In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import urllib.request


url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
file_name = "iris.csv"
urllib.request.urlretrieve(url, file_name)
print(f"Dataset baixado e salvo como {file_name}.")


column_names = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 's
data = pd.read_csv(file_name, header=None, names=column_names)

print("\nDataset - Primeiras Linhas:")
print(data.head())


print("\nInformações Básicas do Dataset:")
print(data.info())


print("\nResumo Estatístico:")
print(data.describe())


print("\nVerificar Valores Ausentes:")
print(data.isnull().sum())


print("\nDistribuição das Classes:")
print(data['species'].value_counts())


sns.pairplot(data, hue='species', diag_kind='kde', palette='husl')
plt.suptitle("Gráfico Pairplot - Dataset Iris", y=1.02)
plt.show()


correlation_matrix = data.iloc[:, :-1].corr()
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title("Matriz de Correlação das Variáveis")
plt.show()


plt.figure(figsize=(12, 6))
sns.boxplot(data=data, x='species', y='sepal_length', palette='Set2')
plt.title("Comprimento da Sépala por Espécie")
plt.show()

sns.boxplot(data=data, x='species', y='sepal_width', palette='Set3')
plt.title("Largura da Sépala por Espécie")
plt.show()
```

```
data.iloc[:, :-1].hist(figsize=(10, 8), bins=15, edgecolor='black')
plt.suptitle("Histogramas das Variáveis", y=0.93)
plt.show()


grouped_stats = data.groupby('species').agg(['mean', 'std'])
print("\nEstatísticas Agrupadas por Espécie (Média e Desvio Padrão):")
print(grouped_stats)


output_file = "iris_summary_stats.csv"
grouped_stats.to_csv(output_file)
print(f"Estatísticas salvas no arquivo '{output_file}'.")
```

```
Dataset baixado e salvo como iris.csv.

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

Informações Básicas do Dataset:
<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
None

Resumo Estatístico:
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
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

Verificar Valores Ausentes:
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64

