FIFA PLAYER STATS ANALYSIS

ARSHAD MANAMAKKAVIL

219033472

am1285

am1285@student.le.ac.uk

1 TABLE OF CONTENTS

List of figures

List of Tables:

2 ABSTRACT:

The main determination of the FIFA Player stats analysis project is to deeply evaluate and analyse all the measures of the player stats. This includes performing an exploratory data analysis on the FIFA 22 dataset. The project aims to help the managers to find the best team and also the FIFA game players to get an optimum result. The second part of the project aims to machine learning and predict the overall scores by training the features of the players as input. At last, the project aims to concentrate mainly on making an online manager develop a predictive analysis to find the best team for some selected countries. These will tell us what the best formation and which are the best eleven players we need to concentrate on in these positions. After answering all the selected questions we get prior knowledge of the selected data and players' features and attributes.

3 PROJECT SUMMARY

My aim for this project is to come up with an optimum solution that can help managers to find the best team and players by predicting and analysing different plots and graphs. This is done by creating different types of questions that every manager comes up with or this is also useful to the players who play the FIFA game. After the proposal, my questions are almost the same as before but there is some slight change in the previous questions. As a part of my research after the project proposal my thoughts about the questions changed and I prefer to add 2 more questions along with the existing questions I also merge some of the questions which sound almost similar to me. For the entire project I use jupyter notebook as my best tool and I hope I completed the project the best I could. I know some of the basics of python and how the Jupyter notebook works, but I need to study some of the machine learning techniques like regression models and prediction techniques. I use some tutorials from Udemy to study the train test split data and prediction. For the questions which I selected I need to look on the internet for different types of approaches to the questions. I need to search for every possible tutorial which is relevant to the project. These projects include some of the ML models and team building some of the tutorials help me to achieve these tasks more easily. The tutorials which I mainly look into are from the Kaggle and towardsdatascience.com helps me more to get the relevant approach to the questions.

4 DATA PROCESSING & ANALYTICAL METHODS

I prepared the data for analysis by web scrapping the sofifa.com using beautiful soup which is a library used to scrap the data from the internet. I also find similar data from Kaggle which can be also used for this project. While I scrape the data from sofifa.com it won't be that simple because I need to spend more time on the data scraping which is a new topic to me. I use Jupyter notebook to scrap the data from the internet. While scraping the data the challenge was that I need to scrap every player's data separately as they are given as an <a> tag which is a hyperlink tag used to get multiple pages. As all the data are given as a table of content it's easy to scrap the exact data from the website. After the extraction of data, I saved it as a CSV file. This CSV file is directly read from the location using a jupyter notebook. Here I am only using the FIFA 22 data only. I am using the Jupyter notebook for the analysis purpose of this particular question because scientific computation and computational data analytics can be flexibly used in the jupyter notebook.

In the first stage of this project I simply just analysed how the data are distributed. Then I thoroughly studied the data and think as a football analyser which columns are needed for answering the particular questions and which columns to be deleted for the data cleaning process. First I just read the data CSV file from the location then I save it to a data frame as fifa_df to be the first data frame. The columns which we get from the data are given below.

'ID', 'Name', 'Age', 'Photo', 'Nationality', 'Flag', 'Overall', 'Potential', 'Club', 'Club Logo', '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', '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', 'GKDiving', 'GKHandling', 'GKKicking', 'GKPositioning', 'GKReflexes', 'Best Position', 'Best Overall Rating', 'Release Clause', 'DefensiveAwareness'

In these columns, I just select the columns which I am using for answering the questions and drop the unwanted columns from the dataset. After that, I just printed the remaining dataset which we are going to analyse. This can be done by sing isnull() function to check for null values which results in many columns having some null values.

Table 1 null values

ShortPassing	False
Volleys	True
Dribbling	False
Curve	True
FKAccuracy	False
LongPassing	False
BallControl	False
Acceleration	False
SprintSpeed	False
Agility	True
Reactions	False
Balance	True
ShotPower	False
Jumping	True
Stamina	False
Strength	False
LongShots	False
Aggression	False
Interceptions	True
Positioning	True
Vision	True
Penalties	False
Composure	True
Marking	True
StandingTackle	False
SlidingTackle	True
GKDiving	False

But only the columns which I am going to use for the analysis need to be cleaned from null values. I got some null values on the following columns.

Here the column name says True is needed to be cleaned from null values. As these all are numerical values I just replace it with the mean values. player_df['Marking'].fillna(player_df['Marking'].mean(), inplace = True) this code is used to replace the mean of Marking column instead of null value. The same method is used for all the other column with null values. I am using the Player's attributes which can influence their performance in the game and which give them how they perform well in recent years. There are 16710 rows and 65 columns for the whole data. Here we drop some unwanted columns which we will not be used in future analysis. For analysis of the player statistics first, we need clean the data and review all the remaining columns.

To review the data based on their value and wages we need to look at how the column value and wage are distributed along with the dataset. By analysing these two column can be given as below

Table 2 Value and Wage

Value	Wage
€107.5M	€250K
€93M	€140K
€44.5M	€135K
€125.5M	€350K
€37M	€45K

Here the data are given as Euro currency with the representation of million for 'M' and thousands for 'K', these make the data more difficult for sorting which they appear to be wrong while we try to sort it. So here I am using some function which I defined as to replace the K with multiply by 1000 and M with multiply by 1000000 so now the data is more efficient and we can sort it to get the more valued player. I am using fifa_df as my initial data frame and each time whenever there is something huge change in the data frame I make a copy of the earlier data frame to prevent data loss even after the analysis.

After that I check is there any other columns need to be changed for analysis and I found that there is some problem with the position column which is distributed as the fig below.

```
fifa_df4["Position"]
```

```
82        <span class="pos pos25">ST
251       <span class="pos pos24">RS
39        <span class="pos pos25">ST
64        <span class="pos pos27">LW
3        <span class="pos pos13">RCM
                    ...
16527                        NaN
16540                        NaN
16572                        NaN
16585    <span class="pos pos28">SUB
16709    <span class="pos pos28">SUB
Name: Position, Length: 16710, dtype: object
```

These type of data is very difficult to analyse, thus we need to remove the part which contain '<span class="pos pos25">' so we get a clean data which position attributes.

After the cleaning of the wage, value and position column, we can see the data as below.

```
#After the first stage of cleaning
fifa_df4[['Value', 'Wage', 'Position', 'Weight']]
```

	Value	Wage	Position	Weight
82	194000000.0	230000.0	ST	73.0
251	137500000.0	110000.0	RS	94.0
39	129500000.0	240000.0	ST	89.0
64	129000000.0	270000.0	LW	68.0
3	125500000.0	350000.0	RCM	70.0
...	...	...	...	...
16527	0.0	0.0	ST	82.0
16540	0.0	0.0	ST	70.0
16572	0.0	0.0	ST	81.0
16585	0.0	0.0	SUB	85.0
16709	0.0	0.0	SUB	69.0

16710 rows × 4 columns

After the first stage of cleaning we can see that the data are distributed approachable for the analysis purpose.

5 QUESTIONS

5.1 EXPLORATORY DATA ANALYSIS ON FIFA22:

For this particular question I am analysing some basic player stats data which will give an overall view of the features of some players.

Firstly I am just checking the no of players from each country which is given below.

Table 3:

Nationality	Players
England	1845
Spain	1151
Germany	1120
France	987
Argentina	846
Brazil	819
Italy	514
Netherlands	443
Portugal	354
United States	341
Mexico	312
Republic of Ireland	308
Scotland	292
Japan	284
Belgium	267

This means that 1845 players are playing for England and 1845 players from Spain which is the second country and so on, as we can see from the no of players and the nationality from the table above.

As we know that these are the country where more players are playing but I am also taking the rated country into account which is not on the list. The selected countries for analysing are 'Brazil', 'England', 'Argentina','Germany', 'Spain', 'France', 'Italy', 'Portugal', 'Belgium'.

First I am analysing the overall scores of players from the selected international teams.

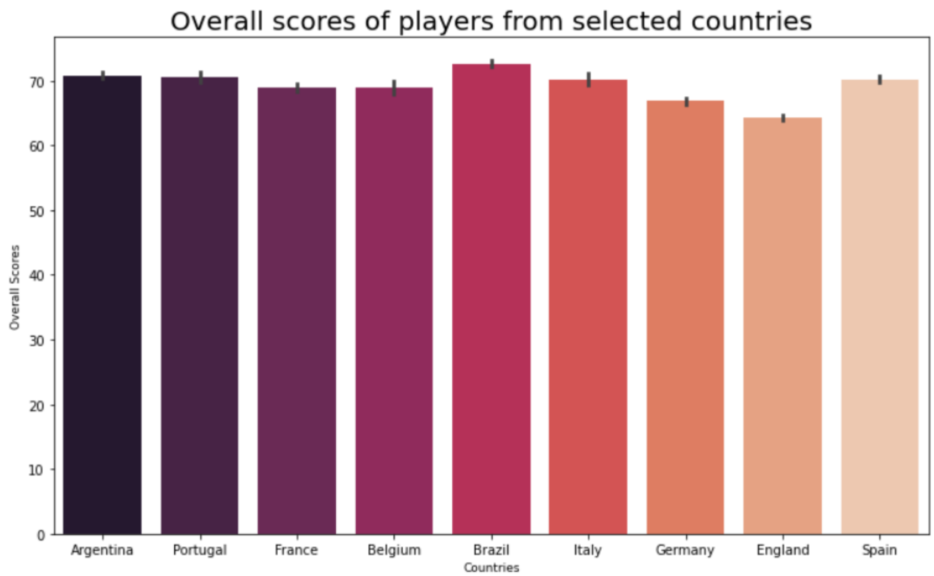


Figure 1 Overall Scores vs Countries

Here we can say that the overall rating of players from these selected countries is almost having an average rating of overall score of about 70. We can say that these teams perform well in every game even though we can't predict who wins when they play against each other because these stats show these teams are equally strong teams among the other teams from the world. After analysing the overall score we are going to analyse the average wage of this same selected country which I mentioned above.

