# 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.

	Α	В	С		D
1	Cliente	Género	Crédito	Мо	nto
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 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 suma de los montos para hombres y la suma de montos para mujeres usando la fórmula =SUMA(FILTRAR(RANGO SUMA, RANGO INCLUIR)).

#### Solución al Problema 1

Para el inciso (a) utilizamos las fórmulas

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

	Α	В	С		D	E	F	G		Н	1	J
Clie	nte	Género	Crédito	Mon	to			USAN	IDO C	ONTAR.SI Y SI	JMAR.SI	
C00	1	Hombre	P0001	\$	14,600.00			Total	Sum	a Monto		
C00	2	Mujer	P0002	\$ :	18,900.00		Hombre	3	\$	50,300.00		
1 C00	3	Hombre	E0004	\$ :	12,000.00		Mujer	3	\$	29,000.00		
C00	4	Mujer	E0032	\$	6,900.00							
C00	5	Mujer	F0005	\$	3,200.00				USA	ANDO FILTRAF	3	
7 C00	6	Hombre	G0128	\$ :	23,700.00							

$f_{x}$	=SUMAR.SI(B2:B7,"HOMBRE",D2:D7)

Figura 2: Solución al Problema 1 inciso (a)

Para el inciso (b) utilizamos las fórmulas

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



Figura 3: Solución al Problema 1 inciso (b)

Nota: Si se utilizamos las fórmulas

=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.

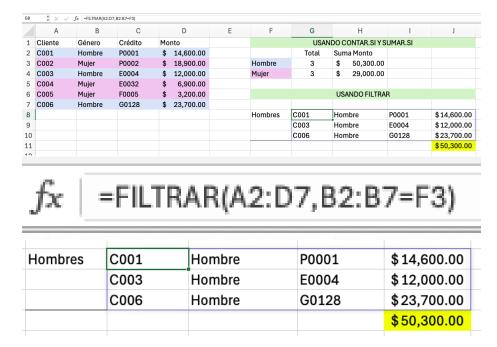


Figura 4: Tabla virtual generada por FILTRAR (c)

### 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.

	Α	В	С	D	Е
1	Cliente	Género	Crédito	Monto	Deuda
2	C001	Hombre	P0001	14,600.00	\$ 3,200.00
3	C001	Hombre	H0014	<b>16,300.00</b>	\$ 2,500.00
4	C002	Mujer	P0002	<b>18,900.00</b>	\$ 1,600.00
5	C003	Hombre	E0004	all \$ 12,000.00	\$ 4,800.00
6	C004	Mujer	E0032	。。。\$ <b>6,900.00</b>	\$ 5,300.00
7	C005	Mujer	F0005	all \$ 3,200.00	\$ 2,100.00
8	C006	Hombre	G0128	<b>11</b> \$ 15,000.00	\$ 1,900.00
9	C005	Mujer	F0095	<b>□</b> \$ 12,500.00	\$ 3,500.00

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

- (a) Crea una tabla dinámica de A1:E9.
- (b) Segmenta las filas por Género y luego por Cliente.
- (c) Cuenta el número de Créditos en valores.
- (d) Calcula la suma del Monto en valores.
- (e) Calcula la suma de la Deuda en valores.
- (f) Calcula el promedio del Monto en valores.
- (g) Calcula el promedio de la Deuda en valores.
- (h) ¿Quién es el cliente de mayor deuda?
- (i) ¿Quién es el cliente al que se le ha prestado más en promedio?
- (j) ¿Quién es el cliente hombre con mayor deuda promedio?
- (k) ¿Es significativamente distinta la deuda entre hombres y mujeres?
- (l) ¿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.

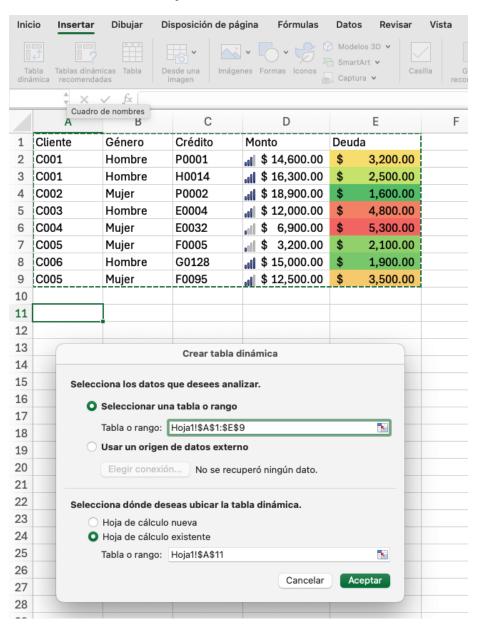


Figura 6: Solución al Problema 2 inciso (a)

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

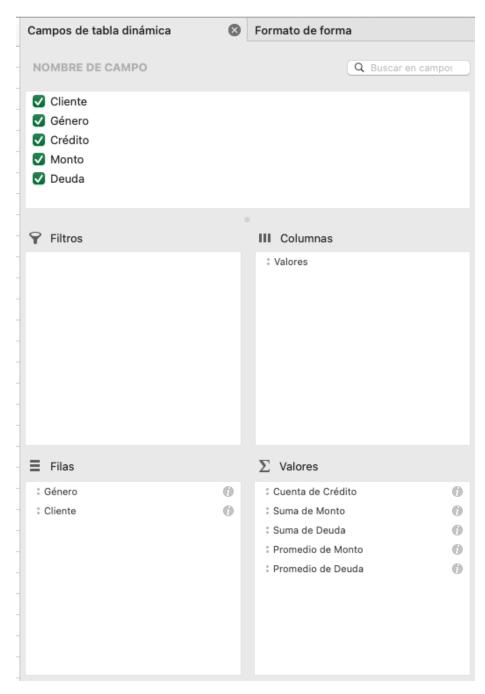


Figura 7: Solución al Problema 2 inciso (b)

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

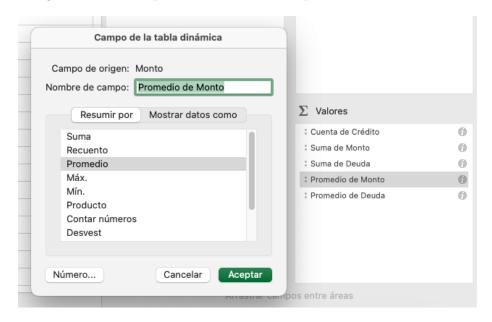


Figura 8: Solución al Problema 2 incisos (c), (d), (e), (f), (g)

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

Etiquetas de fila 🔻 Cuenta de Cré	dito	Sun	na de Monto	Sun	na de Deuda	Pron	nedio de Monto	Pror	nedio de Deuda
<b>■ Hombre</b>	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.



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.

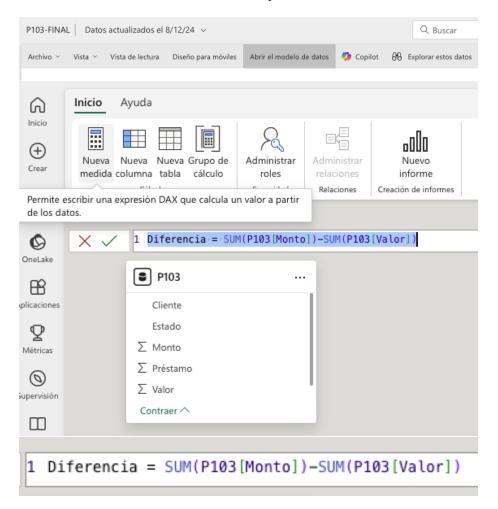


Figura 12: Solución al Problema 3 inciso (b)

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.

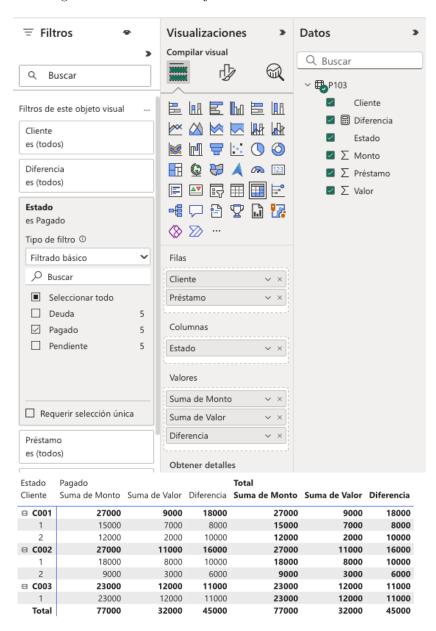
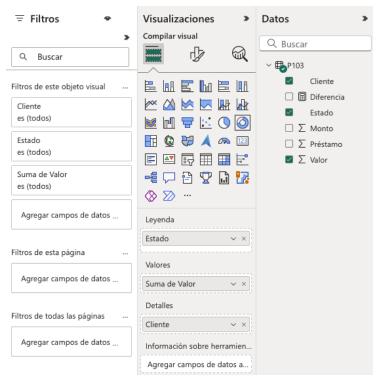


Figura 13: Solución al Problema 3 inciso (c)

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

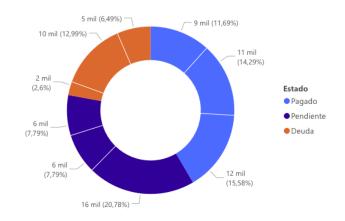


Figura 14: Solución al Problema 3 inciso (d)

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.

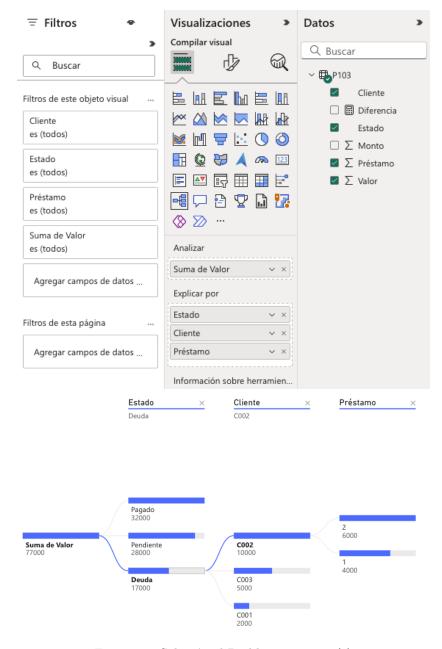


Figura 15: Solución al Problema 3 inciso (e)

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 siguietes. 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"-----")
def reporteFin(codigo, leyenda):
   print(f"-----")
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")
\tt def\ reporteTotalOperacionesDia(numOperaciones):
    reporteTitulo("T-OP", "Total de Operaciones del Día")
    # Imprime el número de operaciones
   reporteFin("T-OP", "Total de Operaciones del Día")
\tt 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 dia
    # 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.

	Α	В
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 del 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  ${\bf 0.34\%}$

### 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 préstamos en los 36 meses.
- (d) Calcula el valor de las cuotas fijas para cubrir el préstamos 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
- 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 Viasualización 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