

VAR CHECKING FOOTBALL MHYTS

Authors
 Oguz Kokes | 20201006
 Serdar Cetiner | 20201016
 Mohamed Elbawab | 20201102

ANSWERING FOOTBALL'S BURNING QUESTIONS WITH *FIFA 21* DATASET

Every football fan will tell you that English can't score from penalties or Neymar falls to the ground on every occasion he gets, but how many of these conceptions are valid? Using *FIFA 21* player dataset, we will analyse some of these common myths to see if the statistical evidence supports these claims or they are just unfounded superstitions.

25 variables
 17,036 players

Personal Data:
Name, Age, Nationality, Reputation.

 Player Skills:
Overall rating & attribute.
Playing position, work rate, weak foot.

 Club Related Information:
Team & league information,
Player wages and transfer values

Checking the correlation heatmap to better understand what we are dealing with; we can already make some interesting observations:

There is a high correlation between dribbling and passing attributes, suggesting an overall ball control ability affects both traits.

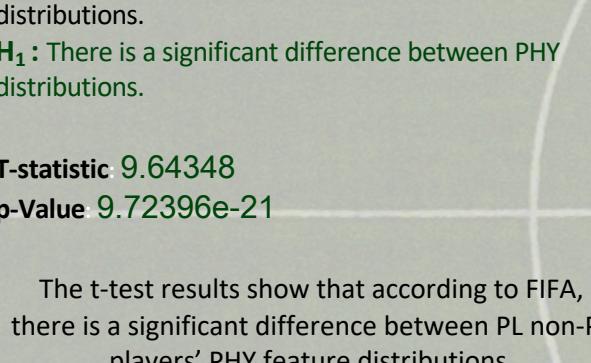
Understandably, potential (Growth) is negatively correlated with Age and overall rating (OVA).

These observations will be useful in later stages, especially during the clustering of the players, all features will be kept as they may be useful for statistical tests & analyses.

IS PREMIER LEAGUE REALLY MORE PHYSICAL?

Pundits and players new to the English league (such as Timo Werner over here) always claim that the play style in England is different than any other league, as it heavily depends on physicality and battles between the players. But does *FIFA* dataset also prove this claim?

Using the physicality (PHY) feature, we can use a 2-way t-test (where $\alpha = 0.05$) by filtering the players as PL vs non-PL and check if the players from PL have in fact a significantly difference between the distributions:



H_0 : There is no difference between the two PHY distributions.
 H_1 : There is a significant difference between PHY distributions.

T-statistic 9.64348
 p-Value 9.72396e-21

The t-test results show that according to *FIFA*, there is a significant difference between PL non-PL players' PHY feature distributions.

Timo, you'll just have to work harder!

[CONFIRMED]

"The Premier League is a little bit different to my old league. It's tougher than I thought."

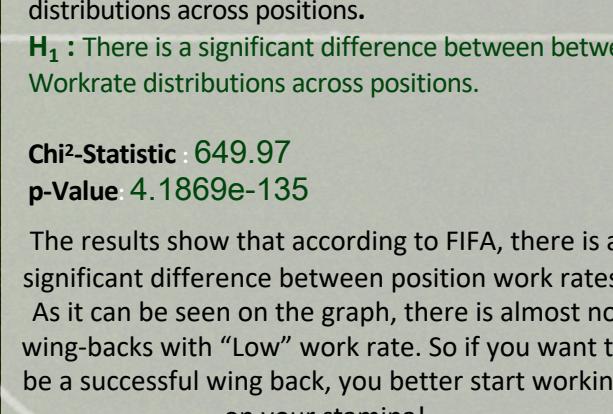
"The contact here is harder than in Germany, it's what I expected but not like this."

AN UNFAIR WORKLOAD?

Ever since Jürgen Klopp's Liverpool became a thing, the wingback position became one of the most crucial roles on the pitch. These players are expected to be an important providers for the attackers as well as being an integral part of the defensive line.

According to some, these new duties caused the wingbacks to be much more engaged and active during a match, causing them to run up and down the pitch for 90 minutes. Others say all players have their own demanding responsibilities.

To check whether there is a significant difference between positions we will use the "Work Rate" feature and perform a Chi-Squared test.



H_0 : There is no difference between Work rate distributions across positions.

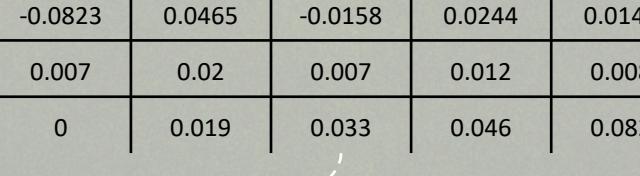
H_1 : There is a significant difference between Workrate distributions across positions.

Chi²-Statistic : 649.97
 p-Value : 4.1869e-135

The results show that according to *FIFA*, there is a significant difference between position work rates. As it can be seen on the graph, there is almost no wing-backs with "Low" work rate. So if you want to be a successful wing back, you better start working on your stamina!

[CONFIRMED]

How Klopp has changed the game using full-backs

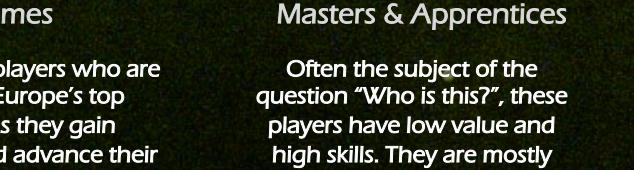


WHAT MAKES A VALUABLE PLAYER?

Saying a player is valuable can mean a lot of things. To find out what attributes of a player affects his monetary value we have computed the linear regression both by using Lasso Regression and Recursive Feature Elimination. Interestingly, height and weight of a player had little to no affect while overall skill and age affect the value significantly.

	Age	OVA	Height	Joined	Total Stats	W/F	SM	IR	PAC	DEF	PHY	GK
coef	-0.3264	0.4366	-0.0161	-0.0823	0.0465	-0.0158	0.0244	0.0142	-0.0299	-0.0399	0.0415	0.0388
std err	0.009	0.015	0.01	0.007	0.02	0.007	0.012	0.008	0.009	0.01	0.011	0.013
p-Value	0	0	0.103	0	0.019	0.033	0.046	0.082	0.001	0	0	0.004

Hierarchical Clustering - Dendrogram



Self Organising Maps and Hierarchical Clustering

were utilized to cluster all players in *FIFA 21*. Five distinct clusters were identified.

U-Matrix: Topographic map representation of football players.



What kind of players are out there?

Football players are often distinguished by their stats and attributes related to the position they play in.

What if we take a slightly different approach and consider their personal & club related information?

Top Dogs

Most talented, known and high valued players respective of their leagues. They are what all players strive to be.

Primes

Highly skilled players who are playing in Europe's top leagues. As they gain reputation and advance their careers they are candidates to be Top Dogs.

Masters & Apprentices

Often the subject of the question "Who is this?", these players have low value and high skills. They are mostly young and undiscovered or good players nearing the end of their careers.

The Regulars

Older players who have reached their skill cap. They don't have particularly high skill yet they have high value due to their experiences.

Wonderkids

Wonderkids are young and talented players who were discovered by top clubs. They made it into the big leagues now they strive to prove themselves.

Top Members

Lionel Messi

Cristiano Ronaldo

Mohammed Salah

Top Members

Mason Mount

Fabián Ruiz Peña

Federico Valverde

Top Members

Santi Cazorla

Thiago Almada

M. Vandeovertdt

Top Members

Martin Dubravka

Tomás Vaclík

Aitor Fernández

Top Members

Ruben Diaz

Sandro Tonali

Çağlar Söyüncü