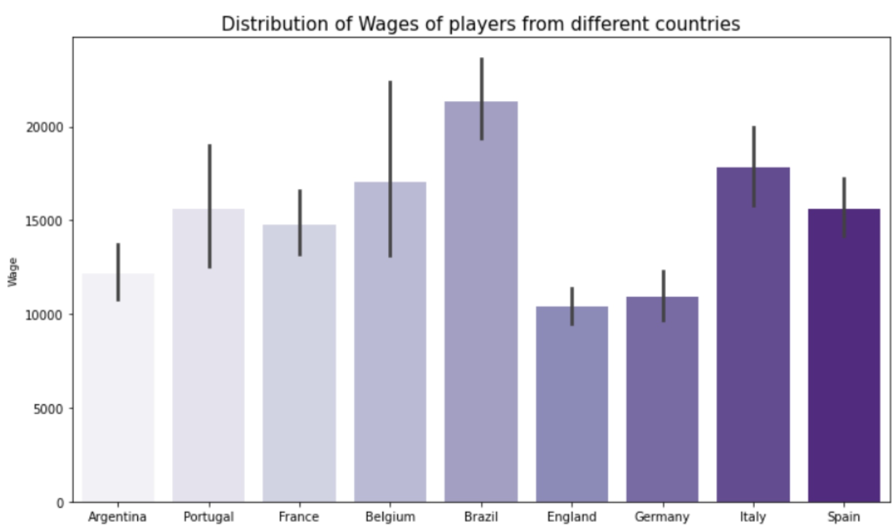


Figure 2: Wages of players from different countries

Here we can see that Brazil is about have an average wage of above 20000euro is the no one country in which more of their players are earning more wages compared to other countries. This rise in the wages of Brazil is due to the players like Neymar Jr, Marquinhos, Ederson, Casemiro, and Vinícius Jr who have been paid more in Brazil international football. The next team after brazil which have an average wage of above 17000euro is Italy these value contribution to Italian football is due to the players like Jorginho, M. Verratti, N. Barella, L. Insigne, C. Immobile, L. Bonucci, G. Donnarum. The third most average wage country is Belgium with an average of above 1500 and it is due to the players like K. De Bruyne, T. Courtois, E. Hazard, R. Lukaku they are the main contribution to the average wage for Belgium international team. The next team's order as per the bar is by the order like Portugal, Spain, France, and Argentina. Here for this, each team for Portugal Cristiano Ronaldo and Bruno Fernandes are the highly paid players and in the case of Argentina, it is Messi who is the player contributing more to this average wage value.

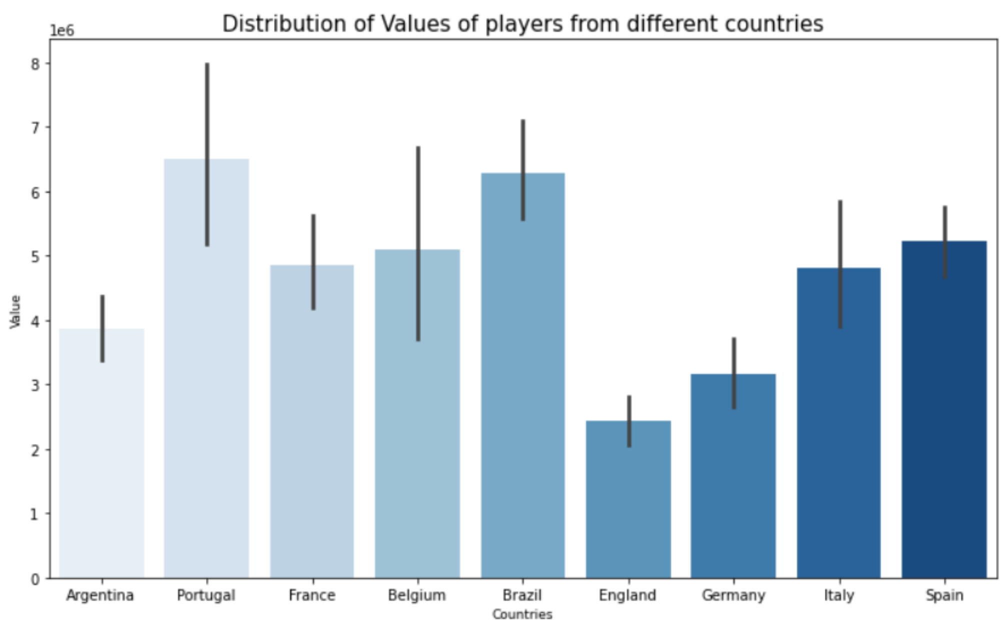


Figure 3 Values of players from different countries

The values and the values are dependent on each other while analysing the bar plot of the value. We can see a quite similar plot for the wages and the values of selected countries. Here the value of Portugal is high compared to the average wage. Here the best average valued team in Portugal and followed by Brazil, Spain, Italy, France, and Argentina

Top Paid Players:

Table 3 Top Paid Players

Name	Wage	Best Overall Rating	Age	Nationality	Potential	International Reputation
K. De Bruyne	350000.0	91.0	30	Belgium	91	4.0
K. Benzema	350000.0	89.0	33	France	89	4.0
L. Messi	320000.0	93.0	34	Argentina	93	5.0
Casemiro	310000.0	89.0	29	Brazil	89	3.0
T. Kroos	310000.0	88.0	31	Germany	88	4.0
R. Sterling	290000.0	88.0	26	England	89	4.0
R. Lewandowski	270000.0	92.0	32	Poland	92	5.0
Cristiano Ronaldo	270000.0	91.0	36	Portugal	91	5.0
Neymar Jr	270000.0	91.0	29	Brazil	91	5.0
M. Salah	270000.0	89.0	29	Egypt	89	4.0

The most valued players:

Table 4: Most valued players

Name	Wage	Value	Best Overall Rating	Age	Nationality	Potential	International Reputation
K. Mbappé	230000.0	194000000.0	92.0	22	France	95	4.0
E. Haaland	110000.0	137500000.0	90.0	20	Norway	93	4.0
H. Kane	240000.0	129500000.0	90.0	27	England	90	4.0
Neymar Jr	270000.0	129000000.0	91.0	29	Brazil	91	5.0
K. De Bruyne	350000.0	125500000.0	91.0	30	Belgium	91	4.0
R. Lewandowski	270000.0	119500000.0	92.0	32	Poland	92	5.0
G. Donnarumma	110000.0	119500000.0	89.0	22	Italy	93	3.0
F. de Jong	210000.0	119500000.0	89.0	24	Netherlands	92	3.0
J. Sancho	150000.0	116500000.0	88.0	21	England	91	3.0
T. Alexander-Arnold	150000.0	114000000.0	87.0	22	England	92	3.0

Histogram of potential scores of players.

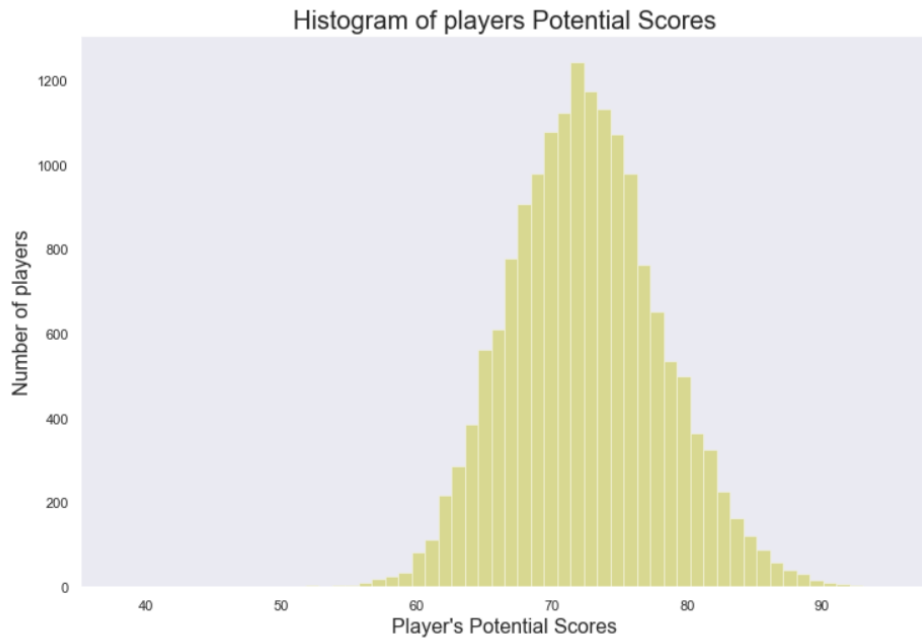


Figure 4: Histogram of player's potential scores

We can see that the 75% of the players having an average of 65 to 85 potential scores.

Top 10 best players with overall scores.

Table 5 Best players with overall scores

Name	Overall	Potential	Club	Preferred Foot	Position	Nationality
L. Messi	93	93	Paris Saint-Germain	Left	RW	Argentina
R. Lewandowski	92	92	FC Bayern München	Right	ST	Poland
Cristiano Ronaldo	91	91	Manchester United	Right	ST	Portugal
J. Oblak	91	93	Atlético de Madrid	Right	GK	Slovenia
K. Mbappé	91	95	Paris Saint-Germain	Right	ST	France
K. De Bruyne	91	91	Manchester City	Right	CM	Belgium
Neymar Jr	91	91	Paris Saint-Germain	Right	LW	Brazil
H. Kane	90	90	Tottenham Hotspur	Right	ST	England
M. ter Stegen	90	92	FC Barcelona	Right	GK	Germany
M. Neuer	90	90	FC Bayern München	Right	GK	Germany

Top 10 players with best potential score:

Table 6 Best players with potential

Name	Overall	Potential	Club	Preferred Foot	Position
K. Mbappé	91	95	Paris Saint-Germain	Right	ST
L. Messi	93	93	Paris Saint-Germain	Left	RW
E. Haaland	88	93	Borussia Dortmund	Left	ST
G. Donnarumma	89	93	Paris Saint-Germain	Right	GK
J. Oblak	91	93	Atlético de Madrid	Right	GK
P. Foden	84	92	Manchester City	Left	SUB
K. Havertz	84	92	Chelsea	Left	LW
R. Lewandowski	92	92	FC Bayern München	Right	ST
F. de Jong	87	92	FC Barcelona	Right	CM
T. Alexander-Arnold	87	92	Liverpool	Right	RB

Now we are analysing the no of players with right footed and left footed. We get the details from the given of each categories from the data below.

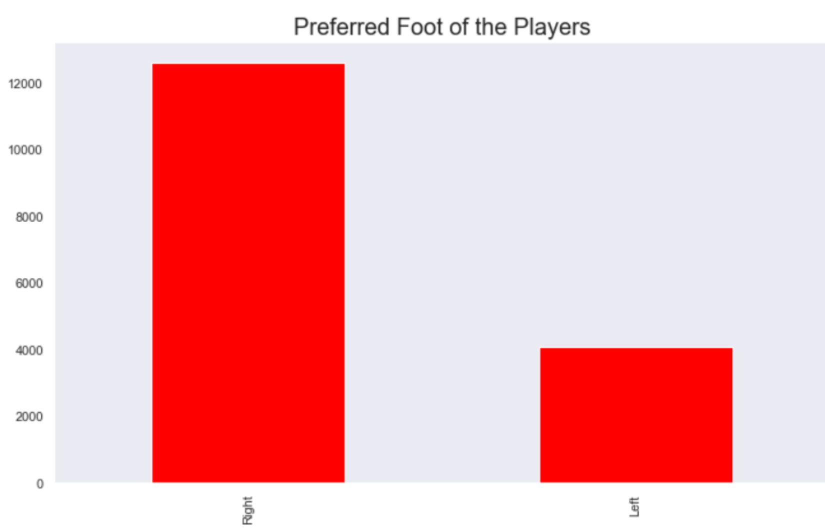


Figure 5 Bar graph of preferred foot of players

Here we can see that there are above 12500 players whose preferred foot is right and the remaining players of about 4000 are left footed. After that, I am some of the best right footed and best left footed footballers based on the dataset we are provided. I am just using the basic sorting operation to get the top players. I am specifically taking the top 10 either left or right-footed players. Let's see if some of the known best left and right-footed players are available in this dataset.

Top 10 Left Footed Players.

Table 7 Top 10 Left footed players

Name	Age	Club	Nationality
L. Messi	34	Paris Saint-Germain	Argentina
T. Courtois	29	Real Madrid CF	Belgium
Ederson	27	Manchester City	Brazil
M. Salah	29	Liverpool	Egypt
R. Lukaku	28	Chelsea	Belgium
E. Haaland	20	Borussia Dortmund	Norway
P. Dybala	27	Juventus	Argentina
A. Robertson	27	Liverpool	Scotland
H. Lloris	34	Tottenham Hotspur	France
A. Di María	33	Paris Saint-Germain	Argentina

Top 10 Right Footed Players:

Table 8: Top 10 right footed players

Name	Age	Club	Nationality
R. Lewandowski	32	FC Bayern München	Poland
Cristiano Ronaldo	36	Manchester United	Portugal
J. Oblak	28	Atlético de Madrid	Slovenia
K. Mbappé	22	Paris Saint-Germain	France
K. De Bruyne	30	Manchester City	Belgium
Neymar Jr	29	Paris Saint-Germain	Brazil
H. Kane	27	Tottenham Hotspur	England
M. ter Stegen	29	FC Barcelona	Germany
M. Neuer	35	FC Bayern München	Germany
N. Kanté	30	Chelsea	France

Top 10 speed players:

Table 9: Top Speed players

Name	Acceleration	Best Position	Age	Nationality	Club	SprintSpeed
K. Mbappé	97.0	ST	22	France	Paris Saint-Germain	97.0
Adama Traoré	97.0	RM	25	Spain	Wolverhampton Wanderers	96.0
A. Davies	96.0	LB	20	Canada	FC Bayern München	96.0
M. Diaby	96.0	LM	21	France	Bayer 04 Leverkusen	92.0
D. James	96.0	RM	23	Wales	Leeds United	95.0
R. Sterling	95.0	LW	26	England	Manchester City	88.0
A. Hakimi	95.0	RB	22	Morocco	Paris Saint-Germain	95.0
Vinícius Jr.	95.0	RM	20	Brazil	Real Madrid CF	95.0
U. Antuna	95.0	RM	23	Mexico	Club Deportivo Guadalajara	90.0
C. Ejuke	95.0	RW	23	Nigeria	PFC CSKA Moscow	93.0

Best defensive players:

Table 10: Top defensive players:

Name	DefensiveAwareness	Best Position	Age	Nationality	Club
G. Chiellini	93.0	CB	36	Italy	Juventus
V. van Dijk	92.0	CB	29	Netherlands	Liverpool
N. Kanté	90.0	CDM	30	France	Chelsea
Rúben Dias	90.0	CB	24	Portugal	Manchester City
K. Koulibaly	90.0	CB	30	Senegal	Napoli
M. Hummels	90.0	CB	32	Germany	Borussia Dortmund
M. Škriniar	90.0	CB	26	Slovakia	Inter
S. Savić	90.0	CB	30	Montenegro	Atlético de Madrid
20 D. De Rossi	90.0	CB	35	Italy	Boca Juniors
Marquinhos	89.0	CB	27	Brazil	Paris Saint-Germain

Overall scores of players based on their Nationality:

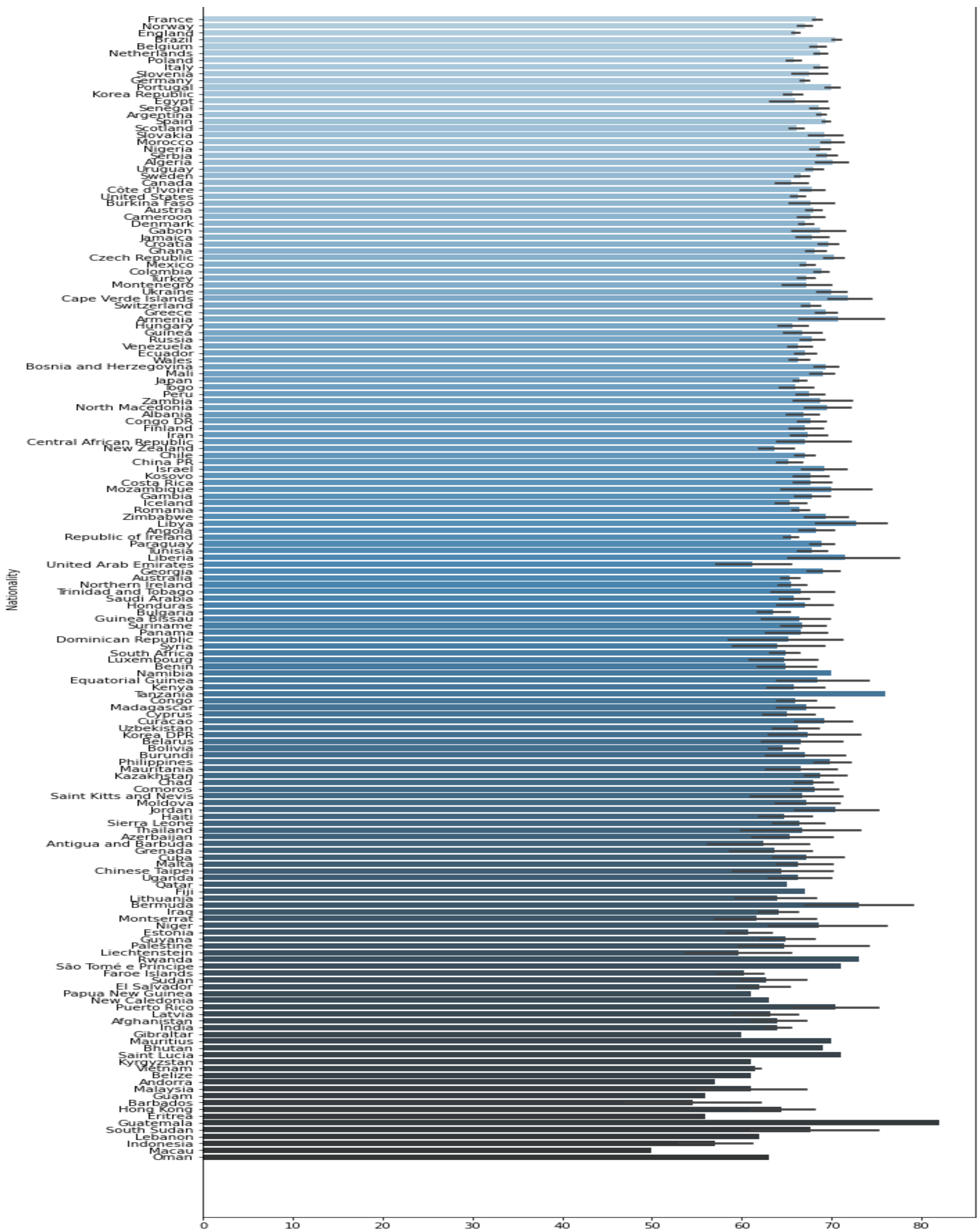


Figure 6 Overall score of players with nationality

5.2 ANALYSE THE FEATURES OF THE PLAYERS AND PREDICTING THE OVERALL RATING BY USING THESE FEATURES.

Here I am using supervised learning and making a Machine Learning Model. Here I am taking the features of a player that are 'Overall','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', 'GKDiving', 'GKHandling', 'GKKicking', 'GKPositioning', 'GKReflexes'. Here this is multiple regression because the input is more for this model. These features give the overall rating of a player. Here we are taking the Overall as the output and the input is taken as the other features. Here we are trying to do some analysis of the input data and find out how the input is related to the overall rating.

First I just check for the null values for the columns I selected using isnull().any() function. Shows the columns with null values are Volleys, Dribbling, Curve, Agility, Balance, Jumping, Interceptions, Positioning, Vision, Composure, Marking, and SlidingTackle. I just replaced these columns with mean values using the code player_df['Marking'].fillna(player_df['Marking'].mean(), inplace = True) in case of Marking and this applies in every columns with null value. After replacing the null I again check for the null values to make sure there is no null value on the columns above.

5.2.1 Splitting the data

For the prediction we must split the data into training and testing sets here I am using train_test_split() which is very reliable for splitting the data into train and test. Here I am taking 20% of data as test and 80% of data as training data. After splitting the data we can see that from the total data of 16710 we get the spitted data as

```
Length of training data: 13368
Length of testing data: 3342
Length of total data: 16710
```

After splitting the data I just look for the histogram of the features of the players.

Histogram for the distribution of player features.

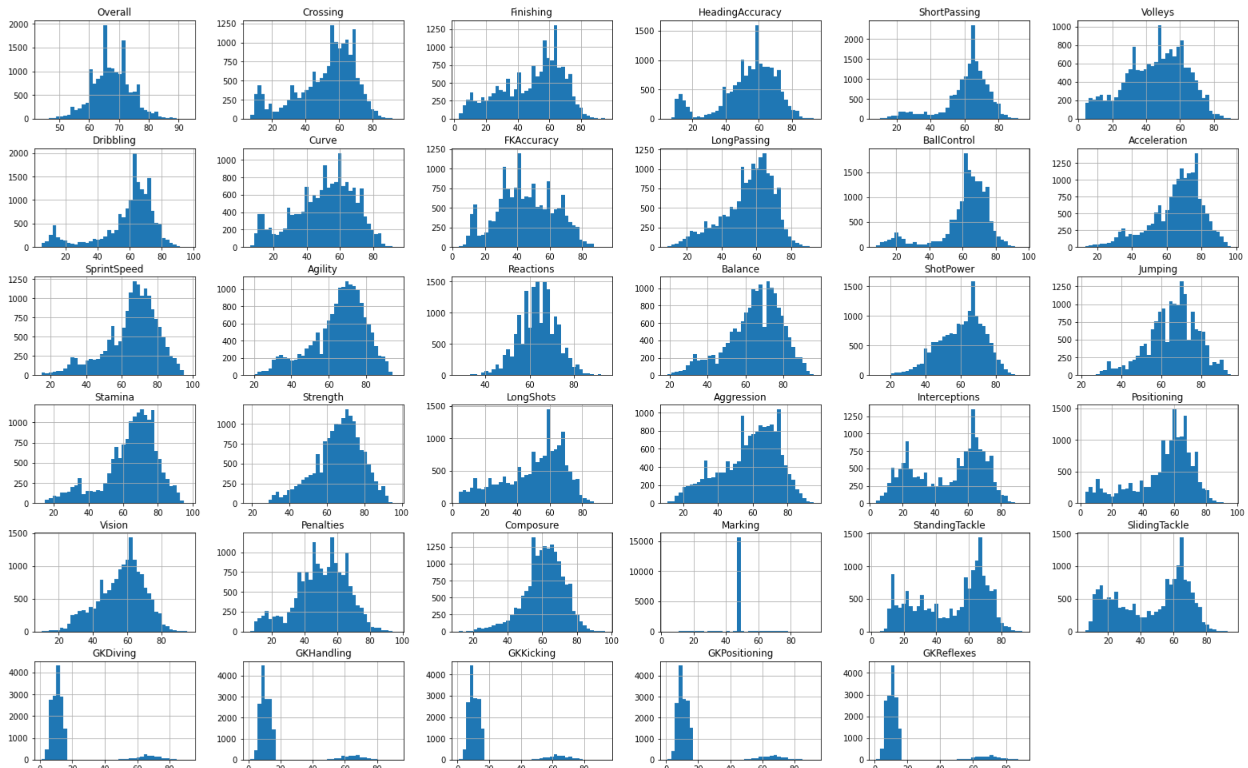


Figure 7: Histogram for features of players

Here we can see that the histogram the values are range from 0 to 100 which means that the scale need not to be done for this data and we can use the data as it is.

5.2.2 Correlation for each attributes:

Using the corr() function I am checking the correlation between the attributes and I just printed the values on the descending order in order to find how the data are ranges from. So we can see the output as

```
Overall            1.000000
Reactions          0.870103
Composure          0.678221
ShotPower          0.521107
Vision             0.502489
ShortPassing       0.494055
LongPassing        0.485379
BallControl        0.422879
Curve              0.377957
Aggression         0.371009
LongShots          0.369671
Crossing           0.366815
```

```
FKAccuracy          0.345707
Volleys             0.342844
Dribbling           0.337564
Positioning         0.325264
Strength            0.322492
Stamina             0.318773
Penalties           0.309715
HeadingAccuracy     0.309546
Interceptions       0.297158
Finishing           0.292795
StandingTackle      0.240179
Jumping             0.237655
SlidingTackle       0.213662
Agility             0.192575
SprintSpeed         0.127681
Acceleration        0.117614
Balance             0.086886
Marking             0.034770
GKPositioning       0.023440
GKKicking           0.022096
GKReflexes          0.020277
GKHandling          0.018943
GKDiving            0.016823
```

We can see that the ranges are from 0 to 1 which says there is a strong correlation when it's close to one. For the further details of correlation we can take some of the best correlated features and draw a scatterplot. Here I am taking only the 5 attributes which have strong relation with overall rating:

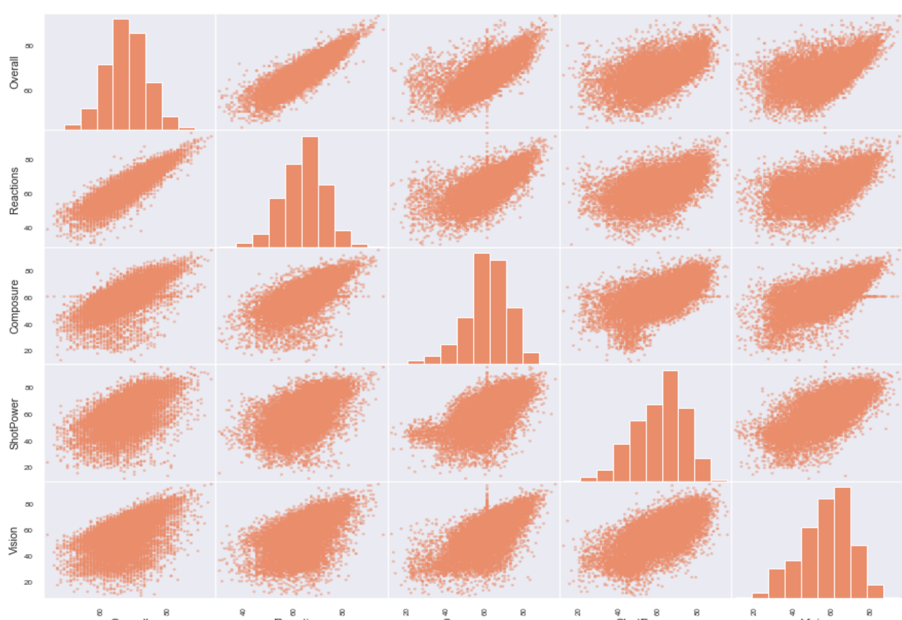


Figure 8: Scatter plot for selected features

Here we can see that all are strongly correlated and the graph itself it shows the reaction is more strongly correlated as this becomes linear when there is a strong relationship. So for that, I am just selecting the reaction scatterplot only to check the relationship with it.

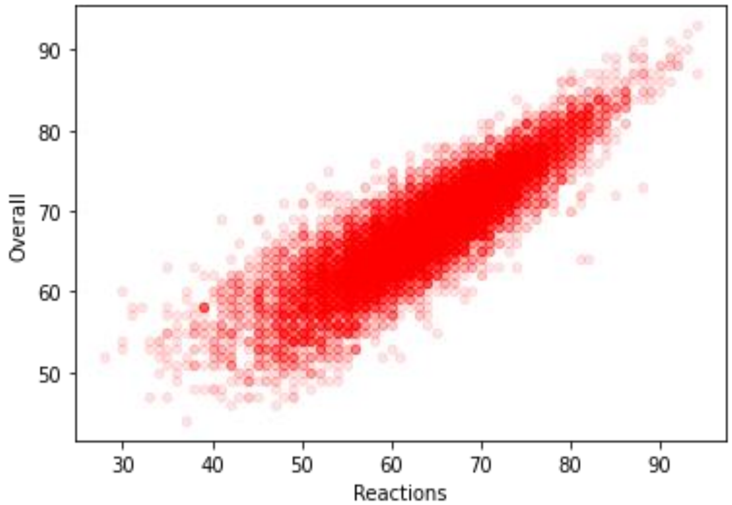


Figure 9: Scatter plot (Overall vs Reactions)

We can see a strong linear relationship for reaction vs overall.

5.2.3 Creating the M L model:

After the splitting of data into training and testing set the target and feature values. Here the overall is selected as the y_train and y_test and the other correlated attributes is taken as the X_train and X_test.
Later we use the Linear regression model on the both training set of the data we get an RMSE value using the Scikit-Learn feature of mean_square_error() which we get the value of 2.318773013162364 as RMSE value. This score is good but as the histogram we can see that there are features with non linear relationship are available so we need a strong regressor method.
Here I just check the different regressor model and find the RMSE value of each the result I got from the data are:

Decision_Tree_Regressor = 0

Random_Forest_Regressor = 0.4763456638850558

The decision tree regressor gives 0 this may be the overfitting of the data and the Random forest gives the value of 0.4763 which is a good model. So for the later methods I am using random forest as my model. For further precision of selecting the model I am just looking for cross validation points of 10 sample values using fold. Here the scores for the models are given below

5.2.4 Cross validation for selected models:

Linear Regression:

```
Scores: [2.29031172 2.37753856 2.33152832 2.30787788 2.2522044  2.31014363
 2.40153034 2.36749112 2.35337611 2.27066769]
Mean: 2.3262669789813994
Standard Deviation: 0.0460682244809026
```

Decision Tree:

```
Scores: [2.02840709 2.11805643 2.08674988 2.04803936 2.01953807 2.07128027
 2.03043411 2.1574175  2.18613623 2.01639983]
Mean: 2.076245876362621
Standard Deviation: 0.05703335502852432
```

Random Forest:

```
Scores: [1.2925973  1.27542106 1.2949312  1.18745046 1.25929558 1.23775438
 1.22093123 1.32138979 1.43111172 1.26880251]
Mean: 1.2789685233518586
Standard Deviation: 0.06271611739648023
```

Here the score obtained for the random forest is the best so in this examine we can conclude that the random forest model is the best model for this project.

5.2.5 Applying on testing set and predict the model:

After selecting the model I apply the reggressor model on the testing set and we obtain the final_rmse values as 1.2427503715294415 this score is good as we know. For the model to get the range of the score I am using the 95 % confidence interval for the calculation. The value obtained from that is given by **array([1.10405271, 1.36745186])** which stats that the rmse value lies between 1.4 and 1.36 which is a good score.

After the model implementation and finding the RMSE value we are almost on the final stage. The prediction is done in this stage of operation it is done by selecting selecting 5 points from the test data using predict() here I am using final_model.predict() to predict the data we get the prediction as the follows.

```
Predictions: [62.03333333 66.03333333 65.06666667 57.23333333 65.76666667]
Labels: [63, 66, 66, 58, 66]
```

These are the prediction value of the 5 data we selected. This tells that this model is far better model and the prediction values are almost closest to the labelled values.

5.3 Analysing the different positions and finding out the best and most skilful players for these positions.

5.3.1 Analysis of different positions in football:

As I am already mentioned there almost 29 positions in this data which is available from the position column of the fifa22 data. Which is given by,

['ST', 'RS', 'LW', 'RCM', 'GK', 'LM', 'RB', 'RDM', 'CAM', 'SUB', 'RCB', 'RW', 'LDM', 'RM', 'CDM', 'LCB', 'LB', 'CM', 'LS', 'LCM', 'CF', 'CB', 'RES', 'RWB', 'RAM', 'LWB', 'RF', 'LAM', 'LF']

(fifauteam, 2021)

Figure 10: differnt positions in football

These are the position distribution among the whole FIFA game. But as a fifa game analyser these positioned not be taken into account there are only 17 selected attributes we need to select from the position from the data. Here we select convert the unwanted position into

RCM, LCM = CM

RS, LS = ST

RDM, LDM = CDM

#RAM, LAM = AM

LF, RF = CF

And we don’t take the ‘SUB’ and ‘RES’ positions also. Now all the position data is completely cleaned and converted into position which we required. After the process we get the available positions as below.

GK GOALKEEPER: Only player allowed to touch the ball who defence most in the defensive position of football. Goalkeeper role is to defence the net without allowing the opponent team to score.

RB, LB (Right Back, Left Back): They positioned on the both side of the centre back and their role to defend the opposition teams wingers from getting to our half.

RWB LWB (Right-Wing-Back, Left-Wing-Back): They need to be more alert to the back and sometimes needs to create chance by making counter attack movements.

CB CENTRE-BACK: Strong position in football which needed more physical character and to provide the centre forward from scoring from the penalty box. They need to be more successful tackles players.

CDM DEFENSIVE MIDFIELDER: They are the key to getting ball from the defence and getting them to the centre mid or even for the long pass to the attacking half thus creating more chances.

CM CENTRE MIDFIELDER: The result of every game depends on the players and how they play on this position. They are the backbone of the game with defence and attacking. Their role is to make more goal scoring chances throughout the game.

CAM CENTRAL ATTACKING MIDFIELDER: Their more ability is to have a vision and passing skill inorder to confuse the opposite defence. And making through passes to be given to the strikers to make more goal scoring opportunity.

RM LM: Right/Left MIDFIELDER: Both this position player are located on the both ends of the side line preferred right foot on the right midfielder and preferred left foot on the left side in order to have high crossing capability.

RW LW (Right/Left-WINGER): Their attributes mainly is speed so they are positioned at the extreme end of the side line which they need to accelerate and make the opposition headache:

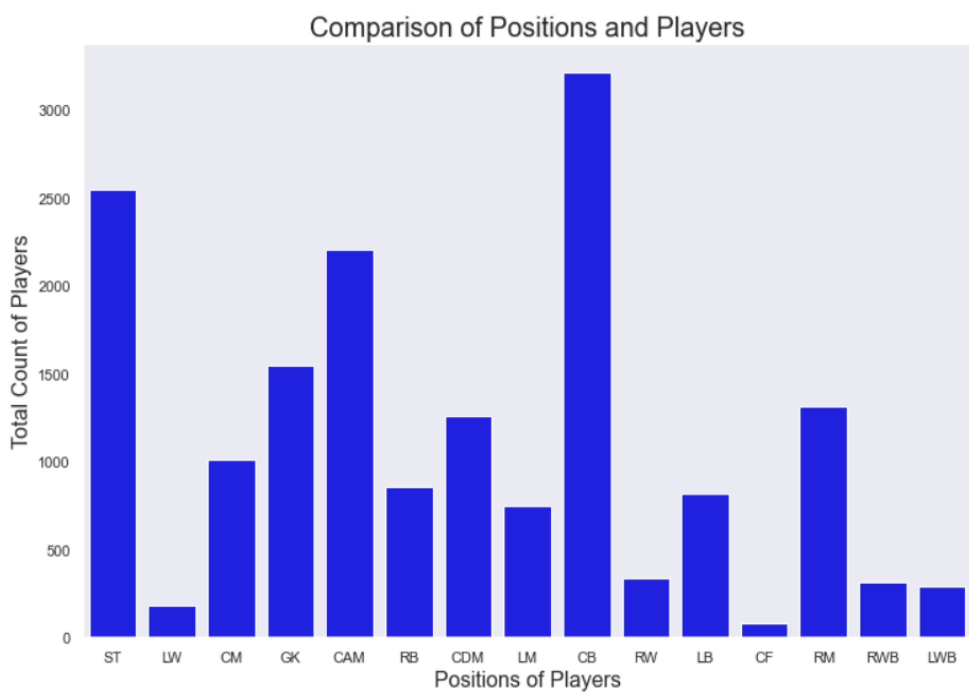
RF LF (Right/Left FORWARD): Right and Left Forwards more like play with wings but they are more reliable on the forward positions whey won't need more skills in crossing.

CF CENTRE FORWARD: They play between the midfield and attack positions. They score goals through the passes and they need creative ability to score goals.

ST STRIKER: They are closest player to the opposing team's goal keeper. They need to have more capability of scoring goals and also has to be capable of winning high ball from the teammates.

5.3.2 Analysing how many player are playing different positions:

I analysed the data with position of the players, here the data are distributed by the bar graph is given below.



From the 17 positions I explained above, here from the data there is only 15 players with available best position they play. There is no player is playing on the position of RF and LF.

We know from the graph that CB and ST are the position where more of the players are players are playing there are almost above 3200 CB and above 2500 above for the position ST. Left Wing and the Centre Forward are the lowest among the position they play.

Later from the information we get I am going to analyse the data by reading the different features of the position. Here I am defining a function which will create a top features for the particular positions we get the output as follows.

```
Best Position CAM: Agility, Balance, Acceleration
Best Position CB: Jumping, Aggression, Marking
Best Position CDM: Aggression, BallControl, Interceptions
Best Position CF: BallControl, Dribbling, Agility
Best Position CM: BallControl, LongPassing, Balance
Best Position GK: GKReflexes, GKDiving, GKPositioning
Best Position LB: Acceleration, Agility, Balance
Best Position LM: Acceleration, Agility, Balance
Best Position LW: Agility, Acceleration, Balance
Best Position LWB: Acceleration, Balance, Agility
Best Position RB: Acceleration, Agility, Balance
Best Position RM: Acceleration, Agility, Balance
Best Position RW: Acceleration, Agility, Balance
Best Position RWB: Acceleration, Agility, Balance
Best Position ST: Jumping, Acceleration, Finishing
```

These are the output which are the top features of each positions. After analysing this data we use the spider plot to plot how all these features are distributed amount the data. Here I am splitting 15 spider plot into 5 different categories.

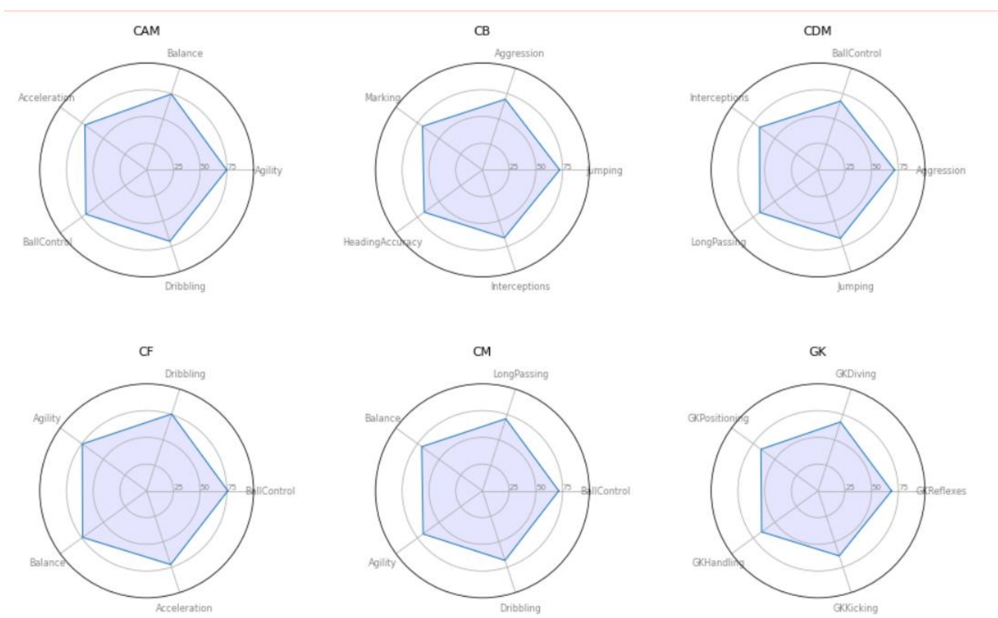


Figure 11: Spider plot for different positions

Here from these plots, we can see how the distribution of each position is classified according to its features. For the Centre Attacking Midfielder, we can see that the top features are Balance, Acceleration, Ball control, dribbling, and Agility. Here the spider plot of every position is neat and almost 75 % because here the features we are taken are the top features. In the case of Goal Keeping, we can see that the top features are GKDiving, GKPositioning, GK Handling, GK Kicking, and GK Reflexes. So we can say the plot is accurate because the plot values are changing according to the change of each position. As a football viewer, we know that the features for the other positions are entirely different from the features of the Goal Keeper. Thus this is very clear in these plots as the positions changing the features also change accordingly. The other remaining position features are given below. These can be analysed as same as the above positions.

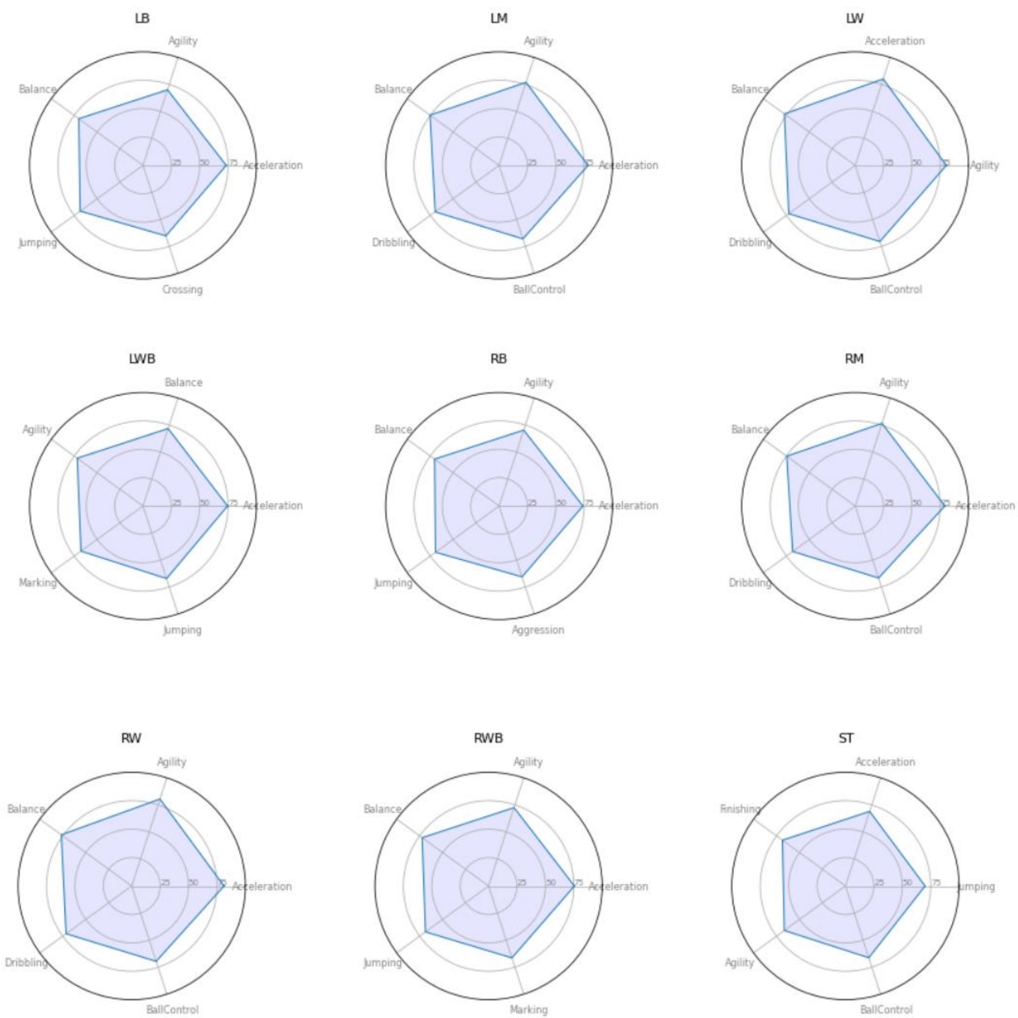


Figure 12: Spider plot for different positions

This spider plot explains the best feature of players playing in different positions. We can analyse any position by using these spider plots and we get a conclusion regarding what are the features required to concentrate for a player to achieve best in their positions.

Best players for each position with their Nationality, age and the club they are playing based on their overall scores is given below.

Table 11: Best players at each position based on overall scores

Position	Name	Age	Club	Nationality	Value	Overall
CAM	Bruno Fernandes	26	Manchester United	Portugal	107500000	88
CB	V. van Dijk	29	Liverpool	Netherlands	86000000	89
CDM	N. KantÃ©	30	Chelsea	France	100000000	90
CF	K. Benzema	33	Real Madrid CF	France	66000000	89
CM	K. De Bruyne	30	Manchester City	Belgium	125500000	91
GK	J. Oblak	28	AtlÃ©tico de Madrid	Slovenia	112000000	91
LB	A. Robertson	27	Liverpool	Scotland	83500000	87
LM	H. Son	28	Tottenham Hotspur	Korea Republic	104000000	89
LW	Neymar Jr	29	Paris Saint-Germain	Brazil	129000000	91
LWB	19 Filipe LuÃs	32	AtlÃ©tico de Madrid	Brazil	21500000	85
RB	T. Alexander-Arnold	22	Liverpool	England	114000000	87
RM	S. Gnabry	25	FC Bayern MÃ¼nchen	Germany	64500000	85
RW	L. Messi	34	Paris Saint-Germain	Argentina	78000000	93
RWB	17 P. Lahm	32	FC Bayern MÃ¼nchen	Germany	29500000	88
ST	R. Lewandowski	32	FC Bayern MÃ¼nchen	Poland	119500000	92

We can create a list of positions with best players by analyzing the overall scores.

Best players for each position with their Nationality, age and the club they are playing based on their Potential scores is given below:

Table 12 : Best position for each player based on Potential

Position	Name	Age	Club	Nationality	Potential
CAM	P. Foden	21	Manchester City	England	92
CB	RÃºben Dias	24	Manchester City	Portugal	91
CDM	N. KantÃ©	30	Chelsea	France	90
CF	K. Benzema	33	Real Madrid CF	France	89
CM	F. de Jong	24	FC Barcelona	Netherlands	92
GK	J. Oblak	28	AtlÃ©tico de Madrid	Slovenia	93
LB	T. HernÃ¡ndez	23	AC Milan	France	90
LM	H. Son	28	Tottenham Hotspur	Korea Republic	89
LW	Neymar Jr	29	Paris Saint-Germain	Brazil	91
LWB	Nuno Mendes	19	Paris Saint-Germain	Portugal	88
RB	T. Alexander-Arı	22	Liverpool	England	92
RM	VinÃcius Jr.	20	Real Madrid CF	Brazil	90
RW	L. Messi	34	Paris Saint-Germain	Argentina	93
RWB	17 P. Lahm	32	FC Bayern MÃ¼ncher	Germany	88
ST	K. MbappÃ©	22	Paris Saint-Germain	France	95

Also we can come up with a list of best players playing different position according to their potential scores. This can be mention as the best players for the best 15 positions.

5.4 CREATING THE BEST TEAM AND TEAM FORMATION FOR SOME INTERNATIONAL TEAMS FOR THE UPCOMING FIFA WORLD CUP 2022

This question is the most interesting question which I came across while doing the project. For this purpose, I am going to take the data which I have been using since the start. These all are cleaned data so we don't need to clean it again. For making the best formation and view the best team we don't need the whole data so we select some columns which we need these columns are 'Name', 'Age', 'Nationality', 'Overall', 'Potential', 'Club', 'Value', 'Wage', 'Position'. These are the columns we have taken.

These tasks are divided into two parts, the first part is to find the best formation and the second part is to find the best team for the best formation.

For finding the best formation I just look into the different formations which more of the team are playing nowadays. The formation which I am taken here is '3-4-3', '4-4-2', '4-3-1-2', '4-3-3', '4-2-3-1'. These are the general formation we have. From this we need to know how this formation is taken into account for that we need to know different positions in football which I already explained in the question where I explained about all the positions and how they work. For every formation mentioned above I create a different function for each formation as follows

343 = ['GK', 'CB', 'CB', 'CB', 'RB|RWB', 'CM|CDM', 'CM|CDM', 'LB|LWB', 'RM|RW', 'ST|CF', 'LM|LW']

442= ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'RM', 'CM|CDM', 'CM|CAM', 'LM', 'ST|CF', 'ST|CF']

4312= ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'CM|CDM', 'CM|CAM|CDM', 'CM|CAM|CDM', 'CAM|CF', 'ST|CF', 'ST|CF']

433 = ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'CM|CDM', 'CM|CAM|CDM', 'CM|CAM|CDM', 'RM|RW', 'ST|CF', 'LM|LW']

4231 = ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'CM|CDM', 'CM|CDM', 'RM|RW', 'CAM', 'LM|LW', 'ST|CF']

Here for every formation, we take the position according to the position of Goal keeper, defenders, midfielders, and attackers. Here we make a list of 11 positions for each formation in which they appear. These can be done by a football viewer for analyzing how this works. For example, take a case for the 342 formations here position of goal keeper always constant and we take three Centre Back and we take 3 as 'RB|RWB', 'CM|CDM', 'CM|CDM', 'LB|LWB' which makes 34 and lastly for the attacking side we take 3 as

'RM|RW', 'ST|CF', 'LM|LW'. So finally we get the formation of 343. This same method is applied to every formation and we get all the formations separately as a list.

After that we create another list which says squad_list as [squad_343_strict, squad_442_strict, squad_4312_strict, squad_433_strict, squad_4231_strict] and also we have a squad_name as ['3-4-3', '4-4-2', '4-3-1-2', '4-3-3', '4-2-3-1'].

As a part of finding the best formation, we create a function to find the best formation in which the function name is get_team_formation. This function reads the data frame and uses appropriate input data to retrieve the best formation for the team which we select. Here we use the overall rating of every player in the international team we select. So we get the best formation after the execution of the function.

The second part is to find the best 11 for the particular formation we get. So we create another function that says best_squad_2022 which takes the value of the overall rating and selects each max value of the column we select. Which is 'Name', 'Age', 'Position', 'Club', 'Value', and'Wage' is taken into the output. After that, if we give the input variable country and call the function we get the best formation. After we get the best team for the particular formation by related function.

We cannot take every team to find the best formation and squat as it is a very time-consuming process. So we select some teams which are qualified for the FIFA world cup 2022. Here I am taking France, England, Brazil, Argentina, Spain, Portugal, Belgium and Netherlands.

5.4.1 France:

They are the previous champions of the FIFA world cup so I decided to select them. After calling the function with nationality = France we get the best formation for France as:

Nationality	Squad	Overall
France	3-4-3	84.64
France	4-4-2	85.09
France	4-3-1-2	85.64
France	4-3-3	84.82
France	4-2-3-1	84.82

The best overall rating for the formation for France is 85.64 which is 4-3-1-2 and by applying the best formation for the best squad we get the best 11 on the 4-3-2-1 formation is:

Table 13: Best 11 (France)

-Overall-
Average rating: 85.6

Position	Player	Overall	Age	Club	Value	Wage
GK	H. Lloris	87	34	Tottenham Hotspur	€13.5M	€125K
RB	N. Mukiele	81	23	RB Leipzig	€34.5M	€65K
RCB	R. Varane	86	28	Manchester United	€68.5M	€180K
RCB	J. Koundé	83	22	Sevilla FC	€53M	€33K
LB	T. Hernández	84	23	AC Milan	€62.5M	€51K
RCM	N. Kanté	90	30	Chelsea	€100M	€230K
CAM	N. Fekir	84	27	Real Betis Balompié	€45M	€42K
LCM	T. Lemar	83	25	Atlético de Madrid	€48M	€71K
CF	K. Benzema	89	33	Real Madrid CF	€66M	€350K
ST	K. Mbappé	91	22	Paris Saint-Germain	€194M	€230K
ST	W. Ben Yedder	84	30	AS Monaco	€41.5M	€88K

These are the best team for the particular formation we get. We can see that the highly rated players like K. Mbappe and K. Benzema are there in the list. But as a practical viewer of footballer the France may not select this formation. Formation in football is related to the manager which is going to play. The team always change according to the manager as a manager point of view we can say that France maybe playing this world cup with 4-3-3 and we can make a team for this formation be like:

-Overall-
Average rating: 84.8

Position	Player	Overall	Age	Club	Value	Wage
GK	H. Lloris	87	34	Tottenham Hotspur	€13.5M	€125K
RB	N. Mukiele	81	23	RB Leipzig	€34.5M	€65K
RCB	R. Varane	86	28	Manchester United	€68.5M	€180K
RCB	J. Koundé	83	22	Sevilla FC	€53M	€33K
LB	T. Hernández	84	23	AC Milan	€62.5M	€51K
RCM	N. Kanté	90	30	Chelsea	€100M	€230K
CAM	N. Fekir	84	27	Real Betis Balompié	€45M	€42K
LCM	T. Lemar	83	25	Atlético de Madrid	€48M	€71K
RW	O. Dembélé	83	24	FC Barcelona	€55M	€165K
ST	K. Mbappé	91	22	Paris Saint-Germain	€194M	€230K
LM	M. Diaby	81	21	Bayer 04 Leverkusen	€52.5M	€52K

5.4.2 England:

The second I am selecting here is England which is a team with probably a chance to lift the world cup the best formation for the England team is given by:

Nationality	Squad	Overall
England	3-4-3	84.00
England	4-4-2	84.18
England	4-3-1-2	83.36
England	4-3-3	84.18
England	4-2-3-1	84.18

Here there are three formation with 84.18 we can select either one of them. I decided to select 4-2-3-1 formation for England team:

Table 14: Best 11 (England)

-Overall-
Average rating: 84.2

Position	Player	Overall	Age	Club	Value	Wage
GK	J. Pickford	83	27	Everton	€33M	€83K
RB	T. Alexander-Arnold	87	22	Liverpool	€114M	€150K
LCB	H. Maguire	84	28	Manchester United	€42.5M	€155K
RCB	J. Tarkowski	81	28	Burnley	€24.5M	€46K
LB	L. Shaw	84	25	Manchester United	€48.5M	€140K
RCM	J. Henderson	84	31	Liverpool	€29.5M	€140K
RCM	J. Ward-Prowse	81	26	Southampton	€33.5M	€69K
RW	M. Mount	83	22	Chelsea	€58.5M	€120K
CAM	J. Maddison	82	24	Leicester City	€41.5M	€100K
LM	J. Sancho	87	21	Manchester United	€116.5M	€150K
ST	H. Kane	90	27	Tottenham Hotspur	€129.5M	€240K

All the highly rated players are available in the list which is a good team to compete at world cup.

5.4.3 Brazil:

Best formation for Brazil team is:

Nationality	Squad	Overall
Brazil	3-4-3	85.73
Brazil	4-4-2	84.55
Brazil	4-3-1-2	84.82
Brazil	4-3-3	85.91
Brazil	4-2-3-1	85.82

Brazil's overall rating for the 4-3-3 formation is high and we can select this formation for the further team building. The best eleven for the 4-3-3 formation is given below.

Table 15: Best 11 (Brazil)

-Overall-
Average rating: 85.9

Position	Player	Overall	Age	Club	Value	Wage
GK	Ederson	89	27	Manchester City	€94M	€200K
RB	19 Dani Alves	82	35	Paris Saint-Germain	€4.8M	€60K
RCB	Marquinhos	87	27	Paris Saint-Germain	€90.5M	€135K
CB	Thiago Silva	85	36	Chelsea	€9.5M	€105K
LB	19 Filipe Luís	85	32	Atlético de Madrid	€21.5M	€80K
CDM	Casemiro	89	29	Real Madrid CF	€88M	€310K
CDM	Fabinho	86	27	Liverpool	€73.5M	€165K
CDM	Fernando	84	33	Sevilla FC	€21M	€43K
RM	Raphinha	82	24	Leeds United	€46M	€89K
CF	Roberto Firmino	85	29	Liverpool	€54M	€185K
LW	Neymar Jr	91	29	Paris Saint-Germain	€129M	€270K

These team also includes the best players with overall rating especially with Neymar Jr which is a good view of the model and good team for the upcoming world cup.

5.4.4 Spain:

Next I am taking Spain which we get the best formation as:

	Squad	Overall
Nationality		
Spain	3-4-3	85.64
Spain	4-4-2	85.36
Spain	4-3-1-2	85.55
Spain	4-3-3	85.82
Spain	4-2-3-1	85.73

The best squad for the formation of 4-3-3 with rating of 85.82 is given by:

Table 16: Best 11 (Spain)

-Overall-
Average rating: 85.8

Position	Player	Overall	Age	Club	Value	Wage
GK	De Gea	84	30	Manchester United	€28M	€120K
RB	Carvajal	85	29	Real Madrid CF	€47.5M	€210K
LCB	Sergio Ramos	88	35	Paris Saint-Germain	€24M	€115K
LCB	A. Laporte	86	27	Manchester City	€77M	€185K
LB	Jordi Alba	86	32	FC Barcelona	€47M	€200K
CDM	Rodri	86	25	Manchester City	€81M	€175K
LCM	Thiago	86	30	Liverpool	€65M	€180K
LCM	Parejo	86	32	Villarreal CF	€53.5M	€64K
RM	Marcos Llorente	86	26	Atlético de Madrid	€88M	€95K
ST	Gerard Moreno	86	29	Villarreal CF	€68M	€73K
LW	Oyarzabal	85	24	Real Sociedad	€77.5M	€57K

Here De Gea and the Sergio Ramos is a highly rated players and all other players we can see from the table above.

5.4.5 Argentina:

The best formation for the Argentina international team is given as:

Nationality	Squad	Overall
Argentina	3-4-3	83.36
Argentina	4-4-2	82.18
Argentina	4-3-1-2	82.73
Argentina	4-3-3	84.18
Argentina	4-2-3-1	84.09

We can see that 4-3-3 formation has best overall rating which is a good view as the general formation of Argentina team is 4-3-3 itself. We can see the best eleven for the 4-3-3 formation is:

Table 17: Best 11 (Argentina)

-Overall-
Average rating: 84.2

Position	Player	Overall	Age	Club	Value	Wage
GK	E. Martínez	84	28	Aston Villa	€33.5M	€81K
RB	F. Bustos	77	25	Club Atlético Independiente	€13M	€14K
RCB	20 E. Garay	83	32	Valencia CF	€16.5M	€48K
LCB	N. Otamendi	81	33	SL Benfica	€12.5M	€18K
LB	M. Acuña	84	29	Sevilla FC	€37M	€45K
LCM	A. Gómez	85	33	Sevilla FC	€33.5M	€51K
CAM	P. Dybala	87	27	Juventus	€93M	€160K
CAM	E. Banega	82	33	Al Shabab	€15M	€46K
RW	L. Messi	93	34	Paris Saint-Germain	€78M	€320K
ST	S. Agüero	87	33	FC Barcelona	€51M	€260K
LW	L. Ocampos	83	26	Sevilla FC	€39.5M	€46K

One of the best player in the world L. Messi is in the list which makes this team prediction is a good team almost all the best rated players in the list. So in my opinion this could be a world cup winning team.

