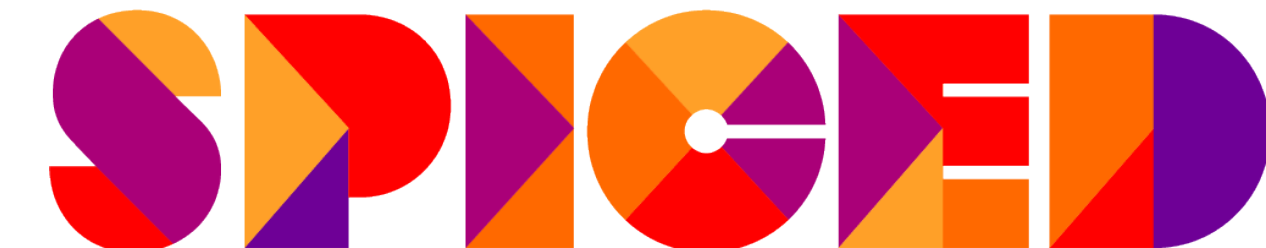


CONFUSION MATRIX

Reality	No Fraud	Fraud
	No Fraud	Fraud
No Fraud		
Fraud		



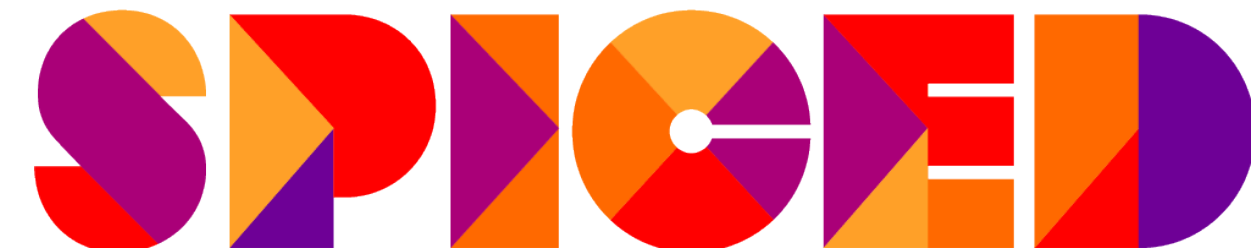
RECALL

Important for many imbalanced datasets, e.g. cancer detection or fraud detection -> should be as high as possible.

Did we catch all positives?

Reality	No Fraud	Fraud
	No Fraud	Fraud
No Fraud	True Negative	False Positive
Fraud	False Negative	True Positive

Recall
 $TP/(TP+FN)$

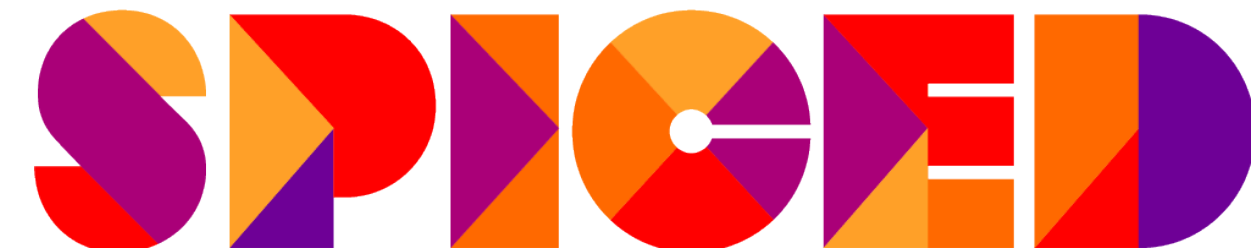


PRECISION

Of all positive predictions (fraud), how many are correct?

Precision
 $TP/(TP+FP)$

Reality	No Fraud	Fraud
	True Negative	False Positive
Fraud	False Negative	True Positive
	No Fraud	Fraud
Prediction		



Metrics Overview

Accuracy: $TP+TN / TP+TN+FP+FN$

Precision: $TP / TP+FP$

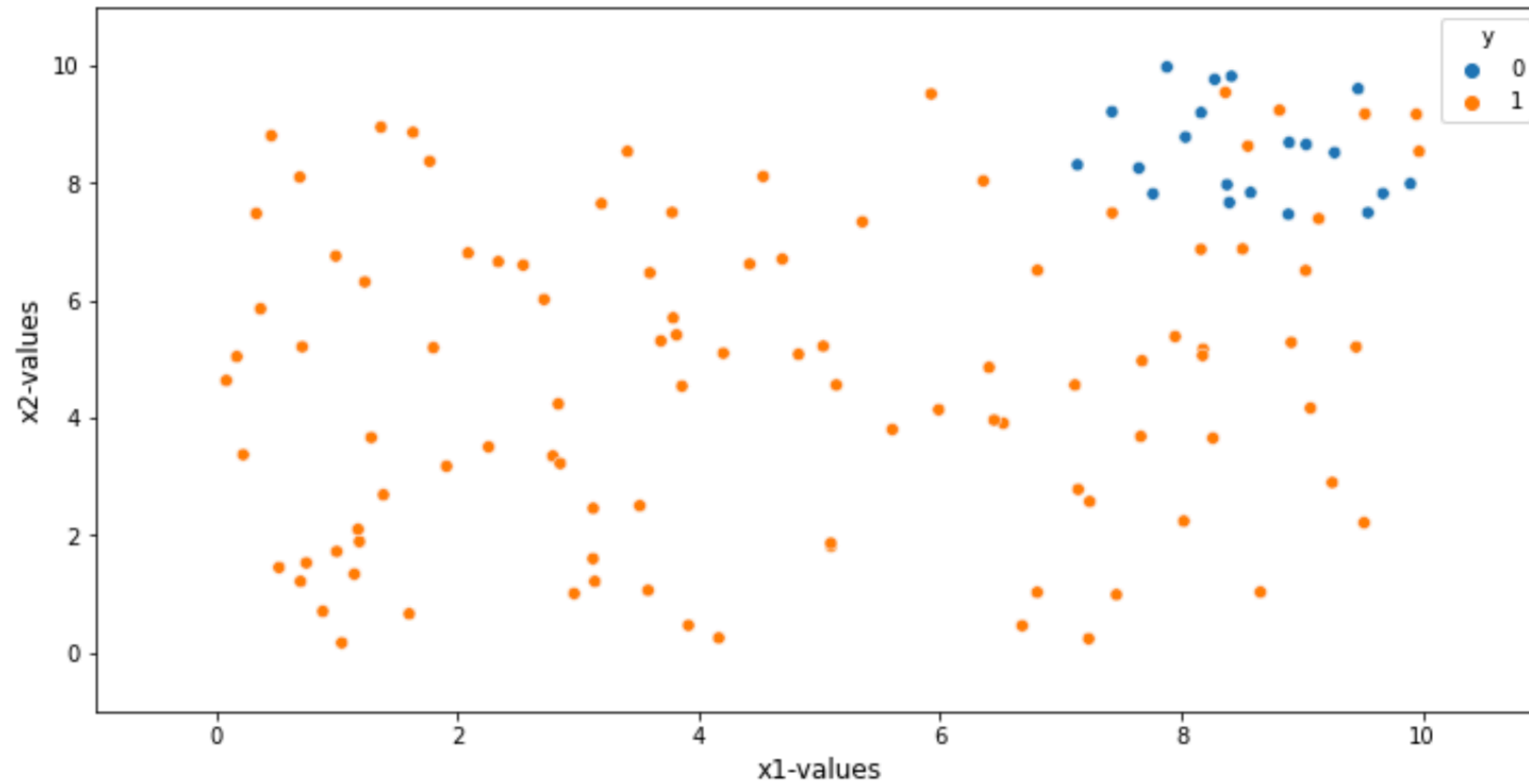
Recall: $TP / TP+FN$

F1-Score: $2*(Precision*Recall) / Precision + Recall$
(Harmonic mean of precision and recall)



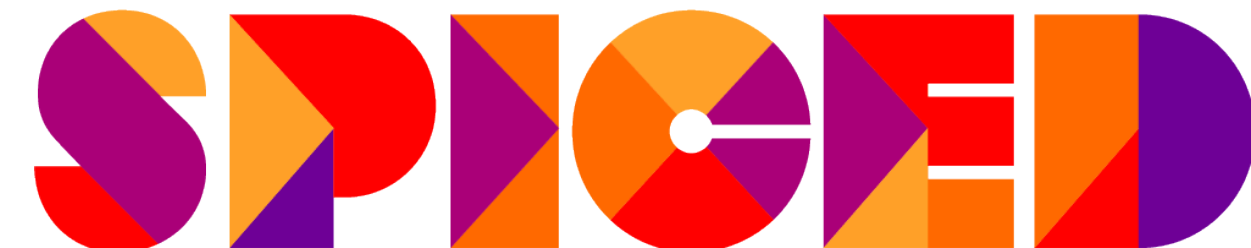
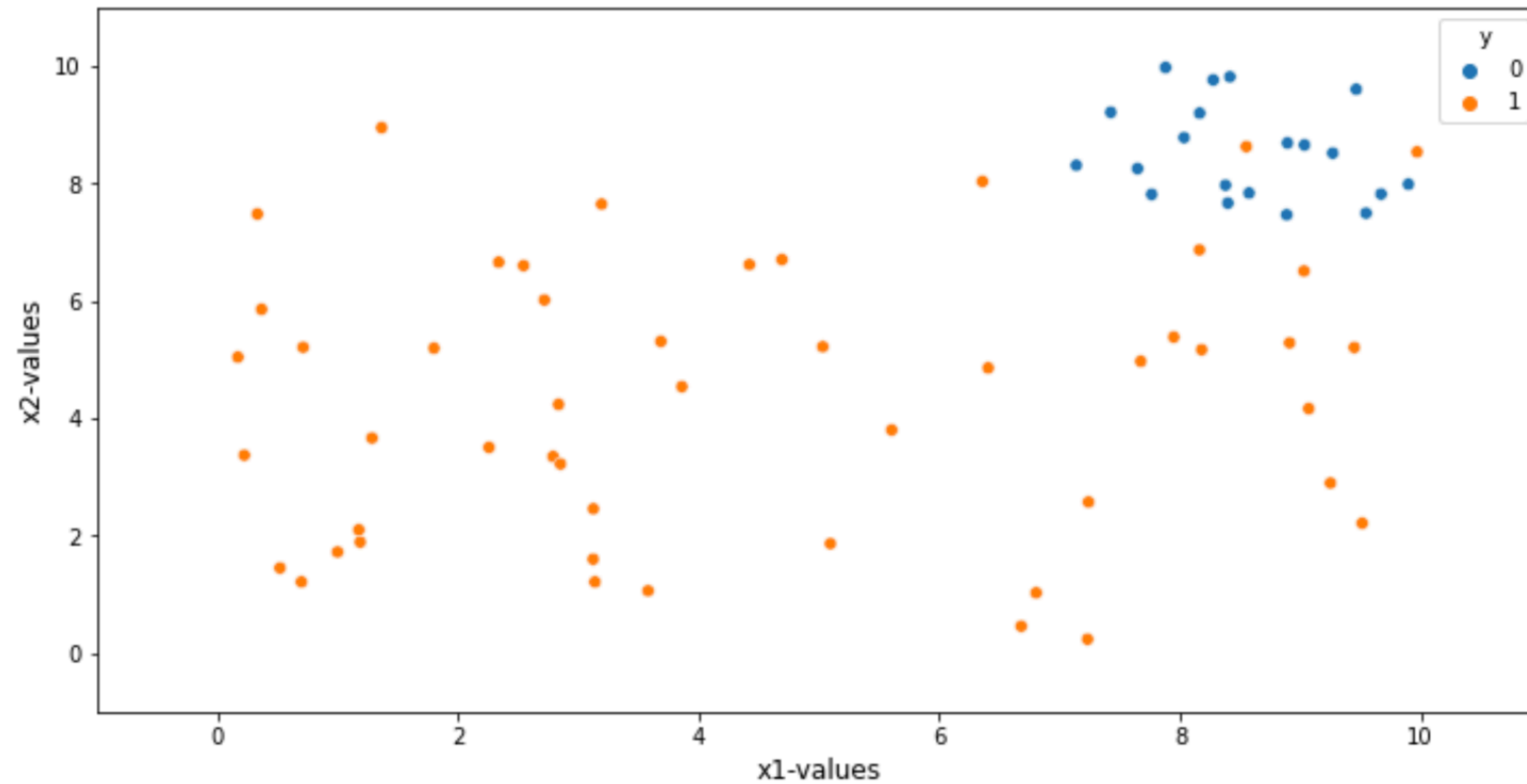
Random Undersampling

