

Ejercicios sobre Archivos CSV y el formato JSON

En las páginas del INE puede encontrarse distintos archivos Excel con los nombres de persona más frecuentes:

https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177009&menu=resultados&idp=1254734710990

Uno de ellos incluye todos los nombres con frecuencia igual o mayor a 20 personas. Contiene dos tablas, una con los masculinos y otra con los femeninos.

Para empezar, abre este archivo Excel con Access y generar dos archivos csv, uno con los nombres masculinos y otro con los femeninos.

Apartado 1

Diseña un programa en Python que cargue ambos archivos en sendas listas, y las almacene en sendas variables, una con los nombres masculinos y otra con los femeninos.

In [1]:

```
import csv
csvFile = csv.reader(open("hombres.csv", "r"), delimiter=";")
for row in csvFile:
    print(row)
```

```
['Frecuencias de nombres masculinos', '', '', '']
['Datos procedentes de la Estadística del Padrón Continuo a fecha 01/01/20
18', '', '', '']
['', '', '', '']
['Nombres masculinos con frecuencia mayor o igual a 20 a nivel nacional.',
'', '', '']
['(*) Edad media de las personas que tienen ese nombre a nivel nacional.',
'', '', '']
['', '', '', '']
['Orden', 'Nombre', 'Frecuencia', 'Edad Media (*)']
['1', 'ANTONIO', '678.425 ', '55,9']
['2', 'JOSE', '594.144 ', '61,1']
['3', 'MANUEL', '590.965 ', '54,9']
['4', 'FRANCISCO', '498.934 ', '57,4']
['5', 'DAVID', '365.196 ', '30,5']
['6', 'JUAN', '346.867 ', '55,7']
['7', 'JOSE ANTONIO', '310.134 ', '49,6']
['8', 'JAVIER', '306.504 ', '33,0']
['9', 'DANIEL', '296.281 ', '27,0']
```

Si intentamos acceder a un elemento de la "lista" csvFile (ej. csvFile[7]) vemos que no es en realidad una lista... por eso la reconvertimos así list(csvFile) .

Vemos que aparecen unas líneas iniciales que no nos interesan, así que nos quedamos con las que sí, y las guardamos en una variable:

Loading web-font STIX-Web/Normal/Italic

In [2]:



```
csvFile = csv.reader(open("hombres.csv", "r"), delimiter=";")
hombres = list(csvFile)[7:]
print(hombres[:20])
```

```
[['1', 'ANTONIO', '678.425 ', '55,9'], ['2', 'JOSE', '594.144 ', '61,1'],
['3', 'MANUEL', '590.965 ', '54,9'], ['4', 'FRANCISCO', '498.934 ', '57,4'],
['5', 'DAVID', '365.196 ', '30,5'], ['6', 'JUAN', '346.867 ', '55,7'], ['7',
'JOSE ANTONIO', '310.134 ', '49,6'], ['8', 'JAVIER', '306.504 ', '33,0'],
['9', 'DANIEL', '296.281 ', '27,0'], ['10', 'JOSE LUIS', '293.564 ', '53,
3'], ['11', 'FRANCISCO JAVIER', '287.711 ', '44,7'], ['12', 'CARLOS', '278.7
52 ', '38,8'], ['13', 'JESUS', '277.718 ', '48,5'], ['14', 'ALEJANDRO', '25
9.146 ', '25,5'], ['15', 'MIGUEL', '248.110 ', '46,5'], ['16', 'JOSE MANUE
L', '242.609 ', '47,2'], ['17', 'RAFAEL', '236.539 ', '50,6'], ['18', 'MIGUE
L ANGEL', '227.734 ', '43,6'], ['19', 'PEDRO', '225.485 ', '53,5'], ['20',
'ANGEL', '220.709 ', '50,1']]
```

In [3]:



```
csvFile = csv.reader(open("mujeres.csv", "r"), delimiter=";")
mujeres = list(csvFile)[7:]
print(mujeres[:20])
```

```
[['1', 'MARIA CARMEN', '656.276 ', '57,0'], ['2', 'MARIA', '606.048 ', '48,
6'], ['3', 'CARMEN', '391.563 ', '60,4'], ['4', 'JOSEFA', '276.682 ', '68,
0'], ['5', 'ANA MARIA', '273.319 ', '51,2'], ['6', 'ISABEL', '266.967 ', '5
7,4'], ['7', 'MARIA PILAR', '263.141 ', '57,0'], ['8', 'MARIA DOLORES', '25
9.216 ', '56,6'], ['9', 'LAURA', '256.381 ', '28,4'], ['10', 'MARIA TERESA',
'251.492 ', '57,1'], ['11', 'ANA', '250.124 ', '43,9'], ['12', 'CRISTINA',
'228.428 ', '33,7'], ['13', 'MARIA ANGELES', '226.047 ', '55,4'], ['14', 'MA
RTA', '225.323 ', '29,3'], ['15', 'FRANCISCA', '213.820 ', '64,9'], ['16',
'ANTONIA', '207.597 ', '64,7'], ['17', 'MARIA ISABEL', '204.354 ', '52,8'],
['18', 'MARIA JOSE', '203.283 ', '46,1'], ['19', 'LUCIA', '202.562 ', '23,
2'], ['20', 'DOLORES', '197.778 ', '67,0']]
```

Apartado 2

La lista con los nombres femeninos puede ampliar cada elemento con un valor adicional representativo del sexo femenino, una “F”. Resp., los nombres masculinos con una “M”.

In [4]:



```
for h in hombres:
    h.append("H")
print(hombres)

print("-----")

for m in mujeres:
    m.append("M")
print(mujeres)
```

```
[['1', 'ANTONIO', '678.425 ', '55,9', 'H'], ['2', 'JOSE', '594.144 ', '61,1', 'H'], ['3', 'MANUEL', '590.965 ', '54,9', 'H'], ['4', 'FRANCISCO', '498.934 ', '57,4', 'H'], ['5', 'DAVID', '365.196 ', '30,5', 'H'], ['6', 'JUAN', '346.867 ', '55,7', 'H'], ['7', 'JOSE ANTONIO', '310.134 ', '49,6', 'H'], ['8', 'JAVIER', '306.504 ', '33,0', 'H'], ['9', 'DANIEL', '296.281 ', '27,0', 'H'], ['10', 'JOSE LUIS', '293.564 ', '53,3', 'H'], ['11', 'FRANCISCO JAVIER', '287.711 ', '44,7', 'H'], ['12', 'CARLOS', '278.752 ', '38,8', 'H'], ['13', 'JESUS', '277.718 ', '48,5', 'H'], ['14', 'ALEJANDRO', '259.146 ', '25,5', 'H'], ['15', 'MIGUEL', '248.110 ', '46,5', 'H'], ['16', 'JOSE MANUEL', '242.609 ', '47,2', 'H'], ['17', 'RAFAEL', '236.539 ', '50,6', 'H'], ['18', 'MIGUEL ANGEL', '227.734 ', '43,6', 'H'], ['19', 'PEDRO', '225.485 ', '53,5', 'H'], ['20', 'ANGEL', '220.709 ', '50,1', 'H'], ['21', 'PABLO', '217.351 ', '27,1', 'H'], ['22', 'JOSE MARIA', '208.379 ', '53,9', 'H'], ['23', 'SERGIO', '206.458 ', '28,2', 'H'], ['24', 'FERNANDO', '204.544 ', '48,7', 'H'], ['25', 'LUIS', '194.269 ', '52,1', 'H'], ['26', 'JORGE', '192.771 ', '34,1', 'H'], ['27', 'ALBERTO', '179.750 ', '36,8', 'H'], ['28', 'JUAN CARLOS', '160.280 ', '45,5', 'H'], ['29', 'ALVARO', '158.234 ', '21,9', 'H'], ['30', 'JUAN JOSE', '153.954 ', '48,5', 'H'], ['31', 'ADRIAN', '152.686 ', '18,9', 'H'], ['32', 'DIEGO', '152.093 ', '3
```

Apartado 3

Ahora deseamos unir ambas listas y convertir los números en números enteros o reales, según el caso.

In [5]:

```

personas = [(int(orden.replace(".", "")), nombre,
              int(posic_punto.replace(".", "")),
              float(real_coma.replace('.', '').replace(',', '.')), sexo)
             for (orden, nombre, posit_punto, real_coma, sexo) in hombres + mujeres]

print(personas)

```

```

[(1, 'ANTONIO', 678425, 55.9, 'H'), (2, 'JOSE', 594144, 61.1, 'H'), (3, 'MANUEL', 590965, 54.9, 'H'), (4, 'FRANCISCO', 498934, 57.4, 'H'), (5, 'DAVID', 365196, 30.5, 'H'), (6, 'JUAN', 346867, 55.7, 'H'), (7, 'JOSE ANTONIO', 310134, 49.6, 'H'), (8, 'JAVIER', 306504, 33.0, 'H'), (9, 'DANIEL', 296281, 27.0, 'H'), (10, 'JOSE LUIS', 293564, 53.3, 'H'), (11, 'FRANCISCO JAVIER', 287711, 44.7, 'H'), (12, 'CARLOS', 278752, 38.8, 'H'), (13, 'JESUS', 277718, 48.5, 'H'), (14, 'ALEJANDRO', 259146, 25.5, 'H'), (15, 'MIGUEL', 248110, 46.5, 'H'), (16, 'JOSE MANUEL', 242609, 47.2, 'H'), (17, 'RAFAEL', 236539, 50.6, 'H'), (18, 'MIGUEL ANGEL', 227734, 43.6, 'H'), (19, 'PEDRO', 225485, 53.5, 'H'), (20, 'ANGEL', 220709, 50.1, 'H'), (21, 'PABLO', 217351, 27.1, 'H'), (22, 'JOSE MARIA', 208379, 53.9, 'H'), (23, 'SERGIO', 206458, 28.2, 'H'), (24, 'FERNANDO', 204544, 48.7, 'H'), (25, 'LUIS', 194269, 52.1, 'H'), (26, 'JORGE', 192771, 34.1, 'H'), (27, 'ALBERTO', 179750, 36.8, 'H'), (28, 'JUAN CARLOS', 160280, 45.5, 'H'), (29, 'ALVARO', 158234, 21.9, 'H'), (30, 'JUAN JOSE', 153954, 48.5, 'H'), (31, 'ADRIAN', 152686, 18.9, 'H'), (32, 'DIEGO', 152093, 32.0, 'H'), (33, 'RAUL', 140247, 31.2, 'H'), (34, 'JUAN ANTONIO', 134403, 48.9, 'H'), (35, 'IVAN', 133802, 25.0, 'H'), (36, 'ENRIQUE', 129768, 49.7, 'H'), (37, 'RUBEN', 128001, 27.1, 'H'), (38, 'RAMON', 123519, 57.5, 'H'), (39, 'OSCAR', 122779, 33.9, 'H').

```

Apartado 4

Ordenamos ahora la lista alfabéticamente...

In [6]:

```

personas.sort(key=lambda p : p[1])

print(personas)

```

```

[(22287, 'AADEL', 22, 30.9, 'H'), (9237, 'AADIL', 67, 32.8, 'H'), (14261, 'AALI', 39, 37.7, 'H'), (10871, 'AALIYAH', 52, 7.4, 'M'), (24601, 'AALTJE', 20, 69.2, 'M'), (19012, 'AAMAR', 27, 32.6, 'H'), (16264, 'AAMER', 33, 31.1, 'H'), (4281, 'AAMIR', 177, 32.3, 'H'), (22288, 'AARAB', 22, 47.0, 'H'), (19553, 'AARAV', 26, 3.2, 'H'), (109, 'AARON', 29056, 14.1, 'H'), (21508, 'AARON ALBERTO', 23, 21.1, 'H'), (8599, 'AARON ALEJANDRO', 73, 12.7, 'H'), (11376, 'AARON ALEXANDER', 52, 9.1, 'H'), (18520, 'AARON ANDRES', 28, 12.9, 'H'), (12492, 'AARON ANTONIO', 46, 17.9, 'H'), (8185, 'AARON DANIEL', 78, 11.8, 'H'), (4776, 'AARON DAVID', 152, 15.3, 'H'), (18521, 'AARON EDUARDO', 28, 18.0, 'H'), (20834, 'AARON ENRIQUE', 24, 17.3, 'H'), (20835, 'AARON FERNANDO', 24, 17.5, 'H'), (12974, 'AARON FRANCISCO', 44, 16.0, 'H'), (10871, 'AARON GABRIEL', 55, 8.9, 'H'), (17059, 'AARON ISAAC', 31, 13.4, 'H'), (24005, 'AARON ISMAEL', 20, 12.8, 'H'), (10872, 'AARON JAVIER', 55, 15.8, 'H'), (4642, 'AARON JESUS', 158, 20.5, 'H'), (23120, 'AARON JOEL', 21, 10.9, 'H'), (4316, 'AARON JOSE', 175, 17.2, 'H'), (12297, 'AARON JOSUE', 47, 18.5, 'H'), (6459, 'AARON MANUEL', 105, 19.0, 'H'), (10703, 'AARON MIGUEL', 56, 19.7, 'H'), (14870, 'AARON MOISES', 37, 20.6, 'H'), (12975, 'AARON NICOLAS', 44, 10.5, 'H'), (24006, 'AARON PAUL', 20, 21.3, 'H'), (17060, 'AARON RAFAEL', 31, 19.1, 'H'), (20836, 'AARON SANTIAGO', 24, 14.

```

Apartado 5

... y finalmente deseamos almacenar el resultado en un archivo cvs

In [7]:



```
ids_columnas = ['Orden', 'Nombre', 'Frecuencia', 'Edad Media (*)', 'Sexo']  
with open("personas.csv", 'w', newline='') as csv_archivo:  
    csvwriter = csv.writer(csv_archivo, delimiter=";")  
    csvwriter.writerow(ids_columnas)  
    csvwriter.writerows(personas)  
  
print("Hecho")
```

Hecho