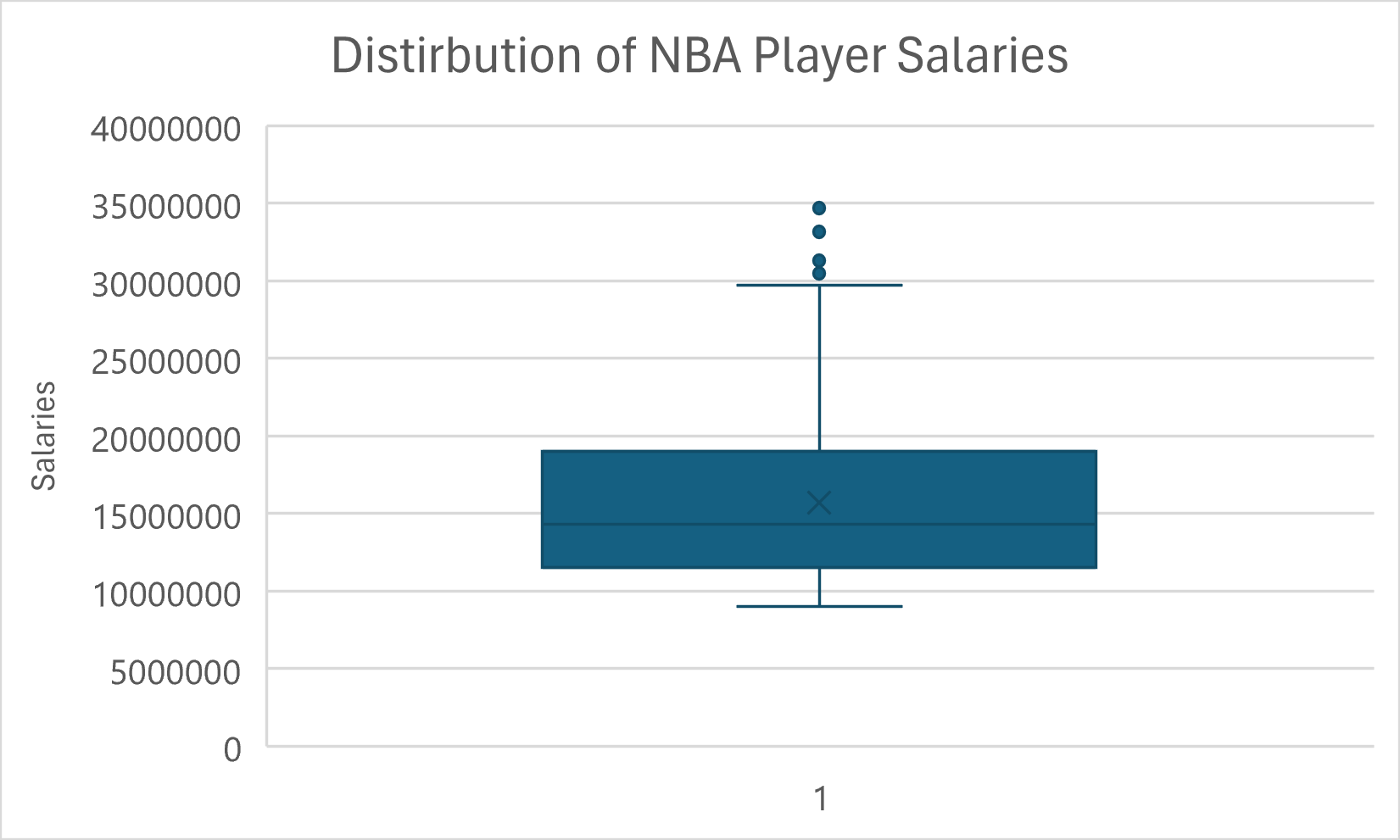
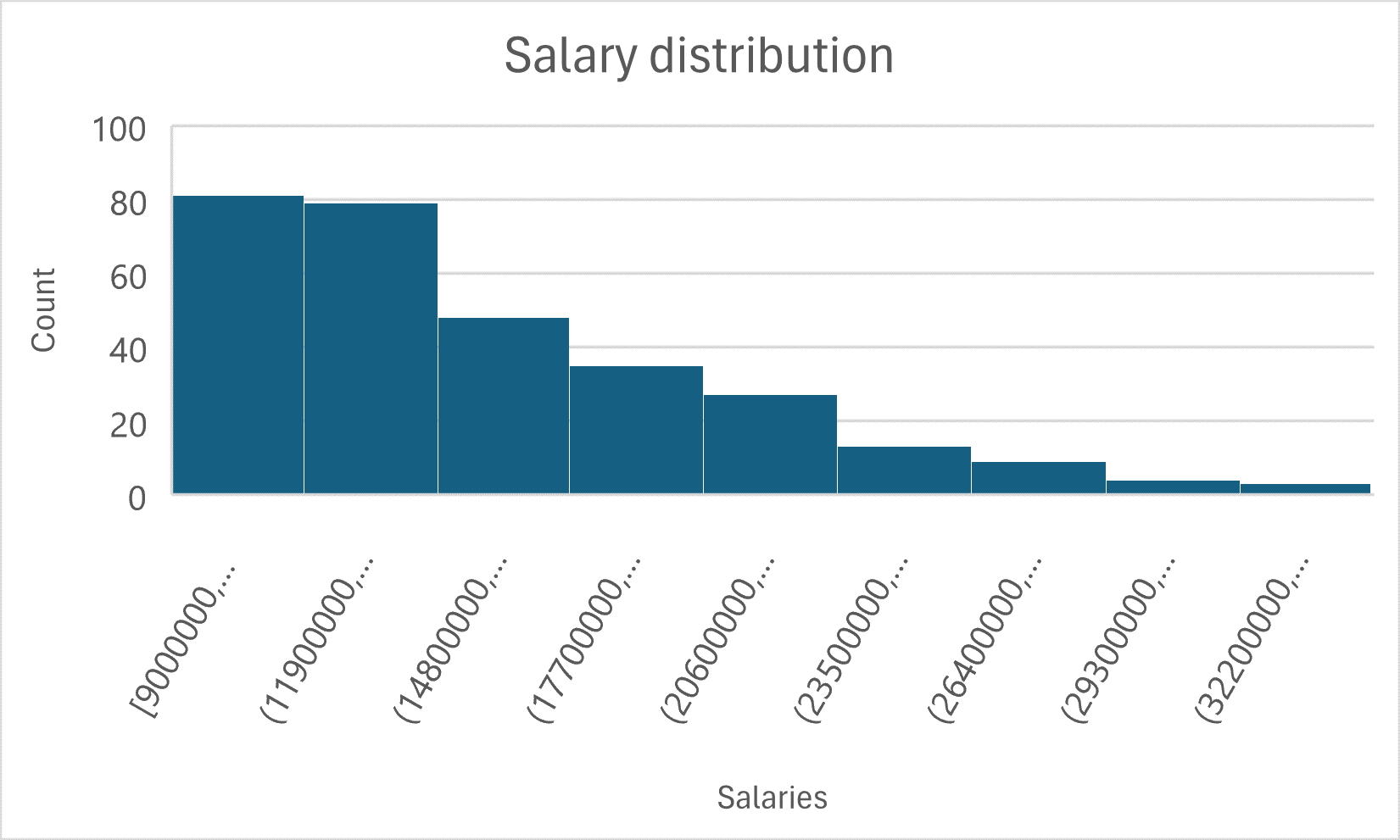
Jonathan Eudja

Data Analysis Final Project Research Memo

* 1. The NBA is a prominent professional basketball league in the United States, known for its highly skilled athletes and competitive environment. One of the main problems I aim to solve with this dataset is to determine what impact do team conferences have on total player salaries in the NBA? This question is important to understand player preferences for team locations and to analyze team financial capacities. It provides insights into conferences that players prefer and teams they are likely to choose during free agency.
  2. In this study, I use descriptive and inferential statistical analysis of data about NBA player salaries we look to see if the given conferences make any type of impact when it comes to total player salaries.

1. 1. The Population for the study is all NBA players that play for different teams and for different seasons. The Data includes specific teams that is listed from 1980s to 2010s but in this case, I will only be using recent teams and not past teams like the Supersonics, Charlotte Bobcats, etc. I would only be using teams from 2010s and onwards.
   2. The sample frame is a dataset containing NBA Player Salary Information from the 1980s to 2010s.
   3. The sample frame I created is that I selected the top 300 most paid players in the NBA during that time frame from a pool if 2,409 unique players from the dataset. This approach ensures that there is no bias as well as representing the population of the dataset.
   4. To avoid any type of error or bias, I attempted to include as many players as possible, whether they have the highest or lowest salary I made sure to include them in the study. As mentioned above I also made sure to remove any duplicates so that there is no bias or duplicate number of players in the dataset.
2. Title of Dataset: NBA Salaries, Source: Data World
   1. Link to Dataset: [NBA Salaries](https://data.world/datadavis/nba-salaries)
   2. Columns of Dataset:
   3. League- Categorizes the type of league the player participates in – categorical – nominal.
   4. playerid - Identifies each player uniquely. – categorical – ordinal
   5. salary – The amount the player earns per year and quantifies the amount of money the player earns per year – numerical – continuous – ratio
   6. season- the start and end of the season – categorical – ordinal.
   7. season\_start- Specifies the year when the season begins. – discrete
   8. season\_end- Specifies the year when the season ends. – discrete
   9. team- Categorizes the team for which the player plays – categorical nominal
   10. A screenshot of a calculator

       Description automatically generatedconference- Categorizes which team belongs to which conference – categorical – nominal
   11. Eastern\_salaries – salaries from eastern conference teams – numerical – continuous
   12. western\_salaries – salaries from western conference teams – numerical – continuous
   13. Descriptive Statistics:
       1. So in our descriptive Statistics we see here that our Average Salary is around 3,651,500$ and our median is lying at around 1,372,960$. We also see are Kurtosis value is above three and is lying at around 6.14 which indicates that are salaries are not normal, which also indicates that the there is a lot more higher salaries that are more shown.
       2. Boxplot:
       3. Histogram:



1. Null and Alternate Hypothesis
   1. Hypothesis for ANOVA Test
      1. Null Hypothesis: The mean salary is the same across All NBA teams
      2. At least one NBA team has a different mean salary then other NBA teams
   2. Hypothesis for Two Sample Test
      1. There is no difference in average salary of NBA Players between the Eastern and Western Conference
      2. There is a difference in average salary of NBA Players between the Eastern and Western Conference
2. Test Justification
   1. ANOVA: So I chose to do ANOVA to test my hypothesis because that will allow me to compare multiple groups and determine if one of the teams have a different mean salary.
   2. Two Sample T-Test: I chose to do a two Sample test because it will allow me to compare the Average salaries of NBA players between the Eastern and Western Conference. It will allow me to record significant difference in the average salary between the two conferences.
3. Test Results
   1. Two Sample T-Test (Minitab):

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a spreadsheet

Description automatically generatedTwo Sample T-Test (Excel):

Interpretation: So with a test stat of 0.60 this indicates that the difference in average salaries for the NBA players in both conferences is pretty small meaning there is not much of a difference in average salaries. A P-value of 0.54 means that we are not rejecting the Null hypothesis as there is not enough evidence to reject meaning there for that we can’t conclude that there is a difference in average salaries between the eastern and western conference teams.

A graph with blue lines and white text

Description automatically generatedA black text with numbers and numbers

Description automatically generated with medium confidenceANOVA Testing (Minitab):

Interpretation: So we get a test statistic of 0.95, this indicates that the difference in average salaries for the NBA players in both conferences is pretty small meaning there is not much of a difference in average salaries; and for our P-value we see that our p-value is at 0.55, which means its way over our level of significance. Therefore we can conclude that there is not enough evidence for us to see that at least one NBA team has a different average salary then others, therefore concluding that there is no significant difference in average salary across all teams.

1. Conclusion
   1. So based on the ANOVA and Two sample test that was ran, we find no significant evidence to suggest that team conferences have a substantial impact on total player salaries in the NBA. This means that any variations in salaries we observed could simply be due to random factors, rather than any meaningful difference between the conferences. However, given the limited data we have to the specific data set and timeframe used in the analysis, this is just a limited conclusion. Future Analyses should be considered on expanding the sample size to include more players as well as teams.
   2. The Limitations provided was of course the sample size used which I selected to be around 300 teams, As well as the statistical tools used as well that have some limitations as wells.
   3. For Future analyses I would probably expand the sample size of my dataset to more players to see if there is a difference in salary, and to try and find current updated data from this decade and to use more statistical models when analyzing the data to find other results with the data.