

# *Feng Chia University*

## *Outstanding Academic Paper by Students*

What Are the Important Factors for NBA Player Salaries in 2017?

**Author(s):** 鄭哲睿(Zhe-Rui Zheng)、朱天婕(Tian-Jie Zhu)、林翰博(Han-Bo Lin)、薛藝湛(Yi-Zhan Xue)、周佳玉(Jia-Yu Zhou)

**Class:** 2nd year of SJSU-FCU 2+2 Bachelor's Program in Business Analytics

**Student ID:** D0571987、D0571926、D0571960、D0571930、D0572026

**Course:** Introduction to Data Analytics

**Instructor:** Dr. Cathy W. S. Chen

**Department:** International School of Technology and Management

**Academic Year:** Semester 1, 2017-2018

## **Abstract**

It is of great interest to identify the factors that influence the salaries of National Basketball Association (NBA) players. This study examines the 2017-2018 wages of 100 NBA players which are randomly selected by the SAS software based on their career performance variables using a multiple linear regression. There are 28 explanatory variables which include age, 3-point field goals per game and free throws per game. The multiple regression analysis is conducted to determine the explanatory variables which are helpful in predicting the salaries of NBA players. Five methods for model selection are used, these include forward selection, backward elimination, stepwise selection, adjusted R-square selection method and C(p) method. All five methods demonstrated similar results. Results indicated that variables such as games started, field goals per game, total rebounds per game, personal fouls per game, also the terms of contract used, have a significant correlation with salary.

**Keyword :** National Basketball Association, Multiple Linear Regression, Model Selection, Multicollinearity, Influential Point, Outliers

## Table of Content

I. Introduction.....	5
II. Method.....	6
i. Data Description.....	6
ii. Scatter Plot and Basic Statistics .....	7
iii. Variable Explanation .....	9
iv. Variable Selection .....	11
v. Model Representation.....	14
III. Model Analysis .....	16
i. Outliers Analysis .....	16
ii. Influential Point Analysis .....	16
iii. Four Assumption Verification .....	18
III. Findings and Discussion .....	21
IV. Appendix .....	22
i. Data Resources .....	22
ii. References .....	22
iii. Outlier and Influential Point Analysis.....	22
iv. Scatter Plot.....	24

## List of Figures

Figure 1: Scatter Plot Between Salary and Age .....	25
Figure 2: Scatter Plot Between Salary and G.....	25
Figure 3: Scatter Plot Between Salary and GS .....	25
Figure 4: Scatter Plot Between Salary and MP .....	25
Figure 5: Scatter Plot Between Salary and FG .....	25
Figure 6: Scatter Plot Between Salary and FGA.....	25
Figure 7: Scatter Plot Between Salary and THP .....	26
Figure 8: Scatter Plot Between Salary and THPA.....	26
Figure 9: Scatter Plot Between Salary and THPPer.....	26
Figure 10: Scatter Plot Between Salary and TWP .....	26
Figure 11: Scatter Plot Between Salary and TWPA.....	26
Figure 12: Scatter Plot Between Salary and TWPPER.....	26
Figure 13: Scatter Plot Between Salary and eFGPer .....	27
Figure 14: Scatter Plot Between Salary and FTA .....	27
Figure 15: Scatter Plot Between Salary and FTPer.....	27
Figure 16: Scatter Plot Between Salary and ORB .....	27
Figure 17: Scatter Plot Between Salary and DRB .....	27
Figure 18: Scatter Plot Between Salary and FGPer .....	27
Figure 19: Scatter Plot Between Salary and TRB .....	28
Figure 20: Scatter Plot Between Salary and FT .....	28
Figure 21: Scatter Plot Between Salary and STL .....	28
Figure 22: Scatter Plot Between Salary and AST .....	28
Figure 23: Scatter Plot Between Salary and BLK.....	28
Figure 24: Scatter Plot Between Salary and TOV .....	28
Figure 25: Scatter Plot Between Salary and PFA.....	29
Figure 26: Scatter Plot Between Salary and PSG .....	29

### List of Tables

Table 1: Descriptive statistics for the 100 NBA players in the 2016-2017 season .....	8
Table 2: Variance inflation factor for the variables .....	14
Table 3: Influential points for the variables .....	17
Table 4: Test for Location $\mu_0=0$ .....	19
Table 5: Heteroscedasticity Test.....	19
Table 6: Durbin-Watson Test.....	20
Table 7: Test for Normality .....	20



## Chapter 1

### Introduction

With more companies sponsoring the NBA and a higher salary cap, an increasing number of players are asking for a higher salary. Furthermore, there is already a huge salary gap between NBA players. In the 2017-2018 season, the lowest salary drawn is 25,000 dollars. However, Stephen Curry earned the highest salary in the NBA at 34,682,550 dollars, which is almost one thousand four hundred times that of the lowest.

Moreover, aside from the huge salary gap, overpaid players exist in every team. Forbes magazine listed Carmelo Anthony as the most overpaid player of 2018. As his pace to produce wins stands at negative 1.3, this means that the team is more likely to lose the game if he is playing. In addition to that, his field goal percentage, rebounds and assists are all below career average. With such lousy performances, his salary remains very high by NBA standards at 26,243,760 dollars, making it the twelfth highest in the NBA. This financial arrangement undoubtedly influences the balance of the team. With an overpaid player on the books, the team manager would find it difficult to trade for better players with the limit on the salary cap. As a fan of the NBA, I am constantly curious whether it is worthwhile to pay such a high salary and whether the players' performance out on the basketball court influences their salary? Hence in this project, I seek to analyze the latest data set which was collected from

Kaggle to find a regression line in the data to predict the players' salaries.

Furthermore, I will also focus on unexpected factors which may influence players' salaries.

Regarding the methodology, due to the large number of variables in the data set, We first selected several variables that might be useful. Then by using a scatter plot, we would have a brief idea of the relationship between the response variable and the explanatory variables. Later on, We use the five methods entirely which includes stepwise selection, forward selection, backward elimination, adjusted R-square selection method and C(p) selection method to choose the variables for the final model. Lastly, we made sure that no multicollinearity exists in the final model and verified all four assumptions.

## **Chapter 2**

### **Method**

#### **2.1 Data Description**

The data set we selected was based on the 2016-2017 season. The website named Basketball Reference provided the entire data set. For players included in the data set, this study will only analyse those whose number of games played is more than 25. Due to the amount of games played being too small, the statistics does not clearly reflect the players' ability, so the regression model cannot accurately predict their salary. Moreover, those who were signed using the FRP (first round pick) will not be

used. According to the rules, the rookies' salaries are constrained by the rookie salary cap. Hence if their performance is better as compared to the latter season, their salaries remain stagnant rather than indicating an increase in wages. Lastly, 100 players will be randomly selected from the entire data set.

## **2.2 Scatter Plot and Basic Statistics**

Figures 4, 5, 6, 10, 11, 14, 17, 20, 24, 26 in the appendix present the MP (minutes played per game), FG (field goals per game), FGA (field goal attempts per game), TWP (2-point field goals per game), TWPA (2-point field goal attempts per game), FTA (free throw attempts per game), DRB (defensive rebounds per game), FT (free throws per game), TOV (turnovers per game) and PSG (points per game), all of which have a positive correlation with salary. For other variables, the scatter plot does not indicate a clear relationship with salary levels. Moreover, the scatter plot did not exhibit any significant outliers and it was not necessary to delete any data points in this step.

Table 1 illustrates the basic statistics which include the mean, average, standard deviation, minimum and maximum for each variable. Several of the minima variable are equal to zero; for example, in the case of the GS (games started) variable, it means that this player was not in the starting line-up for the entire season. Hence in the next



step, a detailed explanation will be outlined along with the processing method for each variable.

