The Gender Gap in Participation and Wages Among the Top Earning Athletes in the World

Abstract

Understanding the data that revolves in the sports industry is vital to many economic and social changes in the world. From sports ranging from Football to Cricket, they all play an important role in improving society in different ways, whether it is in gender equality, lessening corruption, and providing young athletes an important information to the choice of sport they would like to engage in. In this report, I have done different types of analysis, specifically on a dataset representing the top 100 earners in the sports world, to determine which sport on average is the most lucrative entertainment, the differences in the representation of gender in the top sports in the world, and finally the gap in athlete's salary in variation of gender. After completing an analysis on the top 100 most paid athletes, I find that the boxing world on average is the most profitable sport for an athlete, the representation of women in sports among the top paid athletes is at 3%, and that the top paid female athletes earn less than male counterparts in sports.

1. Introduction

In the 21st Century, the sports industry has turned into one of the most lucrative entertainment functions in the world. This industry has hundreds of billions of dollars moving around every day to keep it in working order. Most people are less interested in the operations of a sports enterprise and most interested in the contracts worth millions of dollars signed by athletes, whether they are endorsements or wages they always make the news and catch everybody's attention. This report is an in-depth analysis of the 100 most paid athletes in the world. The following report gives a study on the gaps present at the top in terms of gender and sports.

2. Data and Methods

The study follows a dataset taken from GitHub representing "The 100 Most Paid Athletes". The dataset was in good condition to analyze directly it did not require cleaning as it was a simple dataset with 100 rows of data and 12 attributes. The attributes consist of the rank, name, sport, total pay, salary winnings, endorsements, nation, gender, year of birth, birth date and place of birth. We will specifically work with 5 of the 12 attributes for our study. The sport attribute is a categorical type representing 10 different sports played by the top earning athletes and it identifies as a nominal type. The gender attribute is a categorical type representing the gender of

the athletes present and is identified as a symmetric binary type with both outcomes equally important, whether male or female the athlete's value is equal. The total pay attribute is a continuous numerical type representing the total money earned of the athlete and ranging from $17,300,300 to $105,000,000. It is constituted from the sum of the salary winnings and the endorsements attributes. The salary winnings attribute is a continuous numerical type representing the athlete's amount earned from contracts with respective sports teams and ranges from $200,000 to $105,000,000. Finally, the endorsements attribute is a continuous numerical type representing the amount earned from endorsements such as commercials, brands and so on. The endorsements attribute ranges from $0 to $55,000,000.

For the study conducted the techniques used to evaluate the dataset were statistics and visualization. I first started by examining each attribute separately, the total pay attribute had a standard deviation of about 14378127 and with that I was able to calculate the skewness which gave me a positive value, and so concluded that the total pay was positively skewed (Figure 1). The total pay had a minimum of $17,300,000, a first quartile of $20,400,000, a median of $22,650,000, a mean of $27,513,850, a third quartile of $27,325,000 and a maximum of $105,000,000. Which proves the positive skew because the mean is greater than the median.

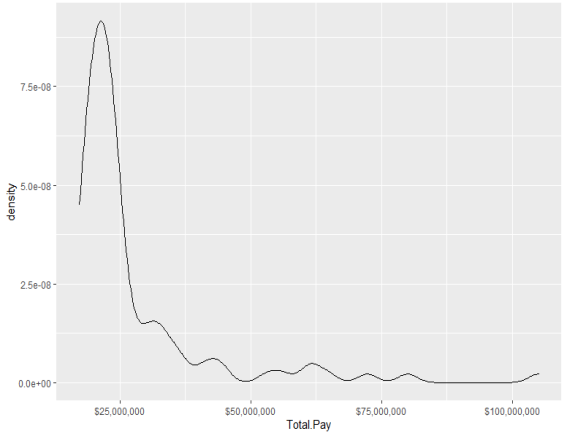

Figure 1: Total Pay Attribute Skewness

The salary winnings attribute had a standard deviation of about 11750861 and with that I was able to calculate the skewness which gave me a 0.3571653 and determined that it was symmetric (Figure 2). The salary winnings attribute had a minimum of $200,000, a first quartile of $16,500,000, a median of $18,700,000, a mean of $20,099,000, a third quartile of $21,925,000 and a maximum of $105,000,000. Which gives an approximation to the symmetrical possibility of the skew because the mean is approximately equal to the median.

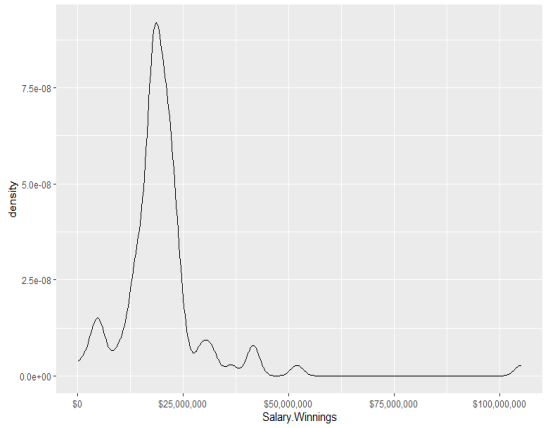

Figure 2: Salary Winnings Attribute Skewness

The endorsements attribute had a standard deviation of about 11802644 and with that I was able to calculate the skewness which gave me a positive value, and so concluded that the endorsements was positively skewed (Figure 3). The endorsements attribute had a minimum of $0, a first quartile of $400,000, a median of $2,500,000, a mean of $7,414,850, a third quartile of $8,250,000 and a maximum of $55,000,000. Which proves the positive skew because the mean is greater than the median.

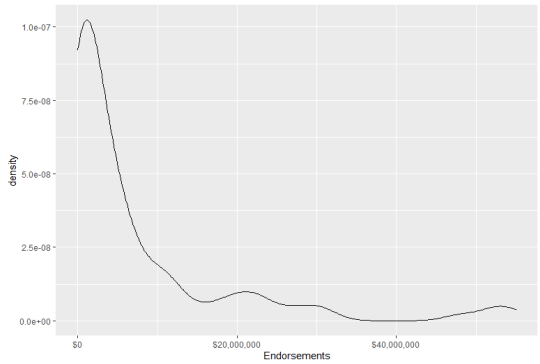

Figure 3:Endorsement Attribute Skewness

To determine the existences of certain outliers in our datasets we used the method of the z-score calculation for each of the three numerical types. Using the z-score method we proved that there were some outliers in our data, for the total pay attributes we found three outliers, $105000000, $80,000,000, and $72,300,000. For the salary winnings attribute we found one outlier, $105,00,000. And finally, for the endorsements attribute we found four outliers, $6,200,000, $53,000,000, $48,000,000, and $52,000,000.

In the results section our visualization will give an insight to the differences the gap between female and male participation in sports, the average difference in salary in each sport presented, and the salary gap between the top earning athlete's in respect to their gender. For this, we will visualize three graphs and get our insights from each one.

3. Results

We will firstly investigate the gender gap in sports, specifically between the top 100 most athletes.

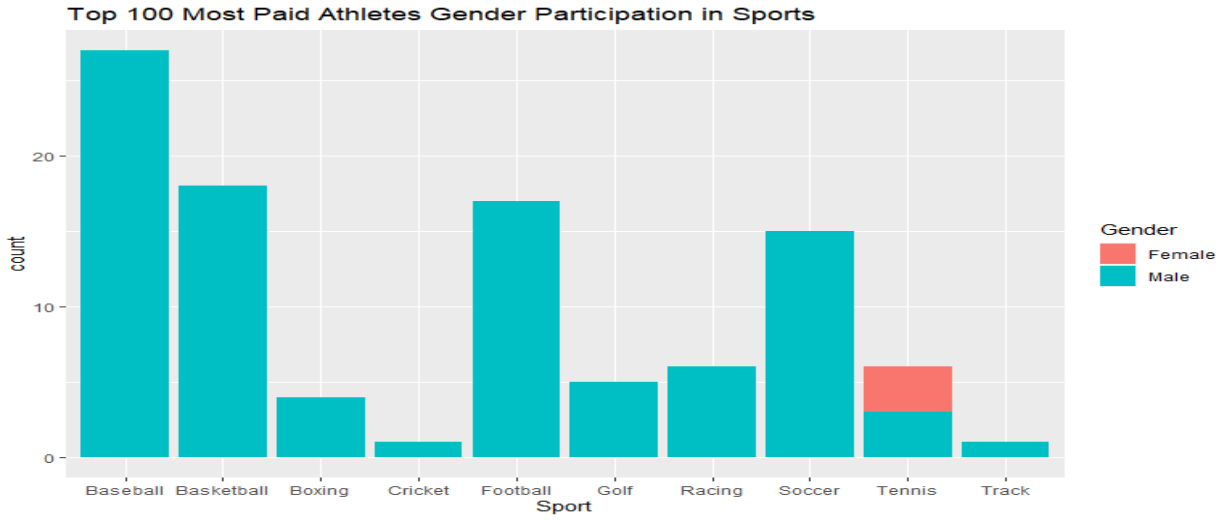

Figure 4: Sports with Respects to Gender Representation

From the following graph (Figure 4) we can visually see the great gap between male and female in the most paid athletes per sport. The number of females on list of the top 100 most paid athletes is low, and the females who made the list are categorized in one sport which is tennis. This translates to a representation of merely 5.9% of female participation in the different sport categories presented in the dataset. We can see that on the overall males constitute the majority in the 100 top earners with a representation of 97%.

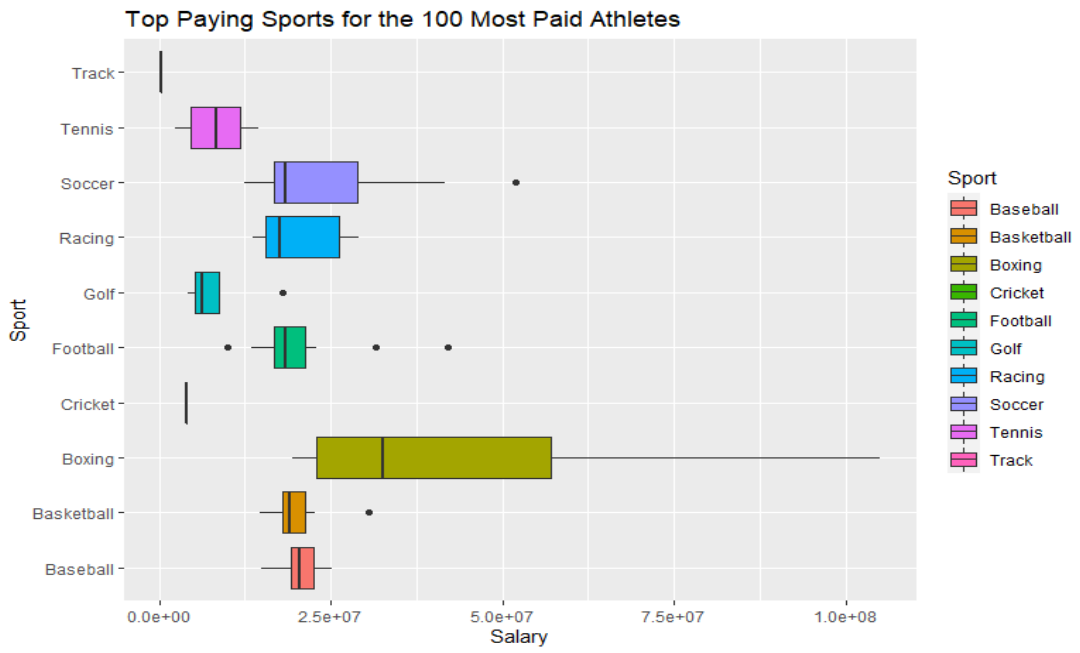

Figure 5: Boxplot Representation of Sports in Function of Salary

Figure 5 visually represents the average salaries in respects to the sports for the 100 most paid athletes. We can visually determine that the Boxing world has by far the highest average compared to the rest of the sports, but this is evidently due to an outlier. I was unable to successfully remove that specific outlier, but with different calculation I was able to determine the highest salary paying sport which is Baseball, followed by Basketball, and Soccer. As you can see on the graph the second greatest average is Baseball right after Boxing and so my calculations we evidently correct.

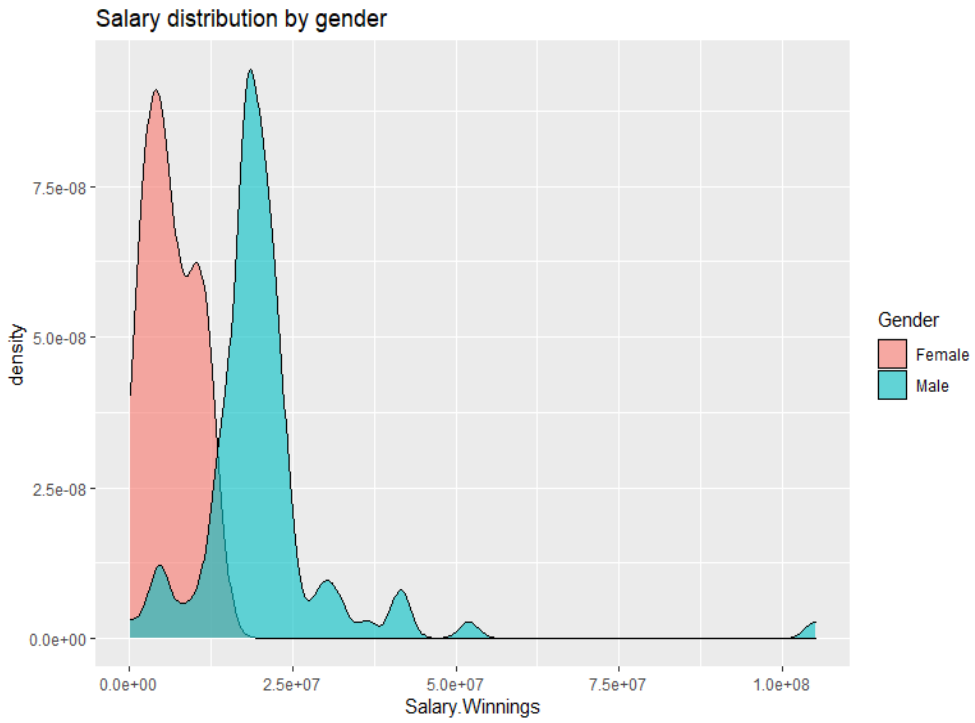

Figure 6: Density Plot Representing Salary in Terms of Gender

Figure 6 shows an evident representation of the salary gap between male and female athletes, most male athletes in the top 100 earners tend to make way more than their counterparts with no regard the sport they play. This insight gives us a good visual representation of the real problem we are facing which is a salary gap between male and female athletes.

4. Additional Analysis

Since females in the top 100 most paid athletes are only represented in one single sport, which is Tennis, I conducted an additional investigation to understand the gap between male and female athletes in this sport. The following are my findings:

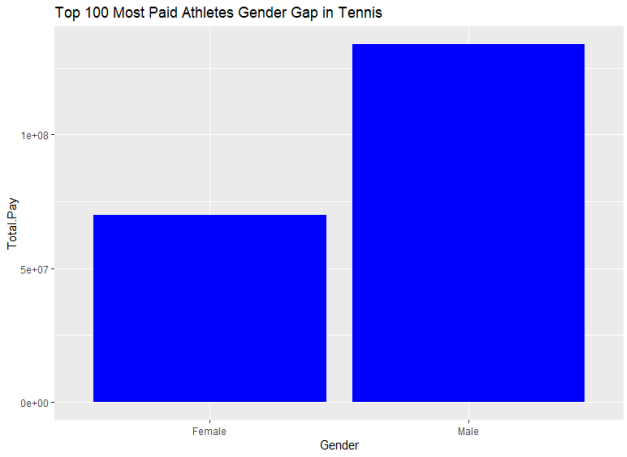

Figure 7: Bar Chart Representing the Total Pay in Tennis Based on Gender

Figure 7 shows an evident gap between the two genders in total pay in Tennis. With Further calculations I was able to determine that on average female athletes approximately made $23,333,333 compared to males who made $44,600,000 which is a difference of 62%.

5. Conclusion

In summary, I used different visualization techniques to make the audience easily see the findings and easily comprehend it. I used density plots and minor calculations to exhibit the skewness of the three attributes, total pay, salary winning, and endorsements. I used a bar chart, a boxplot, and a density plot to convey the problem and visually show the analysis to the audience. We determined the gap between gender participation in respects to sports, which came to about 5.9% for females. We studied the gender gap in sports in respect to the salary winnings to which males heavily dominated the pay coming from sport entities. The following pushed me to discover that the difference of gender total pay, specifically in Tennis was approximately at 62%. And finally, we were able to show that on average the most lucrative sport was Baseball, followed by Basketball and Soccer.

The analysis was limited due to a shortage of necessary skills in statistics and r programming. I had other ideas and problems I wanted to analyze but I was not able to. In the future I hope to get the necessary skills to come back to this dataset and explore more.