**Progress report document structure:**

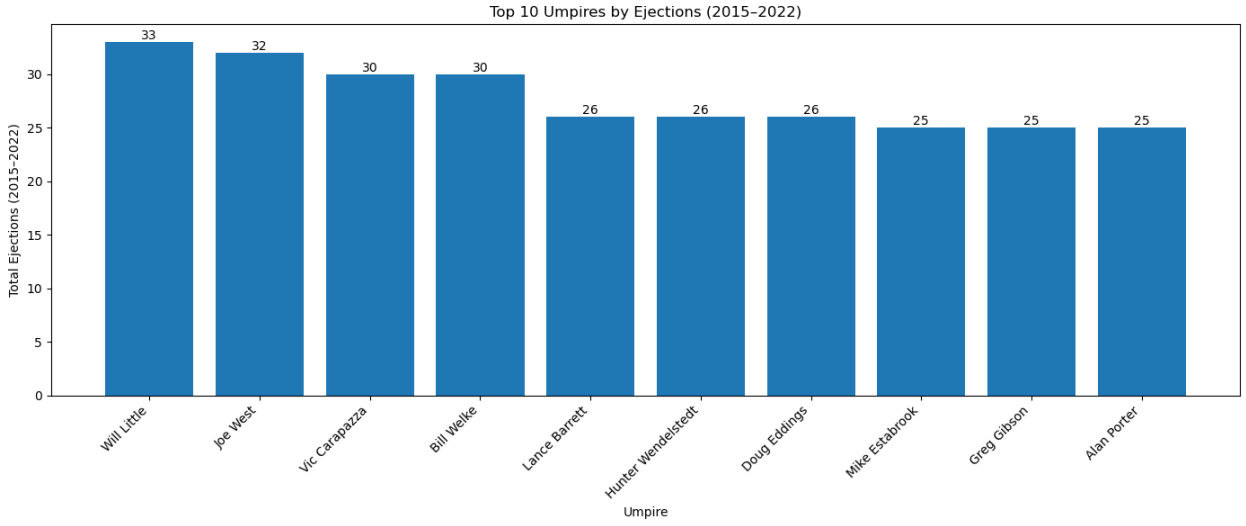
Project scope update:

Originally, my main goal for my project was to analyze umpire ejections and see if the umpire's poor accuracy led to more ejections. After trying to make python codes to help collect and process data, I have hit multiple roadblocks. As a result, I am slightly updating the question for my project. My new question is, **if a player is ejected from a game, is that team more likely to lose the game?**

Right now, I have no trouble working with the csv file with a list of all MLB umpire ejections. I am able to create boxplots easily showing what umpires eject the most, who get ejected the most, and the reason they got ejected the most.

My main challenge right now is getting the result of the game in which a team was ejected. I was able to create a python code that uses MLB API to give me the box score of the game. My main issue is that the format of the GAMEID from retro sheet is different from what the MLB uses for API. Once I figure that out, I will be able to pull the game data and store who the winner is in my database.

Finally, to use my third source, there is a Kaggle database with MLB Umpire scorecards, I can find the top 10 umpires with ejections in the range of the Kaggle Database(2015-2022) and compare their overall accuracies and see if the umpires who love to eject players have a worse accuracy.




**Issues / difficulties**

- Data cleaning – my main hurdle right now is cleaning in a way that lets me determine umpire accuracy. I don't think I have the coding power right now to be able to pull the accuracy of the umpires myself, I will have to use umpscorecards for that data.
  I think what would be in the scope of what I can do is pull the box score data using MLB api and append it to my database of umpire ejections using the game ID. The main issue I am running into right now is that the GAMEID format from the retrosheet csv file, is a different format from the MLB API. Once I figure out a way to convert the gameID format, I can look at MLB ejections and get the box score from that game.