**Social Media and Network Analysis**

**COSC 2671**

**Assignment 2**

**From Goals to Global Fame: How Social Media talks about**

**‘David Beckham’**

**Submitted By – Group 7**

**Preetpal Singh (s4062787)**

**Saurabh Sharma (s4038560)**

**Shivam Shinde (s3994666)**

**Index**

**Pg No:**

1. Introduction
2. Data Description
3. Methodology

* Data Pre-Processing and Exploration
* Word Clouds
* TF-IDF Vectorization

1. Sentiment Analysis (VADER)
2. **Introduction**

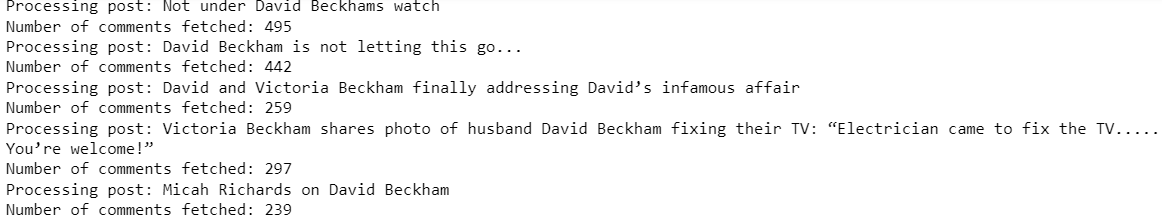
**David Beckham** is not just a footballer, he’s a global icon who presence extends far beyond the pitch. From leading England’s national team to his ventures in fashion and philanthropy, Beckham’s life continues to inspire and intrigue people worldwide. But how do fans and public talk about him today? By analyzing discussions on two major social media platforms – **Reddit and YouTube** – we explore not only what people say about Beckham but also what these conversations reveal about his impact on culture, sports and beyond.

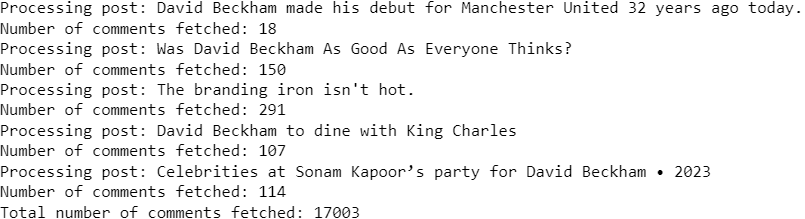
1. **Data Description**

Imagine two arenas of conversation: Reddit, where debates rage about everything from sports statistics to celebrity gossip, and Youtube, where fans relive Beckham’s iconic moments through highlight reels and documentaries. We tapped into these platforms to gather thousands of user comments related to David Beckham. Reddit gave us the inside track of fan debates, while Youtube provided raw, emotional responses to Beckham’s on-screen appearances and football highlights.

Python Reddit API Wrapper **(PRAW)** was used which allowed the extraction of relevant posts and comments. **100** posts were extracted containing **17,003** comments over a period of one yearto get a broader view of public sentiment. **YouTube Data API v3** was used to extract **30 videos** containing a total of **83,825** comments.

David Beckham’s influence extends far beyond the football field, with discussions across both YouTube and Reddit reflecting his multifaceted legacy. On YouTube, fans celebrate his iconic moments, particularly his legendary free kicks and leadership on the pitch, while his media appearances, playful pranks and interviews showcase his charm and sense of humour. His personal life draws admiration, with many praising his supportive relationship with his wife, Victoria (Formerly member of Iconic Pop band group **Spice Girls**) and his role as a family man. On Reddit, conversations delve deeper into his family dynamics, often focusing on humorous and touching moments, such as Beckham’s 25th wedding anniversary and his wife Victoria’s comments about her background. His **football legacy** remains a prominent topic on both the social media platforms, with fans reminiscing about his free kicks and influence on the sport, and comparing him to contemporary players like **Messi** and **Neymar.** Additionally, Reddit comments highlight Beckham’s continued relevance in pop culture and media, with users reflecting on his **Netfilx documentary “Beckham”**, endorsements and legal disputes. **Overall**, comments across both platforms reveal a deep admiration for **Beckham as a football legend, a global celebrity and a cultural icon.**

