```
Start coding or generate with AI.
```

Task 3: Exploratory Data Analysis (EDA)

Dataset: Iris Dataset

Objective:

To understand data patterns, detect outliers, analyze correlations, and identify important features using data visualization.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Load Iris dataset
df = sns.load_dataset("iris")

# Display first 5 rows
df.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Next steps: Generate code with df | New interactive sheet

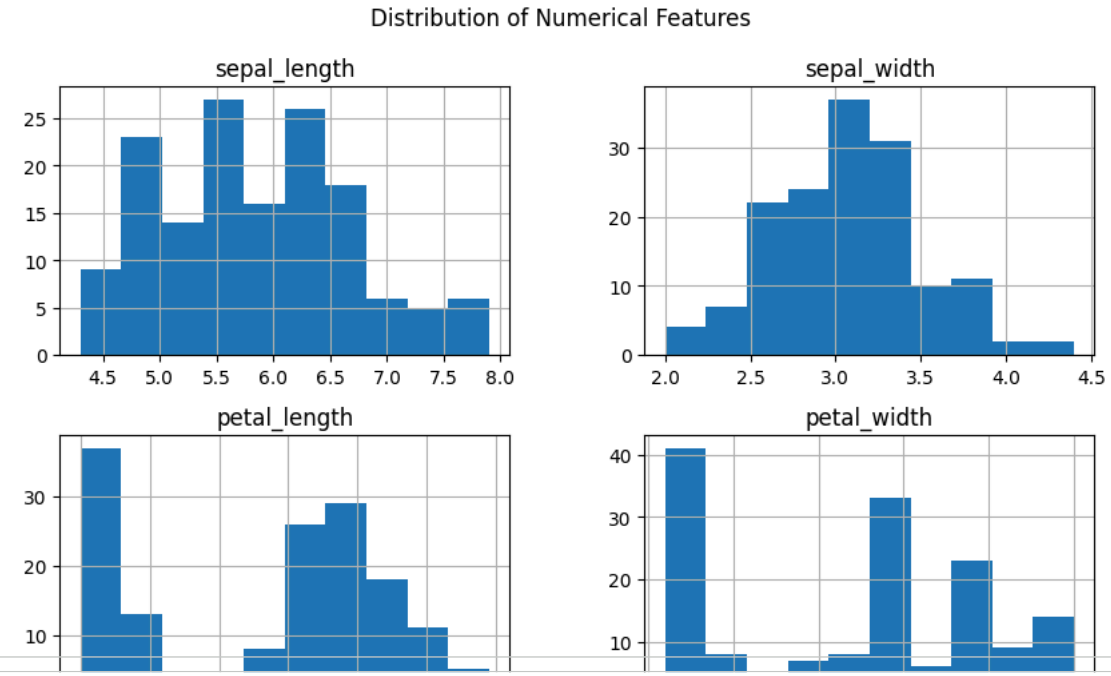
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

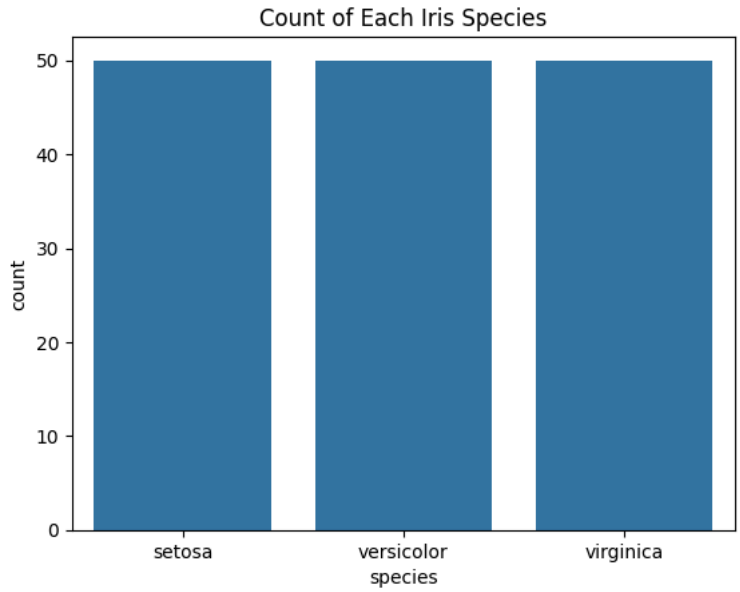
```
df.describe()
```

	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

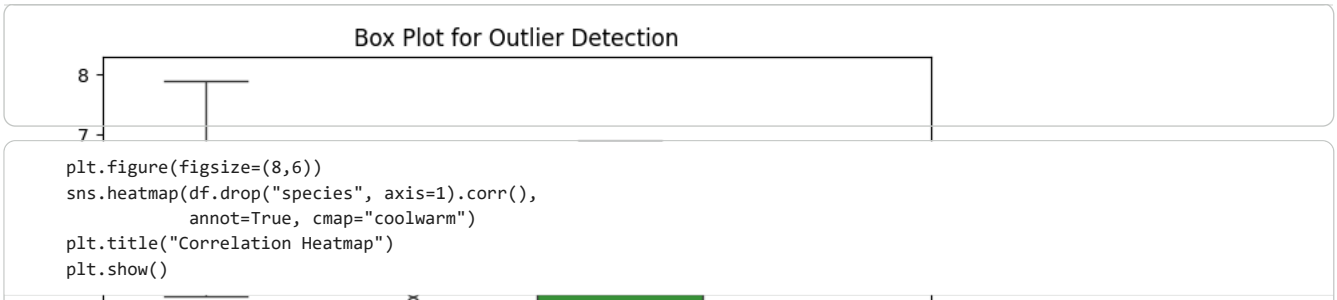
```
df.hist(figsize=(10,6))
plt.suptitle("Distribution of Numerical Features")
plt.show()
```



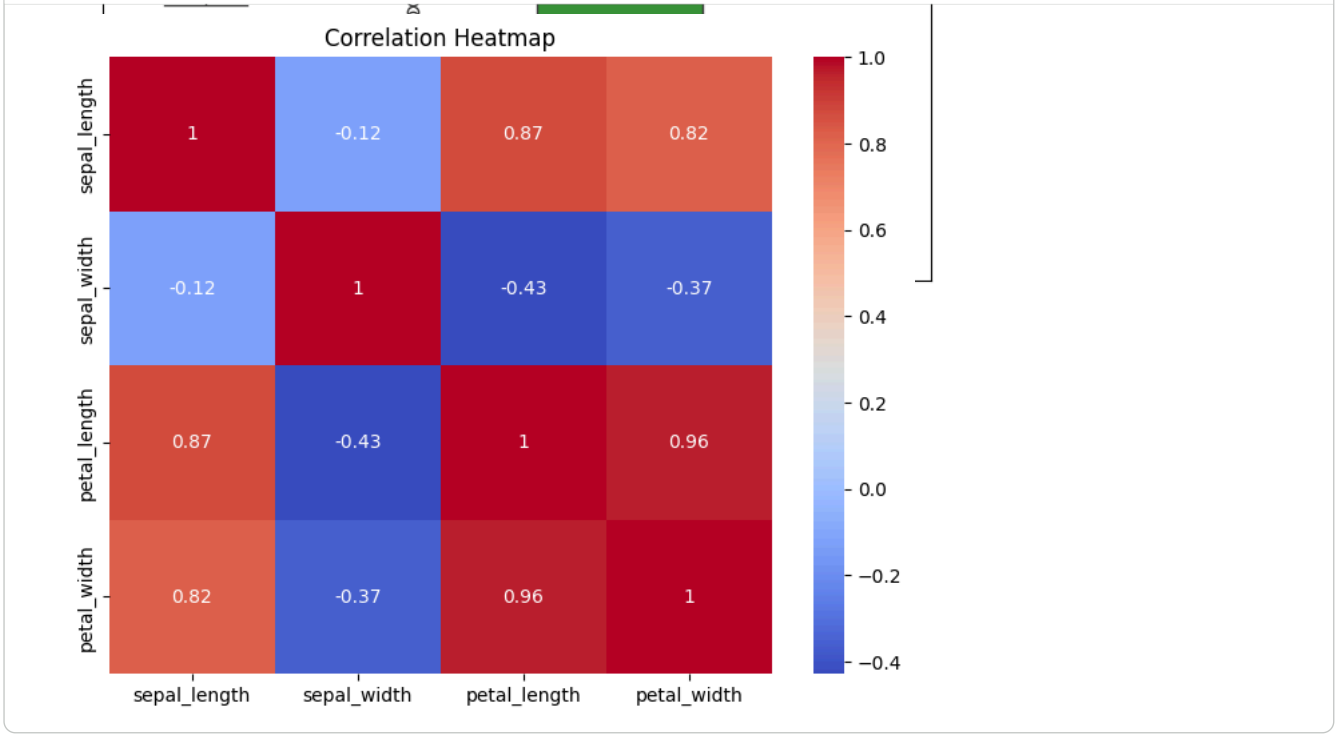

```
sns.countplot(x="species", data=df)
plt.title("Count of Each Iris Species")
plt.show()
```




```
plt.figure(figsize=(8,5))
sns.boxplot(data=df)
plt.title("Box Plot for Outlier Detection")
plt.show()
```




```
plt.figure(figsize=(8,6))
sns.heatmap(df.drop("species", axis=1).corr(),
            annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```



Important Features

Based on EDA, the most important features for prediction are:

- Petal length
- Petal width

Close

```
## Summary

- Iris dataset is clean and balanced
- Petal features are strong predictors
- Minimal outliers exist
- Dataset is suitable for classification
```

Summary

- Iris dataset is clean and balanced
- Petal features are strong predictors
- Minimal outliers exist
- Dataset is suitable for classification