## Training and Testing Set: Size and Composition

The telecom churn dataset was preprocessed and encoded and then split into two: training set and testing set. This partition is essential to make sure that the clustering model is tested on the data that it has not encountered in the training process, which will assist in determining the generalizability of the model. The split was stratified to keep the balance of the target variable (Churn). This makes sure that the training and testing sets both represent the original proportion of churned and non-churned customers without any bias generated by an uneven split. The data was divided in the ratio of 80/20 meaning:

* 80 percent of data was assigned to the training set.
* 20 percent was reserved out as the testing set.