Owen Waller

The Impact of AI calculated statistics on Football

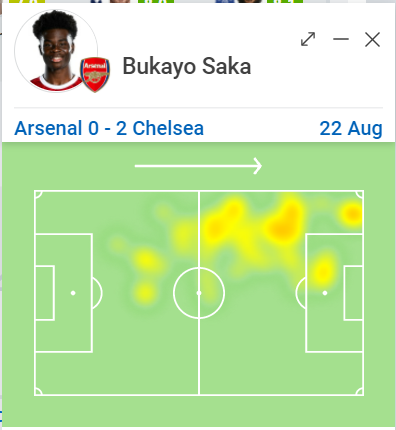




**Stats and their Creation**

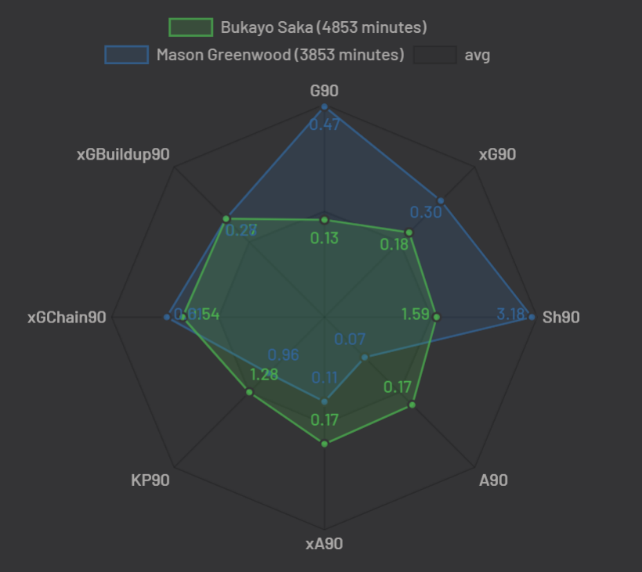
Since the early 1990s, the use of and proper recording of stats in football has become much more widespread, compared to the lack of data previously – the official amount of goals for the Brazilian centre forward Pele is still in question, ranging from 700-1200: a staggering amount (this range alone is more than any individual plyer has ever scored in the Premier League). This shift towards a statistics focused analysis of the game was embodied by the creation of Opta in 1996; its purpose was to record and analyse games in the Premier League, which had replaced the first division a few years prior. At this point the technology and algorithms available for collecting statistics was less developed, but the impact was still huge – for the first-time variables like touches and distance covered were being tracked. Since its creation, Opta has been utilised for a number of high-profile clients, such as Sky Sports as well as Arsenal FC and Manchester City FC. This growth encouraged other companies focusing purely on stats to open, and led to the creation of new stats and ways to measure a player or team’s ability. A great example of this is the possession stat, designed to indicate which team had control of the ball for longer, introduced in 1997 but quickly reinvented several times to help improve accuracy: the formula changed from simply using a clock to working out completed passes and available options for the ball carrier to pass to. The rapid development of the stats industry meant firms had to specialise, such as Wyscout (created in 2004) focusing more on the individual players stats rather than as a team, now having over 550,000 players in their database, and being used by a number of clubs.

This data is now more often used to compare players, whether it be for a club or simply for those interested. This led to applying algorithms, ranging from simply working out the goals scored per 90 minutes (in order to see who is more effective - a player could have scored more goals but at a slower rate) to seeing how they are to create a chance depending on their role on the pitch. Comparing players suitability for a role also spawned the idea of heatmaps, showing where a player spends most of their time on the pitch; these help to distinguish between players who have the same general position but serve a different role to the team:



The Heatmaps provided above (from sofascore.com) demonstrate this concept perfectly: both Bukayo Saka and Jack Grealish are considered left wingers, however the heatmap shows that Grealish prefers to hug the touchline and drive towards the opposition goal, whereas Saka tends to cut inside early and release the ball from the central areas. This would make choosing a player to fit a team’s current system much easier – a team that focuses on heavily possession-based football would opt for Grealish, as his runs to the edge of the pitch would spread the defence, whereas a counter attacking team may choose Saka as he releases the ball quicker in attacks.

