Summary Report: FiveThirtyEight NBA Elo dataset Analysis

Name

Date

# Applied Statistics for STEM 21EW3

## Introduction: Problem Statement

Nowadays almost every team in NBA has at least has one data analyst whose job is to analyze the team performance in previous games. Almost every decision in NBA games is now based on analytics. In this project, it was required to choose any team and do analysis on its performance over the years using statistics techniques. The dataset used for this analysis is FiveThirtyEight NBA Elo dataset. There are four important variables in data set that you have been used to carry out analysis. These variables are:

Table 1: Data Variables

|  |  |
| --- | --- |
| **Variable** | **What does it represent?** |
| pts | Points scored by the team in a game |
| elo\_n | A measure of the relative skill level of the team in the league |
| year\_id | Year when the team played the games |
| fran\_id | Name of the NBA team |

Descriptive statistics and data visualization techniques are used to study distributions of key variables associated with the performance of team. The findings of this analysis will help the management make decisions to further improve team’s performance. The analysis is performed using python programming language.

## Introduction: Your Team and the Assigned Team

The table 2 below shows selected team and assigned team for the analysis in this project.

Table 2. Information on the Teams

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

## Data Visualization: Points Scored by Knicks

Data visualization is a clear representation of information and data. By using visual aids such as charts, graphs, and maps, data recognition tools provide an accessible way to see and understand trends, outliers, and patterns in data. In python script, we created a visualization for the distribution of points by our team. We created two plots one is histogram and other is scatter plot. Histogram has points on x-axis and frequency on y-axis and scatter plot has year id on x-axis and points on y-axis.

*![Chart, histogram

Description automatically generated]()*

Figure 1: Histogram of points scored by your team

The histogram best defines the points scored by our team because it can be very helpful in inspecting team performance as it tells the frequency of different points scored by team between the years 2013 to 2015. Histogram also shows that the distribution of points is normal distribution having peak at 100. The plot also tells that the team has scored very few times around 70 and 125 points.

## Data Visualization: Points Scored by the Bulls

In python script, we have also created a visualization for the distribution of points by assigned team.

*![Chart, histogram

Description automatically generated]()*

Figure 2: Histogram of points scored by Bulls

Histogram plot also seems better choice to visualize points than scatter plot assigned teams as again it shows the frequency of points over the years and gives better interpretation of team’s performance. Bulls point distribution is slightly skewed as the middle point is approximately 100 and it has highest peek at approximately 105. The minimum score for bulls between year 1996 to 1998 had been around 71 and its frequency is approximately 3. The maximum score is about 132 and it has been scored only once.

## Data Visualization: Comparing the Two Teams

A very common application for a data visualization is to compare two data distributions. Through visualizing we can estimates the difference in minimum, maximum, median and sometimes mean of two distributions. In python script we have created two plots i.e., boxplot and histogram to compare two distribution. The box plot seems better choice to compare two distribution than histogram because box plot gives more statistical knowledge which is required to compare two distributions.

![Chart, box and whisker chart

Description automatically generated]()

Figure 3: Box Plot for comparison

The box plot shows that the minimum score of Knicks is about 65 which has occurred only once between year 2013 to 2015. The minimum score of bulls is about 70. The median of points of knicks is less than that of bulls. The box plot also shows that half of times score of knicks had been in range of 88 to 105 approximately and of bulls had been from 92 to 110 approximately

## Descriptive Statistics: Relative Skill of Your Team

Central tendency includes measure of mean, median and mode. It helps in analyzing what is the average, middle and most frequent values in dataset. On the other hand, variability is measure of the spread of dataset. Variability represents how accurately sample represents the entire population. In the Python script, we have calculated descriptive statistics on the relative skill of knicks. The table shows descriptive statistics for relative skill of knicks

Table 3. Descriptive Statistics for Relative Skill of Your Team

|  |  |
| --- | --- |
| **Statistic Name** | **Value** |
| Mean | 1471.29 |
| Median | 1474 |
| Variance | 12288.68 |
| Standard Deviation | 110.85 |

The average points of knicks between the years 2013 to 2015 has been around 1471.29. The median of points was 1474. The variance of points is 12288.68. The amount variance of points had been 110.85. This analysis shows that the difference in expected point and actual points scored by knicks had been 110.85. The distribution of points is slightly skewed to left as the mean is less than median. Note that this difference in mean and median is very small this is why by looking at histogram plot of points we first estimated that the distribution is normal, or bell shaped. The median is usually preferred to other measures of central tendency to represent the center of distribution when your data is skewed, or you are dealing with original data. The mean is best to represent the center of the distribution when data distribution is bell shaped.

## Descriptive Statistics: Relative Skill of the Assigned Team

In the Python script, we have also calculated descriptive statistics on the relative skill of assigned. The table shows Descriptive Statistics for Relative Skill of assigned team which bulls.

Table 4. Descriptive Statistics for Relative Skill of the Assigned Team

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

The average points of knicks between the years 1988 to 1990 has been around 1739.8. The median of points was 1751.23. The variance of points is 2651.55. The amount variance of points had been 51.49. This analysis shows that the difference in expected point and actual points scored by knicks had been 51.49.

The distribution of points is skewed to left as the mean is less than median. The variance and standard deviation of relative skill of bulls is less than knicks. This difference shows that bulls had more consistent relative skill between 1998 to 1990 than knicks between year 2013 to 2015.

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

We need a way to understand the precision of our mean (or measure of central tendency). Range and interquartile range are one way of understanding more about the shape of the curve. Confidence Intervals are used to quantify the uncertainty by providing a lower limit and upper limit that represent a range of values that will represent the true population parameter with a specified level of confidence.

Table 5. Confidence Interval for Average Relative Skill of Teams in Your Team’s Years

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

The 95% confidence interval (rounded) for average relative skill (ELO) of all teams in the years 2013 to 2015 is equal to (1502.02, 1507.18). The average mean is between 1502.2 and 1507.18 with confidence level of 95%. Probability a team has Average Relative Skill LESS than the Average Relative Skill (ELO) of knicks in the years 2013 to 2015 is 0.3841.

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

A confidence interval for the mean is a way of estimating the true population mean. Instead of a single number for the mean, a confidence interval gives you a lower estimate and an upper estimate.

Table 6. 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 (rounded) for average relative skill (ELO) of bulls in the years 1996 to 1998 is equal to (1487.66, 1493.65). Different confidence level will give different intervals of average relative skill. Increasing confidence level will decrees the margin of error and will give more accurate average relative skill interval. The average relative skill of all teams in assigned years (1996 – 1998) is less than chosen years (2013-2015). From the year 1996 to 1998 the average relative skill of teams is between 1487.66 and 1493.65 and from the year 2013 to 2015 the average relative skill is between 1502.02 and 1507.18. This shows that there has been increase of average relative skill of teams between years 1998 to 2013.

## Conclusion

The analysis done above can be helpful in increasing teams’ performance. This analysis shows Knicks and Bulls have performed over a specific time. The analysis can help mangers and team coaches to make decision which results in increase of relative skill of team and points scored match.