# 

Professor Changseung Yoo

Chris Zong 260872947

Ruite Xu 260621929

Junfeng Yang 260696954

Zhentian Jiang 260658190

Junkang Zhang 260867086

Yuan Li 260866398

**A case study of text analytics on user comments in YouTube**

[Introduction 2](#_Toc1463851)

[Market & Competition 2](#_Toc1463852)

[Methodology 4](#_Toc1463853)

[Data Collection 4](#_Toc1463854)

[*Identify the Top 3 Brands* 5](#_Toc1463855)

[Filter Data & Calculate Compound Sentiment Score 6](#_Toc1463856)

[Text Analysis with Descriptive Features 7](#_Toc1463857)

[Identify Frequently Mentioned Models & Calculate Lift Scores 8](#_Toc1463858)

[Limitations 12](#_Toc1463859)

[Data Collection 12](#_Toc1463860)

[Extraction of Models within brands 13](#_Toc1463861)

[Descriptive feature selection 13](#_Toc1463862)

[Sentiment analysis 13](#_Toc1463863)

[Business Insights 13](#_Toc1463864)

[Reference 15](#_Toc1463865)

# Introduction

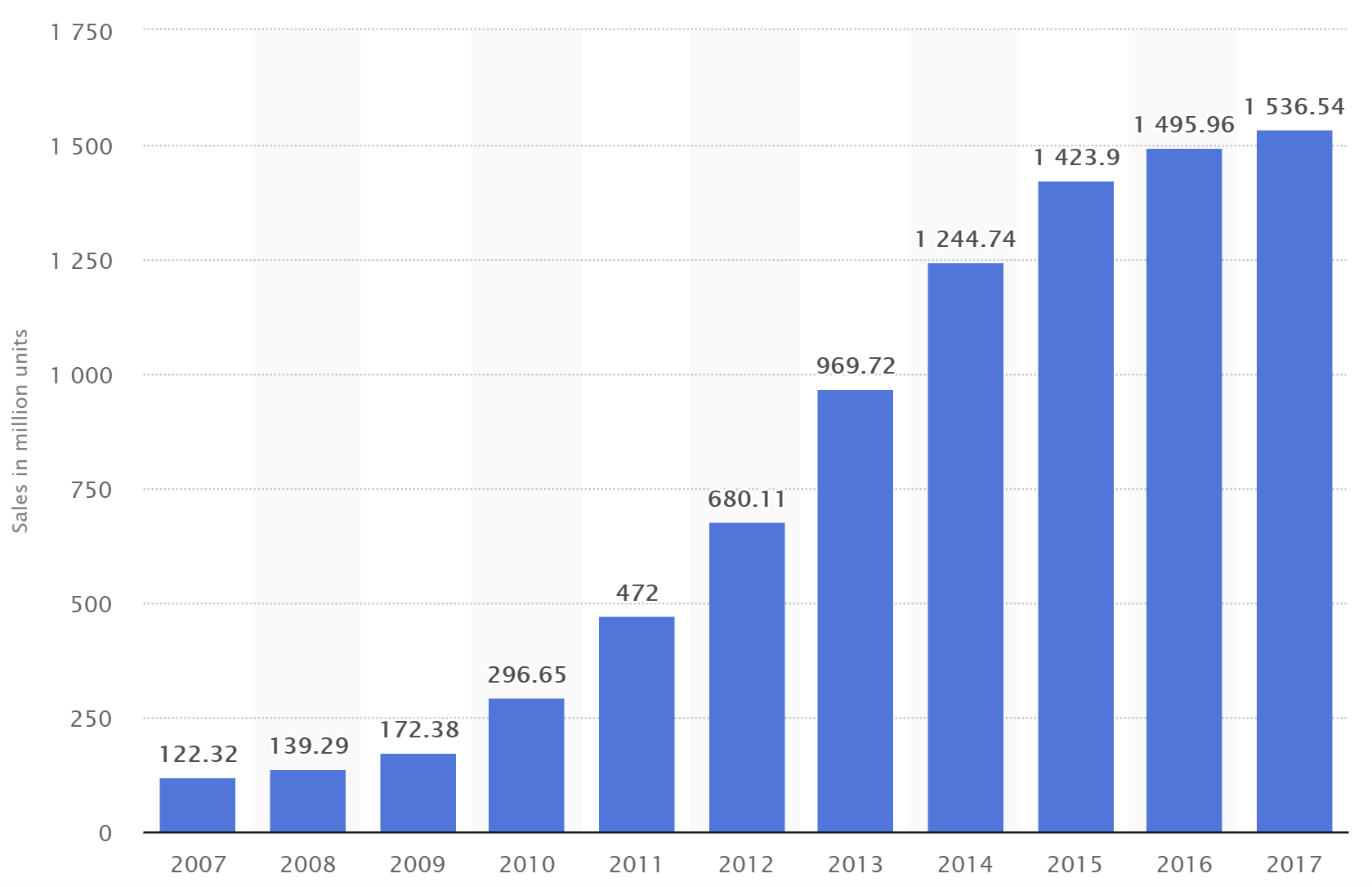
Social media provides an unparalleled platform for consumers to share their product experiences and opinions thanks to the advancement of internet technology, and therefore those online massive consumer reviews are conceded as having unprecedented impact on firm’s marketing strategy. Meanwhile, the text analytics method is believed to have the power to interpret how customers feel and provide better insights to the marketers. In this report, we aimed to study which attributes of a smartphone are most intensely discussed among the customers, and more specifically, how does those attributes affect the sales volume of a particular model of a brand. As the result, we choose Samsung, Apple and Huawei as the three most popular brands based on their frequency of mentions in our YouTube comment data1 , which will be further explained in detail in the methodology part, and studied which attributes associated with different models of those brands can affect the sales figures by performing text analytics on the comment data.

In the following parts, we will discuss the methodology that we implemented in this study, illustrate the important business insights that we gained from the analysis, provide some suggestions to those three brands but not limited to those three based on our results, and list down the limitations that are incurred in this project.

# Market & Competition

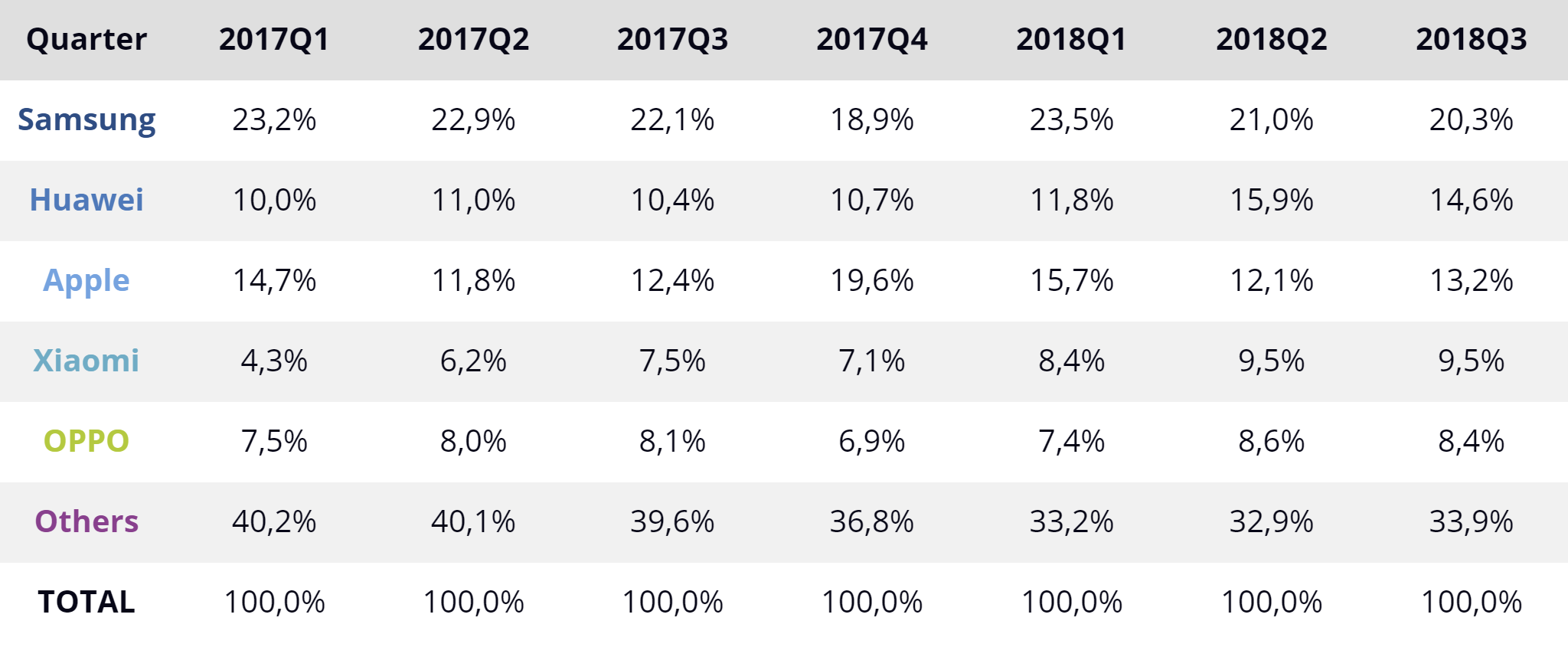
The smartphone industry has been grown exponentially during 2009 to 2014, just after the economic crisis2 (see graph below). The market size has not realized significant growth recently, and this refers to the saturation and stagnation of the market, which companies and brands have to compete more heavily on existing customers instead of new customers.

***(Figure 1)***



To explore smartphone brands in detail, the chart reveals each major brand’s recent ranking, in terms of global market shares (see graph below). The competitive landscape of smartphone has been intense and, surprisingly, Apple has lost the second place, which it has held for long time, to Huawei. The rivalries of Xiaomi, OPPO, and others, such as Vivo, have increasingly become the threat going into the future, as they have grown market shares and taken away market shares from Samsung and Apple.

***(Figure 2)***



# Methodology

## Data Collection

The data-collecting journey started from searching for a suitable platform to extract the valuable information we needed, in this case, any comments regarding the performance, experience, designs and other useful attributes of a smart-phone are critical and meaningful to us with the forthcoming analysis. After careful researches being done, we eventually targeted the comments within the relevant YouTube smartphone videos as the data source because the large scale, high relevance and suitable time horizon.

The chart below briefly demonstrates the characteristics of the data collected and its applications in the later stages.

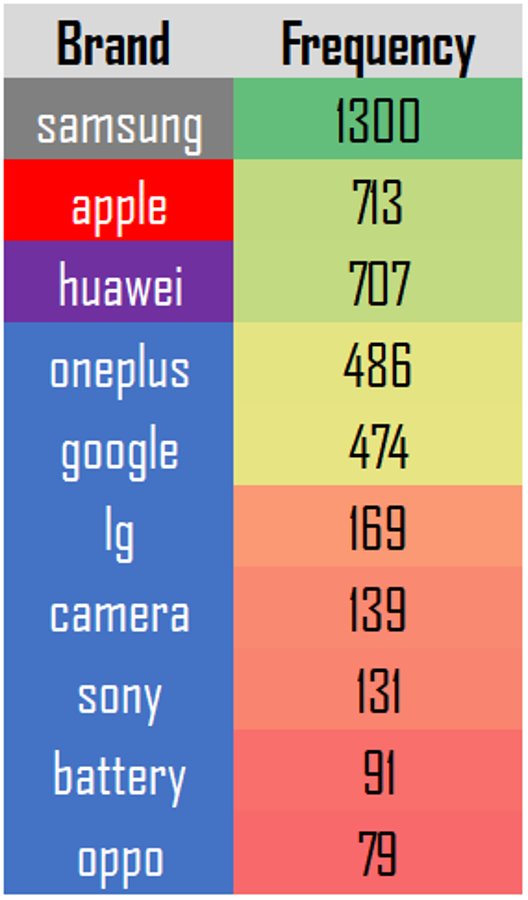
***(Figure 3)***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Source | Size | Description | Usage |
| Top 5 BEST Smartphones of 2018 | YouTube | 3200 Comments | It’s a video describing the top 5 brands in the year of 2018 by an authoritative Youtuber | Identify the top 3 brands from the comments by considering each brand’s occurrences |
| Smartphone Awards 2018! | YouTube | 24413 Comments | It’s a video describing the top awarded brands in the year of 2018 by an authoritative Youtuber | Identify top 5 models within the top 3 brands, then perform sentiment analysis and lift score analysis |
| Model list | CSV File | 17 brands;  335 models | A list of brands with specific models under each brand | Assist to extract models and brands mentioned in the comments |

## *Identify the Top 3 Brands*

Due to the topic of the first YouTube video, most of comments below are follower’s personal top five smartphone of 2018 list and without personal sentimental expression. Thus, we used these 3200 comments to find the top 10 brands from frequency counts distribution. Our results are as follows:

***(Figure 4)***



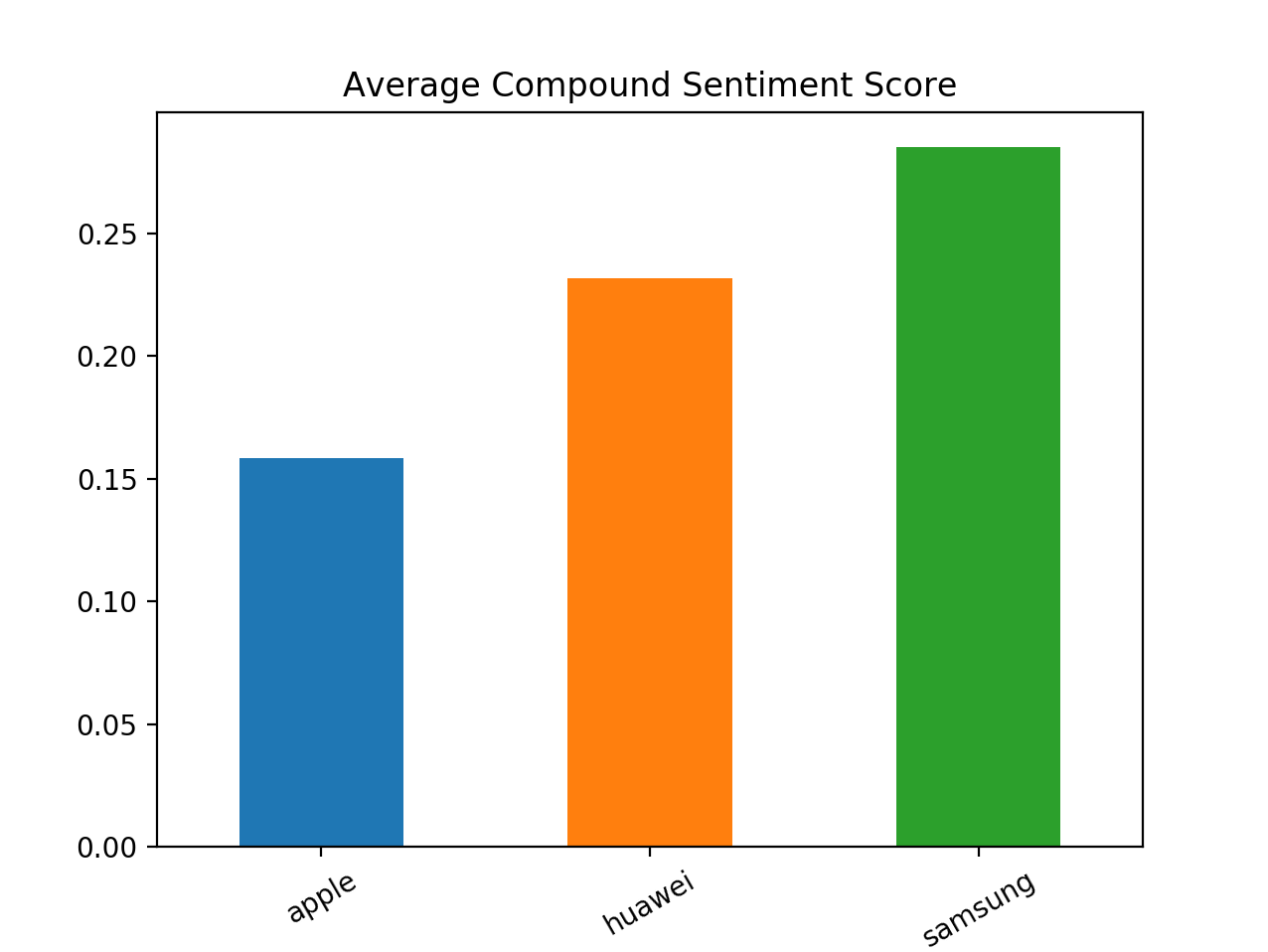
## Filter Data & Calculate Compound Sentiment Score

The analysis continues after identifying the top brands which are Huawei, Apple and Samsung as most of us expected to be. In order to perform sentiment analysis, we cannot use the 3200 comments as they don’t have many descriptive words, instead we scraped comments from another YouTube video called: ‘Smartphone Awards 2018!’, under this video there are over 24000 comments and each of them expressed followers’ (customers’) point of view on 2018’s smartphones.

One important action employed in this stage was we only kept comments with exactly only one brand mentioned in a comment. The reason for it was because the technique applied to calculate the sentiment score of a comment takes the whole piece of it and process. Therefore, in the case where multiple brands appear in a comment would provide no choice but to assume the co-occurring brands share the same sentiment score which is not the situation in most of the cases and would bring inaccurate results in the later analyses. Hence, we only kept single-brand comments and filtered the ones with Huawei, Apple or Samsung existed.

***(Figure 5) (Figure 6)***

|  |  |
| --- | --- |
| Brands | Comments Picked |
| Apple | 1078 |
| Huawei | 898 |
| Samsung | 1378 |



As it is shown above, the graph demonstrates the average compound sentiment score of the “Big 3” brands where Samsung achieves the most positive feedback and Apple with the lowest score.

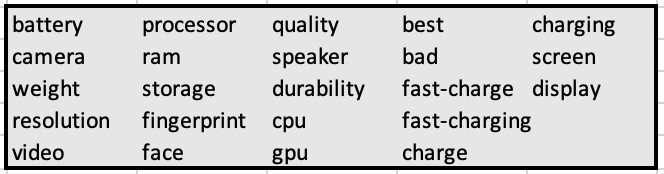
***Note****: The range of compound sentiment score is from -1 to 1 where -1 is the most negative and 1 being the most positive. Compound sentiment score is a score computed from positive, negative and neutral sentiment score with weights correspondingly.*

## Text Analysis with Descriptive Features

To further analyze the features and characteristics associated with each brand and each model. We explored into the text by introducing descriptive features and convert them into dummy variables *(1 if the word appears in the comment, 0 otherwise)* which later provide us the option to calculate the lift scores between them and the models.

The list of features introduced is as following.

***(Figure 7)***

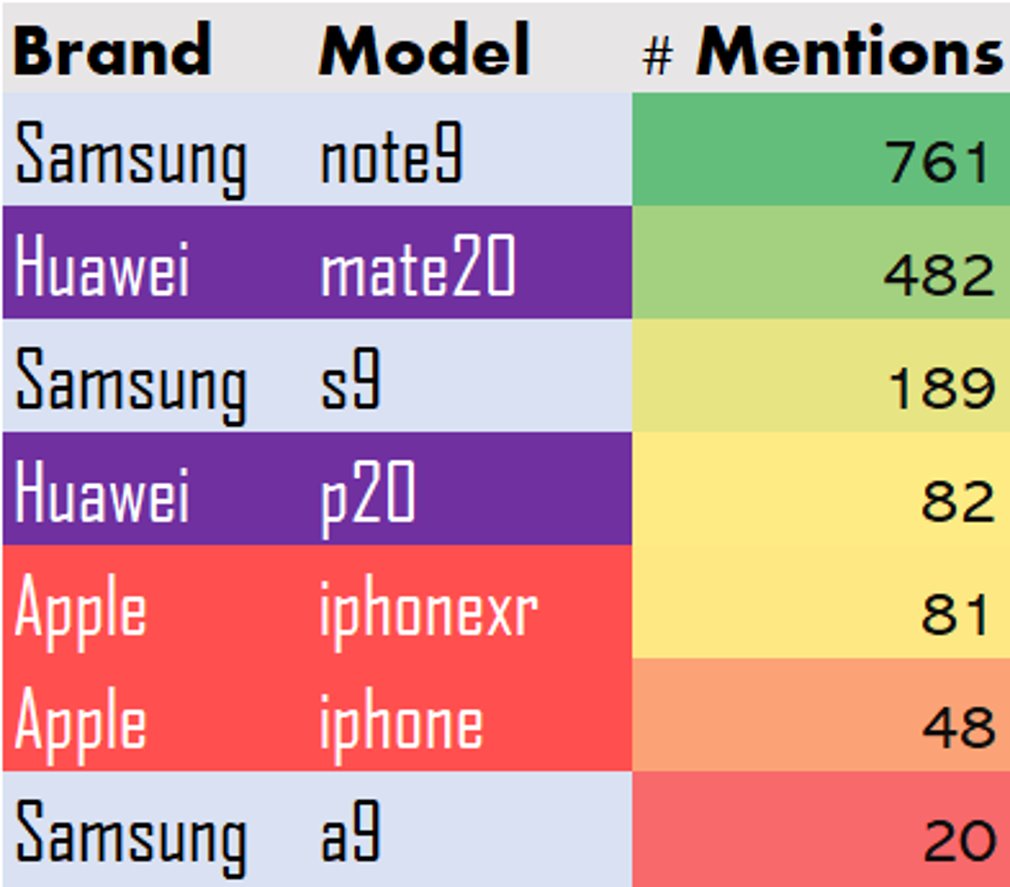


## Identify Frequently Mentioned Models & Calculate Lift Scores

After we chose the top three brands: Samsung, Apple and Huawei, we decided to find top models within these brands to perform further analysis includes sentiment analysis and lift score calculation.

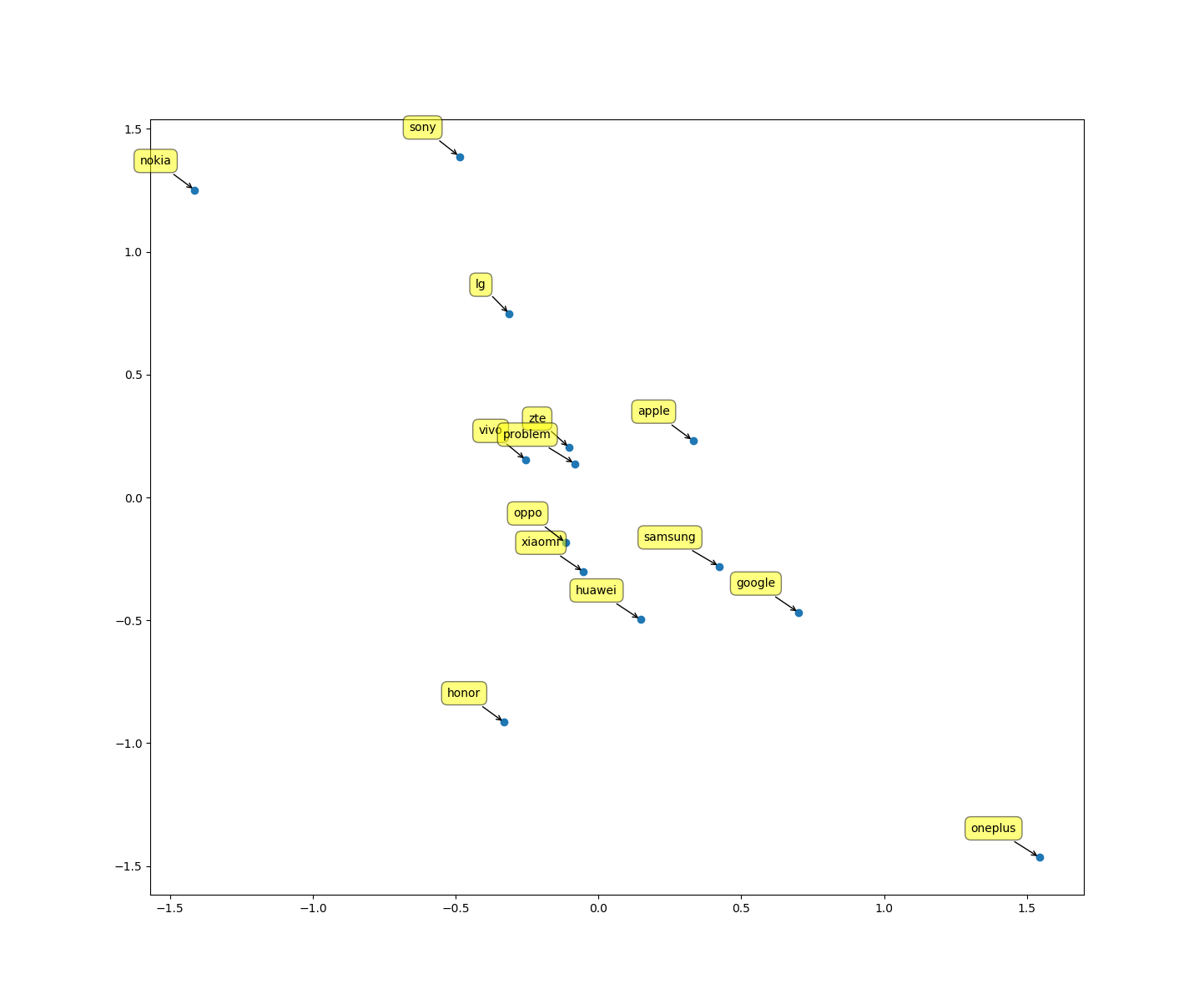
As shown in table below (see Figure 8), Samsung Note9 has received most mentions, 761, in the comments under the smartphone review video, followed by Huawei’s Mate20, and Note9 has been released since the mid-year. The video is uploaded at the end of the year, and therefore comment is, to some degree, fair to reflect all brand models released in 2018. The higher mentions may not mean the best of all, but at least that shows the popularity and the awareness across all the consumers and potential buyers, who are looking for and considering purchasing at that moment. Surprisingly, Apple’s models have been dropped to 4th and 5th place, significantly lower than Samsung’s and Huawei’s models. More importantly, the flagship model, iPhone XR is even lower than Huawei’s P20 Pro, which is not even a flagship model from Huawei and released during early year, and it is rational, since Apple has been consistently dropping its market shares according to the market share chart3 (see Figure 2). It is very important to realize each model’s strengths and weaknesses not only for Apple but also for all other brands.

