Super Bowl Ads Exploratory Data Analysis Report

Ariel Wang, Rheon Jin

October 14, 2022

Abstract

We address the question of how the commercials grouped by brands can receive the best responses. We inspect the data from ten brands that had the most ads in Super Bowls from 2000 through 2020. We apply three matrices to select the most responsive brands, do transformation on variables and linear regression between variables, and perform text analysis on their descriptions. NFL and Bud Light are found to be the two best responsive brands. It turns out that including entertainment, celebrity, danger or patriotism element, and brand's product related information in the description and the ads' content could most effectively help advertisers to increase their responses to a large extent.

1 Introduction

Super Bowl is the annual American football finals of the National Football League (NFL). Its live streaming is the most watched television program in the U.S., and therefore attracts a large number of commercial advertisements at huge costs. In this report, we will analyze the data related to the advertisements for the 10 brands that aired the most spots in all 21 Super Bowls this century, to get insights about how future commercials could get better responses on the Super Bowl. Specifically, the insights could be concluded by deep diving into the following questions:

1. Which brands receive the best responses from their Super Bowl advertisements?
2. Looking into these best responses ads' brands, what major categories are they identified as? How do their characteristics, such as funny, patriotic, etc., change throughout the years?
3. What do all these commercials' descriptions look like on Youtube? How do the most responsive brands link their products to the Super Bowl in their descriptions?

2 Data

The dataset used in this research includes 89 advertisements from the ten brands that had the most ads in Super Bowls from 2000 through 2020, based on the statistics from superbowl-ads.com. There are 25 variables, including the statistics of ads on youtube and 7 defining characteristics of each Superbowl ads. The reader can refer to FiveThirtyEight for characteristics' evaluation taxonomy. Note that in this study, we only focus on the statistics and characteristics given in this dataset, rather than extracting more information from the youtube links to conduct the analysis. So after removing the unused and meaningless variables, we'd like to look into 18 variables.

Our Shiny App link: `https://rheon.shinyapps.io/SuperBowl-Ads/`

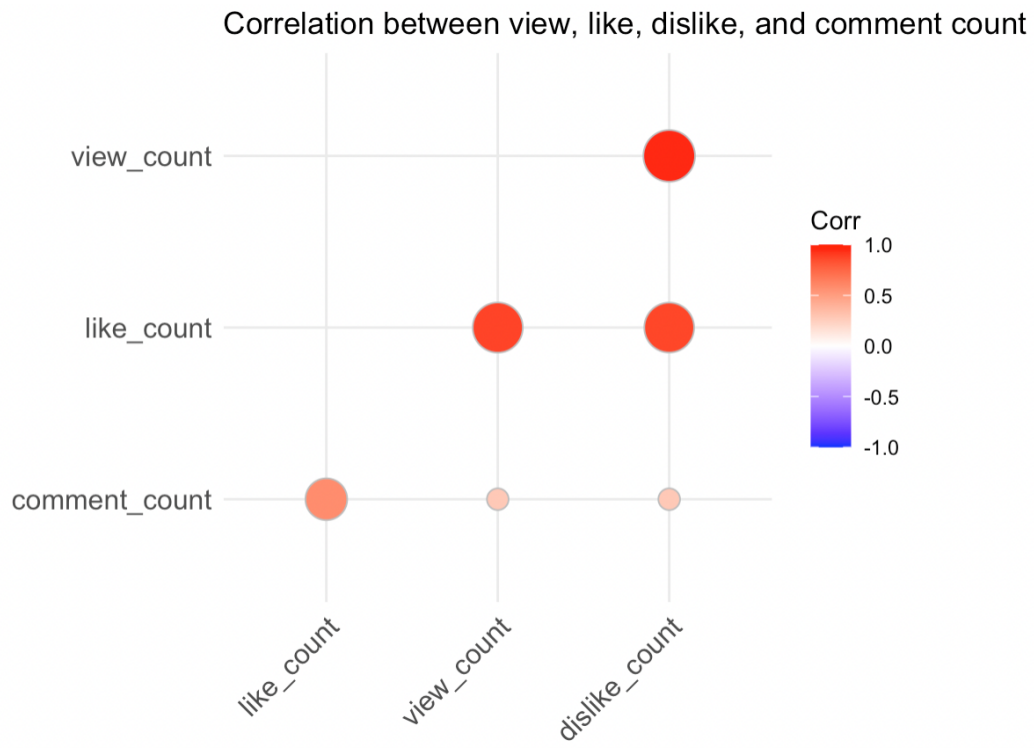

Figure 1: Correlation between view, like, dislike and comment count.

