

Visualisation: [Expected stats in the EPL story | Tableau Public](#)

Project Proposal: Analysing the Correlation Between Expected Statistics (xG and xA) and Actual Goals and Assists in the Premier League

Executive Summary

My project aims to investigate the reliability and effectiveness of expected statistics (xG - Expected Goals, and xA - Expected Assists) in predicting actual goals and assists in the Premier League. I chose this topic because understanding the accuracy of these metrics is crucial for both sports analysts and clubs for player evaluation and tactical decisions. By analysing data, I intend to provide insights into the correlation between xG, xA, and real-life performance.

Project Objectives

- **Assess the Reliability of Expected Statistics:** Determine how well xG and xA align with actual goals and assists over the course of a Premier League season
- **Evaluate the Predictive Power:** Measure the extent to which xG and xA can forecast individual player and team performance, for any anomalous cases, identify causes for them
- **Provide Insights for Clubs and Analysts:** Offer actionable insights to football clubs, analysts, and fans on the utility and limitations of these metrics in assessing player performance and strategy

Audience

The project is intended for football clubs, sports analysts, and data-driven fans interested in player performance evaluation and team strategy improvement. Since the audience will vary from sports enthusiasts to data analysts, it must account for different levels of football knowledge and data literacy.

Dataset

I will use data from the understat.com which includes player data from past and present premier league seasons, including player statistics, match results, and xG and xA values for the entire season. We selected this dataset because it offers comprehensive information for a rigorous analysis of the correlation between expected and actual statistics.

Approach

Data Analysis: Employ statistical techniques to analyse the correlation between xG, xA, and actual goals and assists.

Visualisation: Create interactive dashboards and visualisations to present my findings, making them accessible to various audiences.

Challenges

Foreseeable challenges include:

- Data Quality: Ensuring the accuracy and consistency of data across different sources.
- Interpretability: Making complex statistical findings understandable to a wide range of stakeholders.

I hope to gain more experience in data analysis, model interpretation, and effective data visualisation, specifically interactive dashboards.

A question to my peers would be what are some creative ways to visualise the correlation between xG, xA, and actual goals and assists for non-technical audiences?

Persona Document: Football Enthusiast - Chris

Name: Chris | **Age:** 28 | **Gender:** Male

Notes:

- Chris is a dedicated football fan who watches multiple matches each week.
- He relies on pundits and analysts for post-match insights but wishes to develop a more data-informed perspective.
- Chris is more interested in qualitative insights, such as player tactics and team dynamics, rather than in-depth statistical analysis.



Goals:

1. **Stay Informed:** Chris is passionate about football and follows the Premier League closely. His primary goal is to stay informed about his favourite teams, players, and the league as a whole.
2. **Enjoy the Game:** He wants to enhance his football-watching experience by understanding the dynamics of the game, player performance, and team strategies.
3. **Engage in Football Conversations:** Chris aspires to engage in meaningful discussions about football with friends and colleagues. He wants to contribute to conversations with insights beyond just scores and highlights.

Challenges and Needs:

1. **Limited Data Expertise:** Chris loves football but doesn't have extensive knowledge of data analysis or statistics. He often finds it challenging to interpret complex data-related information.
2. **Data Overwhelm:** The vast amount of football statistics and data available can be overwhelming for Chris. He needs help in distilling the most interesting and relevant insights from the data.
3. **User-Friendly Information:** Chris prefers information that is easy to understand and presented in a visually appealing way. He values simplicity and clarity over technical details.

Context:

- **Viewing Habits:** Chris typically watches Premier League matches on TV or streaming platforms. He also reads football news and follows discussions on social media.
- **Engagement:** He engages with football-related content during matches, halftime, and post-match analysis. Chris occasionally visits football websites and forums to gain insights and share his thoughts.
- **Objective:** While watching matches, Chris wishes to access supplementary information that enhances his understanding of the game, player performances, and team strategies. He often checks post-match analyses to see if they align with his perceptions.

Design choices made

- By utilising a multi-frame data story, an introduction slide can be used to explain concepts which may be foreign to those with little or no football knowledge. The concept of xG is one which requires a little explanation and so an initial slide giving a brief introduction was necessary
- By creating KPI statistics 'Goals KPI' and 'Assists KPI' a easy way to quantify the performance of players is created. This summary statistic is useful for people with limited data expertise since it gives a simple universal scale on which a player's performance can be judged, useful for our audience who may have limited knowledge on data
- Since our audience is a football fan who enjoys using information to understand the game and player performance, a section was added highlighting key individual players who stood out according to the data. This adds a personal touch to the study which the audience is familiar with and will make the presentation more engaging and memorable

Reflection on final project vs original project proposal

Overall the project proposals widely remained the same, the three key objectives presented in the original plan were: Assess the reliability of expected statistics, evaluate the predictive power and provide insights for clubs and analysts. The reliability of expected statistics such as xG and xA were explored on both the player and team level and some outliers were identified, along with possible causes for them.

One area which changed was the aim to provide insights for clubs and analysts. The focus of the project shifted more to the reliability of the statistics and so the only thing this study did was to verify the reliability. No in-depth analysis was conducted on methods to improve the use of these statistics.