CIS 4400
Data Warehousing Project

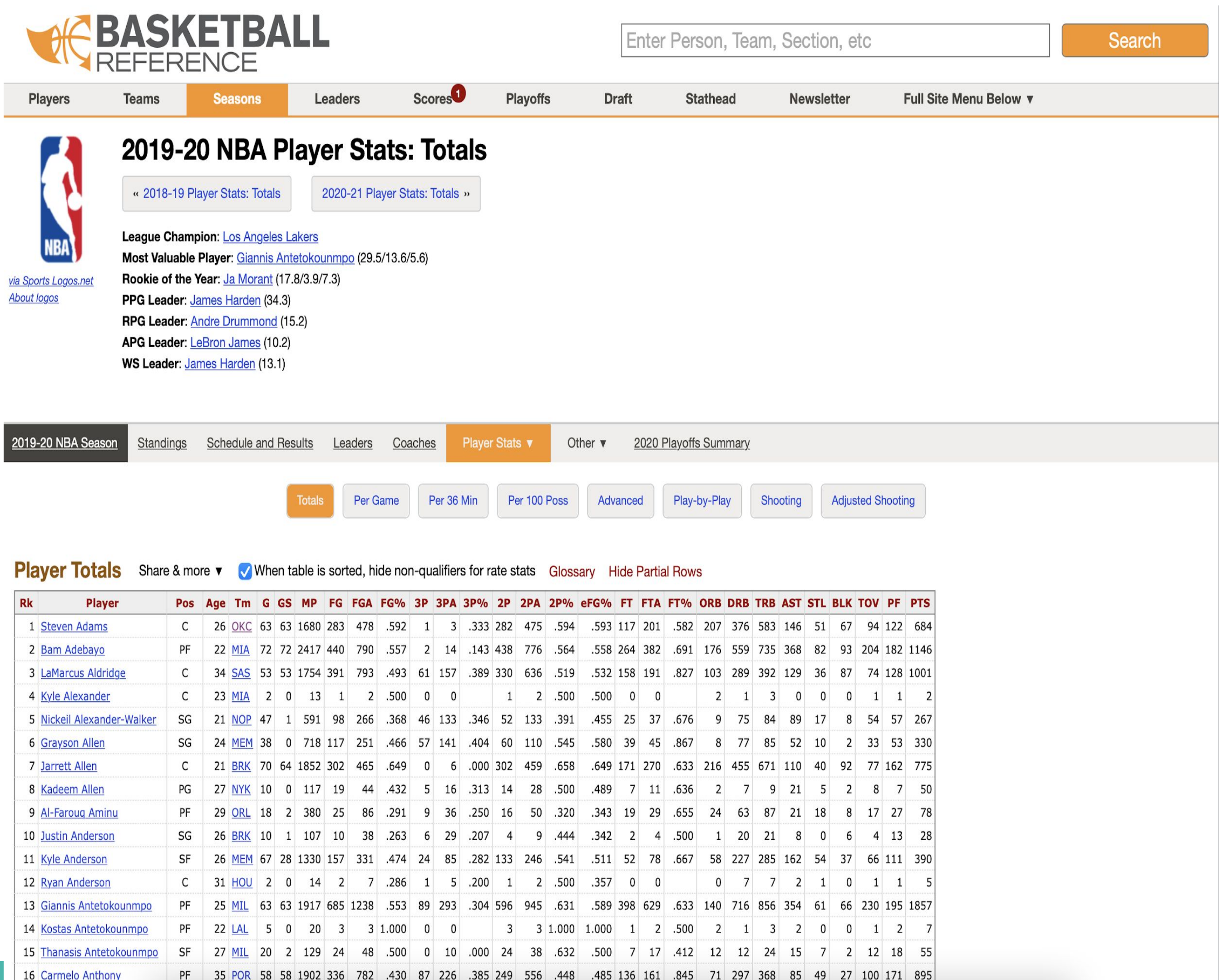
Basketball Reference Analysis

Luyao Chu / Jessie Chen/ Elai Shalev

Source

Basketball reference

Dataset



2019-20 NBA Player Stats: Totals

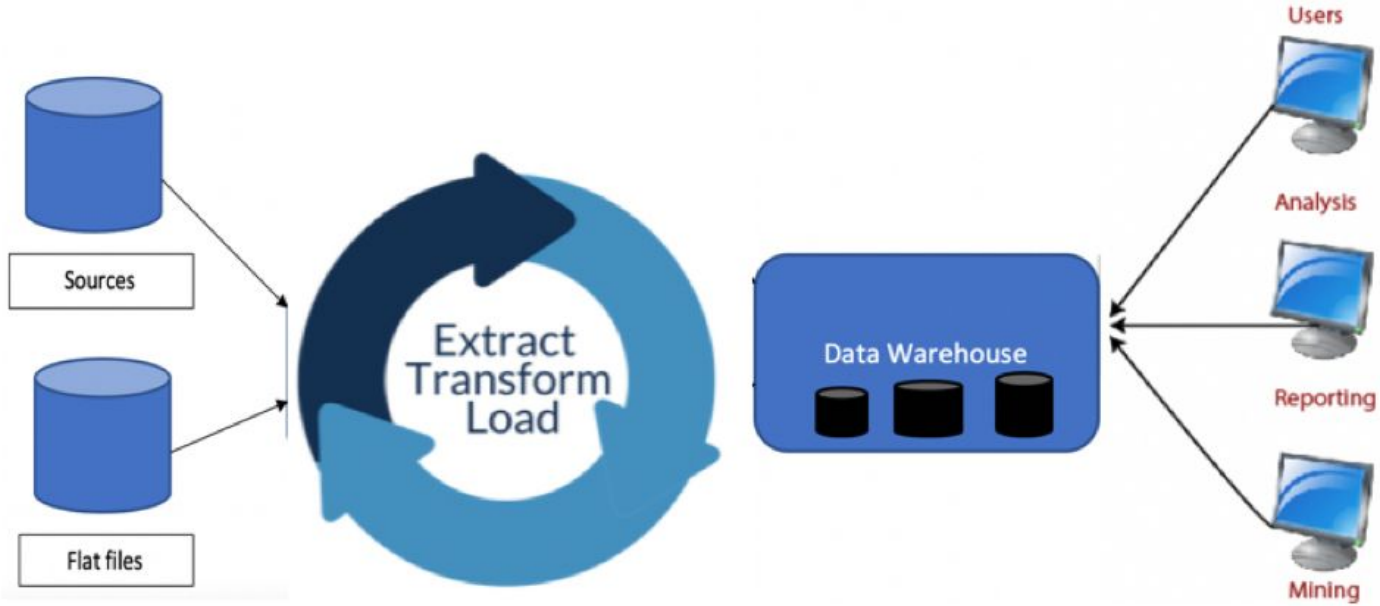
League Champion: Los Angeles Lakers
