----------------IMPORT LIBRARIES-------------------

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
import warnings
warnings.filterwarnings('ignore')
```

---------------IMPORT MODULES--------------------

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import 
accuracy_score,classification_report,confusion_matrix,r2_score
from sklearn.metrics import mean_squared_error
```

LOAD THE DATASET

```
data=pd.read_csv(r"C:\Users\Admin\Downloads\EPL_20_21.csv")
data.head()

             Name     Club Nationality Position  Age  Matches  Starts 
Mins  \
0    Mason Mount  Chelsea         ENG    MF,FW   21       36      32 
2890   
1  Edouard Mendy  Chelsea         SEN       GK   28       31      31 
2745   
2    Timo Werner  Chelsea         GER       FW   24       35      29 
2602   
3   Ben Chilwell  Chelsea         ENG       DF   23       27      27 
2286   
4    Reece James  Chelsea         ENG       DF   20       32      25 
2373   

   Goals  Assists  Passes_Attempted  Perc_Passes_Completed 
Penalty_Goals  \
```

```
0       6        5              1881                     82.300
1
1       0        0              1007                     84.600
0
2       6        8               826                     77.200
0
3       3        5              1806                     78.600
0
4       1        2              1987                     85.000
0

   Penalty_Attempted    xG    xA  Yellow_Cards  Red_Cards
0                  1 0.210 0.240             2          0
1                  0 0.000 0.000             2          0
2                  0 0.410 0.210             2          0
3                  0 0.100 0.110             3          0
4                  0 0.060 0.120             3          0
```

```
data.shape
```

```
(532, 18)
```

```
data.describe()
```

```
          Age  Matches  Starts     Mins   Goals  Assists
Passes_Attempted  \
count 532.000  532.000 532.000  532.000 532.000  532.000
532.000
mean   25.500   19.536  15.714 1411.444   1.853    1.288
717.750
std     4.319   11.840  11.921 1043.172   3.338    2.095
631.373
min    16.000    1.000   0.000    1.000   0.000    0.000
0.000
25%    22.000    9.000   4.000  426.000   0.000    0.000
171.500
50%    26.000   21.000  15.000 1345.000   1.000    0.000
573.500
75%    29.000   30.000  27.000 2303.500   2.000    2.000
1129.500
max    38.000   38.000  38.000 3420.000  23.000   14.000
3214.000

       Perc_Passes_Completed  Penalty_Goals  Penalty_Attempted      xG
\
count                532.000        532.000            532.000 532.000

mean                  77.824          0.192              0.235   0.113

std                   13.012          0.851              0.976   0.148
```

```
min                      -1.000          0.000                0.000   0.000

25%                      73.500          0.000                0.000   0.010

50%                      79.200          0.000                0.000   0.060

75%                      84.625          0.000                0.000   0.150

max                     100.000          9.000               10.000   1.160


           xA  Yellow_Cards  Red_Cards
count 532.000       532.000    532.000
mean    0.073         2.115      0.090
std     0.090         2.269      0.293
min     0.000         0.000      0.000
25%     0.000         0.000      0.000
50%     0.050         2.000      0.000
75%     0.110         3.000      0.000
max     0.900        12.000      2.000
```

```
pd.set_option('display.float_format',lambda x:'%.3f' %x)
data.describe().transpose()
```

```
                        count     mean      std    min     25%
50%  \
Age                   532.000   25.500    4.319 16.000  22.000
26.000
Matches               532.000   19.536   11.840  1.000   9.000
21.000
Starts                532.000   15.714   11.921  0.000   4.000
15.000
Mins                  532.000 1411.444 1043.172  1.000 426.000
1345.000
Goals                 532.000    1.853    3.338  0.000   0.000
1.000
Assists               532.000    1.288    2.095  0.000   0.000
0.000
Passes_Attempted      532.000  717.750  631.373  0.000 171.500
573.500
Perc_Passes_Completed 532.000   77.824   13.012 -1.000  73.500
79.200
Penalty_Goals         532.000    0.192    0.851  0.000   0.000
0.000
Penalty_Attempted     532.000    0.235    0.976  0.000   0.000
0.000
xG                    532.000    0.113    0.148  0.000   0.010
0.060
xA                    532.000    0.073    0.090  0.000   0.000
0.050
```

```
Yellow_Cards          532.000    2.115    2.269  0.000   0.000
2.000
Red_Cards             532.000    0.090    0.293  0.000   0.000
0.000

                          75%      max
Age                    29.000   38.000
Matches                30.000   38.000
Starts                 27.000   38.000
Mins                 2303.500 3420.000
Goals                   2.000   23.000
Assists                 2.000   14.000
Passes_Attempted     1129.500 3214.000
Perc_Passes_Completed  84.625  100.000
Penalty_Goals           0.000    9.000
Penalty_Attempted       0.000   10.000
xG                      0.150    1.160
xA                      0.110    0.900
Yellow_Cards            3.000   12.000
Red_Cards               0.000    2.000
```

```
data.duplicated().sum()
```

```
0
```

```
data.isnull().sum()
```

```
Name                     0
Club                     0
Nationality              0
Position                 0
Age                      0
Matches                  0
Starts                   0
Mins                     0
Goals                    0
Assists                  0
Passes_Attempted         0
Perc_Passes_Completed    0
Penalty_Goals            0
Penalty_Attempted        0
xG                       0
xA                       0
Yellow_Cards             0
Red_Cards                0
dtype: int64
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 532 entries, 0 to 531
```

```
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Name                    532 non-null    object
 1   Club                    532 non-null    object
 2   Nationality             532 non-null    object
 3   Position                532 non-null    object
 4   Age                     532 non-null    int64
 5   Matches                 532 non-null    int64
 6   Starts                  532 non-null    int64
 7   Mins                    532 non-null    int64
 8   Goals                   532 non-null    int64
 9   Assists                 532 non-null    int64
 10  Passes_Attempted        532 non-null    int64
 11  Perc_Passes_Completed   532 non-null    float64
 12  Penalty_Goals           532 non-null    int64
 13  Penalty_Attempted       532 non-null    int64
 14  xG                      532 non-null    float64
 15  xA                      532 non-null    float64
 16  Yellow_Cards            532 non-null    int64
 17  Red_Cards               532 non-null    int64
dtypes: float64(3), int64(11), object(4)
memory usage: 74.9+ KB
```

```
data.duplicated().sum()
```

```
0
```

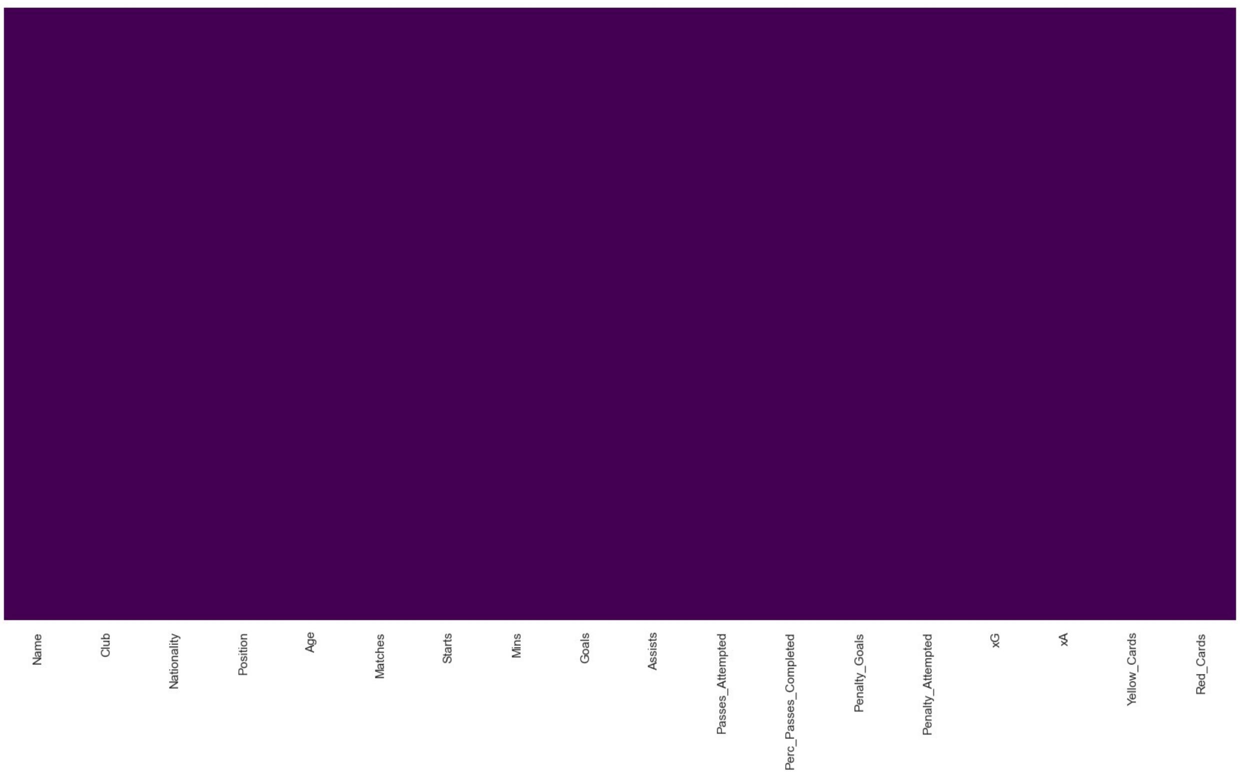
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 532 entries, 0 to 531
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Name                    532 non-null    object
 1   Club                    532 non-null    object
 2   Nationality             532 non-null    object
 3   Position                532 non-null    object
 4   Age                     532 non-null    int64
 5   Matches                 532 non-null    int64
 6   Starts                  532 non-null    int64
 7   Mins                    532 non-null    int64
 8   Goals                   532 non-null    int64
 9   Assists                 532 non-null    int64
 10  Passes_Attempted        532 non-null    int64
 11  Perc_Passes_Completed   532 non-null    float64
 12  Penalty_Goals           532 non-null    int64
 13  Penalty_Attempted       532 non-null    int64
 14  xG                      532 non-null    float64
```

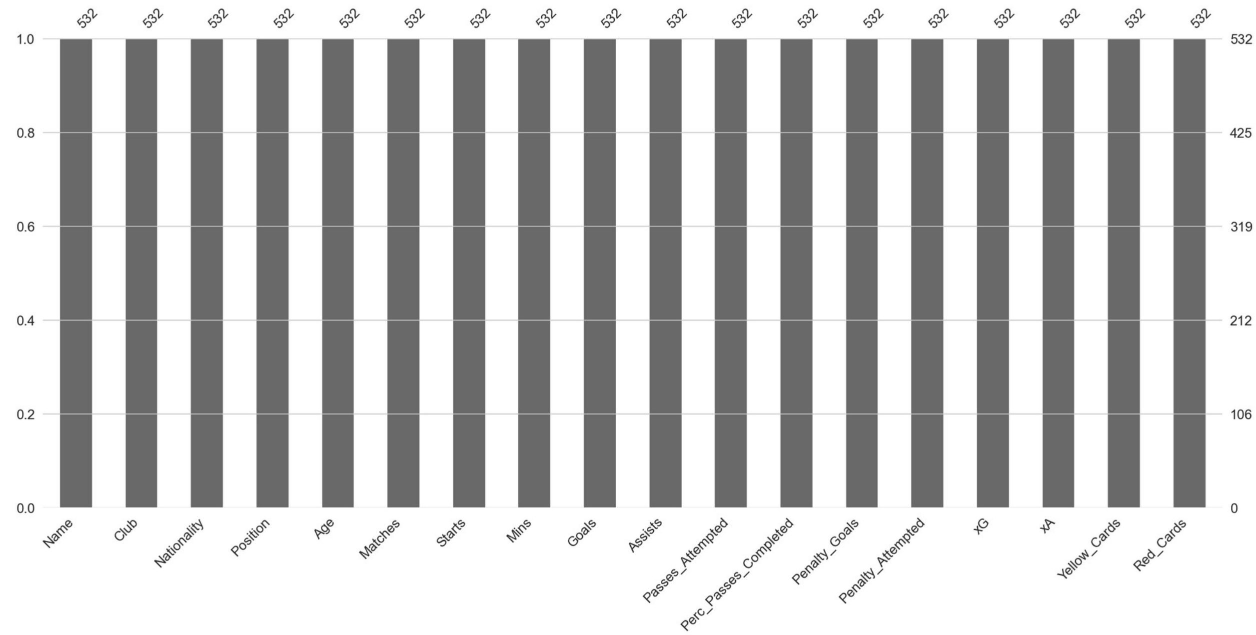
```
 15  xA                     532 non-null    float64
 16  Yellow_Cards           532 non-null    int64
 17  Red_Cards              532 non-null    int64
dtypes: float64(3), int64(11), object(4)
memory usage: 74.9+ KB
```

Data Visualization

```
sns.heatmap(data.isnull(),yticklabels=False,cbar=False,cmap='viridis')
plt.show()
```




```
msno.bar(data)