Table 1 Descriptive statistics for the 100 NBA players in the 2016-2017 season

<b>Variable</b>	<b>N</b>	<b>Mean</b>	<b>Std Dev</b>	<b>Sum</b>	<b>Minimum</b>	<b>Maximum</b>
<b>SAL</b>	100	0.93888	0.81378	93.88761	0.00250	3.32857
<b>Age</b>	100	27.77000	3.53869	2777	19.00000	36.00000
<b>G</b>	100	66.80000	15.19436	6680	25.00000	82.00000
<b>GS</b>	100	38.77000	29.68108	3877	0	81.00000
<b>MP</b>	100	24.32000	7.45576	2432	4.90000	37.80000
<b>FG</b>	100	0.64645	0.17174	64.64514	0.23045	1.03743
<b>FGA</b>	100	8.28300	4.02951	828.30000	1.90000	20.90000
<b>FGPer</b>	100	0.45722	0.06267	45.72200	0.29200	0.65200
<b>THP</b>	100	0.93400	0.73817	93.40000	0	3.20000
<b>THPA</b>	100	2.62800	1.91201	262.80000	0	7.80000
<b>THPPer</b>	100	0.29941	0.12404	29.94100	0	0.45300
<b>TWP</b>	100	1.91894	0.41153	191.89409	1.18322	3.20936
<b>TWPA</b>	100	5.65100	3.30153	565.10000	1.00000	19.20000
<b>TWPPer</b>	100	0.50031	0.05952	50.03100	0.34200	0.65800
<b>eFGPer</b>	100	0.51268	0.04856	51.26800	0.36500	0.65200
<b>FT</b>	100	0.06113	0.34948	6.11274	-1	0.86923

## What Are the Important Factors for NBA Player Salaries in 2017?

<b>FTA</b>	100	2.06700	1.65682	206.70000	0.10000	8.70000
<b>FTPer</b>	100	0.74915	0.12232	74.91500	0.27300	1.00000
<b>ORB</b>	100	0.99400	0.86676	99.40000	0	4.30000
<b>DRB</b>	100	0.44702	0.26002	44.70178	-0.69897	1.01284
<b>TRB</b>	100	4.26800	2.58167	426.80000	0.20000	14.10000
<b>AST</b>	100	1.39256	0.57188	139.25616	0.44721	3.27109
<b>STL</b>	100	0.78600	0.39467	78.60000	0	2.00000
<b>BLK</b>	100	0.44300	0.38462	44.30000	0	2.10000
<b>TOV</b>	100	0.03666	0.24264	3.66553	-0.69897	0.61278
<b>PFA</b>	100	1.89200	0.61162	189.20000	0.20000	3.30000
<b>PSG</b>	100	10.07000	5.39999	1007	2.20000	27.30000

---

### 2.3 Variable Explanation

This section will explain one response variable and twenty-eight explanatory variables which are of interest along with the methodology through which the variables are processed.

#### Response Variable

SAL is considered the response variable. SAL represents the NBA players' annual salary in the denomination of ten million.

#### Explanatory Variables

## What Are the Important Factors for NBA Player Salaries in 2017?

1. Age: Age of Player at the start of the season, February 1st
2. G: Games
3. GS: Games Started
4. MP: Minutes Played Per Game
5. FG: Field Goals Per Game
6. FGA: Field Goal Attempts Per Game
7. FGPer: Field Goal Percentage
8. THP: 3-Point Field Goals Per Game
9. THPA: 3-Point Field Goal Attempts Per Game
10. THPPer: FG% on 3-pt FGAs
11. TWP: 2-Point Field Goals Per Game
12. TWPA: 2-Point Field Goal Attempts Per Game
13. TWPPer: FG% on 2-pt FGAs
14. eFGPer: Effective Field Goal Percentage, this statistic adjusts for the fact that a 3-point field goal is worth one point more than a 2-point field goal.
15. FT: Free Throws Per Game
16. FTA: Free Throw Attempts Per Game
17. FTPer: Free Throw Percentage
18. ORB: Offensive Rebounds Per Game

19. DRB: Defensive Rebounds Per Game

20. TRB: Total Rebounds Per Game

21. AST: Assists Per Game

22. STL: Steals Per Game

23. BLK: Blocks Per Game

24. TOV: Turnovers Per Game

25. PFA: Personal Fouls Per Game

26. PSG: Points Per Game

27. Pos : Position, this is a dummy variable where six positions are included, they are PG, SG, SF, PF, C, PF-C.

28. Signed Using: this is a dummy variable where six signs are included, they are FRP (first round pick), BR (Bird Right), CS (Cap Space), MS (Minimum Salary), Others (Signed like MLE, Bi-annual Exception are included) and None.

However as aforementioned, players with the FRP sign will not be used.

## **2.4 Variable Selection**

Five methods were utilised in the selection of variables; stepwise selection, backward elimination, forward selection, adjusted R-square selection method and C(p) selection method. However, for the adjusted R-square selection method and C(p) selection method, grouping information is ignored. Hence the study primarily uses stepwise

selection, backward elimination and forward selection to choose the variable. To obtain a highly adjusted R-square and to avoid multicollinearity at the same time, it was decided that the variables would be selected through stepwise selection.

### **Backward Elimination**

Backward Elimination begins with a regression on all variables. Following that, every independent variable will be examined, thereby removing the variables with the smallest p-value. This process is repeated until no variables can be removed.

Thirteen variables are chosen through this method. They are Signed Using, GS, MP, FGPer, THPPer, TWPA, eFGPer, FTA, FT, STL, BLK, PFA, PSG. The  $\alpha$  is set at 0.15

### **Forward Selection**

Forward Selection begins with no variables. This method examines every variable and those with the most significant contribution will be added to the model. Following that, a new regression is run with a lesser variable. The step will be repeated until no variables can be selected.

Six variables are chosen through this method. They are Signed Using, GS, THP, FTA, TRB, PFA. The  $\alpha$  is set at 0.15

### **Stepwise Selection**

In Stepwise Selection, a variable can be added or deleted from the model several times before the final model is attained, and is also dependent on the other variables in the model.

Six variables are chosen through this method. They are Signed Using, GS, THP, FTA, TRB, PFA. The  $\alpha$  is set at 0.15 for both sides.

### **Adjusted R-square Selection Method**

Adjusted R-square Selection Method lists all possible models and calculates the adjusted R-square for each model. The model with the largest adjusted R-square will be considered the best model.

PG, SG, CS, MS, GS, MP, FGA, FGPer, THP, THPPer, eFGPer, FTA, FT, STL, BLK, PFA, PSG are the variables chosen through this method. However, in the adjusted R-Square Selection Method, grouping information is not considered.

Thus, the models selected by this method will not be used.

### **C(p) Selection Method**

Similar to the Adjusted R-square Selection model, the C(p) Selection Method also lists all possible models as the C(p) value is calculated for each model. The model with the smallest C(p) value will be considered the best model.

CS, MS, GS, FGA, eFGPer, TRB, PFA, PSG are the variables chosen by this method. During C(p) selection, grouping information is ignored, thus the models

selected by this method will not be used.

## 2.5 Model Representation

The final model was selected through stepwise selection. The final model is shown below.

$$\hat{y}_i = 0.1302 + 0.0074 \text{ GS} + 0.109 \text{ THP} + 0.2581 \text{ FTA} + 0.6653 \text{ TRB} - 0.2319 \text{ PFA} + 0.0572 \text{ BR} + 0.1909 \text{ CS} - 0.3384 \text{ MS}.$$

As the regression model shows, players' salary increased along with an increase in the variables such as GS, THP, FTA, and TRB. However, a higher number of personal fouls in a game actually corresponds with a decrease in salary. Furthermore, players with the sign Cap Space are more likely to earn a higher salary.

### VIF

Multicollinearity exists if there is a substantial correlation between independent variables. To diagnose multicollinearity in this model, the variance inflation factor is used. If the factor is larger than 10, it can be assumed that multicollinearity exists in the model. As shown in table 2, there is no multicollinearity in the final model.

Table 2 Variance inflation factor for the variables

Parameter Estimates					
Variable	Parameter Estimate	Standard Error	t Value	Pr >  t	Variance Inflation

<b>Intercept</b>	0.13020	0.16386	0.79	0.4290	0
<b>GS</b>	0.00735	0.00194	3.80	0.0003	2.03018
<b>THP</b>	0.10898	0.06611	1.65	0.1028	1.46239
<b>FTA</b>	0.25810	0.03178	8.12	<.0001	1.70243
<b>TRB</b>	0.06526	0.02450	2.66	0.0092	2.45638
<b>PFA</b>	-0.23193	0.08840	-2.62	0.0102	1.79504
<b>BR</b>	0.05715	0.15625	0.37	0.7154	2.51203
<b>CS</b>	0.19087	0.13227	1.44	0.1525	2.67373
<b>MS</b>	-0.33838	0.16005	-2.11	0.0373	1.67774
<b>Others</b>	-0.08458	0.18564	-0.46	0.6498	1.57305

### Adjusted R-square

The adjusted R-square is a modified version of R-square for the number of predictors in the model. The adjusted R-square can be calculated with the following formula:

$$adjustedR^2 = 1 - \frac{SSE_m/(n - m - 1)}{SST/(n - 1)}$$

The adjusted R-square for the final model is 0.7565. In other words, 75.65% of the data can be explained by the regression line.



## Chapter 3

### Model Analysis

This section mainly carries out the outliers analysis, the influential point analysis and the verification of four assumptions.

#### 3.1 Outliers Analysis

An observation that is markedly different from, or atypical of, the rest of the observations in a data set is known as an outlier. In the outlier analysis, the RStudent and student residual variables are primarily used in the study. If the RStudent and student residual are in excess of three, it confirms that the point is an outlier. The ninth data point is the only outlier in the data set.

#### 3.2 Influential Point Analysis

An observation that causes the regression estimates to be substantially different from what they would be if the observation was removed from the data set is called an influential observation. Besides that, influential observations are typically outliers that have high leverage. To judge the influential point, four entire methods are used; they are Cook's Distance, COVRATIO, DFFITS and DFBETAS respectively. The table is shown in the appendix.

#### Cook's Distance

If the Cook's Distance is larger than 0.5, the data might be influential. However, in the data set, none of the data points were greater than 0.5. Hence we can

conclude that the 9<sup>th</sup>, 45<sup>th</sup>, 52<sup>nd</sup>, 66<sup>th</sup>, 76<sup>th</sup>, 77<sup>th</sup>, 88<sup>th</sup> are influential points.

### DFFITS

If  $|DFFITS_i| > 2\sqrt{\frac{p}{n}}$ , it can be concluded that the  $i^{th}$  data is influential. Within the data set, if DFFITS is larger than 0.63 or smaller than -0.63, the data can be considered as an influential point. In this data set, we can conclude that the 9<sup>th</sup>, 44<sup>th</sup>, 65<sup>th</sup>, 75<sup>th</sup>, 76<sup>th</sup>, 87<sup>th</sup> are influential points.

### COVRATIO

If  $COVRATIO_i > 1 + 3p/n$  or  $COVRATIO_i < 1 - 3p/n$ , then the  $i^{th}$  point is considered influential. In the data set, if the COVRATIO value is not between 1.3 and 0.7, the data point can be defined as an influential point. Finally, it can be concluded that 9<sup>th</sup>, 13<sup>th</sup>, 51<sup>st</sup>, 96<sup>th</sup>, 98<sup>th</sup>, 99<sup>th</sup> are influential points

### DFBETAS

For these data sets, an observation is deemed influential if  $|DFBETAS_i| > 2/\sqrt{n}$ .

In the data set, if the absolute value of DFBETAS is larger than 0.164, the data point can be concluded as an influential point.

Table 3 Influential points for the variables

Variable	Influential Point
Intercept	9 <sup>th</sup> , 11 <sup>th</sup> , 45 <sup>th</sup> , 52 <sup>nd</sup> , 53 <sup>th</sup> , 56 <sup>th</sup>

<b>GS</b>	40 <sup>th</sup> , 45 <sup>th</sup> , 52 <sup>nd</sup> , 53 <sup>th</sup> , 66 <sup>th</sup> , 76 <sup>th</sup>
<b>THP</b>	9 <sup>th</sup> , 11 <sup>th</sup> , 32 <sup>nd</sup> , 45 <sup>th</sup> , 53 <sup>rd</sup> , 76 <sup>th</sup> , 83 <sup>rd</sup> , 88 <sup>th</sup> ,
<b>FTA</b>	32 <sup>nd</sup> , 45 <sup>th</sup> , 56 <sup>th</sup> , 76 <sup>th</sup> , 77 <sup>th</sup> , 83 <sup>rd</sup> , 88 <sup>th</sup> , 99 <sup>th</sup>
<b>TRB</b>	3 <sup>rd</sup> , 9 <sup>th</sup> , 53 <sup>rd</sup> , 76 <sup>th</sup> , 77 <sup>th</sup> , 88 <sup>th</sup> , 99 <sup>th</sup>
<b>PFA</b>	51 <sup>st</sup> , 54 <sup>th</sup> , 76 <sup>th</sup> , 77 <sup>th</sup>
<b>BR</b>	43 <sup>rd</sup> , 51 <sup>th</sup> , 53 <sup>rd</sup> , 56 <sup>th</sup> , 57 <sup>th</sup> , 66 <sup>th</sup> , 88 <sup>th</sup>
<b>CS</b>	9 <sup>th</sup> , 10 <sup>th</sup> , 51 <sup>st</sup> , 56 <sup>th</sup> , 81 <sup>st</sup>
<b>MS</b>	56 <sup>th</sup>
<b>Others</b>	12 <sup>th</sup> , 25 <sup>th</sup> , 61 <sup>st</sup> , 80 <sup>th</sup>

---

In conclusion, the 9<sup>th</sup> point is the most likely to become the influential point. The player on the ninth point is Chandler Parsons. He is a very outstanding player, especially in the 2014 season. He even set the record for the number of 3-pointers scored in one half. NBA teams offered him a high salary package at that point in time. However, in recent years and due to an injury, his performance has worsened and deteriorated. Hence it is reasonable for this data point to become an influential point and an outlier.

### 3.3 Four Assumption Verification

**First Assumption:**  $E(\varepsilon_i) = 0$

$$H_0: E(\varepsilon_i) = 0 \quad H_1: E(\varepsilon_i) \neq 0$$

Table 4 Test for Location  $\mu_0=0$

Test for Location: $\mu_0=0$		
Test	Statistic	p Value
Student's t	0.004737	0.9962
Sign	1	0.9204
Signed Rank	-63	0.8298

From the Student's t, Sign and Signed Rank test, all the p-values are larger than 0.05, so  $H_0$  cannot be rejected. This assumption is confirmed.

**Second Assumption:  $\text{Var}(\varepsilon_i) = \sigma^2$**

$$H_0: \text{Var}(\varepsilon_i) = \sigma^2 \quad H_1: \text{Var}(\varepsilon_i) \neq \sigma^2$$

Table 5 Heteroscedasticity Test

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variable
SAL	White's Test	21.80	44	0.9980	Cross of all vars
	Breusch-Pagan	9.69	9	0.3759	1, GS, THP, FTA, TRB, PFA, BR, CS, MS, Others

Shown as the table of the Heteroscedasticity test in the appendix, the p-value is larger than 0.05. Hence we can verify this assumption.

**Third Assumption:  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$**

$$H_0: \rho = 0 \quad H_0: \rho = 0$$

$$H_1: \rho > 0 \quad H_1: \rho < 0$$

Table 6 Durbin-Watson Statistics

<b>Durbin-Watson Statistics</b>		
<b>DW</b>	<b>Pr &lt; DW</b>	<b>Pr &gt; DW</b>
2.1781	0.8118	0.1882

According to Durbin-Watson Statistics, both the p-values are larger than 0.05.

Hence there is no positive or negative autocorrelation between the variables.

**Fourth Assumption:**  $\varepsilon_i \sim N(0, \sigma^2)$

**H0: Residuals follow normally distributed**

**H1: Residuals do not follow normally distributed**

Table 7 Test for Normality

<b>Test for Normality</b>		
<b>Test</b>	<b>Statistic</b>	<b>p Value</b>
<b>Shapiro-Wilk</b>	0.977189	0.0801
<b>Kolmogorov-Smirnov</b>	0.052831	>0.1500
<b>Cramer-von Mises</b>	0.041146	>0.2500
<b>Anderson_Darling</b>	0.335538	>0.2500

In order to verify that the residuals follow a normal distribution, four tests are used; Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling. As the table in the appendix shows, the four p-values are

larger than 0.05. Hence  $H_0$  cannot be rejected and the assumption can be verified.

In conclusion, all four assumptions can be verified. This confirms that the model can be used to make an inference or a prediction.

## **Chapter 4**

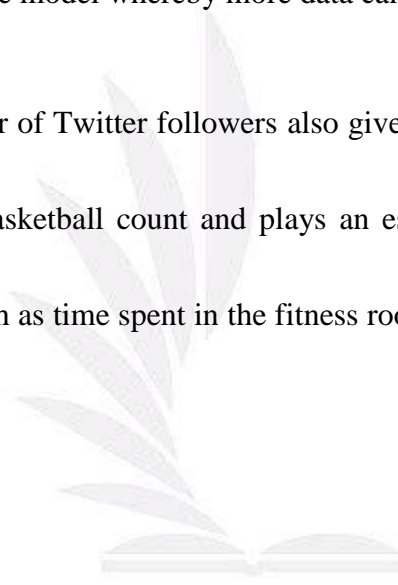
### **Findings and Discussion**

This study primarily highlights the findings through a rigorous analysis of the data set and discusses unexpected discoveries in the following step.

As the regression line is drawn, the salary is mainly related to six variables; they are Games Started (GS), 3-Point Field Goals Per Game (THP), Free Throw Attempts Per Game (FTA), Total Rebounds Per Game (TRB), Personal Fouls Per Game (PFA) and Signed Use. All the variables except for Signed Use and Personal Fouls Per Game have a positive correlation with salary, which means that a higher number for Games Started, 3-Point Field Goals Per Game, Free Throw Attempts Per Game and Total Rebounds Per Game, the higher the salary. Meanwhile, a higher number of Personal Fouls Per Game will correspond with a decrease in salary. Furthermore, Signed Use is a crucial variable in the regression model. As the model shows, players with the sign cap space receive a higher salary.

The research also considered if the players' performance on social media scored a higher R-square. Every players' Twitter account was tracked and an API provided by Twitter allowed revealed the number of followers for each player. Following that, the variable was added into the final model, leading to an increase in the R-square value to 0.7621. It can be concluded that their performance on social media actually increases the accuracy of the model whereby more data can be explained.

The variable for the number of Twitter followers also gives an idea about the players' performance outside the basketball count and plays an essential roll in determining salary levels. Variables such as time spent in the fitness room may also be collected in the future to predict salary.



## **Chapter 5**

### **Appendix**

#### **5.1 Data Resources**

1. [https://www.basketball-reference.com/leagues/NBA\\_2017\\_per\\_game.html](https://www.basketball-reference.com/leagues/NBA_2017_per_game.html) (data resource)
2. <https://www.basketball-reference.com/contracts/players.html> (data resource)
3. <https://www.basketball-reference.com/friv/twitter.html> (NBA players' Twitter accounts)





