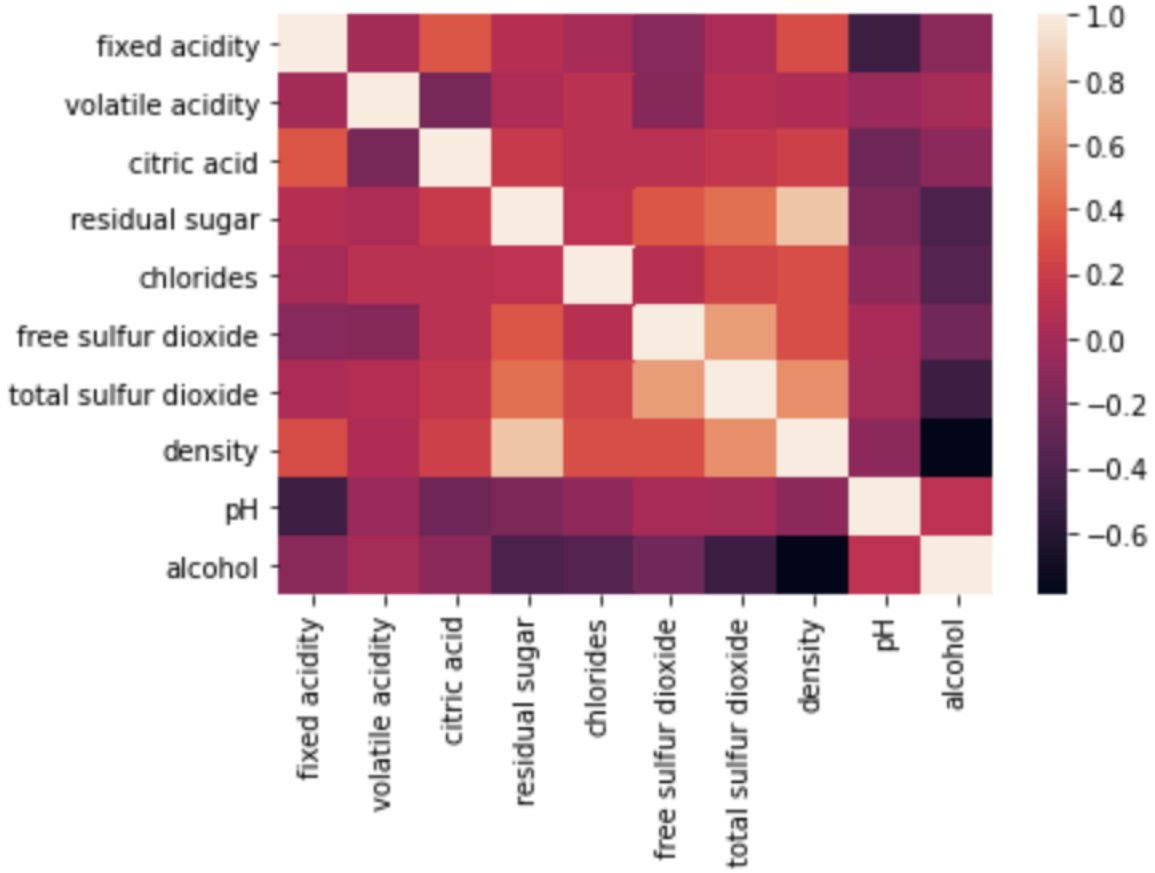
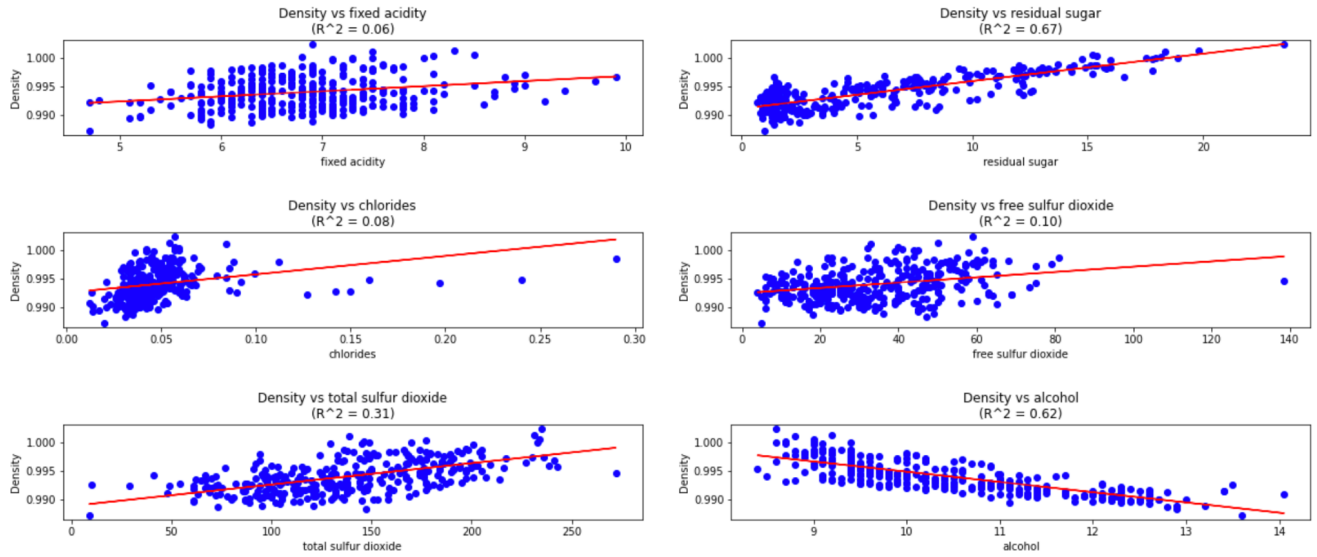
[Q1]
The number of remaining records after duplicate removal : 1617
Heatmap :

```
<AxesSubplot:>
```



[Q2]
R^2 score for “fixed acidity” and “density” : 0.06
R^2 score for “residual sugar” and “density” : 0.67
R^2 score for “chlorides” and “density” : 0.08
R^2 score for “free sulfur dioxide” and “density” : 0.10
R^2 score for “total sulfur dioxide” and “density” : 0.31
R^2 score for “alcohol” and “density” : 0.62
R^2 score for linear combination of the six features and “density” : 0.94
[Q3]



[Q4]

fixed acidity : 3.6849583885179045
volatile acidity : 2.7364663152323674
citric acid : 0.03423537530047388
residual sugar : 250.1380211749334
chlorides : 1.554268192766235
free sulfur dioxide : 0.3348833232370738
total sulfur dioxide : 1573.1849997763782
density : 0.0029298532652383864
pH : 0.3467455894107143
alcohol : 84.76640586655579

[Q5]

Model settings : step_size = [0.1, 0.01, 0.001, 0.0001], r_s = [3456, 4211, 5678],
SGDClassifier(loss = "log", max_iter = 100, ransom_state = r_s, eta0 = step_size,
learning_rate = “constant”, verbose = 0)

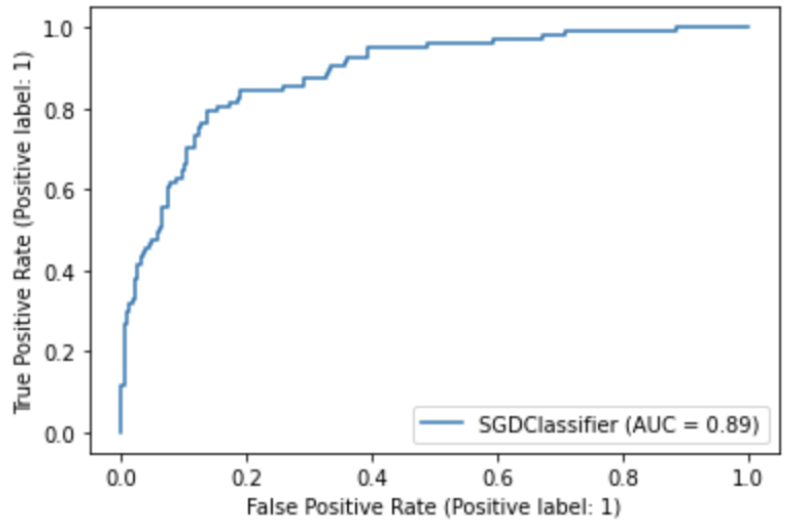
[learning_rate = “constant”, eta0 = 0.1]
Mean training time for step size = 0.1 is 0.0032
Standard deviation for training time for step size = 0.1 is 0.0005
Mean accuracy for step size = 0.1 is 0.8292
Standard deviation for accuracy for step size = 0.1 is 0.0058
Mean F1 score for step size = 0.1 is 0.7311
Standard deviation for F1 score for step size = 0.1 is 0.0136

