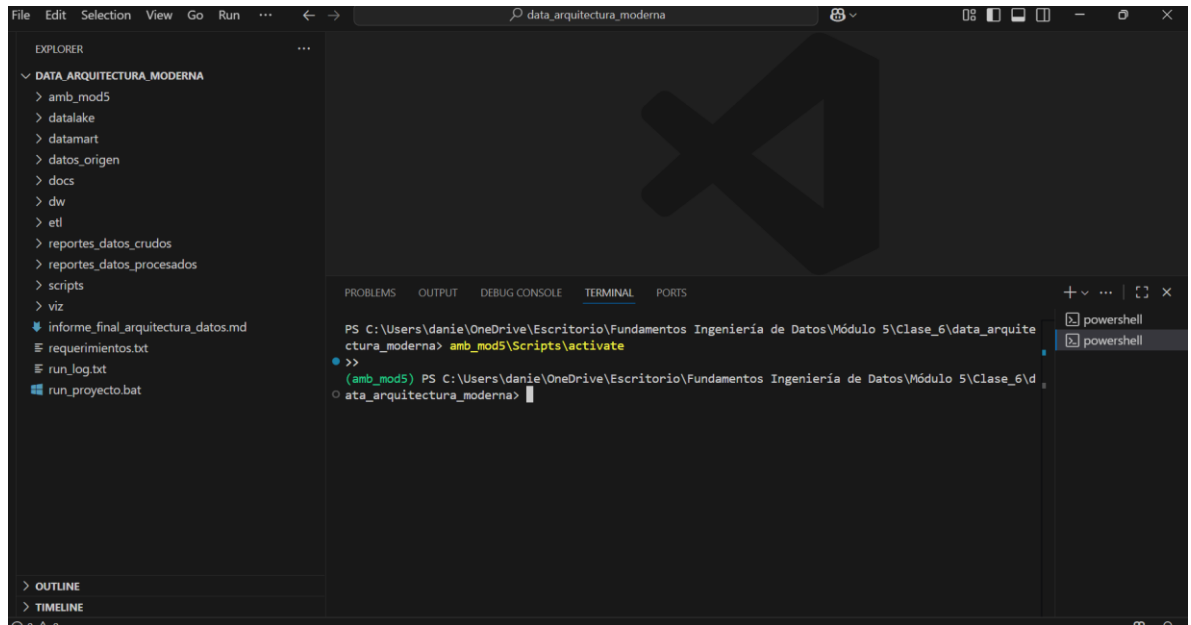


Daniela Méndez Gándara

Evidencia de ejecución arquitectura de datos en Visual Studio Code

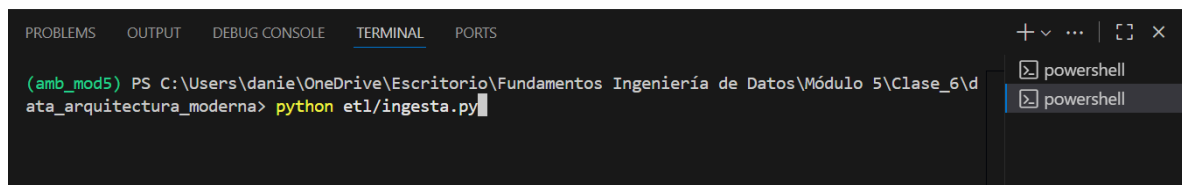
1. Activar entorno virtual en carpeta raíz
amb_mod5\Scripts\activate



