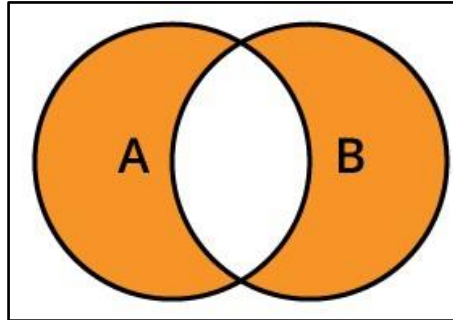


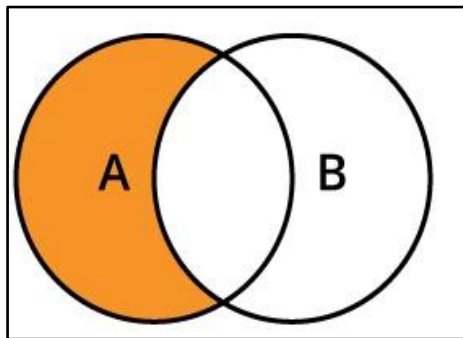
Datathon BanBif

Cuestionario

1. Escriba la sentencia SQL para obtener la información sombreada:



2. Escriba la sentencia SQL para obtener la información sombreada:



3. Escriba la sentencia SQL que muestre el nombre del ejecutivo con más ventas de cada región.

Región	Ejecutivo	Ventas
1	Jorge	1'000
1	Alicia	1'200
1	Miguel	900
2	Nicolás	1'500
2	Rubén	1'600
3	Fernando	800
3	Karina	1'000

4. Dada una matriz cuadrada, calcule el valor absoluto de la diferencia de la suma de las diagonales utilizando pseudocódigo, código R o código Python.

El formato de las variables de entrada constaría de 3 líneas: la primera indica el número de filas y columnas de la matriz.

Las siguientes n-filas indican los elementos de la n-ésima fila.

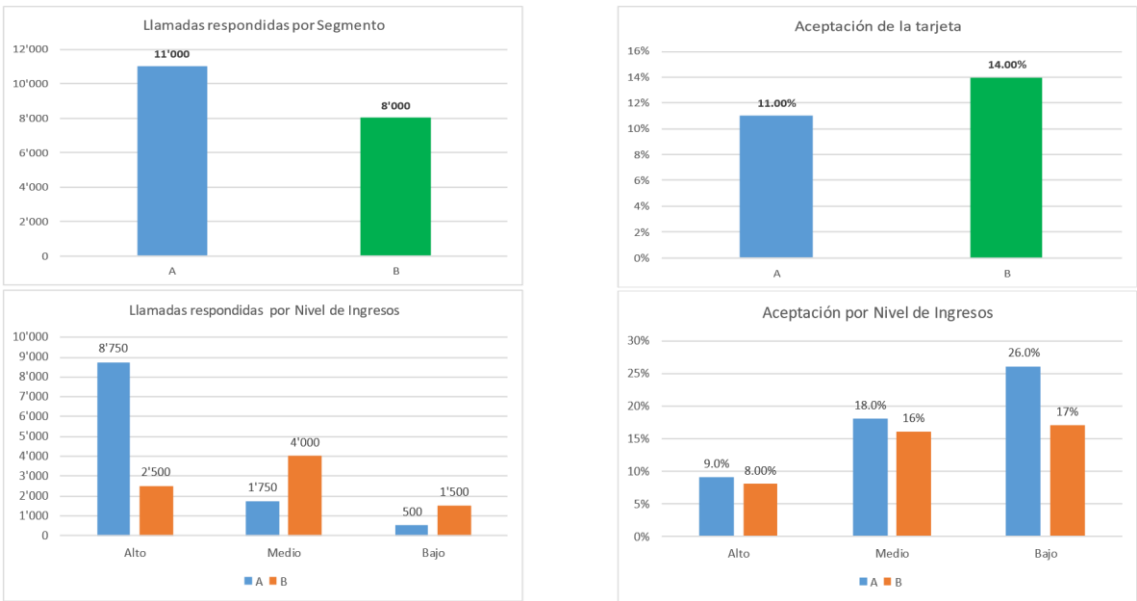
Ejemplo de entrada:

4
11 2 4 6
3 5 9 10
14 8 7 4
3 0 10 8

Ejemplo de salida:

5

5. El algoritmo XGBoost sólo trabaja con datos numéricos. ¿Qué tratamiento hace a sus variables categóricas antes de emplearlo?
6. A continuación, se muestran los resultados de una campaña telefónica dirigida a 2 segmentos de clientes diferentes. ¿Explique para Ud. cuál de las 2 campañas fue más exitosa? ¿por qué?



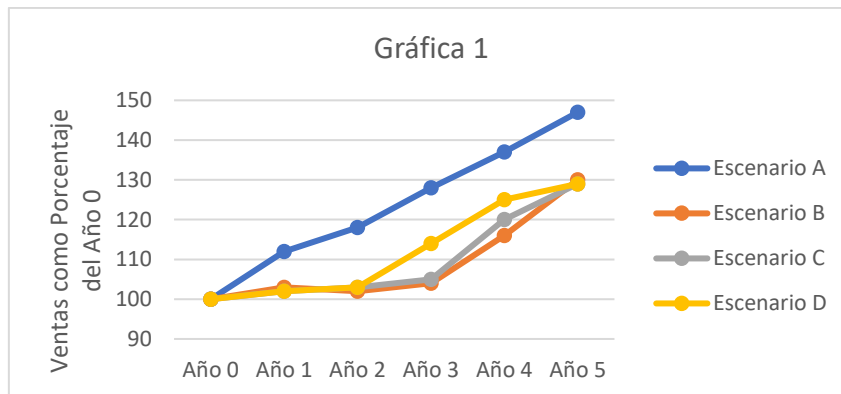
7. 'Firstmilk' es una empresa productora de leche y sus derivados. Sus clientes son supermercados principalmente. La *Tabla 1* muestra los ingresos generados en el último año y el crecimiento promedio obtenido en los últimos 5 años para las categorías de productos.

Tabla 1

	Ingresos presente año	Crecimiento promedio (5 últimos años)
Leche evaporada	S/.370 MM	4.20%

Leche en polvo	S/.65 MM	1.50%
Yogurt	S/.50 MM	1.20%
Quesos	S/.12 MM	-7%

‘Firstmilk’ ha sido adquirida recientemente por ‘Dairy Products Co.’ con el objetivo de impulsar las ventas. El gerente de ‘Dairy Products Co.’ ha trazado como meta que al cabo de 5 años quiere ver un crecimiento de por lo menos el 10% en las ventas de los 2 años últimos. En la *Gráfica 1* se muestran 4 posibles escenarios de crecimiento para ‘Firstmilk’ con el año 0 como el año actual de partida.



- ¿cuál de los escenarios cumple los requerimientos de ‘Dairy Products Co.’? Desarrolle su respuesta.
- ¿cuál es el nivel de ventas que tenía la categoría ‘Quesos’ hace 5 años? Desarrolle su respuesta.
- ¿cuál es el valor estimado aproximado de las ventas en el Año 3 bajo el escenario C? Desarrolle su respuesta.

El gerente de ‘Dairy Products Co.’ quiere probar con un piloto en una ciudad. El piloto consiste en reducir el precio de venta al consumidor final en 6% acompañando la campaña con una estrategia de marketing en diversos medios. El gerente de ventas de ‘First Milk’ le da la siguiente información:

- La campaña de marketing tendrá un costo de S/.2.1 millones.
 - ‘First Milk’ tiene un mercado de 1 millón de compradores quienes compran una caja de leche evaporada por mes.
 - El precio de venta a los supermercados es de S/.10.
 - El precio de venta al consumidor final es de S/.11.
 - ‘First Milk’ obtiene un 20% de margen de utilidad en el producto de leche evaporada.
 - La campaña no deberá afectar el margen de ganancia que tienen los supermercados actualmente.
- ¿cuál es el monto de utilidad promedio por caja de leche evaporada que obtiene ‘First Milk’ antes de la campaña? Detalle su respuesta.

'Dairy Products Co.' requiere que toda inversión se pague al primer año de implementada.

- e) ¿En qué porcentaje se deben incrementar las ventas para cumplir con el requerimiento? Detalle su respuesta.
8. Explique la disyuntiva entre sesgo y varianza al evaluar los resultados de un modelo predictivo y por qué se da.
9. La gerencia de producto le solicita realizar un modelo predictivo para retención de sus clientes ya que necesita identificar qué clientes son los más propensos a dejar el servicio. Explique qué ventajas encuentra en utilizar una regresión logística en lugar de bosques aleatorios para dicha tarea.
10. Ud. ha elaborado un modelo de clasificación y ha conseguido un alto valor de AUC con un algoritmo de machine learning. Cuando le piden la probabilidad de aceptación de un grupo de clientes encuentra que las probabilidades que obtiene con el modelo difieren de las probabilidades reales. ¿Qué pasos y técnicas emplearía para calibrar las probabilidades obtenidas?
11. Los algoritmos de boosting siguen un procedimiento secuencial donde los errores de una clasificación previa sirven de input a la siguiente. Explique cómo es posible que algoritmos de boosting como el XGBoost sean "paralelizables".
12. Calcule y explique el número de capas y el número de neuronas que usaría para resolver el siguiente problema de clasificación binaria, utilizando un ANN (artificial neural network) del tipo MP (multiple perceptron).

