In [1]:
```python
# Import Libraries -

%matplotlib inline
import seaborn as sns
import plotly.express as px
import matplotlib.pyplot as plt

# Data Wrangling
import numpy as np
import pandas as pd
pd.set_option('display.max_columns',50)

# Warnings Ignore
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```python
# Load all relevant datasets
actual_consumption = pd.read_excel('Datsasets/Actual_consumption_Jan2023_Oct2024.xlsx', skiprows=9)
forecasted_consumption = pd.read_excel('Datsasets/Forecasted_consumption_Jan2023_Oct2024.xlsx', skiprows=9)
actual_generation = pd.read_excel('Datsasets/Actual_generation_Jan2023_Oct2024.xlsx', skiprows=9)
forecasted_generation = pd.read_excel('Datsasets/Forecasted_generation_Jan2023_Oct2024.xlsx',skiprows=9)
day_ahead_prices = pd.read_excel('Datsasets/Day-ahead_prices_Jan2023_Oct2024.xlsx', skiprows=9)
cross_border_flows = pd.read_excel('Datsasets/Cross-border_flows_Jan2023_Oct2024.xlsx', skiprows=9)
```

In [3]:
```python
# Display column names to locate the correct time column name
print("Actual Consumption Columns:", actual_consumption.columns)
print("Forecasted Consumption Columns:", forecasted_consumption.columns)
print("Actual Generation Columns:", actual_generation.columns)
print("Forecasted Generation Columns:", forecasted_generation.columns)
print("Day Ahead Prices Columns:", day_ahead_prices.columns)
print("Cross Border Flows Columns:", cross_border_flows.columns)
```

```
Actual Consumption Columns: Index(['Start date', 'End date', 'Total (grid load) [MWh]',
       'Residual load [MWh]', 'Hydro pumped storage [MWh]'],
      dtype='object')
Forecasted Consumption Columns: Index(['Start date', 'End date', 'Total (grid load) [MWh]',
       'Residual load [MWh]'],
      dtype='object')
Actual Generation Columns: Index(['Start date', 'End date', 'Biomass [MWh]', 'Hydropower [MWh]',
       'Wind offshore [MWh]', 'Wind onshore [MWh]', 'Photovoltaics [MWh]',
       'Other renewable [MWh]', 'Nuclear [MWh]', 'Lignite [MWh]',
       'Hard coal [MWh]', 'Fossil gas [MWh]', 'Hydro pumped storage [MWh]',
       'Other conventional [MWh]'],
      dtype='object')
Forecasted Generation Columns: Index(['Start date', 'End date', 'Total [MWh]', 'Photovoltaics and wind [MWh]',
       'Wind offshore [MWh]', 'Wind onshore [MWh]', 'Photovoltaics [MWh]',
       'Other [MWh]'],
      dtype='object')
Day Ahead Prices Columns: Index(['Start date', 'End date', 'Germany/Luxembourg [€/MWh]',
       'Ø DE/LU neighbours [€/MWh]', 'Belgium [€/MWh]', 'Denmark 1 [€/MWh]',
       'Denmark 2 [€/MWh]', 'France [€/MWh]', 'Netherlands [€/MWh]',
       'Norway 2 [€/MWh]', 'Austria [€/MWh]', 'Poland [€/MWh]',
       'Sweden 4 [€/MWh]', 'Switzerland [€/MWh]', 'Czech Republic [€/MWh]',
       'DE/AT/LU [€/MWh]', 'Northern Italy [€/MWh]', 'Slovenia [€/MWh]',
       'Hungary [€/MWh]'],
      dtype='object')
Cross Border Flows Columns: Index(['Start date', 'End date', 'Net export [MWh]',
       'Netherlands (export) [MWh]', 'Netherlands (import) [MWh]',
       'Switzerland (export) [MWh]', 'Switzerland (import) [MWh]',
       'Denmark (export) [MWh]', 'Denmark (import) [MWh]',
       'Czech Republic (export) [MWh]', 'Czech Republic (import) [MWh]',
       'Luxembourg (export) [MWh]', 'Luxembourg (import) [MWh]',
       'Sweden (export) [MWh]', 'Sweden (import) [MWh]',
       'Austria (export) [MWh]', 'Austria (import) [MWh]',
       'France (export) [MWh]', 'France (import) [MWh]',
       'Poland (export) [MWh]', 'Poland (import) [MWh]',
       'Norway (export) [MWh]', 'Norway (import) [MWh]',
       'Belgium (export) [MWh]', 'Belgium (import) [MWh]'],
      dtype='object')
```

