

Análisis de Datos

Campaña Bancaria Resumen del Análisis de Datos

Una entidad bancaria contrata a una empresa de marketing encargada de contactar telefónicamente a posibles clientes para determinar si están interesados o no en adquirir un depósito a plazo fijo con el banco.

La información recolectada por la empresa de marketing se encuentra en un archivo CSV con 45215 filas y 17 columnas.

Cada registro contiene 16 características (las primeras 16 columnas) y una categoría ("yes" o "no" dependiendo de si la persona está o no interesada en adquirir el producto).

1. Carga y limpieza inicial de datos:

Se cargó un dataset con 45215 registros y 17 columnas. Se renombran las columnas para facilitar su lectura. Se identificaron valores nulos y se eliminaron, junto con duplicados.

2. Limpieza de categorías:

Se estandarizaron categorías inconsistentes como "trabajo", "estado_civil", "educacion", y otros campos. Se eliminan columnas poco relevantes como: forma_contacto, duracion_contacto, veces_contactadas, pdays, previous, intentos.

3. Filtrado de valores extremos:

Se filtró la edad máxima a 90 años para evitar registros anómalos (ej: edad=776).

4. Normalización de valores categóricos:

Se estandarizan valores como "div.", "Single", "DIVORCED", etc. También se limpiaron categorías como 'otro_credito' con variantes como 'YES', 'yes', 'No', etc.

5. Análisis de morosidad:

Se identificó que los perfiles con más morosidad son:

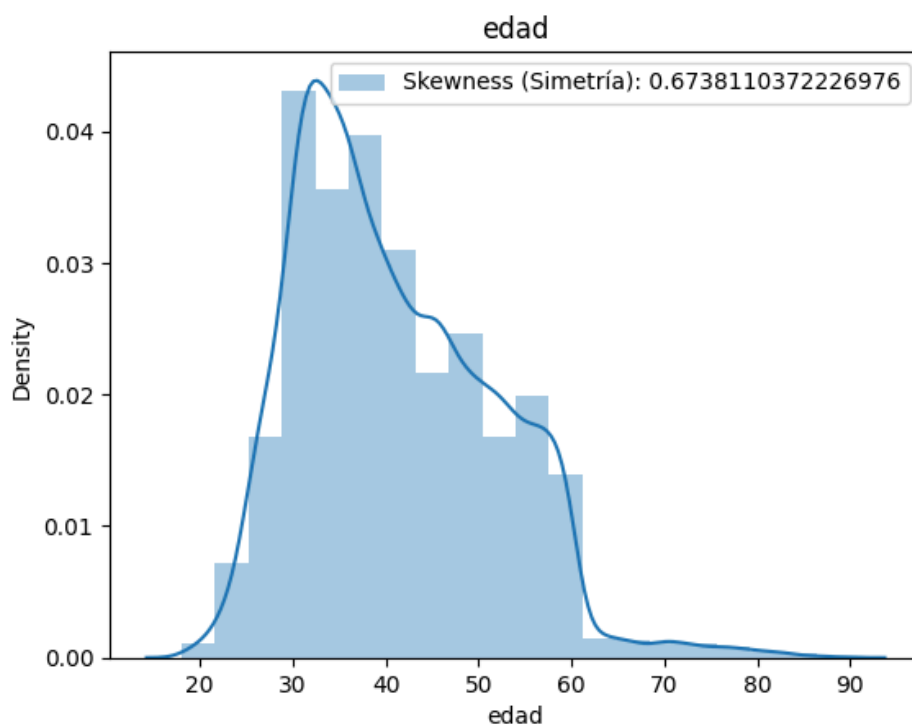
- Estado civil: casado
- Trabajo: no especificado
- Educación: secundaria
- Campaña Bancaria

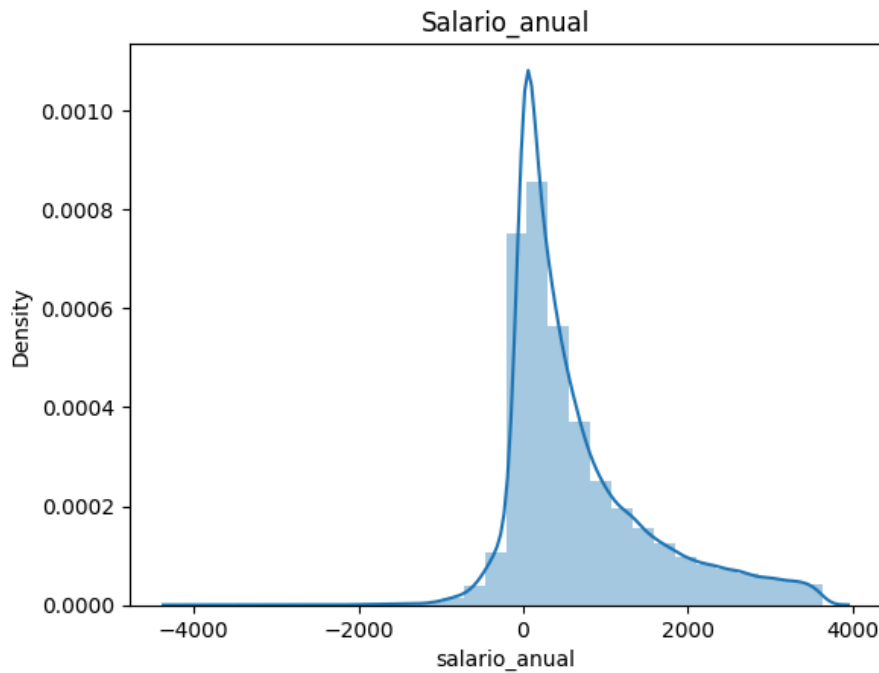
Sin embargo, la proporción de morosos en general es baja (~1.8%).

6. Identificación del perfil ideal:

Se determinó que el perfil más común de cliente sin morosidad y que acepta el producto es:

- Edad: 54 años
- Estado civil: casado
- Educación: secundaria
- Trabajo: no especificado
- Sin hipoteca ni otros créditos





7. Visualizaciones:

Se generaron gráficas de distribución y conteo para variables cualitativas y cuantitativas, ayudando a entender la composición del dataset.

8. Depuración final:

Se eliminaron los registros con morosidad, dejando una base de datos limpia con clientes potenciales fiables. Conclusión: El dataset ha sido limpiado, depurado y analizado, dejando una base sólida para modelos predictivos o segmentación de clientes.

