# MAT 243 Project One Summary Report

[Full Name]

[SNHU Email]

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

The following analysis is based on an NBA data starting on the 1950 until 2020. The data from this dataset will be analyzed using descriptive statistics and data visualization techniques to study distributions of key variables associated with the performance of different teams. Statistical methods such as mean, median, variance, standard deviation and more will be used to better understand the performance of the selected teams.

## Introduction: Your Team and the Assigned Team

Table 1. Information on the Teams

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

The Lakers is the team I chose for this comparison. The assigned years for the Lakers ranges from 2013 – 2015. In contrast, we have the Bulls ranging from 1996 – 1998 as the assigned team.

## Data Visualization: Points Scored by Your Team

Data visualization is a method of representing data in a graphical format to help people see and understand the information. There are many ways that data visualization can be used to study data distributions and trends. One way is by using graphs, charts, or tables to help people visualize the information. Another way is by using interactive graphics to allow people to explore the data in more detail. Data visualization helps us comprehend the information that we are looking at more easily because it makes it easier for us to see patterns and trends in the data. It also helps make it clear when there are inconsistencies in the data or when something seems off about it.

For the team I selected, I chose the histogram as the plot. Histograms are a graphical representation of the distribution of data. They are used to visualize the distribution of data and how it changes over time. Histograms are useful for understanding how data is distributed and can be used for many purposes. They can be used to find outliers, measure variability, or compare distributions. Histograms are easier to read than bar charts because they use two dimensions instead of three dimensions. This means that we can see more information at once in one glance.

The histogram allows us to easily see that the highest point frequency is at 100 points, we can easily deduce this by looking at the center of the histogram. Although there are a few outliers.

*Chart, histogram

Description automatically generated*

## Data Visualization: Points Scored by the Assigned Team

To keep our data visualization uniform, we also used a histogram. Thus, we can easily identify that the highest frequency is around 104 points.

*Chart, histogram

Description automatically generated*

## Data Visualization: Comparing the Two Teams

Data visualization can be used to compare two different data distributions by representing the data in a visual format. Data visualization helps us understand how one distribution compares with another distribution by showing us how the data is distributed. Data visualization can also show us an overview of the entire dataset at once instead of showing individual values for each piece of information. This makes it easier for viewers to see patterns or trends in the dataset without having to look at every single piece of information separately.

The plot selected for this section is the Box Plot. A box plot is a data visualization that uses a box and several lines to depict the distribution of data in a dataset. A box spans the middle 50% of the data, with Q1 as the lower boundary of the box and Q3 as the upper boundary of the box. The median is shown as a line inside the box. Two lines, known as whiskers, extend from the lower boundary of the box to the minimum and from the upper boundary of the box to the maximum. The whiskers represent the lower and upper 25% of the data.

Using the Box plot we can see that the median is really close at 102, although the Q3 for the Bulls is better than the Lakers by a few points. Something else we can clearly see is that the Bulls have more outliers than the Bulls.

*Chart, box and whisker chart

Description automatically generated*

## 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

| **Type** | **Value** |
| --- | --- |
| Statistic  *(for example, Mean)* | X.XX  *\*Round off to 2 decimal places.* |
| Mean | 101.7 |
| Median | 102 |
| Variance | 149.18 |
| Standard Deviation | 12.21 |

* 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.

There are four main measures of central tendency that are used to analyze a data distribution. These include the mean, median, mode, and weighted mean. The most used measure of central tendency is the mean. This is because it is easy to calculate and does not require any special calculations or knowledge about the data set.

The skew for this

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

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

*See Step 7 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 3. Descriptive Statistics for Points Scored by Your Team in Away Games

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

* 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 away games.
  1. Describe the skew: Is it left, right, or bell-shaped?
  2. Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.
* Is your team performing better in games played at home than those played away? Use the mean and the standard deviation to answer this question. What can be deduced by comparing the standard deviation of points scored in home games and points scored in away games?

* 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?

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

* Provide a detailed interpretation of the confidence interval in terms of the average relative skill of teams in the assigned team’s range of years.
* Discuss how your interval would be different if you had used a different confidence level.
* How does this confidence interval compare with the previous one? What does this signify in terms of the average relative skill of teams in the range of years that you picked versus the average relative skill of teams in the assigned team’s range of years?

* 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.*

* What is the practical importance of the analyses that were performed?
* Describe what these results mean for the scenario.

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

## Citations

*You were* ***not*** *required to use external resources for this report. If you did not use any resources, you should remove this entire section. However, if you did use any resources to help you with your interpretation, you* ***must*** *cite them. Use proper APA format for citations.*

Insert references here in the following format:

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.