Code Explanation

1. Importing Libraries

The code begins by importing the necessary libraries:

python

CopyEdit

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inline

import seaborn as sns
```

- `numpy`: A library for numerical computations.
- `pandas`: Used for data manipulation and analysis, particularly to handle DataFrames.
- `matplotlib.pyplot`: A library for plotting graphs and visualizations.
- `%matplotlib inline`: This is a magic command to display the plots inline in Jupyter notebooks.
- `seaborn`: A data visualization library based on matplotlib that provides a higher-level interface for creating informative statistical plots.

2. Loading the Iris Dataset

The dataset is loaded using the seaborn library:

python

CopyEdit

```
dataset = sns.load_dataset('iris')
```

- The `sns.load_dataset('iris')` function loads the famous Iris flower dataset, which contains information about the lengths and widths of the sepals and petals for three different species of Iris flowers.

3. Exploring the Dataset

The following code snippet shows a preview of the dataset and its data types:

python

CopyEdit

```
dataset.head()

dataset.info()
```

- `dataset.head()` displays the first few rows of the dataset, allowing you to get an overview of its structure.
- `dataset.info()` provides information about the DataFrame, such as the number of non-null entries, column data types, and memory usage.

4. Histograms for Each Feature

The next block of code creates histograms for the features of the dataset:

python

CopyEdit

```
fig, axes = plt.subplots(2, 2, figsize=(16, 8))


axes[0,0].set_title("Distribution of Sepal Length")

axes[0,0].hist(dataset["sepal_length"]);


axes[0,1].set_title("Distribution of Sepal Width")
```

```
axes[0,1].hist(dataset["sepal_width"]);


axes[1,0].set_title("Distribution of Petal Length")

axes[1,0].hist(dataset["petal_length"]);


axes[1,1].set_title("Distribution of Petal Width")

axes[1,1].hist(dataset["petal_width"]);
```

- `fig, axes = plt.subplots(2, 2, figsize=(16, 8))`: This line creates a 2x2 grid of subplots (four total), with a specified figure size of (16, 8).
- The next lines plot histograms for each of the four features in the dataset (`sepal_length`, `sepal_width`, `petal_length`, and `petal_width`) in separate subplots.
- `axes[x, y].set_title()` is used to set the title for each subplot, while `axes[x, y].hist()` plots the histogram for the respective feature.

5. Boxplots for Each Feature

The next section of the code creates boxplots for each feature with respect to the species of the Iris flower:

python

CopyEdit

```
fig, axes = plt.subplots(2, 2, figsize=(16,9))

axes[0,0].set_title("Distribution of Sepal Length")

sns.boxplot(y="sepal_length", x="species", data=dataset, orient='v', ax=axes[0, 0])
```

```
axes[0,1].set_title("Distribution of Sepal Width")

sns.boxplot(y="sepal_width", x="species", data=dataset, orient='v', ax=axes[0, 1])


axes[1,0].set_title("Distribution of Petal Length")

sns.boxplot(y="petal_length", x="species", data=dataset, orient='v', ax=axes[1, 0])


axes[1,1].set_title("Distribution of Petal Width")

sns.boxplot(y="petal_width", x="species", data=dataset, orient='v', ax=axes[1, 1])


plt.show()
```

- `sns.boxplot()`: This function is used to create a boxplot for each feature in the dataset, grouped by species (`x="species"`).
- `orient='v'`: The `orient` parameter specifies the orientation of the box plot. 'v' means vertical, and 'h' would mean horizontal.
- `ax=axes[x, y]`: This specifies the subplot in which the boxplot is drawn.
- These boxplots show the distribution of each feature for each Iris species. They help us identify the median, quartiles, and outliers in the data.

6. Inferences

The dataset consists of four features (numeric) and one target variable `species` (nominal). Here are the results based on the analysis:

1. **List of Features and Their Types:**
    - **`sepal_length`: Numeric**
    - **`sepal_width`: Numeric**
    - **`petal_length`: Numeric**
    - **`petal_width`: Numeric**
    - **`species`: Nominal (Categorical)**
2. **Histograms:**
    - **Sepal Length: The distribution is relatively symmetric, with a higher frequency around 5.0 to 6.0.**
    - **Sepal Width: The distribution is slightly skewed with a peak around 3.0.**
    - **Petal Length: There is a peak around 1.5 to 5.0, with some longer petals observed, especially for the "virginica" species.**
    - **Petal Width: Similar to petal length, petal width has a peak around 0.2-1.0, with wider petals in "virginica."**
3. **Boxplots:**
    - **The boxplots show that the "versicolor" species has a similar sepal length to "virginica," but "setosa" has a distinctively smaller range.**
    - **"Virginica" has the largest petal length and width, with a few outliers in both the petal and sepal features.**
    - **"Setosa" tends to have the smallest values for petal length and width compared to the other species.**
4. **Outliers:**

- **The boxplots reveal a few outliers, especially for "virginica," where petal length and width have some extreme values compared to the rest of the data.**