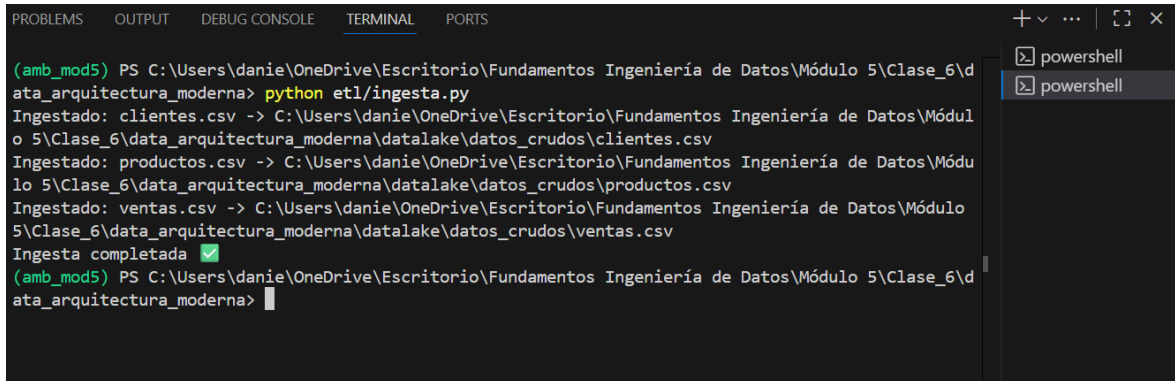
2. Instalar pandas

```
>>>
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> pip install pandas
Requirement already satisfied: pandas in c:\users\danie\onedrive\escritorio\fundamentos ingeniería de datos\módulo 5\clase_6\data_arquitectura_moderna\amb_mod5\lib\site-packages (2.3.1)
Requirement already satisfied: numpy>=1.26.0 in c:\users\danie\onedrive\escritorio\fundamentos ingeniería de datos\módulo 5\clase_6\data_arquitectura_moderna\amb_mod5\lib\site-packages (from pandas)
```

