Importing Libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.axes as ax
import seaborn as sns

sns.set()
```

Loading Data

In [2]:
```
data = pd.read_csv(r'C:\Users\vamsi\Desktop\ML\Data Dimensionality\global_hea
data.head(22)
```

Out[2]:

	Year	Month	Day	Hour	Dew Point	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
0	2011	1	1	0	8	13.522659	986.761841	72.295858	37.288387	3.011042
1	2011	1	1	1	8	12.835814	986.441406	75.376186	37.686718	3.091243
2	2011	1	1	2	8	12.198058	985.736511	78.405198	35.053905	3.007649
3	2011	1	1	3	8	11.583500	985.525696	81.042980	30.135216	2.926715
4	2011	1	1	4	8	11.029578	985.661926	82.548508	24.402969	2.915177
5	2011	1	1	5	1	8.098655	985.723206	65.278494	89.290062	1.937899
6	2011	1	1	6	2	7.806705	986.249146	67.093181	95.835983	1.815175
7	2011	1	1	7	2	9.036408	987.038940	64.035483	105.198181	1.615310
8	2011	1	1	8	3	12.158280	987.897400	55.105656	113.317665	1.439088
9	2011	1	1	9	3	14.965301	988.211914	47.166938	117.327606	1.063374
10	2011	1	1	10	3	18.956082	988.054504	36.747087	72.488327	0.545695
11	2011	1	1	11	2	21.354047	987.463867	29.157015	12.976929	0.980117
12	2011	1	1	12	2	22.161257	986.386169	27.593071	4.777779	1.411915
13	2011	1	1	13	2	22.256216	985.736511	27.274695	5.355347	1.747909
14	2011	1	1	14	2	21.725884	985.521545	27.819289	6.877182	2.005287
15	2011	1	1	15	2	20.507269	985.279663	30.023284	10.448921	2.206879
16	2011	1	1	16	5	18.190106	985.668091	43.187772	16.174072	2.304465
17	2011	1	1	17	5	15.665956	986.169922	49.346720	23.485420	2.516871
18	2011	1	1	18	4	14.728780	986.536682	50.818219	30.906128	2.666706
19	2011	1	1	19	4	13.915211	987.175781	52.122064	38.189991	2.709682
20	2011	1	1	20	3	13.106395	987.245666	53.782549	45.614468	2.713746
21	2011	1	1	21	3	12.326217	986.971374	55.613451	53.939449	2.701125

In [3]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17520 entries, 0 to 17519
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Year                   17520 non-null  int64  
 1   Month                  17520 non-null  int64  
 2   Day                    17520 non-null  int64  
 3   Hour                   17520 non-null  int64  
 4   Dew Point              17520 non-null  int64  
 5   Temperature            17520 non-null  float64
 6   Pressure               17520 non-null  float64
 7   Relative Humidity      17520 non-null  float64
 8   Wind Direction         17520 non-null  float64
 9   Wind Speed             17520 non-null  float64
 10  Solar Radiation (GHI)  17520 non-null  int64  
dtypes: float64(5), int64(6)
memory usage: 1.5 MB
```

In [4]:
```
data.describe()
```

Out[4]:

	Year	Month	Day	Hour	Dew Point	Temperature	
count	17520.000000	17520.000000	17520.000000	17520.000000	17520.000000	17520.000000	1
mean	2011.500000	6.526027	15.720548	11.500000	11.375171	26.953731	
std	0.500014	3.447950	8.796498	6.922384	10.850196	8.417945	
min	2011.000000	1.000000	1.000000	0.000000	-28.000000	5.063506	
25%	2011.000000	4.000000	8.000000	5.750000	3.000000	21.611058	
50%	2011.500000	7.000000	16.000000	11.500000	12.000000	27.455196	
75%	2012.000000	10.000000	23.000000	17.250000	22.000000	32.031030	
max	2012.000000	12.000000	31.000000	23.000000	27.000000	52.157927	

Cleaning Data

In [5]:
```
## Hours where solar radiation is zero
clean1 = data[data['Hour']<6]
clean2 = data[data['Hour']>18]
```

