S06 T01: Tasca dades, probabilitats i estadístiques

In [1]:
```
import random
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
import re
import statistics
from scipy.stats import norm
from scipy import stats
from pandas.plotting import scatter_matrix
import plotly.express as px
import plotly.figure_factory as ff
import plotly.graph_objects as go
from IPython.display import HTML, display_html, display
```

Exercici 1. Agafa un conjunt de dades de tema esportiu que t'agradi i selecciona un atribut del conjunt de dades. Calcula la moda, la mediana, la desviació estàndard i la mitjana aritmètica.

De l'Sprint05, carreguem les dades netes, sense nuls, amb l'històric de jugadors de la selecció espanyola de futbol absolut masculina que han debutat (obtingudes a partir de la web bdfutbol.com). Recordem els noms de les columnes:

- Sobrenom; Nom; Data Naixement; Lloc de Naixament; Província; País; Partits Jugats; Partits Titular; Partits Complets; Partits Suplent; Partits Substituït; Partits Convocats (sense jugar); Partits Guanyats; Partits Empetats; Partits Perduts; Minuts; Goles; Gols Penalt; Goles pròpia porta; Gols Encaixats; Targetes grogues; Targetes vermelles; Edat inicial; Edat final; Alçada; Pes

In [2]:
```
#importem i li assignem un nom de dataframe
jugadors = pd.read_csv('C:\\Users\\Silvia\\Desktop\\rubenIT\\DataSources\\jugadores0
```

In [3]:
```
#Imprimim les dades filtrades per comprovar la importació
print(jugadors.describe())
print(jugadors.head(10))
print(jugadors.tail(10))
```

```
               PJ          PT          PC          PS          PX          PG  \
count  654.000000  654.000000  654.000000  654.000000  654.000000  654.000000
mean    14.155963   11.085627    8.006116    3.070336    3.056575    8.391437
std     22.460518   19.330256   14.271486    5.229901    7.115855   15.330149
min      1.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      2.000000    1.000000    1.000000    0.000000    0.000000    1.000000
50%      5.000000    4.000000    3.000000    1.000000    1.000000    3.000000
75%     16.000000   12.000000    9.000000    3.000000    3.000000    9.000000
max    180.000000  161.000000  125.000000   42.000000   59.000000  131.000000
```

```
               PE          PP           Min           G          GP  \
count  654.000000  654.000000    654.000000  654.000000  654.000000
mean     3.333333    2.431193   1005.507645    1.960245    0.142202
std      4.831199    3.607972   1669.924268    5.165109    0.873092
min      0.000000    0.000000      1.000000    0.000000    0.000000
25%      0.000000    0.000000     90.000000    0.000000    0.000000
50%      1.000000    1.000000    360.000000    0.000000    0.000000
75%      4.000000    3.000000   1129.250000    1.000000    0.000000
max     33.000000   23.000000  13709.000000   59.000000   11.000000

              GPP          GE          TA          TR          EI          EF  \
count  654.000000  654.000000  654.000000  654.000000  654.000000  654.000000
mean     0.019878    0.905199    0.917431    0.032110   23.949541   26.831804
std      0.139687    6.868723    2.419149    0.184904    2.782392    3.488660
min      0.000000    0.000000    0.000000    0.000000   17.000000   17.000000
25%      0.000000    0.000000    0.000000    0.000000   22.000000   25.000000
50%      0.000000    0.000000    0.000000    0.000000   24.000000   27.000000
75%      0.000000    0.000000    1.000000    0.000000   26.000000   29.000000
max      1.000000  100.000000   24.000000    2.000000   34.000000   36.000000

           Altura        Peso
count  654.000000  654.000000
mean   177.594801   73.915902
std      6.021862    5.713472
min    160.000000   60.000000
25%    173.000000   70.000000
50%    178.000000   74.000000
75%    181.750000   77.000000
max    197.000000   95.000000
            Apodo                         Nombre       Fecha             Ciudad  \
0    Marcos Vales         Marcos Vales Illanes  05/04/1975          A Coruña
1           Acuña              Juan Acuña Naya  13/02/1923          A Coruña
2          Martín  José María Martín Rodríguez  25/04/1924          A Coruña
3         Casilla     Francisco Casilla Cortés  02/10/1986           Alcover
4    Juan Sánchez    Juan Ginés Sánchez Romero  15/05/1972            Aldaia
5       Cucurella        Marc Cucurella Saseta  22/07/1998            Alella
6          Piquer          Vicente Piquer Mora  24/02/1935  Algar de Palancia
7             Ito        Antonio Álvarez Pérez  21/01/1975       Almendralejo
8       Planas II           Javier Planas Abad  03/07/1949          Almudévar
9  Josep Martínez         Josep Martínez Riera  27/05/1998             Alzira

   Provincia    País  PJ  PT  PC  PS  ...  G  GP  GPP  GE  TA  TR  EI  EF  \
0   A Coruña  España   1   0   0   1  ...  0   0    0   0   0   0  23  23
1   A Coruña  España   1   0   0   1  ...  0   0    0   1   0   0  18  18
2   A Coruña  España   1   1   1   0  ...  0   0    0   0   0   0  28  28
3  Tarragona  España   1   0   0   1  ...  0   0    0   1   0   0  28  28
4   Valencia  España   1   0   0   1  ...  0   0    0   0   0   0  26  26
5  Barcelona  España   1   1   0   0  ...  0   0    0   0   0   0  22  22
6   Valencia  España   1   1   1   0  ...  0   0    0   0   0   0  26  26
7    Badajoz  España   1   0   0   1  ...  0   0    0   0   0   0  23  23
8     Huesca  España   1   1   1   0  ...  0   0    0   0   1   0  25  25
9   Valencia  España   1   0   0   1  ...  0   0    0   0   0   0  23  23

   Altura  Peso
0   181.0  77.0
1   179.0  88.0
2   176.0  74.0
3   192.0  83.0
4   173.0  72.0
5   172.0  68.0
6   173.0  71.0
7   175.0  70.0
8   174.0  74.0
9   191.0  78.0
```

```
[10 rows x 25 columns]
              Apodo                      Nombre       Fecha           Ciudad  \
644         Fàbregas    Francesc Fàbregas Soler  04/05/1987    Arenys de Mar
645  Fernando Torres  Fernando José Torres Sanz  20/03/1984      Fuenlabrada
646      Xabi Alonso        Xabier Alonso Olano  25/11/1981           Tolosa
647            Silva  David Josué Jiménez Silva  08/01/1986      Arguineguín
648      Zubizarreta  Andoni Zubizarreta Urreta  23/10/1961  Vitoria-Gasteiz
649          Iniesta       Andrés Iniesta Luján  11/05/1984    Fuentealbilla
650         Busquets     Sergio Busquets Burgos  16/07/1988         Sabadell
651             Xavi     Xavier Hernández Creus  25/01/1980         Terrassa
652         Casillas    Iker Casillas Fernández  20/05/1981         Móstoles
653     Sergio Ramos        Sergio Ramos García  30/03/1986            Camas

       Provincia    País   PJ   PT   PC  PS  ...   G  GP  GPP   GE  TA  TR  \
644    Barcelona  España  110   68   22  42  ...  15   0    0    0  15   0
645       Madrid  España  110   75   21  35  ...  38   5    0    0   4   0
646     Gipuzkoa  España  114   86   48  28  ...  16   6    0    0  10   1
647   Las Palmas  España  125   96   37  29  ...  35   2    0    0  10   0
648  Araba/Álava  España  126  125  106   1  ...   0   0    1  100   2   1
649     Albacete  España  131  105   47  26  ...  13   1    0    0   4   0
650    Barcelona  España  133  119   89  14  ...   2   0    0    0  23   0
651    Barcelona  España  133  108   64  25  ...  13   0    0    0   9   0
652       Madrid  España  167  154  125  13  ...   0   0    0   93   2   0
653      Sevilla  España  180  161  118  19  ...  23   8    0    0  24   0

     EI  EF  Altura  Peso
644  18  29   180.0  77.0
645  19  30   186.0  78.0
646  21  32   183.0  75.0
647  20  32   170.0  67.0
648  23  36   187.0  86.0
649  22  34   171.0  68.0
650  20  32   189.0  76.0
651  20  34   170.0  68.0
652  19  35   182.0  80.0
653  18  34   184.0  83.0

[10 rows x 25 columns]
```

