

Fuse Machine Explotory Data Analysis 2021



Submitted to: Mr. Kamal Shrestha

Submitted by: Saijeet Upadhyay

Contents

Introduction	1
About Data	1
Data Analysis	2
1. Overall distribution of player based on Nation.....	2
2. Corelation between the properties of attributes of football players	3
3. Representation of players in different positions	4
4. Dominant foot.....	4
5. Overall rating of players.....	5
6. Age wise distribution of players.....	6
7. Attribution for forwards.....	6
8. Set piece Attributes.....	7
9. Overall relation between Overall, Age and Potential	8
10. Potential vs Age.....	8
11. Potential of Players	9
12. Top 10 overall potential Teams	10
13. Top 5 Overall Clubs	11
14. Wages on Players spend by Club	12
15. Wages based on Nation	13
16. Wages in terms of Penalties Aggression.....	14
17. Real Madrid Country representation.....	14

Introduction

Data scientists utilize exploratory data analysis (EDA) to examine and investigate data sets and describe their major features, typically using data visualization approaches. It aids in determining how to effectively modify data sources to obtain the answers required, making it simpler for data scientists to identify patterns, detect anomalies, test hypotheses, and validate assumptions. EDA is largely used to discover what data may disclose beyond the formal modeling or hypothesis testing tasks, and it offers a deeper knowledge of data set variables and their interactions. It can also assist you evaluate whether the statistical approaches you're thinking about using for data analysis are acceptable.

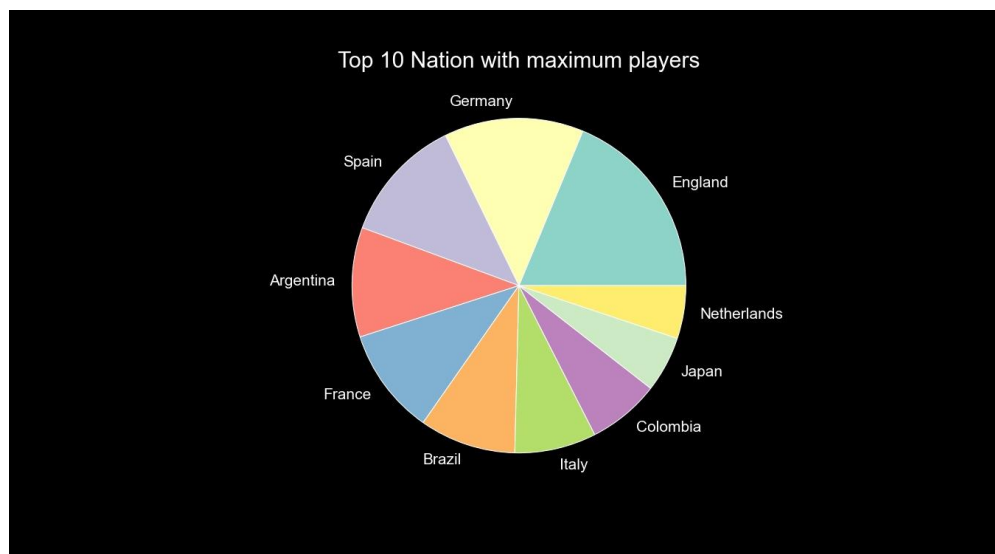
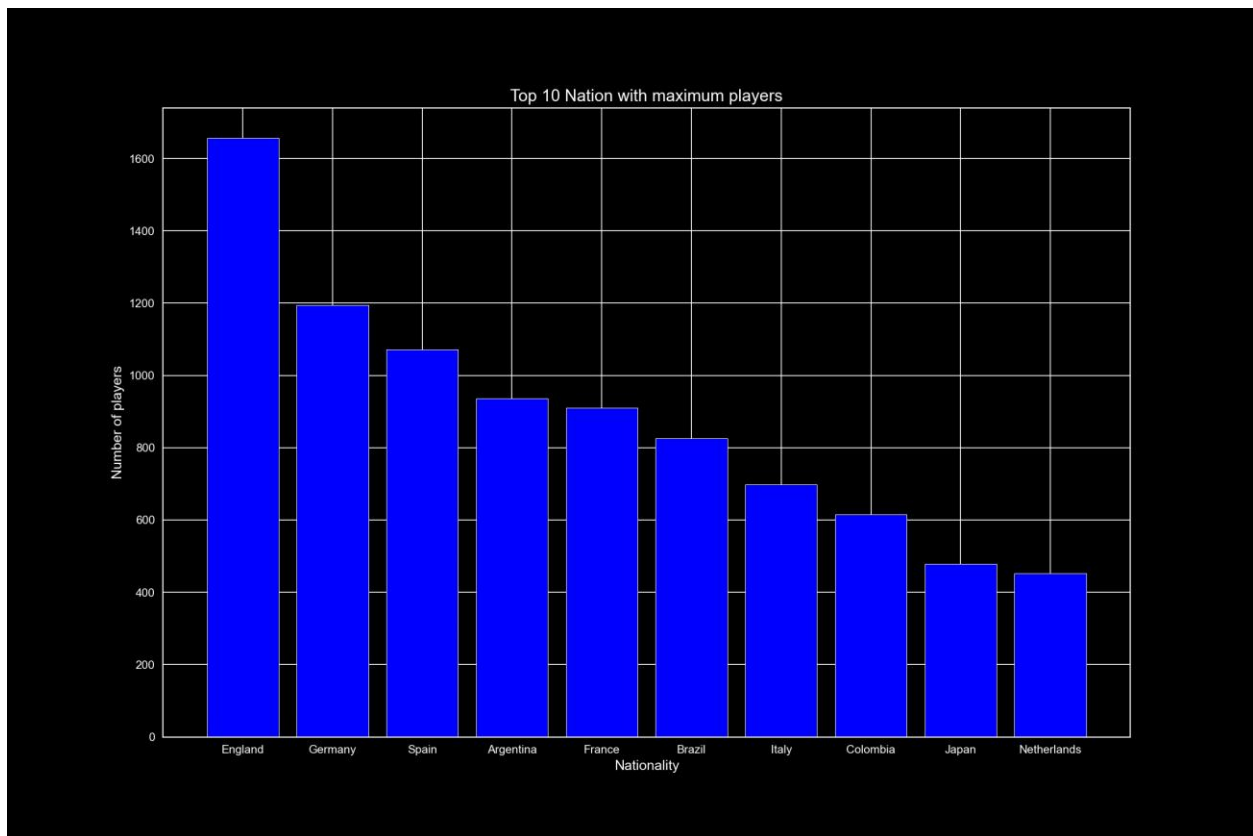
About Data

