

PROYECTO BODEGA DE DATOS

Creación de bodega de datos



UNIVERSIDAD JORGE TADEO LOZANO
BODEGA DE DATOS

David Ricardo Guzmán
Gabriela Cortes Merchán
Jorge Armando Arévalo

1. Proyecto de Curso: Revisión de Alcance

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

1.1. Actividades

En grupos, realice las siguientes actividades:

Revisión de tableros y reportes relacionado con AdventureWorks

- Revise (al menos) tres tableros o reportes similares, disponibles en Internet, sobre consultas y análisis han hecho otras personas sobre AdventureWorks. Revise estos proyectos similares y determine las principales variables y dimensiones que se revisan en cada uno.

Definición del alcance

- Teniendo en cuenta los tableros y reportes revisados, *determine al menos tres variables (hechos, eventos, transacciones)* que se podrían revisar en el caso de estudio.
- Por cada una de las variables (hechos) *determine al menos tres criterios de análisis (dimensiones), adicionales a la dimensión tiempo.*

1.2. Elementos mínimos del informe

Título: Análisis de ventas AdventureGJD

Grupo: David Ricardo Guzmán, Jorge Armando Arévalo, Gabriela Cortes

Descripción: (Descripción textual del alcance inicial. Escriba una introducción al informe y describa cada una de las tablas y figuras a medida que se presentan en el texto)

Tableros y Reportes revisados:

Sitios web, documentos, reportes revisados	
Nombre, URL	Descripción de principales características
Análisis de ventas para AdventureWorks: https://community.fabric.microsoft.com/t5/Data-Stories-Gallery/Sales-Analysis-for-AdventureWorks/mp/2549649	<p>El siguiente panel de control nos brinda una profunda perspectiva a través de un análisis exhaustivo centrado en las ventas, las ganancias y los productos de Adventure Works. A simple vista, se destacan varios aspectos clave, como el crecimiento de las ventas, el total de ventas, los costos totales, las ventas anuales, los ingresos promedio por unidad y las ventas por región. Lo notable de este análisis es que se apoya en tablas geoespaciales, gráficos de barras, gráficos lineales y comentarios textuales que ofrecen una comprensión integral de la contribución regional y un panorama general de la empresa.</p> <p>Este versátil panel de control permite aplicar filtros por año, mes, país y categoría. Además, se sumerge en un análisis detallado de las ganancias y los productos. En cuanto a las ganancias, proporciona una comparación entre el año actual y el año anterior, destacando el beneficio actual. En lo que respecta a los productos, ofrece un desglose de las ventas por categoría, que incluye bicicletas, accesorios y ropa, junto con un ranking de los cinco productos más vendidos y los productos más adquiridos según la ocupación.</p> <p>Cabe destacar que nos sentimos atraídos por este informe de visualización debido a su simplicidad, diseño y la claridad que aporta en la interpretación de datos.</p>

<p>Panel de Control de Adventure Works: https://www.novyprio.com/project/adventure-works-dashboard-1</p>	<p>Este dashboard nos sumerge en un análisis comercial diseñado para proporcionar una visión detallada de las ventas e ingresos de una empresa. Utilizando una temporalidad que, inclusive permite el análisis diario de la información, este panel de control desglosa la información por productos y clientes, ofreciendo una representación visual mediante gráficos de barras, visualizaciones geoespaciales e indicadores clave que revelan el comportamiento tanto de los productos como de los clientes, tanto a nivel de categorías como de subcategorías.</p> <p>Lo que hace que este panel de control destaque es su capacidad para proporcionar información crucial sobre las ganancias mensuales, la rentabilidad y los pedidos mensuales. Esto permite una evaluación precisa del desempeño de las tiendas, países, entre otros, que permiten concluir si se lograron los objetivos de la organización. Además, ofrece una visualización relacionada con el cumplimiento o ejecución frente al presupuesto de ventas, la cual resulta valiosa. También resalta el producto que generó la mayor cantidad de pedidos y ganancias.</p>
<p>Análisis de ingresos y ganancias de AdventureWorks: https://community.fabric.microsoft.com/t5/Data-Stories-Gallery/Adventure-Works-Revenue-and-Profit-Analysis/m-p/2541797</p>	<p>Este panel de control nos invita a un análisis exhaustivo centrado en las ventas anuales y el margen de beneficio, explorando detenidamente las variaciones entre regiones y subcategorías. Además, muestra las preferencias de los clientes en términos de ocupación y edad en relación con las ventas. Asimismo, involucra un análisis detallado de los productos principales, detallando en sus ventas, costos y beneficios. Todo esto se presenta de manera visual a través de gráficos de barras, gráficos de tendencia y gráficos de burbujas vinculados a representaciones geoespaciales.</p> <p>Es importante destacar que, este panel de control no se limita a proporcionar información fragmentada, sino que también ofrece una visión general que permite</p>

	comprender diversos comportamientos en un solo vistazo.
--	---

Listado inicial de hechos y dimensiones:

	Descripción
Variables / Hechos: Análisis de ventas (Hecho: Ventas)	
Criterio de Análisis / Dimensión	1. Dimensión Región: País y estados: Esta dimensión es creada para observar las ventas por región, en las cuales se puede agrupar y analizar el desempeño en ventas por país, entre otros criterios asociados con la rentabilidad de los productos, categorías y subcategorías asociadas.
Criterio de Análisis / Dimensión	2. Dimensión tiempo: Año y mes: Esta dimensión es creada para ver el comportamiento tanto anual como mensual de las ventas, permitiendo ver la cantidad de ventas por año y los ingresos promedio por unidad.
Criterio de Análisis / Dimensión	3. Dimensión categoría: Accesorios, bicicletas, ropa y componentes: Esta dimensión permite observar el comportamiento de ventas según la categoría del producto.
Variables / Hechos: Análisis sobre ventas e ingresos: (Hecho: Ventas)	
Criterio de Análisis / Dimensión	1. Dimensión categoría: categoría y subcategoría: Esta dimensión esta creada agrupar el desempeño de los ingresos agrupados por categoría y subcategoría de producto.
Criterio de Análisis / Dimensión	2. Dimensión tiempo: Días, mes: Esta dimensión esta creada para ver el comportamiento diario y mensual de los ingresos y su desempeño
Criterio de Análisis / Dimensión	3. Dimensión país: Continente, País: Esta dimensión logra visualizar por continente que país genera más ingresos.
Criterio de Análisis / Dimensión	4. Dimensión producto: Productos: Esta dimensión esta creada para mostrar los ingresos por producto.

2. Proyecto de Curso: Revisión de la Base de Datos

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

2.1. Actividades

En grupos, realice las siguientes actividades:

Revisión de tablas relacionadas con los tableros y reportes

- A partir de los tableros y reportes revisados, y de los hechos y dimensiones revisados en la hoja de trabajo anterior, determine cuáles son las tablas que están relacionadas con cada uno.

Creación de consultas

- Considerando cada una de las variables y criterios de análisis, realice una serie de consultas que le permitan extraer los datos relacionados. Estas consultas se usarán más adelante para construir la bodega de datos.
- Determine cuáles variables y dimensiones que se querían revisar inicialmente no se pueden determinar a través de las consultas

2.2. Elementos mínimos del informe

Título: Análisis de ventas AdventureGJD

Grupo: Gabriela Merchán, Jorge Armando Arévalo, David Ricardo Guzmán Mora

Descripción: (Descripción textual de las consultas definidas para el proyecto. El informe puede hacer referencia a archivos en un repositorio de trabajo en Github)

Las consultas definidas con cuatro (4), todas parten de un CTE que relaciona la tabla de ventas (Sales_SalesOrderHeader) con otras tablas relacionadas como clientes, territorio, vendedor, personas y recursos humanos, entre otras.

Consulta 1 (Top 10 de Vendedores)

Esta consulta es un ranking que relaciona el top 10 de los vendedores con el mejor desempeño en ventas. Lo anterior, a través de la definición de una consulta de agrupamiento y utilizando una función de agregación como count().

```
select
    SalesName,
    count(*) as Ventas
from datos
group by SalesName
order by Ventas desc
limit 10
```

SalesName	Ventas
Jillian Carson	473
Michael Blythe	450
Tsvi Reiter	429
Linda Mitchell	418
Jae Pak	348
José Saraiva	271
Shu Ito	242
Garrett Vargas	234
David Campbell	189
Ranjit Varkey Chudukatil	175

El resultado muestra el top 10 de los vendedores que más venden, siendo el primero de ellos Jillian Carson con 473 unidades vendidas.

Consulta 2 (Ventas por País)

Esta consulta hace uso de la sentencia condicional case when con la finalidad de agrupar las distintas regiones geográficas de los Estados Unidos ('Southwest','Northwest','Southeast','Central','Northeast') en un solo país y lo denomina USA. Posteriormente realiza un conteo de las ventas, cuyo propósito es mostrar las unidades vendidas por cada uno de los países.

```
select
```

```

case when Territory in
('Southwest','Northwest','Southeast','Central','Northeast')
then 'USA' else Territory end as Territory2,
count(*) as Ventas
from datos
group by Territory2
order by Ventas desc

```

+-----+-----+	
Territory2	Ventas
+-----+-----+	
USA	2474
Canada	692
France	188
United Kingdom	188
Germany	139
Australia	125
+-----+-----+	

Esta consulta refleja que los Estados Unidos tiene las mayores ventas entre todos los países (2474) y que son superiores en buen grado a las que muestra el país con el segundo nivel de ventas (Canadá).

Consulta 3 (Crecimiento Anual en Ventas)

Esta consulta anida dos consultas adicionales al CTE, siendo su objetivo definir una nueva variable (Año) las cual se obtiene a partir de la fecha de realización de la venta. Posteriormente, la segunda consulta transpone la variable año a nivel de columnas, lo que permite hacer un análisis de la variación de las ventas para cada año.

```

datos2 as (
select Anno as Year,count(*) as Ventas
from datos
group by Year
),

datos3 as(
select
sum(case when year = 2011 then Ventas else 0 end) ventas11,
sum(case when year = 2012 then Ventas else 0 end) ventas12,
sum(case when year = 2013 then Ventas else 0 end) ventas13,
sum(case when year = 2014 then Ventas else 0 end) ventas14
from datos2
)
select ((ventas12 - ventas11)/ventas11)*100 as varventas12,
((ventas13 - ventas12)/ventas12)*100 as varventas13,
((ventas14 - ventas13)/ventas13)*100 as varventas14
from datos3

```


varventas12	varventas13	varventas14
125.3608	46.2031	-60.5757

Esta Consulta refleja que el año 2012 muestra un gran desempeño en ventas (125.3%, en contraste con 2014 que muestra una disminución importante del orden del 60.5%)

Consulta 4 (Promedio de ventas por género)

Esta consulta, partiendo del mismo CTE, genera un análisis adicional el cual muestra el promedio de ventas, pero esta vez ya no por cantidad sino por valor, de lo cual hace uso del atributo 'subtotal' y hace un promedio con la función de agregación avg() y lo agrupa a partir del atributo 'Gender'.

Gender	PromedioVentas
M	20166.27855807
F	22483.18177841

De lo cual se obtiene que las mujeres tienen un promedio de compras mayor al de los hombres.

Listado de Tablas relacionadas

	Descripción
Variables / Hechos:	
Criterio de Análisis / Dimensión	Cliente: Permite relacionar medidas de agregación para el análisis de variables de ventas por cliente.
Criterio de Análisis / Dimensión	Vendedor: permite analizar, mediante el agrupamiento, a los vendedores de mejor desempeño en ventas.
Criterio de Análisis / Dimensión	Persona: permite relacionar características y atributos de la persona, como su nombre, se relaciona con tablas como cliente y vendedores, por lo que se deben manejar joins para traer esta información.
Criterio de Análisis / Dimensión	Territorio: permite relacionar la zona geográfica en la cual se han realizados las ventas de AdventureWorks, permitiendo análisis como la tienda de mayores o menores ventas.
Criterio de Análisis / Dimensión	HumanResources: es un conjunto de tablas que contiene información de los empleados de la empresa, reuniendo características como la fecha de contratación, el género, el departamento para el que trabajan y la periodicidad de pago de la nómina.

Listado de Consultas revisadas

Consulta, URL, nombre	Descripción
Consulta 1: Top 10 de vendedores	Muestra a los 10 vendedores con el mejor desempeño en ventas.
Consulta 2: Ventas por país	Muestra las unidades vendidas por país con presencia de tiendas
Consulta 3: Crecimiento Anual en Ventas	Ejecuta un análisis de desempeño de ventas por año, mediante la comparación respecto del año anterior.
Consulta 4: Promedio de ventas por género	Promedia las ventas por monto y las agrupa para analizar, en promedio, quien realiza más compras si los hombres o las mujeres.

