# Méthodologie

## Pré-traitement

On a importé les données (feuilles app\_test et Oot du classeur prolib).

On a réduit la volumétrie de la bdd en conservant 10% de la base (random\_state=42).   
Nous avons réalisé cette réduction par stratification sur les variables DDefaut\_NDB et datdelhis, de manière à garder les mêmes proportions des taux de défauts par date.

Voici la vérification :

Le taux de défaut pour l’ensemble des données au cours du temps est de l’ordre de 6%.

On a travaillé avec 3 sous échantillons (train, test et ooo). train représente 75% de app\_test et test représente 25% de app\_test (75 et 25% des 10% des données choisies aléatoirement).