



Interacción final de clientes E-Commerce

Profesor: Fernando Carabedo

Tutor: Noelia Ferrero

Alumno: Lucas Andrés Márquez



Introducción

Visión general

El dataset elegido es de las ventas de un E-Commerce del 2023-2024.

Objetivos

- Se busca predecir cuando una transacción va a terminar siendo una compra o no, tomando como variable objetivo a “Interaction type”.
- Buscar cuales son las principales variables que se deben tener en consideración para determinar cuál va a ser la decisión final del cliente.

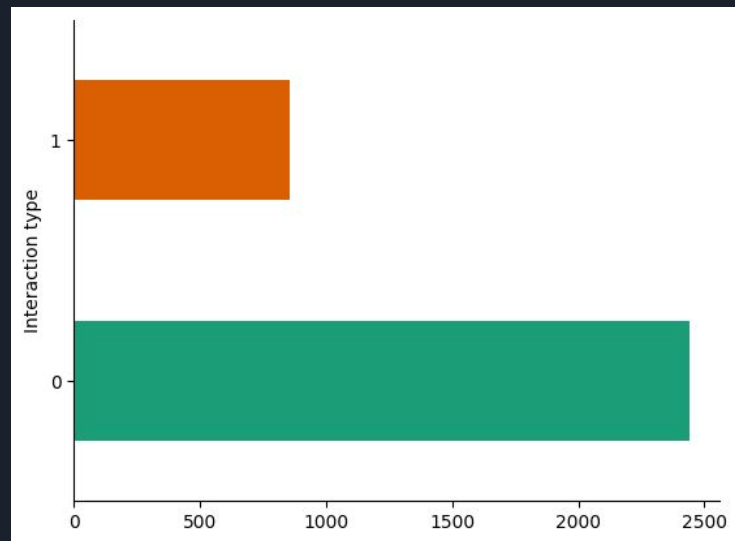
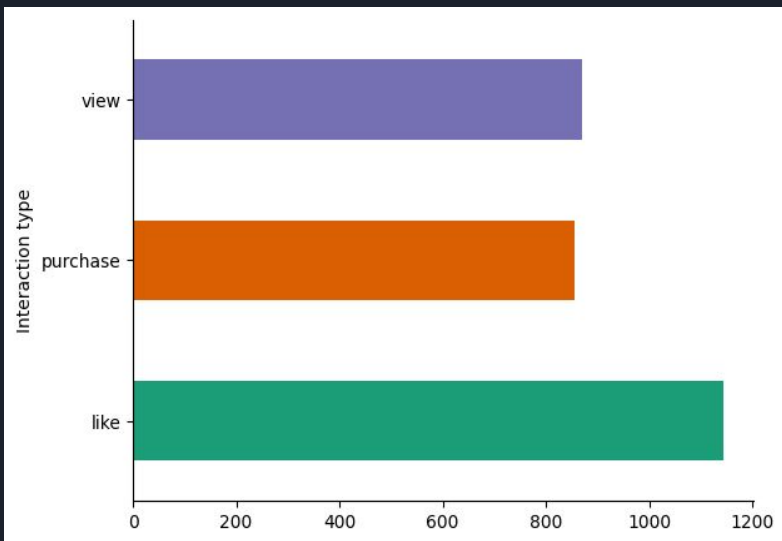


Análisis exploratorio

- Cuenta con tres tablas:
 - E-commerce sales data 2024 (3294 filas, 5 variables).
 - customer_details (3900 filas, 18 variables).
 - product_details (10002 filas, 28 variables).
 - Haciendo un total de 50 variables entre las tres tablas.
 - 31 variables categóricas.
 - 20 variables numéricas.
 - De las cuales 13 variables de la tabla “product_details” estaban vacías.
- Limpieza de datos
 - Se procedió a eliminar 13 variables vacías, 2 con una gran cantidad de nulos y 1 variable que se consideró irrelevante para determinar la variable objetivo de la tabla “product_details”,
 - Se procedió a eliminar 1 variable vacía de la tabla “e_commerce”

Limpieza de datos

- Se procedió a eliminar 13 variables vacías, 2 con una gran cantidad de nulos y 2 variables que se consideraron irrelevantes para determinar la variable objetivo de la tabla “product_details”,
- Se procedió a eliminar 1 variable vacía de la tabla “e_commerce”
- Se transformó la columna 'Interaction type' para que quede purchase por un lado como 1 y agrupar los otros dos segmentos como 0, como se puede visualizar en el gráfico de la derecha





Limpieza de datos

- Se realizó la unión de las tablas por medio de dos merge, un Outer merge entre product_details y e_commerce y un segundo Outer merge del merge anterior y customer_details.
- Se descartaron aquellas filas que no tenían dato en el campo de la variable objetivo ya que no podría determinar si finalmente se concretó o no la compra.
- Armo una lista de columnas de texto con el nombre de las variables que quiero cambiar para cambiar las variables a 0 y 1
- Se transformó la columna “Shipping Weight” para que sea numérica utilizando regex para separar el número de la unidad de medida y en donde figura “pounds”(Libra) que se multiplicará por 16 para pasar a “ounces”(Onza)

Análisis bivariado

Si bien es mayor el número de muestra para hombres, se visualiza una diferencia en cuanto a la totalidad de veces que el hombre no termina comprando con respecto a la mujer.

