Predicting NBA Player Performance: Analyzing the Impact of Minutes, Shot Attempts, and Efficiency on Total Points Per Game Using Simple Linear Regression (SLR) and Multiple Regression (MR)

Angelo Peng

Predictive Analysis (BUSN 6324)

Prof. Nizar Zaarour

Executive Summary

In today's NBA, statistical analysis is a critical component of understanding player performance and optimizing team strategies. Given the abundance of data available from the 2023-2024 NBA season, this report aims to investigate the factors that most effectively predict Points Per Game (PPG) for players, with a focus on understanding the relationship between key variables such as Minutes Played per Game, Field Goal Attempts per Game, Free Throw Attempts per Game, 3-Point Attempts per Game, and Field Goal Percentage (FG%).

Our analysis began with a simple linear regression examining the relationship between Minutes Played per Game (a direct indicator of opportunity) and Points Per Game. While Minutes Played showed a positive and statistically significant relationship with scoring, further investigation revealed that it only explains a portion of a player's scoring ability, capturing about 34.4% of the variation in points. This led us to expand our analysis by incorporating additional variables that reflect different aspects of scoring, such as shot volume, efficiency, and style of play.

Incorporating these variables into a multiple regression model significantly improved the model's accuracy, with an R-squared value of 0.9855, meaning that 98.55% of the variation in scoring can be explained by the combined effect of these variables. Field Goal Attempts per Game emerged as the strongest predictor of points, with Field Goal Percentage playing a crucial role in explaining shooting efficiency. Free Throw Attempts added further predictive power by reflecting scoring opportunities from fouls, while 3-Point Attempts showed variable impact depending on a player's shooting efficiency.

Residual analysis revealed that the model performed well overall but struggled slightly with players who had higher Field Goal Percentages and Minutes Played, suggesting that further refinements may be needed to capture the nuances of highly efficient shooters and players with extensive playing time.

In conclusion, while Minutes Played is an important factor in predicting scoring, our research confirms that Field Goal Attempts per Game and Field Goal Percentage are the most critical indicators of a player's scoring output. Free Throw Attempts and 3-Point Attempts add valuable context by reflecting different styles of play. Combining these factors in a multiple regression model provides a comprehensive understanding of what drives a player's scoring ability. These insights can help the team make more informed decisions regarding player rotations, shot distribution, and identifying key contributors to offensive success.

Further improvements to the model, such as accounting for shooting efficiency and playing time more dynamically, could lead to even more accurate predictions and better on-court strategies for maximizing our players' scoring potential.

Appendix

Appendix A.

Regression Analysis of Points Per Game (Y) & Minutes Per Game (X_1)

SUMMARY OUTPUT

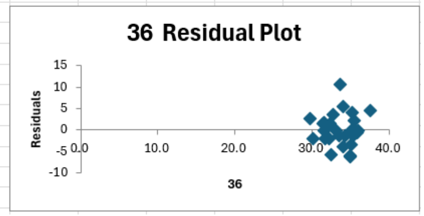
Regression Statistics	
Multiple R	0.586673811
R Square	0.34418616
Adjusted R Square	0.319896759
Standard Error	3.672761595
Observations	29

ANOVA

	df	SS	MS	F	Significance F
Regression	1	191.1446149	191.145	14.1702	0.000822767
Residual	27	364.2077989	13.4892		
Total	28	555.3524138			

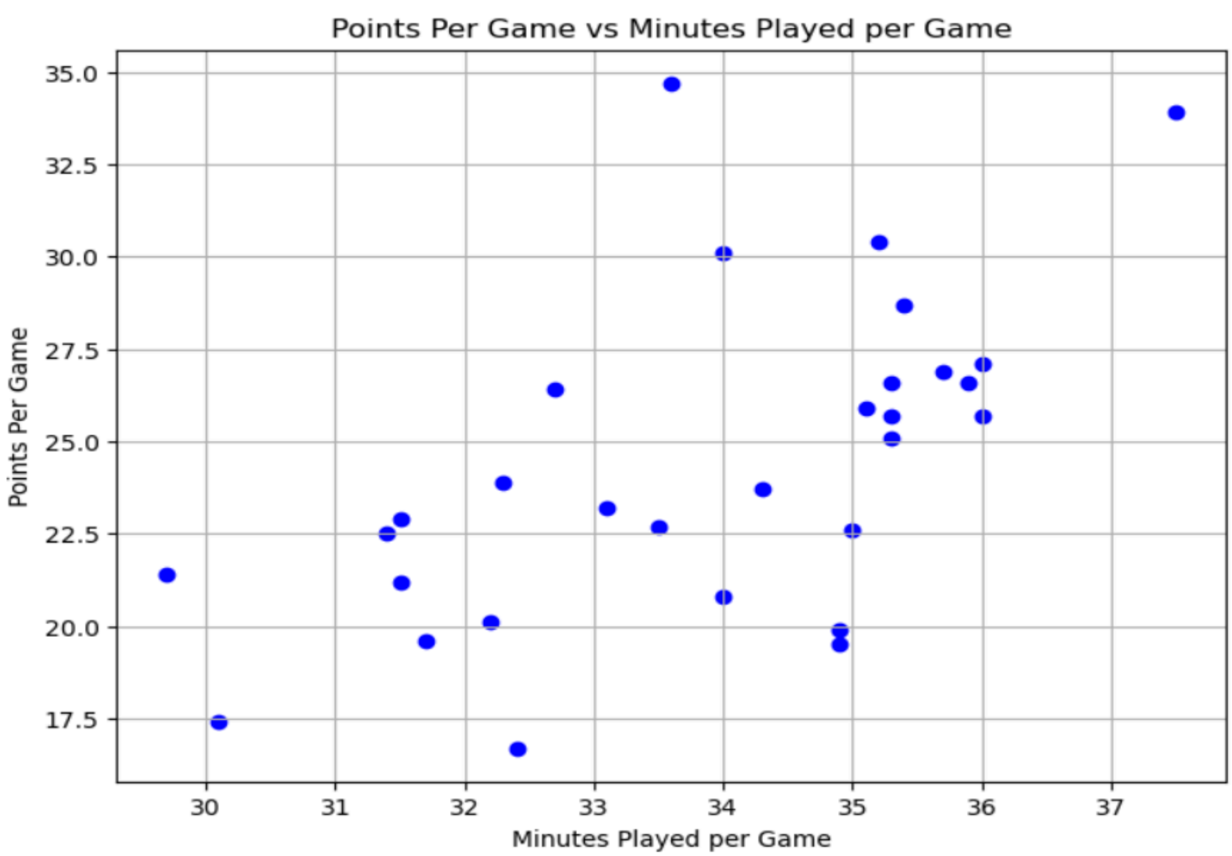
SSR	191.145
SSE	364.208
SST	555.352

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	-21.54142503	12.21062955	-1.76415	0.08902	-46.59556737	3.51272	-46.5956	3.51272
36	1.35875582	0.36095505	3.76434	0.00082	0.618137233	2.09937	0.61814	2.09937



Appendix B.

Scatter Diagram



Appendix C.

Regression Analysis of ALL Variables (PPG Y, PPG X_1, FGAPG X_2, FTAPG X_3, 3PTFGPG X_4, FG% X_5)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.992724622
R Square	0.985502175
Adjusted R Square	0.982350473
Standard Error	0.591659506
Observations	29

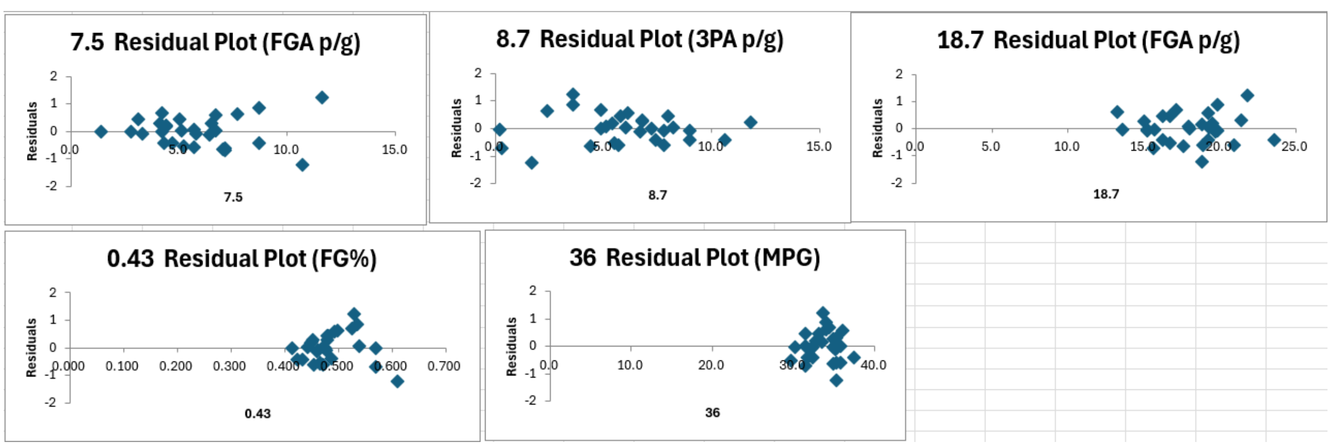
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	547.3010115	109.4602023	312.688964	2.40246E-20
Residual	23	8.051402322	0.350060971		
Total	28	555.3524138			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-15.60286595	2.546668776	-6.12677475	2.9966E-06	-20.87105169	-10.3346802	-20.8710517	-10.3346802
36	-0.05050775	0.073396975	-0.68814484	0.49824392	-0.20234096	0.10132546	-0.20234096	0.10132546
18.7	0.976343988	0.078992988	12.35988175	1.2227E-11	0.812934543	1.13975343	0.812934543	1.13975343
7.5	0.848523123	0.07826773	10.84128959	1.6301E-10	0.686613989	1.01043226	0.686613989	1.01043226
8.7	0.45347394	0.077955957	5.817053088	6.3E-06	0.292209756	0.61473812	0.292209756	0.61473812
0.43	34.04561099	3.939166288	8.642846861	1.1107E-08	25.89682467	42.1943973	25.89682467	42.1943973

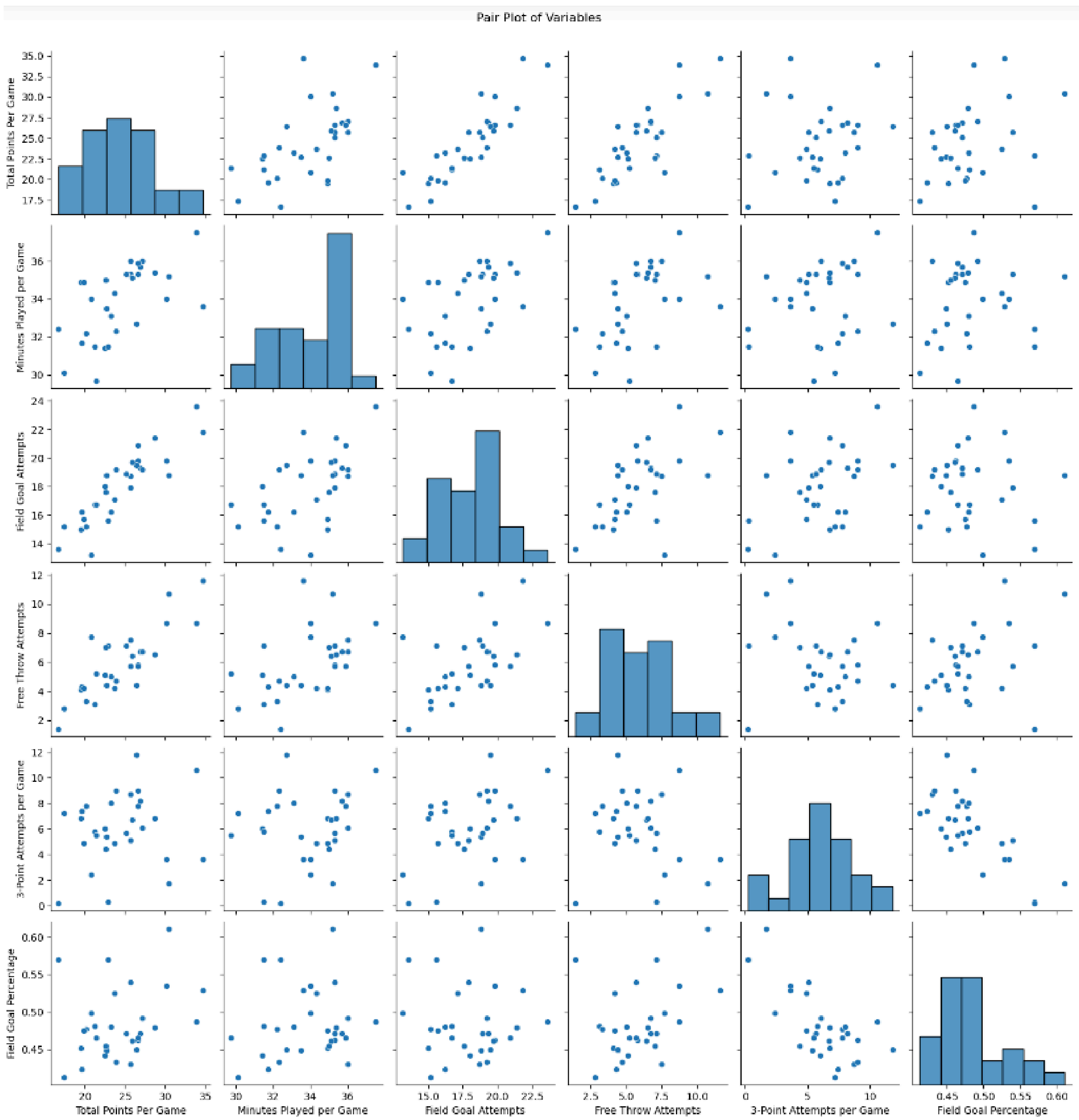
Appendix D.

Residual Plots for Appendix C.



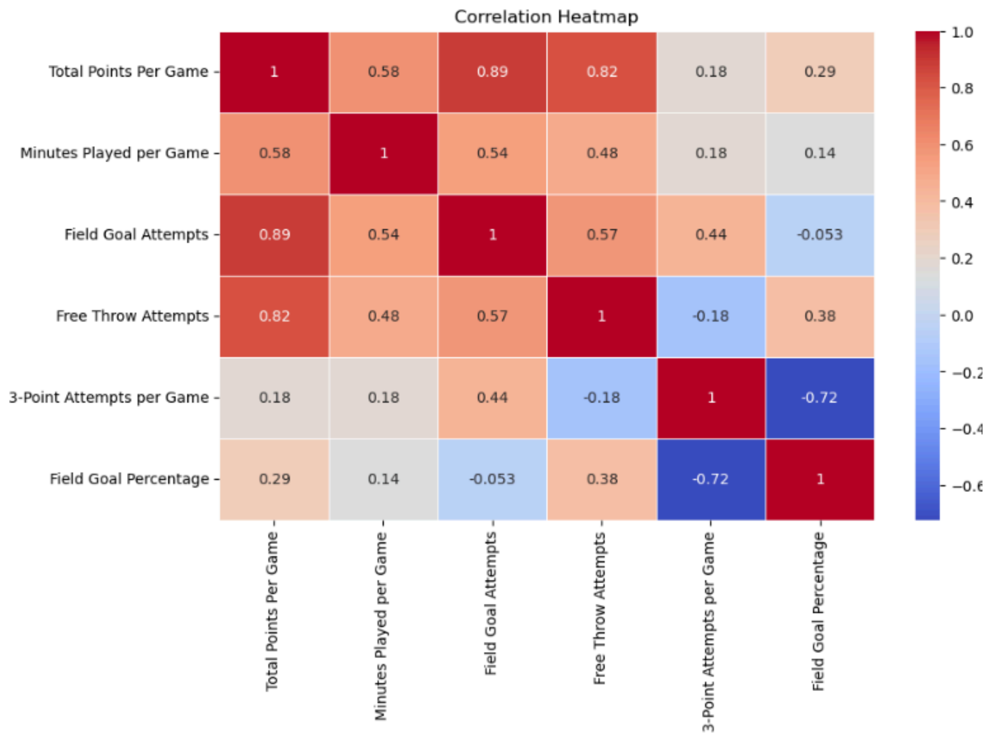
Appendix E.

Pair Plot (Scatter Matrix) of All Variables



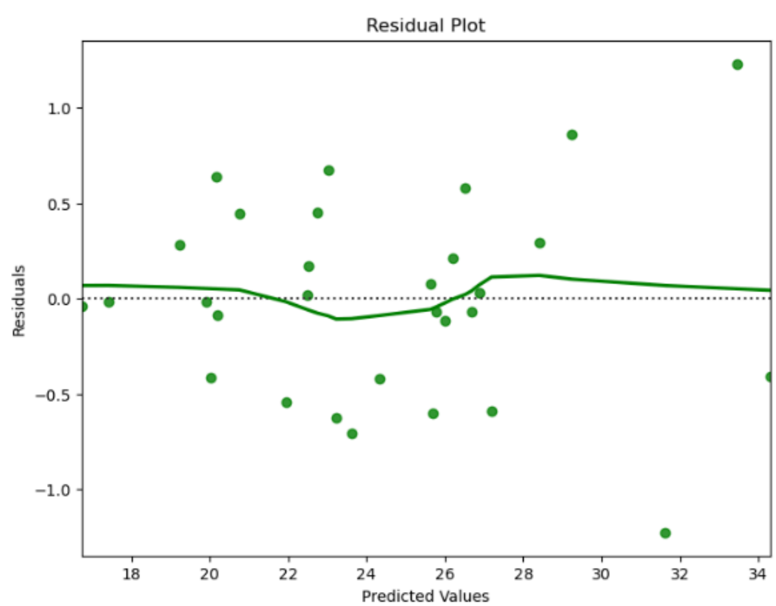
Appendix F.

Correlation Heatmap of All Variables



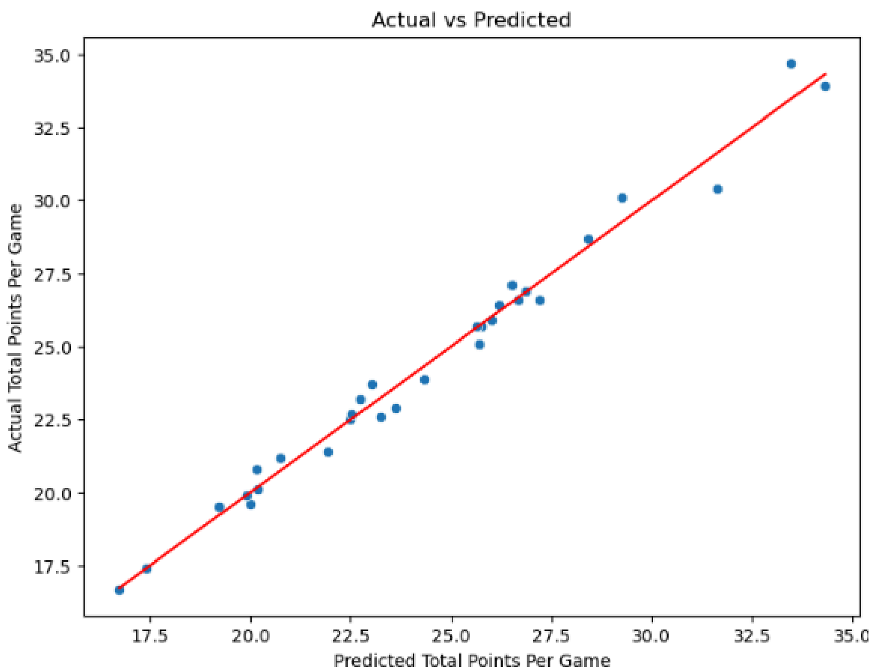
Appendix G.

Residual Plot of All Variables



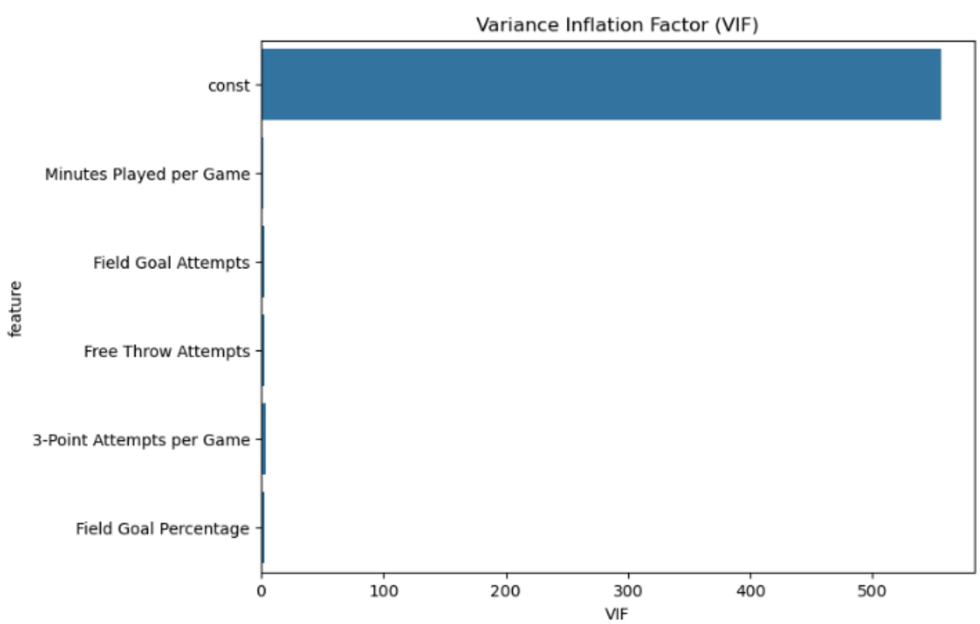
Appendix H.

Actually vs Predicted Plot of All Variables



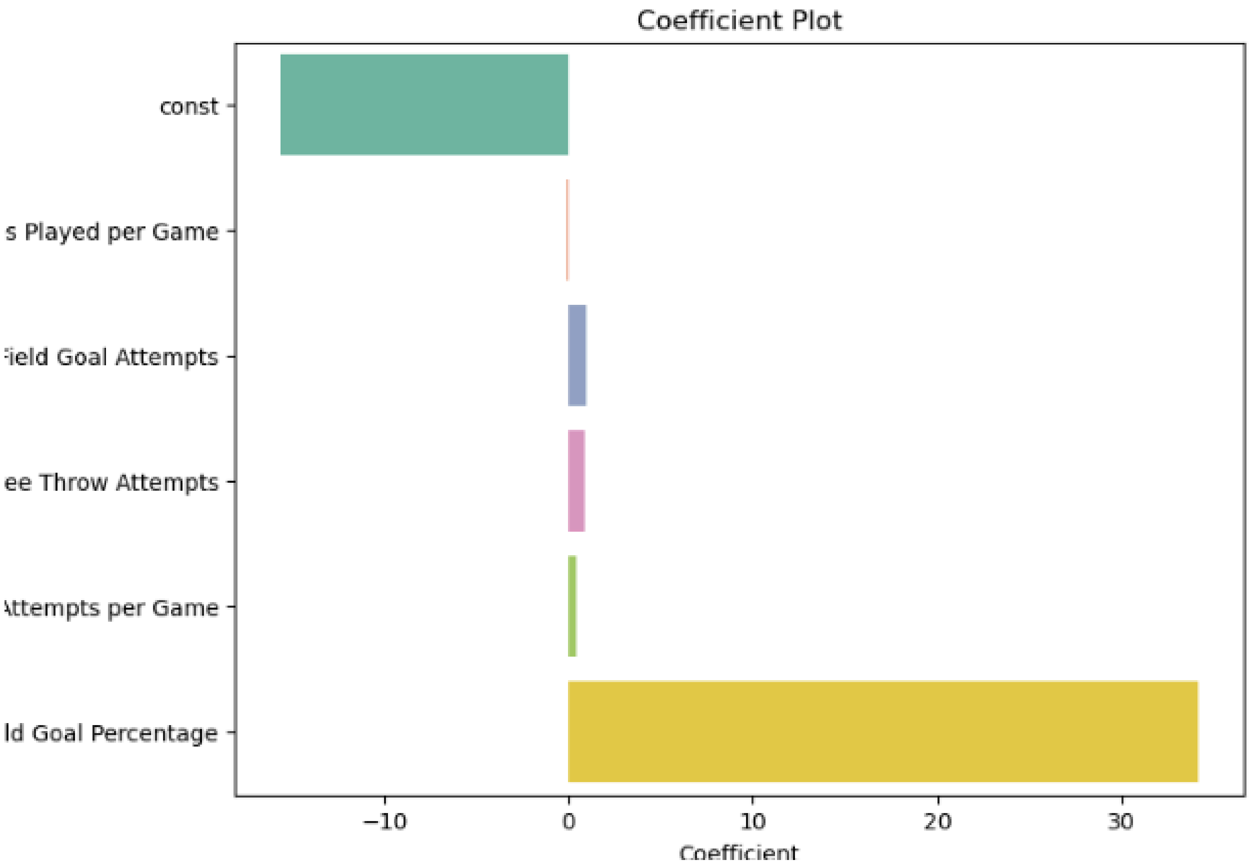
Appendix I.

Variance Inflation Factor (VIF) of All Variables



Appendix J.

Coefficient Plot of All Variables



Body of Report (Simple Linear Regression)

To begin, the analysis focuses on two variables:

- **Minutes Played per Game** (predictor or independent variable)
- **Points Per Game** (outcome or dependent variable)

Minutes Played per Game was selected because it is a fundamental factor in determining a player's opportunity to score. It is intuitive to expect that the more time a player spends on the court, the more chances they have to contribute offensively by taking shots, making plays, or scoring points.

Minutes played also reflect a player's role within the team. Star players or key scorers tend to play more minutes, as coaches rely on them for scoring and playmaking. Players who score the most are often the ones trusted to be on the court during key moments, including the closing minutes of tight games.

Players with consistently high minutes per game are often more integral to the team's strategy, which correlates with consistent performance in scoring. They are the go-to players, expected to perform game after game.

By using Minutes Played per Game as the predictor, we are capturing a player's direct opportunity to score. More time on the court means more chances for shots, assists, free throws, etc., which would naturally translate into more points over time.

Points Per Game are the most common metric used to measure a player's scoring performance. It directly reflects how many points a player contributes on average in each game, making it a critical statistic in understanding a player's impact on the court.

This variable reflects a player's scoring prowess, which is often a key measure of their offensive value. Players who consistently score points are often considered more valuable in offensive strategies.

Points are the direct currency of basketball. The ultimate goal of any game is to outscore the opponent, so a player's contribution to the scoreboard is a central component of their overall performance. Points per game thus directly ties into their offensive contribution to winning games.

Points per game provides a way to measure how consistently a player performs. A player averaging 25 points per game is more reliable than a player whose scoring fluctuates wildly between games.

As mentioned, the more time a player spends on the court, the more they can contribute. This means Minutes Played per Game is a proxy for opportunity—the more opportunities a player has (minutes), the more likely they are to score. Consequently, we expect to see a positive relationship between these two variables.

However, not all players are equally efficient or impactful with their minutes. Some players may score a lot in fewer minutes due to their role (like bench scorers) or because of their shooting efficiency. That's why Minutes Played per Game is a strong indicator but not the only factor that explains Points Per Game. Other factors like shooting percentage, field goal attempts, and free throw efficiency also contribute significantly.

In basketball, a player's scoring ability is often tied to how much they are trusted to be on the court. The best players tend to play more minutes because:

- They are relied upon in crucial moments.
- They have higher endurance and are expected to maintain performance over extended playing times.
- They play key roles in both offense and defense, making them more valuable to the team strategy.

Choosing these variables allows us to investigate a fundamental question in basketball: Does having more playing time lead to more points? The expectation is that a player who plays more minutes will have more opportunities to shoot, which will translate into more points scored per game. However, the regression analysis also allows us to see that this relationship, while significant, is only part of the story. It suggests that other variables could also play a critical role, such as shooting efficiency, shot attempts, and position or role within the team.

By focusing on these variables, we are tapping into two key aspects of a player's performance: their opportunity to score (via minutes played) and their actual scoring performance (via points per game). Together, these metrics provide valuable insights into how a player's contribution is reflected in the number of points they score, as well as how game time allocation influences their ability to perform.

Based on the data provided for 29 observations (players), we observe the following general trends:

- Players with more minutes played generally score more points, but the variation between individuals is notable.
- The average minutes played is concentrated around 35 minutes per game, with points per game ranging from around 17.5 to 35.

