Football Data Analytics

Dahbi El Mehdi

Blabla Conference, 20th December 2022

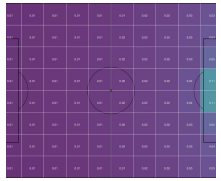
Slides: https://docs.google.com/presentation/d/1Pxi65KwZ7bSHCjlvOrVSi83jebFrTHqWw_IBPDaSKnw/edit#slide=id.g1bc02b52bcc_0_4
Medium: https://medium.com/@dahbielmehdi05
Github repo: https://github.com/Elmehdi9/Football-data-visualization

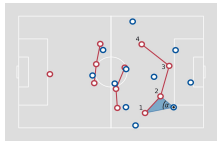
Contents

- Introduction
- Types of data
- The Metrics we can extract from data:

  1) Goal threat metrics
  2) Creativity metrics
  3) Possession metrics
  4) Defensive metrics

- Scouting and recruitment
- Human scouting VS algorithms
- Conclusion





All code for the analysis featured in this pack: github.com/eddwebster/football_analytics.

Introduction & types of data

Data's related fields and football. Different types and sources of football data.

Types of data

Data freely available for all professional matches

Commercially available for professional matches

Proprietary, available for a single team or teams within the same league

High availability

Low availability

Pre 1995

Matchsheet Data

Basic, aggregated stats

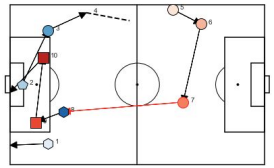
Examples: # goals, subtitutions, cards

1995+

Event Data

Spatio-temporal description of all on-the-ball events)

Examples: Event type, timestamp, spatial location, meta information of on-the-ball actions



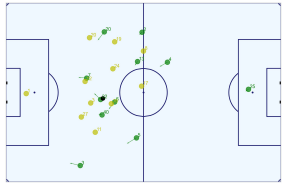
	time	actiontype	player	team
1	16m33s	interception	M. Leckie	Australia
2	16m46s	goalkick	M. Ryan	Australia
3	16m47s	pass	M. Milligan	Australia
4	16m49s	dribble	A. Behich	Australia
5	16m55s	pass	L. Advincula	Peru
6	16m59s	pass	C. Ramos	Peru
7	17m3s	pass	Y. Yotún	Peru
8	17m6s	interception	T. Sainsbury	Australia
9	17m11s	cross	P. Guerrero	Peru
10	17m12s	shot	A. Carrillo	Peru

2005+

Tracking Data

exact spatial locations of the players and the ball

Examples: X, Y coordinates of the 22 players, 3 referees and the ball at every time-step



Low granularity

High granularity

High-level summary

Description of all on-the-ball events

Exact movements of all players and the ball

Source of Event Data

On-the-Ball Event data for football matches is collected by tagging each event that takes place on the pitch, i.e. passes, tackles, aerial-duels, shots, with a timestamp, the player involved, and location on the pitch with X, Y coordinates.

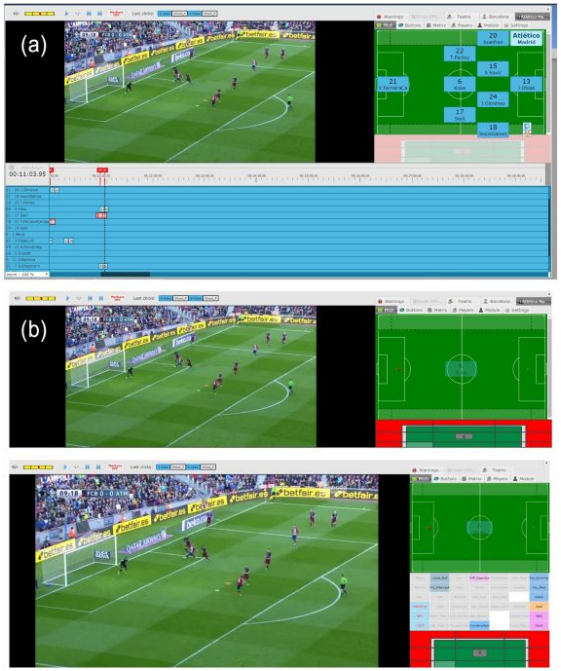
On-the-ball Event data can be provided by:

wyscout

STATSBOMB

InStat

opta



Sources of Tracking Data

Tracking data can be collected in three different ways:

1) In Stadium Optical Tracking:

- Cameras installed in in the stadium at different angles.

2) GPS Tracking:

- Players wear GPS devices to track location at all times.

3) Broadcast Tracking / Single Camera Tracking:

- Optical tracking collected from broadcast feeds.

What metrics can we extract from the data?

The four categories of metrics and how they've been treated mathematically and tactically

Goal threat metrics: Expected goals (xG)

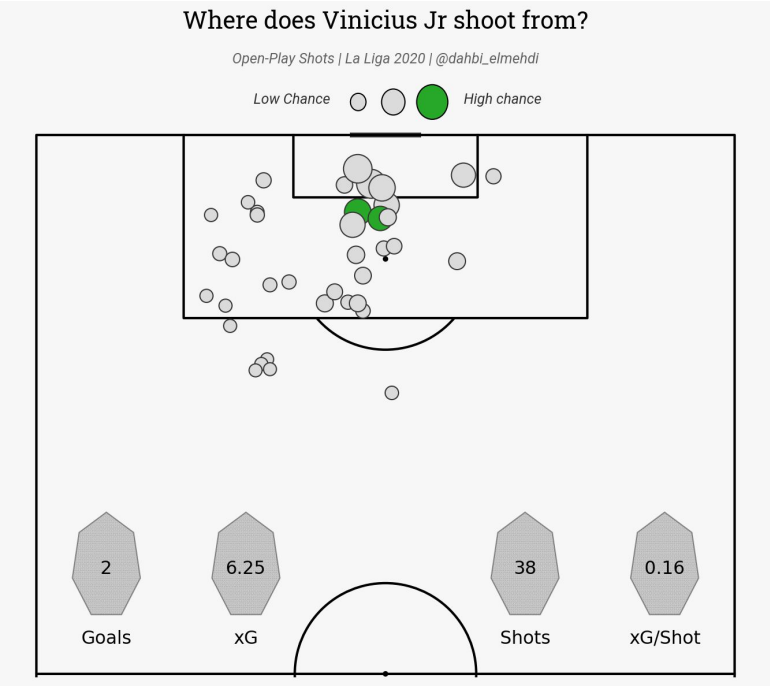
xG is a model that is built using logistic regression algorithm that measures the quality of a shot by calculating the probability that it will be scored from a particular position.

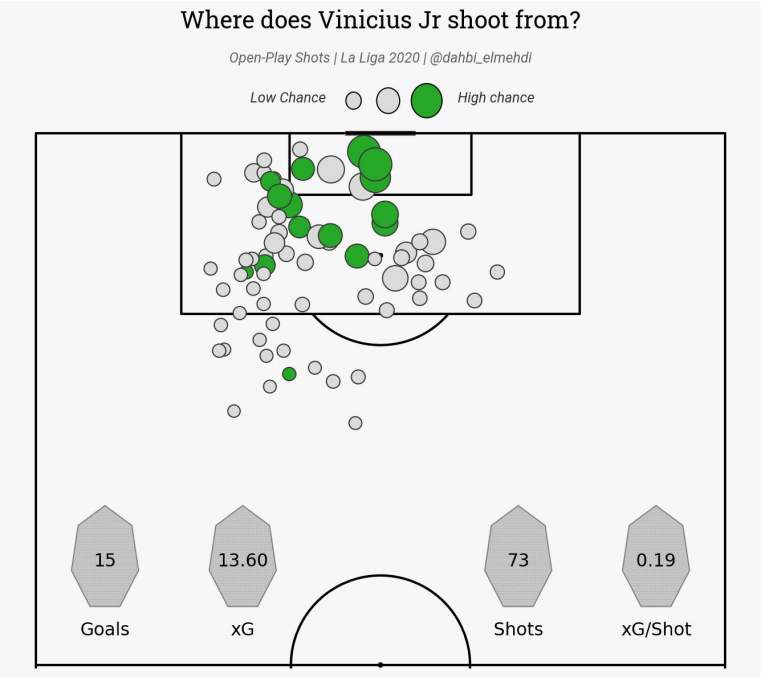
The model is powered by hundreds of thousands of shots and takes into consideration the following features:

- Distance of the goal
- Angle of the goal
- Type of assist
- Pattern of play
- One on One
- Body part

…

Expected goals (xG): Vinicius JR Shotmaps





Goal threat metrics: Expected threat (xT)

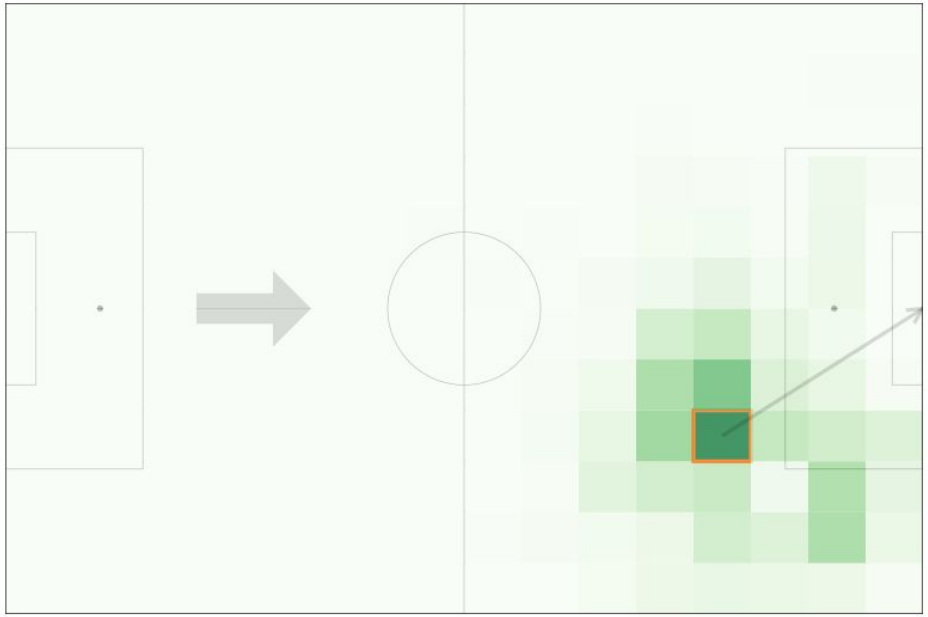
The key intuition, when talking about this metric, is that a player have two options when he has possession in a (x,y) location:

1. He can either shot and score with some probability (xG).
2. He can move the ball by passing or dribbling.

The goal is to calculate the payoff of these actions, as:

xThreat value = shooting payoff + moving payoff

Expected threat (xT): Calculating the payoffs



Premier League Average

When a player has the ball in the highlighted zone, what action will they take next?

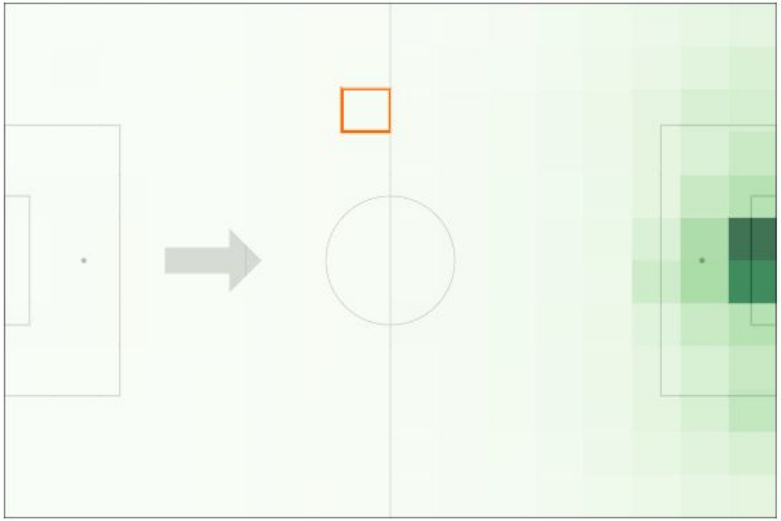
Move: **85%**, according to the map.
Shoot: **15%**, scoring **1%** of shots.

Hover/click to change zones!

Expected threat (xT): Shooting payoff

Based on past data, the probability that a player will tend to shoot is s(x,y), and if you shoot the expected payoff is g(x,y) [xG], so:

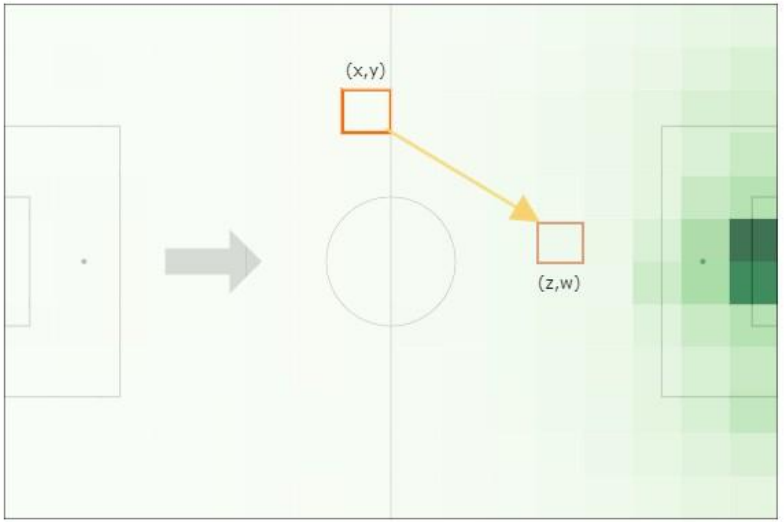
Shooting payoff = s(x,y) x g(x,y)



Expected threat (xT): Moving payoff

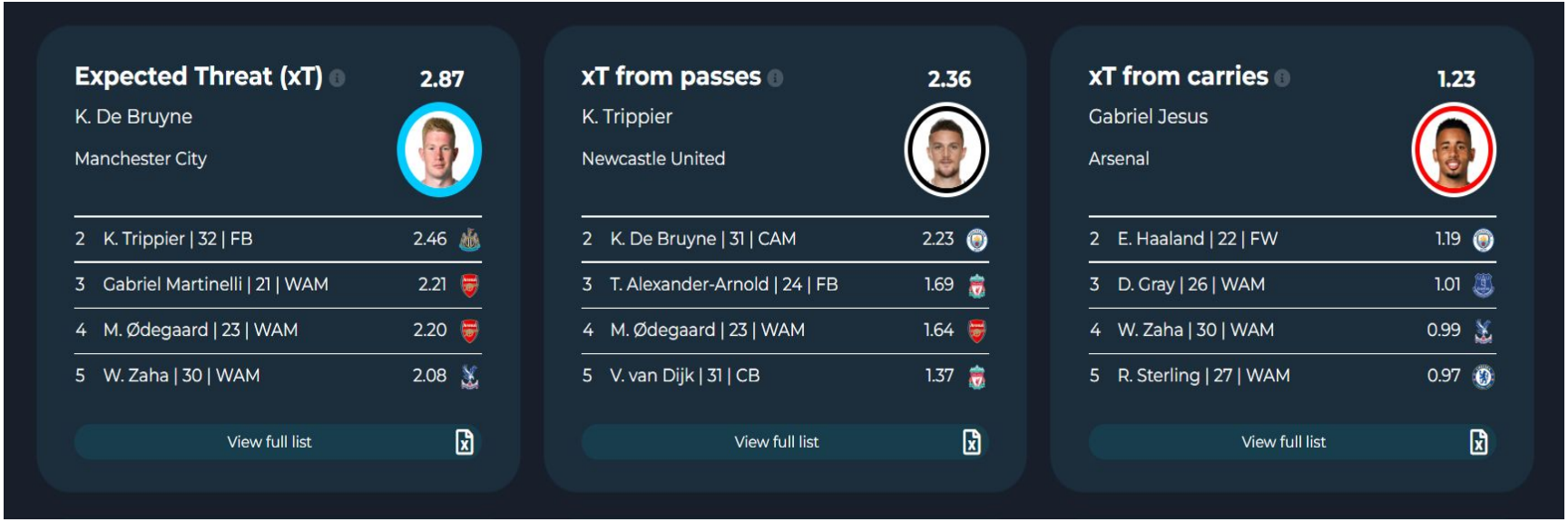
When moving the ball to a (z,w) location the expected payoff is V(z,w), a value that have to be multiplied by the probability of moving the ball to that zone;

$$(m_{x,y} \times \sum_{z=1}^{16} \sum_{w=1}^{12} T_{(x,y)\to(z,w)} V_{z,w})$$



Expected threat (xT): Putting all together

$$V_{x,y} = (s_{x,y} \times g_{x,y}) + (m_{x,y} \times \sum_{z=1}^{16} \sum_{w=1}^{12} T_{(x,y)\to(z,w)} V_{z,w})$$



Goal threat metrics: Expected assists (xA)

xA measures the likelihood that a given pass will become an assist. It considers several factors including:

- Type of pass (e.g., cross, non-cross, header, through ball, etc)
- The pattern of play (e.g., open play, corner, free kick, throw-in, etc)
- Location of where the pass is received
- Location of where the pass is made from
- Distance of the pass

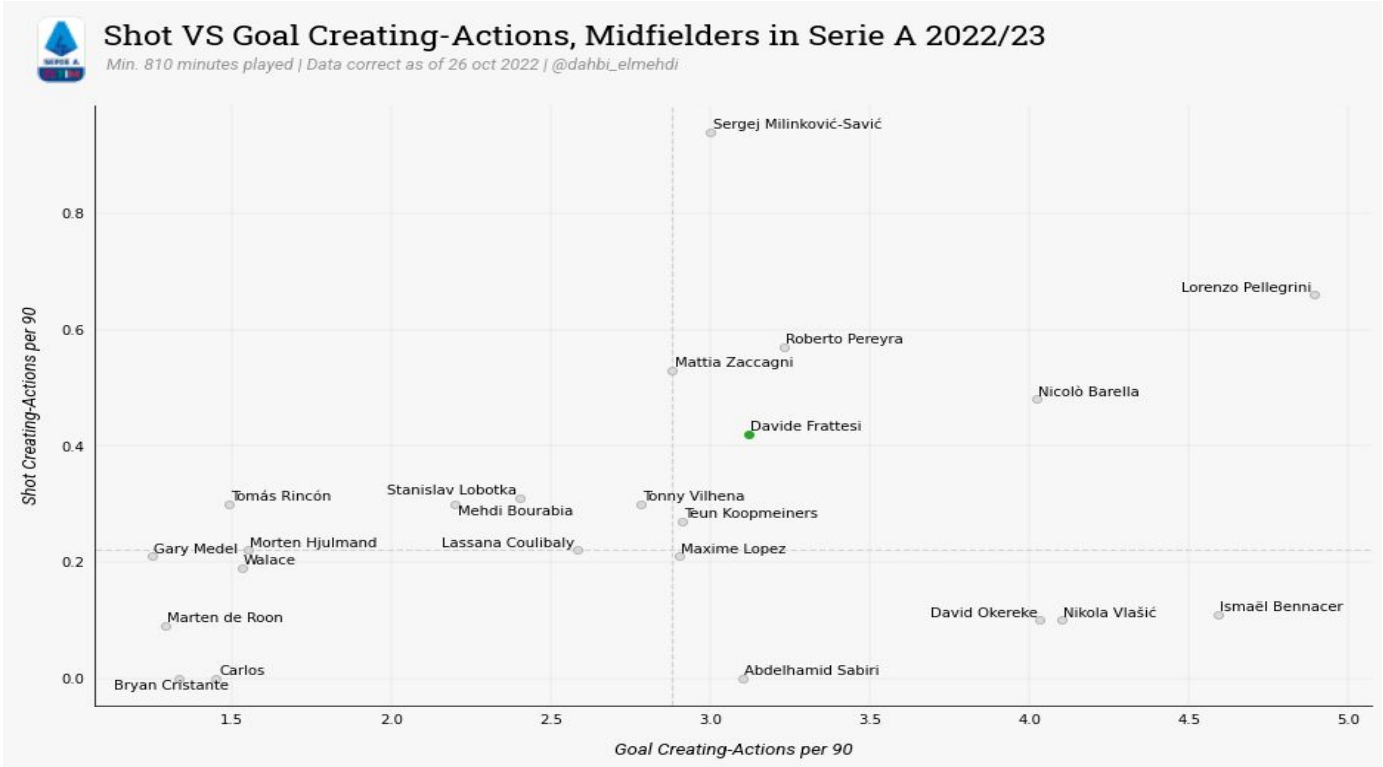
Creativity metrics: Shot/Goal creating actions

