Evaluating players: Chelsea midfielders

Summer 2018 scouting report to analyse defensive metrics for central midfielders

**Objective**

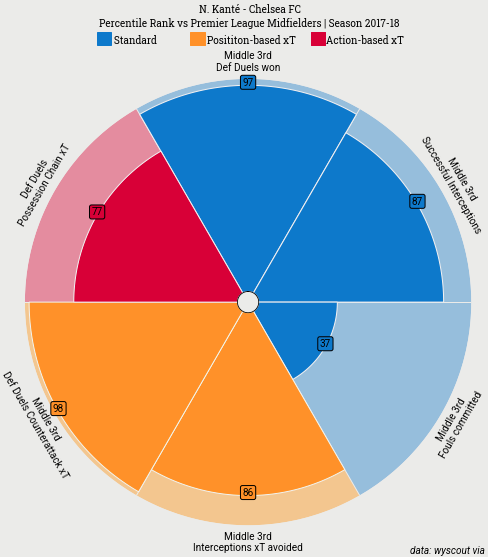
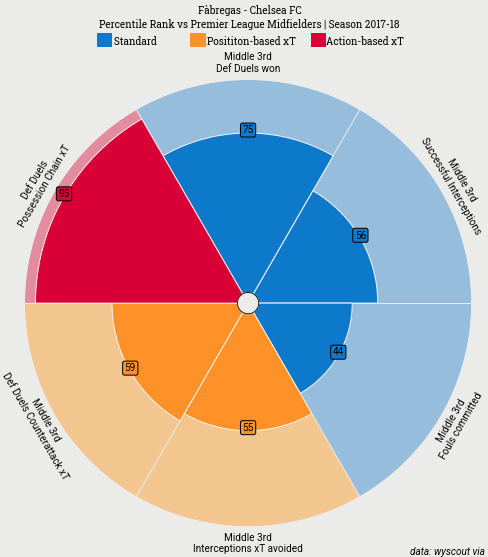
We aim to analyse the current (looking at statistics from the 2017-2018 season) Chelsea midfielders for their contribution or the lack of it to the defensive function of the team. To give some context, Kante had another stellar season as double pivot in the middle of the pitch. On the contrary, Fabregas and Bakayoko couldn’t form a good partnership with Kante leading to a string of poor defensive displays where the Blues were far too easy to play through the midfield block. Chelsea made the call to replace Antonio Conte with Maurizio Sarri who, with Napoli, had managed to make a name for possession and high pressing football. With this study, we want to analyse the underlying stats for Serie A midfielders using some of the defensive metrics and gauge a suitable player to join Sarri at Chelsea this summer.

**Defensive metrics**

We have implemented the following 6 defensive metrics which are then visualized for each of the current midfielders at Chelsea to give us a comparative overview of player’s respective strengths and weaknesses solely from a defensive perspective.

1. Middle 3rd Def Duels Won: This is a standard metric and we simple aggregate the volume of successfully won defensive duels in the middle 3rd by a player across all the matches played in the season. A defensive duel is defined as “when a player attempts to dispossess an opposition player to stop an attack progressing”.
2. Middle 3rd successful Interceptions: Another standard metric where we count the aggregated number of successful interceptions (where an attempted opposition pass is deemed unsuccessful owing to a touch by the current team player) in the middle 3rd.
3. Middle 3rd Fouls committed: This too is a standard metric where we specifically look at the negative aspect of defending in fouls and aggregate the volume of fouls committed in the middle 3rd to allow us to gauge the defensive discipline of a player.
4. Middle 3rd Interceptions xT avoided: xT value avoided by interceptions in the middle 3rd
5. Middle 3rd Def Duels Counterattack xT: xT value associated with potential counterattacks initiated from defensive duels won in the middle 3rd
6. Def Duels Possession Chain xT: xT associated with defensive duel actions in isolated possession chains. Xgboost model trained for Bundesliga possession chains and scored on Serie A midfielders. Filtered down to ‘Ground defending duel’ actions only as the model was skewed towards tall forwards

