Are top NBA performers also the most consistent?

Tyson Zullo, Daniel Dekker, Jesse Olivarez, James Wadsworth

Hypothesis

Alternative:
H_1: Top performers are more consistent than NBA players as a whole

Null:
H_0: There is no difference in the level of consistency of top performers vs. that of the league as a whole

Our Approach (1/2)

- **Data:** Per game team and players stats for 20 seasons (2004 - 2022)

- Limiting dataset to players participating in a ‘meaningful’ number of games
    - Games played = 82

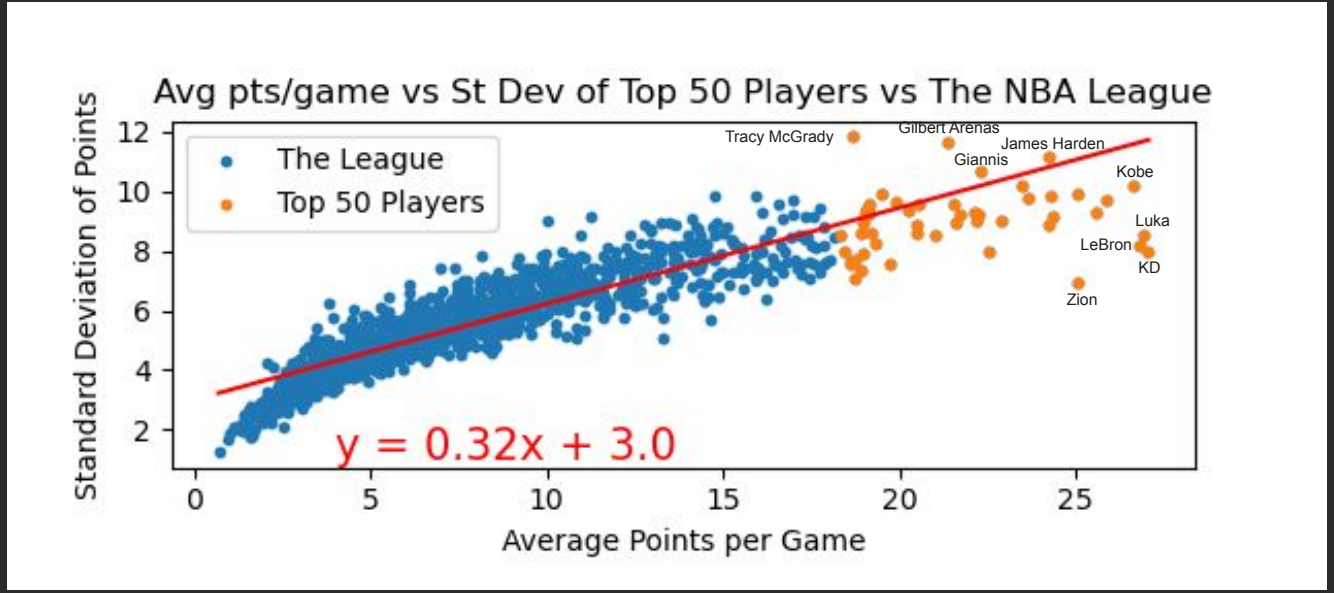
- Statistical categories of interest:
    - *Offensive Stats*
        - Points – **James**
        - Assists – **Daniel**
    - *Defensive Stats*
        - Rebounds - **Tyson**
        - Blocks - **Jesse**
        - Steals - **Jesse**
    - *Aggregate Stats*
        - +/- – **Daniel**

Our Approach (2/2)

For each of the six statistical categories, we did the following:

1. Identified the Top 50 Performers in the category based on per game average and designated them our "Top Performer" dataset
2. Compared the standard deviations of the Top Performer dataset to the standard deviations of the league as a whole for the given statistical category, using t-tests, data visualizations, and linear regression.
3. Drew conclusions about the relationship between performance and consistency based on the results

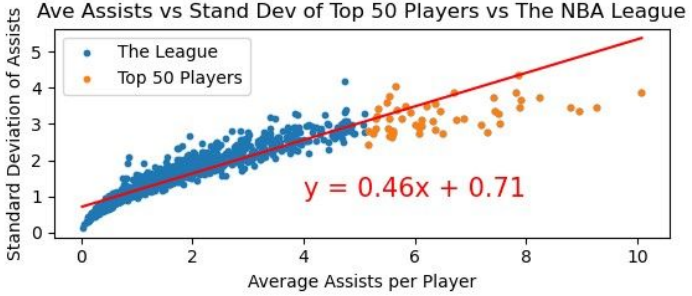
Points



- Top performers have a ***very*** different level of consistency game-to-game than the typical player – **p-value of 3.21e-28**
- **Top performers are far *less consistent*** with their scoring game-to-game than players across the league
    - Typical top performer std for points / game = 9.0 vs. 5.5 for the league as a whole
- However, the linear regression shows that the majority of **top performers are more consistent than expected** based on the league-wide performance (The r-value is: 0.81)
- Notable top performers include KD & LeBron, while TMac & James Harden are among the least consistent

Assists

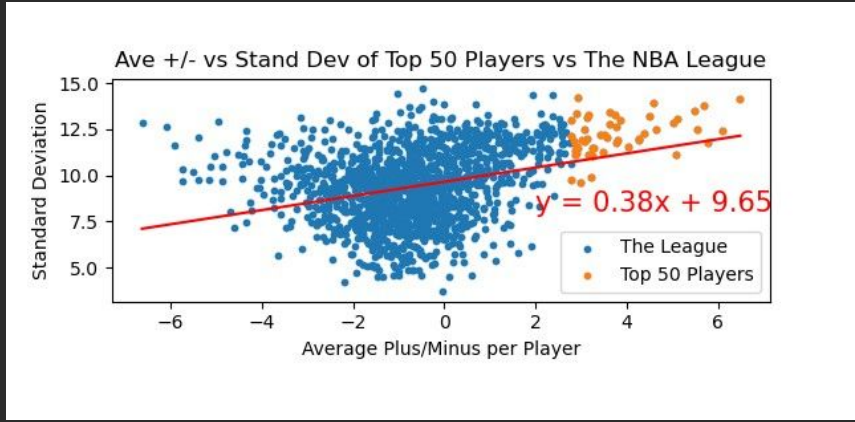
PLAYER_NAME	AST_mean	AST_std
Steve Nash	10.073529	3.876961
Chris Paul	9.268710	3.473118
Trae Young	8.940678	3.347552
John Wall	8.787795	3.454171
Russell Westbrook	8.233607	3.737975
Jason Kidd	7.905639	3.657935
Luka Doncic	7.879056	3.456424
Rajon Rondo	7.850515	4.361560
Deron Williams	7.807654	3.871716
Ben Simmons	7.515670	3.016935
Baron Davis	7.506780	3.294661
Ricky Rubio	7.426184	3.354380
Tyrese Haliburton	7.398844	3.741253
Ja Morant	7.308943	2.780780
LeBron James	7.176064	2.987609
LaMelo Ball	6.879195	3.134190
Allen Iverson	6.735358	3.126922
James Harden	6.690537	3.872836
Damian Lillard	6.490521	2.752530
Stephon Marbury	6.459948	3.353186
Stephen Curry	6.366290	2.826179
Darius Garland	6.366071	3.548499
Jrue Holiday	6.309930	3.089240
Nikola Jokic	6.209375	3.485257



- Top 50 NBA players have a higher standard deviation than the rest of the league, indicating less consistency
  - Top 50: 3.2 assists/game std
  - League: 1.5 assists/game std
- There is a significant difference between the performance of the top 50 players and the league: P-value 6.67e-33
- Most consistent top performers: Steve Nash, Chris Paul, Trae Young, Ja Morant, Lebron James
- There is a strong correlation between assists/game and the standard deviation r = 0.88
- Based on linear regression, top 50 are actually more consistent

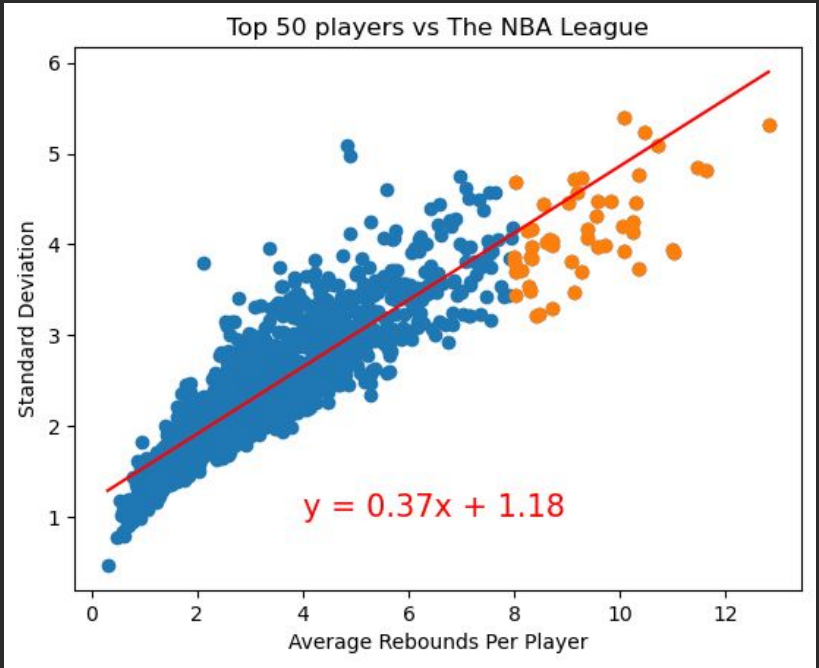
Plus / Minus

PLAYER_NAME	PLUS_MINUS_mean	PLUS_MINUS_std
Stephen Curry	6.489362	14.165745
Kawhi Leonard	6.088353	12.411627
Tim Duncan	5.770390	11.771770
Draymond Green	5.689310	13.801924
Joel Embiid	5.551807	12.520576
Klay Thompson	5.486423	13.530964
Jayson Tatum	5.111111	13.065925
Manu Ginobili	5.090909	11.096973
LeBron James	5.015421	12.881279
Steve Nash	4.636253	12.498781
Kevin Durant	4.592913	13.929794
Chris Paul	4.500371	13.221499
Chauncey Billups	4.372180	12.242759
Tony Parker	4.265541	11.893670
Reggie Miller	4.262857	11.955355
Danny Green	4.022549	11.583739
James Harden	3.860627	12.957973
Nikola Jokic	3.792187	11.904245
Kevin Garnett	3.786008	11.748808
Dirk Nowitzki	3.775781	12.603493
Rudy Gobert	3.752762	13.267059
Donovan Mitchell	3.639080	13.391339



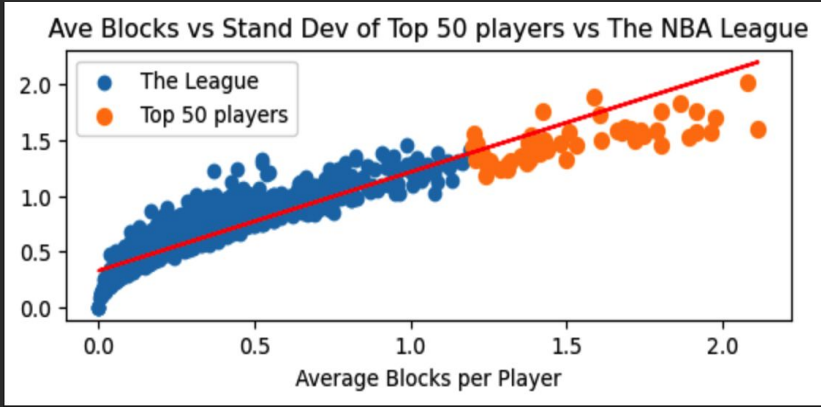
- Top 50 NBA players have a higher standard deviation than the rest of the league, indicating less consistency
  - Top 50: 12.2 +/- std per game
  - League: 9.4 +/- std per game
- There is a significant difference between the performance of the top 50 players and the league: P-value 1.08e-23
- Most consistent top performers: Jose Alvarado, Robert Horry, Tiago Splitter, Manu Ginobili
- There is a weak correlation between average +/- per game and the standard deviation r = 0.099
- Based on linear regression, top performers are less consistent

Rebounds



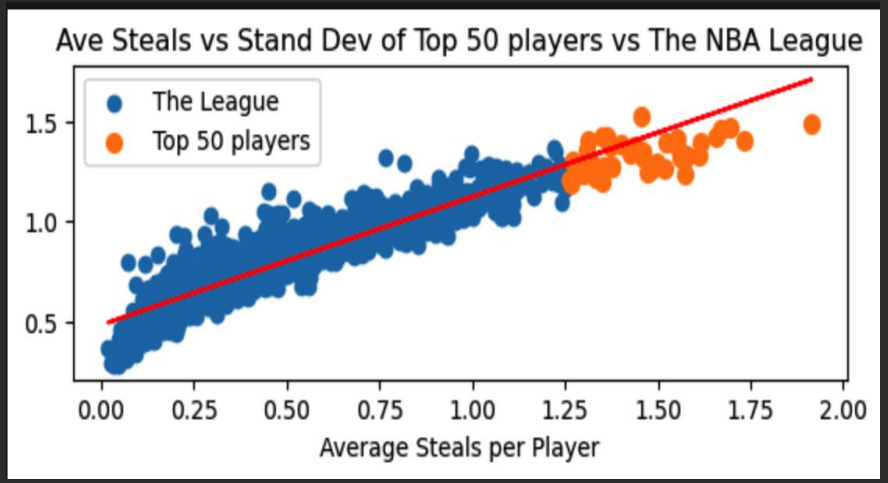
- The top 5 most consistent top performers: Evan Mobley, Luka Doncic, Wendell Carter Jr., Al Horford, and Julius Randle
- Is there a statistical significance between high performance and consistency?
- Ttest Result (statistic=23.16368326535982, pvalue=4.755563100474603e-28, df=49)
- The pvalue suggests strong evidence against the null hypothesis.
- r-value: 0.7906679736573208, indicates there is a very strong correlation between the two variables. The top performers tend to be more consistent when compared to the rest of the league.

Blocks



- The top 5 most consistent top performers were: Myles Turner, Marcus Camby, Anthony Davis, Rudy Gobert, and Alonzo Mourning
- Significance between high performance and consistency?
- `T-statistic: 32.34117091161321`
- `P-value: 2.650445452097839e-37`
- `Reject the null hypothesis: There is a significant difference in standard deviation.`
- r-value: 0.8623664653984808, indicates a strong correlation with both variables

Steals



Top 5 most consistent top performers were: Chris Paul, Jason Kidd, Ricky Rubio, Allen Iverson, and Baron Davis

- Significance between high performance and consistency?
- T-statistic: 42.973780114954124
- P-value: 5.479311290129507e-51
- Reject the null hypothesis: There is a significant difference in standard deviation.
- r-value: 0.8594005618731932, does indeed indicate a strong correlation between both variables

Major Findings & Implications

- Top Performers are not necessarily also the most consistent when individual statistical categories are examined, however:
- Top Performers are more consistent than expected compared to the league as a whole when looking at linear regression models
- We reject the null hypothesis that there is no difference between the top performers and the rest of the league in each of the categories.
- Assists have the highest correlation between performance and consistency
- Statistical analysis can have hidden bias

Appendix

Requirements

Completed Analysis Uploaded to GitHub (20 points)

- Final data analysis contains ample and complete information in README file (10 points)
- Final repository is acceptable for professional quality presentation (10 points)

Visualizations (20 points)

- 6–8 visualizations of data (at least two per question) (10 points)
- Clear and accurate labeling of images (5 points)
- Visualizations supported with ample and precise explanation (5 points)

Analysis and Conclusion (20 points)

- Write-up summarizes major findings and implications at a professional level (5 points)
- Each question in the project proposal is answered with precise descriptions and findings (5 points)
- Findings are strongly supported with numbers and visualizations (5 points)
- Each question response is supported with a well-discerned statistical analysis from lessons (e.g., aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis) (5 points)

Group Presentation (20 points)

- All group members spoke during the presentation (5 points)
- Group was well prepared (5 points)
- Presentation is relevant to material (5 points)
- Presentation maintains audience interest (5 points)

Slide Deck (20 points)

- Slides are visually clean and professional (5 points)
- Slides are relevant to material (5 points)
- Slides effectively demonstrate the project (5 points)
- Slides are clear and maintain audience interest (5 points)

Weighting rubric

****Scoring math:**

3-pointer made = 5 (3 for three points scored, 1 for a 3-pointer, 2 for a field goal made, and -1 for a field goal attempt; 3+1+2-1=5).

2-pointer made = 3 (2 for two points scored, 2 for a field goal made, -1 for a field goal attempt; 2+2-1=3)

Free throw made = 1 (1 for one point scored, 1 for a free throw made, -1 for a free throw attempt; 1+1-1=1)

Weighting rubric

Point = 1
3PM = 1
FGA = -1
FGM = 2
FTA = -1
FTM = 1
REB = 1
AST = 2
STL = 4
BLK = 4
TOV = -2

Point = 1
AST = 2
REB = 1
STL = 4
BLK = 4
TOV = -2

ESPN

Point = 1
REB = 1
AST = 2
STL = 4
BLK = 4
TOV = -2

Point = James
REB = Tyson
AST = 7/gm
STL = 1/gm
BLK = 1/gm
TOV = -3/gm

Yahoo

Points: 1
Rebounds: 1.2
Assists: 1.5
Steals: 3
Blocks: 3
Turnovers: -1

100% = 87

Point = 11.5%
AST = 17.2%
REB = 13.7%
STL = 34.5%
BLK = 34.5%
TOV = -11.5%

- Top Performers by Points - 50 - **James**
  - Are top avg points scorers the most consistent?
- Top Performers by Rebounds - 50 - **Tyson**
  - Are top avg rebounders the most consistent?
- Top Performers by Assists / +/- - 50 - **Daniel**
  - Are top avg assisters the most consistent?
- Top Performers by (Steals + Blocks) - 50 - **Jessie**
  - Are top avg (steals + blockers) the most consistent?

‘Most consistent’ = smallest aggregate std dev across 6 stat categories

- When we examine plus/minus, which is meant to measure the overall impact of an individual player on their team’s success, Top Performers are actually more consistent the higher their average performance – suggesting that in this aggregate measure, Top Performers by Plus/Minus consistently do what it takes to help their team win, regardless of how that shows up in individual stats.

Steps

1. First, find the top performers for your category by
    a. Excluding players who appeared in fewer than 82 games in the dataset
    b. Finding the per-game-average for your statistical category for each player, across the whole dataset
    c. Sorting the dataset by that average to find the top 50 performers for your statistical category
2. Then, calculate the standard deviation to measure consistency of the top 50
    a. Calculating the standard deviation for the statistical category for each of the top 50 performers
3. Finally, compare the consistency of top performers to the consistency of the population
    a. Calculate the standard deviation for that statistical category for the whole dataset of players (excluding those appearing in < 82 games)
    b. Use T-test to compare the top 50 sample standard deviations to the population standard deviations to determine if top performers are more consistent than the overall population