**Project Title**: Player Trends of the Chicago Cubs Across the Years

**Team Members**: Project Group 2:

* Emily Bowers
* Joe Cotten
* Amanda Dunaway
* Luis Martinez
* Carlie Rhoads

**Project Description/Outline**: The driving purpose of this project is to examine comprehensive data and determine trends across players of the Major League Baseball team, the Chicago Cubs. We have access to pitching and batting statistics, and will be looking at trends such as if age has an impact on factors such as runs batted or runs allowed.

**Research Questions to Answer**:

* What impact does age have on baseball performance for players of the Chicago Cubs? We will look at statistics such as runs batted specifically with successful hits, doubles, triples, and home runs. A regression will be run too.
* What hitting statistics (on base percentage, batting average, and slugging average (if any) have a relationship with the number of runs a player scores?
* Does player position have an impact on batter performance? We will look at statistics such as the player’s position in comparison to their runs batted, bases stolen, and home runs.
* Does a pitcher’s overall performance impact game losses and wins? We will compare game losses and wins with pitcher variables such as hits per inning, runs allowed, and walks allowed.
* Does a player’s dominant hand (i.e. left or right) have an impact on the player’s performance? We will do a comparison of hand dominance versus pitcher strikeouts, if the player is a pitcher, and successful hits at bat.

**Datasets to be Used**:

Chicago Cubs Batting & Pitching (1876-2023) – downloaded from Kaggle (<https://www.kaggle.com/datasets/mattop/chicago-cubs-batting-and-pitching-1876-2023>)

This dataset offers a comprehensive archive of batting and pitching performances by the Chicago Cubs in Major League Baseball (MLB), covering every single year from the team's very beginning and creation all the way to the current year. This data set offers a robust number of statistics, including two CSV files that concentrate on batting and pitching respectively. Demographic data of the players is another example of included information, as well as batting and pitching statistics.

**Rough Breakdown of Tasks**:

* Clean and organize the data in both datasets, check for null data and duplicate data
  + Use Pandas to clean and format your dataset or datasets.
  + Create a Jupyter notebook describing the data exploration and cleanup process.
* Analysis of data
  + Create code – include explanation for each section of code.
  + Run code
* Visualization of data
  + Create code
  + Run code
  + Use Matplotlib to create 6 to 8 visualizations of your data (ideally, at least 2 visualizations per “question” that you ask your data). This means we need at least three research questions.
  + Save images as PNG and share with group, especially with members tasked with the report and presentation slide creation
* Create a write-up summarizing your major findings. This should include a heading for each “question” that we asked our data as well as a short description of our findings and any relevant plots
  + Written at professional level
  + Include major findings and implications
  + Strongly supported by our findings, i.e. numbers and visualizations
  + Uses at least some of the following: aggregation, correlation, comparison, summary statistics
* Presentation
  + Create slides
    - Must be visually clean and professional
    - Will use Google Slides for group collaboration
    - Must be relevant to material
    - Must be clear and maintain audience interest
  + Decide who is going to speak for each slide (presentation is ten minutes, so each member should likely plan to speak for two minutes)
  + Practice the presentation