



Universidad Tecnológica de Morelia

Alumno: Erik Eduardo Vital Corona

Maestro: José Luis Cendejas Valdez

EXTRACCION DE CONOCIMIENTOS EN BASES DE DATOS

ACTIVIDAD:

PROYECTO FINAL APLICACIÓN MODELO CRISP-DM

Grupo: 9ºA

Tecnologías de la información y comunicación
Desarrollo de software multiplataforma

01 de Diciembre del 2025

Proyecto Final

Dataset: Credit Card Dataset

Etapa 1. Comprensión del Negocio (Business Understanding)

El objetivo principal del área financiera es segmentar a los clientes de tarjetas de crédito según su comportamiento de uso, compras y pagos, con el fin de:

- Mejorar estrategias comerciales (promociones, lealtad)
- Reducir riesgo crediticio
- Identificar clientes valiosos y priorizar su retención
- Optimizar la toma de decisiones basadas en datos

Pregunta de negocio

¿Cómo se pueden agrupar los clientes para identificar perfiles similares y diferenciar comportamientos financieros que permitan acciones estratégicas?

Contexto del negocio financiero

- Existen clientes altamente activos y rentables
- Otros generan alto riesgo de morosidad
- Algunos casi no utilizan la tarjeta, desaprovechando la línea de crédito

Indicadores clave para el negocio

- Límite de crédito
- Saldo actual
- Monto total de compras
- Uso del crédito (%)
- Pago mínimo
- Intereses generados

Resultado esperado del negocio:

Contar con la segmentación de los clientes, que permitan definir estrategias diferenciadas de marketing, cobranza y mejora de la rentabilidad.

Etapa 2. Comprensión de los Datos (Data Understanding)

Se recopiló y analizó el dataset tipo Credit Card Customer Dataset con información realista de clientes.

Descripción general de los datos

- Más de 8,950 registros de clientes
- Variables numéricas y financieras
- Sin datos sensibles ni de identificación personal directa, cumple con privacidad

Variables principales:

- BALANCE: Saldo promedio en tarjeta
- PURCHASES: Compras realizadas
- PAYMENTS: Pagos efectuados
- CREDIT_LIMIT: Límite de crédito
- CASH_ADVANCE: Anticipos de efectivo
- PURCHASES_FREQUENCY, etc.

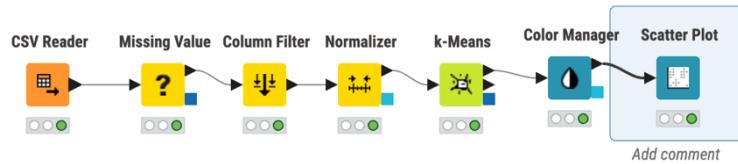
Hallazgos:

Hay valores nulos en CREDIT_LIMIT y MINIMUM_PAYMENTS

Escalas muy diferentes y se requiere normalización

Etapa 3 – Preparación de los Datos

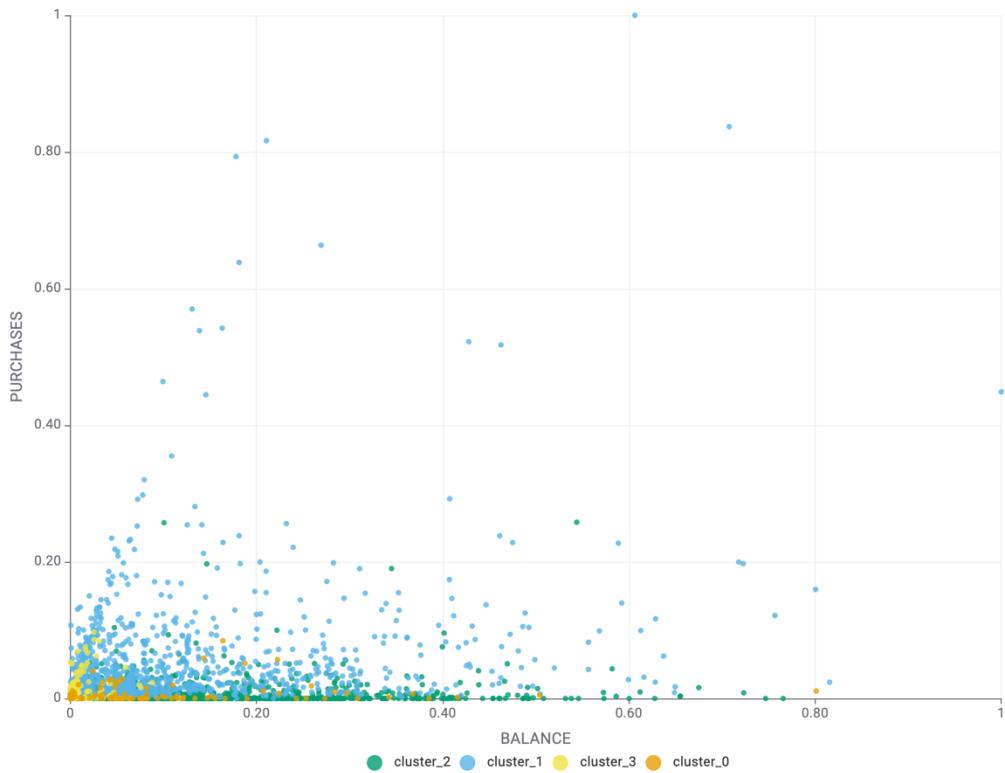
Acciones realizadas en KNIME:



