```
import pandas as pd
d=pd.read_csv('/content/lungcancer.csv')
print(d)
```

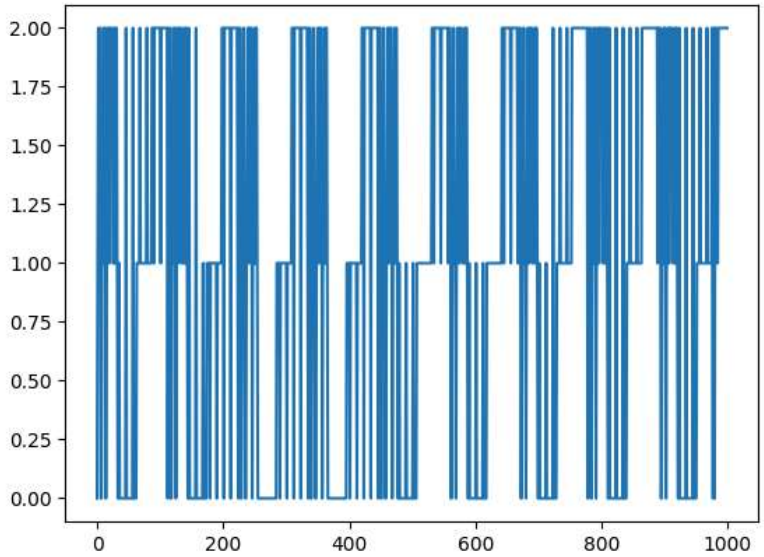
```
      index Patient Id  Age  Gender  Air Pollution  Alcohol use  Dust Allergy  \
0         0         P1   33       1              2            4             5
1         1        P10   17       1              3            1             5
2         2       P100   35       1              4            5             6
3         3      P1000   37       1              7            7             7
4         4       P101   46       1              6            8             7
..      ...        ...  ...     ...            ...          ...           ...
995     995       P995   44       1              6            7             7
996     996       P996   37       2              6            8             7
997     997       P997   25       2              4            5             6
998     998       P998   18       2              6            8             7
999     999       P999   47       1              6            5             6

     OccuPational Hazards  Genetic Risk  chronic Lung Disease  ...  Fatigue  \
0                       4             3                     2  ...        3
1                       3             4                     2  ...        1
2                       5             5                     4  ...        8
3                       7             6                     7  ...        4
4                       7             7                     6  ...        3
..                    ...           ...                   ...  ...      ...
995                     7             7                     6  ...        5
996                     7             7                     6  ...        9
997                     5             5                     4  ...        8
998                     7             7                     6  ...        3
999                     5             5                     4  ...        8

     Weight Loss  Shortness of Breath  Wheezing  Swal0ing Difficulty  \
0              4                    2         2                    3
1              3                    7         8                    6
2              7                    9         2                    1
3              2                    3         1                    4
4              2                    4         1                    4
..           ...                  ...       ...                  ...
995            3                    2         7                    8
996            6                    5         7                    2
997            7                    9         2                    1
998            2                    4         1                    4
999            7                    9         2                    1

     Clubbing of Finger Nails  Frequent Cold  Dry Cough  Snoring  Level
0                           1              2          3        4      0
1                           2              1          7        2      1
2                           4              6          7        2      2
3                           5              6          7        5      2
4                           2              4          2        3      2
..                        ...            ...        ...      ...    ...
995                         2              4          5        3      2
996                         4              3          1        4      2
997                         4              6          7        2      2
998                         2              4          2        3      2
999                         4              6          7        2      2

[1000 rows x 26 columns]
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
from matplotlib import pyplot as pt
x=d['index']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0fca2c20>]
```

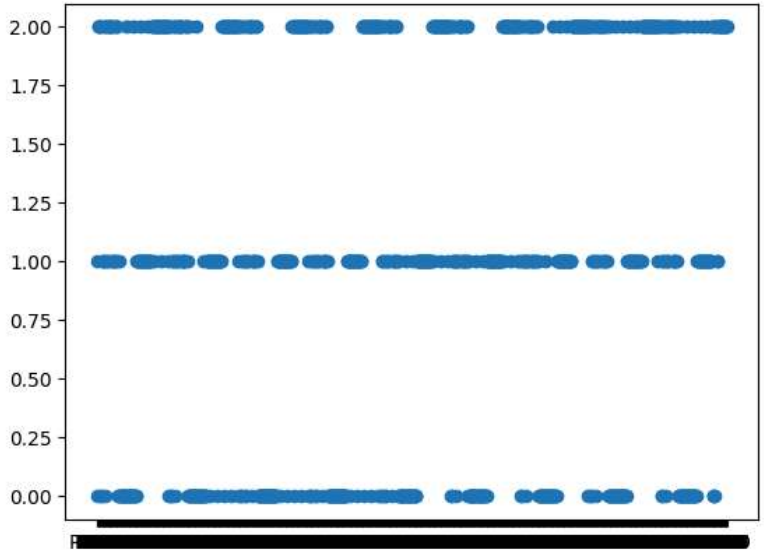



```
pt.scatter(x,y)
```

```
from matplotlib import pyplot as pt
x=d['Patient Id']
y=d['Level']
pt.plot(x,y)
```

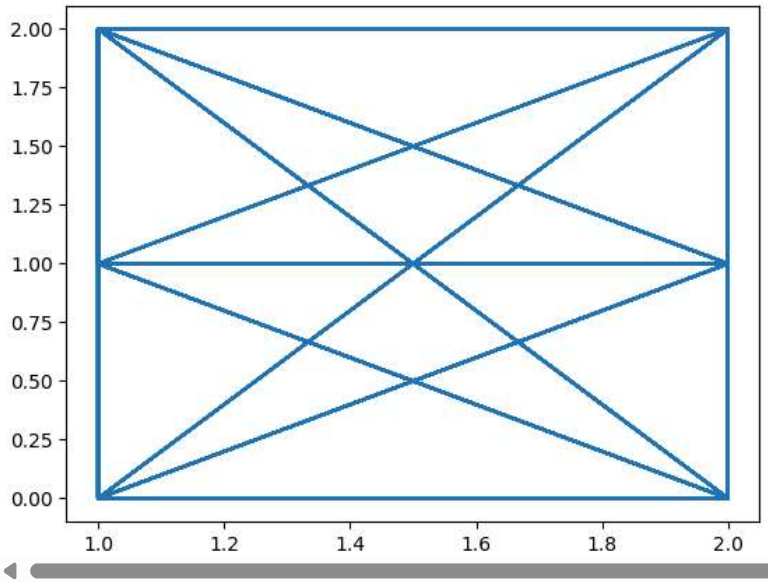
```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c245633d0>
```



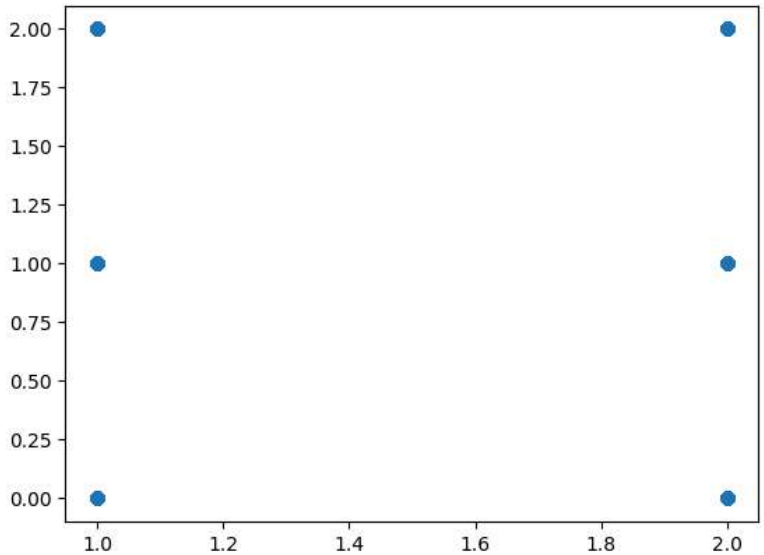

```
from matplotlib import pyplot as pt
x=d['Gender']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c14fe2230>]
```




```
pt.scatter(x,y)
```

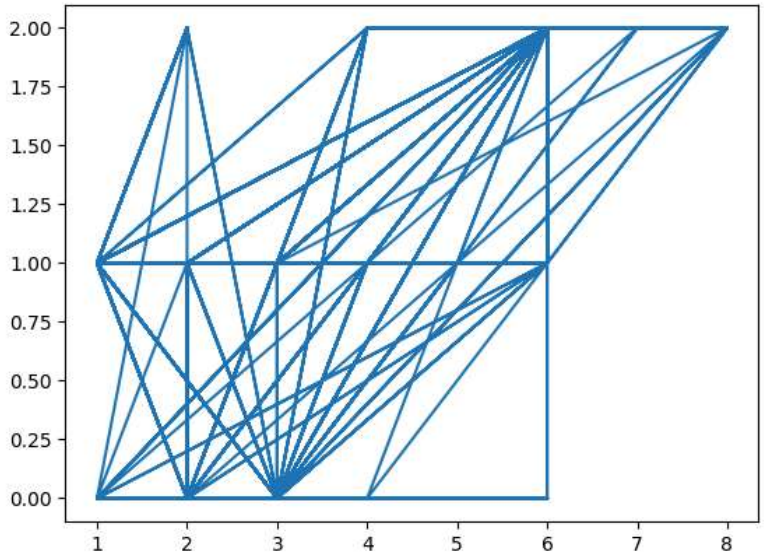
```
<matplotlib.collections.PathCollection at 0x798c0cc2cfd0>
```




Start coding or generate with AI.

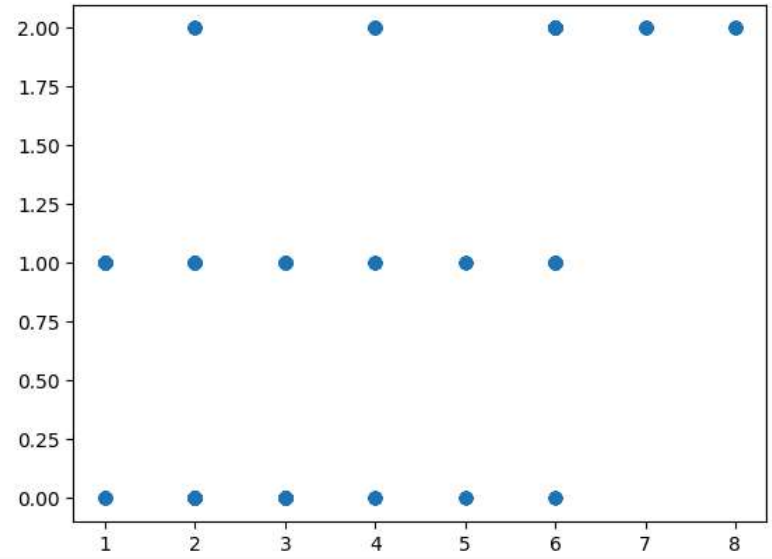
```
from matplotlib import pyplot as pt
x=d['Air Pollution']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0cab4a90>]
```



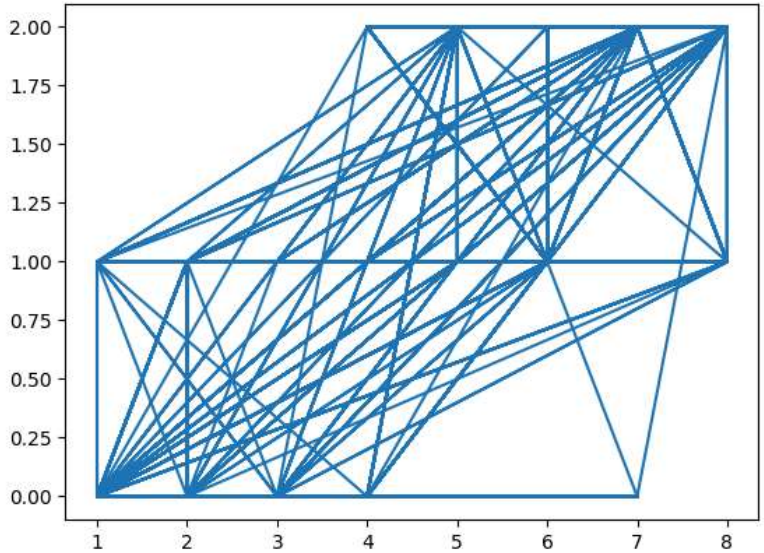
```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0cb20a60>
```



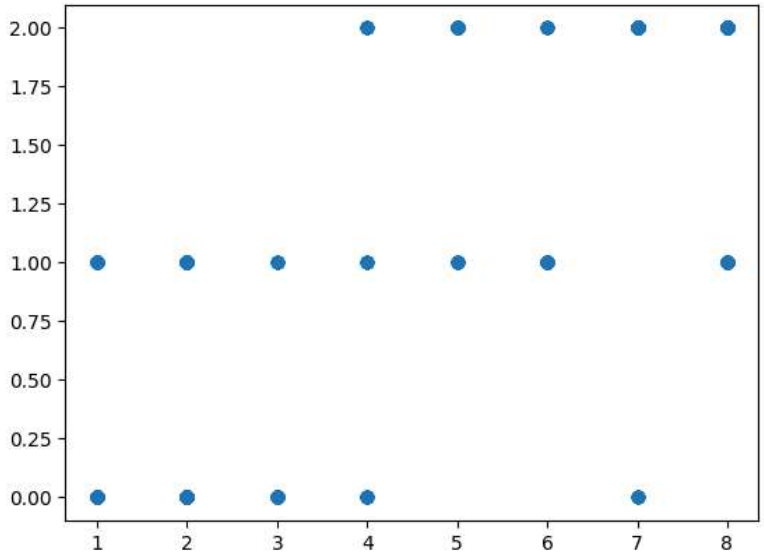

```
from matplotlib import pyplot as pt
x=d['Alcohol use']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c99ea10>]
```




```
pt.scatter(x,y)
```

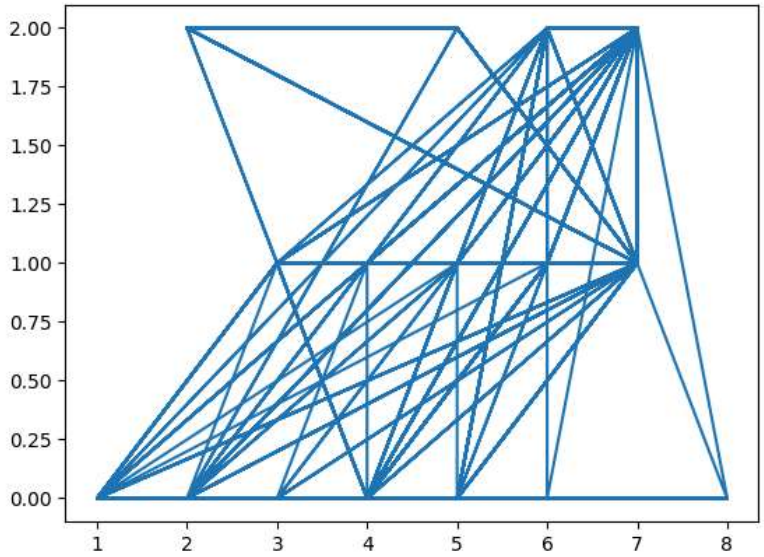
```
<matplotlib.collections.PathCollection at 0x798c0ca178e0>
```




```
from matplotlib import pyplot as pt
x=d['Dust Allergy']
```

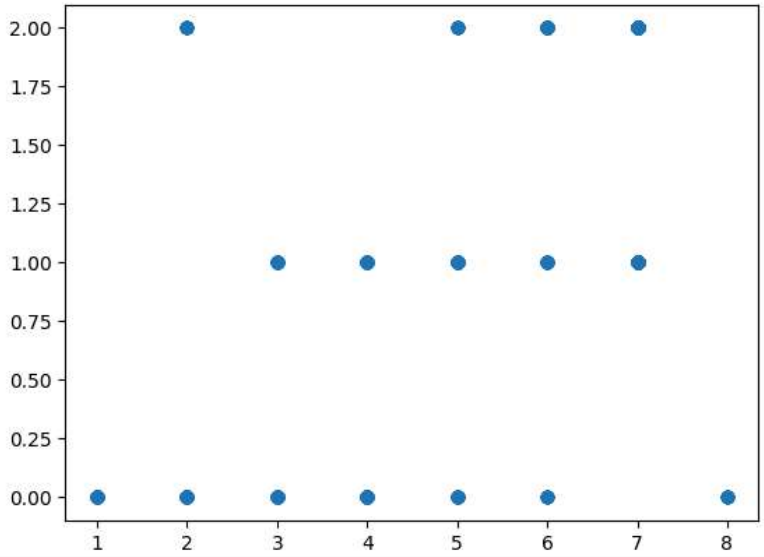
```
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c8b8b20>]
```




```
pt.scatter(x,y)
```

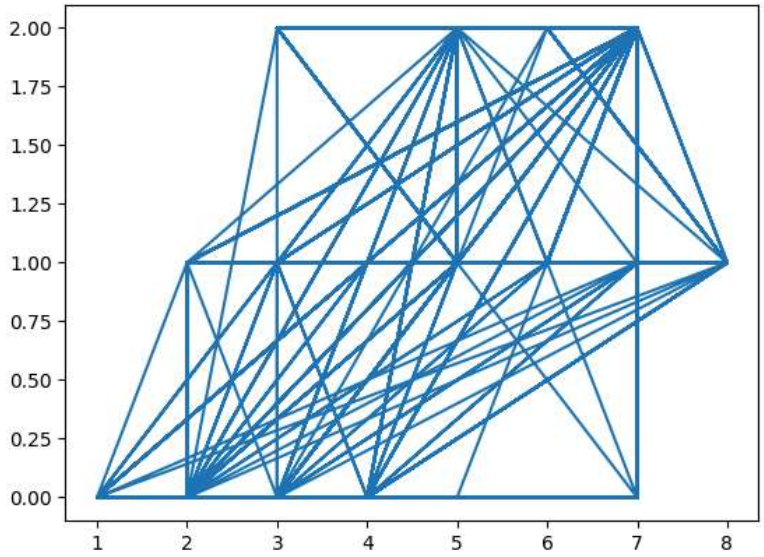
```
<matplotlib.collections.PathCollection at 0x798c0c935330>
```




```
from matplotlib import pyplot as pt
x=d['OccuPational Hazards']
```

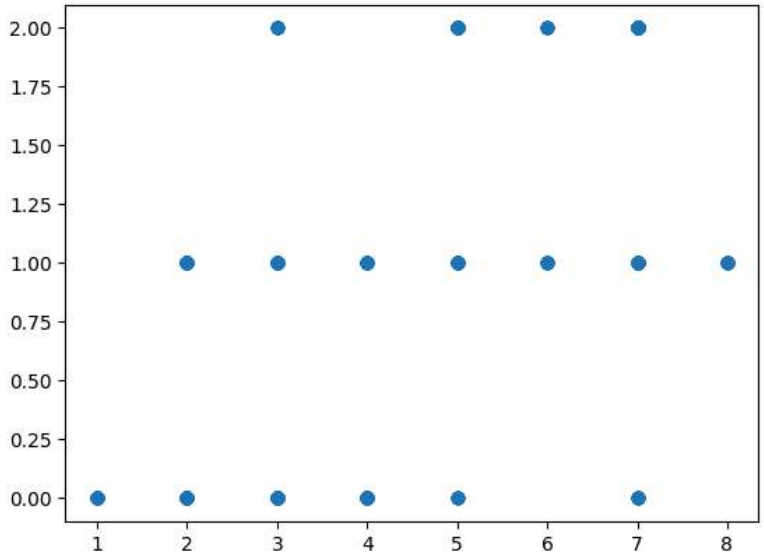
```
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c79eb30>]
```



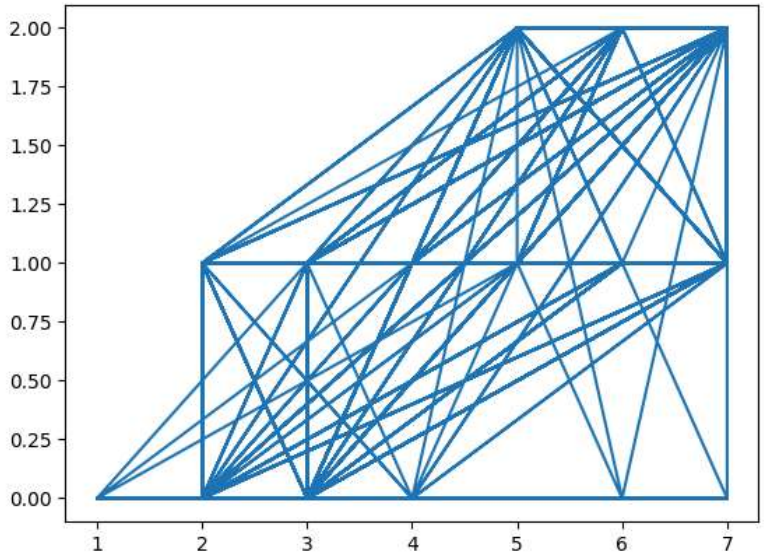

```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0c81b160>
```



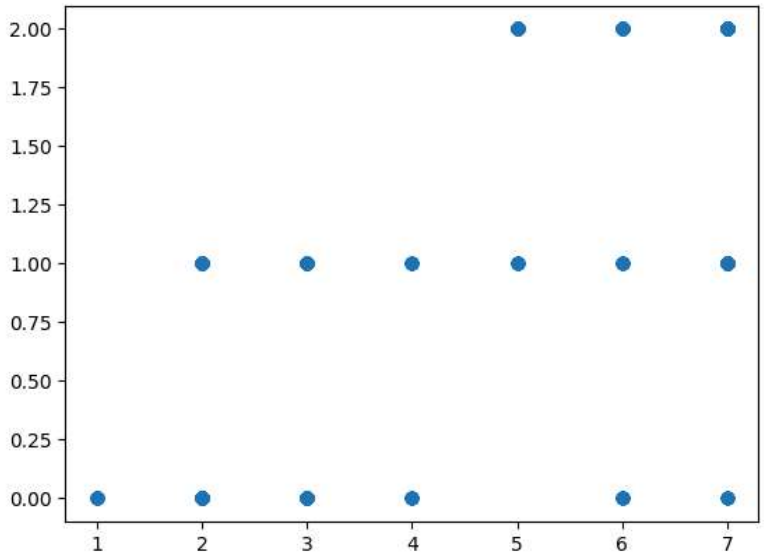
```
from matplotlib import pyplot as pt
x=d['Genetic Risk']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c6bc940>]
```



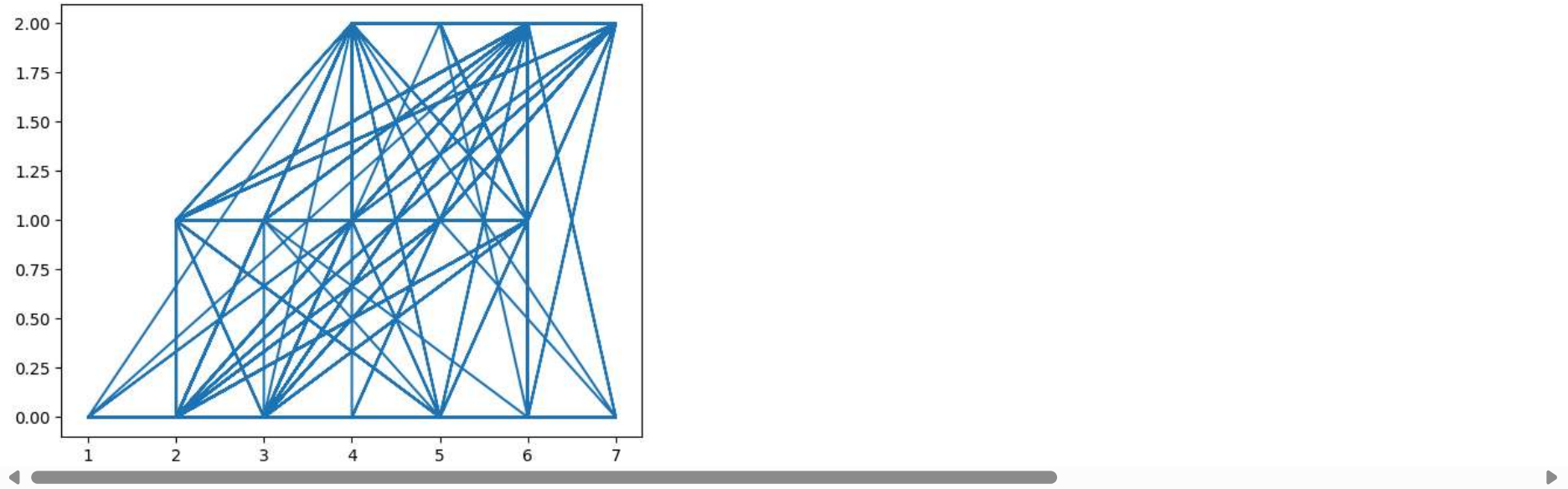

```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0c53cc70>
```



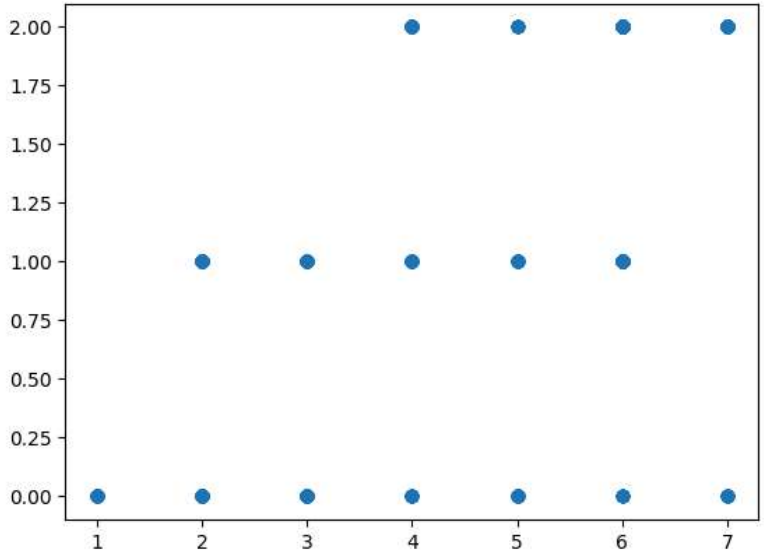
```
from matplotlib import pyplot as pt
x=d['chronic Lung Disease']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c6e5ea0>]
```



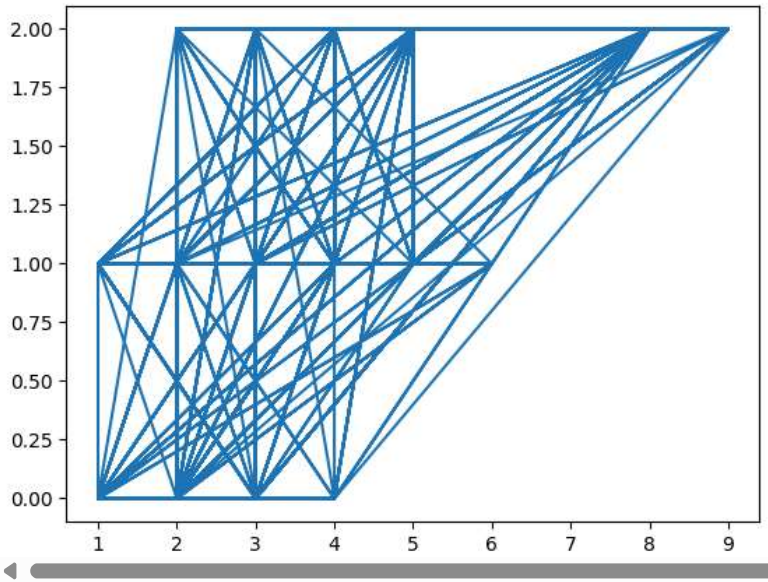

```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0c635e10>
```



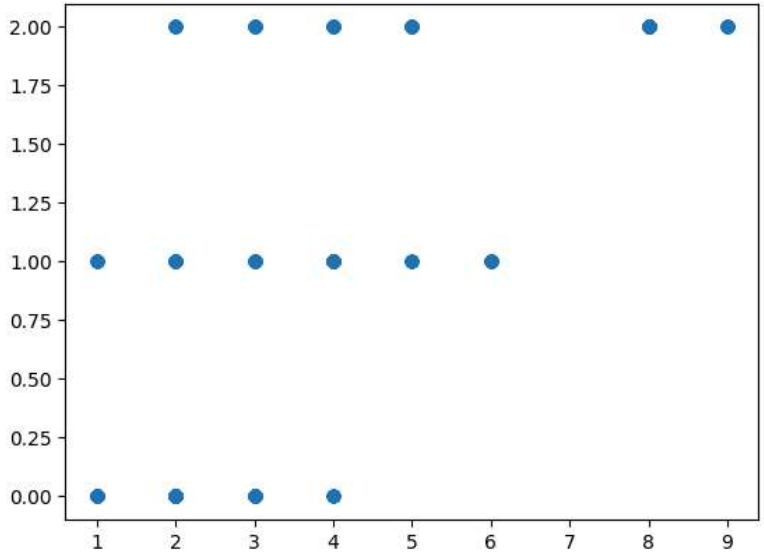

```
from matplotlib import pyplot as pt
x=d['Fatigue']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c4b2a70>]
```



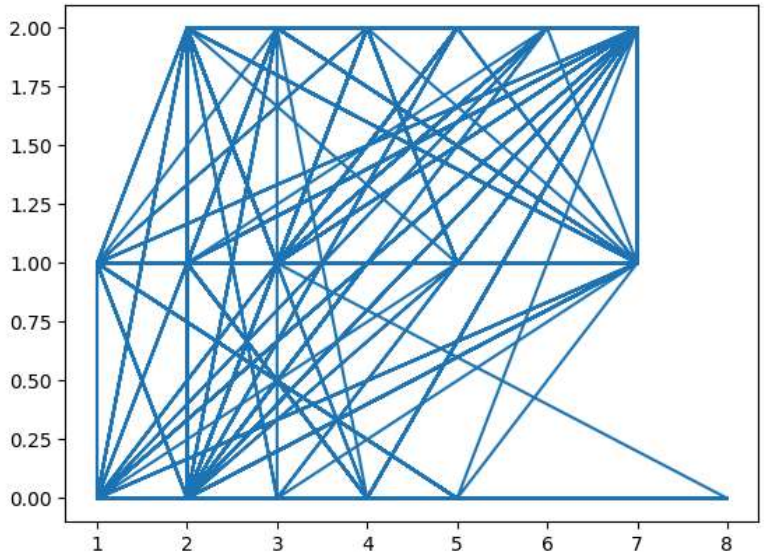

```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0c3542b0>
```



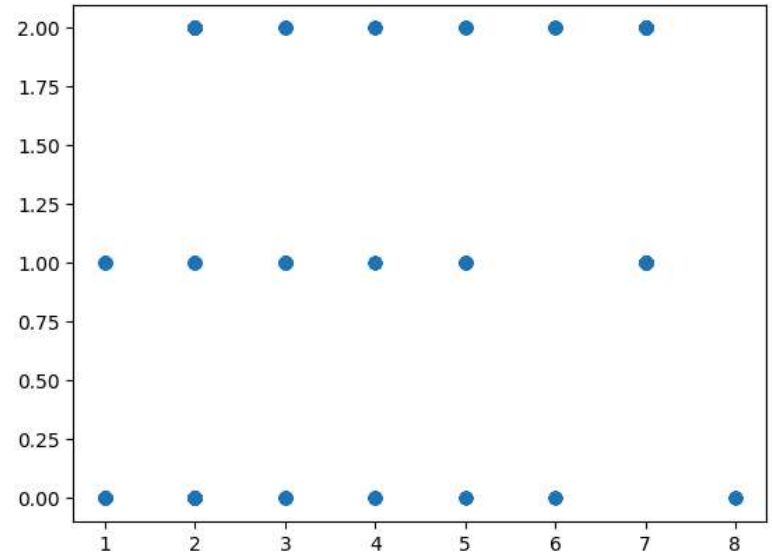

```
from matplotlib import pyplot as pt
x=d['Weight Loss']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c3c57b0>]
```



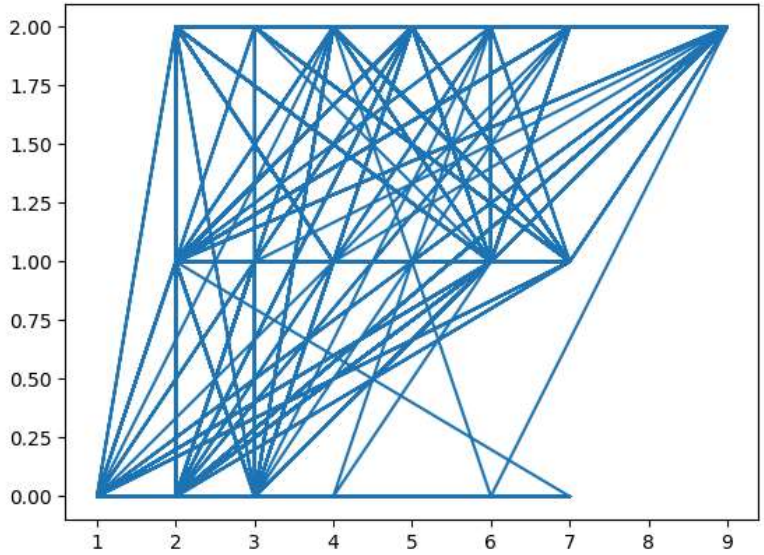
```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0c3c5930>
```



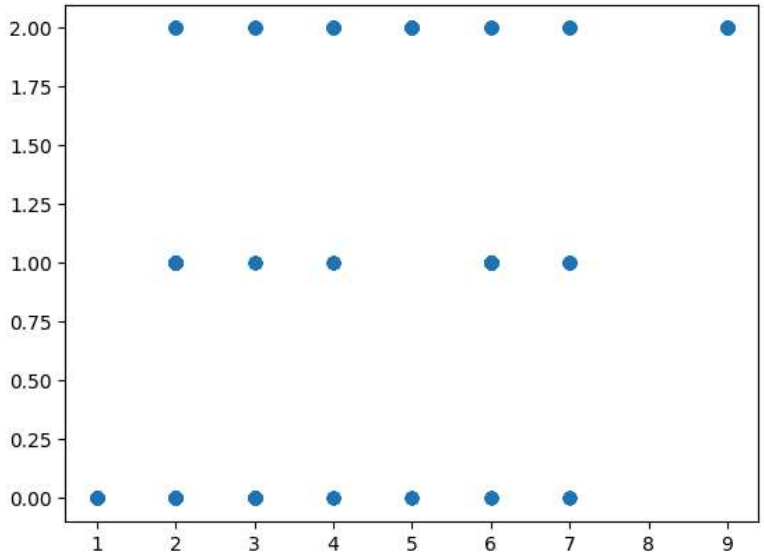

```
from matplotlib import pyplot as pt
x=d['Shortness of Breath']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c2e18a0>]
```




```
pt.scatter(x,y)
```

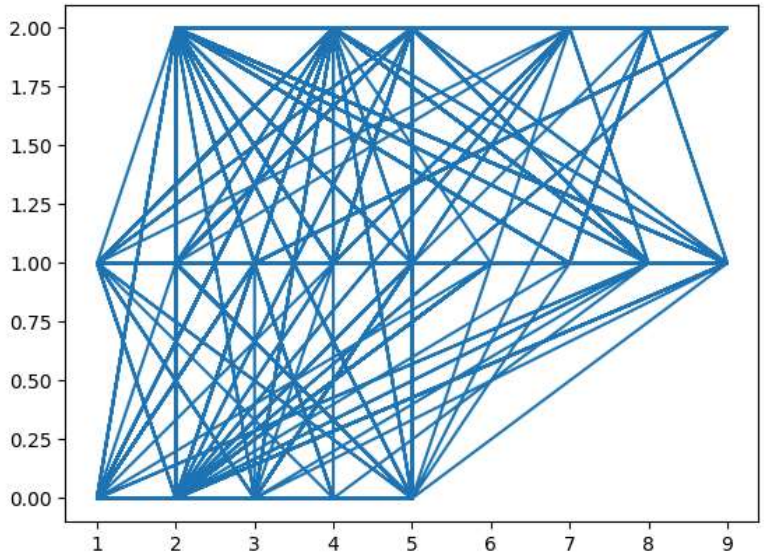
```
<matplotlib.collections.PathCollection at 0x798c0c161cc0>
```




```
from matplotlib import pyplot as pt
x=d['Clubbing of Finger Nails']
```

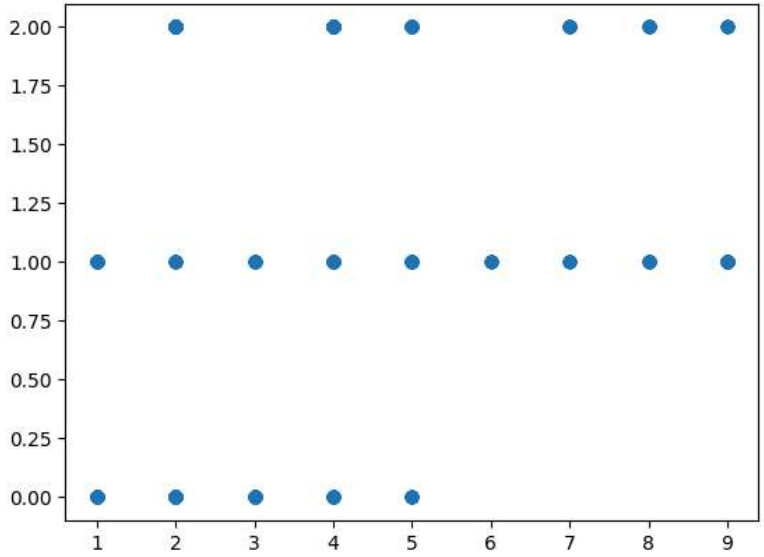
```
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c1dba60>]
```



