

Winter 2020

FIFA: A Complete Player Dataset

18k+ Latest FIFA Players, with ~80 Attributes Extracted from FIFA database
eZvh

FIFA



FIFA: The **Fédération Internationale de Football Association**^[a] (FIFA /ˈfiːfə/ **FEEF**-ə; *French* for 'International Federation of Association Football') is an organization which describes itself as an The Fédération Internationale de Football Association[a] (FIFA /ˈfiːfə/ **FEEF**-ə; *French* for 'International Federation of Association Football') is an organization which describes itself as an international governing body of association football, fútsal, beach soccer, and eFootball. FIFA is responsible for the organization of football's major international tournaments, notably the World Cup which commenced in 1930 and the Women's World Cup which commenced in 1991.



FIFA was founded in 1904 to oversee international competition among the national associations of Belgium, Denmark, France, Germany, the Netherlands, Spain, Sweden, and Switzerland. Headquartered in Zürich, its membership now comprises 211 national associations. Member countries must each also be members of one of the six regional confederations into which the world is divided: Africa, Asia, Europe, North & Central America and the Caribbean, Oceania, and South America.

Although FIFA does not control the rules of football, that being the responsibility of the International Football Association Board, it is responsible for both the organization of a number of tournaments and their promotion, which generate revenue from sponsorship. In 2017, FIFA had revenues of over US \$734 million, for a net loss of \$189 million, and had cash reserves of over US\$930 million.



Reports by investigative journalists have linked FIFA leadership with corruption, bribery, and vote-rigging related to the election of FIFA president Sepp Blatter and the organization's decision to award the 2018 and 2022 World Cups to Russia and Qatar, respectively. These allegations led to the indictments of nine high-ranking FIFA officials and five corporate executives by the U.S. Department of Justice on charges including racketeering, wire fraud, and money laundering. On 27 May 2015, several of these officials were arrested by Swiss authorities, who were launching a simultaneous but separate criminal investigation into how the organization awarded the 2018 and 2022 World Cups. Those among these officials who were also indicted in the U.S. are expected to be extradited to face charges there as well. Many officials were suspended by FIFA's ethics committee including Sepp Blatter and Michel Platini. In early 2017 reports became public about FIFA president Gianni Infantino attempting to prevent the re-elections of both chairmen of the ethics committee, Cornel Borbély and Hans-Joachim Eckert, during the FIFA congress in May 2017. On May 9, 2017, following Infantino's proposal, FIFA Council decided not to renew the mandates of Borbély and Eckert. Together with the chairmen, 11 of 13 committee members were removed.

- **The goal of this project is to use the players' attributes to predict his annual wage in dollars.**

- **About the Data File:**

Attributes:

Data file includes latest edition FIFA players' attributes like:

Age, Nationality, Overall, Potential, Club, Wage, Preferred Foot, International Reputation, Weak Foot, Skill Moves, Work Rate, Position, Jersey Number, Joined, Contract Valid Until, Height, Weight, LS, ST, RS, LW, LF, CF, RF, RW, LAM, CAM, RAM, LM, LCM, CM, RCM, RM, LWB, LDM, CDM, RDM, RWB, LB, LCB, CB, RCB, RB, Crossing, Finishing, Heading, Accuracy, ShortPassing, Volleys, Dribbling, Curve, FKAccuracy, LongPassing, BallControl, Acceleration, SprintSpeed, Agility, Reactions, Balance, ShotPower, Jumping, Stamina, Strength, LongShots, Aggression, Interceptions, Positioning, Vision, Penalties, Composure, Marking, StandingTackle, SlidingTackle, GKDividing, GKHandling, GK Kicking, GKPositioning and GKReflexes.

- **Attributes Values:**

row number	Age	age	Nationality
nationality	Overall	overall rating	Potential
potential rating	Club	current club	Wage (Target Variable)
current wage in thousands of dollars	Special	Preferred Foot	left/right
International Reputation	rating on scale of 5	Weak Foot	rating on scale of 5
Skill Moves	rating on scale of 5	Work Rate	attack work rate/defence work rate
Body Type	body type of player	Real Face	Position
position on the pitch	Jersey Number	jersey number	Joined
joined date	club name if applicable	Contract Valid Until	contract end date
Height	height of the player	Weight	weight of the player
LS	rating on scale of 100	ST	rating on scale of 100
RS	rating on scale of 100	LW	rating on scale of 100
LF	rating on scale of 100	CF	rating on scale of 100
RF	rating on scale of 100	RW	rating on scale of 100
LAM	rating on scale of 100	CAM	rating on scale of 100
RAM	rating on scale of 100	LM	rating on scale of 100
LCM	rating on scale of 100	CM	rating on scale of 100
RCM	rating on scale of 100	RM	rating on scale of 100
LWB	rating on scale of 100	LDM	rating on scale of 100
CDM	rating on scale of 100	RDM	rating on scale of 100
RWB	rating on scale of 100	LB	rating on scale of 100
LCB	rating on scale of 100	CB	rating on scale of 100
RCB	rating on scale of 100	RB	rating on scale of 100
Crossing	rating on scale of 100	Finishing	rating on scale of 100
HeadingAccuracy	rating on scale of 100	ShortPassing	rating on scale of 100
Volleys	rating on scale of 100	Dribbling	rating on scale of 100
Curve	rating on scale of 100	FKAccuracy	rating on scale of 100
LongPassing	rating on scale of 100	BallControl	rating on scale of 100
Acceleration	rating on scale of 100	SprintSpeed	rating on scale of 100
Agility	rating on scale of 100	Reactions	rating on scale of 100
Balance	rating on scale of 100	ShotPower	rating on scale of 100
Jumping	rating on scale of 100	Stamina	rating on scale of 100

Strength	rating on scale of 100	LongShots	rating on scale of 100
Aggression	rating on scale of 100	Interceptions	rating on scale of 100
Positioning	rating on scale of 100	Vision	rating on scale of 100
Penalties	rating on scale of 100	Composure	rating on scale of 100
Marking	rating on scale of 100	StandingTackle	rating on scale of 100
SlidingTackle	rating on scale of 100	GK Diving	rating on scale of 100
GK Handling	rating on scale of 100	GK Kicking	rating on scale of 100
GK Positioning	rating on scale of 100	GK Reflexes	rating on scale of 100

Some Guidance:

- The competition will be opened January 6th 2020 and it will be closed one week 10 of the winter quarter.
- Students will be evaluated based on:
 1. Validity of the model
 2. The R square
 3. The R square adjusted
 4. The simplicity of the model (number of predictors used to create the model).
- Students should report leverage points good ones and bad ones, and explain how they dealt with these kind of points.
- Students should report missing values, and explain how they dealt with these kind of missing data.
- Students should report transformations used to get their last and best Model.
- Students should report their best predictor(s).
- Students are only allowed to use lm, bestpredictor, glm functions in R
- Students are not allowed to use any of the Machine Learning techniques or function to predict the players' annual salary.
- The score of this project will be given to all members of the group assigned at the beginning of the quarter.
- Each group will have 3-4 members. If some students drop from the class, two students are still enough to finish the project.
- Students should have the lecture name and number as part of their group nickname... (such as: Lec 1 AlmoGroup).
- Students are first ranked based on all groups in our three lectures, then the final ranking and grades will be based independently on each lecture.

- **Submission File Format**

The file should contain a header and have the following format:

```
Ob,WageNew
1 ,1690.1
2 ,1877
3 ,1752
.
.
etc.
```

You can download an example submission file (sample_submission.csv) on the [Data page](#).

What's the difference between a private and public leaderboard?

The Kaggle leaderboard has a public and private component to prevent participants from “overfitting” to the leaderboard. If your model is “overfit” to a dataset, then it is not generalizable outside of the dataset you trained it on. This means that your model would have low accuracy on another sample of data taken from a similar dataset.

Public Leaderboard

For all participants, the same 50% of predictions from the test set are assigned to the public leaderboard. The score you see on the public leaderboard reflects your model's accuracy on this portion of the test set.

Private Leaderboard

The other 50% of predictions from the test set are assigned to the private leaderboard. The private leaderboard is not visible to participants until the competition has concluded. At the end of a competition, we will reveal the private leaderboard so you can see your score on the other 50% of the test data. The scores on the private leaderboard are used to determine the competition winners. Getting Started competitions are run on a rolling timeline so the private leaderboard is never revealed.

How do I create and manage a team?

When you accept the competition rules, a team will be created for you. You can invite others to your team, accept a merger with another team, and update basic information like team name by going to the More < [Team](#) page.

We've heard from many Kagglers that teaming up is the best way to learn new skills AND have fun. If you don't have a teammate already, consider asking if anyone wants to team up in the [discussion forum](#).

What are kernels?

Kaggle Kernels is a cloud computational environment that enables reproducible and collaborative analysis. Kernels supports scripts in R and Python, Jupyter Notebooks, and RMarkdown reports. Go to the [Kernels tab](#) to view all of the publicly shared code on this competition. For more on how to use Kernels to learn data science, visit the [Read more about our decision to implement a rolling leaderboard on getting started competitions here](#).

How do I contact Support?

Kaggle does not have a dedicated support team so you'll typically find that you receive a response more quickly by asking your question in the appropriate forum. (For this competition, you'll want to use the [House Prices discussion forum](#)).

Support is only able to help with issues that are being experienced by all participants. Before contacting support, please check the discussion forum for information on your problem. If you can't find it, you can post your problem in the forum so a fellow participant or a Kaggle team

member can provide help. The forums are full of useful information on the data, metric, and different approaches. We encourage you to use the forums often. If you share your knowledge, you'll find that others will share a lot in turn!

If your problem persists or it seems to be affecting all participants then please [contact us](#).

Rules

One account per participant: You cannot sign up to Kaggle from multiple accounts and therefore you cannot submit from multiple accounts.

No private sharing outside teams

Privately sharing code or data outside of teams is not permitted. It's okay to share code if made available to all participants on the forums.

Team Mergers

Team mergers are allowed and can be performed by the team leader. In order to merge, the combined team must have a total submission count less than or equal to the maximum allowed as of the merge date. The maximum allowed is the number of submissions per day multiplied by the number of days the competition has been running.

Team Limits

Maximum of 4 members per team.

Submission Limits

You may submit a maximum of 3 entries per day.

You may select up to 2 final submissions for judging.

Competition Timeline

Start Date: 01/13/2020

Merger Deadline: None

Entry Deadline: None

End Date: 03/23/2020 3:00 AM

Due to the public nature of the data, this competition does not count towards Kaggle ranking points.

- We ask that you respect the spirit of the competition and do not cheat. Hand-labeling is forbidden.

Competition URL

[http://www.kaggle.com/c/](http://www.kaggle.com/c/fifa2019wages) fifa2019wages

Choose a short URL. All links to your competition pages will start with this URL.