**Data Visualization Project Report**

**ISM 6419**

**Sujhan Das**

**Introduction**

**Background**

The Super Bowl is an emotional event involving the annual playoff championship game of the National Football League (NFL). It is one of the most viewed athletic events in the world, and it frequently draws the greatest audience of all American broadcasts during the year. Due to the high viewership, commercial airtime during the Super Bowl broadcast is the most expensive of the year. This prompts corporations to produce their most expensive advertising for the broadcast and thus, commercial viewership has become a vital component of the event. The Super Bowl is also the second-largest food-consuming event in the United States, after Thanksgiving dinner. As an international student currently studying in Tampa, I was privileged to be in a location where the games were to be held (Raymond James Stadium, Tampa) that year. I witnessed firsthand the number of festivities and celebrations that occurred in the city even though I did not actually go watch the actual game for myself. That itself speaks for the volume of awareness USA has for the Super Bowl and how impactful brand associations to the game can be.

Viewers tend to be emotionally invested in such games as they go through feelings of anxiousness, happiness, frustration, and optimism. It is very common for groups of people to watch the game with their friends and family that share this experience together. To be able to present a story about your brand and product to this audience can be genius marketing. Commercials have the strength to make viewers laugh, cry, and invoke feelings and emotion about their product and brand. Super Bowl commercials clearly necessitate meticulous persuasive rhetoric in their planning, design, and execution.

Jamie Turner, the founder of the 60 Second Marketer, in conversation with CNN mentioned that the Superbowl strategy is only recommended for big brands like Budweiser and Microsoft. These companies can afford to invest in marketing so much so spending $5 million for thirty seconds of an advertisement has the potential to bring back a substantial amount of return. In the recent years, Former President Trump and Mayor Michael Bloomberg have reportedly bought $10 million worth of advertisements to run political advertisements during the game. During TV runs of the game, the slot timings of the commercials are also crucial. Former President Trump had slots early in the game which may be because the audience’s attention for any game is at the most during the beginning of the game. If the game isn’t nail biting or doesn’t have your favorable teams playing, a viewer may tend to tune out of the game towards the end.

**Motivation and Research Questions**

For such an important marketing project, commercials need to be respectful to historical data and performance of previous commercials. This project aims to gain insights and address certain research questions regarding the content of the commercials. My end goal for this project is to provide visualizations for stakeholders who are “time-strapped” people, to have sufficient information to make a key decision on the direction of marketing strategies.

The top 10 commercials can be analyzed to see if they had common characteristics in terms of content of the ad. A key question to address is: What kind of story does the audience connect with? Specifically speaking about the top brand in Super Bowl commercials: Budweiser. The audience would expect the Beer brand would produce feel good, funny commercials, however they are famous for bringing in emotion instead of humor to connect with the audience. It would be interesting to detect whether they have a distinct style. A company which intends to launch their commercial at a Super Bowl game may look back at previous trends and be able to realize what kind of content would suit their brand the best. Such insights can help shape creative narratives that would be successful in bringing in viewership.

The scope of the dataset includes whether the commercial was funny, patriotic, uses a celebrity, involves danger, animals or nudity and sensuality. Such characteristics are very informational regarding the kind of stories brands want to produce around their product and how they want their audience to relate to the product.

TV viewings have reduced significantly since the advent of online streaming platforms, their decline is expected in the insights. What also must be taken under consideration is that even after TV viewings of the commercials during Super Bowl, these commercials continue to play on online streaming platforms like YouTube. YouTube viewing analytics could also contribute to finding out popular ad trends and most liked content in commercials. This might help identify patterns for most successful commercials on YouTube and how they have trended across the years.

**Methodology**

**Preparation of Data**

The Original dataset is sourced from <https://github.com/fivethirtyeight/superbowl-ads>. “Superbowl-ads.csv” contains a list of ads from the 10 brands that had the most advertisements in Super Bowls from 2000 to 2020, according to data from superbowl-ads.com, with matching videos found on YouTube. The staff of FiveThirtyEight then came up with seven defining qualities of a Super Bowl ad, watched every video, and graded each one using the following table's taxonomy.

|  |  |
| --- | --- |
| Column | Definition |
| year | Year the spot aired, according to superbowl-ads.com |
| brand | Brand of advertiser, grouped to account for spelling and punctuation differences, and sub-brands (e.g. Coca-Cola Mini is grouped into Coca-Cola) |
| superbowl\_ads\_dot\_com\_url | Link to superbowl-ads.com entry for this ad |
| funny | Was it trying to be funny? Funny commercials are marked True. Anything serious or dramatic is marked False. |
| show\_product\_quickly | If the product was shown within the first 10 seconds of the ad |
| patriotic | If the commercial involved any aspects of being part of USA |
| celebrity | If it involved a celebrity endorsement |
| danger | If the commercial involved aspects of danger, violence, guns etc |
| animals | If any animal (real or through VFX) was present |
| use\_sex | If any subtle suggestions of sex, sexuality, sex appeal or nudity were present |

Additionally, a dataset from <https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2021/2021-03-02/youtube.csv>

was scraped via R Studio that contained extra columns such as Dislike Count, Comment Count, Category ID which I connected to the master dataset linking it through the common field “superbowl\_ads\_dot\_com\_url” which acts as a unique id for each commercial.

