BIG GAME, BIGGER Ads & BIGGEST DATA CHALLENGE

Team 05 – Dawn of Analytics
Debayan Dutta | Dhwani Patel | Disha Tapadiya | Kalyani Joshi

TABLE OF CONTENTS

1. ABSTRACT

This white paper presents an analysis of the ad and tweet data generated during the Super Bowl, one of the largest sporting and advertising events in the world. The study focuses on the relationship between advertising and social media engagement, and how this can be leveraged to maximize brand exposure and engagement during the game. Specifically, we analyze the Twitter activity surrounding Super Bowl commercials, and how this activity correlates with brand recall and purchase intent. We also examine how the length, style, and content of commercials affect Twitter engagement.

Our findings suggest that ads with a high level of emotional engagement and a clear call-to-action are more likely to generate Twitter engagement and positive sentiment. We also find that social media can amplify the reach and impact of commercials, leading to increased brand recall and purchase intent.

This paper provides practical insights for advertisers on how to create effective commercials that can generate buzz and drive engagement on social media during the Super Bowl. Additionally, our study provides guidance on how to use social media analytics to measure the success of advertising campaigns and refine strategies for future events. Overall, this research highlights the importance of integrating social media engagement into advertising strategies and the potential impact it can have on brand perception and ROI.

2. INTRODUCTION

2.1 Objective:

The aim is to conduct an in-depth analysis of Twitter data during the Superbowl game, with the specific goal of extracting valuable insights regarding the efficacy of commercial advertisements. This will be accomplished by closely monitoring Twitter activity during the event, with a focus on identifying emerging trends, key themes, and overall fan sentiment.

The analysis is based on an examination of a wide range of Twitter data, including popular hashtags and other topics of interest that surface during the game. This information is used to generate a comprehensive info graph that offers insightful and actionable recommendations for enhancing commercial strategy in the future.

Ultimately, the goal is to leverage the rich data available through Twitter to provide a nuanced and thorough understanding of the Superbowl advertising landscape, and to use this information to drive informed decision-making and strategic planning for future events.

2.2 Team Structure:

The team responsible for creating this white paper consists of four individuals with complementary skill sets.

Debayan Dutta contributes to the overall strategy and data cleanup. Dhawni Patel is responsible for coordinating the team's data collection efforts and preparing data for analysis. Disha Tapadiya designs and creates infographics that effectively communicate information to the audience. Kalyani Joshi conducts keyword extraction and leads the creation and writing of the white paper to ensure effective structure and content.

Together, the team leverages their diverse skills and expertise to work collaboratively in producing an impactful white paper. They define the strategy, plan and clean the data, create effective infographics, and write a well-structured and informative paper that delivers the key insights to the targeted audience.

2.3 Background and context – SUPERBOWL & Commercial:

The National Football League (SUPERBOWL) is the largest professional American football league in the world, with 32 teams located in major cities across the United States. The SUPERBOWL generates significant revenue through various sources, including ticket sales, merchandise sales, and commercial advertising.

Commercial advertising during SUPERBOWL games has become an integral part of the league's revenue streams, with companies investing millions of dollars each year to advertise their products and services during game broadcasts. These advertisements are typically aired during commercial breaks and are designed to capture the attention of viewers and increase brand awareness.

However, the effectiveness of commercial advertising during SUPERBOWL games is constantly being evaluated, and it is critical for the SUPERBOWL and its commercial partners to understand the impact of these advertisements on viewers. By analyzing social media activity during the game, such as on Twitter, it is possible to gain insights into the effectiveness of these commercial advertisements and inform commercial strategy going forward.

The analysis aims to identify the commercials that are generating the most attention and interest among fans, as well as to gauge the efficacy of these advertisements. Based on the insights gathered, the paper intends to offer actionable recommendations for enhancing the commercial strategy of the SUPERBOWL and its commercial partners.

3. ANALYZING DATA

3.1. Overall RAW Data:

The initial dataset provided for the analysis contained a large amount of data, with over 1.8 million rows and 27 columns. This made the data cleaning and analysis process challenging and required the use of specialized tools and techniques.

