

```
In [5]: import numpy as np
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
import seaborn as sns
import matplotlib.pyplot as plt

In [6]: df=pd.read_csv("BigML_Dataset_5f58a4cc80852e40e6008034.csv")

In [7]: df
```

	Day of Year	Year	Month	Day	First Hour of Period	Is Daylight	Distance to Solar Noon	Average Temperature (Day)	Average Wind Direction (Day)	Average Wind Speed (Day)	Sky Cover	Visibility	Relative Humidity	Average Wind Speed (Period)	Average Barometric Pressure (Period)	Power Generated
0	245	2008	9	1	1	False	0.859897	69	28	7.5	0	10.0	75	8.0	29.82	0
1	245	2008	9	1	4	False	0.628535	69	28	7.5	0	10.0	77	5.0	29.85	0
2	245	2008	9	1	7	True	0.397172	69	28	7.5	0	10.0	70	0.0	29.89	5418
3	245	2008	9	1	10	True	0.165810	69	28	7.5	0	10.0	33	0.0	29.91	25477
4	245	2008	9	1	13	True	0.065553	69	28	7.5	0	10.0	21	3.0	29.89	30069
...
2915	243	2009	8	31	10	True	0.166453	63	27	13.9	4	10.0	75	10.0	29.93	6995
2916	243	2009	8	31	13	True	0.064020	63	27	13.9	1	10.0	66	15.0	29.91	29490
2917	243	2009	8	31	16	True	0.294494	63	27	13.9	2	10.0	68	21.0	29.88	17257
2918	243	2009	8	31	19	True	0.524968	63	27	13.9	2	10.0	81	17.0	29.87	677
2919	243	2009	8	31	22	False	0.755442	63	27	13.9	1	10.0	81	11.0	29.90	0

2920 rows × 16 columns

for showing 5 up data and same as ending 5 data

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

	Day of Year	Year	Month	Day	First Hour of Period	Is Daylight	Distance to Solar Noon	Average Temperature (Day)	Average Wind Direction (Day)	Average Wind Speed (Day)	Sky Cover	Visibility	Relative Humidity	Average Wind Speed (Period)	Average Barometric Pressure (Period)	Power Generated
0	245	2008	9	1	1	False	0.859897	69	28	7.5	0	10.0	75	8.0	29.82	0
1	245	2008	9	1	4	False	0.628535	69	28	7.5	0	10.0	77	5.0	29.85	0
2	245	2008	9	1	7	True	0.397172	69	28	7.5	0	10.0	70	0.0	29.89	5418
3	245	2008	9	1	10	True	0.165810	69	28	7.5	0	10.0	33	0.0	29.91	25477
4	245	2008	9	1	13	True	0.065553	69	28	7.5	0	10.0	21	3.0	29.89	30069

```
In [9]: df.tail()
```

	Day of Year	Year	Month	Day	First Hour of Period	Is Daylight	Distance to Solar Noon	Average Temperature (Day)	Average Wind Direction (Day)	Average Wind Speed (Day)	Sky Cover	Visibility	Relative Humidity	Average Wind Speed (Period)	Average Barometric Pressure (Period)	Power Generated
2915	243	2009	8	31	10	True	0.166453	63	27	13.9	4	10.0	75	10.0	29.93	6995
2916	243	2009	8	31	13	True	0.064020	63	27	13.9	1	10.0	66	15.0	29.91	29490
2917	243	2009	8	31	16	True	0.294494	63	27	13.9	2	10.0	68	21.0	29.88	17257
2918	243	2009	8	31	19	True	0.524968	63	27	13.9	2	10.0	81	17.0	29.87	677
2919	243	2009	8	31	22	False	0.755442	63	27	13.9	1	10.0	81	11.0	29.90	0

show all column name

```
In [10]: df.columns

Out[10]: Index(['Day of Year', 'Year', 'Month', 'Day', 'First Hour of Period',
              'Is Daylight', 'Distance to Solar Noon', 'Average Temperature (Day)',
              'Average Wind Direction (Day)', 'Average Wind Speed (Day)', 'Sky Cover',
              'Visibility', 'Relative Humidity', 'Average Wind Speed (Period)',
              'Average Barometric Pressure (Period)', 'Power Generated'],
              dtype='object')
```