In [4]:

```
# Step 1: Select relevant columns
# Only keep necessary columns to reduce memory usage
actual_consumption = actual_consumption[['Start date', 'End date', 'Total (grid load) [MWh]']]
forecasted_consumption = forecasted_consumption[['Start date', 'End date', 'Total (grid load) [MWh]']]
```

```
actual_generation = actual_generation[['Start date', 'End date', 'Biomass [MWh]', 'Hydropower [MWh]',
                                       'Wind offshore [MWh]', 'Wind onshore [MWh]', 'Photovoltaics [MWh]',
                                       'Nuclear [MWh]', 'Lignite [MWh]', 'Hard coal [MWh]', 'Fossil gas [MWh]']]
forecasted_generation = forecasted_generation[['Start date', 'End date', 'Photovoltaics [MWh]',
                                               'Wind offshore [MWh]', 'Wind onshore [MWh]']]
day_ahead_prices = day_ahead_prices[['Start date', 'End date', 'Germany/Luxembourg [€/MWh]']]
```

In [5]:
```
# Step 2: Rename columns for clarity in the final DataFrame
# Renaming only the columns (except date columns) to match the target format
actual_consumption.rename(columns={'Total (grid load) [MWh]': 'Total load actual'}, inplace=True)
forecasted_consumption.rename(columns={'Total (grid load) [MWh]': 'Total load forecast'}, inplace=True)
day_ahead_prices.rename(columns={'Germany/Luxembourg [€/MWh]': 'Price day ahead'}, inplace=True)
```

In [6]:
```
# Step 3: Merge datasets on 'Start date' and 'End date' using outer joins
merged_df = pd.merge(day_ahead_prices, actual_consumption, on=['Start date', 'End date'], how='outer')
merged_df = pd.merge(merged_df, forecasted_consumption, on=['Start date', 'End date'], how='outer')
merged_df = pd.merge(merged_df, actual_generation, on=['Start date', 'End date'], how='outer')
merged_df = pd.merge(merged_df, forecasted_generation, on=['Start date', 'End date'], how='outer')
merged_df = pd.merge(merged_df, cross_border_flows, on=['Start date', 'End date'], how='outer')
```

In [7]:
```
# Step 4: Drop rows with missing 'Price day ahead' values, as it is the target for prediction
merged_df.dropna(subset=['Price day ahead'], inplace=True)
```

In [8]:
```
merged_df.head()
```

Out[8]:

	Start date	End date	Price day ahead	Total load actual	Total load forecast	Biomass [MWh]	Hydropower [MWh]	Wind offshore [MWh]_x	Wind onshore [MWh]_x	Photovoltaics [MWh]_x	Nuclear [MWh]	Lignite [MWh]	Hard coal [MWh]
0	Jan 1, 2023 12:00 AM	Jan 1, 2023 1:00 AM	-5.17	38346.00	41792.50	4365	1275.25	3059.25	28710.50	1.25	2459.5	3859.25	2067.5
1	Jan 1, 2023 1:00 AM	Jan 1, 2023 2:00 AM	-1.07	37777.25	39621.00	4344.75	1226.5	3586.00	29305.00	1.00	2458.75	3866.5	2052
2	Jan 1, 2023 2:00 AM	Jan 1, 2023 3:00 AM	-1.47	36939.75	38240.75	4333	1222.5	3842.25	29266.00	1.25	2459.75	3860.25	2034.25
3	Jan 1, 2023 3:00 AM	Jan 1, 2023 4:00 AM	-5.08	35932.50	37205.50	4338.75	1223.25	3463.25	27008.50	1.00	2460.5	3864.75	2037
4	Jan 1, 2023 4:00 AM	Jan 1, 2023 5:00 AM	-4.49	35486.25	37326.75	4353.25	1244	3462.25	26438.75	1.50	2461	3841	2040.25

In [9]:

```
# .Describe to show overview of data
merged_df.describe()
```

Out[9]:

	Price day ahead	Total load actual	Total load forecast	Wind offshore [MWh]_x	Wind onshore [MWh]_x	Photovoltaics [MWh]_x	Photovoltaics [MWh]_y	Wind offshore [MWh]_y	[M
count	16204.000000	16204.000000	16204.000000	16204.000000	16204.000000	16204.000000	16204.000000	16204.000000	16204
mean	84.289719	52150.487503	52313.677148	2778.758856	12895.286534	7191.515320	7197.909035	2782.372454	12901
std	46.891651	9171.297052	9112.136495	1857.645810	10009.713755	10715.744998	10717.777823	1757.965261	9996
min	-500.000000	30909.000000	30544.750000	0.000000	145.500000	0.000000	0.000000	24.250000	203
25%	61.347500	44594.500000	44839.062500	1112.500000	4925.750000	2.250000	0.000000	1160.687500	4983
50%	87.285000	52058.250000	52120.000000	2567.500000	10204.000000	324.750000	344.875000	2691.500000	10037
75%	110.052500	59437.937500	59793.500000	4278.562500	18695.187500	12075.812500	12009.000000	4343.250000	18486
max	656.370000	75508.250000	73298.250000	7633.250000	48023.000000	46848.250000	48155.750000	6483.250000	46617

In [10]:

```
# .Info to show datatype, nulls, and count
merged_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16204 entries, 0 to 16203
Data columns (total 40 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Start date                    16204 non-null  object 
 1   End date                      16204 non-null  object 
 2   Price day ahead               16204 non-null  float64
 3   Total load actual             16204 non-null  float64
 4   Total load forecast           16204 non-null  float64
 5   Biomass [MWh]                 16204 non-null  object 
 6   Hydropower [MWh]              16204 non-null  object 
 7   Wind offshore [MWh]_x         16204 non-null  float64
 8   Wind onshore [MWh]_x          16204 non-null  float64
 9   Photovoltaics [MWh]_x         16204 non-null  float64
 10  Nuclear [MWh]                 16204 non-null  object 
 11  Lignite [MWh]                 16204 non-null  object 
 12  Hard coal [MWh]               16204 non-null  object 
 13  Fossil gas [MWh]              16204 non-null  object 
 14  Photovoltaics [MWh]_y         16204 non-null  float64
 15  Wind offshore [MWh]_y         16204 non-null  float64
 16  Wind onshore [MWh]_y          16204 non-null  float64
 17  Net export [MWh]              16204 non-null  object 
 18  Netherlands (export) [MWh]    16204 non-null  object 
 19  Netherlands (import) [MWh]    16204 non-null  object 
 20  Switzerland (export) [MWh]    16204 non-null  object 
 21  Switzerland (import) [MWh]    16204 non-null  object 
 22  Denmark (export) [MWh]        16204 non-null  float64
 23  Denmark (import) [MWh]        16204 non-null  float64
 24  Czech Republic (export) [MWh] 16204 non-null  float64
 25  Czech Republic (import) [MWh] 16204 non-null  float64
 26  Luxembourg (export) [MWh]     16204 non-null  object 
 27  Luxembourg (import) [MWh]     16204 non-null  object 
 28  Sweden (export) [MWh]         16204 non-null  int64  
 29  Sweden (import) [MWh]         16204 non-null  int64  
 30  Austria (export) [MWh]        16204 non-null  float64
 31  Austria (import) [MWh]        16204 non-null  float64
 32  France (export) [MWh]         16204 non-null  int64  
 33  France (import) [MWh]         16204 non-null  int64  
 34  Poland (export) [MWh]         16204 non-null  object 
 35  Poland (import) [MWh]         16204 non-null  object 
 36  Norway (export) [MWh]         16204 non-null  float64
```

```
 37  Norway (import) [MWh]          16204 non-null  float64
 38  Belgium (export) [MWh]         16204 non-null  float64
 39  Belgium (import) [MWh]         16204 non-null  float64
dtypes: float64(19), int64(4), object(17)
memory usage: 4.9+ MB
```