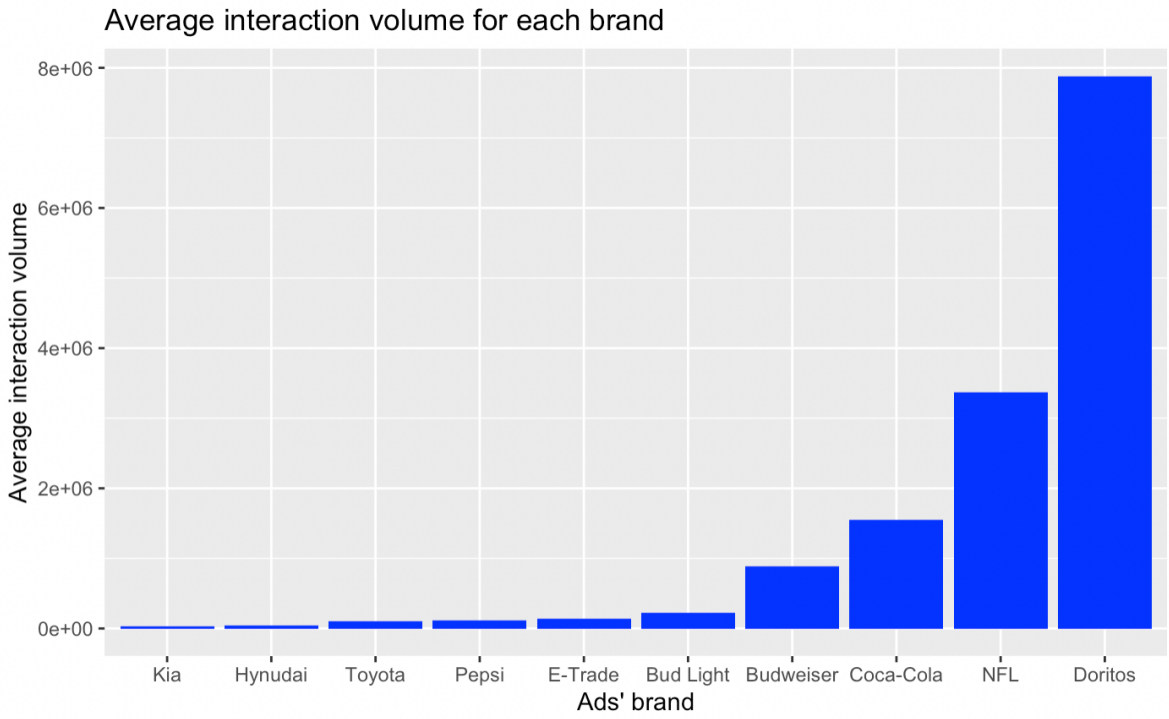

Figure 2: Average interaction volume for each brand.