syntax for null values

```
In [11]: df.isnull().sum()

Out[11]: Day of Year      0
Year      0
Month      0
Day      0
First Hour of Period      0
Is Daylight      0
Distance to Solar Noon      0
Average Temperature (Day)      0
Average Wind Direction (Day)      0
Average Wind Speed (Day)      0
Sky Cover      0
Visibility      0
Relative Humidity      0
Average Wind Speed (Period)      1
Average Barometric Pressure (Period)      0
Power Generated      0
dtype: int64

In [12]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2920 entries, 0 to 2919
Data columns (total 16 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Day of Year         2920 non-null   int64
 1   Year                2920 non-null   int64
 2   Month               2920 non-null   int64
 3   Day                 2920 non-null   int64
 4   First Hour of Period 2920 non-null   int64
 5   Is Daylight          2920 non-null   bool
 6   Distance to Solar Noon 2920 non-null   float64
 7   Average Temperature (Day) 2920 non-null   int64
 8   Average Wind Direction (Day) 2920 non-null   int64
 9   Average Wind Speed (Day) 2920 non-null   float64
10   Sky Cover           2920 non-null   int64
11   Visibility           2920 non-null   float64
12   Relative Humidity     2920 non-null   int64
13   Average Wind Speed (Period) 2919 non-null   float64
14   Average Barometric Pressure (Period) 2920 non-null   float64
15   Power Generated       2920 non-null   int64
dtypes: bool(1), float64(5), int64(10)
memory usage: 345.2 KB

In [13]: df.describe()

Out[13]:
```

	Day of Year	Year	Month	Day	First Hour of Period	Distance to Solar Noon	Average Temperature (Day)	Average Wind Direction (Day)	Average Wind Speed (Day)	Sky Cover	Visibility	Relative Humidity	Average Wind Speed (Period)	Average Barometric Pressure (Period)	Power Generated
count	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2920.000000	2919.000000	2920.000000	2920.000000
mean	183.334247	2008.665753	6.526027	15.720548	11.500000	0.503294	58.468493	24.953425	10.096986	1.987671	9.557705	73.513699	10.129154	30.017760	6979.846233
std	105.769919	0.471807	3.448442	8.797754	6.875041	0.298024	6.841200	6.915178	4.838185	1.411978	1.383884	15.077139	7.261547	0.142006	10312.336413
min	1.000000	2008.000000	1.000000	1.000000	1.000000	0.050401	42.000000	1.000000	1.100000	0.000000	0.000000	14.000000	0.000000	29.480000	0.000000
25%	92.000000	2008.000000	4.000000	8.000000	6.250000	0.243714	53.000000	25.000000	6.600000	1.000000	10.000000	65.000000	5.000000	29.920000	0.000000
50%	183.000000	2009.000000	7.000000	16.000000	11.500000	0.478957	59.000000	27.000000	10.000000	2.000000	10.000000	77.000000	9.000000	30.000000	404.000000
75%	275.000000	2009.000000	10.000000	23.000000	16.750000	0.739528	63.000000	29.000000	13.100000	3.000000	10.000000	84.000000	15.000000	30.110000	12723.500000
max	366.000000	2009.000000	12.000000	31.000000	22.000000	1.141361	78.000000	36.000000	26.600000	4.000000	10.000000	100.000000	40.000000	30.530000	36580.000000

```
In [14]: df['Day of Year'].unique()

Out[14]: array([245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257,
        258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270,
        271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283,
        284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296,
        297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309,
        310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322,
        323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335,
        336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348,
        349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361,
        362, 363, 364, 365, 366,  1,  2,  3,  4,  5,  6,  7,  8,
         9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,
        22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
        35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47,
        48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
        61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73,
        74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86,
        87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99,
        100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112,
        113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125,
        126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138,
        139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151,
        152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164,
        165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177,
        178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190,
        191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203,
        204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216,
        217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229,
        230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242,
        243])

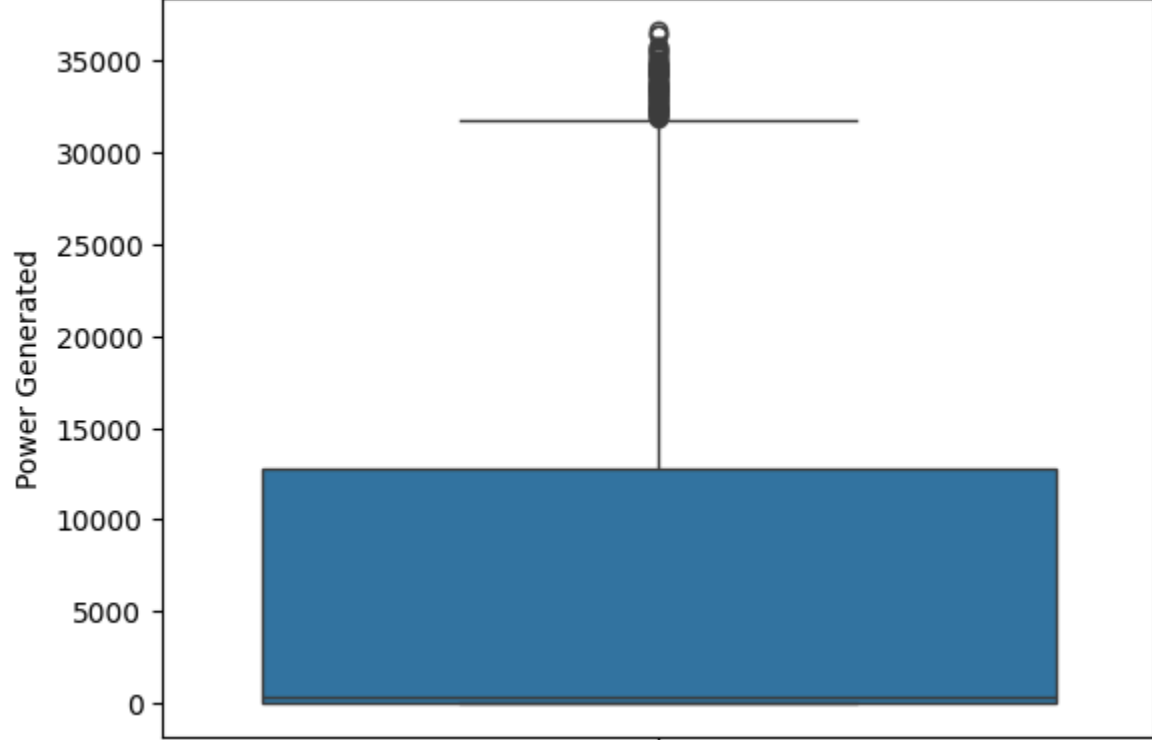
In [20]: df['Is Daylight'].unique()

Out[20]: <bound method Series.unique of 0      False
1      False
2      True
3      True
4      True
...
2915   True
2916   True
2917   True
2918   True
2919   False
Name: Is Daylight, Length: 2920, dtype: bool>
```