SCA is an advanced metric that tracks the two offensive/attacking actions that directly lead to a shot on goal.

The attacking actions can be:

- Live Passes
- Dead-Ball Passes
- Dribbles
- Shots
- Fouls
- Defensive Actions

Creativity metrics: Shot/Goal creating actions



Creativity metrics: Line-breaking passes

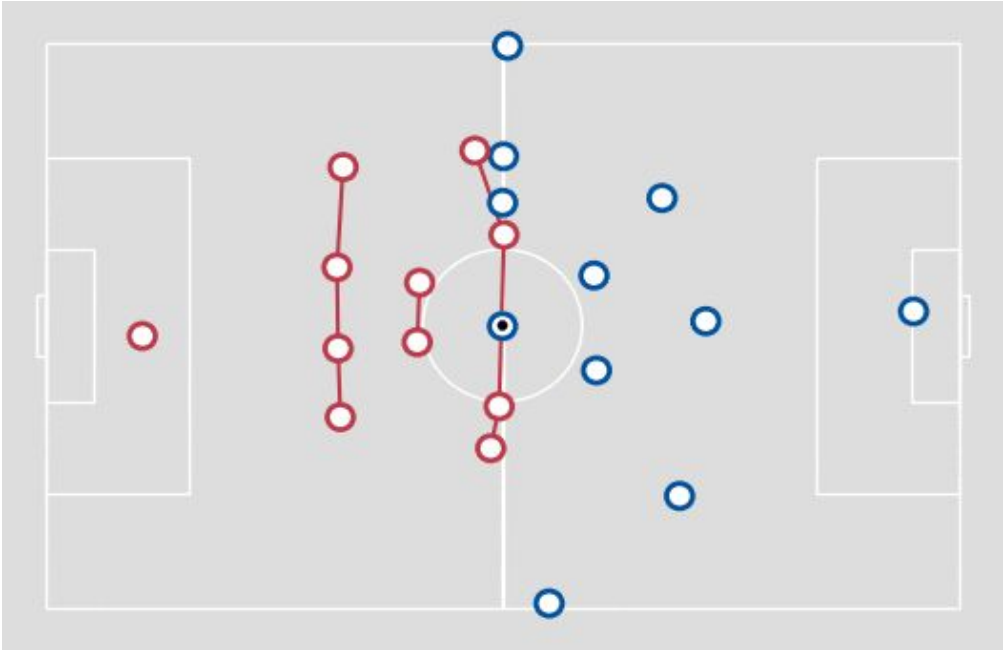
A line-breaking pass is a pass that break a line of the opposition formation. It is also a pass that:

- Progresses the ball 10 meters
- Has a starting point that is at least five meters away from the point of intersection
- Has an ending point of at least two meters beyond the deepest sitting player in the line

Line-breaking passes: Line formation detection

We can detect the formation line by applying a clustering algorithm such as the Jenks natural breaks optimization algorithm. We;

1. Combine both tracking and event data.
2. And passing the X location of players to the model.



Line-breaking passes: Successful/Unsuccessful labeling

In order to detect if an intercepted pass was a breaking-line pass or not, we implement the Weibull survival model.

Line-breaking passes: Quantification

To quantify line-breaking passes we use the VAEP (Valuing Actions by Estimating Probabilities) model that considers the following features:

- Angle view
- The maximum distance between adjacent players in the first opposition line in front of the player with the ball
- Line integrity
- Line compactness
- and other significant features that only tracking data provide

Creativity metrics: Through-balls

An attempted to pass between opposition players in their defensive line to find an on-rushing teammate (running through on goal).

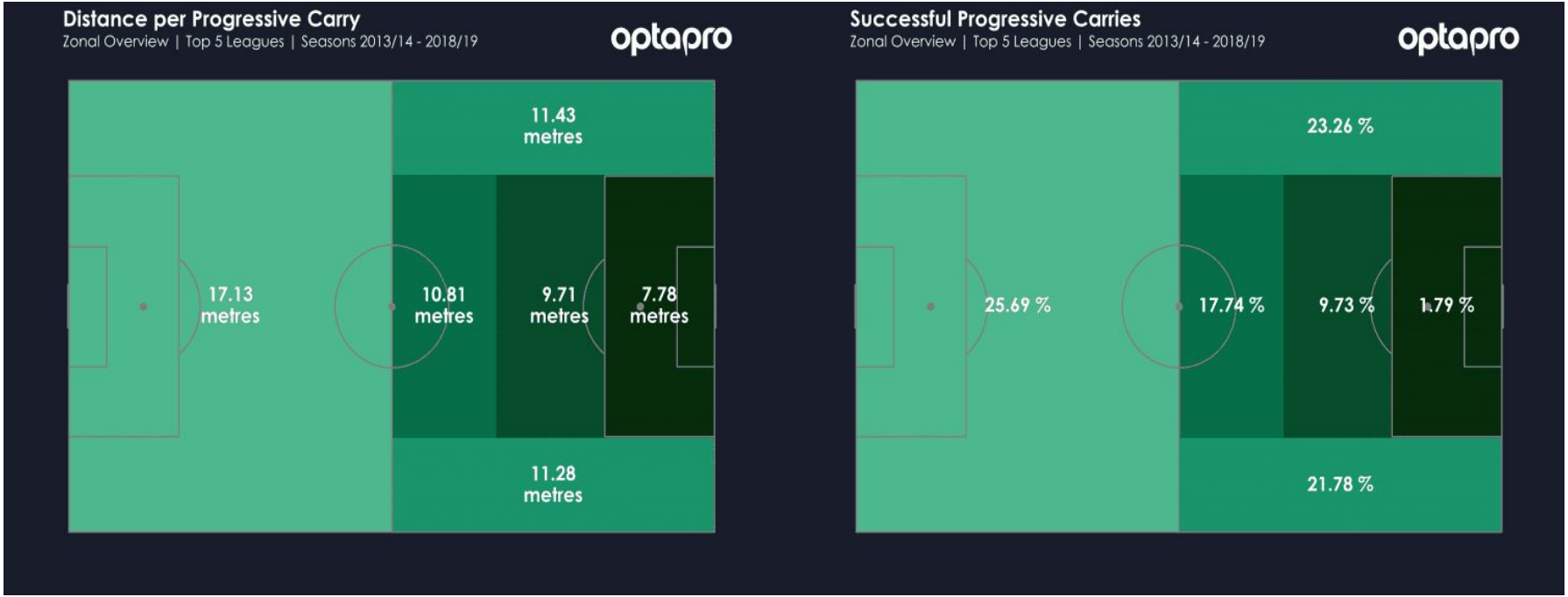
Possession metrics: Progressive passes/carries

Progressive passes are passes that move the ball at least 25% closer to goal.

The use of distance is paramount as it helps differentiate a simple forward pass from a pass that may require skill and vision to spot a teammate.

A carry is defined as any movement of the ball by a player which is greater than five metres from where they received the ball.

