Assignment 3

Problem Statement-

Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for assignment number 1 and 2

Importing python libraries

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

Loading a CSV file into a dataframe

In [2]:
```
A  =  pd.read_csv(r"C:\Users\HP\Downloads\airquality_dataset.csv")
A.head()
```

Out[2]:

	Unnamed: 0	Ozone	Solar.R	Wind	Temp	Month	Day	Humidity
0	1	41.0	190.0	7.4	67	5	1	High
1	2	36.0	118.0	8.0	72	5	2	High
2	3	12.0	149.0	12.6	74	5	3	Low
3	4	18.0	313.0	11.5	62	5	4	NaN
4	5	NaN	NaN	14.3	56	5	5	High

In [3]:
```
A.shape
```

Out[3]:
```
(153, 8)
```

Data Cleaning

In [4]:
```
A = A.drop('Unnamed: 0',axis =1)
A.head()
```

Out[4]:

	Ozone	Solar.R	Wind	Temp	Month	Day	Humidity
0	41.0	190.0	7.4	67	5	1	High
1	36.0	118.0	8.0	72	5	2	High
2	12.0	149.0	12.6	74	5	3	Low
3	18.0	313.0	11.5	62	5	4	NaN
4	NaN	NaN	14.3	56	5	5	High

Replacing numerical null values

In [5]:
```
A['Ozone']=A['Ozone'].fillna(A['Ozone'].mean())

A['Solar.R']=A['Solar.R'].fillna(A['Solar.R'].mean())

A["Wind"]  =  A["Wind"].fillna(A["Wind"].mean())
```

Replacing categorical null values

In [6]:
```
A['Humidity']=A['Humidity'].fillna(A['Humidity'].mode()[0])
```

In [7]:
```
A.isnull().sum()
```

Out[7]:
```
Ozone       0
Solar.R     0
Wind        0
Temp        0
Month       0
Day         0
Humidity    0
dtype: int64
```

Data tranformation

Using Label Encoding for "Humidity" column

In [8]:
```
from sklearn.preprocessing import LabelEncoder

label_encoder=LabelEncoder()

A['Humidity']=label_encoder.fit_transform(A['Humidity'])

A["Humidity"].unique()
```

Out[8]:
```
array([0, 1, 2])
```

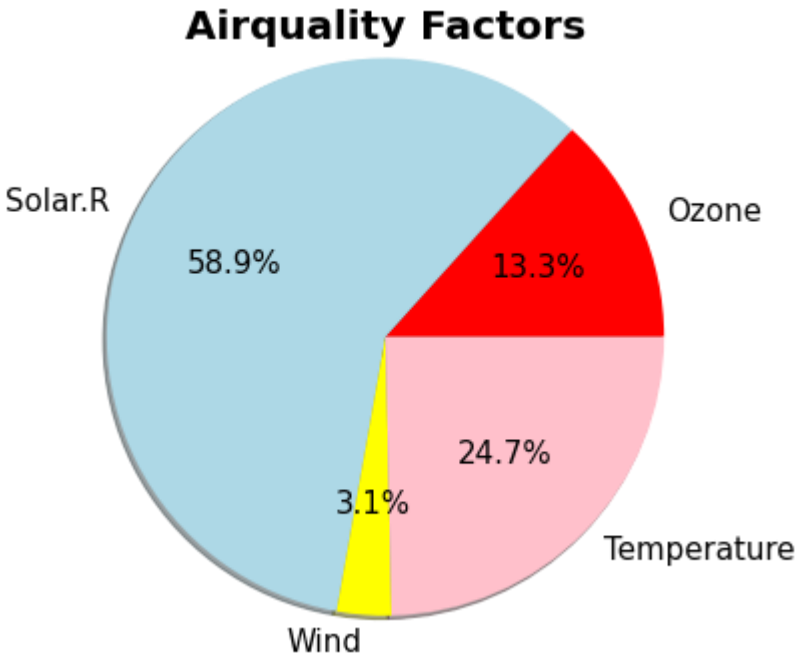
In [9]:
```
A.dtypes
```

Out[9]:
```
Ozone       float64
Solar.R     float64
Wind        float64
Temp          int64
Month         int64
Day           int64
Humidity      int32
dtype: object
```

