

Proyecto Superstore

Ayi Academy





Sobre el Proyecto

Superstore es un dataset de una tienda minorista en Estados Unidos. Vende productos de oficina, tecnología y mobiliario a clientes individuales y corporativos.

Tecnologías utilizadas



Python

Pandas, SQLAlchemy



MySQL

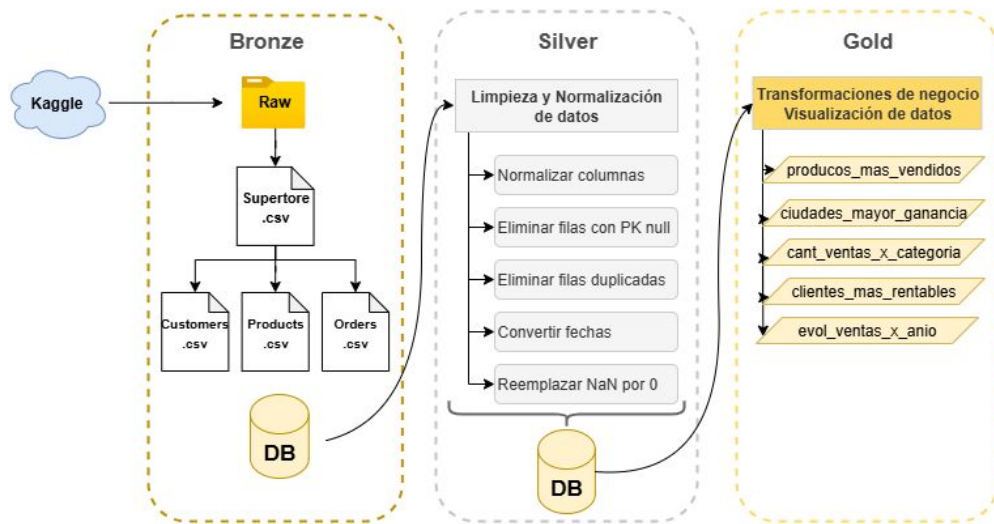


Matplotlib



Seaborn

Arquitectura Medallón



01

Etapa 'Bronze'

Datos Crudos



Bronze

Paso 01



Traer los datos
desde Kaggle

Paso 02



Guardar
dataset en
data/raw

Paso 03



Dividir en tres
dataset

Paso 04



Enviar a la
Base de Datos

Bronze

Conexión con la Base de Datos

```
def get_engine():  
    user = "root"  
    password = "1234"  
    host = "localhost"  
    port = "3306"  
    db = "superstore_db"  
  
    engine = create_engine(f"mysql+pymysql://{user}:{password}@{host}:{port}/{db}")  
    return engine
```

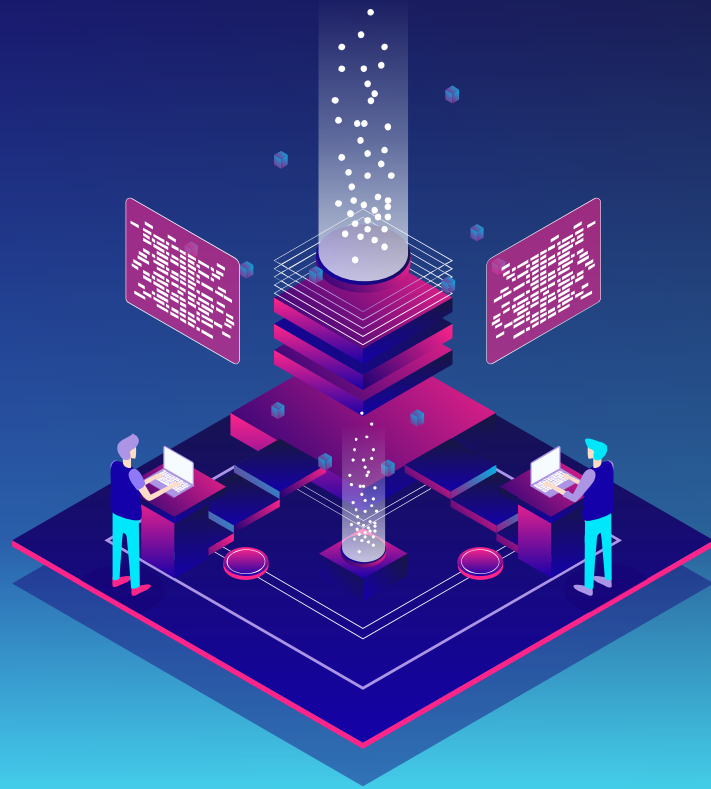
Paso 04

```
# Subir a MySQL (reemplaza tablas si ya existen)  
df_customers.to_sql("customers", engine, if_exists="replace", index=False)  
df_products.to_sql("products", engine, if_exists="replace", index=False)  
df_orders.to_sql("orders", engine, if_exists="replace", index=False)
```

02

Etapa 'Silver'

Limpieza de Datos



Silver

Llamar Base de Datos - Leer Tablas

```
engine = get_engine()

# Leer tablas desde bronze
customers = pd.read_sql("SELECT * FROM customers", engine)
products = pd.read_sql("SELECT * FROM products", engine)
orders = pd.read_sql("SELECT * FROM orders", engine)
```

Normalización - Limpieza

```
# Normalizar columnas
customers.columns = [c.strip().lower().replace(" ", "_") for c in customers.columns]
products.columns = [c.strip().lower().replace(" ", "_") for c in products.columns]
orders.columns = [c.strip().lower().replace(" ", "_") for c in orders.columns]

# Eliminar filas con null en la PK y filas duplicadas
customers = customers.dropna(subset=["customer_id"]).drop_duplicates()
products = products.dropna(subset=["product_id"]).drop_duplicates()
orders = orders.dropna(subset=["order_id"]).drop_duplicates(subset=["order_id", "product_id"])

# Convertir fechas
orders["order_date"] = pd.to_datetime(orders["order_date"])
orders["ship_date"] = pd.to_datetime(orders["ship_date"])
```



Silver

Enviar tablas limpias a la Base de Datos

```
# Guardar tablas limpias en MySQL (silver)
customers.to_sql("customers_silver", engine, if_exists="replace", index=False)
products.to_sql("products_silver", engine, if_exists="replace", index=False)
orders.to_sql("orders_silver", engine, if_exists="replace", index=False)

print("✅ Tablas Silver listas en MySQL")
```



03

Etapa 'Gold'

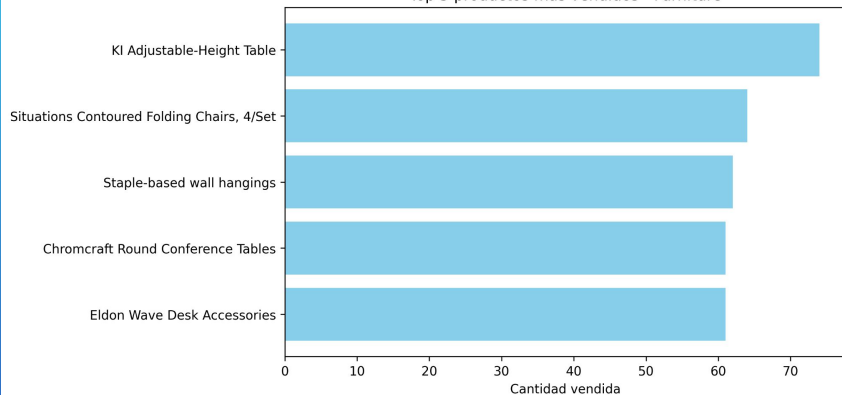
Transformaciones de Negocio
Visualización de Datos