## What Are the Important Factors for NBA Player Salaries in 2017?

53	1.5453	0.9541	0.1400	0.6759	1.2323	0.1093	1.7990	0.5912	0.376	1.571	0.034	1.5841	0.1216	0.9640	0.5895	0.2353	0.2671	-0.3355	0.0534	-0.3030	-0.0865	0.2680	0.0436	-0.0618	0.0934
54	0.8000	1.2257	0.1263	0.9749	1.4786	0.3895	2.0620	-0.4257	0.381	-1.117	0.014	-1.1185	0.0989	1.0793	-0.3705	0.1504	-0.0325	0.0619	-0.0102	0.1049	-0.2398	-0.1035	0.0310	0.0168	0.0217
55	0.1313	0.3980	0.1326	0.1346	0.6615	-0.4421	1.2381	-0.2667	0.379	-0.704	0.006	-0.7018	0.1091	1.1877	-0.2455	-0.0057	-0.1062	-0.0207	0.0559	0.0007	0.0140	0.0431	0.0332	-0.1306	0.0195
56	1.3619	0.7751	0.1215	0.5337	1.0165	-0.0584	1.6086	0.5868	0.383	1.533	0.024	1.5450	0.0916	0.9447	0.4906	0.2153	-0.0957	-0.0336	0.2459	-0.1102	0.0730	-0.2727	-0.3287	-0.2494	-0.2341
57	1.0337	0.5812	0.1028	0.3769	0.7855	-0.2423	1.4046	0.4525	0.388	1.166	0.010	1.1682	0.0656	1.0278	0.3095	-0.0104	-0.0670	0.0108	-0.0482	-0.0690	0.0635	0.2061	0.0572	0.0151	0.0206
58	0.1313	0.6173	0.0906	0.4373	0.7973	-0.2005	1.4351	-0.4860	0.391	-1.242	0.008	-1.2463	0.0509	0.9910	-0.2887	-0.0553	0.1036	0.1285	-0.0012	-0.0017	0.0128	-0.0936	-0.1588	-0.0235	-0.0621
59	0.7465	0.3788	0.1010	0.1782	0.5794	-0.4438	1.2014	0.3677	0.389	0.946	0.006	0.9456	0.0632	1.0802	0.2457	-0.0492	-0.1209	0.0998	-0.0821	-0.0028	0.0676	0.0502	0.1215	0.0284	0.0013
60	0.6606	0.9456	0.1044	0.7382	1.1529	0.1213	1.7698	-0.2850	0.388	-0.735	0.004	-0.7331	0.0676	1.1292	-0.1973	-0.0196	0.1171	-0.0603	-0.0367	-0.0103	0.0389	-0.0483	-0.0955	-0.0309	-0.0119
61	0.3028	0.6859	0.1504	0.3870	0.9848	-0.1659	1.5378	-0.3831	0.372	-1.029	0.017	-1.0294	0.1404	1.1556	-0.4160	-0.0126	-0.0675	-0.0782	0.0657	0.0293	0.0181	0.0193	0.0068	0.0138	-0.2820
62	1.1000	0.9526	0.1176	0.7190	1.1862	0.1214	1.7839	0.1474	0.384	0.384	0.001	0.3820	0.0857	1.2032	0.1170	-0.0055	0.0701	-0.0479	-0.0278	-0.0502	0.0416	-0.0125	0.0180	-0.0187	0.0015
63	1.2000	1.4824	0.0991	1.2856	1.6792	0.6607	2.3040	-0.2824	0.389	-0.726	0.003	-0.7238	0.0609	1.1228	-0.1843	0.0683	0.0755	-0.0483	-0.0939	0.0573	-0.0808	0.0016	-0.0540	-0.0169	0.0154
64	1.6400	1.5247	0.1330	1.2605	1.7889	0.6844	2.3650	0.1153	0.379	0.304	0.001	0.3028	0.1097	1.2431	0.1063	0.0204	0.0267	-0.0431	-0.0142	0.0398	-0.0303	0.0329	-0.0106	-0.0067	0.0063
65	0.2117	0.4782	0.1517	0.1768	0.7796	-0.3746	1.3309	-0.2865	0.372	-0.717	0.009	-0.7148	0.1428	1.2318	-0.2917	-0.0638	-0.1854	0.0628	0.0253	0.0667	0.0456	0.0411	0.0351	-0.1183	0.0017
66	0.3533	1.1796	0.1157	0.9497	1.4095	0.3494	2.0098	-0.8263	0.385	-2.149	0.042	-2.1939	0.0831	0.7202	-0.6603	0.0463	0.3917	-0.1300	-0.1826	-0.0543	-0.0003	-0.3915	-0.0937	-0.0965	-0.0199
67	0.3904	0.8637	0.1096	0.6459	1.0815	0.0368	1.6906	-0.4733	0.386	-1.225	0.012	-1.2287	0.0745	1.0212	-0.3487	-0.0970	0.1278	-0.1400	0.0442	-0.0856	0.1686	-0.0718	-0.1544	-0.0388	-0.0176
68	0.1313	0.1602	0.1179	-0.0739	0.3944	-0.6711	0.9916	-0.0290	0.384	-0.075	0.000	-0.0751	0.0861	1.2228	-0.0231	-0.0017	-0.0025	-0.0014	0.0028	-0.0015	0.0028	0.0006	0.0004	-0.0159	0.0003
69	0.2134	0.4626	0.1359	0.1925	0.7326	-0.3796	1.3048	-0.2492	0.378	-0.660	0.006	-0.6575	0.1146	1.2032	-0.2366	-0.0086	0.0490	-0.0951	0.0412	-0.0563	-0.0873	0.1521	0.1632	0.1079	0.1369
70	0.1577	0.004368	0.1363	-0.2664	0.2751	-0.8381	0.8468	0.1534	0.378	0.406	0.002	0.4041	0.1152	1.2408	0.1458	-0.0180	-0.0614	-0.0123	0.0129	0.0342	0.0146	0.0157	0.0101	0.1033	0.0067
71	0.0500	0.1915	0.1073	-0.0218	0.4047	-0.6343	1.0172	-0.1415	0.387	-0.366	0.001	-0.3638	0.0714	1.1865	-0.1009	-0.0791	0.0015	0.0090	-0.0001	0.0047	0.0156	0.0549	0.0685	0.0622	0.0476
72	0.6021	0.3539	0.1092	0.1370	0.5708	-0.4728	1.1805	0.2483	0.386	0.642	0.003	0.6404	0.0739	1.1532	0.1809	0.1208	-0.0282	0.0071	-0.0092	0.0486	-0.0331	-0.1056	-0.1298	-0.1038	-0.0920
73	2.3500	2.5389	0.1590	2.2230	2.8549	1.6809	3.3970	-0.1889	0.369	-0.512	0.005	-0.5103	0.1569	1.2881	-0.2201	0.0202	0.0163	0.0480	-0.0663	-0.1201	0.0325	0.0375	0.0072	-0.0024	0.0091
74	0.4088	0.3853	0.1336	0.1199	0.6507	-0.4554	1.2260	0.0234	0.379	0.062	0.000	0.0616	0.1107	1.2569	0.0217	-0.0004	-0.0110	-0.0065	-0.0012	0.0028	0.0031	0.0141	0.0047	0.0024	0.0034
75	0.1524	0.1881	0.1452	-0.1003	0.4765	-0.6602	1.0363	-0.0356	0.374	-0.095	0.000	-0.0947	0.1307	1.2850	-0.0367	-0.0355	-0.0074	0.0021	0.0049	-0.0030	0.0231	0.0134	0.0175	0.0179	0.0109
76	0.7421	1.4278	0.1470	1.1357	1.7199	0.5783	2.2773	-0.6857	0.374	-1.835	0.052	-1.8600	0.1341	0.8819	-0.7319	-0.0782	-0.2031	-0.4145	0.3832	-0.3007	0.2936	-0.1227	0.0803	0.0357	0.0988
77	3.3286	2.7771	0.2093	2.3613	3.1928	1.8775	3.6766	0.5515	0.343	1.609	0.097	1.6239	0.2716	1.1465	0.9917	0.0762	0.0680	-0.0318	0.4692	0.2215	-0.3135	-0.1695	-0.1819	-0.0081	0.4565
78	1.0497	0.9462	0.0704	0.8063	1.0861	0.1363	1.7561	0.1035	0.395	0.262	0.000	0.2604	0.0308	1.1449	0.0464	0.0027	-0.0074	0.0029	-0.0170	0.0171	-0.0068	0.0043	0.0217	0.0018	0.0018
79	0.8393	0.8246	0.1024	0.6211	1.0280	0.001320	1.6479	0.0147	0.388	0.038	0.000	0.0377	0.0650	1.1958	0.0099	0.0005	0.0045	-0.0045	-0.0039	-0.0011	0.0016	-0.0003	0.0023	-0.0012	0.0006
80	0.7590	0.4039	0.1534	0.0991	0.7088	-0.4501	1.2579	0.3551	0.371	0.957	0.016	0.9565	0.1460	1.1821	0.3956	-0.0464	-0.0722	-0.0096	-0.0622	0.0744	0.0490	0.0182	0.0134	0.0141	0.2856
81	0.1471	0.8210	0.0855	0.6512	0.9908	0.005402	1.6366	-0.6739	0.392	-1.718	0.014	-1.7367	0.0453	0.8393	-0.3783	-0.0230	0.1046	0.1031	-0.1178	0.1932	-0.0807	-0.0986	-0.2072	-0.0238	-0.0546
82	2.9728	2.5799	0.1254	2.3308	2.8289	1.7442	3.4156	0.3929	0.381	1.030	0.011	1.0304	0.0975	1.1004	0.3386	0.0186	0.0513	0.0834	0.1829	0.0038	-0.1026	-0.0771	0.0064	-0.0069	-0.0433
83	2.8531	2.2724	0.1286	2.0169	2.5278	1.4347	3.1100	0.5807	0.380	1.527	0.027	1.5381	0.1025	0.9685	0.5199	-0.0652	0.0531	0.2189	0.2427	-0.1285	0.0029	-0.1113	0.0305	-0.0069	-0.0868
84	1.0163	0.9772	0.1082	0.7623	1.1921	0.1510	1.8034	0.0391	0.387	0.101	0.000	0.1006	0.0726	1.2043	0.0281	-0.0087	0.0106	-0.0026	-0.0060	-0.0126	0.0171	-0.0039	0.0048	-0.0031	-0.0020
85	0.2329	-0.0503	0.1212	-0.2912	0.1906	-0.8836	0.7830	0.2832	0.383	0.740	0.005	0.7379	0.0912	1.1576	0.2337	-0.0150	-0.0663	0.0220	-0.0014	0.0234	0.0100	0.0203	0.0177	0.1775	0.0037
86	1.4041	1.5867	0.1059	1.3764	1.7970	0.7617	2.4116	-0.1826	0.387	-0.471	0.002	-0.4693	0.0695	1.1724	-0.1282	0.0272	-0.0142	0.0713	-0.0439	0.0205	-0.0505	0.0124	-0.0179	0.0064	-0.0028
87	1.7000	1.6709	0.1248	1.4229	1.9189	0.8355	2.5063	0.0291	0.382	0.076	0.000	0.0758	0.0966	1.2370	0.0248	-0.0097	0.0027	0.0043	0.0094	-0.0149	0.0145	-0.0043	0.0028	-0.0009	-0.0035
88	1.8064	2.4692	0.1682	2.1350	2.8035	1.6043	3.3341	-0.6628	0.385	-1.818	0.070	-1.8420	0.1756	0.9330	-0.8500	-0.0867	-0.1252	0.3336	-0.6610	0.3855	0.0041	-0.2027	0.0659	0.0290	-0.0313
89	0.5192	0.7957	0.1766	0.4449	1.1465	-0.0757	1.6672	-0.2765	0.361	-0.767	0.014	-0.7650	0.1934	1.2984	-0.3746	0.0649	-0.0809	-0.1453	0.0768	0.0603	-0.0692	0.0552	0.0295	0.0189	-0.1804
90	0.2117	0.4492	0.1132	0.2243	0.6741	-0.3796	1.2780	-0.2375	0.385	-0.616	0.003	-0.6144	0.0795	1.1645	-0.1806	-0.1379	-0.0112	0.0629	-0.0254	-0.0012	0.0290	0.1000	0.1271	0.1097	0.0761
91	0.5250	0.9320	0.0757	0.7816	1.0823	0.1202	1.7437	-0.4070	0.394	-1.032	0.004	-1.0324	0.0355	1.0293	-0.1982	-0.0678	-0.0479	0.0132	0.0869	-0.0357	0.0785	-0.0143	-0.0844	0.0093	-0.0149
92	0.1471	0.1988	0.1574	-0.1140	0.5116	-0.6581	1.0557	-0.0517	0.369	-0.140	0.000	-0.1391	0.1537	1.3185	-0.0593	-0.0028	0.0017	0.0199	-0.0041	0.0171	-0.0100	-0.0082	-0.0065	0.0000	-0.0472
93	0.1577	0.2669	0.1139	0.0406	0.4932	-0.5623	1.0961	-0.1092	0.385	-0.284	0.001	-0.2821	0.0805	1.2053	-0.0835	-0.0677	-0.0267	0.0217	0.0052	0.0142	0.0126	0.0490	0.0589	0.0553	0.0369
94	0.3025	0.3579	0.1070	0.1453	0.5705	-0.4676	1.1835	-0.0554	0.387	-0.143	0.000	-0.1424	0.0710	1.2010	-0.0394	-0.0295	0.0002	-0.0029	0.0004	-0.0015	0.0079	0.0239	0.0286	0.0243	0.0207
95	0.4736	0.3427	0.1469	0.0509	0.6344	-0.5068	1.1921	0.1309	0.374	0.350	0.002	0.3487	0.1338	1.2734	0.1370	0.0062	0.0239	0.0357	-0.0234	-0.0333	0.0727	-0.0889	-0.0908	-0.0702	-0.0745
96	2.3776	2.3221	0.1726	1.9793	2.6649																				

# What Are the Important Factors for NBA Player Salaries in 2017?

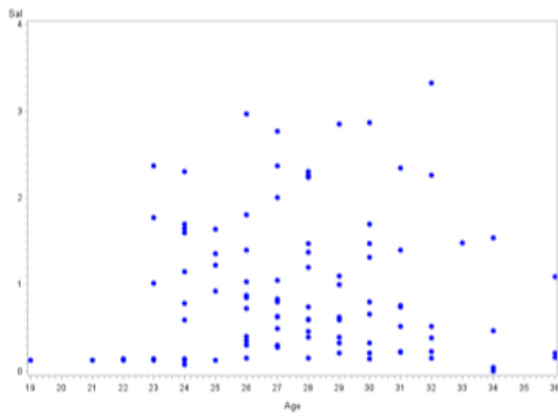


figure 1

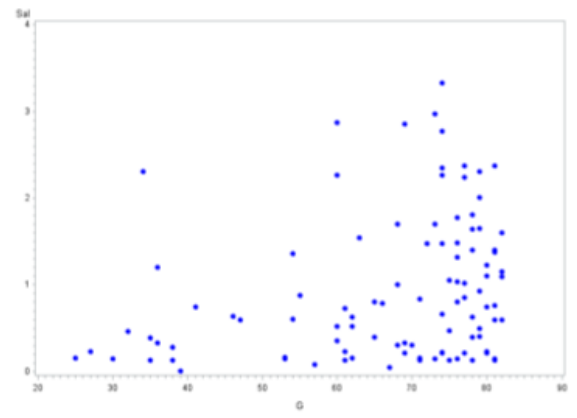


figure 2

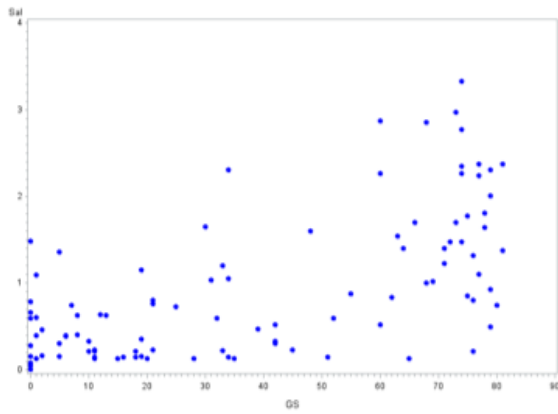


figure 3

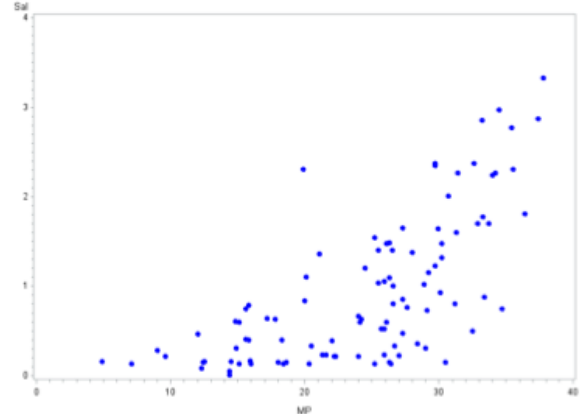


figure 4

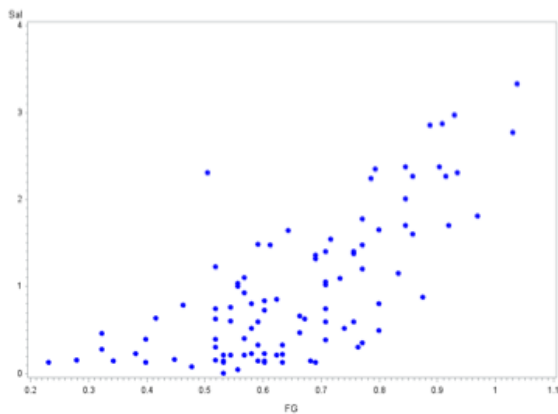


figure 5

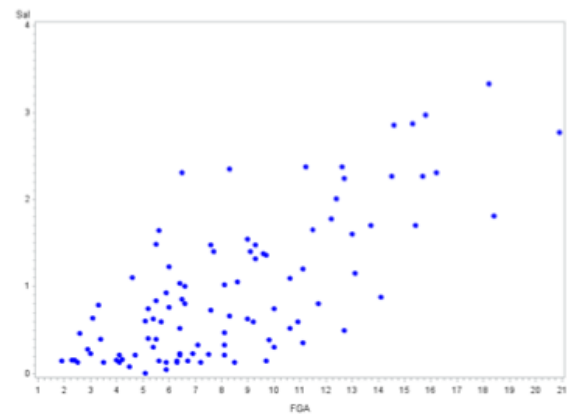


figure 6

# What Are the Important Factors for NBA Player Salaries in 2017?

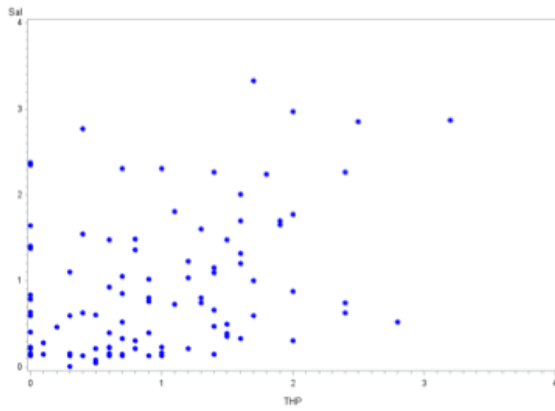


figure 7

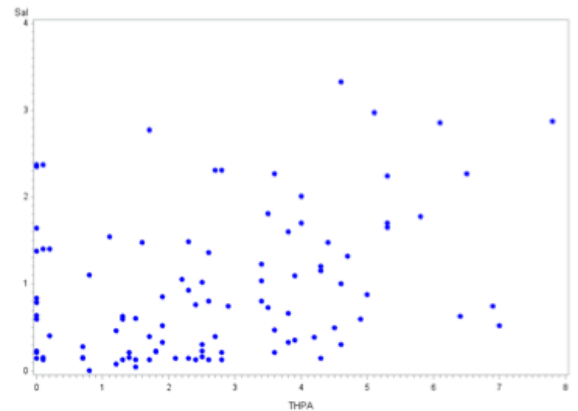


figure 8

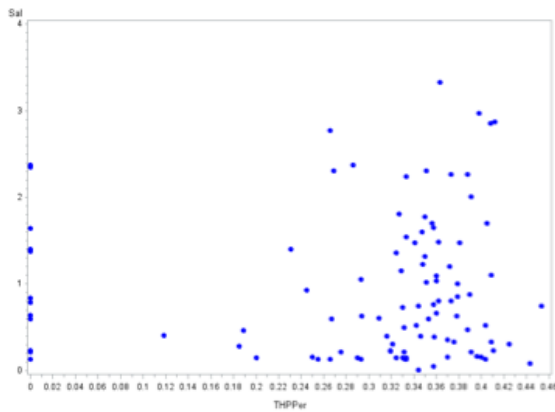


figure 9

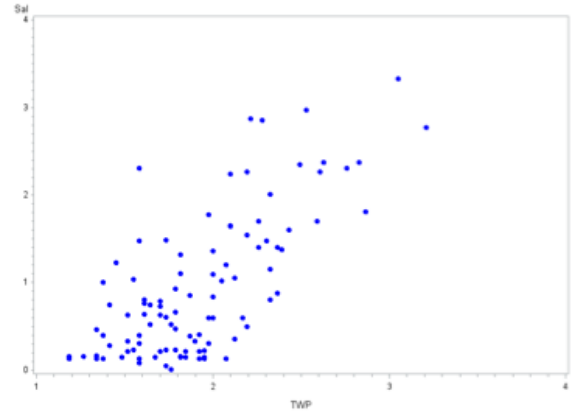


figure 10

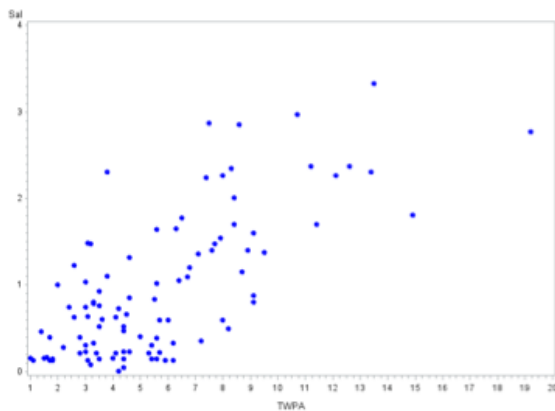


figure 11

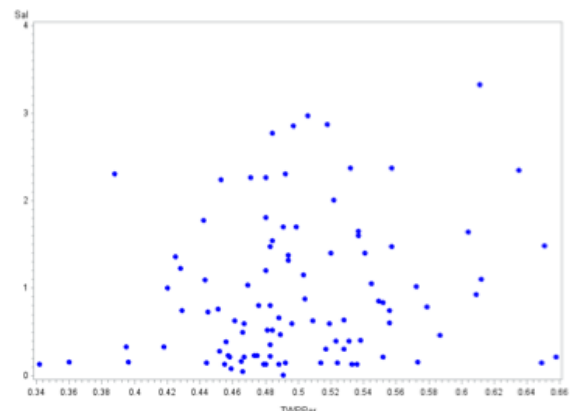


figure 12

What Are the Important Factors for NBA Player Salaries in 2017?

