A comparison of Betting User engagement between Fanduel and Draft Kings Using Twitter Data

Executive Summary

This study investigates how user engagement on Twitter varies between the betting platforms FanDuel and DraftKings. Utilizing python, sentiment analysis and statistical tests such as OLS regression and Mann U Whitney test were conducted on Twitter data from February 27 to March 1, 2021, to compare different metrics like retweets and likes with engagement. Results showed significant differences in user engagement between the sportsbooks, suggesting variations in content appeal and marketing effectiveness. These insights indicate that aligning marketing strategies more closely with user preferences could enhance engagement. The study highlights the limitations of Sentiment analysis, the importance of promotions in optimizing user interaction, and other future implications of the results.

Introduction

Using sentiment analysis, a method of gathering and collecting subjective data from text. The process analyzes and categorizes textual content's ideas, attitudes, feelings, and emotions as positive, negative, or neutral. Given that there is so much user-generated content on internet platforms like social media, blogs, reviews, and discussion forums, sentiment analysis has grown in popularity recently. This analysis is crucial for gaining deeper insights into user engagement with betting platforms on Twitter, specifically comparing FanDuel and DraftKings. It improves comprehension of the elements that impact user interactions and engagement, which is crucial for the marketers, strategists, and decision-makers of these companies. These kinds of insights are crucial for improving user experience, optimizing communication tactics, and customizing material to better suit audiences' preferences and expectations. This research will help shape future marketing and operational initiatives in addition to enhancing engagement approaches. The central question of this study is to assess the differences in user engagement between FanDuel and DraftKings on Twitter, aiming to understand how engagement patterns vary between these two prominent betting platforms. The methodology involves collecting and analyzing Twitter data specific to FanDuel and DraftKings, spanning from February 27, 2021, to March 1, 2021. The analysis assesses metrics including likes, retweets, and engagement using statistical approaches such as OLS regression and Mann U Whitney test as well as sentiment analysis . This approach helps quantify the extent of user engagement and the other variables interact toward each platform. The analysis revealed distinct engagement patterns, indicating significant differences in user engagement (p value = 0.014) between the sportsbooks with Draft Kings having notably higher engagement (Figure 4). Metrics like retweet count and impressions were notably correlated with engagement as well (Figure 2), suggesting tweets with higher impressions and retweet count were more effective with gaining engagement on Twitter. Finally we seen that device type (p value = 0.004) also had a statistically significant relationship with engagement with Androids having higher Engagement. Conclusions and Implications of the results indicate that engagement differences could help influence marketing strategies and operational decisions for FanDuel and DraftKings.  By better understanding these engagement patterns, each platform may be able to better tailor its promotions and content to the interests of its users, which might lead to an increase in user loyalty and activity. To provide a more comprehensive understanding, future study could build on this work by examining longer time periods and more variables across the platform to gain a broader understanding of user engagement trends in the sports betting industry. Additionally, the use of a trained supervised model would bring more significant results and provided a more descriptive explanation of the results.

Conceptual Background

Sentiment analysis, or SA, has grown to be an essential tool for companies, allowing them to gather client input and effectively formulate their marketing plans. By extracting insights from customer sentiments expressed on various platforms, companies can fine-tune their marketing approaches, improve product offerings, and enhance customer service experiences. In a similar way, SA is essential for academics who examine public opinion on a variety of subjects. They can use it to gauge how people are responding to events, identify changes in public opinion, and comprehend societal patterns. This strong analytical capacity helps both private and public companies make well-informed decisions based on consumer sentiment. ( Gunasekaran 2020) concluded that when SA is used rightly, “it gives useful information of public opinion which can result in making better business decisions. It also helps in predicting market trends.” There are a few popular methods of Sentimental analysis according to (Qi & Shabrina) the “There are three main methods to detect and classify emotions expressed in text, which are lexicon-based, machine-learning-based approaches, and hybrid techniques”. To extract meaning from text, “ Sentiment analysis identifies and extracts subjective information from the text using natural language processing and text mining.” ( Wankhade, Rao, & Kulkarni, 2022). Fanduel was founded in 2009 as part of a rebrand from a previous venture called Hubdub, a news prediction site, and started as a Daily Fantasy Sports app allowing for users to play money in fantasy sports lineups that the combined company would hold a monopolistic position in the market. FanDuel leverages their wide variety of props with the use of an amazing same game parlay builder that allow you to combine multiple same game parlays for massive payouts. Draft Kings was founded in 2007 similarly with roots as a Daily fantasy app. New customers may usually take advantage of attractive sign-up bonuses from DraftKings, which include deposit matching or free bets. For instance, a typical promotion might provide a "free square" bet such as over 0.5 points for Lebron that serves as a safety net for first time bettors or a sizable match on a user's initial deposit. These bonuses not only draw in new users, but also entice them to place bets right away, boosting user engagement from the jump. FanDuel and DraftKings announced their intention to merge in 2017. The Federal Trade Commission, fearing that the combined business would hold a monopoly position in the market, had antitrust concerns, which led to the termination of the merger. A year later in 2018 Online sports betting became a statewide issue and both these Sportsbooks took off to new heights being pioneers of in app sports betting. When comparing user engagement between FanDuel and DraftKings on Twitter using sentiment analysis, I hypothesize that user’s engagement with DraftKings-related content will be more frequent than FanDuel-related content. This hypothesis comes from variations in marketing tactics, user interface elements, and promotional strategies.