Consideraciones (datos difíciles de obtener, datos que no se pueden obtener, información adicional)

	Descripción
Consulta 2: Ventas por país.	Unir las distintas regiones de Estados Unidos para unificarlas con una etiqueta de un solo país (USA)

	Descripción
Consulta 3: Crecimiento Anual en Ventas	El reto de esta consulta fue la definición de la variable año y su transposición para obtener los crecimientos en ventas.

3. Proyecto de Curso: Modelo Dimensional

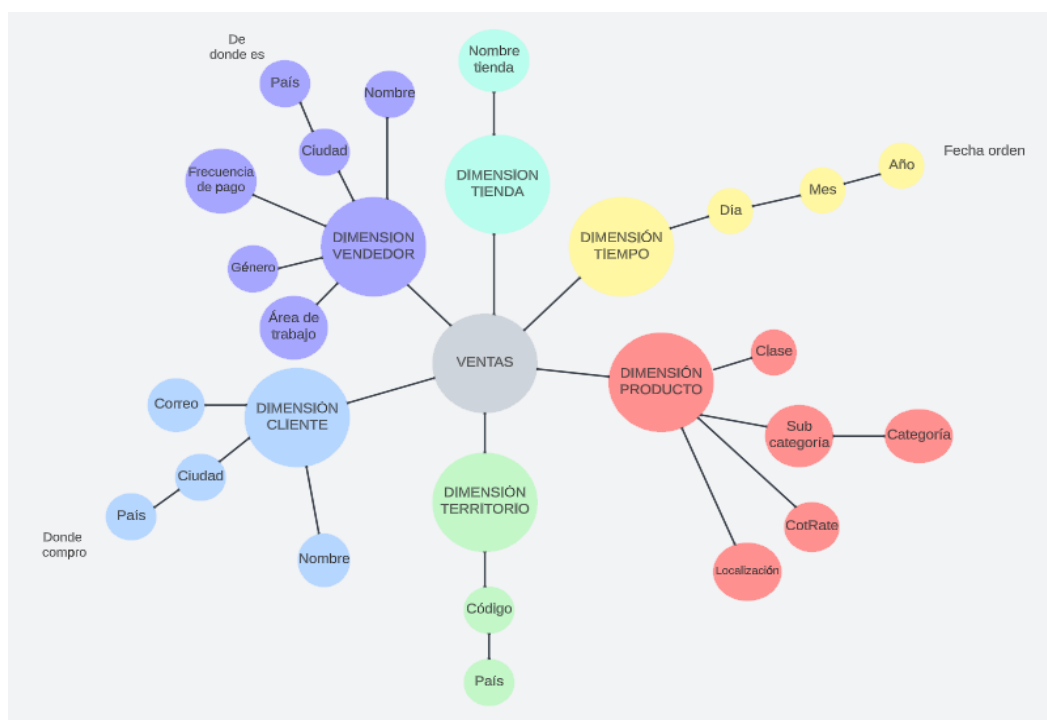
El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

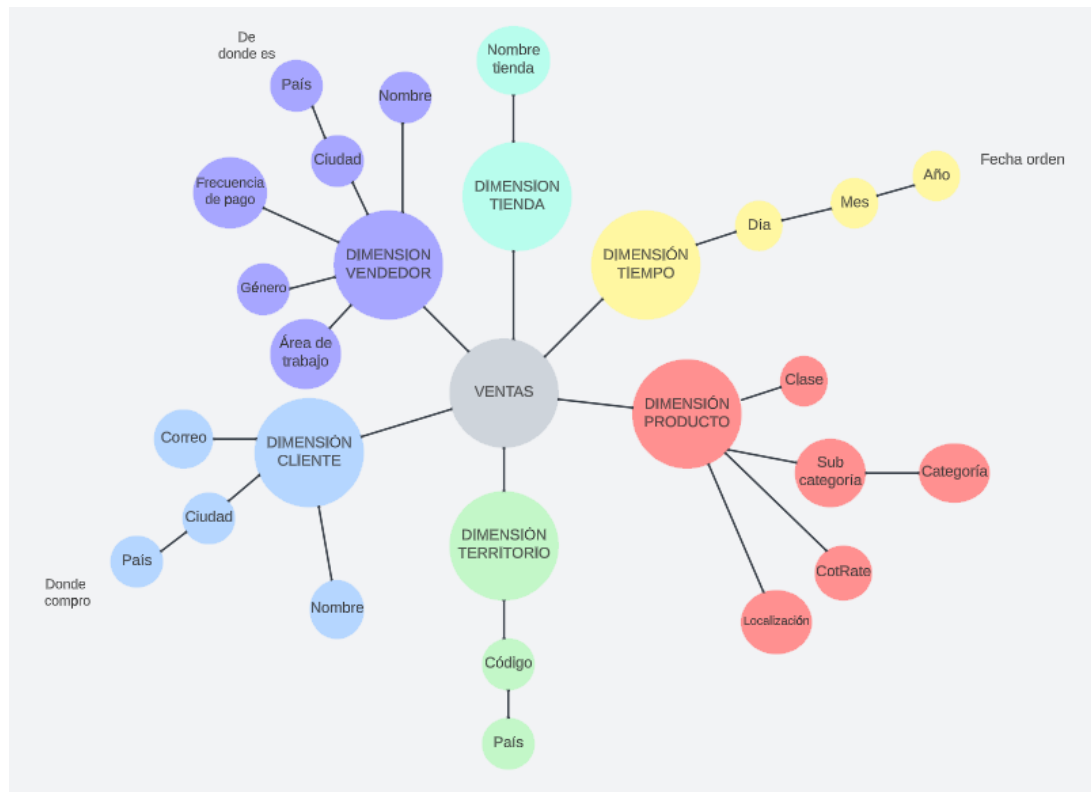
3.1. Actividades

En grupos, realice las siguientes actividades:

Elabore un modelo dimensional para el proyecto

- Considerando las variables que se desean incluir en los tableros y reportes, y las consultas realizadas en la base de datos, elabore un modelo dimensional para el proyecto.
- Elabore una estrella para cada una de las variables (hechos) a analizar. Cree un modelo para cada uno de estas variables, colocando los hechos en el centro y creando líneas y puntos para cada una de las dimensiones y niveles de análisis.
- Considere los datos que sí pudo obtener a través de las consultas. No incluya atributos o dimensiones que no pudo obtener en las consultas





3.2. Elementos mínimos del informe

Título: Análisis de ventas AdventureGJD

Grupo: David Ricardo Guzmán, Jorge Armando Arévalo, Gabriela Cortes

Descripción: (Descripción textual de cada uno de los modelos dimensionales)

Modelo Dimensional: (Análisis de ventas AdventureGJD)

En la tabla de tiempo generaríamos la separación de información con el fin de generar agrupaciones y análisis de información según la fecha o período de tiempo en la cual ocurrieron los hechos del análisis.

dim_time			
primary key(date_key), unique index date_value (date_value)			
date_key	int(8)	not Null	
date_value	date	not Null	
month_number	int(3)	not Null	
year4	int(5)	not Null	
day_of_week	int(5)	not Null	
day_of_week_name	char(12)	not Null	

Para la dimensión vendedor, traemos la información que caracteriza al vendedor y que es correspondiente a la identificación, el nombre, la ciudad, país, frecuencia de pago, departamento de la empresa en el cual labora, con el fin de sacar cifras directas de ventas por vendedor.

dim_vendedor	Tipo Dato		
primary key (salesperson_key), unique index (salesperson_id)			
salesperson_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
salesperson_id	int	not Null	
salesperson_name	varchar(50)	not Null	
city	varchar(30)	not Null	
country	varchar(30)	not Null	
gender	varchar(5)	not Null	
paymentfreq	tinyint(8)	not Null	
department	varchar(30)	not Null	default CURRENT_TIMESTAMP,
last_update	timestamp	not Null	

Esta tabla de tienda se diseña con el fin de tener datos de ventas, por nombre de tienda y validar, por ejemplo, cual tienda es la que más vende.

dim_tienda	Tipo Dato		
primary key (store_key), unique index (store_id)			
store_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
store_id	int	not Null	
store	varchar(100)	not Null	
last_update	timestamp	not Null	

Esta tabla de producto se crea con el fin de realizar los análisis completos de que productos son los más vendidos y realizar el mismo análisis según su categoría y subcategoría.

dim_producto	Tipo Dato		
primary key (product_key), unique index (product_id)			
product_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
product_id	int	not Null	
product	varchar(100)	not Null	
class	char(2),		
category	varchar(100)	not Null	
subcategory	varchar(100)	not Null	
last_update	timestamp	not Null	

La dimensión de territorio se crea con el fin hacer análisis de información relacionados con las zonas geográficas en las cuales la empresa desarrolla su actividad, permitiendo obtener información como, por ejemplo, ventas por país.

dim_territorio	Tipo Dato		
primary key (territory_key), unique index (territory_id)			
territory_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
territory_id	int	not Null	
name_territory	varchar(100)	not Null	
last_update	timestamp	not Null	

La dimensión de clientes es la requerida para realizar un estudio de marketing y ventas, analizar clientes potenciales y facilitar su contacto, por ejemplo, para la definición de campañas de fidelización. que requieren los productos de la empresa.

dim_cliente	Tipo Dato		
primary key (cliente_key), unique index (cliente_id)			
cliente_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
cliente_id	smallint(8)	not Null	
cliente	varchar(200)	not Null	
city	varchar(50)	not Null	
province	varchar(50)	not Null	
country	varchar(50)	not Null	
email	varchar(50)	not Null	
last_update	timestamp	not Null	

En este caso, la tabla de hechos facilita la agrupación de variables de tipo numérico con los hechos que se van a analizar y que se derivan de la actividad de comercialización y de ventas. Se define específicamente con el fin de tener todos los datos de ventas y realizar las conexiones con las distintas dimensiones con el fin de realizar los queries para el análisis de información de la empresa.

fact_ventas	Tipo Dato		
ventas_key	int(8)	not null	AUTO_INCREMENT,
ventas_id	int(11)	not null,	
date_key	int(8)	not Null	
salesperson_key	int(8)	not Null	
store_key	int(8)	not Null	
product_key	int(8)	not Null	
territory_key	int(8)	not Null	
cliente_key	int(8)	not Null	
count_sales	int(11)	not Null	

fact_ventas	Tipo Dato		
SubTotal	decimal(19,4)	not Null	
TaxAmt	decimal(19,4)	not Null	
Unit_Price	decimal(19,4)	not Null	
UnitPriceDiscount	decimal(19,4)	not Null	
Line_Total	decimal(38,6)	not Null	
OrderQty	smallint(8)	not Null	
primary key (ventas_key), unique index ventas_id(ventas_id),			

4. Proyecto de Curso: Diseño de la Bodega de Datos

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

4.1. Actividades

En grupos, realice las siguientes actividades:

Elabore un diseño para la Bodega de Datos

- Considerando los modelos dimensionales planteados en las hojas de trabajo anteriores, *seleccione un solo modelo dimensional*.
- Elabore un *diseño de la base de datos usando un esquema en estrella*. Plantee una tabla para los hechos y una tabla para cada una de las dimensiones.

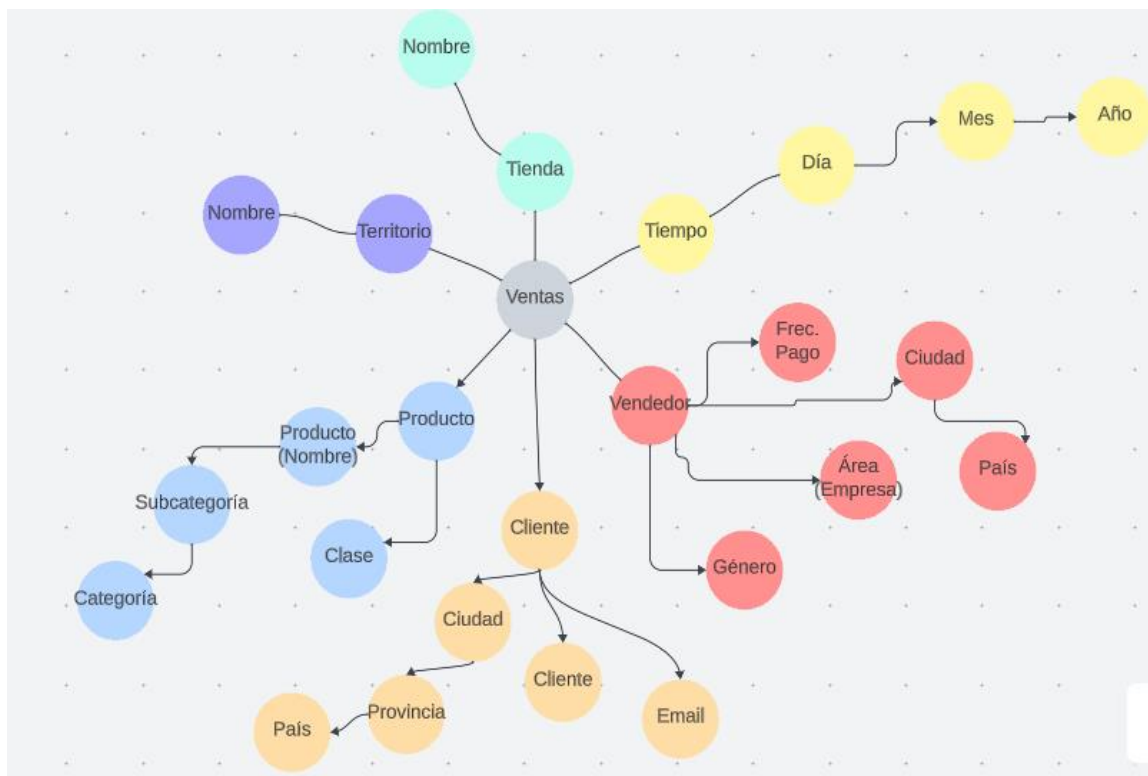


Ilustración 1. Esquema dimensional final.

- Defina los nombres de las tablas, y los nombres y tipos de datos de cada una de las columnas de esas tablas.

dim_time			
primary key(date_key), unique index date_value (date_value)			
date_key	int(8)	not Null	
date_value	date	not Null	
month_number	int(3)	not Null	
year4	int(5)	not Null	
day_of_week	int(5)	not Null	
day_of_week_name	char(12)	not Null	

dim_vendedor			
primary key (salesperson_key), unique index (salesperson_id)			
salesperson_key	int	not Null	AUTO_INCREMENT,
salesperson_id	int	not Null	
salesperson_name	varchar(50)	not Null	
city	varchar(30)	not Null	
country	varchar(30)	not Null	
gender	varchar(5)	not Null	
paymentfreq	tinyint(8)	not Null	
department	varchar(30)	not Null	default CURRENT_TIMESTAMP,
last_update	timestamp	not Null	

dim_tienda			
primary key (store_key), unique index (store_id)			
store_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
store_id	int	not Null	
store	varchar(100)	not Null	
last_update	timestamp	not Null	

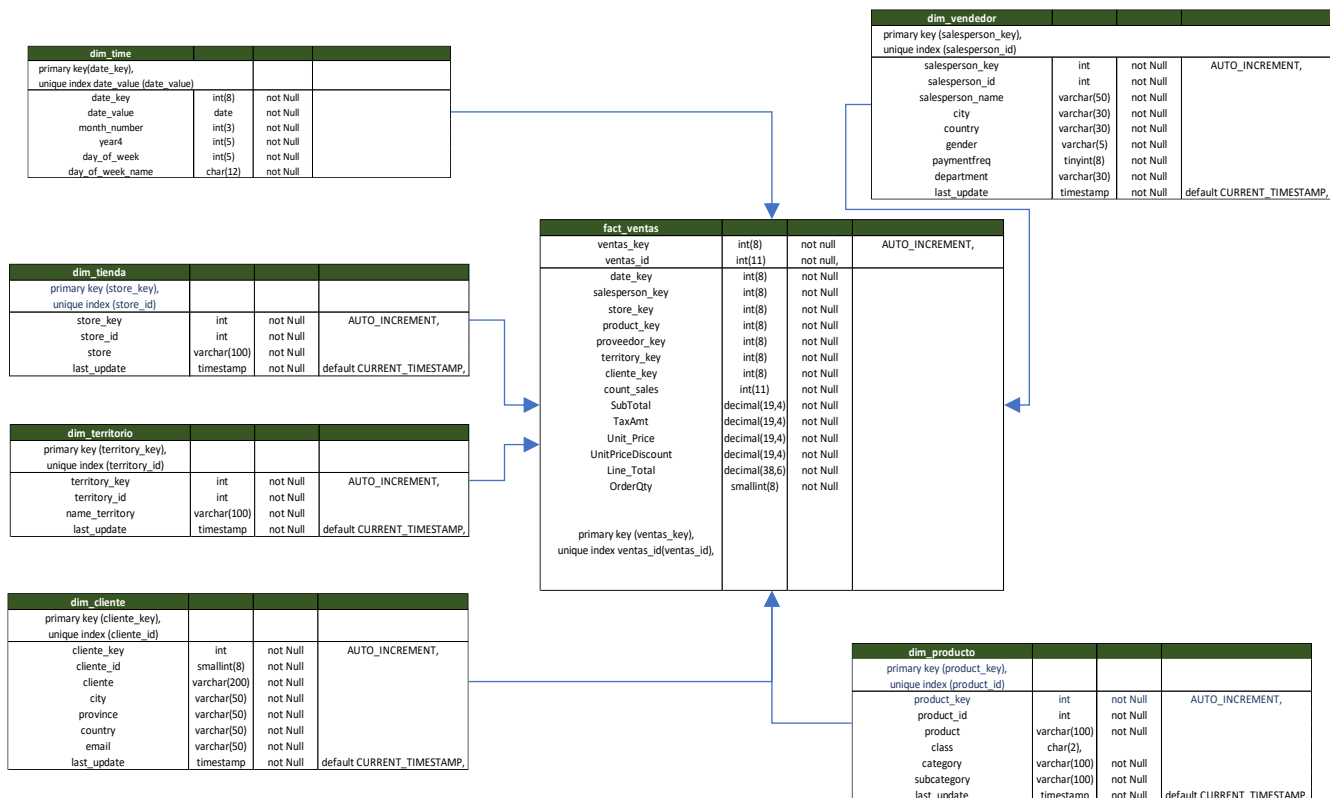
dim_producto			
primary key (product_key), unique index (product_id)			
product_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
product_id	int	not Null	
product	varchar(100)	not Null	
class	char(2)		
category	varchar(100)	not Null	
subcategory	varchar(100)	not Null	
last_update	timestamp	not Null	

dim_territorio			
primary key (territory_key), unique index (territory_id)			
territory_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
territory_id	int	not Null	
name_territory	varchar(100)	not Null	
last_update	timestamp	not Null	

