The best part about working with CSV files and pandas in an assignment like this is that you get to repeatedly learn the syntax and understand each component of the commands as you go. I am also especially interested in college basketball, so being able to complete an assignment while seeing the practical applications of the work is great. This assignment primarily taught me how to create new tables using filtered data from a completely different dataframe. Being able to sift through a data frame and extract the relevant information to create a more relevant chunk of data to analyze is incredibly useful, and proves why pandas is such a vital tool for data scientists.

The most challenging part was certainly learning how to create a new data frame based on the conditionals of an older data frame. I had no idea that this was possible in just one line of code. The conditional statement was so simple, yet it was not what I expected. I expected to have to specify which columns to keep, or a specification of how to order the new dataframe. After a few minutes of research, it was easy to find out what to do, but it was a challenge I didn't expect to see.

Sports data is a great example of the power of data analysis tools like pandas to understand a group of data better. Having tons of data is great, but can become far too large to get valuable insights from without software to help you unwrap it all. There are thousands of college basketball players in the US, but these databases can get much larger when covering different data, such as the US census when there may be hundreds of millions of points of data. Having software that can parse through for you to find patterns, outliers, correlations, etc is vital for efficiency and getting everything you can out of the data.