# sta-5

September 6, 2024

Question 1: 1.Read the data with pandas and find features and target variables 2.Find the target variable 3.Plot a graph between features and target 4.Normalize the give data[0 to 1]

[2]:
```
from typing_extensions import dataclass_transform
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
data = pd.read_csv('/content/Salary_Data.csv.xls')
data.head()
```

[2]:

| | 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 |

[3]:
```
target_var = data['Salary']
target_var.head()
```

[3]:
```
0    39343.0
1    46205.0
2    37731.0
3    43525.0
4    39891.0
Name: Salary, dtype: float64
```

[4]:
```
features=data.drop('Salary',axis=1)
features.head()
```

[4]:

| | YearsExperience |
|---|---|
| 0 | 1.1 |
| 1 | 1.3 |
| 2 | 1.5 |
| 3 | 2.0 |
| 4 | 2.2 |

[5]:
```
plt.scatter(features['YearsExperience'], target_var)
plt.xlabel('Years of Experience')
plt.ylabel('Salary Variable')
plt.title('Relationship between Years of Experience and Salary')
plt.show()
```


[6]:
```
x_normalized = features['YearsExperience']-features['YearsExperience'].min()/
  ↪features['YearsExperience'].max()-features['YearsExperience'].min()
x_normalized.head()
```

[6]:
```
0   -0.104762
1    0.095238
2    0.295238
3    0.795238
4    0.995238
Name: YearsExperience, dtype: float64
```

Question 2: 1.Read the data with pandas and find features and target variables 2.Plot a graph between features and target 3.Normalize the give data[0 to 1] 4.Convert string valued features to numbers

[7]:
```
data2=pd.read_csv('/content/diabetes_data_upload.csv.xls')
data2.head()
```

[7]:

| | Age | Gender | Polyuria | Polydipsia | sudden weight loss | weakness | Polyphagia | Genital thrush | visual blurring | Itching | Irritability | delayed healing | partial paresis | muscle stiffness | Alopecia | Obesity | class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 40 | Male | No | Yes | No | Yes | No | No | No | Yes | No | Yes | No | Yes | Yes | Yes | Positive |
| 1 | 58 | Male | No | No | No | Yes | No | No | Yes | No | No | No | Yes | No | Yes | No | Positive |
| 2 | 41 | Male | Yes | No | No | Yes | Yes | No | No | Yes | No | Yes | No | Yes | Yes | No | Positive |
| 3 | 45 | Male | No | No | Yes | Yes | Yes | Yes | No | Yes | No | Yes | No | No | No | No | Positive |
| 4 | 60 | Male | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Positive |

[8]:
```
target_var=data2['class']
target_var.head()
```

[8]:
```
0    Positive
1    Positive
2    Positive
3    Positive
4    Positive
Name: class, dtype: object
```

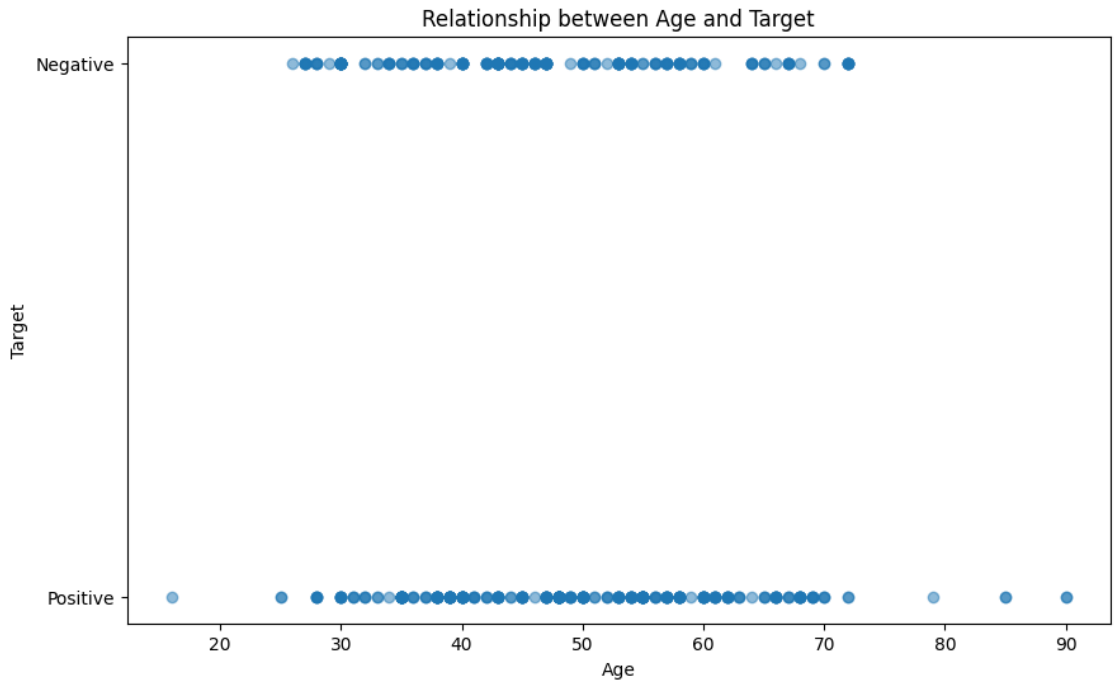
[9]:
```
features=data2.drop('class',axis=1)
features.head()
```