In [4]:

```
#analitzem el percentatge de NaN per cada un dels camps
(jugadors.isnull().sum())*100 / len(jugadors)
```

Out[4]:

```
Apodo        0.0
Nombre       0.0
Fecha        0.0
Ciudad       0.0
Provincia    0.0
País         0.0
PJ           0.0
PT           0.0
PC           0.0
PS           0.0
PX           0.0
PG           0.0
PE           0.0
PP           0.0
Min          0.0
G            0.0
GP           0.0
GPP          0.0
GE           0.0
TA           0.0
```

```
TR          0.0
EI          0.0
EF          0.0
Altura      0.0
Peso        0.0
dtype: float64
```

In [5]:

```python
#Imprimim les dades filtrades
print(jugadors.describe())
print(jugadors.head(10))
print(jugadors.tail(10))
```

```
               PJ          PT          PC          PS          PX          PG  \
count  654.000000  654.000000  654.000000  654.000000  654.000000  654.000000
mean    14.155963   11.085627    8.006116    3.070336    3.056575    8.391437
std     22.460518   19.330256   14.271486    5.229901    7.115855   15.330149
min      1.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      2.000000    1.000000    1.000000    0.000000    0.000000    1.000000
50%      5.000000    4.000000    3.000000    1.000000    1.000000    3.000000
75%     16.000000   12.000000    9.000000    3.000000    3.000000    9.000000
max    180.000000  161.000000  125.000000   42.000000   59.000000  131.000000

               PE          PP           Min           G          GP  \
count  654.000000  654.000000    654.000000  654.000000  654.000000
mean     3.333333    2.431193   1005.507645    1.960245    0.142202
std      4.831199    3.607972   1669.924268    5.165109    0.873092
min      0.000000    0.000000      1.000000    0.000000    0.000000
25%      0.000000    0.000000     90.000000    0.000000    0.000000
50%      1.000000    1.000000    360.000000    0.000000    0.000000
75%      4.000000    3.000000   1129.250000    1.000000    0.000000
max     33.000000   23.000000  13709.000000   59.000000   11.000000

              GPP          GE          TA          TR          EI          EF  \
count  654.000000  654.000000  654.000000  654.000000  654.000000  654.000000
mean     0.019878    0.905199    0.917431    0.032110   23.949541   26.831804
std      0.139687    6.868723    2.419149    0.184904    2.782392    3.488660
min      0.000000    0.000000    0.000000    0.000000   17.000000   17.000000
25%      0.000000    0.000000    0.000000    0.000000   22.000000   25.000000
50%      0.000000    0.000000    0.000000    0.000000   24.000000   27.000000
75%      0.000000    0.000000    1.000000    0.000000   26.000000   29.000000
max      1.000000  100.000000   24.000000    2.000000   34.000000   36.000000

           Altura        Peso
count  654.000000  654.000000
mean   177.594801   73.915902
std      6.021862    5.713472
min    160.000000   60.000000
25%    173.000000   70.000000
50%    178.000000   74.000000
75%    181.750000   77.000000
max    197.000000   95.000000
            Apodo                      Nombre       Fecha             Ciudad  \
0    Marcos Vales        Marcos Vales Illanes  05/04/1975           A Coruña
1           Acuña             Juan Acuña Naya  13/02/1923           A Coruña
2          Martín  José María Martín Rodríguez  25/04/1924           A Coruña
3         Casilla    Francisco Casilla Cortés  02/10/1986            Alcover
4    Juan Sánchez   Juan Ginés Sánchez Romero  15/05/1972             Aldaia
5       Cucurella       Marc Cucurella Saseta  22/07/1998             Alella
6          Piquer         Vicente Piquer Mora  24/02/1935  Algar de Palancia
7             Ito       Antonio Álvarez Pérez  21/01/1975       Almendralejo
8       Planas II          Javier Planas Abad  03/07/1949          Almudévar
9  Josep Martínez        Josep Martínez Riera  27/05/1998             Alzira
```

```
   Provincia    País  PJ  PT  PC  PS  ...  G  GP  GPP  GE  TA  TR  EI  EF  \