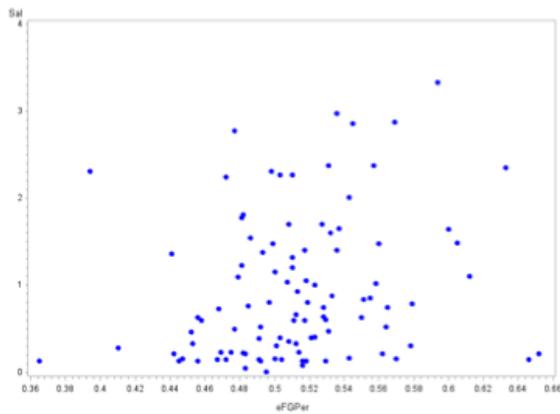


figure 13

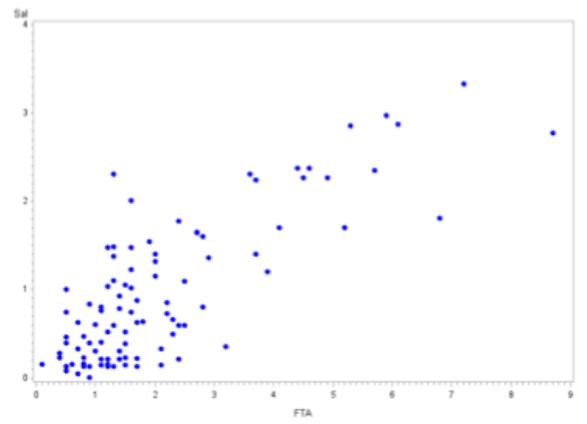


figure 14

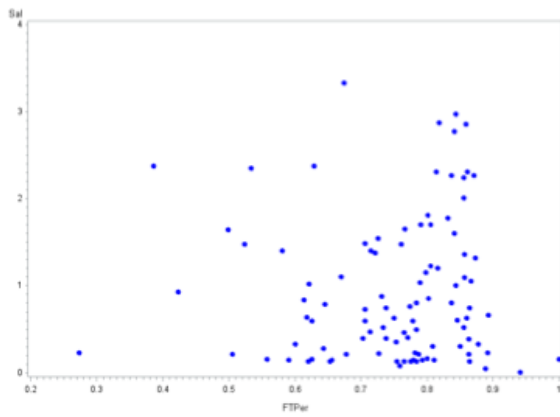


figure 15

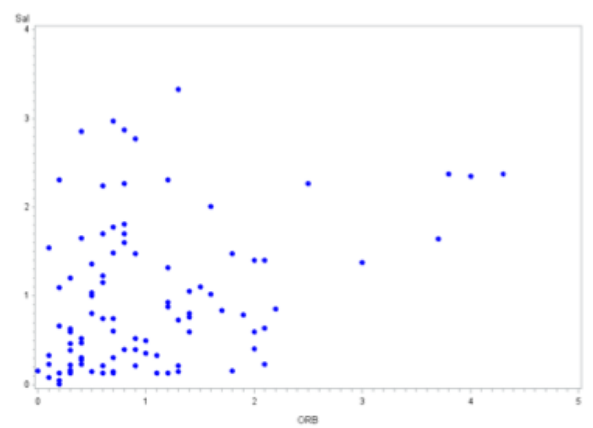


figure 16

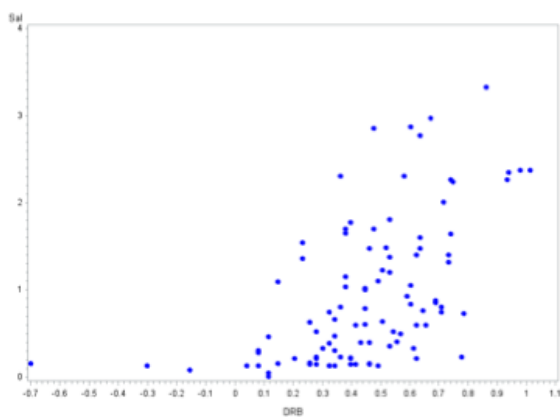


figure 17

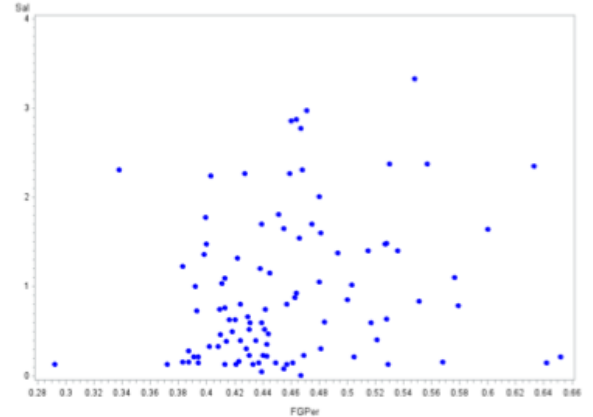


figure 18

What Are the Important Factors for NBA Player Salaries in 2017?

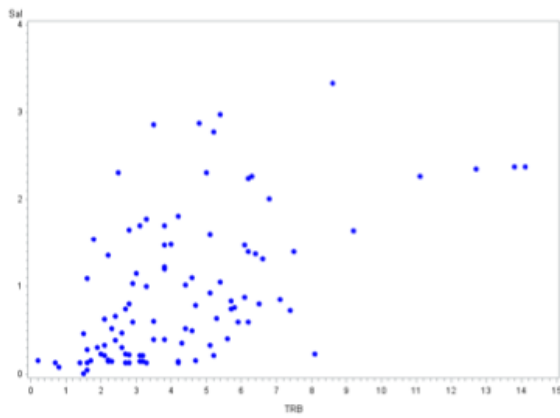


figure 19

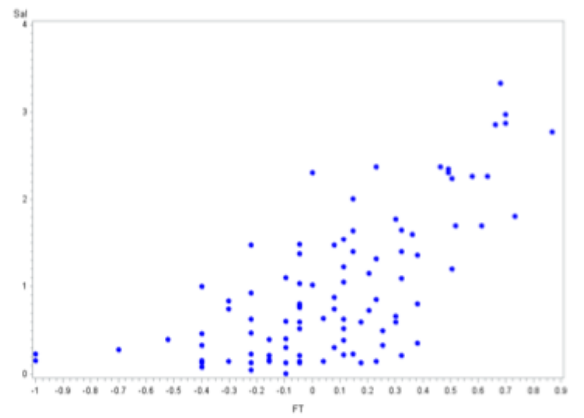


figure 20

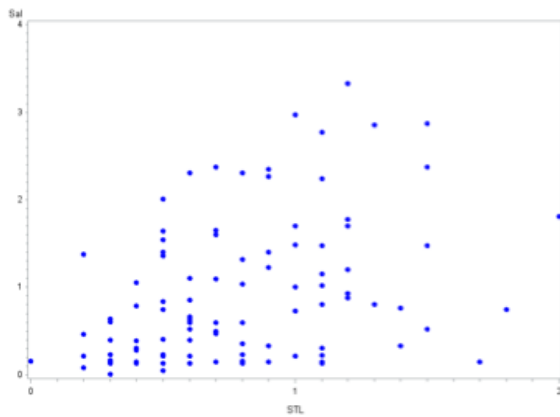


figure 21

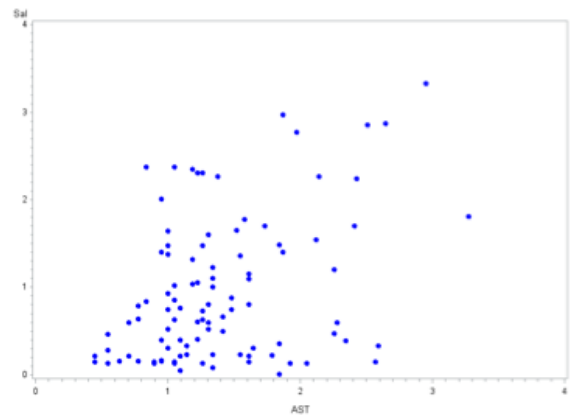


figure 22

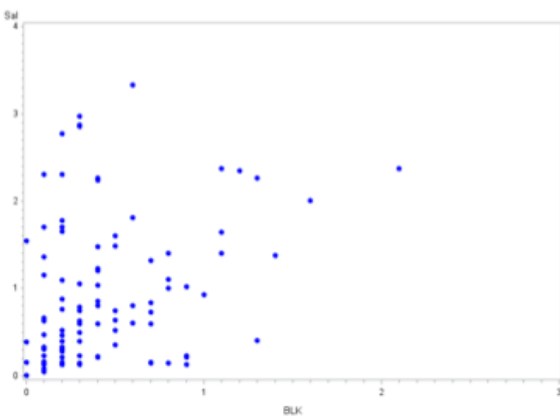


figure 23

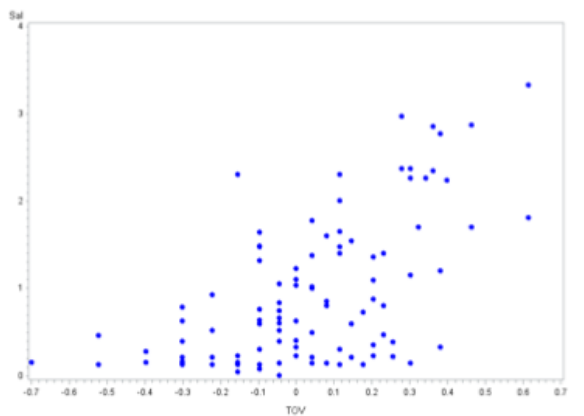


figure 24

What Are the Important Factors for NBA Player Salaries in 2017?

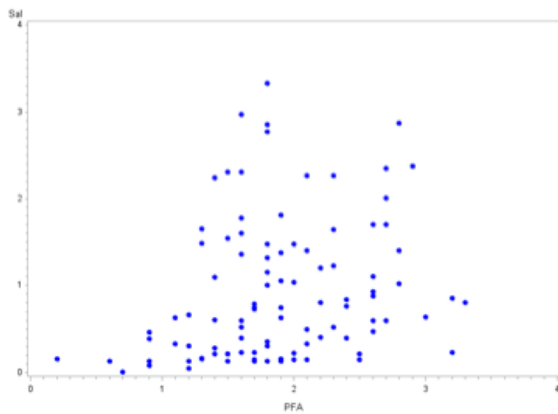


figure 25

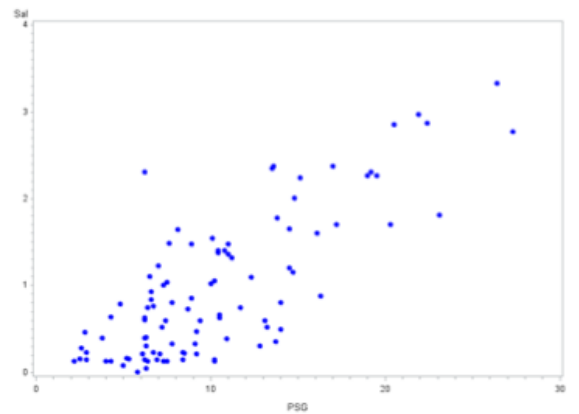


figure 26