

Nashville Soccer Internship Project

<https://github.com/jonkaplan18/nashvillesoccer>

Goal: Extract the list of timestamps of on-ball pressures (defending team player within 5 yards of the player in possession of the ball and moving towards the ball) and export to CSV

1) Offensive and defensive players needed to be assigned

- a. Viewing the video allowed me to see which side of the field each team started on for period 1 and 2. If the period was odd, Atlanta was on the left side of the field, otherwise, Nashville. A player was an offensive player if he was on the opponent's side of the midfield (0,0).

2) Ball possession needed to be defined

- a. Viewing the video allowed me to get a sense of the flow of possessions and how the ball was often within 2 yards away from a player with the ball in possession.
- b. Additionally, I used the *lastTouch* column. If the *lastTouch* was also an offensive player, then this was the second indicator that signified possession.
- c. One challenge was figuring out how to handle the z coordinate in the xyz of the ball location and whether to use the height of the ball as an indicator of ball possession. A concern was to not include accidental on-ball pressures of times when the ball was too high in the air and unreachable for players, despite having a player xy that matched up with the ball. This was relevant mainly for headers. A judgment was made to only include heights below 10 feet. Since player heights were not given, an average height of 6 feet was used based off some of the roster heights. The extra 4 feet gave room for a player within jumping distance of heading a ball which would signify possession. A column with each player's actual height would have made this more accurate.

3) Determine if defensive player was heading towards the ball

- a. Since only a speed was given for the ball and not the direction, it was hard to figure out if a player was moving towards the ball. A player's distance from the ball was determined with the distance formula and looking to see if a player's distance from the ball at that moment got shorter in the next frame compared to the previous. If it did, then this meant they were headed towards the direction of the ball and could be included in on-ball pressure

Cleaning notes:

- Data was in meters and was converted to yards.
- For the game clock, the time was changed from seconds to minutes:seconds in order to sync with the video better since the game clock was in this format. This allowed a way to double check for false positives and missed occurrences of ball pressure visually by fact

checking with the excel sheet while watching. The dataset starts at 0 again once period 2 starts, and the time starts at 0 again for the second half. I would refine this later to more accurately match the game clock in the video.

- The final CSV has the columns *period*, *gameClock*, *wallClock*, *team*, *playerName*, *playerNumber*, *distanceFromBall*, and *changeInDistanceFromBall*. The change in distance column shows how a defender is getting closer to the ball and the distance from the ball column shows how the defender is within 5 yards.