

Análisis de Ventas y Envíos en Superstore

Presentación:

Mi nombre es Juan Segundo Gnarra, soy estudiante de actuar0. He realizado los cursos de Power BI y Data Science en Coderhouse. Con este trabajo práctico, espero seguir aprendiendo sobre el análisis de datos poder desarrollarme profesionalmente, principalmente en SQL

Resumen del Dataset:

El dataset utilizado en este análisis contiene información detallada sobre las transacciones de ventas y envíos realizadas en Superstore. Consta de 10000 filas y 18 columnas que registran datos como las fechas de pedido y envío, modos de envío, categorías de productos, ventas, descuentos, ganancias y costos de envío.

Descripción de las Columnas:

1. Row ID: Identificador único para cada fila.
2. Order ID: Identificador único para cada pedido.
3. Order Date: Fecha del pedido.
4. Ship Date: Fecha de envío.
5. Ship Mode: Modo de envío (por ejemplo, envío estándar, segunda clase, etc.).
6. Segment: Segmento de mercado (por ejemplo, consumidor, oficina en casa, corporativo).
7. City: Ciudad del cliente.
8. State: Estado del cliente.
9. Country: País del cliente.
10. Category: Categoría del producto.
11. Sub-Category: Subcategoría del producto.
12. Product Name: Nombre del producto.
13. Sales: Ventas del producto.
14. Quantity: Cantidad de productos vendidos.
15. Discount: Descuento aplicado.
16. Profit: Ganancia obtenida.
17. Shipping Cost: Costo de envío.
18. Order Priority: Prioridad del pedido (por ejemplo, alta, media, baja).

Descripción de la temática de los datos (Metodología SMART):

- Specific: Análisis de las ventas, costos de envío y ganancias de productos vendidos en Superstore, centrado en diferentes categorías de productos y regiones.
- Measurable: Evaluar el impacto de diferentes categorías de productos, modos de envío y regiones en las ventas y ganancias totales.
- Achievable: Utilizando datos históricos de pedidos desde 2011 hasta 2013.
- Relevant: Entender los patrones de ventas y envíos para optimizar las estrategias de marketing y gestión de inventario.
- Time-bound: Datos desde 2011 hasta 2013.

Hipótesis:

"Las ciudades con mayor volumen de ventas también son las que generan las mayores ganancias para Superstore."

Objetivo del Proyecto:

El objetivo de este proyecto es analizar las ventas y envíos de Superstore para identificar patrones y tendencias que puedan optimizar las estrategias de marketing y gestión de inventario. El análisis se centrará en diferentes categorías de productos y regiones, evaluando el impacto de diversos factores como los modos de envío y los descuentos aplicados.

Alcance:

El análisis abarca los datos de transacciones de ventas y envíos realizados en Superstore desde 2011 hasta 2013. Se evaluarán las ventas, los costos de envío y las ganancias de productos vendidos en diversas categorías y regiones. Este análisis ayudará a entender mejor los patrones de ventas y envíos, facilitando la toma de decisiones estratégicas.

Usuario Final y Nivel de Aplicación del Análisis

Usuario Final:

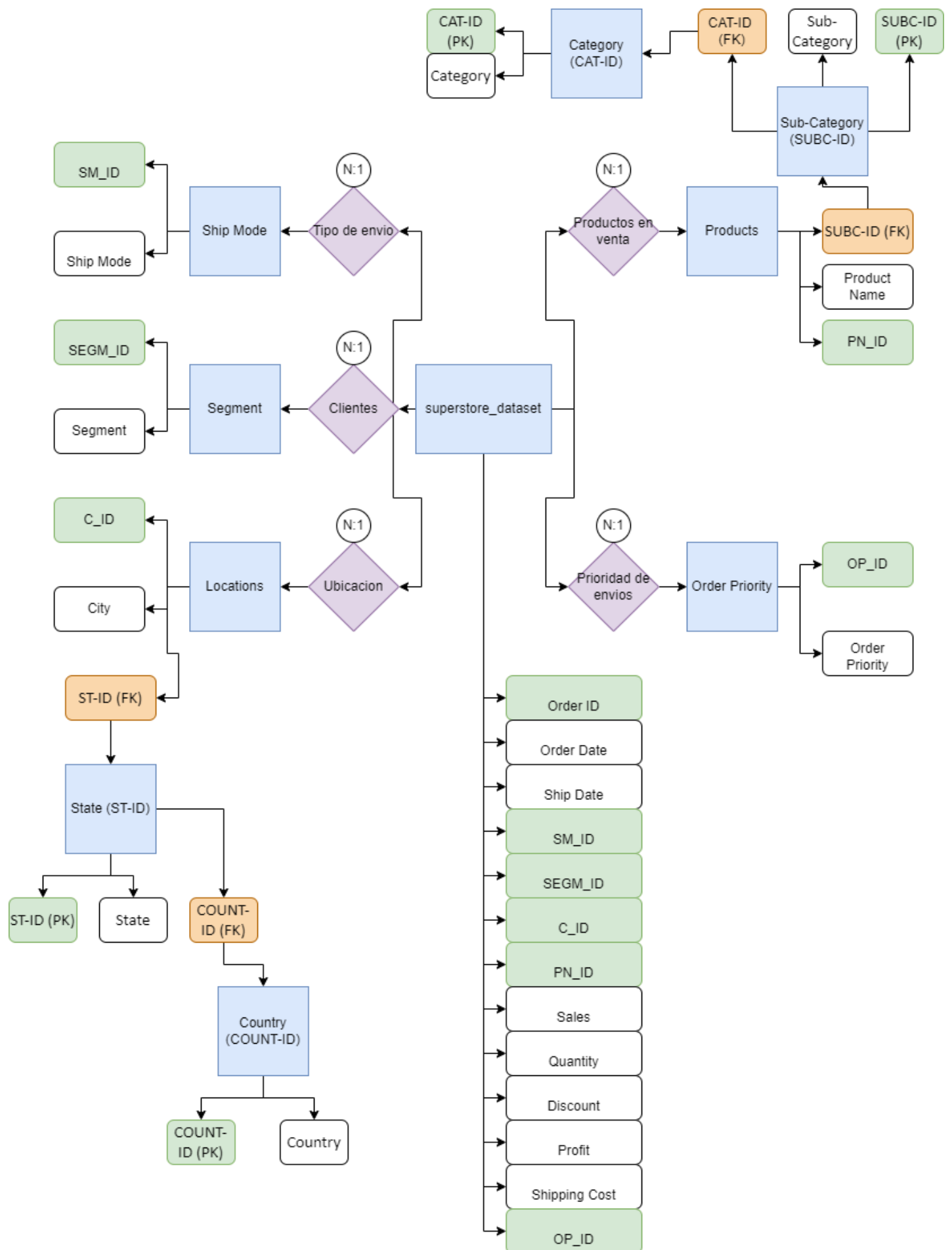
- Gerentes de ventas y marketing.
- Equipo de logística.