Data and Methodology

This analysis explores the variance in engagement between FanDuel and DraftKings on Twitter. We examined 8,406 entries of Twitter data, gathered from February 27, 2021, to March 1, 2021, categorized by the respective sportsbook.

FanDuel Summary Statistics:

Table 1**Polarity and Subjectivity**: On average, the polarity (how positive or negative the sentiment is) is 0.128, indicating a slightly positive sentiment. The average subjectivity (how much opinion vs. fact is in the content) is about 0.296.

**Favorite and Retweet Counts**: Posts have an average of about 1.87 favorites and 17.93 retweets.

**Followers and Friends**: Average followers count is approximately 15,932, suggesting a medium to large audience size for the posts, with an average of 1,409 friends.

Impressions and Reach: Average impressions are about 19.8, and reach is around 17,340.

**Engagement and Engagement Rate**: There are some anomalies in these values (recorded as 'inf'), which might indicate missing or extreme values affecting the calculation.

Draft Kings Summary statistics:

Table 2



**Polarity and Subjectivity**: Similarly, the average polarity is 0.123, also slightly positive, and subjectivity is close to that of DraftKings at 0.297.

Favorite and Retweet Counts: FanDuel posts average about 2.47 favorites and 13 retweets.

**Followers and Friends**: Average followers count is lower at about 7,406, with 949 friends.

**Impressions and Reach**: Average impressions are 15.47, and reach is around 8,355.

**Engagement and Engagement Rate**: These also show anomalies with 'inf' values.

These ‘inf’ values in Tables 1 & 2 show that our data has some division of 0 or other anomalies that can cause misinterpretation of the data which is why is was essential to clean null values when running statistical tests.

As part of my empirical strategy, I analyzed the data gathered for the study using an systematic approach Here's an outline of the methods used:

**Data Collection**: The source of data was Twitter, where I collected FanDuel and DraftKings-related tweets. Metadata like mentions, likes, and retweet counts were included in this.

**Data Cleaning**: In order to handle missing numbers, fix errors, and eliminate any unnecessary information, the raw data was processed. In doing this I combined the two books into a singular table and created binary variables for device type and sports book. In this step I made Draft kings = 0 and Fanduel =1, as well as Android = 0 and iPhone = 1. Additionally, I computed engagement metrics by using the formula, engagement =  "impressions"  / "reach", with reach being a measure of ‘followers’ / ‘friends’  This provided assurance that the analysis would be conducted on reliable and comprehensive data.

**Descriptive statistics**: I performed descriptive statistics to obtain a general picture of the data distribution before moving on to more comprehensive research. This helped me to recognize patterns and irregularities in the data.

**Regression Analysis**: To measure the effect of different factors on user engagement, I used OLS regression models. This helped in modeling and understanding the extent to which sports books influenced engagement.

**Further Hypotheses Testing**: To confirm  if observed differences were statistically significant, specific hypotheses regarding engagement differences were examined using the Mann-Whitney U test.

**Results Interpretation**: The analysis concluded with a thorough explanation of the statistical results, linking them to the study's original theoretical framework and research questions.

Results

When analyzing the data I worked in Jupyter notebook using python to combine the separated twitter datasets, creating a combined dataset named combined books in order to get run statistical testing. Using the binary variable sportsbook I was able to decipher between the Fanduel (1) and Draft kings data (0) and ran my first OLS regression to see the significance on engagement. Upon receiving a perfect R-Squared value (1.00) from my regression in figure 1below I suspected there may be overfitting in my data as its highly unlikely all variables of engagement were encapsulated.

