PYTHON REINFORCEMENT

* Importing Libraries

**Source Code:**

import seaborn as sns

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

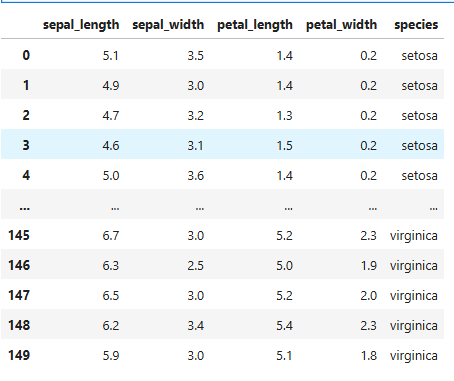
* Loading Dataset

**Source Code:**

iris =sns.load\_dataset("iris")

iris

**Output:**



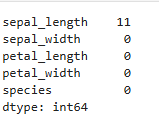
* Introducing NAN values and checking for null values

**Source Code:**

iris.loc[0:10,'sepal\_length']=np.nan

iris.isnull().sum()

**Output:**

****

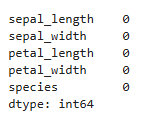
* Filling the null values and checking

**Source code:**

iris['sepal\_length'].fillna(iris['sepal\_length'].mean(),inplace=True)

iris.isnull().sum()

**Output:**



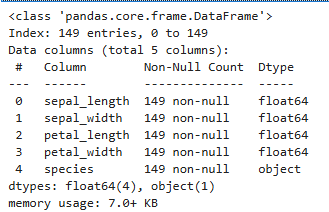
* Dropping duplicates and checking for duplicates

**Source code:**

iris.drop\_duplicates(inplace=True)

iris.info()

**Output:**



* Finding Mean, Median and Variance using Numpy

**Source code:**

feature=['sepal\_length','sepal\_width','petal\_length','petal\_width']

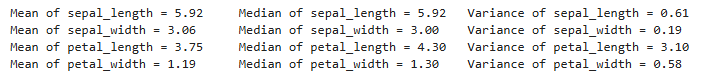
for i in feature:

print(f"Mean of {i} = {np.mean(iris[i]):.2f}","\t"

f"Median of {i} = {np.median(iris[i]):.2f}","\t"

f"Variance of {i} = {np.var(iris[i]):.2f}")

**Output:**



* Filtering rows for each species

**Source code:**

iris['species'].unique()

Setosa = iris[iris['species']=='setosa']

Setosa

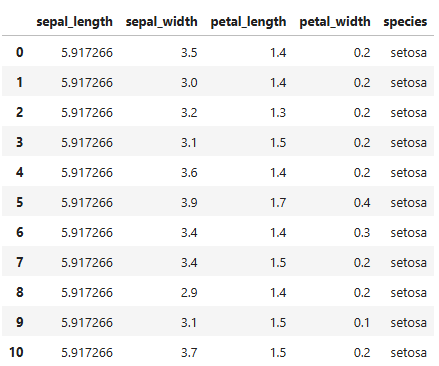
Versicolor = iris[iris['species']=='versicolor']

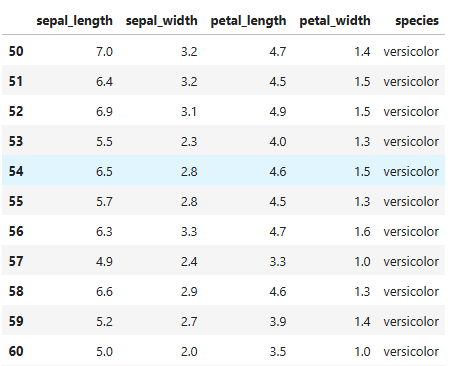
Versicolor

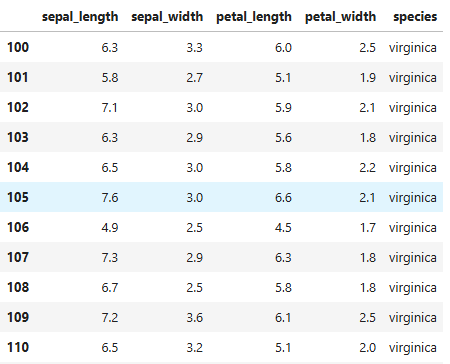
Virginica = iris[iris['species']=='virginica']

