

# README

May 26, 2022

## 1 FIFA 2019 World Cup Dataset

### 1.1 By Paul Anugwom

### 1.2 Overview of Dataset

This dataset consists of all the details of the players who participated in the FIFA World Cup 2019 and their final value at which they were to be sold to other the clubs. Data can be found in <https://www.kaggle.com/datasets/blurredmachine/fifa-2019-world-cup-dataset>. The data originally have 89 attributes(columns) and 18206 rows. After cleaning the data and removing columns not important for my use in this project, I had a dataset with 17899 rows and 18 columns.

This project has two parts that demonstrate the importance of data visualization in the data analysis process. I started by removing the columns not important for my use. I then went ahead to tidy the data by removing wrong values and zeros in the columns. I also converted the datatypes to the appropriate datatypes.

In the first part of exploration, I used Python visualization libraries to systematically explore the dataset, starting from plots of single variables to plots of multiple variables. In the second part, I produced a short presentation that illustrates interesting findings, and relationships that I discovered in the first part. The primary method of conveying my findings will be through transforming the exploratory visualizations from my exploration part into a polished explanatory visualizations.

### 1.3 Summary of Findings

The following are the key insights from the exploration of my dataset:

- 1, Most of the players in this tournament are within the 20s age
- 2, Most of the players in the dataset are from England, and most of the players play in English league
- 3, Left Foot(LF) position players are more valued than the rest position. Also players in Left wing back(LWB) and right wing back(RWB) positions are the least valued.
- 4, Strikers(ST) are the most popular players position in the dataset
- 5, Players with normal body-type are more in the dataset, also most of the players in the datasets are right footed.
- 6, Most of the players work-rate are medium/medium
- 7, More of the players with medium/medium work-rate are more valued and earns high wages
- 8, Players with high international-reputation have high overall rating and potential
- 9, There is a high positive correlation between players values and their wages

10, Players values and wages have a positive correlation with international-reputation, potential and overall rating of players

11, Stocky players are the least popular in this dataset. But Stocky players who are left footed are the most valuable and earn more than the normal and lean players

#### **1.4 Key Insights for Presentation**

For the purpose of the presentation, these major insights will be presented in a polished explanatory visualization:

- 1, Most of the players in the dataset are from Europe, and majority of them are from England
- 2, Players with normal body-type are more in the dataset, also most of the players in the datasets are right footed.
- 3, There is a high positive correlation between players values and their wages
- 4, Players values and wages have moderate positive correlation with international-reputation
- 5, Majority of the right footed players are highly valuable and earn high wages
- 6, More number of the Lean and Normal body-type players are highly valuable and earn high wages as compared to the Stocky players.
- 7, Stocky players are the least popular in this dataset. But few Stocky players who are left footed are the most valuable and earn more than the normal and lean players.