

Descripción de la temática de los datos:

En este proyecto se extrajeron datos cualitativos y cuantitativos de una base de datos disponible perteneciente a la organización “DatData”.

La temática se torna en múltiples datos que surgen por las ventas hechas por una compañía multinacional de sus variados productos en todo el mundo.

Los datos abarcan fechas, canales de distribución, zonas geográficas, canales de promoción, categorías y subcategorías de los productos de dichas ventas.

Los datos corresponden al período 1/1/2005 a 31/12/2011.

Objetivo:

Además de conocer cómo fue el panorama general de las cualidades de las ventas del período, también encontrar los lugares geográficos y canales de distribución donde surgieron mayores y menores transacciones para invertir en mejor infraestructura logística.

Hipótesis:

La multinacional ha observado una disminución significativa en las ventas de sus productos en países europeos de larga distancia, comparados con aquellos más cercanos a su planta en Estados Unidos.

Para evaluar si la distancia geográfica impacta directamente en las ventas, se realizará un análisis comparativo entre Polonia (el país europeo más alejado) y Canadá (el más cercano).

La hipótesis nula (H0) plantea que no existe una diferencia significativa en las ventas entre ambos países. Sin embargo, la hipótesis alternativa (H1) sugiere lo contrario.

En caso de rechazar H0, se implementarán estrategias para fortalecer la logística en los destinos con menor desempeño.

Alcance:

Los aspectos que serán incluídos al estudio principal son los datos correspondientes a las tablas: FactSales, DimGeography, DimChannel, DimProduct, DimProductCategory, DimProductSubcategory y DimStores.

Aquellos tablas que no tornan de gran importancia para el análisis son: DimCalendar (solo servirá como referencia de fechas) y DimPromotion.

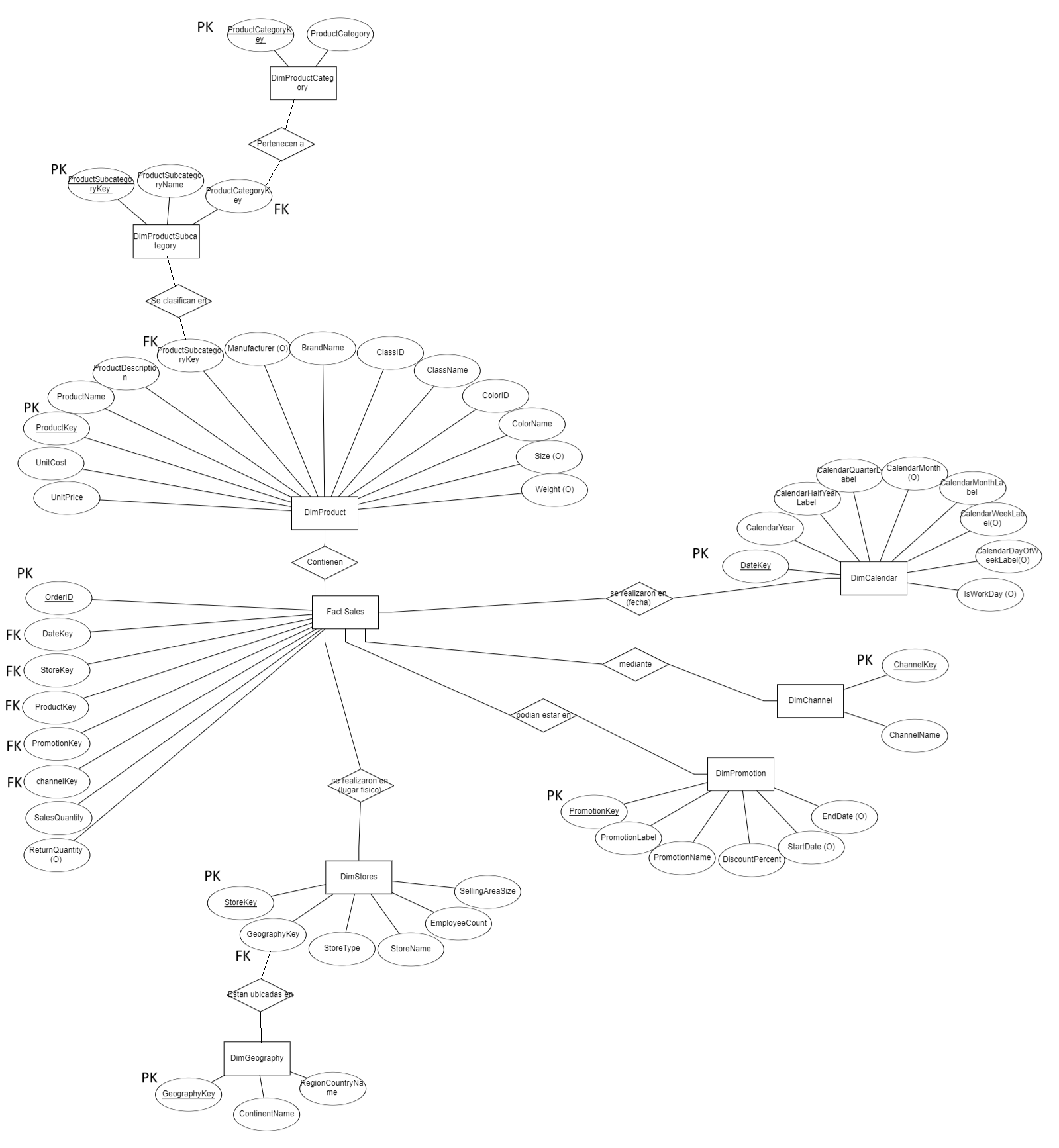
Usuario final y nivel de aplicación por área de organización:

El usuario final será el departamento de logística y supply chain.

Se hará un análisis a nivel táctico ya que el objetivo es guiar al usuario en el proceso de decisión respecto al financiamiento logístico futuro y a mediano plazo.

Diagrama Entidad-Relación:

[Diagrama Entidad-Relación.png](https://drive.google.com/file/d/1WMugycgh8_MTrJ_bVASeLJefPz6uMl87/view?usp=drive_link)



Descripción de tablas

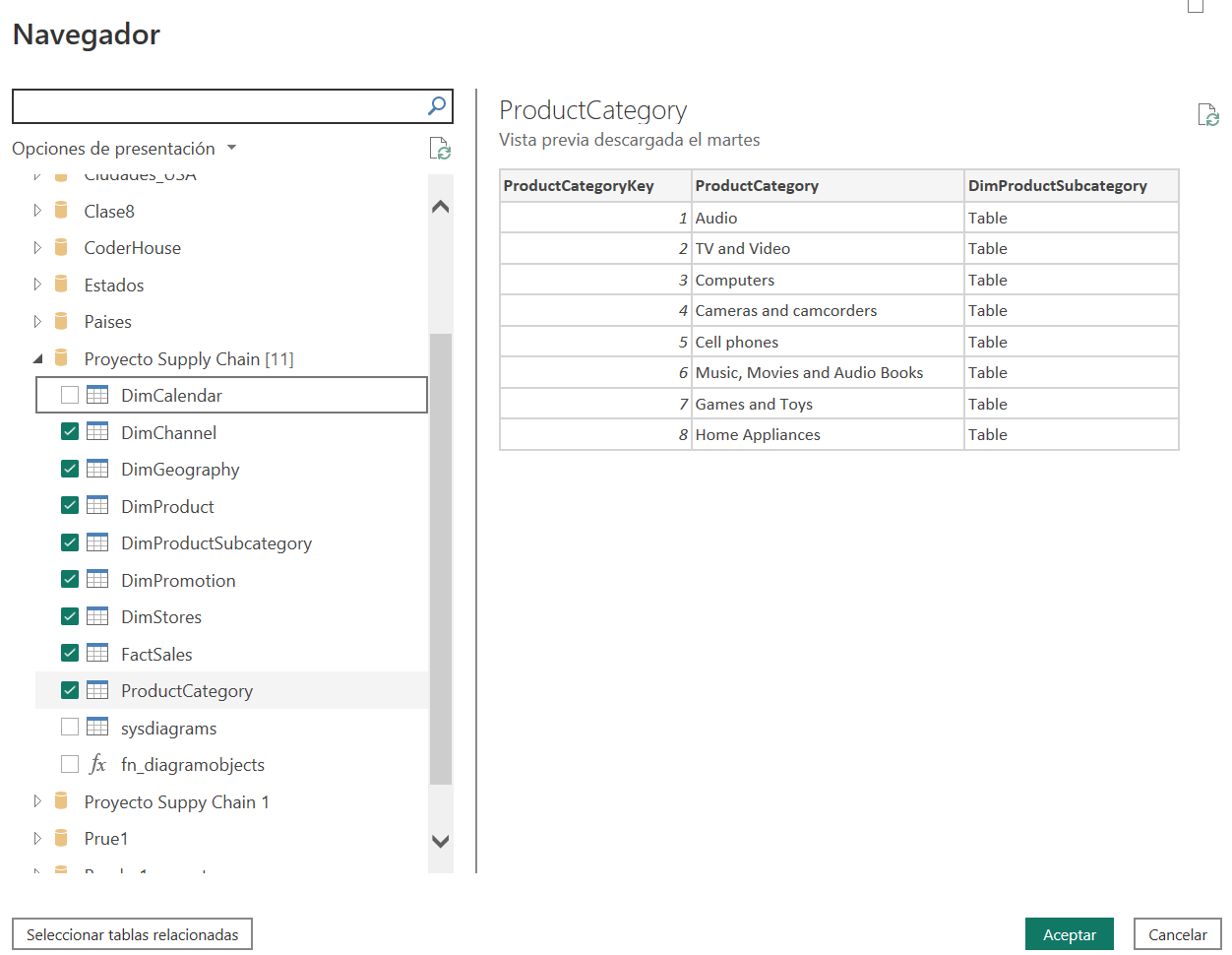
| TABLAS | Descripción | Key |
| --- | --- | --- |
| Tabla FactSales | Contiene cada orden de pedido (venta) que se realizó en el período en sus fechas correspondientes.  Es la tabla de hechos.  Se conecta (FK) con todas las demás tablas dimensiones, menos: Tabla DimProduct, TablaProductCategory y Tabla DimGeography. | OrderID PK  DateKey FK  channelKey FK  StoreKey FK  ProductKey FK  PromotionKey FK |
| Tabla DimProduct | Incluye la descripción específica de cada producto, en cuanto a su calidad, características técnicas y su costo/precio. | ProductKey PK |
| Tabla DimStores | Principalmente describe que tipo de tienda es: física, catálogo, online o revendedor. Y con cuántos empleados cuenta cada uno. También cuenta con una referencia a la ubicación de las tiendas donde nace la relación con la tabla Tabla DimGeography. | StoreKey PK  GeographyKey FK |
| TablaProductCategory | Clasificación de las varias categorías de cada producto único. | ProductCategoryKey PK |
| Tabla DimProductSubcategory | Clasifica los productos en subcategorías únicas.  Incorpora la relación donde nacen las clasificaciones de estas subdivisiones. | ProductSubcategoryKey PK  ProductCategoryKey FK |
| TablaDimCalendar | La columna esencial es la fecha completa en valor numérico, las demás columnas clasifican esta fecha en varios formatos y las subdividen. | DateKey PK |
| Tabla DimPromotion | Se encuentran las campañas de promoción en las que se podían encontrar o no las ventas de los productos. | PromotionKey PK |
| Tabla DimGeography | Describe los lugares donde se hacen las ventas, se clasifican por continente: Asia, Europa, Oceanía y Norteamérica (zona). Además las divisiones en países de esos continentes donde la marca trabaja. | GeographyKey PK |
| Tabla DimChannel | Describe los canales únicos de venta: store  ,online, catalog y reseller. | ChannelKey PK |