The given data is about Fifa 19 which has 18207 rows and 89 columns with zero missing data. There are 38 float 5 int and 45 objects. It has attributes 'Name', 'Age', 'Nationality', 'Overall', 'Potential', 'Club', 'Value', 'Wage', 'Special', 'Preferred Foot', 'International Reputation', 'Weak Foot', 'Skill Moves', 'Work Rate', 'Body Type', 'Real Face', 'Position', 'Jersey Number', 'Joined', 'Loaned From', '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', 'HeadingAccuracy', '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', 'GKReflexes', 'Release Clause'.

Data Analysis

1. Overall distribution of player based on Nation

The given figure below shows the composition of players in Fifa 19 based on nationality. It represents top 10 nations from where most of the players belong who play professional football in elite level who are official representation of real life players.

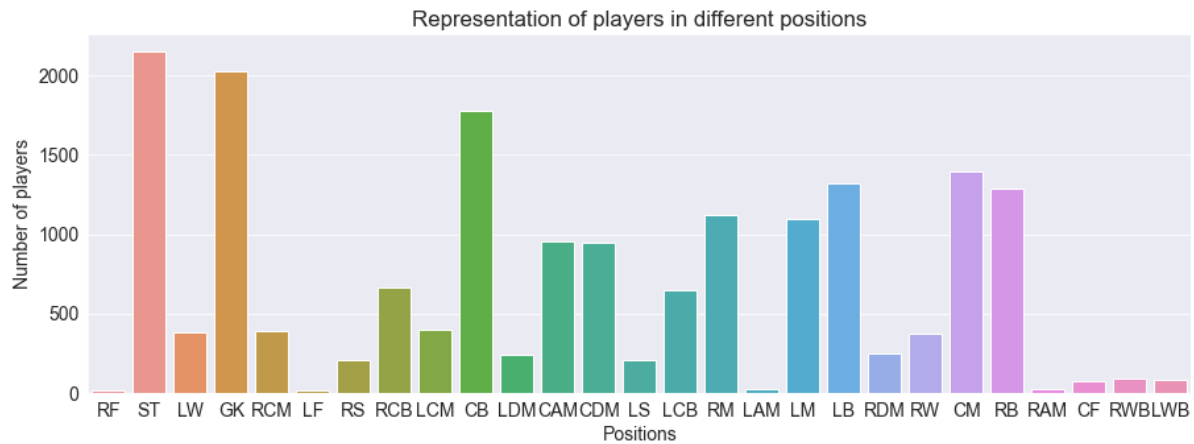


2. Corelation between the properties of attributes of football players

ID	10	-0.7	0.4	0.0	-0.2	-0.4	-0.1	-0.1	0.2	-0.1	-0.1	-0.1	-0.1	-0.2	0.0	-0.2	-0.2	-0.2	-0.1	0.1	0.1	-0.0	-0.4	0.0	-0.2	-0.2	-0.1	-0.3	-0.2	-0.2	-0.2	-0.1	-0.2	-0.1	-0.4	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	
Age	-0.7	1.0	0.5	-0.3	0.2	0.3	0.1	0.0	-0.2	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.2	0.1	-0.2	-0.2	-0.0	0.5	-0.1	0.2	0.2	0.1	0.3	0.2	0.3	0.2	0.1	0.2	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1		
Overall	0.4	0.5	1.0	0.7	0.6	0.5	0.2	0.4	0.2	0.4	0.3	0.5	0.4	0.4	0.4	0.4	0.5	0.5	0.2	0.2	0.3	0.8	0.1	0.4	0.3	0.4	0.4	0.3	0.4	0.3	0.4	0.5	0.3	0.7	0.3	0.3	0.2	-0.0	-0.0	-0.0		
Potential	0.0	-0.3	0.7	1.0	0.4	0.4	0.2	0.4	0.0	0.2	0.2	0.2	0.4	0.3	0.3	0.4	0.2	0.2	0.2	0.2	0.3	0.5	0.1	0.3	0.1	0.2	0.1	0.2	0.2	0.2	0.3	0.2	0.4	0.2	0.1	-0.1	-0.1	-0.1	-0.1	-0.1		
Special	-0.2	0.2	0.6	0.4	1.0	0.3	0.3	0.5	-0.1	0.9	0.7	0.6	0.9	0.8	0.9	0.8	0.8	0.8	0.9	0.7	0.6	0.7	0.8	0.1	0.8	0.3	0.8	0.2	0.8	0.7	0.8	0.8	0.8	0.7	0.8	0.6	0.5	-0.7	-0.7	-0.7	-0.7	
International Reputation	-0.4	0.3	0.5	0.4	0.3	1.0	0.1	0.2	-0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.1	0.4	0.1	0.2	0.1	0.1	0.1	0.2	0.2	0.1	0.2	0.3	0.2	0.4	0.1	0.1	0.1	0.0	0.0	0.0	0.0	
Weak Foot	-0.1	0.1	0.2	0.2	0.3	0.1	1.0	0.3	-0.0	0.3	0.4	0.2	0.3	0.4	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.2	0.3	0.1	0.2	-0.0	0.4	0.1	0.1	0.3	0.3	0.3	0.3	0.1	0.0	0.0	-0.2	-0.2	-0.2	-0.2	-0.2	
Skill Moves	0.1	0.0	0.4	0.4	0.8	0.2	0.3	1.0	0.0	0.7	0.7	0.4	0.7	0.8	0.8	0.7	0.6	0.8	0.7	0.6	0.7	0.4	0.3	0.7	0.1	0.6	0.0	0.3	0.2	0.8	0.7	0.7	0.6	0.2	0.2	0.2	0.6	0.6	0.6	0.6	0.6	
Jersey Number	0.2	-0.2	-0.2	-0.0	-0.1	-0.1	-0.0	-0.0	1.0	0.1	0.0	-0.1	0.0	-0.0	-0.1	-0.1	-0.1	0.0	-0.0	-0.0	-0.2	0.8	0.1	-0.1	-0.2	0.8	-0.1	-0.2	0.8	-0.1	-0.0	-0.2	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0		
Crossing	-0.1	0.1	0.4	0.2	0.9	0.2	0.3	0.7	0.1	1.0	0.7	0.5	0.8	0.7	0.9	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.7	0.4	0.8	0.7	0.1	0.7	-0.0	-0.7	0.5	0.4	0.8	0.7	0.8	0.6	0.4	0.4	-0.7	-0.7	-0.7	-0.7
Finishing	-0.1	0.1	0.3	0.2	0.7	0.2	0.4	0.7	0.0	0.7	1.0	0.5	0.7	0.9	0.8	0.8	0.7	0.5	0.8	0.8	0.8	0.3	0.5	0.8	0.1	0.5	-0.0	0.9	0.2	-0.0	0.9	0.7	0.8	0.5	0.0	-0.0	-0.1	-0.6	-0.6	-0.6	-0.6	
HeadingAccuracy	-0.1	0.1	0.3	0.2	0.6	0.2	0.2	0.4	-0.1	0.5	0.5	1.0	0.6	0.5	0.8	0.4	0.4	0.5	0.7	0.3	0.4	0.3	0.3	0.2	0.6	0.4	0.8	0.5	0.7	0.5	0.5	0.3	0.8	0.5	0.6	0.6	0.5	-0.8	-0.7	-0.7	-0.7	
ShortPassing	0.1	0.1	0.5	0.4</																																						

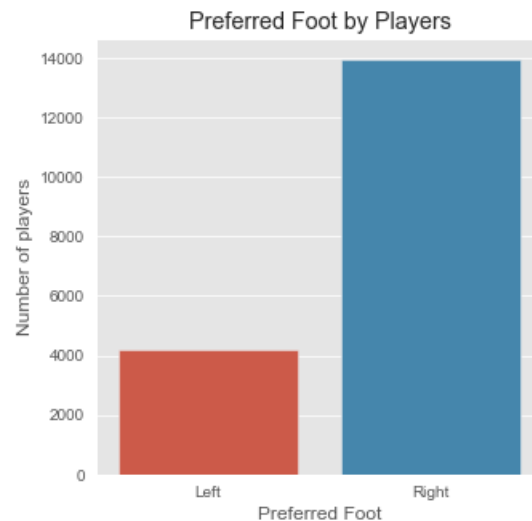
3. Representation of players in different positions

The graph below represents the number of player playing in different positions. The most famous one is of Striker followed by GK with least of left winger and LAM and RAM. The player are divided among different position throughout the data field.

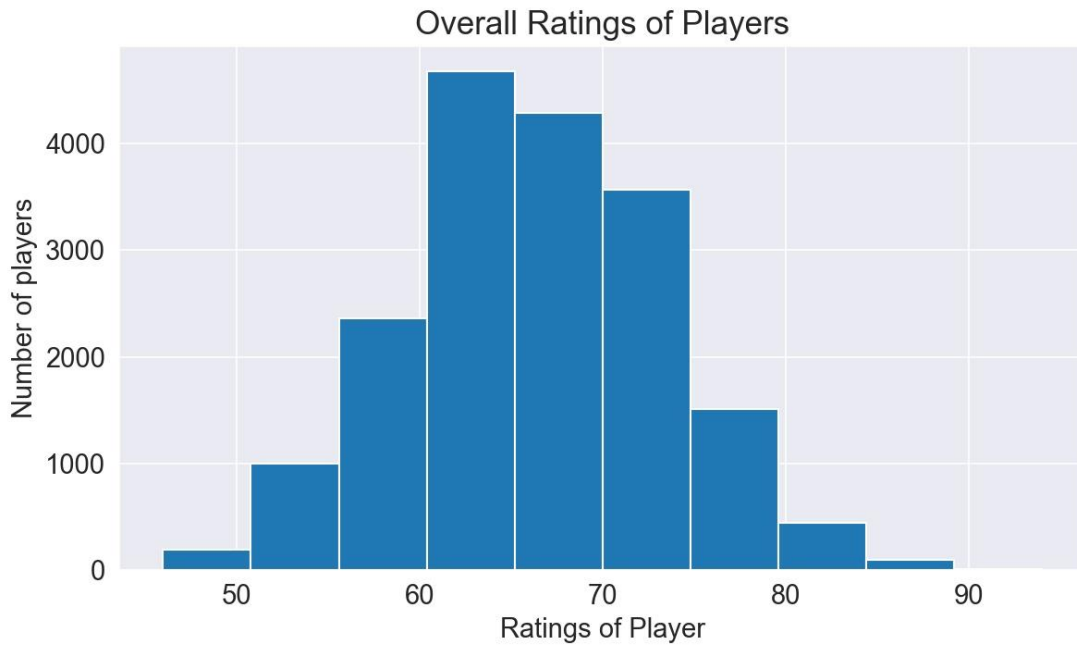


4. Dominant foot

From the graph given below it clearly shows that most player prefer right foot in comparison to their left. In summary most players have dominant right foot.

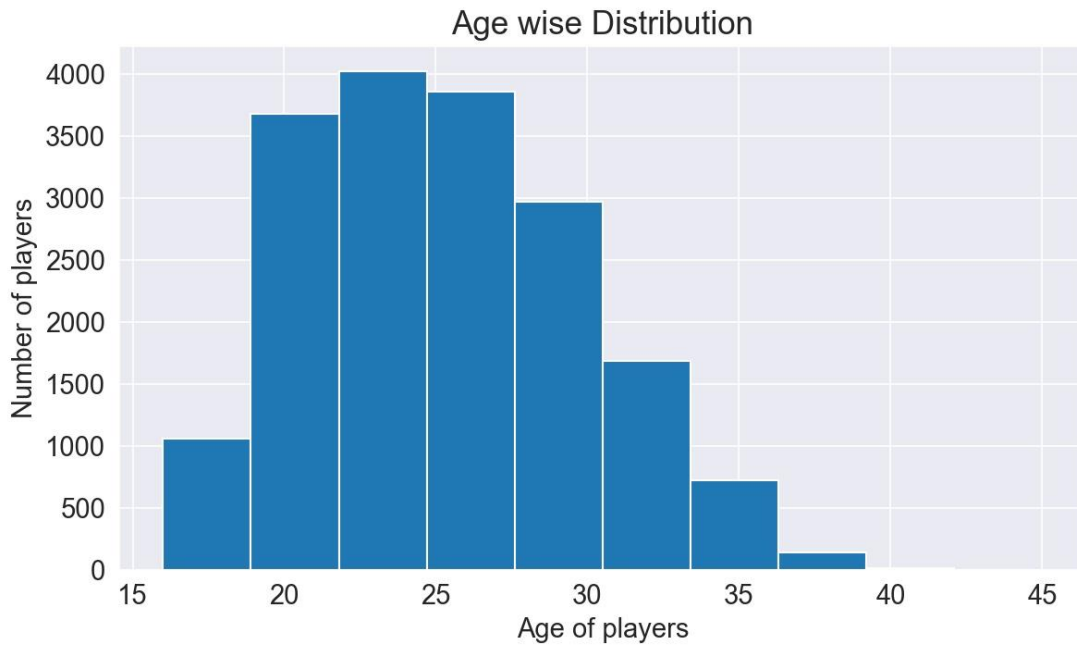


5. Overall rating of players



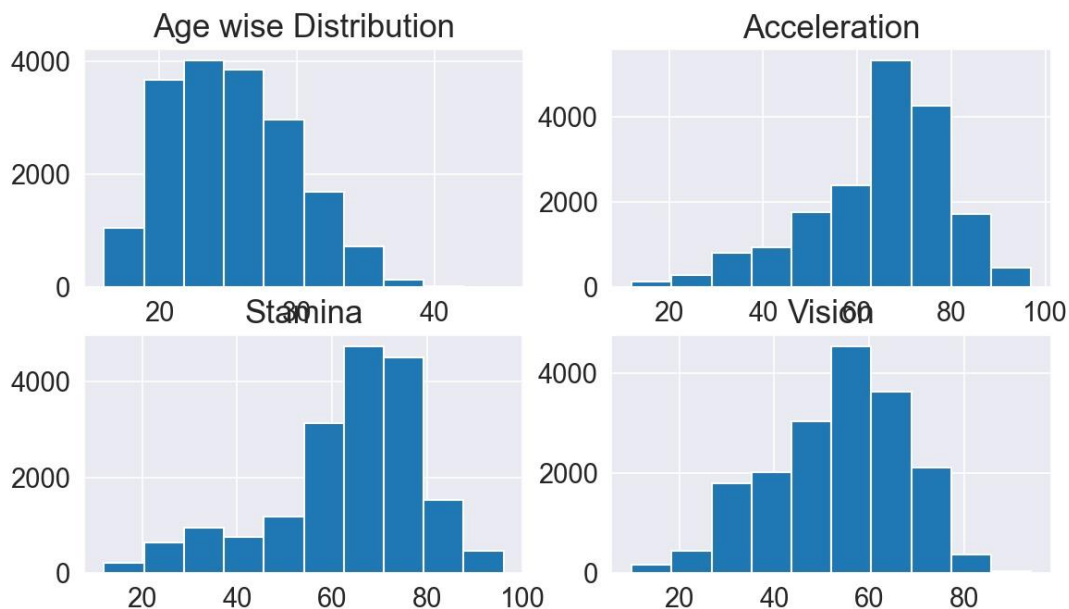
The given diagram shows the histogram of overall rating of player. Most of the player fall from 60-79 which can be concluded as average players. The elite player who can change the course of game can be seen very less falling in league above others which are 80-90. And there even players of very high caliber which represents the face of the game having rating 90+ who are everyone's favorites in very small/least fractions.