Virginica

**Output:**







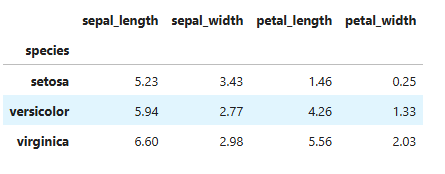
* Grouping data to compute species-wise averages

**Source code:**

species\_average=iris.groupby('species').mean().round(2)

species\_average

**Output:**



* Histogram of sepal\_length

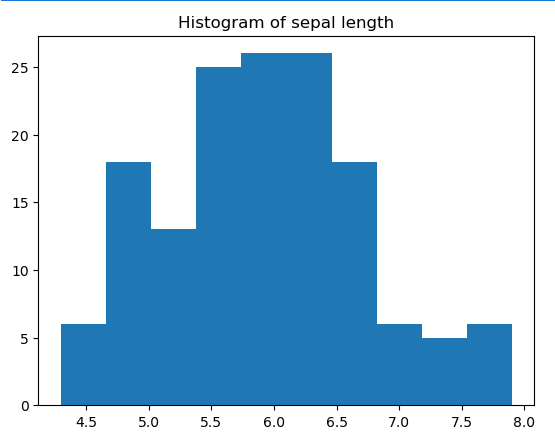
**Source code:**

plt.hist(iris['sepal\_length'],bins=10)

plt.title('Histogram of sepal length')

plt.show()

**Output:**



**Insight:**

1. Sepal length value ranges from 4.3 -7.9
2. The data is concentrated in the range 5.5 – 6.5
3. Only few flowers have very small or very large sepals

* Scatter plot (sepal\_length vs petal\_length)

**Source code:**

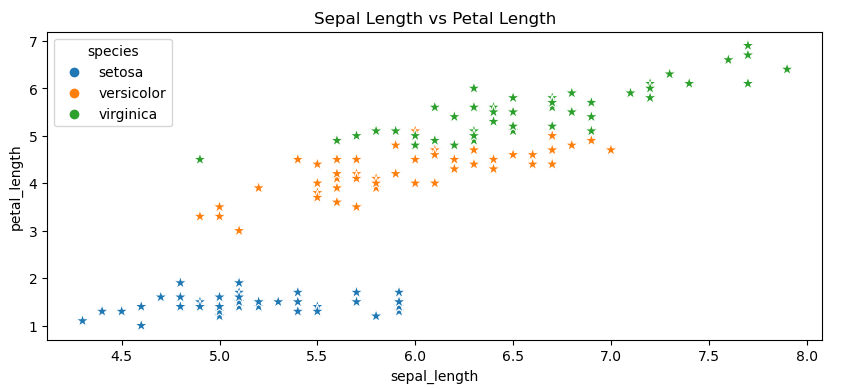
fig=plt.figure(figsize=(10,4))

sns.scatterplot(x='sepal\_length',y='petal\_length',hue='species',data=iris,marker='\*',s=100)

plt.title('Sepal Length vs Petal Length')

plt.show()

**Output:**



**Insights:**

**Setosa**

1. Small sepal and petal length compared to other two species

**Versicolor**

1. Intermediate species
2. Shows slightly an upward trend (positive correlation)
3. Overlap with Virginica

**Virginica**

1. Large sepal and petal compared to others
2. Shows an upward trend (positive correlation)
3. Overlap with Versicolor

* Boxplot grouped by species

**Source code:**

fig=plt.figure(figsize=(20,15))

for i in range(len(feature)):

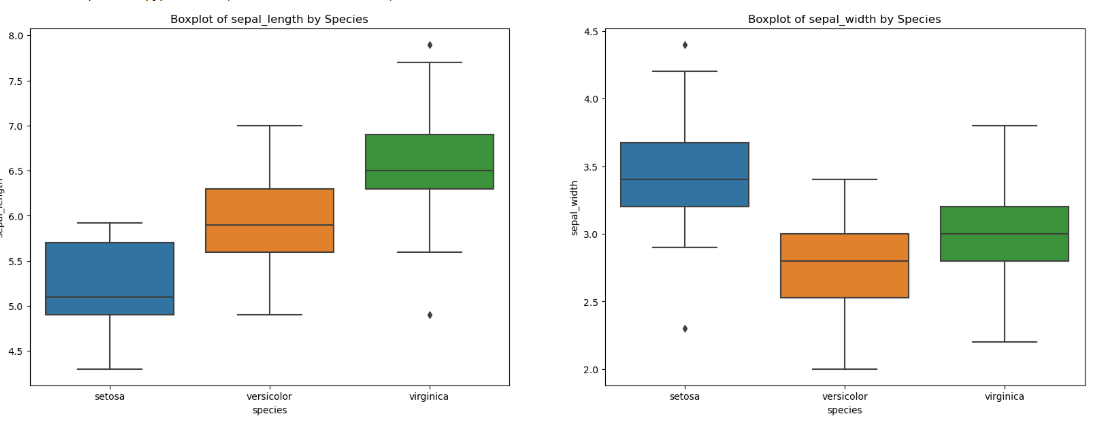
plt.subplot(2,2,i+1)

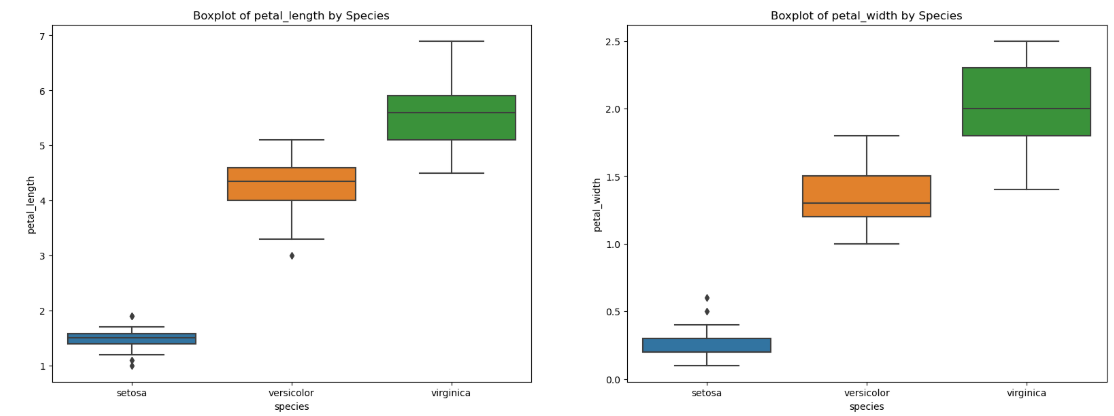
sns.boxplot(x='species',y=feature[i],data=iris)

plt.title(f"Boxplot of {feature[i]} by Species")

plt.show

**Output:**





Insights:

1. Virginica has outlier in sepal length
2. Setosa has outlier in all feature except sepal length
3. Versicolor has an outlier in petal length

* Time series by assigning a date range equal to the dataset size and plot rolling averages.

**Source code:**

dates = pd.date\_range(start='2025-01-01', periods=len(iris))

data = iris.copy()

data.index = dates

window\_size = 10

rolling\_data = data[['sepal\_length', 'sepal\_width', 'petal\_length', 'petal\_width']].rolling(window=window\_size).mean()

plt.figure(figsize=(12,10))

for i in range(len(feature)):

plt.subplot(2,2,i+1)

plt.plot(data.index, data[feature[i]], color="red", label=f"{feature[i]}") # raw data

plt.plot(rolling\_data.index, rolling\_data[feature[i]], color="black", label=f"{feature[i]} (rolling avg)") # rolling average

plt.title(f"{feature[i]} with Rolling Average")

plt.legend()

plt.tight\_layout(pad=4)

plt.show()

**Output:**