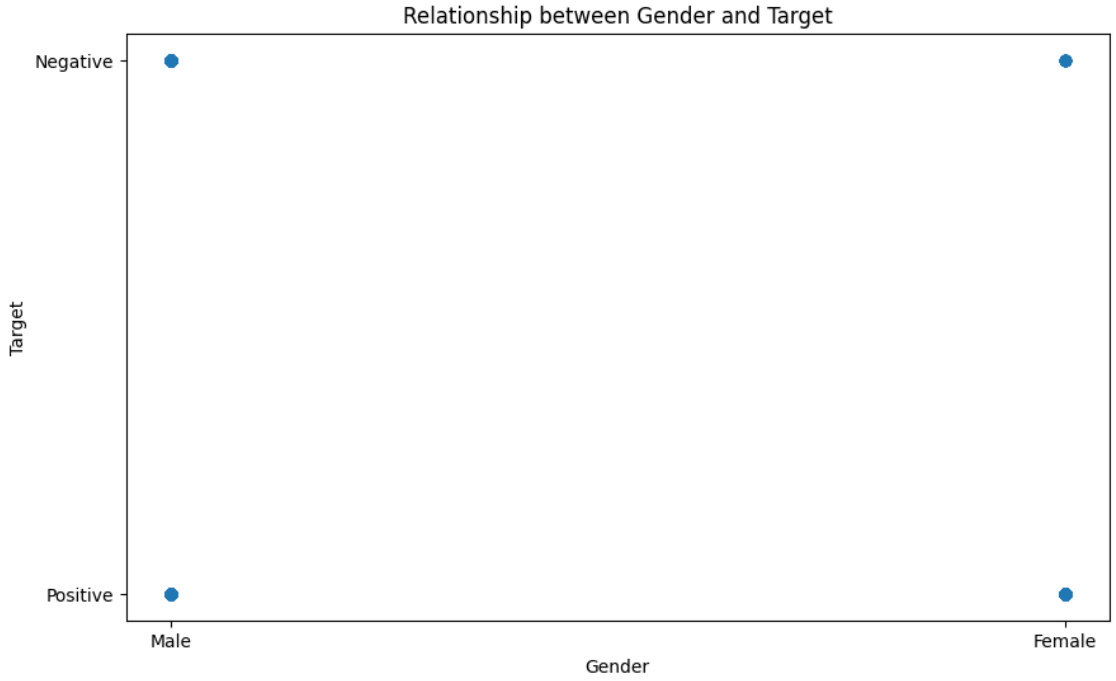
[9]:

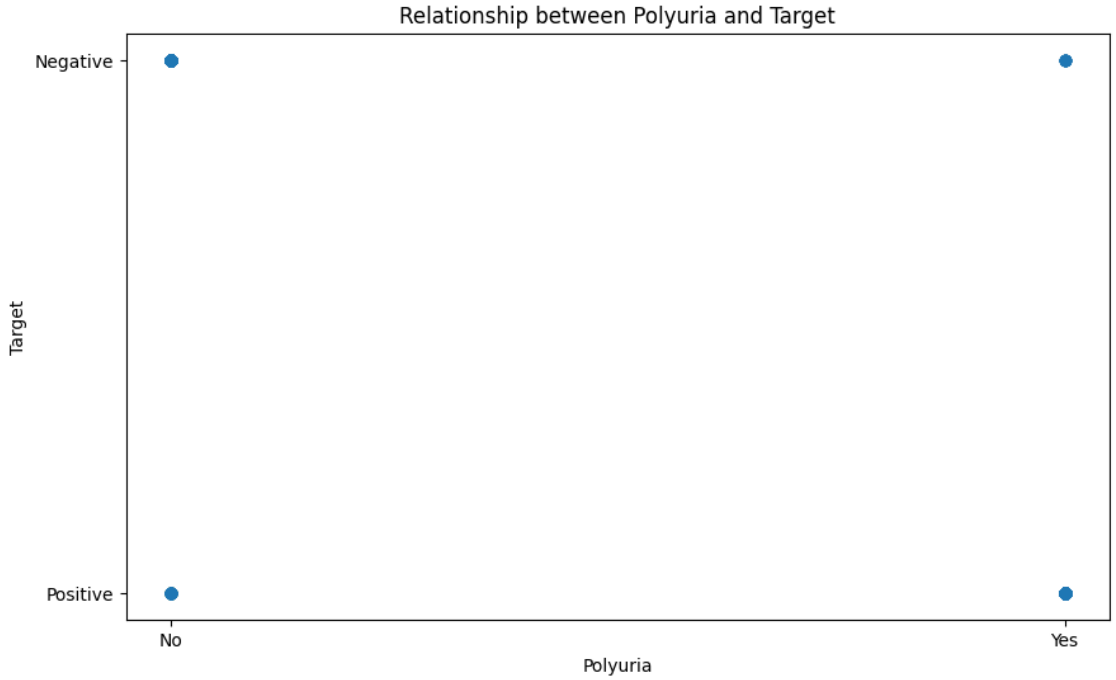
| | Age | Gender | Polyuria | Polydipsia | sudden weight loss | weakness | Polyphagia | Genital thrush | visual blurring | Itching | Irritability | delayed healing |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 40 | Male | No | Yes | No | Yes | No | No | No | Yes | No | Yes |
| 1 | 58 | Male | No | No | No | Yes | No | No | Yes | No | No | No |
| 2 | 41 | Male | Yes | No | No | Yes | Yes | No | No | Yes | No | Yes |
| 3 | 45 | Male | No | No | Yes | Yes | Yes | Yes | No | Yes | No | Yes |
| 4 | 60 | Male | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes |

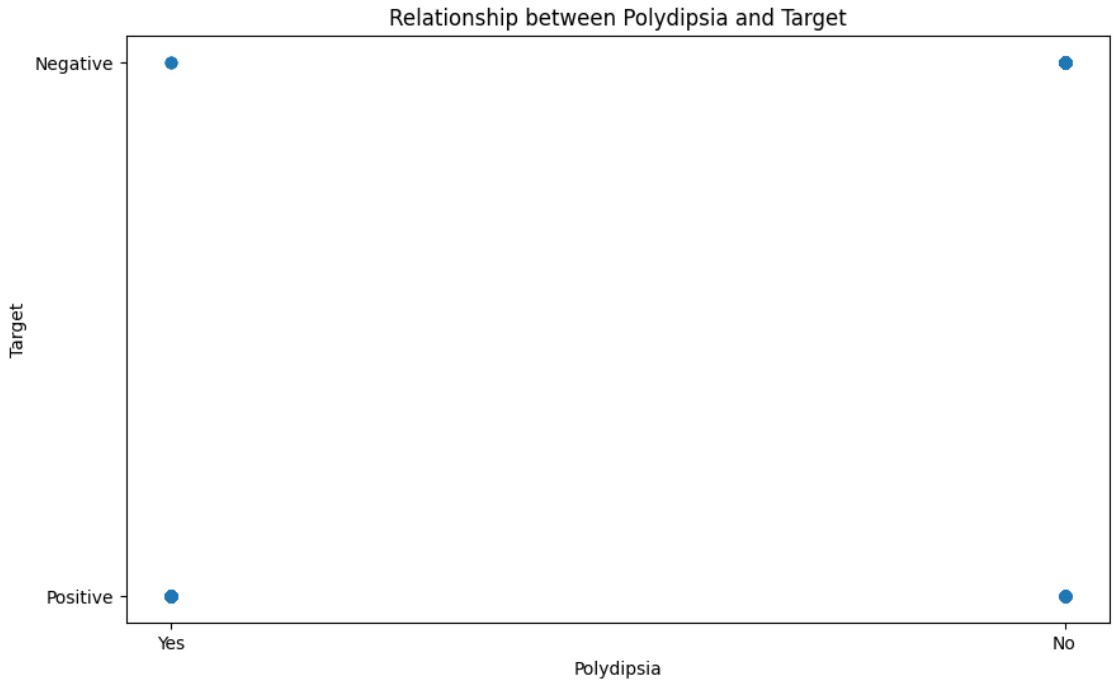
```
2              No              No     Yes             No             Yes
3             Yes              No     Yes             No             Yes
4              No             Yes     Yes            Yes             Yes

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

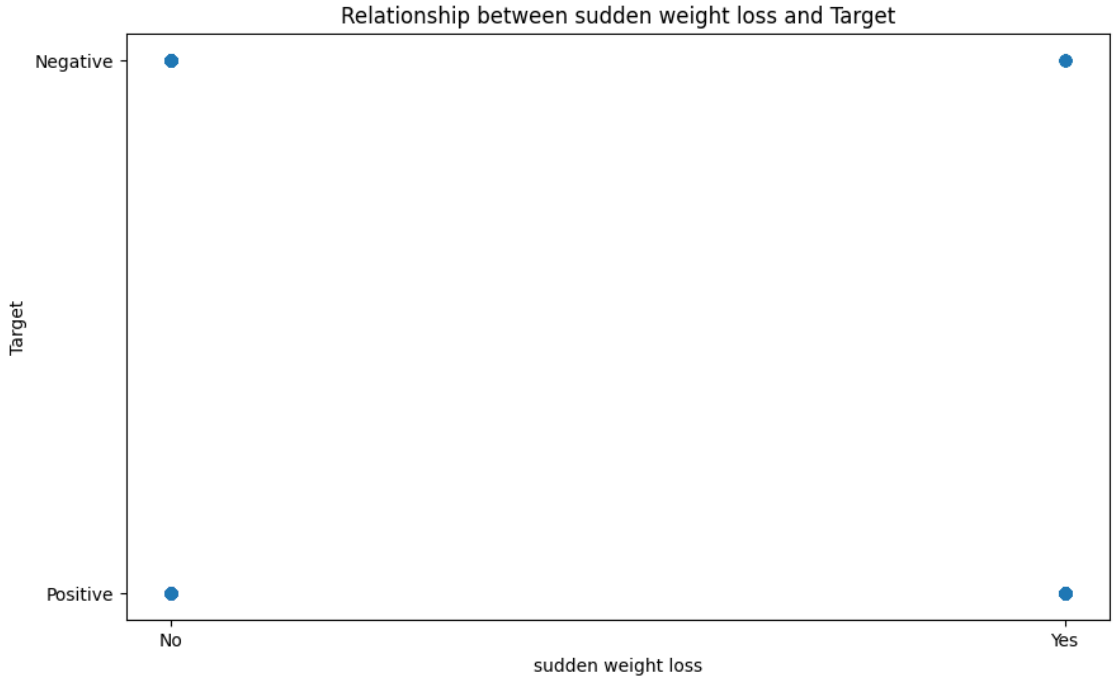
[10]:
```
for column in features.columns:
    plt.figure(figsize=(10, 6))
    plt.scatter(features[column], target_var, alpha=0.5)
    plt.xlabel(column)
    plt.ylabel('Target')
    plt.title(f'Relationship between {column} and Target')
    plt.show()
```

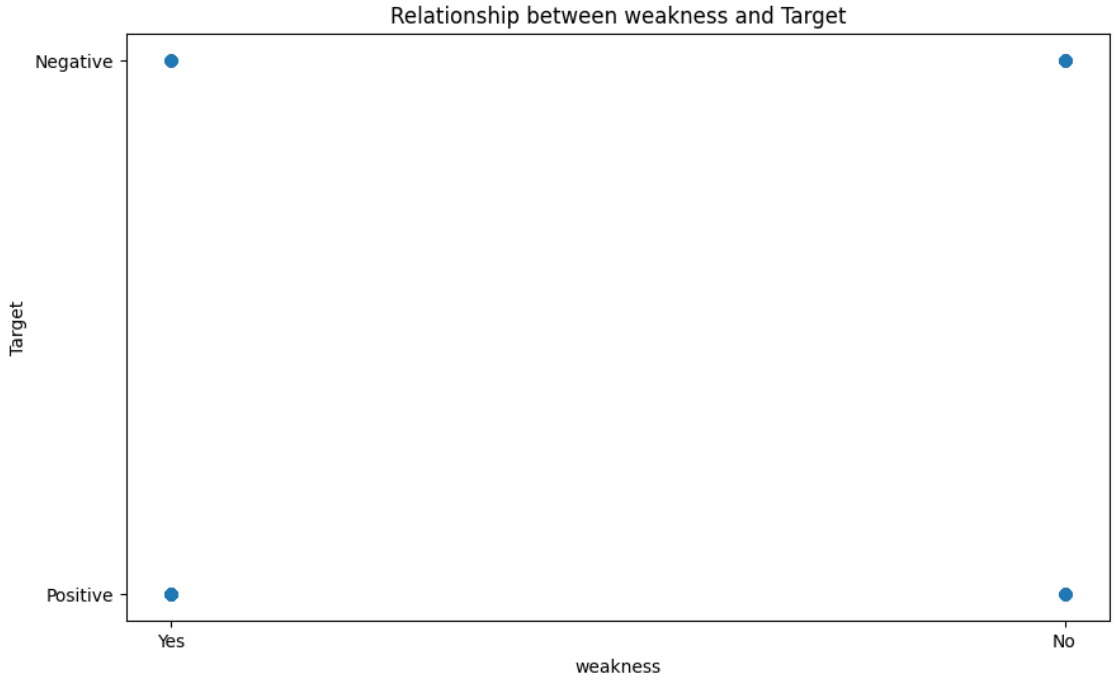


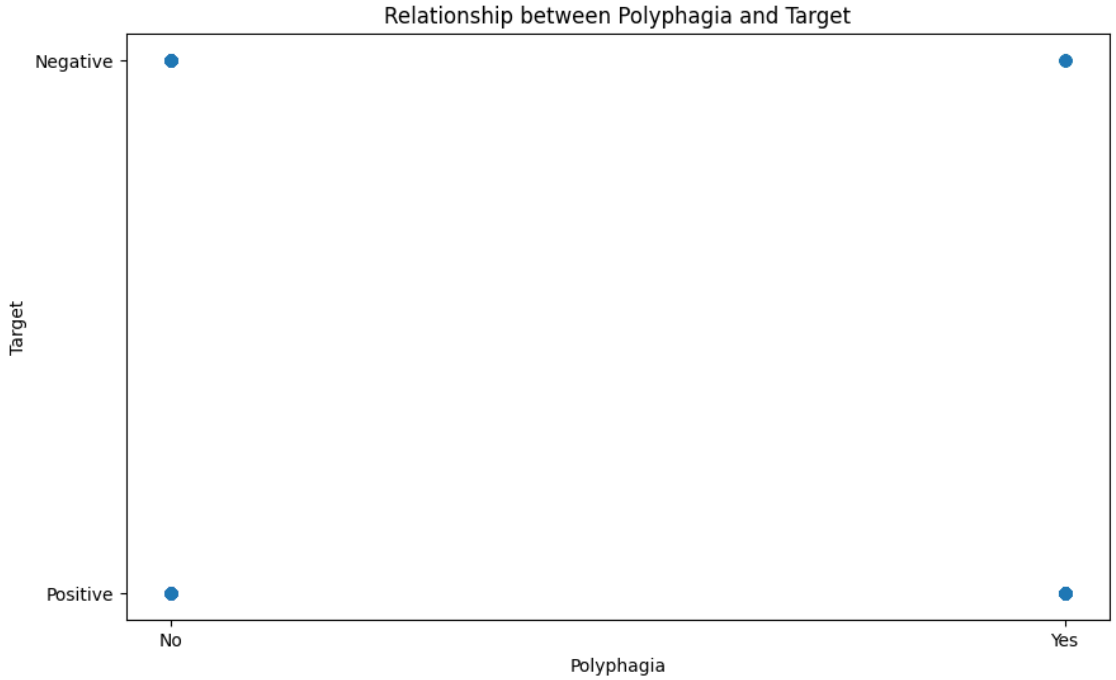

Relationship between Gender and Target
Negative
Positive
Target
Male
Female
Gender

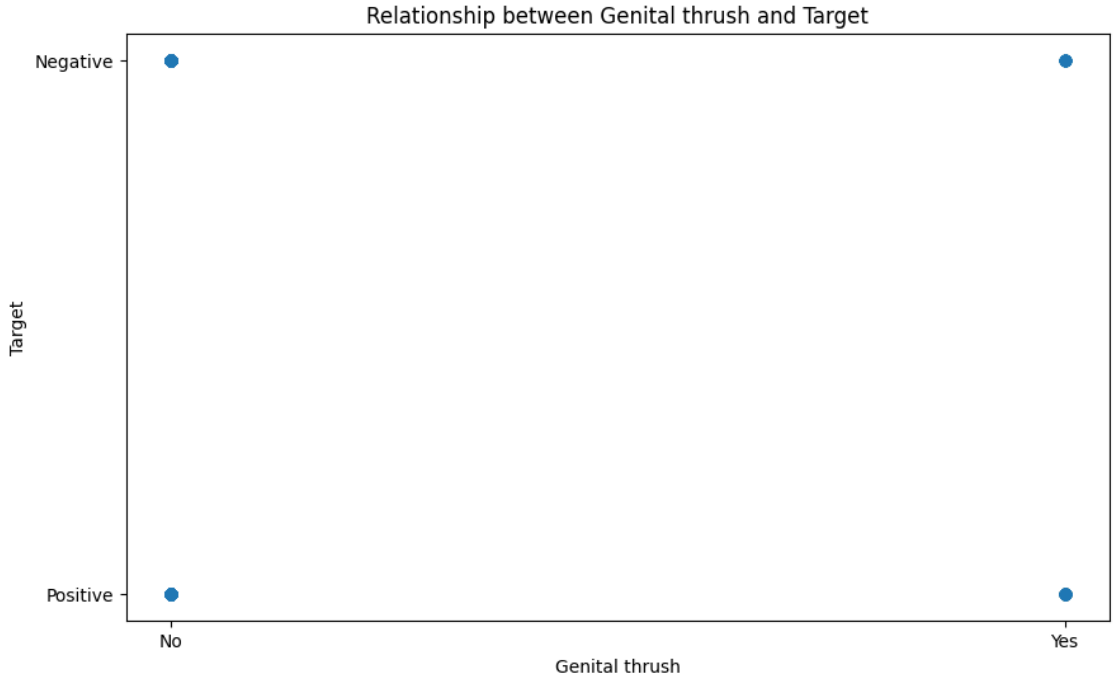

Relationship between Polyuria and Target
Negative
Positive
Target
No
Yes
Polyuria

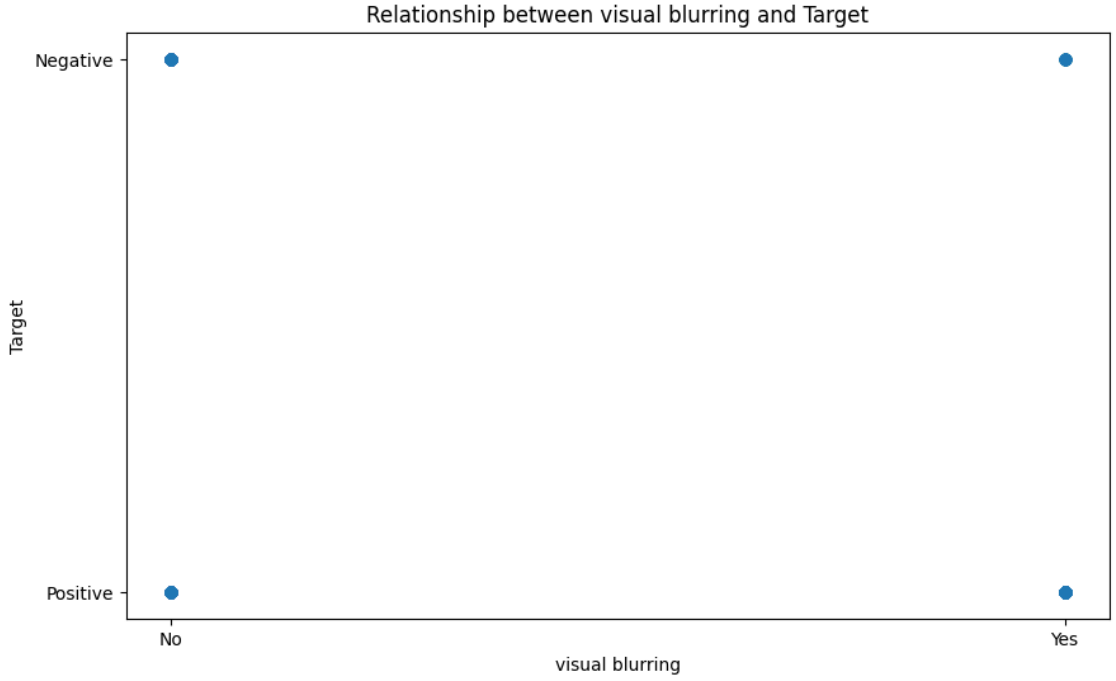

Relationship between Polydipsia and Target
Negative
Positive
Target
Yes
No
Polydipsia

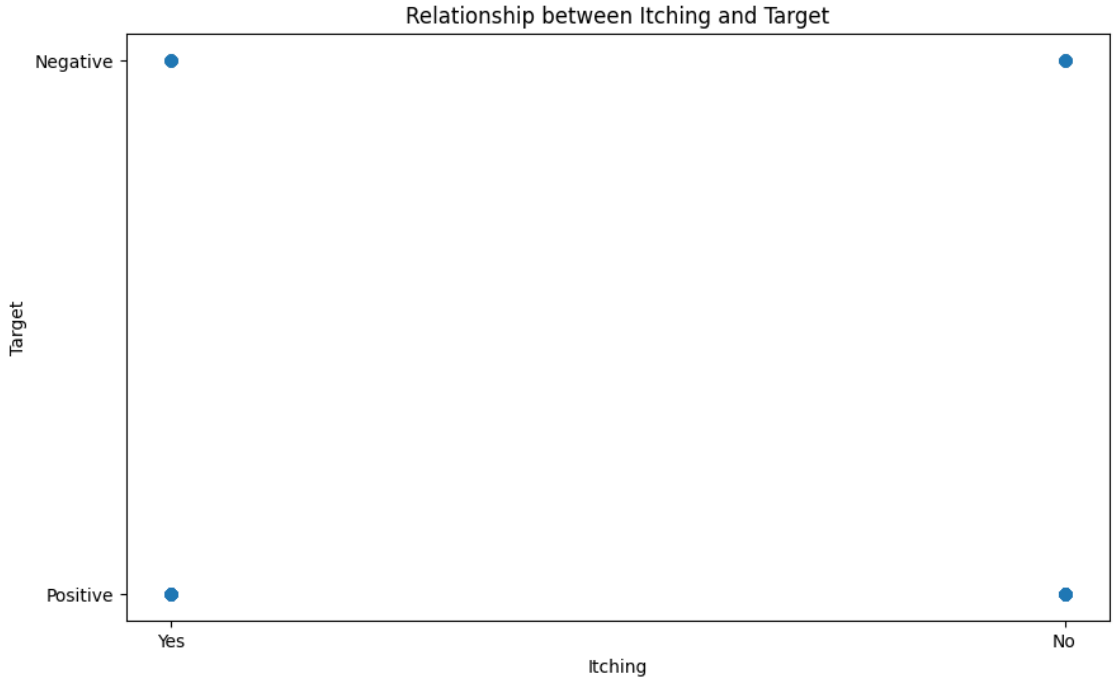

Relationship between sudden weight loss and Target
Negative
Positive
Target
No
Yes
sudden weight loss

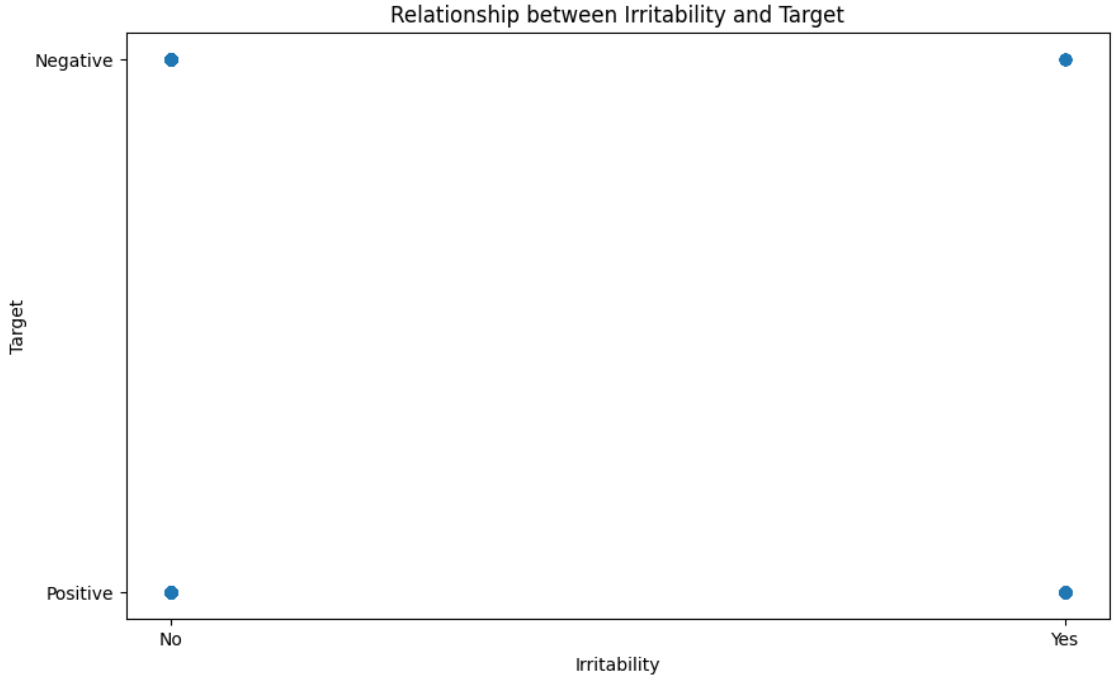

Relationship between weakness and Target
Negative
Positive
Target
Yes
No
weakness

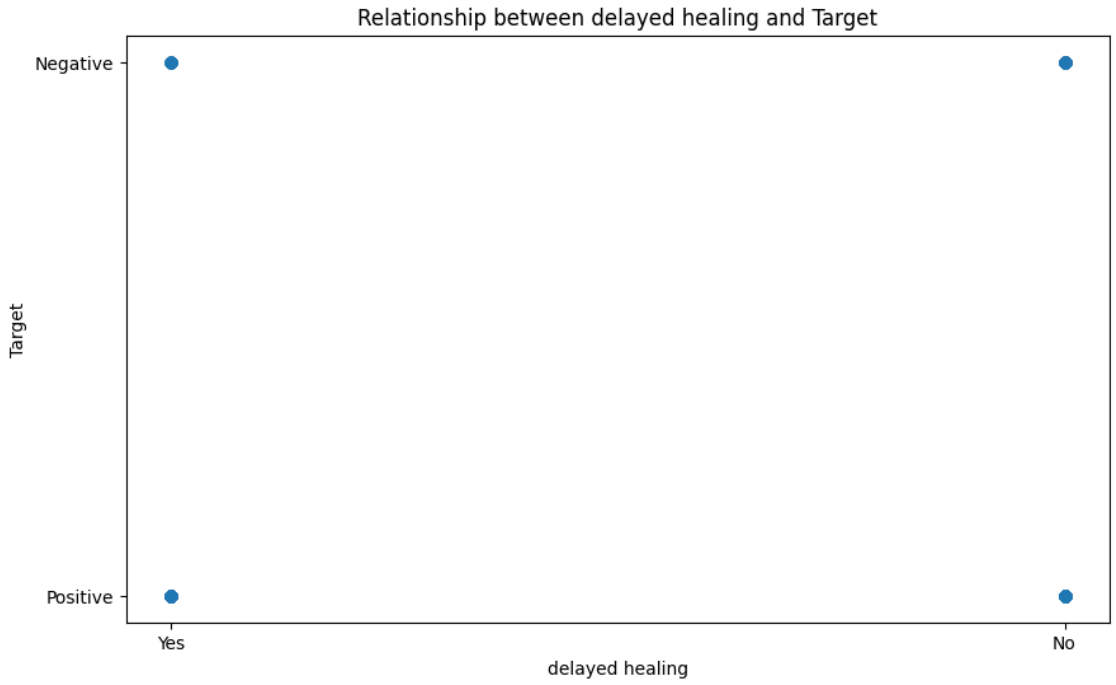

Relationship between Polyphagia and Target
Negative
Positive
Target
No
Yes
Polyphagia

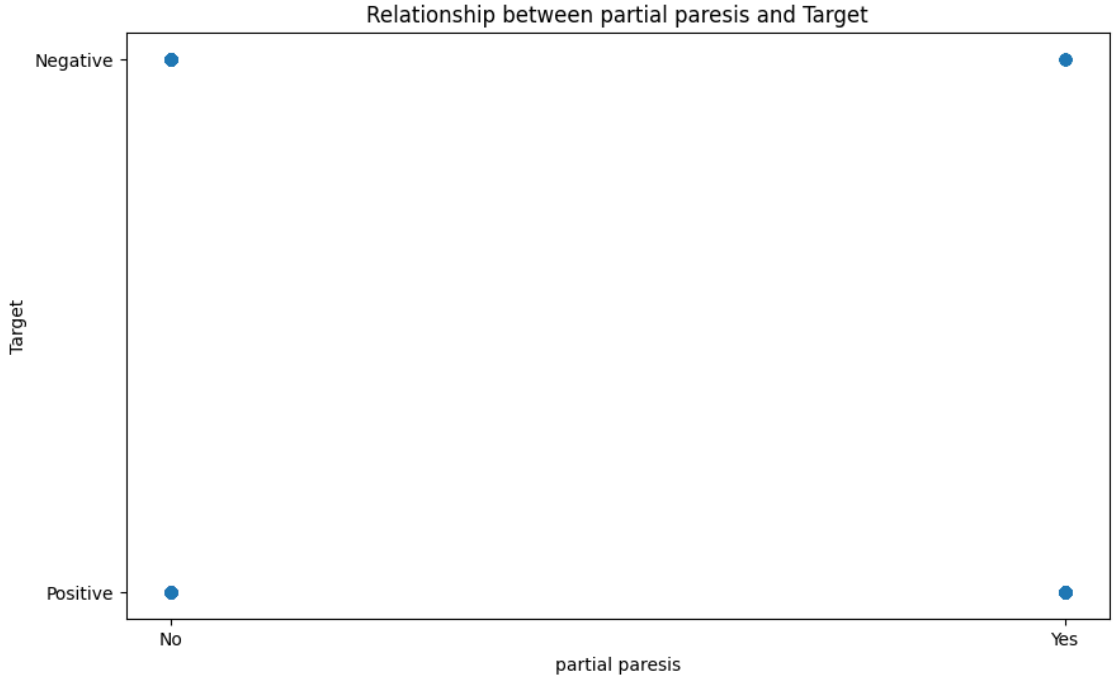

Relationship between Genital thrush and Target
Negative
Positive
Target
No
Yes
Genital thrush

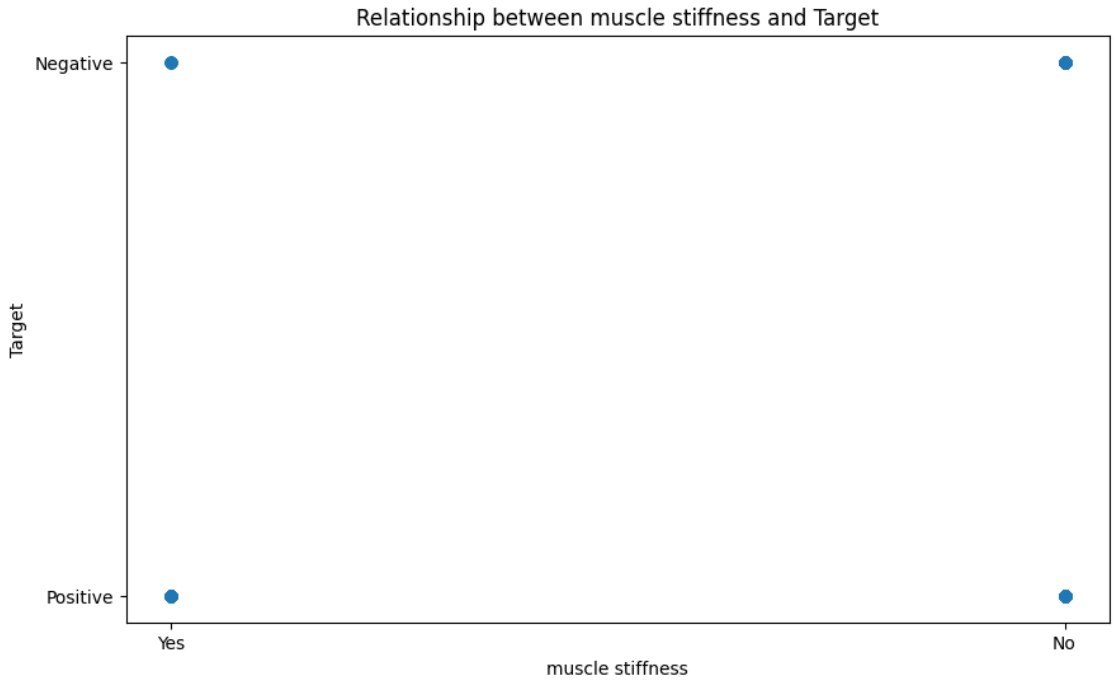

Relationship between visual blurring and Target
Negative
Positive
Target
No
Yes
visual blurring

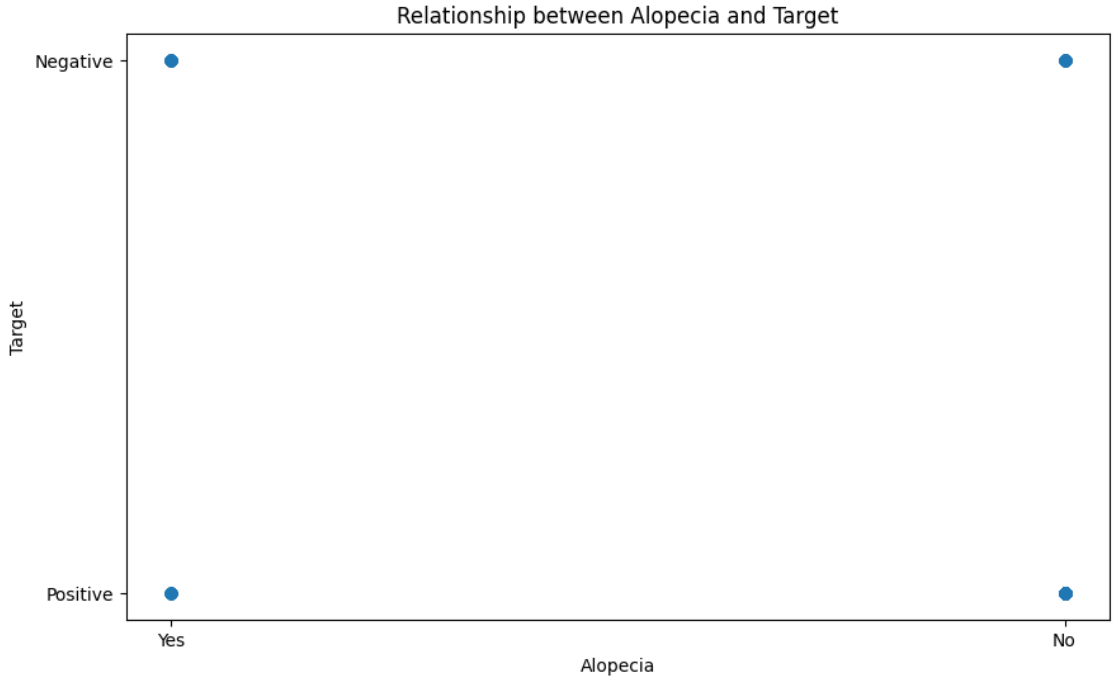

Relationship between Itching and Target
Negative
Positive
Target
Yes
No
Itching

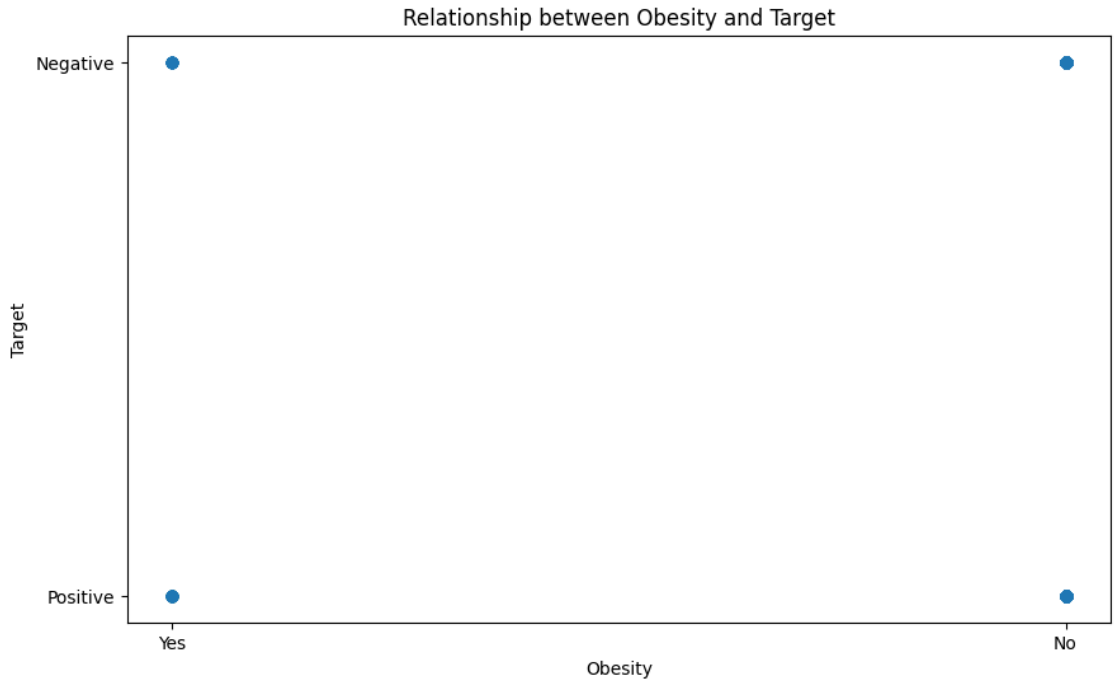

Relationship between Irritability and Target
Negative
Positive
Target
No
Yes
Irritability


Relationship between delayed healing and Target
Negative
Positive
Target
Yes
No
delayed healing


Relationship between partial paresis and Target
Negative
Positive
Target
No
Yes
partial paresis


Relationship between muscle stiffness and Target
Negative
Positive
Target
Yes
No
muscle stiffness


Relationship between Alopecia and Target
Negative
Positive
Target
Yes
No
Alopecia




[11]:
```
one_values = ["Male", "Yes"]
zero_values = ["Female","No"]

