

Máster en Big Data y Data Science

Universidad Internacional de Valencia

Soluciones de Inteligencia de Negocios

Actividad 2

Profesor Arturo Peralta

Estudiante: Christopher Flores

Usuario: christopher.flores

Segunda Convocatoria, 2024

Índice

1. Introducción
2. Proceso ETL
 - 2.1 Proceso E (Extraer)
 - 2.2 Proceso T (Transformar)
 - 2.3 Proceso L (Load / Cargar)
3. Informes/Visualización (Opcional)
4. Conclusiones

Introducción

En la actual dinámica de mercado del sector retail, donde la competencia es feroz y los clientes cuentan con una diversidad de opciones, las cadenas de supermercados enfrentan el desafío constante de innovar y mejorar sus operaciones para mantenerse a la vanguardia. "CostaRicaMart", una cadena de ropa líder en Costa Rica, no es ajena a esta realidad. En respuesta a este desafío, la cadena ha decidido implementar un Sistema de Business Intelligence (BI) robusto y estratégicamente diseñado para transformar datos en acciones concretas que impulsen sus operaciones de marketing y ventas.

El propósito de este Sistema de BI es doble: comprender con mayor precisión el comportamiento de compra de los clientes y, a partir de estos conocimientos, afinar la toma de decisiones estratégicas para aumentar las ventas y la satisfacción del cliente. Esta comprensión se convierte en un elemento transformador, permitiendo a "CostaRicaMart" no solo responder a las tendencias actuales sino anticipar las necesidades futuras de sus clientes.

Para alcanzar este objetivo, "CostaRicaMart" ha planificado utilizar un conjunto diverso de fuentes de datos que proporcionarán una visión integral de la operación del comercio. Desde información detallada de los clientes hasta minuciosos registros de facturación, pasando por un exhaustivo listado de productos y una base de datos completa de empleados y tiendas, cada fuente de datos servirá como una pieza crítica en el rompecabezas de la inteligencia de negocios. Estas fuentes de datos serán integradas en un modelo cohesivo, lo cual facilitará un acceso rápido y eficiente a la información, proporcionando así una base sólida para el análisis y la generación de reportes.

Con estos esfuerzos, "CostaRicaMart" se propone alcanzar varios objetivos estratégicos, incluyendo una mejora en la comprensión del comportamiento de compra de sus clientes, la identificación de patrones de compra y preferencias de productos, la optimización de las estrategias de marketing para mejorar las ventas y, finalmente, un aumento en la eficiencia operativa que se traduzca en una mayor satisfacción del cliente. Este documento detallará el proceso, desde la extracción de datos hasta la visualización de inteligencia, culminando en un conjunto de conclusiones y recomendaciones que impulsarán a "CostaRicaMart" hacia un futuro de éxito y crecimiento sostenido.

NOTA:

ACCESO A LOS DATOS: [Christopherdtacom/Soluciones-de-Inteligencia-de-Negocio](https://christopherdtacom/Soluciones-de-Inteligencia-de-Negocio)
(github.com)

Proceso ETL (Extraer, Transformar, Load/Cargar)

En nuestro proyecto de Business Intelligence, el proceso ETL juega un papel crucial en la consolidación y preparación de los datos para análisis avanzados y toma de decisiones. Este proceso consta de tres etapas esenciales:

1. **Extracción:** Inicialmente, extraemos datos de múltiples fuentes, incluyendo ventas, productos, clientes y proveedores, para reunir la información necesaria que respalde nuestros análisis.
2. **Transformación:** Posteriormente, transformamos estos datos brutos a través de la limpieza, normalización y unificación, asegurando que la información sea precisa, coherente y lista para el análisis. Este paso es esencial para garantizar la calidad de los datos en nuestros informes y paneles.
3. **Carga:** Finalmente, cargamos los datos transformados en nuestro almacén de datos diseñado para soportar consultas rápidas y eficientes, facilitando el acceso a la información clave.

Este ciclo ETL nos permite mantener nuestros datos actualizados y relevantes, soportando así la toma de decisiones estratégicas en nuestra empresa.

Proceso E (Extraer)

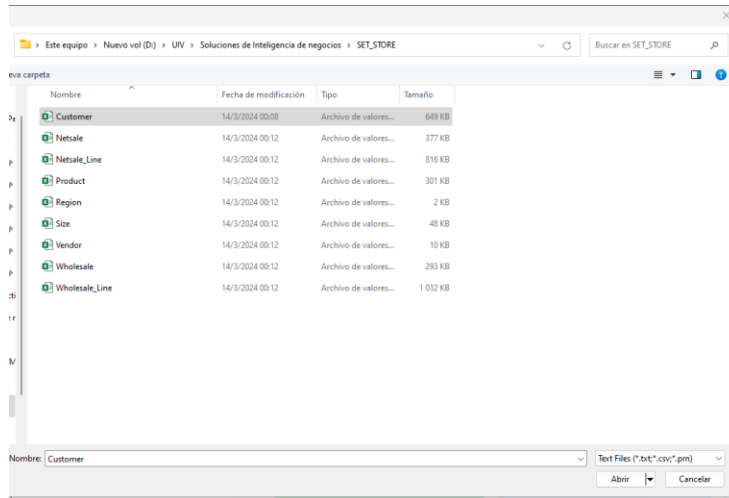
La extracción de datos es el primer paso esencial en nuestro proceso ETL, donde recolectamos información de diversas fuentes para consolidarla en un formato manejable. Utilizaremos las siguientes herramientas para la carga de datos:

- **Carga de datos: Fichero CSV** - Importaremos datos directamente desde archivos CSV, aprovechando su simplicidad y universalidad como formato de intercambio de datos.
- **Carga de datos: Base de datos** - Importaremos información desde SQL Server, lo que nos permite aprovechar los datos almacenados en sistemas de bases de datos relacionales para enriquecer nuestro análisis en Power BI.
- **Carga de datos: Python** - Utilizaremos scripts de Python para cargar datos de manera flexible a nuestro proyecto, demostrando la versatilidad de Power BI para la ingesta de datos de múltiples fuentes.

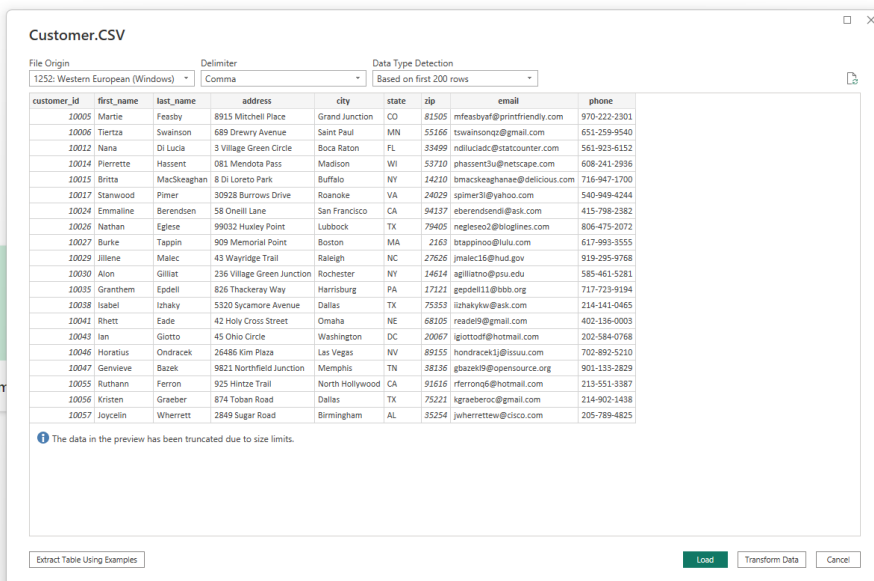
Carga de datos: Fichero CSV

En este caso se presenta como importar un archivo de Excel(CSV) a Power BI:

1. Clic en “Get data”
2. Seleccione “text/csv”
3. Busque la ubicación del archivo y selecciónelo



4. Previsualice la información y corrobore que sea la información adecuada.
Finalmente, Clic en el botón Cargar / Load

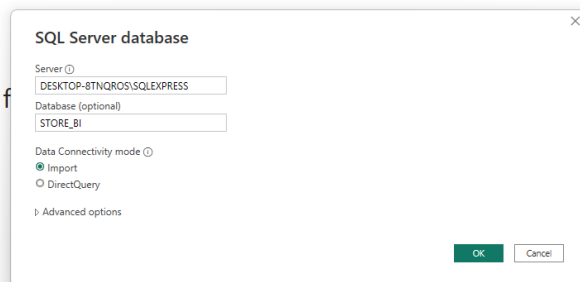


Carga de datos: Base de datos

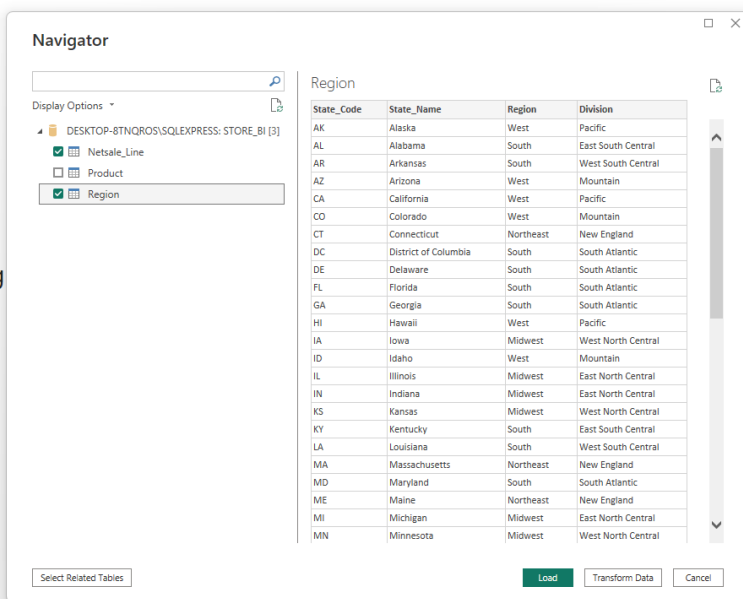
Ahora, vemos como importar una base de datos en Power BI, precisamente una base de datos local en Microsoft SQL Server Management Studio.

1. Clic en “Get data”
2. Clic en “SQL server”
3. Se abrirá una ventana donde tendrá que escribir la información que se le solicita:
 - a. Nombre del servidor
 - b. Nombre de la base de datos a la cual se quiera conectar

Como se observa en la imagen



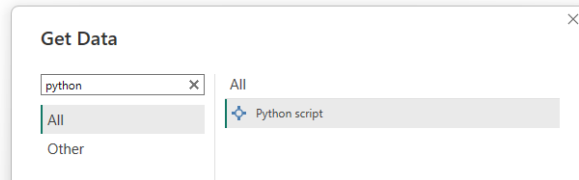
4. Una vez hecho el paso anterior, asegúrese de tener marcada la opción de “Import” y presione el botón “OK”. Además, es posible que aparezca una ventana, donde tiene que asegurarse de tener marcada la opción “Usar mis credenciales” y presione “Conectar”.
5. Por último, se conectará a la base de datos, donde tendrá que seleccionar las tablas que necesite importar y presione “Load”