6. Age wise distribution of players



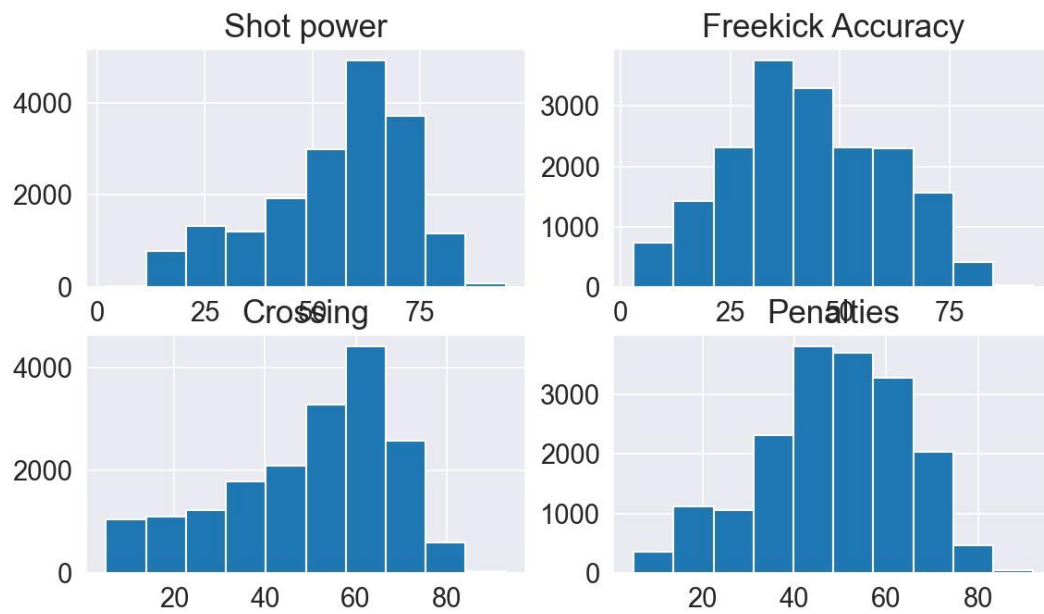
From the given graph we can conclude most players are in 20s peaking at early 20s and thus going downhill towards their late 40s. After 30 most player retire and leaving a very few bare amount in late 30s.

7. Attribution for forwards



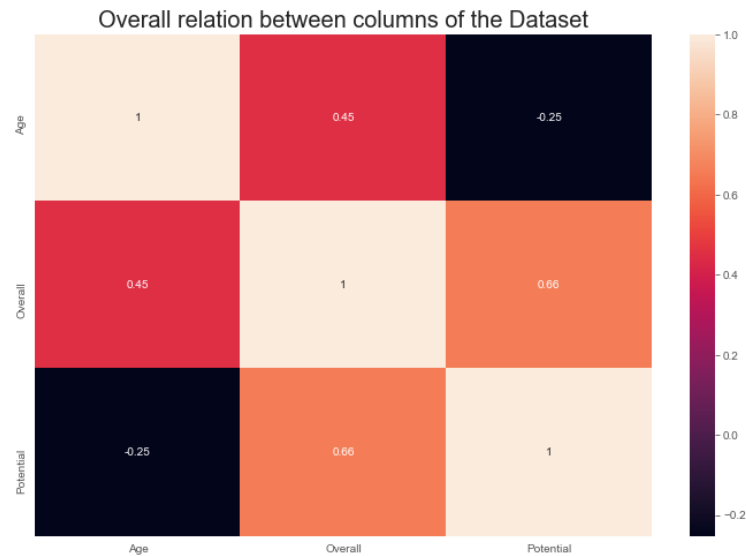
As most of the players are forwards as concluded from about data, these diagrams represent the overall attributes of forward players can change the pace of games.

8. Set piece Attributes



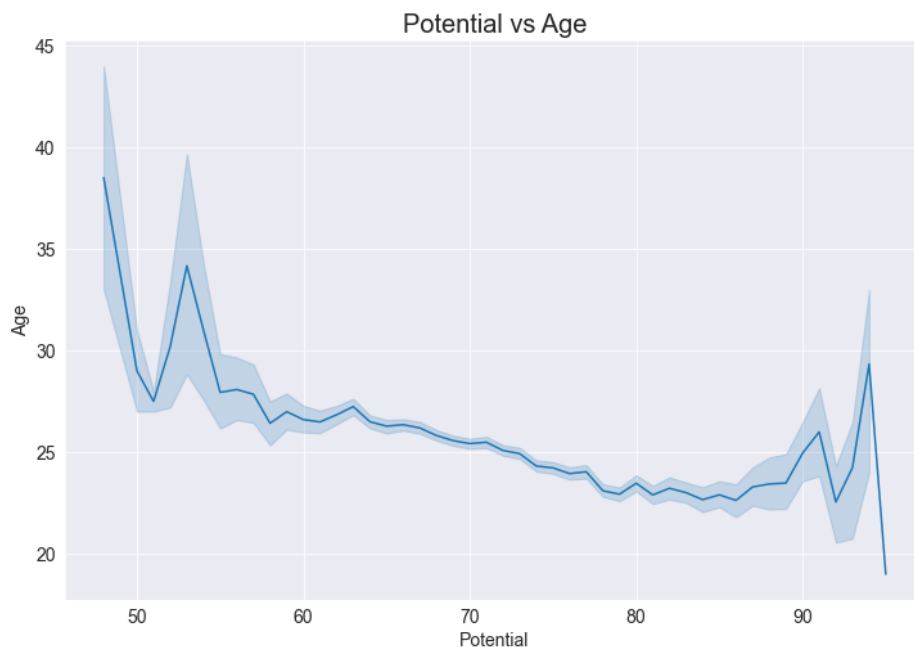
The above diagrams represent the set piece attribute which can be used in multiple set pieces like freekick, penalties and corners. These are the fundamental of football that can determine the end result of the game.

