DSP505: Programming Lab for Data Science and Artificial Intelligence

TPL616: Advanced Programming for DSAI

Lab-4: Practice Problems (Pandas)

Date: 10-Sept-2025

Instructions:

1. Try to complete lab problems during the lab hour and submit it through canvas. If you can't complete it within the lab time, you can submit it by the end of tomorrow.
2. Prepare all your solution files in a zip file and name it as <Name.zip> and submit on canvas.
3. You can use a jupyter notebook to solve the problems.

Problem Set-0:

Practice the codes given in the slides, execute them and make sure you get the correct output. No need to submit this.

Problem Set-1 (Warm up exercises):

Download the Lab_sheet_Notebook.ipynb, corresponding dataset and solve the problems.

Problems Set-2:

Use the Titanic dataset to solve the following tasks:

Step 1: Load Data

- Load the `Titanic.csv` file into a pandas DataFrame.
- Display dataset info, shape, and first 5 rows.

Step 2: Handle Missing Values

- Fill missing **Age** values with the **median Age**.
- Drop the **Cabin** column (too many missing values).

Step 3: Detect Outliers

- Detect outliers in the `Fare` column using:
    1. **Z-score method** (|Z| > 3)
    2. **IQR method** (outside `[Q1 - 1.5×IQR, Q3 + 1.5×IQR]`)
- Create a new column `Outlier_Fare` = True/False.

Step 4: Use a Pipeline

- Create **functions**:
    - `handle_missing(df)` → handles missing values
    - `detect_outliers(df)` → detects outliers in Fare
    - `summarize(df)` → returns average Age, average Fare, and survival rate
- Apply them in a pipeline: