SUPERCARS FANS ON TWITTER

What are They Talking About?

Group 6

Yolanda Kurniarto 480447849 Suya Chen 490247372 Lei Zhang 510221814

Introduction

This project experimented with a dataset related to the Supercar events, which included live tweets and live television viewership ratings. There are two main goals of this Machine Learning (ML) project. First, to investigate the different topics remarks, and sentiments that happen during Supercar events by exploring electronic word-of-mouth (e-WOM) live tweets data. e-WOM is defined as "the exchange of consumption experiences among consumers" (Borgida & Nisbett, 1977, cited in Choi, 2020). Secondly, this project aims to examine the impact of the Second Screen Phenomenon by exploring the impact of live tweets on TV viewership data. Three main ML techniques are used to inspect these two goals: topic modelling, sentiment analysis, and regression analysis.

Problem Definition

The context of this project is the Supercar Championship. There is now a growing trend among Fans of The Supercar Championship to embrace the second screen phenomenon. The second screen phenomenon refers to the simultaneous consumption of two different devices that are used to express opinions and share relevant information (de Meulenaere, Bleumers, & van de Broeck, 2015) online while watching the Supercar Championship broadcasts. This project investigates Supercars fans' online behaviour through Twitter data and its impact on TV viewership. The questions that are explored for this project are listed below:

- Q1: What is the most popular topic for each event?
- Q2: Is there any correlation between sentiment score and popularity of each game?
- Q3: How do eWOM communications impact TV rating and online engagements
    - Q3(1): How do eWOM communications impact TV viewers in different locations?
    - Q3(2): How do eWOM communications impact online engagements on different race weekends?

Other potential questions can be answered by the dataset, such as how the variation in sentiment influences TV viewership, whether there are any events when Supercar's fans are most engaged, and others. However, due to time limitations, it was decided that it was best to explore the four main questions listed above.

Q1: What is the most popular topic for each event?

It is practical to start by investigating the overall Topic Modeling result. Some preprocessing is required. In addition to the English vocabulary stop words, some words in raw materials may not help determine the topics. For instance, the website keywords like "HTTP" and the Twitter keywords like "t" and "rt" do not make any sense in any sentences. Besides, some common words do not refer to a specific topic but appear quite frequently, like "v8sc" and "v8supercars". After that, further analysis requires lowering the characters, tokenizing the texts, and lemmatizing the tokens.

It is then proper to apply the Latent Dirichlet Allocation (Figure 1) on all the tokens and select the top topics. Locations appear quite frequently, also the competitions and racers' names. This model includes all texts, i.e., tweet texts from all 14 events.

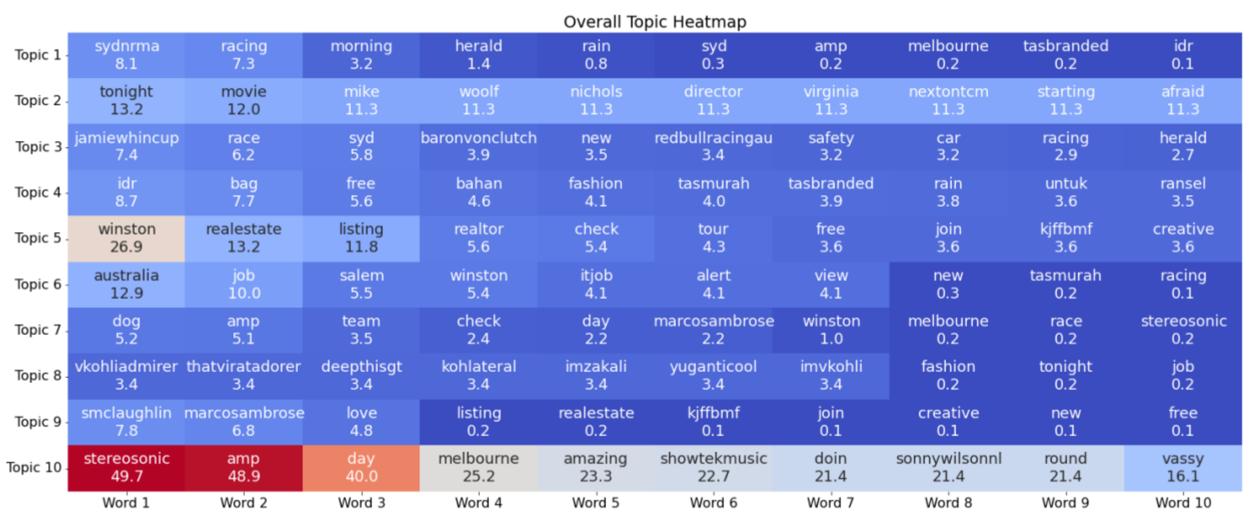

Figure 1

The topic mapping could help identify the inter-topic distance based on the LDA model. As is shown in the figure, there are no significant intersections between any pairs of topics given two principal components as scalers. The mapping also shows the word frequency in a subset compared to the entire set. For example, topic nine (Figure 2) included 10.3% of tokens overall. The word

"Morris," i.e., Paul Morris the racer, appears most frequently. In addition, almost all the appearance of the word "Morris" goes to this topic.

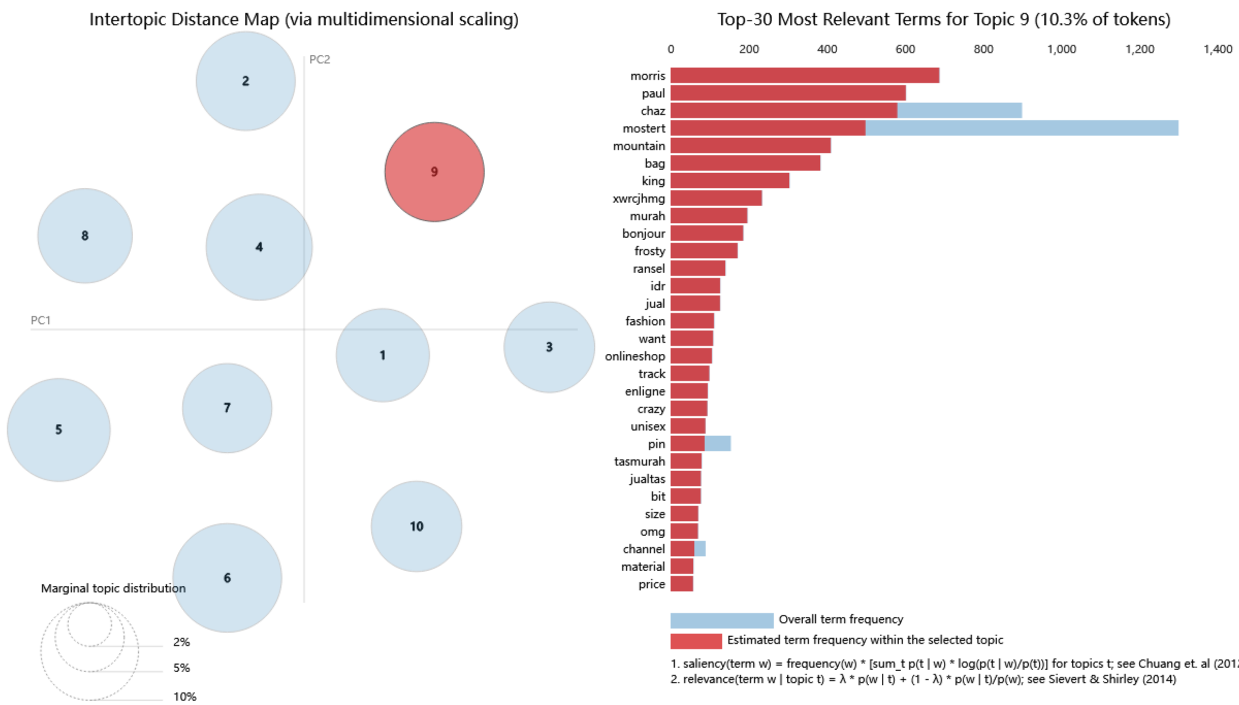

Figure 2

After that, a word cloud (Figure 3) helps visualize the word frequencies. There are several kinds of words that look similar. It is then natural to categorize these frequent words into groups. For example, there are names of racers, like "Paul Morris"; words that are closely related to some specific events, like "mountains"; and some words and phrases that are closely related to racing, like "race," "final lap," and "safety car."