**Fig. 2.1**, shows some sample discussion titles about Beckham as extracted from Reddit.



**Fig 2.1 – Some Fetched Post Titles from Reddit (David Beckham)**

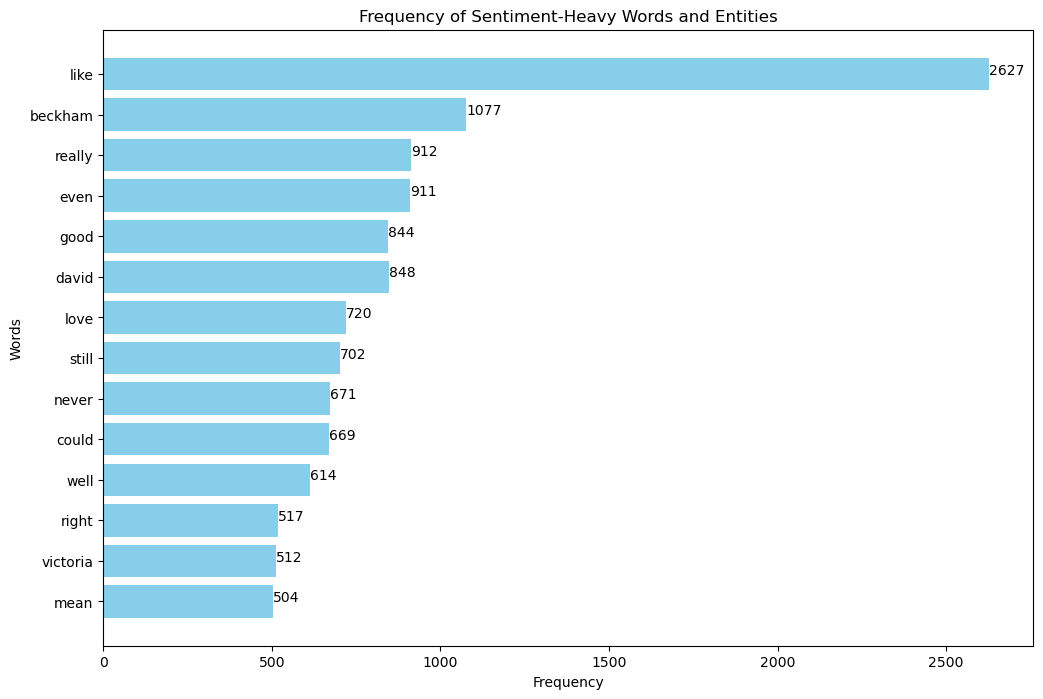
These comments were subjected to text analysis techniques to extract meaningful insights regarding trending topics and public sentiments.

1. **Methodology**

**Data Pre-Processing and Exploration**

The data collected from Reddit and YouTube underwent several preprocessing steps to ensure it was suitable for analysis. These steps included: -

* **Tokenization** – Breaking down text into individual words, allowing for detailed analysis of word frequency and usage.
* **Stopword removal** – Removing common words that do not contribute significantly to the analysis (eg “and, “the”, etc.)
* **Lemmatization** – Converting words to their base forms to standardize the text and reduce dimensionality.
* **Outlier Detection** – Outliers were retained, contributing to rich content representation, balanced sentiment reflection, understanding polarization, strategic insights and enhanced decision making

**Fig 3.1 – Bar Chart - Frequency of Sentiment-Heavy Words and Entities (Reddit)**



**Fig 3.2 – Word Cloud of Sentiment-Heavy Words and Entities (Reddit)**

The bar chart and word cloud together of **Reddit data** provide a comprehensive view of the **sentiment-heavy words** frequently used in the discussions about David Beckham, **focusing on words that carry significant emotional weight**. This approach is particularly effective for sentiment analysis because: -

* **Relevance to Sentiment (Capturing Emotional Intensity)** – Visualizations highlight the emotional tone and opinions expressed, showcasing words like **“like” (2627), “good” (844), “love” (720)** dominates, showing strong admiration for his career, personality and family life.

