

```

# analysis.py
"""
Analyzing Data with Pandas and Visualizing Results with Matplotlib
Assignment Script
"""

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

# -----
# Task 1: Load and Explore Dataset
# -----
try:
    # Load dataset (Iris dataset from seaborn for demo purposes)
    df = sns.load_dataset("iris")

    print("✅ Dataset Loaded Successfully!\n")
    print("First 5 rows of dataset:")
    print(df.head())

    # Check data types and missing values
    print("\nData Info:")
    print(df.info())

    print("\nMissing Values:")
    print(df.isnull().sum())

    # Fill missing values (if any)
    df = df.fillna(df.mean(numeric_only=True))

except FileNotFoundError:
    print("❌ Error: File not found. Please provide a valid dataset.")
    exit()
except Exception as e:
    print(f"❌ An error occurred: {e}")
    exit()

# -----
# Task 2: Basic Data Analysis
# -----
print("\nBasic Statistics:")
print(df.describe())

```

```

# Grouping example (average petal length per species)
grouped = df.groupby("species")["petal_length"].mean()
print("\nAverage Petal Length per Species:")
print(grouped)

# -----
# Task 3: Data Visualizations
# -----

# 1. Line chart (example using cumulative sum to simulate trend)
df["cum_sum"] = df["sepal_length"].cumsum()
plt.figure(figsize=(8, 5))
plt.plot(df.index, df["cum_sum"], label="Cumulative Sepal Length")
plt.title("Line Chart: Cumulative Sepal Length")
plt.xlabel("Index")
plt.ylabel("Cumulative Sum")
plt.legend()
plt.savefig("line_chart.png")
plt.close()

# 2. Bar chart (average petal length per species)
grouped.plot(kind="bar", figsize=(8, 5), title="Average Petal Length per Species")
plt.ylabel("Average Petal Length")
plt.savefig("bar_chart.png")
plt.close()

# 3. Histogram (distribution of sepal length)
df["sepal_length"].plot(kind="hist", bins=20, figsize=(8, 5), title="Sepal Length Distribution")
plt.xlabel("Sepal Length")
plt.savefig("histogram.png")
plt.close()

# 4. Scatter plot (sepal length vs petal length)
plt.figure(figsize=(8, 5))
plt.scatter(df["sepal_length"], df["petal_length"], c="blue", alpha=0.6)
plt.title("Scatter Plot: Sepal Length vs Petal Length")
plt.xlabel("Sepal Length")
plt.ylabel("Petal Length")
plt.savefig("scatter_plot.png")
plt.close()

print("\n✅ Analysis Complete! Charts saved as:")
print(" - line_chart.png")
print(" - bar_chart.png")

```

```
print(" - histogram.png")  
print(" - scatter_plot.png")
```