Minor Project

- our task is to do EDA(exploratry data analysis) on the given data set
- this data have following attributes:-

- Age, Nationality, Overall, Potential, Club, Value, Wage, Preferred Foot, International Reputation, Weak Foot, Skill Moves, Work Rate, 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, Heading, Accuracy, 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, and Release Clause.

In [142…

```
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
plt.figure(figsize=(15,10))
img = mpimg.imread(r'C:\Users\PC-chetan\Downloads\world-cup-g38a43c4e2_1920.jpg')
imgplot = plt.imshow(img)
```

All Insights before cleaning

- there are 18207 rows and 89 columns in this data set
- index start from 0-18207
- in this data set 45 columns with 'object' data type, 6 with 'int' and 38 columns with 'float' data type

- list of column that we'll remove ['Unnamed: 0', 'ID','Photo','Flag','Club Logo','Loaned From','Jersey Number','Real Face']
- 'Unnamed: 0' = its a index of the data set without name
- 'ID' = its a column showing the id of the player and after removing it will not affect our data
- 'Photo' = this column has a lnk of the photo so it will not contribute in our data set for finding some results so we can remove it
- 'Flag' = same like photo column this column has a lnk of the photo so it will not contribute in our data set for finding some results so we can remove it
- 'Club Logo' = same like photo column this column has a lnk of the photo so it will not contribute in our data set for finding some results so we can remove it
- 'Loaned From' = its the info about from where the loan was taken so this is not a column which will help us to find some relation with others columns
- 'Jersey Number' = its a jersey number not specificly related to any of the other column

- 'Real Face' = its also about the yes or no value lke player has real face or not so we can remove this too.

- in this data set out of 89 columns 76 columns have null values
- 43 columns have same number of null values whch is '48'
- may be its a coincedence or we can say he each of these rows of these column are empty
- club column has 241 null values
- again 26 columns have same number of null values which is '2085'
- loaned from column has max number of null values

- here we found out like some data has also null values because count is less
- also we can see most of the players age is around 28
- most of the data is defining good potential but first we have have to clean the data

Insights after cleaning

- now there are 18207 rows and 81 columns in this data set
- index start from 0-18207
- in this data set 40 columns with 'object' data type, 4 with 'int' and 37 columns with 'float' data type
- # we can see clearly all the column with same no. off null value are gone
- also other column are affected by same number of drop
- so its clear by removing that data will effect
- so instead of removing we'll fill those columns
- data set has 'Release Clause' column with '€,M,K' sign we will removw this sign and then fill all the null values with the mean of that column
- we also remove joining date and contract valid until features
- because this columns are not usfull for drawing any result and correlaton

- eldest palyer = O. Pérez

- Youngest Players = G. Nugent
- Best Freekick Takers = L. Messi
- Best players with the Ball Control = L. Messi
- Best Penalty Kick Taker = M. Balotelli
- Fastest Players = L. Messi
- Best Dribbler = L. Messi
- Best Finisher = L. Messi

step 1

- import the required libraries

In [1]:
```
# we use import command to import the libraries
```

In [2]:
```
import pandas as pd # we import pandas to handle the file and its save our lots of time, it handle large data set
import numpy as np #for creating arrays
import seaborn as sns #for data visualzation
import matplotlib.pyplot as plt # for data visualization
```

step 2

- read the given data set

In [3]:
```
# we use (pd.read_csv) for readind csv file
# we (pd.set_optioon) for viewing our data properly we can pass any limit here i passed 500 for each
```

In [4]:
```
df = pd.read_csv(r'C:\Users\PC-chetan\Downloads\data.csv') # to import our data set
pd.set_option('display.max_rows', 500) # for getting the max veiw of raws
pd.set_option('display.max_column', 500) # for getting the max veiw of columns
```

In [5]:
```
#now we will see our data
```

In [6]:
```
df
```

Out[6]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Pot
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	
...	...	...	...	...	...	...	...	...	
18202	18202	238813	J. Lundstram	19	https://cdn.sofifa.org/players/4/19/238813.png	England	https://cdn.sofifa.org/flags/14.png	47	
18203	18203	243165	N. Christoffersson	19	https://cdn.sofifa.org/players/4/19/243165.png	Sweden	https://cdn.sofifa.org/flags/46.png	47	
18204	18204	241638	B. Worman	16	https://cdn.sofifa.org/players/4/19/241638.png	England	https://cdn.sofifa.org/flags/14.png	47	
18205	18205	246268	D. Walker-Rice	17	https://cdn.sofifa.org/players/4/19/246268.png	England	https://cdn.sofifa.org/flags/14.png	47	
18206	18206	246269	G. Nugent	16	https://cdn.sofifa.org/players/4/19/246269.png	England	https://cdn.sofifa.org/flags/14.png	46	

18207 rows × 89 columns

step 3

- now we'll do general analysis of the data.

Exploration of the data

- for understainding the nature of the data and finding out the following:-

> no. of columns. no. of rows. shape of the data. index of the data. data type of the each feature(column). find the list of not use full colmns. no. of null values in each column. after that we will see some vizuals of our data before data cleaning.

In [7]:
```
# df.columns for showing name of the each column of the data set
```

In [8]:
```
df.columns
```

Out[8]:
```
Index(['Unnamed: 0', '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', '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', 'GKDiving', 'GKHandling',
       'GKKicking', 'GKPositioning', 'GKReflexes', 'Release Clause'],
      dtype='object')
```

In [9]:
```
# df.shape for knowing the shape of the data
```

In [10]:
```
df.shape
```

(18207, 89)

Out[10]: (18207, 89)

In [11]:
```
# df.index for knowing the index of our data set
```

In [12]:
```
df.index
```

Out[12]: RangeIndex(start=0, stop=18207, step=1)

In [13]:
```
# df.dtypes for knowing the data type of each column
```

In [14]:
```
df.dtypes
```

Out[14]:
```
Unnamed: 0                  int64
ID                          int64
Name                       object
Age                         int64
Photo                      object
Nationality                object
Flag                       object
Overall                     int64
Potential                   int64
Club                       object
Club Logo                  object
Value                      object
Wage                       object
Special                     int64
Preferred Foot             object
International Reputation  float64
Weak Foot                 float64
Skill Moves               float64
Work Rate                  object
Body Type                  object
Real Face                  object
Position                   object
Jersey Number             float64
Joined                     object
Loaned From                object
Contract Valid Until       object
Height                     object
Weight                     object
LS                         object
ST                         object
RS                         object
LW                         object
LF                         object
CF                         object
RF                         object
RW                         object
LAM                        object
CAM                        object
RAM                        object
LM                         object
LCM                        object
CM                         object
RCM                        object
RM                         object
LWB                        object
LDM                        object
CDM                        object
RDM                        object
RWB                        object
LB                         object
LCB                        object
CB                         object
RCB                        object
RB                         object
Crossing                  float64
Finishing                 float64
HeadingAccuracy           float64
ShortPassing              float64
Volleys                   float64
Dribbling                 float64
Curve                     float64
FKAccuracy                float64
LongPassing               float64
BallControl               float64
Acceleration              float64
SprintSpeed               float64
Agility                   float64
Reactions                 float64
```

```
Balance                   float64
ShotPower                 float64
Jumping                   float64
Stamina                   float64
Strength                  float64
LongShots                 float64
Aggression                float64
Interceptions             float64
Positioning               float64
Vision                    float64
Penalties                 float64
Composure                 float64
Marking                   float64
StandingTackle            float64
SlidingTackle             float64
GKDiving                  float64
GKHandling                float64
GKKicking                 float64
GKPositioning             float64
GKReflexes                float64
Release Clause             object
dtype: object
```

In [15]:
```
# df.select_dtypes for knowing the no. of columns with different datatypes
```

In [16]:
```
df.select_dtypes('object').columns
```

Out[16]:
```
Index(['Name', 'Photo', 'Nationality', 'Flag', 'Club', 'Club Logo', 'Value',
       'Wage', 'Preferred Foot', 'Work Rate', 'Body Type', 'Real Face',
       'Position', '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', 'Release Clause'],
      dtype='object')
```

In [17]:
```
df.select_dtypes('float').columns
```

Out[17]:
```
Index(['International Reputation', 'Weak Foot', 'Skill Moves', 'Jersey Number',
       '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'],
      dtype='object')
```

In [18]:
```
df.select_dtypes('int').columns
```

Out[18]:
```
Index(['Unnamed: 0', 'ID', 'Age', 'Overall', 'Potential', 'Special'], dtype='object')
```

In [19]:
```
# df.info() for getting more clearity on single page
# df.describe() for knowing mean min max values of each colmn in the data set
```

In [20]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18207 entries, 0 to 18206
Data columns (total 89 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Unnamed: 0                18207 non-null  int64
 1   ID                        18207 non-null  int64
 2   Name                      18207 non-null  object
 3   Age                       18207 non-null  int64
 4   Photo                     18207 non-null  object
 5   Nationality               18207 non-null  object
 6   Flag                      18207 non-null  object
 7   Overall                   18207 non-null  int64
 8   Potential                 18207 non-null  int64
 9   Club                      17966 non-null  object
 10  Club Logo                 18207 non-null  object
 11  Value                     18207 non-null  object
```

```
 12  Wage                      18207 non-null  object
 13  Special                   18207 non-null  int64
 14  Preferred Foot            18159 non-null  object
 15  International Reputation  18159 non-null  float64
 16  Weak Foot                 18159 non-null  float64
 17  Skill Moves               18159 non-null  float64
 18  Work Rate                 18159 non-null  object
 19  Body Type                 18159 non-null  object
 20  Real Face                 18159 non-null  object
 21  Position                  18147 non-null  object
 22  Jersey Number             18147 non-null  float64
 23  Joined                    16654 non-null  object
 24  Loaned From               1264 non-null   object
 25  Contract Valid Until      17918 non-null  object
 26  Height                    18159 non-null  object
 27  Weight                    18159 non-null  object
 28  LS                        16122 non-null  object
 29  ST                        16122 non-null  object
 30  RS                        16122 non-null  object
 31  LW                        16122 non-null  object
 32  LF                        16122 non-null  object
 33  CF                        16122 non-null  object
 34  RF                        16122 non-null  object
 35  RW                        16122 non-null  object
 36  LAM                       16122 non-null  object
 37  CAM                       16122 non-null  object
 38  RAM                       16122 non-null  object
 39  LM                        16122 non-null  object
 40  LCM                       16122 non-null  object
 41  CM                        16122 non-null  object
 42  RCM                       16122 non-null  object
 43  RM                        16122 non-null  object
 44  LWB                       16122 non-null  object
 45  LDM                       16122 non-null  object
 46  CDM                       16122 non-null  object
 47  RDM                       16122 non-null  object
 48  RWB                       16122 non-null  object
 49  LB                        16122 non-null  object
 50  LCB                       16122 non-null  object
 51  CB                        16122 non-null  object
 52  RCB                       16122 non-null  object
 53  RB                        16122 non-null  object
 54  Crossing                  18159 non-null  float64
 55  Finishing                 18159 non-null  float64
 56  HeadingAccuracy           18159 non-null  float64
 57  ShortPassing              18159 non-null  float64
 58  Volleys                   18159 non-null  float64
 59  Dribbling                 18159 non-null  float64
 60  Curve                     18159 non-null  float64
 61  FKAccuracy                18159 non-null  float64
 62  LongPassing               18159 non-null  float64
 63  BallControl               18159 non-null  float64
 64  Acceleration              18159 non-null  float64
 65  SprintSpeed               18159 non-null  float64
 66  Agility                   18159 non-null  float64
 67  Reactions                 18159 non-null  float64
 68  Balance                   18159 non-null  float64
 69  ShotPower                 18159 non-null  float64
 70  Jumping                   18159 non-null  float64
 71  Stamina                   18159 non-null  float64
 72  Strength                  18159 non-null  float64
 73  LongShots                 18159 non-null  float64
 74  Aggression                18159 non-null  float64
 75  Interceptions             18159 non-null  float64
 76  Positioning               18159 non-null  float64
 77  Vision                    18159 non-null  float64
 78  Penalties                 18159 non-null  float64
 79  Composure                 18159 non-null  float64
 80  Marking                   18159 non-null  float64
 81  StandingTackle            18159 non-null  float64
 82  SlidingTackle             18159 non-null  float64
 83  GKDiving                  18159 non-null  float64
 84  GKHandling                18159 non-null  float64
 85  GKKicking                 18159 non-null  float64
 86  GKPositioning             18159 non-null  float64
 87  GKReflexes                18159 non-null  float64
 88  Release Clause            16643 non-null  object
dtypes: float64(38), int64(6), object(45)
memory usage: 12.4+ MB
```

In [21]:

```
df.describe()
```

Out[21]:

	Unnamed: 0	ID	Age	Overall	Potential	Special	International Reputation	Weak Foot	Skill Moves	N

count	18207.000000	18207.000000	18207.000000	18207.000000	18207.000000	18207.000000	18159.000000	18159.000000	18159.000000	18147.
mean	9103.000000	214298.338606	25.122206	66.238699	71.307299	1597.809908	1.113222	2.947299	2.361308	19.
std	5256.052511	29965.244204	4.669943	6.908930	6.136496	272.586016	0.394031	0.660456	0.756164	15.
min	0.000000	16.000000	16.000000	46.000000	48.000000	731.000000	1.000000	1.000000	1.000000	1.
25%	4551.500000	200315.500000	21.000000	62.000000	67.000000	1457.000000	1.000000	3.000000	2.000000	8.
50%	9103.000000	221759.000000	25.000000	66.000000	71.000000	1635.000000	1.000000	3.000000	2.000000	17.
75%	13654.500000	236529.500000	28.000000	71.000000	75.000000	1787.000000	1.000000	3.000000	3.000000	26.
max	18206.000000	246620.000000	45.000000	94.000000	95.000000	2346.000000	5.000000	5.000000	5.000000	99.

In [22]:
```
df.describe().T # this command change the position of rows with columns
```

Out[22]:

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	18207.0	9103.000000	5256.052511	0.0	4551.5	9103.0	13654.5	18206.0
ID	18207.0	214298.338606	29965.244204	16.0	200315.5	221759.0	236529.5	246620.0
Age	18207.0	25.122206	4.669943	16.0	21.0	25.0	28.0	45.0
Overall	18207.0	66.238699	6.908930	46.0	62.0	66.0	71.0	94.0
Potential	18207.0	71.307299	6.136496	48.0	67.0	71.0	75.0	95.0
Special	18207.0	1597.809908	272.586016	731.0	1457.0	1635.0	1787.0	2346.0
International Reputation	18159.0	1.113222	0.394031	1.0	1.0	1.0	1.0	5.0
Weak Foot	18159.0	2.947299	0.660456	1.0	3.0	3.0	3.0	5.0
Skill Moves	18159.0	2.361308	0.756164	1.0	2.0	2.0	3.0	5.0
Jersey Number	18147.0	19.546096	15.947765	1.0	8.0	17.0	26.0	99.0
Crossing	18159.0	49.734181	18.364524	5.0	38.0	54.0	64.0	93.0
Finishing	18159.0	45.550911	19.525820	2.0	30.0	49.0	62.0	95.0
HeadingAccuracy	18159.0	52.298144	17.379909	4.0	44.0	56.0	64.0	94.0
ShortPassing	18159.0	58.686712	14.699495	7.0	54.0	62.0	68.0	93.0
Volleys	18159.0	42.909026	17.694408	4.0	30.0	44.0	57.0	90.0
Dribbling	18159.0	55.371001	18.910371	4.0	49.0	61.0	68.0	97.0
Curve	18159.0	47.170824	18.395264	6.0	34.0	48.0	62.0	94.0
FKAccuracy	18159.0	42.863153	17.478763	3.0	31.0	41.0	57.0	94.0
LongPassing	18159.0	52.711933	15.327870	9.0	43.0	56.0	64.0	93.0
BallControl	18159.0	58.369459	16.686595	5.0	54.0	63.0	69.0	96.0
Acceleration	18159.0	64.614076	14.927780	12.0	57.0	67.0	75.0	97.0
SprintSpeed	18159.0	64.726967	14.649953	12.0	57.0	67.0	75.0	96.0
Agility	18159.0	63.503607	14.766049	14.0	55.0	66.0	74.0	96.0
Reactions	18159.0	61.836610	9.010464	21.0	56.0	62.0	68.0	96.0
Balance	18159.0	63.966573	14.136166	16.0	56.0	66.0	74.0	96.0
ShotPower	18159.0	55.460047	17.237958	2.0	45.0	59.0	68.0	95.0
Jumping	18159.0	65.089432	11.820044	15.0	58.0	66.0	73.0	95.0
Stamina	18159.0	63.219946	15.894741	12.0	56.0	66.0	74.0	96.0
Strength	18159.0	65.311967	12.557000	17.0	58.0	67.0	74.0	97.0
LongShots	18159.0	47.109973	19.260524	3.0	33.0	51.0	62.0	94.0
Aggression	18159.0	55.868991	17.367967	11.0	44.0	59.0	69.0	95.0
Interceptions	18159.0	46.698276	20.696909	3.0	26.0	52.0	64.0	92.0
Positioning	18159.0	49.958478	19.529036	2.0	38.0	55.0	64.0	95.0
Vision	18159.0	53.400903	14.146881	10.0	44.0	55.0	64.0	94.0
Penalties	18159.0	48.548598	15.704053	5.0	39.0	49.0	60.0	92.0
Composure	18159.0	58.648274	11.436133	3.0	51.0	60.0	67.0	96.0
Marking	18159.0	47.281623	19.904397	3.0	30.0	53.0	64.0	94.0
StandingTackle	18159.0	47.697836	21.664004	2.0	27.0	55.0	66.0	93.0
SlidingTackle	18159.0	45.661435	21.289135	3.0	24.0	52.0	64.0	91.0
GKDiving	18159.0	16.616223	17.695349	1.0	8.0	11.0	14.0	90.0
GKHandling	18159.0	16.391596	16.906900	1.0	8.0	11.0	14.0	92.0
GKKicking	18159.0	16.232061	16.502864	1.0	8.0	11.0	14.0	91.0

GKPositioning	18159.0	16.388898	17.034669	1.0	8.0	11.0	14.0	90.0
GKReflexes	18159.0	16.710887	17.955119	1.0	8.0	11.0	14.0	94.0

In [23]:
```
# here we found out like some data has also null values because count is less
# also we can see most of the players age is around 28
# most of the data is defining good potential but first we have have to clean the data
# before that we will see some data vizually
```

In [24]:
```
# df.head() for printing top 5 rows of our data set
```

In [25]:
```
df.head()
```

Out[25]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	94	B
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	94	
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	93	Par
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	93	Ma
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	92	Ma

In [26]:
```
# here we saw
#'Unnamed: 0' = its a index of the data set without name
#'ID' = its a column showing the id of the player and after removing it will not affect our data
#'Photo' = this column has a lnk of the photo so it will not contribute in our data set for finding some results
#'Flag' = same like photo column this column has a lnk of the photo so it will not contribute in our data set fo
#'Club Logo' = same like photo column this column has a lnk of the photo so it will not contribute in our data se
#'Loaned From' =  its the info about from where the loan was taken so this is not a column which will help us to
#'Jersey Number' =  its a jersey number not specificly related to any of the other column
#'Real Face' =   its also about the yes or no value lke player has real face or not so we can remove this too.
```

In [27]:
```
# list of column that we'll remove ['Unnamed: 0', 'ID','Photo','Flag','Club Logo','Loaned From','Jersey Number','
```

In [28]:
```
# df.isnull().sum() for getting the count of each in value of each feature
```

In [29]:
```
df.isnull().sum()
```

Out[29]:
```
Unnamed: 0                    0
ID                            0
Name                          0
Age                           0
Photo                         0
Nationality                   0
Flag                          0
Overall                       0
Potential                     0
Club                        241
Club Logo                     0
Value                         0
Wage                          0
Special                       0
Preferred Foot               48
International Reputation     48
Weak Foot                    48
Skill Moves                  48
Work Rate                    48
Body Type                    48
Real Face                    48
Position                     60
Jersey Number                60
Joined                     1553
Loaned From               16943
Contract Valid Until        289
Height                       48
Weight                       48
```