for column in features.columns:
    features[column] = features[column].replace(to_replace=one_values, value=1)
    features[column] = features[column].replace(to_replace=zero_values, value=0)
features
```

[11]:

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

| | Polyphagia | Genital thrush | visual blurring | Itching | Irritability | \ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 0 | |
| 1 | 0 | 0 | 1 | 0 | 0 | |
| 2 | 1 | 0 | 0 | 1 | 0 | |

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

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

[520 rows x 16 columns]
```

[12]:
```
one_values = ["Positive"]
zero_values = ["Negative"]

target_var = target_var.replace(to_replace=one_values, value=1)
target_var= target_var.replace(to_replace=zero_values, value=0)
target_var.head()
```

[12]:
```
0    1
1    1
2    1
3    1
4    1
Name: class, dtype: int64
```

[14]:
```
for column in features.columns:
    features[column] = pd.to_numeric(features[column], errors='coerce')
x_normalized = (features - features.min())/(features.max() - features.min())
x_normalized.head()
```

[14]:
```
        Age  Gender  Polyuria  Polydipsia  sudden weight loss  weakness  \
0  0.324324     1.0       0.0         1.0                 0.0       1.0
1  0.567568     1.0       0.0         0.0                 0.0       1.0
2  0.337838     1.0       1.0         0.0                 0.0       1.0
```

```
3  0.391892     1.0       0.0         0.0                1.0       1.0
4  0.594595     1.0       1.0         1.0                1.0       1.0

   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

   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
```