Exam number: Y3857211

1.

Outputs (all rounded to 3 s.f.) -

MLE:
P(Y=0) = 0.648
P(X1=0|Y=0) = 0.623
P(X1=0|Y=1) = 0.386
P(X2=0|Y=0) = 0.475
P(X2=0|Y=1) = 0.767
P(X3=0|Y=0) = 0.812
P(X3=0|Y=1) = 0.455

Bayesian:
P(Y=0) = 0.647
P(X1=0|Y=0) = 0.623
P(X1=0|Y=1) = 0.388
P(X2=0|Y=0) = 0.475
P(X2=0|Y=1) = 0.764
P(X3=0|Y=0) = 0.810
P(X3=0|Y=1) = 0.455

2.

As the data consisted of multiple continuous variables, I needed a multiple linear regression model. Lasso-regression was found to be less effective at predicting r2 values, so I opted to manually feature-select the most important variables at the risk of overfitting.

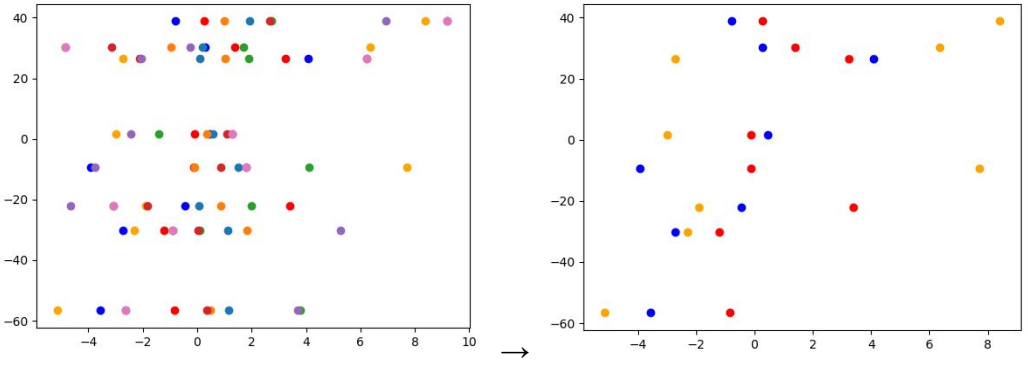
I took the regression coefficients of each variable to test which had the least impact on y (sk-learn doesn't support p-values) and iteratively removed them from the model.

```
      Coefficient
# X1    -0.943629
X2      -4.324248
X3      -1.791904
X4      -2.088912
X5      -2.599120
X6       0.537940
X7       7.963469
X8      -1.096847
X9       1.332416
X10      3.957004
>>>
```

→

```
     Coefficient
X3     -5.343282
X7     12.549229
X10     4.530380
>>>
```

Graphed:

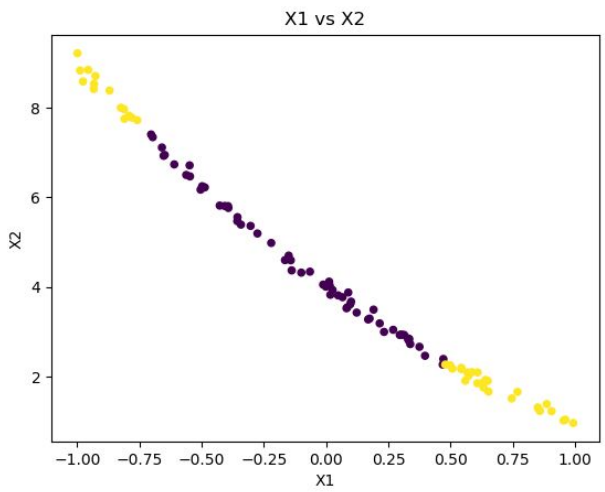


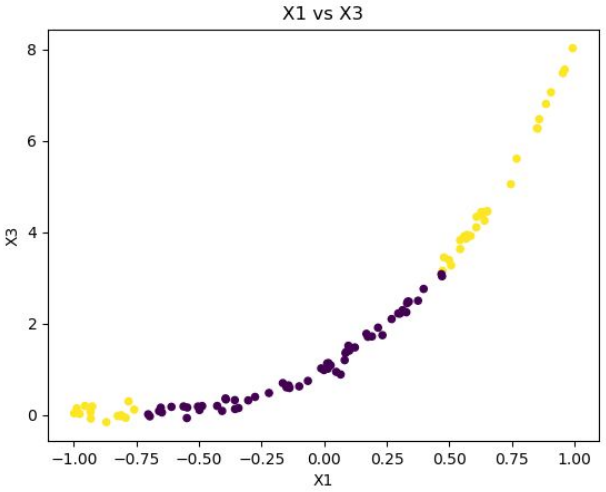
I split the data into test/ training sets and calculated the r2 value to prove it was a good fit by comparing test-y values with predicted-y values.

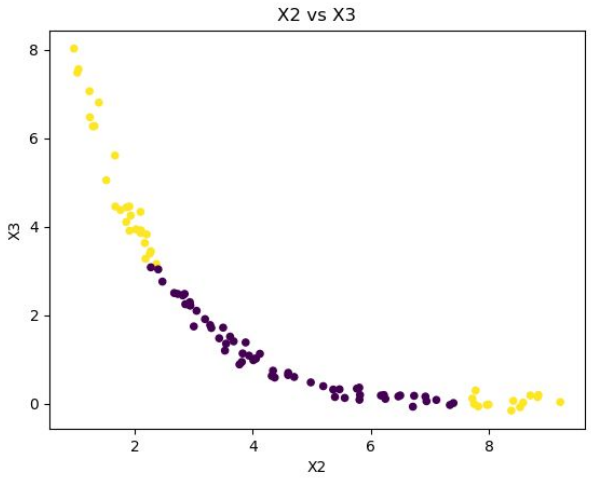
```
      Actual  Predicted
0  -9.223748  -2.456139
6  30.245052  37.222808
7  26.676053  32.454706
r2 value:  0.8658256686273255
```

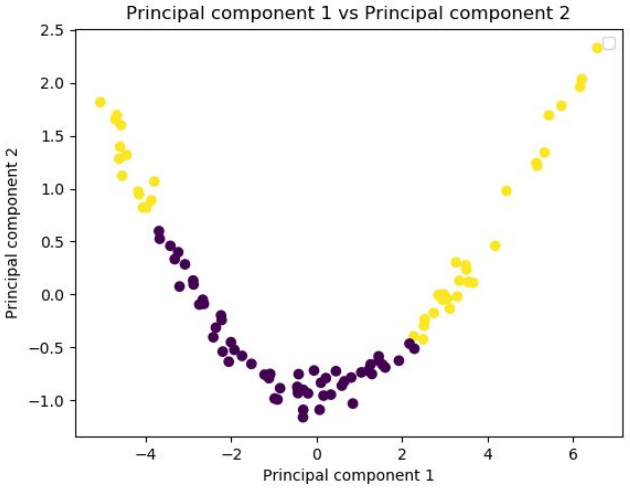
```
       Actual  Predicted
2   38.859968  33.171576
1  -56.634898 -57.000471
5  -30.165856 -30.036073
3  -21.949221 -31.670136
r2 value:  0.9740181616079829
```

3.









Logistic regression would be the most suitable choice of classifier as the predicted Y-values are binary. This will use the X-values to predict a probability that the Y-variable takes a value of 1.

The best pairs of variables will be those that can make a classification with the most certainty. In terms of the graphs, this property can be identified by a distinct separation between colours with no overlap.

The pairs which do this best are (2) X1, X3 and (4) the principal components.

For example, this area of (1) X1, X2 will create some uncertainty surrounding the classification of these values:

And this area of (3) X2, X3: