

1) Indicar lo que mostrarían por pantalla las siguientes instrucciones de shell:

```
var = "DataScience"  
echo ${var//Data/} | tr 'a-z' 'A-Z'
```

- SCIENCE
- 

2) En el contexto de la shell Bash, ¿cuál es el propósito principal del comando cut?  
Seleccione una:

- Pegar texto de diferentes archivos.
  - Comprimir archivos en un formato específico.
  - Cambiar permisos de archivos y directorios.
  - Extraer fragmentos de las líneas de un archivo.
- 

3) En Bash, ¿cuál es el propósito de la siguiente secuencia de comandos usando pipes?

```
grep line archivo.txt | sort
```

Seleccione una:

- Contar cuántas veces aparece repetida la palabra "line" en el archivo "archivo.txt" y almacenar el resultado en la variable "sort".
  - Filtrar las líneas que contienen la palabra "line" en "archivo.txt" y ordenarlas alfabéticamente.
  - Ordenar las líneas de "archivo.txt" en orden alfabético.
  - Concatenar el contenido de los archivos "line" y "archivo.txt" y luego ordenar todas sus líneas alfabéticamente.
- 

4) Indicar cuál es el resultado de la siguiente cadena de comandos:

```
grep DATA fichero.txt | awk '{print $3}' | sort | uniq -c
```

Seleccione una:

- Envía los ficheros DATA y fichero.txt a la salida estándar, imprime la tercera línea de cada fichero, las ordena y cuenta el número de caracteres de cada una.

- Cuenta las veces que se repite una determinada entrada en la tercera columna de aquellas líneas que contienen la palabra DATA en fichero.txt. El resultado está ordenado alfabéticamente por las entradas.
  - Cuenta las veces que se repite una determinada entrada en la tercera columna de aquellas líneas que contienen la palabra DATA en fichero.txt. El resultado está ordenado de mayor a menor número de repeticiones.
  - Copia fichero.txt dentro del directorio DATA y muestra el contenido de este directorio a tres columnas y ordenado únicamente por la fecha de creación.
  - Añade una línea con la palabra DATA al comienzo de fichero.txt y a continuación ordena el fichero alfabéticamente por la tercera columna y elimina las filas repetidas.
- 

5) En Bash, ¿cuál es el propósito de la siguiente secuencia de comandos usando pipes?

grep word archivo.txt | sort

Seleccione una:

- Ordenar las palabras de "archivo.txt" en orden alfabético.
  - Concatenar el contenido de los archivos "word" y "archivo.txt" y luego ordenar todo alfabéticamente.
  - Filtrar las líneas que contienen la palabra "word" en "archivo.txt" y ordenarlas alfabéticamente.
  - Contar cuántas veces aparece la palabra "word" en el archivo "archivo.txt" y mostrar los resultados ordenados alfabéticamente.
- 

6) ¿Cuál es el resultado de la siguiente operación?

```
x <- c(1, 2, 3, 4)
y <- c(2, 3, 4, 5)
print(x*y)
```

- [2, 6, 12, 20]
  - 40
  - Error
  - Ninguna de las opciones.
- 

7) ¿Cual es el output de este código de Python?

```
def f(a=[]):  
    a.append(1)  
    return a
```

```
a = f()  
a = f()  
  
print(sum(a))
```

- Ninguna de las anteriores
  - 1
  - 0
  - 2
- 

8) Indicar para qué sirven cada uno de los siguientes comandos de shell

- tail -> Muestra las últimas líneas de un fichero de texto
  - cut -> Extrae secciones específicas de cada línea de un archivo o de la entrada estándar
  - pwd -> Muestra el directorio de trabajo
  - sed -> Edita automáticamente un fichero de texto o la entrada estándar
  - wc -> Cuenta las líneas, palabras y caracteres de un fichero de texto
- 

9) Imagina un dataset llamado golf que contiene observaciones diarias sobre las actividades realizadas en un club de golf en 2015 (fecha, número de jugadores, condiciones meteorológicas, uso del restaurante, etc.). Si quisiéramos saber en qué momento del año 2015 ha habido una mayor afluencia de clientes utilizando R, podríamos hacer lo siguiente:

- uso <- arrange(golf, desc(number.players))
  - uso <- arrange(golf, number.players)
  - uso <- filter(golf, golf > number players)
- 

10) Imagina un dataset llamado golf que contiene observaciones diarias sobre las actividades realizadas en un club de golf (fecha, número de jugadores, condiciones meteorológicas, uso del restaurante, etc.). Si quisiéramos analizar en R las actividades realizadas durante los días de verano (Junio-Julio-Agosto) en los que ha habido más de 500 jugadores, comenzaríamos por:

- verano <- arrange(golf, month == 6:8, players > 500)
  - verano <- filter(golf, month == 6:8, players > 500)
  - verano <- filter(golf, month != 6:8, players > 500)
- 

11) El siguiente código tiene algunas ambigüedades, ya que al acceder a los elementos (indexing y slicing) se obtienen resultados diferentes.

```
>>> import pandas
>>> s = pandas.Series(["a", "b", "c", "d", "e", "f", "g", "h", "i", "i", "k", "l", "m"],
    index = [1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25])
>>> s
1    a
3    b
5    c
7    d
9    e
11   f
13   g
15   h
17   i
19   j
21   k
23   l
25   m
dtype: object

>>> s[3] # Indexing
'b'

>>> s[3:7] # Slicing
7    d
9    e
11   f
13   g
dtype: object
```

¿Qué está sucediendo? ¿Cómo se puede solucionar/prevenir este problema?

El problema ocurre porque s[3] usa el índice (label), mientras que s[3:7] usa las posiciones (posición 3 a 7).

Una solución es usar .loc[] para acceso por etiqueta y .iloc[] para acceso por posición.

---

12) En R, tenemos un dataset del tipo tibble llamado coches, que contiene precios de distintos modelos de coches en función de sus acabados (basic, standard, premium, excellence). ¿Qué haríamos si quisieramos crear una nueva variable que recoja el precio promedio de cada modelo?

- Cualquiera de las otras dos opciones
  - `coches$precio.medio = mean(coches$precio.basic, mean(coches$precio.basic, coches$precio.standard, coches$precio.excellence))`
  - `coches$precio.medio = apply(cbind(coches$precio.basic, coches$precio.standard, coches$precio.premium, coches$precio.excellence), 1, "mean")`
- 

13) De acuerdo con la Ley 19/2013, ¿a qué tienen derecho las personas en España?

- A ejercer sus derechos relativos a la protección de los datos de carácter personal (derechos de acceso, rectificación, supresión, limitación del tratamiento, portabilidad y de oposición).
  - **Conocer cómo se toman las decisiones que les afectan**
  - A la reutilización y el uso de documentos que obran en poder de las Administraciones y organismos del sector público con fines comerciales o no comerciales.
- 

14) En Bash, ¿cuál es el propósito del siguiente comando find?

```
find /ruta/del/directorio -type f -mtime -3 - size +500kl
```

Seleccione una:

- Buscar y listar todas las carpetas con un tamaño mayor de 500 kilobytes en el directorio especificado.
  - **Buscar y listar todos los archivos con un tamaño mayor de 500 kilobytes que fueron modificados en los últimos 3 días en el directorio especificado.**
  - Buscar y listar todas las carpetas con un tamaño mayor de 500 kilobytes que fueron modificadas en los últimos 3 días en el directorio especificado.
  - Buscar y listar todas las carpetas con un tamaño mayor de 500 kilobytes que fueron modificadas en los últimos 3 minutos en el directorio especificado.
-

15) Se quiere obtener la media de los elementos del siguiente array de numpy. Elige la forma más eficiente de llevarlo a cabo.

Seleccione una:

- import numpy

```
# Generate an array of random integers  
array = numpy.random.randint(100, size=10000000)  
array.mean()
```

- import numpy  
import statistics

```
# Generate an array of random integers  
array = numpy.random.randint(100, size=10000000)  
def mean (a):  
    return sum(a) / len(a)  
mean (array)
```

- import numpy  
import statistics

```
# Generate an array of random integers  
array = numpy.random.randint(100, size=10000000)  
statistics.mean(array)
```

- import numpy  
import statistics

```
# Generate an array of random integers  
array = numpy. random. randint(100, size=10000000)  
statistics.mean(array)
```

---

16) Dada la matriz “m”:

```
m = [,1] [,2] [,3]  
[1,] 1 2 1  
[2,] 4 3 6  
[3,] 7 8 9
```

¿Qué sentencia devolverá el vector con los números 2 y 3?

- m[,2]
  - m[2,3]
  - m[1:2,2]
- 

17) Para la lista alumnos, definida como:

```
alumnos <- data.frame(  
  fisica = c("juan", "rodrigo", "maria", "jose"),  
  matematicas = c("joaquin", "maialen", "jose", NA),  
  biologia = c("ana", "daniel", "markel", "adriana")  
)
```

¿Cuál de los siguientes resultados produciría la instrucción str(alumnos)?

- 'data.frame': 4x3 char matrix: ...
  - '**data.frame'**: 4 obs. of 3 variables: ...
  - 'data.frame': 3 variables: ...
- 

18) ¿Qué implica el Data Act (Reglamento UE 2023/2854) a nivel europeo?

- Regula el acceso y uso compartido de datos generados por dispositivos conectados (IoT).
  - Incentiva el uso de metadatos FAIR y esquemas comunes.
  - Obliga a formatos interoperables, legibles por máquina y accesibles vía API.
- 

19) ¿Cuál de los siguientes repositorios NO está especializado en datos científicos?

- [data.europa.eu](http://data.europa.eu)
- Zenodo
- [datos.gbif.es](http://datos.gbif.es)

## REPASO EXAMEN

### - OPEN DATA

**Ley 19/2013**, de 9 de diciembre, de **transparencia**, acceso a la información pública y buen gobierno. Obligaciones para publicación de información pública. **Conocer cómo se toman las decisiones que les afectan.**

**Ley 37/2007**, de 16 de noviembre, sobre **reutilización** de la información del sector público, que fue adaptada por la Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para incluir los cambios normativos de la Unión Europea (Directiva 2013/37/UE).

Necesidad de **procesar y publicar la información** con unas condiciones que faciliten su reutilización: formatos estructurados, abiertos e interoperables, que garanticen la seguridad, así como la propiedad intelectual e industrial.

### **Data Governance Act (Reglamento UE 2022/868)**

- Establece un marco de **confianza para el intercambio de datos**.
- Promueve la **intermediación de datos** (organismos neutrales que facilitan la compartición segura).
- Introduce el concepto de **altruismo de datos** (ciudadanos y empresas que comparten datos voluntariamente para fines de interés general).
- Impulsa la creación de **espacios europeos de datos sectoriales** (salud, movilidad, energía, agricultura...).

### **Data Act (Reglamento UE 2023/2854, aplicable 2025)**

- Regula el acceso y uso compartido de **datos generados por dispositivos conectados (IoT)**.
- Garantiza que los usuarios (personas y empresas) puedan reutilizar los datos de los productos que poseen.
- Obliga a formatos **interoperables, legibles por máquina y accesibles vía API**.