Whitepaper Analysis - February 2023

Game Day Ad Analytics 2023

Authors: Jacob Horstmann, Travis Swiger, Divyanshu Tripathy, and Kyler Zarate

"What attributes were most beneficial for ads in the 2023 Super Bowl?"

KEY TAKEAWAYS

- **Starpower matters**
- **Run ads in the 1st Quarter, Avoid the 4th Quarter**
- **Community Engagement Pays off**
- **Comedy ads are high-risk, high-reward**

Research Question and Data

Every year, millions of dollars are spent on ad space during the Super Bowl, an event with an estimated 113 million viewers tuning in. Ad space is usually bought in 30, 45, or 60 second intervals, with a 30-second time slot costing $7 million in 2023. With a cost this high, it is crucial that companies understand their market to the best of their abilities and make the most impactful ad they can.

For the Game Day Ad Analytics competition, we wanted to understand what ads in the 2023 Super Bowl were better received and why. To do this, we utilized the Twitter API to track information about tweets containing selected keywords. Doing so returned a set of ad-level data containing variables such as tweet count, quarter, and impressions, which then underwent a thorough cleaning process. In addition, we built two more datasets; one containing ad- and brand-specific data, and the other using publicly-available Google search data.

Overall, our approach to the question of "What makes an ad successful?" was to start at a high-level overview and work our way down to developing measures that would quantitatively define the "best ad." The initial overview was to outline some exploratory analytics which delved into ad content and industry presence. Later, we moved on into consumer-facing measures such as Twitter likes or interest capture, then to business-facing measures which included correlations and genre counts. All of this led up to our final segment: aggregating measures from each of these categories to produce a final, quantitatively-supported ranking of the best ad.

Data Cleaning

The data cleaning was the most difficult and time-consuming part of this challenge. Accomplishing this involved extensive use of Python and the Pandas library. Later on to do the analytics we used the SKLearn library for linear regression and Tableau for the visuals. We knew that our first step was to get a better grasp on our data, so we started by looking at the amount of nulls in each column, and found several where this was more than 25% of the values. After that, we looked for columns that we knew would be irrelevant to our analysis, which happened to coincide with the columns that had a large quantity of nulls in them. Based on these reasons, those problem columns were dropped, making the data set inherently more manageable. Next, we eliminated any rows which should never have had nulls, and could not be used for analysis at all. We determined that it was imperative that every row have a brand name, so if this was not present, the row was dropped.

In cleaning our data, we decided that non-English tweets would be difficult to analyze appropriately, so we made the decision to eliminate any of these observations. They only composed of a considerably small fraction of the data, so the effect on our results can be expected to be negligible.

We then did some more exploring and discovered that some of the ads were either local ads, NFL ads, or shows on Fox that were not full commercials (simply 10 second fillers between plays or commentary). As a result, we removed observations related to these ads.

The last step in our initial cleaning was to merge the data that had brand names that were not exactly the same. Some data was mistyped, so we had to combine observations appropriately.

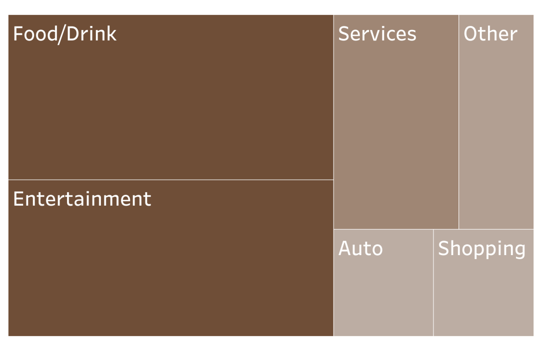
The cleaned data set was at the ad-tweet level, but we wanted to also have it at the brand-tweet level and grouped-by-brand level. In doing so, we used similar data-cleaning methods as before. These transformed data sets also let us add variables that pertained to specific ads such as the genre, if animals were included, and if it included a celebrity.

Our brand-grouped data enabled us to sum metrics such as likes, replies, retweets, and counting the total tweets, which we were able to use for estimations later on, and where we did the bulk of our analytics. This data gave us the flexibility to add variables like company revenue, total runtime, number of ads, and the advertising brand's industry.

Producing this additional data required some research. We used publicly-available information to get the most recent revenue figures, and Owler to get estimates for private companies. We also used YouTube and other websites to verify ad specifics, calculate runtime, and create dummy variables for variables such as celebrity presence. The other website we used to find important info was SmartInsights.com, which included a detailed breakdown of customer conversion rates by industry which were used to calculate revenue attributable to the ad.

Exploratory Analytics

Our first step was creating some simple statistics that described the advertisers.

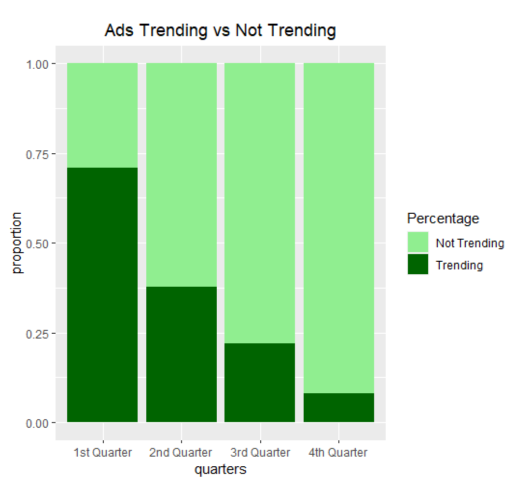


The above graphic depicts the industries represented by the advertisers. Food/Drink and Entertainment were by far the two largest industries, with 20 and 19, respectively. Food and Drink in particular had several large corporations including Annheuser-Busch, Doritos, and Pepsi. Entertainment was mainly represented by new movies such as The Flash and Indiana Jones, but also things like U2's concert, Disney, and Amazon Studios.

Genre of Ad	Percentage
Comedy	68%
Normal	21%
Heartfelt	11%

Looking into genres, it is clear that the go-to strategy for a Super Bowl ad is comedy. However, the ads which aren't particularly funny or heartfelt had an average 15 million impressions, compared to 11 million for comedic and 5 million for heartfelt. It is important to note, though, that if done right, comedic ads can pay off significantly. The company with the most impressions (65 million) was classified under comedic.

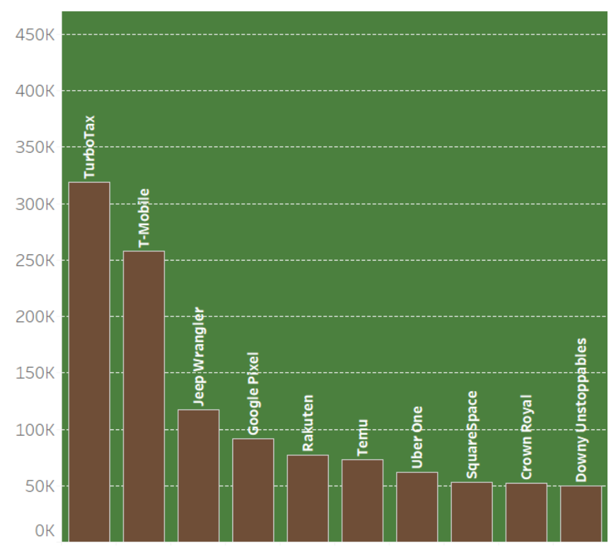
Another interesting relationship is between how popular ads were- according to Twitter data- as the game progressed.



It is very clear from the previous stacked column chart that ads are much more likely to go viral early in the game, and diminish steadily as the game draws on. In the first quarter, 71% of ads began trending, but by the 4th quarter, this dropped to 8%. Based on this, we believe that engagement is likely much lower.

Descriptive Analytics

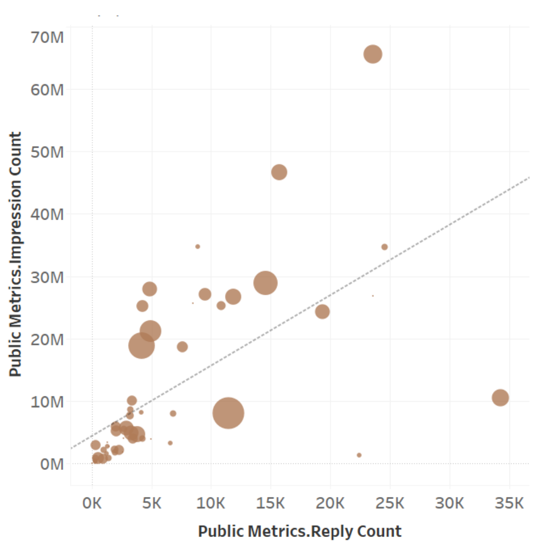
In order to explain the results of various advertisements, we conducted some descriptive analytics. Starting off simply, we ranked ads by those which had the most likes on related tweets.



