

Mostrando análisis de datos

Panorama General de los Datos

Cantidad de observaciones: 16714

Cantidad de columnas: 9

Tipos de datos:

uso_de_credito	float64
edad	float64
radio_deuda	float64
ingreso_mensual	float64
cuentas_abiertas	float64
num_otros_prestamos	float64
num_dependientes	float64
dlq_2yrs	int32
nivel_riesgo	int8
dtype:	object

Cantidad de nulos por columna:

uso_de_credito	0
edad	0
radio_deuda	0
ingreso_mensual	0
cuentas_abiertas	0
num_otros_prestamos	0
num_dependientes	0
dlq_2yrs	0
nivel_riesgo	0
dtype:	int64

Cantidad de duplicados: 2

Cantidad de valores únicos por columna:

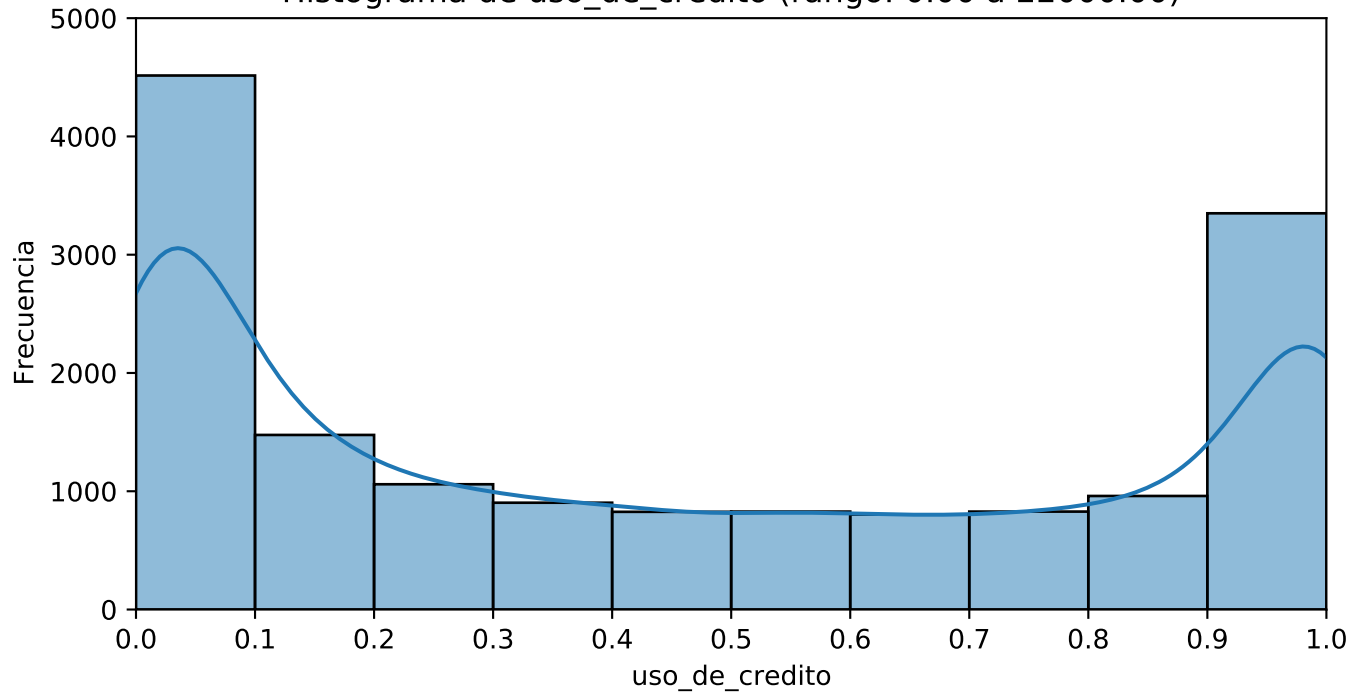
uso_de_credito	13887
----------------	-------

Estadísticas Básicas

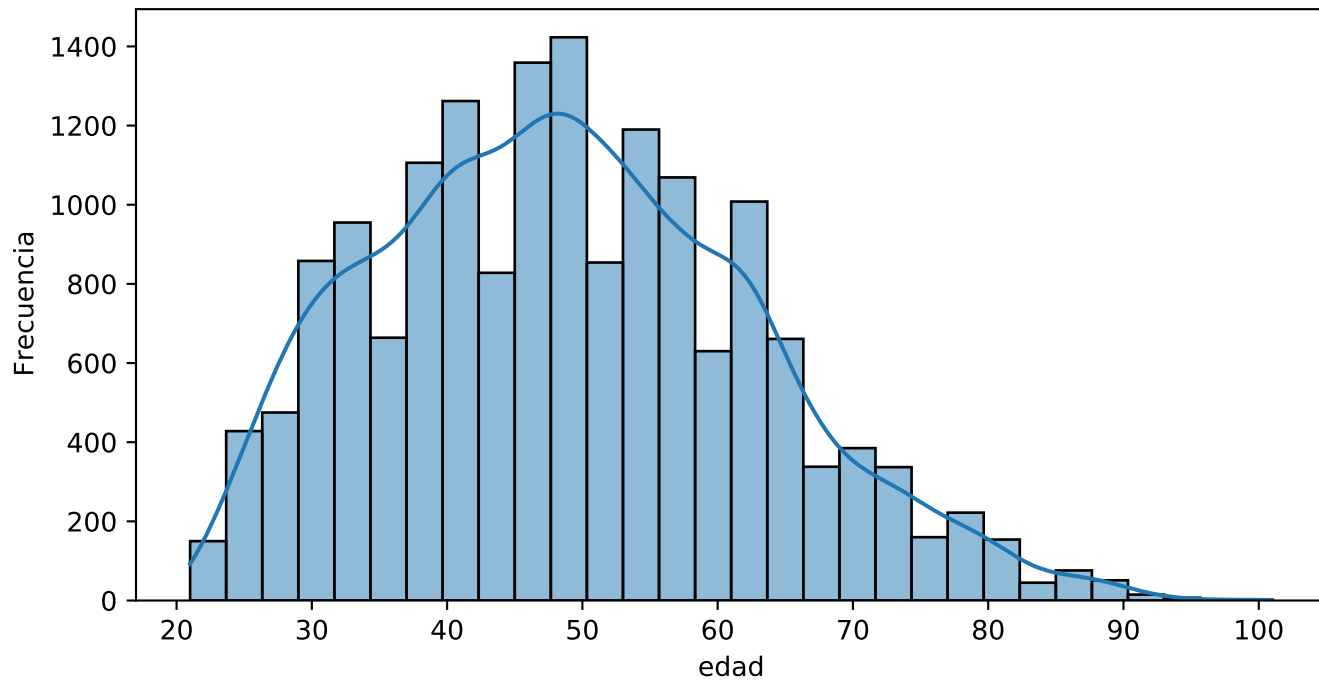
	uso_de_credito	edad	radio_deuda	...	num_dependientes	dlq_2yrs	nivel_riesgo
count	16714.000000	16714.000000	16714.000000	...	16714.000000	16714.000000	16714.000000
mean	4.799862	48.798672	30.980298	...	0.944358	0.500000	2.044334
std	204.062345	13.906078	719.694859	...	1.198791	0.500015	1.216329
min	0.000000	21.000000	0.000000	...	0.000000	0.000000	0.000000
25%	0.082397	38.000000	0.155971	...	0.000000	0.000000	1.000000
50%	0.443080	48.000000	0.322299	...	0.000000	0.500000	3.000000
75%	0.926637	58.000000	0.533426	...	2.000000	1.000000	3.000000
max	22000.000000	101.000000	61106.500000	...	8.000000	1.000000	3.000000

[8 rows x 9 columns]