dim_cliente			
primary key (cliente_key), unique index (cliente_id)			
cliente_key	int	not Null	AUTO_INCREMENT, default CURRENT_TIMESTAMP,
cliente_id	smallint(8)	not Null	
cliente	varchar(200)	not Null	
city	varchar(50)	not Null	
province	varchar(50)	not Null	
country	varchar(50)	not Null	
email	varchar(50)	not Null	
last_update	timestamp	not Null	

fact_ventas			
ventas_key	int(8)	not null	AUTO_INCREMENT,
ventas_id	int(11)	not null,	
date_key	int(8)	not Null	
salesperson_key	int(8)	not Null	
store_key	int(8)	not Null	
product_key	int(8)	not Null	
proveedor_key	int(8)	not Null	
territory_key	int(8)	not Null	
cliente_key	int(8)	not Null	
count_sales	int(11)	not Null	
SubTotal	decimal(19,4)	not Null	
TaxAmt	decimal(19,4)	not Null	
Unit_Price	decimal(19,4)	not Null	
UnitPriceDiscount	decimal(19,4)	not Null	
Line_Total	decimal(38,6)	not Null	
OrderQty	smallint(8)	not Null	
primary key (ventas_key), unique index ventas_id(ventas_id),			

- Construya un modelo entidad relación y/o una especificación de las tablas.



5. Proyecto de Curso: Codificación de la Bodega de Datos

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

5.1. Actividades

En grupos, realice las siguientes actividades:

- Codifique la Bodega de Datos usando un esquema en estrella

Solución:

```
# Creación del esquema de la bodega de datos

create database if not exists adw_dwh;
use adw_dwh;

-- Dimension de tiempo
create table if not exists dim_time(
    date_key          int(8)      not null,
    date_value        date        not null,
    month_number      int(3)      not null,
    year4             int(5)      not null,
    day_of_week       int(5)      not null,
    day_of_week_name  char(12)    not null,

    primary key(date_key),
    unique index date_value (date_value)
);

-- Dimension Vendedor
create table if not exists dim_vendedor(
    salesperson_key    int not null AUTO_INCREMENT,
    salesperson_id     int not null,
    salesperson_name    varchar(50) not null,
    city               varchar(30) not null,
    country            varchar(30) not null,
    gender             varchar(5)  not null,
    paymentfreq        tinyint(8) not null,
    department         varchar(30) not null,
    last_update        timestamp not null default CURRENT_TIMESTAMP,

    primary key (salesperson_key),
    unique index (salesperson_id)
);

-- Dimension Tienda
create table if not exists dim_tienda(
    store_key          int not null AUTO_INCREMENT,
    store_id           int not null,
    store              varchar(100) not null,
    last_update        timestamp not null default CURRENT_TIMESTAMP,

    primary key (store_key),
    unique index (store_id)
);

-- Dimension Producto
create table if not exists dim_producto(
```

```
product_key    int not null AUTO_INCREMENT,
product_id     int not null,
product        varchar(100) not null,
class          char(2),
category       varchar(100) not null,
subcategory     varchar(100) not null,
last_update    timestamp not null default CURRENT_TIMESTAMP,

primary key (product_key),
unique index (product_id)
);

-- Dimension Territorio
create table if not exists dim_territorio(
    territory_key    int not null AUTO_INCREMENT,
    territory_id     int not null,
    name_territory   varchar(100) not null,
    last_update      timestamp not null default CURRENT_TIMESTAMP,

    primary key (territory_key),
    unique index (territory_id)
);

-- Dimension Cliente
create table if not exists dim_cliente(
    cliente_key int not null AUTO_INCREMENT,
    cliente_id  smallint(8) not null,
    cliente     varchar(200) not null,
    city        varchar(50) not null,
    province    varchar(50) not null,
    country     varchar(50) not null,
    email       varchar(50) not null,
    last_update timestamp not null default CURRENT_TIMESTAMP,

    primary key (cliente_key),
    unique index (cliente_id)
);

-- Hechos de las ventas
create table if not exists fact_ventas(
    -- llaves
    ventas_key    int(8)    not null AUTO_INCREMENT,
    ventas_id     int(11)   not null,

    -- foreign keys
    date_key      int(8)    not null,
    salesperson_key int(8)   not null,
```

```

store_key      int(8)      not null,
product_key    int(8)      not null,
territory_key  int(8)      not null,
cliente_key    int(8)      not null,

-- datos
count_sales    int(11)     not null,
SubTotal       decimal(19,4) not null,
TaxAmt         decimal(19,4) not null,
Unit_Price     decimal(19,4) not null,
UnitPriceDiscount decimal(19,4) not null,
Line_Total     decimal(38,6) not null,
OrderQty       smallint(8)  not null,

primary key (ventas_key),
unique index ventas_id(ventas_id),

index date_key (date_key),
foreign key(date_key) references dim_time(date_key),
index salesperson_key (salesperson_key),
foreign key(salesperson_key) references dim_vendedor(salesperson_key),
index store_key (store_key),
foreign key (store_key) references dim_tienda(store_key),
index product_key (product_key),
foreign key(product_key) references dim_producto(product_key),
-- index proveedor_key (proveedor_key),
-- foreign key(proveedor_key) references dim_proveedor(proveedor_key),
index territory_key (territory_key),
foreign key(territory_key) references dim_territorio(territory_key),
index cliente_key (cliente_key),
foreign key (cliente_key) references dim_cliente(cliente_key)
);

```

6. Proyecto de Curso: ETL para las dimensiones

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

1.1. Actividades

En grupos, realice las siguientes actividades:

- **Codifique el ETL para la dimensión tiempo**
- **Codifique el ETL para las otras dimensiones**