In [6]:
```
clean1.head()
```

Out[6]:

	Year	Month	Day	Hour	Dew Point	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
0	2011	1	1	0	8	13.522659	986.761841	72.295858	37.288387	3.011042
1	2011	1	1	1	8	12.835814	986.441406	75.376186	37.686718	3.091243
2	2011	1	1	2	8	12.198058	985.736511	78.405198	35.053905	3.007649
3	2011	1	1	3	8	11.583500	985.525696	81.042980	30.135216	2.926715
4	2011	1	1	4	8	11.029578	985.661926	82.548508	24.402969	2.915177

In [7]:
```
clean2.head()
```

Out[7]:

	Year	Month	Day	Hour	Dew Point	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
19	2011	1	1	19	4	13.915211	987.175781	52.122064	38.189991	2.709682
20	2011	1	1	20	3	13.106395	987.245666	53.782549	45.614468	2.713746
21	2011	1	1	21	3	12.326217	986.971374	55.613451	53.939449	2.701125
22	2011	1	1	22	3	11.570982	986.042419	57.458358	62.491528	2.639709
23	2011	1	1	23	3	10.869849	985.439819	59.446614	71.250389	2.511646

In [8]:
```
data = data.drop(clean1.index,axis=0)
data = data.drop(clean2.index,axis=0)
```

In [9]:
```
data.head()
```

Out[9]:

	Year	Month	Day	Hour	Dew Point	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
6	2011	1	1	6	2	7.806705	986.249146	67.093181	95.835983	1.815175
7	2011	1	1	7	2	9.036408	987.038940	64.035483	105.198181	1.615310
8	2011	1	1	8	3	12.158280	987.897400	55.105656	113.317665	1.439088
9	2011	1	1	9	3	14.965301	988.211914	47.166938	117.327606	1.063374
10	2011	1	1	10	3	18.956082	988.054504	36.747087	72.488327	0.545695

