Q. Using skillcorner broadcast tracking data, create a model that identifies corner kick situations and use it to deliver analytical insights to the coaching staff **in the context of opposition analysis**.

Objective:

When it comes to corner kick situations in terms of opposition analysis what are some things that would interest the coaching staff? Two things come to my mind, first, how does the opposition defend their corners. Second, how do they attack them (take the corners themselves). Since the skillcorner data doesn't have any information about how goals were scored in these games or who scored them, we would not be able to provide those insights to the coaching staff.

Identifying/Extracting Corner situations:

The first method that I tried was to code in the exact rule of corners, i.e., whenever the ball goes outside the goal line, having been last touched by the defending team, the attacking team gets a corner. There were 2 issues with this approach. Firstly, there were instances when the ball wasn't "seen" going past the line (for eg., Game 2068, 1:31), so there was no record of the ball position being beyond the pitch dimensions length wise. The second issue was that it was hard to figure out if the ball hit the defending team before going out of the pitch. The "team in possession flag" or the group in possession in the structured_data files did not update during deflections off defenders, as those sequences were perhaps too quick for the object detectors.

Proposed Solution:

Since we couldn't clearly figure out the event/action of the corner being awarded, we move on to the event/action of the corner being taken. The act of a corner being taken is when the ball is placed on the corner arc (1 meter radius from the corner of the pitch). In terms of code, this equates to looking at all instances where the ball is at a 1 meter distance from the corner & being possessed by the attacking team.

Strictly applying this logic gives us only a limited number of corner situations, as during broadcasting it is very usual to cut away from the game scene when there is a stoppage in play. So there were quite a few corner instances where the ball wasn't seen on the corner arc, either due to camera cuts or due to other occlusions. I opted to slightly overcome this issue by using a tolerance value while computing this distance from the corner. The extra tolerance value did lead to getting a few extra corners, but also led to getting other non-corner situations like throw-ins or clearances, when the ball was seen at a 1 meter (+- tolerance) near the corner arc.

To get rid of non-corner situations I tried to use the ball velocity in those situations, to see if the ball was placed still near the corner arc, as is the case during corners. But since the velocity calculations are based on how the ball locations change per frame & not the actual ball velocity, there was no clear signal of the ball being still during corner situations.

Now till this point we have shortlisted all instances where the ball was found close to the corner arc. Next we try to maximize the probability of this instance being a corner situation. I opted to do that by looking at frames up to 10 seconds from when the ball was found close to the corner and see if a minimum of 4 (parameter) attacking players were found in any of those 10 seconds. For ease of analysis, I have saved all such 10 second frames for the home and away teams in separate csv files for each game.

Analysis

To reiterate, our goal was to give insight to the coaching staff about how the opponent sets up for their offensive as well as defensive corners. In our case we have data from 9 different matches, with a maximum of 2 games per team. To increase the sample size of our analysis, I chose to neglect the actual teams in our dataset and consider that our **"OPPONENT"** played in all of those 9 games. To distribute the corners between attacking and defending, I assumed that all the corners that the actual home team gets in any game, were the **opponent's defensive corners**, and all the corners that the away team gets were the **opponent's offensive corners**. This distribution could be done randomly as well.

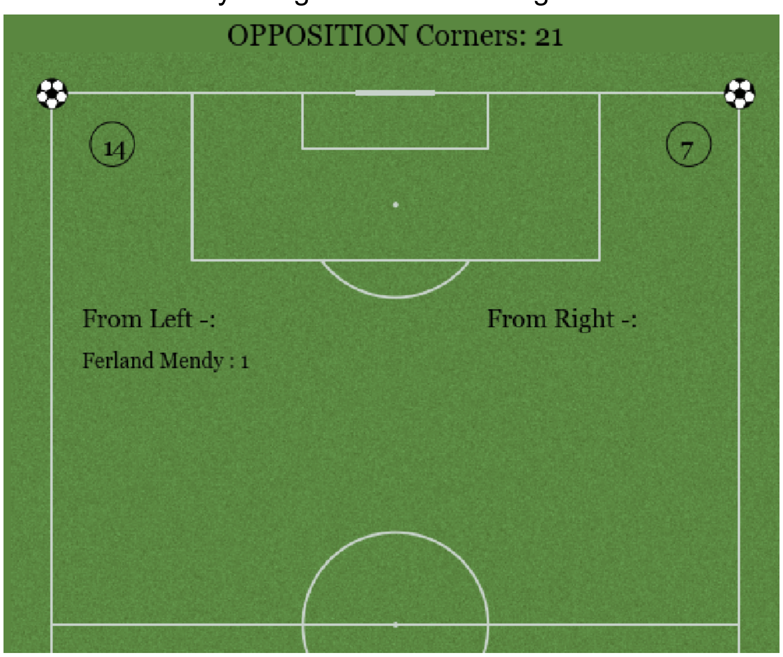
In this part, we go through all our saved frames from the csv files and find the best frame for our analysis, i.e., the frame with the **most number of opposition players** found. The idea is that the frame with the most number of opposition players, be it during an attacking corner or a defensive one, will give the most information about how they set up.

We stack up all of these "best frames" across all games and visualize different attacking & defensive zones of the **opponent** using Kmeans clustering.

Visualizations/Insights:

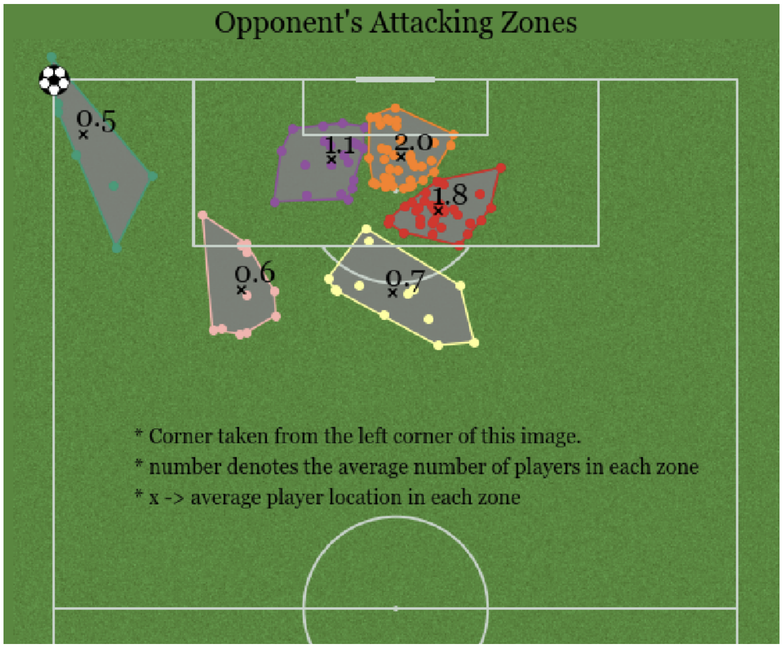
1. **Aggregated corner statistics:**

The code submitted is able to detect the ball location and aggregate the number of corners that were taken from each side and who the takers of the corners were. Due to the limitation of the data, the corner takers were not detected across all corner frames that were selected, but it is something that can be overcome by using events data alongside.