The next step in the analysis is to fit a **simple linear regression** model to describe the relationship between the two variables. The regression equation derived from the analysis is:

Points Per Game= $-21.54 + 1.36 \times$ (Minutes Played Per Game)

This means that for every additional minute a player spends on the court, their points per game are expected to increase by approximately 1.36 points. The intercept, -21.54, while mathematically necessary, has little real-world interpretation since playing 0 minutes is not practical in this context.

In order to assess the goodness of fit of the model, we performed an ANOVA (Analysis of Variance) test. Here are the key results from the ANOVA table (see Appendix A):

- SST (Total Sum of Squares): 555.35
- SSR (Regression Sum of Squares): 191.14 (This is the portion of the variability explained by the model)
- SSE (Error Sum of Squares): 364.21 (This is the portion of the variability not explained by the model)

The R-squared value is 0.344, which means that the model explains about 34.4% of the variation in points per game based on minutes played per game. This is a moderate fit, suggesting that while Minutes Played is a good predictor, there are other factors influencing Points Per Game that are not captured by this model.

To test whether the relationship between **Minutes Played per Game** and **Points Per Game** is statistically significant, we performed an **F-test**:

- The **F-statistic** is **14.17**, and the associated **p-value** is **0.000823**.
- Since the p-value is far below the conventional threshold of 0.05, we can conclude that the relationship between **Minutes Played per Game** and **Points Per Game** is statistically significant.

The **Residual Plot** shows how well the model's predictions align with the actual data. Ideally, residuals (the differences between actual and predicted values) should be randomly scattered around zero, which would indicate a good model fit. In this case, the residuals are mostly small and scattered within a similar range, suggesting that the model performs reasonably well for most predictions. However, there is some clustering of predictions between 30 and 40 points, which may suggest that the model is less effective at predicting points outside of this range. This could mean that other factors influencing scoring are not being captured by the model.

Next, we analyzed the relationship between Points Per Game and Minutes Played Per Game using a scatter diagram (see Appendix B). The scatter plot (as seen in the provided figure) shows that there is a positive relationship between the two variables: as Minutes Played Per Game increases, Points Per Game also tends to increase. However, the points are somewhat scattered, indicating that while the relationship is positive, it is not perfectly linear. The data points suggest that more playing time generally leads to higher scoring, though with some variation.

In conclusion, the analysis shows that there is a statistically significant positive relationship between the number of minutes a player spends on the court and the points they score. Specifically, for every additional minute played, a player's points per game increase by approximately 1.36 points. However, the model explains only 34.4% of the variation in scoring, meaning that other factors (such as shooting efficiency, player role, or team strategy) likely contribute to the variability in points per game that are not accounted for by the model.

While the model offers valuable insights, it is clear that Minutes Played is not the sole factor in determining a player's points per game. To improve the model and gain a better understanding of what drives scoring, it would be beneficial to consider other variables such as Field Goal Attempts, 3-Point Attempts, or Free Throw Attempts.

Overall, the regression model is statistically significant, and **Minutes Played per Game** has a meaningful impact on **Total Points Per Game**. However, the **R-Squared** value of 0.3442 indicates that there are other factors affecting scoring that are not included in this model. The residual analysis also shows that while the model works well for most predictions, it may struggle with extreme values. To improve the model's accuracy, additional variables such as **Field Goal Attempts**, **Free Throw Attempts**, **Three-Point Attempts** or **Shooting Efficiency** could be added to explain more of the variation in scoring. This would increase the model's explanatory power and provide a better fit for the data.

<u>Multiple Regression</u>

After initially analyzing the relationship between **Minutes Played per Game** and **Points Per Game**, we proceeded to incorporate additional independent variables (or predictors) to improve the model's accuracy and capture other aspects of a player's scoring ability. These additional variables include **Field Goal Attempts per Game**, **Free Throw Attempts per Game**, **3-Point Attempts per Game**, and **Field Goal Percentage (FG%)**. Each of these variables contributes uniquely to determining a player's scoring, providing a more complete picture of their offensive performance.

Each of these additional independent variables provides important insights into different aspects of a player's scoring ability. While Minutes Played per Game capture a player's opportunity to score, the inclusion of Field Goal Attempts, Free Throw Attempts, 3-Point Attempts, and Field Goal Percentage allows us to better understand a player's style, efficiency, and scoring versatility.

- Field Goal Attempts per Game captures scoring volume.
- Free Throw Attempts per Game account for a player's ability to draw fouls and score from free throws.
- 3-Point Attempts per Game reflect a player's reliance on long-range shooting, offering a high-reward option for scoring.
- Field Goal Percentage emphasizes scoring efficiency, which helps explain how effectively players convert their opportunities into points.

By including these variables in the model, we gain a more complete understanding of what drives a player's total points per game and can more accurately predict scoring outcomes.

The regression analysis output of all variables from Y, X_1, X_2, X_3, X_4, and X_5 (see Appendix C) provides a comprehensive overview of the relationship between the dependent variable and five independent variables. This analysis helps us assess the significance and strength of these relationships, as well as the overall fit of the model.

The **Multiple R** value of **0.9927** indicates a very strong positive correlation between the observed and predicted values, meaning that the model is highly effective at predicting the dependent variable. The **R-squared** value of **0.9855** suggests that approximately **98.55%** of the variation in the dependent variable is explained by the independent variables. This indicates an excellent model fit. The **Adjusted R-squared** is **0.9824**, which accounts for the number of predictors in the model. Since the adjusted R-squared is very close to the regular R-squared, it confirms that the independent variables included in the model contribute significantly to explaining the variation in the dependent variable. The **Standard Error** of **0.5917** shows that, on average, the predicted values deviate from the actual values by about 0.59 units, suggesting that the model's predictions are quite accurate.