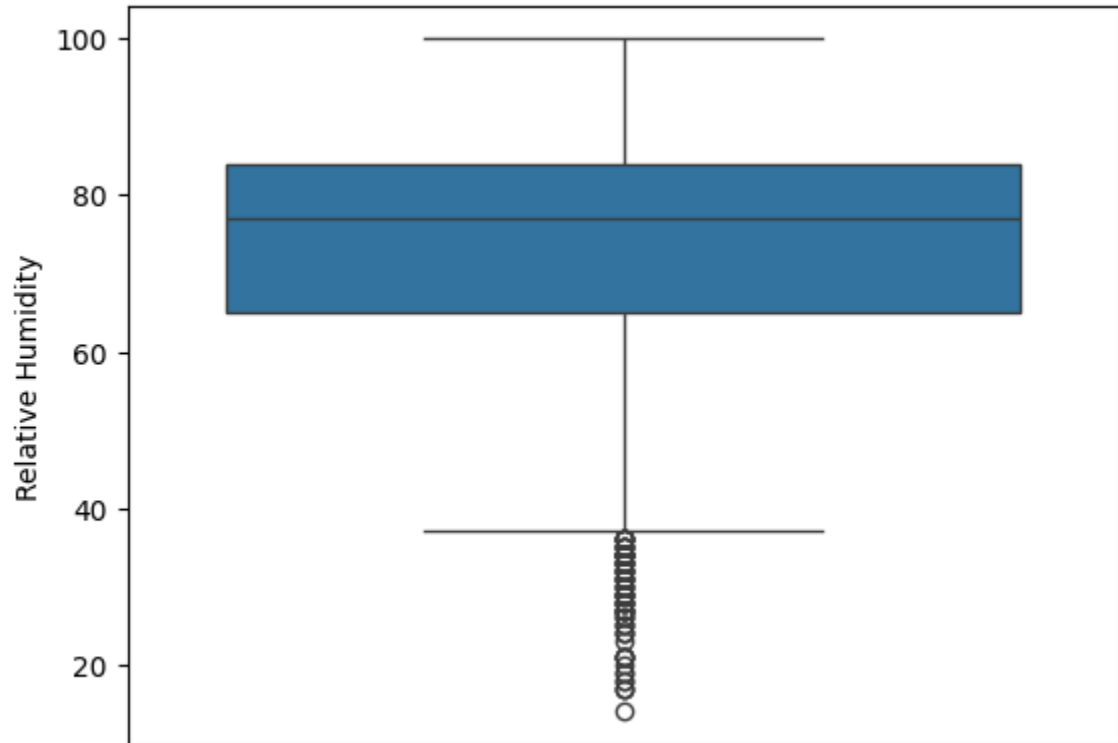
```
In [23]: df.head()
```

	Day of Year	Year	Month	Day	First Hour of Period	Is Daylight	Distance to Solar Noon	Average Temperature (Day)	Average Wind Direction (Day)	Average Wind Speed (Day)	Sky Cover	Visibility	Relative Humidity	Average Wind Speed (Period)	Average Barometric Pressure (Period)	Power Generated
0	245	2008	9	1	1	False	0.859897	69	28	7.5	0	10.0	75	8.0	29.82	0
1	245	2008	9	1	4	False	0.628535	69	28	7.5	0	10.0	77	5.0	29.85	0
2	245	2008	9	1	7	True	0.397172	69	28	7.5	0	10.0	70	0.0	29.89	5418
3	245	2008	9	1	10	True	0.165810	69	28	7.5	0	10.0	33	0.0	29.91	25477
4	245	2008	9	1	13	True	0.065553	69	28	7.5	0	10.0	21	3.0	29.89	30069

```
In [15]: sns.boxplot(df['Power Generated'])
plt.show()
```



```
In [27]: sns.boxplot(df['Relative Humidity'])
plt.show()
```



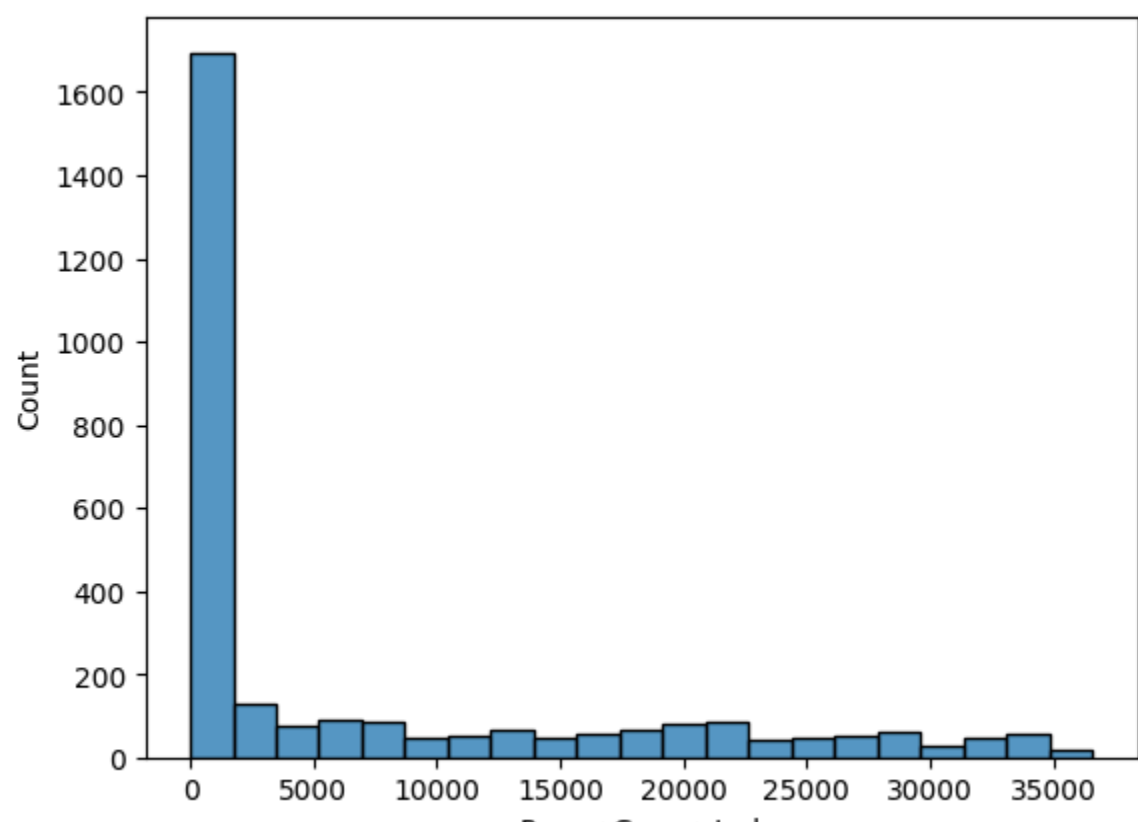
```
In [16]: df.head()
```

	Day of Year	Year	Month	Day	First Hour of Period	Is Daylight	Distance to Noon	Average Temperature (Day)	Average Wind Direction (Day)	Average Wind Speed (Day)	Sky Cover	Visibility	Relative Humidity	Average Wind Speed (Period)	Average Barometric Pressure (Period)	Power Generated
0	245	2008	9	1	1	False	0.859897	69	28	7.5	0	10.0	75	8.0	29.82	0
1	245	2008	9	1	4	False	0.628535	69	28	7.5	0	10.0	77	5.0	29.85	0
2	245	2008	9	1	7	True	0.397172	69	28	7.5	0	10.0	70	0.0	29.89	5418
3	245	2008	9	1	10	True	0.165810	69	28	7.5	0	10.0	33	0.0	29.91	25477
4	245	2008	9	1	13	True	0.065553	69	28	7.5	0	10.0	21	3.0	29.89	30069

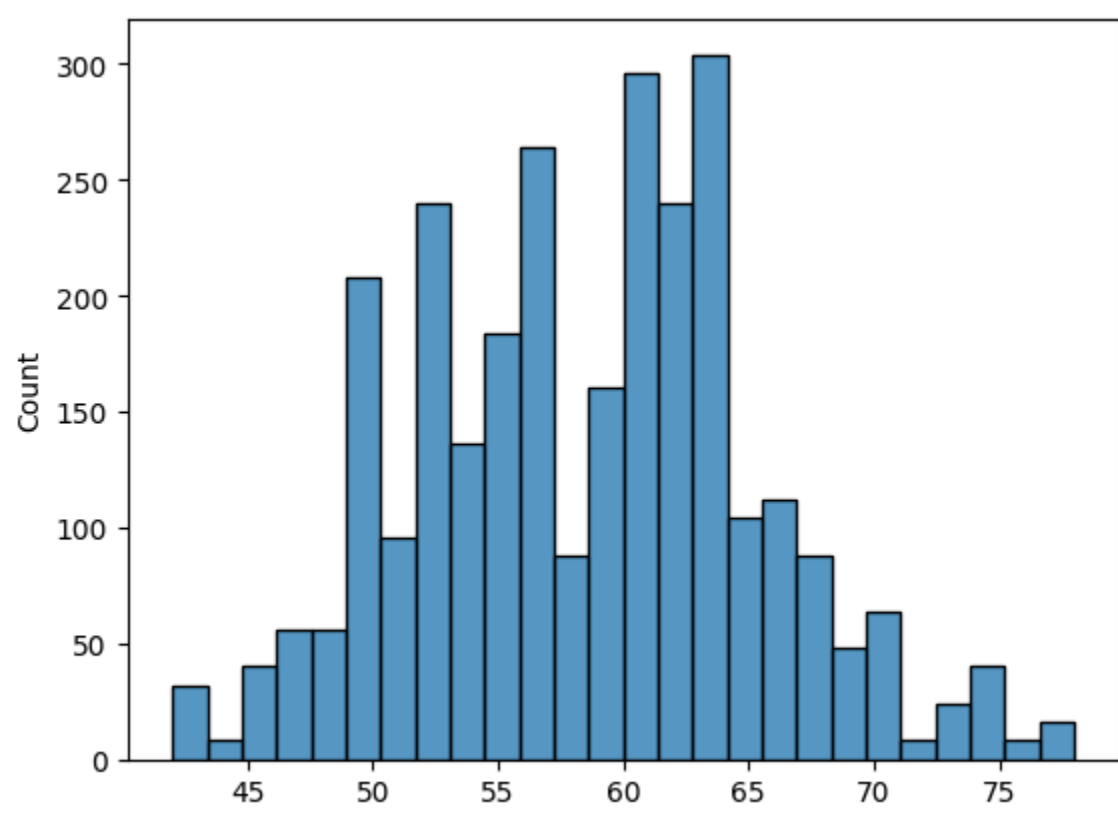
```
In [17]: df.columns

Out[17]: Index(['Day of Year', 'Year', 'Month', 'Day', 'First Hour of Period',
              'Is Daylight', 'Distance to Solar Noon', 'Average Temperature (Day)',
              'Average Wind Direction (Day)', 'Average Wind Speed (Day)', 'Sky Cover',
              'Visibility', 'Relative Humidity', 'Average Wind Speed (Period)',
              'Average Barometric Pressure (Period)', 'Power Generated'],
              dtype='object')
```

```
In [18]: sns.histplot(df['Power Generated'])
plt.show()
```

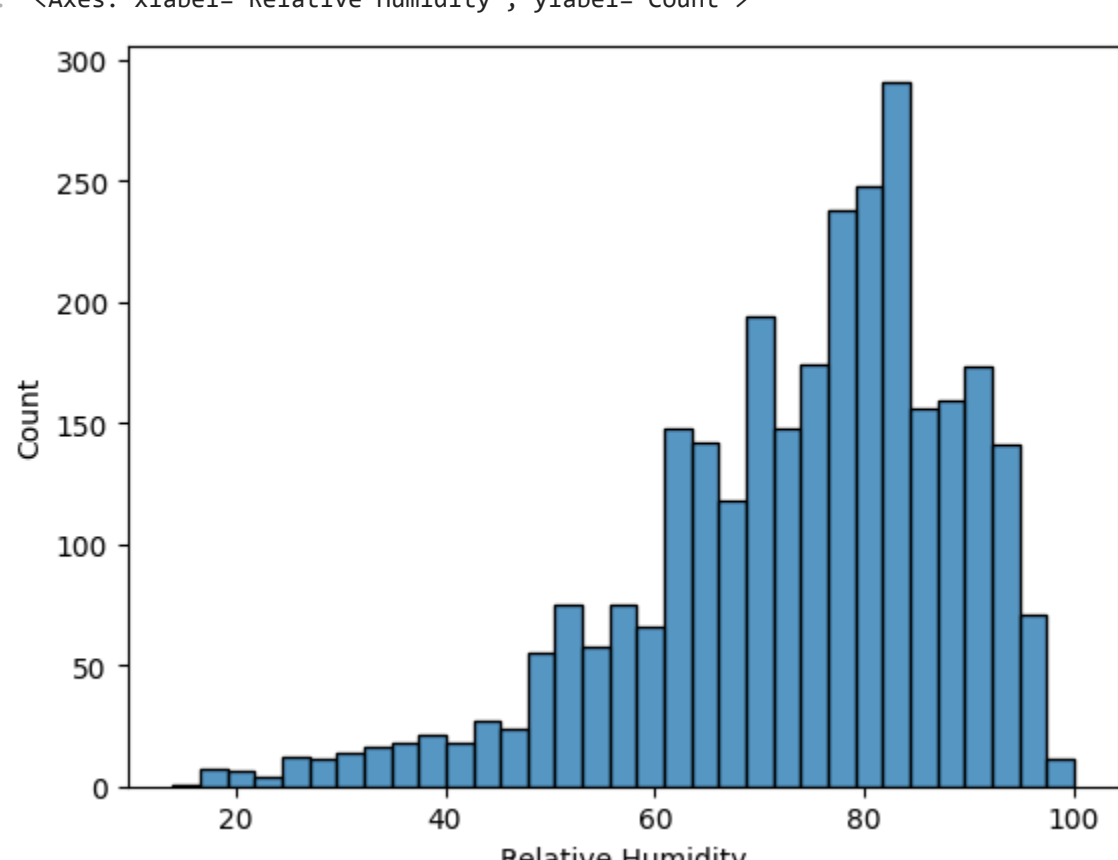


```
In [19]: sns.histplot(df['Average Temperature (Day)'])
plt.show()
```



```
In [26]: sns.histplot(df['Relative Humidity'])

Out[26]: <Axes: xlabel='Relative Humidity', ylabel='Count'>
```



```
In [ ]:
```