Basketball & Finance:

The Correlations Between NBA Team Expenses and Their Performance on Team and Individual Player Level

INF1344: Introduction to Statistics for Data Science
LEC 0102 Group 7
Professor Tao Wang

Adam Borthwick
Iris Di-Er Tu
Sinan Nahel
Yiang Mao
Peter Goldman

1. Introduction

As one of the most-watched sports in the world, the National Basketball Association (NBA) has been established since 1946 and continues to experience increasing viewership each year. With this sustained growth, player salaries have also risen significantly. In addition to league-based compensation, NBA players often hold endorsement contracts with major sports brands for personalized product lines, the value of which depends largely on their popularity and on-court performance.

Motivated by an interest in the financial dimension of the sports industry, this study aims to examine the relationships among player salaries, player performance, team expenses, and team performance to generate insights that could support the development of predictive models for marketing and financial decision-making. To achieve a comprehensive understanding, this research incorporates data analysis from NBA players' statistics and their salaries in the 2024–2025 season, along with analysis of team payrolls, season summaries, and salary cap history data from the 2020–2021 to 2024–2025 seasons. By utilizing these data and findings, we aim to answer the question:

How is on-court performance represented by off-court salary and pay structures within NBA teams?

Additionally, the study builds upon this primary inquiry by further exploring whether player performance and salary demonstrate a positive or negative correlation, thereby contributing to a deeper understanding of the financial dynamics within professional basketball.

2. Literature Reviews

Activities and research have been conducted to study the structure of athletes' salaries and their relationship with team performance in the sports industry. Lazear and Rosen (1981) proposed a rank-based compensation system that can improve work efficiency, and with clear evaluation metrics this type of compensation can help produce optimal financial allocations. This idea further extends to economic theory, where the superstar effect suggests that organizations tend to shift attention and rewards toward a few individuals who receive significant compensation. Simmons and Berri (2011) showed that the superstar effect can significantly influence team outcomes, concluding that pay dispersion has a positive effect not only on team performance but also on individual performance.

In terms of team expenses, professional teams are often founded by private companies seeking financial gain, thus the salary caps are commonly enforced to maintain fairness. Quirk and Fort argued that a salary cap is essential to preserving competitive balance in the NBA; by limiting total payrolls, leagues can prevent destructive competition and ensure teams remain equally attractive to players. Katayama and Nuch (2011) explored wage gaps within and across teams by applying a weighted coefficient of playing time, a Herfindahl index based on playing time, and a Gini coefficient across the full roster. These three measurement approaches concluded that salary dispersion has no causal effect on either game-level or season-level performance.

However, in other sports such as MLB, Tao, Chuang, and Lin (2016) found that salary dispersion does affect team performance, using the Gini coefficient and the Herfindahl–Hirschman Index to measure pay inequality. They reported that greater salary dispersion is generally negatively related to winning, and they further suggest that overall payroll competitiveness matters more for performance than how salaries are distributed inside the team. This contrast highlights how different sports environments place varying emphasis on cooperation versus individual incentives.

3. Data and Methodology

3.1 Topic 1 : Relationship Between Team Spending and Team Performance

3.1.1 Data Description:

Table 1

	NBA Season Data **[***2024_25, 2023_24, 2022_23, 2021_22, 2020_21***]**		
Source	1) NBA Team Payroll	2) NBA Season Summary	*3) NBA Salary Cap History*
Volume	• 30 NBA teams for each dataset • 5 datasets • 150 rows of data in total	• 30 NBA teams for each dataset • 5 datasets • 150 rows of data in total	• 5 seasons of NBA salary cap data • 5 rows in total
Feature	**Rank:** Position of team based on payroll amount (1 = highest payroll) **Team:** Name of the NBA team **Payroll:** Aggregated payroll of	**Team**: Name of the NBA team **W**: Wins **L**: Losses	**Season:**NBA season identifier (e.g., 2020_21, 2021_22) **Cap_Millions:** Official NBA salary cap for the season, expressed in millions of dollars

<table>
<tr><td></td><td>each team for the season</td><td>W/L%: Win percentage

PS/G: Points scored per game

PA/G: Points allowed per game

SRS: Simple Rating System (strength of team based on point differential and schedule)</td><td></td></tr>
<tr><td colspan="4">1. These 5 datasets provide the financial investment each NBA team made in player salaries for the past 5 seasons. The payroll data serves as the independent variable in our analysis.</td></tr>
<tr><td colspan="4">2. This dataset provides performance metrics, including win totals, for the past 5 seasons. The number of wins will serve as the dependent variables in our analysis.</td></tr>
<tr><td colspan="4">3. This dataset contains official salary cap limits set by the NBA for each season over the past five years. The cap provides the maximum allowable team payroll for that season, serving as a reference point for analyzing team spending relative to league constraints.</td></tr>
</table>

Table 2

	Final_df (cleaned dataset)
Volume	• 30 NBA teams for each dataset • 5 datasets • 150 rows of data in total
Features	**Team_Code**: Standardized team identifier combining team short code and season (e.g., GSW_2020_21). **W/L%**: Win percentage **Relative_Payroll**: Payroll Relative to Cap

3.1.2 Data Assessment:

Table 3

#	Task	Description
1	**Handle Missing Value**	• Check for missing entries in key columns **(W, L, W/L%, Payroll).** • Remove rows with missing values.
2	**Remove Duplicates**	• Identify and eliminate duplicate records, retaining only the first occurrence.
3	**Perform Consistency Checks**	• Validate calculated fields: ○ For team summary: Ensure **W/L% = W ÷ (W + L)** in the season summary. ○ For team payroll: Confirm payroll values are numeric and positive. • Remove any rows with inconsistent or invalid data.
4	**Remove Outliers**	• Detect and remove rows with extreme values in **W/L%** and **Payroll** column using the Standard Deviation method. • Threshold: ±3 standard deviations from the mean.
5	**Standardize Team Identifier**	• Generate team identifiers by appending teams' standardized short codes and the season year. Example: Golden State Warriors → GSW_2020_21.
6	**Normalize payroll Value**	• Express payroll figures in millions to simplify numeric representation and improve readability.
7	**Aggregate Multi-Season Data**	• Append data from all five seasons into a single consolidated dataset for longitudinal analysis.

8	**Merge Season and Payroll Data**	• Combine the season summary dataset with the corresponding payroll dataset using the standardized team identifier.
9	**Calculate and add Relative_Payroll column**	• Relative_Payroll = Payroll_Millions/Cap_Millions**.**
10	**Select Relevant Columns**	• Extract essential columns for analysis: **Team_Code, Relative_Payroll, W/L%.**
11	**Randomize row order**	• Randomize row order in final dataset

For topic 1, data was first collected in CSV format from authoritative basketball data platforms. Payroll data was sourced from https://www.hoopshype.com/salaries/teams/, which provides official team salary figures for each NBA season, while performance and salary cap data were obtained from https://www.basketball-reference.com/, a comprehensive repository of team statistics. Each team summary and payroll dataset spans five consecutive seasons (2020–21 through 2024–25), covering all 30 NBA teams per season, resulting in 150 rows per dataset. The salary cap dataset contains only five rows, one for each season, representing the official league-wide cap limit.