Original unbalanced data



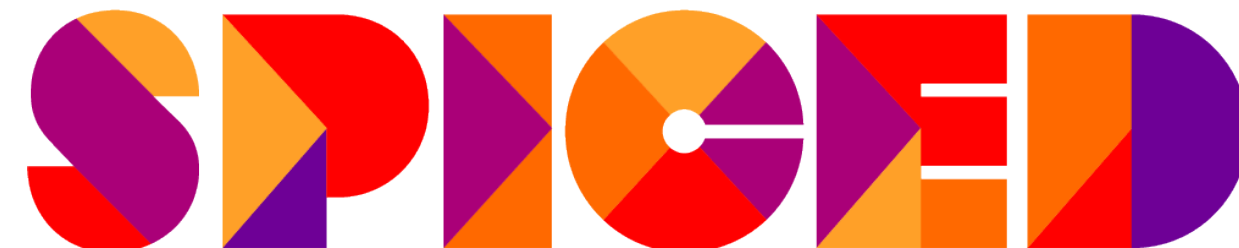
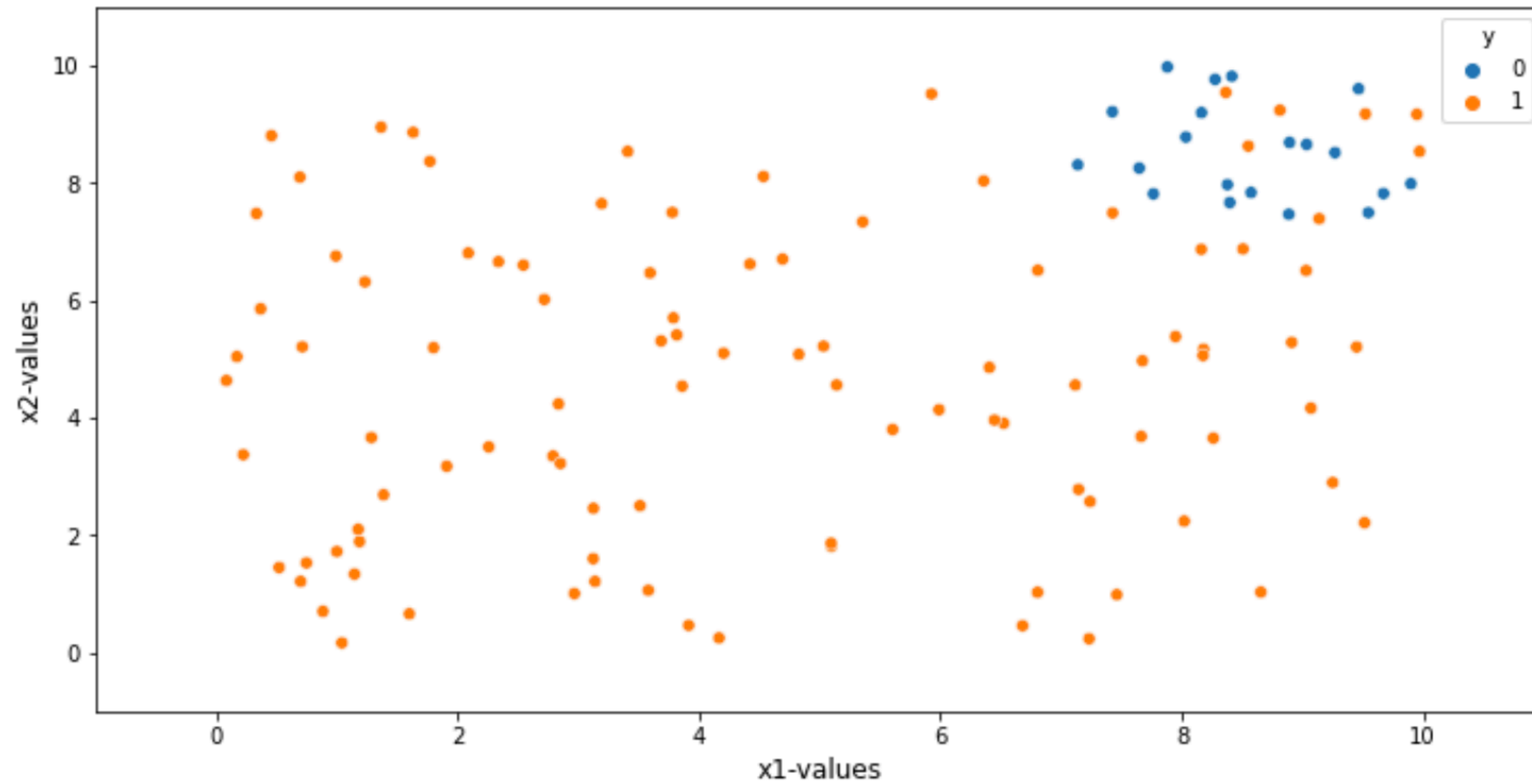
Random Undersampling

Randomly undersampled data



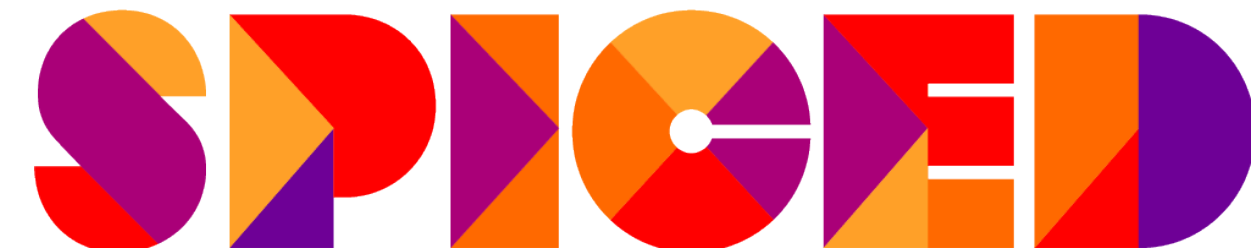
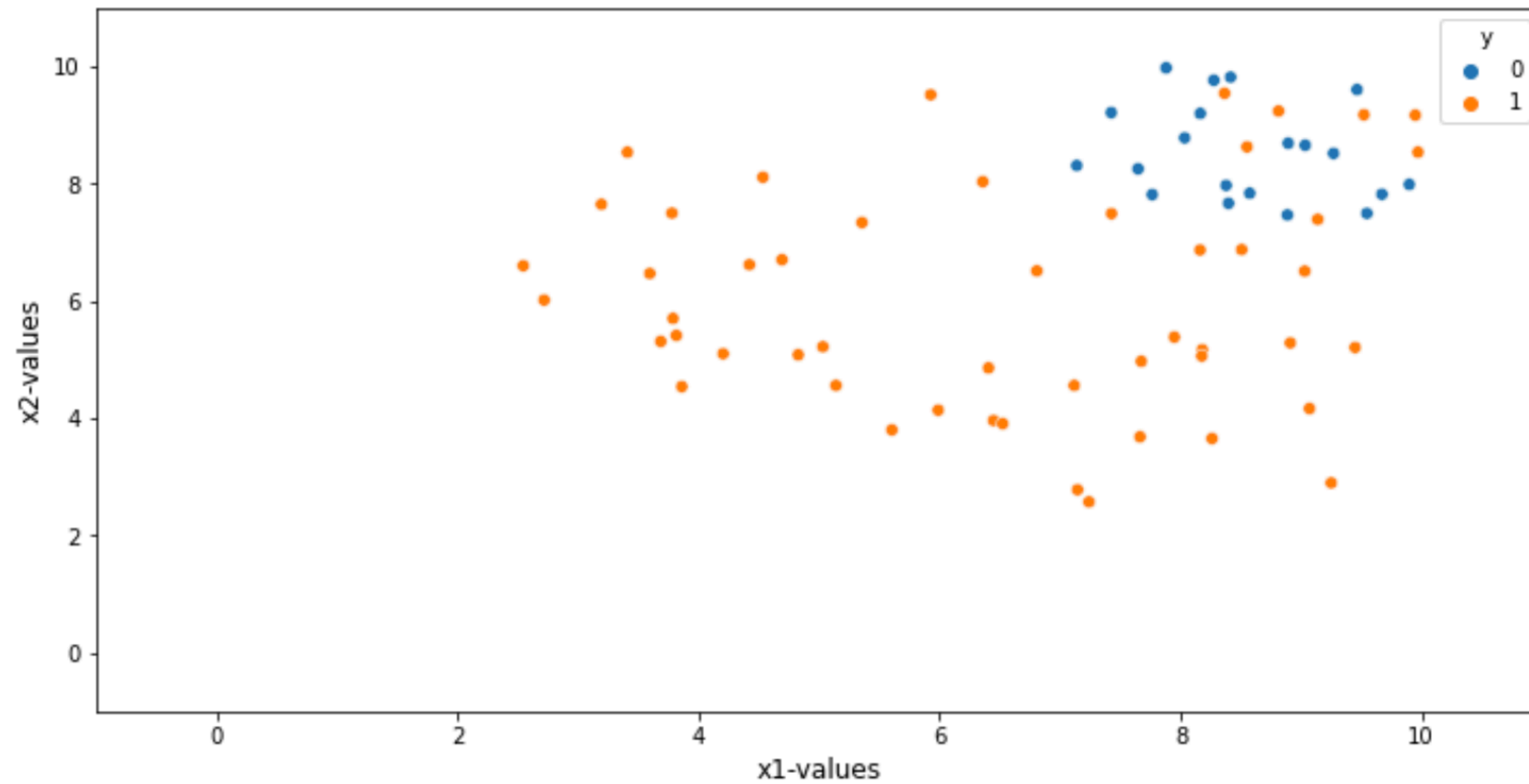
NearMiss Undersampling

