

STATS15 Final Project - FIFA Player Skills Analysis

Thomas Peeler, James Cobb, Mackenzie Smith, Ciara Smith

2022-11-8

Section 1: Introduction and Background

Motivating Question

Soccer is widely considered the most popular sport in the world, with an estimated amount of global fans in the billions. Many make life-changing money due to their skill on the field; Paris Saint-Germain player Kylian Mbappé has a projected salary of over 100 million dollars during the 2022-23 season. Ability on the field is extremely important to both fans and players alike, and with this we aim to answer the motivating question: *what attributes make strong defensive and forward players?*

Background on FIFA and Soccer

General Professional Soccer: In soccer, a *league* is an association made up of *clubs*, or teams, which exists to negotiate matches and competitions between those clubs. Each club can have as many players as they want, but most tend to be around 30. Unlike other sports, a player's team can change once per year via a transfer, in which one club will buy a player from another club for a negotiated amount of money.

In a match, there are 11 players to each team on the field: 10 outfielders and 1 goalkeeper. Outfielders can be generally put into three categories: defenders, midfielders, and forwards. Defenders aim to keep the ball away from their own net, forwards try to get the ball to the other team's net, and midfielders take the role of either position as necessary. Forwards will generally be worse at defending than defenders, and vice versa.

FIFA Series: FIFA (aside from being a real federation) is a series of soccer video games that features many real soccer leagues, clubs, and players from all over the world. One can create their own teams or use ones that already exist to play matches against others or a computer. In the game, each soccer player has a wide range of numerical statistics; these range from specifics such as crossing and heading skill to more general stats like international recognition. FIFA games are released on a yearly basis, and these stats are also updated with each entry to the series.

How Ratings are Collected: The stats for each player are based on real life; EA, the developer of FIFA, employs many scouts to watch, in total, thousands of professional soccer matches and take notes on each player. They then report their own ratings back to EA, who use their own vetting system before implementing the ratings into the game. This process is repeated for each yearly entry of the series, in order to keep up with how players' skills and position in the soccer world can change over time.

##The Data

Structure: Each observational unit is one row of the data, representing one player and their attributes for that year; each player will have one measurement of attributes for each year.

To note, there are 6 summary statistics for each yearly measure of attributes: **defending**, **shooting**, **pasce**, **passing**, **dribbling**, **physic**. These are made up of weighted averages of the other specific attributes:

Pace: Sprint Speed x 0.55 + Acceleration x 0.45 *Shooting:* Finishing x 0.45 + Long shots x 0.2 + Shot power x 0.2 + Positioning x 0.05 + Penalties x 0.05 + Volleys x 0.05 *Passing:* Short Pass x 0.35 + Vision x 0.2 + Crossing x 0.2 + Long Pass x 0.15 + Curve x 0.05 + Freekick Accuracy x 0.05 *Dribbling:* Dribbling x 0.5 + Ball Control x 0.35 + Agility x 0.1 + Balance x 0.05 *Defending:* Marking x 0.3 + Stand Tackle x 0.3 + Interceptions x 0.2 + Heading Acc x 0.1 + Slide Tackle x 0.1 *Physical:* Strength\0.5 + Stamina\0.25 + Aggression\0.2 + Jumping\0.05

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(naniar)
```

```
## Warning: package 'naniar' was built under R version 4.2.2
```

```
#load all csv
players15 <- read.csv("players_15.csv")
players16 <- read.csv("players_16.csv")
players17 <- read.csv("players_17.csv")
players18 <- read.csv("players_18.csv")
players19 <- read.csv("players_19.csv")
players20 <- read.csv("players_20.csv")
players21 <- read.csv("players_21.csv")
players22 <- read.csv("players_22.csv")

#add year column for each and bind
players15$year = 2015
players16$year = 2016
players17$year = 2017
players18$year = 2018
players19$year = 2019
players20$year = 2020
players21$year = 2021
players22$year = 2022

players <- rbind(players15, players16, players17, players18, players19, players20, players21, players22)

#filter out goalkeepers
players_ng <- players %>% filter(!(player_positions=="GK"))
```

```

#select desired columns
players_s <- players_ng %>% select(c(sofifa_id, year, short_name, league_name, club_name, weight_kg, height_cm))

#unique entry id for each year/player combination
players_s <- players_s %>% mutate(entry_id=paste(as.character(sofifa_id), "_", as.character(year), sep="_"))

#replace all empty strings with NA
repl_na <- players_s %>%
  replace(players_s=="", NA)

#demonstrating how all NAs are entries with no league/club; we can keep them
def_nona <- repl_na %>% na.omit()

players_s %>%
  filter(!(entry_id %in% def_nona$entry_id)) %>%
  head()

```

```

##      entry_id sofifa_id year      short_name league_name club_name weight_kg
## 1 209119_2015    209119 2015 F. Amorebielsa                83
## 2 178007_2015    178007 2015 Miguel Veloso                78
## 3 209097_2015    209097 2015   O. Cardoso                84
## 4 20209_2015     20209 2015 Ricardo Costa                80
## 5 49072_2015     49072 2015   D. Forlán                78
## 6 178416_2015    178416 2015     J. Lens                73
##   height_cm age player_positions defending shooting pace passing dribbling
## 1      194  28           CB, LB      81      47  53      56      50
## 2      180  28           CDM, CM      75      72  68      82      78
## 3      188  30              ST      36      87  42      72      69
## 4      183  33           CB, RB      77      37  54      44      52
## 5      180  35              ST      28      81  66      79      80
## 6      178  26           RW, ST      31      75  86      70      81
##   physic
## 1      85
## 2      74
## 3      71
## 4      83
## 5      66
## 6      71

```

```

no_league_club <- repl_na %>% select(-c(league_name, club_name)) %>% na.omit()

players_s %>%
  filter(!(entry_id %in% no_league_club$entry_id)) %>%
  head()

```

```

## [1] entry_id      sofifa_id      year      short_name
## [5] league_name    club_name      weight_kg    height_cm
## [9] age            player_positions defending    shooting
## [13] pace           passing        dribbling    physic
## <0 rows> (or 0-length row.names)

```