

Proyecto: Análisis de Ventas con SQL

Resumen Ejecutivo

Este proyecto busca transformar datos transaccionales del dataset Superstore Sales en información primordial para la toma de decisiones. Para lograrlo se diseñó un modelo analítico optimizado que permite evaluar el desempeño de ventas, analizar tendencias temporales y detectar oportunidades de mejora a nivel de producto, categoría y región.

El resultado final es un dashboard interactivo orientado a negocio, capaz de responder preguntas clave de la gerencia de manera clara, rápida y visual.

Contexto del negocio

Una empresa minorista registra información de sus ventas, productos y clientes.

La gerencia necesita comprender el comportamiento de las ventas para identificar oportunidades de mejora, optimizar el catálogo de productos y apoyar la toma de decisiones estratégicas.

Este proyecto utiliza **SQL** como herramienta principal para analizar los datos y responder preguntas de negocio relevantes.

Objetivo del proyecto

Analizar los datos de ventas para:

- Identificar productos y categorías más rentables
- Analizar la evolución de las ventas en el tiempo
- Detectar clientes con mayor contribución a los ingresos
- Encontrar oportunidades de mejora en productos de bajo rendimiento

El enfoque del proyecto es **analítico y orientado a negocio**, no predictivo.

Herramientas utilizadas

- **SQL (SQLite)**: análisis y consultas de datos
- **Power BI**: visualización de resultados y KPIs
- **GitHub**: control de versiones y documentación

Transformación y modelado de datos

Relacional a Analítico

El proyecto parte de un modelo relacional transaccional compuesto por múltiples tablas (orders, order_details, products, categories, regions). Para facilitar el análisis, se realizó una transformación hacia un Modelo Estrella, separando métricas y atributos de negocio.

Se creó una **tabla de hechos de ventas** y múltiples **tablas de dimensiones**, optimizando el rendimiento y simplificando el análisis en Power BI.

```
-- =====
-- Tabla: customers
-- =====
CREATE TABLE customers(
    customer_id TEXT PRIMARY KEY,
    customer_name TEXT,
    segment TEXT
);

CREATE TABLE regions(
    region_id INTEGER PRIMARY KEY AUTOINCREMENT,
    country TEXT,
    region TEXT,
    state TEXT,
    city TEXT,
    postal_code TEXT
);

CREATE TABLE categories(
    category_id INTEGER PRIMARY KEY AUTOINCREMENT,
    category TEXT,
    sub_category TEXT
);

CREATE TABLE products(
    product_id TEXT PRIMARY KEY,
    product_name TEXT,
    category_id INTEGER,
    FOREIGN KEY (category_id) REFERENCES categories(category_id)
);

CREATE TABLE orders(
    order_id TEXT PRIMARY KEY,
    order_date DATE,
    ship_date DATE,
    ship_mode TEXT,
    customer_id TEXT,
    region_id INTEGER,
    FOREIGN KEY (customer_id) REFERENCES customers(customer_id),
    FOREIGN KEY (region_id) REFERENCES regions(region_id)
);

-- =====
-- TABLA DE HECHOS
-- =====

CREATE TABLE order_details(
    row_id INTEGER PRIMARY KEY,
    order_id TEXT,
    product_id TEXT,
    sales REAL,
    FOREIGN KEY (order_id) REFERENCES orders(order_id),
    FOREIGN KEY (product_id) REFERENCES products(product_id)
);
```

```
1 -- =====
2 -- Paso 0: Activar integridad referencial
3 -- =====
4 PRAGMA foreign_keys = ON;
5
6 ---=====
7 --- Poblar categories
8 ---=====
9 INSERT INTO categories (category,sub_category)
10 SELECT DISTINCT category, sub_category
11 FROM staging_superstore;
12
13 ---=====
14 --- Poblar Regions
15 ---=====
16 INSERT INTO regions (country,region,state,city,postal_code)
17 SELECT DISTINCT country, region , state , city , postal_code
18 FROM staging_superstore;
19
20 ---=====
21 --- Poblar Customers
22 ---=====
23 INSERT INTO customers (customer_id,customer_name,segment)
24 SELECT DISTINCT customer_id, customer_name, segment
25 FROM staging_superstore;
26
27 ---=====
28 --- Poblar products
29 ---=====
30 INSERT INTO products (product_id,product_name,category_id)
31 SELECT s.product_id, s.product_name , c.category_id
32 FROM staging_superstore s
33 JOIN categories c ON s.category = c.category
34 AND s.sub_category = c.sub_category
35 GROUP BY s.product_id;
36
37 ---=====
38 --- Poblar Orders
39 ---=====
40 INSERT INTO orders(order_id,order_date,ship_date,ship_mode,customer_id,region_id)
41 SELECT s.order_id, s.order_date, s.ship_date, s.ship_mode, s.customer_id,r.region_id
42 FROM staging_superstore s
43 JOIN regions r
44 ON s.country = r.country
45 AND s.region = r.region
46 AND s.state = r.state
47 AND s.city = r.city
48 AND s.postal_code = r.postal_code
49 GROUP BY s.order_id;
50
51
52 =====
53 ---Poblar order_details(HECHOS)
54 =====
55 INSERT INTO order_details(row_id,order_id,product_id,sales)
56 SELECT row_id, order_id, product_id, sales
57 FROM staging_superstore;
58
59 ---Conteos Básicos
60 SELECT COUNT(*) FROM order_details;
61 SELECT COUNT(*) FROM orders;
62 SELECT COUNT(*) FROM customers;
63 SELECT COUNT(*) FROM products;
64
65 ---Veremos las 5 filas de la tabla de hechos
66 SELECT * FROM order_details LIMIT 5;
```

CONSULTAS REALIZADAS

```

---Consultas claves

--Ventas Totales
SELECT ROUND(SUM(sales),2) AS total_sales
FROM order_details;

--Ventas por categoria
SELECT c.category,
       ROUND(SUM(od.sales),2) AS total_sales
FROM order_details od
JOIN products p ON od.product_id = p.product_id
JOIN categories c ON p.category_id = c.category_id
GROUP BY c.category
ORDER BY total_sales DESC;

--TOP DE 10 CLIENTES POR VENTAS
SELECT cu.customer_name, ROUND(SUM(od.sales),2) AS total_sales
FROM order_details od
JOIN orders o ON o.order_id = od.order_id
JOIN customers cu ON o.customer_id = cu.customer_id
GROUP BY cu.customer_name
ORDER BY total_sales DESC
LIMIT 10;

--VENTAS POR REGION
SELECT r.region, ROUND(SUM(od.sales),2) AS total_sales
FROM order_details od
JOIN orders o ON od.order_id = o.order_id
JOIN regions r ON o.region_id = r.region_id
GROUP BY r.region
ORDER BY total_sales DESC;

--TENDENCIAS DE VENTAS POR AÑO
SELECT substr(o.order_date,7,4) AS year,
       ROUND(SUM(od.sales),2) AS total_sales
FROM order_details od
JOIN orders o ON od.order_id = o.order_id
GROUP BY year
ORDER BY year;

SELECT sales FROM order_details;
--RANKING DE PRODUCTOS POR VENTAS
SELECT p.product_name,
       ROUND(SUM(od.sales),2) total_sales,
       RANK() OVER(ORDER BY SUM(od.sales) DESC) AS sales_rank
FROM order_details od
JOIN products p ON od.product_id = p.product_id
GROUP BY product_name
ORDER BY sales_rank;

```

DIAGRAMA ER

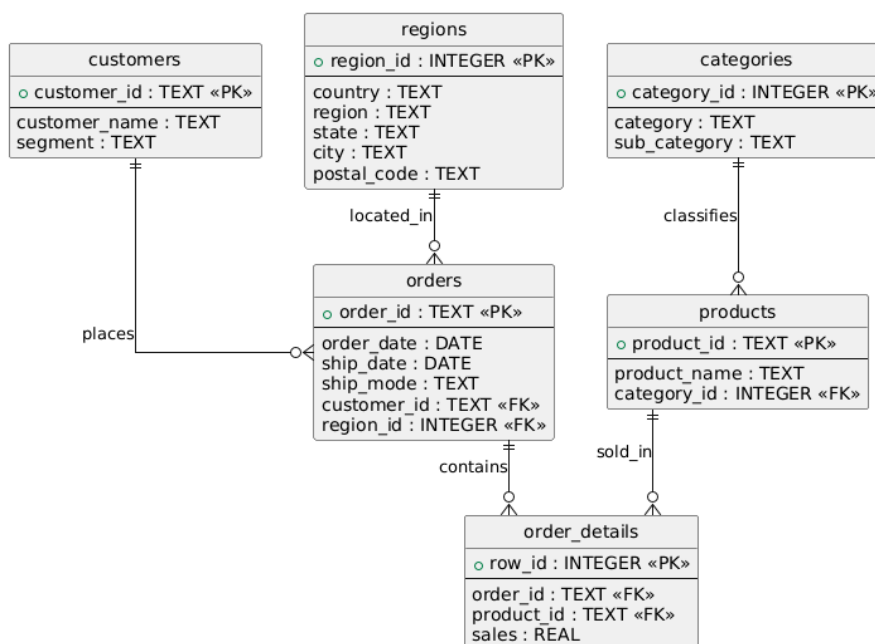
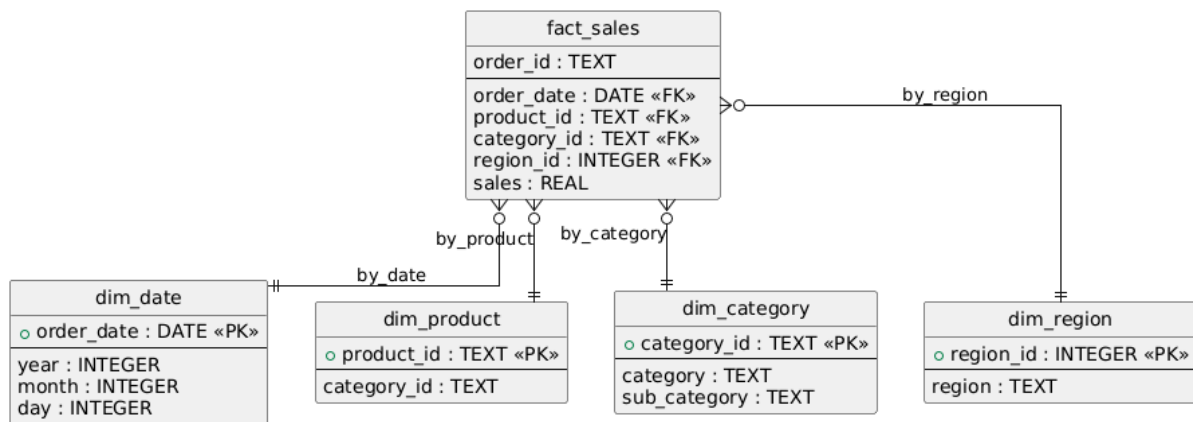
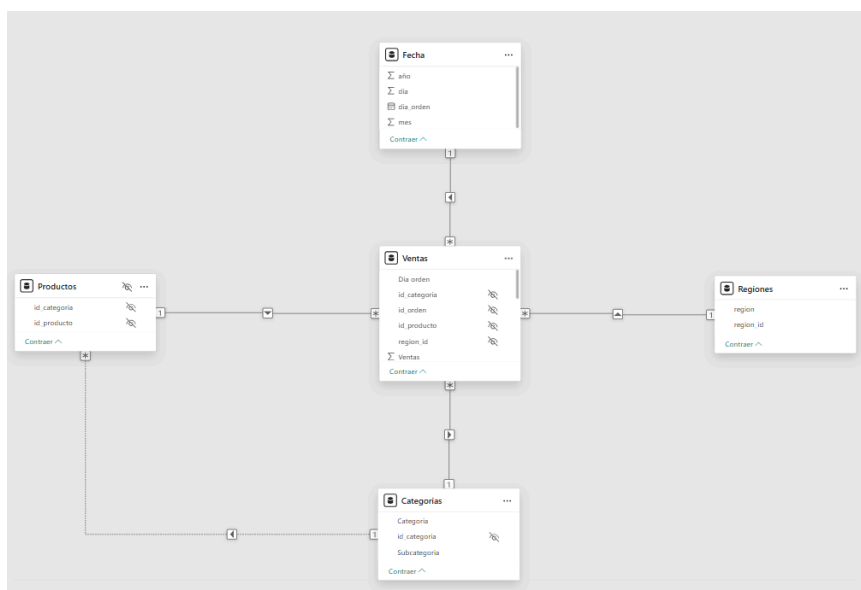


DIAGRAMA ESTRELLA



Se renombro las tablas a términos de negocio como Ventas, Productos, Categorías, Regiones, Fechas, por otro lado se ocultaron IDs técnicos y campos irrelevantes



Desarrollo de KPIs y Lógica DAX

Para responder a las preguntas de negocio, se desarrollaron las siguientes métricas clave utilizando lenguaje DAX:

Análisis de Rentabilidad y Volumen

- **Ventas Totales:** Métrica base para el cálculo de ingresos
- **Total de Pedidos:** Conteo de transacciones únicas para medir volumen.
- **Ticket Promedio:** Relación entre ventas y pedidos para medir el valor medio de compra.

Análisis Comparativo (YoY- Year over Year)

- **Ventas del Año Actual vs Ventas del año anterior:** Permiten la comparación directa de periodos.

- **Variación Ventas Año a Año:** Cálculo porcentual del crecimiento o decrecimiento anual.
- **Indicador de Crecimiento:** Métrica visual para identificar tendencias positivas o negativas rápidamente.

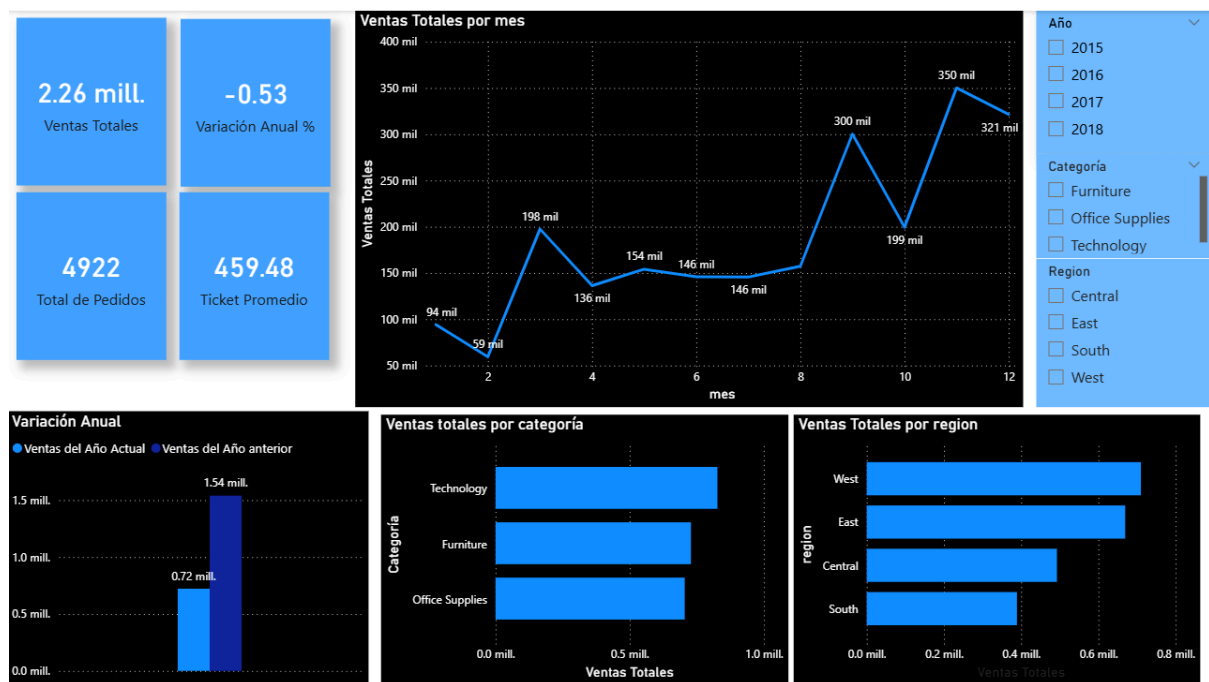
Análisis de Participación Ranking

- **Participación de Ventas % :** Determina el peso de cada categoría o región sobre el total.
- **% Contribución del Periodo:** Analiza la relevancia de un tiempo específico en el total acumulado.
- **Ranking de Categoría:** Clasifica automáticamente las categorías por volumen de ingresos.

Promedios e Históricos

- **Promedios Ventas Mensuales:** Establece una línea base de rendimiento mensual.
- **Ventas Históricas/ Sales All Time:** Acumulado total de la operación registrada.

Análisis de Resultados (Insights de Negocio)



Basado en el análisis del dashboard final, se obtuvieron los siguientes hallazgos:

Ventas Totales

- El negocio alcanzó 2.26 millones en ingresos totales durante el periodo analizado.

Desempeño por Categoría

- La categoría Technology lidera el mercado, con ventas cercanas al millón, consolidándose como el principal motor de ingresos.

Tendencia Temporal

- Se identifica una estacionalidad positiva, con un pico máximo en el mes de noviembre alcanzando aproximadamente 350 mil en ventas.

Análisis Geográfico

- La región **Oeste** presenta el mejor desempeño, mientras que la región **Sur** muestra el menor volumen de ventas y requiere atención estratégica.

Dataset

Dataset de ventas minoristas (Superstore Sales), que contiene información sobre:

- Pedidos
- Productos
- Categorías
- Clientes
- Fechas de venta
- Ingresos y costos

El dataset fue utilizado con fines educativos y de análisis.

Preguntas de negocio

Las principales preguntas que guían el análisis son:

1. ¿Cuáles son los productos con mayor volumen de ventas?

Los productos con mayor volumen de ventas fueron identificados mediante una consulta SQL que sumó las cantidades vendidas. Los principales incluyen:

- Canon imageCLASS 2200 Advanced Copier, con un total de 61599.82.
 - Fellowes PB500 Electric, con un total de 27453.38.
 - Cisco TelePresence System EX90 Videoconferencing Unit, con un total de 22638.48.
 - HON 5400 Series Task Chairs for Big and Tall, con un total de 21870.58.
2. ¿Qué categorías generan más ingresos?
 - La categoría Technology es la que más ingresos genera, con ventas que registran 827455.87 en el total , representando aproximadamente el **36.59 %** de los ingresos totales. Le sigue la categoría **Furniture** y luego **Office Supplies**.
 3. ¿Cómo se comportan las ventas a lo largo del tiempo?
 - Las ventas muestran una clara estacionalidad, con un pico máximo en noviembre, coincidiendo con las campañas de fin de año. Se observa una caída previa en octubre, lo que sugiere una oportunidad para campañas de anticipación.
 4. ¿Qué clientes aportan mayor valor al negocio?

- El análisis de ventas por cliente revela que Sean Miller es el cliente que más valor aporta, con \$25,043.05 en ventas totales, representando aproximadamente el 1.11% de los ingresos acumulados. Le siguen Tamara Chand (\$19,052.22) y Raymond Buch (\$15,117.34). Los 15 principales clientes concentran cerca del 9.37% de las ventas totales, lo que indica una base de clientes diversa pero con claros líderes en contribución.
5. ¿Existen productos con baja rentabilidad que deban revisarse?
- A nivel de categoría, Technology es la más rentable con \$827,455.87, seguida de Furniture (\$728,658.58) y Office Supplies (\$705,422.33). Si bien ninguna categoría es de "baja rentabilidad" en términos absolutos, un análisis a nivel de producto dentro de Furniture y Office Supplies podría revelar artículos con márgenes bajos o ventas estancadas que requieran revisión estratégica, reasignación de inventario o ajustes de precios.

Estas preguntas permiten enfocar el análisis en métricas relevantes para la toma de decisiones.

Enfoque de análisis

El proyecto sigue un enfoque estructurado:

1. Comprensión del negocio
2. Exploración del dataset
3. Análisis mediante consultas SQL
4. Extracción de insights relevantes
5. Visualización de resultados en Power BI

Entregables

- Consultas SQL documentadas
- Insights de negocio basados en datos
- Dashboard interactivo en Power BI
- Documentación clara del proceso y conclusiones

Conclusiones y Recomendaciones

- Este proyecto demostró cómo, a partir de un dataset transaccional, se pueden extraer insights accionables mediante SQL para el análisis y Power BI para la visualización. Se logró:
- Identificar que Technology es el principal motor de ingresos (\$827K), seguida de cerca por Furniture.

- Reconocer a Sean Miller como el cliente de mayor valor, con ventas superiores a los \$25K.
- Visualizar la estacionalidad de ventas con un pico en noviembre (\$350K).
- Detectar que la región Sur requiere atención estratégica por su bajo volumen.

Recomendación Estratégica

Se recomienda implementar campañas de retención y activación de clientes en el mes de octubre, dado que se evitaría la caída de ventas observadas antes del pico de noviembre y mejorar el desempeño anual.