



## پروژه پایانی نیمه اول

گروه سوم

منتور:

خانم شادالین سلطان پور

اعضا:

هدیه، هلیا، غزل، آتوسا و محمد مهدی

زمستان ۱۴۰۴

## فهرست

۳	مقدمه
۴	فاز اول: استخراج داده
۴	جمع‌آوری اطلاعات بازیکن‌ها
۶	بررسی ستون‌ها
۱۰	بازیکن‌های فعال
۱۳	لیست مایکل جردن
۱۶	جمع‌آوری اطلاعات باشگاه‌ها
	فاز دوم: طراحی دیتابیس
	طراحی دیتابیس
	انتخاب ابزار و تکنولوژی دیتابیس
	جمع‌آوری و ذخیره‌ی داده‌ها در دیتابیس
	آماده‌سازی داده برای تحلیل و آموزش مدل
	فاز سوم: تحلیل‌های آماری
	آمار توصیفی
	تست فرض: فرضیه‌ی اول
	تست فرض: فرضیه‌ی دوم



## ❖ فاز اول: استخراج داده

داده‌های خواسته شده را از سایت <https://www.basketball-reference.com> استخراج می‌کنیم.

### ➤ جمع‌آوری اطلاعات بازیکن‌ها:

داده‌های بازیکن‌ها را با استفاده از این لینک <https://www.basketball-reference.com/players/> و با استفاده از کتابخانه selenium جمع‌آوری و ذخیره کردیم.

- **گام اول**، با تعریف یک تابع (get\_alphabet\_link) لینک تمام حروف الفبای اسامی را در یک لیست ذخیره کردیم.
- **گام دوم**، لینک تمام صفحات همه‌ی بازیکن‌های هر حرف الفبا رو با استفاده از یک تابع (get\_player\_page\_links) به دست می‌آوریم و در یک لیست ذخیره می‌کنیم.
- **گام سوم** که مهم‌ترین مرحله هست، به استخراج اطلاعات هر بازیکن از صفحه آن بازیکن می‌پردازیم. این کار توسط یک تابع انجام می‌شود. در ابتدا تلاش شد که با استفاده از کتابخانه BeautifulSoup و requests تمام اطلاعات بازیکن‌ها را از صفحات آن‌ها ذخیره کنیم ولی متأسفانه به علت زیاد بودن تعداد بازیکن‌ها و ارسال زیاد requests از طرف سایت بن شدیم و با توجه به ددلاین پروژه امکان استفاده از این روش دیگر نبود. سپس با استفاده از کتابخانه selenium و استفاده از regex بعضی از اطلاعات خواسته شده را ذخیره کردیم. به علت کند بودن این کتابخانه، زیاد بودن اطلاعات و زمان کم بقیه اطلاعات را از جداول هر لینک حروف الفبا با استفاده از کتابخانه pandas و دستور read\_html به صورت دیتافریم ذخیره کردیم. (داده‌های ستون قد و وزن را به سانتی‌متر و کیلوگرم تبدیل کردیم).
- **گام چهارم**، داده‌هایی که از دو طریق استخراج کرده بودیم را با یکدیگر merge کردیم تا تمام مشخصات در کنار هم قرار بگیرند.
- **گام پنجم**، محاسبه ستون‌های Age، Status و Experience از روی اطلاعات ذخیره شده به این صورت که:
  - محاسبه Age: اگر تاریخ وفات تعریف نشده بود، با استفاده از تاریخ امروز سن محاسبه شود و اگر شخص فوت شده بود، سن از طریق تفاضل تاریخ فوت و تاریخ تولد محاسبه شود.
  - محاسبه Status: اگر ستون To برابر ۲۰۲۶ بود یعنی بازیکن فعال (Active) هست. اگر تاریخ فوت برای بازیکن درج شده بود یعنی بازیکن فوت شده (Deceased) و در غیر اینصورت بازیکن بازنشسته (Retired) شده است.
  - محاسبه Experience: این ستون برای بازیکن‌هایی مقدار دارد که هنوز مشغول بازی (فعال) هستند پس، اگر Status برابر Active بود یعنی بازیکن‌ها فعال هستند و تفاضل ستون To و From برابر میزان تجربه بازیکن هست.

■ گام ششم، مرتب سازی ترتیب ستون‌ها، حذف داده‌های تکراری و ذخیره در فایل اکسل هست.

در نهایت، اطلاعات ذخیره شده برای هر بازیکن به صورت زیر است:

	Name	From	To	Birth date	Death date	Age	Status	Colleges	Height	Weight	Position	Shoots	Current Team	Experience	Pts
0	Alaa Abdelnaby	1991	1995	1968-06-24	NaT	57.0	Retired	Duke	208	109.0	Power Forward	Right	NaN	NaN	5.7
1	Zaid Abdul-Aziz	1969	1978	1946-04-07	NaT	79.0	Retired	Iowa State	205	107.0	Center Power Forward	Right	NaN	NaN	9.0
2	Kareem Abdul-Jabbar*	1970	1989	1947-04-16	NaT	78.0	Retired	UCLA	218	102.0	NaN	NaN	NaN	NaN	NaN
3	Mahmoud Abdul-Rauf	1991	2001	1969-03-09	NaT	56.0	Retired	LSU	185	73.0	Point Guard	Right	NaN	NaN	14.6
4	Tariq Abdul-Wahad	1998	2003	1974-11-03	NaT	51.0	Retired	Michigan, San Jose State	198	101.0	Shooting Guard	Right	NaN	NaN	7.8
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

