

El Modelado de Datos

Fundamentos básicos del modelado de datos

Jose Ignacio González Gómez

Departamento de Economía Contabilidad y Finanzas - Universidad de La Laguna

www.jggomez.eu

V.2.3

Modelo de Datos I: Modelo estrella y copo de nieve

- El modelado de datos como resultado final del proceso de ETL. ¿Qué quiero analizar?
- El modelo de estrella. El modelo copo de nieve

Modelo de Datos II: Tablas de hechos y tablas de dimensiones. Relaciones

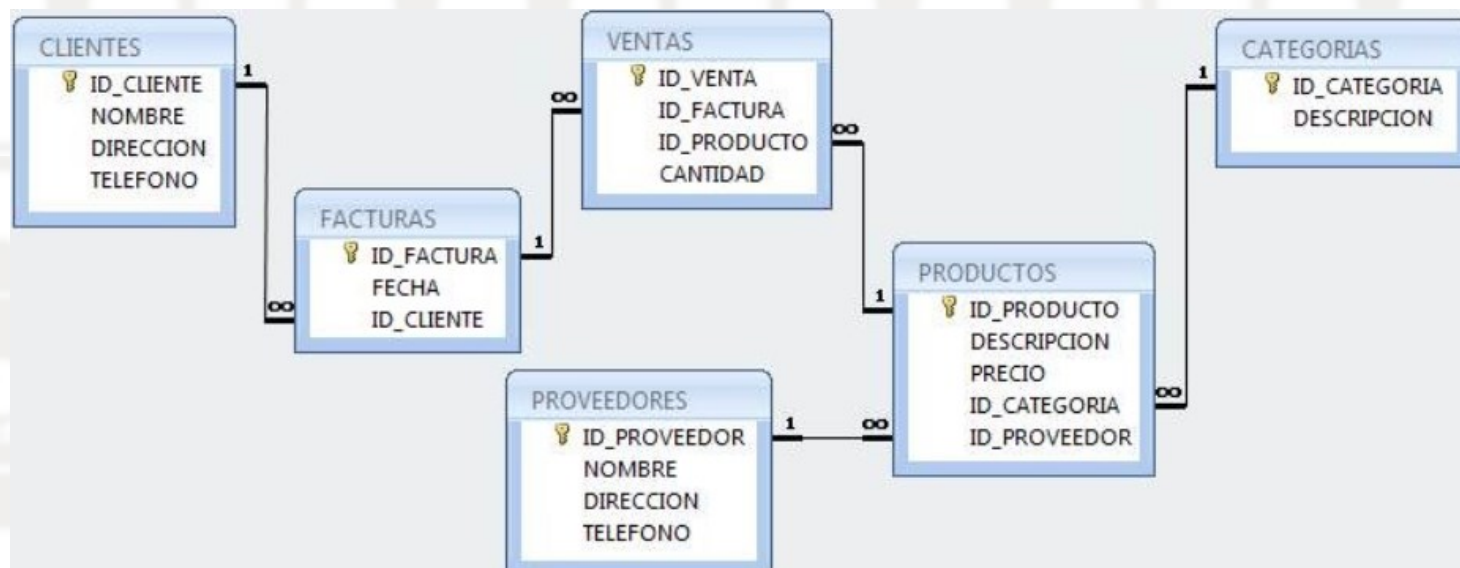
- Las relaciones entre tablas, cardinalidad en las relaciones.
- La propagación de filtros. Dirección de la Relación (del Filtro)
- El contexto de evaluación de las fórmulas

1 Modelo de Datos I: Modelo estrella y copo de nieve

El modelado de datos como resultado final del proceso de ETL. ¿Qué quiero analizar?



En los modelos de datos vamos a trabajar con un conjunto de nuevos conceptos fundamentales como son: Hechos - Dimensiones - Métricas - Claves Principales- Perspectivas- Segmentaciones- Copo de Nieve - Estrellas - etc..



Un Modelo de Datos es un conjunto de tablas organizadas y relacionadas de manera tal que respondan a una necesidad de análisis de negocio

Tipos de tablas

- Tablas de hechos (a partir de las cuales obtenemos las medidas)
- Tablas de dimensiones (nos permiten filtrar la información, por distintos conceptos)

Ejemplos de Modelos de Datos por Procesos de Negocio

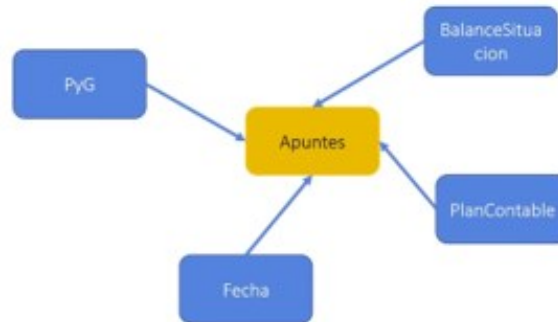
Ejemplo de ventas sencillo



Ejemplo de ventas detallado



Modelo de Datos - Financiero



Modelo de recursos humanos sobre las horas trabajadas



¿Qué quiero analizar?

Debemos comenzar el diseño de un modelo de datos en reflexionar sobre que quiero analizar y esto implica una serie de pasos como son:

1. *Identificar el proceso de negocio* que se ve afectada para el análisis, por ejemplo, ventas, compras, finanzas, gestión de stock, recursos humanos, etc..
2. Contar con una batería de preguntas a las que intentamos que el modelo de respuesta.
3. Identificar donde está la historia de lo que ha ocurrido, en que tablas, y que nivel de detalle necesito, campos que contiene la tabla
4. Y debemos también definir desde que perspectivas o puntos de vista queremos analizar la información, es decir por características del cliente, características del producto, etc.

Estos son los cuatro pasos que debemos definir crear un modelo de datos de un proceso de negocio.

Ejemplo

- ✓ *Proceso de negocio – Ventas.* Contamos con el proceso de negocio en este caso ventas.
 - *Proceso y personas que necesita información* Analizamos que departamentos y personas necesitan información de el
 - *Batería de preguntas.* Hacemos las preguntas a cada persona y a cada departamento. Esta es la parte mas compleja y que mas tiempo requiere pero es la parte fundamental.
- ✓ *Determinamos la variable que queremos analizar y punto de vista.* Por ejemplo, la variable que queremos analizar son las ventas y el punto de vista que queremos que queremos, por ejemplo

Cliente	Fecha	Tienda	Producto	Empleado
¿A <u>quien</u> he vendido?	¿Cuándo?	¿Dónde?	¿Qué?	¿Quién lo ha vendido?

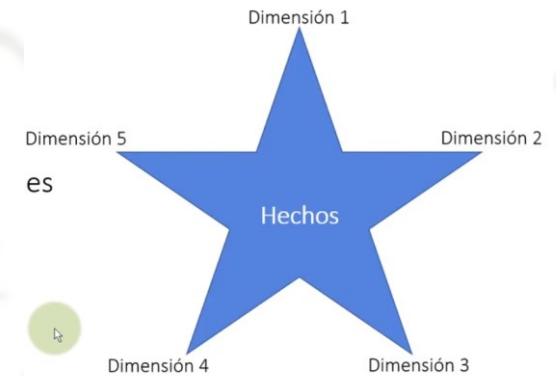
- ✓ *Que vamos a medir por cada venta, por cada hecho que ocurre.* Nos podría interesar el pvp, importe, descuento, coste, beneficio...
- ✓ *Desde que punto de vista vamos a analizar cada venta o hecho.*
 - Cliente: nombre, dirección, edad, salario, población, tipo, estado...
 - Fecha: día, hora, día de la semana, trimestre, es festivo, quincena, semestre, ...
 - Producto: categoría, ubicación, familia, tarifa...

