**Skill vs. Appearance: Analyzing Gender Bias in Fan Commentary in NCAA Basketball**

**Jacob Bronkala**

## **EXECUTIVE SUMMARY**

This study explores how gender influences public discourse surrounding top NCAA basketball players during the 2025 March Madness tournament by analyzing social media posts, specifically tweets. The primary objective was to investigate whether fans describe male and female athletes differently, particularly in terms of their athletic skill versus physical appearance.

Using text mining techniques, including sentiment analysis and topic modeling, the research analyzed 1,000 tweets referencing the top 10 male and 10 female players, capturing a range of opinions on platforms like Twitter. Results revealed that while the sentiment toward male and female athletes was statistically similar, notable differences emerged in the topics discussed. Female athletes were more often associated with appearance-related comments, while male athletes were more frequently discussed in terms of performance and age.

Despite these topic differences, both genders were frequently mentioned in the context of the sport itself, with terms like "game," "season," and "college basketball" appearing regularly in both male and female athletes’ tweets. This indicates a shared focus on the sport, with some remaining gendered narratives regarding appearance and athleticism.

The findings suggest that, while gendered patterns still exist in the types of traits discussed, the overall sentiment and tone of the discourse about male and female athletes show greater equity than might be expected. The growing focus on female athletes' performance in high-visibility tournaments like March Madness reflects an important shift in public perception. However, disparities in how athletes are portrayed across genders persist, especially in appearance-related comments.

This study contributes to the ongoing conversation about gender equality in sports, offering insights into how fans' discussions on social media could shape and reinforce societal views about male and female athletes. Future research could expand this analysis across different sports or media platforms to further explore these dynamics.

## **INTRODUCTION**

Gender bias in sports media and public discourse has long been a topic of concern, particularly when it comes to how female athletes are perceived and described in comparison to their male counterparts. This debate has only increased in the public eye with the emergence of superstar female athletes such as Caitlin Clark, Paige Bueckers, and Juju Watkins. While elite women’s sports have seen rising visibility, particularly during March Madness, fan commentary on social media often reveals lingering stereotypes and imbalances in how athletes are discussed online.

While growing attention to gender disparities in sports coverage has sparked meaningful discussion, much of what we know is based on individual observations or media examples rather than large-scale analysis. This project aims to build on that work by using text mining to explore how fans discuss NCAA men’s and women’s basketball players during the 2025 March Madness tournament. By analyzing fan-generated comments from platforms like Twitter and YouTube, I want to uncover whether fans tend to emphasize different traits, such as athletic skill versus physical appearance based on player gender.

My central research question is: *Do fans describe men’s and women’s college basketball players differently in terms of skill and appearance?*Based on anecdotal evidence, my hypothesis is that *female athletes are more likely to be commented on for their appearance, while male athletes are more likely to be discussed in terms of their athletic skill.*

Using NLP tools like sentiment analysis, topic modeling, and named entity recognition, I aim to quantify these differences and test their statistical significance. This analysis offers a small, data-driven perspective that adds to the broader conversation about gender equality in sports, particularly in how athletes are discussed by fans.

## **DATA DESCRIPTION**

The dataset consists of 1,000 tweets collected between March 18 and April 2, 2025, during the NCAA March Madness tournament. These include 500 tweets referencing the top 10 men’s basketball players and 500 tweets referencing the top 10 women’s basketball players. Tweets were scraped using player name keywords and filtered to include only original, English-language posts. Retweets, duplicates, and promotional content were excluded to preserve authenticity.

Each tweet is associated with a specific athlete, and a separate column includes each athlete’s gender (Male or Female), and playing position—Guard (G), Forward (F), or Center (C). While the tweets themselves do not mention position, this variable enables further analysis by role type when aggregating at the player level.

Although the goal was to collect a larger sample of tweets per athlete, tweet availability was limited in some cases. Many social media posts during the tournament referenced teams rather than individual players, or used pronouns and nicknames instead of full player names, making them difficult to attribute accurately. As a result, each player is represented by 50 tweets, depending on availability.

**METHODS**This study employed text mining techniques to analyze and compare the social media presence of top male and female college basketball players. The analysis focused on sentiment, topical frequency, and language patterns in tweets mentioning each athlete.

#### ***Data Preprocessing***

Text data underwent standard natural language processing (NLP) preprocessing, including:

* Lowercasing all text
* Removal of URLs, hashtags, mentions, emojis, and non-alphabetic characters
* Tokenization and lemmatization using the spaCy library
* Stopword removal

These steps standardized the text and prepared it for downstream analysis.

#### ***Feature Engineering***

To extract meaningful insights, several features were engineered from the cleaned text:

* Sentiment Analysis: VADER sentiment analyzer was used to compute compound polarity scores for each tweet, ranging from –1 (most negative) to +1 (most positive).
* Topic Tagging: Tweets were scanned for the presence of curated keyword lists corresponding to themes such as *performance*, *appearance*, and *age*..

