**Machine Learning in Marketing**

**Supercar Project**

Table of Contents

[**1.**](#_heading=h.gjdgxs) **Introduction** 3

[**2.**](#_heading=h.30j0zll) **Methodology** 4

[**3.**](#_heading=h.1fob9te) **Introduction of the supercars championship** 4

[**4.**](#_heading=h.3znysh7) **Topics that most effective in motivating the audience to engage more** 4

[**4.1**](#_heading=h.2et92p0) **Overall event 11 word cloud** 4

[**4.2**](#_heading=h.tyjcwt) **Positive sentiment analysis** 5

[**4.3**](#_heading=h.3dy6vkm) **Negative sentiment analysis** 8

[**5.**](#_heading=h.1t3h5sf) **Factors that influence TV ratings and their impact on fans' engagement** 10

[**6.**](#_heading=h.4d34og8) **Recommendations** 12

[**7.**](#_heading=h.2s8eyo1) **Limitation** 14

[**8.**](#_heading=h.17dp8vu) **Conclusion** 14

1. **Introduction**

The use of Twitter as a platform for experiencing events and receiving commentary is growing in popularity among sports enthusiasts (Smith et al., 2019). Research findings suggest that sports fans are inclined to actively engage with sporting events by simultaneously watching them on television and posting on Twitter (Murphy, 2018).

A picture containing text, screenshot, font, diagram

Description automatically generated

Figure 1. Twitter makes watching sports on TV more engaging & memorable (Murphy, 2018)

Therefore, the aim of this report is to examine the correlation between television viewing and fans' engagement on Twitter, with a focus on addressing the following two research questions:

Q1: Which topics are most effective in motivating the audience to engage more?

Q2: What factors influence TV ratings and their impact on fans' engagement?

Through this investigation, we aim to identify the factors that significantly impact fans' engagement and shed light on effective strategies to enhance audience interaction.

1. **Methodology**

This report initiates the application of sentiment analysis to classify tweets into positive or negative categories. Subsequently, topic modelling techniques were employed to examine the specific topics that were discussed in these distinct categories of tweets. By analysing both positive and negative sentiments, we aimed to gain insights into the themes and subjects being discussed in relation to the supercars event. Furthermore, regression analysis is conducted to explore the relationship between TV ratings and various relevant variables. Finally, based on the insights gained from the analyses, recommendations and conclusions will be formulated and presented.

1. **Introduction of the supercars championship**

The Repco Supercars Championship is a touring car racing event in Australia and there are 14 racing events every year taking place within Australia. As the premier motorsport category in Australasia and one of Australia’s biggest sports. It also recognised globally as the best touring car category and a leader in motorsport entertainment. With their own social media account, they developed a better channel to interact with their target audience. They have 139K Followers on Twitter, and countless posts in the hashtag under #supercars and #vasc.

Among all the races, BATHURST1000 is the one that got the most attention. This iconic motorsport event is viewed as the most prestigious and challenging touring car race, formerly known as the V8. With over 9000 tweets, Bathurst 1000 as event 11 in our given dataset, accounted for most of the tweets. Therefore, we think it’s the most important event to introduce and analyse for the live motorsport event.

1. **Topics that most effective in motivating the audience to engage more**
   1. **Overall event 11 word cloud**



Figure 2. Overall Word Cloud for Event 11

From the overall word cloud for event 11, we can see that the name of the winners ‘Chaz Mostert’ and ‘Paul Morris’, the name of second player ‘Whincup’ are the most frequent words mentioned in the word cloud, there are also many tweets that mentioned the words like ‘win’, ‘finish’, ‘day’, ‘car’, ‘fuel’ etc. Generally, the event 11’s word cloud is mainly about congratulations on ‘Chaz Mostert’ and ‘Paul Morris’ winning in the last lap in the V8 supercar race, also some people feel pity for the losing of Jamie Whincup in the last lap for lacking fuel.

The sentiment analysis then conducted, as event 11 has 9619 comments with 1592 negative samples and 4275 positive samples, therefore we decided to separate the word cloud event 11 into positive and negative ones to better analyse the audiences’ feelings for the word cloud.

* 1. **Positive sentiment analysis**



Figure 3. Positive Word Cloud for Event 11

From the word cloud for positive words in event 11, we can see that the most frequent words are quite similar to the overall word, mainly for celebrating the winning of ‘Chaz Mostert’ and ‘Paul Morris’, especially the words like ‘win’, ‘well done’, ‘best’, ‘finish’, etc.

