Université USTHB – Bab-Ezzouar Faculté d'Informatique, Département d'IASD, Module : Data Mining Bab-Ezzouar, 2023 / 2024 2ème année Master Informatique Système Informatique Intelligent Semestre 1

Exercice 1 Analyse des Données

Boîte à outils Python:

Ouverture et lecture d'un fichier ligne par ligne :

with open(<path_to_file>, 'r') as file: for line in file: print(line)

Séparation des attributs de chaque ligne :

<IDF_STR_ARRAY> = <IDF_STR>.split(<SEPARATEUR>)

Retourne un tableau de chaînes de caractères après découpage de la chaîne globale selon le séparateur choisie.

Conversion d'une chaîne de caractère selon le type désiré :

<IDF_DOUBLE> = float(<IDF_STR>) <IDF_INTEGER> = int(<IDF_STR>)

Questions:

- 1- Écrire une fonction Python permettant de charger le dataset.
- 2- Écrire une fonction Python permettant d'afficher quelques informations de base sur le dataset.
- 3- Écrire une fonction Python permettant de calculer les tendances centrales d'un attribut.
- 4- Écrire une fonction Python permettant de calculer les quartiles (Q₀, Q₁,Q₂,Q₃,Q₄) d'un attribut.
- 5- Écrire une fonction Python permettant d'afficher le nombre et pourcentage de valeurs manquantes d'un attribut.

Dataset : Dataset-Exos.txt disponible sur ce lien.

Have fun!