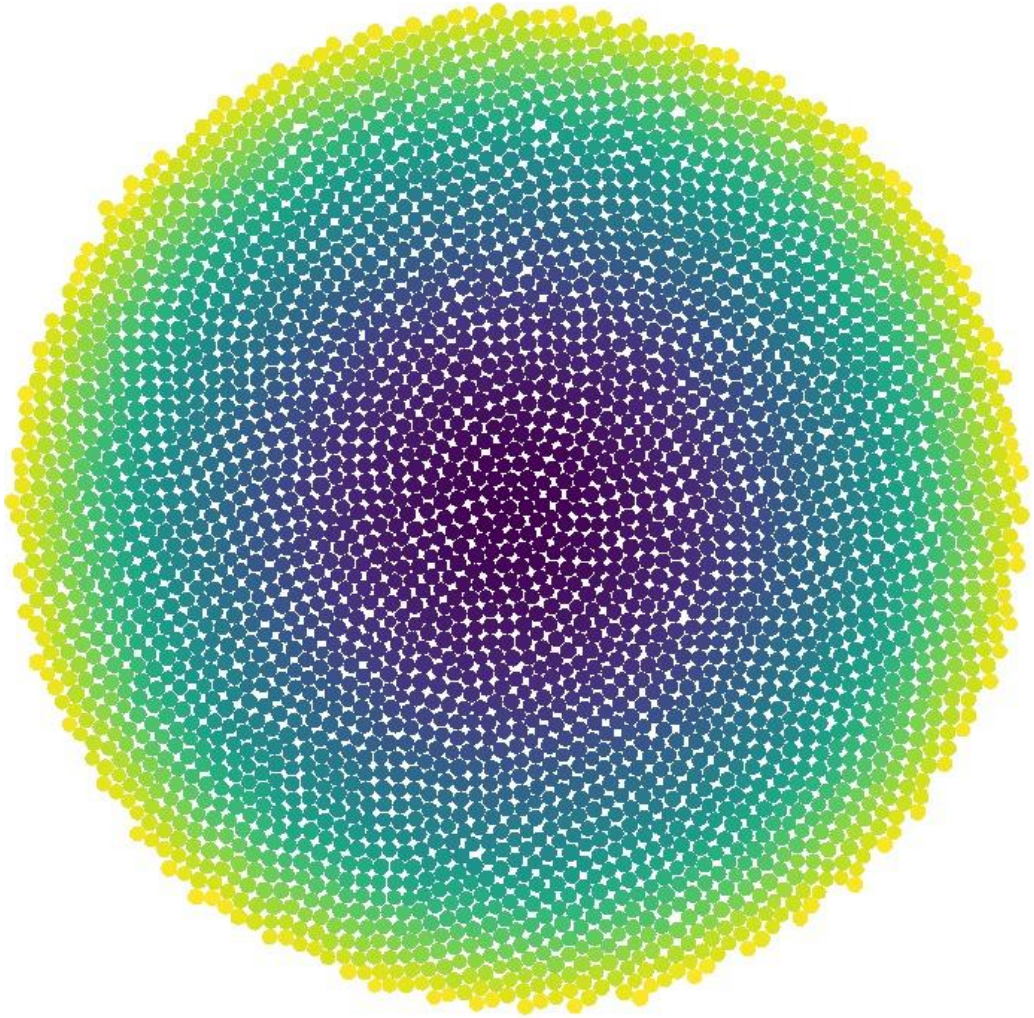
How Accurate Is FIFA At Predicting Future Stars?

And, what can FIFA (the game) tell us about real world football?



Amin Sabri Al-Ait

09.09.2021
Data Science Msc.
Sports Data Seminar

Table of Contents

I INTRODUCTION

The idea of having an almost-monthly updated metrics of football players' abilities, skills and overall scores seemed intriguing and may have laid the foundation for further studies that can empower the entire football sport community as a whole.

Fortunately, football clubs collect this data, analyze it, and make strategic team formation changes and market transfer/loan decisions based on the gained knowledge. These clubs have been doing so for quite a while now. Unfortunately, each club keeps their data as the guarded secret no other team should put their hands on. Especially not their town rivals.

The inspiration for this study came after realizing that there is a *second best* option out there. An option that saves us from having to knock on each football club's door and ask them for their hard earned players' data.

Electronic Arts Inc.

Electronic Arts Inc. (EA) is a video game company that develops a wide range of video game titles. One segment/series of their productions is EA Sports which only focuses on annually releasing sports games with updated metrics and statistics for the players of these sports, and of course, better graphics to catch up with the next generation console gaming machines such as PlayStation and XBox. Examples of these sports games series are: the FIFA series, NHL series, NFL series, Madden series, F1 series etc…

EA FIFA

The first EA FIFA series game was released back in December 15th 1993, and the latest release, FIFA 22, is set to be released on September 27th 2021. Between these years the game has had major changes, especially when it comes to player uniqueness: it is possible to inspect all their abilities and skills' scores (out of 100) in a profile view, unique to every player. The process of how EA FIFA attains all this data and assigns all these scores is explained in articles that can be found on sports websites: Goal (Murphy) and VG247 (Saed).

The Stats

As previously mentioned, every player now has unique profiles with unique scores for skills like agility, jumping, shooting, freekick accuracy, defending etc… There is one special score EA FIFA has been putting onto players' profiles which is **Overall**. As the name suggests, this metric refers to the player's score out of 100 as a complete football player.

Another very interesting and important score is **Potential.** This metric is calculated by the game and it claims that this player's Overall score will reach this Potential score some time in the future. For example, take player John Doe: At EA FIFA 2011 Jon was 22 years old and had an overall score of 75, and potential score of 80. Skip forward 5 years (and 5 EA FIFA games) and John is 27, with overall 83 (> 80). Meaning he has reached and went beyond what EA FIFA has put up for him (5 years ago) by 3 overall points.

II Exploratory Data Analysis Tour

For us to tackle the myth properly, we have to take an exploratory data analysis tour:

The Players

Passing the Potential & Career Length

To tackle the title “How Accurate Is FIFA At Predicting Future Stars?”, we first have to build a foundation in understanding our data. This comes in the form of partitioning the players in two dimensions.

