

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [5]: data=pd.read_csv("data_clean.csv")
data
```

Out[5]:

	Unnamed: 0	Ozone	Solar.R	Wind	Temp C	Month	Day	Year	Temp	Weather
0	1	41.0	190.0	7.4	67	5	1	2010	67	S
1	2	36.0	118.0	8.0	72	5	2	2010	72	C
2	3	12.0	149.0	12.6	74	5	3	2010	74	PS
3	4	18.0	313.0	11.5	62	5	4	2010	62	S
4	5	NaN	NaN	14.3	56	5	5	2010	56	S
...	...	...	...	...	...	...	...	...	...	...
153	154	41.0	190.0	7.4	67	5	1	2010	67	C
154	155	30.0	193.0	6.9	70	9	26	2010	70	PS
155	156	NaN	145.0	13.2	77	9	27	2010	77	S
156	157	14.0	191.0	14.3	75	9	28	2010	75	S
157	158	18.0	131.0	8.0	76	9	29	2010	76	C

158 rows × 10 columns

```
In [6]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 158 entries, 0 to 157
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0      158 non-null   int64
1   Ozone           120 non-null   float64
2   Solar.R         151 non-null   float64
3   Wind            158 non-null   float64
4   Temp C          158 non-null   object
5   Month           158 non-null   object
6   Day             158 non-null   int64
7   Year            158 non-null   int64
8   Temp            158 non-null   int64
9   Weather         155 non-null   object
dtypes: float64(3), int64(4), object(3)
memory usage: 12.5+ KB
```

```
In [7]: print(type(data))
        print(data.shape)
```

```
<class 'pandas.core.frame.DataFrame'>
(158, 10)
```

```
In [8]: data.shape
```

```
Out[8]: (158, 10)
```

```
In [9]: data.dtypes
```

```
Out[9]: Unnamed: 0      int64
        Ozone         float64
        Solar.R       float64
        Wind         float64
        Temp C        object
        Month         object
        Day           int64
        Year          int64
        Temp          int64
        Weather       object
        dtype: object
```

```
In [10]: data.describe
```

```
Out[10]: <bound method NDFrame.describe of      Unnamed: 0  Ozone  Solar.R  Wind  Temp
C Month  Day  Year  Temp Weather
0          1   41.0   190.0   7.4    67    5    1  2010    67    S
1          2   36.0   118.0   8.0    72    5    2  2010    72    C
2          3   12.0   149.0  12.6    74    5    3  2010    74    PS
3          4   18.0   313.0  11.5    62    5    4  2010    62    S
4          5    NaN     NaN  14.3    56    5    5  2010    56    S
..         ...    ...    ...    ...    ...    ...    ...    ...    ...
153        154   41.0   190.0   7.4    67    5    1  2010    67    C
154        155   30.0   193.0   6.9    70    9   26  2010    70    PS
155        156    NaN   145.0  13.2    77    9   27  2010    77    S
156        157   14.0   191.0  14.3    75    9   28  2010    75    S
157        158   18.0   131.0   8.0    76    9   29  2010    76    C

[158 rows x 10 columns]>
```

```
In [ ]: data1=data.drop(["Unnamed:0",])
```

```
In [15]: data['Month']=pd.to_numeric(data['Month'],errors='coerce')
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 158 entries, 0 to 157
Data columns (total 10 columns):
 #   Column      Non-Null Count  Dtype  
---  --
 0   Unnamed: 0   158 non-null   int64  
 1   Ozone        120 non-null   float64
 2   Solar.R     151 non-null   float64
 3   Wind         158 non-null   float64
 4   Temp C       158 non-null   object  
 5   Month        157 non-null   float64
 6   Day          158 non-null   int64  
 7   Year         158 non-null   int64  
 8   Temp         158 non-null   int64  
 9   Weather      155 non-null   object  
dtypes: float64(4), int64(4), object(2)
memory usage: 12.5+ KB
```

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

Out[17]:

Unnamed: 0	Ozone	Solar.R	Wind	Temp C	Month	Day	Year	Temp	Weather
------------	-------	---------	------	--------	-------	-----	------	------	---------

```
In [ ]:
```