FIFA 2019 Player Analysis and Modeling

Anirban Pal  
DSC680

Fall 2019  
<https://anirbanpaldsc.github.io/>

Contents

[Business Question 2](#_Toc20608018)

[Data Source 2](#_Toc20608019)

[Methods used in the project 4](#_Toc20608020)

[Potential Issues? 4](#_Toc20608021)

[GitHub Repository 5](#_Toc20608022)

[Appendix 6](#_Toc20608023)

[Reference 8](#_Toc20608024)

# Business Question

The Fédération Internationale de Football Association or FIFA is a non-profit organization which describes itself as an international governing body of association football, fútsal, beach soccer, and efootball. It is the highest governing body of football. ([Wikipedia](https://en.wikipedia.org/wiki/FIFA))

Using the player stats and data, the project intends to do exploration of facts and questions like:

1. Which Club or National Team has the best-rated players?
2. What is the strength of a team at a particular position?
3. Which team has the best dribbling speed?

Advanced analyses like:

1. Are players priced accurately? i.e. Is there a relationship between Market Value and Wages of Players?
2. What’s the ratio of total wages/ total potential for clubs? Which clubs are the most economical?
3. How is age distribution related to the player’s overall rating?
4. Co-relation between Age and Nationality?
5. Co-relation between Age and Potential?
6. Can we predict a player’s potential based on his skill set?

Machine Learning model like:

1. Can we predict the Value of a player based on its attributes (like accuracy, shot power, reactions, dribbling etc.)?
2. According to the skill ratings of a player, can we predict what position they play in?

# Data Source

Data and stats about players enlisted in FIFA is available online on websites like <https://www.fifaindex.com/>,  [https://sofifa.com](https://sofifa.com/) or FIFA itself. The original dataset has been scraped from FIFA website. But specific to this project, the source of the data is Kaggle:

[https://www.kaggle.com/karangadiya/fifa19/version/4#](https://www.kaggle.com/karangadiya/fifa19/version/4)

**Codebook:**

| **Column  Position** | **Field** | **Details** | **Data Type** | **Example** |
| --- | --- | --- | --- | --- |
| 1 | # | row number | numeric | 0 |
| 2 | ID | unique id for every player | numeric | 158023 |
| 3 | Name | name | string | L. Messi |
| 4 | Age | age | numeric | 31 |
| 5 | Photo | url to the player's photo | url | <https://cdn.sofifa.org/players/1.png> |
| 6 | Nationality | nationality | string | Brazil |
| 7 | Flag | url to player's country flag | url | https://cdn.sofifa.org/flags/52.png |
| 8 | Overall | overall rating | numeric | 94 |
| 9 | Potential | potential rating | numeric | 94 |
| 10 | Club | current club | string | Arsenal |
| 11 | Club Logo | url to club logo | url | <https://cdn.sofifa.org/teams/1.png> |
| 12 | Value | current market value | string | €110.5M |
| 13 | Wage | current wage | string | €565K |
| 14 | Special | special | numeric | 2202 |
| 15 | Preferred Foot | left/right | string | Right |
| 16 | International Reputation | rating on scale of 5 | numeric | 5 |
| 17 | Weak Foot | rating on scale of 5 | numeric | 4 |
| 18 | Skill Move | rating on scale of 5 | numeric | 4 |
| 19 | Work Rate | attack work rate/defense work rate | string | Medium/ Medium |
| 20 | Body Type | body type of player | string | Normal |
| 21 | Real Face | true or false | binary | Yes |
| 22 | Position | position on the pitch | string | RF |
| 23 | Jersey Number | jersey number | numeric | 10 |
| 24 | Joined | joined date | date | 1-Jul-04 |
| 25 | Loaned From | club name if applicable | string | Al Hilal |
| 26 | Contract Valid Until | contract end date | string | 2021 |
| 27 | Height | height of the player | string | 5'7 |
| 28 | Weight | weight of the player | string | 159lbs |
| 29 | LS | rating on scale of 100 | string | 88+2 |
| 30 | ST | rating on scale of 100 | string | 88+2 |
| 31 | RS | rating on scale of 100 | string | 88+2 |
| 32 | LW | rating on scale of 100 | string | 92+2 |
| 33 | LF | rating on scale of 100 | string | 93+2 |
| 34 | CF | rating on scale of 100 | string | 93+2 |
| 35 | RF | rating on scale of 100 | string | 93+2 |
| 36 | RW | rating on scale of 100 | string | 92+2 |
| 37 | LAM | rating on scale of 100 | string | 93+2 |
| 38 | CAM | rating on scale of 100 | string | 93+2 |
| 39 | RAM | rating on scale of 100 | string | 93+2 |
| 40 | LM | rating on scale of 100 | string | 91+2 |
| 41 | LCM | rating on scale of 100 | string | 84+2 |
| 42 | CM | rating on scale of 100 | string | 84+2 |
| 43 | RCM | rating on scale of 100 | string | 84+2 |
| 44 | RM | rating on scale of 100 | string | 91+2 |
| 45 | LWB | rating on scale of 100 | string | 64+2 |
| 46 | LDM | rating on scale of 100 | string | 61+2 |
| 47 | CDM | rating on scale of 100 | string | 61+2 |
| 48 | RDM | rating on scale of 100 | string | 61+2 |
| 49 | RWB | rating on scale of 100 | string | 64+2 |
| 50 | LB | rating on scale of 100 | string | 59+2 |
| 51 | LCB | rating on scale of 100 | string | 47+2 |
| 52 | CB | rating on scale of 100 | string | 47+2 |
| 53 | RCB | rating on scale of 100 | string | 47+2 |
| 54 | RB | rating on scale of 100 | string | 59+2 |
| 55 | Crossing | rating on scale of 100 | numeric | 84 |
| 56 | Finishing | rating on scale of 100 | numeric | 95 |
| 57 | HeadingAccuracy | rating on scale of 100 | numeric | 70 |
| 58 | ShortPassing | rating on scale of 100 | numeric | 90 |
| 59 | Volleys | rating on scale of 100 | numeric | 86 |
| 60 | Dribbling | rating on scale of 100 | numeric | 97 |
| 61 | Curve | rating on scale of 100 | numeric | 93 |
| 62 | FKAccuracy | rating on scale of 100 | numeric | 94 |
| 63 | LongPassing | rating on scale of 100 | numeric | 87 |
| 64 | BallControl | rating on scale of 100 | numeric | 96 |
| 65 | Acceleration | rating on scale of 100 | numeric | 91 |
| 66 | SprintSpeed | rating on scale of 100 | numeric | 86 |
| 67 | Agility | rating on scale of 100 | numeric | 91 |
| 68 | Reactions | rating on scale of 100 | numeric | 95 |
| 69 | Balance | rating on scale of 100 | numeric | 95 |
| 70 | ShotPower | rating on scale of 100 | numeric | 85 |
| 71 | Jumping | rating on scale of 100 | numeric | 68 |
| 72 | Stamina | rating on scale of 100 | numeric | 72 |
| 73 | Strength | rating on scale of 100 | numeric | 59 |
| 74 | LongShots | rating on scale of 100 | numeric | 94 |
| 75 | Aggression | rating on scale of 100 | numeric | 48 |
| 76 | Interceptions | rating on scale of 100 | numeric | 22 |
| 77 | Positioning | rating on scale of 100 | numeric | 94 |
| 78 | Vision | rating on scale of 100 | numeric | 94 |
| 79 | Penalties | rating on scale of 100 | numeric | 75 |
| 80 | Composure | rating on scale of 100 | numeric | 96 |
| 81 | Marking | rating on scale of 100 | numeric | 33 |
| 82 | StandingTackle | rating on scale of 100 | numeric | 28 |
| 83 | SlidingTackle | rating on scale of 100 | numeric | 26 |
| 84 | GKDiving | rating on scale of 100 | numeric | 6 |
| 85 | GKHandling | rating on scale of 100 | numeric | 11 |
| 86 | GKKicking | rating on scale of 100 | numeric | 15 |
| 87 | GKPositioning | rating on scale of 100 | numeric | 14 |
| 88 | GKReflexes | rating on scale of 100 | numeric | 8 |
| 89 | Release Clause | release clause value | string | €226.5M |

# Methods used in the project

This section will be updated as the project progresses.

