# MAT 243 Project One Summary Report

Brandon Petersen

brandon.petersen@snhu.edu

Southern New Hampshire University

## Introduction: Summary Report

I am a Data Analyst hired by the Houston Rockets basketball team and have been given access to a large set of historical data to analyze the performance patterns of the team. I will be using the data set to identify key performance metrics of the Chicago Bulls compared to that of the Houston Rockets. To start, I used the data historical data from 1996 to 1998 of the Chicago Bulls and 2013 to 2015 of the Houston Rockets, and I measured the relative skill level for each team. From here, I will be running descriptive statistics on the relative skills calculated, such as calculating the mean, median, variance, and standard deviation. Once this information has been calculated, I will be calculating the average relative skills for the Chicago Bulls and the Houston Rockets to a 95% confidence interval and comparing their performance between the two time periods.

Table 1. Information on the Teams

| **Name of Team** | **Assigned Years** |
| --- | --- |
| Houston Rockets | 2013 - 2015 |
| Chicago Bulls | 1996 - 1998 |

## Data Visualization: Points Scored by Houston Rockets

Data is everywhere. Everything in some way can be used and interpreted as data. "Data analysis is the collection, transformation, and organization of data to draw conclusions or make predictions(Google)." Data can be used to tell a story and guide business decisions. Data visualization helps add weight to the story being told with data by utilizing graphs and charts.

A scatter plot graph was created while working on this project. The scatterplot was used to represent the scored points for the Houston Rockets between the years 2013 to 2015. The scatterplot graph was chosen over other graphs, such as a histogram. It provides the reader with better visualization and understanding of how many points are being scored each year.

Based on the data provided by the scatterplot below, during the year 2013, the Houston Rockets scored points ranges approximately between 80 and 140 with an approximate average of 108, in 2014 scoring points approximately between 80 and 145 with an approximate average of 107, and 2015 scoring points approximately between 70 and 130 with an approximate average of 105. It can be seen in the scatterplot that each year does contain multiple outliers and a negative slope. This indicates a slight decline in the overall performance of the Houston Rockets between the years 2013 and 2015.

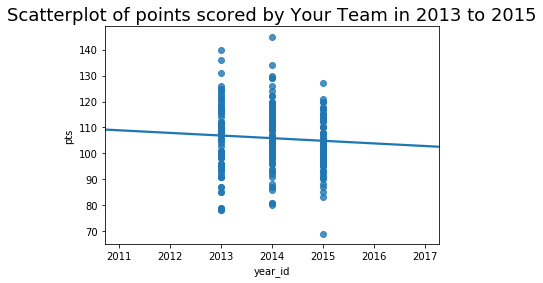
**

Figure 1 Rockets Score 2013-2015

## Data Visualization: Points Scored by Chicago Bulls

Similar to the Houston Rockets, a scatter plot graph was created while working on this project. The scatterplot was used to represent the scored points for the Chicago Bulls between the years 1996 to 1998. The scatterplot graph was chosen over another graph, such as a histogram. It provides the reader with better visualization and understanding of how many points are being scored each year.

Based on the data below provided by the scatterplot, the Bulls during the year of 1996 approximately scored between 70 and 125 points with an approximate average of 106; in 1997, they approximately scored between 70 and 135 with an approximate average of 102, and in 1998 they approximately scored between 70 and 122 with an approximate average of 98. Similar to the Rockets, the Bulls had multiple outliers during the three years with a negative slope. This indicates that the Bull's performance declined over the three years. When you compare the slopes of each scatterplot, the Bulls have a significant decline in performance compared to that of the rockets.

*Chart

Description automatically generated*

Figure 2 Bulls Scored Points 1996-1998

## Data Visualization: Comparing the Houston Rockets and Chicago Bulls

"Data visualization is useful for data cleaning, exploring data structure, detecting outliers and unusual groups, identifying trends and clusters, spotting local patterns, evaluating modeling output, and presenting results (Unwin, 2020)." In addition, data can be quickly gathered from graphs and charts. For instance, comparing two graphs can help to identify patterns and trends effortlessly.

A boxplot chart was created to compare points distribution between the different years to compare these two teams. The boxplot chart was chosen over a histogram because the boxplot provides more information at a glance and is easier for the reader to gather data from. The histogram plots the data on top of each other, unlike where the boxplot separates the two data sets. The boxplot graph data such as minimum points scored, maximum points scored, median points scored, the quartiles for points scored, and whether the data is skewed in a particular direction or not.

When comparing the boxplot graphs of the Chicago Bulls to the Houston Rockets, the Houston Rockets performed better overall. We confirmed this when we compared the two slopes produced from the scatterplot. The Rockets had a higher median of approximately 106 compared to the Bulls of 103. In addition, the Rockets had a higher minimum of 78 and maximum of 135 compared to the Bulls minimum of 71 and maximum of 133. With that said, the Rockets did have outliers above their maximum and below the Bull's minimum. However, even with these outliers, the width of the boxplot graph for the Rockets is smaller than that of the Bulls. This indicates that the Rockets scored more consistently than the Bulls. It can also be seen that both quartiles of the Rockets boxplot are higher than that of the Bulls. To reiterate, based on the boxplots below, we can see that the Houston Rockets between the year 2013 to 2015 scored more consistently and better than the Chicago Bulls during the years between 1996 to 1998.

*Chart, box and whisker chart

Description automatically generated*

Figure 3 Bulls vs. Rockets

## Descriptive Statistics: Relative Skill of the Chicago Bulls

Table 2. Descriptive Statistics for Relative Skill of the Chicago Bulls

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

Descriptive data such as mean, median, and mode are known as a central tendency. The mean calculates the average of the data set, the median is the middle value of the data set when it is ordered from smallest to largest, and the mode is the value that reoccurs most often in the data set. The central tendency helps to provide a clear insight into the data set.

There is a ranking referred to as ELO in many sports, which equates to that team or person's relative skills in that sport. As seen in the table above, the comparative statistics for the Chicago Bulls are calculated from the Bulls ELO ranking during the years between 1996 to 1998. The mean calculated for the Rockets is 1739.8, the median is 1602.14, the variance is 1663.7, and the standard deviation is 40.79. The variance refers to how to spread the data set is from the mean, and the standard deviation refers to how the data set is distributed. Based on all the data provided, we can conclude that the data for the Chicago Bulls is slightly skewed to the right. You can observe that the distribution is skewed to the right first in the boxplot. You can see the "whisker" connected to quartile three, and the maximum is slightly longer than the "whisker" below it.

Along with that, based on the descriptive statistics above, the mean is larger than the median. This indicates that there is a positive skew to the data set. In other words, the data set is skewed to the right instead of being a bell-shaped distribution.

## Descriptive Statistics: Relative Skill of the Houston Rockets

Table 3. Descriptive Statistics for Relative Skill of the Houston Rockets

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1596.29 |
| Median | 1602.12 |
| Variance | 1663.7 |
| Standard Deviation | 40.79 |

Similarly, with the Chicago Bulls, the Houston descriptive scores were calculated on their ELO ranking between the years 2013 to 2015. The mean calculated for the Bulls is 1596.29, the median is 1602.12, the variance is 1663.7, and the standard deviation is 40.79. Based on all the data provided, we can conclude that the data for the Houston Rockets is slightly asymmetrical. In other words, the distribution of the data set is slightly skewed to the left. This is slightly harder to observe on a histogram or the boxplot provided. However, based on the descriptive data provided, the mean is less than the median, indicating a negative skew in the data set.

When comparing the variance between the Rockets and the Bulls, we see the Rockets have a lower score of 1663.7 compared to 2651.55. This indicates that the data set of the Rockets is more trustworthy than that of the Bulls. When comparing the standard deviation of the two teams, we see the Rockets have a score of 40.79 compared to 51.49. This indicates that the Rockets have a more consistent data set than that of the Bulls.

## Confidence Intervals for the Average Relative Skill of All Teams in Houston Rockets

Table 4. Confidence Interval for Average Relative Skill of Teams in Houston Rockets

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

Confidence intervals are often used to measure the uncertainty in a data set. When calculating confidence intervals, the most common to be used is 95%. Based on the data set for the Houston Rockets using a 95% confidence interval, it was found that the average relative skill between the years of 2013 to 2015 falls in the range of 1502.05 to 1507.18 with a mean of 1596.29. Depending on the confidence interval used would indicate a different average relative skill. For instance, if we were to use a confidence interval of 99%, the average relative skill would be 1501.21 to 1507.99, and if we used a confidence interval of 90%, we would have an average relative skill in the range of 1502.44 to 1506.77. IN other words, if we use a lower confidence interval, the average relative skill range lowers, and if we increase the confidence level, the average relative skill range increases.

The probability that a given team in the league has a relative skill less than the Houston Rockets is 0.2086. Since this probability is more significant than 0.05, it can be considered an average probability. On the other hand, if the score were less than that of 0.05, it would be considered unusual.

## 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) |

Based on the data of all teams during the years between 1996 to 1998, it can indicate with a 95% confidence level that average relative skill falls in the range of 1487.66 to 1493.65. Depending on the confidence interval used would indicate a different average relative skill. For instance, if we were to use a confidence interval of 90%, the average relative skill would be in the range of 1488.14 to 1493.17, and if we used a confidence interval of 99%, the average relative skill would be in the range of 1486.72 to 1594.59. If we use a lower confidence interval, the average relative skill range lowers, and if we increase the confidence level, the average relative skill range increases. When you compare the Houston Rockets average relative skill range of 1502.02 to 1505.18 in the years between 2013 to 2015 to the league team's relative skill range of 1487.66 to 1493.65 in the years between 1996 to 1998, we see that the Rockets have a higher relative skill score than the league's teams. Along with that, we can see that the average relative skill range for the Rockets is much smaller than that of the leagues. This indicates that the Houston Rockets have a higher performance than the league's team.

## Conclusion

In conclusion, by analyzing the Houston Rockets and the Chicago Bulls data sets, we notice both similarities and differences. For instance, we notice that both teams for the data gathered in the specific years show that both teams' performances are currently declining and have a negative slope. Along with that, we notice that the Houston Rockets have a more consistent and higher average relative skill than that of the Chicago Bulls. In this project, we used graphs such as the scatterplot and boxplot to understand better what is going on with the data and performance of each team. Then we calculated the central tendency of each team to help support and add validity to what the data was visualizing.

## Citations

* *Google Data Analytics. Coursera. (n.d.). https://www.coursera.org/professional-certificates/google-data-analytics.*
* *Unwin, A. (2020, January 31). Why is Data Visualization Important? What is Important in Data Visualization? · Issue 2.1, Winter 2020. Harvard Data Science Review. https://hdsr.mitpress.mit.edu/pub/zok97i7p/release/3.*