Data Story for Capstone Project 1: Do fans know who the best nba players are?

Story 1: Relationship between stats

I first checked if there is any relationship between the players stats. The first relationship I wanted to see was if the players scored more points per game then do they have a higher true shooting percentage. True shooting percentage is a calculation that uses two-point percentage, three-point percentage, and free throw percentage. I used a scatter plot to see if there was any relationship. According to the scatterplot, there seems like a logarithmic relationship between those two values. This could mean there is a limit on how well a player could shoot the ball. The next question I wanted to ask is what is the relationship between win-shares and real-plus-minus. Do players with higher win shares have high rpm. Again, using a scatter plot, it seems like there is a linear relationship between win shares and rpm. So if a player has high win share then he would also have a high rpm. I also wanted to see the relationship between win shares and the player efficiency rating. Using a scatter plot, I realized there is an exponential like relationship between those values. In addition, I noticed a large amount of players are stuck around a player efficiency rating of 5 and a win share of around 3. Then I wanted to see to the distribution of the players' net rating, player efficiency, and win-shares. I hypothesise that these 3 stats would have similar distribution because they all show how efficient the player was on the court. So, I made histograms and found out that while net rating and player efficiency had a normal distribution, win share was heavily skewed to the right. This could mean that few players have a huge impact on their teams or the good players play on the same team.

Story 2: Relationship how good a player is and their buzz

I wanted to see if there was any relationship on how well a player plays and their social media and fan attention. So, one of the insights that should be checked is the player efficiency rating and win shares and their twitter favorite count. My hypothesis is since the players with highest win-shares and player efficiency play are the best players then the people should notice them and favorite their tweets more often. I made scatter plots and took out outliers like Lebron James and Steph Curry who are way to markertable and have very large social media presence compared to other players. I checked the correlations to see that was true. I noticed that there was a positive,but not a strong correlation between those values. I also checked the correlation with basic stats like PS/G and the twitter favorite count. There a slightly stronger positive correlation between PS/G and twitter favorite count. This could mean players who score more points get noticed easier by fans. There is a strong possibility that other factors play into a twitter favorite count. Some possibilities could be number of tweets a player tweets and the

market size of the team the player plays for. I also ran the same test for player's wikipedia pages views. Instead of using win-shares. I used the steps and came up with the same answer. So the conclusion is that the hypothesis is wrong that the best players are the most popular players. Some data that was not in the data set that could be helpful to figure out on what makes a player popular could be number of twitter followers, other social media numbers, and number of tweets the player tweets or is about the player. I also tested if the best players or the most popular players bring people to games. The hypothesis is that people want to watch the best players or most popular players, so they watch the games that the best/popular players play in. I made bar graph which shows the average attendance and attendance percentage (filled seats/empty seats) for each team. I also made a bar graph for the team wins. I grouped twitter retweets and favorites for each team and see if there is a correlation between the number retweets/favorites tweets and attendance. There is a weak positive correlation between the teams twitter stats and attendance. This could mean other factors that plays into attendance numbers. One possibility could be the brand of the team or the fans' loyalty towards a team.

Conclusion

Basic stats like points per game which are easy stats for fans to follow leads to more popularity of the player. The popularity of the player is not a major factor in determining attendance for a team. However, there is a positive correlation between how good a player is his popularity. There could be other factors in play which this data set does not contain that could lead stronger correlations for a player's popularity. There could have been more digging in the data set to find other relationships.