How the Success and Failures of the US Women's World Cup Team Affects Prejudice

Aastha Jha
Amelie D'Hers
Sunaina Kathpalia

Georgetown University, McCourt School of Public Policy

December 12, 2023

Abstract: Never in the history of policies have we seen more factors affect people's political decisions. As the influence of sports expands in society, we question if it also influences policy via a lens of prejudice. Preliminarily we study the effect of a specific team's success on prejudiced speech and identify flags for further research into political implications. In this research, we analyze how prejudice changes as the US Women's National Team proceeds through four stages of the FIFA Women's World Cup in 2023. Through sentiment analysis models hate comments and comments with high toxicity on YouTube videos are 1. identified at higher rates when the team loses and 2. lower rates as the team wins during the tournament. Additionally, the frequency of politically affiliated words is high in comments identified as hateful. There are however limitations to the models discussed in the research, primarily the fact that YouTube, due to its user standards, preliminarily removes comments containing identity attacks. Overall, the research supports the hypothesis that sports success dictates the direction of shifting prejudice toward marginalized communities. Choosing a team so highly entrenched in political advocacy allows us to develop a framework for further analysis of the effect on politics.

Key Words: Sports, Politics, Prejudice, US Women's National Team, Hate, Toxic

Introduction: Sports Effect on Prejudice and Maybe Politics?

Sports are traditionally only thought of as entertainment, the thing you watch on TV regularly, and occasionally the nationalistic event of the year (depending on the year). Sports are far from the first response to a question about which factors influence politics. However, when critically analyzing the role of sports in society there is a clear link between sports and nationalism. The better a team does the more nationalistic sentiment rises (1). Furthermore, there is preliminary research to support the hypothesis that sports can influence prejudice against marginalized communities. As athletes succeed or fail to succeed, prejudice increases or decreases. (2). These two variables, nationalism, and prejudice very well could be completely separate. But what if they aren't? It seems improbable that there isn't a relationship between these three variables: sports, nationalism, and prejudice.

The observed relationship between sports and societal factors posits that there is a relationship between sports and politics. Our overarching hypothesis is that sports affect politics through prejudice and nationalism. As this is a big question to tackle, we begin with preliminary research on the link between sports and prejudice specifically with a team that represents marginalized communities *and* advocates for policy change for this marginalized community.

For the reasons stated we use the US Women's National team for preliminary research. In 2023 the USWNT, the reigning champions of the FIFA Women's World Cup, were eliminated from the championship in the round of 16, losing at the penalty shootouts against Sweden. Before the 2023 FIFA Women's World Cup, the US Women's National team had won the 2019 and 2015 FIFA World Cup. The public perception of the team going into the 2023 World Cup was that they were very successful as reigning champions. Prior to the 2023 World Cup, the USWNT had been vocal about supporting politics that affected marginalized communities. Specifically key leaders on the team vocally and publicly supported LGBTQ policy as seen here (3). Additionally, team members started a campaign for equal pay between women's and men's soccer teams in the US, further pushing the equal pay policy discussion in all industries (4). Years of advocacy led by USWNT led to actual policy change. In 2023 a new law was passed called The Equal Pay for Team USA Act which "ensures all Team USA athletes—including in future World Cups—receive equal pay and benefits regardless of gender." (5) It is also important to note that the most successful players on this team, including team captain Megan Rapinoe, were the most vocal about policy advocacy in the media.

Fast forward to the 2023 World Cup Round of 16 group stage, the USWNT loses. Immediately the team receives hate speech targeted at their identity. Yet this is not the only attack, their political advocacy and policy ideas also face severe scrutiny. To begin preliminary research on the effects of sports on politics we focus on a team that represents marginalized communities but also advocates significantly for policy solutions. The first of many hypotheses in this line of questioning presents as this: How does a change in a sports team's performance affect prejudiced speech towards the marginalized communities represented by the team? As mentioned, we will specifically be looking at the USWNT FIFA World Cup performance in 2023 to analyze the shift in hate speech and toxicity, specifically as it relates to marginalized communities.

