

Participantes:

- Luca Terranova
- Cristian Ezequiel Albornoz
- Carla S. Centeleghe

Farmacia



7 8 9 10

Introducción

El siguiente proyecto trata sobre la problemática de una farmacia, la cual tiene una base de datos que no satisface sus necesidades. Su base de datos consiste en tablas dedicadas a sus clientes y a sus empleados (las cuales tienen atributos que nos dan la información de los mismos). También cuenta con tablas que especifican productos, proveedores, mercadería, stock y facturación (se explicara más sobre esta base más adelante en el informe) .

Pero los altos mandos necesitaban tomar decisiones con rapidez y teniendo en cuenta el estado actual de la farmacia, es decir deben poder comparar las ventas de los medicamentos en diferentes años o decidir cuál medicamento se vende más en qué mes del año o ver donde están ubicados sus clientes (su área de ventas geográfica). Por estas y más consultas (que se detallaran luego), fue necesario crear un Data Warehouse.

Para la Data Warehouse se decidió utilizar PowerBI y todas sus heramientas.

Indice:

-	Introducción.....	pág 1
-	Proyecto.....	pág 3
	1) Marco teórico.....	pág 3
	2) Base de datos anterior	pág 5
	3) Preguntas no respondidas.....	pág 6
	4) Dimensionar y modelar.....	pág 7
	5) Olap y Cubo.....	pág 10
	6) Power Bi y las respuestas.....	pág 12
-	Conclusión.....	pág 20
-	Links.....	pág 21
-	Anexo.....	pág 21

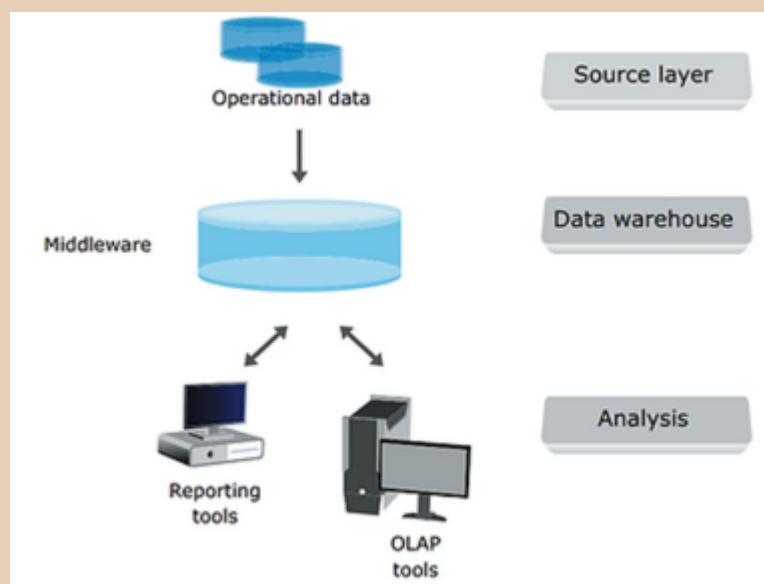
Proyecto:

Este proyecto comenzó tiene como objetivo mostrar la creación de un Data WareHouse para la cual utilizamos el dimensionado, modelado y creación del cubo. Lo que nos permite más rápido acceso a las tablas (donde se encuentran los datos) y la comparación con otros datos, para obtener consultas utilizamos Power Bi.

Seguramente se preguntarán ¿por qué debemos crear un Data WareHouse, si ya tenemos una base de datos existente, a la cual podemos hacer las consultas que deseemos? Esto se debe a que si tenemos una base de datos con gran cantidad de datos las consultas se vuelven lentas, además los altos mandos de las compañías no desean saber datos específicos de cada cliente (por ejemplo su nombre o apellido) o venta; sino una visión más general del panorama de la empresa.

Por ello se empezaron a implementar las primeras Data WareHouse, el Data WareHouse es repositorio centralizado donde se almacena toda la información proveniente de uno o más sistemas, cuyos datos atraviesan un proceso de limpieza y transformación para realizar un análisis sobre ellos. El mismo es una arquitectura bien definida cuyo objetivo es satisfacer la demanda de empresarios de obtener una visión integrada de la empresa y su entorno.

La Estructura de un Data WareHouse se divide en cuatro partes: Hardware, Software de Almacenamiento (SGBD), Software de Extracción y Manipulación de Datos (minería de datos) y Herramientas Middleware.



El **Hardware** es la parte física de la Data Warehouse, ya que se necesita contar con una infraestructura capaz de responder las necesidades avanzadas de los usuarios. En otras palabras, se necesita un equipo capaz de soportar y almacenar la gran cantidad de datos que utiliza el mismo; y sobre todo debe ser capaz de procesar de manera eficaz los pedidos o consultas de los usuarios.

El **Software de Almacenamiento** (SGBD) es necesario ya que procesan una gran cantidad de datos a gran velocidad. Los datos se ven como cubos de información, lo cual añade una nueva dimensión a los datos almacenados en forma de tablas. Esta información suele ser en gran medida consultas multidimensionales, por lo que es importante contar con Bases de Datos Multidimensionales.

Software de Extracción y Manipulación de Datos, su uso principales usos son:

- ★ Control de la extracción de datos y automatización.
- ★ Acceso a diferentes tecnologías.
- ★ Gestión integrada del Data Warehouse y los Data Marts existentes.
- ★ Acceso a gran variedad de fuentes de datos.
- ★ Manejo de excepciones.
- ★ Planificación, logs, interfaces a schedulers de terceros.
- ★ Soporte en la explotación del Data Warehouse

Por último las **Herramientas Middleware**, como bien dice su nombre, son herramientas que se utilizan como soporte entre la fase de gestión de un Data Warehouse y las aplicaciones de usuario.

Para crear una una Data Warehouse hay 5 fases:

- ★ Definición de Objetivos
- ★ Definición de Requerimientos
- ★ Diseño y Modelización
- ★ Implementación
- ★ Revisión

La **definición de los objetivos** consiste en los objetivos que se quieren alcanzar con la base de datos, decir que se quiere obtener y hasta donde se puede o debe llegar.

El siguiente paso es la **definición de requerimientos**, la cual consiste en analizar las necesidades del usuario y ver como pueden ser implementadas por el sistema. El usuario juega un papel muy importante en esta fase, ya que es él quien debe colaborar para decidir las necesidades que deben satisfacerse.

Luego está el **diseño y modelado** de la Data Warehouse, que es la parte donde se crea el modelo lógico de la misma (este modelo está formado por diferentes entidades que permitirán resolver las necesidades del negocio). El

modelo sirve en gran medida para definir cuáles serán las fuentes de datos y las transformaciones necesarias de las mismas

La **implementación** lleva muchos pasos dentro de la misma, los cuales son la extracción de datos y transformación de los mismos, la carga en el Data WareHouse y la explotación del mismo por diversas herramientas como OLAP o DSS, entre muchas otras.