[learning_rate = “constant”, eta0 = 0.01]
Mean training time for step size = 0.01 is 0.0025
Standard deviation for training time for step size = 0.01 is 0.0004
Mean accuracy for step size = 0.01 is 0.8323
Standard deviation for accuracy for step size = 0.01 is 0.0015
Mean F1 score for step size = 0.01 is 0.7260
Standard deviation for F1 score for step size = 0.01 is 0.0061

[learning_rate = “constant”, eta0 = 0.001]

Mean training time for step size = 0.001 is 0.0037
Standard deviation for training time for step size = 0.001 is 0.0005
Mean accuracy for step size = 0.001 is 0.8364
Standard deviation for accuracy for step size = 0.001 is 0.0000
Mean F1 score for step size = 0.001 is 0.7389
Standard deviation for F1 score for step size = 0.001 is 0.0000

[learning_rate = “constant”, eta0 = 0.0001]
Mean training time for step size = 0.0001 is 0.0077
Standard deviation for training time for step size = 0.0001 is 0.0003
Mean accuracy for step size = 0.0001 is 0.8117
Standard deviation for accuracy for step size = 0.0001 is 0.0000
Mean F1 score for step size = 0.0001 is 0.7081
Standard deviation for F1 score for step size = 0.0001 is 0.0000
[Q6]



AUC value : 0.8861892002361597

One advantage presented by ROC curves is that viewing the ROC curve lets you see the tradeoff between sensitivity and specificity for all possible thresholds rather than just the one that was chosen by the modeling technique. Different classification objectives might make one point on the curve more suitable for one task and another more suitable for a different task, so looking at the ROC curve is a way to assess the model independent of the choice of a threshold.
[Q7]
Model settings : num_of_units = [1, 2, 4, 8, 16, 32, 64, 128], r_s = [3456, 4211, 5678], MLPClassifier(hidden_layer_sizes = (num_of_units,), random_state = r_s, max_iter = 500, early_stopping = True)

[number of hidden layer units : 1]
Mean training time for 1 hidden units is 0.04
Standard deviation for training time for 1 hidden units is 0.00
Mean accuracy for 1 hidden units is 0.57
Standard deviation for accuracy for 1 hidden units is 0.19

Mean F1 score for 1 hidden units is 0.15
Standard deviation for F1 score for 1 hidden units is 0.22

[number of hidden layer units : 2]
Mean training time for 2 hidden units is 0.10
Standard deviation for training time for 2 hidden units is 0.05
Mean accuracy for 2 hidden units is 0.45
Standard deviation for accuracy for 2 hidden units is 0.11
Mean F1 score for 2 hidden units is 0.50
Standard deviation for F1 score for 2 hidden units is 0.04

[number of hidden layer units : 4]
Mean training time for 4 hidden units is 0.08
Standard deviation for training time for 4 hidden units is 0.04
Mean accuracy for 4 hidden units is 0.58
Standard deviation for accuracy for 4 hidden units is 0.21
Mean F1 score for 4 hidden units is 0.57
Standard deviation for F1 score for 4 hidden units is 0.08

[number of hidden layer units : 8]
Mean training time for 8 hidden units is 0.14
Standard deviation for training time for 8 hidden units is 0.03
Mean accuracy for 8 hidden units is 0.84
Standard deviation for accuracy for 8 hidden units is 0.00
Mean F1 score for 8 hidden units is 0.74
Standard deviation for F1 score for 8 hidden units is 0.01

[number of hidden layer units : 16]
Mean training time for 16 hidden units is 0.11
Standard deviation for training time for 16 hidden units is 0.03
Mean accuracy for 16 hidden units is 0.83
Standard deviation for accuracy for 16 hidden units is 0.01
Mean F1 score for 16 hidden units is 0.72
Standard deviation for F1 score for 16 hidden units is 0.01

[number of hidden layer units : 32]
Mean training time for 32 hidden units is 0.11
Standard deviation for training time for 32 hidden units is 0.01
Mean accuracy for 32 hidden units is 0.84
Standard deviation for accuracy for 32 hidden units is 0.01
Mean F1 score for 32 hidden units is 0.73
Standard deviation for F1 score for 32 hidden units is 0.01

