

Proyecto: Data Visualization with Power BI

GRUPO #2

- MARIA DE LOS ANGELES GARCIA VERA
- LUIS EDUARDO JARA MEDINA
- MIGUEL SEBASTIÁN OLALLA SACANCELA
- SAMIRA NAOMY PIGUABE ALMEA
- AQUILES ANDERS VALLEJO VASQUÉZ
- NEYVI MAGELES VERA CAJAPE

1. Estado actual del dataset y mejoras de calidad obtenidas tras el preprocesamiento.

El dataset contiene 2205 filas y 39 columnas. Cada fila representa un cliente individual, mientras que las columnas describen sus características sociodemográficas, comportamiento de compra y respuesta a campañas de marketing.

Cabe mencionar que el tipo de las variables identificadas son:

- **Numéricas:** Age, Kidhome, Teenhome, Customer_Days, NumWebPurchases, etc.
- **Monetarias:** Income, MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds, MntTotal, MntRegularProds.
- **Binarias/categóricas:** AcceptedCmp1–5, Response, Complain, marital_, education

Observaciones relevantes:

- El dataset presenta una estructura consistente y no muestra valores nulos significativos.
- Las variables monetarias se encuentran almacenadas como texto debido al símbolo \$, lo que requiere limpieza previa.
- La variable Age no presenta outliers evidentes y se encuentra dentro de rangos coherentes.
- Income y las variables de gasto muestran alta variabilidad, reflejando distintos perfiles de clientes.
- Se detecta un valor negativo en MntRegularProds, lo cual representa una posible inconsistencia de datos.
- Las variables categóricas están codificadas en formato binario y deberán ser reconstruidas.

2. Preprocesamiento de Datos

Es importante señalar que, no se identifican valores nulos en las variables del dataset, por lo que no fue necesario aplicar técnicas de imputación o eliminación de registros.

Por otro lado, se identificaron 184 registros duplicados en el dataset. La presencia de duplicados puede afectar los resultados del análisis al sobrerepresentar ciertos perfiles de clientes, por lo que se procedió a su eliminación.

2.1. Limpieza y transformación de variables monetarias:

'Income','MntWines','MntFruits','MntMeatProducts','MntFishProducts','MntSweetProducts','MntGoldProds','MntTotal','MntRegularProds' se limpiaron eliminando el símbolo \$ y se convirtieron a tipo float64.

2.2. Reconstrucción de Variables Categóricas

Reconstrucción de la variable Marital_Status

Las variables relacionadas con el estado civil se encontraban codificadas mediante columnas binarias (marital_*) . Para mejorar la interpretabilidad del dataset, se reconstruyó una única variable categórica.

La creación de la variable Marital_Status permite representar el estado civil del cliente de forma más clara y reduce la redundancia generada por múltiples columnas binarias.

Reconstrucción de la variable Education

De manera similar, el nivel educativo se encontraba representado mediante variables binarias (education_*). Se procedió a su consolidación en una sola variable categórica.

2.3. Eliminación de columnas originales

Una vez reconstruidas las variables categóricas, se eliminaron las columnas binarias originales para evitar redundancia y optimizar el modelo de datos.

2.4. Tratamiento de valores inválidos

Identificación de valores inválidos

No se encontraron registros con el valor 99999 en la variable Age.

Análisis y decisión de tratamiento

Tras la validación de la variable Age, se confirma que todos los valores se encuentran dentro de rangos coherentes para una población adulta. Por lo tanto, no fue necesario aplicar técnicas de imputación ni eliminación de registros para esta variable.

2.5. Coherencia general del dataset

El dataset presenta coherencia entre las variables sociodemográficas, de comportamiento de compra y respuesta a campañas. Las transformaciones realizadas no afectan la integridad de los datos y permiten un análisis confiable.

Estructura Final del Dataset

- **Filas:** 2,021
- **Columnas:** 31
- **Tipos de datos:**
 - Numéricos (int64 y float64) para Age, gastos, compras, etc.
 - Categóricos (object) para Marital_Status y Education
- Sin valores nulos ni duplicados.

3. Planteamiento de las preguntas de negocio

1. ¿Qué segmentos de clientes generan mayor gasto total?

Se espera identificar los segmentos de clientes con mayor aporte al gasto total, con el fin de priorizar acciones de fidelización, personalización de ofertas y asignación eficiente de recursos comerciales hacia los perfiles de mayor valor para la compañía.

2. ¿Cómo influye la antigüedad del cliente en el gasto total y la frecuencia de compra?

Se espera determinar si una mayor antigüedad se traduce en mayor gasto y frecuencia de compra, permitiendo medir el nivel de lealtad del cliente y diseñar estrategias de retención o reactivación según el ciclo de vida del consumidor.

3. ¿Qué categorías de productos concentran la mayor proporción del gasto total?

Se espera identificar las categorías que generan la mayor proporción de ingresos, con el objetivo de enfocar estrategias comerciales, promociones y optimización del portafolio en los productos de mayor impacto económico.

4. ¿Qué canal de compra está asociado a un mayor ticket promedio?

Se espera identificar el canal de compra que genera tickets promedio más altos, para orientar inversiones, mejorar la estrategia omnicanal y

potenciar aquellos canales que contribuyen con mayor valor por transacción.

5. ¿Existe diferencia en el gasto promedio entre clientes que responden a campañas y los que no?

Se espera medir el impacto de las campañas de marketing sobre el gasto promedio del cliente, con el objetivo de evaluar su efectividad y optimizar la segmentación y el enfoque de futuras acciones comerciales.

4. Planteamiento de KPIs estratégicos del negocio

1. Gasto total por segmento

Permite identificar los segmentos de clientes que generan mayor valor económico y priorizar estrategias de retención y fidelización.

2. Gasto promedio por cliente

Mide el valor medio generado por cliente y facilita comparaciones entre segmentos y períodos.

3. Ticket promedio

Evalúa el valor promedio de cada transacción, clave para estrategias de pricing y cross-selling.

4. Frecuencia promedio de compra

Indica cuántas veces compran los clientes en promedio, ayudando a medir lealtad y comportamiento de recompra.

5. Distribución del gasto por categoría (%)

Identifica las categorías que concentran mayor ingreso y orienta la optimización del portafolio de productos.

6. Ticket promedio por canal

Permite evaluar qué canales generan transacciones de mayor valor y optimizar la estrategia omnicanal.

5. Descripción de los KPIs y justificación de las visualizaciones seleccionadas

Los KPIs definidos para el dashboard fueron seleccionados a partir de las preguntas de negocio planteadas en el punto 3 de este informe, con el propósito de analizar el comportamiento de los clientes, su aporte

económico y sus patrones de compra. Cada KPI fue creado mediante medidas DAX, garantizando así obtener cálculos dinámicos de acuerdo al filtro aplicado. De igual forma, las visualizaciones fueron escogidas tomando en consideración aspectos de claridad, comparabilidad entre segmentos y correcta interpretación de los resultados.

Se utilizaron tarjetas para los indicadores generales, gráfico de barras y columnas para poder realizar comparaciones entre segmentos, facilitando así el análisis interactivo del usuario. De este modo, se pudo transformar los KPIs en información valiosa e interpretable que aporte en la correcta toma de decisiones.

6. Análisis de KPIs: criterios de visualización e insights relevantes

1. Gasto total por segmento

Alineado a la consulta del negocio: *¿Qué segmentos de clientes generan mayor gasto total?* Se realizó la construcción del KPI *Total Gasto* el cual se expone a través de una tarjeta, puesto que permite resumir de forma directa el valor económico total generado por los clientes. Además, funciona como indicador principal de cómo se está desempeñando el negocio. Cabe mencionar que, este KPI se complementa con un gráfico de barras, que permiten comparar de manera clara el gasto entre niveles educativos. De este modo, se pudieron identificar los insights a continuación:

- Los **clientes graduados** concentran el mayor gasto total, lo que sugiere una relación entre el nivel educativo y el comportamiento de consumo, asociada a una mayor estabilidad económica y decisiones de compra de mayor valor.
- Este hallazgo evidencia una oportunidad clara para diseñar estrategias de fidelización y comunicación personalizada dirigidas a este perfil, con el fin de maximizar retornos más altos y estables en el tiempo.
- En este sentido, segmentar por nivel educativo resulta adecuado, ya que es una variable estructural del perfil del cliente que no cambia en el corto plazo y permite pasar de una estrategia general a una estrategia enfocada en los clientes de mayor valor para el negocio.

2. Gasto promedio por cliente

El KPI de gasto promedio por cliente permite identificar cuánto valor económico genera un cliente en promedio. De acuerdo con el gráfico segmentado por estado civil, se observan diferencias relevantes en el nivel de gasto entre los distintos grupos demográficos. Esto evidencia que el comportamiento de compra no es homogéneo y que determinadas características del cliente influyen en el valor económico generado. Algunos segmentos presentan un gasto promedio superior, lo

que sugiere una mayor propensión al consumo o una mayor capacidad de gasto, mientras que otros muestran niveles más moderados y estables.

Este análisis permite comprender cómo variables demográficas, como el estado civil, aportan contexto al comportamiento de compra del cliente y ayudan a identificar patrones de consumo diferenciados. De este modo, se pudieron identificar los insights a continuación:

- Existen diferencias significativas en el gasto promedio entre los distintos estados civiles ya que el análisis evidencia que el gasto promedio por cliente no es homogéneo, sino que varía según el estado civil. Esto confirma que las características demográficas influyen en el comportamiento de consumo y en el valor económico generado por cada cliente.
- El gasto promedio no depende únicamente del volumen de clientes, sino del perfil del consumidor ya que no necesariamente el segmento más numeroso es el que genera mayor gasto promedio. Esto sugiere que la rentabilidad del cliente está más relacionada con su perfil que con la cantidad de clientes en cada segmento.

3. Ticket promedio

El KPI *Ticket Promedio* se toma como indicador clave del comportamiento de compra, puesto que refleja el valor promedio de cada transacción.

4. Frecuencia promedio de compra

Así mismo, la *Frecuencia Promedio de Compra* se representó a través de un gráfico de barras, lo que permite identificar patrones de compra e índices de lealtad entre los distintos grupos de clientes. Cabe mencionar que este KPI complementa la visión del KPI *Ticket Promedio*, puesto que un cliente puede comprar poco pero con alta frecuencia y viceversa.

Finalmente, el KPI *Antigüedad Promedio del Cliente* introduce una visión temporal, este fue creado en base a la variable *Customer_Days* y permite evaluar la relación entre el tiempo de permanencia del cliente y su comportamiento de compra.