

```

import pandas as pd
from sklearn.preprocessing import MinMaxScaler, StandardScaler,
OneHotEncoder

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

# Display basic details about the dataset
print("Dataset Overview:")
print(dataset.info())
print("\nFirst Five Rows:")
print(dataset.head())

# Step 2: Identify Numerical and Categorical Columns
numerical_columns = dataset.select_dtypes(include=['int64',
'float64']).columns
categorical_columns =
dataset.select_dtypes(include=['object']).columns

print("\nNumerical Columns:", numerical_columns.tolist())
print("Categorical Columns:", categorical_columns.tolist())

# Step 3: Feature Scaling
scaler = MinMaxScaler() # MinMaxScaler is a data preprocessing
technique used to scale numerical features to a specific range,
typically between 0 and 1
dataset_scaled = dataset.copy()
dataset_scaled[numerical_columns] =
scaler.fit_transform(dataset[numerical_columns])

print("\nNumerical Features After Scaling:")
print(dataset_scaled[numerical_columns].head())

# Step 4: One-Hot Encoding
encoder = OneHotEncoder(sparse_output=False, drop='first') #
'drop=first' avoids dummy variable trap
categorical_encoded =
encoder.fit_transform(dataset[categorical_columns])

# Convert encoded features to a DataFrame
encoded_df = pd.DataFrame(
    categorical_encoded,
    columns=encoder.get_feature_names_out(categorical_columns)
)

# Combine scaled numerical and encoded categorical data
final_dataset = pd.concat([dataset_scaled[numerical_columns],
encoded_df], axis=1)

```

```
print("\nFinal Dataset After Preprocessing (First Five Rows):")
print(final_dataset.head())
```

*# (Optional) Memory Usage Comparison*

```
print("\nMemory Usage Before Preprocessing:",
dataset.memory_usage(deep=True).sum(), "bytes")
print("Memory Usage After Preprocessing:",
final_dataset.memory_usage(deep=True).sum(), "bytes")
```

Dataset Overview:

```
<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 Five Rows:

|   | 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    |

```
Numerical Columns: ['year', 'data_value']
```

```
Categorical Columns: ['resource', 'variable', 'units', 'magnitude',
'source', 'flag']
```

Numerical Features After Scaling:

|   | year | data_value |
|---|------|------------|
| 0 | 0.0  | 0.243419   |

|   |     |          |
|---|-----|----------|
| 1 | 0.0 | 0.240889 |
| 2 | 0.0 | 0.240881 |
| 3 | 0.0 | NaN      |
| 4 | 0.0 | NaN      |

Final Dataset After Preprocessing (First Five Rows):

|   | year | data_value | resource_Coal | resource_Gas | resource_Geothermal |
|---|------|------------|---------------|--------------|---------------------|
| \ |      |            |               |              |                     |
| 0 | 0.0  | 0.243419   | 0.0           | 0.0          | 0.0                 |
| 1 | 0.0  | 0.240889   | 0.0           | 0.0          | 0.0                 |
| 2 | 0.0  | 0.240881   | 0.0           | 0.0          | 0.0                 |
| 3 | 0.0  | NaN        | 0.0           | 0.0          | 0.0                 |
| 4 | 0.0  | NaN        | 0.0           | 0.0          | 0.0                 |

|   | resource_Hydro | resource_Oil | resource_Renewable | resource_Solar | \ |
|---|----------------|--------------|--------------------|----------------|---|
| 0 | 0.0            | 0.0          | 0.0                | 0.0            |   |
| 1 | 0.0            | 0.0          | 0.0                | 0.0            |   |
| 2 | 0.0            | 0.0          | 0.0                | 0.0            |   |
| 3 | 0.0            | 0.0          | 0.0                | 0.0            |   |
| 4 | 0.0            | 0.0          | 0.0                | 0.0            |   |

|   | resource_Total | generation | ... | variable_Resource | rent | \ |
|---|----------------|------------|-----|-------------------|------|---|
| 0 |                | 0.0        | ... |                   | 0.0  |   |
| 1 |                | 0.0        | ... |                   | 0.0  |   |
| 2 |                | 0.0        | ... |                   | 0.0  |   |
| 3 |                | 0.0        | ... |                   | 0.0  |   |
| 4 |                | 0.0        | ... |                   | 0.0  |   |

|   | units_Gigawatt | hours | units_Number | magnitude_Average | \ |
|---|----------------|-------|--------------|-------------------|---|
| 0 |                | 0.0   | 0.0          | 0.0               |   |
| 1 |                | 1.0   | 0.0          | 0.0               |   |
| 2 |                | 0.0   | 1.0          | 0.0               |   |
| 3 |                | 0.0   | 0.0          | 0.0               |   |
| 4 |                | 0.0   | 0.0          | 0.0               |   |

|            | magnitude_Proportion | magnitude_Thousands | source_Environmental |
|------------|----------------------|---------------------|----------------------|
| Accounts \ |                      |                     |                      |
| 0          | 0.0                  | 1.0                 |                      |
| 1.0        |                      |                     |                      |
| 1          | 0.0                  | 0.0                 |                      |
| 0.0        |                      |                     |                      |
| 2          | 1.0                  | 0.0                 |                      |
| 0.0        |                      |                     |                      |
| 3          | 0.0                  | 1.0                 |                      |
| 1.0        |                      |                     |                      |

```
4          0.0          1.0
1.0
```

|   | source_MBIE | flag_P | flag_R |
|---|-------------|--------|--------|
| 0 | 0.0         | 0.0    | 0.0    |
| 1 | 1.0         | 0.0    | 0.0    |
| 2 | 1.0         | 0.0    | 0.0    |
| 3 | 0.0         | 0.0    | 0.0    |
| 4 | 0.0         | 0.0    | 0.0    |

```
[5 rows x 28 columns]
```

```
Memory Usage Before Preprocessing: 328144 bytes
```

```
Memory Usage After Preprocessing: 179328 bytes
```