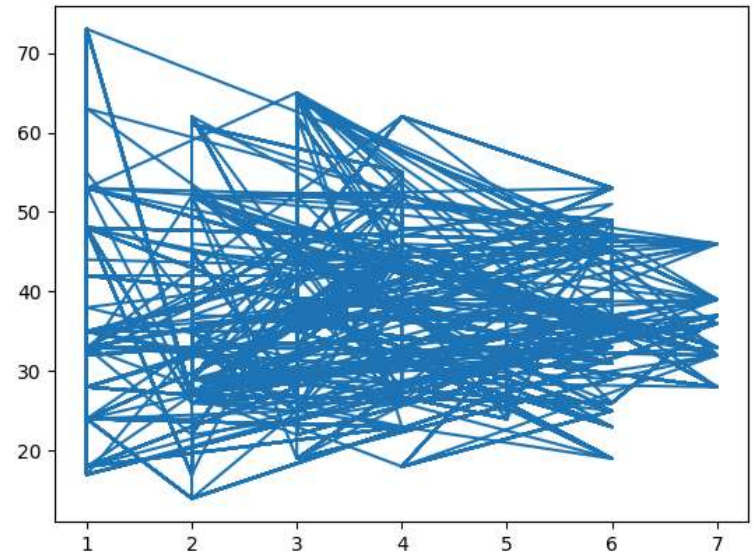

```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c0c0847c0>
```



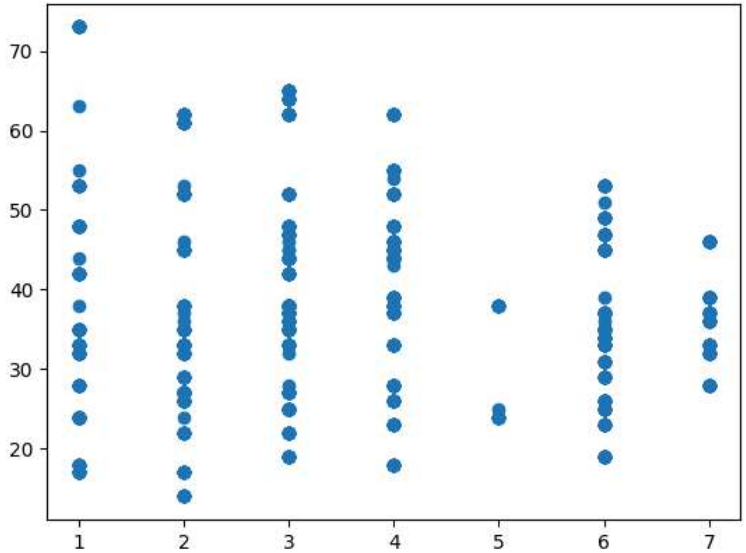
```
from matplotlib import pyplot as pt
x=d['Frequent Cold']
y=d['Age']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c0c0fe140>]
```



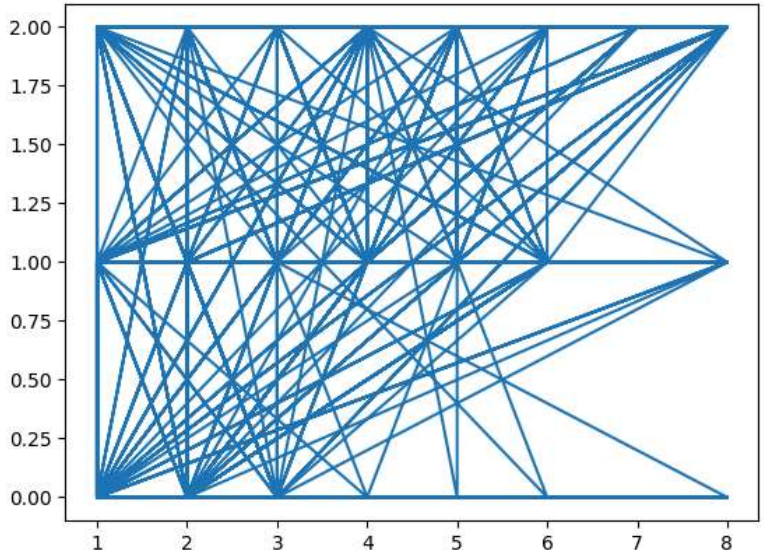

```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c07f3d840>
```



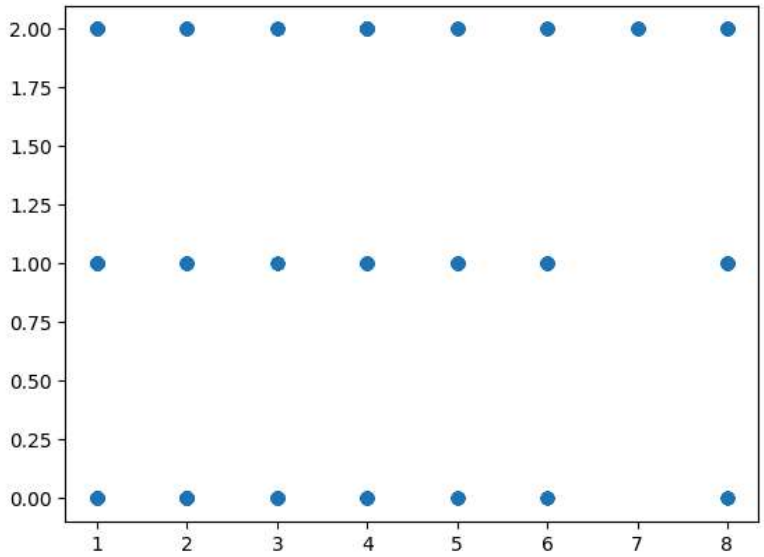

```
from matplotlib import pyplot as pt
x=d['Swal0ing Difficulty']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c07fb8d30>]
```



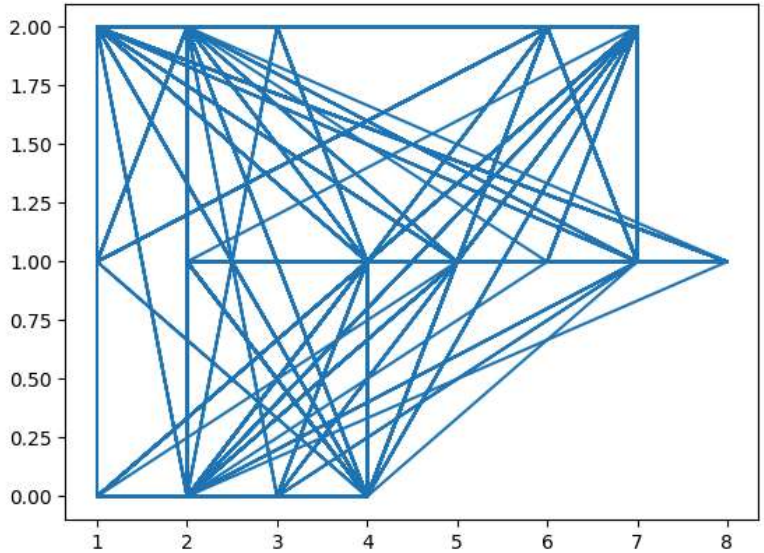
```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c07e2dc30>
```



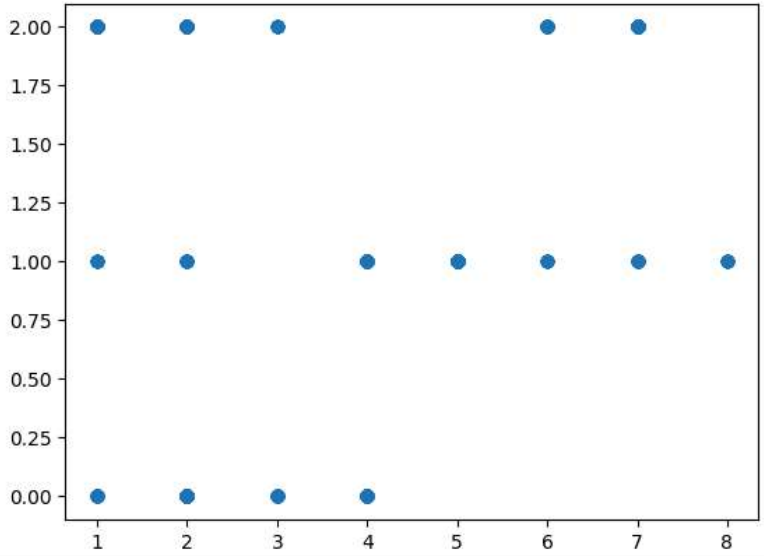

```
from matplotlib  import pyplot as pt
x=d['Wheezing']
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c07eaea40>]
```




```
pt.scatter(x,y)
```

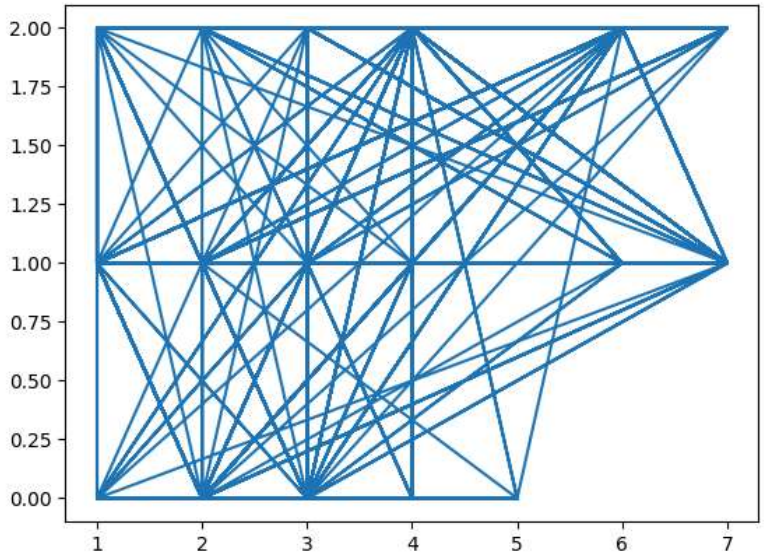
```
<matplotlib.collections.PathCollection at 0x798c07d2fa00>
```




```
from matplotlib import pyplot as pt
x=d['Frequent Cold']
```

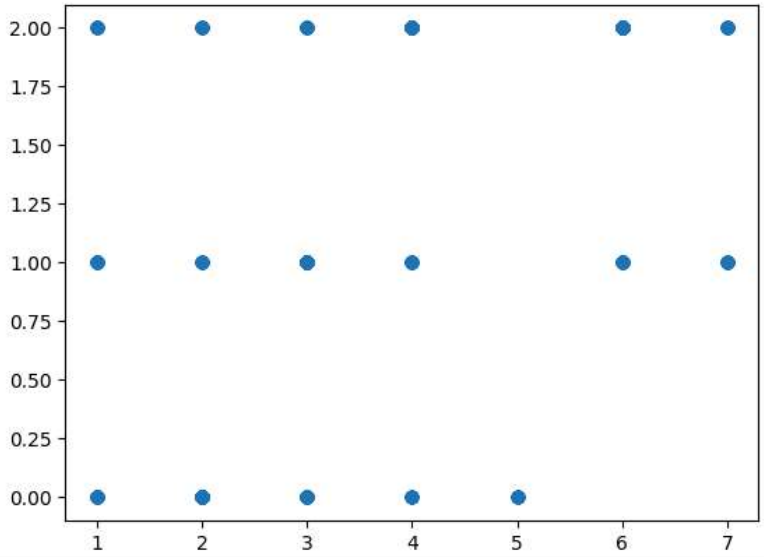
```
y=d['Level']
pt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x798c07dd0df0>]
```




```
pt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x798c07c3cca0>
```



```
print(d.isnull())
```

```
      index  Patient Id    Age  Gender  Air Pollution  Alcohol use  \
0    False       False  False   False          False        False
1    False       False  False   False          False        False
2    False       False  False   False          False        False
3    False       False  False   False          False        False
4    False       False  False   False          False        False
..     ...         ...    ...     ...            ...          ...
995  False       False  False   False          False        False
996  False       False  False   False          False        False
997  False       False  False   False          False        False
998  False       False  False   False          False        False
999  False       False  False   False          False        False

     Dust Allergy  OccuPational Hazards  Genetic Risk  chronic Lung Disease  \
0           False                 False         False                 False
1           False                 False         False                 False
2           False                 False         False                 False
3           False                 False         False                 False
4           False                 False         False                 False
..            ...                   ...           ...                   ...
995         False                 False         False                 False
996         False                 False         False                 False
997         False                 False         False                 False
998         False                 False         False                 False
999         False                 False         False                 False

     ...  Fatigue  Weight Loss  Shortness of Breath  Wheezing  \
0    ...    False        False                False     False
1    ...    False        False                False     False
2    ...    False        False                False     False
3    ...    False        False                False     False
4    ...    False        False                False     False
..   ...      ...          ...                  ...       ...
995  ...    False        False                False     False
996  ...    False        False                False     False
997  ...    False        False                False     False
998  ...    False        False                False     False
999  ...    False        False                False     False

     Swal0ing Difficulty  Clubbing of Finger Nails  Frequent Cold  Dry Cough  \
0                  False                     False          False      False
1                  False                     False          False      False
2                  False                     False          False      False
3                  False                     False          False      False
4                  False                     False          False      False
..                   ...                       ...            ...        ...
995                False                     False          False      False
996                False                     False          False      False
997                False                     False          False      False
998                False                     False          False      False
999                False                     False          False      False

     Snoring  Level
0      False  False
1      False  False
2      False  False
3      False  False
4      False  False
```

```
print(d.describe())
print(d)
```

