“Squid Game: Why Are People Into Squid Game?”

Bolortungalag Enkhjargal

Data Science
Fall 2021

Abstract
According to Netflix's announcement, 142 million households watched the Korean Netflix original series 'Squid Game' during its first 28 days (Tassi, 2021) since its release. Also, last October, it was listed in the top one global rankings in the Netflix TV show category (Lee, 2021). Due to this international phenomenon, I was curious why 'Squid Game' is so popular globally. Therefore, I decided to analyze Twitter's tweets that mentioned 'Squid Game.' I used RStudio to explore the tweets' content to know what factors of 'Squid Game' make people into it.

Data collection
I decided to use a database from the Squid Game series by collecting Twitter information and filtering all of the information by ourselves. However, there is too much information about users, locations, and other data which was related to users' feelings and opinions towards Squid Game. Hence, I changed using a completely collected database of Twitter. The database called 'Kaggle' is a legal source that people were commenting about Squid Game on Twitter in terms of CSV files.

Models
Before I started coding, I first installed required packages: readr, readxl, dplyr, tidyr, ggplot2, tm, tidytext, wordcloud2, widyr, jiebaR, topicmodels, scales, forcats, janitor. Then, before importing the data, I utilized all of the databases that I found in Kaggle by setting a working directory and leading the CSV files with "Squid Games.csv". In part of the data inside the column, I emphasized "date" and "text" and filtered the date year in 2021.

Moreover, in order to analyze the data, I have to clean the data first. However, it is unnecessary to clean all the information from Kaggle. Thus, I picked only a few columns in which I think it will be useful. Since I want to analyze why people are addicted to Squid Game, I only picked text and date columns from the source. Nevertheless, after I cleaned all the articles and the symbols, there are still a lot of not useful words; thus, I use the stopword function to clean more. I manually put all the words that are not necessary to the note and named it as "stopword". Then, I created a new table called "Squid Token", which contained the cleaned data of ID, Text, and Date columns. From this step on, I could analyze the Squid Game data now.

In part of doing a word cloud, I selected 200 words (N=200) out of the whole information which social media users mostly mentioned.

Then, I decided to do the TF-IDF, Term Frequency–Inverse Document Frequency, in which I will use this method to statistically calculate which word is the most important from the Twitter posts that I collected in the specific time. In this project, I initially wanted to select the time before the Halloween period since after that most people will mainly talk about the costume which could affect our result. Luckily, the information that Kaggle has is before Halloween; thus, I did not have to work on filtering the date. As I know the higher the number of the TF-IDF, the more important the word; in this case, the highest number of the TF-IDF is 11.20282.

After doing the TF-IDF, I also did the sentimental analysis in order to let R computationally identify and categorize opinions expressed in a piece of text to determine the writer's attitude

towards a particular topic: positive, negative, or neutral. I also applied this method in order to analyze the tone of Twitter users toward the Squid Game topic. Also, the highest is 4 and the lowest is minus 5.

In part of the Linear discriminant analysis or LDA, I assigned variables of “Squid_dtm” from a data source of “squid_tokens”. I planned to separate all 200 words into 6 bar graphs (k = 6), each graph would only mention the ten most popular words (n = 10). Apart from this, I used a template of LDA codes to execute and implement all of the databases.

Even though I used an LDA function which is able to categorize into 6 graphs to show how 10 words in each graph are similar to each other, I didn’t use this information to analyze it. Although it was able to show 10 significant words, it didn’t mean that most users were commenting and getting into that.

I also did use the correlation method in order to use the statistic to test how large a word relates to the other word in the Twitter document. In this case, I used the code that lets us identify the word I want since it is impossible to let R-studio run all the words inside the collected tweets. Also, I picked three words, at first, yet I could not see any relationship between them at all; I estimated that the problem might occur since I was not the ones who collected the data by ourselves, so the collected data might not be that accurate.

Results

The Netflix series called ‘Squid Game’ has attracted global attention, therefore, the research will find the reason why it was a huge success. In order to analyze the data, first I cleaned our data by removing symbols and articles. The functions I used are shown in (Fig.1).

```
#adding the id column
squid_data$id <- 1:nrow(squid_data)

squid_text <- function(text){
  # Remove "amp;", and "\n"" wich are not part of the tweet
  text <- gsub("amp;", "", text)
  text <- gsub("\\n", "", text)

  text <- gsub("^\\s+", "", text)
  text <- gsub("\\s+$", "", text)
  text <- gsub("[ |\t]+", " ", text)
  # Eliminate URLs
  text <- gsub(" ?(f|ht)(tp)(s?)(://)(.*)[.|/](.*)", " ", text)
  text <- gsub("https?://.+", "", text)
  # Eliminate unnecessary space first
  text <- gsub(' {2,}', ' ', text)
  # tweets.dat$text <- gsub("\\s+", " ", tweets.dat$text )
  # Remove extra characters
  text <- gsub("[^[:alpha:][:space:]]*", "", text)

  text <- gsub("\\d+\\w*\\d*", "", text)
  text <- gsub("[^\x01-\x7F]", "", text)
  text <- gsub("[[:punct:]]", " ", text)


  # Remove references to other twitter users
  #text <- gsub("@\\w+", "", text)
  # Convert to lower case
  text <- sapply(text, tolower)
}
```

