DSPM Final Project

Project Report on NLP Analysis of Consumer posts on Samsung Galaxy S8 vs. iPhone 8, iPhone X

Presented By:

Ruochen Wang

Mengyang Li

Tanaya Seth

Date: March 4, 2020

Contents

1. Introduction

Overview

The fact (Chaffey, 2019) that number of Social Media Users rises almost by 10% every year, makes it obvious that customers will opt for online reviews to convey their feelings about a product, so that their review helps the community at large. The following data snippet shows the rising number of people on social media platforms. More the number of people on these platforms, more data will be generated about them, how they feels and how some reviewers influence what other people feel about the product.

Hence, Social Media Analysis is a great way to see what Consumers think about a product, how they are projecting and influencing through these thoughts in the community via social media platforms. It helps both the company and the customers in the following ways:

1. The company – to see demographics of your social media audience – likes, dislikes, locations etc.
2. To company – to see impact and reach of the content you put online, and on which platform
3. To company – to see what product features people are talking about and how your product compares to others – mentions of similar products along with your product in the same post
4. To customer – to make a decision on whether to buy the product after reading up online reviews about different similar products

Having robust Social Media Analysis can help firms create value and achieve competitive advantage over their competitors

1.1 Problem Statement

The task is to perform Natural Language processing on review data from various social media platforms. The reviews are for top 3 mobile handsets that were released in 2017, namely

- Samsung Galaxy S8, launched on March 29, 2017
- iPhone 8, launched on September 22, 2017
- iPhone X, launched on September 22, 2017

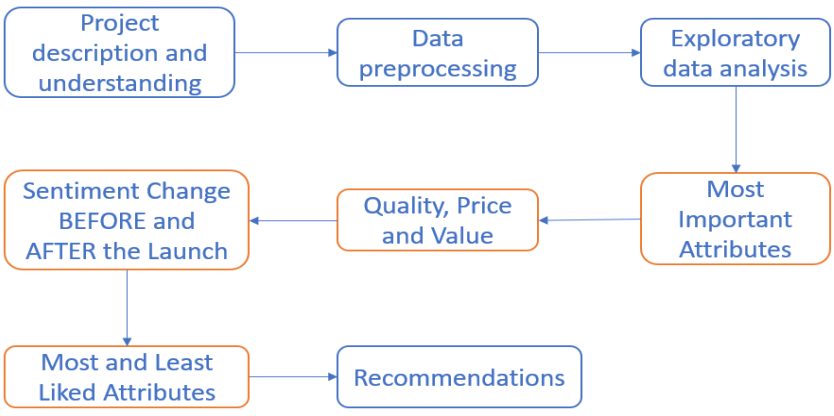
The data is timed from just before the launch of these products, that is from March 2017 to just after the launch, November 2017. Various analyses like Exploratory Data Analysis, Sentiment Analysis and Topic Modelling will help us profile the reviewers and this is turn, will help us make recommendations to respective product companies to improve their product marketing.

1.2 Project Objectives

The major objectives are:

- Gain insights from this data about the most talked about topics for each of these models using topic modelling
- Observe demographics of Social Media Audience by EDA
- Observe outliers for this case and how to handle them
- Find out how the consumers feels about these topics by doing sentiment analysis
- Compare the volume and sentiment of reviews before and after the product launch dates
- Understand how social media platforms play a part in product marketing and how customer reviews impact product sales
- Make recommendations from these major findings
- Explore who can benefits from this analysis solution

1.3 Flowchart of analytics Pipeline



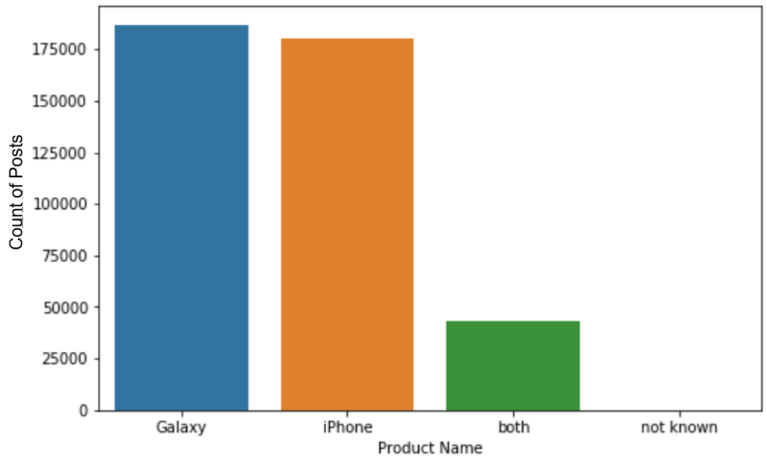
1.4 Data Pre-processing

1. Merge all data sources into one data frame
2. Delete columns that have more than 60% of NaNs, since any column that has a greater than 60% of NaN would potentially bias our analysis. Columns we used are: Sound Bite Text, Source Type, Post Type, Date, No. of Followers/Daily Unique Visitors
3. Drop the duplicate posts so that we can maintain the uniqueness of the data we use
4. Filter the data by only keeping the posts which are original, and filter out posts which are comments or replies. This way we can avoid getting repetitive information to bias the analysis
5. Check the number of followers (a threshold of 2,000). This step aims to filter out posts from celebrities or paid reporters, since we only focus on data from regular consumers
6. Remove special characters and links from Sound Bite Text

2. Exploratory Data Analysis

2.1 Distribution of Number of Posts

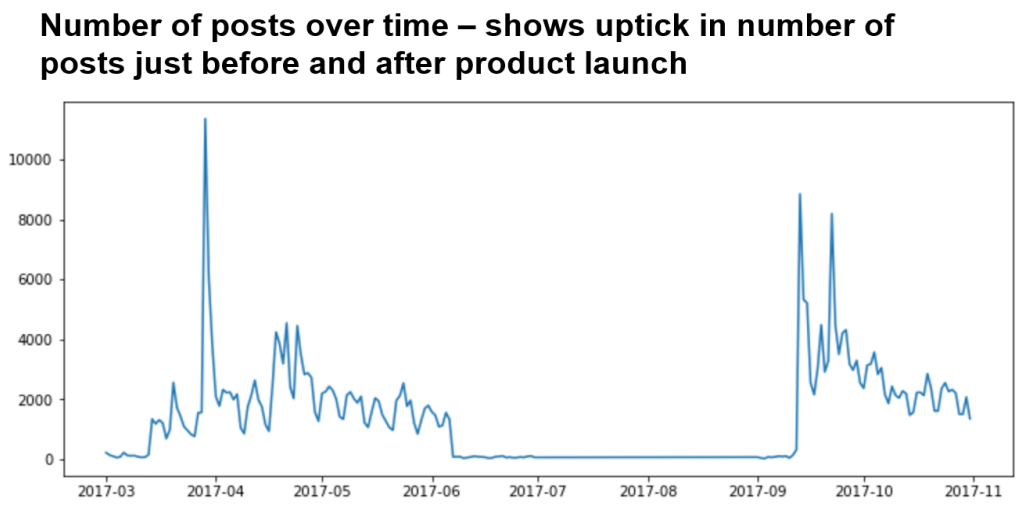
1) Number of Posts by Product



The number of posts by different products are quite the same, people pay similar attention to iPhone and Samsung Galaxy. They would also like to mention these two models together and compare them.