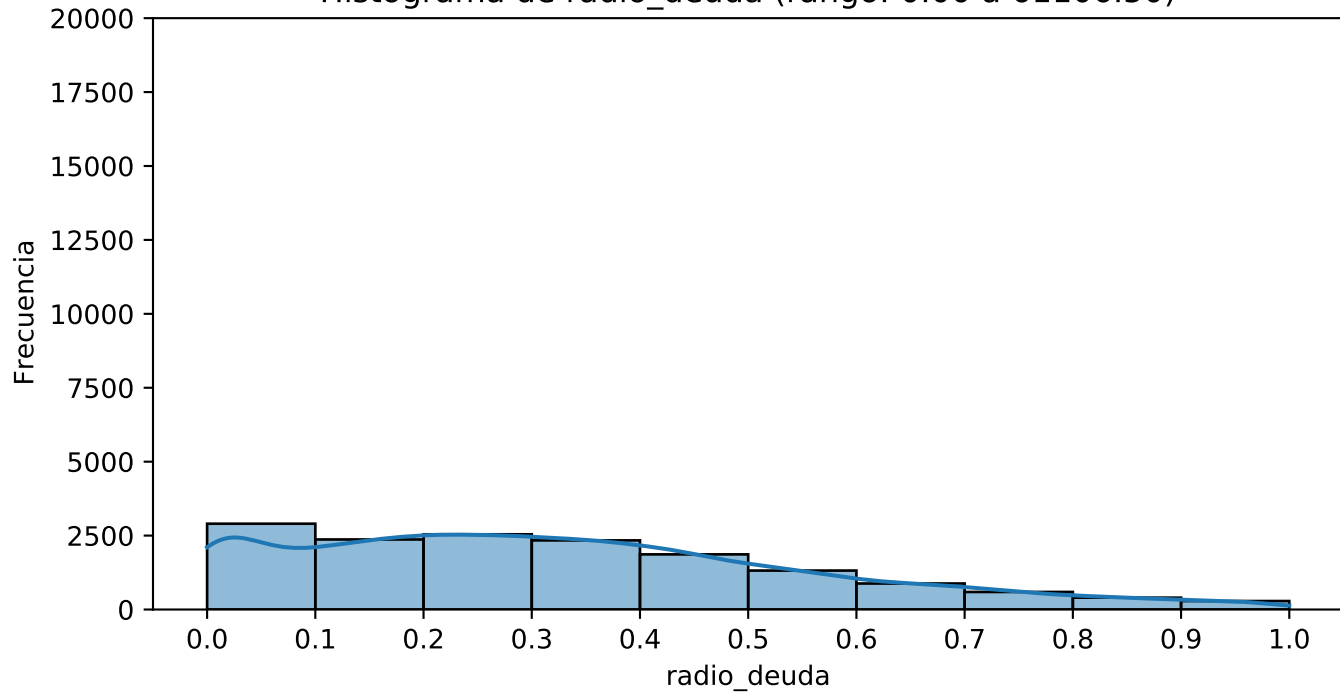
Histograma de uso_de_credito (rango: 0.00 a 22000.00)



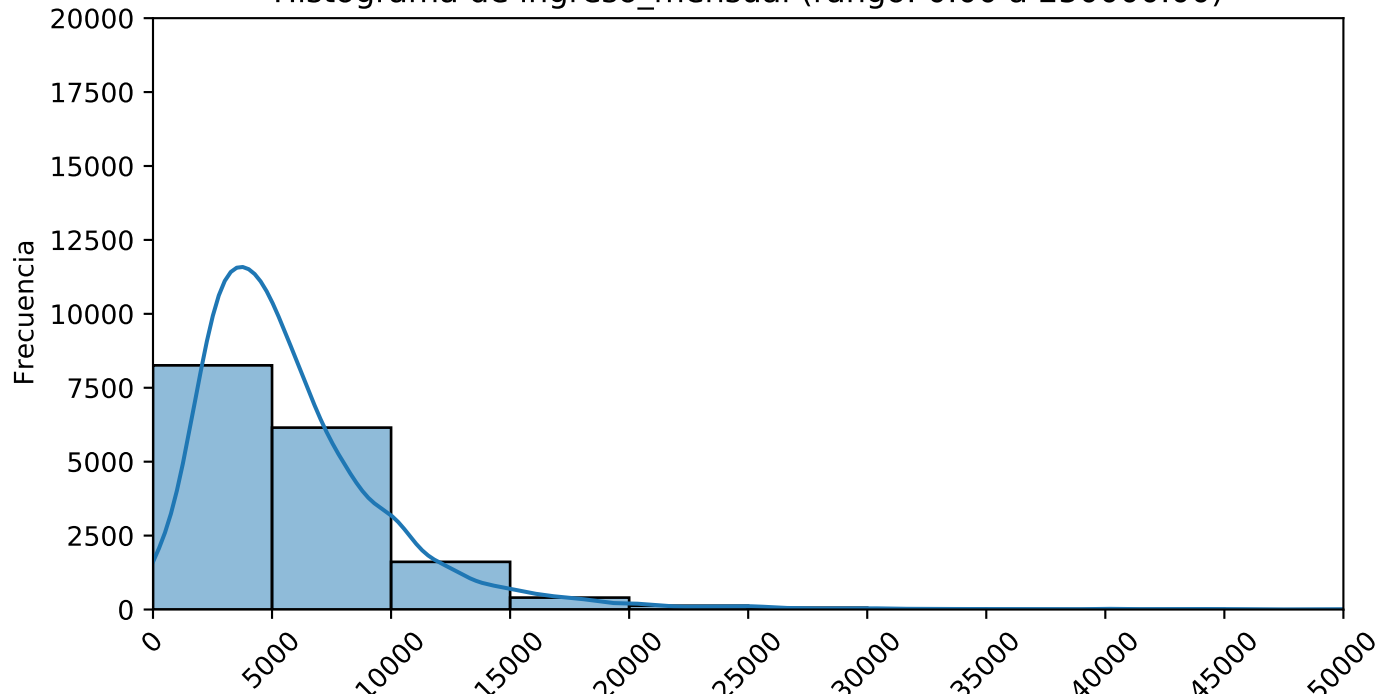
Histograma de edad (rango: 21.00 a 101.00)



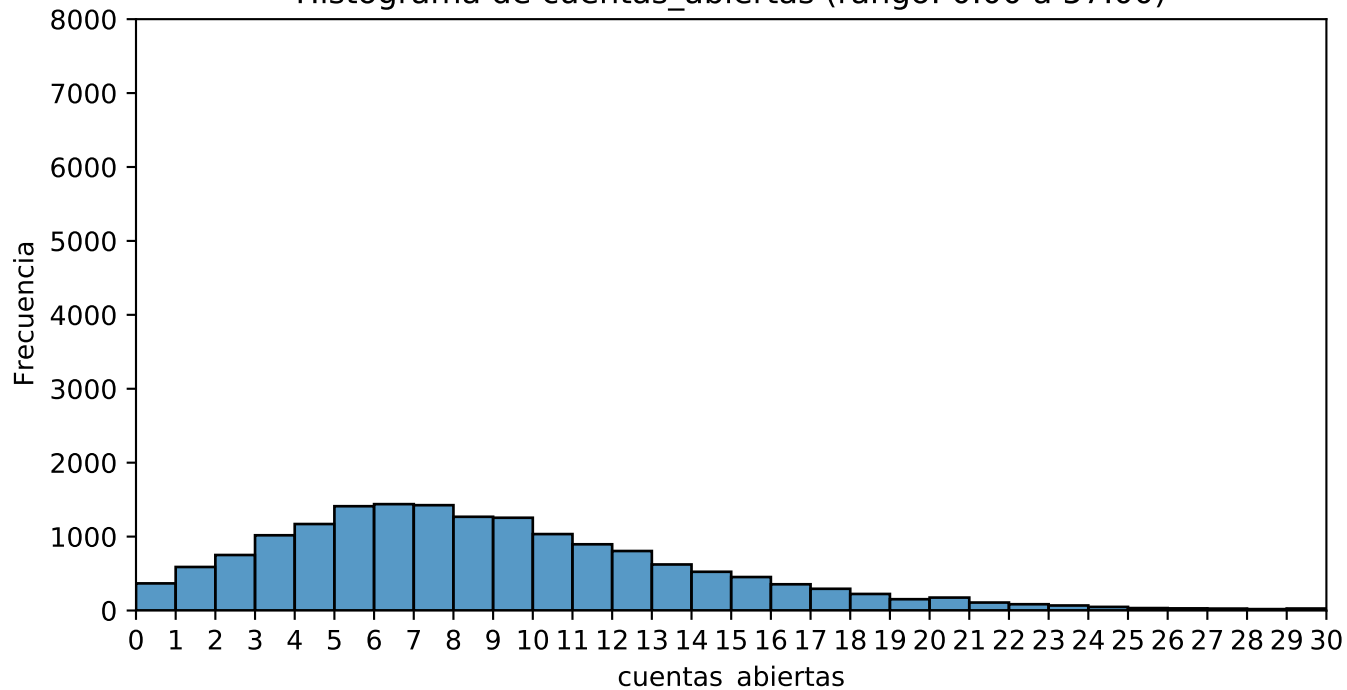
Histograma de radio_deuda (rango: 0.00 a 61106.50)