5.4.6 Portugal:

Nationality	Squad	Overall
Portugal	3-4-3	83.09
Portugal	4-4-2	83.55
Portugal	4-3-1-2	83.64
Portugal	4-3-3	83.73
Portugal	4-2-3-1	83.73

Here I am selecting the formation 4-2-3-1 with rating of 83.77. we can either select 4-3-3 also but as the Portugal team's style is concerned I am selecting 4-2-3-1 and the team corresponding to this formation is given by:

Table 18: Best 11 (Portugal)

-Overall-
Average rating: 83.7

Position	Player	Overall	Age	Club	Value	Wage
GK	A. Lopes	82	30	Olympique Lyonnais	€23M	€60K
RB	Ricardo Pereira	84	27	Leicester City	€40.5M	€130K
RCB	Rúben Dias	87	24	Manchester City	€102.5M	€170K
RCB	Pepe	82	38	FC Porto	€5.5M	€14K
LB	João Cancelo	86	27	Manchester City	€71.5M	€185K
RCM	Rúben Neves	82	24	Wolverhampton Wanderers	€46M	€89K
CDM	Palhinha	82	25	Sporting CP	€41M	€18K
RW	Rafa	82	28	SL Benfica	€30.5M	€22K
CAM	Bruno Fernandes	88	26	Manchester United	€107.5M	€250K
LWB	Rúben Vinagre	75	22	Sporting CP	€12M	€49K
ST	Cristiano Ronaldo	91	36	Manchester United	€45M	€270K

This models makes me happier as we can see the player like Ronaldo on this list, which makes this models a good predictive model with a great combination of players.

5.4.7 Belgium:

Best formation:

Nationality	Squad	Overall
Belgium	3-4-3	83.00
Belgium	4-4-2	81.91
Belgium	4-3-1-2	81.82
Belgium	4-3-3	82.91
Belgium	4-2-3-1	82.64

Best eleven for 3-4-3 formation:

Table 19: Best 11 (Belgium)

-Overall-
Average rating: 83.0

Position	Player	Overall	Age	Club	Value	Wage
GK	T. Courtois	89	29	Real Madrid CF	€85.5M	€250K
RCB	T. Alderweireld	83	32	NaN	€0	€0
RCB	20 V. Kompany	82	33	RSC Anderlecht	€10.5M	€29K
LCB	J. Vertonghen	81	34	SL Benfica	€8.5M	€18K
RB	T. Meunier	77	29	Borussia Dortmund	€9.5M	€48K
RCM	K. De Bruyne	91	30	Manchester City	€125.5M	€350K
CDM	A. Witsel	83	32	Borussia Dortmund	€24M	€72K
LWB	J. Kayembe	74	26	Royal Charleroi S.C.	€4.5M	€13K
RW	A. Januzaj	81	26	Real Sociedad	€28M	€45K
ST	R. Lukaku	88	28	Chelsea	€93.5M	€260K
LM	Y. Carrasco	84	27	Atlético de Madrid	€45M	€81K

5.4.8 Germany:

Best formation:

Nationality	Squad	Overall
Germany	3-4-3	84.73
Germany	4-4-2	84.36
Germany	4-3-1-2	84.45
Germany	4-3-3	85.09
Germany	4-2-3-1	85.09

Best eleven for 4-2-3-1 formation:

Table 20: Best 11 (Belgium)

-Overall-
Average rating: 85.1

Position	Player	Overall	Age	Club	Value	Wage
GK	M. ter Stegen	90	29	FC Barcelona	€99M	€250K
RB	17 P. Lahm	88	32	FC Bayern München	€29.5M	€200K
LCB	M. Hummels	86	32	Borussia Dortmund	€44M	€95K
RCB	M. Ginter	84	27	Borussia Mönchengladbach	€42.5M	€45K
LB	P. Max	80	27	PSV	€22M	€22K
LCM	T. Kroos	88	31	Real Madrid CF	€75M	€310K
LCM	I. Gündoğan	85	30	Manchester City	€51.5M	€185K
RM	S. Gnabry	85	25	FC Bayern München	€64.5M	€110K
CAM	T. Müller	87	31	FC Bayern München	€66M	€140K
LW	K. Havertz	84	22	Chelsea	€94.5M	€130K
ST	16 M. Klose	79	37	Lazio	€1.3M	€80K

5.4.9 Netherlands:

Best formation:

Nationality	Squad	Overall
Netherlands	3-4-3	84.09
Netherlands	4-4-2	83.45
Netherlands	4-3-1-2	84.09
Netherlands	4-3-3	83.82
Netherlands	4-2-3-1	83.82

Best team for 4-3-1-2 formation is given by:

Table 21: Best 11 (Netherlands)

-Overall-

Average rating: 84.1

Position	Player	Overall	Age	Club	Value	Wage
GK	11 E. van der Sar	83	39	Manchester United	€0	€0
RB	R. Karsdorp	78	26	Roma	€16.5M	€51K
LCB	V. van Dijk	89	29	Liverpool	€86M	€230K
LCB	M. de Ligt	85	21	Juventus	€75M	€81K
LB	D. Blind	82	31	Ajax	€21M	€23K
RCM	F. de Jong	87	24	FC Barcelona	€119.5M	€210K
CM	07 P. Cocu	87	35	PSV	€0	€0
RCM	M. de Roon	82	30	Atalanta	€25.5M	€59K
CAM	14 C. Seedorf	82	37	Botafogo	€2.2M	€60K
ST	09 H. de Noteboom	87	32	111648	€0	€0
ST	W. Weghorst	83	28	VfL Wolfsburg	€37M	€95K

So as a conclusion for this particular question we can say that all the team which I predicted to be the best model using their overall rating seems to be more precisely the best. But we can't conclude that this will be the best team which they need to select for the world cup. All the team and formation are entirely dependent on the team's manager. All the manager's playing style is different for every team so we can come to conclude that this will be a helpful for the managers to view the rated players for different positions.

6 CONCLUSION & VALUE STATEMENT

As we finally finished project and come with better solutions for the questions which I selected. The questions which I selected was

- Exploratory Data Analysis on FIFA22: This include features of players and finding the best players with values and different features:
- Analyse the features of the players and predicting the overall rating by using these features: Features of different players are analysed and predict the overall.
- Analysing the different positions and finding out the best and most skilful players for these positions: Here different positions of football is discussed and also the spider plot is plotted.
- Creating the best team and team formation for some international teams for the upcoming FIFA world cup 2022: We have taken some best world cup team and created a best formation and best team for the particular formation.

While examining these kinds of questions we come up with different types of challenges but all the problems get solved to make an optimum solution to the particular problems.

This project benefits the managers of different teams and the FIFA game players who are looking for the best team and formations. We obtained some best formations and teams which can be used by the managers to examine the team's performance. By attaining the solution to these problems we can achieve social and economic values because we will get the best outcome for the managers to select particular players as per their choices.

7 SELF-REFLECTION

The project gives me a positive approach towards challenging data science projects. I have taken this project individually so doing all the codes and creating different visualize on the first stage is little bit tricky but managed to do the project as my own. Google helped me a lot whenever I am stuck in some situations. Doing a project with totally different types of question which includes Machine Learning is quite interesting. If I were given more time and chance to do this project again I will include more questions related to the Artificial Intelligence using different kinds of predictions techniques. Also I would like to take some challenging data set and include prediction of premier league winners.

8 REFERENCES

Chadysz, G. (2020, 01 02). *http://rstudio-pubs-static.s3.amazonaws.com/*. Retrieved from http://rstudio-pubs-static.s3.amazonaws.com/: http://rstudio-pubs-static.s3.amazonaws.com/564818_82ce98b8892f465f945554b35fd6ac7b.html

fifauteam. (2021). *FIFA 2021 Positions*. Retrieved from fifauteam.com: https://fifauteam.com/fifa-21-positions/

fifauteam.com. (2018, 08 31). *fifauteam.com*. Retrieved from fifauteam.com: https://fifauteam.com/fifa-19-attributes-guide/

Koehrsen, W. (2017, Dec 27). *random-forest-in-python*. Retrieved from https://towardsdatascience.com: https://towardsdatascience.com/random-forest-in-python-24d0893d51c0

Le, J. (2018, 05 14). *FIFA World Cup 2018: A Data-Driven Approach to Ideal Team Line-Ups*. Retrieved from https://towardsdatascience.com: https://towardsdatascience.com/fifa-world-cup-2018-a-data-driven-approach-to-ideal-team-line-ups-93505cfe36f8

sharmaroshan. (2019, May 31). *FIFA-2019-Analysis*. Retrieved from github.com: https://github.com/sharmaroshan/FIFA-2019-Analysis

soutanbasak. (2022). *EDA on FIFA22 stats*. Retrieved from www.kaggle.com: https://www.kaggle.com/code/soutanbasak/eda-on-fifa22-stats/notebook

Towards AI. (2021, 07 07). *Exploratory Data Analysis Expounded With FIFA 2021(Part 1)*. Retrieved from towardsai: https://towardsai.net/p/l/exploratory-data-analysis-expounded-with-fifa-2021part-1

Zambrano-Leon, e.-P. (2021, 01 21). *FIFA Ultimate Team Machine Learning Project*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/fifa-ultimate-team-rating-prediction-machine-learning-project-3a02767fcb38

9 APPENDIX

I am attaching the html file for the jupyter notebook code below.