Salary Estimation

Predicting WNBA & NBA salaries from performance statistics

01.

Background

Pay Inequality

NBA Pay Inequality

NBA players are paid 50% of league revenue while WNBA players are only getting about 22% of their league revenue. WNBA players deserve 50% of their league revenue, just as the men do.

USWNT Pay Inequality

US women's soccer team won settlement; US soccer federation pledges to equalize pay for men's and women's teams.

Necessary Perfection

Perpetuating the idea that female athletes must be above and beyond perfect to be considered equal to men.

02.

Hypothesis

Hypothesis

WNBA players' salaries are highly correlated with FG% because society rewards perfection from women. NBA players' salaries are highly correlated with minutes played because pay is often dependent on media coverage and viewership.

03.

Methods

Methods

Scraping

BeautifulSoup package, Python
Websites: CBS Sports, ESPN,
www.basketball-reference.com,
www.spotrac.com
Conversion to CSV
Cleaning in Excel

Testing Regressions

Different regression methods through SKLearn

Correlations

Between individual variables and salaries for NBA and WNBA players to determine significance

Fitting Data

Fitting performance data and salaries to selected regression to see weight each stat has on determining salary

Analysis

Analyzing validity of regression by looking at residual plot to determine if linear regression is best

04.

The Data

All Statistics

‘GP’

Games played

‘Total Mins’

Minutes played (season)

‘Total Points’

Total points scored

‘MPG’

Average minutes played per game

‘PPG’

Average points scored per game

‘GP%’

Field goal percentage (average)

‘GS/GP’

Games started (%)

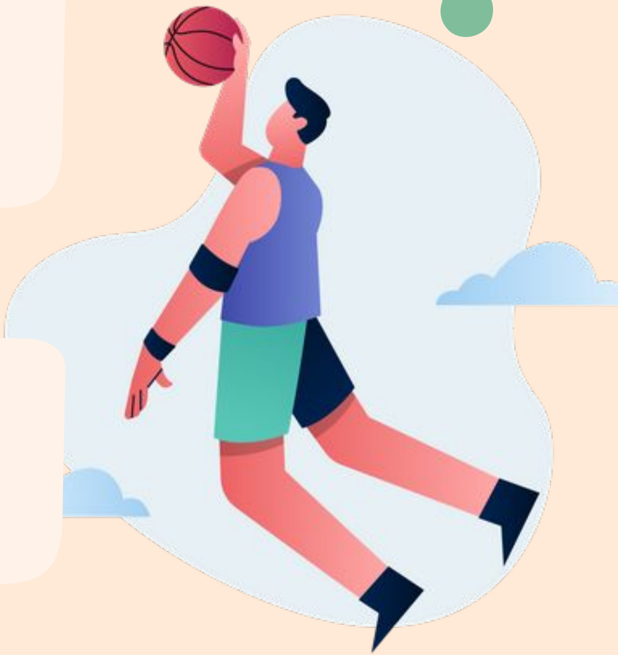
Stats Used in Regression

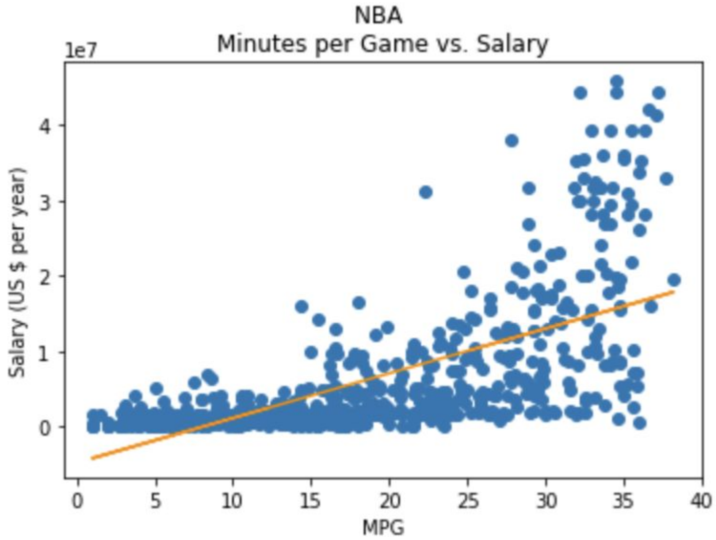
MPG

PPG

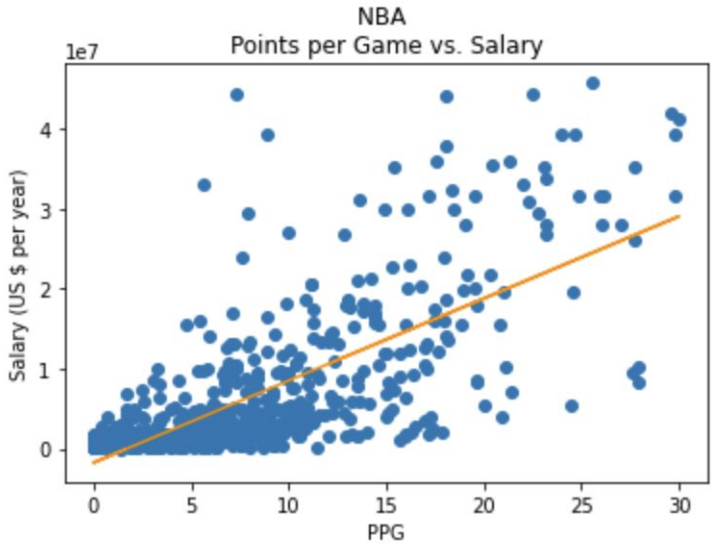
FG%

GS/GP

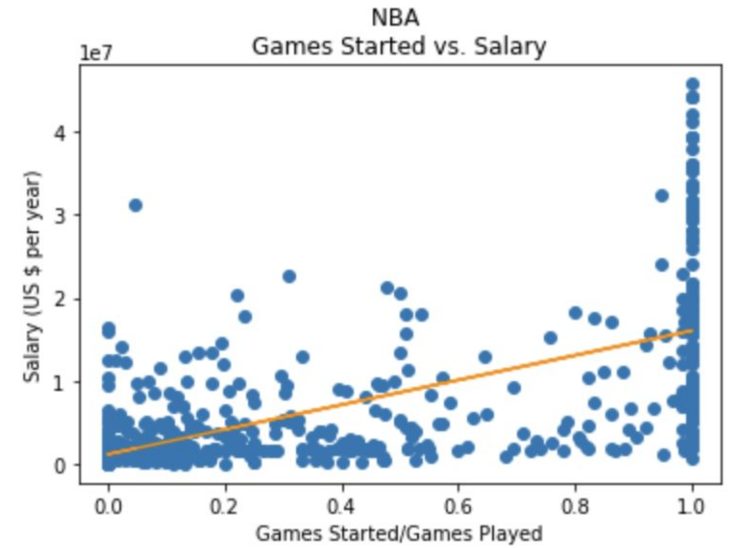




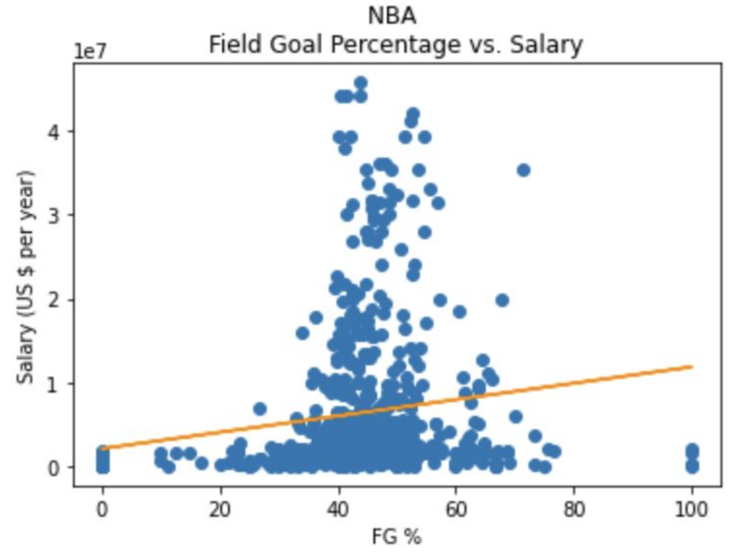
Pearson's correlation coefficient: 0.6505251815336721



Pearson's correlation coefficient: 0.7236124078364726



Pearson's correlation coefficient: 0.6510075843112078



Pearson's correlation coefficient: 0.15534028340766423

Testing regressions for NBA data

Regression	Score
Linear Regression	0.55133133
Ridge	0.55130294
Tweedie Regressor	0.52958898
Lasso	0.55133132
Elastic Net	0.53176275
Lasso Lars	0.55133132

Linear Regression:

Annual Salary = 510089.46 + 6559401.72(**GS/GP**) +
+ -102181.96(**MPG**) + 876426(**PPG**) + -33801.62(**FG%**)

Annual Salary = 510089.46 + 6559401.72(**GS/GP**) + -102181.96(**MPG**) + 876426(**PPG**) + -33801.62(**FG%**)

Interpretations

As the ratio between games started and games played increases by 1, the player's salary is expected to increase by $6,559,401.72.

As the player's average minutes played per game increases by 1 minute, the player's salary is expected to decrease by $102,181.96.

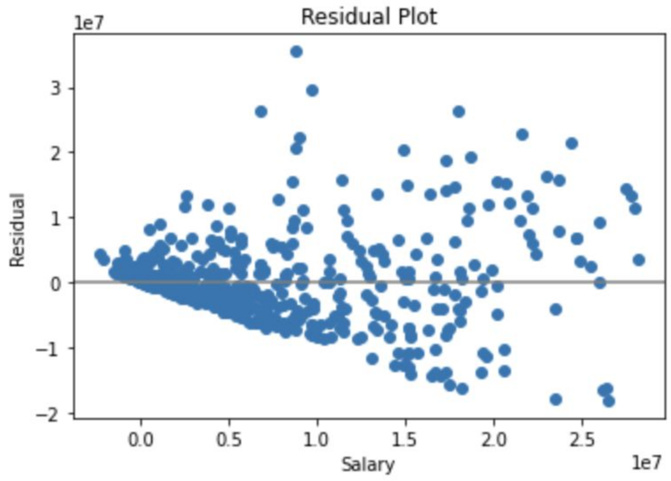
As the player's average points scored per game increases by 1 point, the player's salary is expected to increase by $876,426.52.

As the player's average field goal percentage increases by 1 percentage point, the player's salary is expected to decrease by $33,801.62.

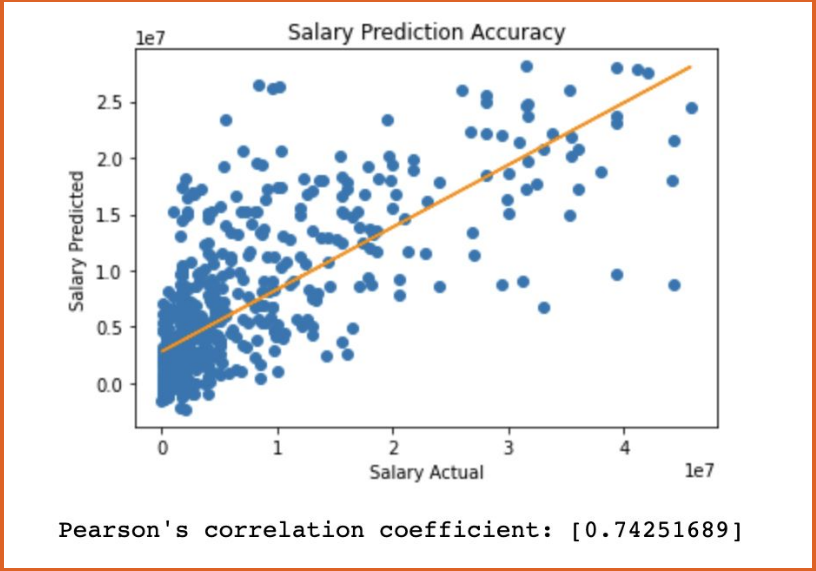
Annual Salary = 510089.46 + 6559401.72(**GS/GP**) + -102181.96(**MPG**) + 876426(**PPG**) + -33801.62(**FG%**)

Analysis of Regression

The residual plot has a very low correlation coefficient. There doesn't seem to be any distinct pattern of residuals in the data.
Any differences between the salary prediction and the actual salary tend to be randomly dispersed.
There is no bias in variability. A linear regression method is a good method to classify the data.



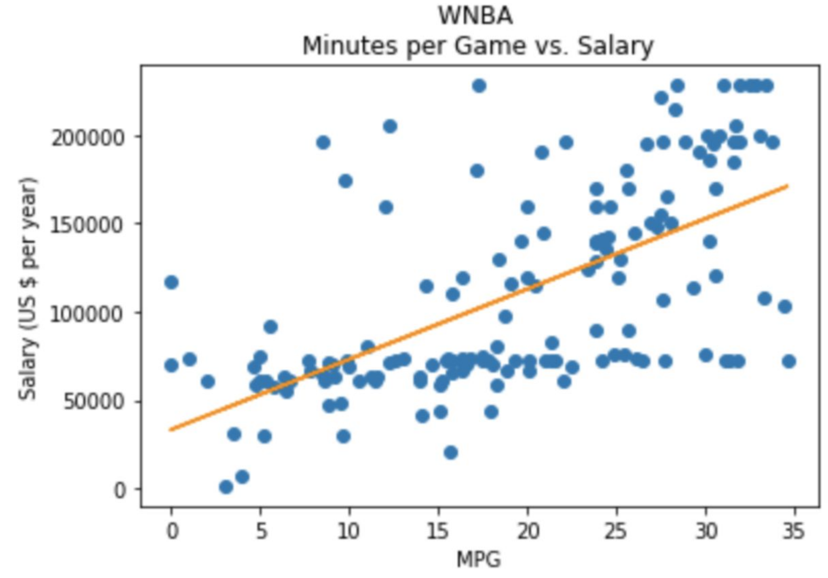
Pearson's correlation coefficient: [5.28595346e-17]



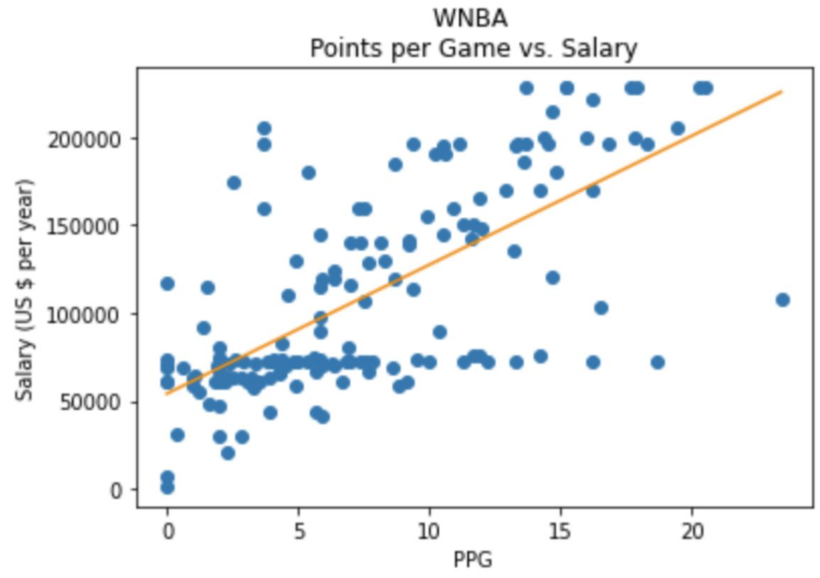
Analysis of Regression

There is a relatively strong, positive correlation between the actual salaries of NBA players and the predicted salaries, as determined by our regression. This indicates that our regression was relatively consistent in predicting salaries.