Vizualizing the data

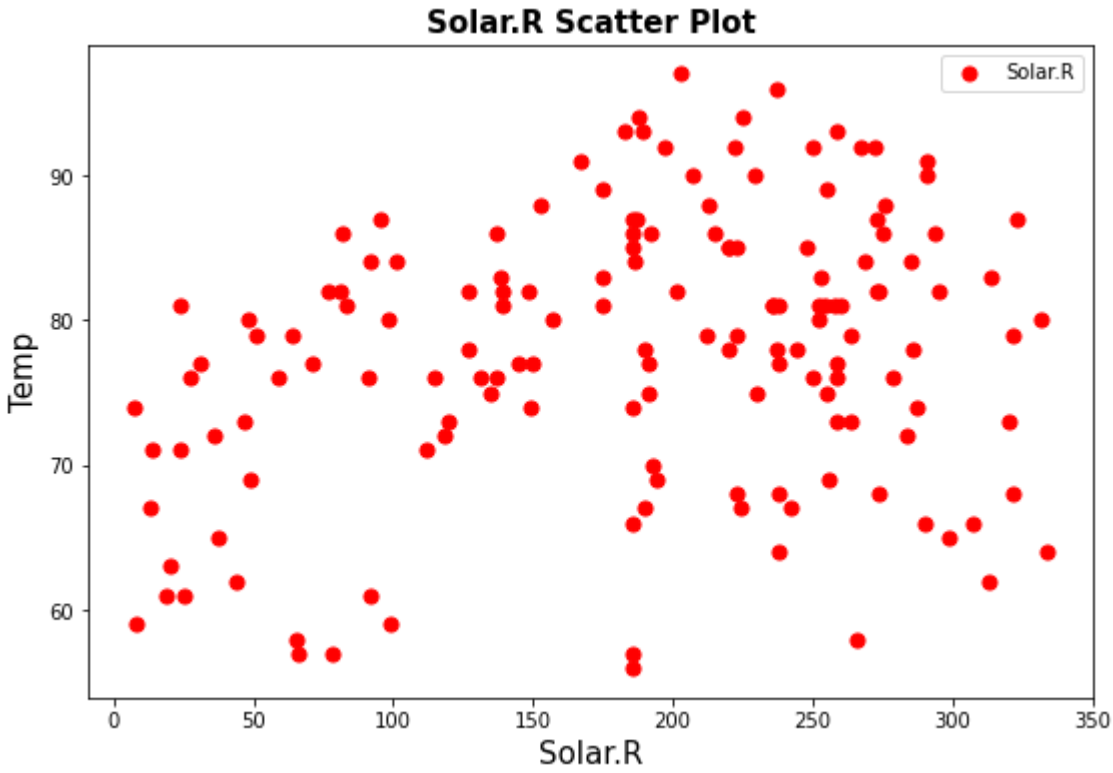
1. Pie chart

In [10]:
```
myplt  =  ['Ozone','Solar.R','Wind','Temperature']

sizes= [A['Ozone'].mean(),A['Solar.R'].mean(),A['Wind'].mean(),A['Temp'].mean()]

colors=['red','lightblue','yellow','pink']
textprops = {"fontsize":15}
plt.title("Airquality  Factors",  fontsize=20,  style="normal",  pad=40,  fontweight  ="

plt.pie(sizes,  labels  =  myplt  ,colors=colors,  autopct='%1.1f%%',radius=1.6,shadow=

plt.show()
```



