

# Final Project Data Checkpoint

1. My github link: [https://github.com/Qingyuan-Liu/507Final\\_project](https://github.com/Qingyuan-Liu/507Final_project)

## 2. Data Source and Database:

- First, I get all the [30 active franchises](https://www.basketball-reference.com/teams/) links from the <a> tag in this page: <https://www.basketball-reference.com/teams/>.

And also I get all the teams' information like team name, league, games, etc in this page and load them to my database nba.sqlite. (Here is my first table: Teams)

The screenshot shows the 'NBA & ABA Team Index' page from basketball-reference.com. It displays a table of teams with columns for League, Year, Games, Wins, Losses, and Championships. The table is filtered for the 1946-74 season. The HTML source code on the right shows the structure of the table, including the 'table' element and the 'tbody' with 'tr' elements for each team.












League	Year	G	W	L	Champ
NBA	1946-74	71	5619	2766	2853
NBA	1946-74	52	4201	2068	2133
NBA	1946-74	13	1005	553	452
NBA	1946-74	4	281	91	190
NBA	1946-74	2	132	54	78
NBA	1946-74	1	82	22	60
NBA	1946-74	8	666	338	328
NBA	1946-74	1	78	36	42
NBA	1946-74	30	2395	1050	1345
NBA	1946-74	6	475	215	260
NBA	1946-74	10	804	293	511
NBA	1946-74	14	1116	542	574
NBA	1946-74	54	4362	2227	2135
NBA	1946-74	50	4035	1867	2168
NBA	1946-74	40	3217	1612	1605
NBA	1946-74	53	4287	2159	2128
NBA	1946-74	44	3543	1746	1797
NBA	1946-74	2	168	125	43
NBA	1946-74	7	576	288	288
NBA	1946-74	72	5680	2753	2927

- Second, I followed the link to the teams' detail page (eg: <https://www.basketball-reference.com/teams/ATL/>) and it include the teams' information in all seasons. So I got the latest season's link of this team through the <a> tag, and scrape that following page.

Please note that players may not be in the uniform of the correct team in these images.

[illegible]

- | No. | Player             | Pos | Ht   | Wt  | Birth Date         | Exp   | College   |
|-----|--------------------|-----|------|-----|--------------------|---|---|
| 12  | De'Andre Hunter    | SF  | 6'7  | 225 | December 2, 1997   |  | <a href="#">R Virginia</a>                            |
| 11  | Tyke Young         | PG  | 6'1  | 180 | September 19, 1998 |  | <a href="#">1 Oklahoma</a>                            |
| 15  | Vince Carter       | PF  | 6'6  | 220 | January 26, 1977   |  | <a href="#">21 UNC</a>                                |
| 22  | Cam Reddish        | SF  | 6'8  | 208 | September 1, 1999  |  | <a href="#">R Duke</a>                                |
| 3   | Kevin Huertel      | SG  | 6'7  | 190 | August 27, 1998    |  | <a href="#">1 Maryland</a>                            |
| 24  | Ramon Fernandez    | C   | 6'23 | 234 | August 15, 1998    |  | <a href="#">R Maryland</a>                            |
| 30  | Damian Jones       | C   | 6'11 | 245 | June 30, 1995      |  | <a href="#">3 Vanderbilt</a>                          |
| 3   | DeAndre' Bembry    | SG  | 6'5  | 210 | July 4, 1994       |  | <a href="#">3 Saint Joseph's</a>                      |
| 20  | John Collins       | PF  | 6'9  | 235 | September 23, 1997 |  | <a href="#">2 Wake Forest</a>                         |
| 2   | Brandon Goodwin    | PG  | 6'10 | 180 | October 2, 1995    |  | <a href="#">1 Central Florida, Florida Gulf Coast</a> |
| 2   | Jeff Teague        | PG  | 6'5  | 195 | June 10, 1988      |  | <a href="#">10 Wake Forest</a>                        |
| 2   | Trevon Graham      | SG  | 6'23 | 193 | October 28, 1993   |  | <a href="#">3 Virginia Commonwealth</a>               |
| 14  | Dewayne Dedmon     | C   | 7'05 | 245 | October 12, 1989   |  | <a href="#">6 USC</a>                                 |
| 4   | Charlie Brown (TW) | SG  | 6'9  | 199 | February 2, 1997   |  | <a href="#">R Saint Joseph's</a>                      |
|     | Clint Capela       | C   | 6'10 | 240 | May 18, 1994       |  |   |
|     | Skyl Labissiere    | C   | 6'10 | 216 | March 18, 1996     |  | <a href="#">3 Kentucky</a>                            |

Tables (3)		
Players		
 Id	INTEGER	"Id" INTEGER PRIMARY KEY AUTOINCR
 Team_id	INTEGER	"Team_id" INTEGER NOT NULL
 Player_id	TEXT	"Player_id" TEXT NOT NULL
 Intro_url	TEXT	"Intro_url" TEXT
 Name	TEXT	"Name" TEXT NOT NULL
 Position	TEXT	"Position" TEXT
 Height	TEXT	"Height" TEXT
 Weight	INTEGER	"Weight" INTEGER
 Birthday	TEXT	"Birthday" TEXT
 Year_of_exp...	TEXT	"Year_of_experience" TEXT
 College	TEXT	"College" TEXT

- \*About the two tables in my database: The Team\_id in my Players table is a **foreign key** which reference to the Id in my Teams table. We could query like:  
'SELECT avg(Players.Height)  
FROM Players,Teams  
WHERE Teams.Id=Players.Team\_id  
GROUP by Players.Team\_id'

New Database
Open Database
Write Changes
Revert Changes
Open Project
Save Project
Attach Database
Close Database

Database Structure
Browse Data
Edit Pragma
Execute SQL

Table: Players

New Record
Delete Record

Id	Team_id	Player_id	Intro_url	Name	Position	Height	Weight	Birthday	Year_of_experience
1	1	12	https://www.ba...	De'Andre Hunter	SF	6-7	225	December 2, 19...	8
2	2	11	https://www.ba...	Tray Young	PG	6-1	180	September 19, ...	1
3	3	15	https://www.ba...	Vinuc Carter	SF	6-6	220	January 26, 1977	21
4	4	22	https://www.ba...	Cam Reddick	SF	6-8	208	September 1, 1...	8
5	5	1	https://www.ba...	Kevin Huarter	SG	6-7	190	August 27, 1998	1
6	6	1	https://www.ba...	Bruno Fernando	C	6-9	233	August 15, 1998	8
7	7	1	https://www.ba...	Dennan Jones	C	6-11	245	June 30, 1995	3
8	8	95	https://www.ba...	De'Andre Bembry	SG	6-5	210	July 4, 1994	3
9	9	1	https://www.ba...	John Collins	PF	6-9	235	September 23, ...	2
10	10	1	https://www.ba...	Brandon Goodwin	PG	6-0	180	October 2, 1995	1
11	11	1	https://www.ba...	JHF Teague	PG	6-3	195	June 10, 1988	10
12	12	1	https://www.ba...	Trevon Graham	SG	6-5	223	October 28, 1993	1
13	13	1	https://www.ba...	Dejounte Davidson	C	7-0	245	August 12, 1980	6
14	14	1	https://www.ba...	Charlie Brown	SG	6-6	199	February 2, 1997	8
15	15	2	https://www.ba...	Brad Wanamaker	PF	6-3	210	July 25, 1989	1
16	16	2	https://www.ba...	Grant Williams	PF	6-6	236	November 30, 1...	8
17	17	37	https://www.ba...	Seini Ojelu	PF	6-6	235	December 5, 19...	2
18	18	2	https://www.ba...	Jayson Tatum	PF	6-8	204	March 3, 1998	2
19	19	2	https://www.ba...	Daniel Theis	C	6-8	215	April 4, 1992	2
20	20	2	https://www.ba...	Marcus Smart	SG	6-3	227	March 6, 1994	5
21	21	2	https://www.ba...	Ers Kanner	C	6-10	262	May 26, 1992	8
22	22	8	https://www.ba...	Keanu Walker	PG	6-0	172	May 8, 1990	8
23	23	2	https://www.ba...	Jaylen Brown	SG	6-7	223	October 24, 1996	3
24	24	2	https://www.ba...	Gordon Hayward	SF	6-7	207	March 23, 1990	9
25	25	2	https://www.ba...	Javonte Green	SF	6-4	220	July 23, 1993	8
26	26	4	https://www.ba...	Carson Galloway	SG	6-11	200	March 22, 1988	8
27	27	2	https://www.ba...	Ramone Langford	SG	6-4	216	October 25, 1999	8
28	28	2	https://www.ba...	Robert Williams	C	7-0	237	October 17, 1997	1
29	29	2	https://www.ba...	Vincent Poirier	C	6-8	235	October 17, 1993	8

1 - 31 of 472

Go to: 1

Edit Database Cell

Mode: Text
Import
Export
Set as NULL

3

Type of data currently in cell: Text / Numeric
1 char(s)

DB Schema

Tables (3)

Players

CREATE TABLE "Players" ("Id" INT "Id" INTEGER PRIMARY KEY AUTO "Intro\_url" TEXT NOT NULL "Player\_id" TEXT NOT NULL "Intro\_url" TEXT "Name" TEXT NOT NULL "Position" TEXT "Height" TEXT "Weight" INTEGER "Birthday" TEXT "Year\_of\_experience" TEXT "College" TEXT

Teams

CREATE TABLE "Teams" ("Id" INT "Id" INTEGER PRIMARY KEY AUTO "Team\_name" TEXT NOT NULL "Intro\_url" TEXT

SQL Log
Plot
DB Schema

- About caching: I stored the url and the html content into the cache.json file and when I use cache, it will print 'Using cache' and when the unique\_key is not in cache, I will fetch it and store it in cache and then it will print 'Fetching'.

[illegible]

Here the first picture is when I delete the cache.json file and run the program, and the second picture is when the second time I run the program after having the cache.json file.





## Welcome to the NBA Team Stats Browser

### Team Stats

**Sort by:**

- ☐ Win-loss Percentage  
☒ Championship

**Sort direction:**

- ☒ High to Low  
☐ Low to High

**Check what you want to see:**

- ☒ League
- ☒ Since
- ☒ Until
- ☒ Year
- ☐ Games
- ☐ Wins
- ☐ Losses
- ☐ Win\_loss\_percentage
- ☐ Playoffs
- ☐ Divisions
- ☐ Conference\_champion
- ☒ Championship

Get teams

### Team Stats

Team name	League	Until	Championship
Boston Celtics	NBA/BAA	2019-20	17
Los Angeles Lakers	NBA/BAA	2019-20	16
Chicago Bulls	NBA	2019-20	6
Golden State Warriors	NBA/BAA	2019-20	6
San Antonio Spurs	NBA/ABA	2019-20	5
Detroit Pistons	NBA/BAA	2019-20	3
Indiana Pacers	NBA/ABA	2019-20	3
Miami Heat	NBA	2019-20	3
Philadelphia 76ers	NBA	2019-20	3
Brooklyn Nets	NBA/ABA	2019-20	2
Houston Rockets	NBA	2019-20	2
New York Knicks	NBA/BAA	2019-20	2
Atlanta Hawks	NBA	2019-20	1
Cleveland Cavaliers	NBA	2019-20	1
Dallas Mavericks	NBA	2019-20	1

Then, they could click the 'Get Heights' button to get the average heights for the five positions in each team, and this information would be shown in Plotly.

### Heights on Court

Click the button to get the average height for each position:

Get Heights

### Player Info

Please select a team:

Player id:

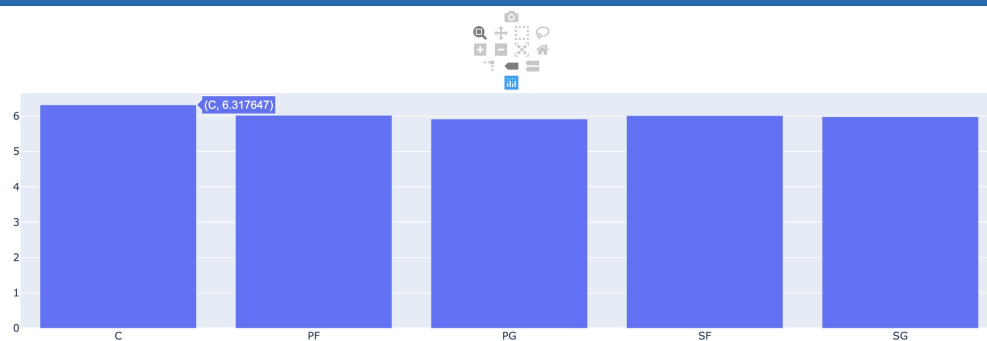
Get Player

### Best Colleges

Click the button to see the top 10 universities with the most drafted players:

Get Colleges

## Heights on Court



There will be 6 different webpage to show the different results, users could access that by clicking the button and enter the required input. If there is no query result that the user want to see, I will show them the error message page (err\_msg.html) and let them go back to the home page.

## Welcome to the NBA Team Stats Browser

**There is no result! Please go back and search again.**

[Home page](#)