



FOOTBALL ANALYTICS: VISUALIZING LEAGUE AND PLAYER PERFORMANCES

Emre ÇORBACI

1. Can we statistically demonstrate the importance of a player?

In this analysis, we explored various metrics to assess a player's impact on the field. We specifically focused on xG (Expected Goals), xA (Expected Assists), xGChain, and xGBuildup, which are advanced statistics that provide deeper insights into a player's contribution to their team's performance. The analysis was conducted using data from players across the **Top 5 European leagues**, ensuring a comprehensive understanding of performance at the highest level. To answer the question, we created several visualizations to illustrate the relationships between these metrics and to highlight the significance of xGChain and xGBuildup as potential indicators of player importance.

xG (Expected Goals)

xG stands for **Expected Goals**, which is a statistical metric that measures the quality of goal-scoring opportunities a player has during a match. It is based on various factors, such as shot location, angle, type of assist, and more. Rather than simply counting goals, **xG** gives an estimate of the probability that a shot will result in a goal, based on historical data.

How it is calculated: Each shot taken during a game is assigned an **xG value** between 0 and 1, where 1 represents a certainty of scoring and 0 represents an impossible chance. These values are derived from historical shot data, considering various aspects like distance, angle, and the situation in which the shot was taken.

xA (Expected Assists)

xA stands for **Expected Assists**, which measures the quality of a pass that leads to a goal-scoring opportunity. It estimates the likelihood that a pass will result in a goal based on factors such as pass type, assist length, and the location of both the passer and the receiver.

How it is calculated: Similar to **xG**, **xA** is determined by evaluating the probability of a pass leading to a goal. It takes into account factors like the position of the passer and the receiver, the type of pass (e.g., through ball, cross), and the shot taken by the player receiving the pass.

xGChain

xGChain is a metric that quantifies a player's involvement in the entire buildup of a goal, from their first contribution to the eventual shot or assist. It includes all passes, dribbles, or actions leading to a shot attempt, regardless of whether the player directly takes the shot or gets credited with the assist.

How it is calculated: **xGChain** is the sum of the **xG** values for all events that occurred during a goal-scoring sequence in which the player was involved. This metric highlights the player's contribution to the overall play and can reflect their importance in offensive buildup, even if they don't directly score or assist.

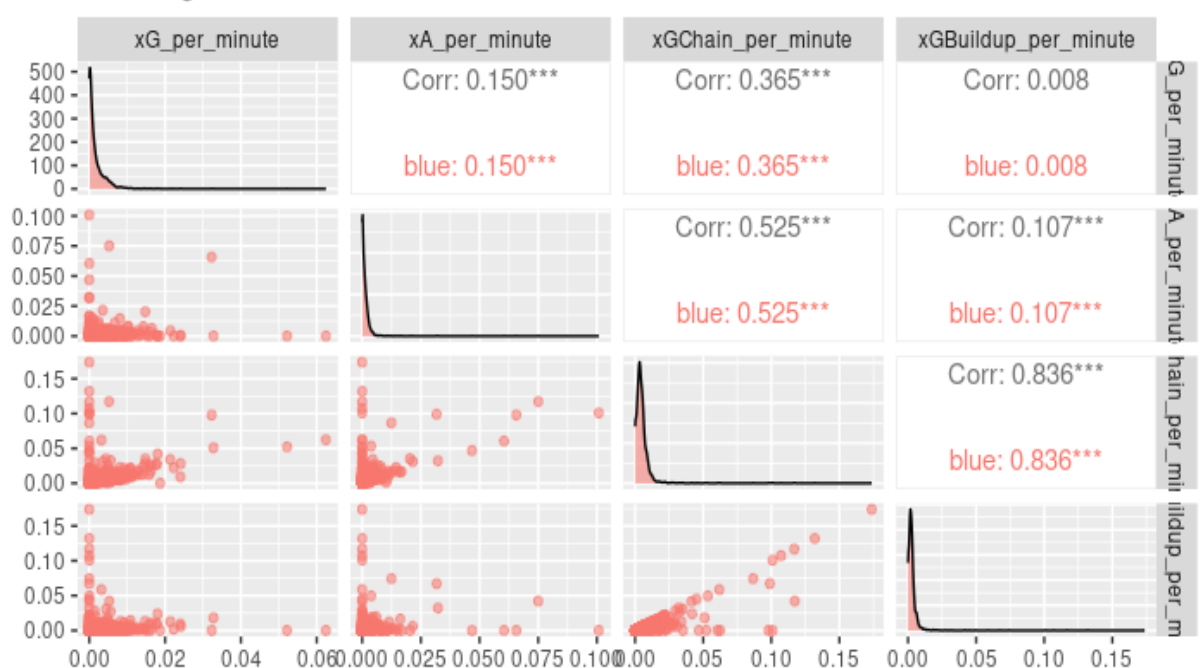
xGBuildup

xGBuildup measures a player's involvement in the buildup phase leading to a goal, excluding the final pass or shot. It focuses on the player's contributions that help move the ball forward and create goal-scoring chances, without directly leading to a shot or assist.

How it is calculated: **xGBuildup** is similar to **xGChain** but excludes the player's direct contribution to the final shot or assist. Instead, it includes the actions leading up to those final moments, such as passing, dribbling, or positioning that helps facilitate the attack.

Correlation Matrix

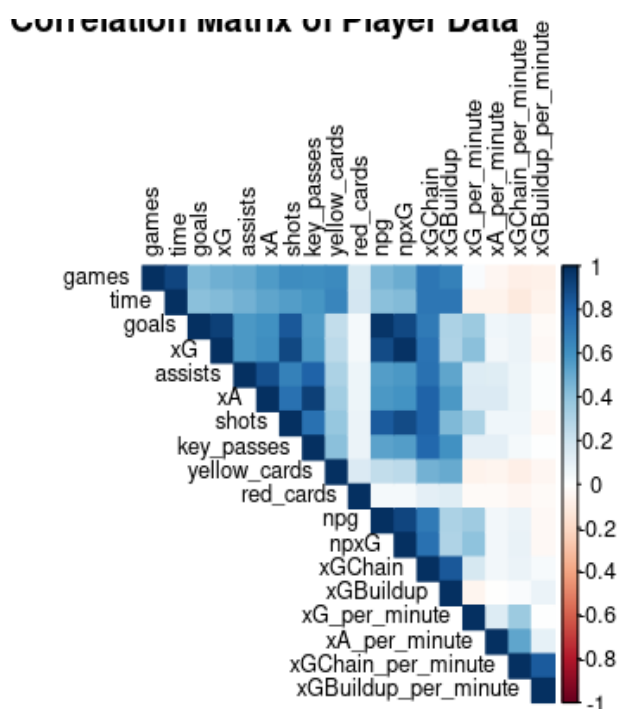
Correlogram of Per Minute Metrics



In the **Correlation Matrix**, we observe the relationships between key metrics like **xG (Expected Goals)**, **xA (Expected Assists)**, **xGChain**, and **xGBuildup**.

- The correlation between **xG** and **xA** is **low**, suggesting that a player's ability to generate goal-scoring opportunities (xA) does not necessarily correlate strongly with their own expected goal-scoring (xG). This indicates that a player may contribute to assists without directly finishing opportunities themselves, or vice versa.
- The correlation between **xGChain** and **xGBuildup** is **high**, which aligns with the expectation that both metrics measure a player's involvement in the buildup phase of a goal. **xGChain** and **xGBuildup** are closely related because they both capture a player's role in facilitating attacking plays, even before the final shot or assist.
- The correlation between **xGChain** and **xA** is **medium**, showing a moderate relationship between a player's contribution to the build-up phase (**xGChain**) and their ability to create assists (**xA**). This suggests that players who are involved in creating goal-scoring opportunities might also be more likely to directly assist in the final play.

This matrix helps us understand that while traditional metrics like **xG** and **xA** are important, advanced metrics like **xGChain** and **xGBuildup** offer valuable additional insights into a player's contribution, especially in terms of their involvement in offensive play prior to goal creation.



In the extended **Correlation Matrix**, we observe a **high correlation** between **key_passes** and **xGChain** and **xGBuildup**. This strong relationship suggests that a player who is actively involved in delivering key passes — passes that lead to goal-scoring opportunities — is also likely to have a significant role in the **xGChain**.

Comparison of xG and xGChain:

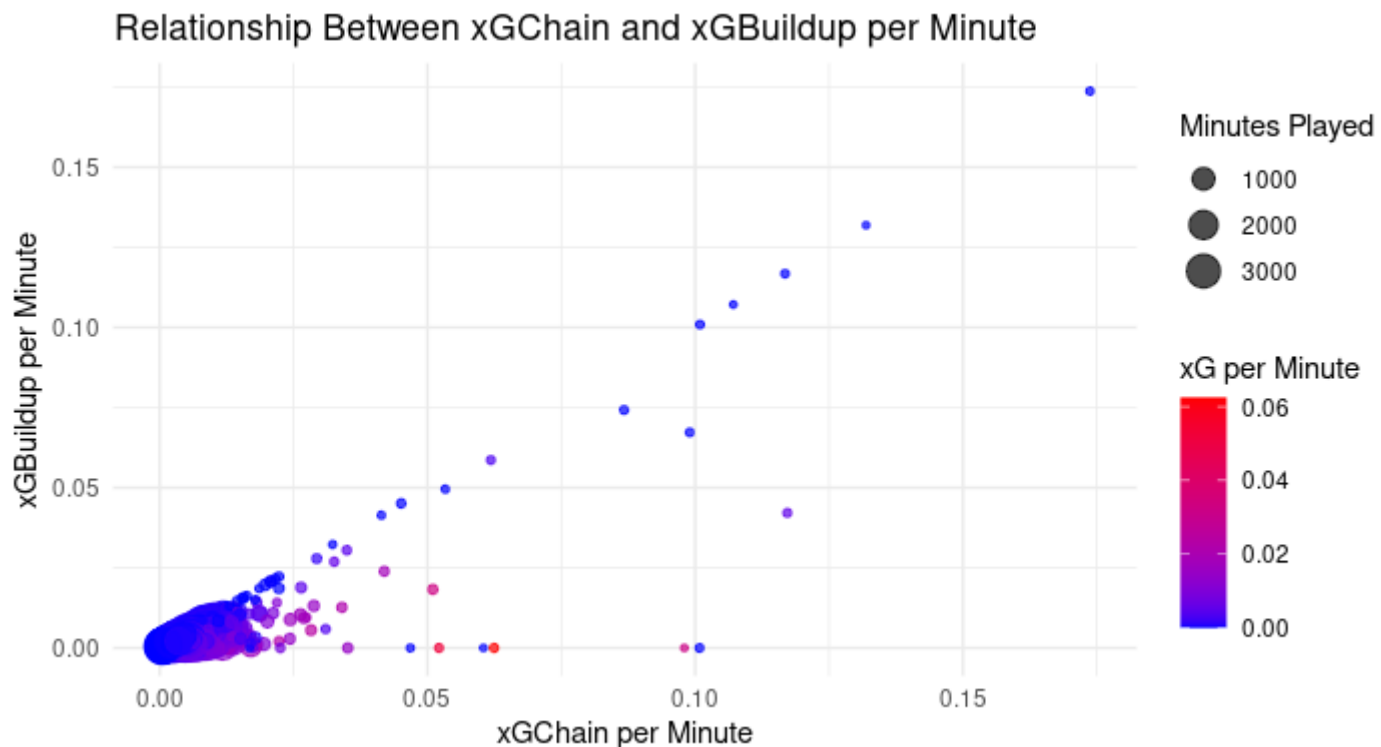
In this analysis, we used a **boxplot** to compare the distributions of **xG (Expected Goals)** and **xGChain**. The boxplot reveals significant differences between these two metrics, specifically in terms of **average values** and **range**.

- The **xGChain** metric shows a **much larger average** and a **wider range** compared to **xG**. This can be attributed to the nature of **xGChain**, which captures a player's involvement in the entire goal creation process, including passes, dribbles, and other contributions leading up to a shot or assist. **xGChain** includes all actions that facilitate a goal-scoring opportunity, regardless of whether the player directly scores or assists. Therefore, players can have high **xGChain** values even if they are not the ones finishing the play, leading to a broader range of values.
- On the other hand, **xG** focuses solely on the probability of a player scoring from their own shot attempts. Since **xG** only accounts for the likelihood of an individual's shot resulting in a goal, its values are generally **more concentrated** and **narrower**. This is because **xG** does not account for the player's contribution to the entire build-up process, only their own scoring chances.

From the **boxplot**, we can observe that the **average value of xGChain** is higher, reflecting that players typically contribute more to the build-up phase than to their own goal-scoring chances. Additionally, the **wider range of xGChain values** indicates that players who play a key role in the attacking build-up (through passing, positioning, and creating opportunities) can have high **xGChain** scores, even if their individual shot attempts are limited.

This analysis highlights the importance of **xGChain** as a more holistic metric that provides insights into a player's involvement in goal creation, emphasizing the value of playmakers and contributors to the offensive play. In contrast, **xG** remains focused on a player's direct goal-scoring potential, making it a more specific but narrower measure.

Relationship Between xGChain and xGBuildup:



In this analysis, we used a **scatter plot** to examine the relationship between **xGChain** and **xGBuildup**, with the color factor representing **xG_per_minute**. The plot shows an overall pattern where most data points are clustered around the **0** mark, indicating that many players contribute modestly to both **xGChain** and **xGBuildup**.