0   A Coruña  España   1   0   0   1  ...  0   0    0   0   0   0  23  23
1   A Coruña  España   1   0   0   1  ...  0   0    0   1   0   0  18  18
2   A Coruña  España   1   1   1   0  ...  0   0    0   0   0   0  28  28
3  Tarragona  España   1   0   0   1  ...  0   0    0   1   0   0  28  28
4   Valencia  España   1   0   0   1  ...  0   0    0   0   0   0  26  26
5  Barcelona  España   1   1   0   0  ...  0   0    0   0   0   0  22  22
6   Valencia  España   1   1   1   0  ...  0   0    0   0   0   0  26  26
7    Badajoz  España   1   0   0   1  ...  0   0    0   0   0   0  23  23
8     Huesca  España   1   1   1   0  ...  0   0    0   0   1   0  25  25
9   Valencia  España   1   0   0   1  ...  0   0    0   0   0   0  23  23

   Altura  Peso
0   181.0  77.0
1   179.0  88.0
2   176.0  74.0
3   192.0  83.0
4   173.0  72.0
5   172.0  68.0
6   173.0  71.0
7   175.0  70.0
8   174.0  74.0
9   191.0  78.0

[10 rows x 25 columns]
              Apodo                      Nombre       Fecha           Ciudad  \
644         Fàbregas    Francesc Fàbregas Soler  04/05/1987    Arenys de Mar
645  Fernando Torres  Fernando José Torres Sanz  20/03/1984      Fuenlabrada
646      Xabi Alonso        Xabier Alonso Olano  25/11/1981           Tolosa
647            Silva  David Josué Jiménez Silva  08/01/1986      Arguineguín
648      Zubizarreta  Andoni Zubizarreta Urreta  23/10/1961  Vitoria-Gasteiz
649          Iniesta       Andrés Iniesta Luján  11/05/1984    Fuentealbilla
650         Busquets     Sergio Busquets Burgos  16/07/1988         Sabadell
651             Xavi     Xavier Hernández Creus  25/01/1980         Terrassa
652         Casillas    Iker Casillas Fernández  20/05/1981         Móstoles
653     Sergio Ramos        Sergio Ramos García  30/03/1986            Camas

       Provincia    País   PJ   PT   PC  PS  ...   G  GP  GPP   GE  TA  TR  \
644    Barcelona  España  110   68   22  42  ...  15   0    0    0  15   0
645       Madrid  España  110   75   21  35  ...  38   5    0    0   4   0
646     Gipuzkoa  España  114   86   48  28  ...  16   6    0    0  10   1
647   Las Palmas  España  125   96   37  29  ...  35   2    0    0  10   0
648  Araba/Álava  España  126  125  106   1  ...   0   0    1  100   2   1
649     Albacete  España  131  105   47  26  ...  13   1    0    0   4   0
650    Barcelona  España  133  119   89  14  ...   2   0    0    0  23   0
651    Barcelona  España  133  108   64  25  ...  13   0    0    0   9   0
652       Madrid  España  167  154  125  13  ...   0   0    0   93   2   0
653      Sevilla  España  180  161  118  19  ...  23   8    0    0  24   0

     EI  EF  Altura  Peso
644  18  29   180.0  77.0
645  19  30   186.0  78.0
646  21  32   183.0  75.0
647  20  32   170.0  67.0
648  23  36   187.0  86.0
649  22  34   171.0  68.0
650  20  32   189.0  76.0
651  20  34   170.0  68.0
652  19  35   182.0  80.0
653  18  34   184.0  83.0

[10 rows x 25 columns]
```

Escollim l'atribut d'alçada Altura. Podem trobar les definicions a internet:

*La **moda** es el valor más repetido (solo aplicable a variables discretas). La **mediana** es el valor que dentro del conjunto de datos es menor que el 50% de los datos y mayor que el 50% restante. La **desviación típica** mide la dispersión de los datos respecto a la media. Se trata de la raíz cuadrada de la varianza, que en sí misma no es una medida de dispersión. Para calcular la desviación típica usamos std y var para la varianza. (ddof=0 es necesario si quieres seguir la definición de desviación típica y varianza de algunas bibliografías, la razón es que hay un parámetro de ajuste que Pandas pone a 1, pero otras librerías ponen a 0). En Excel es la diferencia que hay entre DESVEST.M (ddof=1) y DESVEST.P (ddof=0). La **media aritmética** se define como la suma de N elementos dividida entre N. Se trata una medida bastante conocida entre la gente, aunque tiene el inconveniente de que es muy susceptible a valores extremos.*

In [6]:
```
#imprimim els valors demanats
print("Moda/Mode: " + str(jugadors["Altura"].mode()))
print("Mitjana/Median: "+ str(jugadors["Altura"].median()))
print("Desviació estàndard/std mètode M: " + str(jugadors["Altura"].std(ddof=0)))
print("Desviació estàndard/std mètode P: " + str(jugadors["Altura"].std(ddof=1)))
print("Mitjana aritmètica/Mean: " + str(jugadors["Altura"].mean()))
```

```
Moda/Mode: 0    180.0
dtype: float64
Mitjana/Median: 178.0
Desviació estàndard/std mètode M: 6.0172563169887505
Desviació estàndard/std mètode P: 6.02186194821565
Mitjana aritmètica/Mean: 177.5948012232416
```

Es pot observar com amb el mètode de Pandas *df.describe* també s'obtenen els valors de *mean* i *std*. La desviació de *df.describe* Pandas no coincideix amb la de *df.mean(ddof=0)*, perquè es fan servir metodologies de càlcul diferents. Dibuixem l'histograma i la funció de densitat de probabilitat

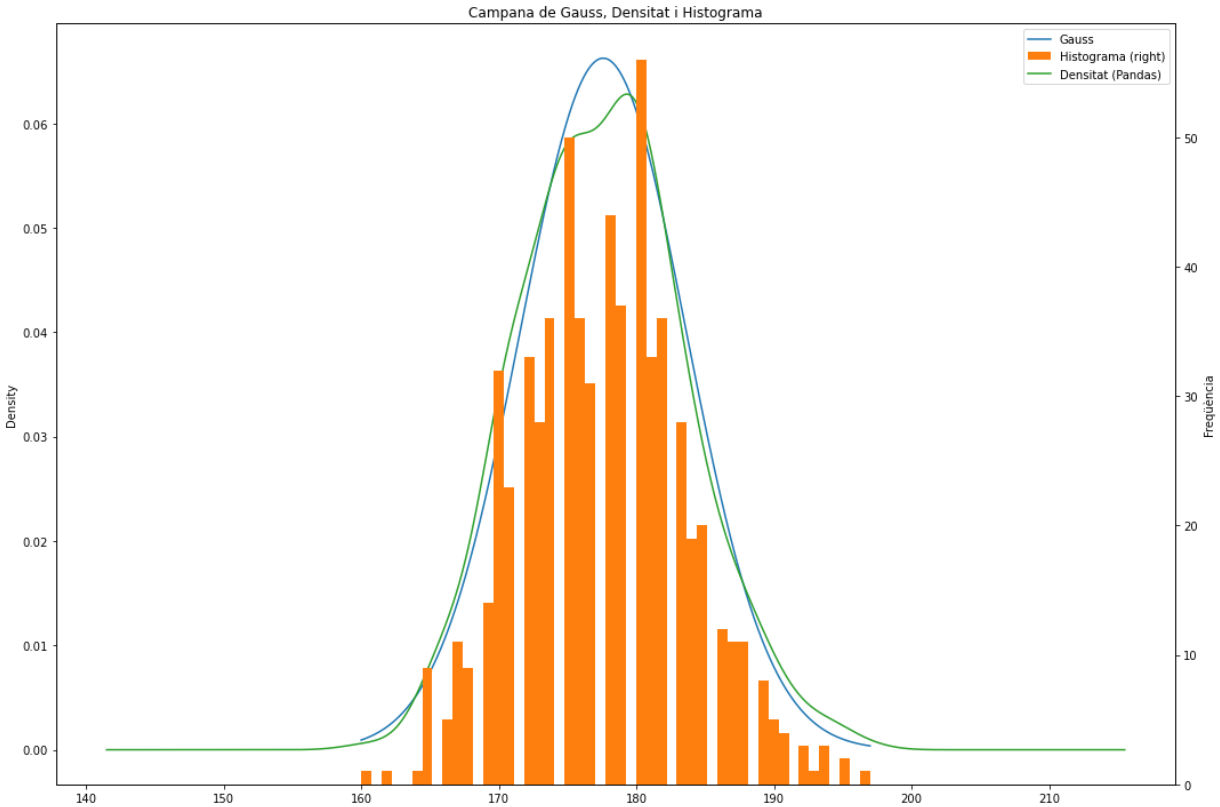
Dibuixem la gràfica de funció de densitat o campana de Gauss. Recordem el seu significat amb un abstracte de la wikipedia.

En estadística y probabilidad se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.1 La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.2 La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos.3Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

In [7]:
```
plt.figure(figsize=(15,10))

sd=jugadors["Altura"].std(ddof=1)
mean=jugadors["Altura"].mean()

#centrem la campana
maxim=jugadors["Altura"].max()
print("Mx: " + str(maxim))
minim=jugadors["Altura"].min()
print("Mn: " + str(minim))
```

```
meitat=maxim-(maxim-minim)/2
print("Meitat: " + str(meitat))

x_axis = np.arange(minim, maxim, 0.01)

ax=plt.plot(x_axis, norm.pdf(x_axis, mean, sd),label="Gauss")
plt.legend(loc="upper left")

ax2=jugadors["Altura"].plot.hist(bins=50,secondary_y=True,label="Histograma",legend=
ax3=jugadors["Altura"].plot.density(label="Densitat (Pandas)",legend=True)

plt.xlabel("Altura [cm]")
plt.ylabel("Freqüència")
plt.title("Campana de Gauss, Densitat i Histograma")

plt.tight_layout()

plt.show()
```

```
Mx: 197.0
Mn: 160.0
Meitat: 178.5
```



Hi ha 3 mètodes per comparar la correlació entre dos atributs.

In [8]:

```
print("Correlació Pearson: " + str(jugadors["Altura"].corr(jugadors["Peso"])))
print("Correlació Spearman: " + str(jugadors["Altura"].corr(jugadors["Peso"], method
print("Correlació Kendall: " + str(jugadors["Altura"].corr(jugadors["Peso"], method=
```

```
Correlació Pearson: 0.7763722499653491
Correlació Spearman: 0.7804597346851507
Correlació Kendall: 0.6141354845527803
```

Exercici 3. Continuant amb les dades de tema esportiu, calcula la correlació de tots els atributs entre

sí i representa'ls en una matriu amb diferents colors d'intensitat.

Podem també incloure tota la taula per trobar la correlació de tots els seus atributs:

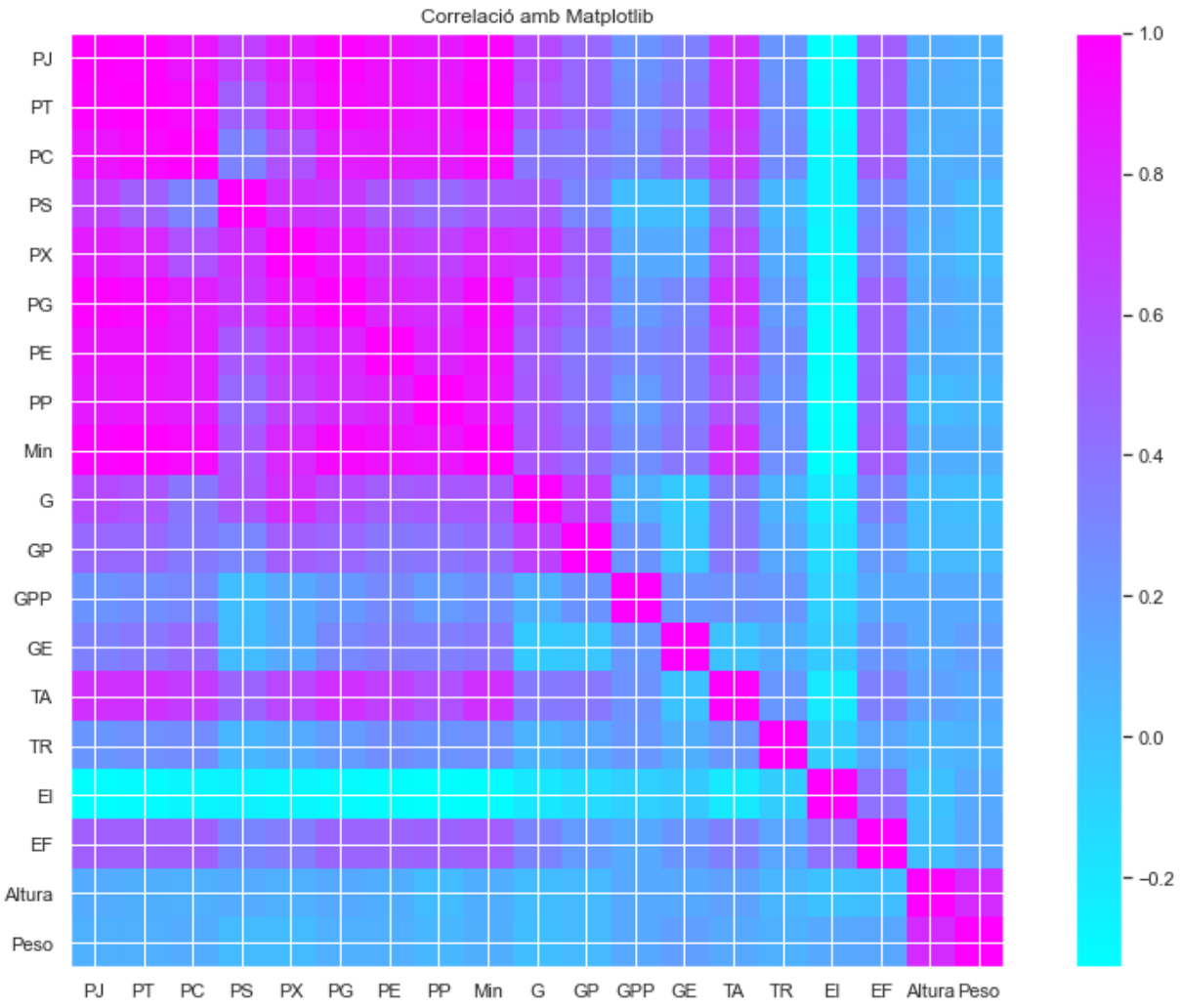
In [9]:
```
jugadors.corr()
```

Out[9]:

	PJ	PT	PC	PS	PX	PG	PE	PP	M
PJ	1.000000	0.979785	0.896164	0.673248	0.858977	0.979742	0.897177	0.861007	0.9841
PT	0.979785	1.000000	0.954129	0.511716	0.796538	0.946709	0.896426	0.876525	0.9973
PC	0.896164	0.954129	1.000000	0.322139	0.579190	0.843171	0.854262	0.852353	0.9575
PS	0.673248	0.511716	0.322139	1.000000	0.744908	0.708498	0.539765	0.457987	0.5402
PX	0.858977	0.796538	0.579190	0.744908	1.000000	0.876191	0.716011	0.665675	0.7829
PG	0.979742	0.946709	0.843171	0.708498	0.876191	1.000000	0.807364	0.769081	0.9505
PE	0.897177	0.896426	0.854262	0.539765	0.716011	0.807364	1.000000	0.815650	0.9017
PP	0.861007	0.876525	0.852353	0.457987	0.665675	0.769081	0.815650	1.000000	0.8806
Min	0.984180	0.997399	0.957517	0.540204	0.782900	0.950511	0.901738	0.880618	1.0000
G	0.612051	0.562434	0.385045	0.549723	0.753880	0.609220	0.515606	0.531202	0.5477
GP	0.465546	0.460589	0.366914	0.296963	0.511649	0.467107	0.384838	0.398102	0.4444
GPP	0.232324	0.265927	0.294920	0.014853	0.126741	0.203034	0.287435	0.198705	0.2637
GE	0.329781	0.376341	0.450643	0.025295	0.116914	0.293794	0.346189	0.341087	0.3805
TA	0.751259	0.745302	0.692993	0.471670	0.629845	0.754638	0.668120	0.575711	0.7460
TR	0.225199	0.248161	0.272679	0.049920	0.103368	0.192210	0.264002	0.231719	0.2481
EI	-0.326374	-0.309647	-0.277818	-0.257169	-0.281861	-0.300254	-0.320239	-0.327180	-0.3152
EF	0.508004	0.507343	0.511220	0.306503	0.350649	0.478332	0.479348	0.488172	0.5160
Altura	0.105709	0.094481	0.085917	0.104770	0.082912	0.116680	0.102767	0.024688	0.0981
Peso	0.084711	0.090873	0.107865	0.027924	0.028819	0.083478	0.090672	0.051238	0.0928

Fem la representació amb la llibreria Matplotlib i, per exemple, el colormap "cool".

In [20]:
```
plt.figure(figsize=(15,10))
jugcorr=jugadors.corr()
plt.imshow(jugcorr, cmap='cool', interpolation='nearest')
plt.colorbar()
plt.xticks(range(len(jugcorr)), jugcorr.columns)
plt.yticks(range(len(jugcorr)), jugcorr.index)
plt.title("Correlació amb Matplotlib")
plt.show()
```



Es pot fer amb la llibreria Pandas.

In [11]:
```
jugcorr01=jugcorr.style.background_gradient(cmap ='viridis').format(precision=2)

#Com és una taula, cal afegir títol des de HTML
display(HTML('<h2>Taula matriu colors amb Panda</h2>'))
display_html(jugcorr01)
```

Taula matriu colors amb Panda

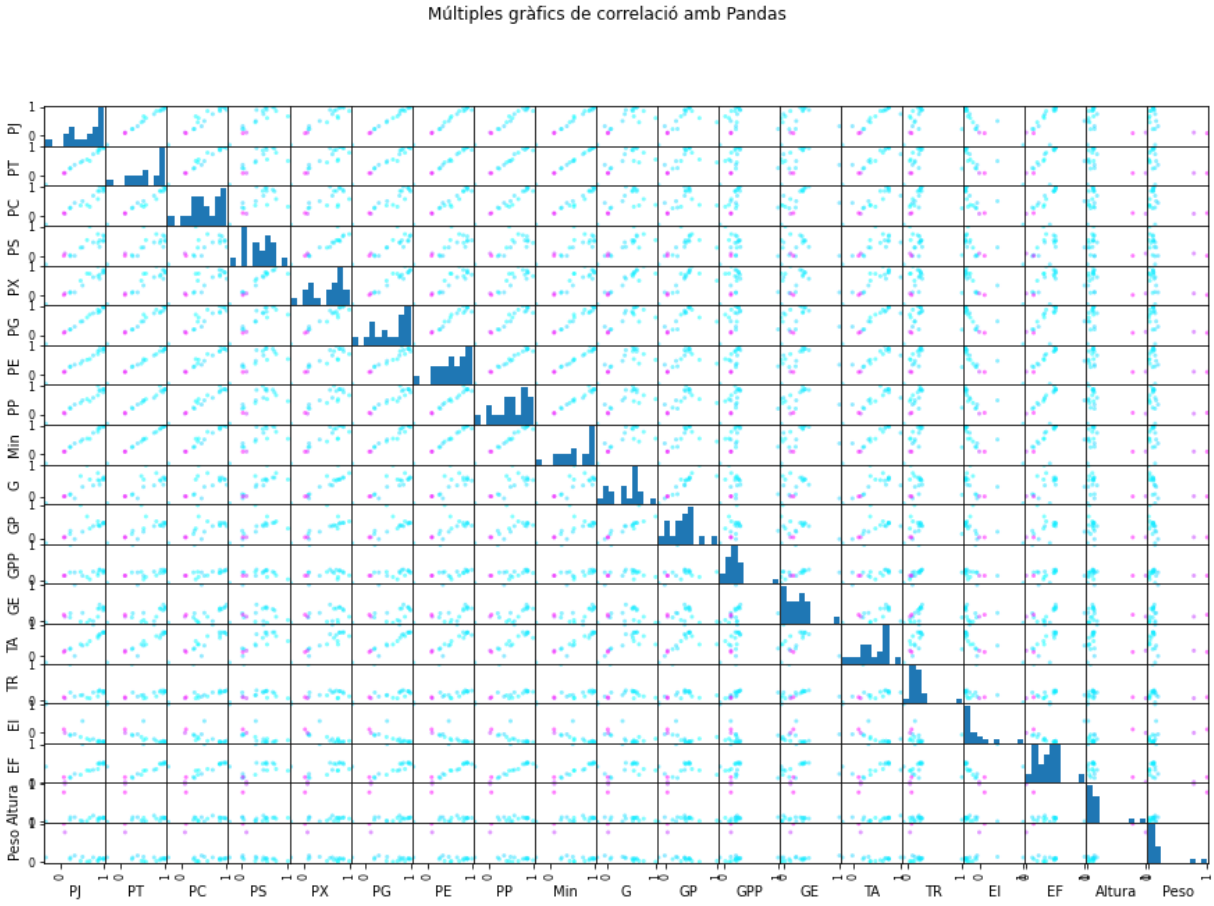
	PJ	PT	PC	PS	PX	PG	PE	PP	Min	G	GP	GPP	GE	TA	
PJ	1.00	0.98	0.90	0.67	0.86	0.98	0.90	0.86	0.98	0.61	0.47	0.23	0.33	0.75	0
PT	0.98	1.00	0.95	0.51	0.80	0.95	0.90	0.88	1.00	0.56	0.46	0.27	0.38	0.75	0
PC	0.90	0.95	1.00	0.32	0.58	0.84	0.85	0.85	0.96	0.39	0.37	0.29	0.45	0.69	0
PS	0.67	0.51	0.32	1.00	0.74	0.71	0.54	0.46	0.54	0.55	0.30	0.01	0.03	0.47	0
PX	0.86	0.80	0.58	0.74	1.00	0.88	0.72	0.67	0.78	0.75	0.51	0.13	0.12	0.63	0
PG	0.98	0.95	0.84	0.71	0.88	1.00	0.81	0.77	0.95	0.61	0.47	0.20	0.29	0.75	0
PE	0.90	0.90	0.85	0.54	0.72	0.81	1.00	0.82	0.90	0.52	0.38	0.29	0.35	0.67	0
PP	0.86	0.88	0.85	0.46	0.67	0.77	0.82	1.00	0.88	0.53	0.40	0.20	0.34	0.58	0
Min	0.98	1.00	0.96	0.54	0.78	0.95	0.90	0.88	1.00	0.55	0.44	0.26	0.38	0.75	0
G	0.61	0.56	0.39	0.55	0.75	0.61	0.52	0.53	0.55	1.00	0.66	0.08	-0.05	0.36	0