Descripción de columnas:

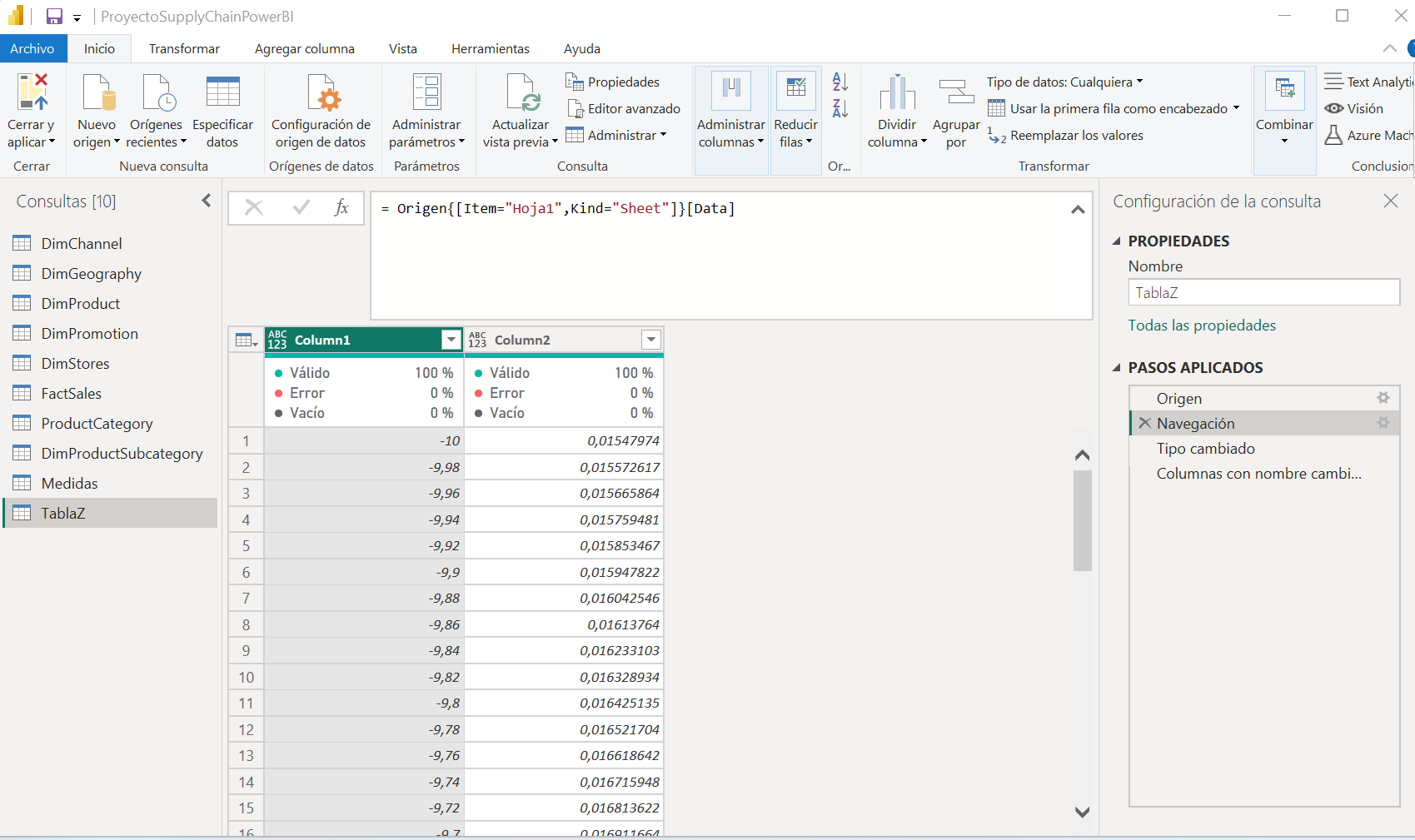
| Tabla FactSales | Tipo de Dato | Tipo de Clave |  | TablaProductCategory | Tipo de Dato | Tipo de Clave |
| --- | --- | --- | --- | --- | --- | --- |
| OrderID | int | PK |  | ProductCategoryKey | int | PK |
| DateKey | datetime | FK |  | ProductCategory | VARCHAR(100) |  |
| channelKey | int | FK |  |  |  |  |
| StoreKey | int | FK |  | Tabla DimProductSubcategory | Tipo de Dato | Tipo de Clave |
| ProductKey | int | FK |  | ProductSubcategoryKey | int | PK |
| PromotionKey | int | FK |  | ProductSubcategoryName | VARCHAR(100) |  |
| SalesQuantity | int |  |  | ProductCategoryKey | int | FK |
| ReturnQuantity | int |  |  |  |  |  |
|  |  |  |  | TablaDimCalendar | Tipo de Dato | Tipo de Clave |
| Tabla DimProduct | Tipo de Dato | Tipo de Clave |  | DateKey | VARCHAR(50) | PK |
| ProductKey | int | PK |  | CalendarYear | VARCHAR(50) |  |
| ProductName | VARCHAR(100) |  |  | CalendarHalfYearLabel | VARCHAR(50) |  |
| ProductDescription | VARCHAR(100) |  |  | CalendarQuarterLabel | VARCHAR(50) |  |
| ProductSubcategoryKey | int | FK |  | CalendarMonth | VARCHAR(50) |  |
| Manufacturer | VARCHAR(50) |  |  | CalendarMonthLabel | VARCHAR(50) |  |
| BrandName | VARCHAR(50) |  |  | CalendarWeekLabel | VARCHAR(50) |  |
| ClassID | int |  |  | CalendarDayOfWeekLabel | VARCHAR(50) |  |
| ClassName | VARCHAR(50) |  |  | IsWorkDay | VARCHAR(50) |  |
| ColorID | int |  |  |  |  |  |
| ColorName | VARCHAR(50) |  |  | Tabla DimPromotion | Tipo de Dato | Tipo de Clave |
| Size | VARCHAR(50) |  |  | PromotionKey | int | PK |
| Weight | float |  |  | PromotionLabel | int |  |
| UnitCost | float |  |  | PromotionName | VARCHAR(100) |  |
| UnitPrice | float |  |  | DiscountPercent | float |  |
|  |  |  |  | StartDate | VARCHAR(50) |  |
| Tabla DimStores | Tipo de Dato | Tipo de Clave |  | EndDate | VARCHAR(50) |  |
| StoreKey | int | PK |  |  |  |  |
| GeographyKey | int | FK |  | Tabla DimGeography | Tipo de Dato | Tipo de Clave |
| StoreType | VARCHAR(50) |  |  | GeographyKey | int | PK |
| StoreName | VARCHAR(100) |  |  | ContinentName | VARCHAR(50) |  |
| EmployeeCount | int |  |  | RegionCountryName | VARCHAR(50) |  |
| SellingAreaSize | int |  |  |  |  |  |
|  |  |  |  | Tabla DimChannel | Tipo de Dato | Tipo de Clave |
|  |  |  |  | ChannelKey | int | PK |
|  |  |  |  | ChannelName | VARCHAR(50) |  |

