

Michael Seman

CTS-2451

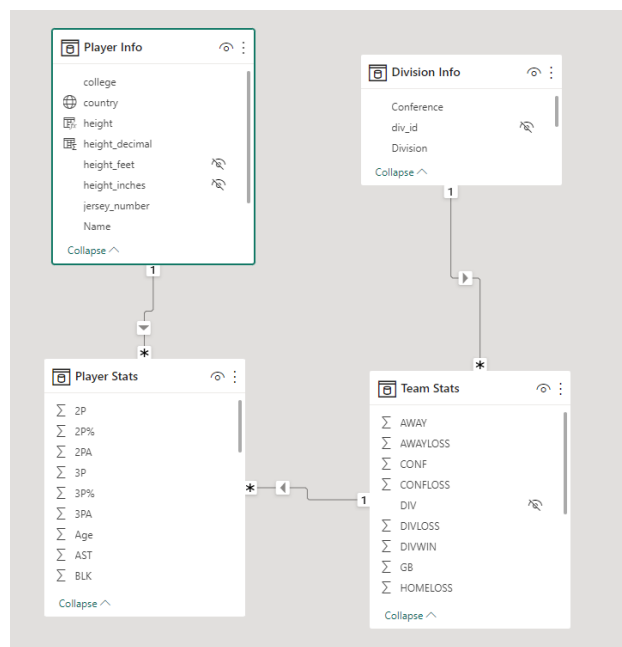
Dr. Ernesto Lee

Final Project

Visualizing the NBA 2020/2021 Season

I decided to do my project in Power BI using a dataset I had previously worked with in SQL Implementation class. I obtained data from the 2021-2022 NBA season using web scraping. All files were .csv files containing individual regular season game stats for the players, team win/loss records, conference and division information, and also individual biographical statistics for players such as height, weight, country, etc. My goal for this project was to create two reports that summarize the main takeaways from these regular season statistics and records. I wanted to highlight the top teams' and players' performances and give the user a way to dive deeper into the data I collected.

The data model is fairly simple. I have four tables: 1 containing player game statistics (named Player Stats), 1 containing player biographical/personal information (named Player Info), 1 containing team records and results (named Team Stats), and 1 table that simple contains the divisions and conferences (named Division Info) along with an id that was then applied to all the teams in that table. Breaking it down now in reverse order, "divisions" had a one-to-many relationship with "teams stats". There were multiple teams in each division. "Team stats" had a one-to-many relationship with "player stats" because players switch teams during the season. As a result some players have many teams. Finally, "player stats" had a many to one relationship to "player info". The stat lines in playerstats are just for that player's stats with one team, so there can be multiple entries per player in that table. But players themselves are only a singular person, so while there can be multiple entries for a single player in playerstats, there is only 1 player thus they appear only once in playerdata, and we have a many to 1 relationship.



In terms of cleaning the data and transforming it there was not too much complex work to be done as I had already done most of this during SQL implementation class. Also, outliers are a thing we want to keep in a dataset like this, so there was no worry about those. That being said, there was a bit of formatting and other issues I faced. One main thing I had to clean was the player height column. The height was listed in the format of feet – inches. Excel and Power BI interpret this as a date. So I initially changed this column to a string in order to keep its original format. In order to use the height for calculations though, I needed to get a column with the height as a numeric. So, I split the column into two, height_feet and height_inches. I then used a simple formula to create a new column where we added height_feet to the height_inches divided by 12 in order to get the height in a usable decimal format. I called this new column height_decimal.

```
height_decimal = ROUND(playerdata20_21[height_feet] + (playerdata20_21[height_inches]/12),2)
```

I also changed the country in the playerdata table to a data category of country. I created a hierarchy for Conference – Division. Finally, I hid all the id columns to make the data easier to use. I also had the player names split between first and last name, so I merged these columns to have a full name for every player.

The biggest issue I faced was the same issue I encountered when I used this dataset previously. The percentage stats cannot be properly aggregated with an average calculation. For example, if a player played on two teams, they would have two different field goal percentages. It is not accurate to simply take the sum of those two values and divide them by 2 (aka just take the average). The accurate way of doing this is to take their total field goals made and divide by total field goals attempted. But there is another issue because all stats are per game stats. So, for example I don't have the total field goals attempted by a player, I have how many field goals they attempted per game. So the final calculation would look like (field goals made * games played)/(field goal attempts*games played). The only issue is this results in a slight difference to the actual percentages due to the rounding done in the data that I mined. There is really no fix for this with the particular data set as it would require getting more complete information from another source. This was an issue for a lot of the columns but my visualizations were only focused on three points, field goals, free throws, and effective field goal percentages, so these were the only columns I fixed for this.

Making my reports was a lot more difficult than I expected it to be. I really struggled going from what I consider to be a Tableau expert to a complete Power BI novice. I struggled with even some of the basic concepts like filtering and more. The more I worked through the project and practiced, the better I got. My goal was to start with a landing page, and then create visualizations/reports for both the team stats overall and then the player stats. I believe I was able to accomplish this to some degree, but I am excited to see how my skills will improve with time.

My landing/home page was very simple. I set up a text box for the title, and then incorporated the NBA logo as an image. I then added navigator buttons. These will be used to go from report to report. Because I wanted my reports to be easily navigated, I added back buttons to every report to make the process easier.

My first dashboard was just two simple text tables. I wanted to present my user with the option to see detailed statistics on both the teams records as well as the individual player stat lines. I added two slicers and edited their interactions so they would not affect each other. The result was a Team Standings table sliced with the conference/division hierarchy and an Individual Player Statistics table

sliced by teams. The user can now define their own level of detail and see a more in-depth look at how teams and players performed in the season.

My next two reports were team based. I want to visualize team wins per conference, so I did this with a tree map. I then wanted to show the best defenses and offenses. I visualized this with a column chart and a bar chart. For these two visualizations I set each bar/block to each team's official hex color value.

My final report was the most fun. I wanted to delve a bit deeper into the percentages. So, I first created 3 cards showing the league average in three point, field goal, and effective field goal shooting percentages. Then below each card, I made a visualization for that percentage. For three point shooting, I wanted to see how height effected 3 point shooting, so I made a scatter plot. For field goal shooting, I wanted to see the effect of age, so I made a line chart. And then for effective field goal percentage I wanted to see the effect of position so I went back to the column chart.

Overall, like I said previously, this was a difficult project for me. It is hard to go from being an expert in one program to being a complete beginner at another. The approach I will take moving forward is this project will serve as a baseline for me. Much as my first Intro to DA tableau project did for me, this project can be saved and will show how much I can grow my skills. The dataset provided some problems that I want to be able to solve in the future. I still have not effectively figured out how to tackle the issue with all my statistics being per game stats. The fields I created still were not giving the results I wanted. Figuring out the answer is just going to require more experience with Power BI. I know how to figure out the answer in general, but I need to learn how to apply that specifically to this program. It would be great to also grow this dataset to include multiple years. The game of basketball has been changing immensely since I first became a fan in the late 80s/early 90s and I would like to be able to visualize this. The best part though, is there are so many tasks that Power BI makes easy. The integration with excel and the fact that a lot of what is done is just an extension of that program makes using it so easy and powerful. While learning more about Power BI, I'm not just honing my skills in this application but also in Excel. Both skills are highly valued in the field, so the benefit is immense.