# MAT 243 Project One Summary Report

Ryan Hatch

ryan.hatch@snhu.edu

Southern New Hampshire University

## Introduction - Problem Statement:

In this project, I was trying to evaluate the performance of a specific basketball team to help make decisions for enhancing the team’s performance. I mainly focused on key performance indicators, like points scored, the teams relative skill level (or ELO rating), and game outcomes over a specific number of years. I used a historical collection of NBA game statistics as the data set. This included information like game scores, ELO ratings, game locations, and the game results for various teams across a period of years. To help solve, I used different statistical methods and descriptive statistics to help break down and summarize the performance metrics, which helped to know the average performance and find the range of variability. I then used data visualization in order to visually represent how the metrics were being distributed with the data provided. I then used confidence interval calculations to get an estimate of the average skill level of the teams, which helped me measure the certainty of my findings.

## Introduction - Your Team and the Assigned Team:

The team that I chose for analysis was the Lakers, using a range of years from 2013 – 2015.

The team that was assigned for the comparative study was the Chicago Bulls, using a range

Between years 1996 – 1998.

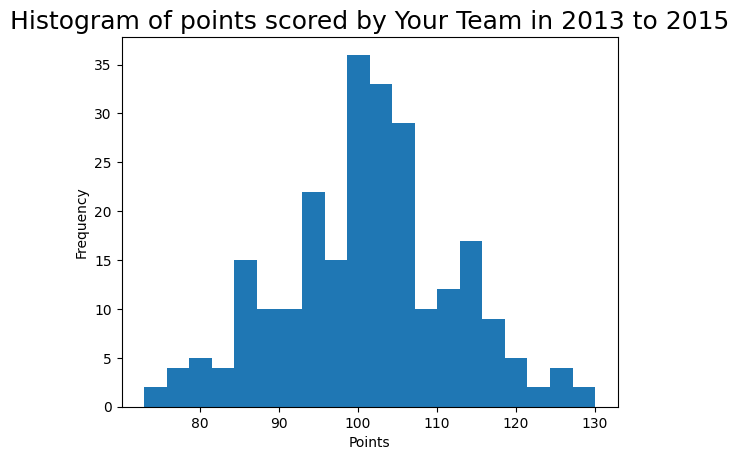
Table 1. Information on the Teams

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

## Data Visualization - Points Scored by Your Team:

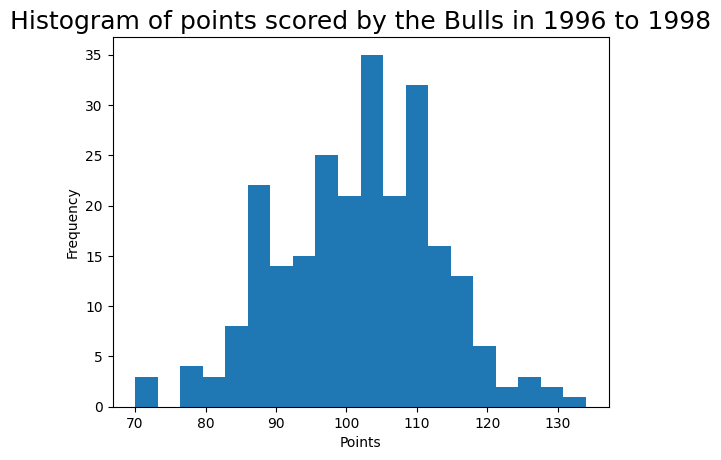
Data visualization is essential for communicating information in a clear and efficient way through graphical representation. Visualizing data can make complex information more understandable and usable. It enables stakeholders to see analytics presented visually, which can help them understand difficult concepts or spot new patterns.

I chose to use a histogram to show the distribution of points scored by the Lakers because it was especially helpful at showing the frequency of game points, which helped to see that the points scored between 2013 – 2015 tend to gather around certain values. The distribution appeared to be normal but with a slight tile to the right, suggesting that there were more games where the Lakes had more higher scoring games than low scoring games. This shows that the team usually scored within a specific range, with not many drastic differences in low or high scores.



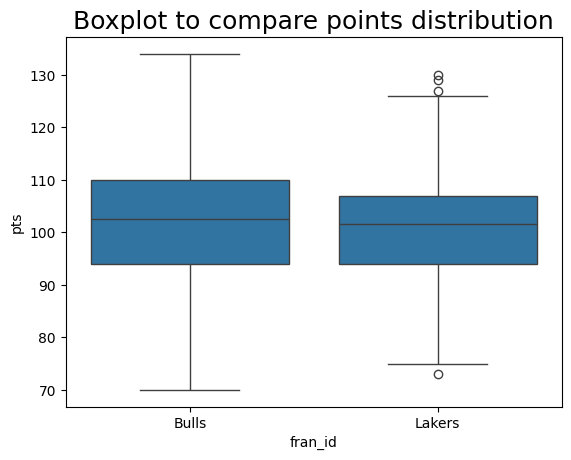
## Data Visualization - Points Scored by the Assigned Team:

The Chicago Bulls performance from 1996 – 1998 was best visualized using a histogram. This representation helps effectively represent the spread of their scores over all the games within that period. The histogram shows the frequency of different scoring intervals, which provided a clear view of the teams scores and their habits.  
  
The Bulls scores follow a bell-shaped curve, which shows a normal distribution. This pattern means that most games had scores that were mostly concentrated around the middle range, with less instances of higher or lower scores. This consistency shows that the team consistently scored within a certain range. The peak in the middle of the plot highlights the most common scoring interval. The pattern in the histogram displays a fairly offensive strategy that led to a consistent scoring range across most of their games.



## Data Visualization - Comparing the Two Teams:

Comparing two sets of data is much easier when you are able to visualize the data. Using visual representations like boxplots, histograms, and scatter plots, is a very beneficial method that helps to make it easier to see how two sets of data relate to each other. For this task, I chose to use my boxplot because it is clear and shows the data’s range, median, quartiles, along with any outliers, making it straightforward to compare between two groups. When comparing the two teams using the boxplot, you can see that the median points scored by the Bulls and the Lakers are pretty close, but the Bulls have a slightly tighter interquartile range, or IQR, which showed that the Bulls are slightly more consistent with their scores. There were also a few outliers in the Bulls box, which also helped to show that the Bulls were more consistent with their playing strategy during their games, in comparison to the Lakers.



## Descriptive Statistics -Points Scored by Your Team in Home Games:

Table 2. Descriptive Statistics for Points Scored by Your Team in Home Games

| Mean | 101.70 |
| --- | --- |
| Median | 102.00 |
| Variance | 149.18 |
| Standard Deviation | 12.21 |

The concepts of central tendency help to understand the central point of a data set by focusing on the mean and median and helping to provide insight into what a typical value would be. None the less, measures of variability, such as variance and standard deviation, help to show how spread out the data is or how much the data points can differ from the average.

* The mean is 101.7 points. This shows that if I was to evenly distribute the points that were scored in all the home games across each game, on average you would end up with about 1017 points scored for each game.
* The median score is 102 points which shows that half the games had scores that were above 102 and half had scores that were below, pinpointing the middle of the data. Its less influenced by extreme values or skewed data.
* The variance, a measure of how much the scores vary, came out to 149.18 points. This indicated a relatively high spread in the game scores.  
  The standard deviation is around 12.21 points, which shows that on average the scores deviate from the mean by about 12 points in either direction.

Since the mean and the median are so close in value, it shows that the distribution of points scored by the Lakers in home games is likely symmetric or bell shaped. This also means that I can assume that there are no extreme outliers which skew the distribution of the scores, leaving the mean to be a suitable measure of central tendency for this data.

## 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 | 100.71 |
| Median | 101.00 |
| Variance | 88.16 |
| Standard Deviation | 9.39 |

The Lakers average score in away games was 100.71 points. I found this mean by adding up all the scores and dividing by the number of games played. The median score is 101 points, which means that in half of the away games, the team scored more than 101 points, and in the other half, they scored less than 101 points. The variance of the scores is 88.16 points, showing how much the scores vary from the average. The standard deviation is 9.39 points, which helps show the average distance of the scores from the mean. The fact that this is lower than the standard deviation for home games can help to prove that the team's scores in away games are more consistent.

The distribution of the team's scores in away games can be understood through the mean and median. Their closeness implies a likely symmetric distribution, or a bell-shaped curve. If the mean were significantly different from the median, it would indicate a skewed distribution. Here, the similar values point to a balanced distribution, with either the mean or median representing the central tendency effectively.

Comparing home and away performances, the team's home mean score of 101.7 is slightly higher than away score of 100.71, hinting at a potential advantage with home games. The standard deviation is also larger at home games, 12.21 compared to away games being 9.39, which shows a more variability in home game scores. This could be because of different strategies used in home versus away games. Overall, the team scores slightly better at home but shows more consistency 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 estimate a population parameter's central tendency with a specified confidence level. A 95% confidence interval means that I am 95% sure the true mean is within that range.

For the Lakers, the 95% confidence interval for the average ELO score is 1502.02 to 1507.18. This indicates we're 95% confident the true average ELO score for all teams during those years is within this interval.

A 99% confidence level would result in a broader interval.

A 90% confidence level would lead to a narrower interval.

## 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% | (1487.66, 1493.65) |

The confidence interval for the average ELO score of both the Lakers and the Bulls between 1996 and 1998 indicates that there is a 95% that the true average falls between 1487.66 and 1493.65, assuming that if I took more samples and calculated the average ELO score for each, 95% of those averages would lie within this range. It helps to give a sense of the uncertainty surrounding the estimate of the average ELO.

Raising the confidence level to 99% would result in a wider interval, which would show a greater certainty that the true mean is within it but also acknowledging more uncertainty about its exact location.

Lowering the confidence level to 90% would make a narrower interval, indicating less certainty but a much more precise and reliable estimate.

Comparing this interval with the one for 2013-2015 shows that they don't overlap, suggesting a difference in average ELO scores between these periods. The earlier period's interval is lower than the later one, hinting at an increase in overall team skill levels over time. This could be due to changes in league dynamics, or even improvements in skill and game strategy. It might also reflect varying levels of competitiveness or different methods to calculate ELO between the periods.

## Conclusion

I was to run analysis on both the Lakers and the Bulls. I parsed all of the data and used the outputs to better understand the basketball team's performance. By looking at the average and spread of the scores, I was able to get insight into the scoring patterns at the home and away games. This helps to make tailored strategic decisions, like where the team needs to improve and how to stabilize their performance.

The stats that I got from the outputs also show the team's skill level compared to the league average, which helps to guide trainers and players to make better decisions based off of their personal needs to develop. I also got from the data that while the teams score more at home, there is more consistency in away games. These two examples that were found in the data could help lead to changes in the way the teams train, practice, and prepare for games. Running this type of analyses can help teams make more informed decisions on where to focus on specifically in order to get better outcomes and more scores in games.