Distribuição das Classes:
species
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
Name: count, dtype: int64
```

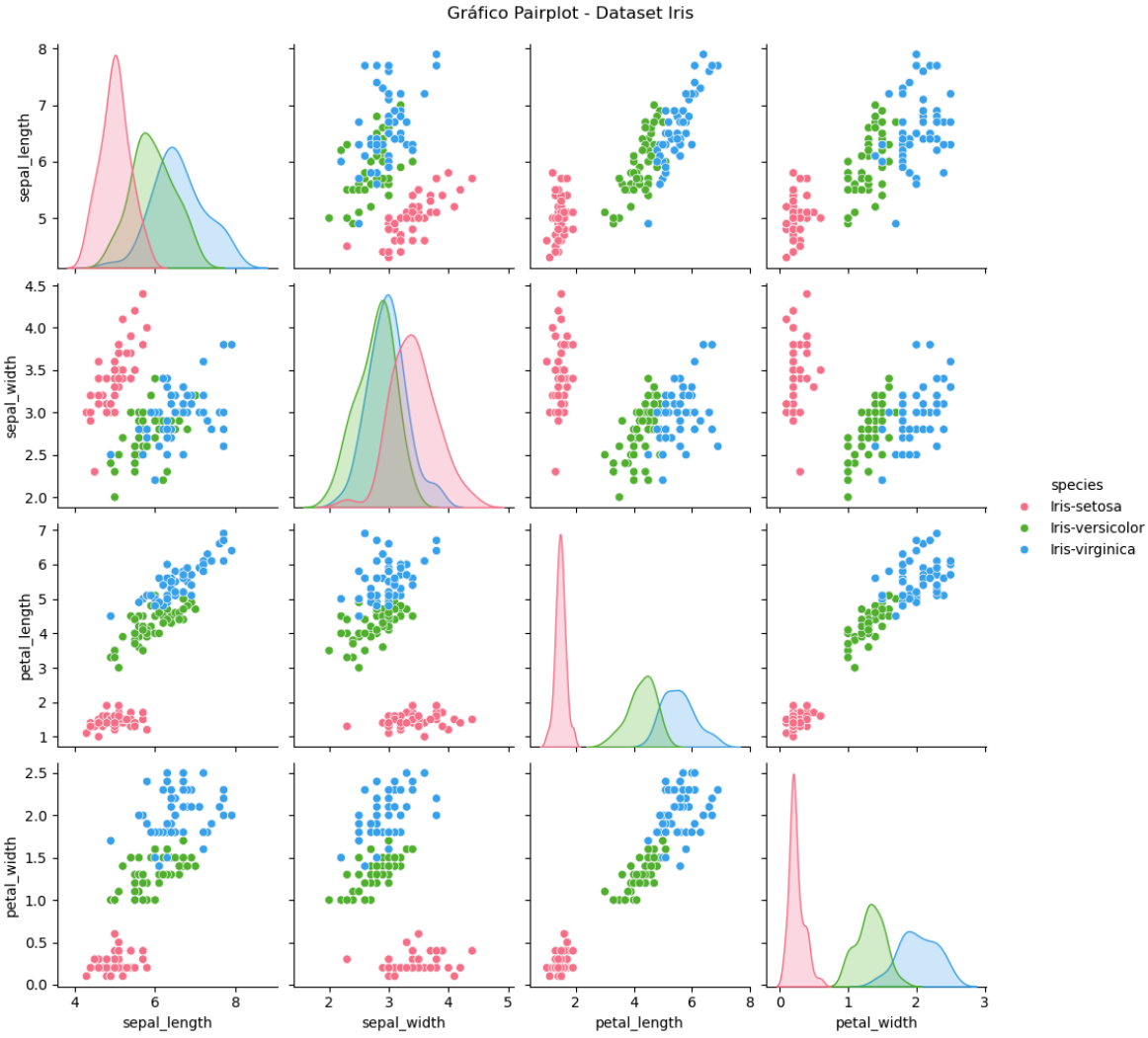
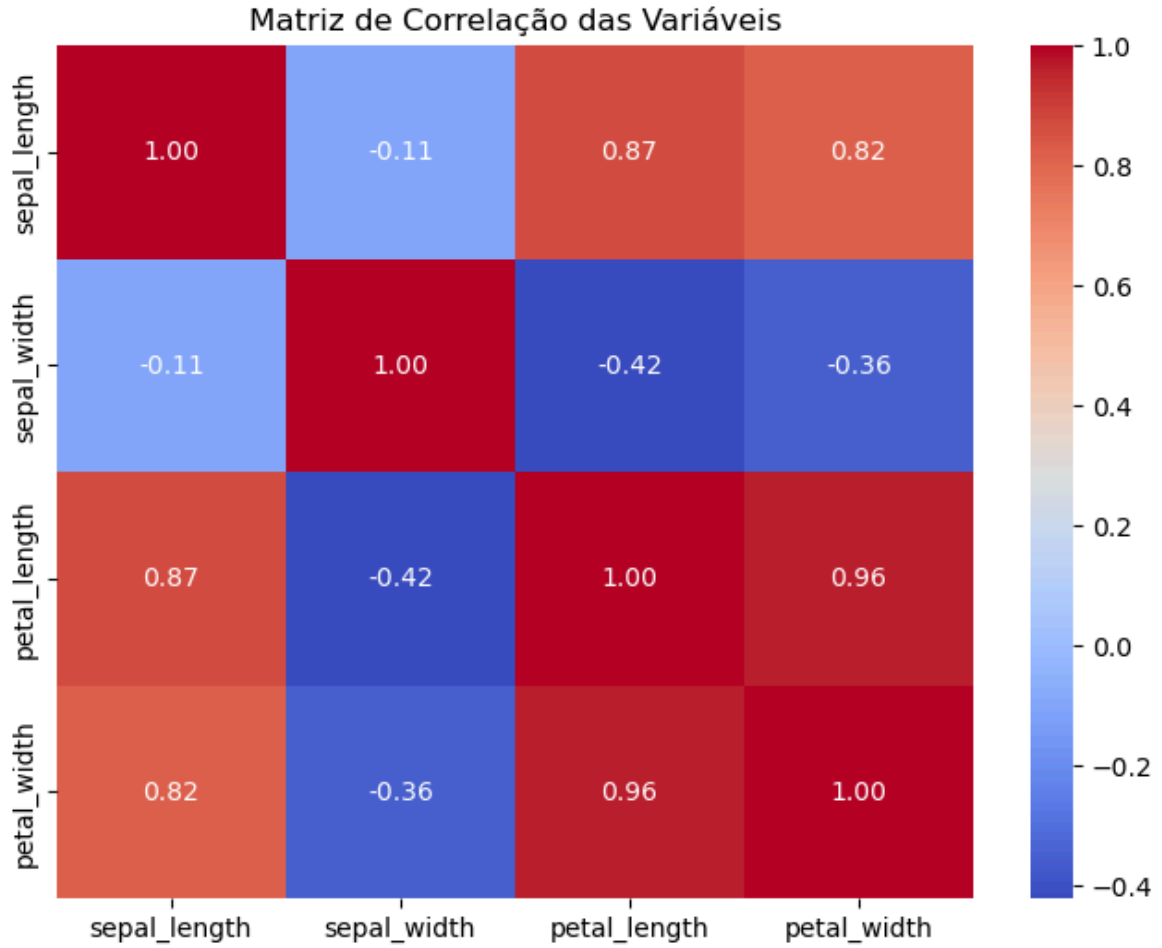
