In [16]:
```
import pandas as pd
import numpy as np
data = pd.read_csv("https://raw.githubusercontent.com/amankharwal/Website-data/master/electricity.csv")
print(data.head())
```

```
           DateTime Holiday  HolidayFlag  DayOfWeek  WeekOfYear  Day  Month  \
0  01/11/2011 00:00    None            0          1          44    1     11
1  01/11/2011 00:30    None            0          1          44    1     11
2  01/11/2011 01:00    None            0          1          44    1     11
3  01/11/2011 01:30    None            0          1          44    1     11
4  01/11/2011 02:00    None            0          1          44    1     11

   Year  PeriodOfDay ForecastWindProduction SystemLoadEA  SMPEA  \
0  2011            0                315.31      3388.77  49.26
1  2011            1                321.80      3196.66  49.26
2  2011            2                328.57      3060.71  49.10
3  2011            3                335.60      2945.56  48.04
4  2011            4                342.90      2849.34  33.75

  ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction SystemLoadEP2  \
0           6.00         9.30       600.71                356.00       3159.60
1           6.00        11.10       605.42                317.00       2973.01
2           5.00        11.10       589.97                311.00       2834.00
3           6.00         9.30       585.94                313.00       2725.99
4           6.00        11.10       571.52                346.00       2655.64

  SMPEP2
0  54.32
1  54.23
2  54.23
3  53.47
4  39.87

C:\Users\HP\AppData\Local\Temp\ipykernel_15892\3353965664.py:3: DtypeWarning: Columns (9,10,11,14,15,16,17) have mixe
d types. Specify dtype option on import or set low_memory=False.
  data = pd.read_csv("https://raw.githubusercontent.com/amankharwal/Website-data/master/electricity.csv")
```

In [17]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38014 entries, 0 to 38013
Data columns (total 18 columns):
 #   Column                   Non-Null Count  Dtype 
---  ------                   --------------  ----- 
 0   DateTime                 38014 non-null  object
 1   Holiday                  38014 non-null  object
 2   HolidayFlag              38014 non-null  int64 
 3   DayOfWeek                38014 non-null  int64 
 4   WeekOfYear               38014 non-null  int64 
 5   Day                      38014 non-null  int64 
 6   Month                    38014 non-null  int64 
 7   Year                     38014 non-null  int64 
 8   PeriodOfDay              38014 non-null  int64 
 9   ForecastWindProduction   38014 non-null  object
 10  SystemLoadEA             38014 non-null  object