Figure 3

Now that the question is about the topics for each event, further research requires dividing the raw materials into 14 subsets according to their corresponding events and repeating the previous process on each set. The results include the top ten topics for each event and their mapping.

After that, the results with the top ten topics in each event, see Appendix Figure 1, indicate some preliminary findings on these topics. For instance, there are some common topics shared between events. Each of these events has at least one topic that a racer's name appears most frequently. Some unique topics appear only in specific events. Most importantly, there is a topic similarity among these events. For example, when a racer's name appears in a topic, it is highly likely to be one of the most frequent words on this topic.

Recommendation

Based on the findings above, we have come up with two recommendations. First, the promotion campaigns in the future shall focus on racers since racers are the most popular amongst all components in each topic. Also, it seems that customized promotions can be helpful for each event based on unique topics since the tweets of these events do not always focus on the same point.

Q2: Is there any correlation between sentiment score and popularity of each game?

In this section, we measure each game's popularity by the number of tweets and viewers during each game. Additionally, we also investigated supporters' sentiments. Since viewers may support different teams and racers, it was hypothesized that they might have different sentiments about the same result.

Simply averaging the sentiment scores of different viewers would give a result close to zero, ignoring the intensity of different viewer sentiments. Therefore, the absolute value is considered a more accurate evaluation of the overall sentiment intensity of the audience.

First, we will use the sentiment analyzer to analyze the sentiment score of each tweet. The compound score ranges from -1 to 1, where -1 is entirely negative, 0 is neutral, and 1 is entirely positive. The absolute value, on the other hand, regardless of positive or negative, only measures the intensity of the sentiment.

Considering that the popularity of each game varies greatly, for example, event 11 was very popular. Over 13,000 tweets had over 9,000 tweets during event 11, and the 70,000+ viewers also accounted for a significant proportion of the 200,000+ viewers. So, we implemented a pivot table like function in python to group all the tweets by event and calculate the total number of tweets during each game, the average of the two Sentiment Scores, and the total number of viewers for each game. Finally, the two data frames are merged to facilitate the extraction of variables later.

Finally, the correlation function is used to find the correlation index (Figure 4) of the different variables.

	number_of_tweet	number_of_tv_viewer	senti_ab_mean	senti_mean
number_of_tweet	1.000000	0.968909	0.721953	0.280913
number_of_tv_viewer	0.968909	1.000000	0.786485	0.404227
senti_ab_mean	0.721953	0.786485	1.000000	0.584623
senti_mean	0.280913	0.404227	0.584623	1.000000

Figure 4

As the table shows, the number of tv viewers showed a high positive correlation with the number of tweets. Sentiment score also shows a positive correlation with these two variables.

The absolute value of the sentiment score, on the other hand, has a higher correlation than the sentiment score itself, somehow indicating that putting aside the positive and negative aspects of emotions, people show stronger emotions in the popular games.

Hence, we can reach a conclusion that, during the more popular event, people get into higher engagement and contribute to higher sentiment scores (evaluated by absolute value).

Recommendation

The primary revenue of a sporting event comes from broadcasting rights, sponsorship and tickets, and Supercar is known to generate revenue from sponsorship, tickets, and media rights. Therefore, we focus on increasing its TV rating by raising the Sentiment Score or, in other words, e-WOM to raise the price of media rights and sponsorship to increase revenue (Cadario, 2015).

Our recommendations are provided below:

• Evaluation in advance could be done to predict which event could be a big hit. They can seize the opportunity to enhance their brand image and commercial value and use their marketing resources more effectively.

• Before the event, which is predicted to be a big hit, prepare more event-relevant official content for social media in advance to improve users' engagement to make the event series a hot topic on social media to improve tv ratings for following events.

• Run a profitable marketing campaign on social media during the popular event. For example, social media challenges that might go viral and reach a wider audience. This can be used as an opportunity to attract high-quality sponsors and raise advertising income.

Q3: How do eWOM communications impact TV viewers and online engagements?

Recent developments have shifted audiences into active media consumption participants (Segijn, et al., 2020). As a result, audiences integrate an additional second screen to engage in social conversations and access real-time information to enhance their experience. Twitter is one of the platforms that is often used for this purpose (Voorveld, van Noort, Muntinga and Bronner, 2018, as cited in Segijn, et al., 2020).

In this section, we aim to delve into the effect of these live tweets on TV ratings in various cities in Australia: Sydney, Brisbane, Melbourne, Adelaide, Tasmania, and Perth. This variable was transformed into dummy variables with Adelaide as the base case. Furthermore, this project also investigates how eWOM communication influences online engagement (i.e., total likes, comments, and retweets) on other race weekends. This variable was also transformed into dummy variables where weekday 5 is set as the base case.

The dataset contains information from a merged dataset that combines live tweets data and TV rating data during each Supercar event throughout the year.

Q3(1): How does eWOM communications impact TV viewers in different locations?

In this section, the goal is to identify whether different locations (i.e., Australian cities) influence the impact of eWOM communications for TV viewers. The table (Figure 5) below summarizes the independent variables in relation to each city. Each city is represented somewhat equally, except for Tasmania.

city	city	tvviewers	number_of_player	number_of_team	total_retweets	total_likes	total_comments
Adelaide	469	16263	385.0	56.0	1159.0	2043.0	409.0
Brisbane	467	47413	388.0	56.0	1157.0	2027.0	406.0
Melbourne	471	46599	386.0	56.0	1159.0	2049.0	409.0
Perth	476	14083	387.0	56.0	1177.0	2079.0	411.0
Sydney	473	53420	387.0	56.0	1166.0	2050.0	412.0
Tasmania	375	7470	353.0	52.0	920.0	1847.0	381.0

Figure 5

Furthermore, the cities that are included in the model (i.e., Adelaide, Brisbane, Melbourne, Perth, Sydney, and Tasmania) have different population sizes. As such, to control for this, we introduce a new variable that is tv_viewers/capita, which comes from the division of total tv viewers divided by the total population of each city in 2014 from the Australian Bureau of Statistics (Australian Bureau of Statistics [ABS], 2016) to match the Twitter data period.

In exploring the effects of e-WOM communications in different cities, first, a Pearson correlation matrix (Figure 6) was generated to see the relationship between the variables in the merged dataset and the target variable: log-transformed TV viewers/capita.

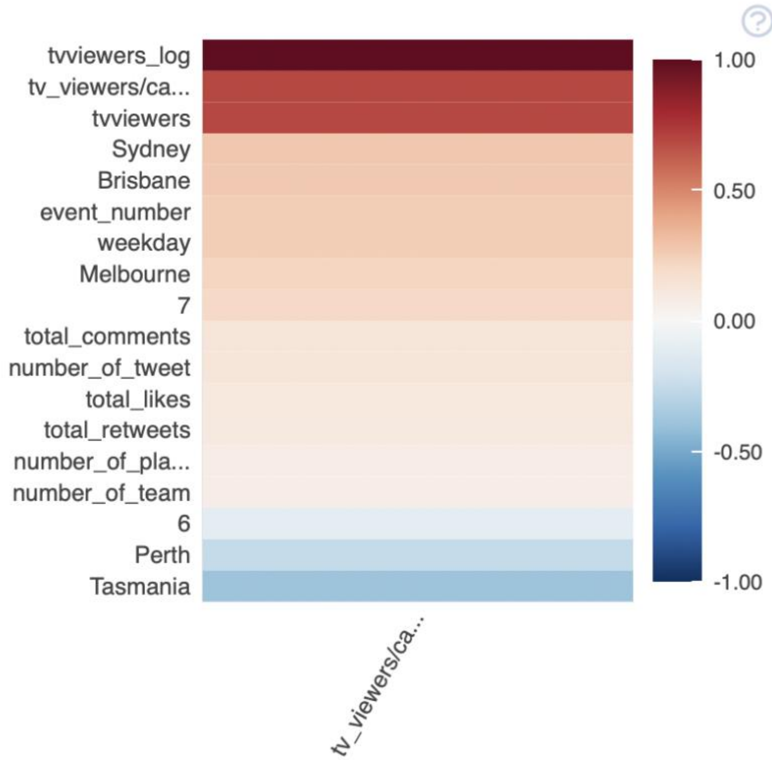

Figure 6

The matrix indicates that e-WOM in Sydney, Brisbane, and Melbourne are positively correlated with viewership while those in Perth and Tasmania are negatively correlated. The OLS regression model is written as:

$$Log(TV\ Viewers/Capita) = -13.02 + 0.01\ Number\ of\ Tweet - 0.05\ Number\ of\ Players + 1.12 Brisbane + 0.98\ Melbourne - 0.22\ Perth + 1.14 Sydney - 0.78\ Tasmania + 0.06\ Number\ of\ Comments$$

Based on the result, we can interpret that eWOM communications in Brisbane, Melbourne, and Sydney are more likely to generate TV viewership than Adelaide. On the other hand, eWOM communications in Perth and Tasmania are more likely to reduce TV viewership compared to Adelaide.

Shifting the focus to e-WOM communications variables, it seems that tweets that contain mention of players negatively impact tv viewership. But, engaging tweets, those with comments, have 6% more odds of increasing tv viewership compared to those with no words. Other measures of engagement for e-WOM which are *total_likes* and *total_retweets* are not significant for the regression model with p-values 0.243 and 0.072 respectively.