Most Valuable Player: Giannis Antetokounmpo (29.5/13.6/5.6)
Rookie of the Year: Ja Morant (17.8/3.9/7.3)
PPG Leader: James Harden (34.3)
RPG Leader: Andre Drummond (15.2)
APG Leader: LeBron James (10.2)
WS Leader: James Harden (13.1)

Player Totals

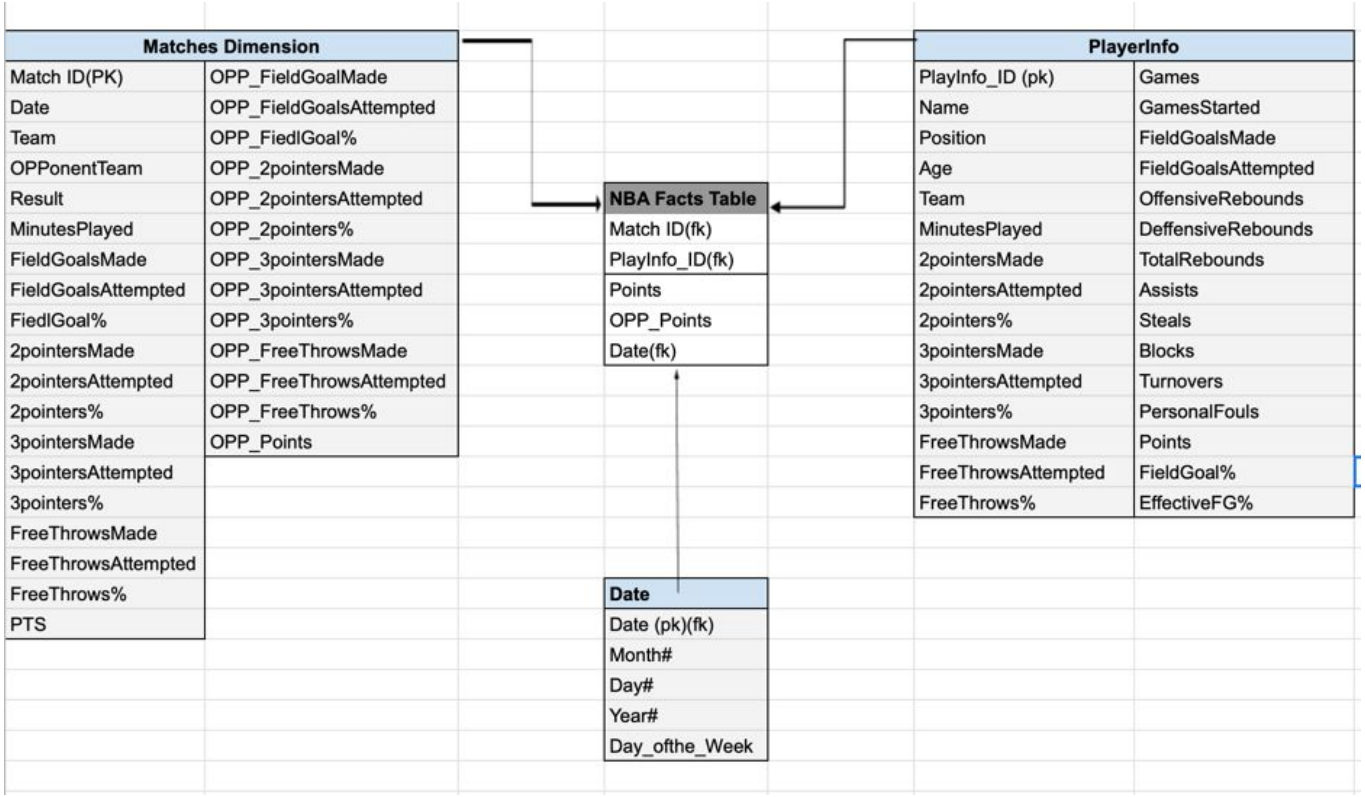
☑ When table is sorted, hide non-qualifiers for rate stats

