

Universidad de Mendoza
Diseño de base de datos II



Business Intelligence

Montaña Agustín
Sanchez Mariano
Orbelli Bruno
Sarmiento Mauro

Índice de contenidos

Índice de contenidos	2
Descripción del proyecto	3
Estructura de base de datos	4
Requerimientos de negocio	5
Creación del warehouse	5
Proceso ETL	6
Cubo OLAP	10
Proyecciones visuales con Power BI	12

Descripción del proyecto

El concepto **inteligencia de negocios** hace referencia al uso de diversas estrategias y herramientas que permiten transformar la información con la que cuenta una organización en conocimiento, buscando la mejoría de su proceso de toma de decisiones. Con este fin, en *Business Intelligence* (BI), se recolectan datos operacionales del negocio y se los reduce de forma tal que puedan emplearse para analizar el comportamiento de la organización y ofrezcan una manera visual de entender su estado actual.

El propósito del siguiente proyecto es desarrollar un proceso de inteligencia de negocios sobre la base de datos de una tienda de insumos informáticos, posibilitando el acceso a estadísticas y datos de interés para la empresa respecto a las conductas de los compradores, requerimientos de insumos, ventas y otras áreas. Se prevé la creación de un *Data Warehouse*, dentro del cual se almacenarán los datos de la base original, para luego construir con ellos un cubo OLAP.

Para ello, se hará uso de las herramientas aquí mencionadas:

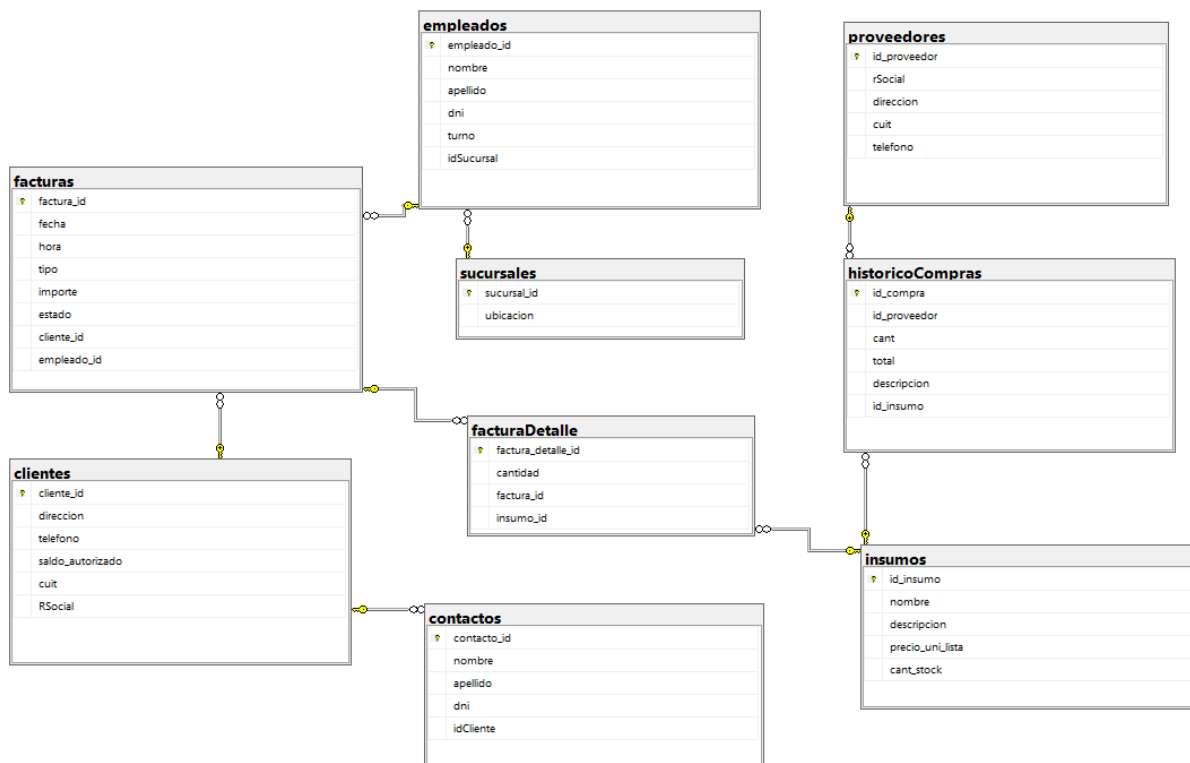
- SQL Server Developer Edition y SQL Server Manager + Integration and Analysis Services (manejo de base de datos y creación del *warehouse*).
- Visual Studio (proceso ETL y construcción del cubo OLAP).
- Microsoft PowerBI (desarrollo de gráficas y proyecciones de inteligencia de negocio).

En términos generales, la creación e implementación del proyecto se presentará a partir de una secuencia de pasos:

1. Creación del *DW* a partir de la base de datos original y los requerimientos del negocio.
2. Población del *DW* mediante flujos de datos y transformación de los mismos (ETL).
3. Creación de las dimensiones y de los cubos OLAP.
4. Desarrollo de proyecciones visuales.

Estructura de base de datos

La base de datos de la tienda de insumos responde a la siguiente estructura:



- Empleados: registro de los empleados de la organización y las sucursales en las que trabajan.
- Proveedores: registro de los proveedores de insumos.
- Facturas: registro de las facturas emitidas, el empleado que realizó la factura y el cliente al que fue emitida.
- Sucursales: registro de las sucursales de la compañía.
- HistoricoCompras: registro de las compras de insumos realizadas a cada proveedor.
- FacturaDetalle: detalles de cada una de las facturas emitidas.
- Clientes: registro de los clientes de la tienda.
- Contactos: registro de los representantes de cada cliente.
- Insumos: registro de los productos ofrecidos con su respectivo precio.

El rango de datos utilizado para rellenarla comprende transacciones entre marzo del 2020 y noviembre del 2022.