The OLS regression model output (Figure 7) is provided below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     tv_viewers/capita_log   R-squared:                       0.401
Model:                               OLS   Adj. R-squared:                  0.399
Method:                    Least Squares   F-statistic:                     165.7
Date:                   Thu, 26 May 2022   Prob (F-statistic):          5.40e-293
Time:                           12:06:15   Log-Likelihood:                -3608.0
No. Observations:                   2731   AIC:                             7240.
Df Residuals:                       2719   BIC:                             7311.
Df Model:                             11
Covariance Type:               nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
const               -13.0258      0.046   -280.913      0.000     -13.117     -12.935
number_of_tweet       0.0144      0.003      4.900      0.000       0.009       0.020
number_of_player     -0.0603      0.014     -4.196      0.000      -0.088      -0.032
number_of_team        0.0127      0.037      0.347      0.729      -0.059       0.085
Brisbane              1.1172      0.059     18.804      0.000       1.001       1.234
Melbourne             0.9799      0.059     16.528      0.000       0.864       1.096
Perth                -0.2242      0.059     -3.791      0.000      -0.340      -0.108
Sydney                1.1313      0.059     19.102      0.000       1.015       1.247
Tasmania             -0.7659      0.063    -12.159      0.000      -0.889      -0.642
total_likes           0.0031      0.003      1.167      0.243      -0.002       0.008
total_retweets       -0.0098      0.005     -1.798      0.072      -0.020       0.001
total_comments        0.0554      0.015      3.822      0.000       0.027       0.084
==============================================================================
Omnibus:                     3129.348   Durbin-Watson:                   0.668
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           587838.507
Skew:                          -5.615   Prob(JB):                         0.00
Kurtosis:                      73.992   Cond. No.                         133.
==============================================================================
```

Figure 7

Recommendation

In light of this, we came up with some recommendations:

- Maintain engagement with Supercars fans by replying/commenting on their tweets. E-WOM communications are often used to express opinions and share relevant information (de Meulenaere, Bleumers, & van de Broeck, 2015). By replying to fans' tweets from the

Supercars official account, Supercar can leverage the persuasive and informative effects (Bae & Hye-Jin, 2020) of e-WOM communications to increase TV viewership. For example, by replying to the fans' tweets, Supercar can persuade fans to watch the next event's live broadcast when they posted a tweet that shows disappointment that their team was losing in the previous event.

- Seeing that Perth and Tasmania have a negative correlation to TV viewership, we recommend that Supercar place more attention on broadcasts in those areas. They can also increase fans' participation by promoting the supercar e-series, the virtual racing competition to offer different types of content for the fans (Colter, 2020).

Q3(2): How does eWOM communications impact online engagements on different race weekends?

In this section, total comments were chosen as the online engagement variable due to two reasons. Firstly, upon regressing the independent variables to tv viewership, the number of comments was a significant variable. As we wanted to maintain the connection between tv viewership and e-WOM communications, total comments were chosen as the target variable. Secondly, from our experiment, we found that using total comments as the dependent variable, the regression model had a higher R-Squared score (Figure 8).

Variable	R–Squared Score
total_likes	0.533
total_retweets	0.535
total_comments	0.657

Figure 8

In exploring the effects of eWOM communications to online engagements on different race weekends, Pearson correlation matrix (Figure 9) was generated to see the relationship between the variables in the merged dataset and the target variable: *total_comments*.

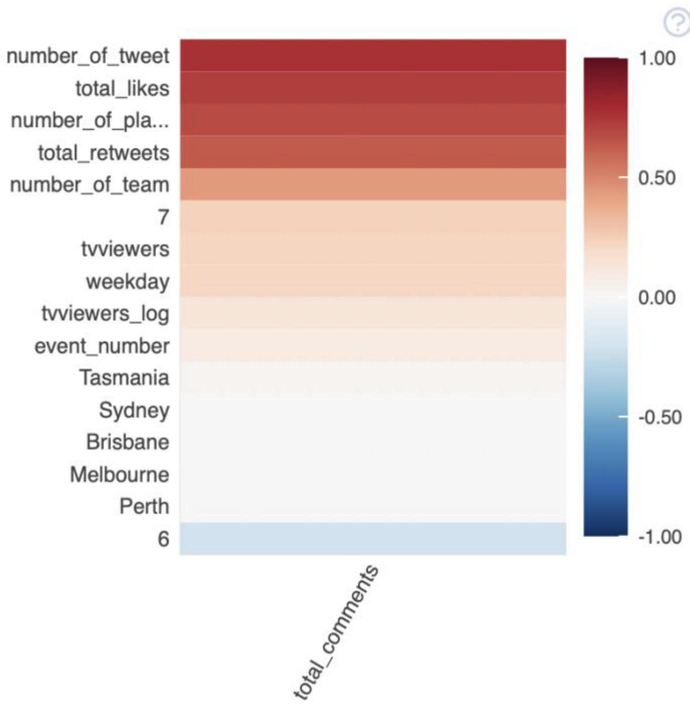

Figure 9

The matrix indicates that tweets that contain mention of players are positively correlated with total comments, but tweets that contain mention of the team have a negative correlation on total comments. This might be because fans are more attached to players rather than teams, and as such, tweets that contain teams do not interest them as much.

The OLS regression model is written as:

$$Total\ Comments = -0.48 + 0.07\ Number\ of\ Tweets + 0.16\ Number\ of\ Players - 0.32\ Number\ of\ Teams + 0.21\ weekday\ 6 + 0.28\ weekday\ 7 + 0.02\ Total\ Retweets + 0.05\ Total\ Likes$$

Based on this, it seems that tweets on weekday 6 and 7, on average, are associated with an increase in total comments compared to weekday 5. Holding all the other variables constant, tweets in weekday 7 are associated with a 28% increase, on average, for total comments. Furthermore, tweets in weekday 6 are associated with a 21% increase, on average, for total comments.

The OLS regression output (Figure 10) is provided below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         total_comments   R-squared:                       0.657
Model:                            OLS   Adj. R-squared:                  0.656
Method:                 Least Squares   F-statistic:                     745.2
Date:                Thu, 26 May 2022   Prob (F-statistic):               0.00
Time:                        11:53:36   Log-Likelihood:                -4367.2
No. Observations:                2731   AIC:                             8750.
Df Residuals:                    2723   BIC:                             8798.
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                       coef    std err          t      P>|t|      [0.025      0
------------------------------------------------------------------------------
const               -0.4828      0.091     -5.300      0.000      -0.661      -(
number_of_tweet      0.0689      0.004     18.673      0.000       0.062      (
number_of_player     0.1560      0.019      8.269      0.000       0.119      (
number_of_team      -0.3177      0.048     -6.601      0.000      -0.412      -(
6                    0.2081      0.095      2.196      0.028       0.022      (
7                    0.2813      0.096      2.942      0.003       0.094      (
total_retweets       0.0160      0.007      2.223      0.026       0.002      (
total_likes          0.0528      0.003     15.627      0.000       0.046      (
==============================================================================
Omnibus:                     1088.164   Durbin-Watson:                   2.154
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11267.234
Skew:                           1.592   Prob(JB):                         0.00
Kurtosis:                      12.428   Cond. No.                         142.
==============================================================================
```

Figure 10

Recommendation

Based on this result, we recommend that Supercars focus on two things on their Twitter account.

- Post live updates on the official account that mention players to promote online engagements.
- Facilitate online discussion on weekdays 6 and 7 to enhance the fans' sense of belonging with the event (Kasavana et al., 2020, Torres, 2017, as cited in Kharouf, et al., 2020).

These two strategies are advised so that Supercar official account can generate higher online engagement by creating posts with content that might interest Supercars' fans more, that is, tweets that mention players such as Paul Morris or Jamie Whincup. By leveraging their support for their favourite racers, fans are more likely to participate in online discussions. As such, Supercars" can maintain consumer engagement with the championship and build "long-term, sustainable customer relationships" (Noh, et al., 2020).

References:

Australian Bureau of Statistics. (2016). *3218.0 - Regional Population Growth, Australia, 2014-15.* https://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/3218.0Main%20Features152014-15?opendocument&tabname=Summary&prodno=3218.0&issue=2014-15&num=&view=

Bae, G., & Hye-Jin, K. (2020). Relation between early e-WOM and average TV ratings. *Asia Pacific Journal of Marketing and Logistics*, *32*(1), 135-148. https://doi.org/10.1108/APJML-10-2018-0402

Cadario, R. (2015). The impact of online word-of-mouth on television show viewership: An inverted U-shaped temporal dynamic. Marketing Letters. 26, pages 411–422 (2015).

Choi, J. (2020). How consumers engage in & utilize the source of electronic Word-of-Mouth (e-WOM)?. *Academy of Strategic Management Journal, 19*(2).

Colter, C. (2020, April 16). Supercars is showing how to win with fans and advertisers during COVID-19. *Mumbrella.* *https://mumbrella.com.au/supercars-is-showing-how-to-win-with-fans-and-advertisers-during-covid-19-624654*

De Meulenaere, J., Bleumers, L., & van de Broeck, W. (2015). Audience perspective on second screen. *The Journal of Media Innovations, 2(*2),6-22, http://dx.doi.org/10.5617/jmi.v2i2.909

Kharouf, H., Biscaia, R., Garcia-Perez, A., & Hickman, E. (2020). Understanding online event experience: The importance of communication, engagement and interaction. *Journal of Business Research, 121*, 735-746. https://doi.org/10.1016/j.jbusres.2019.12.037

Noh, Y., Kim, H., Kim, K., & Cheong, Y. (2020). The relationship between consumer engagements with Super Bowl ad tweets and the advertisers' official Twitter accounts: a panel data analysis of 2019 Super Bowl advertisers. *International Journal of Advertising, 40*(6), 852-869. https://doi.org/10.1080/02650487.2020.1819006

Segijn, C. M., Maslowska, E., Araujo, T., & Viswanathan, V. (2020). Engaging with TV events on Twitter: The interrelations between TV consumption, engagement actors, and engagement content. *Internet Research, 30*(2), 381-401. https://doi.org/10.1108/INTR-08-2018-0389

Appendix

Figure 1: Topic Heatmaps for Each Event