In [14]:

```
# Step 4: Verify data types and view the cleaned DataFrame
print(merged_df.dtypes)
print(merged_df.head())

# View Pearson correlation with 'Price day ahead', ensuring only numeric columns are used
correlation_matrix = merged_df.select_dtypes(include=['number']).corr()
price_correlation = correlation_matrix['Price day ahead'].sort_values(ascending=False)

print("\nPearson correlation with 'Price day ahead':")
print(price_correlation)
```

```
Start date                          object
End date                            object
Price day ahead                    float64
Total load actual                  float64
Total load forecast                float64
Biomass [MWh]                      float64
Hydropower [MWh]                   float64
Wind offshore [MWh]_x              float64
Wind onshore [MWh]_x               float64
Photovoltaics [MWh]_x              float64
Nuclear [MWh]                      float64
Lignite [MWh]                      float64
Hard coal [MWh]                    float64
Fossil gas [MWh]                   float64
Photovoltaics [MWh]_y              float64
Wind offshore [MWh]_y              float64
Wind onshore [MWh]_y               float64
Net export [MWh]                   float64
Netherlands (export) [MWh]         float64
Netherlands (import) [MWh]         float64
Switzerland (export) [MWh]         float64
Switzerland (import) [MWh]         float64
Denmark (export) [MWh]             float64
Denmark (import) [MWh]             float64
Czech Republic (export) [MWh]      float64
Czech Republic (import) [MWh]      float64
Luxembourg (export) [MWh]          float64
Luxembourg (import) [MWh]          float64
Sweden (export) [MWh]                int64
Sweden (import) [MWh]                int64
Austria (export) [MWh]             float64
Austria (import) [MWh]             float64
France (export) [MWh]                int64
France (import) [MWh]                int64
Poland (export) [MWh]              float64
Poland (import) [MWh]              float64
Norway (export) [MWh]              float64
Norway (import) [MWh]              float64
Belgium (export) [MWh]             float64
Belgium (import) [MWh]             float64
dtype: object
             Start date             End date  Price day ahead  \
```

```
0  Jan 1, 2023 12:00 AM  Jan 1, 2023 1:00 AM           -5.17
1   Jan 1, 2023 1:00 AM  Jan 1, 2023 2:00 AM           -1.07
2   Jan 1, 2023 2:00 AM  Jan 1, 2023 3:00 AM           -1.47
3   Jan 1, 2023 3:00 AM  Jan 1, 2023 4:00 AM           -5.08
4   Jan 1, 2023 4:00 AM  Jan 1, 2023 5:00 AM           -4.49

   Total load actual  Total load forecast  Biomass [MWh]  Hydropower [MWh]  \
0           38346.00             41792.50        4365.00           1275.25
1           37777.25             39621.00        4344.75           1226.50
2           36939.75             38240.75        4333.00           1222.50
3           35932.50             37205.50        4338.75           1223.25
4           35486.25             37326.75        4353.25           1244.00

   Wind offshore [MWh]_x  Wind onshore [MWh]_x  Photovoltaics [MWh]_x  \
0                3059.25              28710.50                   1.25
1                3586.00              29305.00                   1.00
2                3842.25              29266.00                   1.25
3                3463.25              27008.50                   1.00
4                3462.25              26438.75                   1.50

   Nuclear [MWh]  Lignite [MWh]  Hard coal [MWh]  Fossil gas [MWh]  \
0        2459.50        3859.25          2067.50           1721.75
1        2458.75        3866.50          2052.00           1568.00
2        2459.75        3860.25          2034.25           1571.50
3        2460.50        3864.75          2037.00           1564.75
4        2461.00        3841.00          2040.25           1550.50

   Photovoltaics [MWh]_y  Wind offshore [MWh]_y  Wind onshore [MWh]_y  \
0                    0.0                3478.25              35515.50
1                    0.0                3390.25              35344.50
2                    0.0                3395.50              35138.75
3                    0.0                3410.25              34441.00
4                    0.0                3431.25              33898.00

   Net export [MWh]  Netherlands (export) [MWh]  Netherlands (import) [MWh]  \
0          12001.75                     1115.50                     -931.75
1          14244.25                     1442.25                      -15.75
2          13897.50                     1452.25                    -1032.00
3          12277.25                     1289.25                    -1317.50
4          12179.00                     1248.00                    -1446.00

