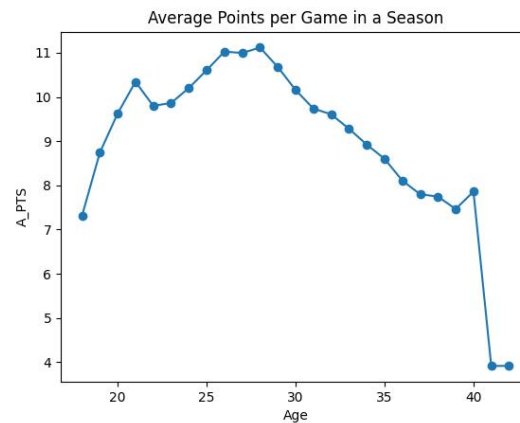# AIML401: Assignment 2 Task 2: NBA Statistical Analysis

Peyton Mou (2212984)
September 6, 2024

1. Average Performance Metrics by Age

Below are six graphs analyzing NBA players' different performance metrics per season by age respectively, and the valid statistics are 9902 rows of data in total, which includes only 2 rows of 18 years old and 3 rows of 41/42 years old.
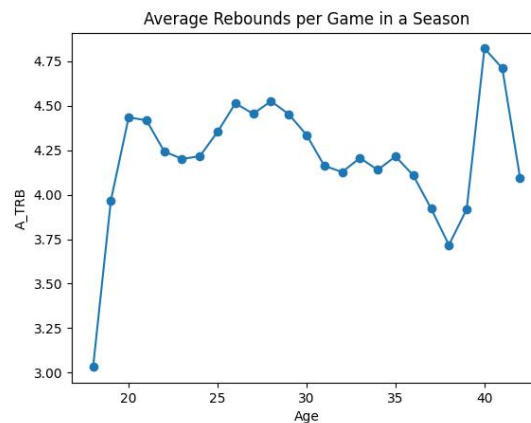

This graph displays players' average points gained per game in a season by age, whose peak is the players at age of 28, the principal value ranges from approximately 7 to 11 within 18-40 age range, and the lowest value is players elder than 40.
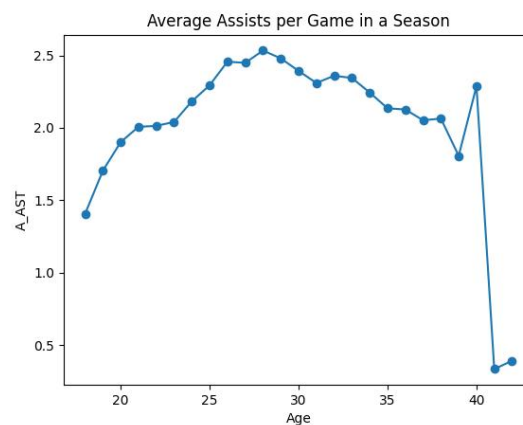

This graph displays players' average number of assists per game in a season by age, shares a similar line shape with the previous graph except a sudden rise from below 2.0 to above 2.0 at age of 40, whose peak is age of 28, the principal value ranges from approximately 1.4 to 2.5 within 18-40 age range, and elder than 40 has the lowest.

Average Blocks per Game in a Season
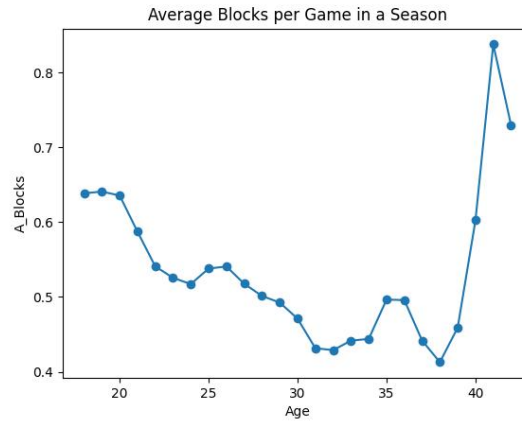
This graph displays players' average number of blocks per game in a season by age, with an extraordinary peak above 0.8 at age of 41 and a sub-maximal value of above 0.7 at age of 41, principal value ranges from approximately 0.4 to 0.65 within 18-40 age range, the age of 38 is the lowest.


Average Rebounds per Game in a Season

This graph displays players' average number of rebounds per game in a season by age, whose peak is above 4.75 at age of 40 and sub-maximal value is close to 4.75 at age of 41, principal value ranges from approximately 3.75 to 4.5 within 19-42 age range (except 40 and 41), and the bottom is around 3 at age of 18.

Average Efficiency Rating per Season
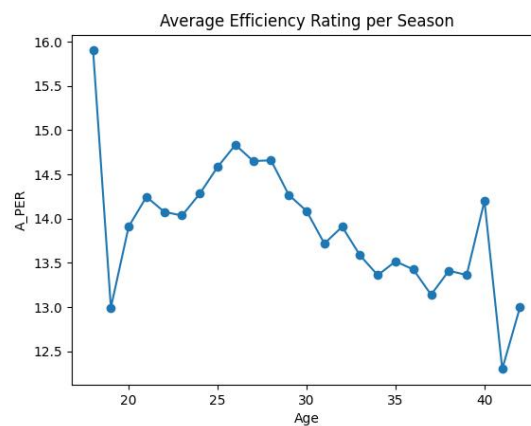
This graph presents players' average efficiency rating per season by age, the highest value is close to 16 at age of 18 and the lowest value is below 12.5 at age of 41, and the principal value range fluctuates from approximately 13 to less than 15 within

19-42 age range (except 41).


Average Shooting Percentage per Season
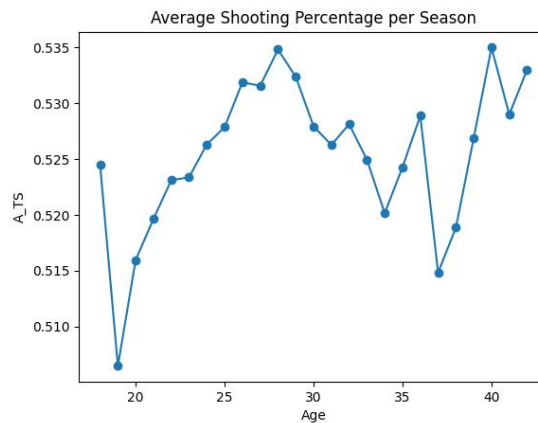
This graph presents players' average shooting percentage per season by age, the highest value is 0.535 at age of 40 and the sub-maximal value is extremely close at the age of 28, the principal value fluctuates from approximately 0.515 to 0.535 within 18-42 age range, except the lowest value around 0.5 at age of 19.

In a conclusion, age of 28 is obviously the golden time for players and gained two peaks and one sub-maximal value out of above six graphs, while age around 40 can bring some surprises in the aspects of A_Blocks, A_TRB, A_TS, A_AST), and the youngest age of 18 outperforms in aspects of A_PER and A_TS.

2. Average Performance Metrics by Age and Position
Below six graphs are NBA players' performance metrics by age and position, different color line represents different position (red is center, blue is small forward, green is shooting guard, yellow is power forward, and black is point guard):

A_PTS by Age and Position
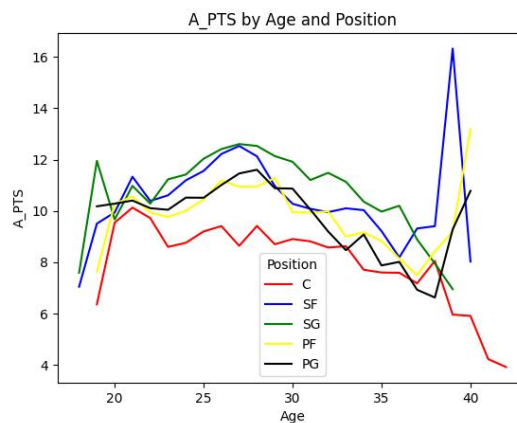
This graph shows average PTS per game in a season by age and position, centers peak at age of 22 and then gradually decline, the other four positions peak within 25-30 age range and decline quickly after the age of 35, and which noticeable is that point guards at age of 39 reached a dramatically high score (probably contributed by a few productive players).

A_AST by Age and Position
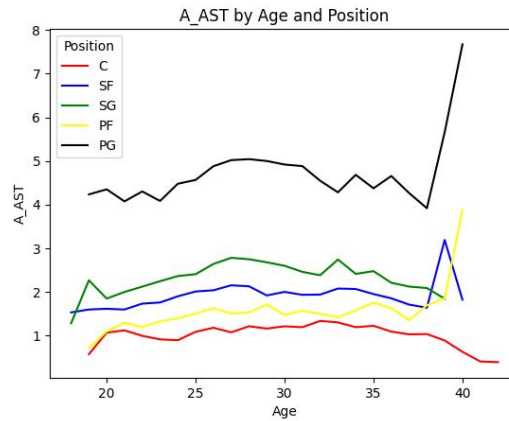
This graph shows the average AST per game in a season by age and position, point guards consistently lead in assists and peak at the age of 40, the other 4 positions maintain relatively low and stable performance across ages, and small forwards and power forwards reach the peak at the age of 39 and 40 respectively.
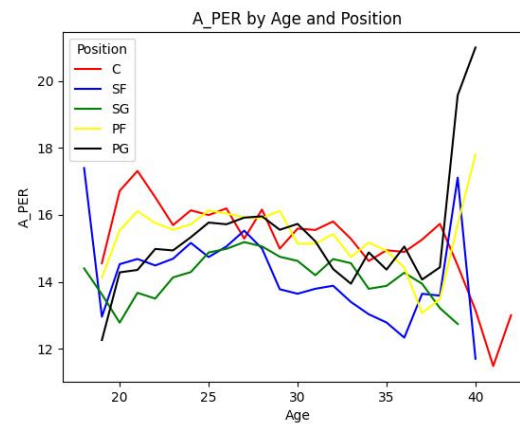

