Nashville SC Technical Analytics Internship Data Project

Report

Prateek Senapati

prateek1210s@gmail.com

My process for extracting the pressures

These are the important considerations in my process for extracting the pressures:

- The ball location is null if the ball is not visible in the tracking area. I have considered only those frames in which the ball location is not null as I have used the ball location to find the home team players who are "on the ball".
- A player is "on the ball" only when he has made the last touch, and with 25 frames captured per second, it's very unlikely to miss out on a player's first touch on receiving the ball. Therefore, another condition that I have used to find an on-ball home team player is to set the "last touch" to "home".
- The tracking data contains a column titled "live". It is set to true if the ball is "in-play" and to false otherwise. I have considered only those frames in which the ball is "in-play".
- For finding whether a home team player is "on the ball" or not, I have used Euclidean distance for an estimate. I have considered a home team player to be "on the ball" if the Euclidean distance between the ball location and his location is less than 1 yard. The z-axis coordinate value for all the players in all the frames is set to 0. As a result, when I use the z-axis coordinate value (which is not 0 for the ball location) for measuring the distance between a player and the ball, a player will not be considered "on the ball" if he's controlling the ball with his head or upper body - as the distance might exceed 1 yard. Therefore, I have not considered the z-axis coordinate values while measuring the distance between the player and the ball. This can be discussed with the team staff, and it can be decided if we need to include the z-axis coordinate values as well.

Challenges and Issues

Here are a few challenges and issues that I faced while working on this data project:

- I have used Euclidean distance to figure out which player is "on the ball" and whether or not any player is "on the ball" in each frame. I believe there are better ways to do this. Any information regarding the "on the ball" player in each frame would have definitely assisted me in making better decisions.
- There's a 1-2 second(s) time lag between the game clock time in the match video and the game clock value (in min:sec format) in the resulting CSV. There might be an actual time lag between the match video game clock and the game clock values in the tracking data, or the new game clock values calculated by me might not have been accurate enough.
- There are instances (which I think are on-ball pressure instances) which didn't make it to the CSV file. For example, Nashville SC's goal #1 and goal #3 were not considered as on-ball pressure instances. Possible reasons include: non-visibility of the ball in the tracking area, no opposition player within 5 yards of on-ball player (to the naked eye, it might seem that opposition players are within 5 yards of him), or the instance might not have been captured by any frame (which is highly unlikely with 25 frames being captured per second).