



## DESARROLLO DE UN MODELO DE RIESGO DE CRÉDITO

Helga Zambrana

# Agenda

- 1. Presentación del problema**
- 2. Objetivo de la investigación**
- 3. Descripción del dataset elegido**
- 4. Hallazgos del análisis exploratorio de datos (EDA)**

# 1 Presentación del problema



Un riesgo de crédito es la probabilidad de pérdida debido al incumplimiento en los pagos de quien solicita un préstamo.

Las instituciones financieras tienen la necesidad de gestionar el riesgo de crédito para mitigar las pérdidas. El objetivo es automatizar la aprobación de un préstamo, identificar los indicadores que diferencian los solicitantes con tendencia a pagar sus deudas de quienes no la pagarán y facilitar las aprobaciones de créditos.

## 2 Objetivo de la investigación



Analizar la información disponible de los aplicantes a un préstamo e identificar las posibles variables que determinen que el solicitante no pagará el préstamo



Analizar diferentes opciones de modelos de clasificación y seleccionar el que tenga la mejor performance en predecir si el solicitante pagará o no el préstamo

### 3 Descripción del dataset elegido



El dataset elegido fue [Home Equity \(HMEQ\)](#) de Kaggle.

HMEQ presenta información sobre las características y la morosidad de 5960 préstamos con garantía hipotecaria. Un préstamo con garantía hipotecaria es un préstamo en el que el deudor utiliza el capital de su vivienda como garantía subyacente.

Los criterios de selección del dataset fueron la claridad de los datos recolectados y su robustez, lo que facilita el análisis, el procesamiento y la generación de un modelo.

## 3 Descripción del dataset elegido

Nombre de la variable	Descripción
BAD	Categoría de la persona candidata al préstamo 1, corresponde al candidato con préstamo incumplido o con mora 0, corresponde al candidato con préstamo al día
LOAN	Monto del préstamo solicitado
MORTDUE	Monto adeudado de la hipoteca existente
VALUE	Valor de la propiedad hipotecada
REASON	Razón para solicitar el préstamo DebtCon, corresponde a consolidación de deudas; HomeImp, corresponde a mejoras para el hogar
JOB	Categoría profesional
YOJ	Cantidad de años en el trabajo actual

# 3 Descripción del dataset elegido

Nombre de la variable	Descripción
DEROG	Número de informes derogatorios importantes. Se refiere a la información proporcionada por una institución financiera a las agencias de crédito y se relaciona con la morosidad o la cancelación de una cuenta de una línea de crédito
DELINQ	Número de líneas de crédito morosas
CLAGE	Tiempo en meses de la línea de crédito de mayor antigüedad
NINQ	Cantidad de consultas de crédito. Son las solicitudes de reportes de crédito hechas por una institución a una agencia de reportes de crédito. 6 es considerado demasiado
CLNO	Número de líneas de crédito
DEBTINC	Relación deuda-ingreso. Es igual a la suma de todos los pagos de deuda del mes, dividida entre el ingreso bruto mensual

Durante el análisis y pre-procesamiento de los datos se eliminaron y transformaron las variables que no aportarán al modelo y se encontraron los siguientes hallazgos

<b>Variable Target</b>	Se considera a 'BAD' como variable target, recordando que el valor 1 corresponde al candidato con préstamo incumplido o con mora y que el valor 0 corresponde al candidato que paga su deuda.
<b>Valores nulos</b>	'DEBTINC' es la columna que cuenta con mayor cantidad de valores nulos con un 21.26% del total, por la cual se la excluye del análisis.
<b>Desbalanceo de datos</b>	Se observa que 840 de 4247 aplicantes tienen al menos una deuda morosa, lo que representa un 19.78% del total.
<b>Correlación</b>	Valor de la propiedad hipotecada



# GRACIAS



Helga Zambrana | Data Science  
[LinkedIn](#)