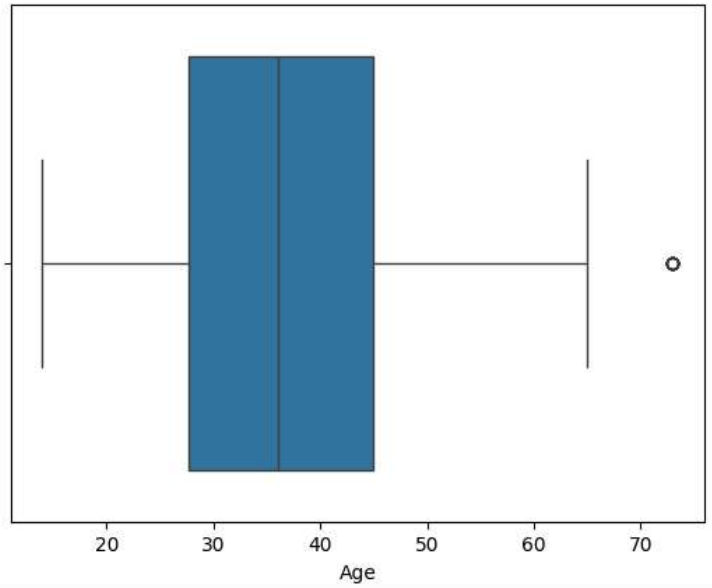
```
             index          Age       Gender  Air Pollution  Alcohol use  \
count  1000.000000  1000.000000  1000.000000      1000.0000  1000.000000
mean    499.500000    37.174000     1.402000         3.8400     4.563000
std     288.819436    12.005493     0.490547         2.0304     2.620477
min       0.000000    14.000000     1.000000         1.0000     1.000000
25%     249.750000    27.750000     1.000000         2.0000     2.000000
50%     499.500000    36.000000     1.000000         3.0000     5.000000
75%     749.250000    45.000000     2.000000         6.0000     7.000000
max     999.000000    73.000000     2.000000         8.0000     8.000000

       Dust Allergy  OccuPational Hazards  Genetic Risk  chronic Lung Disease  \
count   1000.000000           1000.000000   1000.000000           1000.000000
mean       5.165000              4.840000      4.580000              4.380000
std        1.980833              2.107805      2.126999              1.848518
min        1.000000              1.000000      1.000000              1.000000
25%        4.000000              3.000000      2.000000              3.000000
50%        6.000000              5.000000      5.000000              4.000000
75%        7.000000              7.000000      7.000000              6.000000
max        8.000000              8.000000      7.000000              7.000000

       Balanced Diet  ...      Fatigue  Weight Loss  Shortness of Breath  \
count    1000.000000  ...  1000.000000  1000.000000          1000.000000
mean        4.491000  ...     3.856000     3.855000             4.240000
std         2.135528  ...     2.244616     2.206546             2.285087
min         1.000000  ...     1.000000     1.000000             1.000000
25%         2.000000  ...     2.000000     2.000000             2.000000
50%         4.000000  ...     3.000000     3.000000             4.000000
75%         7.000000  ...     5.000000     6.000000             6.000000
max         7.000000  ...     9.000000     8.000000             9.000000

          Wheezing  Swal0ing Difficulty  Clubbing of Finger Nails  \
count  1000.000000          1000.000000               1000.000000
mean      3.777000             3.746000                  3.923000
std       2.041921             2.270383                  2.388048
min       1.000000             1.000000                  1.000000
25%       2.000000             2.000000                  2.000000
50%       4.000000             4.000000                  4.000000
75%       5.000000             5.000000                  5.000000
max       8.000000             8.000000                  9.000000

       Frequent Cold    Dry Cough      Snoring        Level
count    1000.000000  1000.000000  1000.000000  1000.000000
mean        3.536000     3.853000     2.926000     1.062000
std         1.832502     2.039007     1.474686     0.815365
min         1.000000     1.000000     1.000000     0.000000
25%         2.000000     2.000000     2.000000     0.000000
50%         3.000000     4.000000     3.000000     1.000000
75%         5.000000     6.000000     4.000000     2.000000
max         7.000000     7.000000     7.000000     2.000000

[8 rows x 25 columns]
     index Patient Id  Age  Gender  Air Pollution  Alcohol use  Dust Allergy  \
0        0         P1   33       1              2            4             5
1        1        P10   17       1              3            1             5
2        2       P100   35       1              4            5             6
3        3      P1000   37       1              7            7             7
4        4       P101   46       1              6            8             7
```

```
d.columns
```

```
Index(['index', 'Patient Id', 'Age', 'Gender', 'Air Pollution', 'Alcohol use',
       'Dust Allergy', 'OccuPational Hazards', 'Genetic Risk',
       'chronic Lung Disease', 'Balanced Diet', 'Obesity', 'Smoking',
       'Passive Smoker', 'Chest Pain', 'Coughing of Blood', 'Fatigue',
       'Weight Loss', 'Shortness of Breath', 'Wheezing', 'Swal0ing Difficulty',
       'Clubbing of Finger Nails', 'Frequent Cold', 'Dry Cough', 'Snoring',
       'Level'],
      dtype='object')
```

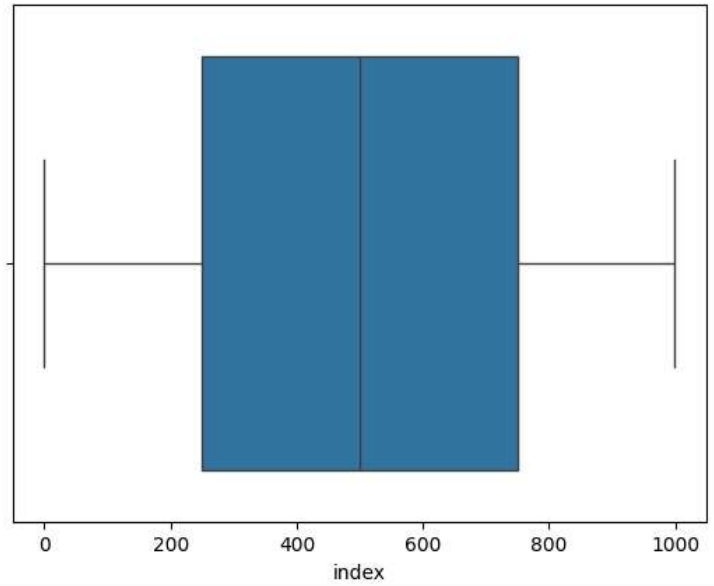
```
import seaborn as sns
sns.boxplot(data=d,x='Age')
```

```
<Axes: xlabel='Age'>
```



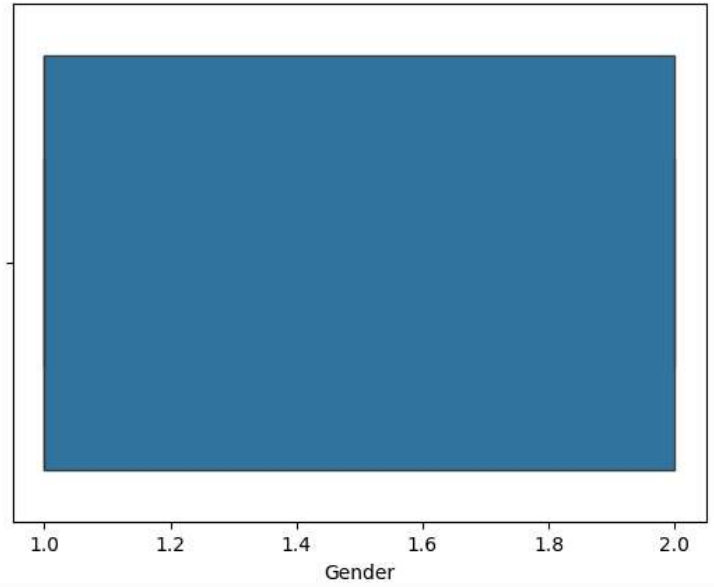

```
import seaborn as sns
sns.boxplot(data=d,x='index')
```

```
<Axes: xlabel='index'>
```



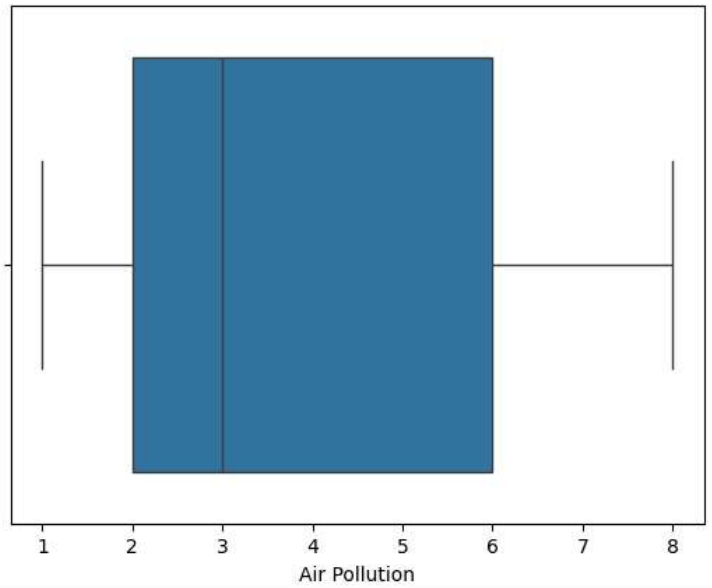

```
import seaborn as sns
sns.boxplot(data=d,x='Gender')
```

```
<Axes: xlabel='Gender'>
```



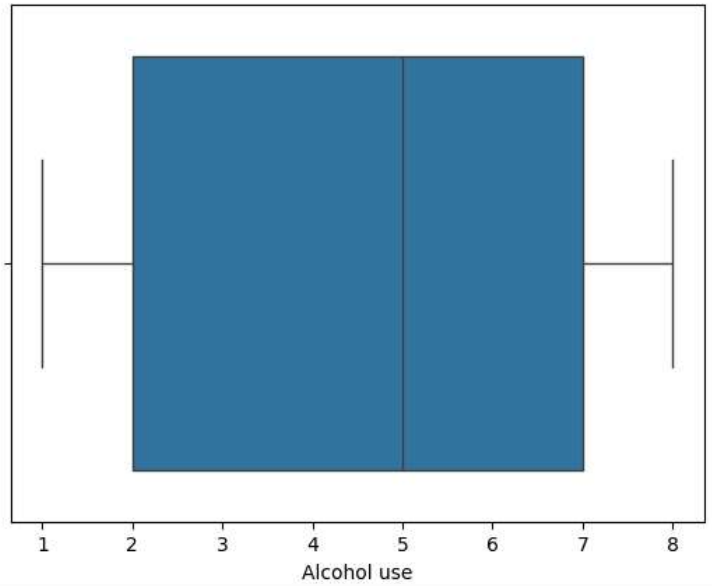
```
import seaborn as sns
sns.boxplot(data=d,x='Air Pollution')
```

```
<Axes: xlabel='Air Pollution'>
```



