📌 Project Title

Level 1 – Task 2 : Data Cleaning and Preprocessing (Iris Dataset)

📖 Project Overview

This project focuses on cleaning and preprocessing a raw dataset to make it suitable for mach analysis.

The dataset used is the Iris dataset.

The objective was to:

- Handle missing data
- Detect and remove outliers
- Convert categorical variables to numerical format
- Normalize/Standardize numerical data

🛠️ Tools Used

- Python
- Pandas
- NumPy
- Scikit-learn
- Matplotlib
- Seaborn

🔍 Step-by-Step Process

1️⃣ Data Inspection

- Loaded dataset using pandas
- Checked data types and missing values
- Generated summary statistics

2️⃣ Missing Data Handling

- Numerical columns were filled using mean imputation
- Categorical columns were filled using mode

3️⃣ Outlier Detection & Removal

- Used Interquartile Range (IQR) method
- Removed values outside $1.5 \times$ IQR

4️⃣ Categorical Encoding

- Applied Label Encoding to convert species column into numerical format

5️⃣ Feature Scaling

- Applied StandardScaler
- Standardized numerical columns to improve model performance

---

📊 Final Output

A cleaned and preprocessed dataset ready for machine learning modeling.

🎯 Key Learning Outcomes

- Importance of data cleaning
- Handling missing values correctly
- Understanding outlier detection using IQR
- Encoding categorical variables
- Importance of feature scaling