

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Project 1 - Team 5

Dataset Name: FIFA23 OFFICIAL DATASET,

Dataset: https://www.kaggle.com/datasets/bryanb/fifa-player-stats-database?select=FIFA18_official_data.csv

Time Frame: 2021 to 2022

Reasons why this data set was used:

The dataset is the "FIFA23 OFFICIAL DATASET," sourced from Kaggle, and it contains player statistics for FIFA 23, a popular football video game. This dataset is valuable for various reasons, making it suitable for data analysis in the context of football player performance and attributes.

Data Source and Authenticity: The dataset is an official FIFA dataset, indicating that it comes from a reliable and authoritative source. This enhances the credibility of the data and ensures that it is accurate and relevant to the FIFA 23 game.

Comprehensive Player Attributes: The dataset includes a wide range of attributes for each player, such as their age, nationality, overall rating, potential, club affiliation, value, wage, preferred foot, height, weight, jersey number, and best overall rating. These attributes provide rich information to analyze and understand player performance and characteristics.

Historical Data: The dataset contains data for FIFA 23, which is the latest edition of the game. As a result, it offers the most recent player statistics, making it useful for conducting analyses based on up-to-date information.

Large Sample Size: The dataset includes data for a significant number of football players, making it suitable for conducting comprehensive statistical analyses and visualizations.

Player Performance Metrics: Attributes such as overall rating, potential, value, and wage provide insights into a player's skills, market value, and earning potential. This information is crucial for evaluating player performance and market influence.

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Visualization Possibilities: With various attributes available, the dataset allows for the creation of diverse visualizations, such as bar charts, scatter plots, pie charts, and geographical maps. These visualizations can aid in understanding player trends and comparisons.

Overall, the FIFA23 OFFICIAL DATASET is a valuable resource for exploring and analyzing player performance and attributes in FIFA 23. It offers comprehensive information on players, club affiliations, and international representation, making it a suitable dataset for various data analysis and visualization tasks in the domain of football analytics.

Objective and questions to be solved:

- Which countries and clubs have the highest and best representation of players?
- Which clubs are the most diverse (representation by Country)?
- How do the economic indicators of players vary based on their position and some characteristics?
- What representation does Mexico have in the dataset?
- What are some fun facts?

Executive technical resume:

The code performs data analysis on the player information from FIFA21 and FIFA22 games. Below is a summary of the main actions and visualizations carried out:

The data of players from FIFA21 and FIFA22 is loaded into two separate DataFrames, and a "Year" column is added to distinguish between both years.

The two DataFrames are combined into a single one, named "merged_df_clean," keeping only relevant columns for analysis.

Data transformations are performed for the "Value" and "Wage" columns, removing symbols and converting them into numeric values.

A bar chart is generated to display the top 5 countries with the most players in the dataset, using the "Nationality" column.

Players with an "Overall" rating greater than 75 are filtered, and a bar chart is created to show the countries with the best players in terms of "Overall Rating."

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

The Geoapify API is utilized to obtain the geographic coordinates (latitude and longitude) of the countries with the best players, and a map is displayed with these locations.

A bar chart is generated to display the distribution of players by their preferred foot ("Preferred Foot").

A scatter plot is created to explore the correlation between "Overall Rating" and "Potential" of the players.

Additional visualizations include histograms of the distribution of values, wages, and ratings of players, as well as a table showing the top 10 players with the highest "Overall" and "Potential" ratings in FIFA22.

Overall, the code performs exploratory data analysis and interesting visualizations on the players from FIFA21 and FIFA22. It provides insights into the distribution of players by country, their skills, wages, and values, as well as the correlation between their ratings. The presented visualizations and charts help gain a better understanding of player characteristics and identify patterns and trends in the data

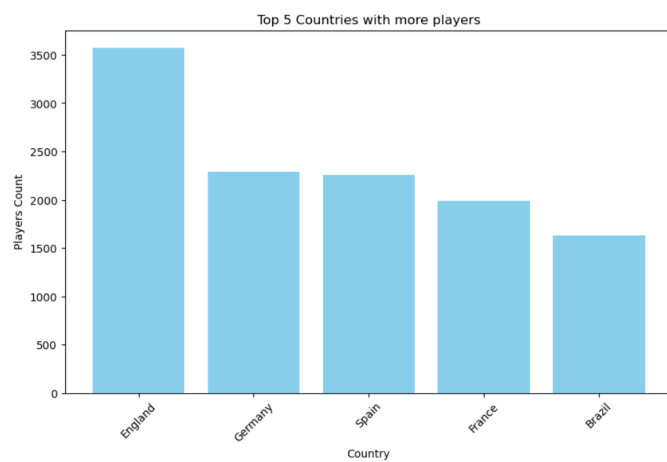
Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Analysis:

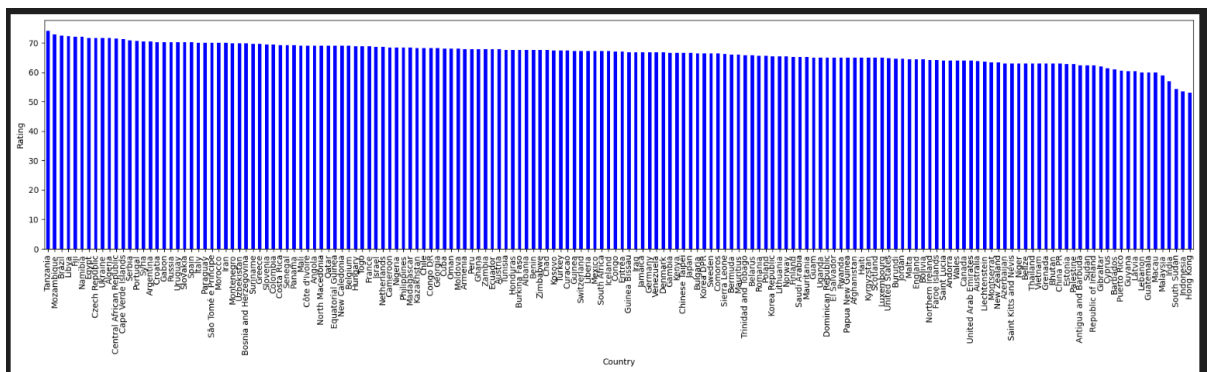
Countries

Count of Total number of countries in FIFA22 : 168

Top 5 countries with more players merging both years are:



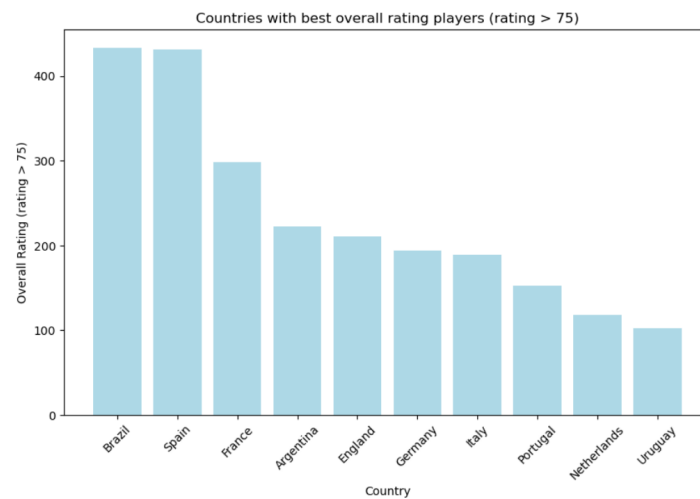
Calculation of the countries with the best avg overall players in the world per year:



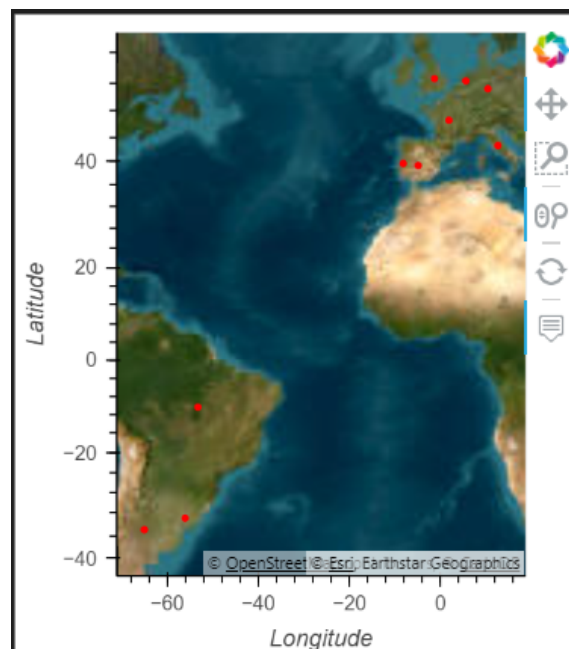
This might sound unexpected, Tanzania is the #1 country with the best players because it has a small number of players with a really good overall rating. Not really an interesting change between countries happened from one year to another

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Calculation of the countries with the best overall rating players (rating > 75)



With the API of Geopify we mapped the top countries with best overall players



Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

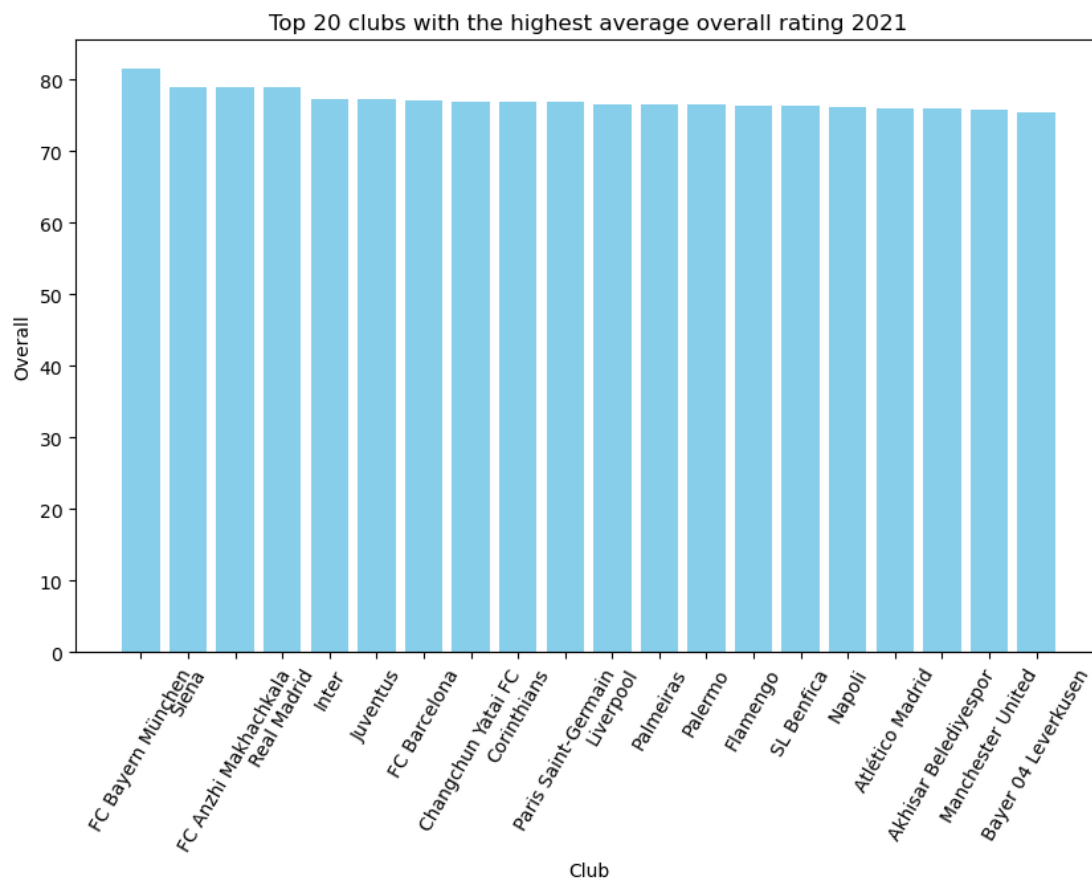
Club

The total number of clubs in FIFA21: 843

The total number of clubs in FIFA22: 869

Analysis of the top 20 clubs with the highest overall rating in FIFA21:

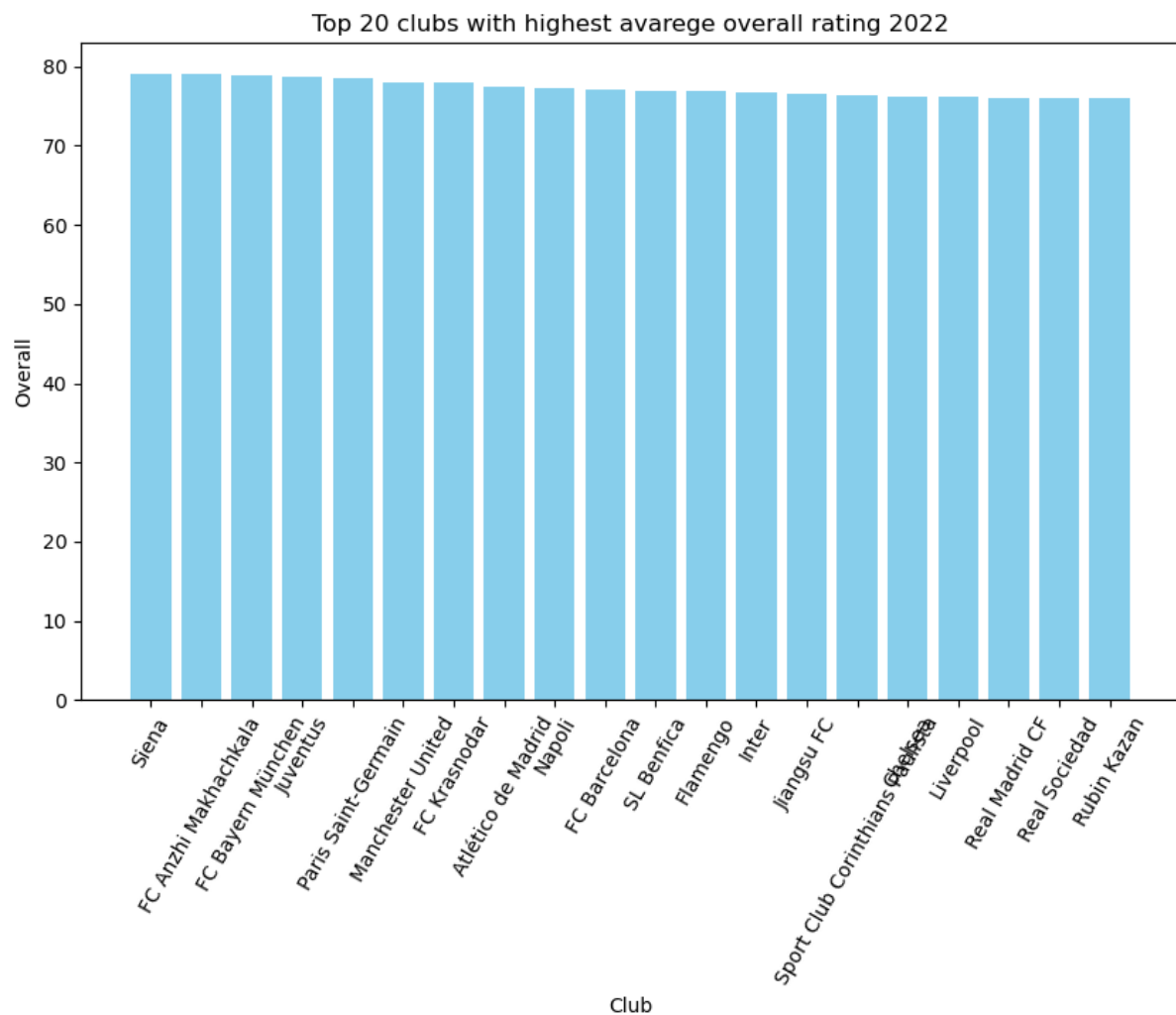
In 2021, the german club Bayern Munich was the club with the highest overall rating in their players, followed by Siena as second and Anzhi Makhachkala as third:



Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Analysis of the top 20 clubs with the highest overall rating in FIFA22.

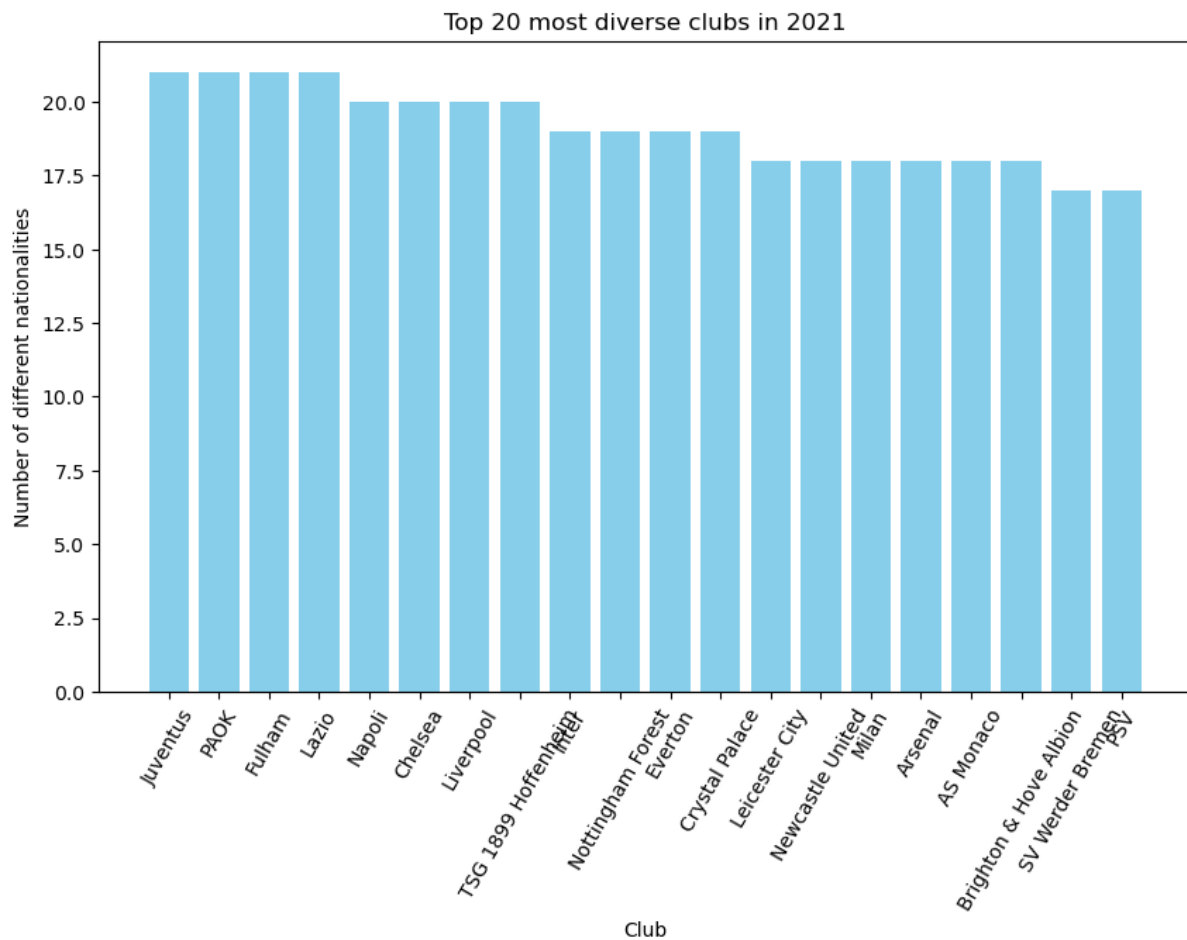
In 2022, the top 3 clubs with the highest overall rating in their players was repeated, only this time Siena was first, Anzhi Makhachkala second and Bayern Munich third.



Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Analysis of the most diverse clubs in FIFA21:

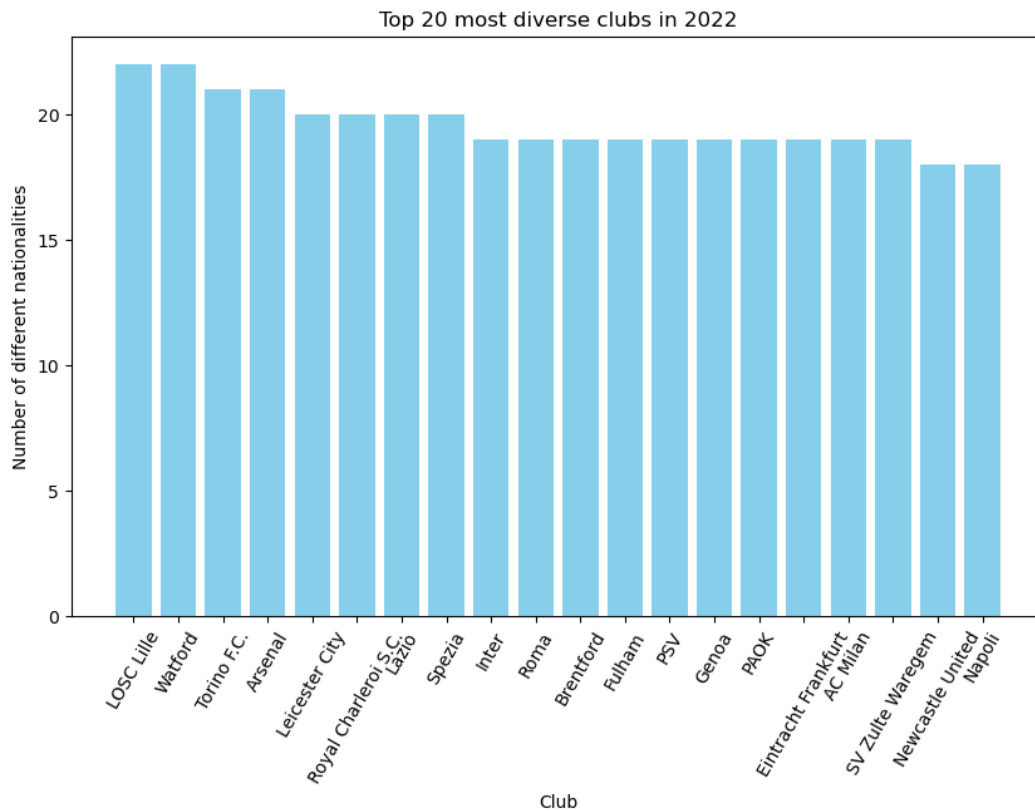
In terms of diversity in FIFA21, Juventus from Italy was the most diverse club along with PAOK from Greece, Fulham from the UK and Lazio from Italy, all of them with 21 players from different nationalities.



Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

Analysis of the most diverse clubs in FIFA22:

In 2022 Lille from France and Watford were the most diverse clubs with 22 different nationalities in their squad:



Players

FIFA 21

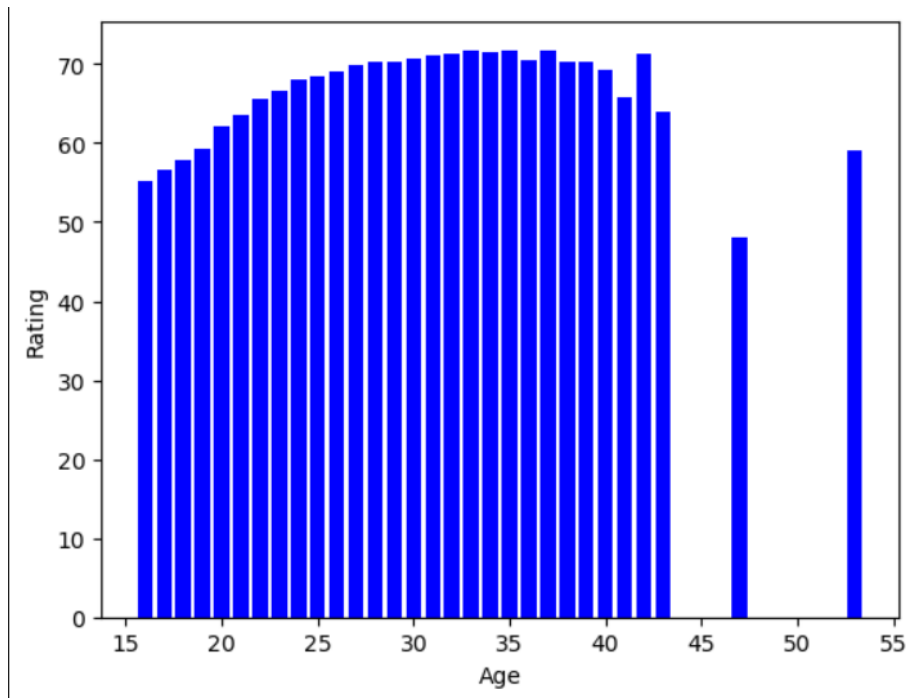
- 17108 players
- the average age is 25.1
- The oldest player was 53 years
- The average rating of players is 66.8
- the worst rated player had 38 rating
- The highest rated player is Messi with 93

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

FIFA 22

- 16710 players
- the average age was 25.7 is older than FIFA 21
- the oldest player was 54 years old the player is still active from FIFA 21
- The average rating of the players was 67.6 the rating goes up from FIFA 21
- the worst rated player had 28 rating
- The highest rated player is Messi with 93 Messi remains as the best player

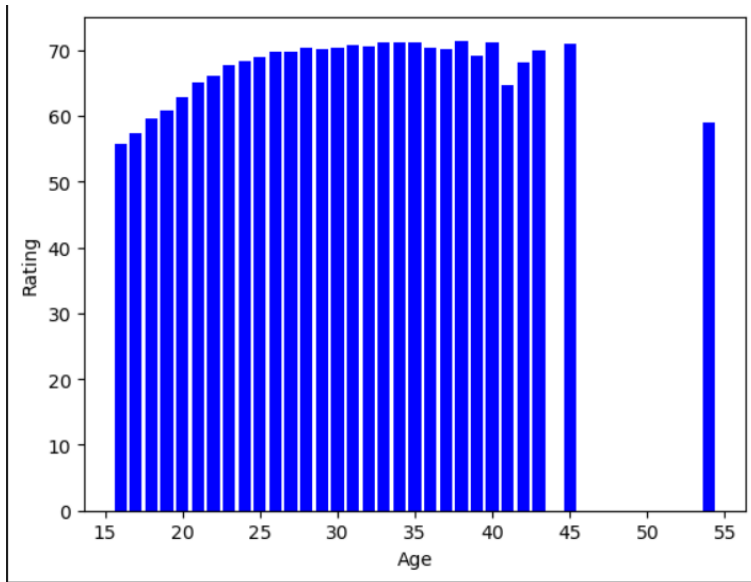
Comparing Age vs Rating FIFA 21



The correlation between both factors is 15 which indicates a low and weak factor but still the older the player is more likely to have better rating. At the age of 35 the rating of the players goes down drastically with exception at age 42 because there are a lot of goalkeepers at that age with good rating.

Comparing Age vs Rating FIFA 22

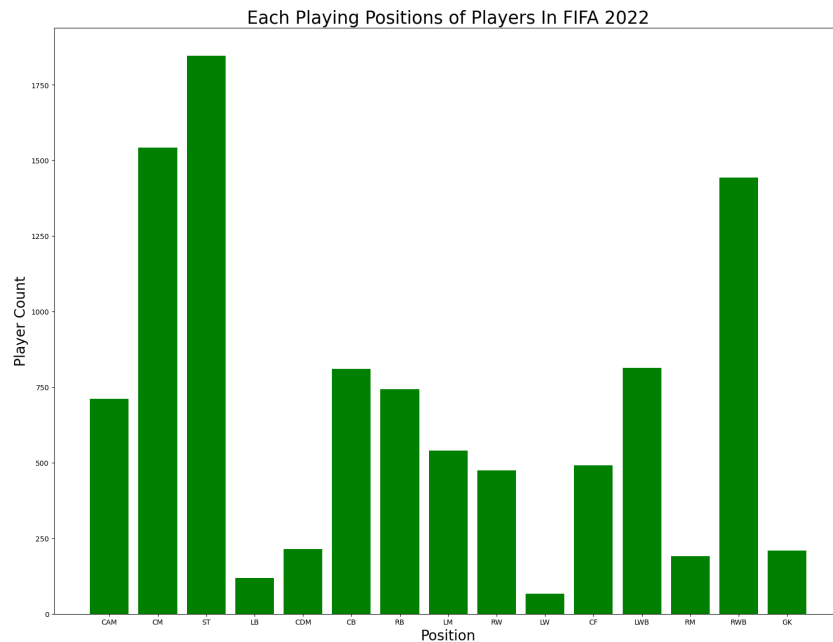
Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea



We can see that younger players have lower ratings than older players, this means that an older brother is likely to be better. The correlation between both factors is 42 which shows a decent correlation. There are few players aged 35 and older but the ones that still play are still at the highest level and are more experienced

Looking at the overall player count by 'Position' we see that the most played positions are "ST", "CM", and "RWB".

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea



Having grouped all positions by 'Defender', 'Forward', 'Goalkeeper', and 'Mid Field', we can look at the 5 best players by said categories.'

		Name	Nationality	Club	Best Position	Height	Overall	Potential
Defender	17349	V. van Dijk	Netherlands	Liverpool	CB	193cm	89	89
	17120	Sergio Ramos	Spain	Paris Saint-Germain	CB	184cm	88	88
	17233	17 P. Lahm	Germany	FC Bayern München	RWB	170cm	88	88
	19862	Rúben Dias	Portugal	Manchester City	CB	187cm	87	91
	17205	A. Robertson	Scotland	Liverpool	LB	178cm	87	88
Forward	17141	R. Lewandowski	Poland	FC Bayern München	ST	185cm	92	92
	17144	Cristiano Ronaldo	Portugal	Manchester United	ST	187cm	91	91
	17190	K. Mbappé	France	Paris Saint-Germain	ST	182cm	91	95
	17147	H. Kane	England	Tottenham Hotspur	ST	188cm	90	90
	17309	K. Benzema	France	Real Madrid CF	CF	185cm	89	89
Goalkeeper	31352	J. Oblak	Slovenia	Atlético de Madrid	GK	188cm	91	93
	29458	M. Neuer	Germany	FC Bayern München	GK	193cm	90	90
	30998	M. ter Stegen	Germany	FC Barcelona	GK	187cm	90	92
	28300	Ederson	Brazil	Manchester City	GK	188cm	89	91
	31716	G. Donnarumma	Italy	Paris Saint-Germain	GK	196cm	89	93
Mid field	17137	L. Messi	Argentina	Paris Saint-Germain	RW	170cm	93	93
	17172	Neymar Jr	Brazil	Paris Saint-Germain	LW	175cm	91	91
	17111	K. De Bruyne	Belgium	Manchester City	CM	181cm	91	91
	17179	N. Kanté	France	Chelsea	CDM	168cm	90	90
	17163	S. Mané	Senegal	Liverpool	LW	175cm	89	89

Fun facts

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

How many players of Mexican nationality participated in Fifa 2022?

	Name	Nationality	Club	Best Position	Height	Overall	Potential
17199	J. Corona	Mexico	FC Porto	RM	173cm	82	82
17214	H. Herrera	Mexico	Atlético de Madrid	CM	185cm	81	81
17330	A. Guardado	Mexico	Real Betis Balompié	CDM	169cm	78	78
17399	R. Jiménez	Mexico	Wolverhampton Wanderers	ST	188cm	83	83
17467	V. Guzmán	Mexico	Pachuca	CAM	175cm	75	78
...
33536	D. Ochoa	Mexico	Real Salt Lake	GK	188cm	66	76
33567	C. Moreno	Mexico	Pachuca	GK	190cm	59	68
33657	J. León	Mexico	Puebla FC	GK	188cm	57	63
33669	R. Castellanos	Mexico	Club Necaxa	GK	184cm	58	68
33677	C. Ramos	Mexico	CF Monterrey	GK	182cm	57	66

312 rows x 7 columns

In total, 312 Mexican players participated in Fifa 2022, and out of those 312, these were the top 10:

	Name	Nationality	Club	Best Position	Height	Overall	Potential
17399	R. Jiménez	Mexico	Wolverhampton Wanderers	ST	188cm	83	83
17967	C. Vela	Mexico	Los Angeles FC	RW	177cm	83	83
17199	J. Corona	Mexico	FC Porto	RM	173cm	82	82
17214	H. Herrera	Mexico	Atlético de Madrid	CM	185cm	81	81
17906	H. Lozano	Mexico	Napoli	RW	175cm	81	82
32353	G. Ochoa	Mexico	Club América	GK	190cm	80	80
17330	A. Guardado	Mexico	Real Betis Balompié	CDM	169cm	78	78
20425	J. Hernández	Mexico	LA Galaxy	ST	175cm	78	78
31985	A. Talavera	Mexico	Club Universidad Nacional	GK	186cm	78	78
17760	L. Rodríguez	Mexico	Tigres U.A.N.L.	RB	177cm	77	77

