## Introduction: Problem Statement

I have been tasked with analyzing a very large data set of historical data and to find patterns within the data set. The team management as well as the coaches of both the Miami Heats and the Chicago Bulls have requested for me to use descriptive statistic with data visualization methods to find distributions of key studies and key variables that are associated with the performance of both teams. With data analytics, it will help both teams’ management to make decisions and make further improvements for both teams. I have been assigned the Chicago Bulls data sets from the year 1996 – 1998 and the data sets of the Miami Heats from the year 2013 – 2015. The methods that I will be using for statistical purposes will be to use data visualization techniques in forms of charts to compare both teams. I as well will use Descriptive statistics to calculate the variance and relevance of both teams. Using Confidence intervals, I can calculate both averages of relative skills on both teams by using the data sets from the years assigned.

## Introduction: Your Team and the Assigned Team

This summery report provides the analysis and my findings of patterns using historical set of data of the professional basketball team performances in the United States. By using descriptive statistics as well as data visualization, the summery report is going to be used by the team management to help aid in improvements in the basketball team performances. Using the team’s data sets. We can compare and calculate the average relative skills that are from the Miami Heat’s using the years 2013 – 2015. We are going to use the Chicago Bulls data set as well from the years 1996 – 1998. I picked the Miami Heats to analyze, and I was assigned the Chicago Bulls for comparative studies.

Table 1. Information on the Teams

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

By using data visualization. You can view the data in formats such as tables or charts that allows for easier interpretations of the data sets. Below is a histogram as well as a scatterplot I used to show spread of points and their frequency for both teams as well as their time frames. Using a histogram and scatterplot you can be difficult to see since larger data sets need a more columns to spread the data sets. Overall, you can use the histogram to view the numbers and frequency to have a good estimate. You can see that the Chicago Bulls has higher frequencies from approximately 110 points compared to the Miami Heats that had frequencies below 110 points. This type of visualization observed tells us the Bulls tend to score above average points compared to the Miami Heats

## Data Visualization: Points Scored by Your Team

Visualizing data conveys information to be used by the viewer by interpretating the displayed data in formats such as tables or charts. Both tables and charts are quick to comprehend the data as well as being effective in analyzing large datasets. Charts helps the viewer find trends or relationships with the data that is charted within a specific timeline. These types of charts show progression and any trend that are either upward or downward trends. I chose to use both charts to show the difference between which data sets are represented to displayed points scored for the Miami Heats from the years 2013 – 2015.

**

**

A scatterplot shows a relationship between two variables. By using the points scored per game and the year it was scored in. It shows a trend of downward, thus, showing the performance of the team. By using the histogram, it shows a continuous data where points scored and the years, they were scored in are represented in such a way that enables the user to view trends. Both charts can be used to show linear regression curves and display trends – downward trend in this case. It helps by analyzing training and its efficiency for the years being displayed. Although, histograms show continuous data, it is limited to the amount of data it can show.

## Data Visualization: Points Scored by the Assigned Team

I chose to use the histogram to display the distribution on points scored by the Chicago Bulls from the years 1996 – 1998. The histogram also shows linear regression curves that displays the distributed points made by the Chicago Bulls from the years 1996 – 1998. It shows a downward trend per score points per year. This downward trend has shown a regression that is quickly regressing compared to the Miami Heats regression curve.

**

**

## Data Visualization: Comparing the Two Teams

The boxplot can be used for a better understanding and visualization of the datasets. It displays the variance, minimum, standard deviation, and the range of the datasets. Maximum is the largest dataset used and the minimum is the smallest dataset. The range represents the deference between the maximum and minimum of the dataset. By using Quartiles. You can use the 25th Quartile and the second can be the 50th Quartile. Q1 is at the bottom of the box, Q2 is the middle of the box, Q3 is the top of the box and all three are represented by the number summaries. The box plot can also be used to detect outliers to be showed.

**

By chosen the box plot. It shows a difference between both teams and their distributions. The box plot shows both teams share similar central tendencies. Furthermore, whiskers indicate different minimums and maximums such as the range of points between both teams. The years 1996 – 1998 for the Bulls shows an upward trend and the years 2013 – 2015 shows a downward trend for the Heats. It as well shows that the Bulls have a scoring average that is more than the Heats average scored points. It is a bell shape

## Descriptive Statistics: Relative Skill of Your Team

Table 2. Descriptive Statistics for Relative Skill of Your Team

| **Statistic Name** | **Value** |
| --- | --- |
| MeanMedianVarianceStandard Deviation | 1617.481652.579987.299.94 |

Measuring the central tendencies and variability are to be used to summarize while analyzing the distributions of very large datasets. The Miami Heats, Using the mean sums all the data value in the datasets provided by both teams, thus, can be divided by the numbers and values within the datasets. The Heats have a mean of 1617.48 through all three seasons. The median was 1652.57 which can be represented as either an ascending or descending order. The variance shows the average of the square difference from both the mean and standard deviation. Thus, resulted in a larger variance that is spread-out by the square root of the variance. The Heats as well had a standard deviation of 99.94 points.

## Descriptive Statistics: Relative Skill of the Assigned Team

Table 3. Descriptive Statistics for Relative Skill of the Assigned Team

| **Statistic Name** | **Value** |
| --- | --- |
| MeanMedianVarianceStandard Deviation | 1739.81751.232651.5551.49 |

The Chicago Bulls have a mean of 1739.8 from the years 1996 – 1998. The median was 1751.23. Variance was 2651.55 and the standard deviation was 51.49. It is a bell-shaped distribution. The variance as well as the standard deviation shows that the Chicago Bulls had lower average which translate to them having low variability and having been consistent with their skills for all three seasons.

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

My calculations shows that 95% confidence intervals are averaged for relative skills of both teams. You can as well calculate the probability of the given teams in the league with relative skills shows that their skills were less than the Miami Heats.

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

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

An estimate can be a single value estimate for using factors. Intervals estimates are ranges of values that contain factors. Combining interval factors and probability shows a statement for the confidence levels. Confidence levels are percentages that shows which intervals contains factors. 95% confidence level shows a team’s average skills of being 1502.02 points as well as 1507.18 points within the rage (years) 2013 – 2015. The probability that teams within the league have relative skills level less than the Miami Heats from the years 2013 – 2015 are 0.3254

## 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% | 1501.98, 1507.23 |

95% confidence level of all the teams’ average skills are between 1501.98 points and 1507.23 points from the years 1996 – 1998. The probability the any given team within the league have relative skills less than the Chicago Bulls from the years 1996 – 1998 are 0.3699. It shows that the probability is 33% that other teams’ skills are less than the Chicago Bulls from the years 1996 – 1998. It means that the Bulls average relative skills from the years 1996 – 1998 are better then the Miami Heats from the years 2013 – 2015.

## Conclusion

In summary, the points scored in the years 1996 – 1998 had similar distribution averages to that of the Miami Heat’s from the years 2013 – 2015. There was a downward trend that showed regression for the Bulls compared to the Miami Heats. The Miami Heat’s and the Chicago Bulls had almost similar central tendencies which are closely symmetrical. The Heats and the bulls have similar 25th Quartile compared to the Bulls 75th Quartile which equaled more points scored by the Heats 75th Quartile. Upon further research, the minimum and the maximum points indicated different ranges which the Miami Heats had and Outliers. Form the years 1996 -1998 it showed that the Bulls had an upward trend which indicated that the Bulls scored more than average points per game than the Heats.

Averaging, the Bull’s had higher skills from the year 1996 -1998 compared to the Heats 2013 – 2015. The variance as well as the standard deviation were lower for the Bulls which indicated that the Bulls had lower variability and were consistent with relative skills throughout all three seasons. The Bull’s mean was 1739.8 compared to the Miami Heats mean of 1617.48. It showed and indicated that the Bulls scored on average more points than the Miami Heats during all three seasons. With a 33% probability it showed that relative skills of another team were less than that of the Bulls from the years 1996 – 1998. With a 26% probability the Miami Heats relative skills were less than other teams within the league from the years 2013 – 2015. The results were that the Bull’s average relative skills from the year 1996 – 1998 were better than that of the Miami Heat’s from the year 2013 – 2015.

## Introduction: Problem Statement

I have been tasked with analyzing a very large data set of historical data and to find patterns within the data set. I will be analyzing relative skills of each team using specified years – comparing wins vs losses. The teams that I am comparing are the Thunders (my team) and the Chicago Bulls (assigned team). The sets that I will be using are datasets using the mean relative skills of each team. The methods that I will use to do my analysis are z-test and t-test

## Introduction: Your Team and the Assigned Team

I picked the Thunders and the years selected to do my analysis are from 2013-2015. The team assigned was the Chicago Bulls with the selected years from 1996-1998 for the comparative studies.

Table 1. Information on the Teams

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

## Hypothesis Test for the Population Mean (I)

* 1. H0: μ> 1342. Null hypothesis is claiming that my teams averaging relative skills are more than 1342 by using the 5% significance level. If not, enough sufficient data shows more evidence.
	2. H0: μ< 1342. Alternative hypothesis is contradicted of the claim.
	3. The level of Significance is 5% or 0.05
	4. Report the Test Statistic and the P-value in a formatted table as shown below:

Table 2: Hypothesis Test for the Population Mean (I)

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 83.0 |
| P-value | 0.0 |

Based on my findings. The p-value of 0.0 and the alpha value of 0.05 would reject the null hypothesis. Since the alpha value is far greater than what the p-value equates.

* Implications within my findings are null hypothesis will be rejected. It as well shows the average relative skills of the team being no greater than 1342. The low significance level of 1342 shows me that the team did not perform well during the specified years.

## Hypothesis Test for the Population Mean (II)

* 1. H0: μ> 110. Null hypothesis is claiming that my teams averaging relative skills are more than 110 by using the 1% significance level. If not, enough sufficient data shows more evidence.
	2. H0: μ<110. Alternative hypothesis is contradicted of the claim.
	3. The level of Significance is 1% or 0.01
	4. Report the Test Statistic and the P-value in a formatted table as shown below:

Table 3: Hypothesis Test for the Population Mean (II)

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | *6.94* |
| P-value | *0.0* |

Based on my findings. The p-value of 0.0 and the alpha value of 0.01 would reject the null hypothesis. Since the alpha value is far greater than what the p-value equates.

* Implications within my findings are null hypothesis will be rejected. It as well shows the average of the team scoring points being no greater than 110. The low significance level of 1342 shows me that the team did not perform well during the specified years.

## Hypothesis Test for the Population Proportion

* 1. H0:  p = 0.50. Null hypothesis is claiming that my teams won games scoring 80 or more using a 5% significance level. If not, enough sufficient data shows more evidence.
	2. H0:  p does not equal 0.50. Alternative hypothesis contradicts the claim.
	3. Level of Significance is 5% or 0.05
	4. Report the Test Statistic and the P-value in a formatted table as shown below:

Table 4: Hypothesis Test for the Population Proportion

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | *0.0* |
| P-value | *1.0* |

Based on my findings of my p-value. It will not be rejecting the null hypothesis since the alpha value is 0.0 of the p-value.

* Implications within my findings are null hypothesis will not be ejected. Proportion that was given was 0.6749. It means that the teams scored more than 80 points for winning the game.

## Hypothesis Test for the Difference Between Two Population Means

* 1. H0: μd = 0. Null hypothesis is claiming that my team and assigned team has the same mean skills level significance. If not, enough sufficient data shows more evidence.
	2. H0: μd does not equal 0. Alternative hypothesis contradicts the claim. The mean skill level of Significance is 1% or 0.01
	3. The level of Significance is 1% or 0.01
	4. Report the Test Statistic and the P-value in a formatted table as shown below:

Table 5: Hypothesis Test for the Difference Between Two Population Means

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | *17.82* |
| P-value | *0.0* |

Based on the p-value of 0.0. It shows that it is less than the alpha value of 0.01. In return, it will reject the null hypothesis.

* Implications within my findings of this hypothesis test will be rejected are the mean skill levels of both teams and showing they are not the same with the skill levels. The skill levels of my team’s vs the assigned team shows that my team had a mean level of 1651.28 and the assigned team had a mean level of 1739.8. The Chicago Bulls was better than my chosen team.

## Conclusion

With my analysis. The importance was to perform and calculate the average points scored by both teams using the specified years. Both mean relative skill level was used to calculate their winning points of 80 or more during the specified years. The information provided to me allowed me to determine the team’s performance during the specified years.

A set of predetermined parameters was used as well to help with the analysis of this data. The predetermined scores average about 110 points per game. By using the predetermined parameters, you can estimate your selected team’s performance.