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Steven Adams	C	26	OKC	63	63	1680	283	478	.592	1	3	.333	282	475	.594	.593	117	201	.582	207	376	583	146	51	67	94	122	684
2	Bam Adebayo	PF	22	MIA	72	72	2417	440	790	.557	2	14	.143	438	776	.564	.558	264	382	.691	176	559	735	368	82	93	204	182	1146
3	LaMarcus Aldridge	C	34	SAS	53	53	1754	391	793	.493	61	157	.389	330	636	.519	.532	158	191	.827	103	289	392	129	36	87	74	128	1001
4	Kyle Alexander	C	23	MIA	2	0	13	1	2	.500	0	0		1	2	.500	.500	0	0		2	1	3	0	0	0	1	1	2
5	Nickeil Alexander-Walker	SG	21	NOP	47	1	591	98	266	.368	46	133	.346	52	133	.391	.455	25	37	.676	9	75	84	89	17	8	54	57	267
6	Grayson Allen	SG	24	MEM	38	0	718	117	251	.466	57	141	.404	60	110	.545	.580	39	45	.867	8	77	85	52	10	2	33	53	330
7	Jarrett Allen	C	21	BRK	70	64	1852	302	465	.649	0	6	.000	302	459	.658	.649	171	270	.633	216	455	671	110	40	92	77	162	775
8	Kadeem Allen	PG	27	NYK	10	0	117	19	44	.432	5	16	.313	14	28	.500	.489	7	11	.636	2	7	9	21	5	2	8	7	50
9	Al-Farouq Aminu	PF	29	ORL	18	2	380	25	86	.291	9	36	.250	16	50	.320	.343	19	29	.655	24	63	87	21	18	8	17	27	78
10	Justin Anderson	SG	26	BRK	10	1	107	10	38	.263	6	29	.207	4	9	.444	.342	2	4	.500	1	20	21	8	0	6	4	13	28
11	Kyle Anderson	SF	26	MEM	67	28	1330	157	331	.474	24	85	.282	133	246	.541	.511	52	78	.667	58	227	285	162	54	37	66	111	390
12	Ryan Anderson	C	31	HOU	2	0	14	2	7	.286	1	5	.200	1	2	.500	.357	0	0		0	7	7	2	1	0	1	1	5
13	Giannis Antetokounmpo	PF	25	MIL	63	63	1917	685	1238	.553	89	293	.304	596	945	.631	.589	398	629	.633	140	716	856	354	61	66	230	195	1857
14	Kostas Antetokounmpo	PF	22	LAL	5	0	20	3	3	1.000	0	0		3	3	1.000	1.000	1	2	.500	2	1	3	2	0	0	1	2	7
15	Thanasis Antetokounmpo	SF	27	MIL	20	2	129	24	48	.500	0	10	.000	24	38	.632	.500	7	17	.412	12	12	24	15	7	2	12	18	55
16	Carmelo Anthony	PF	35	POR	58	58	1902	336	782	.430	87	226	.385	249	556	.448	.485	136	161	.845	71	297	368	85	49	27	100	171	895

