



Análisis de datos de la cadena de hipermercados norteamericana
“SuperStore”

Autor: Martin Aiscar

Institución: CoderHouse

Fecha de presentación: 01/11/2023

Índice

Descripción de la temática de los datos.....	2
Hipótesis.....	2
Dataset.....	2
Diagrama Entidad Relación.....	3
Listado de tablas.....	4
Listado de columnas por tablas.....	5
Tablero Power BI.....	8
Transformación de Datos.....	9
Tabla de Calendario.....	9
Tablas de Medidas.....	9
Medidas Ventas.....	10
Medidas Costos.....	10
Medidas Utilidad.....	11
Medidas Categorías.....	11
Medidas Segmentos.....	12
Visualización de Datos.....	12
Página: “Inicio”	12
Página: “Datos Iniciales”	13
Página: “Ventas”	13
Página: “Segmentación”	14

SuperStore - Proyecto Final

Descripción de la temática de los datos

Para el presente escrito, se extrajo la información de las ventas realizadas entre enero de 2014 y diciembre de 2017, de una cadena de hipermercados de los Estados Unidos.

Dentro de la información recolectada, se puede observar datos relacionados con el origen de las ventas, las categorías y subcategorías de los productos elegidos por los clientes, las cantidades vendidas por cada producto, y el tipo de cliente.

Hipótesis

Este estudio tiene como objetivo comprender qué productos, regiones, categorías o tipos de clientes debe enfocarse la empresa para poder desarrollar nuevas campañas de marketing, como también entender el nivel de ventas y la utilidad generada por las mismas.

Por un lado, para lograr la segmentación del mercado, se verán aquellos estados, ciudades y los productos más demandados que generan el mayor nivel de ventas en estas zonas.

Por otro lado, para visualizar el nivel de ventas de la compañía, se utilizará los datos numéricos ofrecidos por el dataset entregado por la misma.

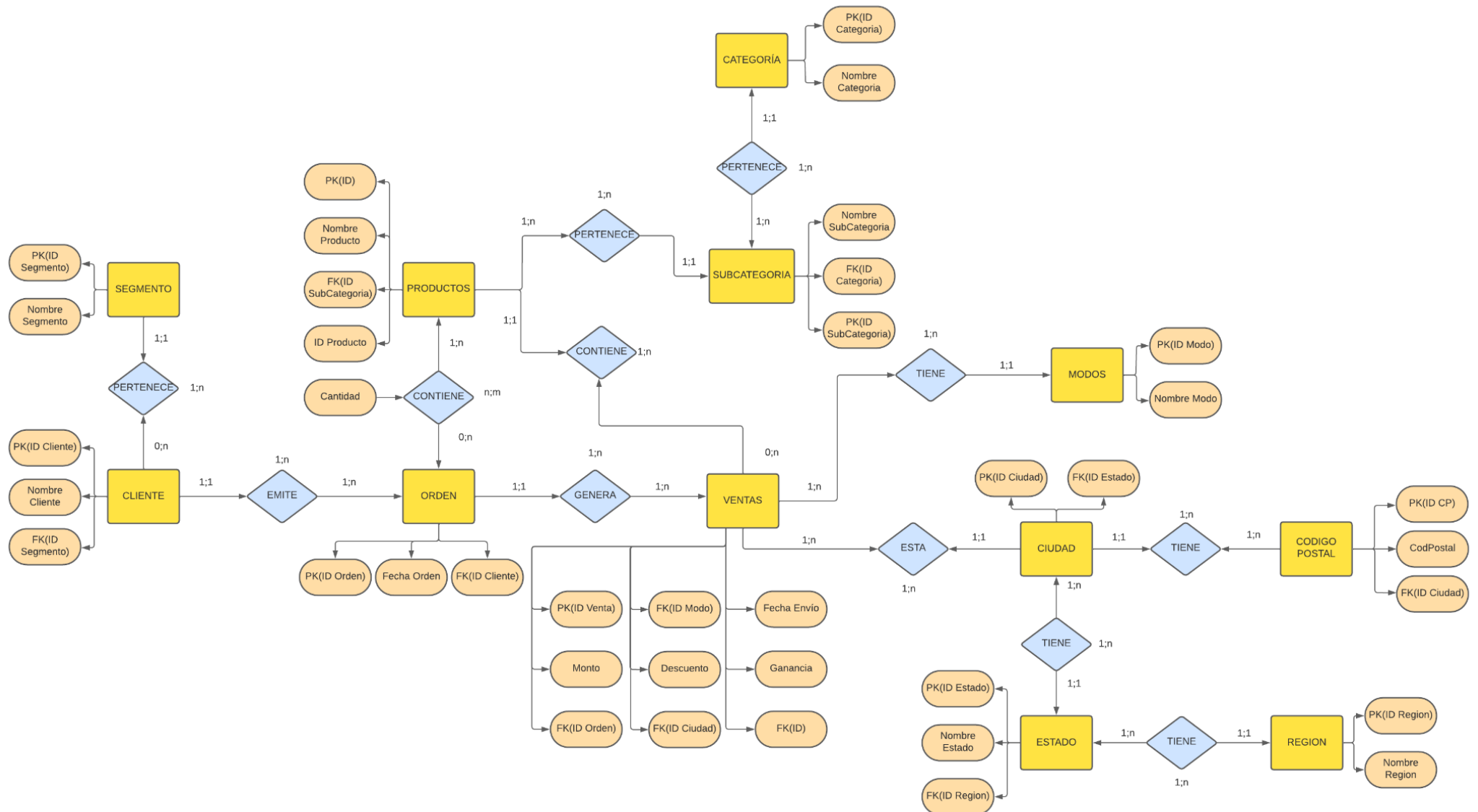
Dataset

El dataset mencionado posee el registro de 10.000 ventas generadas entre 2014 y 2017, donde las mismas están caracterizadas mediante 21 atributos.

- **Row ID:** identificador de venta
- **Order ID:** identificador de la orden de compra generada por el cliente
- **Order Date:** fecha de la orden generada
- **Fecha de Envío:** fecha en la que se envía el producto al cliente
- **Modo de Envío:** forma en la que se realizar el envío al cliente
- **Customer ID:** identificador del cliente
- **Customer Name:** nombre del cliente
- **Segment:** tipo de cliente
- **Country:** país donde se realizó la venta
- **City:** ciudad donde se realizó la venta
- **State:** estado donde se realizó la venta
- **Postal Code:** código postal de la ciudad
- **Región:** región donde se realizó la venta
- **Producto ID:** identificador del producto
- **Product Name:** nombre del producto
- **Category:** categoría del producto
- **Sub-Category:** subcategoría a la que pertenece el producto
- **Sales:** monto total generado por la venta
- **Quantity:** cantidad de producto pedido
- **Discount:** descuento otorgado por la compañía
- **Profit:** utilidad neta de la operación (se restan costos operativos y financieros)

Diagrama Entidad Relación

A partir de los atributos otorgados por el dataset original, se desarrolló el siguiente diagrama entidad relación, el cual servirá como base para el diseño de la estructura de la base de datos.



Listado de tablas

A continuación, se detallará cada una de las tablas que compondrán la base de datos de la compañía, con sus respectivas claves principales, claves foráneas y sus diversos atributos.

