shai

March 23, 2023

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[98]:
```
data=pd.read_csv(r"/content/goals.csv")
data
```

[98]:

	serial	player_name	club	position	goals	right_foot	left_foot	headers	others	inside_area	outside_areas	penalties	match_played
0	1	Benzema	Real Madrid	Forward	15	11	1	3	0	13	2	3	12
1	2	Lewandowski	Bayern	Forward	13	8	3	1	1	13	0	3	10
2	3	Haller	Ajax	Forward	11	3	4	3	1	11	0	1	8
3	4	Salah	Liverpool	Forward	8	0	8	0	0	7	1	1	13
4	5	Nkunku	Leipzig	Midfielder	7	3	1	3	0	7	0	0	12
..	…	…	…	…	…	…	…	…	…	…	…	…	…
178	73	Yansané	Sheriff	Forward	1	0	0	1	0	1	0	0	1
179	73	Yakhshiboev	Sheriff	Forward	1	0	0	1	0	1	0	0	1
180	73	Messias Junior	Milan	Forward	1	0	0	1	0	1	0	0	1
181	73	Morato	Benfica	Defender	1	0	0	1	0	1	0	0	1
182	73	Pedro Porro	Sporting CP	Defender	1	0	0	0	1	1	0	0	1

```
4               6
..            …
178             3
179             3
180             2
181             2
182             7

[183 rows x 13 columns]
```

[100]:
```
data.head()
```

[100]:
```
   serial  player_name         club    position  goals  right_foot  left_foot  \
0       1      Benzema  Real Madrid     Forward     15          11          1
1       2  Lewandowski       Bayern     Forward     13           8          3
2       3       Haller         Ajax     Forward     11           3          4
3       4        Salah    Liverpool     Forward      8           0          8
4       5       Nkunku      Leipzig  Midfielder      7           3          1

   headers  others  inside_area  outside_areas  penalties  match_played
0        3       0           13              2          3            12
1        1       1           13              0          3            10
2        3       1           11              0          1             8
3        0       0            7              1          1            13
4        3       0            7              0          0             6
```

[101]:
```
data.tail()
```

[101]:
```
     serial     player_name         club  position  goals  right_foot  \
178      73         Yansané      Sheriff   Forward      1           0
179      73     Yakhshiboev      Sheriff   Forward      1           0
180      73  Messias Junior        Milan   Forward      1           0
181      73          Morato      Benfica  Defender      1           0
182      73     Pedro Porro  Sporting CP  Defender      1           0

     left_foot  headers  others  inside_area  outside_areas  penalties  \
178          0        1       0            1              0          0
179          0        1       0            1              0          0
180          0        1       0            1              0          0
181          0        1       0            1              0          0
182          0        0       1            1              0          0

     match_played
178             3
179             3
180             2
181             2
```

```
182             7
```

[3]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 183 entries, 0 to 182
Data columns (total 13 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   serial         183 non-null    int64 
 1   player_name    183 non-null    object
 2   club           183 non-null    object
 3   position       183 non-null    object
 4   goals          183 non-null    int64 
 5   right_foot     183 non-null    int64 
 6   left_foot      183 non-null    int64 
 7   headers        183 non-null    int64 
 8   others         183 non-null    int64 
 9   inside_area    183 non-null    int64 
 10  outside_areas  183 non-null    int64 
 11  penalties      183 non-null    int64 
 12  match_played   183 non-null    int64 
dtypes: int64(10), object(3)
memory usage: 18.7+ KB
```

[ ]:
```

```

[102]:
```
data.isnull().sum()
```

[102]:
```
serial          0
player_name     0
club            0
position        0
goals           0
right_foot      0
left_foot       0
headers         0
others          0
inside_area     0
outside_areas   0
penalties       0
match_played    0
dtype: int64
```

[103]:
```
#No null or empty cell
```

[9]:
```
data.shape
```

[9]:
```
(183, 13)
```

[104]:
```
#The file has 13 columns and 183 rows
```

[105]:
```
data.describe()
```

[105]:

	serial	goals	right_foot	left_foot	headers	others
count	183.000000	183.000000	183.000000	183.000000	183.000000	183.000000
mean	55.016393	2.021858	0.972678	0.693989	0.327869	0.021858
std	24.173929	2.005367	1.400281	1.140810	0.612722	0.146621
min	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	43.000000	1.000000	0.000000	0.000000	0.000000	0.000000
50%	73.000000	1.000000	1.000000	0.000000	0.000000	0.000000
75%	73.000000	2.000000	1.000000	1.000000	1.000000	0.000000
max	73.000000	15.000000	11.000000	8.000000	3.000000	1.000000

	inside_area	outside_areas	penalties	match_played
count	183.000000	183.000000	183.000000	183.000000
mean	1.814208	0.207650	0.196721	6.841530
std	1.923943	0.514137	0.549315	2.568196
min	0.000000	0.000000	0.000000	1.000000
25%	1.000000	0.000000	0.000000	5.000000
50%	1.000000	0.000000	0.000000	6.000000
75%	2.000000	0.000000	0.000000	8.000000
max	13.000000	3.000000	3.000000	13.000000

[14]:
```
data['player_name'].count()
```

[14]:
```
183
```

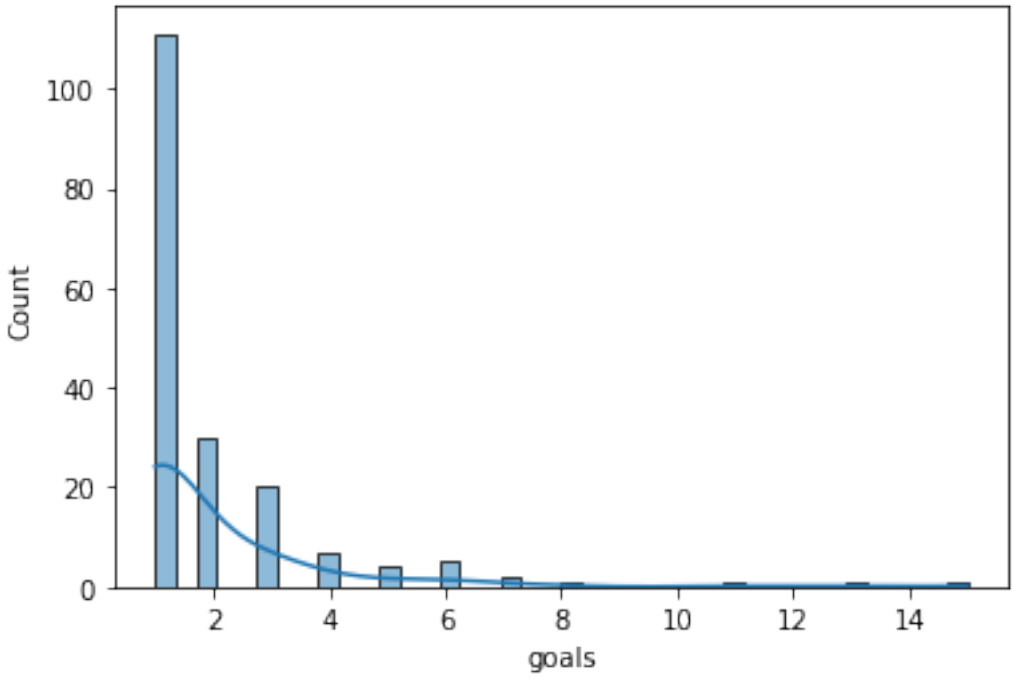
[106]:
```
#There are 183 different players with goals
```

[107]:
```
data["goals"].sum()
```

[107]:
```
370
```

[109]:
```
#Total goals is 370
```

[110]:
```
sns.histplot(x="goals",data=data,kde=True)
plt.show()
```



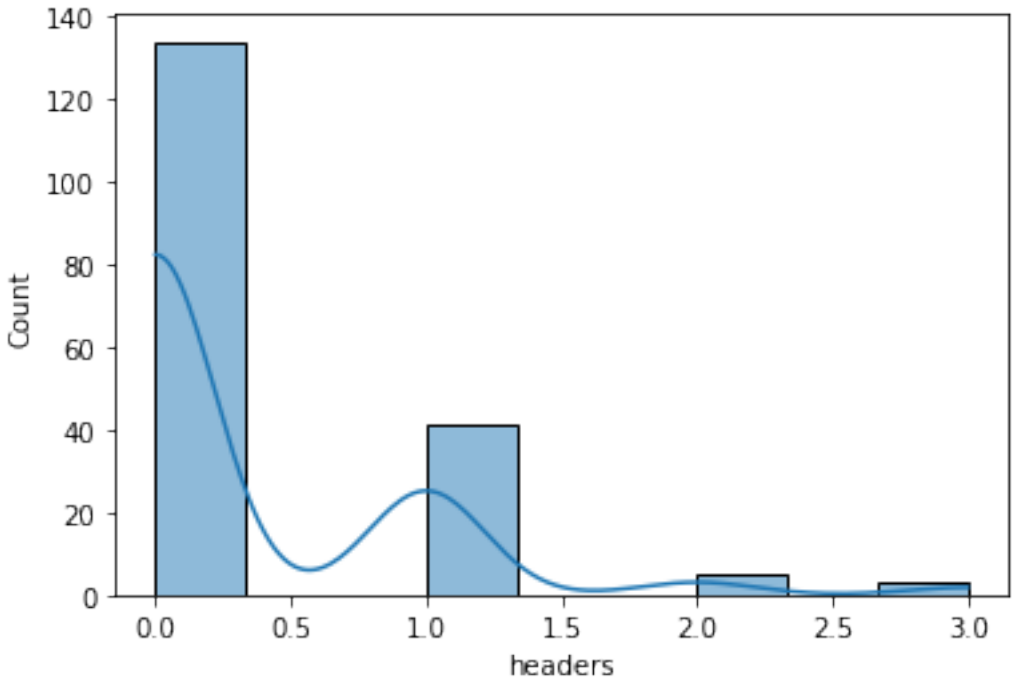
[111]:
```
#Most players have one goal
```

[112]:
```
data['headers'].sum()
```

[112]: 60

[113]:
```
#There are 60 goals scored with the head
```

[114]:
```
sns.histplot(x="headers",data=data,kde=True)
plt.show()
```



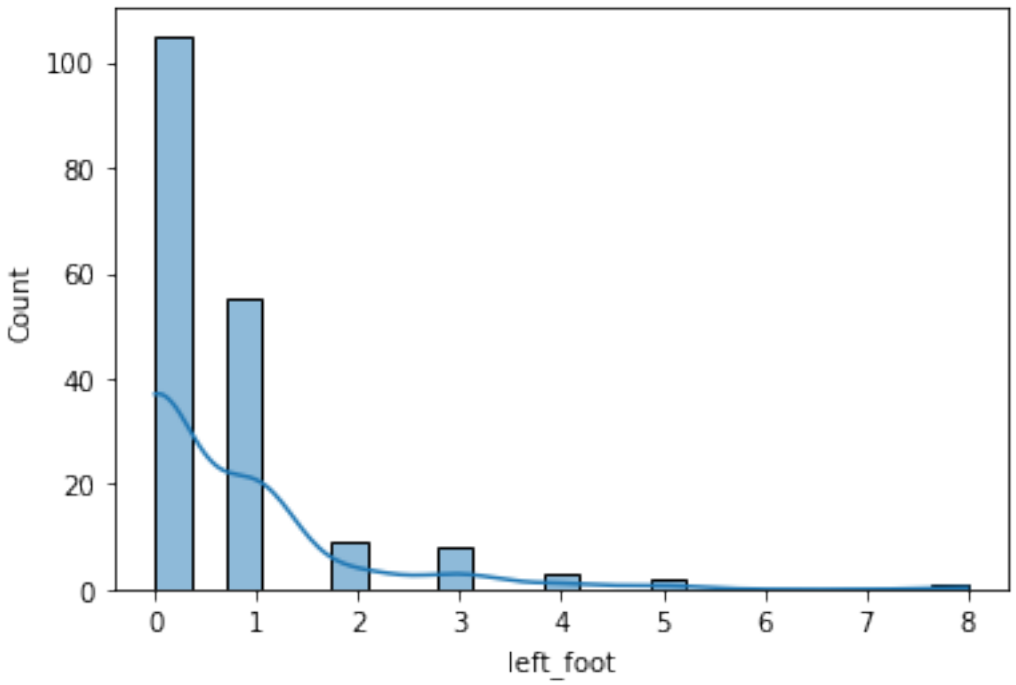
[115]:
```
#Most of the players who scored do not have goals with the head, and most of␣
↪those who scored with the head have a single goal
```

[116]:
```
data['left_foot'].sum()
```

[116]: 127

[117]:
```
#There are 126 goals with the left foot
```

[118]:
```
sns.histplot(x='left_foot',data=data,kde=True)
plt.show()
```



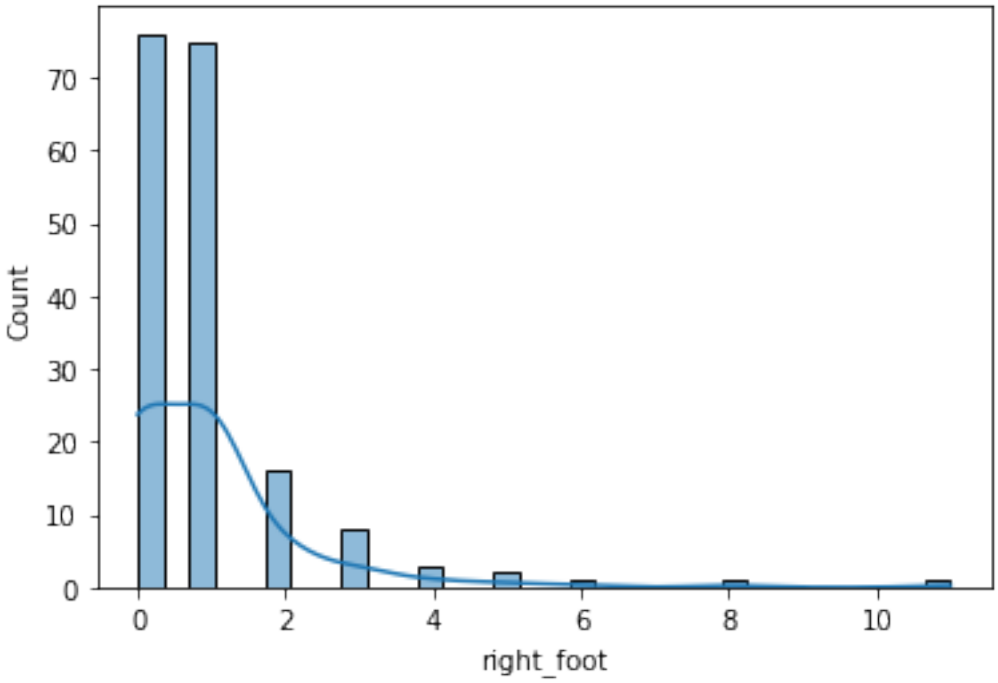
[120]:
```
#Most of the players who scored do not have goals with the left foot, and most␣
↪of those who scored with the left foot have a single goal
```

[121]:
```
data['right_foot'].sum()
```

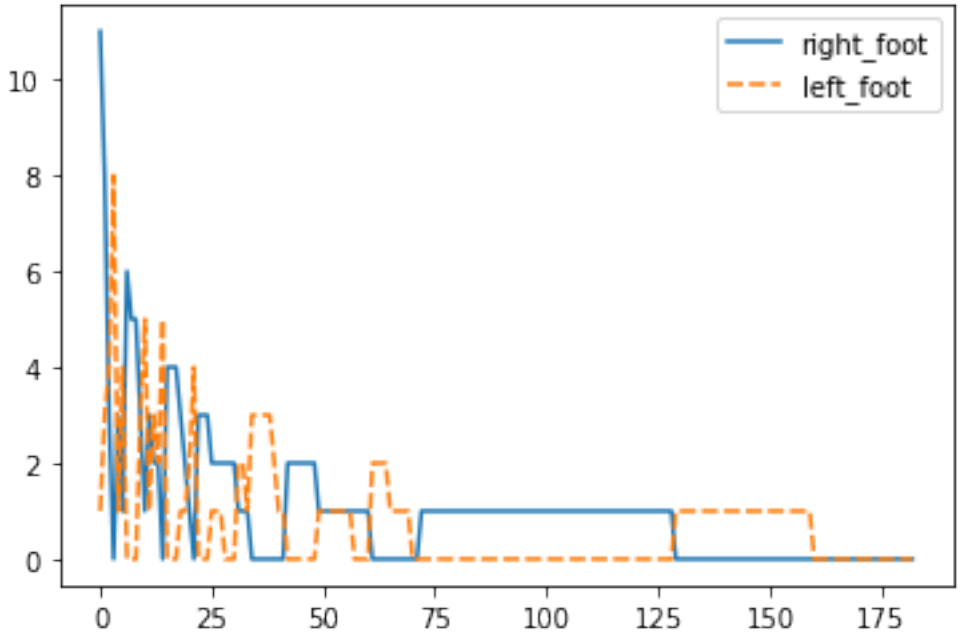
[121]: 178

[122]:
```
#There are 178 goals with the right foot
```

[123]:
```
sns.histplot(x='right_foot',data=data,kde=True)
plt.show()
```



[125]:
```
sns.lineplot(data.
  ↪drop(['goals','headers','serial','others','inside_area','outside_areas','penalties','match_
plt.show()
```



[126]:
```
#Comparison of the number of goals with the right and left foot
```

[127]:
```
data['others'].sum()
```

[127]: 4

[129]:
```
#There are 4 goals with other places of body
```

[130]:
```
data['penalties'].sum()
```

[130]: 36

[131]:
```
#There are 36 goalse git from penalties
```

[132]:
```
data['inside_area'].sum()
```

[132]: 332

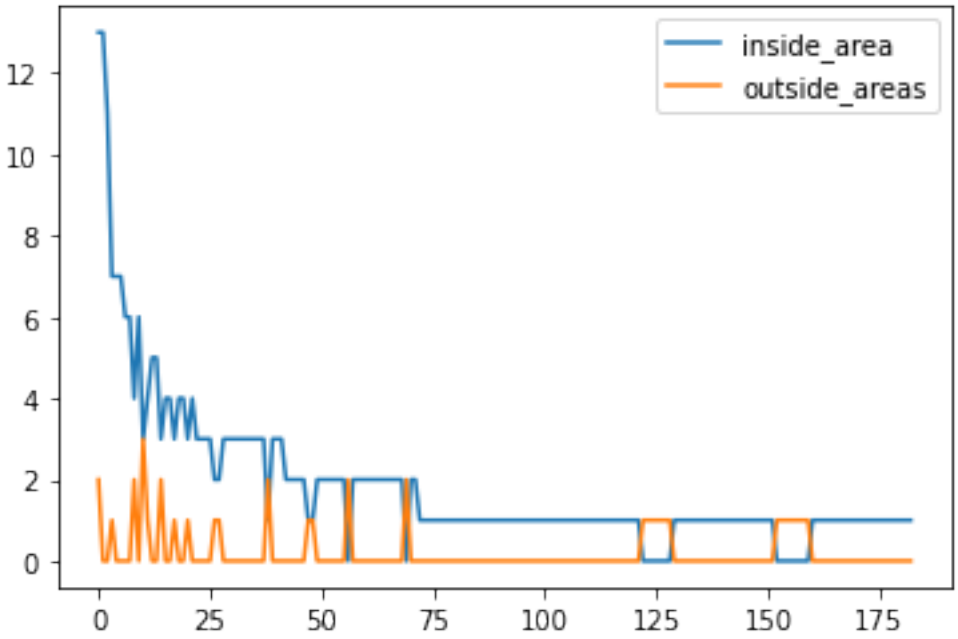
[134]:
```
#There are 332 goals inside the area
```

[135]:
```
data['outside_areas'].sum()
```

[135]: 38

[137]:
```
#There are 38 goals outside the area
```

[139]:
```
plt.plot(data['inside_area'],label="inside_area")
plt.plot(data['outside_areas'],label='outside_areas')
plt.legend()
plt.show()
```



[140]:
```
#Comparison of the number of goalsfrom inside and outside the area
```

[141]:
```
data['goals'].max()
```

[141]:
```
15
```

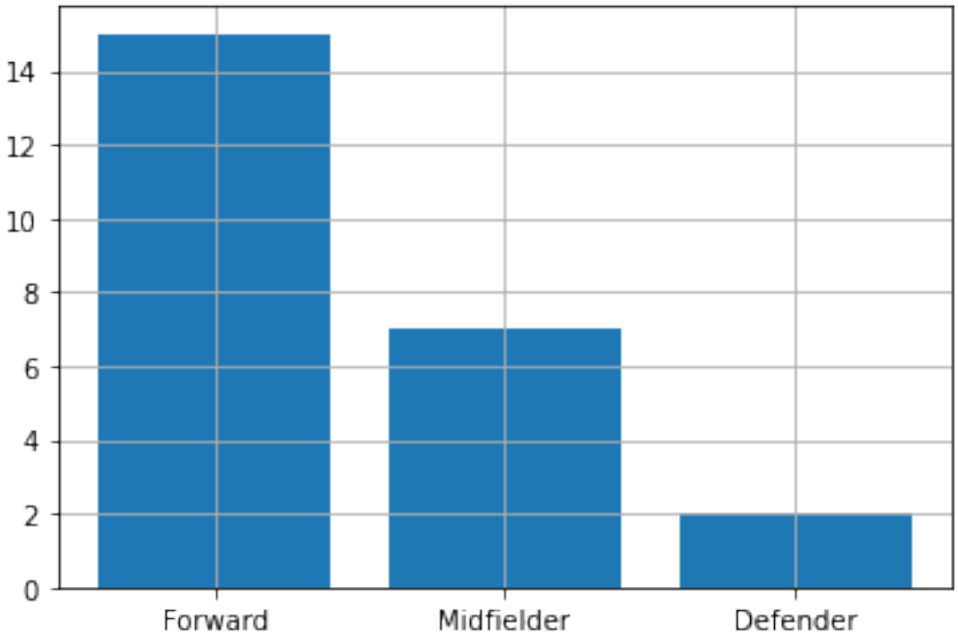
[144]:
```
#Best goaler has 15 goals
```

[145]:
```
data['position'].max()
```

[145]:
```
'Midfielder'
```

[148]:
```
#Most of the goals scored are midfielders
```

[150]:
```
plt.bar(data["position"],data['goals'])
plt.grid()
plt.show()
```



[151]:
```
#The most striker scored 15 goals, the most midfielder scored goals, and the␣
↪most defender scored two goals
```

[152]:
```
maxg=data[['club','goals']].groupby(['club']).sum()
maxg=maxg.sort_values(['goals'],ascending=False).head()
maxg
```

[152]:

club	goals
Bayern	30
Liverpool	30
Real Madrid	28
Man. City	28
Ajax	21

[155]:
```
#These are the 5 most scored teams in the tournament
```

[154]:
```
maxp=data[['goals','player_name']].groupby(['player_name']).sum()
maxp=maxp.sort_values('goals',ascending=False).head()
maxp
```

[154]:

player_name	goals
Benzema	15
Lewandowski	13
Haller	11

```
Salah            8
Nkunku           7
```

[156]:
```
#These are the top 5 scorers
```

[157]:
```
maxb=data[['player_name','match_played']].groupby(['player_name']).sum()
maxb=maxb.sort_values('match_played',ascending=False)
maxb.head()
```

[157]:
```
                 match_played
player_name
Luis Díaz                  13
Salah                      13
Fabinho                    13
Vinícius Júnior            13
Mané                       13
```

[158]:
```
#These are top 5 player played match
```

[159]:
```
maxpgr=data[['player_name','right_foot']].groupby(['player_name']).sum()
maxpgr=maxpgr.sort_values('right_foot',ascending=False)
maxpgr.head()
```

[159]:
```
             right_foot
player_name
Benzema              11
Lewandowski           8
Mbappé                6
Ronaldo               5
Núñez                 5
```

[160]:
```
#The 5 most scored players with the right foot
```

[161]:
```
maxpgl=data[['player_name','left_foot']].groupby(['player_name']).sum()
maxpgl=maxpgl.sort_values('left_foot',ascending=False)
maxpgl.head()
```

[161]:
```
             left_foot
player_name
Salah                8
Messi                5
Sané                 5
Griezmann            4
Haller               4
```

[162]:
```
#The 5 most scored players with the left foot
```

[163]:
```
maxpgh=data[['player_name','headers']].groupby(['player_name']).sum()
maxpgh=maxpgh.sort_values('headers',ascending=False)
maxpgh.head()
```

[163]:

player_name	headers
Haller	3
Nkunku	3
Benzema	3
Havertz	2
Haaland	2

[164]:
```
#The 5 most scored players with the head
```

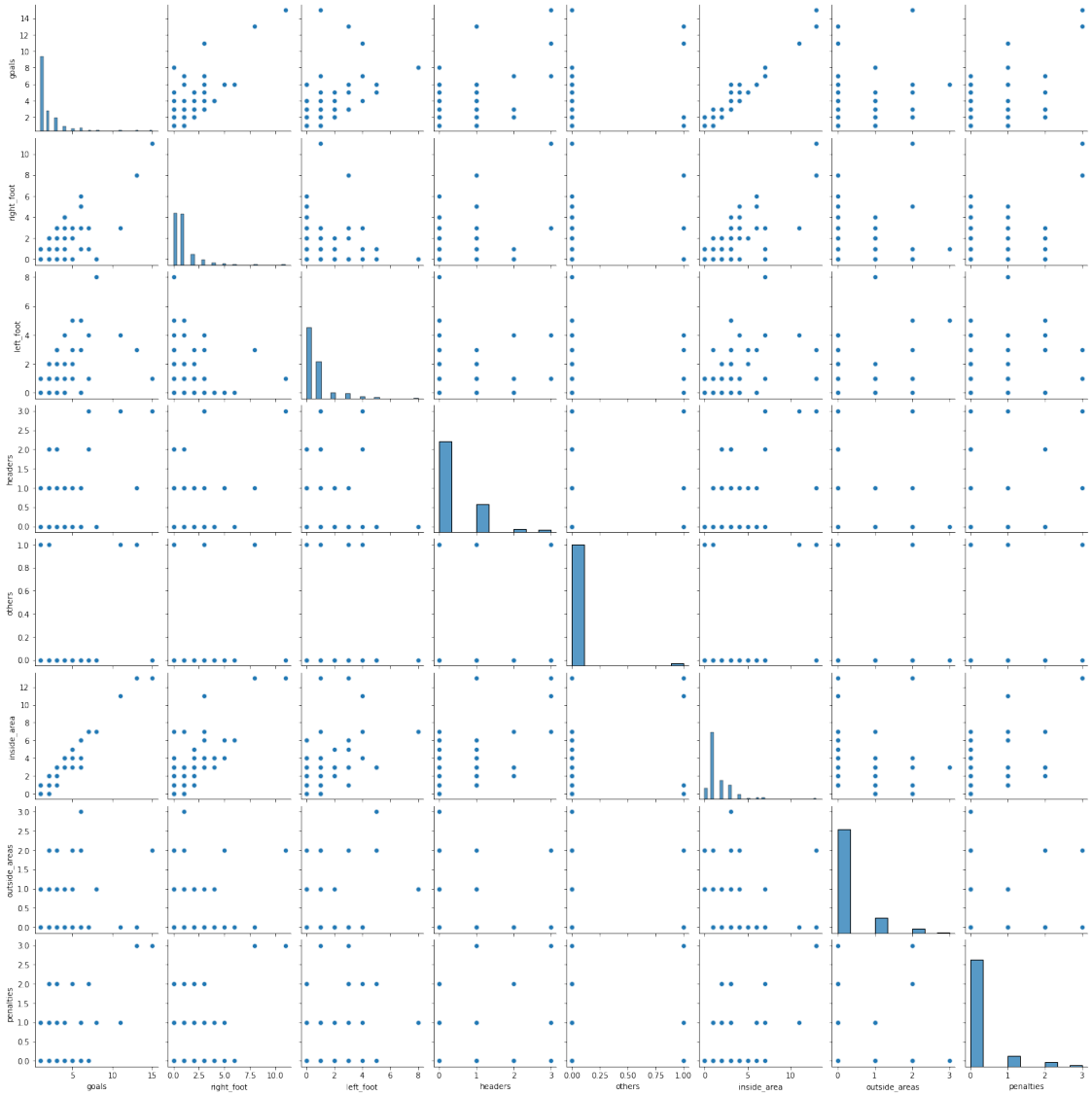
[165]:
```
maxpgp=data[['player_name','penalties']].groupby(['player_name']).sum()
maxpgp=maxpgp.sort_values('penalties',ascending=False)
maxpgp.head()
```

[165]:

player_name	penalties
Lewandowski	3
Benzema	3
Adeyemi	2
Jorginho	2
Rakitić	2

[166]:
```
#The 5 most scored players from penalties
```

[167]:
```
sns.pairplot(data.
  ↪drop(['serial','match_played','player_name','club','position'],axis=1))
plt.show()
```


goals
right_foot
left_foot
headers
others
inside_area
outside_areas
penalties