3.2. Tools and Technologies used & Data Cleanup:

- Tools and Technologies Used:
  - Alteryx, Tableau Prep, Tableau Desktop, R-Studio, Python
- Data Cleanup:
  - Used Alteryx to identify column values and data types.
  - Used Python and pandas to identify missing data.
  - Created a new dataset with relevant columns using Tableau Prep.
  - Split the "location" column into "city", "State" and "Country" using Tableau Prep.
  - Extracted hashtags into new columns from the tweet text column using Tableau Prep
  - Used the count function in Tableau to obtain counts of tweeted brands and ads during the Superbowl.
  - Changed the abbreviations in the "lang" column to full-forms of the languages using R-Studio.
  - Steps involved handling missing data, extracting key information, and reformatting the data for easier analysis using techniques such as column splitting, duplicate removal, and filtering out irrelevant data.

3.3. How data was Analyzed:

After the data cleaning process was complete, we analyzed the data using a range of techniques to identify patterns and trends. This included using Tableau to create visualizations and dashboards, using Python to perform text analysis on the tweet data, and using R-Studio to perform statistical analysis. We identified that the "id" column in the data set was of utmost importance as it was a unique column and the ids were associated with every tweet made including retweets and shares.

3.4. Final Data:

The final dataset used for the analysis was a cleaned and reformatted version of the original data, containing only the relevant columns and data points needed for the analysis. This dataset was used to generate insights and recommendations regarding the efficacy of commercial advertisements during Super Bowl 57, and to provide insights into fan sentiment and emerging trends.

4. VISUALIZATION AND INSIGHTS

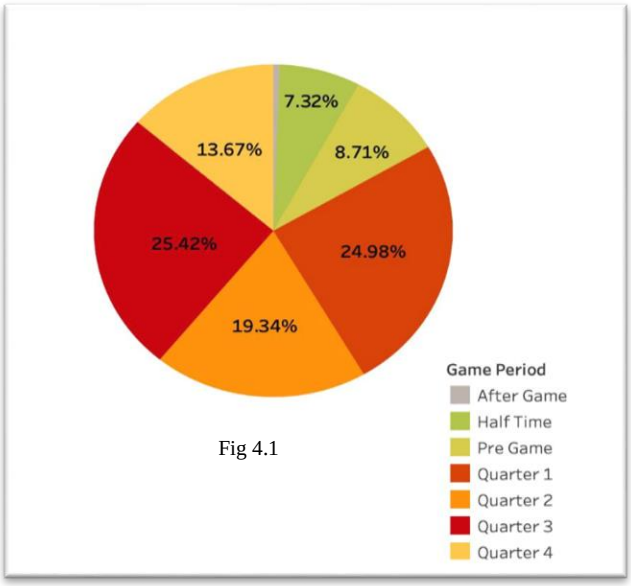


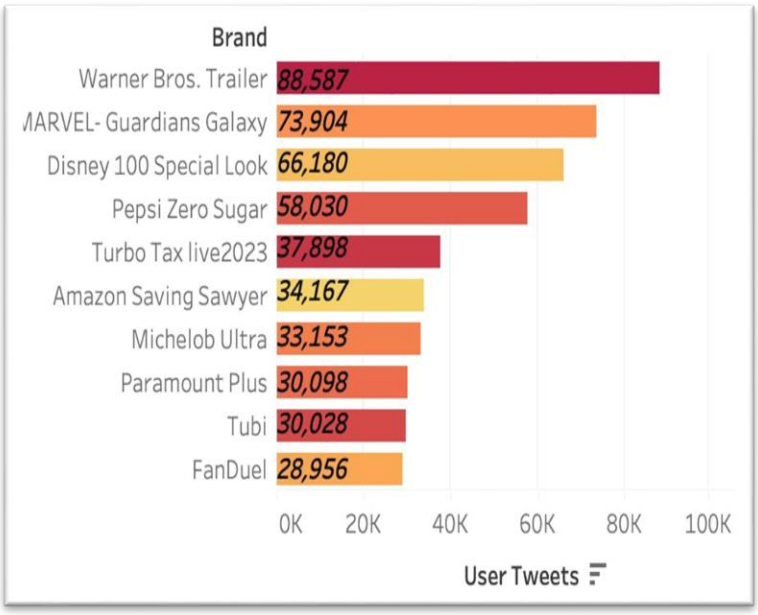
Fig 4.1

4.1 Performance by Quarter:

Our analysis of the Superbowl game involved examining the volume of tweets generated during various time segments, including the pre-game, four quarters, halftime, and post-game periods. Based on our findings, Quarter 3 generated the highest percentage of tweets. We presented this information using a pie chart. To obtain this data, we utilized Alteryx to count the number of tweets per quarter using the id column and time_period_of_the_game field, and then used Tableau to create a visual representation of the data.

