



# ANÁLISIS EXPLORATORIO DE DATOS: RIESGO CREDITICIO

Proyecto Final

Temas Selectos de Estadística  
Noviembre 2024

Alexia Aseneth Díaz Romo  
Expediente: 307924

## Contenido

Análisis Exploratorio de Datos.....	2
Objetivos principales del EDA aplicados al análisis de riesgo crediticio: .....	2
Introducción al riesgo crediticio para empresas .....	3
Conceptos iniciales .....	3
Crédito.....	3
Riesgo Crediticio .....	3
Marco de Basilea.....	3
Asignación de capital económico .....	4
Pérdida esperada (EL).....	4
Pérdida inesperada (UL).....	5
Definición de RWA .....	5
Modelo de probabilidad de incumplimiento PD.....	6
Definición de incumplimiento .....	6
Importancia de la identificación de impagos.....	6
Factores Potenciales de Riesgo .....	7
Factores de riesgo potenciales .....	7
Fuentes de oro para obtener los factores de riesgo para la estimación de PD .....	7
Modelos .....	8
Regresión logística .....	8
Calibración del modelo.....	8
Bibliografía .....	8

# Análisis Exploratorio de Datos

Enfoque fundamental en la ciencia de datos que consiste en analizar, visualizar y resumir los datos para comprender sus principales características, identificar patrones y descubrir relaciones antes de aplicar modelos estadísticos o de machine learning. Esto permite formular hipótesis y tomar decisiones informadas para los pasos posteriores del análisis.

## Objetivos principales del EDA aplicados al análisis de riesgo crediticio:

- **Conocer los datos:** Comprender la estructura y naturaleza del dataset (source: <https://www.kaggle.com/datasets/yashsaxena005/corporate-credit-risk>).
- **Resumir las características principales:** Calcular estadísticas descriptivas como la media, mediana, moda, varianza, percentiles, etc.
- **Visualizar datos:** Usar gráficas de dispersión para visualizar interacción entre diferentes drivers y con ello, generar el mejor modelo de Default Rate.
- **Explorar relaciones entre variables:** Examinar correlaciones y dependencias (matrices de correlación).
- **Preparación de los datos para el modelado:** El EDA es el puente entre la recopilación de datos y la creación de modelos predictivos.
- **Ofrece insights que guían a la toma de decisiones:** Descubrir perfiles de clientes con características asociadas a alto riesgo para ajustar requisitos de aprobación.

Este trabajo proporciona el análisis de riesgo crediticio, el cual es esencial para las instituciones financieras, ya que permite evaluar y gestionar el riesgo de impago de sus clientes. Este proceso se apoya en el uso de herramientas como el análisis exploratorio de datos (EDA), que ayuda a identificar patrones y factores clave relacionados con el incumplimiento. Los conceptos de pérdida esperada (EL), pérdida por estrés (SL), y activos ponderados por riesgo (RWA) son fundamentales para calcular el capital necesario para enfrentar posibles pérdidas. Además, el marco de Basilea proporciona normas internacionales para determinar los requisitos mínimos de capital, mientras que la identificación adecuada de impagos y la selección de factores de riesgo son críticas para desarrollar modelos precisos que estimen la probabilidad de incumplimiento. Con eso en mente, y mediante una serie de códigos en Python se logrará profundizar el tema en cuestión, con el objetivo de generar la aplicación de conocimientos de estadística mediante la creación de un modelo PD que nos permita evaluar y estimar la probabilidad de que un cliente o entidad incumpla con el pago de una deuda en un determinado período de tiempo.

Prácticamente, el EDA no solo proporciona una comprensión profunda de los datos en el análisis de riesgo crediticio, sino que también ayuda a evitar errores en la preparación de datos y asegura que los modelos predictivos sean confiables y robustos. Sin un buen EDA, el análisis de riesgo podría pasar por alto patrones cruciales o estar basado en datos incorrectos, lo que aumentaría los riesgos para la institución financiera.

# Introducción al riesgo crediticio para empresas

## Conceptos iniciales

**Crédito:** Cantidad de dinero u otro medio de pago que una persona o entidad, especialmente bancaria, presta a otro bajo determinadas condiciones de devolución

**Riesgo Crediticio:** posibilidad de sufrir una pérdida como consecuencia de un impago por parte de nuestra contrapartida en una operación financiera. Para el caso de la Institución Financiera, esta debe conocer el monto de capital para poder hacer frente a sus obligaciones por pérdida crediticia.

Los requisitos mínimos de capital han sido definidos internacionalmente y monitoreados por varias autoridades reguladoras:

- **Autoridad de Regulación Prudencial (PRA):** Regulador de servicios financieros en el Reino Unido. Es responsable de la regulación prudencial y la supervisión de bancos, sociedades de construcción, cooperativas de crédito, aseguradoras y grandes empresas de inversión.
- **Autoridad Bancaria Europea (EBA):** Es una autoridad reguladora de la Unión Europea. Sus tareas incluyen realizar pruebas de estrés a los bancos europeos para aumentar la transparencia en el sistema financiero europeo e identificar puntos débiles en la estructura de capital de los bancos.
- **Comisión Nacional Bancaria y de Valores (CNVB):** Es un organismo independiente de la Secretaría de Hacienda y Crédito Público con autonomía técnica y facultades ejecutivas sobre el sistema financiero mexicano.
- **Banco de Pagos Internacionales (BPI):** es una institución financiera internacional propiedad de bancos centrales que promueve la cooperación monetaria y financiera internacional y actúa como banco para los bancos centrales.

## Marco de Basilea

El Marco de Basilea es el conjunto completo de normas del Comité de Supervisión Bancaria de Basilea (BCBS), que es el principal organismo normativo mundial para la regulación prudencial de los bancos. Los miembros del BCBS han acordado implementar plenamente estas normas y aplicarlas a los bancos con actividad internacional en sus jurisdicciones.

Los marcos de Basilea II y III se definieron para tener un enfoque mucho más preciso a la hora de establecer los requisitos de capital para el riesgo crediticio. Las técnicas de gestión del riesgo crediticio se pueden clasificar en:

**Enfoque estandarizado:** implica una categorización simple de los deudores, sin tener en cuenta sus riesgos crediticios reales. Incluye la confianza en las calificaciones crediticias externas.

**Enfoque basado en calificaciones internas (IRB):** los bancos pueden utilizar sus propios modelos internos para calcular el requisito de capital regulatorio para el riesgo crediticio.

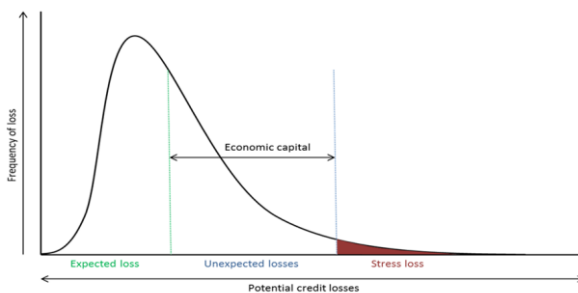
- Básico: solo modelo PD \*enfoque del proyecto
- Avanzado: modelos PD, EAD y LGD

Lo ideal es que los bancos apunten al enfoque IRB para considerar la mejor diferenciación del riesgo crediticio en función de las características de cada cliente. Sin embargo, deben cumplir con los requisitos mínimos para hacerlo.

Los marcos de Basilea están diseñados para llegar a los activos ponderados por riesgo (RWA), el denominador de los principales índices de capitalización.

## Asignación de capital económico

Para estimar la cantidad de capital económico necesario para respaldar sus actividades de riesgo crediticio, los bancos emplean un marco analítico que relaciona el capital económico total requerido para el riesgo crediticio con la función de densidad de probabilidad (PDF) de pérdidas crediticias de su cartera, también conocida como distribución de pérdidas de una cartera crediticia.



### Dónde:

**Pérdida esperada (EL):** monto de pérdida crediticia que el banco esperaría experimentar en su cartera crediticia durante el horizonte temporal elegido (cubierto por políticas de aprovisionamiento y fijación de precios).

**Pérdida inesperada (UL):** monto de pérdida crediticia inesperada dentro del horizonte temporal. Se mantiene capital para compensar la UL y dentro de la metodología IRB.

**Pérdida por estrés (SL):** probabilidad de que las pérdidas excedan la suma de EL y UL, es decir, la probabilidad de que el banco no pueda cumplir con sus obligaciones crediticias con ganancias y capital.

**Pérdida por estrés (SL):** probabilidad de que las pérdidas excedan la suma de EL y UL, es decir, la probabilidad de que el banco no pueda cumplir con sus obligaciones crediticias con ganancias y capital.

\*Las pérdidas pequeñas cercanas o ligeramente inferiores al límite de solvencia se producen con mayor frecuencia que las pérdidas grandes.

\*La probabilidad de que el banco se mantenga solvente durante el horizonte de un año es igual a 1-SL. En el marco de Basilea II, este nivel de confianza se fija en el 99,9%, es decir, se espera que una institución sufra pérdidas que excedan su capital una vez cada 1.000 años.

## Pérdida esperada (EL)

Se supone que la pérdida esperada individual es igual a la probabilidad de que un cliente incumpla en un plazo determinado, multiplicada por la exposición pendiente en el momento del incumplimiento y, una vez más, por la pérdida en caso de incumplimiento, que representa la porción de la exposición que no se recuperará después del incumplimiento.

Luego, la EL de la cartera se puede calcular como la suma de las EL individuales mediante la siguiente fórmula:

$$EL = \sum_{i=1}^N \underbrace{PD_i EAD_i LGD_i}_{\text{Individual EL}}$$

**Donde:**

**PD:** Probabilidad de incumplimiento es el porcentaje promedio de deudores que incumplirán en un período de un año.

**EAD:** Exposición en caso de incumplimiento es la estimación de la deuda pendiente si el prestatario incumple.

**LGD:** Pérdida en caso de incumplimiento representa la proporción de la exposición (EAD) que no se recuperará después del incumplimiento.

### Pérdida inesperada (UL)

Según el Marco de Basilea, la UL puede definirse como:

$$UL = (Total\ Loss - EL) * Maturity\ Adjustment$$

Además, basándose en el modelo de Vasicek para estimar la pérdida total de crédito, la UL puede estimarse finalmente como:

**Donde**  $N$  y  $N-1$  representan la función de distribución normal e inversa, respectivamente.

$$K = UL = \underbrace{\left[ LGD * N \left[ \frac{\sqrt{p} N^{-1}(0.999) + N^{-1}(PD)}{\sqrt{1-p}} \right] - PD * LGD \right]}_{\text{Total Loss (Vasicek model)}} * \underbrace{PD}_{\text{Expected Loss}} * \underbrace{\frac{1+(M-2.5)b}{1-1.5b}}_{\text{Maturity Adjustment}}$$

Correlación de  $p = .12 * \frac{1-e^{-50PD}}{1-e^{-50}} + .24 * \left[ 1 - \frac{1-e^{-50PD}}{1-e^{-50}} \right]$

$M$  es el vencimiento efectivo promedio de la cartera (que se asume comúnmente en 2,5 años).

Ajuste de vencimiento  $b = (0,11852 - 0,05478 * \ln(PD))^2$

### Definición de RWA

Los activos ponderados por riesgo (RWA) se utilizan para determinar la cantidad mínima de capital que deben mantener los bancos y otras instituciones financieras para reducir el riesgo de insolvencia por posibles pérdidas derivadas del riesgo crediticio. El requisito de capital se basa en una evaluación de riesgos para cada tipo de activo bancario.

Según los marcos de Basilea, los RWA para empresas se calculan como:

$$RWA = K * 12.5 * EAD$$

Dependiendo de la composición y uso de los fondos de capital, existen diferentes ratios para calcular los requerimientos finales de capital a través del siguiente método:

$$\frac{Capital}{RWA} = \% \text{Amt of risk covered by capital}$$

$$\text{If } \frac{Capital}{RWA} \geq Threshold \rightarrow Capital \geq Threshold * RWA$$

**Dónde:** **Capital:** cantidad de capital, **Umbral:** límite inferior de producción establecido por el regulador, p. ej., 8 %

# Modelo de probabilidad de incumplimiento PD

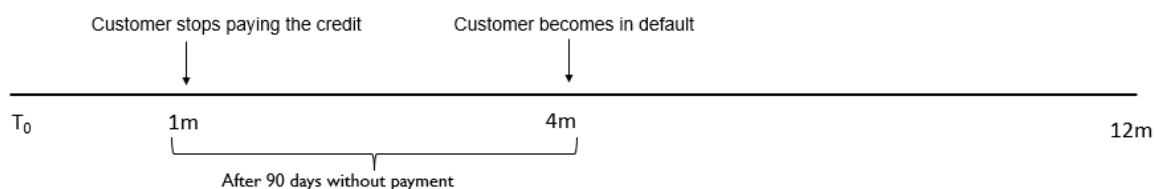
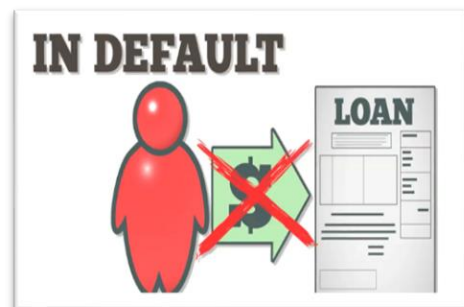
## Definición de incumplimiento

El incumplimiento es la falta de pago de los intereses o del capital de una deuda.

### ¿Cómo se considera que un cliente está en “mora”?

En ausencia de reglas específicas sobre los aspectos de la aplicación de la definición de incumplimiento, se han adoptado diversos enfoques en las instituciones y jurisdicciones que consideran:

**Criterio de morosidad en la identificación del incumplimiento:** p. ej., 90 días de mora en los próximos 12 meses y considerando un umbral de materialidad.



**Indicaciones de improbabilidad de pago:** por ejemplo, quiebra. Generalmente se refleja en la asignación de la peor calificación crediticia al cliente.

## Importancia de la identificación de impagos

La identificación de impagos es un aspecto esencial en el riesgo de crédito, ya que este concepto está presente en prácticamente todas las metodologías de gestión del riesgo de crédito de una institución. En particular, un cliente en mora se utiliza:

- Como variable objetivo del modelo PD: Los clientes clasificados como morosos o no morosos son la variable objetivo para estimar el modelo PD.
- Para construir la base de datos inicial para los modelos LGD y EAD: Los clientes identificados como morosos son el universo que se utilizará para estimar los modelos LGD y EAD.
- Como objetivo para desarrollar otras aplicaciones de riesgo de crédito: Se pueden utilizar diferentes definiciones de mora para gestionar la cartera de riesgo de crédito con modelos de alerta temprana o políticas de toma de decisiones crediticias.
- Debido a la importancia del concepto de mora en el riesgo de crédito, las instituciones deben establecer políticas y procesos estrictos para detectar y rastrear a los clientes en mora con el fin de gestionar todo el riesgo de crédito de la cartera.

Por lo tanto, el concepto de mora se convierte en la base de todos los aspectos del riesgo de crédito.

## Factores Potenciales de Riesgo

Para estimar la probabilidad de incumplimiento, se necesita un conjunto de variables explicativas (factores potenciales de riesgo).

Existe el riesgo de que algunos factores de riesgo puedan capturar de manera inapropiada las características de los deudores morosos. Por lo tanto, al momento de seleccionar los mejores factores de riesgo potenciales, es importante considerar lo siguiente:

- Poder discriminatorio: La variable debe mostrar un buen desempeño en la discriminación de riesgo.
- Tendencias monótonas: La variable debe presentar una tendencia monótona general con respecto a la tasa de incumplimiento.
- Comportamiento intuitivo: La variable debe ser un factor de riesgo intuitivo para predecir el riesgo crediticio.
- Sentido comercial: La variable debe cumplir con el juicio de expertos y el sentido comercial.

Las dos primeras características se pueden medir con pruebas estadísticas y herramientas de visualización, las otras dos están más relacionadas con la experiencia y recomendaciones de las empresas.

Es importante que las variables seleccionadas finalmente cumplan con los cuatro puntos, ya que, por ejemplo, un factor de riesgo que muestra un muy buen desempeño en la discriminación de riesgo pero que no tiene sentido puede sesgar la estimación de la probabilidad de error. Por lo tanto, un modelo con estas características no será aceptado por las empresas.

**Otras recomendaciones:** Las variables que son más fáciles de entender e implementar suelen ser más preferidas que otras variables complejas.

## Factores de riesgo potenciales

Para los fines de la estimación de la probabilidad de impago, los factores de riesgo deben considerarse a nivel de cliente y deben representar características específicas de los clientes. Además, se recomienda tener una amplia variedad de factores de riesgo potenciales para probar. Considerando esto, ¿cuáles podrían ser las fuentes potenciales para obtener factores de riesgo y estimar la probabilidad de impago?

### Fuentes de oro para obtener los factores de riesgo para la estimación de PD

- Información financiera: Por ejemplo, facturación/ventas netas, EBITDA u otras ratios financieras, etc.
- Historial crediticio de la empresa: Por ejemplo, número de créditos vencidos con instituciones bancarias, etc.
- Características demográficas: Por ejemplo, industria, número de empleados, etc.
- Otras características: Por ejemplo, años en el negocio, etc.



## Modelos

**Regresión logística:** Es uno de los algoritmos ML más utilizados para estimar la probabilidad de impago ya que:

- Es sencillo de estimar y entender
- Permite visualizar el desempeño estadístico individual de cada factor de riesgo.
- Sus estimaciones pueden ser bastante precisas en comparación con otros algoritmos más complejos
- Ajusta los valores estimados entre  $[0,1]$

Se basa en la siguiente fórmula:

$$p(x) = \frac{1}{1 + e^{-(Bx)}}$$

## Calibración del modelo

De acuerdo con los estándares del modelo, una vez seleccionado el modelo de PD final, es necesario calibrarlo al menos con respecto a la tendencia central de la cartera. Esto evita la volatilidad cíclica a partir de la estimación del riesgo de impago y permite una evaluación del desempeño de los prestatarios a lo largo del ciclo económico. La tendencia central se define de la siguiente manera:

$$CT = \frac{1}{N} \sum_{t=1}^N \frac{Dt}{Nt}$$

**Donde:**  $Dt$  = número de clientes morosos en el año  $t$ ,  $Nt$  = Número total de clientes en el año  $t$ ,  $N$  = Número total de años

## Bibliografía

IBM. (s.f.). *¿Qué es el análisis exploratorio de datos (EDA)?* Obtenido de <https://www.ibm.com/mx-es/topics/exploratory-data-analysis>

rae. (2024). *Real Academia Española*. Obtenido de <https://www.rae.es/>

Ucha, A. P. (1 de 08 de 2020). *Riesgo de crédito*. Obtenido de <https://economipedia.com/definiciones/riesgo-de-credito.html>

Valadez, A. E. (2023). *Credit Risk Course with python*.