```
"""
Project-1
Apply
logistic Regression
SVM
Decision Tree
RandomForest
on the Loan dataset and check were you will get the best possible
accuracy
project note : Dependent Variable is Loan Status
"""

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import svm
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report
df=pd.read_csv("E:\loan.csv")
df

      Loan_ID  Gender Married Dependents     Education
Self_Employed  \
0    LP001002    Male      No          0      Graduate            No

1    LP001003    Male     Yes          1      Graduate            No

2    LP001005    Male     Yes          0      Graduate           Yes

3    LP001006    Male     Yes          0  Not Graduate            No

4    LP001008    Male      No          0      Graduate            No

..        ...     ...     ...        ...           ...           ...

609  LP002978  Female      No          0      Graduate            No

610  LP002979    Male     Yes         3+      Graduate            No

611  LP002983    Male     Yes          1      Graduate            No

612  LP002984    Male     Yes          2      Graduate            No

613  LP002990  Female      No          0      Graduate           Yes


     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term
\
0               5849                0.0         NaN             360.0
```

```
1                4583             1508.0       128.0             360.0

2                3000                0.0        66.0             360.0

3                2583             2358.0       120.0             360.0

4                6000                0.0       141.0             360.0

..                ...                ...         ...               ...

609              2900                0.0        71.0             360.0

610              4106                0.0        40.0             180.0

611              8072              240.0       253.0             360.0

612              7583                0.0       187.0             360.0

613              4583                0.0       133.0             360.0


     Credit_History Property_Area Loan_Status
0               1.0         Urban           Y
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[614 rows x 13 columns]

df.info()
df.isnull().sum()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Loan_ID            614 non-null    object
 1   Gender             601 non-null    object
 2   Married            611 non-null    object
 3   Dependents         599 non-null    object
```

```
 4   Education          614 non-null    object
 5   Self_Employed      582 non-null    object
 6   ApplicantIncome    614 non-null    int64
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         592 non-null    float64
 9   Loan_Amount_Term   600 non-null    float64
 10  Credit_History     564 non-null    float64
 11  Property_Area      614 non-null    object
 12  Loan_Status        614 non-null    object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
```

```
Loan_ID               0
Gender               13
Married               3
Dependents           15
Education             0
Self_Employed        32
ApplicantIncome       0
CoapplicantIncome     0
LoanAmount           22
Loan_Amount_Term     14
Credit_History       50
Property_Area         0
Loan_Status           0
dtype: int64
```

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

```
Male      489
Female    112
Name: Gender, dtype: int64
```

```
df['Dependents'] = df['Dependents'].fillna('0')
df['Dependents'] = df['Dependents'].replace({'3+':3})
```

```
np.mean(df['Loan_Amount_Term'])
```

```
342.0
```

```
df['Loan_Amount_Term'] = df['Loan_Amount_Term'].fillna(342.0)
df['Loan_Amount_Term'].isnull().sum()
```

```
0
```

```
df['Credit_History'] = df['Credit_History'].fillna(1.0)
df['LoanAmount'] = df['LoanAmount'].fillna(146.4)
df.isnull().sum()
```

```
Loan_ID               0
Gender               13
Married               3
```

```
Dependents            0
Education             0
Self_Employed        32
ApplicantIncome       0
CoapplicantIncome     0
LoanAmount            0
Loan_Amount_Term      0
Credit_History        0
Property_Area         0
Loan_Status           0
dtype: int64
```

```
df=df.fillna(value=0)
```

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

```
Semiurban    233
Urban        202
Rural        179
Name: Property_Area, dtype: int64
```

```
df['Gender'] = df['Gender'].replace({'Male':0, 'Female':1,'unknown' :
2})
```

```
df['Married'] = df['Married'].replace({'Yes' :1, 'No': 0,
'unknown':2})
```

```
df['Education'] = df['Education'].replace ({'Graduate' : 1, 'Not
Graduate' : 0})
df['Self_Employed'] = df['Self_Employed'].replace ({'Yes': 1,'No' : 0,
'unknown':2})

df['Property_Area'] = df['Property_Area'].replace ({'Semiurban':
1,'Urban' : 0, 'Rural':2})

df['Loan_Status'] =df['Loan_Status'].replace({'Y':1, 'N':0})

mylist_train = ['Gender', 'Married', 'Dependents', 'Education',
'Self_Employed', 'Credit_History', 'Property_Area' , 'Loan_Status']

for i in mylist_train:
    df[i] = df[i].astype({i:'category'})

df['Loan_Status'].value_counts(normalize=True)*100
```

```
1    68.729642
0    31.270358
Name: Loan_Status, dtype: float64
```

```
df= df.drop(columns=['Loan_ID'])
```

```
X_train = df.drop(columns=['Loan_Status'])

y_train = df['Loan_Status']

from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()

lr.fit(X_train,y_train)
lr.score(X_train,y_train)
```

```
C:\Users\Renuka\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:814: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\Renuka\anaconda3\lib\site-packages\sklearn\base.py:566:
FutureWarning: Arrays of bytes/strings is being converted to decimal
numbers if dtype='numeric'. This behavior is deprecated in 0.24 and
will be removed in 1.1 (renaming of 0.26). Please convert your data to
numeric values explicitly instead.
  X = check_array(X, **check_params)
```

```
0.8094462540716613
```

```
from sklearn import svm
from sklearn.model_selection import train_test_split
x = df.iloc[:, :-2]
y = df.iloc[:, -1]
x_train, x_test, y_train, y_test = train_test_split(x, y, random_state
= 0, test_size = 0.2)

clf = svm.SVC(kernel='rbf')
clf.fit(x_train,y_train)
y_pred = clf.predict(x_test)

from sklearn.metrics import accuracy_score
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.7317073170731707
```

```
from sklearn.tree import DecisionTreeRegressor
tree=DecisionTreeRegressor()#(criterion="entropy",max_depth=4)
tree
```

```
DecisionTreeRegressor()

tree.fit(x_train,y_train)

DecisionTreeRegressor()

y_pred=tree.predict(x_test)

print(y_pred[0:5])
print(y_test[0:5])

[0. 1. 0. 0. 1.]
454    1
52     0
536    1
469    0
55     1
Name: Loan_Status, dtype: category
Categories (2, int64): [0, 1]

from sklearn import metrics
print("DecisionTrees's Accuracy: ",metrics.r2_score(y_pred,y_test))

DecisionTrees's Accuracy:  -0.145962732919255

from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(random_state=42)
rfc.fit(x_train, y_train)
rfc_pred = rfc.predict(x_test)

from sklearn.metrics import classification_report,confusion_matrix
confusion_matrix(y_test, rfc_pred)

array([[14, 19],
       [ 7, 83]], dtype=int64)

classification_report(y_test, rfc_pred)

'              precision    recall  f1-score   support\n\n           0
0.67      0.42      0.52        33\n           1       0.81      0.92
0.86        90\n\n    accuracy                           0.79
123\n   macro avg       0.74      0.67      0.69       123\nweighted
avg       0.77      0.79      0.77       123\n'

##################################################################
############################################################

"""
Project-2
Apply Exploratory Data Analysis on the FifaDataset
Convert the appropriate columns to machine understandable columns
using label encoding
Handle Null values
```

```
Apply the correlation and remove the column which are more than 60%
correlated
Apply the variance and remove the columns whose variance is less that
20%
Understand the Dataset whether the dataset is corr
"""


import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from collections import Counter

import warnings
warnings.filterwarnings('ignore')
import plotly
sns.set_style('darkgrid')

fifa=pd.read_csv("E:/fifa.csv")
fifa
```

```
           id            name                          full_name
birth_date  \
0      158023        L. Messi    Lionel Andrés Messi Cuccittini  1987-
06-24
1      190460      C. Eriksen      Christian  Dannemann Eriksen  1992-
02-14
2      195864        P. Pogba                       Paul Pogba  1993-
03-15
3      198219      L. Insigne                  Lorenzo Insigne  1991-
06-04
4      201024    K. Koulibaly                Kalidou Koulibaly  1991-
06-20
...       ...             ...                               ...
...
17949  204322     R. McKenzie                   Rory  McKenzie  1993-
10-07
17950  239762       M. Sipľak                    Michal Sipľak  1996-
02-02
17951  235155      J. Bekkema                      Jan Bekkema  1996-
04-09
17952  244883      A. Al Yami               Abdulrahman Al Yami  1997-
06-19
17953  247187  Júnior Brumado  José Francisco dos Santos Júnior  1999-
05-15

       age  height_cm  weight_kgs  positions   nationality
overall_rating  \
```

```
0       31     170.18        72.1   CF,RW,ST     Argentina
94
1       27     154.94        76.2  CAM,RM,CM       Denmark
88
2       25     190.50        83.9     CM,CAM        France
88
3       27     162.56        59.0      LW,ST         Italy
88
4       27     187.96        88.9         CB       Senegal
88
...    ...        ...         ...        ...           ...
...
17949   25     175.26        74.8  RM,CAM,CM      Scotland
67
17950   23     182.88        79.8         LB      Slovakia
59
17951   22     185.42        89.8         GK   Netherlands
59
17952   21     175.26        64.9      ST,LM  Saudi Arabia
59
17953   19     190.50        79.8         ST        Brazil
59

       ...    LWB   LDM   CDM   RDM   RWB    LB   LCB    CB   RCB    RB

0      ...  64+2  61+2  61+2  61+2  64+2  59+2  48+2  48+2  48+2  59+2

1      ...  71+3  71+3  71+3  71+3  71+3  66+3  57+3  57+3  57+3  66+3

2      ...  76+3  77+3  77+3  77+3  76+3  74+3  72+3  72+3  72+3  74+3

3      ...  63+3  58+3  58+3  58+3  63+3  58+3  44+3  44+3  44+3  58+3

4      ...  73+3  77+3  77+3  77+3  73+3  76+3  85+3  85+3  85+3  76+3

...    ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...

17949  ...  53+2  52+2  52+2  52+2  53+2  50+2  46+2  46+2  46+2  50+2

17950  ...  57+2  55+2  55+2  55+2  57+2  57+2  58+2  58+2  58+2  57+2

17951  ...   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN

17952  ...  41+2  35+2  35+2  35+2  41+2  39+2  32+2  32+2  32+2  39+2

17953  ...  41+2  40+2  40+2  40+2  41+2  40+2  40+2  40+2  40+2  40+2


[17954 rows x 92 columns]
```

```
fifa.head()
```

```
       id          name                        full_name  birth_date
age  \
0  158023      L. Messi  Lionel Andrés Messi Cuccittini  1987-06-24
31
1  190460    C. Eriksen    Christian  Dannemann Eriksen  1992-02-14
27
2  195864      P. Pogba                      Paul Pogba  1993-03-15
25
3  198219    L. Insigne                  Lorenzo Insigne  1991-06-04
27
4  201024  K. Koulibaly               Kalidou Koulibaly  1991-06-20
27

   height_cm  weight_kgs  positions nationality  overall_rating  ...
LWB  \
0     170.18        72.1   CF,RW,ST   Argentina              94  ...
64+2
1     154.94        76.2  CAM,RM,CM     Denmark              88  ...
71+3
2     190.50        83.9     CM,CAM      France              88  ...
76+3
3     162.56        59.0      LW,ST       Italy              88  ...
63+3
4     187.96        88.9         CB     Senegal              88  ...
73+3

    LDM   CDM   RDM   RWB    LB   LCB    CB   RCB    RB
0  61+2  61+2  61+2  64+2  59+2  48+2  48+2  48+2  59+2
1  71+3  71+3  71+3  71+3  66+3  57+3  57+3  57+3  66+3
2  77+3  77+3  77+3  76+3  74+3  72+3  72+3  72+3  74+3
3  58+3  58+3  58+3  63+3  58+3  44+3  44+3  44+3  58+3
4  77+3  77+3  77+3  73+3  76+3  85+3  85+3  85+3  76+3

[5 rows x 92 columns]
```

```
fifa.shape
```

```
(17954, 92)
```

```
fifa.describe
```

```
<bound method NDFrame.describe of            id            name
full_name  birth_date  \
0      158023        L. Messi    Lionel Andrés Messi Cuccittini  1987-
06-24
1      190460      C. Eriksen      Christian  Dannemann Eriksen  1992-
02-14
2      195864        P. Pogba                      Paul Pogba  1993-
03-15
```

```
3      198219      L. Insigne                  Lorenzo Insigne  1991-
06-04
4      201024    K. Koulibaly                Kalidou Koulibaly  1991-
06-20
...       ...             ...                               ...
...
17949  204322     R. McKenzie                   Rory  McKenzie  1993-
10-07
17950  239762       M. Sipľak                   Michal Sipľak  1996-
02-02
17951  235155      J. Bekkema                     Jan Bekkema  1996-
04-09
17952  244883      A. Al Yami               Abdulrahman Al Yami  1997-
06-19
17953  247187  Júnior Brumado  José Francisco dos Santos Júnior  1999-
05-15

       age  height_cm  weight_kgs  positions   nationality
overall_rating  \
0       31     170.18        72.1   CF,RW,ST     Argentina
94
1       27     154.94        76.2  CAM,RM,CM       Denmark
88
2       25     190.50        83.9     CM,CAM        France
88
3       27     162.56        59.0      LW,ST         Italy
88
4       27     187.96        88.9         CB       Senegal
88
...    ...        ...         ...        ...           ...
...
17949   25     175.26        74.8  RM,CAM,CM      Scotland
67
17950   23     182.88        79.8         LB      Slovakia
59
17951   22     185.42        89.8         GK   Netherlands
59
17952   21     175.26        64.9      ST,LM  Saudi Arabia
59
17953   19     190.50        79.8         ST        Brazil
59

       ...   LWB   LDM   CDM   RDM   RWB    LB   LCB    CB   RCB    RB

0      ...  64+2  61+2  61+2  61+2  64+2  59+2  48+2  48+2  48+2  59+2

1      ...  71+3  71+3  71+3  71+3  71+3  66+3  57+3  57+3  57+3  66+3

2      ...  76+3  77+3  77+3  77+3  76+3  74+3  72+3  72+3  72+3  74+3
```

```
3      ...  63+3  58+3  58+3  58+3  63+3  58+3  44+3  44+3  44+3  58+3
4      ...  73+3  77+3  77+3  77+3  73+3  76+3  85+3  85+3  85+3  76+3
...    ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
17949  ...  53+2  52+2  52+2  52+2  53+2  50+2  46+2  46+2  46+2  50+2
17950  ...  57+2  55+2  55+2  55+2  57+2  57+2  58+2  58+2  58+2  57+2
17951  ...   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
17952  ...  41+2  35+2  35+2  35+2  41+2  39+2  32+2  32+2  32+2  39+2
17953  ...  41+2  40+2  40+2  40+2  41+2  40+2  40+2  40+2  40+2  40+2

[17954 rows x 92 columns]>
```

```python
for col in fifa.columns:
    print(col)
```

```
id
name
full_name
birth_date
age
height_cm
weight_kgs
positions
nationality
overall_rating
potential
value_euro
wage_euro
preferred_foot
international_reputation(1-5)
weak_foot(1-5)
skill_moves(1-5)
work_rate
body_type
release_clause_euro
club_team
club_rating
club_position
club_jersey_number
club_join_date
contract_end_year
national_team
national_rating
```

national_team_position
national_jersey_number
crossing
finishing
heading_accuracy
short_passing
volleys
dribbling
curve
freekick_accuracy
long_passing
ball_control
acceleration
sprint_speed
agility
reactions
balance
shot_power
jumping
stamina
strength
long_shots
aggression
interceptions
positioning
vision
penalties
composure
marking
standing_tackle
sliding_tackle
GK_diving
GK_handling
GK_kicking
GK_positioning
GK_reflexes
tags
traits
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
```

```
fifa['nationality'].value_counts()[0:10]
```

```
England        1658
Germany        1199
Spain          1070
France          925
Argentina       904
Brazil          832
Italy           655
Colombia        624
Japan           466
Netherlands     441
Name: nationality, dtype: int64
```

```
fifa['nationality'].value_counts()[:-1]
```

```
England             1658
Germany             1199
Spain               1070
France               925
Argentina            904
                    ...
Vietnam                1
Papua New Guinea       1
South Sudan            1
Malta                  1
Yemen                  1
Name: nationality, Length: 159, dtype: int64
```

```
player_salary = fifa[['full_name','wage_euro']]
player_salary.head()
```

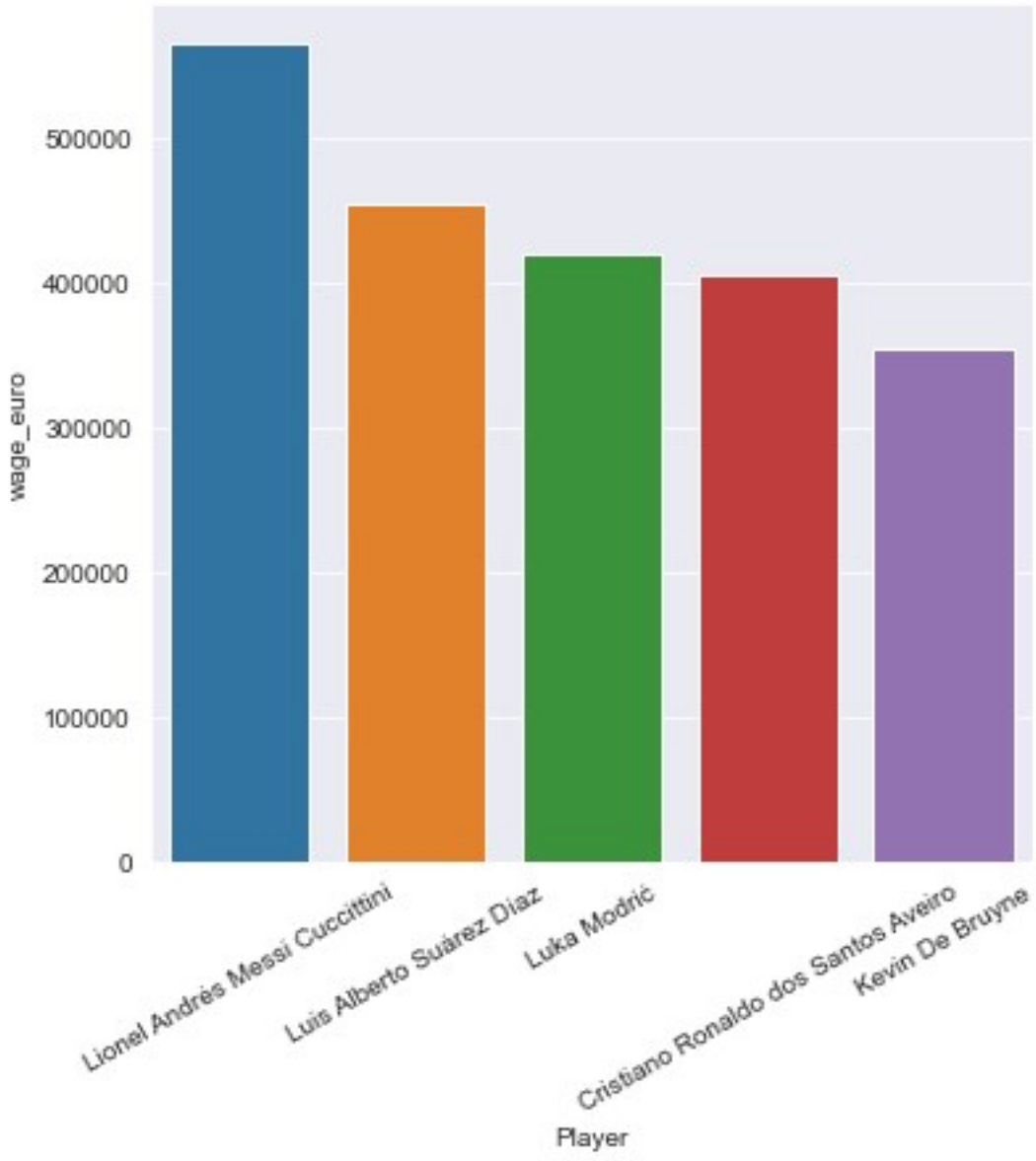
```
                        full_name  wage_euro
0  Lionel Andrés Messi Cuccittini   565000.0
1    Christian  Dannemann Eriksen   205000.0
2                     Paul Pogba   255000.0
```

```
3                 Lorenzo Insigne   165000.0
4               Kalidou Koulibaly   135000.0
```

```
player_salary =
player_salary.sort_values(by=['wage_euro'],ascending=False)
player_salary.head()
```

```
                                 full_name  wage_euro
0          Lionel Andrés Messi Cuccittini   565000.0
17938            Luis Alberto Suárez Díaz   455000.0
17939                        Luka Modrić   420000.0
17944  Cristiano Ronaldo dos Santos Aveiro   405000.0
17941                      Kevin De Bruyne   355000.0
```

```
plt.figure(figsize=(6,6))
sns.barplot(player_salary['full_name'][0:5],player_salary['wage_euro']
[0:5])
plt.xticks(rotation=30)
plt.xlabel('Player')
plt.ylabel('wage_euro');
```




```
Germany = fifa[fifa['nationality']=='Germany']
Germany.head(10)
```

```
         id           name              full_name  birth_date  age
height_cm  \
8    167495       M. Neuer           Manuel Neuer  1986-03-27   32
193.04
12   192448  M. ter Stegen  Marc-André ter Stegen  1992-04-30   26
187.96
36   212188      T. Werner            Timo Werner  1996-03-06   23
154.94
37   212190        N. Süle            Niklas Süle  1995-09-03   23
195.58
49   206113      S. Gnabry           Serge Gnabry  1995-07-14   23
```

```
175.26
56   209658    L. Goretzka          Leon Goretzka  1995-02-06   24
187.96
63   213331         J. Tah           Jonathan Tah  1996-02-11   23
195.58
80   183574       M. Kruse              Max Kruse  1988-03-19   30
154.94
99   204024      C. Kramer       Christoph Kramer  1991-02-19   28
190.50
104  207862      M. Ginter        Matthias Ginter  1994-01-19   25
187.96

     weight_kgs     positions nationality  overall_rating  ...   LWB
LDM  \
8          92.1            GK     Germany              89  ...   NaN
NaN
12         84.8            GK     Germany              89  ...   NaN
NaN
36         74.8            ST     Germany              84  ...  61+3
55+3
37         97.1            CB     Germany              84  ...  69+2
75+2
49         74.8  LM,RM,CAM,LW     Germany              83  ...  65+2
62+2
56         78.9    CM,CAM,CDM     Germany              83  ...  79+3
79+3
63         97.1            CB     Germany              82  ...  70+2
76+2
80         76.2     CF,ST,CAM     Germany              82  ...  64+2
64+2
99         76.2        CDM,CM     Germany              81  ...  76+2
79+2
104        83.0        CB,CDM     Germany              81  ...  74+2
77+2

      CDM   RDM   RWB    LB   LCB    CB   RCB    RB
8     NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
12    NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
36   55+3  55+3  61+3  57+3  49+3  49+3  49+3  57+3
37   75+2  75+2  69+2  73+2  81+2  81+2  81+2  73+2
49   62+2  62+2  65+2  62+2  54+2  54+2  54+2  62+2
56   79+3  79+3  79+3  79+3  77+3  77+3  77+3  79+3
63   76+2  76+2  70+2  73+2  81+2  81+2  81+2  73+2
80   64+2  64+2  64+2  60+2  56+2  56+2  56+2  60+2
99   79+2  79+2  76+2  75+2  75+2  75+2  75+2  75+2
104  77+2  77+2  74+2  76+2  79+2  79+2  79+2  76+2

[10 rows x 92 columns]
```

```
Germany.sort_values(by=['height_cm'],ascending=False).head()
```

```
           id            name         full_name  birth_date  age
height_cm  \
14136  205451       B. Röcker   Benedikt Röcker  1989-11-19   29
198.12
16936  199833   L. Unnerstall   Lars Unnerstall  1990-07-20   28
198.12
14671  158657      T. Kessler    Thomas Kessler  1986-01-20   33
198.12
7839   203835  D. Orlishausen  Dirk Orlishausen  1982-08-15   36
198.12
17736  179783     R. Fährmann     Ralf Fährmann  1988-09-27   30
198.12

       weight_kgs positions nationality  overall_rating  ...   LWB
LDM  \
14136        92.1        CB     Germany              70  ...  56+2
62+2
16936       103.0        GK     Germany              76  ...   NaN
NaN
14671        92.1        GK     Germany              71  ...   NaN
NaN
7839         93.9        GK     Germany              61  ...   NaN
NaN
17736        98.0        GK     Germany              82  ...   NaN
NaN

        CDM   RDM   RWB    LB   LCB    CB   RCB    RB
14136  62+2  62+2  56+2  59+2  69+2  69+2  69+2  59+2
16936   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
14671   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
7839    NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
17736   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN

[5 rows x 92 columns]
```

```
Germany.sort_values(by=['weight_kgs'],ascending=False).head()
```

```
           id           name        full_name  birth_date  age
height_cm  \
8266   239746   L. Watkowiak  Lukas Watkowiak  1996-03-06   23
198.12
16936  199833  L. Unnerstall  Lars Unnerstall  1990-07-20   28
198.12
17736  179783    R. Fährmann    Ralf Fährmann  1988-09-27   30
198.12
37     212190        N. Süle      Niklas Süle  1995-09-03   23
195.58
63     213331         J. Tah     Jonathan Tah  1996-02-11   23
195.58
```

```
       weight_kgs positions nationality  overall_rating  ...   LWB
LDM  \
8266        107.0        GK     Germany              62  ...   NaN
NaN
16936       103.0        GK     Germany              76  ...   NaN
NaN
17736        98.0        GK     Germany              82  ...   NaN
NaN
37           97.1        CB     Germany              84  ...  69+2
75+2
63           97.1        CB     Germany              82  ...  70+2
76+2

        CDM   RDM   RWB    LB   LCB    CB   RCB    RB
8266    NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
16936   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
17736   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
37     75+2  75+2  69+2  73+2  81+2  81+2  81+2  73+2
63     76+2  76+2  70+2  73+2  81+2  81+2  81+2  73+2

[5 rows x 92 columns]
```

```
Germany[['full_name','wage_euro']].sort_values(by=['wage_euro'],ascend
ing=False).head()
```

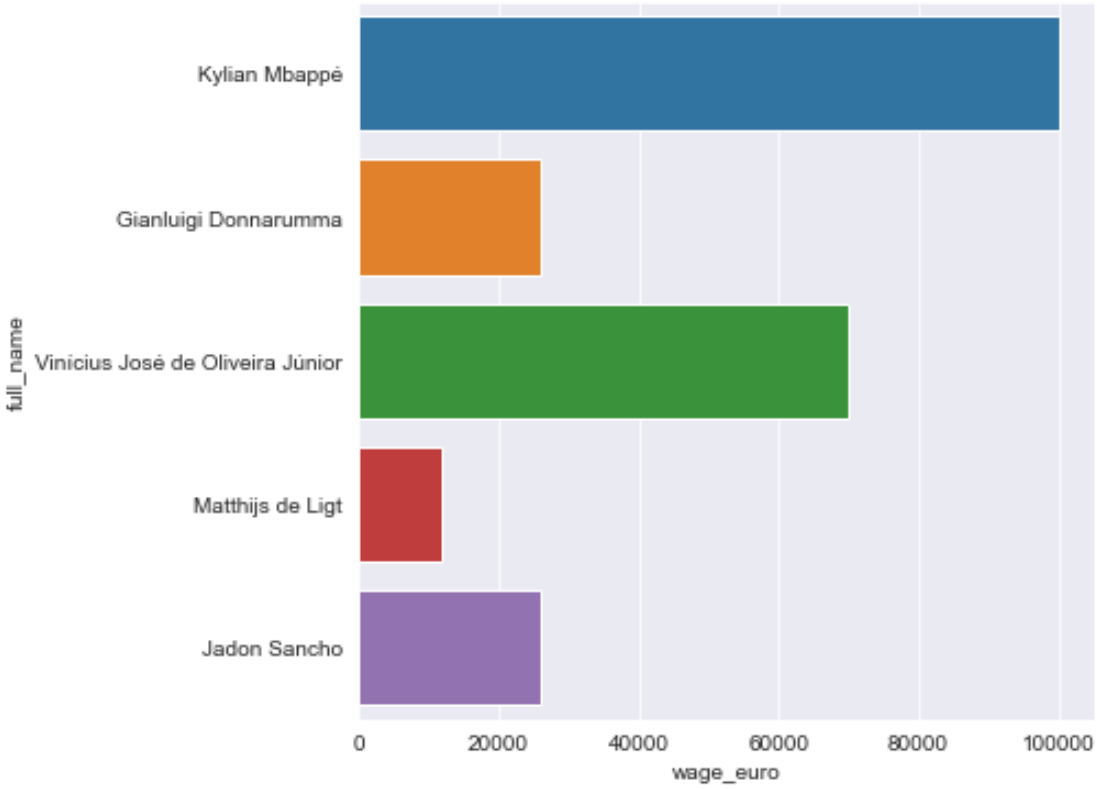
```
                  full_name  wage_euro
17934             Toni Kroos   355000.0
12     Marc-André ter Stegen   240000.0
17915             Leroy Sané   195000.0
17875             Mesut Özil   175000.0
17855         İlkay Gündoğan   175000.0
```

```
#Youngest palyers
player_age = fifa[['full_name','age']]
player_age.sort_values(by=['age']).head()
```

```
             full_name  age
9511        Adil Taoui   17
5637   Gilbert Fuentes   17
5636       Dylan Mbayo   17
5632  Helmer Andersson   17
5617    Adam Ben Lamin   17
```

```
#best youngest players
data_young =
fifa[fifa['age']<=20].sort_values(by=['potential','value_euro'],ascend
ing=[False,True])[0:5]
plt.figure(figsize=(6,6))
sns.barplot(y='full_name',x='wage_euro',data=data_young)
```

```
<AxesSubplot:xlabel='wage_euro', ylabel='full_name'>
```

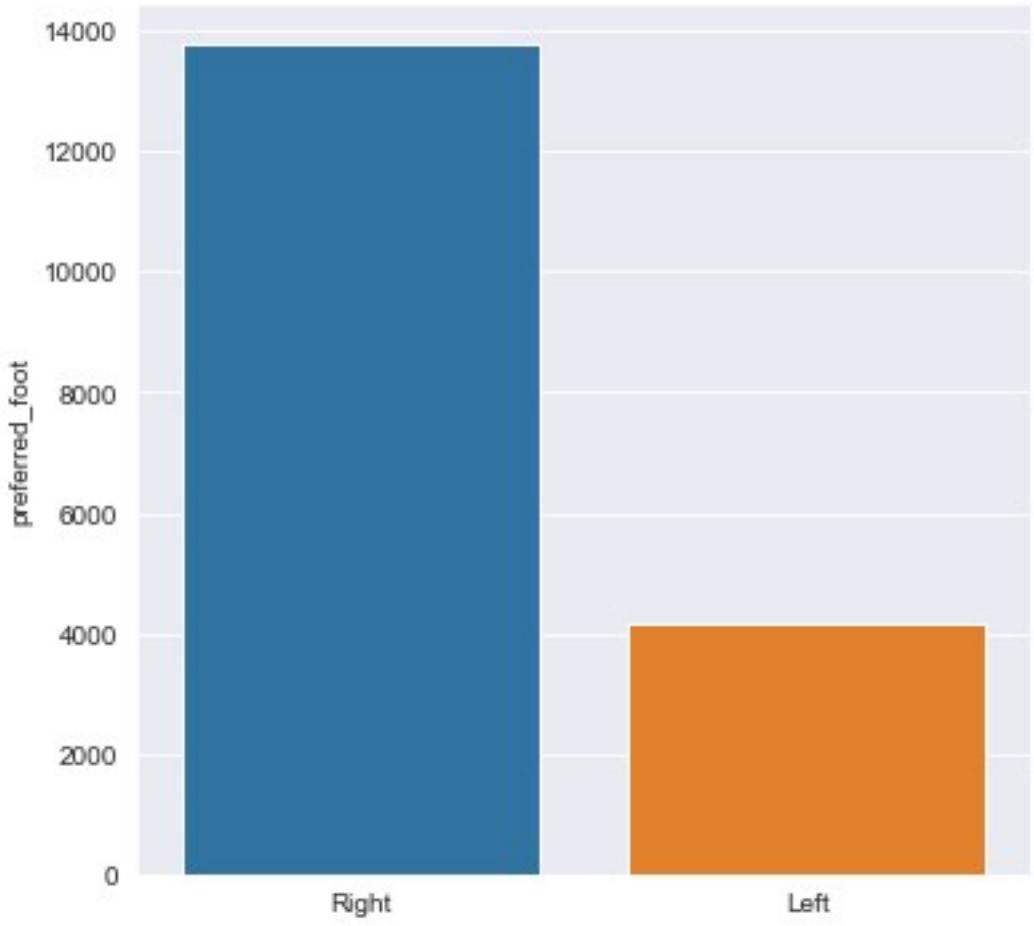



```
#Which foot do footballers prefer?

fifa['preferred_foot'].value_counts()
```

```
Right    13781
Left      4173
Name: preferred_foot, dtype: int64
```

```
#plotting preferred foot

plt.figure(figsize=(6,6))
sns.barplot(fifa['preferred_foot'].value_counts().keys(),fifa['preferr
ed_foot'].value_counts())
```

```
<AxesSubplot:ylabel='preferred_foot'>
```




```
fifa.describe().T
```

	count	mean	std	min
id	17954.0	2.154111e+05	2.975839e+04	16.0
age	17954.0	2.556545e+01	4.705708e+00	17.0
height_cm	17954.0	1.749469e+02	1.402945e+01	152.4
weight_kgs	17954.0	7.530105e+01	7.083684e+00	49.9
overall_rating	17954.0	6.624017e+01	6.963730e+00	47.0
potential	17954.0	7.143093e+01	6.131339e+00	48.0
value_euro	17699.0	2.479280e+06	5.687014e+06	10000.0
wage_euro	17708.0	9.902135e+03	2.199559e+04	1000.0
international_reputation(1-5)	17954.0	1.111674e+00	3.921679e-01	1.0

weak_foot(1-5)	17954.0	2.945695e+00	6.636915e-01	1.0
skill_moves(1-5)	17954.0	2.361034e+00	7.632226e-01	1.0
release_clause_euro	16117.0	4.622522e+06	1.129077e+07	13000.0
club_rating	17940.0	6.936416e+01	5.091678e+00	54.0
club_jersey_number	17940.0	2.029732e+01	1.646473e+01	1.0
national_rating	857.0	7.634189e+01	4.786667e+00	63.0
national_jersey_number	857.0	1.245624e+01	8.098157e+00	1.0
crossing	17954.0	4.969767e+01	1.848331e+01	5.0
finishing	17954.0	4.535825e+01	1.964079e+01	2.0
heading_accuracy	17954.0	5.214821e+01	1.752182e+01	4.0
short_passing	17954.0	5.856606e+01	1.486318e+01	7.0
volleys	17954.0	4.275521e+01	1.780294e+01	3.0
dribbling	17954.0	5.527899e+01	1.909188e+01	4.0
curve	17954.0	4.710115e+01	1.853662e+01	6.0
freekick_accuracy	17954.0	4.268809e+01	1.756563e+01	3.0
long_passing	17954.0	5.266743e+01	1.542155e+01	9.0
ball_control	17954.0	5.822391e+01	1.687572e+01	5.0
acceleration	17954.0	6.469622e+01	1.500134e+01	12.0
sprint_speed	17954.0	6.480350e+01	1.472275e+01	12.0
agility	17954.0	6.337746e+01	1.480695e+01	11.0
reactions	17954.0	6.182188e+01	9.075675e+00	24.0
balance	17954.0	6.387006e+01	1.417404e+01	16.0
shot_power	17954.0	5.531987e+01	1.735130e+01	2.0
jumping	17954.0	6.495522e+01	1.170571e+01	15.0
stamina	17954.0	6.313384e+01	1.591755e+01	12.0

strength	17954.0	6.515885e+01	1.253595e+01	20.0
long_shots	17954.0	4.685246e+01	1.942934e+01	3.0
aggression	17954.0	5.581653e+01	1.739405e+01	11.0
interceptions	17954.0	4.665796e+01	2.075465e+01	3.0
positioning	17954.0	4.985730e+01	1.969431e+01	2.0
vision	17954.0	5.340626e+01	1.415604e+01	10.0
penalties	17954.0	4.835730e+01	1.581084e+01	5.0
composure	17954.0	5.868018e+01	1.162554e+01	12.0
marking	17954.0	4.716286e+01	2.003735e+01	3.0
standing_tackle	17954.0	4.773304e+01	2.167497e+01	2.0
sliding_tackle	17954.0	4.570592e+01	2.128581e+01	3.0
GK_diving	17954.0	1.678445e+01	1.789390e+01	1.0
GK_handling	17954.0	1.655620e+01	1.710823e+01	1.0
GK_kicking	17954.0	1.638866e+01	1.670145e+01	1.0
GK_positioning	17954.0	1.654105e+01	1.723150e+01	1.0
GK_reflexes	17954.0	1.687407e+01	1.818294e+01	1.0

	25%	50%	75%	max
id	201117.25	222919.00	237613.50	2.476070e+05
age	22.00	25.00	29.00	4.600000e+01
height_cm	154.94	175.26	185.42	2.057400e+02
weight_kgs	69.90	74.80	79.80	1.102000e+02
overall_rating	62.00	66.00	71.00	9.400000e+01
potential	67.00	71.00	75.00	9.500000e+01
value_euro	325000.00	700000.00	2100000.00	1.105000e+08
wage_euro	1000.00	3000.00	9000.00	

5.650000e+05				
international_reputation(1-5)	1.00	1.00	1.00	
5.000000e+00				
weak_foot(1-5)	3.00	3.00	3.00	
5.000000e+00				
skill_moves(1-5)	2.00	2.00	3.00	
5.000000e+00				
release_clause_euro	525000.00	1200000.00	3500000.00	
2.265000e+08				
club_rating	66.00	69.00	72.00	
8.600000e+01				
club_jersey_number	9.00	18.00	27.00	
9.900000e+01				
national_rating	73.00	75.00	81.00	
8.500000e+01				
national_jersey_number	6.00	12.00	18.00	
9.900000e+01				
crossing	38.00	54.00	64.00	
9.300000e+01				
finishing	30.00	49.00	62.00	
9.500000e+01				
heading_accuracy	44.00	56.00	64.00	
9.400000e+01				
short_passing	53.00	62.00	68.00	
9.300000e+01				
volleys	30.00	44.00	57.00	
9.000000e+01				
dribbling	49.00	61.00	68.00	
9.700000e+01				
curve	34.00	49.00	62.00	
9.400000e+01				
freekick_accuracy	30.00	41.00	56.00	
9.400000e+01				
long_passing	43.00	56.00	64.00	
9.300000e+01				
ball_control	54.00	63.00	69.00	
9.600000e+01				
acceleration	57.00	67.00	75.00	
9.700000e+01				
sprint_speed	58.00	67.00	75.00	
9.600000e+01				
agility	55.00	66.00	74.00	
9.600000e+01				
reactions	56.00	62.00	68.00	
9.600000e+01				
balance	56.00	66.00	74.00	
9.600000e+01				
shot_power	45.00	59.00	68.00	
9.500000e+01				
jumping	58.00	66.00	73.00	

```
9.500000e+01
stamina                          56.00       66.00       74.00
9.700000e+01
strength                         58.00       66.00       74.00
9.700000e+01
long_shots                       32.00       51.00       62.00
9.400000e+01
aggression                       44.00       59.00       69.00
9.500000e+01
interceptions                    26.00       52.00       64.00
9.200000e+01
positioning                      38.00       55.00       64.00
9.500000e+01
vision                           44.00       55.00       64.00
9.400000e+01
penalties                        38.00       49.00       60.00
9.200000e+01
composure                        51.00       60.00       67.00
9.600000e+01
marking                          30.00       52.50       64.00
9.400000e+01
standing_tackle                  27.00       55.00       66.00
9.300000e+01
sliding_tackle                   24.00       52.00       64.00
9.000000e+01
GK_diving                         8.00       11.00       14.00
9.000000e+01
GK_handling                       8.00       11.00       14.00
9.200000e+01
GK_kicking                        8.00       11.00       14.00
9.200000e+01
GK_positioning                    8.00       11.00       14.00
9.000000e+01
GK_reflexes                       8.00       11.00       14.00
9.400000e+01
```

```
fifa.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17954 entries, 0 to 17953
Data columns (total 92 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   id                            17954 non-null  int64
 1   name                          17954 non-null  object
 2   full_name                     17954 non-null  object
 3   birth_date                    17954 non-null  object
 4   age                           17954 non-null  int64
 5   height_cm                     17954 non-null  float64
 6   weight_kgs                    17954 non-null  float64
 7   positions                     17954 non-null  object
```

```
8   nationality                     17954 non-null  object
9   overall_rating                  17954 non-null  int64
10  potential                       17954 non-null  int64
11  value_euro                      17699 non-null  float64
12  wage_euro                       17708 non-null  float64
13  preferred_foot                  17954 non-null  object
14  international_reputation(1-5)  17954 non-null  int64
15  weak_foot(1-5)                  17954 non-null  int64
16  skill_moves(1-5)                17954 non-null  int64
17  work_rate                       17954 non-null  object
18  body_type                       17954 non-null  object
19  release_clause_euro             16117 non-null  float64
20  club_team                       17940 non-null  object
21  club_rating                     17940 non-null  float64
22  club_position                   17940 non-null  object
23  club_jersey_number              17940 non-null  float64
24  club_join_date                  16018 non-null  object
25  contract_end_year               17593 non-null  object
26  national_team                   857 non-null    object
27  national_rating                 857 non-null    float64
28  national_team_position          857 non-null    object
29  national_jersey_number          857 non-null    float64
30  crossing                        17954 non-null  int64
31  finishing                       17954 non-null  int64
32  heading_accuracy                17954 non-null  int64
33  short_passing                   17954 non-null  int64
34  volleys                         17954 non-null  int64
35  dribbling                       17954 non-null  int64
36  curve                           17954 non-null  int64
37  freekick_accuracy               17954 non-null  int64
38  long_passing                    17954 non-null  int64
39  ball_control                    17954 non-null  int64
40  acceleration                    17954 non-null  int64
41  sprint_speed                    17954 non-null  int64
42  agility                         17954 non-null  int64
43  reactions                       17954 non-null  int64
44  balance                         17954 non-null  int64
45  shot_power                      17954 non-null  int64
46  jumping                         17954 non-null  int64
47  stamina                         17954 non-null  int64
48  strength                        17954 non-null  int64
49  long_shots                      17954 non-null  int64
50  aggression                      17954 non-null  int64
51  interceptions                   17954 non-null  int64
52  positioning                     17954 non-null  int64
53  vision                          17954 non-null  int64
54  penalties                       17954 non-null  int64
55  composure                       17954 non-null  int64
56  marking                         17954 non-null  int64
57  standing_tackle                 17954 non-null  int64
```

```
 58  sliding_tackle                17954 non-null  int64
 59  GK_diving                     17954 non-null  int64
 60  GK_handling                   17954 non-null  int64
 61  GK_kicking                    17954 non-null  int64
 62  GK_positioning                17954 non-null  int64
 63  GK_reflexes                   17954 non-null  int64
 64  tags                          1417 non-null   object
 65  traits                        8137 non-null   object
 66  LS                            15889 non-null  object
 67  ST                            15889 non-null  object
 68  RS                            15889 non-null  object
 69  LW                            15889 non-null  object
 70  LF                            15889 non-null  object
 71  CF                            15889 non-null  object
 72  RF                            15889 non-null  object
 73  RW                            15889 non-null  object
 74  LAM                           15889 non-null  object
 75  CAM                           15889 non-null  object
 76  RAM                           15889 non-null  object
 77  LM                            15889 non-null  object
 78  LCM                           15889 non-null  object
 79  CM                            15889 non-null  object
 80  RCM                           15889 non-null  object
 81  RM                            15889 non-null  object
 82  LWB                           15889 non-null  object
 83  LDM                           15889 non-null  object
 84  CDM                           15889 non-null  object
 85  RDM                           15889 non-null  object
 86  RWB                           15889 non-null  object
 87  LB                            15889 non-null  object
 88  LCB                           15889 non-null  object
 89  CB                            15889 non-null  object
 90  RCB                           15889 non-null  object
 91  RB                            15889 non-null  object
dtypes: float64(9), int64(41), object(42)
memory usage: 12.6+ MB
```

```
fifa.isnull().sum()
```

```
id               0
name             0
full_name        0
birth_date       0
age              0
               ...
LB            2065
LCB           2065
CB            2065
RCB           2065
RB            2065
Length: 92, dtype: int64
```

fifa

```
           id             name                          full_name
birth_date  \
0      158023        L. Messi    Lionel Andrés Messi Cuccittini  1987-
06-24
1      190460      C. Eriksen      Christian  Dannemann Eriksen  1992-
02-14
2      195864        P. Pogba                        Paul Pogba  1993-
03-15
3      198219      L. Insigne                    Lorenzo Insigne  1991-
06-04
4      201024    K. Koulibaly                 Kalidou Koulibaly  1991-
06-20
...       ...             ...                               ...
...
17949  204322     R. McKenzie                     Rory  McKenzie  1993-
10-07
17950  239762       M. Sipľak                     Michal Sipľak  1996-
02-02
17951  235155      J. Bekkema                       Jan Bekkema  1996-
04-09
17952  244883      A. Al Yami               Abdulrahman Al Yami  1997-
06-19
17953  247187  Júnior Brumado  José Francisco dos Santos Júnior  1999-
05-15

       age  height_cm  weight_kgs  positions   nationality
overall_rating  \
0       31     170.18        72.1   CF,RW,ST     Argentina
94
1       27     154.94        76.2  CAM,RM,CM       Denmark
88
2       25     190.50        83.9     CM,CAM        France
88
3       27     162.56        59.0      LW,ST         Italy
88
4       27     187.96        88.9         CB       Senegal
88
...    ...        ...         ...        ...           ...
...
17949   25     175.26        74.8  RM,CAM,CM      Scotland
67
17950   23     182.88        79.8         LB      Slovakia
59
17951   22     185.42        89.8         GK   Netherlands
59
17952   21     175.26        64.9      ST,LM  Saudi Arabia
59
17953   19     190.50        79.8         ST        Brazil
```

```
59

       ...   LWB   LDM   CDM   RDM   RWB    LB   LCB    CB   RCB    RB

0      ...  64+2  61+2  61+2  61+2  64+2  59+2  48+2  48+2  48+2  59+2

1      ...  71+3  71+3  71+3  71+3  71+3  66+3  57+3  57+3  57+3  66+3

2      ...  76+3  77+3  77+3  77+3  76+3  74+3  72+3  72+3  72+3  74+3

3      ...  63+3  58+3  58+3  58+3  63+3  58+3  44+3  44+3  44+3  58+3

4      ...  73+3  77+3  77+3  77+3  73+3  76+3  85+3  85+3  85+3  76+3

...    ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...

17949  ...  53+2  52+2  52+2  52+2  53+2  50+2  46+2  46+2  46+2  50+2

17950  ...  57+2  55+2  55+2  55+2  57+2  57+2  58+2  58+2  58+2  57+2

17951  ...   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN

17952  ...  41+2  35+2  35+2  35+2  41+2  39+2  32+2  32+2  32+2  39+2

17953  ...  41+2  40+2  40+2  40+2  41+2  40+2  40+2  40+2  40+2  40+2

[17954 rows x 92 columns]

fifa2=fifa.fillna(value=0)

fifa2

           id            name                          full_name
birth_date  \
0      158023        L. Messi    Lionel Andrés Messi Cuccittini  1987-
06-24
1      190460      C. Eriksen      Christian  Dannemann Eriksen  1992-
02-14
2      195864        P. Pogba                       Paul Pogba  1993-
03-15
3      198219      L. Insigne                  Lorenzo Insigne  1991-
06-04
4      201024    K. Koulibaly                Kalidou Koulibaly  1991-
06-20
...       ...             ...                               ...
...
17949  204322     R. McKenzie                   Rory  McKenzie  1993-
```

```
10-07
17950  239762       M. Sipľak                     Michal Sipľak  1996-
02-02
17951  235155      J. Bekkema                       Jan Bekkema  1996-
04-09
17952  244883      A. Al Yami               Abdulrahman Al Yami  1997-
06-19
17953  247187  Júnior Brumado  José Francisco dos Santos Júnior  1999-
05-15

       age  height_cm  weight_kgs  positions   nationality
overall_rating  \
0       31     170.18        72.1   CF,RW,ST     Argentina
94
1       27     154.94        76.2  CAM,RM,CM       Denmark
88
2       25     190.50        83.9     CM,CAM        France
88
3       27     162.56        59.0      LW,ST         Italy
88
4       27     187.96        88.9         CB       Senegal
88
...    ...        ...         ...        ...           ...
...
17949   25     175.26        74.8  RM,CAM,CM      Scotland
67
17950   23     182.88        79.8         LB      Slovakia
59
17951   22     185.42        89.8         GK   Netherlands
59
17952   21     175.26        64.9      ST,LM  Saudi Arabia
59
17953   19     190.50        79.8         ST        Brazil
59

       ...   LWB   LDM   CDM   RDM   RWB    LB   LCB    CB   RCB    RB

0      ...  64+2  61+2  61+2  61+2  64+2  59+2  48+2  48+2  48+2  59+2

1      ...  71+3  71+3  71+3  71+3  71+3  66+3  57+3  57+3  57+3  66+3

2      ...  76+3  77+3  77+3  77+3  76+3  74+3  72+3  72+3  72+3  74+3

3      ...  63+3  58+3  58+3  58+3  63+3  58+3  44+3  44+3  44+3  58+3

4      ...  73+3  77+3  77+3  77+3  73+3  76+3  85+3  85+3  85+3  76+3

...    ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

```
17949  ...  53+2  52+2  52+2  52+2  53+2  50+2  46+2  46+2  46+2  50+2

17950  ...  57+2  55+2  55+2  55+2  57+2  57+2  58+2  58+2  58+2  57+2

17951  ...     0     0     0     0     0     0     0     0     0     0

17952  ...  41+2  35+2  35+2  35+2  41+2  39+2  32+2  32+2  32+2  39+2

17953  ...  41+2  40+2  40+2  40+2  41+2  40+2  40+2  40+2  40+2  40+2


[17954 rows x 92 columns]
```

