Terrick Boire

Advanced Sports Analytics

5/22/2023

Team Valuation Project

For this project we were tasked with recommending a team or teams we should buy based on their value. To begin this project, I had to go through the task of cleaning the data and fixing all the missing data. To do this I split the data up by year as you can see in my code. Once the data was split up, I was able to look at the filter data to help me better decide as to how I should clean it. For example, all the data from the year 2023 just needed to be made numeric and have the commas removed. However, the other years were not this simple as there was much more that had to be done to them. For the year 2021 I had to first remove the percent symbols from two columns and then convert those columns into their equal percent value. Once that was done, I was able to remove the commas and change the character data into numeric data. When it came to the 2020 data, I decided I would use that as my base data. What that means is that all the unknown values were made zero. To help me fill in any other unknown information I used the internet and referred to it in my code. For the entire data set population was unknow for the years 2023 and 2021 so I decided to use the population data from 2020 to complete the data set. The reason I used the same population for each year is because it saved me time, instead of looking up populations individually based on year I used what was given to me. Now that the data was cleaned, I was able to make some key comparisons as well as build a couple of regression models.

The first comparison I made was comparing the MLB`s value to revenue ratio to other leagues. As you can see from the two graphs below I first compared the MLB to the NBA.

A screenshot of a graph

Description automatically generated with low confidence

A picture containing text, screenshot, diagram, line

Description automatically generated

I found that the NBA brings in nearly $50 million in revenue and has a higher average value for teams apart from a couple outliers. Now I wanted to see how the MLB`s value to revenue ratio compared to the NHL. As you can see below the MLB has a significantly higher revenue than the NHL and this could simply be because of its popularity compared to the NHL. What surprised me from both boxplots was that even though one league might not have as much revenue that does not mean that the value of the league will be significantly lower as that average value of all the leagues are relatively close. All this information is valuable information as I plan to make a recommendation as to which team or teams my client should invest in.

A picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, diagram, line

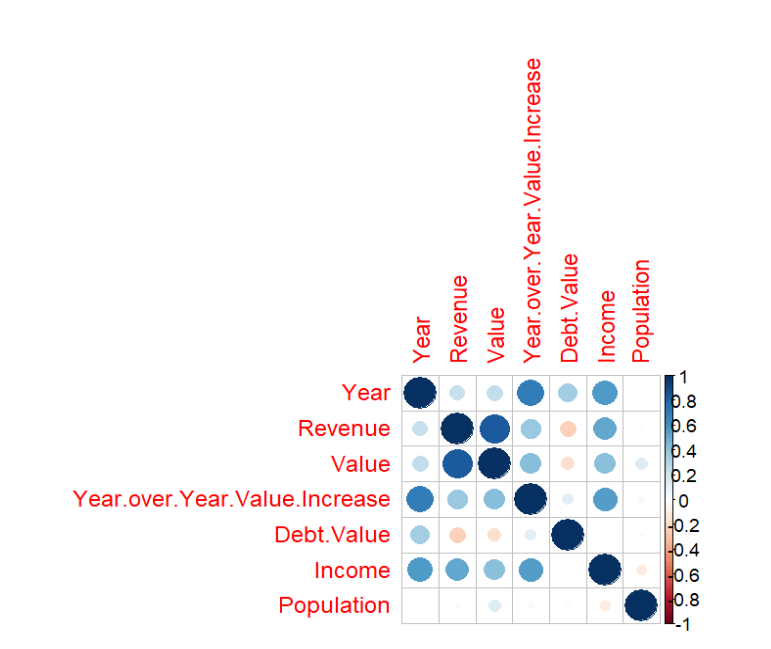
Description automatically generated

Next, I plan on evaluating how the leagues did in other seasons compared to 2021. As we know, on 2021 we were in the middle of a pandemic and that had a huge impact on many businesses including the sports industry since fans could not watch their favorite teams live. Below I compared the three seasons in side-by-side box plots that compared income and income for all of the leagues.