```
LS                        2085
ST                        2085
RS                        2085
LW                        2085
LF                        2085
CF                        2085
RF                        2085
RW                        2085
LAM                       2085
CAM                       2085
RAM                       2085
LM                        2085
LCM                       2085
CM                        2085
RCM                       2085
RM                        2085
LWB                       2085
LDM                       2085
CDM                       2085
RDM                       2085
RWB                       2085
LB                        2085
LCB                       2085
CB                        2085
RCB                       2085
RB                        2085
Crossing                    48
Finishing                   48
HeadingAccuracy             48
ShortPassing                48
Volleys                     48
Dribbling                   48
Curve                       48
FKAccuracy                  48
LongPassing                 48
BallControl                 48
Acceleration                48
SprintSpeed                 48
Agility                     48
Reactions                   48
Balance                     48
ShotPower                   48
Jumping                     48
Stamina                     48
Strength                    48
LongShots                   48
Aggression                  48
Interceptions               48
Positioning                 48
Vision                      48
Penalties                   48
Composure                   48
Marking                     48
StandingTackle              48
SlidingTackle               48
GKDiving                    48
GKHandling                  48
GKKicking                   48
GKPositioning               48
GKReflexes                  48
Release Clause            1564
dtype: int64
```

In [30]:

```
#in this data set out of 89 columns 76 columns have null values
#43 columns have same number of null values whch is '48'
#may be its a coincedence or we can say he each of these rows of these column are empty
#club column has 241 null values
#again 26 columns have same number of null values which is '2085' ]
#loaned from column has max number of null values
```

step 4

- data cleaning

Data Cleaning

- Dealing with Null values of each columns and also with other inappropriate datas in the column

In [31]:

```
df.head()
```

Out[31]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	94	B
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	94	
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	93	Par
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	93	Ma
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	92	Ma

In [32]:
```
# df.drop(column =['c1','c2'], axis= 1, inplace= ture) this is syntex for removing columns from the dataset perma
```

In [33]:
```
df.drop(columns=['Unnamed: 0', 'ID','Photo','Flag','Club Logo','Loaned From','Jersey Number','Real Face'], axis =
```

In [34]:
```
df.head()
```

Out[34]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate	Body Type
0	L. Messi	31	Argentina	94	94	FC Barcelona	€110.5M	€565K	2202	Left	5.0	4.0	4.0	Medium/ Medium	Messi
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	€77M	€405K	2228	Right	5.0	4.0	5.0	High/ Low	C. Ronaldo
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	€118.5M	€290K	2143	Right	5.0	5.0	5.0	High/ Medium	Neymar
3	De Gea	27	Spain	91	93	Manchester United	€72M	€260K	1471	Right	4.0	3.0	1.0	Medium/ Medium	Lean
4	K. De Bruyne	27	Belgium	91	92	Manchester City	€102M	€355K	2281	Right	4.0	5.0	4.0	High/ High	Normal

In [35]:
```
df.shape
```

Out[35]:
```
(18207, 81)
```

In [36]:
```
df.columns
```

Out[36]:
```
Index(['Name', 'Age', 'Nationality', 'Overall', 'Potential', 'Club', 'Value',
       'Wage', 'Special', 'Preferred Foot', 'International Reputation',
       'Weak Foot', 'Skill Moves', 'Work Rate', 'Body Type', 'Position',
       'Joined', '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', 'GKDiving',
       'GKHandling', 'GKKicking', 'GKPositioning', 'GKReflexes',
       'Release Clause'],
      dtype='object')
```

In [37]:
```
df.index
```

Out[37]:
```
RangeIndex(start=0, stop=18207, step=1)
```

In [38]:
```
df.dtypes
```

Out[38]:

```
Name                        object
Age                          int64
Nationality                 object
Overall                      int64
Potential                    int64
Club                        object
Value                       object
Wage                        object
Special                      int64
Preferred Foot              object
International Reputation   float64
Weak Foot                  float64
Skill Moves                float64
Work Rate                   object
Body Type                   object
Position                    object
Joined                      object
Contract Valid Until        object
Height                      object
Weight                      object
LS                          object
ST                          object
RS                          object
LW                          object
LF                          object
CF                          object
RF                          object
RW                          object
LAM                         object
CAM                         object
RAM                         object
LM                          object
LCM                         object
CM                          object
RCM                         object
RM                          object
LWB                         object
LDM                         object
CDM                         object
RDM                         object
RWB                         object
LB                          object
LCB                         object
CB                          object
RCB                         object
RB                          object
Crossing                   float64
Finishing                  float64
HeadingAccuracy            float64
ShortPassing               float64
Volleys                    float64
Dribbling                  float64
Curve                      float64
FKAccuracy                 float64
LongPassing                float64
BallControl                float64
Acceleration               float64
SprintSpeed                float64
Agility                    float64
Reactions                  float64
Balance                    float64
ShotPower                  float64
Jumping                    float64
Stamina                    float64
Strength                   float64
LongShots                  float64
Aggression                 float64
Interceptions              float64
Positioning                float64
Vision                     float64
Penalties                  float64
Composure                  float64
Marking                    float64
StandingTackle             float64
SlidingTackle              float64
GKDiving                   float64
GKHandling                 float64
GKKicking                  float64
GKPositioning              float64
GKReflexes                 float64
Release Clause              object
dtype: object
```

In [39]:

```
df.isnull().sum()
```

Out[39]:

```
Name                         0
Age                          0
Nationality                  0
Overall                      0
Potential                    0
Club                       241
Value                        0
Wage                         0
Special                      0
Preferred Foot              48
International Reputation    48
Weak Foot                   48
Skill Moves                 48
Work Rate                   48
Body Type                   48
Position                    60
Joined                    1553
Contract Valid Until       289
Height                      48
Weight                      48
LS                        2085
ST                        2085
RS                        2085
LW                        2085
LF                        2085
CF                        2085
RF                        2085
RW                        2085
LAM                       2085
CAM                       2085
RAM                       2085
LM                        2085
LCM                       2085
CM                        2085
RCM                       2085
RM                        2085
LWB                       2085
LDM                       2085
CDM                       2085
RDM                       2085
RWB                       2085
LB                        2085
LCB                       2085
CB                        2085
RCB                       2085
RB                        2085
Crossing                    48
Finishing                   48
HeadingAccuracy             48
ShortPassing                48
Volleys                     48
Dribbling                   48
Curve                       48
FKAccuracy                  48
LongPassing                 48
BallControl                 48
Acceleration                48
SprintSpeed                 48
Agility                     48
Reactions                   48
Balance                     48
ShotPower                   48
Jumping                     48
Stamina                     48
Strength                    48
LongShots                   48
Aggression                  48
Interceptions               48
Positioning                 48
Vision                      48
Penalties                   48
Composure                   48
Marking                     48
StandingTackle              48
SlidingTackle               48
GKDiving                    48
GKHandling                  48
GKKicking                   48
GKPositioning               48
GKReflexes                  48
Release Clause            1564
dtype: int64
```

In [ ]:

```
# in this data there are 48 rows those are totally null
Preferred Foot               48
```

```
International Reputation      48
Weak Foot                     48
Skill Moves                   48
Work Rate                     48
Body Type                     48
Position                      60
Joined                      1553
Contract Valid Until         289
Height                        48
Weight                        48
LS                          2085
ST                          2085
RS                          2085
LW                          2085
LF                          2085
CF                          2085
RF                          2085
RW                          2085
LAM                         2085
CAM                         2085
RAM                         2085
LM                          2085
LCM                         2085
CM                          2085
RCM                         2085
RM                          2085
LWB                         2085
LDM                         2085
CDM                         2085
RDM                         2085
RWB                         2085
LB                          2085
LCB                         2085
CB                          2085
RCB                         2085
RB                          2085
Crossing                      48
Finishing                     48
HeadingAccuracy               48
ShortPassing                  48
Volleys                       48
Dribbling                     48
Curve                         48
FKAccuracy                    48
LongPassing                   48
BallControl                   48
Acceleration                  48
SprintSpeed                   48
Agility                       48
Reactions                     48
Balance                       48
ShotPower                     48
Jumping                       48
Stamina                       48
Strength                      48
LongShots                     48
Aggression                    48
Interceptions                 48
Positioning                   48
Vision                        48
Penalties                     48
Composure                     48
Marking                       48
StandingTackle                48
SlidingTackle                 48
GKDiving                      48
GKHandling                    48
GKKicking                     48
GKPositioning                 48
GKReflexes                    48
Release Clause              1564
```

In [40]:
```
#df.dropna()  for remooving all the null values
```

In [41]:
```
# here i picked the random feature and after removing that lets see what will happen
```

In [42]:
```
df.drop(df[df['Height'].isnull()].index,inplace =True)
```

In [43]:
```
df.isnull().sum()
```

Out[43]:
```
Name                        0
Age                         0
Nationality                 0
```

```
Overall                       0
Potential                     0
Club                        241
Value                         0
Wage                          0
Special                       0
Preferred Foot                0
International Reputation      0
Weak Foot                     0
Skill Moves                   0
Work Rate                     0
Body Type                     0
Position                     12
Joined                     1505
Contract Valid Until        241
Height                        0
Weight                        0
LS                         2037
ST                         2037
RS                         2037
LW                         2037
LF                         2037
CF                         2037
RF                         2037
RW                         2037
LAM                        2037
CAM                        2037
RAM                        2037
LM                         2037
LCM                        2037
CM                         2037
RCM                        2037
RM                         2037
LWB                        2037
LDM                        2037
CDM                        2037
RDM                        2037
RWB                        2037
LB                         2037
LCB                        2037
CB                         2037
RCB                        2037
RB                         2037
Crossing                      0
Finishing                     0
HeadingAccuracy               0
ShortPassing                  0
Volleys                       0
Dribbling                     0
Curve                         0
FKAccuracy                    0
LongPassing                   0
BallControl                   0
Acceleration                  0
SprintSpeed                   0
Agility                       0
Reactions                     0
Balance                       0
ShotPower                     0
Jumping                       0
Stamina                       0
Strength                      0
LongShots                     0
Aggression                    0
Interceptions                 0
Positioning                   0
Vision                        0
Penalties                     0
Composure                     0
Marking                       0
StandingTackle                0
SlidingTackle                 0
GKDiving                      0
GKHandling                    0
GKKicking                     0
GKPositioning                 0
GKReflexes                    0
Release Clause             1516
dtype: int64
```

In [44]:

```
# we can see clearly all the column with same no. off null value are gone
# also  other column are affected by same number of drop
# so its clear by removing that data will effect
# so instead of removing we'll fill those columns
```

In [45]:
```
# so here we will going to call our data again
```

In [46]:
```
df = pd.read_csv(r'C:\Users\PC-chetan\Downloads\data.csv') # to import our data set
pd.set_option('display.max_rows', 500) # for getting the max veiw of raws
pd.set_option('display.max_column', 500) # for getting the max veiw of columns
```

In [47]:
```
df
```

Out[47]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Pot
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	
...	...	...	...	...	...	...	...	...	
18202	18202	238813	J. Lundstram	19	https://cdn.sofifa.org/players/4/19/238813.png	England	https://cdn.sofifa.org/flags/14.png	47	
18203	18203	243165	N. Christoffersson	19	https://cdn.sofifa.org/players/4/19/243165.png	Sweden	https://cdn.sofifa.org/flags/46.png	47	
18204	18204	241638	B. Worman	16	https://cdn.sofifa.org/players/4/19/241638.png	England	https://cdn.sofifa.org/flags/14.png	47	
18205	18205	246268	D. Walker-Rice	17	https://cdn.sofifa.org/players/4/19/246268.png	England	https://cdn.sofifa.org/flags/14.png	47	
18206	18206	246269	G. Nugent	16	https://cdn.sofifa.org/players/4/19/246269.png	England	https://cdn.sofifa.org/flags/14.png	46	

18207 rows × 89 columns

In [48]:
```
# now we w'll remove those columns
```

In [ ]:
```
df.drop(columns=['Unnamed: 0', 'ID','Photo','Flag','Club Logo','Loaned From','Jersey Number','Real Face'], axis =
```

In [50]:
```
df
```

Out[50]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate
0	L. Messi	31	Argentina	94	94	FC Barcelona	€110.5M	€565K	2202	Left	5.0	4.0	4.0	Medium/ Medium
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	€77M	€405K	2228	Right	5.0	4.0	5.0	High/ Low
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	€118.5M	€290K	2143	Right	5.0	5.0	5.0	High/ Medium
3	De Gea	27	Spain	91	93	Manchester United	€72M	€260K	1471	Right	4.0	3.0	1.0	Medium/ Medium
4	K. De Bruyne	27	Belgium	91	92	Manchester City	€102M	€355K	2281	Right	4.0	5.0	4.0	High/ High
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
18202	J. Lundstram	19	England	47	65	Crewe Alexandra	€60K	€1K	1307	Right	1.0	2.0	2.0	Medium/ Medium
18203	N. Christoffersson	19	Sweden	47	63	Trelleborgs FF	€60K	€1K	1098	Right	1.0	2.0	2.0	Medium/ Medium
18204	B. Worman	16	England	47	67	Cambridge United	€60K	€1K	1189	Right	1.0	3.0	2.0	Medium/ Medium
18205	D. Walker-Rice	17	England	47	66	Tranmere Rovers	€60K	€1K	1228	Right	1.0	3.0	2.0	Medium/ Medium
						Tranmere								Medium/

18206	G. Nugent	16	England	46	66	Rovers	€60K	€1K	1321	Right	1.0	3.0	2.0	Medium

18207 rows × 81 columns

In [51]:
```
df.isnull().sum()
```

Out[51]:
```
Name                          0
Age                           0
Nationality                   0
Overall                       0
Potential                     0
Club                        241
Value                         0
Wage                          0
Special                       0
Preferred Foot               48
International Reputation     48
Weak Foot                    48
Skill Moves                  48
Work Rate                    48
Body Type                    48
Position                     60
Joined                     1553
Contract Valid Until        289
Height                       48
Weight                       48
LS                         2085
ST                         2085
RS                         2085
LW                         2085
LF                         2085
CF                         2085
RF                         2085
RW                         2085
LAM                        2085
CAM                        2085
RAM                        2085
LM                         2085
LCM                        2085
CM                         2085
RCM                        2085
RM                         2085
LWB                        2085
LDM                        2085
CDM                        2085
RDM                        2085
RWB                        2085
LB                         2085
LCB                        2085
CB                         2085
RCB                        2085
RB                         2085
Crossing                     48
Finishing                    48
HeadingAccuracy              48
ShortPassing                 48
Volleys                      48
Dribbling                    48
Curve                        48
FKAccuracy                   48
LongPassing                  48
BallControl                  48
Acceleration                 48
SprintSpeed                  48
Agility                      48
Reactions                    48
Balance                      48
ShotPower                    48
Jumping                      48
Stamina                      48
Strength                     48
LongShots                    48
Aggression                   48
Interceptions                48
Positioning                  48
Vision                       48
Penalties                    48
Composure                    48
Marking                      48
StandingTackle               48
SlidingTackle                48
GKDiving                     48
GKHandling                   48
GKKicking                    48
```

