



## FUNDAMENTOS DE PROGRAMACIÓN

### Ejercicio de Laboratorio: Nombres

Autor: José A. Troyano. Revisores: Mariano González, Toñi Reina. Última modificación: 27/10/2020

En este proyecto trabajaremos con datos correspondientes a los nombres de las personas nacidas en España desde 2002 a 2017. Los datos están tomados del Instituto Nacional de Estadística, donde se pueden encontrar muchos datos interesantes principalmente sobre la demografía, economía y sociedad españolas. Representaremos la información de entrada mediante listas de tuplas, y a partir de esta estructura implementaremos una serie de funciones que nos permitirán realizar varios tipos de consultas y generar visualizaciones.

Trabajaremos con ficheros en formato CSV. Cada registro del fichero de entrada ocupa una línea y contiene cuatro informaciones sobre los nombres (año, nombre, frecuencia, género). Estas son las primeras líneas de un fichero de entrada:

	A	B	C	D
1	Año	Nombre	Frecuencia	Género
2	2002	ALEJANDRO	8020	Hombre
3	2002	PABLO	5799	Hombre
4	2002	DANIEL	5603	Hombre
5	2002	DAVID	5414	Hombre
6	2002	ADRIAN	4949	Hombre
7	2002	JAVIER	4909	Hombre
8	2002	ALVARO	4595	Hombre
9	2002	SERGIO	3744	Hombre
10	2002	CARLOS	3593	Hombre
11	2002	MARIO	2876	Hombre

Figura 1: fichero de datos

Además de distintos indicadores, generaremos dos gráficas que mostrarán, respectivamente, la evolución del uso de un nombre determinado (Figura 2), o las frecuencias acumuladas de los nombres más populares (Figura 3).

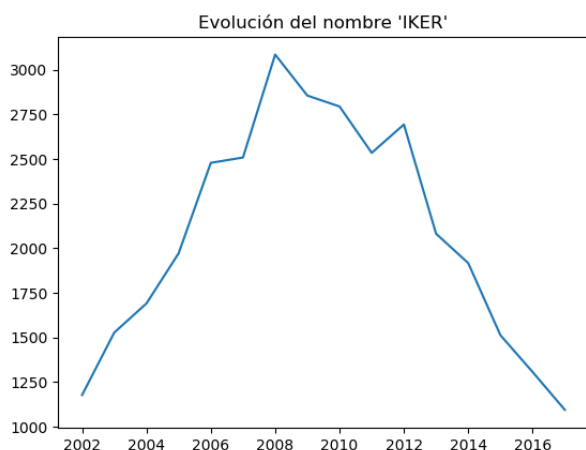


Figura 2: evolución de la frecuencia de un nombre

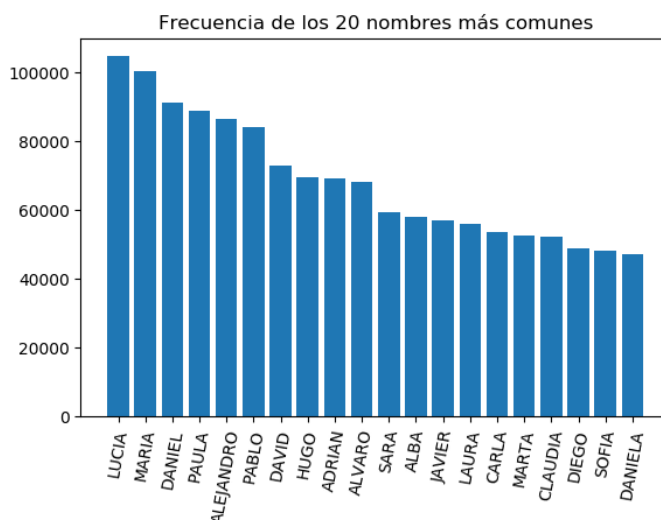


Figura 3: frecuencias de los nombres más populares



Para almacenar en Python la información de cada una de las líneas se usará la siguiente definición de `namedtuple`:

```
FrecuenciaNombre = namedtuple('FrecuenciaNombre', 'año,nombre,frecuencia,genero')
```

Cree un fichero **nombres.py** e incluya en él la definición del `namedtuple` anterior (recuerde que debe importar `namedtuple` del módulo `collections` para poder utilizarlo). A continuación, implemente las funciones que se le piden.

1. **leer\_frecuencias\_nombres**: recibe la ruta de un fichero CSV codificado en UTF-8, y devuelve una lista de tuplas de tipo `FrecuenciaNombre(int, str, int, str)` conteniendo todos los datos almacenados en el fichero.
2. **filtrar\_por\_genero**: recibe una lista de tuplas de tipo `FrecuenciaNombre` y un género de tipo `str`, y devuelve una lista de tuplas de tipo `FrecuenciaNombre` con los registros del género recibido como parámetro.
3. **calcular\_nombres**: recibe una lista de tuplas de tipo `FrecuenciaNombre` y un género de tipo `str`, y devuelve un conjunto `{str}` con los nombres del género recibido como parámetro. El género puede ser 'Hombre', 'Mujer' o tener un valor `None`, en cuyo caso se incluyen en el conjunto todos los nombres. El valor por defecto del género es `None`.
4. **calcular\_top\_nombres\_de\_año**: recibe una lista de tuplas de tipo `FrecuenciaNombre`, un año de tipo `int`, un número *límite* de tipo `int` y un género de tipo `str`, y devuelve una lista de tuplas (`nombre, frecuencia`) de tipo `(str, int)` con los nombres más frecuentes del año y el género dados, ordenada de mayor a menor frecuencia, y con un máximo de *límite* nombres. El género puede ser 'Hombre', 'Mujer' o tener un valor `None`, en cuyo caso se incluyen en la lista todos los nombres. El valor por defecto del límite es 10 y el del género es `None`.
5. **calcular\_nombres\_ambos\_generos**: recibe una lista de tuplas de tipo `FrecuenciaNombre`, y devuelve un conjunto `{str}` con los nombres que han sido utilizados en ambos géneros.
6. **calcular\_nombres\_compuestos**: recibe una lista de tuplas de tipo `FrecuenciaNombre` y un género de tipo `str`, y devuelve un conjunto `{str}` con los nombres que contienen más de una palabra. El género puede ser 'Hombre', 'Mujer' o tener un valor `None`, en cuyo caso se incluyen en el conjunto todos los nombres. El valor por defecto del género es `None`.
7. **calcular\_nombre\_mas\_frecuente\_por\_año**: recibe una lista de tuplas de tipo `FrecuenciaNombre` y un género de tipo `str`, y devuelve una lista de tuplas (`año, nombre, frecuencia`) de tipo `(int, str, int)` ordenada por año con el nombre más frecuente de cada año. El género puede ser 'Hombre', 'Mujer' o tener un valor `None`, en cuyo caso se incluyen en la lista todos los nombres. El valor por defecto del género es `None`. Se calculará en primer lugar la lista de años y, posteriormente, se buscará el nombre más frecuente para cada año.
8. **calcular\_frecuencia\_por\_año**: recibe una lista de tuplas de tipo `FrecuenciaNombre` y un nombre de tipo `str`, y devuelve una lista de tuplas (`año, frecuencia`) de tipo `(int, int)` ordenada por año con la frecuencia del nombre en cada año. En el caso de que un nombre se use para hombres y mujeres, se sumarán ambas frecuencias.
9. **mostrar\_evolucion\_por\_año**: recibe una lista de tuplas de tipo `FrecuenciaNombre` y un nombre de tipo `str`, y genera un gráfico con la evolución de la frecuencia del nombre a lo largo de los años (Figura 2). Se usarán las siguientes instrucciones para generar la gráfica:

```
plt.plot(años, frecuencias)
plt.title("Evolución del nombre '{}'.format(nombre))
plt.show()
```



Donde *años* y *frecuencias* se extraen del resultado de la función *calcular\_frecuencia\_por\_año*.

10. **calcular\_frecuencia\_acumulada**: recibe una lista de tuplas de tipo *FrecuenciaNombre* y un nombre de tipo *str*, y devuelve la frecuencia acumulada del nombre en todos los años.
11. **calcular\_frecuencias\_por\_nombre**: recibe una lista de tuplas de tipo *FrecuenciaNombre*, y devuelve un diccionario *{str: int}* que relaciona cada nombre con la frecuencia acumulada del nombre.
12. **mostrar\_frecuencias\_nombres**: recibe una lista de tuplas de tipo *FrecuenciaNombre* y un número *límite* de tipo *int*, y genera un diagrama de barras con las frecuencias de los nombres más populares, en orden decreciente de popularidad y con un máximo de *límite* nombres (Figura 3). El valor por defecto del límite es 10. Se usarán las siguientes instrucciones para generar la gráfica:

```
plt.bar(nombres, frecuencias)
plt.xticks(rotation=80)
plt.title("Frecuencia de los {} nombres más comunes".format(limite))
plt.show()
```

Donde *nombres* y *frecuencias* se extraen del resultado de la función *calcular\_frecuencias\_por\_nombre*. El cálculo de los nombres más populares se puede realizar ordenando las claves del diccionario devuelto por *calcular\_frecuencias\_por\_nombre* en función de sus valores asociados.

Cree un fichero **nombres\_TEST.py**. Importe todas las funciones del módulo *nombres*. Cargue los datos del fichero CSV y muestre en consola los datos leídos. Incluya llamadas a todas las funciones implementadas, mostrando los resultados en la consola.