

USING DATA TO WIN IN FANTASY FOOTBALL

How data scientists top their
fantasy football league

2020

USING DATA TO WIN IN FANTASY FOOTBALL

HOW DATA SCIENTISTS TOP THEIR FANTASY FOOTBALL LEAGUE

THE BEST STRATEGY? GAMBLE ON MAVERICKS

Let's start by stating the obvious: to overtake the manager at the top of your league, you'll need to pick different players to them. 'Differential' players, as they're known in the fantasy football community, are key to success. But what type of players should you pick?

Those with a high average points tally might seem an easy starting point. But with a battery of statistics on player performance available online, it's unlikely there'll be a pool of untapped, high-scoring players that have not already been snapped up by other managers.

So, with only a few weeks left, it's time to roll the dice: if you want to move up the table, our experts suggest you **pick maverick players with a high volatility of points**. From a statistical point of view, gambling on mercurial players that have the capability of bagging super-high points totals is the most attractive strategy for the chasing pack at this stage of the season.

To explain why, here's an example:

'FATTER TAILS': HOW MAVERICKS CAN HELP YOU CATCH UP

Picture the scene: it's crunch time in your mini-league. To have any chance of catching the leader before the season ends, you need one of your differential players to score at least 12 points in the next Gameweek.

You scan your squad for players that the leader doesn't have. Who might bring in the points you need? You spot David Silva, **the safe bet**. Silva has been a solid pick, with consistent returns of 4 to 8 points most weeks. Sure, this gradual accumulation has helped you stay within touching distance of the top. But time is running out, and gradual accumulation won't help you now - will Silva score the 12 points you need this week?

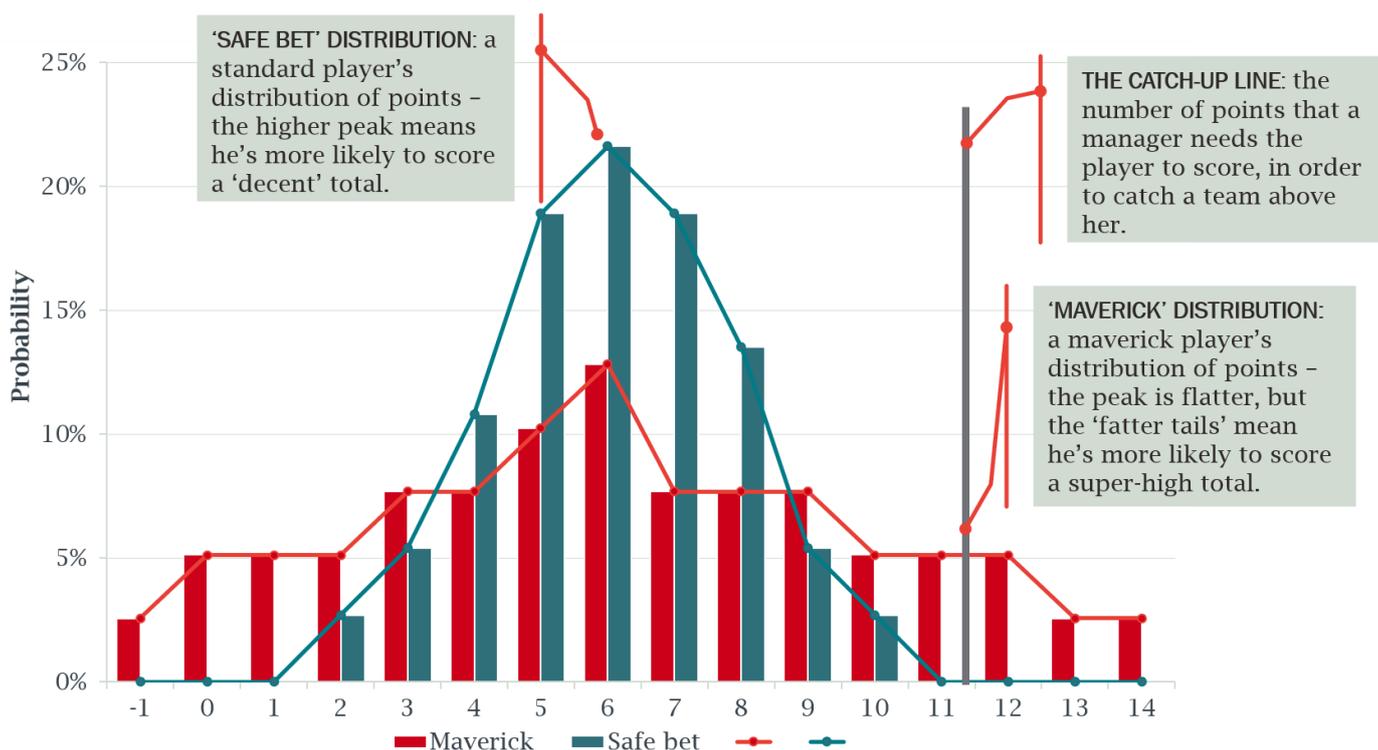
EXEC SUMMARY

It's been a while, but football is back. And with the imminent restart of the Premier League comes a return to action not just for players and fans, but for 7 million fantasy football managers too. With only a handful of weeks left in the season, there's not much time to climb the league table. So what's the best way to pick up points fast? Frontier's data scientists have crunched the numbers to provide the answers: which strategy to follow to catch your league leader, and which players to pick to do it.

The busiest managers can jump straight to the tips, but to any budding data scientist, please do feel free to interrogate our data analysis in the Data Deep Dive further below - and let us know what you think!

Your attention turns to Paul Pogba, **the maverick**. He only bagged 2 points at home to Crystal Palace, but then scored twice away at Burnley and notched up a heavenly 13 points. He's inconsistent, but wouldn't Pogba be more likely than Silva to score those crucial 12 points?

The chart below compares the two players by showing their points distributions: the probability that they'll score a certain amount of points. Silva is more likely to score a decent total - 5, 6 or 7 points - but is unlikely to bag a super-high score. Pogba's distribution has a flatter peak, but 'fatter tails' - he's more likely than Silva to collect a low score, **but also more likely to score a super-high total**.



PLANNING YOUR PICKS

So, picking mavericks is the best strategy if you want to climb the table quickly. But how many should you pick? And how do you find them? Our data scientists have analysed the numbers using statistical methods, and worked out the key steps.

STRUCTURING YOUR TEAM

QUESTION	ANSWER	REASONING	EVIDENCE
When should you start picking mavericks?	ASAP	The probability of catching up falls quickly as the number of Gameweeks remaining falls. Wait too long and you might leave yourself too much to do.	Monte Carlo simulations
How many should you pick?	2 to 4	To a certain extent, more maverick players means greater probability of catching up . But the benefit decreases the more you add, as the probability of low scores starts to offset the probability of high scores. The optimal number is 2 to 4 maverick players – the further behind you are, the higher the number you should pick within this range.	Monte Carlo simulations

IDENTIFYING MAVERICKS

QUESTION	ANSWER	REASONING	EVIDENCE
Where do they play?	Wingers	Goal-scoring midfielders are the most likely to score super-high points totals, and wingers are the most likely midfielder type to score goals.	Machine learning
How much do they cost?	> £6m	High cost players score super-high points more often. But, as long as you stay above £6m, there are also plenty of high-volatility players at the lower end of the scale.	Machine learning
What about fixtures?	They matter – a lot	Super-high points scores tend to result from goals and assists – players facing weaker teams are more likely to score or assist .	Logit regression analysis
What about form?	It matters – a bit	Same reasoning as above – in-form players are more likely to score or assist . But form matters less than fixtures.	Logit regression analysis
What about teams ‘on the beach’?	Doesn’t really matter	At the end of the season, some teams are referred to as ‘on the beach’ – with little left to play for, and lacking motivation. But our analysis shows that, in fact, players lining up against these teams are only slightly more likely to score super-high points totals.	Logit regression analysis

WHO ARE THE BEST MAVERICK PLAYERS?

Now that you’re armed with the strategy, it’s time to start making your picks. To help you, our team at Frontier used a combination of statistical and econometric techniques – **Monte Carlo simulations, machine learning** and **logit regression analysis** – to reveal the players most likely to catapult your team up the table when the season kicks off again on 17 June.

Focusing on goal-scoring midfielders and upcoming fixtures, here is the list of players that we think hold the key to your hopes of clinching the title:

Player	Average Points	Cost (£m)	Position	Highest Points Score	2nd Highest Score	Next Fixture 1	Next Fixture 2	Matches left
Jack Grealish	4.2	6.4	MID	16	13	SHU(H)	CHE(H)	10
Harvey Barnes	4.1	6.2	MID	19	11	WAT (A)	BHA (H)	9
Ismaila Sarr	4.1	6.4	MID	16	14	LEI (H)	BUR (A)	9
Mo Salah	7.2	12.7	MID	16	14	EVE (A)	CRY (H)	9
Matt Doherty	4.3	6.3	DEF	14	11	WHU (A)	BOU (H)	9
Trent Alexander-Arnold	5.7	7.8	DEF	24	12	EVE (A)	CRY (H)	9
Raheem Sterling	4.9	11.7	MID	20	14	ARS (H)	BUR (H)	10
Anthony Martial	5.3	8.0	MID	15	11	TOT (A)	SHU (H)	9

One quick caveat to bear in mind: this list was generated based on analysis of historical Fantasy Premier League data from previous seasons. As with all data analysis, managers should consider how the present Premier League situation is different to that of previous seasons, and whether these insights extend to the current situation.

ENJOY THE RIDE

So there you have it – those are the mercurial points-scorers who we think could help you stage a late comeback in your mini-league. As all fantasy managers know, it will be important to have a strategy for the coming weeks, and the data suggest that gamblers may be rewarded.

Hopefully picking maverick players will be an effective strategy for you, but remember to remain flexible. Whatever happens in the final few weeks, one thing’s for certain: it’s the rollercoaster ride of fantasy football that makes it fun, and throwing in a few maverick picks each week can only add to the entertainment.

THE DATA DEEP DIVE

Our fantasy football tips are based on comprehensive analysis by our data scientists. Here, in our Data Deep Dive, we've highlighted some of the statistical and econometric methods they used.

We see these methods as complementary: **Monte Carlo analysis** explores the impact of a greater volatility of points on a manager's prospects of catch-up by simulating what would happen in thousands of possible ends-of-season. Unsupervised **machine learning** helps identify who the maverick players are by sorting all players into different groups, based on the volatility of their points scores. **Logit regression** analysis isolates the additional impact of form and easy fixtures on the probability of an outcome (in this case 10 points scored) occurring. These methods answer different questions in different ways, but together they provide a range of tips to increase your chances of catching the leader.

Before jumping in, it's also worth reiterating that these are not normal times for the Premier League. We've run checks to make sure our results replicate to other scenarios, but as the season plays out it'll be interesting to see whether insights based on historical data extend to the current situation.

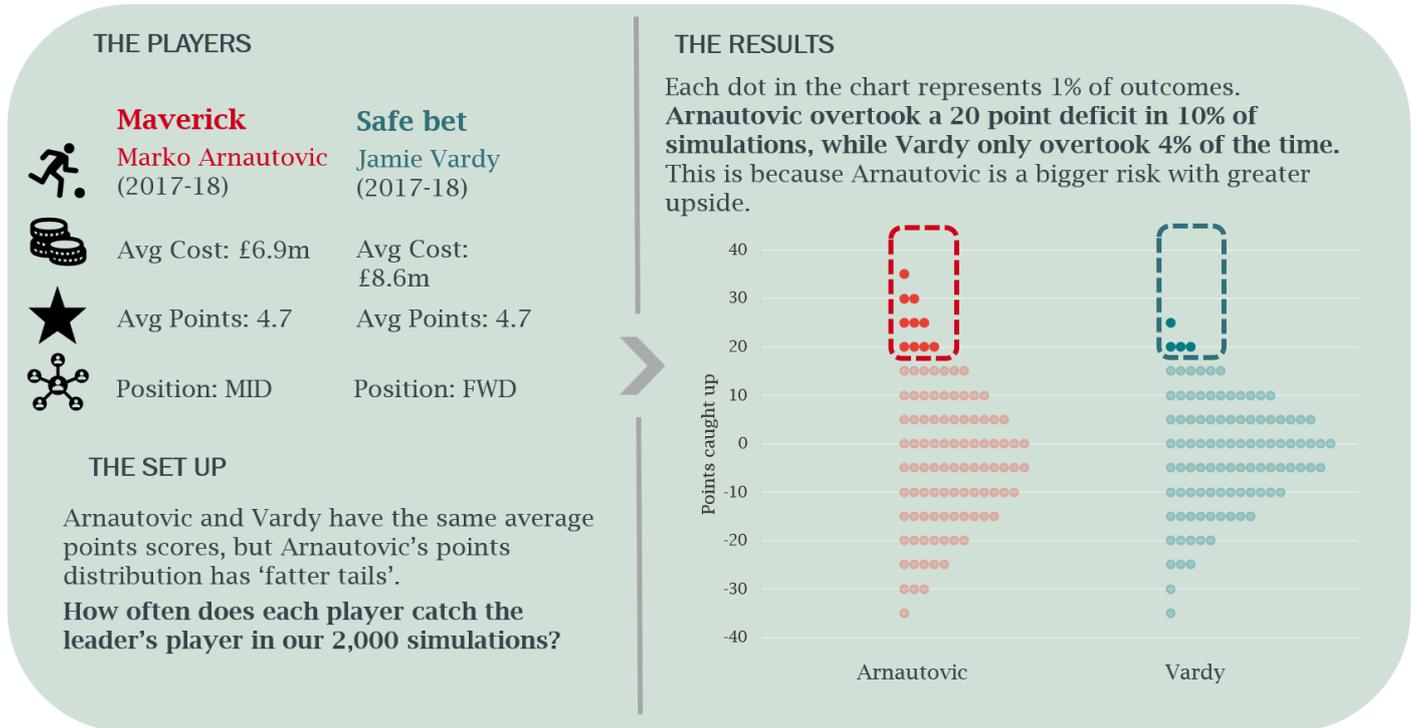
MAVERICKS VS SAFE BETS – MONTE CARLO SIMULATION

We suggested above that maverick players increase your chances of climbing the table when compared with safe bet players. So how did we calculate the odds?

We used a technique called **Monte Carlo simulation**, playing out thousands of possible ends-of-season for pairs of maverick and safe bet players. For each player, we recorded a stat: the proportion of ends-of-season where they overcome a certain points deficit - for example, 20 points - against a base player¹. Since each pair of representative players has the same average points (see the examples of Marko Arnautovic and Jamie Vardy below), any difference in the percentage of catch-ups across players is due to differing points volatility.

¹ In the example below, this base player was Wilfried Zaha from the 2017-18 season. He scored 4.7 points per game.

USING MONTE CARLO ANALYSIS TO CALCULATE HOW OFTEN ARNAUTOVIC AND VARDY CATCH UP



KEY INSIGHTS

- Picking mavericks substantially increases your chances of catching up. Arnautovic (10% chance of catch-up) is more than twice as likely as Vardy (4% chance) to catch up 20 points on his own.
- The ratio of a maverick's chances compared to those of a safe bet is even larger with fewer weeks left and a bigger points deficit.
- But don't wait around – it's best to pick a maverick ASAP, since the probability of catching up a points deficit decreases rapidly as the number of weeks left falls.

HOW MANY MAVERICKS?

Our advice was to pick between two and four maverick players – again, this was based on our Monte Carlo simulations.

A second maverick is almost always helpful: repeating our simulations for two mavericks² catching up the same 20-point deficit across seven weeks, the mavericks succeeded 32% of the time. Two safe bet players³ caught up in 22% of simulations. This means that adding a second maverick increased chances of catching up by 18 percentage points, compared to an additional safe bet increasing chances by 14 percentage points.

² Arnautovic and Pogba catching up Zaha and Costa.

³ Vardy and Eriksen catching up Zaha and Costa.

For managers a modest distance behind the leader, say between 10 and 30 points, it's best not to get too greedy. Our analysis showed that adding more than two mavericks will have a diminishing effect on your chances. Each maverick you add has an offsetting effect: a low score from the second maverick may coincide with a high score from the first maverick, cancelling out the gain from the first player's points volatility. This offsetting effect isn't too large when moving from one to two mavericks, but it becomes greater the more you add.

FINDING THE MAVERICKS – MACHINE LEARNING

To find out what makes a maverick player, we used unsupervised **machine learning** to form groups of similar players. This was to make our lives easier – there are lots of different players to choose from, so grouping similar players made it simpler for us to identify the mavericks.

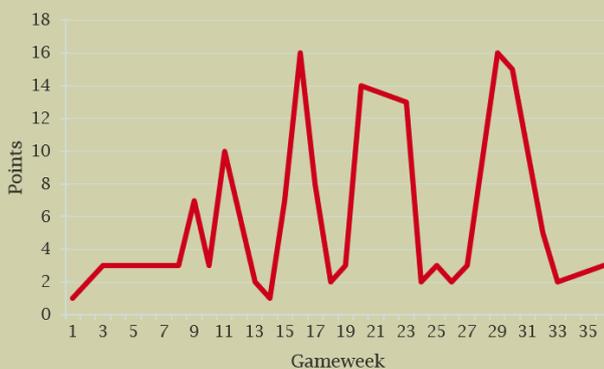
Here is a short description of how the machine learning technique works. For a plot of data for all player-and-season combinations (we call this the “data space”), machine learning uses an algorithm to draw boundaries around data points, thereby identifying a set of groups. This is done for a specified number of groups – we chose four. The boundaries of each group are drawn so that each player-and-season observation is a close distance to the centre of any one of the four groups. Once the boundaries are set, players are allocated to whichever group they sit within – or more precisely, are closest to.

We observed that players in two of the four groups had similar average cost and points, but one of these groups has the players with the greatest points volatility in the dataset – this, therefore, is the maverick group. We compared the characteristics of players (e.g. Position) across the maverick group and the comparison group (which we also refer to as the safe bet group) to determine what is different about mavericks. Below is a summary of the players in these groups.



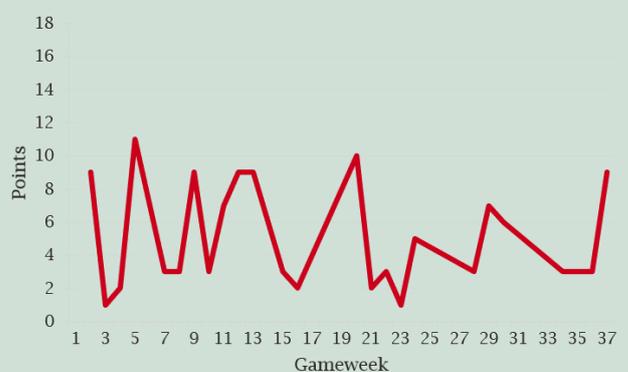
The **maverick** group has **goal-scoring midfielders**, who tend to be **wingers**, or forwards categorised as midfielders. They score 10 points once every 6 games.

A good example of a maverick is **Heung-Min Son**, from the 2017-18 season. His points scores are below; Son scored 10 points 6 times, and a whopping 14 points 4 times.



The **safe bet** group has mainly **forwards and creative midfielders**. They average 4.4 points per game, but score 10 points less than once every 10 games.

David Silva from 2018-19 is a classic safe bet player. He averaged a healthy 5.1 points per game and had lots of scores between 5 and 9 – but he only reached 10 points or more twice.



