NBA AGING CURVE

Introduction

The project is focused on understanding how NBA players progress with age, identifying peak performance ages, and providing developmental suggestions for players at specific ages and positions. It aims to equip teams with strategic insights for proposing the ideal contract lengths and timing for signing players. The project will leverage overall performance metrics such as BPM, OBPM, DBPM, and VORP.

The problem of understanding players aging and performance progression arises in multiple contexts within the NBA. The timing and scenario where this problem becomes particularly crucial include:

- Signing New Players
- Deciding Team Rosters
- Assigning Player Positions or Dropping Players
- Practice
- Draft Period

The development of age curves for basketball players is significantly relevant within the sports industry, offering critical insights for player development, team management, and performance evaluation. Understanding the nuances of how performance evolves with age enables coaches to design training programs that are precisely tailored to the individual needs of players. Meanwhile, team managers can leverage this data to make strategic decisions regarding player contracts and roster composition. By accurately predicting performance track based on age and position, teams can more effectively allocate resources, enhance player longevity, and maximize overall potential.

Data

The data used in the study includes statistics like BPM (Box Plus/Minus), OBPM (Offensive BPM), and DBPM (Defensive BPM) to measure how well basketball players perform. BPM gives a summary of a player's overall impact on the game, considering both offense and defense. OBPM focuses on offense, while DBPM looks at defense.

These statistics give a good picture of how players are doing in games. They cover a lot of important stuff like scoring, passing, rebounding, and defense. Using BPM and its variations gives a broad view of a player's skills. It's like seeing the whole picture of what a player does in a game.By breaking down into OBPM and DBPM, we can see specific strengths and weaknesses in offense and defense. This helps understand players better.

BPM might not include everything. For example, it doesn't count "screen assists," which are important for some players, especially Centers. This means some players might not get enough credit for what they do.

While there's DBPM for defense, it's hard to measure defense with just stats. Things like positioning and teamwork aren't fully covered. So, DBPM might not show the whole picture of a player's defense.
While BPM and these stats are useful, they're not perfect. They give a good overall view of players, but they might miss some important details, especially when it comes to defense.

The data is collected from basketball-refernce.com, 11 seasons player data from 2012-2023. It includes all traditional statistics and advanced stats. The data needs some cleaning up such as dropping rows with Null values, merging metrics, dropping unwanted columns. Rather than the basic cleaning, the project needs filtering out players who played consecutive seasons and with more than 20 games played.

Methodology

The Delta Method is a quantitative analysis technique used to understand the change in performance metrics over time. It's especially useful for studying the phenomena where individual measurements such as the performance of NBA players with age are expected to evolve. Additionally, the delta method gives an accurate result by removing survivorship. Survivorship bias occurs when an analysis only considers those elements that have "survived" or made it past a certain point, potentially leading to skewed or misleading results because it ignores those that did not survive. This can distort the true performance or characteristic of a group by focusing only on the successful case. Under the Delta Method, we made 3 different calculations for each age-bucket and the below is explanation for each calculation

1. **Delta**: For each player and performance metric, we calculate the year-over-year difference, or delta, to gauge how a player's performance changes from one season to the next. This calculation sheds light on the immediate impact of aging on performance.
2. **Chained_Dropoff**: By aggregating these deltas, we track the cumulative effect of aging across an athlete's career. This sequential summation of deltas from season to season illustrates the long-term trajectory of a player's performance, revealing patterns such as peak and decline phases.
3. **Smoothed_Chained_Dropoff**: Real-life data can be noisy. To address this, a smoothing technique like the Savitzky-Golay filter is used on the cumulative delta values. This process helps in identifying the underlying trends by filtering out short-term fluctuations and emphasizing the long-term performance curve.

Exploratory Analysis/Data Visualization

Analytical Question1: How are players at specific positions progressing with age when we look at the key metrics for their specific positions?

