University of California, Berkeley

Performance and Salaries of NBA Players

BY

CHO FUNG CHAN
JUSTIN LAU
JIMMY CHAN
ANTHONY DELA PAZ

FOR
STAT 133 FINAL PROJECTS
FALL 2016
NOVEMBER 30, 2016

Abstract

This project mainly analyze the data about the basketball players from The National Basketball Association (NBA) League in 2015-1016 seasons and the relationship between the salary and performances of NBA players. However, it is essential to understand the efficiency statistics (EFF) to the player in term of his salary and the potential overpayment and underpayment of the players. The data reveals that the players are generally overpaid. However, the project showed that there is correlation coefficient between the on-court contribution of each player and players' salaries.

I. Introduction

The National Basketball Association (NBA) is the top basketball league around the world. From 1940s, the popularity of basketball has risen worldwide. However, the first introduction of salary ceiling is in 1946 season. Since the approach is very restrictive and insufficient, so the policy was abolished in 1947. The first collective bargaining agreement was establish in 1970, the purpose is to set the maximum and minimum salary of the NBA players. The collective bargaining agreement is a legal contract but the salary limitation was regulated around 1985, statically, each team possesses $3.8 million payroll as their entire expenses at the cost of roster. Before the establishment of collective bargaining agreement (CBA), teams had unlimited amount to spend on the roster, and it would disproportionately affect the teams who had less capital so as they are incapable of affording the most talented players in the league. Therefore, the league decided to conduct an audit that would determine the next salary cap in order to create a fair treatment and a competitive balance in the NBA teams.

In order to avoid the rich teams with high monopoly star players, NBA provides each team's salary cap. Therefore, the economic measures are salary distribution, salary ceiling, luxury tax and so on; while the non-economic measures include the procedures for the NBA draft, rules for trades, making regulations and changing rules.

This assignment is to analyze whether the NBA players are reasonable paid in terms of their contribution on their teams. So we will mainly analyze different approaches in the evaluation of the performance of players as to determine if they're receiving a reasonable salary (whether it is underestimated or overestimated). An interesting insight from these studies is that there is a positive correlation between how many point a player gets per game for his team and how much salary that he earns (our statistics revealed that the correlation coefficient is approximately 0.6). This hugely determines the player's value in the NBA league. The other approach is that the offensive and defensive statistics of a player in the team, as our table shows that there is also a positive correlation between the salaries of players and the defensive statistics of the player. Furthermore, we also analyze the on-court contribution and salaries of players.

This project also explores as below. Section II we will analyze the collective data of the NBA player from their salaries, contribution on-court, and statistics. Section III we will explore the methodology, which included Data Acquisition, Data Cleaning, EFF, Shiny app, etc. More importantly, we will describe our data and variable we used in our table. Section IV we will explain results from the collective statistics, and will provide the result of our analysis and the explanation for our results. Section V we will discuss the conclusion, along with our studies and researches.

II. Data

There are 30 different teams throughout the NBA league over the course of NBA season 2015-2016. The primary data sets for our project are mainly from “Basketball Reference”. It provides informative statistics and those data includes each player’s on-court contribution. For instance, points, assists, steal, block, miss field goal, etc and we primarily worked on the “Roster, Totals, Salaries”.

Our first data set contains 471 NBA players and their statistics about efficiency, scores, assists, steals, blocks, offensive, defensive and total rebounds, missed field goal, missed field throw, turnover and game played from 2015-2016 NBA season. We gathered the information from “Basketball Reference” for each individual player. We also used ESPN.com to gather the statistic of the players and we are able to crosscheck that if our salary data set for each individual player is corrected.

Basketball Reference.com is a reliable and informative resource that enables us to collect most of the important statistics for each player.

III. Methodology

In this project, we use all the concepts, skills, and knowledge covered in this course, which included the data acquisition, data cleaning, data analysis. So our methodology starts with operational function and data evaluation, which enable us to identify the performance of players by collecting analysis data for all components involved.

On the Data Acquisition, we have 30 different teams in the NBA league. We scrape the three types of tables which included Roster, Totals, and Salaries for each team, which turned out to be 90 raw data files. We then use the data/rawdata/ directory to store the raw data files, separated in folders roster-data/, stat-data/, and salary-data/ by writing a function in R to scrape.

On the Data Cleaning, we created one CSV file, roster-salary-stats.csv, so that it would store the data in the directory data/cleandata/. This file included all variables from Roster, totals and salary. Additionally, we wrote a data cleaning function in R to produce the "roster-salary-stats.csv" that contains all variables from "Roster", "Totals", and "Salary" and remove the useless contents. Then, we stored the "roster-salary-stats.csv" in "cleandata" folder. After filtering the data, every data analysis project should involve an exploratory phase which known as Exploratory Data Analysis (EDA). On the Exploratory Data Analysis, We wrote an exploratory data analysis function "eda" to calculate descriptive summaries for quantitative variables or frequencies for qualitative variables. We then generate a dedicated R script eda-script R for this analysis. After this, we stored those results to a text file "eda-output.txt" in the folder "cleandata", and used a shiny app that displays a horizontal bar chart. We can compute statistics for each variable and sink() those results to a text file eda-output.txt inside the folder /data/cleandata. We also created histograms, boxplots, and bar-charts for each variable.

On the performance of players, we used the Principal Components Analysis (PCA) to compute a weight for each term in the original EFF formula and created a new table "eff-salary-stats.csv" in the directory "cleandata". We used shiny app to visualize the relationship between all the statistics of players and the salary as to display a scatterplot in which the x-axis corresponds to one statistic, and the y-axis corresponds to salary.

Salaries statistics per team plays an important role in determining on-count contribution of a player. We need to analyze salaries aggregated by each teams. We have to create a table-salaries.csv which containing the summary of salary's statistics for each team. Those statistics included "total payroll, minimum salary, maximum salary, first quartile salary, median salary, third quartile salary, average salary, interquartile range, standard deviation". After that, we stored the table- salaries.csv in the folder data/cleandata/. The next step is to use the shiny app "team-salaries": to visualize the different statistics, which enable us to display a horizontal bar-chart (one bar for each team). We then use one widget to choose the desired statistic, and another widget to

identify if the bars should be appeared in increasing order or decreasing order. The horizontal bar-charts we used in the report are arranged by decreasing order.

The EFF is an important evaluation in determining the value of NBA players and it examines by a formula which calculates what is known as “efficiency statistic”. This computes the on-court contribution as an index that takes into account some other valuable statistics for instance, “points, rebounds, assists, steals, blocks, turnovers, and shot attempts per game”. The formula is represented as follow: EFF = (PTS + REB + AST + STL + BLK - Missed FG - Missed FT - TO) / GP. However, the defensive contribution in a game is difficult to be qualified with tabulated statistics. So we say that the EFF does not take into consideration of the different position of the players.

In order to avoid the drawback of EFF, we created different efficiency which taking the player’s position into consideration. In calculation of these indices, we use the Principle Components Analysis (PCA) which enable us to weight for each team in the league. The EFF* formula is: EFF *= w1(x1)/sd1 + ……+ w8 (x8)/sd8. We then wrote the function that collects the information in “FG” and “FGA” to compute “Missed FG” and likewise, we wrote the function that collects the information in “FT” and “FTA” to compute “Missed FT”. To calculate the separate EFF indices for each position, we creates the subgroup of statistics of player dataset by positions which is a subgroup for, PG, SG, SF, PF, C. we used the function prcomp() to compute the PCA. The next step is to perform a PCA on each subgroup by using the scaled data. So we took the weights of the “first principle component as the coefficients of the EFF and these weights are in the first column of rotation, from the object returned by prcomp().” We represent the weights by taking into account “the standard deviation of each variable” since the PCA weights are received by using “standardized variable”.

