5.EDA-Data Visualization

AIM:

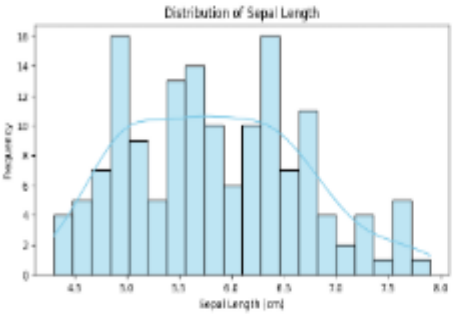
To explore and understand the underlying patterns, distributions, and relationships within the dataset through visual representations, which aids in uncovering insights, detecting anomalies, and guiding further analysis.

PROGRAM:

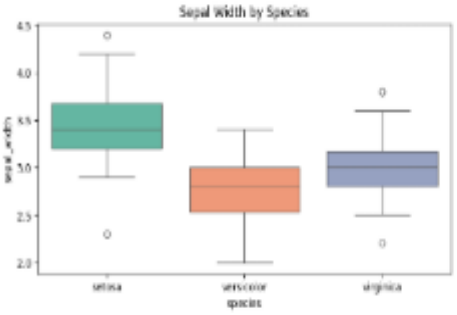
```
import seaborn as sns

import matplotlib.pyplot as plt


# Load iris dataset

df = sns.load_dataset('iris')


# 1. Histogram and KDE plot of sepal_length

plt.figure(figsize=(8, 4))

sns.histplot(df['sepal_length'], kde=True, bins=20, color='skyblue')

plt.title('Distribution of Sepal Length')

plt.xlabel('Sepal Length (cm)')

plt.ylabel('Frequency')

plt.show()


# 2. Boxplot of sepal_width by species (to see spread and outliers)

plt.figure(figsize=(8, 4))

sns.boxplot(x='species', y='sepal_width', data=df, palette='Set2')

plt.title('Sepal Width by Species')

plt.show()
```

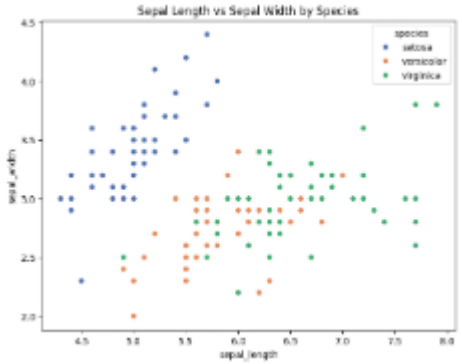
```
# 3. Scatter plot of sepal_length vs sepal_width colored by species
plt.figure(figsize=(8, 6))
sns.scatterplot(x='sepal_length', y='sepal_width', hue='species', data=df, palette='deep')
plt.title('Sepal Length vs Sepal Width by Species')
plt.show()

# 4. Pairplot of all numerical features colored by species
sns.pairplot(df, hue='species', palette='bright')
plt.suptitle('Pairplot of Iris Dataset', y=1.02)
plt.show()

# 5. Correlation heatmap of numeric variables
plt.figure(figsize=(6, 5))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```

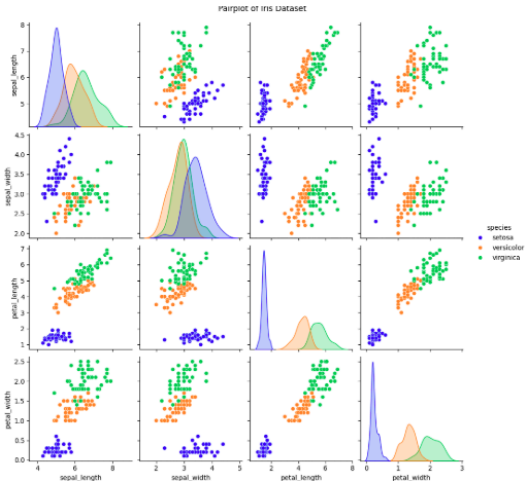
OUTPUT:

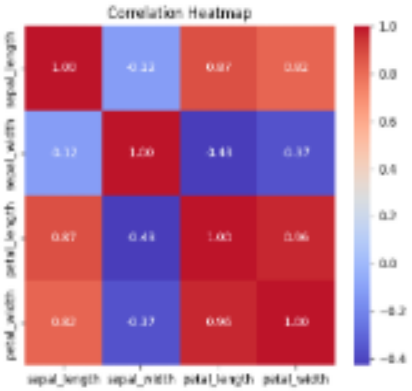



```
tmp/ipython-input-2000117315.py:17: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.11.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

sns.boxplot(x='species', y='sepal_width', data=df, palette='Set2')
```









RESULT:

Thus, the given program was written and executed successfully.