How does an NBA player's social media popularity impact their on-court salary?

Introduction

NBA Players generate a large salary both on and off the court. Many of them have large following on numerous social media platforms Some also try to use their platforms to create a bigger name for themselves. The level of performance of an NBA player normally would have an impact on how high the player's salary is. While is there a relationship between popularity on social media and on-court salary? Should GMs take player's popularity into consideration when they calculate the salary? This was a group project in EXSS327 class. My team member and I collected the data together and analyze the data using SPSS. I used R to analyze the data this time. It would be interesting to use different software and statistic test analyze the same data.

Research Question

How does an NBA player's social media popularity impact their on-court salary?

Related work

Throughout sports there is ample research and literature related to the salaries of players.

"Hot Stove Economics" digs into related topics of valuing long-term contracts, evaluating minor league players, and how to value players both through production and as an asset (Bradbury, 2010). "Predicting Salaries of Major League Baseball players." looks into how salaries can be biased by both current production and previous historical production (Magel et al., 2015). “Computational Estimation of Football Player Wages” examines differing factors such as previous season performance, age, trajectory, personality etc. For their method they create an algorithm that analyzes performance, behavior, and abilities, and uses this algorithm to compare projected vs. actual salaries (Yaldo et al., 2017)

Methodology

Gather data including dependent variable which is salary, and independent variables that could represent performance and social media popularity. Build a linear regression model between salary and performance variables and another model between salary and performance plus popularity variables. Compare the two model and see if the second model is significantly better than the first one.

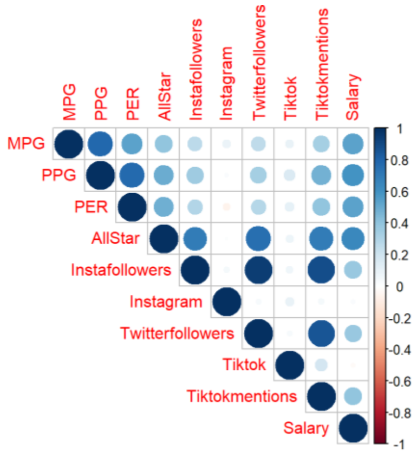
Data Collection

We selected the starting five and the six man in every NBA team from season 2021-2022 as our sample, which is 180 players in total. We collected salary and performance variables from Sports Reference. The performance variables include MPG (minutes per game), PPG (points per game), PER (player efficiency rating), and Allstar times. For social media popularity, we chose three platforms: Instagram, Twitter, and TikTok. We collected whether they have an Instagram account and followers number on Instagram and Twitter. For TikTok, since not a lot of samples have an official account, we collected the binary variable of whether they have an account and the number of mentions of player’s name. We collected popularity data by hand.

Below is some descriptive statistics for the sample.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Salary	180	925258	45780966	14817370.81	11527008.28	1.329E+14
MPG	180	22.1	37.9	30.417	3.8115	14.527
PPG	180	5.7	30.6	15.654	5.7646	33.231
PER	180	7.24	32.94	16.1959	4.64442	21.571
All-star	180	0	18	1.15	2.673	7.145
Twitter Followers	180	0	52300000	1007408.86	4492826.586	2.019E+13
Insta Social?	180	0	1	.97	.180	.032
Tiktok	180	0	1	.22	.417	.174
Insta Followers	180	0	133000000	2702440.48	10860856.71	1.180E+14
Tiktok Mentions	180	1100	5800000000	200285111.7	589633829.2	3.477E+17
Valid N (listwise)	180					



Analyzing Data

Before building the model, starting with a correlation matrix is necessary. It shows the relationship between each two variables and give me a whole picture for the data. Looking at the graph, one can see no variable has negative correlation with Salary. AllStar times has the highest correlation with Salary and Instagram account has the lowest. Since there are only six players that have Instagram account, this variable does not seem to be a strong indicator of salary. There are also three variables that have high correlations between each other: Instagram followers, Twitter followers, TikTok Mentions. This situation might lead to collinearity issue to our model. Based on this, I decide to use stepwise variable selection model building procedure to select best variables.

First, let's look at the performance variables and salary. If using alpha level of 0.01, p value for F-statistics of this model which is 2.2e-16 indicates this model provides a better fit than a model that contains no independent variables. R^2 of 0.5187 shows 51.87% variance in Salary can be explained by this model. To improve this model, I use stepwise variable selection and result in a better model that only has MPG, PPG, and Allstar with smaller AIC and a similar R^2 of 0.5177. By taking out PER, VIF all got smaller than 3 which prevents collinearity issue.

Now, let's look at popularity variables and salary before we combine the model. Model mod_popularity also has a significant p-value for the F-statistics of 4.052e-06 and R^2 is 0.1468. By using stepwise variable selection, I had a better model contains TikTok and TikTok mentions and explains 15.92% of variance in Salary.

Next, I create a combined model containing selected performance variables (MPG, PPG, Allstar) and all popularity variables. Then, I use stepwise variables selection to evaluate this model. As a result, I have mod_combined1 contains MPG, PPG, Allstar, Instagram followers, Twitter followers, TikTok. P value for F-statistics of this model which is 2.2e-16 indicates this model is statistically significant as it provides a better fit than a model that contains no independent variables. R^2 of 0.5187 shows 55.6% variance in Salary can be explained by this model. By looking at the VIF, I noticed the VIF of Instagram and Twitter followers are 10.232132 and 12.593616 which imply collinearity issue as they have high correlation between each other. So, we took out the Instagram followers as Twitter followers has a smaller p value in the model and result in mod_combined2 with similar R^2 of 0.5486.

Finally, I can compare two models: mod_performance1 and mod_combined2. As mod_combined2 contains all variables in mod_performance1, I can compare the nested model using anova test. The p-value of 0.001152 rejects the null hypothesis that mod_performance1 is better than mod_combined2.

Result and implication

By looking at the analysis, I conclude performance variables have a statistically significant relationship with Salary and adding performance variables indeed result in a better model predicting Salary. Although adding popularity result in a statistically better model, the change of R^2 is not big. Only 3.09% of additional variance is explained by adding popularity variables. Performance is a lot more important than popularity when GMs determine salary. For players, they should focus more on performance rather than playing their social media all the time.

Limitation

The sample size is limited since we don't have a lot of time collecting data for all the NBA players. We only consider on-court salary. Nowadays, players are also making money through endorsements, brand deals, social media income, while these data are difficult to acquire for now. We collected popularity data on Sep 20, 2022, while the performance data is for 2021-2022 season. It is difficult to track social media data back in that season, we could only use the data we have right now to analyze.

Discussion

There are quite a few interesting points during analyzing the data. Each variable is positively related to Salary. However, in the final model, two popularity variables (Twitter follower and TikTok account) have negative coefficients, which means when holding other variables unchanged, increasing Twitter fans or creating TikTok account make the predictive salary smaller. This seems a bit strange. It is relatively easier to see why TikTok account would have negative coefficient as many players with high salary don't have accounts and many players with lower salary have accounts. While the pattern in Twitter followers is a bit unclear. There are players like Nikola Jokic and Kawhi Leonard who have no social media appearance and very high salary. Players who have high salary, but no Twitter account would be a reason why this coefficient is negative. What's more, I also use SPSS to analyze the data using "hierarchical" model building procedure. The result shows Sig. F change for adding popularity variables is not significant. The test and procedure are different so different results might happen. It would be interesting in the future to see what cause the difference and which software provides better analysis for this question.