# Potential Issues?

At the beginning of the project, following were identified as potential issues or candidates for considerable data scrubbing effort:

* Since international player and currency data is involved, handling special character in the data is going to be a significant task (Śląsk Wrocław).
* There are a few players missing significant amount of information e.g. Preferred Foot, Skill Move etc.
* The currency fields (e.g. Release Clause) are in thousand or million formats. They need to be converted to numbers.
* Heights column need to be converted to single unit, e.g. inch and treated as numeric for using in the analysis.
* Weight column need to be converted to single unit, e.g. lbs. and treated as numeric for using in the analysis.

Some of the hypothesis evaluations were restricted due to unavailability of data. But all of the partially missing data has been filled up during the data cleansing phase.

# GitHub Repository

<https://github.com/anirbanpalDSC/Fifa-2019-EDA/>

# Appendix

LS – Left Stricker

ST - Striker

RS – Right Striker

LW – Left Winger

Winger (RW-LW)

This position is similar to the side midfielder but have fewer defensive duties. As a winger you will need great dribbling skills to get around your defenders. You will also need to provide your forwards with quality crosses from wide positions.

LF – Left Forward

RF/LF stands for Right Forward and Left Forward. Forward meaning striker. The wide attacking midfielders are usually the ones to feed the ball to forwards and sit deeper in field

CF - Centre forward

Your main task as center forward is to score goals. It doesn't matter how you do it but without scoring goals your status as center forward will be low.

RF - Right Forward

Refer to LF.

RW – Right Winger

Refer to LW

LAM – Left Attacking midfielder

Attacking midfielder (AM)

Attacking midfielders are also called playmakers and their ball skills (shooting, passing, first touch on the ball) should be of high quality.

CAM - Center Attacking midfielder

Refer to LAM

RAM - Left Attacking midfielder

Refer to LAM

LM – Left Midfielder

Side midfielder (LM-RM)

As a side midfielder you will be placed either left or right (depending on your kicking foot) of the midfield. One of your main tasks is to provide your forwards with quality crosses.

LCM – Left Central Midfielder

CM - Central midfielder

As a (CM) you must be ready to play more than one role on the field, depending on your team's strength and tactics. You will be the link between your defense and offense. You must also help in defense when your opponents have possession of the ball.

RCM - Right Central Midfielder

RM – Right Midfielder

Refer to LM

LWB – Left Wing Back

LDM – Left Defensive Midfielder

CDM - Central Defensive Midfielder

RDM - Right Defensive Midfielder

RWB – Right Wing Back

LB – Left Back

As a full back you will be responsible to prevent your opponents from getting the ball into your own 18 yard box (penalty area). You will also need to make offensive runs on the flank especially in a formation like 4-4-2. A full back (FB) could be Left Back (LB) or Right Back (RB).

LCB – Left Center Back

CB – Center Back

If you are playing as CB (even called center halves or central defenders) your main task will be to mark the opposing team forwards (and even their CM (central midfielders).

RCB – Right Center Back

RB – Right Back

Refer to LB

# Reference

1. Tuo Q, Wang L, Huang G, Zhang H and Liu H (2019) Running Performance of Soccer Players During Matches in the 2018 FIFA World Cup: Differences Among Confederations. Front. Psychol. 10:1044. doi: 10.3389/fpsyg.2019.01044
2. *Welcome to FIFA EDA*, Retrieved from <https://www.kaggle.com/ritesaluja/fifa19/>
3. Wiseman, Marc (2019), *Machine Learning using FIFA 2019*, Retrieved from <https://www.linkedin.com/pulse/machine-learning-using-fifa-2019-marc-wiseman/>
4. Verma, Akshaj (2018), *Exploratory Analysis of FIFA 18 dataset using R*, Retrieved from <https://towardsdatascience.com/exploratory-analysis-of-fifa-18-dataset-using-r-ba09aa4a2d3c>
5. Prayag, Devesh (2019), *Insights from FIFA 2019 Player Data Set*, Retrieved from <https://medium.com/@deevesh/insights-from-fifa-2019-player-data-set-cdcfc9304bda>
6. Lu, Pilin (2019), *Does money affect performance? Analysis of the FIFA 19 data.*, Retrieved from <https://medium.com/@u0509421/fifa-19-complete-player-dataset-what-can-we-tell-from-it-f7343fdad020>
7. *Positions of Players on a Soccer Field Sheet*, Retrieved from <http://www.soccer-training-guide.com/positions-of-players-on-a-soccer-field.html#.XZA4L0ZKjMU>