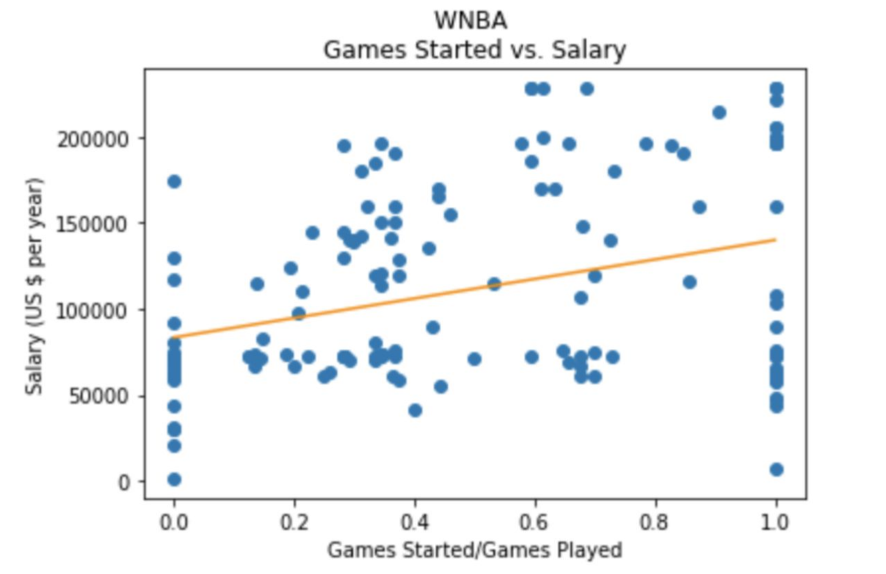
WNBA Variable-Salary Correlations



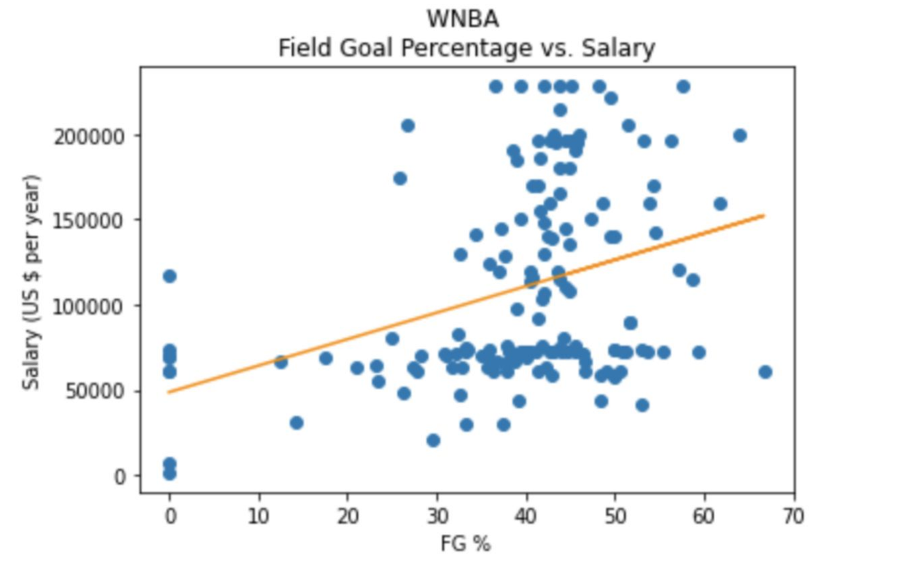
Pearson's correlation coefficient: 0.625832188713717



Pearson's correlation coefficient: 0.682040489012816 9



Pearson's correlation coefficient: 0.349840487358062



Pearson's correlation coefficient: 0.34531952907695157

Testing regressions for WNBA data

Regression	Score
Linear Regression	0.46802465
Ridge	0.46801923
Tweedie Regressor	0.46551012
Lasso	0.46802464
Elastic Net	0.46665415
Lasso Lars	0.46802457

Linear Regression:

$$\textbf{Annual Salary} = 47499.29 + 7112.71(\textbf{GS/GP}) + 582.98(\textbf{MPG}) + 6213.56(\textbf{PPG}) + 11.79(\textbf{FG\%})$$

Interpretations

As the ratio between games started and games played increases by 1, the player's salary is expected to increase by $7,112.71.