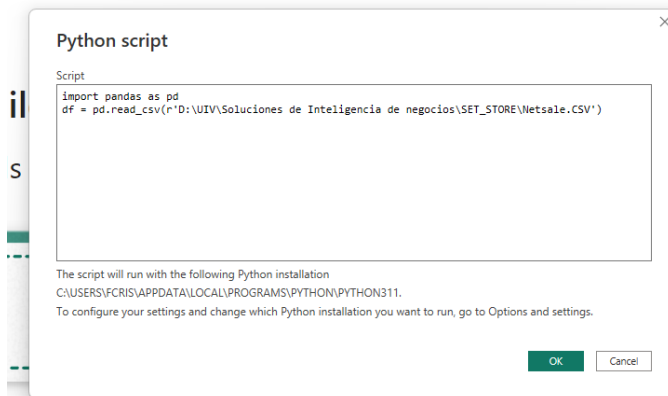
Carga de datos: Python

A continuación, se explica como importar un archivo CSV a través Python:

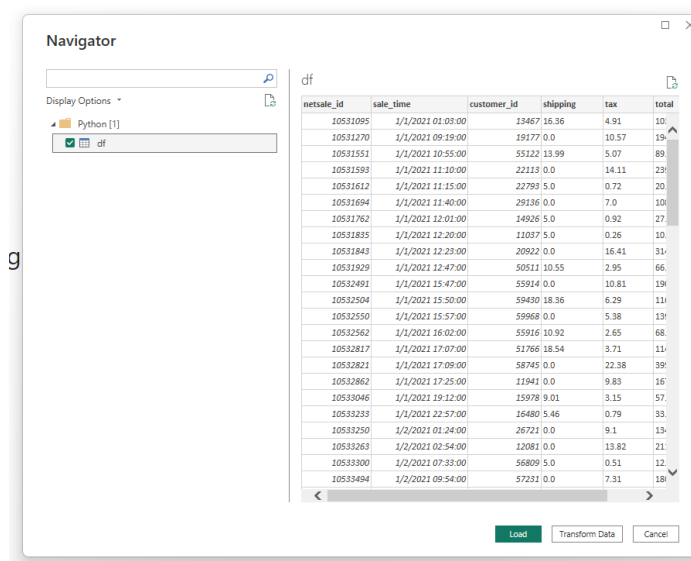
1. Clic en “Get data”
2. Clic en la opción “More”
3. Busque en las opciones: “Python script”



4. Una vez seleccionado, se abrirá una ventana llamada “Python script” donde tendrá que escribir el código de Python necesario para importar la tabla. Y presione el botón “OK”

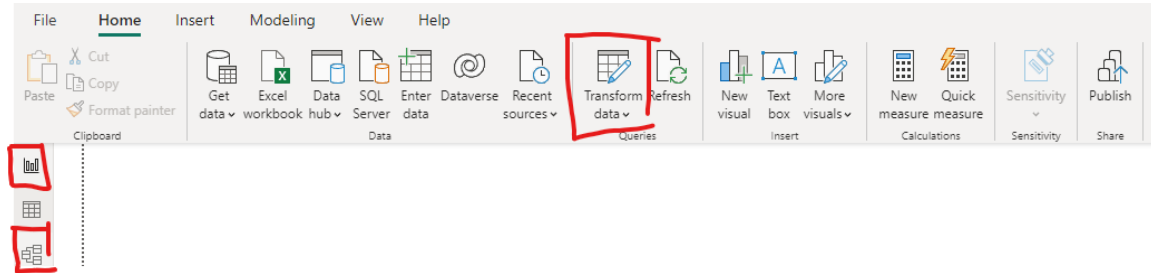


5. Posteriormente se abrirá una venta de navegación donde podrá escoger la tabla con nombre df (así se nombró en el código) y carguela (load button)

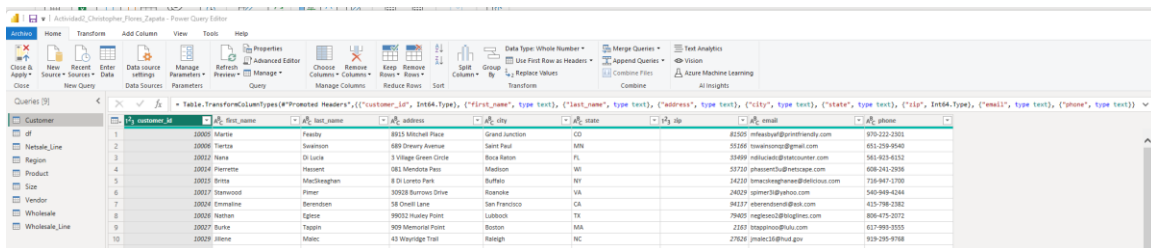


Proceso T (Transformar)

Power BI contiene una sección exclusiva para la transformación de datos. Para acceder a estas funciones: Seleccione “Model” o “Report” y notará que hay una opción “Transform data”



Entonces, se abrirá la siguiente ventana, donde podrá hacer todas las transformaciones



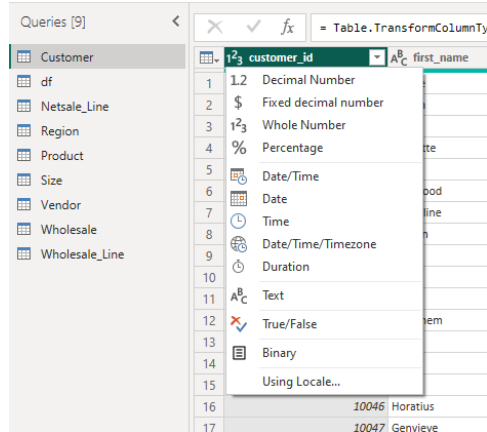
NOTA IMPORTANTE

Los pasos mencionados anteriormente no serán nuevamente mencionados, ya que a continuación se detallarán los procesos de transformación de datos hechos, asimismo, estos procesos se mostrarán una única vez, independientemente de la cantidad de veces que se hicieron, esto para evitar redundancia.