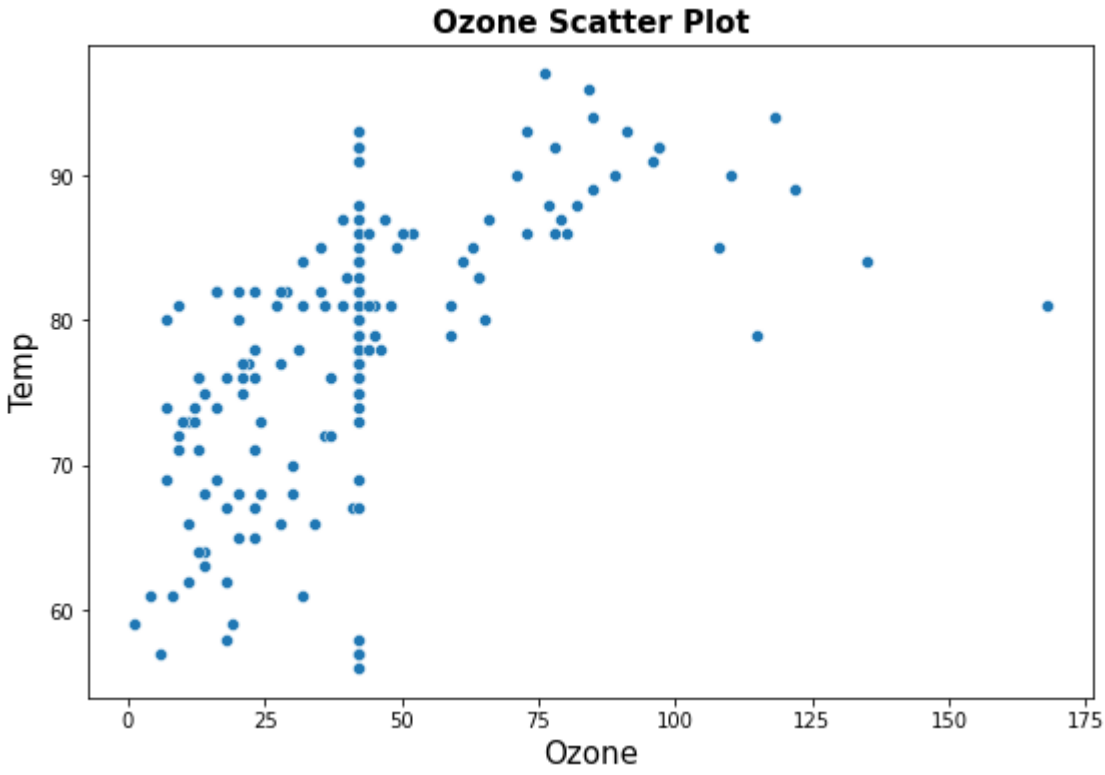
2. Scatter plot

In [11]:
```
plt.figure(figsize=(9,6))
plt.scatter(x  =  A["Solar.R"],y  =  A["Temp"],c="red",marker="o",s=50)     #using matp

plt.xlabel("Solar.R",fontsize=15)
plt.ylabel("Temp",fontsize=15)
plt.title("Solar.R  Scatter  Plot",fontsize=15,fontweight  ="bold")

plt.legend(["Solar.R"])
plt.show()
```



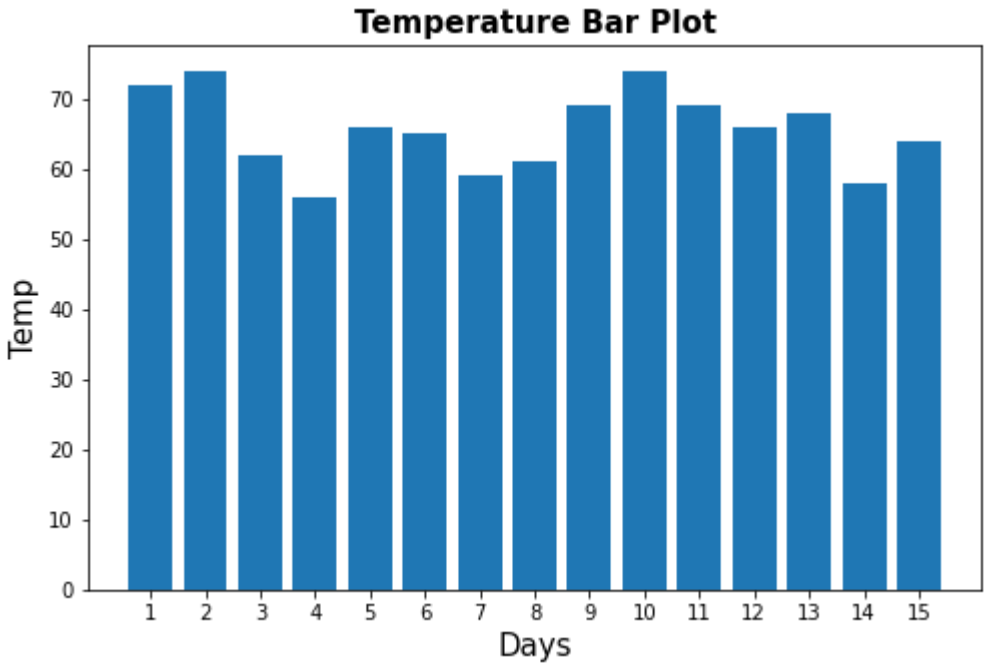
In [12]:
```
plt.figure(figsize=(9,6))
sns.scatterplot(x=A['Ozone'],y=A['Temp'],data  =A)      #using seaborn library

plt.xlabel("Ozone",fontsize=15)
plt.ylabel("Temp",fontsize=15)
plt.title("Ozone  Scatter  Plot",fontsize=15,fontweight  ="bold")
```