We generated a new table for data set “eff-stat-salary.csv” in the directory data/cleandata/. This table contain the chosen variables in order which includes: “players’ name, total points, total rebounds, assists, steals, blocks, missed field goals, missed free throws, turnover, games played, efficiency index, salary”

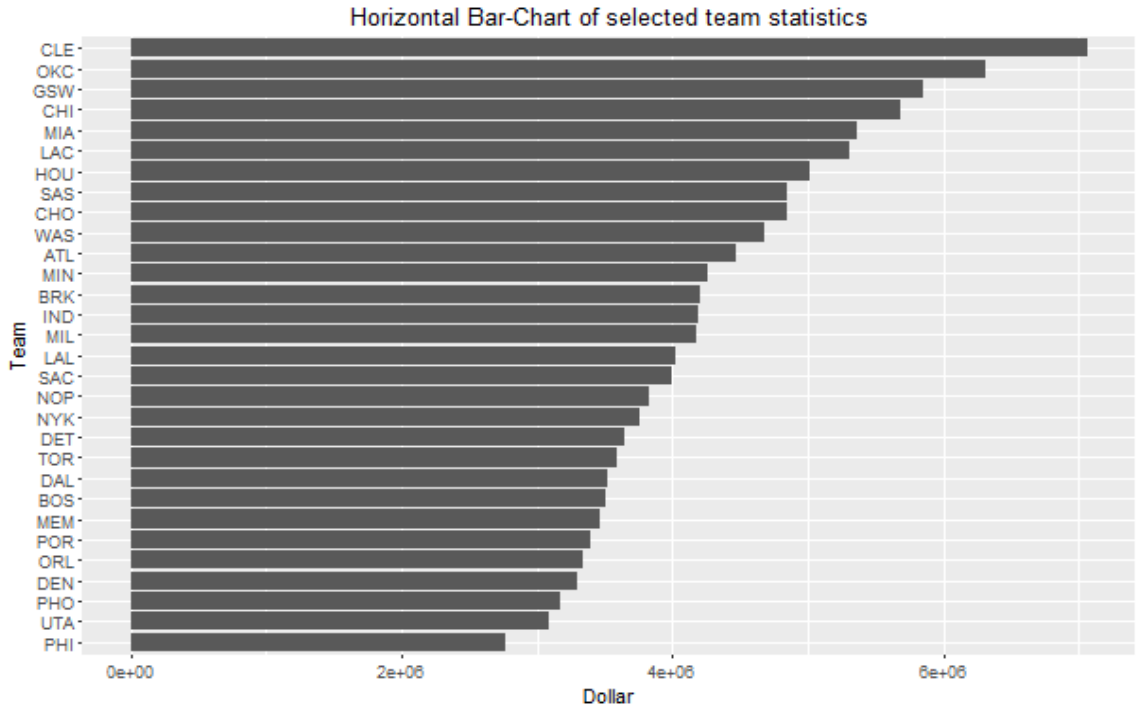
We used the shiny app (“stat salaries”), which enable us to examine the relationship between the on-court contribution (statistics included EFF) and salaries. For the section of salaries, we created a shiny app that generates a scatterplot in which the x-axis represents to one statistic, and the y-axis represents to salary. This includes a widget that allows us to choose the specific statistic for the x-axis and another widget allows us to choose a variable for the y-axis. Another widget is to determine if the dots should be colored by position. In conclusion, this app enables us to create a graphic that can identify the correlation coefficient between the chosen variable.

We also calculated the value of each NBA player as we use the formula (value = efficiency / salary). This allows us to compute the efficiency over their salaries proportion. The approach is that we calculated the ration for each NBA player by using EFF, by dividing it by the salary. We then use this value to analyze the top twenty players in the league and also the bottom twenty players. Lastly, we stores those statistics of 40

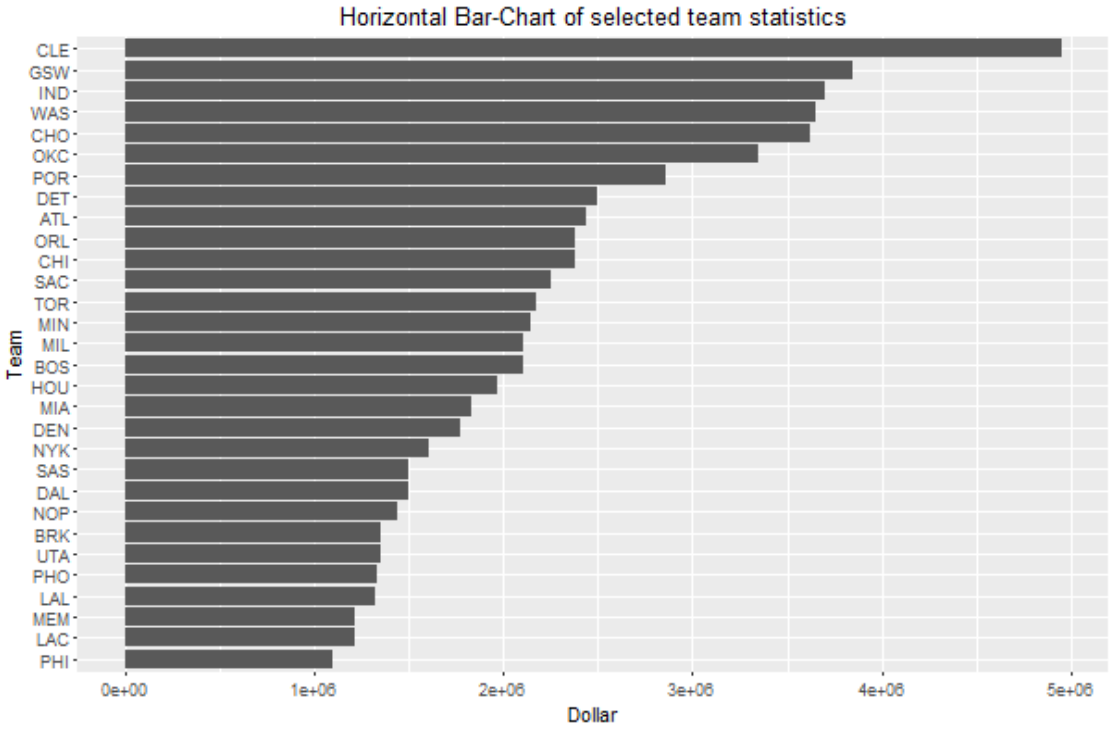
players in the text file called “best-worse-value-players.txt” insider the “data/cleandata directory.

IV. Results

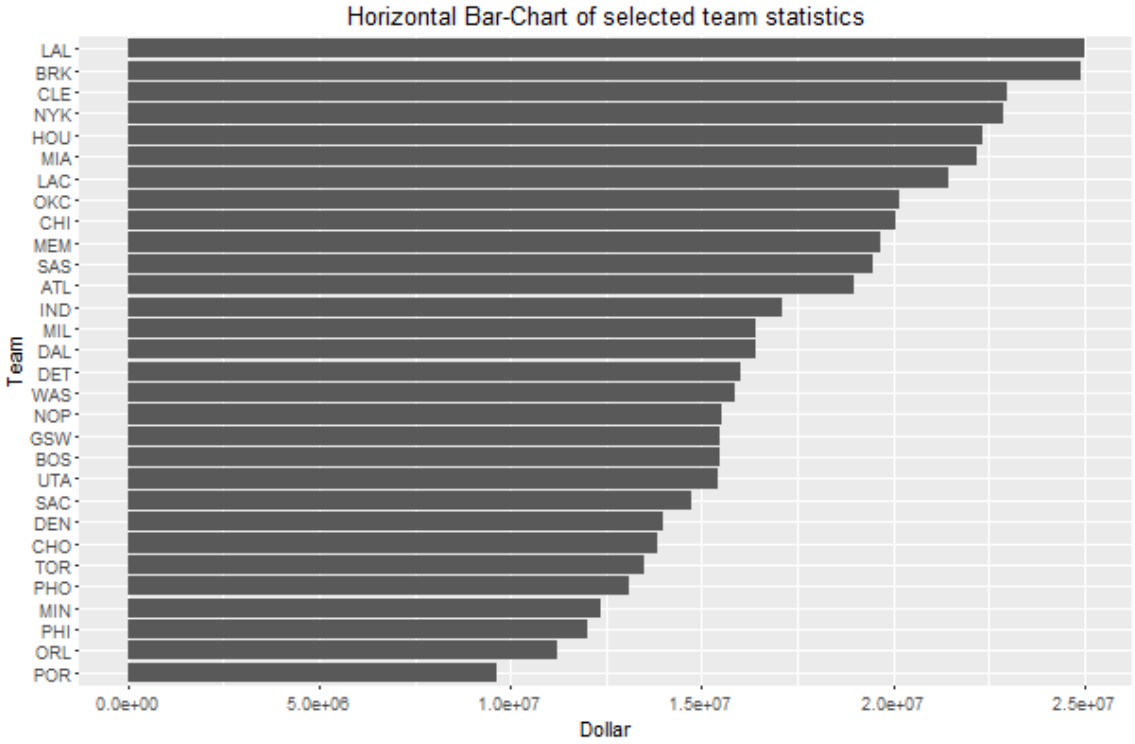
Here are the conclusions of the statistic of NBA players and teams.