Transformación 1 (Cambio de tipo de dato):

En este caso se cambia de la tabla “Customer” la columna “customer_id”. Esto pues, POWER BI detectó el campo como numérico, sin embargo, se necesita de tipo texto para relacionarlo con otras tablas.

1. Clic izquierdo sobre el símbolo de tipo de dato de la columna en este caso “customer_id”
2. Ahora, seleccione el tipo de dato que desea, en este caso “text”



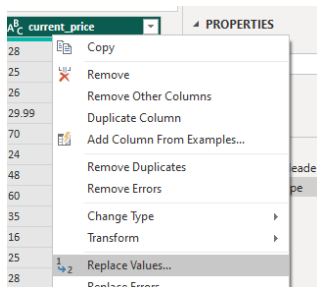
Transformación 2 (Reemplazo de valores):

Como puede observar en la columna “current_price” de la tabla “Product” hay algunos valores con error, esto se debe a que los valores decimales están dados por un punto (23.3), sin embargo, dependiendo del entorno, eso puede variar y en este caso se necesita que esté dado por una coma (23,3).

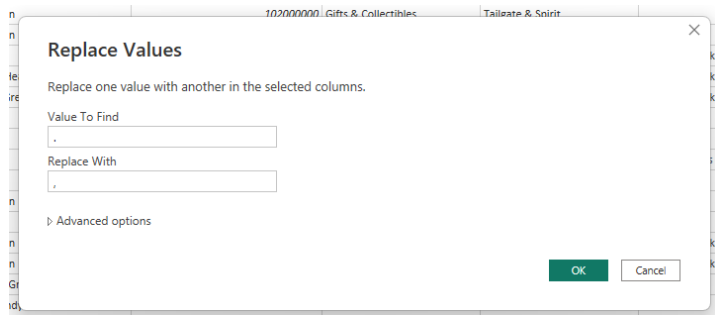
The screenshot shows the 'current_price' column with values 28, 25, 26, and 70. The column is highlighted, and the 'Error' message is visible.

Por lo que se siguen los siguientes cambios para corregir este error:

1. Se cambia la columna a texto (como se mostró anteriormente)
2. Clic derecho y “Replace Values”



3. Se abre una ventana, donde buscamos "." y lo reemplazamos por ",". Y clic en "OK"



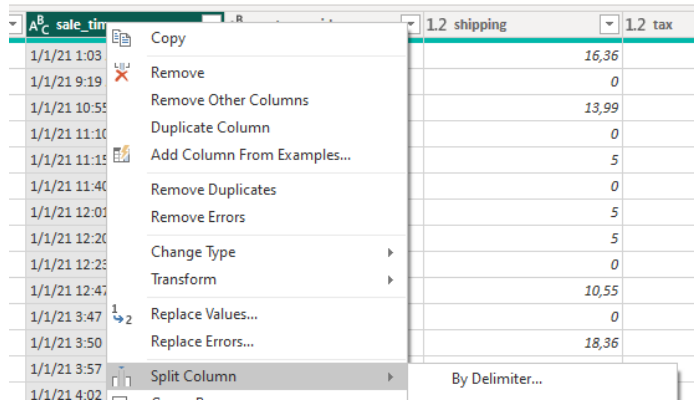
4. Por último, se convierte la columna a decimal.

1.2 current_price
28
25
26
29,99

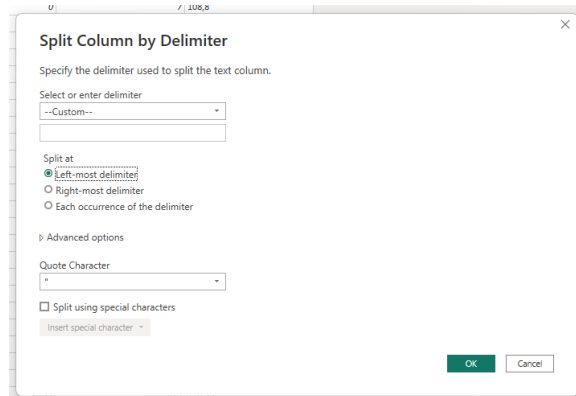
Transformación 3 (División de columna):

En este caso, lo haremos con la columna “sale_time” de nuestra tabla “df”.

1. Primero, vamos a hacer clic derecho y buscar la opción “Split Column” y “By Delimiter”



2. Nos aseguramos de que esté seleccionado “customer” y en la sección de abajo damos un espacio para que sea el delimitar, y nos aseguramos de tener el “Split at” en la opción que necesitamos, en este caso “left-most delimiter”

