

DICCIONARIO DE DATOS

Base RCC - Reporte de Central de Riesgos

Autor: Josef Renato Rodríguez Mallma

Fecha: 20/02/2026

Versión: 1.0

Documento técnico académico / profesional

1. Alcance del Documento

El presente documento describe la estructura, definición y calidad de las variables contenidas en la base de datos RCC. Su finalidad es estandarizar la interpretación de los campos utilizados para análisis estadístico, segmentación y modelamiento predictivo, asegurando consistencia metodológica y gobernanza de datos.

2. Diccionario de Variables

Variable	Tipo	Descripción	Notas / Calidad
id_cliente	String	Identificador único de cliente (ej. C000123)	Único por cliente; 1,500 clientes.
periodo	String (YYYYMM)	Periodo de observación mensual (24 meses, 2 años de historia)	Se usa para splits temporales (Train/Test/OOT).
edad	Numérica (int)	Edad del cliente en años (18-75)	Sin grandes outliers.
sexo	Categórica	Sexo del cliente (M/F)	Puede contener valores nulos.
region	Categórica	Región geográfica: Lima, Centro, Norte, Sur, Oriente	Tiene valores con espacios extras y nulos.
segmento	Categórica	Segmento comercial asignado (A, B, C)	Algunos nulos.
ingreso_mensual	Numérica (float / string sucio)	Ingreso mensual estimado del cliente	Incluye valores sucios: 'N/A', '—', 'missing'.
score_bureau	Numérica (int)	Score de buró (300-950)	Puede tener nulos.
num_lineas	Numérica (int)	Número de líneas de crédito activas	Generado por distribución Poisson.
max_mora_12m	Numérica (int)	Máxima mora registrada en últimos 12 meses	Sesgada a valores bajos; algunos nulos.
mora_30d_12m	Binaria (0/1)	Indicador: si tuvo mora ≥ 30 días en últimos 12 meses	15% aprox. positivos.
saldo_total	Numérica (float)	Saldo total de deudas vigentes	Puede contener nulos.
utilizacion_tc	Numérica (float / string sucio)	Ratio de utilización de tarjeta de crédito (0-1)	Algunos valores como 'N/A', '—'.
antig_laboral_anios	Numérica (float)	Antigüedad laboral en años	Puede tener nulos.
antig_banca_anios	Numérica (float)	Antigüedad como cliente en el banco en años	Generada con distribución normal truncada.

Variable	Tipo	Descripción	Notas / Calidad
producto_principal	Categorica	Producto principal: TC, PL, Hipotecario, Pyme	Algunos nulos, espacios extras.
canal_origen	Categorica	Canal de adquisición: Digital, Tienda, FuerzaVenta, Alianza	Algunos nulos, espacios extras.
mora_30d_next3m	Binaria (0/1)	Target: entrará en mora ≥ 30 días en próximos 3 meses	Desbalanceado (~17.7% positivos).