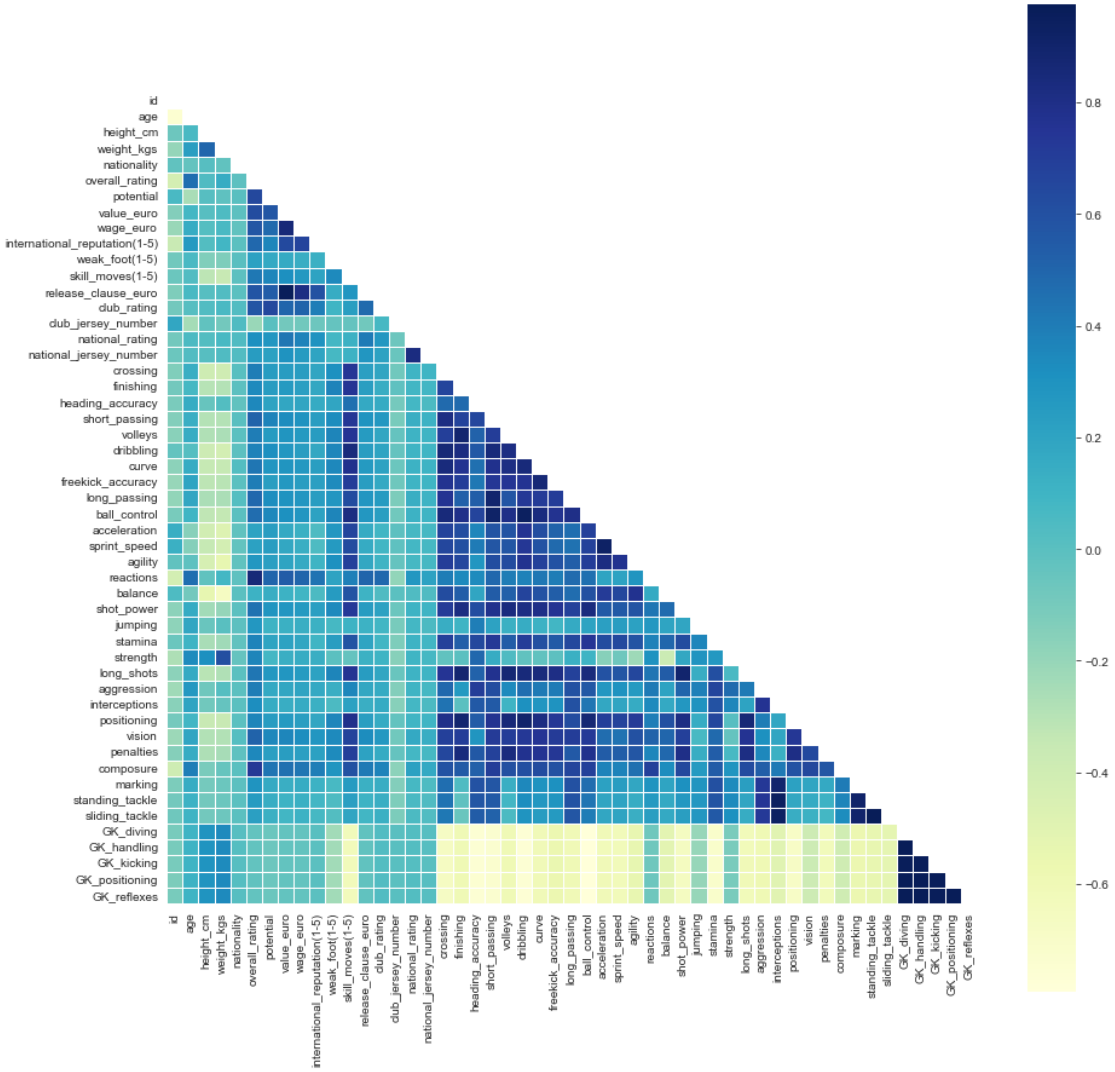
```
from sklearn import preprocessing

# label_encoder object knows how to understand word labels.
label_encoder = preprocessing.LabelEncoder()

# Encode labels in column
fifa2['nationality']=
label_encoder.fit_transform(fifa2['nationality'])

fifa2['nationality'].unique()
```

```
array([  6,  39,  53,  76, 126, 104,  57, 152, 133,  13,  42, 129,
18,
        33,  32,  29, 101, 119, 138,  17,  97,  44,   9,  70,  69,
156,
       149,  25, 127,  77,  22, 115,   8,   2, 121, 122,  75, 139,
27,
       146, 147, 108, 116, 110,  59, 151, 154,  73,  45,  24, 142,
125,
        28,  85,  99,  23,  93,  94,  58,  63, 118,  34, 109,  79,
106,
       131, 120,  41,  20,  37, 130,  12,  54,  49,  35,  38,  67,
128,
        64, 124,  51,  84, 158, 140,   7,  56,   4, 159,  31,  46,
74,
         1,  16,  55,  78,  21, 148,  14, 137,  52,  91, 144,  43,
83,
       113,  98,  66,  30,  10, 145,  88,  19,  82,  26,  50,  81,
47,
       100,  40, 143,  71,   3, 136,  68,  95,   0, 153, 112, 157,
92,
        96,  72,  15,  60, 117,  90,  65, 134,   5,  36, 123, 132,
114,
        87, 155, 107,  48,  11,  80, 135,  62, 103,  61, 102,  86,
89,
       105, 111, 141, 150])
```

```
fifa2['overall_rating']=
label_encoder.fit_transform(fifa2['overall_rating'])

fifa2['overall_rating'].unique()
```

```
array([46, 41, 42, 39, 40, 37, 38, 36, 35, 34, 33, 32, 31, 30, 29, 28,
27,
       26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11,
10,
        9,  8,  7,  6,  5,  4,  3,  2,  0,  1, 43, 44, 45],
dtype=int64)
```

```
corr = fifa2.corr()
mask = np.zeros_like(corr)
mask[np.triu_indices_from(mask)] = True
with sns.axes_style("white"):
    f, ax = plt.subplots(figsize=(15, 15))
    ax =
sns.heatmap(corr,mask=mask,square=True,linewidths=.8,cmap="YlGnBu")
```




```
def correlation(dataset, threshold):#X_train,0.7
    col_corr = set()  # Set of all the names of correlated columns
    corr_matrix = dataset.corr()
    for i in range(len(corr_matrix)): #traverse through the rows
        for j in range(i): #traverse through column
            if abs(corr_matrix.iloc[i, j]) > threshold: # we are
interested in absolute coeff value
                colname = corr_matrix.columns[i]  # getting the name
of column
                col_corr.add(colname)
    return col_corr

corr_features = correlation(fifa2, 0.3)
len((corr_features))
```

```
46
```

```
fifa2
```

```
           id             name                          full_name
birth_date  \
0      158023        L. Messi    Lionel Andrés Messi Cuccittini  1987-
06-24
1      190460      C. Eriksen      Christian  Dannemann Eriksen  1992-
02-14
2      195864        P. Pogba                        Paul Pogba  1993-
03-15
3      198219      L. Insigne                  Lorenzo Insigne  1991-
06-04
4      201024    K. Koulibaly                 Kalidou Koulibaly  1991-
06-20
...       ...             ...                               ...
...
17949  204322     R. McKenzie                    Rory  McKenzie  1993-
10-07
17950  239762       M. Sipľak                     Michal Sipľak  1996-
02-02
17951  235155      J. Bekkema                       Jan Bekkema  1996-
04-09
17952  244883      A. Al Yami               Abdulrahman Al Yami  1997-
06-19
17953  247187  Júnior Brumado  José Francisco dos Santos Júnior  1999-
05-15

       age  height_cm  weight_kgs  positions  nationality
overall_rating  \
0       31     170.18        72.1   CF,RW,ST            6
46
1       27     154.94        76.2  CAM,RM,CM           39
41
2       25     190.50        83.9     CM,CAM           53
41
3       27     162.56        59.0      LW,ST           76
41
4       27     187.96        88.9         CB          126
41
...    ...        ...         ...        ...          ...
...
17949   25     175.26        74.8  RM,CAM,CM          125
20
17950   23     182.88        79.8         LB          129
12
17951   22     185.42        89.8         GK          104
12
17952   21     175.26        64.9      ST,LM          124
12
17953   19     190.50        79.8         ST           18
12
```

```
       ...   LWB   LDM   CDM   RDM   RWB    LB   LCB    CB   RCB    RB
0      ...  64+2  61+2  61+2  61+2  64+2  59+2  48+2  48+2  48+2  59+2
1      ...  71+3  71+3  71+3  71+3  71+3  66+3  57+3  57+3  57+3  66+3
2      ...  76+3  77+3  77+3  77+3  76+3  74+3  72+3  72+3  72+3  74+3
3      ...  63+3  58+3  58+3  58+3  63+3  58+3  44+3  44+3  44+3  58+3
4      ...  73+3  77+3  77+3  77+3  73+3  76+3  85+3  85+3  85+3  76+3
...    ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
17949  ...  53+2  52+2  52+2  52+2  53+2  50+2  46+2  46+2  46+2  50+2
17950  ...  57+2  55+2  55+2  55+2  57+2  57+2  58+2  58+2  58+2  57+2
17951  ...     0     0     0     0     0     0     0     0     0     0
17952  ...  41+2  35+2  35+2  35+2  41+2  39+2  32+2  32+2  32+2  39+2
17953  ...  41+2  40+2  40+2  40+2  41+2  40+2  40+2  40+2  40+2  40+2

[17954 rows x 92 columns]
```

```
fifa2.columns
```

```
Index(['id', 'name', 'full_name', 'birth_date', 'age', 'height_cm',
       'weight_kgs', 'positions', 'nationality', 'overall_rating',
'potential',
       'value_euro', 'wage_euro', 'preferred_foot',
       'international_reputation(1-5)', 'weak_foot(1-5)',
'skill_moves(1-5)',
       'work_rate', 'body_type', 'release_clause_euro', 'club_team',
       'club_rating', 'club_position', 'club_jersey_number',
'club_join_date',
       'contract_end_year', 'national_team', 'national_rating',
       'national_team_position', 'national_jersey_number', 'crossing',
       'finishing', 'heading_accuracy', 'short_passing', 'volleys',
       'dribbling', 'curve', 'freekick_accuracy', 'long_passing',
       'ball_control', 'acceleration', 'sprint_speed', 'agility',
'reactions',
       'balance', 'shot_power', 'jumping', 'stamina', 'strength',
'long_shots',
       'aggression', 'interceptions', 'positioning', 'vision',
'penalties',
       'composure', 'marking', 'standing_tackle', 'sliding_tackle',
```

```
       'GK_diving', 'GK_handling', 'GK_kicking', 'GK_positioning',
       'GK_reflexes', 'tags', 'traits', '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'],
      dtype='object')
```

```
fifa2.select_dtypes('object').columns
```

```
Index(['name', 'full_name', 'birth_date', 'positions',
'preferred_foot',
       'work_rate', 'body_type', 'club_team', 'club_position',
       'club_join_date', 'contract_end_year', 'national_team',
       'national_team_position', 'tags', 'traits', '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'],
      dtype='object')
```

```
fifa2.drop(fifa2.select_dtypes('object').columns,inplace=True,axis=1)
```

```
from sklearn.feature_selection import VarianceThreshold
var_thres=VarianceThreshold(threshold=0.5)
var_thres.fit(fifa2)
```

```
VarianceThreshold(threshold=0.5)
```

```
var_thres.get_support()
```

```
array([ True,  True,  True,  True,  True,  True,  True,  True,  True,
       False, False,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True])
```

```
fifa2.columns[var_thres.get_support() == True]
```

```
Index(['id', 'age', 'height_cm', 'weight_kgs', 'nationality',
'overall_rating',
       'potential', 'value_euro', 'wage_euro', 'skill_moves(1-5)',
       'release_clause_euro', 'club_rating', 'club_jersey_number',
       'national_rating', 'national_jersey_number', 'crossing',
'finishing',
       'heading_accuracy', 'short_passing', 'volleys', 'dribbling',
'curve',
       'freekick_accuracy', 'long_passing', 'ball_control',
'acceleration',
```

```
       'sprint_speed', 'agility', 'reactions', 'balance',
'shot_power',
       'jumping', 'stamina', 'strength', 'long_shots', 'aggression',
       'interceptions', 'positioning', 'vision', 'penalties',
'composure',
       'marking', 'standing_tackle', 'sliding_tackle', 'GK_diving',
       'GK_handling', 'GK_kicking', 'GK_positioning', 'GK_reflexes'],
      dtype='object')
```

```
columns_having_var_more_than_50 =
fifa2.columns[var_thres.get_support() == True]
```

```
len(columns_having_var_more_than_50)
```

```
49
```

```
len(fifa2.columns)
```

```
51
```

```
fifa2.columns[var_thres.get_support() == False]
```

```
Index(['international_reputation(1-5)', 'weak_foot(1-5)'],
dtype='object')
```

```
columns_having_var_less_than_50 =
fifa2.columns[var_thres.get_support() == False]
```

```
len(columns_having_var_less_than_50)
```

```
2
```

```
fifa2.drop(columns_having_var_less_than_50,inplace = True,axis= 1)
```

```
fifa2
```

```
           id  age  height_cm  weight_kgs  nationality  overall_rating
\
0      158023   31     170.18        72.1            6              46

1      190460   27     154.94        76.2           39              41

2      195864   25     190.50        83.9           53              41

3      198219   27     162.56        59.0           76              41

4      201024   27     187.96        88.9          126              41

...       ...  ...        ...         ...          ...             ...

17949  204322   25     175.26        74.8          125              20
```

```
17950  239762   23     182.88        79.8          129              12

17951  235155   22     185.42        89.8          104              12

17952  244883   21     175.26        64.9          124              12

17953  247187   19     190.50        79.8           18              12


       potential   value_euro  wage_euro  skill_moves(1-5)  ...
penalties  \
0             94  110500000.0   565000.0                 4  ...
75
1             89   69500000.0   205000.0                 4  ...
67
2             91   73000000.0   255000.0                 5  ...
82
3             88   62000000.0   165000.0                 4  ...
61
4             91   60000000.0   135000.0                 2  ...
33
...          ...          ...        ...               ...  ...
...
17949         70     975000.0     3000.0                 3  ...
63
17950         67     190000.0     1000.0                 2  ...
32
17951         67     170000.0     1000.0                 1  ...
16
17952         71     280000.0     4000.0                 3  ...
50
17953         75     325000.0     2000.0                 2  ...
52

       composure  marking  standing_tackle  sliding_tackle  GK_diving
\
0             96       33               28              26          6

1             88       59               57              22          9

2             87       63               67              67          5

3             83       51               24              22          8

4             80       91               88              87          7

...          ...      ...              ...             ...        ...

17949         56       40               20              18         15
```

```
17950         52       53                64               60          15

17951         47        9                12               13          63

17952         53       16                18               17          15

17953         45       23                31               21           9
```

```
       GK_handling  GK_kicking  GK_positioning  GK_reflexes
0               11          15              14            8
1               14           7               7            6
2                6           2               4            3
3                4          14               9           10
4               11           7              13            5
...            ...         ...             ...          ...
17949           11          15              12            7
17950            7           7              15            8
17951           57          50              58           63
17952           14          12               6            9
17953           14          12              13           14

[17954 rows x 49 columns]
```

```
fifa2.columns
```

```
Index(['id', 'age', 'height_cm', 'weight_kgs', 'nationality',
'overall_rating',
       'potential', 'value_euro', 'wage_euro', 'skill_moves(1-5)',
       'release_clause_euro', 'club_rating', 'club_jersey_number',
       'national_rating', 'national_jersey_number', 'crossing',
'finishing',
       'heading_accuracy', 'short_passing', 'volleys', 'dribbling',
'curve',
       'freekick_accuracy', 'long_passing', 'ball_control',
'acceleration',
       'sprint_speed', 'agility', 'reactions', 'balance',
'shot_power',
       'jumping', 'stamina', 'strength', 'long_shots', 'aggression',
       'interceptions', 'positioning', 'vision', 'penalties',
'composure',
       'marking', 'standing_tackle', 'sliding_tackle', 'GK_diving',
       'GK_handling', 'GK_kicking', 'GK_positioning', 'GK_reflexes'],
      dtype='object')
```