9. Overall relation between Overall, Age and Potential



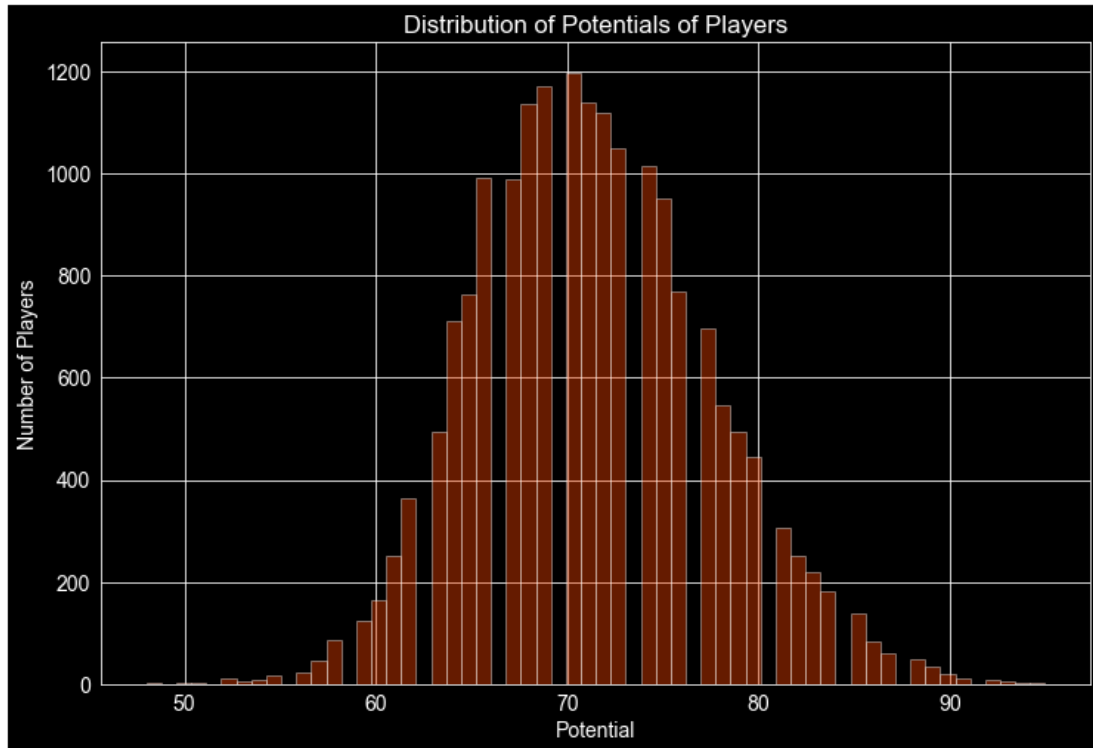
This diagram shows the correlation between age, overall and potential of the players in the dataset. These columns correlate among each other can be seen.

10. Potential vs Age



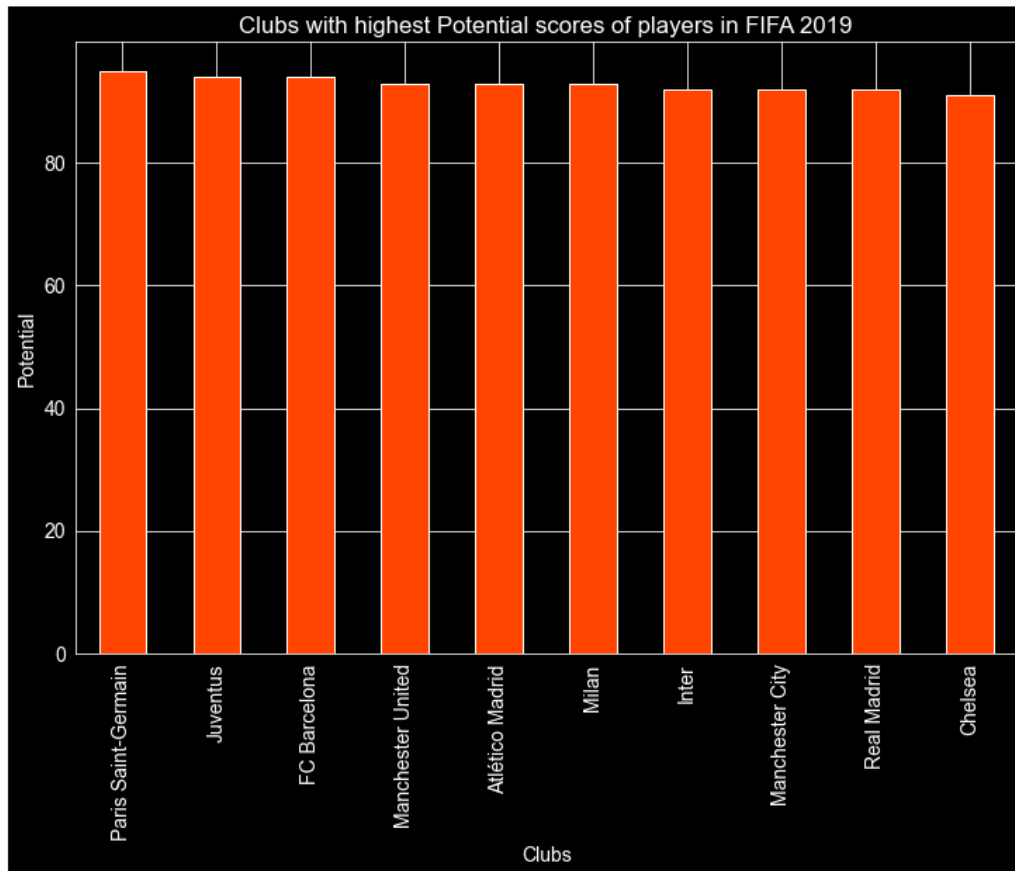
As we can see, the older the player, the less likely a rise in metrics is. Players between the ages of 17 and 22 have more potential, based on their particular stats. We have some severe ups and downs here as well, because some guys are in Club Academies and have the potential to be 90+ ranked players if properly coached.

11. Potential of Players



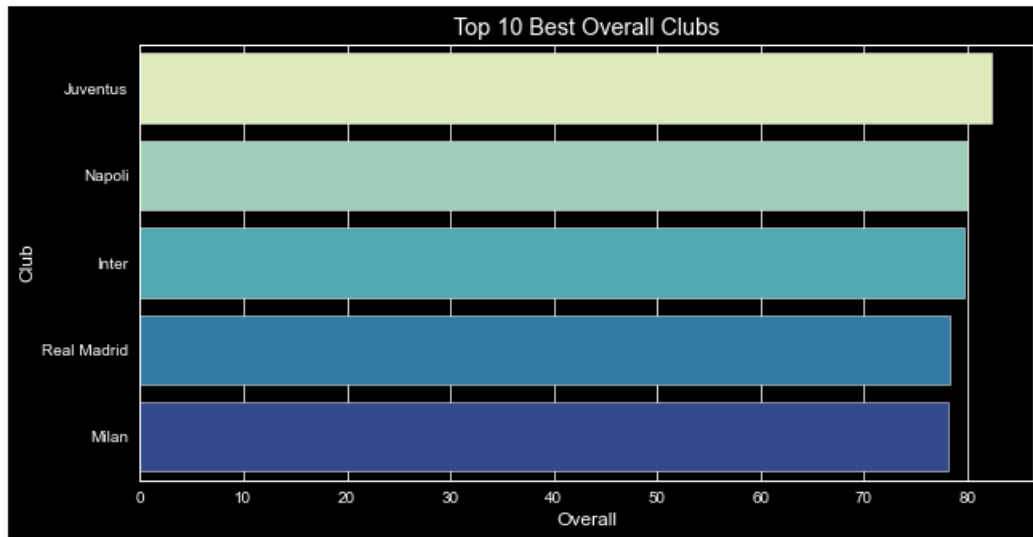
There are just a few players that have the potential for an overall rating of 90 or above. The majority of players have a potential range of 65–80, and the statistics only start going down from there.

12. Top 10 overall potential Teams



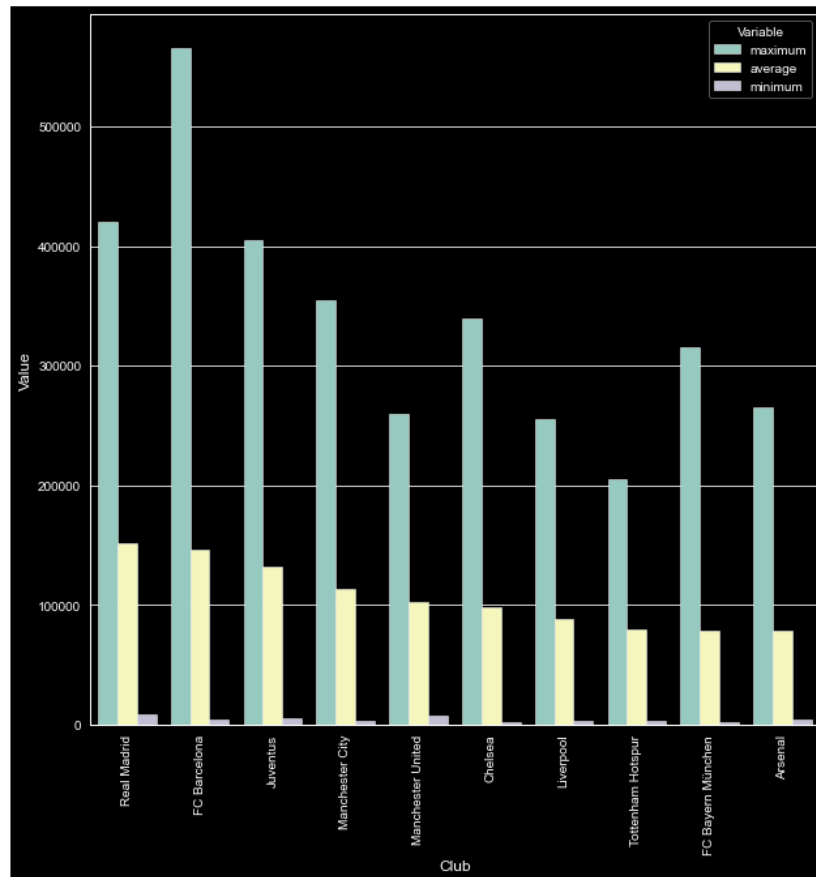
These are the teams that have best potential in which players can raise their potentials in coming future if coached and trained properly. PSG is best team in terms of potential followed by Juventus and Barcelona.

13. Top 5 Overall Clubs



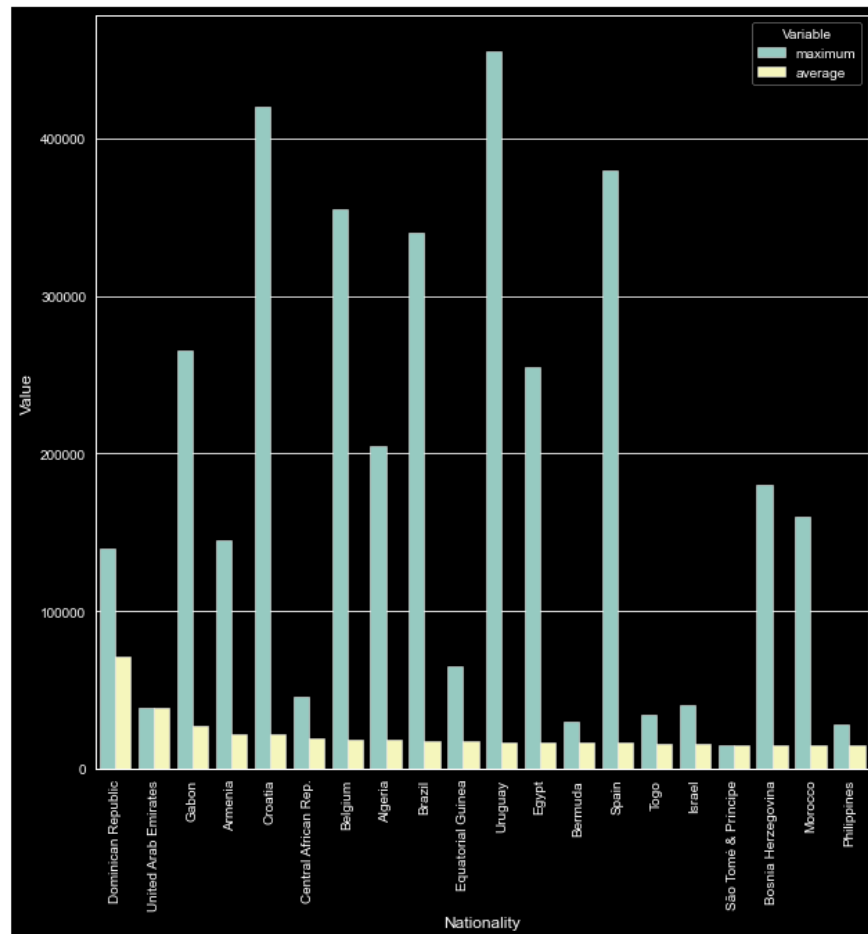
These are the 5 best teams in terms of Fifa 19. They best ratio of overall players. The number team which has can chose is Juventus having overall crossing 80+ which rest of the teams in Fifa 19 don't have. Rest of them are below 80 less Juventus being an Exceptional case.

14. Wages on Players spend by Club



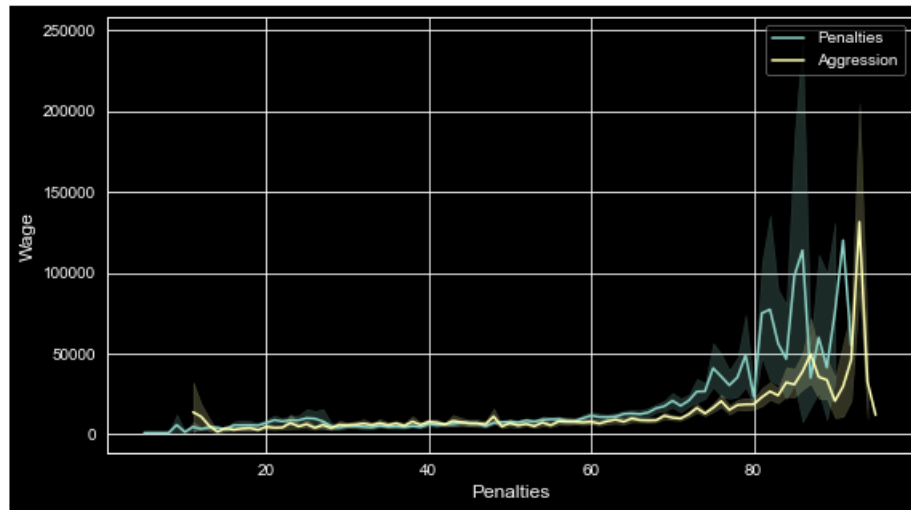
There are total 651 clubs in the dataset. The bar plot represents between the maximum amount of money spent by the clubs on their players, minimum and average amount of money spent on the wages of players by each club. Take note of the salary disparity between the highest and lowest paid athletes. The mean incomes are likewise significantly lower than the maximum earnings, indicating that only a small number of players receive the highest pay. Real Madrid, FC Barcelona, Juventus, Manchester City, and Manchester United are the clubs that spend the most money on their players. Their mean values vary greatly, indicating that a player's club is a significant determining influence in their salary.

15. Wages based on Nation



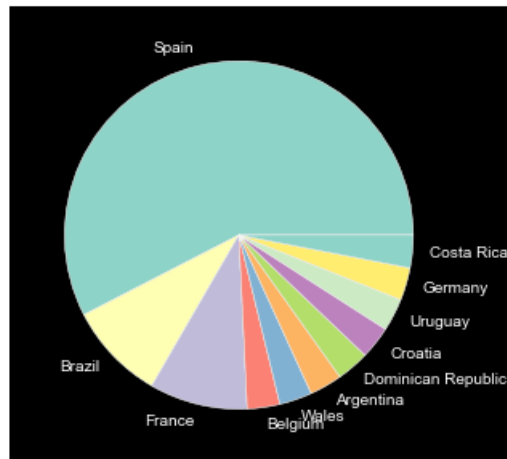
The maximum and mean salaries earned by players from various nations are then shown on a bar chart, which is organized in decreasing order of mean wages. This graphic once again emphasizes the disparity between the highest and mean incomes of the players. This figure also demonstrates how outliers may impact the value of mean salaries. Players from various nations have varying mean pay, showing that a player's nationality may have a role in determining their earnings.

16. Wages in terms of Penalties Aggression



Players' aggression, penalties, and salary are all plotted out. There is no clear association between fines and wages in this graph. Wages appear to be increasing as a result of the fines, but the pattern is not consistent and is constantly shifting. In the same way, the plot for aggression peaks and dips without revealing a clear significant association. As a result, we can deduce that there is little or no relationship between violence and earnings, or penalties and wages.

17. Real Madrid Country representation



This diagram represents the overall player distributed in Real Madrid. Majority comes from Spain followed by Brazil and France.