# MAT 243 Project One Summary Report

Angel Davila

angel.davila@snhu.edu

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## Introduction: Problem Statement

I have been hired to analyze historical data for a basketball team to help the team make better business discissions moving forward. To perform this task, I will be using a basketball statistic dataset that is between the years 1945 and 2015.

*FiveThirtyEight. (April 26, 2019).*FiveThirtyEightNBA Elodataset*. Kaggle. Retrieved from https://www.kaggle.com/fivethirtyeight/fivethirtyeight-nba-elo-dataset/*

The main method is to compare the statistics of a certain period to the great Bulls team in the late 90’s. I will compare look at key metrics like total points scored, elo, etc. Mainly focusing descriptive statics and distribution.

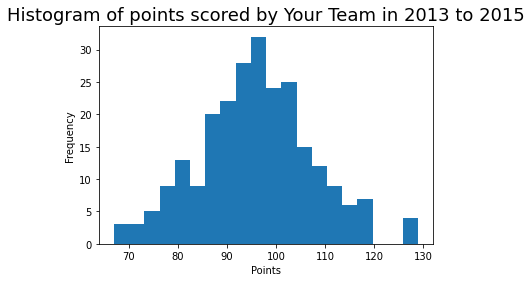
## Introduction: Your Team and the Assigned Team

I grew up in the northwest suburbs of Chicago, so I will choose the Chicago Bulls from 2013 to 2015 and we will be analyzing their data compared to the great team of the 90s.

Table 1. Information on the Teams

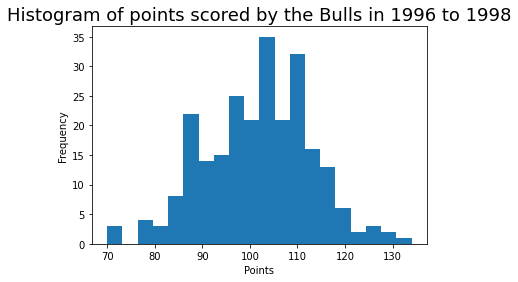
|  | **Name of Team** | **Assigned Years** |
| --- | --- | --- |
| 1. Yours | Bulls | 2013 - 2015 |
| 2. Assigned | Bulls | 1996 - 1998 |

## Data Visualization: Points Scored by Your Team

The first couple of data visualizations was the distribution of points and their frequency. The frequency will give you a good understanding of what is an outlier and what isn’t. You can see that the median of the points distribution was around 95 points. I picked this picture below to as it can give you a good visualization of the frequency of points. This can tell you how frequent they got a certain number of points that year per game. In the below picture, you can see that the median was round 95 or 96 points per game.

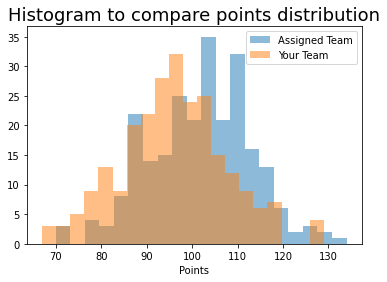
## Data Visualization: Points Scored by the Assigned Team

Since I had already picked a histogram, I decided to use the 96 to 98 histograms as well, to give a better understanding of these point frequencies. Just by looking at this histogram you can see that the point frequency median has shifted to the right. The median being around 105, you can already visualize the difference between a current team against one of the best teams in NBA history.

**

## Data Visualization: Comparing the Two Teams

Having the two histograms overlapping, you can get a better sense of which team was most likely to score more points in a game. The assigned team had more points at more frequent times compared to my team. Because of the left skew, you can most conclude that the 96-98 teams had a better combination of point producers compared to 2013 -2015

****

## Descriptive Statistics: Points Scored By Your Team in Home Games

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1548.57 |
| Median | 1553.64 |
| Variance | 1286.75 |
| Standard Deviation | 35.87 |

Variability shows you the spread of data in your dataset and central tendency shows you the average point where most of the data is measured.

Standard deviation is the difference between each point of data to its mean and is usually used to measure variability within your dataset.

The mean represents the total average of your dataset and in this case, the mean is the complete average of all elo.

Median describes the measure of the center of all your data, where the data is distributed the most. In this case you can see that the most of our data is around 1553.64 elo, which is higher than the average.

When comparing the two histograms, you can see how the 96-98 data is a little bit more skewed left. In this case, the best measurement to represent the center would be the median, as that is where the most frequent the points are scored.

## Descriptive Statistics: Points Scored By Your Team in Away Games

Table 3. Descriptive Statistics for Points Scored by Your Team in Away Games

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1739.8 |
| Median | 1751.23 |
| Variance | 2651.55 |
| Standard Deviation | 51.49 |

For away games point scored, the skew can be described as a left skew, as the median is going positive from the mean. I would say that the team is scoring better in away games.

## Confidence Intervals for the Average Relative Skill of All Teams in Your Team’s Years

Table 4. Confidence Interval for Average Relative Skill of Teams in Your Team’s Years

| **Confidence Level (%)** | **Confidence Interval** |
| --- | --- |
| 95% | ( 1502.02 , 1507.18 ) |

Confidence intervals are used to estimate the central tendency of the population. When the population parameters are not known, samples statistics provide us with the estimation of those parameters. We don’t know exactly how close our sample statistics are to the population. With confidence interval, this helps the sample provide a more precise estimation where the population parameter will fall. With our confidence level being 95%, we can get as close to a more consistent measurement with our sample data. Lowering the confidence level will reduce our level of surety.

## Confidence Intervals for the Average Relative Skill of All Teams in the Assigned Team’s Years

Table 5. Confidence Interval for Average Relative Skill of Teams in Assigned Team’s Years

| **Confidence Level (%)** | **Confidence Interval** |
| --- | --- |
| 95% | (1544.09 , 1553.05) |

Right off the bat, you can see that my team’s confidence margin is larger than the 96-98 Bulls. This margin might be caused because of the number of years, since n grows with the number of years grow.

## Conclusion

The practical importance of the analyses performed was to understand a dataset of statistics for our team compared to a team which has great success. Breaking down all our dataset into descriptive statistics gives us a better visualization of what really is happening under the hood. With this we can make better business decisions when shuffling the team around.

These results mean that although we might not be able to shape our team to exactly the team setup of the 96-98 Bulls, but we can see that, statistically, to increase our potential for winning games, we might need to skew our stats towards the great bulls of the 90’s.