✅ Task 14: ETL Mini Pipeline — Python Extract → Transform → Load

Tools:

- Primary: Python (Colab)
- Libraries: pandas
- Storage: CSV / SQLite

Dataset:

- Retail Sales Dataset
- Customer Churn Dataset
- E-commerce Dataset

Hints / Mini Guide:

1. Load raw dataset from Kaggle CSV.
2. Create folders: raw/, processed/, output/.
3. Clean missing values + duplicates.
4. Standardize column names and datatypes.
5. Create derived columns (margin, segment flags).
6. Split into separate outputs (customers/orders/products).
7. Load outputs into SQLite or export as CSV.
8. Validate counts before/after transformations.
9. Write a README with ETL steps.

Deliverables:

- task14_etl.ipynb
- processed_data.csv (or multiple outputs)
- database.sqlite
- README.md

Final Outcome:
✅ Understand how ETL works practically in analytics workflows.

Interview Questions Related To Above Task:

- What is ETL and why needed?
- ETL vs ELT?
- How do you validate ETL output?
- Why separate tables in pipeline?
- Common ETL errors?

📌 **Task Submission Guidelines**

- ⏰ **Time Window:**

Youcan complete the task anytime between 10:00 AM to 10:00 PM on the given day. Submission link closes at 10:00 PM

- 🔍 **Self-Research Allowed:**

Youare free to explore, Google, or refer to tutorials to understand concepts and complete the task effectively.

- 🛠️ **Debug Yourself:**

Tryto resolve all errors by yourself. This helps you learn problem-solving and ensures you don't face the same issues in future tasks.

- 💸 **No Paid Tools:**

Ifthe task involves any paid software/tools, do not purchase anything. Just learn the process or find free alternatives.

- 📁 **GitHub Submission:**

Create a new GitHub repository for each task.
Add everything you used for the task — code, datasets, screenshots (if any), and a short README.md explaining what you did.

📤 **Submit Here:**

After completing the task, paste your GitHub repo link and submit it using the link below:

- 👉 [Submission Link]