Importación de base de datos normalizada desde SQL

1. Se realiza la importación de la base de datos principal:



1. Se realiza la importación de una tabla desde excel que servirá como referencia estadística en la visualización de datos posterior:

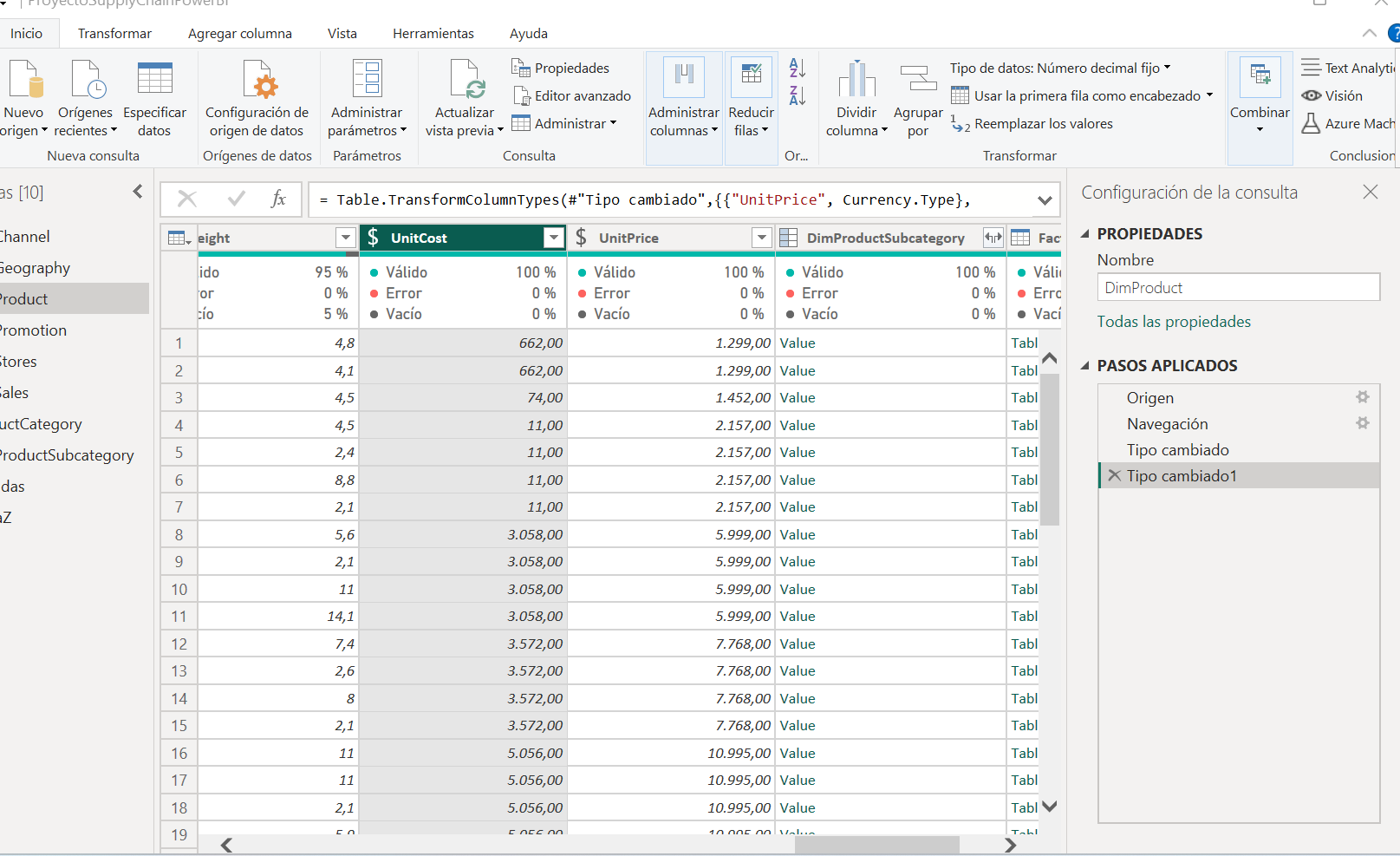


Modelado de datos en Power BI

Transformaciones con Power Query

1. En la tabla Dim Product, la columna de precio de costo y la columna precio de venta se cambia el formato de decimal a decimal fijo

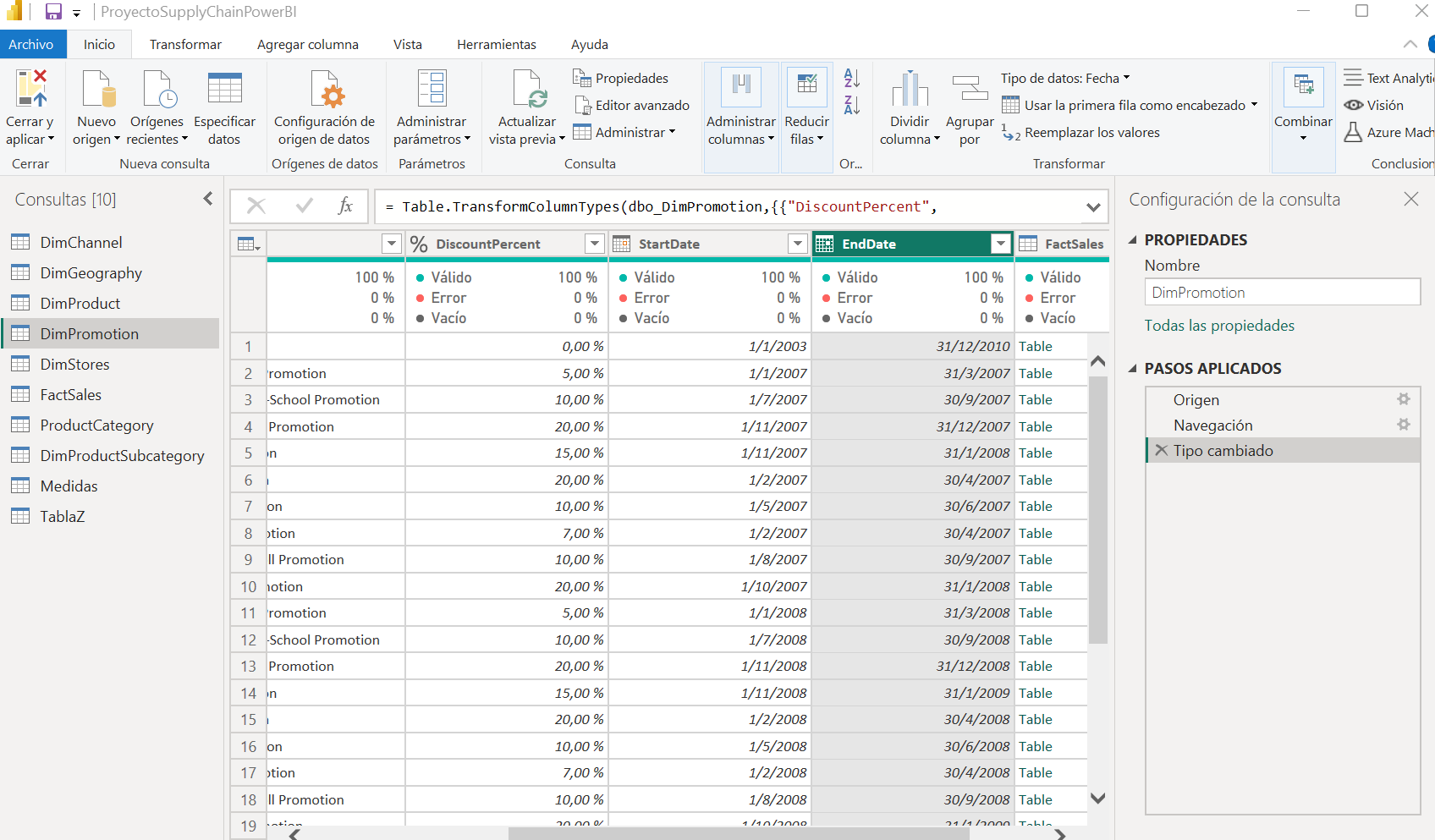
*= Table.TransformColumnTypes(#"Tipo cambiado",{{"UnitPrice", Currency.Type}, {"UnitCost", Currency.Type}})*



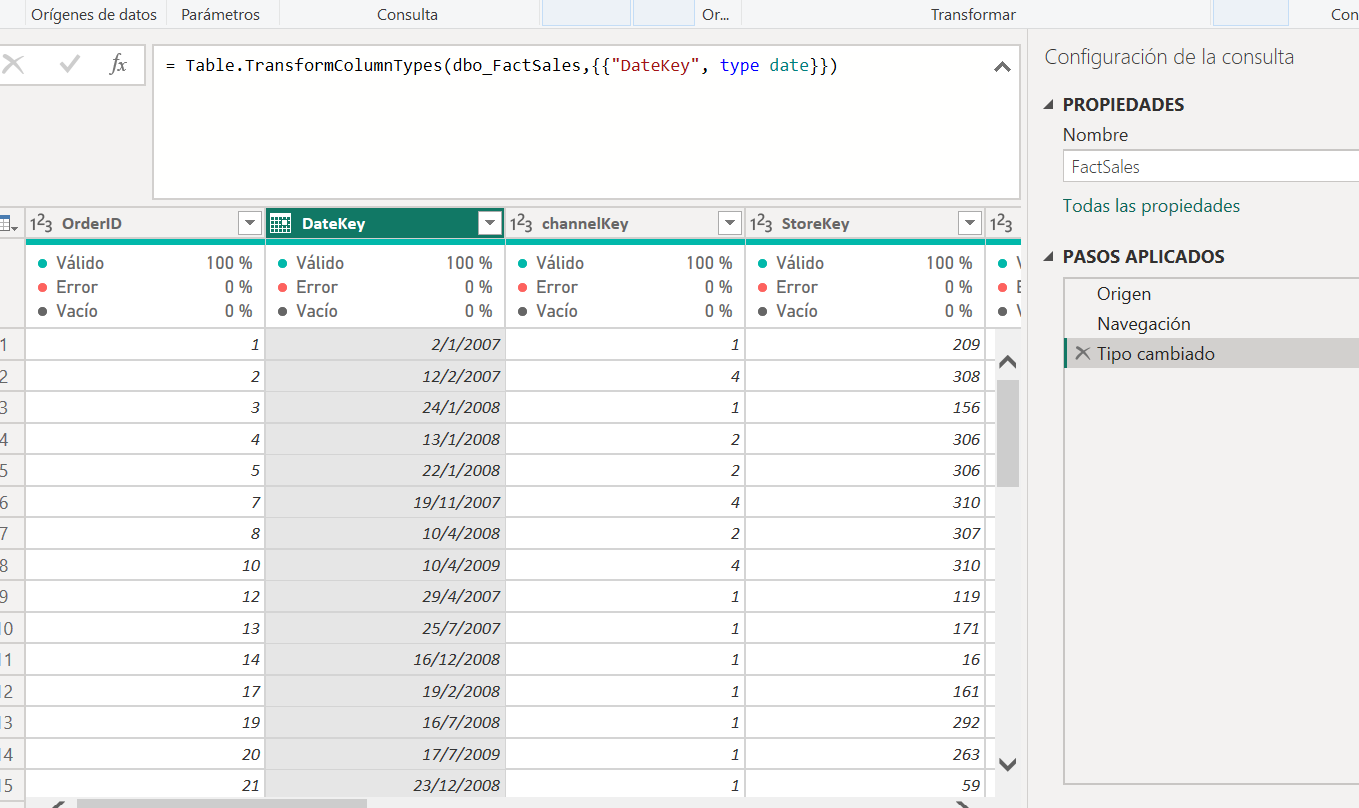
1. En la tabla DimPromotion se cambia el formato de la columna Descuento, de decimal a porcentaje.

Se cambia el formato de fecha y hora a fecha corta en las columnas de start date y end date.

=*Table.TransformColumnTypes(dbo\_DimPromotion,{{"DiscountPercent", Percentage.Type}, {"StartDate", type date}, {"EndDate", type date}})*

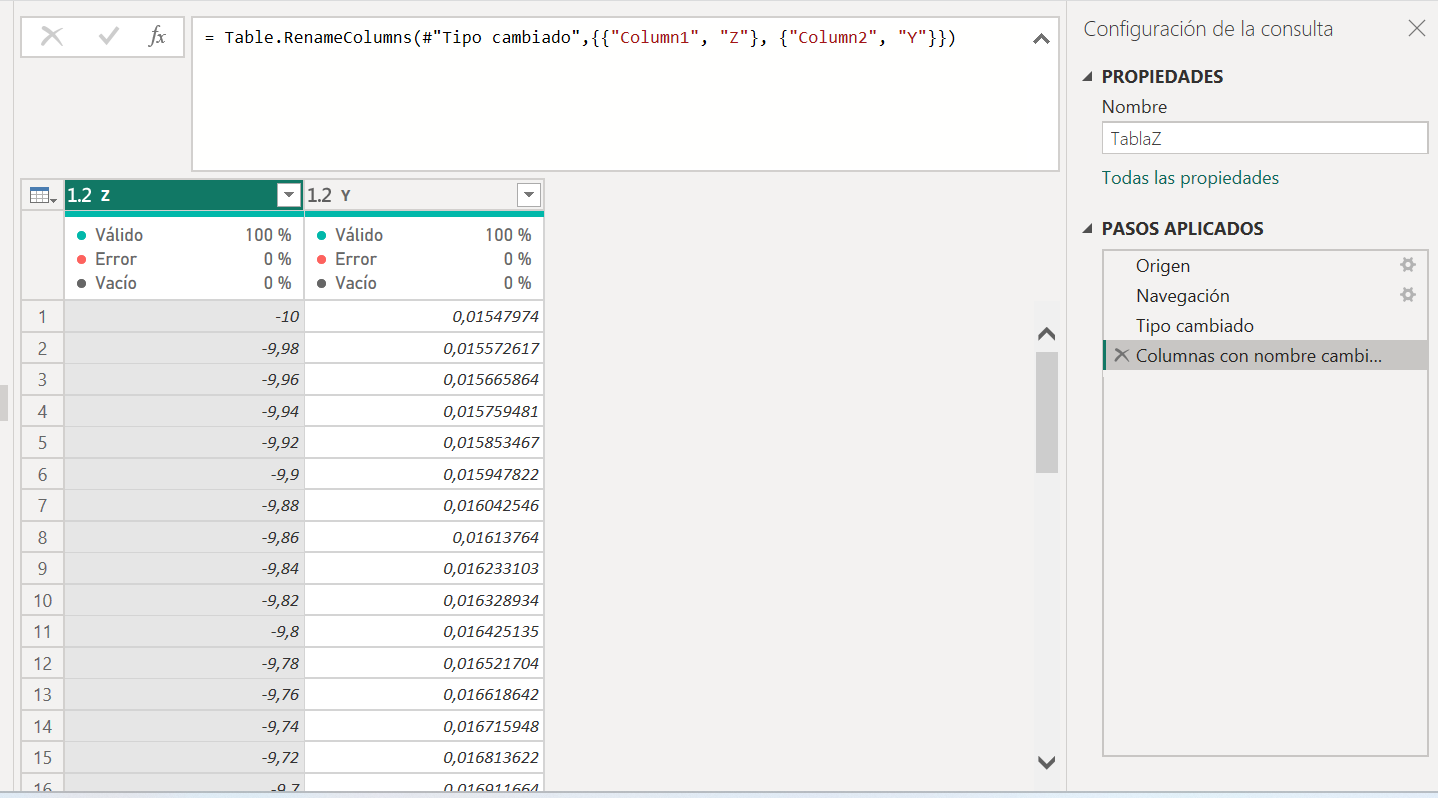


1. En la tabla Fact Sales, se cambia el formato de Date Key, de fecha y hora, a solo fecha corta.

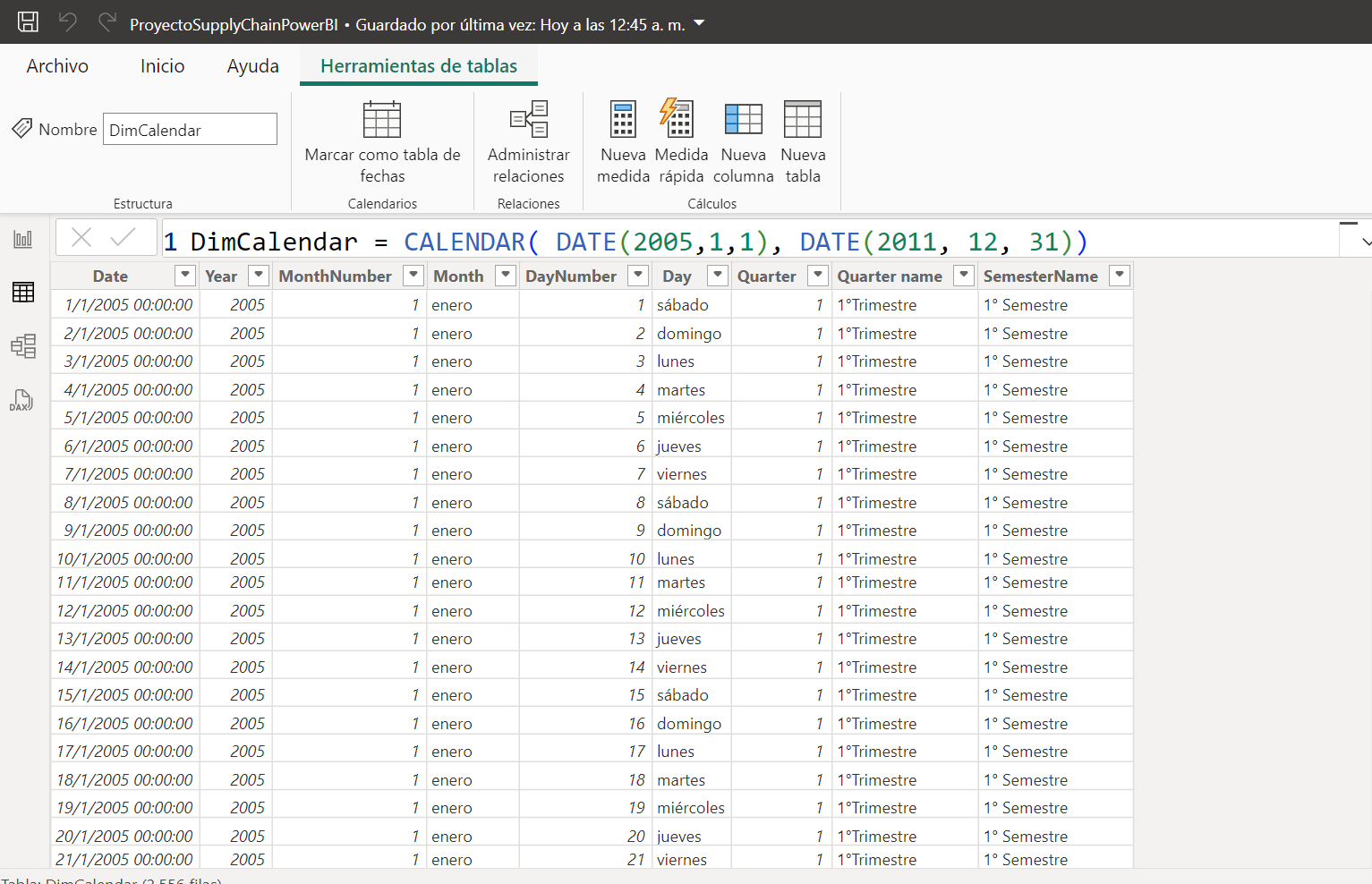


1. En la tabla Z, se nombran las filas de encabezado y se cambian los formatos de ambas columnas de texto a número decimal:

*= Table.TransformColumnTypes(Hoja1\_Sheet,{{"Column1", type number}, {"Column2", type number}})*



Creación de Tabla Calendario



*Columnas de Dim Calendar con código Dax:*

*-Comienzo de operaciones de la empresa: DimCalendar = CALENDAR( DATE(2005,1,1), DATE(2011, 12, 31))*

*-Year = YEAR(DimCalendar[Date])*

*-MonthNumber = MONTH(DimCalendar[Date])*

*-Month = FORMAT(DimCalendar[Date], "MMMM")*

*-DayNumber = DAY(DimCalendar[Date])*

*-Day = FORMAT(DimCalendar[Date], "DDDD")*

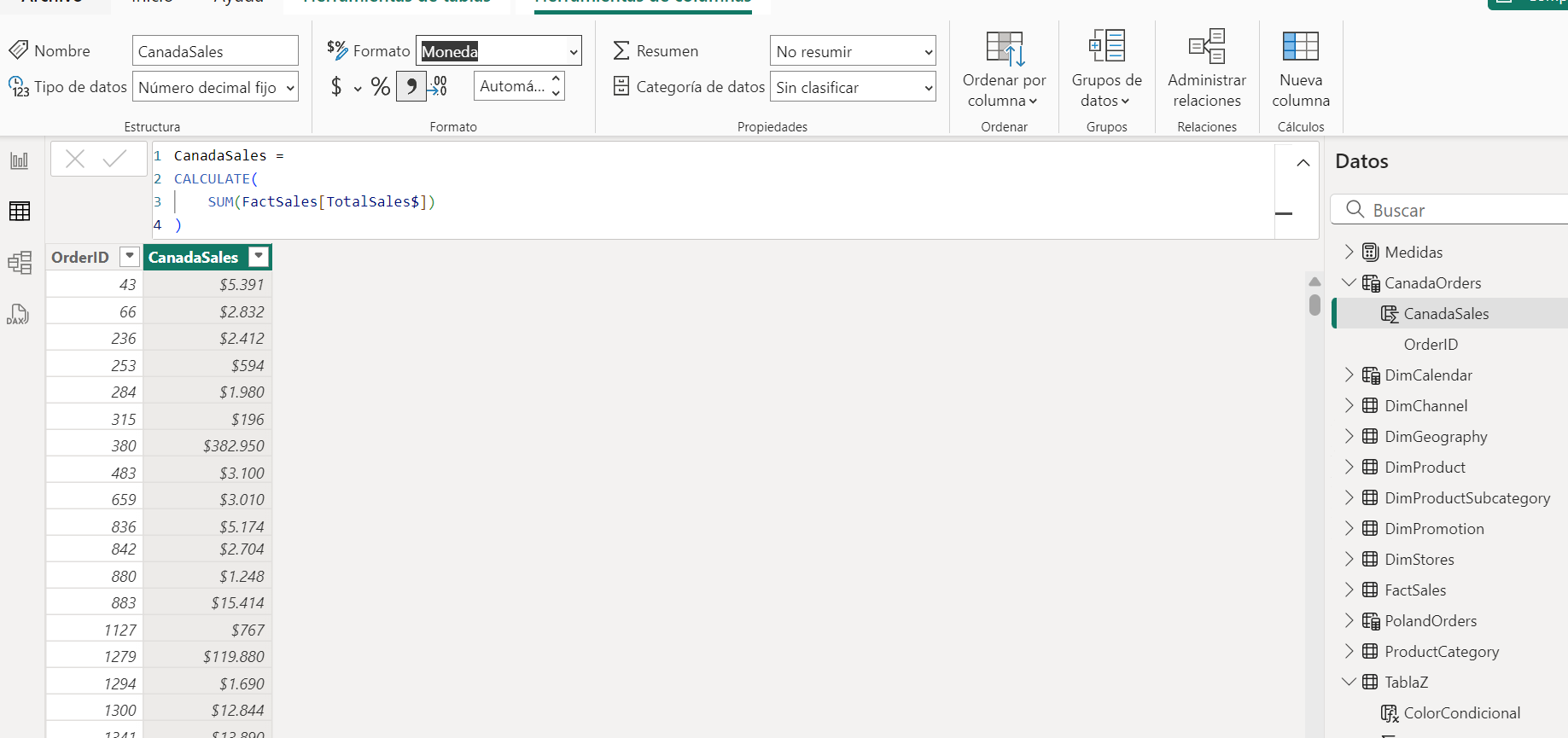
*-Quarter = QUARTER(DimCalendar[Date])*

*-Quarter name = QUARTER(DimCalendar[Date]) & "°Trimestre"*

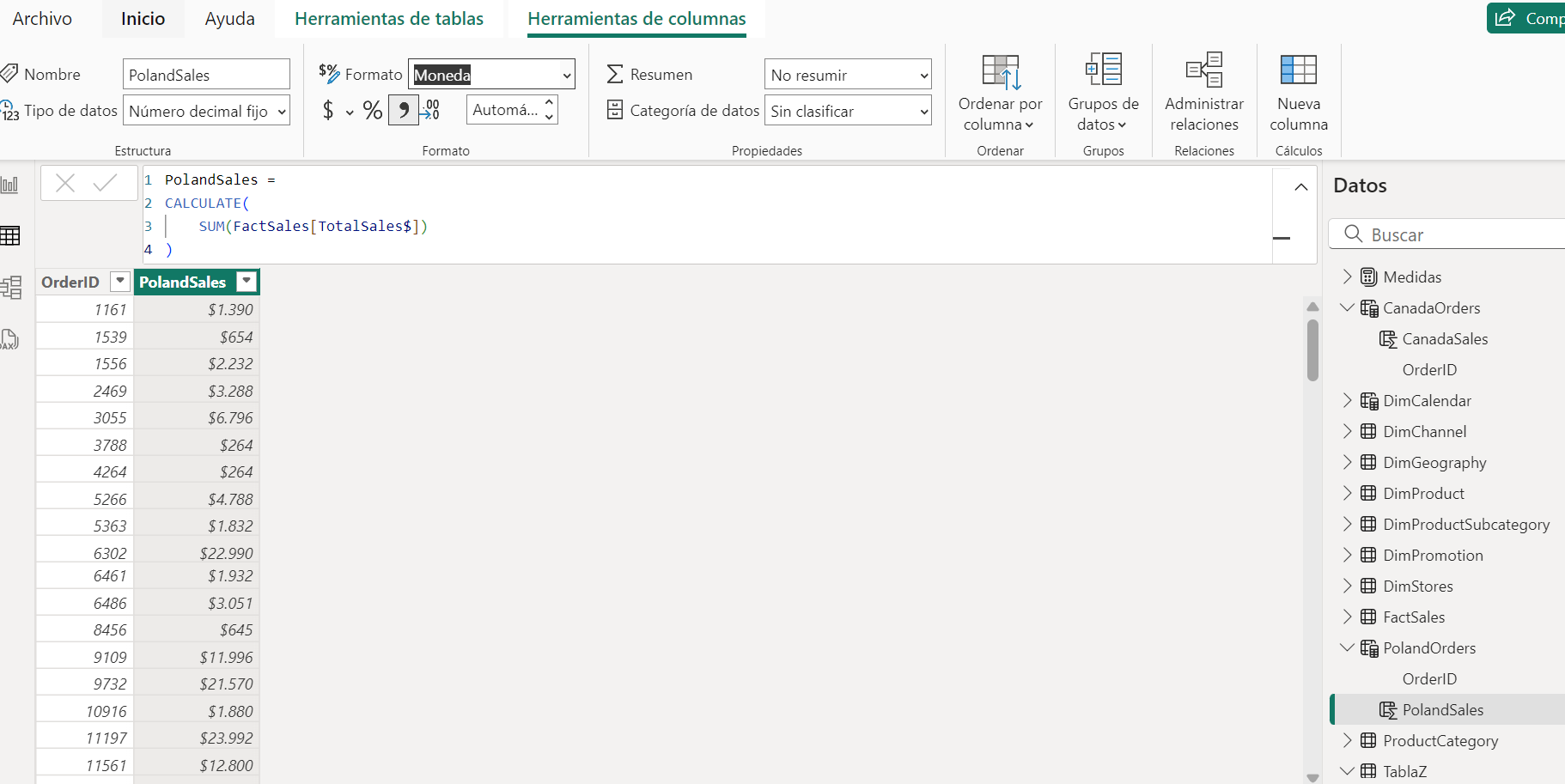
*-SemesterName = IF(DimCalendar[MonthNumber] < 6, "1° Semestre", "2°*

Creación de nuevas tablas con DAX:

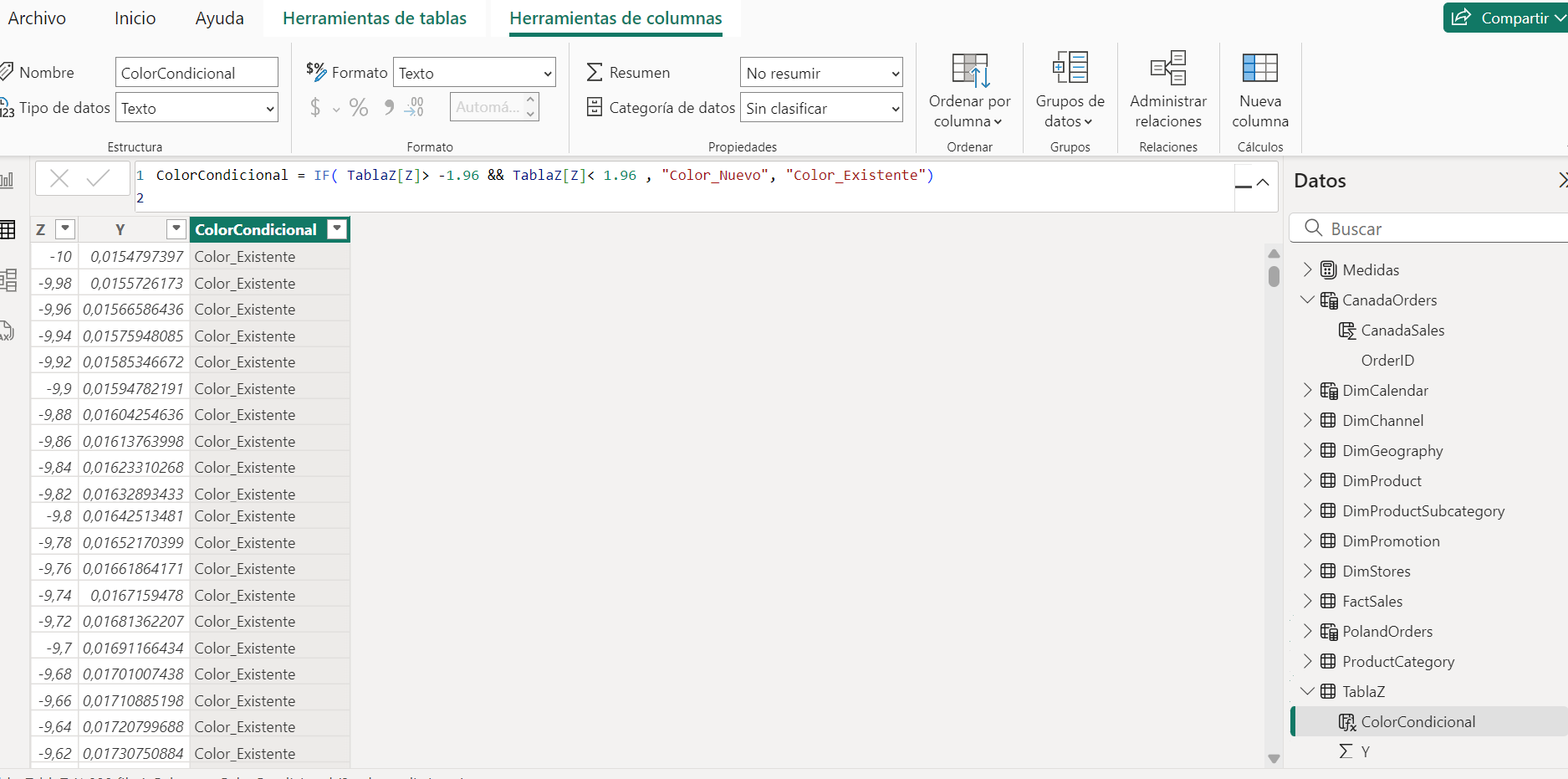
1. Se crea una tabla sobre las órdenes de ventas de Canadá y su monto:

**

1. Se crea una tabla sobre las órdenes de ventas de Polonia y su monto:



1. Se crea una columna condicional con fines de visualización posterior en base al gráfico de distribución normal:

**

1. Traducción de columna de “RegionCountryName” de Inglés a Español

*RegionCountryName =*

*SWITCH(*

*DimGeography[RegionCountryName1],*

*"United States", "Estados Unidos",*

*"Canada", "Canadá",*

*"the Netherlands", "Paises Bajos",*

*"Mexico", "México",*

*"Argentina", "Argentina",*

*"Brazil", "Brasil",*

*"United Kingdom", "Reino Unido",*

*"France", "Francia",*

*"Germany", "Alemania",*

*"Spain", "España",*

*"Italy", "Italia",*

*"Australia", "Australia",*

*"India", "India",*

*"China", "China",*

*"Japan", "Japón",*

*"South Africa", "Sudáfrica",*

*"Russia", "Rusia",*

*"Netherlands", "Países Bajos",*

*"Sweden", "Suecia",*

*"Norway", "Noruega",*

*"Finland", "Finlandia",*

*"Denmark", "Dinamarca",*

*"Belgium", "Bélgica",*

*"Switzerland", "Suiza",*

*"Austria", "Austria",*

*"Greece", "Grecia",*

*"Portugal", "Portugal",*

*"Turkey", "Turquía",*

*"Israel", "Israel",*

*"Saudi Arabia", "Arabia Saudita",*

*"United Arab Emirates", "Emiratos Árabes Unidos",*

*"Singapore", "Singapur",*

*"Malaysia", "Malasia",*

*"Philippines", "Filipinas",*

*"Thailand", "Tailandia",*

*"Vietnam", "Vietnam",*

*"South Korea", "Corea del Sur",*

*"New Zealand", "Nueva Zelanda",*

*"Colombia", "Colombia",*

*"Chile", "Chile",*

*"Peru", "Perú",*

*"Cuba", "Cuba",*

*"Dominican Republic", "República Dominicana",*

*"Puerto Rico", "Puerto Rico",*

*DimGeography[RegionCountryName1]*

*)*

1. Traducción de columna de “ContinentName” de Inglés a Español}

*Continente Nombre =*

*SWITCH(*

*DimGeography[ContinentName],*

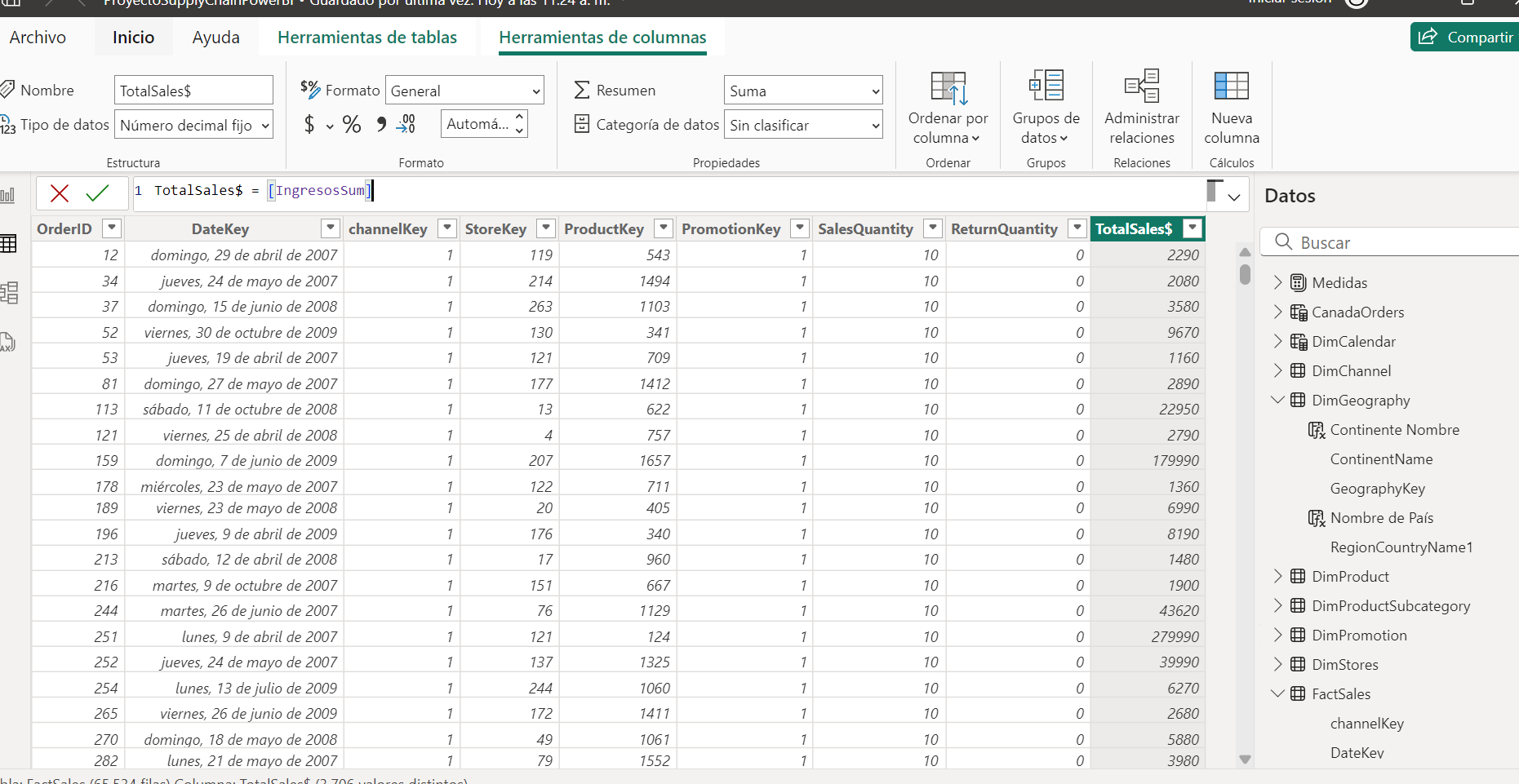
*"North America", "América del Norte",*

*"Asia", "Asia",*

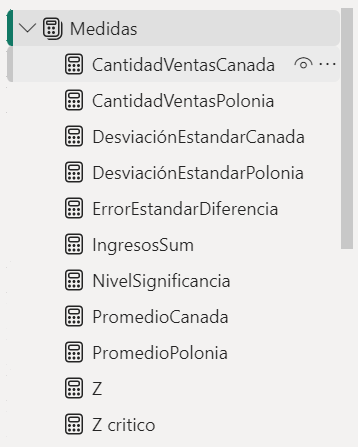
*"Europe", "Europa",*

*"Oceania", "Oceania")*

1. Creación de tabla para determinar los ingresos de cada order ID



Creación Tabla Medidas



*Medidas creadas:*

*-CantidadVentasCanada = COUNT(CanadaOrders[CanadaSales]*

*-CantidadVentasPolonia = COUNT(PolandOrders[OrderID] )*

*-DesviaciónEstandarCanada = STDEV.P(CanadaOrders[CanadaSales])*

*-DesviaciónEstandarPolonia = STDEV.P(PolandOrders[PolandSales])*

*-ErrorEstandarMedias =*

*VAR Var1 =[DesviaciónEstandarPolonia]*

*VAR Var2 = [DesviaciónEstandarCanada]*

*VAR n1 = COUNT(PolandOrders[OrderID] )*

*VAR n2 = COUNT(CanadaOrders[OrderID])*

*RETURN*

*SQRT(((Var1)^2 / n1) + ((Var2)^2 / n2))*

*-IngresosSum = SUM(DimProduct[UnitPrice]) \* SUM(FactSales[SalesQuantity])*

*–NivelSignificancia =*

*VAR NS = 0.05*

*RETURN NS*

*-PromedioCanada = AVERAGE(CanadaOrders[CanadaSales])*

*-PromedioPolonia = AVERAGE(PolandOrders[PolandSales])*

*-Z =*

*VAR mediaPolonia = [PromedioPolonia]*

*VAR mediaCanada = [PromedioCanada]*

*VAR errorestandar = [ErrorEstandarMedias]*

*RETURN*

*(mediaPolonia - mediaCanada ) / errorestandar*

*-Z critico =*

*VAR zcritico = 1.96*

*RETURN "+/-" & zcritico*

*-PromedioEmpleadosPolonia = CALCULATE( AVERAGE(DimStores[EmployeeCount]),*

*FILTER(DimGeography,*

*DimGeography[RegionCountryName] = "Poland"))*

*-PromedioEmpleadosMundial = AVERAGE(DimStores[EmployeeCount])*

*-CostosSum = SUM(DimProduct[UnitCost]) \* SUM(FactSales[SalesQuantity])*

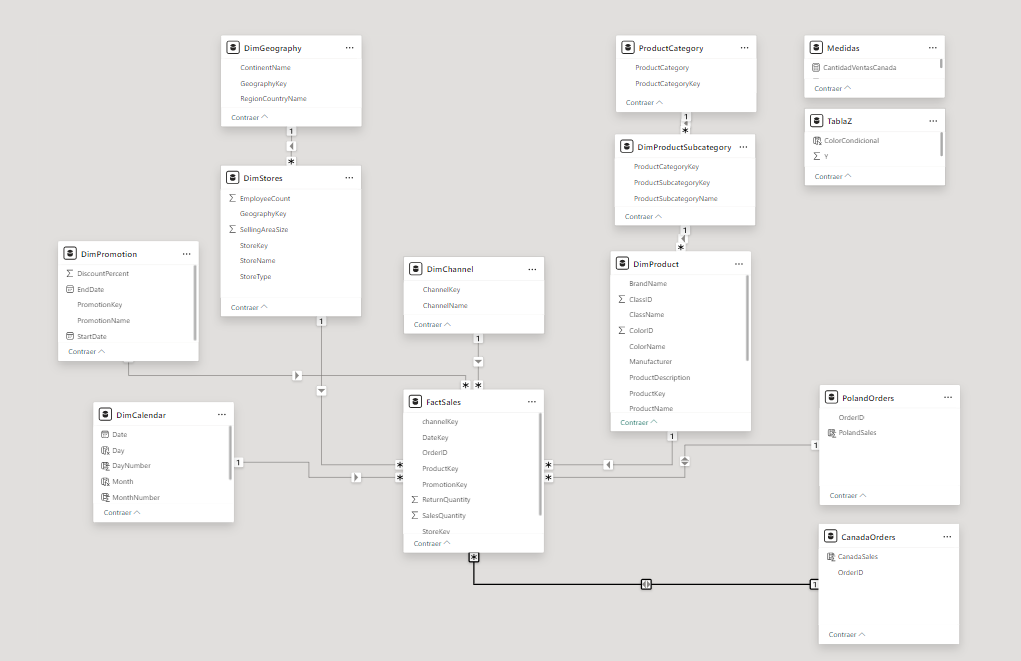
*-Ganancias = [IngresosSum] - [CostosSum]*

*-Max ventas = MAX(FactSales[TotalSales$])*

*-Min ventas = MIN(FactSales[TotalSales$])*

*-Porcentaje de devoluciones = ( SUM(FactSales[ReturnQuantity]) / SUM(FactSales[SalesQuantity]) )*

*Relaciones de tablas:*

**

Tablas Dimensionales:

* DimGeography: Esta tabla contiene información geográfica como país, nombre, código geográfico, etc. Sirve para ubicar geográficamente las ventas.
* DimProductCategory: Contiene información detallada sobre las categorías y subcategorías de los productos, como el nombre de la categoría, el subcategoría y la clave del producto.
* DimProductSubCategory: Esta tabla detalla las subcategorías de los productos, proporcionando una clasificación más granular.
* DimPromotion: Contiene información sobre las promociones, como el nombre de la promoción, la fecha de inicio y fin, etc.
* DimStores: Guarda información sobre las tiendas, en que país se encuentran, etc.
* DimChannel: Contiene información sobre los canales de venta, como el nombre del canal, el tipo de canal, etc.
* DimCalendar: Esta tabla contiene información sobre el tiempo, como fecha, día de la semana, mes, año, etc. Se utiliza para analizar las ventas a lo largo del tiempo.
* DimProduct: Contiene información detallada sobre los productos, como el nombre del producto, la clase, el color, el tamaño, etc.

Tabla de Hechos:

* FactSales: Esta es la tabla central y contiene las medidas numéricas relacionadas con las ventas, como la cantidad de ventas, el precio de venta, la cantidad de productos vendidos, etc. Conecta con todas las tablas dimensionales para proporcionar un contexto a los datos numéricos.

*NOTAS:*

*-No es necesario incluir tablas puentes.*

*-Las transformaciones se realizaron el 18/11/24*

Solapas y su descripción:

1. Se trata de la carátula del proyecto, ofrece una visión general de un proyecto de análisis que busca optimizar una cadena de suministro. Se abordarán aspectos globales, temporales y específicos de un país (Polonia).



1. Esta solapa sienta las bases del estudio al definir claramente el objetivo, las hipótesis y los conceptos estadísticos que se utilizarán para evaluar la relación entre la distancia geográfica y las ventas.

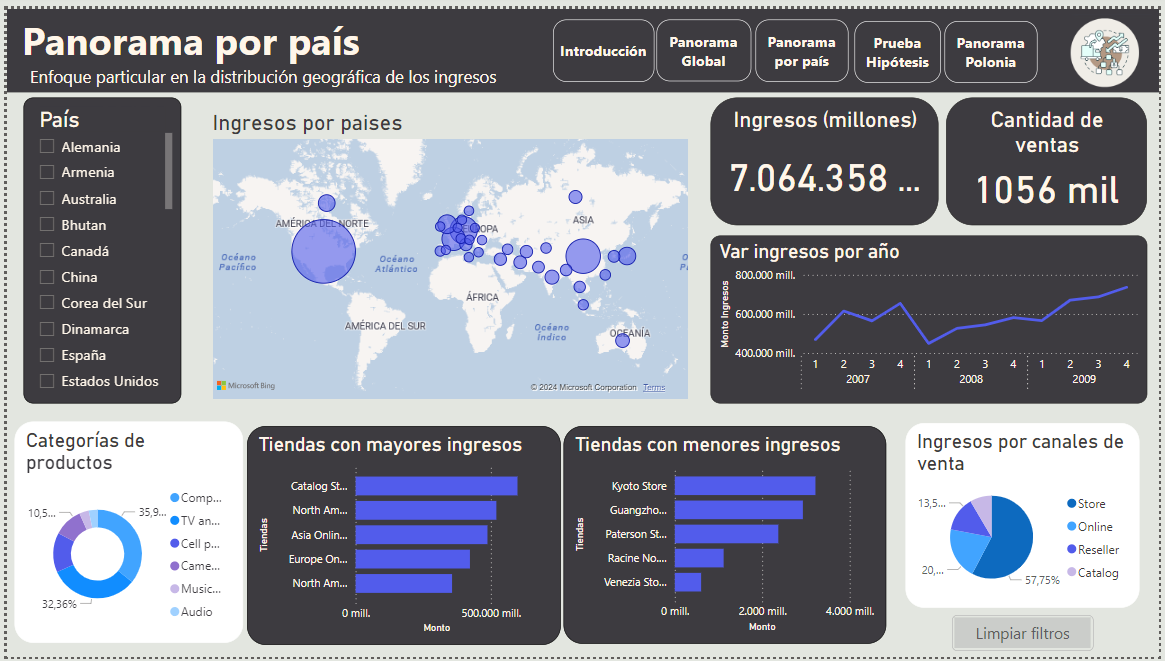


1. Esta solapa proporciona una visión general de los resultados financieros a nivel global, permitiendo identificar tendencias y patrones clave. Sin embargo, solo tiene un fin introductorio para conocer cómo es el desempeño de la multinacional en el mundo.

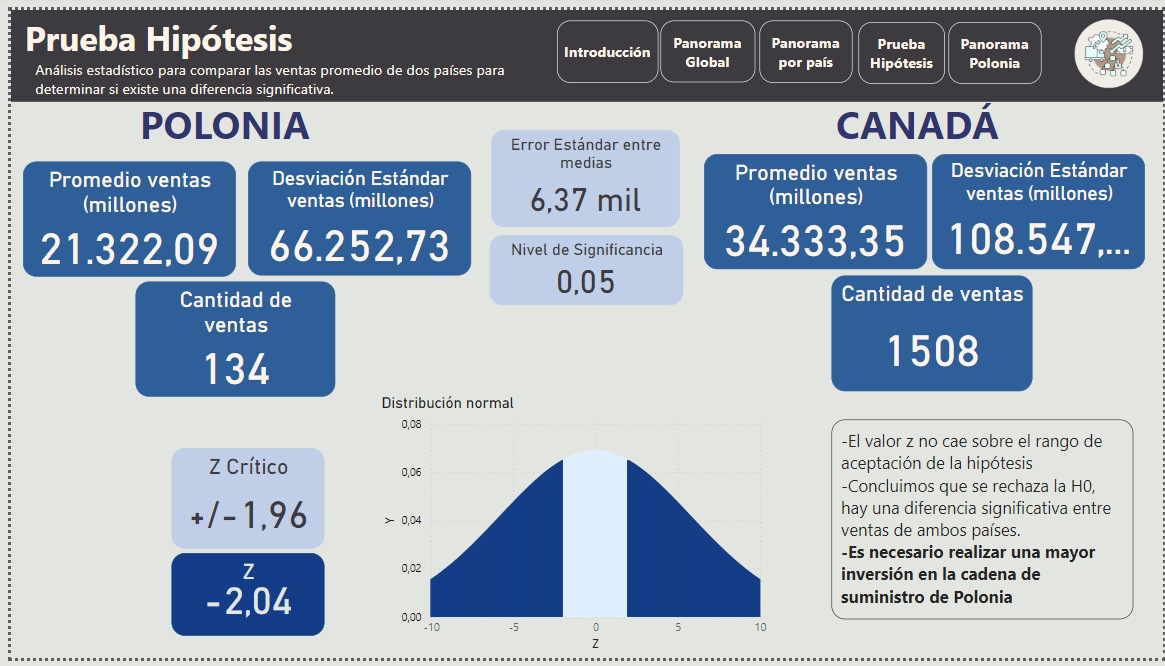


1. Se presenta una visión global de los ingresos de la empresa, desglosados por país y categoría de producto.

El objetivo de esta solapa es poder comparar los destinos en cuanto compete al conocimiento de la cadena de suministro haciendo enfoque en el desempeño de las tiendas físicas y los diferentes canales de venta de los países.



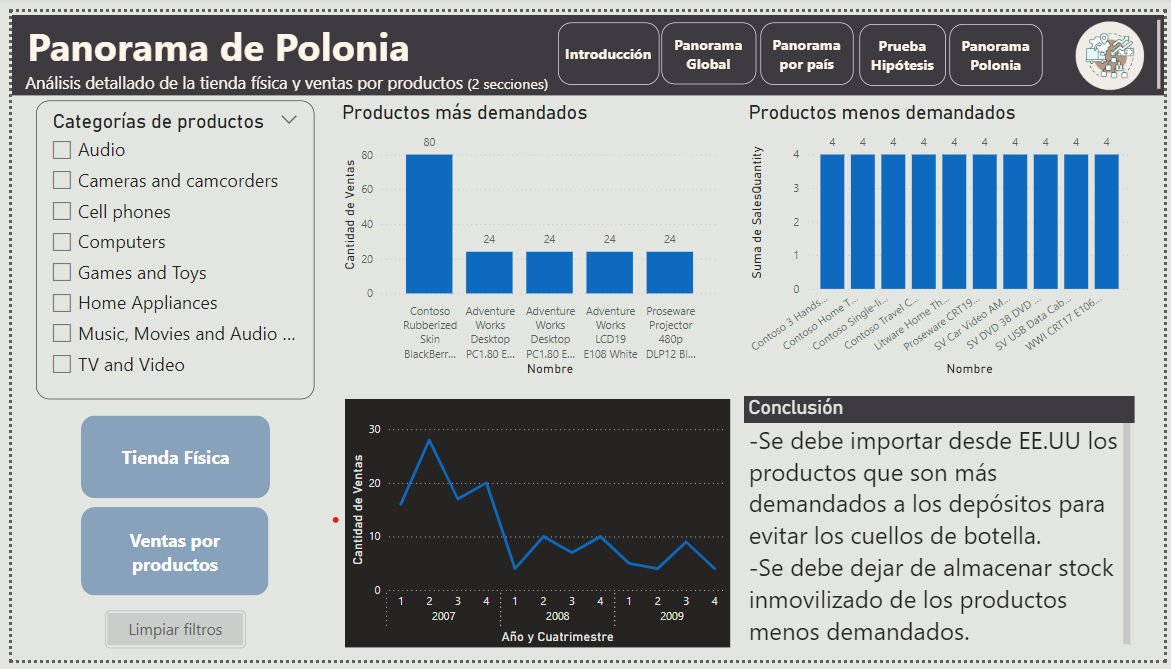
1. Presenta un análisis estadístico comparativo entre las ventas promedio de Polonia y Canadá con el objetivo de determinar si existe una diferencia significativa entre ambos países. Se utiliza la prueba de hipótesis de dos muestras y de dos colas.



1. a) La solapa señala una necesidad de ajustar el personal en la tienda de Warsaw Store para alinearlo con la demanda y mejorar el servicio al cliente. Los datos presentados respaldan la recomendación de contratar más empleados, especialmente para los días de mayor afluencia.



6. b) Se proporciona una visión clara y concisa del desempeño de la tienda en términos de ventas de productos. Los datos presentados permiten identificar oportunidades para mejorar la gestión del inventario, optimizar las compras y aumentar la rentabilidad de la tienda.



**Posibles iniciativas que se pueden llevar a cabo para complementar el proyecto**

En mi opinión se debería ampliar la base de datos implementando un erp en el sector logístico para conocer y analizar:

* Tiempo de ciclo de pedido: Desde que se realiza un pedido hasta que se entrega al cliente.
* Costo de transporte
* Tasa de entrega a tiempo: Porcentaje de pedidos entregados dentro de la fecha prometida.
* Nivel de servicio: Medición de la satisfacción del cliente en cuanto a la logística.
* Inventario en tránsito: Cantidad de inventario que se encuentra en movimiento entre diferentes ubicaciones.
* Tipos de medio de transporte más utilizados.
* Información sobre los proveedores.