Josh Miller
December 14, 2020

Nashville SC Technical Analytics Internship Data Project

GitHub Repository: Miller_Submission

In order to create a CSV file with "a timestamped list of on-ball pressures (defending team player within 5 yards of the player in possession of the ball and moving towards the ball) from your last match," I made the following assumptions:

- "Pressure" is *loosely* defined
    - Pressure can only be applied when the following two conditions are met:
        - The game is "live" (i.e., open play, not during set piece setup, etc.)
        - The ball is within the playing field (i.e., ball's xyz coordinates are non-null)

If the above conditions are met, the rest of the coding logic is as follows:

- The "lastTouch" field is the sole determinant of who has possession of the ball
    - From here, determine which player on the team with ball possession is closest
    - Next, go through every player from the defending team and calculate their distance from the ball at a given moment/frame
        - If within the target distance/radius (i.e., 5 yards for this project), log relevant details into the output file

There are several shortcomings and flaws to the approach I took. Given the preconditions I used, the data that's output into the CSV file contains cases where non-pressure situations that coincidentally meet the other conditions as well. For example, if a CB makes a pass to his CDM and the opposition's CAM is walking nearby but not actively pressing, this would likely still be logged. If a CF takes a long-range effort and his shot whizzes narrowly past a defender, my code might still log this case as it meets the conditions. Additionally, if that same CF's teammate is near that same defender, the log would indicate that for those moments or frames where the ball is within the target radius distance, it is that teammate and not the player who struck the ball who is "in possession" of the ball since he is closest.

I think given more time, I would have been able to discuss in more depth stricter definitions for what constitutes pressure. This would have helped me create additional conditionals to help filter through frames better. I think that the data provided was sufficient; however, if there's a way to quantify a player's "intent to pressure" (more than just physical proximity) that could be even better. One suggestion is using velocity instead of just speed as a player's direction of movement can inform analysts of the player's "intent to pressure."

The biggest challenge was really just breaking the problem down into smaller pieces and then figuring out where they fit and how they were supposed to interact with one another. Other than that, it was just a matter of coding it up.