[number of hidden layer units : 64]
Mean training time for 64 hidden units is 0.09
Standard deviation for training time for 64 hidden units is 0.02
Mean accuracy for 64 hidden units is 0.83
Standard deviation for accuracy for 64 hidden units is 0.01

Mean F1 score for 64 hidden units is 0.71
Standard deviation for F1 score for 64 hidden units is 0.02

[number of hidden layer units : 128]
Mean training time for 128 hidden units is 0.09
Standard deviation for training time for 128 hidden units is 0.01
Mean accuracy for 128 hidden units is 0.85
Standard deviation for accuracy for 128 hidden units is 0.01
Mean F1 score for 128 hidden units is 0.74
Standard deviation for F1 score for 128 hidden units is 0.03

[Q8]

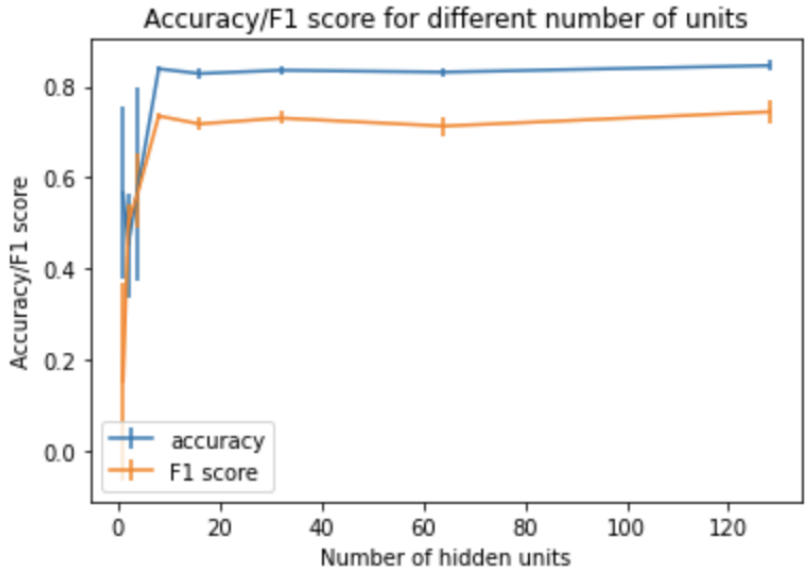
The mean training time for the best logistic regression is 0.0048, and the mean training time for the best neural network model is 0.09.

The mean accuracy score for the best logistic regression is 0.8364, and the mean accuracy score for the best neural network model is 0.85.

The mean F1 score for the best logistic regression is 0.7389, and the mean F1 score for the best neural network model is 0.74.

Hence, for mean training time, the best logistic regression model is shorter, while for accuracy score, the best neural network model is higher, and the mean F1 score for the best neural network model is slightly higher than that of the best logistic regression model.

[Q9]



A possible reason for the gap between the accuracy and the F1 score is that F1 score is the harmonic mean of precision and recall, while the accuracy is high, the recall may be low, which means that the number of correctly identified positive cases is low.

[Q10]

The accuracy, and F1 score all have a trend of increasing while the number of hidden units increase. A reason for it is that as more hidden layer units are used, it may have a better representing power, thus, the accuracy and F1 score is increasing.

[Q11]

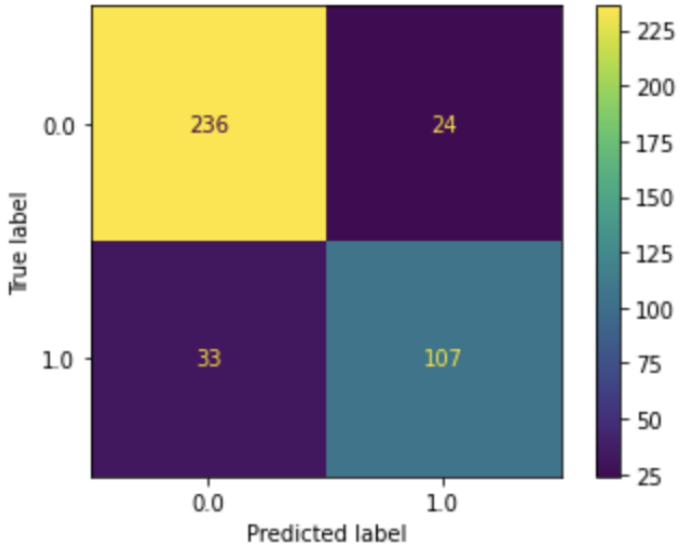
1. {activation = “logistic”, solver = “sgd”, learning_rate_init = 0.1, random_state = 4211, max_iter = 500, early_stopping = True}
2. {activation = “logistic”, solver = “adam”, learning_rate_init = 0.1, random_state = 4211, max_iter = 500, early_stopping = True}
3. {activation = “logistic”, solver = “sgd”, learning_rate_init = 0.01, random_state = 4211, max_iter = 500, early_stopping = True}
4. {activation = “logistic”, solver = “adam”, learning_rate_init = 0.01, random_state = 4211, max_iter = 500, early_stopping = True}
5. {activation = “tanh”, solver = “sgd”, learning_rate_init = 0.1, random_state = 4211, max_iter = 500, early_stopping = True}
6. {activation = “tanh”, solver = “adam”, learning_rate_init = 0.01, random_state = 4211, max_iter = 500, early_stopping = True}
7. {activation = “relu”, solver = “sgd”, learning_rate_init = 0.1, random_state = 4211, max_iter = 500, early_stopping = True}
8. {activation = “relu”, solver = “adam”, learning_rate_init = 0.1, random_state = 4211, max_iter = 500, early_stopping = True}
9. {activation = “relu”, solver = “sgd”, learning_rate_init = 0.01, random_state = 4211, max_iter = 500, early_stopping = True}
10. {activation = “relu”, solver = “adam”, learning_rate_init = 0.01, random_state = 4211, max_iter = 500, early_stopping = True}

[Q12]

1. {'activation': 'relu', 'early_stopping': True, 'learning_rate_init': 0.1, 'max_iter': 500, 'random_state': 4211, 'solver': 'sgd'}
   Mean validation accuracy : 0.854
   Standard deviation of validation accuracy : 0.028
2. {'activation': 'relu', 'early_stopping': True, 'learning_rate_init': 0.01, 'max_iter': 500, 'random_state': 4211, 'solver': 'adam'}
   Mean validation accuracy : 0.850
   Standard deviation of validation accuracy : 0.022
3. {'activation': 'logistic', 'early_stopping': True, 'learning_rate_init': 0.1, 'max_iter': 500, 'random_state': 4211, 'solver': 'adam'}
   Mean validation accuracy : 0.846
   Standard deviation of validation accuracy : 0.026

[Q13]

Accuracy : 0.8575
F1 score : 0.7896678966789668
Confusion matrix :



[Q14]

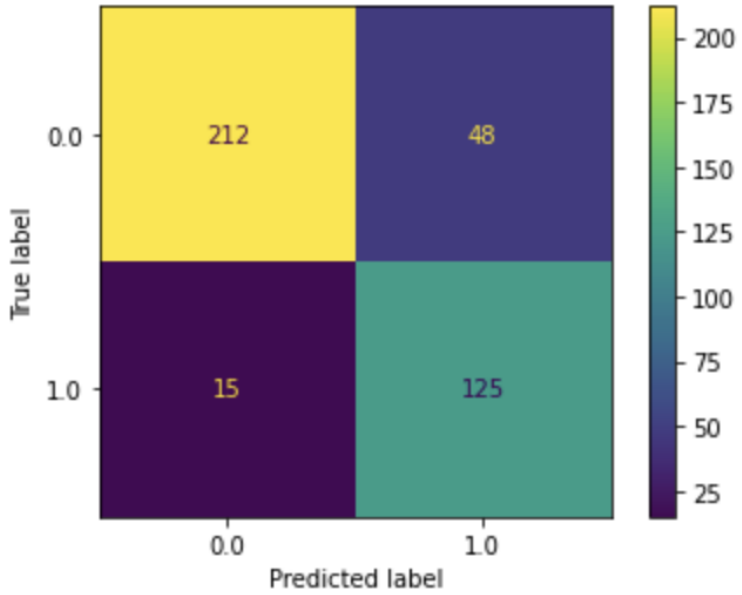
1. Collect more data
2. Undersampling

[Q15]

Accuracy : 0.8425

F1 score : 0.7987220447284346

Confusion matrix :



[Q16]

Accuracy is lower than that in Section 7.1, however, the F1 score is higher than that in Section 7.1. For the confusion matrix, the confusion matrix for oversampling has more true negatives and false positives, while the confusion matrix in Section 7.1 has more true positives and false positives.