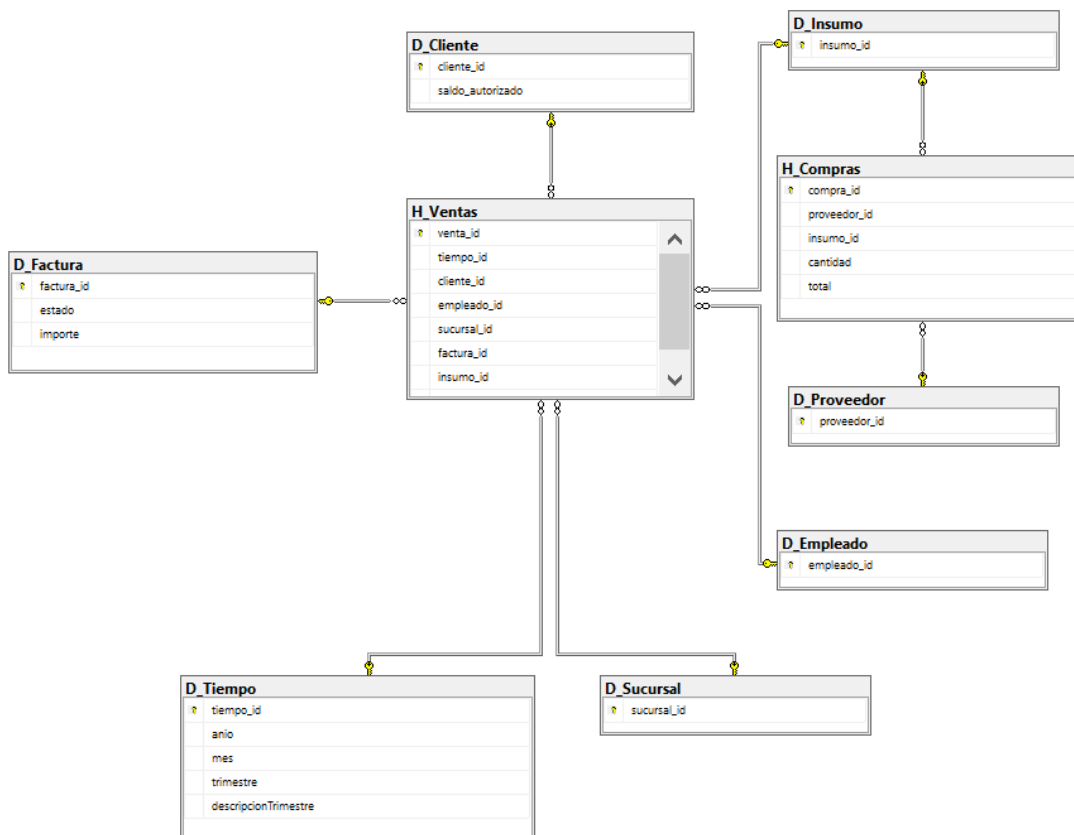
Requerimientos de negocio

El negocio plantea los siguientes requerimientos de información:

- Empleados con mayor recaudación en ventas, mayor cantidad de facturas emitidas e ingreso promedio por factura.
- Ventas mensuales, trimestrales y anuales.
- Cant. de productos vendidos y productos más vendidos en cada sucursal.
- Precio promedio de cada insumo a lo largo de los diferentes proveedores y proveedor con el menor precio promedio para cada insumo.
- Registro de "mejores clientes": clientes ordenados en función de las facturas adeudadas y monto gastado.

Creación del warehouse

En función de los requerimientos previamente planteados, se crea el *DW* con las siguientes dimensiones y hechos:

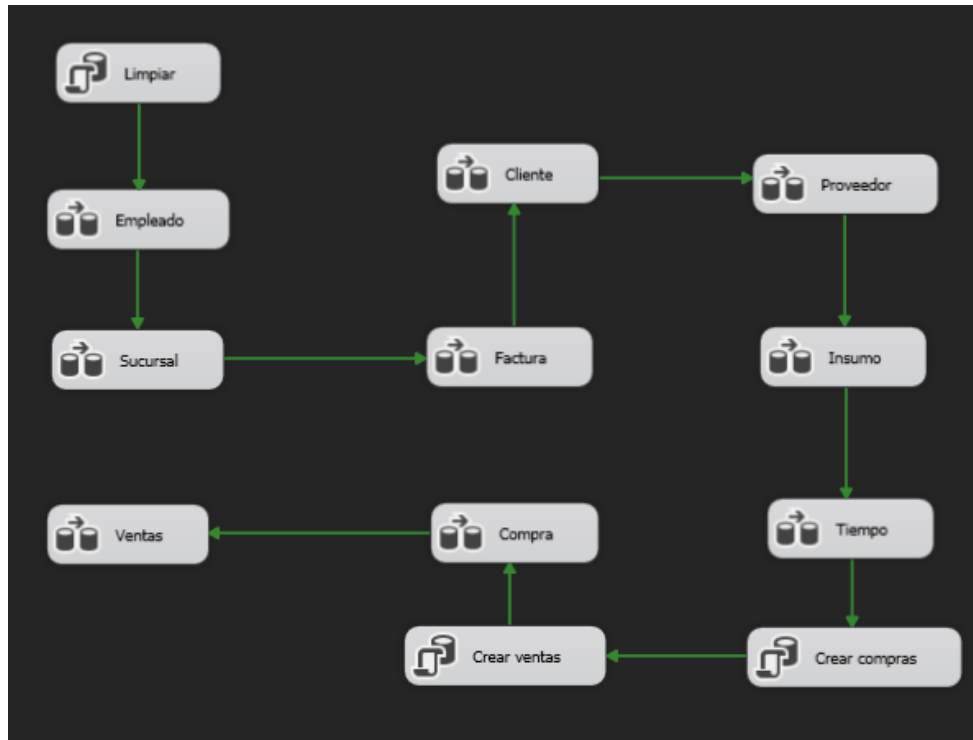


- D_Insumo: dimensión de los insumos involucrados en las compras a proveedores y ventas a clientes (*insumo_id*).
- D_Cliente: dimensión de los clientes involucrados en la venta de productos (*cliente_id*, *saldo_autorizado*).
- D_Facutra: dimensión de las facturas emitidas en cada venta (*factura_id*, *estado*, *importe*).
- D_Proveedor: dimensión de los proveedores involucrados en las compras de insumos (*proveedor_id*).
- D_Empleado: dimensión de los empleados de cada sucursal involucrados en las diferentes ventas (*empleado_id*).
- D_Sucursal: dimensión de las sucursales involucradas en las ventas (*sucursal_id*).
- D_Tiempo: dimensión de tiempo (*tiempo_id*, *anio*, *mes*, *trimestre*, *descripcionTrimestre*).
- H_Compras: tabla de hechos para las compras realizadas a los distintos proveedores (*compra_id*, *proveedor_id*, *insumo_id*, *cantidad*, *total*). Existe una relación 1 a 1 entre los hechos de esta tabla y los registros de HistoricoCompras.
- H_Ventas: tabla de hechos para las ventas realizadas a los distintos clientes (*venta_id*, *tiempo_id*, *cliente_id*, *empleado_id*, *sucursal_id*, *factura_id*, *insumo_id*, *cantidad*). Existe una relación 1 a 1 entre los hechos de esta tabla y los registros de FacturaDetalle.