4.2 Top Ads by Mentions:

The Super Bowl 57 in 2023 lasted for approximately 4 hours, during which we collected data on over 97 advertisements from different brands. Using our dataset, we have identified the top 10 most mentioned ads in user tweets. This information is presented through a word cloud visualization.

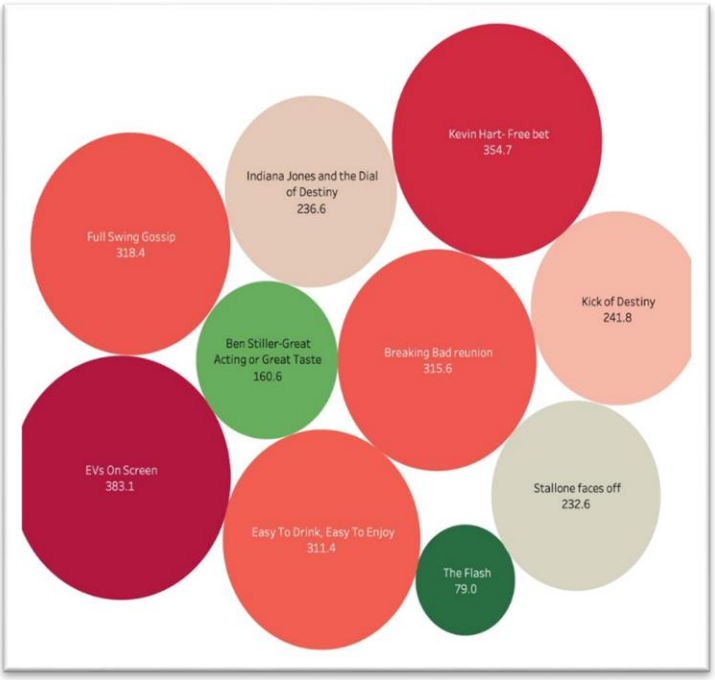


4.3 Most Tweeted Brands:

During the Super Bowl 57 in 2023, companies that advertised during the event received a significant number of tweets and mentions related to their brand. To illustrate this, we extracted the brand names from the "brand_ad_name" column and used the count function to summarize the "id" column and determine the top 10 most tweeted brands. We presented our analysis through a horizontal bar graph. Our findings indicate that Warner Bros. is the most tweeted brand, with approximately 89,000 tweets.

4.4 Cost Analysis of Ad Engagement

During the Super Bowl 57, each ad spent around 7 million dollars for a 30-second slot. The primary goal of each brand was to achieve the highest level of engagement from the viewers. To evaluate this, we analyzed the tweets, retweets, mentions, likes, and total impressions for each ad played during the event, to gain insight into viewer engagement and the overall impression of the brand. We presented our findings using a bubble chart visualization created in Tableau.



Hashtag	
SuperBowl	41,133
TheFlashMovie	24,155
Sweepstakes	23,144
EasyToEnjoy	19,061
CellySweepstakes	18,098
TimeoutForBuds	9,656
ULTRAClub	8,556
DoritosTriangleScheme	6,031
RealOrActing	5,685
AppleMusicHalftime	5,649