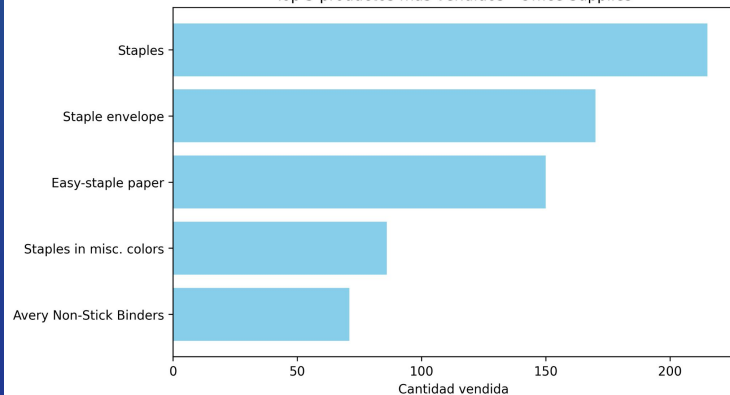
Gold

Productos más vendidos por categoría

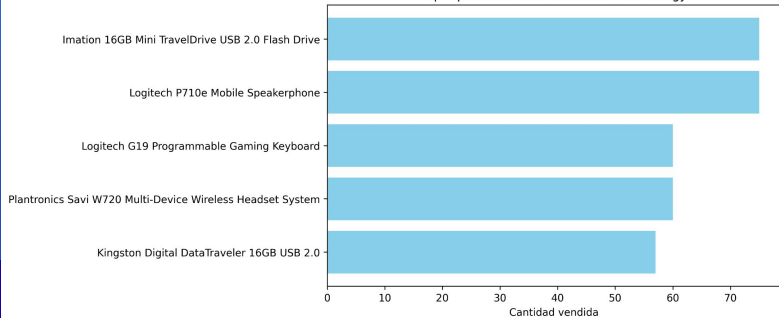
Top 5 productos más vendidos - Furniture



Top 5 productos más vendidos - Office Supplies



Top 5 productos más vendidos - Technology

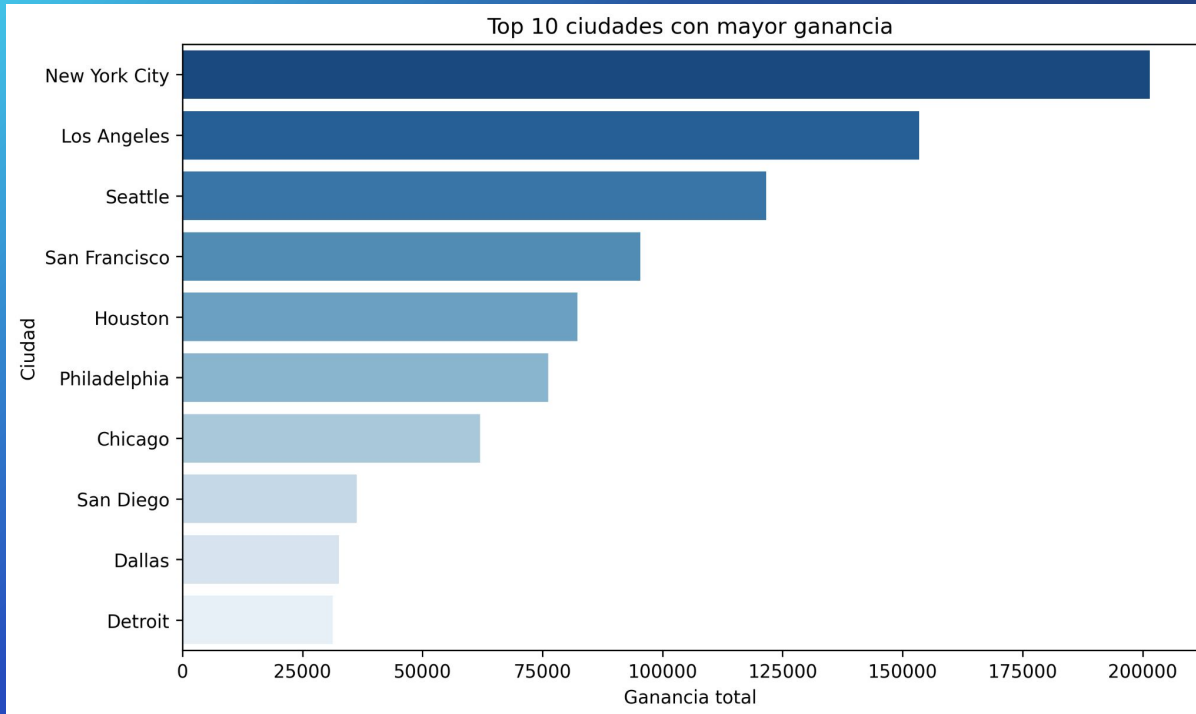


Aportes de valor:

- Productos más demandados
- Ayuda a Optimizar inventario
- Diseñar campañas de marketing
- Tendencias de consumo

Gold

Ciudades con mayor ganancia

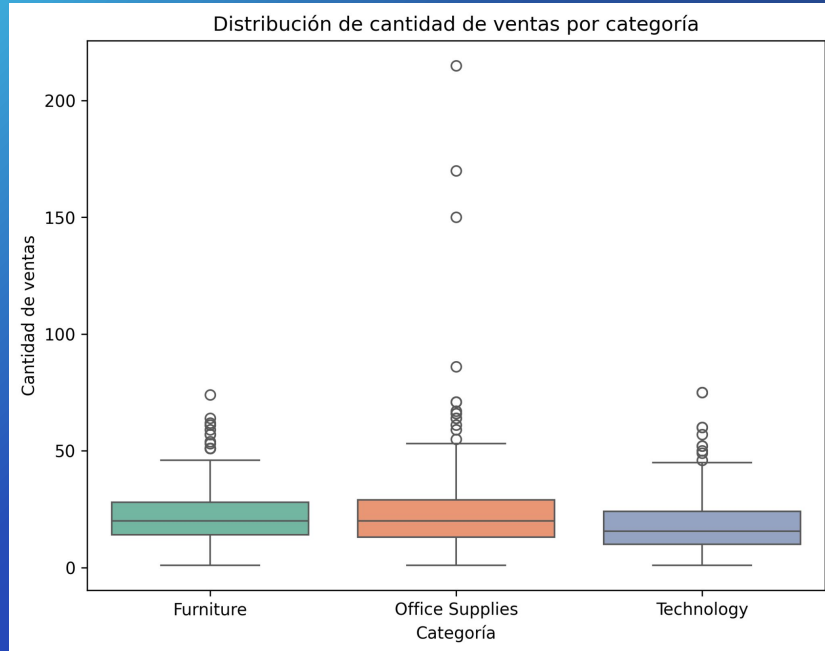


Aportes de valor:

- Zonas geográficas más rentables
- Considerar posibilidad de expansión, logística y distribución
- Identificar oportunidades de negocio

Gold

Cantidad de ventas por Categoría

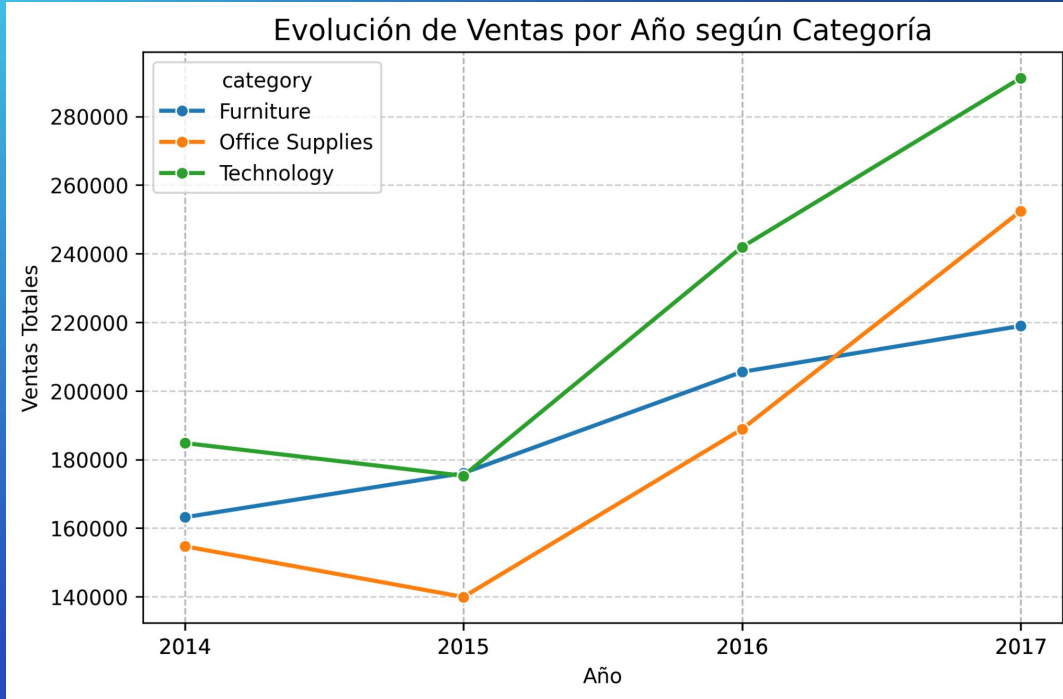


Aportes de valor:

- Visibilidad sobre la distribución de la línea de productos
- Tomar decisiones sobre el balance del catálogo
- Sirve en la planificación de compras y logística

Gold

Evolución de ventas por año

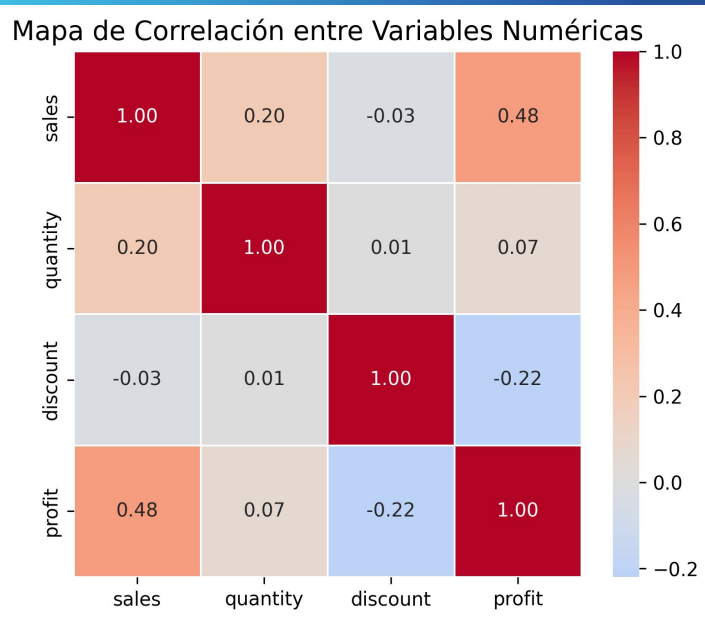


Aportes de valor:

- Analizar tendencia temporal del negocio
- Pronóstico de ventas futuras
- Identificar impacto de estrategias pasadas

Gold

Heatmap



- **Sales - Profit** 0.48 → a mayor venta, mayor ganancia
- **Sales - Quantity** 0.20 → vender más unidades no siempre significa más ventas en dinero
- **Quantity - Profit** 0.07 → la cantidad vendida no garantiza más ganancias
- **Discount - Profit** -0.22 → más descuentos tienden a reducir la ganancia
- **Discount - Sales** -0.03 → aplicar descuento no necesariamente aumenta las ventas en dinero

Aportes de valor:

- Los descuentos no aumentan significativamente las ventas, pero sí afecta negativamente a la ganancia.
- Quantity - Profit → No todos los productos de alto volumen son rentables.
- **Priorización de métricas:** en vez de mirar sólo cantidad de unidades, conviene enfocarse en ventas y márgenes.

**¡Muchas gracias por su
atención!**