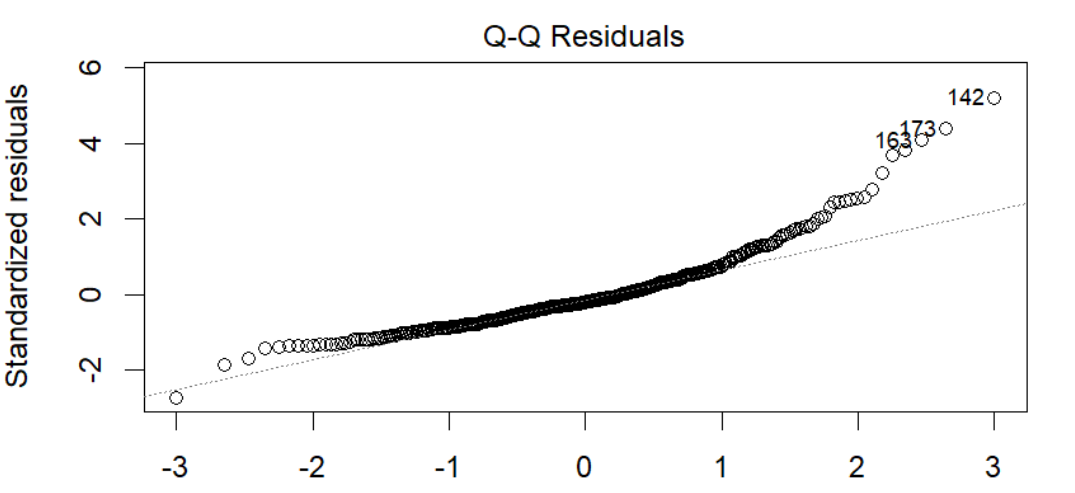
A screenshot of a graph

Description automatically generated with low confidence

As I mentioned before, the year 2020 was used as my baseline since it had a lot of missing data. Covid certainly did have an impact on revenue as you can see a significant increase from all the leagues but the NFL. You can see the same results when it comes to revenue as well. Possible reasons the NFL was not impacted as much could possibly be the time of year the season was played. In the later part of the year covid restrictions did get softer and that is about the time the NFL season is. In addition, games were played outside.

For my regression models I decided to go with linear regression as well as a decision tree. To help me with my linear regression model I thought it would be useful to see which variables were highly correlated to variable Value. I found that revenue, year over year value increase and income were the variables that were highly correlated as you can see below, so I will include them in my model.

Below I included the Q-Q Residuals plot to help show the accuracy of my model. From the plot below I learned that the data is for the most part normally distributed. At the beginning the points fall on or near the 45-degree line and at the end it tails off the line.



My regression trees had similar results as it undervalued as well as overvalued teams. To help me get better results I pruned the tree. The only problem with the regression tree is that there are not unique predictions for each team, that is why I would go with my linear regression model values. However, for me to make a recommendation for a team both models must match. Meaning both models must show that my value of the team is more than the current value. Also the value should increase over time.

A picture containing text, diagram, screenshot, line

Description automatically generated

With all this information gained about the different leagues and my regression models I would recommend that my client at least invest in one NBA team. My recommendation for an NBA team would be the Cleveland Cavaliers at a value of $2.05 billion. The Cavaliers value has only gone up over the years and according to my regression model the valuation of $2.05 billion is lower than my $2.74 billion and $2.9 billion valuations I got from my regression models. With the nearly $1 billion my client has left over I think it would be smart for my client to investigate investing in the NFL or NHL as those are other leagues that bring in the income as well as revenue. However, I would not recommend investing in any teams as there are not enough funds left and there is too much risk with the NHL teams according to my models. My client should be more than happy investing in the Cleveland Cavaliers and could us the other funds for further investments in the team such as a new stadium.