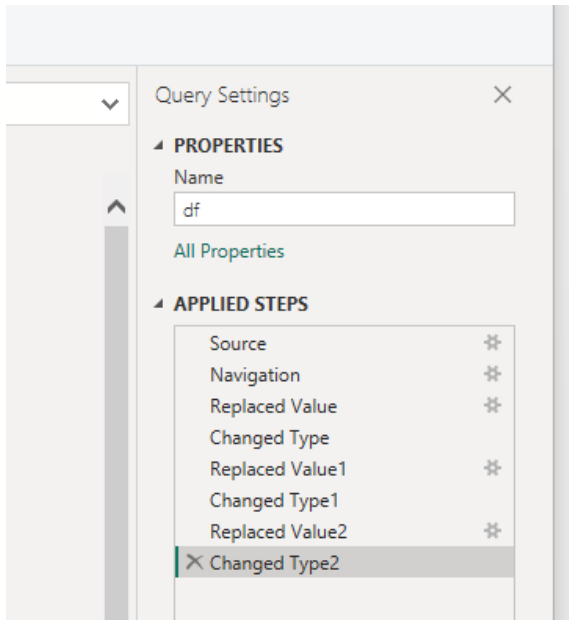


3. Después de presionar “OK”, se harán los cambios

sale_time.1	sale_time.2
1/1/21	01:03:00
1/1/21	09:19:00
1/1/21	10:55:00
1/1/21	11:10:00
1/1/21	11:15:00
1/1/21	11:40:00
1/1/21	12:01:00
1/1/21	12:20:00

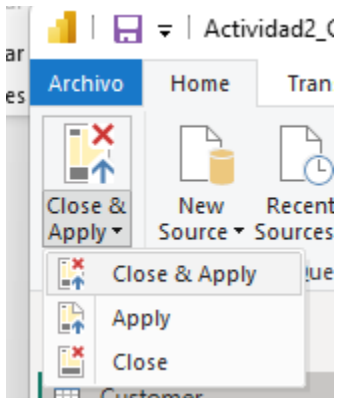
Consideración

En el apartado de la derecha de la ventana de transformaciones, se observan todas las transformaciones hechas para cada tabla, donde se puede regresar a pasos anteriores de ser requerido



De esta misma manera, se da un ejemplo que se hicieron más transformaciones que no se muestran, ya que son análogas a las descritas.

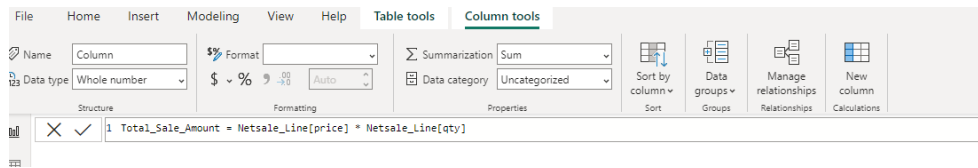
Una vez, se está satisfecho con las transformaciones hechas, se da clic en “close & apply” para realizar todos los cambios



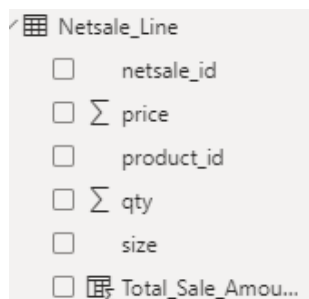
Columna calculada con DAX

En cualquier apartado “Model”, “Report” o “Data”, verá la opción “New Column”, se selecciona y se abrirá un cuadro de comando, donde se usa DAX.

En este caso se está creando la columna “Total_Sale_Amount”



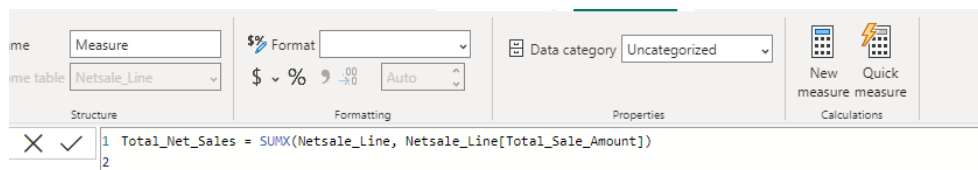
Clic sobre el check una vez escrita y se añadirá a tabla seleccionada



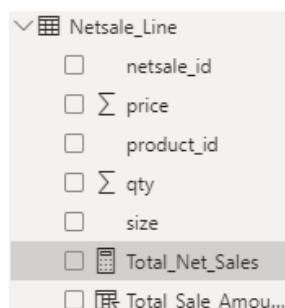
Medida calculada con DAX

Análogamente cualquier apartado “Model”, “Report” o “Data”, verá la opción “New Measure”, se selecciona y se abrirá un cuadro de comando, donde se usa DAX.

En este caso se está creando la medida “Total_Net_Sales”



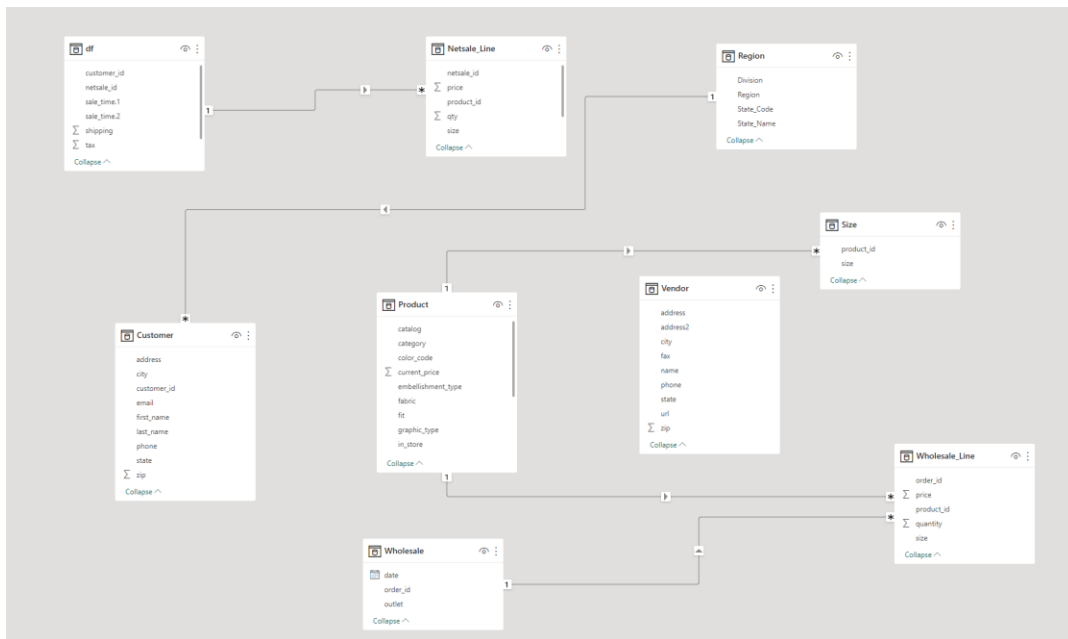
Clic sobre el check una vez escrita y se añadirá a tabla seleccionada



Proceso L (Load / Cargar)

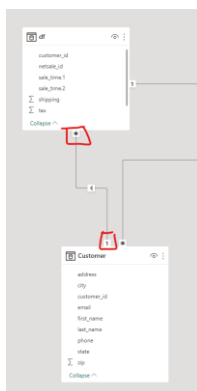
Tras la extracción y transformación de datos, el proceso de Carga representa el paso final en nuestra estrategia ETL para el proyecto de Business Intelligence. En esta etapa, los datos ya procesados y refinados se transfieren al destino final, que puede ser un almacén de datos (data warehouse) o una base de datos diseñada (como es el caso) para consultas analíticas y reportes en Power BI. La carga eficiente y precisa de estos datos es crucial para asegurar la integridad y la disponibilidad de la información para el análisis y la toma de decisiones.

Para este punto podemos ver que POWER BI a hecho algunas conexiones de manera automática por el nombre y tipo de datos de nuestras tablas



Sin embargo, estas conexiones pueden no ser las optimas ni todas las necesarias, por lo que es necesario revisar y cambiar lo que considere penitente.

Para crear nuevas relaciones solo basta con arrastrar la columna (PK) encima de la otra (FK).



En este caso se creó una relación de 1:n entre las tablas “customer” y “df”, esta relación (1:n) fue creada de manera automática, si no satisface con los requerimientos se debe hacer clic izquierdo sobre la relación y “propiedades”, posteriormente se abrirá una ventana donde se podrá modificar la relación

Edit relationship

Select tables and columns that are related.

df

netsale_id	customer_id	shipping	tax	total	sale_time.1	sale_time.2
10548484	59286	0	0	114.55	1/9/21	08:31:00
10573800	57145	0	0	119.1	1/20/21	16:50:00
10576520	15368	0	0	241.2	1/22/21	10:53:00

Customer

customer_id	first_name	last_name	address	city	state	zip	email
10313	Joli	Newbury	4 Merchant Park	Sacramento	CA	94297	jnewbury6@yahoo.
10340	Windy	Maplethorpe	767 Messerschmidt Drive	Sacramento	CA	94207	wmaplethorpe28@
10368	Carrie	Rushby	7595 Maple Wood Center	Sacramento	CA	95818	crushby81@hotmail

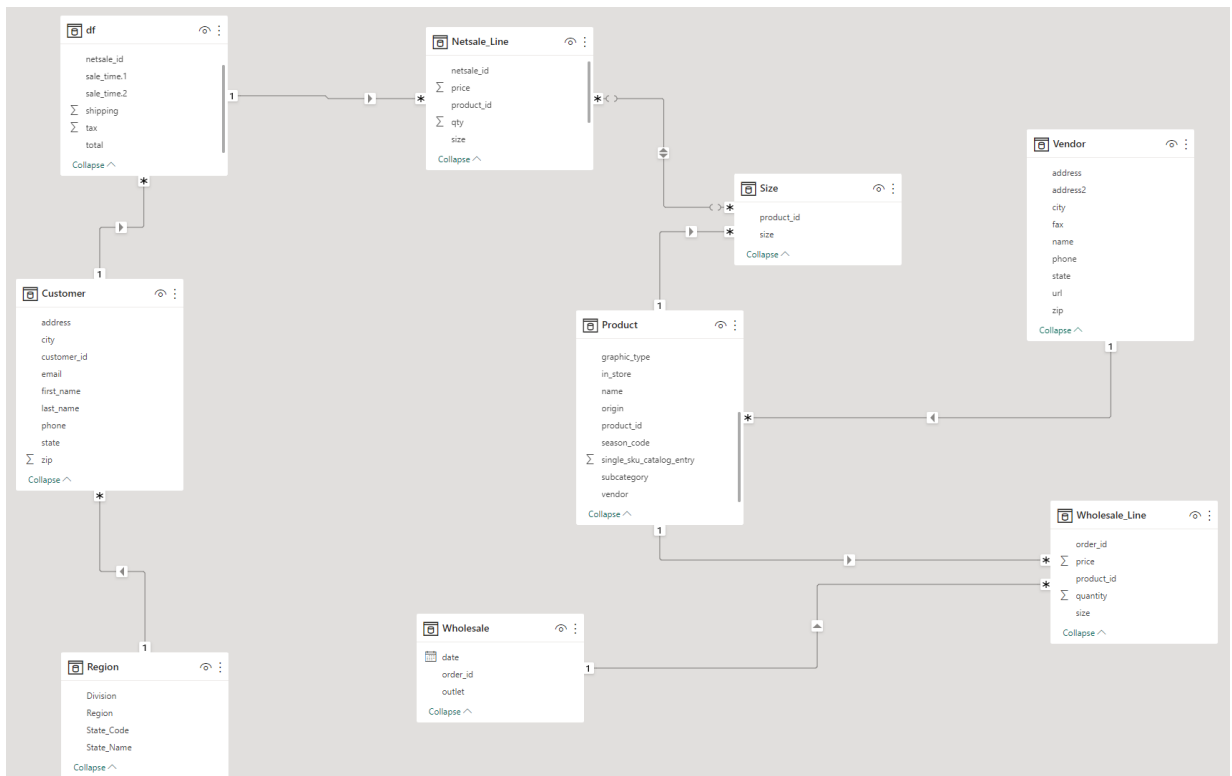
Cardinality: Many to one (*:1) Cross filter direction: Single

☒ Make this relationship active ☐ Apply security filter in both directions

☐ Assume referential integrity

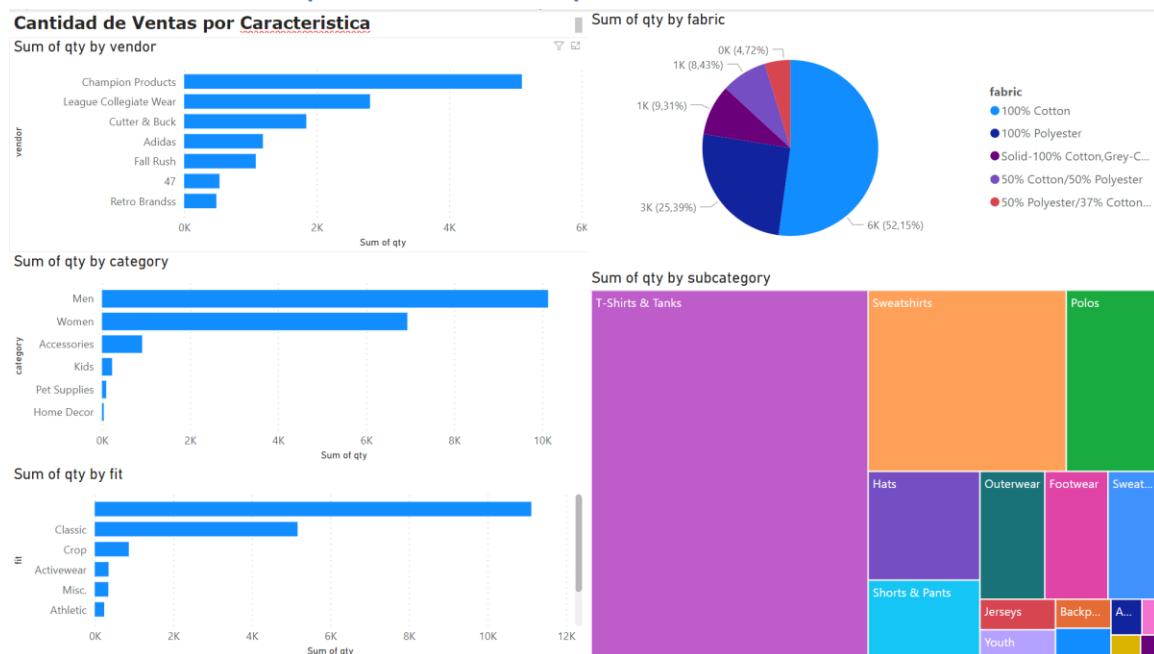
OK Cancel

Finalmente, nuestro modelo de datos es el siguiente:



Informes/Visualización

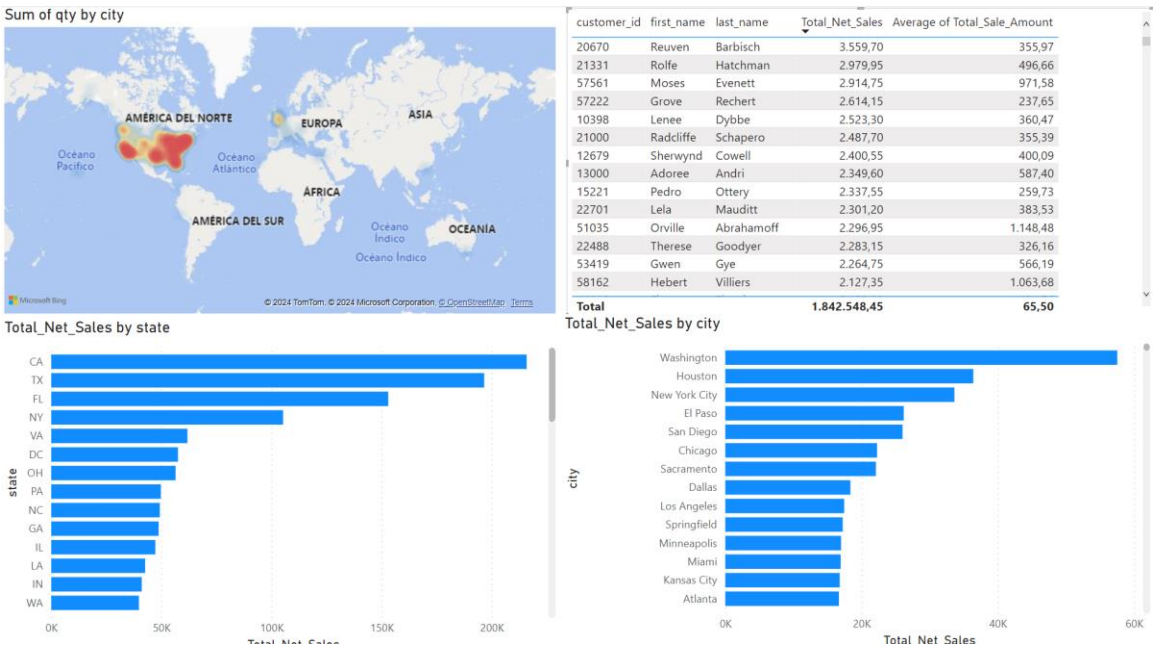
Análisis de ventas netas por características de productos



1. Cantidad de Ventas por Proveedor (Cantidad de qty by vendor): Champion Products parece ser el proveedor con la mayor cantidad de unidades vendidas, seguido de League Collegiate Wear y Cutter & Buck. Esto sugiere que sus productos son populares o bien surtidos.
2. Cantidad de Ventas por Categoría (Sum of qty by category): La categoría 'Men' domina claramente las ventas en términos de cantidad, con 'Women' y 'Accessories' también contribuyendo significativamente. Este patrón indica que los esfuerzos de marketing y la gestión de inventario deberían enfocarse en estos segmentos. Además, puede ser un indicador para explorar la expansión de la línea de productos en estas categorías.
3. Cantidad de Ventas por Tipo de Tela (Sum of qty by fabric): El gráfico de pastel muestra que una gran proporción de las ventas (52.15%) pertenecen a productos hechos con 100% algodón, seguidos por aquellos con una mezcla de 50% poliéster y 37% algodón. Esto sugiere una preferencia de los consumidores por ciertos materiales, lo cual podría influir en decisiones de compra y producción futura.
4. Cantidad de Ventas por Subcategoría (Sum of qty by subcategory): 'T-Shirts & Tanks' es la subcategoría más vendida, lo que indica una alta demanda en este segmento. 'Sweatshirts' y 'Polos' también muestran un buen desempeño. Es recomendable evaluar la estrategia de precios y promociones en estas áreas para maximizar los ingresos.
5. Cantidad de Ventas por Tipo de Ajuste (Sum of qty by fit): El 'Classic fit' lidera con una amplia ventaja en términos de cantidad vendida, lo cual puede indicar

que los consumidores prefieren este tipo de ajuste sobre los demás. Podría ser beneficioso para la empresa invertir en variedades adicionales de 'Classic fit' o realizar más marketing en torno a ellos.

Segmentación de clientes



1. Suma de Cantidad por Ciudad (Sum of qty by city): Aquí se utiliza un mapa de calor para mostrar dónde se concentran las ventas. Parece que hay un enfoque significativo en América del Norte, lo que sugiere que esta es su principal área de mercado.
2. Total de Ventas Netas por Estado (Total_Net_Sales by state): El gráfico de barras muestra que California (CA), Texas (TX) y Florida (FL) son los estados líderes en ventas netas, lo cual puede indicar mercados fuertes o una estrategia de marketing exitosa en estas áreas.
3. Clientes Individuales con Ventas Totales Netas y Promedio de Monto Total de Venta (Total_Net_Sales by customer): Esta tabla ofrece una lista de clientes individuales con sus ventas totales netas y el promedio del monto total de venta. Los clientes con mayores ventas totales deben ser una prioridad para los esfuerzos de retención y desarrollo de relaciones, ya que son probablemente tus clientes más valiosos.
4. Total de Ventas Netas por Ciudad (Total_Net_Sales by city): En el gráfico de barras, Washington lidera en ventas netas por ciudad, seguido de Houston y Nueva York. Esto puede resaltar áreas urbanas con alto potencial de ventas o con una base de clientes leales. También podría influir en las decisiones relacionadas con la asignación de recursos y la planificación de la expansión de la red de ventas.

Conclusiones

Este proyecto de BI ha demostrado ser una exploración completa y multifacética de los datos de ventas de una empresa que opera en múltiples canales y regiones geográficas. A través de un proceso ETL riguroso, hemos extraído datos de varias fuentes, incluyendo archivos planos y bases de datos, asegurando una fuente de información diversa y rica.

Logros Específicos

La integración exitosa de datos de clientes, facturas y productos, junto con información vital de los empleados y tiendas, ha resultado en un modelo de datos coherente y eficiente.

Las visualizaciones interactivas y los informes generados han ofrecido insights significativos sobre patrones de compra y preferencias de productos, facilitando la identificación de oportunidades para la personalización del marketing y promociones dirigidas.

Las medidas y columnas calculadas mediante DAX han permitido la creación de KPIs personalizados y análisis en tiempo real, asegurando que el marketing y las ventas puedan reaccionar dinámicamente a las tendencias emergentes.

Impacto Operativo y Estratégico

La eficiencia operativa se ha visto reforzada al identificar las áreas de alta demanda y optimizar el surtido de productos en las tiendas, así como al ajustar el rendimiento del inventario en función de las tendencias de ventas.

El enfoque en la satisfacción del cliente se ha materializado en estrategias de ventas más adaptadas y en una experiencia de compra mejorada que ha contribuido a fomentar la lealtad del cliente y el crecimiento de las ventas.

Recomendaciones Finales

Se sugiere continuar con la expansión y refinamiento del sistema de BI para incluir análisis predictivos y de inteligencia artificial, lo cual podría anticipar las necesidades del mercado y mejorar aún más la toma de decisiones.

Considerar la incorporación de más fuentes de datos en tiempo real, como el comportamiento en línea y las interacciones en redes sociales, para obtener una comprensión aún más completa del cliente.

Es importante que "CostaRicaMart" mantenga un enfoque en la formación continua y el desarrollo de habilidades en BI dentro de su equipo, asegurando que la organización pueda seguir maximizando el valor de sus datos.