

# stat-assignment-5

September 5, 2024

## 1 Question-02

```
[42]: import pandas as pd
      d=pd.read_csv('/diabetes_data_upload.csv')
      print(d)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	Polyphagia	\
0	40	Male	No	Yes		No	Yes	No
1	58	Male	No	No		No	Yes	No
2	41	Male	Yes	No		No	Yes	Yes
3	45	Male	No	No		Yes	Yes	Yes
4	60	Male	Yes	Yes		Yes	Yes	Yes
..	...	...	...	...	...	...	...	
515	39	Female	Yes	Yes		Yes	No	Yes
516	48	Female	Yes	Yes		Yes	Yes	Yes
517	58	Female	Yes	Yes		Yes	Yes	Yes
518	32	Female	No	No		No	Yes	No
519	42	Male	No	No		No	No	No

	Genital thrush	visual blurring	Itching	Irritability	delayed healing	\
0	No	No	Yes	No	Yes	
1	No	Yes	No	No	No	
2	No	No	Yes	No	Yes	
3	Yes	No	Yes	No	Yes	
4	No	Yes	Yes	Yes	Yes	
..	...	...	...	...	...	
515	No	No	Yes	No	Yes	
516	No	No	Yes	Yes	Yes	
517	No	Yes	No	No	No	
518	No	Yes	Yes	No	Yes	
519	No	No	No	No	No	

	partial paresis	muscle stiffness	Alopecia	Obesity	class
0	No	Yes	Yes	Yes	Positive
1	Yes	No	Yes	No	Positive
2	No	Yes	Yes	No	Positive
3	No	No	No	No	Positive
4	Yes	Yes	Yes	Yes	Positive

..	...	...	...	...	...
515	Yes	No	No	No	Positive
516	Yes	No	No	No	Positive
517	Yes	Yes	No	Yes	Positive
518	No	No	Yes	No	Negative
519	No	No	No	No	Negative

[520 rows x 17 columns]

```
[43]: d.dtypes
```

```
[43]: Age                int64
      Gender            object
      Polyuria          object
      Polydipsia        object
      sudden weight loss object
      weakness          object
      Polyphagia        object
      Genital thrush    object
      visual blurring   object
      Itching           object
      Irritability      object
      delayed healing   object
      partial paresis   object
      muscle stiffness  object
      Alopecia          object
      Obesity           object
      class             object
      dtype: object
```

```
[44]: features=d.columns
      print(features)
```

```
Index(['Age', 'Gender', 'Polyuria', 'Polydipsia', 'sudden weight loss',
      'weakness', 'Polyphagia', 'Genital thrush', 'visual blurring',
      'Itching', 'Irritability', 'delayed healing', 'partial paresis',
      'muscle stiffness', 'Alopecia', 'Obesity', 'class'],
      dtype='object')
```

```
[46]: target=d.columns[-1]
      print(target)
```

```
class
```

```
[47]: mapping={'Female':1,'Male':0,'No':3,'Yes':4,'Positive':5,'Negative':6}
      d=d.replace(mapping)
      print(d)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	\
0	40	0	3	4	3	4	
1	58	0	3	3	3	4	
2	41	0	4	3	3	4	
3	45	0	3	3	4	4	
4	60	0	4	4	4	4	
..	...	...	...	...	...	...	
515	39	1	4	4	4	3	
516	48	1	4	4	4	4	
517	58	1	4	4	4	4	
518	32	1	3	3	3	4	
519	42	0	3	3	3	3	

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	\
0	3	3	3	4	3	
1	3	3	4	3	3	
2	4	3	3	4	3	
3	4	4	3	4	3	
4	4	3	4	4	4	
..	...	...	...	...	...	
515	4	3	3	4	3	
516	4	3	3	4	4	
517	4	3	4	3	3	
518	3	3	4	4	3	
519	3	3	3	3	3	

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity	\
0	4	3	4	4	4	
1	3	4	3	4	3	
2	4	3	4	4	3	
3	4	3	3	3	3	
4	4	4	4	4	4	
..	...	...	...	...	...	
515	4	4	3	3	3	
516	4	4	3	3	3	
517	3	4	4	3	4	
518	4	3	3	4	3	
519	3	3	3	3	3	

	class
0	5
1	5
2	5
3	5
4	5
..	...
515	5
516	5

```
517      5
518      6
519      6
```

```
[520 rows x 17 columns]
```

```
[48]: df=d-d.min()/(d.max()-d.min())
      print(df)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness \
0	39.783784	0.0	0.0	1.0	0.0	1.0
1	57.783784	0.0	0.0	0.0	0.0	1.0
2	40.783784	0.0	1.0	0.0	0.0	1.0
3	44.783784	0.0	0.0	0.0	1.0	1.0
4	59.783784	0.0	1.0	1.0	1.0	1.0
..	...	...	...	...	...	...
515	38.783784	1.0	1.0	1.0	1.0	0.0
516	47.783784	1.0	1.0	1.0	1.0	1.0
517	57.783784	1.0	1.0	1.0	1.0	1.0
518	31.783784	1.0	0.0	0.0	0.0	1.0
519	41.783784	0.0	0.0	0.0	0.0	0.0

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability \
0	0.0	0.0	0.0	1.0	0.0
1	0.0	0.0	1.0	0.0	0.0
2	1.0	0.0	0.0	1.0	0.0
3	1.0	1.0	0.0	1.0	0.0
4	1.0	0.0	1.0	1.0	1.0
..	...	...	...	...	...
515	1.0	0.0	0.0	1.0	0.0
516	1.0	0.0	0.0	1.0	1.0
517	1.0	0.0	1.0	0.0	0.0
518	0.0	0.0	1.0	1.0	0.0
519	0.0	0.0	0.0	0.0	0.0

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity \
0	1.0	0.0	1.0	1.0	1.0
1	0.0	1.0	0.0	1.0	0.0
2	1.0	0.0	1.0	1.0	0.0
3	1.0	0.0	0.0	0.0	0.0
4	1.0	1.0	1.0	1.0	1.0
..	...	...	...	...	...
515	1.0	1.0	0.0	0.0	0.0
516	1.0	1.0	0.0	0.0	0.0
517	0.0	1.0	1.0	0.0	1.0
518	1.0	0.0	0.0	1.0	0.0
519	0.0	0.0	0.0	0.0	0.0

```

      class
0      0.0
1      0.0
2      0.0
3      0.0
4      0.0
..      ...
515     0.0
516     0.0
517     0.0
518     1.0
519     1.0

```

[520 rows x 17 columns]

```
[50]: y=d["class"]
      print(y)
```

```

0      5
1      5
2      5
3      5
4      5
..
515     5
516     5
517     5
518     6
519     6

```

Name: class, Length: 520, dtype: int64

```
[51]: x=d.drop("class",axis=1)
      print(x)
```

```

      Age  Gender  Polyuria  Polydipsia  sudden weight loss  weakness  \
0      40      0         3         4              3          4
1      58      0         3         3              3          4
2      41      0         4         3              3          4
3      45      0         3         3              4          4
4      60      0         4         4              4          4
..      ...      ...      ...      ...      ...      ...
515     39      1         4         4              4          3
516     48      1         4         4              4          4
517     58      1         4         4              4          4
518     32      1         3         3              3          4
519     42      0         3         3              3          3

```

```

      Polyphagia  Genital thrush  visual blurring  Itching  Irritability  \

```

0	3	3	3	4	3
1	3	3	4	3	3
2	4	3	3	4	3
3	4	4	3	4	3
4	4	3	4	4	4
..	...	...	...	...	...
515	4	3	3	4	3
516	4	3	3	4	4
517	4	3	4	3	3
518	3	3	4	4	3
519	3	3	3	3	3

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity
0	4	3	4	4	4
1	3	4	3	4	3
2	4	3	4	4	3
3	4	3	3	3	3
4	4	4	4	4	4
..	...	...	...	...	...
515	4	4	3	3	3
516	4	4	3	3	3
517	3	4	4	3	4
518	4	3	3	4	3
519	3	3	3	3	3

[520 rows x 16 columns]

```
[52]: x = x.replace({'Gender': {'Male': 0, 'Female': 1},
                    'Polyuria': {'Yes': 1, 'No': 0},
                    'Polydipsia': {'Yes': 1, 'No': 0},
                    'sudden weight loss': {'Yes': 1, 'No': 0},
                    'weakness': {'Yes': 1, 'No': 0},
                    'Polyphagia': {'Yes': 1, 'No': 0},
                    'Genital thrush': {'Yes': 1, 'No': 0},
                    'visual blurring': {'Yes': 1, 'No': 0},
                    'Itching': {'Yes': 1, 'No': 0},
                    'Irritability': {'Yes': 1, 'No': 0},
                    'delayed healing': {'Yes': 1, 'No': 0},
                    'partial paresis': {'Yes': 1, 'No': 0},
                    'muscle stiffness': {'Yes': 1, 'No': 0},
                    'Alopecia': {'Yes': 1, 'No': 0},
                    'Obesity': {'Yes': 1, 'No': 0}})

print(x)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	\
0	40	0	3	4	3	4	
1	58	0	3	3	3	4	
2	41	0	4	3	3	4	

3	45	0	3	3	4	4
4	60	0	4	4	4	4
..	...	...	...	...	...	...
515	39	1	4	4	4	3
516	48	1	4	4	4	4
517	58	1	4	4	4	4
518	32	1	3	3	3	4
519	42	0	3	3	3	3

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	\
0	3		3	3	4	3
1	3		3	4	3	3
2	4		3	3	4	3
3	4		4	3	4	3
4	4		3	4	4	4
..	...	...	...	...	...	...
515	4		3	3	4	3
516	4		3	3	4	4
517	4		3	4	3	3
518	3		3	4	4	3
519	3		3	3	3	3

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity
0	4		3	4	4
1	3		4	3	4
2	4		3	4	4
3	4		3	3	3
4	4		4	4	4
..	...	...	...	...	...
515	4		4	3	3
516	4		4	3	3
517	3		4	4	4
518	4		3	3	4
519	3		3	3	3

[520 rows x 16 columns]

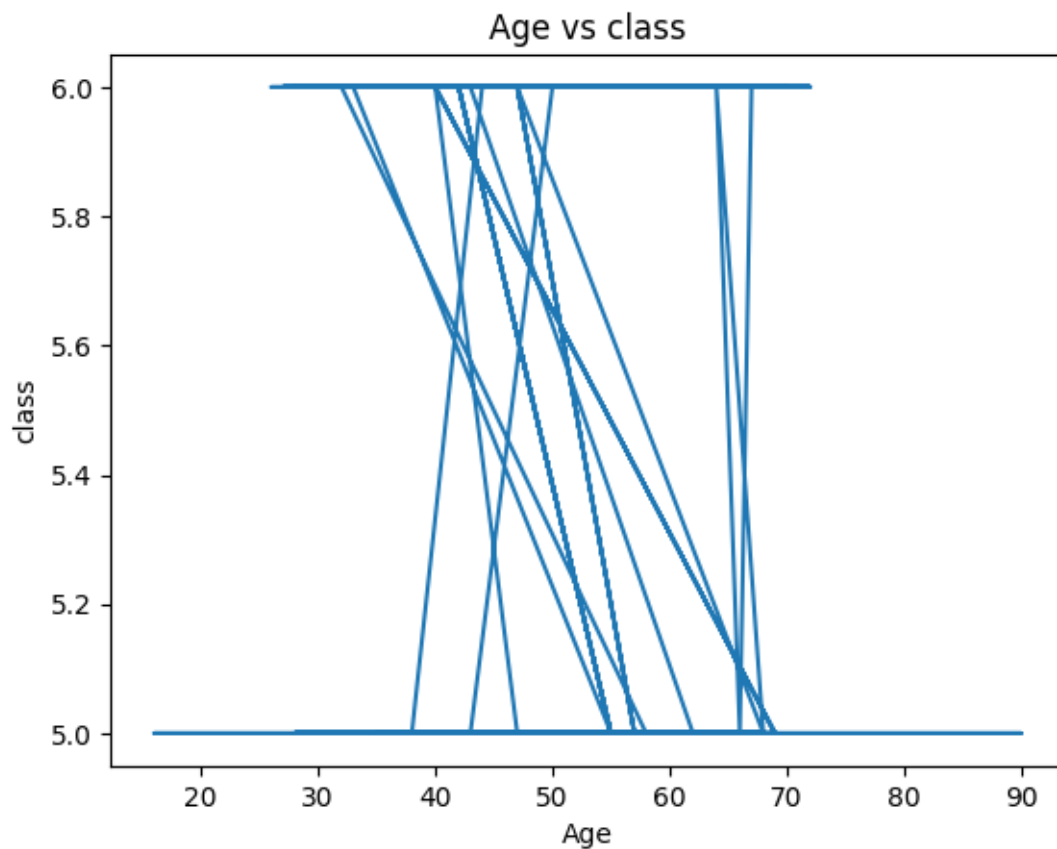
```
[53]: y = y.replace({'Positive': 0, 'Negative': 1})
      print(y)
```

0	5
1	5
2	5
3	5
4	5
..	..
515	5
516	5

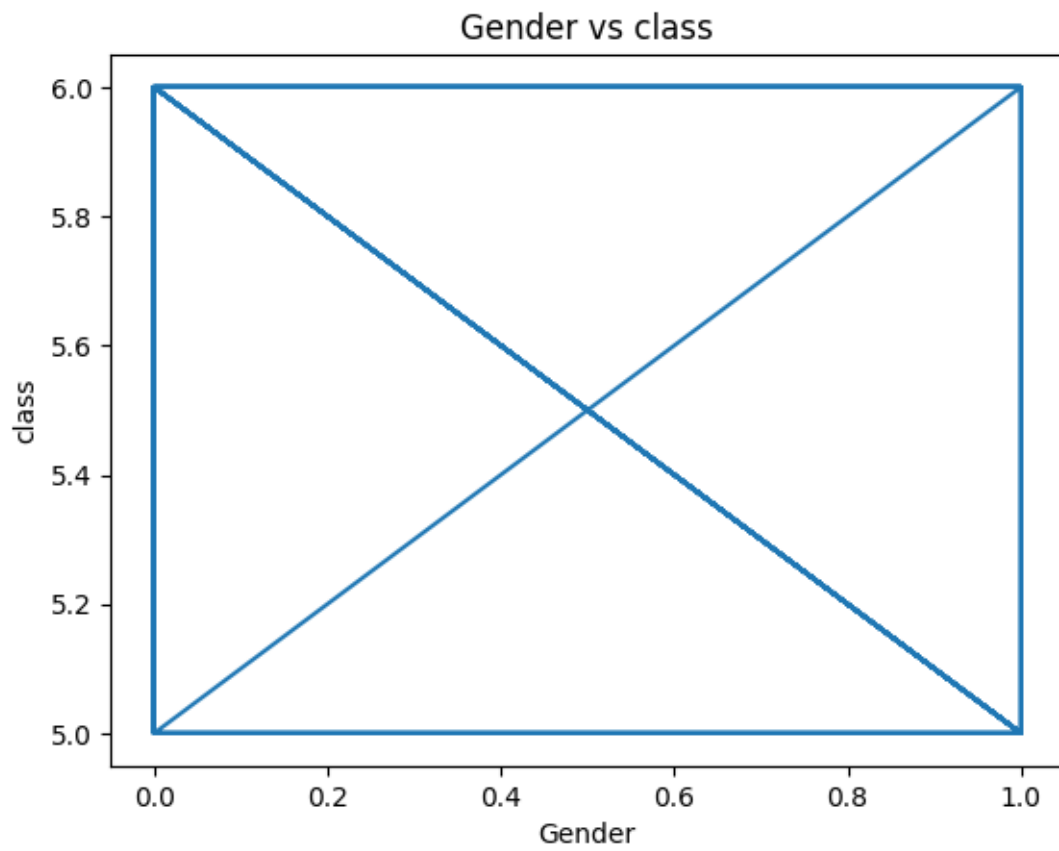
```
517     5
518     6
519     6
Name: class, Length: 520, dtype: int64
```

