Draft Lab

English Explanation:

- The logic for my Draft Pick program can be broken down into two separate cases: Both parties are giving up one pick for another pick (a 1-for-1 swap) or at least one party is giving up more than one pick.
- In the first case, if the pick being given away is higher than the pick received, the trade will always be considered a success (As you can always pick the better player).
- In the second case, To properly rank the draft picks I decided to use both **Value Over Replace Player (VORP)** and **Win Shares** as my evaluation stats.
- I chose these two stats as I felt that they were the two stats which did a good job at stripping out the bias of position, team strength, and other noise. For example, stats that focus on rebounding skew towards centers while three-point percentage stats can hurt them.
- Running a correlation analysis I saw that these two stats had a r-value of .910, which means that they were strongly positively correlated.
- To actually solve the second case, I used **Linear Regression**, where I went through every draft pick in every year (Or more specifically every draft pick that did not have corrupted data) and created training data where the input would be the pick number of the player and the output would be equal to the **WS Scaled VORP** of the year of that player.
    - Scaled in this case would means that each of these values has a maximum value of 1
    - Keep in mind this means that all players would have more than one entry.
- The Primes of the players are weighted more than their other years.
- One last important note is that the output values are also scaled again according to the draft pick. They are scaled in a manner such that the 60th pick should provide an output approximately equal to zero or in other words the 60th pick should have no value whatsoever.
- After training the model with all the points provided from the picks, to figure out whether the trade is good, we simply calculate to see if **the sum of the predicted values of the picks given is greater than the sum of the predicted values of the picks gotten.**

Numerically Sound:

To extract the data, I went through the draftDB sheet from top to bottom. For each player, I would get their draft pick and their name and then use the latter to find all their stats for all their seasons in the playerDB sheet. To expound a bit more on the logic for how the regression works, the meaning of **Scaled WS** and **Scaled VORP** should be explained. What this means is that first I found the max value of WSand VORP for every possible player using excel. After finding the max value for both, for every player I calculated **Scaled VORP =** $\frac{\boldsymbol{VORp_{Player}}}{\boldsymbol{VORP_{MAX}}}$ and **Scaled WS =** $\frac{\boldsymbol{WS_{Player}}}{\boldsymbol{WS_{MAX}}}$. By scaling such that the maximum value is 1 for both, this meant that when added together, VORP would not dominate the calculation as VORP values tended to be about ten times bigger than WS values. To make sure the Linear Regression was not too small as well, I multiplied everything by 100 as well, which wouldn't change any relations it would just make everything larger.

Finally, in order to make sure that the 60th pick were considered worthless, I added a penalty where I multiplied the value above by $(1 - .1694(\#Pick - 1))$. This meant that the first pick in the draft

would get no penalty with each pick below getting an ever-increasing penalty. This would end with the 60th pick having a weight of 0. In total the formula for a player in a single season would be:

$$\left(\frac{\boldsymbol{VORP_{Player}}}{\boldsymbol{VORP_{MAX}}} + \frac{\boldsymbol{WS_{Player}}}{\boldsymbol{WS_{MAX}}}\right) * \mathbf{100} * \left(1 - .1694(\#Pick - 1)\right).$$

One final note is that is the equation is slightly changed if the player's age for that entry is between 25 and 29, the prime of an NBA player. All that changes is that instead of dividing by $\boldsymbol{VORP_{MAX}}$ and $\boldsymbol{WS_{MAX}}$, we divide by $\frac{\boldsymbol{VORP_{MAX}}}{2}$ and $\frac{\boldsymbol{WS_{MAX}}}{2}$ which means that the prime of a player is weighed twice as much as their other years as this is when their production should be the best.

Overall, the data is not perfect and some obscure players like Damian James had to be omitted because their entries had no data associated with them and some players had draft picks higher than 60 which I decided to omit since all these cases came from the 80s, and since the draft has only been 60 picks for the last three decades, I didn't see a good reason to include them. Like, I said the data is not perfect and the program doesn't always return what I think would be a good trade. For example, my program says giving up the 1st pick for the 15th and 30th pick is a good trade. But I more so think that this is because drafting is incredibly hard and while the 1st pick may be good, there has also been a lot of talent drafted in later position and even in the second round. This is why draft picks are also decided by workouts and the "eye test" as well.