2. Zones of attack & defense:

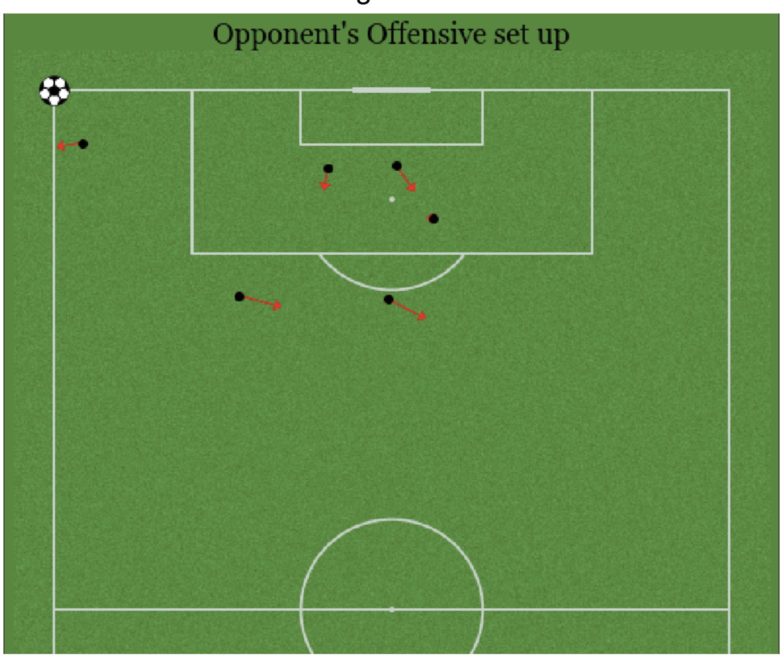
Here we use clustering to see different zones when the **opponent** sets up for attacking corners. From a footballing point of view, we can view these clusters as -: roughly 1 player available for a short corner, 1 player at the near post, 2 players close to the goal, and roughly 2 players far from the goal/runners from deep.



A similar visualization for opponent's defensive zones is shown & explained via a footballing point of view in the presentation created for the coaching staff.

3. Movement patterns:

For each of the above attacking zones we also find the average velocity and the direction of velocity based on all the selected corner frames. This helps us get the knowledge of how the opponent players move while a corner is being taken.



The above visualization is just a sample of a visualization that can be developed and is not something that should be used to take inference from the given data at this point, as the method to calculate these average movement patterns just uses a single frame of reference for any given corner situation.

Also, currently the visualization just shows the center locations of the cluster zones from the previous cluster visualization. In future the idea can be developed and sampling techniques can be used to sample a specific number of players and thus their positions from each zone, thus representing something which is closer to an average offensive/defensive setup.

Pros of this approach:

1) Using our approach we were able to detect 29 (out of 36 detected) actual corner situations and develop visualizations which show how the Opponent sets up on an average during their attacking & defensive corners. The choice to consider that all the 9 games belonged to the "OPPOSITION" team also helped us in getting a bigger sample size for our analysis.

2) The existing analysis code also shows the number of corners taken from either direction (something that is not readily available across well known platforms), and the corner takers.

Cons of this approach:

1) The logic still isn't completely effective in removing non-corner situations from the shortlisted situations. From the total of 36 shortlisted corner situations, 7 of those situations were either not a corner or had an overlap with an existing detected corner.

2) All the visualizations and analysis uses the corner frame with the most number of Opponent's players found in them, to maximize the amount of information that can be gained from the data. This method when used to find the average locations and specially average player velocities is not the right approach, as lots of information is not taken into consideration.

Future tasks - given more data:

1) Trajectory prediction: If given the complete trajectory of the ball, from the point the corner was taken, till it reaches any other player (in the box/outside), we can use that information to cluster the corners into inswinging/outswinging or taken short categories. Another possible solution could be to train a neural network to learn the trajectories, if given the true labels of all the corner situations.

2) We can sync the event data with the existing skillcorner data and finetune our corner detection by using timestamps from the event data of when a corner kick was taken. This will also allow us to integrate shots into our analysis and help create similar visuals as this, by statsbomb.

3) If we have the tracking data available for all players across all frames, we can potentially find finer details in the data like man marking or zonal marking schemes used during corners. “*Players marking man-to-man will typically start in close proximity to an opponent and will often cover a significant distance during the corner kick as they track the opposing player. Zonally marking players, on the other hand, tend to be more stationary.” (Routine Inspection: A Playbook for Corner Kicks - Laurie Shaw)*

Future tasks - given more time:

1) Store & access the data in a more efficient manner, using a database.

2) The linear sum assignment problem (initiated in the analysis code) is one way to potentially dive deeper into analyzing plays where switching of markers occurs. The minimum distance between attacking & defensive players can be used as a feature while developing models for corner situations.