```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud, STOPWORDS

air_quality = pd.read_csv(r'/content/airquality.csv')
heart_disease = pd.read_csv(r'/content/HeartDisease.csv')
forest_fires = pd.read_csv(r'/content/forestfires.csv')

air_quality['Humidity'] = air_quality['Humidity'].str.lower()
#converted to lowercase all values in Humidity

print(air_quality.head())
```

```
   Ozone  Solar.R  Wind  Temp  Month  Day Humidity
0   41.0    190.0   7.4  67.0      5    1     high
1   36.0    118.0   8.0  72.0      5    2   medium
2   12.0    149.0  12.6  74.0      5    3      low
3   18.0    313.0  11.5  62.0      5    4   medium
4    NaN      NaN  14.3  56.0      5    5   medium
```

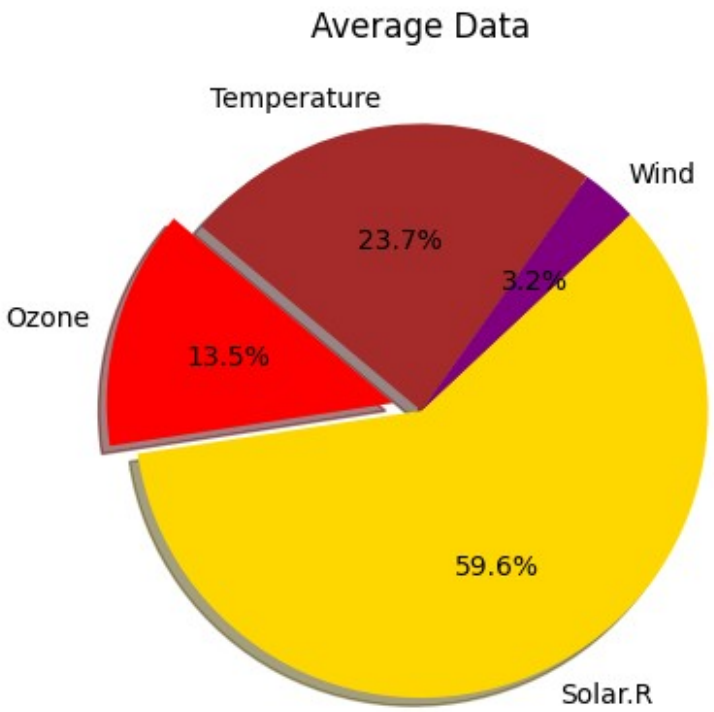
```
print(heart_disease.head())
```

```
   age  sex  chest pain  trestbps  chol  fps  restecg  thalach  exang
\
0   63    1           1       145   233    1        2      150      0

1   67    1           4       160   286    0        2      108      1

2   67    1           4       120   229    0        2      129      1

3   37    1           3       130   250    0        0      187      0

4   41    0           2       130   204    0        2      172      0


   oldpeak  slope  ca thal  num
0      2.3      3   0    6    0
1      1.5      2   3    3    2
2      2.6      2   2    7    1
3      3.5      3   0    3    0
4      1.4      1   0    3    0
```

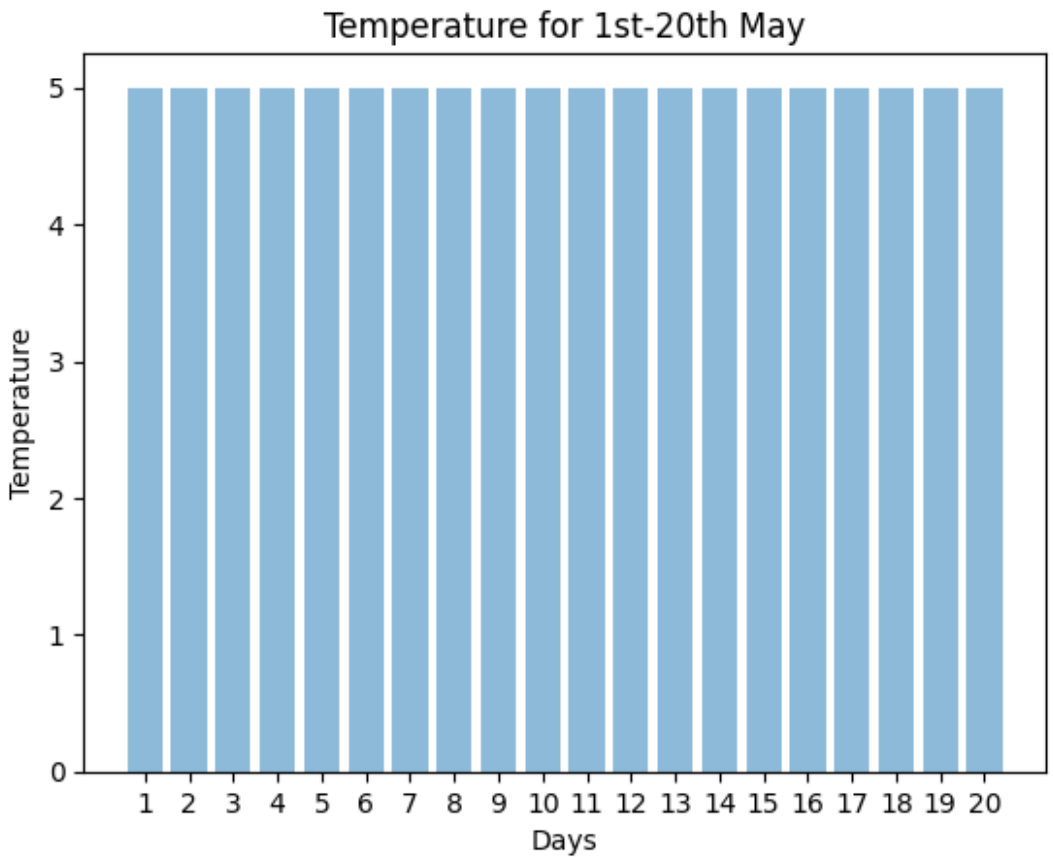
```
print(forest_fires.head())
```

```
   X  Y month  day  FFMC   DMC     DC  ISI  temp  RH  wind  rain  area
0  7  5   mar  fri  86.2  26.2   94.3  5.1   8.2  51   6.7   0.0   0.0
1  7  4   oct  tue  90.6  35.4  669.1  6.7  18.0  33   0.9   0.0   0.0
2  7  4   oct  sat  90.6  43.7  686.9  6.7  14.6  33   1.3   0.0   0.0
```

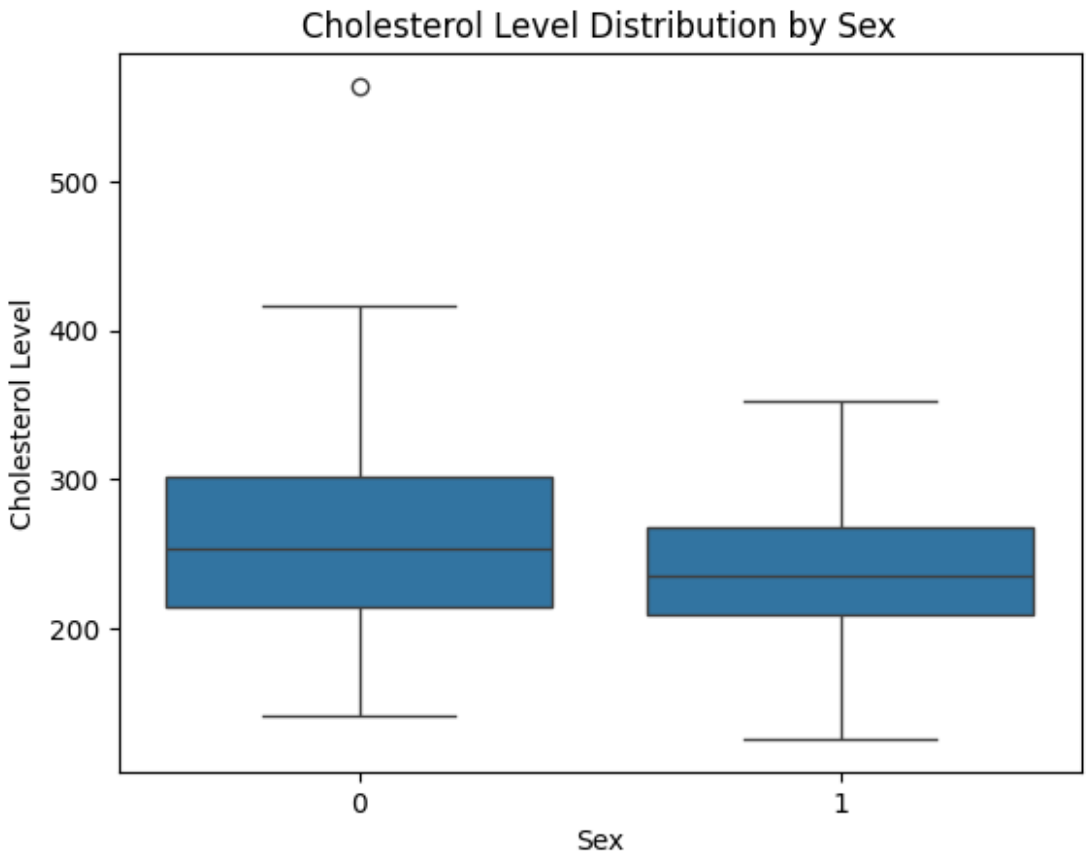
```
3  8  6   mar  fri  91.7  33.3   77.5  9.0   8.3  97   4.0   0.2   0.0