Original unbalanced data



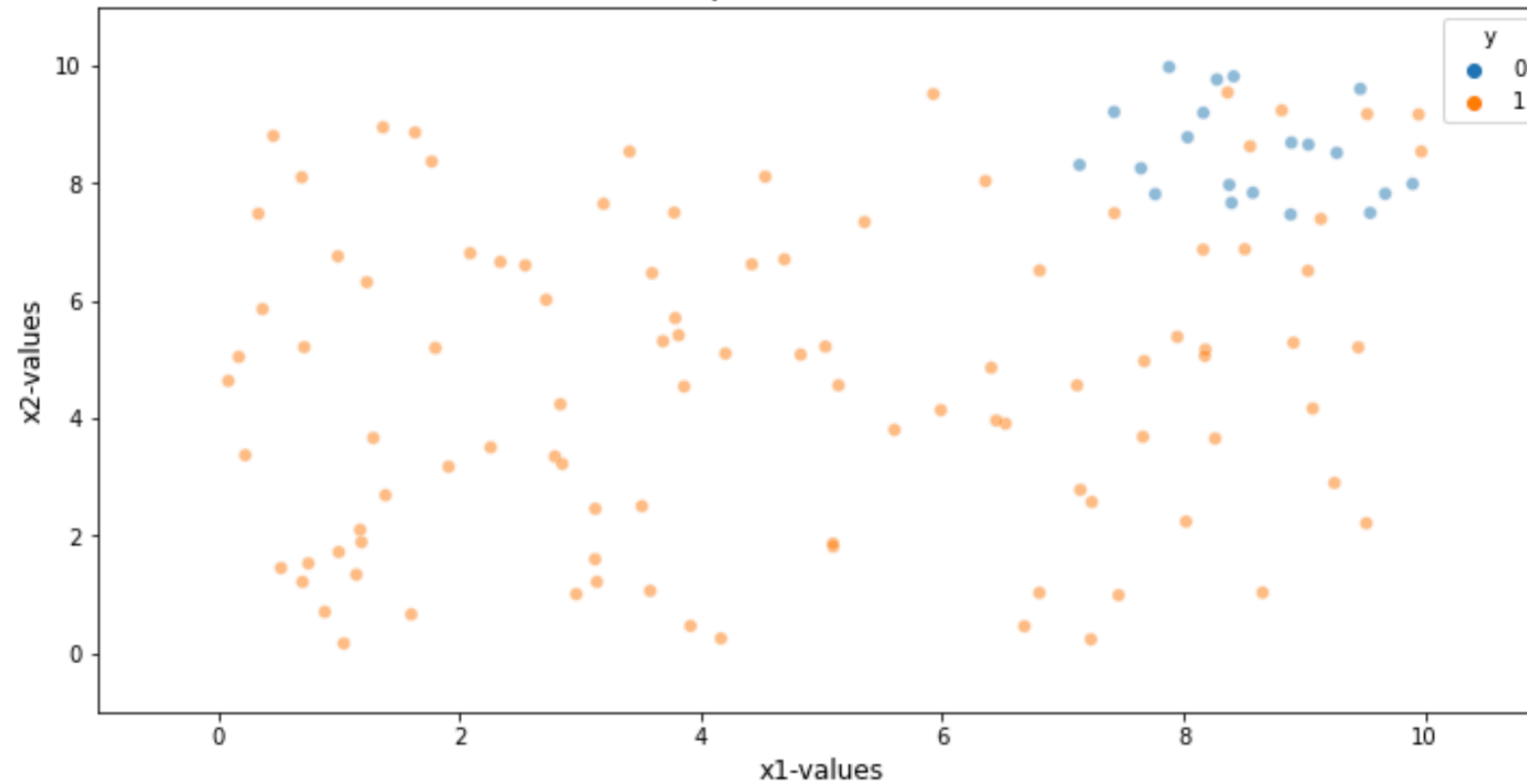
NearMiss Undersampling

With NearMiss undersampling