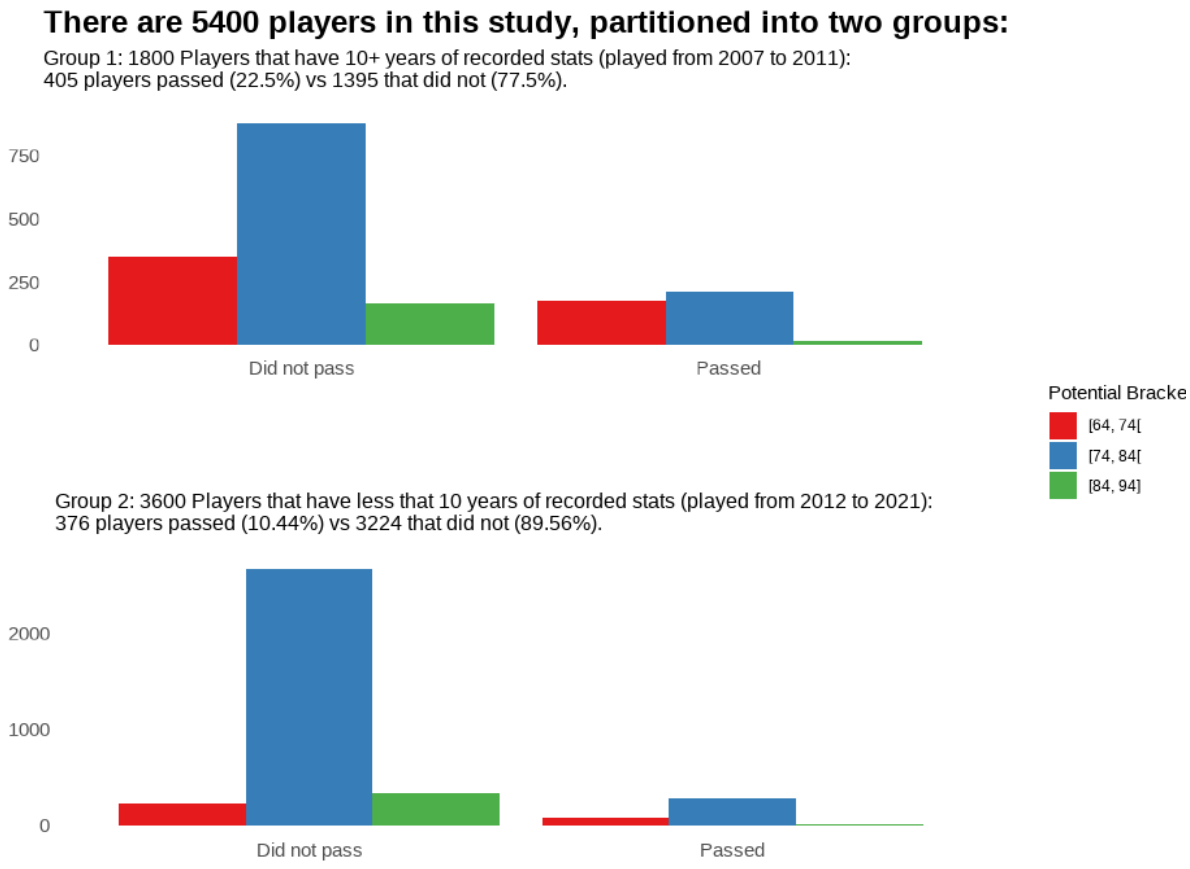


Figure 1

The first is potential passing:
If, throughout the player’s career, the player’s overall reach and bypasses the potential score, then the player is considered to have **passed**. Otherwise, the player is considered as **Did not pass**.

The second partitioning divides on the basis of the player's career length. It puts players that (should) have been playing for 10 years or more into one side, and the rest into the other. This is to make it clear that players of the first partition have had enough time to reach, and perhaps pass, the potential put forth by FIFA.

Figure 1, presents different bar charts that players fall in: We observe an inflated blue **[74, 84[** potential bracket that represents players that did not pass, regardless if they have been playing for 10 years and more or less than 10 years. This tells us where the majority of players' potentials are residing.

Furthermore, we observe that it is almost as likely for players with potential in the **[64, 74[** potential bracket to achieve their potential as it is for players in the **[74, 84[** potential bracket for players that have been playing for 10 years or more.

The same can not be said about players that have been playing for less than 10 years, as only players from the **[74, 84[** potential bracket have achieved their potential with their number being relatively higher than those in the other 2 brackets.

Overall & Potential

Another important aspect we need to touch upon is how are the players' Overall and Potential scores distributed. Figure 2 helps in understanding the bulk mass of players that were situated in the **[74, 84[** potential bracket of figure 1.

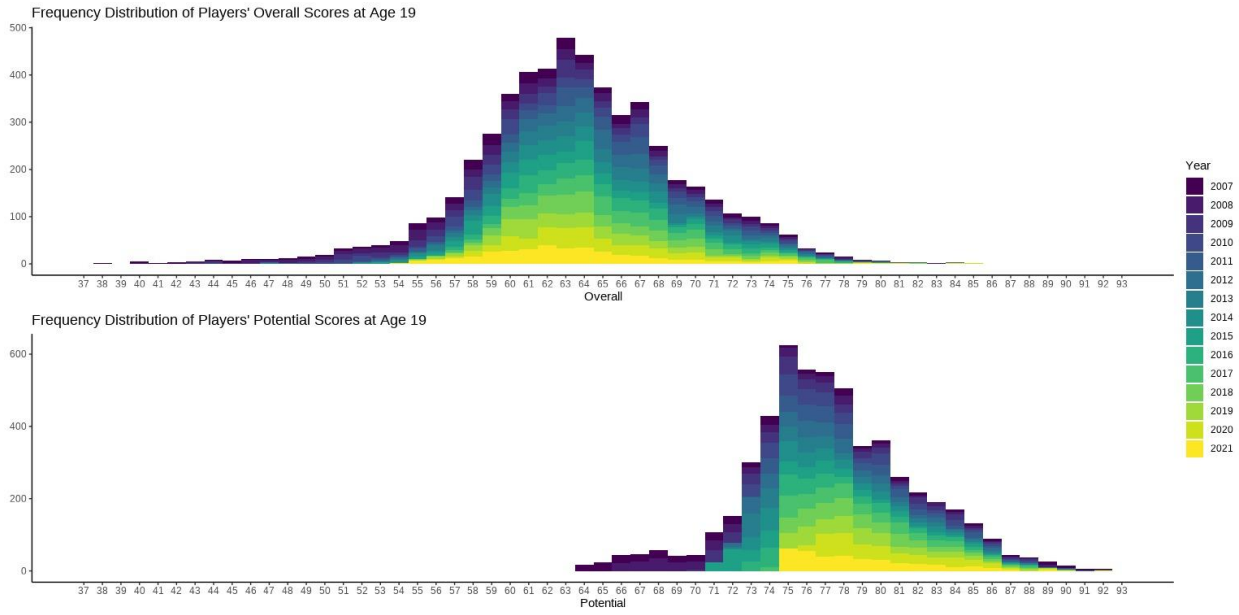


Figure 2

Both distribution plots of figure 2 are stacked barcharts with the Year (year when the players were 19 years old) is the categorical variable. The first plot shows the distribution of all 5,400 players' overalls at their respective age of 19. Similarly, the second plot shows the distribution of the players' potential at their respective age of 19.

It is natural to see that the mode of the first distribution lies on the overall value of 63, while that of the second distribution lies on the potential value of 75, and that is because the game should typically have a player's potential to be higher or equal to the player's overall. Thus the second plot is shifted to the right of the first in terms of x-axis values.

It is also important to note that both distributions are almost bell-shaped, meaning that if we might have had more observations in our study then the distributions would eventually take the shape of a Normal distribution.