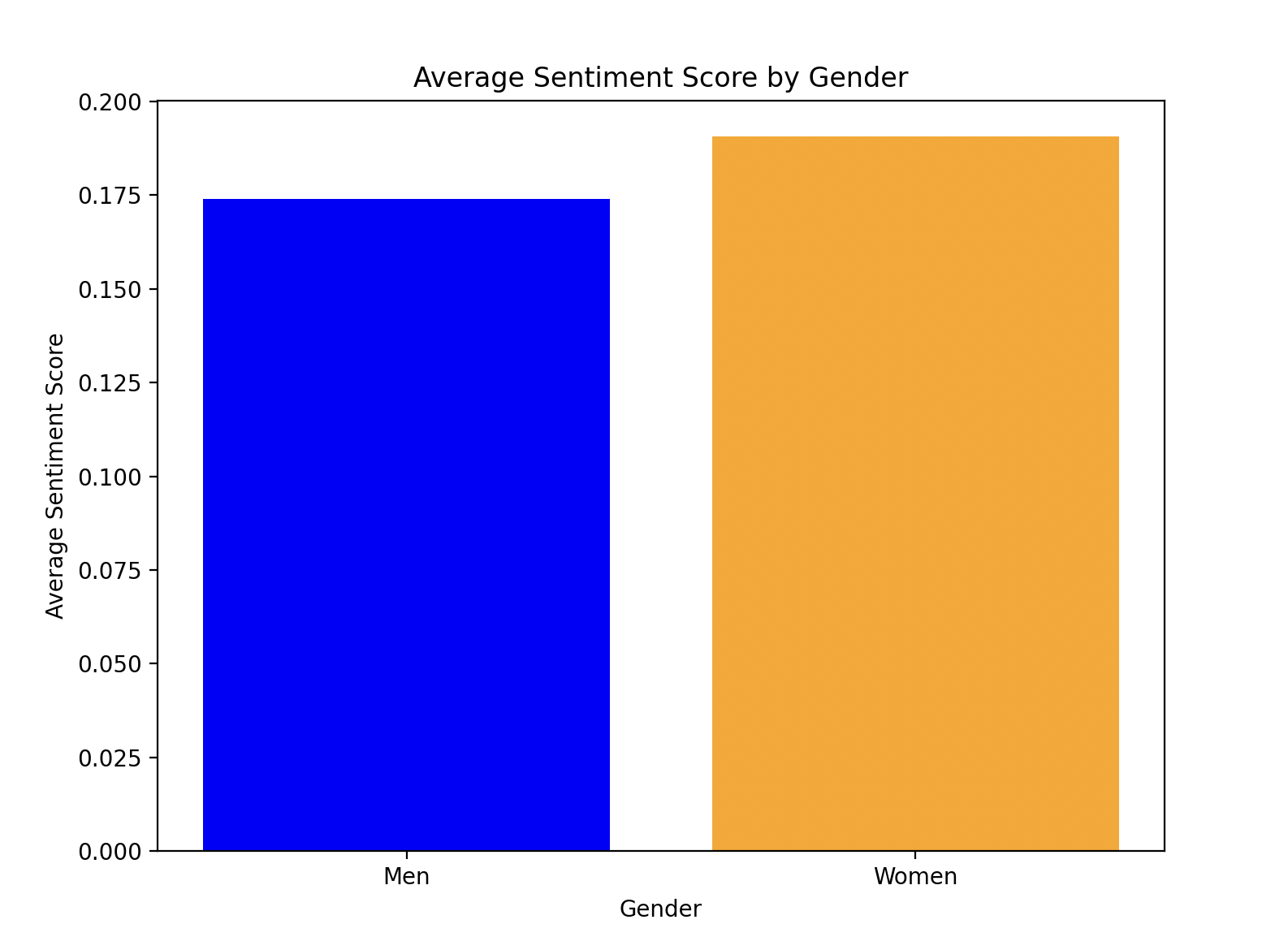
#### ***Analysis***

Descriptive statistics and visualizations (e.g., bar plots, box plots) were used to compare sentiment distributions and keyword prevalence across gender. Welch’s t-tests were conducted to assess statistically significant differences in mean sentiment scores between male and female athletes. Additional qualitative inspection of outlier tweets (particularly those with extreme sentiment scores) was performed to contextualize quantitative findings.

**RESULTS**The analysis revealed consistent patterns in how male and female athletes were discussed, with statistically and thematically distinct trends emerging across sentiment and topic focus.

#### ***Sentiment Analysis***

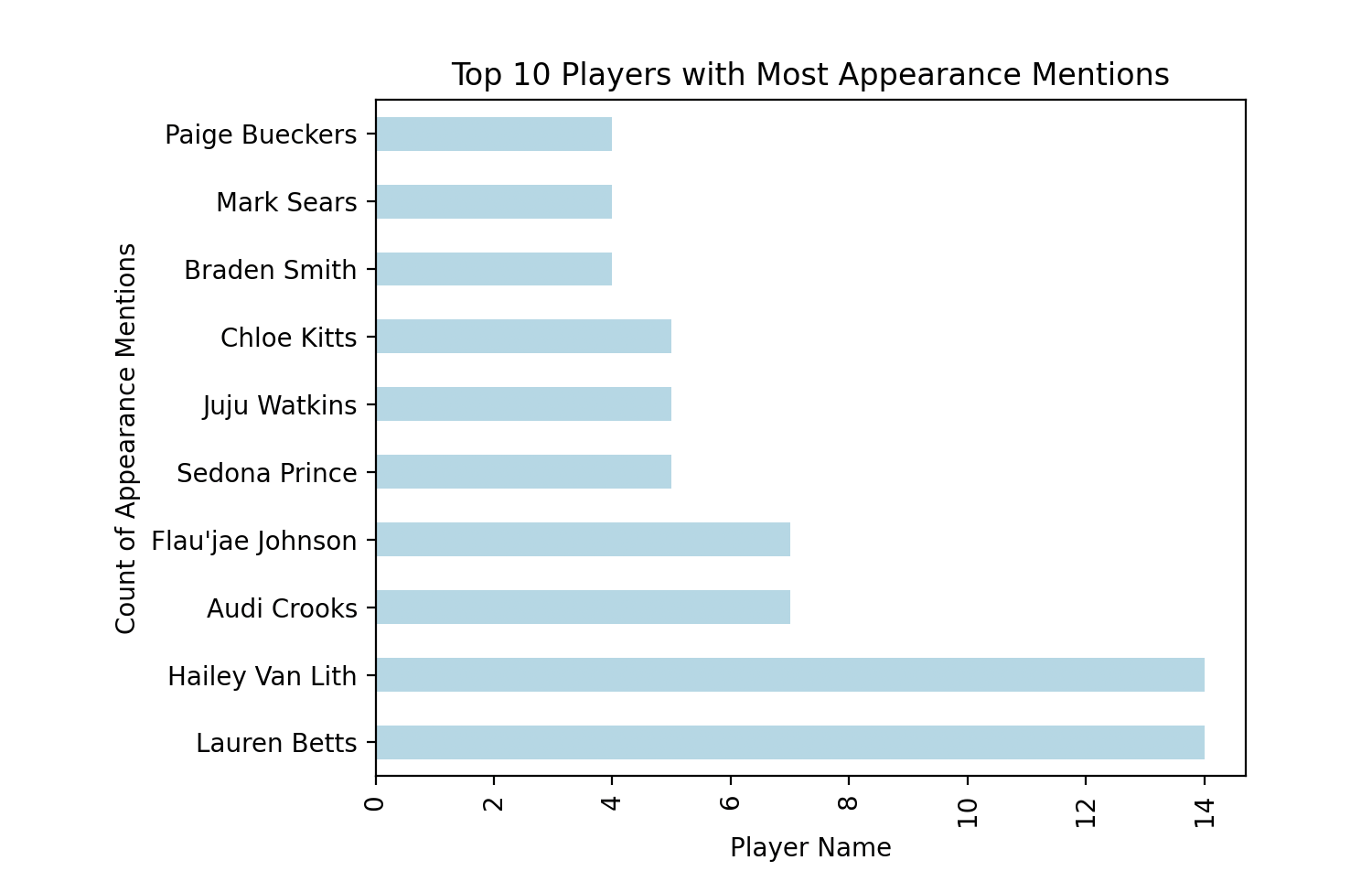
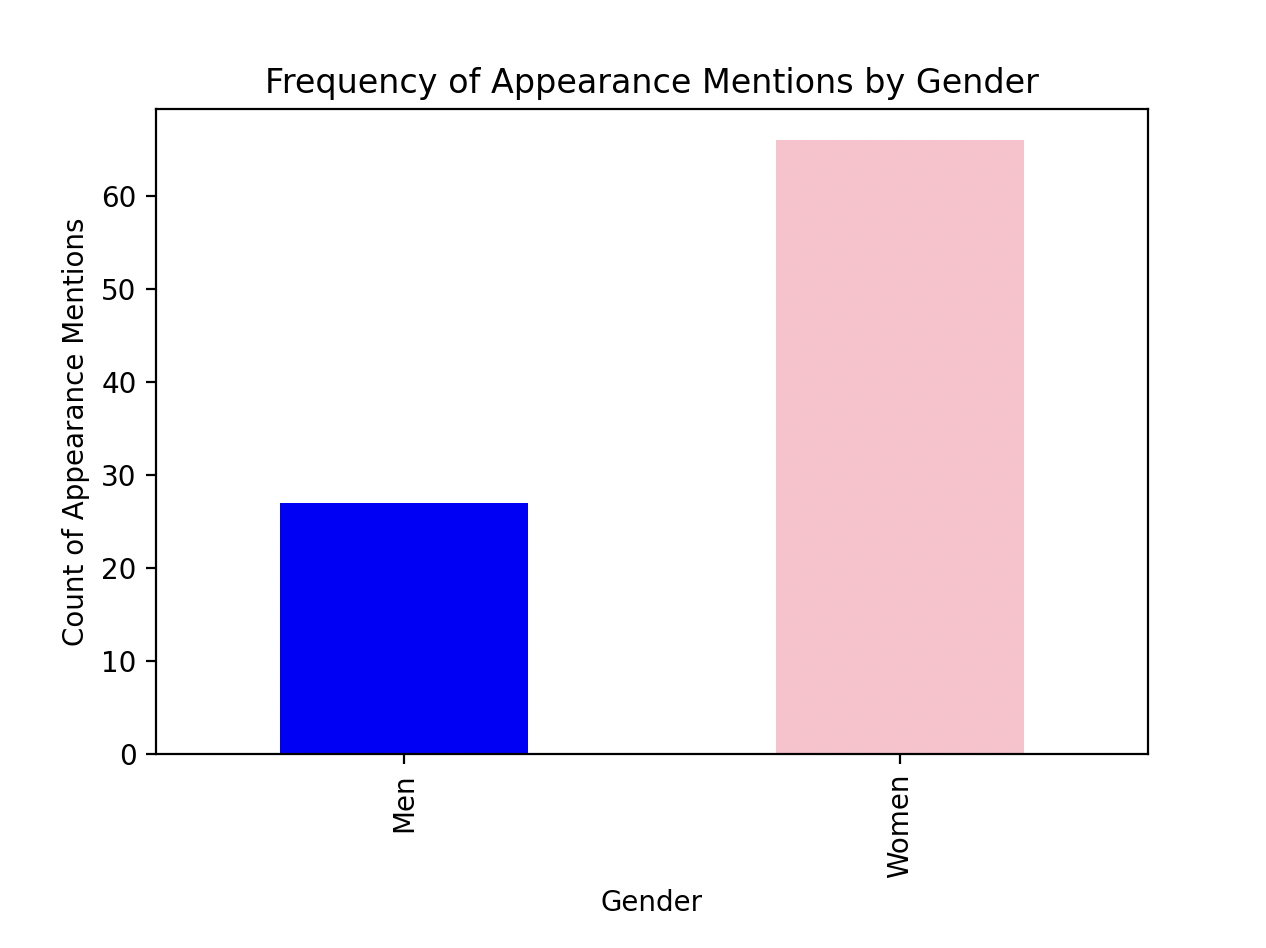
As shown in Figure 1, the distribution of VADER sentiment scores was positively skewed for both genders, but female athletes received a slightly higher mean sentiment (F = 0.190701) than male athletes (M = 0.173911). A Welch’s *t*-test confirmed that there was no statistically significant difference in these scores (*p* = 0.4479), meaning we cannot say either gender’s tweets were more positive.

****

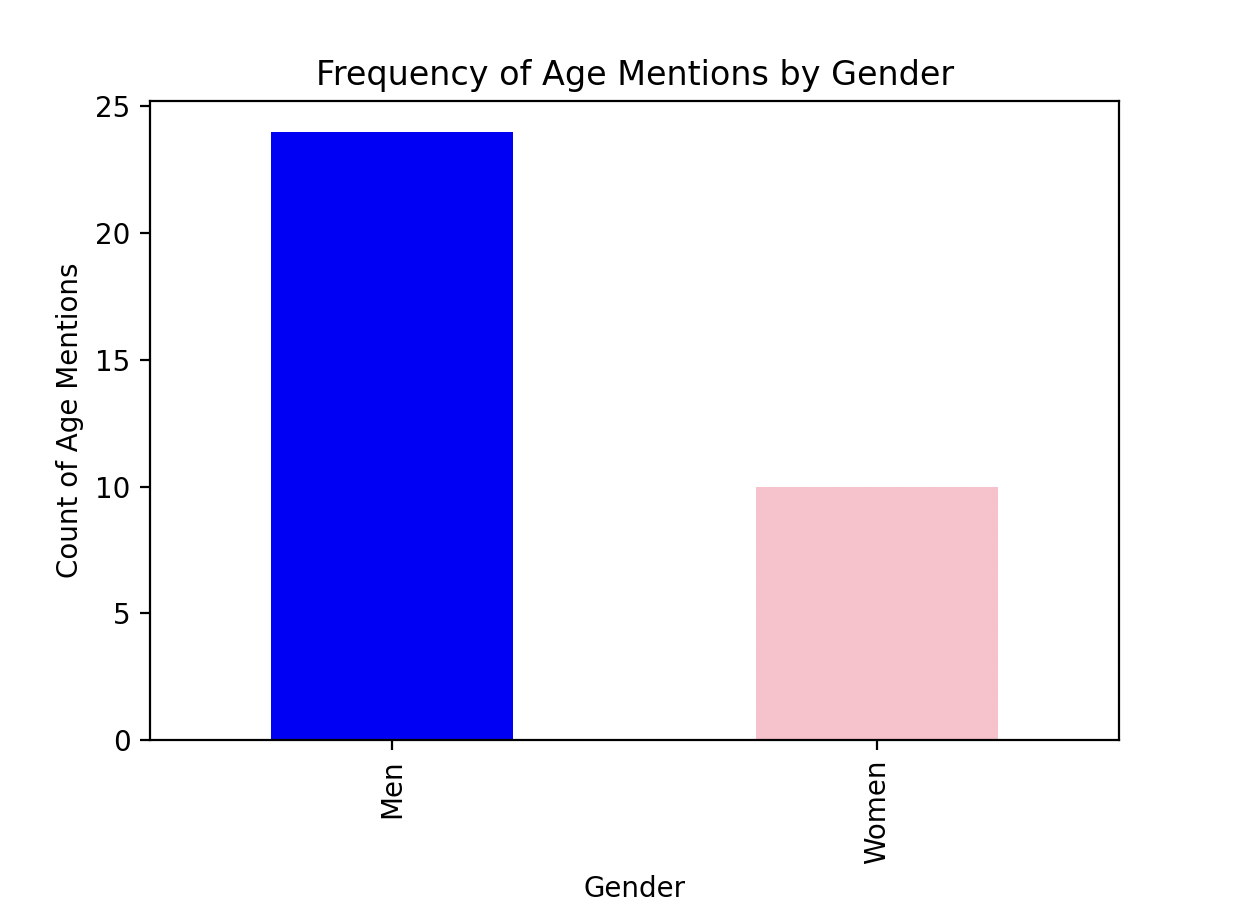
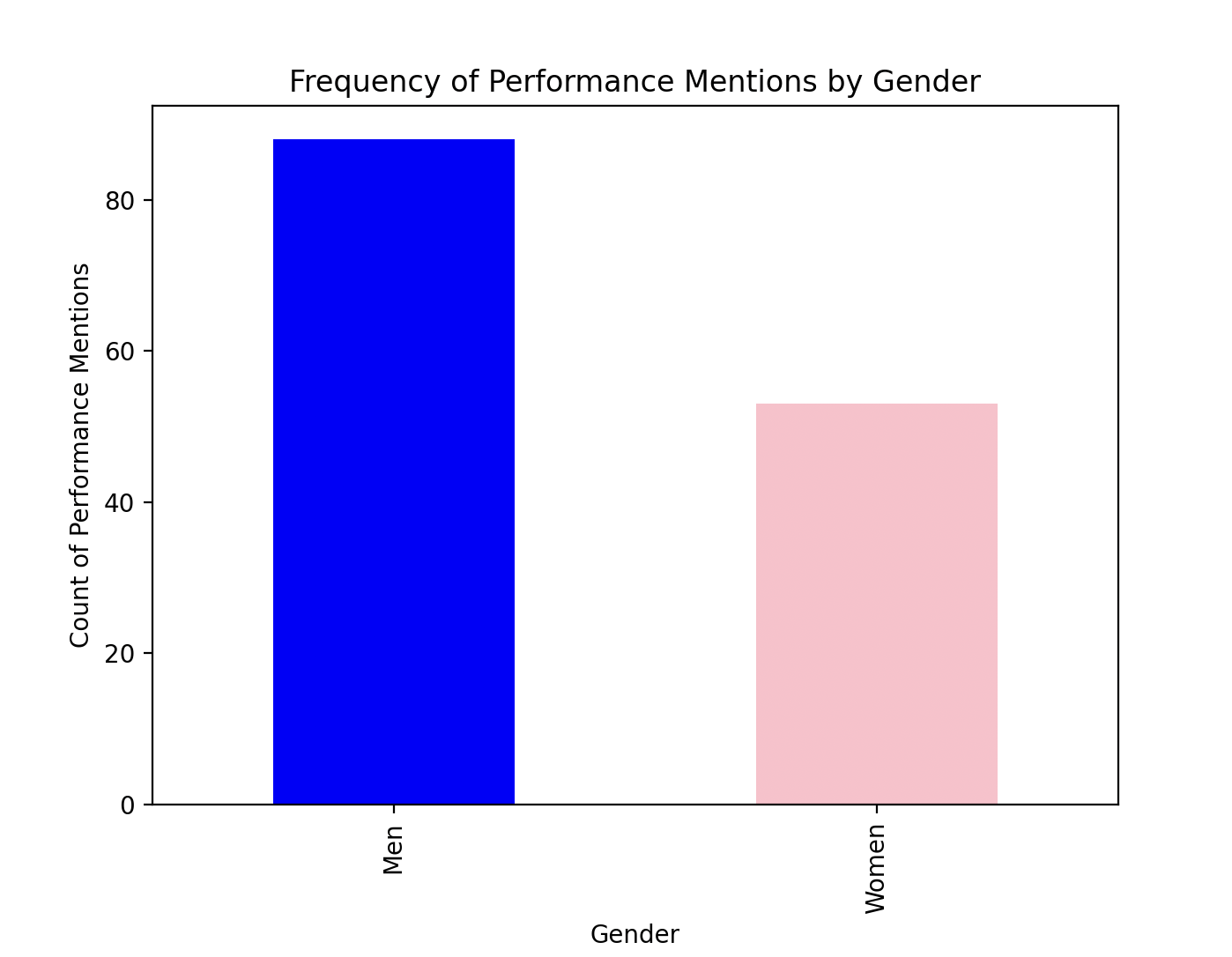
**Figure 1**

#### ***Topic Frequency***

Figure 2 shows a bar chart comparing the number of appearance-related tweets by gender, while Figure 3 highlights the top 10 players who are most frequently mentioned in appearance-related tweets. Appearance-related language was much more common in tweets about female players (13.2%) than male players (5.4%), while 8 of the top 10 most mentioned players by appearance are women, including the top 7. Figure 4 shows performance-related keywords, which were more prevalent in tweets about men (17.6% vs. 10.6% for women). Age-related terms (frequency bar chart in Figure 5) were also more prevalent in tweets about men (4.8%) than women (2%).



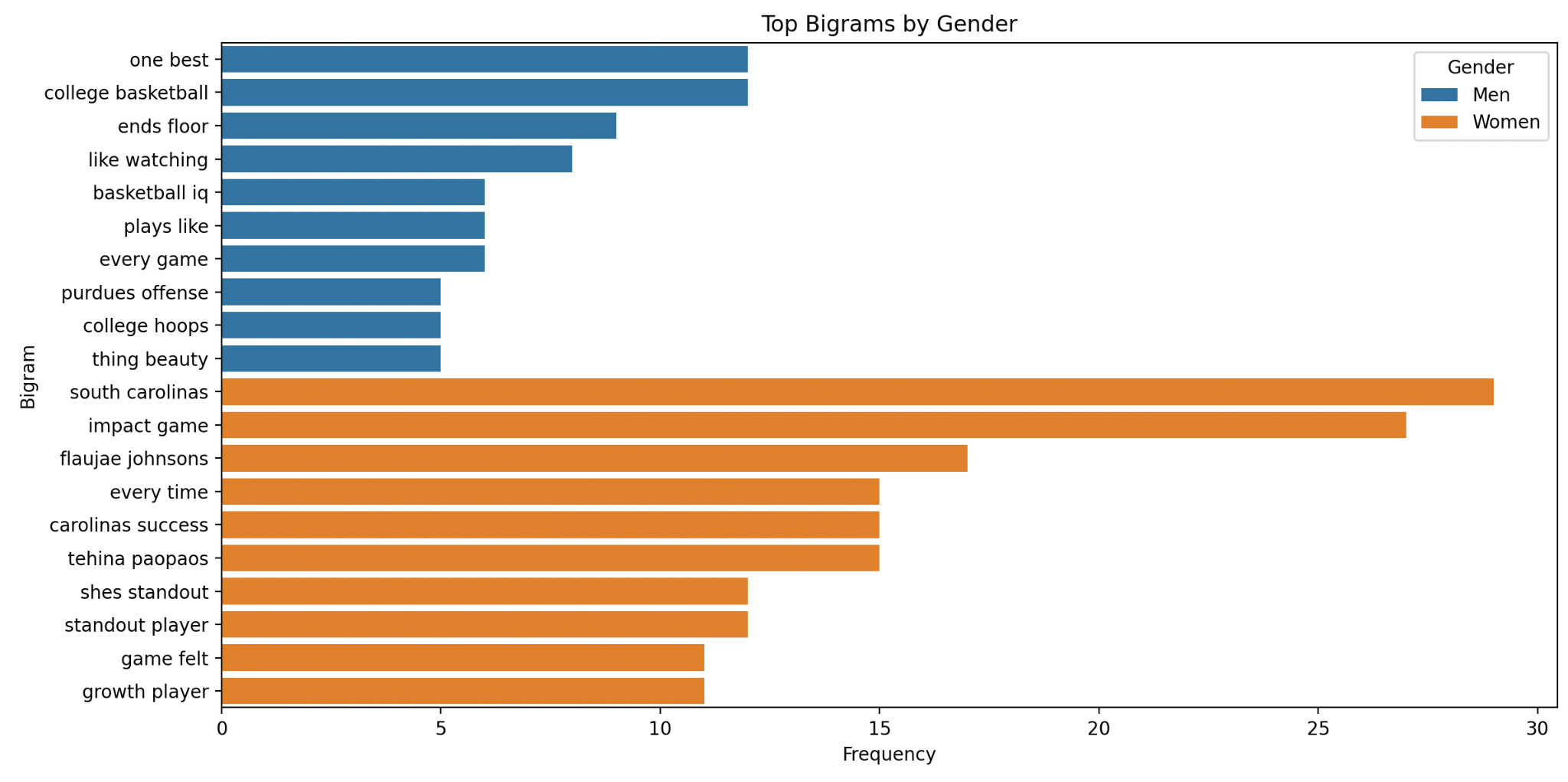
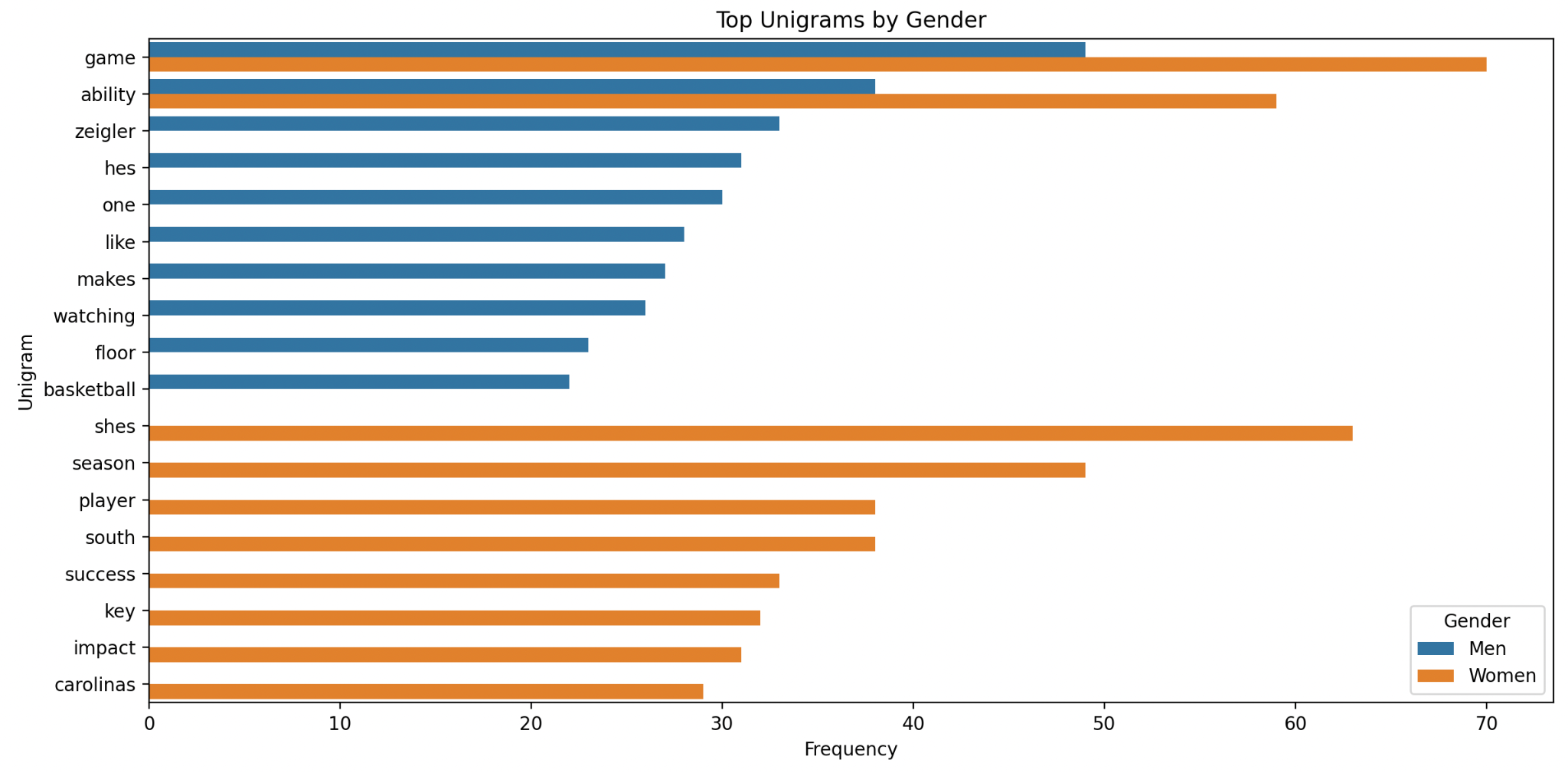
**Figure 2 & 3**



**Figure 4 & 5**

#### ***Word Frequency***

Figure 6 and Figure 7 present the most frequent unigrams and bigrams by gender. For both male and female athletes, terms like "game," "season," and "impact" were commonly mentioned, with bigrams such as "college basketball” appearing frequently for both groups. This suggests that the language used to describe both male and female athletes is relatively similar, with a shared focus on the sport itself, performance, and team associations. Despite some differences in specific terms, the overall patterns in language usage between genders are not as distinct as might be expected.



**Figure 6 & 7**

**DISCUSSION**My findings reveal an interesting contrast: although the topics of tweets about male and female athletes differed in meaningful ways, the overall sentiment of those tweets remained statistically similar. This suggests that while people discuss different aspects of male and female athletes, they tend to do so with comparable emotional tone.

One explanation for this pattern lies in the mix of positive and negative content within each topic category. For example, appearance-related tweets, more common for women, may include both praise (e.g., “so pretty” or “gorgeous”) and criticism (e.g., comments on hair or makeup choices), balancing out in aggregate sentiment. Similarly, performance-related tweets about men can range from admiration (“future NBA star”) to frustration (“terrible shot selection”), which could likewise average out to a similar tone overall.

Another possibility is that social media users tend to express support for top athletes, regardless of gender, especially during high-visibility events like March Madness. This could lead to a general positivity bias that masks underlying differences in the types of attention athletes receive.

Also, the similar use of unigrams and bigrams related to the sport itself, such as “game,” “season,” and “college basketball”, indicates that basketball discourse centers heavily on the shared experience of the sport, even if some gendered narratives persist. This growing focus on the game, especially for well-known women’s players, may reflect progress in how female athletes are covered and perceived in public discourse.

Ultimately, while gendered patterns remain in what aspects of athletes are highlighted, the emotional framing of these discussions appears to be more equitable than might be expected. Continued research could explore whether these trends hold outside the spotlight of major tournaments or among less-followed athletes.

## **CONCLUSION**

This project analyzed public discourse on social media surrounding the top men’s and women’s college basketball players, revealing important insights into how athletes are perceived and discussed. While the overall sentiment of tweets was statistically similar across genders, the topics of discussion diverged. Women were more frequently associated with appearance-related language, while men were more often discussed in terms of performance and age.

At the same time, the prevalence of sport-specific terms and the shared emotional tone suggest a growing convergence in how athletes of all genders are celebrated and critiqued. This reflects both progress and lingering disparity: the attention around women’s basketball is rising, but not always under the same lens as men’s. By highlighting these nuances, our findings underscore the need for more equitable, performance-focused narratives in sports media and public conversation.

Future research could extend this analysis across different sports, time periods, or include media coverage and comment sections on more structured social media platforms to gain a broader understanding of gender dynamics in sports discourse**.**

## **REFERENCES**

(Where I got the top 10 men's players)

<https://www.sbnation.com/college-basketball/2025/3/18/24383198/march-madness-best-players-2025-mens-ncaa-tournament-ranked>

(Where I got the top 10 women’s players)

<https://www.espn.com/womens-college-basketball/story/_/id/44310513/womens-march-madness-2025-ncaa-tournament-ranking-best-players-watkins-bueckers-hidalgo-betts>

**ESPN.** Player positions retrieved from individual athlete profiles on ESPN.com.<https://www.espn.com>

Twitter. (n.d.). *Tweets used in this study were collected from public Twitter profiles of NCAA basketball players*.<https://twitter.com>

## **STATEMENT OF AI HELP**

This study utilized AI and natural language processing (NLP) tools, including the VADER Sentiment Analyzer for sentiment analysis and spaCy for text preprocessing. ChatGPT was also used to assist in generating and refining some of the code for data analysis and text mining. These AI tools enabled the efficient processing of large volumes of text data, allowing for the extraction of insights on sentiment, topics, and language patterns related to gender differences in sports discourse.

## **APPENDICES**

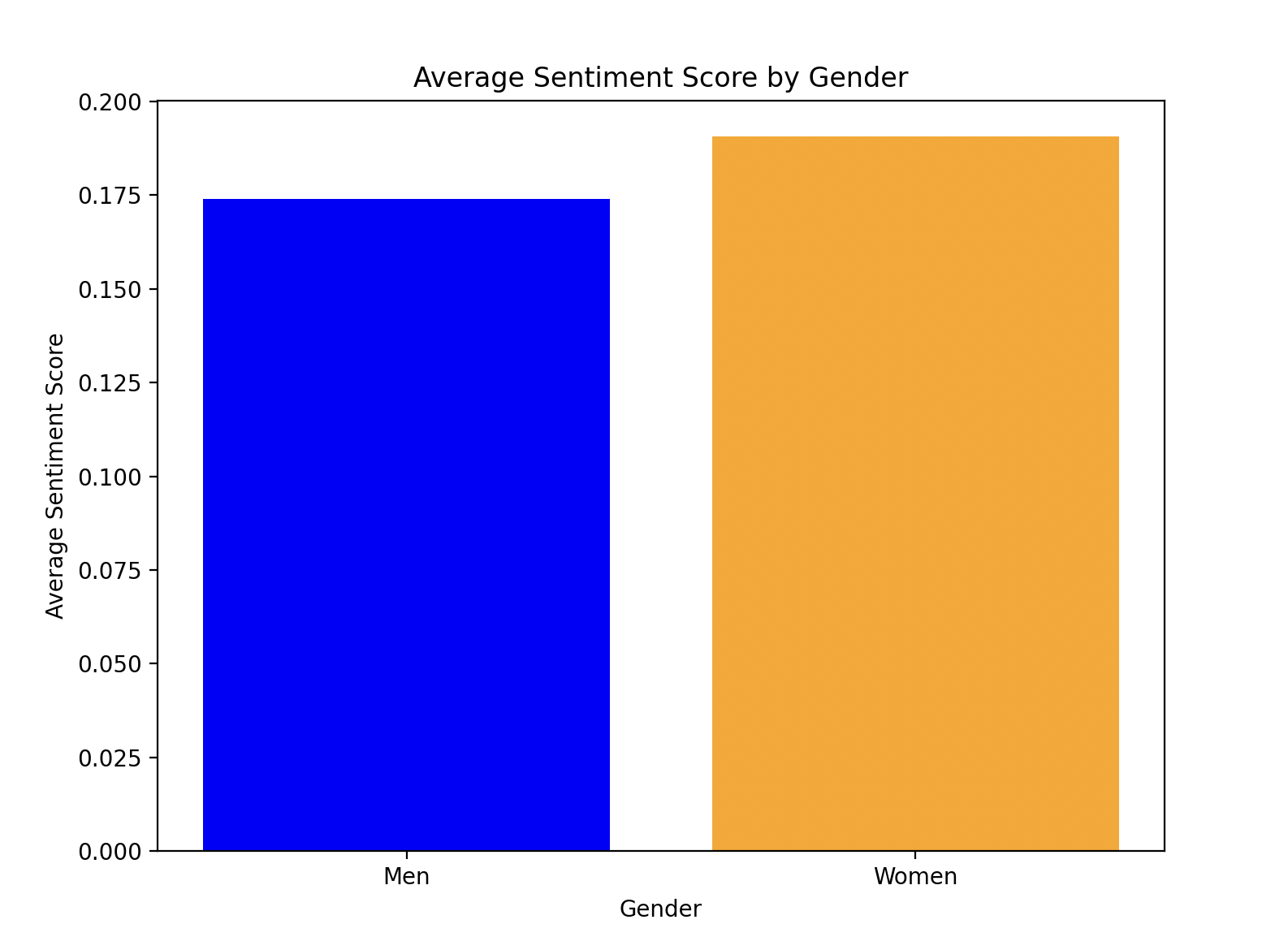
Figure 1: Average Sentiment Score By Gender

Figure 2: Frequency of Appearance Mentions By Gender

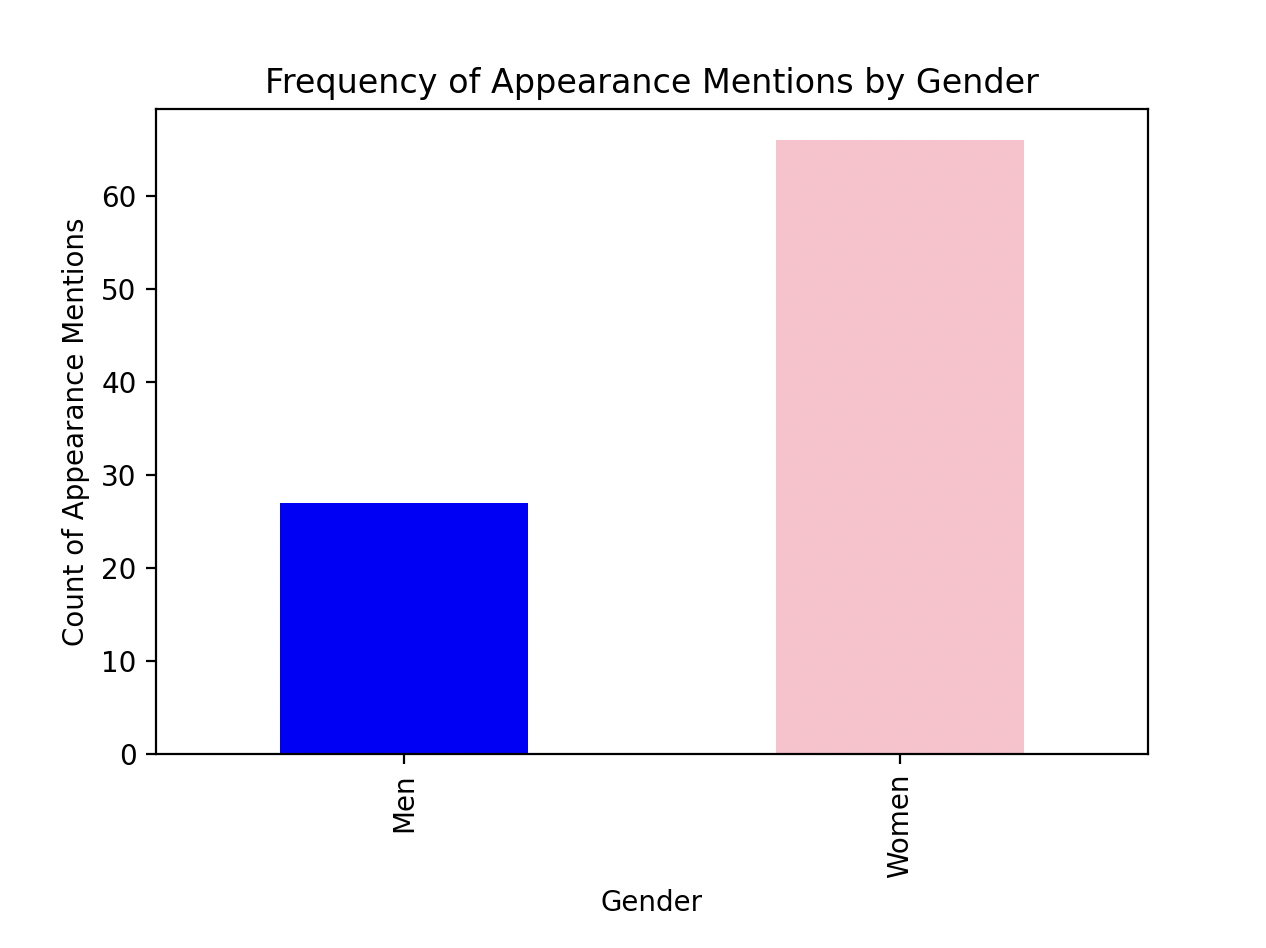


Figure 3:Top 10 players who are most frequently mentioned in appearance-related tweets

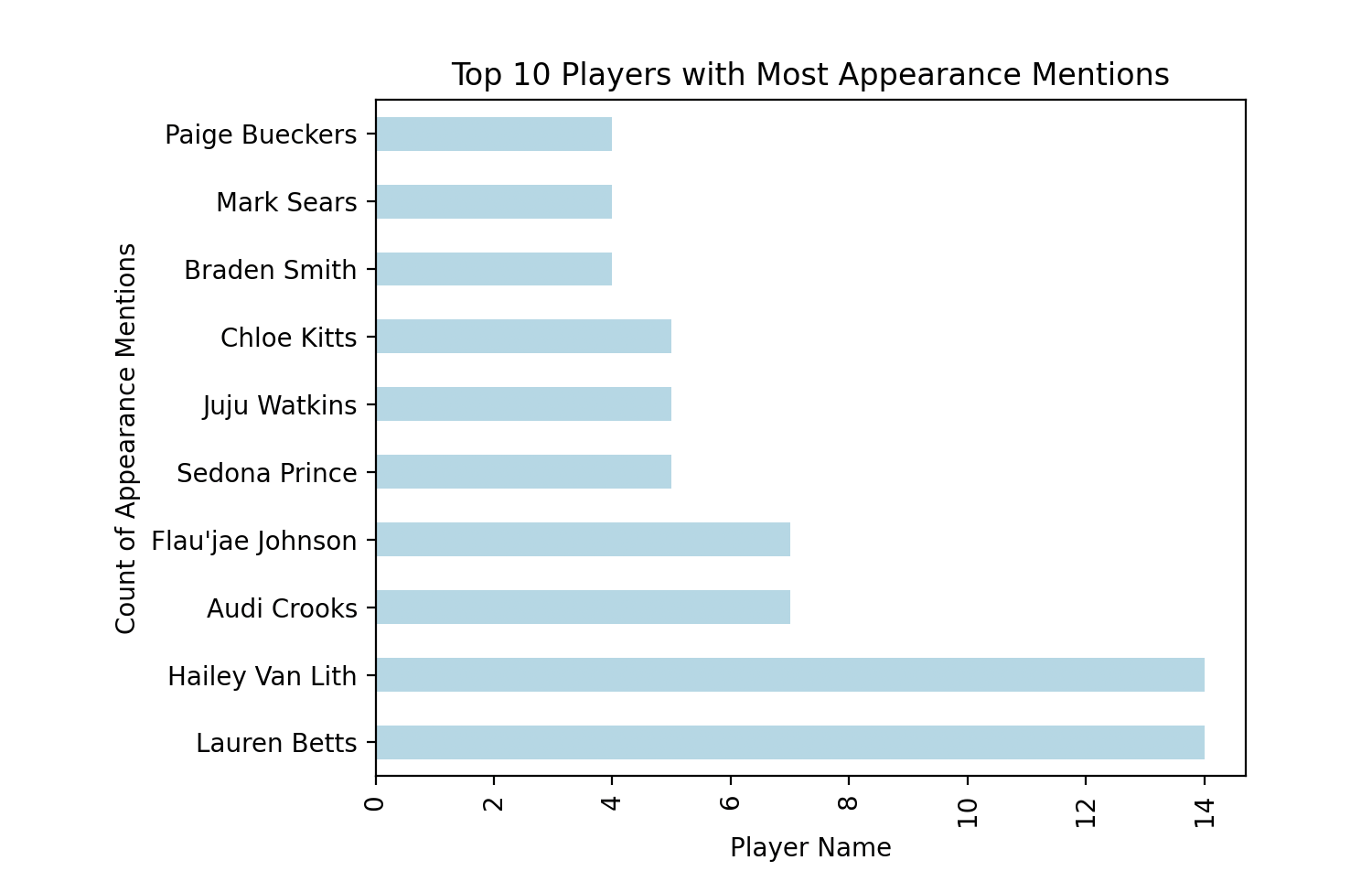


Figure 4: Frequency of Performance Mentions By Gender

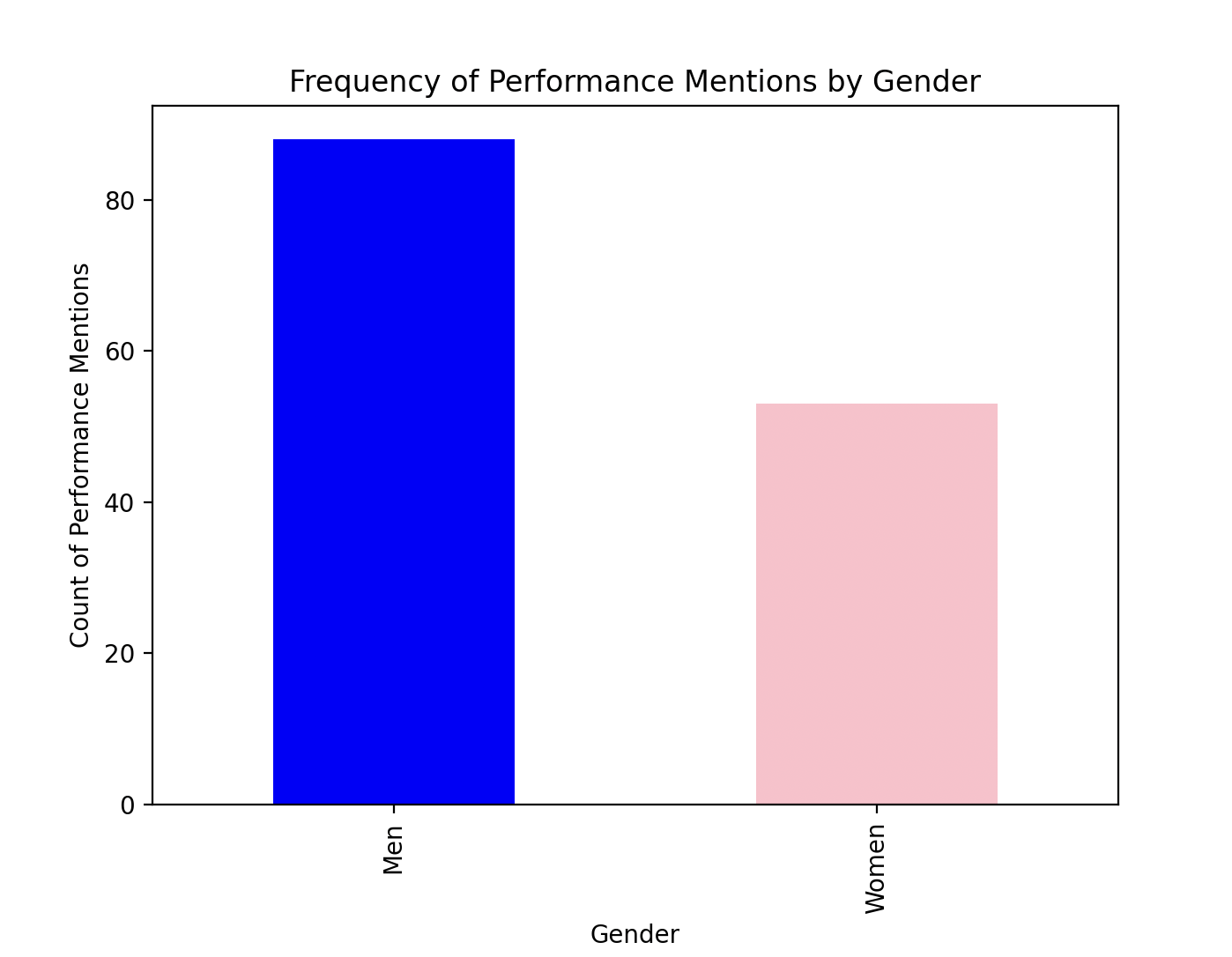


Figure 5: Frequency of Age Mentions By Gender

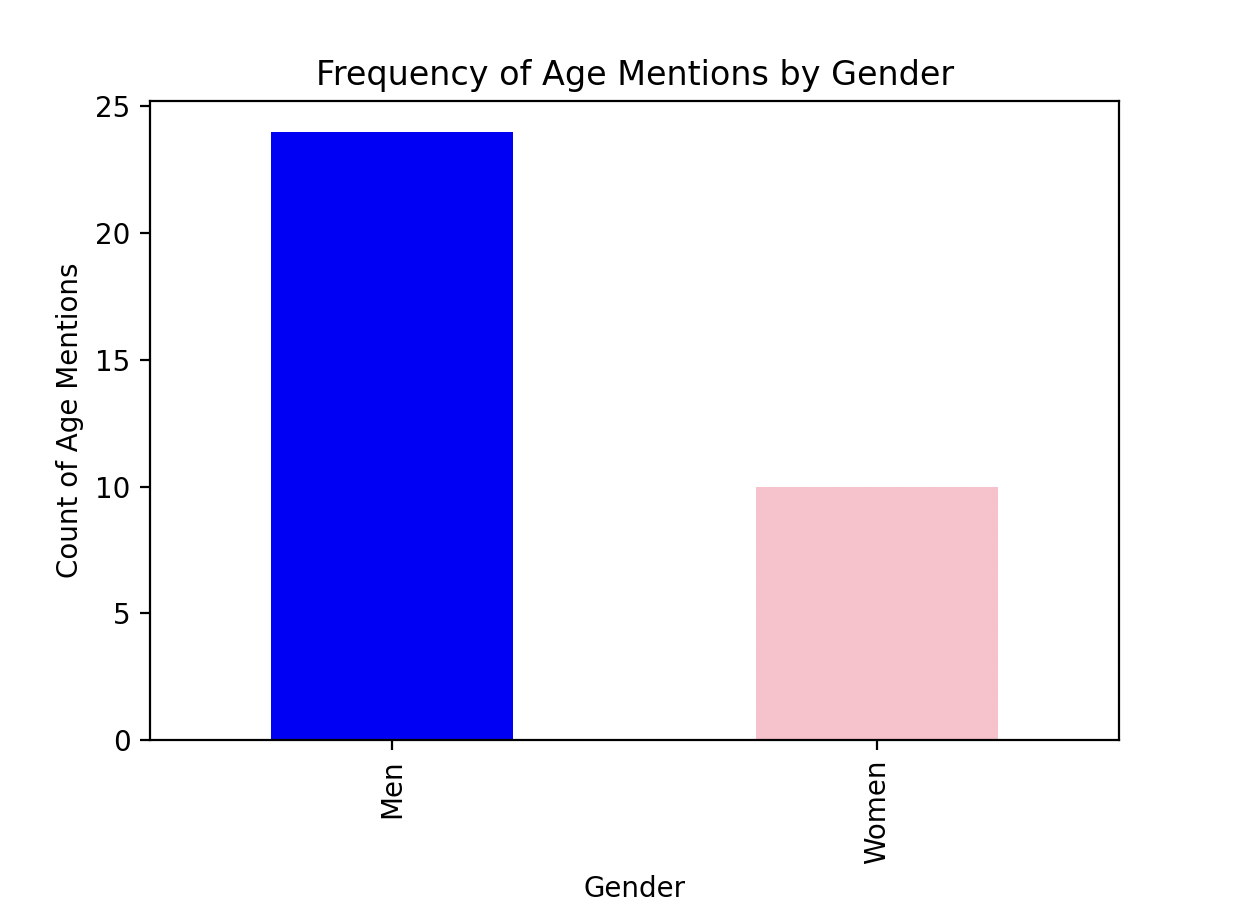


Figure 6: Top 10 Unigrams For Each Gender

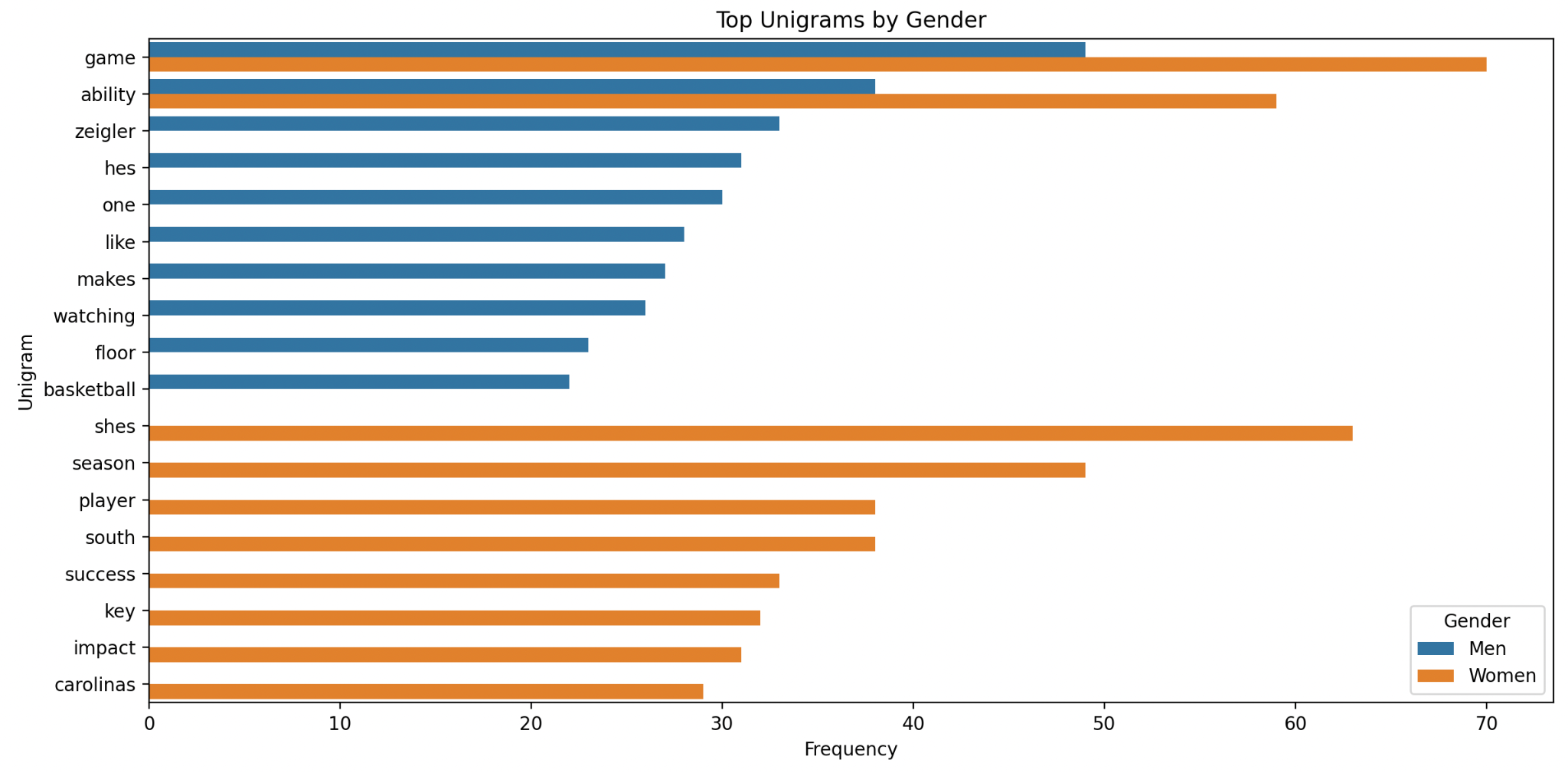


Figure 7: Top 10 Bigrams for each Gender