* **Actionable Insights for Strategic Decisions**

1. **Leverage Positive Sentiment** – The overall positivity can be harnessed for branding opportunities, endorsements and partnerships, particularly with products that align with Beckham’s values and family-oriented image.
2. **Maintain Relevance** – Words like **“still”** and **“even”** indicate Beckham’s ongoing relevance. Strategic efforts would focus on keeping him visible through media and public appearances.
3. **Family Image** – The focus on his wife “Victoria” suggests joined branding opportunities featuring the Beckham family could resonate well with the audiences.
4. **Address Criticism** – Although minimal, potential criticism can be addressed through proactive PR strategies that reinforce positive aspects of Beckham’s legacy.

**By focusing on sentiment heavy words rather than just unique words**, both the visualizations provide a richer understanding of public sentiments surrounding David Beckham. Key topics of discussion are revealed along with emotions that are driving those discussions. This approach **ensures that analysis is both relevant and practical, offering stakeholders clear insights into public opinions and guiding effective strategic communication efforts.**

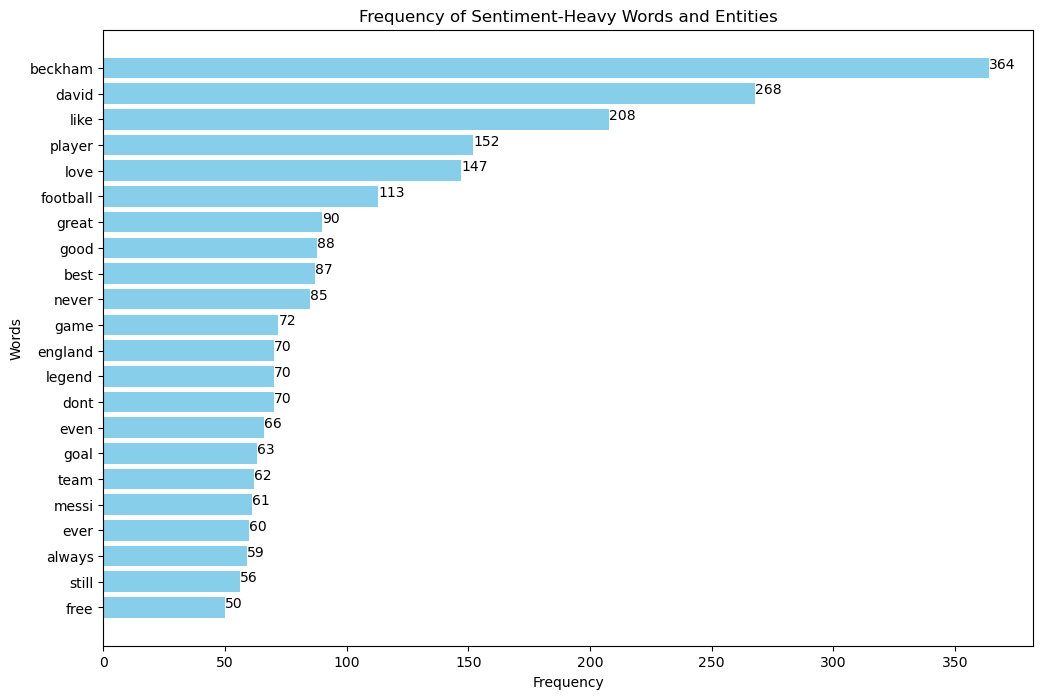
**Alternative Perspective for Stakeholders – Unique Words**

While sentiment heavy words provide critical insights into the emotional tone and intensity of discussions, **identifying unique words offers a complementary view that highlights specific entities, topics or concepts being discussed.**

