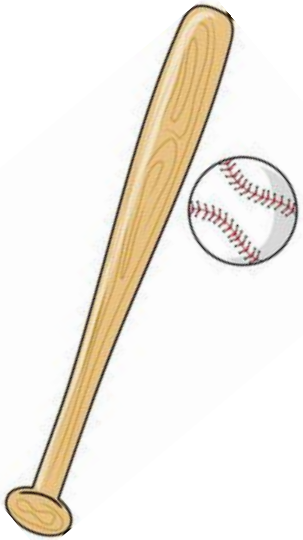
DATABASEBALL

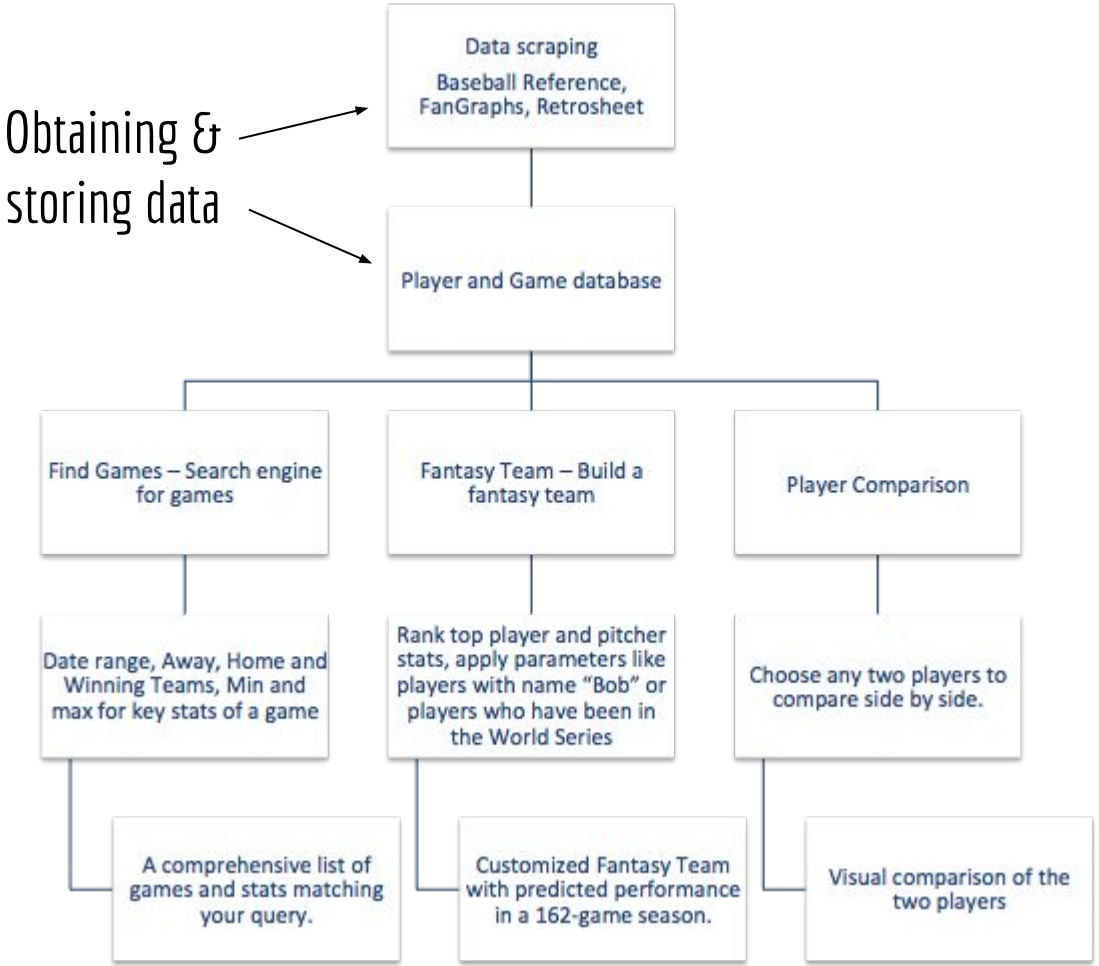


Team JBT: Thomas Dunn, Jessica Lu
& Brianna Pinder

Project Objectives

- ★ Create a search engine for baseball games
- ★ Create fantasy teams with players across eras
- ★ Create a comparison tool for players

Approach


Obtaining & storing data
Data scraping
Baseball Reference, FanGraphs, Retrosheet
Player and Game database
Key Functions
Find Games – Search engine for games
Fantasy Team – Build a fantasy team
Player Comparison
User inputs
Date range, Away, Home and Winning Teams, Min and max for key stats of a game
Rank top player and pitcher stats, apply parameters like players with name "Bob" or players who have been in the World Series
Choose any two players to compare side by side.
Results
A comprehensive list of games and stats matching your query.
Customized Fantasy Team with predicted performance in a 162-game season.
Visual comparison of the two players

