

2eme fichier: Analyse KYC CSP FCA LMA

Comprendre le nom :

KYC (Know Your Customer): mchru7a f fichier 1

CSP (Customer Service Provider)

Le terme **CSP** peut se référer aux fournisseurs de services qui interagissent directement avec les clients. Dans le cadre de l'analyse KYC ou de la conformité réglementaire, ces fournisseurs jouent un rôle clé en tant qu'intermédiaires dans la collecte des données des clients, l'enregistrement de leurs informations et la gestion des demandes de vérification.

Fonction dans l'analyse :

- Collecter et gérer les informations clients à travers divers canaux.
- Faciliter les interactions entre le client et l'institution pour la validation KYC.

FCA (Financial Conduct Authority)

La **FCA** est l'organisme de régulation financière au Royaume-Uni. Son rôle principal est de superviser les sociétés financières pour assurer le respect des règles et des lois, y compris celles liées à la lutte contre le **blanchiment d'argent** et à la fraude.

Fonction dans l'analyse :

- Surveiller la mise en œuvre des processus KYC par les entreprises financières.
- Exiger des audits réguliers pour garantir que les entreprises respectent les réglementations de conformité.

LMA (Loan Market Association)

L'**LMA** est une association qui établit des normes et des pratiques de marché pour le secteur des prêts syndiqués. Elle publie des lignes directrices sur les questions liées à la conformité, y compris les aspects KYC dans le cadre des prêts.

Fonction dans l'analyse :

- Fournir des lignes directrices pour les institutions financières en matière de KYC lors de la syndication de prêts.
- Aider à la gestion des risques et à la conformité des processus KYC lors des transactions financières importantes.

L'analyse **KYC** est un processus essentiel dans le monde financier, notamment dans les secteurs régulés par la **FCA** ou les pratiques de prêts guidées par la **LMA**. Les **CSP** sont des intermédiaires dans la mise en œuvre de ces processus, garantissant que les informations des clients sont collectées et vérifiées conformément aux exigences réglementaires

Comprendre le script :

1.Ouverture des fichiers

- **Contexte réel** : Le script commence par ouvrir les fichiers nécessaires pour l'analyse. Ces fichiers contiennent des informations sur les personnes, leurs professions et des codes de métier associés.

```
OPEN XT050_FCA_PERSONNE_SUM
OPEN XT050_LME_SORT_SUM SECONDARY
```

Explication : Le fichier **XT050_FCA_PERSONNE_SUM** est ouvert en tant que table principale, et **XT050_LME_SORT_SUM** est utilisé comme fichier secondaire pour compléter les informations.

/!\ : had fichier **XT050_LME_SORT_SUM** huwa résultat dial fichier lwl lisiftlk fl pdf

2. Jointure des données

- **Contexte réel** : Cette étape permet de lier les données des deux fichiers (principal et secondaire) à partir du champ **NUM_PERSONNE**, afin d'obtenir une vue complète des personnes et de leurs professions.

```
JOIN PKEY VALUE(NUM_PERSONNE;0) FIELDS NUM_PERSONNE CDE_PROF_ASSURE LIB_PROF_ASSURE
WITH CODE_CSP PROFESSION TO XT_CSP_FCA_LME OPEN PRESORT SECSORT
```

Explication : Le script lie les informations de profession et de métier pour chaque individu en fonction de leur identifiant unique (**NUM_PERSONNE**), et les résultats sont stockés dans le fichier **XT_CSP_FCA_LME**.

Exemple (rir exemple osf machi dakchi lighadi tl9ay f data dialk)

Exemple :

Table principale :

NUM_PERSONNE	CDE_PROF_ASSURE	LIB_PROF_ASSURE
12345	101	Médecin
67890	102	Infirmier

Table secondaire (avec CODE_CSP et PROFESSION):

NUM_PERSONNE	CODE_CSP	PROFESSION
12345	101	Médecin généraliste
67890	103	Infirmier chef

Après jointure des deux tables :

NUM_PERSONNE	CDE_PROF_ASSURE	LIB_PROF_ASSURE	CODE_CSP	PROFESSION
12345	101	Médecin	101	Médecin généraliste
67890	102	Infirmier	103	Infirmier chef

3. Formatage des données des professions

- **Contexte réel :** Le script commence à nettoyer et formater les descriptions des professions pour supprimer les espaces et les caractères spéciaux. Cela permet de normaliser les données avant de procéder à une analyse plus poussée.

```
DELETE FIELD PROFESSION2 OK  
DEFINE FIELD PROFESSION2 COMPUTED  
ALLTRIM(REPLACE(UPPER(PROFESSION), "Á", "A", "È", "E", "Î", "I", ... ))
```

- **Explication :** Les professions sont normalisées en supprimant les accents et en convertissant tout en majuscules. Par exemple, "médecin" devient "MEDECIN", "développeur" devient "DEVELOPPEUR". Cela garantit une cohérence dans la façon dont les professions sont représentées dans le script.