Todo lo anterior implica que disponemos de diferentes tablas que debemos relacionar y que contienen los campos necesarios para dar respuesta a las preguntas planteadas tal y como se muestra en la siguiente ilustración

- Hechos ocurridos → Ventas → 1 Tabla
- Puntos de vista (Perspectivas)
 - Cliente → 1 Tabla
 - Fecha → 1 Tabla
 - Tienda → 1 Tabla
 - Producto → 1 Tabla
 - Empleado → 1 Tabla

El modelo de estrella. El modelo copo de nieve

Un modelo de estrella que es aquel que por su forma relacionamos las tablas siempre permaneciendo en el centro la tabla de hechos y el resto de tablas que son las perspectivas o dimensiones desde que la quiero analizar



Ejemplo del proceso de negocio de las ventas, modelo estrella

Que quiero analizar: Las ventas

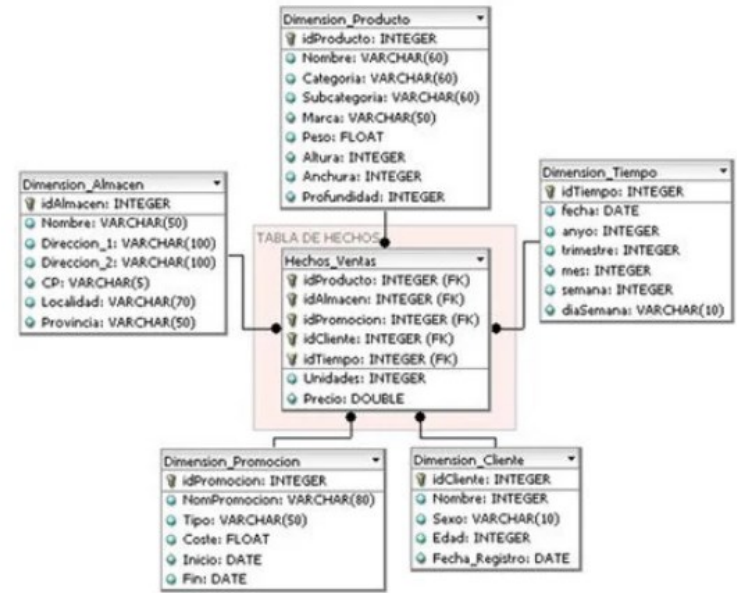
- **Que ha ocurrido ~ Hechos ~ Diario de ventas o histórico de ventas (Tabla de hecho)**
- **Puntos de vista (perspectivas) ~ Dimensiones~ Cliente, Fecha, Tienda, Producto, Empleado.... (Tablas de dimensiones)**

Partiendo de la tabla de hecho diario de ventas o histórico de ventas que contienen diversos campos como son importe de la venta, descuento, fecha, tienda empleada, cliente y producto vamos a analizarla bajo 5 perspectivas.



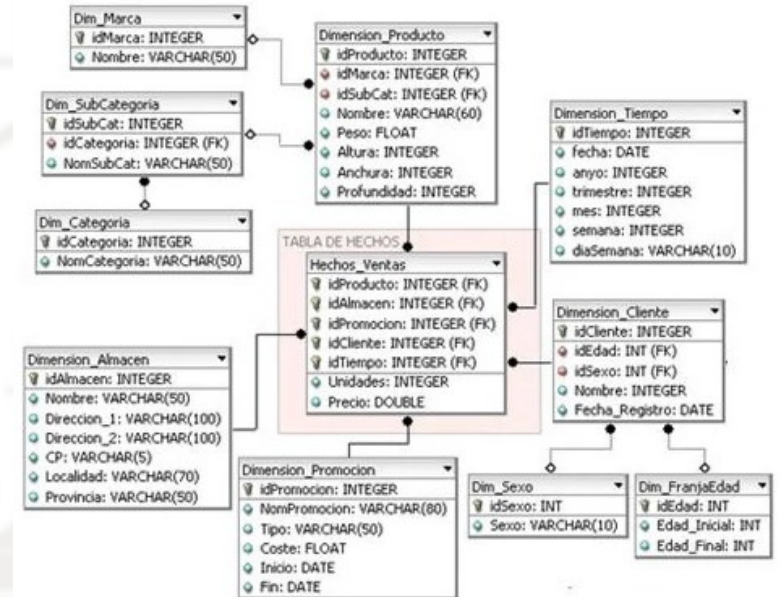
El modelo estrella

En resumen, en los modelos estrellas, la estrella es la raíz, ya que **este modelo parte de una tabla de hechos de la que se desprenden distintas tablas de dimensiones**. La tabla de hechos se queda enraizada en un solo punto y, a partir de allí, se empiezan a producir las tablas de dimensiones.



El modelo copo de nieve

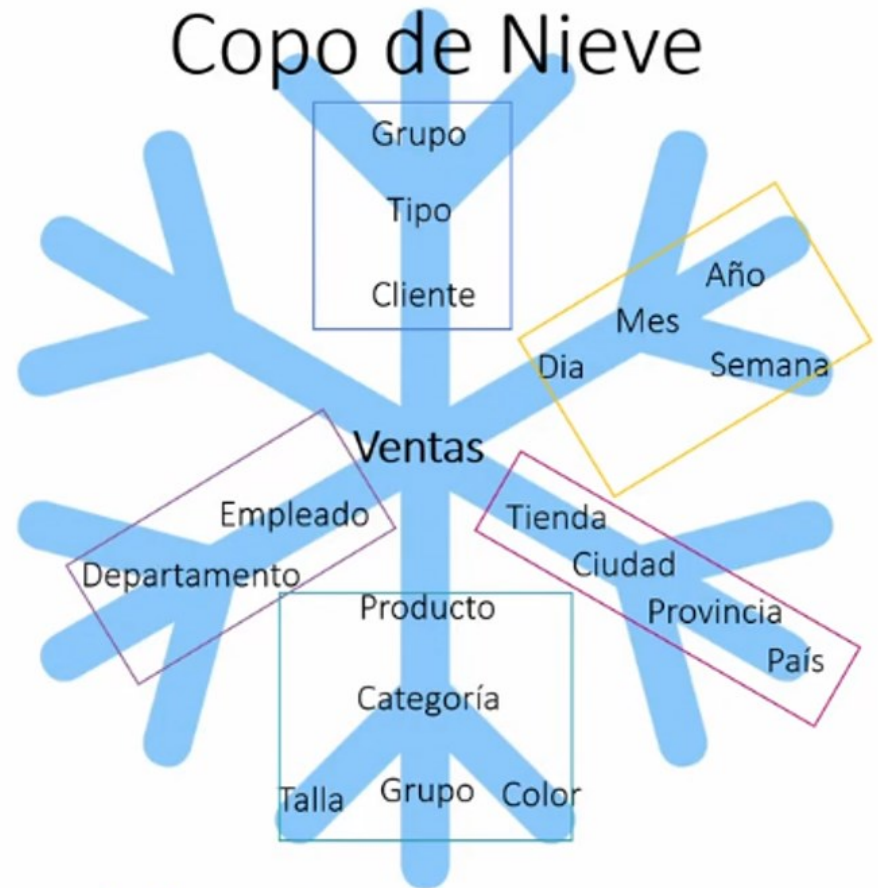
En el modelo copo de nieve **una tabla de dimensiones puede estar interconectada con otra, sin necesidad de recurrir directamente a la tabla de hechos**. Las tablas de dimensiones en el modelo de copo de nieve se interrelacionan entre sí.



Estrella

vs

Copo de Nieve



Hechos: Ventas

2 Modelo de Datos II: Tablas de hechos y tablas de dimensiones. Relaciones

Las relaciones entre tablas, cardinalidad en las relaciones

En general la **tabla de hechos** se suele caracterizar por:

- Pocas columnas
- Casi todas de tipo numérico
- Contiene muchas filas o registros y con crecimiento periódico

Tabla de Dimensiones o de búsqueda. Jerarquías

Son las que nos van a permitir segmentar y filtrar la información por muy diversos criterios. Es decir, son los puntos de vista (perspectiva) desde la que queremos analizar la información



RELACIONANDO LAS TABLAS

Uno a Varios

Varios a Varios

Uno a Uno

Tipo de conector: 1-*

Tipo de conector: *-*

Tipo de conector: 1-1

La cardinalidad de las relaciones se refiere cuáles son los elementos o campos de nuestras tablas que va a estar relacionados entre si

Editar relación

Permite seleccionar tablas y columnas relacionadas.

Ventas

ID Vendedor	ID Venta	ID Factura	Fecha (Factura)	Fecha (Entrega)	Subtotal Venta	Impuesto	Venta
4	125588	39148	03/01/2019	07/01/2019	13	1,95	
4	125957	39268	07/01/2019	09/01/2019	13	1,95	
4	128459	40062	22/01/2019	25/01/2019	13	1,95	

Tabla Calendario

Date	Año	Día	Día de la semana	Mes	Offset de Año	Semestre	cuatrimestre	trimestre
01/07/2017	2017	1	6	7	-5	2	2	3
02/07/2017	2017	2	7	7	-5	2	2	3
03/07/2017	2017	3	1	7	-5	2	2	3

Cardinalidad

- Varios a uno (*:1)
- Varios a uno (*:1)
- Uno a uno (1:1)
- Uno a varios (1:*)
- Varios a varios (*:*)

Dirección del filtro cruzado

- Única
- Única
- Ambas

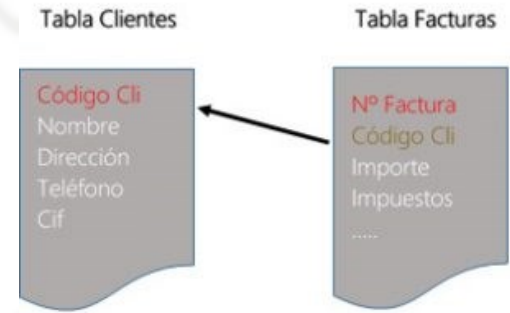
Aplicar filtro de seguridad en ambas direcciones

Aceptar

Cancelar

Uno a Varios, la más frecuente

Es la más frecuente, nos indica que cada uno de los registros de una tabla (Clientes) puede tener muchos registros en la otra (Facturas), pero a la inversa, cada uno de los registros en facturas, solo podría tener un único registro relacionado en la tabla de clientes



Varios a Varios, evitar

También conocida como muchos a muchos, pero este tipo de relaciones deben evitarse por su complejidad.

SKU	Nombre
7	KitchenAid Estufa, Negra
8	KitchenAid Estufa, Roja
8	KitchenAid Estufa, Plateada
9	Apple iPhone X
10	Samsung Galaxy 8
11	Ikea Librero 1.5 metros

Día	SKU	# de productos
30/01/2019	8	105
05/02/2019	8	80
06/02/2019	8	210
08/02/2019	8	85
09/02/2019	8	90

Tenemos una tabla de productos (izquierda) y una tabla de pedidos que se relacionan con su código de producto (SKU), lo que ocurre, que no es frecuente, que disponemos de un producto que tiene el mismo código (8) que solo lo diferencia el color, por eso la relación es de varios a varios.

Uno a Uno, no frecuente

No es muy frecuente, nos indica que cada registro de una tabla tiene un solo valor en la otra, por ejemplo, municipios de España y alcaldes de España

Relaciones generadas automáticamente

Estas relaciones entre se establecen de forma automática al agregar las tablas al modelo de datos. Cuando se importan distintas tablas desde una misma BBDD, es muy común que se arrastran las relaciones igualmente. **Es conveniente repasarlas y confirmar que las relaciones que se hayan generado sean correctas, eliminando aquellas que no correspondan con nuestro modelo para un correcto funcionamiento**

Sobre el editor de relaciones

Para configurar nuestras relaciones contamos con un editor al cual accedemos haciendo doble clic sobre la relación, donde podremos definir la cardinalidad (uno a uno, uno a varios o varios a uno)

Editar relación

Permite seleccionar tablas y columnas relacionadas.

Ventas

ID Vendedor	ID Venta	ID Factura	Fecha (Factura)	Fecha (Entrega)	Subtotal Venta	Impuesto	Venta
4	125588	39148	03/01/2019	07/01/2019	13	1,95	
4	125957	39268	07/01/2019	09/01/2019	13	1,95	
4	128459	40062	22/01/2019	25/01/2019	13	1,95	

Tabla Calendario

Date	Año	Día	Día de la semana	Mes	Offset de Año	Semestre	cuatrimestre	trimestre
01/07/2017	2017	1	6	7	-5	2	2	3
02/07/2017	2017	2	7	7	-5	2	2	3
03/07/2017	2017	3	1	7	-5	2	2	3

Cardinalidad

Varios a uno (*:1)

Varios a uno (*:1)

Uno a uno (1:1)

Uno a varios (1:*)

Varios a varios (*:*)

Dirección del filtro cruzado

Única

Única

Ambas

Aplicar filtro de seguridad en ambas direcciones

Aceptar Cancelar

La propagación de filtros. Dirección de la Relación (del Filtro)

Dirección de la Relación (del Filtro)

Otro de los aspectos relevantes con las relaciones entre las tablas es la dirección de filtrado, es decir es la dirección en la que se van a aplicar los filtros en el momento en que comencemos a analizar nuestra información a través de las diferentes visualizaciones.

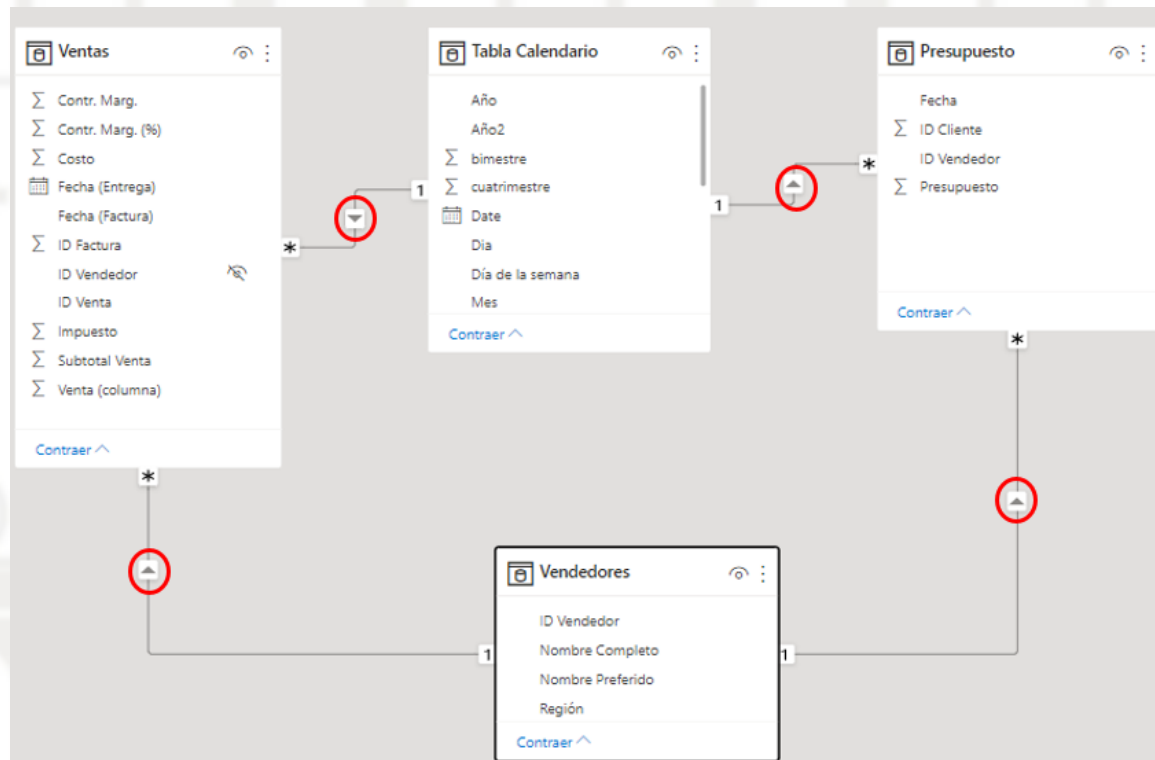
Ejemplo, supongamos que tenemos dos tablas relacionadas, en el 1º caso sin selección (sin filtro) y en el segundo con selección (ID de vendedor 11 y 12) aplicando por tanto filtro y el resultado obtenido es:

Sin Selección (Sin Filtro)

ID Vendedor	ID Vendedor	Cantidad
2	2	1
3	3	1
4	4	1
5	5	1
6	6	1
7	7	1
8	8	1
9	9	1
10	10	1
11	11	1
12	12	1
13	13	1
14	14	1
15	15	1
16	16	1
17	17	1
18	18	1
19	19	1
20	20	1

La Selección Aplica el Filtro

ID Vendedor	ID Vendedor	Cantidad
2	11	1
3	12	1
4	11	2
5	12	2
6	11	3
7	12	3
8	11	4
9	12	4
10	11	5
11	12	5
12	11	6
13	12	6
14	11	7
15	12	7
16	11	8
17	12	8
18	11	9
19	12	9
20	11	10



El contexto de evaluación es la forma en que se evalúan las fórmulas, es decir cómo se aplica las capas de filtrado a las tablas para después hacer sus cálculos.

Así una misma medida como puede ser ventas, puede arrojar distintos resultados dependiendo del contexto de evaluación

The screenshot displays a Power BI report interface. On the left, there are three main sections: 'dClientes' with three filterable items (Cliente 1, Cliente 2, Cliente 3), 'dProductos' with a donut chart showing three segments (24.16%, 51.01%, 24.83%) and three filterable items (Producto 2, Producto 1, Producto 3), and a 'Panel de filtros' section. In the center is a table titled 'Tabla original hPedidos' with columns: Fecha, ID_Producto, ID_cliente, ud, and euros_ud. On the right is a 'Filtros' pane with a search bar and several filterable items: 'euros_ud es (todos)', 'Fecha es (todos)', 'ID_cliente es (todos)', 'ID_Producto es (todos)', and 'ud es (todos)'. A yellow hand cursor is pointing at the 'Fecha es (todos)' filter. Blue arrows point from the 'dClientes' and 'dProductos' sections to the table, and from the 'Panel de filtros' section to the 'Fecha es (todos)' filter.

Fecha	ID_Producto	ID_cliente	ud	euros_ud
01/01/2022	1	1	5	1.000 €
02/01/2022	3	1	5	400 €
01/02/2022	3	1	2	1.000 €
02/02/2022	3	1	2	800 €
01/03/2022	3	3	1	3.000 €
02/03/2022	2	3	88	300 €
02/03/2022	3	2	33	30 €
01/04/2022	3	3	1	3.000 €
02/04/2022	1	3	1	8.000 €
01/05/2022	1	2	10	2.000 €
02/05/2022	1	2	10	3.500 €
01/06/2022	1	1	2	1.000 €
02/06/2022	1	1	2	1.500 €
01/07/2022	2	3	4	3.000 €
02/07/2022	2	3	4	3.000 €
01/08/2022	3	3	8	3.000 €
02/08/2022	2	3	56	3.200 €
01/09/2022	3	2	4	2.000 €
02/09/2022	1	2	4	2.000 €
01/10/2022	1	2	23	2.000 €
02/10/2022	3	2	12	2.200 €
01/11/2022	1	1	1	1.000 €
01/11/2022	1	2	2	2.000 €
02/11/2022	1	2	12	2.400 €
01/12/2022	3	1	1	5.000 €
02/12/2022	1	2	2	800 €
02/12/2022	3	1	3	10.960 €

El Contexto de filtro es el conjunto de datos que están disponibles para ser utilizados para una fórmula (expresión DAX) después de haber superado las restricciones o modificaciones que imponen.

Video recomendado: [Contextos de EVALUACIÓN del lenguaje DAX](#)

Contexto fila o Row Context

Ventas	\$230,294,386
---------------	---------------

Región	Ventas
Región 1	\$55,802,340
Región 2	\$80,240,570
Región 3	\$54,617,792
Región 4	\$39,633,685
Total	\$230,294,386

Contexto de Filtro (Filter Context)

País	Año	Mes
Argentina	2012	Enero
Brasil	2013	Febrero
Colombia		

Categoría de Descuento
None

EL CONTEXTO DE FILTRO

- País: Colombia
- Categoría de Descuento: Cyber Monday
- Mes Nombre: Enero
- SKU: CB01

En general, tenemos 4 elementos del reporte que añaden o remueven filtros del contexto de filtro, estos elementos son:

Objetos visuales. Los objetos visuales son todos aquellos objetos que nos ayudan a construir el reporte mediante comunicación e interacción visual, ejemplos: Matrices, Segmentaciones de datos, tablas y graficos dinámicos, etc..

Año ☰ 🔍

2021 2022

Ptos vending ☰ 🔍

Aeropuerto GC

Aeropuerto TF Sur

Estación Guaguas S.C

Los Cristianos Peatonal

Puerto de Agaete

San Agustin Peatonal

Análisis General de Ventas y del Beneficio Bruto

Ordenar cronologicamente

Trimestre Año	Ingresos	Nº de Tickets	Venta Media p/ Nº Product.Vendidos	Promedio Productos por tick	Beneficio Bruto
T1/21	284.845 €	101.927	2,8 €	1,65	106.314 €
T2/21	350.783 €	114.628	3,1 €	1,57	134.266 €
T3/21	267.529 €	85.616	3,1 €	1,53	100.768 €
T1/22	250.310 €	86.246	2,9 €	1,49	93.697 €
T2/22	13.838 €	5.469	2,5 €	1,47	5.367 €
Total general	1.167.304 €	393.886	3,0 €	1,56	440.412 €

Pto Vending	Ingresos	% Ingresos	Beneficio Bruto
Alimentacion	323.130 €	27,68%	143.613 €
Drogueria y Perfumeria	39.965 €	3,42%	14.987 €
Golosinas y Chocolates	159.888 €	13,70%	71.061 €
Helados	22.233 €	1,90%	10.532 €
Juegos y Complementc	80.079 €	6,86%	22.880 €
Papeleria y Libreria	28.650 €	2,45%	7.428 €
Productos Varios	62.154 €	5,32%	10.359 €
Refrescos y Bebidas	244.069 €	20,91%	100.499 €
Revistas	53.608 €	4,59%	4.873 €
Snacks y Aperitivos	72.546 €	6,21%	27.205 €
SubFamilia 15	77 €	0,01%	7 €
Vestido	80.904 €	6,93%	26.968 €
Total general	1.167.304 €	100,00%	440.412 €

Pto Vending	Benef. Bruto	%Benf.Bruto
Comestibles	352.910 €	80,13%
No comestibles	87.495 €	19,87%
Total general	440.405 €	100,00%

Benef. Bruto

Familia	Benef. Bruto	%
Comestibles	352.910 €	80%
No comestibles	87.495 €	20%

Familia ☰ 🔍

- Comestibles
- No comestibles

Panel de Filtros. Disponible tanto en Excel como en Power BI

Interacciones Visuales

Administración de Roles (RLS) — Este es un caso especial

En resumen