Heat maps don’t serve as the only graphic when comparing players – attribute charts use players numbers to give a visual representation.



In this example once again both players are seen as wingers, however the graphic demonstrates Mason Greenwood is a more effective goal scorer, whilst Saka is better at creating opportunities for the team. The visualisation of these stats helps give a clearer image of who excels in which areas – as you can see the graphs for these two players are very different in shape, more noticeably than their relative playstyles would be on the pitch.

During the 2010s, player comparisons became a more global tool, rather than focusing on selection for clubs only – casual fans could use the comparison feature on the Premier League app or website in debates, whether it be for face stats or underlying numbers.

**What are the main xStats?**

Over the last 5 years in particular, the idea of using stats which can not be measured but only calculated has grown in popularity, introducing the growing idea of xStats (x for expected). This revolutionised the world of football stats, as it sought to remove luck or anomalies from the data – in other words, it acted as a representation of a player or teams stats in perfect and fair conditions, often revealing what bad luck hadn’t. The main implementation of this is xGoals, arguably the most important stat in deciding which team plays better in a match. The premise is simple: the xG of a shot is how likely the shot is to result in a goal, based on factors such as angle from goal, distance from goal and many more depending on the formula used. For instance, a shot from outside the box cutting in may only have a value of 0.1, whereas a close-range tap in could be as high as 0.9, and a penalty 0.76. Calculating the total xG from all a team’s shots is a good way to tell which team had better chances to score, even if they ended up losing the game in the end. The introduction of this stat was a game changer for clubs – they could see past the harsh realities of luck deciding football matches to see who the better player was. As one might expect, there are some scenarios where the formula will need to be tweaked. An example of this is when a team takes two high quality shots in quick succession – to sum them may result in an xG greater than 1 whereas in reality they could not have scored more than once in that attack. To work around this, the probabilities of each shot not resulting in a goal are multiplied, then this number subtracted from 1 to work out the xG for the whole attack. Adaptations like these for a number of scenarios helped to build xGoals into a very reliable and accurate statistic.

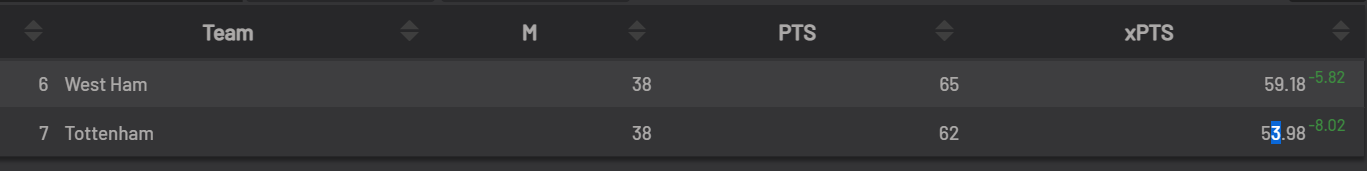
Luck is not the only factor to affect xGoals however – if a player is a very gifted finisher, they may consistently perform at a rate better than their xGoals suggests. A great example of this theory is Heung Min Son’s 20/21 season: for the previous 3 years he had the best xGoals overperformance in all of Europe, however, was not seen as a clinical finisher due to the lack of chances created for him. This changed when he burst onto the scene in the 20/21 season, leading the goalscoring charts for a good amount of the season and scoring 4 goals in one game – from only 2.1 xGoals. Harry Kane acting as a creator in this game meant Son was given opportunities and proved the value of xGoals once again.

