

# MoneyBall

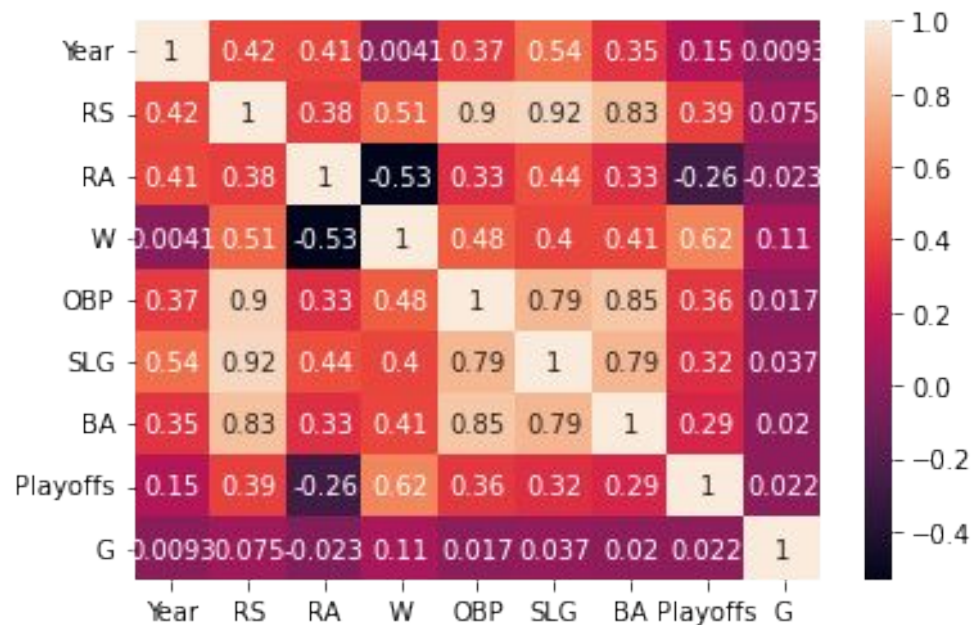
Nick Pipal

# Introduction

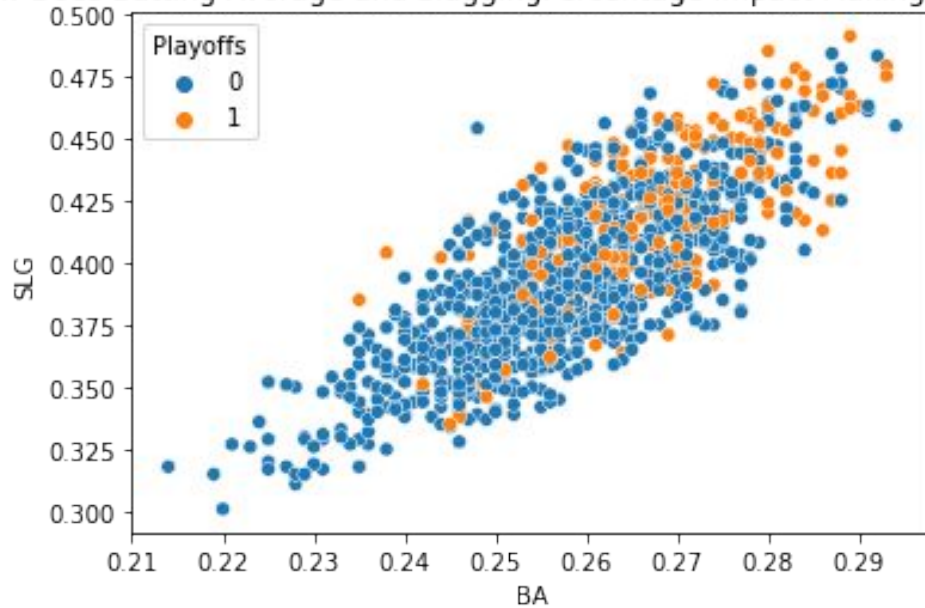
- Where was this dataset found?
  - This dataset was found on Kaggle.
  - Made by Wes Duckett
  - Data was from 1962-2012
- What is the Machine Learning Problem?
  - What statistics determine whether or not a team makes the Playoffs or not.
- Feature Columns
  - League, Year, Runs Scored, Runs Allowed, Wins, On-Base Percentage, Slugging Percentage, Batting Average and Games Played
- Target Column
  - Playoffs

How can we better understand the data?

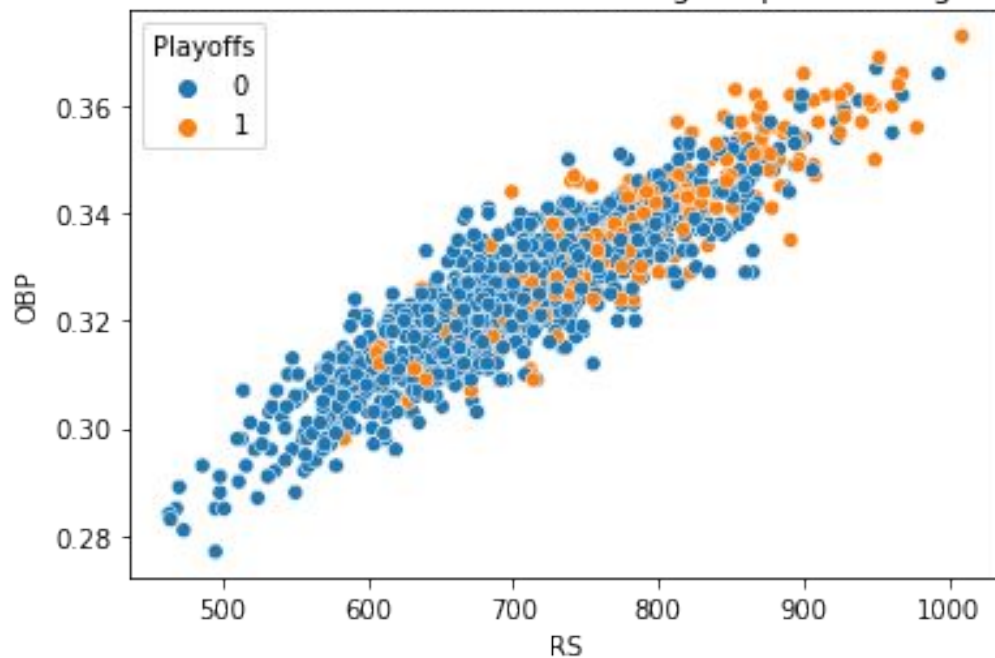
# Exploratory Visuals



### How Does Batting Average and SluggingPercentage Impact Making the Playoffs



How Does RunsScored and OnBasePercentage Impact Making the Playoffs



# Production Models

- What models did I use?
  - Random Forest, KNeighbors, Bagging Tree, and a Decision Tree
- Which model performed the best?
  - The Random Forest model did end up performing the best
  - First glance:
    - Train Score = 92%
    - Test Score = 57%
  - After Tuning the Hyperparameters:
    - Train Score = 69%
    - Test Score = 56%

# Final Recommendations

- What does my final model tell me?
  - First need to get more data that includes the years 2013-2021.
  - More data can help me improve this model.
  - Need to focus on players that have a:
    - High On-Base Percentage
    - High Batting Average
    - High Slugging Percentage