KEY INSIGHTS

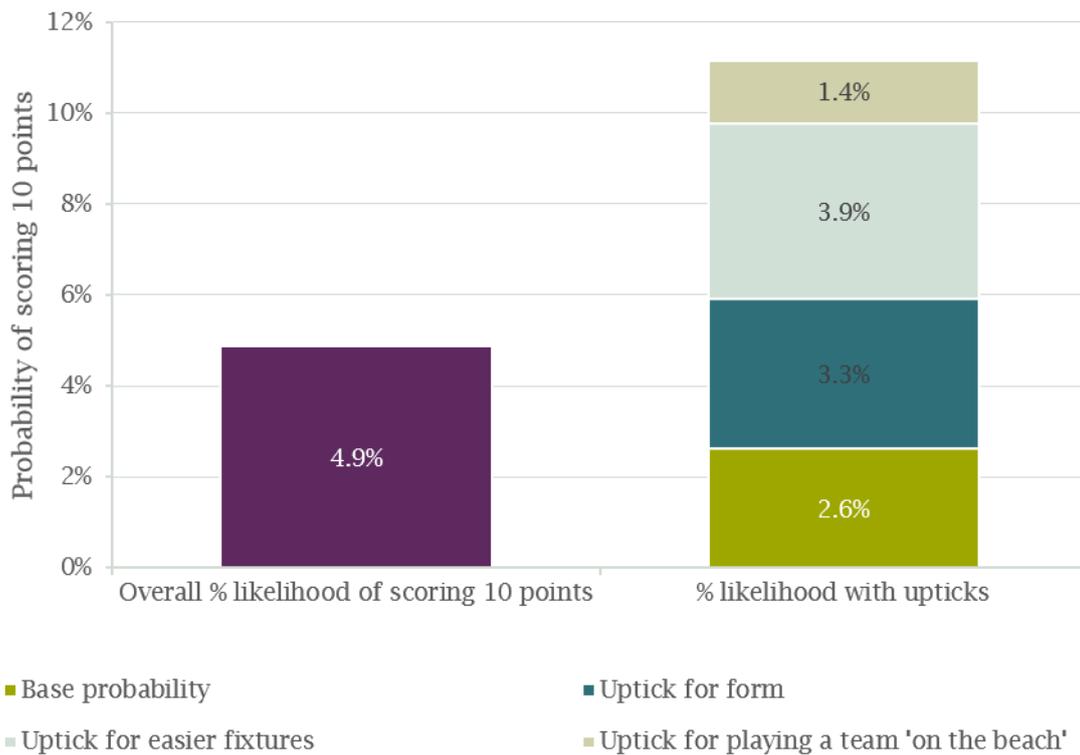
- What differentiates midfielders in the maverick group is that they score more goals. We checked where these midfielders play on the pitch – they’re often wingers.
 - Forwards are split evenly across both the maverick and safe bet groups, so tread carefully and be sure your hunch is right when presuming a forward is a maverick.
 - High-cost players score super-high points more often. But, as long as you stay above £6m, there are also plenty of good-value high-volatility players.
-

ANALYSING FORM AND FIXTURES – LOGIT REGRESSION

To calculate the impact of form and ‘easy’ fixtures (those against weaker teams), we employed **logit regression** analysis. This technique is used to estimate the probability of a particular outcome occurring and to identify whether different factors affect the likelihood of an outcome. It also estimates the size of the individual impact that each factor has on the probability of the outcome.

Our data scientists used logit regression analysis to estimate the likelihood that different players score 10 points in a certain Gameweek. This included estimating the uptick in probability for in-form players and players with easier fixtures scoring 10 points. We also analysed whether playing against a team that’s ‘on the beach’ – one with nothing left to play for – increases the likelihood of a player scoring 10 points. Our definition for being in ‘good form’ was scoring 8 points or more in the previous week; we categorised easier fixtures as those with an Fantasy Premier League difficulty rating of 2 or 3 (as opposed to 4 or 5); and we manually identified teams that could have been ‘on the beach’.

The chart below presents our results. The left-hand bar is the overall probability that any player will score 10 points, irrespective of form or fixtures – so roughly 5% of players score 10 points in any week. The right-hand bar starts with the conditional probability that a player that is not in form with a hard fixture scores 10 points (2.6%), and then adds on the impact of a player being in form and having easier fixtures. Our results suggest that form and fixtures should not be ignored – players that are both in form and have an easy fixture could be twice as likely as the average player to score 10 points. Playing against ‘on the beach’ teams, however, has a limited impact.



KEY INSIGHTS

- Players with 'easier' fixtures are **2.5 times as likely** to score 10 points as the base player.
- In-form players are **2.3 times as likely** to score 10 points as the base player. But view this stat with caution: when we controlled for the underlying ability of players in our analysis, the impact of form fell substantially.
- Players facing 'on the beach' teams are only slightly more likely to score 10 points – a testament, perhaps, to footballers' professionalism.

AUTHORS

ROBERT WOODHARD
Consultant

WANT TO KNOW MORE?

WWW.FRONTIER-ECONOMICS.COM

HELLO@FRONTIER-ECONOMICS.COM

+44 (0) 207 031 7000