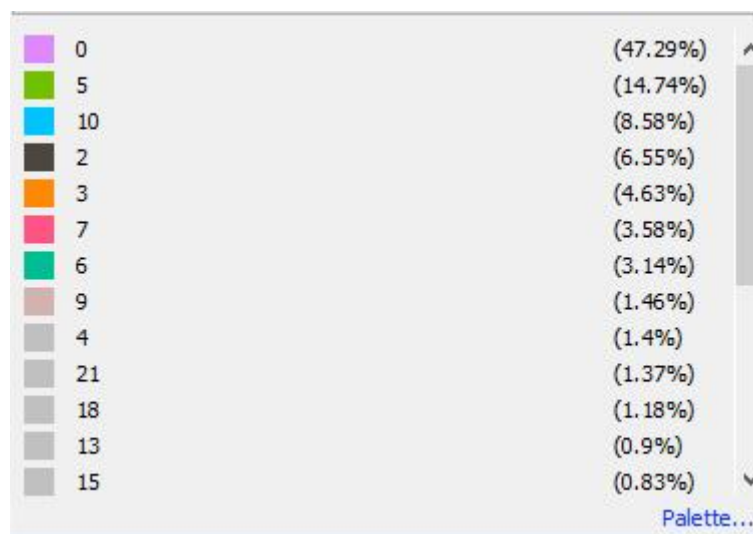


Figure 13: Top community form in Huawei Network

Figure above show the top community from in these network. We will be discussing 1 by 1 of each of the top form community



Rhadebelihle

Figure 14: Top 1 community form in Huawei Network

Figure above show the biggest community form in the network. About 4992 out of 10556 nodes in this network form this community and in the center which is the. This user is social influencer talking about wellbeing and religions stuff. This user have 124.7k followers and we only extract 5k due to computation power. From the 5k nodes, some of them are student, some of them are small startup, some of them are news reporting channel, some also the same as him which is social influencer talking about well being stuff.

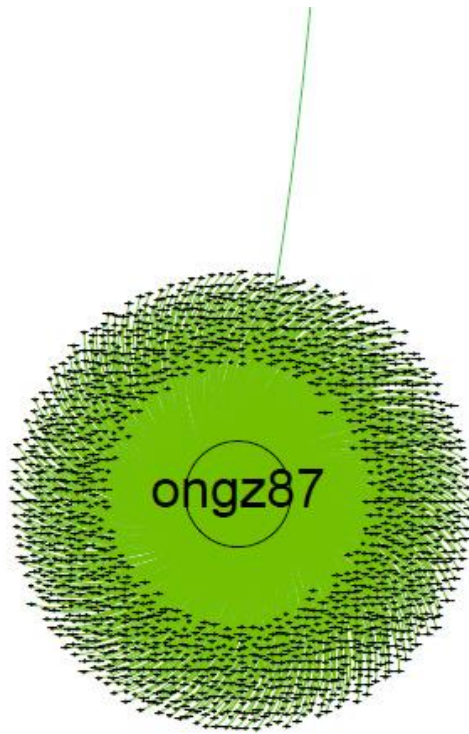


Figure 15: Top 2 community form in Huawei Network

Figure above show the 2nd biggest community form in the network. About 1556 out of 10556 nodes in this network form this community and in the center which is @ongz87. This user is a football fans and movie critic and most of the follower is from Indonesia.

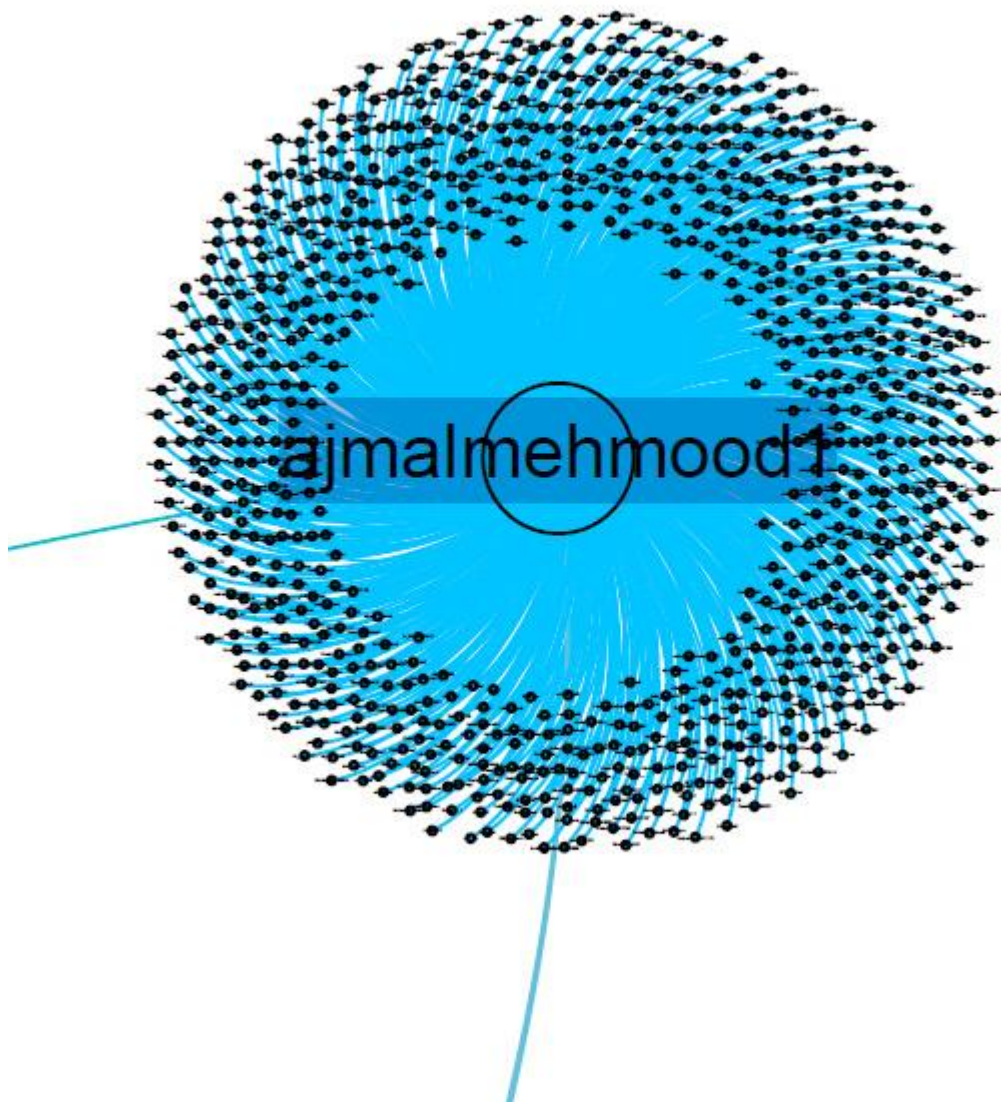


Figure 16: Top 3 community form in Huawei Network

Figure above show the 3rd biggest community form in the network. About 906 out of 10556 nodes in this network form this community and in the center which is @ajmalmehmood1. This user is a radiologist which his profession is related to technology stuff. Most of his followers, is either a politician, journalist or tech nerds.

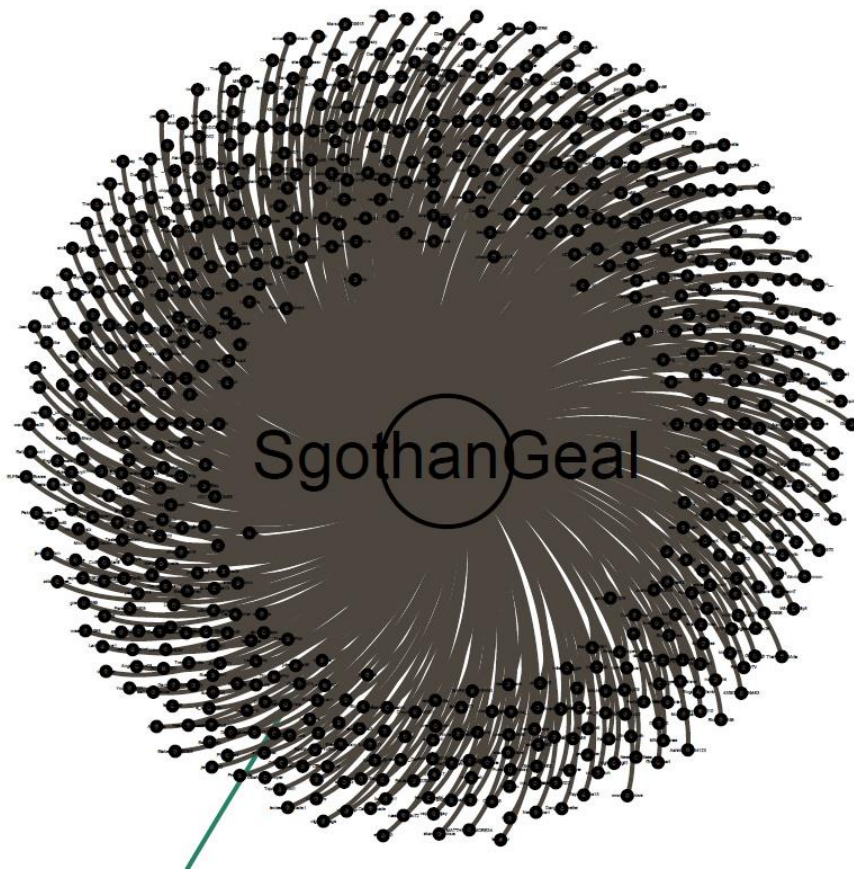


Figure 17: Top 4 community form in Huawei Network

Figure above show the 4th biggest community form in the network. About 906 out of 10556 nodes in this network form this community and in the center which is @SgothanGeal. This user is seller on vape product and flavor. Most of his follower are also that like to vape.

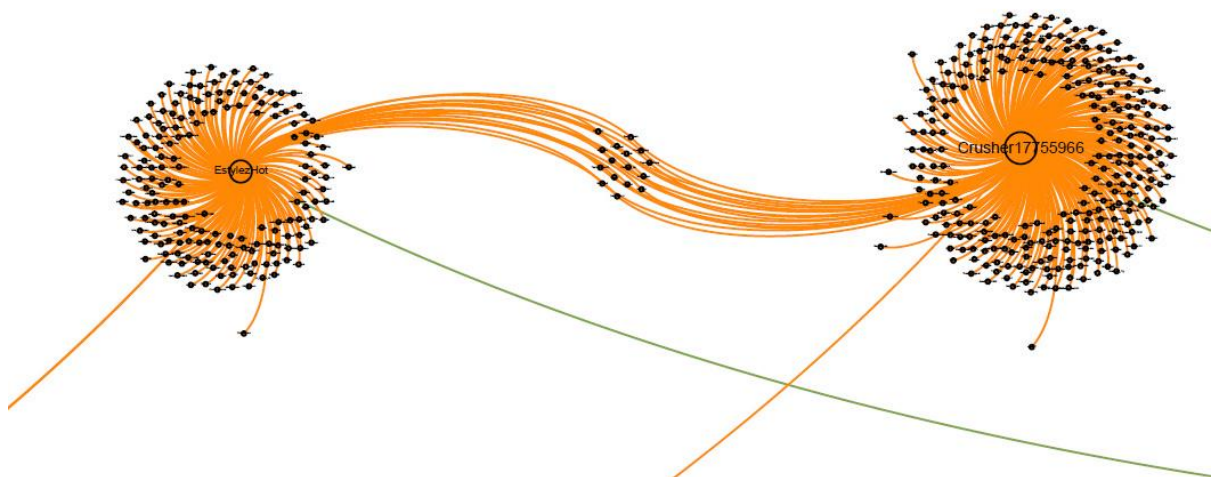


Figure 18: Top 5 community form in Huawei Network

Figure above show the 5th biggest community form in the network. About 489 out of 10556 nodes in this network form this community with 2 main nodes which is @EstylezHot and @Crusher17755966 . @EstylezHot is a pages that sell fashion accessories while @Crusher17755966 doesn't have much information about him other than most of his retweet are about shoes. This community is form due to their background as most of the user is from Africa side.

