Aror University of Art, Architecture, Design & Heritage Sukkur.

Department of Artificial Intelligence and Multimedia Gaming

Programming for AI (Spring-2025)

LAB#06

Prepared by: Abdul Haseeb Shaikh

Objective of Lab No. 6:

After performing lab6, students will be able to:

- Implement dataframes using pandas
- Perform Data Analysis using pandas
- Perform Data Cleaning using pandas

Task# 01: Basics of Data Frames

1. Import the Pandas library.
2. Create a dictionary with at least **four keys** representing different attributes of students (e.g., "Name", "Age", "Department", "GPA").
3. The values for each key should be a **list** containing at least **five entries**.
4. Convert the dictionary into a Pandas DataFrame.
5. Print the DataFrame to verify the structure.

Example Output:

If done correctly, the DataFrame should look something like this:

Name	Age	Department	GPA
Alice	20	AI	3.8
Bob	22	CS	3.5
Carol	21	Data Sci	3.9
David	23	IT	3.7
Eve	22	AI	3.6

- Add a new column "Graduation Year" by assuming all students graduate at the age of **24**.
- Display only the students from the "AI" department.

Task#02: Data Analysis

Read the student dataset provided, and perform the following data analysis tasks

Basic Analysis

1. **What is the average score for each subject (math, reading, and writing)?**
2. **How many students passed (score ≥ 50) in each subject?**

Demographic Insights

3. **Is there a significant difference in scores between male and female students?**
4. **How does parental education level affect student performance?**

Comparative Analysis

5. **Do students who complete the test preparation course perform better than those who don't?**
6. **Which subject has the highest average score, and which one has the lowest?**

Correlation Analysis

7. **Is there a correlation between math and reading scores?**
8. **How strongly are reading and writing scores related?**

Task#03: Data Cleaning

Import the train.csv file which is provided to you and do the following tasks:

1. **Handling Missing values:**
    a. Identify columns with missing values.
    b. Decide whether to **drop** or **fill** missing values (e.g., using mean, median, or mode).
2. **Converting data types:**
    a. Convert numerical columns stored as text (if any) to numeric.

3. **Removing Duplicates:**
    a. Check for duplicate rows and remove them.
4. **Standardizing the data:**
    a. Ensure consistency in categorical values (e.g., "yes"/"Yes"/"YES" should be uniform).
    b. Strip extra spaces from text columns.