**Exercice semaine 2- data science**

**Date limite de soumission des réponses: *Jeudi 30 mars 2023 à 12h***

**Discussion en ligne : *Vendredi 31 mars 2023 de 16h à 18h***

**Thématiques abordées : 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 : Numpy (chaque question est notée sur 5pts)**

1. Importer l’objet (nommez le « objet») localisé à l’adresse URL : <https://raw.githubusercontent.com/pefura/IFPERA/main/photo_semaine_2.png>
2. Quel est le type de « objet» ?
3. Déterminer les valeurs minimale et maximale des pixels de « objet ».
4. Déterminer la taille, la forme et le nombre de dimension de « objet ».
5. Exécuter la fonction : plt.imshow(objet) [il s’agit d’une fonction du package « matplotlib.pyplot »]. Quelle est votre conclusion ?
6. Utiliser la librarie open CV pour transformer cette image en image à 3 canaux RGB.

**Partie 2 : Pandas (chaque question est notée sur 5pts sauf les questions 12 et 13 qui sont notées respectivement sur 10 et 15 pts)**

Voici une base des données disponible à l’adresse : <https://raw.githubusercontent.com/pefura/IFPERA/main/Cameroon_lung_function.csv>

Cette table de données sera nommée « dataset » après importation dans votre IDE.

Le tableau de codage de dataset est fourni ci-dessus :

|  |  |  |
| --- | --- | --- |
| Variables | signification | code |
| ID | identification | / |
| sex | Genre/sexe des sujets | 1= sexe masculin  2= sexe féminin |
| age | Age en années | / |
| ethnicity | Groupe ethnique | 1= bantou  2= Soudano-sahelien  3= mixte |
| height | Taille en cm | / |
| fvc | Capacité vitale force en L | / |
| fev | Volume expiratoire maximal à la première seconde en L | / |
| fef2575 | Débit expiratoire médian | / |
| fevfvc | Rapport fev/fvc | / |

1. Importer dataset à l’aide du package pandas et afficher les 5 premières lignes de dataset.
2. Rendre conforme la typologie des variables de dataset et ne contenant que les 3 premières colonnes.
3. Sélectionner les 100 dernières observations de dataset.
4. Calculer le logarithme naturel de l’âge de tous les individus et inclure cette colonne dans dataset.
5. Décrire numériquement à l’aide d’une fonction de pandas les variables quantitatives de dataset
6. Définissez une fonction générique de python (nommée num\_cat) permettant de décrire numériquement les variables catégorielles de dataset (effectifs et proportions). Les effectifs et les proportions doivent être combinés dans un tableau avec les modalités des variables catégorielles correspondantes.
7. Définissez une fonction de python nommée « ratio » permettant de calculer le rapport fev/fvc des individus situés respectivement à la nième position et la nième + 1 position.

Vérifier la véracité de la fonction pour les individus de la 10ème et 11ème position.

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

**Good luck**