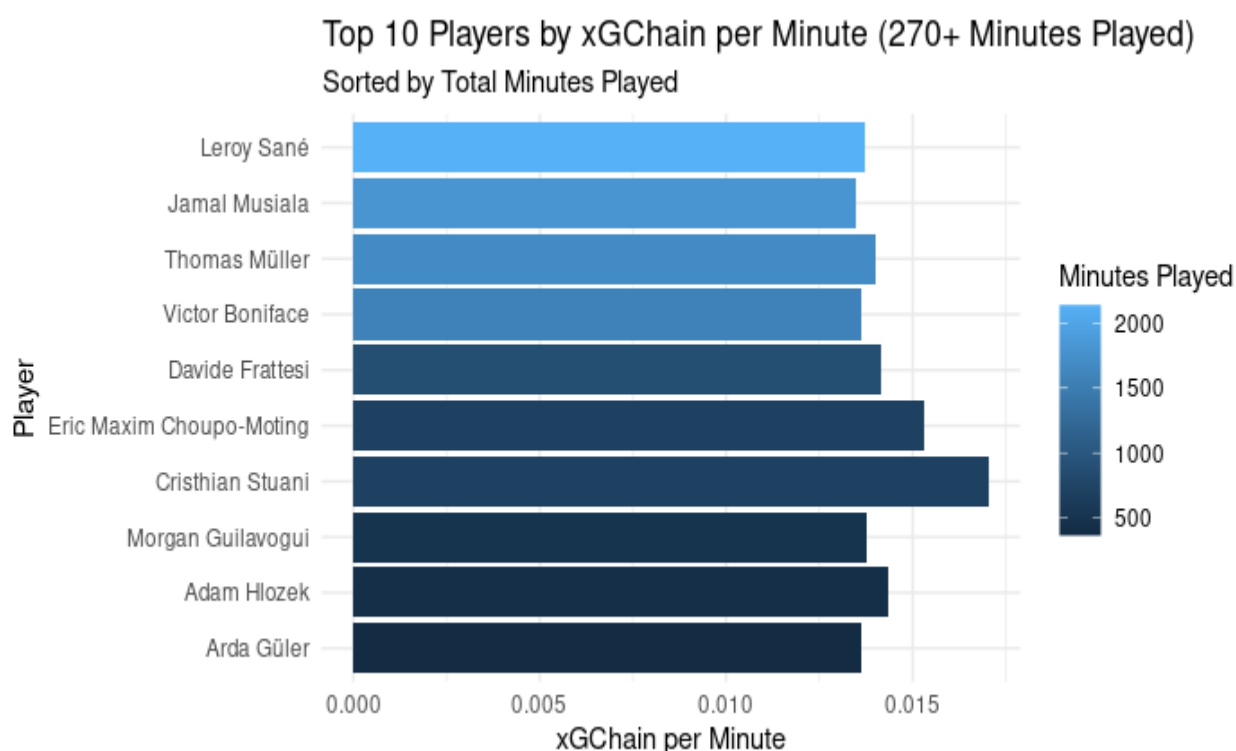
However, some observations in the scatter plot stand out, as they show a **linear upward trend**. This trend suggests that, for certain players, as their **xG_per_minute** increases, both their **xGChain** and **xGBuildup** values tend to rise as well. In other words, players who have higher **xG per minute** values — meaning they are more involved in generating goal-scoring opportunities — also tend to have higher **xGChain** values, indicating a greater involvement in the overall goal creation process.

This relationship highlights that players with higher goal-scoring potential, as represented by **xG per minute**, often play a more active role in building up goal-scoring opportunities, which is captured by **xGChain**. These players contribute not only to their own shots but also to the overall offensive play, moving the ball and creating chances for others.

Thus, the scatter plot emphasizes the importance of considering both **xGChain** and **xGBuildup** when evaluating a player's offensive involvement. While **xGBuildup** measures a player's contribution to the build-up phase, **xGChain**

takes into account the player's role in the entire sequence leading to a goal, revealing a deeper level of involvement, especially for players with high **xG**.