Data preparation and cleaning were performed in R. Initially, team names were standardized using a mapping dictionary that converted full team names into short codes (e.g., “Golden State Warriors” → “GSW”). Separate lists of data frames were created for team summary and payroll data, and empty lists were initialized to store processed outputs. Cleaning steps were automated using loops and dplyr functions. For team summary data, rows with missing values in W, L, or W/L% were removed, duplicates were eliminated, and outliers in W/L% were detected using the standard deviation method (threshold: ±3 SD from the mean) and excluded. Consistency checks ensured that calculated win-loss percentages matched reported values within a tolerance of 0.001. Team identifiers were standardized by concatenating short codes with season labels, and cleaned data across seasons was aggregated using bind_rows().

Payroll data underwent similar cleaning: missing values were removed, duplicates eliminated, and payroll figures validated as numeric and positive. Payroll values were normalized by converting them to millions and rounding to three decimal places. After cleaning, payroll datasets were aggregated across seasons. The salary cap data required minimal preparation, as it contained only season and cap values; these were converted to millions for consistency.

Finally, the combined team summary and payroll datasets were merged on the standardized team identifier using inner_join(). A derived feature, Relative_Payroll, was computed by dividing each team's payroll by the corresponding season's salary cap. Before applying the model, the rows in the final dataset were randomized because the original data was sorted by team name when collected. Randomization was performed to avoid introducing a fake autocorrelation pattern that could bias the regression analysis. The final dataset retained 3 key variables, Team_Code, W/L%, and Relative_Payroll, for subsequent modeling and analysis.

3.1.3 Testing Assumptions: Topic 1

After organizing the data linear regression assumptions were tested with the following results:

3.1.3.1 Independence of Residuals

Durbin-Watson test:

DW = 2.3197, p-value = 0.9759

After conducting a Durbin-Watson test, Independence holds true $P > 0.05$.

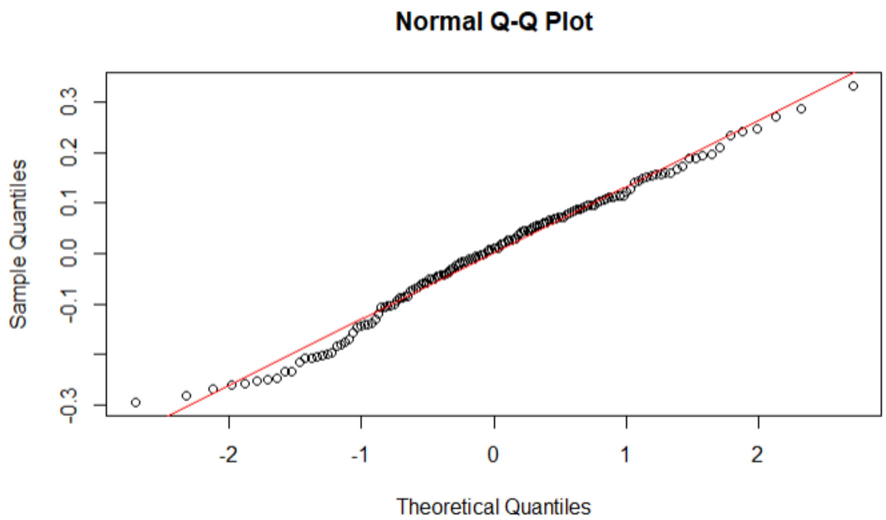
3.1.3.2 Normality of Errors

Shapiro-Wilk normality test:

W = 0.98689, p-value = 0.1726

After conducting Shapiro-Wilk test and conducting a visual Q-Q plot test normality holds true $P > 0.05$.

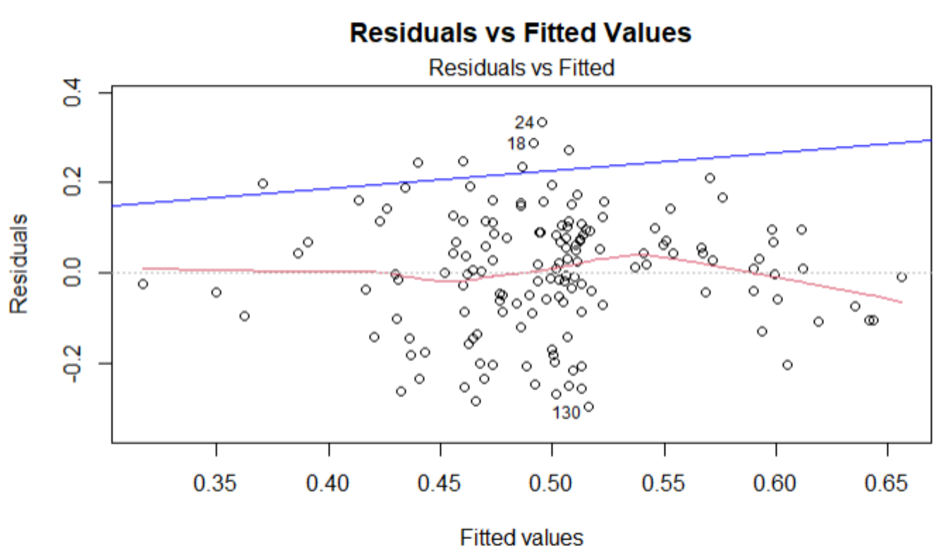
Figure 1



3.1.3.3 Linearity & Homoscedasticity

Based on the residuals vs fitted values results, homoscedasticity holds due to normal distribution of residuals. Additionally Linearity holds based on the red LOESS line within the residuals vs fitted graph.

Figure 2



3.2 Topic 2 : Relationship Between Player Salary and Performance

Data Description:

Table 4

	NBA Player Data **[***2024_25***]**	
Source	1. NBA Player Stats	2. NBA Player Salaries
Volume	• 16,512 entries • 25 total columns	• 16,512 entries • 25 total columns
Features	**Player**: Name of the player. **Tm**: Abbreviation of the player's team **Opp**: Abbreviation of the opposing team. **Res**: Result of the game for the player's team. **MP**: Minutes played during the game. **FG, FGA, FG%**: Field goals made, attempted and percent **3P, 3PA 3P%**: 3-point field goals made, attempted and percent. **FT, FTA, FT%**: Free throws made, attempted, and percent. **ORB, DRB, TRB**: Offensive, Defensive and total rebounds. **AST**: Assists. **STL**: Steals. **BLK**: Blocks.	**Player**: Name of the player. **Team**: Abbreviation of the player's team. **Salary**: Salary of the player in USD

	TOV: Turnovers. **PF**: Personal fouls. **PTS**: Total points scored. **GmSc**: Game Score, a metric summarizing player performance for the game	
1. This dataset provides comprehensive performance statistics for NBA players throughout the 2024/2025 season. It includes a mix of stats allowing for use for player performance analysis.		
2. This dataset provides individual player salaries in USD throughout the 2024-25 season.		