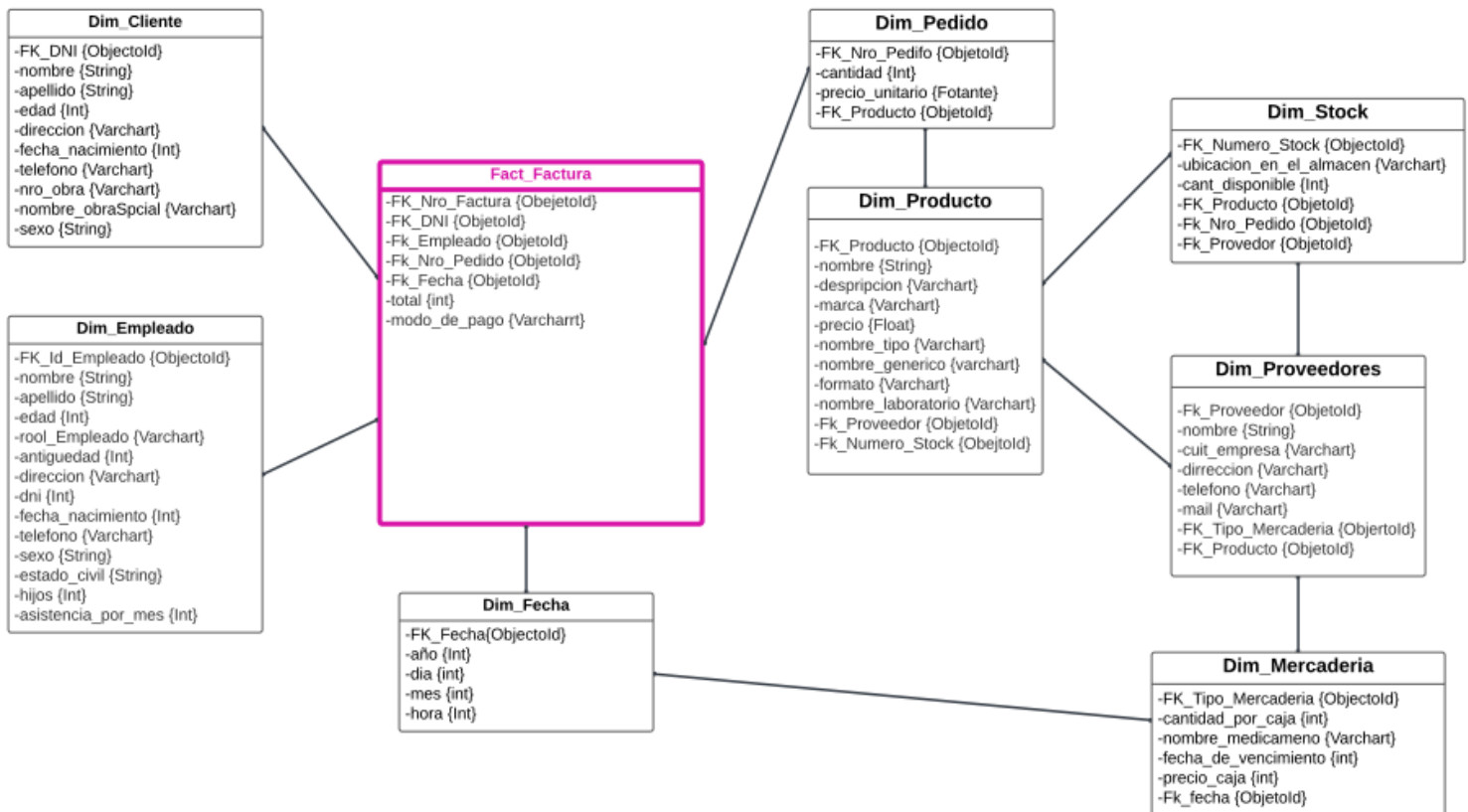
DATA WAREHOUSE:

Teniendo en cuenta lo explicado anteriormente, ya sabemos que es una Data WareHouse, como se compone y cómo se implementa. Ahora pasaremos a aplicar lo aprendido en la creación de una Data WareHouse para una farmacia.

La farmacia (que en este caso es nuestro cliente) nos planteó que su actual base de datos nos satisface las demandas de los altos mandos de la empresa, los cuales desean analizar a la empresa desde un punto más objetivo y poder hacer grandes comparaciones de los datos obtenidos con datos de años pasados.

La farmacia utiliza una base de datos con motor SQL Server, la misma almacena datos en diferentes tablas como se muestra a continuación.

BASE DE DATOS 2 - FARMACIA



La misma contiene diferentes tablas con diferentes atributos los cuales sirven para almacenar datos específicos de cada venta realizada. Muchas de ellas están unidas entre sí con una llave foránea, como por ejemplo la tabla Dim_Producto está unida a Dim_Stock y a Dim_Provedores, y está a su vez está unida a la tabla Dim Mercadería.

Una base de datos con tantas tablas unidas puede generar retrasos a la hora de consultas, más si estas están llenas de muchos datos, los cuales puede que no sean necesarios.

Luego de analizar la base de datos existente, se pasó a realizar una captura de requerimientos y una definición de objetivos. Lo cual consistió en entrevistar a los altos mandos de la farmacia para saber qué deseaban de la Data Warehouse.

De las entrevistas se lograron establecer estas preguntas como objetivos que la Data Warehouse debía responder:

1. ¿Cuál fue de los últimos cinco años, el año con mayor facturación, y cuánto se facturó?
2. ¿Cuál es el mes del año con mayor volumen de ventas y cuánto se facturó ese mes?
3. ¿Cuál es el día de la semana con más ventas?
4. ¿En el primer trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?
5. ¿En el segundo trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?
6. ¿En el tercer trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?
7. ¿En el cuarto trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?
8. ¿En los últimos tres años cuáles fueron los veinte artículos más vendidos y qué cantidades?
9. ¿En los últimos tres años se facturó más a clientes en forma particular o a través de obras sociales?
10. ¿Cuál es el ranking de ventas realizadas a través de las obras sociales en los últimos tres años?
11. ¿Nuestra clientela son más mujeres, hombres o es indistinto?
12. ¿Cuáles fueron los diez artículos más vendidos según el género?
13. ¿Cuáles son los diez artículos que más compran las personas mayores de sesenta años? ¿Cuál es la facturación semanal de esta franja etaria?
14. ¿Cuáles son los diez artículos que más compran las personas entre treinta y sesenta años? ¿Cuál es la facturación semanal de esta franja etaria?

15. ¿Cuáles son los diez artículos que más compran las personas menores de treinta años? ¿Cuál es la facturación semanal de esta franja etaria?
16. ¿En qué zona viven los clientes que más compraron en el mes? Hacer ranking de zonas
17. ¿En el último año qué empleado realizó más ventas? ¿Cuál es su perfil (soltero-casado-hijos-etc.)

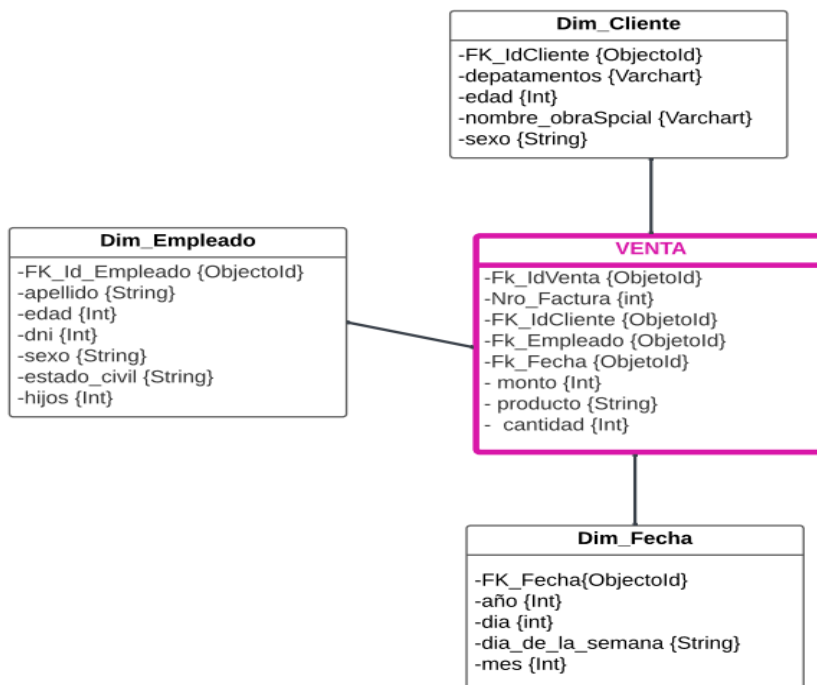
Luego del análisis podemos decir que nuestra Data WareHouse está orientada al departamento de ventas de la empresa.

Teniendo claro los objetivos pasamos al diseño y modelado de la Data WareHouse. Para ello debemos tener en cuenta que al diseñar y modelar, también hay que dimensionar la base de datos. Es decir ver cuales son los eventos que responden a nuestras preguntas planteadas por el cliente (en este caso la farmacia). Cada pregunta o combinación de ellas, representan una dimensión.

Además no debemos perder de vista donde se guardaran los datos, cuáles datos nos son útiles y cuáles no, cómo se van a definir las relaciones de las llaves primarias, que pasa si un dato es nulo (vacío) y que pasa si en la base de datos se modifica un dato.

En nuestro caso decidimos realizar el siguiente diagrama (modelo multidimensional del tipo estrella), se puede ver cuáles datos consideramos importantes extraer de la base de datos origen, y cuales descartamos ya que no nos servían para responder los objetivos planteados por el cliente. S

BASE DE DATOS 2 - FARMACIA



Como se puede observar en el diagrama eliminamos varias tablas, como por ejemplo la tabla Stock, la tabla Producto, la tabla Proveedores y la tabla Mercadería. Esto se debe a que luego de desnormalizar las tablas, encontramos varios datos que eran irrelevantes en nuestro Data Warehouse, los cuales podían ser omitidos en la misma.

También al analizar las tablas eliminamos o cambiamos algunos atributos de las tablas originales. Para poder así agilizar el tiempo de respuesta. Otro cambio que realizamos fue cambiar una llave foránea por el nombre del producto, siendo así más ágil el procesamiento de datos.

Además decidimos lo que sucedería con los datos nulos, los mismos serán reemplazados por “no disponible” o en algunos casos se eliminarán. Ya que el modelado no aceptara datos nulos.

La base de datos puede cambiar constantemente por lo que el Data Warehouse requiere un método para reflejar estos cambios dentro de las dimensiones, en nuestro caso usamos el Tipo 1. El cual consiste en sobrescribir el atributo viejo por el nuevo valor, por lo que no queda ningún registro del valor anterior y todos los datos siempre muestran los valores más actualizados.

En otras palabras lo que sucede si un dato es modificado en la base de datos original es lo siguiente:

1. Se guardan los cambios en la base de datos
2. Se actualiza el cubo (donde se guarda la información)
3. Se guardan los cambios en el cubo
4. Se actualiza el Power BI
5. Se cambian las tablas o los reportes automáticamente

Para la fase de implementación, la cual consiste en primer lugar en exportar los datos a una base de datos destino. Para ello realizamos el proceso de ETL y la creación de una tabla de hecho.

¿En qué consiste el proceso de ETL? El mismo consiste en mover datos desde múltiples fuentes, formatearlos y limpiarlos, y cargarlos en un almacén de datos.

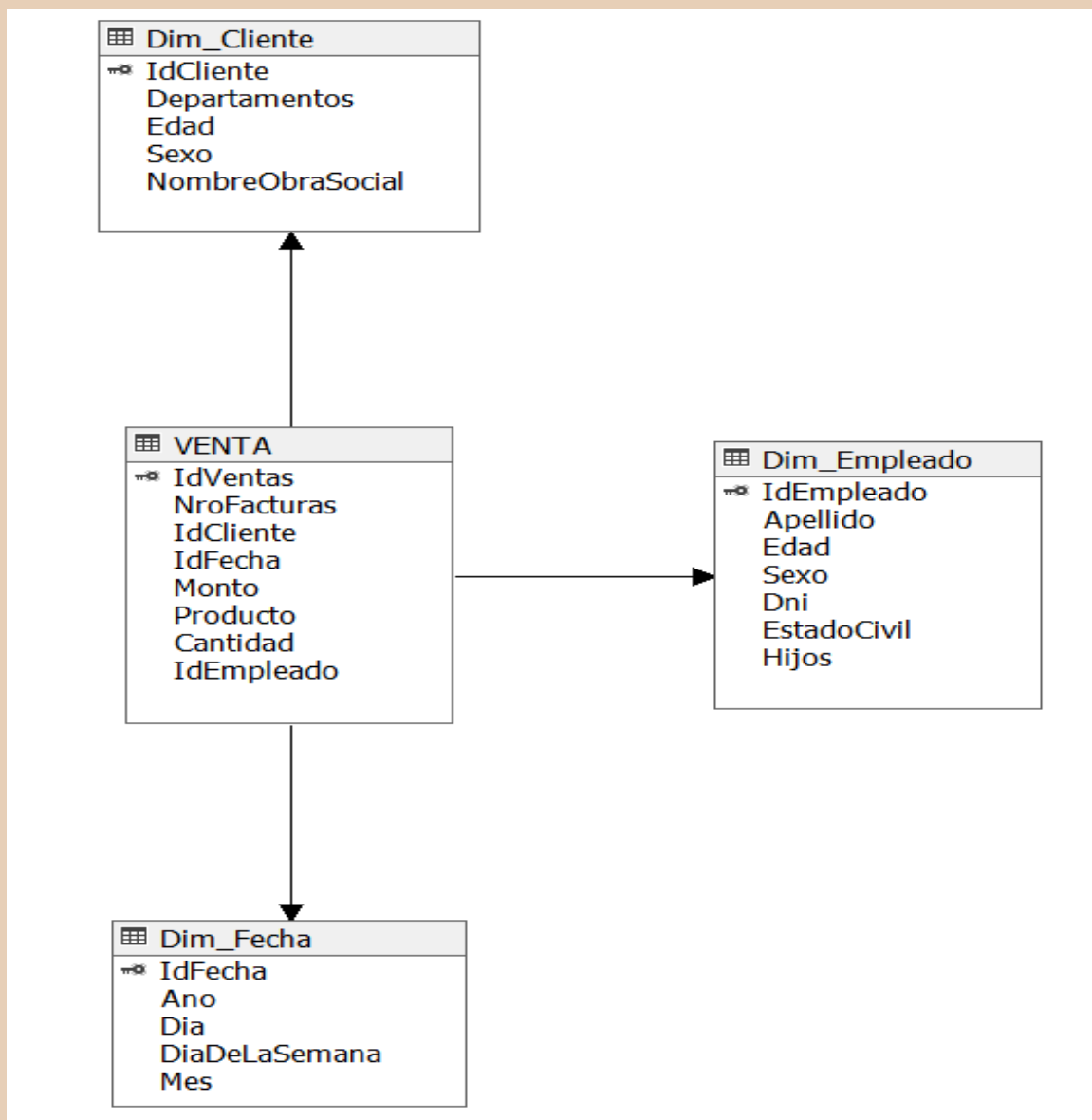
Es decir que el proceso se divide en tres grandes bloques:

- ★ **Extracción de datos:** consiste en determinar qué datos se procesan, la velocidad y prioridad de extracción de los mismos. También consiste en el análisis y limpieza de datos para poder diferenciarlos, por lo cual se realizan variedad de acciones sobre las tablas y filas de datos específicas.
- ★ **Transformación:** los datos, transformarlos, se controlan a través de diversos conjuntos de reglas (establecidas en su mayoría por la compañía) para proporcionar calidad, orden y sobre todo evitar errores (los cuales se corrigen y se resuelven todas las diferencias).

Con el fin de dar claridad a cómo deben ser integrados los datos en la base de datos. Todo este proceso se realiza generalmente por una validación, eliminación de duplicados, codificación y filtrado en el formato deseado, lo cual permite conocer cuáles datos tiene alguna diferencia para ver si se omiten o se hacen a un lado para un análisis más profundo.

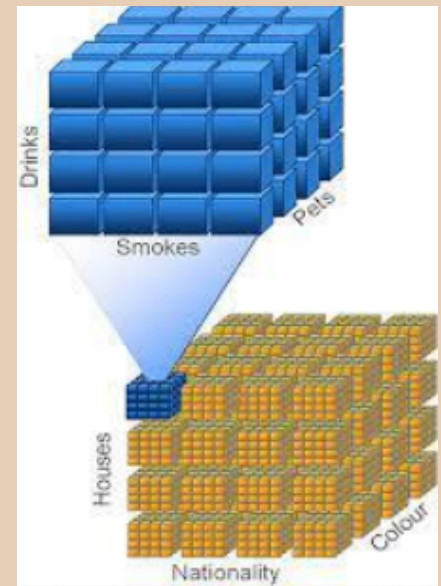
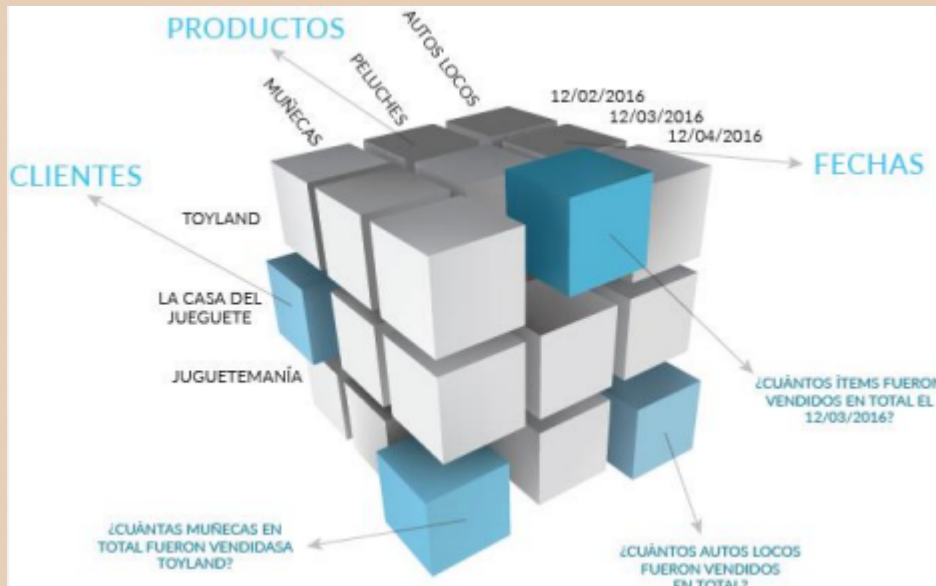
- ★ **Carga:** es el almacenamiento de los datos en una base de datos destino, la cual es llamada Almacén de Datos o Repositorio centralizado.