Top 10 Players by xGChain and xGBuildup:



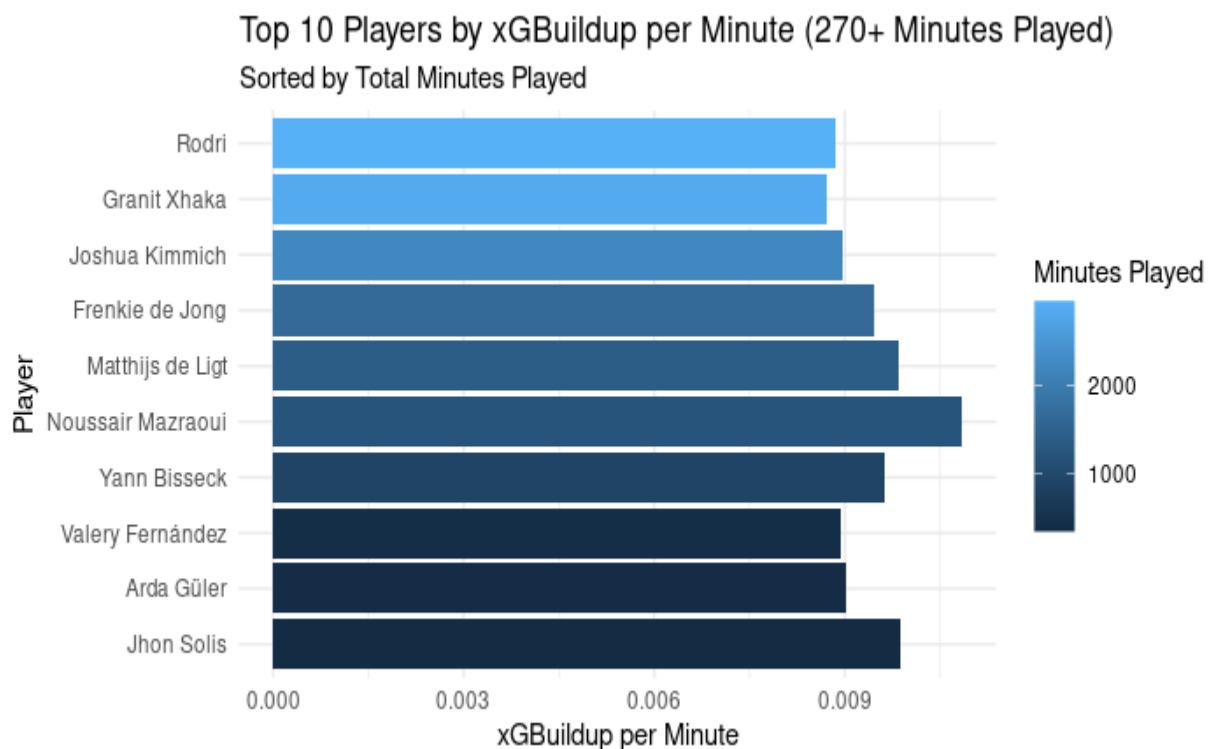
For this analysis, we focused on players with at least **270 minutes of playing time** to ensure that per-minute statistics are meaningful and not overly influenced by small sample sizes. Among the top 10 players ranked by **xGChain**, the list was further sorted by total playing time to better understand their overall contribution.

From this sorting, **Leroy Sané** emerged as the top-ranked player in **xGChain**. This ranking can be attributed to his high playing time, which allows his cumulative contributions to stand out. However, it's important to note that **Cristhian Stuani**, despite producing the **highest xGChain per minute**, ranked lower on the list due to his relatively limited playing time (around 500 minutes). While Stuani's efficiency in generating xGChain is remarkable, his overall ranking reflects the smaller sample of minutes he played.

This comparison highlights the trade-off between **total contribution** and **efficiency** in cumulative metrics like **xGChain**. Players like **Leroy Sané** exemplify consistency and high involvement over extended playing periods, whereas

players like **Cristhian Stuani** demonstrate exceptional impact in shorter bursts of playing time.

By accounting for both total playing time and **xGChain**, this analysis provides a nuanced view of player contributions, emphasizing the need to balance cumulative metrics with per-minute efficiency.



For the **xGBuildup** analysis, we again focused on players who played at least **270 minutes**, ensuring reliable per-minute statistics. The list was ranked by **xGBuildup_per_minute**, and then further sorted by total playing time to provide additional context.

- **Rodri** ranked first in this category, thanks to his **nearly 3000 minutes** of playing time, which highlights his consistent involvement in the build-up phase throughout the season.
- On the other hand, **Noussair Mazraoui** recorded the **highest xGBuildup_per_minute**, reflecting his exceptional efficiency in contributing to the build-up. However, due to his **relatively lower playing time (around 1200 minutes)**, he ranked in the middle of the top 10.
- Notably, **Arda Güler** appears in the top 10 for both **xGChain** and **xGBuildup**, showcasing his dual importance in attack organization and play-building. This highlights his versatility and value as a player who can contribute both to the build-up and final phases of attacking plays.

