

SEGMENTACIÓN DE CLIENTES BANCARIOS

Identificando segmentos para una
toma de decisiones más acertada



Escenario de negocio/problema

En el sector bancario actual, tratar a todos los clientes por igual resulta ineficiente y costoso. La segmentación permite:



Diseñar ofertas de alto valor que se apeguen al perfil de grupos de clientes aumentando el nivel de confianza en la institución y con ello la fidelidad.



Identificar grupos de alto riesgo con alta tasa de utilización de crédito y bajos ingresos, el banco puede actuar proactivamente para prevenir el impago.



Escenario de negocio/problema

Objetivo: Desarrollar un modelo no supervisado de Machine Learning para agrupar a clientes similares y basado en sus características, entonces ofrecer productos personalizados.



Variable	Tipo de Dato	Descripción
customer_id	int64	Identificador único del cliente.
age	int64	Edad del cliente en años.
gender	object	Género del cliente (M = Masculino, F = Femenino).
dependent_count	int64	Número de personas que dependen económicamente del cliente.
education_level	object	Nivel educativo máximo alcanzado.
marital_status	object	Estado civil del cliente.
estimated_income	int64	Ingreso anual estimado del cliente.
months_on_book	int64	Antigüedad del cliente en el banco (meses).
total_relationship_count	int64	Número de productos contratados por el cliente.
months_inactive_12_mon	int64	Meses de inactividad en el último año.
credit_limit	float64	Límite de crédito máximo en la tarjeta.
total_trans_amount	int64	Monto total transaccionado en los últimos 12 meses.
total_trans_count	int64	Cantidad total de transacciones en los últimos 12 meses.
avg_utilization_ratio	float64	Índice de utilización de la línea de crédito (0 a 1).
avg_transaction_value	float64	<i>Feature Engineering:</i> Monto promedio por transacción individual.

EDA: Variables numéricas



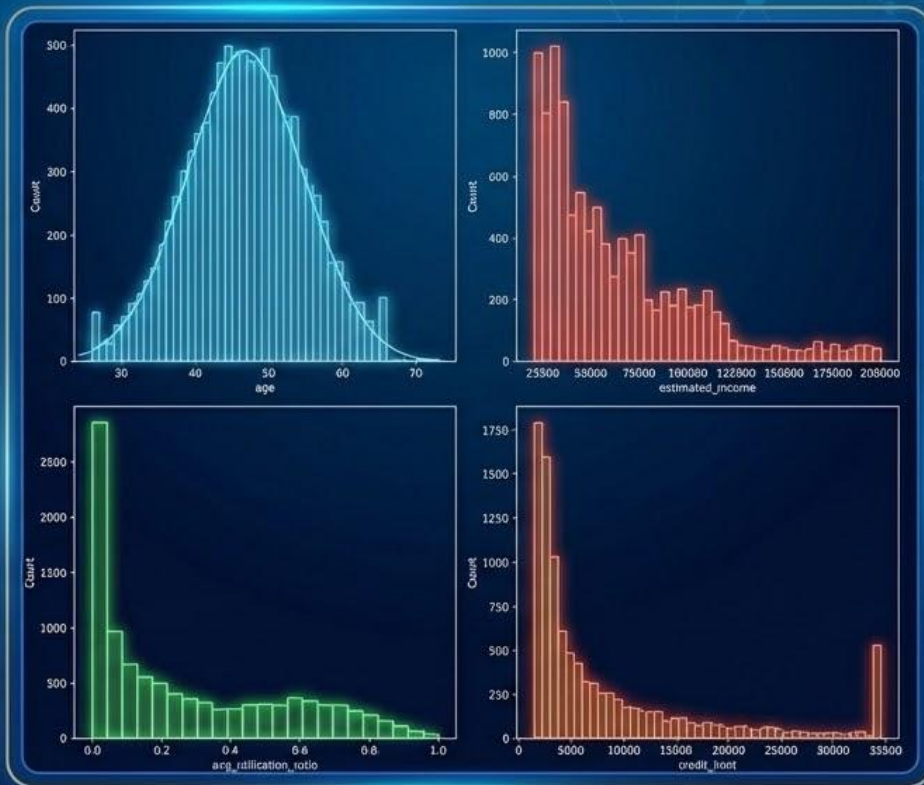
- La edad se distribuye normalmente.



- `estimated_income` y `credit limit` se encuentran sesgadas hacia la izquierda.

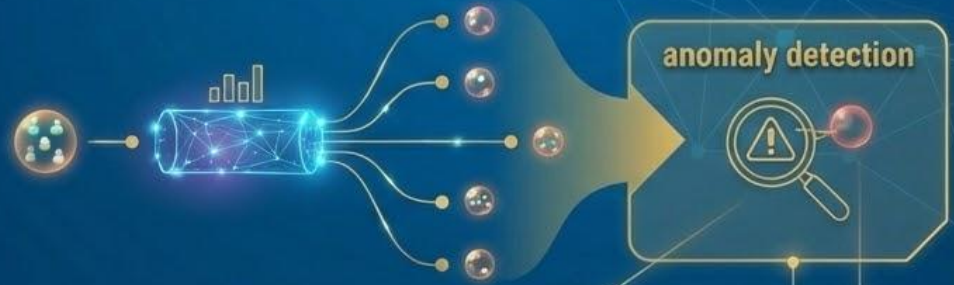


- `avg_utilization_ratio` muestra a muchos clientes utilizando un porcentaje bajo de su línea de crédito y pocos clientes utilizando la totalidad de su línea de crédito.

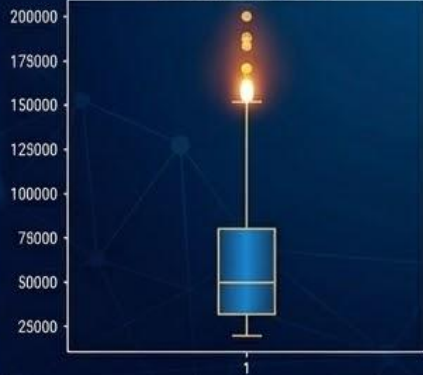


EDA: Valores atípicos

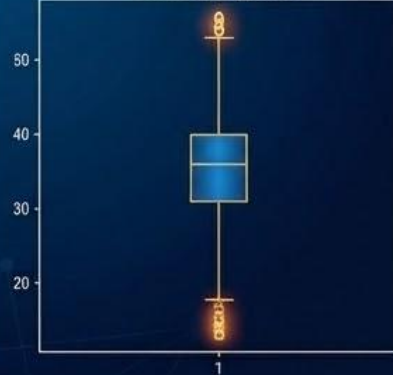
Identificamos datos atípicos extremos en las variables `estimated_income`, `credit_limit` y `total_trans_amount`.



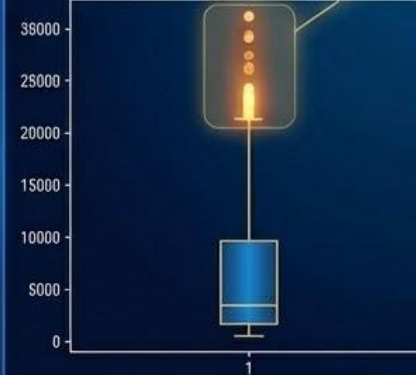
estimated_income



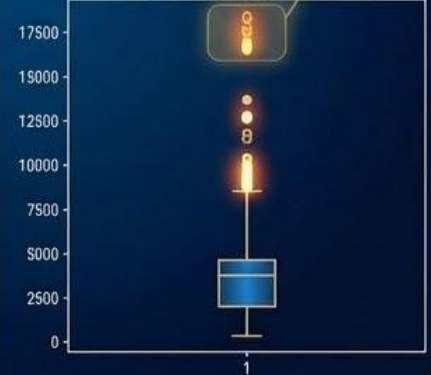
months_on_book



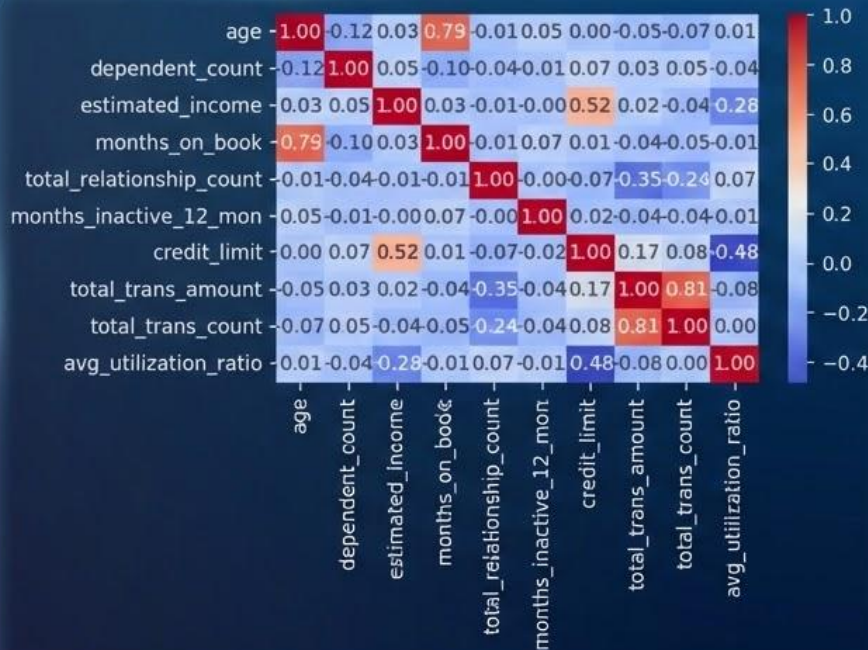
credit_limit



total_trans_amount



EDA: Correlación entre variables



Alta correlación entre las siguientes variables:



total_trans_amount &
total_trans_count -> 0.81



age & months_on_books ->
0.79



credit_limit & estimated_income
-> 0.52

Feature Engineering



1. Eliminación de columnas

'customer_id',
'months_on_book'



2. Creación de característica

'avg_transaction_value' =
 $(\text{total_trans_amount}) /$
 $(\text{total_trans_count})$



3. Mapeo y Dummies

education_level
(Mapeo), gender &
marital_status
(Dummies)



4. Transformación Logarítmica

estimated_income,
credit_limit,
avg_transaction_value



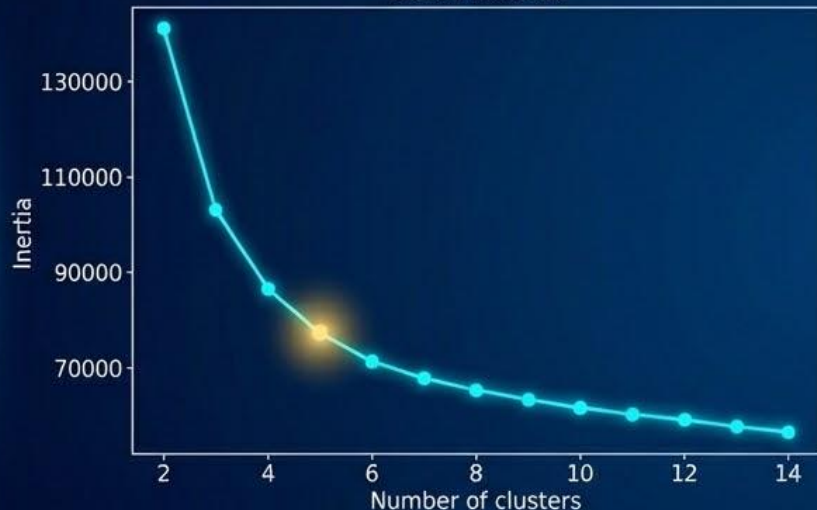
5. Estandarización de datos

StandardScaler()

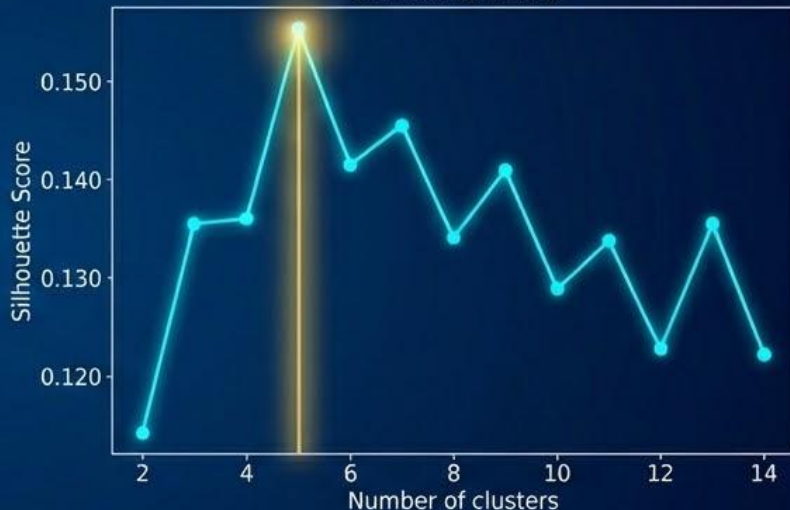
Construcción de KMeans

Una vez evitados valores atípicos y correlaciones entre variables. El análisis del método del codo y el coeficiente de silueta sugiere establecer como 5 el número de centroides.

