

Trabajo Práctico Integrador

Laboratorio de Recopilación e Integración de Datos

Estudiante

Ponce, Carolina

Docente

Jáuregui, Santiago

Índice

Índice.....	1
Introducción.....	2
Hipótesis.....	2
Alcance.....	2
Modelado de Datos.....	3
Definición de tablas.....	3
Relaciones Físicas y Lógicas.....	4
Transformaciones del dataset.....	5
Métricas y Dimensiones.....	6
Dashboard.....	7
Portada.....	7
Customers Analysis.....	8
Sales Analysis.....	11
Conclusión.....	14
Anexo.....	15

Introducción

Este proyecto se basa en un conjunto de datos proporcionado por el profesor de la asignatura 'Laboratorio de Recopilación e Integración de Datos', el cual nos brindará la oportunidad de asumir el papel de consultores en análisis de datos para un e-commerce en pleno crecimiento, que busca evaluar la evolución de sus ventas en términos reales. De esta forma nos permite simular situaciones y desafíos reales del mundo empresarial a través de la herramienta Tableau.

Hipótesis

El objetivo de este análisis, es más bien descriptivo, se quiere conocer los clientes de este e-commerce y cómo evolucionan las ventas a lo largo del año 2023.

La información que se busca de este análisis desde el lado del cliente, es conocer su público, la cantidad total de clientes; la cantidad de clientes activos, es decir aquellos que han hecho una transacción o más, y la cantidad de clientes inactivos, es decir aquellos que no han hecho una transacción aún; cuál es el poder adquisitivo de los clientes; cuál es la distribución según género y grupos etarios; y quienes son aquellos clientes que más gastan en este e-commerce.

Por otro lado, desde la mirada de las ventas, se busca averiguar la cantidad de transacciones que se han realizado; cuál es la suma total y el promedio de aquellas transacciones realizadas; qué rubro o categoría es más rentable y cuáles no, y cómo fueron evolucionando las ventas por categoría, a lo largo del año 2023, del mes de enero al mes de octubre.

Alcance

El análisis de este dataset podría ser valioso para líderes empresariales y ejecutivos ya que les permiten conocer el estado del e-commerce en el año 2023, teniendo dos miradas, desde el lado del cliente, como de las ventas.

Modelado de Datos

El dataset está compuesto por cuatro archivos en formato csv, los cuales son, 'customers', 'customers purchasing power', 'transactions' y 'merchants'.

Definición de tablas

Tabla 'Customers': dicha tabla contiene los datos personales de los distintos clientes, como número de identificador, nombre, apellido, género y fecha de nacimiento.

Customers		
Campo	Tipo de campo	Clave
Customer ID	int	PK
Name	varchar	
Surname	varchar	
Gender	varchar	
Birthdate	date	

Tabla 'Customers Purchasing Power': indica el nivel socioeconómico de cada cliente a través de su número identificador.

Customers Purchasing Power		
Campo	Tipo de campo	Clave
Customer ID	int	PK/FK
Purchasing Power	varchar	

Tabla 'Transactions': en dicha tabla se almacenan aquellos datos perteneciente a la venta, como, el número identificador de transacción, identificador del cliente, fecha, identificador del vendedor y el monto de la venta.

Transactions		
Campo	Tipo de campo	Clave
Transaction ID	int	PK
Customer ID	int	FK
Date	date	
Merchant ID	int	FK
Transaction Amount	int	

Tabla 'Merchants': en ella se registran el identificador de cada vendedor, su respectivo nombre y categoría.

Merchants		
Campo	Tipo de campo	Clave
Merchant ID	int	PK
Merchant Name	varchar	
Category	varchar	

Relaciones Físicas y Lógicas

Entre las tablas 'customers' y 'customers purchasing power', se estableció una relación física de tipo left join, a través de la cláusula 'Customer ID' de las respectivas tablas.

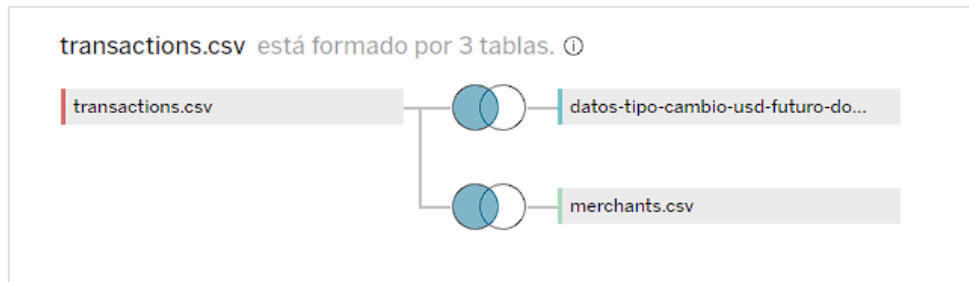
customers

customers.csv está formado por 2 tablas. ⓘ



Entre las tablas 'transactions' y 'merchants', se estableció una relación física de tipo left join, a través de la cláusula 'Merchant ID' de las respectivas tablas. A su vez, para poder obtener un tipo de cambio de moneda, se agregó de la página datos.gob.ar una tabla con tipos de cambios. Para poder relacionar dicha tabla con 'transactions', se estableció una relación física de tipo left join.

customers



De esta manera, se obtuvo una tabla que contiene los datos personales de los clientes y su nivel socioeconómico y por otro lado, una tabla que contiene las ventas realizadas, algunos pocos datos del vendedor y el tipo de cambio de moneda. Para unir estas dos tablas importantes se relacionaron de manera lógica.

customers



Transformaciones del dataset

Las transformaciones efectuadas fueron las siguientes:

- ➔ En las tablas en las que se estableció una relación física, como sucede con la conexión entre 'customers' y 'customers purchasing power', se optó por ocultar la columna 'Customer ID' en la segunda tabla con el fin de evitar redundancias.
- ➔ Lo mismo sucede con la conexión entre las tablas 'transactions' y 'merchants', donde la columna 'Merchant ID' de la segunda tabla también se optó por ocultar.
- ➔ Es el mismo caso con la conexión de las tablas 'transactions' y la tabla de tipos de cambio de moneda, en el cual, en la segunda tabla, se ocultaron todas las columnas que no iban a ser utilizadas, en este caso, la columna seleccionada para unirse a 'transaction' es 'tipo cambio implícito en adrs', la cual fue renombrada como 'valor dolar'.
- ➔ En la tabla 'Customers' en la columna 'Gender' se modificaron los valores nulos, por el valor 'Other'.

Métricas y Dimensiones

Las medidas calculadas que se usaron para este dataset son:

- 'Age': gracias a la fórmula DATEDIFF, se calculó la diferencia entre la fecha de nacimiento y la fecha actual, para obtener la edad de cada cliente.

```
DATEDIFF('year', [Birthdate], TODAY())
```

- 'Transaction ARS': a través de esta métrica podemos obtener el monto de las ventas en pesos argentinos para poder usarlos en el parámetro 'Currency'.

```
IF [Currency] = 'ARS' THEN [Transaction Amount] ELSE [transactions USD] END
```

- 'Transaction USD': esta métrica fue creada para poder realizar cambio de moneda, de pesos argentinos a dólar.

```
[Transaction Amount]/[dollar value]
```

Las dimensiones calculadas que se usaron en este caso son:

- 'Age Group': se creó un campo que representa grupos de edades, el cual nos permite organizar y categorizar las edades.

```
IF [Age] >= 18 AND [Age] < 26 THEN '18-25'  
ELSEIF [Age] >= 26 AND [Age] < 36 THEN '26-35'  
ELSEIF [Age] >= 36 AND [Age] < 46 THEN '36-45'  
ELSEIF [Age] >= 46 AND [Age] < 56 THEN '46-55'  
ELSEIF [Age] >= 56 AND [Age] < 66 THEN '56-65'  
ELSEIF [Age] > 65 THEN '+66'  
END
```

- 'Gender Group': se creó este campo para poder analizar la cantidad de clientes por género, a través de la operación lógica IF.

```
IF [Gender] = 'M' THEN 'Male'  
ELSEIF [Gender] = 'F' THEN 'Female'  
ELSE 'Other'  
END
```

- 'Purchasing Power Levels': esta dimensión fué creada para poder ajustar los gráficos dependiendo del poder adquisitivo se quiera visualizar.

- 'Status Customer': dimensión creada para poder agrupar a los clientes como activos o inactivos, dependiendo si habían realizado una transacción o más, o ninguna.

```
IF [Customer ID] = [Customer ID1] THEN 'Active'  
ELSE 'Inactive'  
END
```

Dashboard

El dashboard está compuesto por una portada y dos pestañas, las cuales están destinadas al análisis del cliente y la siguiente está destinada al análisis de la evolución de las ventas. Presentan la misma gama de colores, pero tratando de diferenciarse uno del otro.

Portada

Consta de dos botones, los cuales permiten dirigirse a la pestaña 'Customers Analysis' y a la pestaña 'Sales Analysis'.

E-commerce analysis



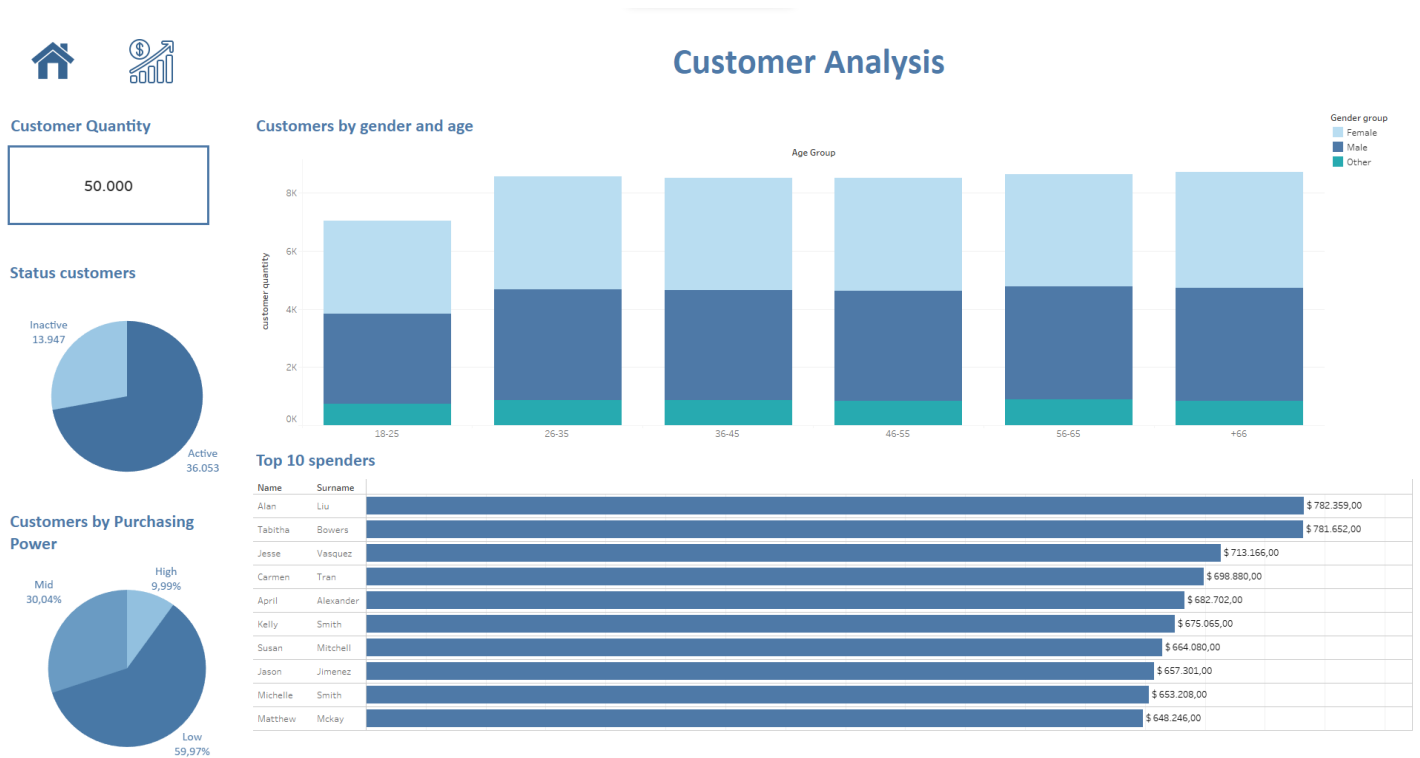
Customer Analysis



Sales Analysis

Customers Analysis

La pestaña 'Customers Analysis' presenta dos botones en el extremo superior izquierdo, uno para volver al home y otro para poder acceder al análisis de las ventas.



A su vez presenta:

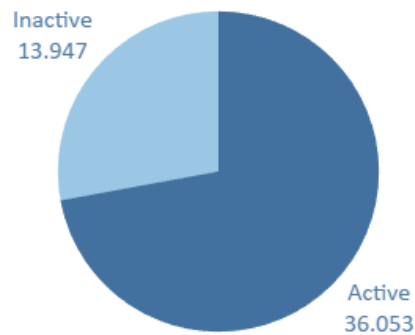
→ 'Customer Quantity': indica la cantidad total de clientes.

Customer Quantity

50.000

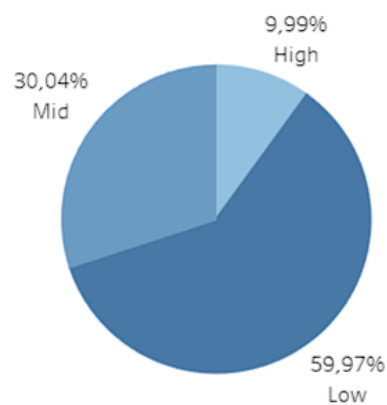
- 'Status customers': es un gráfico de composición, de tipo circular, para poder comparar la porción de clientes activos, es decir aquellos que realizaron una o más transacciones, o inactivos, aquellos que no realizaron ninguna transacción.

Status customers



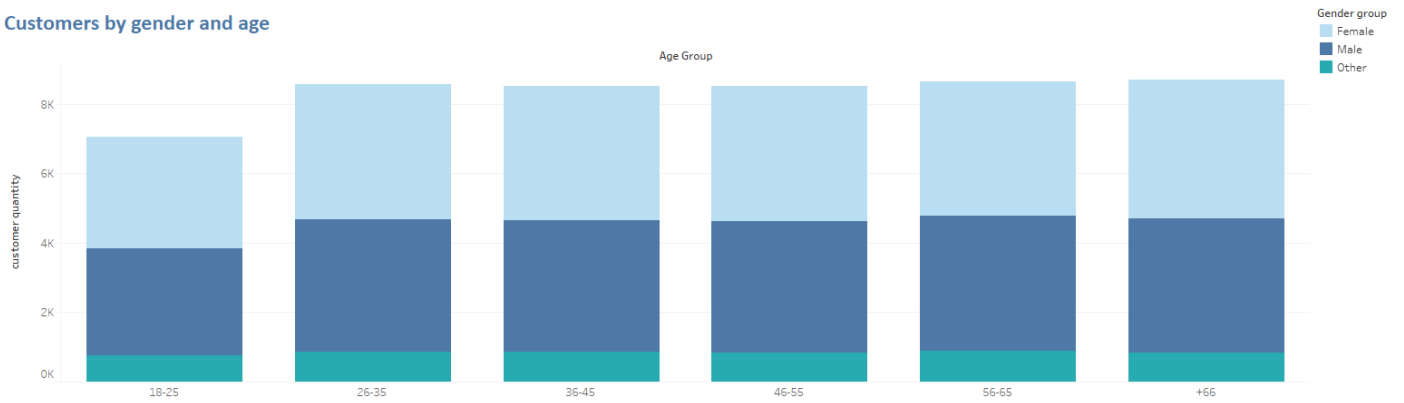
- 'Percentage of customers by purchasing power': es un gráfico de composición, en este caso circular, para poder comparar los distintos porcentajes de los clientes según su poder adquisitivo (alto, medio y bajo).

Percentage of customers by purchasing power



→ 'Customers by gender and age': es un gráfico de comparación, de tipo columnas apiladas, el cual permite comparar la cantidad de clientes por rango etario y género.

Customers by gender and age



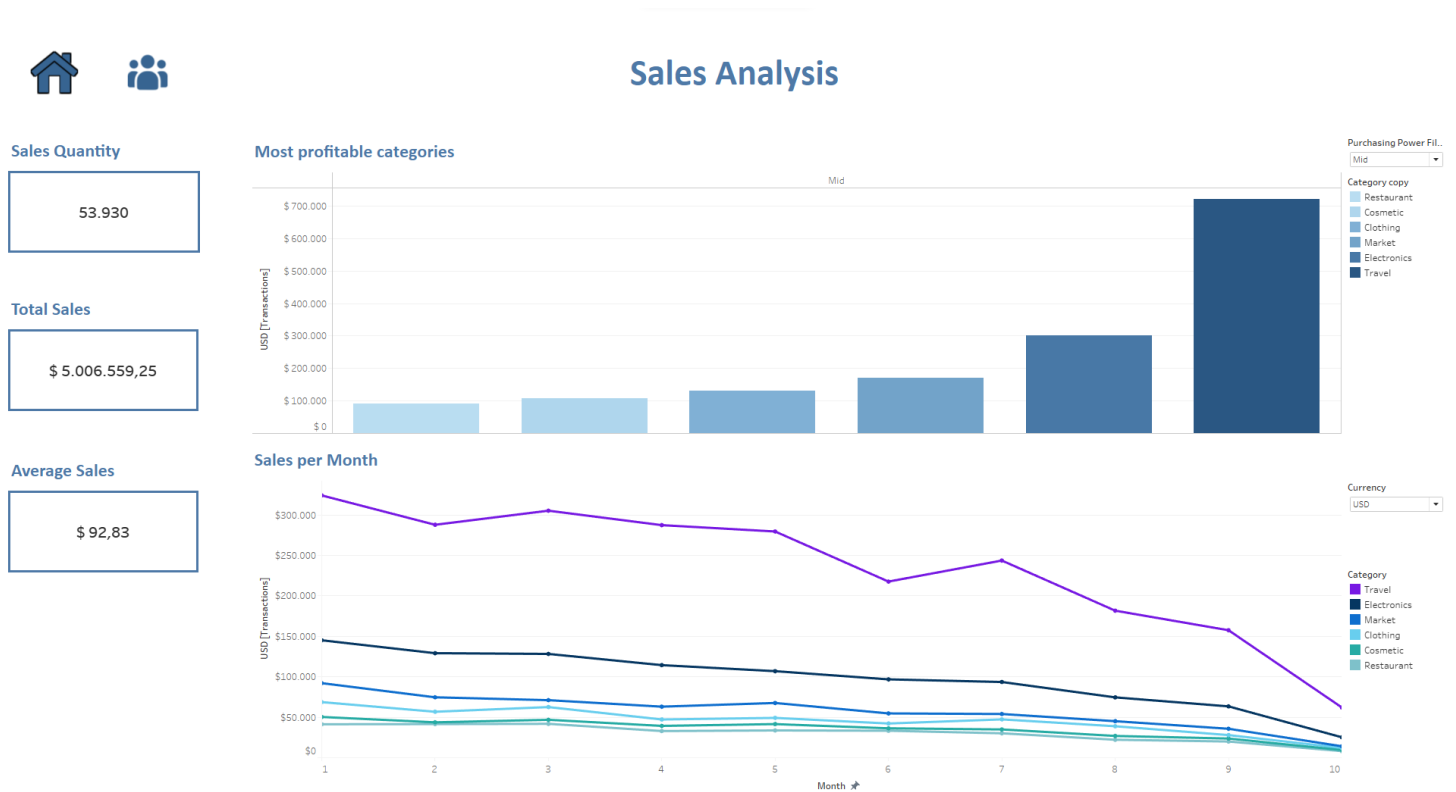
→ 'Top 10 spenders': es un gráfico de comparación, en este caso de barras, filtrado solo por los 10 clientes que más gastaron.

Top 10 spenders

Name	Surname	
Alan	Liu	\$ 782.359,00
Tabitha	Bowers	\$ 781.652,00
Jesse	Vasquez	\$ 713.166,00
Carmen	Tran	\$ 698.880,00
April	Alexander	\$ 682.702,00
Kelly	Smith	\$ 675.065,00
Susan	Mitchell	\$ 664.080,00
Jason	Jimenez	\$ 657.301,00
Michelle	Smith	\$ 653.208,00
Matthew	Mckay	\$ 648.246,00

Sales Analysis

La pestaña 'Sales Analysis' presenta dos botones en el extremo superior izquierdo, uno para volver al home y otro para poder acceder al análisis de los clientes.



Presenta:

→ 'Sales quantity': indica la cantidad de transacciones realizadas.

Sales Quantity

53.930

→ 'Total sales': indica la suma total de transacciones realizadas. Su valor puede cambiar dependiendo de qué moneda se elija.

Pesos ARG:

Total Sales

\$ 2.383.486.655,00

Pesos Dólar:

Total Sales

\$ 5.006.559,25

→ 'Average sales': indica el promedio de las transacciones realizadas. Su valor puede cambiar dependiendo de qué moneda se elija.

Pesos ARG:

Average Sales

\$ 44.195,93

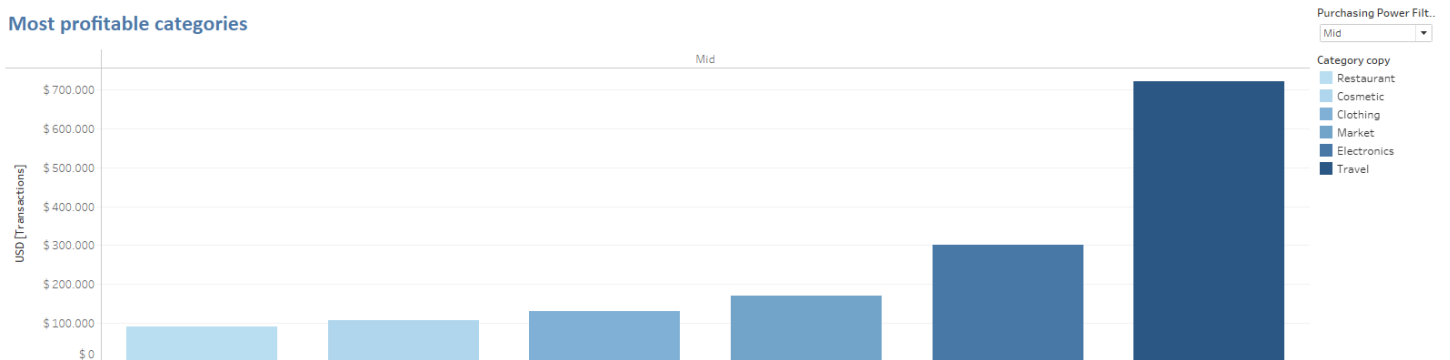
Pesos Dólar:

Average Sales

\$ 92,83

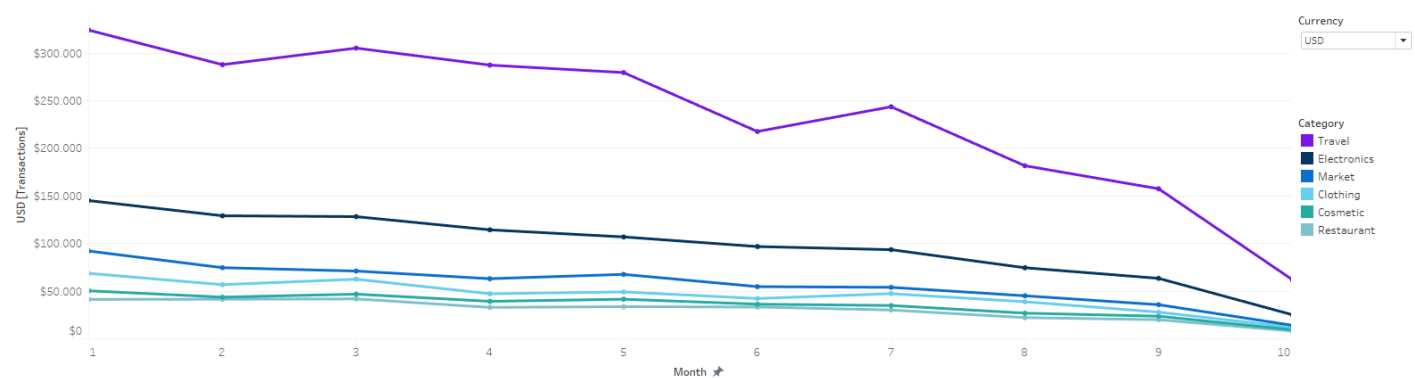
→ 'Most profitable categories': es un gráfico de comparación, de tipo columnas, el cual permite comparar las categorías y poder observar cuales de ellas son las más rentables, filtrados por el poder adquisitivo de los clientes a través del parámetro Purchasing Power. Su valor puede cambiar dependiendo de qué moneda se elija.

Most profitable categories



→ 'Sales per Month': es un gráfico de tendencia, de tipo línea, el cual permite ver cómo varían las ventas a lo largo del año 2023, entre los meses de enero y octubre. Se puede observar su evolución en pesos argentinos y dólar.

Sales per Month



Conclusión

A partir de las hipótesis planteadas en el apartado de clientes, se puede observar que hay un registro de 50.000 clientes, del cual 36.053 clientes (72% aproximadamente) al menos ha realizado una transacción o más de una, mientras que 13.947 clientes (28% aproximadamente) aún no ha realizado alguna transacción. Con respecto al poder adquisitivo de los clientes, se puede observar que el grupo 'Low', es decir aquellos clientes de bajo poder adquisitivo, es el que más consume y a su vez, el grupo 'High', es decir aquellos clientes de alto poder adquisitivo, es el que menos consume. También se puede decir que, desde el gráfico clientes según género y edad, los grupos de género son distribuidos equitativamente según los distintos grupos etarios, a excepción del grupo etario que abarca de los 18 a 25 años de edad, que se encuentra levemente disminuído.

No es un detalle menor que casi el 28% de los clientes no haya hecho una transacción aún, para lo cual se pueden aplicar estrategias comerciales para atraer o convencer a ese porcentaje de clientes de que realicen su primera compra. Como también aplicar más estrategias comerciales para poder atraer tanto a los niveles 'High' y 'Middle' (alto y medio) de poder adquisitivo. También se calculó un top 10 de aquellos clientes que más han gastado, para poder reconocer su lealtad y ofrecerles un beneficio.

Por otro lado, desde la perspectiva de las ventas, se han realizado a lo largo del año 2023, del mes de enero al mes de octubre, 53.930 transacciones, las cuales suman \$2.383.486.655,00 pesos argentinos, con un promedio de \$44.195,93. En los tres niveles de poder adquisitivo, la categoría Travel es la que rinde más ganancias, mientras que Restaurant y Cosmetic son los que presentan menos ganancias. Y de los tres niveles de poder adquisitivo, el grupo 'Low' (bajo poder adquisitivo), es quien genera más ganancias.

El e-commerce a simple vista presenta desde enero a septiembre un balance estable, pero desde octubre descendió drásticamente. Si analizamos el mismo gráfico en dólares podemos apreciar que la empresa no obtuvo ganancias reales, sino más bien se descapitalizó.

Anexo

Tableau: [Dashboard](#)