Data Assessment:

Table 5

#	Task	Description
1	**Clean Column Names**	• Clean column names to be consistent across data sets - “Team” -> “tm”
2	**Clean Player Names**	• Clean the names of players to avoid misalignment of data with names that contain unique characters i.e "Dennis Schröder" to "Dennis Schroder" using a custom function. • Clean inconsistent naming conventions within player names “Jr.” vs “Jr”.
3	**Summarize Game Statistics to Average**	• Start by summarizing each game stat into an average of player performance i.e FT% ->avg_FT% • Converts 16 512 rows to 559 rows for each player • Creating a count of total games played and including additional stats like win_pct.
4	**Clean Player Salaries Database**	• Convert salaries to numeric integers - removing ‘$’ and ‘,’.
5	**Consolidate Data Sets**	• Left join data sets based on both team and players.

6	**Handle Missing Values**	• Through this process there were a few major discrepancies which led to unclean data - which had to be handled, many of these discrepancies related to players who were logged • Most commonly these were players who get traded mid season and then are part of multiple teams - the salary data set had only one team registered per player. • For these players we'll use the team for which the contract was made with them on the "player_performance_salaries" data set. ○ Step 1 - Find players who have 2 teams ○ Step 2 - Filter those players ○ Step 3 - For each player, remove the NA salary row • Secondary merge between the original salary data set to ensure all players are included, irrespective of team • After this process remove rows still missing values.

3.3 Initial Analysis

To examine the relationship between player salary and performance we initially cleaned the data for both NBA player performance and salary as discussed within the data section. After the initial cleaning and consolidation of these data sets initial graphs were made, initial graphs were built to help give a visual view of distributions within the dataset.

Figure 3

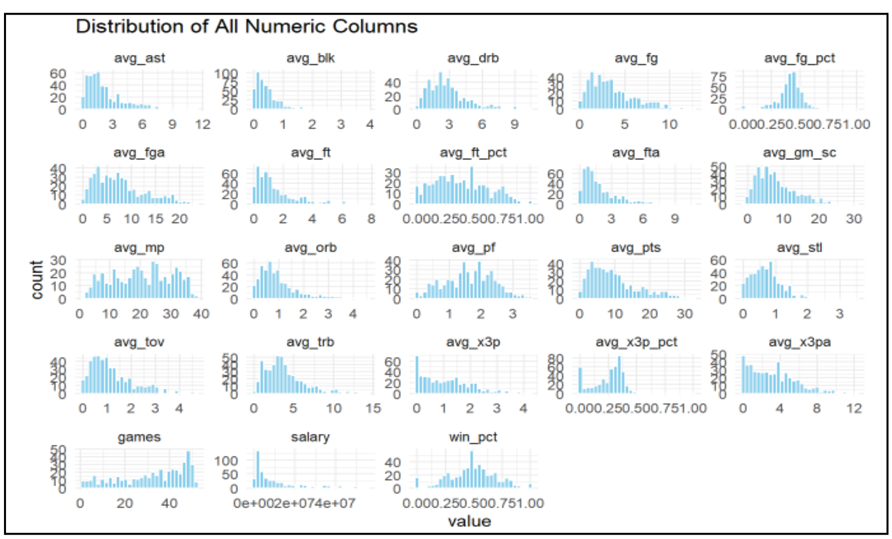
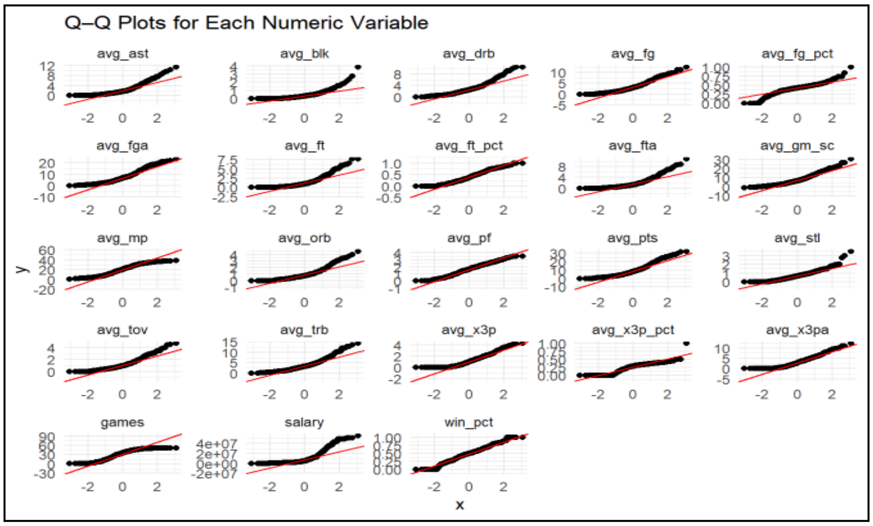


Figure 4



These initial graphs were utilized to identify any major issues within the data set, based on these our team moved forward with multiple linear regression.

3.4 Included Data

Tables 6 & 7

Name	Description
salary	Dependent Variable
win_pct	Win percent
avg_fg_pct	Field goal percent
avg_x3p_pct	3 point shooting percent
avg_ft_pct	Free throw percent
avg_trb	Average total rebounds per game

Name	Description
avg_ast	Average assists per game
avg_stl	Average steals per game
avg_blk	Average block per game
avg_tov	Average turnovers per game
avg_pf	Average personal fouls per game

3.5 Excluded Data

For multiple linear regression it's key that data that doesn't lead to multicollinearity - for this reason the following stats were excluded outright due to their obvious threats to multicollinearity

- FG, FGA, 3P, 3PA, FT, FTA, DRB, ORB - were excluded because they are highly correlated with fg_pct, x3p_pct, ft_pct, TRB
- PTS(points) is a combination of fg + x3p + ft
- GmSc - excluded since its a makeup of multiple other columns
- avg_mp is excluded because it is correlated with stats measuring averages per game

win_pct and pf are included within the multiple linear regression at this point but may have weak correlation leading to unexpected noise during multiple linear regression. These two stats in particular will be kept in mind during subsequent VIF checks.

3.6 Testing Assumptions: Topic 2

Multiple Linear regression works under certain assumptions for verifying these assumptions several tests were completed and visual checks were completed:

3.6.1 Independence of Residuals

Durbin-Watson test:

Based on the results DW 1.9288 - residuals are independent with low positive autocorrelation