	PJ	PT	PC	PS	PX	PG	PE	PP	Min	G	GP	GPP	GE	TA	
GP	0.47	0.46	0.37	0.30	0.51	0.47	0.38	0.40	0.44	0.66	1.00	0.23	-0.02	0.37	0
GPP	0.23	0.27	0.29	0.01	0.13	0.20	0.29	0.20	0.26	0.08	0.23	1.00	0.21	0.24	0
GE	0.33	0.38	0.45	0.03	0.12	0.29	0.35	0.34	0.38	-0.05	-0.02	0.21	1.00	-0.01	0
TA	0.75	0.75	0.69	0.47	0.63	0.75	0.67	0.58	0.75	0.36	0.37	0.24	-0.01	1.00	0
TR	0.23	0.25	0.27	0.05	0.10	0.19	0.26	0.23	0.25	0.06	0.13	0.21	0.10	0.21	1
EI	-0.33	-0.31	-0.28	-0.26	-0.28	-0.30	-0.32	-0.33	-0.32	-0.21	-0.14	-0.09	-0.05	-0.22	-0
EF	0.51	0.51	0.51	0.31	0.35	0.48	0.48	0.49	0.52	0.32	0.20	0.12	0.22	0.33	0
Altura	0.11	0.09	0.09	0.10	0.08	0.12	0.10	0.02	0.10	0.02	0.03	0.12	0.12	0.16	0
Peso	0.08	0.09	0.11	0.03	0.03	0.08	0.09	0.05	0.09	0.01	0.03	0.12	0.17	0.12	0

Es podria fer amb Pandas i una matriu de dispersió (scatter), amb un gràfic per a cada una de les correlacions.

In [12]:

```
#Cal donar-li un valor variable a c per assignar-li els colors posteriorment amb col
pd.plotting.scatter_matrix(jugcorr, c=jugcorr["Peso"],cmap="cool",figsize=(15, 10))
plt.suptitle("Múltiples gràfics de correlació amb Pandas")
plt.show()
```