***(Figure 8)***



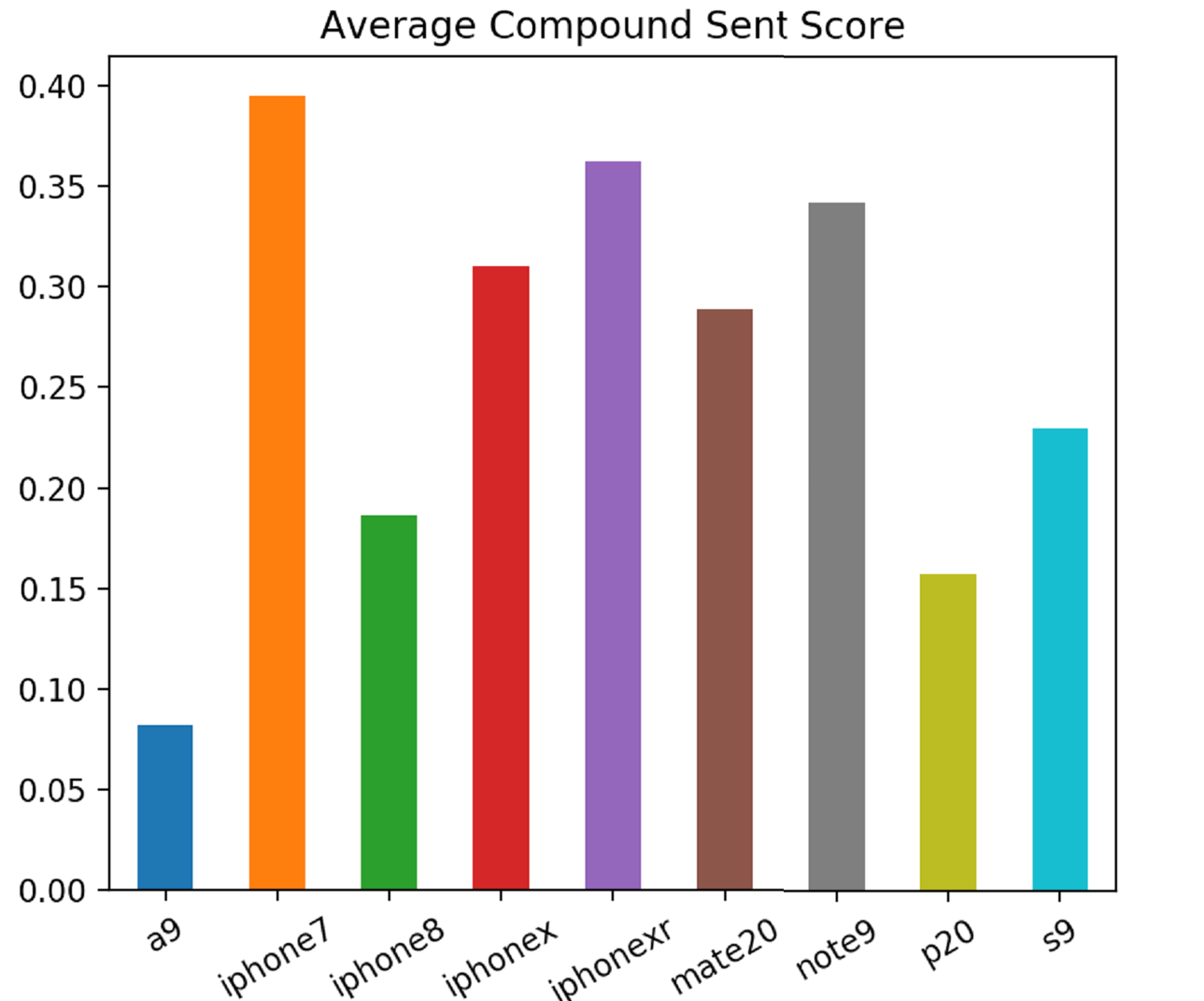
Moreover, from analyzing the more brands and locate them on the multidimensional diagram (MDS), and we have found different distances among brands and problem. There are three clusters, one at the left top corner, one at the center, and one at right bottom corner. Major brands are on at center. Sony and Nokia are too far from the center, which may mean they are not directly competing with the major flagship brands, and therefore they are not in the consideration for comparison. Despite OnePlus is recommended in the video, it is not within the center cluster, which means less similarity for comparison. Vivo, OPPO, Zte and Xiaomi are closer to “problem” compared to other major brands, since they are mainly targeting on the compact smartphone market, which means cheaper and best-value performance. Hence, they may lack in after-sales services and quality. Apple is closer to “problem” than Samsung, Huawei, Google, and Honor, and this may mean Apple associate with problems, which can be technical, quality, and other issues, more often than others. This may infer the reason behind the less popularity and declining market shares.

***(Figure 9)***



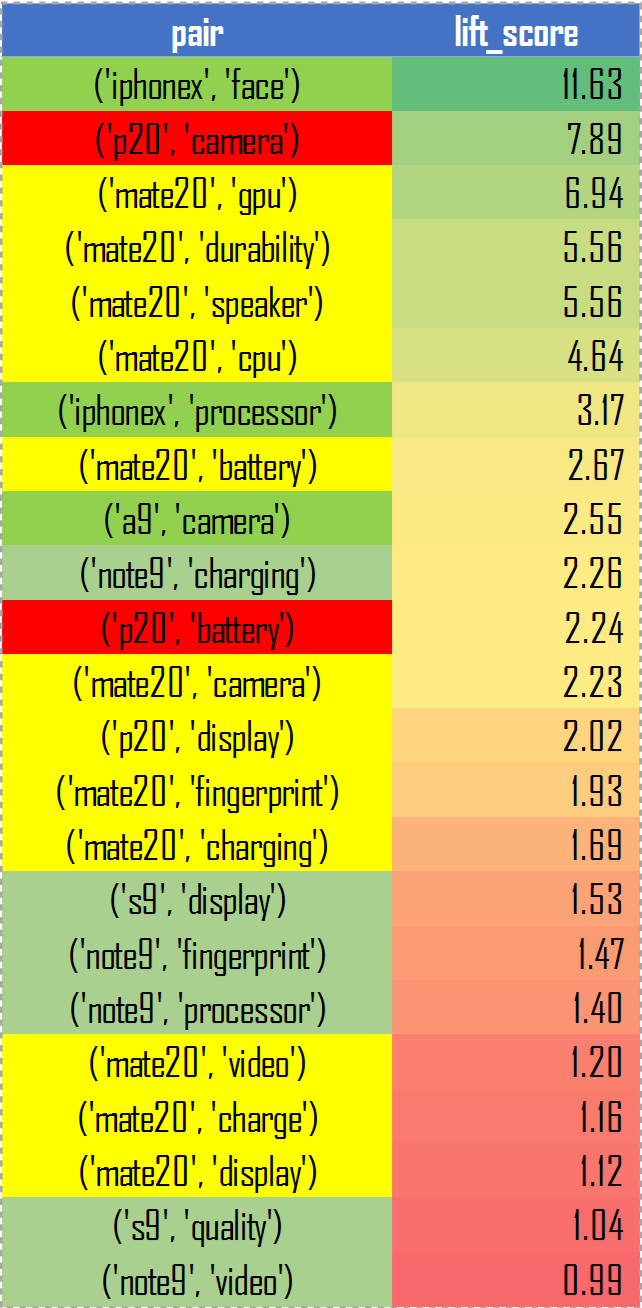
With the top models selected, we have also plotted “Average Compound Sentiment Score for Top models”. We have found that, in general, Apple’s models have received better sentiment scores on average, meaning positive comments and impression over Apple’s products, as shown in the bar chart (see Figure 10). iPhone 7 and iPhone XR are listed as 1st and 2nd, in terms of sentiment scores. iPhone 7 is the first phone to cancel the physical home button and iPhone XR is a cheaper version but better value, which is approximately 75% of the flagship model iPhone XS’s price but includes most features.