4  8  6   mar  sun  89.3  51.3  102.2  9.6  11.4  99   1.8   0.0   0.0

forest_fires['month'] =
forest_fires['month'].astype('category').cat.codes
forest_fires['day'] = forest_fires['day'].astype('category').cat.codes

print(forest_fires.head())

   X  Y  month  day  FFMC   DMC     DC  ISI  temp  RH  wind  rain
area
0  7  5      7    0  86.2  26.2   94.3  5.1   8.2  51   6.7   0.0
0.0
1  7  4     10    5  90.6  35.4  669.1  6.7  18.0  33   0.9   0.0
0.0
2  7  4     10    2  90.6  43.7  686.9  6.7  14.6  33   1.3   0.0
0.0
3  8  6      7    0  91.7  33.3   77.5  9.0   8.3  97   4.0   0.2
0.0
4  8  6      7    3  89.3  51.3  102.2  9.6  11.4  99   1.8   0.0
0.0

# 1. PIE CHART - Visualizing average values from airquality.csv
data = pd.read_csv("airquality.csv")
labels = ['Ozone', 'Solar.R', 'Wind', 'Temperature']
sizes = [data['Ozone'].mean(), data['Solar.R'].mean(),
data['Wind'].mean(), data['Temp'].mean()]
colors = ['red', 'gold', 'purple', 'brown']
explode = (0.1, 0, 0, 0)
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=140)
plt.title('Average Data')
plt.savefig('pie_chart.png')
plt.show()
```



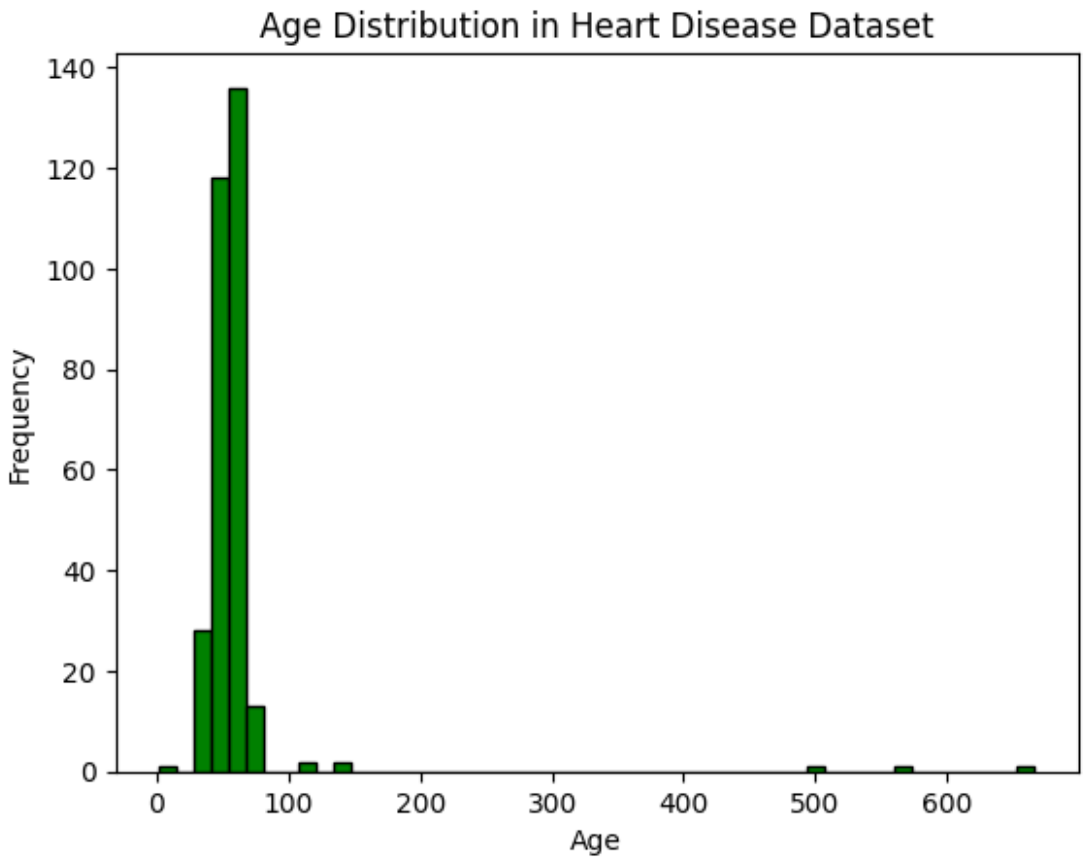

```
# 2. BAR PLOT - Temperature variation for the first 20 days
data = pd.read_csv("airquality.csv")
temp_values = data.iloc[1:21, 4]
y_pos = np.arange(len(temp_values))
days = range(1, 21)
plt.bar(y_pos, temp_values, align='center', alpha=0.5)
plt.xticks(y_pos, days)
plt.ylabel('Temperature')
plt.xlabel('Days')
plt.title('Temperature for 1st-20th May')
plt.savefig('bar_plot.png')
plt.show()
```



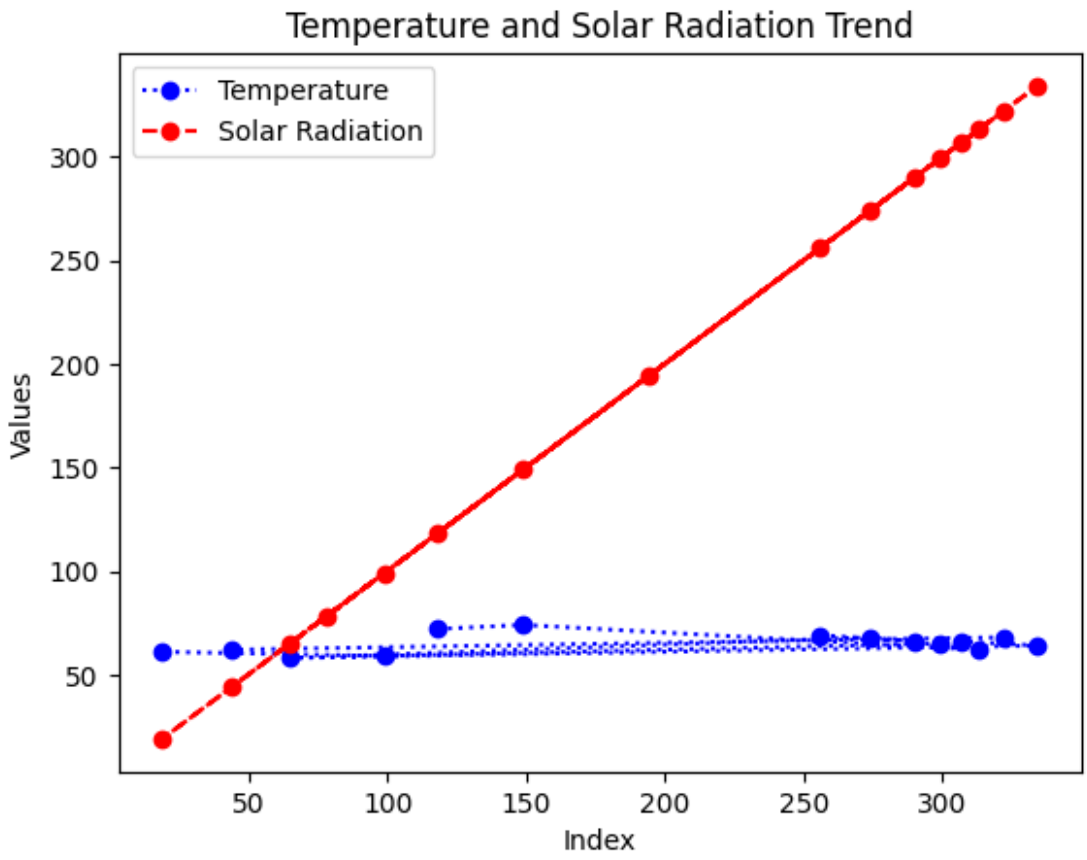

```
# 3. BOXPLOT - Heart disease dataset (example on cholesterol levels)
sns.boxplot(x='sex', y='chol', data=heart_disease)
plt.xlabel("Sex")
plt.ylabel("Cholesterol Level")
plt.title("Cholesterol Level Distribution by Sex")
plt.savefig('boxplot.png')
plt.show()
```



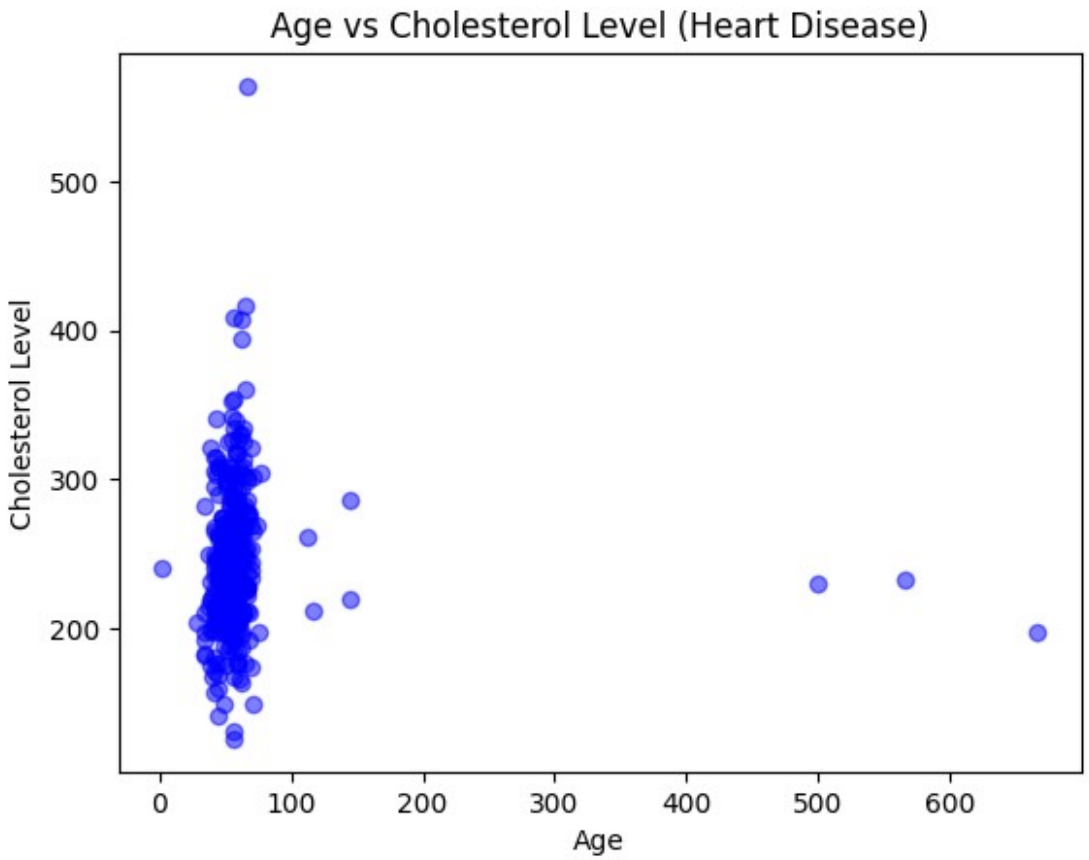

```
# 4. HISTOGRAM - Age distribution in heart disease dataset
plt.hist(heart_disease['age'], bins=50, color='green',
edgecolor='black')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Age Distribution in Heart Disease Dataset')
plt.savefig('histogram.png')
plt.show()
```



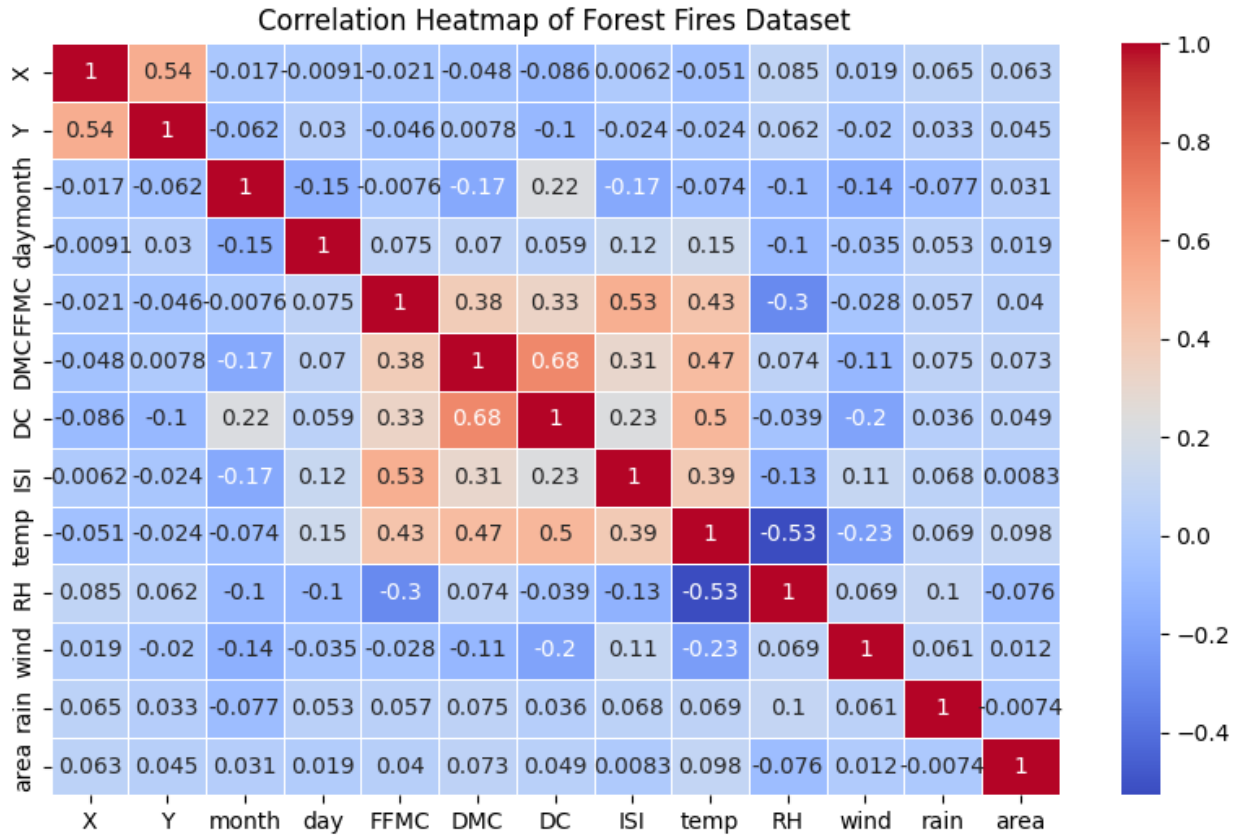

```
# 5. LINE GRAPH - Temperature trend in air quality dataset

plt.plot(air_quality.iloc[1:20, 1], air_quality.iloc[1:20]['Temp'],
         color='blue', marker='o', linestyle="dotted",
label="Temperature")

plt.plot(air_quality.iloc[1:20, 1], air_quality.iloc[1:20]['Solar.R'],
         color='red', marker='o', linestyle="dashed", label="Solar
Radiation")

plt.xlabel('Index')
plt.ylabel('Values')
plt.title('Temperature and Solar Radiation Trend')
plt.legend()
plt.savefig('line_graph.png')
plt.show()
```




```
# 6. SCATTER PLOTS
# Cholesterol vs Age (Heart Disease dataset)
plt.scatter(heart_disease['age'], heart_disease['chol'], color='blue',
alpha=0.5)
plt.xlabel('Age')
plt.ylabel('Cholesterol Level')
plt.title('Age vs Cholesterol Level (Heart Disease)')
plt.savefig('scatter_plot_heartdisease.png')
plt.show()
```




```
# 7. HEAT MAP - Correlation matrix of forest fires dataset
plt.figure(figsize=(10, 6))
sns.heatmap(forest_fires.corr(), annot=True, cmap='coolwarm',
linewidths=0.5)
plt.title("Correlation Heatmap of Forest Fires Dataset")
plt.savefig('heatmap.png')
plt.show()
```




```
# 8. WORD CLOUD -
wordcloud = WordCloud(width=800, height=400,
background_color='white').generate('
'.join(air_quality['Humidity'].astype(str)))

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title("Word Cloud of Humidity in air quality Dataset")
plt.savefig('wordcloud.png')
plt.show()
```


Word Cloud of Humidity in air quality Dataset
high
low
nan
medium