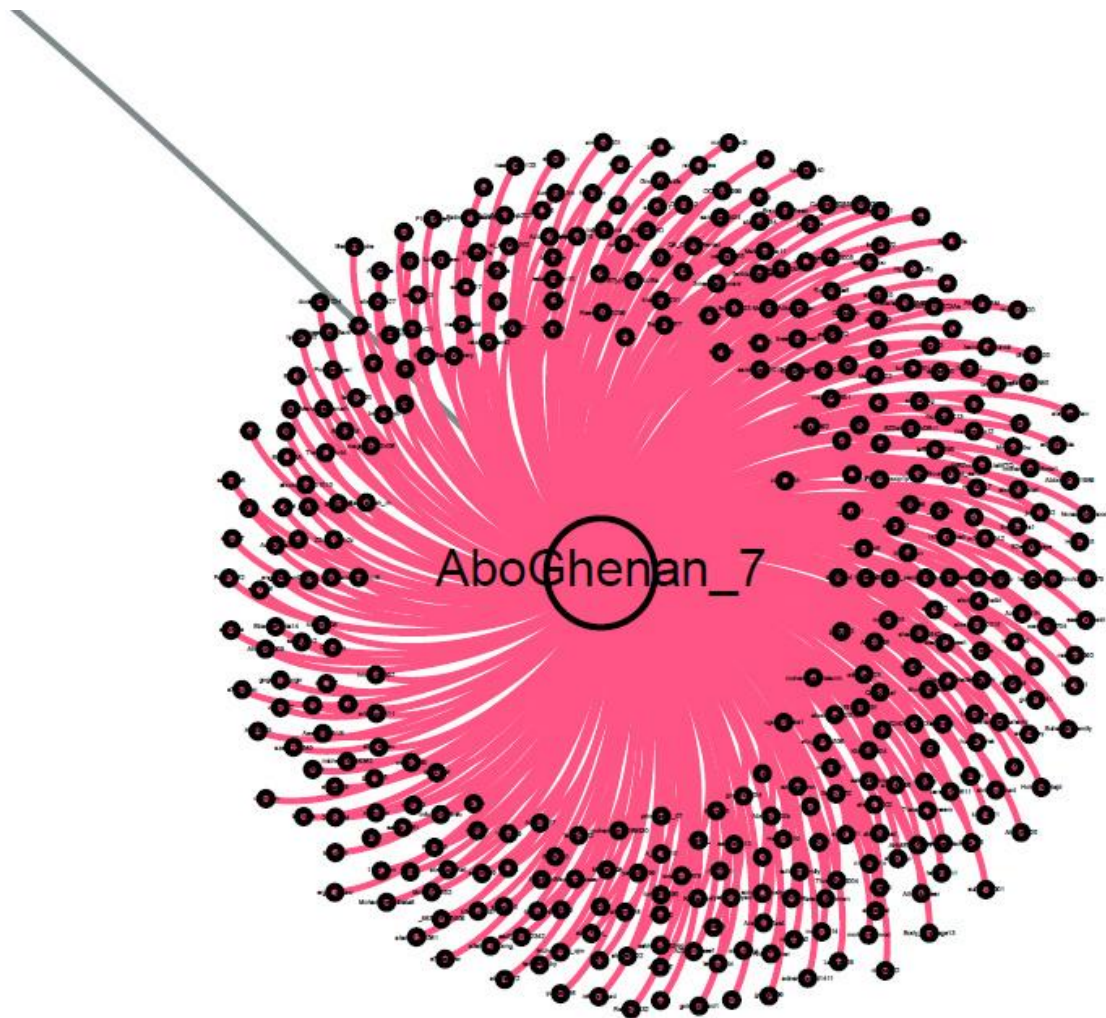


Figure 19: Top 6 community form in Huawei Network

Figure above show the 6th biggest community form in the network. About 378 out of 10556 nodes in this network form this community with the main node AboGhenan_7. @AboGhenan_7 is a tech nerds where most of his retweet are about mobile phone. Most of the followers of his are from the same country as him.

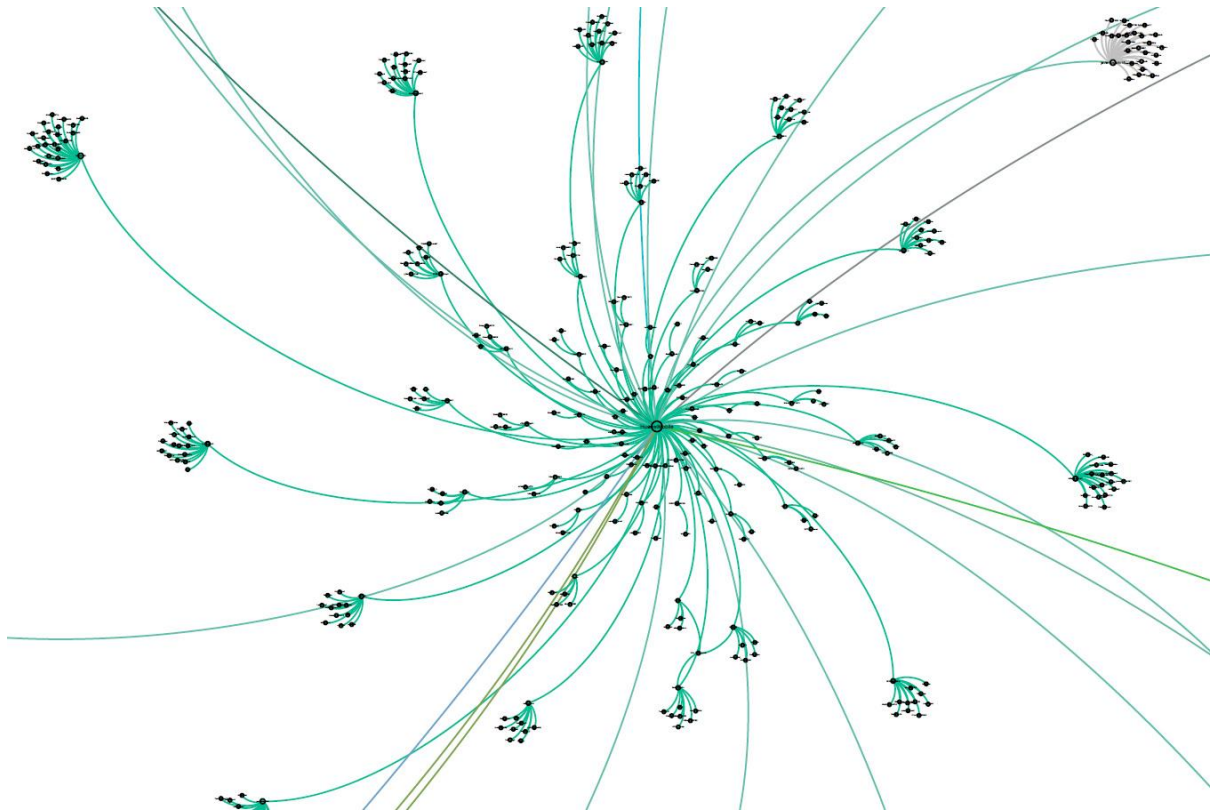


Figure 20: Top 7 community form in Huawei Network

Figure above show the 7th biggest community form in the network. About 331 out of 10556 nodes in this network form this community and in the center which is our main target @HuaweiMobile. This community is form because each of Huawei's follower have their own followers at least 4 to 20 followers.

SAMSUNG

SamsungMobile has a total 5940 nodes, 5941 edges and average degree of 2.0003 in the network. From there we found about 18 communities has been form over the network.



Figure 21: Top Community form in Samsung Network

Figure above show the top community from in these network. We will be discussing 1 by 1 of each of the top form community

ismailduran66

Figure 22: Top 1 Community form in Samsung Network

Figure above show the biggest community form in the network. About 3816 out of 5940 nodes in this network form this community and in the center which is the @ismailduran66. This user is from Germany and is an Islamic person. Most of his follower is from the same country as him which is Germany.

