

Date attendue de rendu : 17 mars 2020 à 18h

5 Itérations et modifications du code existant

La source de données initiale a été modifiée.

Parmi les modifications majeures, le fichier d'entrée est en parquet (auparavant CSV), et certains noms de colonnes ont été modifiés. Ces modifications ne sont pas exhaustives, nous vous laissons découvrir les autres changements apportés à ces données d'entrée.

Ces variations de format sont très fréquentes lors de développements liés à la data. Ainsi, en phase de POC (Proof of Concept), les données utilisées sont souvent des extractions des systèmes. Lors d'un passage en pilote, ce sont des extractions différentes ou les connecteurs doivent gérer directement les données en sortie d'un flux.

L'objectif de ce travail est donc d'**adapter votre code existant** afin qu'il gère ce nouveau format.

A la fin de ce *refactoring*, nous ferons un récapitulatif des points de douleur rencontrés par tous les étudiants (pensez-y lors de votre *refactoring*).