هر ستون از جدول بیان کننده اطلاعات زیر است:

توضیحات	اسم ستون
نام و نام خانوادگی	Name
سال شروع بازی کردن	From
آخرین سال بازی کردن	To
تاریخ تولد	Birth date
تاریخ فوت	Death date
سن	Age
وضعیت بازیکن	Status
دانشگاه محل تحصیل	Colleges
قد (سانتی متر)	Height
وزن (کیلوگرم)	Weight
پست بازی	Position
پرتاب دست	Shoots
تیم فعلی بازیکن	Current Team
میزان تجربه بازیکنان فعال	Experience
میانگین امتیاز بازی	Pts

## بررسی ستون‌ها

Status		
بازنشسته، فوت شده، فعال		
Status: ['Retired' 'Deceased' 'Active']		
Status	count	
0	Retired	3757
1	Deceased	1151
2	Active	505

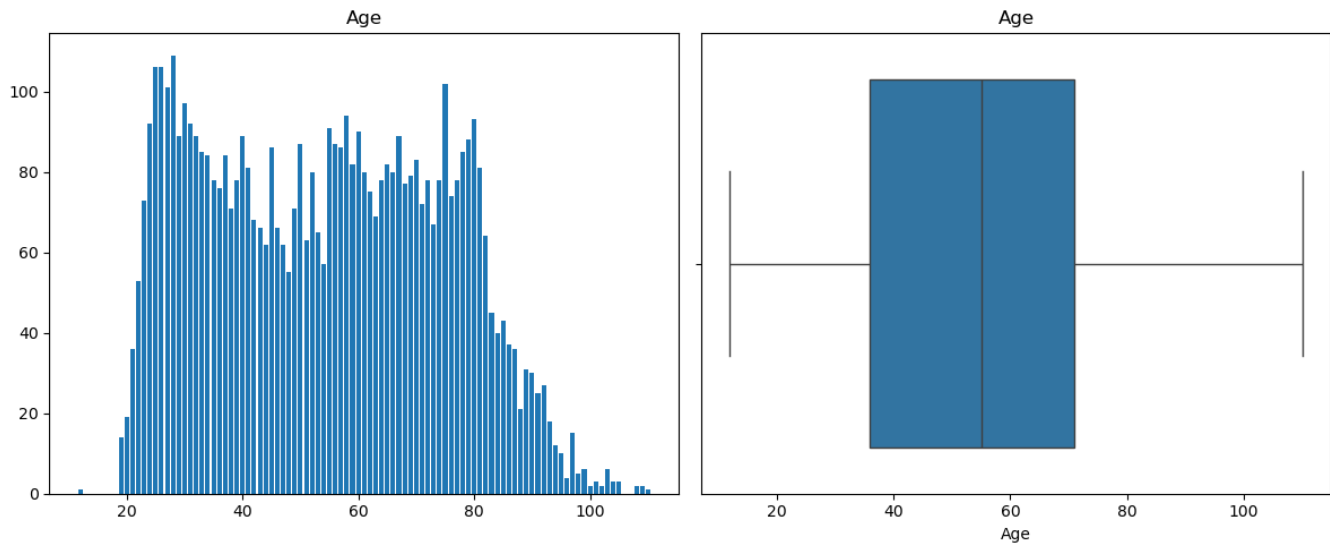
Shoots		
راست، چپ، راست و چپ، چپ و راست		
Shoots: ['Right' 'Left' 'Right Left' 'Left Right']		
Shoots	count	
0	Right	4878
1	Left	356
2	Right Left	1
3	Left Right	1

Current Team		
تعداد تیم‌های موجود به همراه نام تیم‌ها و تعداد بازیکن‌های هر تیم		
count of team: 30		
['Sacramento Kings' 'Houston Rockets' 'Miami Heat' 'Toronto Raptors' 'Memphis Grizzlies' 'New Orleans Pelicans' 'Atlanta Hawks' 'Phoenix Suns' 'Cleveland Cavaliers' 'Utah Jazz' 'Milwaukee Bucks' 'New York Knicks' 'Portland Trail Blazers' 'Los Angeles Lakers' 'Washington Wizards' 'Charlotte Hornets' 'Orlando Magic' 'Philadelphia' 'San Antonio Spurs' 'Oklahoma City Thunder' 'Los Angeles Clippers' 'Minnesota Timberwolves' 'Boston Celtics' 'Indiana Pacers' 'Denver Nuggets' 'Golden State Warriors' 'Chicago Bulls' 'Dallas Mavericks' 'Brooklyn Nets' 'Detroit Pistons']		