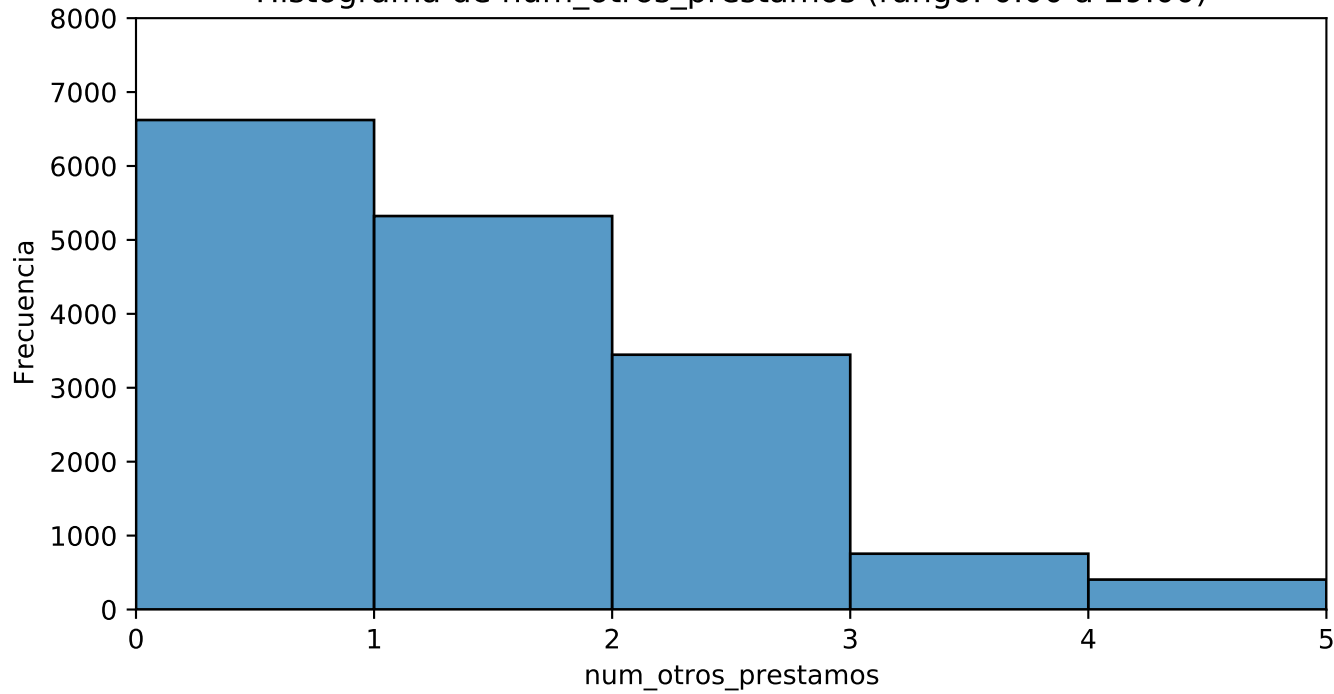
Histograma de ingreso_mensual (rango: 0.00 a 250000.00)



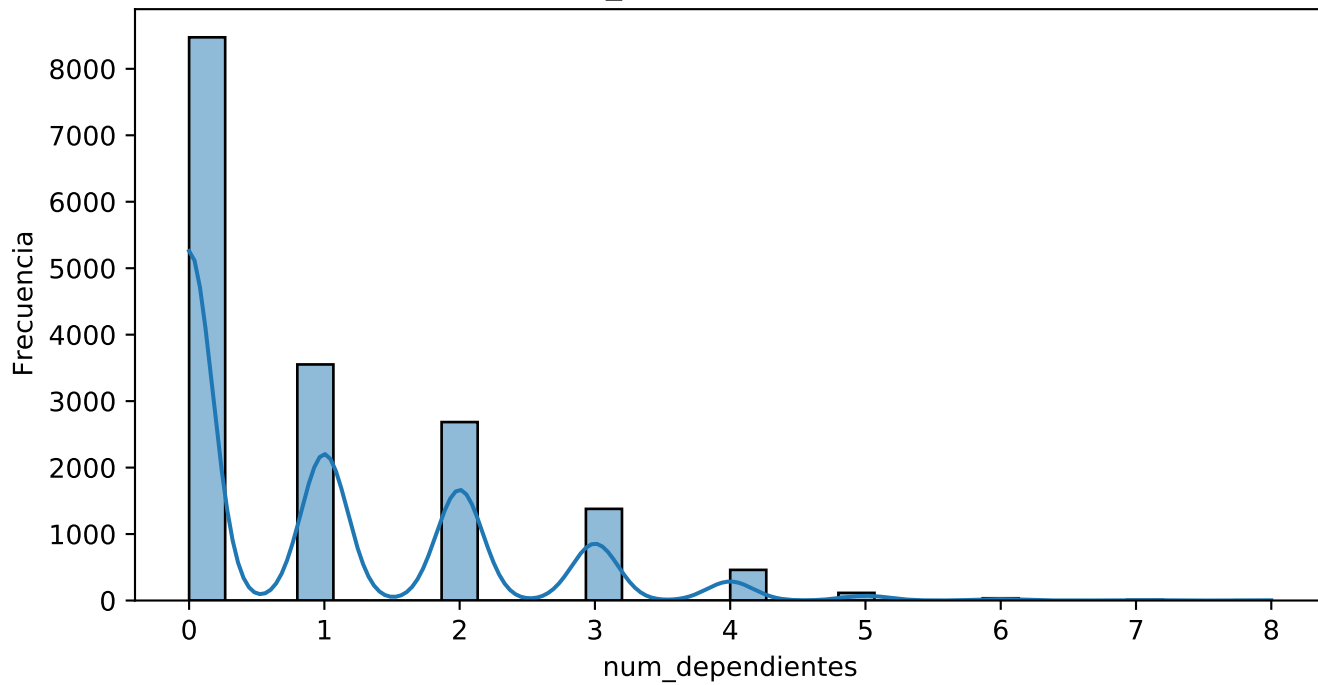
Histograma de cuentas_abiertas (rango: 0.00 a 57.00)



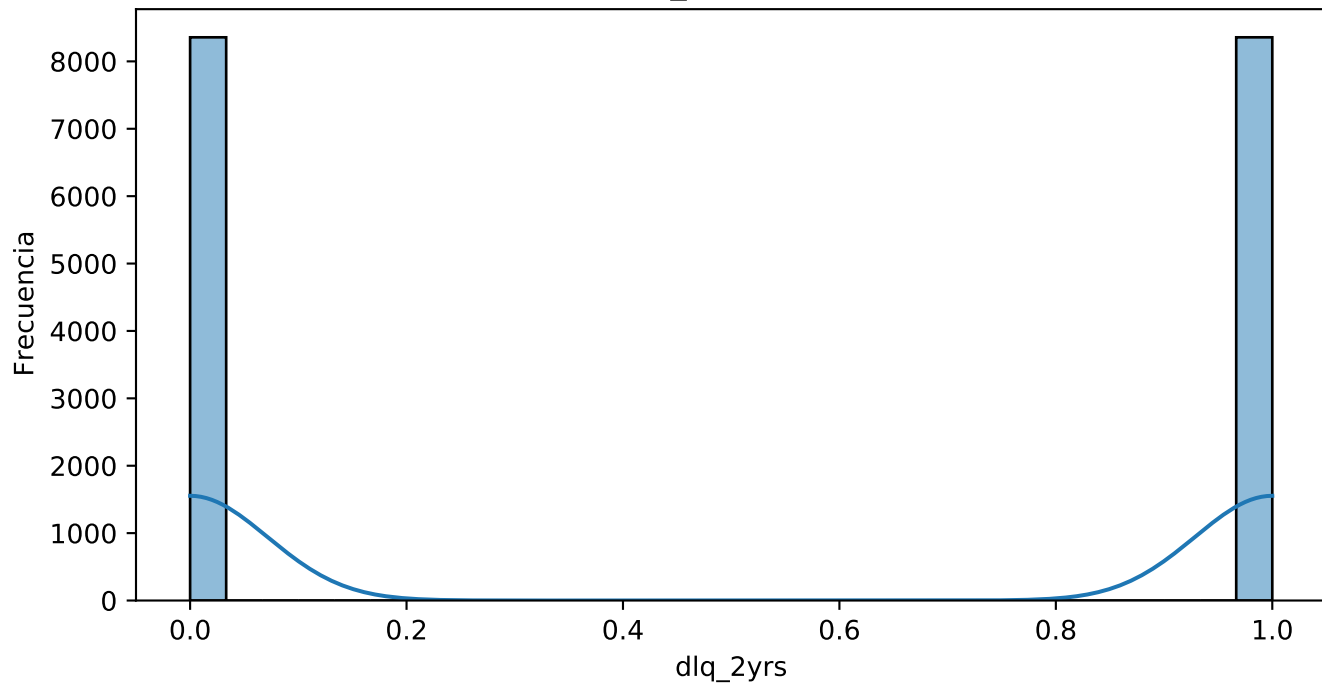
Histograma de num_otros_prestamos (rango: 0.00 a 29.00)



