

INFORME EMPRESARIAL

CALIDAD DE LOS DATOS DE ORIGEN

- Este informe tiene el objetivo de brindar una perspectiva de la calidad de los datos de origen para las tablas correspondientes a **Venta y Compra**.

1. Diccionario de Datos:

- Ventas.

Nombre de la Columna	Descripción
Id_venta	Índice que determina como único a cada registro en venta.
Fecha	Fecha de realización de la venta.
Fecha_entrega	Fecha en que se realiza la entrega del producto.
Id_canal	Identificador del canal de venta, telefónica, online o presencial.
Id_cliente	Identificador del cliente que realiza la compra.
Id_sucursal	Identificador de la sucursal a cargo de la venta.
Id_empleado	Identificador del empleado a cargo de la venta.
Id_producto	Identificador del producto que se ha vendido.
Precio	Precio del producto vendido.
Cantidad	Cantidades del determinado producto que se vendieron.

- Compra.

Nombre de la columna	Descripción
Id_compra	Índice que determina como único a cada registro de compra.
Fecha	Fecha de realización de la compra.
Id_producto	Identificador del producto que se ha comprado.
Cantidad	Cantidades del determinado producto que se compraron.
Precio	Precio del producto comprado.
Id_proveedor	Identificador del proveedor al que se compra el producto

- Gasto.

Nombre de la columna	Descripción
Id_gasto	Índice que determina como único a cada registro de gasto.
Id_sucursal	Identificador de la sucursal a cargo del gasto.
Id_tipo_gasto	Identificador del tipo de gasto realizado.
Fecha	Fecha de realización del gasto.
Monto	Costo del gasto realizado.

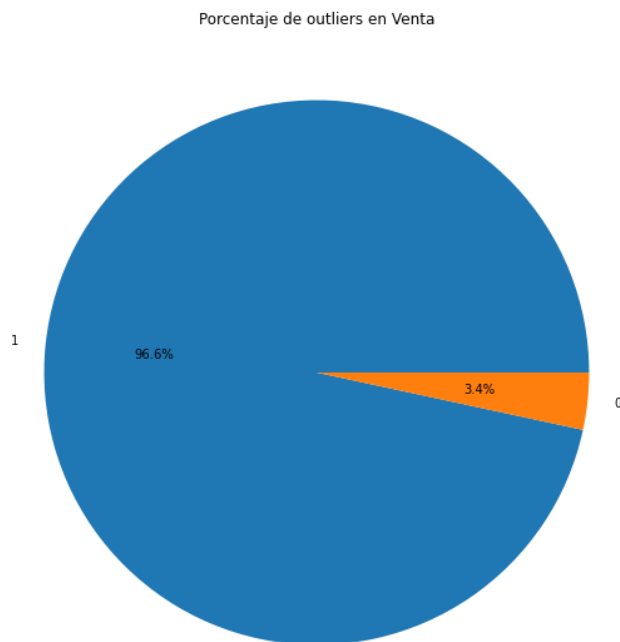
2. Informe grafico de calidad del dato de origen.

OUTLIERS DE VENTA.

- Cantidad de outliers en venta

Outliers	1576
No Outliers	45069

- Porcentaje de Outliers en Venta



Nota: Los datos clasificados como outliers están representados con el numero cero y el color naranja, los datos no outliers con el numero 1 y el color azul.

Los 1576 elementos clasificados como outliers representan un 3.4% del total de los datos.

Más adelante en este análisis se determina una solución arbitraria al respecto de ese 3.4% de valores outliers, sin embargo, ahora se puede decir que con la cantidad de datos inicial, no se clasifica como significativo para el correcto análisis estadístico.

- **Gráfico de caja, visualización de outliers de Precio por Canal de Venta:**

Grafico de caja, visualización de outliers de Precio por Canal de Venta

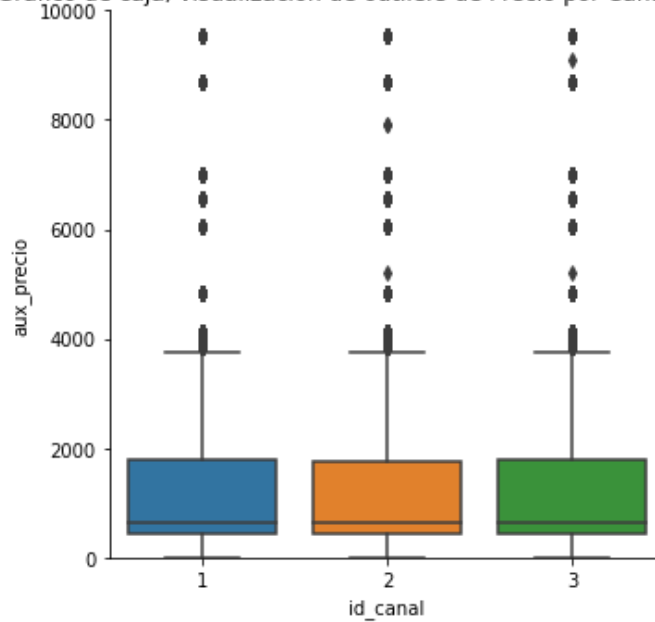
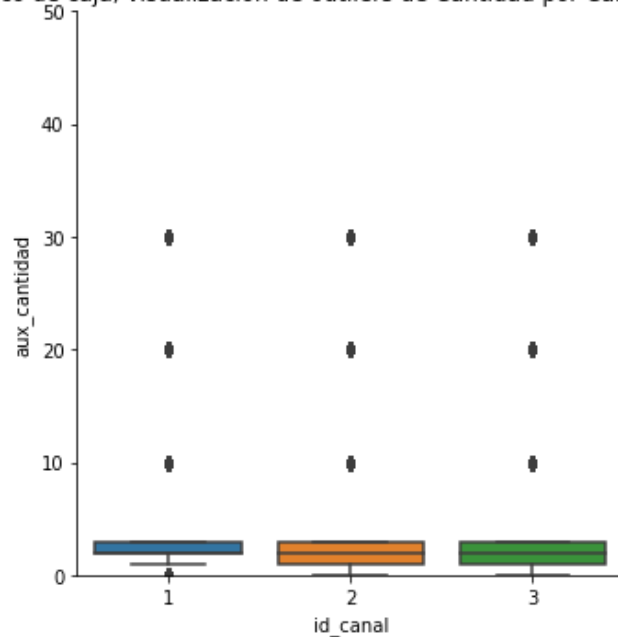


Grafico de caja, visualización de outliers de Cantidad por Canal de Venta



Nota: Los datos se clasifican en función del canal de venta correspondiente a compra, donde 1 es telefónica, 2 es online y 3 es presencial.

La primera grafica corresponde a la distribución de los datos de Precio de Venta y la segunda a la de Cantidad de Venta.

La distribución de los outliers por Canal de Venta es similar, más, el objetivo de la caracterización es visualizar de manera mas eficiente el comportamiento de los outliers.

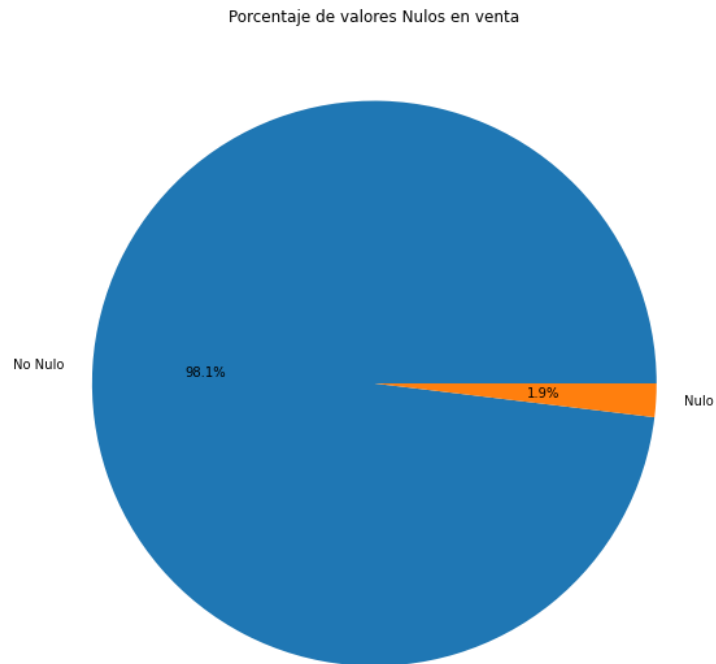
VALORES NULOS EN VENTAS

- Cantidad de valores Nulos en Venta

Nulos	892
-------	-----

No Nulos	45753
----------	-------

- **Porcentaje de valores Nulos en Venta:**



Nota: Los datos clasificados como nulos están representados con el color naranja, los datos no nulos con el color azul.