By far the winner was TurboTax with almost 330k likes, followed by other major brands: T-Mobile and Jeep. Naturally, this begs the question- what were they doing right? We explored a lot of different variables that may have contributed, but by far the most important was how active the brand was on Twitter during the game. We found that for each additional tweet sent out by the company, it was correlated with, on average, 509 thousand more impressions. Diving further, 8 of these top 10 companies tweeted more than the median company. Using the median, we could rule out outliers such as Avocados from Mexico and sportsbook companies, which were far outside the normal range of mid-game tweets.

Another area we investigated was the correlation between the number of impressions and the number of replies. In the scatter plot below, we showcase this relationship, with the size of the point depending on the number of tweets sent out by the company's Twitter page. There is a clear positive and linear relationship between these two variables, indicating that increases in the number of replies by a company results in an increase in the number of impressions they will get. While there are some exceptions, this would imply that companies that have an involved social media page generally benefit from their efforts.

It is also interesting to note that some companies seem to prioritize replying over likes, retweets, or posts (indicated by small points with high reply counts). This implies that their ratio of replies is higher compared to companies with bigger points but less replies. It would be interesting to find a ratio of replies, likes, retweets, and posts that would maximize the number of impressions a company gets.



Predictive and Prescriptive Analytics

Change in Interest Capture

For our first measure, we wanted to understand how well an ad retained attention pre- and post-Super Bowl. To do this, we gathered data from Google Trends to measure search interest based on the occurrence of brand-specific keywords over a three-week period. For each day in this period, the numbers ranged from 0 to 100, with 100 indicating the day with the most search traffic. Using this format allows us to look at the daily changes in search query interest.

In order to make sure pre-Super Bowl trends were constant, we looked at keyword occurrences two weeks prior to airing. Good counterfactuals included those which had relatively constant trends before the game. On the day of the Super Bowl, nearly every brand saw a sharp spike in Google searches. What we were interested in was what the trends would return to, come postgame. For many brands, the pregame and postgame average searches look similar before and after the spike. This indicates that while the ad attracted users to search for their brand-specific keywords, the ad was not effective enough to sustain a higher level of interest. A good ad would see higher postgame average searches than pregame average searches. Certain brands were eliminated from consideration in this process if their trends were wildly inconsistent and didn't have good treatment counterfactuals. After recording the three-week trends for each brand, we calculated the average pregame searches from January 29th to February 11th and the average postgame searches from February 14th to February 17th (the latest data available). To identify the change between pre- and post-game interest, we took the difference between the two average values of the periods which gives us the change in search interest level.

From these measurements, we ranked the brands from largest to smallest increase. Our next step was to omit television shows that came from the Fox Network, because these ads did not have to compete or pay for ad space, and did not serve to answer the true question of this study. After omitting this factor (along with those already ruled out based on poor trend patterns), we found that the top performer in this metric was PopCorners, with an increase of 19.3 percentage points: from a pregame average of 12.7% to a postgame average of 32%. We also found that based on the top rankings, 40% of the brands were able to sustain a 5 percentage point (or greater) increase in average Google search interest after coming off the spike caused by the Super Bowl. It is reasonable to expect that the brands that managed to sustain that level of increase were likely able to profit from the increased traffic to their websites and social media because of increased sales from the additional attention and brand awareness.

Percentage Increase in Attributable Revenue

For this metric, we sought to quantify the effect of increased brand-specific Google search interest on weekly revenue for different companies, estimating the change in revenue that can be solely attributed to running the Super Bowl ad campaign. To start, we built on the previous pregame average, postgame average and the difference between the two. Each brand was then sorted into five industries: Entertainment, Food/Drink, Service, Shopping, and Auto. We then utilized the average conversion rates for each industry, giving us benchmarks for the expected share of people who end up buying from an advertiser website compared to the total number of Twitter impressions.

After gathering the annual revenues for each company, we converted to weekly revenue. Using this, we calculate ad-attributed revenue by multiplying the weekly revenue by the conversion rate and then multiplying the product by the pre- and post-game percentage point difference. Dividing this attributable revenue by the weekly revenue yields the percentage of weekly revenue that is reasonably attributable to each company's ad campaign. For this ranking, we again eliminated Fox shows. Based on the percentage increase of weekly revenue attributed to their ads, "Ant-Man" enjoyed the most success with a 1.91% increase.

Predicted Campaign

For our final measure, we sought to get a better idea of which brand had the best predicted campaign. To achieve this, we built an ordinary least squares (OLS) regression model that explored several different factors that affected Twitter impressions as regressors. For these regressors, we included: total screen time, number of tweets sent by a brand, presence of a celebrity, presence of an animal, and ad genre. From this model, we found that for every additional minute of runtime, the average number of Twitter impressions increased by 211 thousand, all else held constant. This would imply that the longer the ads run for during the Super Bowl, the more likely that they will be remembered by fans.

Another significant factor is the number of tweets sent by a brand's Twitter page mid-game. Every additional tweet sent resulted in an average increase of 411 thousand impressions, all else held constant. Further, genre played a significant role. In this model, the effect of making the genre of your ad typical (non-comedic or heartfelt) results in an average of 8.5 million additional impressions, all else held constant..The celebrity dummy variable was found to be significant at the 10% level, and was positively correlated with impression counts. The animal dummy variable was not statistically significant.

This OLS model gives us an equation with which we can plug in values into each of the selected variables. This equation then gives us a predicted value that takes each of the factors into account. For this criteria, we applied this equation to each brand and their metrics to give us predicted values and to compute residuals. Using these, we selected the brands with the highest predicted values, which implies that they fit the line of best fit the closest and maximized their Twitter impressions. The winner based on this criteria was FanDuel, with a predicted value of 38.4 million impressions.

Best Ad Rankings

Brand	Points
T-Mobile	18
PopCorners	17
Avocados From Mexico	14
The Flash	12
Tubi	11

The points were aggregated across three categories. First, the top ten in captured interest (based on Google Trends), were given 10 points if placing 1st, 9 points for 2nd, all the way to 1 point for 10th. The winner of this first category was PopCorners. The second category was based on predicted impressions, in which FanDuel won. Finally, we used our own metric of ad-attributable revenue, in which Ant-Man was the top scorer. Adding these point values yields the table above.

Conclusion

Analytic Limitations

While we are happy with our analysis, there was more we wish we could have done. The main limitation was time. With more time, we would have wanted to get to other measures like sentiment scoring. We were prepared to do it, however, none of us had access to a powerful enough computer. We considered doing it in batches, but by then we decided our time was better invested elsewhere. We were also curious to see if all press really is good press using impression counts and Google search trends. Time also limited what other variables we could consider adding, such as popularity of the celebrity? Unfortunately, we were not sure how to automate this and did not have time to do this manually.

Other Observations and Trends

One of the things we noticed was that a lot of the ads played on nostalgia. This has been a trend for a while. Many tv shows are reboots of old shows, Rihanna played many hits from the 2000s/2010s, and many news agencies reported on this as well. This may be due to increasing buying power of younger generations including Millennials and Gen Z. Another observation we had was that some companies did not see the same benefit of tweeting during the Super Bowl. The ones who did ran several polls or tweeted at other companies.

Recommendations

Through this process, we were able to come up with key aspects of running a successful ad campaign that is backed up by data. The first thing we would recommend is that if possible, run the ad in the first quarter. The social media engagement is much higher in that quarter than any other. Try to avoid being in the fourth quarter as much as possible, where only 8% of ads trended. Next, consider the content of the ad. The most clear answer is that it should include a celebrity, and note that including animals will usually not affect the outcome in terms of engagement. It is also important to decide what type of ad to run. There are two options that should be considered depending on risk tolerance. A comedic ad is a higher risk, but also offers a higher upside. On average, these underperform compared to a typical ad. If you are to do a comedic theme, make sure it gets people talking. The last recommendation we have is for the brand to engage with the viewers before, during, and after the Super Bowl. Some of the most successful ads ran deals, polls, and tweeted at other brands. Not all engagement is created equal however- some companies went for quantity over quality and did not see the same uptick in impressions. Considering all these factors, T-Mobile is who we consider the most successful, because they followed all of these recommendations.

Citations

Owler, https://corp.owler.com/.

Bond, Conor. “Conversion Rate Benchmarks: Find out How Your Conversion Rate Compares.” *WordStream*, 22 Oct. 2022, https://www.wordstream.com/blog/ws/2019/08/19/conversion-rate-benchmarks.

“Digital Marketing Strategy Advice - Smart Insights Digital Marketing.” *Smart Insights*, 25 Nov. 2022, https://www.smartinsights.com/.