Progressive passes/carries: Pitch categorization

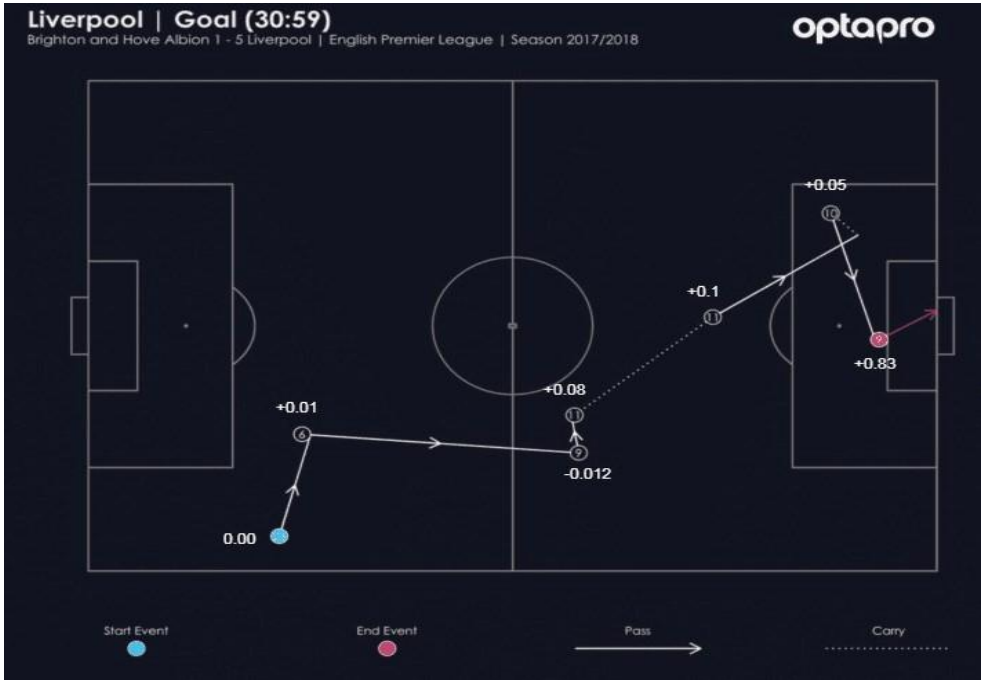


Possession metrics: VAEP

Valuing Actions by Estimating Values framework that provide a simple approach to valuing actions relying on machine learning predictive models.

$$V(S_i) = P^k_{scores}(S_i) - P^k_{concedes}(S_i)$$

VAEP: Use case study



Jersey No	Player	Action	Value	P scores
5	Roberstson	Pass	0.00	0.01
6	Winjaldum	Pass	+0.01	0.02
9	Firmino	Pass	-0.012	0.002
11	Salah	Carry	+0.08	0.05
11	Salah	Pass	+0.1	0.04
10	Mané	Pass	+0.05	0.19
9	Firmino	Shot	+0.83	1.00

Defensive metrics: PPDA

Passes Allowed per Defensive Actions is a metric that measure the intensity of a high press and that by;

PPDA = Number of Passes made by Attacking Team / Number of Defensive Actions

The defensive actions includes:

- tackles
- interceptions
- Challenges (failed tackles)
- Fouls

Defensive metrics: Forced Turnovers

Show when possession is lost due to pressure from an opponent. The more pressure teams and players apply to an opponent, the more likely they are to force a turnover of possession.

Scouting and recruitment

The main elements of scouting. What platforms football clubs use to monitor potential signings?

Scouting and recruitment: Approach

The main elements of scouting and player recruitment using cutting edge data

1) Squad Building
- Build a balanced football squad using data profiles

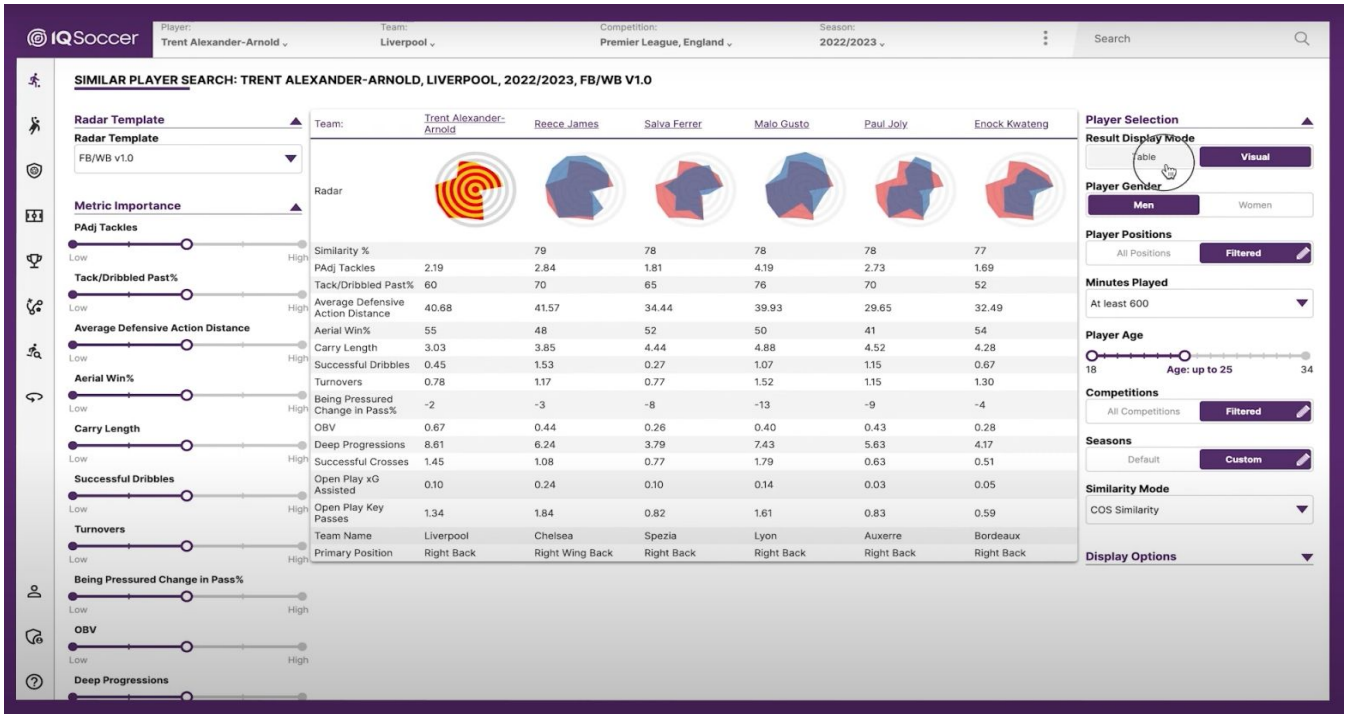
2) Scouting integration and workflow
- Integrating data into scouting, manage the data scouting system