Elbow Method



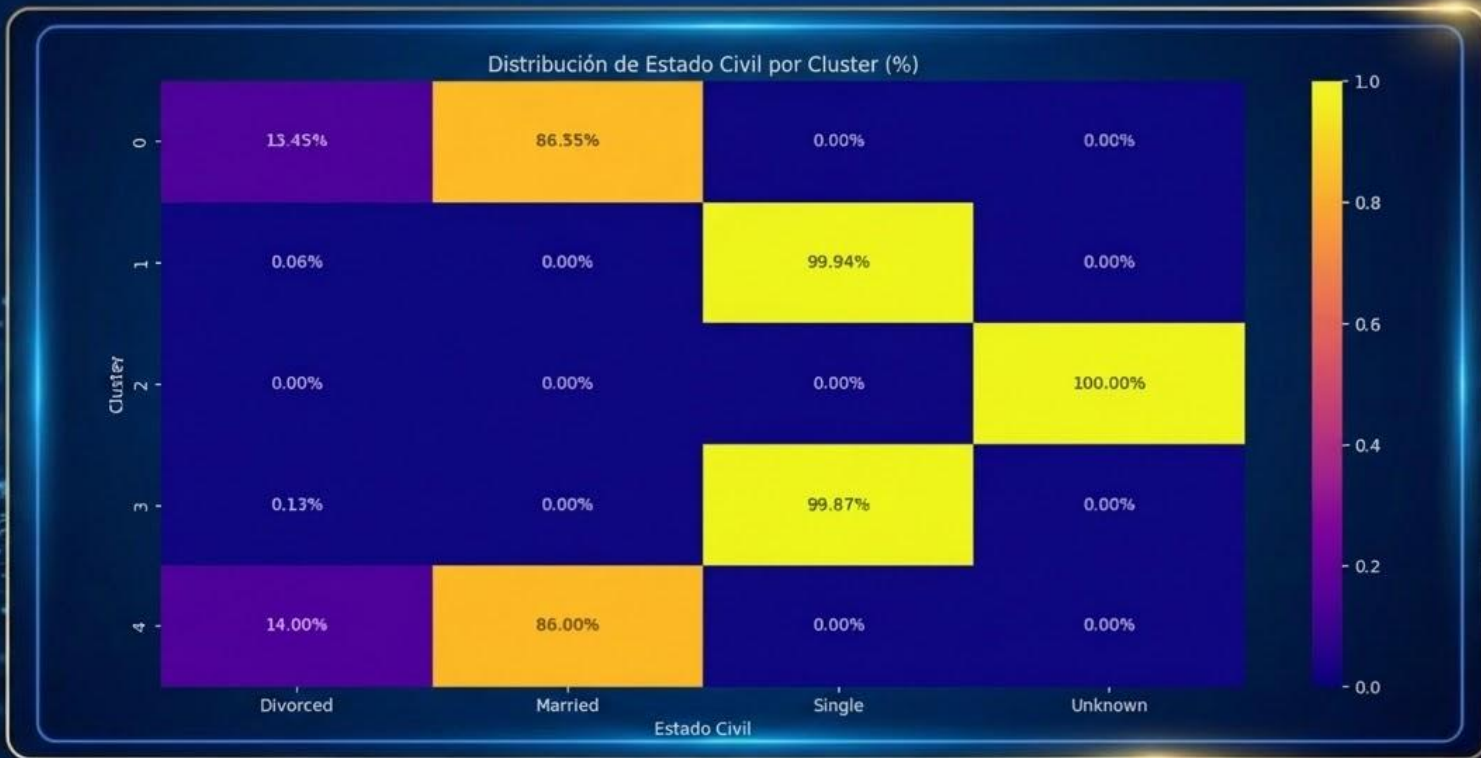
Silhouette Score



Interpretación del modelo: Perfilamiento de clusters por los valores promedio de las variables numéricas

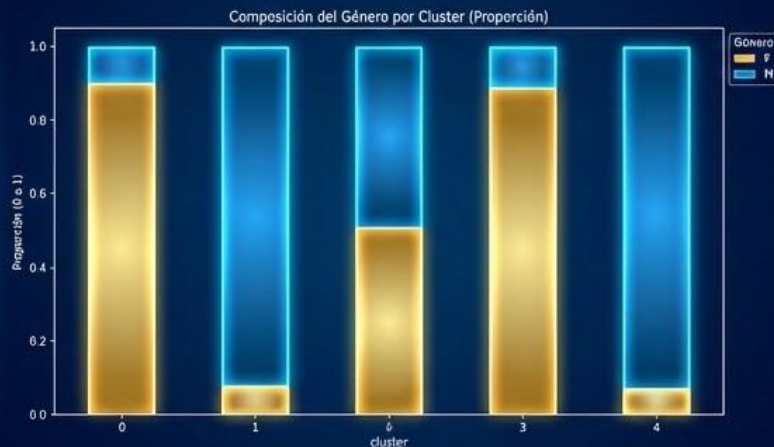


Interpretación del modelo: Perfil demográfico por cluster

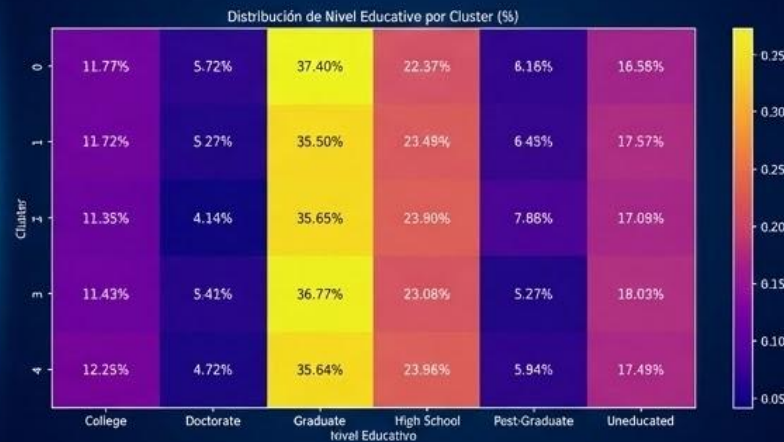


Interpretación del modelo: Perfil demográfico por cluster

Composición del Género por Cluster (Proporción)



Distribución de Nivel Educativo por Cluster (%)



Interpretación del modelo

Cluster 0: Segmento Mayoritario & Alto Uso



Segmento con más miembros. Ingresos y líneas de crédito bajos. Edad promedio más alta. Tasa de utilización y promedio de productos más altos.

Perfil demográfico: Mayormente mujeres casadas (86.55%) y divorciados (13.45%).

Cluster 1: Altos Ingresos & Bajo Uso



Posee uno de los ingresos más altos y el límite de crédito más alto. Tasa de utilización más baja, nivel de transacción sólido.

Perfil demográfico: Se constituye mayormente por hombres solteros (99.94%).

Cluster 2: Jóvenes & Datos Desconocidos



Grupo con la menor edad promedio y mayor número de dependientes. El cluster más pequeño.

Perfil demográfico: Estado civil desconocido, constituido en igual cantidad por hombres y mujeres.

Interpretación del modelo

Cluster 3: Ingresos bajos y límites de crédito reducidos



A diferencia del **cluster 0** este segmento tiene una **inactividad ligeramente mayor** y es el segundo segmento más numeroso. También tiene una utilización alta del crédito.

Perfil demográfico:



Mujeres solteras
(99.87%)



Divorciadas
(0.013%)

Cluster 4: Es el grupo con el ingreso promedio más alto **(\$92,827)**



y un **límite de crédito generoso (\$13,460)**. Tienen una tasa de utilización baja y son los que menos transacciones realizan.

Perfil demográfico:



Hombres casados
(86%)



Divorciados
(14%)

Estrategias de acción

Cluster 0: Alto Riesgo / Uso Diario



Debido a su **alta** tasa de utilización (39%) e ingresos bajos, este grupo es el más vulnerable a impagos.

Monitorear y Ofrecer **productos de consolidación de deuda** para reducir su carga financiera.

Cluster 1: Mayor Límite / Menor Uso



Representan el segmento con **mayor límite de crédito** pero **menor uso**. En este segmento, la estrategia debe enfocarse en **retención** mediante **beneficios premium** (acceso a salas VIP, seguros de viaje) para asegurar que mantengan su capital en la institución bancaria.

Estrategias de acción

Cluster 2: Jóvenes y Datos Incompletos



Al ser el grupo más joven y con datos incompletos (100% Unknown en estado civil), la prioridad es el **enriquecimiento de datos**.

Cluster 3: Uso Constante / Campañas Cashback



Es el segundo grupo más numerosos y con un uso constante del crédito, son ideales para **campañas de cashback** (p.e. supermercados y/o gasolineras) para **cimentar su lealtad**.

Cluster 4: Altos Ingresos / Baja Frecuencia



Este grupo posee los ingresos más altos pero baja frecuencia de uso. Se recomiendan ofrecer **productos de inversión** o tarjetas de crédito con **recompensas por volumen de gasto**.