Age

1. At which age are players reaching (and passing) their potential?
2. At which age are players reaching their prime?

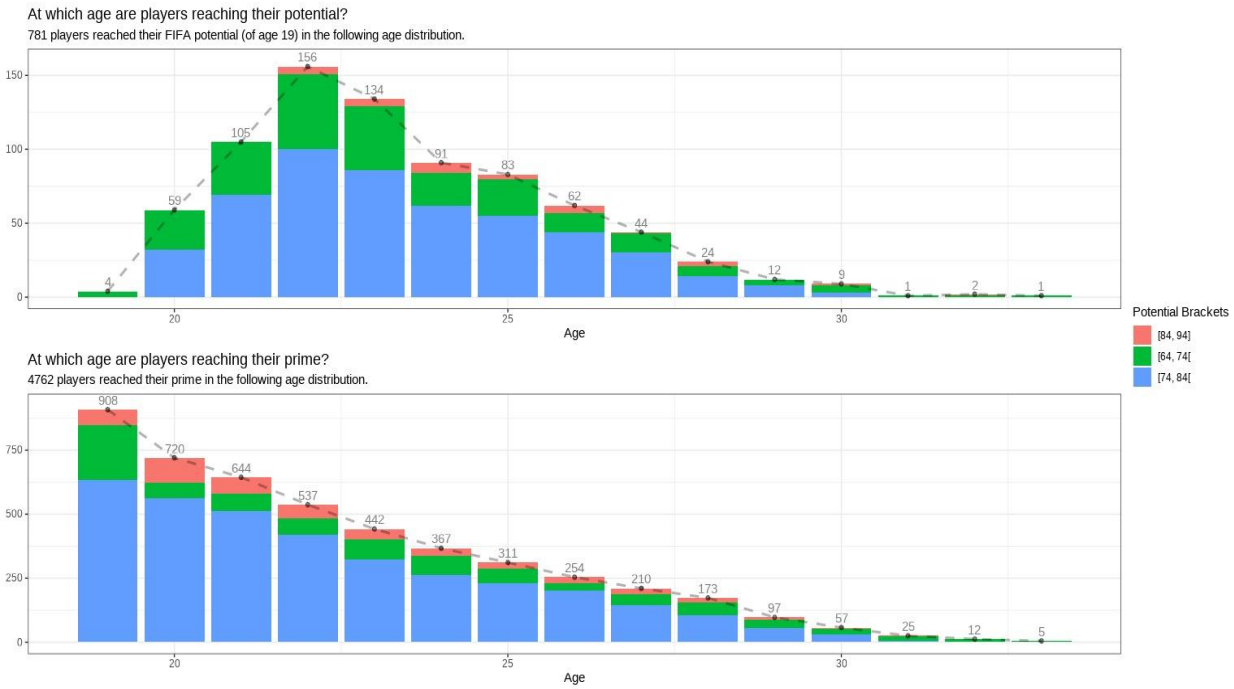


Figure 3

First plot of Figure 3 helps us visualize the ages of when players reach (or even pass) their potential. Peaking at the age of 22 then slowly decreasing till the age of 33. The few observations of age 31, 32 and 33 are very remarkable since in the world of football, one would not expect any player to shine if they are 30 or older, but these players have defied

expectation*.

Second plot of figure 3 shows a smooth decline in numbers of players that have reached their prime when moving from left to right. Perhaps the most remarkable insight we get from the graph is its shape that basically indicates: regardless of which potential bracket a player falls in, the younger he is the more likely it gets for him to reach his prime overall. This is remarkable because it is widely known in the world of football that players reach their prime in the age 27-29. This BBC article goes in-depth to explain why.

Again, the bulk quantity of players that fall in the [74, 84[potential bracket can not be neglected, making the majority of each bar of both plots of figure 3.

What does it take for players to reach their potential?

For what given potential are players most likely to reach it?

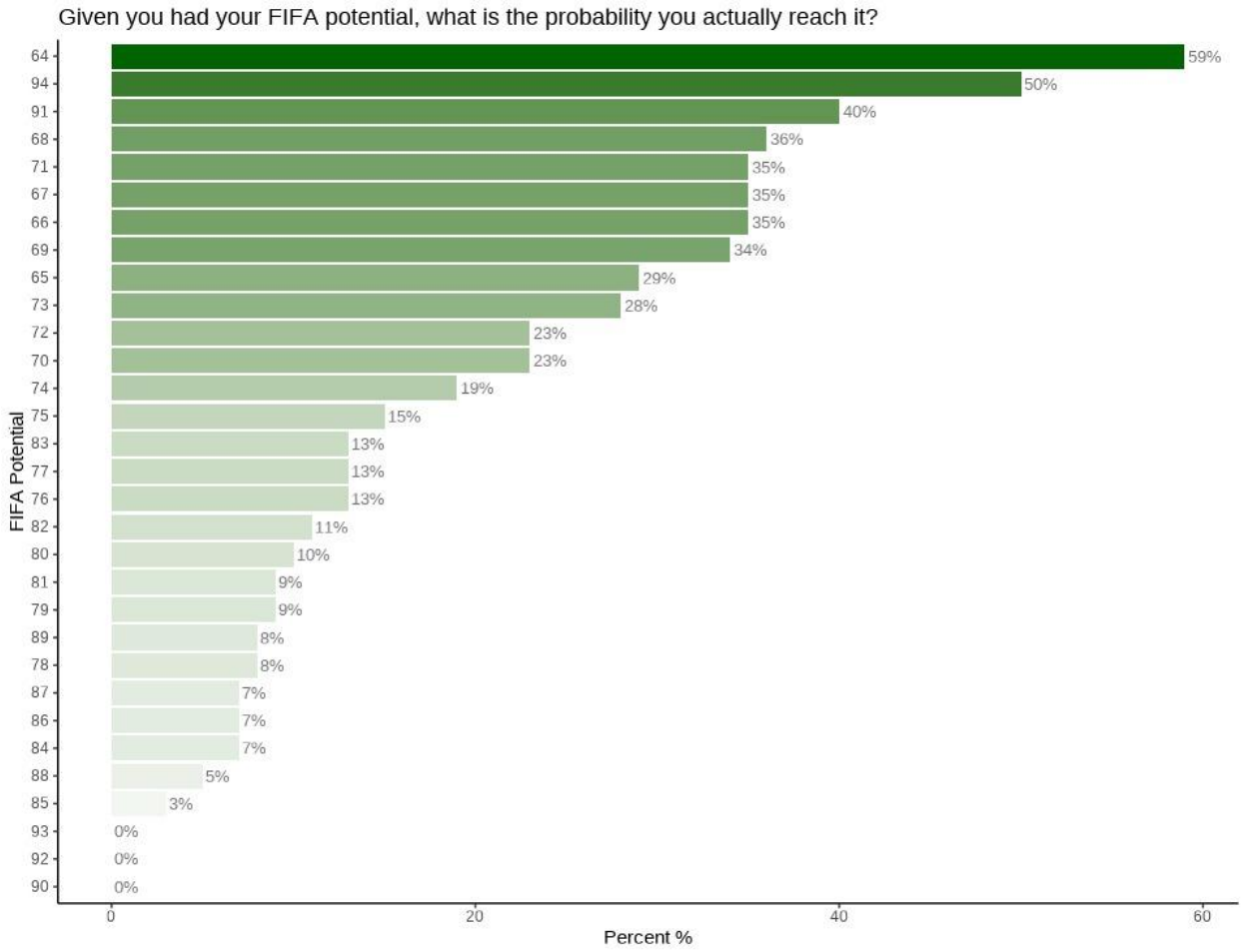


Figure 4

Figure 4 lays out the percentages for each potential score to be achieved. Basically, if we add a new player (5,400 + 1), without knowing his overall score, if he gets a potential score of 64, then he is 59% certain of reaching it. A remarkable insight is players that get a potential of 94 have a 50% chance of reaching it. Another thing worth noting is that 8 out of 10 of the top potential scores (64, 68, 71, 67, 66, 69, 65, 73) all lie in the lowest potential bracket [64, 74[. Proving that players with low potential scores, regardless of their overall, are relatively more likely to reach them.

This is further proven when we look at the bottom of the figure, specifically at the potential scores of 93, 92 and 90. All have a percentage of 0%. Meaning that players that are assigned these potential scores at their age of 19, have no chance of reaching it.

For what given overall are players most likely to achieve a given potential?

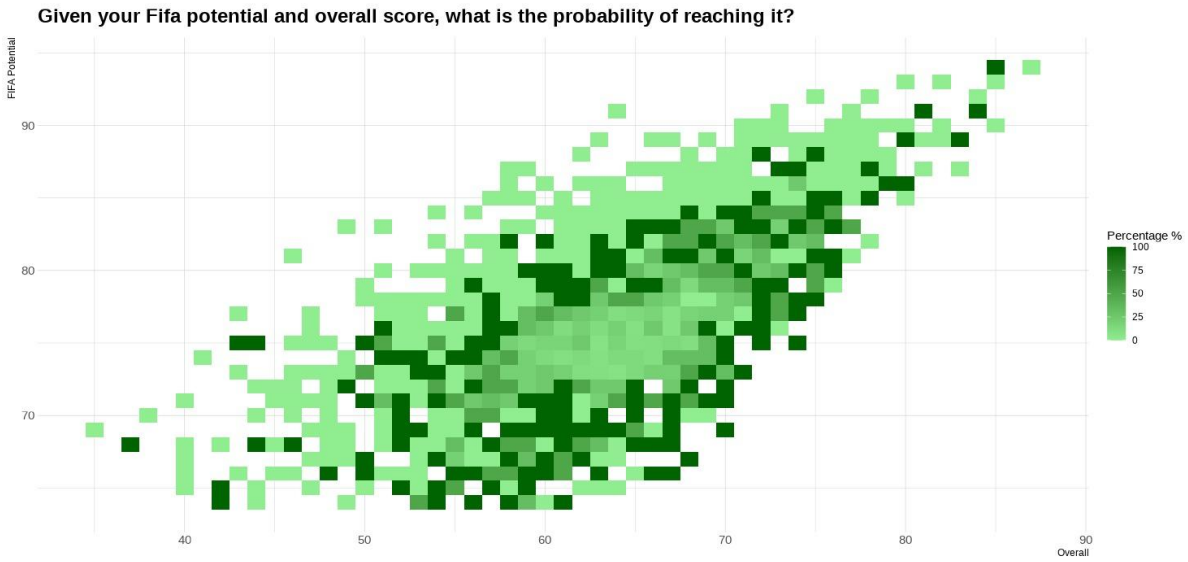


Figure 5

Building on the knowledge from figure 4, where we saw what percentage each potential score had for players to reach it, neglecting the player's overall score. Figure 5 adds the neglected overall as a dimension to the X-axis to depend on.

Figure 5 is a heatmap where the darker the color is, the more likely a player is to achieve the potential on the Y-axis, given he has the overall score on the X-axis.

One can notice the dark green circulating a middle-shade green zone that is from overall 58 to 70 and potential between 70 and 80 (tip: this is very close to the middle potential bracket [74, 84[). Players that fall into these segments have around a 50% chance of

reaching their potential. While players with higher or lower potential scores get the dark green color.

A few 100% outliers lie on both ends:

- High overall and high potential
- Low overall and low potential

Does the club matter?

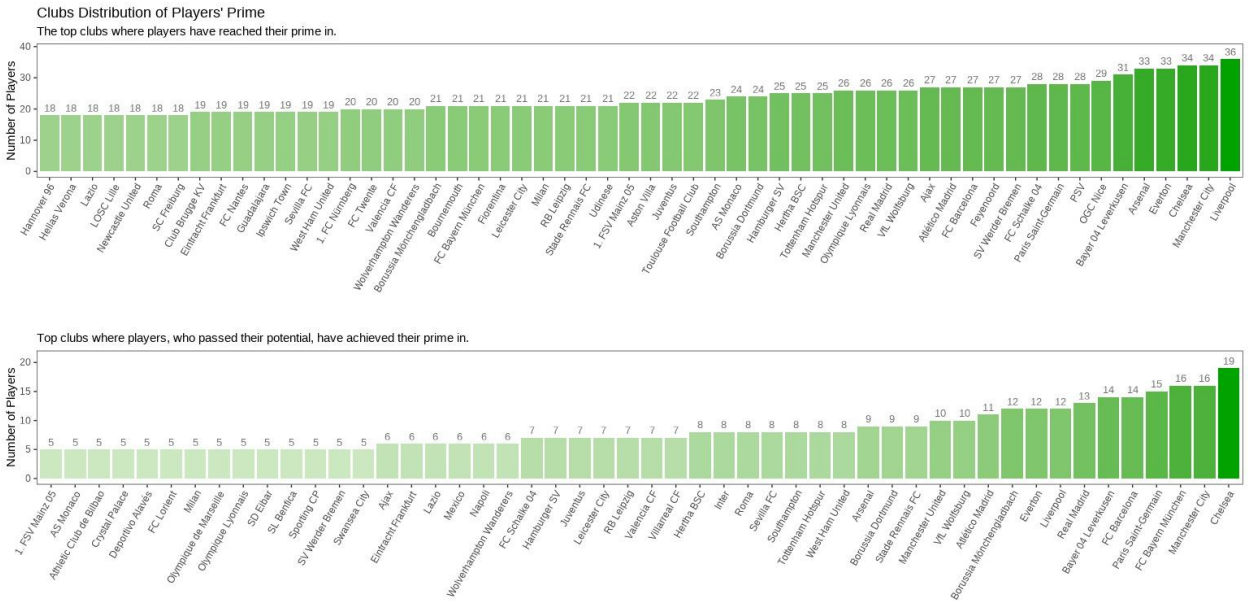


Figure 6

Football clubs continuously strive to bring the best young talents in football to enrich their lineups, hoping that these players assist their club in winning trophies and establishing national league dominance.

Figure 6 gives us two bar-plots. The first displays which clubs are players reaching their prime in the most. While the second displays which clubs, where players who have reached (or even passed) their potential, are reaching their prime in the most.

The second bar-plot of figure 6 is the same as the first but with the filter of “only players that passed their potential”. The addition of this filter gave the bar-plot a smoother decline which further highlights the number one club: Chelsea.

Does nationality matter?

To further explore what it takes to be a potential achiever, we have to explore the nationality background of the players. Figure 7 lays out the top countries where the potential achievers are from.

It is not surprising to see that European countries take 7 out of the top 10 countries and the other 3 are from the North & South American continents.

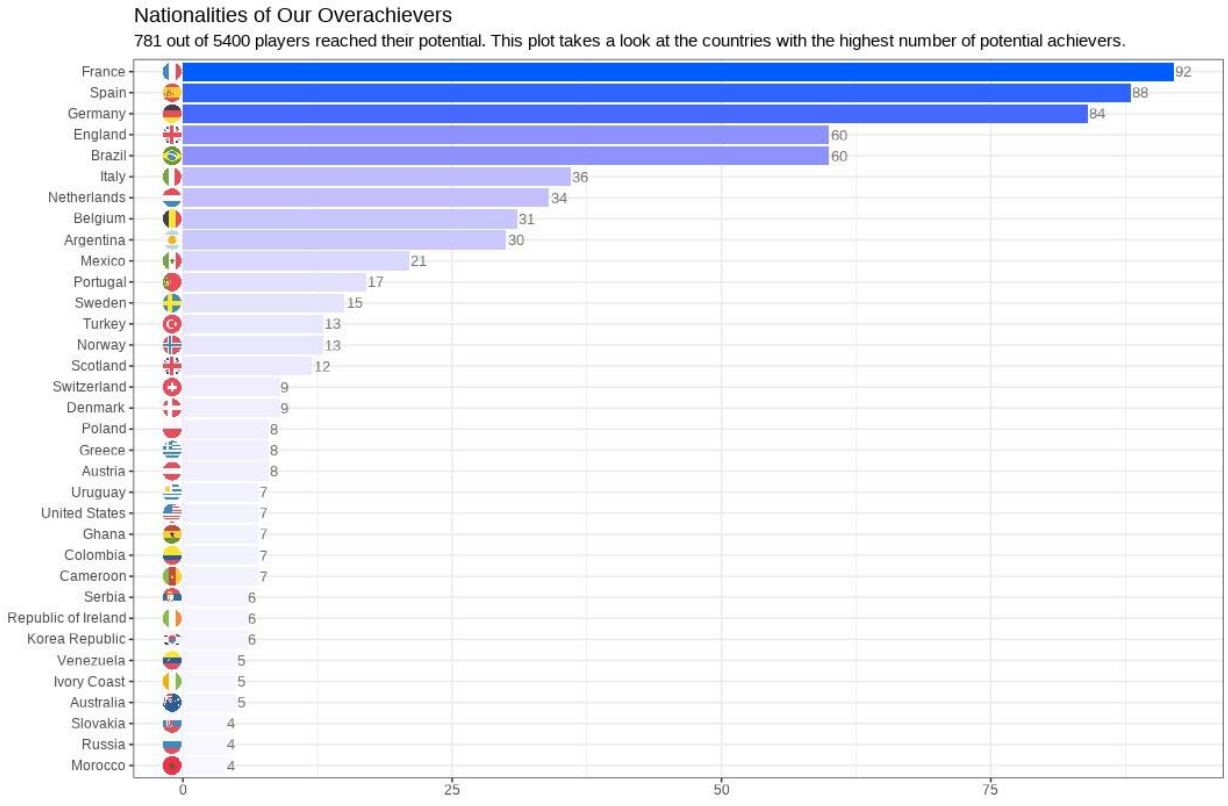


Figure 7

How much did EA FIFA get it right?

