```
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")
                        Date HomeTeam AwayTeam FTHG FTAG FTR HTHG HTAG HTR ... BbAv<2.5 BbAH BbAHh BbMxAHH BbMxAHA BbAvAHA PSCH PSCD PSCA
Out[57]:
               Div
            0 SP1 17/08/2018
                                                                                   Α ...
                                                                                                                               1.85
                                                             3
                                                                        0
                                                                                              2.06
                                                                                                      20
                                                                                                           -0.75
                                                                                                                      1.89
                                                                                                                                         2.07
                                                                                                                                                   2.00
                                                                                                                                                         1.59
                                                                                                                                                               4.42
                                                                                                                                                                      5.89
                                   Betis
                                           Levante
                                                      0
                                                                              1
            1 SP1 17/08/2018
                                          Valladolid
                                                                 D
                                                                                   D ...
                                                                                              1.71
                                                                                                           -0.75
                                                                                                                      2.06
                                                                                                                               2.01
                                                                                                                                         1.90
                                                                                                                                                               3.57
                                                                                                                                                                      5.62
                                  Girona
                                                                                                      20
                                                                                                                                                   1.85
                                                                                                                                                         1.76
            2 SP1 18/08/2018
                                                             0
                                                                        0
                                                                                   D ...
                                                       3
                                                                 Η
                                                                              0
                                                                                                           -2.50
                                                                                                                      1.95
                                                                                                                               1.91
                                                                                                                                         2.00
                                                                                                                                                   1.95
                                                                                                                                                         1.10
                                                                                                                                                              11.85
                                                                                                                                                                     32.17
                               Barcelona
                                            Alaves
                                                                                              3.18
                                                                                                      19
            3 SP1 18/08/2018
                                                                        0
                                   Celta
                                           Espanol
                                                                                              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
                                                                 Α
                                                                        1
                                                                              1
                                                                                   D ...
                                                                                              1.81
                                                                                                      18
                                                                                                           -0.25
                                                                                                                      1.76
                                                                                                                               1.74
                                                                                                                                         2.23
                                                                                                                                                         2.32
                                                                                                                                                               3.21
                                                                                                                                                                      3.53
                                                                                                                                                   2.14
          375 SP1 18/05/2019
                                                                        2
                                                                                   Н ...
                                         Ath Madrid
                                                      2
                                                             2
                                                                 D
                                                                              0
                                                                                              2.23
                                                                                                           0.50
                                                                                                                      1.91
                                                                                                                               1.87
                                                                                                                                         2.04
                                                                                                                                                   1.99
                                                                                                                                                                      1.81
                                 Levante
                                                                                                      18
                                                                                                                                                         4.34
                                                                                                                                                               4.10
          376 SP1 18/05/2019
                                                                                                                               2.60
                                  Sevilla
                                         Ath Bilbao
                                                                                              2.08
                                                                                                      19
                                                                                                           -1.00
                                                                                                                                         1.60
                                                                                                                                                   1.53
                                                                                                                                                         2.17
                                                                                                                                                               3.08
                                                                                                                                                                      4.15
          377 SP1 18/05/2019
                                                             2
                                                                        0
                                Valladolid
                                           Valencia
                                                      0
                                                                                   Α ...
                                                                                              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
                                                                                   D ...
                                                                                              2.65
                                                                                                      19
                                                                                                            0.50
                                                                                                                      2.03
                                                                                                                               1.98
                                                                                                                                         1.92
                                                                                                                                                   1.88
                                                                                                                                                         4.96
                                                                                                                                                                4.55
                                                                                                                                                                      1.65
                                                                        0
                                                                                   D ...
          379 SP1 19/05/2019 Real Madrid
                                             Betis
                                                       0
                                                                                              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
          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',
                  'BbOU', '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', 'BbAvAHH', 'BbAvAHH', '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
                        Date HomeTeam AwayTeam partits_jugats
                                                               gols_casa_fora mitjana_gols_casa mitjana_gols_fora mitjana_gols_temp_ant_casa mitjana_gols_temp_ant_fora
               Div
Out[60]:
            0 SP1 17/08/2018
                                                            20
                                                                        -0.75
                                                                                         1.89
                                                                                                         1.85
                                                                                                                                  2.07
                                                                                                                                                           2.00
                                   Betis
                                           Levante
            1 SP1 17/08/2018
                                          Valladolid
                                                            20
                                                                        -0.75
                                                                                         2.06
                                                                                                         2.01
                                                                                                                                  1.90
                                                                                                                                                          1.85
                                  Girona
            2 SP1 18/08/2018
                               Barcelona
                                            Alaves
                                                            19
                                                                        -2.50
                                                                                         1.95
                                                                                                         1.91
                                                                                                                                  2.00
                                                                                                                                                          1.95
            3 SP1 18/08/2018
                                                            18
                                                                                         2.26
                                                                                                                                  1.74
                                                                                                                                                          1.71
                                   Celta
                                           Espanol
                                                                        -0.75
                                                                                                         2.18
            4 SP1 18/08/2018
                                Villarreal
                                          Sociedad
                                                            18
                                                                        -0.25
                                                                                         1.76
                                                                                                         1.74
                                                                                                                                  2.23
                                                                                                                                                          2.14
          375 SP1 18/05/2019
                                        Ath Madrid
                                                            18
                                                                        0.50
                                                                                         1.91
                                                                                                         1.87
                                                                                                                                  2.04
                                                                                                                                                          1.99
                                 Levante
                                                            19
                                                                        -1.00
                                                                                         2.70
                                                                                                         2.60
                                                                                                                                                          1.53
          376 SP1 18/05/2019
                                  Sevilla
                                         Ath Bilbao
                                                                                                                                  1.60
          377 SP1 18/05/2019
                                Valladolid
                                                                        0.50
                                                                                         2.03
                                                                                                         1.98
                                                                                                                                  1.92
                                                                                                                                                          1.88
          378 SP1 19/05/2019
                                         Barcelona
          379 SP1 19/05/2019 Real Madrid
                                                                        -1.00
                                                                                         1.94
                                                                                                         1.91
                                                                                                                                  2.01
                                                                                                                                                          1.98
                                              Betis
         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_for
          Categorical columns are: ['AwayTeam', 'Div', 'HomeTeam']
In [62]:
           # visualize a standard scaler transform of the sonar dataset
           from pandas import read_csv
           from pandas import DataFrame
           from pandas.plotting import scatter_matrix
           from sklearn.preprocessing import StandardScaler
           from matplotlib import pyplot
           df4= df3[['gols_casa_fora', 'mitjana_gols_casa', 'mitjana_gols_fora', 'mitjana_gols_temp_ant_casa',
                  'mitjana_gols_temp_ant_fora']].copy()
           # load dataset
           url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/sonar.csv"
           dataset = read_csv(url, header=None)
           # retrieve just the numeric input values
           data = dataset.values[:, :-1]
           # perform a robust scaler transform of the dataset
           trans = StandardScaler()
           df4 = trans.fit_transform(df4)
           # convert the array back to a dataframe
           df4 = DataFrame(df4)
           # summarize
           print(df4.describe())
           # histograms of the variables
           df4.hist()
           pyplot.show()
          count 3.800000e+02 3.800000e+02 3.800000e+02 3.800000e+02 3.800000e+02
          mean -8.472755e-18 5.079270e-16 -6.274221e-16 2.414005e-16 1.368788e-15
                 1.001318e+00 1.001318e+00 1.001318e+00 1.001318e+00 1.001318e+00
                -3.086977e+00 -1.889708e+00 -1.956498e+00 -2.258154e+00 -2.303907e+00
          min
                -7.093711e-01 -7.833712e-01 -7.950867e-01 -7.431495e-01 -7.519530e-01
          25%
                 1.822310e-01 -9.667902e-02 -8.994391e-02 -9.386192e-02 -1.129131e-01
          50%
                 7.766323e-01 5.518635e-01 5.737199e-01 6.852831e-01 7.087097e-01
          75%
                 2.857037e+00 4.633867e+00 4.265350e+00 3.801863e+00 3.447452e+00
          max
                                     100
          100
                                      50
           50
                                     100 -
           50
                       4 2
           50
In [63]:
           from sklearn.decomposition import PCA
           pca = PCA(n_components=5)
           principalComponents = pca.fit_transform(df4)
           principalDf = pd.DataFrame(data = principalComponents
                         , columns = ['gols_casa_fora', 'mitjana_gols_casa', 'mitjana_gols_fora', 'mitjana_gols_temp_ant_casa',
                  'mitjana_gols_temp_ant_fora'])
           principalDf
Out[63]:
               gols_casa_fora mitjana_gols_casa mitjana_gols_fora mitjana_gols_temp_ant_casa mitjana_gols_temp_ant_fora
                    -0.905016
            0
                                    0.626190
                                                    -0.180066
                                                                             -0.019585
                                                                                                      0.041740
                    0.430162
                                    0.337630
                                                    -0.245093
                                                                             -0.020100
                                                                                                      0.007891
            1
                                                    -0.136864
            2
                    0.030346
                                     2.550732
                                                                             -0.062111
                                                                                                      -0.020544
            3
                    1.797625
                                    0.037704
                                                    -0.184090
                                                                              0.018403
                                                                                                      -0.040041
            4
                    -2.147548
                                    0.280960
                                                    -0.007732
                                                                             -0.061195
                                                                                                      0.073106
                                                    -0.249844
          375
                    -1.057622
                                    -0.860060
                                                                             -0.004502
                                                                                                      -0.008577
                    4.227370
                                    -0.226778
                                                     0.809627
                                                                             -0.108891
          376
                                                                                                      0.037079
                                    -1.774476
                                                    -0.135063
          377
                    -2.467119
                                                                              0.024790
                                                                                                      0.025127
          378
                    -0.111865
                                    -1.063701
                                                    -0.309260
                                                                              0.006089
                                                                                                      -0.018734
          379
                    -0.455880
                                    0.829732
                                                    -0.140437
                                                                             -0.074532
                                                                                                      -0.079487
```

```
380 rows × 5 columns
```

```
In [76]:
    cmap = sns.cubehelix_palette(rot=-.2, as_cmap=True)
    g = sns.relplot(
        data=principalDf,
        x="mitjana_gols_casa", y="mitjana_gols_fora",

        palette=cmap, sizes=(10, 200),
    )
    g.set(xscale="log", yscale="log")
    g.ax.xaxis.grid(True, "minor", linewidth=.25)
    g.ax.yaxis.grid(True, "minor", linewidth=.25)
    g.despine(left=True, bottom=True)
```

```
Out[76]: <seaborn.axisgrid.FacetGrid at 0x1d78758a940>
```