<AxesSubplot:>
```



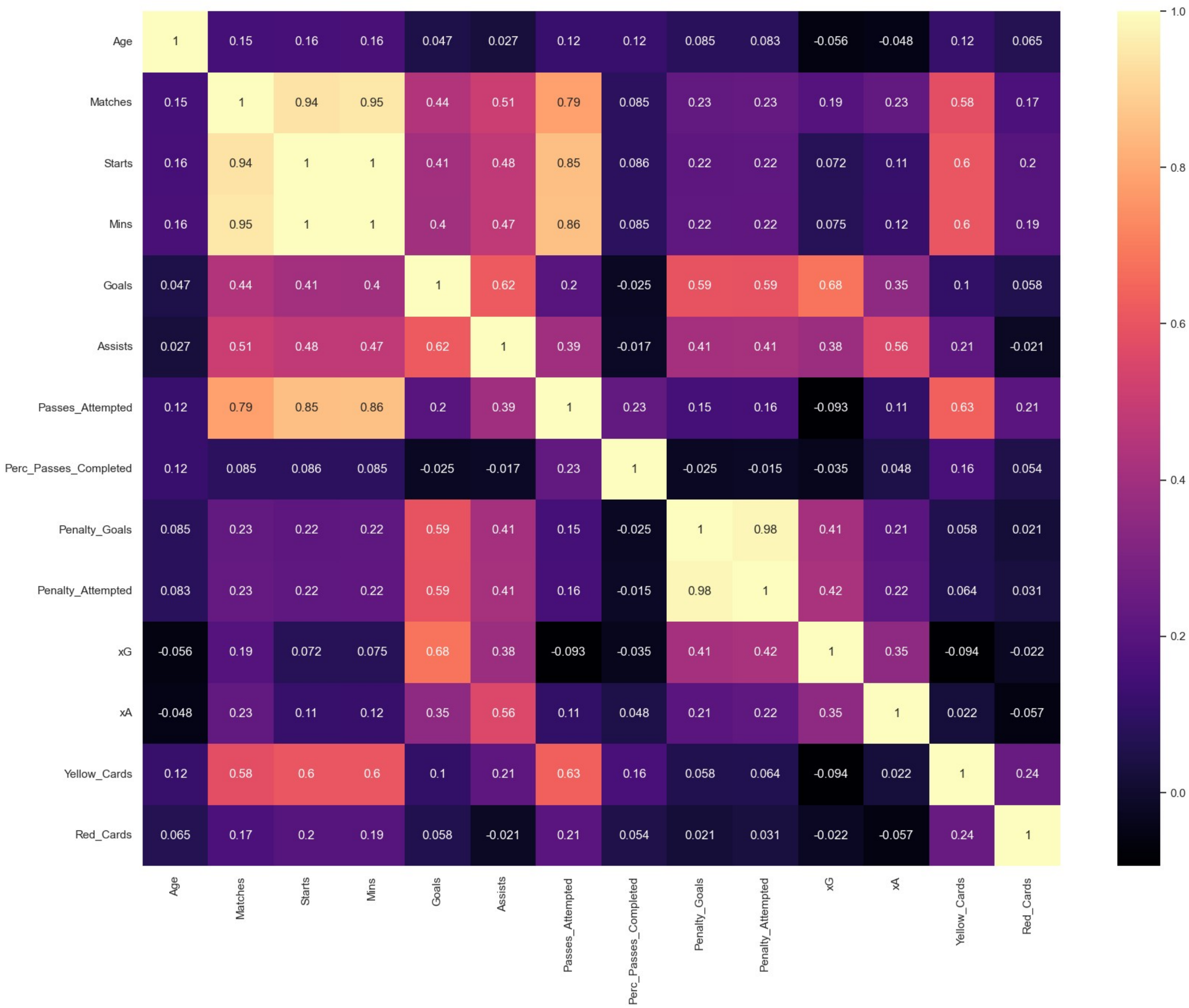

```
data.corr()
```

```
                         Age  Matches  Starts  Mins  Goals  Assists  \
Age                    1.000    0.150   0.156 0.159  0.047    0.027
Matches                0.150    1.000   0.938 0.947  0.442    0.508
Starts                 0.156    0.938   1.000 0.997  0.405    0.478
Mins                   0.159    0.947   0.997 1.000  0.400    0.475
Goals                  0.047    0.442   0.405 0.400  1.000    0.618
Assists                0.027    0.508   0.478 0.475  0.618    1.000
Passes_Attempted       0.120    0.785   0.853 0.856  0.202    0.394
Perc_Passes_Completed  0.123    0.085   0.086 0.085 -0.025   -0.017
Penalty_Goals          0.085    0.228   0.217 0.219  0.595    0.408
Penalty_Attempted      0.083    0.234   0.221 0.224  0.590    0.412
xG                    -0.056    0.192   0.072 0.075  0.682    0.383
xA                    -0.048    0.228   0.113 0.118  0.348    0.559
Yellow_Cards           0.115    0.578   0.603 0.605  0.104    0.214
Red_Cards              0.065    0.168   0.195 0.193  0.058   -0.021

                       Passes_Attempted  Perc_Passes_Completed
Penalty_Goals  \
Age                               0.120                  0.123
0.085
Matches                           0.785                  0.085
0.228
Starts                            0.853                  0.086
0.217
Mins                              0.856                  0.085
0.219
Goals                             0.202                 -0.025
0.595
```

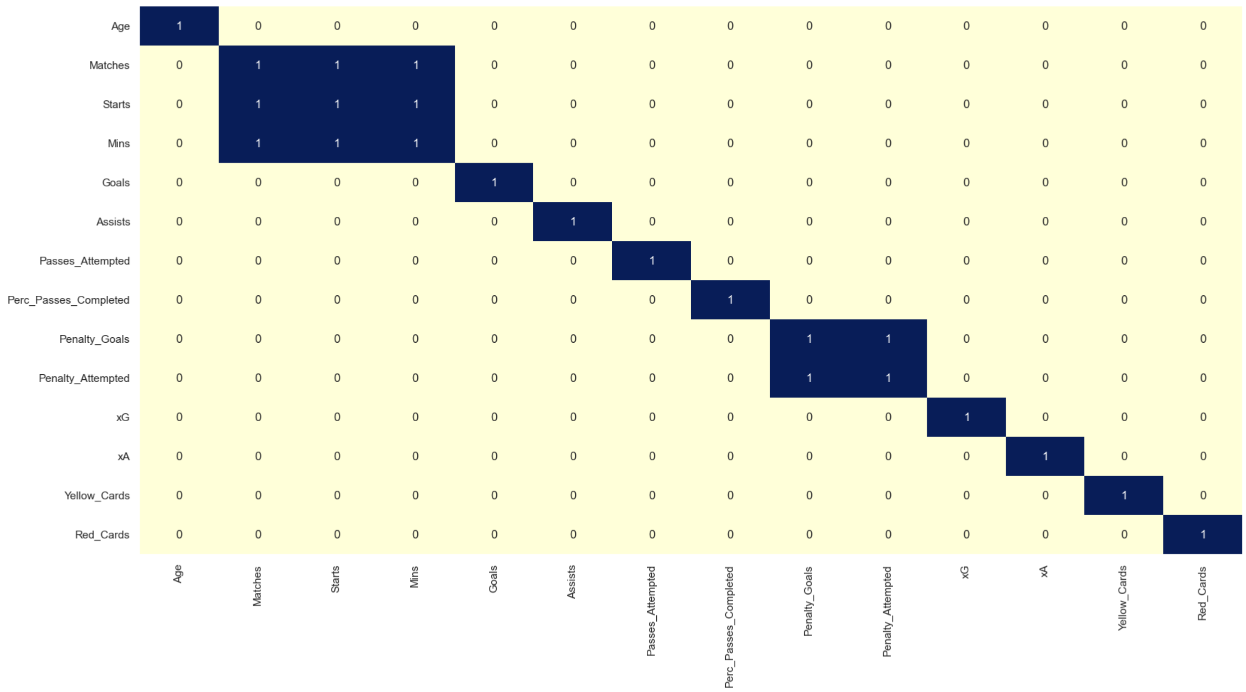
```
Assists                                     0.394                   -0.017
0.408
Passes_Attempted                            1.000                    0.228
0.150
Perc_Passes_Completed                       0.228                    1.000
-0.025
Penalty_Goals                               0.150                   -0.025
1.000
Penalty_Attempted                           0.157                   -0.015
0.982
xG                                         -0.093                   -0.035
0.407
xA                                          0.105                    0.048
0.205
Yellow_Cards                                0.635                    0.162
0.058
Red_Cards                                   0.208                    0.054
0.021

                       Penalty_Attempted     xG     xA
Yellow_Cards  \
Age                                0.083 -0.056 -0.048              0.115

Matches                            0.234  0.192  0.228              0.578

Starts                             0.221  0.072  0.113              0.603

Mins                               0.224  0.075  0.118              0.605

Goals                              0.590  0.682  0.348              0.104

Assists                            0.412  0.383  0.559              0.214

Passes_Attempted                   0.157 -0.093  0.105              0.635

Perc_Passes_Completed             -0.015 -0.035  0.048              0.162

Penalty_Goals                      0.982  0.407  0.205              0.058

Penalty_Attempted                  1.000  0.420  0.219              0.064

xG                                 0.420  1.000  0.347             -0.094

xA                                 0.219  0.347  1.000              0.022

Yellow_Cards                       0.064 -0.094  0.022              1.000

Red_Cards                          0.031 -0.022 -0.057              0.245


                       Red_Cards
```

```
Age                         0.065
Matches                     0.168
Starts                      0.195
Mins                        0.193
Goals                       0.058
Assists                    -0.021
Passes_Attempted            0.208
Perc_Passes_Completed       0.054
Penalty_Goals               0.021
Penalty_Attempted           0.031
xG                         -0.022
xA                         -0.057
Yellow_Cards                0.245
Red_Cards                   1.000
```

```
plt.figure(figsize=(20,15))
corr = data.corr()
sns.heatmap(data.corr(), cmap="magma", annot=True)
plt.show()
```




```
sns.heatmap(data.corr() > 0.9, annot=True, cbar=False,cmap="YlGnBu")
plt.show()
```

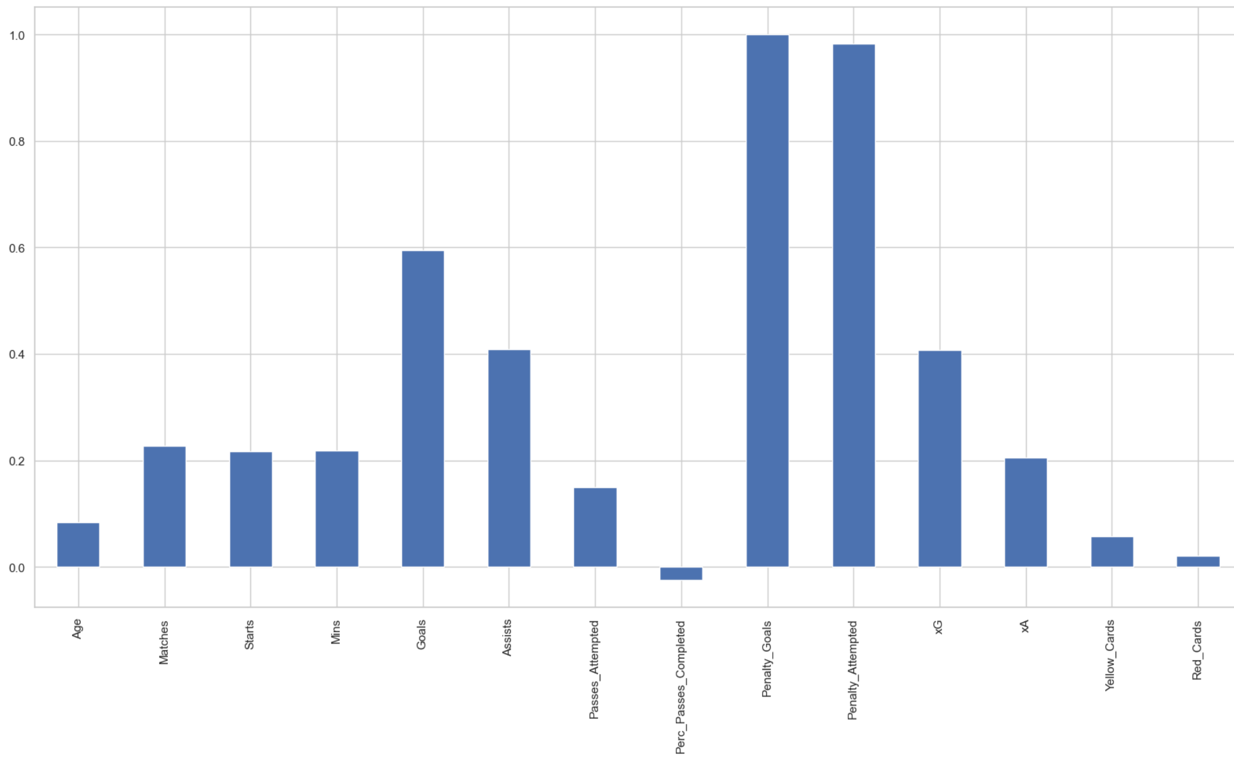