The riveting success of xGoals led to further xStats being created – showing players true ability in areas other than goalscoring. A good example of this is the xAssists stat which shows how likely a player is to get an assist from any given pass:

xA = likelihood of player shooting \* xG if they had shot

Given the unreliable nature of the assist stat this proved very useful in seeing who was really creating more for a team – a player could get an assist for a short pass where the receiving player beats 4 players and scores, or not get one for beating 4 players and playing an inch perfect ball: there is too much reliance on the receiving player’s abilities. xAssists was quickly taken up and used by most stats companies as a replacement when scouting, alongside stats like Key Passes and Big Chances Created (the more traditional stats for measuring creativity outside of assists).

The other main xStat used is xPoints, which determines how many points a team would expect to win based on the chances they created in a game. Due to the discrete values of points normally (0, 1 or 3) this stat works best over a longer period of time, where it has time to balance out. A good example of this is Tottenham during the 20/21 campaign – they massively overperformed their xPoints for the first 6-7 weeks landing them in 1st place, however eventually it balanced out and they finished 7th: below West Ham who had been outperforming them on xPoints all season but were often behind on actual points.

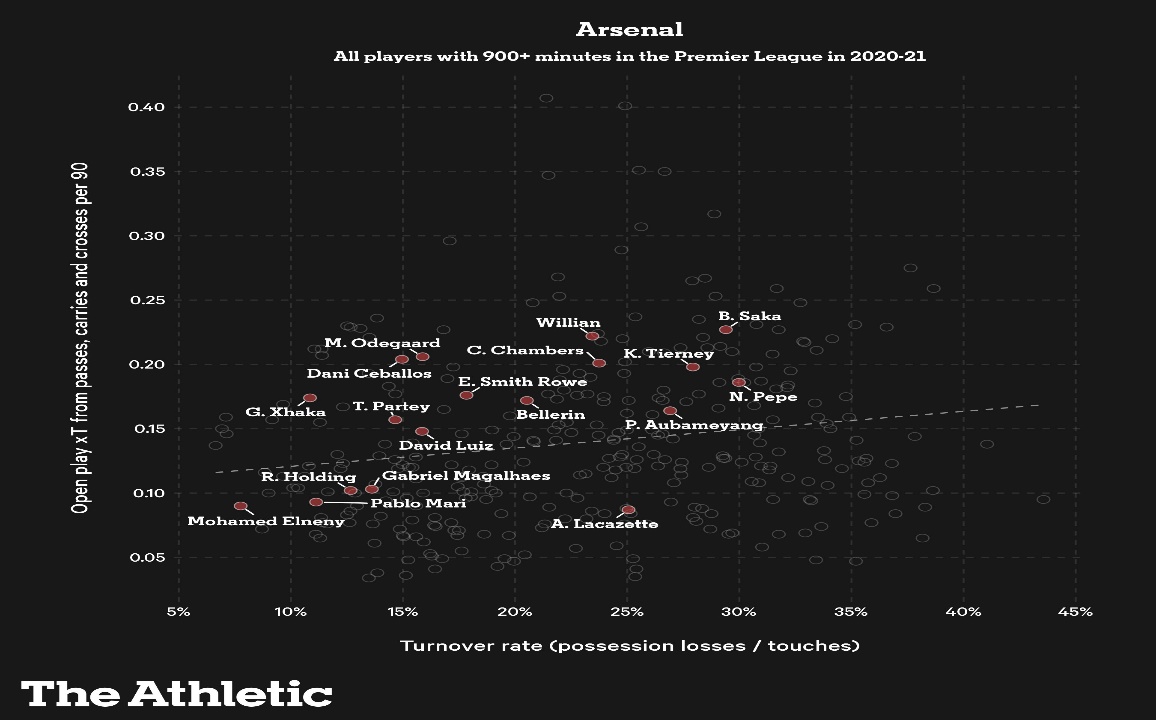


The main use of this stat is to assess a manager or coach’s ability, or a philosophy as a whole – if for instance using wingbacks results in higher xPoints but it is not currently working, the team may choose to stick with it for a while anyway.