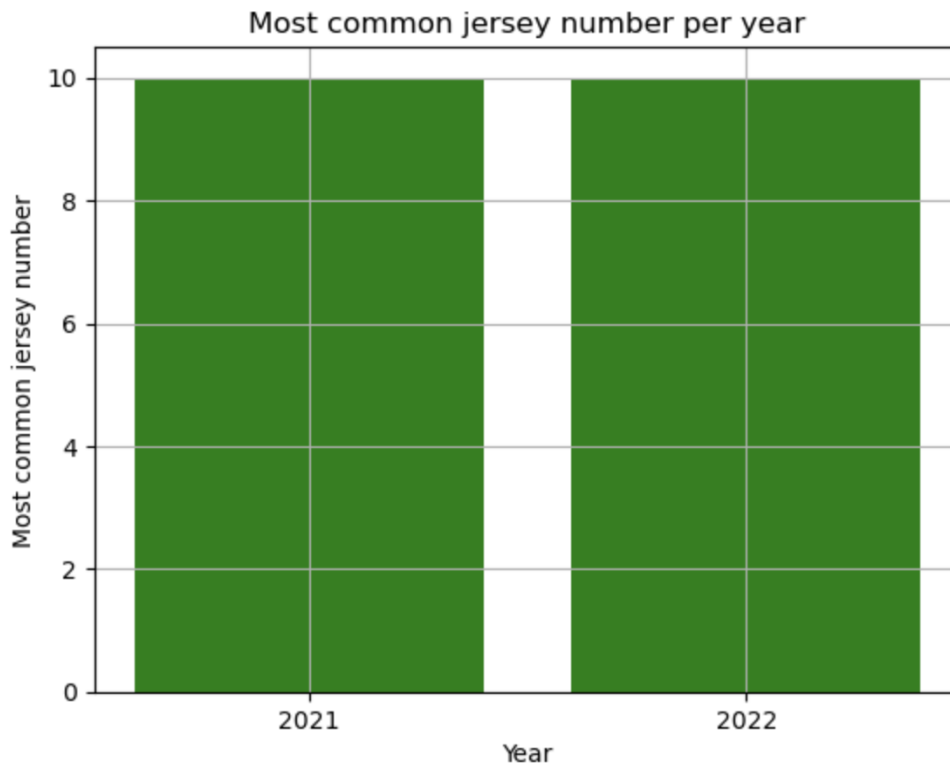
Most common jersey number per year

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

In both years, the most common jersey number were the number 10:

Año 2021: Número de Jersey más Común = 10.0

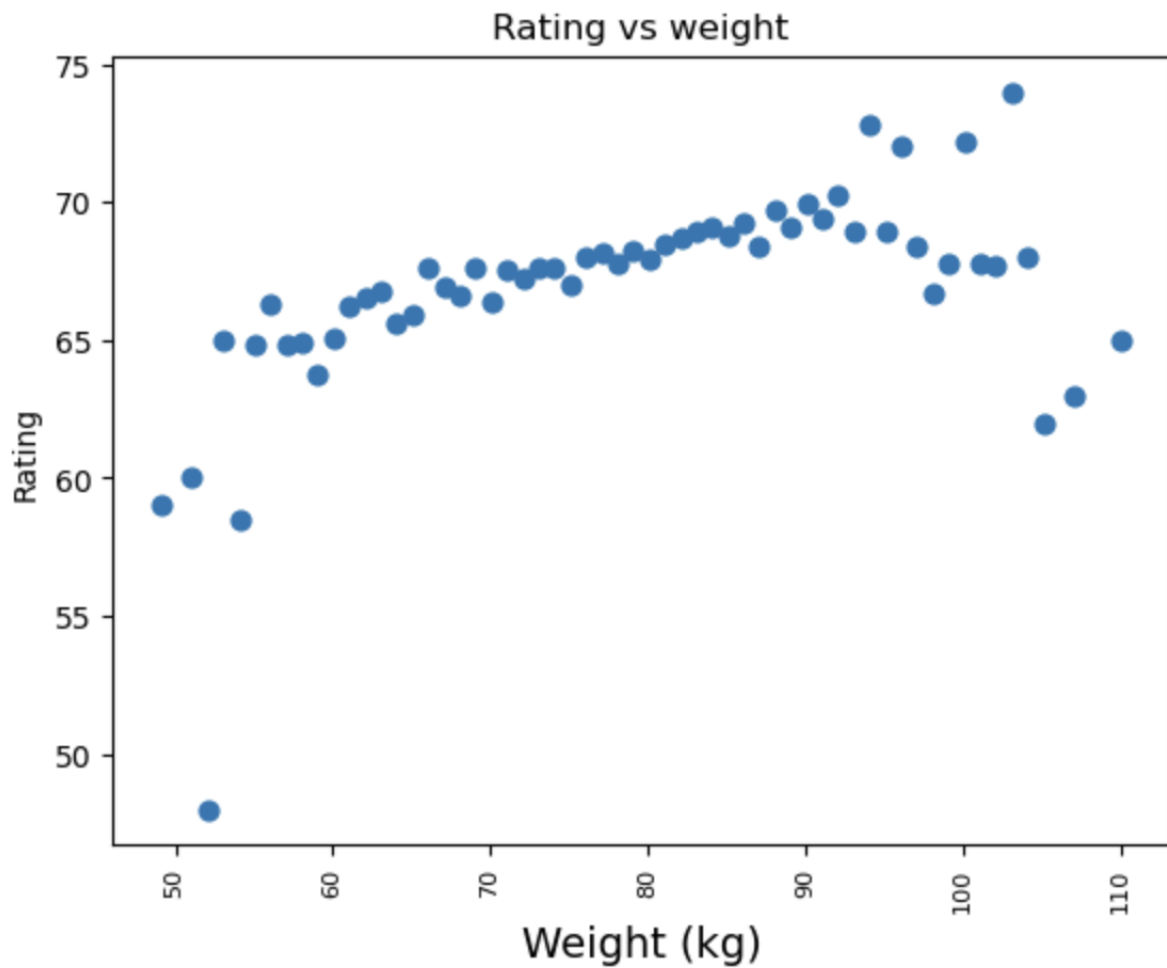
Año 2022: Número de Jersey más Común = 10.0



Is there a relationship between a player's weight and their overall rating?

We generated a scatterplot graph to analyze the relationship between the overall rating and the weight of a player, getting the following figure for 2022:

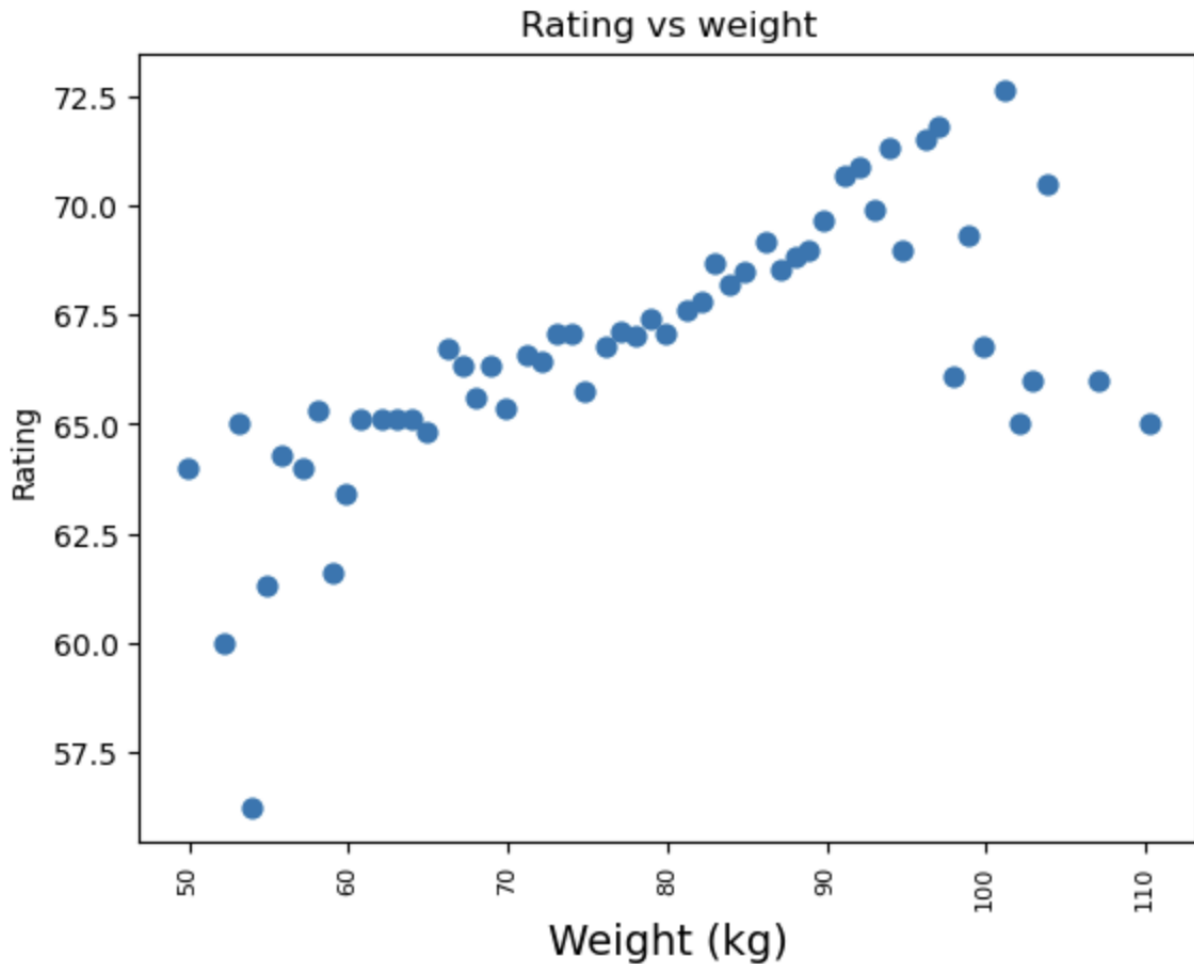
Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea



The Pearson correlation of both factors was 0.54, which means that there is not a strong positive correlation.

Meanwhile for 2021, the figure was as follows:

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea

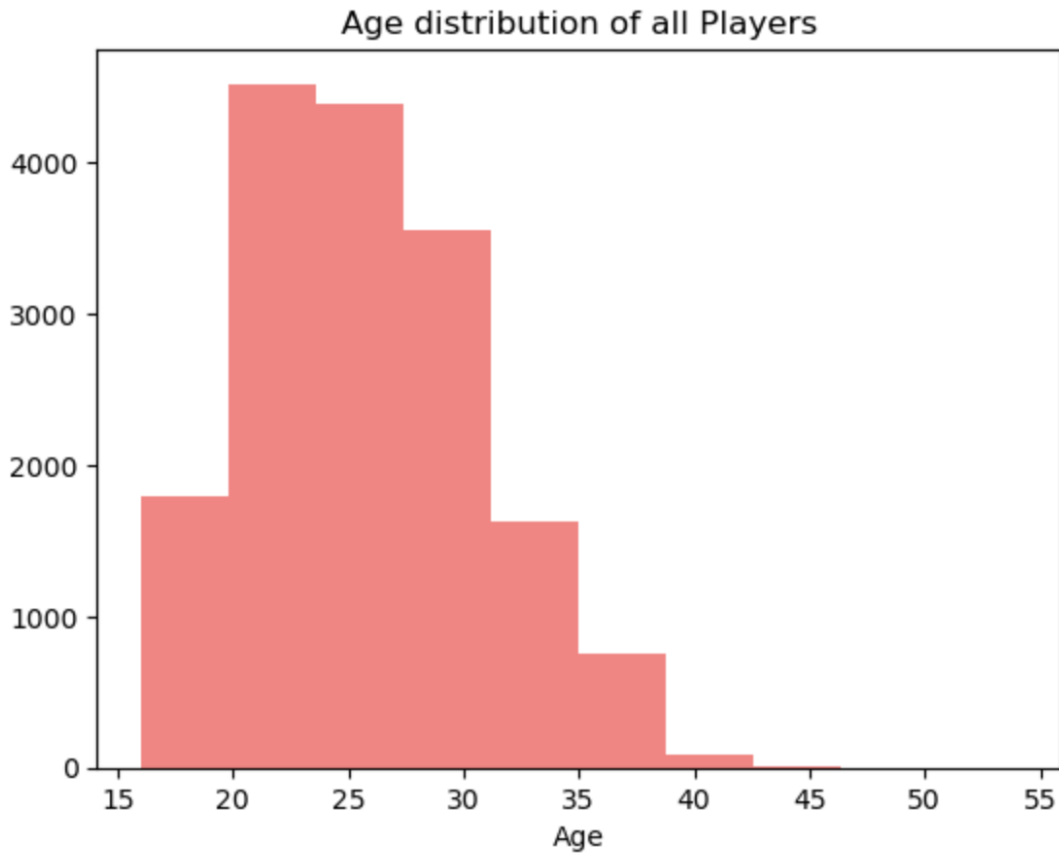


With a Pearson correlation of 0.69, meaning that there is a strong positive correlation between the weight and the rating of a player.

Analyzing the distribution of all players' age

A histogram was generated to visualize the distribution of the ages of all the players, obtaining the following figure:

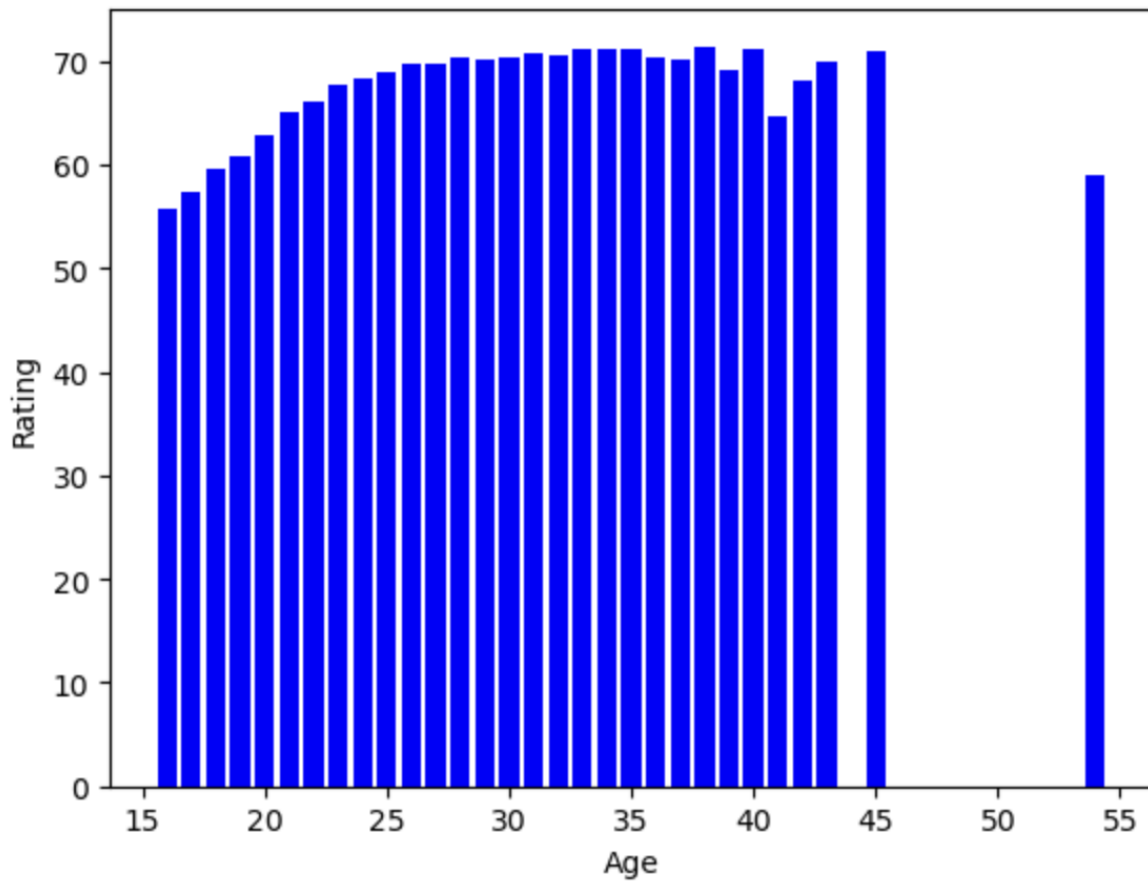
Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea



Then, the correlation between the ages and the overall ratings was obtained, first visualizing the relation in a bar graph per year:

2021:

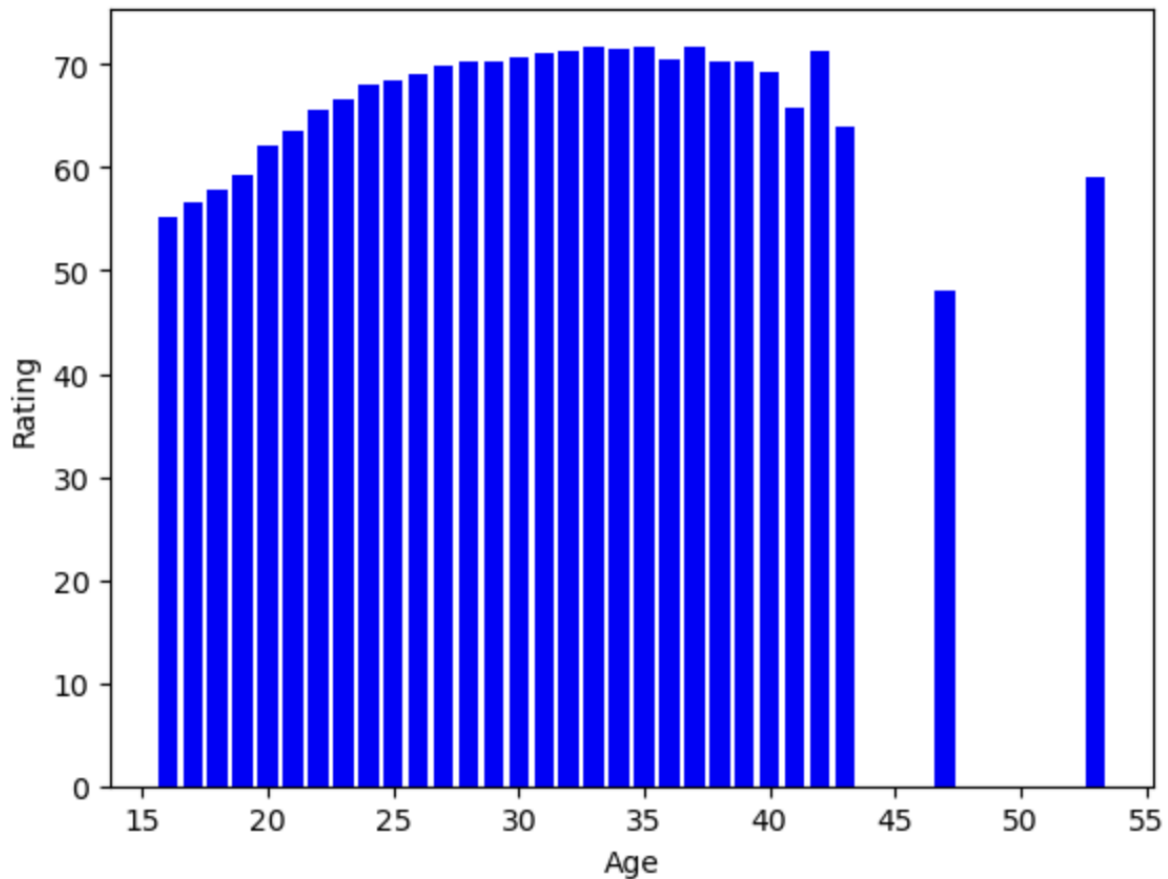
Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea



In 2021, the Pearson correlation between both factors was 0.42 which means there is not a strong positive correlation.

2022:

Ma. Fernanda Arana
Fabián Moreno
Víctor Olguín
Pablo Borboa
Eduardo Rea



While the Pearson correlation between both factors in 2022 was 0.15. Resulting in a weaker positive correlation.

Conclusions:

In conclusion, the presented code offers a comprehensive analysis of the FIFA datasets, providing valuable insights into player representation, economic indicators, player preferences, and top-ranking players. It showcases the power of data analytics in exploring and understanding complex datasets, offering valuable information to football enthusiasts and analysts alike.