- **Segmento:** contiene todos los tipos de clientes que se pueden presentar dentro de la base de datos.
 - **ID Segmento:** este identificador será la **clave primaria** de la tabla.
 - **Nombre Segmento**
- **Cliente:** contiene todos los clientes de la compañía que realizaron ventas.
 - **ID Cliente:** este identificador será la **clave primaria** de la tabla.
 - **ID Segmento:** este identificador será la **clave foránea** que hará referencia a la tabla Segmento.
 - **Nombre Cliente**
- **Orden:** contiene los datos más importantes de todas las orden emitidas por los clientes junto con su fecha de emisión
 - **ID Orden:** este identificador será la **clave primaria** de la tabla.
 - **ID Cliente:** este identificador será la **clave foránea** que hará referencia a la tabla Cliente.
 - **Fecha Orden:** aquellas órdenes emitidas entre el año 2014 y 2017.
- **Categoría:** contiene todas las categorías de productos existentes.
 - **ID Categoría:** este identificador será la **clave primaria** de la tabla.
 - **Nombre Categoría**
- **Subcategoría:** contiene todas las subcategorías de productos existentes, las cuales cada una de ellas están relacionadas a una categoría.
 - **ID SubCategoría:** este identificador será la **clave primaria** de la tabla.
 - **ID Categoría:** este identificador será la **clave foránea** que hará referencia a la tabla Categoría
 - **Nombre Categoría**
- **Productos:** en esta tabla se detalla cada uno de los productos vendidos por la compañía entre el año 2014 y 2017. Cabe aclarar que se debe agregar un campo adicional a los ofrecidos por el dataset, ya que existen productos diferentes que tienen el mismo código de producto.
 - **ID:** este identificador será la **clave primaria** de la tabla.
 - **ID Producto:** hará referencia al código de producto interno.
 - **ID SubCategoría:** este identificador será la **clave foránea** que hará referencia a la tabla Subcategoría.
 - **Nombre Producto**
- **Producto Orden:** para salvar la relación muchos a muchos existente entre la tabla Productos y Orden, se debe crear una tabla adicional para poder guardar los datos en la base de datos de manera correcta.
 - **ID Producto:** este identificador será la **clave foránea** que hará referencia a la tabla Producto.
 - **ID Orden:** este identificador será la **clave foránea** que hará referencia a la tabla Orden
 - **Cantidad:** cantidad de productos pedidos.

Tanto la clave foránea ID Producto como la ID Orden, forman de manera concatenada la clave principal de la tabla Producto Orden.

- **Modos:** en esta tabla se detalla la formas o modos en la que se realiza una venta.
 - **ID Modo:** este identificador será la **clave primaria** de la tabla.
 - **Nombre Modo**
- **Región:** contiene todas las regiones de estados unidos donde se realizaron ventas
 - **ID Región:** este identificador será la **clave primaria** de la tabla.
 - **Nombre Región**
- **Estado:** en esta tabla se detalla todos los estados donde se realizaron ventas.
 - **ID Estado:** este identificador será la **clave primaria** de la tabla.
 - **ID Región:** este identificador será la **clave foránea** que hará referencia a la tabla Región
 - **Nombre Estado**
- **Ciudad:** en esta tabla se detalla todas las ciudades donde se realizaron ventas
 - **ID Ciudad:** este identificador será la **clave primaria** de la tabla.
 - **ID Estado:** este identificador será la **clave foránea** que hará referencia a la tabla Estado
 - **Nombre Ciudad**
- **Código Postal:** contiene todos los códigos postales de cada una de las ciudades donde se realizaron ventas. Se debe tener en cuenta que en una ciudad puede haber más de un código postal, sin embargo se optó por agregar un atributo adicional como clave primaria para salvar el caso en donde existan más de un código postal igual para cada ciudad.
 - **ID CP:** este identificador será la **clave primaria** de la tabla.
 - **ID Ciudad:** este identificador será la **clave foránea** que hará referencia a la tabla Ciudad
 - **CodPostal:** código postal numérico.
- **Ventas:** contiene cada una de las ventas realizadas en el periodo de análisis, con la información del monto obtenido, descuentos ofrecidos y la utilidad neta por cada operación realizada.
 - **ID Venta:** este identificador será la **clave primaria** de la tabla.
 - **ID Modo:** este identificador será la **clave foránea** que hará referencia a la tabla Modo.
 - **ID Orden:** este identificador será la **clave foránea** que hará referencia a la tabla Orden.
 - **ID Ciudad:** este identificador será la **clave foránea** que hará referencia a la tabla Ciudad.
 - **ID:** este identificador será la **clave foránea** que hará referencia a la tabla Producto.
 - **Fecha Envío:** fecha en la que se realizó el envío de la venta realizada.
 - **Monto:** monto obtenido por la compañía debido a la operación realizada.
 - **Descuento:** descuento otorgado por la empresa.
 - **Ganancia:** utilidad neta de la venta.

Listado de columnas por tablas

Una vez definidas las tablas con su claves y atributos, se especificará el tipo de variable que tendrá cada columna.

Segmento

Campo	Tipo de variable	Tipo de clave
ID Segmento	int	PK (auto incremental)
Nombre Segmento	varchar(50)	

Cliente

Campo	Tipo de variable	Tipo de clave
ID Cliente	varchar (50)	Primary Key
ID Segmento	int	Foreign Key
Nombre Cliente	varchar (50)	

Orden

Campo	Tipo de variable	Tipo de clave
ID Orden	varchar(50)	Primary Key
ID Cliente	varchar(50)	Foreign Key
Fecha Orden	date	

Categoría

Campo	Tipo de variable	Tipo de clave
ID Categoría	int	PK (auto incremental)
Nombre Categoría	varchar(50)	

Subcategoría

Campo	Tipo de variable	Tipo de clave
ID SubCategoría	int	PK (auto incremental)
ID Categoría	int	Foreign Key
Nombre Categoría	varchar(50)	

Productos

Campo	Tipo de variable	Tipo de clave
ID	int	Primary Key
ID SubCategoría	int	Foreign Key
ID Producto	varchar(50)	
Nombre Producto	varchar(50)	

Producto Orden

Campo	Tipo de variable	Tipo de clave
ID Producto	int	Foreign Key (CK)
ID Orden	varchar(50)	Foreing Key (CK)
Cantidad	int	

Modos

Campo	Tipo de variable	Tipo de clave
ID Modo	int	Primary Key
Nombre Modo	varchar(50)	

Región

Campo	Tipo de variable	Tipo de clave
ID Region	int	Primary Key
Nombre Región	varchar (50)	

Estado

Campo	Tipo de variable	Tipo de clave
ID Estado	int	Primary Key
ID Region	int	Foreign Key
Nombre Estado	varchar(50)	

Ciudad

Campo	Tipo de variable	Tipo de clave
ID Ciudad	int	Primary Key
ID Estado	int	Foreign Key
Nombre Ciudad	varchar(50)	

Código Postal

Campo	Tipo de variable	Tipo de clave
ID CP	int	Primary Key
ID Ciudad	int	Foreign Key
CodPostal	int	

Ventas

Campo	Tipo de variable	Tipo de clave
ID Venta	int	Primary Key
ID Modo	int	Foreign Key
ID Orden	varchar(50)	Foreign Key
ID Ciudad	int	Foreign Key
ID Producto	int	Foreign Key
Monto	decimal(18,4)	
Descuento	decimal(18,4)	
Ganancia	decimal(18,4)	
Fecha Envío	date	

Tablero Power BI

Mediante la herramienta Power BI de Microsoft, se realizó un tablero donde se plasmaron los datos necesarios para responder a las hipótesis planteadas al inicio de este escrito.

En primer lugar, en este informe, se detallará todas las segmentaciones y transformaciones de datos realizadas, y las medidas calculadas para desarrollar cada una de las páginas del

tablero de visualización. En segundo lugar, se desarrollará cada una de las páginas del dashboard y cada uno de los datos necesarios.

Transformación de Datos

En este caso de estudio, la única transformación de datos que se realizó fue el agregado de una nueva columna en la Tabla "Ventas" llamada "Costo_Financiero", donde en cada registro se calcula el descuento otorgado por la empresa para todas las ventas realizadas por todos los clientes. Esta columna se calcula como: "Monto x Descuento"

Tabla de Calendario

El objetivo de esta tabla es poder realizar las segmentaciones de datos tanto por año como por mes en el tablero. También, sirve para realizar gráficos más claros para mostrar la evolución de determinadas variables a lo largo del período de análisis (2014-2017).

Las columnas de esta tabla son las siguientes:

- **Date:** se muestra la fecha completa
- **Año**
- **Mes:** mes expresado en números.
- **Mes_Letras:** mes expresado en letras.
- **Día:** día expresado en números
- **Día_Letras:** día expresado en letras
- **Trimestre:** muestra el número de trimestre del año.

Cada uno de estas columnas se generaron mediante las siguientes líneas de código:

- **Date:** "Calendario = CALENDARAUTO()"
- **Año:** "AÑO = YEAR(Calendario[Date])"
- **Mes:** "MES = MONTH(Calendario[Date])"
- **Mes_Letras:** "MES_LETRAS = FORMAT(Calendario[Date], "MMMM")"
- **Día:** "DIA = DAY(Calendario[Date])"
- **Día_Letras:** "DIA_LETRAS = FORMAT(Calendario[Date], "DDDD")"
- **Trimestre:** "TRIMESTRE = ROUNDUP(Calendario[MES]/4,0)"

Tablas de Medidas

Para generar las visualizaciones de datos necesarias y tener de manera ordenada cada una de las medidas, se generaron diferentes tablas de medidas en función de qué entidad o variable se quiso medir.

Las tablas de medidas generadas fueron:

- **Medidas Ventas**
- **Medidas Costos**
- **Medidas Utilidad**
- **Medidas Categorías**
- **Medidas Segmentos**

Medidas Ventas

Dentro de la tabla de medidas de Ventas, se crearon las siguientes medidas:

- **VCategoria:** se muestra las ventas totales de la categoría más vendida
- **Ventas LY:** muestra las ventas totales del año anterior
- **VSegmento:** muestra las ventas totales realizadas por el tipo de cliente que más ventas realizó
- **VTotales:** muestra las ventas totales del periodo de análisis
- **VVariación %:** variación de ventas respecto del año anterior

Cada una de estas medidas se calcularon de la siguiente manera:

- **VCategoria:** `"VCategoria = if([Furniture]>[Office_Supplies] && [Furniture]>[Technology], [Furniture], if([Office_Supplies]>[Furniture] && [Office_Supplies]>[Technology], [Office_Supplies], if([Technology]>[Office_Supplies] && [Technology]>[Furniture], [Technology]))"`
- **VentasLY:** `"VentasLY=CALCULATE([VTotales],SAMEPERIODLASTYEAR(Calendario [Date]))"`
- **VSegmento:** `"VSegmento = IF([Consumer]>[Corporate] && [Consumer]>[Home_Office], [Consumer], IF([Corporate]>[Consumer] && [Corporate]>[Home_Office], [Corporate], IF([Home_Office]>[Corporate] && [Home_Office]>[Consumer], [Home_Office]))"`
- **VTotales:** `"VTotales = SUM(Ventas[Monto])"`
- **VVariación%:** `"VVariacion%=DIVIDE(MedidasVentas[VTotales]-MedidasVentas[Ventas LY],MedidasVentas[Ventas LY],BLANK())"`

Medidas Costos

Dentro de la tabla de medidas de Costos, se crearon las siguientes medidas:

- **CTotales:** calcula el costo total de la operaciones realizadas en el periodo de análisis
- **Costo LY:** calcula el costo total del año anterior
- **CVariacion%:** calcula la variación del costo total respecto del año anterior
- **CFinanciero:** muestra el costo total financiero que incurrió la empresa por otorgar descuentos a sus clientes.
- **CFinanciero%:** muestra la proporción del costo financiero respecto del costo total
- **COperativo:** muestra el costo operativo total de la empresa, el cual se calcula deduciendo el costo financiero al costo total.

Para calcular cada una de estas medidas, se desarrollaron las siguientes líneas de código:

- **CTotales:** `"CTotales = SUM('Ventas'[Monto]) - SUM('Ventas'[Ganancia])"`
- **CostoLY:** `"CostoLY=CALCULATE([CTotales],SAMEPERIODLASTYEAR(Calendario [Date]))"`
- **CVariacion%:** `"CVariacion % = DIVIDE([CTotales]-MedidasCostos[Costo LY],MedidasCostos[Costo LY],BLANK())"`
- **CFinanciero:** `"CFinanciero = SUM(Ventas[Costo_Financiero])"`
- **CFinanciero%:** `"CFinanciero % = CALCULATE([CFinanciero]/[CTotales])"`

- **COperativo:** “COperativo = [CTotales]-[CFinanciero]”

Medidas Utilidad

Dentro de la tabla de medidas de Utilidad, se crearon las siguientes medidas:

- **UTotal:** calcula la utilidad total de la empresa en el periodo de análisis.
- **Utilidad LY:** muestra la utilidad del año anterior
- **Utilidad %:** muestra la proporción de la utilidad respecto de las ventas totales
- **UVariacion%:** calcula la variación de la utilidad respecto del año anterior

Las medidas antes mencionadas se calcularon de la siguiente manera:

- **UTotal:** “UTotal = sum(Ventas[Ganancia])”
- **UtilidadLY:** “UtilidadLY=CALCULATE(SUM(Ventas[Ganancia]),SAMEPERIODLASTYEAR(Calendario[Date]))”
- **Utilidad %:** “Utilidad % = CALCULATE(sum(Ventas[Ganancia])/[VTotales])”
- **UVariacion%:** “UVariacion % = DIVIDE([UTotal]-MedidasUtilidad[Utilidad LY],MedidasUtilidad[Utilidad LY],BLANK())”

Medidas Categorías

Dentro de la tabla de medidas de Categorías, se desarrollaron las siguientes medidas:

- **Furniture:** calcula la cantidad de ventas totales de la categoría de producto “Furniture”
- **Office_Supplies:** calcula la cantidad de ventas totales de la categoría de producto “Office_Supplies”
- **Technology:** calcula la cantidad de ventas totales de la categoría de producto “Technology”
- **Categoría:** muestra la categoría más vendida en el periodo de análisis
- **ColorPreferencia:** esta muestra será utilizada para colorear un determinado estado en un mapa, y en función del color, mostrará la categoría más vendida en ese estado.

Cada uno de estas medidas se generaron mediante las siguientes líneas de código:

- **Furniture:** “Furniture=CALCULATE([VTotales],Categoria[Nombre_Categoria]="Furniture")”
- **Office_Supplies:** “Office_Supplies=CALCULATE([VTotales],Categoria[Nombre_Categoria]="Office Supplies")”
- **Technology:** “Technology=CALCULATE([VTotales],Categoria[Nombre_Categoria]="Technology")”
- **Categoría:** “Categoria = IF([Furniture]>[Office_Supplies] && [Furniture]>[Technology], "Furniture", IF([Office_Supplies]>[Furniture] && [Office_Supplies]>[Technology], "Office Supplies", IF([Technology]>[Office_Supplies] && [Technology]>[Furniture], "Technology"))”
- **ColorPreferencia:** “ColorPreferencia = if([Furniture]>[Office_Supplies] && [Furniture]>[Technology], "#118DFF", if([Office_Supplies]>[Furniture] && [Office_Supplies]>[Technology] , "#12239E", if([Technology]>[Office_Supplies] && [Technology]>[Furniture], "#E66C37"))”

Medidas Segmentos

Dentro de la tabla de medidas de Segmentos, se desarrollaron las siguientes medidas:

- **Corporate:** calcula la cantidad de ventas totales por tipo de cliente "Corporate"
- **Home_Office:** calcula la cantidad de ventas totales por tipo de cliente "Home Office"
- **Consumer:** calcula la cantidad de ventas totales por tipo de cliente "Consumer"
- **Segmento:** muestra el tipo de cliente que más ventas realizó en el periodo de análisis.

Cada uno de estas medidas se generaron mediante las siguientes líneas de código:

- **Corporate:** "Corporate=CALCULATE([VTotales],Segmento[Nombre_Segmento]="Corporate")"
- **Home_Office:** "Home_Office=CALCULATE([VTotales],Segmento[Nombre_Segmento]="Home Office")"
- **Consumer:** "Consumer=CALCULATE([VTotales],Segmento[Nombre_Segmento]="Consumer")"
- **Segmento:** "Segmento=IF([Consumer]>[Corporate]&&[Consumer]>[Home_Office],"Consumer",IF([Corporate]>[Consumer]&&[Corporate]>[Home_Office],"Corporate",IF([Home_Office]>[Corporate]&&[Home_Office]>[Consumer],"Home Office")))"

Visualización de Datos

A continuación, se hará una breve descripción de cada una de las páginas del tablero de datos desarrollado.

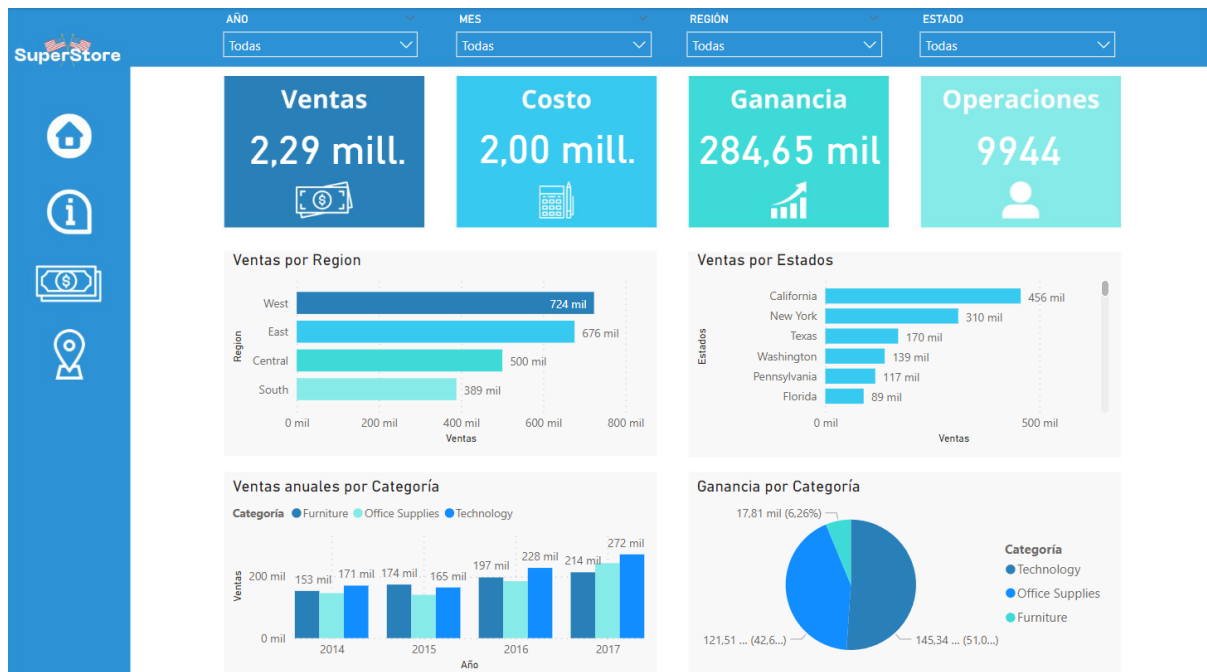
Página: "Inicio"

En esta página, se muestra el logo de la compañía y los botones para acceder a las diferentes páginas:



Página: “Datos Iniciales”

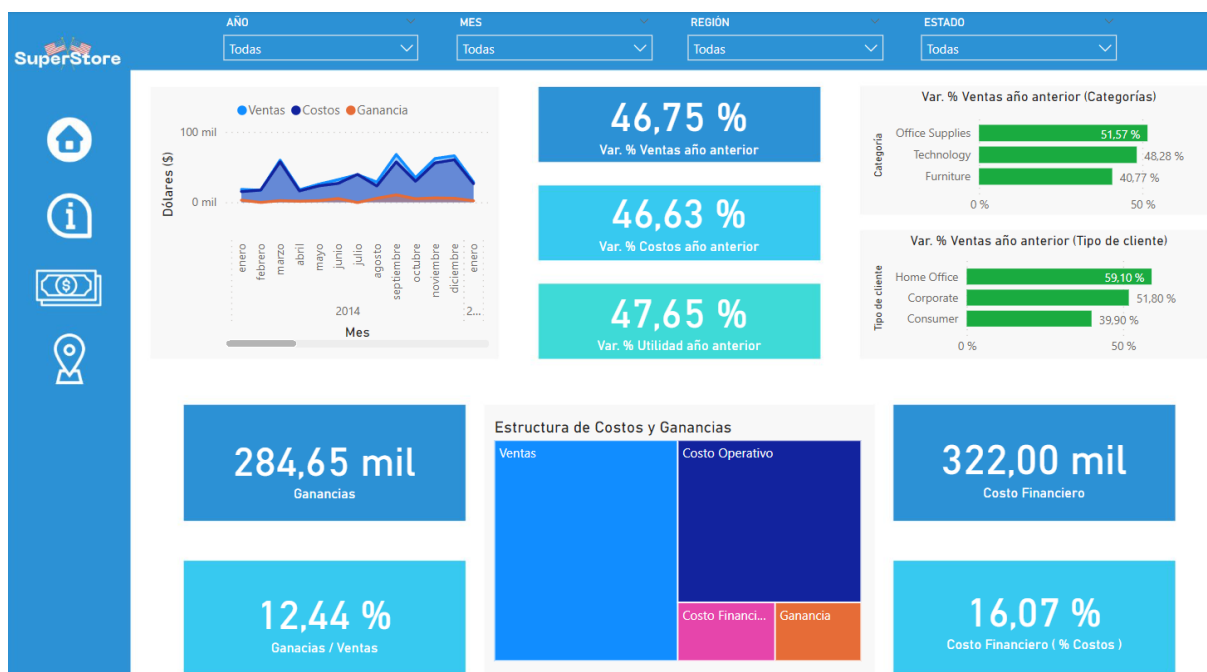
En la página de datos iniciales, se muestran datos fundamentales para realizar un análisis superficial de la empresa, donde se pueda segmentar por año, mes, región y estado.



Página: “Ventas”

Esta página responde a la segunda hipótesis planteada al principio de este escrito: “...para visualizar el nivel de ventas de la compañía, se utilizará los datos numéricos ofrecidos por el dataset entregado por la misma.”

Este tablero muestra la evolución de las ventas en el periodo de análisis, la variación de las ventas, costos y utilidad respecto del año anterior, como varían las ventas por tipo de cliente y categoría de producto, y como es la composición visual del costo total de la compañía.



Página: “Segmentación”

Esta página responde a la primera hipótesis planteada al inicio: “...para lograr la segmentación del mercado, se verán aquellos estados, ciudades y los productos más demandados que generan el mayor nivel de ventas en estas zonas.”

En este tablero, se puede observar aquellos productos más vendidos, qué categorías se consumieron más en cada estado, qué proporción de las ventas corresponde a cada tipo de cliente, y se muestra visualmente las categorías más vendidas y el tipo de cliente que más productos compró, con su respectivo nivel de venta.