P value of 0.2161

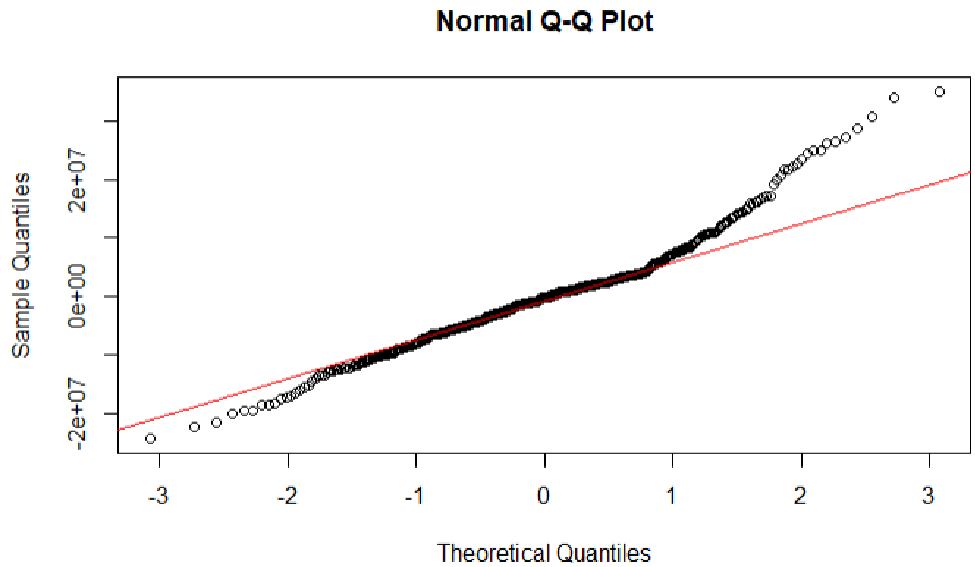
3.6.2 Normality of Errors

Shapiro-Wilk normality test:

W = 0.96262, p-value = 1.516e-09

P = 1.516e-09 value result is very low which could be a result of larger sample size since SW tests are built for $n < 200$

Figure 5



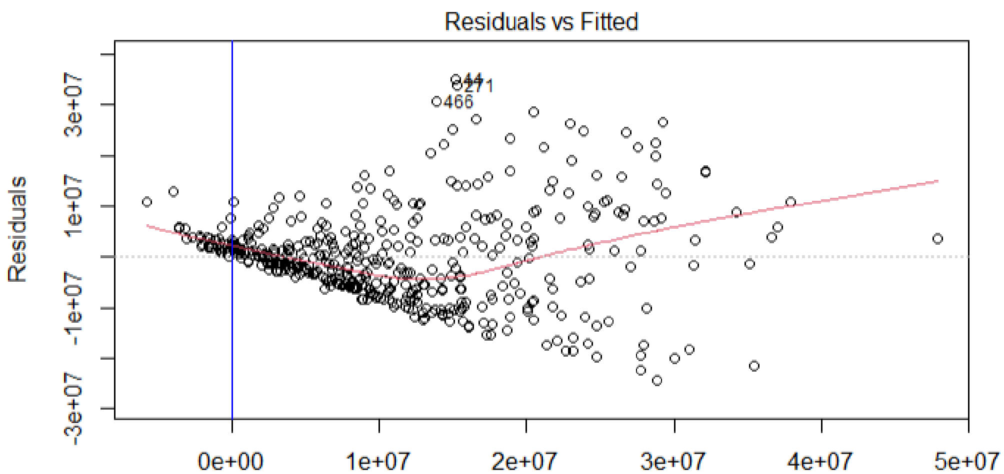
Additionally, this could be a result of players at the extreme ends of the graph who have a disproportionately large salary i.e Steph Curry, Joel Embiid. To verify whether this is true QQ-plot of residuals will be utilized to visually check

3.6.3 Linearity & Homoscedasticity

To test linearity and Homoscedasticity, a residuals vs fitted graph was utilized with a LOESS (red) line to identify lack of linearity. Based on the results of this test there is a significant lack of linearity - signified by the clear curve within the LOESS line.

Based on this graph, homoscedasticity also fails. This graph has a significant clear heteroscedastic behavior - identified by the significant funnel shape. To resolve this, the linear model was rebuilt utilizing a log transformation for salary.

Figure 6



3.7 Re-modelling With Log of Salary:

3.7.1 Independence of Residuals

Durbin-Watson test:

P value of 0.01469, DW 1.8007 - significant positive autocorrelation with lower score than original modelling, log transformation compresses extreme values which leads to expected results. However, because the player data are cross-sectional and not time-ordered, the Durbin–Watson test does not have a strict temporal interpretation. Due to the lack of relevance of this test an alternative method was utilized to ensure independence. Specifically, players that are within the same group are highly correlated with each other, players were clustered based on team since performance was highly correlated within the same cluster.

Figure 7

```
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 13.164437   0.266977 49.3092 < 2.2e-16 ***
avg_fg_pct   0.182255   0.534240  0.3411  0.733149    
avg_x3p_pct  1.784586   0.356397  5.0073 7.891e-07 ***
avg_ft_pct   1.425036   0.288511  4.9393 1.101e-06 ***
avg_trb      0.104490   0.022328  4.6797 3.788e-06 ***
avg_ast      0.139938   0.049842  2.8077  0.005203 ** 
avg_stl      0.202717   0.148427  1.3658  0.172682    
avg_blk      0.138289   0.149959  0.9222  0.356921    
avg_tov     -0.112133   0.117469 -0.9546  0.340297    
avg_pf       0.160742   0.082266  1.9539  0.051317 .  
win_pct      0.407580   0.363129  1.1224  0.262276    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

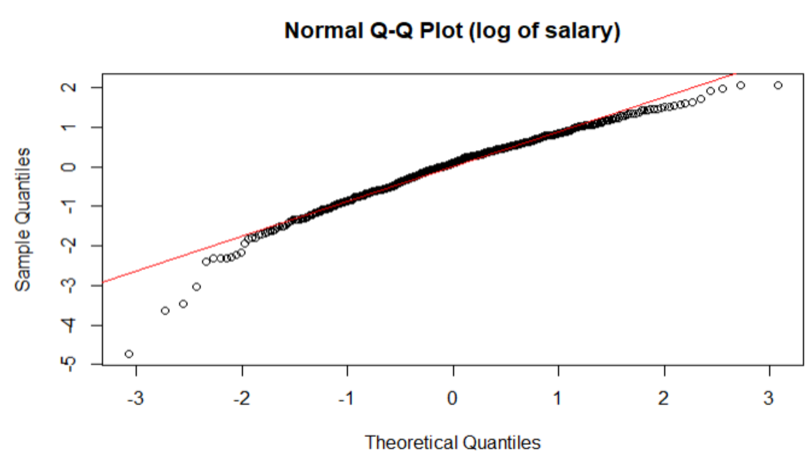
3.7.2 Normality of Errors

Shapiro-Wilk:

W = 0.96729, p-value = 1.011e-08

P value remains low, Q-Q plot is utilized for visual check and the Normal Q-Q Plot (log of salary) below indicates normality.

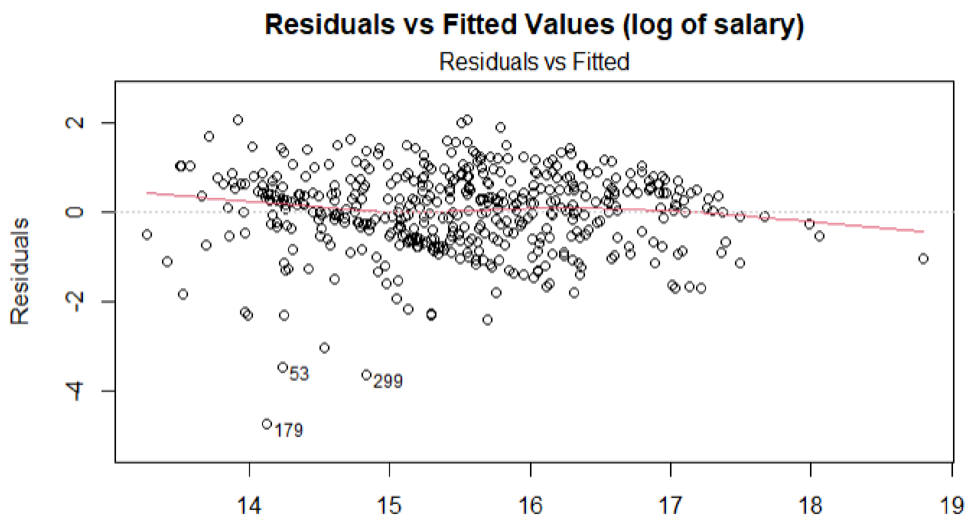
Figure 8



3.7.3 Linearity & Homoscedasticity

Utilizing the log of salary, Linearity and Homoschedasticity holds true, residuals vs fitted chart LOESS line is approximately straight and distribution is significantly.

Figure 9



3.7.4 Multicolinearity

VIF test was utilized to check for multicollinearity. Lack of multicollinearity assumption holds in this case, all variables have VIF < 10.

Figure 10

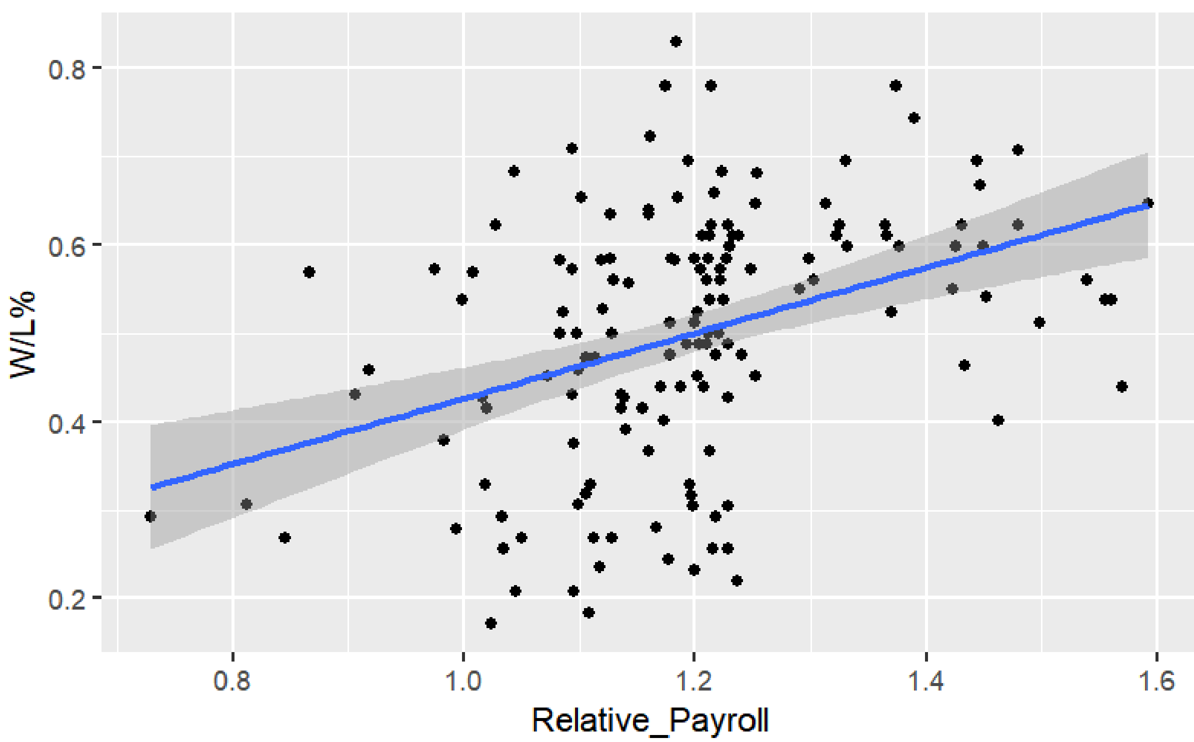
avg_ag_pct	avg_x3p_pct	avg_ft_pct	avg_trb	avg_ast	avg_stl	avg_blk	avg_tov	avg_p
1.741471	1.404378	2.278374	3.083636	4.138304	2.160149	2.031319	4.745035	2.406468

4. Results and Interpretation

4.1 Topic 1 : Relationship Between Team Spending and Team Performance

Figure 11

□[38;5;232m`geom_smooth()` using formula = 'y ~ x'□[39m



4.1.1 Represented Relationship

The scatterplot illustrates the relationship between team win percentage (W/L%) and payroll spending relative to the salary cap, which measures how much a team spends on player salaries compared to the league's salary cap in that season. For example, a value of 1.20 means the team spent 20% above the salary cap, usually by paying the luxury tax, while a value of 0.90 means the team spent 10% below the cap. Using this measure allows us to compare teams across different seasons even when the cap changes each year.

In the graph, each point represents a team in a particular season, and the blue line shows the overall trend. The clear upward slope indicates that teams who spend more relative to the cap tend to have higher win percentages. Although the individual points are scattered widely meaning that teams with similar spending can still perform differently, the general direction of the trend remains positive. This reflects what we learned in class, which is that even when there is a relationship, it does not have to be perfect because real-world data contain noise, unmeasured variables, and natural variation. In simple terms, the graph shows that teams that spend more money on their roster usually win more games. Spending does not guarantee success, but it does increase the chances of winning.

This relationship also makes logical sense in the context of the modern NBA. The teams that spend far above the cap are usually those with multiple star players on max contracts like teams like the Warriors, Clippers, Bucks, and Celtics in recent years. High-spending teams are generally investing heavily in top-tier talent, veteran depth, and roster stability while teams that spend far below the cap tend to be rebuilding teams, teams full of young players on rookie contracts, or teams intentionally avoiding large salaries and thus lack the experience or star power needed to consistently win games.

The widening grey confidence band at the far ends of the graph also fits NBA reality. Extremely high-spending teams can either dominate (if their stars are healthy) or underperform (if injuries or chemistry problems arise), creating higher uncertainty. A real NBA example that fits this pattern is the 2021 to 2022 Los Angeles Lakers. They had one of the highest payrolls in the league that season but finished with a losing record. Despite investing heavily in stars, the team struggled with injuries and poor roster fit, which caused them to underperform relative to their spending level. This type of outcome explains why the uncertainty increases among the highest spending teams and why the relationship between payroll and performance, while positive, is not perfect.

On the contrary very low-spending teams can occasionally exceed expectations, but most perform poorly, which creates additional variation. For example consider the 2022 to 2023 Oklahoma City Thunder. They were one of the youngest and lowest-payroll teams in the league, yet they finished with a winning record. Their success came from strong player development and an efficient roster built around emerging stars on rookie contracts. A team like this would appear on the left side of the graph with a relatively low payroll but a win percentage that is higher than expected based on their spending level. Cases like the Thunder show that even low-spending teams can outperform their financial limitations when they have a talented young core, good coaching, and a roster that fits well together.

Overall, the graph provides clear descriptive evidence that payroll spending is positively related to team performance. Teams that spend a greater share of the salary cap tend to win more, although spending alone does not fully determine outcomes. Factors such as injuries, coaching, roster balance, and player development all play an important role as well.

4.1.2 Regression Output

Figure 12

```
Call:
lm(formula = `W/L%` ~ Relative_Payroll, data = final_df)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.29377 -0.08762  0.01300  0.09136  0.33472 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.05541    0.08723   0.635    0.526    
Relative_Payroll  0.37075    0.07217   5.137 8.68e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1337 on 148 degrees of freedom
Multiple R-squared:  0.1513,	Adjusted R-squared:  0.1456 
F-statistic: 26.39 on 1 and 148 DF,  p-value: 8.676e-07
```

Here is the regression output summary of the graph. The slope coefficient for Relative Payroll, which is estimated at 0.37075, is the most important result in the regression output because it tells us how win percentage is expected to change when team spending relative to the salary cap increases. A one-unit increase in Relative Payroll means going from something like 1.0 to 2.0, which would indicate that a team went from spending 100 percent of the salary cap to spending 200 percent of the salary cap. In the NBA, this kind of jump is extremely unrealistic because no team can legally double the salary cap under current league rules. Because this kind of increase does not reflect how NBA payrolls behave in the real world, it is more meaningful to interpret the slope in terms of smaller changes, such as a 0.10 increase, which represents a 10 percent rise in spending relative to the cap. Interpreting the effect at this scale provides results that match actual team spending patterns and allows for more practical and understandable conclusions.

For example, a 0.10 increase in Relative Payroll, meaning that a team spends 10 percent more of the salary cap, is associated with an increase of approximately 0.037 in win percentage, which is about 3.7 percentage points. This effect size indicates that even relatively small increases in payroll relative to the salary cap can have a meaningful impact on how often a team wins games.

In simpler terms, the model suggests that teams that allocate a higher share of the salary cap toward player salaries tend to perform better, and each additional 10 percent of spending typically translates into about 3 to 4 percentage points more wins over the course of a season. This pattern aligns with real NBA dynamics as teams with multiple star players on expensive contracts often exceed the salary cap in pursuit of competitiveness, while rebuilding teams tend to spend less because they rely more heavily on young players on rookie contracts. As a result, the positive slope reflects the practical reality that investing more in player talent is usually associated with better on-court performance.

The regression output summary also shows that the p-value for the Relative Payroll variable is extremely small at 8.676e-07, which is far below the common significance threshold of 0.05. This indicates that the relationship between team spending and win percentage is very unlikely to be due to random chance. Based on this result, we reject the null hypothesis that spending has no relationship with performance and accept the alternative hypothesis that spending and winning are related. The very small p-value provides strong statistical evidence that payroll spending relative to the salary cap is associated with higher win percentages. Although this does not prove that spending directly causes teams to win more games, it does confirm that the connection observed in the data is statistically reliable.

The R-squared value of 0.1513 as indicated on the regression output summary indicates that about 15 percent of the variation in team win percentage can be explained by how much teams spend relative to the salary cap. While this may seem low at first, our lectures emphasized that R-squared values in real-world human performance data are rarely high because sports outcomes are influenced by many unpredictable factors. Even with this limitation, a moderate R-squared is still meaningful when the model is statistically significant, since it shows that Relative Payroll explains a real and measurable part of team performance. In other words, spending is not the whole story, but it is clearly one slice of the puzzle. This makes sense in the NBA, where success also depends on injuries, roster depth, coaching quality, travel demands, schedule strength, and countless other variables that money cannot directly control. The adjusted R-squared of 0.1456, which is very close to the original value, supports this interpretation. Because the model uses only one predictor, the similarity between R-squared and adjusted R-squared confirms that the model is not overfitting and that payroll spending relative to the cap genuinely helps explain differences in win percentage across teams.

Lastly, the residuals in the regression range from approximately –0.29 to 0.33, meaning that some teams performed about 29 percentage points worse than the model predicted, while others exceeded their predicted win percentage by roughly 33 percentage points. This wide range shows that although the model successfully captures a real upward trend between spending and performance, many teams do not perform exactly as their payroll level would suggest. This aligns with the reality of the NBA, where team success depends on far more than spending alone. For example, a high-spending team like the 2021–2022 Los Angeles Lakers underperformed largely due to injuries, aging stars, and roster imbalance, resulting in a win percentage far below what their payroll would predict. In contrast, a low-spending team such as the 2022–2023 Oklahoma City Thunder exceeded expectations because of strong player development, an improving young core, and effective coaching. These deviations are precisely what the residuals capture. They highlight that while payroll plays a role, NBA outcomes are influenced by many unpredictable factors that cause teams to either outperform or underperform relative to financial expectations.

4.2 Topic 2 : Relationship Between Player Salary and Performance

Using our finalized dataset, we ran a multiple linear regression using R, the output of which can be seen below.

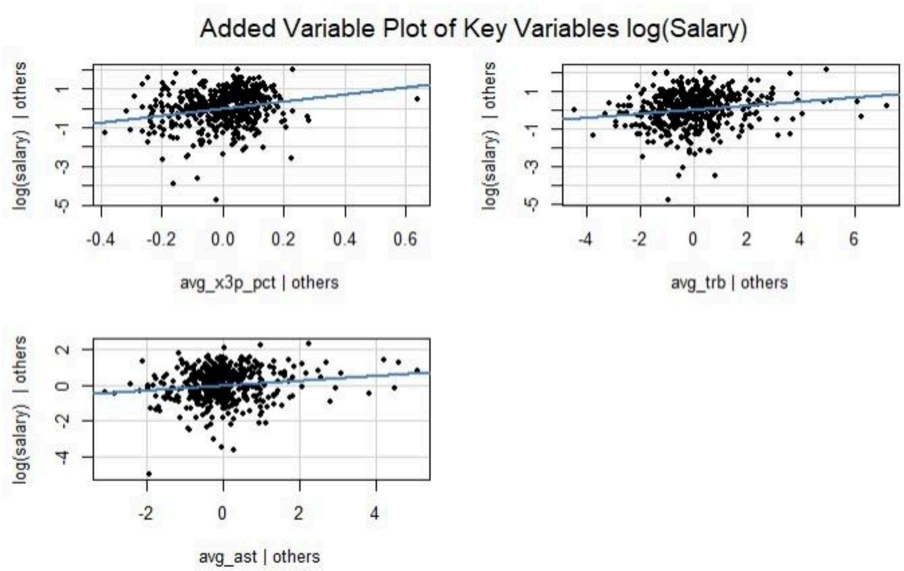
Figure 13

```
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 13.164437   0.266977 49.3092 < 2.2e-16 ***
## avg_fg_pct   0.182255   0.534240  0.3411  0.733149
## avg_x3p_pct  1.784586   0.356397  5.0073 7.891e-07 ***
## avg_ft_pct   1.425036   0.288511  4.9393 1.101e-06 ***
## avg_trb      0.104490   0.022328  4.6797 3.788e-06 ***
## avg_ast      0.139938   0.049842  2.8077  0.005203 **
## avg_stl      0.202717   0.148427  1.3658  0.172682
## avg_blk      0.138289   0.149959  0.9222  0.356921
## avg_tov     -0.112133   0.117469 -0.9546  0.340297
## avg_pf       0.160742   0.082266  1.9539  0.051317 .
## win_pct      0.407580   0.363129  1.1224  0.262276
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9175 on 458 degrees of freedom
Multiple R-squared:  0.5197, Adjusted R-squared:  0.5092
F-statistic: 49.55 on 10 and 458 DF,  p-value: < 2.2e-16

##     Estimate         Std. Error        t value          Pr(>|t|)
##  Min.   :-0.1121   Min.   :0.02233   Min.   :-0.9546   Min.   :0.000e+00
##  1st Qu.: 0.1391   1st Qu.:0.09987   1st Qu.: 1.0223   1st Qu.:2.440e-06
##  Median : 0.1823   Median :0.14996   Median : 1.9539   Median :5.132e-02
##  Mean   : 1.5998   Mean   :0.21632   Mean   : 6.4995   Mean   :1.747e-01
##  3rd Qu.: 0.9163   3rd Qu.:0.32245   3rd Qu.: 4.8095   3rd Qu.:3.013e-01
##  Max.   :13.1644   Max.   :0.53424   Max.   :49.3092   Max.   :7.331e-01
```

The adjusted R^2 value was 0.5092, indicating that the model accounts for around 51% of the variation found in the dataset. As well, the F-statistic was 49.55 on 10 and 458 degrees of freedom, with an extremely low p-value (< .001), indicating that the regression model is significant.

The most significant predictors of player salary were the average percent of the free throws made (avg_ft_pct), the percent of 3-point field goals made (avg_x3p_pct), and the average number of rebounds (avg_trb), all of which were highly significant at $p < .001$. Players' average number of assists (avg_ast) was also significant at $p < 0.01$. Since a log transformation was applied to our salary data, our regression output can't be easily read to say that an X percent increase in a particular variable results in a Y increase in salary dollars, but we nonetheless would expect to see an increase in pay as most variables increase. These positive relationships can be seen in the added variable plots below.

Figure 14



These plots display the relationship between the salary and the key variables discussed above, when controlling for all other variables, by plotting the residuals of each factor after regression. We can see that all exhibit a positive relationship (as indicated by the blue regression line), although the strength of this varies. Average percent of 3-point field goals made appears to be the strongest relationship, followed by average amount of rebounds. The regression line for average number of assists is more slight, which we might expect given that the relationship was found to be significant, but less so than other variables.

The least significant factors were players' average number of blocks, steals, and turnovers. It is interesting to observe that the least impactful variables are those that involve direct interaction with the opposing team: by preventing them from scoring, taking the ball from an opponent, or losing possession of it. The most significant factors relate to a player's shooting and ball handling skills. A player who is consistently able to make their free throws and 3-point goal attempts, or to gain possession of the ball for their team is an asset, is likely to be perceived as a strong player by those making hiring and salary decisions. These factors are extremely salient parts of player performance that clearly contribute to a team's winning, as opposed to factors representing player actions that work to counteract the opponent and prevent them from scoring.

While both free throw and field goal percent are related to a player's scoring ability, only free throw percent was found to be a significant predictor of salary. Free throws are given to teams in certain circumstances (such as after a foul by their opponent), and are made freely by players without interference by the opposing team. Field goals, on the other hand, are more general shots made at any time during the game. Again, we might see this as reflective of which player attributes stand out more to observers. When making a free throw, the player is in the spotlight and it is solely their shooting skills that are in question. Field goals, on the other hand, take place during the rest of the game, when other players are in movement on the court and there are many factors that can affect a player's attempt to shoot.

It is also of note that while generally all variables had a positive relationship to salary, even if slight, turnover had a negative relationship, whereby an increase in turnover rate decreases salary. When we take into account that turnovers involve losing possession of the ball, this makes sense, as they would never be advantageous to a team. However, as noted above, this was one of the least significant factors, meaning that its impact on salary is very slight.

5. Discussions

5.1 Topic 1 : Relationship Between Team Spending and Team Performance

As we arrive at the finding of a slight positive relationship between NBA team spending and their overall performance, the lack of support for a strong positive correlation falls short of our original expectation, where we believed higher spending would lead to better performance. The expectation is genuine, since it is normal to have players receiving big contracts when they outperform others during the NBA season or even before they enter the professional league (e.g., NCAA or CUBA). Our study takes into consideration the business strategies and evaluations that each team makes before signing their contracts with particular players, since it is the only reasonable way for the team to calculate the detailed and precise worth of a player. However, the findings present an otherwise answer, resulting in an opportunity for further discussion on the potential factors that influence those with big contracts to perform poorly on the court. The factors of this can range from player experience, health conditions, injuries, personal life, and other things that our current dataset has not yet taken into consideration. In all, it is difficult to evaluate or make a conclusion based on the pure math that statistical assumptions present to us; it is also necessary to consider the changing factors and influences that athletes are prone to.

Another important consideration is the role of salary cap mechanisms and luxury tax penalties in shaping team spending behavior. While our analysis shows a positive relationship between spending and performance, this relationship is moderated by league rules that discourage excessive spending through financial penalties. Teams that exceed the salary cap face luxury tax obligations, which can deter further investment in player contracts and introduce a strategic trade-off: while higher spending may enhance on-court performance, it also raises significant financial risks for the organization. Additionally, the presence of salary cap exceptions, such as the mid-level exception or Bird rights, allows certain teams to retain players without fully adhering to cap limits, introducing complexity into the spending-performance dynamic. These structural factors suggest that team spending decisions are not purely performance-driven but are also shaped by financial constraints and long-term roster planning. Future research could incorporate these variables to better understand how league policies mediate the link between team payroll and success.

5.2 Topic 2 : Relationship Between Player Salary and Performance

Overall, we observed that there is a relationship between player performance and salary, with players that perform better generally being paid more. However, as with our first topic, this is not a universal rule, and not all player statistics we looked at have an equal effect on salary. Determining which factors are most relevant gives us insight into which elements of player performance might be seen as most valuable by the NBA. While our analysis here was focused on salary, investigating the relationship between win percentage and player statistics would also give interesting insights into which factors actually help to win the most games, which may or may not line up with those most correlated with higher salary. Our study was limited to the 2024-25 season, but an analysis over a longer span of time could also shed greater light on how these factors may change over time. It could also be valuable to see if different teams value different player stats, as NBA teams are not all the same. Further analysis would also be affected by the fact that players move around, playing on different teams at different points in time, even within the same season. This was something we also had to account for in our data preparation.

Initially, our analysis included minutes played as a variable, but this was removed due to its relationship with other variables. As a number of other variables measure the average amount of a particular action performed per game (such as amount of assists, steals, etc.), this is naturally correlated with the amount of time they spend on the court, as longer play time means that they have more chances for each of these actions.