```
data.corr()['Goals']
```

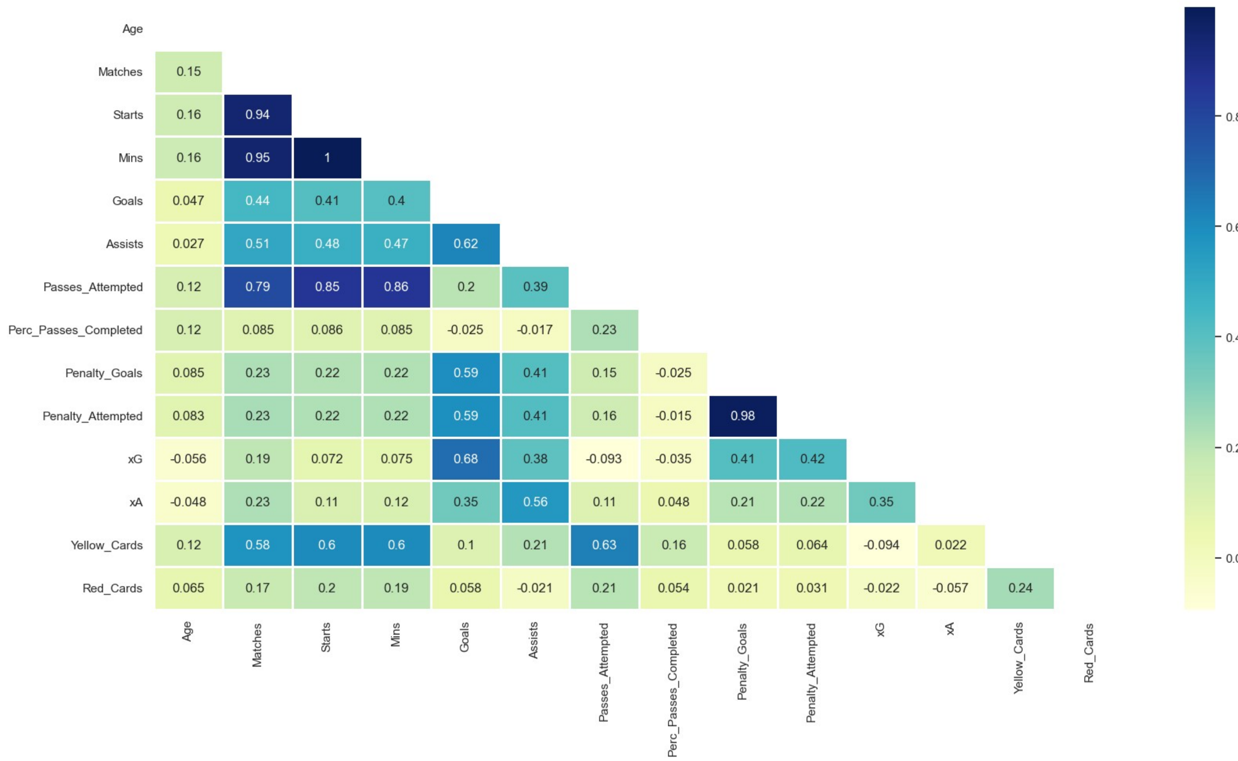
```
Age                      0.047
Matches                  0.442
Starts                   0.405
Mins                     0.400
Goals                    1.000
Assists                  0.618
Passes_Attempted         0.202
Perc_Passes_Completed   -0.025
Penalty_Goals            0.595
Penalty_Attempted        0.590
xG                       0.682
xA                       0.348
Yellow_Cards             0.104
Red_Cards                0.058
Name: Goals, dtype: float64
```

```
data.corr()['Penalty_Goals'].plot(kind='bar')
```

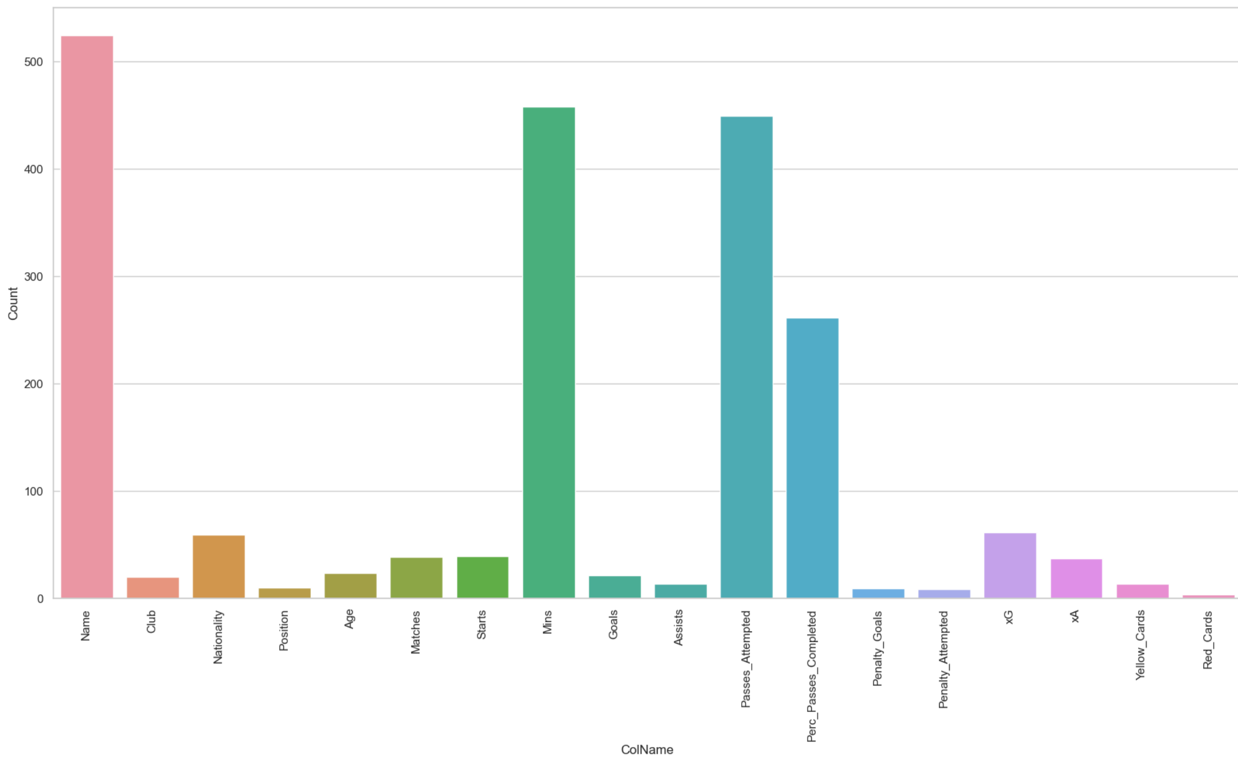
```
<AxesSubplot:>
```



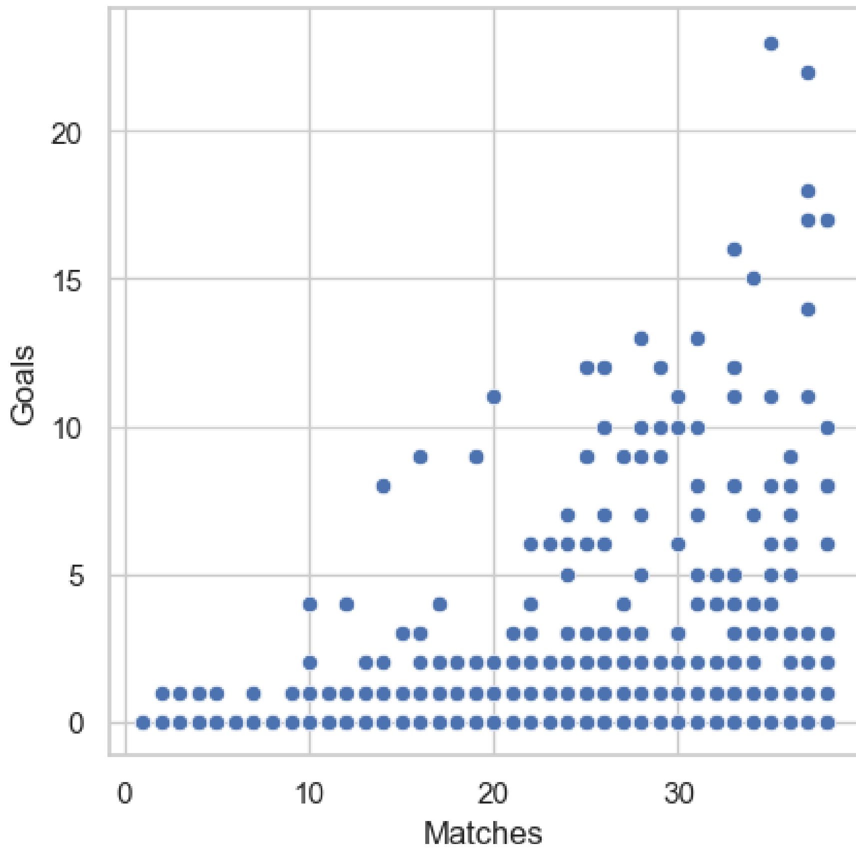

```
plt.figure(figsize=(20,10))
corr = data.corr()
mask=np.triu(np.ones_like(corr,dtype=bool))

sns.heatmap(data=corr, mask=mask,
cmap="YlGnBu",annot=True,linewidth=2)
plt.show()
```



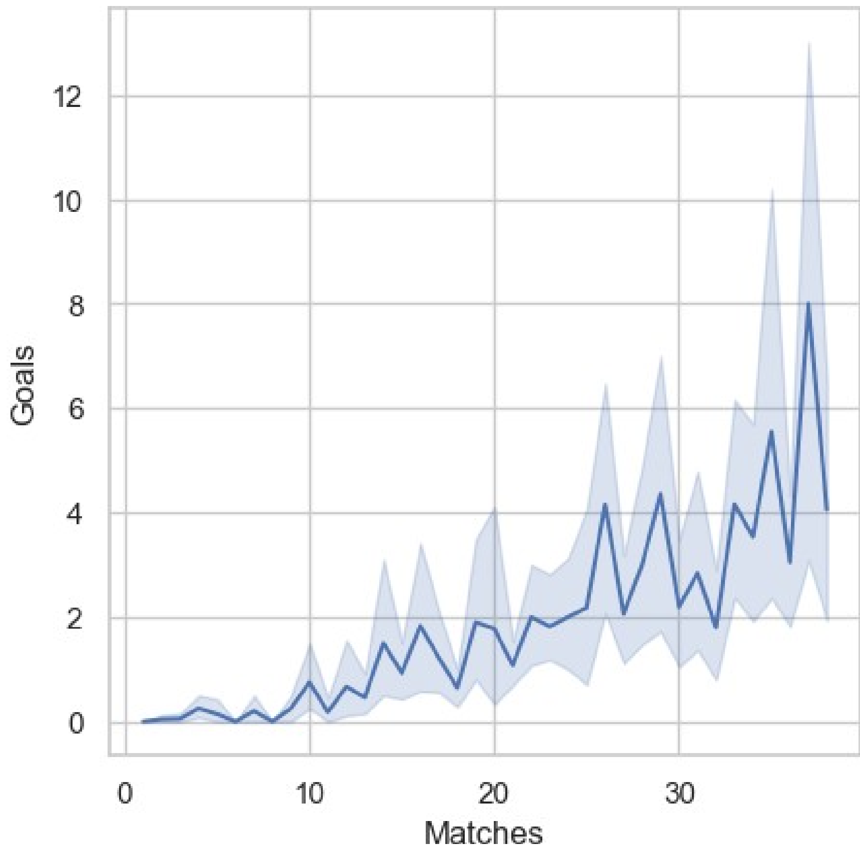

```
unique=data.nunique().to_frame()
unique.columns=['Count']
unique.index.names=['ColName']
unique=unique.reset_index()
sns.set(style='whitegrid',color_codes='True')
sns.barplot(x='ColName', y = 'Count', data = unique)
plt.xticks(rotation=90)
plt.show()
```




```
plt.figure(figsize=(5,5))
sns.scatterplot(x=data['Matches'],y=data['Goals'])
plt.show()
```

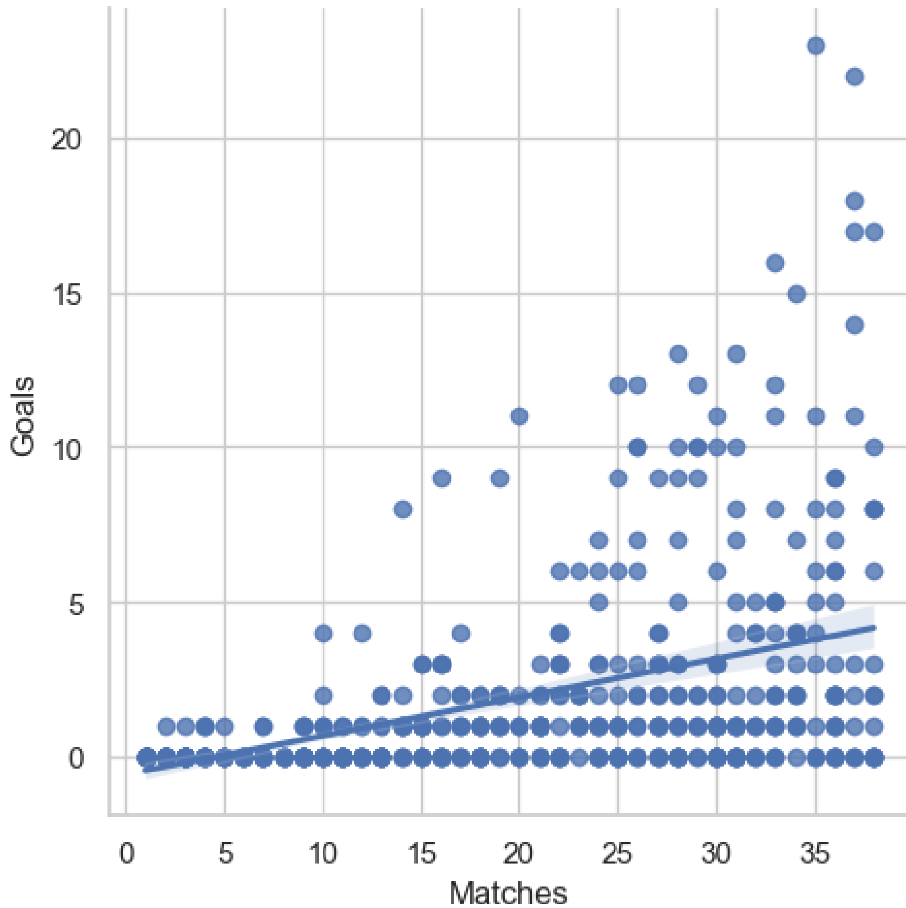



```
plt.figure(figsize=(5,5))
sns.lineplot(x=data['Matches'],y=data['Goals'])
plt.show()
```

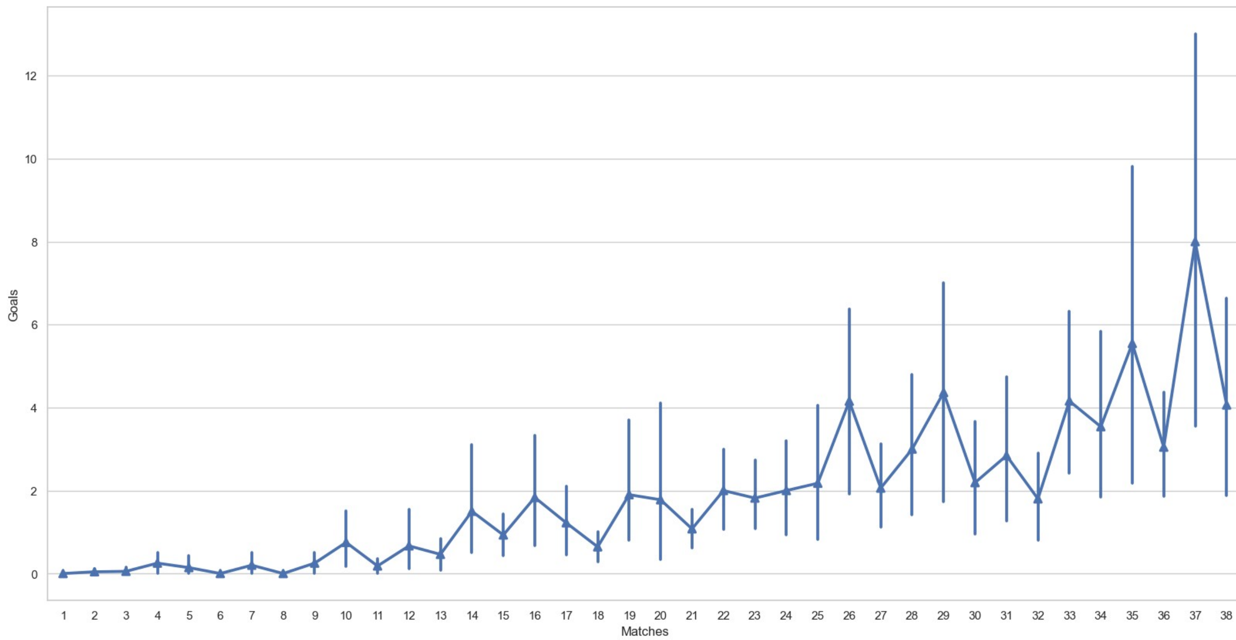



```
sns.lmplot(x='Matches', y='Goals', data=data)
```

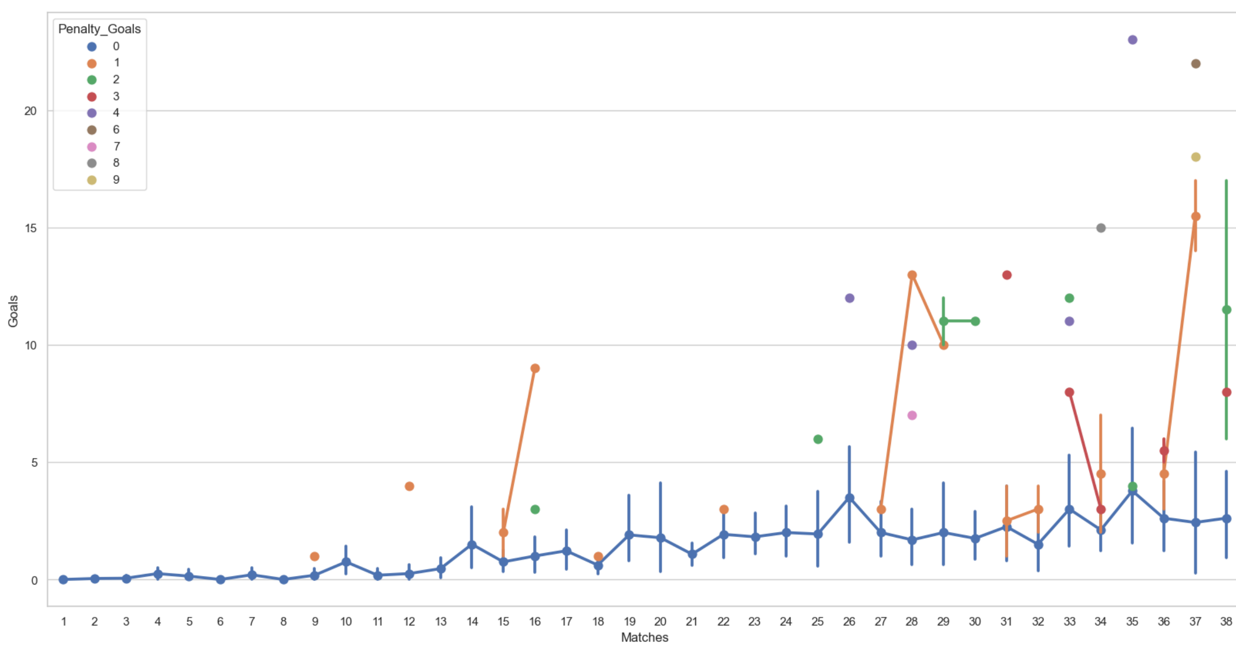
```
<seaborn.axisgrid.FacetGrid at 0x2494e2e4e50>
```



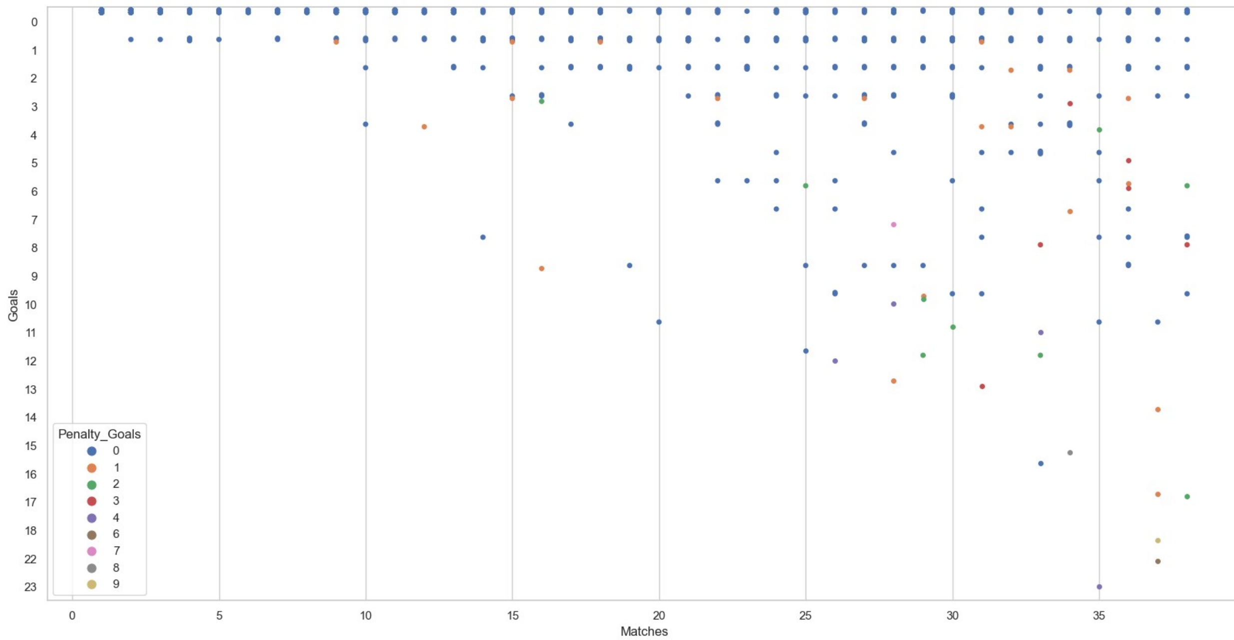

```
plt.figure(figsize=(20,10))
#sns.factorplot(x='Matches',y='Goals',data=data)
sns.pointplot(x='Matches',y='Goals',markers='^',data=data)
plt.show()
```



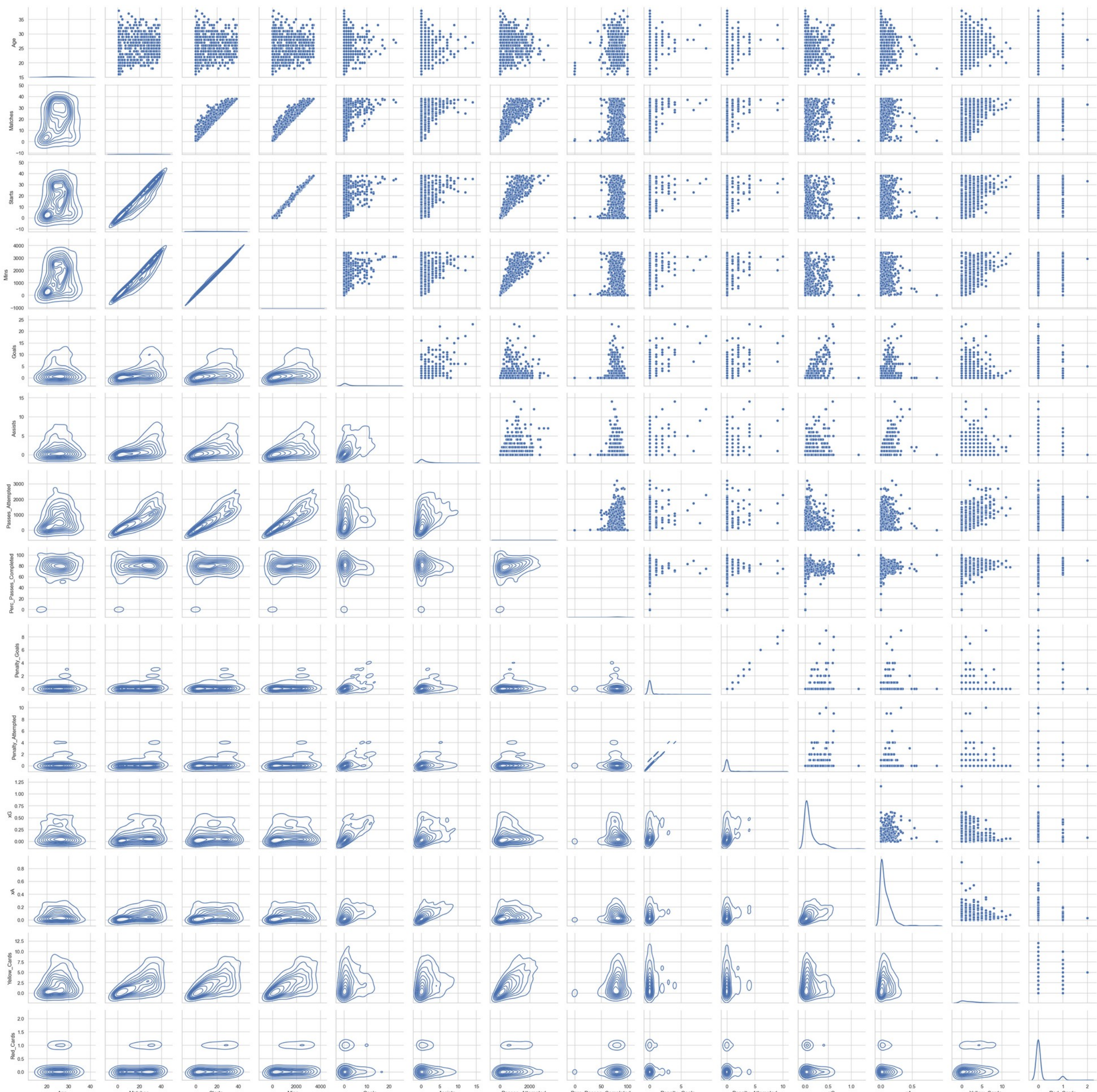

```
plt.figure(figsize=(20,10))

sns.pointplot(x='Matches',y='Goals',hue='Penalty_Goals',data=data)
plt.show()
```



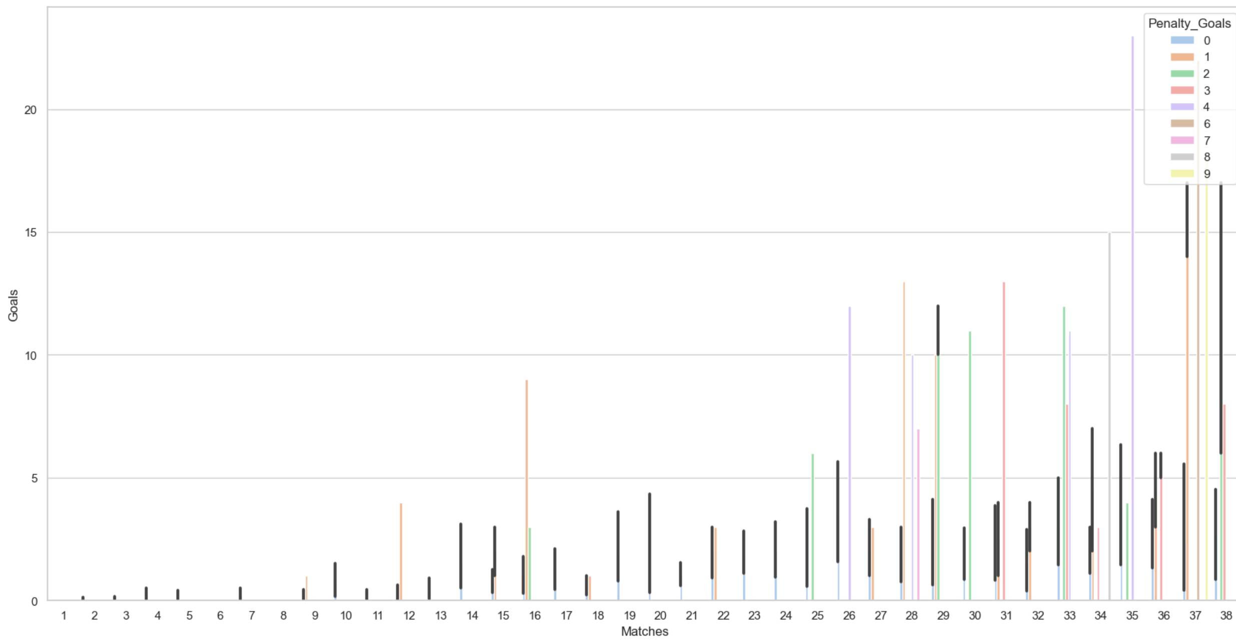

```
plt.figure(figsize=(20,10))

sns.swarmplot(x='Matches',y='Goals',hue='Penalty_Goals',orient='h',dat
a=data,dodge=True)
plt.show()
```




```
graph=sns.PairGrid(data)

graph=graph.map_upper(sns.scatterplot)

graph=graph.map_lower(sns.kdeplot)

graph=graph.map_diag(sns.kdeplot,lw=2)

plt.show()
#kernel distribution estimation plot
```



```
plt.figure(figsize=(20,10))
sns.barplot(x='Matches',y='Goals',hue='Penalty_Goals',palette='pastel'
,data=data)
plt.show()
```




```
data['MinsPerMatch']=(data['Mins']/data['Matches']).astype(int)

data['GoalsPerMatch']=(data['Goals']/data['Matches']).astype(float)

data.head()

ColName            Name     Club Nationality Position  Age  Matches
Starts  \
0          Mason Mount  Chelsea         ENG    MF,FW   21       36
32
1        Edouard Mendy  Chelsea         SEN       GK   28       31
31
2          Timo Werner  Chelsea         GER       FW   24       35
29
3         Ben Chilwell  Chelsea         ENG       DF   23       27
27
4          Reece James  Chelsea         ENG       DF   20       32
25

ColName  Mins  Goals  Assists  Passes_Attempted  Perc_Passes_Completed
\
0        2890      6        5              1881                 82.300

1        2745      0        0              1007                 84.600

2        2602      6        8               826                 77.200

3        2286      3        5              1806                 78.600

4        2373      1        2              1987                 85.000
```

```
ColName  Penalty_Goals  Penalty_Attempted    xG    xA  Yellow_Cards  \
0                   1                  1 0.210 0.240             2
1                   0                  0 0.000 0.000             2
2                   0                  0 0.410 0.210             2
3                   0                  0 0.100 0.110             3
4                   0                  0 0.060 0.120             3

ColName  Red_Cards  MinsPerMatch  GoalsPerMatch
0                0            80          0.167
1                0            88          0.000
2                0            74          0.171
3                0            84          0.111
4                0            74          0.031
```

```
#total Goals
TotalGoals=data['Goals'].sum()
TotalGoals
```

```
986
```

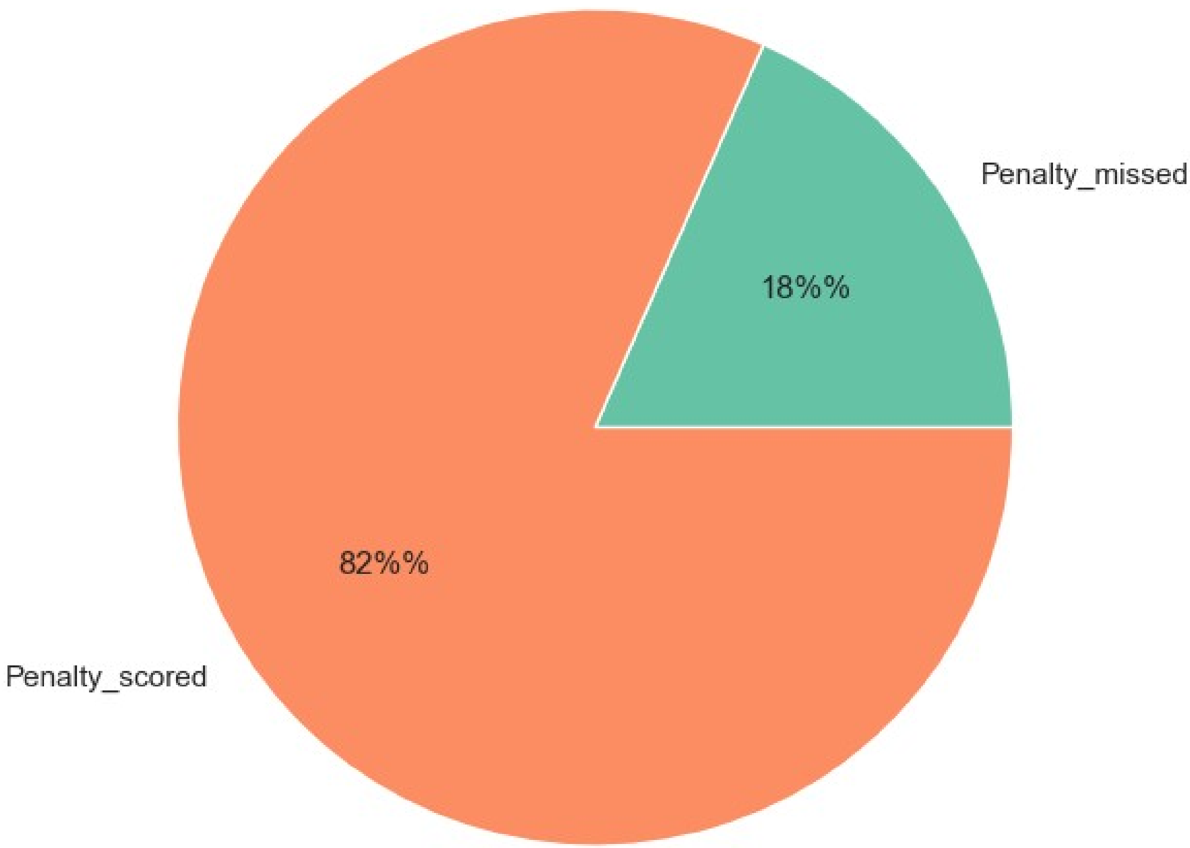
```
#Penalty Goals

Total_PenaltyGoals=data['Penalty_Goals'].sum()
Total_PenaltyGoals
```

```
102
```

```
#Penalty Attempts
Total_PenaltyAttempts=data['Penalty_Attempted'].sum()
Total_PenaltyAttempts
```

```
125
```

```
#Pie chart for penalty missed and scored

plt.figure(figsize=(7,7))
Penalty_not_scored=data['Penalty_Attempted'].sum()- Total_PenaltyGoals
d=[Penalty_not_scored,Total_PenaltyGoals]
labels=['Penalty_missed','Penalty_scored']
color=sns.color_palette('Set2')
plt.pie(d,labels=labels,colors=color,autopct='%1.0f%%%%')
plt.show()
```




```
#Unique Positions

data['Position'].unique()

array(['MF,FW', 'GK', 'FW', 'DF', 'MF', 'FW,MF', 'FW,DF', 'DF,MF',
       'MF,DF', 'DF,FW'], dtype=object)

#Total Forward Players

data[data['Position']=='FW']

ColName                Name              Club Nationality Position
Age  \
2               Timo Werner           Chelsea         GER       FW
24
16            Tammy Abraham           Chelsea         ENG       FW
22
19           Olivier Giroud           Chelsea         FRA       FW
```

```
33
23       Ruben Loftus-Cheek          Chelsea         ENG       FW
24
30         Raheem Sterling   Manchester City         ENG       FW
25
..                     ...               ...         ...      ...  ..
.
516           Oliver Burke  Sheffield United         SCO       FW
23
518         Oliver McBurnie  Sheffield United         SCO       FW
24
519         Rhian Brewster  Sheffield United         ENG       FW
20
523            Billy Sharp  Sheffield United         ENG       FW
34
526        Daniel Jebbison  Sheffield United         ENG       FW
17

ColName  Matches  Starts  Mins  Goals  Assists  Passes_Attempted  \
2             35      29  2602      6        8               826
16            22      12  1040      6        1               218
19            17       8   748      4        0               217
23             1       1    60      0        0                16
30            31      28  2536     10        7              1127
..           ...     ...   ...    ...      ...               ...
516           25      14  1269      1        1               262
518           23      12  1324      1        0               426
519           27      12  1128      0        0               225
523           16       7   735      3        0               123
526            4       3   284      1        0                34

ColName  Perc_Passes_Completed  Penalty_Goals  Penalty_Attempted    xG
xA  \
2                       77.200              0                  0 0.410
0.210
16                      68.300              0                  0 0.560
0.070
19                      74.200              0                  0 0.580
0.090
23                      68.800              0                  0 0.000
0.000
30                      85.400              0                  1 0.430
0.170
..                         ...            ...                ...   ...
...
516                     70.600              0                  0 0.170
0.130
518                     62.900              0                  0 0.210
0.070
```

```
519                        69.300            0                    0 0.140
0.130
523                        69.900            2                    2 0.330
0.070
526                        70.600            0                    0 0.500
0.010

ColName  Yellow_Cards  Red_Cards  MinsPerMatch  GoalsPerMatch
2                   2          0            74          0.171
16                  0          0            47          0.273
19                  1          0            44          0.235
23                  0          0            60          0.000
30                  4          0            81          0.323
..                ...        ...           ...            ...
516                 2          0            50          0.040
518                 2          0            57          0.043
519                 1          0            41          0.000
523                 1          0            45          0.188
526                 0          0            71          0.250

[81 rows x 20 columns]
```

```
#Total Goal Keepers

data[data['Position']=='GK']
```

```
ColName                        Name                     Club Nationality
Position  \
1                  Edouard Mendy                  Chelsea         SEN
GK
20             Kepa Arrizabalaga                  Chelsea         ESP
GK
22               Willy Caballero                  Chelsea         ARG
GK
27                      Ederson          Manchester City         BRA
GK
48                 Scott Carson          Manchester City         ENG
GK
49                 Zack Steffen          Manchester City         USA
GK
58                 David de Gea        Manchester United         ESP
GK
64               Dean Henderson        Manchester United         ENG
GK
84                      Alisson             Liverpool FC         BRA
GK
101                      Adrián             Liverpool FC         ESP
GK
106           Caoimhín Kelleher             Liverpool FC         IRL
GK
```

108	Kasper Schmeichel	Leicester City	DEN	GK
137	Łukasz Fabiański	West Ham United	POL	GK
155	Darren Randolph	West Ham United	IRL	GK
160	Hugo Lloris	Tottenham Hotspur	FRA	GK
183	Bernd Leno	Arsenal	GER	GK
206	Mathew Ryan	Arsenal	AUS	GK
211	Rúnar Alex Rúnarsson	Arsenal	ISL	GK
215	Illan Meslier	Leeds United	FRA	GK
230	Kiko Casilla	Leeds United	ESP	GK
238	Jordan Pickford	Everton	ENG	GK
251	Robin Olsen	Everton	SWE	GK
259	João Virgínia	Everton	POR	GK
264	Emiliano Martínez	Aston Villa	ARG	GK
290	Karl Darlow	Newcastle United	ENG	GK
306	Martin Dúbravka	Newcastle United	SVK	GK
315	Rui Patrício	Wolverhampton Wanderers	POR	GK
338	John Ruddy	Wolverhampton Wanderers	ENG	GK
342	Vicente Guaita	Crystal Palace	ESP	GK
364	Jack Butland	Crystal Palace	ENG	GK
369	Alex McCarthy	Southampton	ENG	GK
383	Fraser Forster	Southampton	ENG	GK
402	Robert Sánchez	Brighton	ESP	GK
409	Mathew Ryan	Brighton	AUS	GK
426	Nick Pope	Burnley	ENG	GK
439	Bailey Peacock-Farrell	Burnley	NIR	

```
GK
444                 Will Norris                  Burnley         ENG
GK
447             Alphonse Areola                   Fulham         FRA
GK
470                 Marek Rodák                   Fulham         SVK
GK
475               Sam Johnstone     West Bromwich Albion         ENG
GK
500                David Button     West Bromwich Albion         ENG
GK
505              Aaron Ramsdale         Sheffield United         ENG
GK

ColName  Age  Matches  Starts  Mins  Goals  Assists  Passes_Attempted
\
1         28       31      31  2745      0        0              1007

20        25        7       6   585      0        0               243

22        38        1       1    90      0        0                26

27        26       36      36  3240      0        1              1090

48        34        1       1    90      0        0                16

49        25        1       1    90      0        0                28

58        29       26      26  2295      0        0               594

64        23       13      12  1125      0        0               314

84        27       33      33  2970      1        0              1137

101       33        3       3   270      0        0                99

106       21        2       2   180      0        0                62

108       33       38      38  3420      0        0              1218

137       35       35      35  3150      0        0              1002

155       33        3       3   270      0        0                66

160       33       38      38  3420      0        0              1067

183       28       35      35  3131      0        0              1156

206       28        3       3   270      0        0                67

211       25        1       0    16      0        0                11
```

```
215       20        35       35  3150      0        0                1348
230       33         3        3   270      0        0                  94
238       26        31       31  2742      0        0                1152
251       30         7        7   630      0        0                 199
259       20         1        0    48      0        0                  17
264       27        38       38  3420      0        0                1295
290       29        25       25  2250      0        0                 726
306       31        13       13  1170      0        0                 427
315       32        37       37  3329      0        0                 801
338       33         2        1    91      0        0                  24
342       33        37       37  3330      0        0                1080
364       27         1        1    90      0        0                  21
369       30        30       30  2700      0        0                1069
383       32         8        8   720      0        0                 274
402       22        27       27  2430      0        0                1095
409       28        11       11   990      0        0                 399
426       28        32       32  2880      0        0                 979
439       23         4        4   360      0        0                 113
444       26         2        2   180      0        0                  56
447       27        36       36  3240      0        0                1001
470       23         2        2   180      0        0                  46
475       27        37       37  3330      0        1                1282
500       31         1        1    90      0        0                  37
505       22        38       38  3420      0        0                1141

ColName  Perc_Passes_Completed  Penalty_Goals  Penalty_Attempted    xG
xA  \
```

1	84.600	0	0	0.000	0.000
20	81.500	0	0	0.000	0.000
22	92.300	0	0	0.000	0.000
27	83.100	0	0	0.000	0.010
48	93.800	0	0	0.000	0.000
49	82.100	0	0	0.000	0.000
58	77.100	0	0	0.000	0.000
64	75.200	0	0	0.000	0.000
84	85.200	0	0	0.000	0.000
101	76.800	0	0	0.000	0.000
106	82.300	0	0	0.000	0.000
108	72.700	0	0	0.000	0.000
137	60.500	0	0	0.000	0.000
155	54.500	0	0	0.000	0.000
160	71.500	0	0	0.000	0.000
183	79.800	0	0	0.000	0.000
206	92.500	0	0	0.000	0.000
211	63.600	0	0	0.000	0.000
215	80.900	0	0	0.000	0.000
230	87.200	0	0	0.000	0.000
238	66.100	0	0	0.000	0.000
251	71.400	0	0	0.000	0.000
259	52.900	0	0	0.000	0.000
264	65.600	0	0	0.000	0.010
290	50.100	0	0	0.000	

```
0.000
306                     72.600           0                   0 0.000
0.000
315                     66.700           0                   0 0.000
0.000
338                     79.200           0                   0 0.000
0.000
342                     55.400           0                   0 0.000
0.000
364                     28.600           0                   0 0.000
0.000
369                     64.300           0                   0 0.000
0.000
383                     56.200           0                   0 0.000
0.000
402                     71.600           0                   0 0.000
0.000
409                     78.700           0                   0 0.000
0.000
426                     50.700           0                   0 0.000
0.010
439                     51.300           0                   0 0.000
0.000
444                     48.200           0                   0 0.000
0.040
447                     73.600           0                   0 0.000
0.000
470                     80.400           0                   0 0.000
0.000
475                     49.900           0                   0 0.000
0.010
500                     43.200           0                   0 0.000
0.000
505                     49.100           0                   0 0.000
0.000

ColName  Yellow_Cards  Red_Cards  MinsPerMatch  GoalsPerMatch
1                   2          0            88          0.000
20                  1          0            83          0.000
22                  0          0            90          0.000
27                  3          0            90          0.000
48                  0          0            90          0.000
49                  0          0            90          0.000
58                  0          0            88          0.000
64                  3          0            86          0.000
84                  1          0            90          0.030
101                 0          0            90          0.000
106                 0          0            90          0.000
108                 0          0            90          0.000
```

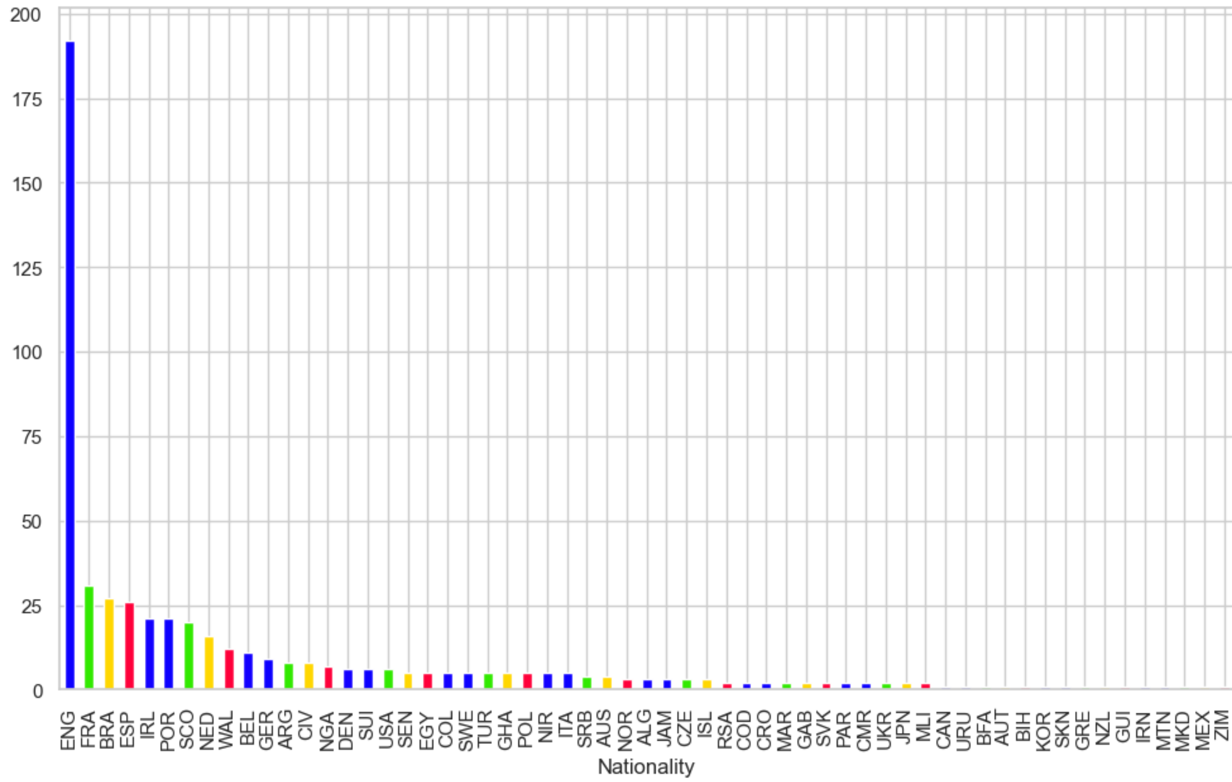
137	2	0	90	0.000
155	0	0	90	0.000
160	0	0	90	0.000
183	0	1	89	0.000
206	0	0	90	0.000
211	0	0	16	0.000
215	0	0	90	0.000
230	0	0	90	0.000
238	1	0	88	0.000
251	1	0	90	0.000
259	0	0	48	0.000
264	1	0	90	0.000
290	3	0	90	0.000
306	0	0	90	0.000
315	1	0	89	0.000
338	0	0	45	0.000
342	2	0	90	0.000
364	0	0	90	0.000
369	2	0	90	0.000
383	0	0	90	0.000
402	2	0	90	0.000
409	1	0	90	0.000
426	1	0	90	0.000
439	0	0	90	0.000
444	0	0	90	0.000
447	2	0	90	0.000
470	0	0	90	0.000
475	1	0	90	0.000
500	0	0	90	0.000
505	1	0	90	0.000

```
#Players from Different Nation

np.size((data['Nationality'].unique()))
```

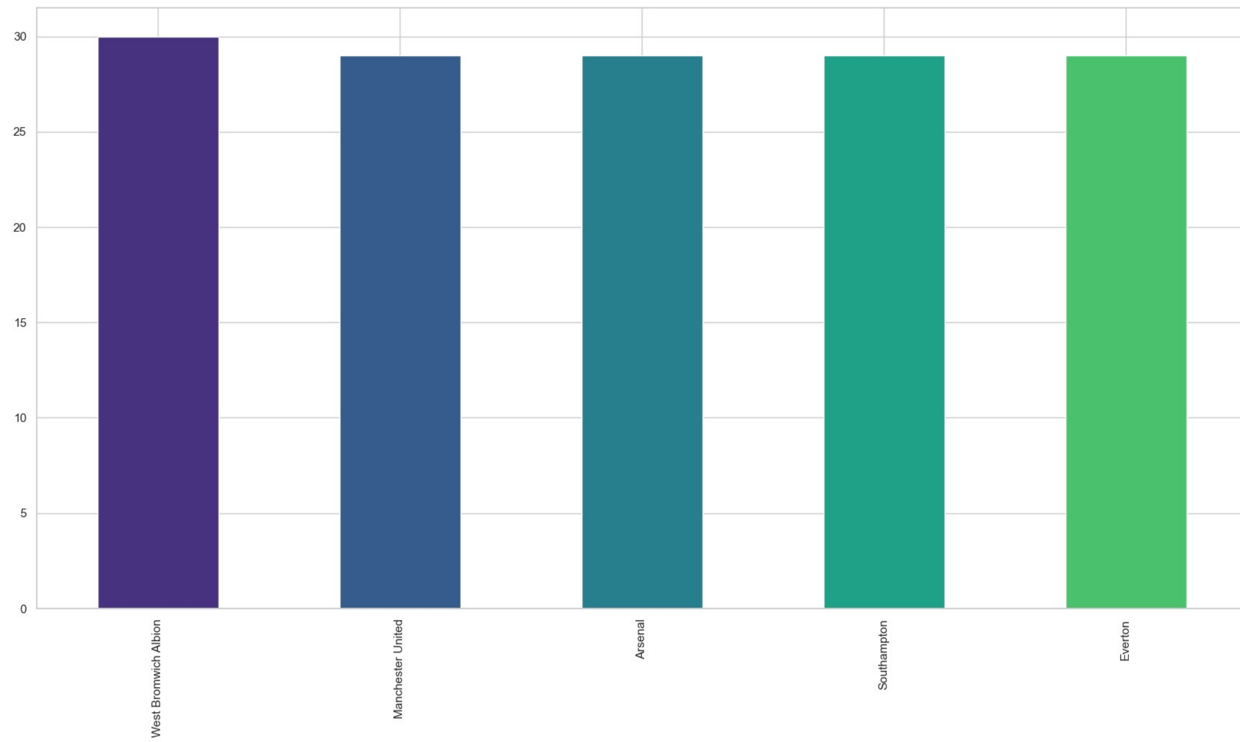
```
59
```

```
#Most players came from which countries

Nationality=data.groupby('Nationality').size().sort_values(ascending=F
alse)
Nationality.plot(kind='bar',figsize=(12,7),color=sns.color_palette('pr
ism',5))
```

```
<AxesSubplot:xlabel='Nationality'>
```

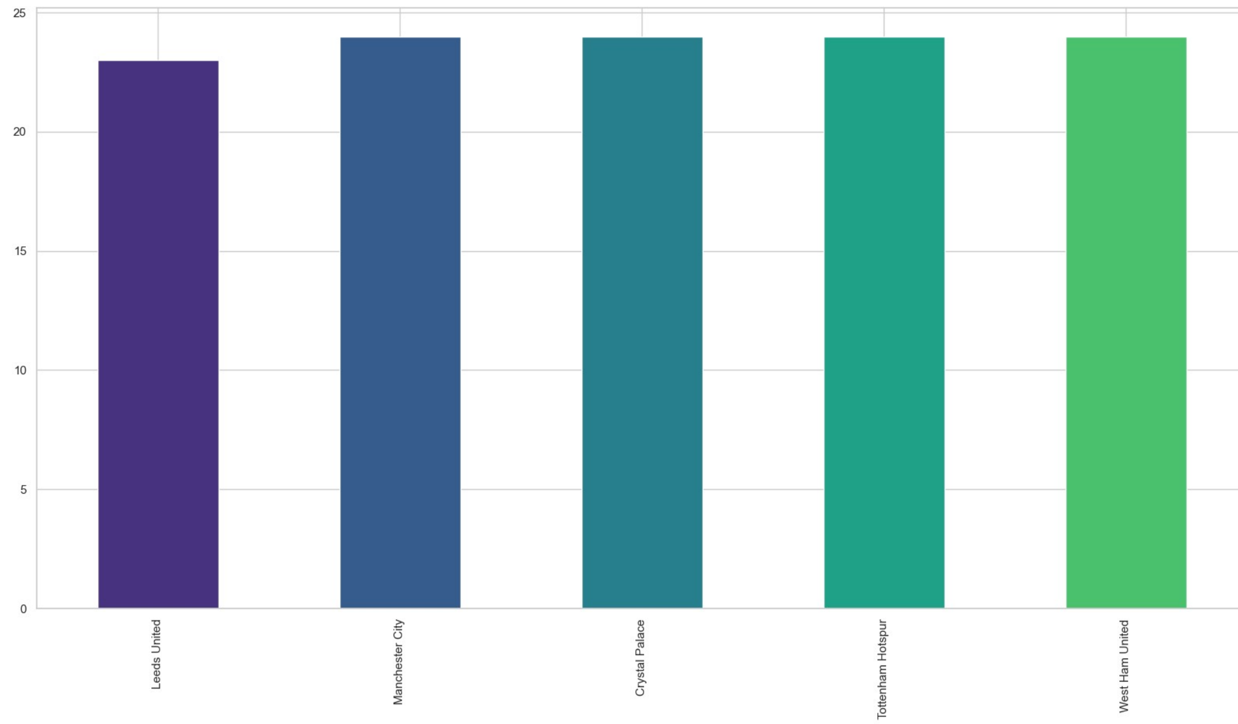



```
#Clubs with maximum players in their squard
data['Club'].value_counts().nlargest(5).plot(kind='bar',color=sns.colo
r_palette('viridis'))

#nlargest=head=top5

<AxesSubplot:>
```

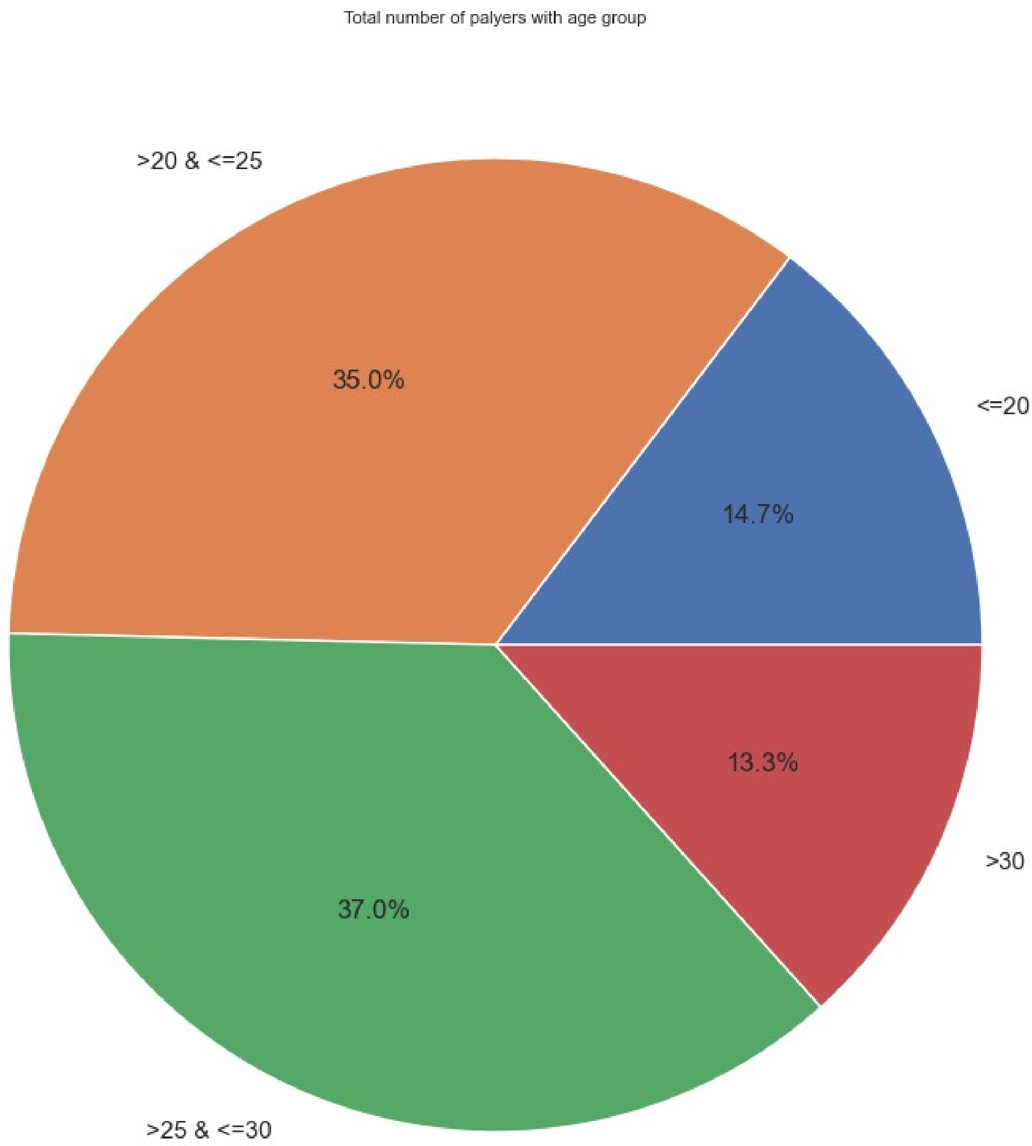



```
#club with least player in their squad

data['Club'].value_counts().nsmallest(5).plot(kind='bar',color=sns.col
or_palette('viridis'))

#nsmallest=tail=last5
```

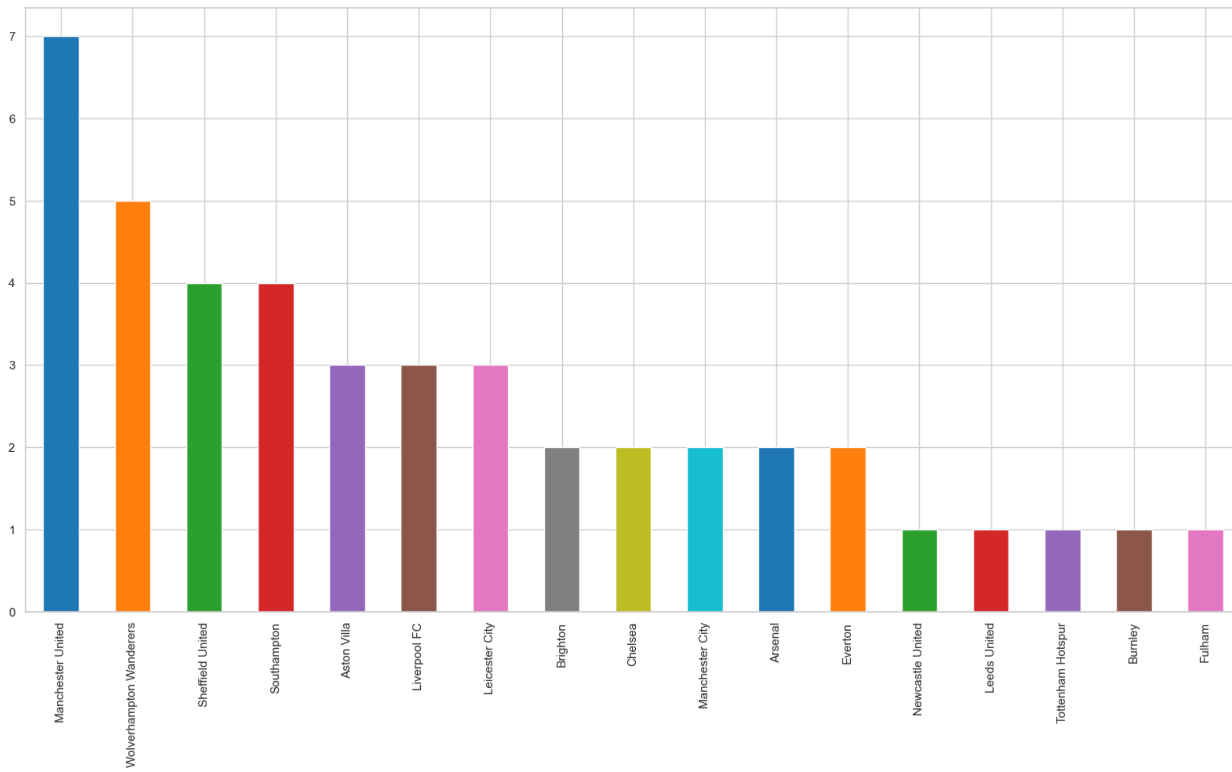
```
<AxesSubplot:>
```




```
#players based on age group

under20=data[data['Age']<=20]

age20_25=data[(data["Age"] > 20) & (data["Age"] <= 25)]

age25_30=data[(data["Age"] > 25) & (data["Age"] <= 30)]

above30=data[(data["Age"] > 30)]

x=np.array([under20['Name'].count(),age20_25['Name'].count(),age25_30[
'Name'].count(),above30['Name'].count()])
labels=['<=20','>20 & <=25','>25 & <=30', '>30']
plt.title('Total number of palyers with age group', fontsize=8)
plt.pie(x,labels=labels,autopct='%.1f%%')
plt.show()
```




```
#Total under20 players in each club

Players_under20=data[data['Age'] < 20]

Players_under20['Club'].value_counts().plot(kind='bar',color=sns.color
_palette('tab10'))

<AxesSubplot:>
```



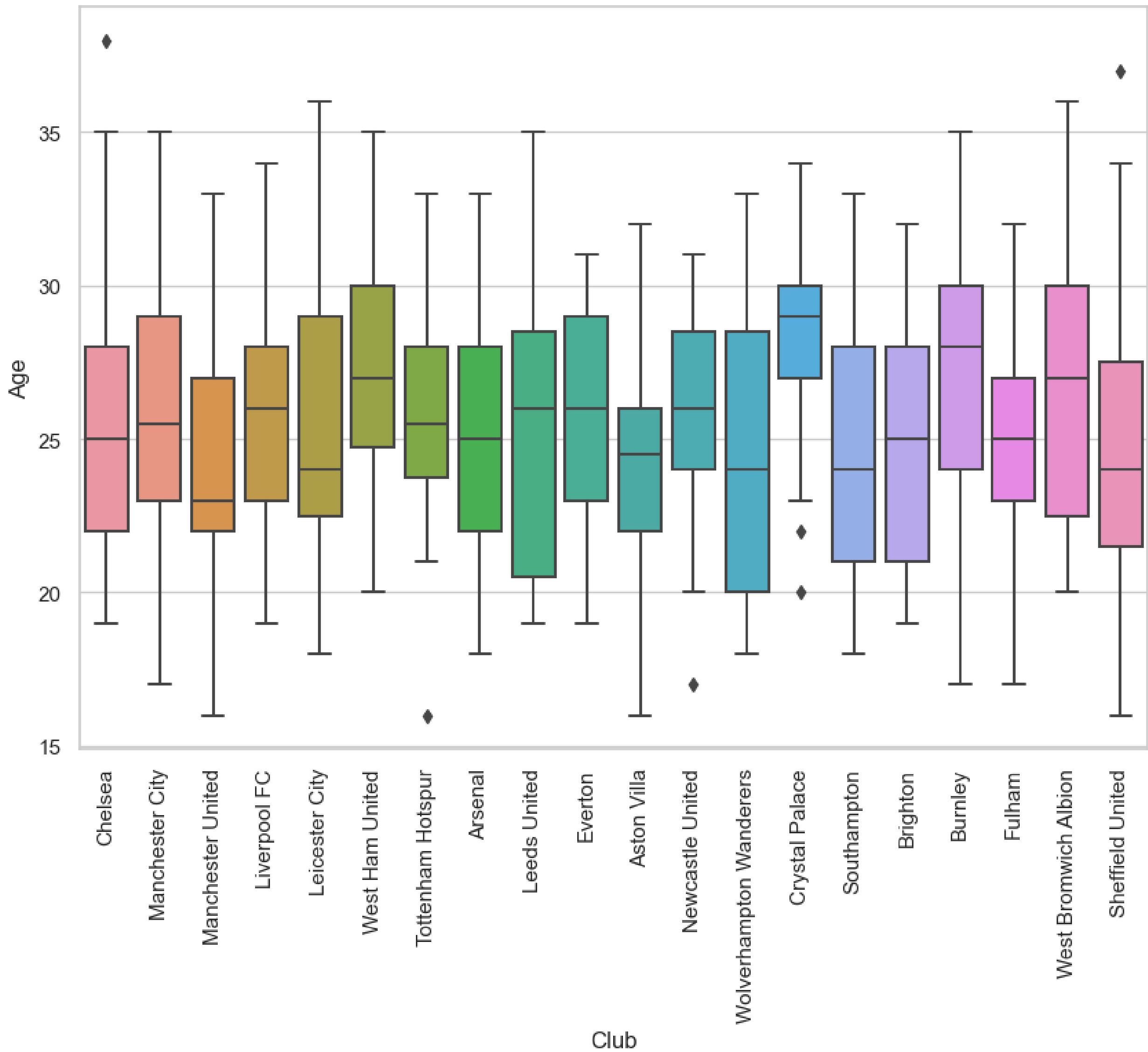

```
# Under 20 players in manchester city

Players_under20[Players_under20['Club']=='Manchester City']

ColName         Name             Club Nationality Position  Age
Matches  \
47       Eric García  Manchester City         ESP       DF   19
6
50        Liam Delap  Manchester City         ENG       FW   17
1

ColName  Starts  Mins  Goals  Assists  Passes_Attempted  \
47            3   383      0        0               344
50            0    40      0        0                 7

ColName  Perc_Passes_Completed  Penalty_Goals  Penalty_Attempted    xG
xA  \
47                      93.600              0                  0 0.030
0.020
50                      71.400              0                  0 0.060
0.000

ColName  Yellow_Cards  Red_Cards  MinsPerMatch  GoalsPerMatch
47                  0          0            63          0.000
50                  0          0            40          0.000

Players_under20[Players_under20['Club']=='Chelsea']
```

```
ColName                 Name     Club Nationality Position  Age
Matches  \
18       Callum Hudson-Odoi  Chelsea         ENG    FW,DF   19
23
21            Billy Gilmour  Chelsea         SCO       MF   19
5

ColName  Starts  Mins  Goals  Assists  Passes_Attempted  \
18           10  1059      2        3               659
21            3   261      0        0               215

ColName  Perc_Passes_Completed  Penalty_Goals  Penalty_Attempted    xG
xA  \
18                      82.200              0                  0 0.120
0.260
21                      89.300              0                  0 0.010
0.040

ColName  Yellow_Cards  Red_Cards  MinsPerMatch  GoalsPerMatch
18                  0          0            46          0.087
21                  0          0            52          0.000
```

```
#average age of players in each club

plt.figure(figsize=(10,7))
sns.boxplot(x='Club', y='Age', data=data)
plt.xticks(rotation=90)
```

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14,
15, 16,
        17, 18, 19]),
 [Text(0, 0, 'Chelsea'),
  Text(1, 0, 'Manchester City'),
  Text(2, 0, 'Manchester United'),
  Text(3, 0, 'Liverpool FC'),
  Text(4, 0, 'Leicester City'),
  Text(5, 0, 'West Ham United'),
  Text(6, 0, 'Tottenham Hotspur'),
  Text(7, 0, 'Arsenal'),
  Text(8, 0, 'Leeds United'),
  Text(9, 0, 'Everton'),
  Text(10, 0, 'Aston Villa'),
  Text(11, 0, 'Newcastle United'),
  Text(12, 0, 'Wolverhampton Wanderers'),
  Text(13, 0, 'Crystal Palace'),
  Text(14, 0, 'Southampton'),
  Text(15, 0, 'Brighton'),
  Text(16, 0, 'Burnley'),
  Text(17, 0, 'Fulham'),
```

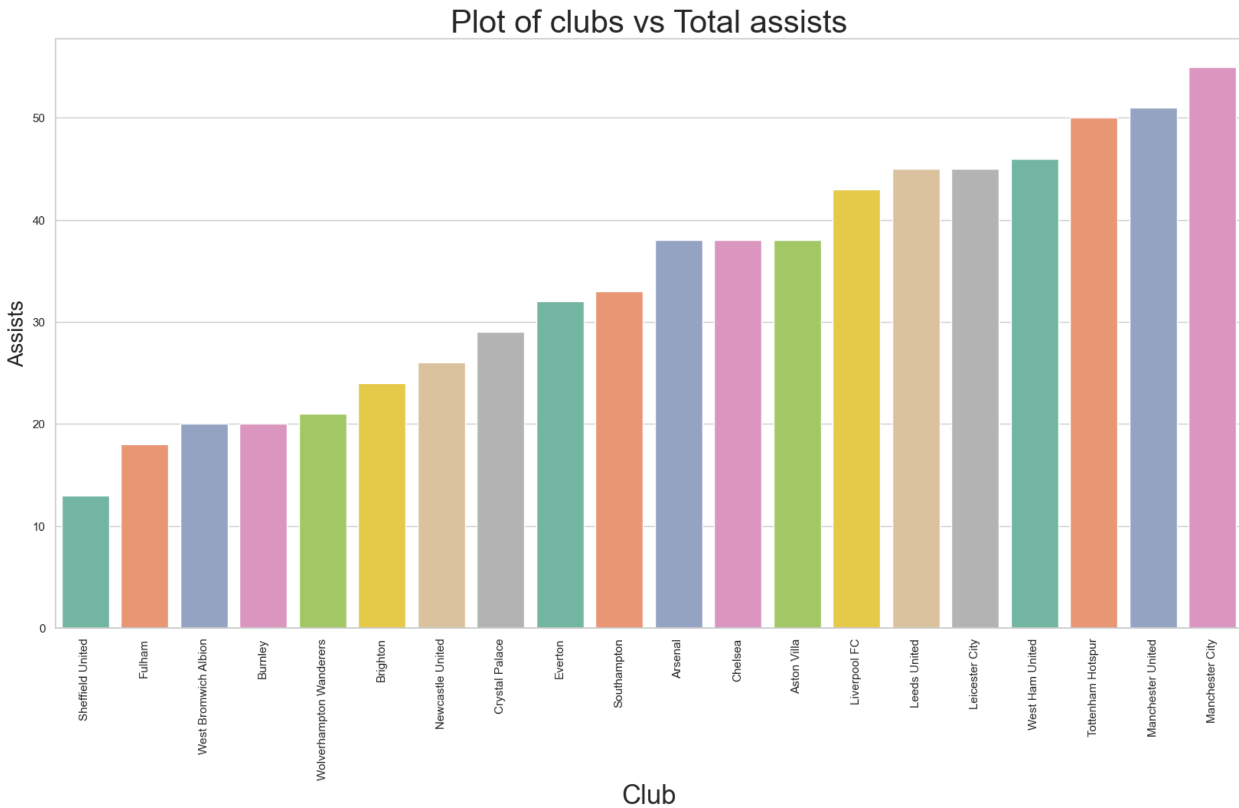
```
 Text(18, 0, 'West Bromwich Albion'),
 Text(19, 0, 'Sheffield United')])
```




```
#total assesment from each club

assists_by_club=pd.DataFrame(data.groupby('Club',as_index=False)
['Assists'].sum())
sns.set_theme(style='whitegrid',color_codes=True)
ax=sns.barplot(x='Club',y='Assists',data=assists_by_club.sort_values(b
y='Assists'),palette='Set2')
ax.set_xlabel('Club',fontsize=25)
ax.set_ylabel('Assists',fontsize=20)
plt.xticks(rotation=90)
```

```
plt.rcParams['figure.figsize']=(20,10)
plt.title('Plot of clubs vs Total assists', fontsize=30)
```

```
Text(0.5, 1.0, 'Plot of clubs vs Total assists')
```



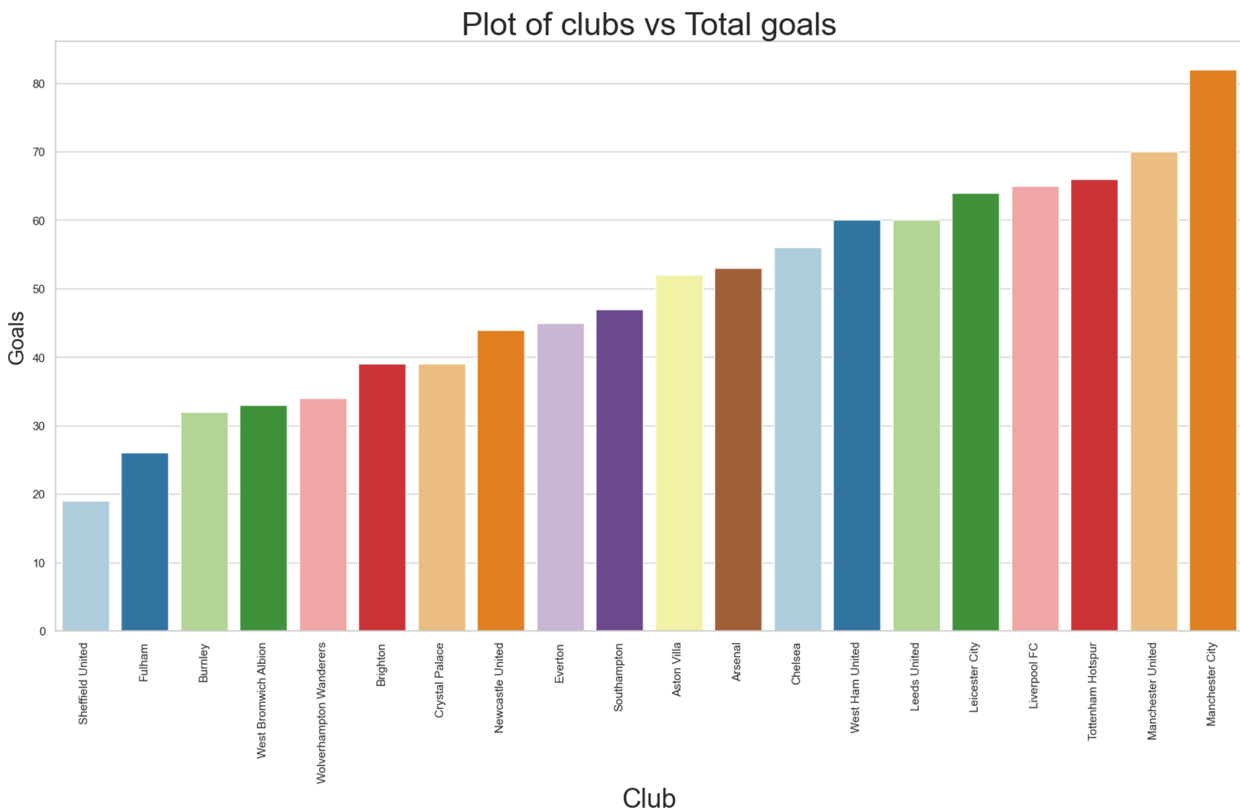

```
#Top 10 assists
assists=data[['Name','Club','Assists','Matches']].nlargest(n=10,column
s='Assists')
assists
```

ColName	Name	Club	Assists	Matches
162	Harry Kane	Tottenham Hotspur	14	35
34	Kevin De Bruyne	Manchester City	12	25
51	Bruno Fernandes	Manchester United	12	37
161	Son Heung-min	Tottenham Hotspur	10	37
273	Jack Grealish	Aston Villa	10	26
54	Marcus Rashford	Manchester United	9	37
110	Jamie Vardy	Leicester City	9	34
220	Raphael Dias Belloli	Leeds United	9	30
2	Timo Werner	Chelsea	8	35
136	Aaron Cresswell	West Ham United	8	36

```
goals_by_club=pd.DataFrame(data.groupby('Club',as_index=False)
['Goals'].sum())
sns.set_theme(style='whitegrid',color_codes=True)
```

```python
ax=sns.barplot(x='Club',y='Goals',data=goals_by_club.sort_values(by='Goals'),palette='Paired')
ax.set_xlabel('Club',fontsize=25)
ax.set_ylabel('Goals',fontsize=20)
plt.xticks(rotation=90)
plt.rcParams['figure.figsize']=(20,10)
plt.title('Plot of clubs vs Total goals', fontsize=30)
```

```
Text(0.5, 1.0, 'Plot of clubs vs Total goals')
```




```python
#top 10 goals by player

goals=data[['Name','Club','Goals','Matches']].nlargest(n=10,columns='Goals')
goals
```

ColName	Name	Club	Goals	Matches
162	Harry Kane	Tottenham Hotspur	23	35
81	Mohamed Salah	Liverpool FC	22	37
51	Bruno Fernandes	Manchester United	18	37
161	Son Heung-min	Tottenham Hotspur	17	37
214	Patrick Bamford	Leeds United	17	38
237	Dominic Calvert-Lewin	Everton	16	33
110	Jamie Vardy	Leicester City	15	34
267	Ollie Watkins	Aston Villa	14	37

```
33              İlkay Gündoğan    Manchester City     13        28
191        Alexandre Lacazette            Arsenal     13        31

#top 10 goals per match

goals_per_match=data[['Name','GoalsPerMatch','Matches','Goals']].nlargest(n=10,columns='GoalsPerMatch')
goals_per_match
```

ColName	Name	GoalsPerMatch	Matches	Goals
162	Harry Kane	0.657	35	23
81	Mohamed Salah	0.595	37	22
307	Joe Willock	0.571	14	8
145	Jesse Lingard	0.562	16	9
175	Gareth Bale	0.550	20	11
74	Anthony Elanga	0.500	2	1
51	Bruno Fernandes	0.486	37	18
237	Dominic Calvert-Lewin	0.485	33	16
120	Kelechi Iheanacho	0.480	25	12
92	Diogo Jota	0.474	19	9

Dependent and independent features

```
X = data[['Matches']]

y = data['Goals']
```

Splitting the dataset into training and testing

```
from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=1)

from sklearn.preprocessing import StandardScaler
ss=StandardScaler()
X_train=ss.fit_transform(X_train) # xtrain = training input samples
X_test=ss.transform(X_test) # xtest - testing input samples
```

Model Development and Model Training

```
from sklearn.ensemble import RandomForestClassifier

clfr=RandomForestClassifier(n_estimators=10,criterion='entropy',random
_state=0)

clfr.fit(X_train,y_train)
```

```
RandomForestClassifier(criterion='entropy', n_estimators=10,
random_state=0)
```

```
from sklearn.ensemble import RandomForestClassifier

clfr1=RandomForestClassifier(n_estimators=10,criterion='gini',random_s
tate=0)

clfr1.fit(X_train,y_train)
```

```
RandomForestClassifier(n_estimators=10, random_state=0)
```

Model Prediction

```
from sklearn.metrics import
confusion_matrix,classification_report,accuracy_score

ypre=clfr.predict(X_test)# entropy ypre calculation

ypre1=clfr1.predict(X_test)# gini ypre calculation

#entropy ypre calculation

data = pd.DataFrame({'Actual': y_test, 'Predicted': ypre})
data.head()
```

	Actual	Predicted
110	15	1
244	0	1
430	0	2
438	0	0
233	0	0

```
# gini ypre calculation

data = pd.DataFrame({'Actual': y_test, 'Predicted': ypre1})
data.head()
```

	Actual	Predicted
110	15	1

```
244        0          1
430        0          2
438        0          0
233        0          0
```

```
print('entropy Accuracy Score:')
accuracy_score(y_test,ypre)*100
```

```
entropy Accuracy Score:
```

```
44.85981308411215
```

```
print('gini Accuracy Score:')
accuracy_score(y_test,ypre1)*100
```

```
gini Accuracy Score:
```

```
44.85981308411215
```

```
print('entropy - confusion matrix\n--------------------\n')
print(confusion_matrix(y_test,ypre))
print('gini - confusion matrix\n--------------------\n')
print(confusion_matrix(y_test,ypre1))
```

```
entropy - confusion matrix
--------------------

[[41  7  1  0  0  0  0  0  0  0  0  0  0  0  0]
 [17  7  2  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 7  2  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 4  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 3  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]]
gini - confusion matrix
--------------------

[[41  7  1  0  0  0  0  0  0  0  0  0  0  0  0]
 [17  7  2  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 7  2  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 4  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
```

```
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 3  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]]
```

```
print('entropy result\n--------------')
print(classification_report(y_test,ypre))
print('gini index result\n---------------------')
print(classification_report(y_test,ypre1))
```

```
entropy result
--------------
              precision    recall  f1-score   support

           0       0.52      0.84      0.64        49
           1       0.29      0.27      0.28        26
           2       0.00      0.00      0.00         9
           3       0.00      0.00      0.00         5
           4       0.00      0.00      0.00         2
           5       0.00      0.00      0.00         3
           6       0.00      0.00      0.00         2
           7       0.00      0.00      0.00         1
           8       0.00      0.00      0.00         2
           9       0.00      0.00      0.00         3
          10       0.00      0.00      0.00         1
          14       0.00      0.00      0.00         1
          15       0.00      0.00      0.00         1
          16       0.00      0.00      0.00         1
          17       0.00      0.00      0.00         1

    accuracy                           0.45       107
   macro avg       0.05      0.07      0.06       107
weighted avg       0.31      0.45      0.36       107

gini index result
---------------------
              precision    recall  f1-score   support

           0       0.52      0.84      0.64        49
           1       0.29      0.27      0.28        26
           2       0.00      0.00      0.00         9
           3       0.00      0.00      0.00         5
           4       0.00      0.00      0.00         2
           5       0.00      0.00      0.00         3
           6       0.00      0.00      0.00         2
```

```
           7       0.00      0.00      0.00         1
           8       0.00      0.00      0.00         2
           9       0.00      0.00      0.00         3
          10       0.00      0.00      0.00         1
          14       0.00      0.00      0.00         1
          15       0.00      0.00      0.00         1
          16       0.00      0.00      0.00         1
          17       0.00      0.00      0.00         1

    accuracy                           0.45       107
   macro avg       0.05      0.07      0.06       107
weighted avg       0.31      0.45      0.36       107
```

```
plt.scatter(X_train,y_train)
plt.plot(X_train,clfr.predict(X_train))
```

```
[<matplotlib.lines.Line2D at 0x2493b22ec70>]
```