Cabe aclarar que, considerando que las compras de insumos a proveedores y las ventas a clientes constituyen transacciones independientes, se decidió tratarlos como hechos separados, estableciendo un modelo estrella para cada uno de ellos.

Proceso ETL

Creando un proyecto 'Integration Services' en Visual Studio, se lleva a cabo el flujo y transformación de los datos de la base original al *warehouse* de la siguiente manera:



- Scripts

- Limpiar - elimina las tablas de hechos (ventas y compras) y trunca las tablas de dimensiones, vaciando cualquier entrada previa/accidental en el *warehouse*.
- Crear compras y crear ventas - crea las dos tablas de hechos, con sus respectivas claves foráneas.

- Flujos de datos

- Empleado, sucursal, factura, (...) - realiza el flujo de información hacia la dimensión indicada, mapeando únicamente los campos necesarios (indicados en el diagrama del *warehouse*).
- Compras y ventas - realiza el flujo de información hacia las tablas de hechos, mapeando las claves foráneas con las claves primarias de las diferentes entidades.

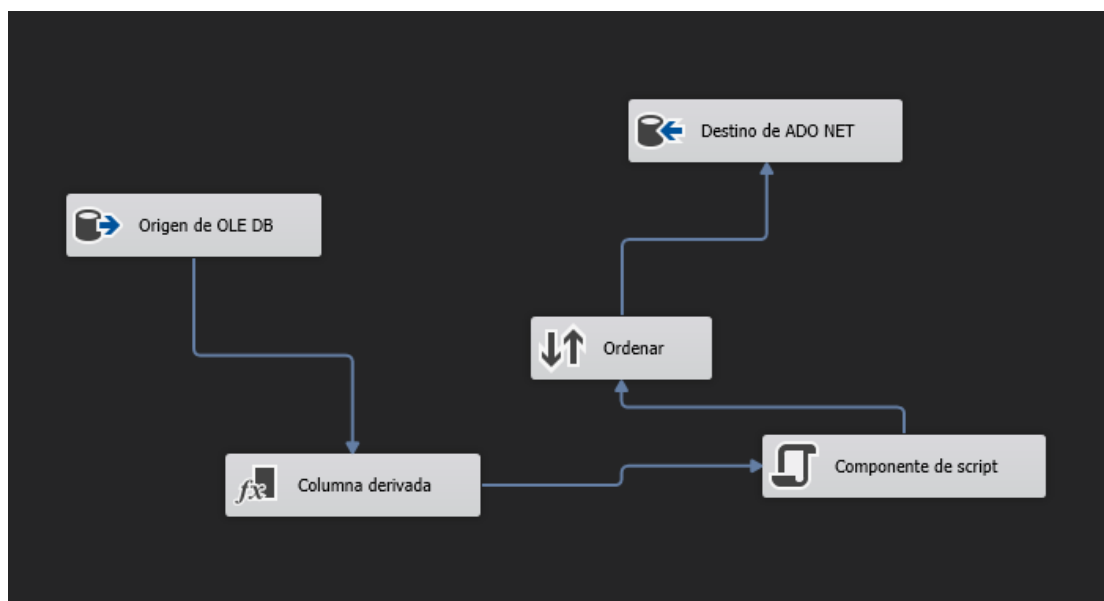
* Para formar correctamente la dimensión tiempo, se hizo uso de un script que permite obtener el nombre (en texto) de un trimestre a partir del número de *quarter* del año.

```
public override void Entrada0_ProcessInputRow(Entrada0Buffer Row)
{
    Console.WriteLine(Row.identificador);

    switch (Row.quarter) {

        case 1:
            Row.DescrpcionTrimestre = "Primer Trimestre";
            break;
        case 2:
            Row.DescrpcionTrimestre = "Segundo Trimestre";
            break;
        case 3:
            Row.DescrpcionTrimestre = "Tercer Trimestre";
            break;
        case 4:
            Row.DescrpcionTrimestre = "Cuarto Trimestre";
            break;
    }
}
```

El identificador de cada fecha se obtiene creando una columna derivada donde se combinan el año, mes y día en una única clave, permitiendo, a partir de cualquier fecha dada, obtener su id en la dimensión tiempo.



Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

+

 Variables y parámetros

+

 Columns

+

 Funciones matemáticas

+

 Funciones de cadena

+

 Funciones de fecha y hora

+

 Funciones NULL

+

 Conversiones de tipo

+

 Operadores

Descripción:

Derived Column Name	Derived Column	Expression	Data Type	Le
identificador	<add as new column>	(DT_I4)((DT_STR,4,1252)DATEPART("year",fecha) + (...)	entero de cuatro bytes ...	
day	<add as new column>	DATEPART("day",fecha)	entero de cuatro bytes ...	
month	<add as new column>	DATEPART("month",fecha)	entero de cuatro bytes ...	
year	<add as new column>	DATEPART("year",fecha)	entero de cuatro bytes ...	
quarter	<add as new column>	DATEPART("quarter",fecha)	entero de cuatro bytes ...	

Configure Error Output...

OK

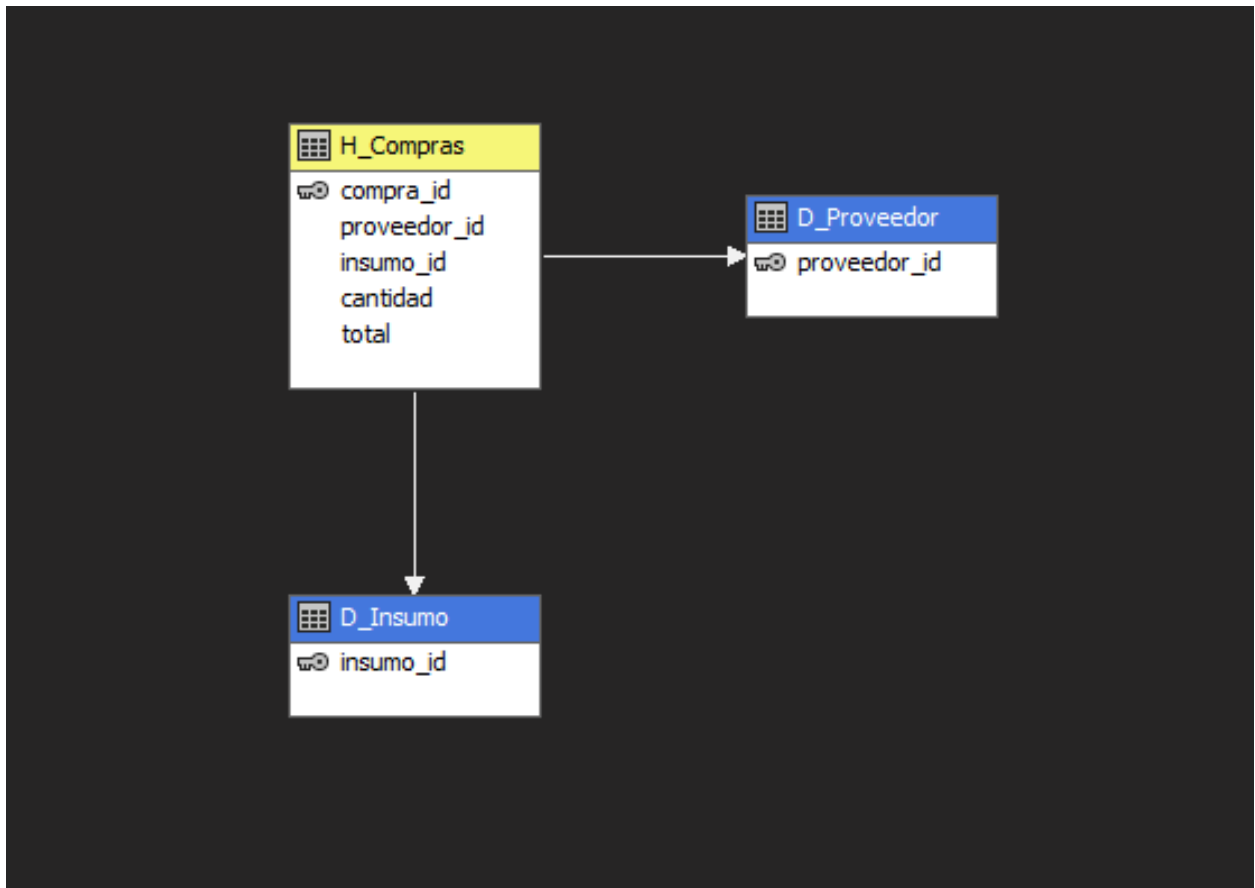
Cancel

Ayuda

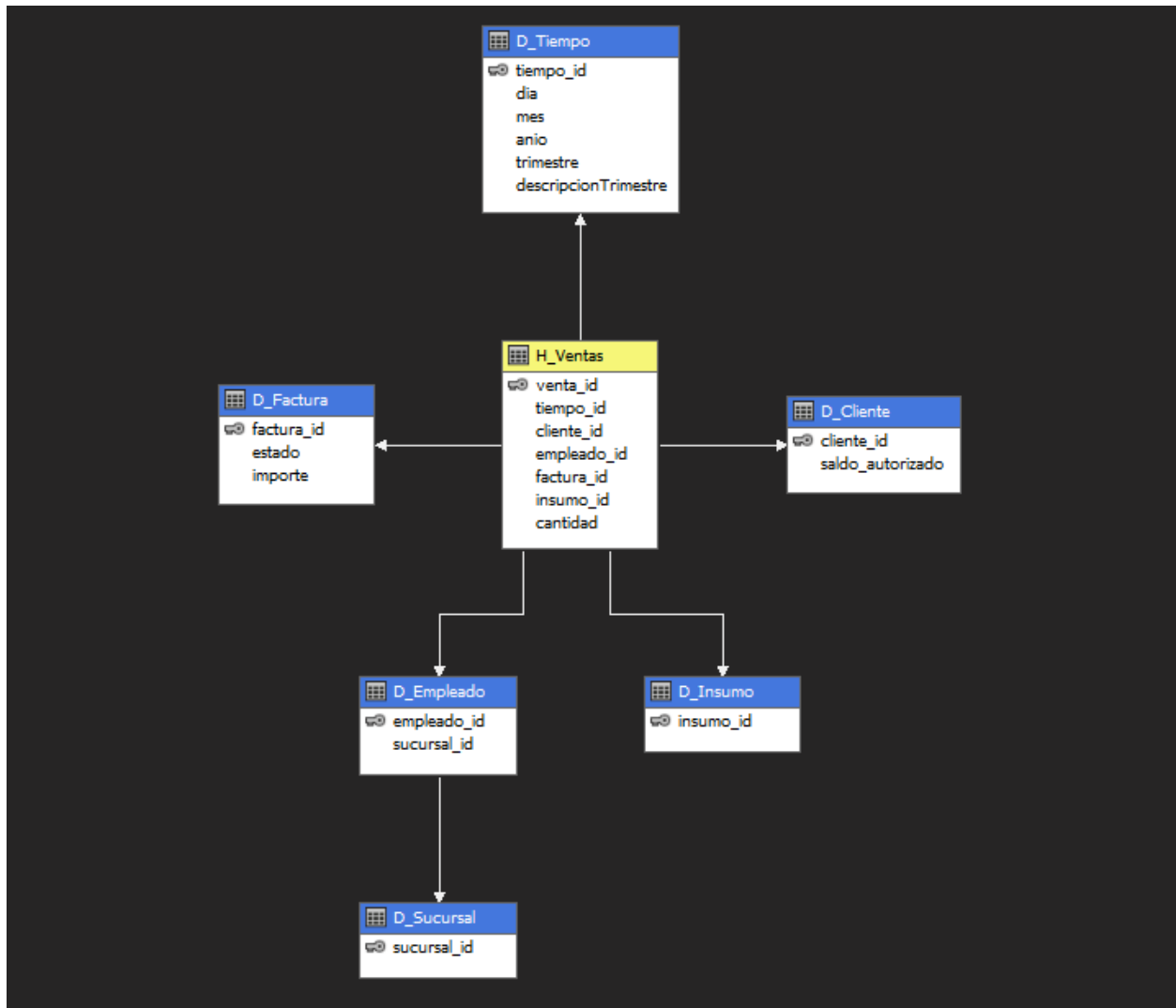
Cubo OLAP

Creando un proyecto 'multidimensional y minería de datos en Analysis Services' en Visual Studio, se lleva a cabo el establecimiento de las dimensiones y la creación de dos cubos OLAP (uno para cada una de las tablas de hechos desarrolladas). Esto nos permitirá tener una base de datos más optimizada (desnormalizada) para poder utilizar desde la aplicación *PowerBi*.

La forma final de los cubos responde a los siguientes diagramas:



Cubo compras



Cubo ventas

- Dimensiones:

Las dimensiones en *Analysis Services* simplifican a las tablas originales dentro del *DW*. Se especifican los campos que se van a utilizar y se lleva a cabo el mapeo en base a las claves foráneas existentes.

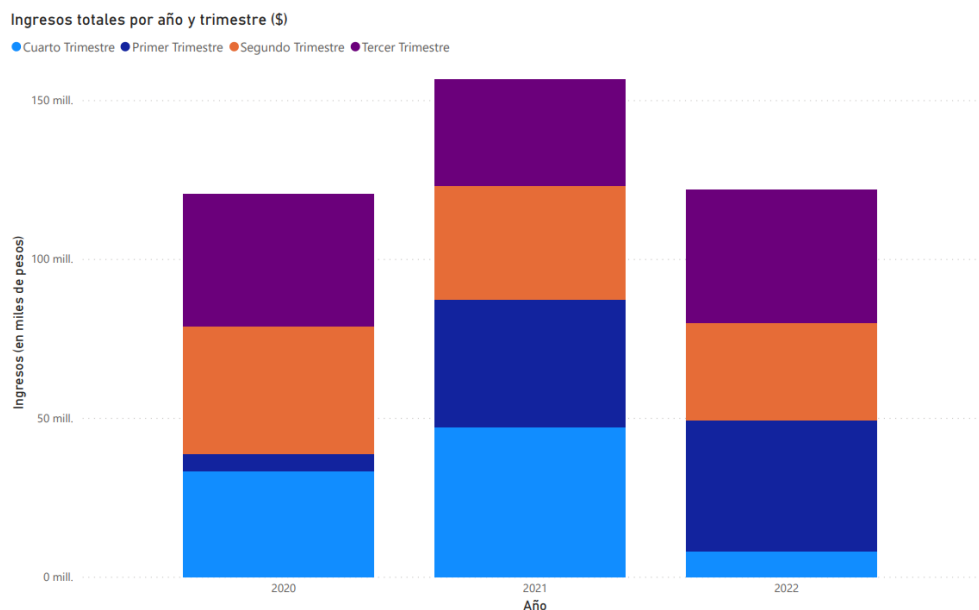


Proyecciones visuales con Power BI

Una vez concluida la elaboración y el procesamiento de los cubos OLAP, se los importa dentro de PowerBI para elaborar los reportes gráficos correspondientes. Para ello, se aplica a las tablas de hechos desnormalizadas las operaciones de agregación, agrupación y transformación necesaria.

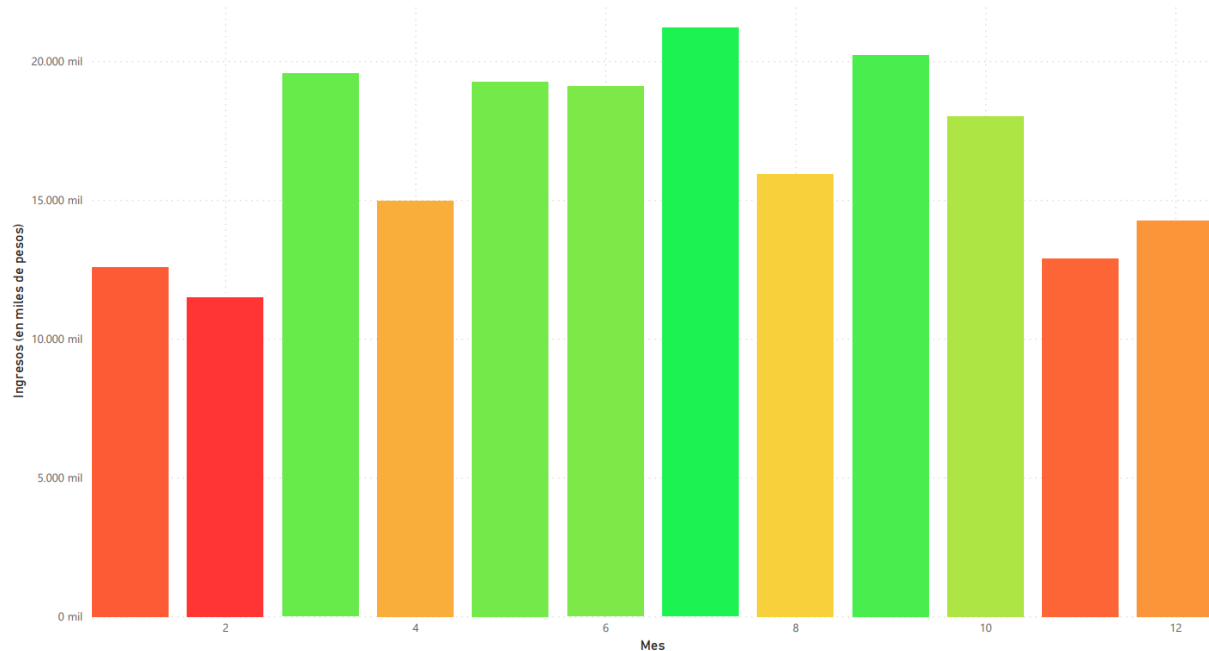
De esa forma, se obtienen los resultados de los requisitos de negocio:

- Ingresos totales por año y trimestre (\$)



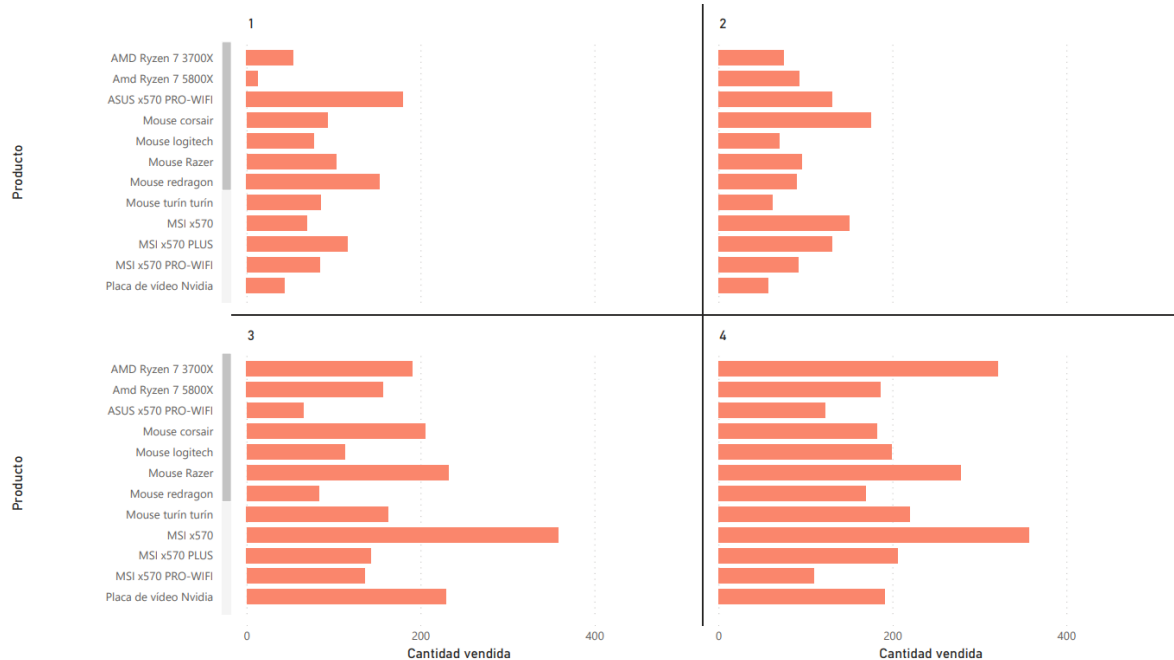
- Ingresos totales por mes (\$)

Ingresos totales por mes (\$)



- Venta de productos totales por sucursal

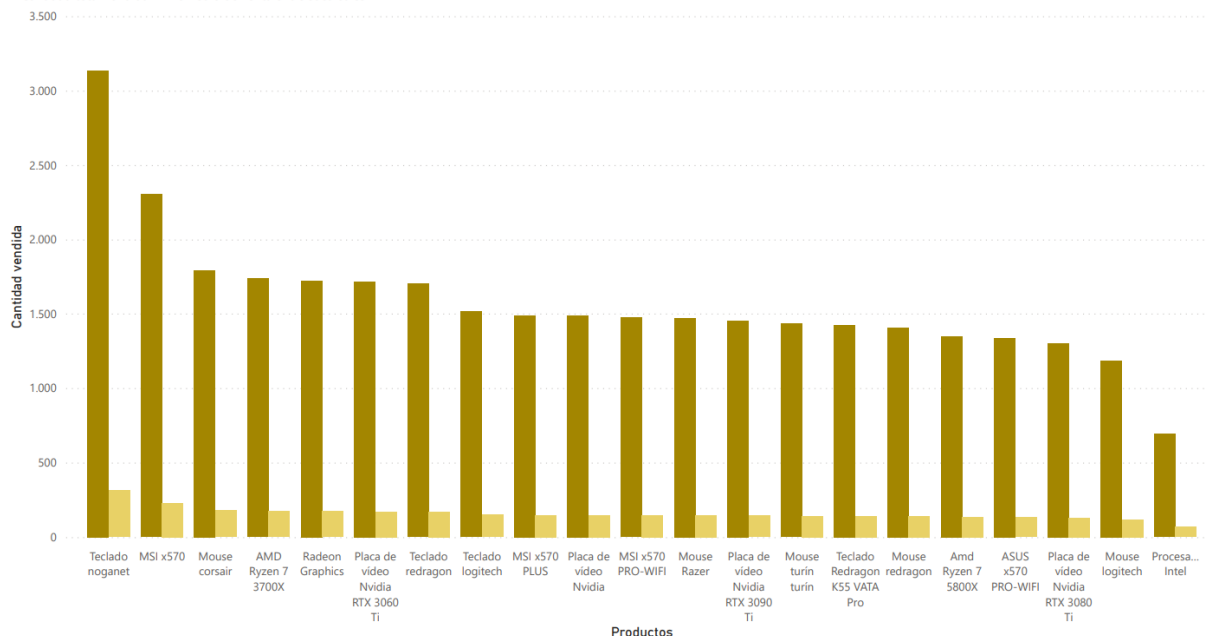
Venta de productos por sucursal



- Cant. total vendida de cada prod. y promedio de venta entre sucursales

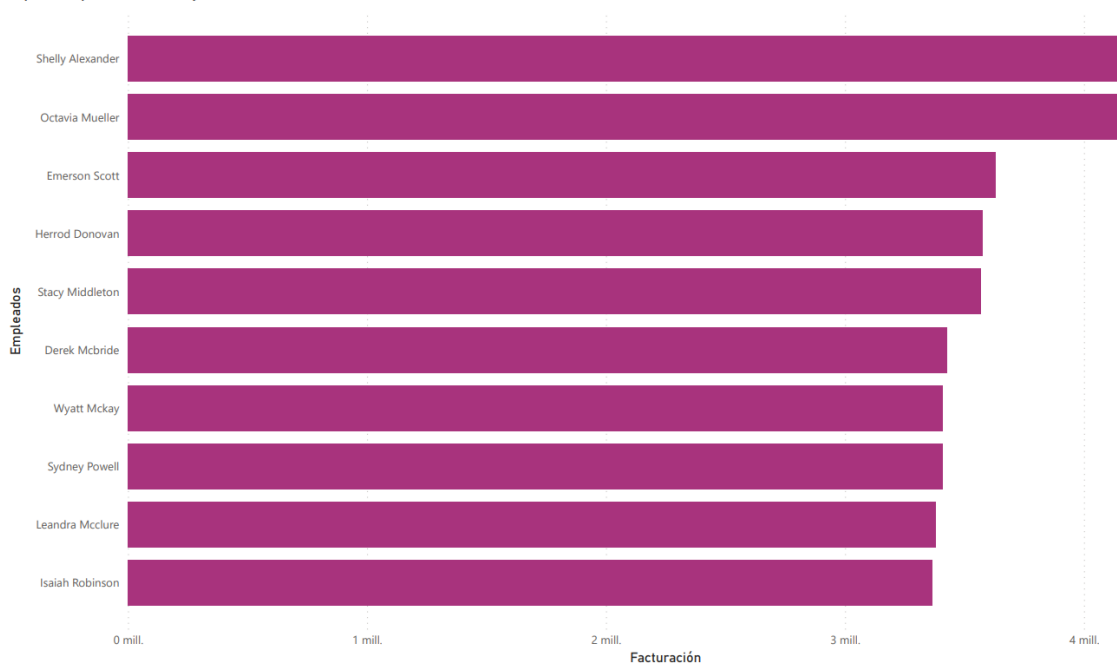
Cantidad total vendida y promedio de venta entre sucursales