3) Player Analysis
- How to analyze players at all types

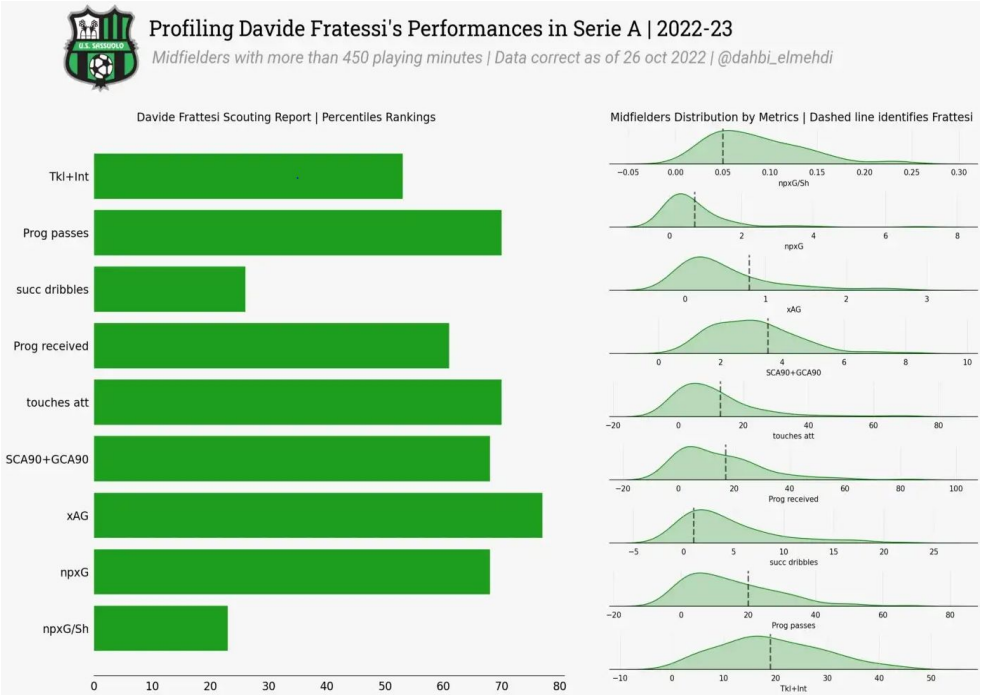
4) Player Evaluation
- Build a balanced football squad using data profiles

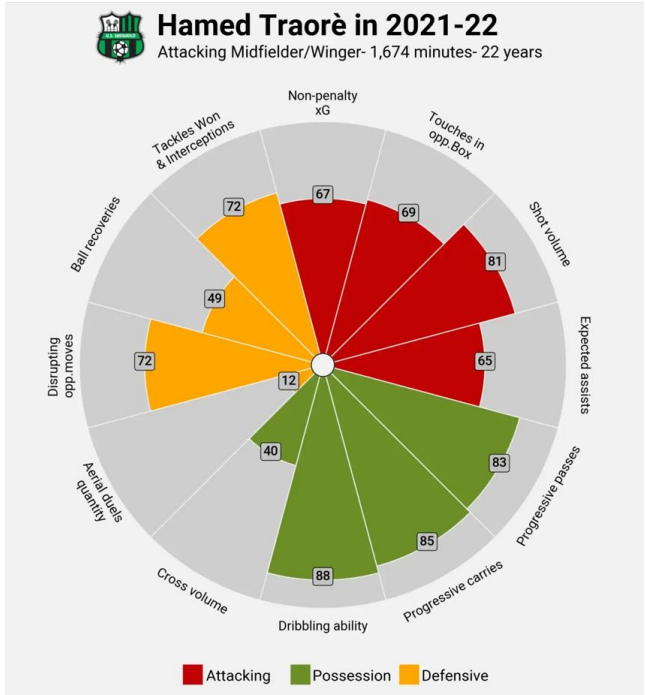
Scouting and recruitment: StatsBomb platform

Scouting and recruitment: StatsBomb platform



Scouting and recruitment: StatsBomb platform





Human Scouting vs Algorithms

Discussing their co-existence

Conclusion

Bibliography and Edd Webster’s living Legacy