Exemple :

Avant nettoyage :

"médecin généraliste"

Après nettoyage :

"MEDECIN GENERALISTE"

4. Nouvelle définition des professions

- **Contexte réel :** Dans cette étape, le script redéfinit et nettoie encore plus les professions en supprimant les préfixes inutiles, comme "PROF." ou d'autres chaînes de texte spécifiques, pour avoir une version encore plus standardisée.

```
ALLTRIM(REPLACE(ALLTRIM(PROFESSION2), "PROF.", "PROFESSION"))
```

- **Explication :** Toutes les occurrences de "PROF." dans les professions sont remplacées par "PROFESSION". Par exemple, "PROF. MEDECIN" devient "PROFESSION MEDECIN".

Exemple :

Avant : "PROF. DEVELOPPEUR"

Après : "PROFESSION DEVELOPPEUR"

5. Extraction des professions différentes

- **Contexte réel :** Cette étape permet d'extraire uniquement les lignes où les professions et les codes métiers ne correspondent pas exactement. Cela permet de détecter les incohérences entre les professions enregistrées.

```
EXTRACT FIELDS NUM_PERSONNE CDE_PROF_ASSURE LIB_PROF_ASSURE CODE_CSP  
PROFESSION3 AS "PROFESSION" TO XT_CSP_FCA_LME2
```

```
EXTRACT FIELDS ALL IF CDE_PROF_ASSURE <> CODE_CSP AND  
ALLTRIM(LIB_PROF_ASSURE) <> ALLTRIM(PROFESSION) TO XT_CSP_FCA_LME3
```

Explication : Les individus ayant des professions ou des codes de métier différents sont extraits pour une analyse ultérieure. Cela est particulièrement utile pour détecter des doublons ou des erreurs de saisie dans les professions.

Exemple :

Si une personne a "MEDECIN" comme profession dans un fichier, mais "MEDECIN GENERALISTE" dans un autre, cette différence sera capturée ici.

6. Calcul du coefficient de similarité (DICE)

- **Contexte réel :** Le coefficient DICE mesure la similarité entre deux professions. Un score proche de 1 indique une forte similarité, tandis qu'un score plus bas signale une différence importante.

```
DICECOEFF(LIB_PROF_ASSURE; PROFESSION)
```

- **Explication :** Le script compare deux professions et calcule leur similarité. Cela permet de comprendre si deux professions sont fondamentalement les mêmes avec des libellés différents.

Exemple :

Comparaison de "MEDECIN" et "MEDECIN GENERALISTE" :

DICE Coefficient = 0,85 (ce qui indique une forte similarité, mais pas une identité totale).

7. Extraction des professions dissimilaires

- **Contexte réel :** Une fois que le coefficient DICE est calculé, cette étape extrait uniquement les lignes où la similarité est faible (coefficient inférieur à 0,7). Ces professions sont considérées comme étant suffisamment différentes pour nécessiter une investigation.

```
EXTRACT FIELDS NUM_PERSONNE CDE_PROF_ASSURE LIB_PROF_ASSURE CODE_CSP  
PROFESSION IF DICE_COEF < 0.7 TO RE506_CSP_FCA_LME
```

- **Explication :** Les professions ayant un coefficient de similarité faible sont extraites pour une analyse plus approfondie. Cela permet de vérifier les cas où les professions sont réellement différentes et pas seulement des variations mineures.

Exemple :

Comparaison de "MEDECIN" et "CHIRURGIEN" :

DICE Coefficient = 0,6 (ce qui indique une différence significative).

8. Finalisation

À la fin du processus, le script a nettoyé, standardisé et comparé les professions, et a extrait celles qui nécessitent une analyse supplémentaire. Ce type de traitement est particulièrement utile pour vérifier les incohérences dans les bases de données contenant des informations sur des métiers ou des professions.

9. En général

1. **Ouverture des fichiers et jointure :** Le script ouvre deux fichiers contenant des informations sur des personnes et leurs professions. Il relie ces fichiers pour obtenir des

informations complètes en associant les identifiants de chaque personne à leurs professions dans les deux bases de données.

2. **Nettoyage et normalisation des professions** : Le script procède ensuite à une étape de nettoyage des données pour standardiser les descriptions des professions. Cela inclut la suppression des accents, la conversion en majuscules, et l'élimination des espaces ou préfixes inutiles. Cela permet d'avoir des libellés de professions uniformes pour faciliter les comparaisons.
3. **Comparaison des professions** : Le script compare les professions des individus dans les deux bases de données, en vérifiant si le code métier et le libellé de la profession sont identiques ou non.
4. **Calcul de la similarité (coefficient DICE)** : Pour les professions qui ne sont pas exactement identiques, le script calcule un **coefficient DICE**, qui mesure la similarité entre deux professions. Cela permet d'évaluer si des professions apparemment différentes sont en réalité très proches (par exemple, "Médecin" et "Médecin Généraliste").
5. **Extraction des différences significatives** : Le script extrait les professions où la similarité est faible (c'est-à-dire que le coefficient DICE est inférieur à 0,7), ce qui indique des différences significatives. Ces lignes sont ensuite sauvegardées pour être analysées plus en profondeur.

Objectif global :

Le but de ce script est d'identifier les incohérences ou les variations dans les descriptions des professions et de corriger ou analyser ces différences pour garantir une meilleure qualité des données dans les systèmes métiers.