Results: Ordinary least squares

============================================================================

Model: OLS Adj. R-squared: 1.000

Dependent Variable: engagement AIC: -457308.2832

Date: 2024-05-09 00:02 BIC: -457244.9948

No. Observations: 8367 Log-Likelihood: 2.2866e+05

Df Model: 8 F-statistic: 2.069e+28

Df Residuals: 8358 Prob (F-statistic): 0.00

R-squared: 1.000 Scale: 1.0720e-25

----------------------------------------------------------------------------

Coef. Std.Err. t P>|t| [0.025 0.975]

----------------------------------------------------------------------------

const 0.0000 0.0000 0.0388 0.9691 -0.0000 0.0000

polarity -0.0000 0.0000 -0.0221 0.9824 -0.0000 0.0000

subjectivity -0.0000 0.0000 -0.1240 0.9013 -0.0000 0.0000

engagement\_rate 0.0100 0.0000 403476521933915.4375 0.0000 0.0100 0.0100

friends -0.0000 0.0000 -88.7663 0.0000 -0.0000 -0.0000

followers 0.0000 0.0000 4.9400 0.0000 0.0000 0.0000

impressions -0.0000 0.0000 -0.0255 0.9796 -0.0000 0.0000

favorite\_count -0.0000 0.0000 -0.0089 0.9929 -0.0000 0.0000

retweet\_count -0.0000 0.0000 -0.0377 0.9700 -0.0000 0.0000

sportsbook -0.0000 0.0000 -0.0319 0.9745 -0.0000 0.0000

----------------------------------------------------------------------------

Omnibus: 20088.531 Durbin-Watson: 1.863

Prob(Omnibus): 0.000 Jarque-Bera (JB): 178562607.431

Skew: 24.857 Prob(JB): 0.000

Kurtosis: 716.947 Condition No.: 1549467422265496064

============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is

correctly specified.

[2] The smallest eigenvalue is 9.98e-23. This might indicate that

there are strong multicollinearity problems or that the design

matrix is singular.

Figure 1

When examining the dataset deeper I decided to remove engagement rate from my analysis as it directly correlates with engagement. Instead, I introduced a binary variable titled 'device type' to see the significance of device on user engagement and provide a more comprehensive look into the influencing factors of the data. After this second look, I observed the R-squared value of 0.030 from the regression in Figure 2 below indicating that the model explains only 3% of the variance in engagement. This implies that the current model isn’t effectively capturing all factors of engagement between the datasets, suggesting a more well-trained model or more diverse dataset would be better explain the variable.

OLS Regression Results

==============================================================================

Dep. Variable: engagement R-squared: 0.030

Model: OLS Adj. R-squared: 0.029

Method: Least Squares F-statistic: 23.18

Date: Wed, 08 May 2024 Prob (F-statistic): 2.79e-35

Time: 23:20:36 Log-Likelihood: -7972.7

No. Observations: 6004 AIC: 1.596e+04

Df Residuals: 5995 BIC: 1.602e+04

Df Model: 8

Covariance Type: nonrobust

==================================================================================

coef std err t P>|t| [0.025 0.975]

----------------------------------------------------------------------------------

const 0.1249 0.031 4.073 0.000 0.065 0.185

polarity 0.0365 0.043 0.855 0.393 -0.047 0.120

subjectivity 0.0657 0.040 1.662 0.097 -0.012 0.143

favorite\_count -0.0005 0.000 -2.580 0.010 -0.001 -0.000

retweet\_count 0.0010 0.000 8.159 0.000 0.001 0.001

followers -9.366e-08 2.44e-07 -0.384 0.701 -5.72e-07 3.84e-07

friends -3.093e-06 2.1e-06 -1.471 0.141 -7.22e-06 1.03e-06

impressions 0.0005 9.71e-05 5.343 0.000 0.000 0.001

sportsbook -0.0589 0.024 -2.464 0.014 -0.106 -0.012

device\_type -0.0806 0.028 -2.845 0.004 -0.136 -0.025

==============================================================================

Omnibus: 18969.568 Durbin-Watson: 1.610

Prob(Omnibus): 0.000 Jarque-Bera (JB): 2417282762.392

Skew: 49.977 Prob(JB): 0.00

Kurtosis: 3109.879 Cond. No. 1.64e+16

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The smallest eigenvalue is 5.42e-20. This might indicate that there are

strong multicollinearity problems or that the design matrix is singular.

Figure 2

Statistic: 9565693.0, P-value: 1.597507156865993e-12

Figure 3

The Mann U Whitney test (Figure 3) further confirms our findings that there is a significant difference in engagement between the Draft Kings and Fanduel furthermore confirms We can accept the null hypothesis that the engagement between Draft kings is statically significant based of the results from our Mann Whitney U test.

When discussing these results, it’s important to note during the time of collection there was a limit to the number of sports being played constraining the amount of bets able to be placed and possibly affecting engagement. A lack of variables also noticeably limited the ability to discern the largest factors of engagement as only about 3% of the causes (figure 2) are being accounted for. One of the largest limitations of our analysis is domain dependence, when models used for sentiment analysis are trained, they use specific datasets for the opinionative words when in the real-world language is very domain dependent. For example, the word "unpredictable" might have a negative connotation in the context of car performance reviews but could be positive in a movie review discussing a plot twist (Wunderlich & Memmert, 2020). Sentiment analysis models trained on one domain might misinterpret the sentiment of terms in another similarly here that could be happening as the model isn’t specifically trained for sports or betting analysis. Additionally, only a few days (February 27, 2021, to March 1, 2021) went into the dataset. This minimal period may not have allowed for the capturing of larger trends or variations in engagement which could have been influenced by specific events, seasonal sports activities, or marketing campaigns that occurred outside of these dates. I would love to see similar data during the month of October when all major professional sports are going as this may cause for higher results and better analysis due to a larger and more diverse sample set.

Key Findings

A graph of a number of blue squares

Description automatically generated with medium confidence

Figure 4

The key finding that allows us to accept our null hypothesis that DraftKings has more user engagement is the variable are our p value of 0.014 in figure 1 and our Mann U Whitney score in Figure 2. sportsbook also demonstrated a negative coefficient, indicating noticeable fluctuations in engagement across the platforms. This variability suggests that there is a significant relationship between user engagement and sportsbook which can be seen above in figure 4.

A graph with green and black dots

Description automatically generated

Figure 5

A graph with blue and black lines

Description automatically generated

Figure 6

On the flip side, we also discovered that the variables retweet count and impressions had a clear positive relationship on engagement based off their respective coefficients (0.0010, 0.0005) and p values ( 0.00) (Figure 2). This relationship is shown at the 0.01 confidence level suggesting a high statical correlation that as these metrics increase, so does user engagement. This trend was consistently observed, and reinforced when the regression line is put against a graph(Figures 5& 6 ) .

A graph of a bar graph

Description automatically generated with medium confidence

Figure 7

Furthermore, our analysis revealed that the type of device used to access Twitter, separated by Android (0) and iPhone (1), had a discernible correlation on engagement as viewed in figure 7, with a regression coefficient of -0.0806 and p value of 0.014(Figure 2). This suggests that engagement levels can significantly vary at a confidence interval of 0.05 depending on the technology through which users interact with content.

Conclusion & Future work

An extended period of data collecting is necessary to capitalize on the use of the sentiment analysis fully. An approach where all key sports events throughout the year are captured would uncover seasonal patterns that could significantly influence user engagement. Looking into the Graphical interface differences may be something that can help explain the drop in engagement between android and iPhone users and is a cheap method which testing could be implemented. In addition, the accuracy of the sentiment analysis could be improved with the implementation of a supervised sentiment model that has been specially trained on sports betting data. The distinctive vocabulary and idioms frequently used in sports can be misinterpreted by traditional sentiment analysis algorithms. A model with direct training on sports betting dialogue would be better able to identify the subtleties of user sentiment in this situation, offering more in-depth understanding of the feelings and preferences of the target audience. Finally, it's critical to consider how particular marketing and promotions influence user engagement. DraftKings and FanDuel can use the correlation between retweet count and engagement to their advantage by employing additional influence marketing, giving promo codes for professional sports betters to post to intern increase app engagement. By examining the kinds of promos that people respond to the most and figuring out which athletes have the highest sentiment Sports books would make certain that marketing funds are spent as effectively as possible.

Table of Figures

[Table 1 5](#_Toc166190172)

[Table 2 5](#_Toc166190173)

[Figure 1 9](#_Toc166194102)

[Figure 2 10](#_Toc166194103)

[Figure 3 10](#_Toc166194104)

[Figure 4 11](#_Toc166194105)

[Figure 5 12](#_Toc166194106)

[Figure 6 13](#_Toc166194107)

[Figure 7 14](#_Toc166194108)

# Works Cited

Wunderlich, F., & Memmert, D. (2020, January 7). *Innovative approaches in sports science-lexicon-based sentiment analysis as a tool to analyze sports-related Twitter communication*. MDPI. <https://www.mdpi.com/2076-3417/10/2/431#B28-applsci-10->

00431 Bradley, A., & James, R. J. E. (2019). How are major gambling brands using Twitter? *International Gambling Studies*, *19*(3), 451–470. <https://doi.org/10.1080/14459795.2019.1606927>

Gunasekaran, K. P. (2023, May). *Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review*. arrive.org. https://arxiv.org/pdf/2305.14842

Qi, Yuxing, and Zahratu Shabrina. "Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach." *Social Network Analysis and Mining* 13.1 (2023): 31.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780.