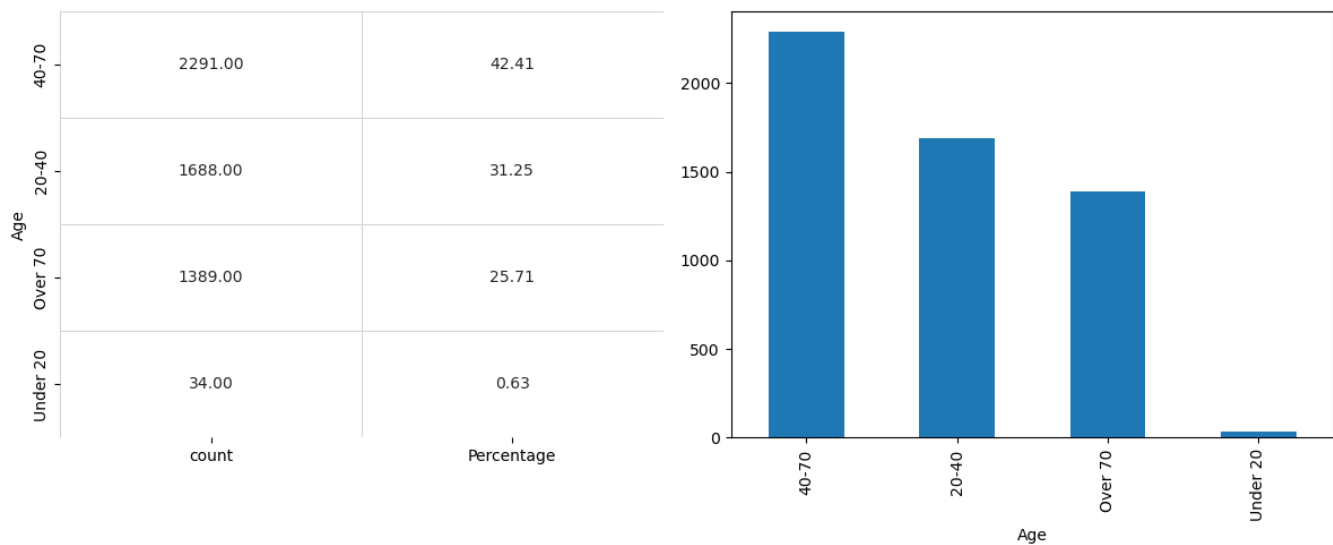
	Current Team	count
0	Memphis Grizzlies	19
1	Sacramento Kings	18
2	New Orleans Pelicans	18
3	Toronto Raptors	18
4	New York Knicks	18
5	Cleveland Cavaliers	18
6	Indiana Pacers	18
7	Chicago Bulls	18
8	Charlotte Hornets	18
9	Detroit Pistons	18
10	Washington Wizards	18
11	Miami Heat	17
12	Utah Jazz	17
13	Houston Rockets	17
14	Dallas Mavericks	17
15	Brooklyn Nets	17

16	Portland Trail Blazers	17
17	Los Angeles Lakers	17
18	Phoenix Suns	17
19	San Antonio Spurs	17
20	Orlando Magic	17
21	Philadelphia	17
22	Los Angeles Clippers	17
23	Golden State Warriors	17
24	Minnesota Timberwolves	16
25	Oklahoma City Thunder	16
26	Denver Nuggets	16
27	Boston Celtics	16
28	Milwaukee Bucks	15
29	Atlanta Hawks	15

## بررسی ستون سن بازیکن ها - Age

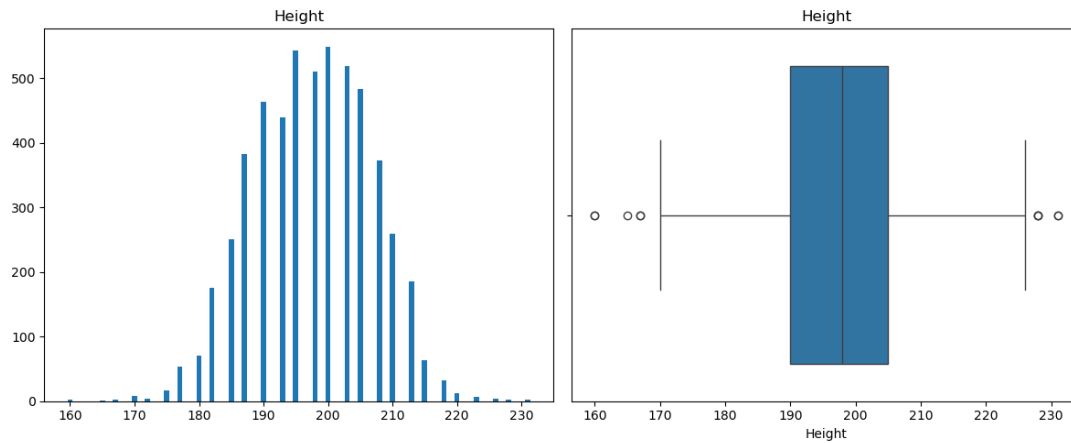


با توجه به نمودارها متوجه می شویم که فراوانی رده سنی بازیکن ها بین ۴۰ تا ۷۰ سال هست.

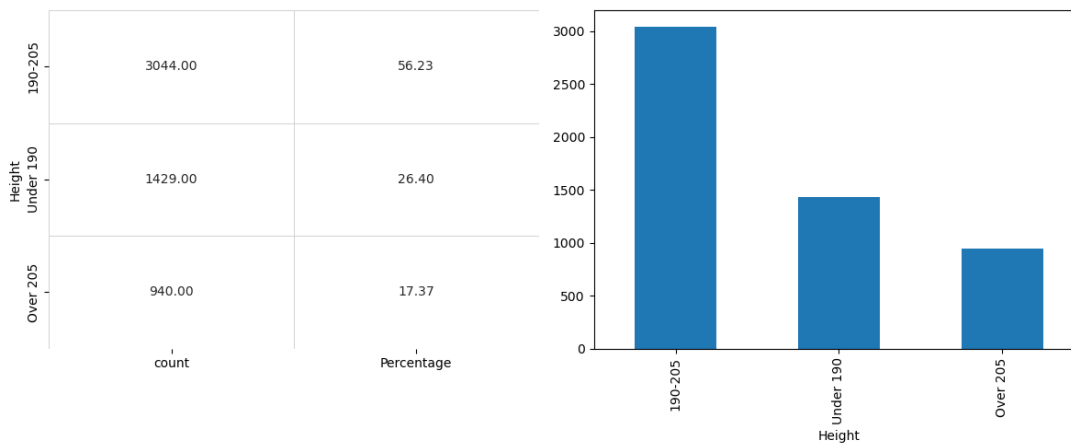


و بعد از دسته بندی کردن مشخص شد که ۴۲ درصد از بازیکن ها بین ۴۰ تا ۷۰ سال سن دارند و ۲۵ درصد بالای ۷۰ سال سن دارند و ۳۱ درصد از بازیکن ها بین ۲۰ تا ۴۰ سال سن دارند که بازیکن های فعال در این رده سنی قرار دارند.

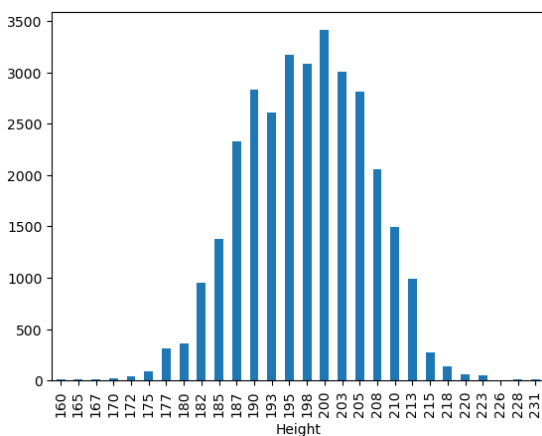
## بررسی ستون قد بازیکن ها - Height



با توجه به نمودارها متوجه می شویم که فراوانی قد بازیکن ها بین ۱۹۰ تا ۲۰۵ سانتی متر هست.



مشخص می شود که قد بیش از ۵۰ درصد از بازیکن ها بین ۱۹۰ و ۲۰۵ سانتی متر هست و تنها ۱۷ درصد از بازیکن ها قد بالای ۲۰۵ سانتی متر دارند.

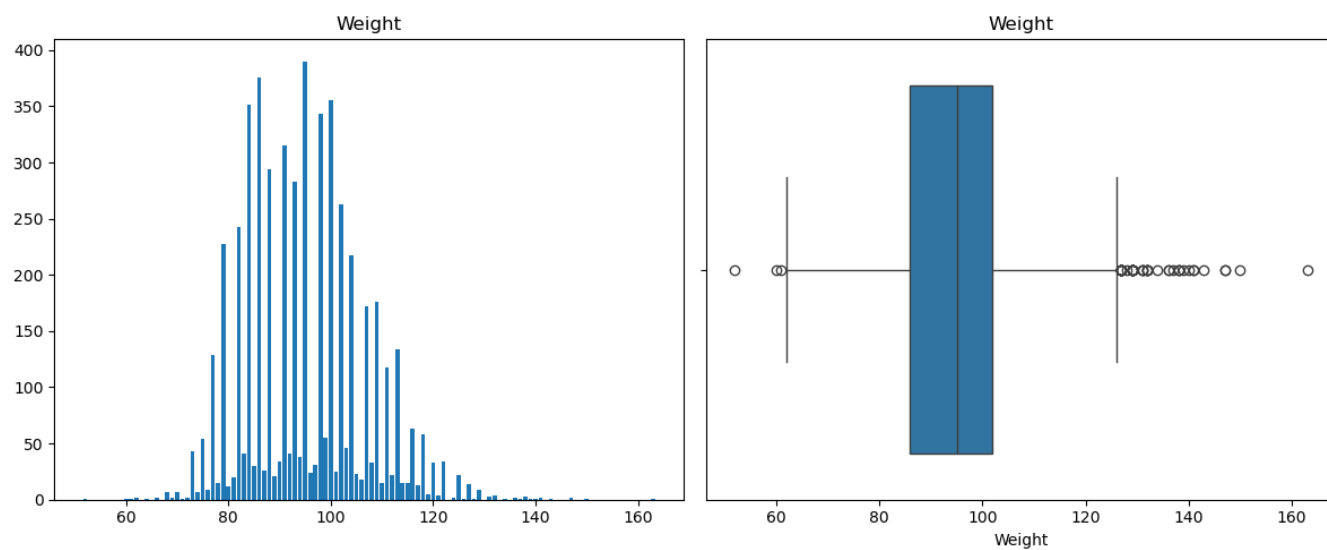


