**Jointures effectuées :**

~~Deep\_course\_cf\_themes.csv : base des thèmes des textes~~

~~Deep-course\_cf\_types.csv : id du type de document, applicabilité ou non~~

* Ces deux bases peuvent être fusionnées à Deep\_course\_cf\_text\_versions.csv  pour avoir toutes les caractéristiques des textes (appelée df\_text dans le notebook)
* Pour avoir les textes appliqués, on joint également deep\_course\_cf\_text\_version\_nodes.csv à df\_text

~~Deep\_course\_formulaire\_questionnaire.csv : grandes parties du questionnaire~~

* Base jointe à Deep\_course\_formulaire\_questionnaire\_pages.csv

Deep\_course\_cm\_nodes.csv : structure du questionnaire (numéro de la question, id page de la question, id question) erreur dans le document word fourni ? confondue avec deep\_course\_formulaire\_page\_questions.csv ? car deep\_course\_cm\_nodes.csv correspond apparemment à la nature des clients.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3515 textes, 53 thèmes regroupés en 8 domaines. Un texte peut être applicable ou pour information.

Pourquoi il y a 7369 textes uniques dans deep\_course\_cf\_text\_version\_nodes.csv ? cela ne devrait pas excéder 3515

597 clients dans la base de données, mais seulement 573 qui ont des textes appliqués dans deep\_course\_cf\_text\_version\_nodes.csv. Les autres n’ont pas encore répondu ?

3 statuts différents : SITE (586), GROUP (9) et ABOVE\_SITE (2). Signification exacte ?

329942 textes appliqués sur l’ensemble des clients.

316 questions : apparemment 2 questionnaires différents : enonce\_site et enonce\_groupe, selon le type du client. 282 questions pour un groupe et 312 pour un site.

Type 0 (question ouverte) 1 (yes no question) 2(yes/no , QCM ? pas clair) 3 (divers) 4 (quantité ?) 5 (informations complémentaires)

264 questions à choix : 5 questions à 6 choix, 4 questions à 5 choix, 3 questions à 4 choix, 4 questions à 3 choix, 248 à 2 choix.

9 questions demandent une unité précise.

Parties du questionnaire dans l’ordre : info générales, HSE, environnement, sécurité, surete-malveillance et energie.

Dans deep\_course\_formulaire\_page\_questions.csv, on a 321 id de questions uniques, ce qui est étrange car on ne devrait pas en avoir plus de 316 d’après la base de questions.

A quoi sert Pages\_id exactement ? On a 44 Pages\_id mais il n’y aurait que 11 pages au questionnaire d’après page\_order… Nous n’avons pas bien compris.