

Estudios: Máster Universitario en Ingeniería y Ciencia de Datos

Asignatura: Programación en Entornos de Datos

TEST

Cada pregunta del test **sumará 1 punto** en caso de contestarse de forma correcta y **restará 0.3 puntos** en caso de contestarse de forma errónea. Las preguntas en blanco no restarán. Para superar esta parte, será necesario obtener **al menos 2 puntos sobre 4**.

1. Dado el siguiente fragmento de código, cuál será el contenido de a3 tras ejecutar dicho código:

```
a1 = np.arange(0,10)
a2 = np.arange(0,10,3)
a3 = a1[:] * a2[a2[1:2]]
```

- a) [0,3,6,9,12,15,18,21,24,27]
 - b) [0,9,18,27,36,45,54,63,72,81]
 - c) [0,0,0,0,0,0,0,0,0,0]
 - d) a3 no tendrá ningún contenido porque se genera un error al realizar la multiplicación
2. Que un tipo de datos sea inmutable significa que:
- a) Una vez se asigna un valor de ese tipo a una variable, ésta nunca podrá cambiar de valor mediante otra asignación
 - b) Los objetos de dicho tipo no disponen de métodos que permitan modificar su valor
 - c) Sólo existe un único valor de ese tipo
 - d) El valor del objeto desaparece tras ser consultado por primera vez
- Dado el siguiente fragmento de código:

3. Dado el siguiente fragmento de código:

```
import pandas as pd
import numpy as np
data = pd.DataFrame(np.arange(20).reshape((5,4)),
                    index=['X','Y','Z','W','V'],
                    columns=['X','Y','Z','W'])
```

¿Cuál de las siguientes opciones devolverá como resultados un DataFrame de 6 elementos?

- a) data.iloc[:, :3][:1]

- b) `data.iloc[:, :3][data.Z % 3 == 0]`
- c) `data.iloc[:, :2][data.Z % 2 == 0]`
- d) `data.iloc[:3, :][data.Z % 3 == 0]`

4. Dado el siguiente código:

```
import sympy
A = sympy.Matrix([[2,3],[5,4]])
b = sympy.Matrix([4,3])
```

que representa el sistema de ecuaciones:

$$\begin{aligned}2x_1 + 3x_2 &= 4 \\ 5x_1 + 4x_2 &= 3\end{aligned}$$

Para obtener la solución $[x_1, x_2]$ al sistema de ecuaciones hay que utilizar:

- a) `A.solve(b)`
- b) `A.solve_matrix(b)`
- c) `b.solve(A)`
- d) `solve(A,b)`

Problemas

Cada uno de estos problemas se puntuará sobre 3 puntos. Para superar esta parte será necesario obtener **al menos 2 puntos sobre 6**. Se recuerda a los estudiantes que los pequeños errores en el código (como nombres exactos de funciones, orden de los parámetros y similares) no se penalizarán.

Problema 1. El gobierno nos ha contratado para trabajar en un proyecto analizando los datos de los nacimientos de los últimos 100 años almacenados en un fichero CSV con la siguiente estructura:

- En cada línea se encuentran los campos de cada nacimiento, separados por comas.
- En la primera línea se encuentran los nombres de los campos, separados por comas, a modo de cabecera.
- De cada nacimiento se dispone de los siguientes datos: día, mes, año, día de la semana, hora, minuto, peso al nacer, sexo y nombre del bebé. Los campos día de la semana, sexo y nombre del bebé se encuentran codificados como texto, el peso como un real y el resto de campos como enteros.

Dada la gran cantidad de nacimientos durante un siglo, se hace necesario comprimir los datos de forma que puedan ser cargados y procesados, por lo que nuestro primer trabajo consiste en hacer algo para que ocupe lo mínimo posible en memoria.

Para realizar el ejercicio se pide explicar a alto nivel qué acciones tomaríamos para realizar esta compresión. A continuación, se pide escribir el código asociado a ello. Finalmente, añadir código para averiguar cuál es el día de la semana en el que se han producido un mayor número de nacimientos, así como el nombre de niña más común a lo largo de los 100 años de los que se guarda registro.

Problema 2. El juego de los euromillones es un juego de azar que consiste en elegir en cada apuesta 5 números de entre 50 (del número 1 al 50 inclusivos) sin repeticiones, y dos números de estrellas de entre 12 (del 1 al 12 inclusivos) sin repeticiones. En total, tenemos para cada apuesta tenemos 7 números (por ejemplo: 1, 2, 3, 48, 49, 50 – 3, 6).

Queremos generar aleatoriamente 100 combinaciones distintas y guardarlas a fichero para poder procesarlas. ¿Qué estructura de datos, de entre las estudiadas, usaría y cómo lo haría? Justifique sus decisiones.

Escriba el código necesario para crear la estructura y generar las distintas combinaciones. También escriba el código para guardar el contenido en un fichero. Justifique el formato de salida usado en el fichero.

¿Cómo obtendría el número más repetido, de los 50, así como el número de estrellas, de los 12, más repetido? Escriba el código.