● Cantidad total vendida ● Promedio de venta entre sucursales



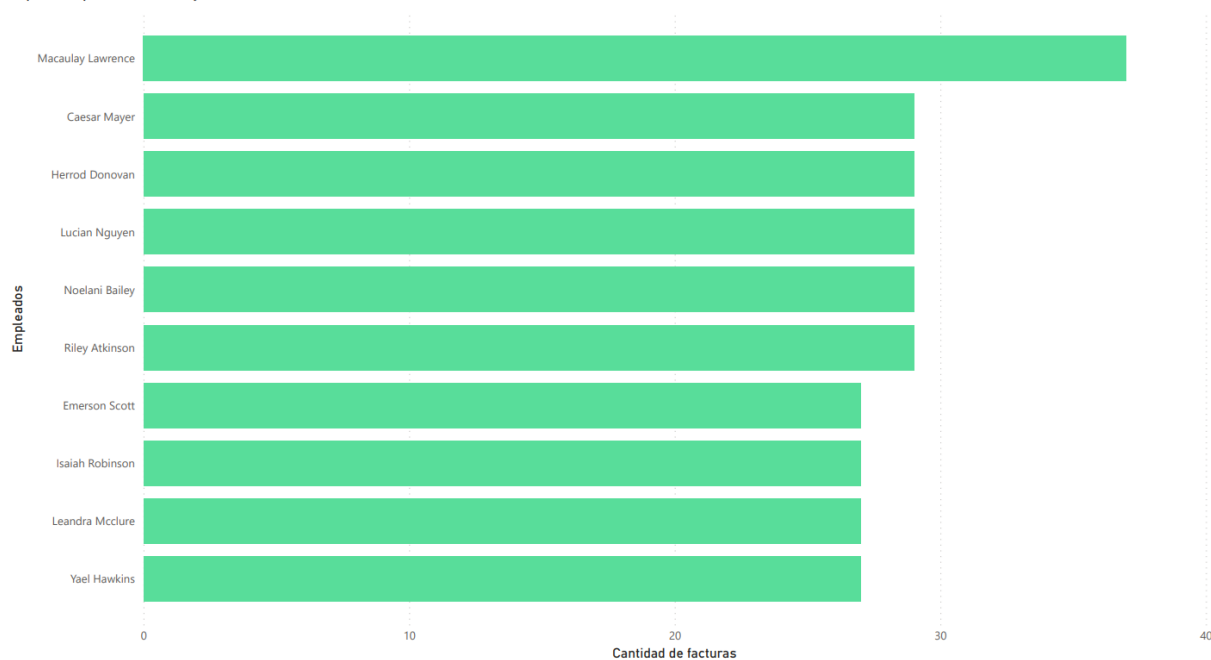
- Top 10 empleados con mayor facturación

Top 10 empleados con mayor facturación



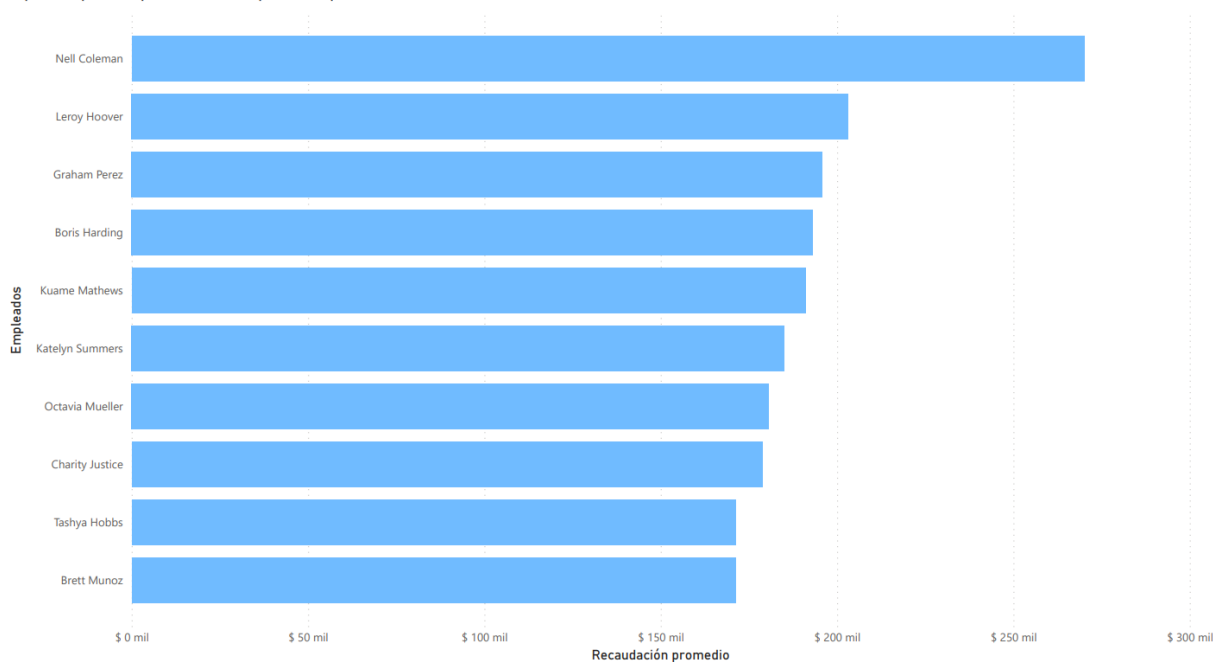
- Top 10 empleados con mayor cantidad de facturas

Top 10 empleados con mayor cantidad de facturas



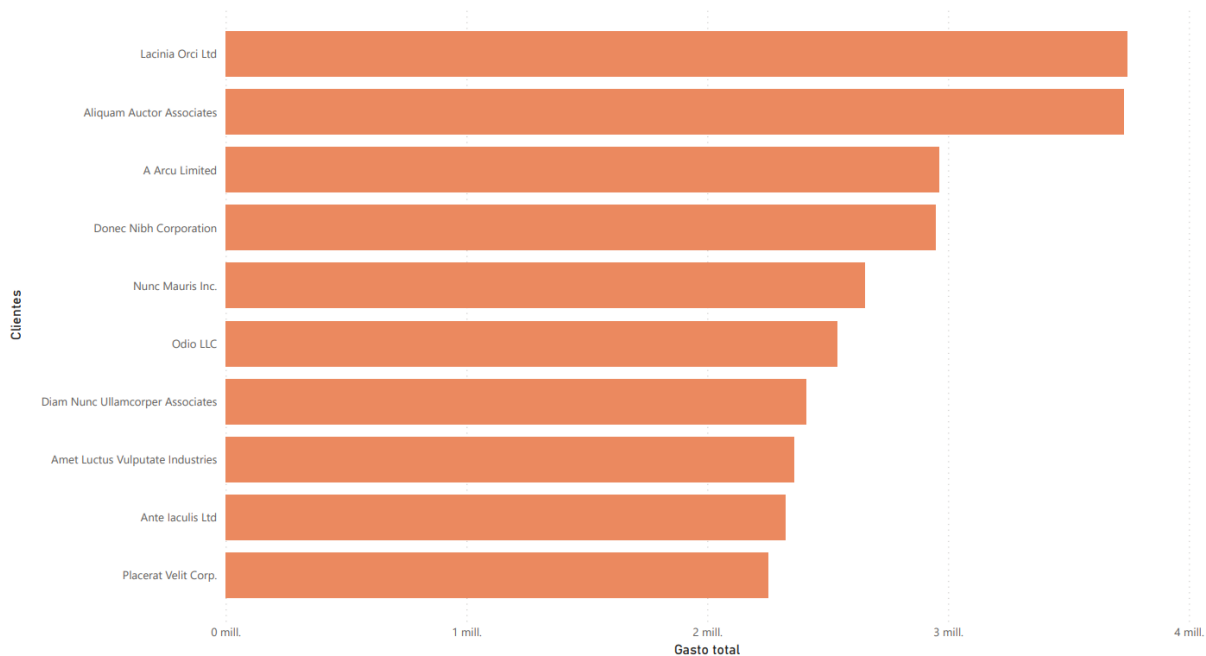
- Top 10 empleados por recaudación promedio por factura

Top 10 empleados por recaudación promedio por factura



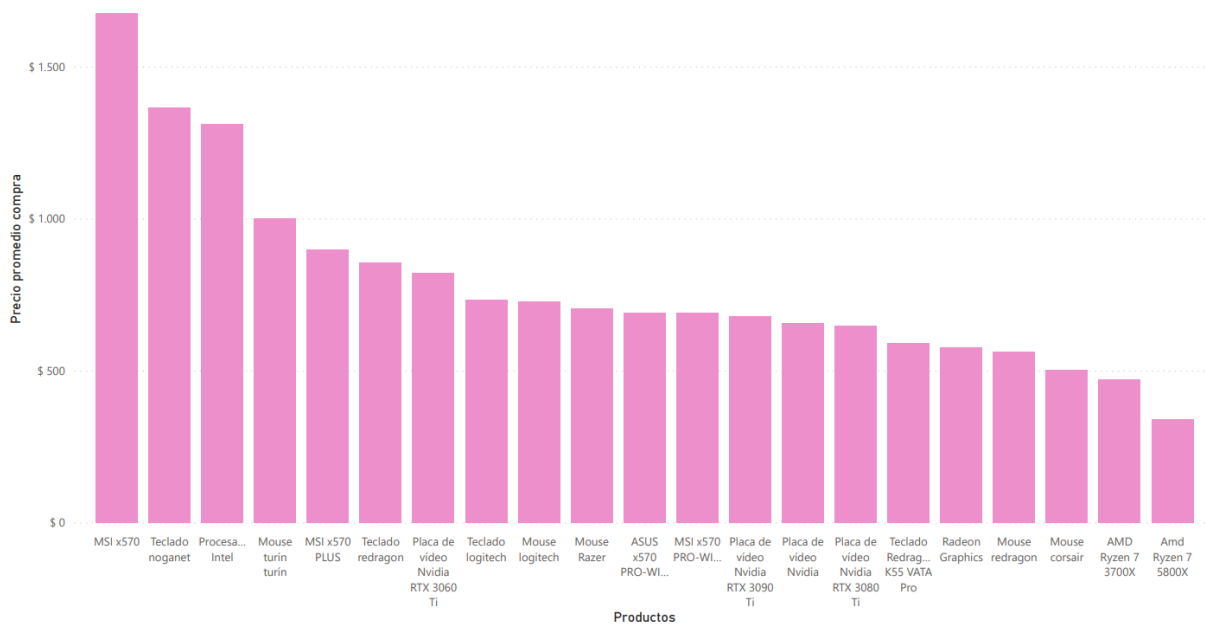
- Top 10 clientes con mayor monto de compra

Top 10 clientes con mayor monto de compra



- Precio promedio de insumos entre proveedores

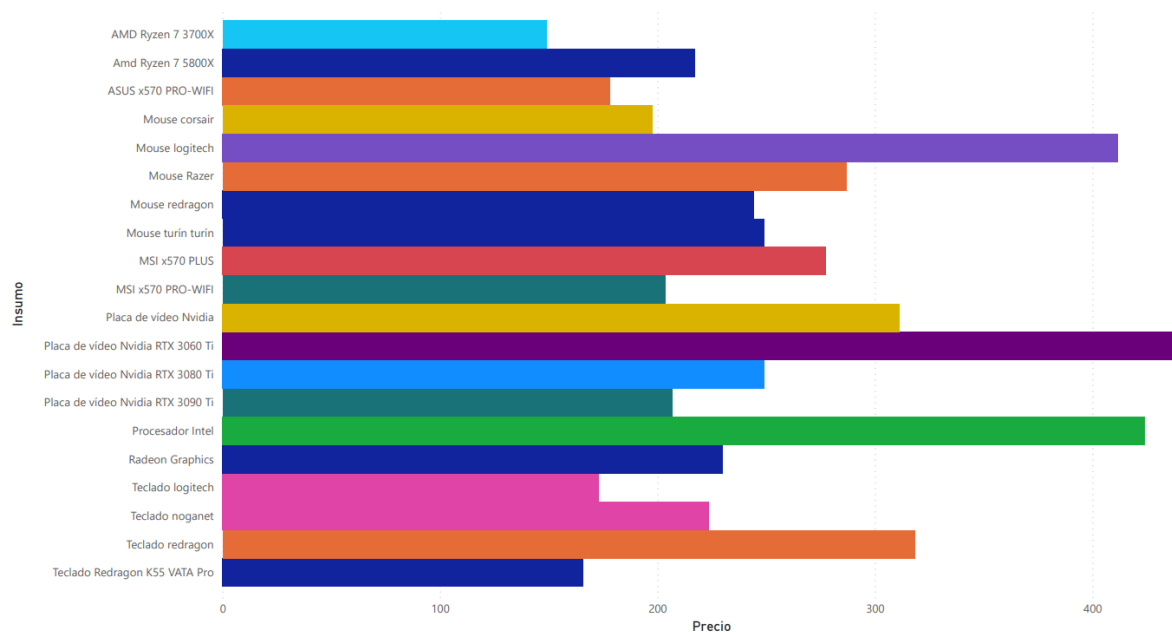
Precio promedio de insumos entre proveedores



- Precio promedio mínimo de insumos

Precio promedio mínimo de insumos

Proveedor 2 3 4 5 7 8 9 10 11 16 17



- Proporción de facturas pagadas en base al cliente con mayor monto de compra

Proporción de facturas pagadas

