

Nombre y apellido: [REDACTED]

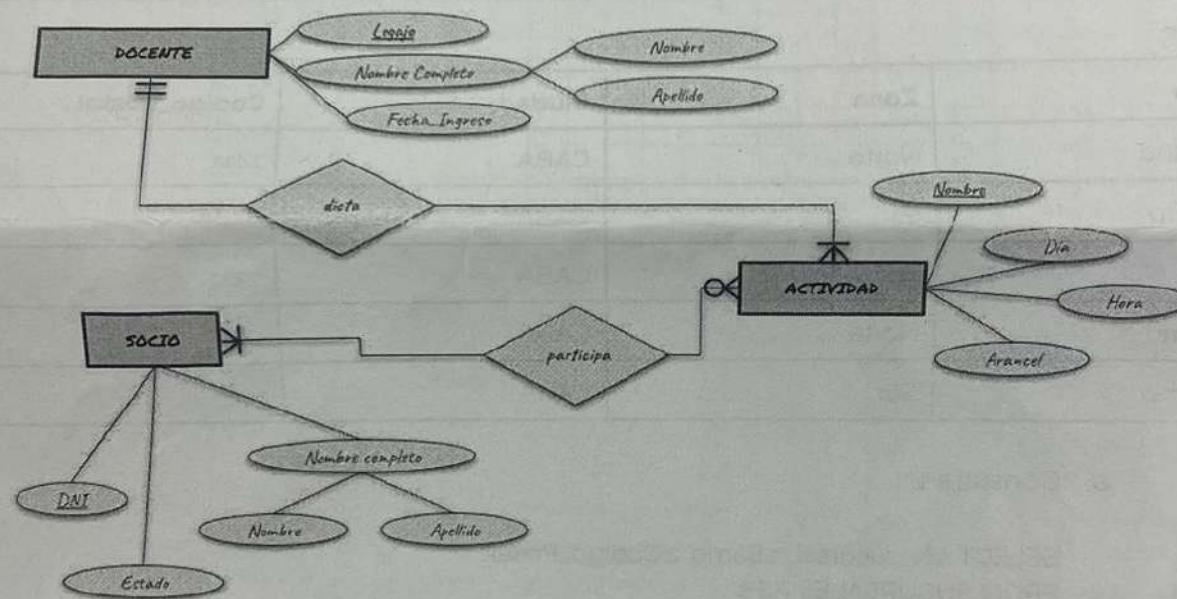
LU: [REDACTED]

Grilla de puntuajes. No completar.

Ej 1	Ej 2	Ej 3	Ej 4	Ej 5	Ej 6	Ej 7	Ej 8	Total	Condición
15	12	14	0	10	14	2	0	67	APROB.

**Aclaraciones:** El parcial NO es a libro abierto. Para aprobar se requieren al menos 60 puntos. Cualquier decisión de interpretación que se tome debe ser aclarada y justificada. Todas las respuestas tienen que estar justificadas de manera concisa. Agregue nombre, apellido, LU y nro. de hoja (empezando a numerar en las hojas de respuesta) en el extremo superior izquierdo de cada hoja.

- (15 p) Dado el siguiente DER mapearlo al modelo relacional. No olvide indicar en todos los casos nombre de esquema, sus atributos, clave primaria y foreign keys (las FK con flechas). En el caso de existir más de una opción para implementar una relación, elegir aquella que evite los Nulls a través del diseño.



- (15 p) Dado el siguiente esquema, correspondiente a un catálogo de libros, decir si está en 2FN y/o en 3FN (justificar). En caso de no estarlo, proponer una descomposición que se encuentre en 3FN y preserve las dependencias funcionales. Marcar las claves primarias (PK) y las dependencias funcionales en los esquemas surgidos por la descomposición.

#### Esquema

LIBRO(Título, Nombre\_Autor, Apellido\_Autor, Nacionalidad\_Autor, Editorial, Idioma, Número\_Edición, Año\_Edición, Precio)

#### Dependencias Funcionales

DF1: [Título] → [Nombre\_Autor, Apellido\_Autor, Idioma]

DF2: {Número\_Edición} → {Año\_Edición}

DF3: {Título, Editorial, Número\_Edición} → {Precio}

DF4: {Nombre\_Autor, Apellido\_Autor} → {Nacionalidad\_Autor}

3. (15 p) Dadas las tablas SUCURSALES y ZONAS, con el contenido que se muestra a continuación, si se ejecutan las siguientes consultas SQL ¿qué se obtiene como resultado? Escribir la tabla resultante con su contenido, es decir tanto filas como columnas.

SUCURSALES

N_sucursal	Barrio	Cant_Cajeros	Cant_Cajas
123	Palermo	2	3
52	Villa Crespo	4	3
107	Belgrano	3	2
141	Palermo	3	4

ZONAS

Barrio	Zona	Ciudad	Codigo_Postal
Belgrano	Norte	CABA	1411
Caballito	Sur	CABA	1406
Flores	Sur	CABA	1321
Palermo	Norte	CABA	1418
Palermo	Sur	CABA	1425

a. Consulta 1

```
SELECT s.N_sucursal, s.Barrio, z.Codigo_Postal  
FROM SUCURSALES AS s  
LEFT OUTER JOIN ZONAS AS z  
ON s.Barrio = z.Barrio  
ORDER BY s.N_Sucursal DESC;
```

b. Consulta 2

```
SELECT s.Barrio, COUNT(*) AS TotalSucursales, SUM(s.CantCajeros) AS TotalCajeros  
FROM SUCURSALES AS s  
INNER JOIN ZONAS AS z  
ON s.Barrio = z.Barrio  
GROUP BY s.Barrio  
HAVING TotalCajeros > 3  
ORDER BY s.Barrio ASC;
```

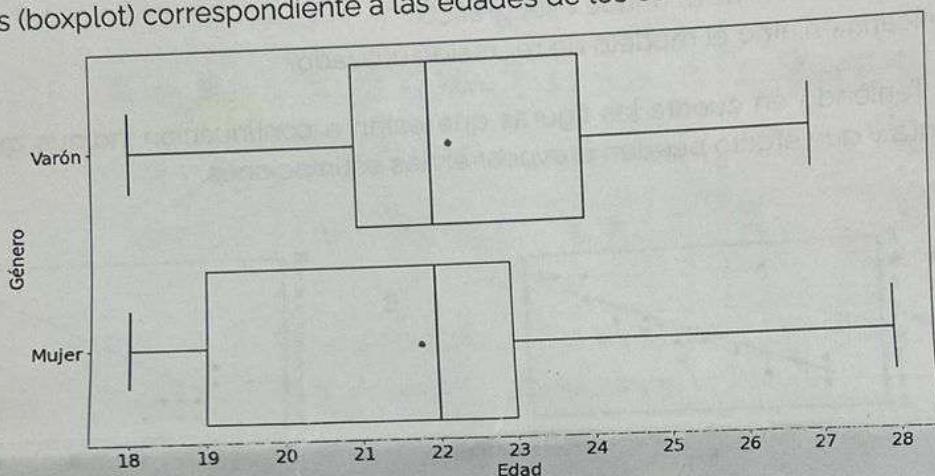


Nombre y apellido: [REDACTED]

LU: [REDACTED]

4. (10 p) Respecto del proyecto de actualización de la ley de protección de datos personales, indicar:
- Qué significa el principio de neutralidad tecnológica.
  - Dar un ejemplo de un tipo de dato nuevo considerado dentro de los datos personales, que no estuviera contemplado en la ley previa.
5. (10 p) Se entrevistó a 100 mujeres y 100 varones estudiantes de la Licenciatura en Ciencia de Datos de la FCEN. En base a la información obtenida, se construyó el siguiente gráfico de cajas (boxplot) correspondiente a las edades de los estudiantes.