As the player's average minutes played per game increases by 1 minute, the player's salary is expected to increase by $582.98.

As the player's average points scored per game increases by 1 point, the player's salary is expected to increase by $6,213.56.

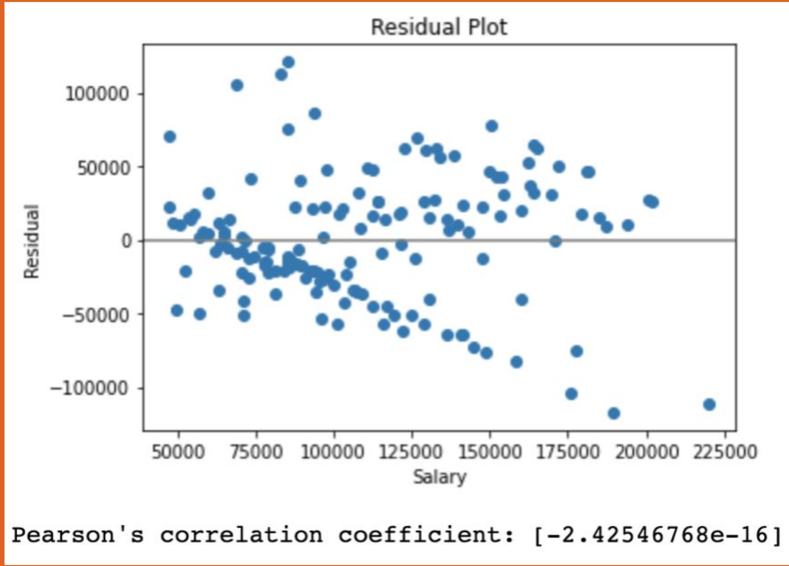
As the player's average field goal percentage increases by 1 percentage point, the player's salary is expected to increase by $11.79.

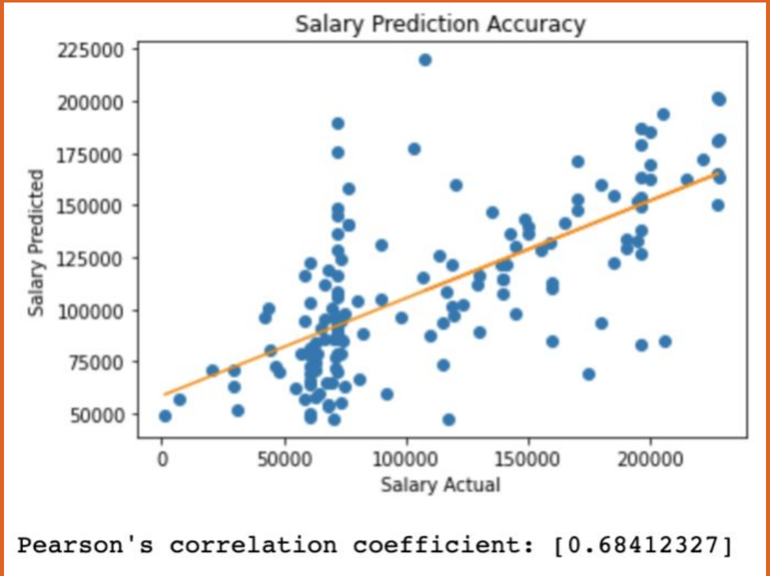
Annual Salary = 47499.29 + 7112.71(**GS/GP**) + 582.98(**MPG**) + 6213.56(**PPG**) + 11.79(**FG%**)

Annual Salary = 47499.29 + 7112.71(**GS/GP**) + 582.98(**MPG**) + 6213.56(**PPG**) + 11.79(**FG%**)

Analysis of Regression

The residual plot has a very low correlation coefficient. There doesn't seem to be any distinct pattern of residuals in the data.
Any differences between salary prediction and actual salary tend to be randomly dispersed.
There is no bias in variability. A linear regression method is a good method to classify the data.





Analysis of Regression

There is a relatively strong, positive correlation between the actual salaries of WNBA players and the predicted salaries. This indicates that our regression was relatively consistent in predicting salaries.

05.

Discussion

of data

Inclusion of Positions

Reason to include players' positions: points may vary based on position.

Would need an interacted linear regression to weigh player's PPGs differently based on position.

One Way ANOVA test of players' average PPG based on position gave a p-value >0.05. No statistically significant difference in PPG based on position.

Ultimately, we decided not to include players' positions in regression.

```
PPG grouped by position:           F_onewayResult(statistic=1.0790117583238445, pvalue=0.3659707325798397)
Total Points grouped by position:  F_onewayResult(statistic=0.8932073784676631, pvalue=0.4676005644913293б)
```

Accounting for Veteran Status

There is no real consistent, systematic way of determining a player’s veteran status.

There was no data on veteran status available for NBA players, and it would be even more difficult to find that data for WNBA players.

Veteran status would not have a large impact on individual performance statistics.

Ultimately, we decided not to include players’ veteran status in regression.

NBA "Field Goal Percentage vs. Salary"

It is important to acknowledge that in plotting FG% v. Salary for NBA data, there appears to be a bell-shaped distribution with a correlation of only 0.15.

There may be other drivers behind why this could be the case.

Despite not having a large correlation, we decided to keep it in our regression of NBA data for sake of continuity.

NBA Regression

Coefficients for MPG and FG% were both negative in NBA regression.

This is expected, since an increase in MPG or FG% would result in a salary increase. It is justifiable within context of research question.

Similarly to the findings in the WNBA data, MPG and FG% do not have as much of a positive impact on salary as percentage of GS/GP and PPG do.

06.

Discussion

of findings

	NBA	WNBA
Average GS/GP	0.346	0.473
Average MPG	18.697	19.319
Average PPG	7.852	7.635
Average FG%	43.151	39.605
Average Salary	$6290373.82	$110030.90
# Players used in regression	638	155

Findings Summarized

	NBA	WNBA
Intercept	510089.46	47499.29
GS/GP	6559401.72	7112.71
MPG	-102181.96	582.98
PPG	876426.52	6213.56
FGP	-33801.62	11.79

WNBA

GS/GP goes up with salary, Average points scored goes up with salary

Average minutes played goes up with salary but only a little

FG% has little impact on salary

NBA

GS/GP goes up with salary, Average points scored goes up with salary

FG% and Average minutes played are negatively correlated.

Our hypothesis was wrong.

Neither FG% nor minutes played increased with salary for NBA or WNBA.

Recall: Our hypothesis...
- WNBA salaries correlated with FG%;
- NBA salaries correlated with minutes played.

Significance of Findings

Seeing as WNBA and NBA players' salaries are related to performance stats relatively similarly, this analysis did not have any satisfying conclusions in terms of pay inequality.

Perhaps a better or more appropriate method to explore our topic and research question would be to look at social factors rather than performance statistics.

Because performance statistics shows to be similarly related to NBA vs WNBA salaries, this analysis may reveal that performance stats are arbitrary in determining or examining pay disparities between NBA and WNBA athletes.

07.

Thank You!

08.

Sources

Data

www.basketball-reference.com
www.spotrac.com
https://www.cbssports.com/
https://www.espn.com/

Background

https://www.nbcsports.com/bayarea/warriors/warriors-draymond-green-sounds-wnba-nba-pay-gap-taking-action
https://medium.com/nightingale/visualizing-the-gap-eccb912d75e2
https://medium.com/@yannickondoa/behind-the-figures-1-why-wnba-players-are-not-paid-as-much-as-nba-players-51c1574a6d7
https://www.nytimes.com/2022/02/22/sports/soccer/us-womens-soccer-equal-pay.html
https://www.nbcnews.com/think/opinion/u-s-women-s-soccer-team-equal-pay-settlement-sets-ncna1290309

Shaye as NBA wife → $$$$$ Anna as WNBA wife → 00000$