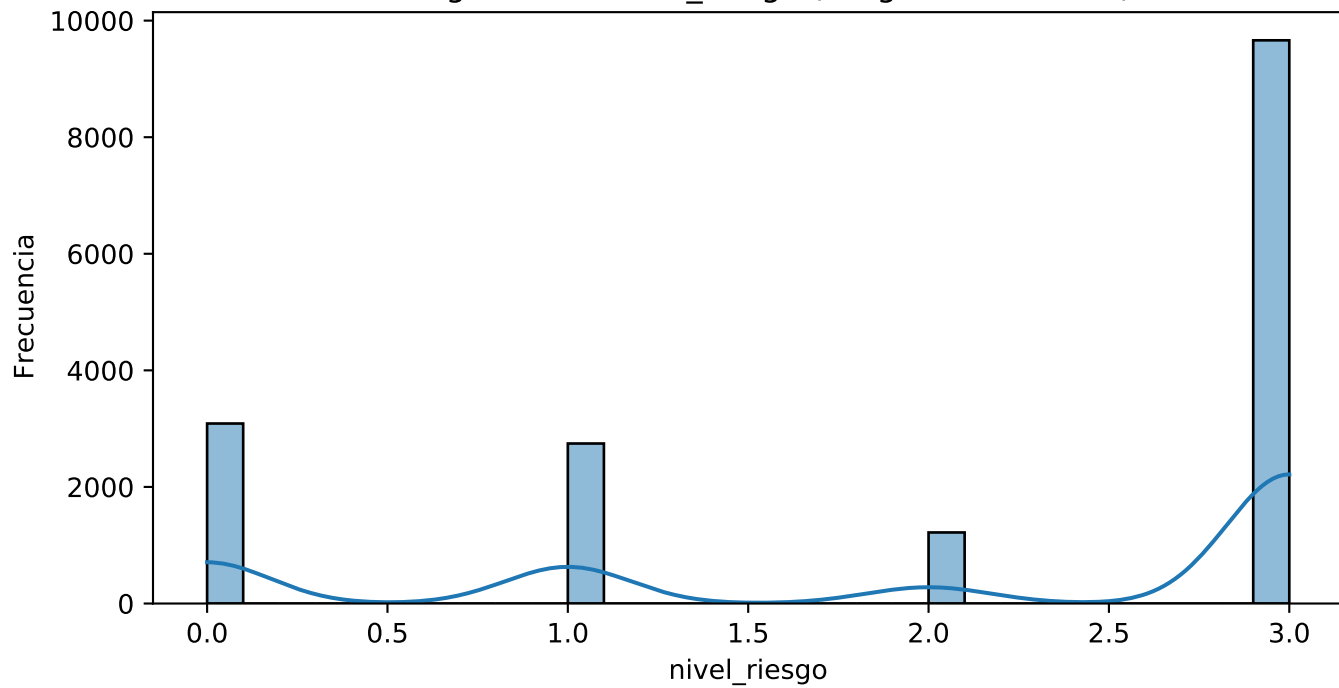
Histograma de num_dependientes (rango: 0.00 a 8.00)



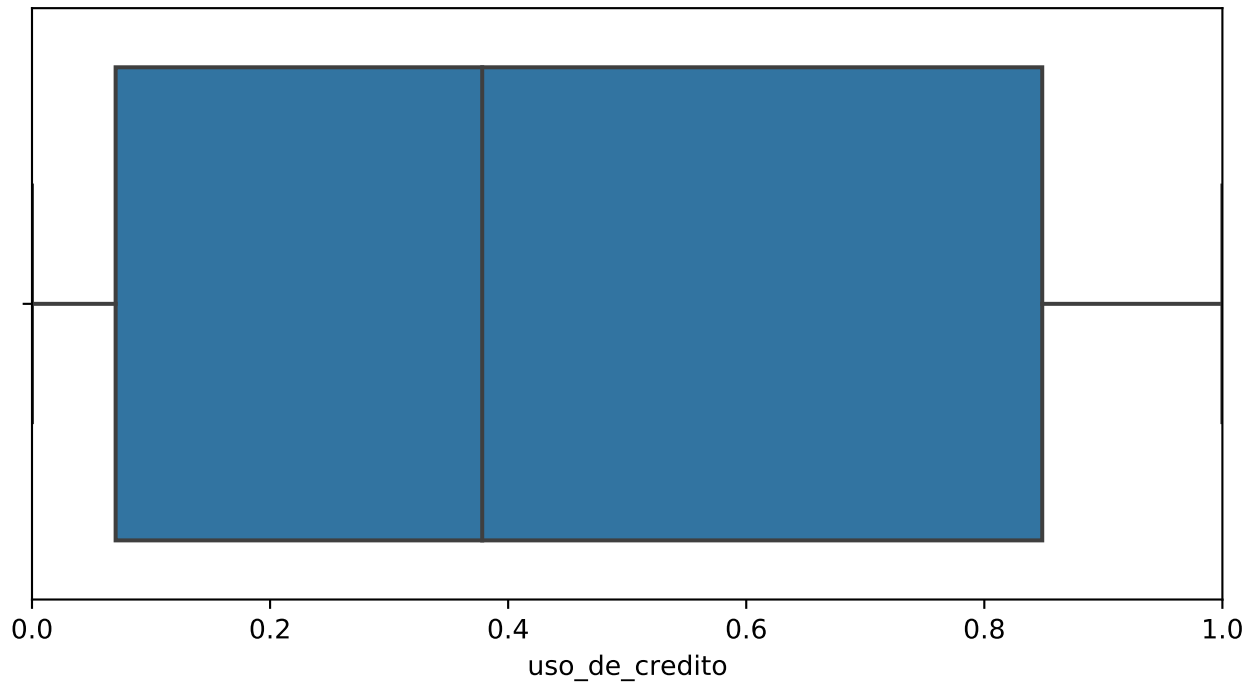
Histograma de dlq_2yrs (rango: 0.00 a 1.00)



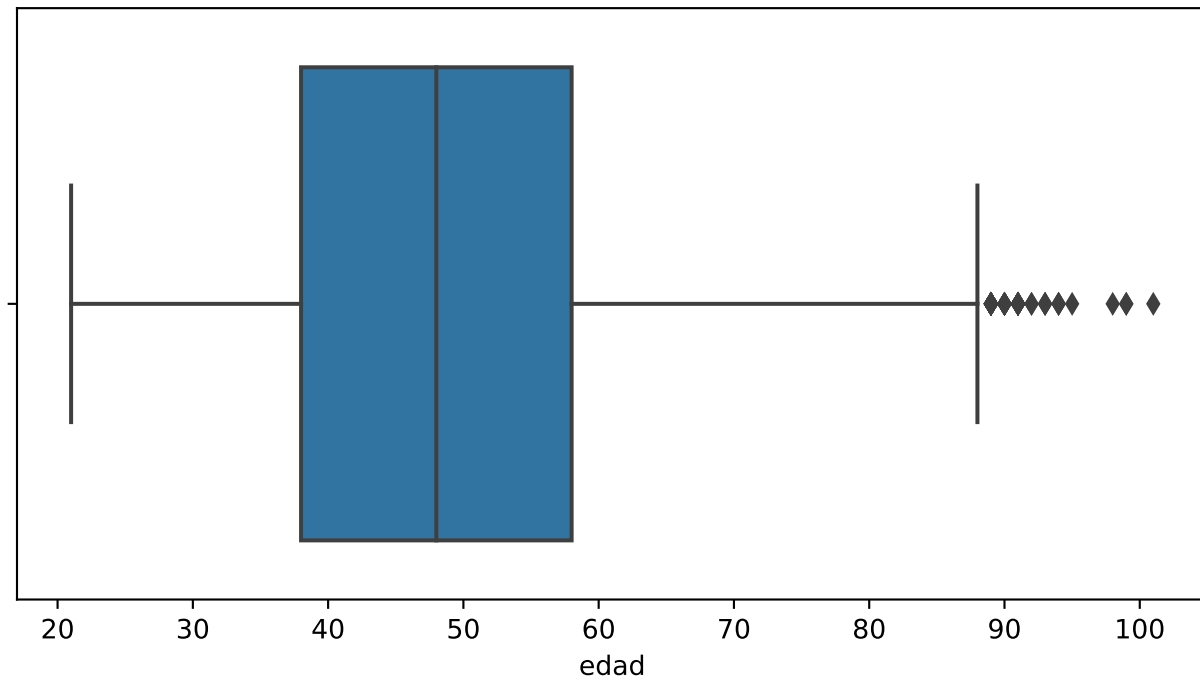
Histograma de nivel_riesgo (rango: 0.00 a 3.00)



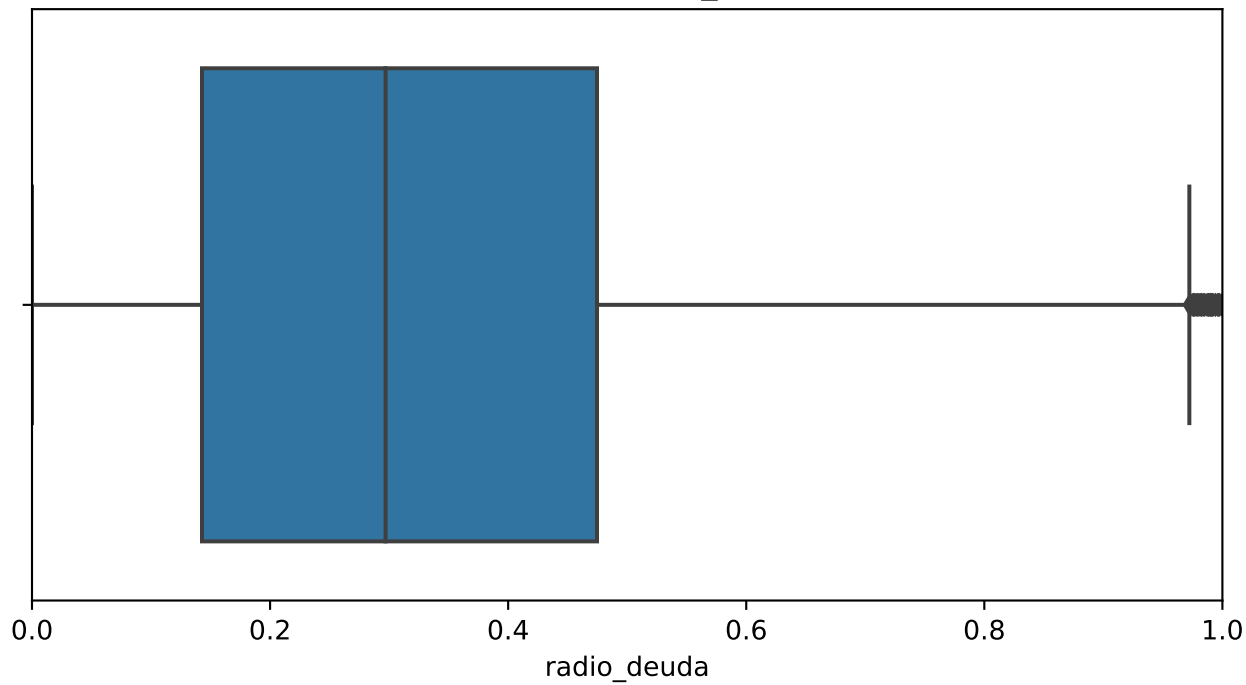
Boxplot de uso_de_credito



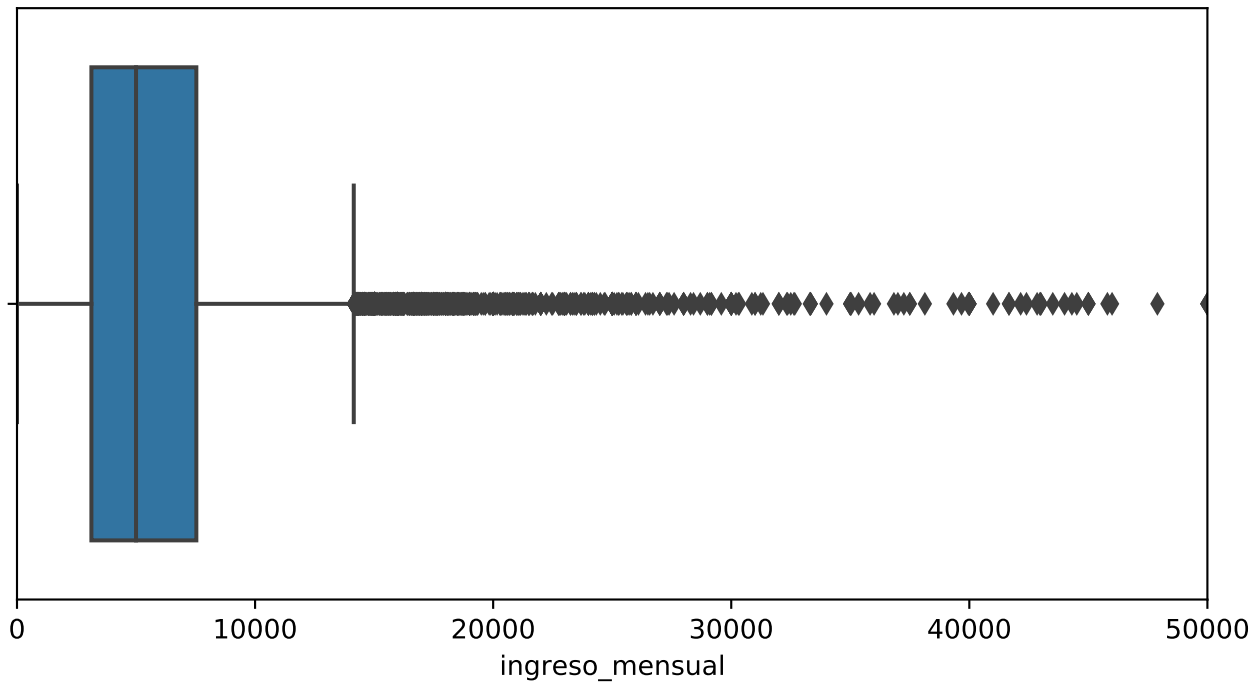
Boxplot de edad



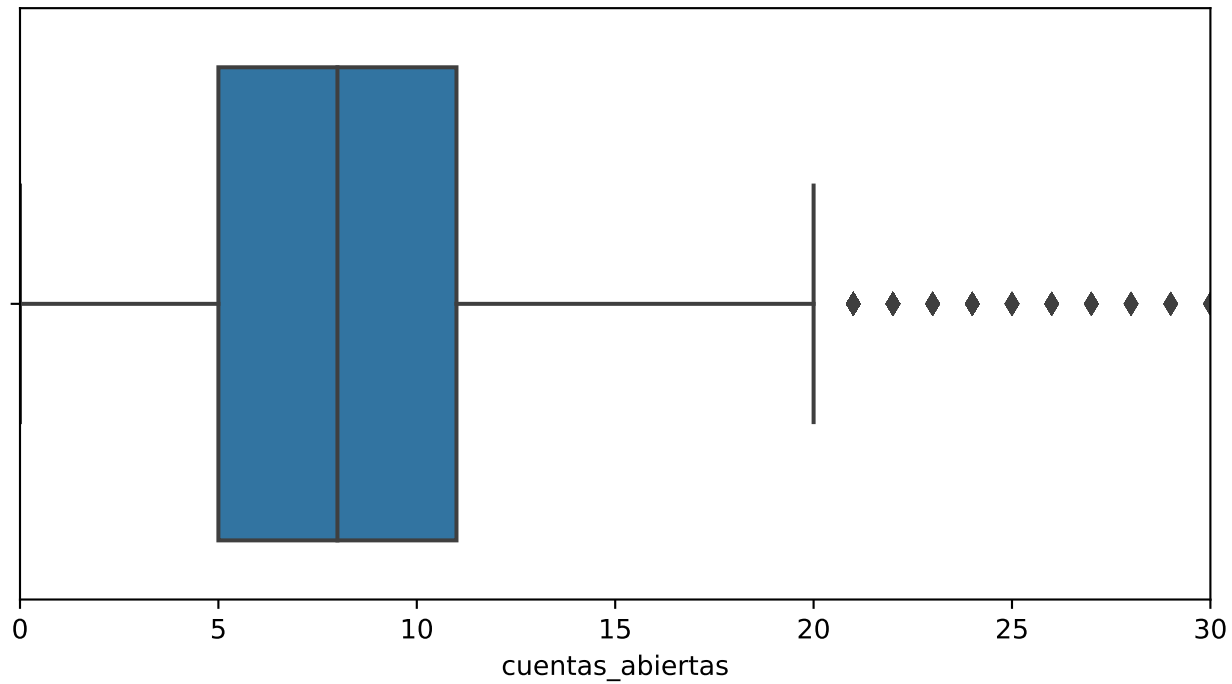
Boxplot de radio_deuda



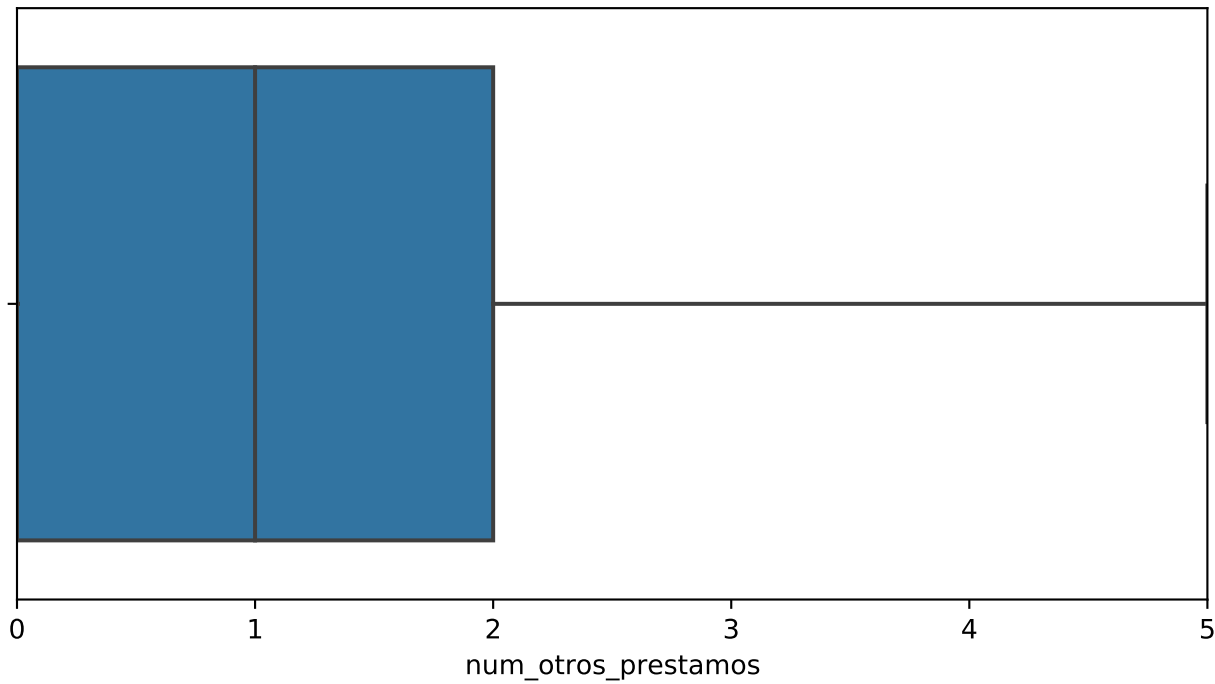
Boxplot de ingreso_mensual



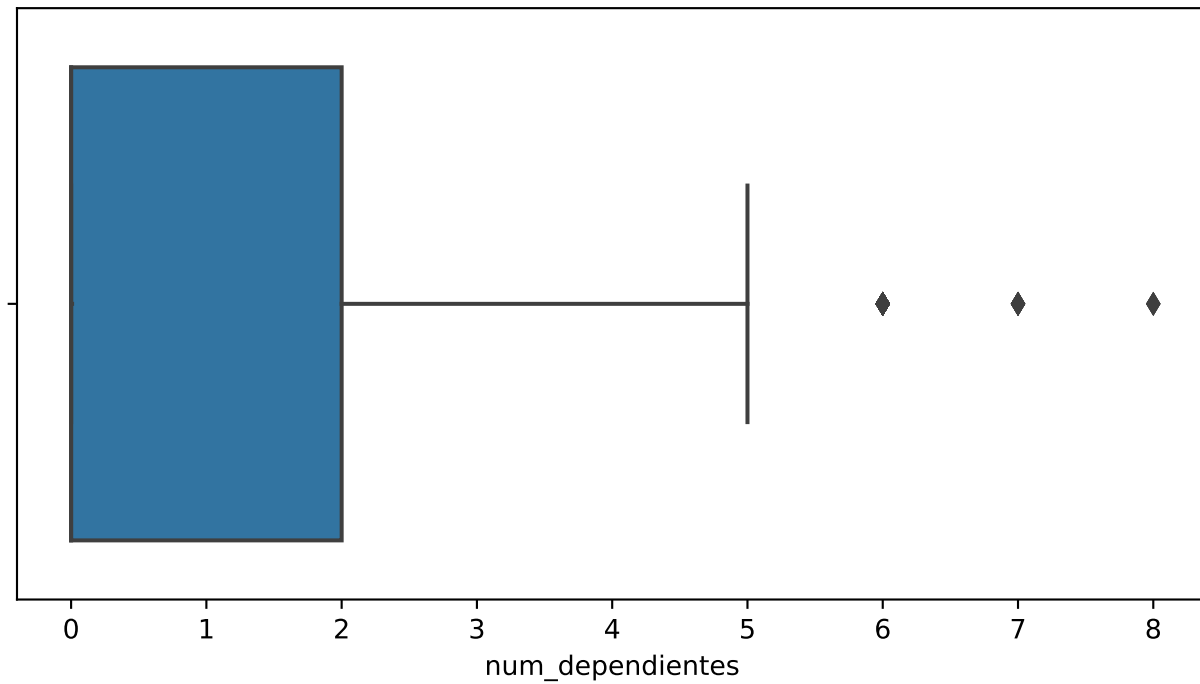
Boxplot de cuentas_abiertas



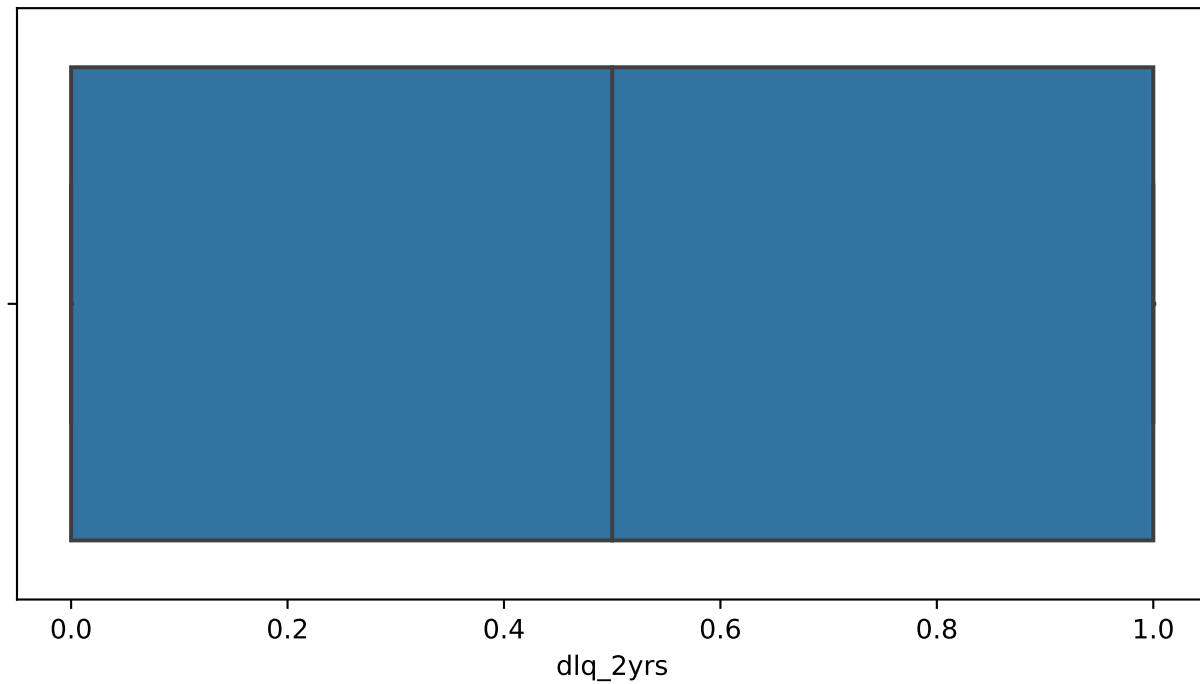
Boxplot de num_otros_prestamos



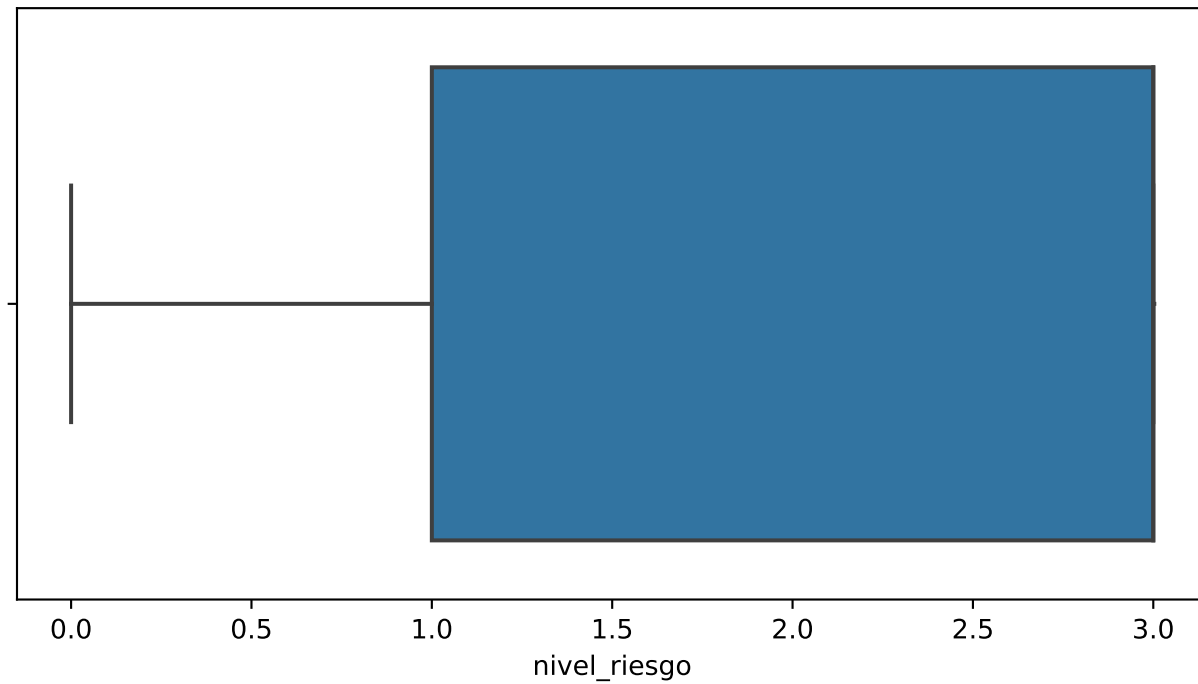
Boxplot de num_dependientes



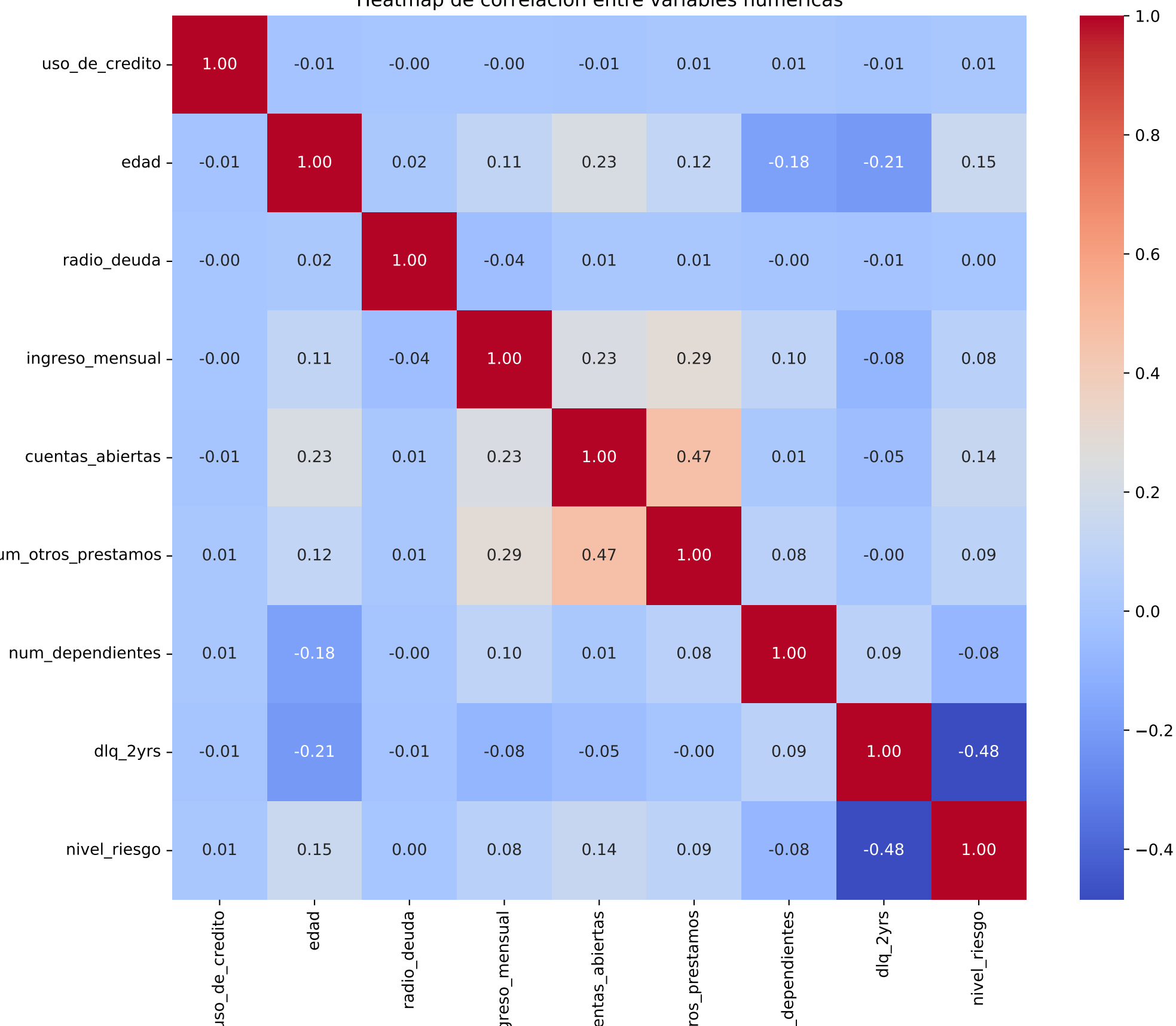
Boxplot de dlq_2yrs



Boxplot de nivel_riesgo

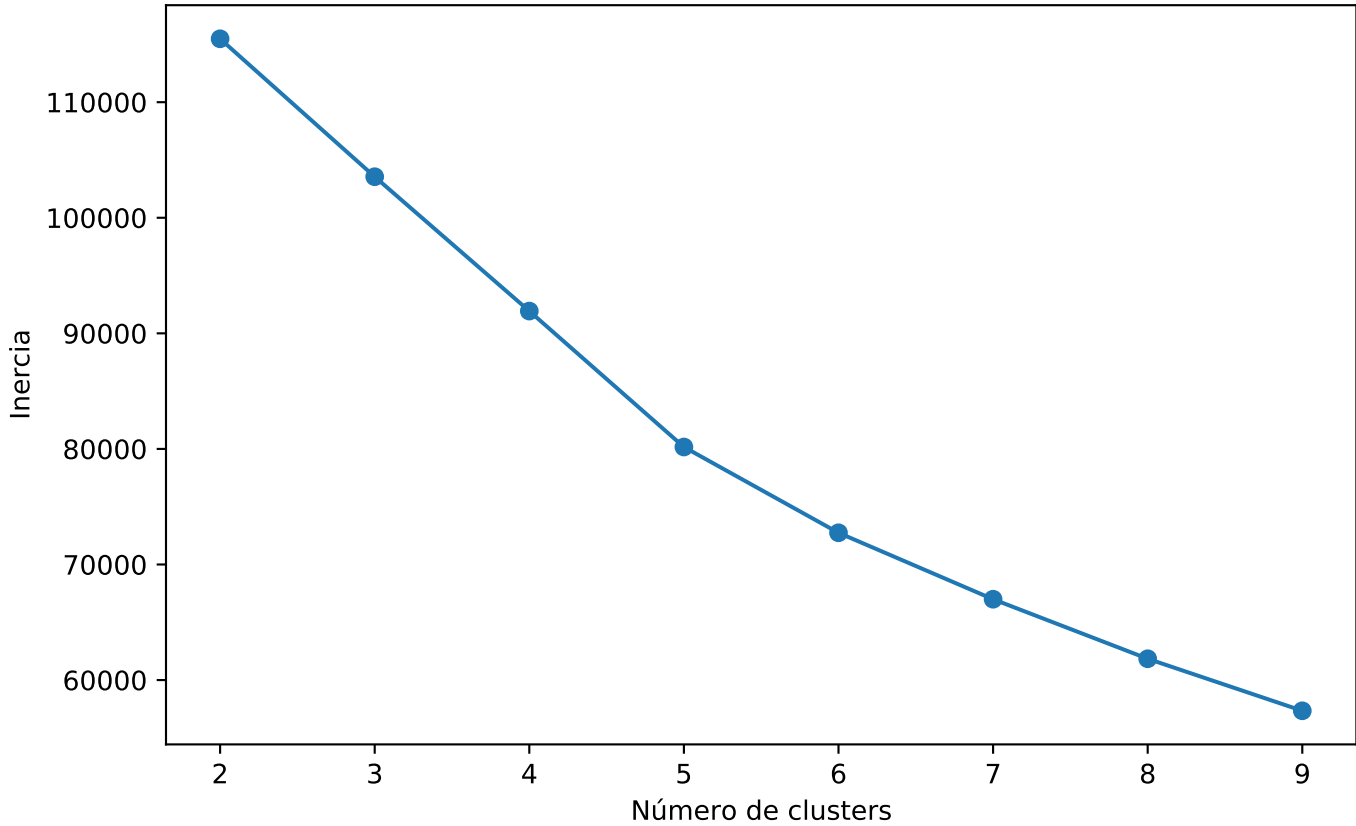


Heatmap de correlación entre variables numéricas



Manejo de modelos

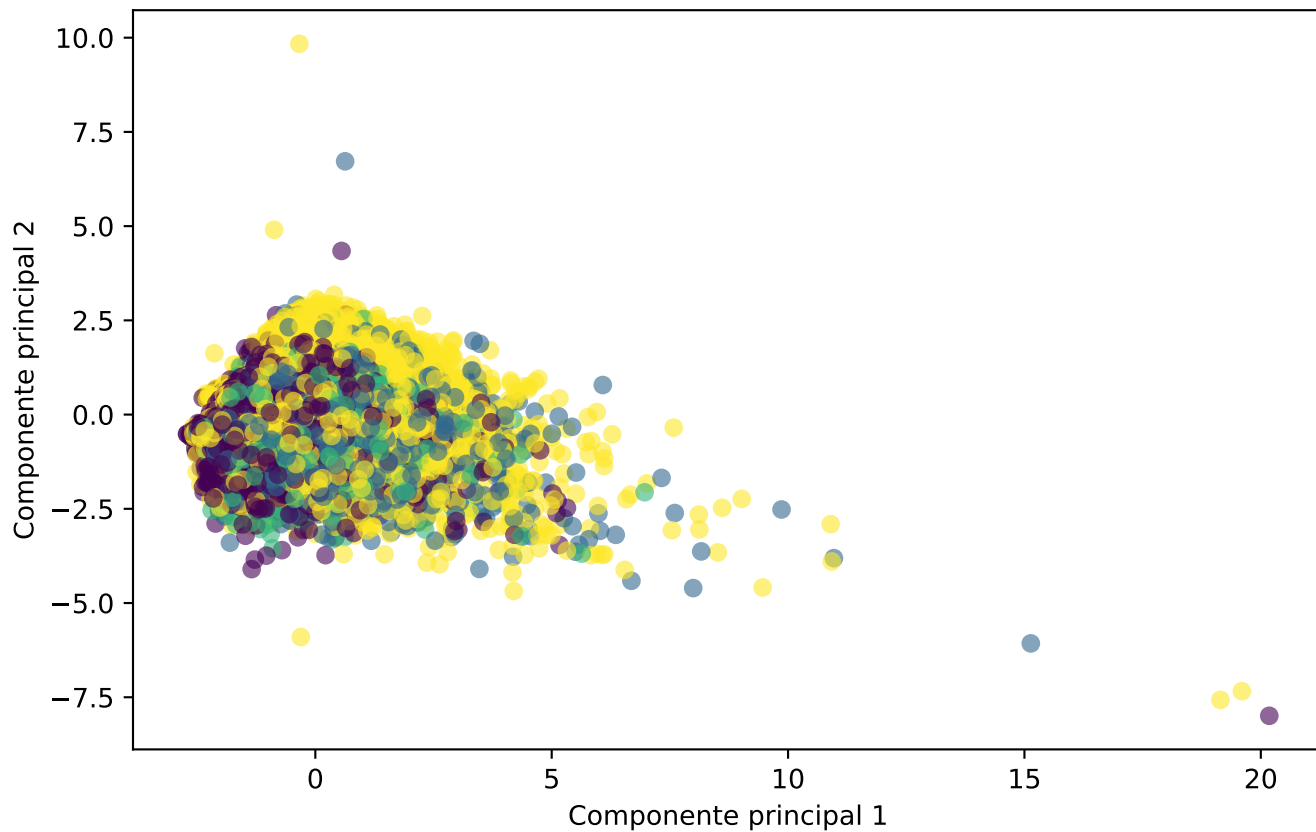
Método del codo



Interpretación del método del codo:

El punto donde la curva deja de bajar abruptamente sugiere el número óptimo de clusters.

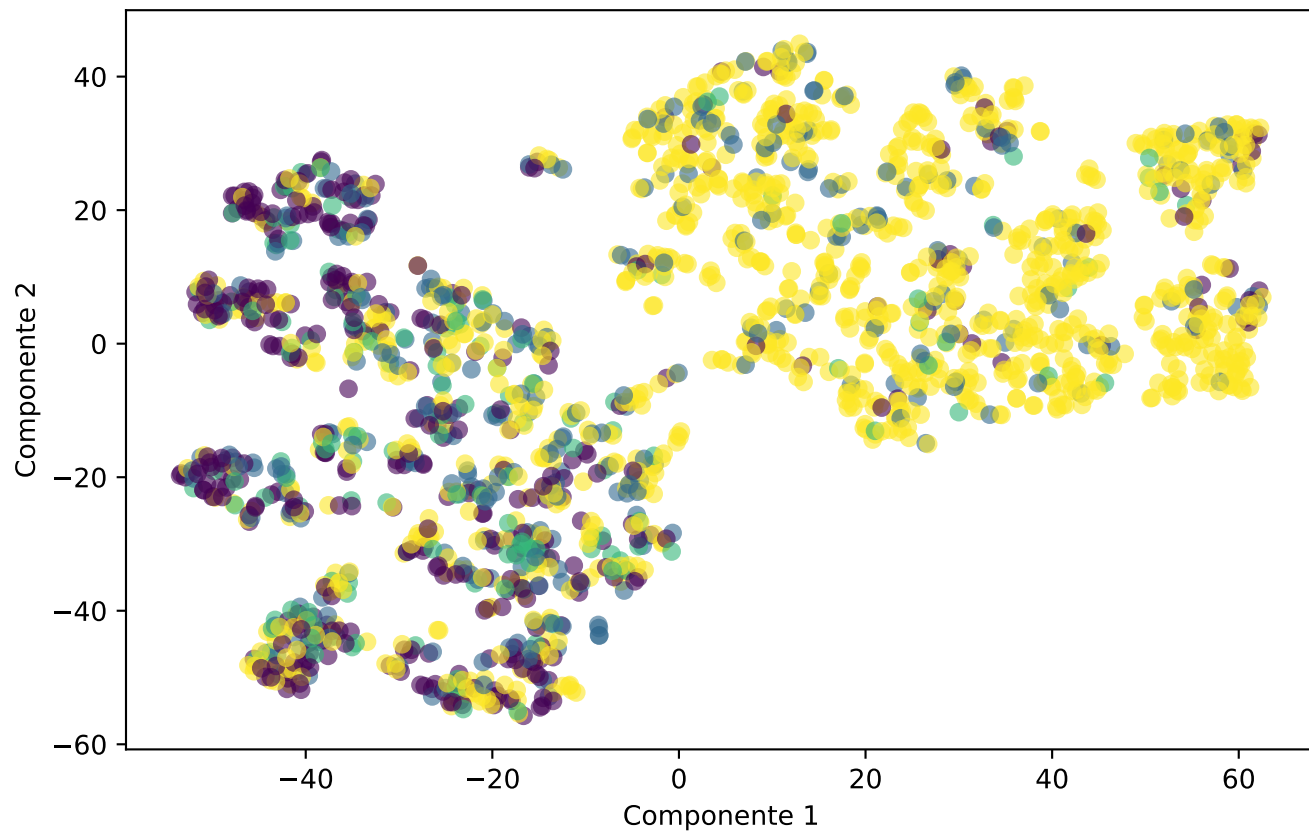
Reducción de dimensionalidad con PCA



Interpretación de PCA:

Observa si los grupos de riesgo se separan visualmente en el espacio reducido.

Reducción de dimensionalidad con t-SNE



Interpretación de t-SNE:

Permite visualizar agrupamientos no lineales. Compara con PCA.

Regresión Lineal Simple (edad -> ingreso_mensual): $R^2 = 0.013$
Regresión Lineal Múltiple (todas -> ingreso_mensual): $R^2 = 0.116$
Regresión Logística: No aplica
LDA: 0.612
KNN: 0.694
KMeans Silhouette: 0.214
K-Medoids Silhouette: 0.190
DBSCAN Silhouette: -0.01196582470200573
GMM Silhouette: 0.215
Mejor modelo según silhouette: GMM (score: 0.215)

Interpretación de Regresión Lineal Simple:

ca (ej. ingreso mensual) usando solo una variable (ej. edad). El R^2 indica qué tan bien la variabl

Interpretación de Regresión Lineal Múltiple:

érica usando varias variables al mismo tiempo. El R^2 muestra la proporción de la variabilidad ex

Interpretación de Regresión Logística:

Entre dos categorías (por ejemplo, riesgo alto/bajo). El score indica la proporción de aciertos en l

Interpretación de LDA (Análisis Discriminante Lineal):

clases (niveles de riesgo) encontrando combinaciones lineales de variables. El score es la preci

Interpretación de K-Vecinos más Cercanos (KNN):

registro según los 'vecinos' más cercanos en el espacio de variables. El score es la precisión de

Interpretación de K-Means:

clusters según similitud. El silhouette score mide qué tan bien están separados los grupos (más

Interpretación de K-Medoids:

Similar a K-Means pero más robusto a valores atípicos. El silhouette score tiene la misma interpretación.

Interpretación de DBSCAN:

densidad, detectando grupos de cualquier forma y valores atípicos. El silhouette score mide la

Interpretación de GMM (Gaussian Mixture Model):

asumiendo que cada grupo sigue una distribución normal. El silhouette score mide la separación

Conclusiones:

El modelo con mejor desempeño de acuerdo al silhouette score fue: GMM (score: 0.215).

Se recomienda analizar más a fondo las variables que más influyen en la segmentación y considerar la recolección de más datos si es posible. para mejorar la prediccion
Además, los resultados muestran el desempeño de todos los modelos supervisados y no supervisados aplicados al dataset.