In [1]:

```
import numpy as np
```

In [ ]:

```
# Names: Aley Elfouly, Hussein Moataz; #ID: 900211655, 900211243
```

In [ ]:

```
# Abstract: We will be showing statistcal data about all our continues variables and
```

In [ ]:

```
e what we are wroking on the outcomes should be great, the data set is about the eng
```

In [3]:

```
import pandas as pd
```

In [4]:

```
import os
```

In [5]:

```
os.getcwd()
```

Out[5]:

```
'C:\\Users\\Moataz Wahba\\Desktop\\python'
```

In [6]:

```
os.chdir('C:\\Users\\Moataz Wahba\\Desktop\\python')
```

In [9]:

```
df=pd.read_csv("EPL_20_21.csv")
df.head()
```

Out[9]:

	Name	Club	Nationality	Position	Age	Matches	Starts	Mins	Goals	Assists	Passes_
0	Mason Mount	Chelsea	ENG	MF,FW	21	36	32	2890	6	5	
1	Edouard Mendy	Chelsea	SEN	GK	28	31	31	2745	0	0	
2	Timo Werner	Chelsea	GER	FW	24	35	29	2602	6	8	
3	Ben Chilwell	Chelsea	ENG	DF	23	27	27	2286	3	5	
4	Reece James	Chelsea	ENG	DF	20	32	25	2373	1	2	

In [18]:

```
df['Age'].mean()
```

Out[18]:

```
25.5
```

In [19]:

```
df['Mins'].mean()
```

Out[19]:

```
1411.4436090225563
```

In [21]:

```
df['Goals'].mean()
```

Out[21]:

```
1.8533834586466165
```

In [ ]:

```

```

In [22]:

```
df['Assists'].mean()
```

Out[22]:

```
1.287593984962406
```

In [23]:

```
df["Passes_Attempted"].mean()
```

Out[23]:

```
717.75
```

In [25]:

```
df["Yellow_Cards"].mean()
```

Out[25]:

```
2.1146616541353382
```

In [26]:

```
df["Red_Cards"].mean()
```

Out[26]:

```
0.09022556390977443
```

In [27]:

```
df['Age'].median()
```

Out[27]:

```
26.0
```

In [28]:

```
df['Goals'].median()
```

Out[28]:

```
1.0
```

In [29]:

```
df['Assists'].median()
```

Out[29]:

```
0.0
```

In [31]:

```
df["Passes_Attempted"].median()
```

Out[31]:

```
573.5
```

In [32]:

```
df["Yellow_Cards"].median()
```

Out[32]:

```
2.0
```

In [33]:

```
df["Red_Cards"].median()
```

Out[33]:

```
0.0
```

In [44]:

```
df["Age"].quantile(0.25)
```

Out[44]:

```
22.0
```

In [43]:

```
df["Age"].quantile(0.5)
```

Out[43]:

```
26.0
```

In [42]:

```
df["Age"].quantile(0.75)
```

Out[42]:

```
29.0
```

In [41]:

```
df["Age"].quantile(1)
```

Out[41]:

```
38.0
```

In [47]:

```
df["Goals"].quantile(0.25)
```

Out[47]:

```
0.0
```

In [48]:

```
df["Goals"].quantile(0.5)
```

Out[48]:

```
1.0
```

In [49]:

```
df["Goals"].quantile(0.75)
```

Out[49]:

```
2.0
```

In [50]:

```
df["Goals"].quantile(1)
```

Out[50]:

```
23.0
```

In [51]:

```
df["Assists"].quantile(0.25)
```

Out[51]:

```
0.0
```

In [52]:

```
df["Assists"].quantile(0.50)
```

Out[52]:

```
0.0
```

In [53]:

```
df["Assists"].quantile(0.75)
```

Out[53]:

```
2.0
```

In [54]:

```
df["Assists"].quantile(1)
```

Out[54]:

```
14.0
```

In [55]:

```
df["Passes_Attempted"].quantile(0.25)
```

Out[55]:

```
171.5
```

In [56]:

```
df["Passes_Attempted"].quantile(0.50)
```

Out[56]:

```
573.5
```

In [57]:

```
df["Passes_Attempted"].quantile(0.75)
```

Out[57]:

```
1129.5
```

In [58]:

```
df["Passes_Attempted"].quantile(1)
```

Out[58]:

```
3214.0
```

In [59]:

```
df["Yellow_Cards"].quantile(0.25)
```

Out[59]:

```
0.0
```

In [60]:

```
df["Yellow_Cards"].quantile(0.5)
```

Out[60]:

```
2.0
```

In [61]:

```
df["Yellow_Cards"].quantile(0.75)
```

Out[61]:

```
3.0
```

In [62]:

```
df["Yellow_Cards"].quantile(1)
```

Out[62]:

```
12.0
```

In [63]:

```
df["Red_Cards"].quantile(0.25)
```

Out[63]:

```
0.0
```

In [67]:

```
df["Red_Cards"].quantile(0.5)
```

Out[67]:

```
0.0
```

In [68]:

```
df["Red_Cards"].quantile(0.75)
```

Out[68]:

```
0.0
```

In [69]:

```
df["Red_Cards"].quantile(1)
```

Out[69]:

```
2.0
```

In [71]:

```
df["Age"].var()
```

Out[71]:

```
18.657250470809792
```

In [79]:

```
df["Goals"].var()
```

Out[79]:

```
11.142304914829387
```

In [73]:

```
df["Assists"].var()
```

Out[73]:

```
4.389826968551317
```

In [74]:

```
df["Passes_Attempted"].var()
```

Out[74]:

```
398631.26129943505
```

In [75]:

```
df['Yellow_Cards'].var()
```

Out[75]:

```
5.148786514308378
```

In [77]:

```
df['Red_Cards'].var()
```

Out[77]:

```
0.08600597539045352
```

In [80]:

```
df["Age"].std()
```

Out[80]:

```
4.319403948556999
```

In [83]:

```
df["Goals"].std()
```

Out[83]:

```
3.338009124437707
```

In [86]:

```
df["Assists"].std()
```

Out[86]:

```
2.095191391866461
```

In [87]:

```
df["Passes_Attempted"].std()
```

Out[87]:

```
631.3725218121509
```

In [88]:

```
df['Yellow_Cards'].std()
```

Out[88]:

```
2.2690937649882117
```

In [90]:

```
df['Red_Cards'].std()
```

Out[90]:

```
0.2932677537515052
```

In [91]:

```
df.head()
```

Out[91]:

	Name	Club	Nationality	Position	Age	Matches	Starts	Mins	Goals	Assists	Passes_
0	Mason Mount	Chelsea	ENG	MF,FW	21	36	32	2890	6	5	
1	Edouard Mendy	Chelsea	SEN	GK	28	31	31	2745	0	0	
2	Timo Werner	Chelsea	GER	FW	24	35	29	2602	6	8	
3	Ben Chilwell	Chelsea	ENG	DF	23	27	27	2286	3	5	
4	Reece James	Chelsea	ENG	DF	20	32	25	2373	1	2	

In [93]:

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

Out[93]:

```
West Bromwich Albion        30
Manchester United           29
Arsenal                     29
Southampton                 29
Everton                     29
Liverpool FC                28
Fulham                      28
Chelsea                     27
Newcastle United            27
Brighton                    27
Wolverhampton Wanderers     27
Sheffield United            27
Leicester City              27
Burnley                     25
Manchester City             24
Crystal Palace              24
Tottenham Hotspur           24
West Ham United             24
Aston Villa                 24
Leeds United                23
Name: Club, dtype: int64
```

In [94]:

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

Out[94]:

```
ENG    192
FRA     31
BRA     27
ESP     26
IRL     21
POR     21
SCO     20
NED     16
WAL     12
BEL     11
GER      9
ARG      8
CIV      8
NGA      7
USA      6
DEN      6
SUI      6
SEN      5
EGY      5
ITA      5
POL      5
SWE      5
GHA      5
COL      5
NIR      5
TUR      5
AUS      4
SRB      4
NOR      3
ISL      3
ALG      3
JAM      3
CZE      3
GAB      2
SVK      2
MLI      2
COD      2
PAR      2
JPN      2
RSA      2
CMR      2
CRO      2
UKR      2
MAR      2
ZIM      1
MTN      1
NZL      1
IRN      1
CAN      1
MEX      1
BFA      1
AUT      1
MKD      1
BIH      1
URU      1
```

```
GUI       1
GRE       1
KOR       1
SKN       1
Name: Nationality, dtype: int64
```

In [12]:

```
df.head()
```

Out[12]:

	Name	Club	Nationality	Position	Age	Matches	Starts	Mins	Goals	Assists	Passes_
0	Mason Mount	Chelsea	ENG	MF,FW	21	36	32	2890	6	5	
1	Edouard Mendy	Chelsea	SEN	GK	28	31	31	2745	0	0	
2	Timo Werner	Chelsea	GER	FW	24	35	29	2602	6	8	
3	Ben Chilwell	Chelsea	ENG	DF	23	27	27	2286	3	5	
4	Reece James	Chelsea	ENG	DF	20	32	25	2373	1	2	

In [1]:

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

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
 last)
C:\Users\MOATAZ~1\AppData\Local\Temp/ipykernel_21868/2338253866.py in
<module>
----> 1 df['Position'].value_counts()

NameError: name 'df' is not defined
```

In [95]:

```
df = df.sort_values(["Goals", "Name"], ascending = (False,True ))
print(df)
```

```
49          Zack Steffen    Manchester City          USA        GK    25
1
137  Łukasz Fabiański    West Ham United          POL        GK    35
35

     Starts  Mins  Goals  Assists  Passes_Attempted  Perc_Passes_Compl
eted  \
162      35  3082     23       14               937
70.1
81       34  3078     22        5              1288
83.2
51       35  3099     18       12              2283
74.6
214      37  3050     17        7               506
76.3
161      36  3114     17       10              1199
76.7
..      ...   ...    ...      ...               ...
...
22         1    90      0        0                26
```

In [16]:

```
df
```

Out[16]:

	Name	Club	Nationality	Position	Age	Matches	Starts	Mins	Goals	Assists	Passes_Attempted	
0	Mason Mount	Chelsea	ENG	MF,FW	21	36	32	2890	6	5	1881	
1	Edouard Mendy	Chelsea	SEN	GK	28	31	31	2745	0	0	1007	
2	Timo Werner	Chelsea	GER	FW	24	35	29	2602	6	8	826	
3	Ben Chilwell	Chelsea	ENG	DF	23	27	27	2286	3	5	1806	
4	Reece James	Chelsea	ENG	DF	20	32	25	2373	1	2	1987	
...	...	...	...	...	...	...	...	...	...	...	...	
527	Lys Mousset	Sheffield United	FRA	FW,MF	24	11	2	296	0	0	50	

In [22]:

```
Goals_Scored=df['Goals'].sum()
```

In [24]:

```
Penelty_Goals_Scored=df['Penalty_Goals'].sum()
```

In [20]:

```
df['Penalty_Attempted'].sum()
```

Out[20]:

```
125
```

In [29]:

```
Penelty_Goals_Scored/Goals_Scored*100
```

Out[29]:

```
10.344827586206897
```

In [32]:

```
Penelty_Goals_Scored=df['Penalty_Goals'].sum()
Goals_Without_pens=Goals_Scored-Penelty_Goals_Scored
print(Penelty_Goals_Scored,Goals_Without_pens)
```

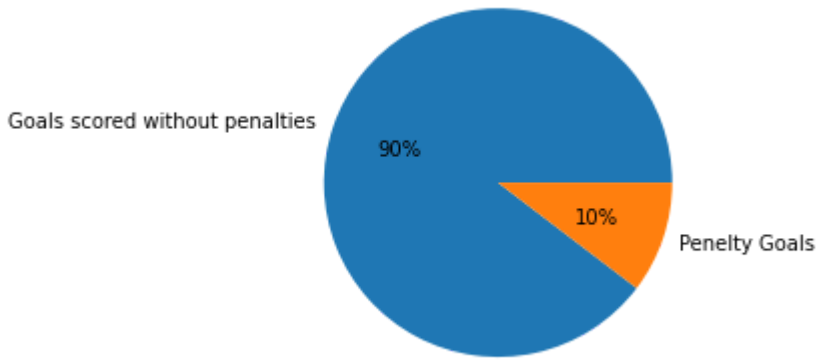
```
102 884
```

In [56]:

```
import matplotlib.pyplot as plt
data=[Goals_Without_pens,Penelty_Goals_Scored]
labels=['Goals scored without penalties','Penelty Goals']
plt.pie(data,labels=labels,autopct='%.0f%%')
plt.show
```

Out[56]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [ ]:

```

```

In [53]:

```
Pens_Scored=df['Penalty_Goals'].sum()
Pens_Attempted=df['Penalty_Attempted'].sum()
Pens_Missed=Pens_Attempted - Pens_Scored
print(Pens_Scored,Pens_Missed)
```

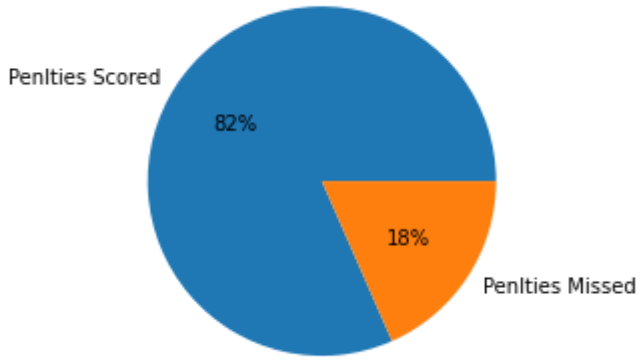
```
102 23
```

In [55]:

```
data=[Pens_Scored,Pens_Missed]
labels=['Penlties Scored', 'Penlties Missed']
plt.pie(data,labels=labels,autopct='%.0f%%')
plt.show
```

Out[55]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



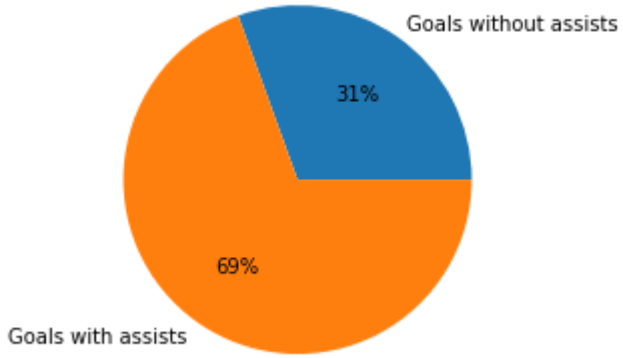
In [67]:

```
max(df['Goals','Name'])
```

```
---------------------------------------------------------------------
-----
KeyError                                  Traceback (most recent call
 last)
~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(s
elf, key, method, tolerance)
   3360             try:
-> 3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:

~\anaconda3\lib\site-packages\pandas\_libs\index.pyx in pandas._libs.i
ndex.IndexEngine.get_loc()

~\anaconda3\lib\site-packages\pandas\_libs\index.pyx in pandas._libs.i
ndex.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyOb
jectHashTable.get_item()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyOb
jectHashTable.get_item()

KeyError: ('Goals', 'Name')

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call
 last)
C:\Users\MOATAZ~1\AppData\Local\Temp/ipykernel_21868/3119828578.py in
<module>
----> 1 max(df['Goals','Name'])

~\anaconda3\lib\site-packages\pandas\core\frame.py in __getitem__(sel
f, key)
   3456             if self.columns.nlevels > 1:
   3457                 return self._getitem_multilevel(key)
-> 3458             indexer = self.columns.get_loc(key)
   3459             if is_integer(indexer):
   3460                 indexer = [indexer]

~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(s
elf, key, method, tolerance)
   3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:
-> 3363                 raise KeyError(key) from err
   3364
   3365         if is_scalar(key) and isna(key) and not self.hasnans:

KeyError: ('Goals', 'Name')
```

In [72]:

```
Assists=df['Assists'].sum()
Goals= df['Goals'].sum()
data= [Goals-Assists,Assists]
labels=['Goals without assists','Goals with assists']
plt.pie(data, labels=labels,autopct='%.0f%%')
plt.show()
```



In [83]:

```
English_Player= (df['Nationality']=='ENG').sum()
NON_English_Player= (df['Nationality']!='ENG').sum()
print(NON_English_Player)
```

```
340
```

In [85]:

```
data=[English_Player,NON_English_Player]
labels=
```

```
  File "C:\Users\MOATAZ~1\AppData\Local\Temp/ipykernel_21868/62354234
4.py", line 2
    labels=
           ^
SyntaxError: invalid syntax
```

In [92]:

```
import numpy as s
```

In [94]:

```
s.random.random([2,2])
```

Out[94]:

```
array([[0.82154444, 0.44502466],
       [0.36817253, 0.58940727]])
```

In [96]:

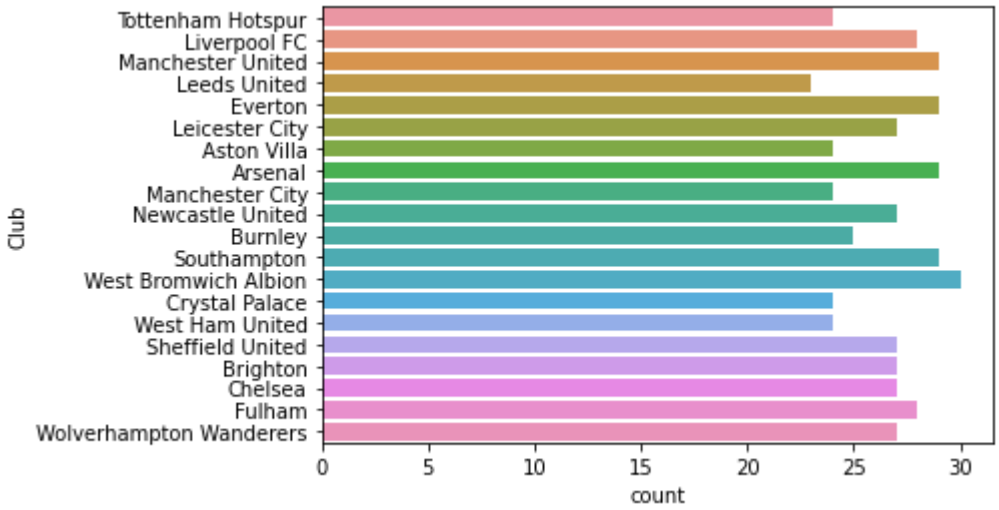
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [97]:

```
sns.countplot(y=df['Club'],)
```

Out[97]:

```
<AxesSubplot:xlabel='count', ylabel='Club'>
```

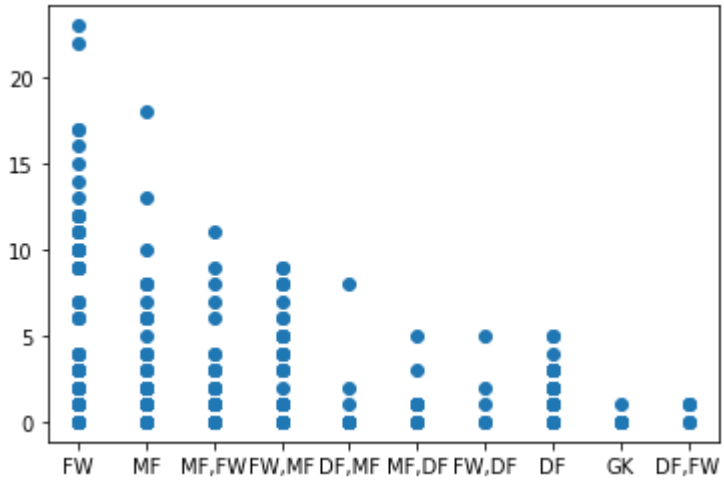


In [105]:

```
plt.scatter(df['Position'],df['Goals'])
# Here we can see that as expected the forwards have most goals and defenders and ke
```

Out[105]:

```
<matplotlib.collections.PathCollection at 0x1a8d4c85640>
```

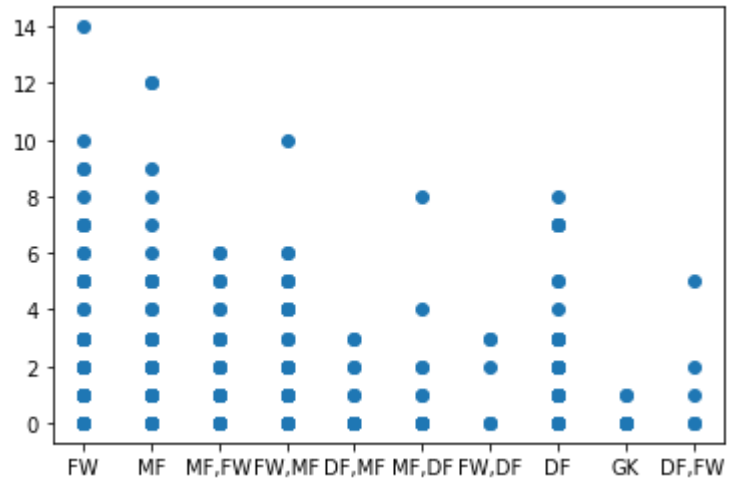


In [106]:

```
plt.scatter(df['Position'],df['Assists'])
#here we can see that its expected that the forwards and midfeilders have the higest
```

Out[106]:

```
<matplotlib.collections.PathCollection at 0x1a8d4cf4550>
```



In [111]:

```
C:\Users\Moataz Wahba\anaconda3\lib\site-packages\seaborn\_decorators.
py:36: FutureWarning: Pass the following variable as a keyword arg: vm
in. From version 0.12, the only valid positional argument will be `dat
a`, and passing other arguments without an explicit keyword will resul
t in an error or misinterpretation.
  warnings.warn(

---------------------------------------------------------------------
-----
IndexError                                Traceback (most recent call
 last)
C:\Users\MOATAZ~1\AppData\Local\Temp/ipykernel_21868/3363123468.py in
<module>
----> 1 sns.heatmap(df['Position'],df['Passes_Attempted'])

~\anaconda3\lib\site-packages\seaborn\_decorators.py in inner_f(*args,
**kwargs)
     44             )
     45         kwargs.update({k: arg for k, arg in zip(sig.parameters
, args)})
---> 46         return f(**kwargs)
     47     return inner_f
     48

~\anaconda3\lib\site-packages\seaborn\matrix.py in heatmap(data, vmin,
vmax, cmap, center, robust, annot, fmt, annot_kws, linewidths, linecol
or, cbar, cbar_kws, cbar_ax, square, xticklabels, yticklabels, mask, a
x, **kwargs)
    538     """
    539     # Initialize the plotter object
--> 540     plotter = _HeatMapper(data, vmin, vmax, cmap, center, robu
st, annot, fmt,
    541                           annot_kws, cbar, cbar_kws, xticklabe
ls,
    542                           yticklabels, mask)

~\anaconda3\lib\site-packages\seaborn\matrix.py in __init__(self, dat
a, vmin, vmax, cmap, center, robust, annot, fmt, annot_kws, cbar, cbar
_kws, xticklabels, yticklabels, mask)
    109         mask = _matrix_mask(data, mask)
    110
--> 111         plot_data = np.ma.masked_where(np.asarray(mask), plot_
data)
    112
    113         # Get good names for the rows and columns

~\anaconda3\lib\site-packages\numpy\ma\core.py in masked_where(conditi
on, a, copy)
   1927     (cshape, ashape) = (cond.shape, a.shape)
   1928     if cshape and cshape != ashape:
-> 1929         raise IndexError("Inconsistent shape between the condi
tion and the input"
   1930                          " (got %s and %s)" % (cshape, ashap
e))
   1931     if hasattr(a, '_mask'):
```

```
IndexError: Inconsistent shape between the condition and the input (go
t (532, 1) and (532,))
```

In [117]:

```
import scipy as stats
pd.corr(df['Goals'],df['Assists'])
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call
 last)
C:\Users\MOATAZ~1\AppData\Local\Temp/ipykernel_21868/1332537138.py in
<module>
      1 import scipy as stats
----> 2 pd.corr(df['Goals'],df['Assists'])

~\anaconda3\lib\site-packages\pandas\__init__.py in __getattr__(name)
    242         return _SparseArray
    243
--> 244     raise AttributeError(f"module 'pandas' has no attribute
 '{name}'")
    245
    246

AttributeError: module 'pandas' has no attribute 'corr'
```

In [119]:

```
df[['Goals','Assists']].corr()
# As we can see from the correlation is 0.6 which is almost a strong positive correl
```

Out[119]:

	Goals	Assists
Goals	1.000000	0.617831
Assists	0.617831	1.000000