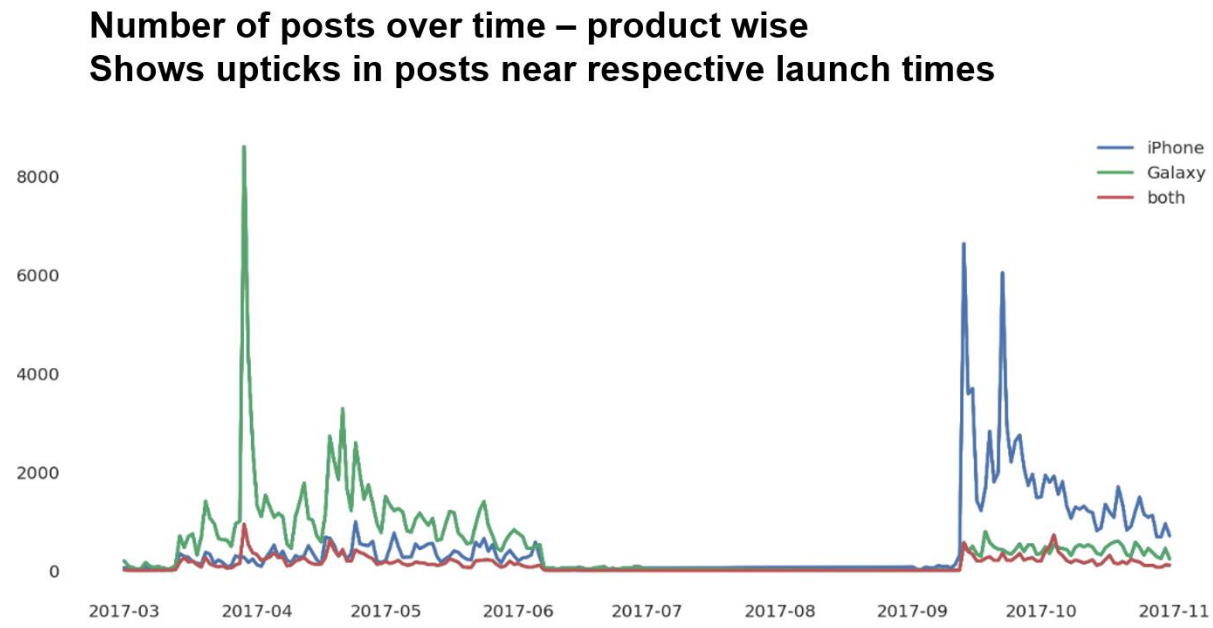
2.2 Time Series Analysis

1) Number of posts over time



It shows the trend that people would like to discuss about these products just before and after release date.

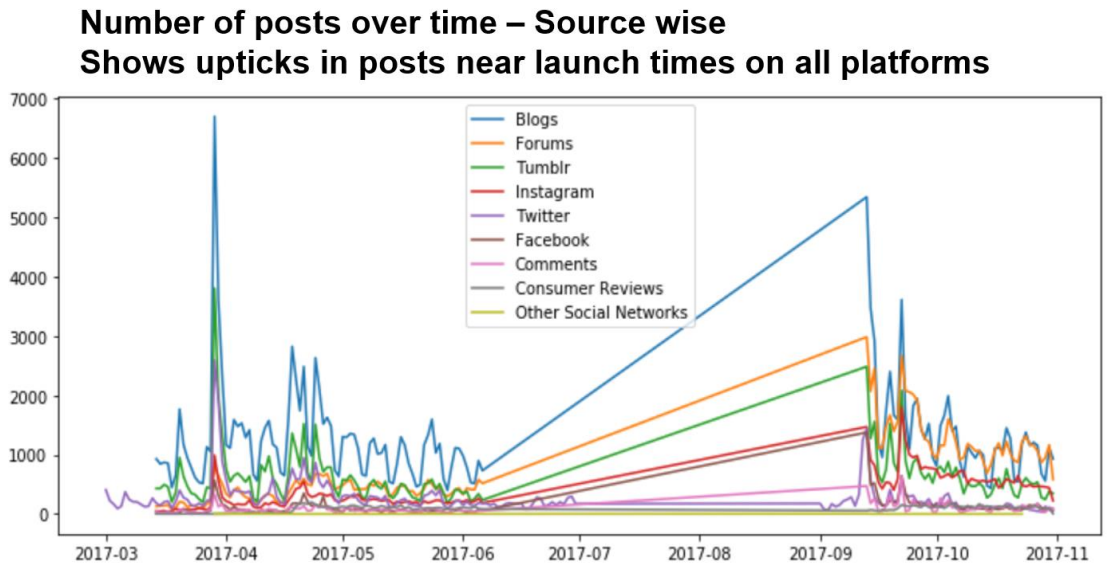
2) Number of posts over time by product



Since the Samsung's release data is 29th March and the release date of iPhone is 22nd Sep, this trend exactly demonstrate that information heat rate fluctuation on social media platforms. There two peaks for Galaxy and iPhone after their launch time. The difference here is that iPhone trend has experienced another peak one month after the first peak. It might be people are discussing their new technology implementations or there might be some problems coming

out when they use the phone for a while (such as screen bleak). On the other hand, people's interests on Galaxy go down quickly compare to the iPhone.

3) Number of posts over time – by different social media platforms



This graph above shows that similar trend of number of posts across different social media platforms. People prefer to show their opinions on Blogs, Forums and Tumblr. The largest peak was found in Blogs just after the launch time, it might be because there are some technology professionals who write analysis reports about the products after phone release.

3. Topic Modeling

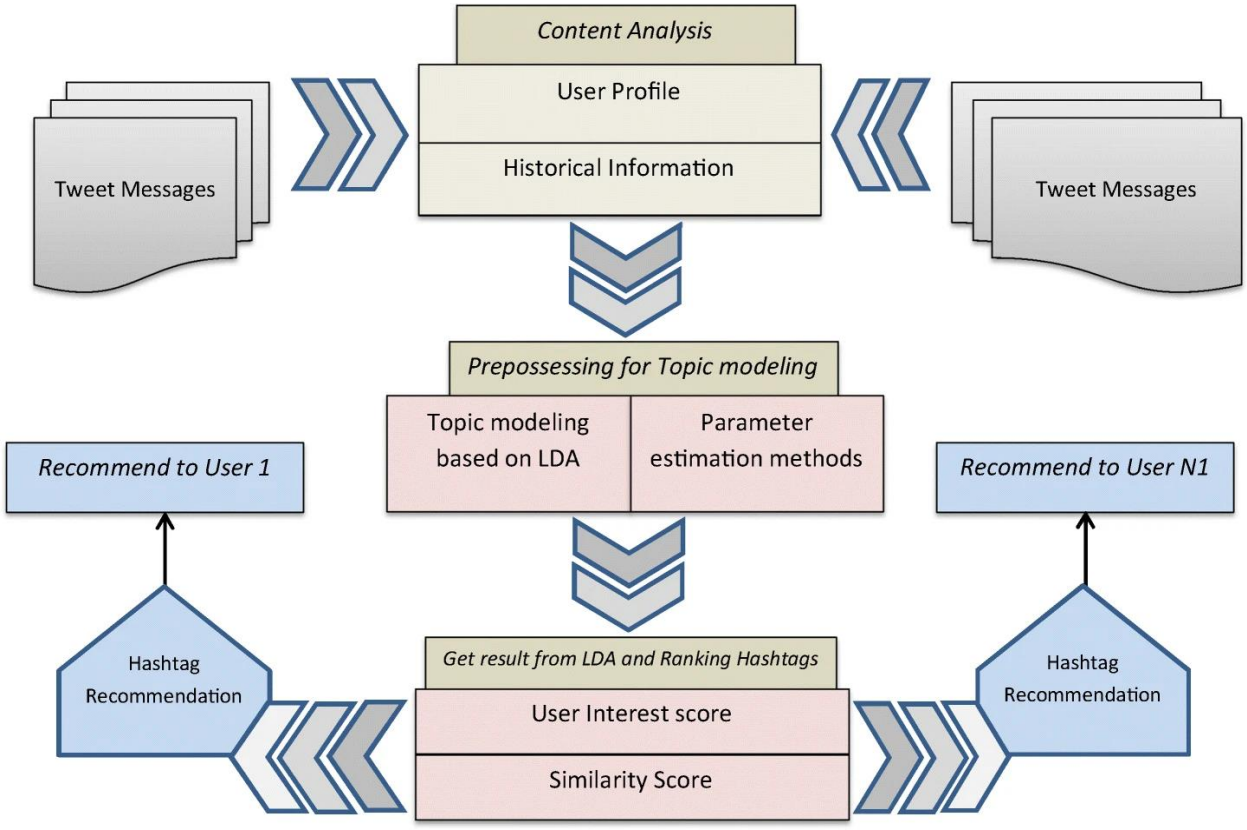
3.1 Methodology

Topic modeling is a type of statistical modeling for discovering the abstract "topics" that occur in a collection of documents. Latent Dirichlet Allocation (LDA) is an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions.

Here we apply the LDA model to a set of post texts from different social media platforms, split them into different topics. Then we find out what kinds of key topics they are considering on social media. For the dataset we used, we have the original dataset with all three products (after preprocessing) and the three subset datasets as we separated total data into three parts: iPhoneX, iPhone8 and Samsung 8 by filtering the key words in their text. We used these four datasets to go through the LDA model to find out the most important attributes and its connection to different phone models.

We assume that there are at least 6 clusters for the total dataset, since we have three different models of phone and different features of phones such as quality, price and value. For discovering more connections between top features and different phone models, we set the K as 8, put it into 8 clusters.

Below is one example of how LDA mode works for content analysis.



A simple framework based on LAD to generate tag as a recommendation system on twitter

Jelodar, H., Wang, Y., Yuan, C. et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 78, 2019

3.2 LDA Result

Through the LDA model, we got 8 clusters. Five of them are highly related to Samsung Galaxy, and the other three are connected to iPhone.

A. Samsung Galaxy

	Cluster 1	Cluster 3	Cluster 5	Cluster 6	Cluster 7
Focus	Camera	Security	Protector	Platform	Bixby
Top Features	galaxy, display, camera, screen, pixel, design	galaxy, unlock, camera, scanner, security, recognition, fingerprint	galaxy, case, plus, screen, protect, glass, cover, protector	Twitter, galaxy, tag, iPhone, tech, Samsung, YouTube	galaxi, Samsung, launch, android, Bixby, note, Mobil
Business Insights	People cares more about the camera and the appearance of the phone	For galaxy, people are concerned about the new facial recognition technology's security, and compared that with the traditional fingerprint unlock method	Consumers are concerned about its large screen's maintenance, whether it's strong enough, whether it needs extra protection	Which platforms did those tow companies placing their ads on (twitter, YouTube) / People cares most of these new techs embedded in the phones	People care about the Android system and voice assitance Bixby embedded in Samsung

B. iPhone

	Cluster 2	Cluster 4	Cluster 8
Focus	Unlock	Rumor	Time
Top Features	Iphone, gold, black, color, unlock, plus	Iphone, gold, black, color, unlock, plus	phone, charge, work, iphone, galaxi, download, like, time
Business Insights	People are more curious about the color of the iPhone, and whether iPhone will have the unlock button, whether it will have a plus version	Since the name and the design of iPhone X is not publicized before the launch, consumers are curious about the name and the appearance of it	People concerns about the battery level, the charger, and compare iphone's work time with Samsung

3.3 Topmost Liked Features:

To find the top most and least liked features for each model, we divided each dataset into two attitudes groups: positive and negative based on other sentiment score. Then we extract top features (top frequent words, top frequent nouns, top frequent adjectives) from these groups.

	Iphone X	Iphone 8	Galaxy 8
Focus	Accessories	Design	Design
The Most Liked Attributes	wireless case accessary car fast compatible audible safe	camera price design feature screen wireless fit	camera screen display dech
Business Insights	Consumers like the accessories comes along with the phone. They feel comfortable about the price, and they like the idea of wireless charger. The compatibility with cars is also a feature that consumers value most	For IP8, consumers like the price, the camera, the screen, the design of the size since it fits the body mechanics, and the wireless charger	Consumers like the camera, the screen layout and the high-tech feature come along with the phone

3.4 Top Least Liked Features:

	Iphone X	Iphone 8	Galaxy 8
Focus	Window	Screen	Game
The Least Liked Attributes	hold window bored	bleak store	game
Business Insights	Consumers think the phone is too large to hold, and the screen layout may contradict with some consumers' aesthetics, and consumers are sort of get bored of the design	the storage is small, and the screen easily goes to bleak	It's might not a good device for playing games

3.5 Findings and Recommendations

Based on the top modeling result we got above, we found there are certain features for iPhone and Samsung users care about.

