|  |
| --- |
| **RAPPORT SUR LA CONSERVATION/EXCLUSION DES METRIQUES** |

**Récapitulatif**

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

‘’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 d’établir le niveau d’importance des variables (colonnes) vis-à-vis de l’analyse ainsi que la liste des possibles nettoyages à faire sur les données en pratique (lignes).

Dans le tableau suivant :

* Si la variable est barrée, elle est exclue
* Si elle ne l’est pas elle est conservée.
* Le commentaire indique la justification du choix à chaque fois

Une des étapes futures concernant ces attributs, sera la sélection des différentes valeurs que peuvent prendre ces attributs (élimination d’une part moins importante pour chaque attribut).

**Actions envisagées pour le nettoyage**

* Création d’un nouvel attribut commun aux deux tables **DEMISSION** à 0 si le client n’est pas (encore) démissionnaire, à 1 s’il l’est (cet attribut vient remplacer **CDMOTDEM** et est commun aux deux tables) 🡪 légère perte d’information sur le motif
* Laisser la variable **CDMOTDEM** et ajouter ND pour les non-démissionnaires de table2 🡪 plus précis (attention au nom de l’attribut qui peut porter à confusion sur les clients non-démissionnaires)

|  |  |  |
| --- | --- | --- |
| **Attribut** | **Signification** | **Commentaire** |
| **TABLE 1** | | |
| **~~ID~~** | Identifiant unique | Inutile et non-commun aux tables |
| **CDSEXE** | Code relatif au sexe | Marqueur descriptif direct commun à tous les clients |
| **MTREV** | Montant des revenus | Essentiel, mais traitement à envisager (grosse quantité de 0) |
| **NBENF** | Nombre d’enfants | Marqueur fort (>80% n’ont pas d’enfant) |
| **CDSITFAM** | Situation familiale | Marqueur descriptif direct sur le client |
| **DTADH** | Date d’adhésion à l’organisme bancaire | Permet de recalculer plusieurs autres variables |
| **CDTMT** | Code représentant le statut  du sociétaire (catégorie) | Marqueur descriptif direct du client (commun à tous les clients) |
| **~~CDDEM~~** | Code de démission | Pas commun aux deux tables et inutile pour la prédiction (ne concerne pas les clients non-démissionnaires ou les futurs démissionnaires) + voir actions envisagées |
| **DTDEM** | Date de démission | Permet de calculer plusieurs autres attributs (**ADH**…) |
| **~~ANNEEDEM~~** | Année de démission | Possibilité de la calculer à partir de DTDEM |
| **(CDMOTDEM)** | Motif de la démission (catégorie) | Voir actions envisagées |
| **CDCATCL** | Type de client (catégorie) | Marqueur descriptif direct du client (commun à tous les clients) |
| **AGEAD** | Âge du client à l’adhésion, en années | Permet de calculer d’autres variables et de décrire de nouveaux intervalles |
| **~~RANGAGEAD~~** | Tranche d’âge du client à l’adhésion | Redondance avec **AGEAD** |
| **~~AGEDEM~~** | Âge du client à la démission, en années | Retrouvable à partir de **AGEAD** et **DTDEM** |
| **~~RANGAGEDEM~~** | Tranche d’âge du client à la démission | Redondance avec **AGEDEM** |
| **~~RANGDEM~~** | Date de la démission au format N AAAA (code puis année) | Code non-pertinent + date retrouvable à partir de **DTDEM** |
| **~~ADH~~** | Durée de la période d’adhésion, en années | Retrouvable à partir de **DTDEM** et **DTADH** |
| **~~RANGADH~~** | Tranche de la durée de la période d’adhésion | Redondance avec **ADH** (donc calculable) |
| **TABLE 2** | | |
| **ID** | Identifiant unique | Inutile et non-commun aux tables |
| **CDSEXE** | Code relatif au sexe | Marqueur descriptif direct commun à tous les clients |
| **DTNAIS** | Date de naissance | Permet de calculer l’âge différentes valeurs pour rendre commun certains attributs entre les tables |
| **MTREV** | Montant des revenus | Essentiel et commun aux deux tables |
| **NBENF** | Nombre d’enfants | Marqueur descriptif fort et commun à tous les clients |
| **DCSITFAM** | Situation familiale | Marqueur descriptif fort et commun à tous les clients |
| **DTADH** | Date d’adhésion à l’organisme bancaire | Permet de recalculer certains attributs pour les rendre communs aux deux tables |
| **CDTMT** | Code représentant le statut du sociétaire (catégorie) | Marqueur descriptif direct du client (commun à tous les clients) |
| **(CDMOTDEM)** | Motif de la démission, ou rien si non-démissionnaire | Voir actions envisagées |
| **CDATCL** | Type de client (catégorie) | Marqueur descriptif direct du client (commun à tous les clients) |
| **~~BPADH~~** | Signification inconnue | Inconnu |
| **DTDEM** | Date de démission, ou 31/12/1900 si non-démissionnaire | Commun aux deux tables + (envisager une fusion de tables 🡪 concordance) |