



95851 DSPM - Final Project Report

Apple Product Management Team

Social Media Listening Evaluation Report

Team Members: Ashlyn Im, Linda Hung, Shih-Ting Sun, Wan-Ju Yu

andrewID: manyii, tzutingh, shihtins, wyu2

Motivation: Business Problem

The number of internet users has exploded tremendously in the last decade as a result of technological advancements. The amount of data generated by people's engagement on the internet has increased dramatically. Amidst the noise is a wealth of knowledge regarding customers' opinions and feedback that we can tap into. Apple's core values have always been prioritizing people over products. We're eager to hear from users through social media listening, a new way to analyze and collect customers' thoughts via natural language processing. Particularly, we are interested in answering these questions:

- How do the customers feel generally towards the products?
- What attributes drive customers' interest in purchasing the product?
- Could we predict whether a person adopts/uptakes the product based on the post?

To conduct this, we will mainly be looking at the social media data in 2017, when our major competitor, Samsung, launched their Samsung Galaxy S8 flagship product. Later in the same year, we also launched iPhone 8 and iPhone X. The time frame provides us with a fair foundation for comparison. Regarding the sources, we have collected data from the following sources: Instagram, Tumblr, Facebook, Twitter, blogs, forums, comments, consumer reviews, and other social media.

Project Overview

The whole project pipeline will be divided into 8 steps:

Step 0: Load the package

Step 1: Load the File

Step 2: Data Preprocessing

- 2.1 Remove Irrelevant Rows
- 2.2 Summary Statistics
- 2.3 Exploratory Data Analysis
- 2.4 Feature Engineering: Category and Before/After Launch
- 2.5 Handle Null Values
- 2.6 Preprocessing Text Data

Step 3: Demographic Analysis

- 3.1 Overview of Demographics
- 3.2 Compare US vs. Non-US

Step 4: Sentiment Analysis

- 4.1 Overall Sentiment
- 4.2 Sentiment Before VS After Launch

Step 5: Attribute Analysis

- 5.1 Topic Modeling



- 5.2 POS Analysis
- 5.3 Hashtag Frequency
- 5.4 Sentiment Analysis for Product Attributes

Step 6: Predictive Analysis

- 6.1 Feature Engineering
- 6.2 Prediction of Product Uptake
- 6.3 Model Fitting

Step 7: Comparison between Twitter and non-Twitter Dataset

- 7.1 Exploratory Data Analysis of Twitter and Non-Twitter Dataset
- 7.2 Word Frequency for Twitter and Non-Twitter Dataset
- 7.3 Hashtag for Twitter and Non-Twitter Dataset
- 7.4 Sentiment Analysis for Product Attributes for Twitter and Non-Twitter Dataset

Step 8: Recommendation

- 8.0 Preparation: Attribute Enhancement
- 8.1 Introduction: Informative Marketing
- 8.2 Growth: Persuasive Marketing: Persuasive Marketing
- 8.3 Maturity: Product Consolidation
- 8.4 Decline: Pivoting

Tasks

Extract information for the following products: Galaxy S8, iPhone X, iPhone 8.

Question 1. What were the most important attributes for each product?

Answered in step 5.

Question 2. What attributes were liked the most and what attributes weren't?

Answered in step 5.

Question 3. In particular, how did consumers feel about the Quality, Price and “Value” of each product?

Answered in step 5.

Question 4. What was the sentiment for these products? Did the sentiment change BEFORE and AFTER the launch of the products? How?

Answered in step 4.

Question 5. Extra Credit: Can you predict the uptake or adoption of the product by looking solely at the pre-release data? Can you “validate” that by looking at the post-release data?

Answered in step 6.

Question 6. Extra Credit: Process the Twitter and non-Twitter datasets separately for some of the tasks. Can you draw any conclusions on the ease or accuracy of processing and analysis for these two sets? Was one set of sources easier, more accurate than the other?

Answered in step 7.



Step 0: Load the Package

(Please refer to the code file.)

Step 1: Load the File

(Please refer to the code file.)

Step 2: Data Preprocessing

(Please refer to the code file.)

Preview the data columns and sample rows for both tweets and non-tweets.

2.1 Remove Irrelevant Rows

Before any preprocessing, we know that some of the rows are irrelevant. To avoid misinformation, we will remove them first.

2.1.1 Excel Sources

Eliminate the details, sources on the dates, etc. at the bottom of each excel file.

Process only the original posts rather than replies and comments (Column: Post Type).

2.1.2 Professional and Celebrities Endorsement

Eliminate posts from authors that have a large number of followers as they will likely be professionals or celebrities.

According to the research from the Advertising Standards Authority in the UK, which could be applied to the US context since social media are boundaryless, anyone with over 30,000 social media followers as a "celebrity". Therefore, we will use 30,000 as a threshold. Source:

<https://www.thedrum.com/news/2019/07/04/asa-defines-anyone-with-over-30000-social-media-followers-celebrity>

2.1.3 Duplicated Rows

Lastly, we will drop the duplicated rows as they are likely to be advertisements.

2. Summary Statistics

We would like to take a look at the high-level statistics first regarding data shape, column statistics and missing values.

Upon removal of the unnecessary information above, we have in total 330447 rows for the remaining analysis.

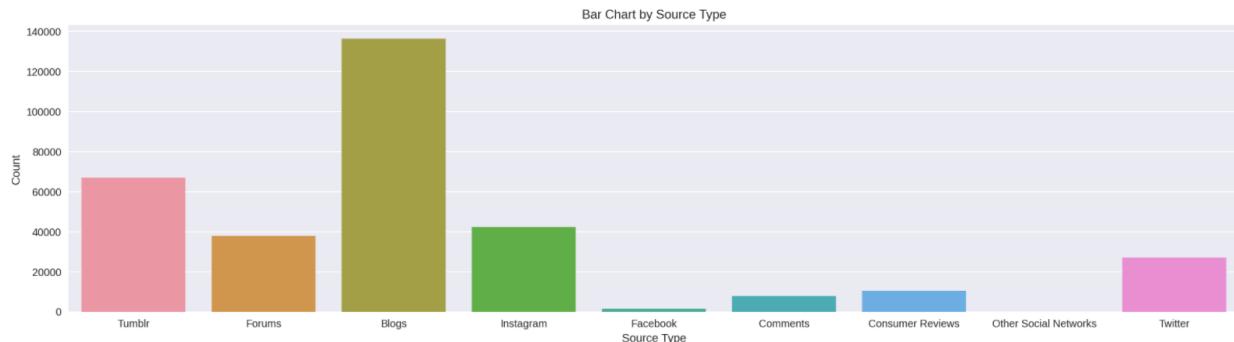
The summary statistics(see in the code file) show that the data frame has some large missing values in certain columns. Fortunately, our focus of the analysis "Sound Bite Text" does not have any null values.



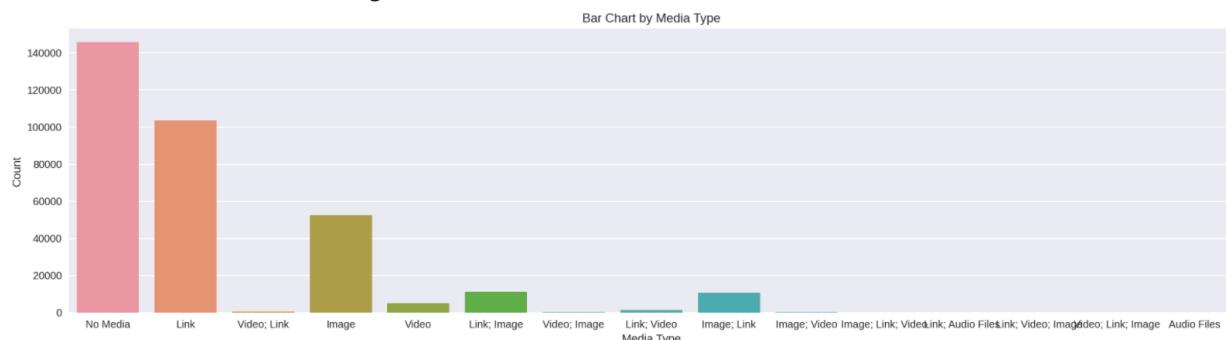
2.3 Exploratory Data Analysis

Visualizing the data could give us some interesting insights into the overall user portfolio of our data. In the following section, we will visualize categorical variables in the form of bar charts and numerical variables in the form of histograms.

We have eight kinds of social media sources, and their counts vary:



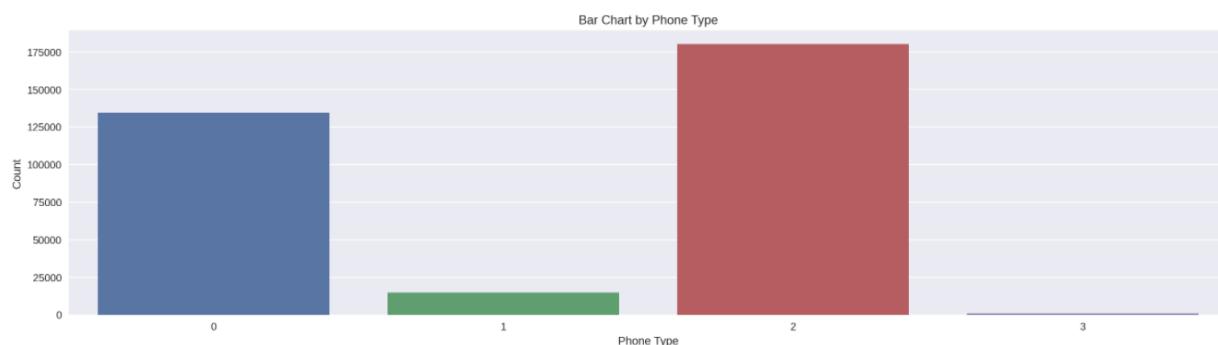
There are several combinations of media types mentioned in the Sound Bite Text, but the most common ones are Link and Images.



2.4 Feature Engineering: Category and Before/After Launch

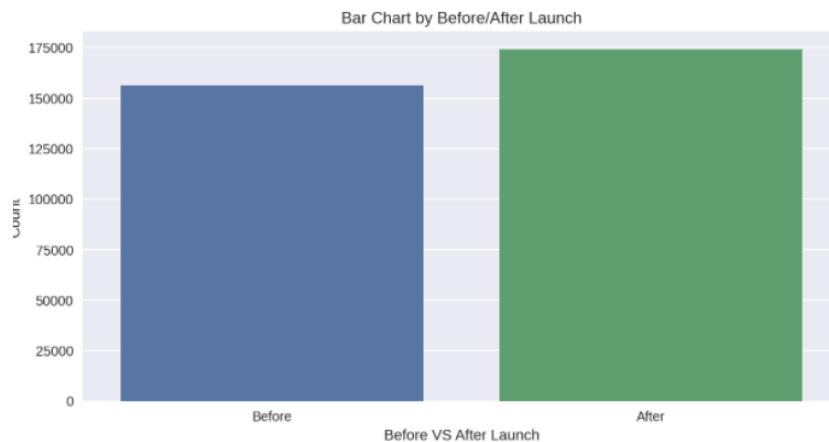
After performing some exploratory data analysis, we will proceed on to feature engineering. For this part, we will extract the Category (iPhone or Samsung or both) from "Sound Bite Text" and Before/After Launch from the "Published Date" for convenient identification.

In order to know how iPhone and Samsung cell phones have been mentioned in posts, we analyze different counts of Phone Types. It shows that we have plenty of data mentioning Samsung compared to the iPhone. (0 represents None, 1 represents iPhone, 2 represents Samsung Galaxy 8, and 2 represents both)





We also like to find out the number of differences before and after launch. The plot below shows that the amount of data in these two groups is similar.



2.5 Handle Null Values

The main focus is "Sound Bite Text" for natural language processing. But as mentioned, there is no null value in that column, so we do not have to further process this.

2.6 Preprocessing Text Data

In this section, we cleaned the original text data by removing the stop words and retaining words with characters between 3 and 13 with the help of spaCy library and TweetTokenizer. It will return the string representation of clean text for every observation. At this stage, we do not tokenize the text yet as we want to preserve a general format for attribute analysis and sentiment analysis.

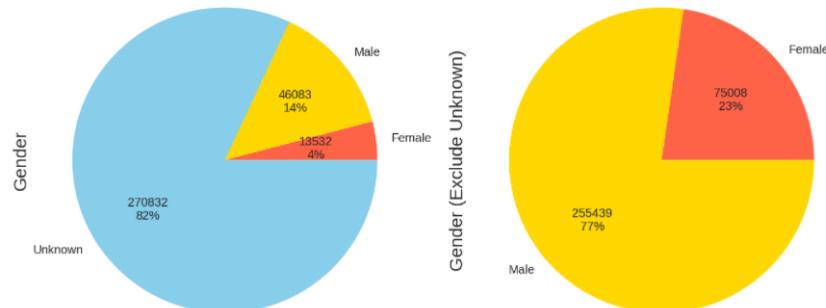
For sentiment analysis, we will be parsing the entire string representation into the model to preserve the contextual meaning of the text data.

For attribute analysis, we will only be tokenizing the sentence there against a specific and personalized set of stopwords.

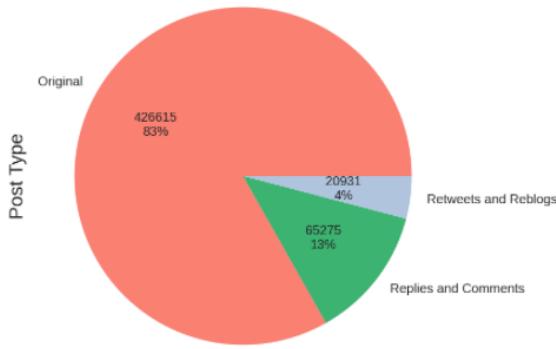
Step 3: Demographic Analysis

3.1 Overview of Demographics

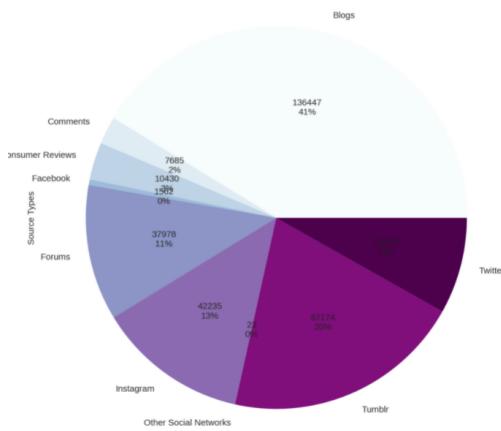
To explore the dataset and analyze demographic features, we first overview the data.



After excluding the unknowns in the "gender" column, we can find out that the majority of text data is composed by male users.



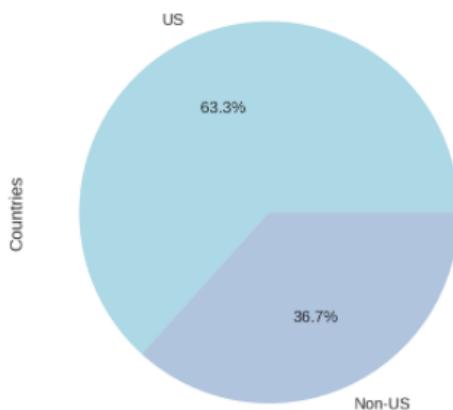
As for the post type, most of the posts are original and only a small portion of them are comments, replies, or retweets.



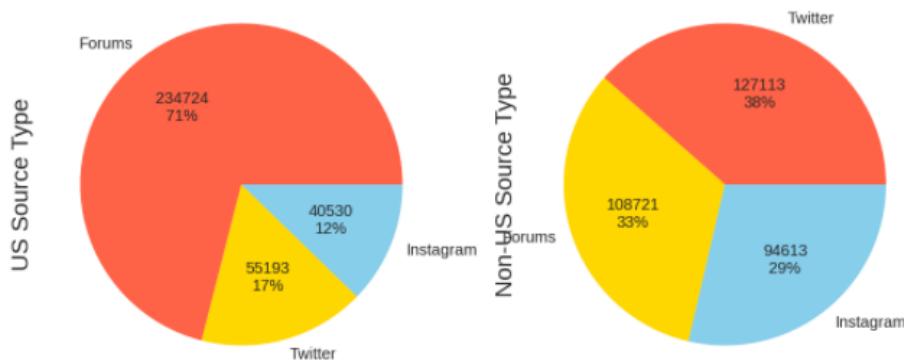
We can find out that the majority of source types are Blogs, Forum, and Tumblr, which is surprising to us, because we expect that Instagram, Twitter, or Facebook will have a large amount of discussion about Apple, Samsung, and their products as well.

3.2 Compare US v.s Non-US

We are specifically curious about whether there are any differences between the US market and Non-US markets. Therefore, we conduct a different analysis comparing these two groups.

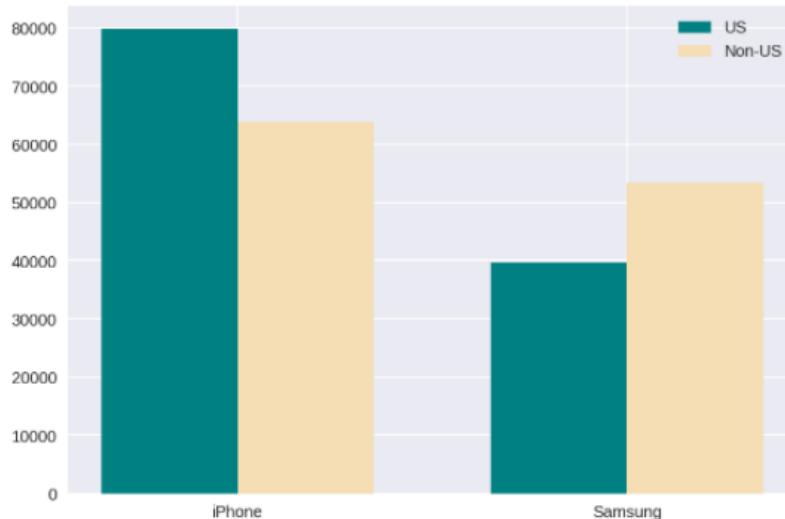


For the dataset we have, around 63.3% are posts from the US, the rest are 36.7%.



It seems that most of the posts on the abovementioned platforms like Facebook, Tumblr are from users who are not labeled where they are located, so we will be focusing on these three platforms. We can find out that there are more non-US posts from Twitter and Instagram compared to the US posts.

To further investigate the volume of discussion in the US and Non-US regions, we calculated the term frequency of iPhone keywords and Samsung keywords mentioned in these three platforms.



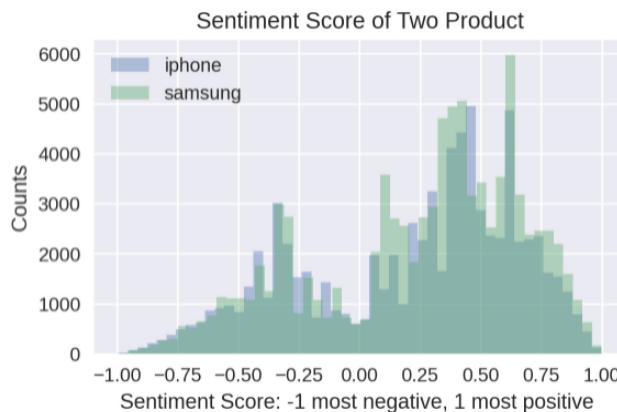
We discovered that people in the US discuss iPhone more often than Samsung, however, in Non-US countries, Samsung has a higher discussion volume than iPhone in 2017. Since higher social media volume is associated with brand recognition, we would suggest that Apple's marketing strategies towards the overseas market can use social media listening to keep up with the trending topics and create opportunities for their customers to share their thoughts about the product on social media.



Step 4: Sentiment Analysis

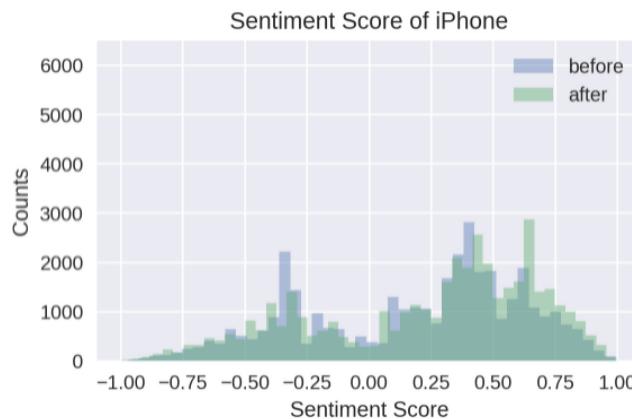
In this section, we will be conducting sentiment analysis on the text data. We aim to find out the respective sentiments for iPhone and Samsung before and after the launch, as well as the overall sentiments. To do so, we utilize a Sentiment Intensity Analyzer and create a function to get the compound polarity scores for the sentences in Soundbite Text.

4.1 Overall Sentiment

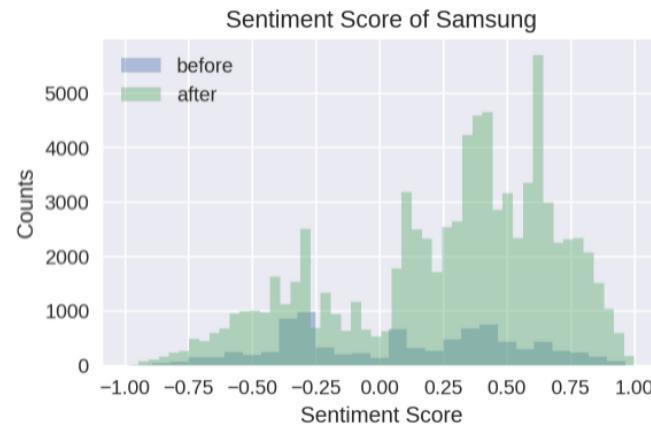


The above plot is the sentiment score for the whole period of time for the two products. Since there are many neutral texts, we would hide the neutral counts to focus on the comparison between positive and negative text for these two products. In general, comments are mostly neutral to positive. Besides, Samsung has a slightly higher positive sentiment score from the Soundbite Text.

4.2 Sentiment Before VS After Launch



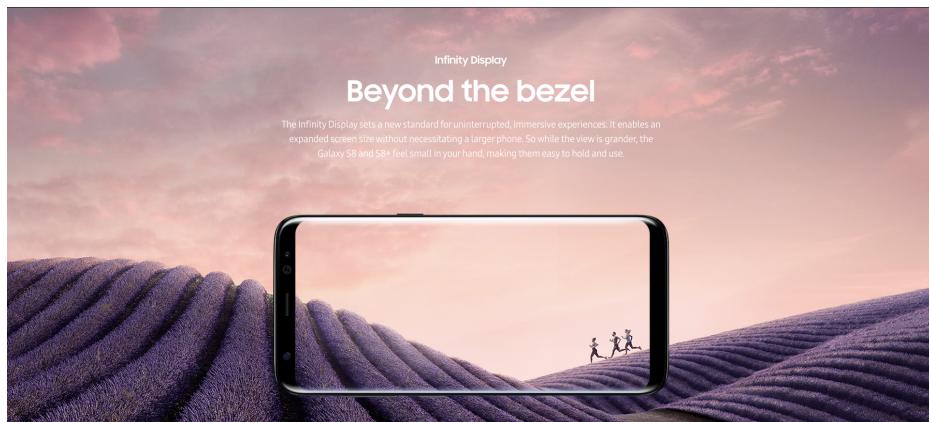
The above plot is the sentiment score for the whole period of time for the iPhone products. The amount of voice before launching is similar to the amount of voice after launching. The positive review around 0.5-0.7 sentiment score is slightly higher than before launching and the negative review around -0.4 sentiment score is also lower than before launching. In sum, our iPhone product meets the expectations of users. Before launching, customers have a positive expectation, and after launching they still feel positive about our products.



The above plot is the sentiment score for the whole period of time for the Samsung s8 products. The amount of voice after launching is significantly higher than the amount of voice before launching. Especially in positive scores, we can see lots of positive voices on social media.

When comparing the before and after launch sentiment scores of the two respective phones, we found that Samsung's scores increased significantly after the launch with a higher number of posts while iPhones maintained similar levels after the launch.

We are thus interested to find out how and why Samsung is able to generate the amount of buzz that it did. Our hypothesis is that Samsung has many innovations, such as the infinity display on screen, which captures customers' great attention on social media.



Example of Samsung S8 innovation: Infinity Edge

In the following step, we will look more into the product attributes to hopefully find a link to this occurrence.



Step 5: Attribute Analysis

We will firstly conduct topic modeling to find out the most important attributes customers care about for both iPhone and Samsung. For these topics, we will further investigate using POS to figure out what customers think about the attributes. We will then perform further market research based on the attributes that the customers mention. Furthermore, we will do text sentiment analysis for posts involving those attributes.

5.1 Topic Modeling

In order to identify what most customers are concerned about, we performed topic modeling on the Sound Bite Text. This will cluster and return the top words per topic and their probabilities within the topic. From this result, we can then interpret certain attributes that the consumers are interested in and do further market research on them.

Since every time generating the topics, we may get different results. Here are some overall general interesting attributes for the iPhone.

- Cluster 1: cases, outlook
- Cluster 2: colors, storage
- Cluster 3: camera/ photography
- Cluster 4: battery

For our iPhones, we can see that most of the attributes are familiar attributes that have consistently been sought after by our customers.

An interesting thing to note is that the attributes that show up are due to changes in the iPhone 8 and iPhone X. For example, cases might be due to the increase in iPhone screen size, and colors would be due to the release of the new Gold color. As for cameras, the iPhone X was the first iPhone with 2 back cameras, creating buzz. Lastly, for battery, it could be due to the first time iPhone 8 and X were released with wireless charging capabilities.

For reference, please refer to:

- Gold Color
- Camera, Wireless Charging

Since every time generating the topics, we may get different results. Here are some overall general interesting attributes for Samsung.

- Cluster 1: competitors
- Cluster 2: Bixby voice assistant and snapdragon 835
- Cluster 3: iris scanning
- Cluster 4: screen (due to Infinity Display)
- Cluster 5: protection, cases
- Cluster 6: operating systems
- Cluster 7: leaking news in 2017 about the launch



Key Takeaway: An interesting thing that we observed is that Samsung actually made a great breakthrough in 2017 with regard to its phone's newest features: bixby voice assistant, snapdragon 835, infinity edge, and some external factors including leaked news regarding 2017 launch. All these could possibly explain the reason for increased buzz after the launch.

For reference, please refer to:

- [Bixby](#)
- [Snapdragon 835](#)
- [Iris Scanning](#)
- [Infinity Display](#)
- [2017 Leaked News](#)

As a result, together with the previous sentiment section, we again see a hypothesis that the heat in social media buzz of Galaxy S8 could have been generated by the new features. And we would like to verify this in the following analysis.

5.2 POS Analysis

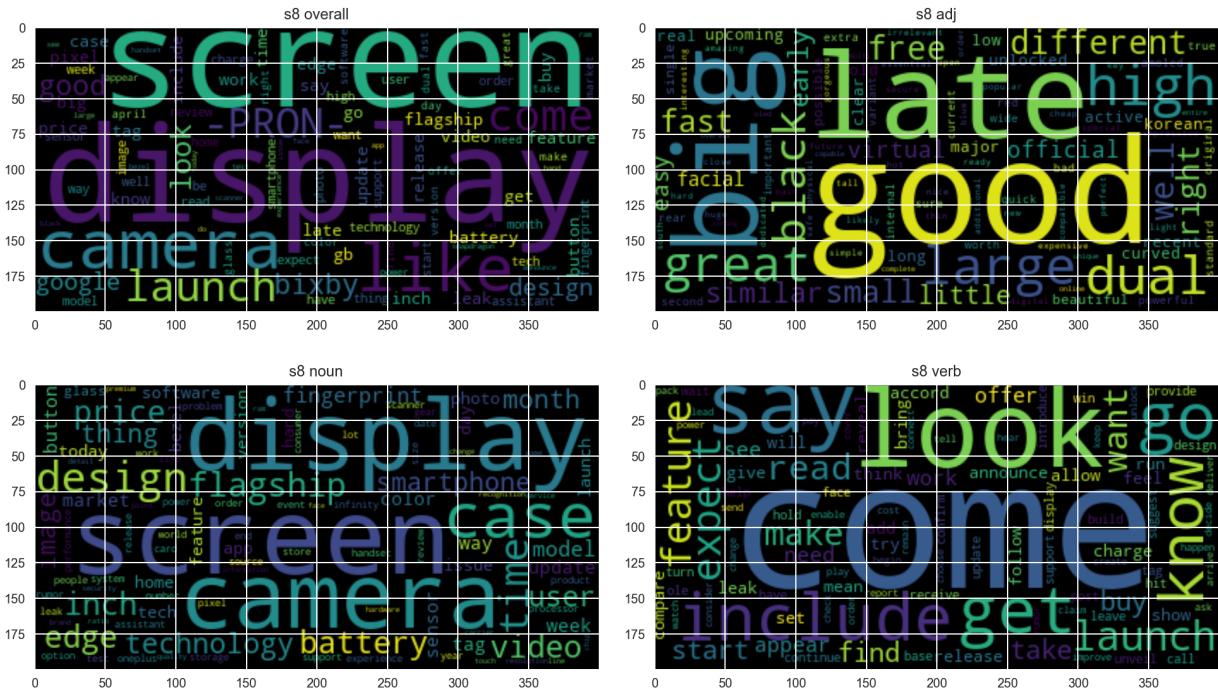


From the cloud graph above we can capture some keywords that people talked about iPhone 8/X.

Noun words: Camera, case, photography, photo, silver, design

Adj words: beautiful, good

We discovered that the words associated with iphones are more related to the photo taking features, and the stylish, good-looking appearance of the product.



From the cloud graph above we can capture some keywords that people talked about Galaxy s8.

Noun words: Screen, Camera, display, case, battery, time

Adj words: good, high, big, different, large

Matching what we expected, the new features of Galaxy s8 are the most discussed topics, in section 5.4 we will explore what do people think about those attributes.

5.3 Hashtag Frequency

We collect the top 100 highest frequency words for both iphone and samsung. One interesting finding is that there are several special and only-for iphone hashtags, such as “iphoneography”, “iphoneonly”, “iphoneedition”, “shotoniphe8plus”, “iphone8murah”, and “iphonemurah”. Some words like “murah” are even not English but Malay. This could be related to the special iphone culture, since there aren’t any kinds of this word in Samsung hashtags pool. Another interesting finding is that there are brands beyond apple and samsung mentioned in Samsung hashtags pool, such as “huawei” and “htc”.

iphone 8/X hashtags	Samsung s8 hashtags
<pre>['iphone8', 'iphone8plus', 'apple', 'iphone', 'iphonex', 'technology', 'iphone7', 'tech', 'ios', 'iphone7plus', 'photography', 'smartphone', 'instagood', 'applewatch', 'appleiphone', 'phone', 'love', 'new', 'shotoniphone', 'ios11', 'photooftheday', 'gold', 'mobile', 'design', 'newiphone', 'beautiful', 'like4like', 'iphoneography', 'portraitmode', 'jualiphone8', 'iphoneonly', 'iphoneedition', 'portrait', 'instagram',</pre>	<pre>['samsung', 'iphone8', 'apple', 'galaxys8', 'android', 'tech', 'iphone8plus', 'galaxy', 'iphone', 'ifttt', 'technology', 's8', 'smartphone', 'samsunggalaxys8', 'news', 'iphonex', 'samsunggalaxy', 'phone', 'twitter', 'google', 'galaxys8plus', 'mobile', 's8plus', 'instagood', 'photography', 'gadgets', 'love', 'ios', 'samsungs8', 'unboxyourphone', 'photooftheday', 'samsunggalaxys8plus', 'picoftheday', 'iphone7',</pre>



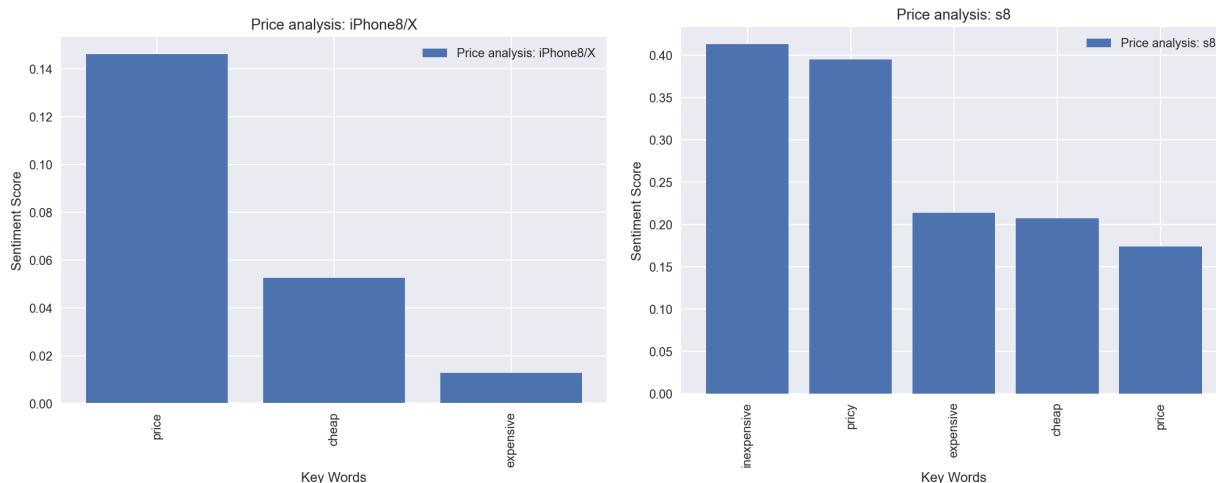
'instadaily', 'picoftheday', 'nofilter', 'iphonecase', 'follow', 'samsung', 'iphonephotography', 'jualiphone8plus', 'gadgets', 'followme', 'luxury', 'camera', 'ipad', 'newphone', 'shotoniphone8plus', 'macbook', 'fashion', 'jualiphone', 'photo', 'applestore', 'airpods', 'teamiphone', 'style', 'iphonesia', 'black', 'selfie', 'iphone10', 'electronics', 'instalike', 'iphone6', 'likeforlike', 'iphone8case', 'nature', 'happy', 'phonecase', 'art', 'photographer', 'silver', 'iphone8murah', 'amazing', 'portraitphotography', 'technews', 'news', 'jualphonex', 'gadget', 'applewatchseries3', 'lifestyle', 'concept', 'iphonemurah', 'follow4follow', 'like', 'instaiphone', 'followforfollow', 'ipod', 'case', 'accessories', 'mac', 'cute', 'stevejobs', 'appleiphone8', 'jualiphone8murah', 'fall', 'usa', 'me', 'jualiphonemurah', 'iphone6s']

'pinterest', 'new', 'galaxys7', 'like4like', 'canada', 'technews', 'trending', 'beautiful', 'iphone7plus', 'video', 'instagram', 'youtube', 'note', 'design', 'followme', 'follow', 'amazing', 'huawei', 'androbrix', 'ftag', 'instadaily', 'smartphones', 'nature', 'instalike', 'appleiphone', 'applewatch', 'cell', 'instatech', 'samsungblog_', 'gadget', 'style', 'motorola', 'selfie', 'caseme', 'galaxys7edge', 'like', 'nofilter', 'photo', 's7edge', 'digital', 'camera', 'great', 'infinitydisplay', 'happy', 'exo', 'hot', 'follow4follow', '2017', 'fashion', 'newphone', 'windows', 'cute', 'case', 'shotoniphone', 'ios11', 'gold', 'likeforlike', 's', 'htc', 'oneplus', 'me', 'innovation', 'nougat', 'black', 'lg', 'electronics']

5.4 Sentiment Analysis for Product Attributes

We are going to use Price-Quality-Value(PQV) Framework for sentiment analysis. Based on the attribute analysis(POS) and topic modeling results, we discovered some key words in PQV and would like to learn more about consumer preferences.

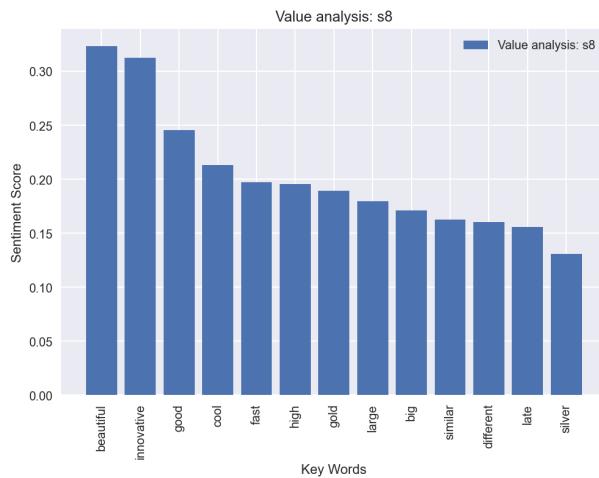
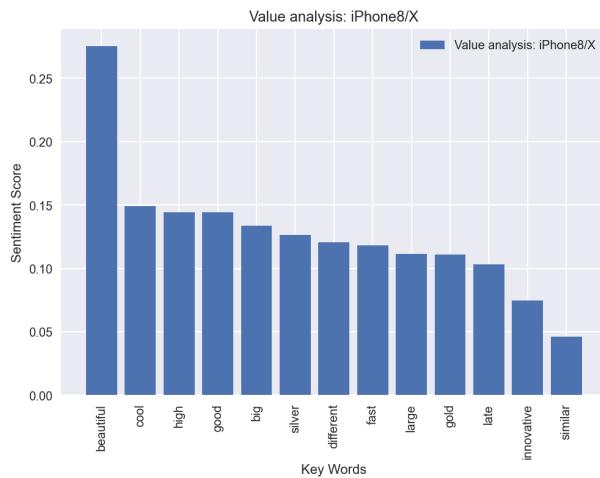
5.4.1 Price Analysis:



In terms of "Price", we found that both iPhone and Galaxy s8 fall around positive 0.15 (iPhone even has a slightly larger figure) which means that overall we are performing pretty well in price setting. One part worth noticing is that Galaxy s8 is having a positive 0.40+ in "inexpensive" which might infer that although we are actually selling at a similar price, consumers might acknowledge Galaxy s8 having better value for money.



5.4.2 Value Analysis:



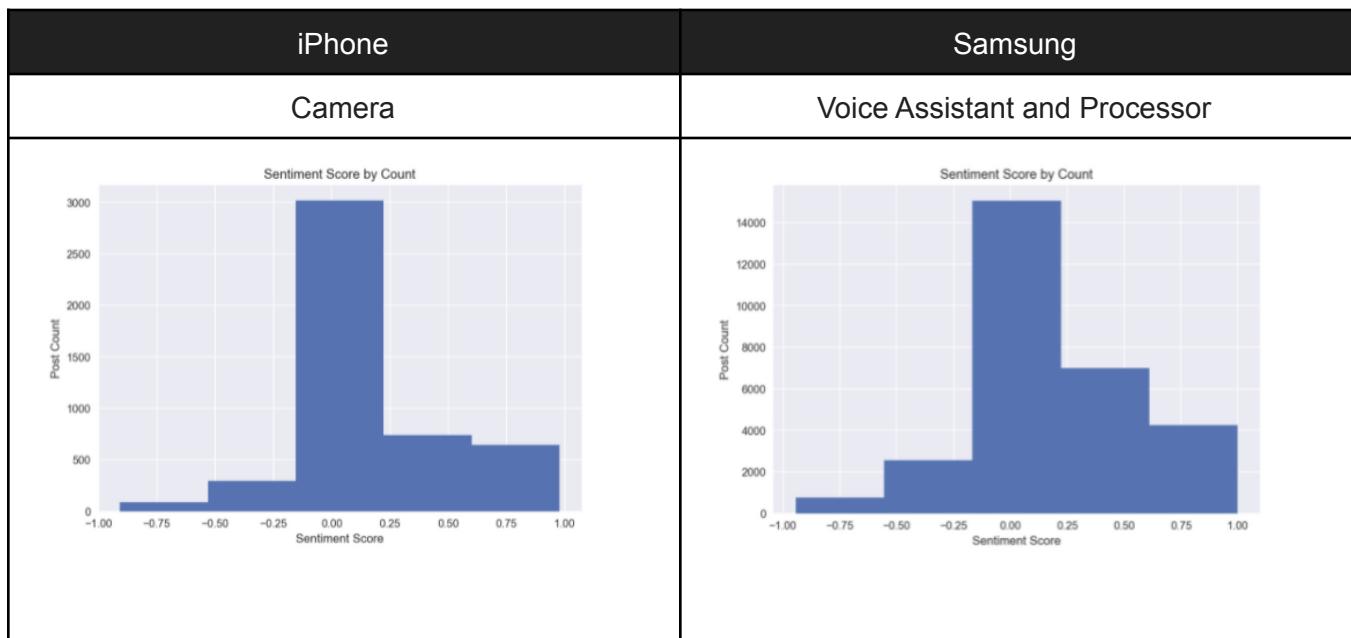
"Beautiful" is one of iPhone's most frequent words, and not surprisingly, we are having a pretty good sentiment score on it. However, with new designs in display and features, Galaxy s8 outperformed us in almost every aspect, including appearance, innovation, overall feelings in "good" and "different".

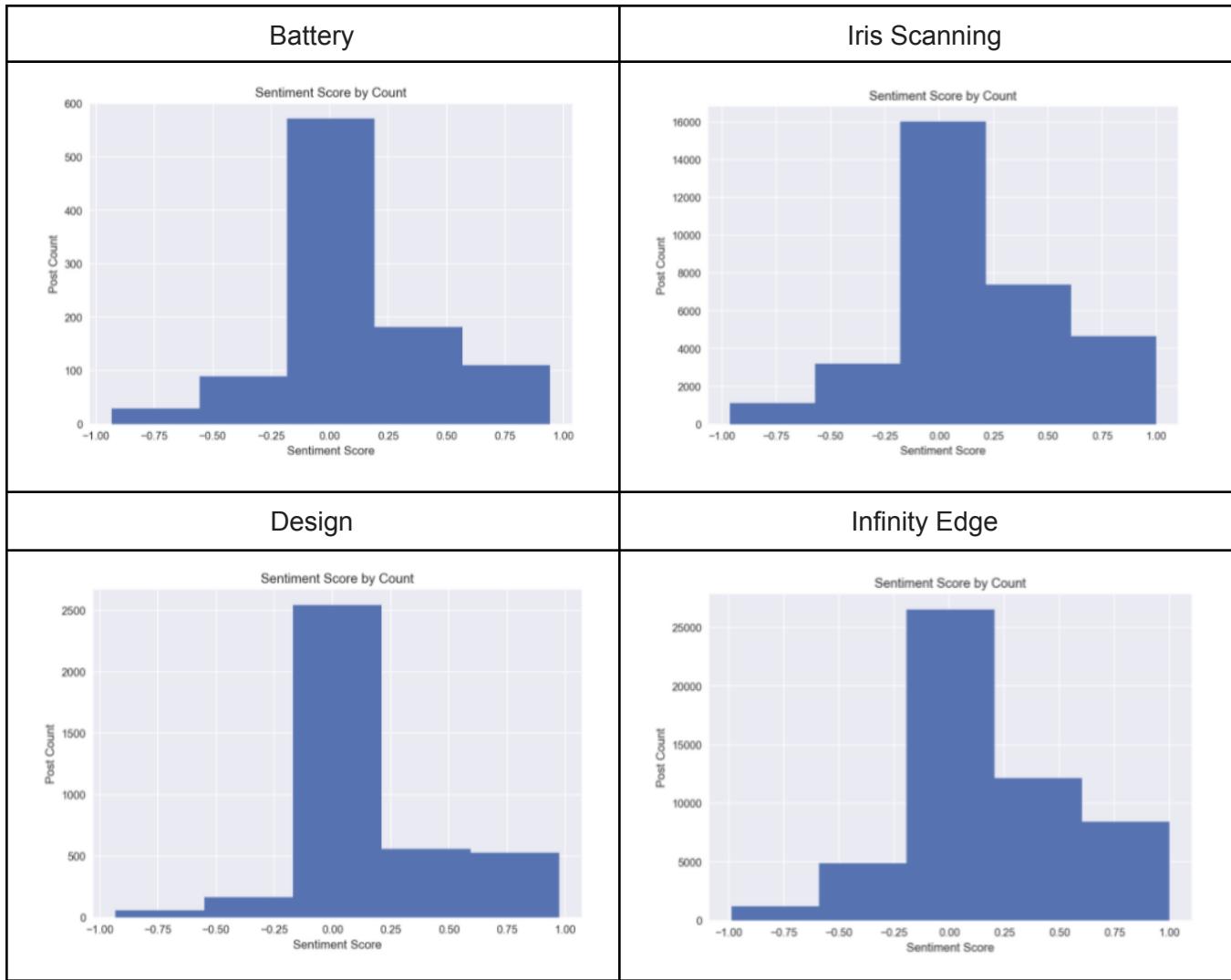
5.4.3 Quality Analysis:

This would be a special focus on analyzing the device features that customers discuss for each brand. We will pick the top 3 features based on the topic modeling and then conduct sentiment analysis. If necessary, market research will be used to explain certain sentiment trends.

iPhone: The three selected attributes that we want to focus on are camera, battery and design.

Samsung: The three selected attributes that we want to focus on are: bixby voice assistant and snapdragon 835, iris scanning as well as screen (due to Infinity Display).





To sum up, we can see that Samsung's feature generates a great deal of positive sentiments compared to iPhone's. iPhone's performance has been stable and has maintained its strength in providing beautiful and stylish designs as well as keeping its adequate pricing strategy. Nonetheless, playing safe didn't seem like a perfect idea in creating an online buzz in this competition since Samsung successfully won attention from the market by its highly unexpected features.

Step 6: Predictive Analysis

In this part, we conduct the predictive analysis using self-defined features with generated labels to predict the uptake of iPhone and Samsung products. By trying out two different machine learning models, we evaluate each of their performances with different metrics and provide our suggestion of which model will be a more suitable predictor.

6.1 Feature Engineering

To predict the adoption, we extract features that we guess might be associated with uptake of the products, such as text length, sentiment score, and media types mentioned in the post.



	Features								Label	
	Phone_Type	After	Length	Sentiment_Score	Image	Video	Audio_Files	Link	Adoption	
1	1	43		-0.3252	0	0		0	1	False
1	1	4		0.0000	1	0		0	0	False
1	1	35		0.0516	1	0		0	0	False
1	1	623		0.0000	0	0		0	0	False
1	1	33		-0.1779	0	0		0	0	False

6.2 Prediction of Product Uptake

During the model training process, we conducted K-fold validation to choose our hyperparameter and use the confusion matrix to evaluate our classification results.

For the classification problem, we will use logistic regression as our baseline model, then using random forest classifier to benchmark our performance. In particular, we will tune our random forest hyperparameters to use 5-fold cross-validation.

6.3 Model Fitting and Evaluation

	Accuracy	Precision	Recall	ROC_AUC
Logistic Regression	0.622	0.654	0.518	0.622
Random Forest Classifier	0.605	0.571	0.84	0.605

Logistic Regression		Random Forest Classifier	
	(Actual) Positive	(Actual) Negative	
(Predicted) Positive	4183	1826	(Actual) Positive
(Predicted) Negative	3204	3453	(Actual) Negative

	(Actual) Positive	(Actual) Negative
(Predicted) Positive	2465	4192
(Predicted) Negative	1062	5595

While comparing the evaluation result of Logistic Regression and Random Forest Classifier, we discover that although Logistic Regression performs better in terms of Accuracy and Precision, it has a higher False Negative rate than the Random Forest model. We believe that false negatives are more sensitive than false positives because that means we will possibly miss a potential customer while he/she will actually adopt our product. In conclusion, we believe that using Random Forest Classifier will better help us predict the uptake of our products.



Step 7: Comparison between Twitter and non-Twitter Dataset

In this section, we aim to come up with conclusions on the ease or accuracy of processing and analysis for these two sets. Through EDA, word frequency, hashtag analysis, and sentiment analysis, we are able to find the common points and differences in these two datasets.

7.1 Exploratory Data Analysis of Twitter and Non-Twitter Dataset

We first get the basic information of these two datasets and check if there are some columns without many non-value that we can look at. Then, we count the number of data related to iphone and samsung, count the value of columns "Post Type", "Media Type" and "Author Location - Country 1".

The results show that 60% of non-twitter text comes from the US, but the author countries in twitter are more diverse US: 0.38, India 0.12, UK: 0.09, Canada 0.03. The total No. of text in non-tweet is also larger than the total No. of text in tweet (198384 v.s 19100). In both dataset tweet and non-tweet, Samsung s8 has more text compared to iphone.(No. of iPhone text in non-tweet: 7947; No. of Samsung text in non-tweet: 104771; No. of iPhone text in tweet: 220; No. of Samsung text in tweet: 11207)

Thus, when we want to hear customer opinions on the US market, we could use non-twitter sources. However, when we want to have more diverse or global customer opinions, we should use twitter sources.

7.2 Word Frequency for Twitter and Non-Twitter Dataset

In the code file, we tokenize words in four groups (iphone related text and s8 related text in twitter or non twitter dataset). Then, for each group we do POS and get the highest frequency adj, noun, and verb in each group.

The following table contains the 10 highest frequency of POS words in each group. One obvious thing we can see is that on non twitter platforms, people care more about iphone appearance and color (e.g. black, gray, gold, silver, gold and beautiful). On twitter, people mention using more function-based adj (e.g. oleddisplay, facial for facial recognition) and attach more youtube links.

	Top 10 words in Twitter	Top 10 words in Non Twitter
iPhone (adj.)	[('@apple', 43), ('good', 7), ('real', 3), ('blue', 3), ('low', 3), ('easyphone', 3), ('free', 3), ('waterproof', 3), ('severe', 2), ('minimal', 2)]	[('black', 653), ('beautiful', 467), ('free', 433), ('amazing', 403), ('good', 393), ('gray', 355), ('ready', 332), ('gold', 293), ('unlocked', 285), ('late', 284)]
iPhone (noun.)	[('leak', 10), ('case', 10), ('tech', 9), ('design', 9), ('ios', 9), ('rumor', 8), ('technology', 6), ('iphone7', 6), ('display', 5), ('technew', 5)]	[('camera', 887), ('gold', 862), ('case', 797), ('silver', 671), ('technology', 647), ('tag', 587), ('design', 543), ('tech', 517), ('photo', 466), ('model', 441)]



Samsung s8 (adj.)	[('good', 330), ('late', 147), ('big', 147), ('free', 124), ('dual', 101), ('@youtube', 98), ('black', 97), ('facial', 97), ('unlocked', 93), ('official', 85)]	[('good', 17504), ('late', 10973), ('big', 10674), ('high', 7734), ('large', 6964), ('dual', 6708), ('great', 6540), ('black', 4900), ('different', 4885), ('well', 4801)]
Samsung s8 (noun.)	[('case', 459), ('video', 436), ('screen', 334), ('cnet', 276), ('tech', 261), ('test', 220), ('leak', 212), ('price', 185), ('technology', 185), ('order', 178)]	[('screen', 28620), ('display', 23394), ('camera', 20434), ('case', 14586), ('design', 14082), ('time', 13443), ('flagship', 12530), ('battery', 10210), ('price', 10174), ('inch', 9956)]

7.3 Hashtag for Twitter and Non-Twitter Dataset

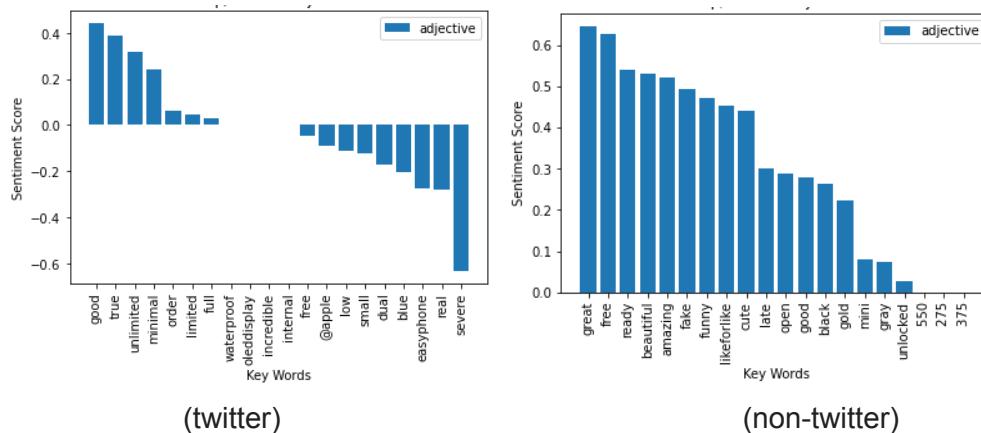
In the code file, we get the highest frequency hashtag in four groups (iphone related text and s8 related text in twitter or non twitter dataset). Then, we compare the iphone/s8 hashtag in both twitter and nontwitter platforms. From hundreds of hashtag in the two datasets, the unique one in each dataset is as following table:

	Unique Hashtag on Twitter	Unique Hashtag on non Twitter
iPhone	{'smartphones', 'video', 'youtube'}	{'ipad', 'iphoneonly', 'jualiphonex'}
Samsung s8	{'followback', 'ipad', 'jualiphonex', 'swag', 'tagforlikes'}	{'note', 'otterbox', 'selfie', 'smartphones', 'video'}

Surprisingly, though two dataset have quite different numbers and demographics of text and author, people use similar words of hashtags on both twitter and non twitter.

7.4 Sentiment Analysis for Product Attributes for Twitter and Non-Twitter Dataset

In this section, we compute sentiment scores to specific iphone/s8 product attributes. We use the top20 and bottom 20 frequency of POS words to do the comparison between two products in the code file. Due to the pages limited, we focus on the adj word for iphone in these two datasets. From the following plots, we could see Twitter has both positive and negative adj, but nonTwitter only has positive adj.



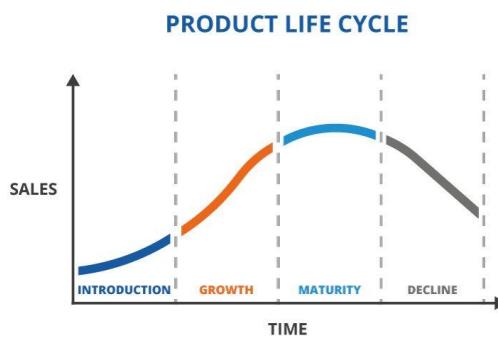


To sum up, twitter dataset has more ease and accuracy of processing and analysis. The twitter dataset owns customers with more diverse demographics and provides a more diverse voice. This would help the analysis get insights that are more accurate and closer to the entire market. Besides, the size of twitter dataset is smaller, lowering the computational cost.

Step 8: Recommendation

A Strategy for Every Phase

From the preceding analysis, we have identified five key recommendations, which we will execute at each stage of the product development life-cycle for each new product.



8.0 Development: Attribute Enhancement

The first recommendation is that we should be more proactive in improving our product attributes. This strategy will be executed during the product development phase. We discovered that Samsung and Apple have similar market shares and product price points after doing more investigation. Despite the fact that we had fantastic features such as great product appearance and great cameras, Samsung was able to outperform us due to a few key big flagship advancements, such as the infinity display and the current Snapdragon 835 CPU, among other things. Our iPhone X debut was quite similar, with more people praising the new Gold look, as well as the multiple cameras and wireless charging battery capabilities. These features give customers a sense of “upgrade” and freshness, which creates excitement and buzz for Samsung and Apple, and creates anticipation for the next launch. As a result, we should focus on improving year after year in order to keep existing clients and attract new ones, particularly with an emphasis to create a sleek design and be different, as extracted from the analysis results.

8.1 Preparation: Informative Marketing

Apple should focus on informative marketing in the introduction phase. At this stage, our product is still new and customers need time to learn about it. We'll invite influencers to launch events and use community engagement to drive excitement. During this phase, we will target early adopters and loyal customers, and engage them in focus groups. We will use the information we extracted from attribute and sentiment analysis to find out quality information on what customers truly think about those aspects, and make changes on the products as much as possible.



8.2 Growth: Persuasive Marketing

In addition to the above, we should incorporate advanced marketing tools to further tap on the launch of our products during the Growth stage. The product is beginning to gain traction and getting noticed by consumers. Understanding our customers' needs and wants is essential for us to succeed. Our predictive algorithm allows us to forecast whether a potential customer will purchase the product when it launches. We recognize the importance of privacy and take it very seriously. To that end, we will only extract publicly available posts incorporating the use of relevant key hashtags, for example #iphone etc. We will aggregate the proportion of potential buyers against the entire market using our predictive model. We will then tweak our marketing strategy accordingly, say to promote more features that the customers care about, or further spread positive anticipated features that customers are excited about to maximize the network externality effect.

8.3 Maturity: Product Consolidation

We'll start consolidating market share after we've reached maturity, and we'll start summarizing critical takeaways. The market is becoming saturated, and there is a lot of critical information available. We can figure out what people liked about the product and what their pain points were using our analysis. We may begin studying new product development ideas and including the community in the process using this knowledge. Focus groups could be formed to get their feedback on what they desire from their product. We can organize a design/features competition, for example, and collect useful information for future product attribute enhancements.

8.4 Decline: Pivoting

Lastly, as the product reaches the end of its life, we will try to pivot the remaining customers over to our newer products. We will communicate early to give users time and prepare them for the termination of this product. Apart from that, we can provide incentives for people to change, with discounts from trade-in and provide additional benefits such as a month subscription to Apple Music etc.

Conclusion

Overall, every product is a marketing opportunity, and we will do everything we can to maximize its benefits. We believe that if these suggestions are implemented correctly, Apple will be able to reach new heights and maintain its position as the world's leading phone and technological manufacturer.

Contribution

Ashlyn Im: Data Preprocessing, Topic Modeling, Predictive Modeling, Recommendation

Linda Hung: Data Preprocessing, Hashtag Analysis, Sentiment Analysis, Twitter Comparison, PDF

Shih-Ting Sun: POS Tagging, Word Frequency, In-depth Attribute Insights, Attribute Sentiment Analysis

Wan-Ju Yu: Demographic Analysis, Feature Engineering, Prediction of Product Update, PDF