

Informe de Análisis Exploratorio de Datos (EDA) - Marketing Campaign

Hallazgos y Observaciones Clave

1. Descripción General del Conjunto de Datos

El conjunto de datos contiene información sobre una campaña de marketing con 29 columnas y 2240 entradas.

Las variables incluyen datos demográficos (edad, educación, estado civil, ingresos), comportamiento de compra (gastos en diferentes categorías, frecuencia de visitas web) y respuestas a campañas anteriores.

2. Visualizaciones Importantes

Distribución de Edad (Year_Birth): Se observa que la mayoría de los clientes nacieron entre 1960 y 1980, lo que sugiere una audiencia predominantemente de mediana edad.

Nivel de Educación: La mayoría de los clientes tienen educación universitaria (Graduation) o superior (PhD, Master).

Ingresos: La distribución de ingresos muestra una variedad, con algunos valores atípicos en el extremo superior.

Relación entre Ingresos y Gastos: Los clientes con mayores ingresos tienden a gastar más en productos como vinos y carnes.

3. Comportamiento de Compra

Gastos por Categoría: Los clientes gastan más en vinos, seguido de productos cárnicos. Los gastos en frutas y dulces son significativamente menores.

Frecuencia de Compra: La mayoría de las compras se realizan en tiendas físicas (NumStorePurchases), seguidas de compras en línea (NumWebPurchases) y catálogos (NumCatalogPurchases).

4. Respuestas a Campañas Anteriores

La mayoría de los clientes no respondieron a las campañas anteriores (AcceptedCmp1-5), con una baja tasa de aceptación general.

La campaña más exitosa fue la última (AcceptedCmp5), aunque aún con una baja participación.

5. Datos Faltantes

La columna "Income" tiene algunos valores faltantes (24 entradas), lo que podría requerir imputación o eliminación dependiendo del análisis posterior.

Comentarios en el Código

Carga de Datos:

```
df = pd.read_csv("/content/marketing_campaign.csv", sep='\t')
```

Se cargan los datos desde un archivo CSV, especificando el separador como tabulación (\t).

Análisis Exploratorio:

```
df.info()
```

```
df.describe().T
```

info() proporciona un resumen de las columnas, tipos de datos y valores no nulos.

describe().T muestra estadísticas descriptivas para las variables numéricas.

Visualizaciones:

```
sns.pairplot(df)
```

Se utiliza pairplot para visualizar relaciones entre variables numéricas. Esto ayuda a identificar correlaciones y patrones.

Generación de Reporte Automatizado:

```
report = ProfileReport(df)
```

```
report.to_file(output_file="report.html")
```

Se genera un reporte EDA automatizado utilizando ydata-profiling, que incluye estadísticas, distribuciones y correlaciones.

Conclusiones

Perfil del Cliente:

La audiencia objetivo es principalmente de mediana edad, con educación universitaria y niveles de ingresos variados.

Los clientes con mayores ingresos muestran un mayor gasto en categorías premium como vinos y carnes.

Efectividad de Campañas:

Las campañas anteriores tuvieron una baja tasa de aceptación, lo que sugiere la necesidad de optimizar las estrategias de marketing para aumentar la participación.

Recomendaciones:

Segmentación: Dirigir campañas específicas a grupos demográficos clave (por ejemplo, clientes con altos ingresos para productos premium).

Personalización: Utilizar datos de comportamiento de compra para personalizar las ofertas y aumentar la relevancia.

Tratamiento de Datos Faltantes: Considerar imputar los valores faltantes en "Income" para evitar sesgos en análisis futuros.

Próximos Pasos:

Realizar un análisis más detallado de las correlaciones entre variables para identificar factores clave que influyen en las compras.

Implementar modelos predictivos (por ejemplo, clasificación para predecir la respuesta a campañas) basados en los insights obtenidos.

Visualizaciones Clave (Resumen)

Distribución de Edad y Educación: Muestra la concentración de clientes por edad y nivel educativo.

Gastos por Categoría: Comparación de gastos en diferentes categorías de productos.

Relación Ingresos vs. Gastos: Gráfico de dispersión que ilustra cómo los ingresos afectan los patrones de gasto.

Este EDA proporciona una base sólida para entender el conjunto de datos y guiar decisiones estratégicas en futuras campañas de marketing.