

	Colegio Universitario de Cartago	
	BIG DATA (BD)	

**I Cuatrimestre 2025**  
**Lenguaje de Minería de Datos**  
**BD-123**

---

**Nombre Docente: Osvaldo González Chaves**

**Fecha: 14-02-2025**

**Valor de la prueba: 60 puntos**

---

**INSTRUCCIONES:**

- 1.- Favor de leer con cuidado toda la prueba. Para lo anterior, cuenta con 10 minutos.
  - 2.- Proceda luego a contestar cada pregunta, en el orden en que aparece en el presente examen. Asegúrese de fundamentar sus respuestas y de contestar únicamente lo que se le pide.
- 

**ESTIMADO ESTUDIANTE: PARA SU MEJOR DESARROLLO DE LA PRESENTE PRUEBA, ASEGURESE DE LEER CON CUIDADO CADA PREGUNTA, RAZONE SUS RESPUESTAS Y FUNDAMENTE LAS MISMAS.**

- Deberá entregar un archivo con el contenido de las respuestas de la primera parte etiquetado de la siguiente manera: <NombreEstudiante-preguntas\_examen>.docx
  - Deberá entregar un script en R etiquetado con su nombre conteniendo las respuestas a los ejercicios etiquetado de la siguiente manera: <NombreEstudiante-Script>.r
  - Sea claro, ordenado y conciso en sus respuestas

### Primera Parte: (20 puntos)

Lea las siguientes preguntas y responda de acuerdo con el enunciado:

1. Indique 4 recomendaciones para escribir código en R y muestre un ejemplo de cada uno(4pts)
2. Mencione los tipos de fuentes de datos y de 2 ejemplos para cada uno (3 pts.)
3. ¿Cuál es la diferencia entre el clustering jerárquico y el particional? Justifique su respuesta (2 pts.)
4. Explique en que consiste la maximización de distancias inter-cluster(2 pts.)
5. Explique que es cluster contiguo (nearest neighbour) y de dos ejemplos (3 pts.)
6. ¿Qué es la apofenia y en que podría afectar en un análisis de clustering? (2 pts.)
- 7.Cuál es la diferencia entre un vector y una matriz (2 pts.)
8. Que es un vector de tipo factor (2 pt)

### Segunda Parte (40 pts.) 2 pts. c/u

En RStudio realizar lo siguiente:

#### Vectores

1. Declare un vector de notas de un estudiante, nombrado `notasEstudiante` con los siguientes valores:  
90.5, 80.0, 75.0, 65.5, 75.5, 82.0, 99.5, 89.0, 83.6, 76.5
2. Declare un vector con los siguientes valores y nombrarlo `vectorMaterias`:  
"Matematica", "Espanol", "Ingles", "Frances", "Folosofia", "Tecnologia", "Fisica",  
"Biologia", "Sociales", "Quimica"
3. Asigne los nombres `vectorMaterias` al vector `notasEstudiante` e imprima el resultado.
4. Muestre el tamaño del vector e imprima el resultado.
5. Obtenga el promedio de las notas del estudiante en variable `promedioNotas` e imprima el resultado.
6. Muestre la nota de la materia "Ingles" del estudiante.
7. Obtenga la nota mínima y máxima del vector, declare dos variables e imprimirla el resultado.

## Matrices

8. Construya una matriz en R, empleando la función `matrix` con el parámetro `nrow`, que construya por columnas la matriz de la siguiente manera:

$$\begin{pmatrix} 12 & 45 & 78 \\ 90 & 34 & 56 \end{pmatrix}$$

9. Sume la matriz anterior con la función `sum ()` e imprima el resultado con la función `cat`.  
10. Obtenga el promedio de la matriz del punto 1 e imprima el resultado con la función `cat`.  
11. Sume las columnas de la matriz del punto 1 e imprima el resultado con la función `cat`.  
12. Sume las filas de la matriz del punto 1 e imprima el resultado con la función `cat`.  
13. Añada la suma de las columnas a la matriz del punto 1 e imprima el resultado con la función `cat`.

Cree la siguiente matriz del medallero de juegos nacionales 2023:

```
datos_juegosNacionales<-  
matrix(c(23,12,13,48,12,2,1,15,10,7,11,28,8,7,9,24,8,3,4,15,7,7,5,19,7,7,2,16,5,4,6,15),byr  
ow= TRUE ,nrow=8)
```

Cree los siguientes vectores:

```
cuadroMedallero <- c("Oro", "Plata", "Bronce", "Total")
```

```
comiteCantonal <- c("San José", "Escazú", "Alajuela", "Santa  
Cruz", "Oreamuno", "Limon", "Grecia", "San Ramón")
```

14. Asigne nombres a las columnas de la matrix `datos_juegosNacionales`, use el vector `cuadroMedallero`  
15. Asigne nombres a las filas de la matrix `datos_juegosNacionales`, use el vector `comiteCantonal`  
16. Muestre los datos del cantón Oreamuno  
17. Muestre el cantón que tiene menos medallas de plata e imprima el resultado. (Puede mostrar la fila, la posición de la matriz o el nombre del cantón)

## Factores

18. Cree un factor con el nivel de géneros musicales:  
("rock", "pop", "jazz", "salsa", "merengue", "pop ", " rock", "regueton", "jazz ", "salsa")  
19. Cree el factor ordenado con el nivel de género musical.  
20. Muestre el resultado o summary del factor.