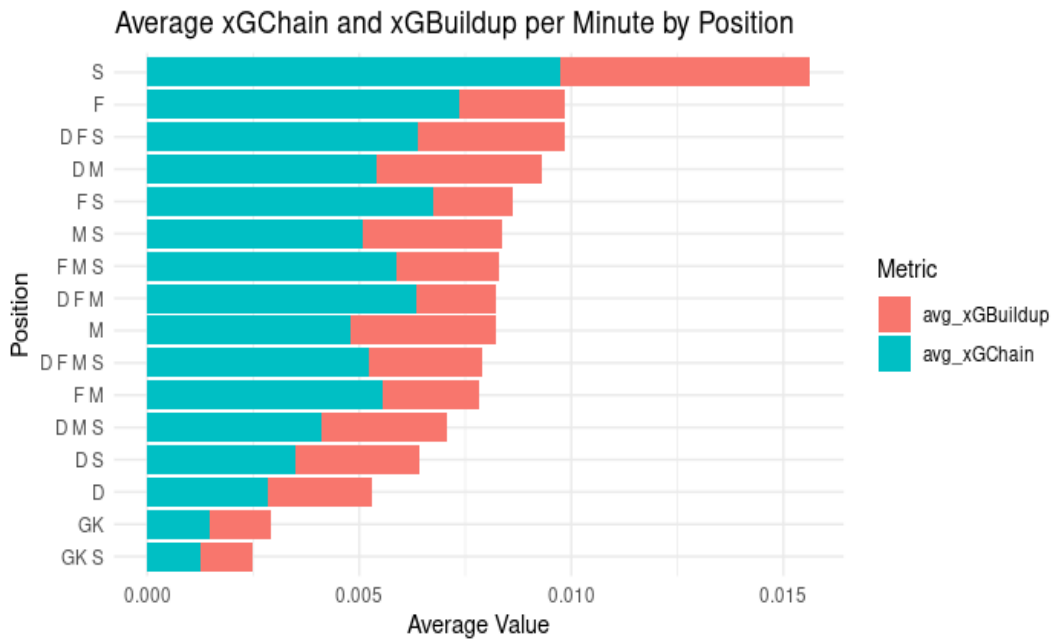
This analysis emphasizes the distinction between cumulative contributions and efficiency. While players like **Rodri** demonstrate sustained impact over long playing periods, others like **Mazraoui** excel in making a significant impact in shorter bursts. The inclusion of **Arda Güler** in both lists underscores his comprehensive value to the team, excelling in both play creation and attack orchestration.

Conclusion:

Traditional statistics, while valuable, often fall short in fully capturing a player's overall impact on the game. Metrics like **xGChain** and **xGBuildup** provide a more nuanced understanding by quantifying contributions that go beyond direct goal-scoring or assisting. These advanced metrics allow us to evaluate a player's involvement in the build-up and progression of offensive plays, offering a deeper insight into their importance within a team.

As demonstrated through this analysis, players who excel in **xGChain** and **xGBuildup** significantly influence their team's offensive structure, whether by creating chances or contributing to the flow of the game. Incorporating such metrics into player evaluation enables a more comprehensive understanding of their role and value, moving beyond simplistic measures to embrace the complexity of modern football. This highlights the growing need to integrate advanced analytics into football analysis, ensuring that players' contributions are assessed with the depth they deserve.

2. Does Player Position Significantly Influence xGChain and xGBuildup Contributions?



Average xGChain and xGBuildup by Position:

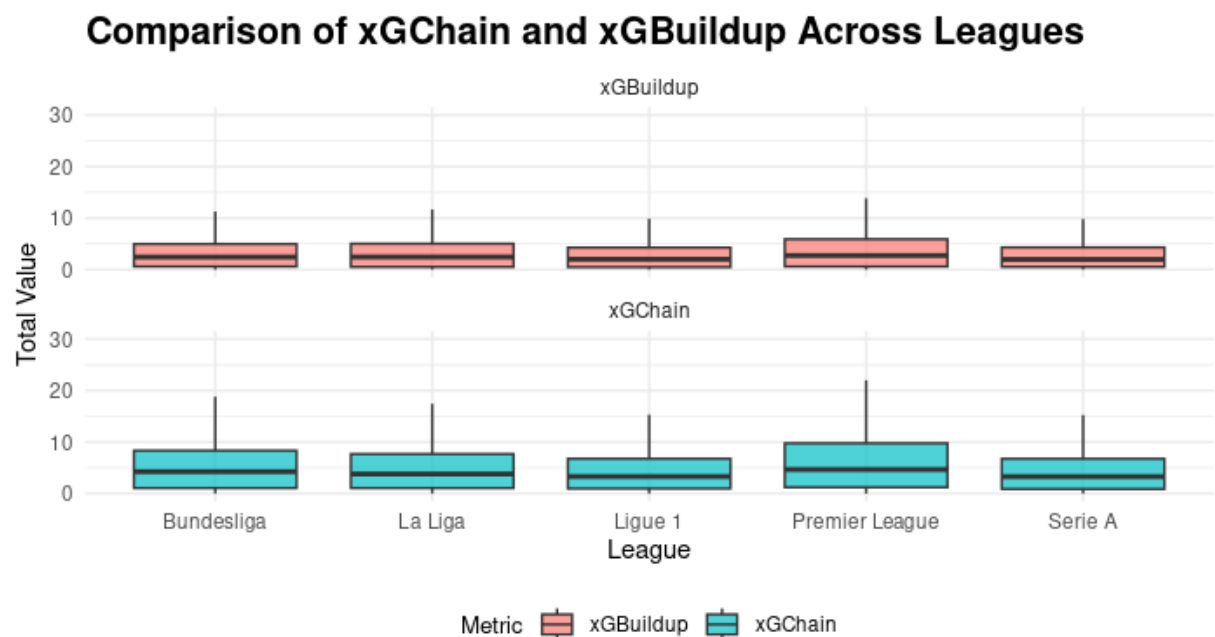
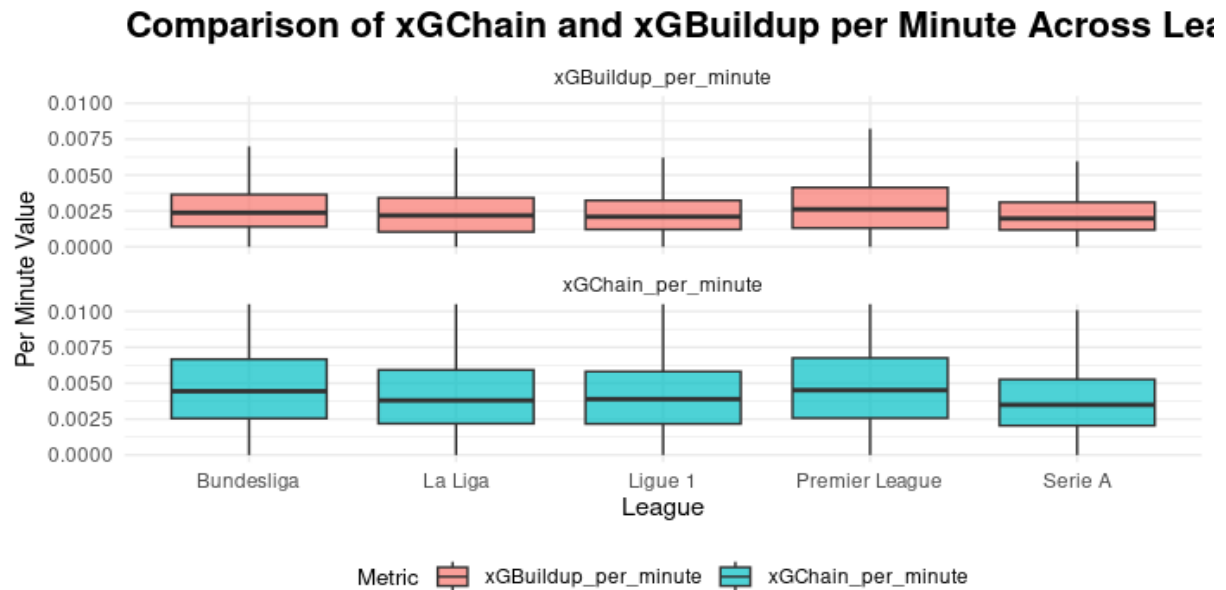
This bar plot showcases the average **xGChain** and **xGBuildup** values for different positions across the **Top 5 European leagues**. The analysis reveals some interesting patterns:

- **Strikers (Forwards)** produce the **highest average xGChain values**, which aligns with expectations. As primary goal-scorers, strikers are naturally more involved in the sequences leading directly to goals.
- Additionally, **Strikers** also generate the **highest average xGBuildup values**, which may seem surprising at first glance. This indicates that forwards in the top leagues often contribute significantly to the build-up phase as well, possibly through link-up play or holding up the ball in attacking positions.
- It's important to note that these averages are calculated for players across the entire Top 5 leagues. The patterns observed here might differ for individual teams, as team tactics and roles assigned to positions vary widely.