**Advanced xStats**

Despite being more accurate than basic recorded numbers, these xStats aren’t perfect, and so even more developed stats have been developed, including xKP90, which shows how many key passes a player is likely to make per 90 minutes: a ridiculous amount of detail for something the naked eye may not even pick up. However, over time these stats also line up, for example the likes of Kevin De Bruyne having higher numbers of these whilst also being visibly one of the best creators in world football. Stats like these attempt to show as clearly as possible who is better to have on the pitch; another example would be xGoalChain, which shows the probability of an attack resulting in a goal based on their completed passes or dribbles.

Expected threat (xThreat) is the “next big football metric”, according to athletic journalist Tom Worville, and hopes to combine all of the key attacking metrics in order to formulate a picture of how effective a player is at progressing the ball into more dangerous spaces – either through ball carrying or dribbling.



This graph plots xThreat against turnover rate, which once again helps when deciding a player to start – from this diagram we can see that Martin Odegaard is very effective at contributing to the team’s threat without losing the ball, whereas Saka is more effective however also more of a risk taker, demonstrated by his higher turnover rate. Alex Lacazette’s location on the graph is also worth discussing, however in his role as a hold up play striker, he is likely to be under a lot of pressure when receiving the ball, accounting for the high turnover rate. On top of this, as a striker he is unlikely to be heavily involved in attacks until the end. We can use other stats to support this conclusion, for instance this heatmap and his goal stats for the season:

Alex Lacazette Heatmap (Premier League 2019/20)


Graphical user interface, application

Description automatically generated

This combination clearly demonstrates his position as a striker to us, and it a good example of how combining different metrics and diagrams can be used to give a full outlook on a player – an approach which has had a huge impact on how premier league clubs attract players.

* Impact on Transfers – Case Study: Aaron Wan Bissaka
* Brentford and the Moneyball Model
* Creation of Stat Industry

Impact on the Industry

**Impact on Transfers**

Prior to the popularisation of stats, players purchased from other clubs were mostly assessed based on how good their performances were to the naked idea, combined with very light stats – simple goals and assists, games won whilst they were present. Money was also much more present, with clubs basing a lot of their judgement based on the pricing of the player. This approach was still effective, due to the scouts expert knowledge often proving itself when selecting players. However, the introduction and popularisation of advanced and calculated stats has proven to massively improve recruitment at the top level, with many clubs dedicating an entire department to analysing transfer targets, or outsourcing to scouting agencies (see *Creation of the Stat Industry*).

Case Study: Aaron Wan Bissaka

Aaron Wan Bissaka in a Congolese right back who currently plays for Manchester United. He was signed by the club in June 2019 from Crystal Palace for £50m , in what was described as “the most thorough scouting process you could imagine”. After coming 2nd in the 2017/18 season, United needed a right back to replace the aging Antonio Valencia, a key weakness in the team. This led to the creation of a highly stat driven shortlist, including of course the Congolese defender. The player had a quietly good season, which may have gone under the radar if it were before the age of stats, especially given he only made 7 appearances. Despite this, his numbers were still better than United’s current right back, with more tackles, blocks and interceptions per game. Whilst this is likely to be skewed due to lower table teams having less possession, even when these stats are normalised by applying it per perentage of time defending , Wan Bissaka turns out much better numbers. Had these two players had seasons like this in the early 2000s, it is unlikely even a few would realise that the crystal palace man was a better fit for United.

Whilst United didn’t opt to sign the right back just yet, he was added to the shortlist as a serious candidate, and next season began to show why. He came 3rd in the league for total tackles, this year making 35 appearances and solidifying himself as one of the top right backs in the whole competition. His market value rose from being £4.5m in June 2018 to being £32m when United ending up signing him, where they were forced to cough up £50m to sign the player. This instance perhaps shows stats should be trusted more, as signing the Congolese defender the year before would’ve saved United big money.