Compared to the positive word cloud, the overall word cloud is more diverse than the positive one as there are also many people supporting Jamie Whincup and expressing their sadness for his loss.

All positive tweets for event 11 are grouped into 10 topics by topic modelling. Topic 6, topic 10 and topic 2 are the most discussed amongst all tweets during event11, the proportions of the three topics are 16.3%, 14% and 11%, separately.

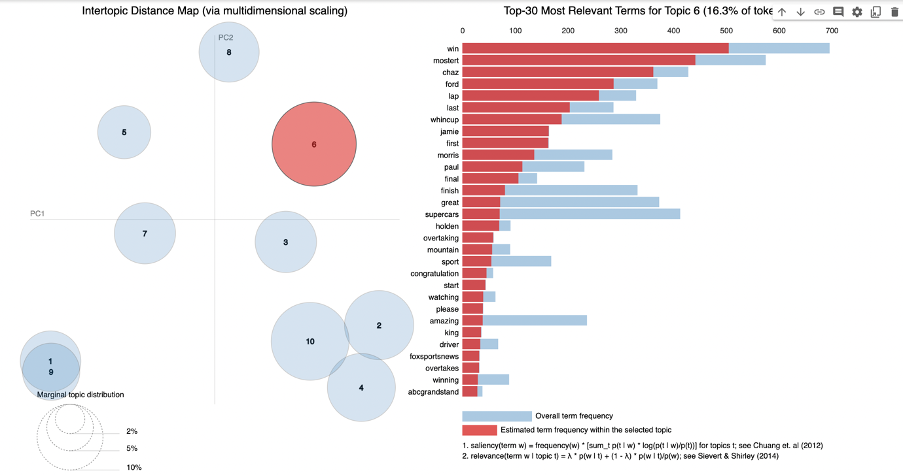


Figure 4. Event 11 Top-30 Most Relevant Terms for Positive Topic 6

For topic 6, the tweets include the topic specific words ‘win’, ‘Chaz’, ‘Whincup’, ‘lap’, it is mainly about congratulating on Ford Performance Racing Team, especially the Australian player Chaz Mostert winning on final lap on an incredible 2014 Bathurst 1000 for Ford as Jamie Whincup ran low on fuel. This topic also mentioned the words like good, unbelievable, great, awesome, love, showing the audience are very excited about Mostert’s victory.

A screenshot of a graph

Description automatically generated with low confidence

Figure 5. Event 11 Top-30 Most Relevant Terms for Positive Topic 10

For topic 10, The tweets mainly mentioned words like Chaz mozzie, “fpr” (Ford Performance Racing), Paul Morris, win, Australia etc. What deserves to be mentioned is that despite the similarities between topic 6 and topic 7 in congratulating on the winner, topic 7 is more focused on the winning team and the Ford Performance Brand. The tweets also used the pet name Chaz mossie instead of Chaz Mostert, which might shows that the fans feel closer to the players and are willing to post more tweets.

There are also words that mentioned the names of other players like Svg(Shane Van Gisbergen), Nick Percat, Oliver Gavin, showing their supportiveness for their own favourite teams and members. What’s more, companies or medias like AMP (Action Media Partners), Holden Fans, Foxsportsnews, Abcgrandstand are also mentioned in the tweets.

A screenshot of a graph

Description automatically generated with low confidence

Figure 6. Event 11 Top-30 Most Relevant Terms for Positive Topic 2

For topic 2, the tweets content is more related to showing people’s surprise and excitement about the championship. Generally, it is similar to topic 6 but without mentioning the champions’ name.

Above all, from the word cloud and positive analysis for event 11, we can conclude that the tweets are mainly about excitement for the race, discussing favourite teams, highlighting spectacular moments as well as their thanks for the organizers and agencies related to the race.

* 1. **Negative sentiment analysis**

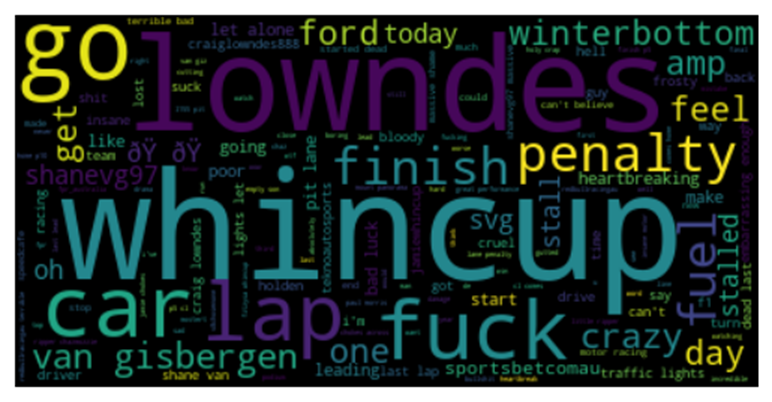


Figure 7. Negative Word Cloud for Event 11

In negative analysis, from the word cloud, same as the positive tweets the players’ names are the most frequent words that clearly to be seen. These names can refer to the incidents that were caused by these players during the race. In addition, the words " penalty, crazy, poor even swear words appeared in the negative sentiment word cloud. These words indicate negative feelings and thoughts that the audience got when specific incidents happened in the supercar race.

In terms of topic modelling for negative analysis, the three most discussed topics will be introduced in the following paragraph.

A screenshot of a graph

Description automatically generated with low confidence

Figure 8. Event 11 Top-30 Most Relevant Terms for Negative Topic 6

First, from topic 6, the words “lap” and “last” were shown above other words, therefore, we assumed that this topic is mostly about the last lap of the race. One of the dramas that happened at the last lap and made 2014 Bathurst 1000 excited was Team Ford driver Mostert and his co-driver Paul Morris, who began the race from the last position on the starting grid, managed to take the lead only on the last lap. Furthermore, since it also mentioned “svg” which was considered to be the abbreviation of Shane van Gisbergen, because of a late pitstop on the 151st lap, van Gisbergen experienced his car dying on him as he pulled in for a late fuel top-up. This accidental happening generated a flurry of "unbelievable" comments from the audience on Twitter.

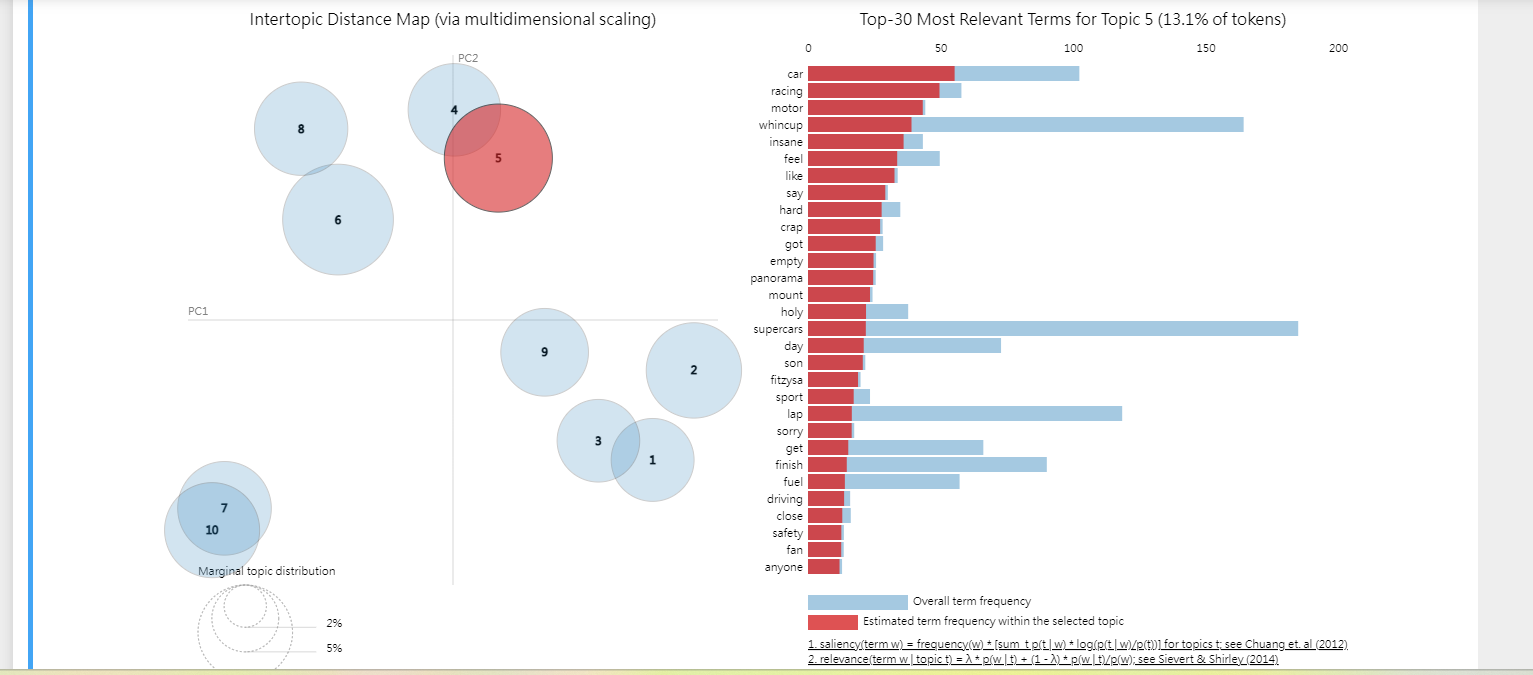


Figure 9. Event 11 Top-30 Most Relevant Terms for Negative Topic 5

For topic 5, the Red bull driver Jamie Whincup ignored instructions to save fuel in the final stages of the race, which caused the Red Bull Holden car to run out of fuel on the final lap, dropping him from the lead to fifth place. Whincup had been told to save fuel by his engineer but refused to save fuel and went for the win instead of second place. This was only the second time in 10 years that Triple Eight has failed to achieve a podium finish in the Bathurst 1000.

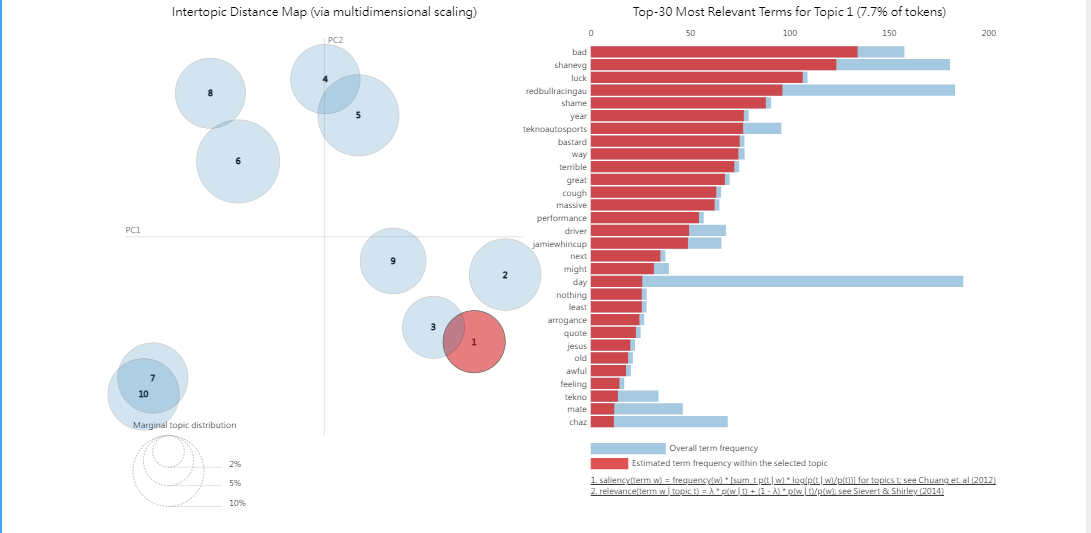


Figure 10. Event 11 Top-30 Most Relevant Terms for Negative Topic 1

In topic 1, 'bad, 'shanevg', 'luck', 'redbullracingau', 'shame', 'year', 'teknoautosports', 'bastard', 'way', 'terrible' are the ten relevant terms of topic 1. According to these words we found that the topic one is possibly discussed about the van gisbergen’s mistake of over-fuelled. The kiwi driver van Gisbergen and his team were on track to win the race but this minor error in the last stages of the race caused the car to stall and they finished at 16th. The team had been attempting to save fuel in order to make it to the end of the race, but this proved in vain. After the race, the driver was spotted alone in the Tekno Autosport transporter. Not only did this result leave the driver disappointed, but it also disappointed the supporters who were hoping for a team victory.

For both positive and negative analysis, the occurrence of unexpected incidents often prompts individuals to share their thoughts on Twitter, thereby amplifying engagement during live sport events.

**Recommendation**

In conclusion, it is suggested that Supercar hold a campaign which enables fans to vote in Twitter Polls to support the teams and comment under the polls to increase audience engagement and interactions. Unveiling the winners’ information, even their training story, pictures and videos on Supercar’s official account are also considered to be an effective approach to communicate with the target audience. Based on the ratio of positive tweets, sharing more positive user generated content is recommended to increase the engagement.

1. **Factors that influence TV ratings and their impact on fans' engagement**

To answer the question of what variables have a great impact on the TV rating, dataset from Tv rating and tweet sports were merged and selected numeric variables such as number of players, number of teams,number of tweets, number of retweets and number of likes, event and weekdays were taken to find the correlation as shown in the below heatmap to find the strength of relationship between the variables.

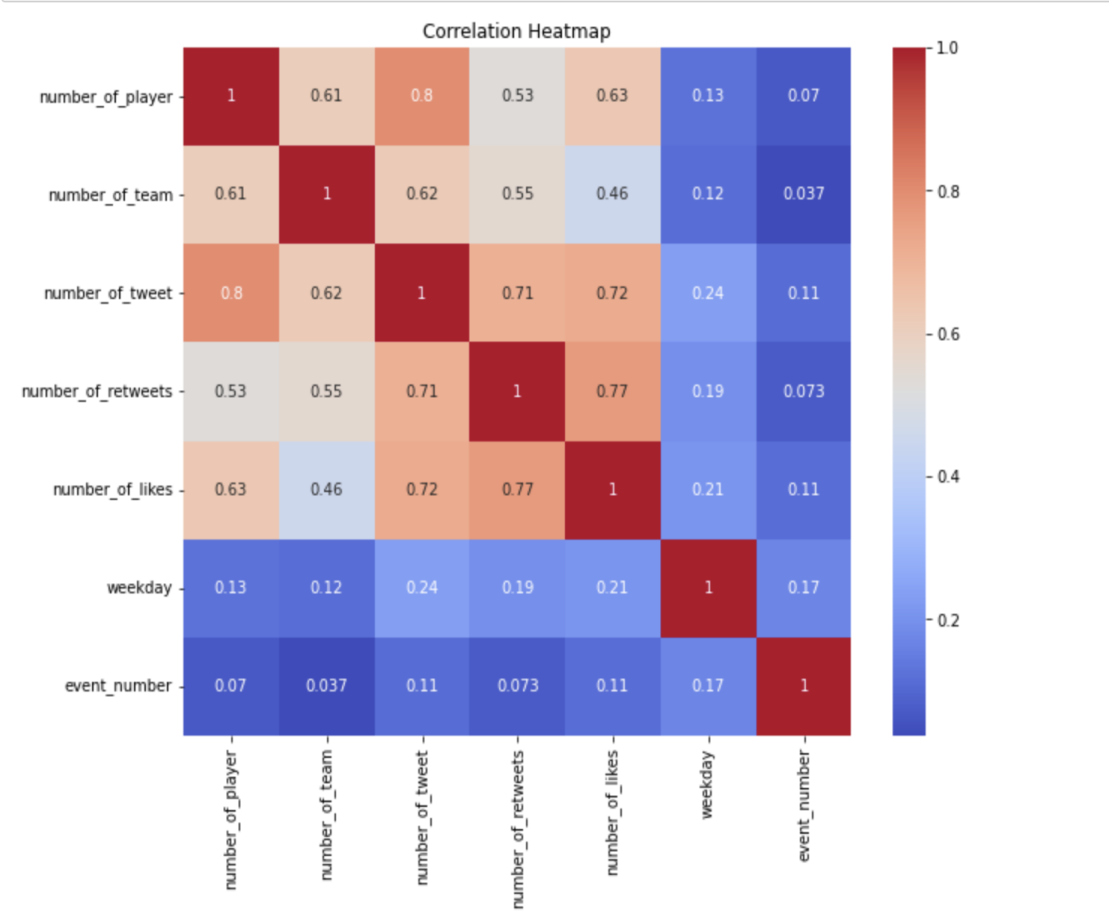


Figure 11. TV Rating Correlation Heatmap

As the number of players, number of team, number of tweets, number of retweets, number of likes have the highest correlation than weekday and event number hence, highly correlated variables were considered as independent variables and TV viewers are the dependent variables.

