For my final project, I decided to take a deeper look into NCAA Men’s Basketball to see what factors led to teams being successful. I also wanted to analyze which conferences perform the best. As a Minnesota Golden Gopher fan, I am a bit biased towards the Big 10 and I wanted to test to see if they were better than the competition or at least were as successful as the other major conferences.

My dataset consisted of NCAA regular season results by game from 1996-2017. I trimmed this down to consist of only the Big10, Big12, Big East, ACC, SEC, and PAC12 conferences and their opponents. This would allow me to more easily work with the data and also was necessary as the data I was working with was in a online data table that would only allow you to export 16,000 rows at a time. This was the first hurdle I needed to jump over, but I think I still got what I needed, and these were the conferences I wanted to focus on anyways, so it all worked out.

Based on my EDA, the ACC had the highest win percentage (wins divided by total games), so I was technically wrong about the Big10 being the best or most “successful” conference. But, after running a hypothesis test on likelihood of winning, there was no statistical difference between the Big10 and all other conferences when it came to winning games. From that standpoint, I felt confident that my initial assertion was at least warranted. Some other takeaways were that attendance didn’t have a very good fit predicting wins. I was a little surprised because I would have thought having a larger attendance would have maybe given a team more of an advantage.

Some challenges I faced were more based on my dataset. There were large parts incomplete or missing such as attendance date of the game missing from the years 1996-00. I had to go in and replace the missing values with the average attendance by team. In hindsight, I’m not sure this data lent itself well to all of the different steps of analysis outlined in the project, but I think that is also a learning experience that was good for me to have. There rarely is a perfect dataset just existing out there that will have no missing values and will fit all of the analysis you want to run. Learning how to adjust and adapt to the data you have been given is a big takeaway that I’ve had from this course.

Another hurdle I had was initially I wanted to do an analysis on the NCAA tournament results, but I had such a tough time with the data that it never came to fruition. This was a bummer for me as I really wanted to get my hands on that data and run an analysis on it, but I think I had my fill with the regular season data that it all kind of worked itself out.

Overall, I really enjoyed the course and this final project. I’m glad it challenged me in the ways that it did, as I find adversity to be one of the greatest teachers. Thanks for a great semester!