

Data Visualization for FIFA 20 Players Data

Venkata Ravi Teja Takkella

113219890

venkataravite.takkella@stonybrook.edu

Video Link :

https://drive.google.com/file/d/1fJNLn9mJP_z1BtyF0PTyoQWfoFCA6bNa/view?usp=sharing

Data source :

https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset?select=players_15.csv

The above mentioned link contains 104 columns, but I have pre-processed the above data using a **python script** to just bring out the below mentioned columns for this assignment. I've also removed the data points for which few of the fields were missing to bring out a more understandable visualization.

The above data has been scraped from **sofifa** website by the composer to obtain all the below mentioned values.

Data points :

There are 15466 data points available from the above link. For the ease of processing and better visualization, I've taken the 500 top rated players to proceed with this assignment.

Features :

Field Name	Description	Variable
short_name	Name of the player	Categorical
age	Age of the player	Categorical
height_cm	Height of the player	Numerical
weight_kg	Weight of the player	Numerical
nationality	Nationality of the player	Categorical
club	Playing football club of the player	Categorical
overall	Overall rating of the player	Numerical

preferred_foot	Preferred foot of the player	Categorical
skill_moves	Skill rating of the player	Categorical
body_type	Body type of the player	Categorical
team_position	Position in the team of the player	Categorical
team_jersey_number	Jersey Number of the player	Numerical
pace	Pace rating of the player	Numerical
shooting	Shoot rating of the player	Numerical
passing	Pass rating of the player	Numerical
dribbling	Dribble rating of the player	Numerical
defending	Defence rating of the player	Numerical
physic	Physic rating of the player	Numerical

Motive :

I generally play a lot of FIFA on my PS, making me compare the ratings, age, weight, physique of the players and find the interesting aspects which category/value makes the player better. It always used to be a huge task to check the ratings on the fifa data websites. But this course gave me a chance to use my knowledge to come up with a platform to show the above data in a much better way (Bar/Histogram/Scatter plots).

The data obtained has many features which bring out the exact representation of the player in the real world. It does show the comparisons between age/pace (which generally decreases while other increases), country/rating (May be depending on the facilities available), club/players, height/pace etc. This representation can also help us to generalise few features for any further assessment of the new players.

This motive helped me to pick the mentioned data to proceed with this assignment.

Noteworthy points :

- 1) I've done the **pre-processing** to just obtain the columns required to display. This helps to decrease the load of reading the data while generating the bar/histogram/scatter plots.
- 2) As part of the pre-processing, I added a new line at the end of the CSV to categorize the features to either **categorical/numerical**.
- 3) I've also done the processing of **removing** the data points from the visualisation where few feature values are not available. This helps in normalizing the data by removing the outliers.

- 4) The code I've written is so **generic** such that it just depends on the CSV. The dropdown values, labels of the axis, scales are all made generic just based on the CSV.
- 5) I've also added an **extra feature** of filtering the data for huge categorical data points.
For example:
The number of clubs is so huge that it is pretty difficult to visualise on a barchart. I've added a new filter "nationality" for now, so we can visualise only the clubs of the players from the provided country. This helps to provide more insights on each player and country.
- 6) The above mentioned filter is also generic and brought from a **specific field** provided in the CSV which is added during pre-processing.

Setup and Implementation :

- 1) The original data csv is placed under **data/players_15.csv**.
- 2) Pre-processed data csv is placed in the **data/fifa_processed.csv** folder.
- 3) The preprocessing python script is placed under the **scripts/processing.py**.
- 4) The entire project is built on **Javascript/Jquery** using **D3 Version 4**.
- 5) The javascript files are placed under **js/**, css files under **css/**, images under **img/**, d3 libraries under **lib/**.
- 6) This project needs to be brought up on a **server** of any choice to make the code work. I've used the "**Live Server**" extension of VS code for building.
- 7) Code related to barchart is in **js/barchart.js**, histogram is in **js/histogram.js**, scatter plot in **js/scatterplot.js**, rest in **js/index.js**.