

Task 14 – ETL Mini Pipeline

Objective

To build a simple ETL (Extract, Transform, Load) pipeline using Python and pandas and understand real-world analytics workflows.

Tools & Technologies

Python, Pandas, SQLite, CSV Files, Jupyter Notebook / Google Colab

Dataset

Retail Sales Dataset (CSV format)

ETL Workflow

Extract: Loaded raw retail sales data from CSV using pandas.

Transform: Removed duplicates, handled missing values, standardized column names, converted data types, created derived columns like margin and customer flags, and split data into customers, orders, and products tables.

Load: Exported transformed data into CSV files and loaded it into a SQLite database.

Validation

Validated record counts before and after transformation and ensured schema consistency.

Output Files

task14_etl.ipynb
processed_data.csv
customers.csv
orders.csv
products.csv
database.sqlite

Learning Outcome

This task provided hands-on experience in building ETL pipelines commonly used in data analytics projects.