

```
from sklearn.datasets import load_iris
import pandas as pd

# Load the Iris dataset
iris = load_iris()

# Convert the dataset into a pandas DataFrame
data = pd.DataFrame(data=iris.data, columns=iris.feature_names)
data['target'] = iris.target # Add target column

# Save the DataFrame to a CSV file
data.to_csv('iris.csv', index=False)
```

Data Cleaning and Preprocessing:

```
# Load data from the CSV file
data = pd.read_csv('iris.csv')

# Remove duplicates
data = data.drop_duplicates()

# Fill missing values with the mean of the column
data.fillna(data.mean(), inplace=True)

# Remove rows with missing values
data = data.dropna()

# Convert string column to datetime (not applicable to Iris dataset)
# Example: data['date_column'] = pd.to_datetime(data['date_column'])
```

```
print(data.columns)

Index(['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',
      'petal width (cm)', 'target'],
      dtype='object')
```

Data Selection and Filtering:

```
# Select rows where a condition is met
filtered_data = data[data['sepal length (cm)'] > 5.0]

# Select rows based on multiple conditions
filtered_data = data[(data['sepal length (cm)'] > 5.0) & (data['sepal width (cm)'] == 3.0)]

# Select specific columns
selected_columns = data[['sepal length (cm)', 'sepal width (cm)']]

# Select rows based on index
row = data.loc[5] # Select row with index 5
```

Grouping and Aggregation:

```
# Group data by species
grouped_data = data.groupby('target')

# Calculate mean, sum, etc. within each group
mean_values = grouped_data.mean()
sum_values = grouped_data.sum()
```

Merging and Joining DataFrames:

```
# Merge two DataFrames on a common column
# (Example: Merge with another copy of the same dataset for illustration purposes)
merged_data = pd.merge(data, data, on='target')

# Concatenate DataFrames along rows or columns
concatenated_data = pd.concat([data.head(), data.tail()], axis=0) # Concatenate along rows
concatenated_data = pd.concat([data.head(), data.tail()], axis=1) # Concatenate along columns
```

Data Visualization:

```
import matplotlib.pyplot as plt
```

```
# Plot histogram
data['sepal length (cm)'].plot(kind='hist', bins=20)
plt.title('Histogram of Sepal Length')
plt.xlabel('Length (cm)')
plt.ylabel('Frequency')
plt.show()

# Plot line chart (not applicable to Iris dataset)
# Example: data.plot(x='date_column', y='numeric_column')
```