R Script Snippet:

**library (readr)**

**urlfile="https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2021/2021-03-02/youtube.csv"**

**mydata<-read\_csv(url(urlfile))**

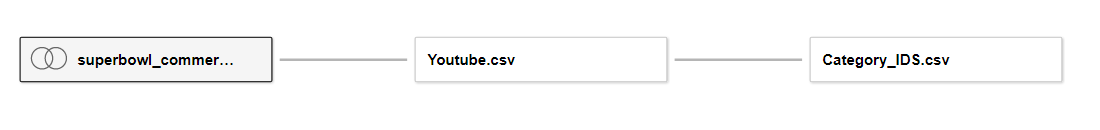
**write.csv(mydata,"C:/Users/sujha/OneDrive/Desktop/Data Visualisation/Super+Bowl+Commercials/Youtube.csv", row.names = FALSE)**

I additionally created a dataset from scraping data from <https://gist.github.com/dgp/1b24bf2961521bd75d6c> that had a list of Youtube Api video category ids. This would contribute heavily to specific YouTube analysis.

I further downloaded csv from <https://www.pro-football-reference.com/super-bowl/>

To get information about the stadium, city, and location of the game. The dataset also includes the winning team and losing team along with the points scored in the game.

This addition is to stretch the ambition of the analysis a bit. The aim is to explore whether location of the game or the players of the finals of the game show a trend with the brands showcasing commercials that year.

This is a snippet of my connections of the datasets in Tableau. 

Diagram

Description automatically generated

Superbowl\_commercials.csv consists of data from FivethirtyEight. I have joined it with the Youtube scraped dataset by joining the fields containing Superbowl Urls. The youtube.csv consists of category ids column which I have joined with the third dataset to get the category names. The superbowl\_commercials.csv file has been inner joined to superbowl\_locations by connecting them through the Year field.

Tableau has been used as the visualization tool for creating charts and dashboards of this project.

**Visualizations and Analysis**

Number of Ads and TV Viewers trends

Chart, line chart

Description automatically generated

This visualization is a display of the number of commercials from the top 10 brands per year. We can see here that the average number of ads per year is 12. I have used the Forecast option here from Tableau to predict the values for the next four years based on the historical data and a 95% confidence interval. It’s interesting to see that there is an upward trend in the forecast which may be true for the current year 2022 due to post-Covid recovery.

Chart, line chart

Description automatically generated

We can see a rapid decline in TV viewing trends as one would have expected. TV streaming will possibly decline even further as online streaming services gain more and more popularity. This should be an important factor for brands as they will not only target the Super Bowl audience but also online platform audiences after the game stops airing. YouTube is a strong online platform for advertisements. Considering its analytics is also vital.

**Top Brands**

Chart, bar chart

Description automatically generated

The top brands according to the number of commercials are as per the visualization above. Bud Light (62) and Budweiser (11) seem to have the greatest number of commercials by a bid margin. The rest of the brands have commercials almost less than half of the top two giant brands. Among the top six brands, four belong to drinking products, a common complementary when audiences watch the game.

Chart, bar chart

Description automatically generated

The next visual is the estimated cost expenditure from brands for commercials. Here the color legend plays a crucial role as it is immediately easy for us to identify that the brands have been shuffled. This means just because a brand has the most number of commercials, doesn’t mean it made the greatest investment. Brands like Coca-Cola and NFL have spent more with respect to the number of ads. Brands like Pepsi, Doritos and E-trade have spent less with respect to the number of commercials they have released.

**Average Cost per Brands**

The average cost spent by brands on commercials have an expectedly upwards trends. Here, I have tried to study the different trends by brand groups.

Chart, line chart

Description automatically generated

This line chart shows the upward trend of amount of capital invested in these commercials for drinking brands (Bud Light, Budweiser, Coca-Cola and Pepsi). There is a steep rise from 2014 and then a gradual fall till 2019. There is a rise again from that year till the present.

Chart, line chart

Description automatically generated

This line chart shows the general upward trend of average cost of capital invested in advertisements from motor brands like Hyundai, Kia and Toyota. The trend is visibly different from that of the drinking brands. There is a peak at 2017 and then a decline right after and then a stagnant trail thereafter. Unlike the drinking brands, motor brands haven’t invested more after the steep fall.

