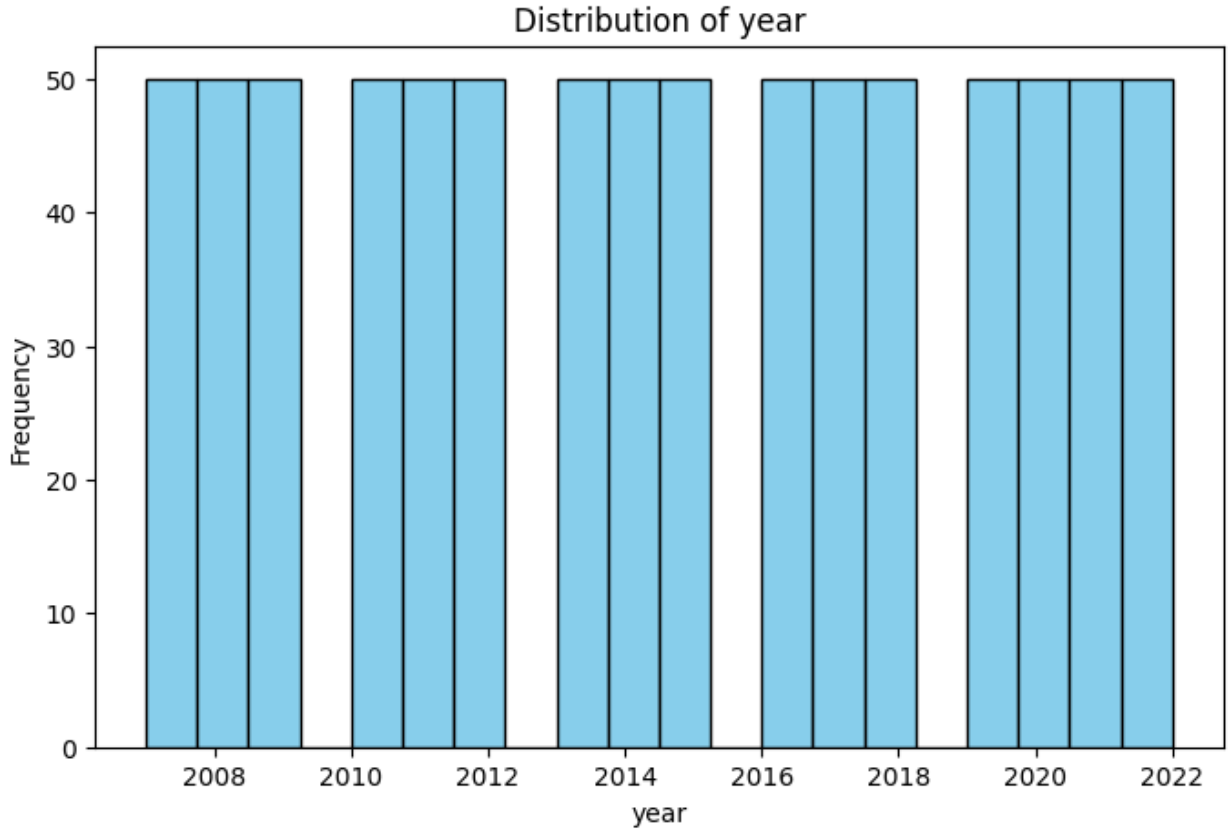
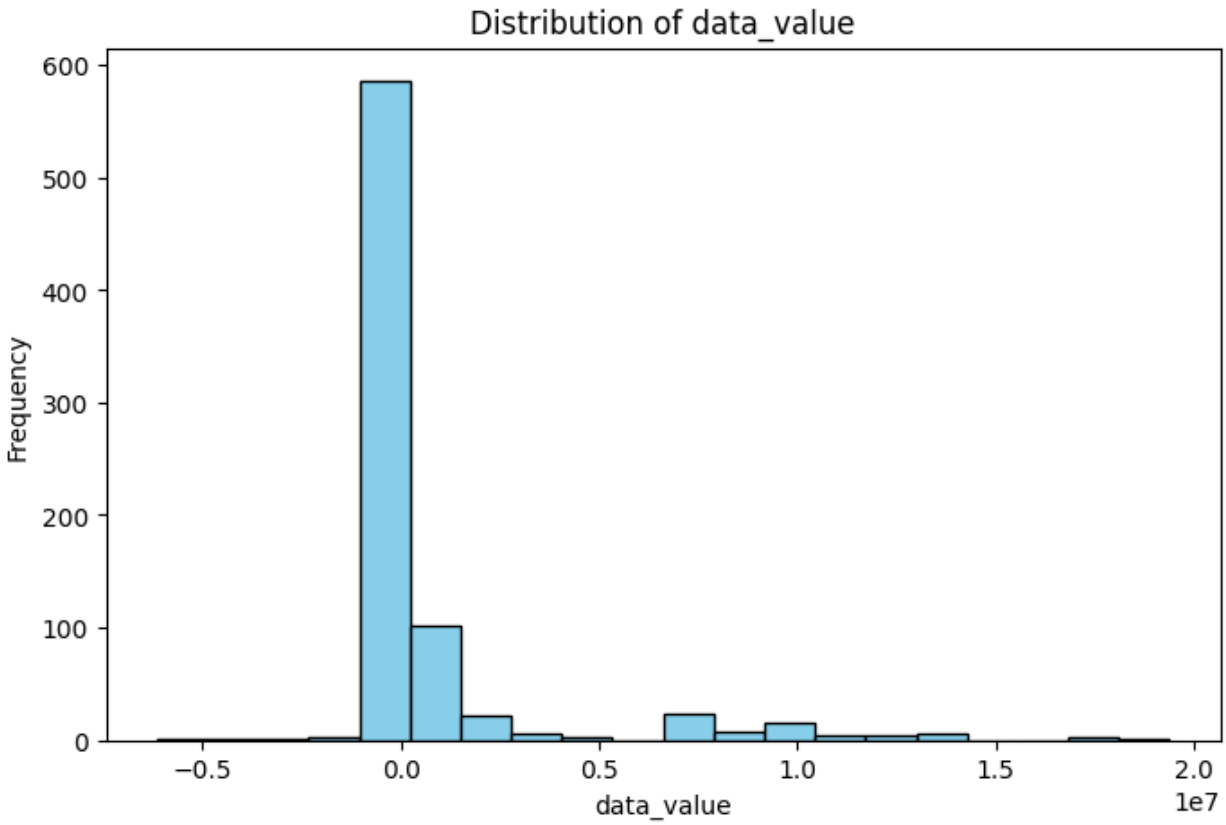
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Import the dataset
# Load the dataset
file_path = "renewable-energy-stock-account (1).csv"
data = pd.read_csv('/content/renewable-energy-stock-account (1).csv')

# Display dataset structure
print("Dataset Structure:")
print(data.info())
print("\nFirst few rows of the dataset:")
print(data.head())

# Step 2: Basic Visualizations (Matplotlib)
# Histogram for numerical features
numerical_cols = data.select_dtypes(include=['float64',
'int64']).columns
for col in numerical_cols:
    plt.figure(figsize=(8, 5))
    plt.hist(data[col].dropna(), bins=20, color='skyblue',
edgecolor='black')
    plt.title(f'Distribution of {col}')
    plt.xlabel(col)
    plt.ylabel('Frequency')
    plt.show()

# Bar Plot for categorical data
categorical_cols = data.select_dtypes(include=['object',
'category']).columns
for col in categorical_cols:
    plt.figure(figsize=(8, 5))
    data[col].value_counts().plot(kind='bar', color='orange')
    plt.title(f'Category Distribution for {col}')
    plt.xlabel(col)
    plt.ylabel('Count')
    plt.show()

# Pie Chart for a categorical feature
for col in categorical_cols:
    if data[col].nunique() <= 5:  # Limit pie charts to columns with
fewer categories
        plt.figure(figsize=(8, 5))
        data[col].value_counts().plot(kind='pie', autopct='%1.1f%%',
colors=sns.color_palette('pastel'))
        plt.title(f'Proportion of Categories in {col}')
        plt.ylabel('')
        plt.show()
```

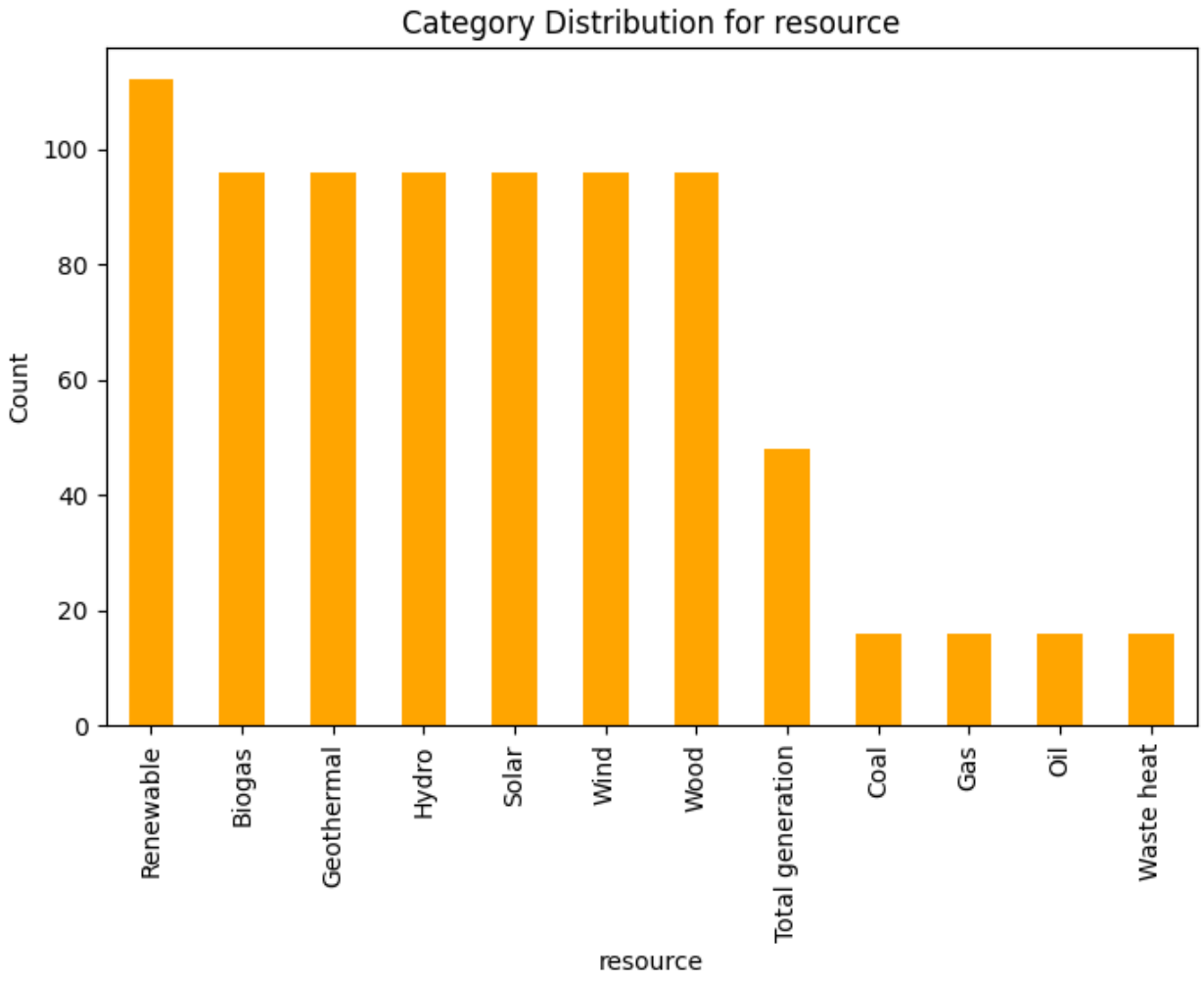
```python
# Step 3: Advanced Visualizations (Seaborn)
# Box Plot for numerical features
for col in numerical_cols:
    plt.figure(figsize=(8, 5))
    sns.boxplot(data[col], color='skyblue')
    plt.title(f'Box Plot of {col}')
    plt.show()

# Violin Plot
for col in numerical_cols:
    plt.figure(figsize=(8, 5))
    sns.violinplot(data[col], color='lightgreen')
    plt.title(f'Violin Plot of {col}')
    plt.show()

# Pair Plot for relationships
sns.pairplot(data.select_dtypes(include=['float64', 'int64']))
plt.suptitle("Pairwise Relationships")
plt.show()

# Heatmap for correlations
plt.figure(figsize=(10, 8))
# Select only numerical columns for correlation
numerical_data = data.select_dtypes(include=['float64', 'int64'])
correlation_matrix = numerical_data.corr()  # Calculate correlation
for numerical data
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Heatmap of Correlations")
plt.show()

# Step 4: Customizations
# Titles, labels, and legends are added in individual plots.

# Step 5: Conclusion
print("\nConclusion:")
print("Based on the visualizations, we can observe the distributions
and relationships among the features. Significant trends and anomalies
are highlighted in the visual outputs.")
```

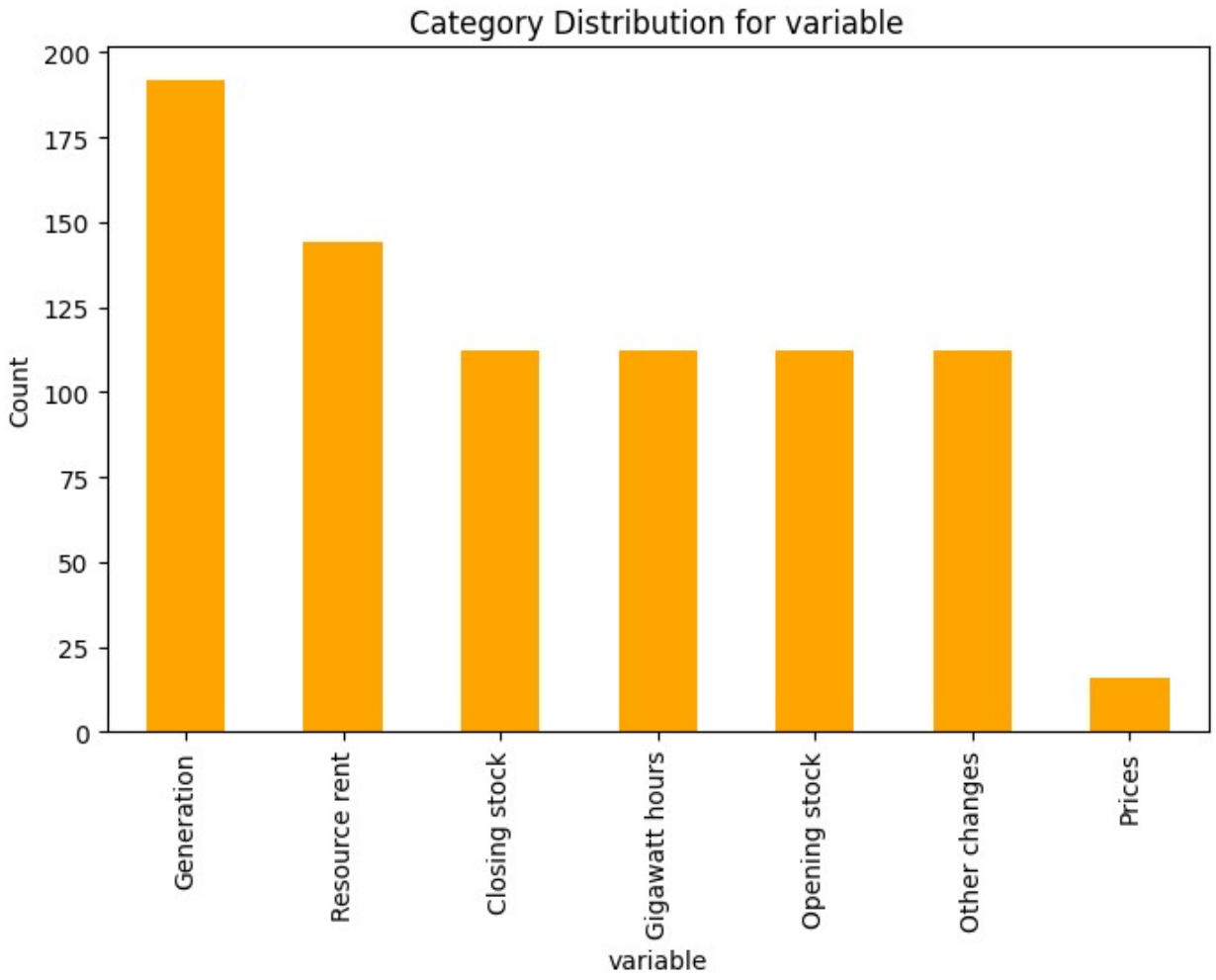
```
Dataset Structure:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 800 entries, 0 to 799
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   year        800 non-null    int64
 1   resource    800 non-null    object
 2   variable    800 non-null    object
 3   units       800 non-null    object
 4   magnitude   800 non-null    object
```

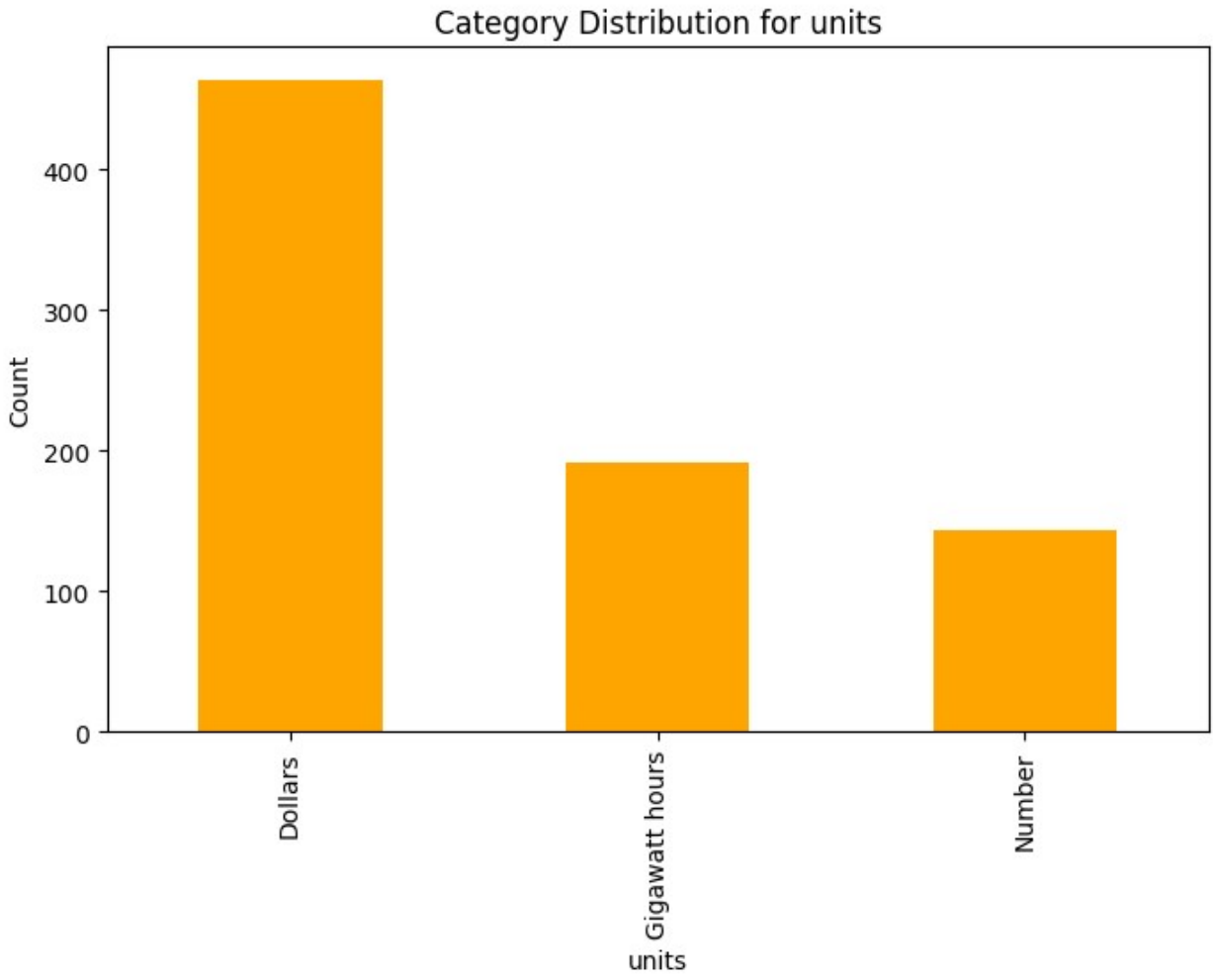
```
 5   source      800 non-null    object
 6   data_value  786 non-null    float64
 7   flag        800 non-null    object
dtypes: float64(1), int64(1), object(6)
memory usage: 50.1+ KB
None

First few rows of the dataset:
   year resource        variable           units   magnitude  \
0  2007   Biogas   Closing stock         Dollars   Thousands
1  2007   Biogas      Generation  Gigawatt hours      Actual
2  2007   Biogas  Gigawatt hours          Number  Proportion
3  2007   Biogas   Opening stock         Dollars   Thousands
4  2007   Biogas   Other changes         Dollars   Thousands

                    source  data_value flag
0  Environmental Accounts    64848.00    F
1                    MBIE      224.00    F
2                    MBIE        0.01    F
3  Environmental Accounts         NaN    F
4  Environmental Accounts         NaN    F
```

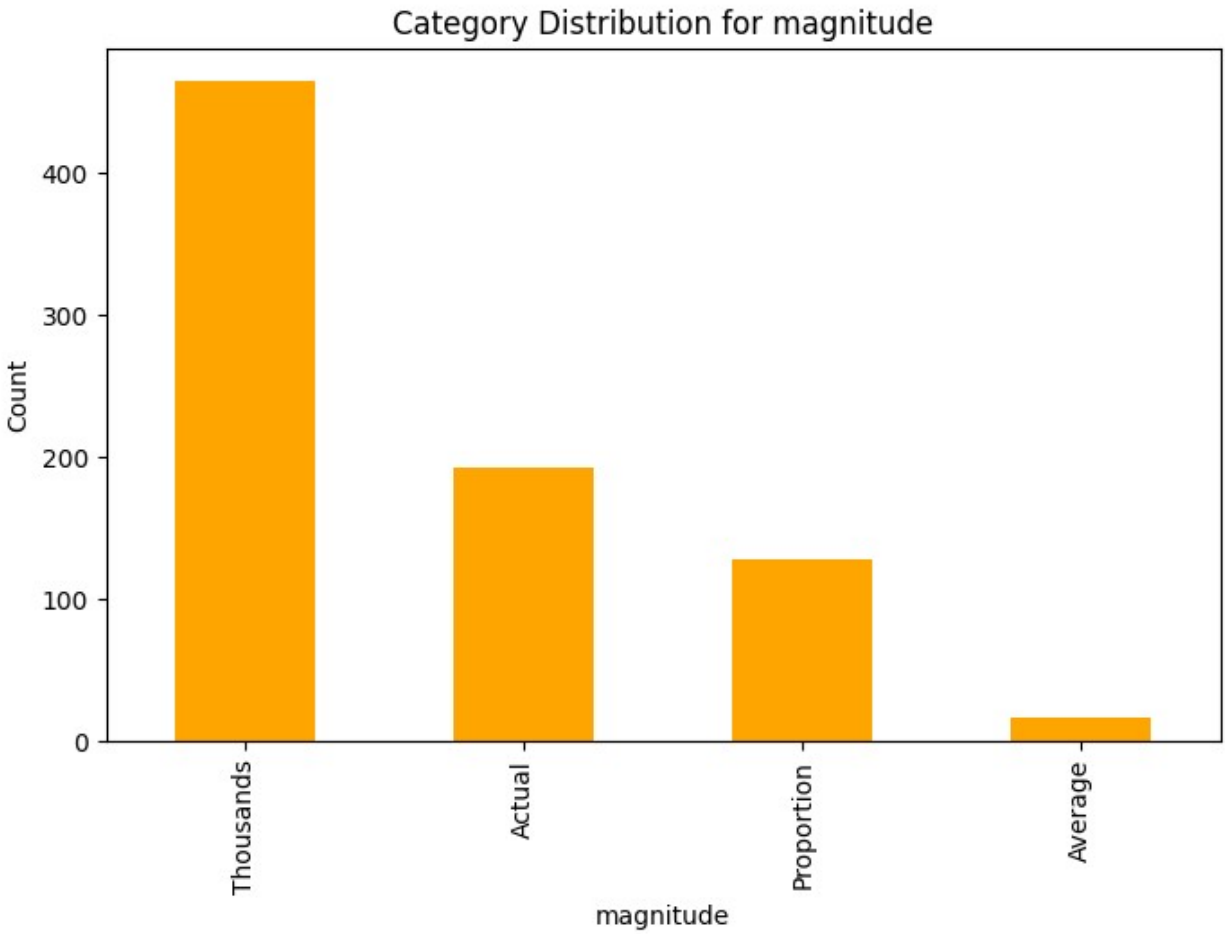


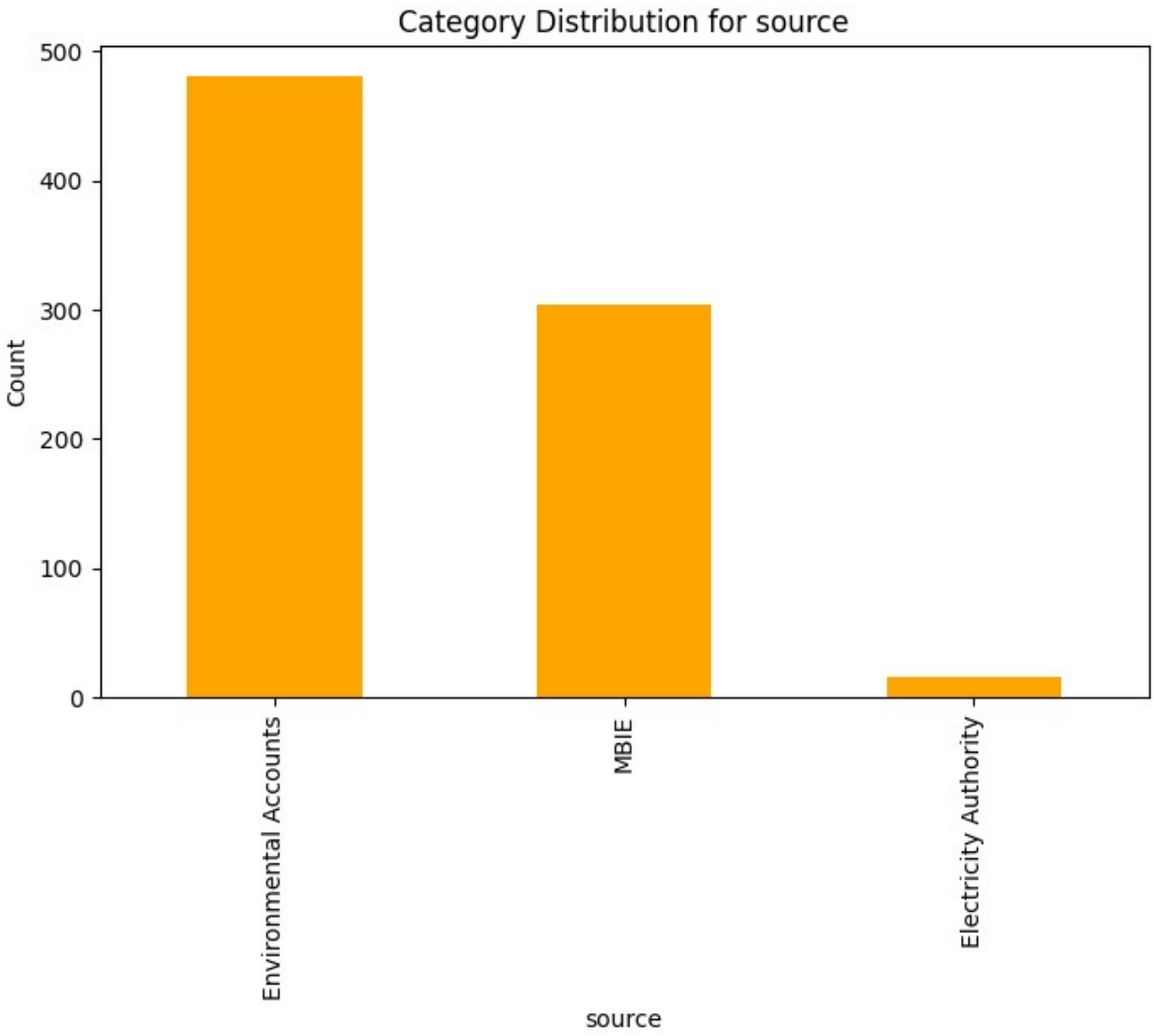

Distribution of data_value
Frequency
600
500
400
300
200
100
0
−0.5
0.0
0.5
1.0
1.5
2.0
data_value
1e7

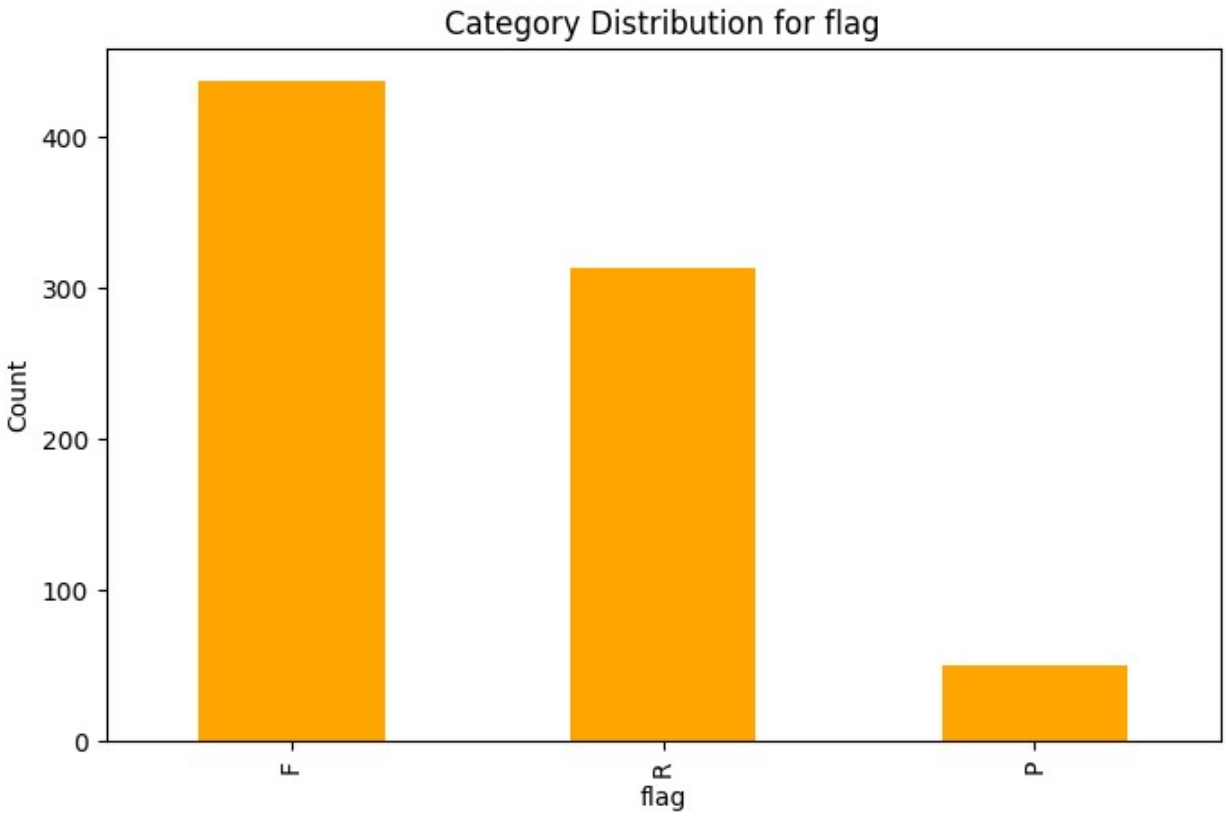

Category Distribution for resource
Count
0
20
40
60
80
100
Renewable
Biogas
Geothermal
Hydro
Solar
Wind
Wood
Total generation
Coal
Gas
Oil
Waste heat
resource

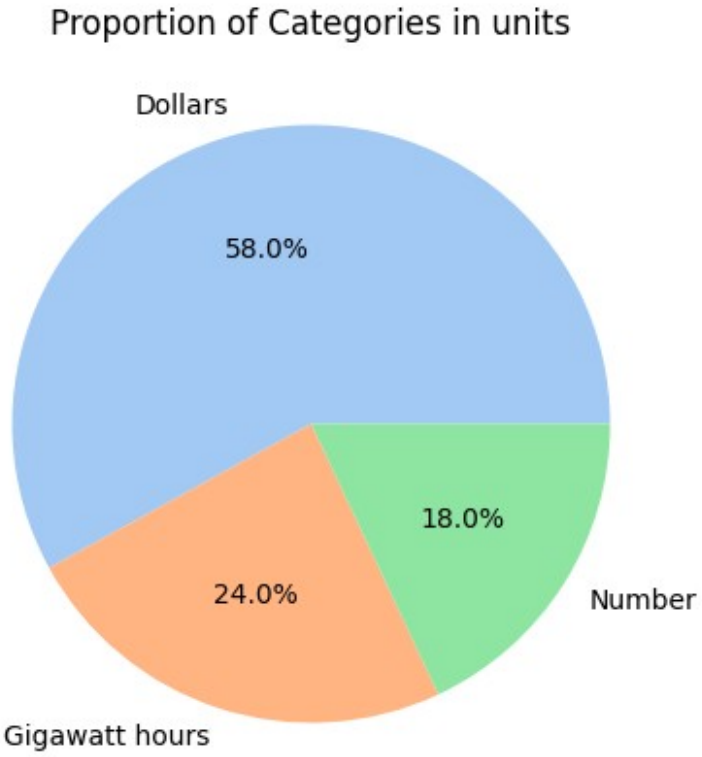

Category Distribution for variable
Count
200
175
150
125
100
75
50
25
0
Generation
Resource rent
Closing stock
Gigawatt hours
Opening stock
Other changes
Prices
variable

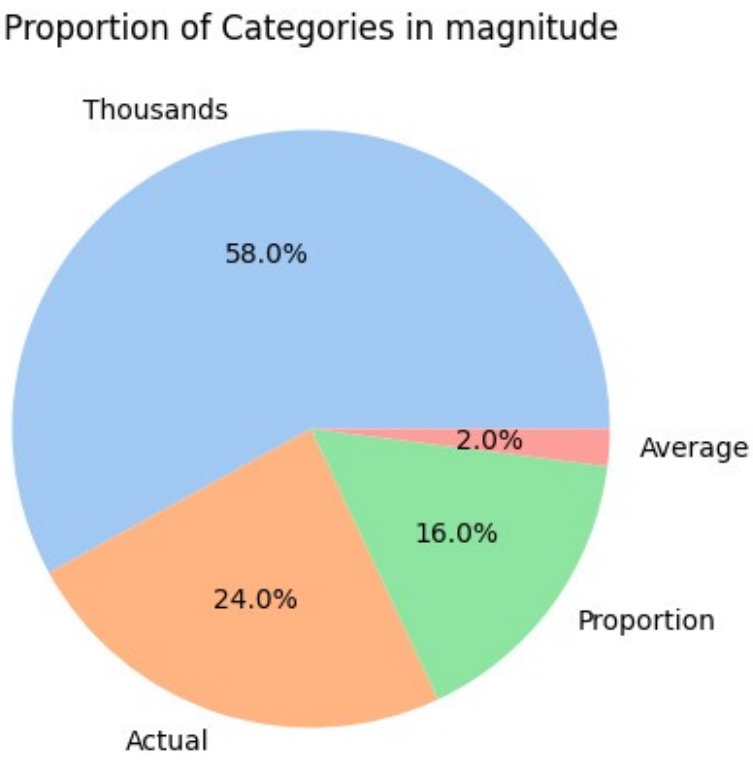

Category Distribution for units
Count
400
300
200
100
0
Dollars
Gigawatt hours
Number
units

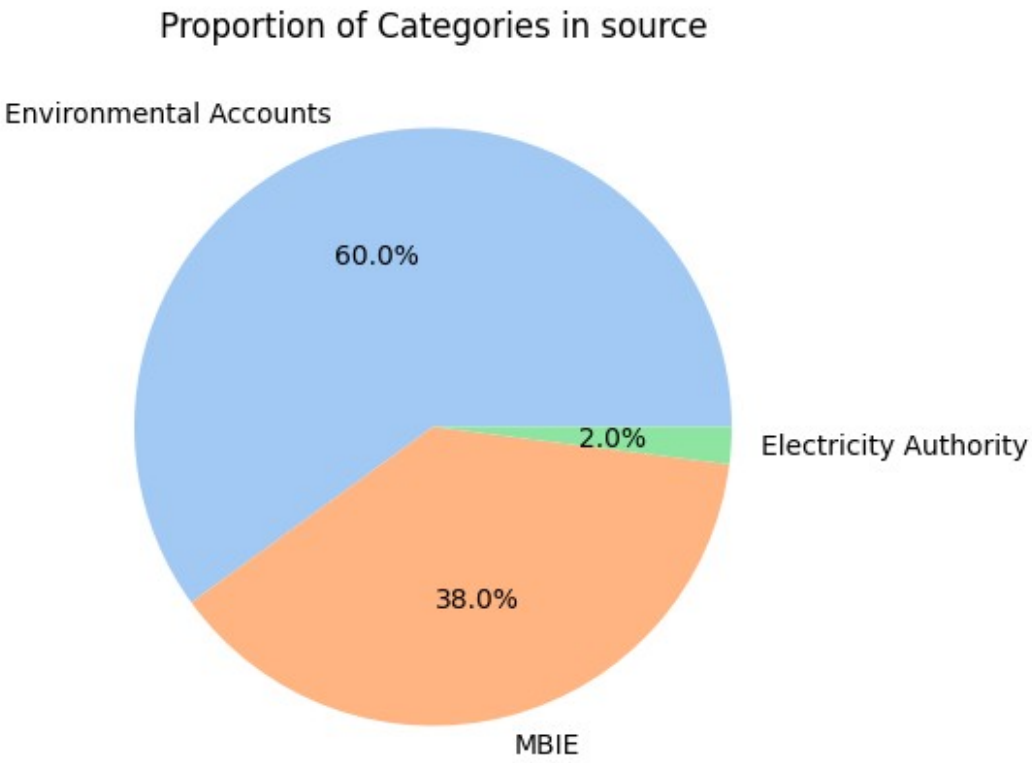

Category Distribution for magnitude
Count
400
300
200
100
0
Thousands
Actual
Proportion
Average
magnitude

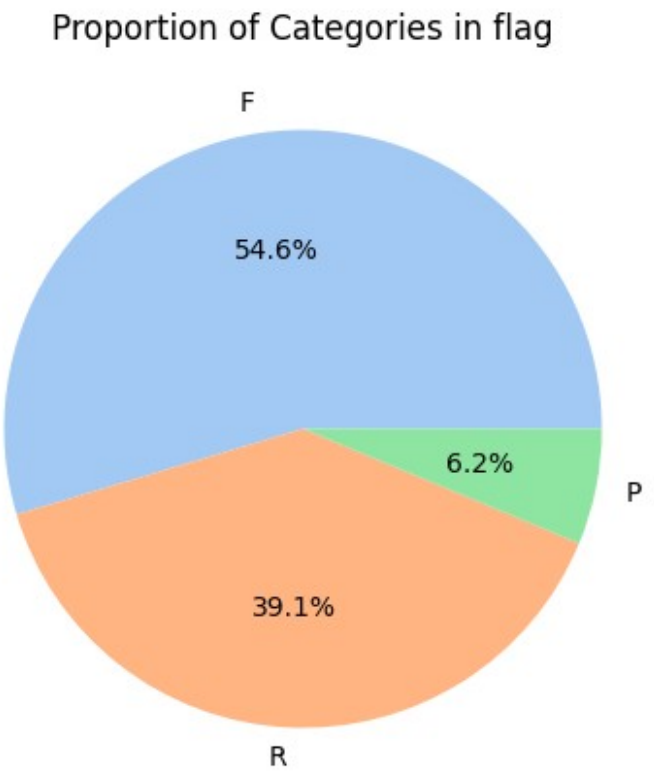

Category Distribution for source
500
400
300
200
100
0
Count
Environmental Accounts
MBIE
Electricity Authority
source

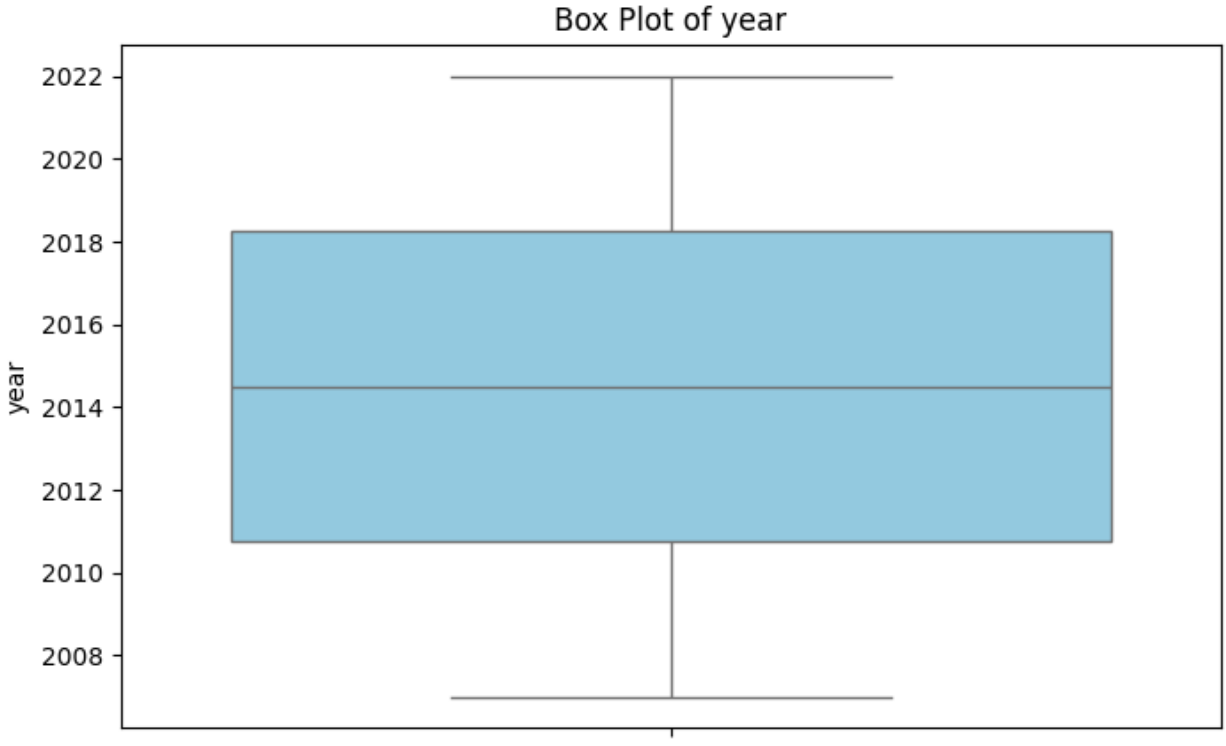

Category Distribution for flag
Count
400
300
200
100
0
F
R
P
flag

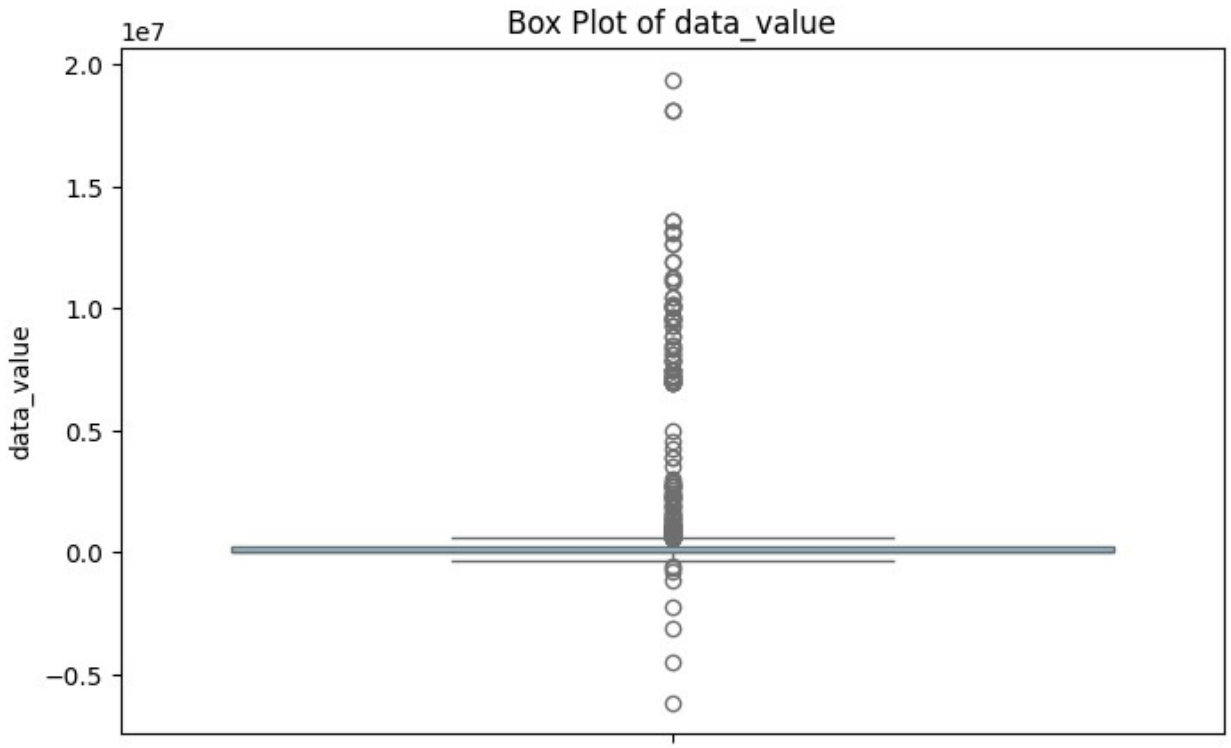

Proportion of Categories in units
Dollars
58.0%
18.0%
24.0%
Number
Gigawatt hours

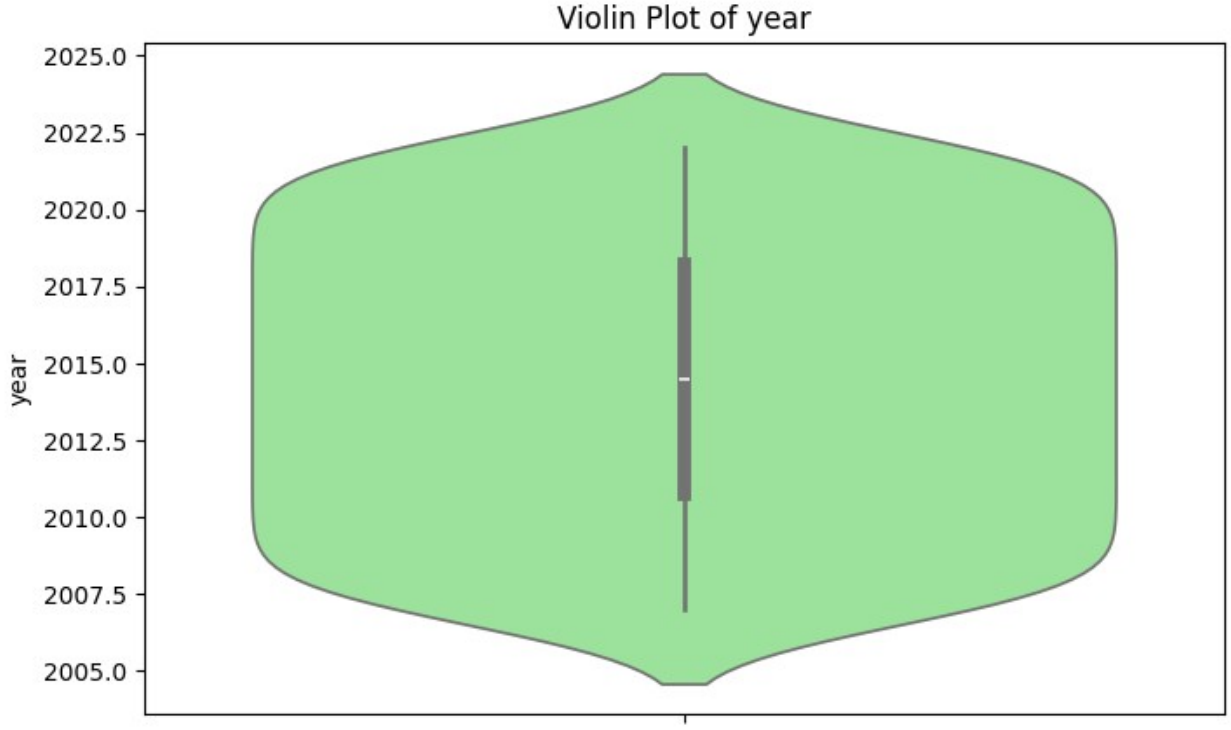

Proportion of Categories in magnitude
Thousands
58.0%
2.0%
Average
16.0%
Proportion
24.0%
Actual

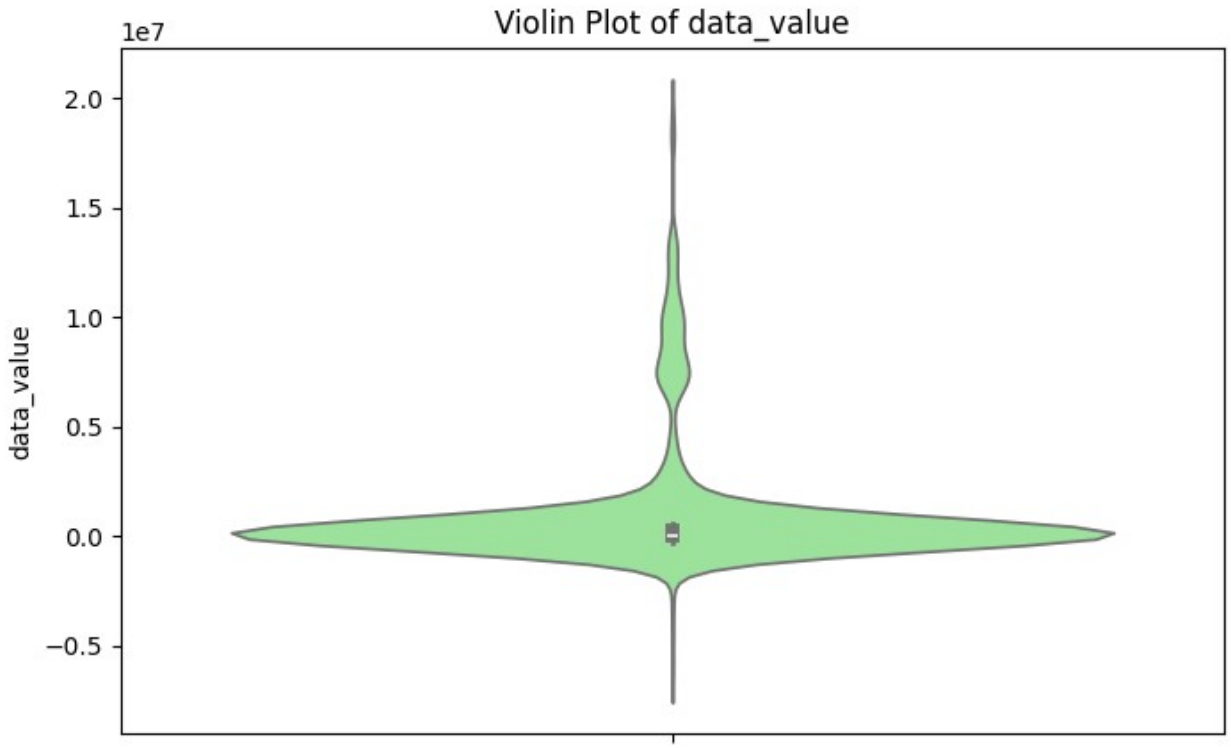

Proportion of Categories in source
Environmental Accounts
60.0%
2.0%
Electricity Authority
38.0%
MBIE

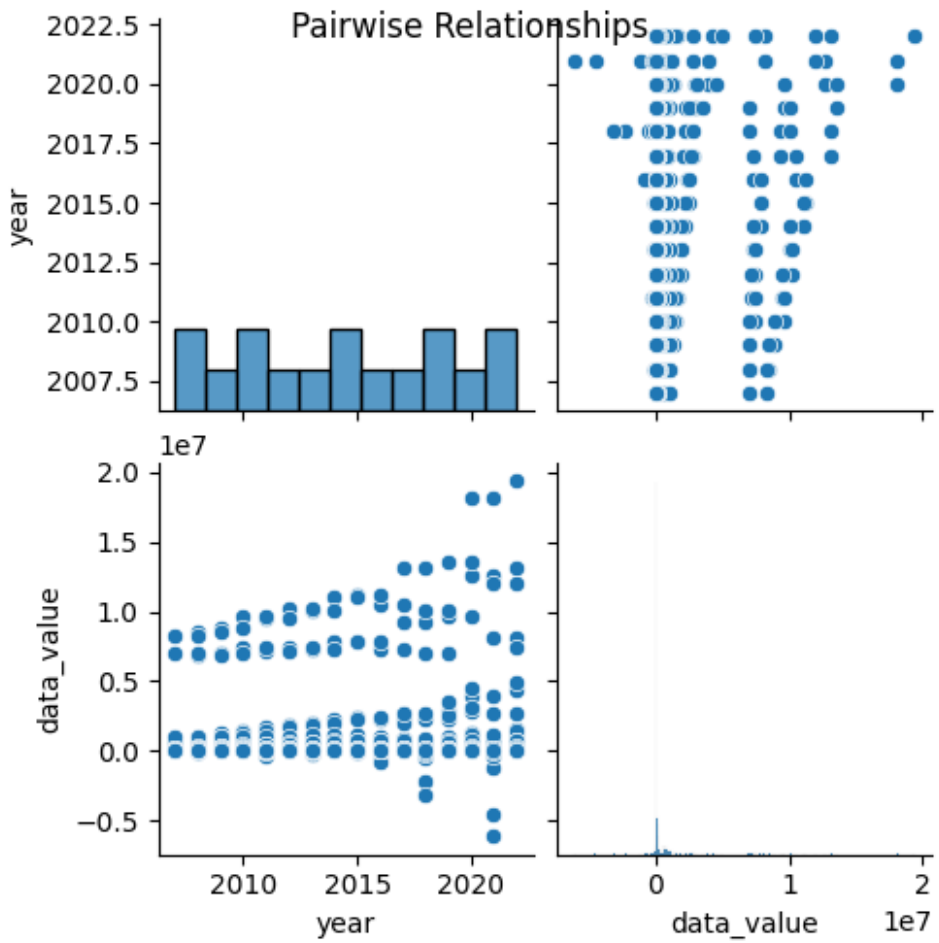

Proportion of Categories in flag
F
54.6%
6.2%
P
39.1%
R

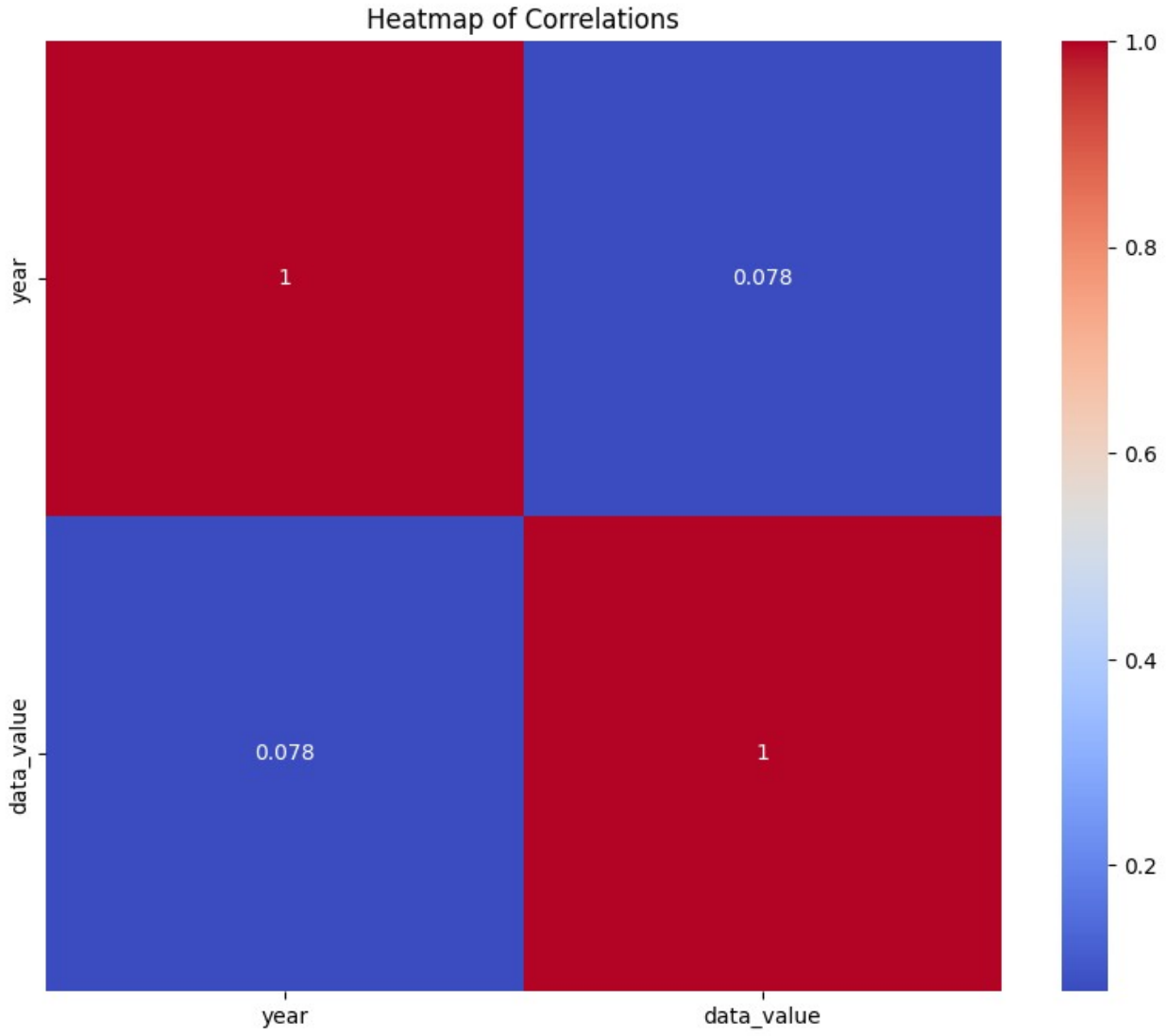

Box Plot of year
year
2022
2020
2018
2016
2014
2012
2010
2008


Box Plot of data_value
1e7
2.0
1.5
1.0
0.5
0.0
−0.5
data_value


Violin Plot of year
2025.0
2022.5
2020.0
2017.5
2015.0
2012.5
2010.0
2007.5
2005.0
year


Violin Plot of data_value
1e7
2.0
1.5
1.0
0.5
0.0
−0.5
data_value


Pairwise Relationships
year
data_value
1e7
2022.5
2020.0
2017.5
2015.0
2012.5
2010.0
2007.5
2.0
1.5
1.0
0.5
0.0
−0.5
2010
2015
2020
0
1
2
1e7




```
Conclusion:
Based on the visualizations, we can observe the distributions and
relationships among the features. Significant trends and anomalies are
highlighted in the visual outputs.
```