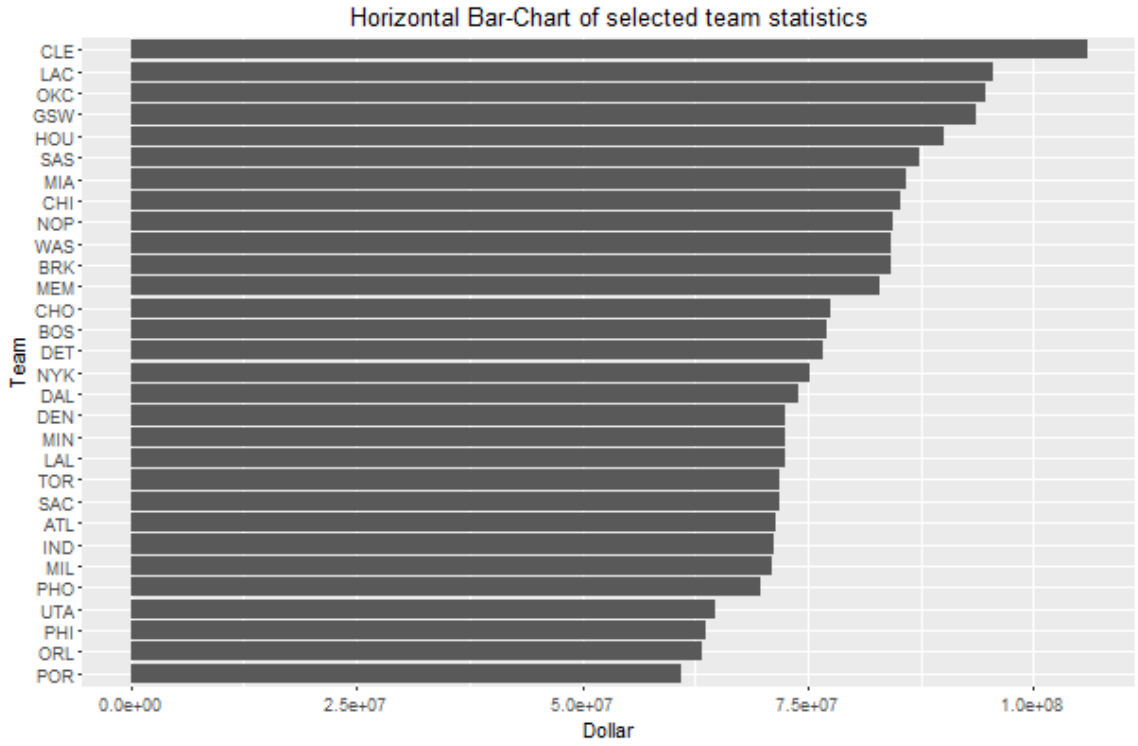
This table shows that average salary for each team. We concluded that the Cleveland has the highest average salary. Conversely, the Philadelphia has the lowest average salary. This indicates Cleveland has more capital to pay for the roster, which allow them to enhance their competiveness. As a result, Cleveland was the Champion in NBA season 2015-2016. This can be illustrated that the more capital a team possess, the higher possibility for them to hire greater player as they have higher chance to win the games.



This table shows that median salary for each team. We can see that the Cleveland also has the highest median salary. The ranking is fairly similar to the table of average salary for teams.

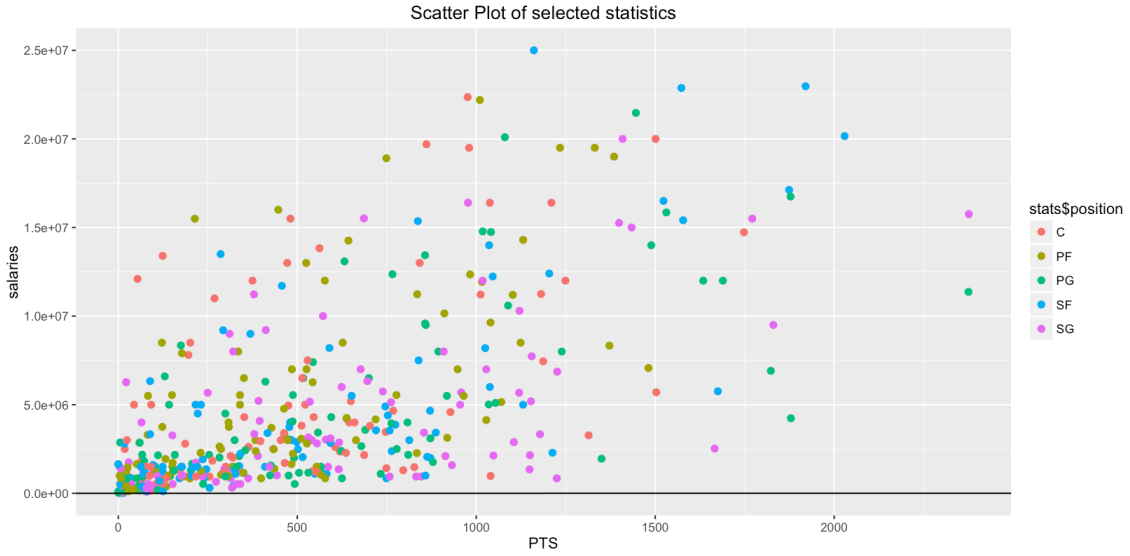


This chart shows that the maximum payroll for each team. It can be seen that Lakers has the highest single player payroll. But ironically, Lakers maintain a relatively low percentage of winning throughout NBA 2015-2016. The amount Lakers invested in that particular player is not worth.

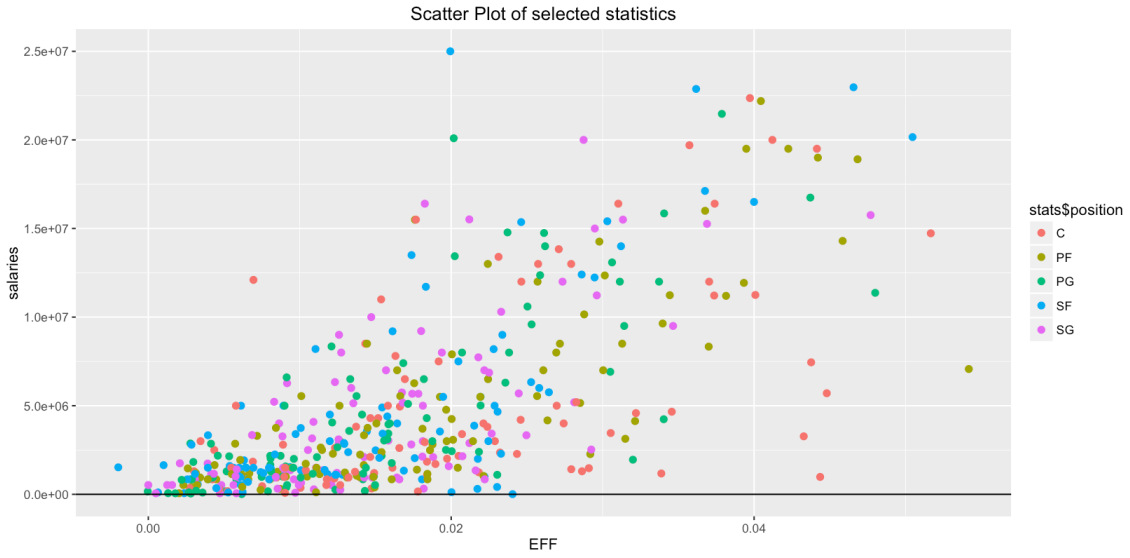


The last table displays that the total amount of salary for each team. This can be seen that Cleveland has the highest total payroll among all teams. On the other hand, Portland has the lowest total payroll and the ranking is relatively similar to the table of maximum payroll for each team. Since Portland has less capital, so they are incapable of affording superstars.

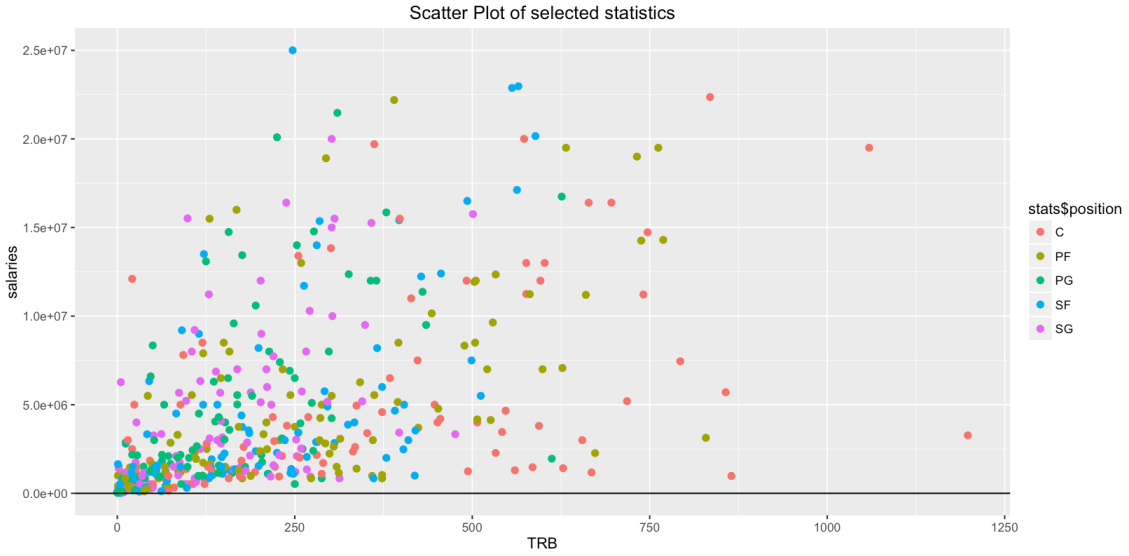
The table on the following page displays a positive correlation between the salaries of a player and the point he made per game. We applied the efficiency and salary.



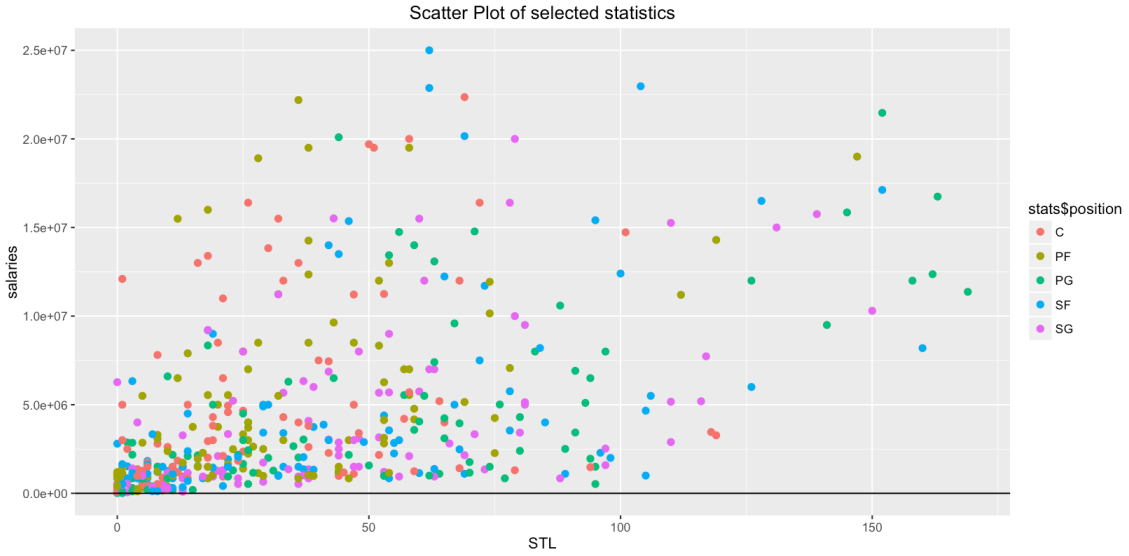
From what we have observed above, there is positive correlation between the salary and the point made by a player. The correlation is 0.6400332. We concluded that a player's revenue proportionally determined how many points he made per game.



Our second model shows that there is positive correlation between the salary and the EFF for players. The correlation is 0.703521. We also concluded that a player's revenue proportionately determined by his efficiency statistics per game.



We also analyze the relationship between the salary of a player and the total rebounds. We found that there is also a strong correlation as it appeared to be 0.5298866. This result tells us that the offensive and defensive rebound skills proportionately determined one's salary.



This table reveals the relationship between the salary of a player and the total of stealing of player. In this table, the correlation is 0.4857321. Interestingly, the graph shows that a large amount of players received low salaries and it matched with their stealing skills.

However, when we rank players according to their value, we see that the players who scored high in EFF are unseen at the top of the value chart. This simply illustrates that for the amount these star players are paid, they are underperforming.

V. Conclusion

Our project provides us an insight of the correlation coefficient between the on-court contribution of each player and their salaries. The statistics revealed in our studies showed NBA players are generally overpaid when we are comparing their on-court contribution (EFF) and their revenues. Additionally, we found that the top 20 efficient players receive a reasonable salary relative to their EFF. However, those statistics are incapable of describing the entirety of salaries of player because we also need to take into consideration of the positions of player in order to avoid the drawback of EFF (this is considered as the simplest EFF). The other interesting factor from our models illustrates that there is a strong correlation coefficient between the point scored and the salary. As a result, we can say that a player's revenue is greatly determined by his scoring power. On the other hand, we found that centers receive more offensive and defensive rebound statistically. It can be explained that centers have more control of rebound, whereas they also have a higher chance of turnover. However, from our data set, it shows that point guard has a stronger ability of stealing. These skills play an important role in determining the payroll of players. We found that a certain amount of the players are over paid. So players' practical ability has not always matched their salaries. Our studies enable us to evaluate the on court contribution of a player, and to improve our knowledge about the factors that contribute to the salaries of a player.

Work Cited

Coon, Larry. "2011 Collective Bargaining Agreement." NBA Salary Cap FAQ. N.p., n.d. Web. 29 Nov. 2016.

Hill, J Richard, and Peter A. Groothuis. “The New NBA Collective Bargaining Agreement, the Median Voter Model, and a Robin Hood Rent Redistribution.”

Journal of Sports Economics, May 2001. Web. 29 Nov. 2016. “NBA &ABA Player Directory.” Basketball-Reference.com. N.P. n.d. Web. 29 Nov. 2016.

“NBA Player Scoring Statistics.” ESPN. ESPN Internet Ventures. N.d. Web. 29 Nov. 2016.

"NBA Salary Cap History." RealGM.com, n.d. Web. 29 Nov. 2016.

“NBA Stats: Totals.” Basketball-Reference.com. Basketball-Reference. n.d. Web. 29 Nov. 2016.