Sales Data ETL and Analysis Project

-----------------------------------

1️⃣ Folder Setup

Create a project folder on your computer:

sales\_data\_project/

├── sales\_data.xlsx

├── etl\_script.py

-----------------------------------

2️⃣ Sample Excel Data

Create 'sales\_data.xlsx' with a sheet named 'Sheet1'.

Example columns:

| Date | Product | Region | Quantity | Unit\_Price |

|------------|--------------|-------------|----------|------------|

| 2025-06-01 | Widget A | East | 10 | 25.00 |

| 2025-06-02 | Widget B | West | 5 | 15.00 |

| 2025-06-03 | Widget A | East | 8 | 25.00 |

-----------------------------------

3️⃣ Install Required Python Packages

In your terminal or command prompt:

pip install pandas openpyxl sqlalchemy

-----------------------------------

4️⃣ Python ETL Script ('etl\_script.py')

import pandas as pd

from sqlalchemy import create\_engine

# Step 1: Read Excel data

file\_path = 'sales\_data.xlsx'

df = pd.read\_excel(file\_path)

# Step 2: Clean/transform data

df['Quantity'] = df['Quantity'].fillna(0)

df['Unit\_Price'] = df['Unit\_Price'].fillna(0)

df['Date'] = pd.to\_datetime(df['Date'])

df['Total\_Sale'] = df['Quantity'] \* df['Unit\_Price']

# Step 3: Create SQLite connection

engine = create\_engine('sqlite:///sales.db')

# Step 4: Load data into SQL

df.to\_sql('sales', con=engine, if\_exists='replace', index=False)

print("Data loaded successfully into 'sales.db'.")

# Step 5: Run SQL queries

with engine.connect() as connection:

# Top 3 Products by Revenue

query\_top\_products = """

SELECT Product, SUM(Total\_Sale) AS Revenue

FROM sales

GROUP BY Product

ORDER BY Revenue DESC

LIMIT 3;

"""

top\_products = pd.read\_sql(query\_top\_products, connection)

print("\nTop 3 Products by Revenue:\n", top\_products)

# Monthly Sales Summary

query\_monthly\_sales = """

SELECT strftime('%Y-%m', Date) AS Month, SUM(Total\_Sale) AS Total\_Sales

FROM sales

GROUP BY Month;

"""

monthly\_sales = pd.read\_sql(query\_monthly\_sales, connection)

print("\nMonthly Sales Summary:\n", monthly\_sales)

# Export results to Excel

with pd.ExcelWriter('sales\_summary.xlsx') as writer:

top\_products.to\_excel(writer, sheet\_name='Top\_Products', index=False)

monthly\_sales.to\_excel(writer, sheet\_name='Monthly\_Sales', index=False)

print("\nReports exported to 'sales\_summary.xlsx'.")

-----------------------------------

5️⃣ How to Run the Script

In your terminal:

python etl\_script.py

-----------------------------------

6️⃣ What This Does

✅ Reads sales data from Excel

✅ Cleans and processes the data

✅ Loads into SQLite database

✅ Runs SQL analysis queries

✅ Exports summary reports back to Excel

-----------------------------------

7️⃣ How to List on Your Resume

Sales Data ETL and Analysis Project

- Built an automated data pipeline integrating Python, SQL, and Excel to clean, analyze, and report sales data.

- Implemented SQL queries to identify top-selling products and monthly revenue trends.

- Automated report generation and export to Excel.