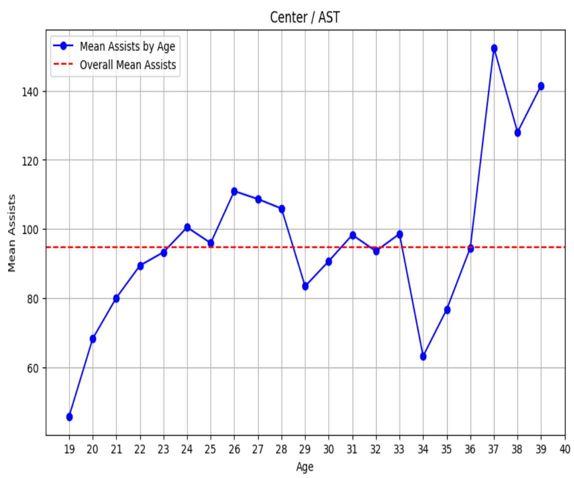
We utilize correlation analysis to identify the top 10 skills that exhibit the highest correlation with BPM, shedding light on the most relevant skills for players. Initially, we partition the dataset into five distinct dataframes, each corresponding to a different position (C, PF, SF, PG, SG). Subsequently, we rank the correlations separately for each position. Following our correlation analysis, we present the ranking table displaying the top 10 metrics with the strongest correlation to BPM:

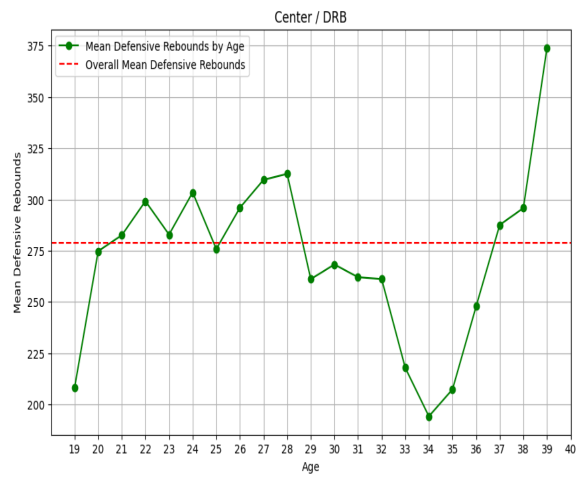
New Correlation Ranking with domain knowledge

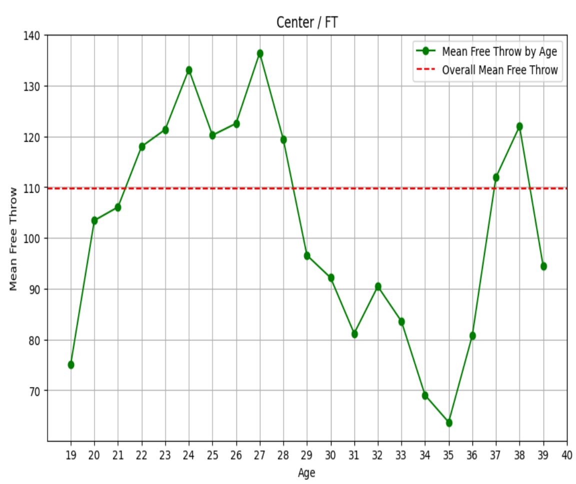
Rank	Center	Power Forward	Small Forward	Point Guard	Shooting Guard
1	AST	FT	AST	FT	FT
2	FT	FTA	FT	FTA	AST
3	DRB	AST	FTA	AST	FTA
4	STL	2P	DRB	DRB	STL
5	2P	DRB	TRB	TRB	TRB
6	FTA	2PA	2P	3P	DRB
7	TRB	TRB	STL	2P	2P
8	2PA	STL	2PA	3PA	2PA
9	BLK	BLK	3P	STL	3P
10	3P	ORB	BLK	2PA	3PA

Then, let's zoom in to Center and take the top 3 important ability for example (AST, FT DRB)

Simply plot line charts to see ability changes with age



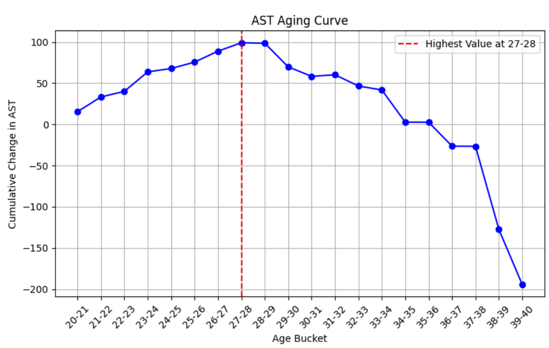


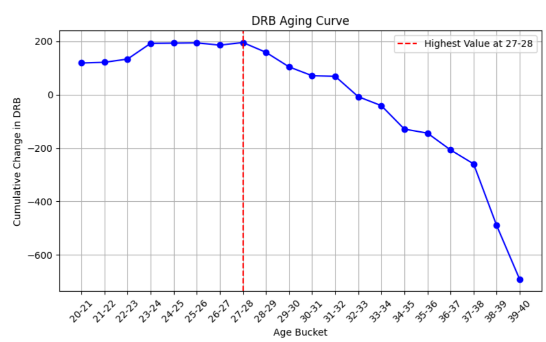


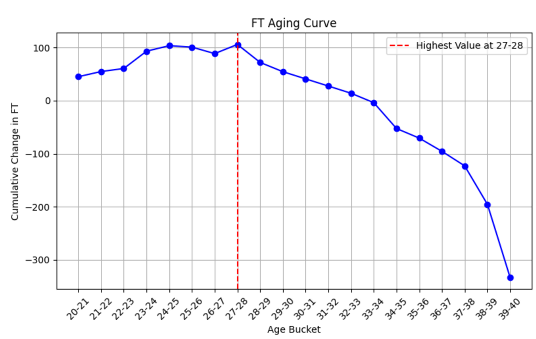
It's fascinating to observe the developmental trajectory of center players in their early career stages, particularly in aspects such as AST, DRB, and FT, where significant improvements are evident. However, as they reach the age of 30, we notice a decline in these areas. Notably, there's a conspicuous outlier at the age of 37, indicating a rare occurrence where only a select few players remain in the league at this age. These outliers typically comprise superstar players, leading to what's known as "survivorship bias". Below is a table displaying all players aged 37 who remained in the league from 2012 to 2023:

	Player	Age	Pos	G	AST	TRB	BLK
609	David West	37	C	73	138.0	238.0	75.0
610	LeBron James	37	C	56	349.0	459.0	59.0
611	Tim Duncan*	37	C	74	220.0	721.0	139.0
612	Chris Andersen	37	C	27	6.0	49.0	6.5
613	Kevin Garnett*	37	C	54	82.0	358.0	40.0
614	Taj Gibson	37	C	49	34.0	93.0	12.0
615	Pau Gasol*	37	C	77	238.0	619.0	79.0

To address this issue, we've employed the delta method to mitigate survivorship bias. As a result, the plots below exhibit a smoother trend, free from outliers:







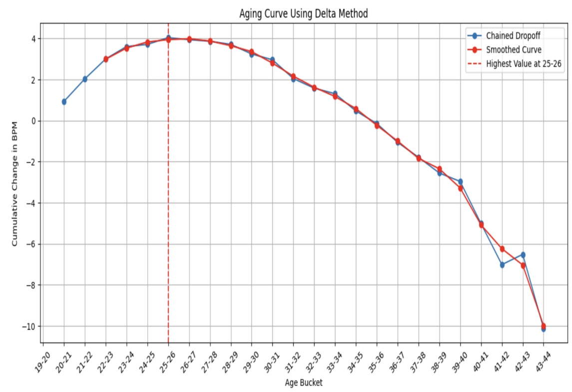
Analytical Question2: Using aging curves derived from player evaluation metrics(such a OBPM, DBPM, BPM, and VORP), what is the optimal age range representing a player's peak performance and the age range indicating deterioration?

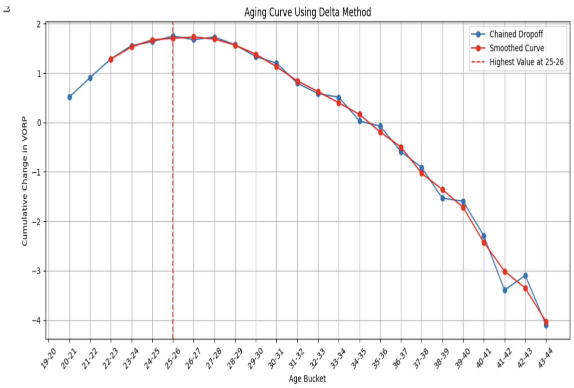
By applying the delta method to 4 different performance metrics, which are BPM, OPBM, VORP, and DPBM.

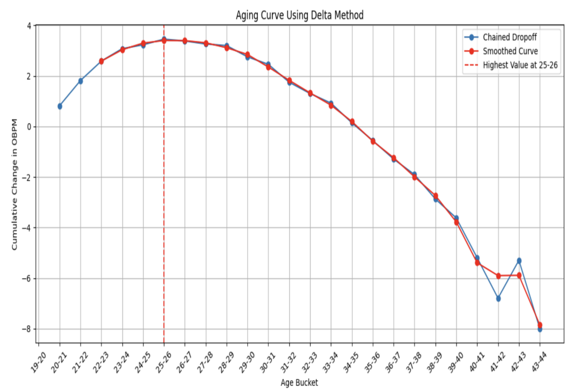
	Age_Bucket	BPM_Delta	OBPM_Delta	VORP_Delta	DBPM_Delta
0	19–20	NaN	NaN	NaN	NaN
1	20–21	0.933962	0.820755	0.516981	0.103774
2	21–22	1.096241	0.996617	0.391353	0.106767
3	22–23	0.969505	0.769231	0.376648	0.201923
4	23–24	0.604973	0.493481	0.270296	0.107728
5	24–25	0.107414	0.153046	0.081897	−0.045230
6	25–26	0.326610	0.223559	0.110452	0.103616
7	26–27	−0.097661	−0.070877	−0.070234	−0.020526
8	27–28	−0.075779	−0.120144	0.051379	0.040947
9	28–29	−0.153397	−0.062981	−0.157340	−0.089904
10	29–30	−0.465665	−0.444421	−0.239485	−0.023820
11	30–31	−0.284433	−0.297970	−0.131218	0.015059
12	31–32	−0.912297	−0.707520	−0.404268	−0.206301
13	32–33	−0.469084	−0.446056	−0.211832	−0.017303
14	33–34	−0.271635	−0.380929	−0.073558	0.102083
15	34–35	−0.839341	−0.778295	−0.478295	−0.061822
16	35–36	−0.610526	−0.702339	−0.109942	0.099708
17	36–37	−0.915217	−0.731522	−0.509783	−0.194565
18	37–38	−0.727083	−0.616667	−0.327083	−0.102083
19	38–39	−0.767857	−0.964286	−0.617857	0.203571
20	39–40	−0.412500	−0.750000	−0.062500	0.350000
21	40–41	−2.050000	−1.575000	−0.700000	−0.475000
22	41–42	−2.000000	−1.600000	−1.100000	−0.400000
23	42–43	0.500000	1.500000	0.300000	−1.000000
24	43–44	−3.600000	−2.700000	−1.000000	−0.900000

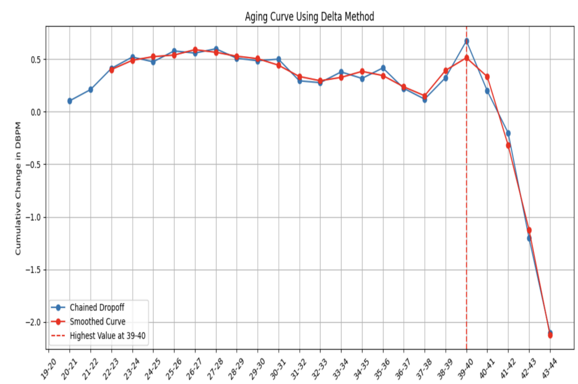
First, by focusing solely on the delta change, we observed three significant shifts in players' performance. The first is in the age bucket of 20-21, the second is 29-30, and the third is 34-35. It makes sense that in reality, players improve their performance in the early stages of their careers and experience a negative shift around the age of 29-30, leading some to even start retiring. Finally, at around the age of 34-35, most players retire from their careers.

