

PROYECTO PERSONAL

ANALISIS VENTAS MINORISTAS

INDICE

| | |
|---|-----------|
| PROYECTO PERSONAL | |
| ANALISIS VENTAS MINORISTAS..... | 1 |
| INFORME EJECUTIVO..... | 2 |
| Introducción:..... | 2 |
| Descripción de la temática de los datos:..... | 3 |
| Objetivo del Proyecto:..... | 3 |
| Alcance del Proyecto:..... | 3 |
| Nivel de Aplicación (Operativo/Táctico):..... | 5 |
| Observaciones:..... | 6 |
| INFORME TECNICO..... | 7 |
| Introducción:..... | 7 |
| Detalles del Dataset:..... | 8 |
| Normalización de Datos en SQL:..... | 11 |
| Explicación del Diagrama Obtenido en SQL:..... | 12 |
| Organización de Datos en POWER BI:..... | 15 |
| Explicación del Diagrama Obtenido en POWER BI:..... | 17 |
| Herramientas Utilizadas:..... | 19 |
| ANEXO..... | 20 |
| DRIVE | |
| GIT..... | 20 |

INFORME EJECUTIVO

Introducción:

- **Análisis Ventas Minoristas**
 - El presente informe proporciona un contexto básico del conjunto de datos analizado, el alcance del análisis, el objetivo inicial y el público objetivo al que está dirigido, dicho conjunto de datos de ventas minoristas, se encuentra disponible en [Kaggle](#).
Por último se brindarán algunas conclusiones a partir de un tablero interactivo creado en Power BI.

PROYECTO PERSONAL

Descripción de la temática de los datos:

- **“Sample Sales Data”**
 - Este dataset¹ contiene información de una empresa minorista dedicada a la venta de vehículos y transportes varios; algunas variables son la línea del producto (autos clásicos, motocicletas, aviones, etc.), la empresa compradora, la cantidad de vehículos comprados, el tamaño del acuerdo, etc.

Objetivo del Proyecto:

1. El objetivo del proyecto es proporcionar un análisis enfocado en el rendimiento de las ventas por año, territorio, país, cliente y productos. El análisis incluye:
 - a. Ventas por año y por país teniendo en cuenta ingresos brutos, costos y ganancias brutas.
 - b. Rendimiento de los productos y su estado (enviado, cancelado, etc.) por territorio y país.
 - c. Participación de los clientes en función del tamaño del acuerdo, sus productos comprados, ingresos brutos generados, etc.
 - d. Factores que influyen en las ventas y en el tamaño de los acuerdos.

Alcance del Proyecto:

1. **Cantidad de Registros:**
 - a. Posee una cantidad total de ‘2.823’ filas.
 - b. Posee una cantidad total de ‘25’ columnas, de las cuales se han utilizado ‘16’ debido a que las restantes no quitan valor al análisis.
2. **Variables Principales:**
 - a. **Ventas**
 - i. Esta variable es la más significativa.
 - b. Fecha
 - c. Línea del Producto
 - d. País
 - e. Empresa Cliente

¹ **Dataset:** conjunto de datos almacenados que poseen relación entre sí y que, generalmente, se encuentran en forma de tablas (filas y columnas).

PROYECTO PERSONAL

3. Rango Temporal:

- a. El dataset abarca un rango temporal desde el 01 de enero del 2003 hasta el 31 de mayo del 2005.

4. Análisis a Realizar:

- a. El análisis se enfocará en los siguientes puntos clave:
 - i. Ventas por año y país:
 1. Se analizarán las ventas a lo largo del tiempo y en los diferentes países a fin de descubrir patrones que ayuden en la toma de decisiones.
 2. Se presentarán en forma de etiqueta los ingresos brutos, costos y las ganancias brutas en cada lugar y año brindando información extra en función de las segmentaciones entre gráficos.
 - ii. Rendimiento por producto:
 1. Se analizarán los productos por cantidad vendida y el estado de los mismos (enviado, cancelado, etc.) por país y/o región a fin de poder observar el comportamiento por producto en función de diferentes segmentaciones.
 - iii. Clientes y acuerdos:
 1. Se analizará a los clientes a través del tamaño de los acuerdos, la cantidad y los productos que compran y la ganancia generada.
 - iv. Factores que influyen en las ventas:
 1. Se analizará a través de un gráfico de elementos influyentes, las variables que aumentan o disminuyen las ventas, a partir del tamaño de los acuerdos, el país, el MSRP (precio recomendado por fabricante) y el producto comprado.
 - v. Factores que influyen en los acuerdos:
 1. Se analizará a través de un gráfico de elementos influyentes diferentes variables (Cliente, Línea del producto, Total de órdenes, País, etc.) que aumentan las posibilidades de firmar un tamaño de acuerdo determinado.

5. Limitaciones:

- a. Solamente se cuenta con información hasta el 31 de mayo del 2005, por lo que algunos análisis realizados sobre este mismo año podrían no ser los esperados.
- b. El dataset posee solamente el costo por unidad, por lo que se desconoce completamente algunos costos adicionales (como impuestos, salarios, mantenimiento, etc.), los cuales servirán para evidenciar gráficamente la ganancia neta de la empresa.
 - i. Sería recomendable en el futuro incluir otros costos o impuestos que puedan afectar en los ingresos.

Nivel de Aplicación (Operativo/Táctico):

- Usuario objetivo

- Dicho análisis está dirigido a los responsables del área de ventas, quienes podrán utilizar un tablero operativo para optimizar su toma de decisiones y mejorar las estrategias empleadas.

PROYECTO PERSONAL

Observaciones:

1. Se observó que los clientes que más **ganancia bruta** generaron son:
 - a. Euro Shopping Channel (+100 mil)
 - i. Ubicada en España.
 - b. Mini Gifts Distributors Ltd. (+100 mil)
 - i. Ubicada en USA.
1. a. Sería recomendable explorar la posibilidad de ofrecerles tamaños de acuerdos personalizados o descuentos acordes a fin de fortalecer relaciones y su fidelidad a largo plazo.
2. Se observó que los países con más **ingresos brutos** generados (mayor a un millón) son USA, España y Francia; sería recomendable mantener u optimizar la disponibilidad de los productos en estos mercados clave. Además, se sugiere implementar incentivos, como descuentos para nuevos clientes, con el fin de fortalecer la presencia en estas regiones y fomentar el crecimiento a largo plazo.
3. Se ha observado que la línea de productos más vendida son los **Autos Clásicos** seguida de los **Autos Vintage**; se sugeriría explorar la posibilidad de intentar llegar a un acuerdo beneficioso con los proveedores de dichas líneas con el objetivo de reducir costos y por consiguiente aumentar las ganancias.
4. Se observó que el tamaño del acuerdo 'Medium' generó la venta de más de 50 mil productos, obteniendo una ganancia bruta de más de un millón de dólares; sin embargo, los acuerdos denominados como '**Large**' generaron más de **500 mil dólares** con tan solo la venta de **7.414 productos**.
 - a. Es posible que los acuerdos 'Large' incluyan productos de mayor margen de ganancia o condiciones especiales.
 - b. Sería recomendable investigar la posibilidad de ofrecerlos más a menudo a los clientes.
 - c. Se observó en el gráfico de 'elementos influyentes clave' en la página "Factores de Acuerdo" que las posibilidades de pactar un acuerdo '**Large**' aumentan si la línea del producto son **Autos Clásicos**.
5. Se ha observado que el promedio de **ventas aumenta** siempre y cuando:
 - a. El tamaño del acuerdo es '**Large**'.
 - b. La línea del producto es '**Classics Cars**'.
 - c. Cuando el '**MSRP**' (precio recomendado por el fabricante) es alto de por sí.






























INFORME TECNICO

Introducción:

- **Proceso madurativo del proyecto**
 - En este informe se detalla el proceso completo de desarrollo del proyecto, desde la recopilación inicial de datos hasta el análisis final realizado en Power BI. A lo largo del documento, se explican las decisiones tomadas y los pasos realizados en cada fase del proyecto, con el objetivo de proporcionar una visión más amplia y técnica de todo el trabajo realizado.

Detalles del Dataset:

- [ARCHIVO DATASET CSV INICIAL](#)
 - 'sales_data_sample.csv'
- [ARCHIVO DATASET SQL NORMALIZADO](#)
 - 'VentasMinoristas.bacpac'
- [ARCHIVO POWER BI](#)
 - 'VentasMinoristas.pbix'
- **VISUALIZACIÓN DE COLUMNAS DEL DATASET INICIAL**

| | | |
|---|---|---|
|  |  | dbo.sales_data_sample |
|  |  | Columns |
|  | | ORDERNUMBER (nvarchar(50), not null) |
|  | | QUANTITYORDERED (nvarchar(50), not null) |
|  | | PRICEEACH (nvarchar(50), not null) |
|  | | ORDERLINENUMBER (nvarchar(50), not null) |
|  | | SALES (nvarchar(50), not null) |
|  | | ORDERDATE (datetime2(7), not null) |
|  | | STATUS (nvarchar(50), not null) |
|  | | QTR_ID (nvarchar(50), not null) |
|  | | MONTH_ID (nvarchar(50), not null) |
|  | | YEAR_ID (nvarchar(50), not null) |
|  | | PRODUCTLINE (nvarchar(50), not null) |
|  | | MSRP (nvarchar(50), not null) |
|  | | PRODUCTCODE (nvarchar(50), not null) |
|  | | CUSTOMERNAME (nvarchar(50), not null) |
|  | | PHONE (nvarchar(50), not null) |
|  | | ADDRESSLINE1 (nvarchar(50), not null) |
|  | | ADDRESSLINE2 (nvarchar(50), null) |
|  | | CITY (nvarchar(50), not null) |
|  | | STATE (nvarchar(50), null) |
|  | | POSTALCODE (nvarchar(50), null) |
|  | | COUNTRY (nvarchar(50), not null) |
|  | | TERRITORY (nvarchar(50), not null) |
|  | | CONTACTLASTNAME (nvarchar(50), not null) |
|  | | CONTACTFIRSTNAME (nvarchar(50), not null) |
|  | | DEALSIZE (nvarchar(50), not null) |

PROYECTO PERSONAL

• SELECCION DE LAS COLUMNAS PARA TABLAS DEL DIAGRAMA

| | | |
|--|----------|------------|
| dbo.sales_data_sample | ORDENES | PRODUCTOS |
| Columns | CLIENTES | LOCACIONES |
| ORDERNUMBER (nvarchar(50), not null) | | |
| QUANTITYORDERED (nvarchar(50), not null) | | |
| PRICEEACH (nvarchar(50), not null) | | |
| ORDERLINENUMBER (nvarchar(50), not null) | | |
| SALES (nvarchar(50), not null) | | |
| ORDERDATE (datetime2(7), not null) | | |
| STATUS (nvarchar(50), not null) | | |
| QTR_ID (nvarchar(50), not null) | | |
| MONTH_ID (nvarchar(50), not null) | | |
| YEAR_ID (nvarchar(50), not null) | | |
| PRODUCTLINE (nvarchar(50), not null) | | |
| MSRP (nvarchar(50), not null) | | |
| PRODUCTCODE (nvarchar(50), not null) | | |
| CUSTOMERNAME (nvarchar(50), not null) | | |
| PHONE (nvarchar(50), not null) | | |
| ADDRESSLINE1 (nvarchar(50), not null) | | |
| ADDRESSLINE2 (nvarchar(50), null) | | |
| CITY (nvarchar(50), not null) | | |
| STATE (nvarchar(50), null) | | |
| POSTALCODE (nvarchar(50), null) | | |
| COUNTRY (nvarchar(50), not null) | | |
| TERRITORY (nvarchar(50), not null) | | |
| CONTACTLASTNAME (nvarchar(50), not null) | | |
| CONTACTFIRSTNAME (nvarchar(50), not null) | | |
| DEALSIZE (nvarchar(50), not null) | | |

• EXISTENCIA DE NULOS POR COLUMNA

- "ADDRESSLINE2" = 2521
- "STATE" = 1468
- "TERRITORY" = 1074
- "POSTALCODE" = 76

Normalización de Datos en SQL:

- CAMBIOS REALIZADOS CON RESPECTO AL DATASET INICIAL

- Se han removido las columnas:

- “QTR_ID” por falta de utilidad ya que se podrían extraer dichos datos de la columna “ORDERDATE”.
- “MONTH_ID” por falta de utilidad ya que se podrían extraer dichos datos de la columna “ORDERDATE”.
- “YEAR_ID” por falta de utilidad ya que se podrían extraer dichos datos de la columna “ORDERDATE”.
- “ADDRESSLINE1” por falta de utilidad debido a que hay 63 clientes en total y obviamente todos poseen una dirección diferente; es mucho mas optimo analizar por ejemplo por país.
- “ADDRESSLINE2” por falta de utilidad y debido a la basta cantidad de datos faltantes.
- “PHONE” por falta de utilidad; no aporta valor al análisis.
- “STATE” por falta de utilidad ya que la cantidad de datos está reducida a la mitad y solamente cuenta con estados de EEUU.

- MANEJO DE NULOS:

- Variables Categóricas

- “TERRITORY”

- Dicha variable categórica no poseía realmente valores nulos, ya que Pandas interpretó “NA” (NORTH AMERICA) como un valor vacío, por lo que se decidió renombrar todos los registros “NA” por “NORTH AMERICA”.

- “POSTALCODE”

- Se rellenaron los valores faltantes con los códigos postales más comunes correspondientes de dichas ciudades:
 - “SAN FRANCISCO = 94102”
 - “LOS ANGELES = 90001”

PROYECTO PERSONAL

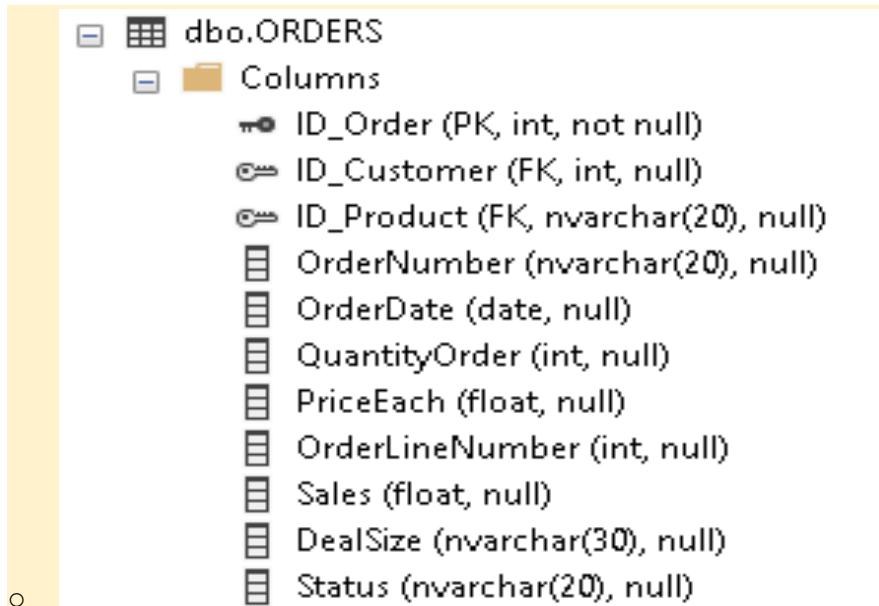
Explicación del Diagrama Obtenido en SQL:

- DESCRIPCIÓN GENERAL

- El diagrama contiene cuatro(4) tablas en total, las cuales son tablas dimensionales que representan relaciones normalizadas, formando así un **esquema copo de nieve**.

- DESCRIPCIÓN DE TABLAS

- ORDERS

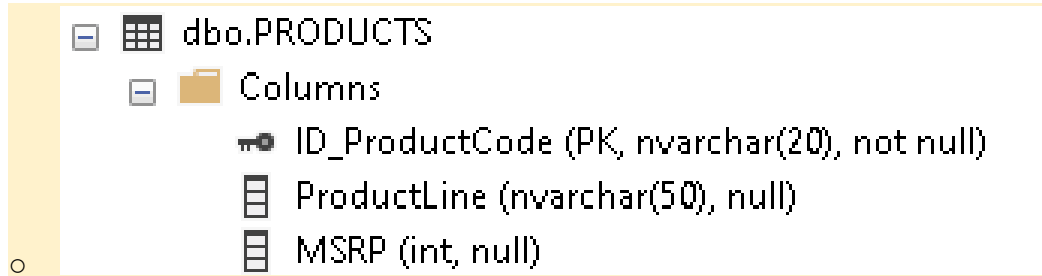


| |
|-------------------------------------|
| dbo.ORDERS |
| Columns |
| ID_Order (PK, int, not null) |
| ID_Customer (FK, int, null) |
| ID_Product (FK, nvarchar(20), null) |
| OrderNumber (nvarchar(20), null) |
| OrderDate (date, null) |
| QuantityOrder (int, null) |
| PriceEach (float, null) |
| OrderLineNumber (int, null) |
| Sales (float, null) |
| DealSize (nvarchar(30), null) |
| Status (nvarchar(20), null) |

- Tabla central que almacena la cantidad total de registros (2.823) y relaciona cada uno de ellos de manera correspondiente con las demás tablas.
 - ID_ORDER = Identificador único del pedido.
 - ID_CUSTOMER = Identificador único del cliente.
 - ID_PRODUCT = Identificador único del producto.
 - OrderNumber = Número de orden del pedido.
 - OrderDate = Fecha realizada del pedido.
 - QuantityOrder = Cantidad ordenada.
 - PriceEach = Precio unitario de cada producto.
 - OrderLineNumber = Número de línea del pedido.
 - Sales = Ventas generadas.
 - DealSize = Tamaño del acuerdo del pedido.
 - Status = Estado del pedido.

PROYECTO PERSONAL

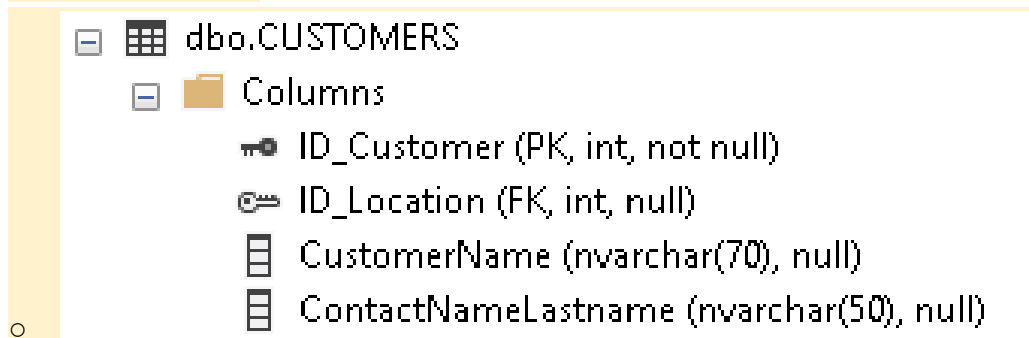
○ PRODUCTS



| |
|---|
| dbo.PRODUCTS |
| Columns |
| ID_ProductCode (PK, nvarchar(20), not null) |
| ProductLine (nvarchar(50), null) |
| MSRP (int, null) |

- Tabla que almacena de manera única todos los productos existentes en el dataset.
 - ID_ProductCode = Identificador único del producto.
 - ProductLine = Línea del producto.
 - MSRP = Precio de venta sugerido por el fabricante.

○ CUSTOMERS



| |
|--|
| dbo.CUSTOMERS |
| Columns |
| ID_Customer (PK, int, not null) |
| ID_Location (FK, int, null) |
| CustomerName (nvarchar(70), null) |
| ContactNameLastname (nvarchar(50), null) |

- Tabla que almacena de manera única todos los datos relacionados a los clientes.
 - ID_Customer = Identificador único del cliente.
 - ID_Location = Identificador único de la ubicación.
 - CustomerName = Nombre de la empresa cliente.
 - ContactNameLastname = Nombre y apellido del contacto del cliente.

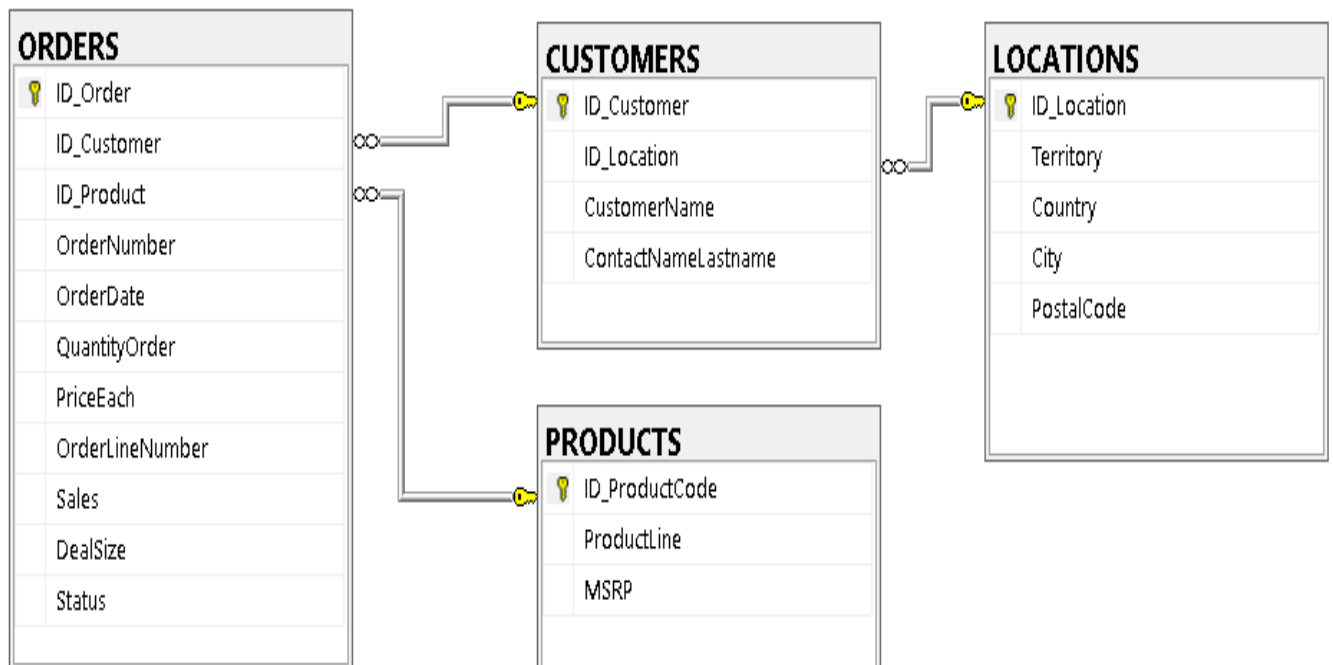
PROYECTO PERSONAL

LOCATIONS

| |
|---------------------------------|
| dbo.LOCATIONS |
| Columns |
| ID_Location (PK, int, not null) |
| Territory (nvarchar(50), null) |
| Country (nvarchar(50), null) |
| City (nvarchar(50), null) |
| PostalCode (nvarchar(20), null) |

- Tabla que almacena de manera única todas las ubicaciones halladas en el dataset.
 - ID_Location = Identificador único de la ubicación.
 - Territory = Territorio donde se encuentra el cliente.
 - Country = País donde se encuentra el cliente.
 - City = Ciudad donde se encuentra el cliente.
 - PostalCode = Código postal del cliente.

DIAGRAMA RESULTANTE EN SQL:



PROYECTO PERSONAL

Organización de Datos en POWER BI:

● CAMBIOS REALIZADOS CON RESPECTO AL ESQUEMA COPO DE NIEVE OBTENIDO EN SQL

- Implementación de tablas
 - Tabla “MEDIDAS”
 - Tabla que contiene seis(6) medidas necesarias para la realización del análisis.
 - Tabla “CALENDARIO”
 - Tabla creada a partir de la función “CALENDAR()” que abarca un periodo desde la primera fecha encontrada en el dataset (01 de enero del 2003) hasta el 31 de mayo del 2005; Dicha tabla se conecta con la tabla de “ORDERS” a través de la columna “OrderDate”.
- Columnas agregadas
 - No se han implementado nuevas columnas para dicho análisis.

● MEDIDAS

- TOTAL_VENTAS
 - Medida que calcula el total de ventas a partir de la suma de la variable “Sales”.
 - ```
1 TOTAL_VENTAS = CALCULATE(SUM(ORDERS[Sales]))
```
- COSTOS
  - Medida que calcula el costo total de los pedidos a partir de la cantidad ordenada multiplicada por el precio unitario del producto.
  - ```
1 COSTOS = SUMX(ORDERS, ORDERS[QuantityOrder] * ORDERS[PriceEach])
```
- GANANCIA_BRUTA
 - Medida que calcula la ganancia bruta obtenida a partir de la resta del total de ventas y el costo.
 - ```
1 GANANCIA_BRUTA =
2 VAR ganancias_netas = [TOTAL_VENTAS] - [COSTOS]
3 RETURN IF(ganancias_netas < 1, 0, ganancias_netas)
```
- KPI\_OBJETIVO\_ANUAL
  - Dicha medida es un KPI ficticio creado únicamente para fines académicos y no representa información alguna.

# PROYECTO PERSONAL

```
1 KPI_OBJETIVO_ANUAL = 75300
```

- TOTAL\_ORDENES

- Medida que calcula el total de órdenes a partir de un identificador único; la misma es utilizada a la hora de observar la cantidad de órdenes por cliente, país, territorio, etc.

```
1 TOTAL_ORDENES = CALCULATE(COUNT(ORDERS[ID_Order]))
```

- CANTIDAD\_PRODUCTOS

- Medida que calcula el total de productos vendidos.

```
1 CANTIDAD_PRODUCTOS = CALCULATE(SUM(ORDERS[QuantityOrder]))
```



# PROYECTO PERSONAL

## Explicación del Diagrama Obtenido en POWER BI:

- DESCRIPCIÓN GENERAL

- El diagrama se mantiene idéntico al de “SQL” y solamente se ha implementado la tabla “CALENDARIO” y la tabla “MEDIDAS”, aunque esta última no tiene relación directa con ninguna otra entidad.

- DESCRIPCIÓN DE TABLAS

- MEDIDAS

- Tabla creada para almacenar medidas utilizadas en los diferentes gráficos, etiquetas y tooltips para realizar diversos análisis.

- CALENDARIO

- Tabla creada para almacenar toda la información relacionada a fechas.

- LOCATIONS

- Tabla que almacena de manera única todas las ubicaciones halladas en el dataset.

- CUSTOMERS

- Tabla que almacena de manera única todos los datos relacionados a los clientes.

- PRODUCTS

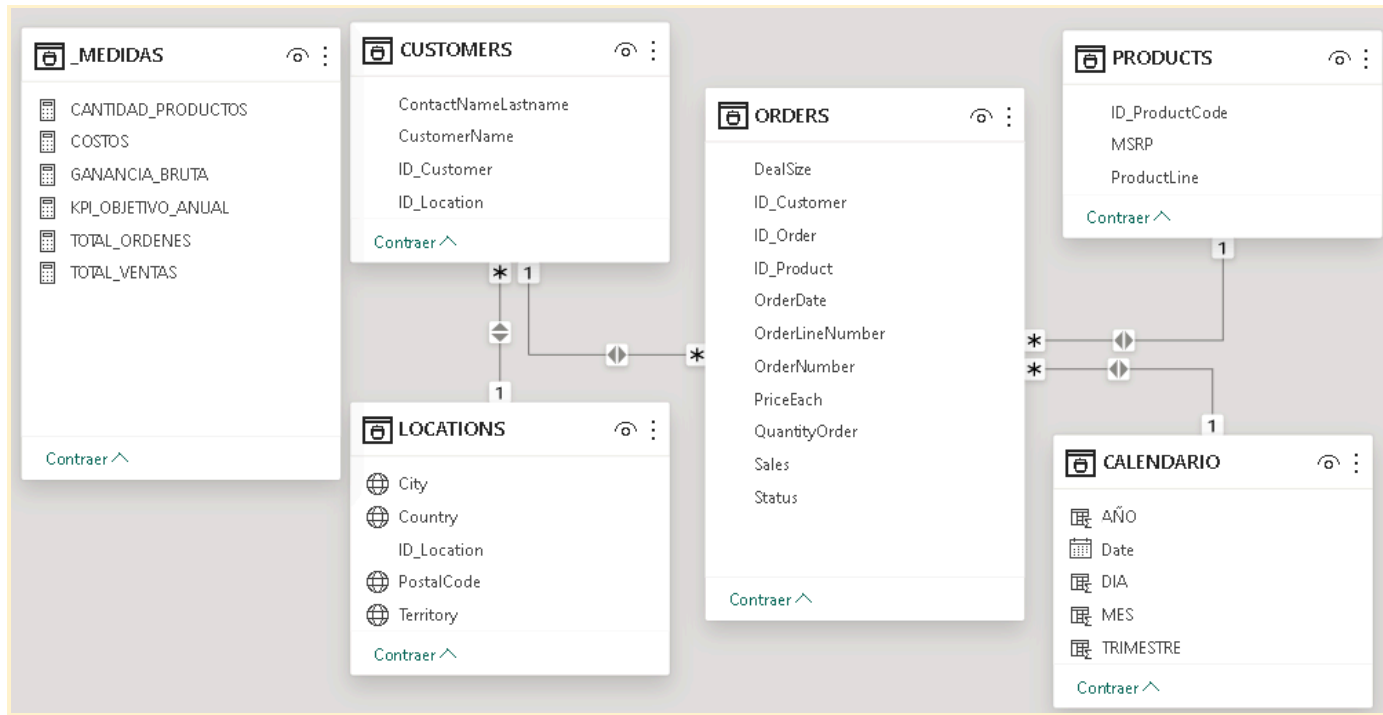
- Tabla que almacena de manera única todos los productos existentes en el dataset.

- ORDERS

- Tabla central que almacena la cantidad total de registros (2.823) y relaciona cada uno de ellos de manera correspondiente con las demás tablas.

# PROYECTO PERSONAL

- DIAGRAMA RESULTANTE EN POWER BI:



## Herramientas Utilizadas:

- Pandas v2.2.2
  - Utilizada para la lectura de registros nulos.
- SQL Server v16.0.1000.6
  - Utilizada para la lectura, limpieza y transformación de datos.
- Power BI v2.124.2028.0 32-bit (diciembre de 2023)
  - Utilizada para la realización del dashboard y el análisis de datos.

**PROYECTO-PERSONAL**

**ANEXO**

[DRIVE](#)  
[GIT](#)