

SCORING

Loic Hannequin M2 DS Groupe 2

Gabriel Lebovic 2024-2025

# Contexte

## Présentation des données

## Regroupement de données

Pour la variable « marital-status » : le regroupement a été fait de la manière suivante :

* Marié 🡪 Married-AF-spouse ; Married-civ-spouse ; Married-spouse-absent
* Plus marié 🡪 Divorced ; Separated ; Widowed
* Jamais marié 🡪 Never-married

Pour la variable native-country, elle est représentée par la modalité « United-States » à hauteur de 90 % et ensuite aucunes autres modalités dépassent les 2 %, donc on décide de regrouper tous les pays ensemble et de composer la modalité « Le reste du monde », ce qui nous fait une modalité qui contient maintenant 10 % de notre population, ce qui est acceptable.