Luego de realizar todos los pasos nuestro Almacén de Datos que daría de la siguiente manera.



El siguiente paso en la implementación de un Data WareHouse consiste en la construcción del cubo, un proceso analítico.

Un procesamiento analítico en línea, es una tecnología que organiza grandes bases de datos empresariales, y proporciona un análisis complejo. Se basa en **Estructuras Multidimensionales**, las cuales consisten en agrupar datos en diferentes dimensiones y métricas.



Un cubo en pocas palabras es un análisis multidimensional que representa los datos mediante dimensiones, cada dimensión corresponde a una jerarquía de datos. Esto nos permite llevar a cabo un análisis completo de diversas situaciones, para obtener las soluciones correctas de los problemas planteados.

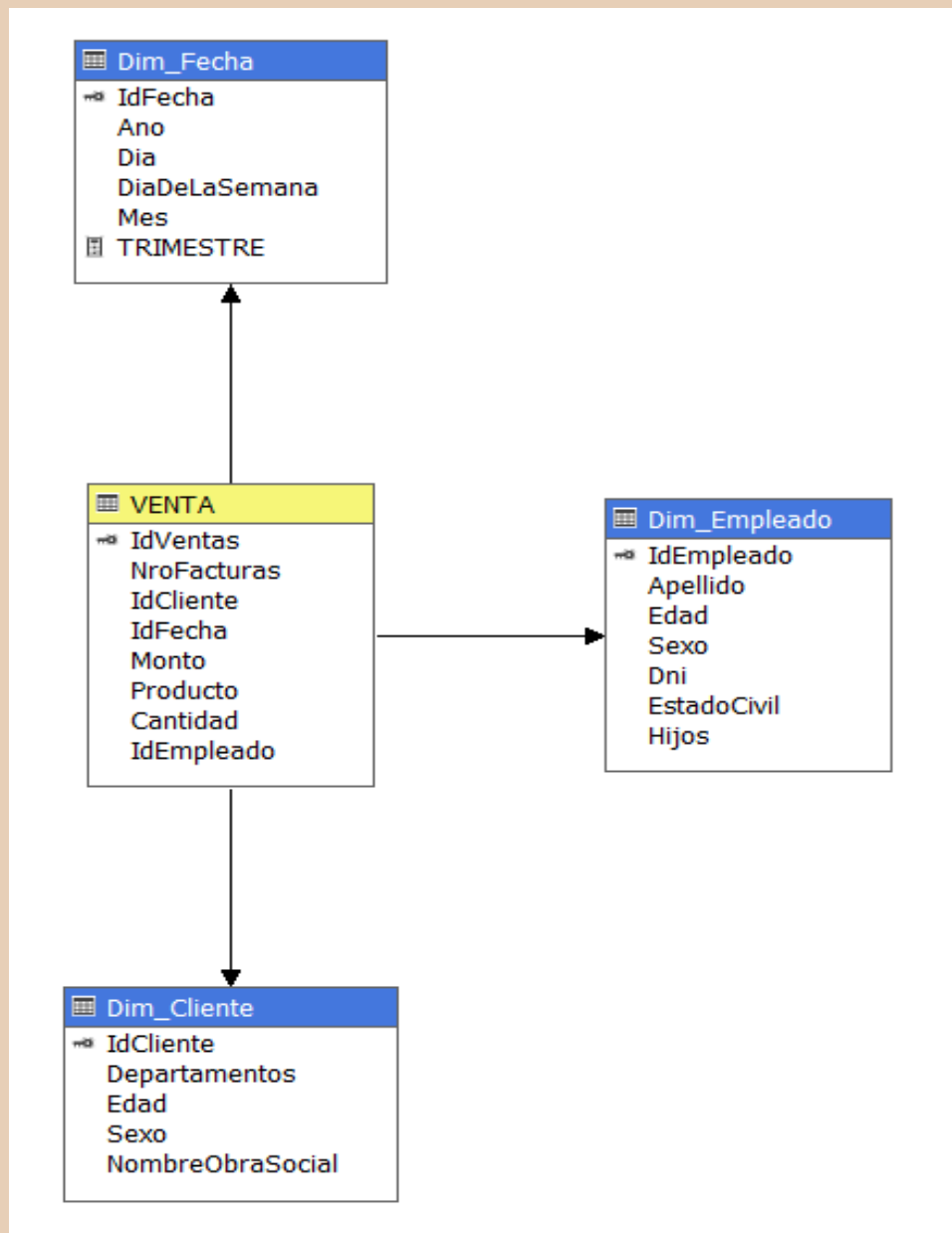
Existen tres tipos de procesos de análisis:

- ★ MOLAP: utiliza datos totales precalculados
- ★ ROLAP: Trabaja directamente sobre las bases relacionales
- ★ HOLAP: híbrido entre los dos anteriores

Para crear el cubo que utilizaremos en el proyecto utilizaremos OLAP, que es una base de datos que posee diversas dimensiones.

Cuando creamos las dimensiones (antes de crear y procesar el cubo), agregamos una columna más en la tabla Dim_Fecha. La cual consiste en agrupar los meses en semestres.

Ahora si cuelgo del dimensionado tabla por tabla, unimos todo en un cubo y nos queda de la siguiente manera.

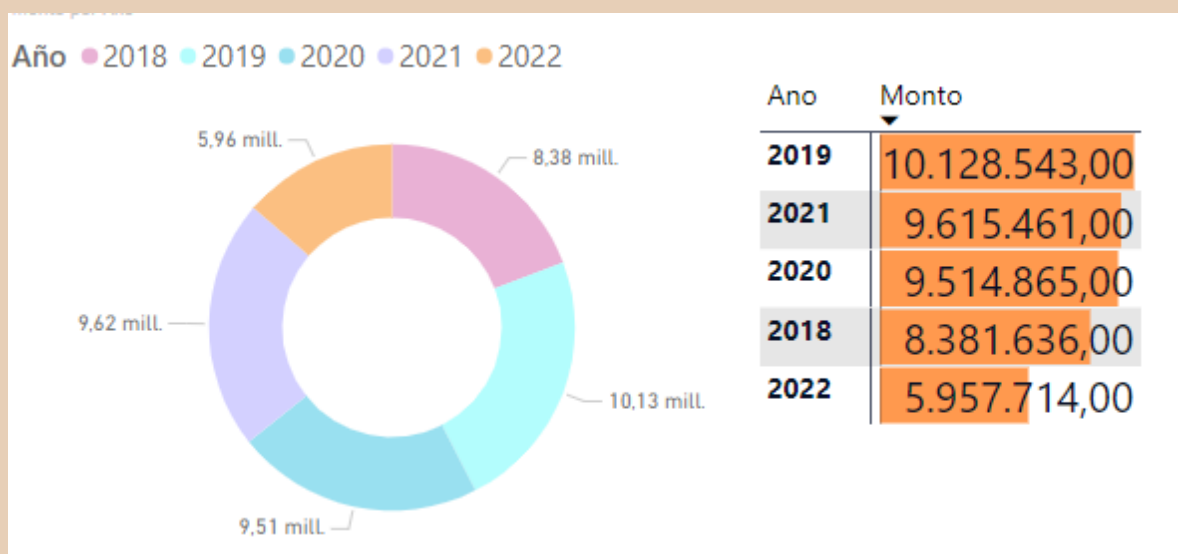


Con el cubo en pleno funcionamiento, lo utilizamos para extraer datos y obtener las respuestas a las preguntas planteadas por el cliente. Es decir, ver si hemos llegado al objetivo de nuestra Data Warehouse.

El cubo se une a la aplicación Power Bi y nos permite obtener la siguiente información:

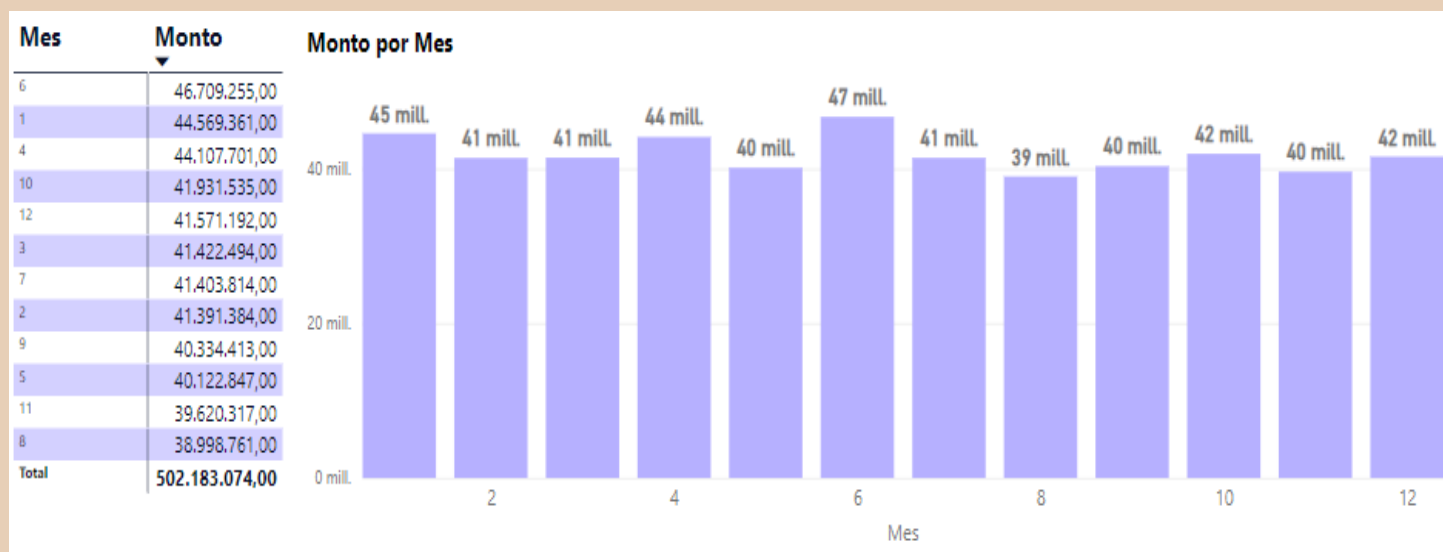
1. ¿Cuál fue de los últimos cinco años, el año con mayor facturación, y cuánto se facturó?

En los últimos 5 años el año con más facturación fue el 2019 con un monto total de ventas por \$ 10.128.543



2. ¿Cuál es el mes del año con mayor volumen de ventas y cuánto se facturó ese mes?

El mes con mayor facturación de este año fue el Junio (mes 6) , con una monto total de ventas por \$ 46.709.255



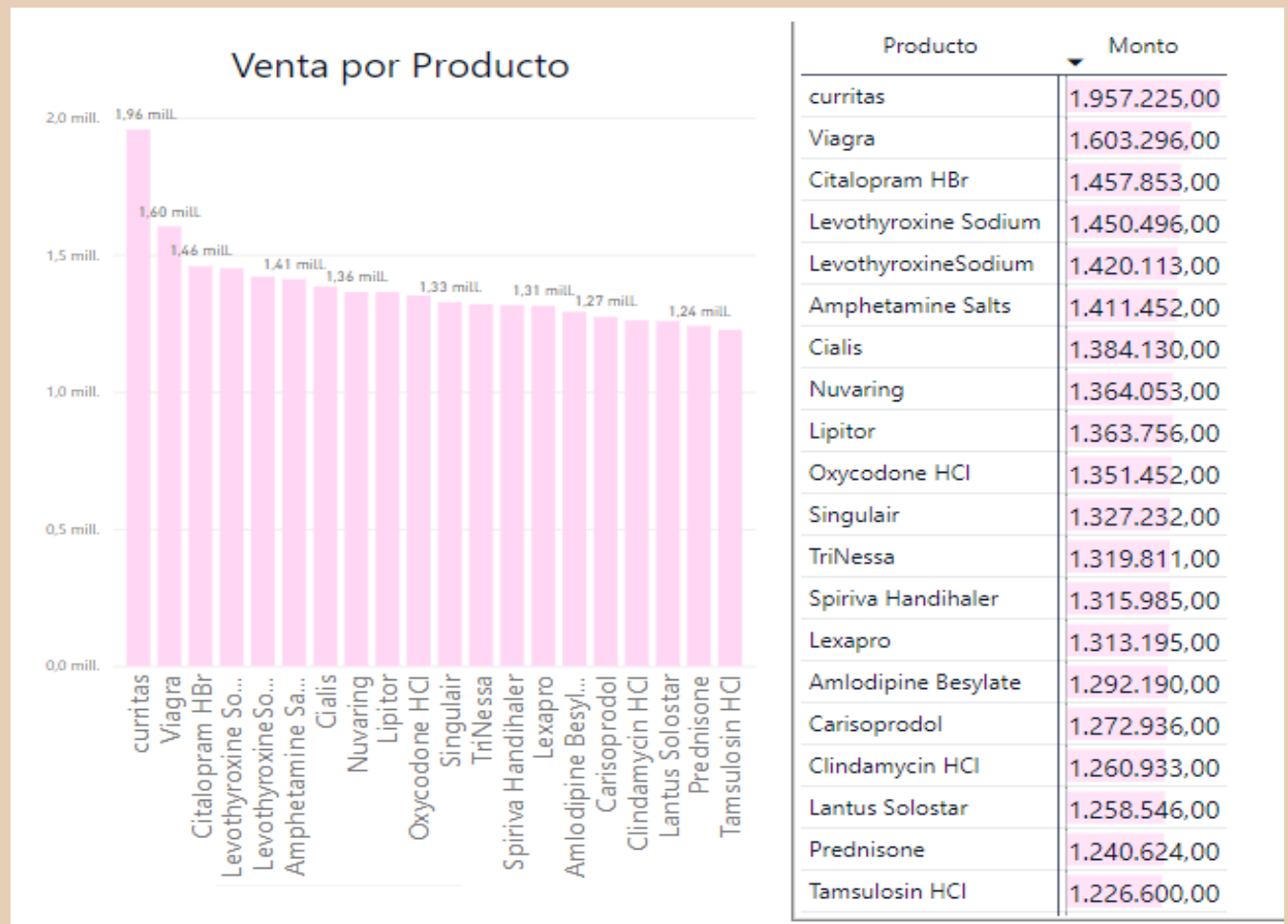
3. ¿Cuál es el día de la semana con más ventas?

El día con más ventas en la semana es el día lunes, con un monto de ventas por \$ 8.245.436



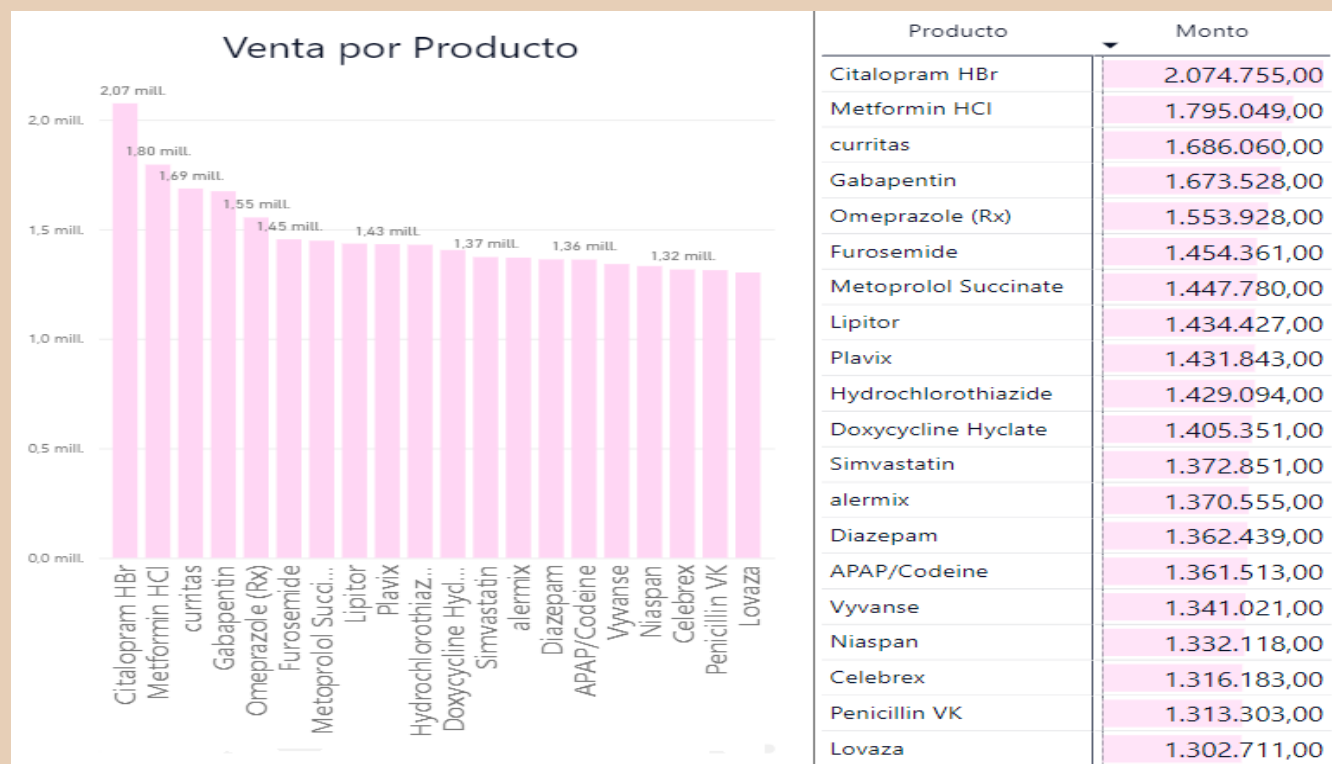
4. ¿En el primer trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?

En el primer trimestre del año los 20 productos más vendidos fueron:



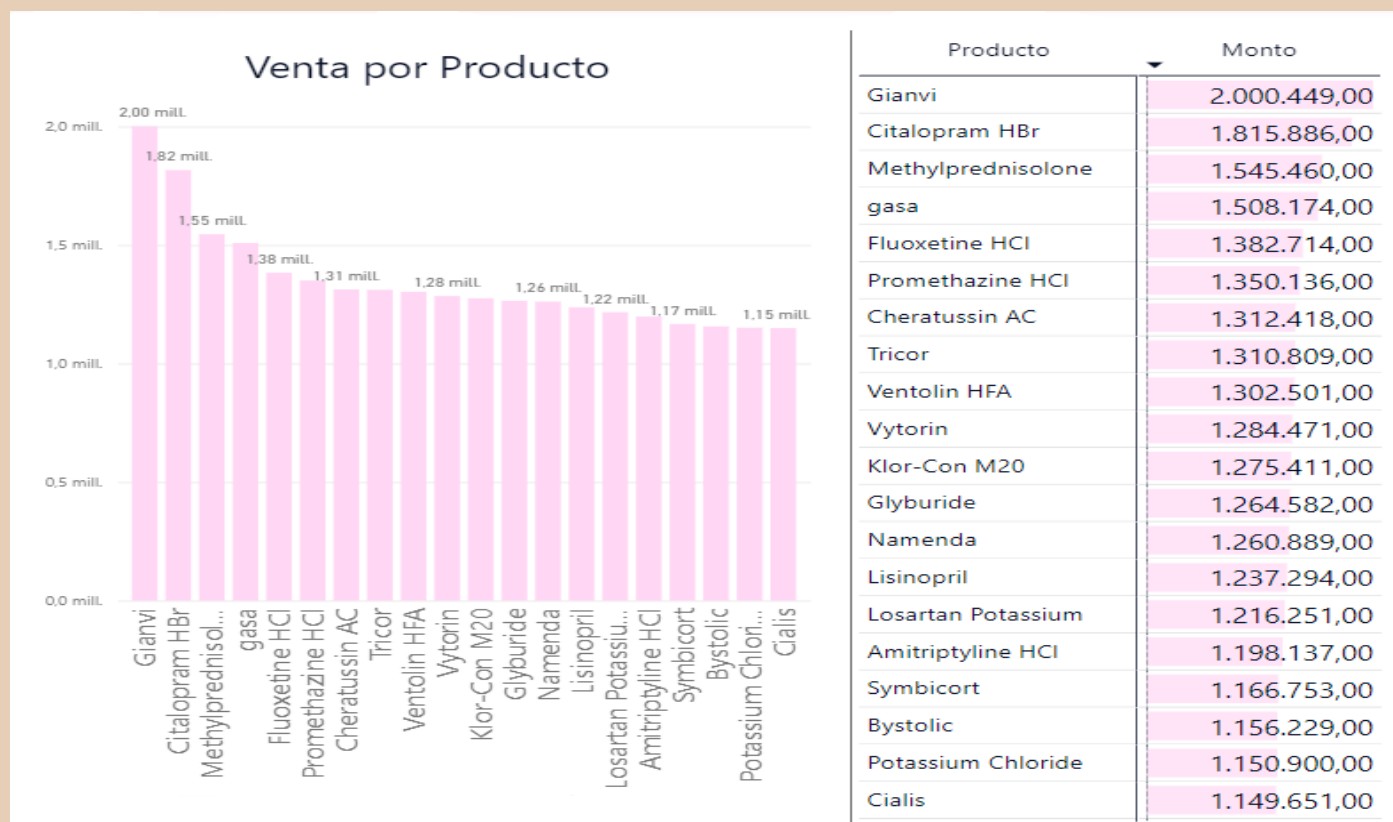
5. ¿En el segundo trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?

En el segundo trimestre del año los 20 productos más vendidos fueron:



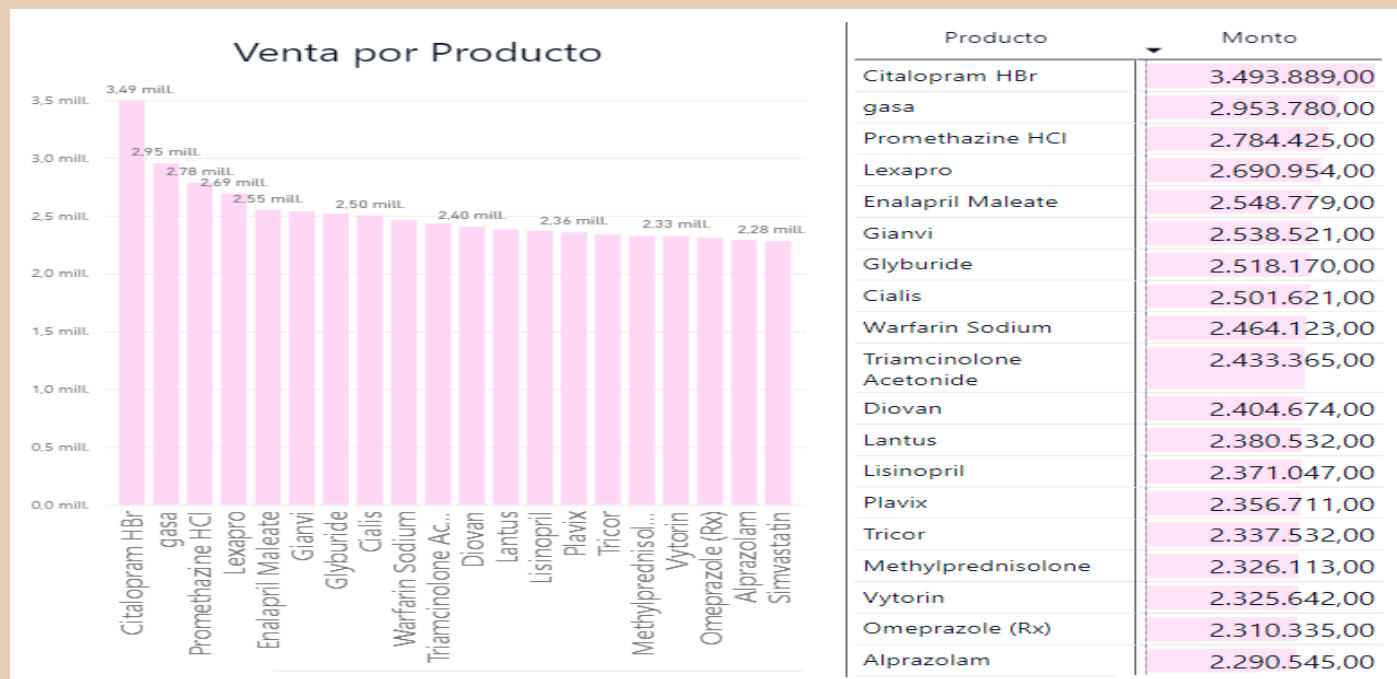
6. ¿En el tercer trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?

En el tercer trimestre del año los 20 productos más vendidos fueron:



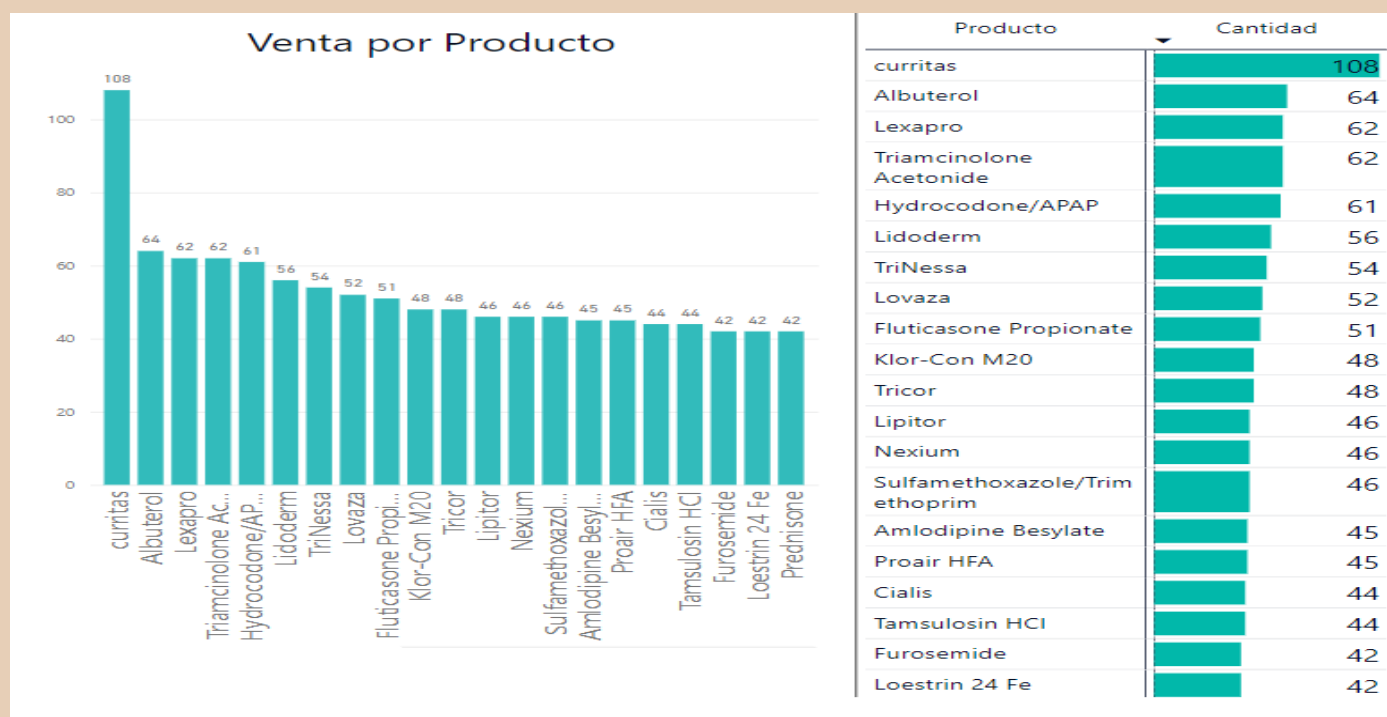
7. ¿En el cuarto trimestre del año cuáles fueron los veinte artículos más vendidos y los montos recaudados?

En el cuarto trimestre del año los 20 productos más vendidos fueron:



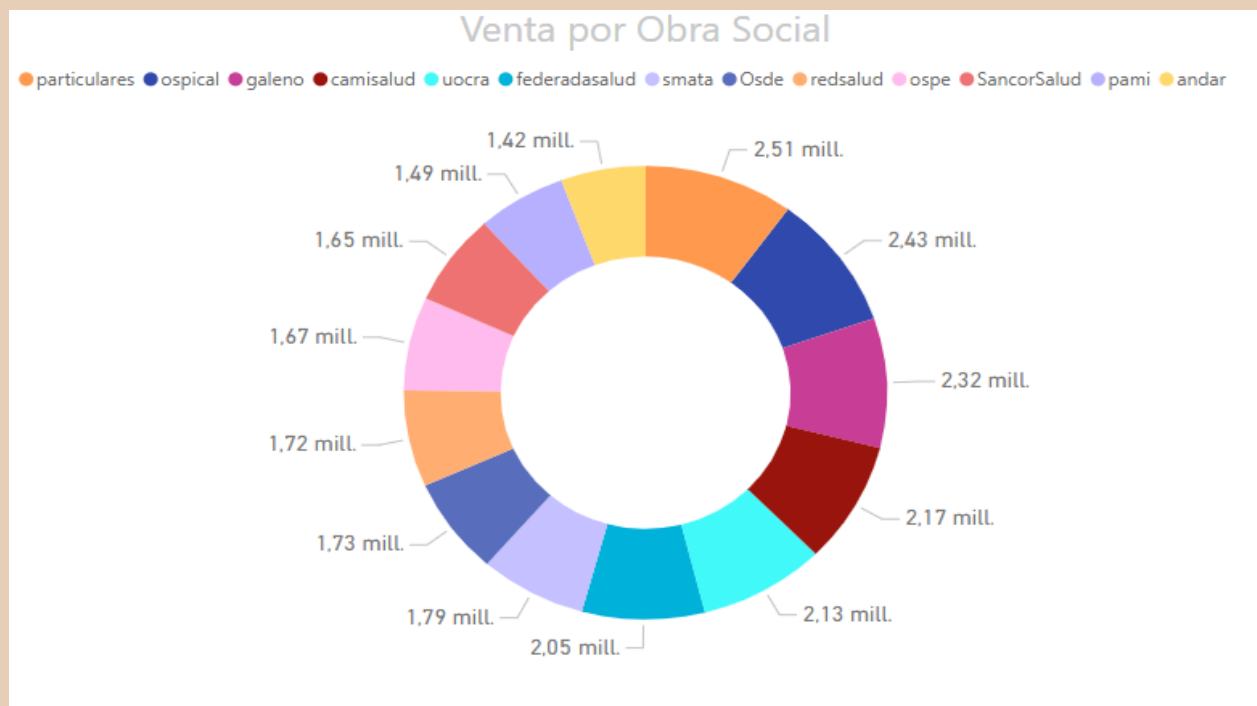
8. ¿En los últimos tres años cuáles fueron los veinte artículos más vendidos y qué cantidades?

En los últimos tres años los 20 productos más vendidos fueron:



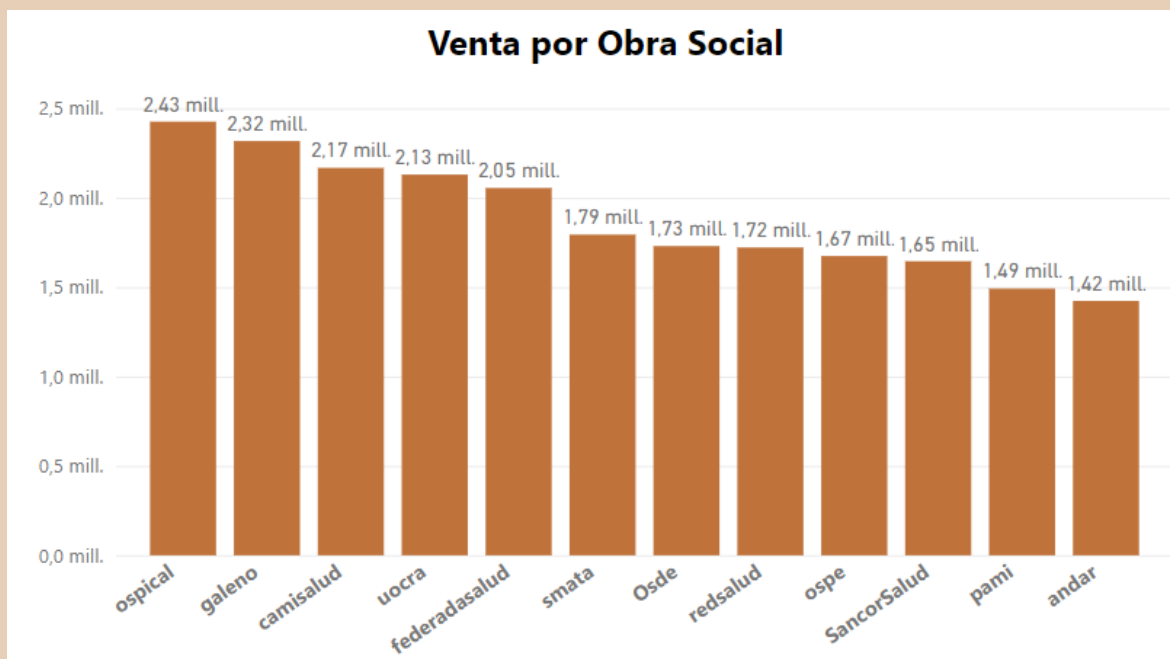
9. ¿En los últimos tres años se facturó más a clientes en forma particular o a través de obras sociales?

En los últimos tres años se facturó más a través de particulares.



10. ¿Cuál es el ranking de ventas realizadas a través de las obras sociales en los últimos tres años?

El ranking está compuesto por:



11. ¿Nuestra clientela son más mujeres, hombres o es indistinto?

Nuestra clientela está compuesta mayoritariamente por hombres, aunque la diferencia es relativamente poca.



12. ¿Cuáles fueron los diez artículos más vendidos según el género?

Los 10 artículos mas vendidos (cantidad) fueron:

Producto	femenino
Citalopram HBr	465
curritas	402
Plavix	377
Albuterol	333
Viagra	328
Januvia	327
Lidoderm	323
Promethazine HCl	323
Sulfamethoxazole/Trimethoprim	322
Ventolin HFA	322

Producto	masculino
Citalopram HBr	468
curritas	350
Bystolic	349
Klor-Con M20	348
Spiriva Handihaler	345
Advair Diskus	336
Omeprazole (Rx)	336
Triamcinolone Acetonide	328
Diazepam	323
Warfarin Sodium	323

13. ¿Cuáles son los diez artículos que más compran las personas mayores de sesenta años? ¿Cuál es la facturación semanal de esta franja etaria?

Los 10 artículos que compran los mayores de 60 son los siguientes y tienen una facturación total semana de:



14. ¿Cuáles son los diez artículos que más compran las personas entre treinta y sesenta años? ¿Cuál es la facturación semanal de esta franja etaria?

Los 10 artículos que compran las personas entre 30 y 60 son los siguientes y tienen una facturación total semana de:



15. ¿Cuáles son los diez artículos que más compran las personas menores de treinta años? ¿Cuál es la facturación semanal de esta franja etaria?

Los 10 artículos que compran las personas que tiene menos de 30 son los siguientes y tienen una facturación total semana de:



16. ¿En qué zona viven los clientes que más compraron en el mes? Hacer ranking de zonas

Los cliente que más compraron en el último mes (noviembre) viven en las siguientes zonas:

TOTAL VENTAS

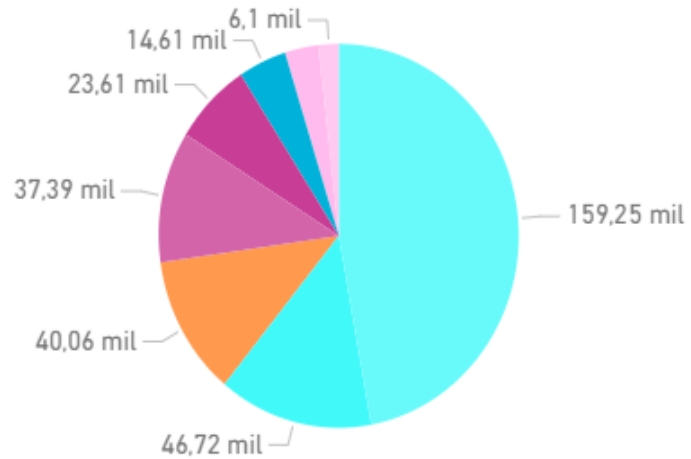
337,99 mil

Departamentos Monto

Santa Rosa	159.251,00
San Martin	46.716,00
Tunuyan	40.059,00
General Alvear	37.387,00
Lavalle	23.614,00
Rivadavia	14.612,00
Capital	10.253,00
Maipu	6.100,00

Monto por Departamentos

● Santa Rosa
 ● San Martin
 ● Tunuyan
 ● General Al...
 ● Lavalle
 ● Rivadavia
 ● Capital
 ● Maipu



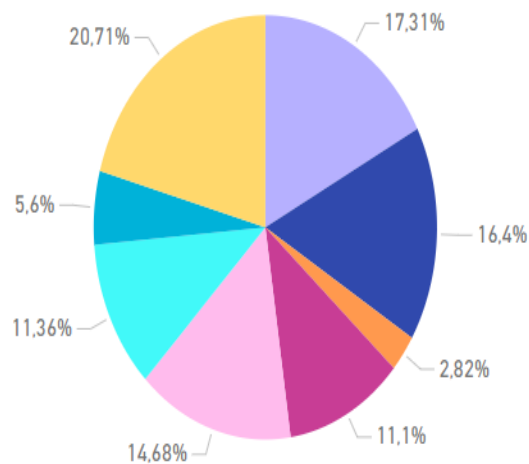
17. ¿En el último año qué empleado realizó más ventas? ¿Cuál es su perfil (soltero-casado-hijos-etc.)

En el último año el empleado que realizó más ventas fue Rodrigues, quien tiene 45 años, está soltero y no tiene hijos.

Monto por Apellido

● Gomez
 ● Lopez
 ● Mrmartinez
 ● Pepe
 ● Perez
 ● Ramirez
 ● Ramon
 ● Rodrigues

Apellido	Monto
Rodrigues	1.234.003,00
Gomez	1.031.547,00
Lopez	977.301,00
Perez	874.754,00
Ramirez	677.040,00
Pepe	661.336,00
Ramon	333.442,00
Mrmartinez	168.291,00
Total	5.957.714,00



Conclusión:

Dado los datos recabados podemos afirmar algunas de las tendencias de nuestros clientes en función de nuestras ventas fueron las siguientes:

Junio es el mes en el que más ventas se realizan, tiene sentido que sea así dado que en este mes comienza el invierno.

El día con mayor ventas a lo largo de la semana es el Lunes.

Para cada uno de los trimestres, el ítem más vendido son curitas en el primero, en el segundo y cuarto trimestre es Citalopram, y por último para el tercer trimestre es Gianvi.

El ítem más vendido de los últimos 3 años son curitas, por lo que se recomienda tener una gran disponibilidad de esto en las farmacias para evitar posibles faltas de stock y ampliar la venta de productos dedicados a los politraumatismos.

La tendencia en ventas muestra que desde hace tres años se vienen vendiendo más ítems a particulares que a clientes con una obra social en concreta.

Para los clientes que compraron usando una obra social, el ítem que más venta tuvo es Ospical.

No hay una tendencia significativa entre la cantidad de ventas entre hombres y mujeres. Para ambos sexos el ítem más vendido se trata del mismo, Citalopram.

Para el grupo etario de mayores de 60 años el ítem más vendido es Metrioprotol. El cual es un beta-bloqueador, medicamento usado para el control de algunos problemas coronarios.

En el rango de entre 30 y 60 años, el ítem más vendido es el Advair Diskus. Su monodroga es fluticasona , un antiinflamatorio esteroideo.

Y por último los menores de 30 tienen más ventas en el ítem Enalapril Maleate.

Con respecto a los departamentos, la localidad con más ventas, por lejos es Santa Rosa.

El empleado-vendedor con más ventas es Rodrigues.

Links:

❖ https://github.com/Carla-Sofia-Centeleghe/BD2_Farmcia

Anexos:

En las siguientes hojas se mostrarán ejemplos de los informes de ventas que se pueden realizar con Power Bi y la Data Warehouse.