A. Findings

- For iPhone, consumers mainly care about their design and new technologies implementation. They pay attention to the iPhone's size, color, unlock button and screen. It's interesting that people don't show too much negative attitude when it comes to iPhone's price. It might be because the value and quality of iPhone meet their needs and the expectation. Thus, consumers think it's "valuable" instead of "expensive".
- They really like those new techs such as wireless charger and accessories with cars. On the other hand, consumers criticize and get bored about their design, it might be because they don't have too much creative change on design when it compares to the previous model. For iPhone X, people think the big screen is hard to hold and the screen layout may contradict with some consumers' aesthetics. For iPhone 8, people don't like the storage and it seems to have some screen bleak problems sometime.

- For Samsung Galaxy 8, consumers care about their camera, appearance, system, security, facial recognition, protector, and voice assistance Bixby.
- They really like the camera and the screen layout with the Samsung. From the result, it seems like the consumer's game experience on Samsung is not satisfied. Besides, people show more concerns about the security and finding the phone protector.

B. Recommendations

- ➢ For the iPhone, they should maintain their appearance advantage and further develop their design rather than using a design similar to previous models or other phones.
- ➢ For the iPhone, they should use more creative technology implementations to attract consumers. They made some achievements based on some new technology's utilization like facial recognition and wireless charger.
- ➢ For the iPhone, consumers show high tolerances about the price, so the company doesn't need to worry too much about the negative impact of high price setting. They should focus more on the design and new techs.

- ➢ For Samsung, their strengths mainly were shown on their camera and screen. Samsung has the industry-leading camera because of its high pixel and photographic effect.

For Samsung, people show more concerns about the protection of the screen. They might need to focus on strengthening their screen. Moreover, the game experience on the Samsung platform should be improved.

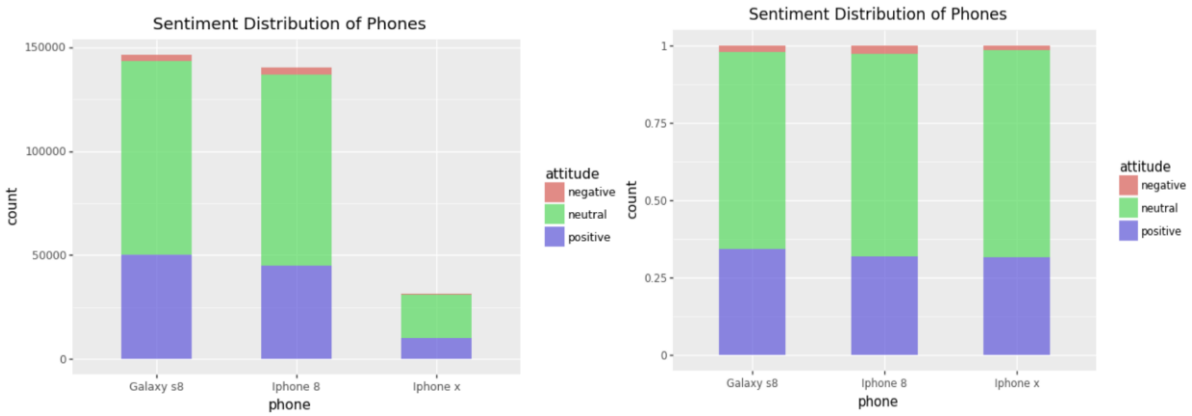
4. Sentiment Analysis - Sentiment Change Before and After Launch

4.1 General sentiment analysis for all the models

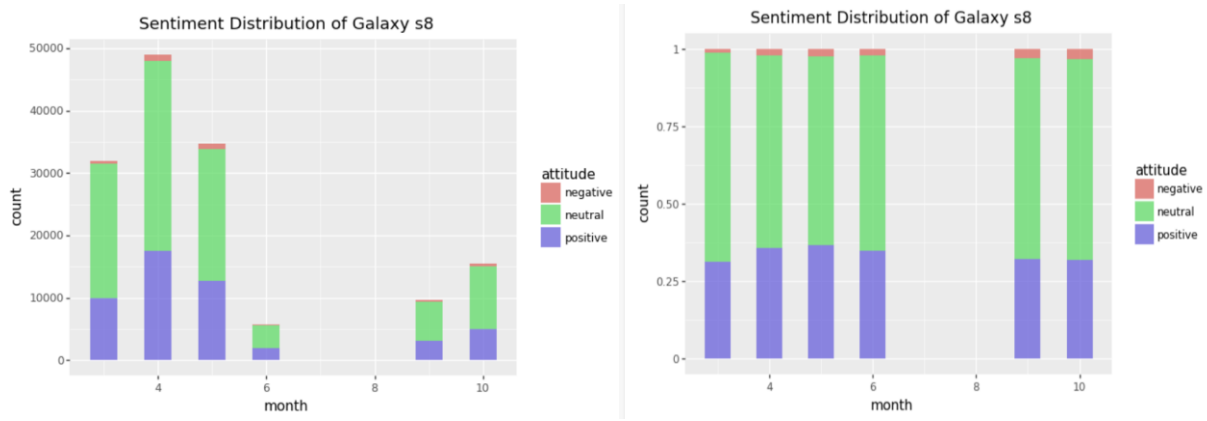
For sentiment analysis, we proceed the analysis using four data frames, which are the total data frame, the Galaxy S8 data frame, the iPhone 8 data frame, the iPhone X data frame. Thus, we can generate insights both from a large picture as well as a detailed perspective. Then we perform a general sentiment analysis and a sentiment change before and after the launch, in an attempt to discover some correlations between the sentiment relates the launch.

First, we conduct a general sentiment analysis for all the models. We use Textblob to generate sentiment score for the posts we are analyzing. The score ranges from -1 to 1, with a higher number suggesting a more positive attitude. And we decide a score among -0.2 to 0.2 as neutral, a score below -0.2 as negative, and a score above 0.2 as positive. Please note that the data for July and August are not available for our analysis.

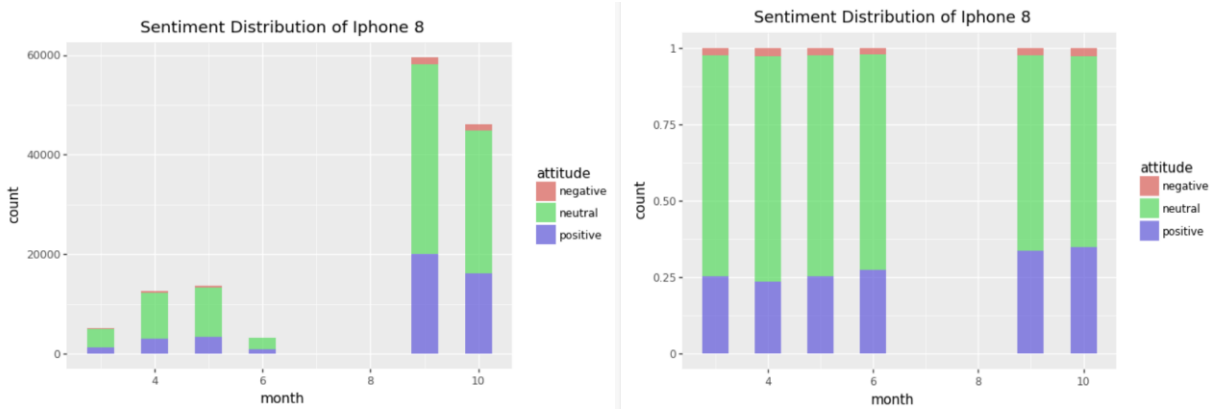
We find that iPhone X didn't get as much attention as the other two models. From a percentage point of view, iPhone X got the least negative attitude, while iPhone 8 got the most negative attitude. The positive and neutral attitudes were quite similar across three models.



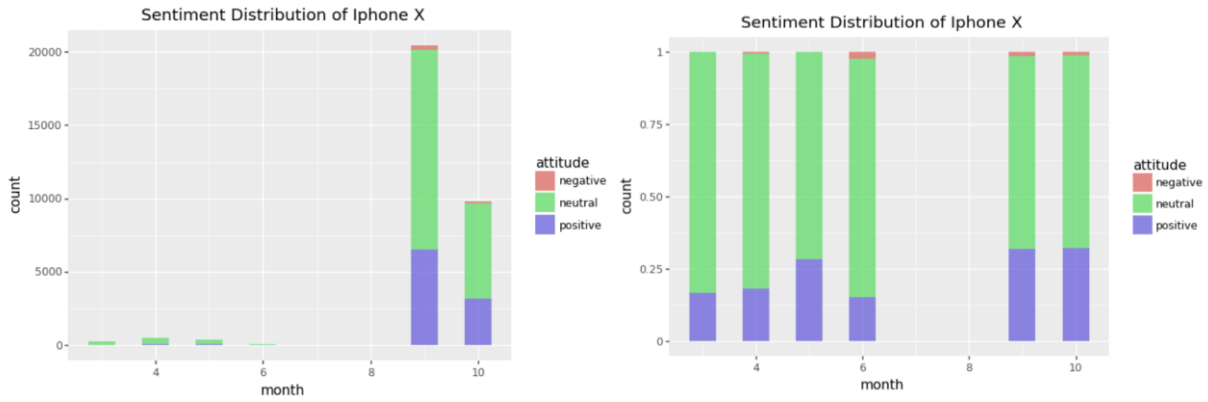
Then we breakdown the data and analyze the data by product in a monthly basis. For Samsung Galaxy S8, the attention dropped dramatically after 3 months of its launch, but raised a bit after iPhone launched its products. Consumers developed a gradual, though minor, negative attitude toward Galaxy S8. The positive attitude stayed constant. This demonstrates that people who were initially holding a neutral attitude changed their attitude into a negative one as they get to know more about the product.



For iPhone 8, it received exploded attention right before its launch. In terms of the attitude, consumers developed a positive attitude toward iPhone 8, though the negative attitude did not decrease. This demonstrates that only those people who were initially holding a neutral attitude towards iPhone 8 changed their thoughts as they know more about the product. Those who did not like the product in the first place, did not change their attitude.

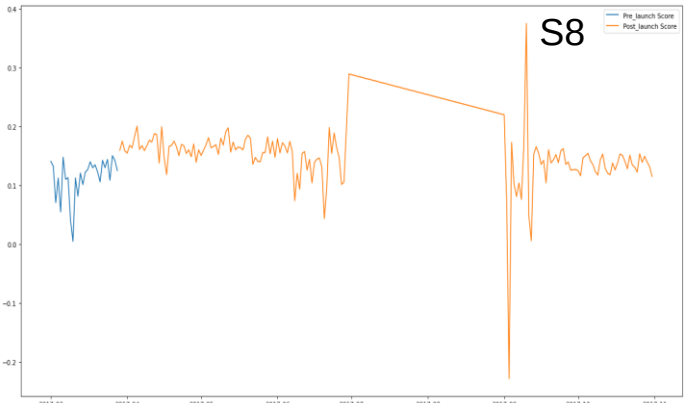


For iPhone X, it almost got no attention before the launch, and the reason behind this is because Apple did not release the name of this product. So, when we were filtering data under the key word “iPhone X”, we cannot detect many posts. One thing noteworthy is that the attention for iPhone X dropped dramatically after its launch, about 50%, compared to the other two models. This may due to the high suspension Apple set up for this product, which over-raised people’s discussion about the product, so it is understandable that the attention dropped. In terms of the percentage, we ignored the data before launch, because the size of the data is too small. The negative attitude decreased after the launch while the positive attitude and the neutral attitude stayed the same. This demonstrates that only those people who were initially holding a negative attitude towards iPhone X changed their thoughts to neutral as they know more about the product.

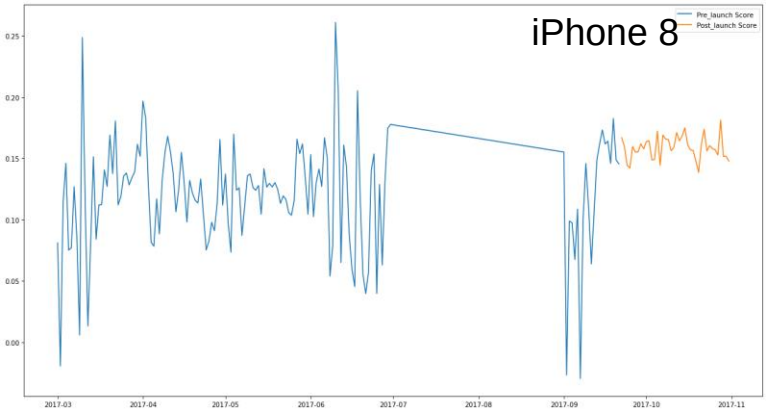


Then we conduct the sentiment change before and after the launch for each of the models. The blue line indicates the sentiment score for before launch. The orange line indicates the sentiment score for post launch. And the sentiment score is calculated by the mean of the score for that day.

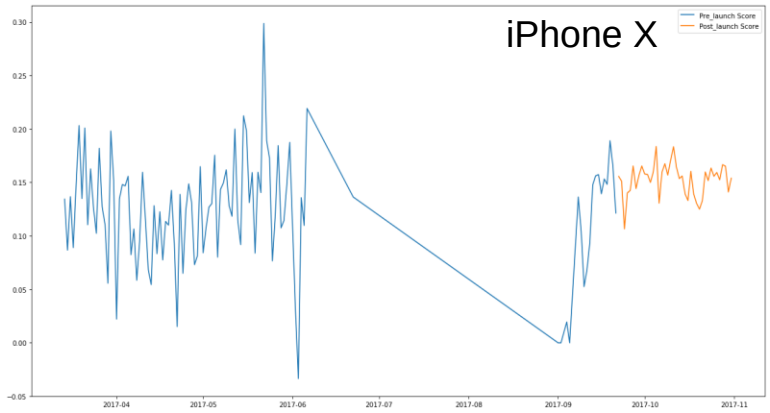
Overall, Samsung fluctuates more after its launch, while iPhone fluctuates rapidly before its launch. The sentiment scores for all three products rise after their launch.



For Galaxy 8, it has a relatively smooth graph compares to the iPhone products. One thing which is noteworthy is that the sentiment score dropped to negative rapidly before iPhone launched their products and then raised when they were launched. This may because people were overestimating the products which iPhone was going to launch, but when they got more information about the product, they felt a bit disappointed and turned to Galaxy 8.

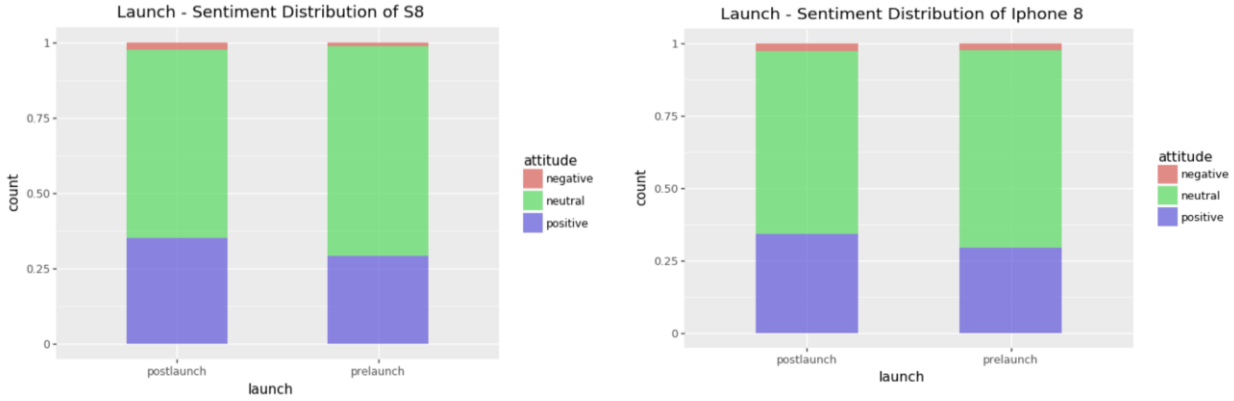


For iPhone 8, it fluctuates tremendously before the launch, and comes to a smoother pattern after the launch. Consumers felt highly about the product especially before Samsung launched its Galaxy 8. But then the score dropped dramatically right before they launched the iPhone 8 (but still above 0 unlike what happened for Galaxy 8). This may be due to misleading information sent out by the media, or other unreal assumptions made toward the product. Then the sentiment score raised steadily after the launch. This may due to the high quality of the product and the clarification of all the misleading information before the launch.



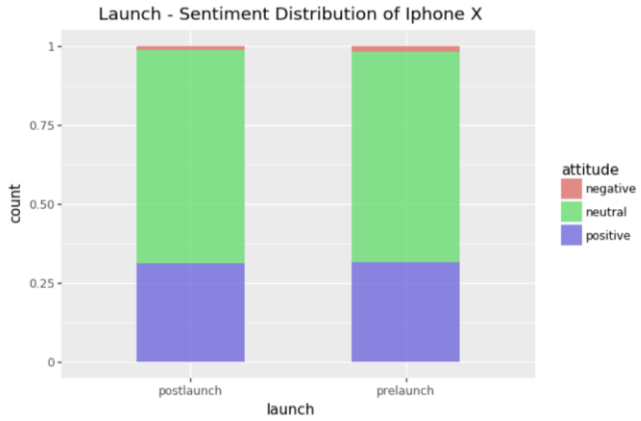
For iPhone X, it also fluctuates greatly before launch and the launch of Galaxy 8 did not have any impact on its sentiment score. This probably because at that time, consumers did not know the name of iPhone X, so they did not attach their emotions with the iPhone X. As the launch date approaching, consumers feeling toward the product went stable and things continued this way after the launch.

Besides analyzing the data from a chronical basis, we also generate insights from the light of percentage. Please note that all the left bars are POST launch, and all the right bars are PRE launch.



For Samsung Galaxy 8, the attitudes polarized after the launch—both the negative attitude and the positive attitude increased. This demonstrates that consumers had strong feeling about the new features attached this model. For iPhone 8, there is more positive attitude after the launch while the proportion of negative attitude stays the same. This indicates that iPhone 8 did not address the problem coming from consumers who did not like the product. But this product definitely established some features which the neutral, general consumers liked. For iPhone X,

there is less negative attitude after the launch while the positive attitude stays the same. This indicates that iPhone X did have features that were disruptive to the market and changed consumers' perspective towards the product.

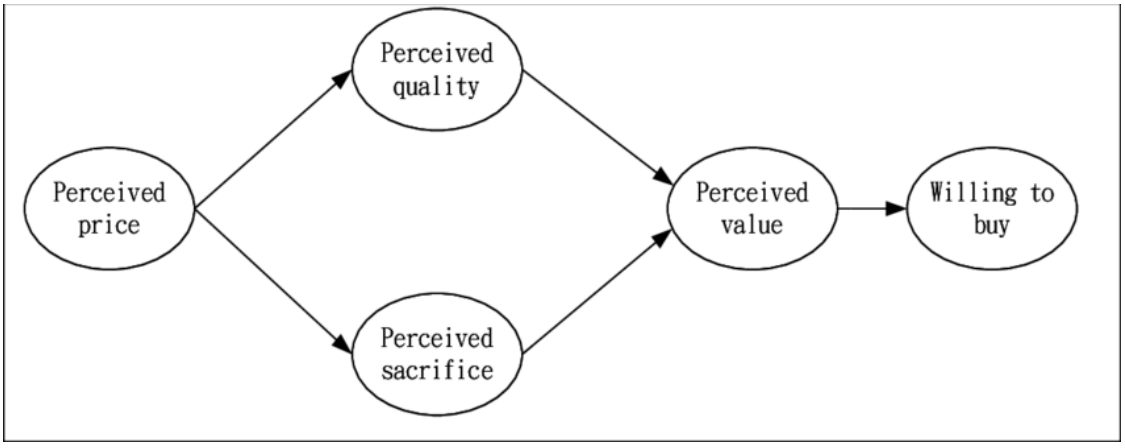


4.2 Recommendation:

Companies can consider launching two models in a same product line at the same time, so that consumers will unconsciously compare the two models, and the less well-performed model will automatically serve as an anchor which will drive consumers' expectation lower than they should.

5. Why is quality-price-value model important?

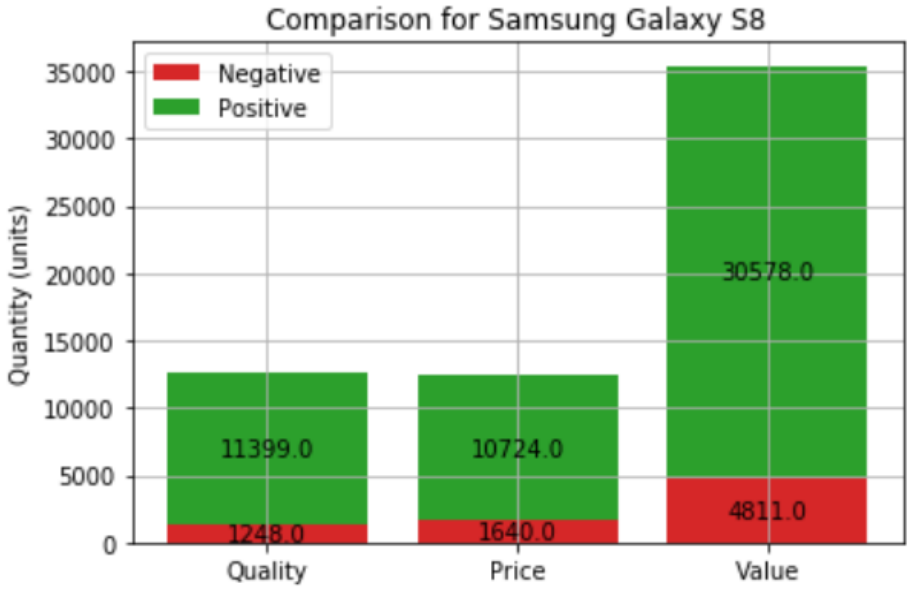
(Monroe & Krishnan, 1985)



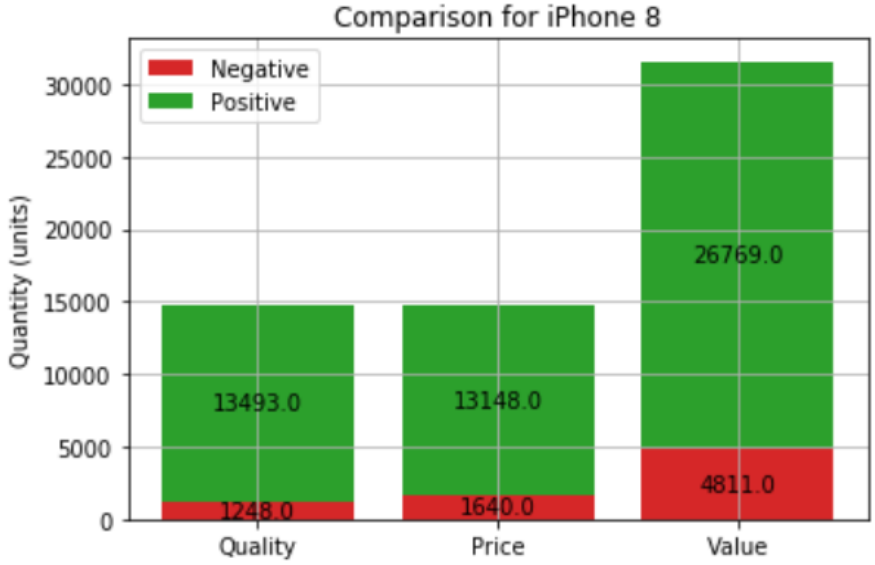
- Consumer perceptions of Price, Quality and Value are considered pivotal determinants of Consumer behavior.
- Buyers' perceptions of value represent a trade-off between the quality or benefits they perceive in the product relative to the sacrifice they perceive by paying the price:
- Perceived Value = (Perceived Benefits)/(Perceived Sacrifice).
-
- When the price is high, consumers perceive that the quality of the product is high

- Highly discounted products are perceived to be of lower quality.
- Willingness to buy can be determined from this model

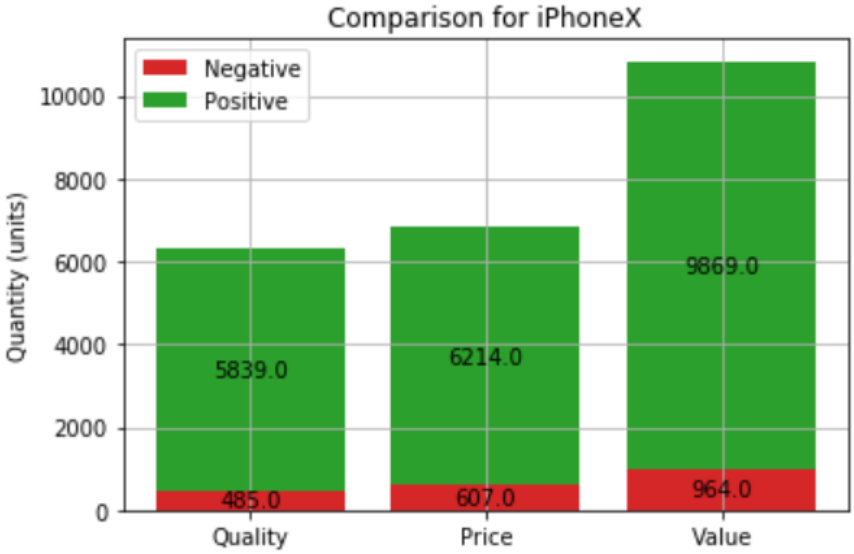
5.1 QPV model for Samsung Galaxy S8



5.2 QPV model for iPhone 8



5.3 QPV Model for iPhone X



We can see that for each product, consumers are talking more about value rather than price or quality.

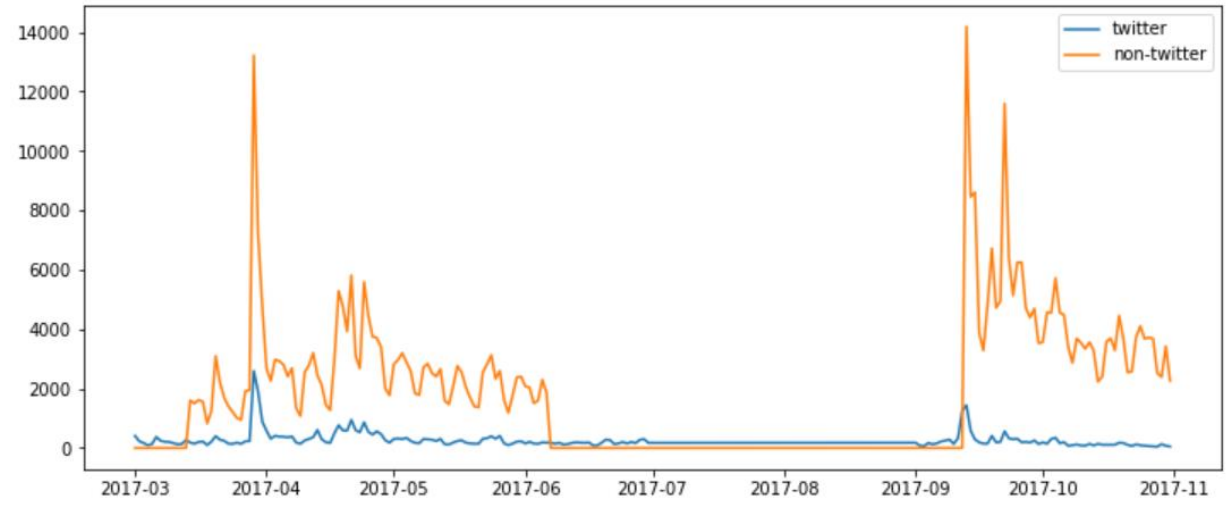
As seen from Monroe and Krishnan's QPV model explained above, the results of our analysis coincide with our expectations, that people want to see value in their investment.

Models iPhone 8 and iPhone X have a higher price as compared to Samsung Galaxy S8. Therefore, customers have a higher perceived value and quality of Apple products.

We can attest to this by observing that in case of iPhone, customers are talking relatively more about quality and price, compared to Samsung Galaxy S8.

Another observation is that value of iPhone 8 has more negative sentiment for value that iPhone X, which has the highest price in all three models. But consumers think that the value is worth the price.

6.1 Extra Credit – Twitter v/s Non-Twitter data

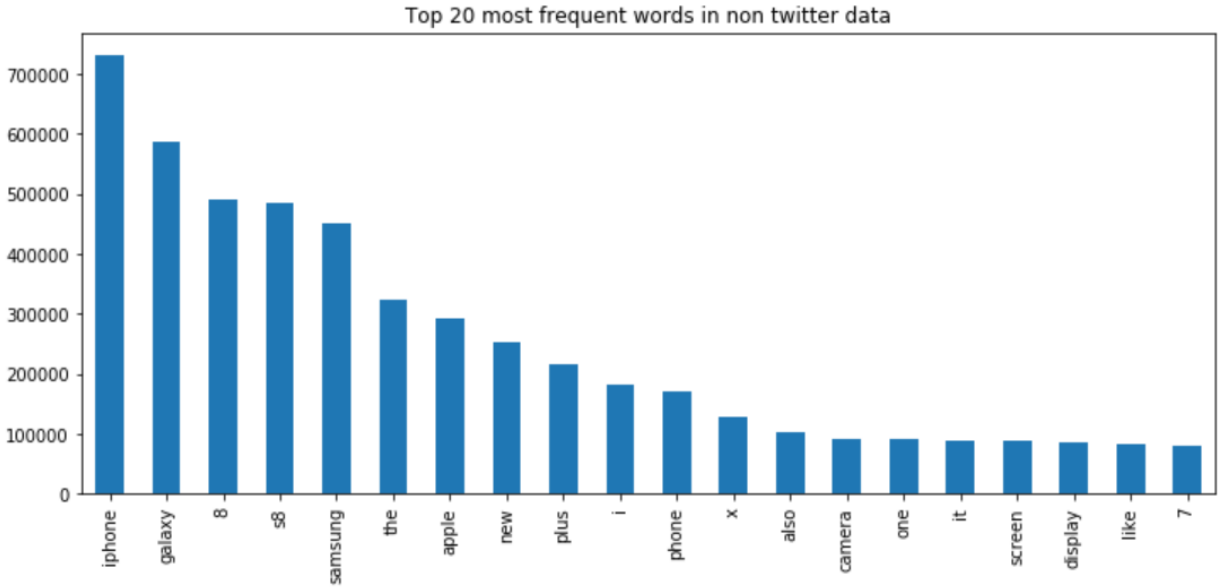


Visualizing number of posts overtime, for twitter and non-twitter data shows that customer engage more on non-twitter social media platforms for reviewing products.

The flat line between the 2 peak regions is the time for which we do not have data.

The peak regions surround the launch times of these phone models.

Looking closely at the data from twitter and non-Twitter platforms, we see the following top 20 words:



We can see certain differences in top 20 words observed in Twitter v/s top 20 words observed in non-twitter data. We can see that Twitter data gives a more general idea about the phone models, but other social media platforms give a more specific idea about the model features. This could be helpful in choosing a social media platform for both companies and Customers as:

1. Customers can choose where to look for reviews, for specific tagged reviews about a specific feature, they can go to social media platforms other that Twitter. For a general review, they can go to Twitter
2. Companies can choose platforms other that twitter to look for insights about specific features about their products and make decisions accordingly about how to proceed with the features
3. Companies can opt for targetted marketing by observing trends on various social media platforms.

6.2 Extra Credit

Based on the graph we got (the blue and orange line graph), we are not sure about whether people will buy the Samsung Galaxy 8 since there isn't a salient trend and consumers' attitude are quite ambiguous, and the post release line supports this argument. For the two Apple products, we predict a growth in the sentiment score, since the score rises dramatically right before the launch, and the post release line supports this argument.

Bibliography

Chaffey, D. (2019, February 12). *Global social media research summary 2019*. Retrieved from smart insights: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Monroe, & Krishnan. (1985). Retrieved from https://www.researchgate.net/figure/Figure-1-Perceived-value-model-Monroe-and-Krishnan-1985_fig1_304247936