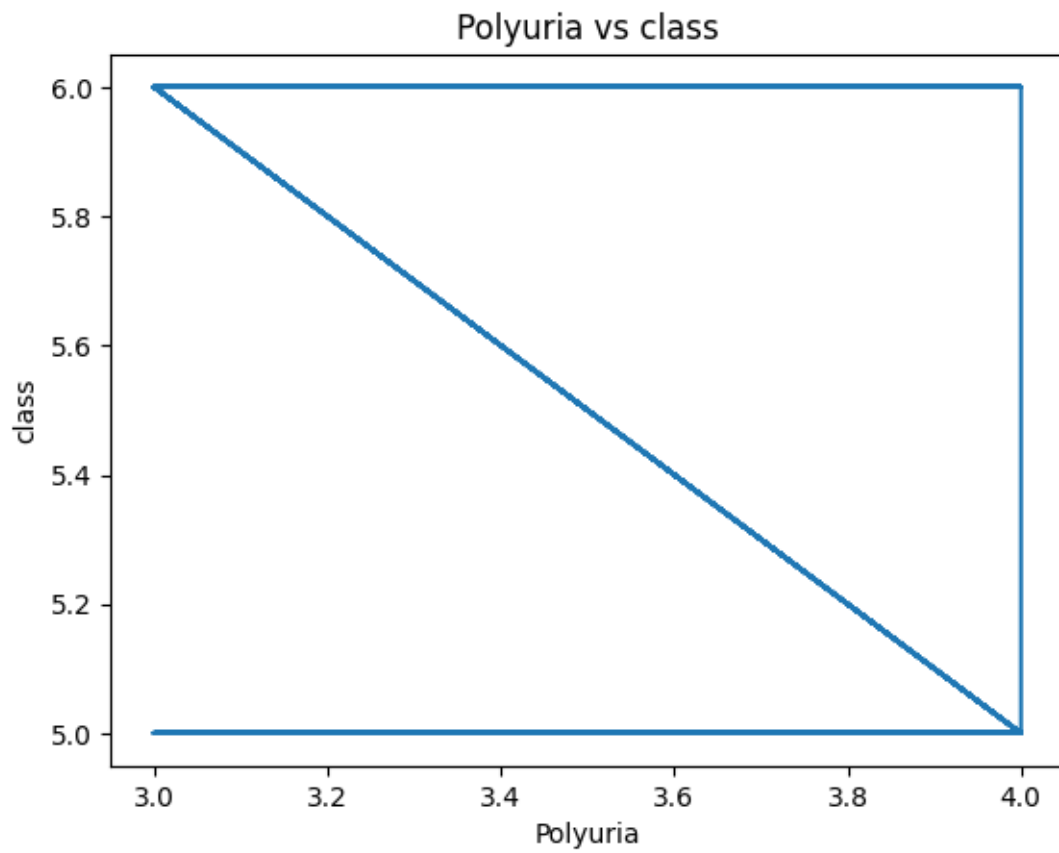
```
[54]: import pandas as pd
import matplotlib.pyplot as plt
```

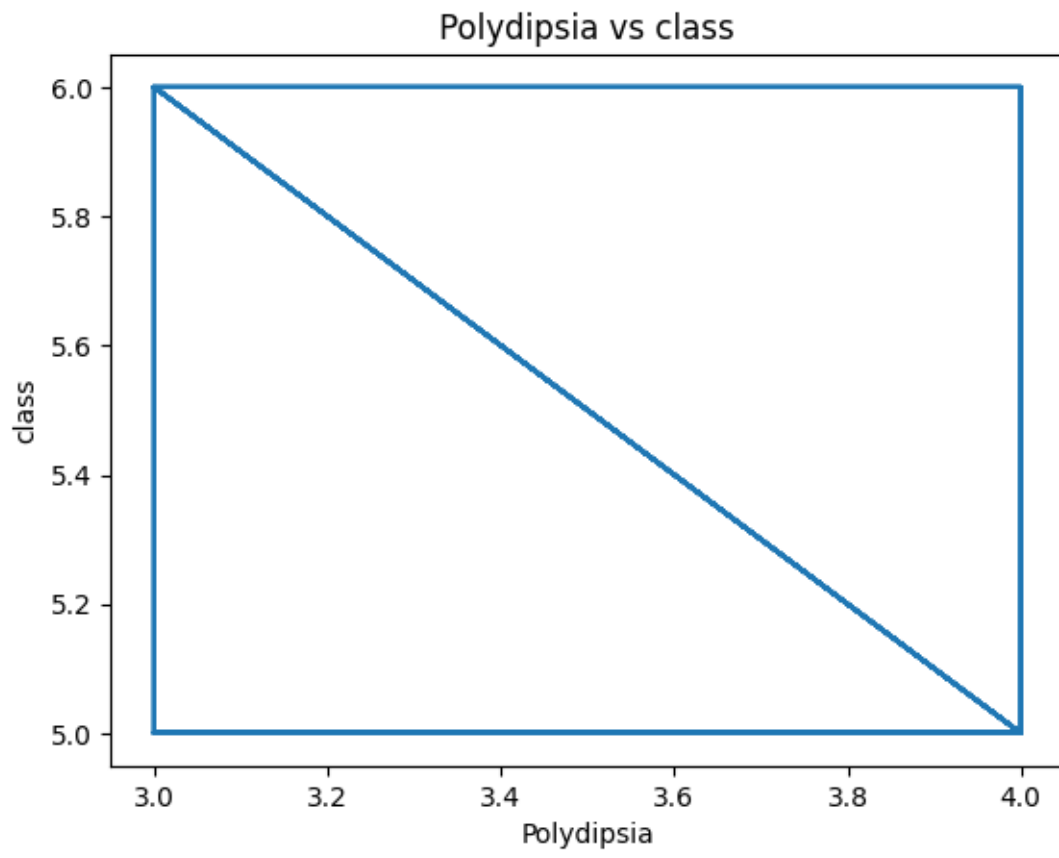
```
[55]: for column in x.columns:
    plt.plot(x[column], y)
    plt.xlabel(column)
    plt.ylabel('class')
    plt.title(f'{column} vs class')
    plt.show()
```

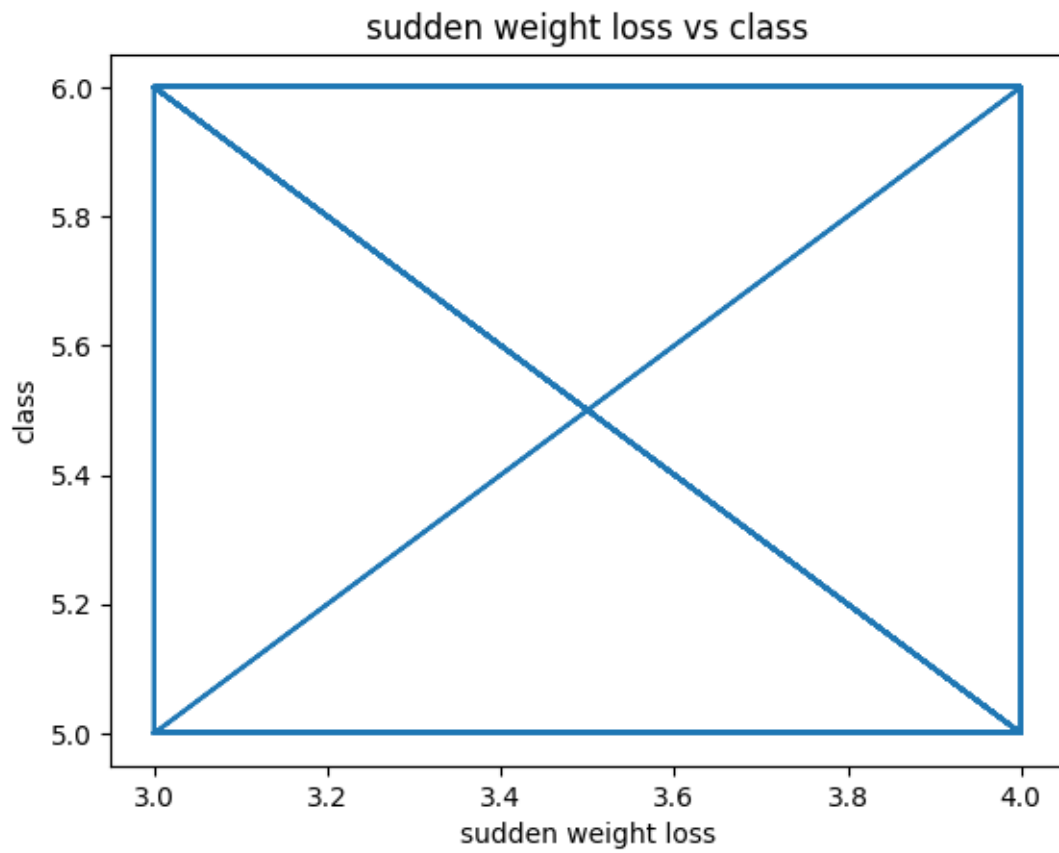


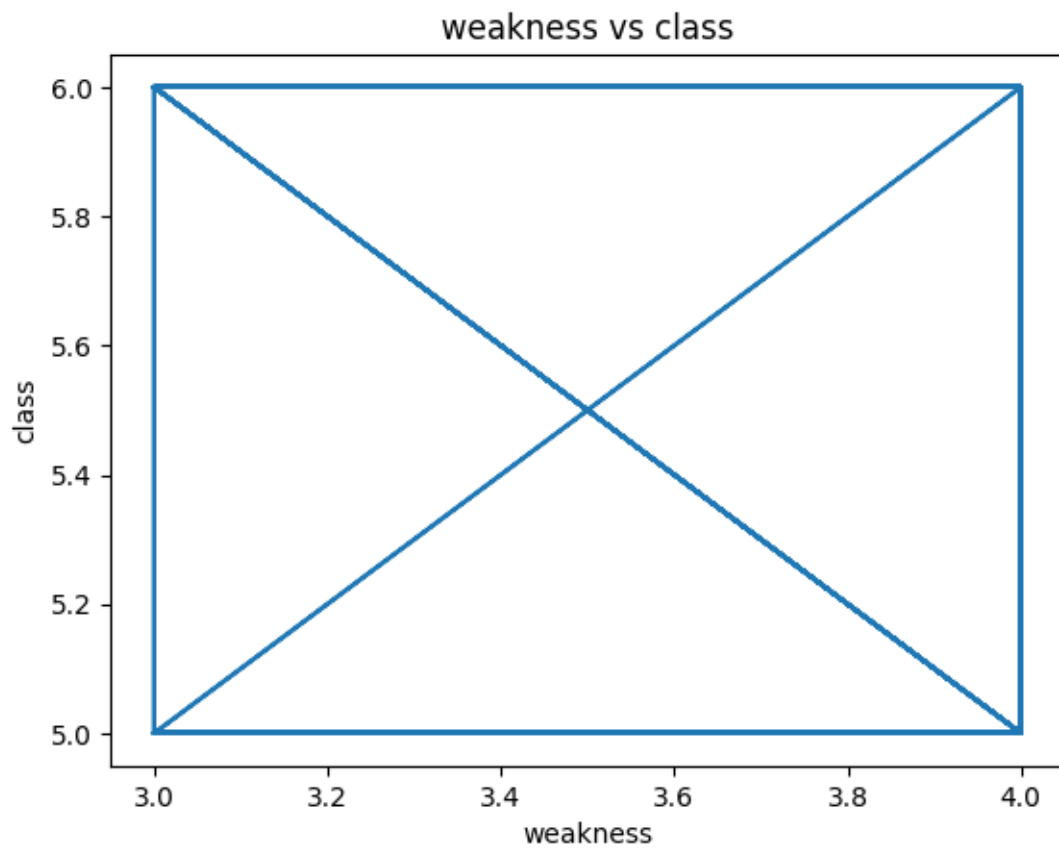


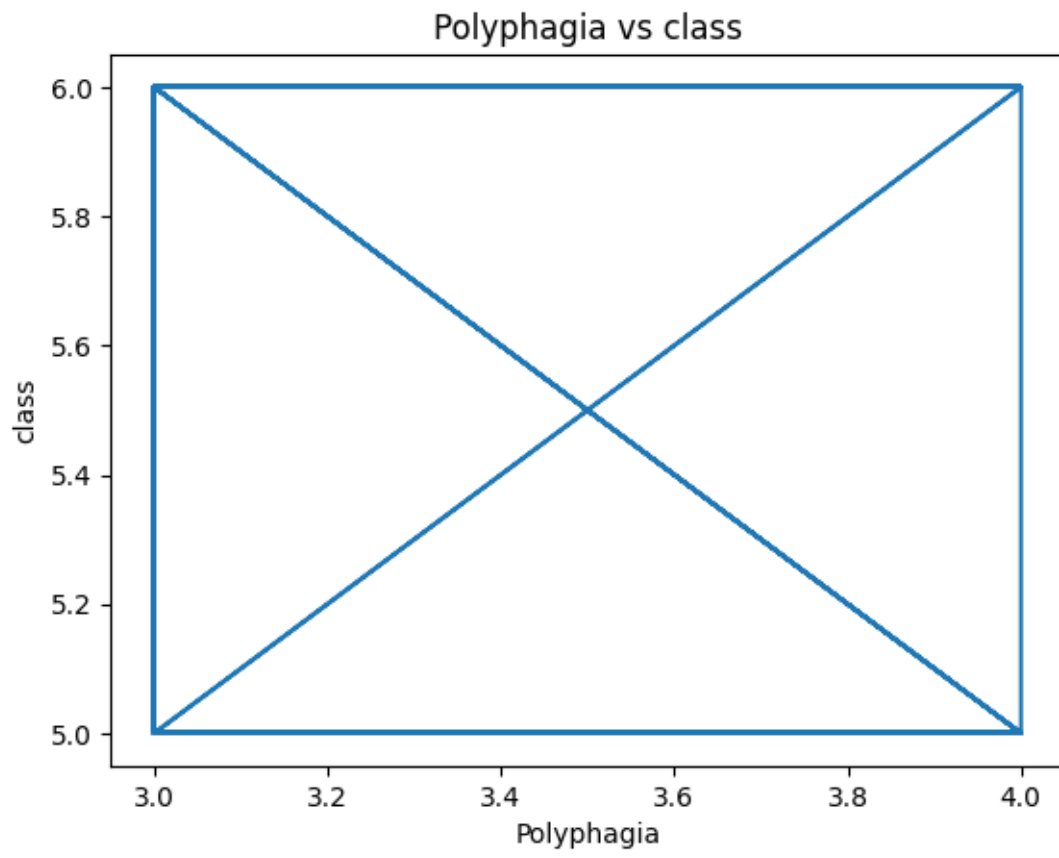


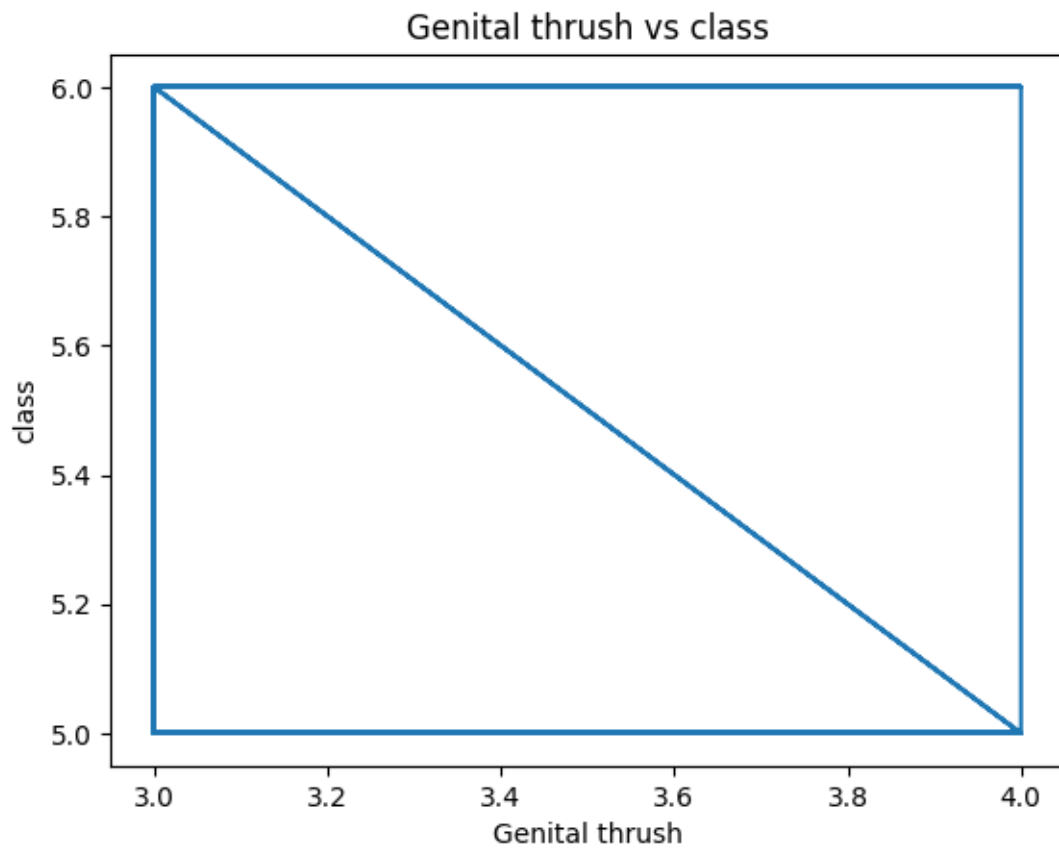


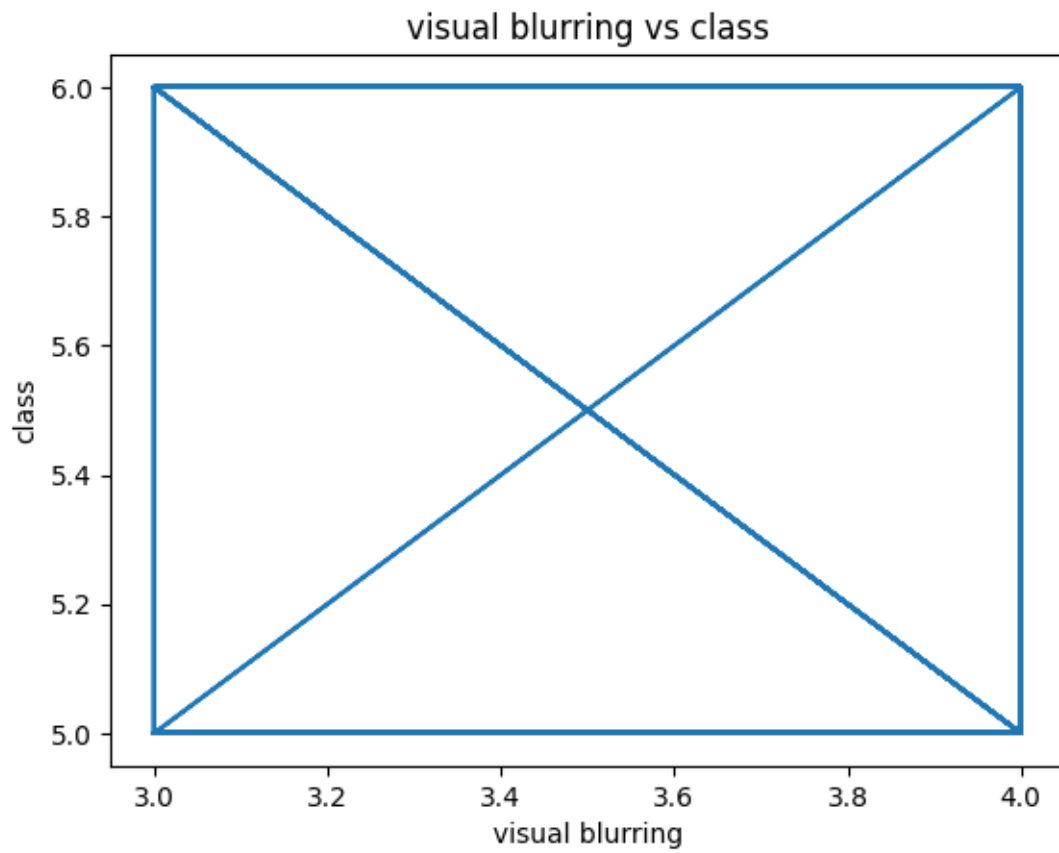




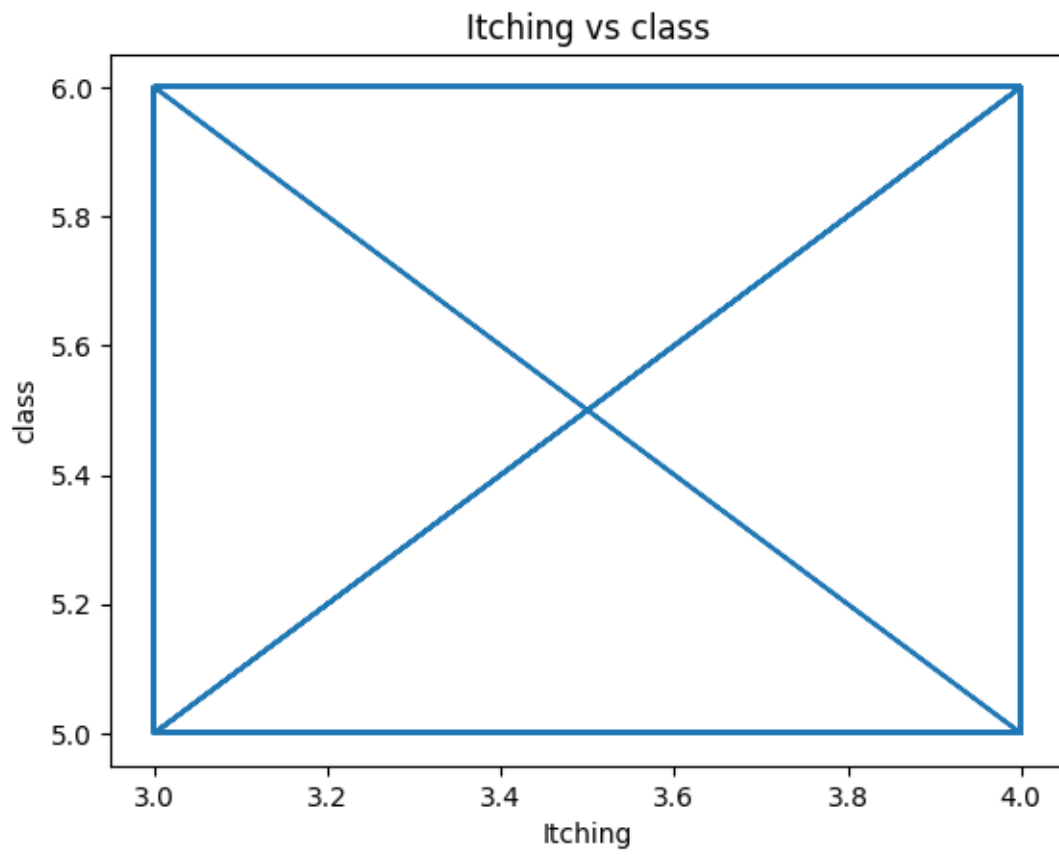


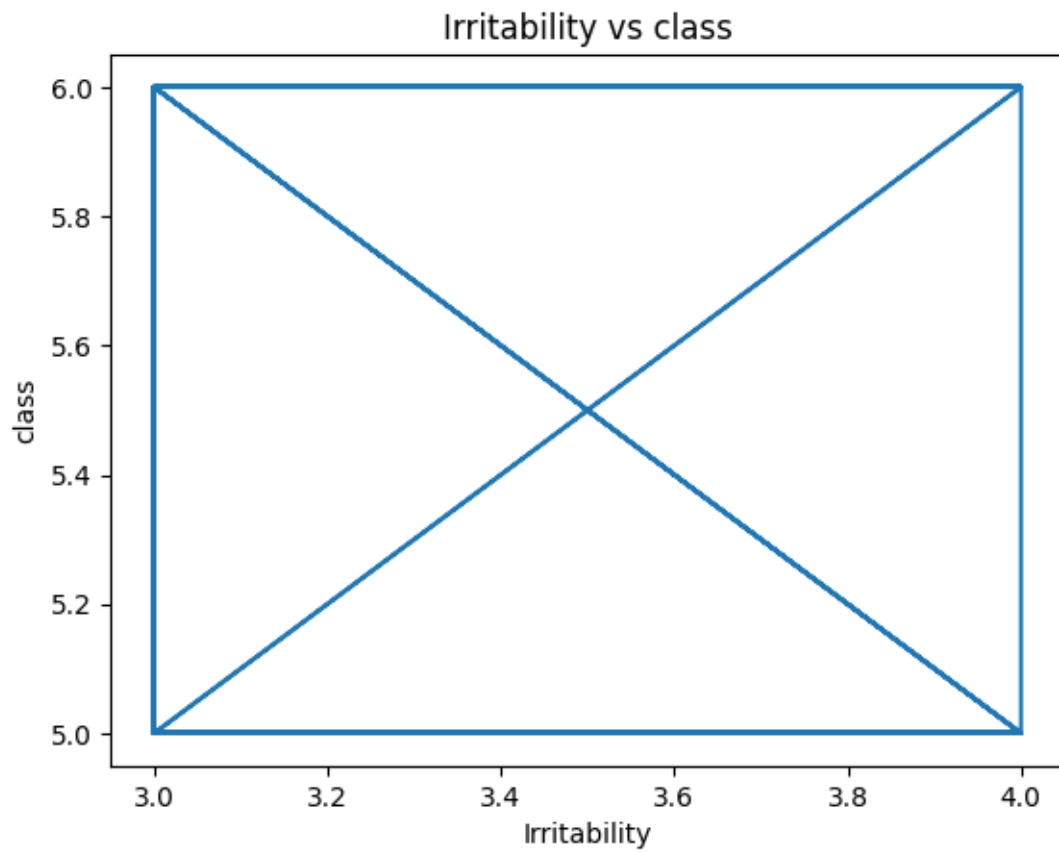


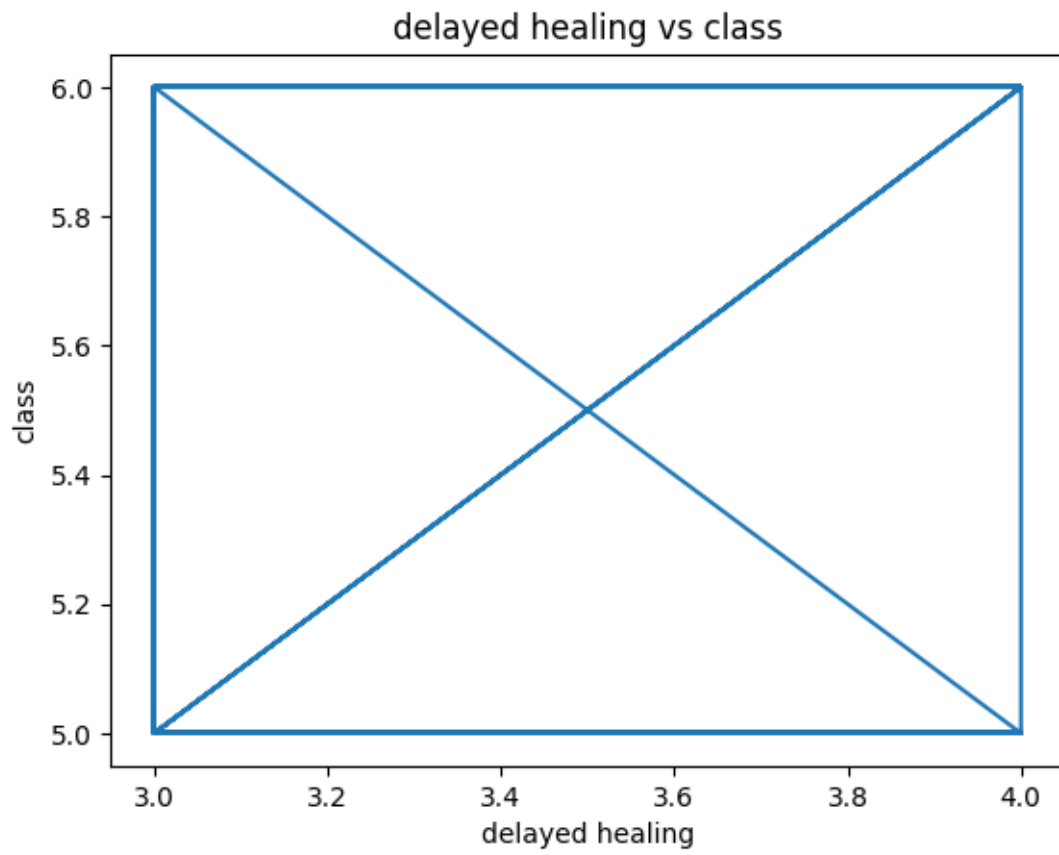


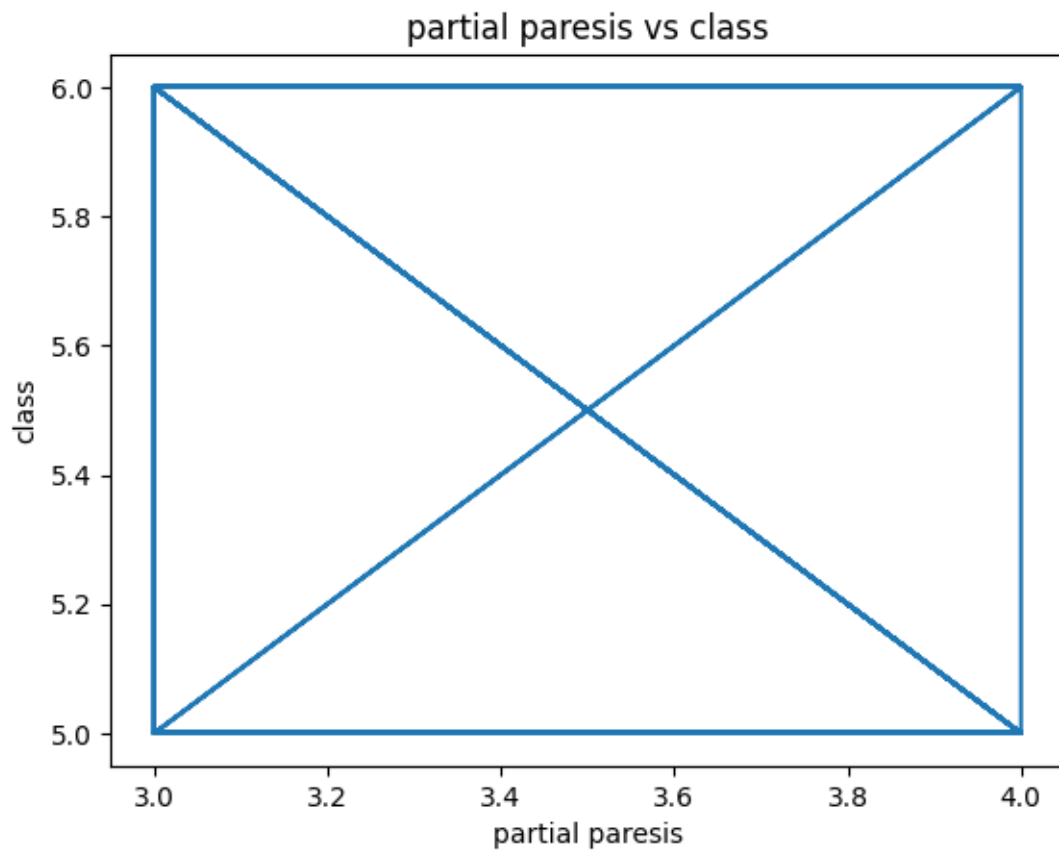


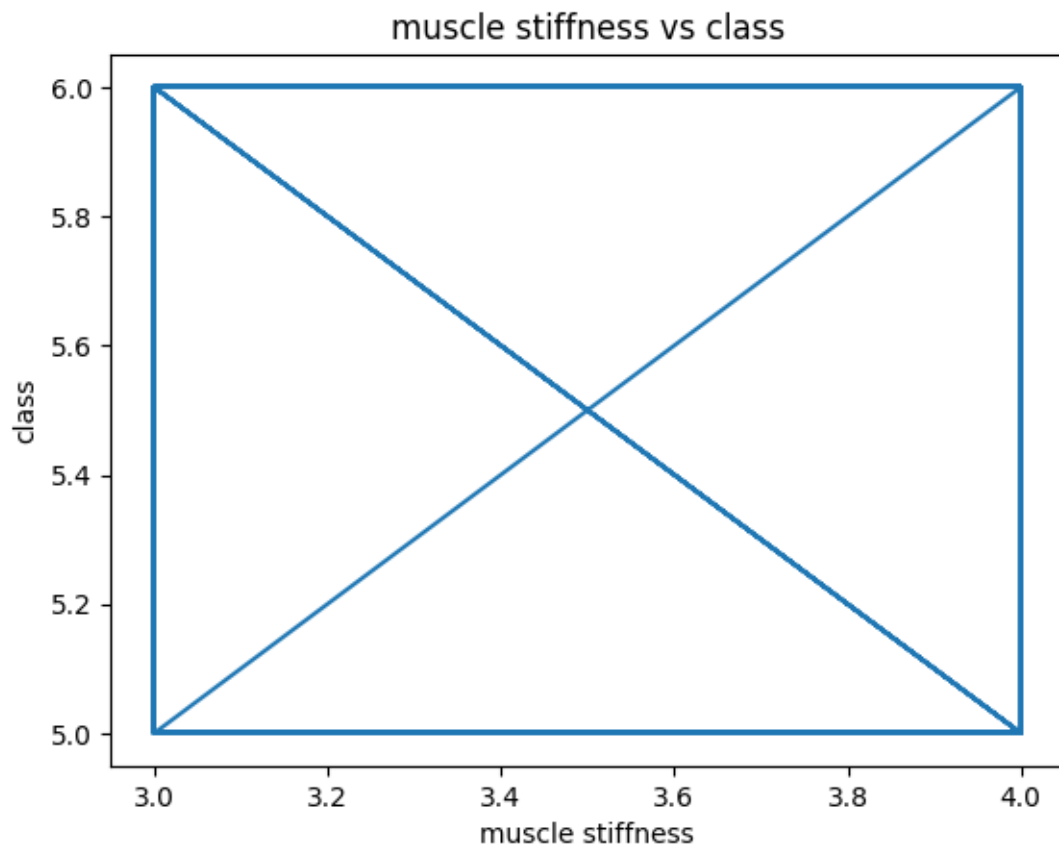


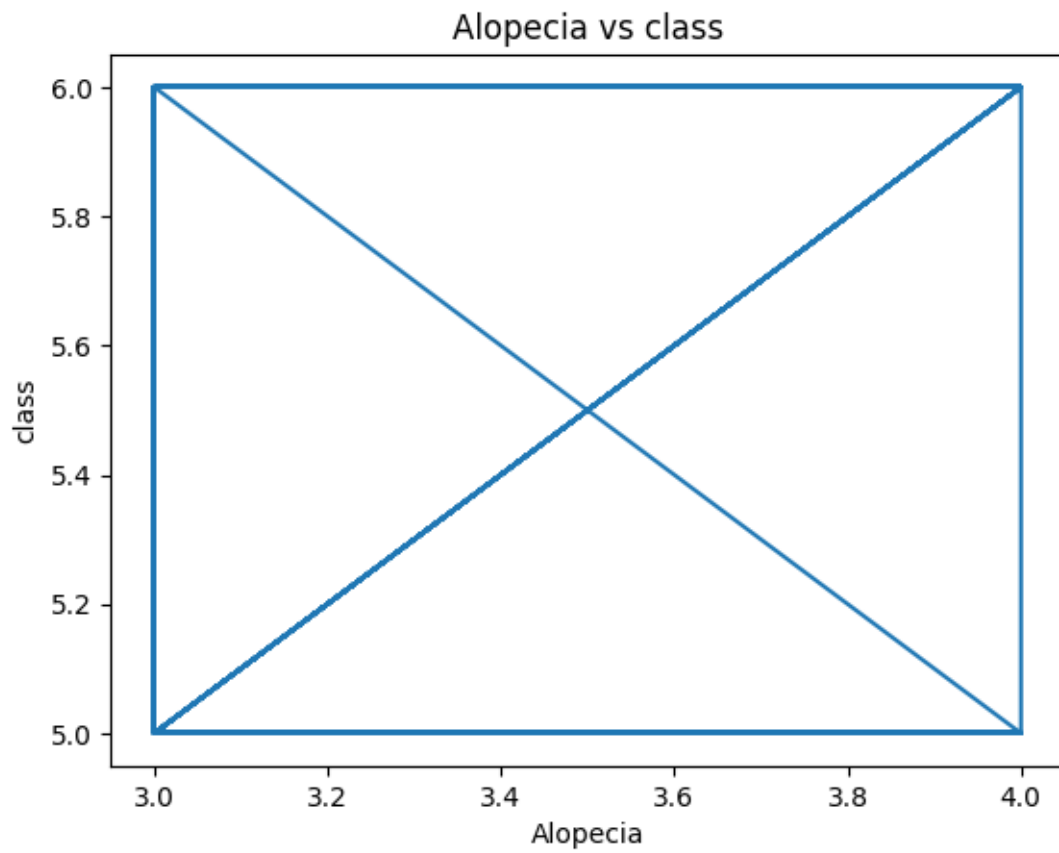


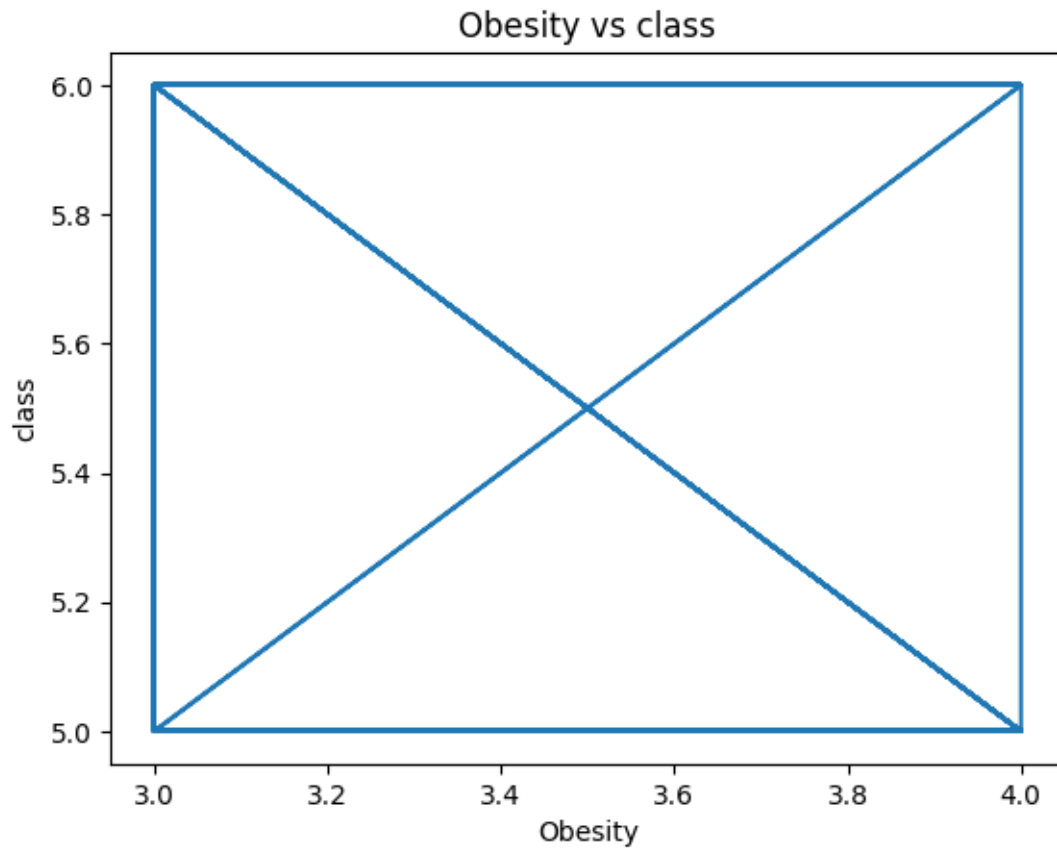












```
[56]: x.isnull().sum()
      x.fillna(0, inplace=True)
      print(x)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	\
0	40	0	3	4	3	4	
1	58	0	3	3	3	4	
2	41	0	4	3	3	4	
3	45	0	3	3	4	4	
4	60	0	4	4	4	4	
..	..	..	..	..	..	..	
515	39	1	4	4	4	3	
516	48	1	4	4	4	4	
517	58	1	4	4	4	4	
518	32	1	3	3	3	4	
519	42	0	3	3	3	3	

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	\
0	3	3	3	4	3	
1	3	3	4	3	3	

2	4	3	3	4	3
3	4	4	3	4	3
4	4	3	4	4	4
..	...	...	...	...	...
515	4	3	3	4	3
516	4	3	3	4	4
517	4	3	4	3	3
518	3	3	4	4	3
519	3	3	3	3	3

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity
0	4	3	4	4	4
1	3	4	3	4	3
2	4	3	4	4	3
3	4	3	3	3	3
4	4	4	4	4	4
..	...	...	...	...	...
515	4	4	3	3	3
516	4	4	3	3	3
517	3	4	4	3	4
518	4	3	3	4	3
519	3	3	3	3	3

[520 rows x 16 columns]

```
[57]: from sklearn.preprocessing import MinMaxScaler
      from sklearn.model_selection import train_test_split
```

```
[58]: scaler = MinMaxScaler()
      x_norm = pd.DataFrame(scaler.fit_transform(x), columns=x.columns)
      print(x_norm)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness \
0	0.324324	0.0	0.0	1.0	0.0	1.0
1	0.567568	0.0	0.0	0.0	0.0	1.0
2	0.337838	0.0	1.0	0.0	0.0	1.0
3	0.391892	0.0	0.0	0.0	1.0	1.0
4	0.594595	0.0	1.0	1.0	1.0	1.0
..	...	...	...	...	...	...
515	0.310811	1.0	1.0	1.0	1.0	0.0
516	0.432432	1.0	1.0	1.0	1.0	1.0
517	0.567568	1.0	1.0	1.0	1.0	1.0
518	0.216216	1.0	0.0	0.0	0.0	1.0
519	0.351351	0.0	0.0	0.0	0.0	0.0

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability \
0	0.0	0.0	0.0	1.0	0.0
1	0.0	0.0	1.0	0.0	0.0



2	1.0	0.0	0.0	1.0	0.0
3	1.0	1.0	0.0	1.0	0.0
4	1.0	0.0	1.0	1.0	1.0
..	...	...	...	...	...
515	1.0	0.0	0.0	1.0	0.0
516	1.0	0.0	0.0	1.0	1.0
517	1.0	0.0	1.0	0.0	0.0
518	0.0	0.0	1.0	1.0	0.0
519	0.0	0.0	0.0	0.0	0.0

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity
0	1.0	0.0	1.0	1.0	1.0
1	0.0	1.0	0.0	1.0	0.0
2	1.0	0.0	1.0	1.0	0.0
3	1.0	0.0	0.0	0.0	0.0
4	1.0	1.0	1.0	1.0	1.0
..	...	...	...	...	...
515	1.0	1.0	0.0	0.0	0.0
516	1.0	1.0	0.0	0.0	0.0
517	0.0	1.0	1.0	0.0	1.0
518	1.0	0.0	0.0	1.0	0.0
519	0.0	0.0	0.0	0.0	0.0

[520 rows x 16 columns]

```
[59]: xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.2, random_state=30)
print(xtrain)
print(xtest)
print(ytrain)
print(ytest)
```

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	\
195	31	0	4	3	3	3	
349	37	0	3	3	3	3	
257	48	1	4	4	3	4	
157	48	0	4	4	4	3	
459	57	0	4	4	4	4	
..	...	...	...	...	...	...	
430	32	0	3	3	3	3	
145	61	0	4	3	3	4	
140	47	0	4	4	3	3	
500	66	0	4	3	4	3	
421	61	1	4	3	3	3	

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	\
195	4	3	4	3	3	
349	3	3	3	3	3	
257	3	3	4	4	3	

157	4	4	3	3	3
459	4	3	4	3	3
..	...	...	...	...	...
430	3	4	3	3	4
145	4	4	4	4	4
140	3	3	3	3	3
500	3	4	3	4	4
421	4	3	3	3	4

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity
195	3	4	3	4	3
349	3	3	3	3	3
257	4	4	3	3	3
157	4	3	3	3	3
459	3	4	3	3	3
..	...	...	...	...	...
430	4	3	3	3	4
145	4	4	3	3	3
140	3	3	4	3	3
500	3	3	3	4	3
421	3	3	3	4	3

[416 rows x 16 columns]

	Age	Gender	Polyuria	Polydipsia	sudden weight loss	weakness	\
248	42	0	3	3	3	3	
60	65	1	4	4	3	4	
162	35	1	3	3	3	3	
391	58	0	3	4	3	3	
493	44	0	4	3	4	4	
..	...	...	...	...	...	...	
188	70	0	4	3	4	4	
158	56	0	4	3	4	4	
363	68	1	4	4	3	4	
164	48	1	4	4	4	4	
42	50	1	3	4	3	4	

	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	\
248	3	3	3	3	3	
60	4	3	3	4	3	
162	3	3	3	3	3	
391	3	3	4	4	3	
493	3	4	3	4	3	
..	...	...	...	...	...	
188	4	4	3	3	4	
158	3	4	3	4	4	
363	4	3	4	4	3	
164	4	3	4	4	4	
42	4	3	4	4	4	

	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity
248	3	3	3	4	3
60	3	4	4	3	3
162	3	3	3	3	3
391	3	3	4	3	3
493	4	3	3	4	4
..	...	...	...	...	...
188	4	4	4	4	3
158	3	3	3	4	3
363	4	4	3	3	3
164	4	4	4	3	3
42	4	4	4	3	3

[104 rows x 16 columns]

```

195    5
349    6
257    5
157    5
459    5
..
430    5
145    5
140    5
500    5
421    5
Name: class, Length: 416, dtype: int64
248    6
60     5
162    5
391    6
493    6
..
188    5
158    5
363    5
164    5
42     5
Name: class, Length: 104, dtype: int64

```

## 2 Question-01

```

[60]: import pandas as pd
      d=pd.read_csv('/content/Salary_Data.csv')
      print(d)

```

YearsExperience	Salary
-----------------	--------

0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

```
[61]: d.dtypes
```

```
[61]: YearsExperience    float64
Salary                  float64
dtype: object
```

```
[62]: features=d.columns
print(features)
```

```
Index(['YearsExperience', 'Salary'], dtype='object')
```

```
[63]: target=d.columns[-1]
print(target)
```

```
Salary
```

```
[64]: y=d['Salary']  
      print(y)
```

```
0      39343.0  
1      46205.0  
2      37731.0  
3      43525.0  
4      39891.0  
5      56642.0  
6      60150.0  
7      54445.0  
8      64445.0  
9      57189.0  
10     63218.0  
11     55794.0  
12     56957.0  
13     57081.0  
14     61111.0  
15     67938.0  
16     66029.0  
17     83088.0  
18     81363.0  
19     93940.0  
20     91738.0  
21     98273.0  
22    101302.0  
23    113812.0  
24    109431.0  
25    105582.0  
26    116969.0  
27    112635.0  
28    122391.0  
29    121872.0
```

Name: Salary, dtype: float64

```
[65]: x=d.drop('Salary',axis=1)  
      print(x)
```

```
      YearsExperience  
0              1.1  
1              1.3  
2              1.5  
3              2.0  
4              2.2  
5              2.9  
6              3.0  
7              3.2  
8              3.2
```

9	3.7
10	3.9
11	4.0
12	4.0
13	4.1
14	4.5
15	4.9
16	5.1
17	5.3
18	5.9
19	6.0
20	6.8
21	7.1
22	7.9
23	8.2
24	8.7
25	9.0
26	9.5
27	9.6
28	10.3
29	10.5

```
[66]: plt.scatter(d['YearsExperience'], d['Salary'])  
      plt.xlabel('YearsExperience')  
      plt.ylabel('Salary')  
      plt.title('YearsExperience vs Salary')  
      plt.show()
```



```
[67]: d_norm = (d - d.min()) / (d.max() - d.min())
      print(d_norm)
```

	YearsExperience	Salary
0	0.000000	0.019041
1	0.021277	0.100094
2	0.042553	0.000000
3	0.095745	0.068438
4	0.117021	0.025514
5	0.191489	0.223376
6	0.202128	0.264812
7	0.223404	0.197425
8	0.223404	0.315545
9	0.276596	0.229837
10	0.297872	0.301051
11	0.308511	0.213359
12	0.308511	0.227097
13	0.319149	0.228561
14	0.361702	0.276163
15	0.404255	0.356804
16	0.425532	0.334255

17	0.446809	0.535755
18	0.510638	0.515379
19	0.521277	0.663938
20	0.606383	0.637928
21	0.638298	0.715119
22	0.723404	0.750898
23	0.755319	0.898665
24	0.808511	0.846917
25	0.840426	0.801453
26	0.893617	0.935956
27	0.904255	0.884763
28	0.978723	1.000000
29	1.000000	0.993870