This is an interesting insight about reluctance from motor industry to invest in marketing possibly a result of Covid effects on the motor industry.

YouTube Analytics

Chart, bar chart

Description automatically generated

This chart is in exclusion of the most viewed commercial by Doritos (which has approximately 180M) to have a better view of the rest next ten commercials in the top eleven watched commercials on YouTube. Almost all brands are food and drink commercials. I have added the Like Counts for each commercial as well. I didn’t want to do an extensive analysis on the likes as every YouTube viewer doesn’t necessarily press the like or dislike button on the YouTube video. It may be a faulty parameter as to determine whether the video was actually “liked” or how much it was liked. However, view count is a parameter about the reach of the commercial and whether it has become a talking point among individuals and should be taken under strong consideration.

For the next analysis I created a new parameter called “LikeDislikeRatio”. This is to support my theory that only a certain population of YouTube viewers engage with the “Like” and “Dislike” buttons. This parameter can be considered to determine the most well perceived advertisements and those can be assumed to have less criticism.

Chart, bar chart

Description automatically generated

Interestingly, the motor brands Kia, Hyundai and Toyota have more well perceived advertisements than the drinking brands who otherwise have been topping in most of the visualizations.

Chart, bubble chart

Description automatically generated

This visual is a bubble chart for the category ids of all the YouTube videos. This shows the number of commercials per category. Entertainments, Comedy and People seem to be the top three categories.

**Map Visualization**

Map

Description automatically generated

This visualization shows us the number of commercials each brand has when the Super Bowl takes place in each region.

Map

Description automatically generatedInterestingly, E-trade commercials only aired when the Super Bowl took place in the slight eastern region.

Map

Description automatically generatedThe motor brands i.e Hyundai, Kia and Toyota have a greater number of commercials playing when the games take place in the Southern States.

This may not be a very strong insight or prejudgment for forecasting a successful commercial. However, there is a slight trend and would be interesting to gain more insight if there were more data about other brands.

Dashboard: Characteristics that make a Commercial

Chart, bubble chart

Description automatically generated

This dashboard includes filters for Brand and Year. By selecting the brands and years we can get insights about the kind of commercial it is. Humor and showing the product quickly seem to be top contenders in characteristics.

Chart, bubble chart

Description automatically generated

These are the defining characteristics for a commercial about a drink and food brand: Funny, includes animals and shows the product within ten seconds of the commercial. Most of these commercials are not patriotic.

Chart

Description automatically generated

Focusing on the non- alcoholic beverages: Coca Cola and Pepsi, almost half the advertisements use celebrities which may be a reason why Coca Cola spends more on commercials compared to the number of commercials it has. They are less funny than the alcoholic brands.

Chart, bubble chart

Description automatically generated

Most motor brands show the product relatively slower than the drink and food brands. Studying all the brand type groups show that they all have similar characteristics. Thus, Super Bowl brand commercials do not always have brand specific content. Further insights can be drawn when filtering through the years.

**Conclusion**

The Super Bowl is the year's most important advertising event. Companies queue up to spend millions of dollars on commercial time during the game, and they work on the advertisements for months, researching customers and experimenting with new creative concepts.

However, unlike most advertising these days, Super Bowl commercials are made to appeal to a wide range of people. With almost 100 million people expected to watch the game, this is hardly the place to send out tailored messaging to a small group of people. Super Bowl commercials must reflect broader social, cultural, and economic trends.

The analysis conducted through visualizations show how the different brands have faired in terms of commercials throughout the years. The beer brands Budweiser and Bud Light seem to be giants in the Super Bowl Commercial space and because of the relevance of their product with the event of viewing the game, rightly so. Commercials give a good understanding about the current state of the citizens of America. The commercials in 2021 seem to have a more positive base which may be a calculated response after Covid.

A limitation of this analysis was the number of brands associated with the dataset. Cryptocurrency companies have reportedly started putting up commercials for Super Bowl but that does not reflect on the dataset. A future scope for this analysis could be a study on how the stock prices of brands increase of decrease after their Super Bowl commercial airs in the game.

*References:*

<https://www.youtube.com/watch?v=3xJallX4_Lk>

<https://en.wikipedia.org/wiki/Super_Bowl#Entertainment>

<https://en.wikipedia.org/wiki/Super_Bowl_commercials>

<https://www.edutopia.org/discussion/lets-analyze-super-bowl-commercials>

<https://www.cnn.com/2022/02/14/perspectives/super-bowl-ads-2022/index.html>

<https://finance.zacks.com/invest-jse-8740.html>

<https://www.thestreet.com/lifestyle/sports/super-bowl-ads-stock-prices-14854700>