4.5 Trending Hashtags:

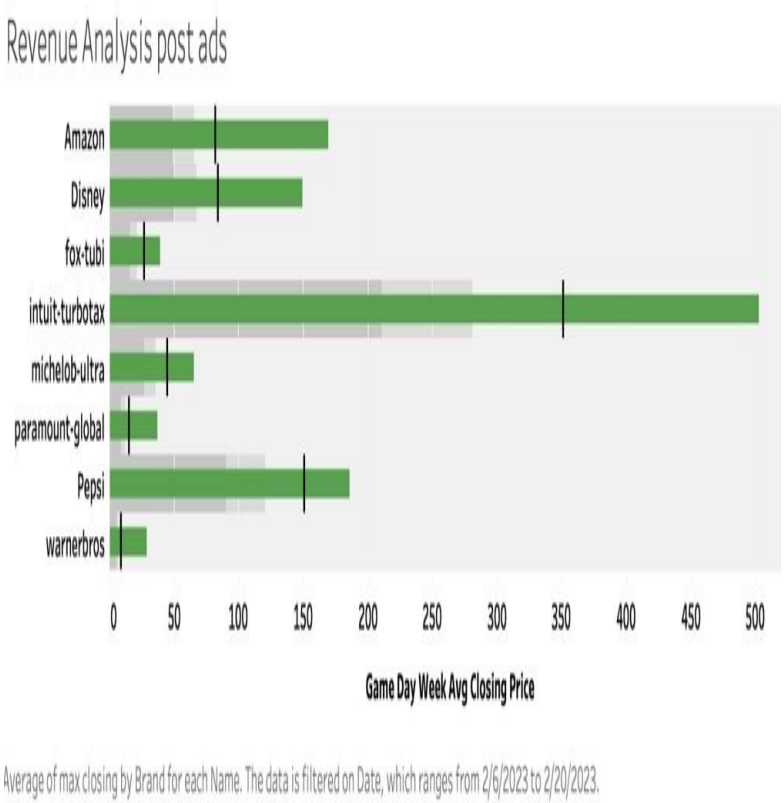
One of the most engaging activities on Twitter is the use of hashtags, where viewers can express their feelings and highlight specific aspects that caught their attention during the game. To identify the most trending hashtags during the Super Bowl 57, we cleaned the "text" column containing the text of each tweet and extracted the hashtags. We used Alteryx's regular expression feature to parse words starting with a pound symbol (#) and then counted the number of times each word appeared during the event. We presented our findings using a table to display the top 10 trending keywords.

4.6 Stock Price Trends

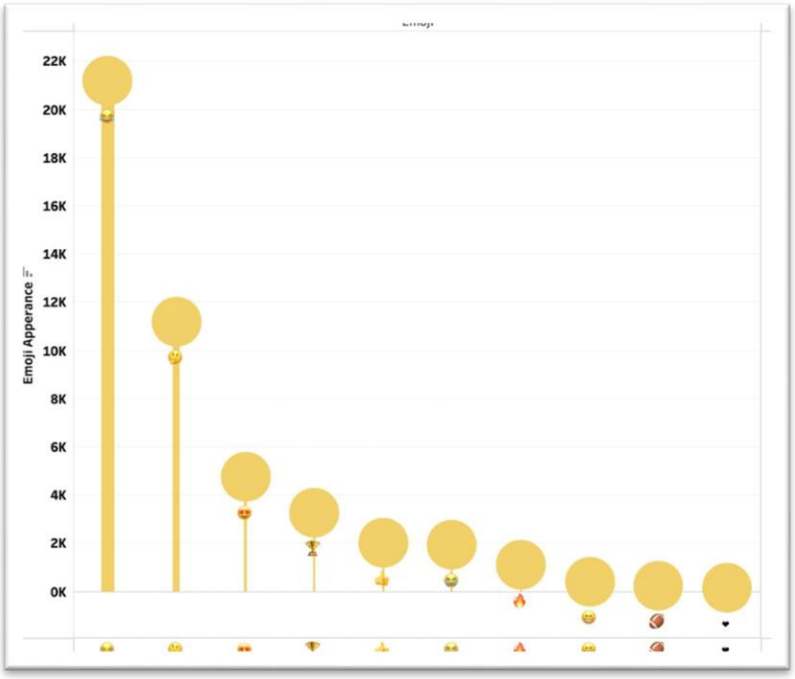
Many of the brands that advertised during the Super Bowl 57 are publicly listed on major stock exchanges in the US. To evaluate

the impact of the event on these brands, we gathered the average closing stock prices for the top 10 most tweeted brands before and after the Super Bowl, and visualized the differences using a bar graph created in Tableau.

We collected this data from Yahoo Finance website.



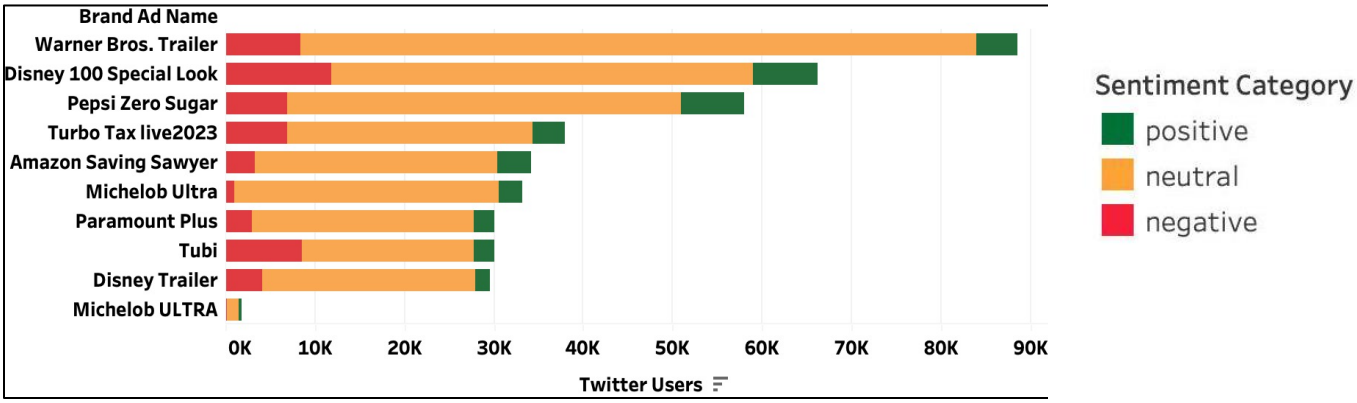
4.7 Emoji Usage Analysis:



Tweets are virtual expressions of viewers that are filled with feelings and emotions, which are best conveyed using emojis. To explore the emotions expressed during Super Bowl 57, we used the emoji package in Python to extract emojis from the "text" column in our dataset and determine how frequently each emoji appeared. We visualized the most commonly used emojis during the event, using the visualization mentioned above.

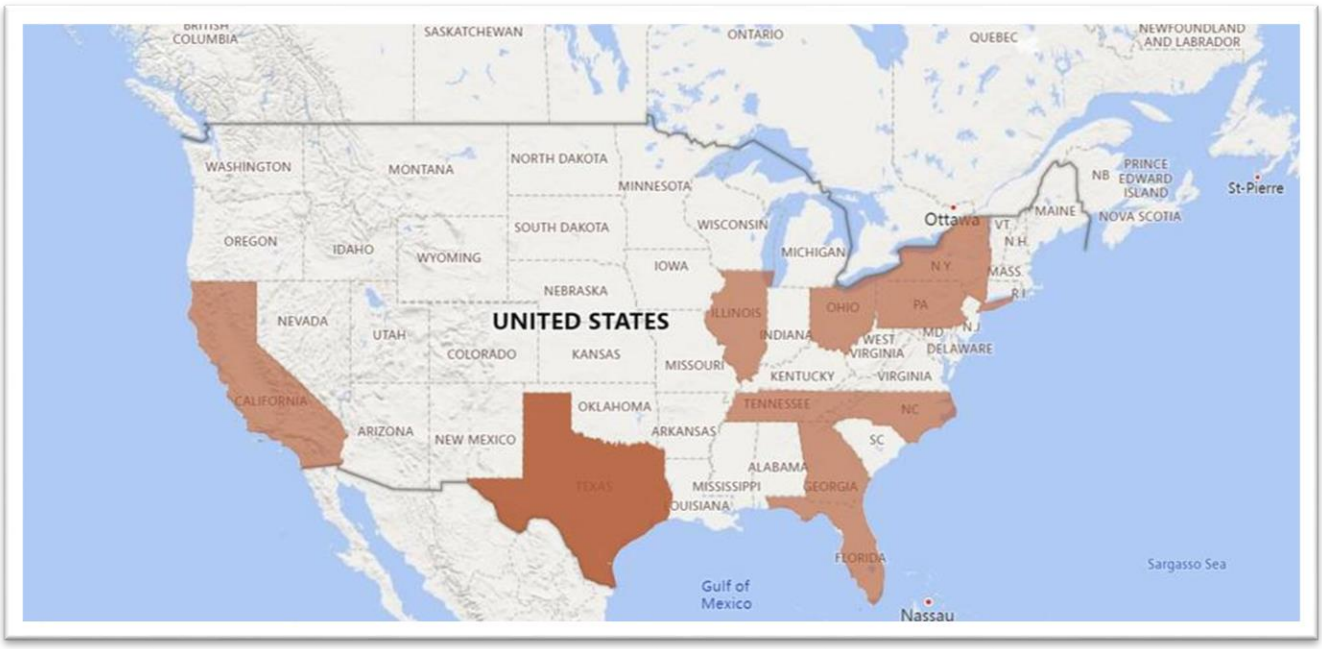
4.8 Sentiment Analysis and Brand Perception:

Sentiment analysis involves analyzing the sentiment or mood conveyed by the text in a given dataset. We conducted sentiment analysis on a clean dataset of tweets and found that over 50% of the tweets were neutral in nature. This analysis provides insight into the overall sentiment expressed by viewers during Super Bowl 57.



4.9 Most responsive States:

We conducted data cleaning and analysis on the "id" column to determine the states with the highest number of tweets during Super Bowl 57. Our findings showed that Texas had the most expressive state, followed by California and Pennsylvania. This information provides valuable insight into which states were most engaged with the event and its associated content.



5. CONCLUSION

5.1 Summary of Findings:

Based on our analysis of the Super Bowl 57 event, we have identified the following key findings:

1. The top 10 most trending ads during the event have been identified using a world cloud visualization.
2. The top 10 most tweeted brands during the event have been determined, with Warner Bros. being the most tweeted brand with around 89k tweets.
3. We have analyzed the engagement levels for each ad during the Super Bowl 57, and visualized the findings using a bubble chart in Tableau.
4. Using Alteryx's regular expression feature, we have identified the top 10 trending hashtags during the event.
5. We have conducted a stock price analysis on the top 10 tweeted brands before and after the Super Bowl, and visualized the difference using a bar graph created in Tableau.
6. Our emoji analysis has identified the most commonly used emojis during the Super Bowl 57.
7. Our sentiment analysis has revealed that over 50% of the tweets during the event were neutral in nature.
8. We have determined the top states with the highest number of tweets during the event, with Texas being the most expressive state followed by California and Pennsylvania.

These findings provide valuable insights into the viewers' engagement and sentiment during the Super Bowl 57 event, and can be used by brands to optimize their advertising strategies for future events.

5.2 Limitations of this study:

The limitations of this study include:

- The use of Twitter data for analysis may suffer from selection bias, as the opinions and sentiments expressed on Twitter may not be representative of the wider population. For example, Twitter users may be more likely to be younger, more urban, and more tech-savvy than the general population, which may influence their views on advertising and other topics. This means that the findings of this study may not be generalizable to the wider population.
- The study only focuses on the Superbowl game, which may not be representative of other sporting events or advertising campaigns. Different sporting events may have different audiences, advertising budgets, and promotional strategies, which may influence the effectiveness of advertising. Similarly, different advertising campaigns may have different target audiences, messages, and creative executions, which may affect their impact.
- The accuracy and reliability of the analysis are dependent on the completeness and accuracy of the Twitter data that was collected and analyzed. For example, if there were technical issues or gaps in

the data collection process, the analysis may be incomplete or biased. Similarly, if there were inaccuracies in the data cleaning or analysis process, the findings may be compromised.

- The study does not account for other factors that may influence the effectiveness of advertising during the Superbowl game, such as brand reputation, advertising budget, and the quality of the advertisement. For example, a well-established brand with a strong reputation may have an advantage over a lesser-known brand in terms of generating buzz on social media, even if their advertisement is not as effective. Similarly, a high-quality advertisement with a compelling message and creative execution may be more impactful than a low-quality advertisement with a weak message, even if the former does not generate as much social media buzz.

5.3 Future Scope:

Based on the findings of our study, there are several potential areas for future research and analysis. These include:

1. Conducting a deeper analysis of the engagement levels for each ad during the Super Bowl, such as by tracking user behavior before and after the ads.
2. Examining the impact of social media activity during the Super Bowl on the overall brand image and reputation of the brands that participated in the event.
3. Investigating the correlation between the top trending ads and the sentiment of the tweets during the Super Bowl, and the potential impact on viewer engagement.
4. Comparing the social media activity during Super Bowl 57 with that of previous years to identify trends and patterns in user behavior.
5. Conducting further analysis of the top states with the highest number of tweets during the Super Bowl to identify potential regional variations in viewer engagement and sentiment.

By conducting further research in these areas, we can gain a more comprehensive understanding of the impact of social media activity during major events such as the Super Bowl, and provide valuable insights for brands to optimize their marketing strategies in the future.

6. CHALLENGES FACED

- Keyword collection: Ensuring relevance and avoiding biases.
  - We had to ensure that relevant keywords are collected for the accuracy of the analysis.
  - We also had to ensure that the keywords were unbiased and representative of the most significant trends and themes.
- Data cleaning and preparation: Filtering and standardizing data for efficient analysis.
  - The task of cleaning and preparing a substantial dataset comprising 1.8 million rows for analysis. However, the large size of the file posed a challenge as it led to a significant decrease in the system's processing speed.
  - This involved filtering out any irrelevant or extraneous data points and standardizing the data to enable efficient analysis.
- Identifying trends and patterns: Creating charts and analyzing data to extract meaningful insights.
  - The process of creating charts and analyzing data to derive valuable insights necessitated an exhaustive data analysis approach that emphasized thoroughness, accuracy, and attention to detail.
- Mitigating potential biases and distortions: Ensuring the accuracy and reliability of the final analysis.
  - We had to be mindful of any potential biases and distortions in the data.
  - This involved carefully examining each data point and addressing any inconsistencies or inaccuracies to ensure the accuracy and reliability of the final analysis.
- Creating effective infographics and presentations: Communicating insights in a clear, engaging, and accessible manner.
  - We had to create effective infographics and a PowerPoint presentation that clearly and engagingly communicated the insights gleaned from the analysis.
  - A key aspect of this endeavor was the need to prioritize the design, structure, and overall effectiveness of the communication to facilitate easy comprehension and engagement with the findings among the intended audience. Given the abundance of data and analysis generated, deciding what to include and what to omit posed a challenge.

7. ACKNOWLEDGEMENT & THANKS

We would like to extend our sincere thanks and appreciation to our sponsors for their support during this competition. Their contributions have helped us to conduct research and analysis to produce valuable insights and recommendations.

Also, we would like to express our appreciation to our professors for their guidance and mentorship throughout the research process. Their expertise and feedback have been invaluable in helping us to refine our research methodology and analysis.

Finally, we would like to thank University of Utah for providing us with the resources and facilities necessary to conduct this analysis. The university's commitment to research excellence and innovation has been a driving force behind our efforts to produce a comprehensive and informative analysis.

Thank you once again to all of our sponsors, professors, and UofU for their support and dedication to this project. We could not have done it without them.

8. REFERENACES

1. Twitter. (n.d.). About Twitter. Retrieved from https://about.twitter.com/
2. Super Bowl Retrieved from https://www.nfl.com/super-bowl/
3. Alteryx. (n.d.). Alteryx Designer. Retrieved from https://www.alteryx.com/products/alteryx-platform/alteryx-designer
4. Tableau. (n.d.). Tableau Desktop. Retrieved from https://www.tableau.com/products/desktop
5. Tableau. (n.d.). Tableau Prep. Retrieved from https://www.tableau.com/products/prep
6. RStudio. (n.d.). RStudio. Retrieved from https://www.rstudio.com/
7. Python Software Foundation. (n.d.). Python. Retrieved from https://www.python.org/
8. Pandas. (n.d.). Pandas Documentation. Retrieved from https://pandas.pydata.org/docs/
9. Twitter Developer Documentation: https://developer.twitter.com/en/docs
10. Previous white paper and info graphs: https://utah.instructure.com/courses/866565/pages/past-white-paper?module_item_id=19567064