

Alumno: Federico Gauna

Título: Análisis de la evolución de las ventas de un e-commerce

Introducción

En este proyecto, se analizarán las ventas de un e-commerce en crecimiento. Se nos presentan datasets con información sobre los clientes, vendedores, las transacciones entre estos y un dataset provisto por un tercero con el poder de compra de cada cliente. También, se usará un dataset con el tipo de cambio a dólar para poder evaluar este crecimiento en términos reales debido al contexto inflacionario.

El periodo de tiempo sobre el cual se encuentran los datos es entre enero y octubre del 2023. Pero, se trabajarán los datos de enero hasta septiembre del 2023 *-explicado a más a detalle en el preprocesamiento-*.

Alcance

El trabajo de análisis de estos datos tiene como principal meta armar un dashboard a nivel operacional, que tiene como finalidad mostrar la evolución de las ventas del e-commerce en términos reales e identificar tanto los puntos destacados que deben ser preservados como las áreas de oportunidad que requerirán una atención especial. Estos últimos serán el enfoque primordial para implementar mejoras significativas. Por esto, este análisis será dirigido a los Managers del emprendimiento.

Objetivos

En términos generales, lo que se hará es realizar un análisis sobre estos datos para visualizar como es la evolución de las ventas. Los objetivos a los que se pretenden llegar son los siguientes.

- Revisar la cantidad de ventas y sus respectivos montos para identificar patrones, puntos altos y bajos o valles.
- Examinar el crecimiento o decrecimiento de las ventas en pesos y dólares para comprender como el contexto inflacionario afecta al e-commerce.
- Evaluar la distribución de oferta de productos por categoría para identificar áreas de fortaleza y posibles oportunidades de expansión o mejora.
- Reconocer cuales son los clientes que compran con más frecuencia y cuáles son los vendedores lideres en términos de volumen de venta.

- Analizar los grupos etarios y su relación con respecto a las ventas y categorías, para saber quiénes son los que más compran y qué compran.
- Relacionar la cantidad de compras de los clientes con el poder de compra y encontrar patrones de comportamiento o segmentos de clientes relevantes.

Hipótesis

Teniendo en cuenta cuales son los objetivos de este análisis, las hipótesis que se van a plantear, probar y validar a lo largo del proyecto son las siguientes:

- Se espera encontrar patrones estacionales en las ventas, con aumentos notables durante ciertos períodos del año.

Variables a Evaluar: Fecha de la transacción, montos de venta.

- Se postula que las variaciones en el tipo de cambio pueden influir en el comportamiento de compra, afectando el crecimiento o decrecimiento de las ventas tanto en pesos como en dólares.

Variables a Evaluar: Tipo de cambio, montos de ventas en pesos y dólares.

- La diversificación de la oferta de productos por categoría contribuirá positivamente al rendimiento general, atrayendo a un público más amplio y generando mayores oportunidades de venta.

Variables a Evaluar: Cantidad de productos por categoría, monto por categoría.

- Existe una relación directa entre la frecuencia de compra de los clientes y el rendimiento de los vendedores, ya que ciertos vendedores pueden tener una base de clientes leales.

Variables a Evaluar: Cantidad de compras de clientes, Cantidad de ventas de vendedores.

- Se supone que el poder de compra de un cliente impacta su comportamiento de compra, por lo cual se espera que clientes de mayor poder adquisitivo realicen compras más frecuentes o de mayor valor.

Variables a Evaluar: Poder de compra de los clientes, frecuencia y monto de compras.

Dataset

Contamos con cuatro datasets: *customers*, *customers_purchasing_power*, *merchants* y *transactions*.

El dataset de “customers” contiene información sobre los clientes que utilizan el e-commerce para comprar productos.

El dataset de “customers_purchasing_power” contiene información sobre el segmento socioeconómico de los clientes, indicando “low”, “mid” y “high” según tengan más o menos poder de compra.

El dataset de “merchants” contiene la información de los vendedores que utilizan el e-commerce para publicar y vender sus productos.

El dataset de “transactions” contiene toda la información sobre las transacciones que se realizan dentro del e-commerce entre los vendedores y los clientes.

El dataset de “datos-tipo-cambio-usd” contiene la información sobre el tipo de cambio a dólar desde el 2002 hasta el 2023.

Columnas

Dataset “customers”:

- CustomerID (INT): Clave primaria.
- Name (STRING): Nombres de los clientes.
- Surname (STRING): Apellidos de los clientes.
- Gender (STRING): Género de los clientes.
- Birthdate (DATE): Fecha de nacimiento de los clientes.

Dataset “customers_purchasing_power”:

- CustomerID (INT): Clave primaria.
- Purchasing_Power (STRING): Indicador de poder de compra.

Dataset “merchants”:

- MerchantID (INT): Clave primaria.
- Merchant_Name (STRING): Nombres de los vendedores.
- Category (STRING): Categoría de los productos que se venden.

Dataset “transactions”:

- TransactionID (INT): Clave primaria.
- Customer_ID (INT): Clave foránea. ID de los clientes.
- Date (STRING): Fecha de la transacción.
- MerchantID (INT): Clave foránea. ID de los vendedores.
- Transaction_Amount (INT): Monto de la venta en pesos.

Modelo de Datos



Se hicieron 3 relaciones lógicas. Por un lado, se relacionó la tabla de “transactions_ecommerce_clean” con “customers_clean” utilizando las columnas “Customer ID” y “Customer ID1”. También se la relacionó con “merchants” utilizando las columnas “Merchant ID” y “MerchantID”. Por último, se la relacionó con “datos-tipo-cambio-usd” utilizando las columnas “Date” e “Indice Tiempo”.



Se hizo una relación física con las tablas “customers_clean” y “customers_purchasing_power”. Se utilizó un inner join relacionando las columnas “Customer ID1” con “Customer ID”.

Preprocesamiento

Dataset “customers”:

- La columna ‘Gender’ tiene registros faltantes.
- Hay 158 registros duplicados.
- Se filtran los índices de los registros duplicados y se eliminan del DataFrame.
- Los registros faltantes se rellenan con un string.
- En Tableau, se crea una columna nueva con un campo calculado en la que se juntan las columnas ‘Name’ y ‘Surname’.

Dataset “merchants”:

- No tiene registros faltantes.
- No tiene duplicados.

Dataset “transactions_ecommerce”:

- No tiene registros faltantes.
- Tiene 3930 registros duplicados.

- Se filtran los índices de los registros duplicados y se eliminan del DataFrame.
- En Tableau, se filtraron todos los registros cuya fecha sea anterior al “01/10/2023”, por dos motivos:
 - En primer lugar, las ventas de octubre solo llegan hasta el día 14, siendo este faltante de información el causante de un análisis incorrecto de los datos, ya que se están trabajando TODAS las ventas de TODO el mes.
 - En segundo lugar, no se dispone del cambio a dólar en las fechas de octubre, por lo cual, no es posible analizar estas ventas en un contexto real.
-

Dataset “customers_purchasing_power”:

- No tiene registros faltantes
- No tiene duplicados.

Dataset “datos-tipo-cambio-usd”:

- En Tableau, se ocultaron todas las columnas, excepto por “Índice Tiempo” y “Tipo Cambio Implícito En Adrs”.
- En Excel, se reemplazó los registros de la columna “Tipo Cambio Implícito En Adrs” que tenían un punto (.) por una coma (,), para que Tableau los reconozca como valor numérico.

Métricas

Las métricas creadas para este análisis son las siguientes:

- Cantidad de Ventas:

$\text{Cantidad de ventas} = \text{COUNT}([\text{Transaction ID}])$

- Promedio de Ventas:

$\text{Prom_Ven_Men} = \text{COUNT}([\text{Transaction ID}]) / 9$

- Promedio de Monto (ARS):

$\text{Prom_Mnt_Men} = \text{SUM}([\text{Transaction Amount}]) / 9$

- Ventas (USD):

$[\text{Transaction Amount}] / [\text{Tipo Cambio Implícito En Adrs}]$

- Promedio de monto (USD):

$$\text{Prom_Mnt(USD)_Men} = \text{SUM}([\text{Ventas (USD)}]) / 9$$

Visualización

Con respecto a la visualización, lo primero que tenemos es una página principal con un menú para ir a los distintos tableros: vendedores – clientes – ventas. Esta división fue hecha para agrupar a los gráficos en correspondencia a la temática que explican.

En la parte de “Vendedores”, utilizan gráficos de barras para visualizar los mayores vendedores y una línea de tiempo que muestra el crecimiento y decrecimiento, junto a sus distintos máximos y mínimos. Se representa tanto en cantidad como en monto (ARS y USD).

En la parte de “Clientes”, se usan gráficos de barras para ver cuáles son los clientes que más compras y cuales los que menos. También, se usa una tabla en la que se visualiza una matriz RFM, en la cual se puede ver la distribución de los clientes más valiosos. Y, por último, con un gráfico de torta se muestra el porcentaje de clientes según su poder de compra.

En la parte de “Ventas”, se muestran gráficos de barras que se representa los ingresos totales de cada categoría en pesos y en dólares, y la cantidad de ventas segmentado por poder de compra para cada categoría.

Conclusiones

Los insights que se encontraron después de hecho el análisis son los siguientes:

- Existió un crecimiento y picos en la cantidad de ventas y en las ganancias en pesos, pero, al analizarlo en dólares, se ve un decrecimiento.
- La categoría “Travel” es de las que tiene mayor monto mientras que “Restaurant” es de las que menos.
- La cantidad de ventas por categoría está distribuida de forma bastante uniforme siendo “Electronics” la que más vende y “Travel” la que menos.
- Hay una gran cantidad de clientes, pero muchos de ellos tienen una frecuencia y recencia en la aplicación muy baja.

- La mayor parte de las transacciones son hechas por clientes con poco poder de compra, mientras que los que tienen alto poder de compra son la menor proporción.

Como conclusión, se puede decir que, durante el periodo de enero a septiembre, aunque hubo un crecimiento en lo que respecta a las ventas, esto solo es cierto si se habla de cantidad de ventas, ya que las ganancias en pesos crecen, pero si se ven en un término real como pueden ser los dólares, hay un decrecimiento bastante notable debido al contexto inflacionario.

Se podría adjudicar el crecimiento en ventas y la gran cantidad de clientes a la novedad del e-commerce, ya que gran parte de estos clientes lo utilizan muy poco o, directamente, no vuelven después del primer uso de la aplicación. Una recomendación sería aplicar una estrategia de marketing para recuperar los clientes a punto de perderse y mantener a los pocos que son más activos.

Conocer a los mayores vendedores y a los clientes que más compran de la aplicación es importante para saber quiénes son los usuarios que hay que cuidar para mantener activo el e-commerce.

Por otro lado, saber quiénes son los clientes con menos compras también es importante, ya que es una potencial oportunidad del negocio. Se tendrá que hacer un análisis más profundo sobre estos y decidir si vale la pena invertir recursos en recuperarlos o no.

Con respecto al poder de compra de los clientes, nos encontramos lo que probablemente sea la mayor oportunidad para el e-commerce, ya que los clientes con poder de compra "High" y "Mid" abarcan el 40% de total. Llevar a cabo las estrategias necesarias para incentivar el uso de la aplicación a este grupo sería fundamental.

Por último, hay una oferta de categorías bastante uniforme, pero lo que más resalta es lo que sería la otra oportunidad de negocio, la categoría de "Travel". Es la que acumula mayores ingresos por mucha diferencia, pero al mismo tiempo es la que menos se compra. Se recomienda comunicarse con los vendedores y llegar a un acuerdo para promocionar más esta categoría, especialmente en los clientes con alto poder de compra.