Name: Kush Parihar

Class: IS-1 Batch A

Roll no: 2203023

Subject: Big Data Analytics Lab

BDA Experiment 1

# Problem Statement:

1. Getting data to work with: Download the sample Iris dataset locally (Kaggle)

2. Setting up the working directory.

3. Unpacking the data. Decompress the file locally.

4. Looking at the data. Display the top (10) and bottom (10) of the file.

5. Measuring the length of the data set. Count the number of lines in the file.

6. Encode the categorical data

7. Plot a graph stating the flower like sepal’s width, height, etc.,

# Objective:

Applying data analysis methods to the iris dataset is the primary goal of this project. You will preprocess the data, investigate its properties, visualize essential metrics, and reach conclusions about the flower details using Python programming and pertinent tools.

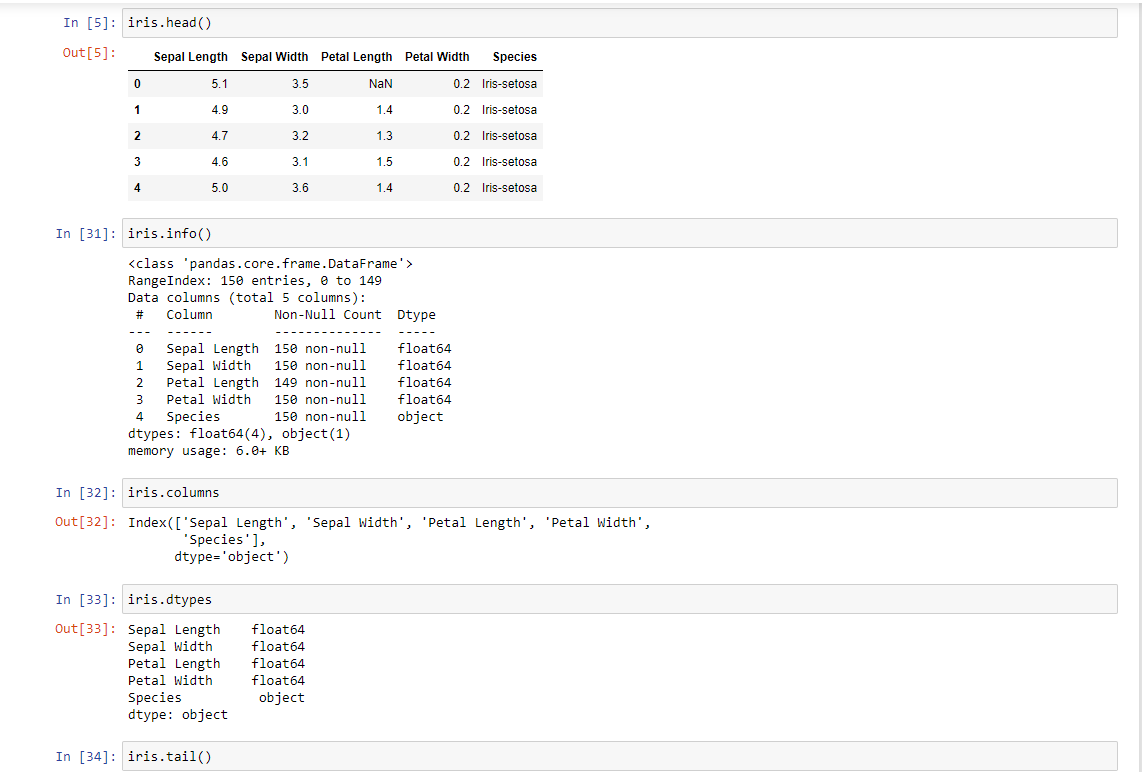
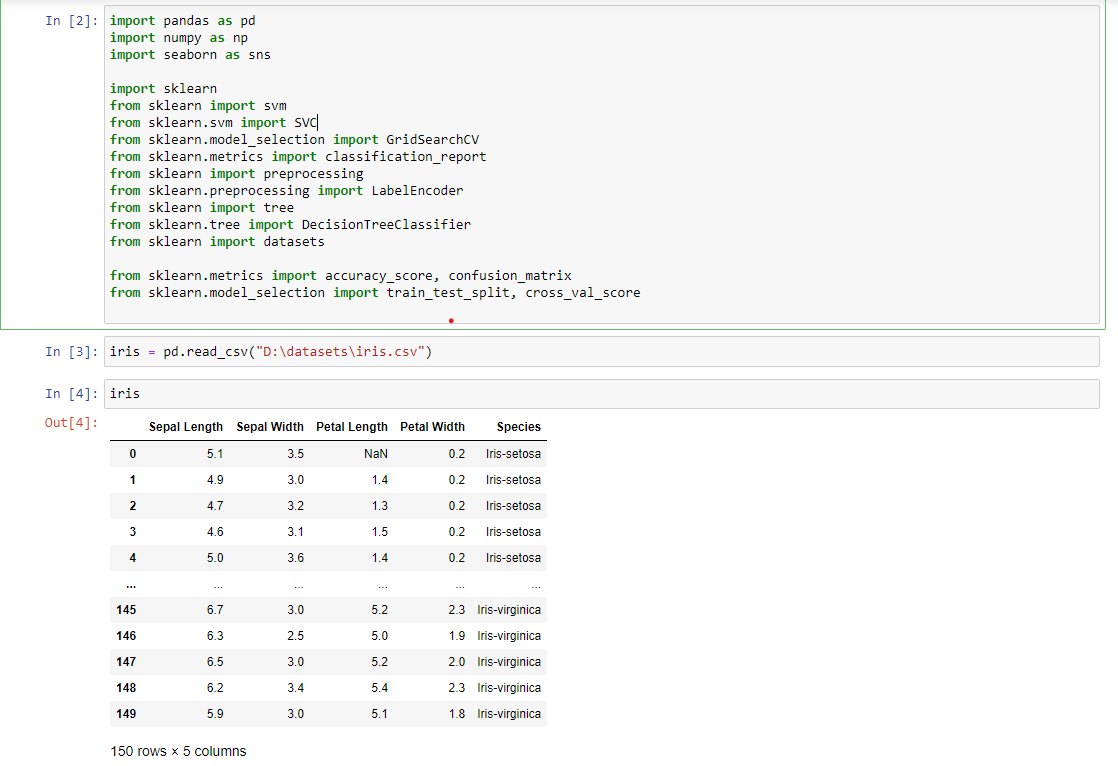
# Theory:

1. **Data Preprocessing:** Describe the importance of data preprocessing and its steps, including handling missing values, outliers, and transforming data to a suitable format.
2. **Exploratory Data Analysis (EDA):** Explain the role of EDA in uncovering trends, patterns, and relationships within the dataset. Discuss the use of summary statistics and data visualization techniques.
3. **Python and Data Libraries:** Introduce Python programming and relevant libraries like Pandas, Matplotlib, and Seaborn that you'll use for data analysis and visualization.

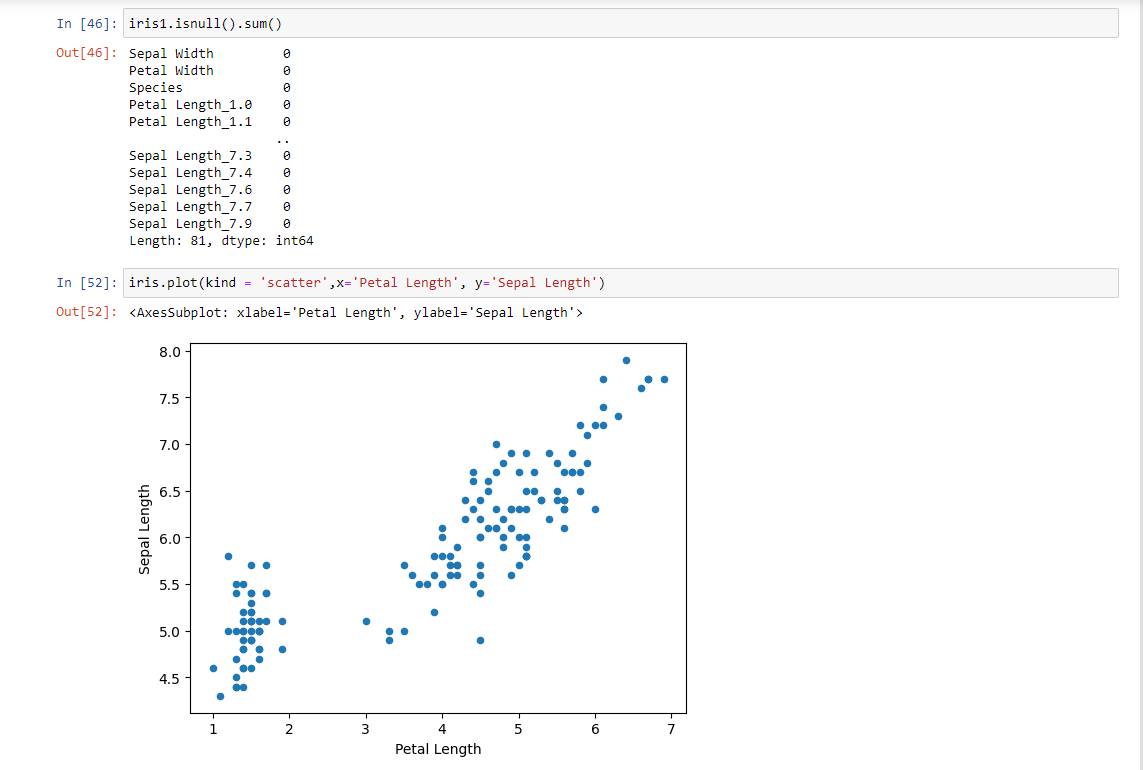
**Algorithm:**

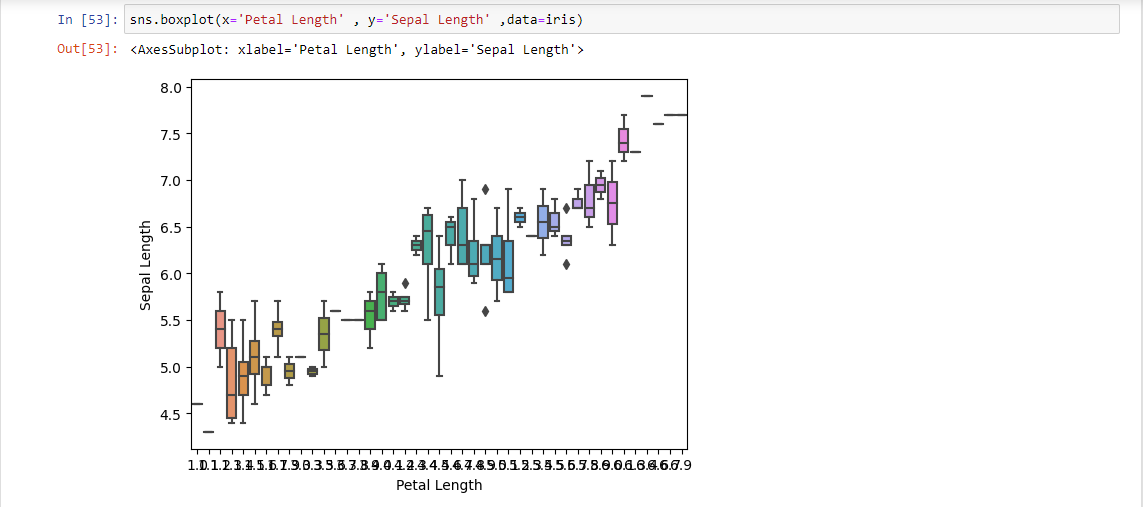
1. Getting Data to Work With: Download the Sample iris Dataset Locally (Kaggle)
2. Download the provided iris dataset from Kaggle and save it to your local machine.
3. Setting Up the Working Directory
4. Create a directory where you will store the dataset and analysis files.
5. Unpacking the Data: Decompress the File Locally.
6. If the dataset is compressed (e.g., in a ZIP file), extract its contents to the working directory.
7. Looking at the Data: Display the Top (10) and Bottom (10) of the File
8. Use Python and Pandas to load the dataset and display the first and last 10 rows to get an initial overview.
9. Measuring the Length of the Dataset: Count the Number of Lines in the File
10. Use Python to count the number of rows in the dataset, representing the total number of detailing.
11. Encode the Categorical Data
12. If the dataset contains categorical data (e.g., states), encode them using suitable techniques like label encoding.
13. Plot a Graph Stating the petal length, width, and height

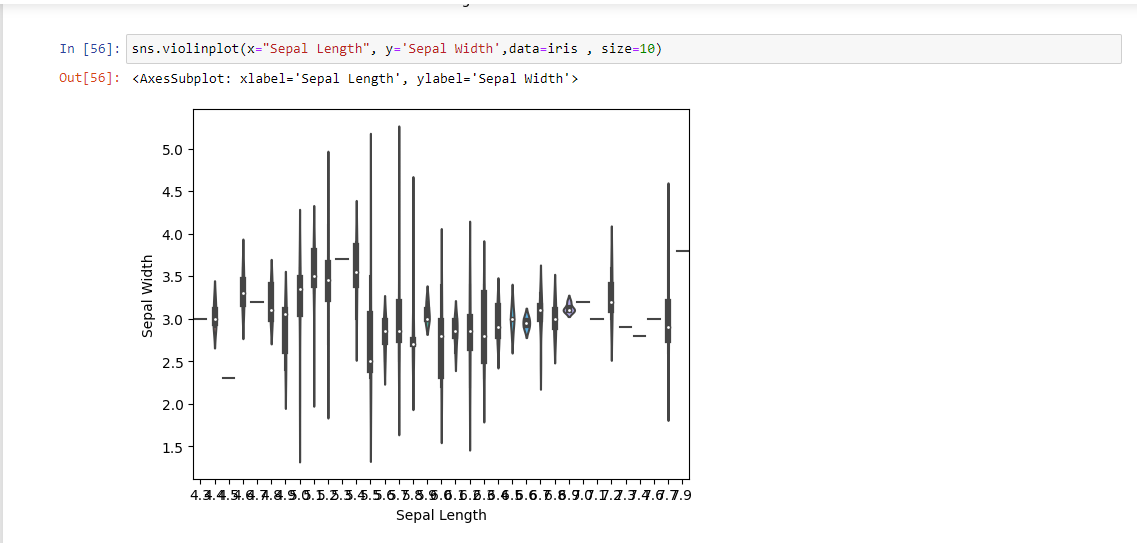
# Code:













# Conclusion:

In this assignment, you demonstrated your ability to work with a real-world dataset and apply essential data analysis techniques. By following the provided steps, you were able to preprocess the data, visualize key metrics, and draw insights regarding the flowers details. This experience highlights your proficiency in data processing and visualization, which are crucial skills in understanding and addressing real-world challenges.