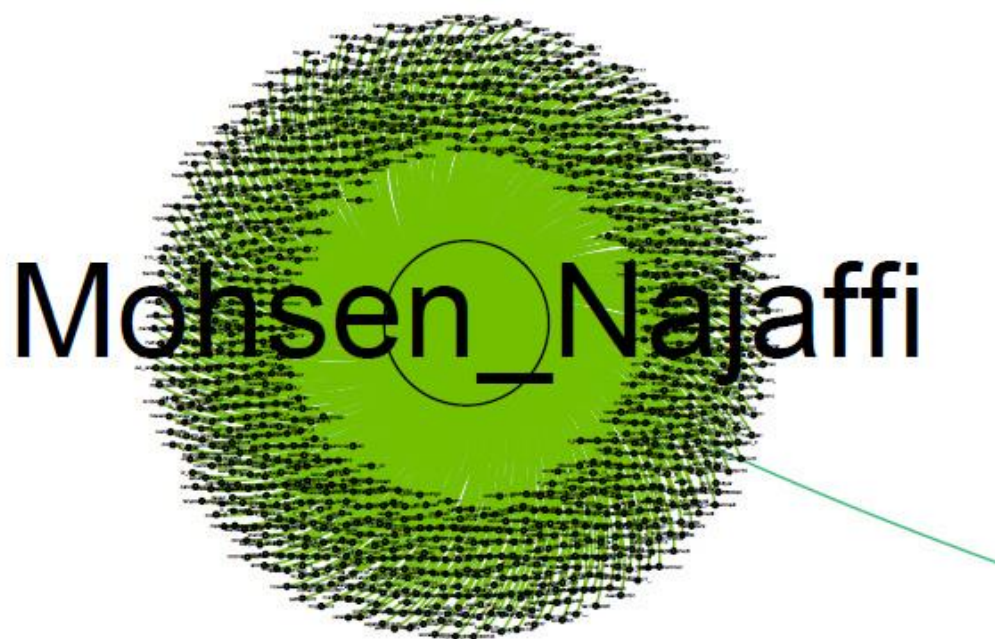


Figure 23: Top 2 Community form in Samsung Network

Figure above show the 2nd biggest community form in the network. About 1210 out of 5940 nodes in this network form this community and in the center which is the @Mohsen_Najaffi. This user is from Iran most of his tweet post is regarding about politician issue in Iran. Most of his followers are from the same country and has the same interest as he is which is regarding politician issue in Iran.

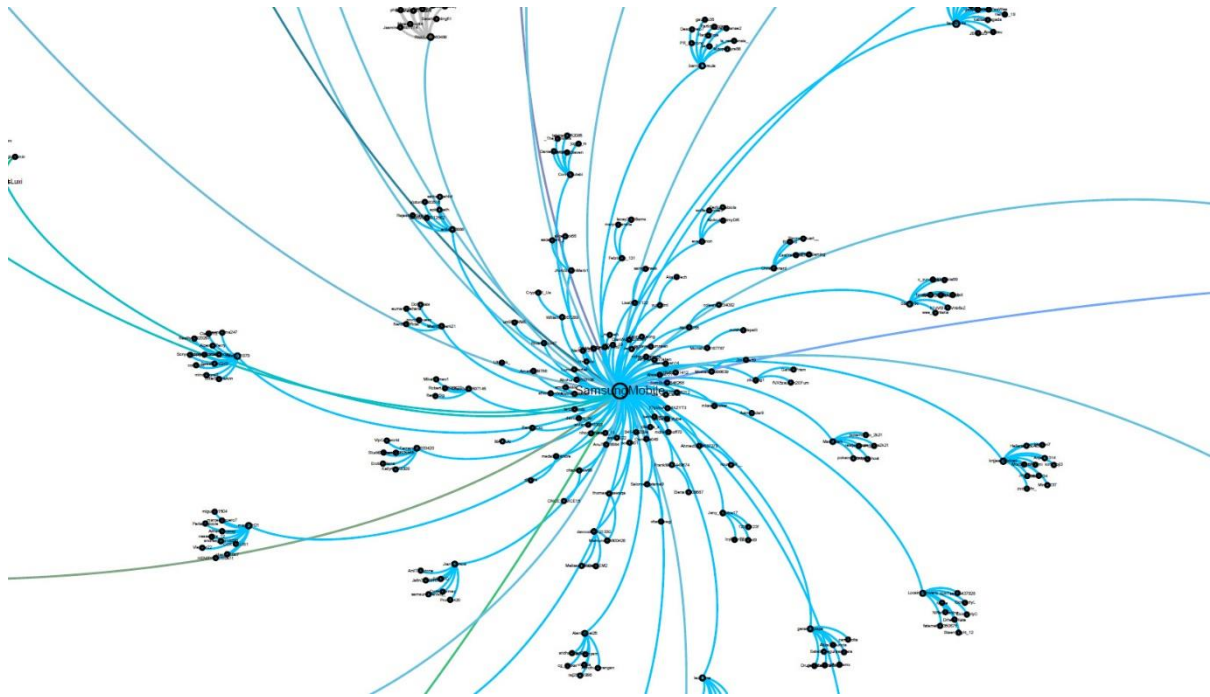


Figure 24: Top 3 Community form in Samsung Network

Figure above show the 3rd biggest community form in the network. About 299 out of 5940 nodes in this network form this community and in the center which is also our main target @SamsungMobile. Why this community is form is because each followers of SamsungMobile have at least 2 to 10 of their own followers.

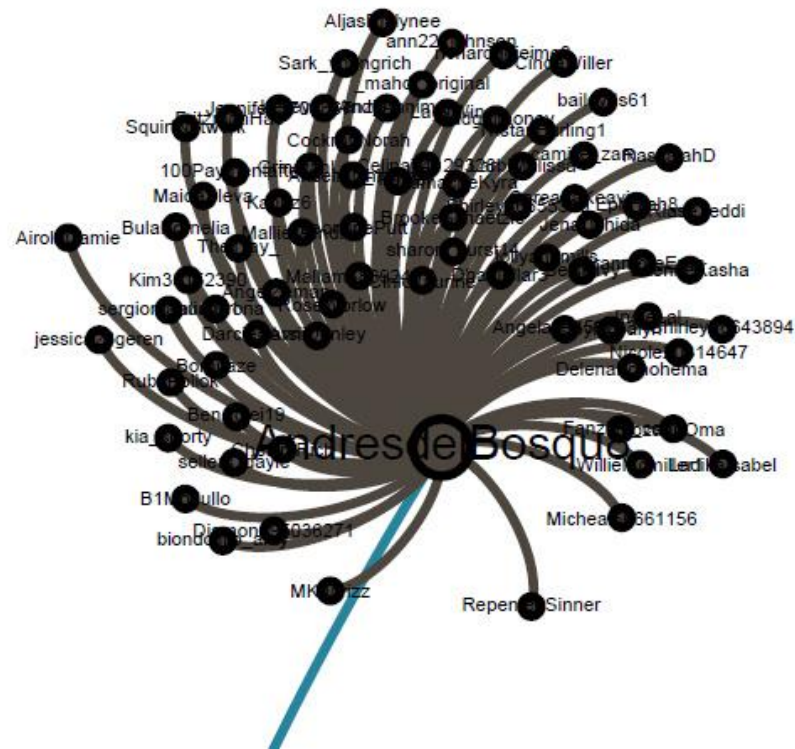


Figure 25: Top 4 Community form in Samsung Network

Figure above show the 4th biggest community form in the network. About 77 out of 5940 nodes in this network form this community and in the center is @AndresdelBosqu8. This user is a football fans. Most of his followers is either bot or adult content creator.

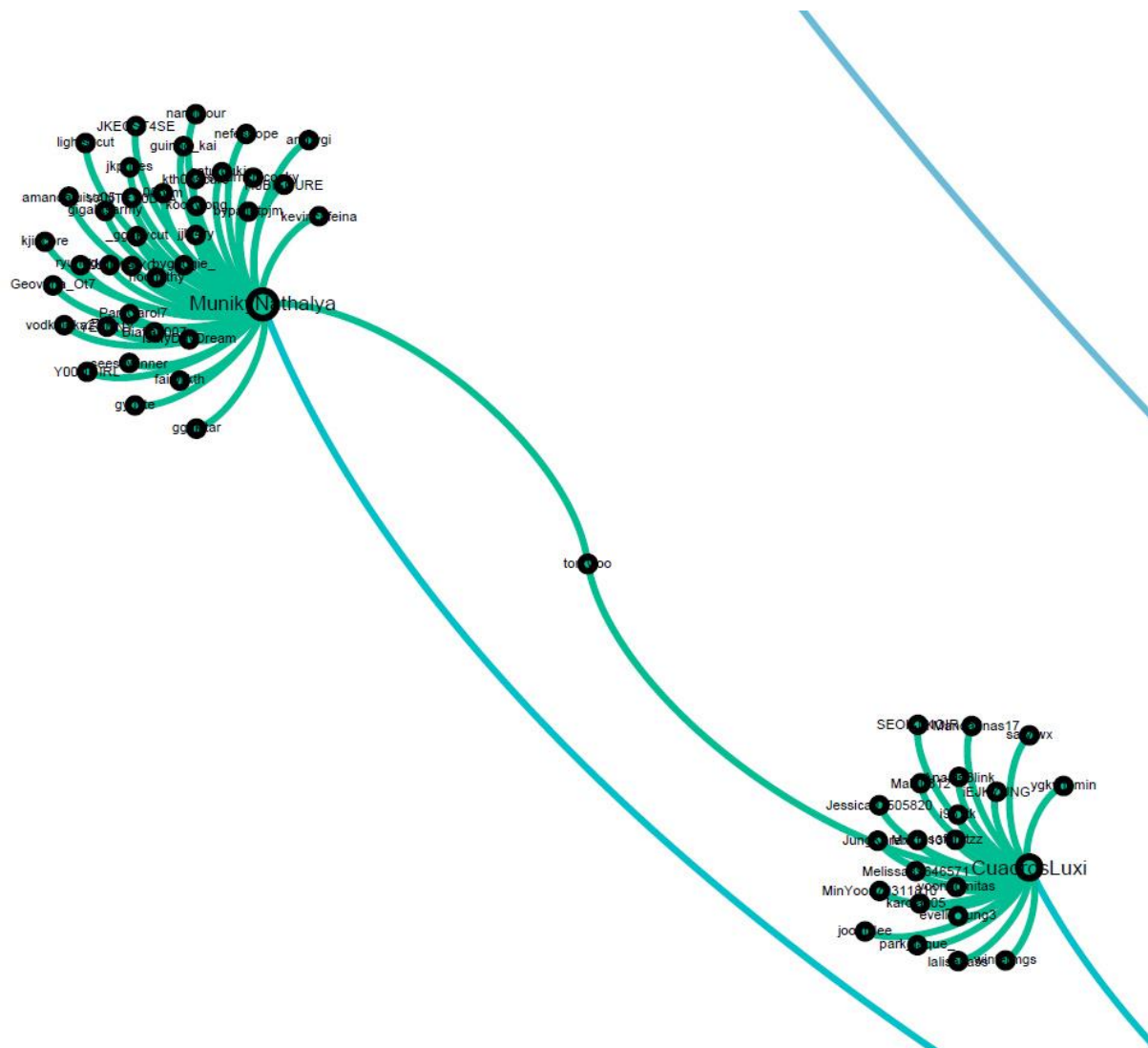


Figure 28: Top 7 Community form in Samsung Network

Figure above show the 7th biggest community form in the network. About 61 out of 5940 nodes in this network form this community and there is 2 center nodes which is @MunikyNathalya and @CuadrosLuxi. @MunikyNathalya is a BTS fans and @CuadrosLuxi is also a BTS fans but this user loves all korean bands like blackpink and other solo korean actor. Both of these account speak in the same language which is spanish.

References

- Chan, T. (2021, August 27). *Power of BTS Compels You: These BTS-Backed Phones Broke Pre-Order Records for Samsung*. From Rolling Stone: <https://www.rollingstone.com/product-recommendations/electronics/bts-samsung-phone-galaxy-z-flip-fold-3-1218006/>
- Jones, R. (2021, September 22). *Samsung adds 5G and Windows 11 to the Galaxy Book Pro 360*. From Trusted Reviews: <https://www.trustedreviews.com/news/samsung-adds-5g-and-windows-11-to-the-galaxy-book-pro-360-4169003>
- Patel, N. (2021, June 12). *Social Media Marketing: How to do it, Types, Tools & Tips*. From Neil Patel: <https://neilpatel.com/what-is-social-media-marketing/>
- ThriveHive. (2015, June 25). *How to Use Hashtag Marketing to Grow Your Business* . From ThriveHive: <https://thrivehive.com/how-to-use-hashtag-marketing-to-grow-your-business/>
- Wikipedia. (2019, May 20). *HarmonyOS*. From Wikipedia: <https://en.wikipedia.org/wiki/HarmonyOS>