```
GKPositioning                48
GKReflexes                   48
Release Clause             1564
dtype: int64
```

In [52]:
```
# we w'll fill all the values with mean of the each
```

In [53]:
```
df['Club'].fillna('No Club', inplace = True)
df['Preferred Foot'].fillna('Right', inplace = True)
df['International Reputation'].fillna(df['International Reputation'].mean(), inplace = True)
df['Weak Foot'].fillna(df['Weak Foot'].mean(), inplace = True)
df['Skill Moves'].fillna(df['Skill Moves'].mean(), inplace = True)
df['Work Rate'].fillna('Medium/ Medium', inplace = True)
df['Body Type'].fillna('Normal', inplace = True)
df['Position'].fillna('unknown', inplace = True)

df['Crossing'].fillna(df['Crossing'].mean(), inplace = True)
df['Finishing'].fillna(df['Finishing'].mean(), inplace = True)
df['HeadingAccuracy'].fillna(df['HeadingAccuracy'].mean(), inplace = True)
df['ShortPassing'].fillna(df['ShortPassing'].mean(), inplace = True)
df['Volleys'].fillna(df['Volleys'].mean(), inplace = True)
df['Dribbling'].fillna(df['Dribbling'].mean(), inplace = True)
df['Curve'].fillna(df['Curve'].mean(), inplace = True)
df['FKAccuracy'].fillna(df['FKAccuracy'].mean(), inplace = True)
df['LongPassing'].fillna(df['LongPassing'].mean(), inplace = True)
df['BallControl'].fillna(df['BallControl'].mean(), inplace = True)
df['Acceleration'].fillna(df['Acceleration'].mean(), inplace = True)
df['SprintSpeed'].fillna(df['SprintSpeed'].mean(), inplace = True)
df['Agility'].fillna(df['Agility'].mean(), inplace = True)
df['Reactions'].fillna(df['Reactions'].mean(), inplace = True)
df['Balance'].fillna(df['Balance'].mean(), inplace = True)
df['ShotPower'].fillna(df['ShotPower'].mean(), inplace = True)
df['Jumping'].fillna(df['Jumping'].mean(), inplace = True)
df['Stamina'].fillna(df['Stamina'].mean(), inplace = True)
df['Strength'].fillna(df['Strength'].mean(), inplace = True)
df['LongShots'].fillna(df['LongShots'].mean(), inplace = True)
df['Aggression'].fillna(df['Aggression'].mean(), inplace = True)
df['Interceptions'].fillna(df['Interceptions'].mean(), inplace = True)
df['Positioning'].fillna(df['Positioning'].mean(), inplace = True)
df['Vision'].fillna(df['Vision'].mean(), inplace = True)
df['Penalties'].fillna(df['Penalties'].mean(), inplace = True)
df['Composure'].fillna(df['Composure'].mean(), inplace = True)
df['Marking'].fillna(df['Marking'].mean(), inplace = True)
df['StandingTackle'].fillna(df['StandingTackle'].mean(), inplace = True)
df['SlidingTackle'].fillna(df['SlidingTackle'].mean(), inplace = True)
df['GKDiving'].fillna(df['GKDiving'].mean(), inplace = True)
df['GKHandling'].fillna(df['GKHandling'].mean(), inplace = True)
df['GKKicking'].fillna(df['GKKicking'].mean(), inplace = True)
df['GKPositioning'].fillna(df['GKPositioning'].mean(), inplace = True)
df['GKReflexes'].fillna(df['GKReflexes'].mean(), inplace = True)
```

In [54]:
```
df.isnull().sum()
```

Out[54]:
```
Name                         0
Age                          0
Nationality                  0
Overall                      0
Potential                    0
Club                         0
Value                        0
Wage                         0
Special                      0
Preferred Foot               0
International Reputation     0
Weak Foot                    0
Skill Moves                  0
Work Rate                    0
Body Type                    0
Position                     0
Joined                    1553
Contract Valid Until       289
Height                      48
Weight                      48
LS                        2085
ST                        2085
RS                        2085
LW                        2085
LF                        2085
CF                        2085
RF                        2085
RW                        2085
LAM                       2085
CAM                       2085
```

```
RAM                       2085
LM                        2085
LCM                       2085
CM                        2085
RCM                       2085
RM                        2085
LWB                       2085
LDM                       2085
CDM                       2085
RDM                       2085
RWB                       2085
LB                        2085
LCB                       2085
CB                        2085
RCB                       2085
RB                        2085
Crossing                     0
Finishing                    0
HeadingAccuracy              0
ShortPassing                 0
Volleys                      0
Dribbling                    0
Curve                        0
FKAccuracy                   0
LongPassing                  0
BallControl                  0
Acceleration                 0
SprintSpeed                  0
Agility                      0
Reactions                    0
Balance                      0
ShotPower                    0
Jumping                      0
Stamina                      0
Strength                     0
LongShots                    0
Aggression                   0
Interceptions                0
Positioning                  0
Vision                       0
Penalties                    0
Composure                    0
Marking                      0
StandingTackle               0
SlidingTackle                0
GKDiving                     0
GKHandling                   0
GKKicking                    0
GKPositioning                0
GKReflexes                   0
Release Clause            1564
dtype: int64
```

In [55]:
```
df.head()
```

Out[55]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate	Body Type
0	L. Messi	31	Argentina	94	94	FC Barcelona	€110.5M	€565K	2202	Left	5.0	4.0	4.0	Medium/ Medium	Messi
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	€77M	€405K	2228	Right	5.0	4.0	5.0	High/ Low	C. Ronaldo
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	€118.5M	€290K	2143	Right	5.0	5.0	5.0	High/ Medium	Neymar
3	De Gea	27	Spain	91	93	Manchester United	€72M	€260K	1471	Right	4.0	3.0	1.0	Medium/ Medium	Lean
4	K. De Bruyne	27	Belgium	91	92	Manchester City	€102M	€355K	2281	Right	4.0	5.0	4.0	High/ High	Normal

In [56]:
```
df['Release Clause'].unique()
```

Out[56]:
```
array(['€226.5M', '€127.1M', '€228.1M', ..., '€74K', '€101K', '€147K'],
      dtype=object)
```

In [ ]:
```
def value_conversion(Value):
```

```
    if isinstance(Value,str):
        out = Value.replace('€','')
        if 'M' in out:
            out = float(out.replace('M',''))*1000000
        elif 'K' in Value:
            out = float(out.replace('K',''))*1000
        return float(out)

df['Value'] = df['Value'].apply(lambda x: value_conversion(x))
df['Wage'] = df['Wage'].apply(lambda x: value_conversion(x))
df['Release Clause'] = df['Release Clause'].apply(lambda x: value_conversion(x))
df['Release Clause'].fillna(df['Release Clause'].mean(), inplace = True)
```

In [59]:

```
df.head()
```

Out[59]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate	
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	2202	Left	5.0	4.0	4.0	Medium/ Medium	
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	2228	Right	5.0	4.0	5.0	High/ Low	Rc
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	2143	Right	5.0	5.0	5.0	High/ Medium	N
3	De Gea	27	Spain	91	93	Manchester United	72000000.0	260000.0	1471	Right	4.0	3.0	1.0	Medium/ Medium	
4	K. De Bruyne	27	Belgium	91	92	Manchester City	102000000.0	355000.0	2281	Right	4.0	5.0	4.0	High/ High	N

In [60]:

```
df.isnull().sum()
```

Out[60]:

```
Name                          0
Age                           0
Nationality                   0
Overall                       0
Potential                     0
Club                          0
Value                         0
Wage                          0
Special                       0
Preferred Foot                0
International Reputation      0
Weak Foot                     0
Skill Moves                   0
Work Rate                     0
Body Type                     0
Position                      0
Joined                     1553
Contract Valid Until        289
Height                       48
Weight                       48
LS                         2085
ST                         2085
RS                         2085
LW                         2085
LF                         2085
CF                         2085
RF                         2085
RW                         2085
LAM                        2085
CAM                        2085
RAM                        2085
LM                         2085
LCM                        2085
CM                         2085
RCM                        2085
RM                         2085
LWB                        2085
LDM                        2085
CDM                        2085
RDM                        2085
RWB                        2085
LB                         2085
LCB                        2085
CB                         2085
RCB                        2085
RB                         2085
Crossing                      0
Finishing                     0
HeadingAccuracy               0
```

```
ShortPassing              0
Volleys                   0
Dribbling                 0
Curve                     0
FKAccuracy                0
LongPassing               0
BallControl               0
Acceleration              0
SprintSpeed               0
Agility                   0
Reactions                 0
Balance                   0
ShotPower                 0
Jumping                   0
Stamina                   0
Strength                  0
LongShots                 0
Aggression                0
Interceptions             0
Positioning               0
Vision                    0
Penalties                 0
Composure                 0
Marking                   0
StandingTackle            0
SlidingTackle             0
GKDiving                  0
GKHandling                0
GKKicking                 0
GKPositioning             0
GKReflexes                0
Release Clause            0
dtype: int64
```

In [ ]:
```
# we also remove joining date and contract valid until features
# because this columns are not usfull for drawing any result and correlaton
```

In [64]:
```
df.drop(columns =['Joined','Contract Valid Until'], axis = 1, inplace= True )
```

step 5

- visualizations using univariate bivariate and multivariate analysis

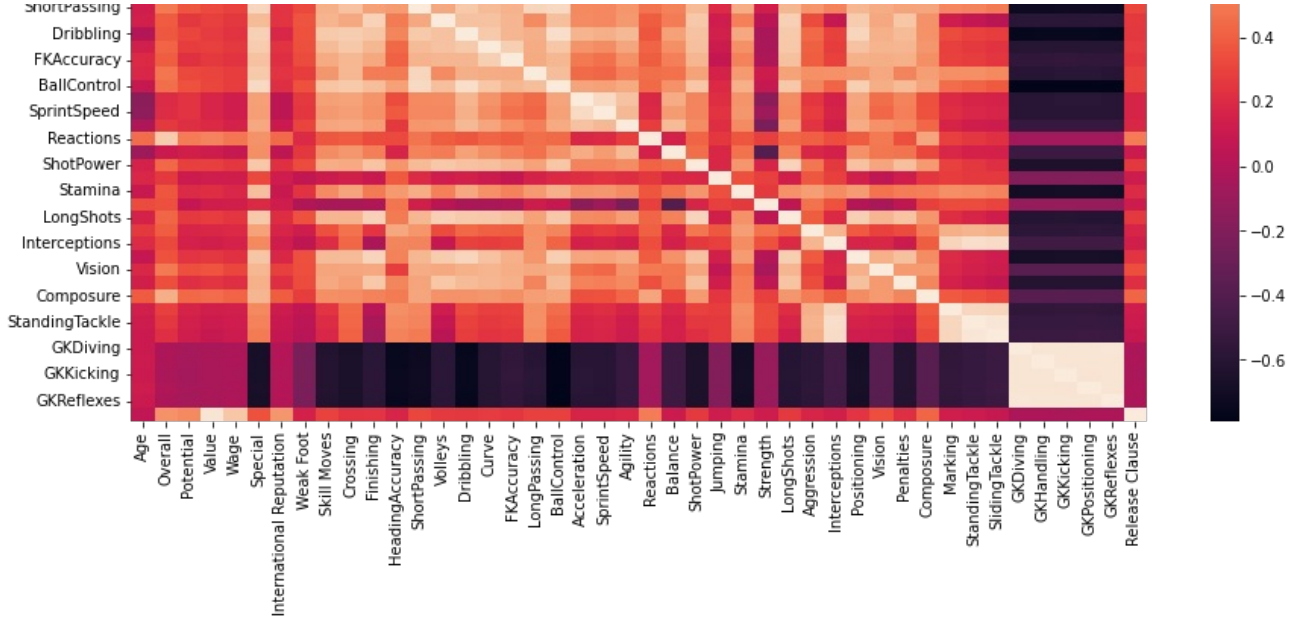
visualization

In [65]:
```
df.head()
```

Out[65]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate	
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	2202	Left	5.0	4.0	4.0	Medium/ Medium	
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	2228	Right	5.0	4.0	5.0	High/ Low	Rc
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	2143	Right	5.0	5.0	5.0	High/ Medium	N
3	De Gea	27	Spain	91	93	Manchester United	72000000.0	260000.0	1471	Right	4.0	3.0	1.0	Medium/ Medium	
4	K. De Bruyne	27	Belgium	91	92	Manchester City	102000000.0	355000.0	2281	Right	4.0	5.0	4.0	High/ High	N

In [97]:
```
plt.figure(figsize=(15,7))
sns.heatmap(df.corr())
plt.show()
```

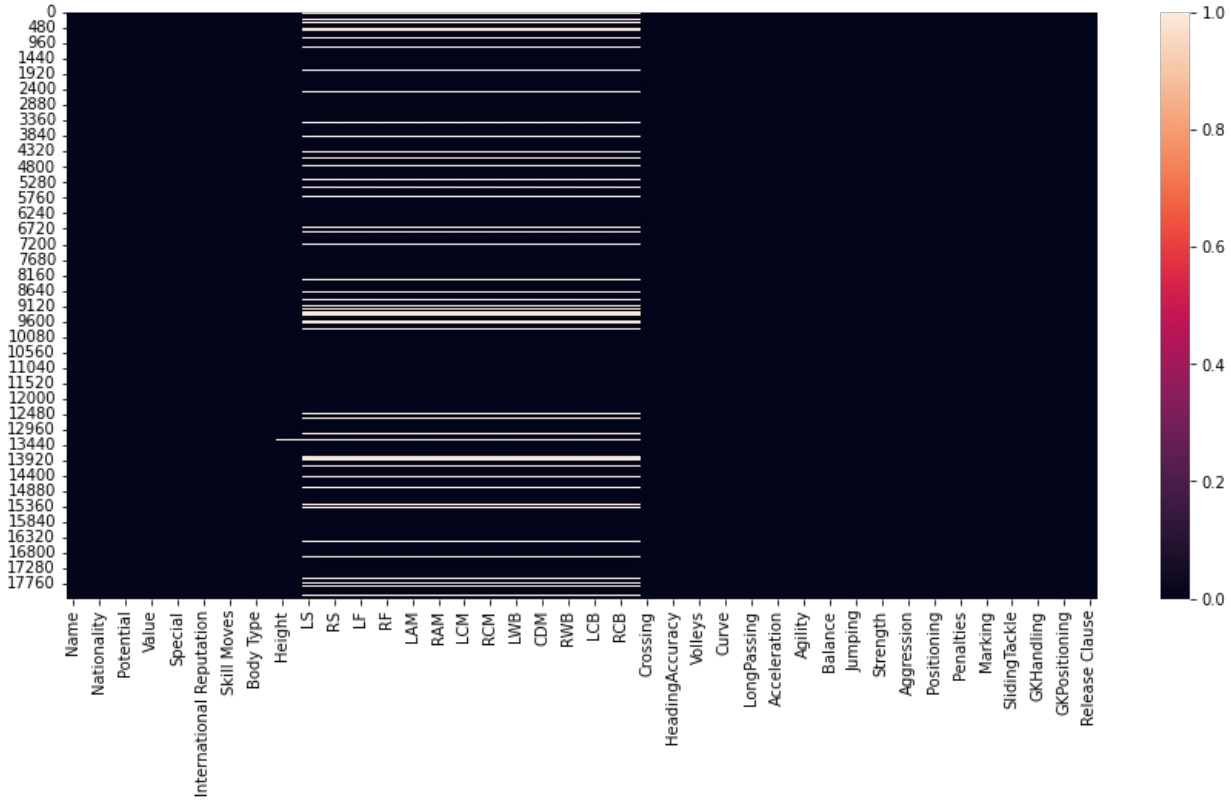


In [ ]:

```
# here we are seening most of our data is corelated
# we can easly observe that by seeing the light coloured boxex
```

In [99]:

```
plt.figure(figsize=(15,7))
sns.heatmap(df.isnull())
plt.show()
```



1. Different Clubs in the data

In [68]:

```
df.Club
```

Out[68]:

```
0               FC Barcelona
1                   Juventus
2        Paris Saint-Germain
3          Manchester United
4            Manchester City
                ...
18202        Crewe Alexandra
18203         Trelleborgs FF
```

```
18204       Cambridge United
18205        Tranmere Rovers
18206        Tranmere Rovers
Name: Club, Length: 18207, dtype: object
```

In [75]:
```
df['Club'].unique()
```

Out[75]:
```
array(['FC Barcelona', 'Juventus', 'Paris Saint-Germain',
       'Manchester United', 'Manchester City', 'Chelsea', 'Real Madrid',
       'Atlético Madrid', 'FC Bayern München', 'Tottenham Hotspur',
       'Liverpool', 'Napoli', 'Arsenal', 'Milan', 'Inter', 'Lazio',
       'Borussia Dortmund', 'Vissel Kobe', 'Olympique Lyonnais', 'Roma',
       'Valencia CF', 'Guangzhou Evergrande Taobao FC', 'FC Porto',
       'FC Schalke 04', 'Beşiktaş JK', 'LA Galaxy', 'Sporting CP',
       'Real Betis', 'Olympique de Marseille', 'RC Celta',
       'Bayer 04 Leverkusen', 'Real Sociedad', 'Villarreal CF',
       'Sevilla FC', 'SL Benfica', 'AS Saint-Étienne', 'AS Monaco',
       'Leicester City', 'Atalanta', 'Grêmio', 'Atlético Mineiro',
       'RB Leipzig', 'Ajax', 'Dalian YiFang FC', 'Everton',
       'West Ham United', '1. FC Köln', 'TSG 1899 Hoffenheim',
       'Shanghai SIPG FC', 'OGC Nice', 'Al Nassr',
       'Wolverhampton Wanderers', 'Borussia Mönchengladbach',
       'Hertha BSC', 'SV Werder Bremen', 'Cruzeiro',
       'Athletic Club de Bilbao', 'Torino', 'Medipol Başakşehir FK',
       'Beijing Sinobo Guoan FC', 'Crystal Palace', 'PFC CSKA Moscow',
       'VfL Wolfsburg', 'Shakhtar Donetsk', 'Toronto FC',
       'Lokomotiv Moscow', 'Sassuolo', 'New York City FC', 'Fluminense',
       'PSV', 'Levante UD', 'Fulham', 'Watford', 'Atlanta United',
       'Montpellier HSC', 'Galatasaray SK', 'Fenerbahçe SK', 'SD Eibar',
       'Los Angeles FC', 'Sampdoria', 'Al Hilal', 'VfB Stuttgart',
       'SC Braga', 'River Plate', 'Deportivo Alavés', 'No Club',
       'Eintracht Frankfurt', 'Girona FC', 'Guangzhou R&F; FC', 'Burnley',
       'Stoke City', 'Southampton', 'Tianjin Quanjian FC', 'Getafe CF',
       'Beijing Renhe FC', 'Montreal Impact', 'Chievo Verona', 'Genoa',
       'Portland Timbers', 'Tigres U.A.N.L.', 'RCD Espanyol',
       'Hebei China Fortune FC', 'Cagliari', 'Chicago Fire', 'DC United',
       'Sagan Tosu', 'Dynamo Kyiv', 'Santos', 'Internacional',
       'América FC (Minas Gerais)', 'Independiente', 'Boca Juniors',
       'Cruz Azul', '1. FSV Mainz 05', 'Bournemouth', 'Spartak Moscow',
       'Racing Club', 'FC Augsburg', 'Fiorentina', 'FC Nantes',
       'Feyenoord', 'Club Brugge KV', 'Brighton & Hove Albion', 'Al Ahli',
       'Jiangsu Suning FC', 'SC Freiburg', 'PAOK', 'Stade Rennais FC',
       'Trabzonspor', 'SPAL', 'Portimonense SC', 'Olympiacos CFP',
       'Club Atlético Huracán', 'Kasimpaşa SK', 'Newcastle United',
       'Frosinone', 'Querétaro', 'KRC Genk', 'Hannover 96',
       'Stade Malherbe Caen', 'Godoy Cruz', 'Toulouse Football Club',
       'RSC Anderlecht', 'Huddersfield Town', 'CD Tondela',
       'Seattle Sounders FC', 'Hamburger SV', 'FC Red Bull Salzburg',
       'Rio Ave FC', 'FC Girondins de Bordeaux', 'Melbourne Victory',
       'Parma', 'FC Basel 1893', 'Al Wehda', 'BSC Young Boys', 'KAA Gent',
       'Al Ittihad', 'Standard de Liège', 'Shanghai Greenland Shenhua FC',
       'Colo-Colo', 'Junior FC', 'West Bromwich Albion',
       'RC Strasbourg Alsace', 'Göztepe SK', 'Deportivo Cali',
       'Deportivo Toluca', 'Bologna', 'Nagoya Grampus', 'Amiens SC',
       'Changchun Yatai FC', 'Club Atlético Lanús', 'Botafogo',
       'Club América', 'Udinese', 'Real Valladolid CF', 'CD Leganés',
       'Club Atlético Banfield', 'Celtic', 'Vitória Guimarães',
       'FC København', 'UD Las Palmas', 'Deportivo de La Coruña',
       'Universidad Católica', 'San Lorenzo de Almagro', 'Rayo Vallecano',
       'Monterrey', 'Columbus Crew SC', 'MKE Ankaragücü',
       'Guizhou Hengfeng FC', 'Swansea City', 'Tianjin TEDA FC',
       'Chongqing Dangdai Lifan FC SWM Team', 'AEK Athens', 'Al Taawoun',
       'Melbourne City FC', 'En Avant de Guingamp',
       'Akhisar Belediyespor', 'Foggia', 'LOSC Lille', '1. FC Nürnberg',
       'Clube Sport Marítimo', 'Real Sporting de Gijón', 'BB Erzurumspor',
       'Shandong Luneng TaiShan FC', 'Club Atlético Colón', 'Bahia',
       'Once Caldas', 'FC Groningen', 'Angers SCO', 'Paraná',
       'Antalyaspor', 'Minnesota United FC', 'Club León', 'Empoli',
       'VVV-Venlo', 'Leeds United', 'Viktoria Plzeň', 'Alanyaspor',
       'Atlético Paranaense', 'Derby County', 'Kawasaki Frontale',
       'Cardiff City', 'Aston Villa', 'Guadalajara', 'Dijon FCO',
       'Santos Laguna', 'Málaga CF', 'Vitória', 'Çaykur Rizespor',
       'U.N.A.M.', 'Nottingham Forest', 'Royal Antwerp FC',
       'Club Tijuana', 'Sport Club do Recife', 'Real Salt Lake',
       'AZ Alkmaar', 'SK Slavia Praha', 'Willem II', 'Middlesbrough',
       'Dinamo Zagreb', 'Club Atlas', 'Granada CF', 'Sydney FC',
       'Sporting Kansas City', 'SV Zulte-Waregem', 'Philadelphia Union',
       'Real Oviedo', 'Pachuca', 'Boavista FC', 'Atiker Konyaspor',
       'Kaizer Chiefs', 'GD Chaves', 'Palermo', 'Atlético Nacional',
       'Puebla FC', 'Perth Glory', 'Panathinaikos FC', 'FC Sion',
       'Vitória de Setúbal', 'New York Red Bulls', 'Al Shabab',
       'Monarcas Morelia', 'Albacete BP', 'Rangers FC', 'Sparta Praha',
       'Legia Warszawa', 'Urawa Red Diamonds', 'Rosario Central',
       'Stade de Reims', 'ADO Den Haag', 'Chapecoense', 'FC Midtjylland',
       'San Jose Earthquakes', 'Belgrano de Córdoba', 'Brescia',
```

'Kashima Antlers', 'CD Everton de Viña del Mar',
'Fortuna Düsseldorf', 'SD Huesca', 'Preston North End',
'Club Atlético Talleres', 'Benevento', 'Vitesse',
'Gimnasia y Esgrima La Plata', 'Houston Dynamo', 'Club Necaxa',
'Norwich City', 'Holstein Kiel', 'Ettifaq FC', 'Kayserispor',
'1. FC Heidenheim 1846', 'Brentford', 'Yeni Malatyaspor',
'Lobos BUAP', 'Bursaspor', 'Ceará Sporting Club',
'Sheffield United', 'FC Ingolstadt 04', 'Estudiantes de La Plata',
'AIK', 'Queens Park Rangers', 'Suwon Samsung Bluewings',
'Heart of Midlothian', 'Reading', 'FC Dallas', 'Heracles Almelo',
'Venezia FC', 'CD Lugo', 'Henan Jianye FC', 'Orlando City SC',
'CA Osasuna', 'NAC Breda', 'Livorno', 'Universidad de Chile',
'Brøndby IF', 'Aberdeen', 'Defensa y Justicia', 'Atlético Tucumán',
'Blackburn Rovers', 'SV Darmstadt 98', 'Moreirense FC',
'Sanfrecce Hiroshima', 'CD Numancia', 'KV Oostende', 'FC Utrecht',
'Vancouver Whitecaps FC', 'Odense Boldklub', 'SC Heerenveen',
'Racing Club de Lens', 'Independiente Santa Fe',
'Sporting de Charleroi', 'Millonarios FC', 'Sheffield Wednesday',
'Perugia', 'Daegu FC', 'Vélez Sarsfield',
'Grasshopper Club Zürich', 'Sivasspor', 'Nîmes Olympique',
'Rosenborg BK', 'SK Sturm Graz', 'FC Metz',
'CD Universidad de Concepción', 'Hellas Verona', 'Brisbane Roar',
'CD Feirense', 'Hull City', 'Waasland-Beveren', 'Neuchâtel Xamax',
'Real Zaragoza', 'CD Aves', 'Millwall', 'Unión de Santa Fe',
'KAS Eupen', 'Cádiz CF', 'FC Tokyo', 'CD Tenerife',
'1. FC Union Berlin', 'Al Fayha', 'AJ Auxerre',
'Patriotas Boyacá FC', 'Molde FK', 'Bristol City', 'CD Nacional',
'Sporting Lokeren', 'FC St. Pauli', 'Deportes Iquique',
'Al Qadisiyah', 'Atlético Bucaramanga', 'Club Atlético Tigre',
'FK Austria Wien', 'Patronato', 'Malmö FF', 'Kashiwa Reysol',
'US Cremonese', 'VfL Bochum 1848', 'SK Rapid Wien',
'KSV Cercle Brugge', 'Rionegro Águilas', 'Gimnàstic de Tarragona',
'Lecce', 'Santa Clara', 'BK Häcken', 'New England Revolution',
'Orlando Pirates', 'Atlético Huila', 'Western Sydney Wanderers',
'Kalmar FF', 'Independiente Medellín', 'Fortuna Sittard',
'Lech Poznań', 'Djurgårdens IF', 'CF Reus Deportiu', 'SK Brann',
'Ulsan Hyundai FC', 'Sint-Truidense VV', 'Carpi', 'Al Fateh',
'Royal Excel Mouscron', 'AC Ajaccio', 'PEC Zwolle', 'Sunderland',
'Club Atlético Aldosivi', 'US Salernitana 1919', 'FC Lorient',
'Argentinos Juniors', 'AD Alcorcón', 'Crotone', 'Excelsior',
'KV Kortrijk', 'IFK Norrköping', 'Adelaide United',
'FC St. Gallen', 'Tiburones Rojos de Veracruz', 'CD Palestino',
'Jeju United FC', 'Deportes Tolima', 'Jeonbuk Hyundai Motors',
'Birmingham City', 'América de Cali', 'La Equidad', 'Spezia',
'Aalborg BK', 'Le Havre AC', 'Górnik Zabrze',
'Central Coast Mariners', 'Wigan Athletic',
'Jagiellonia Białystok', 'Cittadella', 'Hibernian', 'FC Lugano',
'San Martín de San Juan', 'Strømsgodset IF', 'Júbilo Iwata',
"Newell's Old Boys", 'Al Faisaly', 'Colorado Rapids',
'IF Elfsborg', 'SV Sandhausen', 'Al Batin', 'Stade Brestois 29',
'UD Almería', 'Gyeongnam FC', 'Yokohama F. Marinos', 'Kilmarnock',
'Pescara', 'Newcastle Jets', 'Córdoba CF', 'RCD Mallorca',
'Hammarby IF', 'Cerezo Osaka', 'KFC Uerdingen 05',
'Shimizu S-Pulse', 'MSV Duisburg', 'Os Belenenses',
'DSC Arminia Bielefeld', 'Ipswich Town', 'FC Seoul',
'Lechia Gdańsk', 'Gamba Osaka', 'CF Rayo Majadahonda', 'LASK Linz',
'Bolton Wanderers', 'Al Raed', 'Extremadura UD', 'SC Paderborn 07',
'Wellington Phoenix', 'Unión Española', 'Alianza Petrolera',
'Cracovia', 'Gangwon FC', 'Elche CF', 'ESTAC Troyes', 'AS Béziers',
'La Berrichonne de Châteauroux', 'Clermont Foot 63',
'1. FC Magdeburg', 'Pohang Steelers', 'Örebro SK', 'Arka Gdynia',
'SG Dynamo Dresden', 'SpVgg Greuther Fürth', 'CD Huachipato',
'Wisła Kraków', 'Stabæk Fotball', 'Eintracht Braunschweig',
'Valenciennes FC', 'FC Thun', 'San Luis de Quillota',
' SSV Jahn Regensburg', 'Cosenza', 'FC Nordsjælland',
'FC Erzgebirge Aue', 'Jeonnam Dragons', 'Wolfsberger AC',
'Chamois Niortais Football Club', 'Club Deportes Temuco',
'AS Nancy Lorraine', 'Red Star FC', 'Al Hazem', 'Pogoń Szczecin',
'Charlton Athletic', 'Grenoble Foot 38', 'FC Hansa Rostock',
'San Martin de Tucumán', 'Incheon United FC', 'Śląsk Wrocław',
'GFC Ajaccio', '1. FC Kaiserslautern', 'Deportivo Pasto',
'Lincoln City', 'Motherwell', 'Rotherham United', 'Burton Albion',
'Wisła Płock', 'FC Wacker Innsbruck', 'Peterborough United',
'Ascoli', 'FC Zürich', 'Fleetwood Town', 'Padova',
'FC Sochaux-Montbéliard', 'SV Wehen Wiesbaden', 'Unión La Calera',
'Scunthorpe United', "CD O'Higgins", 'CD Antofagasta',
'Plymouth Argyle', 'Aarhus GF', 'Lillestrøm SK', 'Karlsruher SC',
'GIF Sundsvall', 'FC Emmen', 'Barnsley', 'Audax Italiano',
'V-Varen Nagasaki', 'Paris FC', 'SpVgg Unterhaching', 'Hobro IK',
'De Graafschap', 'Hokkaido Consadole Sapporo', 'Tromsø IL',
'FC Luzern', 'FK Haugesund', 'Zagłębie Lubin', 'VfR Aalen',
'Dundalk', 'Oxford United', 'Piast Gliwice', 'Ohod Club',
'Östersunds FK', 'Vegalta Sendai', 'Crawley Town',
'FC Admira Wacker Mödling', 'Vålerenga Fotball', 'Dundee FC',
'Portsmouth', 'Envigado FC', 'Miedź Legnica', 'Odds BK',
'SC Fortuna Köln', 'US Orléans Loiret Football', 'Sarpsborg 08 FF',
'Jaguares de Córdoba', 'Bradford City', 'Accrington Stanley',
'St. Johnstone FC', 'Boyacá Chicó FC', 'Luton Town',
'SV Mattersburg', 'Kristiansund BK', 'Sangju Sangmu FC',

```
       'Rochdale', 'Walsall', 'Korona Kielce', 'Shonan Bellmare',
       'FC Würzburger Kickers', 'FSV Zwickau', 'St. Mirren', 'AC Horsens',
       'Esbjerg fB', 'HJK Helsinki', 'Southend United', 'Bristol Rovers',
       'Hamilton Academical FC', 'TSV 1860 München', 'Curicó Unido',
       'SCR Altach', 'Ranheim Fotball', 'Stevenage',
       'SG Sonnenhof Großaspach', 'Oldham Athletic', 'Milton Keynes Dons',
       'FK Bodø/Glimt', 'SC Preußen Münster', 'Wycombe Wanderers',
       'Vejle Boldklub', 'Bury', 'Randers FC', 'VfL Osnabrück',
       'SønderjyskE', 'IFK Göteborg', 'Mansfield Town', 'Coventry City',
       'Waterford FC', 'Shrewsbury', 'IK Start', 'Gillingham',
       'FC Energie Cottbus', 'FC Carl Zeiss Jena', 'Hallescher FC',
       'SV Meppen', 'AFC Wimbledon', 'Blackpool', 'Doncaster Rovers',
       'Sandefjord Fotball', 'VfL Sportfreunde Lotte', 'Cheltenham Town',
       'IK Sirius', 'Vendsyssel FF', 'Swindon Town', 'Notts County',
       'SKN St. Pölten', 'Exeter City', 'Northampton Town',
       'Shamrock Rovers', 'Colchester United', 'Livingston FC',
       'TSV Hartberg', 'Tranmere Rovers', 'Cambridge United',
       'Grimsby Town', 'Port Vale', 'Itagüí Leones FC',
       'Forest Green Rovers', 'Dalkurd FF', 'Zagłębie Sosnowiec',
       'Carlisle United', 'Trelleborgs FF', "St. Patrick's Athletic",
       'Morecambe', 'Cork City', 'IF Brommapojkarna', 'Crewe Alexandra',
       'Yeovil Town', 'Bohemian FC', 'Macclesfield Town',
       'Newport County', 'Sligo Rovers', 'Derry City', 'Limerick FC',
       'Bray Wanderers'], dtype=object)
```

In [84]:
```
print('Total no of clubs :')
df.Club.nunique()
```

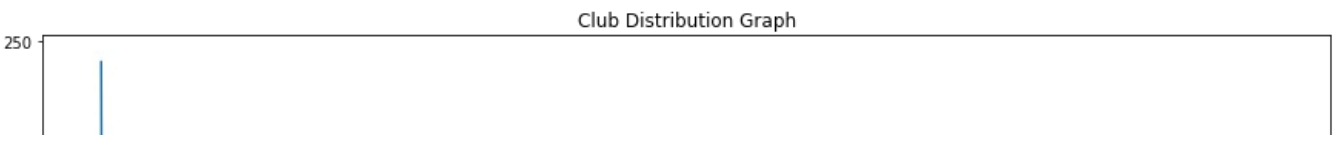
```
Total no of clubs :
652
```

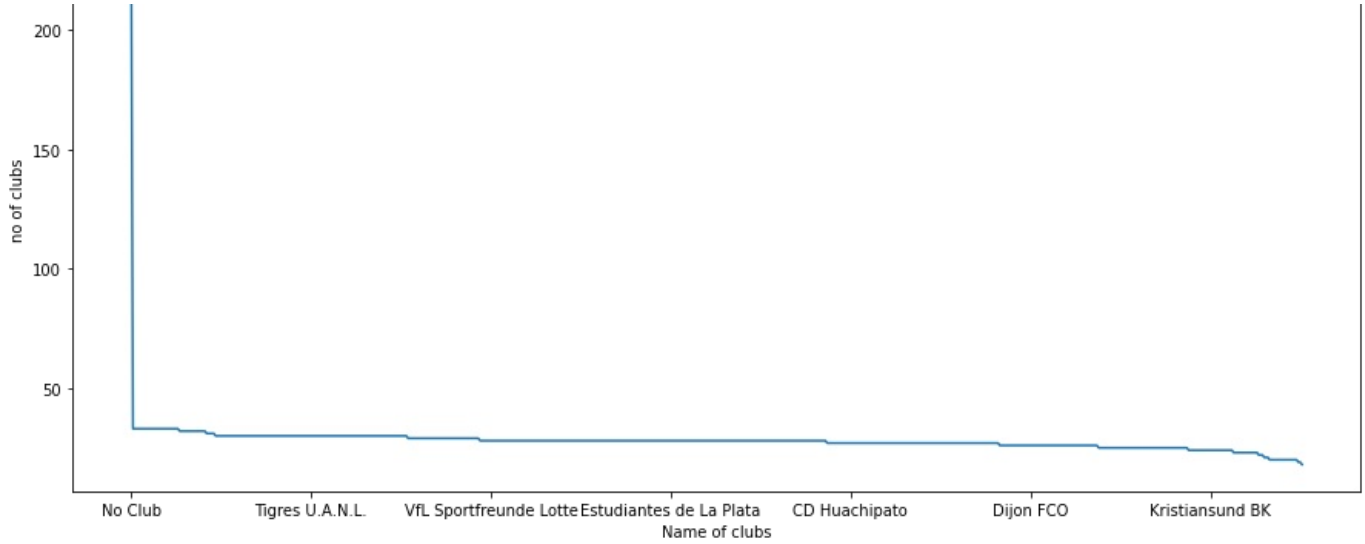
Out[84]:

In [78]:
```
df.Club.value_counts().head(30)
```

Out[78]:
```
No Club                    241
FC Barcelona                33
Burnley                     33
AS Monaco                   33
Everton                     33
TSG 1899 Hoffenheim         33
Wolverhampton Wanderers     33
Eintracht Frankfurt         33
Southampton                 33
Valencia CF                 33
Newcastle United            33
Frosinone                   33
CD Leganés                  33
Rayo Vallecano              33
Cardiff City                33
Fortuna Düsseldorf          33
RC Celta                    33
Empoli                      33
Atlético Madrid             33
Arsenal                     33
Real Madrid                 33
Tottenham Hotspur           33
Manchester United           33
Borussia Dortmund           33
Liverpool                   33
Chelsea                     33
Manchester City             33
Crystal Palace              32
SV Werder Bremen            32
Bournemouth                 32
Name: Club, dtype: int64
```

In [82]:
```
plt.figure(figsize=(15,7))
df.Club.value_counts().plot()
plt.xlabel('Name of clubs')
plt.ylabel('no of clubs')
plt.title('Club Distribution Graph')
plt.show()
```

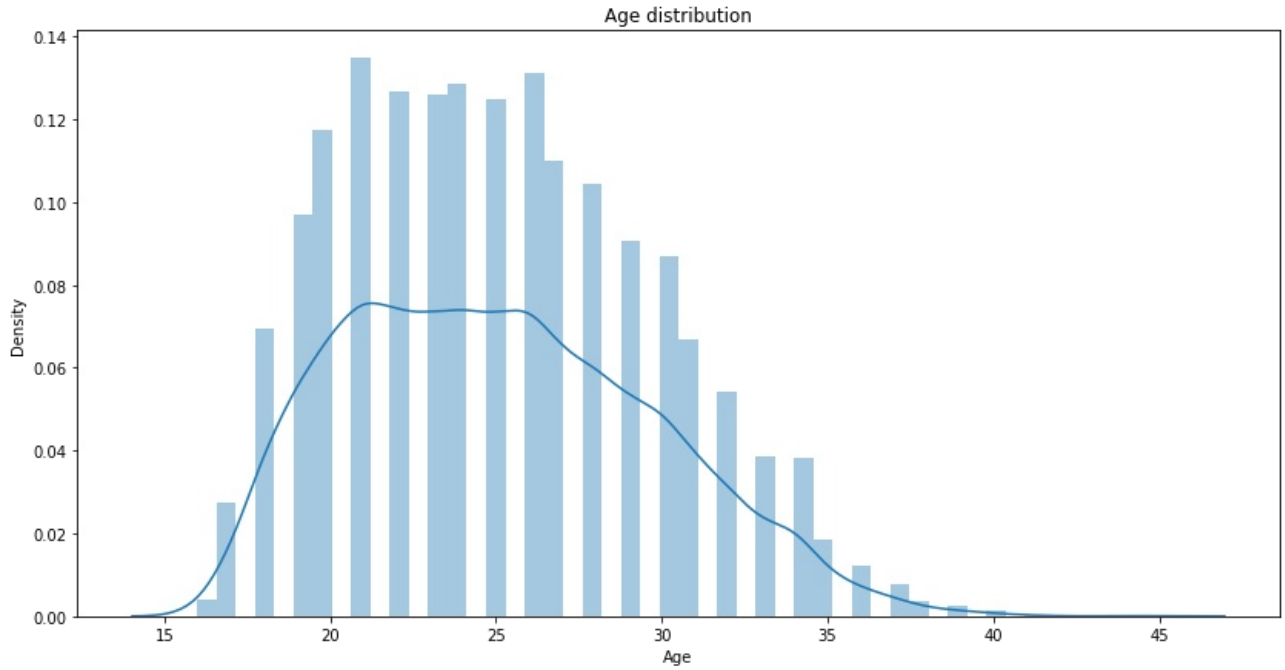




2. Distribution of the players on the bases of Age

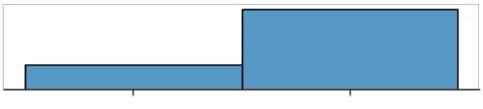
In [89]:
```
plt.figure(figsize=(14,7))
sns.distplot(x=df['Age'], bins =50)
plt.title('Age distribution')
plt.xlabel('Age')
plt.show()
```

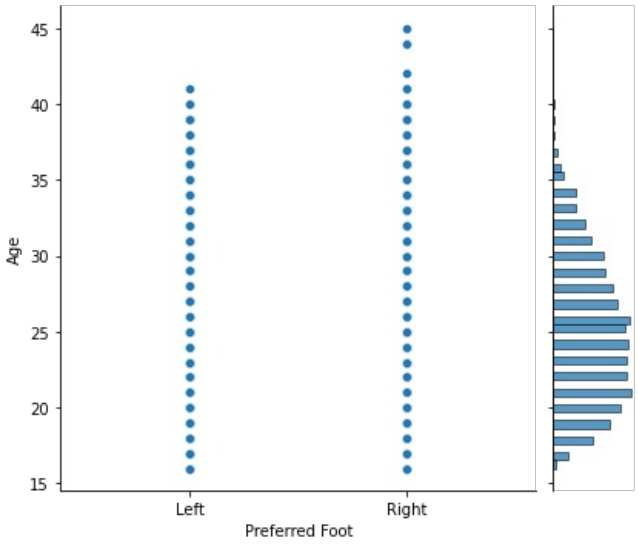
```
c:\python\python39\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated fun
ction and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level fun
ction with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



In [133…
```
plt.figure(figsize=(15,9))
sns.jointplot(df['Preferred Foot'], df.Age)
plt.show()
```

```
c:\python\python39\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as ke
yword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other argumen
ts without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
<Figure size 1080x648 with 0 Axes>
```





In [90]:
```
df.Age.max()
```

Out[90]:
```
45
```

In [91]:
```
df.Age.min()
```

Out[91]:
```
16
```

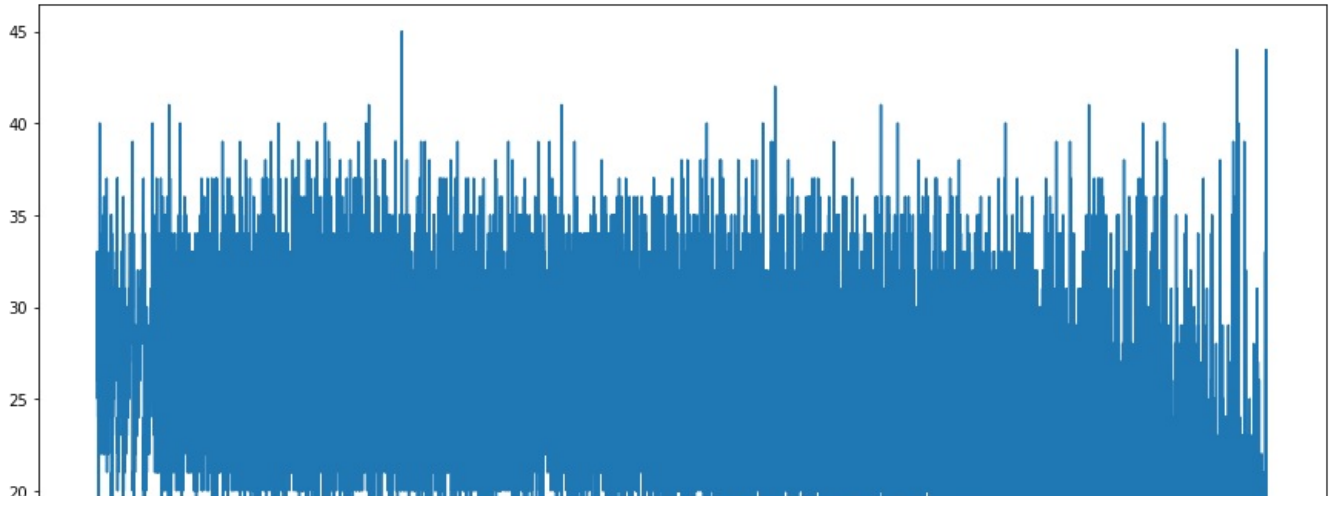
In [92]:
```
df.Age.mean()
```

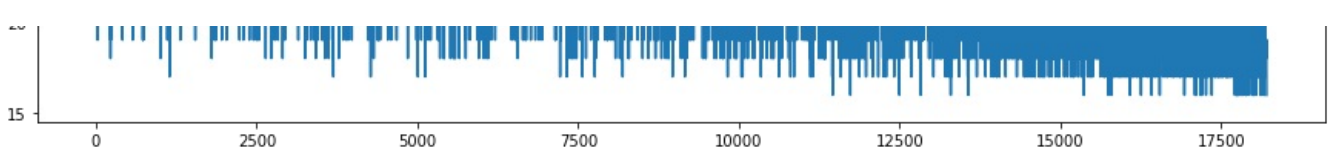
Out[92]:
```
25.122205745043114
```

In [93]:
```
df.Age.describe()
```

Out[93]:
```
count    18207.000000
mean        25.122206
std          4.669943
min         16.000000
25%         21.000000
50%         25.000000
75%         28.000000
max         45.000000
Name: Age, dtype: float64
```

In [95]:
```
plt.figure(figsize=(15,7))
df.Age.plot()
plt.show()
```





3. distribution on the bases of Nationality

In [100…
```
df.head()
```

Out[100…

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate	
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	2202	Left	5.0	4.0	4.0	Medium/ Medium	
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	2228	Right	5.0	4.0	5.0	High/ Low	Rc
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	2143	Right	5.0	5.0	5.0	High/ Medium	N
3	De Gea	27	Spain	91	93	Manchester United	72000000.0	260000.0	1471	Right	4.0	3.0	1.0	Medium/ Medium	
4	K. De Bruyne	27	Belgium	91	92	Manchester City	102000000.0	355000.0	2281	Right	4.0	5.0	4.0	High/ High	N

In [101…
```
df.Nationality.unique()
```

Out[101…
```
array(['Argentina', 'Portugal', 'Brazil', 'Spain', 'Belgium', 'Croatia',
       'Uruguay', 'Slovenia', 'Poland', 'Germany', 'France', 'England',
       'Italy', 'Egypt', 'Colombia', 'Denmark', 'Gabon', 'Wales',
       'Senegal', 'Costa Rica', 'Slovakia', 'Netherlands',
       'Bosnia Herzegovina', 'Morocco', 'Serbia', 'Algeria', 'Austria',
       'Greece', 'Chile', 'Sweden', 'Korea Republic', 'Finland', 'Guinea',
       'Montenegro', 'Armenia', 'Switzerland', 'Norway', 'Czech Republic',
       'Scotland', 'Ghana', 'Central African Rep.', 'DR Congo',
       'Ivory Coast', 'Russia', 'Ukraine', 'Iceland', 'Mexico', 'Jamaica',
       'Albania', 'Venezuela', 'Japan', 'Turkey', 'Ecuador', 'Paraguay',
       'Mali', 'Nigeria', 'Cameroon', 'Dominican Republic', 'Israel',
       'Kenya', 'Hungary', 'Republic of Ireland', 'Romania',
       'United States', 'Cape Verde', 'Australia', 'Peru', 'Togo',
       'Syria', 'Zimbabwe', 'Angola', 'Burkina Faso', 'Iran', 'Estonia',
       'Tunisia', 'Equatorial Guinea', 'New Zealand', 'FYR Macedonia',
       'United Arab Emirates', 'China PR', 'Guinea Bissau', 'Bulgaria',
       'Kosovo', 'South Africa', 'Madagascar', 'Georgia', 'Tanzania',
       'Gambia', 'Cuba', 'Belarus', 'Uzbekistan', 'Benin', 'Congo',
       'Mozambique', 'Honduras', 'Canada', 'Northern Ireland', 'Cyprus',
       'Saudi Arabia', 'Curacao', 'Moldova', 'Bolivia',
       'Trinidad & Tobago', 'Sierra Leone', 'Zambia', 'Chad',
       'Philippines', 'Haiti', 'Comoros', 'Libya', 'Panama',
       'São Tomé & Príncipe', 'Eritrea', 'Oman', 'Iraq', 'Burundi',
       'Fiji', 'New Caledonia', 'Lithuania', 'Luxembourg', 'Korea DPR',
       'Liechtenstein', 'St Kitts Nevis', 'Latvia', 'Suriname', 'Uganda',
       'El Salvador', 'Bermuda', 'Kuwait', 'Antigua & Barbuda',
       'Thailand', 'Mauritius', 'Guatemala', 'Liberia', 'Kazakhstan',
       'Niger', 'Mauritania', 'Montserrat', 'Namibia', 'Azerbaijan',
       'Guam', 'Faroe Islands', 'India', 'Nicaragua', 'Barbados',
       'Lebanon', 'Palestine', 'Guyana', 'Sudan', 'St Lucia', 'Ethiopia',
       'Puerto Rico', 'Grenada', 'Jordan', 'Rwanda', 'Qatar',
       'Afghanistan', 'Hong Kong', 'Andorra', 'Malta', 'Belize',
       'South Sudan', 'Indonesia', 'Botswana'], dtype=object)
```

In [105…
```
print('Total no. of Countries: ')
df.Nationality.nunique()
```

```
Total no. of Countries:
```

Out[105…
```
164
```

In [103…
```
df.Nationality.value_counts()
```

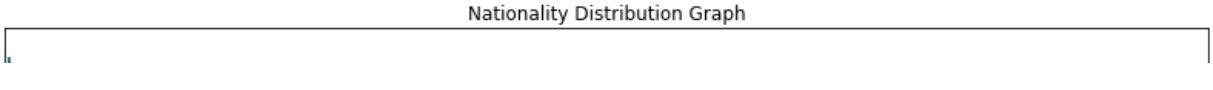
Out[103…
```
England                 1662
```

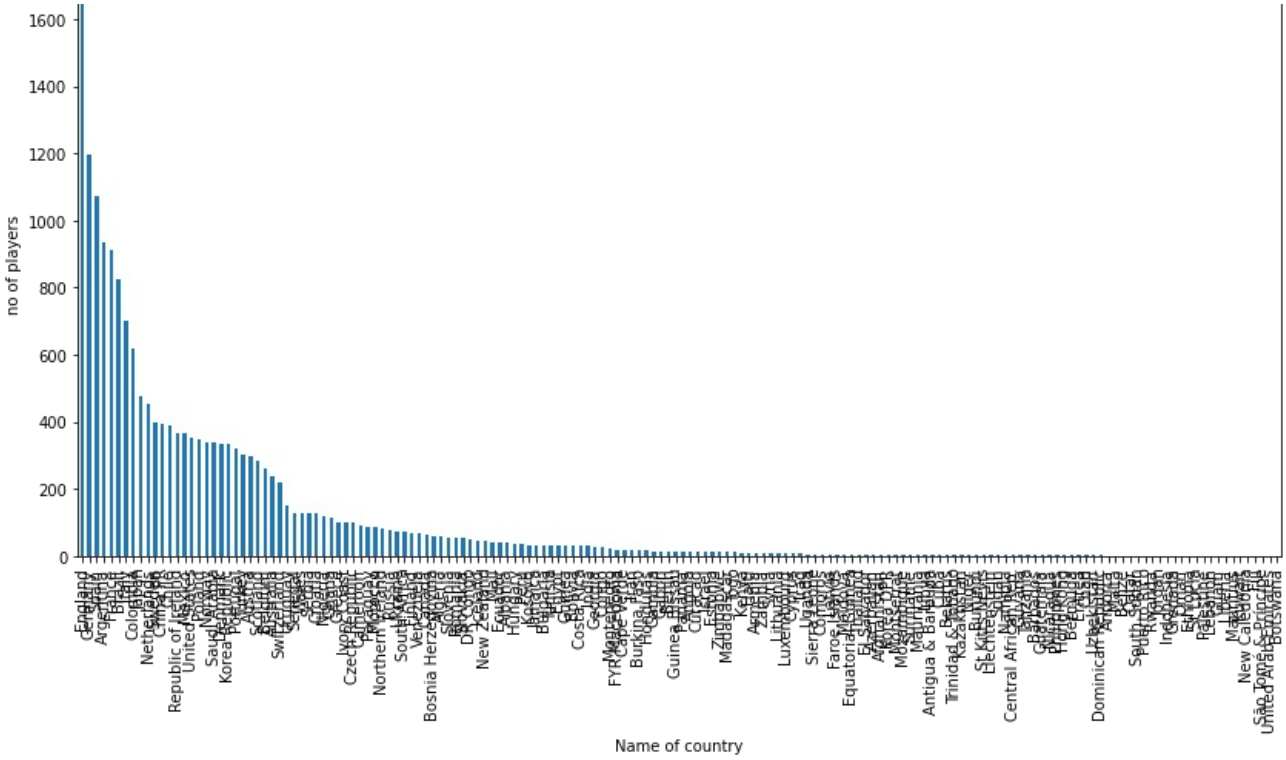
Germany	1198
Spain	1072
Argentina	937
France	914
Brazil	827
Italy	702
Colombia	618
Japan	478
Netherlands	453
Sweden	397
China PR	392
Chile	391
Republic of Ireland	368
Mexico	366
United States	353
Poland	350
Norway	341
Saudi Arabia	340
Denmark	336
Korea Republic	335
Portugal	322
Turkey	303
Austria	298
Scotland	286
Belgium	260
Australia	236
Switzerland	220
Uruguay	149
Senegal	130
Wales	129
Serbia	126
Croatia	126
Nigeria	121
Ghana	114
Greece	102
Ivory Coast	100
Czech Republic	100
Cameroon	90
Paraguay	85
Morocco	85
Northern Ireland	80
Russia	79
Ukraine	73
South Africa	71
Finland	67
Venezuela	67
Canada	64
Bosnia Herzegovina	61
Algeria	60
Slovenia	55
Romania	54
Slovakia	54
DR Congo	52
Iceland	47
New Zealand	44
Mali	43
Ecuador	43
Albania	40
Hungary	38
Peru	37
Kosovo	33
Jamaica	32
Bulgaria	32
Tunisia	32
Egypt	31
Guinea	31
Bolivia	30
Costa Rica	30
India	30
Georgia	26
Congo	25
Montenegro	23
FYR Macedonia	20
Cape Verde	19
Iran	17
Burkina Faso	16
Honduras	16
Gambia	15
Gabon	15
Benin	15
Guinea Bissau	15
Panama	15
Angola	15
Curacao	14
Israel	14
Estonia	13
Zimbabwe	13
Madagascar	12
Togo	12

```
Kenya                     10
Haiti                     10
Armenia                   10
Zambia                     9
Syria                      9
Lithuania                  8
Luxembourg                 8
Cyprus                     8
Iraq                       7
Uganda                     6
Sierra Leone               6
Comoros                    6
Latvia                     6
Faroe Islands              6
Moldova                    5
Equatorial Guinea          5
Thailand                   5
El Salvador                5
Azerbaijan                 5
Afghanistan                4
Korea DPR                  4
Montserrat                 4
Mozambique                 4
Suriname                   4
Mauritania                 4
Libya                      4
Antigua & Barbuda          4
Cuba                       4
Belarus                    4
Trinidad & Tobago          4
Kazakhstan                 4
Niger                      3
Burundi                    3
St Kitts Nevis             3
Liechtenstein              3
Sudan                      3
Namibia                    3
Central African Rep.       3
Guyana                     3
Tanzania                   3
Barbados                   3
Guatemala                  3
Nicaragua                  2
Philippines                2
Hong Kong                  2
Bermuda                    2
Eritrea                    2
Chad                       2
Uzbekistan                 2
Dominican Republic         2
Andorra                    1
Malta                      1
Belize                     1
Qatar                      1
South Sudan                1
Puerto Rico                1
Rwanda                     1
Jordan                     1
Indonesia                  1
Grenada                    1
Oman                       1
Ethiopia                   1
St Lucia                   1
Palestine                  1
Lebanon                    1
Guam                       1
Liberia                    1
Mauritius                  1
Kuwait                     1
New Caledonia              1
Fiji                       1
São Tomé & Príncipe        1
United Arab Emirates       1
Botswana                   1
Name: Nationality, dtype: int64
```

In [110…

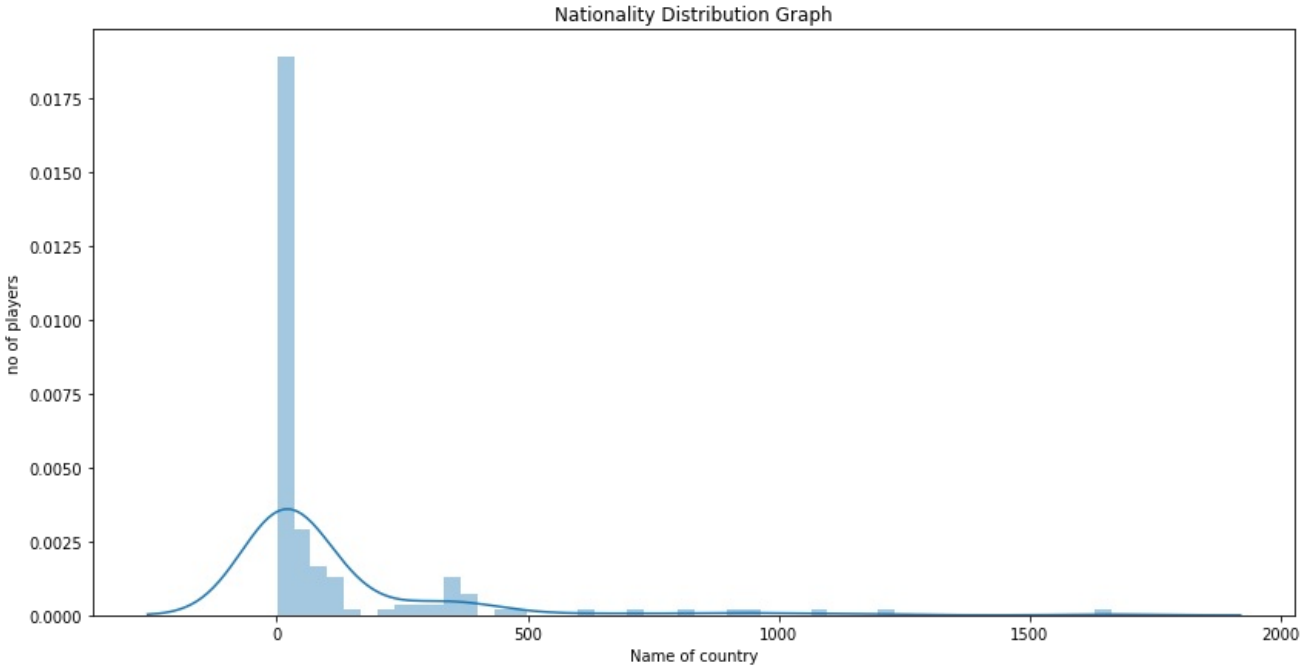
```
plt.figure(figsize=(14,7))
df.Nationality.value_counts().plot(kind = 'bar')
plt.xlabel('Name of country')
plt.ylabel('no of players')
plt.title('Nationality Distribution Graph')
plt.show()
```





In [113…

```
fig = plt.subplots(figsize=(14,7))
sns.distplot(x=df.Nationality.value_counts(), bins =50)
plt.xlabel('Name of country')
plt.ylabel('no of players')
plt.title('Nationality Distribution Graph')
plt.show()
```



4.Analysis of individual players

In [118…

```
df.sort_values(by='Overall', ascending=False).head(5).loc[:,['Name','Overall']]
```

Out[118…

	Name	Overall
0	L. Messi	94
1	Cristiano Ronaldo	94
2	Neymar Jr	92
3	De Gea	91
4	K. De Bruyne	91

In [119…
```
df.sort_values(by='Potential', ascending=False).head(5).loc[:,['Name','Potential']]
```

Out[119…

	Name	Potential
25	K. Mbappé	95
0	L. Messi	94
15	P. Dybala	94
1	Cristiano Ronaldo	94
2	Neymar Jr	93

In [120…
```
df.sort_values(by='Age', ascending=False).head(5).loc[:,['Name','Age']]
```

Out[120…

	Name	Age
4741	O. Pérez	45
18183	K. Pilkington	44
17726	T. Warner	44
10545	S. Narazaki	42
7225	C. Muñoz	41

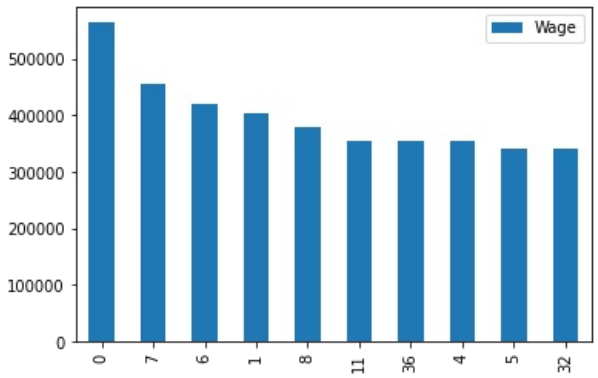
In [121…
```
df.sort_values(by='Value', ascending=False).head(5).loc[:,['Name','Value']]
```

Out[121…

	Name	Value
2	Neymar Jr	118500000.0
0	L. Messi	110500000.0
4	K. De Bruyne	102000000.0
5	E. Hazard	93000000.0
15	P. Dybala	89000000.0

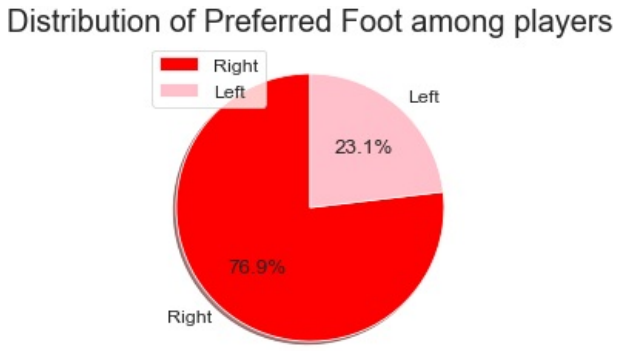
In [128…
```
df.sort_values(by='Wage', ascending=False).head(10).loc[:,['Name','Wage']].plot(kind='bar')
```

Out[128…
```
<AxesSubplot:>
```



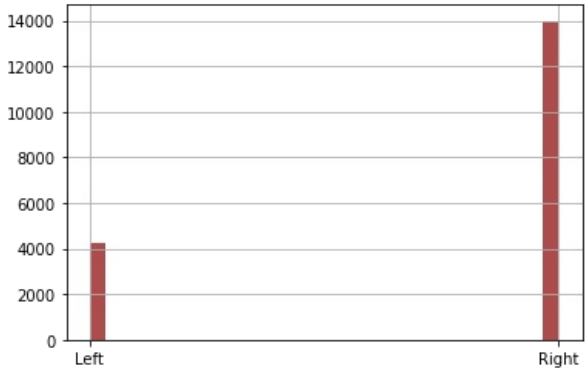
4. Preferred Foot

In [157…
```
labels = df['Preferred Foot'].value_counts().index
size = df['Preferred Foot'].value_counts()
colors=['red','pink']
plt.pie(size, labels = labels, colors = colors, shadow = True, autopct='%1.1f%%',startangle = 90)
plt.title('Distribution of Preferred Foot among players', fontsize = 20)
plt.legend()
plt.show()
```



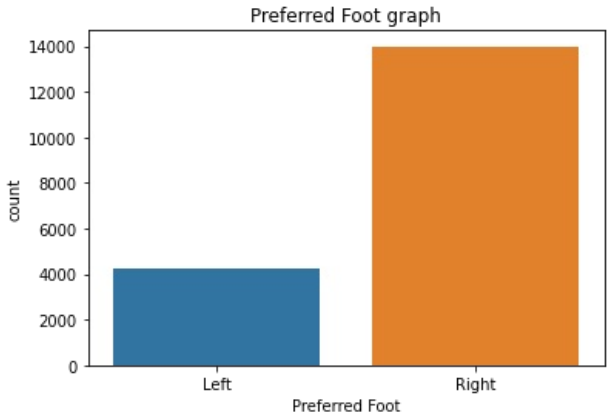
In [130…
```
df['Preferred Foot'].hist(bins = 30, color = 'darkred', alpha = 0.7)
```

Out[130…
```
<AxesSubplot:>
```



In [131…
```
sns.countplot(df['Preferred Foot'])
plt.title('Preferred Foot graph')
plt.show()
```

```
c:\python\python39\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a k
eyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



5.Visualization on the bases of position

In [134…
```
df.head()
```

Out[134…

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Special	Preferred Foot	International Reputation	Weak Foot	Skill Moves	Work Rate
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	2202	Left	5.0	4.0	4.0	Medium/ Medium
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	2228	Right	5.0	4.0	5.0	High/ Low

2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	2143	Right	5.0	5.0	5.0	High/ Medium
3	De Gea	27	Spain	91	93	Manchester United	72000000.0	260000.0	1471	Right	4.0	3.0	1.0	Medium/ Medium
4	K. De Bruyne	27	Belgium	91	92	Manchester City	102000000.0	355000.0	2281	Right	4.0	5.0	4.0	High/ High

In [135…
```
df.Position.nunique()
```

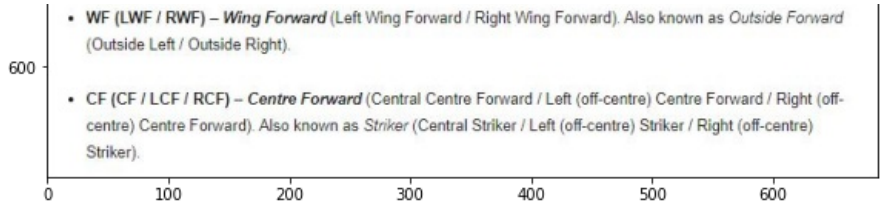
Out[135…
```
28
```

In [136…
```
df.Position.value_counts()
```

Out[136…
```
ST          2152
GK          2025
CB          1778
CM          1394
LB          1322
RB          1291
RM          1124
LM          1095
CAM          958
CDM          948
RCB          662
LCB          648
LCM          395
RCM          391
LW           381
RW           370
RDM          248
LDM          243
LS           207
RS           203
RWB           87
LWB           78
CF            74
unknown       60
RAM           21
LAM           21
RF            16
LF            15
Name: Position, dtype: int64
```

- meaning of these latters

In [145…
```
plt.figure(figsize=(25,10))
img = mpimg.imread(r'C:\Users\PC-chetan\Desktop\positons_name.JPG')
imgplot = plt.imshow(img)
```



- visual

In [146…

```
plt.figure(figsize=(25,10))
img = mpimg.imread(r'C:\Users\PC-chetan\Desktop\football_position.JPG')
imgplot = plt.imshow(img)
```

In [147…

```
Midfielders = ["CAM", "CDM", "CM", "LAM", "LCM", "LM", "LDM", "RAM", "RCM", "RDM", "RM"]
Defence = ["CB", "LB", "LCB", "LWB", "RB", "RCB", "RWB", ]
Forwards= ["CF", "LF", "LS", "LW", "RF", "RS", "RW", "ST"]
Goalkeepers = ['GK']
Not_Specific = ['unknown']
conditions = [(df['Position'].isin(Midfielders)),
              (df['Position'].isin(Defence)),
              (df['Position'].isin(Forwards)),
              (df['Position'].isin(Goalkeepers)),
              (df['Position'].isin(Not_Specific))]
values = ['Midfielders', 'Defence', 'Forwards', 'Goalkeepers','Not Specific']
df['Category'] = np.select(conditions, values)
```

In [148…

```
df['Category'].value_counts()
```

Out[148…

```
Midfielders     6838
Defence         5866
Forwards        3418
Goalkeepers     2025
Not Specific      60
Name: Category, dtype: int64
```

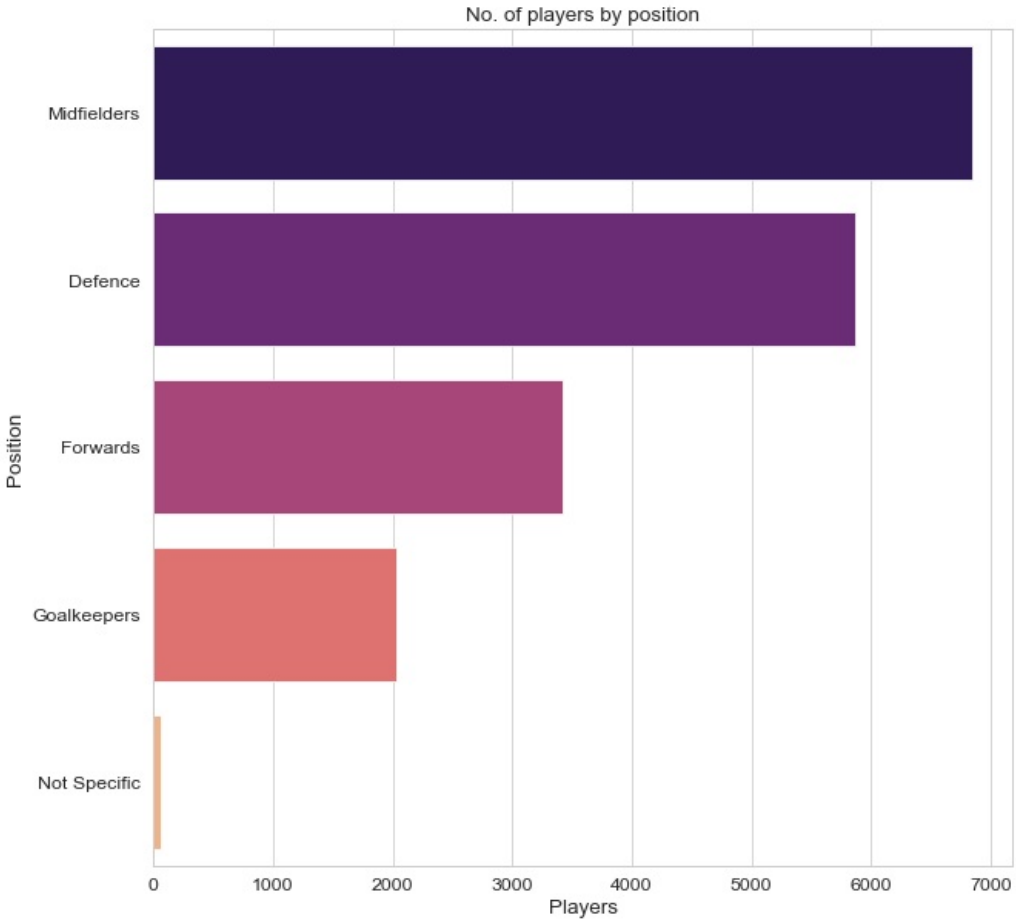
In [151…

```
sns.set_style('whitegrid')
sns.set_context('paper', font_scale=1.4)
cat_count = df['Category'].value_counts()
```

```
plt.figure(figsize=(10, 10))
sns.barplot(cat_count.values, cat_count.index, palette="magma")
plt.title('No. of players by position')
plt.ylabel('Position')
plt.xlabel('Players')

plt.show()
```

```
c:\python\python39\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as ke
yword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other argumen
ts without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

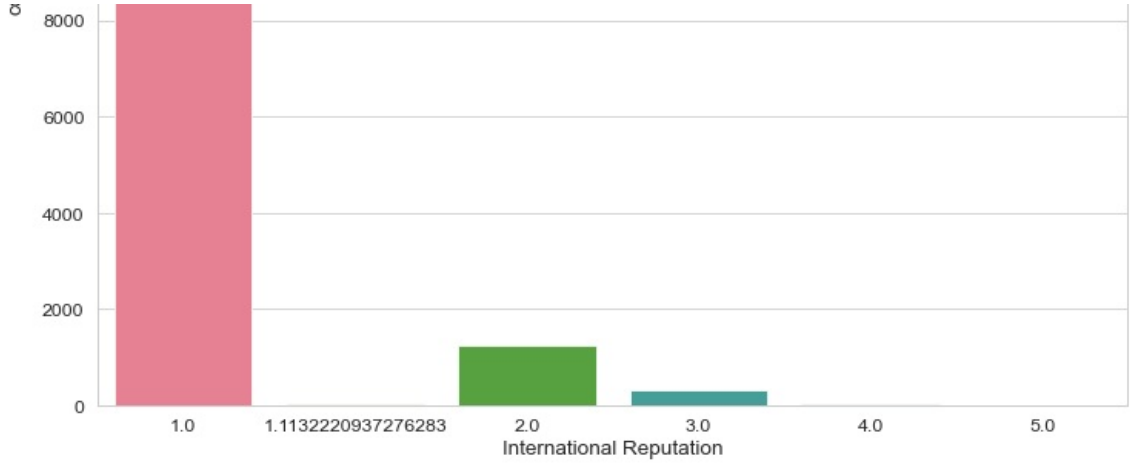


6. International Reputation

In [158…

```
plt.figure(figsize=(12,10))
sns.countplot(df['International Reputation'], palette = 'husl')
plt.title('International Reputation of the Players', fontsize = 20)
plt.show()
```

```
c:\python\python39\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a k
eyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

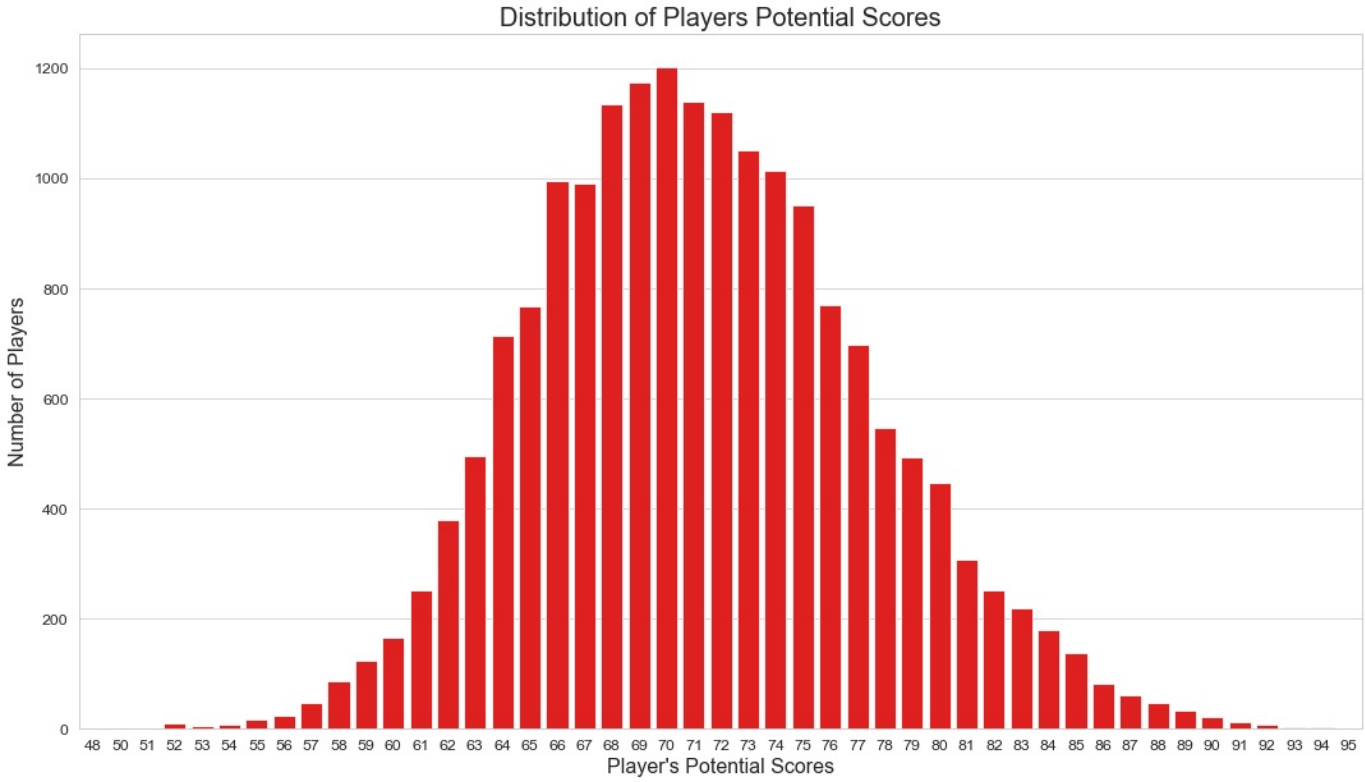


7. Visualization of Potential Score by players

In [162…

```
x = df['Potential']
plt.figure(figsize=(18,10))
ax = sns.countplot(x, color='r')
ax.set_xlabel(xlabel = "Player's Potential Scores", fontsize = 16)
ax.set_ylabel(ylabel = 'Number of Players', fontsize = 16)
ax.set_title(label = 'Distribution of Players Potential Scores', fontsize = 20)
plt.show()
```

```
c:\python\python39\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a k
eyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



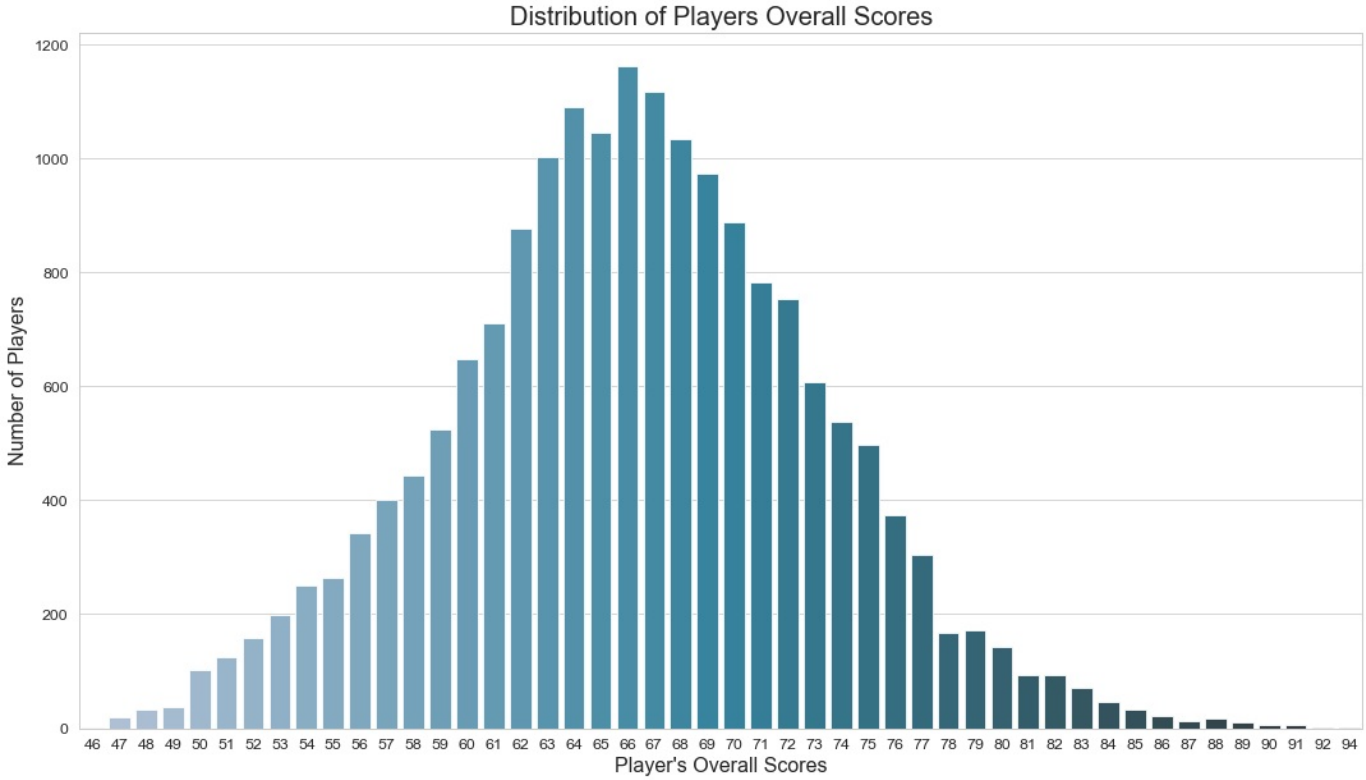
8. Visualization of Overall Score by players

In [169…

```
x = df['Overall']
plt.figure(figsize=(18,10))
ax = sns.countplot(x, palette='PuBuGn_d')
ax.set_xlabel(xlabel = "Player's Overall Scores", fontsize = 16)
ax.set_ylabel(ylabel = 'Number of Players', fontsize = 16)
ax.set_title(label = 'Distribution of Players Overall Scores', fontsize = 20)
```

```
plt.show()
```

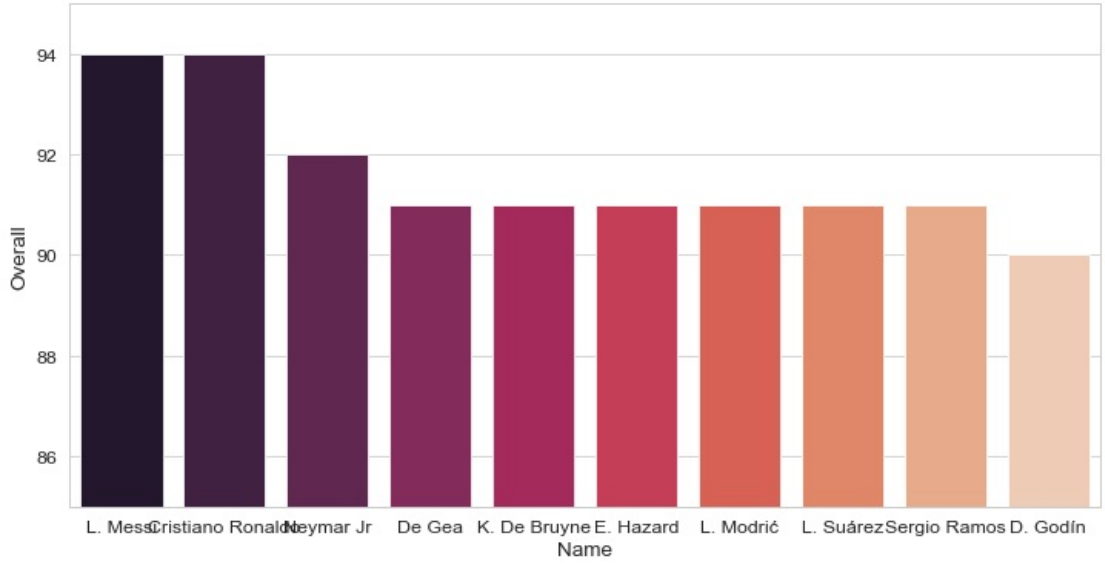
```
c:\python\python39\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a k
eyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



9. Visualization of top players

In [170…

```
df_best_players = pd.DataFrame.copy(df.sort_values(by ='Overall',ascending = False ).head(10))
plt.figure(1,figsize = (12,6))
sns.barplot(x ='Name' , y = 'Overall' , data = df_best_players, palette='rocket')
plt.ylim(85,95)
plt.show()
```



Step 6

- data analysis

> Best Finisher

> Best Dribbler
>
> Fastest Players
>
> Best Penalty Kick Taker
>
> Best players with the Ball Control
>
> Best Freekick Takers
>
> Youngest Players
>
> Eldest Players

In [173…
```
df.columns
```

Out[173…
```
Index(['Name', 'Age', 'Nationality', 'Overall', 'Potential', 'Club', 'Value',
       'Wage', 'Special', 'Preferred Foot', 'International Reputation',
       'Weak Foot', 'Skill Moves', 'Work Rate', 'Body Type', 'Position',
       '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', 'GKDiving', 'GKHandling',
       'GKKicking', 'GKPositioning', 'GKReflexes', 'Release Clause',
       'Category'],
      dtype='object')
```

Data Analysis

- Best Finisher

In [172…
```
df.sort_values(by = 'Finishing' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']].
```

Out[172…

	Name	Club	Nationality	Overall	Age	Finishing
0	L. Messi	FC Barcelona	Argentina	94	31	95.0
16	H. Kane	Tottenham Hotspur	England	89	24	94.0
1	Cristiano Ronaldo	Juventus	Portugal	94	33	94.0
7	L. Suárez	FC Barcelona	Uruguay	91	31	93.0
23	S. Agüero	Manchester City	Argentina	89	30	93.0
38	G. Higuaín	Milan	Argentina	88	30	92.0
10	R. Lewandowski	FC Bayern München	Poland	90	29	91.0
43	M. Icardi	Inter	Argentina	87	25	91.0
26	M. Salah	Liverpool	Egypt	88	26	90.0
17	A. Griezmann	Atlético Madrid	France	89	27	90.0
48	C. Immobile	Lazio	Italy	87	28	90.0
98	E. Džeko	Roma	Bosnia Herzegovina	85	32	89.0
21	E. Cavani	Paris Saint-Germain	Uruguay	89	31	89.0
87	A. Lacazette	Arsenal	France	85	27	88.0
252	A. Belotti	Torino	Italy	82	24	88.0
350	A. Milik	Napoli	Poland	81	24	88.0
33	P. Aubameyang	Arsenal	Gabon	88	29	88.0
25	K. Mbappé	Paris Saint-Germain	France	88	19	88.0
145	Jonas	SL Benfica	Brazil	84	34	88.0
130	Iago Aspas	RC Celta	Spain	84	30	88.0

- Best Dribbler

In [175…
```
df.sort_values(by = 'Dribbling' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']].
```

Out[175…

	Name	Club	Nationality	Overall	Age	Finishing
0	L. Messi	FC Barcelona	Argentina	94	31	95.0
2	Neymar Jr	Paris Saint-Germain	Brazil	92	26	87.0
5	E. Hazard	Chelsea	Belgium	91	27	84.0
30	Isco	Real Madrid	Spain	88	26	79.0
94	Y. Brahimi	FC Porto	Algeria	85	28	76.0
56	Bernardo Silva	Manchester City	Portugal	86	23	75.0
65	Douglas Costa	Juventus	Brazil	86	27	67.0
15	P. Dybala	Juventus	Argentina	89	24	84.0
32	Coutinho	FC Barcelona	Brazil	88	26	79.0
84	R. Mahrez	Manchester City	Algeria	85	27	80.0
371	J. Corona	FC Porto	Mexico	81	25	76.0
50	D. Mertens	Napoli	Belgium	87	31	86.0
185	Felipe Anderson	West Ham United	Brazil	83	25	71.0
67	Thiago	FC Bayern München	Spain	86	27	69.0
155	O. Dembélé	FC Barcelona	France	83	21	78.0
76	Iniesta	Vissel Kobe	Spain	86	34	70.0
270	W. Zaha	Crystal Palace	Ivory Coast	82	25	77.0
6	L. Modrić	Real Madrid	Croatia	91	32	72.0
129	J. Cuadrado	Juventus	Colombia	84	30	67.0
63	M. Verratti	Paris Saint-Germain	Italy	86	25	60.0

- Fastest Players

In [176…
```
df.sort_values(by = 'Dribbling' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']].
```

Out[176…

	Name	Club	Nationality	Overall	Age	Finishing
0	L. Messi	FC Barcelona	Argentina	94	31	95.0
2	Neymar Jr	Paris Saint-Germain	Brazil	92	26	87.0
5	E. Hazard	Chelsea	Belgium	91	27	84.0
30	Isco	Real Madrid	Spain	88	26	79.0
94	Y. Brahimi	FC Porto	Algeria	85	28	76.0
56	Bernardo Silva	Manchester City	Portugal	86	23	75.0
65	Douglas Costa	Juventus	Brazil	86	27	67.0
15	P. Dybala	Juventus	Argentina	89	24	84.0
32	Coutinho	FC Barcelona	Brazil	88	26	79.0
84	R. Mahrez	Manchester City	Algeria	85	27	80.0
371	J. Corona	FC Porto	Mexico	81	25	76.0
50	D. Mertens	Napoli	Belgium	87	31	86.0
185	Felipe Anderson	West Ham United	Brazil	83	25	71.0
67	Thiago	FC Bayern München	Spain	86	27	69.0
155	O. Dembélé	FC Barcelona	France	83	21	78.0
76	Iniesta	Vissel Kobe	Spain	86	34	70.0
270	W. Zaha	Crystal Palace	Ivory Coast	82	25	77.0
6	L. Modrić	Real Madrid	Croatia	91	32	72.0
129	J. Cuadrado	Juventus	Colombia	84	30	67.0
63	M. Verratti	Paris Saint-Germain	Italy	86	25	60.0

- Best Penalty Kick Taker

In [177…
```
df.sort_values(by = 'Penalties' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']].
```

Out[177…

	Name	Club	Nationality	Overall	Age	Finishing
206	M. Balotelli	OGC Nice	Italy	83	27	85.0
118	Fabinho	Liverpool	Brazil	84	24	66.0
16	H. Kane	Tottenham Hotspur	England	89	24	94.0
823	R. Jiménez	Wolverhampton Wanderers	Mexico	78	27	76.0
945	L. Baines	Everton	England	77	33	63.0
507	R. Boudebouz	Real Betis	Algeria	80	28	77.0
384	D. Perotti	Roma	Argentina	81	29	66.0
297	M. Kruse	SV Werder Bremen	Germany	82	30	83.0
109	Z. Ibrahimović	LA Galaxy	Sweden	85	36	86.0
3199	Edinho	CD Feirense	Portugal	73	35	76.0
68	M. Reus	Borussia Dortmund	Germany	86	29	87.0
1274	M. Noble	West Ham United	England	76	31	60.0
3161	M. Jedinák	Aston Villa	Australia	73	33	59.0
315	David Villa	New York City FC	Spain	82	36	86.0
10	R. Lewandowski	FC Bayern München	Poland	90	29	91.0
204	B. Dost	Sporting CP	Netherlands	83	29	87.0
240	A. Kramarić	TSG 1899 Hoffenheim	Croatia	82	27	86.0
456	S. Haller	Eintracht Frankfurt	France	80	24	84.0
2039	D. Ba	Shanghai Greenland Shenhua FC	Senegal	75	33	74.0
610	L. Milivojević	Crystal Palace	Serbia	79	27	54.0

- Best players with the Ball Control

In [178…
```
df.sort_values(by = 'BallControl' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']
```

Out[178…

	Name	Club	Nationality	Overall	Age	Finishing
0	L. Messi	FC Barcelona	Argentina	94	31	95.0
2	Neymar Jr	Paris Saint-Germain	Brazil	92	26	87.0
30	Isco	Real Madrid	Spain	88	26	79.0
13	David Silva	Manchester City	Spain	90	32	76.0
5	E. Hazard	Chelsea	Belgium	91	27	84.0
1	Cristiano Ronaldo	Juventus	Portugal	94	33	94.0
29	L. Insigne	Napoli	Italy	88	27	77.0
6	L. Modrić	Real Madrid	Croatia	91	32	72.0
76	Iniesta	Vissel Kobe	Spain	86	34	70.0
35	Marcelo	Real Madrid	Brazil	88	30	70.0
15	P. Dybala	Juventus	Argentina	89	24	84.0
32	Coutinho	FC Barcelona	Brazil	88	26	79.0
31	C. Eriksen	Tottenham Hotspur	Denmark	88	26	80.0
4	K. De Bruyne	Manchester City	Belgium	91	27	82.0
25	K. Mbappé	Paris Saint-Germain	France	88	19	88.0
56	Bernardo Silva	Manchester City	Portugal	86	23	75.0
65	Douglas Costa	Juventus	Brazil	86	27	67.0
280	M. Götze	Borussia Dortmund	Germany	82	26	75.0
11	T. Kroos	Real Madrid	Germany	90	28	76.0
28	J. Rodríguez	FC Bayern München	Colombia	88	26	83.0

- Best Freekick Takers

In [179…
```
df.sort_values(by = 'FKAccuracy' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']]
```

Out[179…

	Name	Club	Nationality	Overall	Age	Finishing
0	L. Messi	FC Barcelona	Argentina	94	31	95.0
293	S. Giovinco	Toronto FC	Italy	82	31	80.0
72	M. Pjanić	Juventus	Bosnia Herzegovina	86	28	60.0
1113	E. Bardhi	Levante UD	FYR Macedonia	77	22	78.0
449	H. Çalhanoğlu	Milan	Turkey	80	24	65.0
90	Parejo	Valencia CF	Spain	85	29	79.0
3550	S. Nakamura	Júbilo Iwata	Japan	72	40	65.0
15	P. Dybala	Juventus	Argentina	89	24	84.0
124	M. Depay	Olympique Lyonnais	Netherlands	84	24	79.0
2	Neymar Jr	Paris Saint-Germain	Brazil	92	26	87.0
166	Anderson Talisca	Guangzhou Evergrande Taobao FC	Brazil	83	24	80.0
932	R. Malinovskyi	KRC Genk	Ukraine	77	25	69.0
36	G. Bale	Real Madrid	Wales	88	28	86.0
31	C. Eriksen	Tottenham Hotspur	Denmark	88	26	80.0
211	Juan Mata	Manchester United	Spain	83	30	78.0
291	A. Kolarov	Roma	Serbia	82	32	55.0
292	G. Sigurðsson	Everton	Iceland	82	28	77.0
200	Marcos Alonso	Chelsea	Spain	83	27	68.0
873	K. Honda	Melbourne Victory	Japan	78	32	65.0
4470	Y. Endo	Gamba Osaka	Japan	71	38	67.0

- Youngest Players

In [180…
```
df.sort_values(by = 'Age' , ascending = True)[['Name','Club','Nationality','Overall', 'Age','Finishing']].head(20
```

Out[180…

	Name	Club	Nationality	Overall	Age	Finishing
18206	G. Nugent	Tranmere Rovers	England	46	16	34.0
17743	J. Olstad	Sarpsborg 08 FF	Norway	52	16	53.0
13293	H. Massengo	AS Monaco	France	62	16	39.0
16081	J. Italiano	Perth Glory	Australia	58	16	50.0
18166	N. Ayéva	Örebro SK	Sweden	48	16	47.0
17712	K. Broda	Wisła Kraków	Poland	53	16	8.0
17354	L. D'Arrigo	Adelaide United	Australia	54	16	40.0
13567	Y. Verschaeren	RSC Anderlecht	Belgium	62	16	54.0
16927	B. Nygren	IFK Göteborg	Sweden	56	16	60.0
18018	B. O'Gorman	Bray Wanderers	Republic of Ireland	50	16	48.0
17091	A. Doğan	Kayserispor	Turkey	55	16	22.0
17115	C. Bassett	Colorado Rapids	United States	55	16	44.0
15793	J. Lahne	IF Brommapojkarna	Sweden	59	16	57.0
15746	Y. Begraoui	AJ Auxerre	France	59	16	68.0
17175	B. Mumba	Sunderland	England	55	16	33.0
18003	J. Imbrechts	Sporting de Charleroi	Sweden	51	16	4.0
18106	J. Cleary	Bohemian FC	Republic of Ireland	50	16	7.0
17177	R. Gómez	Belgrano de Córdoba	Argentina	55	16	53.0
17200	H. Andersson	Örebro SK	Sweden	55	16	24.0
17976	M. Larsen	Esbjerg fB	Denmark	51	16	42.0

- Eldest Players

In [181…
```
df.sort_values(by = 'Age' , ascending = False)[['Name','Club','Nationality','Overall', 'Age','Finishing']].head(2
```

Out[181…

	Name	Club	Nationality	Overall	Age	Finishing
4741	O. Pérez	Pachuca	Mexico	71	45	14.0
18183	K. Pilkington	Cambridge United	England	48	44	12.0
17726	T. Warner	Accrington Stanley	Trinidad & Tobago	53	44	13.0
10545	S. Narazaki	Nagoya Grampus	Japan	65	42	11.0
7225	C. Muñoz	CD Universidad de Concepción	Argentina	68	41	15.0
1120	J. Villar	No Club	Paraguay	77	41	12.0
12192	H. Sulaimani	Ohod Club	Saudi Arabia	63	41	43.0
15426	M. Tyler	Peterborough United	England	59	41	8.0
4228	B. Nivet	ESTAC Troyes	France	71	41	69.0
10356	F. Kippe	Lillestrøm SK	Norway	65	40	46.0
16264	P. van der Vlag	FC Emmen	Netherlands	58	40	11.0
9484	B. Castillo	Atlético Huila	Colombia	66	40	10.0
4187	C. Lucchetti	Atlético Tucumán	Argentina	71	40	13.0
2821	S. Bertoli	Patronato	Argentina	73	40	19.0
3550	S. Nakamura	Júbilo Iwata	Japan	72	40	65.0
1294	A. Bizzarri	Foggia	Argentina	76	40	17.0
14129	Y. Nakazawa	Yokohama F. Marinos	Japan	61	40	40.0
41	G. Buffon	Paris Saint-Germain	Italy	88	40	15.0
864	Hilton	Montpellier HSC	Brazil	78	40	45.0
12453	W. Díaz	Jaguares de Córdoba	Colombia	63	40	19.0

--------End--------