Architectural Design



Dimensional Model



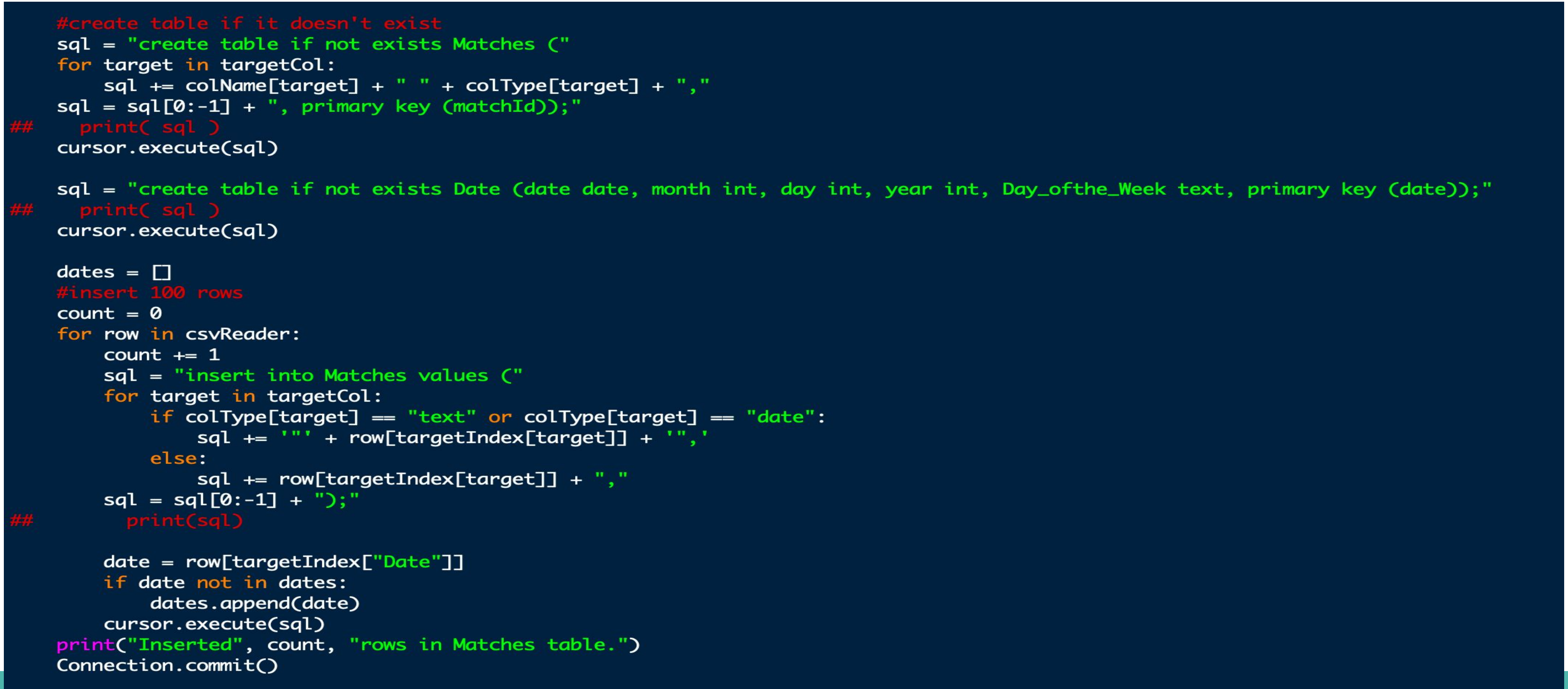
Matches Dimension	
Match ID(PK)	OPP_FieldGoalMade
Date	OPP_FieldGoalsAttempted
Team	OPP_FiedlGoal%
OPPonentTeam	OPP_2pointersMade
Result	OPP_2pointersAttempted
MinutesPlayed	OPP_2pointers%
FieldGoalsMade	OPP_3pointersMade
FieldGoalsAttempted	OPP_3pointersAttempted
FiedlGoal%	OPP_3pointers%
2pointersMade	OPP_FreeThrowsMade
2pointersAttempted	OPP_FreeThrowsAttempted
2pointers%	OPP_FreeThrows%
3pointersMade	OPP_Points
3pointersAttempted	
3pointers%	
FreeThrowsMade	
FreeThrowsAttempted	
FreeThrows%	
PTS	

NBA Facts Table
Match ID(fk)
PlayInfo_ID(fk)
Points
OPP_Points
Date(fk)

Date
Date (pk)(fk)
Month#
Day#
Year#
Day_ofthe_Week

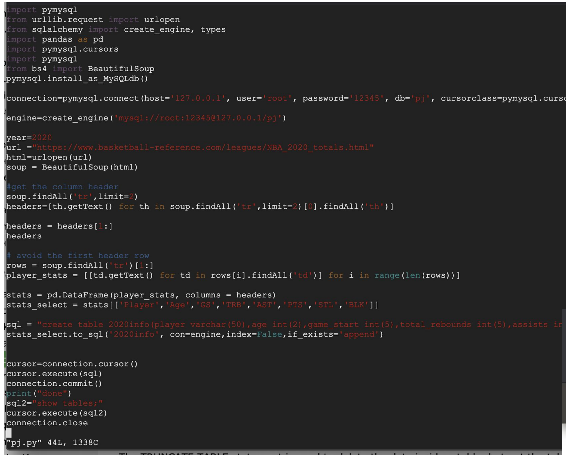
PlayerInfo	
PlayInfo_ID (pk)	Games
Name	GamesStarted
Position	FieldGoalsMade
Age	FieldGoalsAttempted
Team	OffensiveRebounds
MinutesPlayed	DeffensiveRebounds
2pointersMade	TotalRebounds
2pointersAttempted	Assists
2pointers%	Steals
3pointersMade	Blocks
3pointersAttempted	Turnovers
3pointers%	PersonalFouls
FreeThrowsMade	Points
FreeThrowsAttempted	FieldGoal%
FreeThrows%	EffectiveFG%

Matches and Date Table



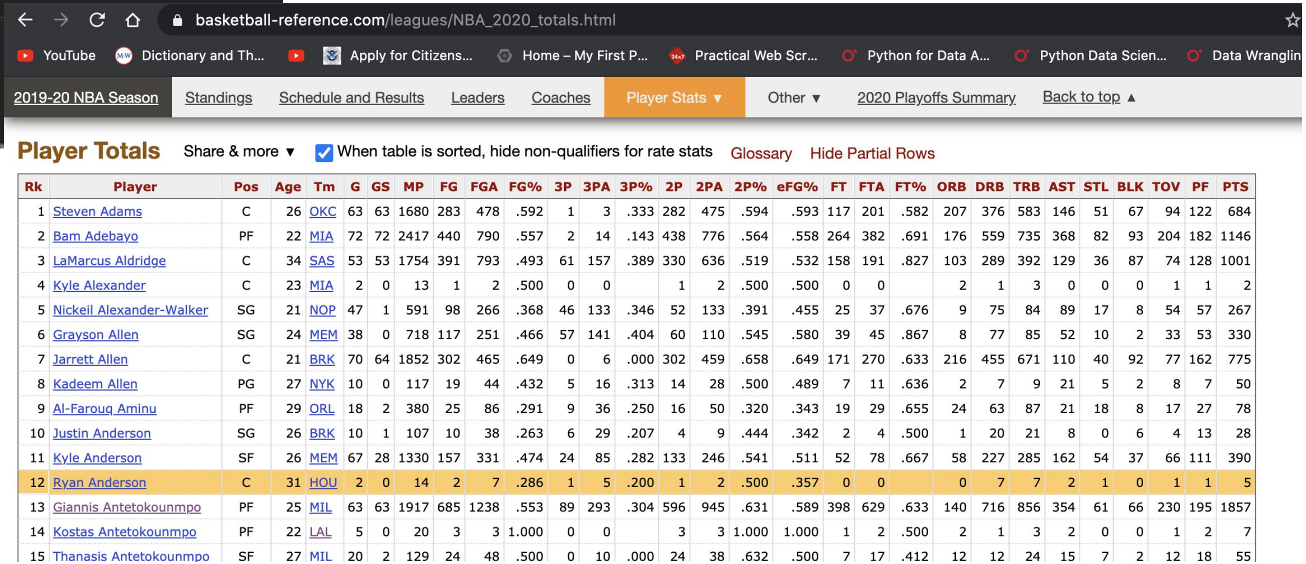
```
    #create table if it doesn't exist
    sql = "create table if not exists Matches ("
    for target in targetCol:
        sql += colName[target] + " " + colType[target] + ","
    sql = sql[0:-1] + ", primary key (matchId));"
##    print( sql )
    cursor.execute(sql)

    sql = "create table if not exists Date (date date, month int, day int, year int, Day_ofthe_Week text, primary key (date));"
##    print( sql )
    cursor.execute(sql)

    dates = []
    #insert 100 rows
    count = 0
    for row in csvReader:
        count += 1
        sql = "insert into Matches values ("
        for target in targetCol:
            if colType[target] == "text" or colType[target] == "date":
                sql += '"' + row[targetIndex[target]] + '",'
            else:
                sql += row[targetIndex[target]] + ","
        sql = sql[0:-1] + ");"
##        print(sql)

        date = row[targetIndex["Date"]]
        if date not in dates:
            dates.append(date)
        cursor.execute(sql)
    print("Inserted", count, "rows in Matches table.")
    Connection.commit()
```

Web Scrap & Aggregate the Data



```
import pymysql
from urllib.request import urlopen
from sqlalchemy import create_engine, types
import pandas as pd
import pymysql.cursors
import pymysql
from bs4 import BeautifulSoup
pymysql.install_as_MySQLdb()

connection=pymysql.connect(host='127.0.0.1', user='root', password='12345', db='pj', cursorclass=pymysql.cursors.DictCursor)

engine=create_engine('mysql://root:12345@127.0.0.1/pj')

year=2020
url ="https://www.basketball-reference.com/leagues/NBA_2020_totals.html"
html=urlopen(url)
soup = BeautifulSoup(html)

#get the column header
soup.findAll('tr',limit=2)
headers=[th.getText() for th in soup.findAll('tr',limit=2)[0].findAll('th')]

headers = headers[1:]
headers

# avoid the first header row
rows = soup.findAll('tr')[1:]
player_stats = [[td.getText() for td in rows[i].findAll('td')] for i in range(len(rows))]

stats = pd.DataFrame(player_stats, columns = headers)
stats_select = stats[['Player','Age','GS','TRB','AST','PTS','STL','BLK']]

sql = "create table 2020info(player varchar(50),age int(2),game_start int(5),total_rebounds int(5),assists in
stats_select.to_sql('2020info', con=engine,index=False,if_exists='append')


cursor=connection.cursor()
cursor.execute(sql)
connection.commit()
print("done")
sql2="show tables;"
cursor.execute(sql2)
connection.close

"pj.py" 44L, 1338C
```

Use python web scraping the player_stats from 2019-2020 season



Player Totals

☑ When table is sorted, hide non-qualifiers for rate stats

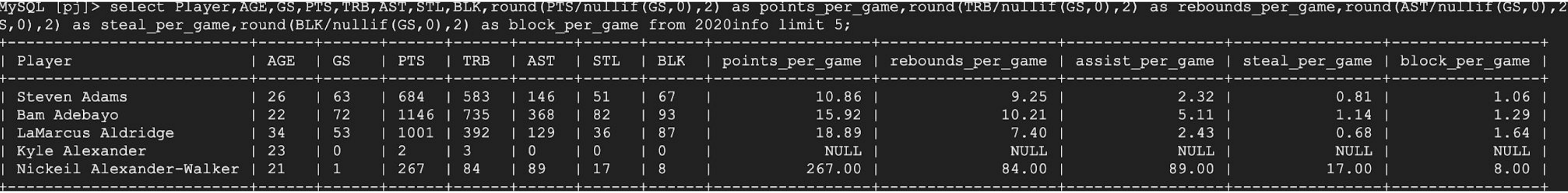
Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Steven Adams	C	26	OKC	63	63	1680	283	478	.592	1	3	.333	282	475	.594	.593	117	201	.582	207	376	583	146	51	67	94	122	684
2	Bam Adebayo	PF	22	MIA	72	72	2417	440	790	.557	2	14	.143	438	776	.564	.558	264	382	.691	176	559	735	368	82	93	204	182	1146
3	LaMarcus Aldridge	C	34	SAS	53	53	1754	391	793	.493	61	157	.389	330	636	.519	.532	158	191	.827	103	289	392	129	36	87	74	128	1001
4	Kyle Alexander	C	23	MIA	2	0	13	1	2	.500	0	0		1	2	.500	.500	0	0		2	1	3	0	0	0	1	1	2
5	Nickeil Alexander-Walker	SG	21	NOP	47	1	591	98	266	.368	46	133	.346	52	133	.391	.455	25	37	.676	9	75	84	89	17	8	54	57	267
6	Grayson Allen	SG	24	MEM	38	0	718	117	251	.466	57	141	.404	60	110	.545	.580	39	45	.867	8	77	85	52	10	2	33	53	330
7	Jarrett Allen	C	21	BRK	70	64	1852	302	465	.649	0	6	.000	302	459	.658	.649	171	270	.633	216	455	671	110	40	92	77	162	775
8	Kadeem Allen	PG	27	NYK	10	0	117	19	44	.432	5	16	.313	14	28	.500	.489	7	11	.636	2	7	9	21	5	2	8	7	50
9	Al-Farouq Aminu	PF	29	ORL	18	2	380	25	86	.291	9	36	.250	16	50	.320	.343	19	29	.655	24	63	87	21	18	8	17	27	78
10	Justin Anderson	SG	26	BRK	10	1	107	10	38	.263	6	29	.207	4	9	.444	.342	2	4	.500	1	20	21	8	0	6	4	13	28
11	Kyle Anderson	SF	26	MEM	67	28	1330	157	331	.474	24	85	.282	133	246	.541	.511	52	78	.667	58	227	285	162	54	37	66	111	390
12	Ryan Anderson	C	31	HOU	2	0	14	2	7	.286	1	5	.200	1	2	.500	.357	0	0		0	7	7	2	1	0	1	1	5
13	Giannis Antetokounmpo	PF	25	MIL	63	63	1917	685	1238	.553	89	293	.304	596	945	.631	.589	398	629	.633	140	716	856	354	61	66	230	195	1857
14	Kostas Antetokounmpo	PF	22	LAL	5	0	20	3	3	1.000	0	0		3	3	1.000	1.000	1	2	.500	2	1	3	2	0	0	1	2	7
15	Thanasis Antetokounmpo	SF	27	MIL	20	2	129	24	48	.500	0	10	.000	24	38	.632	.500	7	17	.412	12	12	24	15	7	2	12	18	55

Analysis Player's Performance

Analysis how players perform by calculate in different metrics;

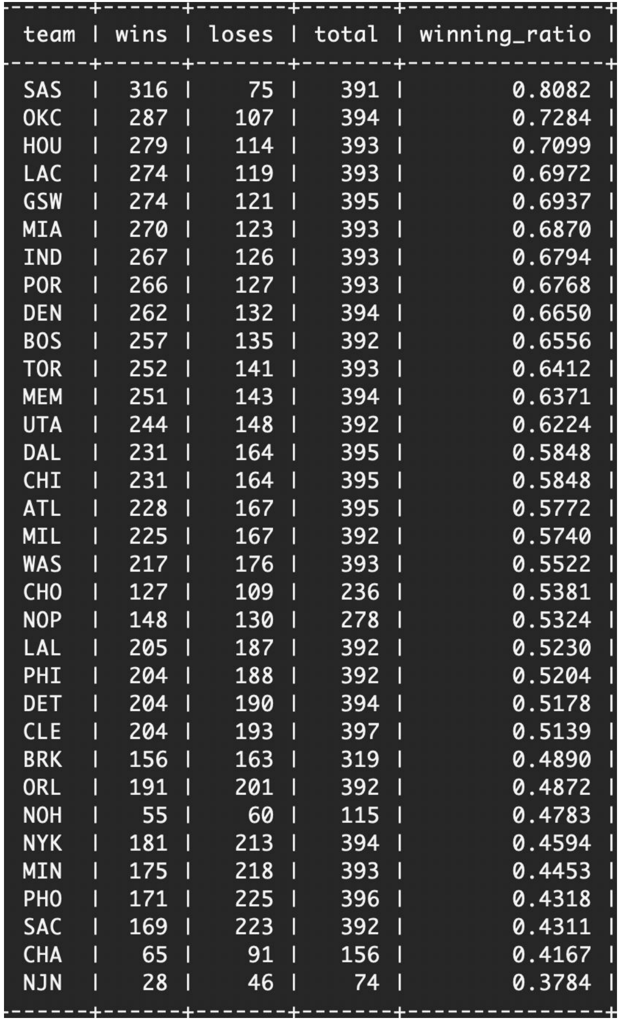
Know better about their strength and weakness;

Help them practice in furthur season



```
MySQL [pj]> select Player,AGE,GS,PTS,TRB,AST,STL,BLK,round(PTS/nullif(GS,0),2) as points_per_game,round(TRB/nullif(GS,0),2) as rebounds_per_game,round(AST/nullif(GS,0),2
S,0),2) as steal_per_game,round(BLK/nullif(GS,0),2) as block_per_game from 2020info limit 5;
+--------------------------+------+------+------+------+------+------+------+-----------------+-------------------+-----------------+----------------+----------------+
| Player                   | AGE  | GS   | PTS  | TRB  | AST  | STL  | BLK  | points_per_game | rebounds_per_game | assist_per_game | steal_per_game | block_per_game |
+--------------------------+------+------+------+------+------+------+------+-----------------+-------------------+-----------------+----------------+----------------+
| Steven Adams             | 26   | 63   | 684  | 583  | 146  | 51   | 67   |           10.86 |              9.25 |            2.32 |           0.81 |           1.06 |
| Bam Adebayo              | 22   | 72   | 1146 | 735  | 368  | 82   | 93   |           15.92 |             10.21 |            5.11 |           1.14 |           1.29 |
| LaMarcus Aldridge        | 34   | 53   | 1001 | 392  | 129  | 36   | 87   |           18.89 |              7.40 |            2.43 |           0.68 |           1.64 |
| Kyle Alexander           | 23   | 0    | 2    | 3    | 0    | 0    | 0    |            NULL |              NULL |            NULL |           NULL |           NULL |
| Nickeil Alexander-Walker | 21   | 1    | 267  | 84   | 89   | 17   | 8    |          267.00 |             84.00 |           89.00 |          17.00 |           8.00 |
+--------------------------+------+------+------+------+------+------+------+-----------------+-------------------+-----------------+----------------+----------------+
```

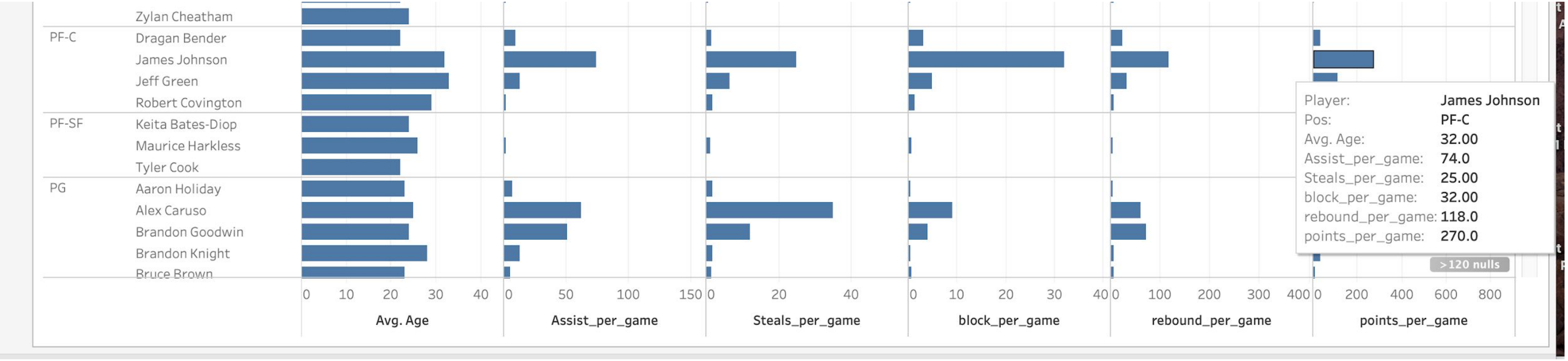
Winning Ratio



```
+------+------+-------+-------+---------------+
| team | wins | loses | total | winning_ratio |
+------+------+-------+-------+---------------+
| SAS  |  316 |    75 |   391 |        0.8082 |
| OKC  |  287 |   107 |   394 |        0.7284 |
| HOU  |  279 |   114 |   393 |        0.7099 |
| LAC  |  274 |   119 |   393 |        0.6972 |
| GSW  |  274 |   121 |   395 |        0.6937 |
| MIA  |  270 |   123 |   393 |        0.6870 |
| IND  |  267 |   126 |   393 |        0.6794 |
| POR  |  266 |   127 |   393 |        0.6768 |
| DEN  |  262 |   132 |   394 |        0.6650 |
| BOS  |  257 |   135 |   392 |        0.6556 |
| TOR  |  252 |   141 |   393 |        0.6412 |
| MEM  |  251 |   143 |   394 |        0.6371 |
| UTA  |  244 |   148 |   392 |        0.6224 |
| DAL  |  231 |   164 |   395 |        0.5848 |
| CHI  |  231 |   164 |   395 |        0.5848 |
| ATL  |  228 |   167 |   395 |        0.5772 |
| MIL  |  225 |   167 |   392 |        0.5740 |
| WAS  |  217 |   176 |   393 |        0.5522 |
| CHO  |  127 |   109 |   236 |        0.5381 |
| NOP  |  148 |   130 |   278 |        0.5324 |
| LAL  |  205 |   187 |   392 |        0.5230 |
| PHI  |  204 |   188 |   392 |        0.5204 |
| DET  |  204 |   190 |   394 |        0.5178 |
| CLE  |  204 |   193 |   397 |        0.5139 |
| BRK  |  156 |   163 |   319 |        0.4890 |
| ORL  |  191 |   201 |   392 |        0.4872 |
| NOH  |   55 |    60 |   115 |        0.4783 |
| NYK  |  181 |   213 |   394 |        0.4594 |
| MIN  |  175 |   218 |   393 |        0.4453 |
| PHO  |  171 |   225 |   396 |        0.4318 |
| SAC  |  169 |   223 |   392 |        0.4311 |
| CHA  |   65 |    91 |   156 |        0.4167 |
| NJN  |   28 |    46 |    74 |        0.3784 |
+------+------+-------+-------+---------------+
```

```
select team , wins , loses , total , (wins / total) as winning_ratio from (select team , count(if
(result like "%W%" , 1, NULL)) as wins , count(if(result like "%L%" , 1, NULL)) as loses ,
count(result) as total  from Matches group by team) as tb1 order by winning_ratio desc;
```

Data Visualization



Player:	James Johnson
Pos:	PF-C
Avg. Age:	32.00
Assist_per_game:	74.0
Steals_per_game:	25.00
block_per_game:	32.00
rebound_per_game:	118.0
points_per_game:	270.0

Ethical or not?

From our dataset?

All the data is visible to the public

No personal gathered from user

From Website?

Asked for personal email, address, credit/debit card info

Based on privacy statement from website, the visitors' home servers might be provided to third party.

Be Careful with it!