

REPORTE_LIMPIEZA_BANCOS_2024

Optimización y Estructuración de Datos de Marketing Bancario

Preparado por: Jorge Carlos Cuevas Noriega

1. Resumen Ejecutivo

El objetivo de este proyecto fue transformar un dataset de campañas de marketing bancario con formatos inconsistentes en un modelo de datos relacional y limpio. La calidad de los datos es el pilar fundamental para cualquier análisis predictivo posterior; sin este proceso, los resultados de negocio carecerían de precisión.

2. Desafíos Técnicos Identificados

Antes de la intervención, los datos presentaban los siguientes problemas:

Inconsistencia de Formatos: Columnas booleanas (como hipotecas o préstamos) almacenadas como texto ("yes"/"no"), lo que impedía cálculos lógicos.

Ruido en Categorías: Datos con puntos innecesarios o valores "unknown" que fragmentaban las agrupaciones de clientes.

Falta de Relacionalidad: Datos masivos en un solo bloque sin una clave primaria (ID) definida.

3. Resultados del Proceso de Limpieza

Tras ejecutar el script `limpieza_bancos.py`, se lograron los siguientes hitos:

Normalización Completa: Se procesaron 41,188 registros, asegurando que el 100% de las columnas críticas tengan el tipo de dato correcto (Boolean, Integer, Object).

Estructura Relacional: Se segmentó la información en tres tablas lógicas (client, campaign y economics), optimizando el almacenamiento y la velocidad de consulta.

Integridad de Datos: Se generó un archivo comparativo de datos sucios vs. limpios para auditoría, garantizando que no hubiera pérdida de información durante la limpieza.

4. Impacto en el Negocio

Con este dataset optimizado, el departamento de analítica ahora puede:

Identificar con precisión el perfil del cliente con mayor tasa de conversión.

Reducir el tiempo de preparación de datos en un 80% para futuros modelos de propensión de compra.

Garantizar que las decisiones estratégicas se basen en datos íntegros y validados.