In a brief writeup, 500 words or less, please explain your process for extracting the fullback overlaps, quantifying the effectiveness of those overlaps, any missing information that would have assisted you, and any challenges you faced. Please do not use any data other than the provided files.

Basically, I leverage the jupyter notebook to leverage my analysis, which I import both csv file and json file into the jupyter notebook, and filter the data with only home players in Json file then join two tables where player_number within json file and number in csv file. Therefore I can see the position column within the new dataframe, which makes the process easier for filtering both left and right back players.

The fullback overlap was develop for check the data through a left or right back (fullback) runs around a teammate who has the ball and moves ahead of them. The code does this by looping through the position data of side fullbacks, identified by player IDs, and checking if they meet the conditions for an overlap in each frame (a snapshot of player positions).

To determine if an overlap occurred, the code checks if the fullback is on the same side of the field as the teammate, ahead of the teammate, and ahead of the ball, which the ball is in possession of team. The function loops through each fullback player and counts the number of overlaps they made throughout the game. The final result is a dictionary that shows the number of overlaps for each fullback player. which I compare these counts to evaluate the effectiveness of the Fullback Overlaps, the highest number of counts refer to the high effectiveness of fullback overlaps.

The code is also optimized to minimize the number of calculations needed, reducing the time it takes to process a large dataset.

One of the major challenges encountered during the project was that the available data did not provide any indication of whether a player was in possession of the ball. Even though the possession could be referred through last touch team id, it does not accurately deliver the correct answer. Without exact data on possession, it was challenging to accurately measure the effectiveness of the fullback overlap.

Furthermore, measuring the effectiveness of the fullback overlap through available data was limited by the absence of crucial information. For instance, the data did not capture whether a player made an assist, delivered a high-quality cross, or scored a goal. These are essential factors in analyzing the effectiveness of overlap, and their absence limited the analysis process.

Besides, two periods have different positions on the field. After watching the game video, I saw the home team was attacking from right to left in the first half and reversed the position in the second half. Align with coordination system, the touch line is in ascending order from left to right, which refer to that home team’s position in reverse order in the first half. So I set the if function before the count_overlaps function to ensure the fullback overlap calculated correctly.