Data Scraping - player data

Scraped by first letter of last name from baseball-reference and fangraphs

Made dictionaries of all data by first letter of last name

Parsed out data necessary for each table into a csv file

Combined the letter csv files to make one csv file for each table

Converted each csv file into a sqlite table in the all_players database

Data Scraping - game data

Downloaded all game files from retrosheet.com

Looped through text files for regular seaon games and wrote the necessary data into a csv file

Looped through text files for post seaon games and wrote the necessary data into a csv file

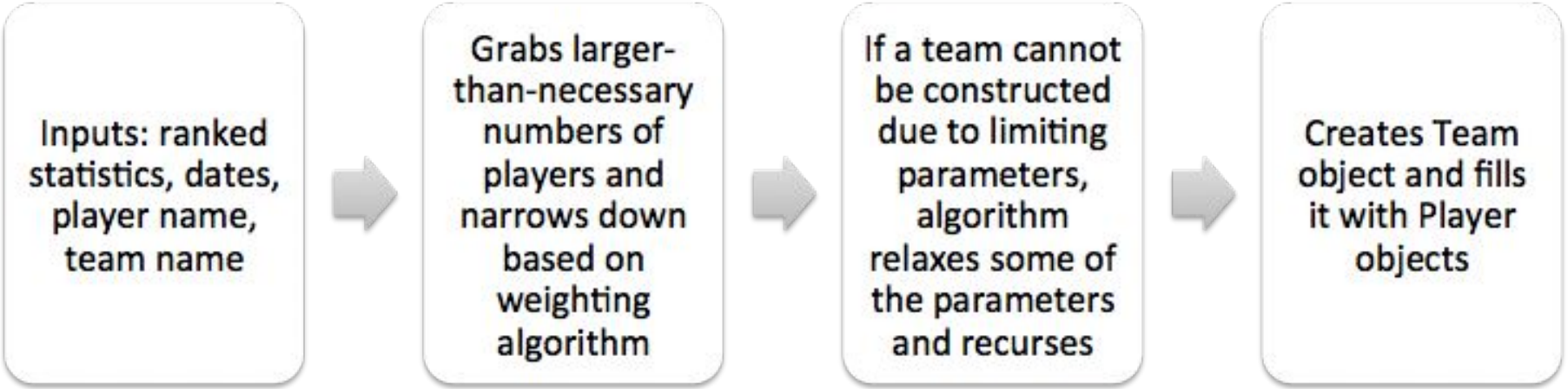
Combined the two csv files and converted that file into a sqlite table in the all_games database

Website views

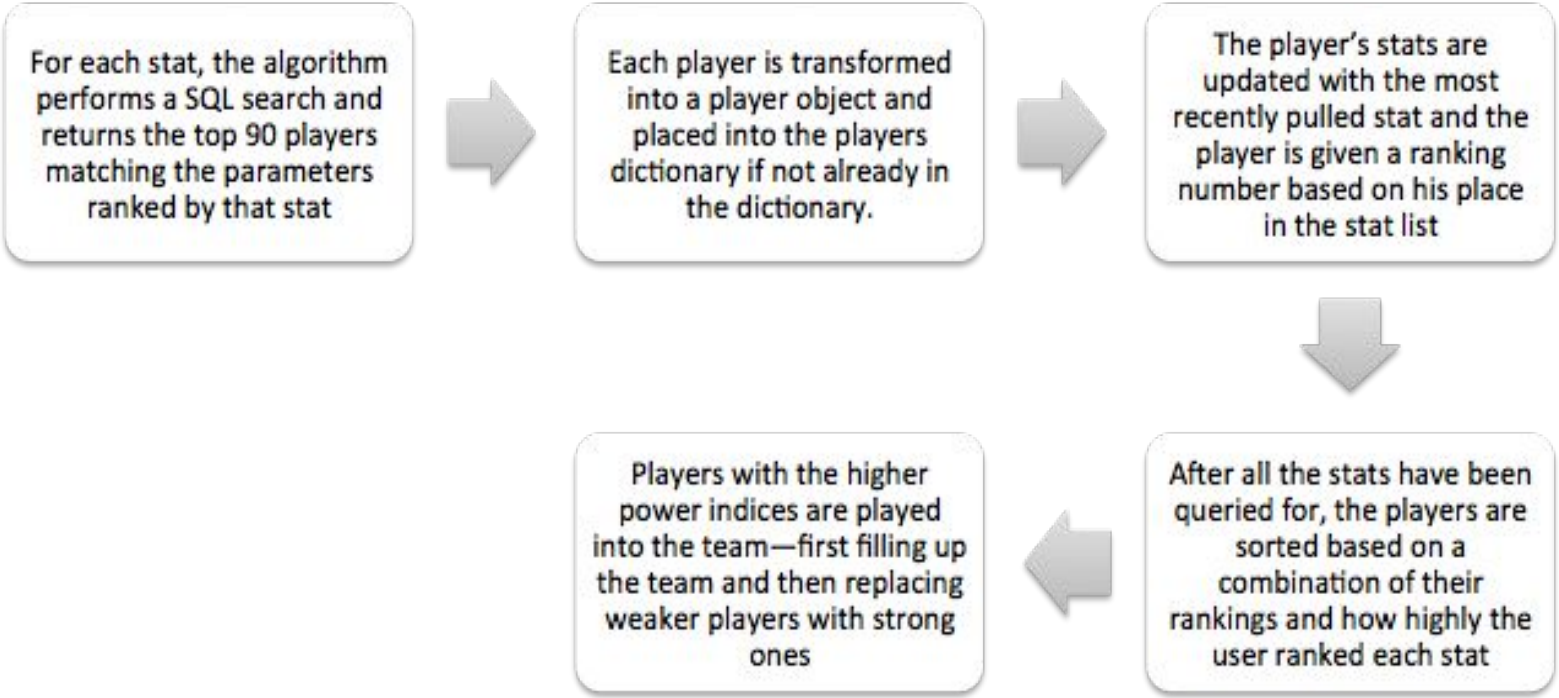
Our website is built with Django and has 3 main views:

- Find Games
  - Parameterized SQL query of ‘all_games’ database for essential game facts
- Fantasy
  - Statistics selection, unique parameters, customized team & predicted performance over the season
- Players
  - Matplotlib generates visual comparison of players

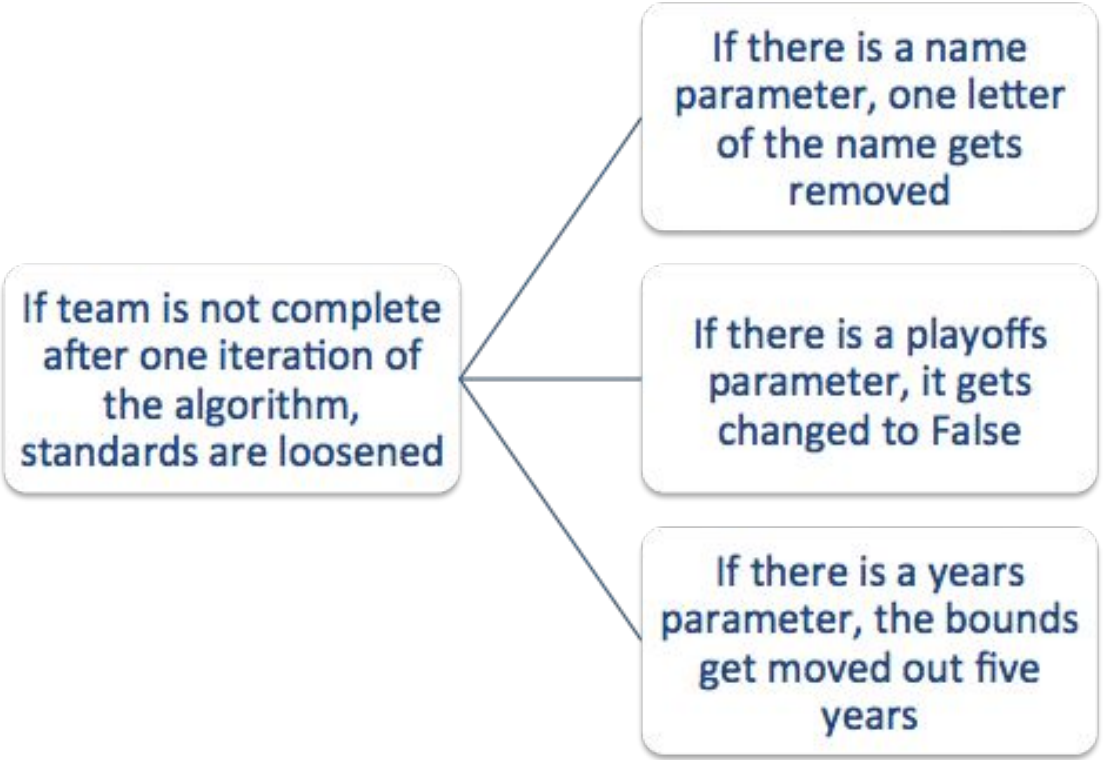
Fantasy Teams--Algorithm



Fantasy Teams--Weighting Algorithm



Fantasy Teams--Team Completion



If name parameter and playoffs parameter cannot be changed, the function goes through its accumulated player list and see if any more players can be added.

Fantasy Teams--Stat Selection and Calculations

- The stats that we selected were chosen to represent two conflicting schools of thought in baseball: old-school style stats that used to be the measure for a player's skill such as batting average and ERA, and sabermetric stats created in the last 20 years that attempt to more fully capture a player's value
- Wins above replacement, FIP, and WRC+ in particular are classic statistics finely tuned to represent a player's full value
- A team's winning percentage is calculated by applying its wins above replacement to a 0-win team's benchmark winning percentage (32%)
- Runs per game are calculated via a formula based on WRC

Challenges & Surprises

- ★ Dealing with inconsistencies on web pages while scraping data
- ★ Working with Django - formatting pages with HTML and CSS, sqlite database connection, integration with Python
- ★ Figuring out how to rank the players based on their proficiency in multiple statistics
- ★ Preventing the team from double-adding players, and generally making sure the best players actually ended up on the team with all positions filled