Potential brackets' accuracy evolution

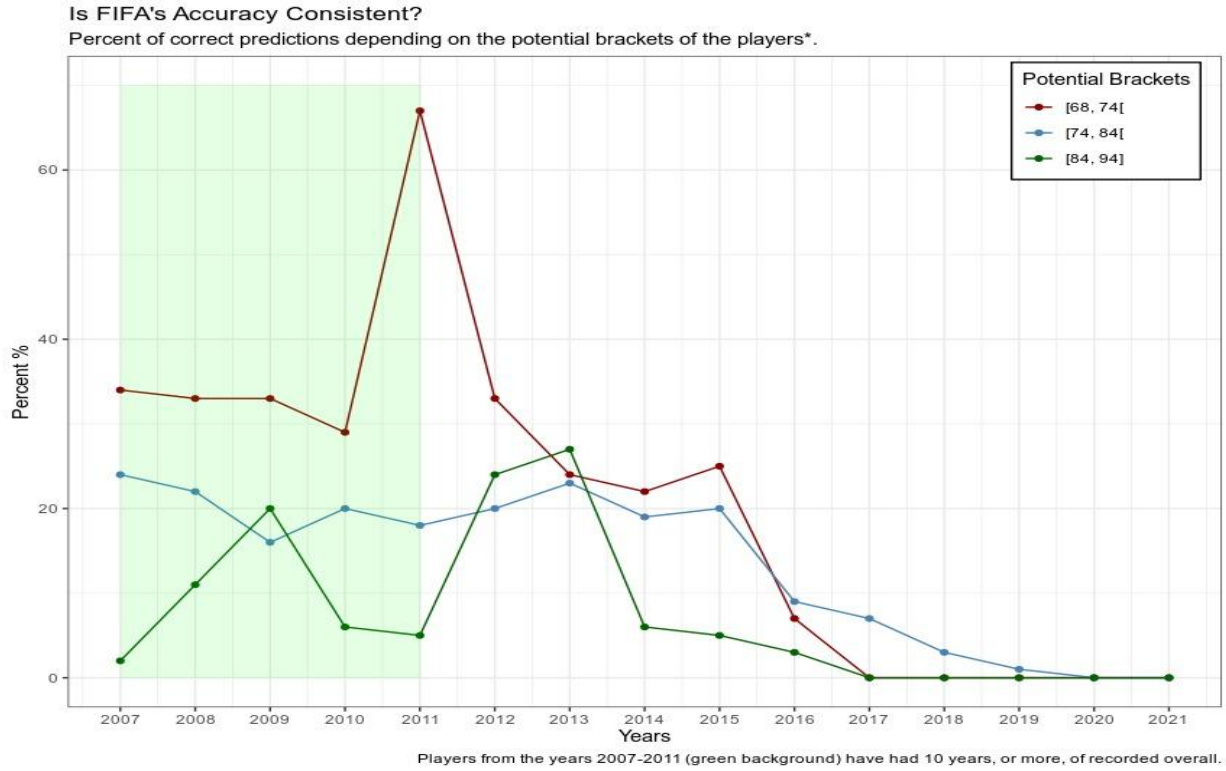


Figure 8

In the aim to analyze the game's accuracy at predicting future stars given all the installments (versions) it had from 2007 till 2021, we only need to focus and judge the highest potential bracket **[84, 94]**.

That is because the players that achieve their potential of values 84 till 94 mean that their overall will reach at least 84.
Moreover, the word "stars" in the title "How Accurate Is FIFA At Predicting Future Stars" refers to world-class players, i.e. players with high overall scores.

Figure 8 helps us track the accuracy percentage of each potential bracket throughout the years.

Recall that in figure 1 we partitioned the players into those that played for 10 years or more, and others that have played less than 10 years. This is presented in figure 8 by the green background that indicates potential brackets (which indicates players) of these years (including 2011) (should) have had 10 years or more of recorded overall scores in

the EA FIFA games.
That said, it is very important to note that all potential bracket accuracy percentages are decreasing as we move to the right, that is because players of these later years (2015 till 2021) are having less and less time to prove themselves, i.e. less and less time to cross their potential.

And as previously mentioned, our focus will be on the **[84, 94]** potential bracket. For this, we use figure 9 that only tracks this potential bracket. We observe the bracket's minimum value is 2% in 2007 and then reaches it's all time high of around 27% in 2013, then the percentage decreases to reach 0% in 2017.

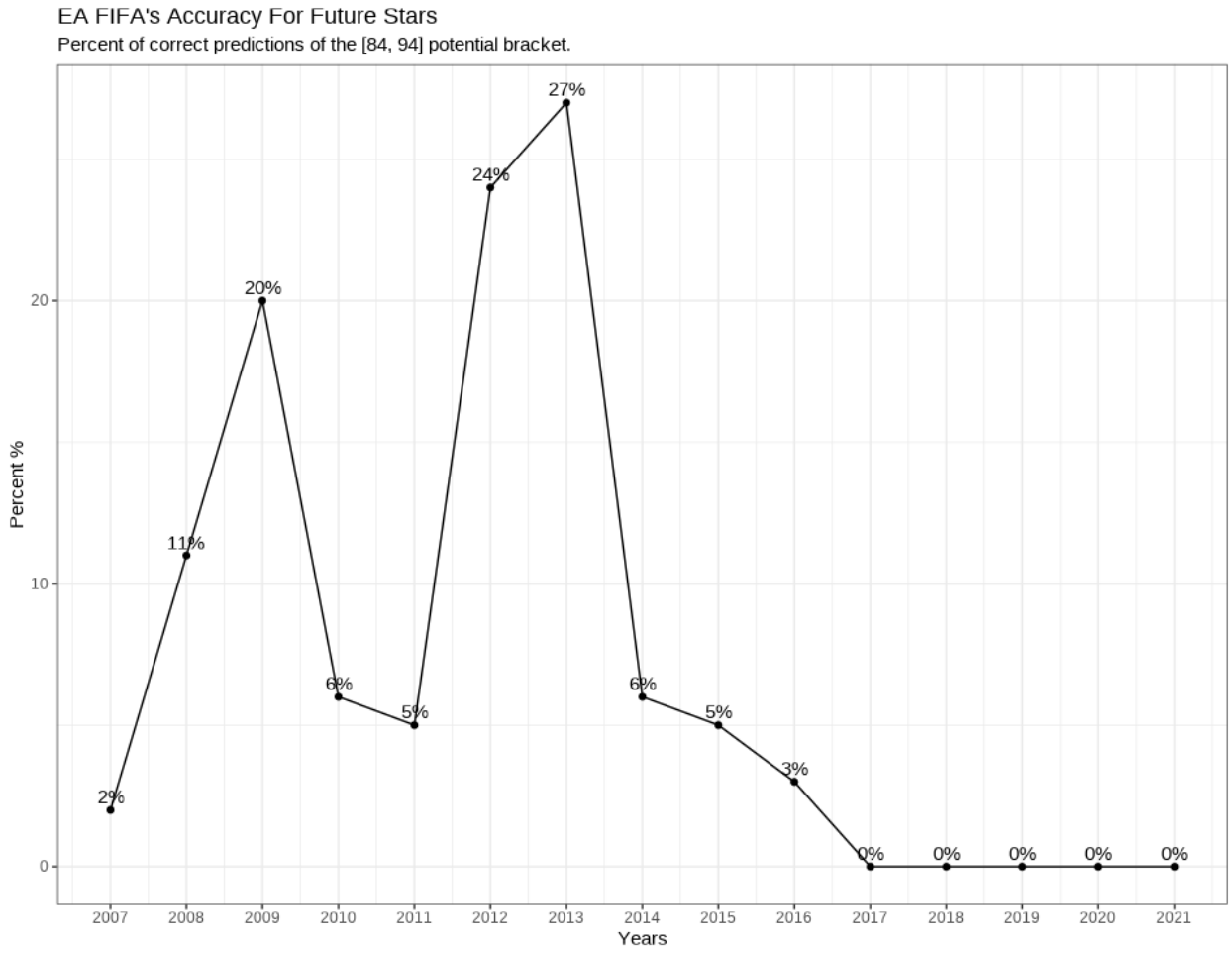


Figure 9

Unlike the [74, 84[potential bracket in figure 8 where its percentage fluctuation is fairly fixed between 2007 and 2015, the percentage fluctuations of the [84, 94[are more extreme thus we can not conclude a fixed behavior for this potential bracket.

To finally conclude whether EA FIFA is a good predictor we have to find its one true

accuracy percentage. To do that we have to consider a tuning parameter:
Years of football playing.

1. 10 years:
   If we consider the minimum playing time is 10 years thus we should calculate the average percentage for the years between 2007 and 2011, which results in:
   $\frac{2+11+20+6+5}{5} = 8.8\%$
2. 8 years:
   If we take into account the maximum accuracy reached in 2013, and thus push down the minimum playing years to 8, i.e. players should be playing for at least 8 years (and of course be of the [84, 94] potential bracket) to be considered for this percentage. Which results in:
   $\frac{2+11+20+6+5+24+27}{7} = 13.6\%$
3. 6 years:
   If we take into account the 6% and 5% accuracies of 2014 and 2015 respectively under the excuse that this might be a pattern i.e. year 2009 had a local maximum of 20% then dropped to 5% and 6% in 2010 and 2011 respectively then rose to 27% (global maximum) then dropped to 5% and 6% in 2014 and 2015 respectively. If we attribute these similarities in percentages to some pattern we do not know yet of, and not to the fact that players in 2014 and 2015 have not had enough time to prove themselves, then the minimum playing time for players (that lie in the [84, 94] potential bracket) should play for at least 6 years, and this results in:
   $\frac{2+11+20+6+5+24+27+5+6}{9} = 11.8\%$

Conclusion

To tidy our numbers up, EA FIFA's accuracy for players that lie in the [84, 94] potential bracket are:

- 8.8% if players play a minimum of 10 years
- 13.6% if players play a minimum of 8 years
- 11.8% if players play a minimum of 6 years

Moreover, as we have touched upon all aspects of a potential achiever in previous sections, we saw that playing time can't be the sole characteristic that affects the accuracy:

- Figures 4 and 5 showed us which overall and potential scores have the highest probability.

- Figure 6 told us which clubs have been best for players to reach their Prime Overall score after reaching (or passing) their potential.
- Figure 7 told us which countries are best at producing potential achievers.

Educated Guessing

Now, to estimate when these players would reach their potential, let us work with what we have and let us associate figure 3 with the [84, 94] potential bracket of figure 8. To see that potential bracket up close, figure 10 displays the age distribution of players, strictly from the [84, 94] potential bracket, that reached their potential and prime.

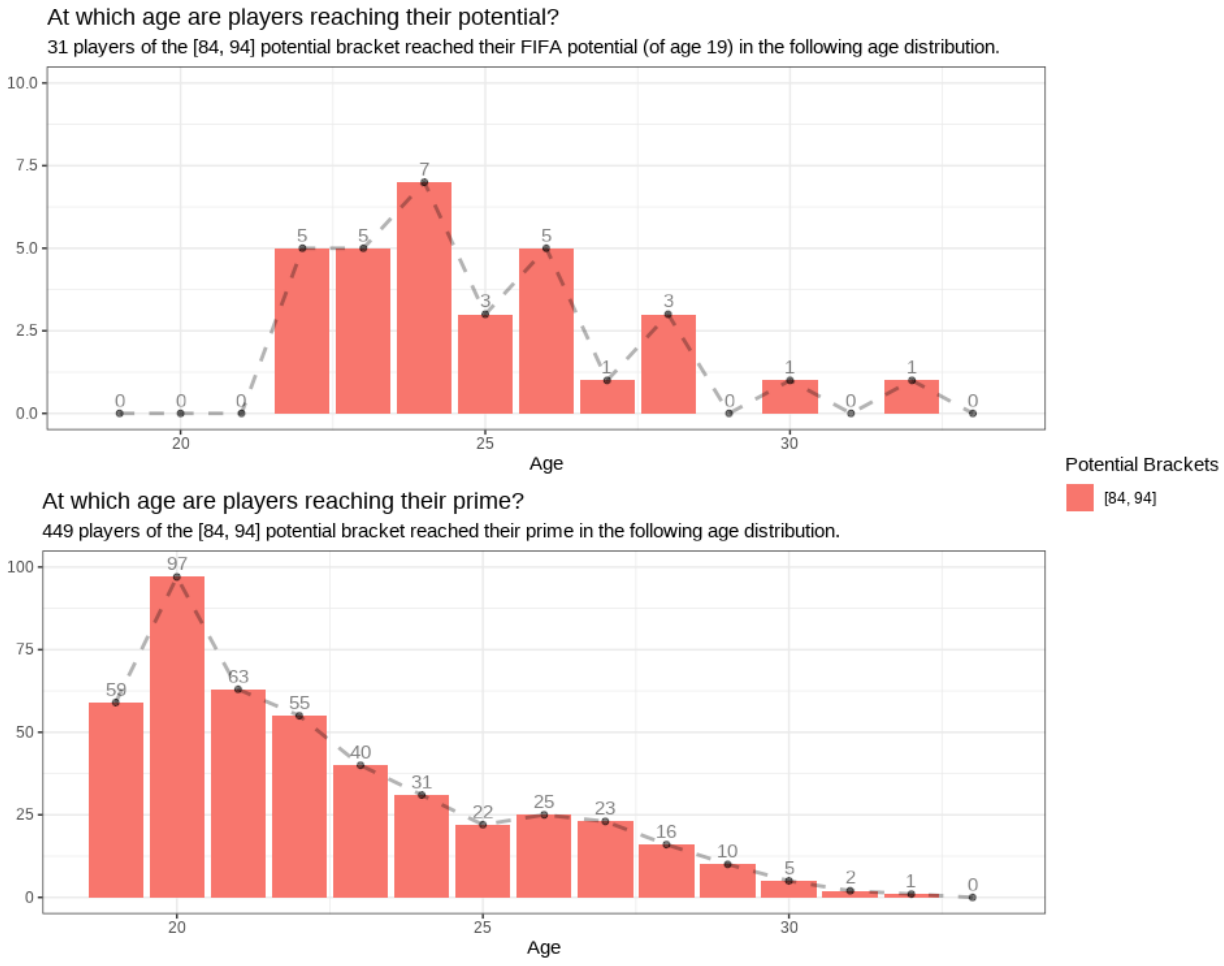


Figure 10

After observing the first plot of figure 10, choosing 6 years as a minimum playing time does not sound that bad since the plot tells us by the age the players are 25, 20 (5+5+7+3) out of 31 players (20/31) have already reached their potential, which is more than half.

Is there a curse for the top 1 player with the highest potential of every year?

The aforementioned description of how the data was scraped specified that the players of each year were sorted by highest potential to lowest, then the top 360 players were chosen for scraping.

That being said, how does the top one player in terms of highest potential score of every year compare to one another? Does ranking as the teenager with the highest potential for this year by EA FIFA cast a weight on the players? Or does this have a motivational effect on the player?

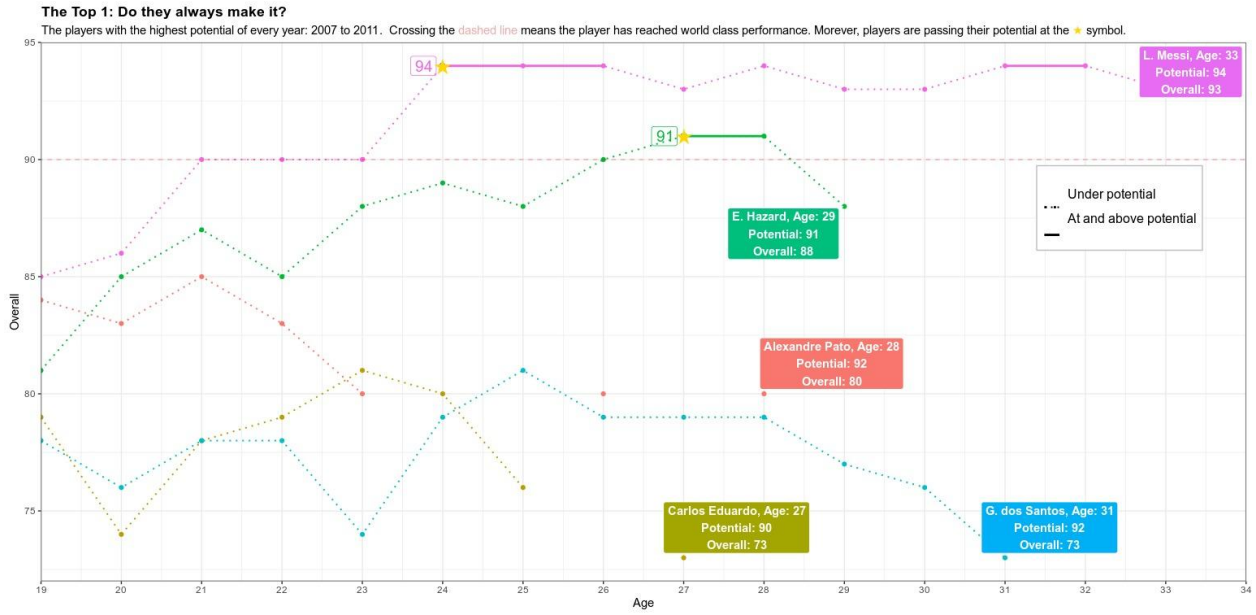


Figure 11

Figure 9 lays out the overall evolution of the top 1 players in highest potential for the years 2007 (Lionel Messi*), 2008 (Carlos Eduardo), 2009 (Dos Santos), 2010 (Alexander Pato) and 2011 (Eden Hazard) as they group up from age 19.

These players have had 10 years or more to prove themselves against their potential. We can see that only 2 out of 5 players have reached their potential while the rest were at least 12 overall points away from their potential i.e. nowhere near!

One insight is Alexander Pato's overall line which is interrupted between the age of 23 and 26, and then agen between 26 and 28. That is because Pato has been transferred to leagues that EA FIFA does not track. Same can be said about Carlos Eduardo.

Meanwhile figure 10 lays out the overall evolution of players that have had less than 10

years to play. Specifically for years 2012 (Neymar), 2013 (Lucas), 2014 (Marquinhos), 2015 (Berrardi), 2016 (Brandt), 2017 (Dembele), 2018 (Rashford), 2019 (Mbappe), 2020 (Felix) and finally 2021 (Vinicius).

We see that only Neymar was able to reach his potential at the age of 24 and have so far managed to maintain an overall it. While Mbappe has been having a higher overall year

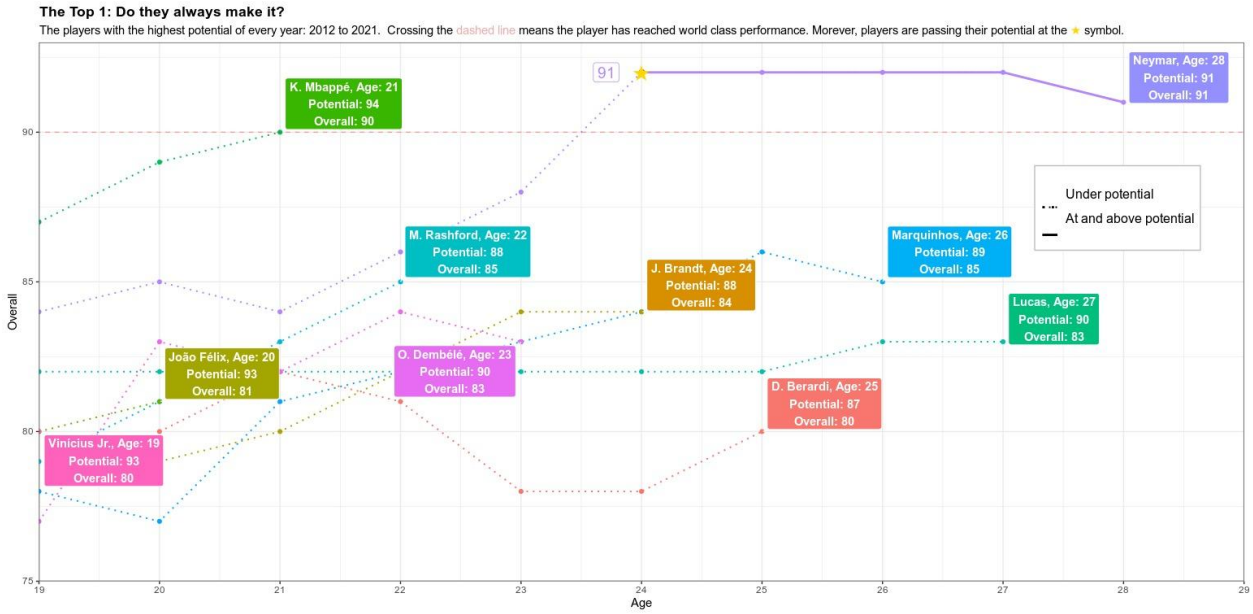


Figure 12

by year and looks set to break his potential, 94, at an earlier age than Neymar's. While Rashford, Brandt and Marquinhos are only 3, 4 and 4 overall points behind their potential respectively. Whereas Dembele, Lucas and Berrardi are falling behind their expectations and it is looking harder for them to achieve their potential. Finally, it is very early to put any judgements on Felix and Vinicius.

Taking all insights from both figures 11 and 12 into consideration, one can not properly declare a "curse" on the highest potential players of every year. That is because it seems there will always be little while till one of these players will actually reach their potential: Mbappe is currently a prime candidate to do so since the last one that did (Neymar from 2012). Moreover, the ratio of players that achieved their potential scores and have been playing for 10 years or more is 2/5 while that of the players that have been playing for less than 10 years is 1/9. So, perhaps the longevity of play time should not be neglected when assessing the top 1 players or any player for that matter.

Interesting insights

Figures 9 and 10 had witnessed rises and declines in overall scores as players grew up. This was the inspiration of Figures 11 and 12.

