**METHODOLOGY**

**Overview of Research Design**

The research design for this project employs discourse analysis, which involves an in-depth examination of written language in the context of social reality. Specifically, the study focuses on comments made by football fans on Facebook, with a particular emphasis on posts related to Arsenal and Real Madrid fan clubs, analyzing user comments to identify patterns in sentiment and engagement. The discourse analysis takes into account various factors such as societal norms, political climate, time, place, intended audience, and socio-cultural background of the speaker. The dataset obtained from Apify includes information about the comments, including author details, comment counts, and reaction counts.

**Definitions of Controversial Terms**

**User Verification:** Refers to the process of confirming the authenticity of a user's account on Facebook. Verified users typically have a higher level of credibility.

**Sentiment Analysis:** Involves the use of natural language processing techniques to determine the emotional tone of comments—whether they are positive, negative, or neutral

**Author Details:** The classification of author gender involves potentially contentious decisions as it may rely on assumptions or algorithms.

**Engagement Metrics:** The term "controversial notion of online influence" refers to the debate around the true impact and significance of comments and reactions on social media platforms.

**Textual Content:** Refers to the content of comments, which could be a source of contentious sentiments and opinions.

**Data and Data Collection**

**Data Source**

The dataset is sourced from Apify and includes information such as author details, comment counts, reaction counts, and the text of comments. The dataset is focused on posts related to Arsenal and Real Madrid in Facebook fan clubs.

**Data Cleaning**

After loading the dataset using pandas in Python, irrelevant columns (e.g., URL, author ID) are dropped, and missing values are handled through data cleaning procedures. The primary data attributes include author details, comment and reaction counts, and the sentiment of comments.

**Exploratory Data Analysis**

**Descriptive Analysis**

The descriptive analysis was conducted to provide a foundational understanding of key characteristics within the dataset. The analysis encompassed several aspects, starting with an exploration of user demographics, including verification status and gender distribution, and engagement metrics including comment\_count and reaction\_count.

**Discourse Analysis (Sentiment Analysis)**

In the sentiment analysis process, the Natural Language Toolkit's SentimentIntensityAnalyzer is employed to classify comments as positive, negative, or neutral based on their compound sentiment scores. The **find\_sentiment** function is created to classifier the text, it utilizes this analyzer, categorizing comments with a compound score greater than 0 as "Positive," those with a score less than 0 as "Negative," and those with a score of 0 as "Neutral." This method provides a nuanced understanding of the emotional tone of the comments.

**Data Visualizations**

In the context of the football fan comments dataset, data visualization played a crucial role in presenting key insights and patterns to facilitate a better understanding of user engagement and sentiment. A series of visuals including count plots, categorical plots, horizontal bar chart, bar charts and word clouds, are employed to present sentiment patterns and communicate findings.

**User Segmentation Analysis**

User segmentation analysis is aimed at uncovering nuanced insights into how demographic factors influence user engagement on social media platforms, contributing to a more holistic understanding of fan behaviour.

By integrating sentiment analysis with user segmentation, insights into how sentiments are expressed within specific user segments are uncovered, providing a more comprehensive understanding of the dynamics of football fan interactions on social media.

**SUMMARY**

* **Research Design:** Employing discourse analysis, the study delves into written language nuances, considering contextual factors such as societal norms and political climate.
* **Controversial Terms Defined:** Clear definitions are provided for terms like User Verification, Sentiment Analysis, Author Details, Engagement Metrics, and Textual Content.
* **Data and Collection:** Sourced from Apify, the dataset undergoes meticulous cleaning, focusing on key attributes, including author details, comment counts, reaction counts, and textual content.
* **Exploratory Data Analysis:** Unveiling insights into user demographics, verification status, gender distribution, and engagement metrics.
* **Omitted Sections:** The summary excludes detailed sections on Data Cleaning, Discourse Analysis (Sentimental Analysis), and Data Visualization.
* **Key Contributions:** The methodology involves refining the dataset, extracting sentiment insights, and presenting findings through visualizations.
* **Holistic Understanding:** The research aims to provide a comprehensive understanding of football fan interactions on social media, integrating sentiment analysis with user segmentation.

Top of Form

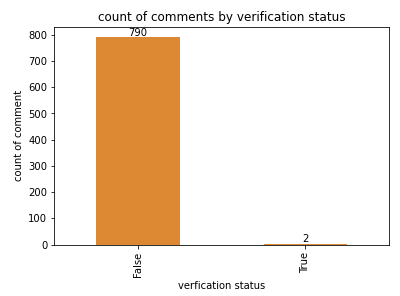
**RESULT AND FINDINGS**

**Overview**

In this section, we aim to present the findings obtained from the analysis and share valuable insight, contributing to the understanding of football fans' engagement on Facebook. The analysis reveals insights into the engagement patterns of football fans on Facebook, focusing on sentiments, user demographics, and verification status.

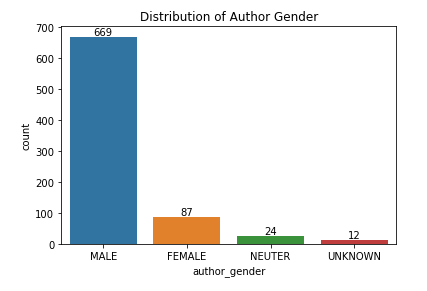
**FINDINGS**

**Count of Comments by Verification Status**

****

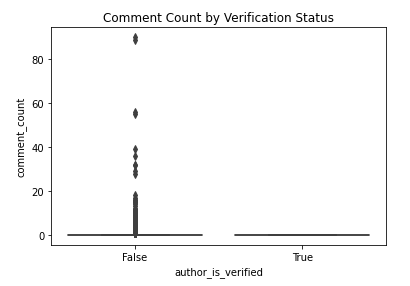
The above visual shows that over 99% of the users are using unverfied account while less than 1% of the users are using verified account

**Distribution of Author Gender**

****

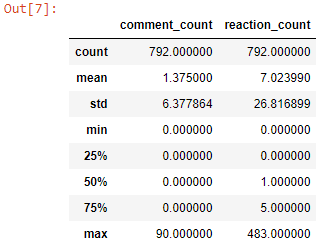
The above visuals shows that the gender with highest no of comments is the male gender with more than 60 counts, then the female gender and neuter gender while there are very few comments with unknown gender.

**Comment Count by Verification Status**

****

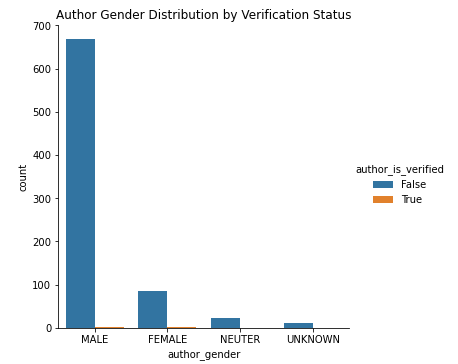
Most of the authors are not verified while only few has verified accounts.

**Statistical Summary of Comment Count and Reaction Count**



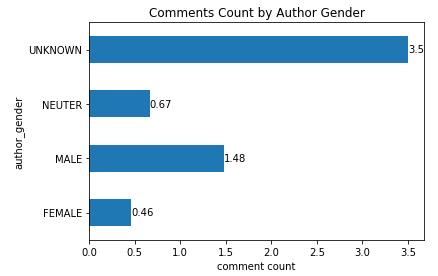
The statistic summary shows the average comment\_count to be 1.375 and reaction count 7,023, it also shows that more than 75% of the comment\_count is 0 and about 50% of the reaction ranges from 0 to 1 and remaining from 5 to 483 reaction\_ counts.

**Author Gender Distribution by Verification Status**



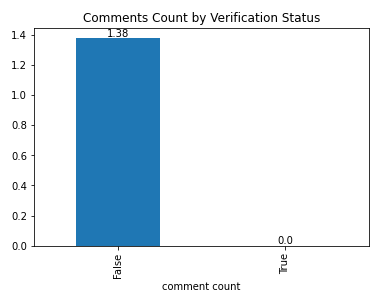
The plot shows that only the female and male gender authors has very few verified account while none of the neuter and unknown gender has a verified account

**Comments Count by Author Gender**



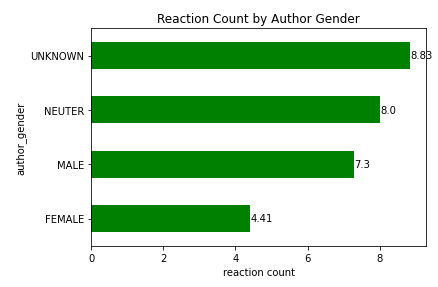
On average, female authors contribute around 0.46 comments, indicating a relatively lower engagement level compared to other gender groups. Male authors, on the other hand, have a higher average comment count of approximately 1.48, suggesting a more active participation in commenting. Authors with a neuter gender have an average comment count of about 0.67. Authors with an unknown gender, surprisingly, have the highest average comment count of 3.5. This could be due to various reasons, such as a smaller sample size or specific trends within this category.

**Comments Count by Verification Status**



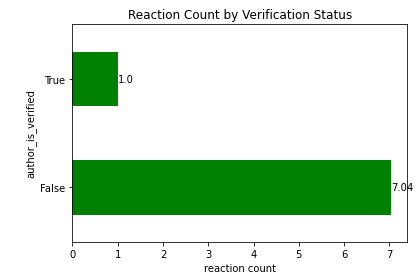
Unverified authors have an average comment count of 1.38, showcasing consistent engagement across the majority of users. Verified authors, however, have an average comment count of 0. This is an interesting finding that could be attributed to the small sample size of verified authors in the dataset.

**Reaction Count by Author Gender**



Female authors receive an average of approximately 4.41 reactions per comment. Male authors receive a higher average of about 7.30 reactions per comment, indicating a potentially more engaged audience. Authors with a neuter gender receive an average of 8 reactions per comment. Authors with an unknown gender receive the highest average reactions, around 8.83 per comment.

**Reaction Count by Verification Status**

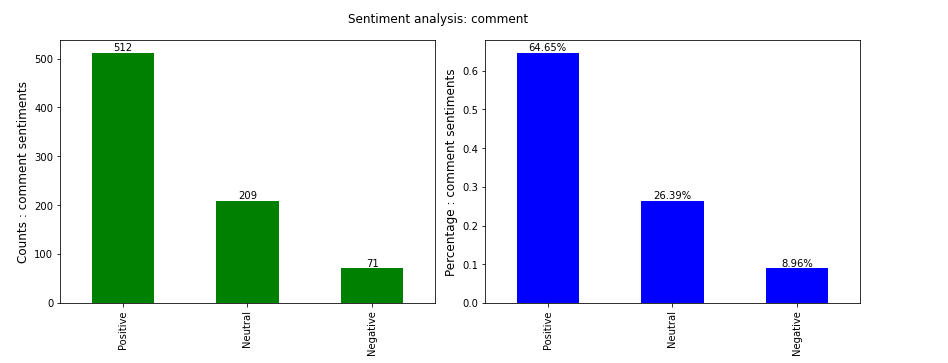


Unverified authors receive an average of 7.04 reactions per comment. Verified authors, similar to the comment count, have an average reaction count of 1. This again may be influenced by the small sample size of verified authors.

**Insightful Questions**

Below are insights obtained from the data to answer the questions of this assignment.

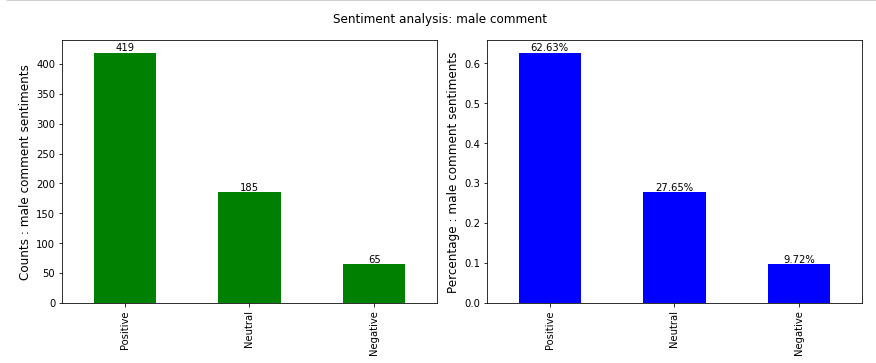
1. **What is the percentage of different discourse expressed by football fans on Facebook**

****

The visual above shows the result of the sentimental analysis on whole dataset with 64.65% of positive comments (512 counts), 26.39% of neutral comments (209 counts), and 8.96% of negative comments (71 counts)

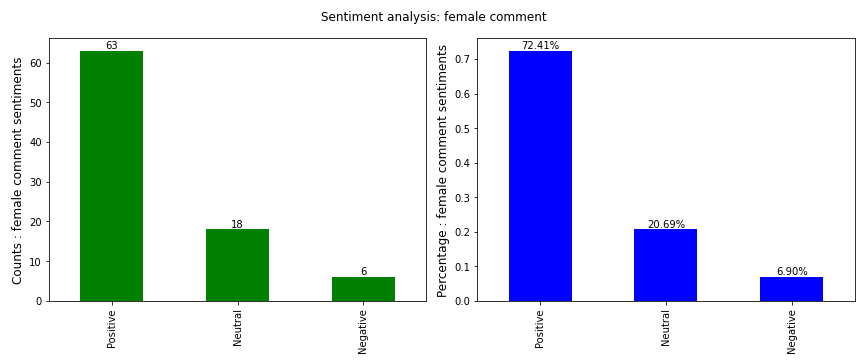
1. **Does Football Fans Gender influence the discourse of comments on Facebook**

**Sentimental Analysis of Comment for Male Gender**



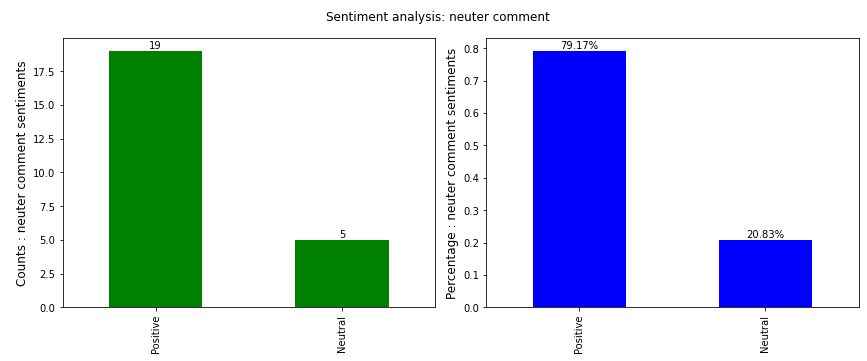
The visual above shows the result of the sentimental analysis on comments for male gender with 62.63% of positive comments, 27.65% of neutral comments, and 9.72% of negative comments

**Sentimental Analysis of Comment for Female Gender**



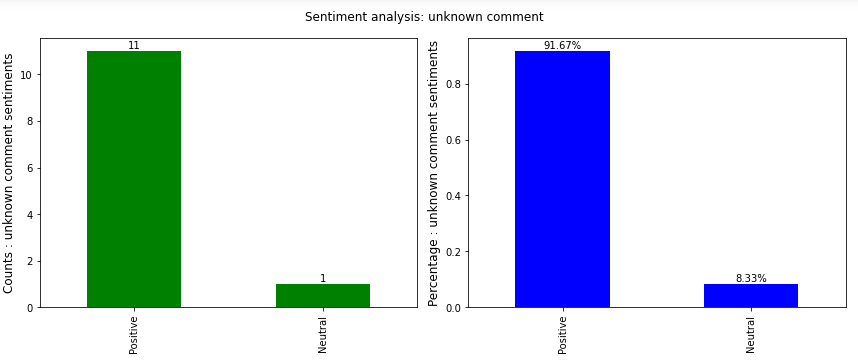
The visual above shows the result of the sentimental analysis on comments for female gender with 72.41% of positive comments, 20.69% of neutral comments, and 6.90% of negative comments

**Sentimental Analysis of Comment for Neuter Gender**



The visual above shows the result of the sentimental analysis on comments for neuter gender with 79.17% of positive comments and 20.83% of neutral comments with no negative comments

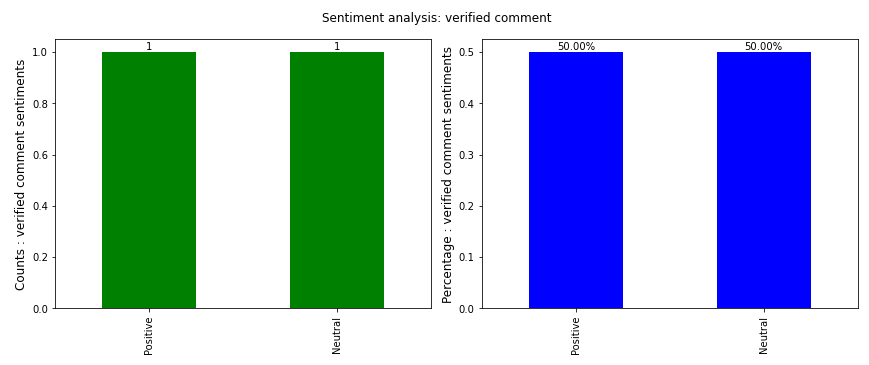
**Sentimental Analysis of Comment for Uknown Gender**



The visual above shows the result of the sentimental analysis on comments for unknown gender with 91.67% of positive comments and 8.33% of neutral comments with no negative comments

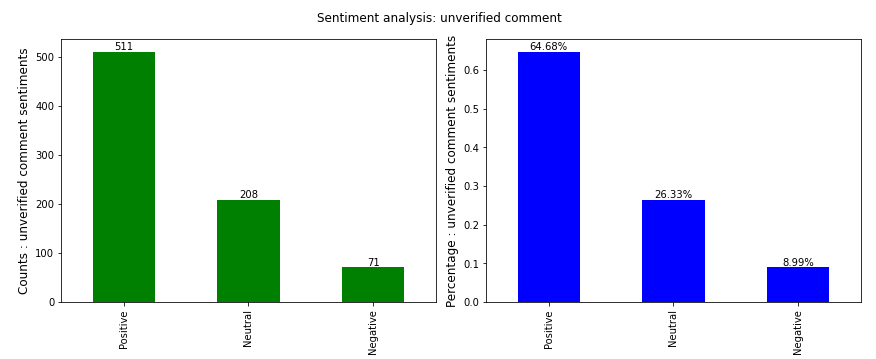
1. **Does Verification Status of Author influence the discourse of comment on Facebook**

**Sentimental Analysis of Comment for Verified Comments**



Since we have only 2 verified user one of it has a positive comment while the other has a negative comment

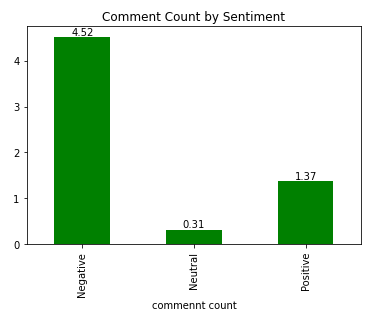
**Sentimental Analysis of Comment for Unverified Comments**



In over 99% of unverified users, the result of the sentimental analysis on its comments is shown above with 64.68% of positive comments, 26.33% of neutral comments and 8.99% of negative comments.

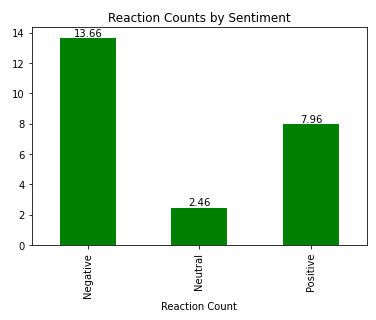
1. **What Demographic Engagement Pattern can be identified with respect to different discourse of comment on Facebook**

**Comment Count by Sentiment**

****

Negative sentiment comments have the highest average comment count (4.52), indicating more engagement for negative sentiments. Positive sentiment comments have the lowest average comment count (1.37), suggesting comparatively less engagement for positive sentiments. Neutral sentiment comments fall in between, with an average comment count of 0.31.

**Reaction Counts by Sentiment**



Negative sentiment comments have the highest average reaction count (13.66), indicating a higher level of reaction (likes, etc.) for negative sentiments. Positive sentiment comments have a lower average reaction count (7.96), suggesting relatively fewer reactions for positive sentiments. Neutral sentiment comments fall in between, with an average reaction count of 2.46.

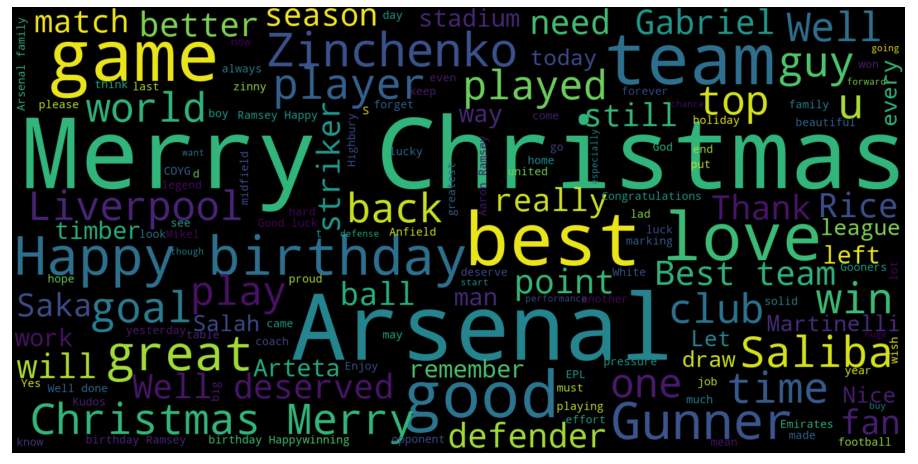
1. **What are the Prevalent words expressed in comments by football fans on Facebook**

**Prevalent Words in Comments**



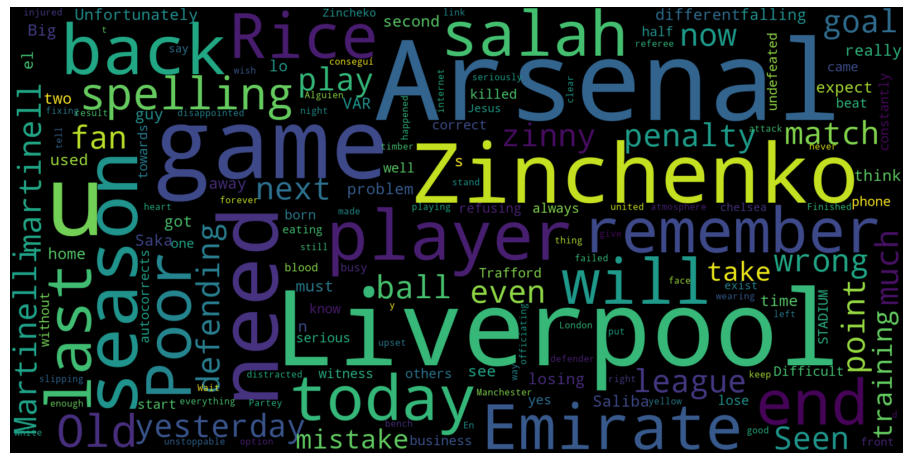
1. **What are the common words expressed in positive comments by football fans**

**Prevalent words in comments (Positive Sentiments)**

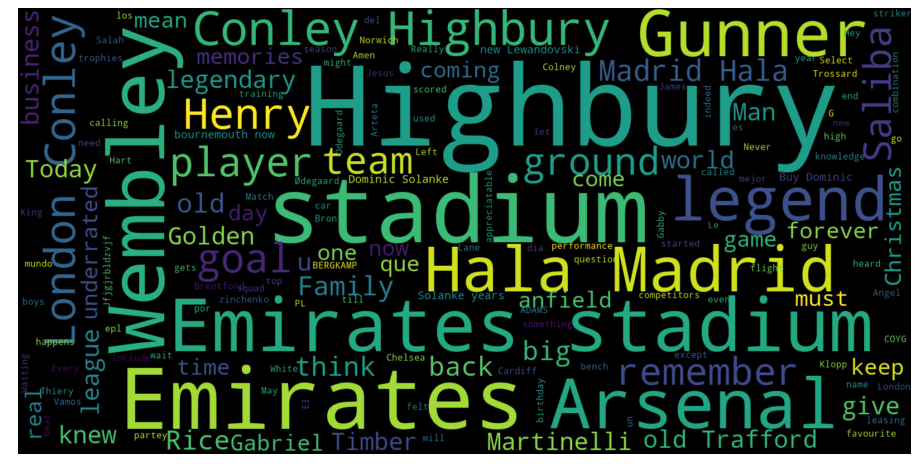


1. **What are the common words in hate/negative discourse expressed by football fans against soccer players on Facebook?**

**Prevalent words in comments (Negative Sentiments)**

****

**Prevalent words in comments (Neutral Sentiments)**



The above Word clouds visually offered a representation of prevalent words in comments categorized by sentiment.

**Key Insights**

**Account Verification:** Over 99% of users have unverified accounts, emphasizing the prevalence of non-verified profiles on social media.

**Gender Distribution:** Male users dominate the comment section, followed by females, while neuter and unknown gender users contribute less.

**Verification Status:** The majority of authors, regardless of gender, are not verified. Analysis suggests that verification status does not strongly correlate with sentiment expression.

**Sentiment Analysis:** Approximately 64.65% of comments are positive, 26.39% are neutral, and 8.96% are negative. Male and female users express similar sentiment distributions, with neuter and unknown genders predominantly expressing positive sentiments.

**Verified Authors:** The limited number of verified users includes both positive and negative sentiments. Verified authors have a comment count of 0 and a reaction count of 1, indicating a notable difference in engagement compared to unverified authors.

**Demographic Engagement Patterns**: Female authors receive an average of 4.41 reactions per comment, while male authors have a higher engagement with approximately 7.30 reactions. Neuter and unknown gender authors receive notably high average reactions of 8 and 8.83 per comment, respectively.

**Unknown Gender Category:** The unknown gender category predominantly consists of positive comments, warranting further investigation into this group's characteristics and engagement patterns.

**Sentiment and Engagement:** Negative sentiment comments attract the highest average comment count (4.52) and reaction count (13.66), while positive sentiment comments have lower averages (1.37 for comments, 7.96 for reactions), indicating strong engagement for negative sentiments.

**Word Cloud Analysis:** Word cloud visualizations offer insights into prevalent words in comments, categorized by sentiment, guiding content creators to understand fan language.

**SUMMARY**

* Gender and verification status play a role in shaping user engagement metrics.
* Verified authors exhibit a distinct engagement pattern with a comment count of 0, raising questions about the nature of their social media interactions.
* The unknown gender category's predominantly positive comments warrant further investigation to uncover unique characteristics and engagement patterns within this group.
* Negative sentiment comments attract higher engagement, both in terms of comment count and reaction count. Positive sentiment comments, while prevalent, exhibit relatively lower engagement.
* This analysis provides a foundation for understanding football fan engagement. .

**Conclusions**

In summary, the analysis of football fan engagement on Facebook highlights significant patterns related to sentiments, user demographics, and verification status. With over 99% of users having unverified accounts, a gender-based disparity in participation is evident, with male users dominating the comment section. Despite the prevalence of unverified accounts, the study suggests that verification status does not strongly correlate with sentiment expression.

Sentiment analysis reveals that around 64.65% of comments are positive, 26.39% are neutral, and 8.96% are negative. Male and female users express similar sentiment distributions, while neuter and unknown genders predominantly express positive sentiments.

The unknown gender category stands out for predominantly positive comments, warranting further investigation. Negative sentiment comments attract higher engagement, suggesting a need to address negative sentiments and foster positive community engagement in social media strategies.

The study lays a foundation for understanding football fan engagement, emphasizing the importance of recognizing diverse user segments. Continuous monitoring and additional analyses will contribute to a more nuanced understanding of fan behavior over time.