

viu
.es

2023 - 2024



ACTIVIDAD 2

Master en Big Data y Data Science

09MBID – Soluciones de Inteligencia de Negocio

Nombre: GEMMA ARLANDIS ESTEVE

Fecha: 15/01/2024

Contenido

Introducción.....	3
Proceso ETL (Extraer, Transformar, Load/Cargar)	4
Proceso E (Extraer).....	4
Proceso T (Transformar).....	4
Proceso L (Load/Cargar)	9
Visualización	9
Conclusiones.....	11
ANEXO I.....	12
Descripción de los datos	12

Introducción

En esta actividad se va a realizar la implementación de un sistema BI para la empresa Chocolates Valor.

Chocolates Valor es una empresa de producción y comercialización de chocolates con más de 130 años de historia. Su historia empieza en 1881 donde el “Señor Valor” comenzó como chocolatero en el municipio alicantino de la Villajoyosa. A lo largo de su historia, la empresa evolucionó tecnológicamente con la llegada de los diferentes avances tecnológicos en la sociedad, como la llegada de la electricidad. También ha superado desafíos durante la Guerra Civil y las décadas de los 40 y 50, consolidándose con una nueva planta en los años 60, esta planta ha ido ampliándose hasta llegar a ocupar 22.000m2 en la actualidad. La primera chocolatería fue creada en el año 1984, dando lugar a la creación de la primera franquicia en 1993. La empresa destaca por su compromiso con la calidad y la innovación en la producción de chocolates.

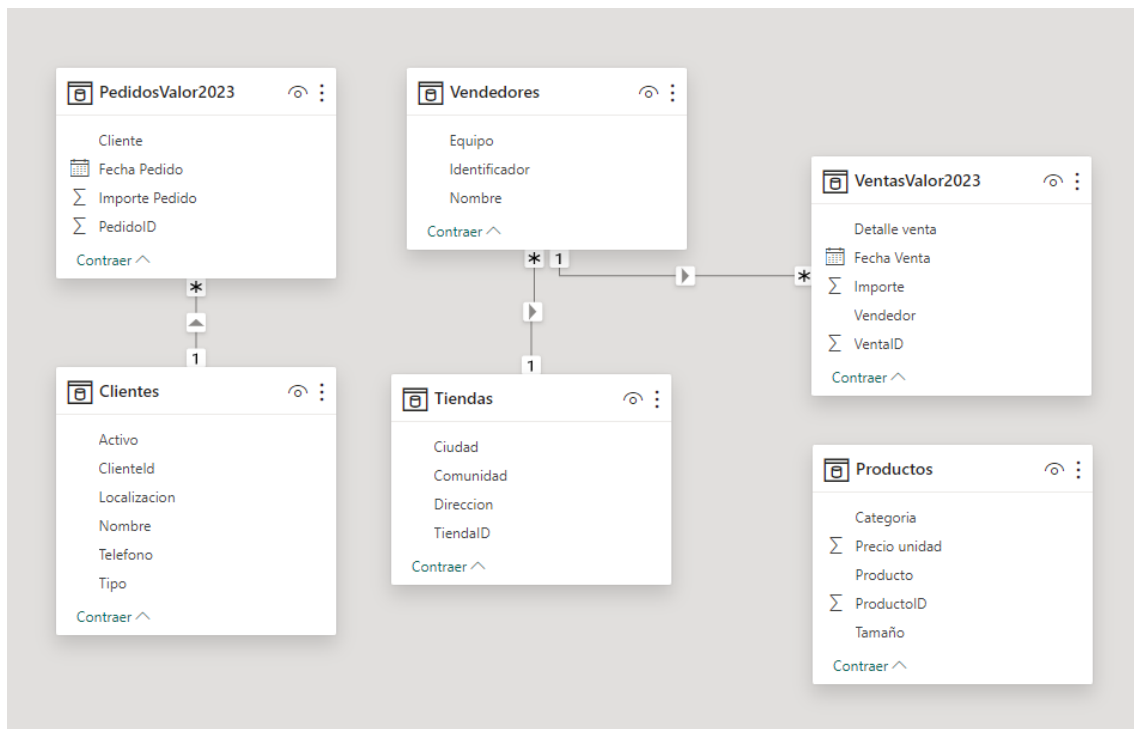
Para la implantación del sistema BI, se ha creado un modelo de datos simulado y simplificado de la empresa. Este modelo de datos será descrito en el punto [Descripción de los datos](#).

Proceso ETL (Extraer, Transformar, Load/Cargar)

Proceso E (Extraer)

Durante el proceso de extracción, se ha realizado la creación de las diferentes fuentes de datos (Descripción de los datos - [ANEXO I](#)). Una vez la base de datos con sus tablas y los ficheros CSV han sido creados, se han cargado en Power BI.

Tras el proceso de carga de los datos, se han administrado las relaciones entre tablas, para el modelo de datos.

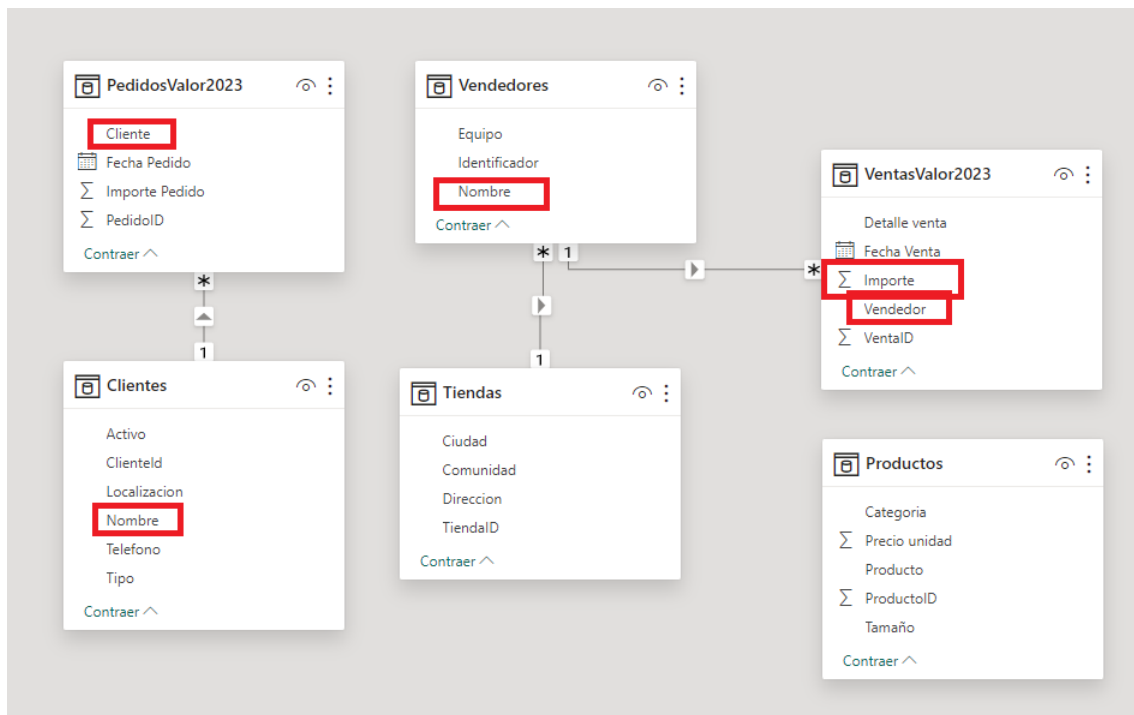


La relación Productos con Ventas no se ha podido realizar desde un primer momento, ya que el detalle de la venta contiene una lista de productos vendidos con sus cantidades.

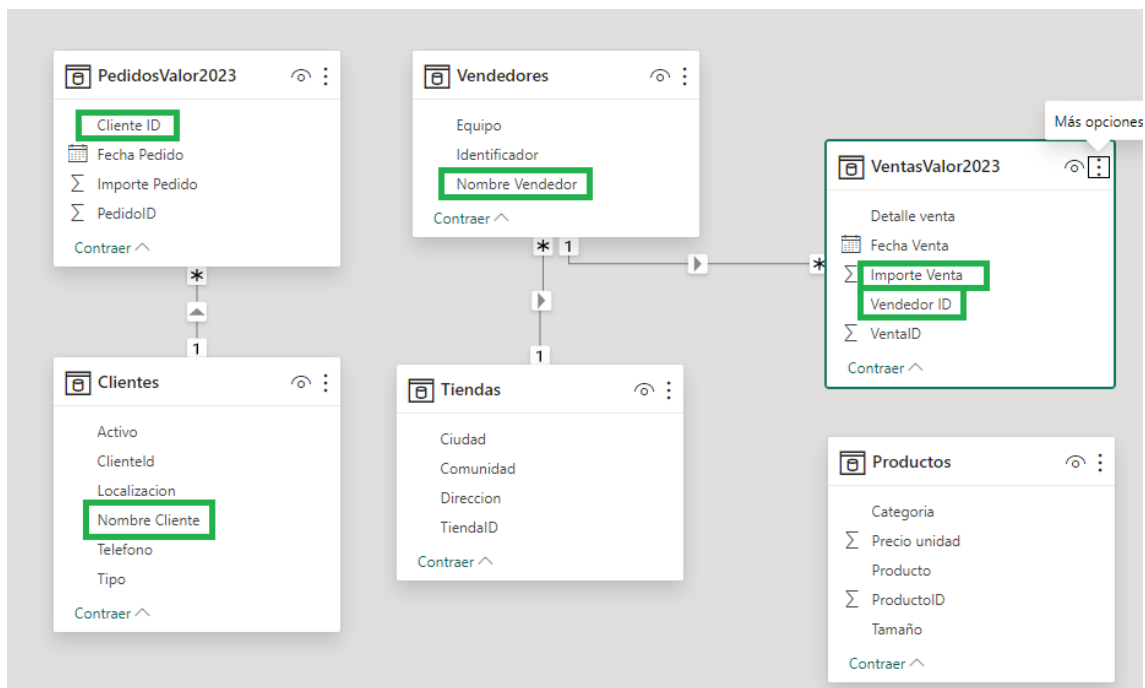
Proceso T (Transformar)

En el proceso de transformación se han realizado diferentes acciones con diferentes herramientas.

Primero se han modificado algunos nombres de columnas, ya que se repetían en diferentes tablas.



Tras estas modificaciones el modelo de datos ha quedado de la siguiente forma:



A continuación, se ha detectado una fila errónea en la tabla VentasValor2023, y con un script Python se ha eliminado.

227	2023227	VAL40	3x6	5,16	29/12/2023
228	2023228	VAL25		null	null

El script Python utilizado es el siguiente:



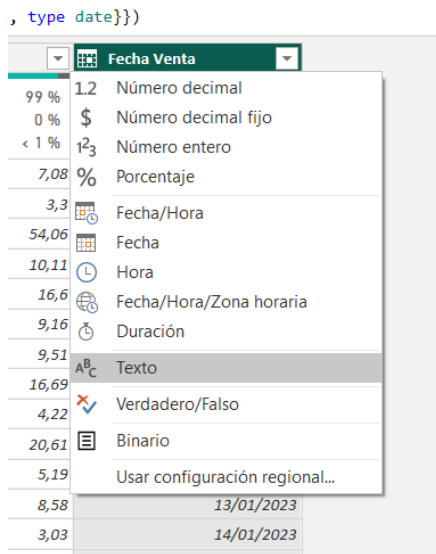
Tras aplicar el script, se ha comprobado que la fila se ha eliminado pero las columnas de fecha e importe han sido modificadas:

213	2023213	VAL30	3x19	535	Microsoft.OleDb.Date
214	2023214	VAL24	2x10	819	Microsoft.OleDb.Date
215	2023215	VAL43	2x19	914	Microsoft.OleDb.Date
216	2023216	VAL6	4x18	505	Microsoft.OleDb.Date
217	2023217	VAL44	3x17	798	Microsoft.OleDb.Date
218	2023218	VAL15	2x26	458	Microsoft.OleDb.Date
219	2023219	VAL39	3x8	646	Microsoft.OleDb.Date
220	2023220	VAL6	3x9	876	Microsoft.OleDb.Date
221	2023221	VAL15	2x27	991	Microsoft.OleDb.Date
222	2023222	VAL31	2x19	643	Microsoft.OleDb.Date
223	2023223	VAL22	4x26	359	Microsoft.OleDb.Date
224	2023224		1x6,1x27,3x23	2072	Microsoft.OleDb.Date
225	2023225	VAL17	4x6	417	Microsoft.OleDb.Date
226	2023226	VAL22	1x21,2x5	1344	Microsoft.OleDb.Date
227	2023227	VAL40	3x6	516	Microsoft.OleDb.Date

ción de oerfiles de columnas basada en las 1000 primeras filas

Por lo que se han tenido que deshacer los cambios producidos del script y realizar una pequeña transformación previa.

Se han modificado las columnas 'Fecha Venta' de tipo Date a tipo Texto:



Después de cambiar el tipo de datos, se ha lanzado de nuevo el script Python.

	Vendedor ID	Detalle venta	Importe Venta	Fecha Venta
206	2023206 VAL1	1x6,2x15	12,04	28/11/2023
207	2023207 VAL20	3x18	3,47	28/11/2023
208	2023208 VAL34	2x25	5,99	30/11/2023
209	2023209 VAL6	1x13,2x15	13,28	04/12/2023
210	2023210 VAL26	2x2	6,73	05/12/2023
211	2023211 VAL46	1x10,1x14,3x11	20,1	07/12/2023
212	2023212 VAL7	3x25	8,2	08/12/2023
213	2023213 VAL30	3x19	5,35	09/12/2023
214	2023214 VAL24	2x10	8,19	10/12/2023
215	2023215 VAL43	2x19	9,14	16/12/2023
216	2023216 VAL6	4x18	5,05	17/12/2023
217	2023217 VAL44	3x17	7,98	18/12/2023
218	2023218 VAL15	2x26	4,58	18/12/2023
219	2023219 VAL39	3x8	6,46	21/12/2023
220	2023220 VAL6	3x9	8,76	22/12/2023
221	2023221 VAL15	2x27	9,91	22/12/2023
222	2023222 VAL31	2x19	6,43	24/12/2023
223	2023223 VAL22	4x26	3,59	25/12/2023
224	2023224	1x6,1x27,3x23	20,72	27/12/2023
225	2023225 VAL17	4x6	4,17	27/12/2023
226	2023226 VAL22	1x21,2x5	13,44	29/12/2023
227	2023227 VAL40	3x6	5,16	29/12/2023

La fila errónea ha sido eliminada, y las fechas siguen siendo fechas y con el tipo de datos correcto al igual que el importe.

Otra transformación aplicada a esta tabla VentasValor2023, ha sido para sustituir los valores vacíos de la columna 'Vendedor ID' por un valor 'KO'.

Table.ReplaceValue("#Tipo cambiado2","", "KO", Replacer.ReplaceValue,{"Vendedor ID"})

	VentaID	A8 Vendedor ID	A8 Detalle venta	1.2 Importe Venta	Fecha Venta
	● Válido 100 % ● Error 0 % ● Vacío 0 %	● Válido 100 % ● Error 0 % ● Vacío 0 %	● Válido 100 % ● Error 0 % ● Vacío 0 %	● Válido 100 % ● Error 0 % ● Vacío 0 %	● Válido 100 % ● Error 0 % ● Vacío 0 %
1	202301	VAL1	2x24	7,08	01/01/2023
2	202302	VAL20	3x17	3,3	02/01/2023
3	202303	VAL14	1x20,1x2,1x17,1x16	54,06	04/01/2023
4	202304	VAL8	1x8,3x21	10,11	05/01/2023
5	202305	VAL41	1x18,2x24	16,6	06/01/2023
6	202306	VAL19	4x27	9,16	07/01/2023
7	202307	VAL24	2x14	9,51	09/01/2023
8	202308	VAL9	1x26,3x6	16,69	10/01/2023
9	202309	VAL29	3x17	4,22	11/01/2023
10	202310	VAL35	1x7,1x21,2x17	20,61	12/01/2023
11	202311	VAL21	3x11	5,19	13/01/2023
12	202312	VAL39	2x13	8,58	13/01/2023
13	202313	KO	4x15	3,03	14/01/2023
14	202314	VAL19	3x23	7,8	14/01/2023
15	202315	VAL15	2x23	7,41	18/01/2023
16	202316	VAL23	4x4	9,28	20/01/2023
17	202317	VAL23	1x14,2x27	14	21/01/2023
18	202318	VAL43	1x7,2x15	19,59	23/01/2023
19	202319	VAL10	3x13	9,31	25/01/2023
20	202320	VAL27	1x18,2x2	14,68	26/01/2023
21	202321	VAL42	3x7	3,92	26/01/2023
22	202322	VAL11	3x13	8,3	26/01/2023
23	202323	VAL6	3x21	8,04	30/01/2023
24	202324	VAL39	2x20	5,17	31/01/2023
25	202325	VAL31	2x5	3,69	31/01/2023

Para seguir con las transformaciones, se ha creado una columna con DAX en la tabla PedidosValor2023 y en la tabla VentasValor2023 para obtener el mes de las transacciones para realizar un análisis mensual posterior.

La fórmula DAX que se ha utilizado es la siguiente:

1	MesPedido2023 = MONTH('PedidosValor2023'[Fecha Pedido])
1	MesVenta2023 = MONTH('VentasValor2023'[Fecha Venta])

Dando como resultado:

PedidoID	Cliente ID	Importe Pedido	Fecha Pedido	MesPedido2023
2023026	6	7166	martes, 24 de enero de 2023	1
2023027	33	872	jueves, 26 de enero de 2023	1
2023028	6	26110	jueves, 26 de enero de 2023	1
2023029	7	14010	jueves, 26 de enero de 2023	1
2023030	21	23895	viernes, 27 de enero de 2023	1
2023031	12	7361	sábado, 28 de enero de 2023	1
2023032	10	11522	sábado, 28 de enero de 2023	1
2023033	3	19079	sábado, 28 de enero de 2023	1
2023034	28	2564	sábado, 28 de enero de 2023	1
2023035	12	17437	domingo, 29 de enero de 2023	1
2023036	3	8642	domingo, 29 de enero de 2023	1
2023037	24	1550	lunes, 30 de enero de 2023	1
2023038	28	3102	martes, 31 de enero de 2023	1
2023039	13	29965	martes, 31 de enero de 2023	1
2023040	2	17586	martes, 31 de enero de 2023	1
2023041	20	28391	miércoles, 1 de febrero de 2023	2
2023042	6	10034	jueves, 2 de febrero de 2023	2
2023043	15	17188	jueves, 2 de febrero de 2023	2
2023044	9	19401	jueves, 2 de febrero de 2023	2
2023045	32	4477	viernes, 3 de febrero de 2023	2
2023046	12	8470	viernes, 3 de febrero de 2023	2
2023047	8	24288	viernes, 3 de febrero de 2023	2
2023048	25	4598	sábado, 4 de febrero de 2023	2
2023049	26	1451	martes, 7 de febrero de 2023	2
2023050	19	8172	viernes, 10 de febrero de 2023	2
2023051	26	4695	viernes, 10 de febrero de 2023	2

VentaID	Vendedor ID	Detalle venta	Importe Venta	Fecha Venta	MesVenta2023
202301	VAL1	2x24	7,08	domingo, 1 de enero de 2023	1
202302	VAL20	3x17	3,3	lunes, 2 de enero de 2023	1
202303	VAL14	1x20,1x2,1x17,1x16	54,06	miércoles, 4 de enero de 2023	1
202304	VAL8	1x8,3x21	10,11	jueves, 5 de enero de 2023	1
202305	VAL41	1x18,2x24	16,6	viernes, 6 de enero de 2023	1
202306	VAL19	4x27	9,16	sábado, 7 de enero de 2023	1
202307	VAL24	2x14	9,51	lunes, 9 de enero de 2023	1
202308	VAL9	1x26,3x6	16,69	martes, 10 de enero de 2023	1
202309	VAL29	3x17	4,22	miércoles, 11 de enero de 2023	1
202310	VAL35	1x7,1x21,2x17	20,61	jueves, 12 de enero de 2023	1
202311	VAL21	3x11	5,19	viernes, 13 de enero de 2023	1
202312	VAL39	2x13	8,58	viernes, 13 de enero de 2023	1
202313	KD	4x15	3,03	sábado, 14 de enero de 2023	1
202314	VAL19	3x23	7,8	sábado, 14 de enero de 2023	1
202315	VAL15	2x23	7,41	miércoles, 18 de enero de 2023	1
202316	VAL23	4x4	9,28	viernes, 20 de enero de 2023	1
202317	VAL23	1x14,2x27	14	sábado, 21 de enero de 2023	1
202318	VAL43	1x7,2x15	19,59	lunes, 23 de enero de 2023	1
202319	VAL10	3x13	9,31	miércoles, 25 de enero de 2023	1
202320	VAL27	1x18,2x2	14,68	jueves, 26 de enero de 2023	1
202321	VAL42	3x7	3,92	jueves, 26 de enero de 2023	1
202322	VAL11	3x13	8,3	jueves, 26 de enero de 2023	1
202323	VAL6	3x21	8,04	lunes, 30 de enero de 2023	1
202324	VAL39	2x20	5,17	martes, 31 de enero de 2023	1
202325	VAL31	2x5	3,69	martes, 31 de enero de 2023	1
202326	VAL34	2x2	8,49	martes, 31 de enero de 2023	1
202327	VAL37	3x25	8,52	lunes, 6 de febrero de 2023	2
202328	VAL21	3x10	8,77	miércoles, 8 de febrero de 2023	2

Por último se han calculado dos medidas. Una medida en la tabla PedidosValor2023 y otra medida en la tabla VentasValor2023. Estas dos medidas serán el total de euros obtenidos en el año 2023.

✕ ✓	1 TotalVentas = SUM(VentasValor2023[Importe Venta])
-----	---

✕ ✓	1 TotalPedidos = SUM(PedidosValor2023[Importe Pedido])
-----	--

Proceso L (Load/Cargar)

Como el modelo de datos ha sido creado expresamente para la práctica, no existe otra base de datos o un Datawarehouse donde almacenar el modelo ya transformado.

Lo que se ha realizado, en cambio, es una visualización.

Visualización

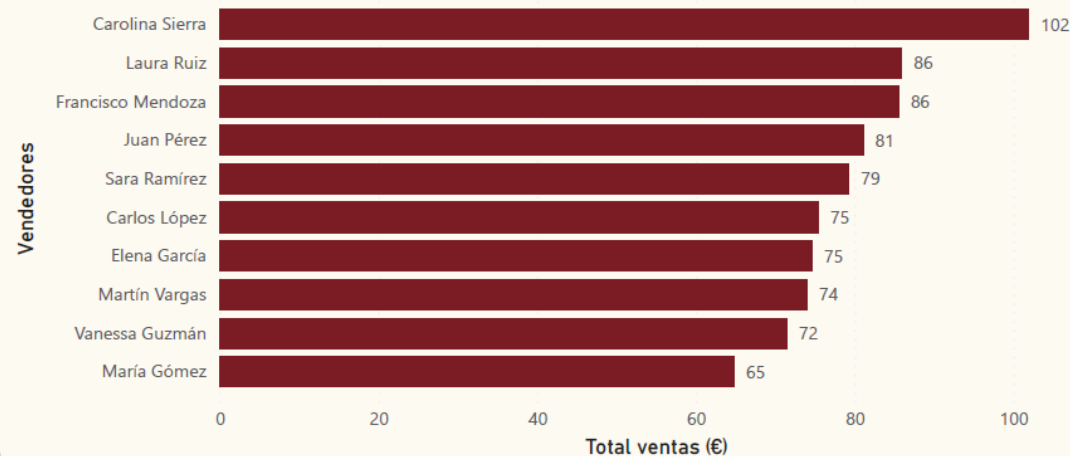
Se ha realizado una visualización simple con dos gráficos y dos tarjetas.

- Gráfico de columnas sobre el total (€) de pedidos en cada mes del 2023.
- Gráfico de barras sobre el top 10 de vendedores, siendo los diez vendedores que más han facturado en sus ventas en el 2023.
- Tarjeta sobre el total de pedidos (€) en 2023.
- Tarjeta sobre el total de ventas (€) en 2023.

TOTAL IMPORTE PEDIDOS 2023 POR MESES



TOP 10 Vendedores 2023



CHOCOLATES VALOR

RESULTADOS FACTURACIÓN 2023



Total Pedidos 2023 (€)

6 mill.



Total Ventas 2023 (€)

2,10 mil



Conclusiones

En esta actividad se ha hecho una simulación de implantación de un sistema BI para la empresa Chocolates Valor, a partir de datos ficticios.

Aunque los datos eran bastante simplificados, se ha podido demostrar que es posible implementar un sistema BI y que se pueden obtener grandes beneficios de ello.

Esta actividad puede interpretarse como un POC para analizar la viabilidad de la implantación.

Se ha podido comprobar que con poco tiempo, la implantación es posible y se pueden obtener resultados interesantes, como se puede ver en la visualización.

Esta es solo una primera fase, lo ideal sería seguir alimentando el modelo con más fuentes de datos, obtener mejores relaciones entre tablas y así poder obtener análisis y visualizaciones más relevantes para los objetivos de la empresa.

ANEXO I

Descripción de los datos

El modelo de datos presentado para implantar el sistema BI ha sido creado a mano para esta actividad. Es un modelo de datos simulado y simplificado con información referente al 2023.

Este modelo de datos está formado por cuatro tablas dentro de una base de datos y dos tablas en ficheros CSV.

A continuación el detalle de cada campo de cada tabla:

- Tabla Productos: Contiene la información de los productos que produce y comercializa Chocolates Valor.

PRODUCTOS		
Variable	Tipo	Definición
ProductoID	<i>int</i>	Identificador único del producto
Producto	<i>texto</i>	Nombre del producto
Categoría	<i>texto</i>	Grupo de productos al que pertenece el producto (Tradicionales, Sin Azucares...)
Tamaño	<i>float</i>	Cantidad en gramos del producto

- Tabla Tiendas: Contiene la información de las chocolaterías que tiene la empresa por el territorio español.

TIENDAS		
Variable	Tipo	Definición
TiendaID	<i>int</i>	Identificador único del producto
Dirección	<i>texto</i>	Dirección postal de la chocolatería
Ciudad	<i>texto</i>	Localidad donde se encuentra la chocolatería
Comunidad	<i>texto</i>	Comunidad Autónoma a la que pertenece la chocolatería

- Tabla Vendedores: Contiene los datos de los empleados que trabajan en las chocolaterías.

VENDEDORES		
Variable	Tipo	Definición
Identificador	<i>código alfanumérico (VAL00)</i>	Identificador de cada empleado
Nombre	<i>texto</i>	Nombre completo del empleado
Equipo	<i>texto</i>	Chocolatería en la que trabaja el empleado

- Tabla Clientes: Contiene la información de los comercios a los que Chocolates Valor distribuye sus productos para su posterior venta.

CLIENTES		
Variable	Tipo	Definición
ClienteID	<i>int</i>	Identificador del cliente
Nombre	<i>texto</i>	Nombre del cliente
Tipo	<i>texto</i>	Tipo de cliente (Gran empresa o PYME)
Localización	<i>texto</i>	Ciudad donde se encuentra el cliente
Telefono	<i>int</i>	Telefono de contacto del cliente
Activo	<i>texto</i>	Variable que indica si el cliente esta activo

- Tabla Ventas Valor: Contiene el registro de las ventas realizadas en todas las chocolaterías durante el año 2023.

VENTAS VALOR		
Variable	Tipo	Definición
VentaID	<i>int</i>	Identificador de la transacción
Vendedor	<i>código alfanumérico (VAL00)</i>	Identificador del empleado que ha realizado la venta
Detalle venta	<i>Texto</i>	Elementos que se han vendido con sus cantidades (ejemplo: 2xIDProducto)
Importe	<i>numérico</i>	Importe total obtenido de la venta
Fecha Venta	<i>fecha</i>	Fecha en la que se ha realizado la venta

- Tabla Pedidos Valor: Contiene el detalle de los pedidos realizados por los clientes durante el año 2023.

PEDIDOS VALOR		
Variable	Tipo	Definición
PedidoID	<i>int</i>	Identificador del pedido
Cliente	<i>int</i>	Identificador del cliente que realiza el pedido
Importe Pedido	<i>numérico</i>	Importe total del pedido
Fecha Pedido	<i>fecha</i>	Fecha del pedido