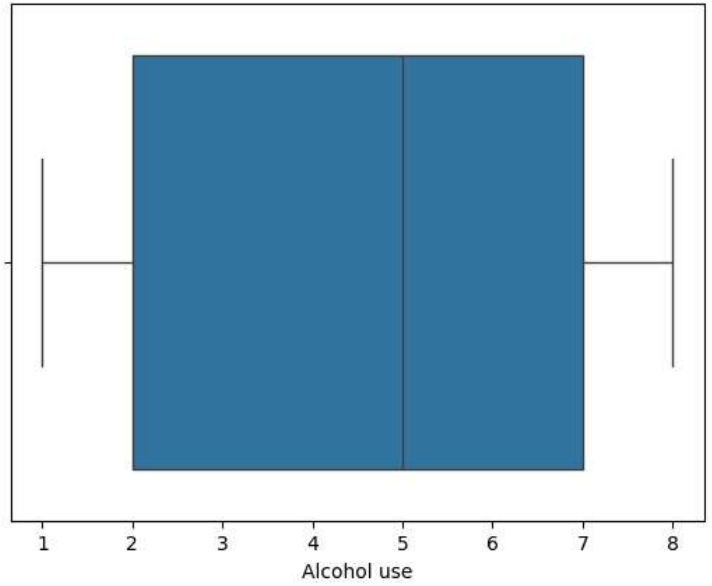

```
import seaborn as sns
sns.boxplot(data=d,x='Alcohol use')
```

<Axes: xlabel='Alcohol use'>



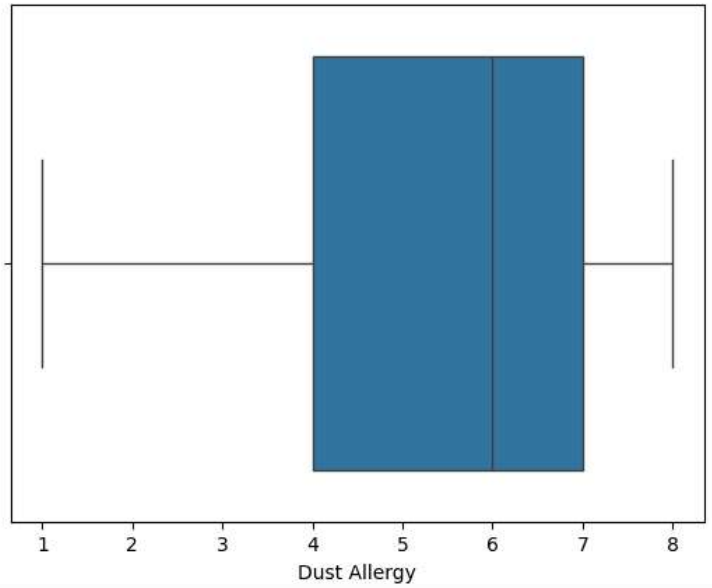

```
import seaborn as sns
sns.boxplot(data=d,x='Alcohol use')
```

<Axes: xlabel='Alcohol use'>



```
import seaborn as sns
sns.boxplot(data=d,x='Dust Allergy')
```

```
<Axes: xlabel='Dust Allergy'>
```




```
import seaborn as sns
sns.boxplot(data=d,x='OccuPational Hazards')
```

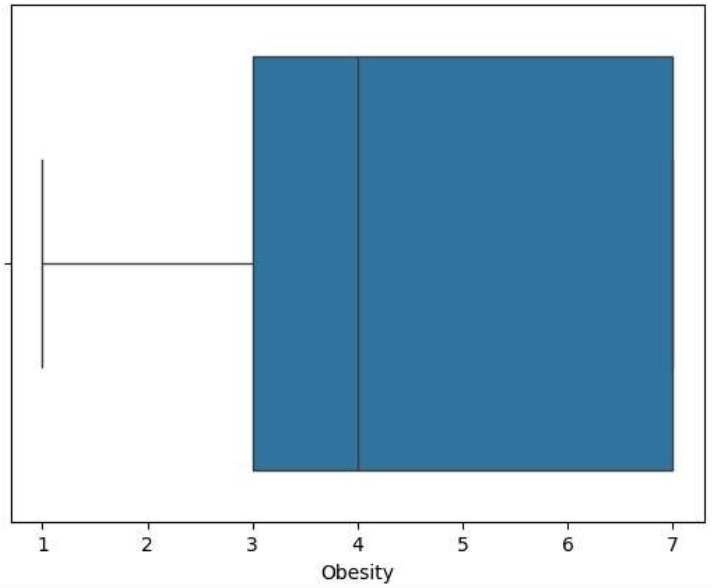
```
import seaborn as sns
sns.boxplot(data=d,x='Genetic Risk')
```

```
import seaborn as sns
sns.boxplot(data=d,x='chronic Lung Disease')
```

```
import seaborn as sns
sns.boxplot(data=d,x='Balanced Diet')
```

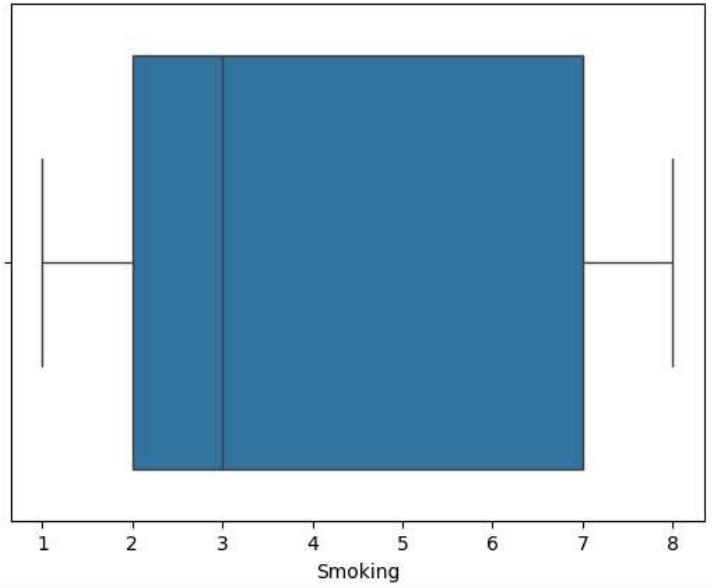
```
import seaborn as sns
sns.boxplot(data=d,x='Obesity')
```

```
<Axes: xlabel='Obesity'>
```



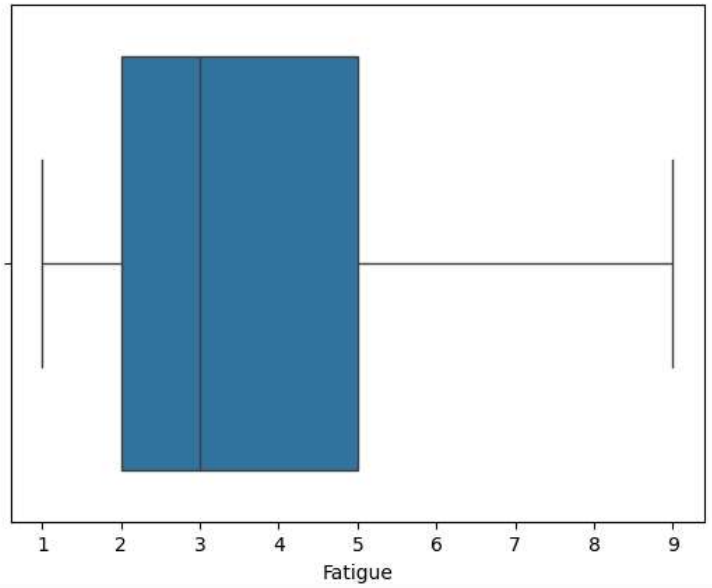

```
import seaborn as sns
sns.boxplot(data=d,x='Smoking')
```

```
<Axes: xlabel='Smoking'>
```



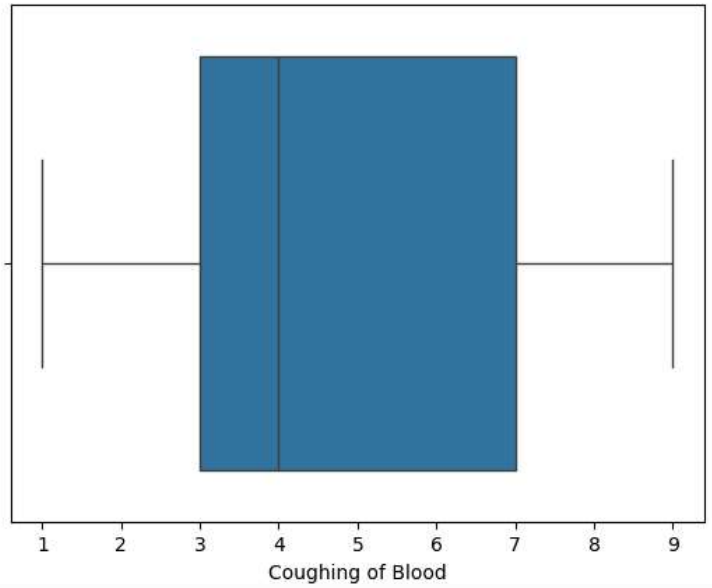
```
import seaborn as sns
sns.boxplot(data=d,x='Fatigue')
```

```
<Axes: xlabel='Fatigue'>
```



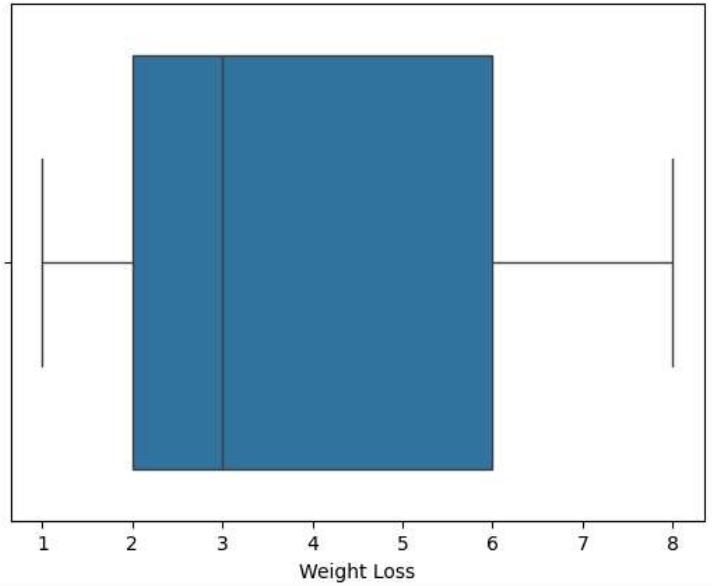

```
import seaborn as sns
sns.boxplot(data=d,x='Coughing of Blood')
```

```
<Axes: xlabel='Coughing of Blood'>
```



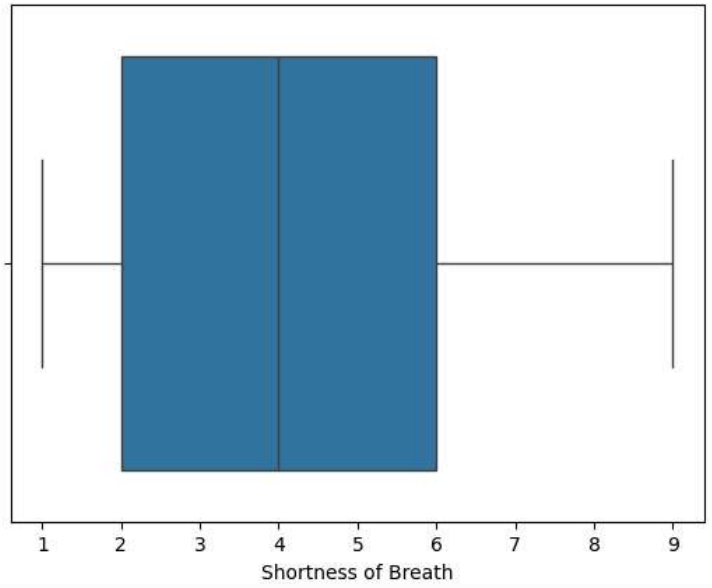

```
import seaborn as sns
sns.boxplot(data=d,x='Weight Loss')
```

```
<Axes: xlabel='Weight Loss'>
```



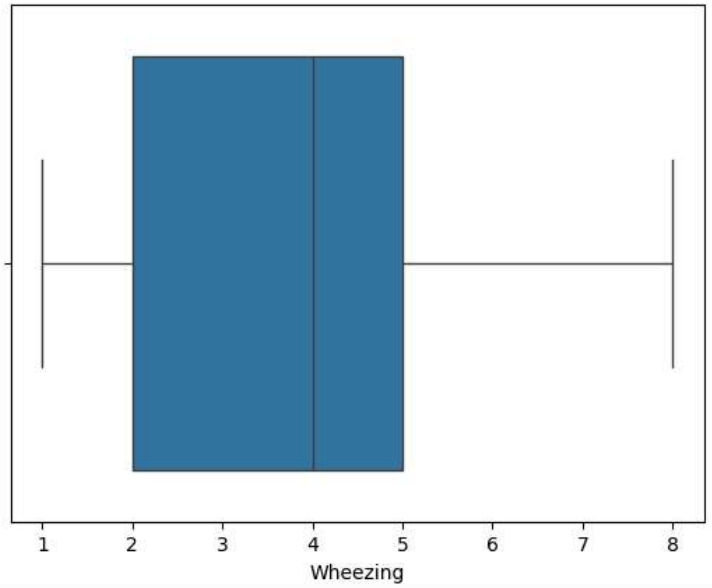
```
import seaborn as sns
sns.boxplot(data=d,x='Shortness of Breath')
```

```
<Axes: xlabel='Shortness of Breath'>
```



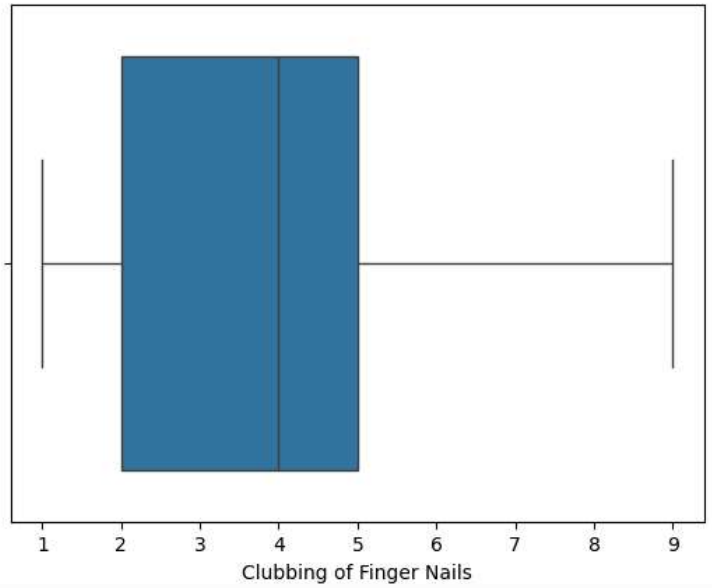

```
import seaborn as sns
sns.boxplot(data=d,x='Wheezing')
```

<Axes: xlabel='Wheezing'>



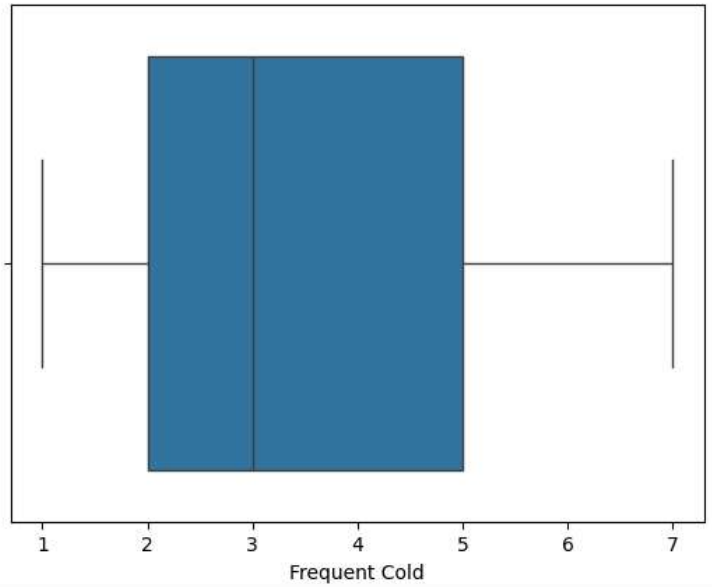

```
import seaborn as sns
sns.boxplot(data=d,x='Clubbing of Finger Nails')
```

<Axes: xlabel='Clubbing of Finger Nails'>



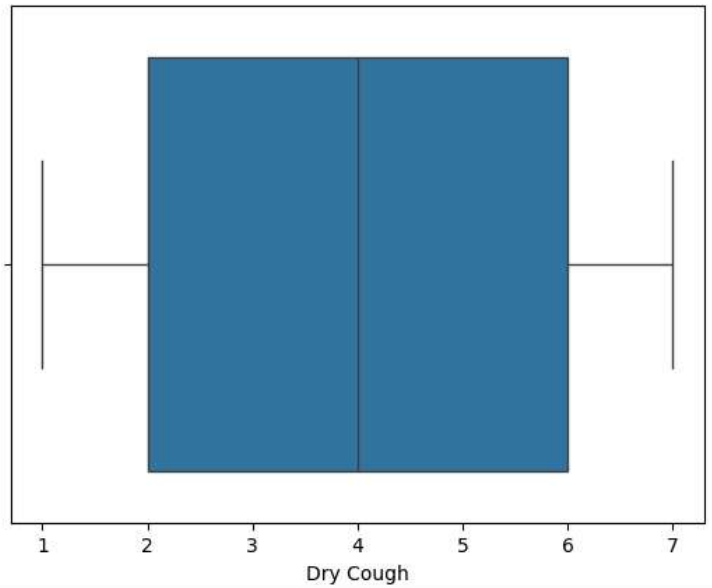

```
import seaborn as sns
sns.boxplot(data=d,x='Frequent Cold')
```

<Axes: xlabel='Frequent Cold'>



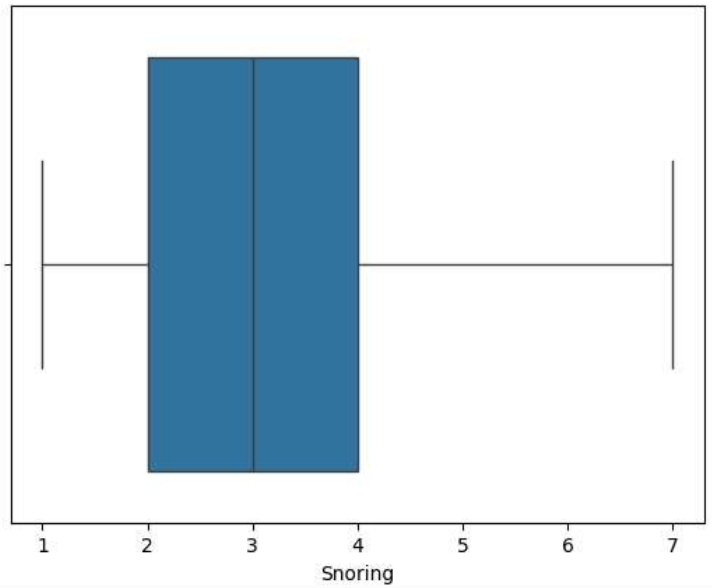

```
import seaborn as sns
sns.boxplot(data=d,x='Dry Cough')
```

<Axes: xlabel='Dry Cough'>



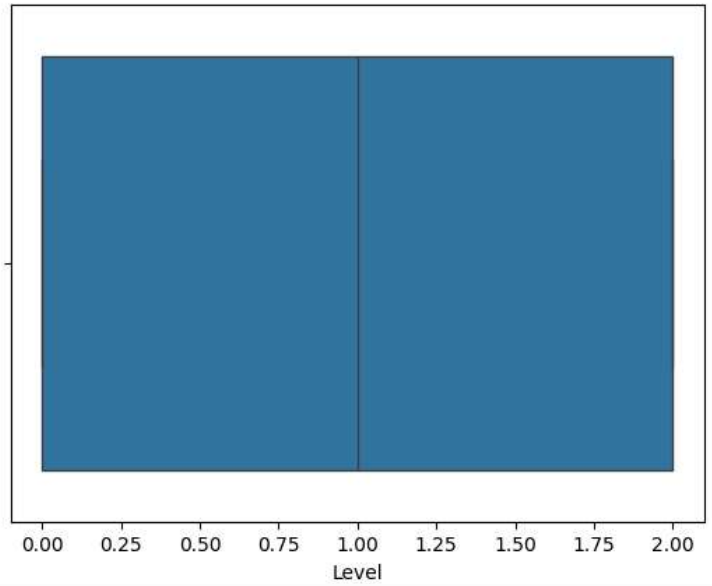
```
import seaborn as sns
sns.boxplot(data=d,x='Snoring')
```

```
<Axes: xlabel='Snoring'>
```




```
import seaborn as sns
sns.boxplot(data=d,x='Level')
```

<Axes: xlabel='Level'>




```
print(d.dtypes)
```

```
index                      int64
Patient Id                object
Age                        int64
Gender                     int64
Air Pollution              int64
Alcohol use                int64
Dust Allergy               int64
OccuPational Hazards       int64
Genetic Risk               int64
chronic Lung Disease       int64
Balanced Diet              int64
Obesity                    int64
Smoking                    int64
Passive Smoker             int64
Chest Pain                 int64
Coughing of Blood          int64
Fatigue                    int64
Weight Loss                int64
Shortness of Breath        int64
Wheezing                   int64
Swal0ing Difficulty        int64
Clubbing of Finger Nails   int64
```

```
Frequent Cold               int64
Dry Cough                   int64
Snoring                     int64
Level                       int64
dtype: object
```

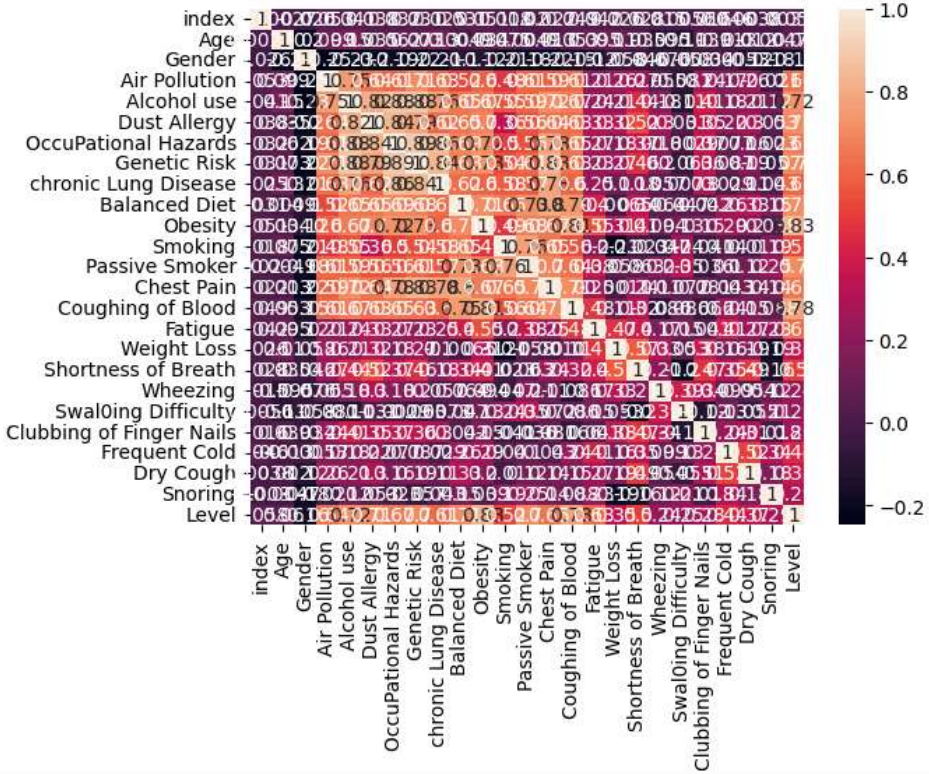
```
import numpy as np

d_numeric = d.select_dtypes(include=[np.number])
corr_matrix = d_numeric.corr()
print(corr_matrix)
```

```
                            index       Age    Gender  Air Pollution  \
index                    1.000000  0.002674 -0.025739       0.053307
Age                      0.002674  1.000000 -0.202086       0.099494
Gender                  -0.025739 -0.202086  1.000000      -0.246912
Air Pollution            0.053307  0.099494 -0.246912       1.000000
Alcohol use              0.041374  0.151742 -0.227636       0.747293
Dust Allergy             0.037960  0.035202 -0.204312       0.637503
OccuPational Hazards     0.032355  0.062177 -0.192343       0.608924
Genetic Risk             0.030725  0.073151 -0.222727       0.705276
chronic Lung Disease     0.025177  0.128952 -0.205061       0.626701
Balanced Diet            0.030743  0.004863 -0.099741       0.524873
Obesity                  0.050584  0.034337 -0.123813       0.601468
Smoking                  0.018407  0.075333 -0.206924       0.481902
Passive Smoker           0.019517  0.004908 -0.184826       0.606764
Chest Pain               0.022210  0.012864 -0.218426       0.585734
Coughing of Blood        0.049401  0.053006 -0.146505       0.607829
Fatigue                  0.042346  0.095059 -0.116467       0.211724
Weight Loss              0.026393  0.106946 -0.057993       0.258016
Shortness of Breath      0.027950  0.035329 -0.045972       0.269558
Wheezing                 0.015078 -0.095354 -0.076304       0.055368
Swal0ing Difficulty      0.005573 -0.105833 -0.058324      -0.080918
Clubbing of Finger Nails 0.015706  0.039258 -0.034219       0.241065
Frequent Cold            0.045687 -0.012706 -0.000526       0.174539
Dry Cough                0.003793  0.012128 -0.123001       0.261489
Snoring                 -0.002957 -0.004700 -0.181618      -0.021343
Level                    0.057758  0.060048 -0.164985       0.636038

                         Alcohol use  Dust Allergy  OccuPational Hazards  \
index                       0.041374      0.037960              0.032355
Age                         0.151742      0.035202              0.062177
Gender                     -0.227636     -0.204312             -0.192343
Air Pollution               0.747293      0.637503              0.608924
Alcohol use                 1.000000      0.818644              0.878786
Dust Allergy                0.818644      1.000000              0.835860
OccuPational Hazards        0.878786      0.835860              1.000000
Genetic Risk                0.877210      0.787904              0.893049
chronic Lung Disease        0.763576      0.619556              0.858284
Balanced Diet               0.653352      0.647197              0.691509
Obesity                     0.669312      0.700676              0.722191
Smoking                     0.547035      0.358691              0.497693
Passive Smoker              0.592576      0.560002              0.555311
Chest Pain                  0.717242      0.639983              0.775619
Coughing of Blood           0.667612      0.625291              0.645947
Fatigue                     0.237245      0.332472              0.267844
Weight Loss                 0.207851      0.321756              0.176226
Shortness of Breath         0.435785      0.518682              0.366482
```

```
Wheezing                    0.180817      0.304850             0.178925
Swal0ing Difficulty        -0.114073      0.031141            -0.002853
Clubbing of Finger Nails    0.414992      0.345714             0.366447
Frequent Cold               0.180778      0.219389             0.077166
Dry Cough                   0.211277      0.300195             0.159887
Snoring                     0.122694      0.052844             0.022916
Level                       0.718710      0.713839             0.673255

                          Genetic Risk  chronic Lung Disease  Balanced Diet  \
index                         0.030725              0.025177       0.030743
Age                           0.073151              0.128952       0.004863
Gender                        0.223727              0.205961       0.009741
```

```
import seaborn as sn
import matplotlib.pyplot as plt
sn.heatmap(corr_matrix,annot=True)
```

```
<Axes: >
```




```
import numpy as np
import pandas as pd

# Assuming 'd' is your DataFrame
d_numeric = d.select_dtypes(include=[np.number])
```

```
cov_matrix = d_numeric.cov()
print(cov_matrix)
```

```
                              index         Age    Gender  Air Pollution  \
index                     83416.666667    9.273273 -3.646647      31.260260
Age                           9.273273  144.131856 -1.190138       2.425265
Gender                       -3.646647   -1.190138  0.240637      -0.245926
Air Pollution                31.260260    2.425265 -0.245926       4.122523
Alcohol use                  31.313814    4.773812 -0.292619       3.976056
Dust Allergy                 21.717217    0.837127 -0.198529       2.563964
OccuPational Hazards         19.696697    1.573413 -0.198879       2.606006
Genetic Risk                 18.874875    1.867948 -0.232392       3.045846
chronic Lung Disease         13.441441    2.861742 -0.185946       2.352152
Balanced Diet                18.961461    0.124691 -0.104486       2.275836
Obesity                      31.044545    0.875966 -0.129059       2.594995
Smoking                      13.269269    2.257305 -0.253349       2.442122
Passive Smoker               13.031532    0.136206 -0.209600       2.848048
Chest Pain                   14.626627    0.352140 -0.244320       2.711792
Coughing of Blood            34.642142    1.545079 -0.174492       2.996436
Fatigue                      27.452452    2.561618 -0.128240       0.964925
Weight Loss                  16.820320    2.833063 -0.062773       1.155956
Shortness of Breath          18.446446    0.969209 -0.051532       1.250651
Wheezing                      8.892392   -2.337536 -0.076430       0.229550
Swal0ing Difficulty           3.654655   -2.884689 -0.064957      -0.373013
Clubbing of Finger Nails     10.832332    1.125524 -0.040086       1.168849
Frequent Cold                24.180180   -0.279544 -0.000472       0.649409
Dry Cough                     2.233734    0.296875 -0.123029       1.082563
Snoring                      -1.259259   -0.083207 -0.131383      -0.063904
Level                        13.601602    0.587800 -0.065990       1.052973

                          Alcohol use  Dust Allergy  OccuPational Hazards  \
index                       31.313814     21.717217             19.696697
Age                          4.773812      0.837127              1.573413
Gender                      -0.292619     -0.198529             -0.198879
Air Pollution                3.976056      2.563964              2.606006
Alcohol use                  6.866898      4.249354              4.853934
Dust Allergy                 4.249354      3.923699              3.489890
OccuPational Hazards         4.853934      3.489890              4.442843
Genetic Risk                 4.889349      3.319620              4.003804
chronic Lung Disease         3.698759      2.268569              3.344144
Balanced Diet                3.656223      2.737723              3.112673
Obesity                      3.726932      2.949224              3.234635
Smoking                      3.577854      1.773353              2.618298
Passive Smoker               3.589805      2.564389              2.705906
Chest Pain                   4.285692      2.890621              3.727808
Coughing of Blood            4.247631      3.007272              3.305746
Fatigue                      1.395467      1.478238              1.267227
Weight Loss                  1.201837      1.406331              0.819620
Shortness of Breath          2.609489      2.347748              1.765165
Wheezing                     0.967517      1.233028              0.770090
Swal0ing Difficulty         -0.678677      0.140050             -0.013654
Clubbing of Finger Nails     2.596948      1.635340              1.844525
Frequent Cold                0.868100      0.796356              0.298058
Dry Cough                    1.128890      1.212467              0.687167
Snoring                      0.474136      0.154364              0.071231
```

```
Level                          1.535630      1.152923              1.157077

                          Genetic Risk  chronic Lung Disease  Balanced Diet  \
index                        18.874875             13.441441      18.961461
Age                           1.867948              2.861742       0.124691
```

```
import seaborn as sn
import matplotlib.pyplot as plt
sn.heatmap(cov_matrix,annot=True)
```

```
<Axes: >
```


```
x=d.iloc[:,3:8]
y=d.iloc[:,8:9]
print(x)
print(y)
```

```
     Gender  Air Pollution  Alcohol use  Dust Allergy  OccuPational Hazards
0         1              2            4             5                     4
1         1              3            1             5                     3
2         1              4            5             6                     5
3         1              7            7             7                     7
4         1              6            8             7                     7
..      ...            ...          ...           ...                   ...
995       1              6            7             7                     7
996       2              6            8             7                     7
```

```
997       2             4           5            6                   5
998       2             6           8            7                   7
999       1             6           5            6                   5

[1000 rows x 5 columns]
     Genetic Risk
0               3
1               4
2               5
3               6
4               7
..            ...
995             7
996             7
997             5
998             7
999             5

[1000 rows x 1 columns]
```

```
from sklearn.model_selection import train_test_split
```

```
from imblearn.over_sampling import SMOTE
sm=SMOTE(random_state=42)
x,y=sm.fit_resample(x,y)
print(x.shape)
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state=True,test_size=0.20)
```

```
(2289, 5)
```

```
from sklearn.linear_model import LogisticRegression
```

```
lr=LogisticRegression()
```

```
mm=lr.fit(x_train,y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/validation.py:1183: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:460: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

```
yp=mm.predict(x_test)
```

```
print(yp)
```

```
[2 1 1 4 7 2 2 3 5 5 7 2 7 2 4 2 7 1 7 4 5 3 1 4 1 3 5 7 3 7 5 1 1 7 5 2 7
 5 6 1 2 5 5 7 6 6 7 7 5 2 4 1 1 5 5 1 1 5 2 5 4 2 2 2 1 7 3 5 2 4 4 7 4 1
 4 1 3 7 3 1 2 2 1 5 2 6 1 7 5 4 4 4 2 3 7 4 1 5 5 6 1 7 1 7 6 5 7 4 3 5 1
 1 5 2 5 3 5 5 4 4 7 1 2 3 4 2 5 3 2 1 1 4 3 7 7 1 2 7 5 2 4 1 5 1 1 1 5 2
 1 2 4 7 5 5 5 2 2 1 2 1 7 1 3 7 2 1 2 2 7 1 1 1 3 5 7 7 5 1 2 3 7 1 1 1 3
 2 5 1 2 5 1 5 1 2 1 4 5 2 1 7 6 4 7 5 7 5 1 7 5 2 4 5 4 7 7 7 5 1 4 2 1 7
 1 2 3 5 7 4 2 4 7 5 1 2 7 6 1 2 5 4 1 7 2 7 1 5 4 1 5 6 1 5 5 4 7 1 2 5 1
 7 7 1 1 3 7 5 7 7 1 5 2 7 4 7 5 7 5 5 5 1 2 1 5 4 5 7 2 1 7 7 2 5 6 1 4 1
 7 2 1 7 1 7 5 5 1 1 6 5 5 3 5 7 7 7 7 7 2 1 4 5 7 3 7 4 4 7 1 4 1 2 2 2 7
 5 7 1 4 4 4 1 7 1 6 2 7 4 5 1 4 1 5 5 7 1 5 1 3 2 2 1 2 3 2 1 7 1 1 1 7 2
 1 5 2 1 1 7 7 3 1 2 4 2 7 1 1 1 7 2 4 1 4 2 4 1 5 7 7 7 1 1 2 3 1 7 5 2 2
 2 1 4 5 4 2 2 1 1 5 5 1 2 2 2 5 4 2 4 4 5 7 1 5 2 1 4 1 1 2 5 7 7 5 1 5 7
 5 7 5 1 1 3 7 2 2 6 6 1 1 4]
```

```
from sklearn.metrics import accuracy_score
print(accuracy_score(yp,y_test))
```

```
0.6615720524017468
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
  classifier=DecisionTreeClassifier(criterion='entropy',random_state=0)
  mm=classifier.fit(x_train,y_train)
```

```
yp=mm.predict(x_test)
```

```
print(yp)
```

```
[4 1 3 4 7 2 3 1 5 5 7 2 7 4 2 3 7 1 6 4 5 3 1 2 1 1 5 7 3 7 5 1 1 7 6 3 7
 5 6 3 3 5 5 6 6 6 7 7 5 2 2 1 1 5 5 1 3 5 2 5 2 4 4 4 1 7 1 5 2 2 4 7 2 1
 4 1 3 7 3 3 3 3 1 5 2 6 3 6 5 4 4 2 2 3 6 4 1 5 5 6 1 7 1 7 6 5 7 2 3 5 1
 1 5 2 5 3 5 5 4 4 7 1 4 3 4 2 5 3 2 1 1 2 3 7 7 1 4 7 5 4 2 1 5 1 1 3 5 2
 1 2 4 7 5 5 5 3 4 1 2 1 7 1 3 7 4 1 2 4 6 3 1 1 3 6 7 7 6 1 2 3 7 1 1 3 3
 2 5 1 2 5 1 5 3 2 1 4 6 2 1 7 3 4 7 5 7 5 3 7 5 3 2 5 4 7 7 7 5 1 4 3 1 6
 1 2 3 5 7 6 4 4 7 5 1 3 7 6 1 3 5 2 1 7 4 7 1 5 2 3 5 6 1 5 5 2 7 2 4 5 1
 7 7 1 1 3 6 5 7 7 3 5 2 7 2 7 5 7 5 5 5 3 2 1 5 4 5 7 4 1 7 7 4 5 6 1 2 1
 7 3 1 7 1 7 5 5 1 1 6 5 5 3 5 7 7 7 6 7 4 1 4 5 7 3 7 2 2 7 3 2 1 3 2 4 7
 5 7 1 4 4 4 1 7 1 3 2 6 2 5 1 2 2 5 5 7 1 5 1 3 3 2 1 4 3 6 1 7 1 3 1 7 2
 1 5 4 1 1 7 7 2 2 4 2 2 7 2 2 1 7 2 2 1 2 4 4 2 5 7 6 7 1 2 4 2 1 7 5 6 2
```