In [10]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9490 entries, 6 to 17514
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Year                   9490 non-null   int64  
 1   Month                  9490 non-null   int64  
 2   Day                    9490 non-null   int64  
 3   Hour                   9490 non-null   int64  
 4   Dew Point              9490 non-null   int64  
 5   Temperature            9490 non-null   float64
 6   Pressure               9490 non-null   float64
 7   Relative Humidity      9490 non-null   float64
 8   Wind Direction         9490 non-null   float64
 9   Wind Speed             9490 non-null   float64
 10  Solar Radiation (GHI)  9490 non-null   int64  
dtypes: float64(5), int64(6)
memory usage: 889.7 KB
```

Preparing input and output data

In [11]:
```
x = data.iloc[:,0:10]
y = data.iloc[:,-1]
```

In [12]:
```
x.head()
```

Out[12]:

	Year	Month	Day	Hour	Dew Point	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
6	2011	1	1	6	2	7.806705	986.249146	67.093181	95.835983	1.815175
7	2011	1	1	7	2	9.036408	987.038940	64.035483	105.198181	1.615310
8	2011	1	1	8	3	12.158280	987.897400	55.105656	113.317665	1.439088
9	2011	1	1	9	3	14.965301	988.211914	47.166938	117.327606	1.063374
10	2011	1	1	10	3	18.956082	988.054504	36.747087	72.488327	0.545695

In [13]:
```
y.head()
```

Out[13]:
```
6       0
7       0
8     159
9     363
10    533
Name: Solar Radiation (GHI), dtype: int64
```

Feature Selection with Correlation Matrix

Correlation matrix with HeatMap

In [14]:
```
corr_matrix = data.corr()
```

In [15]:
```
corr_matrix
```

Out[15]:

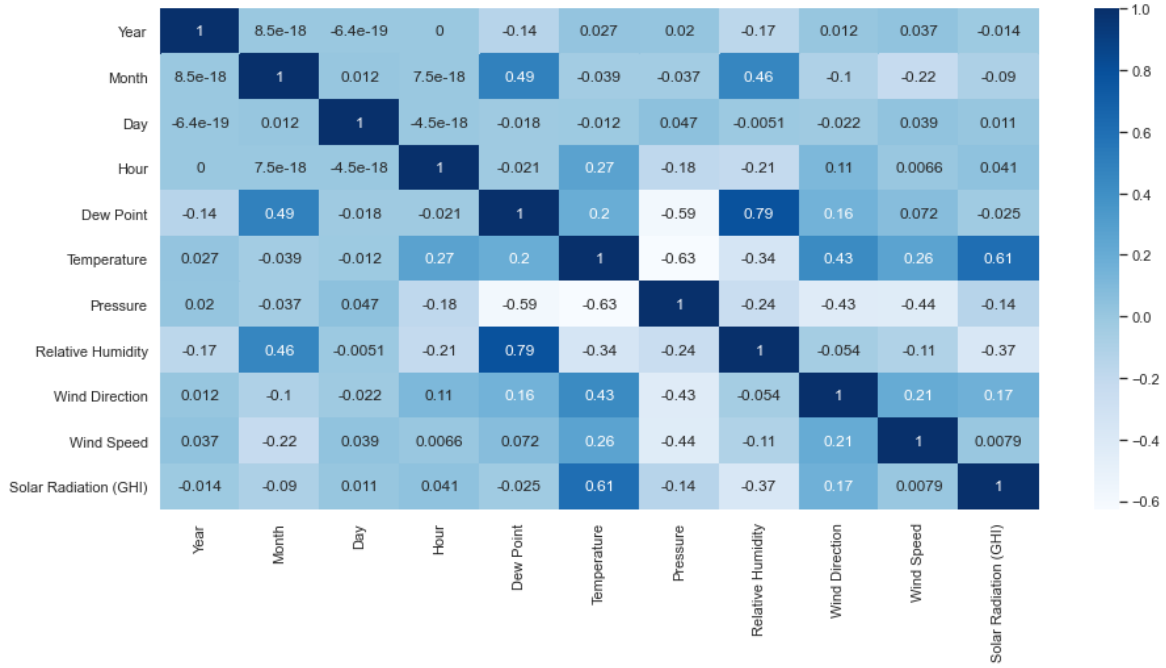
	Year	Month	Day	Hour	Dew Point	Temperature
Year	1.000000e+00	8.523444e-18	-6.383926e-19	0.000000e+00	-0.139773	0.026887
Month	8.523444e-18	1.000000e+00	1.189318e-02	7.486885e-18	0.491577	-0.038576
Day	-6.383926e-19	1.189318e-02	1.000000e+00	-4.458807e-18	-0.017772	-0.011873
Hour	0.000000e+00	7.486885e-18	-4.458807e-18	1.000000e+00	-0.020686	0.269499
Dew Point	-1.397731e-01	4.915774e-01	-1.777210e-02	-2.068626e-02	1.000000	0.198330
Temperature	2.688710e-02	-3.857553e-02	-1.187262e-02	2.694991e-01	0.198330	1.000000
Pressure	2.025241e-02	-3.665724e-02	4.670753e-02	-1.834597e-01	-0.587488	-0.627170
Relative Humidity	-1.663735e-01	4.609927e-01	-5.073932e-03	-2.118687e-01	0.790023	-0.344609
Wind Direction	1.199631e-02	-1.026477e-01	-2.171216e-02	1.136411e-01	0.160133	0.431627
Wind Speed	3.745498e-02	-2.169017e-01	3.854214e-02	6.593433e-03	0.072254	0.263510
Solar Radiation (GHI)	-1.389141e-02	-9.022667e-02	1.142019e-02	4.057306e-02	-0.024650	0.608713

In [16]:
```
features = corr_matrix.index
features
```

Out[16]:
```
Index(['Year', 'Month', 'Day', 'Hour', 'Dew Point', 'Temperature', 'Pressur
e',
       'Relative Humidity', 'Wind Direction', 'Wind Speed',
       'Solar Radiation (GHI)'],
      dtype='object')
```

In [17]:
```
plt.figure(figsize=(15,7))
sns.heatmap(data[features].corr(),annot = True,cmap='Blues')
```

Out[17]: <AxesSubplot:>



In [ ]:

In [ ]:

In [ ]:

In [ ]: