1. Task 1 (Completed):

it is completed by choosing the dataset I choosed the famous iris dataset for this assignment

2. Task 2 (Completed):

Task 2 is completed by uploading the dataset to google colab.

3. Task 3 (Completed):

By performing the descriptive analysis on the dataset you can see the code in the below cell I have mentioned Task3 at the top of cell in the comment

4. Task 4 (Completed):

Choose any two variables and perform the pearson corelation coefficient and comment on the findings see the below cell I have mentioned Task 4 at the top of cell in the comment

5. Task 5 (Completed):

Create atleast 4 visualizations See the below cell I have mentioned Task 5 at the top of cell in the comment

6. Task 6 (Completed):

I have properly commented each and every cell and output of that cell

7. Task 7 (Completed):

Downloading the notebook

```
#task3
import pandas as pd

# Loading the dataset into a Pandas DataFrame
file_path = 'iris.csv'
df = pd.read_csv(file_path)

# Displaying the descriptive statistics for relevant numeric variables
numeric_df = df.select_dtypes(include='number')
descriptive_stats = numeric_df.describe()
print(descriptive_stats)

# Additional statistics: median and mode
median_values = numeric_df.median()
mode_values = numeric_df.mode().iloc[0]

print("\nMedian values:")
print(median_values)

print("\nMode values:")
print(mode_values)

# Analysis: Comment on the results
analysis = """
Analysis of the Iris Dataset:

1. Sepal Length:
    - Mean: {:.2f}
    - Median: {:.2f}
    - Mode: {:.2f}
    - Standard Deviation: {:.2f}

2. Sepal Width:
    - Mean: {:.2f}
    - Median: {:.2f}
    - Mode: {:.2f}
    - Standard Deviation: {:.2f}

3. Petal Length:
    - Mean: {:.2f}
    - Median: {:.2f}
    - Mode: {:.2f}
    - Standard Deviation: {:.2f}
```

```
4. Petal Width:
    - Mean: {:.2f}
    - Median: {:.2f}
    - Mode: {:.2f}
    - Standard Deviation: {:.2f}


Observations:
- Sepal length has a relatively higher mean and median compared to sepal width, petal length, and petal width.
- The standard deviation for petal length and petal width is higher compared to sepal length and sepal width, indicating greater variability :
- The mode values for sepal width and petal width show some repetitive measurements in the dataset.
""".format(
    numeric_df['sepal_length'].mean(), numeric_df['sepal_length'].median(), mode_values['sepal_length'], numeric_df['sepal_length'].std(),
    numeric_df['sepal_width'].mean(), numeric_df['sepal_width'].median(), mode_values['sepal_width'], numeric_df['sepal_width'].std(),
    numeric_df['petal_length'].mean(), numeric_df['petal_length'].median(), mode_values['petal_length'], numeric_df['petal_length'].std(),
    numeric_df['petal_width'].mean(), numeric_df['petal_width'].median(), mode_values['petal_width'], numeric_df['petal_width'].std()
)

print(analysis)
```

```
           sepal_length  sepal_width  petal_length  petal_width
count        150.000000   150.000000    150.000000   150.000000
mean           5.843333     3.057333      3.758000     1.199333
std            0.828066     0.435866      1.765298     0.762238
min            4.300000     2.000000      1.000000     0.100000
25%            5.100000     2.800000      1.600000     0.300000
50%            5.800000     3.000000      4.350000     1.300000
75%            6.400000     3.300000      5.100000     1.800000
max            7.900000     4.400000      6.900000     2.500000

Median values:
sepal_length    5.80
sepal_width     3.00
petal_length    4.35
petal_width     1.30
dtype: float64

Mode values:
sepal_length    5.0
sepal_width     3.0
petal_length    1.4
petal_width     0.2
Name: 0, dtype: float64

Analysis of the Iris Dataset:

1. Sepal Length:
    - Mean: 5.84
    - Median: 5.80
    - Mode: 5.00
    - Standard Deviation: 0.83

2. Sepal Width:
    - Mean: 3.06
    - Median: 3.00
    - Mode: 3.00
    - Standard Deviation: 0.44

3. Petal Length:
    - Mean: 3.76
    - Median: 4.35
    - Mode: 1.40
    - Standard Deviation: 1.77

4. Petal Width:
    - Mean: 1.20
    - Median: 1.30
    - Mode: 0.20
    - Standard Deviation: 0.76

Observations:
- Sepal length has a relatively higher mean and median compared to sepal width, petal length, and petal width.
- The standard deviation for petal length and petal width is higher compared to sepal length and sepal width, indicating greater variabi
- The mode values for sepal width and petal width show some repetitive measurements in the dataset.
```

```python
#task4

# Correlation Analysis
# Select two variables: petal_length and petal_width
correlation = numeric_df[['petal_length', 'petal_width']].corr(method='pearson')

print("\nCorrelation Analysis:")
print(correlation)

correlation_analysis = """
Correlation Analysis:
The Pearson correlation coefficient between petal length and petal width is {:.2f}. This indicates a strong positive correlation, suggesting
""".format(correlation.loc['petal_length', 'petal_width'])

print(correlation_analysis)
```

```
Correlation Analysis:
             petal_length  petal_width
petal_length     1.000000     0.962865
petal_width      0.962865     1.000000

Correlation Analysis:
The Pearson correlation coefficient between petal length and petal width is 0.96. This indicates a strong positive correlation, suggesti
```

```python
#task 5
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Data Visualization
# 1. Histogram for Sepal Length
plt.figure(figsize=(10, 6))
sns.histplot(df['sepal_length'], bins=20, kde=True)
plt.title('Histogram of Sepal Length')
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Frequency')
plt.show()

# Analysis: The histogram shows that sepal length is normally distributed with a slight skew towards the longer lengths. The most common sepa:

# 2. Scatter Plot for Petal Length vs Petal Width
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='petal_length', y='petal_width', hue='species')
plt.title('Scatter Plot of Petal Length vs Petal Width')
plt.xlabel('Petal Length (cm)')
plt.ylabel('Petal Width (cm)')
plt.show()

# Analysis: The scatter plot shows a strong positive correlation between petal length and petal width. Different species of Iris are clearly :

# 3. Box Plot for Sepal Width by Species
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='species', y='sepal_width')
plt.title('Box Plot of Sepal Width by Species')
plt.xlabel('Species')
plt.ylabel('Sepal Width (cm)')
plt.show()

# Analysis: The box plot indicates that the species 'setosa' has a higher median sepal width compared to 'versicolor' and 'virginica'. Additi

# 4. Pair Plot for All Numerical Variables
plt.figure(figsize=(10, 6))
sns.pairplot(df, hue='species')
plt.suptitle('Pair Plot of All Numerical Variables', y=1.02)
plt.show()

# Analysis: The pair plot provides a comprehensive view of the relationships between all pairs of numerical variables. It reinforces the earl:
```
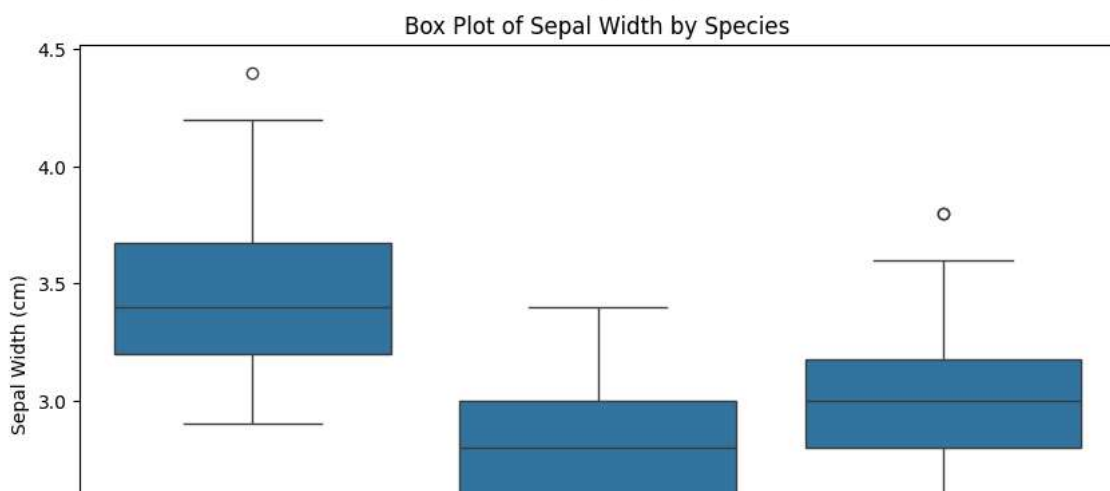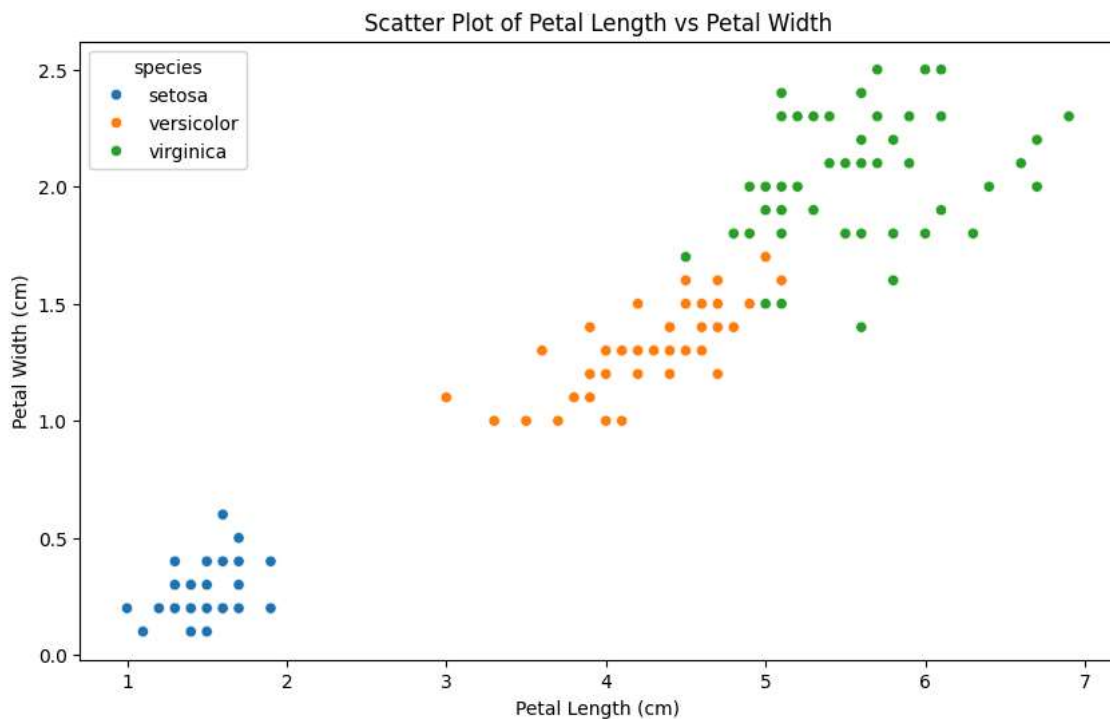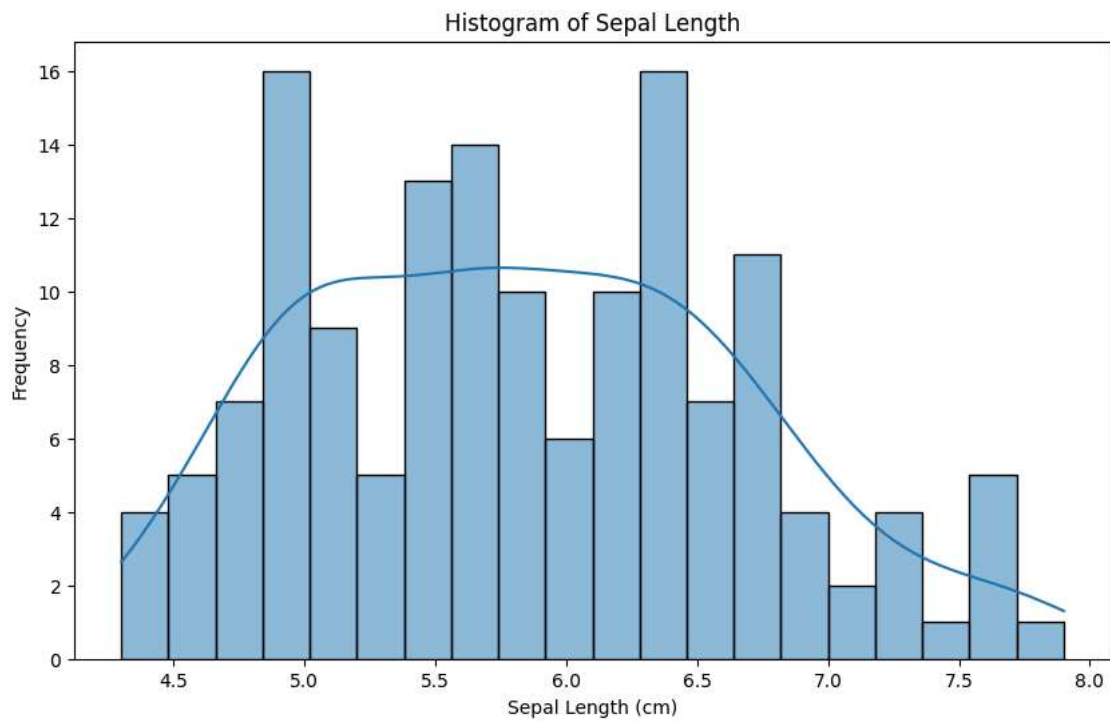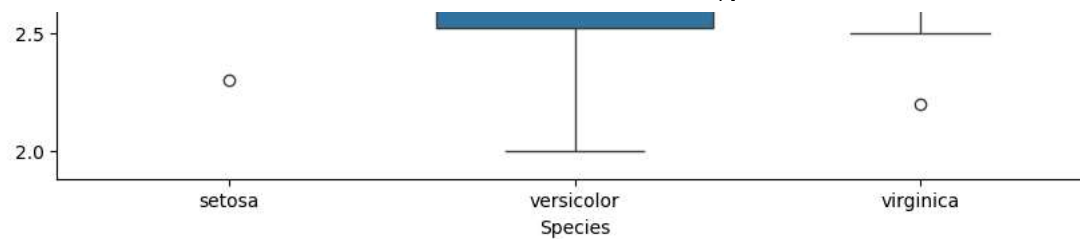
Histogram of Sepal Length

Scatter Plot of Petal Length vs Petal Width

Box Plot of Sepal Width by Species

<Figure size 1000x600 with 0 Axes>

Pair Plot of All Numerical Variables