Question 1:

I chose the soccer dataset because soccer is a game I have played and watched my entire life, so I wanted to analyze something that genuinely interests me. I expected to find that squad experience and rotation patterns would have clear relationships with team performance, assuming that more experienced teams would perform better. As a huge CR7 fan, seeing his longevity made me wonder if it was his talent or his experience that makes him so good, which inspired me to investigate whether squad age correlates with success.

Question 2:

The biggest challenge was cleaning the data, which forced me to look online and use external resources to figure out the right functions and methods, as well as learning more about r in general with new functions like casewhen and ifelse. I also realized partway through that I needed an entirely new dataset to properly answer my research questions, which set me back significantly. Next time, I plan to make sure I fully understand the data I am using and what exactly I need before I start working. I learned that I need a set plan before I start anything as a data science student, otherwise I tend to falter and waste time.

Question 3:

My personal background as a longtime soccer player and fan influenced me to focus on squad age and rotation because I have always been curious about how experience impacts performance at the highest level. As someone who started out not too great at soccer, I always wondered if experience could ever beat just raw talent, which shaped my research questions, while someone without a soccer background might have focused more on purely statistical metrics like goal differences or possession statistics. The way I approached this project was in genuine interest of how important other factors besides talent would matter, as I figured it would motivate me to keep playing as well.