**Unique words – based on context of discussions and not on frequency**

* **Victoria** (Specific Person) - Frequent mentions of Victoria, highlight the public’s interest in his relationship with his wife, adding to his image as a family man.
* **David** (Specific Person) – Represents a more personal reference to Beckham himself, indicating intimate or individual focused conversations.
* **Right** (Affirming statements or comparison) – Used in debate or affirmations about Beckham’s abilities or decisions.
* **Well** (Performance or outcome) – Reflects commentary on how Beckham has performed or handled situations.
* **Could** (Speculative discussions) – Suggests hypothetical discussions about Beckham’s career, life choices, or potential future roles.
* **Really** (Emphasis or intensity) – Used to amplify or emphasize sentiments often reinforcing admiration or posing questions about his legacy.
* **Mean** (Understanding context) – Suggests where people seek clarity or questions aspects of Beckham’s life.

This approach provides a holistic understanding of both **what is being talked about (such as Beckham’s personal and professional life) and how it is being talked about (through speculative, affirming or critical discussions)** on **Reddit** enabling more strategic and targeted actions in response to public disclosure.

**Fig 3.3 – Bar Chart - Frequency of Sentiment-Heavy Words and Entities (YouTube)**

**Fig 3.4 – Word Cloud of Sentiment-Heavy Words and Entities (YouTube)**



The bar chart and word cloud together of **YouTube data** provide a comprehensive view of the **sentiment-heavy words** frequently used in the discussions about David Beckham, **focusing on words that carry significant emotional weight**. This approach is particularly effective for sentiment analysis because: -

* **Relevance to Sentiment (Capturing Emotional Intensity)** – Visualizations highlight the emotional tone and opinions expressed, showcasing words like **“beckham” (364), “david” (268), “like” (208), “player” (152), “love” (147)** dominates, showing strong admiration for his football career, personality and legacy. Words like **“great” (90), “good” (88) and “best” (87),** further reinforce the positive sentiment surrounding his accomplishments as a footballer.

* **Actionable Insights for Strategic Decisions**

1. **Leverage Positive Sentiment** – The overall positivity can be harnessed for branding opportunities, endorsements and partnerships, particularly with products that align with Beckham’s image as a football legend and sports icon.
2. **Maintain Relevance** – Words like **“still”** and **“ever”** indicate Beckham’s ongoing relevance in football. Strategic efforts would focus on keeping him visible through media and public appearances.
3. **Football Legacy** – The focus on terms like **“football” (113), “goal” (63), “team” (62) and “game” (72)** suggests endorsements or partnerships with sports brands would strongly resonate with the audiences on YouTube platform.
4. **Address Criticism** – Although minimal, words like **“never” (85) and “dont” (70),** suggest some level of criticism, likely comparisons with other footballers. These can be addressed through proactive PR strategies that reinforce positive aspects of Beckham’s legacy and achievements.

**By focusing on sentiment heavy words rather than just unique words**, both the visualizations provide a richer understanding of public sentiments surrounding David Beckham. Key topics of discussion are revealed along with emotions that are driving those discussions. This approach **ensures that analysis is both relevant and practical, offering stakeholders clear insights into public opinions and guiding effective strategic communication efforts.**

**Alternative Perspective for Stakeholders – Unique Words**

While sentiment heavy words provide critical insights into the emotional tone and intensity of discussions, **identifying unique words offers a complementary view that highlights specific entities, topics or concepts being discussed.**

**Unique words – based on context of discussions and not on frequency**

* **Football** (Specific Sport) - Frequent mentions of **“football”**, highlight the public’s interest in Beckham’s legacy as a legendary football player.
* **Player** (Football Role) – Represents Beckham’s professional role, reinforcing admiration for his on-filed performance and legendary status.
* **Goal** (Achievements) – Words like **“goal”** emphasize Beckham’s contributions and memorable moments in football, showcasing his excellence in the sport.
* **England** (National Identity) – Discussions around **“England”** refer Beckham’s role as an English football star, often linking his career to national pride.
* **Team** (Collaborative Role) – The word **“team”** reflects Beckham’s association with multiple successful football clubs and highlights his role in a team context.
* **Messi** (Comparison) – Mention of **“Messi”** suggests comparison between Beckham and other iconic players which can represent opportunities to discuss Beckham’s unique legacy.
* **Legens** (Status) – Beckham’s status as a **“legend”** reflects his enduring influence in football and pop culture, highlighting his global impact.

This approach provides a holistic understanding of both **what is being talked about (such as Beckham’s professional life as a footballer) and how it is being talked about (through discussions of his role as a player, his legendary status and comparison with other players)** on **YouTube** enabling more strategic and targeted actions in response to public disclosure.

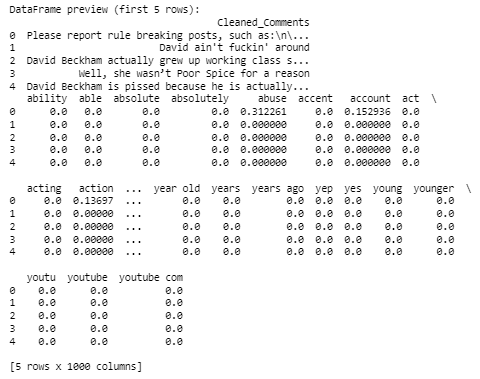
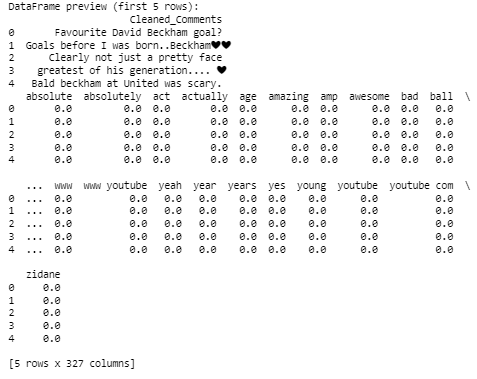
**Overall Conclusion of Sentiments on Reddit and YouTube**

On **Reddit**, Beckham’s **family image** and **celebrity status** are discussed alongside his career, reflecting a **more diverse admiration** for him as a public figure.

On **YouTube**, the focus is primarily on Beckham’s **football career**, with a strong emphasis on his **athletic accomplishments** and **legendary status** in the sport.

Both platforms show **positive sentiments** but then nature of discussions reflect the **specific interests** of each audience: personal life on Reddit and football achievements on YouTube.

**TF-IDF Vectorization**

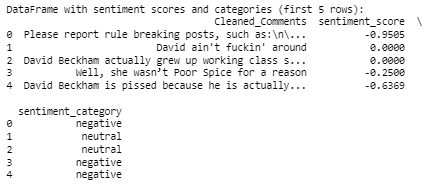
****The TF-IDF vectorization process transforms the textual data into a numerical representation that highlights the importance of specific words in a comment, providing stakeholders with a focused view of the most relevant terms in discussion about David Beckham. TF-IDF assigns a score to each word, balancing how often it appears in a single comment (term frequency) and how rare it is across all the comments (inverse document frequency). Words that occur frequently in a specific comment but are rare across the dataset receive higher scores. By analyzing the TF-IDF scores, stakeholders can prioritize content, tailor communication strategies and gain deeper insights into the specific subjects that resonate within the conversations.

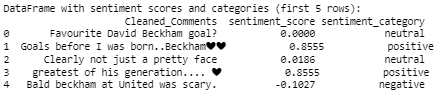
**Table 3.5 – TF-IDF Vectors – Dataframe (Head) (Reddit) Table 3.6 – TF-IDF Vectors – Dataframe (Head) (YouTube)**

**Reddit** has a broader range of discussions (1000 columns), including both personal life and casual comments

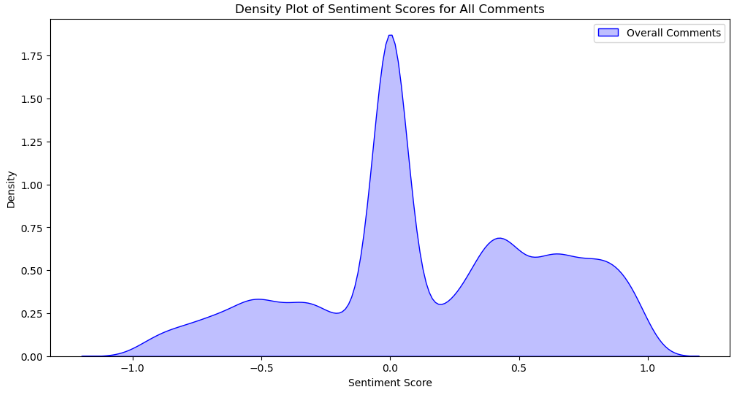
**YouTube** has a more focused set of terms (327 columns), concentrating on Beckham’s football legacy, goals and achievements.

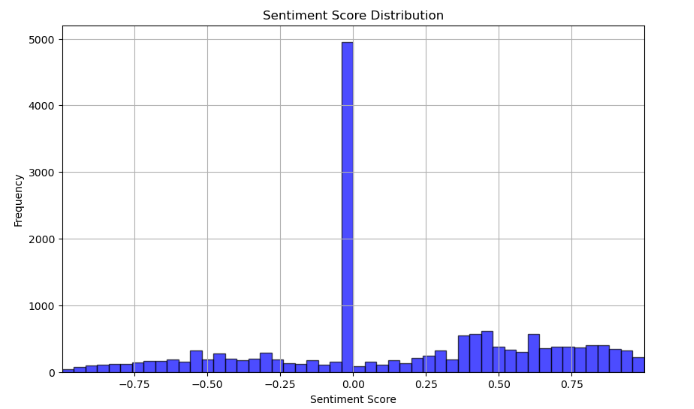
1. **Sentiment Analysis (VADER)**

****The VADER sentiment analyzer was used to assess the sentiment of each comment. VADER is particularly suited for social media text as it accounts for the context of the words, including slang and emoticons, providing a compound score that reflects the overall sentiment.

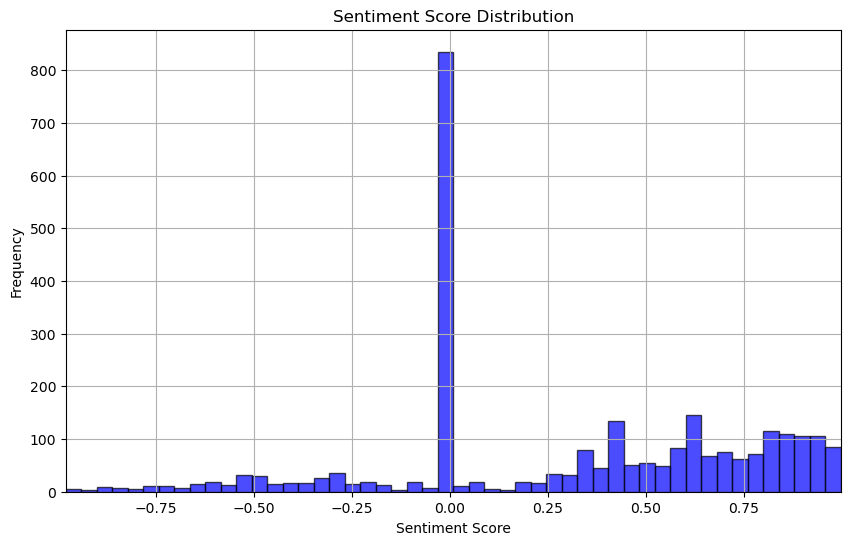
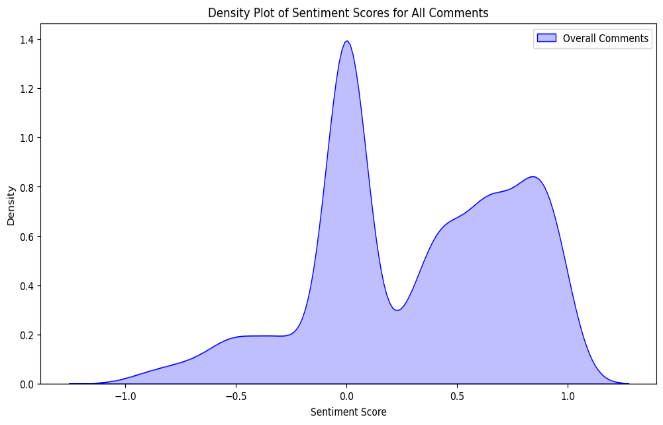
****

**Table 4.1 – VADER Analysis – Dataframe (Head) (Reddit) Table 4.2 – VADER Analysis– Dataframe (Head) (YouTube)**

****Each comment has been assigned a sentiment score ranging from -1 to 1, where negative value indicates a negative sentiment, positive value indicates positive sentiment and scores close to zero (0) indicates neutral sentiment. These scores are then categorized into **three sentiment categories of negative, positive and neutral.** This gives a clear understanding of how different comments reflect varying degrees of sentiment towards Elon Musk. **This categorization is crucial for stakeholders to identify areas of concern, positivity or neutrality in public discussions.**

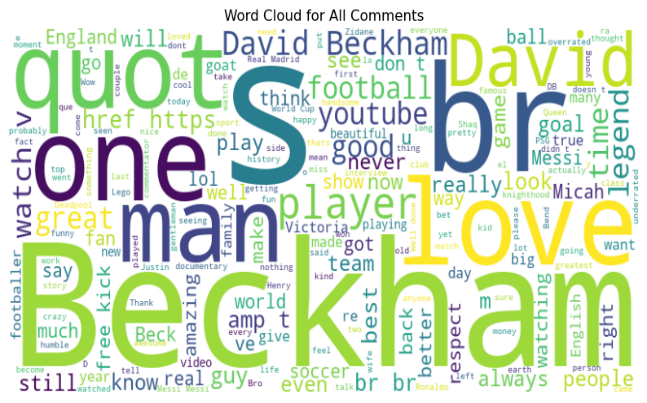
****

**Fig.4.1–Histogram-Sentiment Score-All Comments (Reddit) Fig.4.2–Density Plot-Sentiment Score-All Comments (Reddit)**



**Fig.4.3–Histogram-Sentiment Score-All Comments (YouTube) Fig.4.4–Density Plot-Sentiment Score-All Comments (YouTube)**

On **Reddit**, the discussions are broader and more varied, with a notable presence of both **neutral and positive sentiment** but also a higher incidence of **negative sentiment** compared to **YouTube**. The diversity likely reflects Reddit’s tendency for more debates and discussions that cover Beckham’s **personal life, public image and career highlights**. Meanwhile, **YouTube** shows a more focused discussion around Beckham’s **football legacy** with a **clear positive sentiment bias**. The majority of YouTube comments are neutral or positive, with a relatively little negative sentiment, reflecting admiration for his **career achievements** and iconic moments on the field. **Overall**, YouTube portrays Beckham as a revered football figure, while Reddit showcases a more subtle conversation that blends **praise with personal critique**. This comparative sentiment analysis suggests that **Beckham’s football career** resonates more positively on YouTube whereas, Reddit users engage in **more balanced discussions** involving both his **personal and professional life**.



**Fig 4.5 – Word Cloud of All Comments (Reddit) Fig 4.6 – Word Cloud of All Comments (YouTube)**

**On Reddit**, conversations are more **opinion-driven,** with prominent words like **“people”, “think” and “know”,** indicating a focus on sharing thoughts and reflecting on public perceptions. Reddit users also delve into Beckham’s personal life, with mentions of his wife **Victoria** and **relationship**. In contrast, YouTube discussions focus heavily on Beckham’s football career, with words like **“love”, “football”, “player” and “legend”** dominating. YouTube users admire Beckham’s achievements, highlighting his football legacy and individual accomplishments. **Overall**, Reddit offers more personal discussions, wile YouTube centers on Beckham’s **sports legacy** and **public admiration**.

Above approach gives stakeholders a comprehensive understanding of both the content and tone of the discussions, enabling more informed decision making and strategic communication.