 11  SMPEA                    38014 non-null  object
 12  ORKTemperature           38014 non-null  object
 13  ORKWindspeed             38014 non-null  object
 14  CO2Intensity             38014 non-null  object
 15  ActualWindProduction     38014 non-null  object
 16  SystemLoadEP2            38014 non-null  object
 17  SMPEP2                   38014 non-null  object
dtypes: int64(7), object(11)
memory usage: 5.2+ MB
```

In [18]:
```
data["ForecastWindProduction"] = pd.to_numeric(data["ForecastWindProduction"], errors= 'coerce')
data["SystemLoadEA"] = pd.to_numeric(data["SystemLoadEA"], errors= 'coerce')
data["SMPEA"] = pd.to_numeric(data["SMPEA"], errors= 'coerce')
data["ORKTemperature"] = pd.to_numeric(data["ORKTemperature"], errors= 'coerce')
data["ORKWindspeed"] = pd.to_numeric(data["ORKWindspeed"], errors= 'coerce')
data["CO2Intensity"] = pd.to_numeric(data["CO2Intensity"], errors= 'coerce')
data["ActualWindProduction"] = pd.to_numeric(data["ActualWindProduction"], errors= 'coerce')
data["SystemLoadEP2"] = pd.to_numeric(data["SystemLoadEP2"], errors= 'coerce')
data["SMPEP2"] = pd.to_numeric(data["SMPEP2"], errors= 'coerce')
```

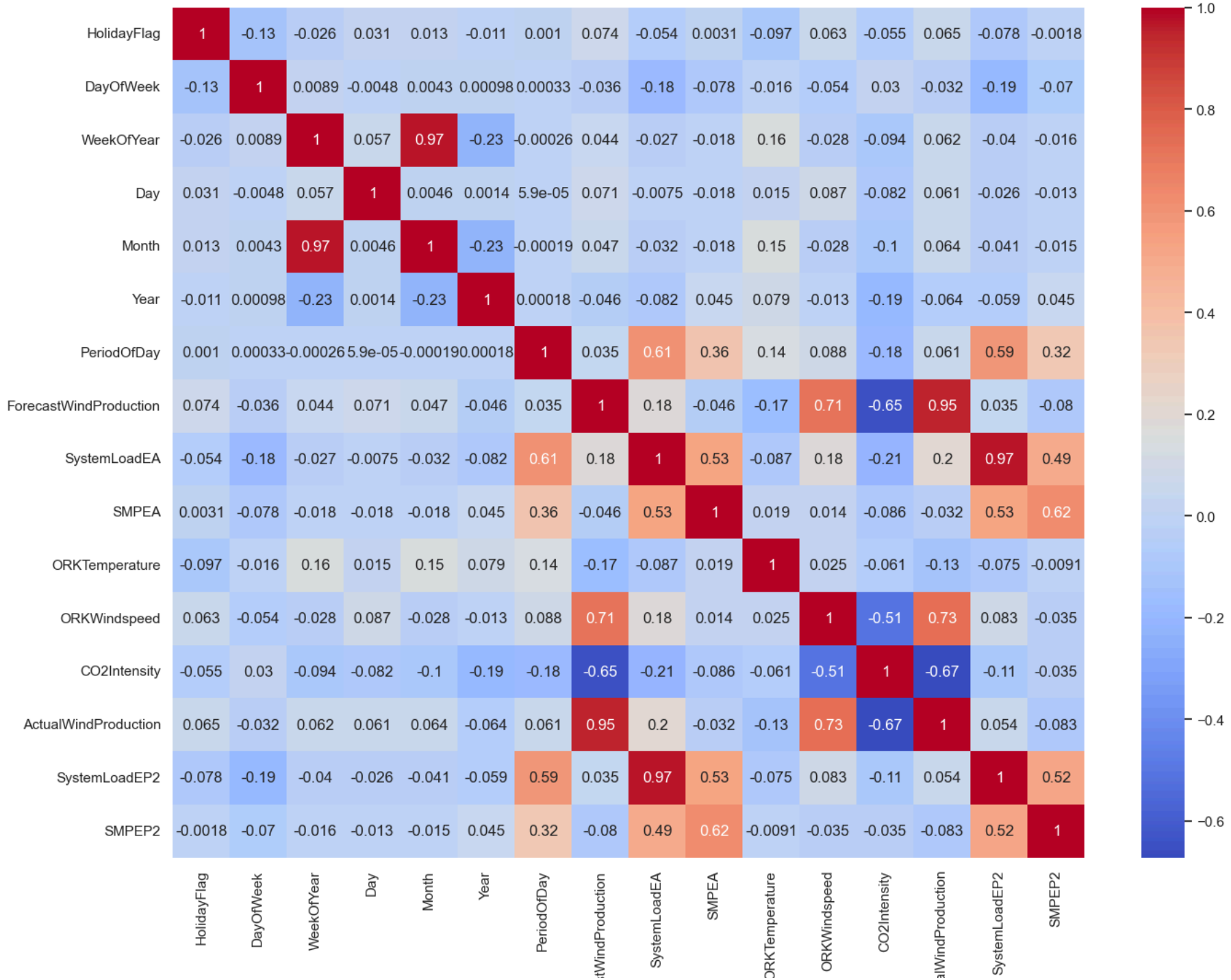
In [19]:
```
data.isnull().sum()
```

Out[19]:
```
DateTime                   0
Holiday                    0
HolidayFlag                0
DayOfWeek                  0
WeekOfYear                 0
Day                        0
Month                      0
Year                       0
PeriodOfDay                0
ForecastWindProduction     5
SystemLoadEA               2
SMPEA                      2
ORKTemperature           295
ORKWindspeed             299
CO2Intensity               7
ActualWindProduction       5
SystemLoadEP2              2
SMPEP2                     2
dtype: int64
```

In [20]:
```
data = data.dropna()
```

In [21]:
```
import seaborn as sns
import matplotlib.pyplot as plt
correlations = data.corr(method='pearson')
plt.figure(figsize=(16, 12))
sns.heatmap(correlations, cmap="coolwarm", annot=True)
plt.show()
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_15892\2280798960.py:3: FutureWarning: The default value of numeric_only in D
ataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  correlations = data.corr(method='pearson')
```


HolidayFlag 1 -0.13 -0.026 0.031 0.013 -0.011 0.001 0.074 -0.054 0.0031 -0.097 0.063 -0.055 0.065 -0.078 -0.0018
DayOfWeek -0.13 1 0.0089 -0.0048 0.0043 0.00098 0.00033 -0.036 -0.18 -0.078 -0.016 -0.054 0.03 -0.032 -0.19 -0.07
WeekOfYear -0.026 0.0089 1 0.057 0.97 -0.23 -0.00026 0.044 -0.027 -0.018 0.16 -0.028 -0.094 0.062 -0.04 -0.016
Day 0.031 -0.0048 0.057 1 0.0046 0.0014 5.9e-05 0.071 -0.0075 -0.018 0.015 0.087 -0.082 0.061 -0.026 -0.013
Month 0.013 0.0043 0.97 0.0046 1 -0.23 -0.00019 0.047 -0.032 -0.018 0.15 -0.028 -0.1 0.064 -0.041 -0.015
Year -0.011 0.00098 -0.23 0.0014 -0.23 1 0.00018 -0.046 -0.082 0.045 0.079 -0.013 -0.19 -0.064 -0.059 0.045
PeriodOfDay 0.001 0.00033 -0.00026 5.9e-05 -0.00019 0.00018 1 0.035 0.61 0.36 0.14 0.088 -0.18 0.061 0.59 0.32
ForecastWindProduction 0.074 -0.036 0.044 0.071 0.047 -0.046 0.035 1 0.18 -0.046 -0.17 0.71 -0.65 0.95 0.035 -0.08
SystemLoadEA -0.054 -0.18 -0.027 -0.0075 -0.032 -0.082 0.61 0.18 1 0.53 -0.087 0.18 -0.21 0.2 0.97 0.49
SMPEA 0.0031 -0.078 -0.018 -0.018 -0.018 0.045 0.36 -0.046 0.53 1 0.019 0.014 -0.086 -0.032 0.53 0.62
ORKTemperature -0.097 -0.016 0.16 0.015 0.15 0.079 0.14 -0.17 -0.087 0.019 1 0.025 -0.061 -0.13 -0.075 -0.0091
ORKWindspeed 0.063 -0.054 -0.028 0.087 -0.028 -0.013 0.088 0.71 0.18 0.014 0.025 1 -0.51 0.73 0.083 -0.035
CO2Intensity -0.055 0.03 -0.094 -0.082 -0.1 -0.19 -0.18 -0.65 -0.21 -0.086 -0.061 -0.51 1 -0.67 -0.11 -0.035
ActualWindProduction 0.065 -0.032 0.062 0.061 0.064 -0.064 0.061 0.95 0.2 -0.032 -0.13 0.73 -0.67 1 0.054 -0.083
SystemLoadEP2 -0.078 -0.19 -0.04 -0.026 -0.041 -0.059 0.59 0.035 0.97 0.53 -0.075 0.083 -0.11 0.054 1 0.52
SMPEP2 -0.0018 -0.07 -0.016 -0.013 -0.015 0.045 0.32 -0.08 0.49 0.62 -0.0091 -0.035 -0.035 -0.083 0.52 1
HolidayFlag
DayOfWeek
WeekOfYear
Day
Month
Year
PeriodOfDay
SystemLoadEA
SMPEA
ORKWindspeed
CO2Intensity
SystemLoadEP2
SMPEP2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

In [22]:
```
x = data[["Day", "Month", "ForecastWindProduction", "SystemLoadEA",
          "SMPEA", "ORKTemperature", "ORKWindspeed", "CO2Intensity",
          "ActualWindProduction", "SystemLoadEP2"]]
y = data["SMPEP2"]
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y,
                                                test_size=0.2,
                                                random_state=42)
```

In [23]:
```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
model.fit(xtrain, ytrain)
```

Out[23]:
```
RandomForestRegressor

RandomForestRegressor()
```

In [25]:
```
#features = [["Day", "Month", "ForecastWindProduction", "SystemLoadEA", "SMPEA", "ORKTemperature", "ORKWindspeed", "CO
features = np.array([[10, 12, 54.10, 4241.05, 49.56, 9.0, 14.8, 491.32, 54.0, 4426.84]])
model.predict(features)
```

```
C:\Users\HP\anaconda3\lib\site-packages\sklearn\base.py:420: UserWarning: X does not have valid feature names, but Ra
ndomForestRegressor was fitted with feature names
  warnings.warn(
```

Out[25]:
```
array([68.7649])
```