Introduction

The supercars championship is widely regarded as the best touring car category worldwide and a leader in motorsport entertainment since 1960 (Supercars Championship, n.d.). These high-speed events feature cutting-edge sports cars, pushing the limits of performance to the extreme. Nowadays, social media platforms like Twitter have reshaped the ways fans engage during live broadcasts. This report aims to explore the impact of live tweets from fans during the games, utilising topic modelling, sentiment analysis and regression analysis to examine their influence on online user engagement and television ratings.

Two datasets were utilised for analysis. The first dataset provides comprehensive details on tweets, and the second dataset illustrates relevant information on television viewership counts. Both datasets were sourced from the year 2014 and underwent data cleaning procedures to address missing values prior to the analysis.

The analysis classifies the 14 games in 2014 into four distinct segments based on their averaged TV viewership (Appendix 1). This categorisation helps to understand the varying levels of fan engagement and popularity associated with each segment. By examining these segments, tailored managerial recommendations can be provided to effectively target specific fan groups, ultimately enhancing their overall gaming experiences.

By assessing the relationship between live tweets and television ratings, this report aims to provide insights for event organisers, broadcasters, and marketers. These findings can assist in optimising strategies to enhance fan experiences as well as TV viewership of supercars racing in Australia.

Main Analysis

Segment 1

Ten topics were generated through the topic modelling analysis to find the most frequently tweeted contents during game 11. These topics, combined with the background research of the game, clarify 3 major incidents that occurred during the game. Each of these incidents is associated with a specific topic that emerged from the analysis.

Topic 1 revealed the biggest twist of the game, with three names repeatedly mentioned: Jamie Whincup, Chaz Mostert and Paul Morris (Appendix 2). According to an article from Autosport (2014), Mostert and Morris, representing the "Ford" team, surpassed Whincup on the final lap of the Bathurst 1000 race as Whincup encountered fuel problems. This sensation propelled the topic to become the most popular and widely discussed subject of the game.

Topic 4 disclosed the rivalry between Mark Frosty Winterbottom and Craig Lowndes (Appendix 3). Winterbottom was collided by Lowndes from behind with only 8 laps remaining (Dale, 2015). This resulted in Winterbottom ending up in the 6th place, while Lowndes received a penalty for his action. The unexpected accident left fans in disbelief, expressing their sympathy for Winterbottom on Twitter. Consequently, this incident emerged as another popular topic within the game.

Topic 9 involved Shane Van Gisbergen (Appendix 4). As reported by V8 Supercars (2014), the engine of his car failed to start from the pits with 10 laps to go, losing him valuable time and ended up 16th. This turn of events led to an overwhelming amount of sympathy from fans expressing their disappointment of him not securing a victory.

This race contained a series of unpredictable moments among competing drivers. It is such unpredictability in outcome that contributed to the game receiving the highest TV viewership among all events. The dynamic of the competition made it challenging to predict the ultimate champion, further amplifying the interest and engagement of viewers.

Subsequently, sentiment analysis was conducted to gain insight into the sentiments surrounding the event. The analysis yielded a negative score of 0, a neutral score of 0.644, a positive score of 0.356, and a compound score of 0.8881. While the neutral score outweighed the positive score, the compound score suggested an overall positive sentiment among Twitter users regarding event 11.

The integration of insights derived from topic modelling and sentiment analysis suggested that players generate significant engagement on Twitter. Furthermore, there is a prevailing positive sentiment expressed by users towards these players. Based on this comprehensive understanding, 2 managerial implications are proposed to target segment 1 and maximise TV viewership.

First, invite popular racers to do product endorsement. According to La Ferle and Choi (2005) , the use of a celebrity endorser creates and maintains consumer attention to advertisements. In this case, The Supercars Championship could effectively capitalise on this concept by collaborating with its current sponsors, namely Panasonic, Schick, Repco Authorised Service, and Boost Mobile (Brittle, 2023). For example, creating television-exclusive commercials that feature prominent racers to enhance overall viewership. Second, host post-game analysis featuring popular racers on TV. The inclusion of popular racers in the post-game analysis attracts a larger audience and fosters viewer engagement beyond the primary event. It also adds value and interest among fans who are keen to gain insights from their favourite racers. Overall, this approach could potentially enhance viewership and increase audience retention.

Segment 2

The analytical exploration of Segment 2 primarily centred on the critical determinants contributing to television viewership during events 12-14, utilising the robust method of four distinct linear regression models. In the first two models, the dependent variables were classified as "TV views," with the independent variables comprising sentiment scores, likes, retweets, and comments. All these independent variables were found to hold statistical significance, as evidenced by their respective P-values falling below the 0.05 threshold (Appendix 5). In a subsequent investigation, sentiment scores were adjusted to serve as the dependent variables. In this configuration, only the quantity of comments and tweets were found to hold significant statistical weight (Appendix 6). Based on these findings, it's suggested to deploy a blend of automated response technology along with direct human interaction to address user comments. This approach is likely to stimulate an increase in the volume of comments and tweets, thereby positively affecting sentiment scores and, by extension, television viewership.

Nevertheless, it is crucial to avoid over-communication to prevent the onset of marketing fatigue. Thus, the timing of engagements warrants careful consideration. It should be noted that there was no significant divergence observed in the trend of sentiment scores during the day and night timeframes. Here, 'daytime' was operationally defined as commencing at 6:00:00, while 'nighttime' was deemed to initiate at 17:59:59 (*Figure 1*).