Out[12]:
```
Text(0.5, 1.0, 'Ozone Scatter Plot')
```



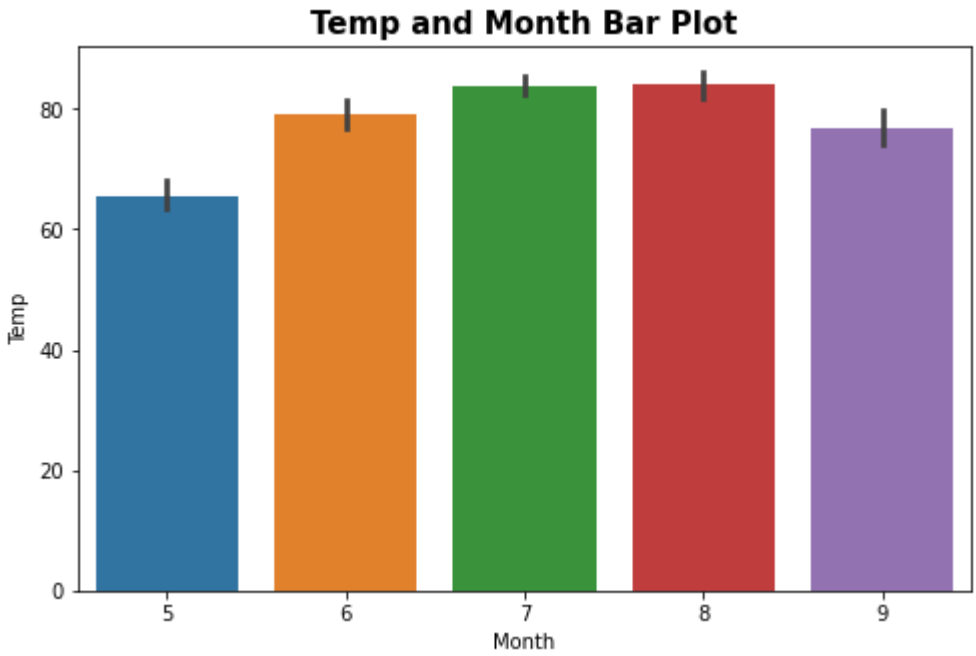
3. Bar Plot

In [13]:
```
h= A.iloc[1:16,3]     #taking first 15 rows of temp
y_pos=np.arange(len(h))
v=range(1,16)

plt.figure(figsize=(8,5))
plt.bar(y_pos,height=h,width=0.8,  bottom=None,align='center',  data=None)
plt.xticks(y_pos,v)
plt.title('Temperature  Bar  Plot',fontsize=15,fontweight  ="bold")
plt.xlabel("Days",fontsize=15)
plt.ylabel("Temp",fontsize=15)

plt.show()
```



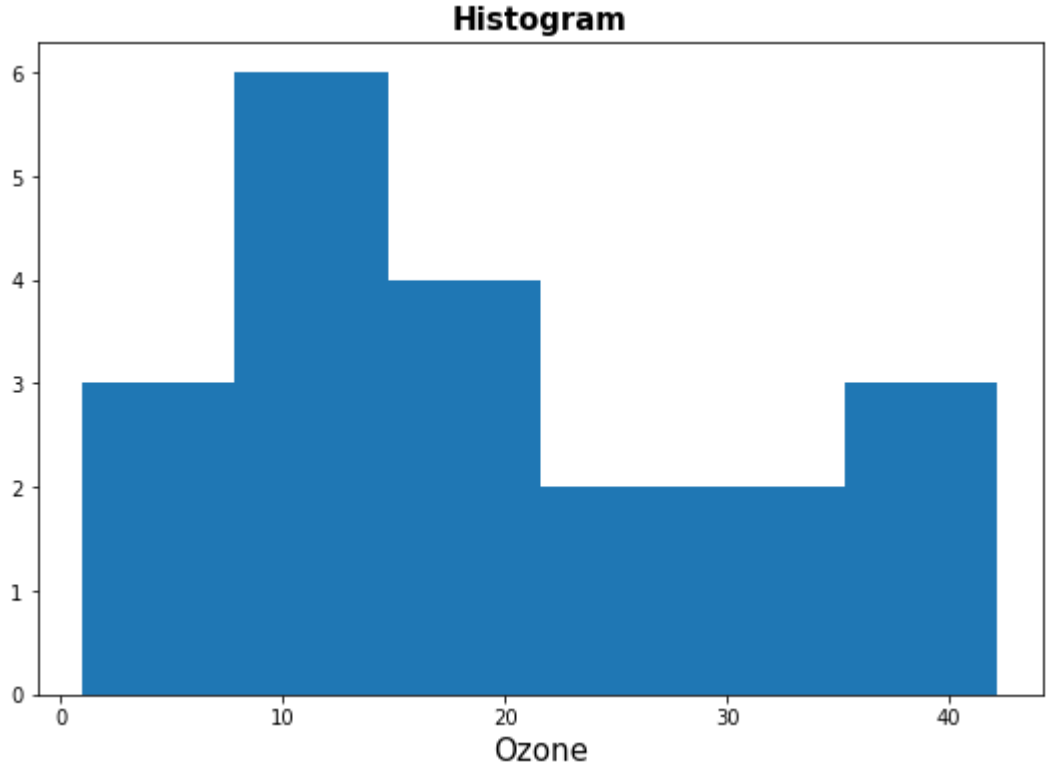
In [14]:
```
plt.figure(figsize=(8,5))
sns.barplot(A["Month"],A["Temp"])
plt.title('Temp  and  Month  Bar  Plot',fontsize=15,fontweight  ="bold")
plt.show()
```

```
c:\users\hp\appdata\local\programs\python\python39\lib\site-packages\seaborn\_deco
rators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. F
rom version 0.12, the only valid positional argument will be `data`, and passing o
ther arguments without an explicit keyword will result in an error or misinterpret
ation.
  warnings.warn(
```



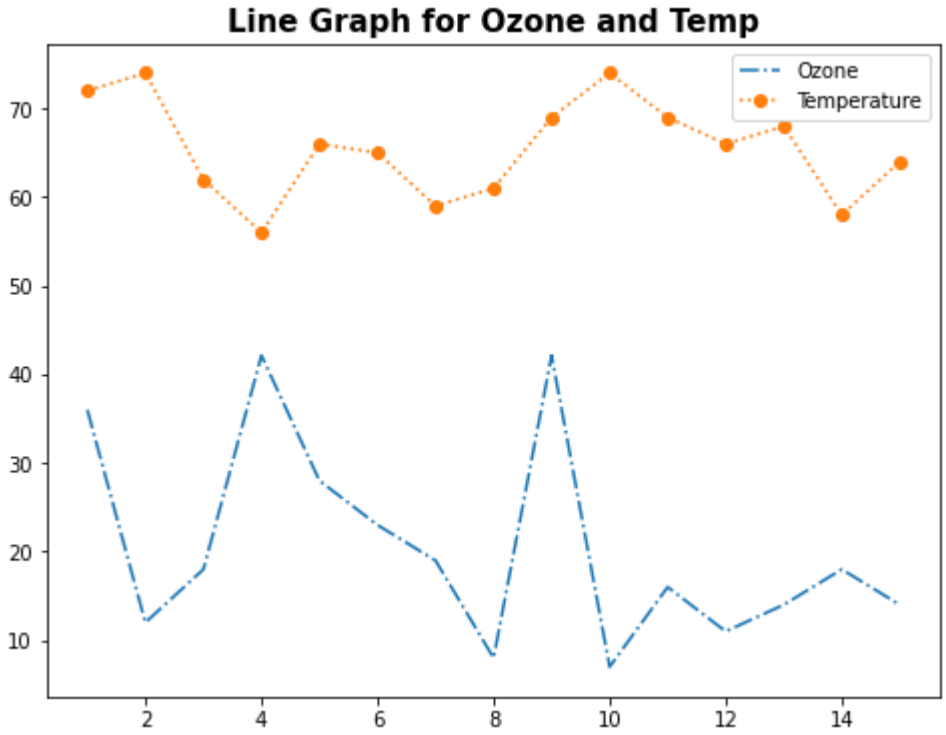
4. Histogram

In [15]:
```
x=A.iloc[1:21,0] #taking first 20 row of Ozone column

plt.figure(figsize=(9,6))
plt.hist(x,bins='auto')
plt.title('Histogram',fontsize=15,  fontweight  ="bold")
plt.xlabel("Ozone",fontsize=15)
plt.show()
```



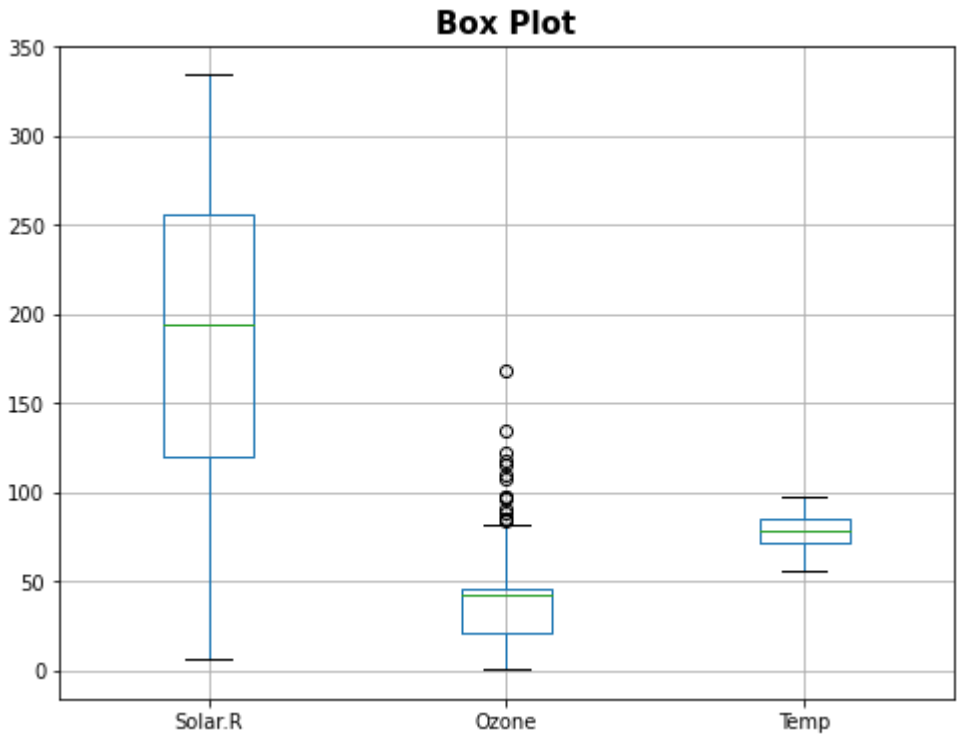
5. Line Graph

In [16]:
```
O=A.iloc[1:16,0]
T=A.iloc[1:16,3]

plt.figure(figsize=(8,6))
plt.plot(O,label="Ozone",linestyle="dashdot")
plt.plot(T,label="Temperature",marker="o",linestyle="dotted")
plt.title("Line Graph for Ozone and Temp",fontsize=15,fontweight="bold")

plt.legend()
plt.show()
```



6. Box Plot

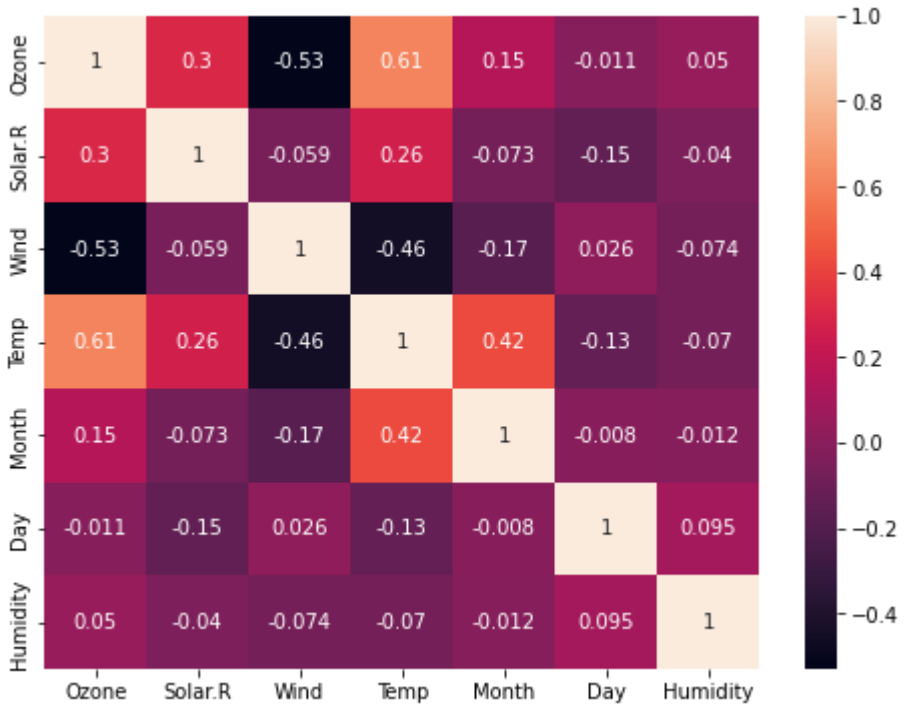
In [17]:
```
plt.figure(figsize=(8,6))
A.boxplot(column=["Solar.R","Ozone","Temp"],grid='True')
plt.title("Box Plot",fontsize=15,fontweight="bold")
plt.show()
```



7. Heat Map

In [18]:
```
correlation = A.corr()
plt.figure(figsize=(8,6))
sns.heatmap(correlation, annot = True)
```

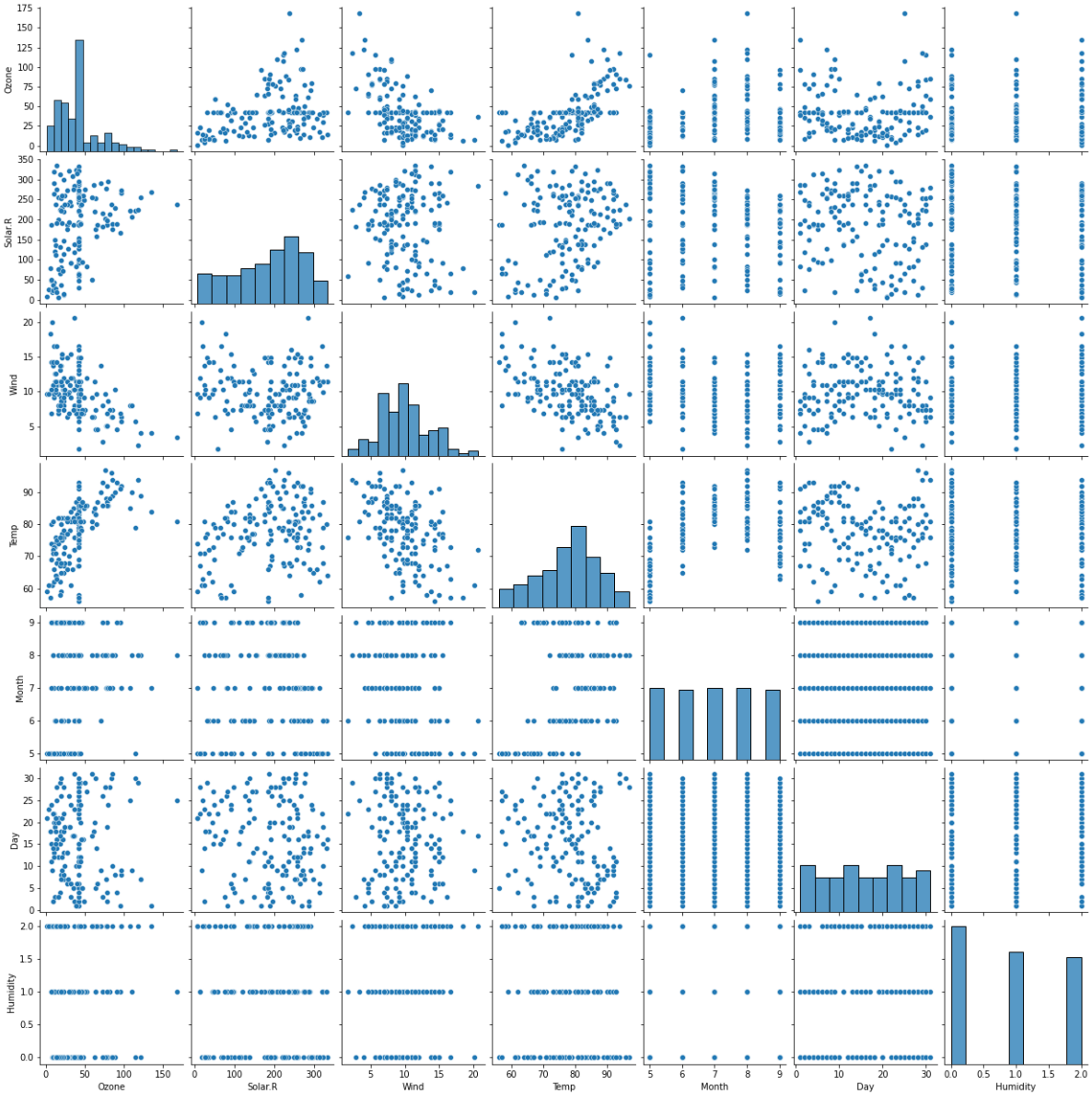
Out[18]: <AxesSubplot:>



8. Pair Plot

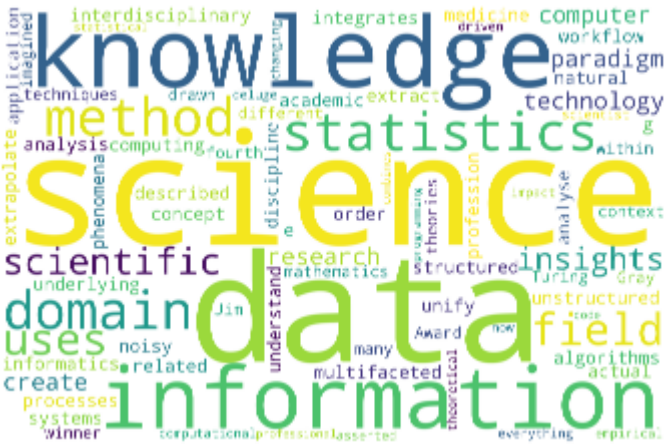
In [19]:
```
sns.pairplot(A)
```

Out[19]: <seaborn.axisgrid.PairGrid at 0x2b2143f5550>

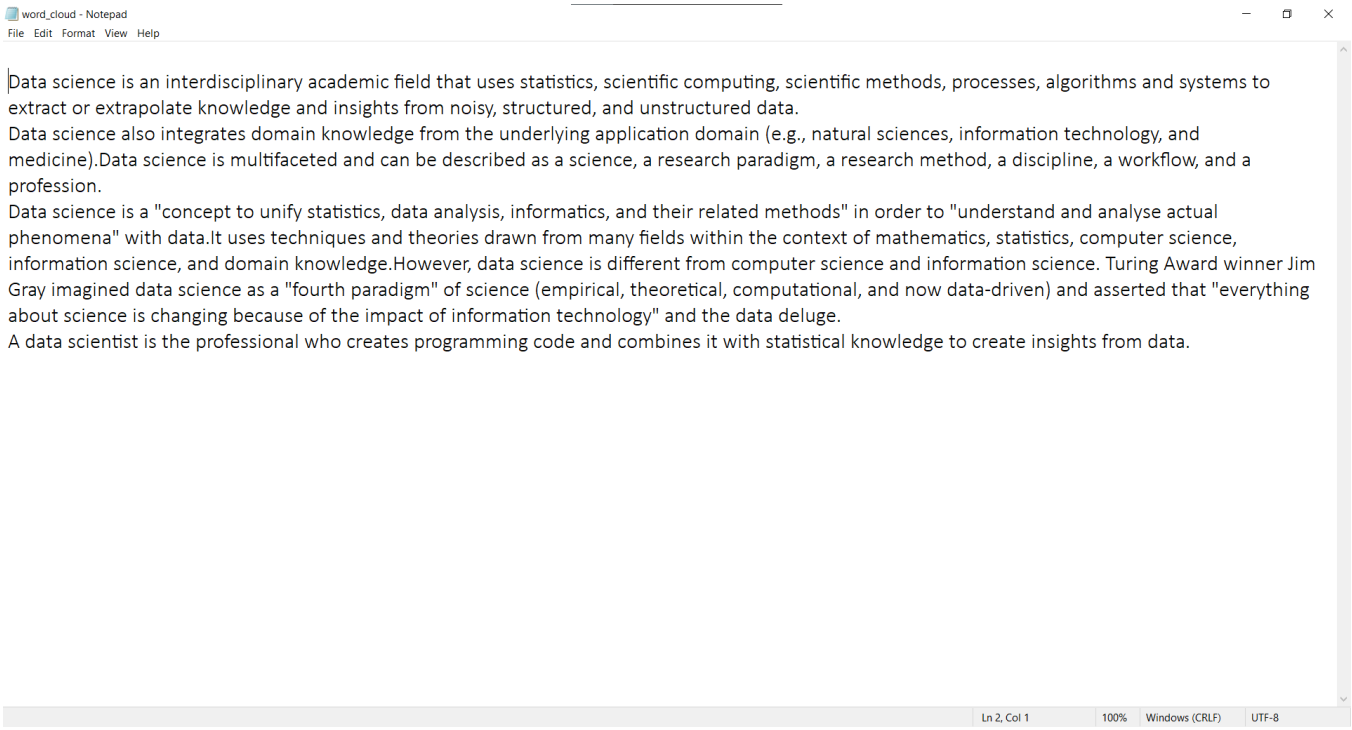


9. Word Cloud

In [20]:
```
from wordcloud import WordCloud, STOPWORDS
text = open("word_cloud.txt").read()
wc = WordCloud(background_color="white", height=6225, width=9450).generate(text)
plt.imshow(wc)
plt.axis("off")
plt.show()
```



Word_Cloud txt file:



word_cloud - Notepad

File Edit Format View Help

Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data.
Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine).Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession.
Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" in order to "understand and analyse actual phenomena" with data.It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge.However, data science is different from computer science and information science. Turing Award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational, and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.
A data scientist is the professional who creates programming code and combines it with statistical knowledge to create insights from data.

Ln 2, Col 1 | 100% | Windows (CRLF) | UTF-8