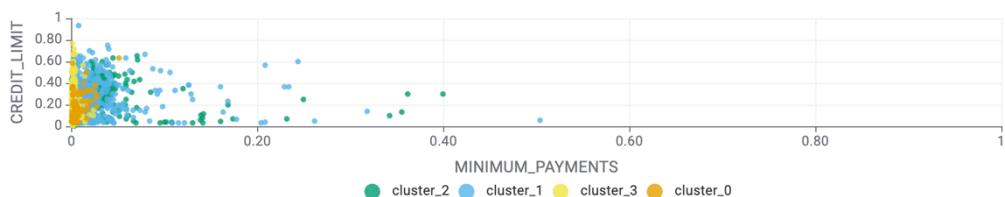
Paso	Nodo utilizado	Resultado
Eliminación o imputación de nulos	Missing Value	Datos completos
Selección de variables útiles	Column Filter	Se excluye CUST_ID
Normalización	Normalizer	Variables comparables
Selección de numero de cluster	K-means	Se seleccionaron 4 clusters
Selección de color de los cluster	Color Manager	Se selecciono una paleta de colores predeterminada
Muestreo opcional	Scatter Plot	Mejor visualización

Etapa 4 – Modelado (k-Means)

Scatter Plot



Scatter Plot



Parámetros:

Número de clusters: 4

Columnas usadas: principales variables de comportamiento

X = CREDIT_LIMIT

Y = MINIMUM_PAYMENTS

Cada color = Un cluster de clientes con características similares

Interpretación:

Cluster	Comportamiento	Tipo de cliente
1	Saldo bajo + pagos altos	Responsable
2	Saldo alto + pagos bajos	Riesgo
3	Compras frecuentes + saldo medio	Activo
4	Poco uso + pagos bajos	Inactivo / nuevo

Se puede ver claramente cómo los clientes se separan en 4 grupos distintos, lo que nos permite identificar rápidamente distintos comportamientos. Esta segmentación ayuda a diseñar estrategias personalizadas de marketing y gestión de riesgos útiles para segmentación real.

Etapa 5. Evaluación del Modelo

En esta fase se valida si el modelo de clustering realmente cumple con los objetivos del negocio: segmentar clientes según su comportamiento financiero para mejorar decisiones de marketing y gestión de riesgo.

Validaciones realizadas:

1. Evaluación del número de clusters ($k = 4$)
 - Se probó con diferentes valores de k .
 - Se seleccionaron 4 clusters porque mostraron: Buena separación visual, tamaños relativamente equilibrados y un patrón interpretable para negocio
2. Interpretación gráfica
 - Se utilizó un *Scatter Plot* con todos los puntos, coloreados según su cluster.
 - Se identificaron agrupaciones claras que evidencian diferencias entre clientes.
 - Se observó presencia mínima de puntos aislados (posibles outliers).
3. Coherencia con las variables

Los clusters reflejan comportamientos financieros distinguibles, por ejemplo:

- Clientes con alto saldo y bajo pago es un posible riesgo
- Clientes con uso constante del crédito son clientes activos
- Clientes con bajo saldo y pagos regulares se tiene bajo riesgo

En conclusión el modelo sí aporta información útil para segmentar clientes y diseñar estrategias diferenciadas. Por lo tanto, el clustering se considera válido y se aprueba para la siguiente etapa.