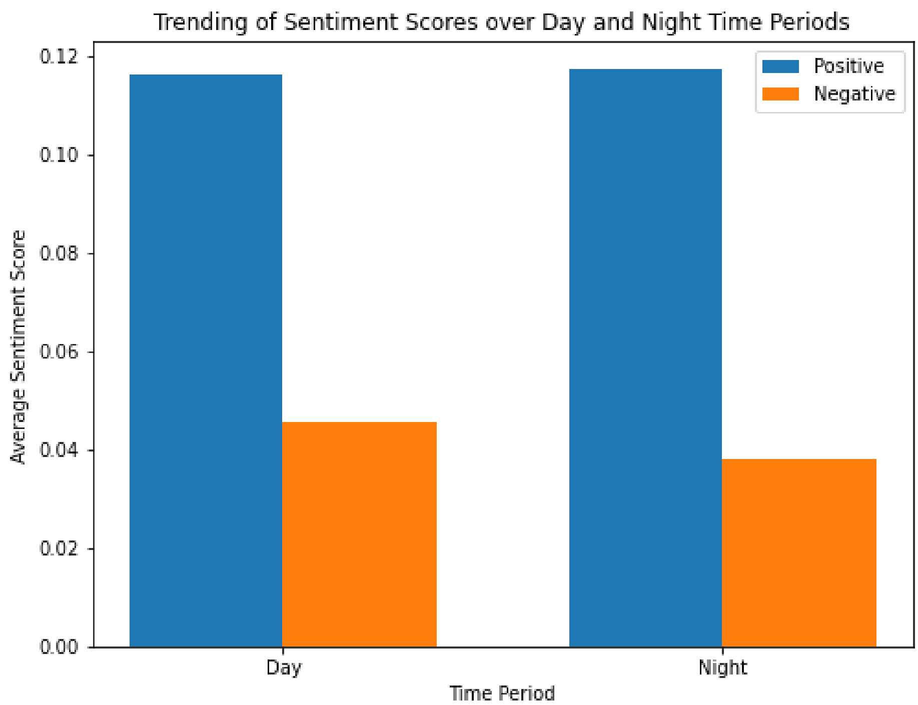


Figure 1. Trending of Sentiment Scores over Day and Night Time Periods.

Consequently, the evaluation of average sentiment scores within each temporal segment becomes an essential task. The pattern of sentiment score fluctuations across ascending chronological intervals does not display a transparent or predictable trend for the first 30 values (*Figure 2)* and the last 30 values (*Figure3*). Despite this, it allows for the establishment of a benchmark for automated responses to user commentary. For instance, we can program chatbots to initiate interactions with Twitter users when the average positive sentiment score dips below the threshold of 0.1. This strategic engagement could potentially amplify positive sentiment and enhance overall user interaction.

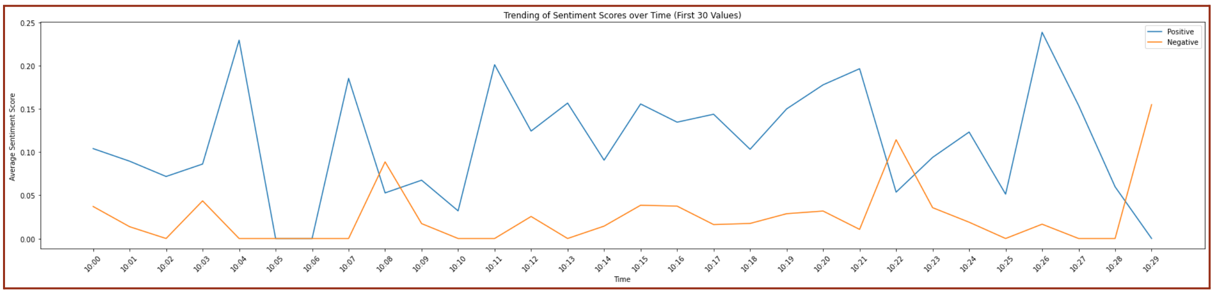


Figure 2. Trending of Sentiment Scores over Time (First 30 Values)

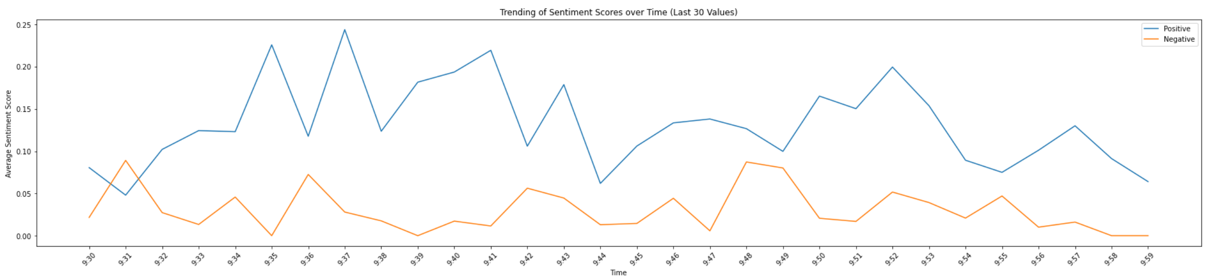


Figure 3. Trending of Sentiment Scores over Time (Last 30 Values)

Moreover, the thematic substance of conversations can carry profound implications. As illustrated in the subsequent donut plot featuring the twenty most distinct words in positive tweets (*Figure 4*), such insights can be harnessed for more sophisticated and resonant interactions . Leveraging this information, chatbots can intentionally incorporate these trending terms into their responses. This strategic approach will make the engagement more contextually relevant and potentially more compelling for users, thereby fostering enriched discussions.

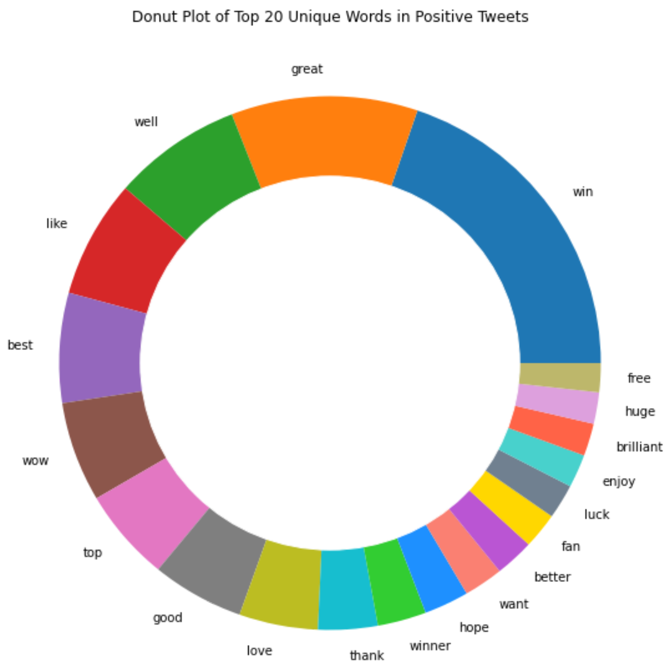


Figure 4. Donut Plot of Top 20 Unique Words in Positive Tweets.

Segment 3

Segment 3 is defined by ordinary ratings, representative of supercar competitions under normal conditions, namely events 1, 2, 7, and 10 in the dataset. As shown in *Figure 5*, segment 3 shows a smaller number of tweets, players, and teams compared to other segments. Therefore, it is necessary to find out the underlying factor contributing to the low TV ratings by connecting these features.

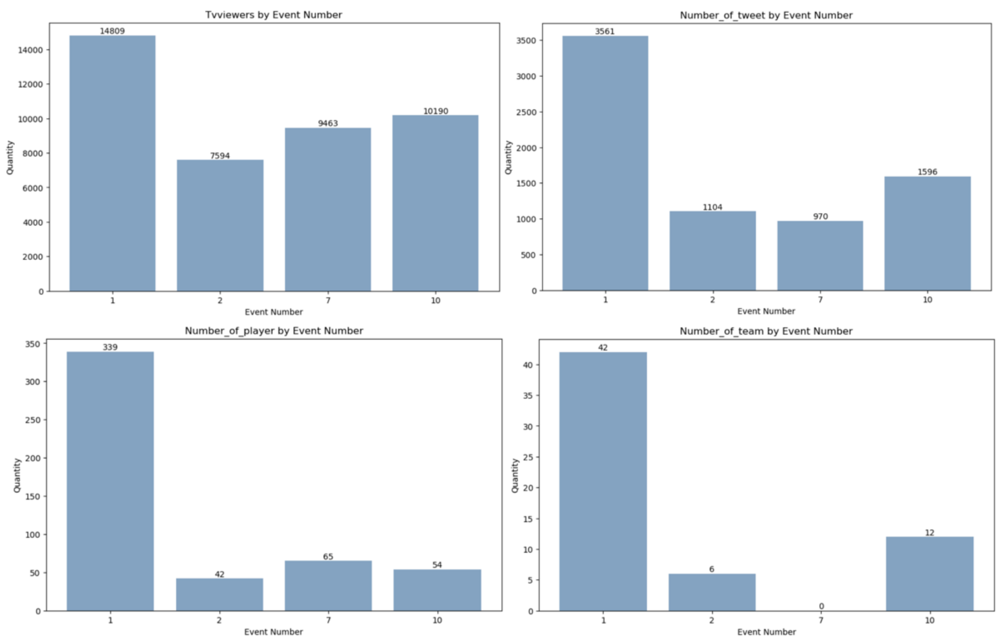


Figure 5. Overview of segment 3

Topic modelling could be used to identify the topics of a set of customer reviews by detecting patterns (Federico, 2019). *Figure 6* reveals that most popular topics in segment 3 revolve around celebrities in the game, and its importance is even greater than that of players and teams. For instance, Winston, the most frequently mentioned, refers to Winston Good, a renowned author and sports car photographer with over two million social media followers (Christina et al., 2016). His presence at an event could have made it a talking point. Additionally, terms like 'muerte', 'hijos', and 'bonjour' refer to a French team who became popular in 2015. In essence, not just the race content and players but also invited celebrities and popular teams can influence TV ratings, potentially having an even more significant marketing promotion force. However, to quantify the extent of this impact, data such as the number of celebrities may be needed to support further research.

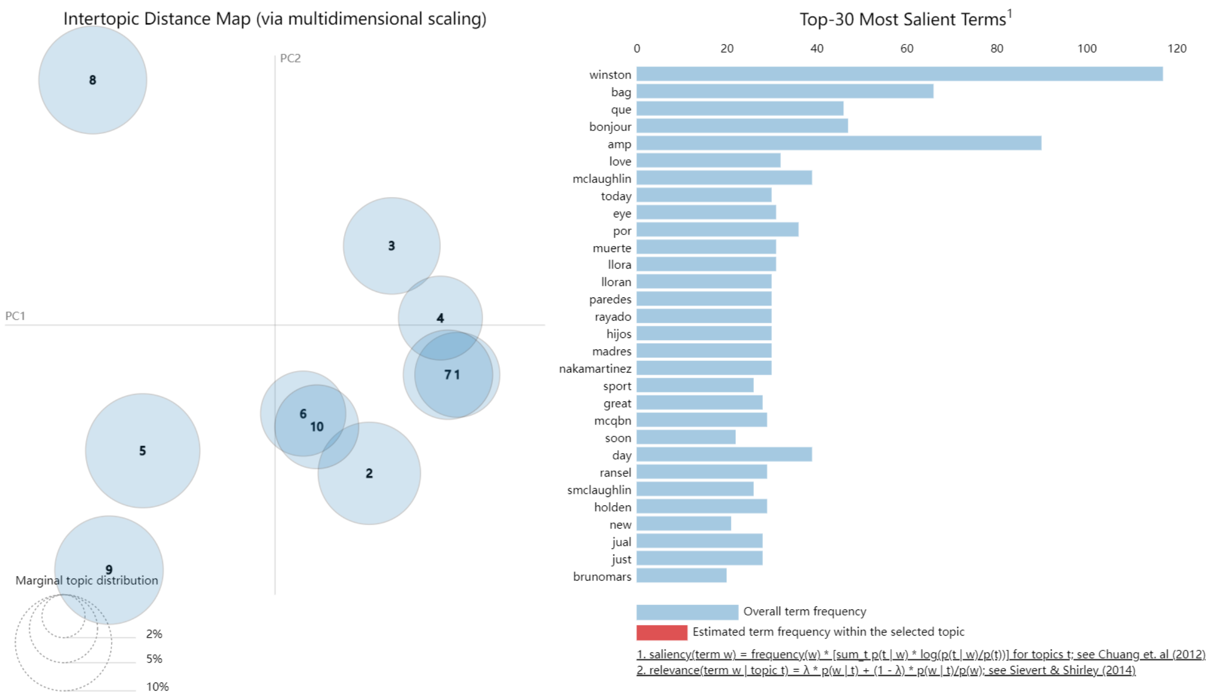
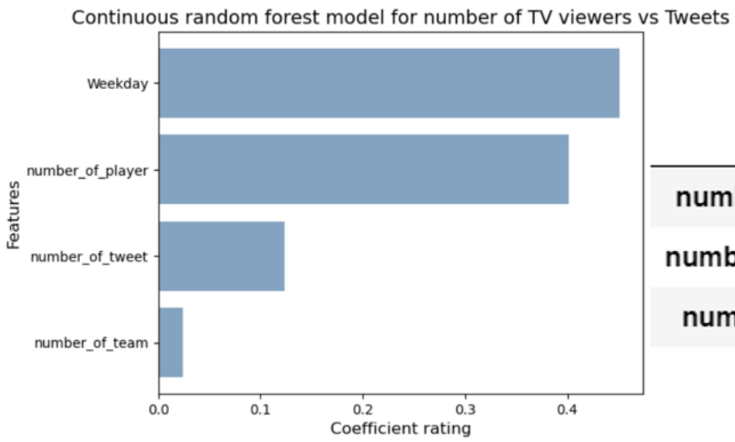