The research presented in this analysis clearly shows that as the Women's World Cup Team experienced success, meaning they won the first three games of the World Cup, hate speech decreased. When the team lost in the final game they played, Round of 16, the hate speech increased significantly. Toxicity levels of speech followed a very similar pattern. Additionally, the research delves into word frequency which demonstrates that some of the most common words in hateful comments were surrounding political themes. The research supports the hypothesis that team success increases negative public sentiment. Furthermore, through word frequency analysis, we can draw a new hypothesis about the link between this speech and politics.

Data and Methods: Exploring YouTube Comments

This research relied on data from YouTube comments. YouTube is a commonly used social media video-sharing platform owned by Alphabet (the parent company of Google). The videos were not live streams of the matches, but rather recordings posted by both individual users of the platform and major broadcasters such as FOX Sports. We utilized the YouTube API to scrape data from the four matches played by the US Women's National team during the 2023 FIFA World Cup. The four matches included three group stage matches against Vietnam, the Netherlands, and Portugal respectively and the first-round match against Sweden in which the USWNT lost. Using the API's internal search feature, we collected the first 20 videos that came up when searching for each match title. Given that YouTube sorts by relevance anyway, we decided utilizing their relevance algorithm would best suit the needs for our project, since our search results were likely to be the most common videos that most people were viewing and commenting on as well. These videos included both full streams of the matches as well as highlight reels and additional commentary. From these 80 videos, we decided to scrape the entirety of the comments.

From the YouTube API, we collected the video ID, the like count, the reply count, the text of the comment, and the date of the comment and stored it in a Pandas data frame. During the collection process, we added the title of the game to correspond with each comment and converted the date to a DateTime data type. We collected approximately 71,000 comments resulting in a data frame of that many rows. Our dataset was at the comment level, with each comment representing one row in our data frame, a necessary element for conducting proper sentiment analysis. We encountered many comments that included characters beyond those in the English language including emojis. We decided to keep the comments with emojis and other extraneous characters because we felt that the other text surrounding these items would still be pertinent to our analysis even if the classification models, we were using later on would not recognize them. We do, however, acknowledge that emojis may change the meaning of a comment and can portray negative sentiment even on their own, and keep this in mind when continuing our analysis. In the same way, models are restricted by nuances such as sarcasm, emojis are another area in which the complication of this nuance is accepted within the context of our analysis.

By far the most important variable in the analysis is the text of our comments. Most of the conclusions presented in this research rely on the analysis of the text sentiment and content. Although game titles become important for grouping analysis, we still consider comment text the most interesting variable. Additionally, some analysis is presented around the number of comments' likes and replies.

Analysis: RoBERTa and Perspective Models

In this research two methods were used for sentiment analysis: RoBERTa and Perspective (6)(7). The models were applied to the comment text to determine the sentiment of the comments.

RoBERTa, Facebook's RoBERTa model fine-tuned on the DynaHate dataset, is a pre-trained model that classifies input text as either hate speech or non-hate speech. RoBERTa "uses self-attention to process input sequences and generate contextualized representations of words in a sentence" (8). This model is unique because it was trained on a larger data set than both Meta's RoBERTa and Google's BERT, theoretically giving it higher classification accuracy. The functionality extracted from the application of this model is hate speech classification and RoBERTa score. Hate speech classification classified comments into a dichotomous variable with values of 0 (False, interpreted as not hate speech) and 1 (True, interpreted as hate speech). RoBERTa score indicates the accuracy of the classification. This score is a probability as well. The model will classify a comment as hate speech if it's more than .5. A higher RoBERTa score approaching 1 indicates a more accurate classification. The bounds of this score are between 0 and 1.

Perspective is a pre-trained model originally meant to help content moderators identify the toxic directionality of conversations. "Perspective uses machine learning models to identify abusive comments. The models score a phrase based on the perceived impact the text may have in a conversation. Developers and publishers can use this score to give feedback to commenters, help moderators more easily review comments, or help readers filter out "toxic" language." (9). The functionality extracted from the application of this model is the toxicity score and identity score. A higher toxicity score indicates a higher likelihood that a user would perceive a comment as toxic. For example, the comment "this woman is ugly" with a toxicity score of 90% means that 9 out of 10 individuals would identify the comment as toxic (10). In addition to the toxicity attribute, we used the identity attack score which very similarly identifies the probability an individual would classify the comment as an identity attack or not. Perspective documentation recommends using a threshold of 70% for classification. All comments with a score above 0.7 (70%) are considered toxic comments. Any comment with a score below 0.7 is considered a non-toxic comment.

Both models provided us with a holistic understanding of the profile of comments on the top 20 YouTube videos of each game the US Women's National team played in the 2023 Women's World Cup. The scores provided inform our analysis of prejudice against marginalized groups, specifically women and LGBTQ+ people, before and after the success of the US Women's National team during the 2023 World Cup.

Results: Sports Success Direct Effect on Prejudice

I. Hate Score

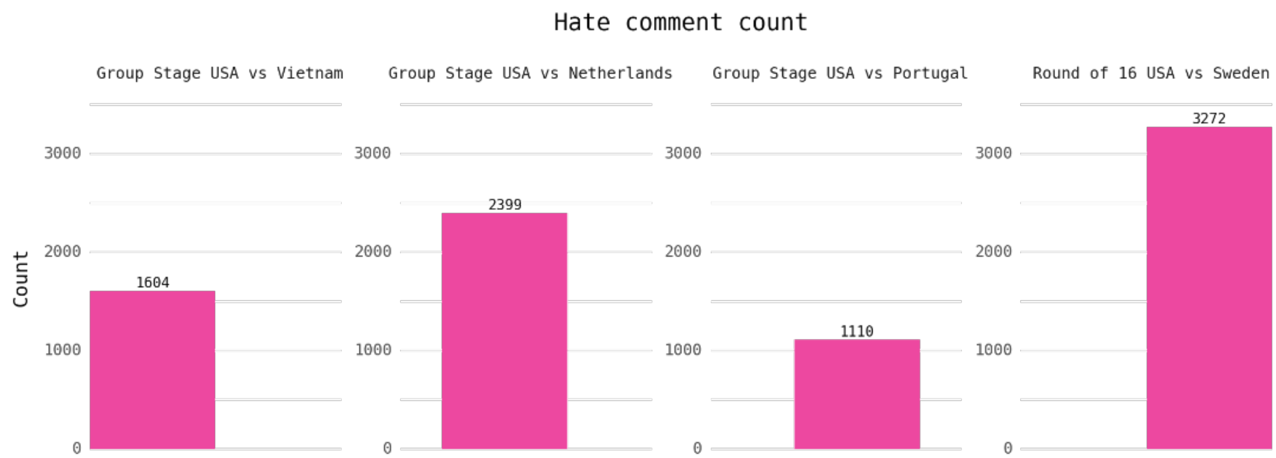


Figure 1.1

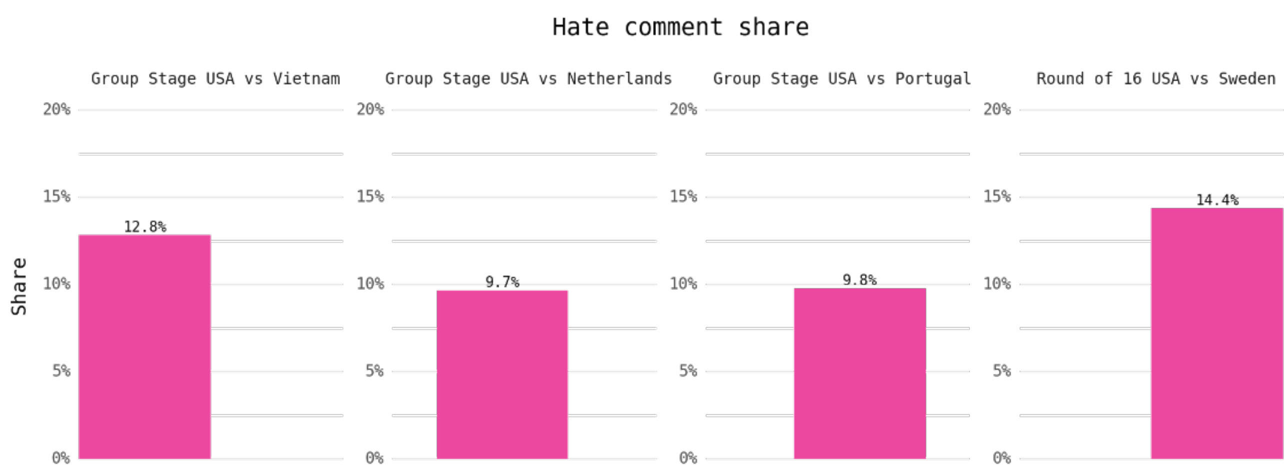


Figure 1.2

Substantively, Figure 1.1 demonstrates there is a clear difference between the number of hate comments on the match where the women's team loses. Being cognizant of the fact that there are bound to be more comments on a video of the team losing we shift our attention to Figure 1.2, which shows that the share of hate comments declines as the team wins matches in Vietnam, then the Netherlands, then Portugal. When the team loses the hate comment share rises by 5%, which is significantly more than any other percentage change between previous matches. This supports the idea that positive results decrease prejudice and negative results increase prejudice. In Figure 1.3 we see that even though there are significantly more comments on the Netherlands match videos, the percentage of hate comments is still less than on the Vietnam match videos where the USWNT lost.

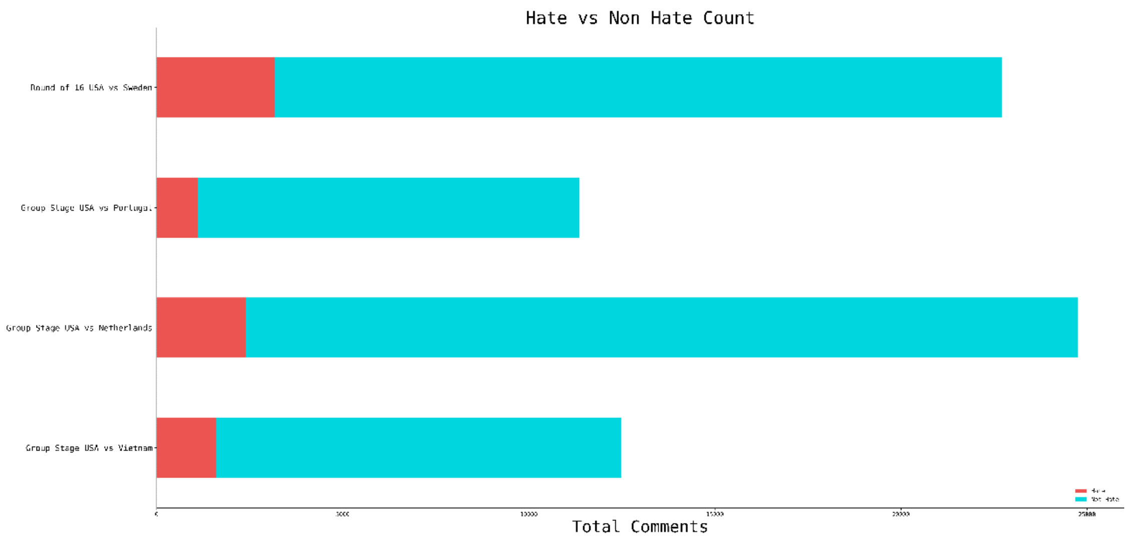


Figure 1.3

II. Reply and Like Count

An identical trend is represented in the analysis of likes and replies to hate comments. In Figures 2.1 and 2.2 as the women's team wins the first three matches, the replies and likes on hate comments decrease. When the team loses there is a stark rise in replies and likes on hate comments. Again, this further supports the hypothesis by showing the interaction with hate comments (also an indicator of prejudice expression) follows a similar trend dependent on the team's success.

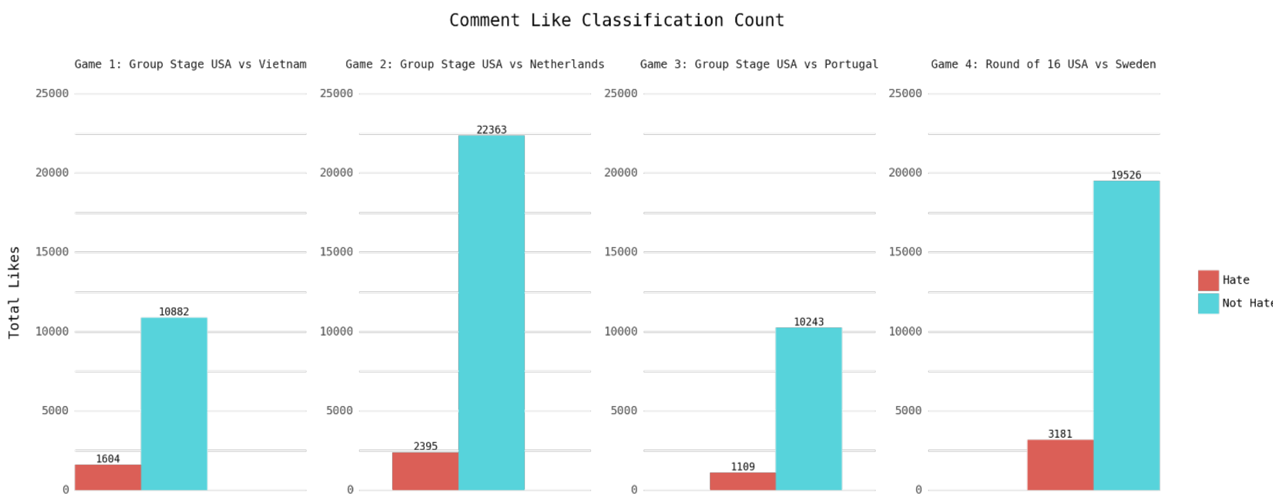


Figure 2.1

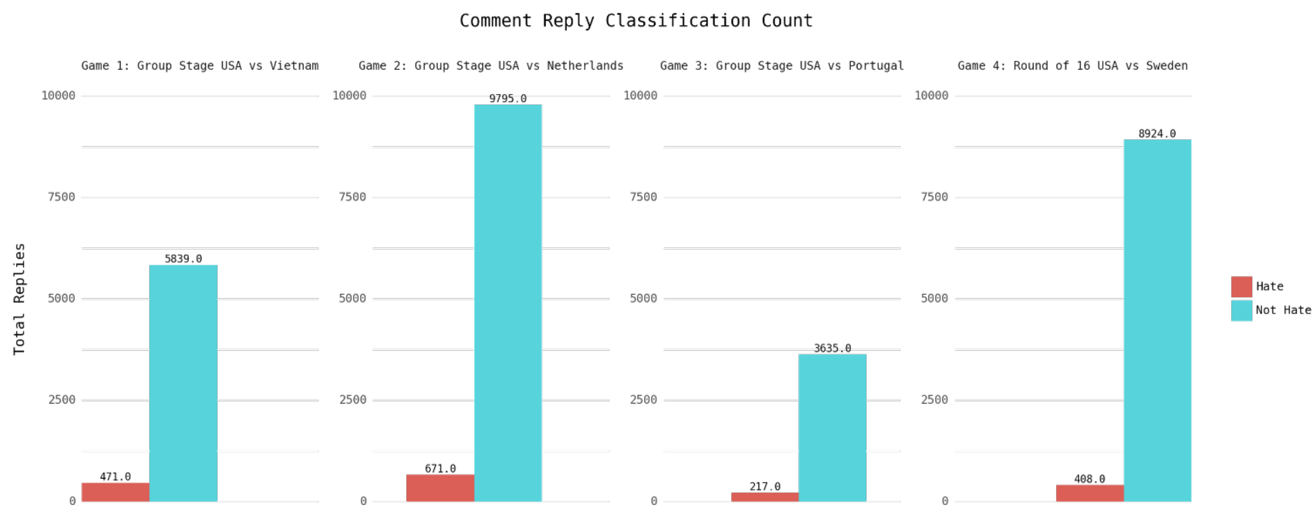
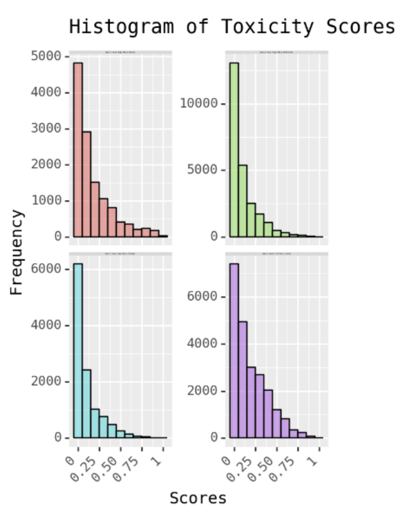


Figure 2.2

III. Toxicity & Identity Score

As seen in Figure 3.1 & Figure 3.2 the toxicity score reflects higher scores for the last game against Vietnam. Similarly, to what we saw with hate speech the lowest toxicity scores were in the middle two matches, against the Netherlands and Portugal, as the team was winning and progressing in the World Cup. This supports our research conclusion on hate speech analysis. We use this model to analyze toxic speech but also to confirm previous findings that prejudice does decrease as the team wins and increases when they lose.



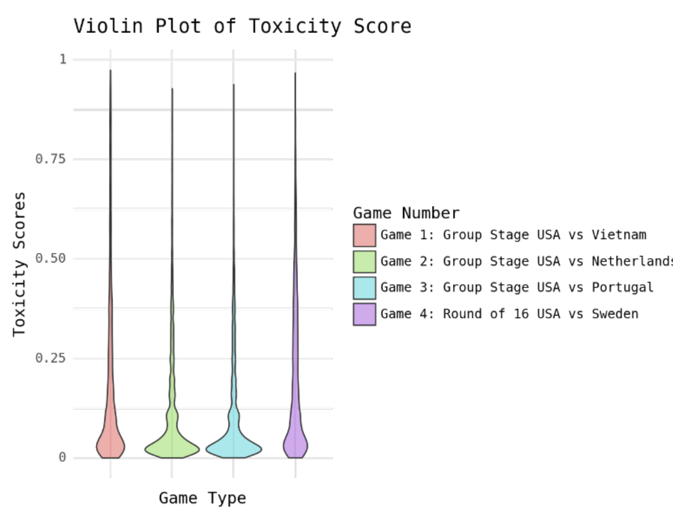


Figure 3.1

Figure 3.2

The identity attack scores in Figure 3.3 produce far less information. In terms of YouTube user agreement, targeting certain protected communities is not allowed. Because of this, we suspect many comments containing identity attacks likely were already removed by YouTube's content moderation practices. Therefore, although we present identity attack visually here, it is only to deem that it will not be used in our final research conclusion due to its lack of significance.

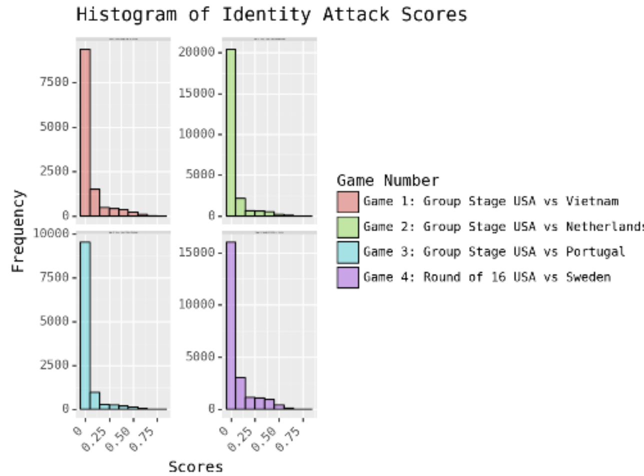


Figure 3.3

IV. Frequency of Words

To further analyze the content categories in hate comments we analyzed the frequency of certain words used. After removing filler words, including a list of words that were soccer related (e.g. goal, game, soccer, etc.) and gender identifying (e.g. women, women's, girl), we arrived at the results displayed below. One key indicator in Figure 4.1 is that “woke” is the third most-used word in hate comments. “Woke” is an ironically derogatory term used in the US to denote a person or concept who is liberal leaning politically. We observe that the third most frequent word in hate speech indicates political affiliation. This conclusion prompts an obvious hypothesis as to how the effect of politics factors into this analysis of the relationship between sports and politics. We will discuss this more in further analysis below, but it is an excellent indicator for investigating a new further-fetched hypothesis that sports influence politics via prejudice. Similarly, Rapinoe, the team captain of the US Women's National team and the staunchest policy advocate is mentioned significantly more in hate comments than in positive comments. In preliminary research, we found that Rapinoe was often criticized specifically for her politics over the mention of her athleticism. Her frequent mention of hate comments supports further research into her role as a policy advocate in sports and supports the idea that athletes as a whole influence policy. Finally, another important observation is that a lot of the same words used in hate comments are also used in non-hate comments. This is an interesting observation that helps shape our understanding of hate comments. Additionally, this has implications for large language models, including the ones we use, which aim to correctly classify different speech patterns with similar verbiage.

Hate Comments

Figure 4.1

Non-Hate Comments

Figure 4.2

Discussion: The Future Role of Athletes in Policy

To recap, the main findings of the research are as follows. As the USWNT wins, hate comments decrease substantially in quantity and proportionally. When the USWNT loses, the hate comments increase significantly. Toxicity levels support this finding by following the same trend. Additionally, like and reply to hate comments follow the same pattern, again reinforcing our primary observations. Finally, the word frequency analysis shows that words involving politics are used frequently in hate comments.

As discussed previously, our overarching hypothesis is that there is a relationship between sports and politics. We believe that this correlation is mediated partially through prejudice, so we have tested prejudice in the model. To analyze the effort of sports on the increase and decrease of prejudice, we model the patterns of hate speech on the team winning versus losing. Very simply put, when the USWNT team wins the hate speech decreases by approximately a few percentage points. When the team loses, they experience more hate comments than all other winning games. The relationship between the team's success and prejudice is clear. To solidify these findings, we will need to diversify the research. Meaning we must look at other trends amongst different types of teams, potentially with different communities. This would make our predictions more robust if the findings were consistent with the findings above.

Furthermore, we found in the research that politics was a frequently mentioned topic in hate comments. The USWNT was chosen because of how highly political they are. If hate comments here were just about prejudice, we would have observed words associated with prejudice as the most frequent. But this is not the observation. Instead, we observe politically affiliated words to be the most frequently used words. This supports the furthering of this research into the crux of our initial question: is there an effect of sports on politics? There is enough preliminary evidence in this research for the hypothesis that there is a connection between these two. The next step would be to conduct research in this area perhaps with a linear regression to identify if the relationship exists or the creation of another model tuned to identify political sentiment in comments.

Sports are traditionally thought of as exclusively entertainment, however, there is evidence of how important they may be in policy, nationalism, and society as a whole. Sports success determining prejudice can have huge implications for the treatment of marginalized communities as discussed in the articles cited above. But even more surprisingly sports *influences* what people think about their political motivations and decisions. Advocacy and policy work would benefit from the social leverage sports teams and players provide to social issues. As we see in the USWNT there is much backlash in hate comments but that means policy advocacy is getting the attention of viewers. People care about what players have to say, more accurately, what their idols have to say about politics. There is a huge opportunity for advocacy research to be done to use this societal power for positive impacts on marginalized communities and policy changes. Additionally, athletes are often cast aside as influential figures in society. Yet, athletes are becoming important moral figures to their communities and thus can no longer be cast aside. American society has already begun to see the effects of athletes in policy such as football players in vaccine campaigns or the MLB hosting special LGBTQ advocacy events at baseball games. However, as stated, this advocacy can also turn negative, especially after varying performances. Analyzing all these implications through a data science lens gives us the confidence to say that this is an important area of policy research that should be continued. Future campaigns, government program implementation plans, and social movement awareness are all examples of what athletes could be leveraged for to drive effective policy/government change. More importantly, it is time to stop pretending that athletes are only entertainment and treat them like the truly important members of history, culture, and society that they are.

Works Cited

1) "Olympic Athletes Compete for National Pride. - Kathleen E. Powers." *The Washington Post*, 26 July 2021, www.washingtonpost.com/politics/2021/07/26/do-olympics-promote-nationalism-international-conflict-heres-research/.
2) Alrababa'H, Ala', et al. "CAN Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes: American Political Science Review." *Cambridge Core*, Cambridge University Press, 7 June 2021, www.cambridge.org/core/journals/american-political-science-review/article/can-exposure-to-celebrities-reduce-prejudice-the-effect-of-mohamed-salah-on-islamophobic-behaviors-and-attitudes/A1DA34F9F5BCE905850AC8FBAC78BE58.
3) "'Go Gays!': Why USA's LGBTQ World Cup Stars Are More Important than Ever." *The Guardian*, Guardian News and Media, 5 July 2019, www.theguardian.com/football/2019/jul/05/usa-gay-players-womens-world-cup-megan-rapinoe.
4) Das, Andrew. "U.S. Soccer and Women's Players Agree to Settle Equal Pay Lawsuit." *The New York Times*, The New York Times, 22 Feb. 2022, www.nytimes.com/2022/02/22/sports/soccer/us-womens-soccer-equal-pay.html.
5) "Icymi: Cantwell's New Law Guarantees Equal Pay for All Future World Cups." *U.S. Senate Committee on Commerce, Science, & Transportation*, 21 July 2023, www.commerce.senate.gov/2023/7/icymi-cantwell-s-new-law-guarantees-equal-pay-for-all-future-world-cups#:~:text=New%20law%20ensures%20all%20Team,newly%20signed%20national%20law.
6) *Roberta - Hugging Face*, huggingface.co/docs/transformers/model_doc/roberta. Accessed 12 Dec. 2023.
7) "Using Machine Learning to Reduce Toxicity Online." *Perspective*, perspectiveapi.com/. Accessed 12 Dec. 2023.
8) "Overview of Roberta Model." *GeeksforGeeks*, GeeksforGeeks, 10 Jan. 2023, www.geeksforgeeks.org/overview-of-roberta-model/.
9) "Using Machine Learning to Reduce Toxicity Online." *Perspective*, perspectiveapi.com/how-it-works/. Accessed 12 Dec. 2023.
10) "Developers." *Perspective*, developers.perspectiveapi.com/s/about-the-api-score?language=en_US#:~:text=A%20higher%20score%20indicates%20a,perceive%20that%20comment%20as%20toxic. Accessed 12 Dec. 2023.