

02_feature_analysis

November 30, 2024

Looking for data file at:

/Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/processed_data/final_processed_data.csv

Dataset Overview:

=====

Shape: (613, 8)

Features:

- Year: float64 (Missing: 0)
- Month: float64 (Missing: 0)
- Hydroelectric Power: float64 (Missing: 0)
- Solar Energy: float64 (Missing: 0)
- Wind Energy: float64 (Missing: 0)
- Geothermal Energy: float64 (Missing: 0)
- Biomass Energy: float64 (Missing: 0)
- Total Renewable Energy: float64 (Missing: 0)

Normality Test Results:

	statistic	p_value
Year	409.837029	1.011626e-89
Month	469.167425	1.323086e-102
Hydroelectric Power	NaN	NaN
Solar Energy	174.988568	1.003957e-38
Wind Energy	NaN	NaN
Geothermal Energy	5325.538313	0.000000e+00
Biomass Energy	NaN	NaN
Total Renewable Energy	46.595575	7.619025e-11

Highly Correlated Feature Pairs ($|\text{correlation}| > 0.8$):

Year - Solar Energy: 0.827

Year - Geothermal Energy: 0.930

Solar Energy - Year: 0.827

Geothermal Energy - Year: 0.930

Available columns in dataset:

Index(['Year', 'Month', 'Hydroelectric Power', 'Solar Energy', 'Wind Energy',
'Geothermal Energy', 'Biomass Energy', 'Total Renewable Energy'],

```

dtype='object')
Warning: renewable_generation not found. Available columns:
Index(['Year', 'Month', 'Hydroelectric Power', 'Solar Energy', 'Wind Energy',
      'Geothermal Energy', 'Biomass Energy', 'Total Renewable Energy'],
      dtype='object')

```

Feature Importance Rankings:

None

Principal Component Loadings:

	PC1	PC2	PC3 \
Year	5.876726e-01	1.950889e-02	7.680819e-02
Month	5.166614e-03	8.877180e-01	-4.592721e-01
Hydroelectric Power	6.938894e-18	-8.326673e-17	-1.110223e-16
Solar Energy	5.461301e-01	7.722898e-02	9.875103e-02
Wind Energy	-0.000000e+00	0.000000e+00	-0.000000e+00
Geothermal Energy	5.721294e-01	3.289291e-02	9.144373e-02
Biomass Energy	-0.000000e+00	0.000000e+00	-0.000000e+00
Total Renewable Energy	-1.703647e-01	4.522497e-01	8.746747e-01

	PC4	PC5	PC6 \
Year	-2.376770e-01	7.693312e-01	4.893183e-33
Month	-2.993566e-02	1.014670e-02	-4.081425e-33
Hydroelectric Power	1.110223e-16	3.330669e-16	-2.436590e-17
Solar Energy	8.086565e-01	-1.791662e-01	-2.626447e-33
Wind Energy	-0.000000e+00	0.000000e+00	-1.110215e-16
Geothermal Energy	-5.364555e-01	-6.127312e-01	-2.807496e-33
Biomass Energy	-0.000000e+00	0.000000e+00	1.000000e+00
Total Renewable Energy	-3.006059e-02	2.205669e-02	6.457784e-34

	PC7	PC8
Year	-5.756718e-17	-1.924235e-16
Month	-9.155562e-17	-1.556198e-17
Hydroelectric Power	2.866481e-01	9.580359e-01
Solar Energy	1.374089e-17	7.405168e-17
Wind Energy	9.580359e-01	-2.866481e-01
Geothermal Energy	4.121389e-17	1.375242e-16
Biomass Energy	2.220446e-16	-5.551115e-17
Total Renewable Energy	-3.749220e-17	-1.622267e-17

<IPython.core.display.HTML object>