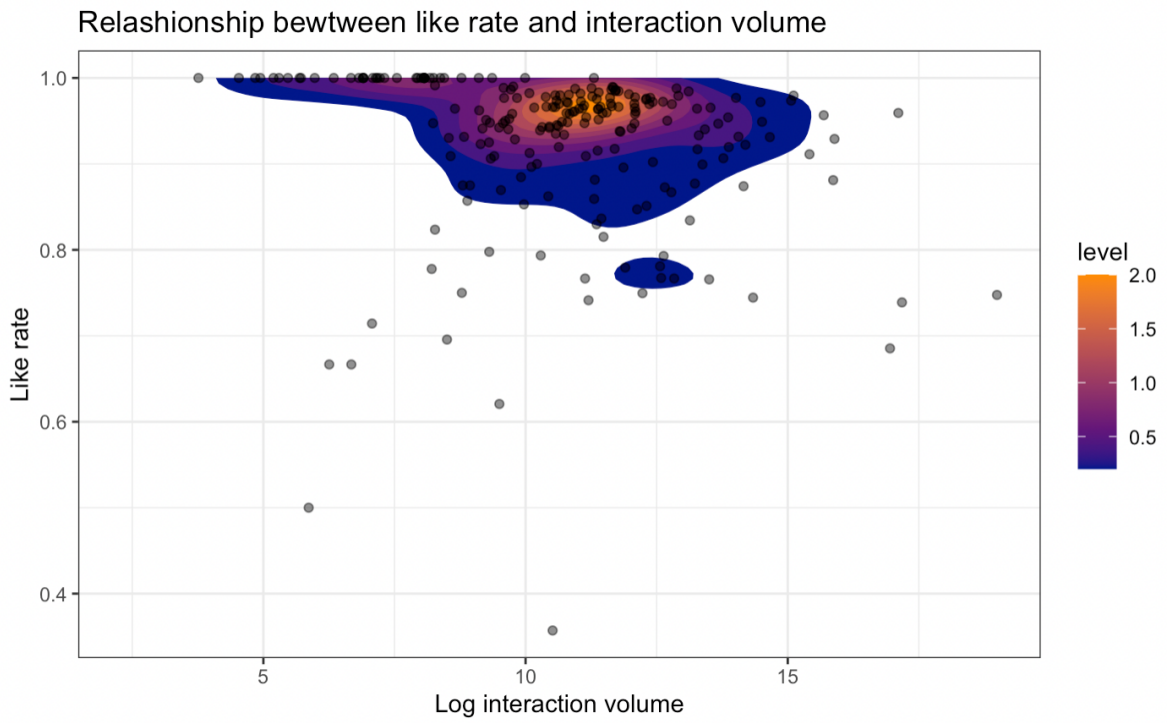

Figure 3: Heatmap for relationship between like rate and interaction volume.

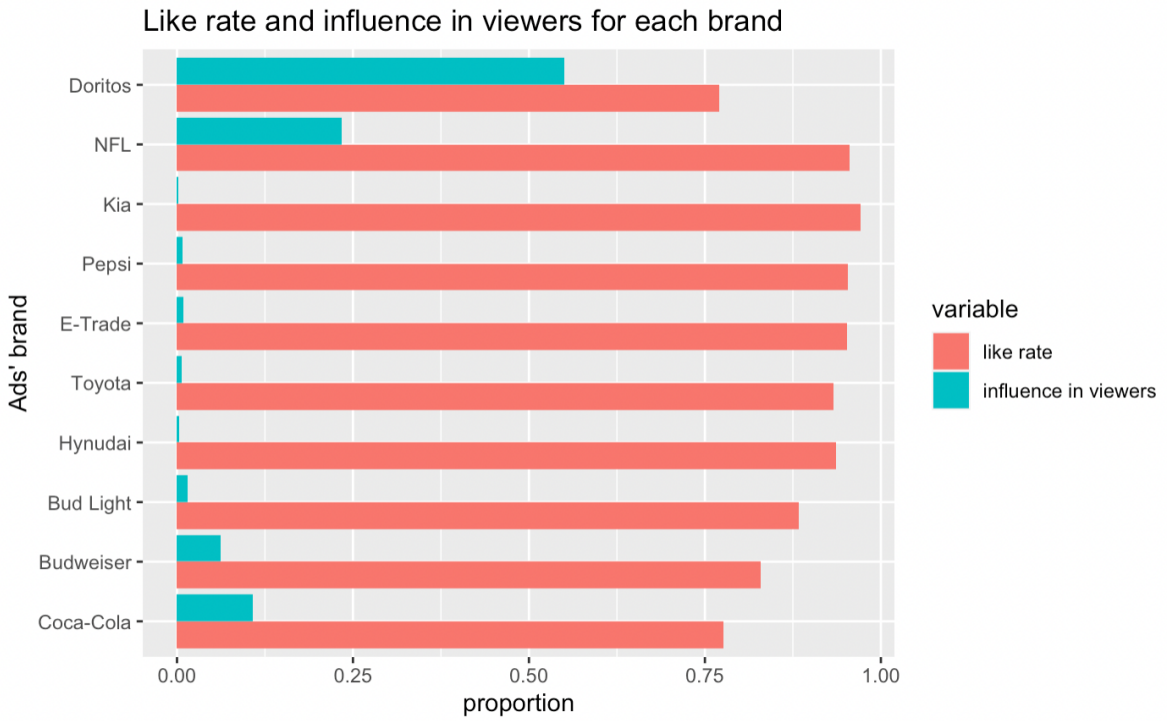

Figure 4: Like Rate and Influence in Viewers for each brand.

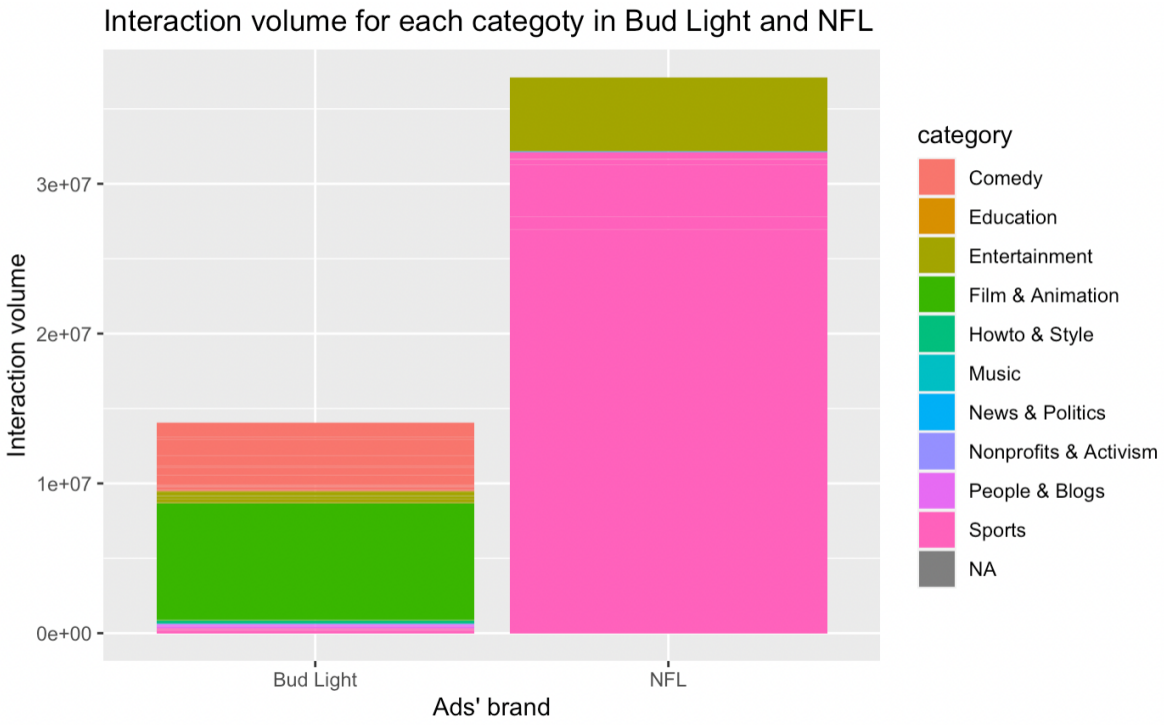

Figure 5: Interaction volumes for each category in Bud Light and NFL.

Figure 6: WordCloud for all advertisements' descriptions.

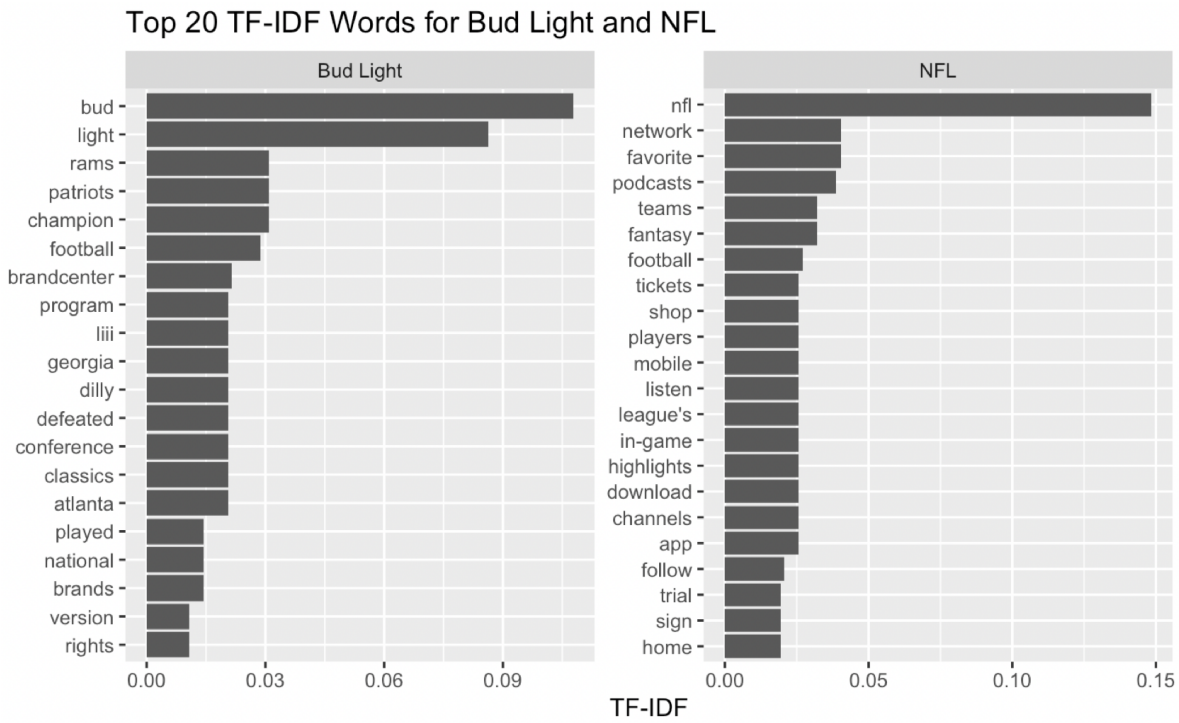

Figure 7: Top 20 TF-IDF Words for Bud Light and NFL.

3 Method

For our analysis, we relied on graphs produced by R, especially the tidyverse package in it. The responses for the placement of ads could be determined by **3 matrices** regarding this dataset:

- Interaction Volume = view count + like count + dislike count + comment count
- Like Rate = like count / (like count + dislike counts)
- Influence in Viewers = view count for a brand / total view count

Based on these matrices, our selection criteria for brands that have good responses is brands whose both average Interaction Volume and Influence in Viewers are top five, with a Like Rate greater than 87.5%. For the Interaction Volume matrix, we did a log transformation on it. Then, the interpretation of it becomes the percentage in Interaction Volume. The reason for that is the data are left-skewed and log transformation makes it approximately normal. Therefore, it could improve the linearity between variables and make the graph readable. To visualize the result of our chosen best response brands, we applied correlogram, heat map and bar charts here.

To investigate the effect of different features on response variables which are Interaction Volume or Like Rate, we use shiny app to create two simple linear regression lines to fit the group of videos that have a certain theme and the group of videos that do not throughout the selected time period, with response variable to be log Interaction Volume or Like Rate. Hence, we could easily tell how the interaction effect of features changes throughout certain time frame from the distance of two regression lines. Besides, a bar chart is applied to visualize the categories contained in best response brands versus each category's corresponding Interaction Volume on y-axis.

To have a glance of commercials' descriptions, we want to find the most frequently appeared keywords by word cloud. So we preprocessed the description texts first, by lower-casing and tokenizing them. Specifically, we removed punctuation, url, and numbers within all texts, and stored the rest of tokens and their corresponding brand name into an one-token-per-document-per-row table. Then, we removed all stop words from the table and counted the frequency of each unique word, as well as

words like “superbowl”, “commercial”, “ads”, etc. which were not meaningful to our study since they were just the reiterations of dataset’s topic, returning a word-frequency table for us to make a word cloud.

To see how the most responsive brands wrote their descriptions for ads and how they related their product with the Super Bowl specially, we figured out the most important keywords by computing and joining the TF-IDF values for each word treating the different brands as documents with the `bind_tf_idf()` function. Using the table returned with columns of all brands, word and tf-idf, we only visualized the good responses brands determined by our selection criteria to see their top 20 words sorted by tf-idf values.

4 Result

From the correlogram for variables to calculate Interaction Volume *(Figure 1)*, we could clearly see that the correlations among view count, like count, dislike count, and comment count are positively correlated by pairs, since all points are either red or white. Therefore, it is reasonable for us to add all of them together and create a matrix called Interaction Volume for it includes all information from these four variables.

From the histogram of average interaction for each brand *(Figure 2)*, it shows Doritos have the highest average interaction volume, followed by NFL which is little less than half that of Doritos’, while Coca-Cola, Budweiser, and Bud Light’s average interaction volumes are ranked behind.

From the heat map of relationship between like rate and interaction volume *(Figure 3)*, we could tell that most videos have high Like Rate and medium Interaction Volume. Thereby, Interaction Volume could not be the only criteria to determine whether the brand has good responses or not, so we should combine Like Rate with Interaction Volume to determine the best responses brand.

From *Figure 4*, the bar chart shows Like Rate and Influence in Viewers for each brand. Doritos, NFL, Coca-Cola, Budweiser, and Bud Light are top 5 highest Influence in Viewers. However, Doritos has the lowest Like Rate. Within these top 5 brands, Bud Light and NFL have Like Rate over 87.5%, which therefore could be determined as two most responsive brands.

From the bar of the NFL in *Figure 5*, we can clearly see that videos that are Sports and Entertainment could receive the highest Interaction Volume. From the bar of the Bud Light, two main categories that contribute to the total Interaction Volume are Film & Animation and Comedy. Besides, Sports and Entertainment videos also have a certain amount of Interaction Volume, while other types of videos almost have no contribution to the Interaction Volume for these two brands.

For the interactive Shiny app, as an example we could choose the theme of patriotic and the response variable as the Interaction Volume. We could tell that videos that contain the patriotic theme would receive more Interaction Volume than videos that do not contain the patriotic theme. In addition, we could also see the Interaction Volume of patriotic themes actually decreased year by year. We could use the interactive widgets to determine the effect of different themes within a designated time frame.

From the WordCloud *(Figure 6)*, we can examine that the words being frequently mentioned in all descriptions are brand names where NFL appeared the most frequently. Other than that, words related to games’ result, month held, famous persons appeared on Superbowl games, for instance “fails”, “january”, “game”, and “arnold” appear frequently as well.

Regarding the top 20 important words for Bud Light and NFL *(Figure 7)*, in Bud Light’s keywords histogram, the words features the product promotion in the context of Super Bowl, with keywords ordered by “bud”, “light”, “rams”, “patriots”, “champion”, etc.. In NFL’s keyword histogram, the most important words were “nfl”, “network”, “favorite”, “podcast”, “teams”, etc..

5 Discussion

5.1 Best Responses Ads' Brands

Brands that have the high Interaction Volume are almost all beverage or snack companies such as Doritos, Coca-Cola, and Budweister while brands that have the low Interaction Volume are almost all car companies such as Toyota and Hynudai. The first reason for this might be that beverages and snacks are much cheaper than cars and have a higher purchase rate, so advertisements of them will receive higher interaction. Another reason may be people will drink or eat something when they are watching the superbowl, so they of course attract more attention.

By our selection criteria that brands have good responses is top five average interaction volume and influence in viewers with a like rate greater than 87.5%, NFL and Bud Light are only two brands that satisfy these two criteria *(From Figure 2 and 4)*. In addition, NFL has higher values on all three matrices, so it is the brand with the best response. This conclusion also substantiates the information we get from heat map *(Figure 3)* that videos with high Interaction Volume do not have a high Like Rate. For instance, Doritos have the highest Interaction Volume but the lowest Like Rate.

NFL has both high average Interaction Volume and Like Rate might because Super Bowl is the actual the final playoff game of the NFL to determine the champion. Then it and its advertisements will definitely receive high attention and good Like Rate. However, we could not find an explanation as to why most videos that have high Interaction Volume also have a low Like Rate. To give a reasonable account for this question, we need to investigate more videos or other features of videos in the future, like the scripts of ads.

5.2 Categories and Themes of Ads

Most types of videos are Sports, Film & Animation, Entertainment, and Comedy for the two brands NFL and Bud Light that have the best response *(From Figure 5)*. Advertisements of NFL are identified as a sports category for sure and Film & Animation is commonly used by big brands. Noteworthy thing here is the Entertainment and Comedy categories. These two types of videos could receive such a high Interaction Volume because most people use Youtube as a platform for videos that could entertain them *(Reference 3)*. Therefore, brands might put more funny stuff in advertisements to increase the Interaction Volume.

If we just focus on the recent data to investigate the effect of different characteristics, take the funny characteristics as an example, videos that contain this characteristic have higher Interaction Volume and Like Rate. This corresponds to the conclusion we made from *Figure 5*, where entertainment causes high interaction volume. However, the effect of funny videos on Like Rate becomes less obvious. This might be because there are less funny topics or memes in recent years.

Characteristics that could increase both Interaction Volume and Like Rate from 2000 to 2020 are funny, celebrities, and danger. The reason why celebrities and danger have high Interaction Volume and Like Rate might be they are not common in daily lives. Therefore, people are likely to be curious about them. A possible recommendation we would give to companies is in order to improve the response, they should include these three characteristics in their advertisements.

5.3 Text Analysis of Ads' Descriptions on Youtube

Regarding the findings in word cloud *(Figure 6)* for all ads' descriptions, we could observe that most ads clearly stated their brand, month of games which were mostly January, and failed game results. Other than these, funny was the most frequently mentioned ads characteristic, indicating that most ads were based on the funny theme. Another high frequency word "arnold" was the name of the actor Arnold Schwarzenegger that was invited to film the Super Bowl commercial ads, which showed that it was quite eye-catching to have some famous celebrities appear in the ads and emphasize their names in the descriptions.

Looking into the top 20 important words in Bud Light, we could get a sense that they spent most of the effort on promoting its product, Limited-edition Rams Bud Light Championship Bottles, with a theme of patriotism. Imagine a scenario where Super Bowl fans are watching the games' live streaming with their friends or family, which is a typical American style culture, and drinking while watching. Thereby, the patriotic theme and the Bud Light's product fitted so well with the atmosphere of the audience watching the Super Bowl that the Bud Light's ads undoubtedly received great attention and responses.

Another brand with the best response was NFL, National Football League with 32 teams and delivering sports news, games and entertainment in its app and online platforms. Super Bowl is completely in the NFL's scope of business, so that its ads received high responses without doubt. Its ads' descriptions greatly incorporate online sources of NFL, including NFL network, NFL podcast, NFL app, and football.

Therefore, from all text analysis we did so far, the ads that received great responses mostly had the properties in ads' descriptions of clearly stating games related information, mentioning their brands' name, promoted products and names of popular celebrities appeared, and making the audience immerse in the theme of sports, funny and patriotism. Also, it was beneficial to summarize some useful sources related to the brand's product in the description as NFL did, to provide the audience with easier access to the product if they were interested in. If we can get more data like sales data for products promoted in these commercials, and investigate into the high sales brand, we could figure out more useful findings about the relationship between products and Super Bowl commercials.

References

1. Wikimedia Foundation. (2022, October 12). Arnold Schwarzenegger. Wikipedia. Retrieved October 14, 2022, from https://en.wikipedia.org/wiki/Arnold_Schwarzenegger

2. TechPostPlus. (2022, July 9). YouTube video categories List (complete guide). TechPostPlus. Retrieved October 14, 2022, from https://techpostplus.com/youtube-video-categories-list-faqs-and-solutions/

3. Bump, P. (2021, June 14). Why people visit youtube & how to engage them [new data]. HubSpot Blog. Retrieved October 14, 2022, from https://blog.hubspot.com/marketing/why-people-visit-youtube