

Mineria de Datos

Análisis Exploratorio de Datos y Predicción de Incumplimiento de Préstamos

Aina Llaneras Casas, Alejandro Arcas Alberti, Alessandro Natali Vilamú,
Berta Moyano Núñez, Blanca Romero Sainz, Iker Meneses Sales, Ismael
Argemí Fernández, Iván Martínez Yates, Marta Gomez de la Tia Privat, Mireia
Bohils Tenas, Mireia Bolívar Rubia, Oscar Arroyo Luque, Letizia

GRUPO 1: Aina Llaneras, Blanca Romero, Iván Martínez
GRUPO 2: Alejandro Arcas Alberti, Alessandro Natali, Iker Meneses, Letizia
GRUPO 3: Ismael Argemí, Mireia Bohils, Oscar Arroyo
GRUPO 4: Berta Moyano, Marta Gómez, Mireia Bolívar

23 de Septiembre del 2023

Definición del proyecto y asignación

Fuente de obtención de los datos

Los datos se han extraído del repositorio de bases de datos Kaggle. El enlace de la página web es el siguiente: https://www.kaggle.com/datasets/gauravduttakiit/loan-defaulter?select=application_data.csv

Descripción de los datos

Esta base de datos está diseñada para abordar el desafío de identificar posibles incumplimientos de préstamos en un entorno empresarial real. El conjunto de datos contiene información relacionada con préstamos otorgados a diversos prestatarios, junto con detalles financieros y personales de los solicitantes. El objetivo principal de esta base de datos es permitir a las instituciones financieras o analistas de riesgos realizar un análisis exploratorio de datos completo para evaluar la probabilidad de que un prestatario incumpla con sus obligaciones financieras.

Estructura e información de la matriz de datos

Filas (individuos)	Columnas (variables)	Nro. variables numéricas	Nro. variables categóricas	Nro. variables respuesta u objetivo
5000	15	7	8	1

VARIABLES EXPLICATIVAS

Nombre	Descripción	Tipo	Diccionario y dominio
CODE_GENDER	Género del cliente	Categórica	M—Male, F—Female
NAME__INCOME__TYPE	Tipo de ingresos	Categórica	1-Businessman, 2-Commercial associate, 3-Pensioner, 4-State servant, 5-Working
NAME__EDUCATION__TYPE	Nivel de estudios del cliente	Categórica	1-Academic degree, 2-Higher education, 3-Incomplete higher, 4-Lower secondary, 5-Secondary special
NAME__FAMILY__STATUS	Estado civil	Categórica	1-Married, 2-Single/not married, 3-Civil marriage, 4-Separated, 5-Widow
OCCUPATION__TYPE	Actividad laboral	Categórica	1-Laborers, 2-Sales staff, 3-Core staff, 4-Managers, 5-Drivers, 6-Accountants, 7-Cleaning staff, 8- High skill tech staff, 9-HR staff, 10-IT staff, 11-Cooking staff, 12-Low-skill Laborers, 13-Medicine sraff, 14-Private service staff, 15-Realty agents, 16-Security staff, 17-Secretaries, 18-Waiters/barmen staff

Nombre	Descripción	Tipo	Diccionario y dominio
ORGANIZATION _TYPE	Tipo de organización donde trabaja el cliente	Categórica	1-Advertising, 2-Agriculture, 3-Bank, 4-Business Entity Type 1, 5-Business Entity Type 2, 6-Business Entity Type 3, 7-Cleaning, 8-Construction, 9-Culture, 10-Electricity, 11-Emergency, 12-Government, 13-Hotel, 14-Housing, 15-Industry: type 1, 16-Industry: type 10, 17-Industry: type 11, 18-Industry: type 12, 19-Industry: type 13, 20-Industry: type 2, 21-Industry: type 3, 22-Industry: type 4, 23-Industry: type 5, 24-Industry: type 6, 25-Industry: type 7, 26-Industry: type 9, 27-Insurance, 28-Kindergarten, 29-Legal Services, 30-Medicine, 31-Military, 32-Mobile, 33-Other, 34-Police, 35-Postal, 36-Realtor, 37-Restaurant, 38-School, 39-Security, 40-Security Ministries, 41-Self-employed, 42-Services, 43-Telecom, 44-Trade: type 1, 45-Trade: type 2, 46-Trade: type 3, 47-Trade: type 4, 48-Trade: type 6, 49-Trade: type 7, 50-Transport: type 1, 51-Transport: type 2, 52-Transport: type 3, 53-Transport: type 4, 54-University, 55-XNA 1, 2, 3
REGION_RATING _CLIENT	Nuestra calificación de la región donde vive el cliente	Categórica	
AMT_INCOME _TOTAL	Ingresos totales del cliente	Numérica	[29250,2250000]
AMT_CREDIT	Importe de crédito del préstamo	Numérica	[45000,3375000]
AMT_ANNUITY	Anualidad del préstamo	Numérica	[2673,177827]
DAYS_BIRTH	Edad del cliente en número de días en el momento de pedir el préstamo	Numérica	[-25159,-7711]
OWN_CAR_AGE	Edad en años del coche del cliente	Numérica	[0,65]

Nombre	Descripción	Tipo	Diccionario y dominio
AMT_GOODS_PRICE	Para préstamos al consumo, es el precio de los bienes para los cuales se otorga el préstamo	Numérica	[45000,3375000]
CNT_FAM_MEMBERS	Número de familiares del cliente	Numérica	[1,8]

VARIABLE OUTPUT

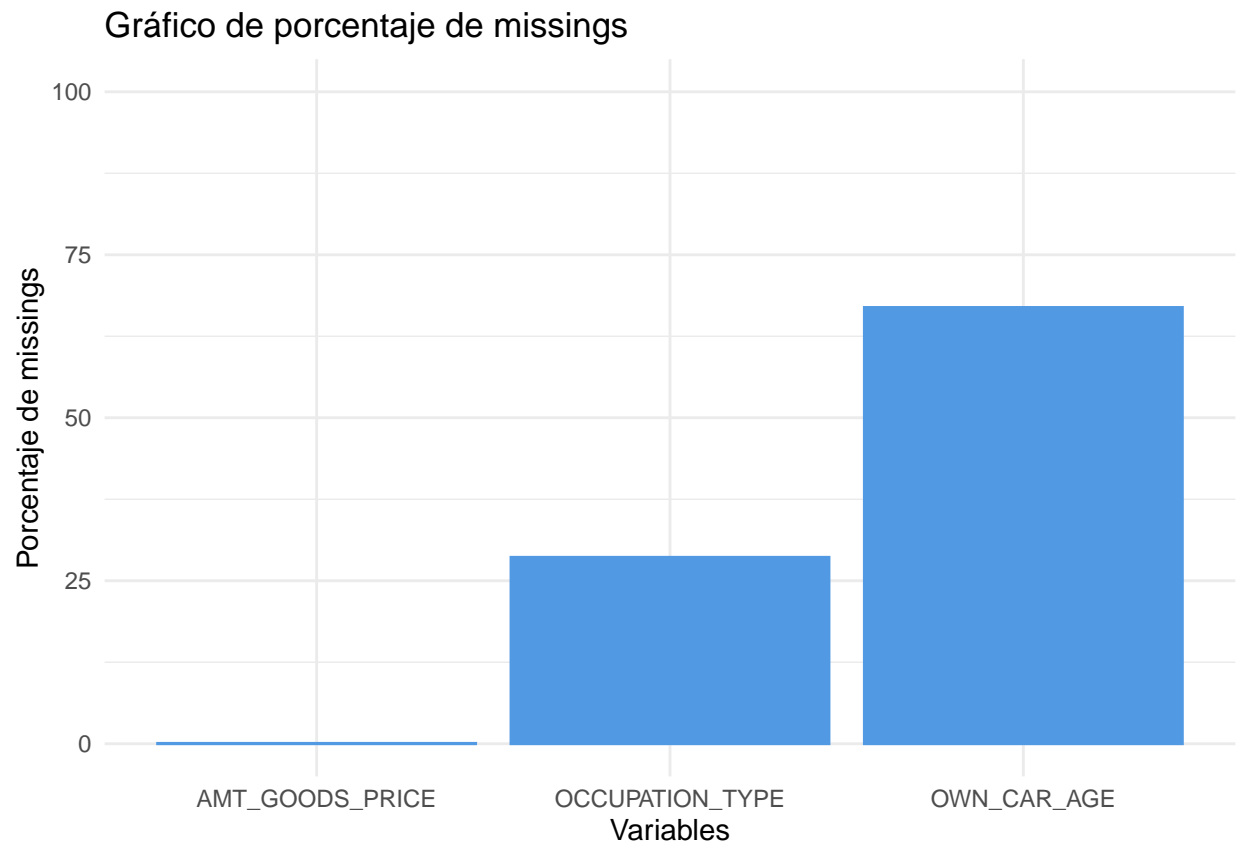
Nombre	Descripción	Tipo	Diccionario y dominio
Target	Target	Categórica	1 - Cliente con dificultades de pago: él/ella tuvo pagos atrasados de más de X días en al menos una de las primeras Y cuotas del préstamo en nuestra muestra. 0 - Todos los demás casos

VARIABLES MISSING

Nro. de casillas missings	Respeto del total de la matriz datos
4779	6.37 %

Porcentaje de missings por variable (tabla y histograma):

	Nro. de missings	Porcentaje de missings
OWN_CAR_AGE	3346	66.92 %
AMT_GOODS_PRICE	3	0.06 %
OCCUPATION_TYPE	1430	28.6 %



Únicamente se han representado las variables que tienen algún valor faltante.