1.2. Elementos mínimos del informe

```
2.      use adw;
3.
4.      insert into adw_dwh.dim_cliente (
5.          cliente_id,
6.          cliente,
7.          city,
8.          province,
9.          country,
10.         email
11.     )
12.     select
13.         Customer.PersonID as cliente_id,
14.         CONCAT(Person.FirstName, ' ', Person.LastName) AS cliente,
15.         Address.City as city,
16.         Prov.Name as province,
17.         Region.Name as country,
18.         Email.EmailAddress as email
19.
20.     from Sales_Customer as Customer
21.         join Person_Person as Person
22.             on Customer.PersonID = Person.BusinessEntityID
23.         join Person_EmailAddress as Email
24.             on Person.BusinessEntityID = Email.BusinessEntityID
25.         join Person_Address as Address
26.             on Person.BusinessEntityID = Address.AddressID
27.         join Person_StateProvince as Prov
28.             on Address.StateProvinceID = Prov.StateProvinceID
29.         join Person_CountryRegion as Region
30.             on Prov.CountryRegionCode = Region.CountryRegionCode
31.     ;
32.
```

```
33.     use adw;
34.
35.     insert into adw_dwh.dim_producto(
36.         product_id,
37.         product,
38.         localizacion,
39.         costrate,
40.         class,
41.         category,
42.         subcategory
43.     )
44.     select
45.         Product.ProductID as product_id,
46.         Product.Name as product,
47.         Location.Name as localizacion,
48.         Location.CostRate as costrate,
```

```
49.         Product.Class as class,
50.         Category.Name as category,
51.         Subcategory.Name as subcategory
52.
53.     from Production_Product as Product
54.         join Production_ProductSubcategory as Subcategory
55.             on Product.ProductSubcategoryID = Subcategory.ProductSubcategoryID
56.         join Production_ProductCategory as Category
57.             on Subcategory.ProductCategoryID = Category.ProductCategoryID
58.         join Production_ProductInventory as Inventory
59.             on Product.ProductID = Inventory.ProductID
60.         join Production_Location as Location
61.             on Inventory.LocationID = Location.LocationID
62.     ;
63.
```

```
64.     use adw;
65.
66.     INSERT INTO adw_dwh.dim_territorio (
67.         territory_id,
68.         name_territory
69.     )
70.     select
71.         TerritoryID,
72.         case when Name in ('Southwest','Northwest','Southeast','Central','Northeast') then
73. 'USA' else Name end as name_territory
74.     from Sales_SalesTerritory
75.     ;
```

```
76.     use adw;
77.     INSERT INTO adw_dwh.dim_tienda (
78.         store_id,
79.         store
80.     )
81.
82.     select
83.         BusinessEntityID,
84.         Name
85.     from Sales_Store
86.     ;
87.
88.
```

```
89.     use adw;
90.
91.     INSERT INTO adw_dwh.dim_time (
92.         date_key,
93.         date_value,
94.         month_number,
95.         year4,
96.         day_of_week,
```

```
97.         day_of_week_name
98.     )
99.     SELECT
100.         TO_DAYS(fecha) AS date_key,
101.         fecha,
102.         MONTH(fecha) AS month,
103.         YEAR(fecha) AS year4,
104.         DAY(fecha) as day,
105.         DATE_FORMAT(fecha, '%W')
106.     FROM (
107.         select distinct DATE(OrderDate) as fecha
108.     from Sales_SalesOrderHeader
109.     ) as fechas
110.
```

```
111.     use adw;
112.
113.     insert into adw_dwh.dim_vendedor (
114.         salesperson_id,
115.         salesperson_name,
116.         city,
117.         country,
118.         gender,
119.         paymentfreq,
120.         department
121.     )
122.     select
123.         BusinessEntityID as salesperson_id,
124.         CONCAT(person.FirstName, ' ', person.LastName) as salesperson_name,
125.         ad.City as city,
126.         cregion.Name as country,
127.         hr.gender as gender,
128.         hrpay.PayFrequency as paymentfreq,
129.         hrdept.Name as department
130.     from Sales_SalesPerson as sp
131.         join HumanResources_Employee as hr using (BusinessEntityID)
132.         join Person_Person as person using (BusinessEntityID)
133.         join Person_BusinessEntityAddress as badd using (BusinessEntityID)
134.         join Person_Address as ad using (AddressID)
135.         join Person_StateProvince as sprovd using (StateProvinceID)
136.         join Person_CountryRegion as cregion using (CountryRegionCode)
137.         join HumanResources_EmployeePayHistory as hrpay using (BusinessEntityID)
138.         join HumanResources_EmployeeDepartmentHistory as hrhist using (BusinessEntityID)
139.         join HumanResources_Department as hrdept using (DepartmentID)
140.     ;
```


7. Proyecto de Curso: ETL para los hechos

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

7.1 Actividades

En grupos, realice las siguientes actividades:

- **Codifique el ETL para las tablas de hechos**

7.2 Elementos mínimos del informe

```
use adw

insert into adw_dwh.fact_ventas (
ventas_id,
date_key,
salesperson_key,
store_key,
product_key,
territory_key,
cliente_key,
count_sales,
SubTotal,
TaxAmt,
Unit_Price,
UnitPriceDiscount,
Line_Total,
OrderQty
)
select
    detail.SalesOrderDetailID as detailorder,
    to_days(OrderDate) as date_key,
    salesperson_key,
    store_key,
    product_key,
    territory_key,
    cliente_key,
    1 as count_sales,
    SubTotal,
    TaxAmt,
    UnitPrice,
    UnitPriceDiscount,
```

```
LineTotal,  
OrderQty  
  
from Sales_SalesOrderDetail as detail  
  join Sales_SalesOrderHeader as sales  
    on detail.SalesOrderID = sales.SalesOrderID  
  join Sales_Customer as customer  
    on sales.CustomerID = customer.CustomerID  
  join Sales_Store as store  
    on customer.StoreID = store.BusinessEntityID  
  join adw_dwh.dim_tienda as dimt  
    on store.BusinessEntityID=dimt.store_id  
  join adw_dwh.dim_vendedor as dimv  
    on sales.SalesPersonID=dimv.salesperson_id  
  join adw_dwh.dim_producto as dimp  
    on detail.ProductID=dimp.product_id  
  join Sales_SalesTerritory as st  
    on sales.TerritoryID=st.TerritoryID  
  join adw_dwh.dim_territorio as dimty  
    on st.TerritoryID=dimty.territory_id  
  join adw_dwh.dim_cliente as dimcl  
    on sales.CustomerID=dimcl.cliente_id  
;
```

8. Proyecto de Curso: Consultas sobre la bodega de datos

El proyecto del curso busca desarrollar una bodega de datos para AdventureWorks, una multinacional dedicada a la venta de bicicletas.

8.1 Actividades

En grupos, realice las siguientes actividades:

- **Elabore consultas sobre la bodega de datos**

Respuesta al punto 8.1

```
use adw_dwh  
  
-- Ventas por País  
select name_territory as Pais,round(avg(SubTotal),0) as Pomedio_Ventas,  
sum(count_sales) as Num_Ventas
```

```
from fact_ventas as fv
    join dim_territorio as dimt
        on dimt.territory_key=fv.territory_key
group by name_territory
order by Pomedio_Ventas desc

;

-- Top 10 Vendedores
select salesperson_name as Vendedor, sum(count_sales) as Num_Ventas
from fact_ventas as fv
    join dim_vendedor as dimv
        on dimv.salesperson_key=fv.salesperson_key

group by salesperson_name
order by Num_Ventas desc
limit 10

;

-- Top 5 de Tiendas con mayores ventas
select store as Tienda, round(sum(SubTotal),0) as Ventas, sum(count_sales) as
Num_Ventas
from fact_ventas as fv
    join dim_tienda as dimtienda
        on dimtienda.store_key=fv.store_key
group by store
order by Ventas desc
limit 5

;

-- Promedio Compras por Genero
select case when gender='M' then 'Hombres'
            when gender='F' then 'Mujeres'
            else gender end as Genero,
        round(avg(SubTotal),0) as Promedio_Ventas
from fact_ventas as fv
    join dim_vendedor as dimv
        on dimv.salesperson_key=fv.salesperson_key

group by genero

;

-- Ventas por Año
select year4 as Anno, sum(count_sales) as Num_Ventas
from fact_ventas as fv
    join dim_time as dimtime
```

```
        on fv.date_key=dimtime.date_key
group by year4

;

-- Variación ventas por año
with datos as
(select sum(case when year4=2011 then count_sales else 0 end) as ventas11,
       sum(case when year4=2012 then count_sales else 0 end) as ventas12,
       sum(case when year4=2013 then count_sales else 0 end) as ventas13,
       sum(case when year4=2014 then count_sales else 0 end) as ventas14
from fact_ventas as fv
join dim_time as dimtime
  on fv.date_key=dimtime.date_key
)
select ((ventas12 - ventas11)/ventas11)*100 as varventas12,
       ((ventas13 - ventas12)/ventas12)*100 as varventas13,
       ((ventas14 - ventas13)/ventas13)*100 as varventas14
from datos

;

-- Impuesto promedio por Subcategoría de producto
select Subcategory as Subcategoría, avg (TaxAmt) as Promedio_Impuesto
from fact_ventas as fv
  join dim_producto as dimp
    on fv.product_key=dimp.product_key
group by Subcategory
order by Promedio_Impuesto desc
```

8.2 Elementos mínimos del informe

Título: AdventureWorksGJD

Grupo: David Ricardo Guzmán, Jorge Armando Arévalo, Gabriela Cortes

Repositorio:

[adw-proyectoGJD/scripts/Consultas_DWH.sql at main · drguzman18/adw-proyectoGJD \(github.com\)](https://github.com/drguzman18/adw-proyectoGJD/tree/main/scripts/Consultas_DWH.sql)

1. Principales problemas:

Respuesta: Uno de los principales inconvenientes durante el proceso de creación del proyecto se presentó en el levantamiento del proceso ETL principal de

fact_ventas, teniendo en cuenta que el orden si fue determinante en la creación del query para realizar el proceso, ya que teníamos varias formas de realizar el proceso, pero generaba duplicidad en los registros, por lo cual presentaba errores.

2. Principales Aprendizajes y Conclusiones:

Fue muy interesante el proceso de creación de ETL, análisis de bodegas de datos, y los tipos de conexiones que existen. Adicionalmente, comprender que, en este caso, el orden de los factores sí altera los resultados, pues dependiendo de esto, se evitan duplicidades en la información y errores en el cargue de la información en la bodega de datos.

Inclusive sirvió para entender las conexiones con las distintas alternativas para la visualización de datos como PowerBI, realizando algunos análisis preliminares a través de un dashboard como se muestra a continuación:

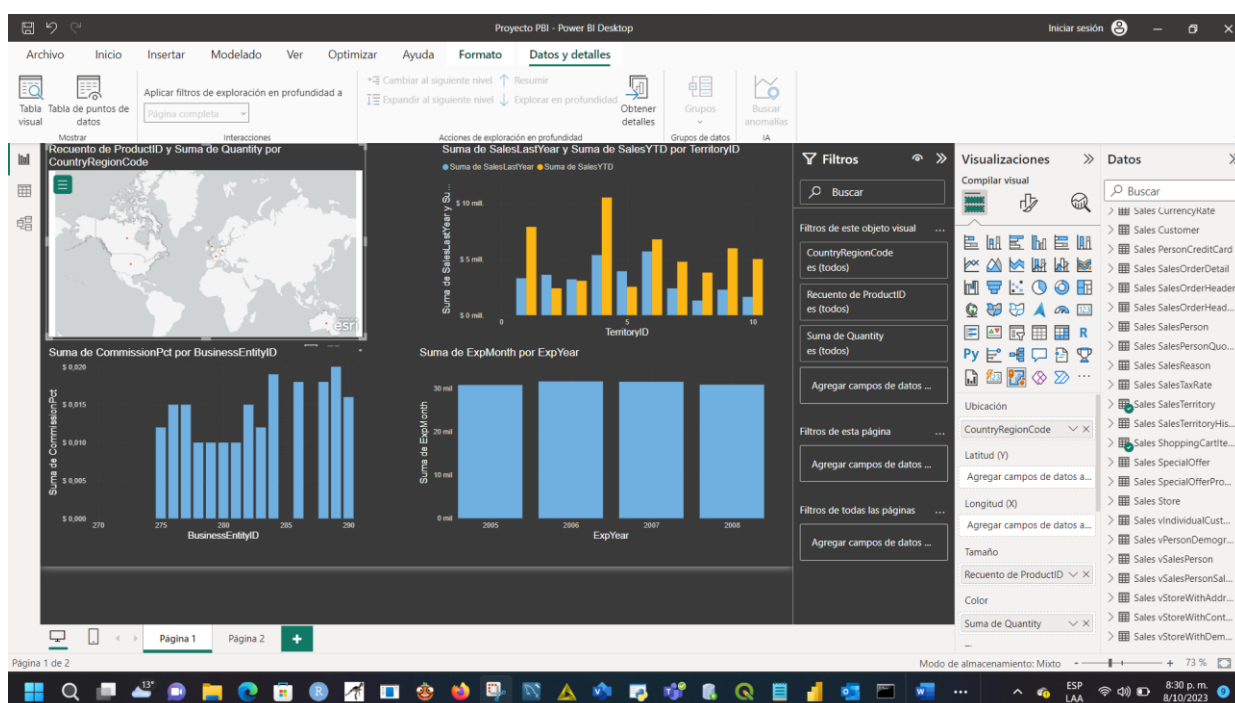


Ilustración 2. Visualización preliminar realizada sobre la base de AdventureWorks y con la DWH ejecutada.