Project Proposal

Date: 10/31/2023

Authors: Amelie D'Hers, Aastha Jha, Sunaina Kathpalia

Topic:

How does a change in a sports team's performance affect prejudiced speech towards the marginalized communities represented by the team?

To do this we will specifically be looking at the US women's world cup team performance in 2023 to analyze the shift in sentiment towards three categories of communities: gender, sexual orientation and race.

Background:

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. This outcome was entirely unexpected and was the first time since the 2016 Summer Olympics that the team didn't reach the semi finals of a major tournament. The team's elimination was met with outrage on social media platforms, and included a significant amount of negative comments regarding women and LGBTQ+ members. Prior to the 2023 world cup, the US Women's team had leveraged their success as reigning champions in order to become advocates of social policies such as LGBTQ+ protection and elimination of the gender wage gap.

Data:

We intend to create our own data set that contains comments from videos or streams of every match the US Women's National Team played during the 2023 FIFA world cup and the qualifying matches just before. In order to create this data set, we will have to scrape sites that host videos and streaming such as Youtube and Twitch for the comments around each match. We also want to try pulling comments from gambling platforms like DraftKings that also host live commenting during matches.

Methodology:

The first method we will employ is data scraping. Using existing packages like the Youtube API wrapper and HTML packages like beautiful soup we will scrape Youtube, Twitch and other streaming platforms to collect data from comments, and sort it by match played.

Next we will employ data wrangling by cleaning up the data we scrape in order to fit the format needed for our analysis. This will include filtering comments to only be about the US women's

national team and not their opponents. The next step will be to conduct text analysis by conducting sentiment analysis on the data we have collected. We will have to come up with a classification system for whether a comment contains positive and negative sentiment about a marginalized community. At this point we can conduct EDA or exploratory data analysis of our cleaned up data set. We will employ data visualization packages such as plotnine, ggplot and seaborn to create visualizations that characterize and summarize our data best. These will include plots that show how comments changed over time. Finally, we will conduct statistical analysis in the form of an interrupted time series analysis to show whether there was a change in sentiment before and after the loss during the round of 16 at the 2023 FIFA world cup.

Definition of Success/Objectives:

The ultimate objective of this project is to be able to use data to confirm a long standing hypothesis. As diligent sports fans and policy fanatics, we all have a deep curiosity surrounding the seemingly nationalistic link between the two. We have all experienced the increase in prejudice targeted at groups that the US women world cup team represented and advocated for. This hypothesis is strong but lacks evidence base data analysis. Success would be first and foremost proving or disproving our hypothesis.

From this analysis we would like to be able to show in general how sports success or decline in success changes prejudice around certain marginalized communities. Specifically as stated above we will be using the 2023 women's world cup to understand how this phenomenon manifests. We will be producing multiple visuals to demonstrate trends in our sentiment analysis and further explain insights. We will also be presenting analysis on some of the most common comments before and after the decline in success of the US Women's team during the 2023 World Cup.

In order to be successful we want to not only be able to show the relationship between sports and prejudice but also to use these insights for further research. Sports are traditionally thought of as exclusively entertainment, however more research shows how important they can become in policy, nationalism and now prejudice. Advocacy and policy work would benefit from the social leverage sports teams and players provide to social issues. Our objective is not only to be successful in our analysis but to go further and provide tangible evidence for future advocacy strategies that include athletes. Additionally, athletes are often cast aside as influential figures in society. We expect this analysis to provide evidence that athletes are important advocates in society. 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 before and after performance that are important to account for. We plan to analyze all of these implications through a data backed lens in our final project.