Chart, line chart

Description automatically generated

Nevertheless, he repaid his price the next season, topping the leaderboards for total tackles with 129 once again, helping United to cement 3rd place and a spot in the champions league. Unsurprisingly, his attacking returns were limited with 0 goals and just 4 assists, however this was expected from his previous 2 seasons, and the player was not signed to be attacking like other fullbacks.

Overall, the case of Aaron Wan Bissaka’s transfer to United is a fascinating example of how stats can be used to get ahead of the curve, and should perhaps be relied on more heavily, to save money and provide teams with quality players that are the right fit for their system.

**Brentford and the Moneyball Model**

No team takes this sentiment more seriously than newly promoted Brentford, who adopt a “Moneyball” approach to recruitment. The definition of Moneyball, according to Billy Bean, is “Overlooking traditional received wisdom around player scouting to instead use data and find what the market undervalues.

Brentford’s moneyball philosophy began when the club was taken over by Mathew Benham, a stat fanatic who made money gambling through his firms “SmartOdds” and “Matchbook”. This involved using highly detailed statistical analysis to find fine margins which could tilt the balance of a game. This inspired Benham to use these tools elsewhere, rather than purely for betting gains.

The most famous tool used by Benham was xGoals for strikers rather than basic goals or goals per 90 stats. In short, Brentford are more likely to sign a striker with a lower goals tally and higher expected goals than one with more goals but lower expected goals, as the former is likely to be as good if not better, and will definitely be cheaper.

However, there are two key features which have to be considered with this approach, as stated by Brentford’s head of player recruitment Lee Dykes in an interview with the Athletic. The first is that the teams’ compositions need to be considered. On a simple level, if a players xG are mostly coming from crosses, signing this player to a team that doesn’t use wide players or is generally bad at crossing will not replicate the same results. The second is that a players development needs to be considered, both in who the player is playing with, and also if the current system will allow for further growth. The position this best applies to is centre back, and Brentford did bring in Pontus Jansson to help guide their younger defenders through games and further their development.

This may all sound like an ideological theory, but has already had several successes in the real world:



**Said Benrahma**

Signed from: Nice, £3 million

Sold to: West Ham, £23 million

Profit: £20 million

Position: Winger / AM

Key stat: xThreat

**Ezri Konsa**

Signed from: Charlton Athletic, £2.85 million

Sold to: Aston Villa, £13 million

Profit: £10.25 million

Position: Centre Back

Key stat: Post shot xG prevented

**Ollie Watkins**

Signed from: Exeter City, £7.2 million

Sold to: Aston Villa, £34 million

Profit: £26.8 million

Position: Striker

Key stat: xGoals

There are many more successes, and also players still at Brentford who have already proven their worth, most notably Ivan Toney, who scored 32 goals and got 10 assists in the 2020/1 championship season and is now valued at approximately £30m despite only being signed for £5m.

Chart, line chart

Description automatically generated

Brentford’s focus on statistical analysis is not purely used in recruitment however, and they employ Bernardo Cueva to give Thomas Frank (current manager) real time stat updates to allow for changes in the match. This was seen vs Arsenal on the opening day of the 2021/22 season, with the resulting formation change resulting in Brentford winning the match.

This stat focused model is undoubtedly the reason Brentford have progressed so far with little investment, and is likely to help them progress further and continue to stay in the top division of English Football.

**Creation of the Stat Industry**

The clubs themselves are not the only ones using new advanced metrics to generate more profit. Properly understanding and utilising these stats takes a great deal of time and resources, both in the technology and the analysts using it. For a lot of clubs, this setup cost is too great to open an advanced analytics department, and so they look to outsource all their analysis. This has created a gap in the market for firms to step in and provide the clubs with this service, and there are 2 main types of firm which do this.

1. Pure Stat Companies

These companies focus purely on collecting and generating the stats required by clients, whilst remaining completely objective. This makes them slightly more reliable, as there is no incentive for them to skew stats by normalisation: this could happen in stat companies which make a commission whenever their preferred player is signed by a client club. However, pure stat companies often provide the data in a more raw form, meaning further analysis may need to be done to figure out who the right fit for the club is

The best example of this type of company would be Opta stats, whose clients include the Premier League itself, who they now record stats such as distance travelled, and minutes played for. In 2013, Opta was sold for 60 million, and is now valued at approximately 75 million, showing the recent growth in the industry.

1. Statistical Scouting Companies

These companies don’t just collect raw data, but instead use it to draw conclusions and make recommendations to clubs as to which players or systems are best. Whilst they are more likely to have slightly skewed stats, they also provide context and draw parallels to the client clubs current players to make a decision. These types of companies are used by every club in the top divisions of Europe, and the football scouting industry is valued at other £1 billion as of 2018.

A good example of this type of company is driblab, which provides easily adaptable tools to be used to find the right players.

A person in a blue shirt

Description automatically generated with medium confidence

Overall, the creation of the stat industry and advanced stats have had huge impacts on one another – the design of new algorithms to refine stats drove the creation of firms and the industry, and this industry then drove the development of new stats and algorithms in turn. The football stats and scouting industry is worth over £1.5 billion today and this huge influx of money into football has impacted the sport down to its core – it makes up around 4% of the sport’s total industry net worth.

Impact on Fans

* Fantasy Football and Gambling
* Ruining the magic of football?

**Fantasy Football and Gambling**

Although naturally in a less intense and fine-tuned manner than professional clubs, fans of the sport can still use these new statistics and methods of analysis to their advantage. One of the most notable instances of this is seen in the top levels of the “Fantasy Football” league hosted by the Premier League itself.

A picture containing graphical user interface

Description automatically generatedFantasy Football, to give a brief summary, is where players select their 15-man squad from a starting budget to try and earn the highest points tally possible. Points are awarded to players for goals, assist, clean sheets and other key events throughout a match. During this current season, over 8 million people are signed up to compete, with numbers growing every season – this figure is over double the tally in the 2015/16 season.

The style of this game, in requiring players to think ahead to try and gain an advantage, lends itself well to the use of computer calculated stats in order to try and gain an edge over others in your league. The recent influx in stats has led to a much more high-level skill ceiling for fantasy football players, and the top competitive leagues will all be full of those using these stats. These often lead to players earning real winnings through the game, with competitive leagues set up with real cash prizes.

In a survey carried out with members of the top fantasy league “Always Cheating”, which offers up to £500 prize money and includes the current number 1 player, all 12 players surveyed agreed that xStats were key to modern day Fantasy Football, with 9 selecting the strongly agree option.

Chart, pie chart

Description automatically generated

The idea of using reward money is expanded on when talking about using these stats to earn money by gambling. Whilst there is always a risk associated, individuals or even firms may use these stats to spot “banker” odds to take advantage of – where the risk is much lower than the potential reward.

As mentioned earlier in *Brentford and the Moneyball Model*, the chairman of Brentford FC, Mathew Benham used to do just this, owning two firms which aimed to use stats to generate revenue from gambling, by providing or using advantageous odds. SmartOdds, the first of his companies, used statistical research to give advice to the highest level of gamblers, and Benham used this to turn £700,000 into excess of 3 million pounds. He was not the only one sharing in the profits, with many football fans also taking advantage of the software and making themselves profit.

His second company, Matchbook, was made for the benefit of fans, and acts as a betting exchange service which provides the best odds based on analysis and allows betters to trade bets with one another. The company is now estimated to be worth over £200,000, making a profit whilst also benefitting fans of the game.

It doesn’t take a huge company to effectively use these stats however. Any fan with a little time and faith can do research into a players stats or patterns of play in order to capitalise on good odds. For instance, arsenal were 17/20 favourites to beat Brentford, however analysis of Ivan Toney’s playstyle and the team’s xGoals from set pieces would’ve shown the casual better that Brentford would be a good bet vs a depleted Arsenal squad.