

Chelsea Group Project

October 2022

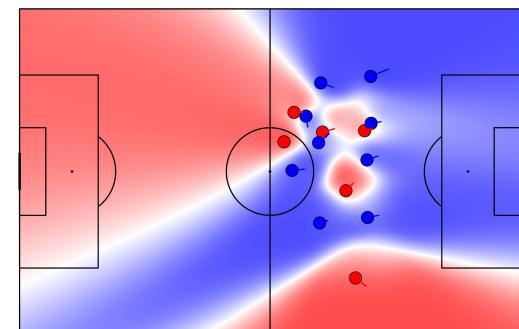
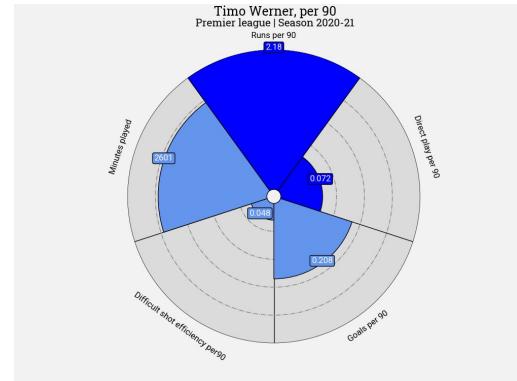
David Andersson, Edd Webster, Leyla Abdul Kader, Lukas Schmid, Mattias Holmström, Paula Borst, Tim Solig, Yash Karle

Slides: <https://docs.google.com/presentation/d/1rGntP2QkR1ubtB2pdVVz1Inf05eK7eVA2tUy7XggirE/edit?usp=sharing>
Code:



Contents

- Introduction (slides 3-4)
- Analytical Projects (3):
 - 1) Club Performance (slides 5-9)
 - 2) Set Pieces (slides 10-16)
 - 3) Tracking Data Scouting (slides 17-24)
- Conclusion (slide 25-26)
- References (slide 27-28)



Introduction

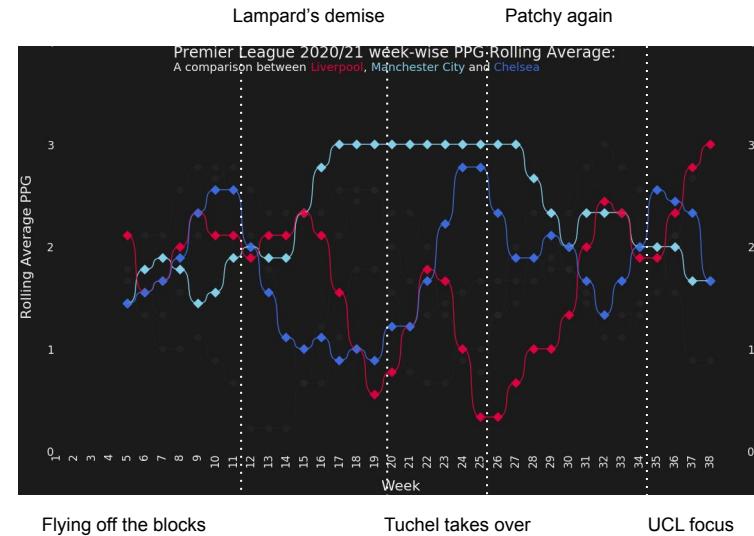
Introduction

End of 20-21 EPL season review and look forward at ways to improve for the 21-22 season

1 How will Chelsea bridge the gap to Man City & Liverpool?

2 How can Chelsea be unpredictable in the final 3rd and unlock low-block opponents?

3 How can Chelsea manage to put games to bed (2+ goals lead) and avoid high pressure situations at the end?



Project 1: Club Performance

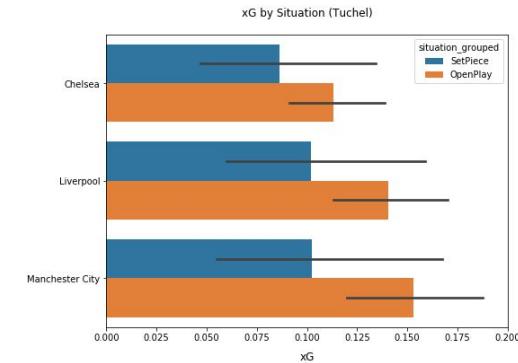
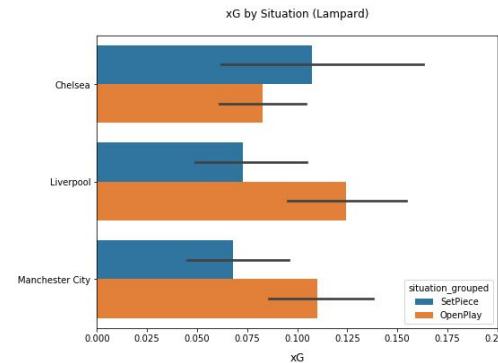
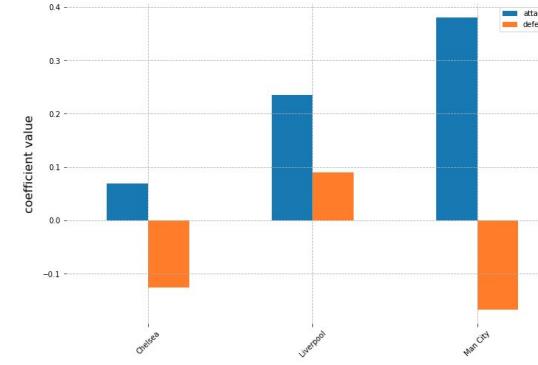
Through the evaluation of a prediction model



In search of marginal gains...

Narrowing down our focus on areas where there's most potential value to be gained

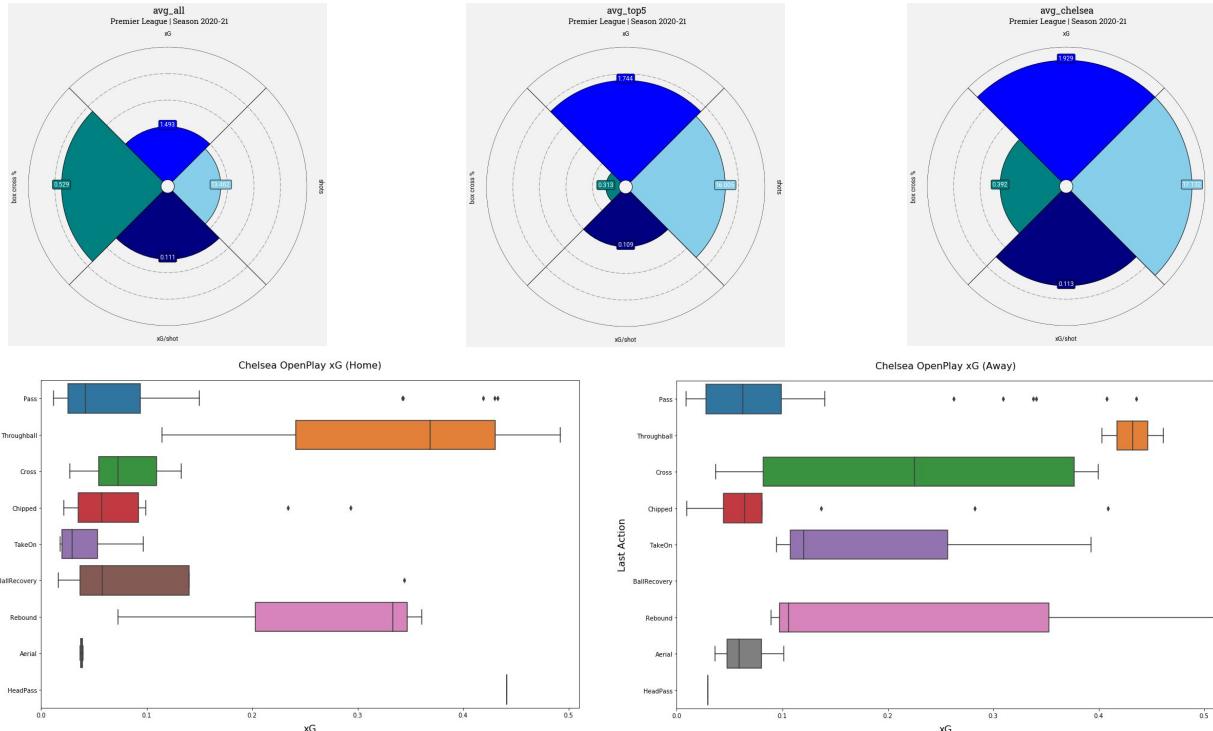
- Naïve first pass regression model to gauge offensive vs defensive shortcomings
 - Are the strikers being wasteful or do the numbers paint a different story?
 - Are Chelsea under Tuchel solid defensively or were they just lucky in keeping clean sheets?
- How does the 1st half of the season compare with the 2nd half?
- Can we spot any major shift in trends for xG avenues?
- Are we already catching-up with City and Liverpool?



Chelsea's attacking style under Tuchel

What are the key differentials when it comes to offensive team attributes?

- Analysing team KPIs to identify specific areas for improvement and correlations
 - Box cross %
 - Shots
 - xG
 - xG/shot
- Better understand open play chance creation; home vs away differences
- This feeds into fine-tuning the recommendations of the tracking data scouting



"LastAction" : [Pass, ThroughBall, Cross, Chipped, TakeOn, BallRecovery, Rebound, Aerial, HeadPass] refer to the penultimate player action before leading to a shot



Building a performance-based league simulation model

Baseline Poisson regression model built and 21-22 season simulated using Monte Carlo Simulations (20,000 iterations)

