

Why did you choose the dataset you used for your EDA? What about the context of the data interested you? What relationships did you expect to find among your variables of interest?

I chose to do the ATP Tennis Tour dataset because I am familiar with the ATP Tour and have watched some of the games. I also used to play tennis so I have familiarity with the topic. Seeing how different factors in tennis such as opponent rank, court surface, etc can affect match results was very interesting to me. I expected to see that lower (better) ranked ATP players would win more often, and that court surface did not affect match time much.

What challenges did you face while completing the assignment, and how did you overcome those challenges? What do you plan to do differently next time as a result? List at least one specific thing you learned about yourself as a data science student as a result.

One of the challenges I faced during the project were values being missing which affected the graphs. To overcome this, I omitted the NA values which worked. In the future to avoid this, I will clean my data properly before knitting the file. One specific thing I learned is that I need to be more careful when cleaning data because any small mistake can change the results drastically.

Describe at least one way your personal background influenced how you analyzed this data or presented your analysis. What might someone else have done differently if analyzing the same dataset?

My background in playing/watching tennis influenced how I knew which variables should be the ones to analyze. I looked at player rank, court surface, and number of aces because I know that those can affect match outcomes. If someone without a tennis background analyzed this they may have analyzed variables that would not matter as much.