This analysis highlights the dual contribution of strikers in both **goal-scoring sequences (xGChain)** and **offensive build-up (xGBuildup)**. While strikers naturally dominate xGChain due to their role, their significant involvement in

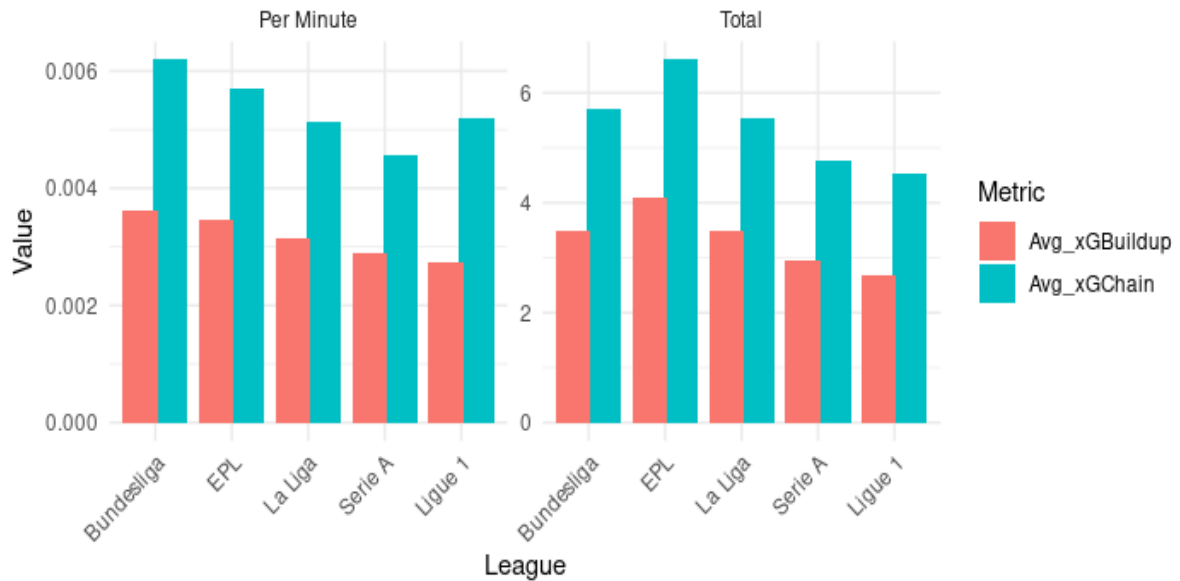
xGBuildup suggests a broader impact in team dynamics beyond just scoring goals.

3. Is there any difference between the leagues by XGChain or XGBuildup Contributions?



Comparison of xGChain and xGBuildup Metrics Across League

Includes Total and Per Minute Averages, Sorted in Descending Order



Looking at both the **grouped barplot** and the **boxplots**, we can draw several important conclusions about the performance of players across the top 5 European leagues.

1. **Premier League (EPL)** stands out as the league with the highest overall averages for both **xGChain** and **xGBuildup**, as shown in the barplot. This suggests that players in the EPL generally have higher involvement in attacking and build-up plays compared to other leagues, reflecting a more dynamic and offensive playing style.
2. **Serie A**, on the other hand, has the lowest average values in both metrics, which can be seen both in the barplot and the boxplot. This suggests that Serie A players, as a group, may be less involved in build-up play or goal-scoring opportunities than players in other leagues. It could indicate a more defensive or slower-paced style of play in Serie A.
3. When we shift to the **per minute statistics**, the **Bundesliga** emerges as the league with slightly higher values than the EPL. Despite the EPL having higher total averages, the Bundesliga's players show more effective output in both **xGChain** and **xGBuildup** when measured per minute. This could indicate that the Bundesliga is played at a faster pace, with players more frequently contributing to attacking sequences, albeit over a shorter amount of time.
4. The **boxplots** reinforce these findings, with the EPL showing a wider range of values, suggesting more variability in individual player contributions. The boxplot for Serie A shows a narrower distribution, indicating more consistency in lower values. Meanwhile, the Bundesliga's boxplot shows a similar spread to the EPL,

further supporting the observation that per-minute values in the Bundesliga are comparable, if not slightly better, than those in the EPL.

These observations provide a deeper understanding of how player contributions in terms of build-up and attacking play can vary across leagues. The **EPL** may have higher total contributions, but the **Bundesliga** shows a slight edge in per-minute performance, while **Serie A** lags behind both in terms of both total and per-minute metrics.