Tejasvi Bhagwatkar

Dr. Byrd

CGT 270 – 007

24 August 2021

Acquiring Data – Sport Player’s Salaries

My first dataset came from Tableau, and it consisted of the Global Sport Finances. Tableau is a trusted source used by over a million users worldwide. Tableau allows you to download data as a excel file in a structured format. Within the data, the salaries of the players are listed in a float format. For example, in the year 2014, Russell Westbrook made $19.2 by playing basketball. The rest of the data is listed using various characters. The independent variable of this database are the players, and the dependent variable of the database would be the player’s salary. Since, the salaries are based on the player’s rank, what sport they play and how many endorsements they get based on their salaries. I would describe this data source as a secondary source. Since, the data came from the ESPN and then got rearranged and reorganized in a structured format with many of the earnings being rounded up. The target audience of this database would be sport enthusiastic and investors looking at how much each player costs. This target would mainly consist of educated and some uneducated males, ages 30 and up. Since this database is mainly there to inform, it does not consist of any ideology slant. The database went through an editing process by the community of Tableau, which is a website. The database also seems to be free of any logical fallacies. Finally, the research that went behind putting this data together came from a primary source making it accurate to reality.

My second database came from the website Kaggle, and it was written by a Data Scientist, named Parul Pandey. The data consisted, of a list of the richest athletes from the year 1990-2020. The data came in a structured format, through an excel file. Within the data, the name of the players, and the name of the sport is written using characters. The rank of the athletes is listed using integers and the earning of the athletes is written using floats. The independent variable of the data would be the athletes’ earnings and the dependent data would be the athletes rank since the ranks are based on the earnings. I would describe this data set as a secondary source. Because this data was taken from Forbes magazine and then got reorganized into a structured format with some additional information added to it. The target audience for this data set would be any individual wanting to get more information about a certain player’s earnings each year. Majority of the times these individuals would consist of males in the ages of 30 and up with some sort of an educational background. The data set also went through an editing and a peer review process by the website and other peers and is on its 10th version. The database also seems to be free of any logical fallacies. Finally, the data came directly from a reputable magazine, which gets its information from real primary sources, making it reliable and accurate.

My third database came from the website Kaggle, and it was gathered by a user named Kaggle Novice. It lists the salaries of every NBA player from the year 2000-2019. The data came in an excel file in a structured format. The data set mainly consists of integers and characters. The year that they players joined the NBA, and their salaries and ranks are listed in integers, and the names of the players were listed using characters. The independent variable of this data set would be the rank of the players and the dependent variable would be the player’s salary. Because the higher-ranking players made more than the lower ranking players. I would consider this data set as a secondary source because it seems like it was collected from different websites and got put together in a structured format by someone. The target audience for this data set would be middle-aged males and anyone who wants to be informed about the rankings and the earnings of NBA players. There doesn’t seem to be any logical fallacies in this dataset. However, I do question the accuracy of this data. Because it does not seem to have been edited or reviewed by someone else. And the writer of this data also does not have any credentials attached to his name.