Additionally, when comparing four different overall performance metrics, the magnitude of shift in VORP is relatively smaller than in BPM and OBPM, indicating that BPM and OBPM have higher potentials for performance improvement as well as decline. For players, teams, and coaches, if they establish strategy effectively, there is a high chance for players to improve their performance or to prevent a quick or sharp decline. However, if we look at DBPM, we can observe that there is no shift throughout the entire 11 seasons, implying that DBPM may be less susceptible to the rapid declines seen in other performance metrics









As evident from the aging curves for each of these three metrics, the Chained Drop-off values continue up to the 25-26 age bucket, while Smoothed curves seem to increase up to between 25-26 and 26-27, indicating that NBA players typically peak in performance around the ages of 25-27. After this period, there are some fluctuations, but the trend begins to consistently decline after the 28-29 age bucket. Consequently, we conclude that the optimal age range for peak performance in terms of BPM, OBPM, and VORP appears to be between 25-26 and 28-29 years. After 29, there is a more consistent drop in the Smoothed Chained Drop-off values, signaling the start of performance decline.

Unlike BPM, OBPM, and VOPR, DBMP shows a very different aging curve, characterized by smaller and slower fluctuations throughout a player's career. Notably, there is no consistent pattern of significant declines; instead, the deltas hover around minor improvements and sligh declines, with the most substantial drop occurring between age 42-43 at -1. Overall, DBPM suggests a more stable performance with less dramatic changes. However, DBMP has a limitation as it is simply calculated by calculating OBPM subtracted from BPM. Therefore, for more accurate prediction, we need advanced statistical metrics to assess defensive skills.

Analytical Question3: What advice can we offer to teams and agents regarding the best age for signing new contracts to maximize team benefit? Furthermore, when teams sign players around the peak or post-peak ages, what is the recommended optimal contract length to avoid financial inefficiency?

- Best Age for Signing New Contracts:

    - Before Peak: If the data shows that players improve up to a certain age, teams could aim to sign players just before they reach their peak. This could maximize the value the team gets from the player during the upswing of their career trajectory.
    - At Peak: Signing players at their peak might be most beneficial for immediate team performance, especially if the team is in 'win-now' mode, though the financial terms need to be carefully considered to ensure they match the expected contribution.

- Optimal Contract Length:
    - Players: For players at or near their peak, the contract length should ideally not extend far beyond their expected peak performance years to prevent overpaying for declining performance.
    - Younger Talent: For younger players with potential, longer contracts could lock in talent at a more favorable price, anticipating the player's improvement.
    - Veteran Players: Short-term contracts might be more appropriate for players past their peak to mitigate risks with potential decline and injuries.

Conclusion

- **NBA Teams:**
    - Decision-Making: Teams can use the insights to make informed decisions about draft picks, trades, and free agency signings, leading to a better-constructed roster and improved team performance.
    - Contract Negotiation: By understanding when a player is likely to peak or decline, teams can negotiate contracts that are financially efficient, ensuring they pay for performance that meets or exceeds the contract value.
    - Player Development: Insights into age progression can help teams create personalized training programs to maximize players' skills and potentially extend their peak performance period, leading to a more competitive team.

- **Team Sponsors:** Using insights from peak performance to estimate team strengths and make deals with top performers
- **Player Sponsors**: Player sponsors can craft marketing strategies that coincide with a player’s peak performance, leveraging their heightened visibility and success.
- **Event Sponsors:** Event Planning and Promotion: By knowing when players are likely to peak, event sponsors can plan promotions around games that are expected to showcase high-level performances, potentially increasing viewership and ticket sales.
    - Content Focus: Media outlets can use the insights to focus content on rising stars, veteran players who are maintaining performance, or discussing the impact of age on player’s careers.