

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 fx =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										

fx	=FILTRAR(A2:D7,B2:B7=F3)
----	--------------------------

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 range A1:E9. 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

	A	B	C	D	E	F
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	
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						

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 *i*
: Cliente *i*

Valores

: Cuenta de Crédito *i*
: Suma de Monto *i*
: Suma de Deuda *i*
: Promedio de Monto *i*
: Promedio de Deuda *i*

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

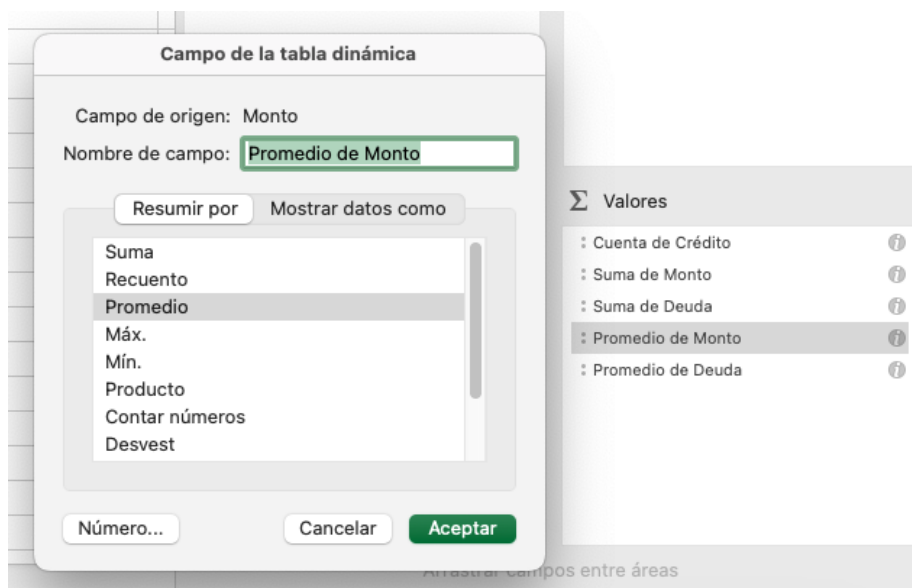


Figura 8: Solución al Problema 2 incisos (c), (d), (e), (f), (g)

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
Hombre	4	\$ 57,900.00	\$ 12,400.00	\$ 14,475.00	\$ 3,100.00
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
Mujer	4	\$ 41,500.00	\$ 12,500.00	\$ 10,375.00	\$ 3,125.00
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
Total general	8	\$ 99,400.00	\$ 24,900.00	\$ 12,425.00	\$ 3,112.50

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.

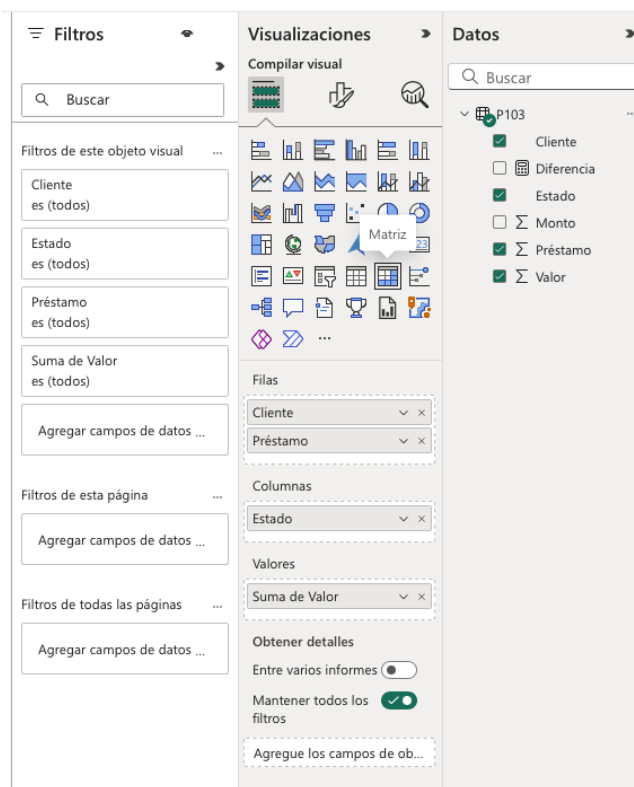
(g) De la matriz del inciso (b) responde cuál el es el préstamo con menor deuda total y usando el inciso (a) cómo es su deuda respecto al préstamo.

(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
C001	2000	9000	16000	27000
1	2000	7000	6000	15000
2	0	2000	10000	12000
C002	10000	11000	6000	27000
1	4000	8000	6000	18000
2	6000	3000	0	9000
C003	5000	12000	6000	23000
1	5000	12000	6000	23000
Total	17000	32000	28000	77000

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.

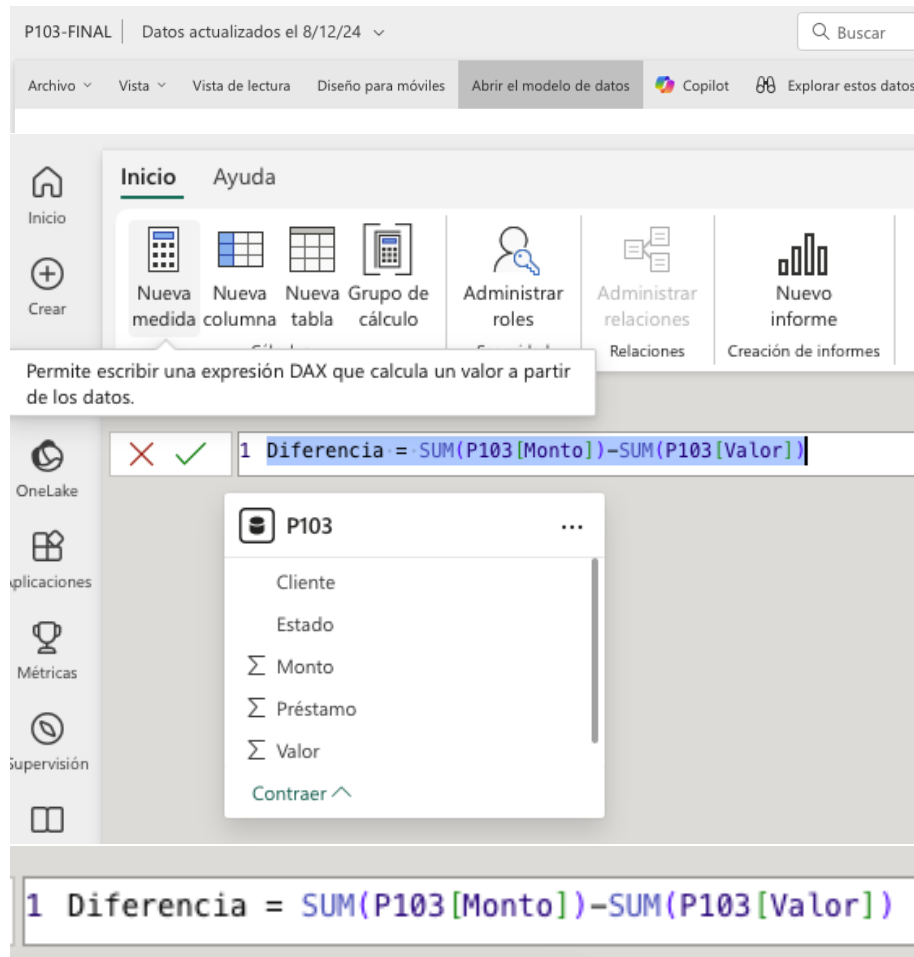


Figura 12: Solución al Problema 3 inciso (b)

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.

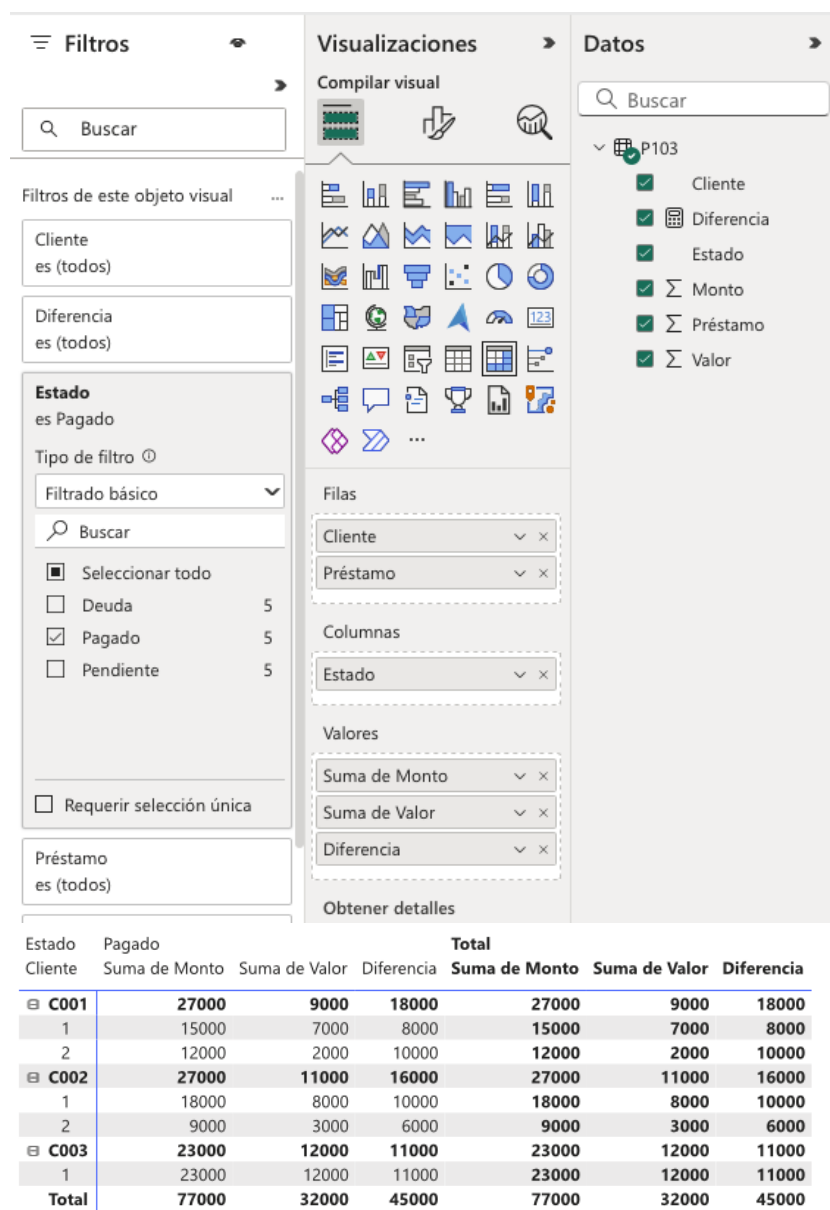
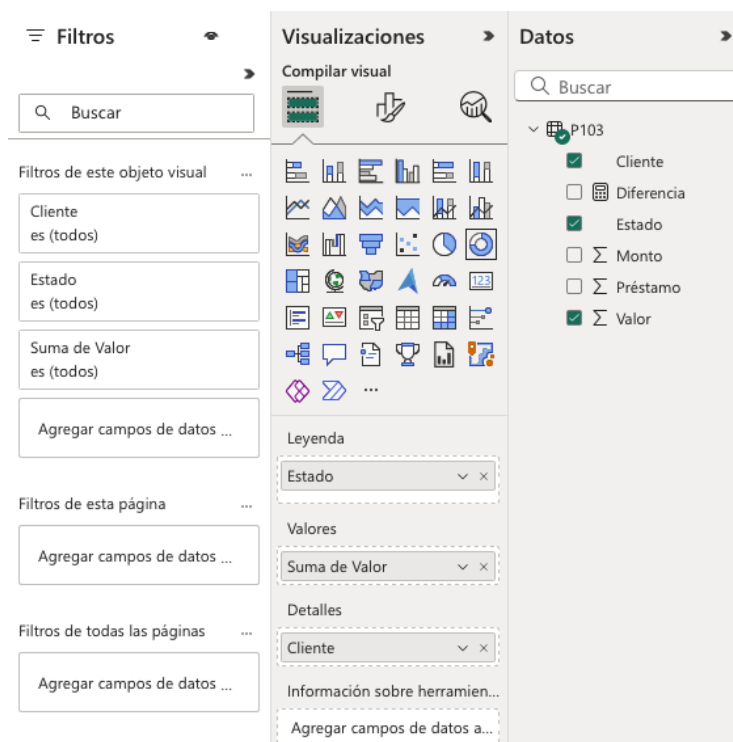


Figura 13: Solución al Problema 3 inciso (c)

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

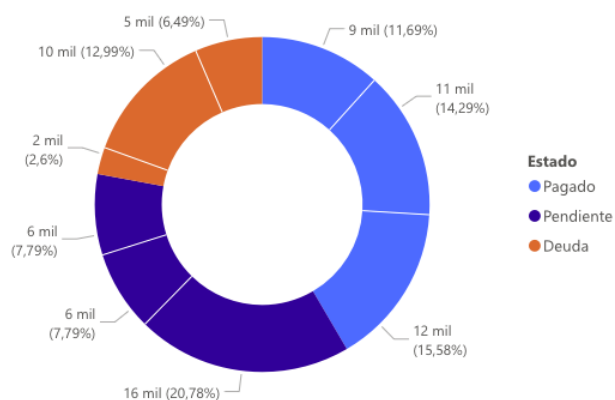


Figura 14: Solución al Problema 3 inciso (d)

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.

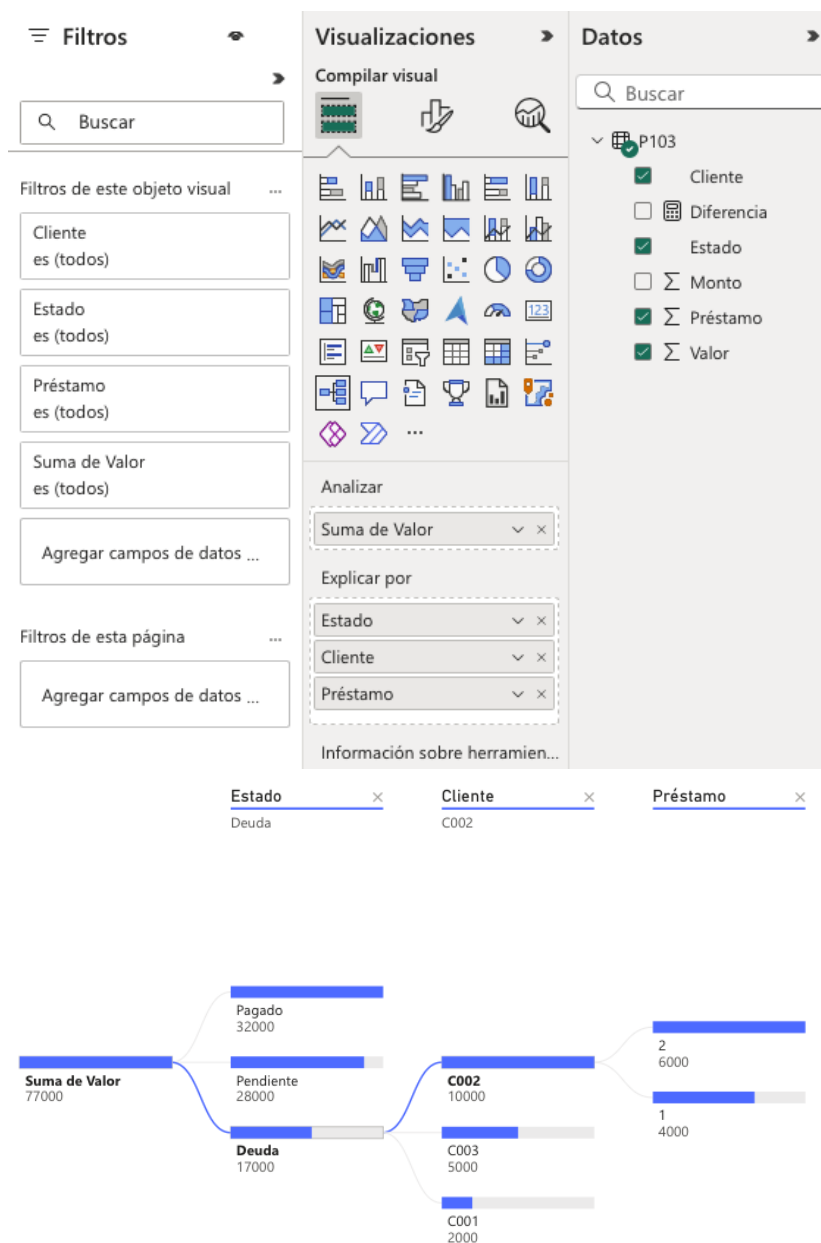


Figura 15: Solución al Problema 3 inciso (e)

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

def reporteBalanceDiaFin(valance):
    reporteTitulo("B-OUT", "Balance al Final del Día")
    print(f"${valance:.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
- Problema 6 — Interés Compuesto
- Problema 7 — Tasa de Interés Variable
- Problema 8 — Tabla de Amortización Simple
- Problema 9 — Interés Simple Futuro
- Problema 10 — Punto de Liquidación

Módulo II — Estadística y Probabilidad con Python

Problema 1 — Probabilidad de un Evento Simple

Problema 2 — Tabla conjunta de Dos Eventos

Problema 3 — Probabilidad Conjunta de Dos Eventos

Problema 4 — Probabilidad Condicional de Dos Eventos

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

Problema 2 — Dataset de indicadores en CSV

Problema 3 — Adquisición de un CSV en Python

Problema 4 — Indicadores en Series de Pandas

Problema 5 — Indicadores en DataFrames de Pandas

Problema 6 — Filtrar columnas en DataFrames

Problema 7 — Filtrar filas en DataFrames

Problema 8 — Agrupación de datos en DataFrames

Problema 9 — Combinación de datos en DataFrames

Problema 10 — Visualización de datos

Módulo IV — Machine Learning

Problema 1 — Aprendizaje Supervisado

Problema 2 — Regresión Simple

Problema 3 — Regresión Múltiple

Problema 4 — Clasificación Simple

Problema 5 — Clasificación Múltiple

Problema 6 — Aprendizaje No Supervisado

Problema 7 — Reducción de Dimensionalidad

Problema 8 — Matriz de Adyacencia

Problema 9 — Matriz de Similitud

Problema 10 — Clusterización

Módulo IV — Deep Learning

Problema 1 — Modelo Generalizado

Problema 2 — Perceptrón Simple

Problema 3 — Capa de Perceptrones Múltiples

Problema 4 — Modelo Secuencial Simple

Problema 5 — Capas Densas Ocultas

Problema 6 — Regresión Generalizada

Problema 7 — Clasificación Generalizada

Problema 8 — Exportación de una Red Neuronal

Problema 9 — Importación de una Red Neuronal

Problema 10 — Predicción y Evaluación