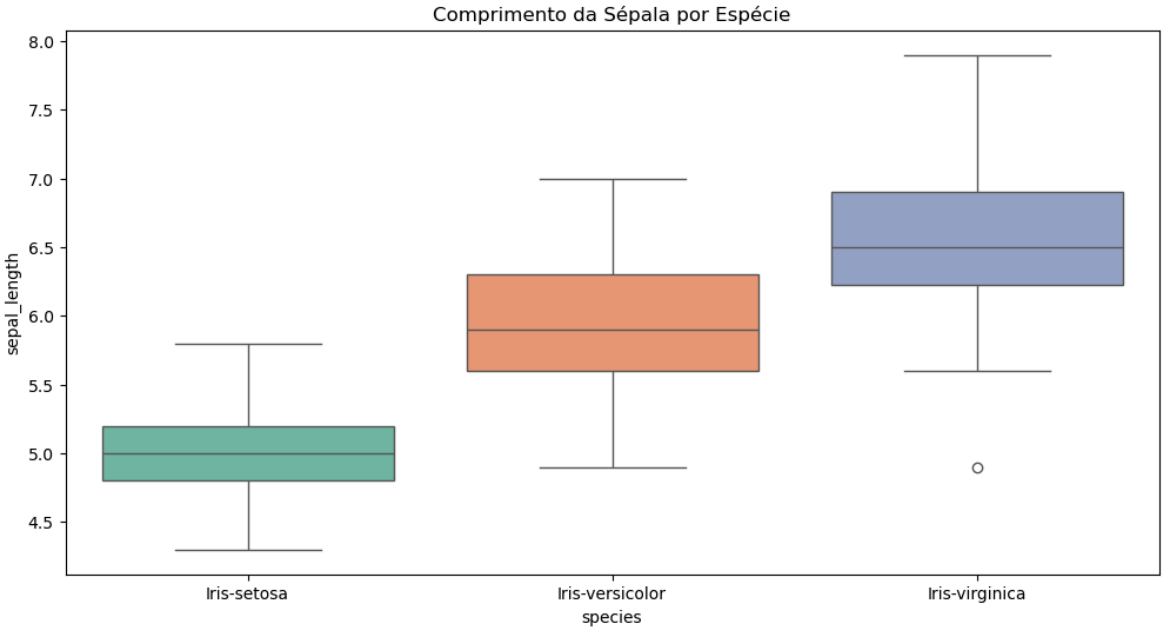


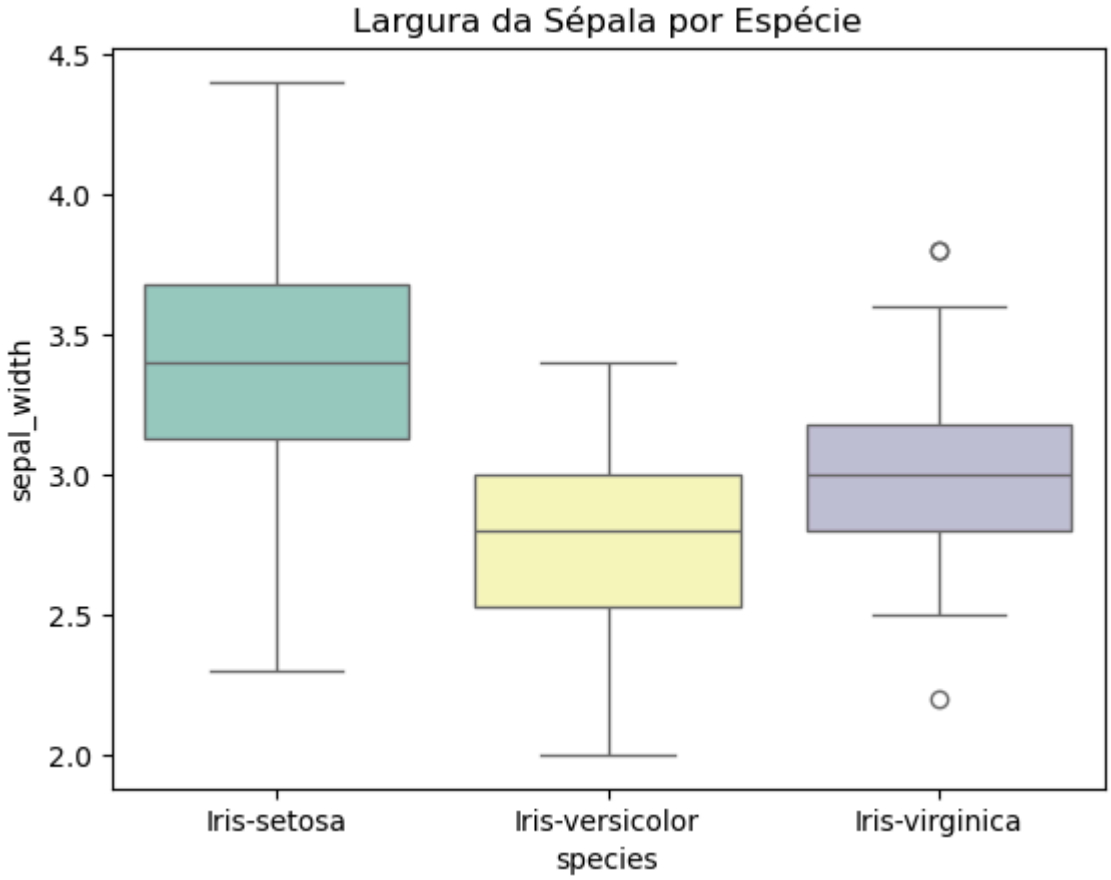
Gráfico Pairplot - Dataset Iris
sepal_length
sepal_width
petal_length
petal_width
species
Iris-setosa
Iris-versicolor
Iris-virginica

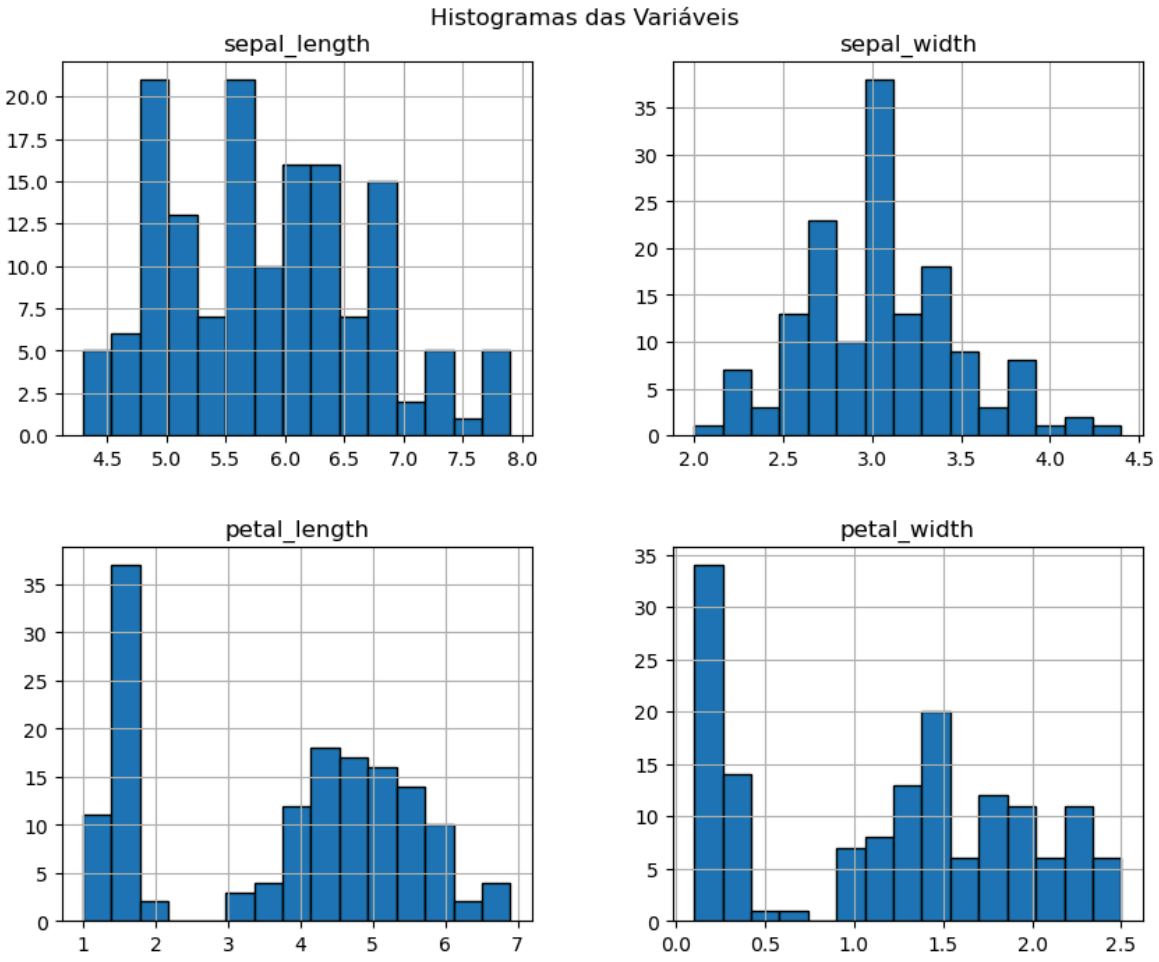

Matriz de Correlação das Variáveis
sepal_length
sepal_width
petal_length
petal_width
1.00
-0.11
0.87
0.82
-0.11
1.00
-0.42
-0.36
0.87
-0.42
1.00
0.96
0.82
-0.36
0.96
1.00