Nivel de Aplicación del Análisis:

- **Operativo:** Optimización de las operaciones diarias de ventas y envíos.
- **Táctico:** Mejora de estrategias de marketing y promociones a corto y mediano plazo.
- **Estratégico:** Planeación y gestión a largo plazo para aumentar la rentabilidad y eficiencia.

Diagrama de Entidad Relación

El diagrama de entidad relación (ERD) ilustra cómo se relacionan las distintas entidades dentro del dataset de Superstore. A continuación se presenta el diagrama que muestra las tablas y sus relaciones:



Listado de Tablas

1. Superstore_dataset

Descripción: Esta tabla contiene información sobre cada pedido realizado en Superstore.

- **Llave Primaria:** `Order ID`
- **Llaves Foráneas:**
 - `Customer ID` (relacionada con la tabla Customers)
 - `Ship Mode ID` (relacionada con la tabla Ship Modes)
 - `Segment ID` (relacionada con la tabla Segments)
 - `Order Priority ID` (relacionada con la tabla Order Priorities)
- **Relaciones:**
 - Cada pedido está asociado a un cliente específico.
 - Cada pedido tiene un modo de envío, un segmento de mercado y una prioridad.

2. Customers (Clientes)

Descripción: Esta tabla contiene información sobre los clientes que realizan pedidos en Superstore.

- **Llave Primaria:** `Customer ID`
- **Relaciones:**
 - Un cliente puede realizar múltiples pedidos.

3. Ship Modes (Modos de Envío)

Descripción: Esta tabla contiene los diferentes modos de envío disponibles en Superstore.

- **Llave Primaria:** `Ship Mode ID`
- **Relaciones:**
 - Un modo de envío puede estar asociado a múltiples pedidos.

4. Segments (Segmentos)

Descripción: Esta tabla contiene los diferentes segmentos de mercado (por ejemplo, consumidor, oficina en casa, corporativo).

- **Llave Primaria:** `Segment ID`
- **Relaciones:**
 - Un segmento puede estar asociado a múltiples pedidos.

5. Order Priorities (Prioridades de Pedido)

Descripción: Esta tabla contiene las diferentes prioridades que se pueden asignar a un pedido.

- **Llave Primaria:** `Order Priority ID`
- **Relaciones:**
 - Una prioridad de pedido puede estar asociada a múltiples pedidos.

6. Products (Productos)

Descripción: Esta tabla contiene información sobre los productos disponibles en Superstore.

- **Llave Primaria:** `Product Name ID`
- **Relaciones:**
 - Un producto puede estar asociado a múltiples pedidos.

7. Locations (Ubicaciones)

Descripción: Esta tabla contiene información sobre las ubicaciones de los clientes (ciudad, estado, país).

- **Llave Primaria:** `City ID`
- **Relaciones:**
 - Una ubicación puede estar asociada a múltiples clientes.

Listado de Columnas

- **Superstore_dataset**

Descripción: Esta tabla contiene información sobre cada pedido realizado en Superstore.

- **Order ID:** Identificador único para cada pedido (PK)
- **Order Date:** Fecha del pedido
- **Ship Date:** Fecha de envío
- **SM_ID:** Identificador del modo de envío (FK)
- **SEGM_ID:** Identificador del segmento de mercado (FK)
- **C_ID:** Identificador único del cliente (FK)
- **PN_ID:** Identificador único del cliente (FK)
- **Sales:** Ventas del producto
- **Quantity:** Cantidad de productos vendidos
- **Discount:** Descuento aplicado
- **Profit:** Ganancia obtenida
- **Shipping Cost:** Costo de envío
- **Order Priority ID:** Identificador del producto (FK)

- **Ship Modes (Modos de Envío)**

Descripción: Esta tabla contiene los diferentes modos de envío disponibles en Superstore.

- **SM_ID:** Identificador del modo de envío (PK)
- **Ship Mode:** Modo de envío (por ejemplo, envío estándar, segunda clase, etc.)

- **Segments (Segmentos)**

Descripción: Esta tabla contiene los diferentes segmentos de mercado (por ejemplo, consumidor, oficina en casa, corporativo).

- **SEGM_ID:** Identificador del segmento de mercado (PK)
- **Segment:** Segmento de mercado (por ejemplo, consumidor, oficina en casa, corporativo)

- **Locations (Ubicaciones)**

Descripción: Esta tabla contiene información sobre las ubicaciones de los clientes (ciudad, estado, país).

- **C_ID:** Identificador único de la ciudad (PK)
- **City:** Ciudad del cliente
- **State:** Estado del cliente
- **Country:** País del cliente

- **Products (Productos)**

Descripción: Esta tabla contiene información sobre los productos disponibles en Superstore.

- **PN_ID:** Identificador único del nombre del producto (PK)
- **Product Name:** Nombre del producto
- **Category:** Categoría del producto
- **Sub-Category:** Subcategoría del producto

- **Order Priorities (Prioridades de Pedido)**

Descripción: Esta tabla contiene las diferentes prioridades que se pueden asignar a un pedido.

- **OP_ID:** Identificador de la prioridad del pedido (PK)
- **Order Priority:** Prioridad del pedido (por ejemplo, alta, media, baja)

Transformaciones Realizadas

Limpieza de datos

1. **Encabezados promovidos:** De todas las tablas cargadas se utilizó la primera fila como encabezado
2. **Borrar filas vacías:** Utilizando la función de conservar las filas superiores se borraron las celdas vacías que se trajeron desde la fuente de datos.
3. **Controlar los tipos de datos:** Se controló que los formatos de las columnas estén bien asignadas para el tipo de dato
4. **Nueva columna:** Se agregó una nueva columna en la hoja “superstore_dataset” llamada “TiempoDeEnvio” en el cual se restan las columnas de “Order Date” y “Ship Date” para calcular el tiempo que tarda en llegar el producto al cliente.

Medidas calculadas generadas

- Costo Envio Promedio = $\text{[Costos Envios]} / 9998$
- Costos Envios = $\text{SUM}(\text{superstore_dataset}[\text{Shipping Cost}])$
- Countrows = $\text{COUNTROWS}(\text{'superstore_dataset'})$
- MaxVentas = $\text{MAXX}(\text{ALLSELECTED}(\text{'superstore_dataset'}), [\text{Sales}])$
- MinVentas = $\text{MINX}(\text{ALLSELECTED}(\text{'superstore_dataset'}), [\text{Sales}])$
- PromedioTiempoDeEnvio = $\text{AVERAGE}(\text{'superstore_dataset'}[\text{TiempoDeEnvio}])$
- Rentabilidad media = $\text{[Rentabilidad Total]} / \text{[Recuento de ventas]}$
- Recuento de ventas = $\text{COUNT}(\text{superstore_dataset}[\text{Order ID}])$
- Rentabilidad Total = $\text{SUM}(\text{superstore_dataset}[\text{Profit}])$
- Suma DiaEnvio = $\text{SUM}(\text{superstore_dataset}[\text{Ship Date}])$
- Suma orden = $\text{SUM}(\text{superstore_dataset}[\text{Order Date}])$
- Total de Compras = $\text{COUNTROWS}(\text{superstore_dataset})$
- Total de Descuento = $\text{SUM}(\text{superstore_dataset}[\text{Discount}])$
- Total de Rentabilidad Año Anterior = $\text{CALCULATE}(\text{[Rentabilidad total]}, \text{DATEADD}(\text{FILTER}(\text{DATESYTD}(\text{'Dim Fechas'}[\text{ Date Orden}]), \text{'Dim Fechas'}[\text{ Date Orden}] \leq \text{TODAY}()), -1, \text{YEAR}))$
- Total de Ventas Año Anterior = $\text{CALCULATE}(\text{[Ventas totales]}, \text{DATEADD}(\text{FILTER}(\text{DATESYTD}(\text{'Dim Fechas'}[\text{ Date Orden}]), \text{'Dim Fechas'}[\text{ Date Orden}] \leq \text{TODAY}()), -1, \text{YEAR}))$
- Última Actualización = $\text{FORMAT}(\text{NOW}(), \text{"dd/mm/yyyy hh:mm:ss"})$
- Ventas totales = $\text{SUM}(\text{superstore_dataset}[\text{Sales}])$