**Chelsea midfielders**

 ­­­­­­

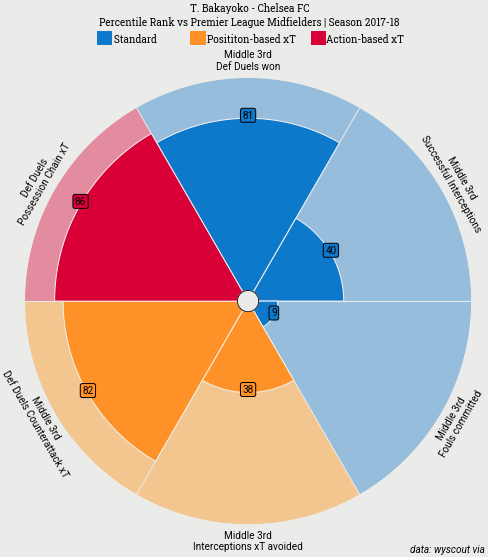
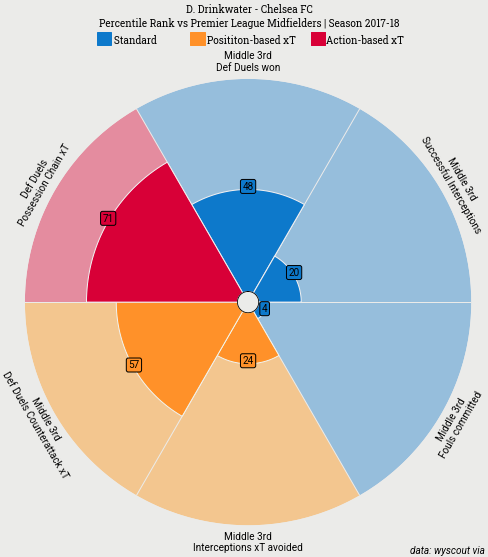
 

Figure: Chelsea midfielders for 2017-2018 season - percentile radar charts comparing the 5 metrics

**Data assumptions:**

Middle 3rd event’s locations filtering for metrics 1-5. “Midfielder” role name filtering to calculate percentile rankings for all metrics. Normalization for minutes played (all metrics presented as “per 90 minutes”), opposition possession % (metrics 1-5) and own team possession % (metric 6). As mentioned above we rely on a percentile-based ranking score and hence we use a negative axis for metric 3 (given that we hope to minimize the fouls committed metric).

**Seria A midfielders**

We narrow down our search on one metric to analyse players in Serie A: Action-based Expected Threat for successful defensive duels in isolated possession chains. Please refer the below table for the top 10 players for this metric in this league for their performance in the 2017-2018 season. Again, all the data assumptions mentioned above hold valid for the numbers in this table.

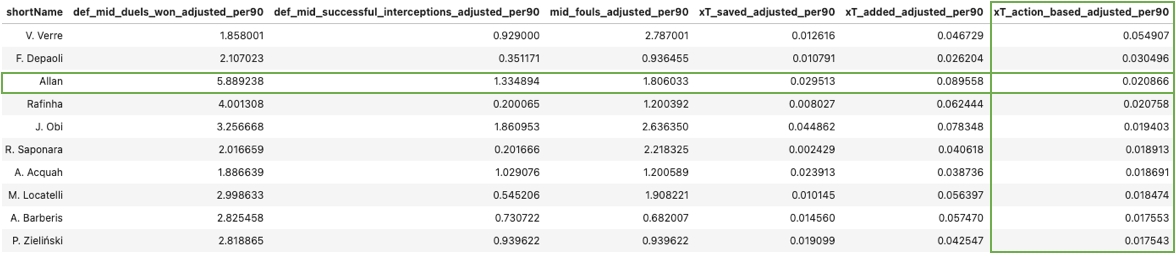


Figure: Top 10 Seria A midfielders for ‘Def Duels Possession Chain xT’ metric

Next, we look at the radar chart comparison between two Napoli midfielders who played together under Sarri in the 2017-2018 season.

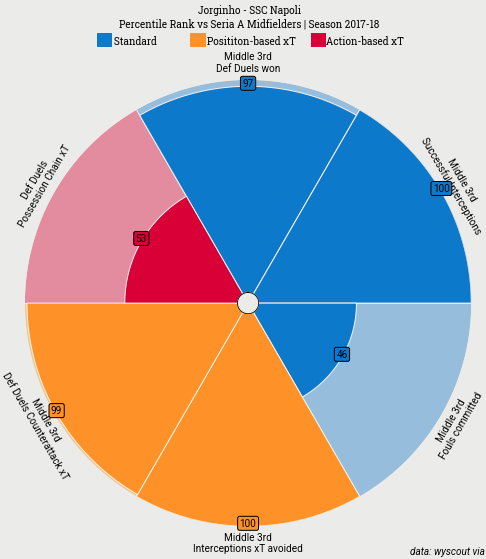
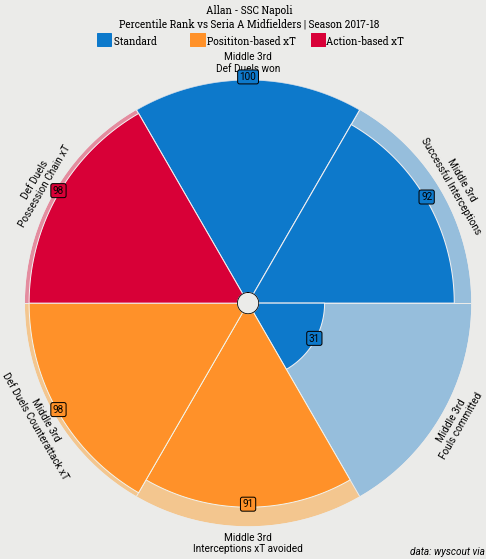
 

Figure: Jorginho vs Allan for SSC Napoli 2017-2018 season - percentile radar charts comparing the 5 metrics

**Search criteria:**

1. Age – 27 vs 26 and heading into the peak for both. Minutes played – 2949 vs 2722
2. Club & manager – Napoli teammates. Our aim is to find a player who contrasts Kante’s playing style and provides something extra in a collective pairing with him.
3. Playing style – Double pivot deep-lying playmakers who supply the team with quick exchanges on the wings in a Maurizio Sarri managed Napoli team that dominated over 60% of the possession in the recently completed season and hence would be a natural fit in transitioning over to Chelsea.
4. Contract tenure left at the club at that time – Signed by Sarri for €10m in 2015
5. Market value context as seen in the graph below - €25m compared to €40m

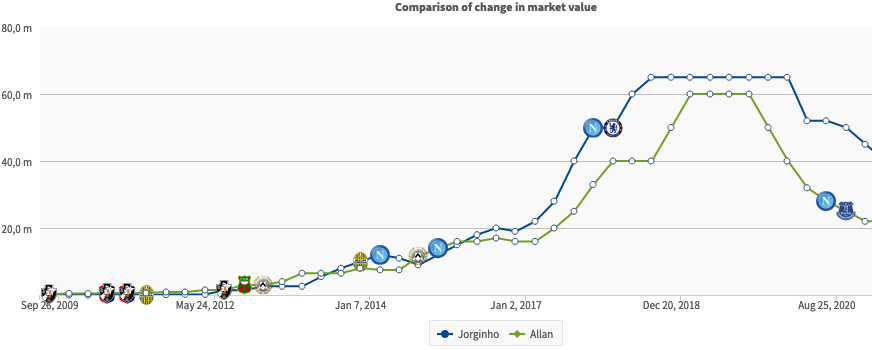


Figure: Jorginho vs Allan Transfermarkt comparison of market value over time

**Final recommendation:**

Common consensus with the group given the detailed investigation of the current Chelsea situation as was observed in the summer of 2018. Emphasis of this analysis on the search for a midfielder who could ably partner Kante at the heart of Chelsea’s midfield was discussed in brief which all of the group agreed upon. We discussed that this player would ideally proved defensive solidarity as well as creativity in spawning attacks from deep making him an all-round number 8/6 midfielder. Defenders vs midfielders variance in the underlying data and need to develop calibration across different metrics for a fair comparison. Additionally all of us had analysed players from different leagues which also suggestss the need for normalization for relative league strength. Owing to all of these complexities and citing the most relevance in terms of forming that strong parternship with Kante in the midfield, the group agreed that Allan was the most suitable midfielder amongst the top 10 players presented when sorted by the narrowed down metric. Also, the search criteria factors mentioned above deemed him to be differential compare to Jorginho.

**Code:** Create possession chains and train action-based xT model using this notebook: <https://github.com/yashkarle/soccer-ds-stats/blob/main/Projects/4UppsalaMMS/4_ValuingActions/possession_chains.ipynb>   
Create final metrics and plot radar charts using this notebook:   
<https://github.com/yashkarle/soccer-ds-stats/blob/main/Projects/4UppsalaMMS/3_StatsScouting/assignment_2_evaluating_players.ipynb>

**Data:** Wyscout data, supporting data dump and saved pre-trained models can be found here:<https://github.com/yashkarle/soccer-ds-stats/tree/main/Projects/4UppsalaMMS/data/Wyscout>

**Future work:**

Look into more details of the under the hood implementations of the xG model used for the action-based xT metric calculations. This will allow us to customize the feature set for midfielders looking specifically from defensive duels point of view. We can also utilize feature attribution techniques like Shapley index to understand how exactly the xG model relies on individual features to better tweak the same for our specific use case. Use of clustering techniques to add player roles and position-based groupings in the data containing the metrics when doing the final ranking.

**References:**

1. Wyscout data glossary: <https://dataglossary.wyscout.com/>
2. Peak age of players based on their position: <https://theathletic.com/2935360/2021/11/15/what-age-do-players-in-different-positions-peak/>
3. Transfermarkt Jorginho vs Allan player comparison: <https://www.transfermarkt.com/vergleich/spielervergleich/statistik/def/spieler/102017&126422>