- Training data used from 19-20 season and predictions made based on performance metrics from the 20-21 season
- Simulated for a 17-team league; leaving out the newly promoted teams owing to data limitations
- “A top six team playing at home against a team other than City or Liverpool with more shots on target” scores the most

| | coef | std err | z | P> z |
|-------------------|---------|---------|--------|-------|
| Intercept | -0.4484 | 0.095 | -4.745 | 0.000 |
| shots | -0.0380 | 0.010 | -3.797 | 0.000 |
| shots_ot | 0.1395 | 0.017 | 8.070 | 0.000 |
| corners | -0.0252 | 0.012 | -2.020 | 0.043 |
| xG | 0.3424 | 0.046 | 7.435 | 0.000 |
| SetPiece_xG | -0.0862 | 0.147 | -0.587 | 0.557 |
| home | 0.1715 | 0.063 | 2.703 | 0.007 |
| big_six | 0.2048 | 0.069 | 2.960 | 0.003 |
| vs_city_liverpool | -0.3087 | 0.130 | -2.370 | 0.018 |

```

* Game # 3 *
* Home team: Chelsea
* Away team: Crystal Palace
2.1647832351671084 0.5001200497337408
*****
*
* SIM STATS *
*
*****
| Total # of sims | Total time (s) for sims | HOME WINS | AWAY WINS | DRAWS |
| 20000 | 0.007 | 15299 | 1404 | 3297 |
| - | - | 76.5% | 7.02% | 16.48% |
*****
*
* SCORE MATRIX (% PROBABILITY) *
*
*****
| | 0 | 1 | 2 | 3 | 4 |
| 0 | 6.82 | 3.59 | 0.81 | 0.15 | 0.01 |
| 1 | 14.99 | 7.22 | 1.76 | 0.29 | 0.02 |
| 2 | 16.39 | 8.48 | 2.19 | 0.3 | 0.03 |
| 3 | 11.88 | 5.64 | 1.4 | 0.24 | 0.03 |
| 4 | 6.48 | 3.31 | 0.78 | 0.12 | 0.01 |
*****
*
* SUMMARY *
*
Chelsea win probability %: 76.5 xPts = 2.46
Crystal Palace win probability %: 7.02 xPts = 0.31

```

| | 2021-22 Expected Points | | |
|----------------|-------------------------|------------|----------------|
| | h_baseline | a_baseline | total_baseline |
| Man City | 34.1 | 28.8 | 62.8 |
| Liverpool | 31.2 | 27.4 | 58.5 |
| Chelsea | 31.5 | 25.6 | 57.1 |
| Man United | 28.2 | 27.5 | 55.7 |
| Tottenham | 29.8 | 18.7 | 48.5 |
| Arsenal | 24.6 | 20.2 | 44.8 |
| West Ham | 23.8 | 19.8 | 43.6 |
| Everton | 22.5 | 19.9 | 42.4 |
| Leicester | 21.1 | 19.8 | 40.9 |
| Aston Villa | 23.3 | 17.6 | 40.8 |
| Brighton | 24.2 | 16.3 | 40.5 |
| Leeds | 24.4 | 15.5 | 39.9 |
| Newcastle | 22.3 | 14.4 | 36.7 |
| Wolves | 20.9 | 14.8 | 35.8 |
| Burnley | 19.2 | 15.9 | 35.1 |
| Southampton | 21.2 | 13.1 | 34.2 |
| Crystal Palace | 18.6 | 13.9 | 32.5 |



Simulate Alternative Scenarios to Determine Improvements to Chelsea's Attacking Output

Determined by simulating “what-if” scenarios for both quantitative and qualitative features

- 1) Scaling up home team's xG by 0.1% of stadium capacity in 1000s for all teams e.g. Chelsea playing at Stamford bridge (capacity 40k) expect to have +0.4xG added for their home matches;
- 2) Chelsea have a 10% increase in overall shots on target per match;
- 3) Chelsea have a 10% increase in xG per match against the teams outside of top six; and
- 4) All of the above 3 scenarios come true: results in 65%(+35%) chances of a top 2 finish

| 2021-22 Simulated xP (home) | | | | | |
|-----------------------------|------------|------|------|------|------|
| | h_baseline | h_s1 | h_s2 | h_s3 | h_s4 |
| Man City | 34.1 | 36.4 | 34.0 | 34.1 | 31.1 |
| Chelsea | 31.4 | 33.3 | 32.4 | 32.2 | 30.9 |
| Liverpool | 31.2 | 33.6 | 31.2 | 31.2 | 28.3 |
| Tottenham | 29.8 | 32.6 | 29.8 | 29.8 | 27.4 |
| Man United | 28.2 | 31.3 | 28.1 | 28.1 | 25.8 |
| Arsenal | 24.7 | 27.6 | 24.6 | 24.6 | 22.4 |

| 2021-22 Simulated xP (away) | | | | | |
|-----------------------------|------------|------|------|------|------|
| | a_baseline | a_s1 | a_s2 | a_s3 | a_s4 |
| Man City | 28.8 | 27.3 | 28.8 | 28.8 | 29.9 |
| Man United | 27.5 | 25.7 | 27.3 | 27.5 | 28.8 |
| Liverpool | 27.3 | 25.7 | 27.3 | 27.3 | 28.9 |
| Chelsea | 25.6 | 23.9 | 26.6 | 26.4 | 28.6 |
| Arsenal | 20.2 | 18.3 | 20.1 | 20.1 | 21.7 |
| Tottenham | 18.8 | 16.7 | 18.7 | 18.8 | 21.2 |

| 2021-22 Simulated xP (overall) | | | | | |
|--------------------------------|----------------|----------|----------|----------|----------|
| | total_baseline | total_s1 | total_s2 | total_s3 | total_s4 |
| Man City | 62.8 | 63.7 | 62.9 | 62.9 | 60.9 |
| Liverpool | 58.5 | 59.3 | 58.4 | 58.5 | 57.3 |
| Chelsea | 57.0 | 57.2 | 59.1 | 58.6 | 59.5 |
| Man United | 55.7 | 56.9 | 55.5 | 55.6 | 54.6 |
| Tottenham | 48.5 | 49.3 | 48.5 | 48.5 | 48.6 |
| Arsenal | 44.9 | 45.9 | 44.7 | 44.7 | 44.2 |



