

## Análisis de Campaña: Exploración y Modelado de Datos

Este proyecto realiza un análisis exploratorio y descriptivo de las campañas de marketing de una entidad bancaria. El objetivo es identificar tendencias, patrones y realizar un análisis descriptivo del funcionamiento de las campañas y el comportamiento de los clientes y sus características para medir el éxito de la campaña de marketing.

### Estructura del Proyecto

- data/ # Datos crudos y procesados
- notebooks/ # Notebooks de Jupyter con el análisis
- README.md # Descripción del proyecto e informe

### Instalación y Requisitos

Este proyecto usa Python 3.13.5 y requiere las siguientes bibliotecas:

- pandas
- numpy
- matplotlib
- seaborn

### Resultados y Conclusiones

Se realizó un proceso de limpieza y transformación de datos que incluyó:

- Estandarización de los nombres de las columnas ( que empiecen por mayúscula)
- Unificación de los ID para más adelante hacer merge - Conversión de columnas con valores numéricos almacenados como texto (por ejemplo: 'Euribor3M').
- Tratamiento de valores nulos como por ejemplo:
  - En la columna 'Age' se imputó la mediana.
  - En 'Education' y 'Default' se completaron con valores como 'Unknown'.
  - En las columnas 'Cons.Price.Idx' y 'Date', dado que el porcentaje de nulos era muy bajo (1.09% y 0.57%), se dejaron sin modificar.

Un análisis descriptivo que permitió identificar características de los clientes como:

- La edad de los clientes se distribuye mayormente entre 25 y 60 años, habiendo algunos picos por encima de esta edad
- En 'Marital', la mayoría de los clientes están casados, seguidos de solteros y divorciados.
- En 'Job', las ocupaciones más frecuentes son 'blue-collar', 'management' y 'technician'.
- En df2, los ingresos presentan gran dispersión con valores entre aproximadamente 8,000 y 161,000. Con una media de 93.087 y la mediana de 92.962. Hay poca diferencia entre una y otra por lo tanto no hay muchos valores atípicos

- La variable 'Campaign' muestra que a mayor número de contactos, menor probabilidad de obtener respuesta positiva , insistir produce rechazo.
- En cuanto a la duración de las llamadas al inicio hay tasa baja de éxito (colgar nada más llamar) pero ésta aumenta según avanza la llamada.
- En estado civil, los solteros presentan un porcentaje de aceptación mayor en comparación con casados y divorciados.
- La variable 'Job' muestra diferencias claras: estudiantes y jubilados tienen tasas de respuesta 'sí' relativamente más altas.
- En df2, el ingreso se relaciona con el nivel educativo: quienes poseen estudios universitarios muestran mayores ingresos promedio.

### Conclusiones

- Factores con mayor relación con el éxito:
  - Estado civil: solteros > casados/divorciados.
  - Ocupación: estudiantes y técnicos con mayor aceptación.
  - Número de campañas: a partir de 3 intentos, la tasa de éxito cae.
  - Duración de llamada: cuanto más larga, mayor probabilidad de conversión.
- La mayoría de clientes dijeron **no**, lo que indica espacio de mejora en la segmentación por ejemplo dirigir la campaña a gente joven o jubilados para que acepten, o bien a casados o divorciados.
- Los clientes con mayor nivel educativo tienden a tener ingresos más altos, lo que podría ser un factor relevante para segmentación de campaña

### Próximos Pasos

Realizar otro tipo de análisis predictivo como por ejemplo qué probabilidades hay de que un tipo de cliente se suscriba

### Autores

- Maria Cristina Martinez Gutierrez
- [@Mariacris155](https://github.com/Mariacris155)