Figure 6. Main topics of segment 3

Several variables were used to predict TV ratings by using the OLS regression method. The results demonstrated a significant relationship between the TV ratings of segment 3 and three key variables, namely the number of players, number of teams and weekday. These variables exhibited statistical significance with p values lower than the 5% threshold (Appendix 7).

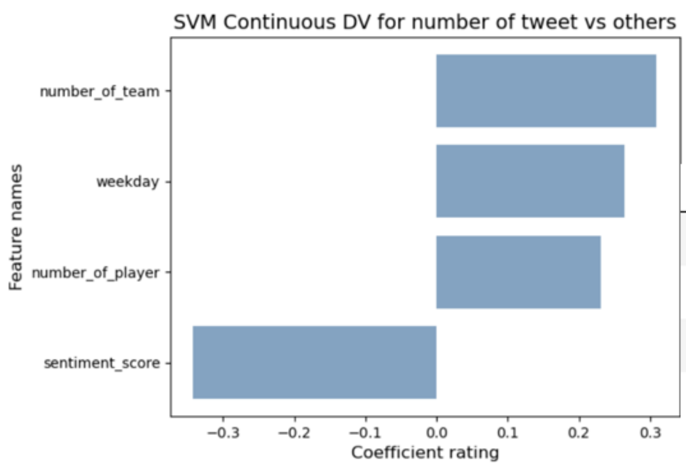
Furthermore, a continuous random forest model was implemented to explain the degree of influence of various factors on TV ratings. As shown in *Figure 7*, weekday and number of players received substantial influence on the TV ratings of segment 3, accounting for 45.11% and 40.16% respectively. This can be attributed to the preference of people to watch games on weekends, particularly Sundays, when they have more spare time. Additionally, an interesting finding emerged indicating that tweets had a considerable impact on TV viewership, accounting for 12.29%, underscoring the significant marketing potential in social media platforms.



	Feature Importance
number_of_tweet	0.122948
number_of_player	0.401611
number_of_team	0.024283
Weekday	0.451158

Figure 7. Continuous random forest model of TV viewers in segment 3

Subsequently, another regression model was performed to explore the relationship between the number of tweets and various other factors. The results revealed that the number of players, number of teams, sentiment score and weekday hold statistical significance with p value less than 0.05 (Appendix 8). Then, an SVM model was performed to assess the extent to which tweets were influenced by these aforementioned factors. *Figure 8* illustrates that an increase in week day is estimated to yield a rise of approximately 0.26 units in the number of tweets. Moreover, every additional player could lead to a 0.23 unit increase in tweets, while every additional team may result in 0.3 more tweets. Thus, these findings indicate that the quantity and diversity of players and teams have a huge influence on social media engagement.



	SVR coefficient rating
weekday	0.264625
number_of_player	0.231321
number_of_team	0.309337
sentiment_score	-0.341848

Figure 8. SVM model for number of tweets in segment 3

However, the R-squared of the regression model in segment 3 is much less than 1 and the error rates of the random forest model and SVM are both much higher than 0.3, indicating that the underfitting may be biased due to the small number of observations.

Overall, there are two recommendations for this segment. First, influencer marketing might help attract more potential viewers. For example, inviting celebrities with a large following and a genuine interest in the game to attend the race could prompt their fans to tune into TV (Lee, 2013). Second, the Supercars Championship could build a social media community that promotes TV content discussion and idea sharing. Marketing teams may also actively interact with the community, addressing queries and comments. In other words, this strategy not only increases social media engagement but also to get valuable feedback from customers (*Vantage unveils Supercar Blondie as Brand Ambassador*, 2023).

Segment 4

The analysis of topic modelling indicated notable trends among the audience. Particularly, Adelaide and the 2-in-1, 3-in-1, and 4-in-1 events stood out (*Figure 9*). There is a significant fan preference for Adelaide, which ironically had been an area of low TV ratings. The factors contributing to this preference could be traced back to the occurrence of the second youngest winner in Adelaide and the fact that the event concluded the 2022 season(Pavey, 2022). Simultaneously, the Winston Cup NASCAR series was identified as a prominent competitor. The name 'Winston' appeared frequently in the supercar fans' tweets, suggesting a shared fan base between supercar racing and NASCAR, which boasts a vast pool of ardent followers in the United States (Bonkowski, 2022).

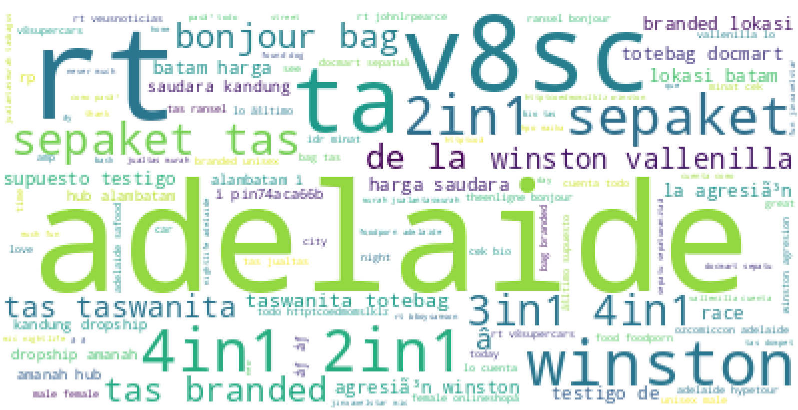


Figure 9. Word Cloud in segment 4

Sentiment analysis further enriched our understanding of the audience. As shown in *Figure 10*, a significant portion, 46.1%, demonstrated positive sentiment, while 40.8% were neutral, leaving a smaller fraction expressing negative sentiment. Segment 4 witnessed a higher concentration of positive tweets, which could be attributed to the Adelaide event, the fan-voted Best Event of the season for its unique viewer experience, intense competition, and high-calibre contenders.

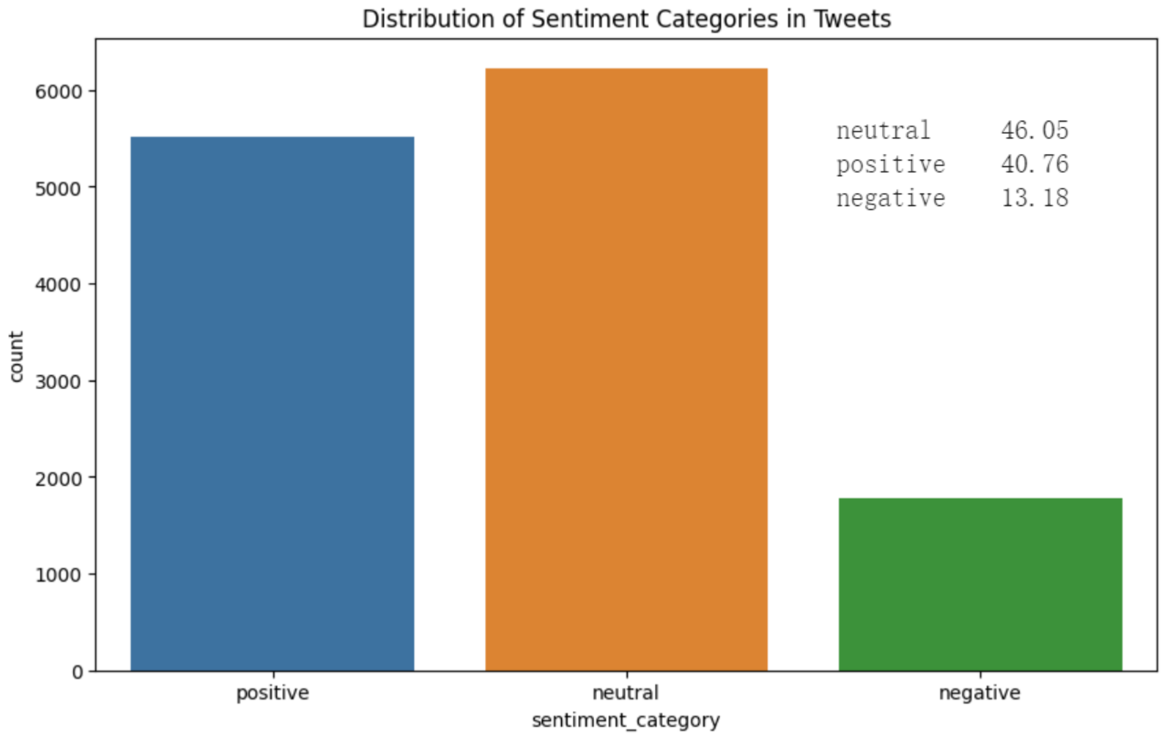


Figure 10. Distribution of sentiment categories in tweets.

To increase viewer engagement and ratings, we propose a dual-strategy approach. Firstly, we recommend revitalising event cities, including upgrades to existing circuits and local infrastructure to enhance the overall event appeal. Secondly, we suggest strategic differentiation of event timings to avoid overlap with our primary competitor, Winston Cup NASCAR. This tactic aims to maximise viewership by aligning with peak viewer availability and considering global time zones, hence avoiding a direct scheduling conflict with our competition. This approach will allow us to capture a broader audience, driving up our viewership and ratings.

Recommendation summary

In conclusion, TV viewership across most segments is significantly influenced by the number of tweets, team and player engagement, and sentiment. Moreover, the volume of tweets exerts the most considerable influence, presenting an opportunity to boost TV ratings through fostering more vibrant engagement on cross platforms. Harnessing the influence of teams and players can further help attract potential viewers via strategic marketing partnerships (Lee, 2013). Moreover, Supercars can enhance social media community engagement, which can lead to higher TV ratings. Finally, by collaborating with local communities to develop localised marketing campaigns, Supercars can seize the interest of local residents, thereby further increasing their ratings.

Limitations

The analysis of supercar racing viewership using Twitter data has some inherent limitations that are important to note. The accuracy of our models suffered due to a limited number of observations for certain ranges, and the exclusion of influential factors like player performance, event competitiveness, and external conditions like weather and concurrent events. Finally, Twitter data may not fully represent the diverse viewer base as it could be biased towards younger users, and sentiment scores could be skewed by highly influential users.

Further Research

Going forward, we propose several avenues for in-depth exploration. It is important to recommend segmenting the viewer base for targeted engagement, studying social media strategies of competitor racing events for effective adaptation, and devising a model for predicting viewership trends based on social media sentiment and activity. It could also be insightful to identify and analyse influencers among followers, who could be leveraged to boost the brand's visibility. By addressing the above limitations and implementing these further research, the Supercars Championship can enhance the understanding of their viewers and improve the social media strategy for future supercar racing events.

Reference

Autosport (2014) "Bathurst 1000: Chaz Mostert/Paul Morris win from last for Ford," *Autosport*. Available at: https://www.autosport.com/supercars/news/bathurst-1000-chaz-mostert-paul-morris-win-from-last-for-ford-5046355/5046355/.

Bonkowski, J. (2022, December 14). How R.J. Reynolds' Sponsorship, Winston Cup Became NASCAR Game Changer in 1971.Available at: How R.J. Reynolds' Sponsorship, Winston Cup Became NASCAR Game Changer in 1971 (autoweek.com)

Brittle, C. (2023) *Supercars ties up four partnerships for 2023*. Available at: https://www.blackbookmotorsport.com/news/supercars-partnerships-panasonic-schick-repco-boost-2023-sponsorship#:~:text=Panasonic%2C%20Schick%20and%20Repco%20Authorised,Boost%20Mobile%20extends%20existing%20deal.&text=The%20Supercars%20Championship%20has%20signed,existing%20deal%20with%20Boost%20Mobile.

Christina Rogers; Gautham Nagesh, Rogers, C., & Gautham Nagesh. (2016). Business News: Ford to Audition Supercar Buyers --- Shoppers should be prepared to provide videos, social media to get limited-edition GT. The Wall Street Journal. Eastern Edition. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_proquest_newspapers_1780734653

Dale, W. (2015) "V8 Supercars: Mark Winterbottom and Craig Lowndes' controversial clashes since 2013," *Fox Sports*. Available at: https://www.foxsports.com.au/motorsport/supercars/v8-supercars-mark-winterbottom-and-craig-lowndes-controversial-clashes-since-2013/news-story/48bd7cd6a431c

6a73e7a4b07729489fa.

Federico Pascual. (2019). Topic Modeling: An Introduction. Available at: https://monkeylearn.com/blog/introduction-to-topic-modeling/

La Ferle, C. and Choi, S.M. (2005) "The Importance of Perceived Endorser Credibility in South Korean Advertising," *Journal of Current Issues and Research in Advertising*, 27(2), pp. 67–81. Available at: https://doi.org/10.1080/10641734.2005.10505182.

Lee, J. G. (2013). Supercar Lifestyle promotion challenged. Journal of Direct, Data and Digital Marketing Practice, 15(1), 83–84. https://doi.org/10.1057/dddmp.2013.41

Pavey, J. (2022, June 23). Adelaide 500 locked in as finale until 2026. Supercars.Available at: Adelaide 500 locked in as finale until 2026 - Supercars

Supercars Championship (n.d.) *Our History*. Available at: https://www.supercars.com/championship/our-history/.

V8 Supercars (2014) *SVG "gutted" after starter motor failure - Supercars*. Available at: https://www.supercars.com/news/championship/svg-gutted-after-starter-motor-failure/

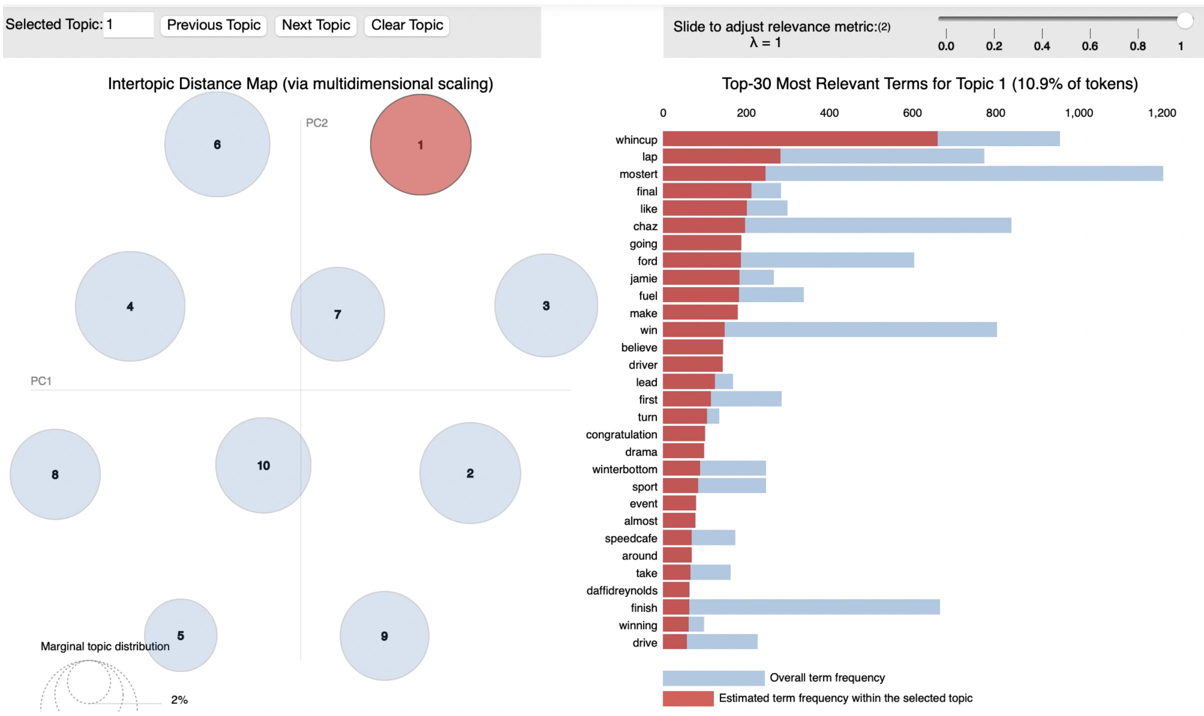
Vantage unveils Supercar Blondie as Brand Ambassador. (2023). PR Newswire Association LLC. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_proquest_wirefeeds_2766349409

Appendix 1 (Ranking of average TV viewership)

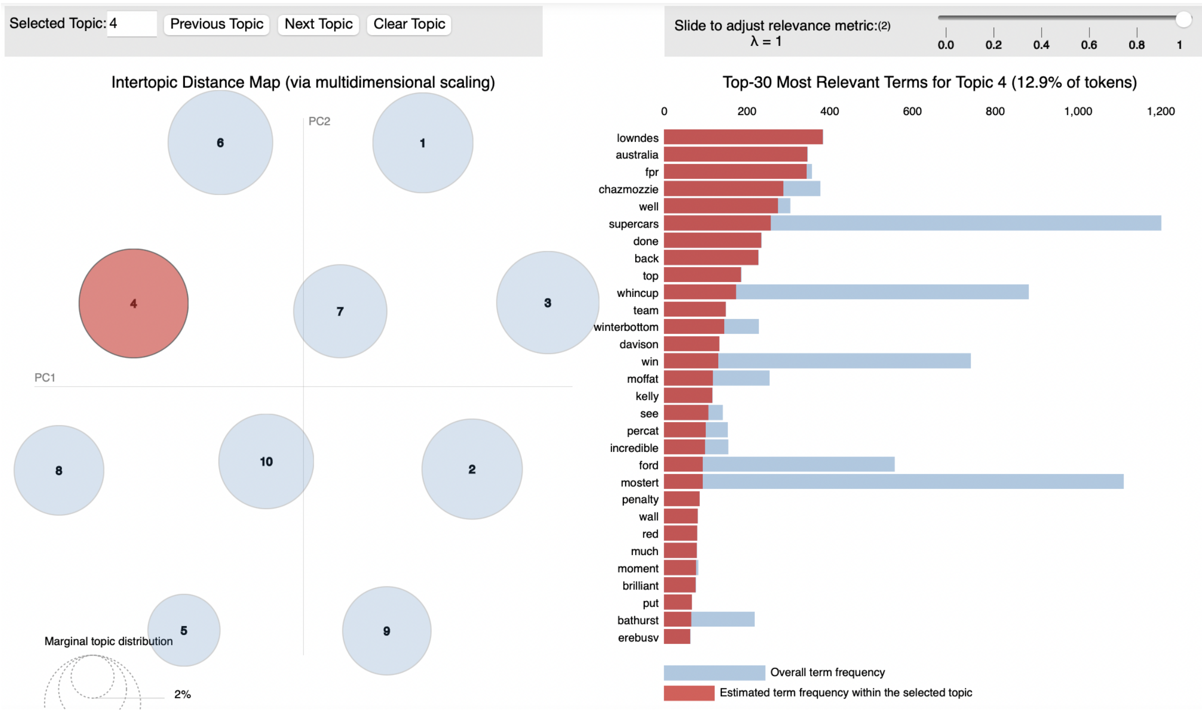
```
event_number
11    139.534791
14     73.388889
12     62.917857
13     59.634615
7      57.700000
1      57.613811
10     52.112745
2      50.865385
9      47.957317
3      45.067358
5      40.953608
8      40.851351
6      40.091398
4      33.386266
Name: tvviewers, dtype: f
```

Segment 1 refers to the event that has the highest average TV viewership, specifically event 11. Segment 2 comprises events 12, 13, and 14. Segment 3 includes events 1, 2, 7, and 10. The remaining events are categorised as segment 4.

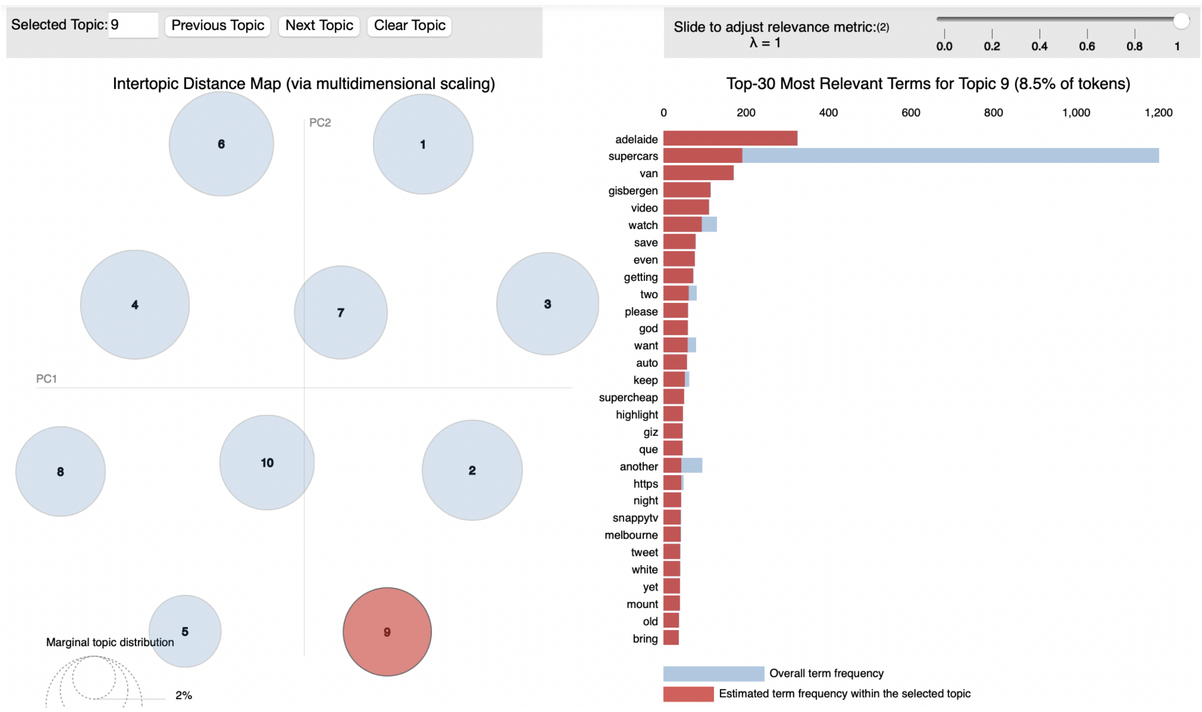
Appendix 2 (Topic modelling in segment 1 - topic 1)



Appendix 3 (Topic modelling in segment 1 - topic 4)



Appendix 4 (Topic modelling in segment 1 - topic 9)



Appendix 5 (Results of model 1 & 2 in segment 2)

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):	0.045
Model:	OLS	Adj. R-squared (uncentered):	0.040
Method:	Least Squares	F-statistic:	10.14
Date:	Fri, 12 May 2023	Prob (F-statistic):	1.56e-06
Time:	01:23:39	Log-Likelihood:	-3783.5
No. Observations:	652	AIC:	7573
Df Residuals:	649	BIC:	7587
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x2	1.8524	0.676	2.739	0.006	0.525	3.180
x3	6.7308	2.130	3.160	0.002	2.549	10.913
x4	23.9729	6.866	3.491	0.001	10.490	37.456

Omnibus:	24.527	Durbin-Watson:	0.144
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26.289
Skew:	0.471	Prob(JB):	1.96e-06
Kurtosis:	3.280	Cond. No.	10.2

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):	0.225
Model:	OLS	Adj. R-squared (uncentered):	0.220
Method:	Least Squares	F-statistic:	46.95
Date:	Fri, 12 May 2023	Prob (F-statistic):	1.14e-34
Time:	01:03:12	Log-Likelihood:	-3715.5
No. Observations:	652	AIC:	7439.
Df Residuals:	648	BIC:	7457.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x2	0.8039	0.613	1.312	0.190	-0.399	2.007
x3	2.3239	1.677	1.386	0.166	-0.969	5.617
x4	19.9446	6.530	3.054	0.002	7.123	32.767
x1	86.5800	7.000	12.368	0.000	72.834	100.326

Omnibus:	13.678	Durbin-Watson:	0.486
Prob(Omnibus):	0.001	Jarque-Bera (JB):	14.246
Skew:	0.361	Prob(JB):	0.000806
Kurtosis:	2.945	Cond. No.	11.9

Appendix 6 (Results of model 3 & 4 in segment 2)

and should_run_async(code)

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):	0.181
Model:	OLS	Adj. R-squared (uncentered):	0.177
Method:	Least Squares	F-statistic:	38.96
Date:	Fri, 12 May 2023	Prob (F-statistic):	5.19e-16
Time:	02:00:33	Log-Likelihood:	-140.97
No. Observations:	354	AIC:	285.9
Df Residuals:	352	BIC:	293.7
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
number_of_tweet	0.0234	0.003	8.050	0.000	0.018	0.029
number_of_player	-0.0013	0.021	-0.064	0.949	-0.043	0.040

Omnibus:	3.759	Durbin-Watson:	1.864
Prob(Omnibus):	0.153	Jarque-Bera (JB):	3.164
Skew:	0.135	Prob(JB):	0.206
Kurtosis:	2.624	Cond. No.	7.94

Notes:

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):	0.021
Model:	OLS	Adj. R-squared (uncentered):	0.016
Method:	Least Squares	F-statistic:	4.533
Date:	Fri, 12 May 2023	Prob (F-statistic):	0.00372
Time:	01:31:27	Log-Likelihood:	-304.79
No. Observations:	652	AIC:	615.6
Df Residuals:	649	BIC:	629.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x2	0.0022	0.003	0.670	0.503	-0.004	0.009
x3	0.0414	0.018	2.294	0.022	0.006	0.077
x4	0.0920	0.034	2.687	0.007	0.025	0.159

Omnibus:	12.522	Durbin-Watson:	1.492
Prob(Omnibus):	0.002	Jarque-Bera (JB):	10.088
Skew:	0.219	Prob(JB):	0.00645
Kurtosis:	2.577	Cond. No.	10.6

Appendix 7 - Results of regression analysis in segment 3

OLS Regression Results

Dep. Variable:	tvviewers	R-squared:	0.052
Model:	OLS	Adj. R-squared:	0.046
Method:	Least Squares	F-statistic:	8.612
Date:	Tue, 23 May 2023	Prob (F-statistic):	5.86e-08
Time:	13:00:59	Log-Likelihood:	-1228.8
No. Observations:	797	AIC:	2470.
Df Residuals:	791	BIC:	2498.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.8695	0.428	4.368	0.000	1.029	2.710
number_of_tweet	-0.0349	0.008	-4.535	0.000	-0.050	-0.020
number_of_player	-0.0583	0.040	-1.466	0.143	-0.136	0.020
number_of_team	0.3168	0.130	2.435	0.015	0.061	0.572
sentiment_score	-0.0782	0.258	-0.303	0.762	-0.584	0.428
weekday	0.3262	0.070	4.682	0.000	0.189	0.463

Omnibus:	911.251	Durbin-Watson:	0.884
Prob(Omnibus):	0.000	Jarque-Bera (JB):	112091.575
Skew:	-5.373	Prob(JB):	0.00
Kurtosis:	60.096	Cond. No.	138.

Appendix 8 - Results of regression analysis in segment 3

Dep. Variable:	number_of_tweet	R-squared:	0.329
Model:	OLS	Adj. R-squared:	0.326
Method:	Least Squares	F-statistic:	97.28
Date:	Tue, 23 May 2023	Prob (F-statistic):	2.43e-67
Time:	13:00:59	Log-Likelihood:	-754.63
No. Observations:	797	AIC:	1519.
Df Residuals:	792	BIC:	1543.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0045	0.234	-0.019	0.985	-0.463	0.454
number_of_player	0.2347	0.020	11.906	0.000	0.196	0.273
number_of_team	0.3019	0.067	4.492	0.000	0.170	0.434
sentiment_score	-0.4724	0.142	-3.337	0.001	-0.750	-0.194
weekday	0.2874	0.037	7.772	0.000	0.215	0.360