Project 2: Set Pieces

To investigate clubs strengths and weaknesses with
free-kicks, corners and throw-ins



***Chelsea scored 12 out of 58 goals in the
20/21 Premier League season from a set
piece****

*A set piece is defined as a goal coming directly from a set piece or initiated from a dead-ball situation and later resulting in a goal from the subsequent actions.
Reference: whoscored.com/Regions/252/Tournaments/2/Seasons/8228/Stages/18685/TeamStatistics/England-Premier-League-2020-2021

Objectives of this Set Pieces Analysis

Goals & Methods

Objectives:

- Analyse set pieces; free-kicks, throw-ins and corners, and how they contributed to Chelsea's success in the 2020/2021 season.
- Identify possible weaknesses in the set pieces.
- Compare Chelsea's set piece performance (mainly focused on the offense) with the rest of the league, also focusing on the top 6, based on expected goal (xG) and expected threat (xT) values.
- Did switching from Lampard to Tuchel mid-season have any impact for set-pieces?

Methods used:

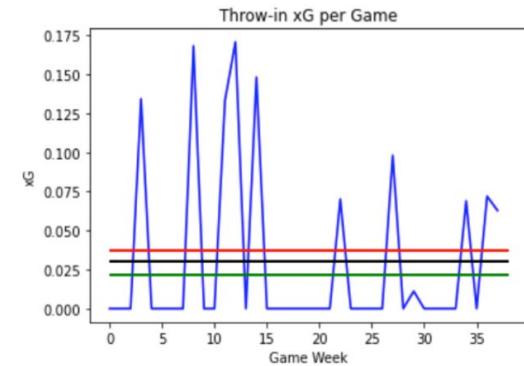
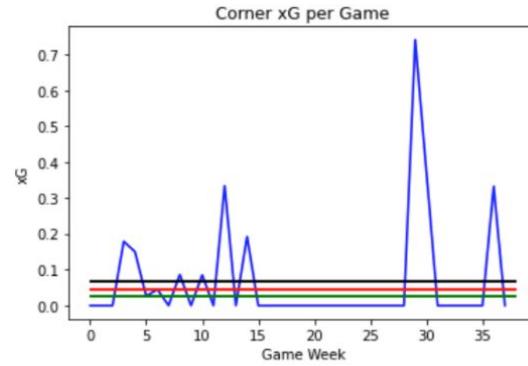
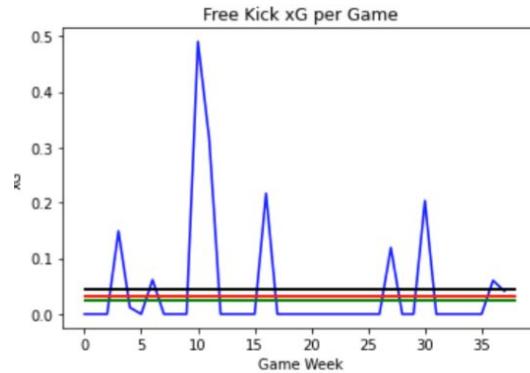
- Heatmaps to investigate possible trends in how set pieces are carried out.
- Possession chains & their visualizations to isolate specific event chains that led to shots (whether successful or not).
- Line graphs indicating Chelsea's xG and xT performance per game week & average lines for the xGs of the top 6 and all teams in the league.
- Possession chains with their calculated xT values.
- Ranked list of Chelsea's top players based on xG & xT values.

Notes:

- Possession chains were filtered to the following:
 - Only chains with duration <25s
 - Only chains with a shot distance <30m



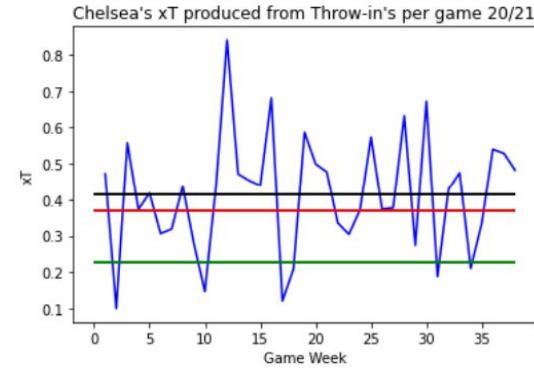
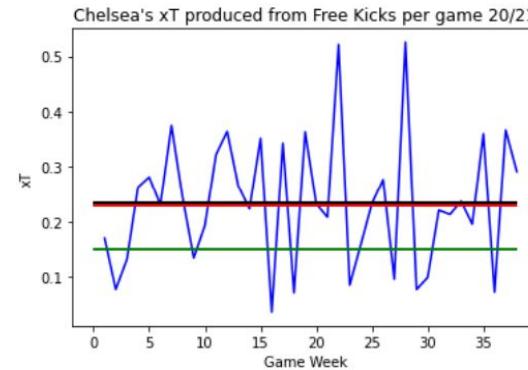
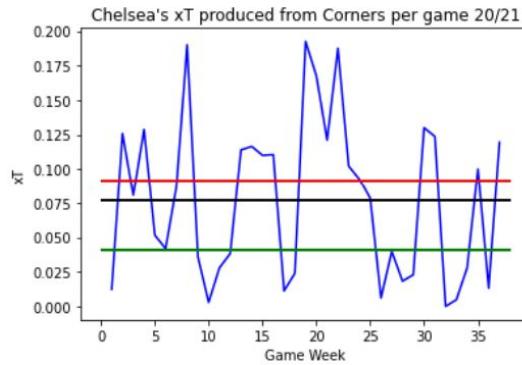
League Comparison of Expected Goals (xG) Generated from Set-Pieces



— xG Chelsea
— Avg xG All teams (excl. Top 6)
— Avg xG Top 6 (excl. Chelsea)
— Avg xG Chelsea



League Comparison of Expected Threat (xT) Generated From Set-pieces

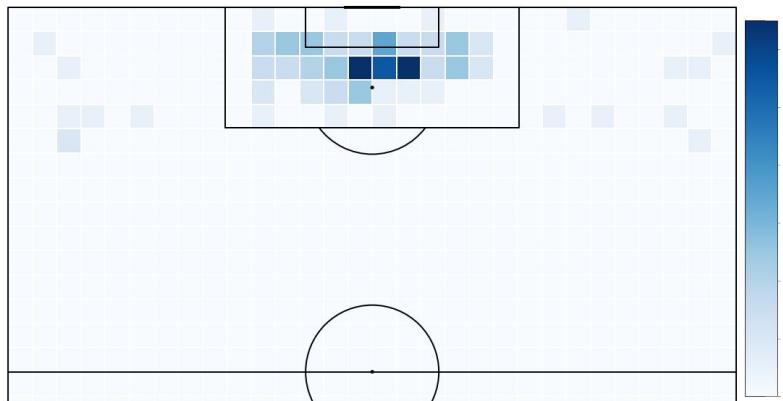


— xT Chelsea per game
— Avg xT Top 6 (excl. Chelsea)
— Avg xT Chelsea
— Avg xT All teams (excl. Top 6)

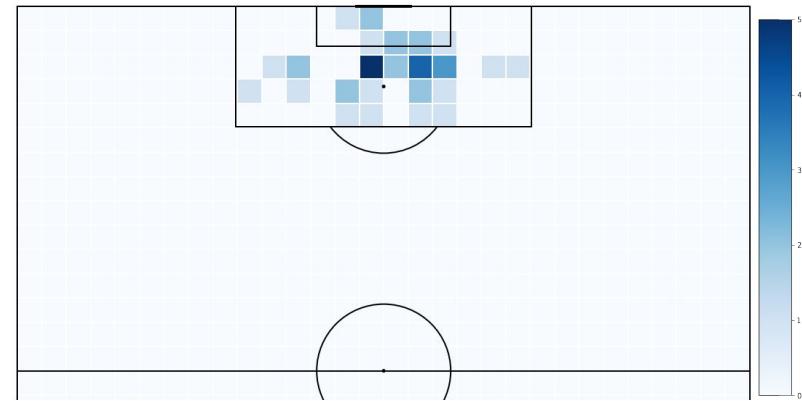


Heatmaps of Corners and Free-kicks for Chelsea

Chelsea Corners - 2020/2021 Premier League Season



Chelsea Freekicks - 2020/2021 Premier League Season



Player Rankings (xT / xG Chain) generated from Set-pieces

Expected Threat (xT)

| | player_id | xT | PlayerName | | player_id | xT | PlayerName |
|-----|-----------|-----------|------------------------|----|-----------|----------|-------------------|
| 263 | 141746 | 11.703309 | Bruno Fernandes | 20 | 184341 | 7.846725 | Mason Mount |
| 253 | 122798 | 11.432426 | Andrew Robertson | 17 | 172850 | 6.076113 | Ben Chilwell |
| 314 | 169187 | 11.166984 | Trent Alexander-Arnold | 24 | 225796 | 5.463850 | Reece James |
| 102 | 60551 | 11.089565 | Ashley Westwood | 16 | 165153 | 4.413832 | Timo Werner |
| 315 | 169359 | 9.191580 | Matt Targett | 1 | 41328 | 4.322621 | César Azpilicueta |
| 188 | 101178 | 9.059363 | James Ward-Prowse | 19 | 176413 | 3.673523 | Christian Pulisic |
| 237 | 114283 | 8.491716 | Jack Grealish | 7 | 91651 | 3.672549 | Mateo Kovacic |
| 306 | 166989 | 8.303798 | Youri Tielemans | 5 | 85955 | 3.489482 | Jorginho |
| 357 | 184341 | 7.846725 | Mason Mount | 13 | 124183 | 3.248819 | Hakim Ziyech |
| 190 | 101188 | 7.775756 | Lucas Digne | 12 | 116594 | 3.221781 | N'Golo Kanté |

Expected Goal (xG) Chain

| | player_id | xG | PlayerName | | player_id | xG | PlayerName |
|----|-----------|-----------|-------------------|-----|-----------|-----------|------------------------|
| 20 | 184341 | 12.732077 | Mason Mount | 263 | 141746 | 16.186447 | Bruno Fernandes |
| 17 | 172850 | 7.186730 | Ben Chilwell | 102 | 60551 | 14.146523 | Ashley Westwood |
| 24 | 225796 | 6.434027 | Reece James | 314 | 169187 | 14.069247 | Trent Alexander-Arnold |
| 7 | 91651 | 5.446151 | Mateo Kovacic | 237 | 114283 | 13.269937 | Jack Grealish |
| 16 | 165153 | 5.059755 | Timo Werner | 357 | 184341 | 12.732077 | Mason Mount |
| 5 | 85955 | 5.041397 | Jorginho | 253 | 122798 | 12.665500 | Andrew Robertson |
| 1 | 41328 | 4.915851 | César Azpilicueta | 107 | 61366 | 12.040358 | Kevin De Bruyne |
| 13 | 124183 | 4.700083 | Hakim Ziyech | 77 | 55459 | 11.263566 | Aaron Cresswell |
| 19 | 176413 | 4.472233 | Christian Pulisic | 315 | 169359 | 10.606970 | Matt Targett |
| 12 | 116594 | 3.933868 | N'Golo Kanté | 306 | 166989 | 10.385418 | Youri Tielemans |



Project 3: Tracking Data

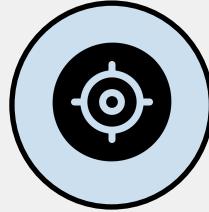
Scouting players to correct Chelsea's weaknesses



Three Key Metrics for Consideration

Derived from the tracking data to assess player performance and determine suitable candidates for recruitment

Measuring direct play leading up to shot



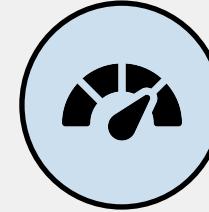
Through the identification of possession chains leading to shots.

Runs at high speed in the final third



Player speeds are calculated and offensive players in the final third are identified.

Ratio of difficult shots on target



The shots that players are able to place on target from bad positions.

*All metrics are adjusted for minutes played.

Analysis conducted using SkillCorner Broadcast Tracking data and Opta Event data.

opta

 SkillCorner



Metric 1: Measuring Direct Play Leading up to a Shot



Definition: A metric that identifies possession chains leading to shots. By pass value packing we identify how many defenders are defeated. Value of the packing is then adjusted for directness of play which is the amount of passes in the chain.

Advantages

- + Takes into account the amount of defenders defeated - the quality of shots
- + Identifies players heavily involved in possession leading to shots
- + Adjusts for the directness of possession

Disadvantages

- The metric is weighted towards stronger teams and generally midfielders
- Actions where ball is played further back the pitch will be lesser valued,
- The quality of shots is more complicated than the metric can handle

| # | Player |
|----|----------------|
| 1 | Steven Alzate |
| 2 | Dennis Praet |
| 3 | De Bruyne |
| 4 | Connolly |
| 5 | Phil Foden |
| 6 | Ross Barkley |
| 7 | Bernardo Silva |
| 8 | Tyler Roberts |
| 9 | Adam Lallana |
| 10 | Neal Maupay |



Metric 2: Runs at High Speed in the Final Third



Definition: Player speeds are calculated and offensive players in the final third are identified. The amount of runs a player makes is summed for the whole season.

Advantages

- + Identifies players with good physical capabilities in terms of speed.
- + Shows players that take initiative.
- + Players at a high level will make runs into space.

Disadvantages

- Does not differentiate between bad runs and good runs.
- Does not identify what other options the player had.
- Weighted towards stronger teams.

| # | Player | Runs per 90 |
|----|----------------|-------------|
| 1 | Timo Werner | 2.17 |
| 2 | Sterling | 2.15 |
| 3 | Firmino | 2.06 |
| 4 | Matej Vydra | 2.0 |
| 5 | Edinson Cavani | 1.9 |
| 6 | Nathan Redmond | 1.85 |
| 7 | Jamie Vardy | 1.83 |
| 8 | Erik Lamela | 1.78 |
| 9 | Salah | 1.69 |
| 10 | Danny Ings | 1.68 |



Metric 3: Ratio of Difficult Shots on Target



Definition: The metric calculates the ratio of shots that players are able to place on target from bad positions. Bad positions are defined as shots at bad angles.

Advantages

- + Identifies players able to get to shooting opportunities.
- + Adjusts for missed shots, thus finding players skilled at getting shots on target.

Disadvantages

- Players of strong teams likely to find themselves in better positions.
- Does not account for how pressured the player is.

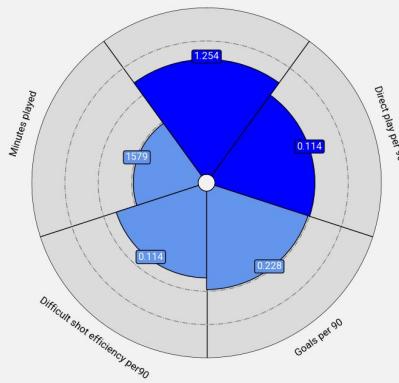
| # | Player | Metric |
|----|---------------|--------|
| 1 | Ryan Fraser | 0.61 |
| 2 | Dennis Praet | 0.49 |
| 3 | Nathan Tella | 0.41 |
| 4 | Joe Bryan | 0.37 |
| 5 | Ayoze Pérez | 0.34 |
| 6 | Jordan Ayew | 0.33 |
| 7 | Ashley Barnes | 0.28 |
| 8 | Martial | 0.25 |
| 9 | Steven Alzate | 0.24 |
| 10 | Batshuayi | 0.23 |



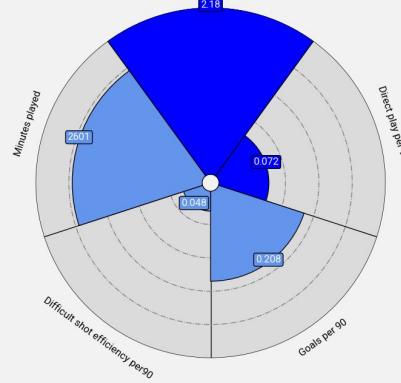
Identified Players in the Current Squad as Potential Areas to Upgrade

Based on the three derived metrics from the broadcast tracking data

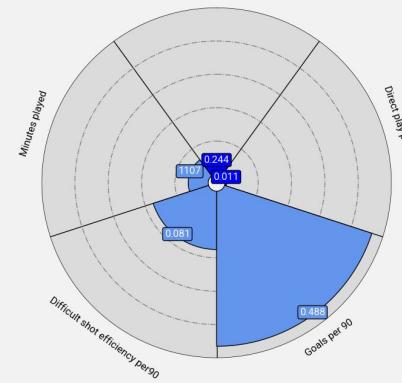
Kai Havertz, per 90
Premier league | Season 2020-21
Runs per 90



Timo Werner, per 90
Premier league | Season 2020-21
Runs per 90



Tammy Abraham, per 90
Premier league | Season 2020-21
Runs per 90



Kai Havertz

| | |
|--------------|--------------------|
| Nationality | |
| Age | 22 |
| Position | Attacking Midfield |
| Market value | 70 m. € |



Timo Werner

| | |
|--------------|----------------|
| Nationality | |
| Age | 25 |
| Position | Centre-Forward |
| Market value | 65 m. € |



Tammy Abraham

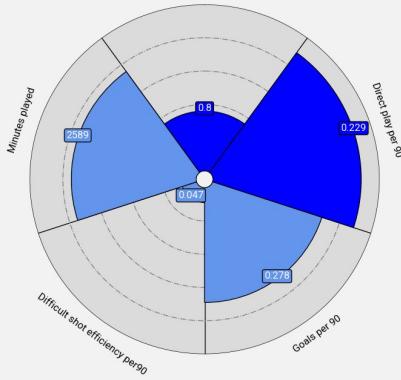
| | |
|--------------|----------------|
| Nationality | |
| Age | 24 |
| Position | Centre-Forward |
| Market value | 38 m. € |



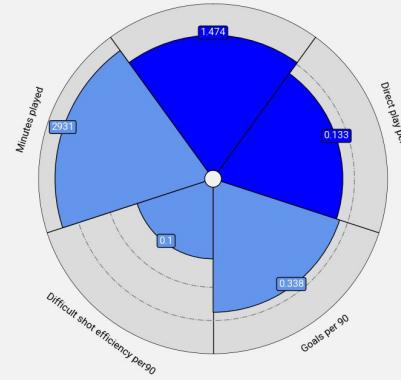
Shortlisted Players

Based on the three derived metrics from the broadcast tracking data

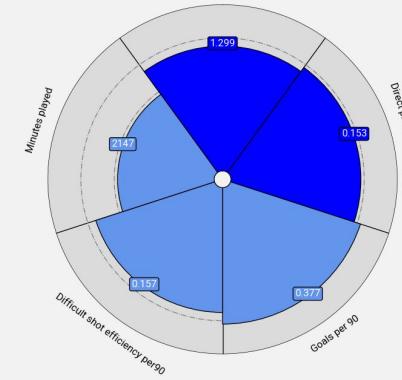
Neal Maupay, per 90
Premier league | Season 2020-21
Runs per 90



Marcus Rashford, per 90
Premier league | Season 2020-21
Runs per 90



Gabriel Jesus, per 90
Premier league | Season 2020-21
Runs per 90



Neal Maupay



Nationality



Age

25

Position

Centre-Forward

Market value

20 m. €



Marcus Rashford



Nationality



Age

24

Position

Left Winger

Market value

85 m. €



Gabriel Jesus



Nationality



Age

24

Position

Centre-Forward

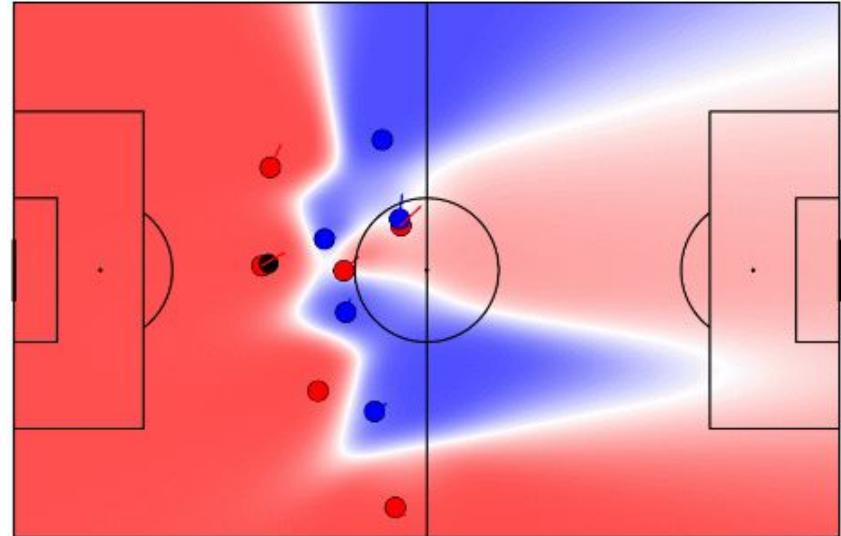
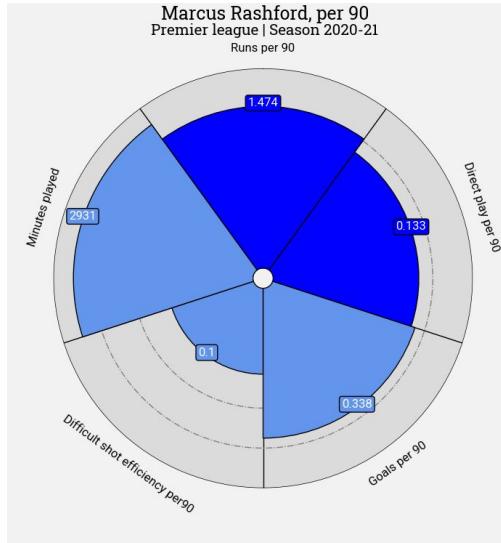
Market value

60 m. €



Marcus Rashford is Our Ideal Pick for Recruitment

The final recommendation of recruitment to improve Chelsea's effectiveness going forward



Example of the use of Pitch Control, to demonstrate how Rashford is able to create space through his runs.



Conclusion



Conclusion of this Analysis

Data-driven insights determined

- 1 Identified the need to focus on offensive enhancements to the team based on club performance metrics. Simulations of alternative scenarios suggest signing an appropriate forward, thereby uplifting both quantity and quality of attempts on target, would be a key first step towards narrowing down the gap to Man City and Liverpool.
- 2 Determined that Chelsea are generally performing well within set pieces through the application of Expected Goals and Expected Threat, and are often managing to get the ball into dangerous areas. Ranked joint fifth highest in the league in amount of goals from set pieces (6, excluding penalties).
- 3 Application of Skill Corner Broadcast Tracking data to derive metrics for player scouting recommendations for Chelsea's front line, measuring directness of play, high-speed runs into the final third and the ability to take difficult shots. The resulting analysis found Marcus Rashford to be the top recommendation.



References



References and Further Reading

Data Sources:

- Football-data.co.uk recorded results data
- Understat aggregated player and team performance data
- Transfermarkt player bio and evaluation data
- Opta Event data (Stats Perform)
- SkillCorner Broadcast tracking data

Tutorials

- [How Our Club Soccer Predictions Work | FiveThirtyEight](#)
- [Premier League Projections and New Expected Goals - Cartilage Free Captain](#)
- Friends of Tracking Tracking data tutorials by [Laurie Shaw](#):
 - Part 1 – Introduction to analysing tracking data in Python: youtube.com/watch?v=8TrleFkIEsE
 - Part 2 – Measuring the physical performance of players: youtube.com/watch?v=VX3T-4IB2o0
 - Part 3 – Advanced metrics: Pitch Control: youtube.com/watch?v=5X1cSehLq6s
 - Part 4 – Evaluating player actions and passing options: youtube.com/watch?v=kXSLKwADXKI

GitHub Repositories

- [LaurieOnTracking](#) by [Laurie Shaw](#) for Tracking data implementation

Seminars and Videos

- How Tracking Data is Used in Football and What are the Future Challenges with [Javier Fernández](#), [Sudarshan 'Suds' Gopaladesikan](#), [Laurie Shaw](#), [Will Spearman](#) and [David Sumpter](#) for Friends of Tracking: youtube.com/watch?v=kHTq9 cwdkGA.



Thank You!