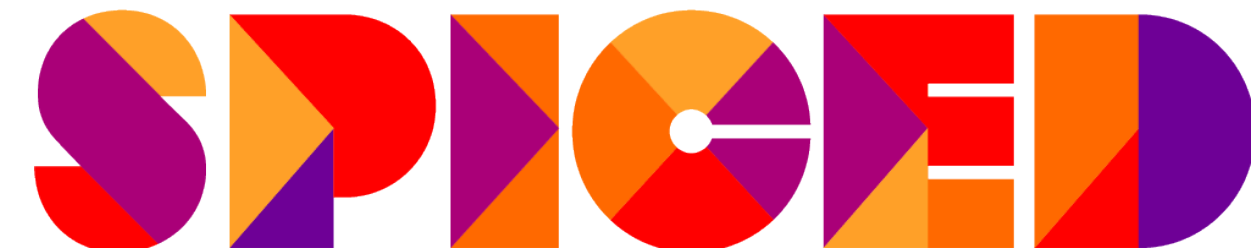
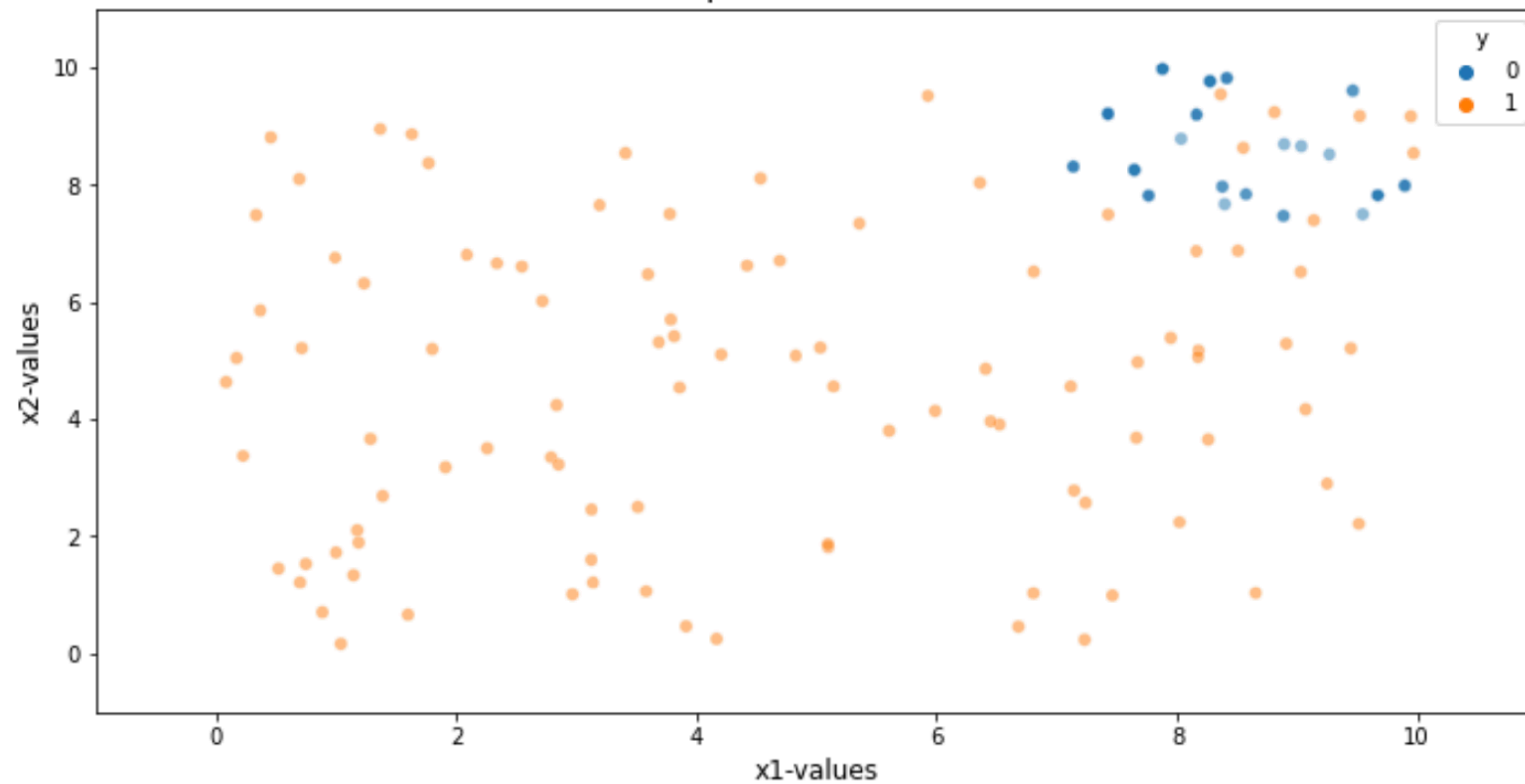
Random Oversampling

Original unbalanced data



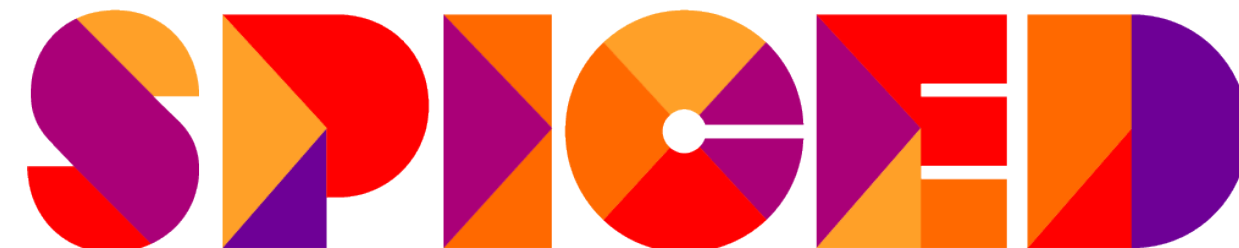
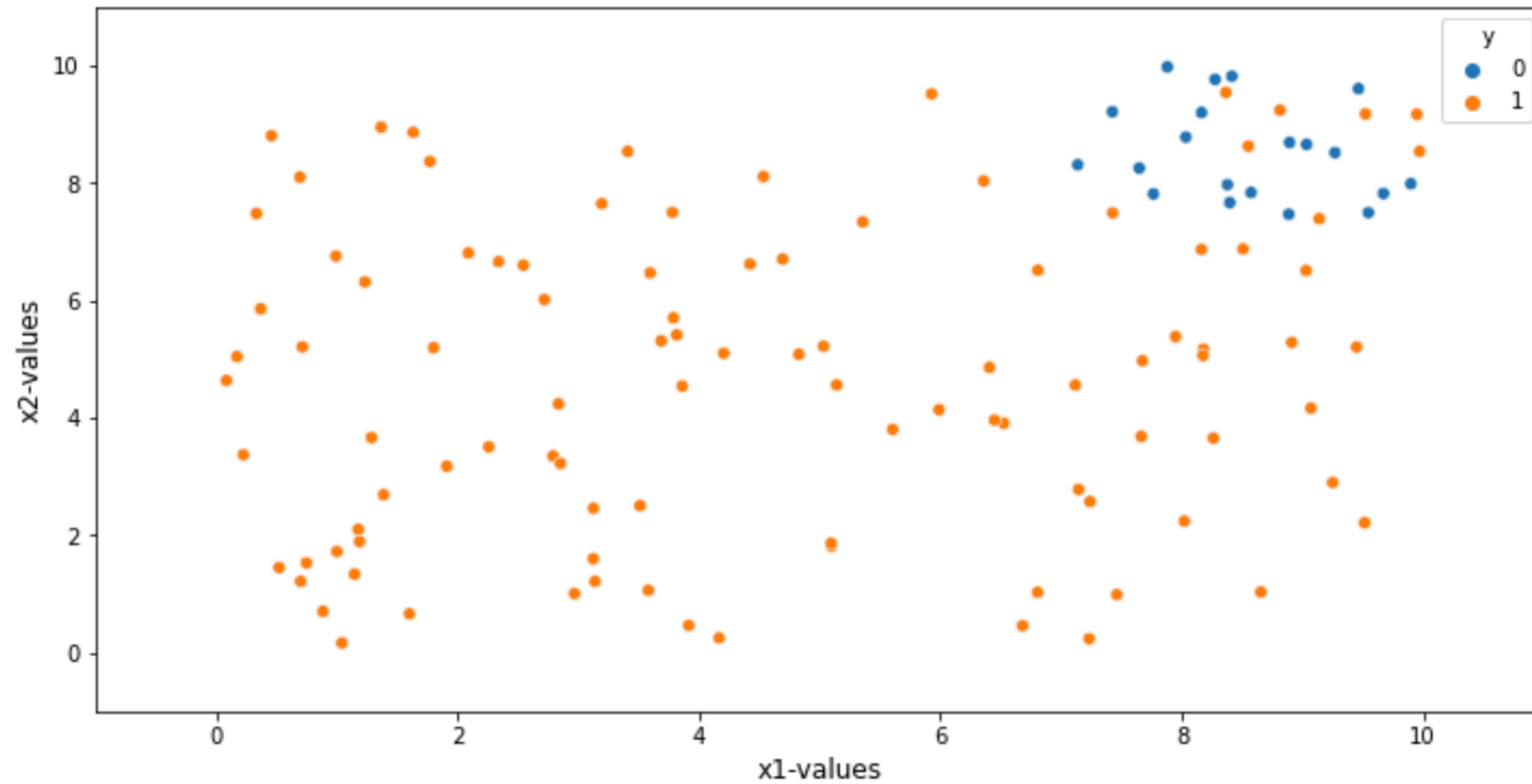
Random Oversampling

With Random oversampling



SMOTE Oversampling

Original unbalanced data



SMOTE Oversampling

SMOTE Oversampling

