Milestones

- Pre-Processing Data 11/21
  - Feature selection
  - Script for combining all CSV's into one that is useful for models
    - Cutoff for certain games played
    - Exclude pitchers for project
    - Exclude 2020 for COVID (less games played)
- Building Model(s) 11/30
  - More pre-processing?
    - Scaling
    - Standardization
    - Transformations
    - Pre-pruning
  - Build Classifier Models
    - PCA - Andrew
    - Decision Trees
    - Ensemble Methods
      - Majority Vote
      - Bagging
      - Boosting
      - Random Forests
      - Stacking
    - K Nearest Neighbors
  - Model Tuning and Evaluation
    - Post-pruning
    - Check for overfitting or underfitting
    - Cross-validation
      - Hyperparameter tuning
- Report Writeup 12/3

Features
- Games played (G)
- Runs (R)
- Hits (H)
- Home runs (HR)
- Batting Average (BA)
- Slugging Percentage (SLG)
- RBI (RBI)
- Stolen Bases (SB)
- Win Probability Added (WPA)**
- Runs Scored % (RS%)**
- Strikeouts (SO)
- Hard Hit %**

- OPS
- Jersey Sales