**Exercice semaine 1- data science**

**Date limite de soumission des réponses: *Jeudi 23 mars 2023 à 12h***

**Discussion en ligne : *Vendredi 24 mars 2023 de 16h à 18h***

**Thématiques abordées : Bases de python, numpy, pandas**

**NB :** Vous êtes priés d’utiliser Jupyter notebook comme environnement de développement intégré (IDE). Vous pouvez soumettre votre travail en Jupyter notebook ou en htlm.

**Partie 1 : Bases de python (chaque question est notée sur 5pts)**

1. Déclarer les variables x, y et z respectives suivantes dans votre IDE : 2, 4.3 et <https://raw.githubusercontent.com/pefura/IFPERA/main/Base%20_pleuresie_ML_2.csv>
2. Calculer x/y, x-2y.
3. Déterminer la partie entière de la division de y par x.
4. Déterminer le reste de la division de y par x.
5. En utilisant une fonction native précise de python, créer la liste nommée « List » de tous les éléments de z séparés par « / ».
6. Extraire l’élément en deuxième position de « List ».
7. Supprimer l’élément « ' ' » de la liste « List ».
8. Extraire le dernier élément de « List ».
9. Construire un dictionnaire nommé « dictionnaire » avec les clés suivantes : oiseau, chat, chien. Les valeurs de ces clés étant la traduction anglaise des clés.

**Partie 2 : Numpy (chaque question est notée sur 5pts)**

Déclarer les objets suivants (poids en kg, taille en cm et sexe [1= sexe masculin, 2 = sexe féminin]):

poids= [50, 55, 80, 70, 86, 64]

taille = [170, 170, 170, 162, 171, 148]

sexe = (1, 2, 2, 2, 1, 1)

1. Comment appelle-t-on les structures de données utilisées pour chacun des objets poids, taille et sexe définis dans la vignette de la partie 2 dans python (**dresser dans votre IDE simplement un tableau contenant les réponses**)?
2. En utilisant la librarie numpy [Numerical Python], transformer les objets poids, taille, sexe en tableaux numpy de dimension 1 (1 D array).
3. Créer une matrice (mat) numpy dont les objets poids, taille et sexe sont des colonnes.
4. Quelle est la forme de « mat » ?

**Partie 3 : Pandas (chaque question est notée sur 5pts)**

1. Transformer mat en un dataframe de pandas nommé « dataset » en indiquant clairement le nom des colonnes.
2. Sélectionner la sous-base « dataset\_female » de dataset ne contenant que les sujets de sexe féminin.

**Partie 4 : Fonctions**

1. En utilisant la boucle for, créer une liste contenant les logarithmes naturels des éléments de la liste « poids » définie à la partie 2 **(10 pts).**
2. Définissez une fonction générique nommée « obesity\_classification » qui permet à partir du poids et de la taille d’un individu de déterminer s’il est maigre, de poids normal, en surpoids ou obèse. Il est nécessaire d’utiliser l’indice de masse corporelle (IMC) avec les cut-off suivants 18.5 kg/m2, 26 kg/m2 et 30 kg/m2 pour définir respectivement les classes d’IMC. Il est également conseiller d’afficher la classe correspondante avec l’IMC correspondant **(10 pts).**
3. Exécuter la fonction « obesity\_classification » pour les deux premières observations de dataset (question 14), **5 pts.**

**The right way to learn data science is to do data science**

**Good luck**