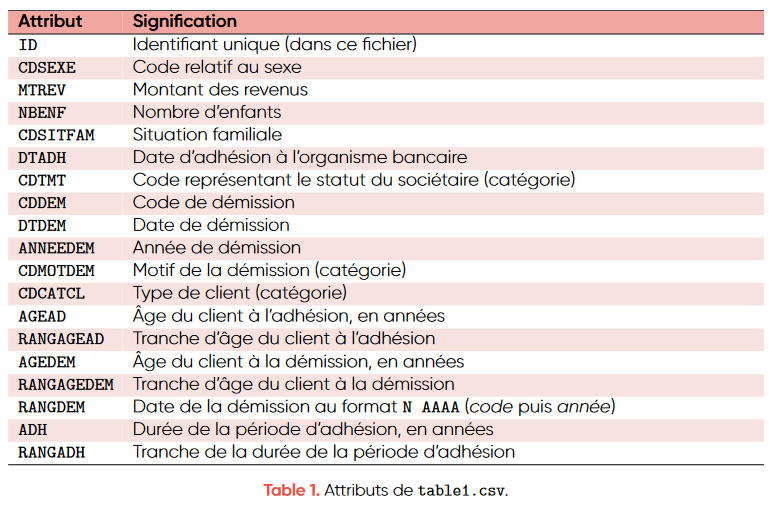
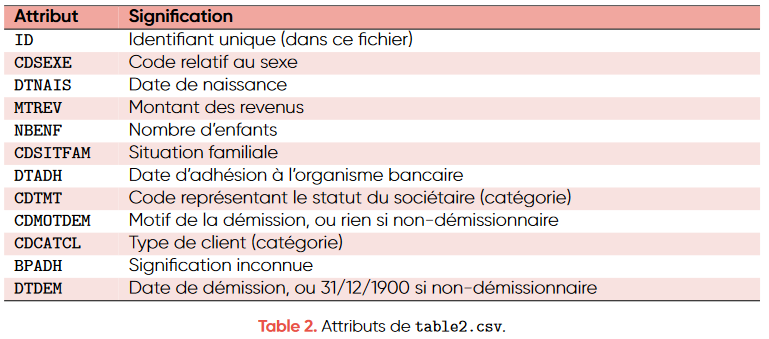
|  |
| --- |
| **ANALYSE ET INTERPRETATION DES FREQUENCES DE VALEURS**  **PRISES PAR LES METRIQUES** |

Attributs/Colonnes du fichier ‘’table1.csv’’ (contient les 30 332 démissionnaires de l’organisme bancaire, pour la période allant de 1999 à 2006) :



Attributs/Colonnes du fichier ‘’table2.csv’’ (échantillon aléatoire de 15 022 sociétaires de la banque, incluant des démissionnaires et des sociétaires actuels) :



Le but de cette partie de l’analyse et de lister les valeurs prises par les variables ainsi que leurs fréquences, permettant ainsi d’isoler les variables pertinentes et de réduire le nombre de variables total afin de simplifier l’analyse.

Pour voir les sorties et les fréquences calculées associées de table1.csv, voir les documents.txt contenus dans les dossiers **Analyse/Mesures\_frequences\_valeurs/** et **Analyse/Plots\_frequences\_valeurs/**

Ainsi on peut envisager ces hypothèses/actions sur ‘’table1.csv’’ :

* Exclure **ID** (pas d’information sur la démission 🡪 simple identifiant de ligne)
* Exclure **CDDEM** -> code de démission (type de démission pas important ici)
* Si garder **RANGAGEAD** 🡪 enlever **AGEAD** ?
* Si garder **RANGAGEDEM** 🡪 enlever **AGEDEM** ?
* Exclure **ADH** ? 🡪 Possibilité de le calculer avec **DTADH** et **RANGDEM**
* Exclure **CDSEXE** ? seulement 3 valeurs et une distribution assez équivalente entre 2/3.
* Changer **MTREV** en plage plutôt qu’en valeurs uniques
* Exclure **CDMOTDEM** ? 🡪 Motif codifié de la démission mal réparti est peu descriptif
* Traitement des valeurs **manquantes** ou **aberrantes**