

# ANÁLISIS EXPLORATORIO DE DATOS (EDA)

## Campaña Bancaria de Depósito a Plazo

Informe profesional elaborado a partir del análisis completo del dataset proporcionado. Incluye transformación, limpieza, análisis descriptivo, visualización e interpretación de resultados.

## 1. Introducción

Este informe presenta un análisis exploratorio de datos (EDA) aplicado a un conjunto de información proveniente de una campaña real de marketing bancario centrada en la contratación de un depósito a plazo. El propósito principal es comprender el comportamiento de los clientes, identificar patrones, detectar grupos de interés y analizar los factores que influyen en la respuesta positiva a la campaña.

## 2. Objetivos del Proyecto

- Comprender la estructura del dataset y evaluar la calidad de los datos.
- Realizar una limpieza completa y homogénea para asegurar consistencia.
- Analizar variables demográficas, económicas y operativas.
- Estudiar la variable objetivo ("yes" / "no") y su nivel de desbalance.
- Identificar segmentos con mayor o menor probabilidad de conversión.
- Generar visualizaciones claras, profesionales y orientadas al negocio.

## 3. Metodología Aplicada

La metodología utilizada se basa en el proceso estándar de análisis exploratorio de datos:

- 1. Carga y exploración inicial:** comprensión de tipos, valores únicos y estructura general.
- 2. Limpieza y transformación:** imputación, normalización, corrección de formatos y detección de outliers.
- 3. Análisis univariante:** revisión individual de cada variable.
- 4. Análisis bivariante:** comparación con la variable objetivo y búsqueda de relaciones relevantes.
- 5. Visualización:** gráficos de barras, histogramas, donuts, dispersión y mapas de correlación.
- 6. Conclusiones y recomendaciones:** hallazgos estratégicos para decisiones futuras.

## 4. Transformación y Limpieza

Después de la carga del archivo, se aplicaron diversas tareas de depuración:

- **Corrección de formatos:** conversión adecuada de fechas y valores decimales que provenían con coma.
- **Valores ausentes:** imputación mediante mediana en variables numéricas y categoría 'unknown' en categóricas.
- **Normalización de categorías:** unificación de profesiones, estados civiles y niveles educativos.
- **Detección de outliers:** variables como duración de llamadas y número de contactos mostraron colas largas.
- **Creación de variable binaria:** "y\_bin" para representar la respuesta de manera numérica.

## 5. Análisis Descriptivo del Dataset

El dataset final contiene más de 40.000 registros y 23 variables. Se clasifican en:

- **Demográficas:** edad, estado civil, educación.
- **Económicas:** ingresos, situación financiera (hipoteca, préstamo, impago).
- **Operativas:** número de contactos, duración de llamada, visitas web.
- **Macroeconómicas:** índice de precio, euribor, índice de confianza.
- **Objetivo:** contratación del depósito ("yes" / "no"). **Conclusión general:** existe un fuerte desbalance en la variable objetivo, con apenas un 11% de respuestas positivas.

## 6. Análisis por Variables

**Edad:** distribución amplia entre 18 y 90 años, con concentración entre 30 y 50 años.

**Ingresos:** distribución sesgada hacia ingresos altos, indicando presencia de outliers naturales.

**Estado civil:** predominio de clientes casados.

**Profesión:** mayor presencia de administrativos, técnicos y servicios.

**Duración:** variable operativa relevante; llamadas largas favorecen la conversión.

**Visitas web:** clientes digitalmente activos muestran mayor interacción.

## 7. Comportamiento y Conversión

La variable objetivo muestra:

- **'No':** 88,7% de los registros.

- **'Sí':** 11,3% del total. El análisis bivalente revela:
- **Edad:** jóvenes (18–29) y seniors (50+) presentan mayor probabilidad de conversión.

- **Profesión:** estudiantes y jubilados tienen tasas más altas de contratación.

- **Contactos previos:** la saturación reduce la tasa de éxito si se supera determinado umbral.

- **Ingresos:** los segmentos de ingresos medios son más receptivos.

## 8. Insights de Negocio

A partir de los resultados: **1. Foco en segmentos rentables:** jóvenes, seniors y grupos específicos como estudiantes y jubilados.

**2. Optimizar número de contactos:** evitar sobrecontacto que disminuye la conversión.

**3. Estrategias diferenciadas:** perfiles administrativos y operarios requieren campañas más específicas.

**4. Marketing digital:** clientes con alta actividad web pueden ser más receptivos a campañas digitales.

## 9. Conclusiones Finales

El EDA confirma la importancia de segmentar y personalizar campañas bancarias. La combinación de variables demográficas, financieras y operativas permite identificar de manera clara los grupos con mayor probabilidad de contratación. Este análisis constituye una base sólida para fases posteriores de modelización (Machine Learning) o diseño de

estrategias comerciales.