```
C:\Users\avani\AppData\Local\Temp\ipykernel_10568\4041631069.py:50: FutureWarnin
g:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effe
ct.

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




```
C:\Users\avani\AppData\Local\Temp\ipykernel_10568\4041631069.py:54: FutureWarnin
g:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.boxplot(data=data, x='species', y='sepal_width', palette='Set3')
```


Largura da Sépala por Espécie
sepal_width
4.5
4.0
3.5
3.0
2.5
2.0
Iris-setosa
Iris-versicolor
Iris-virginica
species


Histogramas das Variáveis
sepal_length
sepal_width
petal_length
petal_width

```
Estatísticas Agrupadas por Espécie (Média e Desvio Padrão):
                sepal_length           sepal_width           petal_length  \
                        mean       std        mean       std         mean
species
Iris-setosa            5.006  0.352490       3.418  0.381024        1.464
Iris-versicolor        5.936  0.516171       2.770  0.313798        4.260
Iris-virginica         6.588  0.635880       2.974  0.322497        5.552

                          petal_width
                     std        mean       std
species
Iris-setosa      0.173511       0.244  0.107210
Iris-versicolor  0.469911       1.326  0.197753
Iris-virginica   0.551895       2.026  0.274650
Estatísticas salvas no arquivo 'iris_summary_stats.csv'.
```

In [ ]: