Assignment : 02

Problem Statement:

Write a python script to find basic descriptive statistics using summary, quartile function, etc on iris datasets.

Objective:

The goal is to write a Python script that calculates key descriptive statistics for the Iris dataset. Using libraries like `pandas` or `sklearn`, the script will compute mean, median, mode, standard deviation, range, quartiles, and minimum/maximum values to provide insights into the data's distribution and spread, aiding analysis and exploration.

Prerequisite:

1. Python environment
2. Required libraries
3. Basic Python knowledge
4. Iris dataset familiarity
5. Pandas DataFrames
6. Statistical concepts

Theory:

The Iris dataset is a widely used dataset in machine learning and statistics, containing measurements of three Iris flower species: Setosa, Versicolor, and Virginica. Each entry includes four features: sepal length, sepal width, petal length, and petal width. Due to its simplicity and clear species differentiation, it's ideal for demonstrating data analysis and machine learning techniques.

Descriptive Statistics:
Descriptive statistics summarize a dataset's key characteristics, quantifying central tendency, variability, and distribution. Important measures include:

Mean: Average value, indicating the central point of the data.

Median: Middle value in sorted data, useful for understanding distribution, especially with outliers.

Mode: Most frequent value, identifying common measurements.

Standard Deviation: Measure of data spread around the mean, showing variability.

Range: Difference between maximum and minimum values, showing data extent.

Algorithm:

1. Load the Iris Dataset

2. Explore the Dataset

3. Calculate Descriptive Statistics

4. Determine Quartiles

5. Display Results

6. Identify Outliers (Optional)

7. End Process

Reference:

1) Iris Dataset Overview

2) Python Libraries

3) Descriptive Statistics

Conclusion:

Analyzing the Iris dataset using descriptive statistics provides key insights into the data's distribution and variations across different flower species. By calculating measures like mean, median, standard deviation, and quartiles, we can summarize the dataset effectively, uncovering patterns and trends in the features. Utilizing Python libraries such as pandas and numpy simplifies this analysis, making it efficient and accessible. This exercise not only strengthens fundamental data analysis skills but also serves as a foundation for more advanced techniques in statistics and machine learning