A_Blocks by Age and Position

This graph shows the average A_Blocks per game in a season by age and position, centers distinguish across ages and peak at age of 21, power forwards outperform and peak at age of 21 as well, small forwards peak at age of 18 and sharply decline, shooting guards and point guards consistently maintain low performance with minimal variation.

A_TRB by Age and Position

This graph show the average A_TRB per game in a season by age and position, centers and power forwards perform extraordinarily and shooting forwards reach

medium rebounds performance, which generally peak within 35-40 age range, shooting guards and point guards remain low and stable performance with minimal variation.

This graph displays the average A_PER per season by age and position, generally five lines fluctuated and coiled together within 25-37 age range, point guards, power forwards and small forwards increase sharply and peak within 38-40 age range, centers peak at age of 22 and shooting guards peak at age of 27.


This graph displays the average A_TS per season by age and position, whose outlook shares some similarities with the previous graph, generally five lines fluctuated and coiled together within 28-35 age range, point guards, power forwards and small forwards increase sharply and peak within 39-40 age range, centers peak at age of 22 and shooting guards peak at age of 37.

In a conclusion, centers and power forwards peak at their early 20s and distinguish in A_Blocks and A_TRB, small forwards and shooting guards peak late 30s and outperform in A_PTS, point guards tend to bring surprise around 40s and outperform in A_AST. Regarding A_TS and A_PER, influence from different positions are weakened, players attain high scores during their gold time around the age of 28, and some dramatic spikes around age of 40 can be observed.

3. Overall Performance Metrics by Position with 95% Confidence Interval
Below six bar graphs present the mean value of NBA players' overall performance

metric by position with 95% confidence interval (which represents the range we can be 95% confident that the true mean of the data lies, based on the sample data):

O_PTS Mean by Position with 95% Confidence Interval
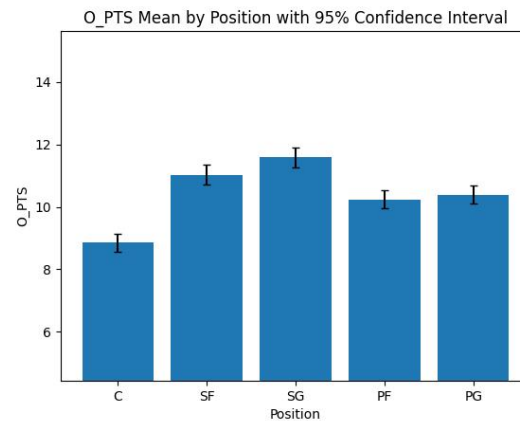
This graph shows O_PTS mean by position with 95% CI, shooting guards attain the maximal score and centers have the minimal score, while the max-min difference is less than 4, which does not construct a considerable gap.

O_AST Mean by Position with 95% Confidence Interval
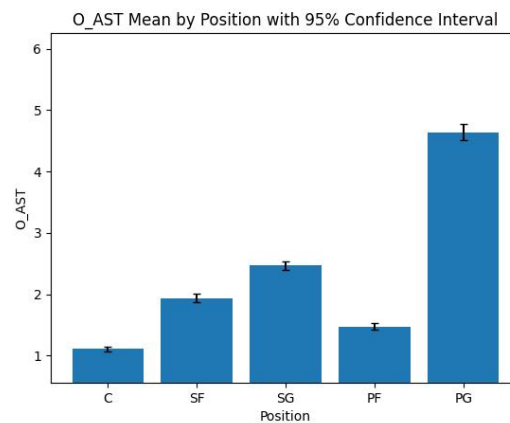
This graph shows O_AST mean by position with 95% CI, point guards achieve the maximal score close to 5, followed by shooting guards with around 2.5, and centers only have slightly above 1.

O_Blocks Mean by Position with 95% Confidence Interval
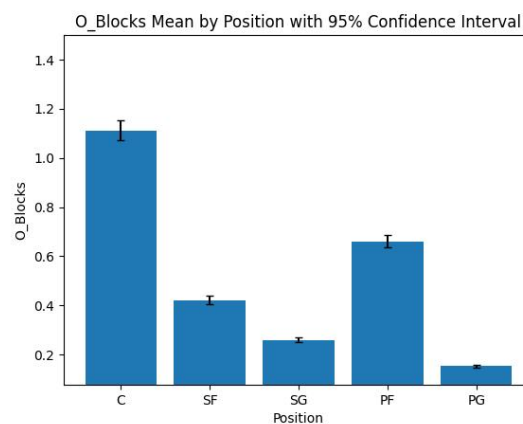
This graph shows O_Blocks mean by position with 95% CI, centers obviously distinguish with approximately 1.5, followed by power forwards with above 0.6, and the lowest one of less than 0.2 belongs to point guards.

O_TRB Mean by Position with 95% Confidence Interval
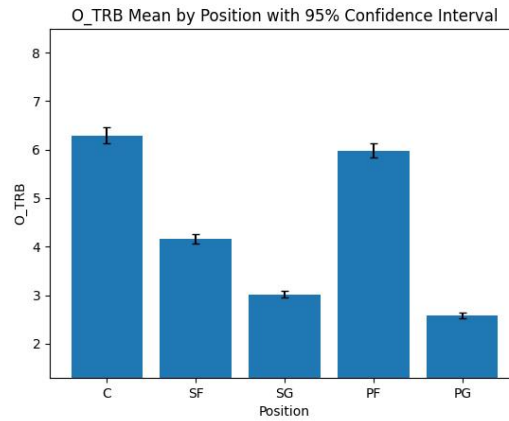
This graph shows O_TRB mean by position with 95% CI, centers and power forwards (which rank the top 1 and 2 respectively with very close gap) have exceeded the other 3 positions with a wide margin, and the lowest score belongs to point guards.

O_PER Mean by Position with 95% Confidence Interval
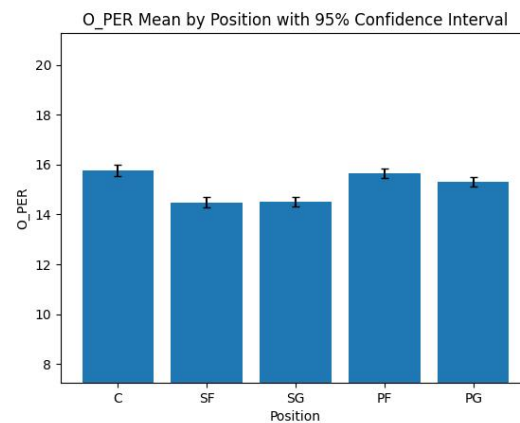
This graph shows O_PER mean by position with 95% CI, all 5 positions are neck-and-neck, centers slightly peak with a score almost 16, small forwards slightly bottom out with a score above 14.

O_TS Mean by Position with 95% Confidence Interval
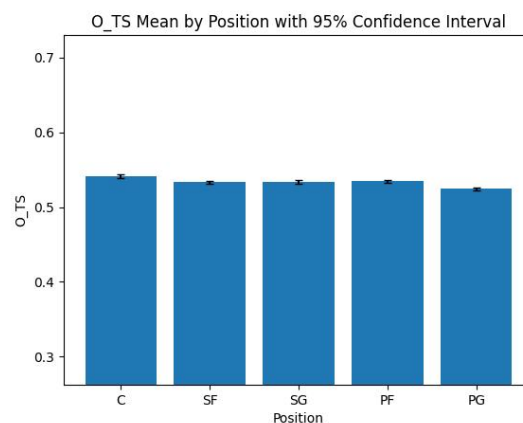
This graph shows O_TS mean by position with 95% CI, similar to the previous graph, at first glance it seems that all 5 positions are equally matched, and statistically the best performer is centers (0.5412), and the lowest one is point guards (0.5241), and the max-min difference is extremely small.

In a conclusion, we can tell players in different positions generally have neck-and-neck O_TS and O_PER, however their other performance levels vary when play in different positions, e.g. centers tend to stand out in O_Blocks and O_TRB, point guards usually excel at O_AST, and shooting guards exceed in O_PTS.

4. Westbrook's Triple-Doubles

Chi-squared test is utilized to evaluate whether Event X (team won) and Event Y (Westbrook got triple doubles) are independent, statistical results as below:

Items Calculated	Returned Statistics
Probability of Y	16.53%
Probability of X and Y	9.26%
Conditional Probability of X given Y	55.98%
Probability of X not Y	2.77%
Conditional Probability of X given not Y	3.32%
Chi-squared Value	6.0537541087487985e-34
P-Value	1.0
Degrees of Freedom	1
Expected Frequencies	[[0.01988363 0.14546447] [0.10036953 0.73428237]]

According to the specific sample data and calculation results above, chi-squared value is around 6.14, meaning there is discrepancy between the observed and the expected if the null hypothesis were true. However, p-value is 1.0, which is much greater than 0.05 (the standard significance level) indicating that Westbrook's triple-doubles do not significantly increase the likelihood of his team winning, and the null hypothesis (X and Y are independent) cannot be rejected.