Doug Marcum
DSC 530 – Final Project Summary
2/29/2020

Statistical/Hypothetical Question
For this term project, an evaluation of defensive importance within the National Basketball Association (NBA) will be conducted, to determine if the adage of 'Defense Wins Championships!' remains accurate in today's game. Since sports have been played and a score kept, there have been coaches, sideline parents, armchair experts, and others declaring the importance of defense. While defense is an integral part of most team-oriented sports, is it the most important aspect of winning? Does the best defense have a better chance at winning a championship in comparison to the best offense?

Outcome of your EDA
The outcome of the EDA is interesting. The idea that teams are either offensively or defensively focused is a bit misleading. From the initial analysis, it was clear that most playoff and championship caliber teams consistently have higher Defensive and Offensive Ratings (DRtg, ORtg). Additionally, a unison or balance is often found in teams that are delivering results on both ends of the court.

After discovering this, there was still enough information and data to explore that helped move closer to answering the overall question. Two models for each prediction were created to focus on defense and offense respectively.

First, looking at teams that make the playoffs, the models were built playoffs are the dependent variable and either defensive or offensive ratings as the explanatory. With this, the model was able to accurately predict playoff teams based on those ratings with 75.67% (D) and 73.67% (O) respectively.

In predicting a champion, the model considered additional explanatory variables (pace of play and simple rating system) in order to remove some of the unknown. From this the model returned a rather high accuracy of 96.67% for the defensive model and 96.33% for the offensive model. To test the predictions, the variables used in each model looked as so:

<u>Defense Prediction</u>
```
columns_d = ['DRtg', 'SRS', 'Pace']
def_champ = pd.DataFrame([[Champs.DRtg.min(), Champs.SRS.max(), Champs.Pace.min()]],
                columns=columns_d)
def_champ_percent_to_win = results_dc.predict(def_champ) * 100
```

<u>Offense Prediction</u>
```
columns_o = ['ORtg', 'SRS', 'Pace',]
off_champ = pd.DataFrame([[Champs.ORtg.max(), Champs.SRS.max(), Champs.Pace.max()]],
                columns=columns_o)
off_champ_percent_to_win = results_oc.predict(off_champ) * 100
```

This allowed us to take the best ratings for offense and defense from championship teams over the past ten seasons to create 'super' teams dominating one side of the court. By doing this the model predicted that the defense focused team would have a 70.2% chance at capturing a championship, and the offense focused team would have a 68.1% chance of capturing a championship.

While these finds are not definitive, it can be implied that yes, defense does win championships.

What do you feel was missed during the analysis? Were there any variables you felt could have helped in the analysis? Were there any assumptions made you felt were incorrect?
In terms of looking at the project to see if there were missed steps during analysis, incorrect assumptions, or additional variables that could have benefitted the analysis, the answer is yes. As with any model, more data and input could help strengthen the predictive power. Additionally, two main assumptions made (offense is dominating modern game play and rule changes effecting game play over time) could have been worked through before being narrowed down. The project does a good job of understanding the past ten seasons, but in retrospect, it would be interesting to trend the data to see if there are changes over time. Additionally, if changes are found, dig deeper to discover the fundamental cause of those changes.

What challenges did you face, what did you not fully understand?
Luckily, there were not many challenges faced along the way that would fall outside of the scope of any project. By working with sports data, the data was very clean from the beginning and thus allowed for more time on understanding what the data contained. There was worry when bringing the entire project together that areas of unknown would cloud and cause missteps to be taken, but that does not seem to be the case. Certainly there are spots that can be improved upon and maybe errors made that I do not realize I have made, but all and all, focusing on a topic of interest helped in my understanding of each criteria needed to complete the project.