Machine Learning – Practice Sheet No. 1

Objectives & Outcomes	
Pre-knowledge	• Basic Python syntax • Jupyter Notebook usage • NumPy arrays (data handling) • Pandas basics (DataFrame ops) • Matplotlib plotting • Intro to Scikit-Learn datasets
Objectives	• Practice dataset loading in Scikit-Learn • Explore & analyze Iris dataset • Apply Pandas for summarization • Visualize data with scatter plots & histograms • Calculate descriptive statistics by species
Outcomes	• Ability to load ML datasets in Python • Skill in data exploration using Pandas • Competence in visualization with Matplotlib • Understanding species–feature relationships • Preparedness for classification tasks

Illustrative Examples

1	Q1: What is the Iris dataset, and why is it famous in Machine Learning? A1: The Iris dataset is one of the most famous datasets in machine learning. • It contains 150 samples of iris flowers. • Each sample has 4 features (measurements): 1. Sepal length (cm) 2. Sepal width (cm) 3. Petal length (cm) 4. Petal width (cm) • The flowers belong to 3 species (target labels): ○ Setosa ○ Versicolor ○ Virginica It's popular because it's small, easy to understand, and often used to learn and test machine learning algorithms.

2	Answer The following Questions: a) Install those important Libraries in Machine Learning: (Numpy-matplotlib-Scikit-Learn-Pandas) b) Import those Libraries.
3	Iris Dataset: a) Load the iris dataset in Python using Scikit-Learn. b) Identify which species each target number (0, 1, 2) represents Each species. c) Analyze the data using Pandas. d) Visualize the relationship between Sepal Length and Sepal Width. e) Calculate the average sepal length for each of the three species.

Assignments	
1	a) Modify the scatter plot to visualize petal length vs. petal width.
2	a) Create a histogram for a single feature (e.g., petal width). What does the distribution tell you about the data?

	كود الطالب		اسم الطالب
Comments			Grades