THE REAL GAME

SUPER BOWL ANALYSIS

BUSY MINIONS – UNDERGRAD 06

LINSEY YOONJI LEE

SARAH KIM

SEO JUN CHUNG

TABLE OF CONTENTS

INTRO

The Super Bowl is not only one of the most-watched sporting events in the world, but it is also a major platform for brands to showcase their advertisements. With 15 to 90 seconds **51 commercials ads** during the 2023 Super Bowl costing **$7 million for every 30 seconds**, it is imperative for brands to monitor the performance of their ads. While traditional mechanisms such as the USA Today AD METER ratings are still acceptable measuring yardsticks, social media channels like Facebook and Twitter provide digital footprints that offer an additional in-depth understanding of consumer feedback.

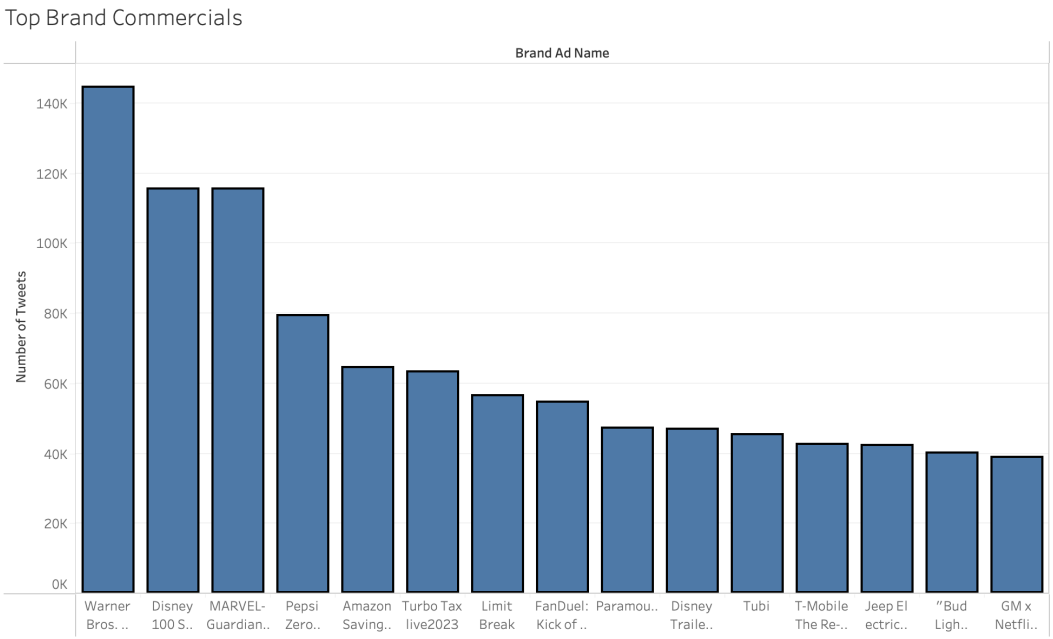
ASSUMPTIONS

This white paper will analyze the effectiveness of Super Bowl ads based on metadata collected from approximately **1.9 million Tweets** during the game and 1.5 hours afterward. Our analysis will provide insights that will help advertisers decide how to approach their Super Bowl ad spending in the future.

TOP BRAND COMMERCIALS

All of the companies advertise their products at the Super Bowl as one of their biggest marketing strategies. To see which ads performed the best on Twitter for the Super Bowl ads, we analyzed the amount of tweets tweeted for each of the 51 commercials featured during the Super Bowl.

TWEET VOLUME BY ADS



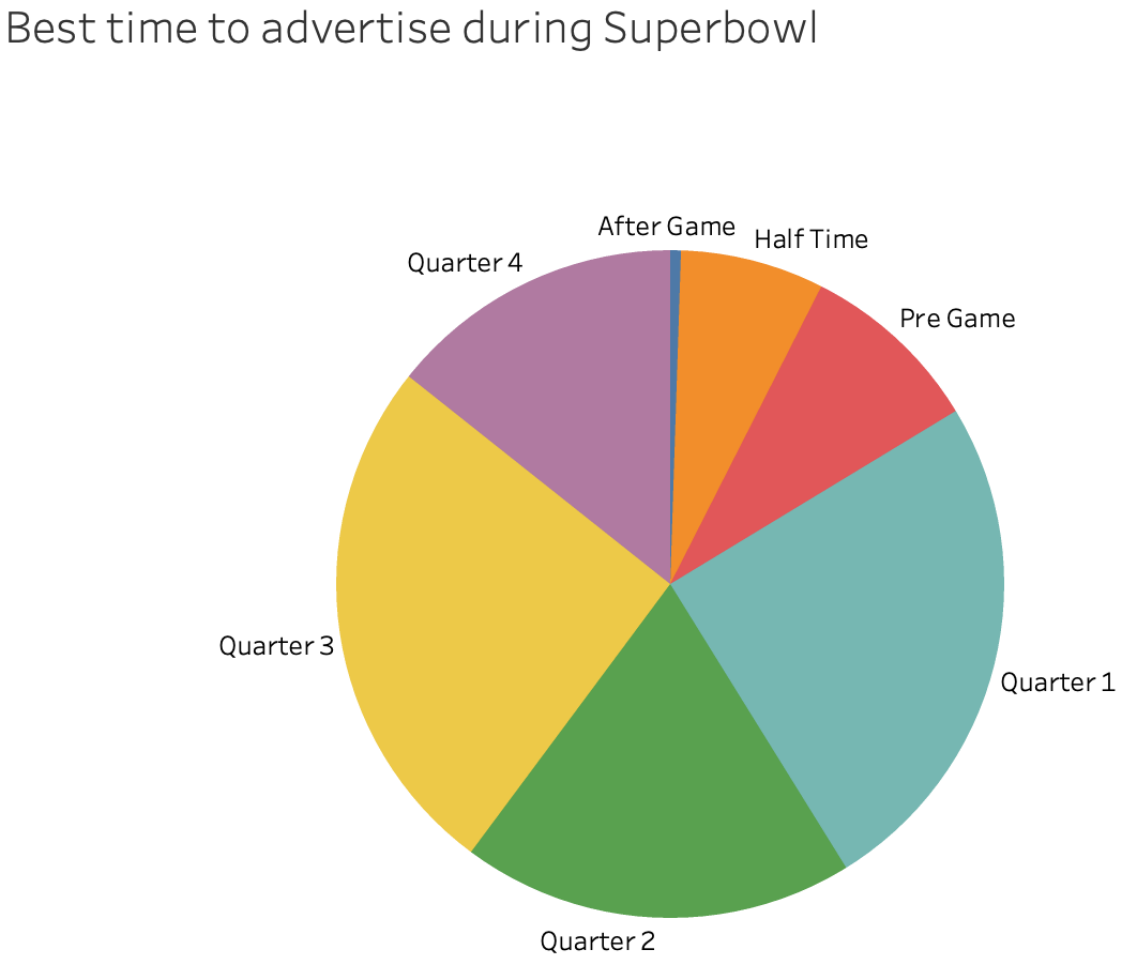
- **METHOD:**

1) Used Tableau Prep to Clean up the data so that only the Number of Tweets for the Super Bowl Ads were extracted
2) Used Tableau Desktop to create a bar graph to compare the Number of Tweets for each Brand Ad Name to observe which Ads performed the best on twitter

TOP 5 ADS BY TWEETS

- With the observation of the data and the graph above, we found that the top 5 Brands by the Number of Tweets were…
    - Warner Bros: The Flash - 144,205 tweets
    - Disney 100 Special Look - 115,311 tweets
    - Marvel: Guardians of the Galaxy Vol.3 - 115,118 tweets
    - Pepsi Zero Sugar: Ben Stiller - 79,030 tweets
    - Amazon: Saving Sawyer - 64,384 tweets

BEST TIME TO ADVERTISE



- **METHOD:**

1) Used Tableau Prep to Clean Up the data so that only the number of tweets that holds the value of either one of the 7 time fields were counted for.
2) Used Tableau Desktop to create a pie chart to display the Number of Tweets tweeted about the Super Bowl ads during each time frame of the game

ANALYSIS

- With the observation of the data and the graph above, we found that the best time to advertise during the Super Bowl is during the Quarter 1 and Quarter 3.
    - Pre-Game: 167,062
    - Quarter 1: 469,031
    - Quarter 2: 358,926
    - Half-Time: 131,417
    - Quarter 3: 481,796
    - Quarter 4: 270,243
    - After Game: 9,756
    - Total: 1,888,231
- With our observations, we can possibly infer that Quarter 1 is the best to advertise as it is when the game first begins with most viewer's focus entirely put on the game. Quarter 3 is also a great time to advertise as most people are watching the game and interacting on social media after watching the highlight of the Super Bowl, the Half-Time show.
- With similar reasoning, Quarter 4 has been seen to be the worst time to advertise, most likely because the people are more focused on which team is going to win the Super Bowl, or because some viewers have lost interest in the game and/or the commercials.

TOP KEYWORDS

Tweets to wordcloud is a great way to analyze the language people use on Twitter. Word clouds provide a quick and visual representation of the topics and sentiments being discussed.

In this context, we analyzed tweets during the SuperBowl to create two word clouds, one highlighting the most frequently used words and the other showing the most popular hashtags.

TOP WORDS MENTIONED

- **METHOD:**

1) Used Python in JupyterNotebook to clean data and save it as a new CSV file.

a)

```
df.dropna(subset=['text'], inplace=True)  # delete NaN
df_clean = df[df['text'].str.contains('id')].index
df.drop(df_clean, inplace=True)
```

```
df_clean = df[df['text'].str.contains('type')].index
df.drop(df_clean,inplace=True)
```

```
df_clean = df[df['text'].str.contains('https')].index
df.drop(df_clean,inplace=True)
```

2) Use the STOPWORDS function to create a word cloud

a)

```
stopwords = set(STOPWORDS)
mask = np.array(Image.open('twitter.png'))

#we will read the text
df = pd.read_csv('cleaned.csv')

def data_processing(text):
    text = text.lower()
    text = re.sub(r"https\S+|www\S+https\S+", '',text, flags=re.MULTILINE)
    text = re.sub(r'[^\w\s]','',text)
    text_tokens = word_tokenize(text)
    filtered_text = [w for w in text_tokens if not w in stop_words]
    return " ".join(filtered_text)

#wordcloud
wordcloud = WordCloud(stopwords = stopwords , width=1600 , height=800,mask=mask,background_color="Black",colormap="Set2
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)

#saving the image of wordcloud
wordcloud.to_file ('wordcloud.png')
plt.show()
```

TOP HASHTAGS MENTIONED

- **METHOD:**

1) Use the same dataset, erase emotions from the 'text' column

a)

```
# Define the regular expression to match emoticons
emoticon_pattern = re.compile("["u"\U0001F600-\U0001F64F"  # emoticons
                              u"\U0001F300-\U0001F5FF"  # symbols & pictographs
                              u"\U0001F680-\U0001F6FF"  # transport & map symbols
                              u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
                              u"\U00002702-\U000027B0"  # other emoticons
                              u"\U000024C2-\U0001F251""]+", re.UNICODE)

# Remove emoticons from the text column
df['text'] = df['text'].apply(lambda x: emoticon_pattern.sub(r'', x))
```

2) Extract the top 30 words with hashtags

a)

```
# Extract words with hashtags from the 'text' column and store in a new column 'words_with_hashtags'
df['words_with_hashtags'] = df['text'].apply(lambda x: [word for word in x.split() if '#' in word])

# Flatten the list of words with hashtags and create a new dataframe with two columns 'word' and 'count'
words_with_hashtags_df = pd.DataFrame(df['words_with_hashtags'].sum(), columns=['word'])
words_with_hashtags_df['count'] = 1

# Group the words by count and sort in descending order to get the top 10 most popular words with hashtags
top_30_words_with_hashtags = words_with_hashtags_df.groupby('word').count().sort_values('count', ascending=False)[:30]

print(top_30_words_with_hashtags)
```

```
                          count
word
#SuperBowl                11304
#ZeroDebate               10749
#RealOrActing              5620
#GreatActing               3891
#SuperBowlLVII             2620
#TheFlashMovie             1826
#GreatTaste                1669
#TheFlash                  1139
#                           861
#Sweepstakes                693
#SBLVII                     683
#SuperBowl2023              628
#SuperBowlCommercials       567
#Rihanna                    559
#HalftimeShow               540
#eBay                       401
#ChiefsKingdom              387
#DoritosTriangleScheme      384
#SuperBowlAds               380
#NFL                        344
#NextLevelChef              333
"#SuperBowl                 322
#Disneyland                 321
#GuardiansOfTheGalaxyVol3   305
#artificialintelligence     304
#ai                         304
#TimeoutForBuds             294
#nationalsecurityagency     294
#parody                     294
#nsa                        294
```

3) Use the STOPWORDS function to create a word cloud

a)

```
stopwords = set(STOPWORDS)
mask = np.array(Image.open('twitter.png'))

#we will read the text
df = pd.read_csv('top_30_words_with_hashtags.csv')

def data_processing(text):
    text = text.lower()
    text = re.sub(r"https\S+|www\S+https\S+", '',text, flags=re.MULTILINE)
    text = re.sub(r'[^\w\s]','',text)
    text_tokens = word_tokenize(text)
    filtered_text = [w for w in text_tokens if not w in stop_words]
    return " ".join(filtered_text)

#wordcloud
wordcloud = WordCloud(stopwords = stopwords , width=1600 , height=800,mask=mask,background_color="Black",colormap="Set2
plt.figure(figsize=(20,10),facecolor='k')
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.tight_layout (pad=0)

#saving the image of wordcloud
wordcloud.to_file ('wordcloud2.png')
plt.show()
```

ANALYSIS

- Top Brands that included a hashtag in their ads & hash:
  - Warner Bros: #TheFlashMovie(1826) #TheFlash(1139)
  - Disney: #Disneyland(321)
  - Marvel: #GuardiansOfTheGalaxyVol3(305)
  - PepSi: #ZeroDebate(10749) #RealOrActing(5620) #GreatActing(3891) #GreatTaste(1669)

- Which Ad Had the Most Tweeted Keywords?
  - Our data analysis demonstrates that the #ZeroDebate hashtag, associated with the Pepsi ad, had the highest keywords used in tweets. This finding implies that the Pepsi ad was particularly effective in engaging social media users, as evidenced by the high level of discussion generated by the #ZeroDebate hashtag

TOP BRAND SENTIMENT

For brand owners, the primary goal of advertising is not only to reach a large audience and leave them with a positive sentiment but also to create a lasting association with the brand in the minds of customers. By conducting top brand sentiment tweets analysis, companies can identify how their customers perceive their brand in comparison to their competitors.

To measure the sentiment, we have defined the three categories of metrics: positive, neutral, and negative. This analysis can help brand owners develop effective marketing strategies that enhance their brand image and ensure that customers associate their brand with positive feelings.

SENTIMENT BY EACH BRAND

- **METHOD:**

1) Used Python in JupyterNotebook to clean data and save it as a new CSV file.
2) Extract the exact brand name in the dataset and create a new CSV file

```
df = pd.read_csv('brandtextclean.csv')
```

```
df.head(10)
```

	brand_ad_name	text
0	Bass Pro Shops	"2023 MAYFEST planning has begun! Apply now to...
1	'url': 'https://t.co/CSipnt5TKp'	'media_key': '3_1624986202037157893'}]"
2	"[{'start': 3	0.0
3	Bass Pro Shops	"50 ವರ್ಷಗಳ ಹಿಂದೆ: ಮಂಗಳವಾರ
4	'url': 'https://t.co/hWk7fe9vLM'	13–2–1973'
5	Bass Pro Shops	"#Golden #jubilee #weddinganniversary #celebra...
6	'url': 'https://t.co/RL3BYwhkn1'	'description': '#goldenjubilee of #weddinganni...
7	"[{'start': 3	0.0
8	Bass Pro Shops	"Oloriburuku ni omo Marvin yi
9	"[{'start': 3	0.0

```
df.dropna(subset=['text'], inplace=True)  # Remove NaN value in 'text'column
```

```
df = df.dropna()

# "Warner Bros. Trailer: The Flash"included row extract "text" column
new_df = df.loc[df['brand_ad_name'].str.contains('Pepsi Zero Sugar: Ben Stiller-Great Acting or Great Taste'), 'text']
print(new_df)
```

```
270                                        "@Snehaaahere
276                Ngl I really miss her. Just her friendship
277         Huge meltdown of Priyanka fans after she finis...
278         RT @DeJLoaf: No one can break up a true bond/f...
279                 "Mozart somefascistregimeAnd administered
                                  ...
809502      i hate patrick mahomes i think we should shove...
809510                              Mahomes is from outer space
809513                              Someone start the space lol
809521                                                    "Oh
809523      "Heavens Slipping Advice space.""Lincoln 'aspi...
Name: text, Length: 16956, dtype: object
```

a)

```
new_df.to_csv("pepsi.csv", index = False)
```

3) Do some text preprocessing steps

```
text_df = pd.read_csv('flash.csv')
```

```
def data_processing(text):
    text = text.lower()
    text = re.sub(r"https\S+|www\S+https\S+", '',text, flags=re.MULTILINE)
    text = re.sub(r'[^\w\s]','',text)
    text_tokens = word_tokenize(text)
    filtered_text = [w for w in text_tokens if not w in stop_words]
    return " ".join(filtered_text)
```

```
stemmer = PorterStemmer()
def stemming(data):
    text = [stemmer.stem(word) for word in data]
    return data
```

a)

```
text_df['text'] = text_df['text'].apply(lambda x: stemming(x))
```

4) Use the TextBlob library to calculate the polarity score of each text and then maps the polarity score to a string label representing the sentiment of the text.

```
def polarity(text):
    return TextBlob(text).sentiment.polarity
```

```
text_df['polarity'] = text_df['text'].apply(polarity)
```

```
def sentiment(label):
    if label <0:
        return "Negative"
    elif label ==0:
        return "Neutral"
    elif label>0:
        return "Positive"
```

```
text_df['sentiment'] = text_df['polarity'].apply(sentiment)
```

```
text_df.head(10)
```

a)

	text	polarity	sentiment
0	"🔥CARAJO	0.000	Neutral
1	"El mejor #Batimovil de todos los tiempos #Bat...	0.000	Neutral
2	"HE SAYS THE WORDS. HE ACTUALLY SAYS THEM.	0.000	Neutral
3	""Yeah…	0.000	Neutral
4	"Ladies and gentlemen of the world...	0.000	Neutral
5	"人気の同人⭐	0.000	Neutral
6	DC는 에즈라 밀러 교체할 결단력을 배트걸 폐기할 때 다 써버린 것입니다	0.000	Neutral
7	"Ezra Miller Is INNOCENT ?! + Ezra Will Remain...	0.625	Positive
8	"I'm Batman!!! PORRA	0.000	Neutral
9	"o quão inamorável vc é?	0.000	Neutral

i) The polarity function takes a string of text as input and returns the polarity score of the text using the TextBlob library. The polarity score ranges from -1 to 1, where -1 represents a very negative sentiment and 1 represents a very positive sentiment.

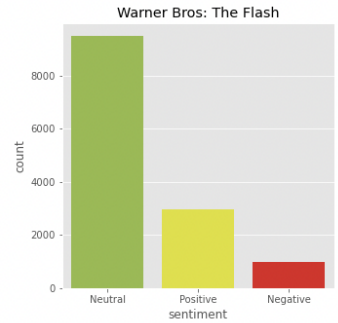
ii) The sentiment function takes a numerical polarity score as input and returns a string label representing the sentiment of the text. If the polarity score is less than 0, the function returns "Negative"; if it is equal to 0, the function returns "Neutral"; and if it is greater than 0, the function returns "Positive".

5) Make two charts

a)

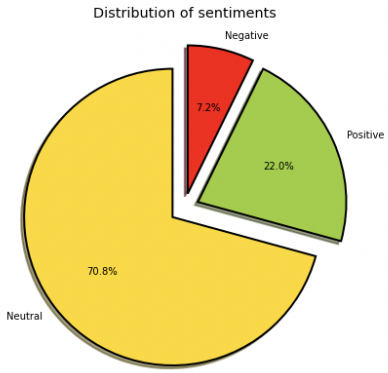
```
fig = plt.figure(figsize=(5,5))
colors = ["yellowgreen", "yellow", "red"] # specify colors for Negative, Neutral, Positive categories
sns.countplot(x='sentiment', data=text_df, palette=colors)
plt.title('Warner Bros: The Flash')
```

```
Text(0.5, 1.0, 'Warner Bros: The Flash')
```



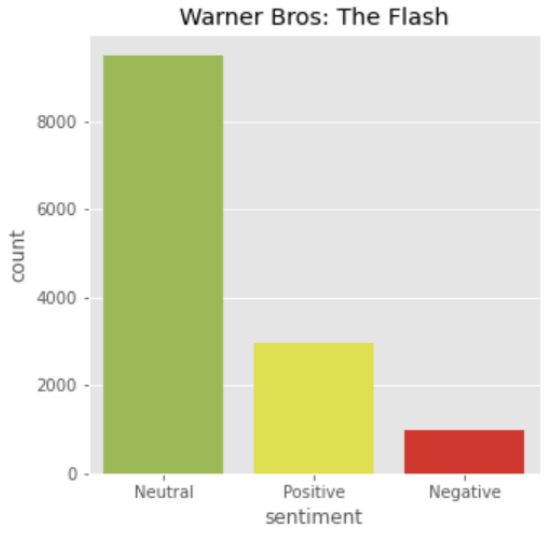

```
fig = plt.figure(figsize=(7,7))
colors = ("gold", "yellowgreen", "red")
wp = {'linewidth':2, 'edgecolor':"black"}
tags = text_df['sentiment'].value_counts()
explode = (0.1,0.1,0.1)
tags.plot(kind='pie', autopct='%1.1f%%', shadow=True, colors = colors,
          startangle=90, wedgeprops = wp, explode = explode, label='')
plt.title('Distribution of sentiments')
```

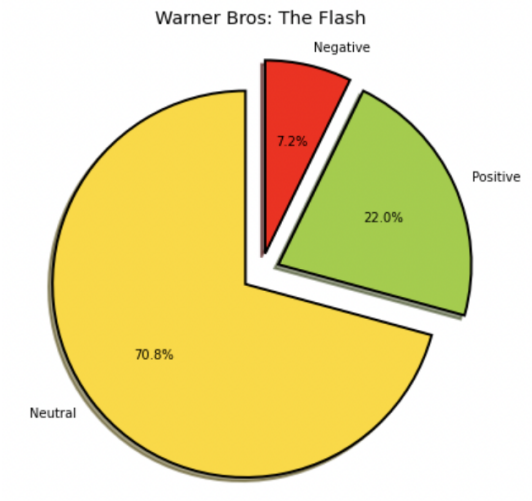
```
Text(0.5, 1.0, 'Distribution of sentiments')
```



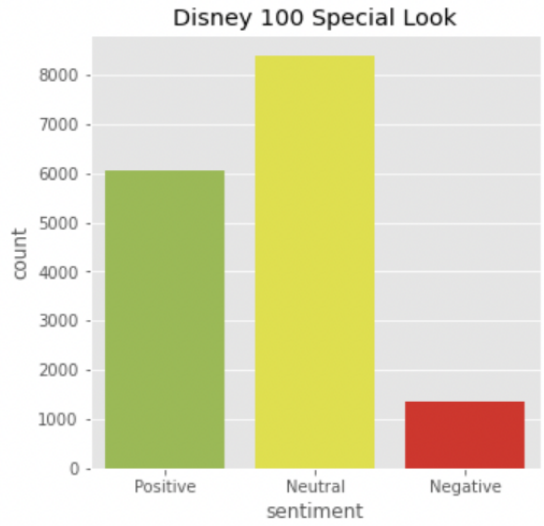
6) Do the same thing for top 4 brands

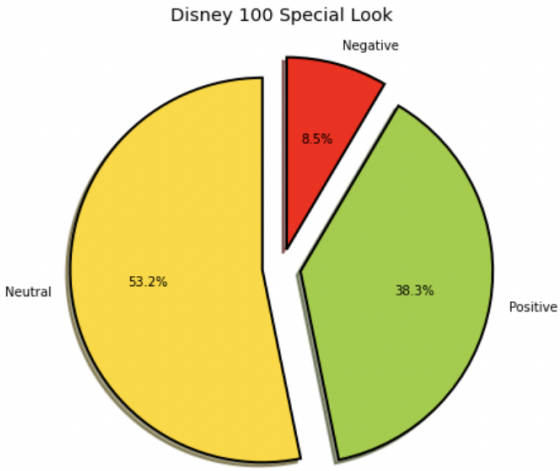
Warner Bros: The flash



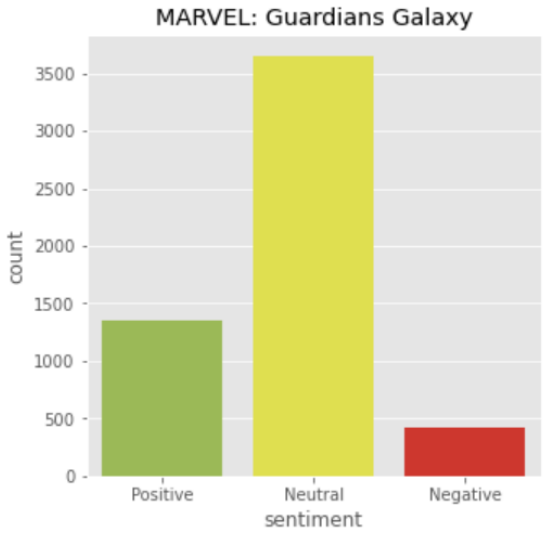


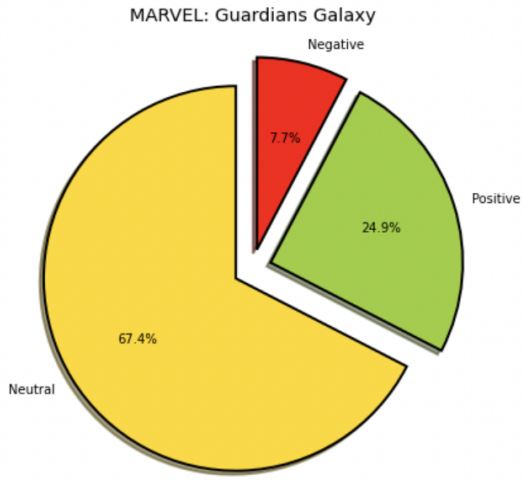
Disney 100 Special Look



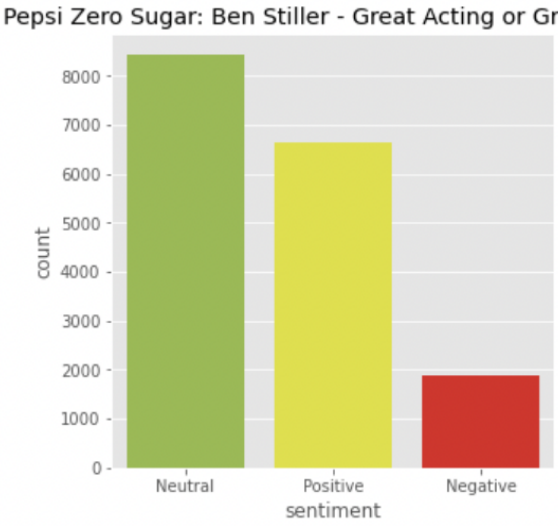


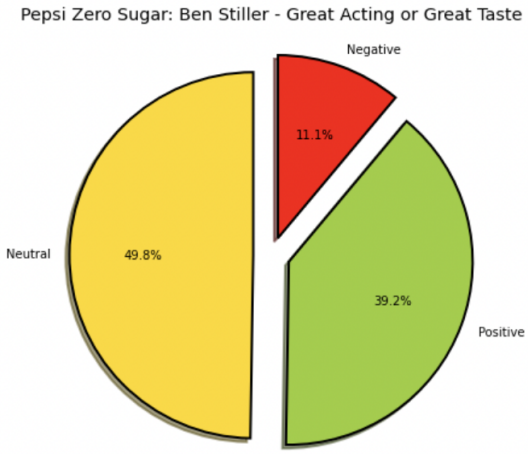
Marvel: Guardians Galaxy





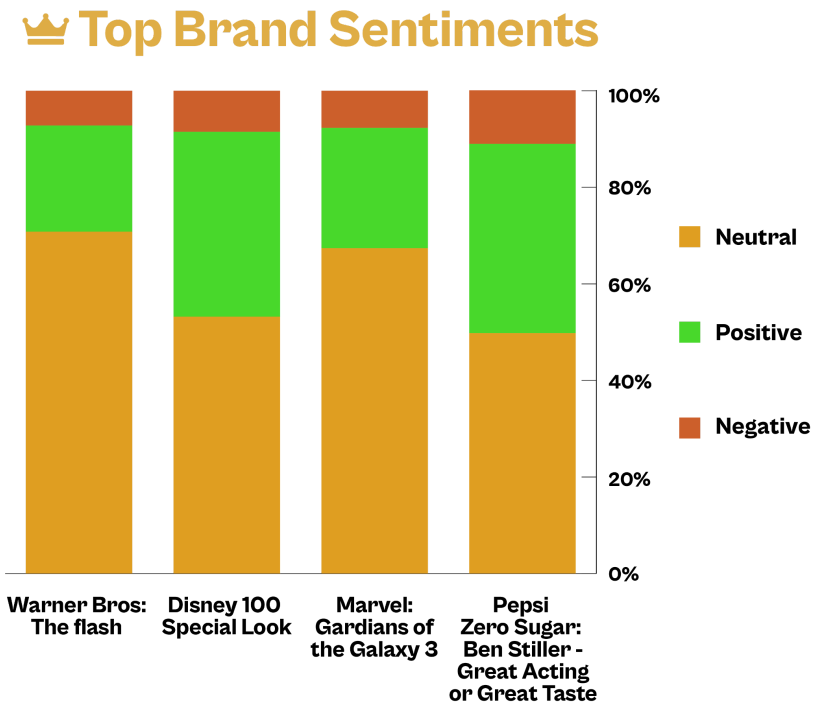
Pepsi Zero Sugar: Ben Stiller - Great Acting or Great Taste





BRAND COMPARISON

We combined each graph into one for easy comparison.



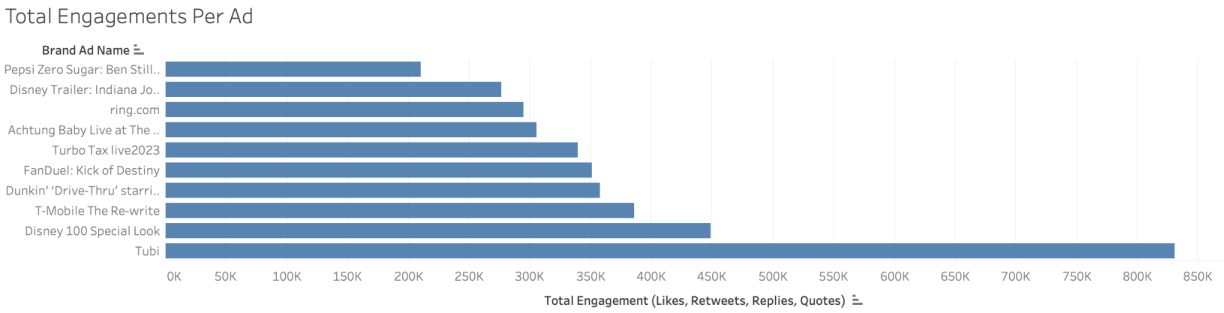
ANALYSIS

- The Pepsi Zero Sugar ad received the highest positive response at 39.2%, while Warner Bros' movie "The Flash" received the lowest positive response at 22%.
- Based on the data and the sentiment graphs of the top brands above, we infer that brands that are the most familiar to the audiences hold the most positive responses. Disney, Marvel, Warner Bros, and Pepsi all are brands very well known to the American audience and all hold characters that have been familiar to us for a long time.

TOTAL ENGAGEMENTS

Not only is it important to focus on the number of tweets tweeted about each ad but it is also important to focus on other types of engagements done on Twitter posts like the number of Likes, Retweets, Replies, and Quotes

LIKES, RETWEETS, REPLIES AND QUOTES

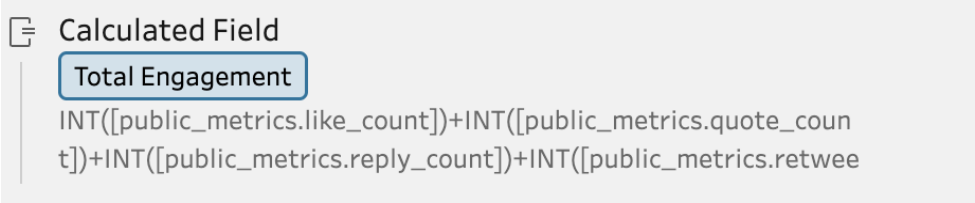


- **METHOD:**

1) Used Tableau Prep to Clean up the data so that the messy data for the public_metrics.retweet_count, reply_count, like_count, and quote_count.

a)

2) Then continue using Tableau Prep to find the Total Engagements for each ad by adding all the values of Likes, Retweets, Replies, and Quotes.
   a)



3) Using Tableau Desktop, we then create a bar graph representing the Total Engagements of each ads, and displayed the Top 10 ads with the most engagements for the Super Bowl

TOP 10 ENGAGEMENTS

- With the observation of the data and the graph above, we found that the Top 10 Ads by the Number of Engagements were…
  - Tubi: 830,490
  - Disney: 448,720
  - T-mobile: 386,240
  - Dunkin: 358,000
  - FanDuel: 351,360
  - TurboTax: 339,660
  - Achtung: 305,410
  - Ring.com: 294,420
  - Indiana Jones: 276,500
  - Pepsi Zero: 210,230

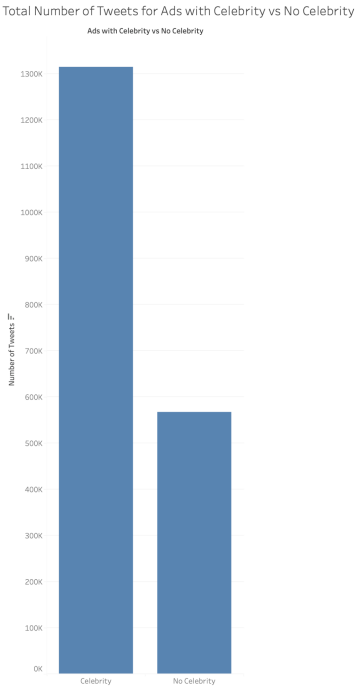
ANALYSIS

- This data serves a significant purpose as the Top 10 Ads for the Number of Engagements appear to be slightly different from the Top 10 Ads for the Number of Tweets.
- Tubi held the highest number of engagements, despite not even being on the Top 10 list for the Number of Tweets, most likely because of the unique ad Tubi performed by tricking many of the audience members that their TV was being controlled by someone else. Proving that factors like celebrities, comedy, or compassion isn't the only source of factor that can determine how well an advertisement performs.

CELEBRITIES

Although there are many factors in an advertisement that can boost its marketing effect, many companies take advantage of the fame of celebrities from famous movies, tv shows, sports, or the music industry to ensure that the advertisement reaches its target audience with the maximum positive marketing effect.

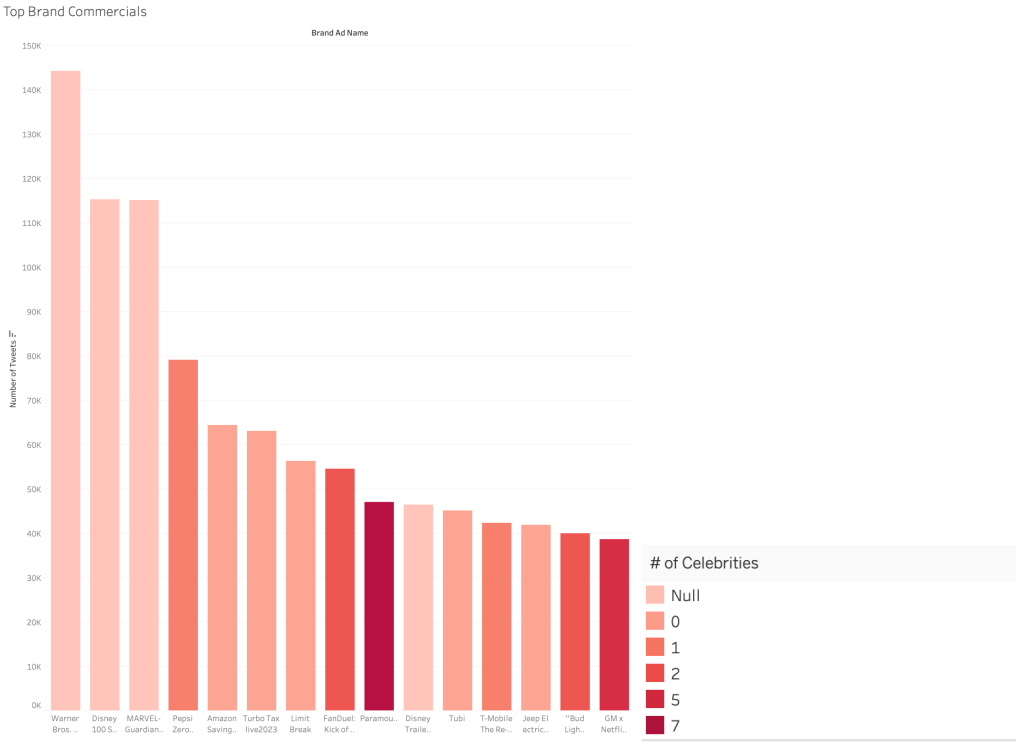
AD WITH & WITHOUT CELEBRITIES



- **METHOD:**

1) Used Tableau Prep to Clean up the data so that only the Number of Tweets for the Super Bowl Ads were extracted
2) Used Tableau Desktop to split the data for each 51 ads depending on whether or not a celebrity has appeared in its ad
3) Used Tableau Desktop to then create a bar graph comparing the Total Sum of the number of tweets tweeted for the ads featuring a celebrity and ads not featuring a celebrity

CORRELATION BETWEEN TWEETS & CELEBS



- **METHOD:**

1) Used Tableau Prep to Clean up the data so that only the Number of Tweets for the Super Bowl Ads were extracted
2) Used Tableau Desktop to give each advertisement a new table of values that determines the number of celebrities featured in the ad
3) Used Tableau Desktop to then create a bar graph comparing the Total Sum of the number of tweets tweeted for the ads while indicating how many celebrities were featured in it with the use of a color palette.
    a) Darker color indicates that more celebrities were featured
    b) Null values indicate that the ad was for a movie, meaning that countless numbers of celebrities were featured in the ad

ANALYSIS

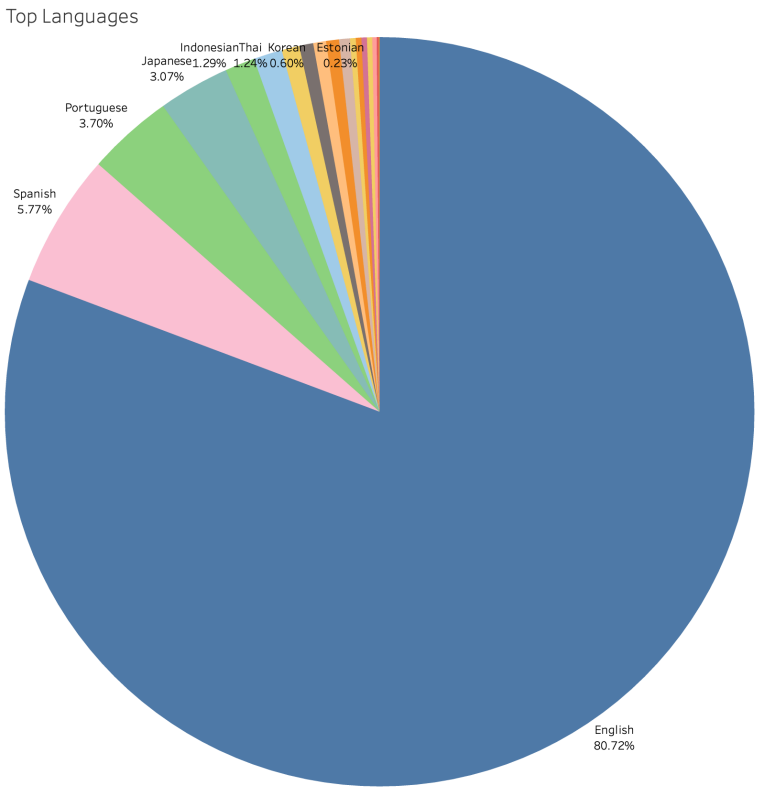
- With the observation of the data and the graph above, we found that the more tweets were tweeted for ads that featured a celebrity. However, for ads, with the exception of movie trailers, was not much affected by the number of celebrities.
- So we can infer that although it may be important for companies to use celebrities in their commercials, it is not necessary for them to use a large amount of celebrities, as Super Bowl commercials are all

relatively short and is more important to create a strong and persuasive impression with the minimum amount of information and factors for the audience.

LANGUAGES

As the Super Bowl is the biggest sports event for American Football, it is most likely that English will be the most commonly used language. So it will be important for companies to mainly focus on English speakers with their ads. However, that does not mean that companies should exclude other language speakers, so we will analyze what other languages are spoken on Twitter for the Super Bowl ads to see which shadow audiences companies should consider targeting.

TOP LANGUAGES



- **METHOD:**

3) Used Tableau Prep to Clean up the data so that only the Number of Tweets for the Super Bowl Ads and the Top Languages used are all valid data
4) Used Tableau Desktop to create a pie chart to compare the percentage of languages used on the tweets related to the Super Bowl ads

ANALYSIS

- With the observation of the data and the graph above, we found that the top language used during the Super Bowl was, as expected, English, at 80.72%, with other languages like Spanish, Portuguese, and Japanese being the next top 3 languages used.
- We can infer that Spanish was the next most used language for the fact that the Hispanic population is the biggest population in America that is known to also use a language different from English, making it pretty obvious data and knowledge for companies to take advantage of. But with American Football not being too popular in Portugal and Japan, companies most likely would not have expected this result, so companies for next year's Super Bowl may want to take in consideration of targeting audiences from those language regions.

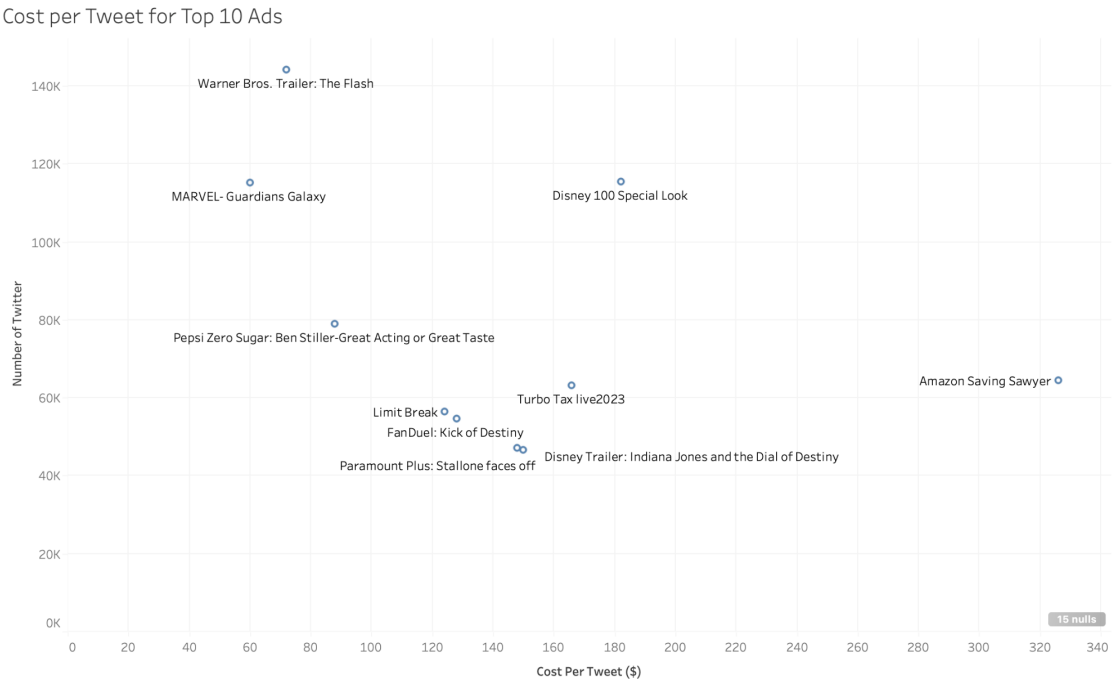
RETURN ON INVESTMENT

No matter how well a company's ad is engaged upon on Twitter, it doesn't really matter if they spent several times more than companies that performed similarly to gain that outcome, which is why the return on investment is an important factor to take in consideration when analyzing Super Bowl commercials.

Ad Name	Tweets	AirTime (sec)	Total Cost	Cost per Tweet
Warner Bros. Trailer : The Flash	144,205	45	$10,500,000	$72.81
Disney 100 Special Look	115,311	90	$21,000,000	$182.12
Marvel - Guardians of Galaxy Vol.3	115,118	30	$7,000,000	$60.81
Pepsi Zero: Ben Stiller - Great Acting or Great Taste	79,030	30	$7,000,000	$88.57
Amazon Saving Sawyer	64,384	90	$21,000,000	$326.17
Turbo Tax live 2023	63,129	45	10,500,000	$166.33
Limit Break	56,299	30	$7,000,000	$124.34
FanDuel: Kick of Destiny	54,496	30	$7,000,000	$128.45
Paramount Plus: Stallone Faces off	46,984	30	$7,000,000	$148.99
Disney Trailer: Indiana Jones and the Dial of Destiny	46,513	30	$7,000,000	$150.50
Tubi	45,111	15	$3,500,000	$77.59

T-Mobile The Rewrite	42,294	60	$14,000,000	$331.02
Jeep Electric Boogie	41,901	60	$14,000,000	$334.12
Bud Light: Easy to Drink	39,976	60	$14,000,000	$350.21
GM x Netflix: EVs on Screen	38,667	60	$14,000,000	$362.07
Michelob Ultra: Full Swing Gossip	37,624	30	$7,000,000	$186.05
RAM Premature Electrification	36,850	60	$14,000,000	$379.92
Dunkin' 'Drive-Thru' starring Ben	36,411	30	$7,000,000	$192.25
Heinz - LVII Meanz 57	36,371	60	$14,000,000	$384.92
Pop Corner: Breaking Bad reunion	35,546	30	$7,000,000	$196.93
DraftKing: Kevin Hart-Free bet	35,037	30	$7,000,000	$199.79
Squarespace: The Singularity	29,496	30	$7,000,000	$237.32
Crown Royal x Super Bowl LVII- Thank you Canada	27,279	60	$14,000,000	$513.22
Pepsi Zero Sugar: Steve Martin - Great Acting or Great Taste	24,390	30	$7,000,000	$287.00
H&R Block	24,089	60	$14,000,000	$581.18

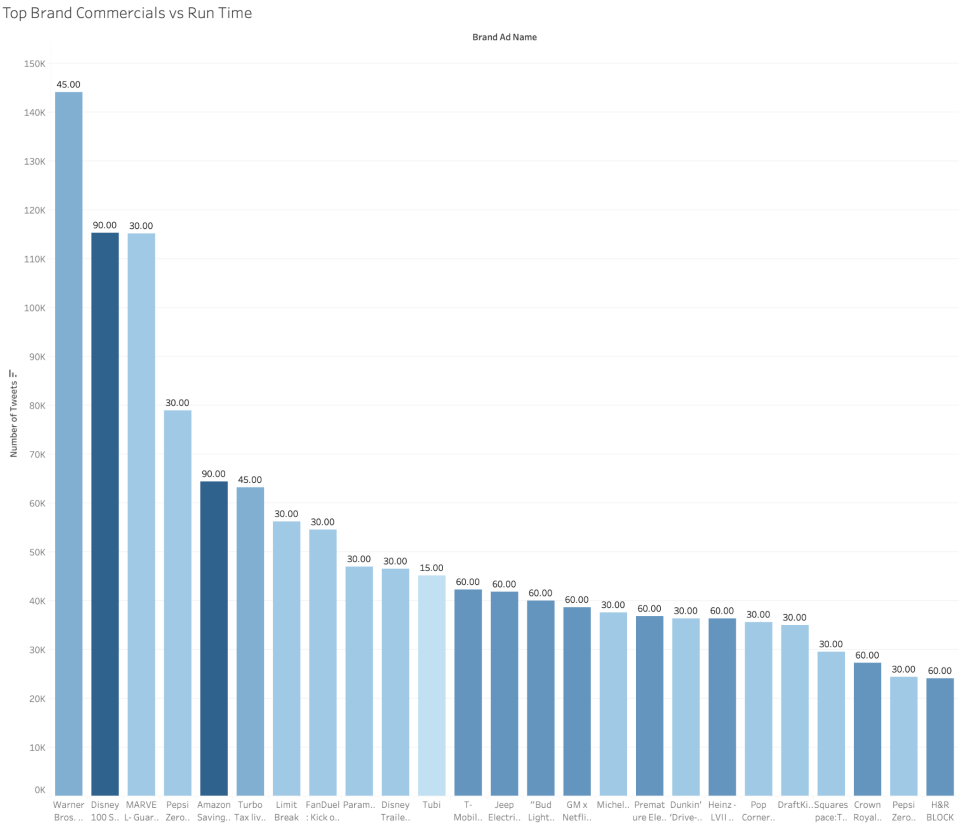
COST PER TWEET



- **METHOD:**

1) Used Tableau Prep to Clean up the data so that only the valid Number of Tweets for the Super Bowl Ads were extracted
2) Searched online to find the Airtime for each ad to calculate the Cost for each ad and the Cost per Tweet for each ad as well
3) Used Tableau Desktop to create a scatter plot comparing the number of tweets compared to the cost per tweet ($) for the top 10 ads.

AIRTIME



ANALYSIS

- With the observation of the data and the graph above, we found that the companies that had the best ROI value were…
    - Marvel - Guardians of Galaxy Vol.3 - $60.81 per tweet
    - Warner Bros - The Flash - $72.81 per tweet
    - Tubi - $77.59 per tweet
- and the companies that had the worst ROI value were…
    - H&R Block - $581.18
    - Crown Royal - $513.22
    - Heinz - $384.92
- We can infer that movie trailers perform the best for the amount of money invested as Marvel and Warner Bros' two movie trailers performed for the best ROI value. And we can infer that Tubi has performed exceptionally well, especially for a commercial that did not take advantage of a celebrity, and for being only 15 seconds long, with its unique ideas in tricking the audiences in a fresh way that no other company has attempted.