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

# Step 1: Data Collection
# Assuming you have a CSV file named 'power_plant_data.csv' in the same directory
data = pd.read_csv('power_plant_data.csv')

# Step 2: Data Preprocessing
# Perform any necessary data cleaning and transformation here
# Handle missing values, outliers, data type conversions, etc.

# Step 3: Exploratory Data Analysis (EDA)
# Example code for EDA

# Display the first few rows of the dataset
print(data.head())

# Check the dimensions of the dataset (rows, columns)
print(data.shape)

# Summary statistics of the dataset
print(data.describe())

# Correlation matrix
correlation_matrix = data.corr()
print(correlation_matrix)

# Example visualization

# Distribution of capacity MW
```

```
plt.figure(figsize=(8, 6))
sns.histplot(data['capacity_mw'], kde=True)
plt.xlabel('Capacity (MW)')
plt.ylabel('Frequency')
plt.title('Distribution of Power Plant Capacity')
plt.show()

# Boxplot of primary fuel types by country
plt.figure(figsize=(12, 8))
sns.boxplot(data['country'], data['primary_fuel'])
plt.xticks(rotation=90)
plt.xlabel('Country')
plt.ylabel('Primary Fuel')
plt.title('Primary Fuel Types by Country')
plt.show()

# Step 4: Perform further analysis based on your research questions, such as analyzing the distribution of power plant types or comparing capacity by renewable vs. non-renewable sources.

# ... (Continue with your specific analysis steps)

# Step 10: Conclusion
# Summarize your findings and insights.
```