Los 892 elementos clasificados como nulos representan un 1.9% del total de los datos.

Los valores nulos en venta tampoco parecen catalogar como un problema para la continuidad del análisis, sin embargo, adelante se propone una alternativa a la ausencia de estos valores en venta, ya que al familiarizar con la naturaleza de los datos se puede fijar en alternativas alrededor de producto y su identidad, como su valor relacional o las cantidades representativas de venta.

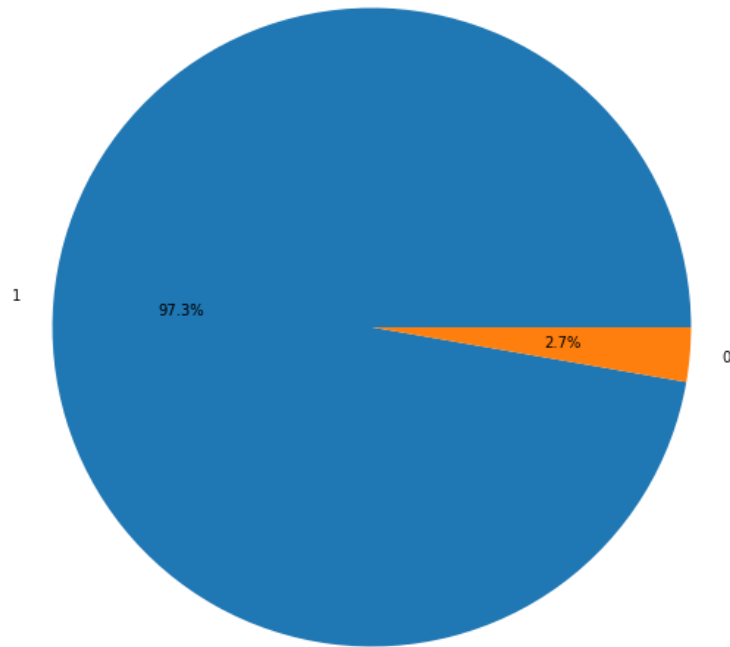
OUTLIERS EN COMPRA

- **Cantidad de outliers en compra**

Outliers	317
No Outliers	11222

- **Porcentaje de outliers en Compra:**

Porcentaje de outliers en Compra

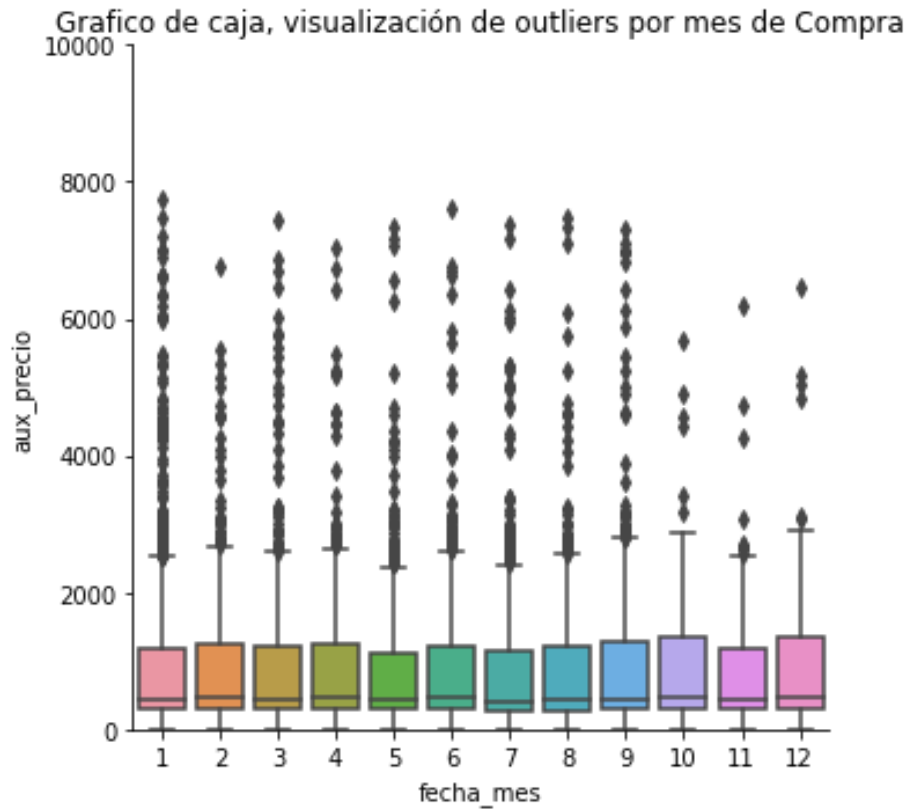


Nota: Los datos clasificados como outliers están representados con el numero cero y el color naranja, los datos no outliers con el numero 1 y el color azul.

Los 317 elementos clasificados como outliers representan un 2.7% del total de los datos.

Las alternativas al intentar solucionar un problema de outliers en las compras, pasan por la ausencia de una referencia a la hora de determinar el valor de la compra, en función de eso se examinan cuales son los comportamientos de los datos, tanto en precio como en cantidad, todo con el fin de no llegar a generar un sesgo sobre el análisis final y alejarse del concepto de la compra para la empresa.

- **Gráfico de caja, visualización de outliers por mes de Compra:**



Nota: En función de una mejor visualización y comprensión del comportamiento de los valores outliers en las compras, se categoriza en meses del año, no para demostrar su comportamiento a lo largo del año, sino para entender la distribución de los datos. Cada numero representa el correspondiente mes del año en orden ascendente.

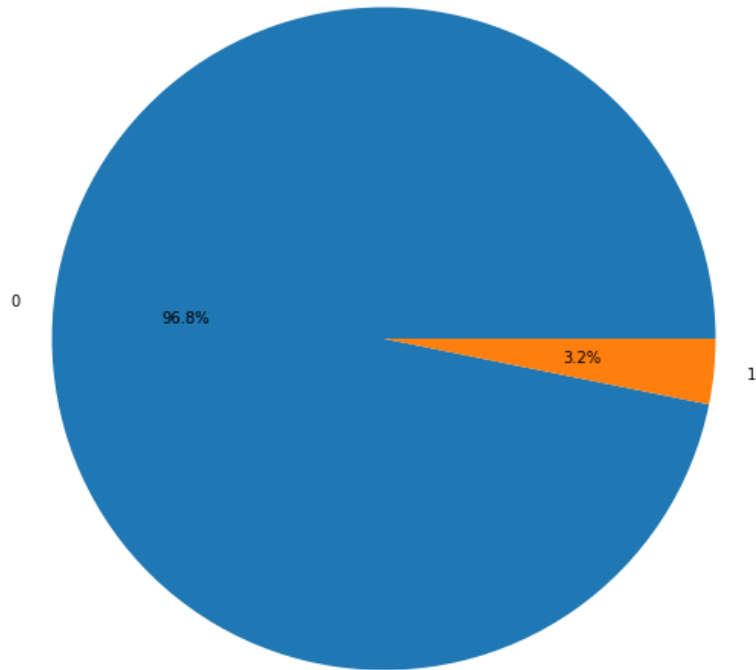
VALORES NULOS EN COMPRAS

- **Cantidad de valores Nulos en Compra:**

Nulos	367
No Nulos	11172

- **Porcentaje de valores Nulos en Compra:**

Porcentaje de valores Nulos en Precio de Compra



Nota: Los datos clasificados como nulos están representados con el numero cero y el color naranja, los datos no nulos con el numero 1 y el color azul.

Los 367 elementos clasificados como nulos representan un 3.2% del total de los datos.

Al avanzar en el análisis de la calidad del dato se nota como entre valores outliers y nulos la cifra de datos a corregir o depurar aumenta, la propuesta alrededor de ello se determina en función de no arriesgar a la pérdida de más información de la esperada.

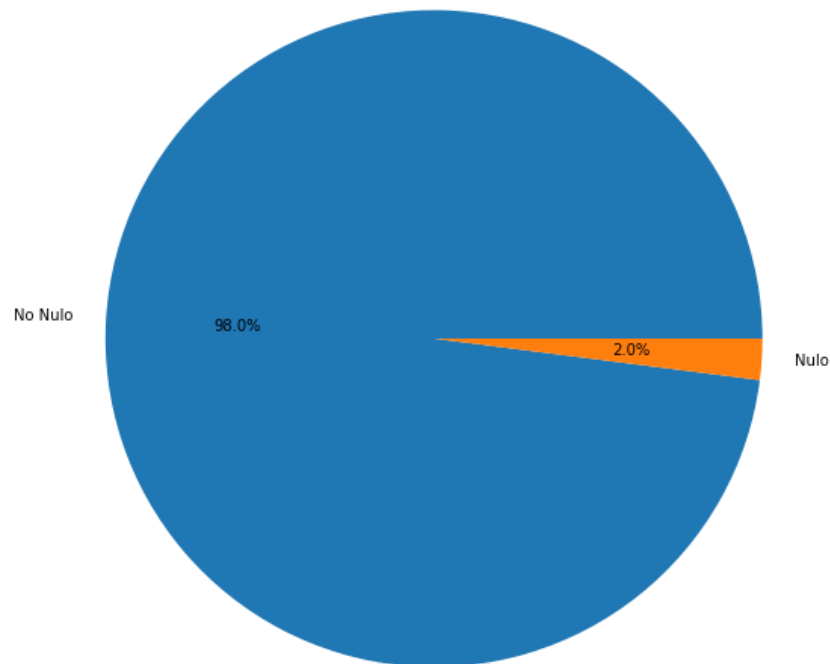
VALORES NULOS EN CLIENTE

- Cantidad de valores nulos en cliente**

Nulos	54
No Nulos	2618

- Porcentaje de valores Nulos en provincia de cliente**

Porcentaje de valores Nulos en cliente de cliente



Nota: Los datos clasificados como nulos están representados con el color naranja, los datos no nulos con el color azul.

Los **54** elementos clasificados como nulos representan un **2.0%** del total de los datos.

La naturaleza de estos datos nulos es categórica en algunos casos se llega a determinar la ausencia total del dato y en otros se puede valer de la referenciación de otras fuentes para determinar el posible valor correspondiente.

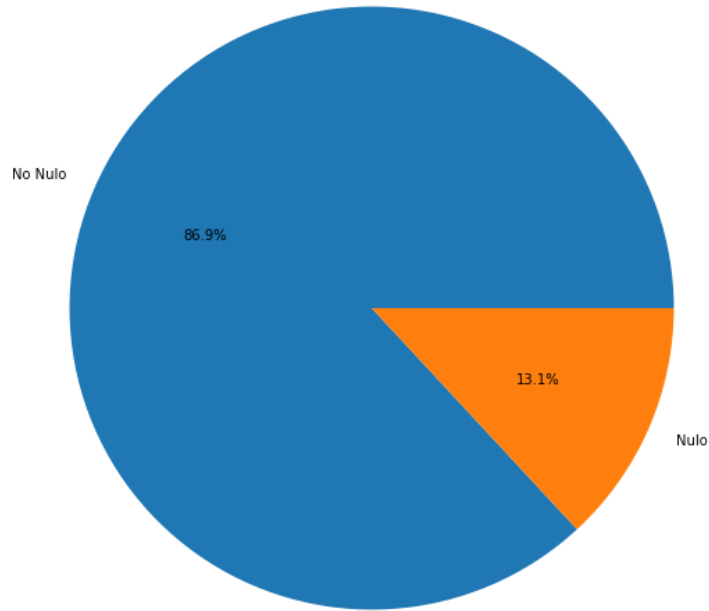
VALORES NULOS EN LOCALIDAD

- Cantidad de valores Nulos en localidad**

Nulos	543
No Nulos	3599

- Porcentaje de valores Nulos en localidad**

Porcentaje de valores Nulos en localidad



Nota: Los datos clasificados como nulos están representados con el color naranja, los datos no nulos con el color azul.

Los **543** elementos clasificados como nulos representan un **13.1%** del total de los datos.

Este set de datos en particular tiene la característica de albergar una gran cantidad de información de cada identificador de localidad, por tanto se cae en el problema de perder en el camino información que detalla a fondo la localidad, sin embargo, en función de suplir de información al resto del contenido no representa un peligro para el análisis.

VALORES NULOS EN PROVEEDOR

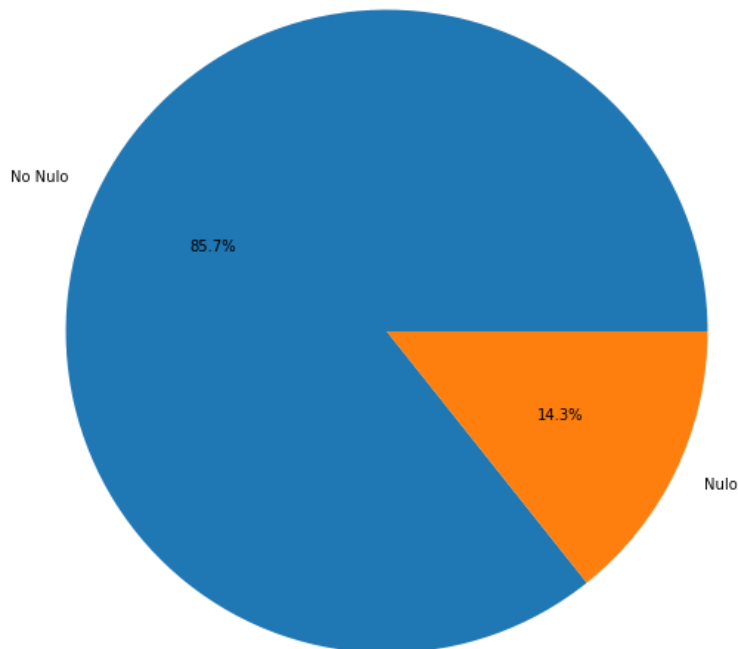
- **Cantidad de valores Nulos en proveedor**

Nulos	2
No Nulos	12

Dos datos nulos que pertenecen a la columna nombres.

- **Porcentaje de valores Nulos en proveedor**

Porcentaje de valores Nulos en proveedor



Nota: Los datos clasificados como nulos están representados con el color naranja, los datos no nulos con el color azul.

Los **2** elementos clasificados como nulos representan un **14.3%** del total de los datos.

La ausencia de estos elementos se escapa a alternativas de subsanación, por tanto se determina simplemente cambiar a la 'Sin Dato' los dos espacios vacíos.

PROPUESTA PARA SUBSANAR VALORES NULOS

Dependiendo de la columna y su tabla correspondiente se opta por determinadas soluciones, en función de llenar los valores nulos.

- Venta

Nombre de columna	Determinación de valores nulos
Precio	Se ingesta la tabla tipo producto, con la cual se igualan los precios en función del id_producto .
Cantidad	Se determina llenar los valores nulos con el valor correspondiente a la mediana de la columna.

- Compra

Nombre de columna	Determinación de valores nulos
-------------------	--------------------------------

Precio	Se determina dentro de la tabla producto el valor aproximado del precio de compra.
Cantidad	Se determina llenar los valores nulos con el valor correspondiente a la mediana de la columna.

En cuanto a las demás columnas dentro del Dataset con valores faltantes, al ser categóricas se determina llenar los valores nulos con la premisa ***'Sin Dato'***.

En cuanto a las demás columnas de tipo numerico / categorico, se opta por remplazar los valores nulos por **ceros**.

PROPUESTA PARA OUTLIERS

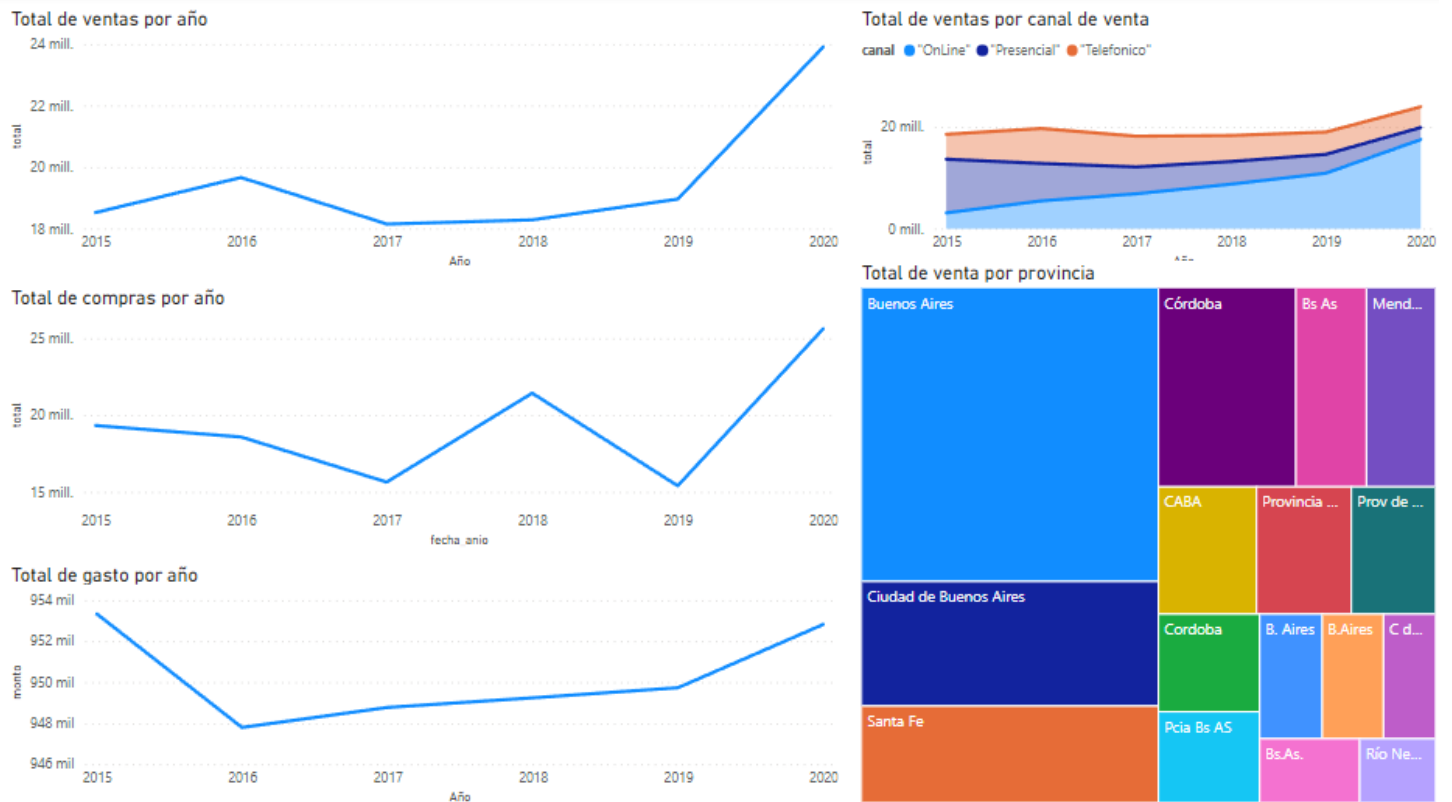
- Venta.

Nombre de columna	Determinación de outliers
Precio	Se toman los valores de la tabla producto para determinar los precios.
Cantidad	Se determina cambiar los valores outliers por el valor de la mediana de la columna.

- Compra.

Nombre de columna	Determinación de outliers
Precio	Se toman los valores de la tabla producto, para valores catalogados como outliers.
Cantidad	Se determina cambiar los valores outliers por el valor de la mediana de la columna.

KPI



Esta serie de KPI arrojan luz sobre el comportamiento de los tres principales segmentos de la empresa, venta, compra y gasto.

Las tres graficas de la izquierda representan el comportamiento de los mismos en el periodo de tiempo de estudio. Un comportamiento similar es el aumento en el ultimo tramo, explicado en el la siguiente grafica **Total de ventas por canal de venta**, donde se observa la consolidación y ascenso del canal de venta online en 2020.

Por último, se grafica la distribución de las ventas en el territorio nacional, donde el sesgo de distribución refleja la importancia de **Buenos Aires** en las ventas.



Además, al contrastar la información de clientes con la de sucursales, se podría a partir de las cifras originales determinar y recomendar la apertura de una nueva sucursal en la ciudad Mendoza, pues en la relación cantidad de clientes con cantidad de sucursales, Mendoza es la ciudad mas cercana a los estándares que maneja la ciudad de Buenos Aires. Cosa que podría cambiar con la ingesta de nuevos datos.

- **Mapa de sucursales**

