Contents

0. Individual Role

1. Project overview

2. Data Crawling

3. Project Pipeline

4. Brand Image Analysis

5. Conclusion

0. Individual Role

LEE NAYOUNG	Data Preprocessing + Sentiment Analysis
CHUNG INYOUNG	Pipeline configuration + Insight provision from result
HAN SOOMIN	Data crawling

1. Project Overview

Brand image refers to the perception, especially the emotional part, that consumers have formed about a particular brand. Consumers form an image by experiencing a product or service of a specific brand for a long time, and the brand image formed in this way has a great influence on a company's market position and competitiveness. Furthermore, consumers make purchasing decisions considering the brand's image in

the market. Therefore, it is important for companies to manage and form their own brand image according to their goals. To do this, we use a variety of strategies to create the desired perception among consumers.

Samsung, one of the representative companies in Korea. Born in the country that is the home of Samsung, we can all answer Samsung's brand image well. Because Samsung's products have been with the majority of people's lives, the thoughts we feel about Samsung soon become the brand image of Samsung. Then, what about Samsung's brand image abroad? Probably not many people can give a clear answer. However, for some domestic conglomerates targeting overseas, the brand image abroad is just as important as the brand image in Korea. Here, we thought we could use natural language processing techniques to approach the brand image of domestic companies in foreign countries, which we do not know well. And as the subject of the analysis, we chose Samsung, a company most familiar to the people of Korea.

2. Data Crawling

2-1 Dataset construction

We looked at several communities where we could find datasets. Among them, Reddit, a large English-American community, provides an API, and since there are many subreddits and a wide variety of users, we decided to use reddit to crawl the data.

2-2 Data Crawled Target

Market Capitalization		
1	삼성전자	3,886,328
2	LG에너지솔루션	1,319,760
3	SK하이닉스	645,738
4	삼성바이오로직스	567,257
5	LG화학	516,030
6	삼성SDI	470,349
7	삼성전자우	449,296
8	현대차	421,371
9	기아	348,207
10	NAVER	321,700

Actually, we initially tried to analyze the perception of Korean companies, not just Samsung. In order to select large companies that are familiar to us, we decided to use the market capitalization with clearer figures as the standard. When Korean companies were listed in order of market capitalization as of 23.5.6, the following results were obtained.

As a result of examining all subreddits of the derived companies, we set target words such as "samsung", "hyundai", "kia", "sk", and "lg", which are the most active and have the most subscribers. In order to get words that can represent the image of the company and have clear emotions, we decided to use a dataset from a subreddit named after the company among numerous subreddits. (Samsung subreddit, Hyundai subreddit, Kia subreddit) The reason we said general here is that there are so many subreddits for specific items such as 'Samsung Galaxy subreddit', 'Galaxy Watch subreddit', and 'Samsung Pay subreddit'. Because of this, in order to obtain universal results, we decided to use only the data from the subreddit that only has the name of the company in question.

However, on 23.6.12, there was a notice that reddit's API would be charged, and many subreddit owners objected to this, and there was a case of changing the subreddit to private mode. And the same goes for the Samsung and Hyundai subreddits. Therefore, we decided to analyze **only the brand image of Samsung** companies with pre-crawled data.

2-3 Data crawling method

And we tried to take an approach with the following criteria, but in the end, all approaches were unsuccessful.

Here are our some failed data crawling approaches:

1. Hot posts, top posts, new posts, and 'controversial' posts, which are the criteria provided by the reddit community
2. Comments on posts classified as controversial were crawled (Posts are classified as 'controversial' if there are more than a certain number of comments)
3. All posts and comments within 1 year

All the results obtained these ways were only very ambiguous results that were difficult to derive insights from. From repeated failures, we took a direct look at the text dataset we crawled and found a few problems:

1. There are many articles that contain humor rather than information.
2. There are several articles for advertising and dominating purposes.
3. There are writings that are composed of profane words and abbreviations that lose their meaning when tokenized.

   In order to escape from these problems, we decided to **directly select a dataset** that is not related to the above three problems by using Reddit API.

3. Project Pipeline

   After data crawling, only noun-type roots are extracted from the corpus of each established company, and the term-frequency is calculated and sorted in high order. After data processing such as unification of pronouns and possessives, statistics are extracted to extract the main keywords for the company. Based on the keywords extracted in this way, we are going to go through sentiment analysis on keywords, but before that, we went through data filtering. Through sentence-by-sentence sentiment analysis, only sentences that exceed a certain emotional threshold score were primarily filtered, and data were secondarily filtered through whether or not they contained keywords related to the company.

   In this filtered data, we went through sentiment analysis on keywords, and data was classified into positive corpus and negative corpus. We extracted keywords again for

the sentiment-classified text, and backtracked the data by closely examining the data based on the extracted keywords to clearly know the positive and negative attributes. By doing so, it was possible to analyze the brand image by giving insights on positive/negative attributes.

4. Samsung brand image analysis

4-1 Keyword Extraction

 The following is the term-frequency result and word cloud for the 'Samsung' corpus.
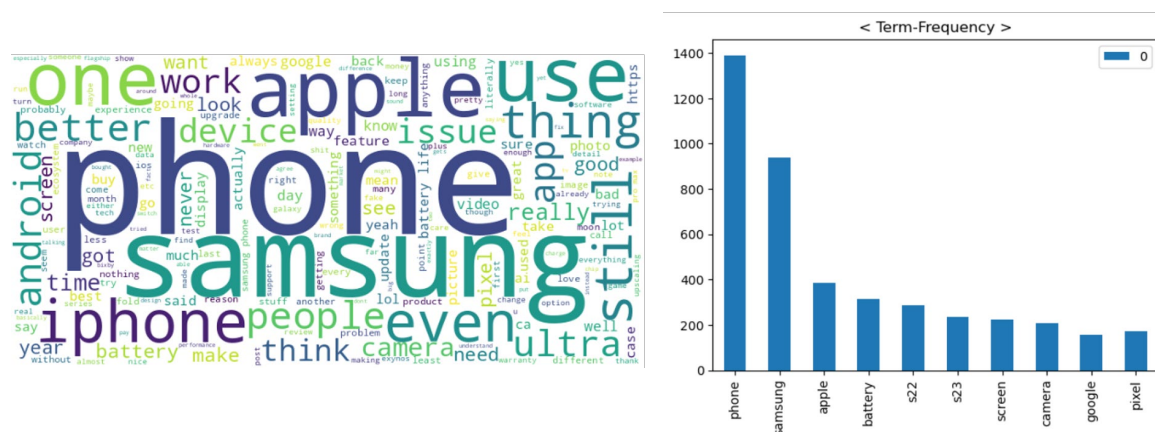
   We realized the following results from these keyword extractions:

1. When you talk about Samsung on Reddit, most of the talk is about phones and their performance.

2. Apple is the company most often mentioned in samsung.

3. On Reddit Google's Pixel appears to be rivaling Samsung's Galaxy model.

So, in order to understand the in-depth brand image of Samsung, we extracted insights from texts in which emotions were classified and analyzed in-depth sentiment on 'phone' and its synonyms 's23, s23', which was extracted as the most keyword. For a more detailed analysis, we analyzed Samsung's brand image through analysis of 'battery', 'camera', 'screen', and 'ui', which are representative attributes of mobile phone functions.

4-2 Sentiment analysis and insights for keywords that have undergone text filtering

Note: The overall positive rate for major keywords is very low, which seems to be caused by the pessimistic nature of the Reddit community.

A. 'Phone' , 's23', 's22', 's22u,'s23u'

Positive : 128 Negative: 369

→ Positive rate : **25.75 %**
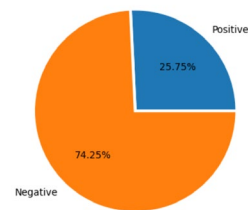
P/N	KEYWORD	Sentences with keywords
Positive	Foldable	foldables is where you can look to for innovation
		i'd say folding touchscreen phones are pretty innovative.

|		
Negative	snapdragon	Also if you really want one of the s22 series phones you could always import it, the snapdragon version is usually much better in every way, look into it.
	display	i have the s22 ultra for about 9months (my first curved display, and first ever android) and i can say i just like flat displays more.
	chip	even if it's the snapdragon model you still wouldn't want to play on a s22 cuz it's simply not built for gaming due to its tiny size and **** chip

Word cloud based on keywords of positive and negative corpus
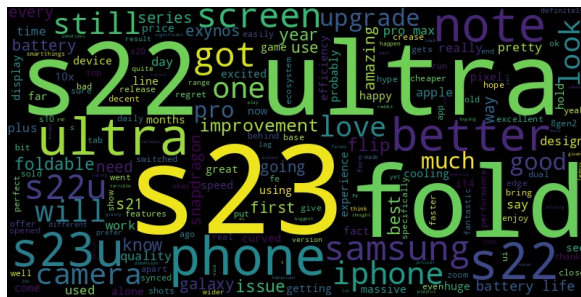
< Positive corpus >
<Negative corpus>
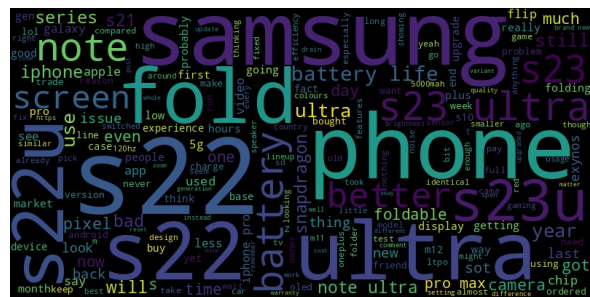
Insight

The majority of people are positive about Samsung's innovation that made foldable phones, and see Samsung as the number one player in the foldable industry. In terms of functionality, most people think it is inferior to Apple's

iPhone, and it is also compared to the pixel, which was previously extracted as a major keyword, but it seems to give a high value to Samsung's innovation.

Interestingly, snap-dragon came out as a positive keyword, but chip came out as a negative keyword. Since Snap-dragon is a mobile SoC adopted by Samsung as a chip for its phone, it is safe to say that chip and Snap-dragon are synonyms. In order to interpret these results, while researching data on Samsung's past chips, we found that Samsung used its own chip called 'exynos' and snapdragon chip depending on the region. However, the exynos series could not be installed on the s23 due to the heat issue and the performance degradation of exynos chip model of s22. As a result, we concluded that negative perceptions of Exynos and user needs for integrating Snapdragon models were reflected, resulting in a dual result for Samsung's chips.

B. 'Battery'

Positive : 61 Negative: 235

→ Positive rate : **20.61 %**
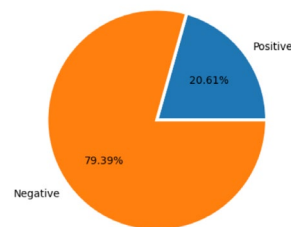
P/N	KEYWORD	Sentences with keywords
Positive	Battery life	The battery life absolutely does improve over time compared to day one.
		if he's coming from any other iphone, the battery on the s22u won't be disappointing to him at all.

Negative	Battery life	The battery life absolutely does improve over time compared to day one.
		if he's coming from any other iphone, the battery on the s22u won't be disappointing to him at all.
	heat	they overheat and run out of battery relatively fast due to the inefficient processor (snapdragon 8 gen 1).
	charging speed	they just need to increase charging speed & battery sizes.

< Positive corpus >

<Negative corpus>

Insight

Battery's positivity rate is very low compared to other attributes. People seem to think very negatively about batteries as a whole. Since 'battery life' is present in both positive and negative keywords, it was found that the most important attribute of a battery is battery capacity. Looking at the negative corpus, there was a negative perception of Samsung's battery capacity through comparison with the iPhone and Pixel, and criticism of the heat problem that has plagued Samsung for years. Also, there were complaints about the charging speed. Interestingly, there were quite a few voices arguing for the introduction of other technologies that could efficiently use the current battery capacity.

C. 'Camera'

Positive : 61 Negative: 235

→ Positive rate : **26.88 %**

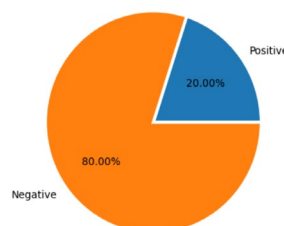| P/N	KEYWORD	Sentences with keywords
Positive	Zoom	my s23 non ultra can take a pretty good photo of the moon with the 3x camera, but the 10x looks really good.
		i love the mega zoom on the camera of the s22 ultra
Negative	Software(app)	i couldn't even record 4k for a few min without the camera app crashing.
		thats why a lot of android phone users tend to download gcam instead of their native camera app because google tend to do coding better
	Marketing	please note that it's a smartphone and not a video camera.. but clearly professionals expected to be able to use it as such, based on the marketing.
		if that's the case, lay off the marketing and expenditure on camera and focus on the general smartphone experience like battery life and usability.
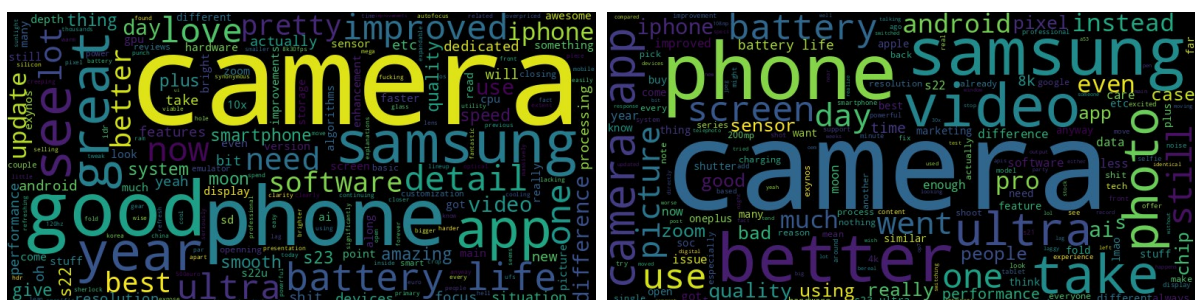
< Positive corpus >
<Negative corpus>

Insight

From the positive corpus of cameras, it was possible to achieve one keyword, 'ZOOM IN'. As much as users feel that the representative identity of Galaxy is the zoom-in function of the camera, users on Reddit shared reviews of shooting Moon using the 100x zoom function, or generally shared positive comments about the 100x zoom-in function. . In this galaxy s23 ad, there was a scene where the moon was clearly taken with a camera, but judging from the reactions of users on reddit, we thought this was Samsung's successful branding marketing.

Interestingly, the negative corpus took a critical stance against Samsung's marketing strategy. To sum up the content of the overall criticism, it is as follows: The commercials claim to be able to shoot professional video, but in reality it is not, and it makes no sense to expect such performance from a smartphone camera, so the focus should be on other smartphone experiences rather than camera performance. Also, the most cited negative attribute of Samsung's camera was Samsung's default camera app itself. Criticism about the instability of Samsung's default camera app and its usability has been overwhelming.

D. 'screen' (curved vs flat)

Positive : 40 Negative: 160

→ Positive rate : **20 %**

P/N	KEYWORD	Sentences with keywords
Positive	Large, big	fold has a much larger and more importantly wider outer screen.
		the z-fold 5 looks to be a real big improvement over the 4 based on the leaks, thinner, lighter, less of a crease, wider front screen, etc.
	Curved	the curved screen was a massive selling point for me when i switched to samsung in 2015
		there's nothing wrong with preferring a curved screen.
Negative	Crack	my friend also had problems with her fold 3, after 3 or so months the screen started to crack in the middle a
	Curved	I've used several curved screen phones and i definitely prefer a flat screen.
		make it easier for the screen to break as the curved glass is a bigger stress point if the phone is dropped.

< Positive corpus >   <Negative corpus>

Insight

There were large and curve as positive keywords for screen. Perhaps, compared to other companies in the smartphone market, from the time Samsung branded its
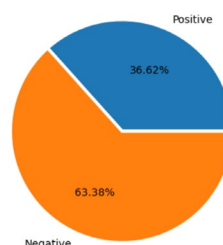
products through a uniquely large screen, users seem to have a general brand perception of Samsung's smartphones as wide, large, and thin. Most of the people prefer these properties, but there was also criticism that usability was poor due to large properties.

What is interesting is that the slightly rounded property of the side of the mobile phone that only Samsung has is called 'curved screen', and the keyword 'curved' was extracted from both positive and negative corpus, and we found that the likes and dislikes of the curved screen are very divided. In the case of the positive corpus, it was mainly about the attractive part of the design of the curved screen, and in the case of the negative corpus, the main story was that a flat screen was preferred over a curved screen in that the curved screen was vulnerable to cell phone damage. In order to find out the exact preference ratio for curved screens, we separately extracted keywords for curved screens and conducted sentiment analysis on them.

As a result of sentiment analysis on the curved screen, the (positive) rate was 30%, and below is the word cloud about positive, negative corpus of 'curved'.

Judging from these results, we could know that there are more users who think negatively about 'curved screen'. In addition, there was 'crack' as a negative keyword, and the part of the sentence containing the keyword was related to the screen damage problem of Fold, a foldable phone developed by Samsung. From these results, we were able to gain an insight that Samsung's foldable model is not yet functionally complete. In addition, it was easy to find articles expressing dissatisfaction with Samsung's service center's response to cell phone damage.

13

E. 'UI'

Positive : 26 Negative: 45

→ Positive rate : **36.62 %**

P/N	KEYWORD	Sentences with keywords
Positive	oneUI, software	personally, what i wanted was the stability and ecosystem of an iphone with the software ethos of android.
		samsung has tremendously improved on software updates over the last few years.
		oneui is its an os built on top of andriod. it has its own app store, features and ecosystem.
Negative	Compared to iphone, apple	apple is currently only ahead in terms of software stabilization, which agreed, is cinematically smooth. oneui looks like a toy os with all the weird icons.
	Ecosystem	Samsung makes great hardware, but does anyone actually use Samsung's app store or Samsung's assistant? It is kind of hard to be an ecosystem when so few people use those
		As someone who has owned an entirely all-Samsung house (phone, laptop, tablet, TVs, etc...) and someone who currently has mostly Apple products in now, there is no comparison. Apple's ecosystem is much better. It is consistent, always works, and works across generations of product.

Insight

 Overall, people who use existing Galaxy models seem to be satisfied with the usability of oneUI, although there is a difference between liking and disliking oneUI, which is adopted by Samsung Galaxy, depending on personal preference. In the negative corpus, there were many articles criticizing Samsung's UI through comparison with Apple and iPhone, and criticism for not having an ecosystem as good as Apple's was the main focus.

5. Conclusion

5-1 Result of Analysis

	Galaxy	Battery	Camera	Screen	UI
Positive rate	25.75	20.61	26.88	20	36.62

Based on the positive ratio of the Galaxy series, which falls in the middle among the five keywords, it appears that the most urgent issues Samsung needs to address right now revolve around the **battery** and **screen**.

Based on previous analysis, the direction Samsung should pursue is as follows:

1) Select better chips to reduce battery heating

2) Abandon curved screens

3) Focus on improving camera UI rather than solely enhancing camera performance

4) Invest more in the Fold and Flip designs to establish an innovative brand image and become a leading player in that sector

5-2 Potential for Development

Our project holds potential for utilization by Samsung and other companies in the following ways:

1) Companies can develop internal content moderation AI programs that allow them to assess their reputation more objectively.

2) Business consultants can utilize our project to identify vulnerabilities within a company and gauge public sentiment towards the company.