3. Ejecutar ingesta.py de carpeta etl

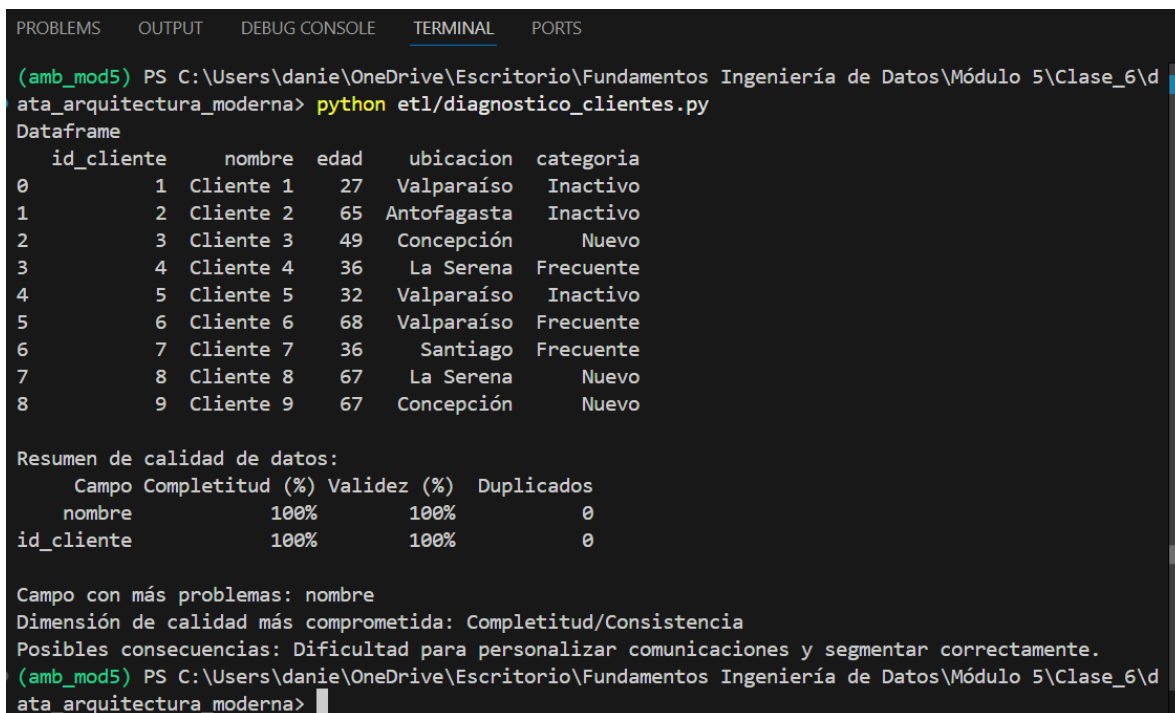


The screenshot shows the Visual Studio Code Terminal view. The command prompt is running 'PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl\ingesta.py'. There are two 'powershell' tabs open on the right.



```
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl/ingesta.py
Ingestado: clientes.csv -> C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\datalake\datos_crudos\clientes.csv
Ingestado: productos.csv -> C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\datalake\datos_crudos\productos.csv
Ingestado: ventas.csv -> C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\datalake\datos_crudos\ventas.csv
Ingesta completada ✓
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

4. Ejecutar diagnóstico de datos crudos de la carpeta datos_crudos mediante `diagnostico_clientes.py`, `diagnostico_productos.py` y `diagnostico_ventas.py` en carpeta `etl`



```
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl/diagnostico_clientes.py
Dataframe
  id_cliente  nombre  edad  ubicacion  categoria
0           1  Cliente 1    27  Valparaíso  Inactivo
1           2  Cliente 2    65  Antofagasta  Inactivo
2           3  Cliente 3    49  Concepción   Nuevo
3           4  Cliente 4    36  La Serena  Frecuente
4           5  Cliente 5    32  Valparaíso  Inactivo
5           6  Cliente 6    68  Valparaíso  Frecuente
6           7  Cliente 7    36  Santiago   Frecuente
7           8  Cliente 8    67  La Serena   Nuevo
8           9  Cliente 9    67  Concepción   Nuevo

Resumen de calidad de datos:
  Campo  Completitud (%)  Validez (%)  Duplicados
nombre              100%          100%           0
id_cliente          100%          100%           0

Campo con más problemas: nombre
Dimensión de calidad más comprometida: Completitud/Consistencia
Posibles consecuencias: Dificultad para personalizar comunicaciones y segmentar correctamente.
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_architectura_moderna> python etl/diagnostico_productos.py

Diagnóstico productos.csv
  Campo Completitud (%) Validez (%) Duplicados
  id_producto          100%      100%         0
nombre_producto        100%      100%         0
  categoria            100%      100%         6
  proveedor            100%      100%         7

Duplicados de filas completas: 0

Campo con más problemas: id_producto
Dimensión afectada: Integridad/Unicidad de claves
Impacto potencial: IDs inválidos o duplicados afectan joins y el DW.
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_architectura_moderna>
```

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_architectura_moderna> python etl/diagnostico_ventas.py

Dataframe
  id_venta id_producto id_sucursal fecha cantidad monto
0         1           9           3 2023-01-01      12 22557
1         2           2           1 2023-01-02      10 24278
2         3           6           4 2023-01-03      17 39119
3         4           6           4 2023-01-04       7 17835
4         5           5           2 2023-01-05       7 49733
5         6           6           4 2023-01-06      17 7426
6         7           5           3 2023-01-07      16 10024
7         8           7           3 2023-01-08       7 20807
8         9           2           2 2023-01-09       2 43804

Diagnóstico ventas.csv
  Campo Completitud (%) Validez (%) Duplicados
  id_venta          100%      100%         0
id_producto         100%      100%        90
id_sucursal         100%      100%        95
  fecha            100%      100%         0
  cantidad         100%      100%        81
  monto            100%      100%         0

Duplicados de filas completas: 0

Campo con más problemas: id_venta
Dimensión afectada: Integridad/Unicidad de claves
Impacto potencial: Ventas duplicadas o inválidas distorsionan todas las métricas.
```

5. Ejecutar limpieza_*.py para limpiar datos crudos y generar datos procesados

Daniela Méndez Gándara

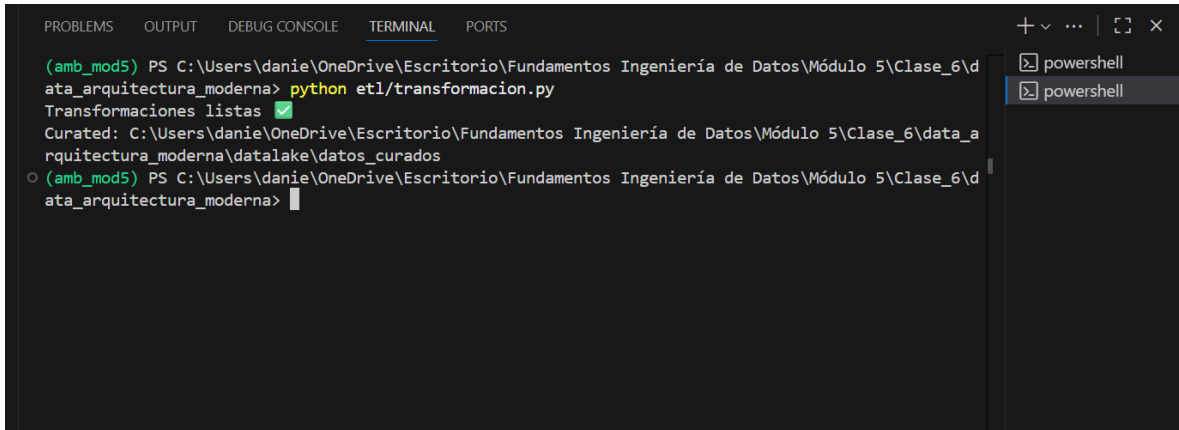
```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl/limpia_clientes.py
C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\etl\limpia_clientes.py:7: SyntaxWarning: invalid escape sequence '\c'
python etl\clean_clientes.py [--in ruta_csv] [--out ruta_csv] [--sep ,] [--enc utf-8]
Filas originales: 50 | tras deduplicar por id_cliente: 50
Nulos en columnas clave -> id_cliente: 0, nombre: 0
Guardado: C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\data lake\datos_procesados\clientes_limpio.csv
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl/limpia_productos.py
C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\etl\limpia_productos.py:7: SyntaxWarning: invalid escape sequence '\l'
python etl\limpia_productos.py [--in ruta_csv] [--out ruta_csv] [--sep ,] [--enc utf-8]
Filas originales: 10 | tras deduplicar por id_producto: 10
Nulos clave -> id_producto: 0, nombre_producto: 0
Guardado: C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\data lake\datos_procesados\productos_limpio.csv
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl/limpia_ventas.py
C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\etl\limpia_ventas.py:7: SyntaxWarning: invalid escape sequence '\c'
python etl\clean_ventas.py [--in ruta_csv] [--out ruta_csv] [--sep ,] [--enc utf-8]
Validez fecha tras limpieza: 100.0%
Filas originales: 100 | tras deduplicar por id_venta: 100
Validez fecha (no nulos): 100%
Guardado: C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\data lake\datos_procesados\ventas_limpio.csv
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

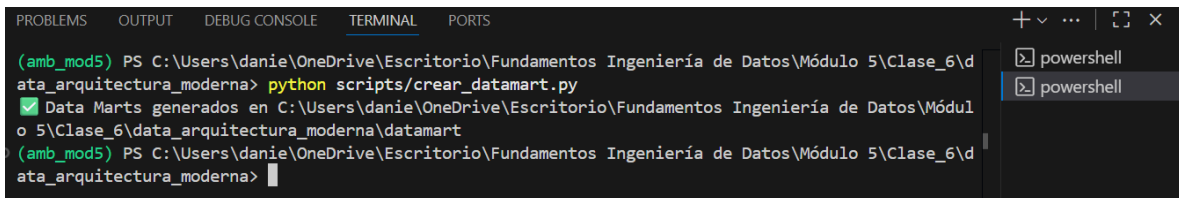
6. Transformar datos procesados en dimensiones y tabla de hecho ventas mediante la ejecución de transformación.py en carpeta etl

Daniela Méndez Gándara



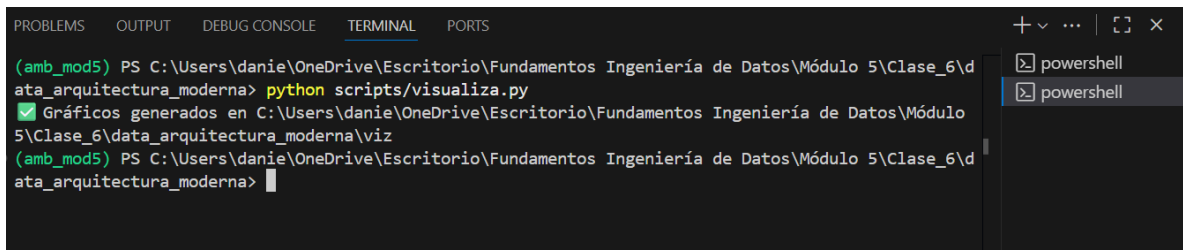
```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python etl/transformacion.py
Transformaciones listas ✓
Curated: C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\datalake\datos_curados
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

7. Se crea datamart mediante script crear_datamart.py en carpeta scripts



```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python scripts/crear_datamart.py
✓ Data Marts generados en C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\datamart
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

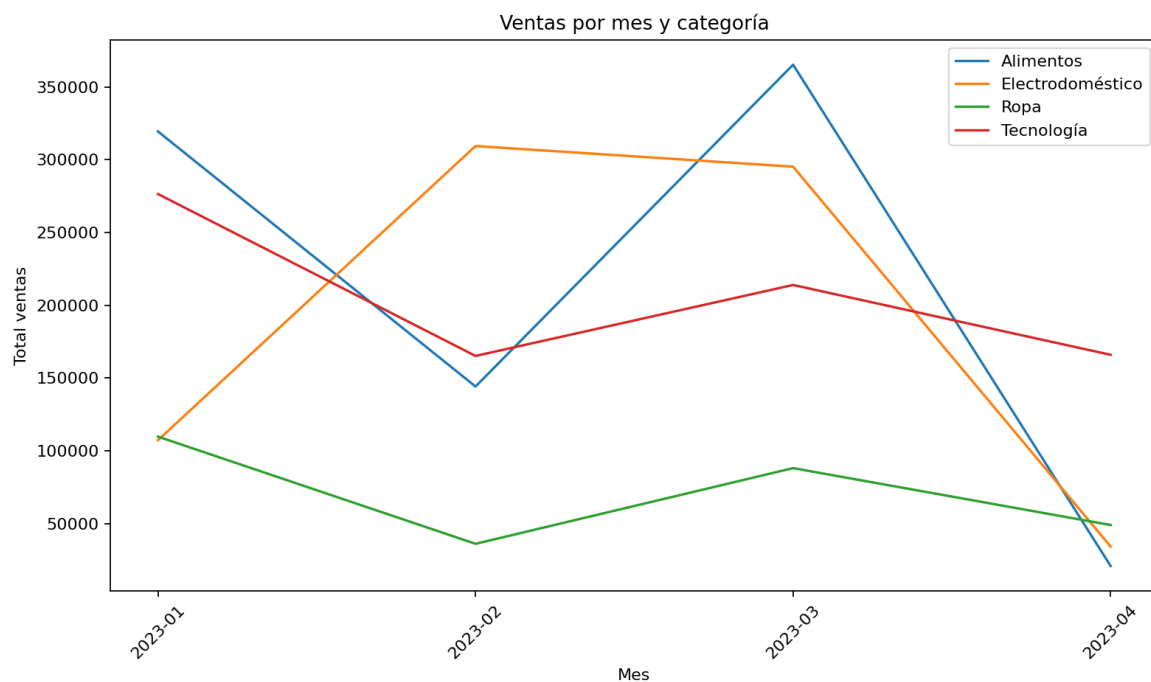
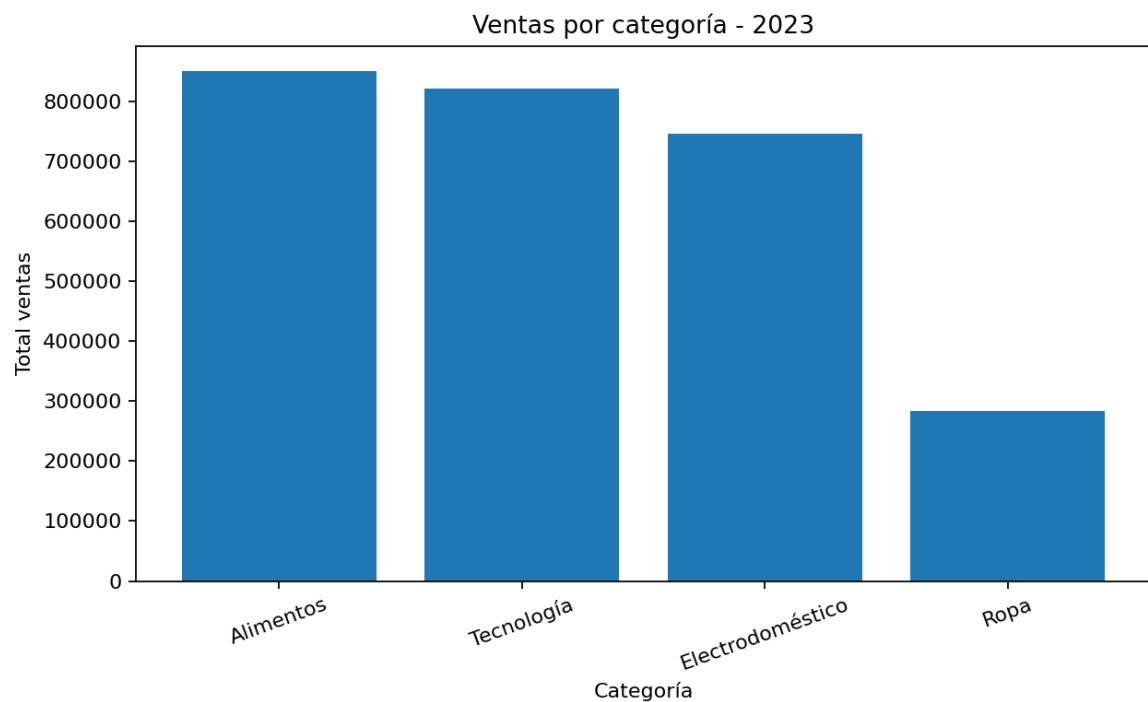
8. Visualizar mediante script visualiza.py en carpeta scripts



```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> python scripts/visualiza.py
✓ Gráficos generados en C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\viz
(amb_mod5) PS C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>
```

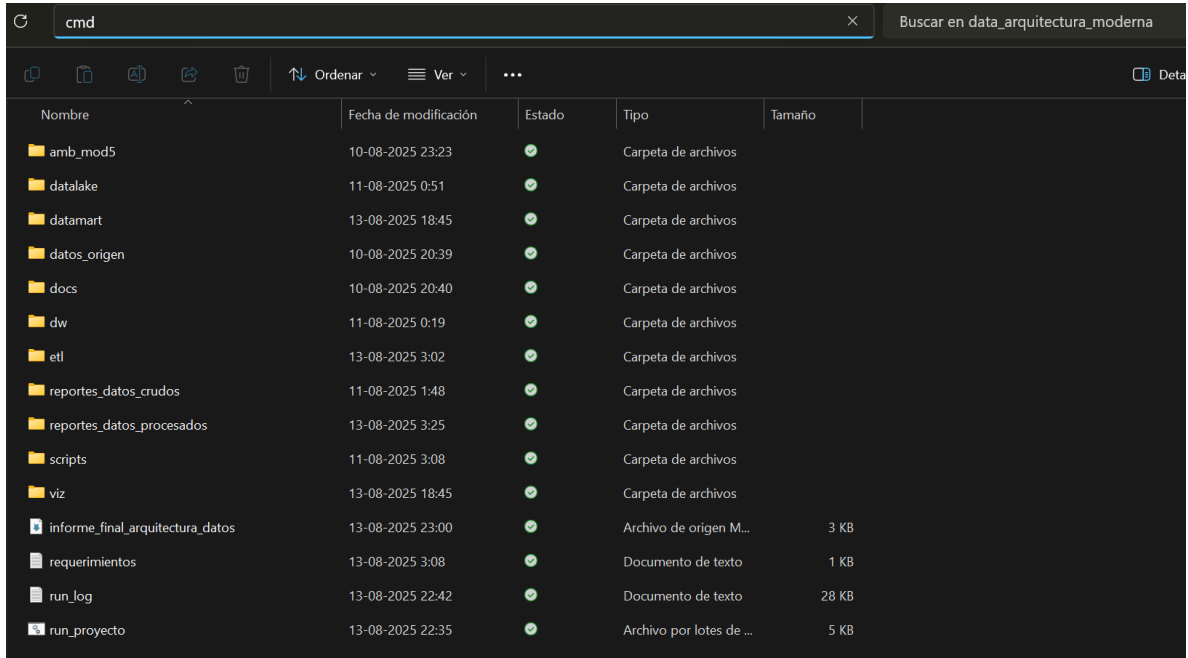
Gráficos generados

Daniela Méndez Gándara



Daniela Méndez Gándara

Se puede ejecutar todo el proyecto , es decir, los pasos descritos en el informe mediante run_proyecto.bat abriendo cmd desde carpeta raíz y ejecutando desde esa misma.



Nombre	Fecha de modificación	Estado	Tipo	Tamaño
amb_mod5	10-08-2025 23:23	✓	Carpeta de archivos	
datalake	11-08-2025 0:51	✓	Carpeta de archivos	
datamart	13-08-2025 18:45	✓	Carpeta de archivos	
datos_origen	10-08-2025 20:39	✓	Carpeta de archivos	
docs	10-08-2025 20:40	✓	Carpeta de archivos	
dw	11-08-2025 0:19	✓	Carpeta de archivos	
etl	13-08-2025 3:02	✓	Carpeta de archivos	
reportes_datos_crudos	11-08-2025 1:48	✓	Carpeta de archivos	
reportes_datos_procesados	13-08-2025 3:25	✓	Carpeta de archivos	
scripts	11-08-2025 3:08	✓	Carpeta de archivos	
viz	13-08-2025 18:45	✓	Carpeta de archivos	
Informe_final_arquitectura_datos	13-08-2025 23:00	✓	Archivo de origen M...	3 KB
requerimientos	13-08-2025 3:08	✓	Documento de texto	1 KB
run_log	13-08-2025 22:42	✓	Documento de texto	28 KB
run_proyecto	13-08-2025 22:35	✓	Archivo por lotes de ...	5 KB



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Versión 10.0.26100.4946]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingenieria de Datos\Módulo 5\Clase_6\data_arquitectura_moderna> run_proyecto.bat
```

Daniela Méndez Gándara

```
[13/14] Data Mart (OBLIGATORIO)

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>if exist "scripts\crear_datamart.py" ("amb_mod5\Scripts\python.exe" scripts\crear_datamart.py || goto :ERROR ) else (echo [13/14] Falta scripts\crear_datamart.py )
✅ Data Marts generados en C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\data mart

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>REM === 11) VISUALIZACION (OBLIGATORIO)

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>echo [14/14] Visualizacion (OBLIGATORIO)
[14/14] Visualizacion (OBLIGATORIO)

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>if exist "scripts\visualiza.py" ("amb_mod5\Scripts\python.exe" scripts\visualiza.py || goto :ERROR ) else (echo [14/14] Falta scripts\visualiza.py )
✅ Gráficos generados en C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna\viz

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>echo.

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>echo ==== Flujo completado con exito ====
==== Flujo completado con exito ====

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>echo Revisa run_log.txt para mas detalle de diagnosticos.
Revisa run_log.txt para mas detalle de diagnosticos.

C:\Users\danie\OneDrive\Escritorio\Fundamentos Ingeniería de Datos\Módulo 5\Clase_6\data_arquitectura_moderna>pause
Presione una tecla para continuar . . . |
```