

III B.Tech I Semester

23A39504	DATA WRANGLING LAB (Professional Core)	L	T	P	C
		0	0	3	1.5

Course Objectives:

- To provide hands-on experience in collecting, cleaning, transforming, and preparing data for analysis.
- To enable students to handle missing, inconsistent, and unstructured data using Python tools.
- To familiarize students with data integration from multiple sources including APIs, databases, and web scraping.
- To enhance skills in preparing datasets for Machine Learning and Data Analytics tasks.

Course Outcomes:

After successful completion of this lab, students will be able to:

- Perform data cleaning operations using Python and related libraries.
- Handle missing, duplicate, and inconsistent data in real-world datasets.
- Integrate and transform data from heterogeneous sources.
- Conduct exploratory analysis and reshape datasets as per analytical needs.
- Apply preprocessing techniques to make data suitable for machine learning models.

List of Experiments (with Cognitive Levels):

1. Load and explore datasets using Pandas: Shape, info, describe, data types
2. Handle missing data using techniques: fillna, dropna, interpolation
3. Detect and remove duplicate and inconsistent records in real datasets
4. Data type conversions and formatting (e.g., datetime parsing, string manipulation)
5. Merge, join, and concatenate multiple datasets
6. Normalize and standardize numerical features using sklearn
7. Apply label encoding and one-hot encoding to categorical variables
8. Perform data binning, transformation, and discretization
9. Web scraping using BeautifulSoup or Selenium to extract tabular data
10. Read and write data using CSV, Excel, JSON, and SQLite/SQLAlchemy
11. Visualize data quality and outliers using seaborn/matplotlib
12. Mini Project: Real-world data wrangling and cleaning pipeline for an open dataset

Software/Tools Required:

- Python (3.7+), Jupyter Notebook / Google Colab
- pandas, NumPy, seaborn, matplotlib
- scikit-learn (for preprocessing utilities)
- BeautifulSoup, requests, Selenium (for web scraping)
- SQLAlchemy / SQLite (for basic data import/export)

Textbooks & Reference Books:

- Wes McKinney, Python for Data Analysis, 3rd Edition, O'Reilly, 2022
- Hadley Wickham, R for Data Science (for concepts, applicable to Python too)
- Jake VanderPlas, Python Data Science Handbook, O'Reilly
- Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly
- Joel Grus, Data Science from Scratch, O'Reilly