The ANOVA table breaks down the variation in the model into regression and residual components. The Regression Sum of Squares (SSR) is 547.3010, which indicates that a substantial portion of the total variation is explained by the model. The Residual Sum of Squares (SSE) is 8.0514, meaning only a small portion of the variation is left unexplained. This small residual error further supports the idea that the model provides a good fit to the data. The F-statistic is 312.69, with a Significance F (p-value) of 2.40E-20, which is much smaller than 0.05. This indicates that the overall regression model is highly statistically significant, meaning that the independent variables, taken together, provide a meaningful explanation of the variation in the dependent variable.

- At $\alpha = 0.1$, the p-value (2.40E-20) is far below 0.1, indicating that the model is highly significant.
- At $\alpha = 0.05$, the p-value is still far below 0.05, confirming the model's significance.
- At $\alpha = 0.01$, the p-value remains well below 0.01, reinforcing that the model is statistically significant.

We can conclude that regardless of the significance level, the F-test shows that the independent variables, together, have a significant relationship with Points Per Game.

The residual plots (see Appendix D) show the differences between the actual and predicted values for several variables in the regression model, including Field Goal Attempts per Game (FGA p/g), 3-Point Attempts per Game (3PA p/g), Minutes Played per Game (MPG), and Field Goal Percentage (FG%). Residual plots help us understand how well the model is predicting the outcomes for each variable. Ideally, the points should be randomly scattered around the zero line, meaning the model is making errors in a random, unpredictable way. Below is an explanation of what each plot shows and what we can conclude from them.

For Field Goal Attempts per Game (7.5), the points are mostly scattered around the zero line without any clear pattern. This suggests that the model is doing a good job of predicting how field goal attempts influence the outcome. While there are a few points that deviate a little, overall, the errors are small, and the model seems to handle this variable well.

In the plot for 3-Point Attempts per Game (8.7), the points are also scattered randomly around zero, without any obvious trend or pattern. This tells us that the model fits this variable pretty well, too. There may be a slight spread of points at the extremes (players who take a lot of 3-point shots), but overall, the model is doing a good job of incorporating 3-point attempts into the predictions.

For Field Goal Attempts per Game (18.7), the points are again small and well-distributed around the zero line. This means the model is doing a solid job of predicting outcomes based on field goal attempts. The errors are small and balanced across the range of values, so the model seems to accurately capture the impact of field goal attempts.

The residuals for Field Goal Percentage (FG%) tell a different story. While many points are close to zero, there's an increasing spread of points as FG% goes up. This suggests that as players become more efficient with their shots, the model's errors get larger. This could mean the model is not capturing the full complexity of how shooting efficiency impacts the outcome. The model may need some adjustment to better handle this variable, as the increasing spread of errors suggests the predictions become less accurate as field goal percentage rises.

For Minutes Played per Game (MPG), the points are mostly clustered toward the higher end of the plot (around 30 to 40 minutes). While most residuals are small, this clustering suggests that the model may not be as reliable when predicting players who play a lot of minutes. The model might struggle a bit to capture the full impact of playing time, especially for those who are on the court more frequently.

The p-values for each predictor allow us to determine which independent variables are statistically significant in predicting Points Per Game.

- 36 (p-value = 0.4982): Not statistically significant at any commonly used significance level (α = 0.1, 0.05, or 0.01).
- 18.7 (p-value = 7.89E-12): Statistically significant at all levels, showing a strong relationship with the dependent variable.
- 7.5 (p-value = 1.63E-10): Also statistically significant at all levels, meaning it strongly contributes to the prediction of Points Per Game.
- 8.7 (p-value = 6.30E-06): Statistically significant at all levels.
- 0.43 (p-value = 1.11E-08): Highly significant at all levels.

Four of the five independent variables (18.7, 7.5, 8.7, and 0.43) are highly significant predictors of Points Per Game, while 36 is not statistically significant. Therefore, 36 may not contribute much to predicting the dependent variable and could potentially be removed or further investigated.

Overall, the model does a good job for most of the variables. For Field Goal Attempts per Game (FGA p/g), 3-Point Attempts per game (3PA p/g), and Minutes Played per Game (MPG), the errors are small and random, showing that the model handles these variables well. However, the plot for Field Goal Percentage (FG%) shows signs that the model has more difficulty accurately predicting outcomes for this variable, especially as players become more efficient with their shots. This might mean that the model needs some improvements to better account for changes in shooting efficiency and playing time.

The Pair Plot/Scatterplot Matrix (see Appendix E) provides a visual representation of the relationships between all pairs of variables in the dataset. It allows us to observe how the independent and dependent variables relate to each other. If we notice that the points form a clear linear pattern, it suggests a strong relationship between those variables. On the other hand, if the points are randomly scattered with no clear trend, it indicates that there is little to no relationship. For example, if there is a visible linear trend between "Minutes Played per Game" and "Total Points Per Game," we can conclude that players who play more minutes tend to score more points. If no pattern is seen between other variables and the target variable, these may not be useful for predicting total points, and we might consider excluding them from the model.

The Correlation Heatmap (see Appendix F) shows how strongly different variables are related to each other. The colors in the heatmap tell us how strong the relationship is, with values closer to 1 meaning a strong positive relationship (as one variable increases, the other does too) and values closer to -1 meaning a strong negative relationship (as one increases, the other decreases). For example, if Minutes Played per Game and Total Points Per Game have a strong positive correlation, it means that playing more minutes is closely tied to scoring more points. However, if two independent variables, like Field Goal Attempts and 3-Point Attempts, are too closely related, it can create problems for the model. In this case, one of the variables might need to be removed to keep the model stable.

The Residual Plot (see Appendix G) helps us see how well the model is doing by showing the difference between the actual values and the predicted values. Ideally, these differences (called residuals) should be randomly scattered around the zero line, meaning the model is doing a good job of predicting without making any systematic mistakes. If the residuals start forming a pattern, it could mean the model is missing some important relationships. If the errors get larger as the predicted values increase, this could indicate a problem called heteroscedasticity, where the model's accuracy becomes inconsistent, which can affect its reliability.

The Actual vs. Predicted Plot (See Appendix H) compares the real values of Total Points Per Game to what the model predicts. If the points on this plot are close to the diagonal line, it means the model is making accurate predictions. If many points are far away from the line, it indicates the model is less reliable for those data points. In that case, the model may need improvement by adding more variables or trying a different approach to better capture the relationships between variables.

The Variance Inflation Factor (VIF) (see Appendix I) helps check if there’s high correlation between variables in the model, meaning some variables are too closely related to each other. High VIF values (above 5 or 10) suggest that high correlation is a problem, which makes it hard to determine the true effect of each variable on the outcome. If the VIF values are low (below 5), it means high correlation isn't an issue, and each variable provides unique, useful information to the model.

The Distribution of Residuals/Histogram (see Appendix J) shows how the errors (residuals) are spread out. Ideally, the residuals should form a bell-shaped curve centered around zero, meaning the model is making predictions that are balanced—sometimes it over-predicts, sometimes it under-predicts, but on average, it gets things right. If the residuals are skewed or have multiple peaks, it suggests the model might be missing some important patterns in the data. In such cases, the model may need refining or a different type of model may be needed to capture the underlying relationships.

The combination of these graphs provides a clear view of how well the model is working and the relationships between different factors. The Coefficient Plot and Correlation Heatmap helps identify which factors (like Minutes Played or Field Goal Percentage) have the most impact on predicting points per game. The Residual Plot and Distribution of Residuals show how well the model fits the data and whether the errors are distributed in a balanced way. The Actual vs. Predicted Plot gives a clear picture of the model's accuracy, while the VIF Plot ensures that the predictors aren't too closely related to each other. Together, these visualizations help us evaluate how effective the model is and point out areas where it can be improved, leading to better predictions for total points per game.

Conclusion

The analysis of both simple and multiple regression models for predicting Points Per Game (Y) using variables such as Minutes Played per Game, Field Goal Attempts per Game, Free Throw Attempts per Game, 3-Point Attempts per Game, and Field Goal Percentage reveals several key insights. While Minutes Played per Game consistently showed a positive relationship with scoring, it was not the strongest predictor; Field Goal Attempts per Game emerged as a primary driver of point totals, with players who take more shots generally scoring more. Free Throw Attempts per Game also added predictive power by reflecting additional scoring opportunities, while 3-Point Attempts per Game showed variable effects, with greater impact depending on shooting efficiency. Field Goal Percentage (FG%), an important measure of efficiency, significantly contributed to the model, showing that players with higher shot accuracy tend to score more points even with fewer attempts. Although there were some signs of overlap or strong correlation among variables like Field Goal Attempts, Minutes Played, and 3-Point Attempts, it wasn't severe enough to impact model stability. The multiple regression model provided the most complete picture, balancing both shot volume and efficiency to explain scoring performance. The high R-squared values indicated a strong model fit, but residual analysis suggested some areas for improvement, particularly with FG% and Minutes Played at higher values. Overall, the model captured key aspects of scoring, but refining it further could improve its performance with more complex relationships.

The analysis of both the Simple Linear Regression (SLR) and Multiple Regression (MR) models revealed some limitations that could affect the insights gained from the study. In the SLR model, Minutes Played per Game was the sole predictor of Points Per Game, explaining only 34.4% of the variation in scoring. This means that while minutes are important, they do not capture all the factors influencing a player's scoring ability, such as shot efficiency and shot volume. This limited scope leaves other important variables unexplored. In the MR model, which included variables like Field Goal Attempts, 3-Point Attempts, Free Throw Attempts, and Field Goal Percentage, some strong correlations were observed, particularly between Minutes Played, Field Goal Attempts, and 3-Point Attempts. Although these correlations were not severe enough to impact model stability, they make it difficult to isolate the effect of each variable on scoring. Additionally, the residual analysis showed some clustering of errors, particularly for players with higher Field Goal Percentage and Minutes Played, suggesting the model might struggle to capture the full impact of highly efficient shooters or players with significant playing time.

Moreover, the model does not account for team dynamics (e.g. chemistry and morale) or player roles/positions, which are critical in understanding individual performance. For example, some players may score efficiently with fewer minutes because they have specific offensive roles, while others may play more minutes in a defensive capacity (e.g. 3&D players), contributing less to scoring. Another limitation is that external factors such as game tempo, matchup strength, and team strategies are not considered, despite their significant influence on player performance. This absence leaves key questions unanswered about how such external factors affect scoring.

Several questions remain unanswered: How do the interactions between different variables affect scoring? How does a player's role, position, and team strategy influence their scoring ability? How do situational factors like game tempo and opponent strength impact player performance? And lastly, can this model be used to predict future performance based on the data?

To address these limitations and unanswered questions, the model could be extended to include interaction terms or nonlinear relationships, which would capture the combined effects of variables like Field Goal Attempts and Field Goal Percentage on scoring. Additionally, incorporating player position and player role as categorical variables would add important context to the analysis, differentiating the performance of key scorers from role players. Furthermore, including external factors like game tempo, matchups, and team strategies would enhance the model’s ability to explain variations in scoring based on different game conditions. Finally, developing a predictive model based on past data, accounting for factors like fatigue and injury, could provide more actionable insights for forecasting player performance in future games.

By addressing these limitations, we can improve the model’s accuracy and offer more comprehensive insights into the factors driving scoring performance in the NBA, helping the team make better-informed decisions around player utilization and game strategy.

In conclusion, while Minutes Played per Game is an important factor in predicting Points Per Game, it’s clear that Field Goal Attempts per Game and Field Goal Percentage are the most critical predictors. Free Throw Attempts and 3-Point Attempts also add value, reflecting different styles of play. Combining these variables in a multiple regression model provides the best insight into what drives a player’s scoring, with both shot volume and efficiency playing key roles. To further improve the model, addressing some of the limitations identified in the residual analysis (like the challenges with FG% and higher MPG) could lead to even more accurate predictions.