Title

Interested in football at all?

You’re welcome to ask questions at any time.

Summary of the Data

Original dataset contained all the players in the game, and contained several rows and columns that we weren’t particularly interested in.

Out of 18k players, we only wanted the top 25 (sorted by Overall Rating). The other rows were removed.

Some columns also features data we weren’t interested in - such as Club, Nationality, Value and Wages.

Purpose

Sample size is only 25 because to focus purely on the elite players

Expected Results

GK - Reflexes, Handling, Diving, Positioning and Kicking. Also Reactions.

CB - Such as tackling, marking and interceptions.

CM - Will have very different stats depending on their main role, hope to find out core stats that are vital to all midfielders. These are likely to be Passing, Composure and Vision - stats that help create goals and dictate the game.

WM - Wide players are expected to be fast and able to get past players with the ball.

ST - Needs to be very good at getting into the right positions and scoring goals.

27 years old seems like a good point where players have developed technical and mental attributes, but they are still in top physical condition.

Analysis Process

General information about the analysis process.

Python because: comfortable with it, capable of doing what I needed, a skills to develop further.

Club, Nationality, Image links, etc were removed, only really interested in attributes.

Overall Rating and Age are relevant to all players. A simple visualization was produced using PyPlot.

Players were sorted into groups by main playing position.

In a category, if no players had a rating of above 85 in a particular attribute column, that column was removed for that group.

Analysed and visualised the data.

Discussed the things I expected, and things that I was surprised about + any reflections I had

Overall Rating and Age

All the attribute related stats are out of a maximum of 99 - including the Overall Rating stat here.

For the Overall Rating, it’s important to note that the mean and median values are much closer to the minimum.

Meanwhile for age, it is closer to the middle. (Midpoint is 28.5)

Players by Overall Rating

Visualisation of players by Overall.

We can see why the mean and median were much closer to the minimum overall.

Messi and Ronaldo much better than the others.

Most players between 89 and 91, with only 3 players between 92 and 94.

Age Distribution

Small sample size with lots of bins in the histogram means findings might not be totally accurate - but still gives us a rough idea.

With more time, a full attribute analysis would’ve been done for players near the peaks. For now, used football knowledge to come to an well educated guess.

We assume that players are more likely to be near their peak in order to be in the top 25 players. This may not be true for players that are much better than the other players in the sample - such as Messi and Ronaldo.

Sorting Players by Position

Pick their favourite player from the list.

No full backs, irrelevant category.

Central midfielders include both defensive and attacking midfielders.

Griezmann could be in CM, Winger or Striker. He's listed as CAM in the dataset so grouped with CM.

It might’ve been a good idea to separate CMs into multiple categories.

Removing Irrelevant Data by Position

All the players are in groups, more data can be removed.

e.g. not interested in the outfield attributes of a goalkeeper.

If at least one player from a set has a particular stat that is above 85, then it is worth looking at. If that’s not the case the we aren’t particularly interested in the attribute.

Positional Analysis (Title)

Main analysis part of the presentation.

There will be visualizations - that I made in Tableau - to look at any stats that are particularly high or low for each player

This is followed by a table where we look at the mean values of every attribute. Which will help us decide which stats are the most important.

We will do this for every positional group that we are analysing.

Goalkeepers

Grouped by attribute and colour coded by player.

GK Reflexes stats are consistently very high (De Gea’s stat is exceptionally high).

GK Kicking has the widest range, and contains the lowest and 2nd lowest value. It’s a nice stat to have for GKs but not vital.

If a GK had 72 reflexes for example (like Courtois’ kicking), then they probably wouldn’t be top 25

Average stats for GKs

Ordered by the mean value

Top goalkeepers are more likely to have very high reflexes.

All these stats are important, however, reflexes are probably the most important and kicking the least.

Central Defenders

Reactions are consistently fairly low. (35 stats in total most were dropped though)

Standing tackle and jumping are the most vital.

All stats are important, Reactions and Strength less important.

Average Stats for CBs

Higher on the table = more likely to be key stats

Central Midfielders

Graph is too big to reliably extract information.

Easier to see variances on the colour coded table

Jumping doesn’t matter too much. Griezmann is probably only good at it because he’s a very explosive player

Interceptions and Standing tackle are important for CDMs, and more important than sliding tackle.

Reactions, Short Passing, Ball Control, Vision and Composure for all CMs

Average Stats for CMs

Table of averages gives a lot of info on core stats

Reactions, Short Passing, Ball Control, Vision and Composure are core attributes for midfielders of any type.

Physical attributes, aside from Stamina, in general appear to be less important.

Technical attributes are key.

Wingers

Dybala is noticeably less proficient than the other 3.

Dribbling and ball control are key technical skills.

Agility and acceleration are key physical skills.

All of these are vital when taking on defenders.

Average Stats for Wingers

Surprising that Crossing did not make it onto the list.

Probably because all the players featured are Inside Forwards rather than Traditional Wingers.

Balance is high on average, but a couple players had low balance whilst other had very high

Stats are mainly geared towards dribbling past players.

Finishing, composure and reactions are very high too, suggesting that modern wide players are expected to score goals whilst under pressure.

Strikers

Cristiano has really well rounded stats. Balance doesn’t seem as important as the others.

Would like to investigate the relationship between Aggression and Work Rate.

Some strikers are good at jumping and heading. Probably depends on their body type.

Average Stats for Strikers

All about getting into the right positions and scoring goals.

Dribbling is surprisingly high (almost same as wingers), would’ve expected the average to be around 90 (below finishing).

Summary

Messi and Ronaldo skew the the mean slightly, since they are significantly better than everyone else in the dataset.

Multiple peak ages for different players.

The majority of stats per position were roughly what we expected.

Conclusion

Bimodal distribution in the age distribution histogram not considered. First peak at 27 was expected. But a second peak was much later at 32 was not.

Reactions were vital for more attacking players, but still important all round. Even for defenders and goalkeepers.

Crossing was less important than expected since there were no Traditional Wingers, only wide players who play the Inside Forward role.

Reflection

Stats by Age to see when stats peak. Could look at historical data.

Central midfielders could’ve been split into CDMs and CAMs.

See what stats young players lack in. Would help influence what they train.

Not experienced with Tableau.

Data libraries such as PyPlot’s histograms and customising the number of bins - needed to google

Wrote reports and made presentations in university, and not so much since