**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**: The driving purpose of this project is to examine pitching and batting data to determine trends across players of the Major League Baseball team, the Chicago Cubs.

**Research Questions**:

* 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.

* 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**:

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 year from the team's 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.
  + Use Pandas to clean and format datasets.
  + Create a Jupyter notebook describing the data exploration and cleanup process.
  + Explore our data types, columns, etc.
  + Check for null data and duplicate data.
  + Perform any exploratory data analysis to possibly reveal any additional research questions.
* Analyze and visualize the data.
  + For each research question:
    - Pseudocode: outline our desired outcome and route.
    - Create code.
      * Use functions (.groupby(), .mean(), .loc[], etc.) to examine relationships.
      * Use Matplotlib to create 6 to 8 visualizations (ideally, at least 2 visualizations per “question” that we ask our 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.
      * Include explanation for each section of code.
* Create a write-up summarizing our major findings.
  + Include a heading for each “question” that we asked of our data, as well as a short description of our findings and any relevant plots.
  + Include major findings and implications.
  + Include numbers and visualizations to strongly support our findings.
  + Uses at least some of the following: aggregation, correlation, comparison, summary statistics.
* Presentation
  + Create slides
    - Will use Google Slides for group collaboration.
  + Decide who is going to speak for each slide (The presentation is ten minutes, so each member should likely plan to speak for two minutes).
  + Practice the presentation.