

Banjercito

Diplomado en Ciencia de Datos

## Guía de Estudio

### Evaluación Final

Alan Badillo Salas

Diciembre 2024

## Introducción

El Diplomado en Ciencia de Datos ha llegado a su etapa final, dentro del diplomado hemos aprendido a lo largo de 5 módulos, el uso de las herramientas más importantes usadas en el campo de la ciencia de datos. En el primer módulo hemos introducido el análisis financiero a través de Excel, Power BI y Python, en el segundo módulo hemos reunido los conceptos más importantes de probabilidad y estadística, en el tercer módulo hemos trabajado más a detalle con las librerías de Numpy y Pandas y el flujo de la ciencia de datos, en el cuarto módulo hemos aprendido las tres piezas fundamentales del Machine Learning, que son la Regresión, la Clasificación y la Clusterización, así como la simulación estadística y probabilística, finalmente, en el quinto módulo hemos generalizado todos los modelos, para que una Red Neuronal pueda hacer los pronósticos generalizados a través del Deep Learning.

En esta guía repasaremos los fundamentos de cada módulo, a través de preguntas y reactivos similares a los de la evaluación final del diplomado, cubriendo los conceptos vistos en cada módulo y detallando su solución. Comprender esta guía debería ser suficiente para tener un buen desempeño como Científico de Datos y estar preparado para problemas del mercado actual y aplicaciones reales en la industria y el gobierno.

## Módulo I — Introducción a Python con Finanzas

### Problema 1 — Uso de Excel

Una empresa financiera posee una tabla con los montos de créditos que ha aprobado a sus clientes. La tabla posee el género de cada cliente y se desea calcular el monto total de los créditos aprobados para hombres y mujeres.

En la Figura 1 se muestra la tabla con el cliente, género, crédito y monto aprobado.

	A	B	C	D
1	Cliente	Género	Crédito	Monto
2	C001	Hombre	P0001	\$ 14,600.00
3	C002	Mujer	P0002	\$ 18,900.00
4	C003	Hombre	E0004	\$ 12,000.00
5	C004	Mujer	E0032	\$ 6,900.00
6	C005	Mujer	F0005	\$ 3,200.00
7	C006	Hombre	G0128	\$ 23,700.00

Figura 1: Créditos aprobados a clientes por género

(a) Calcule la suma de los montos para hombres y la suma de montos para mujeres usando la fórmula `=SUMA.SI(RANGO CRITERIO, CRITERIO, RANGO SUMA)`.

(b) Calcule la suma de los montos para hombres y la suma de montos para mujeres usando la fórmula `=SUMA(FILTRAR(RANGO SUMA, RANGO INCLUIR))`.

Para el inciso (a) utilizamos las fórmulas

=SUMAR.SI(B2:B7,"MUJER",D2:D7)

Figura 2: Solución al Problema 1 inciso (a)

=SUMA (FILTRAR (D2:D7, B2:B7="MUJER"))

Figura 3: Solución al Problema 1 inciso (b)

=FILTRAR(A2:D7,B2:B7="HOMBRE")

=FILTRAR(A2:D7,B2:B7="HOMBRE")

Se generará la tabla virtual con todos los créditos correspondientes a las filas de los hombres o de las mujeres.

G8 <span>fx</span>   =FILTRAR(A2:D7,B2:B7=F3)										
	A	B	C	D	E	F	G	H	I	J
1	Cliente	Género	Crédito	Monto			USANDO CONTAR.SI Y SUMAR.SI			
2	C001	Hombre	P0001	\$ 14,600.00			Total	Suma Monto		
3	C002	Mujer	P0002	\$ 18,900.00		Hombre	3	\$ 50,300.00		
4	C003	Hombre	E0004	\$ 12,000.00		Mujer	3	\$ 29,000.00		
5	C004	Mujer	E0032	\$ 6,900.00						
6	C005	Mujer	F0005	\$ 3,200.00						
7	C006	Hombre	G0128	\$ 23,700.00			USANDO FILTRAR			
8						Hombres	C001	Hombre	P0001	\$14,600.00
9							C003	Hombre	E0004	\$12,000.00
10							C006	Hombre	G0128	\$23,700.00
11										\$50,300.00
12										

<b>fx</b>	<b>=FILTRAR(A2:D7,B2:B7=F3)</b>									
-----------	---------------------------------	--	--	--	--	--	--	--	--	--

Hombres	C001	Hombre	P0001	\$14,600.00
	C003	Hombre	E0004	\$12,000.00
	C006	Hombre	G0128	\$23,700.00
				\$50,300.00

Figura 4: Tabla virtual generada por FILTRAR (c)

## Problema 2 — Uso de Tablas dinámicas en Excel

Una empresa financiera posee una tabla con los montos de créditos que ha aprobado a sus clientes y su deuda actual. La tabla posee el género de cada cliente y se desea calcular la deuda total y deuda promedio.

En la Figura 5 se muestra la tabla con el cliente, género, crédito, monto aprobado y deuda actual.









	A	B	C	D	E
1	Cliente	Género	Crédito	Monto	Deuda
2	C001	Hombre	P0001	 \$ 14,600.00	\$ 3,200.00
3	C001	Hombre	H0014	 \$ 16,300.00	\$ 2,500.00
4	C002	Mujer	P0002	 \$ 18,900.00	\$ 1,600.00
5	C003	Hombre	E0004	 \$ 12,000.00	\$ 4,800.00
6	C004	Mujer	E0032	 \$ 6,900.00	\$ 5,300.00
7	C005	Mujer	F0005	 \$ 3,200.00	\$ 2,100.00
8	C006	Hombre	G0128	 \$ 15,000.00	\$ 1,900.00
9	C005	Mujer	F0095	 \$ 12,500.00	\$ 3,500.00

Figura 5: Créditos aprobados con deuda actual a clientes por género

- Crea una tabla dinámica de A1:E9.
- Segmenta las filas por Género y luego por Cliente.
- Cuenta el número de Créditos en valores.
- Calcula la suma del Monto en valores.
- Calcula la suma de la Deuda en valores.
- Calcula el promedio del Monto en valores.
- Calcula el promedio de la Deuda en valores.
- ¿Quién es el cliente de mayor deuda?
- ¿Quién es el cliente al que se le ha prestado más en promedio?
- ¿Quién es el cliente hombre con mayor deuda promedio?
- ¿Es significativamente distinta la deuda entre hombres y mujeres?
- ¿Cuánto se ha recuperado en total de todos los créditos?

## Solución al Problema 2

Para el inciso (a) creamos la tabla dinámica para el rango A1:E9 y la insertamos en una celda de la misma hoja.

The screenshot shows the Excel interface with the 'Insertar' tab selected. A dynamic table is visible in the background, spanning rows 1 to 9 and columns A to E. The table has the following data:

Cliente	Género	Crédito	Monto	Deuda
C001	Hombre	P0001	\$ 14,600.00	\$ 3,200.00
C001	Hombre	H0014	\$ 16,300.00	\$ 2,500.00
C002	Mujer	P0002	\$ 18,900.00	\$ 1,600.00
C003	Hombre	E0004	\$ 12,000.00	\$ 4,800.00
C004	Mujer	E0032	\$ 6,900.00	\$ 5,300.00
C005	Mujer	F0005	\$ 3,200.00	\$ 2,100.00
C006	Hombre	G0128	\$ 15,000.00	\$ 1,900.00
C005	Mujer	F0095	\$ 12,500.00	\$ 3,500.00

The 'Crear tabla dinámica' dialog box is open, showing the following configuration:

- Selecciona los datos que desees analizar:**
  - ☒ Seleccionar una tabla o rango
  - Tabla o rango: Hoja1!\$A\$1:\$E\$9
  - ☐ Usar un origen de datos externo
  - Elegir conexión... No se recuperó ningún dato.
- Selecciona dónde deseas ubicar la tabla dinámica.**
  - ☐ Hoja de cálculo nueva
  - ☒ Hoja de cálculo existente
  - Tabla o rango: Hoja1!\$A\$11

Buttons: Cancelar, Aceptar

Figura 6: Solución al Problema 2 inciso (a)

Para el inciso (b) arrastramos los campos del género y cliente a las filas

Campos de tabla dinámica

Formato de forma

NOMBRE DE CAMPO

Buscar en campo:

☒ Cliente
 ☒ Género
 ☒ Crédito
 ☒ Monto
 ☒ Deuda

Filtros

Columnas

: Valores

Filas

: Género

: Cliente

Valores

: Cuenta de Crédito

: Suma de Monto

: Suma de Deuda

: Promedio de Monto

: Promedio de Deuda

Figura 7: Solución al Problema 2 inciso (b)

Para el inciso (c), (d), (e), (f), (g) arrastramos los campos como valores y configuramos cada campo si es recuento, suma o promedio

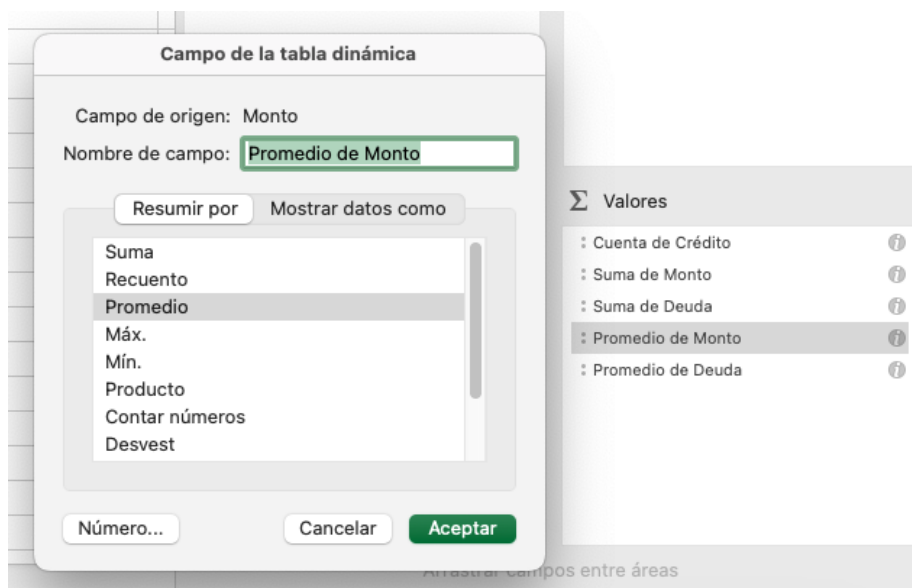


Figura 8: Solución al Problema 2 incisos (c), (d), (e), (f), (g)

Para el inciso (h), (i), (j), (k), (l) observamos los valores en la tabla dinámica

Etiquetas de fila	Cuenta de Crédito	Suma de Monto	Suma de Deuda	Promedio de Monto	Promedio de Deuda
<b>Hombre</b>	<b>4</b>	<b>\$ 57,900.00</b>	<b>\$ 12,400.00</b>	<b>\$ 14,475.00</b>	<b>\$ 3,100.00</b>
C001	2	\$ 30,900.00	\$ 5,700.00	\$ 15,450.00	\$ 2,850.00
C003	1	\$ 12,000.00	\$ 4,800.00	\$ 12,000.00	\$ 4,800.00
C006	1	\$ 15,000.00	\$ 1,900.00	\$ 15,000.00	\$ 1,900.00
<b>Mujer</b>	<b>4</b>	<b>\$ 41,500.00</b>	<b>\$ 12,500.00</b>	<b>\$ 10,375.00</b>	<b>\$ 3,125.00</b>
C002	1	\$ 18,900.00	\$ 1,600.00	\$ 18,900.00	\$ 1,600.00
C004	1	\$ 6,900.00	\$ 5,300.00	\$ 6,900.00	\$ 5,300.00
C005	2	\$ 15,700.00	\$ 5,600.00	\$ 7,850.00	\$ 2,800.00
<b>Total general</b>	<b>8</b>	<b>\$ 99,400.00</b>	<b>\$ 24,900.00</b>	<b>\$ 12,425.00</b>	<b>\$ 3,112.50</b>

Figura 9: Solución al Problema 2 incisos (h), (i), (j), (k), (l)

Respuestas: (h) C001 \$5,700, (i) C002 \$18,900, (j) C003 \$4,800, (k) No, la diferencia es de \$100 menor al 1% de diferencia, (l)  $\$74,500 = \$99,400 - \$24,900$ .



### Problema 3 — Uso de Power BI

Una empresa financiera posee una tabla con los préstamos que ha aprobado a sus clientes indicando el monto, el estado y el valor. En cada préstamo se reporta el estado de sobre lo que se ha pagado del préstamo, lo que se debió haber cubierto actualmente (deuda) y lo que se debería cubrir en el futuro (pendiente). La financiera requiere una tabla que resuma de cada cliente sus préstamos y en cada préstamo se desglose la suma de cuánto adeuda, cuánto ha pagado y cuánto está pendiente, así como el total. También requiere un indicador que muestre la diferencia entre la suma del monto del préstamo y el valor del monto pagado, para saber cuánto se adeuda en total del préstamo, igualmente desglosado por cliente y préstamo. Finalmente la financiera requiere dos reportes visuales, un anillo que muestre el estado de los préstamos (pagado, pendiente y deuda) segmentado por cliente. Y un esquema jerárquico que permita desglosar el valor del préstamo por su estado, cliente y número de préstamo.

En la Figura 10 se muestra la tabla de préstamos, con la columna del cliente, el número de préstamo, el monto, el estado (pagado, deuda, pendiente) y el valor asociado al estado.

	ABC 123 Cliente	ABC 123 Préstamo	\$ Monto	ABC 123 Estado	\$ Valor	+
1	C001	1	15000	Pagado	7000	
2	C001	1	15000	Deuda	2000	
3	C001	1	15000	Pendiente	6000	
4	C001	2	12000	Pagado	2000	
5	C001	2	12000	Deuda	0	
6	C001	2	12000	Pendiente	10000	
7	C002	1	18000	Pagado	8000	
8	C002	1	18000	Deuda	4000	
9	C002	1	18000	Pendiente	6000	
10	C002	2	9000	Pagado	3000	
11	C002	2	9000	Deuda	6000	
12	C002	2	9000	Pendiente	0	
13	C003	1	23000	Pagado	12000	
14	C003	1	23000	Deuda	5000	
15	C003	1	23000	Pendiente	6000	
+						

Figura 10: Préstamos aprobados a clientes desglosado por cliente, número de préstamo, estado (pagado, deuda y pendiente) y valor asociado al estado

(a) Crea una matriz que reporte las filas para el cliente y número de préstamo en ese orden, el estado como columna y la suma del valor en los valores.

(b) En el modelo de datos crea una nueva medida con la fórmula

**Diferencia = SUM(P103[Monto])-SUM(P103[Valor])**

para reportar la diferencia entre la suma del monto menos la suma del valor asociado al estado. Donde *P103* es el nombre de la tabla creada al inicio.

(c) Crea una matriz que reporte las filas para el cliente y número de préstamo en ese orden, el estado como columna y la suma del monto, suma del valor y suma de diferencia en los valores en ese orden. Reporta solo el estado Pagado en los filtros del objeto visual.

(d) Crea una gráfica de anillo cuya leyenda este dada por el estado del préstamo, en valores reporte la suma del valor asociado al estado, y en detalles segmente al cliente.

(e) Crea un esquema jerárquico para analizar la suma del valor valor asociado al estado, explicado por el estado del préstamo, el cliente y el número de préstamo en ese orden.

(f) De la matriz del inciso (a) responde si cada desglose corresponde al monto reportado en cada préstamo.

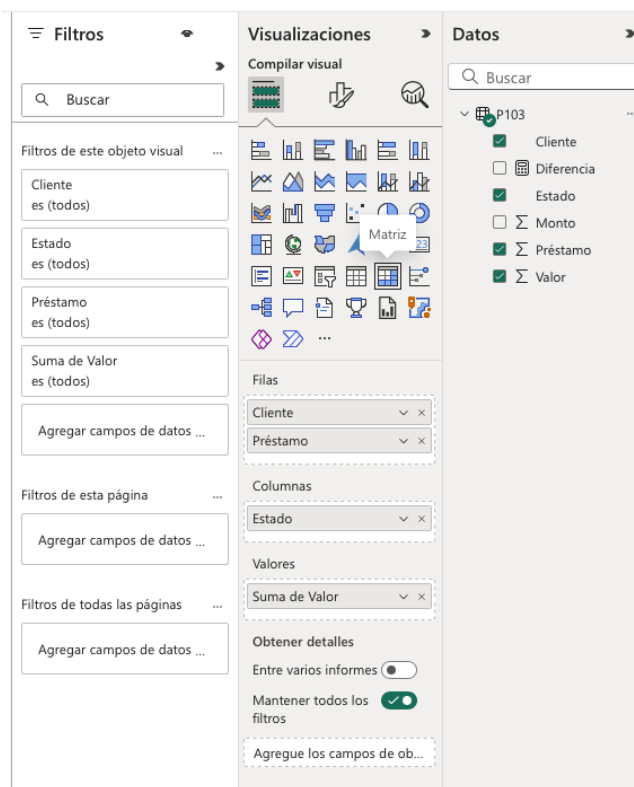
(g) De la matriz del inciso (b) responde cuál el es el préstamo con menor deuda total y usando el inciso (a) cómo es su deuda respecto al préstamo.

(h) Del anillo identifica al cliente de mayor deuda y el cliente con mayor valor pendiente.

(i) Del esquema jerárquico identifica al cliente de mayor deuda y el número de préstamo de mayor deuda para ese cliente.

### Solución al Problema 3

Para el inciso (a) creamos una matriz que reporte las filas para el cliente y número de préstamo en ese orden, el estado como columna y la suma del valor en los valores.



Cliente	Deuda	Pagado	Pendiente	Total
<b>C001</b>	<b>2000</b>	<b>9000</b>	<b>16000</b>	<b>27000</b>
1	2000	7000	6000	15000
2	0	2000	10000	12000
<b>C002</b>	<b>10000</b>	<b>11000</b>	<b>6000</b>	<b>27000</b>
1	4000	8000	6000	18000
2	6000	3000	0	9000
<b>C003</b>	<b>5000</b>	<b>12000</b>	<b>6000</b>	<b>23000</b>
1	5000	12000	6000	23000
<b>Total</b>	<b>17000</b>	<b>32000</b>	<b>28000</b>	<b>77000</b>

Figura 11: Solución al Problema 3 inciso (a)

Para el inciso (b) abrimos el modelo de datos y dentro del modelo de datos creamos una nueva medida con la fórmula para calcular la diferencia.

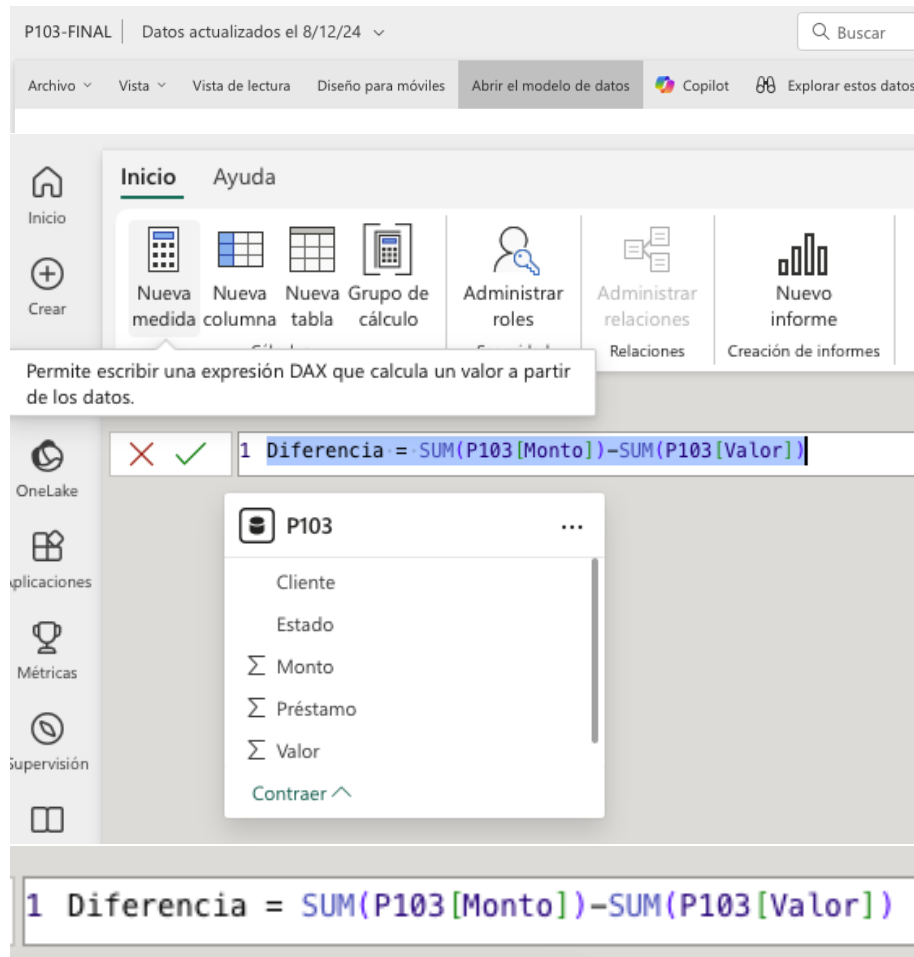


Figura 12: Solución al Problema 3 inciso (b)

Para el inciso (c) creamos una matriz que reporte las filas para el cliente y número de préstamo en ese orden, el estado como columna y la suma del monto, suma del valor y suma de diferencia en los valores en ese orden. Reporta solo el estado Pagado en los filtros del objeto visual.

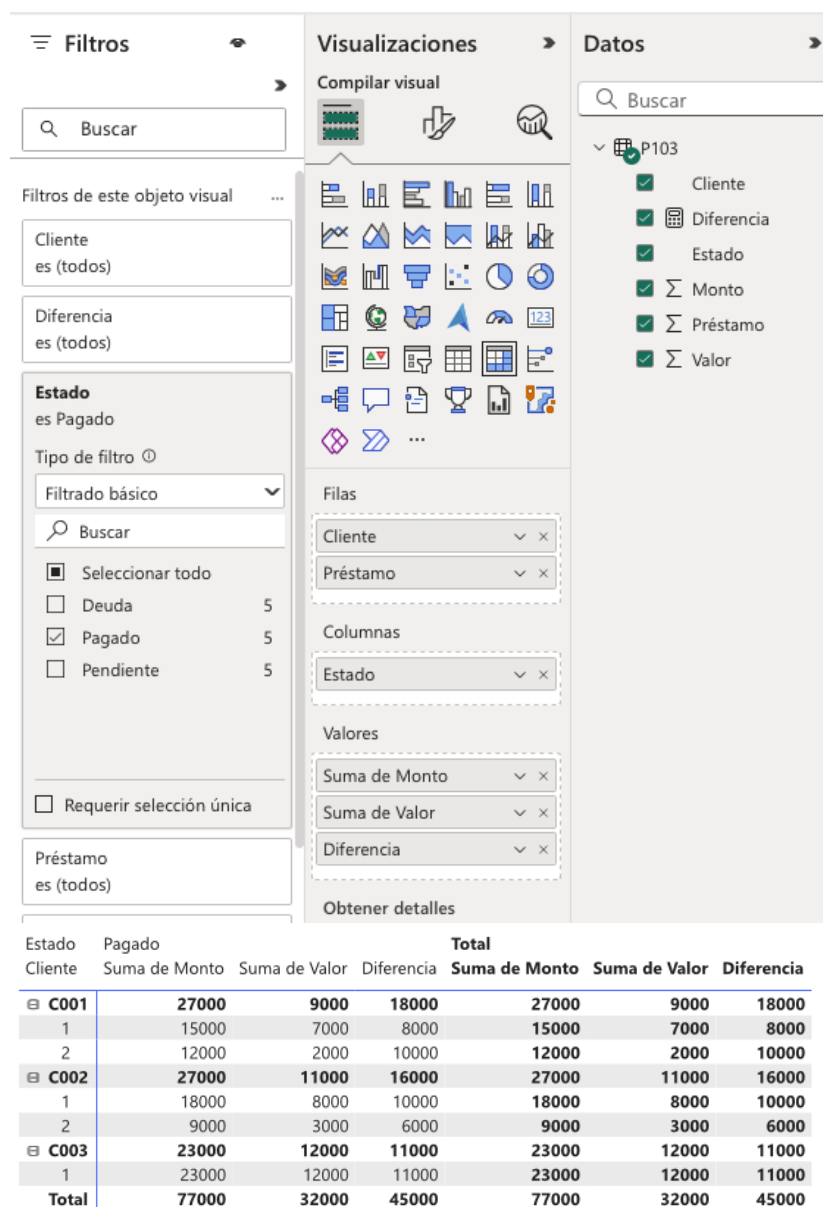
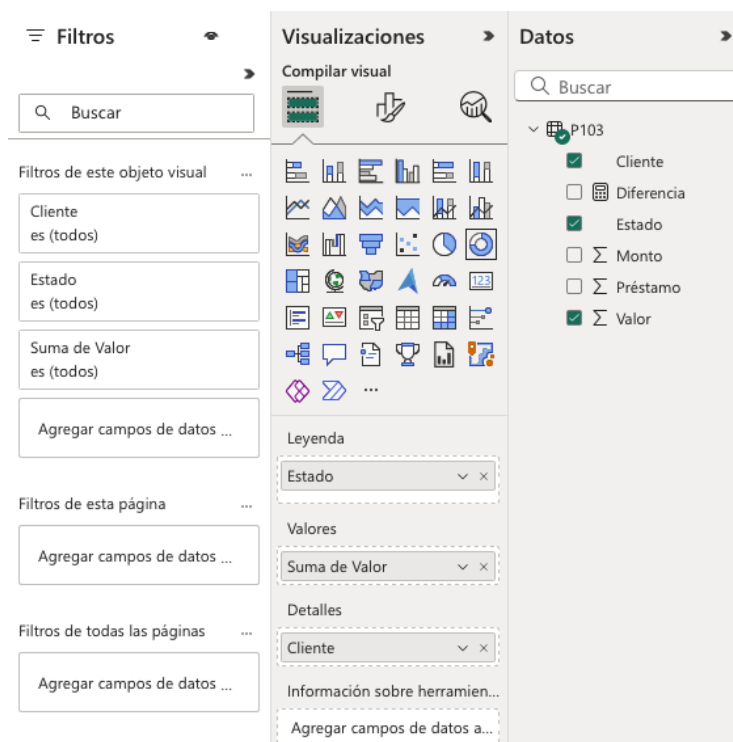


Figura 13: Solución al Problema 3 inciso (c)

Para el inciso (d) creamos una gráfica de anillo cuya leyenda este dada por el estado del préstamo, en valores reporte la suma del valor asociado al estado, y en detalles segmente al cliente.



Suma de Valor por Estado y Cliente

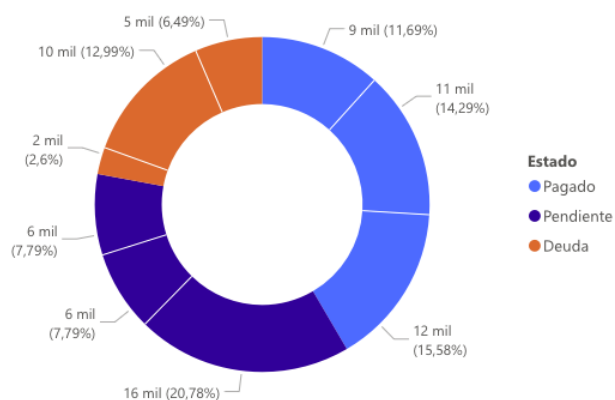


Figura 14: Solución al Problema 3 inciso (d)

Para el inciso (e) creamos un esquema jerárquico para analizar la suma del valor valor asociado al estado, explicado por el estado del préstamo, el cliente y el número de préstamo en ese orden.

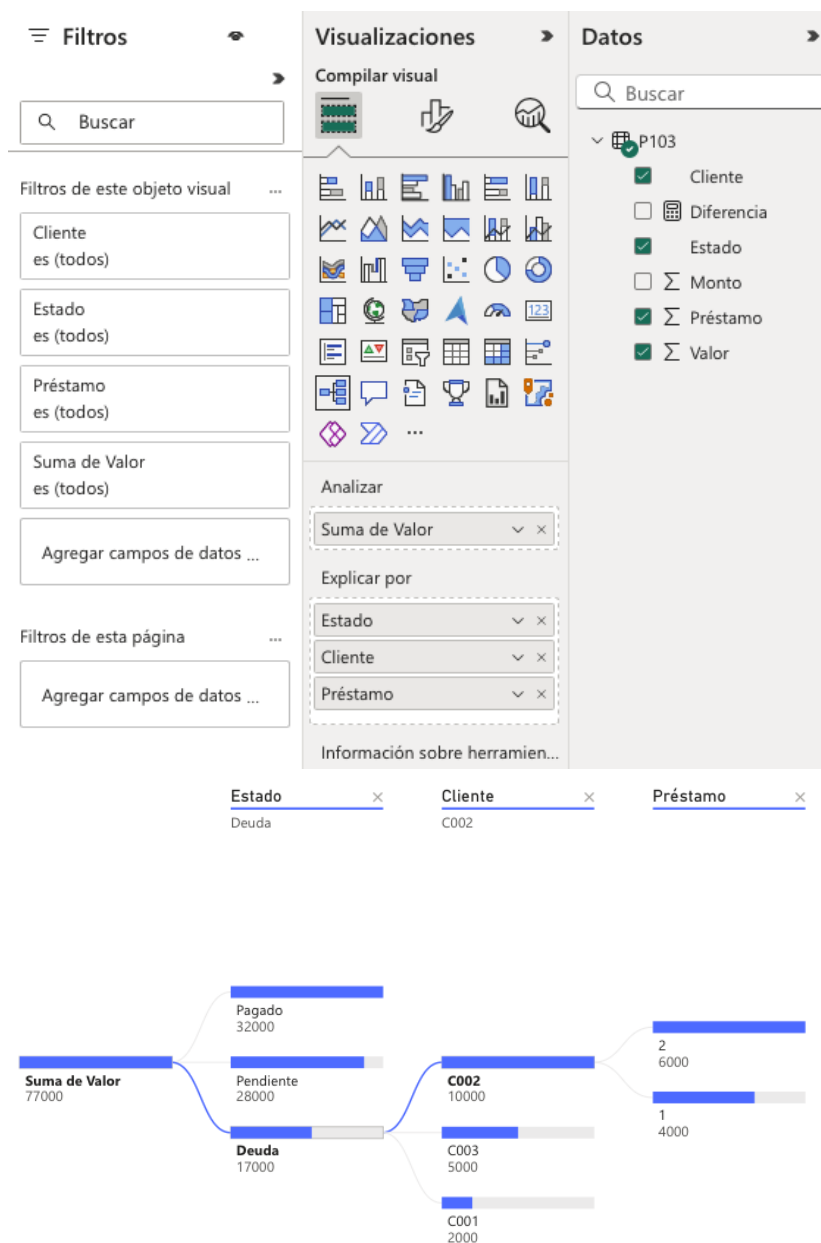


Figura 15: Solución al Problema 3 inciso (e)

Para los otros incisos las respuestas son: (f) Sí, la suma de montos en cada préstamo coincide con la suma de su deuda, lo pagado y lo pendiente. (g) Para el cliente *C002* en el préstamo número 2 la diferencia total es de **\$6,000** y es el préstamo de menor deuda, ya que lo pendiente para ese préstamo es de **\$0**, sin embargo, respecto a los otros préstamos es el de mayor deuda. (h) **C002 \$10,000 12.99%** es el cliente de mayor deuda y **C001 \$16,000 20.78%** es el cliente de mayor valor pendiente. (i) **C002 \$10,000 2 \$10,000** es el cliente de mayor deuda, con su préstamo número dos.



## Problema 4 — Uso de Python

Un banco ha decidido usar Python en un ambiente controlado para poder hacer operaciones y cálculos de riesgo, para ello ha instalado un ambiente virtual cerrado en computadoras sin acceso a internet para que no se filtren los conjuntos de datos utilizados, ni los reportes generados.

El banco comenzará por automatizar una serie de funciones útiles que pueda aplicar para cálculos simples, para de esta forma crear una metodología de trabajo cuando deseen incrementar el volumen de las operaciones.

Para ello, el banco requiere que se generen los siguientes reportes automatizados para saber que todo está funcionando correctamente.

Implementa las funciones que se detallan a continuación y resuelve los incisos siguientes. Completa el código que haga falta para que las funciones operen correctamente siguiendo las pistas de los comentarios

```
def reporteTitulo(codigo, leyenda):
    print(f"Banco Nacional - Reporte {codigo} / {leyenda}")

def reporteCorte(codigo, leyenda):
    print(f"----- {codigo} / {leyenda} -----")

def reporteFin(codigo, leyenda):
    print(f"----- {codigo} / {leyenda} -----")

def reporteBalanceDiaInicio(balance):
    reporteTitulo("BIN", "Balance al Inicio del Día")
    print(f"${balance:.2f}")
    reporteFin("BIN", "Balance al Inicio del Día")

def reporteBalanceDiaFin(balance):
    reporteTitulo("B-OUT", "Balance al Final del Día")
    print(f"${balance:.2f}")
    reporteFin("B-OUT", "Balance al Final del Día")

def reporteTotalOperacionesDia(numOperaciones):
    reporteTitulo("T-OP", "Total de Operaciones del Día")
    # Imprime el número de operaciones
    reporteFin("T-OP", "Total de Operaciones del Día")

def reporteTotalOperacionesRelativasMes(numOperacionesDia, numOperacionesMes):
    reporteTitulo("T-OR", "Tasa de Operaciones del día respecto al mes")
    tasa = 100 * numOperacionesDia / numOperacionesMes
    # Imprime el número de operaciones del día
    # luego una diagonal y el número de operaciones del mes
    # Haz un corte del reporte
    # Imprime la tasa de operaciones del día respecto al mes a un dígito
    # Ejemplo: 41/200 (20.5%)
    reporteFin("T-OR", "Tasa de Operaciones del día respecto al mes")
```

- (a) Reporta el balance de inicio del día con **\$ 185,000.32**
- (b) Reporta un total de **247** operaciones del día
- (c) Reporta un total de **247** de **192,316** operaciones relativas del mes
- (d) Reporta el balance de fin del día con **\$ 232,827.71**

### Problema 5 — Interés Simple

Una financiera posee montos de préstamos a proyectos y la tasa anual fija que le cobrará sus clientes. La financiera requiere calcular la suma de montos y proponer una misma tasa de interés anual única para todos los proyectos, así como encontrar la tasa anual en caso de perder el monto por algún proyecto.

En la Figura 16 se muestran los montos y la tasa anual de los montos de préstamo destinados a proyectos financiados.

	A	B
1	Monto	Tasa Anual
2	\$ 16,300.00	8.0%
3	\$ 15,200.00	13.0%
4	\$ 25,400.00	6.5%
5	\$ 184,900.00	8.3%
6	\$ 192,000.00	7.1%

Figura 16: Montos de préstamos a proyectos de una financiera y su tasa anual

- (a) Calcula el interés simple generado por la tasa anual y reporta cada interés y la suma de todos estos intereses.
- (b) Calcula la tasa anual promedio y de todas las tasas y la suma de todos los montos y calcula el interés de la suma total con la tasa anual promedio. ¿Este interés es el mismo que el de la suma del inciso (a)?
- (c) ¿Cuál debería ser la tasa anual de todos los montos para que la suma del interés sea la misma?
- (d) ¿Cuál debería ser la tasa anual de todos los montos, si no existiera el monto de \$ 25,400.00 para que la suma del interés sea la misma a las anteriores?

### Problema 6 — Interés Compuesto

Una financiera desea calcular el interés compuesto que generaría un monto de \$ **4,312,939.83** con una tasa anual de **8%**, si se tuviera que pagar mensualmente a una tasa compuesta de **0.67%** y si tuviera que pagar a tasa compuesta de **0.34%**.

- (a) Calcula el interés compuesto a 12 meses con una tasa mensual de **0.67%**.
- (b) Calcula el interés compuesto a 24 quincenas con una tasa quincenal de **0.34%**

### Problema 7 — Tasa de Interés Variable

Una financiera desea calcular el interés que generaría un monto de **\$ 2,617,800.01** con una tasa mensual variable que comienza en **6%** el primer mes y se actualiza a **5%** para el segundo mes, luego **4%** al tercer mes, **3%** para el cuarto mes y luego **2%** y **1%** para el quinto y sexto mes, luego de ahí se aplica el **0.5%** en los meses restantes.

- (a) Calcula el interés compuesto para los 12 meses usando el interés variable.
- (b) Si se usara un interés anual fijo, ¿De cuánto sería?

### Problema 8 — Tabla de Amortización Simple

Una financiera desea calcular una tabla de amortización para un monto financiado de **\$ 6,500.00**, con un enganche inicial de **\$ 2,300.00**, el cual pagará los intereses amortizados. Los intereses están diseñados para que bajen linealmente durante 48 semanas y recuperen **80%** del monto financiado. Es decir, se necesitan recuperar intereses por el **80%** de los **\$ 6,500.00**, diseñando intereses que se reduzcan linealmente en 48 pagos y luego se considere el enganche inicial como descuento inmediato a los primeros intereses.

- (a) Calcula la tabla de amortización sin considerar el enganche.
- (b) ¿Cuántos pagos iniciales cubre el enganche?
- (c) Determina un abono fijo para terminar de pagar los intereses más el monto financiado en 48 semanas.

### Problema 9 — Interés Simple Futuro

Un banco financió un proyecto durante 36 meses por **\$ 8,496,500.00**, con una tasa anual fija y simple del **12%** y desea calcular los intereses que se generarán al mes 25 y al mes 30. También desea saber en qué mes los intereses superarán el valor de los **\$ 40,000.00**.

- (a) Calcula los intereses para el mes 25 y 30.
- (b) Determina en qué mes los intereses superarán los **\$ 40,000.00**.
- (c) Calcula el valor de las cuotas fijas para cubrir el préstamo en los 36 meses.
- (d) Calcula el valor de las cuotas fijas para cubrir el préstamo en los 36 meses, si se da un adelanto del **10%** del valor del proyecto (los intereses deberían ser los mismos).

### Problema 10 — Punto de Liquidación

Un banco financió un proyecto durante 48 meses por \$ **6,232,800.00**, con una tasa anual fija y simple del **12%** y desea calcular el valor de liquidación que considere pagar los intereses futuros y la deuda total, la liquidación se llevará a cabo en el mes 28.

- (a) Calcula los intereses totales.
- (b) Calcula los intereses faltantes a partir del mes 28.
- (c) Calcula la deuda de liquidación para el mes 28.

## Módulo II — Estadística y Probabilidad con Python

### Problema 1 — Probabilidad de un Evento Simple

Una financiera posee una tabla con el género de sus clientes, si tiene trabajo, si está casado y si tiene alguna deuda.

En la Figura 17 se muestra la tabla con el género, trabajo, casado y deuda de los clientes.

	A	B	C	D
1	Género	Trabajo	Casado	Deuda
2	HOMBRE	SI	NO	SI
3	MUJER	NO	SI	NO
4	MUJER	SI	SI	NO
5	HOMBRE	NO	SI	SI
6	MUJER	SI	NO	SI
7	HOMBRE	SI	SI	NO
8	HOMBRE	NO	NO	SI
9	MUJER	SI	SI	SI
10	HOMBRE	SI	SI	NO

Figura 17: Tabla de deuda por género, trabajo y casado

- (a) Calcula la probabilidad de ser hombre
- (b) Calcula la probabilidad de ser mujer
- (c) Calcula la probabilidad de tener trabajo
- (d) Calcula la probabilidad de no tener trabajo
- (e) Calcula la probabilidad de estar casado
- (f) Calcula la probabilidad de no estar casado
- (g) Calcula la probabilidad de tener deuda
- (h) Calcula la probabilidad de no tener deuda



## Problema 2 — Tabla conjunta de Dos Eventos

Un restaurante está inspeccionando la bebidas asociadas al pago con tarjeta de sus clientes, porque sospecha que los clientes que pagan con tarjeta de crédito consumen más cerveza que aquellos que pagan con tarjeta de débito.

En la Figura 18 se muestra la tabla con los datos de los clientes si utilizaron tarjeta de crédito o débito y el tipo de bebida que consumieron.

	A	B	C
1	Cliente	Tarjeta	Bebida
2	C001	Débito	Refresco
3	C002	Crédito	Cerveza
4	C003	Débito	Refresco
5	C004	Crédito	Cerveza
6	C005	Crédito	Refresco
7	C006	Débito	Cerveza
8	C007	Débito	Refresco
9	C008	Débito	Refresco
10	C009	Crédito	Cerveza
11	C010	Débito	Refresco
12	C011	Crédito	Cerveza
13	C012	Crédito	Refresco
14	C013	Débito	Refresco
15	C014	Débito	Cerveza
16	C015	Débito	Cerveza

Figura 18: Tabla de tipo bebida consumida asociada al tipo de tarjeta

- (a) Genera la tabla conjunta donde las filas sean el tipo de tarjeta (Débito, Crédito) y las columnas el tipo de bebida (Refresco, Cerveza). En cada celda pon el número total de registros que cumplen la fila y la columna, por ejemplo, para la fila Débito y la columna Refresco, hay 6 registros.
- (b) Agrega una columna adicional con los totales marginales para la fila, sumando todas las columnas.
- (c) Agrega una fila adicional para los totales marginales para la columna, sumando todas las filas.
- (d) Indica el total de registros en la intersección de la columna marginal y la fila marginal.
- (e) Genera la tabla de probabilidad conjunta, que es la misma que la tabla conjunta, solo que cada valor está dividido sobre el total de registros.
- (f) Calcula la probabilidad de tener una tarjeta de débito
- (g) Calcula la probabilidad de tener una tarjeta de crédito
- (h) Calcula la probabilidad de haber consumido refresco
- (i) Calcula la probabilidad de tener consumido cerveza

### Problema 3 — Probabilidad Conjunta de Dos Eventos

Una empresa ha recuperado la tabla conjunta del personal que fue promovido a un mejor puesto y los que no a lo largo de 3 años. La tabla divide los hombres y mujeres contra si fueron o no promovidos, también cuenta con sus totales marginales. La empresa necesita encontrar la probabilidad conjunta de que los hombres hayan sido promovidos o no.

En la Figura 19 se muestra la tabla conjunta de promoción del personal por género.

	A	B	C	D
1	Personal	Promovido	No Promovido	
2	Hombre	172	85	257
3	Mujer	140	67	207
4		312	152	464

Figura 19: Tabla conjunta de promoción del personal por género

- (a) Divide cada valor de la tabla conjunta por el total marginal de 464.
- (b) Indica cual es la probabilidad total de que un hombre haya sido promovido.
- (c) Indica cual es la probabilidad total de que un hombre no haya sido promovido.
- (d) Indica cual es la probabilidad total de que una mujer haya sido promovida.
- (e) Indica cual es la probabilidad total de que una mujer no haya sido promovida.
- (f) Indica cual es la probabilidad total de que alguien sea hombre.
- (g) Indica cual es la probabilidad total de que alguien sea mujer.
- (h) Indica cual es la probabilidad total de que alguien sea promovido.
- (i) Indica cual es la probabilidad total de que alguien no sea promovido.

## Problema 4 — Probabilidad Condicional de Dos Eventos

Una fábrica de cosméticos está diseñando un nuevo programa para que sus empleados trabajen dos horas menos de lo convencional. Para determinar si su modelo funciona, ha contratado empleados temporales, a los cuales se les pregunta si quieren ganar \$ 2,800 por 6 horas diarias durante los 6 días de jornada de la semana, o ganar \$ 4,080 por 8 horas. Al cabo de la semana, el empleado temporal tiene la oportunidad de tener un contrato de 3 meses o renunciar.

Después de construir la tabla conjunta con el número de empleados que trabajaron 6 horas u 8 horas y que renunciaron o no renunciaron, la empresa decide generar la tabla de probabilidad conjunta para no exponer los números netos sobre los empleados que contrató. Ahora requiere calcular la probabilidad de haber trabajado 6 horas dado que renunció y haber trabajado 8 horas dado que renunció, para poder entender cuál era turno que eligió una persona que decidió renunciar al final.

Además la empresa necesita saber cuál es la probabilidad de renunciar, dado que se trabaja 6 horas y la probabilidad de renunciar dado que se trabaja 8 horas, para entender quién tiene mayor probabilidad de renunciar y decidir si vale la pena mantener un turno de 8 horas en el futuro.

En la Figura 20 se muestra la tabla de probabilidad de los empleados que deciden renunciar o no en turnos de 6 y 8 horas.

	A	B	C	D
1		6 horas	8 horas	
2	Renuncia	0.15	0.34	0.49
3	No Renuncia	0.24	0.27	0.51
4		0.39	0.61	1

Figura 20: Tabla de probabilidad de renuncia por turno

- (a) Indica cual es la probabilidad que alguien renuncie  $P(Renuncia)$
- (b) Indica la probabilidad de que alguien trabaje 6 horas  $P(6h)$
- (c) Indica la probabilidad de que alguien trabaje 8 horas  $P(8h)$
- (d) Indica la probabilidad de que alguien trabaje 6 horas y renuncie  $P(6h \cap Renuncia)$
- (e) Indica la probabilidad de que alguien trabaje 8 horas y renuncie  $P(8h \cap Renuncia)$
- (f) Calcula la probabilidad de trabajar 6 horas dado que se renuncia  

$$P(6h|Renuncia) = \frac{P(6h \cap Renuncia)}{P(Renuncia)}.$$
- (g) Calcula la probabilidad de trabajar 8 horas dado que se renuncia  

$$P(8h|Renuncia) = \frac{P(8h \cap Renuncia)}{P(Renuncia)}.$$
- (h) Calcula la probabilidad de renunciar dado que se trabajó 6 horas  

$$P(Renuncia|6h) = \frac{P(6h \cap Renuncia)}{P(6h)}.$$
- (i) Calcula la probabilidad de renunciar dado que se trabajó 8 horas  

$$P(Renuncia|8h) = \frac{P(8h \cap Renuncia)}{P(8h)}.$$

Problema 5 — Probabilidad Conjunta de Tres Eventos  
Problema 6 — Probabilidad Condicional de Tres Eventos  
Problema 7 — Estadísticos Principales de Un Eje  
Problema 8 — Estadísticos Principales de Dos Ejes  
Problema 9 — Matriz de Correlación entre Tres Ejes  
Problema 10 — Prueba F entre Dos Hipótesis de Un Eje

## Módulo III — Ciencia de Datos

### Problema 1 — Estructuración de un Dataset en Excel

Crea un archivo Excel que represente la siguiente información:

País	Año	Población (millones)	PIB per cápita (USD)
México	2022	126	10000
Estados Unidos	2022	331	63000
Canadá	2022	38	48000
Brasil	2022	213	7500
Japón	2022	126	40000

Table 1: Datos para estructurar

Guarda el archivo con el nombre `indicadores.xlsx`.

## Problema 2 — Dataset de indicadores en CSV

Convierte el archivo `indicadores.xlsx` del Problema 1 en un archivo CSV llamado `indicadores.csv`. Usa el siguiente formato de columnas:

```
País,Año,Población,PIB_per_capita
México,2022,126,10000
Estados Unidos,2022,331,63000
Canadá,2022,38,48000
Brasil,2022,213,7500
Japón,2022,126,40000
```



### Problema 3 — Adquisición de un CSV en Python

Escribe un script en Python que lea el archivo `indicadores.csv` del Problema 2 y lo muestre en la consola. Completa el siguiente código:

```
import pandas as pd

# Ruta del archivo
file_path = "...".

# Cargar el CSV
df = pd.read_csv(file_path)

# Mostrar el DataFrame
print(...)
```

## Problema 4 — Indicadores en Series de Pandas

Usando el DataFrame cargado en el Problema 3, crea una Serie con los valores del PIB per cápita. Completa el siguiente código:

```
# Crear la Serie
pib_per_capita = ...

# Mostrar la Serie
print(pib_per_capita)
```

## Problema 5 — Indicadores en DataFrames de Pandas

Agrega una nueva columna llamada `PIB_total` que calcule el PIB total como el producto de `Población` y `PIB per cápita`. Completa el código:

```
# Agregar la columna
df['PIB_total'] = ...

# Mostrar el DataFrame
print(df)
```

## Problema 6 — Filtrar columnas en DataFrames

Filtra únicamente las columnas `País` y `PIB_total`. Completa el código:

```
# Filtrar columnas
filtered_df = df[...]

# Mostrar las columnas filtradas
print(filtered_df)
```

## Problema 7 — Filtrar filas en DataFrames

Filtra las filas donde el PIB total sea mayor a 1 billón de USD (**1e12**). Completa el código:

```
# Filtrar filas
high_pib = df[...]  
  
# Mostrar las filas filtradas  
print(high_pib)
```

## Problema 8 — Agrupación de datos en DataFrames

Agrupar los datos por año y calcula el promedio de PIB per cápita. Completa el código:

```
# Agrupar y calcular promedio
average_pib = df.groupby(...)[...].mean()

# Mostrar el resultado
print(average_pib)
```

## Problema 9 — Combinación de datos en DataFrames

Combina el DataFrame actual con otro que contenga información adicional, como índice de desarrollo humano (IDH). Completa el código:

```
# Crear otro DataFrame
idh_data = {
    "País": ["Mexico", "Estados Unidos", "Canada", "Brasil", "Japon"],
    "IDH": [0.758, 0.926, 0.929, 0.765, 0.919]
}
idh_df = pd.DataFrame(idh_data)

# Combinar DataFrames
combined_df = pd.merge(df, idh_df, on=...)

# Mostrar el resultado
print(combined_df)
```

## Problema 10 — Visualización de datos

Crea un gráfico de barras que muestre el PIB total por país usando la librería `matplotlib`. Completa el código:

```
import matplotlib.pyplot as plt

# Crear grafico de barras
plt.bar(..., ...)

# Configurar etiquetas
plt.xlabel("País")
plt.ylabel("PIB Total (USD)")
plt.title("PIB Total por País")

# Mostrar grafico
plt.show()
```

## Módulo IV — Machine Learning

### Problema 1 — Aprendizaje Supervisado

Dado el siguiente conjunto de datos en formato CSV:

```
Edad,Salario,Compra
25,50000,No
35,60000,Sí
45,80000,Sí
30,70000,No
40,85000,Sí
```

Carga estos datos en un `DataFrame` de `pandas` y utiliza la librería `sklearn` para entrenar un modelo de clasificación que prediga si una persona comprará un producto basado en su edad y salario.

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
import pandas as pd

# Cargar los datos
data = ...

# Separar variables independientes y dependientes
X = ...
y = ...

# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(...)

# Entrenar el modelo
model = DecisionTreeClassifier()
model.fit(...)

# Evaluar el modelo
accuracy = model.score(...)
print(f"Accuracy: {accuracy}")
```



## Problema 2 — Regresión Simple

Usa el siguiente conjunto de datos para predecir el precio de una casa basado en su tamaño en pies cuadrados:

Tamaño	Precio
1000	150000
1500	200000
2000	250000
2500	300000
3000	350000

Crea un modelo de regresión lineal y muestra la ecuación de la recta ajustada.

```
from sklearn.linear_model import LinearRegression

# Datos
X = ...
y = ...

# Crear modelo
model = LinearRegression()
model.fit(...)

# Mostrar ecuacion de la recta
print(f"Coeficiente: {model.coef_}")
print(f"Intercepto: {model.intercept_}")
```

### Problema 3 — Regresión Múltiple

Amplía los datos del Problema 2 agregando una columna de número de habitaciones:

```
Tamaño,Habitaciones,Precio
1000,2,150000
1500,3,200000
2000,4,250000
2500,4,300000
3000,5,350000
```

Entrena un modelo de regresión múltiple para predecir el precio de una casa.

## Problema 4 — Clasificación Simple

Utilizando el conjunto de datos del Problema 1, usa una regresión logística para predecir la compra de un producto.

```
from sklearn.linear_model import LogisticRegression

# Crear modelo
model = LogisticRegression()
model.fit(...)

# Predicción
pred = model.predict(...)
print(f"Predicciones: {pred}")
```

### Problema 5 — Clasificación Múltiple

Dado el siguiente conjunto de datos, crea un modelo de clasificación que prediga la categoría del cliente (A, B o C) basado en el ingreso y la edad:

Edad	Ingreso	Categoría
25	30000	A
35	50000	B
45	70000	C
30	40000	A
40	60000	B

## Problema 6 — Aprendizaje No Supervisado

Usa el siguiente conjunto de datos para aplicar el algoritmo de K-Means y agrupar los puntos en 2 clústeres:

X,Y  
1,2  
2,1  
3,4  
5,6  
6,5

```
from sklearn.cluster import KMeans

# Datos
X = ...

# Crear modelo
kmeans = KMeans(n_clusters=2)
kmeans.fit(...)

# Mostrar centroides
print(f"Centroides: {kmeans.cluster_centers_}")
```

### **Problema 7 — Reducción de Dimensionalidad**

Aplica PCA (Análisis de Componentes Principales) a un dataset con 3 dimensiones y reduce el espacio a 2 dimensiones. Usa datos sintéticos generados con `numpy`.

## Problema 8 — Matriz de Adyacencia

Crea una matriz de adyacencia que represente el siguiente grafo:

$$A \leftrightarrow B, A \leftrightarrow C, B \leftrightarrow D, C \leftrightarrow D$$

```
import numpy as np

# Matriz de adyacencia
adj_matrix = np.array([
    [0, 1, 1, 0],
    [1, 0, 0, 1],
    [1, 0, 0, 1],
    [0, 1, 1, 0]
])

print(adj_matrix)
```

### **Problema 9 — Matriz de Similitud**

Dado un conjunto de puntos en un espacio 2D, calcula una matriz de similitud basada en la distancia euclidiana inversa.



## Problema 10 — Clusterización

Aplica el algoritmo DBSCAN para encontrar clústeres en el siguiente conjunto de datos:

X,Y  
1,2  
2,2  
3,3  
8,8  
8,9  
25,30

```
from sklearn.cluster import DBSCAN

# Datos
X = ...

# Aplicar DBSCAN
dbscan = DBSCAN(eps=2, min_samples=2)
dbscan.fit(...)

# Mostrar etiquetas de cluster
print(f"Etiquetas: {dbscan.labels_}")
```

## Módulo IV — Deep Learning

### Problema 1 — Modelo Generalizado

Construye un modelo generalizado con Keras que tenga:

- Una capa de entrada con 10 neuronas.
- Una capa oculta con 16 neuronas y activación ReLU.
- Una capa de salida con una sola neurona y activación sigmoidal.

```
from keras.models import Sequential
from keras.layers import Dense

# Crear el modelo
model = Sequential([
    Dense(16, activation='relu', input_shape=(10,)),
    ...
])

# Compilar el modelo
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

## Problema 2 — Perceptrón Simple

Implementa un perceptrón simple para resolver el problema lógico AND:

Input: [0,0], [0,1], [1,0], [1,1]

Output: [0, 0, 0, 1]

```
import numpy as np
from keras.models import Sequential
from keras.layers import Dense

# Datos
X = np.array([[0,0], [0,1], [1,0], [1,1]])
y = np.array([0], [0], [0], [1])

# Modelo
model = Sequential([
    Dense(1, activation='sigmoid', input_shape=(2,))
])

model.compile(optimizer='sgd', loss='binary_crossentropy', metrics=['accuracy'])

# Entrenar
model.fit(X, y, epochs=100, verbose=1)
```

### Problema 3 — Capa de Perceptrones Múltiples

Crea un modelo con tres capas densas: la primera de 8 neuronas, la segunda de 4 neuronas, y la tercera de 2 neuronas, todas con activación ReLU.

```
model = Sequential([
    Dense(8, activation='relu', input_shape=(10,)),
    Dense(4, activation='relu'),
    Dense(2, activation='relu')
])
```

## Problema 4 — Modelo Secuencial Simple

Entrena un modelo secuencial para predecir el precio de una casa basada en el tamaño:

Tamaño (m<sup>2</sup>): [50, 60, 70, 80, 90]

Precio (USD): [100, 120, 140, 160, 180]

```
X = np.array([[50], [60], [70], [80], [90]])
y = np.array([[100], [120], [140], [160], [180]])

model = Sequential([
    Dense(1, activation='linear', input_shape=(1,))
])

model.compile(optimizer='sgd', loss='mean_squared_error')
model.fit(X, y, epochs=100, verbose=1)
```

### **Problema 5 — Capas Densas Ocultas**

Modifica el modelo del problema 4 para agregar una capa oculta con 10 neuronas y activación ReLU.

## Problema 6 — Regresión Generalizada

Entrena un modelo para realizar una regresión en el siguiente conjunto de datos:

x: [1, 2, 3, 4, 5]

y: [2.1, 4.2, 6.3, 8.4, 10.5]

Ajusta la salida usando una función de pérdida `mean_absolute_error`.

### Problema 7 — Clasificación Generalizada

Usa un modelo de clasificación para identificar si un punto en el plano pertenece al primer cuadrante. Los datos son:

x:  $[(1,1), (-1,1), (1,-1), (-1,-1)]$

y:  $[1, 0, 0, 0]$

Entrena el modelo usando activación sigmoideal y optimizador Adam.



## Problema 8 — Exportación de una Red Neuronal

Guarda el modelo creado en el Problema 7 en un archivo llamado `modelo.h5`.

```
model.save('modelo.h5')
```

## Problema 9 — Importación de una Red Neuronal

Carga el modelo guardado en el Problema 8 y realiza predicciones en los datos:

```
from keras.models import load_model

# Cargar modelo
model = load_model('modelo.h5')

# Predicciones
pred = model.predict([[1,1]])
print(pred)
```

## Problema 10 — Predicción y Evaluación

Evalúa el modelo del Problema 9 usando los siguientes datos de prueba:

X\_test: [(1,1), (0,0), (-1,-1)]  
y\_test: [1, 0, 0]

```
# Evaluar modelo  
loss, accuracy = model.evaluate(X_test, y_test)  
print(f"Loss: {loss}, Accuracy: {accuracy}")
```