One area that this research could be extended revolves around how factors related to perceived performance affect the salary of a player, these include years played, height, weight and college of the players. While salary specifically grows over time, players who have been playing for longer tend to be paid more - adjusting for experience could provide a more robust analysis of these factors and verify the effect of experience on salary.

Additionally, height and weight within the NBA have a social and positional connotations that influence salary beyond measurable performance metrics. Incorporating how these metrics affect player salary could lead to a more thorough understanding of this area of research. This could motivate central research questions including "*how does player height affect salary?*" and *"how does player height affect individual player performance in relation to salary?"*

The results of our multiple linear regression give us insight into which player statistics are correlated with higher salary, and may be considered more valuable in the NBA. Statistics that reflect how well players are able to shoot, particularly high-point value and unimpeded shots, and gain possession of the ball for their team, are significantly correlated with an increase in salary. Although our models and results are one piece of a larger puzzle, this analysis gives us insight into how statistical analysis can be used to inform decision making and describe patterns that might not otherwise be apparent, something that is highly relevant to NBA teams and beyond.

5.3 Additional Considerations and Limitations

Beyond the results discussed above, several important factors help explain why our models do not fully capture the relationship among spending, performance, and salary in the NBA. These factors either do not appear in the dataset or are shaped by rules that limit how well statistical models can reflect real-world basketball behavior.

5.3.1 Unobserved Variables

One major limitation is that our analysis includes only numerical variables, such as points, rebounds, and payroll totals. However, in professional basketball, many important influences cannot be easily quantified. These include the quality of a team's coach, how well players work together, team chemistry, personal motivation, leadership, mental health, or life events outside basketball. A team might perform poorly even if it has talented players, simply because of internal disagreements or a coaching change. Likewise, a player might earn a high salary because they are seen as a strong leader or a positive influence, not just because of their box score statistics. Since these types of variables are not included in our dataset, our models cannot account for them, limiting the variation we can explain.

5.3.2 Salary Cap Restrictions Distort "True" Value

The NBA uses a salary cap system that places strict limits on how much teams can spend on player salaries. Some rules set a maximum amount that any one player can earn (Akabas, 2024). This means the league's best players cannot be paid their true market value. If teams were allowed to pay whatever they wanted, the top players would earn far more than they do now. Because the maximum salary rule prevents this, many elite players end up earning similar amounts even if their performance levels are very different. This makes it harder for statistical models to detect strong links between performance and salary because the salaries of top players are artificially compressed together.

5.3.3 Rookie Scale Contracts Distort the Low End

New players entering the league are paid using a fixed rookie scale. This means their salary depends mostly on where they were selected in the NBA draft rather than on their performance. Some rookies perform at a very high level right away, but they are still paid a low salary because their contract is predetermined. This results in many young players being underpaid relative to their true value (McKeone, 2024; Paine, 2015). It also weakens the statistical relationship between salary and performance because their performance is high while their salary remains low.

5.3.4 Veteran Minimum and Mid-Level Contracts

There are also salary rules that affect players who are neither rookies nor superstars. Many average or slightly below-average players are signed to veteran minimum contracts. These contracts pay players similar amounts regardless of their exact performance levels (Magaro-George, 2023). There are also mid-level contracts that teams commonly use for reliable role players, and these contract amounts tend to cluster. Because many players fall into these fixed contract categories, their salaries do not always reflect their exact level of performance. This creates plateaus where players with very different statistics earn similar salaries (Paine, 2015). As a result, the model sees less salary variation and has a harder time finding strong statistical patterns.

5.3.5 Garbage Time Inflation

Player statistics can also be affected by the game's context. When a game is essentially decided, and one team has a large lead, coaches usually rest their top players and let bench players play more minutes. During this lower-intensity part of the game, often called "garbage time", it is easier for players to score points and collect rebounds because the defense is less focused. Since our dataset uses season averages, it treats garbage-time statistics the same as statistics earned during competitive moments. This can make some players appear slightly better statistically than they actually are in high-pressure situations, introducing additional noise into the model.

Appendix

R Code: 1344 Final Code Group 7

References

Akabas, L. (2024, July 5). *How Do NBA Contracts and the Salary Cap Work?* Sportico. https://www.sportico.com/feature/nba-salaries-explained-salary-cap-1234786618/

Basketball-Reference. (n.d.). *NBA & ABA League index*. https://www.basketball-reference.com/leagues/

Basketball-Reference. (n.d.). *NBA salary cap history*. https://www.basketball-reference.com/contracts/salary-cap-history.htm

Falk, B. (n.d.). *Guide - Garbage Time.* Cleaning the Glass. cleaningtheglass.com/stats/guide/garbage_time

Frick, B., Prinz, J., & Winkelmann, K. (2003). Pay inequalities and team performance: Empirical evidence from the North American major leagues. *International Journal of Manpower, 24*(4), 472–488. https://doi.org/10.1108/01437720310485942

HoopsHype. (n.d.). *NBA Team Salaries & Contracts*. https://www.hoopshype.com/salaries/teams/

Lazear, E. P., & Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, *89*(5), 841–864. http://www.jstor.org/stable/1830810

Magaro-George, N. (2023, July 24). *How Does the Veteran Minimum Salary Work in the NBA?* Air Alamo. airalamo.com/posts/how-does-the-veteran-minimum-salary-work-in-the-nba

McKeone, L. (2024, June 26). *How Much Do NBA Rookies Make? Breaking down NBA Draft Pick Salary.* SI. www.si.com/nba/nba-draft-pick-rookie-salary-breakdown.

NBAstuffer. (2023, July 13). Player Evaluation Metrics. https://www.nbastuffer.com/analytics-101/player-evaluation-metrics/.

Paine, N. (2015, July 15). *No Matter How Much They Make, the Best Players in the NBA Are Vastly Underpaid*. FiveThirtyEight. fivethirtyeight.com/features/kawhi-leonard-like-all-the-best-nba-players-is-vastly-underpaid/

Palmieri, E. (2025, February 8). *NBA - player stats - season 24/25*. Kaggle. https://www.kaggle.com/datasets/eduardopalmieri/nba-player-stats-season-2425/data

Pradhan, S., & Leshchinskii, D. (2025). Violating the salary cap: exploring performance gains in the National Basketball Association. *Frontiers in sports and active living*, *7*, 1625458. https://doi.org/10.3389/fspor.2025.1625458

Ratin21. (2025, May 6). *NBA player salaries 2024-25*. Kaggle. https://www.kaggle.com/datasets/ratin21/nba-player-salaries-2024-25

Quirk, J., & Fort, R. D. (1992). *Pay Dirt: The Business of Professional Team Sports*. Princeton University Press. https://doi.org/10.2307/j.ctv346rfd

Simmons, R., & Berri, D. J. (2011). *Mixing the princes and the paupers: Pay and performance in the National Basketball Association.* Labour Economics, 18(3), 381-388. https://doi.org/10.1086/261010

Tao, Y. L., Chuang, H. L., & Lin, E. S. (2016). Compensation and performance in Major League Baseball: Evidence from salary dispersion and team performance. International Review of Economics & Finance, 43, 151-159. https://doi.org/10.1016/j.iref.2015.10.037

U.S. Bureau of Labor Statistics. (n.d.). Consumer Price Index for All Urban Consumers (CPI-U) [Data table]. Accessed September 22, 2025. https://data.bls.gov/timeseries/CUUR0000SA0.