Afterwards, Multiple regression analysis was utilised to find the relationship between dependent and independent variables as down in the following table:

A screenshot of a computer

Description automatically generated with low confidence

Figure 12. Multiple Regression Analysis of TV Viewers

From the above table we can conclude that the number of tweets have the highest coefficient of 1.8556 with p-value>0.05 being the most statistically significant variable which is positively correlated with the TV viewers. The number of team have the highest negative coefficient value of -5.6275 with p-value of 0.068 however since the p-value is more than 0.05 hence it not statistically significant to define the relationship with TV viewers as a result the second highest coefficient which is number of player(-2.7346) have p-value<0.05 signifies it statistically significant to have negative relation with the Tv viewers.

1. **Recommendations**

Based on our above analysis, we can conclude that to increase the Tv rating, we need to increase the number of tweets. Hence, we recommend running a social campaign to increase social engagement for twitter users. The social campaign will be called Watch and Win which will be divided into 3 stages. They are:

| **Stages** | **Description** |
| --- | --- |
| 1 | In the first half of the game, a promotion code will be shared with a random question about the game which has to be answered. |
| 2 | Then, the twitter users have to send the code along with the answer with the #suwatwin and @reply in the sponsor message. |
| 3 | After the campaign is over, the winners will be randomly chosen who have sent the message in the promotion code and hashtag. 10 winners will get rewarded with to meet the final match winners and the second and third winner will get gifts from the two main sponsors. |

Table 1. ‘Watch and Win’ Social Campaign

With this campaign, twitter users can interact more by participating in the game by tweeting while watching the game with the hope of winning exciting prizes. Suh and Liu’s (2016) report emphasises that gamification enhances user engagement as it brings enjoyment about what they are doing and game dynamics like rewards, competitions, altruism and self-expression to satisfy their psychological needs. Further, Cheng et al.(2016) emphasised in their report that the higher the user engagement of social media, the higher the TV rating. On the other hand, Supercars have introduced competition for the users like Thrifty Thrills, KingGee, win a supercar but not a social media campaign like Watch and Win which will be a new and exciting campaign for the users. Hence, this will be beneficial to increase the number of tweets furthermore improving the TV rating.

**Overall recommendation:**

This report concludes that conducting Twitter Polls to support the teams and allowing comment under the polls and sharing winner’s information such as pictures, videos etc on Supercar’s official account helps to surge audience engagement and interactions. Further, running the Watch and Win social campaign, to increase the tweet number through social engagement results in increase of the TV rating of Supercars. Based on the recommendations, the manager should target their audiences and allocate budget, time and manpower to carry out the campaign and advertise the campaign well to make it successful.

There are three significant limitations related to this report. Firstly, the report is constrained by the availability of a restricted amount of data from Twitter, as the dataset may not be fully representative. Secondly, due to technical constraints or lack of access to relevant data, the research cannot accurately gauge user engagement on the livestream platform, potentially limiting the comprehensive understanding of user interactions and their impact. Lastly, regression analysis was drawn from using only one model without comparing with other models which couldn’t guarantee whether the utilised model is the best model or not. Hence, for future research, data from other social media such as Facebook and Instagram should be considered to cover a larger sample size for better and reliable results (CWauthors, 2022) and also to authenticate the prediction model, we may further use more models to verify the accuracy of the model and use benchmarks to evaluate.

**Reference**

Cheng, M.-H., Wu, Y.-C., & Chen, M.-C. (2016). Television Meets Facebook: The Correlation between TV Ratings and Social Media. *American Journal of Industrial and Business Management*, *06*(03), 282–290. <https://doi.org/10.4236/ajibm.2016.63026>

CWauthors (2022). *The importance of large sample sizes in research | CW Authors*. [online] Available at: <https://www.cwauthors.com/article/importance-of-having-large-sample-sizes-for-research#:~:text=Larger%20studies%20provide%20stronger%20and>.

Murphy, G. (2018). Twitter changes the live TV sports viewing experience Twitter Marketing <https://marketing.twitter.com/en/insights/twitter-changes-the-live-tv-sports-viewing-experience>

Smith, L. B., Pegoraro, A., & Cruikshank, S. A. (2019). Tweet, Retweet, Favorite: The Impact of Twitter Use on Enjoyment and Sports Viewing. *Journal of Broadcasting & Electronic Media*, *63*(1), 94–110. <https://doi.org/10.1080/08838151.2019.1568805>

Suh, A., Wagner, C., & Liu, L. (2016). Enhancing User Engagement through Gamification. *Journal of Computer Information Systems*, *58*(3), 204–213. <https://doi.org/10.1080/08874417.2016.1229143>

‌

‌