Fig.1

After the first step, there were still unwanted words and phrases, thus, I used the ‘stopword’ function and made a new table called ‘Squid Token’ (Fig.2). This included the cleaned data of ID, Text, and Date columns.

```
squid_tokens <- squid_data %>%
  unnest_tokens(word, text) %>%
  filter(!word %in% stopwords("english")) %>%
  filter(!word %in% c(stopwords2$a)) %>%
  filter(!word %in% c("https", "t.co", "rt"))
```

Fig.2

I created a visualization (Fig.3) to see the most frequently appearing terms to the less frequently appearing within 200 words.

Fig.3

Following, here are the results with the highest TF-IDF (Fig.4). I manually filtered 300 words with the highest TF-IDF to get the 50 best words to interpret their relationship.

	id	word	n	tf	idf	tf_idf
1	18002	light	10	0.5000000	4.004637	2.0023185
2	46372	mfs	8	0.8888889	7.801623	6.9347761
3	14062	candy	7	0.5833333	5.734760	3.3452769
4	66992	ot	7	0.5384615	9.123379	4.9125887
5	3017	ha	6	0.4615385	6.539381	3.0181760
6	396	play	5	0.3333333	3.898977	1.2996591
7	15606	please	5	0.5000000	4.727388	2.3636939
8	18002	green	5	0.2500000	4.048987	1.0122467
9	18002	red	5	0.2500000	3.823812	0.9559531
10	33969	mug	5	0.4166667	7.983945	3.3266436
11	1654	bts	4	0.5000000	6.096875	3.0484375

Fig.4

I found the writer's attitude towards a particular topic is positive (on left), negative (on right), or neutral from sentimental analysis (Fig.5).

Fig.5

Lastly, here is the result of the correlation method (Fig.6) that I used the statistic to test how big is the relationship of the words between each other.

	item1	item2	correlation
1	colore	squirty	1
2	squirty	colore	1
3	malek	rami	1
4	rami	malek	1
5	dxdonmega	amiontheair	1
6	amiontheair	dxdonmega	1
7	ashaangi	usopen	1
8	usopen	ashaangi	1
9	tahun	offernetflix	1
10	aplikasi	offernetflix	1
11	offernetflix	tahun	1
12	aplikasi	tahun	1
13	blnready	sharingfree	1
14	sharingfree	blnready	1
15	offernetflix	aplikasi	1
16	tahun	aplikasi	1
17	vuconcept	wowcdtdeja	1
18	wowcdtdeja	vuconcept	1
19	gregorelli	rosmello	0.99457812
20	rosmello	gregorelli	0.99457812
21	prelemi	rosmello	0.99456352

Fig.6

Insights: **Why are people into Squid Game?**

A. Marathon viewing

Into can be defined as strongly interested in or involved with something (“into”, Cambridge Dictionary). Meanwhile, there are two perspectives to explain this behavior: (1) addictive to watch, and (2) addictive to talk about it. One behavior that can define the first perspective is “marathon viewing”. Psychologically, marathon viewing is a situation in which people consume –which in this case is a movie– in excessive amounts in a short period of time in order to escape reality and elicit psychological comfort (Greene & Maggs, 2017). The word “extraction” with a high value of TF-IDF ‘shows how many people want to dig deeper into the story, “ensnare” shows how the people are trapped with the story, leading to addiction to

watching more. The storyline of the Squid game that is tied to social reality can be a source of learning for people. Presented by the character emotions and the storyline –survival– wherein the real world I also have to struggle to live in an increasingly competitive world, filled with various religious human emotions ranging from cunning, doubt, fear, to help each other, all of which can be seen directly by the audience and a lot of life lessons can be taken throughout the scenes. Thus an explanation can answer why people become "extract" and further "ensnare" with the storyline of Squid Game.

Other studies found that people are motivated to watch more and more as an: (1)enjoyment, (2)recommendation of others, (3)perceived control –control of social media– , and (4)fandom (Shim & Kim, 2018). The game itself which represents the movie the most creates an extravagant and "hyperventilating" sensation, and this is what makes us like and 'enjoy' the movie more. The games inside of squid games present a thrilling sensation mainly because of the cruelty factor. Studies found that people like to watch and can enjoy watching media violence (Weaver & Kobach, 2012). One explanation for this is that it may take viewers to meet their social norms expectations due to morbid curiosity –curiosity of negative topics– (Goldstein & Zillmann, 1998).

B. Social Impact.

The other points can help to clarify the second point, "why do people talk about it?" According to the research of Tukachinsky and Eyal (2018), media marathoning can lead to more profound types of media involvement. These points show how to develop social relationships. Thanks to the advancement of technology and social media, it can be utilized as a highly open location for them to express their sentiments, as well as a venue for sharing viewpoints. According to the findings, people talk a lot about social media platforms like Tik-Tok and YouTube, and a lot of creativity is generated by the general population, and these ideas and inventions emerge from the scenes and characters in this series. For example, "memes", many people create memes and share them on social media; in the future, this type of information serves as inspiration for others, piques people's interest, and even inspires others to be more creative. As a result of this social action, a trend has emerged in which many people discuss how they may incorporate this content into their social lives. The reason for all of this is that individuals want to have control over social media – or at least perceive to have control. They cannot resist the emergence of a trend, even if they do not want to follow it.

Conclusion

While analyzing Twitter’s tweets about ‘Squid Game,’ I was able to draw three significant conclusions on why ‘Squid Game’ is popular. First, ‘Squid Game’ is popular because of its interior story plot. When analyzing, it can be seen that many words related to the game in ‘Squid Game,’ such as “licky” and “loafing” are mentioned. Through this, it could be analyzed that the impressive games in the drama itself aroused interest in viewers. When analyzing the emotional sentiments on tweets, I could see how viewers feel about ‘Squid Game’ through words such as “awesome” “fantastic” and “insane.” Also, according to Kim’s research, anyone in any country has more accessible access to topics familiar to them (Kim, 2009). The main topics covered in the ‘Squid Game’ are survival and capitalism. Survival is one of the natures that humans naturally have, and capitalism is the aspect of the world I

currently live in. People were attracted to topics familiar with them, and through this, I could see that ‘Squid Game’ became more popular.

Second, when I analyzed Twitter’s tweets, I could see that Korean words in ‘Squid Game’ are mentioned in the original form. For example, the words ‘Hiong/Hyung (Brother)’, ‘Dalgona (Korean Traditional Candy),’ ‘Kkanbu (Old Friends),’ and other characters were mentioned in Korean pronunciation. It is analyzed that the content that utilizes Korea’s unique culture has come fresh to foreigners. Korea’s unique culture is mentioned a lot in this drama. Still, at the same time, ‘survival,’ the story’s main concept, is a subject that humans can naturally sympathize with regardless of country. This combination of familiar topics and unfamiliar materials can be thought of as a factor in the popularity of the ‘Squid Game.’

Lastly, ‘Squid Game’ has gained a lot of approval for its own characteristics. Still, since then, it has become more popular as it has been actively mentioned on various social media. Through coding, I could see that multiple social media, namely YouTube, TikTok, and Twitter, were actively mentioned. For example, TikTok also created a ‘Squid Game Challenge’ that utilizes one of the various games in the drama. There was also an Internet game website that imitated ‘Squid Game’ (http://mydirect.yiv.com/Squid-Challenge?gclid=CjwKCAiA866PBhAYEiwANkIneDOsquusys4Zy4VP8lSJkU_WJLYQUKXnGTLM6vAa-8RAydW6XuHGjxoCqvIQAvD_BwE).
While analyzing, I could see that unexpected words were often mentioned. It was the NFT. NFT stands for the non-fungible token, which basically means that it’s a one-of-a-kind digital asset that belongs to you and you only. People were also using the popularity of ‘Squid Game’ as one of the elements of virtual transactions. These social impacts helped ‘Squid Game’ increase its popularity even more. People became more enthusiastic about ‘Squid Game’ to communicate with others and gain capital gains.

References

Kaggle data source:
https://www.kaggle.com/datasets/deepcontractor/squid-game-netflix-twitter-data

Goldstein, J. H., & Zillmann, D. (1998). *Why we watch: The attractions of violent entertainment*. Oxford University Press on Demand.

Greene, K. M., & Maggs, J. L. (2017). Academic time during college: Associations with mood, tiredness, and binge drinking across days and semesters. *Journal of Adolescence*, *56*, 24-33. https://doi.org/10.1016/j.adolescence.2016.12.001

Into. (n.d.). In *Cambridge dictionary | English dictionary, translations & thesaurus*. https://dictionary.cambridge.org/dictionary/english/into

Kim, G. (2009). In Search of Blockbuster Drama: The Effects of Genre Familiarity, Star Power, and Internet Accounts on Rating Performance. *Korean Strategic Management Society, 12(1),* 25-54.

Lee, H. (2021). *Top 10 Korean dramas on Netflix. . .The heyday of K-content*. Etoday. https://www.etoday.co.kr/news/view/2072396

Shim, H., & Kim, K. J. (2018). An exploration of the motivations for binge-watching and the role of individual differences. *Computers in Human Behavior*, *82*, 94-100. https://doi.org/10.1016/j.chb.2017.12.032

Tassi, P. (2021). *Squid Game's New Netflix Viewer Numbers Seem Almost Impossible*. Forbes.
https://www.forbes.com/sites/paultassi/2021/10/20/squid-games-new-netflix-viewer-numbers-seem-almost-impossible/?sh=14f53fdb5506

Tukachinsky, R., & Eyal, K. (2018). The psychology of marathon television viewing: Antecedents and viewer involvement. *Mass Communication and Society*, *21*(3), 275-295. https://doi.org/10.1080/15205436.2017.1422765

Weaver, A. J., & Kobach, M. J. (2012). The relationship between selective exposure and the enjoyment of television violence. *Aggressive Behavior*, *38*(2), 175-184. https://doi.org/10.1002/ab.21417