Top 25 rises in overall

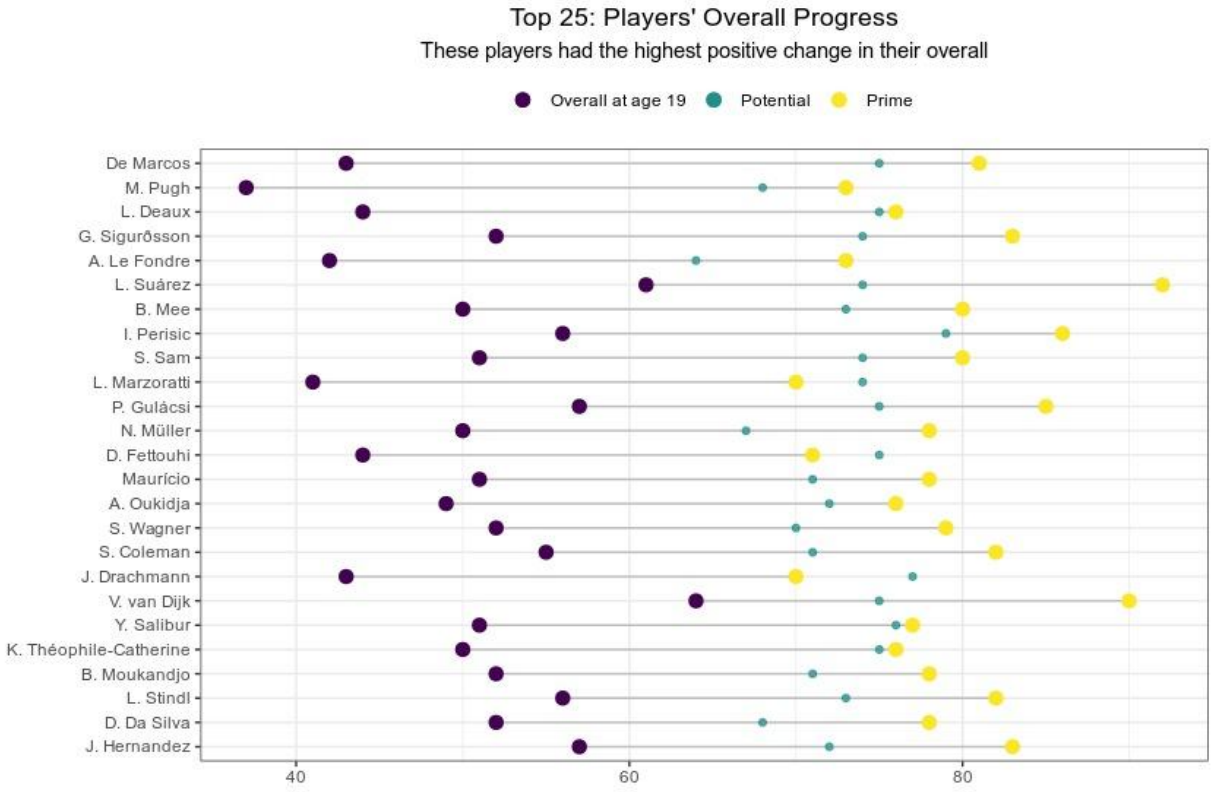


Figure 13

First, figure 11 lists the highest rises of overall scores in a player's career regardless of its length, sorted by highest rise at the top. De Marcos has certainly surprised many with his overall rise as it started at around the value of 44, and then primed at around 82, passing his EA FIFA potential along the way.

Another surprising insight for ex-Barcelona number 9 striker Luis Suarez* who, at age 19, had an overall just a bit higher than 60. This score then skyrocketed to over 90, passing his potential by a fairly large margin.

Top 25 steepest declines in overall

Second, figure 12 shows the highest negative changes in their overall. It is important to note that this plot should be read from right to left, unlike figure 11.

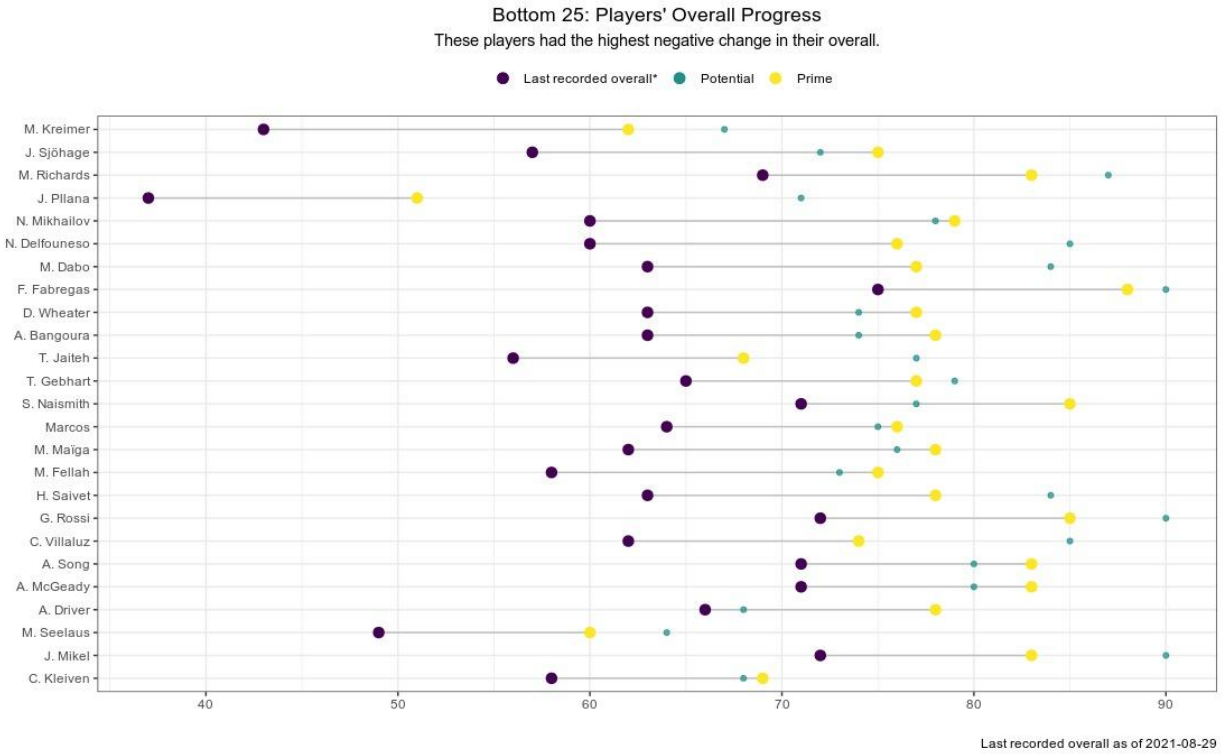


Figure 14

These declines in overall scores happen after players reach their prime score, and on that basis the top 25 highest negative declines were chosen, and displayed and sorted by the highest to lowest.

M. Kreimer reached his prime at around 62, but it was not enough for him to pass his potential (around 67) then his overall declined to the last recorded overall at around 43.

Notable mentions are ex-Chelsea and ex-Barcelona playmaker Fabregas* who, at his prime was a couple overall points behind his potential. He then declined from an overall score of around 88 down to 75.

III References

1. Murphy, Ronan. “FIFA player ratings explained: How are the card number & stats decided?” *Goal*, 2019, https://www.goal.com/en-ae/news/fifa-player-ratings-explained-how-are-the-card-number-stats/1hszd2fgr7wgf1n2b2yjdpgynu. Accessed 15 8 2021.
2. Saed, Sherif. “EA explains how FIFA player ratings are calculated.” *VG247*, 27 9 2016, https://www.vg247.com/how-ea-calculates-fifa-17-player-ratings. Accessed 15 8 2021.

*Profiles of these players are explored more in the “Player Profiles.pdf”