

```
In [57]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
df = pd.read_csv("season-1819.csv")
df
```

Out[57]:

	Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	...	BbAv<2.5	BbAH	BbAHh	BbMxAHH	BbAvAHH	BbMxAHA	BbAvAHA	PSCH	PSCD	PSCA
0	SP1	17/08/2018	Betis	Levante	0	3	A	0	1	A	...	2.06	20	-0.75	1.89	1.85	2.07	2.00	1.59	4.42	5.89
1	SP1	17/08/2018	Girona	Valladolid	0	0	D	0	0	D	...	1.71	20	-0.75	2.06	2.01	1.90	1.85	1.76	3.57	5.62
2	SP1	18/08/2018	Barcelona	Alaves	3	0	H	0	0	D	...	3.18	19	-2.50	1.95	1.91	2.00	1.95	1.10	11.85	32.17
3	SP1	18/08/2018	Celta	Espanol	1	1	D	0	1	A	...	1.76	18	-0.75	2.26	2.18	1.74	1.71	2.18	3.26	3.85
4	SP1	18/08/2018	Villarreal	Sociedad	1	2	A	1	1	D	...	1.81	18	-0.25	1.76	1.74	2.23	2.14	2.32	3.21	3.53
...
375	SP1	18/05/2019	Levante	Ath Madrid	2	2	D	2	0	H	...	2.23	18	0.50	1.91	1.87	2.04	1.99	4.34	4.10	1.81
376	SP1	18/05/2019	Sevilla	Ath Bilbao	2	0	H	1	0	H	...	2.08	19	-1.00	2.70	2.60	1.60	1.53	2.17	3.08	4.15
377	SP1	18/05/2019	Valladolid	Valencia	0	2	A	0	1	A	...	2.25	20	1.50	1.78	1.74	2.20	2.13	8.01	5.13	1.40
378	SP1	19/05/2019	Eibar	Barcelona	2	2	D	2	2	D	...	2.65	19	0.50	2.03	1.98	1.92	1.88	4.96	4.55	1.65
379	SP1	19/05/2019	Real Madrid	Betis	0	2	A	0	0	D	...	3.42	23	-1.00	1.94	1.91	2.01	1.98	1.33	6.38	8.09

380 rows × 61 columns

```
In [58]: df.columns

Out[58]: Index(['Div', 'Date', 'HomeTeam', 'AwayTeam', 'FTHG', 'FTAG', 'FTR', 'HTHG',
      'HTAG', 'HTR', 'HS', 'AS', 'HST', 'AST', 'HF', 'AF', 'HC', 'AC', 'HY',
      'AY', 'HR', 'AR', 'B365H', 'B365D', 'B365A', 'BWH', 'BWD', 'BWA', 'IWH',
      'IWD', 'IWA', 'PSH', 'PSD', 'PSA', 'WHH', 'WHD', 'WHA', 'VCH', 'VCD',
      'VCA', 'Bb1X2', 'BbMxH', 'BbAvH', 'BbMxD', 'BbAvD', 'BbMxA', 'BbAvA',
      'Bb0U', 'BbMx>2.5', 'BbAv>2.5', 'BbMx<2.5', 'BbAv<2.5', 'BbAH', 'BbAHh',
      'BbMxAHH', 'BbAvAHH', 'BbMxAHA', 'BbAvAHA', 'PSCH', 'PSCD', 'PSCA'],
      dtype='object')
```

```
In [59]:
```

```
In [60]: # Agafa un conjunt de dades de tema esportiu que t'agradi i normalitza els atributs categòrics en dummy.
# Estandaritza els atributs numèrics amb StandardScaler.

df3= df[['Div', 'Date', 'HomeTeam', 'AwayTeam', 'BbAH', 'BbAHh', 'BbMxAHH', 'BbAvAHH', 'BbMxAHA', 'BbAvAHA']].copy()
df3 = df3.rename(columns={'BbAH': 'partits_jugats', 'BbAHh': 'gols_casa_fora',
      'BbMxAHH': 'mitjana_gols_casa', 'BbAvAHH': 'mitjana_gols_fora', 'BbMxAHA': 'mitjana_gols_temp_ant_casa',
      'BbAvAHA': 'mitjana_gols_temp_ant_fora'})

df3
```

Out[60]:

	Div	Date	HomeTeam	AwayTeam	partits_jugats	gols_casa_fora	mitjana_gols_casa	mitjana_gols_fora	mitjana_gols_temp_ant_casa	mitjana_gols_temp_ant_fora
0	SP1	17/08/2018	Betis	Levante	20	-0.75	1.89	1.85	2.07	2.00
1	SP1	17/08/2018	Girona	Valladolid	20	-0.75	2.06	2.01	1.90	1.85
2	SP1	18/08/2018	Barcelona	Alaves	19	-2.50	1.95	1.91	2.00	1.95
3	SP1	18/08/2018	Celta	Espanol	18	-0.75	2.26	2.18	1.74	1.71
4	SP1	18/08/2018	Villarreal	Sociedad	18	-0.25	1.76	1.74	2.23	2.14
...
375	SP1	18/05/2019	Levante	Ath Madrid	18	0.50	1.91	1.87	2.04	1.99
376	SP1	18/05/2019	Sevilla	Ath Bilbao	19	-1.00	2.70	2.60	1.60	1.53
377	SP1	18/05/2019	Valladolid	Valencia	20	1.50	1.78	1.74	2.20	2.13
378	SP1	19/05/2019	Eibar	Barcelona	19	0.50	2.03	1.98	1.92	1.88
379	SP1	19/05/2019	Real Madrid	Betis	23	-1.00	1.94	1.91	2.01	1.98

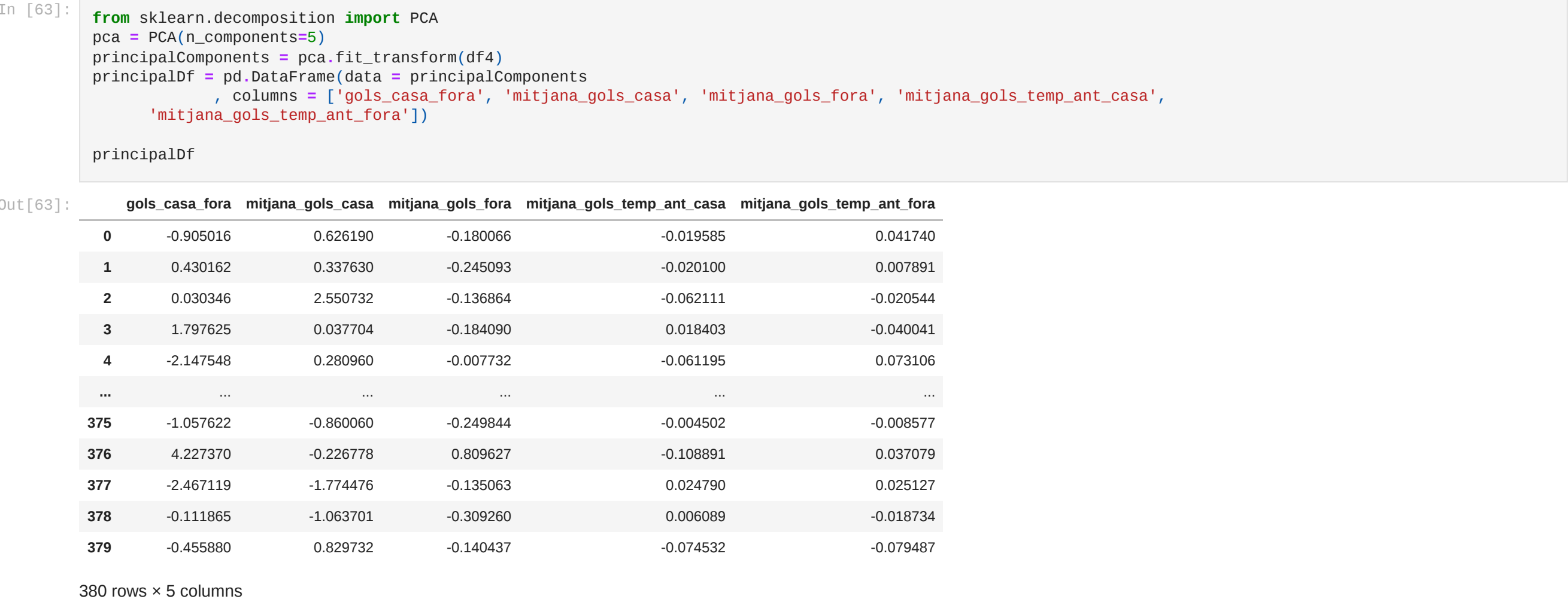
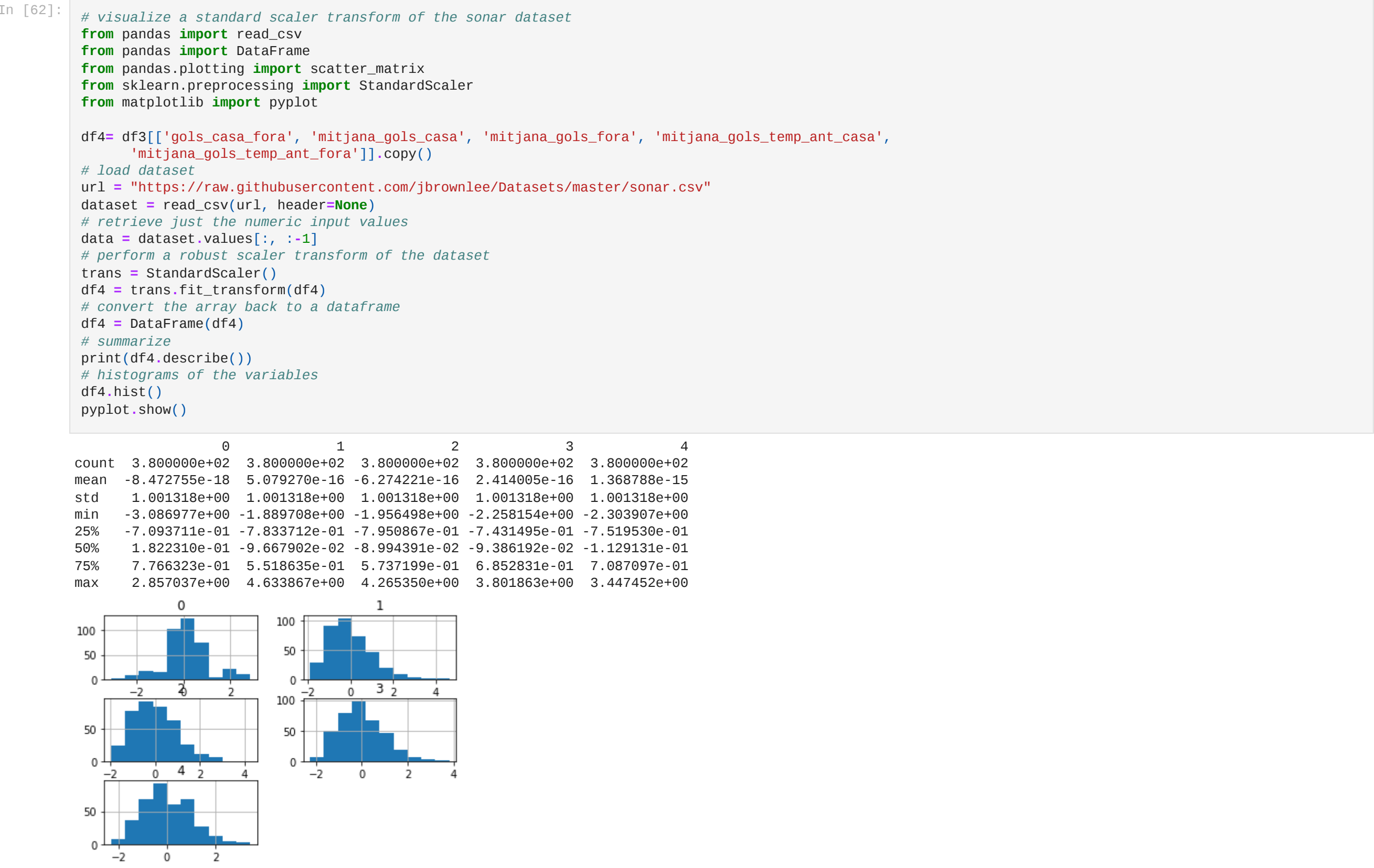
380 rows × 10 columns

```
In [61]: # Columna Target
targ_column = ['Date']
print(f'The target column is: {targ_column}')

# Columnes Numèriques
num_columns = list(df3.select_dtypes('number').columns)
print(f"Numerical columns are: {num_columns}")

# Columnes Categòriques (= les que no són numèriques i que no són el target)
cat_columns = list(set(df3.columns) - set(num_columns) - set(targ_column))
print(f"Categorical columns are: {cat_columns}")

The target column is: ['Date']
Numerical columns are: ['partits_jugats', 'gols_casa_fora', 'mitjana_gols_casa', 'mitjana_gols_fora', 'mitjana_gols_temp_ant_casa', 'mitjana_gols_temp_ant_fora']
Categorical columns are: ['AwayTeam', 'Div', 'HomeTeam']
```



```
In [ ]:
```