   Switzerland (export) [MWh]  Switzerland (import) [MWh]  \
```

```
0                   2135.25                        -2.25
1                   2526.50                       -88.00
2                   2263.50                       -41.00
3                   2180.00                        -3.00
4                   2447.50                        -0.25

   Denmark (export) [MWh]  Denmark (import) [MWh]  \
0                 2606.25                    -0.5
1                 2315.00                     0.0
2                 2596.25                     0.0
3                 2732.75                     0.0
4                 2567.00                     0.0

   Czech Republic (export) [MWh]  Czech Republic (import) [MWh]  \
0                         1393.0                         -845.0
1                         1540.0                         -802.0
2                         1526.0                         -898.0
3                         1374.0                         -871.0
4                         1282.0                         -962.0

   Luxembourg (export) [MWh]  Luxembourg (import) [MWh]  \
0                     376.50                        0.0
1                     379.00                        0.0
2                     182.50                        0.0
3                     113.75                        0.0
4                     101.25                        0.0

   Sweden (export) [MWh]  Sweden (import) [MWh]  Austria (export) [MWh]  \
0                    469                      0                 2393.75
1                    506                      0                 2603.25
2                    506                      0                 2392.50
3                    506                      0                 2298.25
4                    506                      0                 2357.75

   Austria (import) [MWh]  France (export) [MWh]  France (import) [MWh]  \
0                  -79.75                    861                     -4
1                  -45.00                   2008                      0
2                  -61.75                   1970                      0
3                  -68.75                   1443                      0
4                  -87.50                   1473                      0

   Poland (export) [MWh]  Poland (import) [MWh]  Norway (export) [MWh]  \
```

```
0              1065.0                   0.0               1273.25
1              1173.0                   0.0               1289.00
2              1038.0                   0.0               1298.00
3               681.0                   0.0               1298.75
4               558.0                   0.0               1308.25

   Norway (import) [MWh]  Belgium (export) [MWh]  Belgium (import) [MWh]
0                    0.0                  176.50                     0.0
1                    0.0                    0.00                  -587.0
2                    0.0                  705.25                     0.0
3                    0.0                  620.75                     0.0
4                    0.0                  826.00                     0.0

Pearson correlation with 'Price day ahead':
Price day ahead                  1.000000
Lignite [MWh]                    0.706410
Fossil gas [MWh]                 0.616888
Hard coal [MWh]                  0.573374
Luxembourg (export) [MWh]        0.432364
Biomass [MWh]                    0.420294
Total load actual                0.289017
Total load forecast              0.278979
Nuclear [MWh]                    0.265016
Netherlands (import) [MWh]       0.151035
France (export) [MWh]            0.096997
Luxembourg (import) [MWh]        0.085328
Hydropower [MWh]                 0.036067
Poland (import) [MWh]            0.032173
Poland (export) [MWh]           -0.020908
France (import) [MWh]           -0.083219
Netherlands (export) [MWh]      -0.104530
Belgium (export) [MWh]          -0.111760
Belgium (import) [MWh]          -0.124363
Switzerland (export) [MWh]      -0.146772
Sweden (import) [MWh]           -0.207934
Wind offshore [MWh]_x           -0.252183
Switzerland (import) [MWh]      -0.301798
Wind offshore [MWh]_y           -0.327093
Austria (import) [MWh]          -0.329806
Austria (export) [MWh]          -0.337749
Czech Republic (import) [MWh]   -0.340299
Czech Republic (export) [MWh]   -0.344022
```

```
Photovoltaics [MWh]_x          -0.383821
Wind onshore [MWh]_x           -0.386911
Photovoltaics [MWh]_y          -0.388841
Wind onshore [MWh]_y           -0.406169
Net export [MWh]               -0.419468
Denmark (import) [MWh]         -0.421173
Sweden (export) [MWh]          -0.470143
Denmark (export) [MWh]         -0.522718
Norway (import) [MWh]          -0.536448
Norway (export) [MWh]          -0.552707
Name: Price day ahead, dtype: float64
```

In [15]:

```
# Step 5: Calculate Pearson correlation with the target 'Price day ahead'
# Select only numeric columns to avoid errors with non-numeric data
numeric_cols = merged_df.select_dtypes(include=['number'])

# Calculate the correlation matrix on numeric columns only
correlation_matrix = numeric_cols.corr()

# Get correlation values with 'Price day ahead' and sort
if 'Price day ahead' in correlation_matrix.columns:
    price_correlation = correlation_matrix['Price day ahead'].sort_values(ascending=False)
    print("Pearson correlation with 'Price day ahead':")
    print(price_correlation)
else:
    print("Column 'Price day ahead' is not found in the numeric data for correlation.")
```

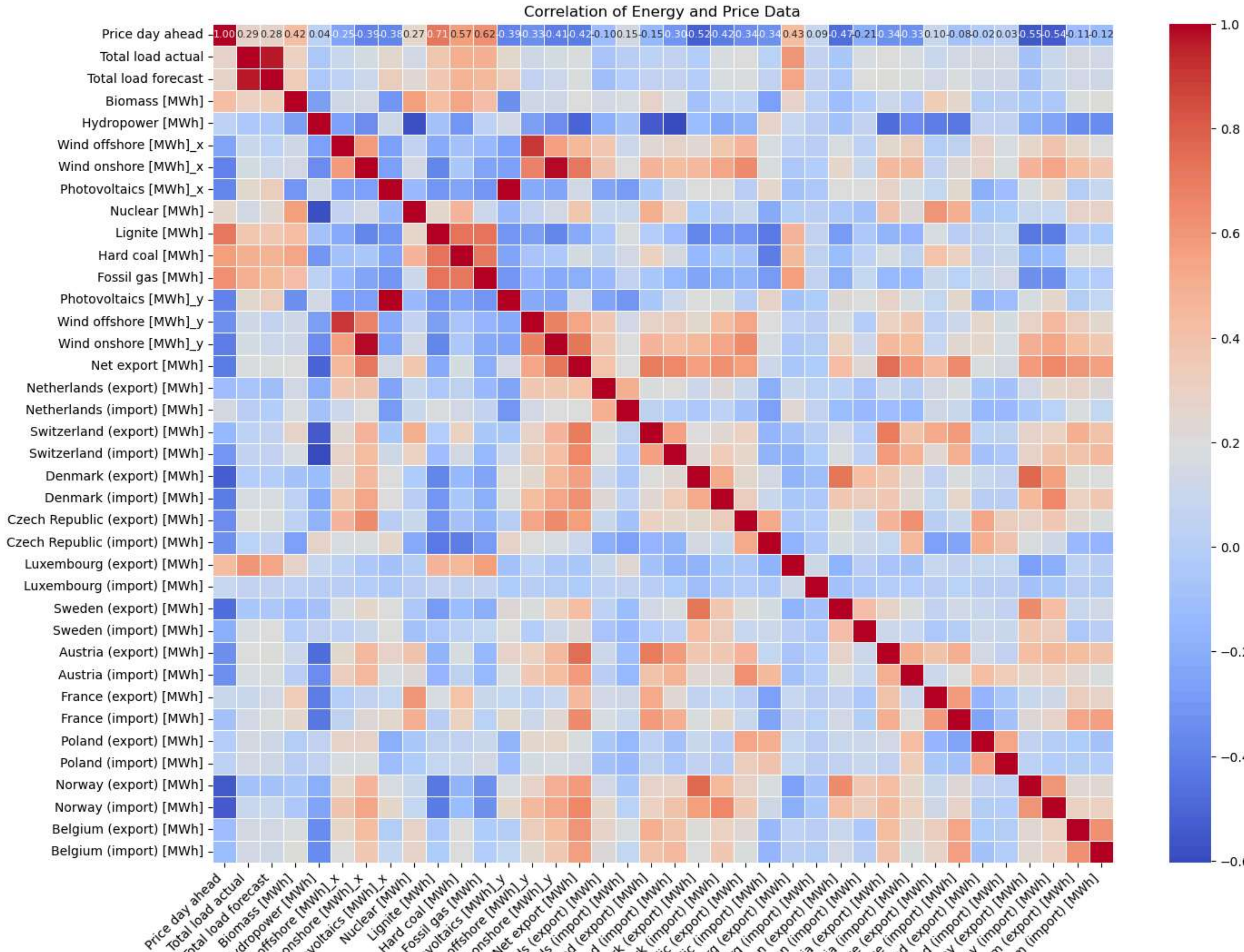
```
Pearson correlation with 'Price day ahead':
Price day ahead                   1.000000
Lignite [MWh]                     0.706410
Fossil gas [MWh]                  0.616888
Hard coal [MWh]                   0.573374
Luxembourg (export) [MWh]         0.432364
Biomass [MWh]                     0.420294
Total load actual                 0.289017
Total load forecast               0.278979
Nuclear [MWh]                     0.265016
Netherlands (import) [MWh]        0.151035
France (export) [MWh]             0.096997
Luxembourg (import) [MWh]         0.085328
Hydropower [MWh]                  0.036067
Poland (import) [MWh]             0.032173
Poland (export) [MWh]            -0.020908
France (import) [MWh]            -0.083219
Netherlands (export) [MWh]       -0.104530
Belgium (export) [MWh]           -0.111760
Belgium (import) [MWh]           -0.124363
Switzerland (export) [MWh]       -0.146772
Sweden (import) [MWh]            -0.207934
Wind offshore [MWh]_x            -0.252183
Switzerland (import) [MWh]       -0.301798
Wind offshore [MWh]_y            -0.327093
Austria (import) [MWh]           -0.329806
Austria (export) [MWh]           -0.337749
Czech Republic (import) [MWh]    -0.340299
Czech Republic (export) [MWh]    -0.344022
Photovoltaics [MWh]_x            -0.383821
Wind onshore [MWh]_x             -0.386911
Photovoltaics [MWh]_y            -0.388841
Wind onshore [MWh]_y             -0.406169
Net export [MWh]                 -0.419468
Denmark (import) [MWh]           -0.421173
Sweden (export) [MWh]            -0.470143
Denmark (export) [MWh]           -0.522718
Norway (import) [MWh]            -0.536448
Norway (export) [MWh]            -0.552707
Name: Price day ahead, dtype: float64
```

In [ ]:

In [16]:

```
# Step 1: Select only numeric columns to calculate the correlation matrix
numeric_cols = merged_df.select_dtypes(include=['number'])
correlation_matrix = numeric_cols.corr()

# Step 2: Plot the heatmap with clearer annotation settings
plt.figure(figsize=(16, 12))  # Increase figure size for clarity
sns.heatmap(
    correlation_matrix,
    annot=True,
    fmt=".2f",
    cmap="coolwarm",
    linewidths=0.5,
    annot_kws={"size": 8}  # Adjust font size for annotations
)
plt.title("Correlation of Energy and Price Data")
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for readability
plt.yticks(rotation=0)  # Keep y-axis labels horizontal
plt.show()
```


Correlation of Energy and Price Data
Price day ahead
Total load actual
Total load forecast
Biomass [MWh]
Hydropower [MWh]
Wind offshore [MWh]_x
Wind onshore [MWh]_x
Photovoltaics [MWh]_x
Nuclear [MWh]
Lignite [MWh]
Hard coal [MWh]
Fossil gas [MWh]
Photovoltaics [MWh]_y
Wind offshore [MWh]_y
Wind onshore [MWh]_y
Net export [MWh]
Netherlands (export) [MWh]
Netherlands (import) [MWh]
Switzerland (export) [MWh]
Switzerland (import) [MWh]
Denmark (export) [MWh]
Denmark (import) [MWh]
Czech Republic (export) [MWh]
Czech Republic (import) [MWh]
Luxembourg (export) [MWh]
Luxembourg (import) [MWh]
Sweden (export) [MWh]
Sweden (import) [MWh]
Austria (export) [MWh]
Austria (import) [MWh]
France (export) [MWh]
France (import) [MWh]
Poland (export) [MWh]
Poland (import) [MWh]
Norway (export) [MWh]
Norway (import) [MWh]
Belgium (export) [MWh]
Belgium (import) [MWh]
1.00 0.29 0.28 0.42 0.04 -0.25 -0.39 -0.38 0.27 0.71 0.57 0.62 -0.39 -0.33 -0.41 -0.42 -0.10 0.15 -0.15 -0.30 -0.52 -0.42 -0.34 -0.34 0.43 0.09 -0.47 -0.21 -0.34 -0.33 0.10 -0.08 -0.02 0.03 -0.55 -0.54 -0.11 -0.12
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Strong Predictors for Price These features showed high correlation with Price day ahead in the heatmap and have logical significance based on how electricity markets operate. Including them in the model is justified as they likely capture key drivers of price variations.

Total load actual and Total load forecast:

Justification: These represent the demand for electricity, which is one of the most fundamental factors influencing price. When demand is high, prices tend to increase due to limited supply. Reasoning: The strong correlation with price (0.71 for actual load and 0.62 for forecasted load) indicates that both variables are essential predictors. However, they are also highly correlated with each other (~0.93), meaning they both reflect similar information. Therefore, it might be practical to use one of them (preferably forecast) or employ techniques to handle multicollinearity if both are included.

Fossil gas (MWh):

Justification: Natural gas is often used in electricity production, especially for balancing peak demand. Gas-fired plants are flexible and can be ramped up or down as needed. Reasoning: Gas has a moderate to strong positive correlation (~0.57) with price, suggesting that gas generation is a cost-sensitive factor. When gas usage rises, prices often increase due to the relatively high variable costs of gas generation compared to renewables.

Wind onshore (MWh)_x, Wind offshore (MWh)_x, and Photovoltaics (MWh)_x:

Justification: Renewable energy sources like wind and solar have low marginal costs and can contribute to reducing electricity prices when their production is high. Reasoning: The negative correlation between these renewable sources and the price (correlations around -0.30 to -0.40) suggests that increased production from renewables reduces the reliance on more expensive generation sources. Including these features helps capture price reductions driven by higher renewable output.

Moderate Predictors

These features had moderate correlations with Price day ahead and also align with known industry dynamics, making them potentially useful but less essential than the strong predictors. Including them could improve model performance, but they may not be as critical.

Nuclear (MWh):

Justification: Nuclear power provides a stable baseload, which is relatively inexpensive in terms of operational costs once plants are running. Reasoning: Nuclear generation has a moderate correlation with price (~0.3-0.4), reflecting its role in reducing price variability by providing steady output. However, because nuclear plants generally operate continuously, their contribution to price fluctuations may be limited.

Lignite (MWh) and Hard coal (MWh):

Justification: Coal-based energy production can be more expensive and less flexible, but it often provides significant output in certain grids. Reasoning: Both lignite and hard coal show moderate positive correlations with price. This correlation is likely because coal plants contribute to price increases when other, cheaper sources are limited. However, their impact is generally less than that of natural gas and renewables.

Import/Export Variables for Certain Countries (e.g., Netherlands, Switzerland, Denmark):

Justification: Cross-border electricity flows can help balance supply and demand, especially when domestic production falls short. Imports from neighboring countries can lower prices, while high exports can drive prices up. Reasoning: Countries like the Netherlands, Switzerland, and Denmark showed moderate correlation with Price day ahead in the matrix. This suggests that cross-border flows with these countries have a notable impact on prices, possibly due to interconnected market dynamics in the European grid. Their influence makes sense as these countries may have complementary energy profiles or act as significant trading partners.

Weak Predictors (Consider Excluding)

These variables had low correlation with Price day ahead, meaning their impact on price is likely negligible or inconsistent. Including them might introduce noise into the model without significantly improving prediction accuracy.

Import/Export Variables for Luxembourg, Sweden, Norway, and Belgium: Justification: These countries showed weak or near-zero correlation with day-ahead prices, indicating limited influence on price fluctuations. Reasoning: Electricity imports/exports with these countries may be less frequent, smaller in volume, or less directly tied to price-sensitive resources. For example: Luxembourg: Being a smaller country with a limited impact on larger energy markets. Sweden and Norway: Often rely heavily on hydropower, which is renewable and less variable in cost, reducing its impact on price fluctuations in other regions. Belgium: May have lower import/export volumes or a market structure that makes it less influential on domestic price setting.