

## Data Analysis Project Report

The Super Bowl is the yearly championship game in the National Football League. It is one of the most highly anticipated events of the year. It has evolved from just a national event into a cultural phenomenon, bringing in multimillions of viewers every year. One of the more interesting aspects of the Super Bowl is the quick, attention-grabbing commercials that air during the game. These commercials have become just as much of an experience as the game itself.

Companies shell out incredible amounts of money for coveted ad placements, so the question arises of what makes the audience return to these commercials after they air on television? In addition, how does the return viewership help these businesses or companies be more successful?

“Superbowlcomm.csv” holds datasets detailing the information needed to answer what it is that gives these companies more viewership, which in turn can lead into more business. A comprehensive analysis has been carried out to answer that question with the data dictionary being listed below:

Variable	Definition
Year	Year the advertisement aired
Brand	Brand of advertiser
Super Bowl Ads Link	Link to SuperBowl-ads.com entry for the ad
YouTube Link	Link to YouTube video for the ad
Funny	Was the ad trying to be funny? Is it jokey, goofy, or silly?
Shows Product Quickly	Did the commercial show the product right away?
Patriotic	Was the ad patriotic?
Danger	Did the add involve danger?
Animals	Did the advertisement involve animals?
Uses Sex	Was sex used to sell its product?
Length	Runtime for the ad, in seconds.
Estimated Cost	Cost of the TV spot for the ad, in millions of dollars.
YouTube Views	YouTube video view count as of 01/11/2022
YouTube Likes	YouTube video likes as of 01/11/2022
TV Viewers	Millions of viewers, according to SportsMediaWatch.com

The data is from SuperBowl-ads.com, from 10 brands that had the most advertisements on television during the Super Bowl from 2000 to 2020. FiftyThree staffers then came up with characteristics of Super Bowl ads that fit the commercials best.

When it came to the data, cleaning was not the easiest with what was given. There was a good amount of missing data that needed cleaning. I dropped some rows that had missing values for specific columns. From there I dropped columns that were not necessary to data, or the question being asked. The main one that was dropped was YouTube likes because YouTube hides dislikes now and it would not be an accurate representation of what viewers enjoy. Sorting by YouTube views was necessary as it is the primary source or piece of information needed to answer the

questions at hand. Due to the dataset being mostly Boolean in response to the data dictionary, I changed the True or False values into integers for plotting purposes.

From a visualization standpoint, I created a bar chart to represent celebrity influence on returning views of the advertisement. It is shown that there were more celebrities not in these commercials than there were in them. Based on the data and the bar graph of the relationship between the two, the trend of having a celebrity in an advertisement would bring more viewers to return to the commercial on YouTube if more data was given. In this case of more data, it would be more celebrities in more Super Bowl commercials.

A boxplot was displayed for the Patriotic and Animals relationships against YouTube views. Due to the large disparity of the data, the boxplot does not show as most normal boxplots do. Both seem to represent about the same amount when it comes to the correlation with the views. The plot shows that just because a commercial included something patriotic or had animals in it, does not mean viewers were more interested in it.

Plotting a strip plot showcased the idea that whether the product was shown relatively soon, the data failed to show an established side that indicated more of a positive relationship with a YouTube return. There was one outlier that might have been a commercial that was widely enjoyed by the masses.

Finally, a joint plot and a heatmap were displayed in the code. The joint plot was used to try to represent a positive relationship between comedy, or if the advertisement was funny to how that affected the audience to return to watch the video. The heat map is used to present the hot and cold spots of each column and row.

As a result of data analysis, viewers are going to be attracted to what captivates them. It seems that celebrities entice the audience to return to these advertisements in terms of trends more likely. Overall, it seems that an advertisement being funny or including celebrities yielded the most viewers on return after viewing it live during the game. This makes sense as celebrities are such a strong part of pop culture along with the national event that the Super Bowl is.

Entertainment is what people watch tv for, so comedy also plays a big factor in making someone feel good or good enough to want to return to a commercial. So as companies see these results and compare them to advertisements of years previous, they might be more inclined to put celebrities in the commercials with an element of comedy to endorse their products to promote business and increase sales.