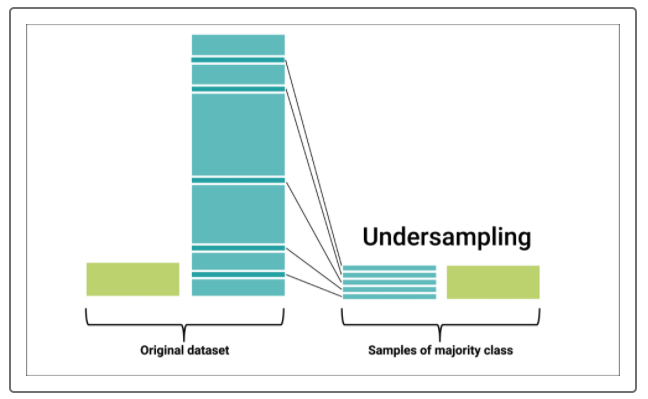
**You've** learned that in oversampling, the smaller class is resampled to make it larger. Undersampling, in contrast, takes the opposite tack.

Undersampling is another technique to address class imbalance. Undersampling takes the opposite approach of oversampling. Instead of increasing the number of the minority class, the size of the majority class is decreased.



Keep in mind that both oversampling and undersampling involve tradeoffs. Oversampling addresses class imbalance by duplicating or mimicking existing data. In contrast, undersampling only uses actual data. On the other hand, undersampling involves loss of data from the majority class. Furthermore, undersampling is practical only when there is enough data in the training set. There must be enough usable data in the undersampled majority class for a model to be useful.

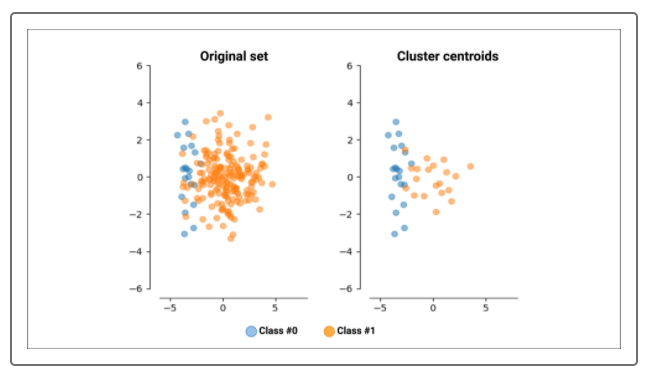
We'll discuss two approaches to undersampling: random and cluster centroid. Both are similar to the oversampling methods we've seen.

**Random Undersampling**

In random undersampling, randomly selected instances from the majority class are removed until the size of the majority class is reduced, typically to that of the minority class. The dataset used in this example contains information on credit card default.

## Cluster Centroid Undersampling

Cluster centroid undersampling is akin to SMOTE. The algorithm identifies clusters of the majority class, then generates synthetic data points, called centroids, that are representative of the clusters. The majority class is then undersampled down to the size of the minority class.



While resampling can attempt to address imbalance, it does not guarantee better results.