BBall Analytics

by Alex Xu, Calder Lund, Dhanish Kakkar, Tianchang Zhang

Introduction:

BBall Analytics is a basketball data analyzing tool. Basketball is a sport that lends itself to record a lot of data. Each player has metrics for their shooting accuracy from various locations on the floor, for their passing abilities, their defensive statistics, etcetera. BBall Analytics provides website access for easy queries of player and team statistics from 1937 to 2017. In addition to querying "raw" data, BBall Analytics provides advanced analytics tools to measure a players contributions with many well known basketball metrics (see TS%, PER, etc.).

Who will use this product?

This tool can be very useful for NBA teams to identify the types of players that add the most value to a team. Sports analysts who do reporting on players will want easy ways to bring up interesting facts about players and teams. Our program will address this, see below. In addition, anyone interested in sport betting and fantasy sports will be very interested in BBall Analytics' functionalities as it can provide a competitive advantage over others they are competing against.

How will the user interact with the product?

The most straightforward (albeit complicated to implement) features provided by BBall Analytics will allow the user to view standings of teams in a given year. Then, the user can select players in the given time period by clicking on teams. Each player will display stats.

East			W	L
MIL (1)	F	$	43	7
TOR (2)	F	$	36	14
BOS (3)	F	$	34	15
MIA (3)	F	$	34	15
IND (5)	F	$	31	19
PHI (6)	F	$	31	20
BRK (7)	F	$	22	27
ORL (8)	F	$	22	28
CHI (9)	F	$	19	33
WAS (10)	F	$	17	32
DET (11)	F	$	18	34
CHO (12)	F	$	16	35
NYK (13)	F	$	15	36
CLE (14)	F	$	13	38
ATL (14)	F	$	13	38

→

Player	Pos	Ht	Wt	Birth Date
Terence Davis	SG	6-4	205	May 16, 1997
OG Anunoby	SF	6-7	232	July 17, 1997
Chris Boucher	PF	6-9	200	January 11, 1993
Fred VanVleet	PG	6-1	195	February 25, 1994
Serge Ibaka	C	7-0	235	September 18, 1989
Pascal Siakam	PF	6-9	230	April 2, 1994
Kyle Lowry	PG	6-0	196	March 25, 1986

↓

Season	Age	Tm	Lg	Pos	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%
2016-17	22	TOR	NBA	PF	55	38	15.6	1.9	3.7	.502	0.0	0.1	.143	1.9	3.6	.515
2017-18	23	TOR	NBA	PF	81	5	20.7	3.1	6.1	.508	0.4	1.6	.220	2.8	4.5	.612
2018-19	24	TOR	NBA	PF	80	79	31.9	6.5	11.8	.549	1.0	2.7	.369	5.5	9.1	.602
2019-20 ★	25	TOR	NBA	PF	39	39	34.8	8.8	19.1	.462	2.2	5.9	.367	6.6	13.2	.504
Career			NBA		255	161	25.3	4.8	9.4	.509	0.8	2.3	.332	4.0	7.1	.567

. . .

Another notable way the user can interact with our database is by filtering top player by category. For example, simple queries may involve returning all the Shooting Guards from 2000 to 2010 who shot over 40% from beyond the 3-point line or returning the top 5 players sorted by their assists per game metric. A more complicated query may consist of the rebound rate a team consisting of the best starting lineup (one from each position) of all time according to their Player Efficiency rating (PER is a calculated value based on many attributes).

Next, as mentioned in the introduction, we will include common advanced metrics. These metrics are not found in the data, but rather are calculated using the data we have. Advanced metrics are very helpful is interpreting most given data for each player and team. Some advanced metrics include, PER, TS%, offensive rating, defensive rating, rebound rate, game score, etc.

Sports betters and fantasy sports players need to keep track of individual players and teams to make predictions, our software will allow users to maintain accounts that track their favourite teams and players. In addition, they will be able to assemble their own teams from preexisting players. This functionality can then be combined with other features.

The most ambitious feature we would like to implement a predictor that will determine odds the one team wins over another team. This will be based on the teams prior performance against similar teams and how their players have performed on a given season. This will tie together almost all of our data from each table and then feed that into a simple machine learning model (not the focus). For now, accuracy is not our primary focus. It will be to have functionality to make predictions while using all relevant data.

Lastly, we will have a feature that allows the user to see all time leaders. For example, the Raptor's all time assist leader is Kyle Lowrey and the NBA's all time most points is set by Wilt Chamberlain where he scored 100 points in one game. This feature is targeted at more of a sports analyst or TV sports talk show host. They require easy queries to compare current players to well-known players from the past.

Updating:

There will be a specific user with administrative rights who can create players and teams. This will allow for many updating scripts to be executed.

Datasets:

We have collected data from two Kaggle datasets. The first dataset contains individual player basketball statistics from 1950 to 2017. The second dataset contains stats on each team and each player from 1937 to 2012. As mentioned above, this data will be used to do in depth analysis of player statistics in attempt to find correlation between player tendencies and team winning. We plan on populating our database by using Python to transfer relevant data from CSV format to SQL, this will probably be done with Pandas, a data frame tool in Python.

Sample Dataset:

A sample dataset will be created using all players from the Toronto Raptors and Boston Celtics in the the years 1999, 2003, 2014 and 2017. This will contain about 150 players on two teams over time. There will be overlap where certain players play on the same team for multiple years.

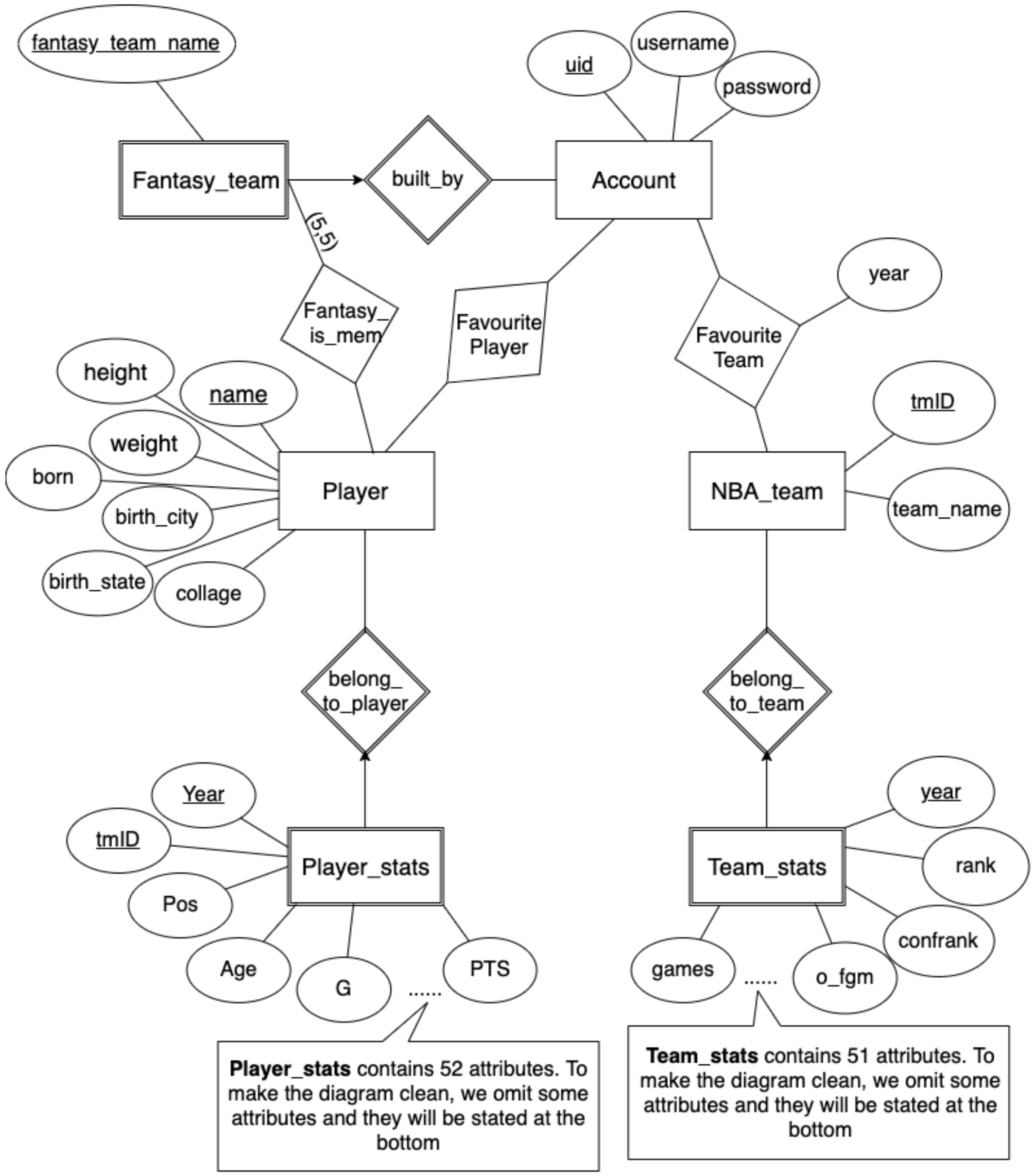
System Support:

The backend code that will call the PostgreSQL queries will be implemented Python. Django will be used to connect the Python functionality to our front end Javascript and HTML website design.

Database Schema

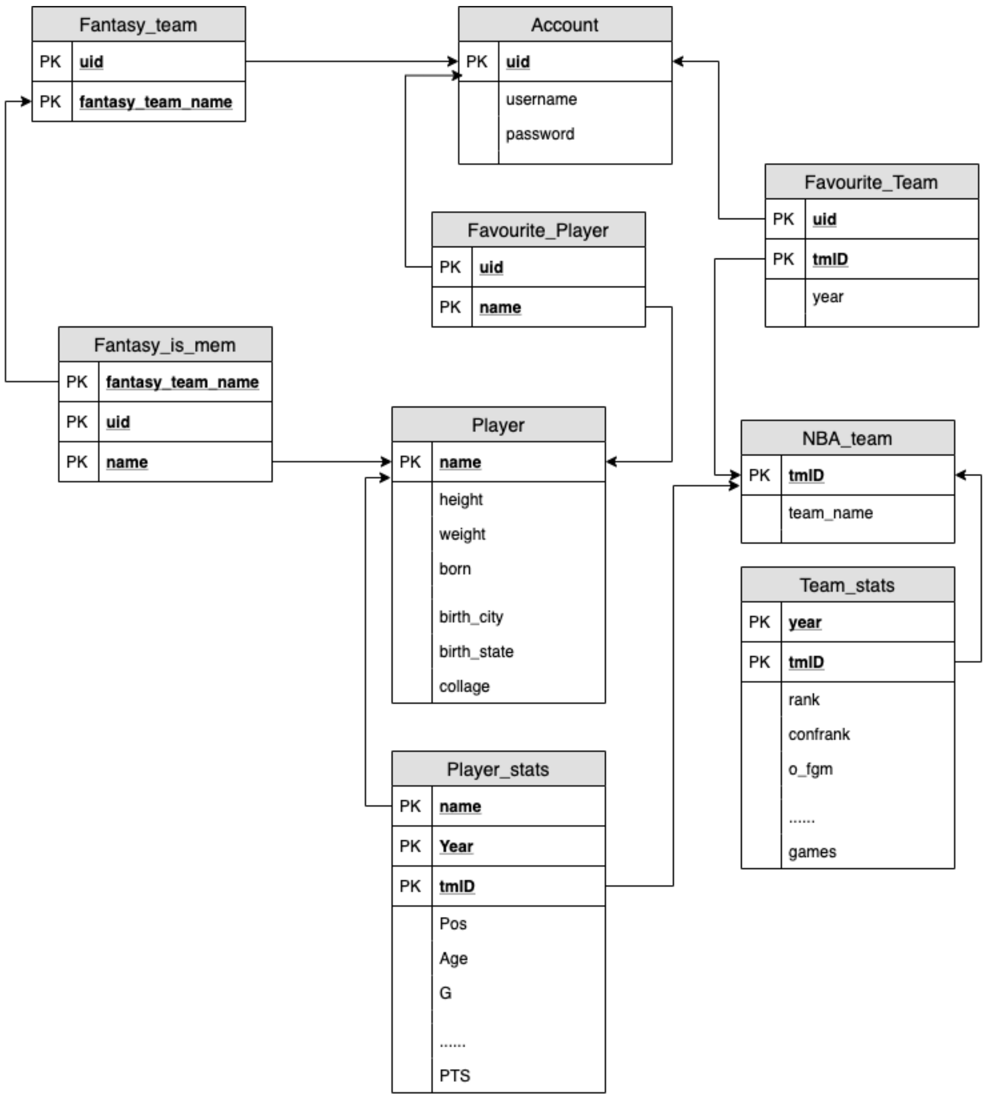
Assumption:

- All players in our data have a unique name ('Firstname + Lastname').
- In the same year, a player can switch between different teams but does not change his position within the same team.
- Each NBA team has a unique team ID.
- **account.uid** is unique.
- Users can have duplicated username.
- Each account cannot build multiple fantasy teams with the same name but different accounts can.
- User built fantasy teams must have a fixed team size of 5.
- User cannot add the same team to favourite-team list twice (even with different years).



Player_stats Attributes: Year, tmID, Pos, Age, G, GS, MP, PER, TS%, 3PAr, FTr, ORB%, DRB%, TRB%, AST%, STL%, BLK%, TOV%, USG%, blanl, OWS, DWS, WS, WS/48, blank2, OBPM, DBPM, BPM, VORP, FG, FGA, FG%, 3P, 3PA, 3P%, 2P, 2PA, 2P%, eFG%, FT, FTA, FT%, ORB, DRB, TRB, AST, STL, BLK, TOV, PF, PTS

Team_stats Attributes: year, rank, confRank, playoff, name, o_fgm, o_fga, o_ftm, o_fta, o_3pm, o_3pa, o_oreb, o_dreb, o_reb, o_asts, o_pf, o_stl, o_to, o_blk, o_pts, d_fgm, d_fga, d_ftm, d_fta, d_3pm, d_3pa, d_oreb, d_dreb, d_reb, d_asts, d_pf, d_stl, d_to, d_blk, d_pts, o_tmRebound, d_tmRebound, homeWon, homeLost, awayWon, awayLost, neutWon, neutLoss, confWon, confLoss, divWon, divLoss, pace, won, lost, games


Fantasy_team
PK uid
PK fantasy_team_name
Account
PK uid
username
password
Favourite_Team
PK uid
PK tmID
year
Favourite_Player
PK uid
PK name
Fantasy_is_mem
PK fantasy_team_name
PK uid
PK name
Player
PK name
height
weight
born
birth_city
birth_state
collage
NBA_team
PK tmID
team_name
Team_stats
PK year
PK tmID
rank
confrank
o_fgm
......
games
Player_stats
PK name
PK Year
PK tmID
Pos
Age
G
......
PTS