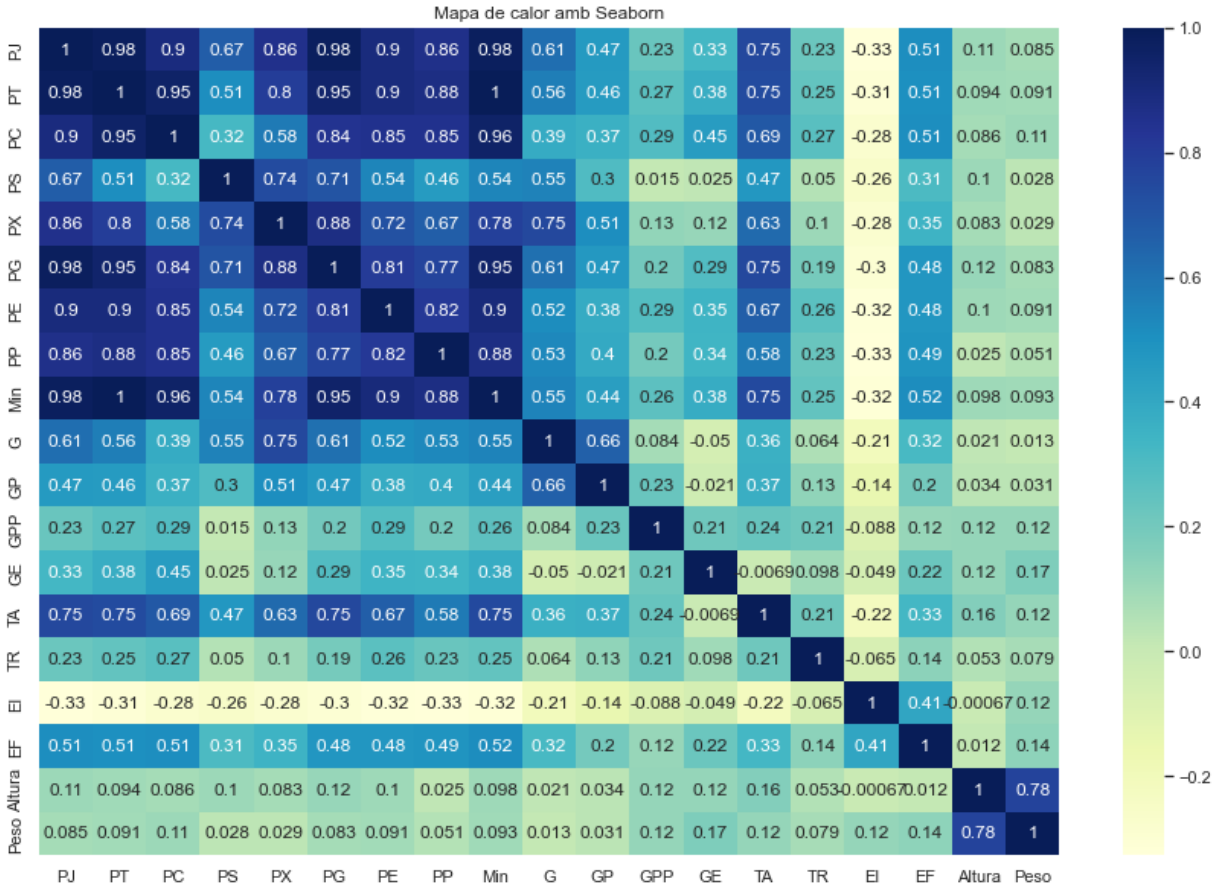
Podem fer-ho també amb la llibreria Seaborn.

In [13]:

```
sns.set(rc = {'figure.figsize':(15,10)})
sns.heatmap(jugadors.corr(),cmap="YlGnBu",annot=True)
plt.title("Mapa de calor amb Seaborn")
```

Out[13]:

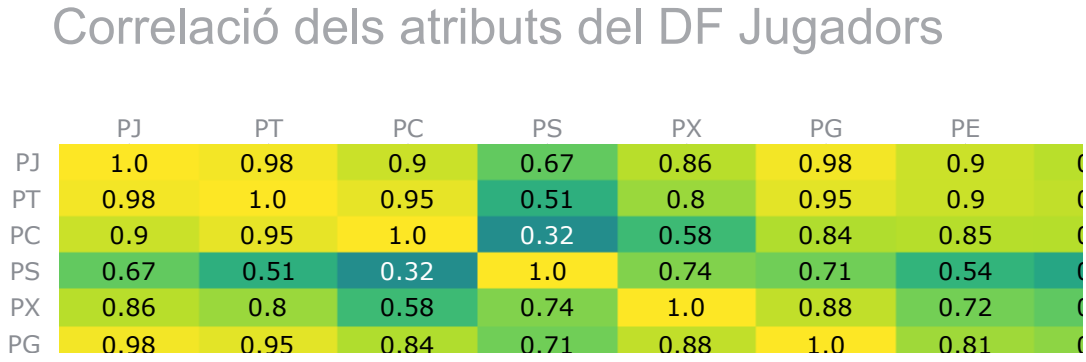
```
Text(0.5, 1.0, 'Mapa de calor amb Seaborn')
```



També es podria fer un mapa intereactiu de calor amb Plotly, tot i que els números es superposen, els colors es poden percebre.

In [14]:

```
correlation = jugadors.corr().values # obtener los numeros de la correlacion
names = list(jugadors.corr().columns.values) # obtener los nombres de las columnas
transposed_corr = correlation[::-1] # es necesario transponer la matriz
fig=ff.create_annotated_heatmap(transposed_corr, x = names,y = names[::-1], colorsca
                                annotation_text=transposed_corr.round(2),showscale=T
fig.update_layout(title= "Correlació dels atributs del DF Jugadors",
                  plot_bgcolor="#2d3035", paper_bgcolor="#FECB52",
                        title_font=dict(size=25, color="#a5a7ab", family="Muli, sans
                                        font=dict(color="#8a8d93"))
```



Es pot observar en els diferents mapes de calor com el pes i l'alçada tenen una alta correlació (s'apropen al color groc=1). També hi ha una correlació notable (colors verds clarets) entre partits jugats i minuts, amb targetes grogues rebudes. La relació entre partits jugats i guanyats, perduts, empetats o minuts també té una correlació molt alta (tons groguencs): a més partits, més probabilitats de tenir un desenllaç d'aquests tres, però com que les victòries predominen, els partits guanyats tenen major correlació.

Exercici 4. Continuant amb les dades de tema esportiu, selecciona un atribut i calcula la mitjana geomètrica i la mitjana harmònica.

Ara seleccioanarem l'atribut pes per calcular la mitjana geomètrica i harmònica. De la wikipedia trobem:

- La mitjana geomètrica o proporcional d'una quantitat finita de n nombres reals és l'arrel n-èsima del producte de tots els nombres.

$$Media\ geométrica = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_N}$$

- La mitjana harmònica d'una quantitat finita de n nombres a1,a2, ..., an, és igual a:

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

In [15]:
```
print("Mitjana geomètrica: " + str(round(statistics.geometric_mean(jugadors["Altura"
```

```
Mitjana geomètrica: 177.49
```

In [16]:
```
print("Mitjana harmònica: " + str(round(statistics.harmonic_mean(jugadors["Altura"])
```

```
Mitjana harmònica: 177.39
```

In [ ]: