**College Basketball Analysis**

Every year March Madness is one of the most talked about sporting events of the year and millions of people make brackets trying to predict who will win each of the games. So, I thought it would be a perfect time to look deeper into the numbers and see what leads a team to victory. We will see what basketball stats correlate with winning the most, and by the end of this I should be able to compare my results with the games that are happening and see if the outputs I got would help make a better bracket.

To get these outputs I will be using data from Data Camps website. This data has information about all regular season and tournament games from 2003 to 2018. Each games have most of the basic basketball stats like team score, assists, rebounds, turnovers, fouls, and more of the common basketball stats. I will use these to try and make predictions based of the team’s stats and compare how basketball is played today versus back in 2003 and see if there are a lot of differences.

The first thing I looked at after organizing and cleaning up the data was to try and find which variables would make the best model in correlation to winning games. I found that these values make up the model of best fit to this data those stats are score, fouls, field goals attempted, defensive rebounds, steals, turnovers, offensive rebounds, free-throws attempted, and blocks the order listed is the order of importance to the model. So, I decided to start looking more into those values that have a bigger impact on the model, the first one I looked into was personal fouls I choose this one first because it has the second biggest impact on the model behind team score. This graph on the right shows the win percentage at each number of Chart, scatter chart

Description automatically generatedpersonal fouls in a game, so teams with 5 fouls a game have over an 80% chance to win the game but teams with 30 fouls only have a 20% chance to win. This graph isn’t too surprising to me the more you foul the other team the more points they will get, and the goal of basketball is to get the most points. When I looked at the next most important value in the model (field goals attempted) I was Chart, scatter chart

Description automatically generatedsurprised with the results. the data shows that you have a higher chance to win the game when you attempt less shots, I would have thought that the more shots you attempt would lead to more made shots which would lead to more wins, but this graph shows that teams with less shots had a better chance to win.

The next thing I investigated was to see how much the game has changes over the years, the graph from the paragraph above shows you that teams are taking a lot more shots in recent year and if we are going to be predicting which teams are going to win right now the data from back in 2003 could be hurting out models if the game has changed a lot over the years. I made a model that just look the last three years to see if it is different than the overall model. Both models are similar there are a couple small differences in which variables have more correlation to winning but the difference is small. I am surprised that the difference isn’t Chart, line chart

Description automatically generatedbigger with how the game has changed and example of this is how many more three pointers are attempted now verse 15 years ago. This graph on the right shows three-point attempts per game throughout the years, I would have thought since the three-point shot is so much bigger in today’s game it would have had more of a correlation to winning games.

**To DO:**

* **Compare regular season results to tournament results**
* **Make predictions based off these results**
* **Compare my predictions to what actually happened**
* **Conclusion**

**Publication:**

<https://journalofsportsanalytics.com/>