***(Figure 10)***



To further investigate features associated with each model, we have calculated the lift scores between the top models and the descriptive features introduced from the previous step (see Figure11). We have only kept pairs of models and their associated features with lift scores above 0.99, which we feel some level of significance to investigate and analyze models. We have found that iPhone X is good at face recognition and processor. iPhone X, released in 2017, is a special case in this video and comments, since it has been the first to implement to face recognition on a smartphone and it is expected to have strong association with this feature. P20 Pro and Mate20 Pro represent Huawei’s technology and brand’s perception. Huawei has strengths on camera, speaker, battery, durability, fast-charging, and display. Furthermore, Note9 and S9 are representations of Samsung galaxy brand. They are good at features of fast-charging, fingerprint, processor, video, display, and overall quality. In addition, the instructor has called note9 the model of the year in the video. On the other hand, some of the feature associations may refer to negative correlation, and we have researched on external sources and found that Huawei’s Mate20 Pro has weaknesses on processor and video, and P20 Pro is not very impressive on battery and display4.

***(Figure 11)***



# Limitations

## Data Collection

Social media platforms such as Instagram and Twitter don’t have enough number of comments for a post of certain topic like ‘top 5 smartphones in 2018’, most of posts on Instagram and Twitter mentioned more about newest smartphones or upcoming ones. There isn’t any big forum for smartphones, not like forum of cars (Edmunds), so it limited our choice to YouTube video comments.

## Extraction of Models within brands

Different Spelling styles of reviewers, Ex. “Samsung S9”, “S9”, “GalaxyS9” or “S 9”. Hard to cover all possibility in our model collection phase. Also, there are many models in different brands but with same names. We cannot tell a model's pros and cons by calculating only lift score with some attributes, we need do more research to judge. For example, GPU is Mate20's weak point, but we cannot get to know that only by seeing 6.94 lift score of ‘Mate20’ and ‘GPU’.

## Descriptive feature selection

We could only manually think about attributes such as battery, screen, fast-charging, etc. But it’s hard for us to cover every possible attribute, for example, we forgot considering gaming.

## Sentiment analysis

Usually a comment will mention more than one model or brand, so getting compound sentiment score cannot show anything, for example, when analyzing a sentence: ‘I think iPhone8 sucks while Mate20 is awesome.’, a compound sentiment score cannot give us a correct view of people’s sentiment on this two models (brands), since they are in one sentence. We thought one solution, which is select only comments that mentioned only one brand, but due to limited number of samples, there are only 3000 comments satisfied this condition, so the result may not be too reliable.

# Business Insights

The rivalries amongst Samsung, Huawei, and iPhone should help each other know their brand’s and product’s perception better from consumer’s view point, which can be drawn from comments and ratings. There are also learning opportunities to improve on their weaknesses to match opponent’s features, and they can also extend their strengths to more specifically target consumers who are seeking these features.

Huawei can spend some budget on improving on processor and display, and, at the same time, extend the leading on photography side, camera, speaker, and durability. Meanwhile, Samsung has overall best models, but it can focus on display and fingerprint, which is more on the security side. On the other side, iPhone can put more effort on cutting costs and model prices as well as applying more AI on screen interface, such as face recognition and other fun features.

With all these improvement suggestions, they can increase their associated positive lift scores and reduce negative ones. The sentiment scores will also become more equal according to comments made by users. They will also add and change perceptions from consumer view point and, hopefully, capture more market shares and revenues.

# Reference

**1. *List of best-selling mobile phones****.* (2019, January 23). Retrieved from <https://en.wikipedia.org/wiki/List_of_best-selling_mobile_phones#2017>

**2.** ***Cell phone sales worldwide 2007-2017****. (2018, February).* Retrieved from[*https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/*](https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/)

**3. *IDC - Smartphone Market Share - Vendor.*** (n.d.). Retrieved from <https://www.idc.com/promo/smartphone-market-share/vendor>

**4.** ***Huawei HiSilicon Kirin 980 more than a year behind Apple's A12 Bionic in performance.*** (2018, October 18). Retrieved from <https://www.google.ca/amp/s/appleinsider.com/articles/18/10/18/huawei-hisilicon-kirin-980-more-than-a-year-behind-apples-a12-bionic-in-performance/amp/>