

Projet entreprise multinational corporation – Part 1

Contexte :

Une entreprise multinational spécialisée dans la data définit une grille salariale pour ces employés à travers le monde en fonction de leurs niveaux d'étude, leurs postes, leurs compétences entre autres. Elle réunit dans un fichier Excel toutes les informations sur ses employés ainsi que les réponses à plusieurs questions auxquels ils ont répondu.

Pour traiter ces données avec la plus grande facilité il conviendra de les intégrer dans une base de données structurées.

Pour ce faire :

- Il faudra dans un premier temps mener une brève analyse exploratoire qui aura seulement pour but d'identifier le type des variables (et donc le type des colonnes de la base de données que vous allez créer et leurs longueurs) et de déterminer s'il y a des valeurs manquantes.
 - Pour cela utiliser ma méthode `pandas.read_excel()` (pensez à ignorer les 3 premières lignes)
 - Avec l'attribut `dtypes` de votre dataframe vous obtiendrez le type de toutes les colonnes.
 - Avec la méthode `Series.value_counts()` vous aurez les différentes modalités des variables de types object. Ce qui vous donnera un bon aperçu.
- Vous devriez ensuite modéliser la base de données que vous jugeriez la plus adaptée à vos données en utilisant la méthode MERISE.
 - Idéalement parmi les colonnes de type int il y aura « TelecommuteDaysPerWeek ».
 - Les colonnes « OtherJobDuties » et « KindsOfTasksPerformed » contenant plusieurs modalités dans une cellule donneront naissance à des tables qui leur sont dédiées ainsi qu'à des tables de liaisons.

Les rendus attendus :

- Un MCD, MLD-MRD
- Le script SQL permettant la création de la base. Vous pouvez utiliser l'interface graphique de Workbench pour créer votre base, ou rédiger directement le script.

A vous de jouer.