# MAT 243 Project One Summary Report

Brandon Taylor

Brandon.taylor6@snhu.edu

Southern New Hampshire University

**Notes:**

* Replace the bracketed text on page one (the cover page) with your personal information.
* You will use your selected team for all three projects

## Introduction: Problem Statement

*Discuss the statement of the problem in terms of the statistical analyses that are being performed. In your response, you should address the following questions:*

* What is the problem you are going to solve?
  + I am a data analyst for a basketball team. The problem I must solve is to analyze the team’s performance. Particularly, I am reviewing my team, the Sixers
* What data set are you using?
  + I am using a historical basketball data set, specially the nbaallelo.csv file obtained on Kaggle.
* What statistical methods will you be using to do the analysis for this project?
  + I will be using descriptive statistics like the mean, median, variance and standard deviation, as well as utilize data visualization techniques such as histograms and scatterplots. We will use this to obtain the confidence interval for this assignment.

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

## Introduction: Your Team and the Assigned Team

*In this project, you picked a team and you were assigned a team to do comparative analysis.*

*See Steps 1 and 2 in the Python script to address the following items:*

* What team did you pick and what years were picked to do the analysis?
  + The team that I had chosen for this assignment is the Sixers. The analysis period for this team is 2013-2015.
* What team and range of years were you assigned for the comparative study? (Hint: This is called the assigned team in the Python script.) Present this information in a formatted table as shown below.

Table 1. Information on the Teams

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

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

## Data Visualization: Points Scored by Your Team

*In the Python script, you created a visualization for the distribution of points scored by your team.*

*See Step 3 in the Python script to address the following items in a paragraph response:*

* In general, how is data visualization used to study data distributions and trends?
  + Data visualizations, like histograms and scatterplots, are used to graphically represent data, making it easier to identify patterns, understand the shape of the distribution, detect outliers, and observe relationships between variables. Histograms are particularly useful for understanding the distribution of a single variable, while scatterplots can show trends between two variables.
* In this activity, you were asked to pick one of the two plots that best describes the data distribution of the variable for your team. Include a screenshot of this plot in your report.
  + A graph of points scored by a team

    AI-generated content may be incorrect.
* Why did you pick this plot?
  + I picked this plot because I felt that it best showcased the frequency of how much they scored and how often
* What can you say about the distribution of the variable by visually inspecting this plot? What does this signify? I would say that it is center is symmetric or nearly bell shaped.

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

## Data Visualization: Points Scored by the Assigned Team

*In the Python script, you created a visualization for the distribution of points scored by the assigned team.*

*See Step 4 in the Python script to address the following items in a paragraph response:*

* In this activity, you were asked to pick one of the two plots that best describes the data distribution of the variable for the assigned team. Include this plot in your report.
  + A graph of points scored by the bulls

    AI-generated content may be incorrect.
* Why did you pick this plot? Again, this is a good graph to give the general frequency of the points that they score
* What can you say about the distribution of the variable by visually inspecting this plot? What does this signify? This seems to somewhat skew right, but it seems to ultimately be a bell shaped curve

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

## Data Visualization: Comparing the Two Teams

*In the Python script, you created a visualization for the difference in the distributions of points scored by your team and the assigned team.*

*See Step 5 in the Python script to address the following items in a paragraph response:*

* In general, how is data visualization used to compare two different data distributions?
* In this activity, you were asked to pick one of the two plots that best compares the data distributions of your team with the assigned team. Include a screenshot of this plot in your report.
  + A graph of a diagram

    AI-generated content may be incorrect.
* Why did you pick this plot? This plot shows the perfect distribution rate between the team. This version of the chart now only shows how their shapes overlap, but it also shows which side of the chart either team is skewed one, since they both appear bell curved otherwise.
* How do the two distributions compare to each other? The Bulls skew to the right with more points, the Sixers screw left with less point. They overlap around the middle often, though.

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

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

*In the Python script, you calculated descriptive statistics on the points scored by your team in games played at home venue. These included the mean, median, variance, and standard deviation for the relative skill of your team.*

*See Step 6 in the Python script to address the following items:*

* Summarize **all** statistics in a formatted table as shown below. Use one row for each statistic. You will need to add rows to the table in order to include all of your statistics.

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

| **-** | **Value** |
| --- | --- |
| Statistic  *(for example, Mean)* | X.XX  *\*Round off to 2 decimal places.* |

* In general, how are the measures of central tendency and variability used to analyze a data distribution?
* Interpret each statistic in detail and explain what it represents in this scenario.
* Use the mean and the median to describe the distribution of points scored by your team in home games.
  + Describe the skew: Is it left, right, or bell-shaped?
  + Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.

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

* During the 2013-2015 period, the Sixers' performance in home games, as measured by points scored, showed a central tendency around the mid-90s. The average (mean) points scored per home game was 96.74, while the median score was 95.00 points. This median value indicates that in half of their home games, the Sixers scored 95 points or fewer, and in the other half, they scored 95 points or more. The mean being slightly higher than the median suggests a mildly right-skewed distribution in their home game scores, possibly due to a few games with exceptionally high point totals. In such a skewed distribution, the median (95.00 points) often provides a more robust representation of the typical central performance as it's less influenced by these extreme values. In terms of scoring consistency, the standard deviation was 10.25 points. This means that individual game scores typically varied from the average score by about 10.25 points, reflecting the spread of their scoring performance at home; the variance was 105.05 (squared points), which also measures this dispersion.

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

| **Statistic Name** | **Value** |
| --- | --- |
| Mean  Median  Variance  Standard Deviation | 96.74  *95.0*  *105.05*  *10.25* |

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

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

*In the Python script, you calculated a 95% confidence interval for the average relative skill of all teams in the league during the years of your team. Additionally, you calculated the probability that a given team in the league has a relative skill level less than that of the team that you picked.*

*See Step 8 in the Python script to address the following items:*

* Report the confidence interval in a formatted table as shown below.

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

| **Confidence Level (%)** | **Confidence Interval** |
| --- | --- |
| XX% (for example, 95%) | (X.XX, X.XX)  *\*Round off to 2 decimal places.* |

* Describe how confidence intervals are generally used in estimating the measures of central tendency for a population.
* Provide a detailed interpretation of the confidence interval in terms of the average relative skill of teams in the range of years that you picked.
* How would your interval be different if you had used a different confidence level?
* What is the probability that a given team in the league has a relative skill level less than that of the team that you picked? Is it unusual that a team has a skill level less than your team?

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

*In the Python script, you calculated a 95% confidence interval for the average relative skill of all teams in the league during the years of the assigned team. Additionally, you calculated the probability that a given team in the league has a relative skill level less than that of the assigned team.*

*See Step 9 in the Python script to address the following items:*

* Report the confidence interval in a formatted table as shown below.

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 95% confidence interval for the average relative skill (ELO) of all NBA teams was calculated to be (1487.66, 1493.65). This interval suggests that we can be 95% confident that the true average ELO for all teams in the league during those years fell between 1487.66 and 1493.65. If a different confidence level had been used, the interval would change; for example, a 99% confidence level would produce a wider interval, providing greater certainty but less precision, whereas a 90% confidence level would result in a narrower, more precise interval but with less certainty.

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

## Conclusion

*Describe the results of your statistical analyses clearly, using proper descriptions of statistical terms and concepts.*

This project really dug into the team's performance by looking at the historical data. When we put the scoring patterns of our team and the assigned team side-by-side visually, some clear differences in how they put points on the board emerged, making for a straightforward comparison of their offensive styles. Drilling down into our team's own numbers, we saw how their average scores and consistency varied between home and away games, and their overall scoring distribution had a particular shape. Beyond just our team, looking at the average skill level across the league for the two different time periods studied suggested that the general level of play in the NBA may have shifted over the years.

Ultimately, this kind of statistical analysis is really valuable because it gives the coaching staff and management solid, fact-based information. Knowing these specific performance details, like any differences in home and away scoring or how consistently the team played, can help shape game plans and training focus. Clear visuals and stats also make it easier to share these complex performance insights with everyone involved. Plus, seeing how things stacked up in different NBA eras gives us a better yardstick for evaluating our team.