**📑 INFORME DE ANÁLISIS EXPLORATORIO DE DATOS (EDA)**

**🎯 1. Introducción**

El presente análisis explora datos de campañas de marketing de una entidad bancaria portuguesa. El objetivo es identificar patrones en los clientes que suscriben depósitos a plazo y comprender las variables que influyen en su decisión, para apoyar futuras estrategias comerciales.

**🔍 2. Datos analizados**

**Datasets:**

1. **bank-additional.csv**  
   Información sobre campañas de marketing directo: edad, profesión, estado civil, educación, préstamos, duración de llamadas, resultado de campañas anteriores, indicadores macroeconómicos y suscripción al producto (y).
2. **customer-details.xlsx**  
   Información demográfica y de comportamiento de clientes: ingresos, número de hijos y adolescentes en el hogar, fecha de alta como cliente y visitas mensuales al sitio web. Contiene tres hojas con clientes de diferentes años.

**🛠 3. Preparación y limpieza de datos**

* **Lectura de archivos** usando pandas.
* **Unión de datasets** con merge en la columna id\_ (bank) y ID (customer-details).
* Conversión de columnas de fecha a formato datetime.
* Revisión de **valores nulos y duplicados**, eliminando los registros duplicados detectados.
* Normalización de nombres de columnas para consistencia.

**📊 4. Análisis descriptivo**

**4.1 Estadísticas generales (numéricas)**

| **Variable** | **Media** | **Mediana** | **Desviación estándar** |
| --- | --- | --- | --- |
| Edad | ~40 | 38 | 10.5 |
| Ingreso (€) | ~52,000 | 51,000 | 12,000 |
| Duración | ~258 seg | 180 seg | 259 seg |
| Campaña | ~2.6 | 2 | 2.6 |

*(Valores aproximados, ver notebook para cifras exactas).*

**4.2 Distribución de variables categóricas**

* **Profesión:** mayor proporción en *blue-collar*, *management* y *technician*.
* **Estado civil:** mayoría *casados*, seguidos de *solteros*.
* **Educación:** predominan *secondary* y *tertiary*.
* **Resultado de campañas anteriores (poutcome):** la mayoría *unknown*, luego *failure* y *success*.

**📈 5. Visualizaciones clave**

1. **Histograma de edad**  
   Distribución normal con ligera asimetría a la derecha, concentrada entre **30 y 40 años**.
2. **Boxplot de ingresos según suscripción**  
   La mediana de ingresos es similar para clientes que suscriben o no el producto, mostrando que **el ingreso no es un factor determinante** por sí solo.
3. **Countplot de ocupación**  
   Refuerza que *blue-collar* es la categoría dominante, pero la tasa de suscripción es mayor en *management* y *admin*.
4. **Heatmap de correlación**
   * Correlación positiva entre **euribor3m** y suscripción.
   * **Duración de llamada** correlaciona con suscripción (pero cuidado con esta variable ya que no es usable en predicción antes de llamar).
5. **Distribución de la variable objetivo**  
   El dataset está desbalanceado con mayor proporción de *no suscripción*.

**💡 6. Insights y conclusiones**

✅ La duración de la llamada y la tasa **euribor3m** son variables altamente asociadas a la suscripción.  
✅ Clientes con resultado de campañas previas exitosas (*poutcome = success*) presentan mayor probabilidad de suscribirse nuevamente.  
✅ El ingreso anual no muestra diferencias significativas entre los que suscriben y los que no.  
✅ Las profesiones *management* y *admin* tienen mayor tasa de conversión comparado con otras.

**🚀 7. Recomendaciones estratégicas**

🔹 **Segmentación dirigida**: enfocar campañas en ocupaciones con mayor tasa de conversión.  
🔹 **Optimización de contactos**: priorizar a clientes con resultados positivos en campañas previas.  
🔹 **Modelado predictivo**: implementar modelos de clasificación (Logistic Regression, Random Forest) para predecir suscripción y optimizar recursos.  
🔹 **Dashboard ejecutivo**: crear visualizaciones dinámicas en Power BI o Tableau para seguimiento mensual.

**👤 Autor**

* **Nombre:** Aida Herrero de Diego