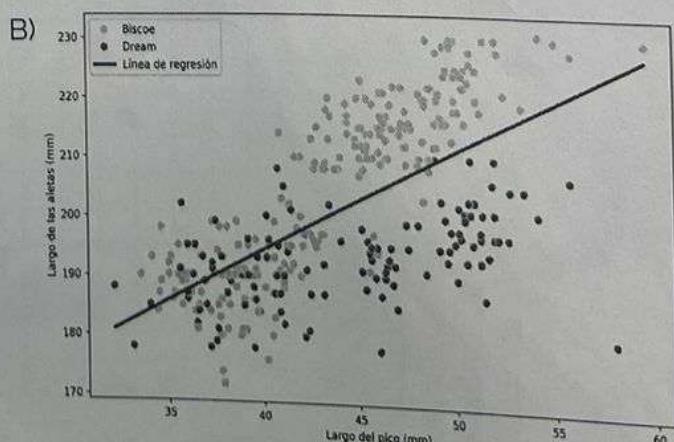
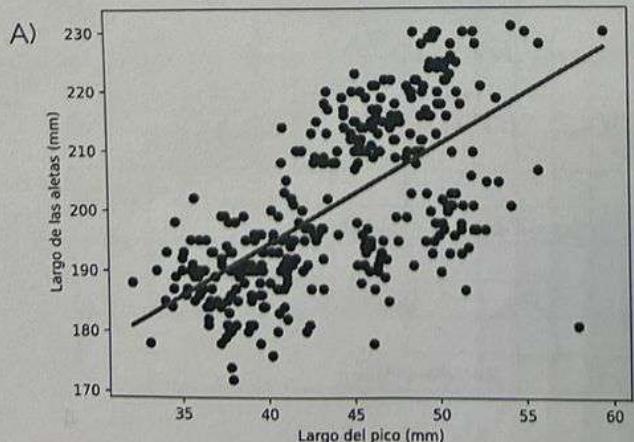
10/10



Responda las siguientes preguntas de manera concisa y justificando sus respuestas:

- ¿Qué medida de tendencia central tiene el mismo valor en varones y mujeres?
- ¿Qué género presenta mayor rango de edades? ¿Y mayor rango intercuartílico (IQR)?
- ¿Cuántas de las mujeres entrevistadas tienen 19 años o menos?
- ¿Es cierto que hay más mujeres entrevistadas de 19 años o menos que varones?

6. (20 p) El conjunto de datos denominado 'Palmer Penguins' es un dataset que contiene información morfológica de diferentes representantes de tres especies de pingüinos que habitan en las islas cercanas a la *Palmer Station* (Biscoe y Dream), ubicada en la península antártica. A cada individuo, al capturarlo, se le tomaron medidas del largo y ancho del pico, largo de las aletas, sexo, especie, entre otras. Con el fin de estudiar aspectos alométricos (como cambia el tamaño de una parte del cuerpo en relación a otra parte), se realizó una regresión cuyo resultado se muestra en la parte A:

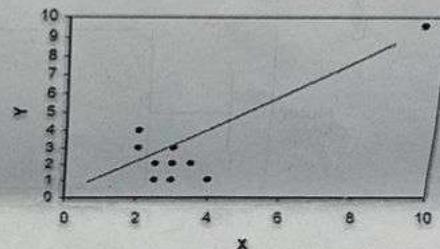
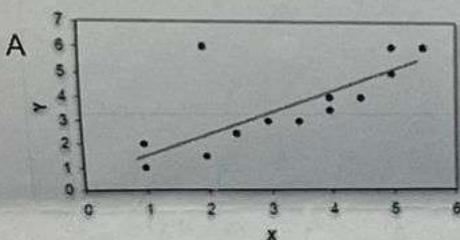


Responda las siguientes preguntas de manera concisa y justificando sus respuestas:

- A. ¿Cuál es la variable dependiente? ¿Y la independiente?
- B. ¿Cuál es el algoritmo de regresión que se utilizó para estudiar esta relación?
- C. ¿Considera que se puede hablar de causalidad en la relación?
- D. ¿Cómo se distribuyen los puntos en el espacio?
- E. Teniendo en cuenta que se obtuvo un valor de  $R^2$  de 0.43. Explique qué significa, qué información brinda.
- F. El valor de la pendiente fue de 1.69 y la ordenada al origen de 126.68. Explique estos valores y qué significan.

Luego, se incorporó información valiosa respecto de la isla de origen de donde se habían tomado las muestras obteniéndose el gráfico de la parte B):

- G. ¿Qué información nueva brinda este gráfico?
  - H. ¿Modificaría en algo el modelo de regresión utilizado?
7. (5 p) Teniendo en cuenta las figuras que están a continuación indique qué problema presenta y qué efecto pueden provocar en las estimaciones.



8. (10 p) Decidir V o F y justificar de manera concisa en ambos casos.
- A. El algoritmo K-means puede manejar clusters de forma no esférica de manera eficiente.
  - B. El valor de precisión siempre es mayor que el valor de recall en un modelo que tiene muchos falsos negativos.
  - C. En el algoritmo DBSCAN, el parámetro 'minPts' define el número mínimo de puntos para formar un cluster.
  - D. El recall mide la capacidad de un modelo para detectar instancias positivas.