



مینی پروژه سوم

گروه پنجم

اعضای گروه : پوریا ذره پرور ، ماهان رضایی ، نیلوفر نینوایی ، علی قاسمی نژاد ، طناز حافظی

منتور : مصطفی محمدی

لینک گیت هاب : <https://github.com/QBC-DA10-Group5/basketball-reference>

پیشگفتار

در دنیای امروز، داده‌ها نقشی حیاتی در تصمیم‌گیری‌های ورزشی و مدیریتی ایفا می‌کنند. لیگ NBA به‌عنوان یکی از معتبرترین لیگ‌های بسکتبال جهان، حجم عظیمی از اطلاعات در مورد بازیکنان، تیم‌ها، مسابقات و عملکرد فصل‌ها را تولید می‌کند. استفاده‌ی بهینه از این داده‌ها نه تنها به تحلیلگران و مربیان درک عمیق‌تری از نقاط قوت و ضعف تیم‌ها و بازیکنان می‌دهد، بلکه می‌تواند برای هواداران، سرمایه‌گذاران و مدیران باشگاه‌ها نیز ابزار ارزشمندی باشد.

پروژه‌ی حاضر با هدف طراحی و پیاده‌سازی یک پایگاه داده‌ی جامع برای اطلاعات NBA شکل گرفته است. در فاز نخست، موجودیت‌های اصلی همچون بازیکنان، تیم‌ها، فصل‌ها و مسابقات شناسایی شده و جداول موردنیاز همراه با کلیدهای اصلی و روابط بین آن‌ها طراحی گردید. در ادامه و طی فاز دوم، با بهره‌گیری از ابزارهای نوین مانند SQL و SQLAlchemy، دیتابیس ایجاد شده و داده‌های واقعی استخراج‌شده از وبسایت [Basketball Reference](#) توسط تیم جمع‌آوری و در جداول وارد شد. در نهایت، این پایگاه داده زمینه‌ساز انجام تحلیل‌های آماری و پاسخ به پرسش‌های کلیدی درباره‌ی عملکرد بازیکنان و تیم‌ها خواهد بود.

این گزارش در واقع داستانی از مسیر طراحی تا تحلیل است؛ سفری که از تعریف مسئله و طراحی منطقی پایگاه داده آغاز شده، با جمع‌آوری و ذخیره‌سازی داده‌ها ادامه یافته و در نهایت به مرحله‌ی تحلیل و پاسخ به فرضیات آماری ختم می‌شود.

پیشگفتار	1
فهرست	2
مقدمه	3
NBA دانش پایه ای راجع به	4
فاز اول : طراحی شمای اصلی دیتابیس و استخراج داده	5
طراحی شمای اصلی دیتابیس	5
players	7
teams	7
seasons	8
awards	8
awards_season	8
player_seasons	9
team_seasons	10
mvp_candidates	10
web scraping (وب اسکرپینگ)	12
فاز دوم : طراحی دیتابیس	13
فاز سوم: تحلیل های آماری	14
آمار توصیفی) سوال 1 :	14
آمار توصیفی) سوال 2 :	15
آمار توصیفی) سوال 3 :	16
فرضیه ها) تست فرض :	16
فرضیه ها) فرضیه 1 :	17
فرضیه ها) فرضیه 2 :	18
Visualization	19
فرضیه های بیشتر	24

مقدمه

امروزه تحلیل داده در حوزه‌ی ورزش به یکی از ارکان اصلی موفقیت تیم‌ها و بازیکنان تبدیل شده است. حجم بالای اطلاعات تولیدشده در مسابقات ورزشی، فرصتی فراهم کرده تا با استفاده از روش‌های نوین علم داده و طراحی پایگاه‌های اطلاعاتی، بتوان از دل این داده‌ها بینش‌های ارزشمند استخراج کرد. لیگ NBA به‌عنوان بزرگ‌ترین و محبوب‌ترین لیگ بسکتبال جهان، نمونه‌ای بارز از محیطی است که در آن داده‌ها به‌صورت گسترده در اختیار پژوهشگران، تحلیلگران و مدیران ورزشی قرار دارد.

با توجه به اهمیت این داده‌ها، پروژه‌ی حاضر با هدف ایجاد یک پایگاه داده ساخت‌یافته و یکپارچه از اطلاعات NBA تعریف شده است. این پایگاه داده شامل موجودیت‌های اصلی مانند بازیکنان، تیم‌ها، فصل‌ها، مسابقات و جوایز بوده و به‌گونه‌ای طراحی شده است که بتواند روابط پیچیده‌ی بین آن‌ها را به‌صورت دقیق و بدون افزونگی نمایش دهد.

طراحی چنین پایگاه داده‌ای، علاوه بر امکان ذخیره‌سازی مطمئن و مدیریت بهینه‌ی داده‌ها، بستری فراهم می‌آورد تا در مراحل بعد بتوان تحلیل‌های آماری متنوعی انجام داد؛ از بررسی عملکرد بازیکنان و مقایسه‌ی تیم‌ها گرفته تا آزمون فرضیه‌های ورزشی و ارائه‌ی پیشنهادهای مبتنی بر داده برای تصمیم‌گیری‌های آینده.

در مجموع، این پروژه نشان می‌دهد که چگونه ترکیب ابزارهای پایگاه داده، وب‌اسکرپینگ و تحلیل آماری می‌تواند به خلق یک منبع ارزشمند برای درک عمیق‌تر دنیای بسکتبال منجر شود.

دانش پایه ای راجع به NBA

بسکتبال یکی از محبوب‌ترین ورزش‌های تیمی در جهان است که در آن دو تیم پنج نفره تلاش می‌کنند توپ را به داخل سبد حریف پرتاب کنند. این ورزش در سال ۱۸۹۱ توسط جیمز نایس‌اسمیت (James Naismith) ابداع شد و به سرعت در ایالات متحده و سپس در سراسر جهان گسترش یافت.

NBA (National Basketball Association) معتبرترین و حرفه‌ای‌ترین لیگ بسکتبال جهان است که در سال ۱۹۴۶ تأسیس شد. این لیگ شامل ۳۰ تیم است (۲۹ تیم از ایالات متحده و ۱ تیم از کانادا) که در دو کنفرانس شرق و غرب دسته‌بندی می‌شوند.

فصل‌های NBA شامل سه بخش اصلی هستند: فصل عادی که برای رتبه‌بندی تیم‌ها برگزار می‌شود، مرحله پلی‌آف که تیم‌های برتر هر کنفرانس در آن رقابت می‌کنند، و نهایتاً فینال که قهرمان کنفرانس شرق و غرب برای کسب عنوان قهرمانی لیگ با یکدیگر بازی می‌کنند. هر تیم ترکیبی از بازیکنان اصلی و ذخیره دارد و موقعیت‌های اصلی بازیکنان شامل Point Guard، Shooting Guard، Small Forward، Power Forward و Center است. از ویژگی‌های منحصربه‌فرد NBA، ثبت و انتشار دقیق آمار بازیکنان و تیم‌هاست؛ از جمله امتیاز، ریباند، پاس گل، توپ‌قاپی، بلاک و درصد شوت‌ها. این حجم از داده‌های ساختاریافته موجب شده NBA به یکی از غنی‌ترین منابع داده ورزشی در جهان تبدیل شود و تحلیل آن در حوزه‌های ورزشی، مدیریتی و اقتصادی ارزش فراوانی داشته باشد.

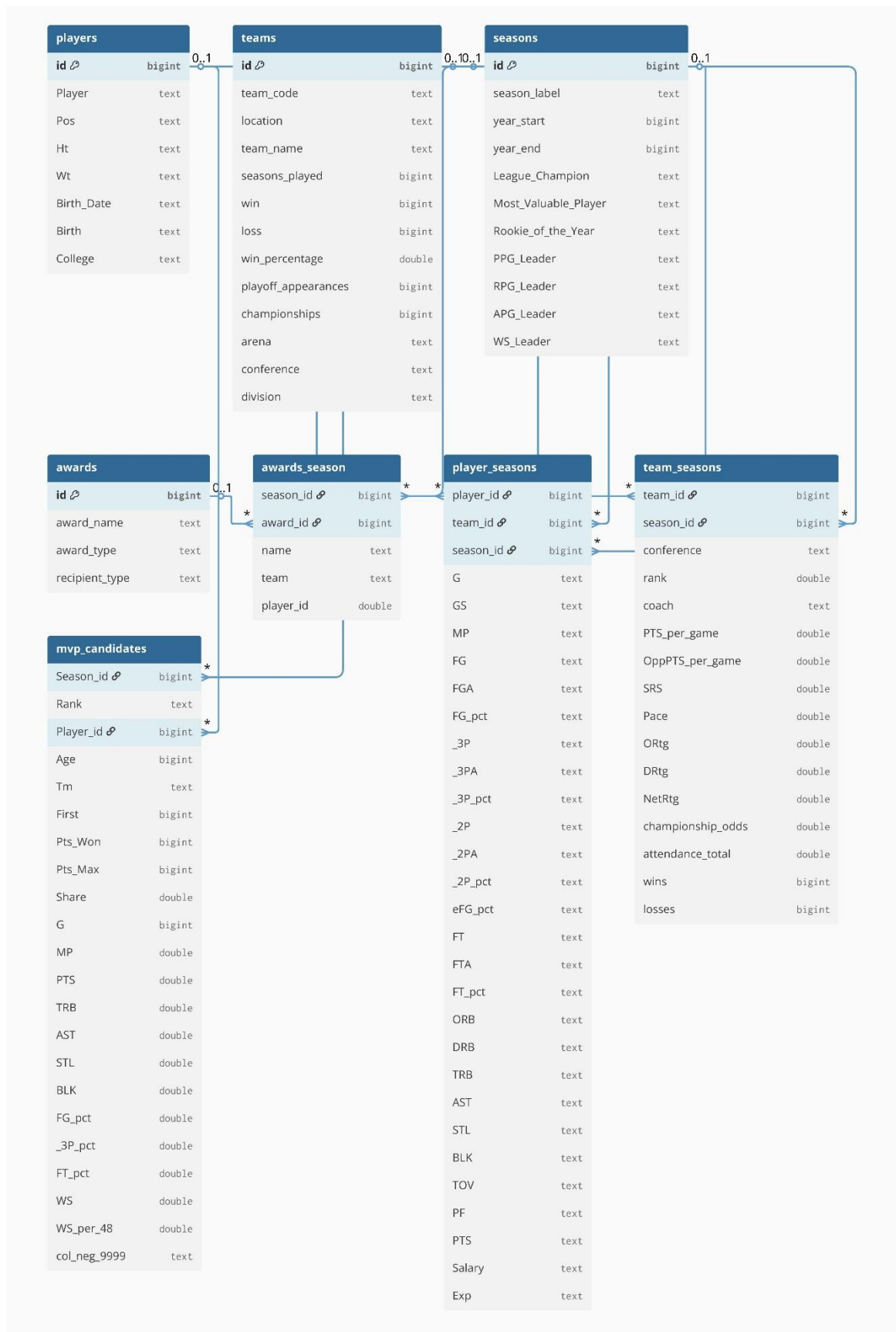
فاز اول : طراحی شمای اصلی دیتابیس و استخراج داده

طراحی شمای اصلی دیتابیس

پس از کسب اطلاعات اولیه درباره لیگ NBA ، وارد فاز ابتدایی پروژه شدیم. پیش از شروع فرایند استخراج داده از وبسایت مرجع، لازم بود شمای کلی از نیازهای پروژه و داده‌های مورد نیاز به دست آوریم. به همین منظور، با توجه به سؤالات مطرح شده در فاز سوم و همچنین فرضیه‌هایی که تصمیم داشتیم به آن‌ها بپردازیم، اقدام به طراحی دیتابیس کردیم تا ساختار داده‌ها و روابط میان آن‌ها به‌طور شفاف مشخص شود.

دیتابیس نهایی شامل موجودیت‌های زیر است که هر کدامشان ویژگی‌های منحصر به فرد خود را دارد:

- جدول **players** : شامل اطلاعات کلی هر بازیکن مثل نام و نام خانوادگی ، تاریخ تولد ، قد ، وزن و مکان تولد هر بازیکن است.
- جدول **player_seasons** : شامل اطلاعات هر بازیکن در هر فصل می باشد. مثل تعداد هایی که تجربه ی حرفه ای داشته است ، درآمد بازیکن ، تعداد دقایقی که بازی کرده است و
- جدول **teams** : شامل اطلاعات کلی هر تیم می باشد. مثل نام تیم ، موقعیت مکانی تیم ، ایالتی که تیم در آن است ، کنفرانس تیم (شرق یا غرب) و قسمت تیم (division)
- جدول **team_seasons** : شامل اطلاعات هر تیم در هر فصل است. مثل تعداد برد و باخت هر تیم در هر فصل ، نام مربی ، رتبه و رنک تیم در فصل و
- جدول **awards_season** : شامل هر جایزه و تیم هایی که جایزه ها را برنده شده اند است.
- جدول **awards** : جوایزی که در لیگ NBA وجود دارد.
- جدول **seasons** : اطلاعات کلی راجع به هر فصل مثل زمان شروع و پایان ، قهرمان فصل ، ارزشمندترین بازیکن و
- جدول **mvp_candidates** : این جدول نشان میدهد در هر فصل چه کسانی کاندید مایکل جردن تروفي بودن. چون توی سؤالات به این اطلاعات نیاز داشتیم به همین خاطر اضافه کردیم و با Player_id به جدول players و با season_id به جدول seasons وصل می شود.



در ادامه به جزییات جداولی که ساختیم میپردازیم :

players

Attribute	Explanation
id (bigint)	شناسه یکتای بازیکن (کلید اصلی)
Player (text)	نام کامل بازیکن
Pos (text)	موقعیت بازیکن
Ht (text)	قد بازیکن
Wt (text)	وزن بازیکن
Birth_Date (text)	تاریخ تولد بازیکن
Birth (text)	محل تولد
College (text)	دانشگاه یا کالجی که بازیکن در آن تحصیل کرده یا بازی کرده

teams

Attribute	Explanation
id (bigint)	شناسه یکتای تیم (کلید اصلی)
team_code (text)	کد اختصاری تیم (مثلاً BOS برای Boston Celtics)
location (text)	شهر/مکان تیم
team_name (text)	نام تیم
seasons_played (bigint)	تعداد فصل‌هایی که تیم بازی کرده
win (bigint)	کل بردها
loss (bigint)	کل باخت‌ها
win_percentage (double)	درصد بردها
playoff_appearances (bigint)	تعداد حضور در پلی‌آف
championships (bigint)	تعداد قهرمانی‌ها

arena (text)	سالن ورزشی خانگی
conference (text)	کنفرانس تیم (East یا West)
division (text)	دسته یا بخش تیم

seasons

Attribute	Explanation
id (bigint)	شناسه فصل (کلید اصلی)
season_label (text)	برچسب فصل (مثل 2020-21)
year_start (bigint)	سال شروع فصل
year_end (bigint)	سال پایان فصل
League_Champion (text)	تیم قهرمان لیگ
Most_Valuable_Player (text)	MVP بازیکن
Rookie_of_the_Year (text)	بهترین تازه وارد
PPG_Leader (text)	رهبر میانگین امتیاز
RPG_Leader (text)	رهبر ریباند
APG_Leader (text)	رهبر پاس گل
WS_Leader (text)	Win Shares رهبر شاخص

awards

Attribute	Explanation
id (bigint)	شناسه جایزه (کلید اصلی)
award_name (text)	نام جایزه
award_type (text)	نوع جایزه (فصلی، سالانه، ...)
recipient_type (text)	نوع دریافت کننده (بازیکن، مربی، تیم)

awards_season

Attribute	Explanation
-----------	-------------

season_id (bigint)	شناسه فصل (FK → seasons)
award_id (bigint)	شناسه جایزه (FK → awards)
name (text)	نام برنده
team (text)	تیم مربوط به جایزه
player_id (double)	شناسه بازیکن در صورت وجود (FK → players)

player_seasons

Attribute	Explanation
player_id (bigint)	شناسه بازیکن (FK → players)
team_id (bigint)	شناسه تیم (FK → teams)
season_id (bigint)	شناسه فصل (FK → seasons)
G (text)	تعداد بازی‌ها
GS (text)	تعداد بازی‌های شروعی
MP (text)	میانگین دقایق بازی
FG, FGA, FG_pct (text)	شوت‌های موفق، تلاش و درصد
_3P, _3PA, _3P_pct (text)	سه امتیازی‌ها و درصد آن
_2P, _2PA, _2P_pct (text)	دو امتیازی‌ها و درصد آن
eFG_pct (text)	درصد شوت موثر
FT, FTA, FT_pct (text)	آمار پرتاب آزاد
ORB, DRB, TRB (text)	ریباند هجومی، دفاعی و کل
AST, STL, BLK, TOV, PF (text)	پاس گل، توپ‌قاپی، بلاک، از دست دادن توپ، خطاها
PTS (text)	امتیاز
Salary (text)	حقوق
Exp (text)	تجربه بازیکن

team_seasons

Attribute	Explanation
team_id (bigint)	شناسه تیم (FK → teams)
season_id (bigint)	شناسه فصل (FK → seasons)
conference (text)	کنفرانس
rank (double)	رتبه تیم
coach (text)	نام سرمربی
PTS_per_game (double)	میانگین امتیاز در هر بازی
OppPTS_per_game (double)	میانگین امتیاز خورده
SRS (double)	شاخص قدرت ساده
Pace (double)	سرعت بازی
ORTg, DRTg, NetRtg (double)	رتبه هجومی، دفاعی و خالص
championship_odds (double)	شانس قهرمانی
attendance_total (double)	کل تماشاگران
wins, losses (bigint)	برد و باخت‌ها

mvp_candidates

Attribute	Explanation
Season_id (bigint)	شناسه فصل (FK → seasons)
Rank (text)	MVP رتبه بازیکن در لیست
Player_id (bigint)	شناسه بازیکن (FK → players)
Age (bigint)	سن بازیکن
Tm (text)	تیم بازیکن
First (bigint)	تعداد رأی اول
Pts_Won (bigint)	امتیاز کسب شده
Pts_Max (bigint)	حداکثر امتیاز ممکن

Share (double)	درصد سهم رأی
G (bigint)	تعداد بازی‌ها
MP (double)	میانگین دقیقه بازی
PTS (double)	امتیاز
TRB (double)	ریباند
AST (double)	پاس گل
STL (double)	توپ‌قاپی
BLK (double)	بلاک
FG_pct (double)	درصد پرتاب
_3P_pct (double)	درصد سه‌امتیازی
FT_pct (double)	درصد پرتاب آزاد
WS (double)	Win Shares
WS_per_48 (double)	در هر ۴۸ دقیقه Win Shares
col_neg_9999 (text)	ستون اضافی برای داده‌های خاص/نامشخص

وب اسکرپینگ (web scraping)

پس از مشخص شدن جداول و ویژگی‌های هر جدول، شروع به جمع‌آوری اطلاعات از سایت مرجع کردیم. برای جدول‌های player و player_season از کتابخانه‌ی Selenium استفاده شد. برای استخراج داده‌ها، وارد قسمت teams سایت شدیم و تیم‌های فعال را انتخاب کردیم. به ازای هر تیم، از سال ۲۰۱۹-۲۰۱۸ تا ۲۰۲۴-۲۰۲۵ وارد هر season شدیم و از جداول Roster و Table per Game استفاده کردیم تا اطلاعات مورد نیاز برای جدول player و player_season به دست آید.

برای مابقی جداول، شامل اطلاعات مربوط به تیم‌ها، فصل‌ها و جایزه MVP، تلاش شد مطابق با اسکیمای ما تا حد امکان فانکشنال نوشته شوند. در این بخش از کتابخانه‌های Requests و BeautifulSoup استفاده شد و برای عبور از DDoS Protection سایت از کتابخانه‌ی Cloudscraper بهره برده شد. همچنین برای جلوگیری از Rate Limit، تاخیرهایی در کد اعمال شد و در نهایت با استفاده از توابع پایتون و کتابخانه‌ی Regex، مقادیر مربوط به تگ‌ها استخراج شدند.

یکی از چالش‌های این بخش این بود که نام بازیکنان به زبان‌های غیرانگلیسی نوشته شده بود، که در نهایت همه کاراکترها به زبان انگلیسی نرمال‌سازی شدند.

فاز دوم : طراحی دیتابیس

حتما به این نکته توجه داشته باشد که فاز اول و دوم پروژه هم پوشانی بسیار زیادی دارند و در هم آمیخته اند و ما برای مدیریت این مسئله به صورت رفت و برگشتی این دو قسمت را حل کردیم.

پس از اینکه تیم وب اسکریپینگ جداولی که طراحی کرده بودیم را در قالب CSV به دست آوردند زمان آن فرارسیده بود که دیتابیس بسازیم تا بتوانیم با استفاده از آن به سوالات و تحلیل های آماری بپردازیم.

برای ساختن دیتابیس با استفاده از `sqlite` یک دیتابیس به صورت لوکال بر روی سیستم ایجاد می شود. سپس با استفاده از `sqlalchemy` یک اینجین می سازیم و به دیتابیس وصل می کنیم. با کمک اینجین دیتا را از فایل های CSV به دیتابیس منتقل می کنیم. در نهایت با همین اینجین هم به دیتابیس کویری می زنیم و دیتای مورد نظر را دریافت می کنیم.

در تمامی مراحل ساخت پایگاه داده، به ساختار و قواعد صحیح طراحی توجه ویژه ای داشتیم. کلیدهای اصلی و فرعی به دقت انتخاب و تعریف شده اند و روابط بین جداول به درستی برقرار شده اند. تلاش کردیم از روابط یک به چند غیرضروری پرهیز کنیم و در نهایت، به نرمال سازی داده ها نیز توجه شده است.

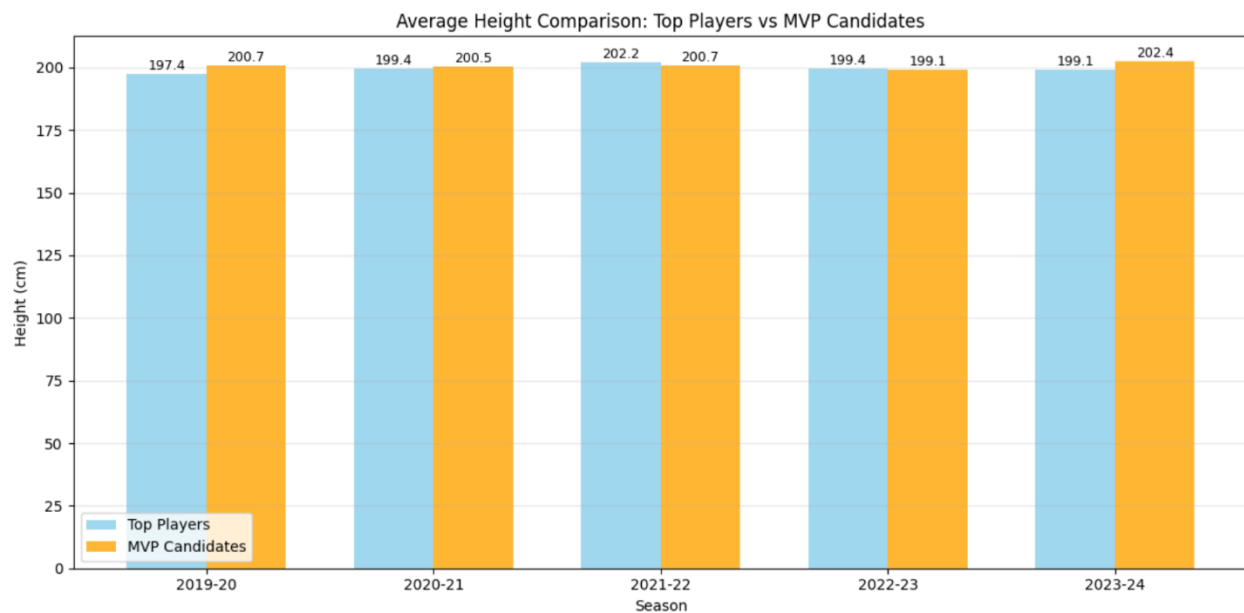
توضیحات درباره ی SQLite:

SQLite یک سیستم مدیریت پایگاه داده ی سبک و Embedded است که برخلاف پایگاه های داده ی سنتی مانند MySQL یا SQL Server، نیاز به نصب سرور جداگانه ندارد. به عبارت دیگر، SQLite یک پایگاه داده سرورلس است که به صورت یک فایل واحد روی دیسک کار می کند.

فاز سوم: تحلیل های آماری

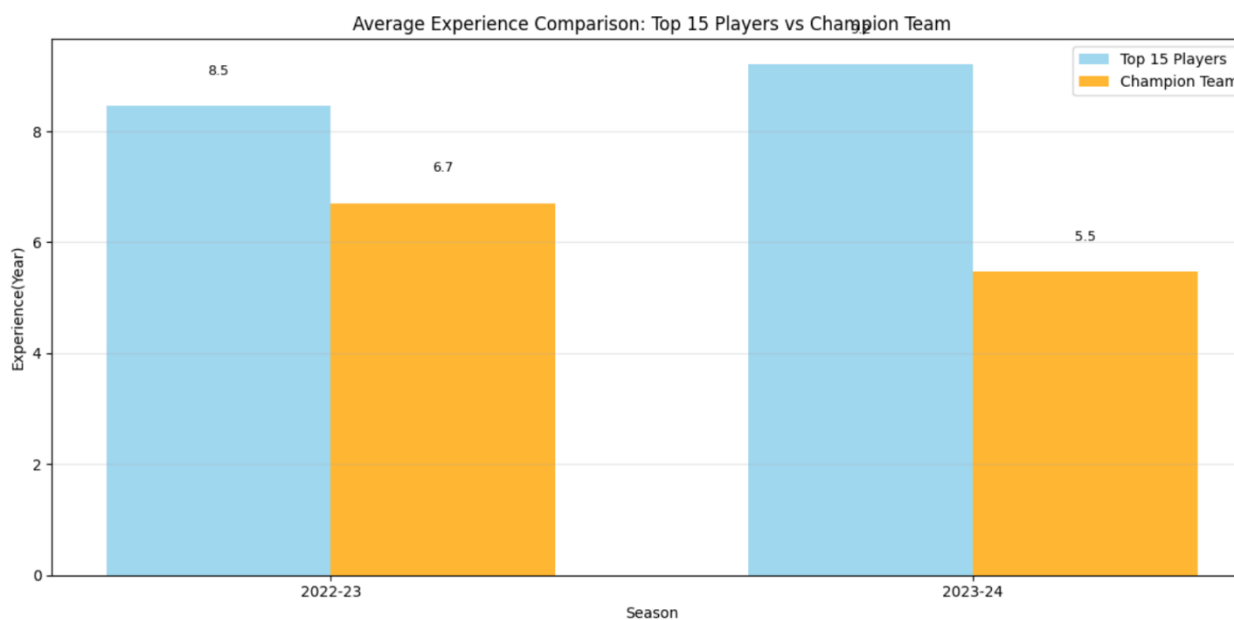
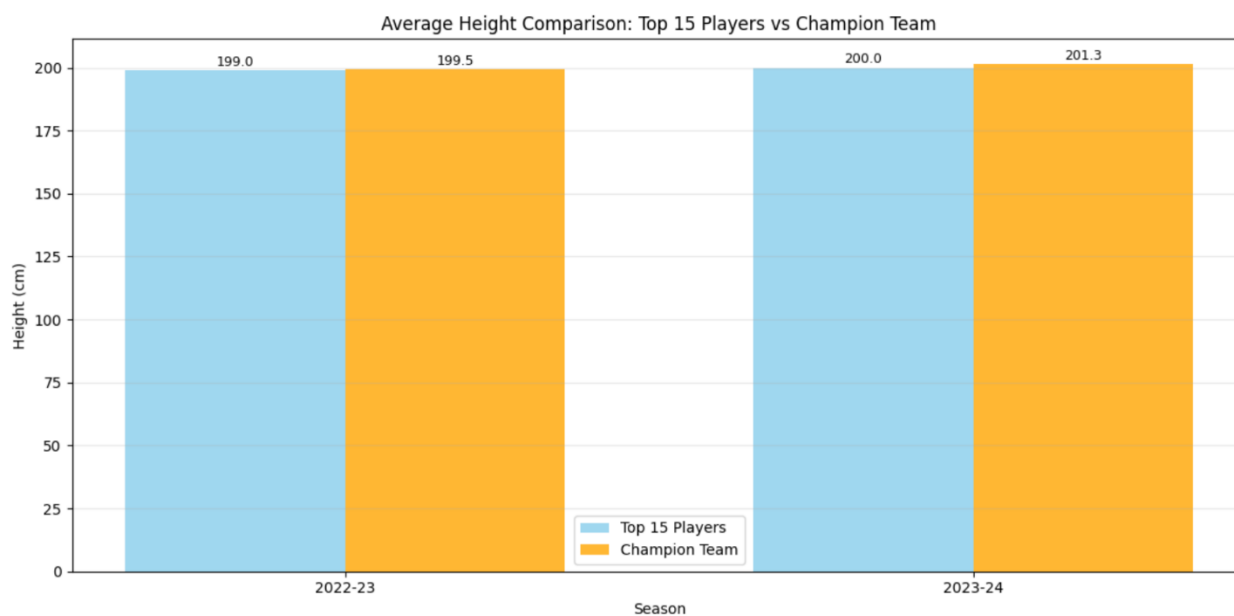
آمار توصیفی) سوال 1 :

در هر فصل میانگین قد ده بازیکن برتر فصل با کاندیداهای mvp مقایسه شده است.



آمار توصیفی) سوال 2 :

در دو فصل آخر میانگین تجربه و قد، ۱۵ بازیکن برتر فصل با بازیکنای تیم قهرمان مقایسه شده است.



آمار توصیفی) سوال 3 :

بازیکنایی که در پست point guard هستند و بیشترین نامزدی mvp رو داشتند و برای خرید به باشگاه پیشنهاد می شوند : (از فصل های ۲۰۱۹-۲۰ تا ۲۰۲۳-۲۴)

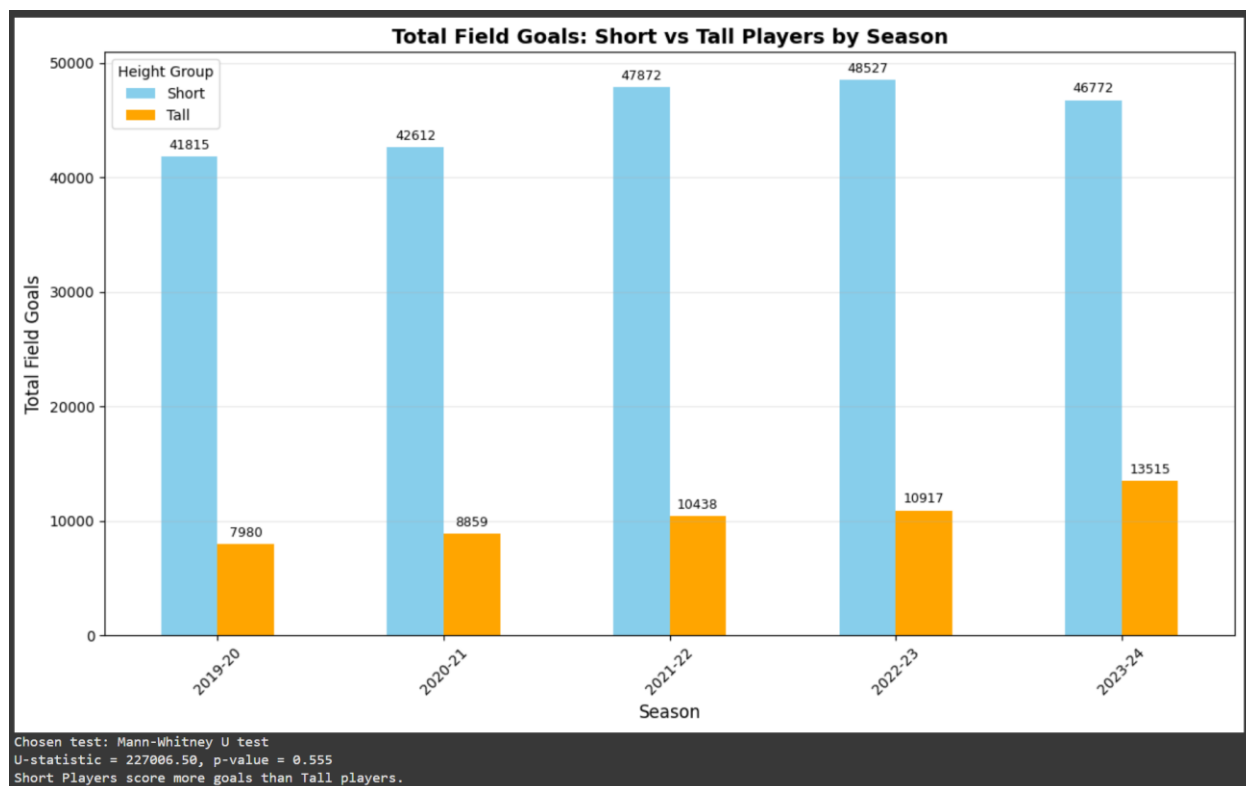
Player	count
Luka Dončić	5
Chris Paul	3
Stephen Curry	3
James Harden	2
Damian Lillard	2

dtype: int64

فرضیه ها) تست فرض :

فرضیه مورد بررسی :

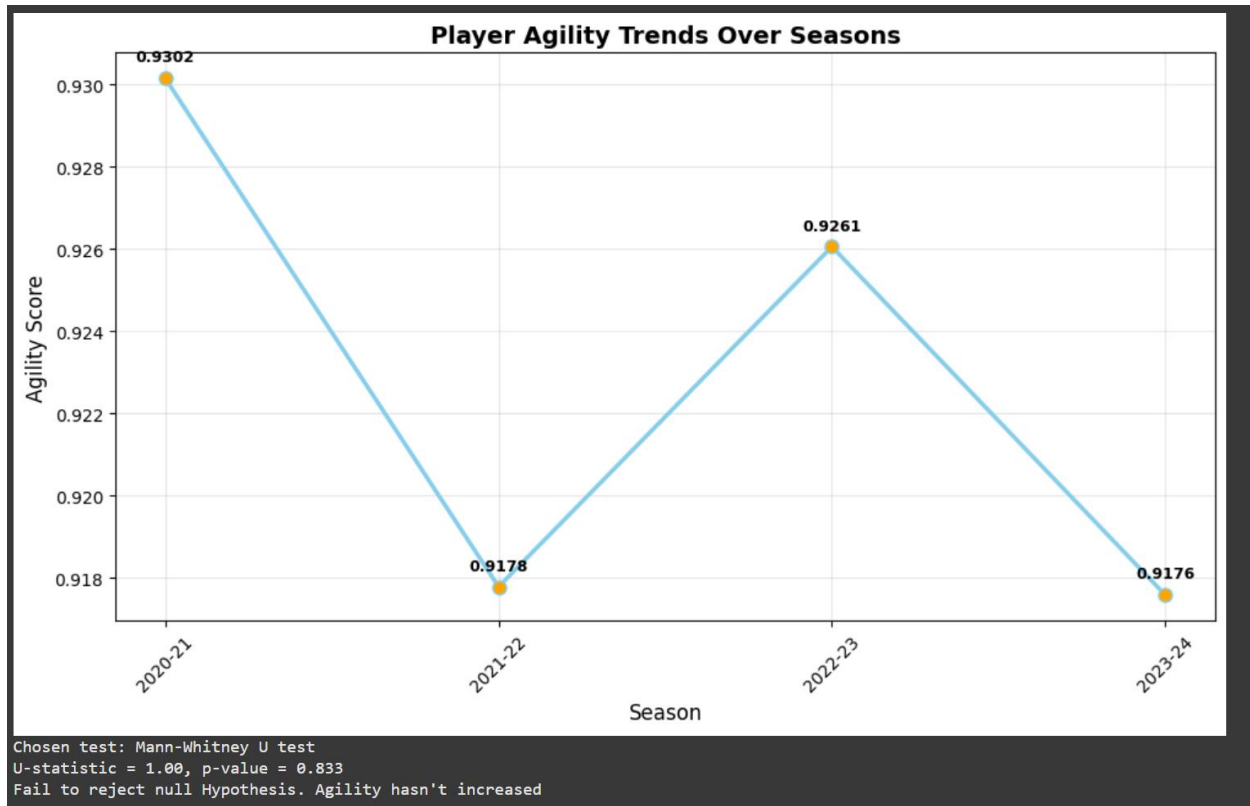
فرض صفر : مهاجم های قد بلند، گلهای بیشتری به نسبت مهاجم های قد کوتاه به ثمر می‌رسانند.



فرضیه ها) فرضیه 1 :

فرضیه مورد بررسی :

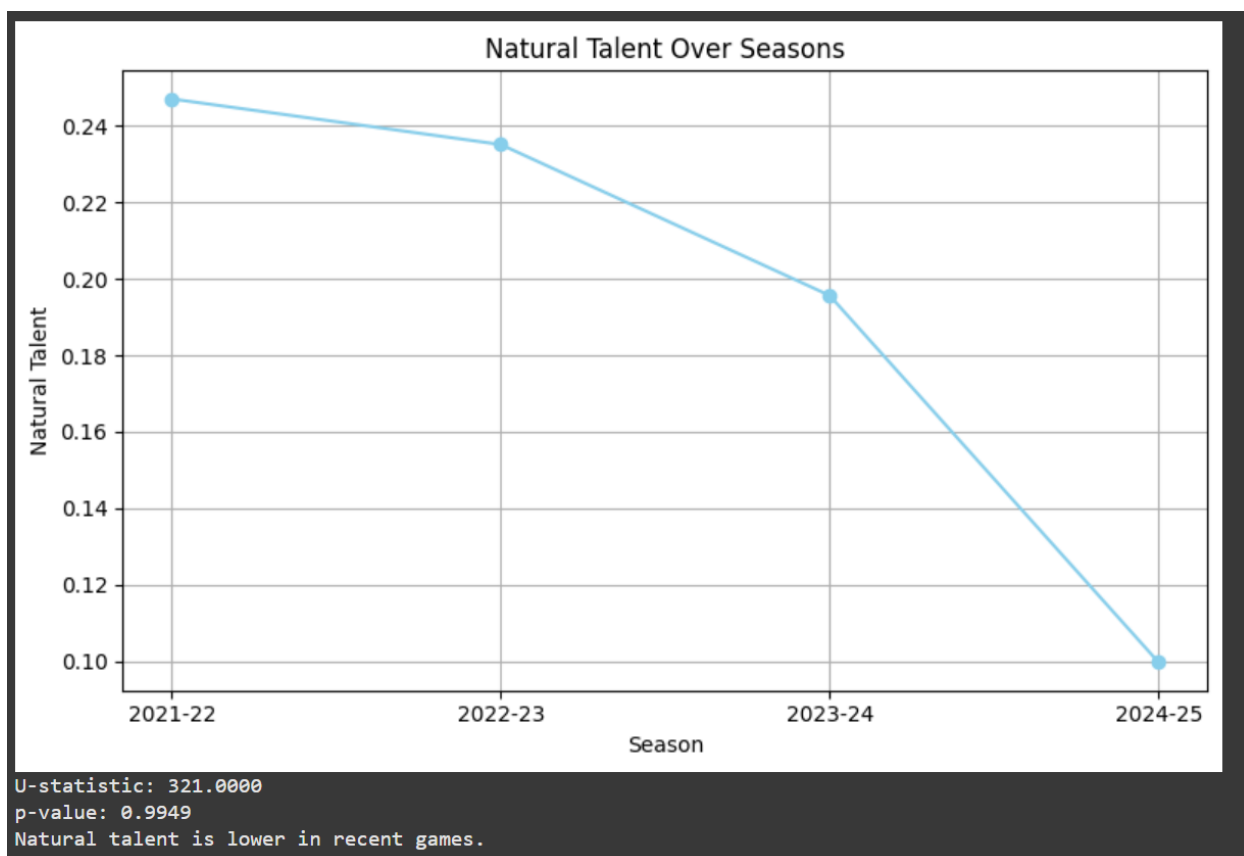
میانگین چابکی افرادی که در ۲۰ نفر اول هر فصل حضور دارند نسبت به گذشته افزایش یافته است.



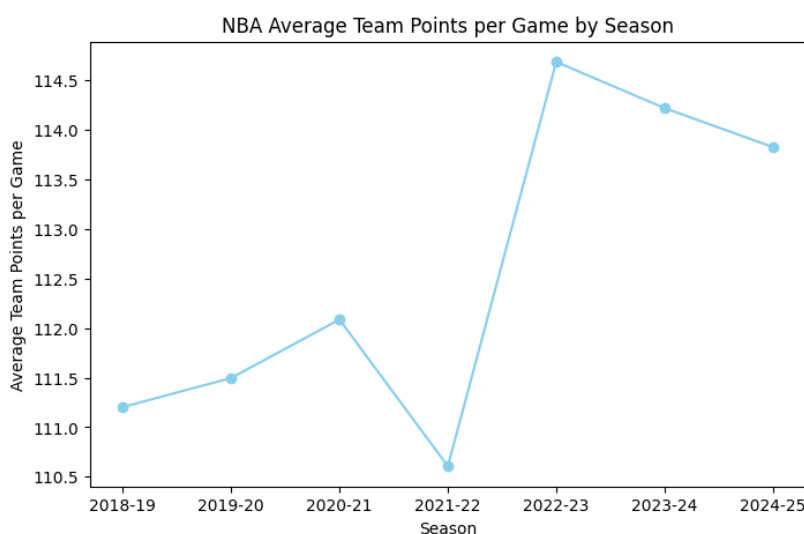
فرضیه ها) فرضیه 2 :

فرضیه مورد بررسی :

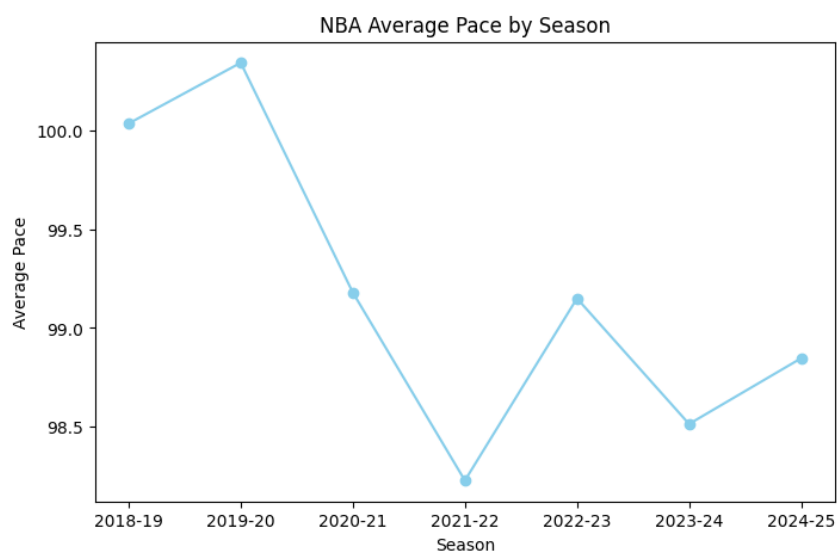
میانگین میزان توانایی بازیکنان تیم قهرمان نسبت به گذشته افزایش یافته است.



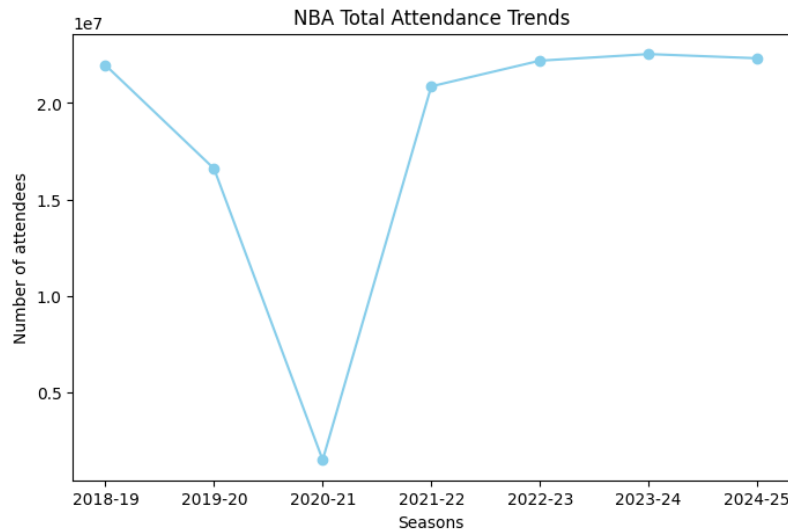
Visualization



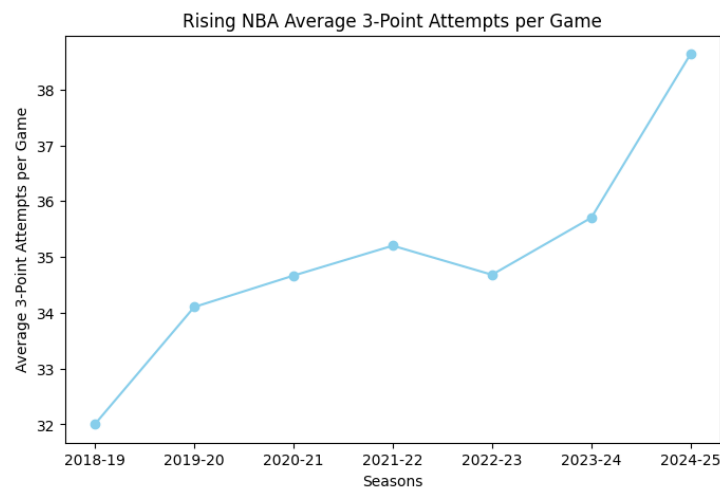
این نمودار روند میانگین امتیاز هر تیم در فصل‌های 2018-2019 تا 2024-2025 را نشان می‌دهد. میزان امتیازدهی از حدود 111 امتیاز در شروع دوره به 114-115 امتیاز در فصل‌های اخیر افزایش یافته و تنها در فصل 2021-2022 افت محسوسی داشته است.



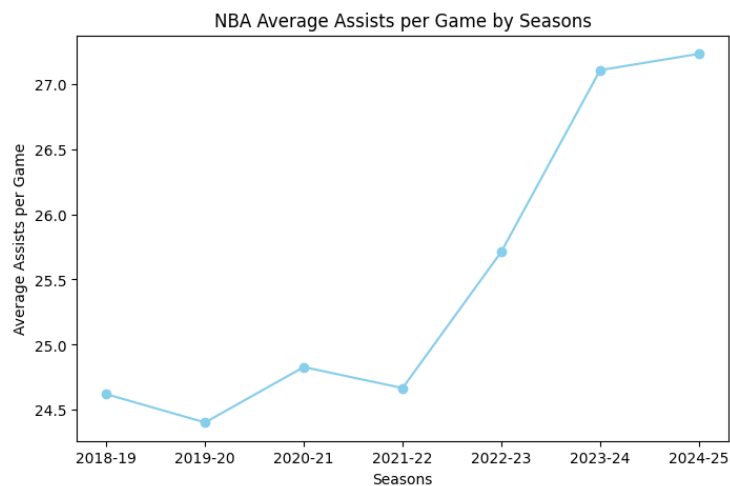
این نمودار نشان می‌دهد که تندی بازی در هر فصل چگونه تغییر کرده است که مشاهده می‌شود با وجود افزایش امتیازها، سرعت بازی اندکی کاهش یافته است.



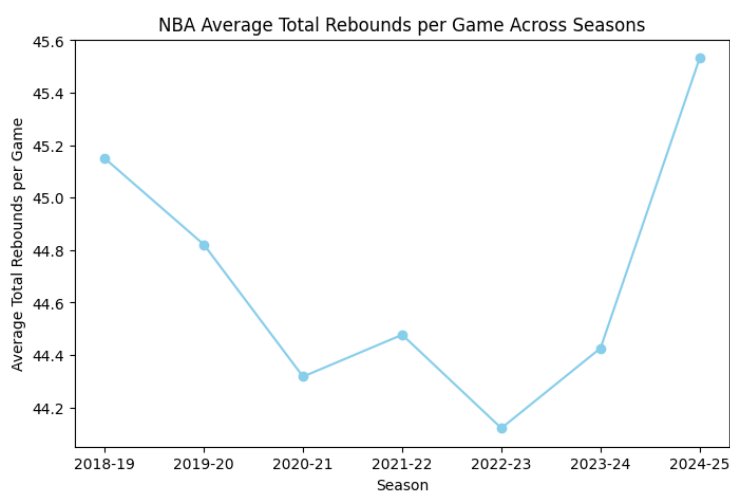
این نمودار نشان می‌دهد که با شروع بیماری کرونا و همه گیر شدن آن، حضور تماشاچی‌ها در سال 2020-2021 به شدت افت کرد (حدود 1 میلیون نفر). اما در فصل‌های بعدی، تعداد تماشاچی‌ها به حدود 21 میلیون نفر رسیده است.



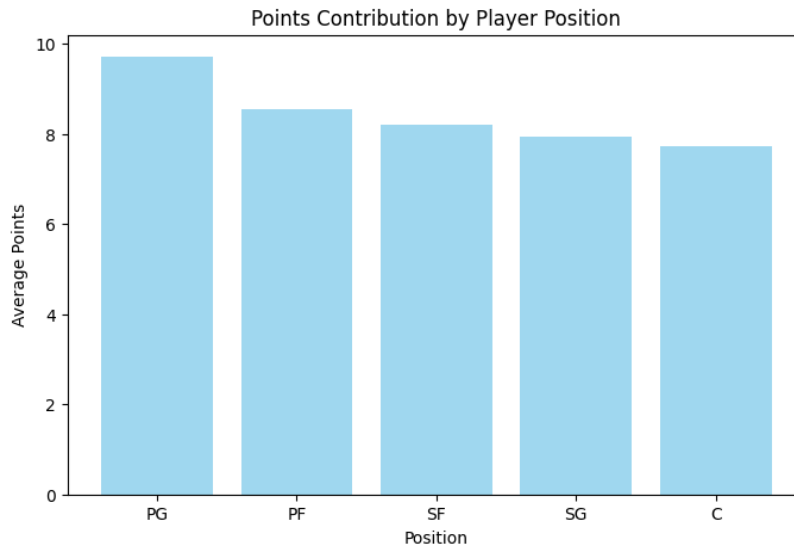
این نمودار نشان می‌دهد که میانگین تلاش تیم‌ها برای پرتاب سه‌امتیازی از حدود 32 در سال 2018-2019 به نزدیک 40 در سال 2024-2025 رسیده است. نتیجه این است که تیم‌ها در طراحی حمله بیشتر به خط سه‌امتیاز تکیه می‌کنند.



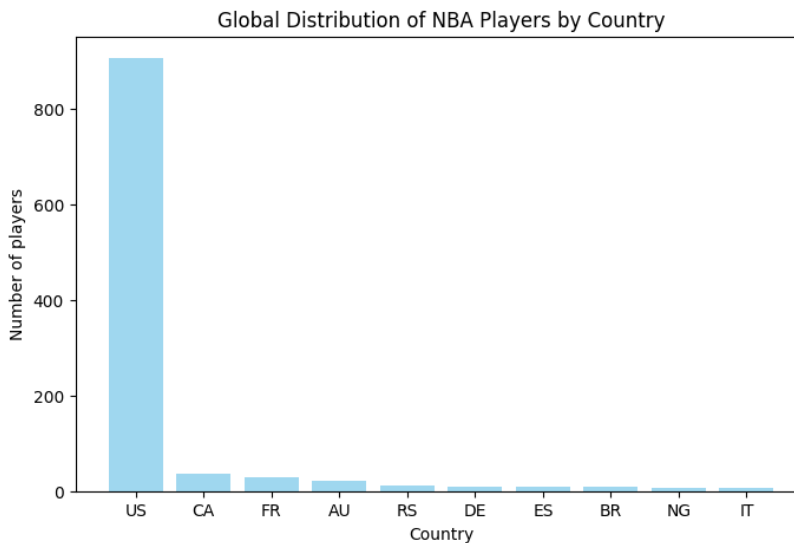
این نمودار نشان می‌دهد که میانگین پاس گل از حدود 24.5 در سال 2018-2019 به نزدیک 27.5 در سال 2024-2025 رسیده است. این افزایش می‌تواند بازتابی از افزایش همکاری تیمی باشد.



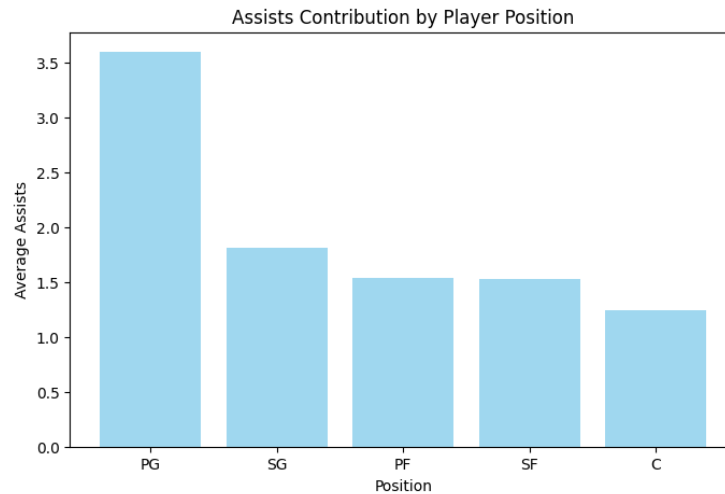
این نمودار نشان می‌دهد که میانگین ریباند کل تیم‌ها از حدود 45.1 در سال 2018-2019 به 44.1 در سال 2024-2025 کاهش یافته اما در سال 2024-2025 به حدود 45.5 رسیده است که این نشان دهنده افزایش قدرت ریباند کلی در سال‌های اخیر می‌باشد.



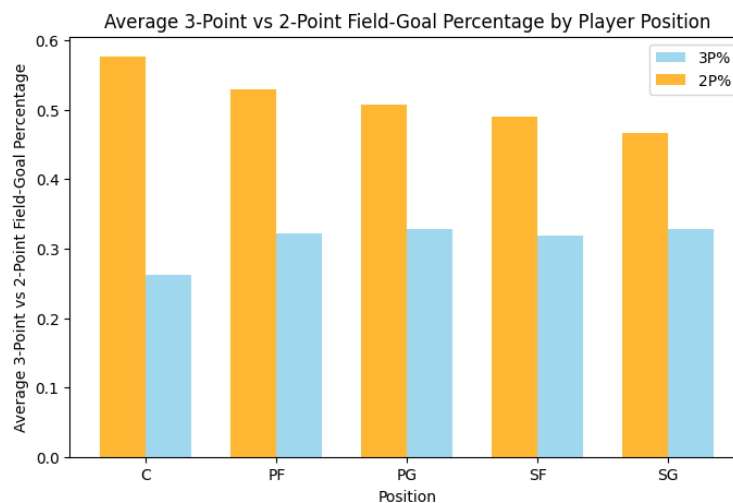
این نمودار، مقایسه نقش‌های مختلف در امتیازگیری می‌باشد. مقایسه میانگین امتیازها نشان می‌دهد که نقش‌های point guard (PG) با میانگین حدود 9.8 امتیاز بیشترین سهم را دارند، در حالی که نقش‌های center (C) حدود 7.5 امتیاز کسب می‌کنند.



این نمودار، سهم کشورهای مختلف از بازیکنان NBA در فصل‌های اخیر را بررسی می‌کند. مشاهده می‌شود که کشور آمریکا با اختلاف بیشترین تعداد بازیکنان را دارد. پس از آن، کشورهای کانادا، استرالیا و فرانسه قرار دارند.



این نمودار، مقایسه نقش‌های مختلف در پاس گل می‌باشد. در این نمودار نقش‌های point guard (PG) با میانگین حدود 3.6 پاس گل در هر بازی فاصله زیادی نسبت به سایر پست‌ها دارند.

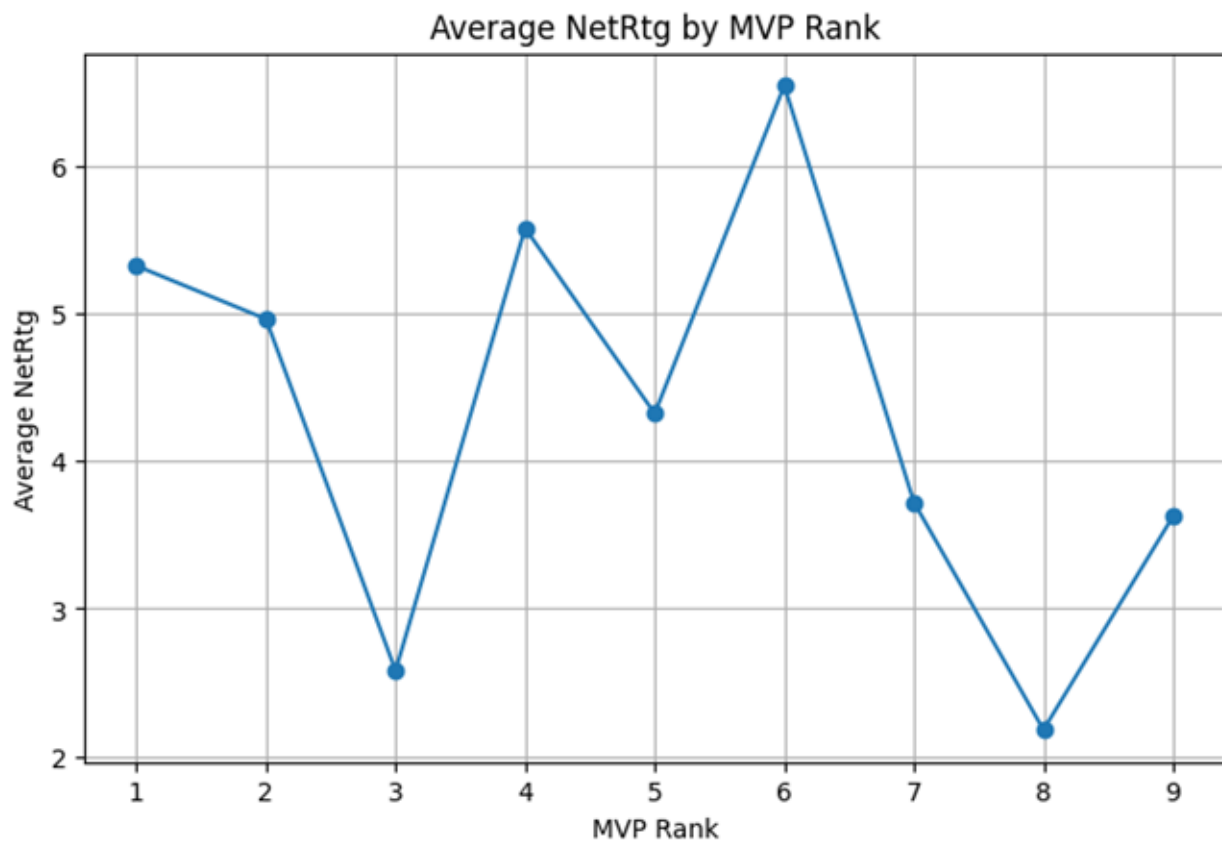


این نمودار میانگین درصد موفقیت شوت‌های دو امتیازی و سه امتیازی را برای هر پست نشان می‌دهد. مشاهده می‌شود که نقش‌های center (c) بیشترین موفقیت در شوت‌های دو امتیازی و کمترین موفقیت در شوت‌های سه امتیازی را دارند.

فرضیه های بیشتر

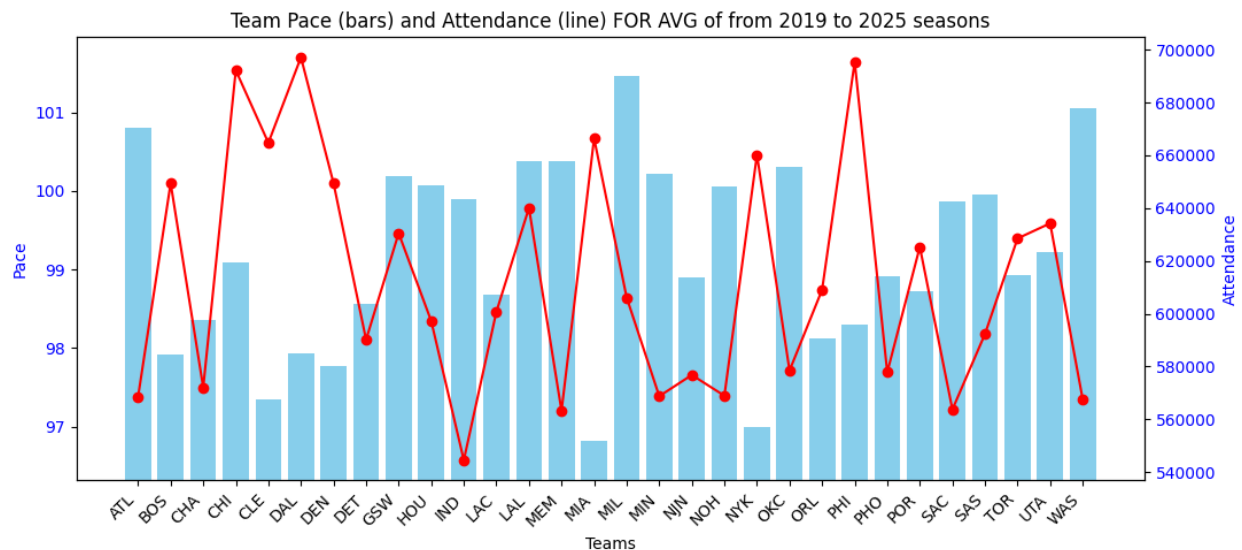
فرضیه 1: MVP در تیم هایی با NetRtg بالاتری هستند.

من ابتدا انتظار داشتم که بازیکن هایی که در 9 رتبه MVP هستند در تیم هایی بازی کرده باشند که NetRtg بالاتری داشته باشند. NetRtg یک شاخص هست که با توجه به عوامل مختلفی به تیم ها داده می شود و هرچه بیشتر باشد نشاگر بهتر بودن آن تیم است. با توجه به اینکه به طور چشمی رابطه ای در نمودار دیده می شود ولی بعد از تست مشاهده می شود که رابطه معناداری بین این دو متغیر نیست.



فرضیه 2: تیم هایی که سریع تر بازی می کنند، تماشاگر بیشتری دارند.

در ابتدا فکر میکردم مانند بازی فوتبال هرچه تیم ها سریع تر بازی کنند، تماشاگران بیشتری جذب تیم ها می شوند اما با توجه به داده هایی که از دیتابیس گرفتم متوجه شدم تعداد تماشاگرانی که بازی های تیم را مشاهده می کنند ارتباط معناداری با سرعت بازی تیم ندارد.

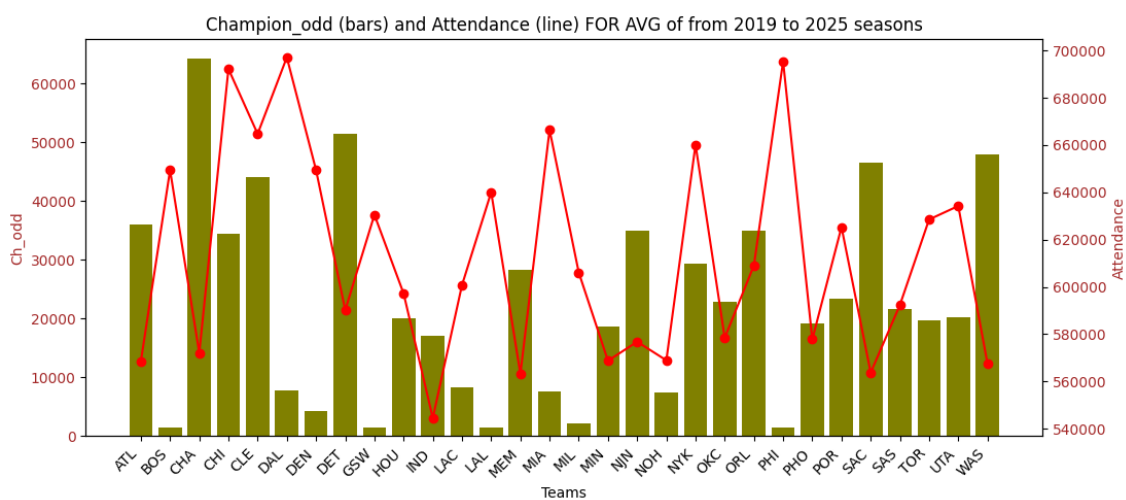


فرضیه 3: تیم هایی که ریسک بیشتری (سود بیشتر در شرط بندی) تماشاگر بیشتری دارند.

یک پارامتر جذاب که کمتر به آن توجه می شود، پارامتر **Champion_odd** هست. این پارامتر نشان دهنده این هست که اگر شما به طور میانگین یک دلار روی برد یک تیم شرط بندی کنید و همه بازی ها را ببرد، آن یک دلار شما به چه عددی میرسد. خب قاعدتا برای تیم هایی که از مدعیان سنتی NBA هستند، این پارامتر کمتر هست چون تیم قوی تری هست و احتمال برد آنها بیشتر است و سود کمتری در شرط بندی نصیب شما میشود.

به طور کلی هرچه این مقدار بیشتر باشد نشان دهنده این است که تیم در شرط بندی ریسکی تر هست. حال با این فرضیه قصد داشتم ببینم آیا اینکه یک تیم که ریسک بیشتری در شرط بندی دارد می تواند بازی های پرتماشاگر تری داشته باشد یا نه؟

با توجه به نمودار و آزمون انجام شده روی داده ارتباط معناداری بین این دو پارامتر نیست و تعداد تماشاگران به عوامل دیگری بستگی دارد.





پایان