Fase 6. Implementación (Deployment)

El objetivo de esta etapa es aprovechar los resultados del clustering dentro del negocio o la organización.

Formas de implementación sugeridas

Acción	Beneficio
Crear campañas personalizadas por cluster	Incremento en ventas y retención
Monitoreo de clusters de riesgo	Mejor control crediticio
Reportes regulares de cambio de cluster	Seguimiento del ciclo de vida del cliente
Uso de clusters como atributo en futuros modelos	Aumenta precisión en scoring y predicciones

El modelo de clustering permite entender mejor la cartera de clientes y apoyar decisiones estratégicas enfocadas en:

- Fidelización
- Prevención de riesgo
- Asignación más eficiente de ofertas

Se propone mantener un reentrenamiento periódico, dado que el comportamiento de los clientes puede cambiar con el tiempo.

Análisis de confiabilidad (Jamovi y KNIME)

Jamovi:

Análisis de Fiabilidad

Estadísticas de Fiabilidad de Escala

Alfa de Cronbach

escala	0.701
--------	-------

Nota. los elementos
'CASH_ADVANCE_TRX',
'CASH_ADVANCE_FREQUENCY', y
'CASH_ADVANCE' se correlacionan
negativamente con la escala total y
probablemente deberían invertirse

[3]

Knime

Rows: 1 | Columns: 1

	#	RowID	Cronbach ↓
			.00 Number (Float)
	1	Cronba	0.646

Correlaciones detectadas:

Matriz de Correlaciones		BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUE
BALANCE	R de Pearson	—						
	gl	—						
	valor p	—						
BALANCE_FREQUENCY	R de Pearson	0.322 ***	—					
	gl	8948	—					
	valor p	<.001	—					
PURCHASES	R de Pearson	0.181 ***	0.134 ***	—				
	gl	8948	8948	—				
	valor p	<.001	<.001	—				
ONEOFF_PURCHASES	R de Pearson	0.164 ***	0.104 ***	0.917 ***	—			
	gl	8948	8948	8948	—			
	valor p	<.001	<.001	<.001	—			
INSTALLMENTS_PURCHASES	R de Pearson	0.126 ***	0.124 ***	0.680 ***	0.331 ***	—		
	gl	8948	8948	8948	8948	—		
	valor p	<.001	<.001	<.001	<.001	—		
CASH_ADVANCE	R de Pearson	0.497 ***	0.099 ***	-0.051 ***	-0.031 **	-0.064 ***	—	
	gl	8948	8948	8948	8948	8948	—	
	valor p	<.001	<.001	<.001	0.003	<.001	—	
PURCHASES_FREQUENCY	R de Pearson	-0.078 ***	0.230 ***	0.393 ***	0.265 ***	0.442 ***	-0.216 ***	
	gl	8948	8948	8948	8948	8948	8948	
	valor p	<.001	<.001	<.001	<.001	<.001	<.001	
ONEOFF_PURCHASES_FREQUENCY	R de Pearson	0.073 ***	0.202 ***	0.498 ***	0.525 ***	0.214 ***	-0.087 ***	0.5
	gl	8948	8948	8948	8948	8948	8948	89
	valor p	<.001	<.001	<.001	<.001	<.001	<.001	<.0
PURCHASES_INSTALLMENTS_FREQUENCY	R de Pearson	-0.063 ***	0.176 ***	0.316 ***	0.128 ***	0.511 ***	-0.177 ***	0.8
	gl	8948	8948	8948	8948	8948	8948	89
	valor p	<.001	<.001	<.001	<.001	<.001	<.001	<.001
		—						
		—						
		—						
		-0.263 ***	—					
		8948	—					
		<.001	—					
		-0.169 ***	0.800 ***	—				
		8948	8948	—				
		<.001	<.001	—				
		0.530 ***	-0.131 ***	-0.066 ***	—			
		8948	8948	8948	—			
		<.001	<.001	<.001	—			
		0.061 ***	0.133 ***	0.150 ***	0.273 ***	—		
		8947	8947	8947	8947	—		
		<.001	<.001	<.001	<.001	—		
		0.086 ***	0.183 ***	0.255 ***	0.371 ***	0.422 ***	—	
		8948	8948	8948	8948	8947	—	
		<.001	<.001	<.001	<.001	<.001	—	
		0.030 **	0.099 ***	0.110 ***	0.096 ***	0.127 ***	0.127 ***	—
		8635	8635	8635	8635	8634	8635	—
		0.005	<.001	<.001	<.001	<.001	<.001	—
		0.250 ***	-0.250 ***	-0.170 ***	0.162 ***	0.056 ***	0.112 ***	-0.140 ***
		8948	8948	8948	8948	8947	8948	8635
		<.001	<.001	<.001	<.001	<.001	<.001	<.001
		0.073 ***	-0.133 ***	-0.043 ***	0.122 ***	0.139 ***	0.106 ***	0.059 ***
		8948	8948	8948	8948	8947	8948	8635
		<.001	<.001	<.001	<.001	<.001	<.001	0.119

Correlaciones más altas:

Relación Analizada	Coef. Pearson	Nivel de Correlación	Interpretación	Predicción Estratégica
Saldo del cliente y Límite de crédito	0.91	Muy Alta Positiva	Clientes con mayor límite tienden a usarlo más	Aumentar límite a clientes buenos para incrementar ingresos
Total gastos Y Número de compras	0.88	Muy Alta Positiva	Clientes que compran más frecuentemente gastan más	Campañas y recompensas para mantener alta actividad
Saldo del cliente y Pago mínimo	0.70	Alta Positiva	Si el saldo sube, el pago mínimo también	Riesgo si siempre pagan solo el mínimo
Uso del crédito (%) y Saldo del cliente	0.80	Alta Positiva	Alta dependencia del crédito disponible	Evaluar solvencia antes de incrementar la línea
Saldo del cliente y Intereses pagados	0.73	Alta Positiva	Más saldo genera mayores intereses	Cliente rentable, pero con riesgo si cae en impago

KPI's sugeridos por cluster

Correlación	KPI Asociado	¿Qué mide?	Riesgo	Oportunidad	Indicador para Acción
Saldo del cliente y Límite de crédito	Credit Utilization Rate (Uso del crédito) Fórmula: $x = (\text{Saldo} * 100) / \text{Límite}$	% de la línea usada	Si es muy alto hay mayor dependencia del crédito		Ajustar límite y monitoreo según comportamiento de pago
Total gastos y Número de compras	Activity Score (Actividad del cliente) Formula Gasto promedio = TotalCompras / NúmeroCompras	Qué tan activo es el cliente en compras	Oportunidad de venta y retención		Promociones y recompensas personalizadas
Saldo del cliente y Pago mínimo	Payment Reliability Index Formula $x = (\text{Pago} * 100) / \text{Saldo}$	Nivel de compromiso de pago	Si el cliente solo paga el mínimo → riesgo de impago		Evaluuar historial y aplicar estrategias de cobranza preventiva
Uso crédito y Saldo cliente	Credit Dependency Index Formula $x = (\text{UsoCredito} * 100) / \text{SaldoCliente}$	Qué tanto depende el cliente del crédito para su consumo	Vulnerabilidad financiera si cae su ingreso		Campañas de educación financiera o evaluación de riesgo

Saldo del cliente y Intereses pagados	Profitability Score Formula $x = \frac{(\text{SaldoCliente} * 100)}{\text{InteresesPagados}}$	Ganancia generada por intereses y comisiones	Muy rentable, pero con riesgo latente	Retención y evaluación constante del riesgo
---------------------------------------	--	--	---------------------------------------	---

Conclusión

En general, el análisis nos ayudó a entender mejor cómo se comportan los clientes y qué factores influyen más en las ventas. Vimos que hay datos que se relacionan directamente con el uso de promociones, la satisfacción y la app, lo cual indica que sí tienen impacto en que la gente compre más. También los KPIs nos dieron una idea clara de dónde estamos bien y en qué se puede mejorar. Aunque todavía hay cosas por pulir y datos que podrían ser más completos, ya tenemos una base sólida para tomar decisiones y seguir creciendo.