**Web Analytics**

**Nintendo Switch**

**Product & Strategy Analysis**

**Group Project Report**

Ji Lyu

Chang Liu

Tianyi Hu

Ting Ding

Dec 9 2019

**Executive Summary**

Nintendo released its next generation game console Nintendo Switch in Mar 2017, almost 6 years after its previous Wii U. Fans and players were so eager to see the surprise with such a long-time preparation. With more news coming out, many people held negative attitudes on it. However, when Nintendo Switch finally came out, it received much more likes than dislikes. It successfully become one of the three most popular game consoles, together with Xbox and PlayStation.

This project aims to find out the key successful factors behind Nintendo Switch and make comparison with the other two consoles. Youtube and Twitter are the main data source for this project, as there is a large amount of social comment there. We crawled the Youtube comments and tweets with their locations using Python and some essential packages. Based on the data we crawled, we conducted sentiment analysis, unsupervised cluster, keywords analysis with Wordcloud and Tableau Map to reach the conclusion.

In general, Nintendo Switch did make a success and went on its unique style compared with the other two consoles. Yet there are still more to improve for its future product to keep its market.

1. **Business Goal Analysis**

The business goals of this project are just as mentioned in the project title. There are two main parts of the goals. One is the Product - Nintendo Switch, and the other one is the strategy for future generations of Nintendo Switch. We need to know people’s attitude towards Nintendo Switch and what are people talking about Nintendo Switch and other game consoles. We choose Youtube and Twitter as the main data source because the information there are from various of different groups including both players and viewers. The business goals are as follows:

**How do people like Nintendo Switch?**

This goal is aiming to find out the overall sentiment on Nintendo Switch. We will solve this problem mainly based on Youtube comments from Nintendo Switch Official Channel.

**What’s the Strategy for next generation of Nintendo Switch?**

Based on the previous one, this goal is aiming to analysis advantages, disadvantages and most attractive points of Nintendo Switch. We will solve this problem on both Youtube comments and tweets.

**How to carry on the Differentiation Strategy against PlayStation & Xbox?**

As we all know that PlayStation and Xbox are the two giants in game console markets. If Nintendo Switch wants to stand together with these two, it must deliver its own different characters to win part of the market. Thus, Nintendo should carry on a differentiation strategy to take its advantage. We will mostly focus on twitter data to solve this problem.

1. **Datasets Description**

In our project, Twitter and YouTube are two platforms that we collect user’s evaluation and location about Nintendo Switch. About Twitter, we mainly use Python package “tweepy.cursor” to crawl latest week’s tweets with specific keywords.

We use Twitter API to get permission of crawling and get user name, location, followers number and tweet from Twitter by using 12 different keywords – “nintendo Switch”, “switch price”, “ps4 price”, “xbox price”, “switch graphics”, “ps4 graphics”, “xbox graphics”, “switch game”, “ps4 game”, “xbox game”, “switch controller”, “ps4 controller”, “xbox controller”. Finally, we totally collect 30503 data. Following are one of our code with keyword “nintendo switch”.

On the other hand, we find the hottest video in Nintendo YouTube channel which first describe this NS product. Under this video, there are around 140 thousand comments. At first, we try to use Python package “selenium” to crawl these comments, but only nearly three thousand of them have been collected, which is not enough for us to analyze. Then we change to apply for YouTube Data API and mainly use Python package “googleapiclient” to crawl more data from the Internet. We request for the permission of YouTube first, insert this video’s ID to find it and crawl 100 comments at a time. After crawling process finished, we save them into a txt file. Finally, we totally crawl 73679 comments under this video. The detail of our code is shown below.

A screenshot of a cell phone

Description automatically generated

*Figure 1 Python Code for tweepy.cursor*







*Figure 2 Python Code for googleapiclient*

After preparing these two original files, we begin with the process of data cleaning.

For the YouTube part, because we want to analyze the sentiment of user’s comments, so we need to remain the words that show people’s sentiment like adjectives, nouns and verbs, instead of some meaningless words like conjunctions, pronouns and auxiliary verbs. So, we first change all text into lower case to make it normative and remove the stopwords which are listed in NLTK package, punctuations which we use in the daily life, then do lemmatization work to change all words into the original form. Following is our code and part of the result.

A screenshot of a cell phone

Description automatically generated

*Figure 3 Python Code for YouTube comments data cleaning*

A screenshot of a cell phone

Description automatically generated

*Figure 4 Output of YouTube comment after data cleaning*

Tweets have some differences from YouTube comments because tweets are web texts, which have html characters, emoji, URL and hyperlink. So, we also need to deal with these unrelated contents. Following is our code and part of the result.

A screenshot of a social media post

Description automatically generated

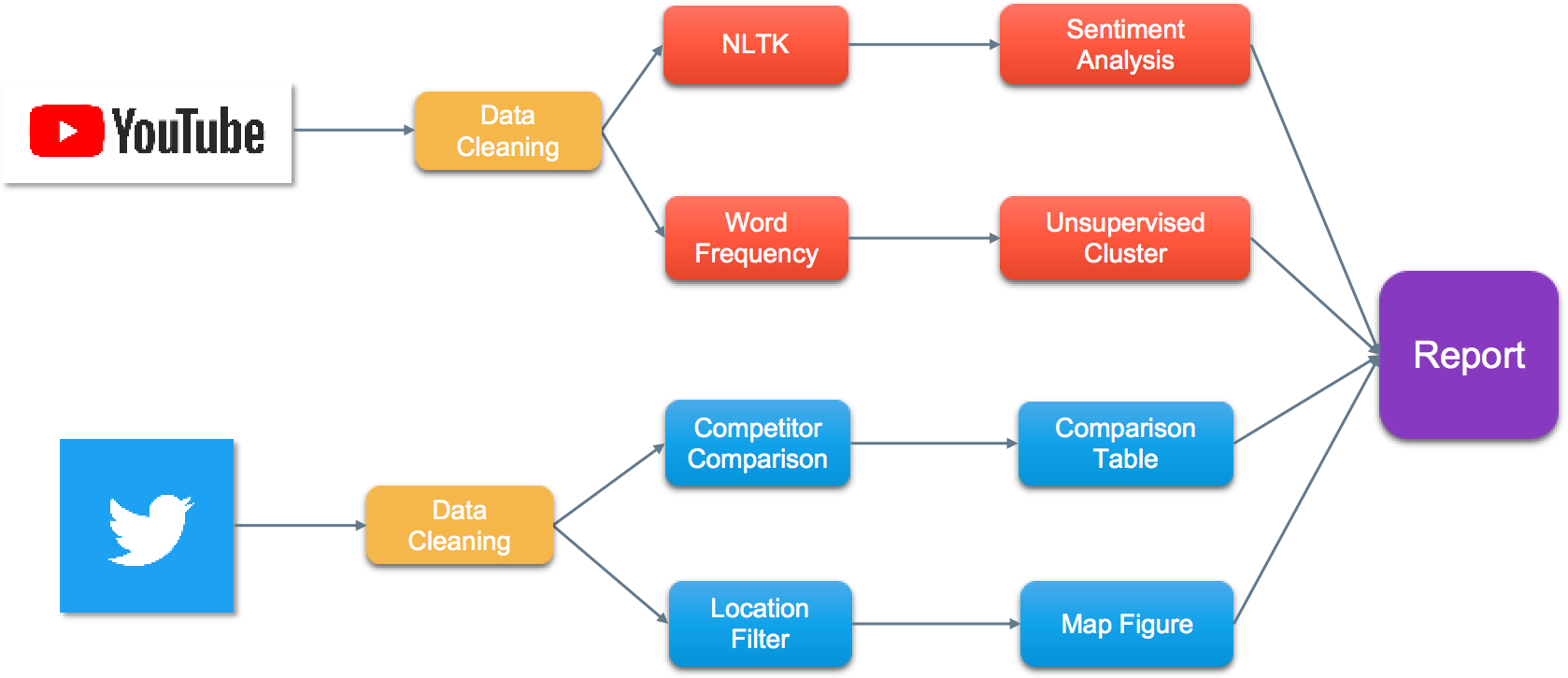
*Figure 5 Python Code for web content of Tweet data cleaning*

A close up of text on a black background

Description automatically generated

*Figure 6 Result after Tweet data cleaning*

1. **System Design (Methodology)**



*Figure 7 Methodology Map*

About the project system design, after collecting data from the Internet and finishing data cleaning process, we choose several analysis tools to analyze and present our results, such as Python, SPSS, Excel and Tableau.

We use Python package “NLTK” to collect word frequency list and sentiment analysis results including scores of positive, neutral, negative and compound. We plan to use visualization figures, pie chart and histogram to analyze the results. In another part, select SPSS to run the unsupervised clustering model. We want to understand user’s concerns about the NS console by clustering. We also focus on competitor comparison by using twitter data. We find most frequent words in each text and calculate their proportions in each text file to compare user’s different opinions. The next point we want to know is user distribution. We use the locations we get from Twitter and Tableau map tool to show the visualization results.

1. **System implementation & Evaluation**

**4.1 Statistics Analysis of Sentiment Scores**

After crawling more than 70,000 comments from YouTube, we applied NLTK to calculate the sentiment scores of each record. To understand these records more clearly, we made a statistical analysis of the comments and their sentiment scores. We only focused on the compound scores in this case. The results are shown below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Average | Mode | Median | Max | Min | Std |
| 0.1957 | 0 | 0 | 0.9983 | -0.9862 | 0.3678 |

From this table, We found the average value of sentiment scores is 0.1957, larger than 0, which means the whole attitude towards this video is positive, and most comments expressed their love for the first look at Nintendo Switch. The mode and median are both 0. This fact told us there were many comments having sentiment scores at 0, and we could also notice the information from behind figures. Besides, the maximum and minor values are very close to 1 and -1, but extreme values do not play an important role in statistics, so we do not talk about them in this case.

We used 0.2 as a limit to split the records and observe their distribution. Let us see the following bar chart.

A screenshot of a cell phone

Description generated with very high confidence

*Figure 8 Compound Sentiment Score Distribution*

From Figure 8, we could find most comments concentrate on the interval which is from 0~0.2, and it explains why the mode and median values are both 0. However, too many bars could not tell us the probable attitudes of these comments. We divided the records into three groups: -1.0~-0.3 as negative, -0.3~0.3 as neutral, and 0.3~1.0 as positive and generated a pie chart.

A picture containing screenshot, drawing

Description automatically generated

*Figure 9 Sentiment Distribution Pie Chart*

From Figure 9, we observe that only 7% of comments are in the negative group, and the largest one is the neutral group since more than 40,000 comments are in 0~0.2 from Figure 3.1. The positive group still occupies 39% of all the comments in such a case, which means commenters have an optimistic attitude towards Nintendo Switch.

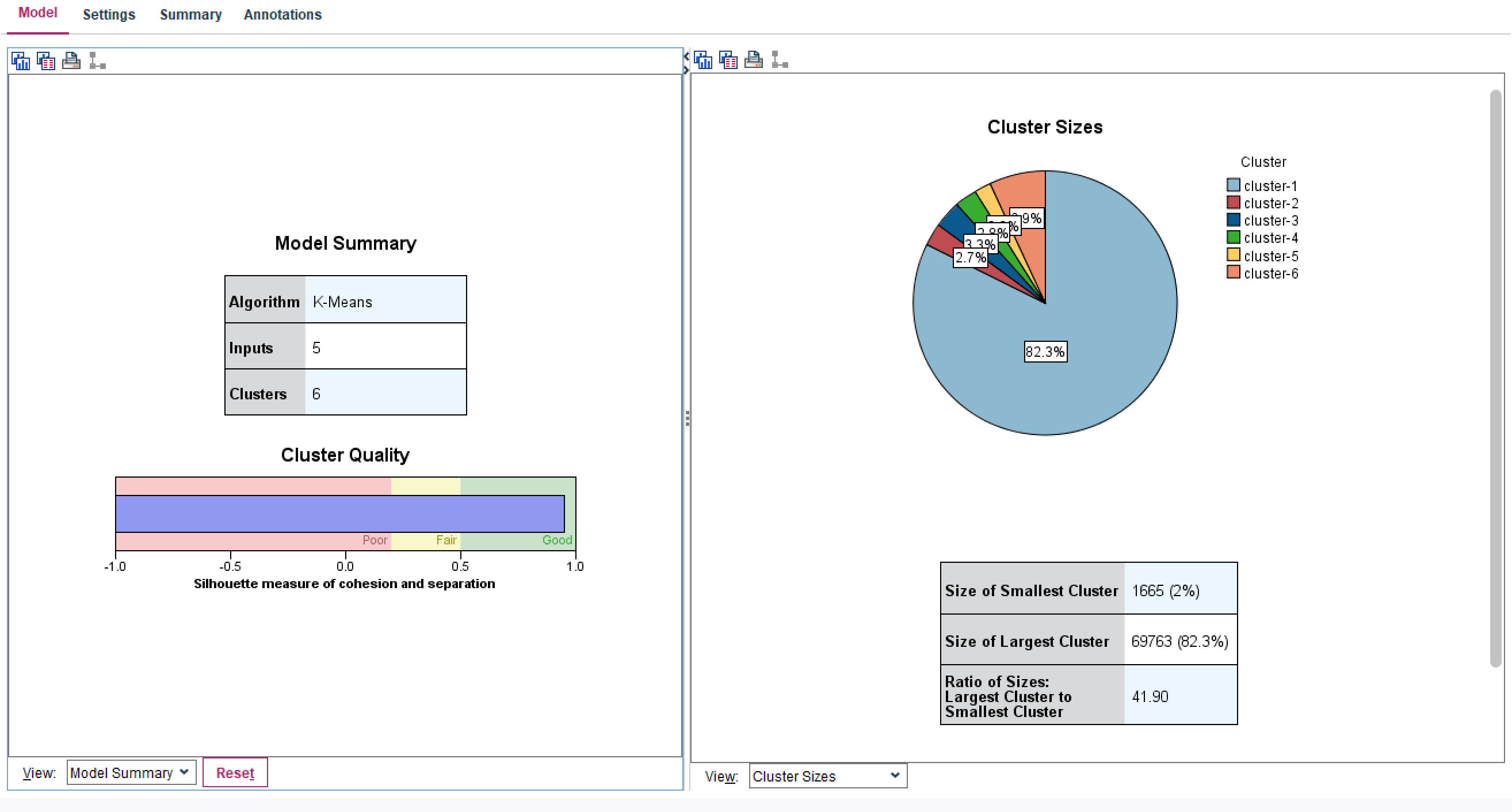
**4.2 Data Mining of Word Frequency**

After sentiment scores, we would focus on the content itself of comments and calculate the frequencies of words in records, then we chose 5 important adjectives and verbs to calculate their frequencies in each record and generated a dataset. The code shows below.



*Figure 10 Python code for Word Frequency*

Since we did not know the relationships and rules among the data, we used an unsupervised data mining method: K-means Cluster. The next question was to decide the K. Since we only chose 5 words from more than 70,000 records, we could expect that the largest cluster must be a cluster in which all the frequencies of chosen words are 0, and the best K could make the largest cluster as small as possible. After several times of trial, we found the best K is 6.



*Figure 11 Cluster Result from SPSS Modeler*

From Figure 11, we found the average silhouette of this model is about 1.0, which means every record is well matched to the cluster it assigned to and weak matched to neighboring clusters. The clustering configuration of this model is appropriate.

On the other hand, we cared about the clusters themselves and created a table of clusters information.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Cluster1 | Cluster2 | Cluster3 | Cluster4 | Cluster5 | Cluster6 |
| Size | 82.3% | 2.7% | 3.3% | 2.8% | 2.0% | 6.9% |
| Most Frequent Values | | | | | | |
| new | 0 | 0 | 1 | 0 | 0 | 0 |
| like | 0 | 0 | 0 | 0 | 0 | 1 |
| want | 0 | 1 | 0 | 0 | 0 | 0 |
| buy | 0 | 0 | 0 | 1 | 0 | 0 |
| wait | 0 | 0 | 0 | 0 | 1 | 0 |

We went back to the content of the cleaned comments and found which cluster they were assigned to. First, we found the comments in the second largest cluster, Cluster6, in which most frequent value of like is 1, are such as ‘It seem like good console’ or ‘I love switch guy play switch nonstop like switch dog’ (The comments may look strange since we have cleaned them). These records are very likely to have a positive attitude towards Nintendo Switch.

Also, comments in the third-largest cluster, Cluster3, are such as ‘I believe 3 years since Nintendo show us new console.’ or ‘Wow! Where I buy new console?!’. The most frequent value of new is 1 in Cluster3, and we found these commenters are excited about the new console, Nintendo Switch, after such a long time that Nintendo has not launched a new product.

Besides, comments in Cluster4, in which most frequent value of buy is 1, are such as ‘Hmm look interest might buy,’ ‘ah yes, first buy console nintendo’ and ‘buy mine 2 days ago ... love already!’. These commenters look like existed and potential customers for Nintendo.

Finally, we used K-means Cluster to get separated groups of comments and understand what each group is caring about. Nintendo could divide those commenters into different fields and design different marketing strategies for them. However, from this model, the information may be not enough, and further studies are needed.

**4.3 Wordclouds of Nintendo, PlayStation and Xbox**

As mentioned above, we crawled 30,503 tweets and cleaned them. These tweets fall into 12 data sets, including 4 aspects of 3 different game consoles. The 4 aspects are ‘controller’, ‘games’, ‘price’ and ‘graphic’. The 3 game consoles are ‘Nintendo Switch’, ‘PlayStation’ and ‘Xbox’.

With the help of NLTK, we can calculate the most frequent words to build the word clouds, so we have 12 Wordclouds, below are three of them. For more details, we will do the analysis in a comparison table.

A close up of a newspaper

Description automatically generated

*Figure 12 3 of the 12 Wordclouds*

From the comparison table below, we can see there are 3 products in the columns, Nintendo Switch, PlayStation and Xbox. Each product has 4 aspects, ‘controller,’ ‘games’, ‘price’, and ‘graphic’. In this report, we will elaborate some keywords.

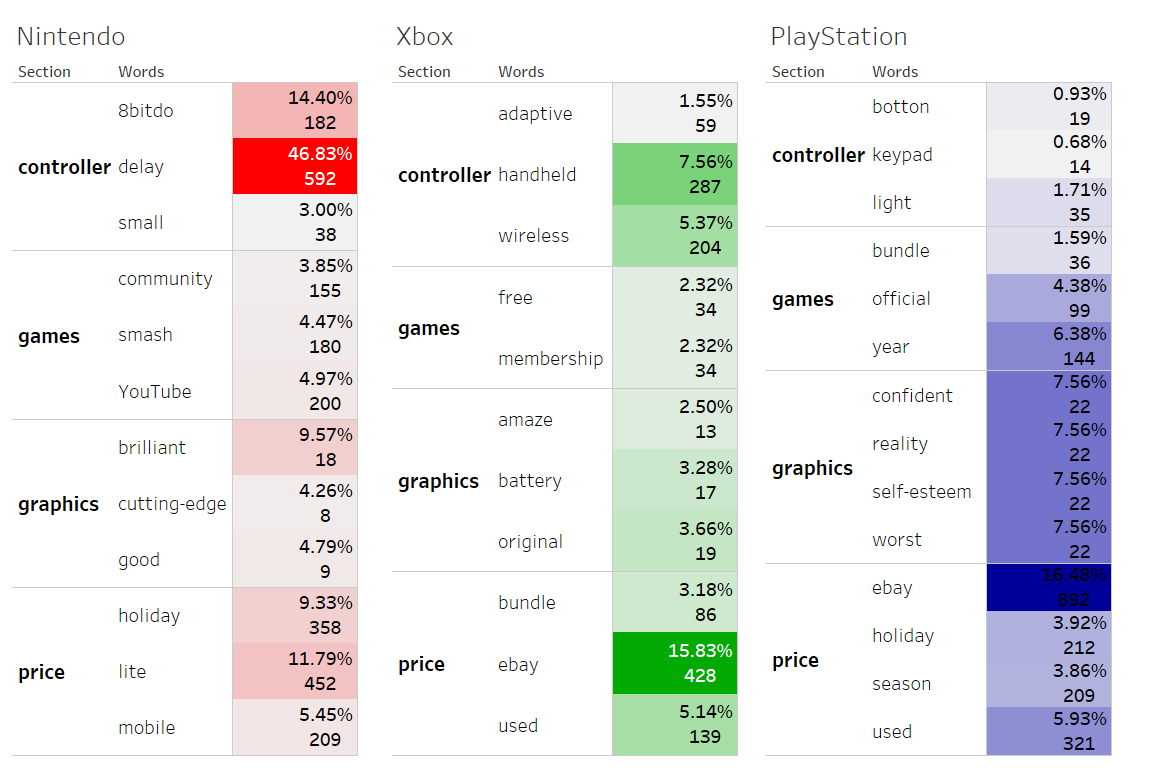
We will check ‘Nintendo Switch’ column first. In 1,264 tweets mentioned ‘switch controller’, the word ‘delay’ appeared 592 times, so obviously there is a problem about the controller quality. Also, 38 tweets mentioned ‘small’. We can say that Nintendo may need to improve its controller, no matter in size or in quality.

Next, we will turn to the games in switch, in 4,024 tweets that mentioned ‘switch games’, the word ‘YouTube’ appeared 200 times and the word ‘community’ appeared 155 times. So, we can see the players really care about the interactions between each other. Based on this, Nintendo can unit some YouTubers to play their games and upload videos. Also, Nintendo can organize some activities in the community to maintain the customer stickiness.

Then we come to price. In 3,834 tweets mentioned ‘switch price’, the word ‘holiday’ appeared 358 times, so we can see that switch is really a good present for family and friends. Also, CNN news reported on December 2, 2019 said that ‘Nintendo Switch Was Black Friday’s Big Winner’. So, we can see how popular the switch is. To keep this popularity, Nintendo may need to provide some coupons or discounts when holiday.

Although we have data set of graphics, the number of tweets about this topic is just 188, we barely can draw any conclusion based on this, maybe with more time, we can crawl more tweets and do the analysis. This time, we will just skip this and just do the comparison.

|  |  |  |  |
| --- | --- | --- | --- |
| Word Frequency | Nintendo Switch | Xbox | PlayStation |
| Controller | 1264 | 3798 | 2044 |
| Games | 4024 | 1464 | 2258 |
| Price | 3834 | 2704 | 5413 |
| Graphics | 188 | 519 | 291 |



*Figure 13 Matrix of Keywords*

Comments about controllers of PlayStation and Xbox are good. Users believe them are ‘light’ and ‘handheld’, so Nintendo can learn from the competitors and do some improvements.

Next, we focus on ‘price’. In tweets about PlayStation and Xbox, many mentioned ‘Ebay’ and ‘Used’. This may because people are selling their used game consoles. Also, since PlayStation and Xbox have been released for many years, we can say that users may get a little bit tired of the old consoles. So, Nintendo can take this opportunity and expand market to explore the new users.

Although we crawled many tweets, due to the drawback of technology and limit of time, we don’t have many information about some certain aspects. Given more time, we can crawl tweets when PlayStation and Xbox just released, maybe the data will be balanced.

**4.4 Tweet Heat Map of US and World**

Based on the location data of tweets, which focused on keywords of ‘Nintendo Switch’, we used Tableau to draw heat maps of US and the whole world. Before drawing the map, we removed those meaningless location data, such as ‘None’, ‘On Earth’, ‘Behind you’, etc.

From the Heat Map of US, we could see that California has the most heat discussion about Nintendo Switch on twitter. New York goes right behind and Texas and Florida rank the third together. The heat map could provide useful marketing information that helps Nintendo plan its future market campaign. For example, to set up the next Release Event for next generation of Nintendo Switch in Los Angeles may have a much better result than other States.

A picture containing toy, black

Description automatically generated

*Figure 14 Heat Map of US*

Then we take a further look at the Heat Map of the World. Besides US, the United Kingdom is another heat point for Nintendo Switch, despite the small population. On the other hand, Canada and Brazil also bring heat discussions on Nintendo Switch. As we all know that the initial release country set of the product is quite essential. This heat map offers useful information for how to choose the initial release countries.

It might be strange that Japan on this heat map is quite cool for the discussion. The reason is that the Unicode and language we filtered for the crawling. In this way, Twitter in Japanese may be filtered off. Meanwhile, we have to say that twitter is banned in China, so we have almost no result from China. There are indeed a great many Nintendo fans and discussion there, and the very Official Chinese Version of Nintendo Switch has just been released on Dec 4.

This worldwide heat map may not cover all the countries, but it will still help interpret the discussion distribution in the world.

A picture containing map, food

Description automatically generated

*Figure 15 Heat Map of World*

1. **Conclusion and Future Direction**

According to all the analysis and result, we can make conclusions to interpret our business goals:

* Most people hold neutral attitude towards Nintendo Switch and the general attitude is positive.
* For next generation of Nintendo Switch, Nintendo should carefully choose its initial release region (North America, UK, Brazil must be included) and run effective market campaign in US states (California, New York, Texas, Florida are ideal choices). Also, Nintendo must update its controller.
* Nintendo can cooperate with or support Youtubers to make more game streaming or videos, together with its own unique games like *Mario*. Meanwhile, to build up and maintain the community for players is necessary. Nintendo Switch is a mobile like portable console, much different from the other two consoles. Different and proper iteration pace plays important role.

For future direction and improvement, we still need to consider:

* Choose the time period after each console’s release rather than all the same period.
* Balance data amount for each console to avoid bias.
* Try more Unicode for twitter data to cover most of the languages.
* Crawl more comments from more videos of different time to improve data size.