### تاثیر قد بر روی مجموع امتیازهای کسب شده بازیکن ها

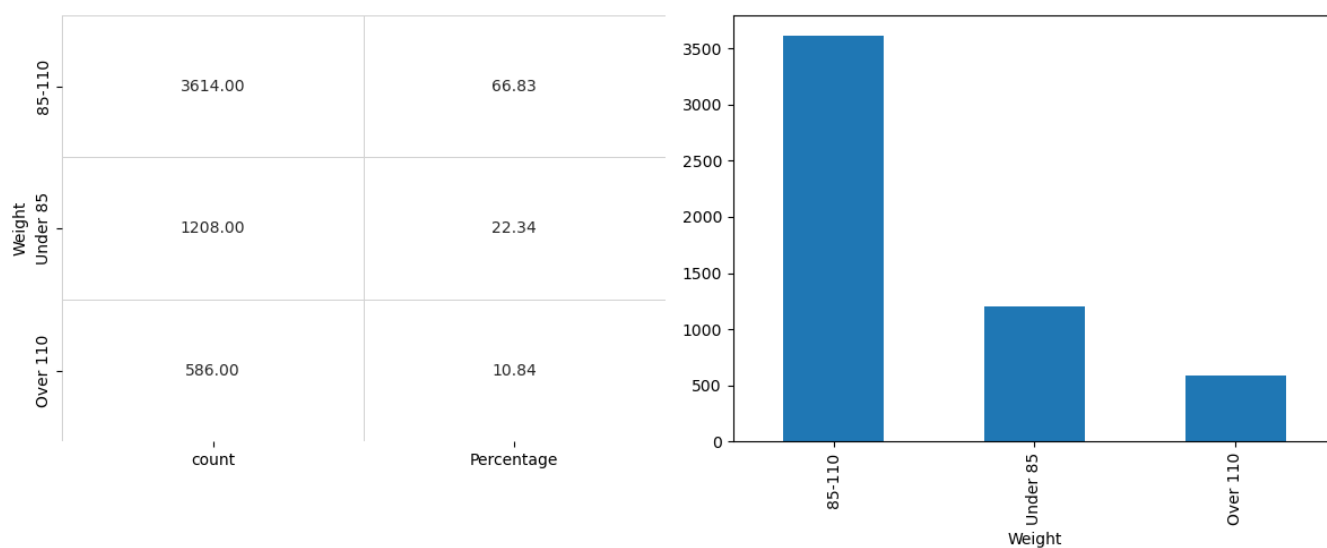
از روی نمودار متوجه می شویم که بازیکن هایی که قد بین ۱۹۰ و ۲۰۵ سانتی متر دارند و ۵۶ درصد از بازیکن ها را شامل می شوند توانسته اند امتیاز بیشتری کسب کنند.



## بررسی ستون وزن بازیکن ها – Weight



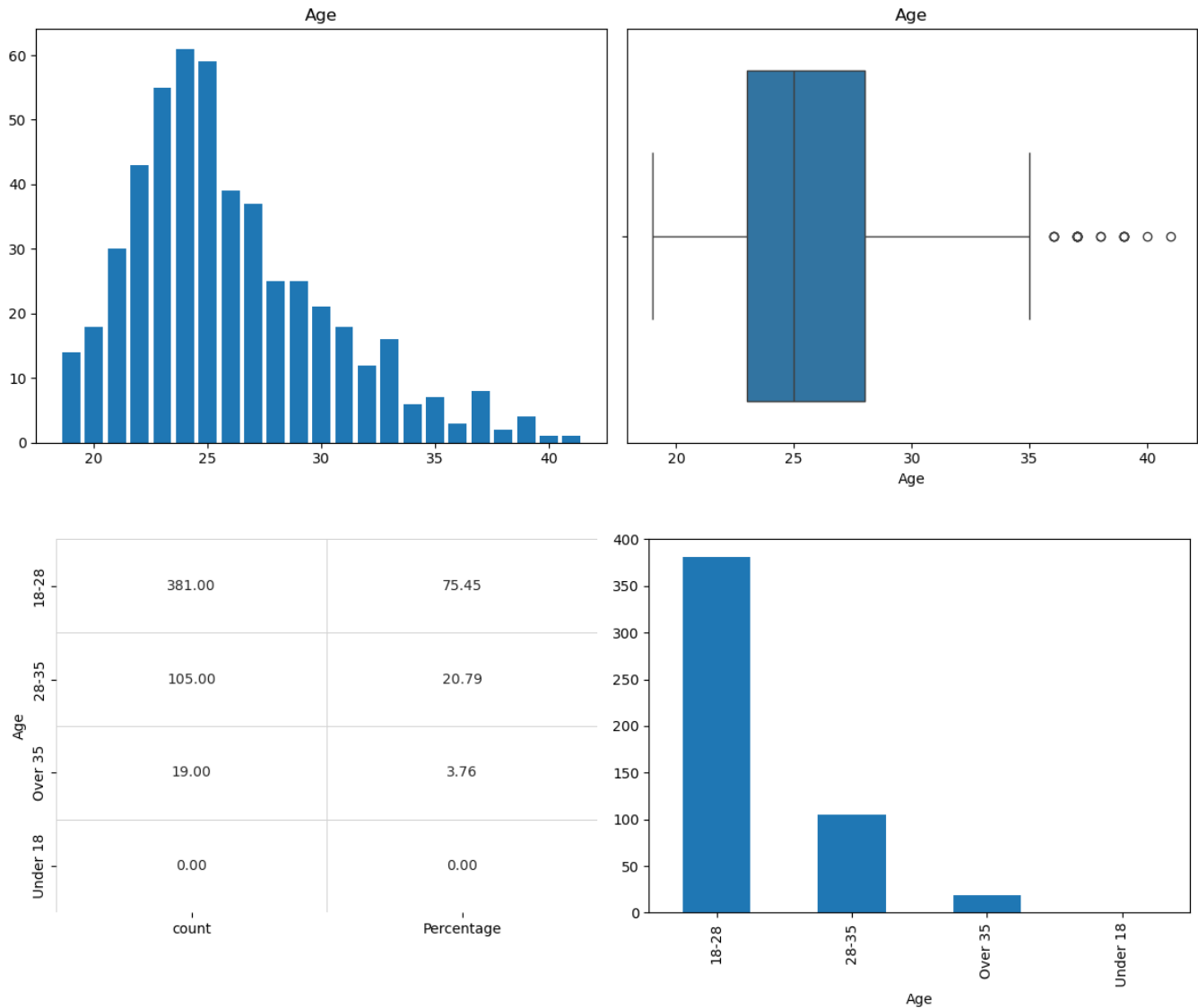
با توجه به نمودار متوجه می شویم که فراوانی وزن بازیکن ها حدودا بین ۸۵ تا ۱۱۰ کیلوگرم هست.



پس از دسته بندی مشخص می شود که وزن حدود ۶۷ درصد از بازیکن ها بین ۸۵ تا ۱۱۰ کیلوگرم هست و فقط ۱۱ درصد بازیکن ها وزن بالای ۱۱۰ کیلوگرم دارند.

## بازیکن‌های فعال (Status == Active)

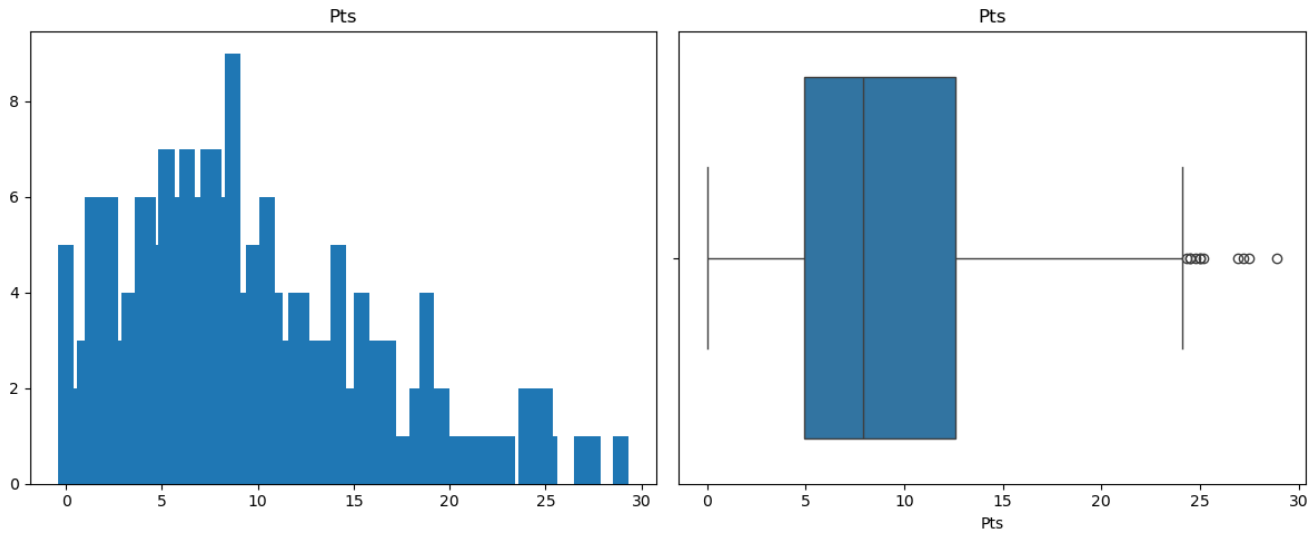
بررسی سن بازیکن‌های فعال



با توجه به نمودارها متوجه می‌شویم که در بین بازیکن‌های فعال، بازیکن زیر ۱۸ سال نداریم. رده سنی ۱۸ تا ۲۸ سال بیش‌ترین فراوانی را دارند و ۷۵ درصد بازیکن‌های فعال در این رده سنی قرار دارند و ۲۰ درصد بین ۲۸ و ۳۵ سال سن دارند و تنها حدود ۴ درصد از بازیکن‌های فعال بالای ۳۵ سال سن دارند.

## بازیکن‌های فعال (Status == Active)

بررسی میانگین امتیاز بازیکن‌های فعال



Average Of Pts Active Player: 9.18

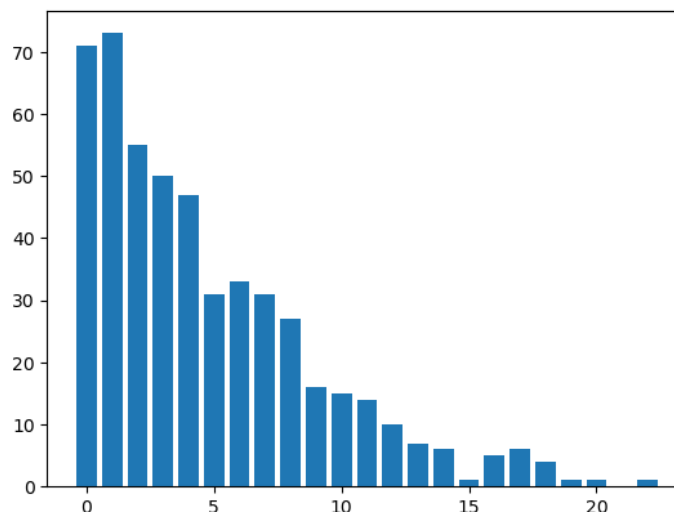
میانگین امتیاز بازیکن‌های فعال حدود ۹ است و از روی نمودار متوجه می‌شویم که بازه امتیاز کسب شده توسط این بازیکن‌ها بین ۵ تا ۱۲ هست.

## بازیکن‌های فعال (Status == Active)

بررسی میزان تجربه افراد فعال

Experience	Count	Percentage
0.0	71	14.059406
1.0	73	14.455446
2.0	55	10.891089
3.0	50	9.900990
4.0	47	9.306931
5.0	31	6.138614
6.0	33	6.534653
7.0	31	6.138614
8.0	27	5.346535
9.0	16	3.168317
10.0	15	2.970297
11.0	14	2.772277
12.0	10	1.980198
13.0	7	1.386139
14.0	6	1.188119
15.0	1	0.198020
16.0	5	0.990099
17.0	6	1.188119
18.0	4	0.792079
19.0	1	0.198020
20.0	1	0.198020
22.0	1	0.198020

Average Of Experience Active Player: 4.71



میانگین میزان تجربه بازیکن‌های فعال حدود ۵ است و از روی نمودار و جدول متوجه می‌شویم که بیشتر بازیکن‌های فعال تجربه بازی حرفه‌ای طولانی مدت ندارند و بیش‌ترین فراوانی بین بازه صفر تا هشت سال هست.

## بازیکنان جدید ("Experience" == 0.0)

Count of New Player(Join 2026): 71

Age	Count	Percentage
23.0	17	23.943662
19.0	14	19.718310
24.0	12	16.901408
20.0	10	14.084507
22.0	9	12.676056
21.0	6	8.450704
26.0	2	2.816901
25.0	1	1.408451

۷۱ بازیکن جدید داریم که در سال ۲۰۲۶ تازه بازی حرفه‌ایی خود را شروع کرده‌اند و تجربه‌ایی ندارند.

سن این بازیکن‌ها را بررسی کردیم که متوجه شویم بین چه رده‌سنی هستند و فراوانی هر سن به چه صورت هست. متوجه شدیم بین ۱۹ تا ۲۵ سال سن دارند و بیش از ۵۰ درصد از بازیکن‌های جدید بین ۱۹ تا ۲۲ سال سن دارند.

## لیست مایکل جردن

داده‌های بازیکن‌هایی که برنده این جایزه شده‌اند را با استفاده از کتابخانه pandas از این لینک <https://www.basketball-reference.com/awards/mvp.html> به دست آوردیم و ابتدا در یک لیست ذخیره کردیم و سپس به دیتافریم تبدیل و در فایل اکسل ذخیره کردیم.

	Season	Lg	Player	Voting	Age	Tm	G	MP	PTS	TRB	AST	STL	BLK	FG%	3P%	FT%	WS	WS/48
0	2024-25	NBA	Shai Gilgeous-Alexander	(V)	26	OKC	76	34.2	32.7	5.0	6.4	1.7	1.0	0.519	0.375	0.898	16.7	0.309
1	2023-24	NBA	Nikola Jokić	(V)	28	DEN	79	34.6	26.4	12.4	9.0	1.4	0.9	0.583	0.359	0.817	17.0	0.299
2	2022-23	NBA	Joel Embiid	(V)	28	PHI	66	34.6	33.1	10.2	4.2	1.0	1.7	0.548	0.330	0.857	12.3	0.259
3	2021-22	NBA	Nikola Jokić	(V)	26	DEN	74	33.5	27.1	13.8	7.9	1.5	0.9	0.583	0.337	0.810	15.2	0.296
4	2020-21	NBA	Nikola Jokić	(V)	25	DEN	72	34.6	26.4	10.8	8.3	1.3	0.7	0.566	0.388	0.868	15.6	0.301

هر ستون از دیتافریم بیان‌گر اطلاعات زیر است:

نام ستون	توضیحات
Season	سال فصل بازیکن
Lg	نام لیگ
Player	نام بازیکن
Age	سن
Tm	نام تیم بازیکن در آن فصل
G	تعداد بازی‌ها
MP	دقایق بازی بازیکن در هر بازی
PTS	امتیاز بازیکن در هر بازی
TRB	ریباند کل در هر بازی
AST	پاس گل در هر بازی
STL	توپ‌ربایی در هر بازی
BLK	بلاک در هر بازی
FG%	درصد موفقیت شوت‌های دو و سه امتیازی
3P%	درصد موفقیت شوت‌های سه امتیازی
FT%	درصد موفقیت پرتاب‌های آزاد
WS	سهم بازیکن در تعداد بردهای تیم
WS/48	سهم برد در هر ۴۸ دقیقه بازی

با توجه به اطلاعات بازیکن های این لیست، به بررسی چند مورد پرداختیم:

### لیست ۱۰ بازیکن برتر

بر اساس میانگین امتیازهای کسب شده

	Player	PTS
16	Kareem Abdul-Jabbar	172.2
24	Michael Jordan	155.7
36	Wilt Chamberlain	119.5
22	LeBron James	112.0
1	Bill Russell	83.3
25	Moses Malone	80.4
26	Nikola Jokić	79.9
21	Larry Bird	78.7
23	Magic Johnson	68.7
11	Giannis Antetokounmpo	57.2

بر اساس تعداد گل های زده شده

	Player	G
16	Kareem Abdul-Jabbar	490
24	Michael Jordan	408
1	Bill Russell	379
36	Wilt Chamberlain	314
22	LeBron James	295
21	Larry Bird	241
25	Moses Malone	241
23	Magic Johnson	236
26	Nikola Jokić	225
33	Tim Duncan	163

بر اساس تعداد دفعات کسب این جایزه

	Player	Season
16	Kareem Abdul-Jabbar	6
24	Michael Jordan	5
1	Bill Russell	5
36	Wilt Chamberlain	4
22	LeBron James	4
21	Larry Bird	3
23	Magic Johnson	3
26	Nikola Jokić	3
25	Moses Malone	3
32	Steve Nash	2

براساس این جداول کریم عبدالجبار را می توان بهترین بازیکن در هر زمینه دانست و پس از آن مایکل جردن مشهور قرار دارد.

	Season	Lg	Player	Voting	Age	Tm	G	MP	PTS	TRB	AST	STL	BLK	FG%	3P%	FT%	WS	WS/48
45	1979-80	NBA	Kareem Abdul-Jabbar	(V)	32	LAL	82	38.3	24.8	10.8	4.5	1.0	3.4	0.604	0.0	0.765	14.8	0.227
48	1976-77	NBA	Kareem Abdul-Jabbar	(V)	29	LAL	82	36.8	26.2	13.3	3.9	1.2	3.2	0.579	NaN	0.701	17.8	0.283
49	1975-76	NBA	Kareem Abdul-Jabbar	(V)	28	LAL	82	41.2	27.7	16.9	5.0	1.5	4.1	0.529	NaN	0.703	17.0	0.242
51	1973-74	NBA	Kareem Abdul-Jabbar	(V)	26	MIL	81	43.8	27.0	14.5	4.8	1.4	3.5	0.539	NaN	0.702	18.4	0.250
53	1971-72	NBA	Kareem Abdul-Jabbar	(V)	24	MIL	81	44.2	34.8	16.6	4.6	NaN	NaN	0.574	NaN	0.689	25.4	0.340
54	1970-71	NBA	Kareem Abdul-Jabbar	(V)	23	MIL	82	40.1	31.7	16.0	3.3	NaN	NaN	0.577	NaN	0.690	22.3	0.326

	Season	Lg	Player	Voting	Age	Tm	G	MP	PTS	TRB	AST	STL	BLK	FG%	3P%	FT%	WS	WS/48
27	1997-98	NBA	Michael Jordan	(V)	34	CHI	82	38.8	28.7	5.8	3.5	1.7	0.5	0.465	0.238	0.784	15.8	0.238
29	1995-96	NBA	Michael Jordan	(V)	32	CHI	82	37.7	30.4	6.6	4.3	2.2	0.5	0.495	0.427	0.834	20.4	0.317
33	1991-92	NBA	Michael Jordan	(V)	28	CHI	80	38.8	30.1	6.4	6.1	2.3	0.9	0.519	0.270	0.832	17.7	0.274
34	1990-91	NBA	Michael Jordan	(V)	27	CHI	82	37.0	31.5	6.0	5.5	2.7	1.0	0.539	0.312	0.851	20.3	0.321
37	1987-88	NBA	Michael Jordan	(V)	24	CHI	82	40.4	35.0	5.5	5.9	3.2	1.6	0.535	0.132	0.841	21.2	0.308

می توان متوجه شد که در چه فصل هایی، در چه تیم هایی، با چه امتیازهایی و در چه سنی توانستند این جایزه را کسب کنند.

## لیست ۱۰ تیم برتر

بر اساس مجموع گل های بازیکن ها

	Tm	G
1	BOS	766
10	LAL	643
16	PHI	539
3	CHI	489
12	MIL	379
9	HOU	315
20	SAS	244
15	OKC	238
17	PHO	230
7	DEN	225

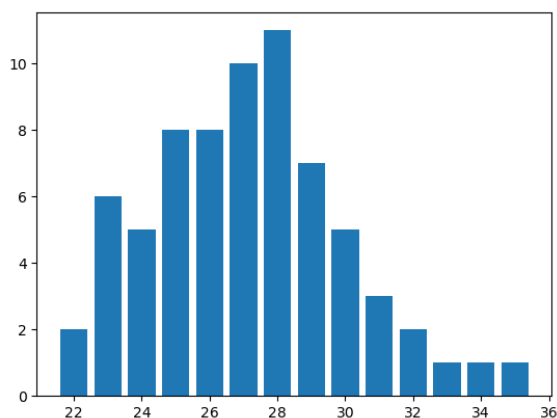
بر اساس تعداد دفعات کسب این جایزه توسط بازیکن های آنها

	Tm	Player
1	BOS	10
10	LAL	8
16	PHI	7
3	CHI	6
12	MIL	5
9	HOU	4
20	SAS	3
15	OKC	3
17	PHO	3
7	DEN	3

بهترین تیم، تیم بوستون هست. حالا می خواهیم بدانیم چه بازیکن هایی برای این باشگاه این افتخارات را کسب نموده اند.

	Season	Lg	Player	Voting	Age	Tm	G
39	1985-86	NBA	Larry Bird	(V)	29	BOS	82
40	1984-85	NBA	Larry Bird	(V)	28	BOS	80
41	1983-84	NBA	Larry Bird	(V)	27	BOS	79
52	1972-73	NBA	Dave Cowens	(V)	24	BOS	82
60	1964-65	NBA	Bill Russell	(V)	30	BOS	78
62	1962-63	NBA	Bill Russell	(V)	28	BOS	78
63	1961-62	NBA	Bill Russell	(V)	27	BOS	76
64	1960-61	NBA	Bill Russell	(V)	26	BOS	78
67	1957-58	NBA	Bill Russell	(V)	23	BOS	69
68	1956-57	NBA	Bob Cousy	(V)	28	BOS	64

## نمودار فراوانی سن همه ی بازیکن ها در این لیست



بر اساس این نمودار می توان نتیجه گرفت که بیشتر بازیکن ها در این لیست بین ۲۵ تا ۳۰ سال سن دارند.

➤ جمع آوری اطلاعات باشگاه‌ها:



