**Exercice semaine 2- data science**

**Date limite de soumission des réponses: *Jeudi 06 avril 2023 à 12h***

**Discussion en ligne : *Vendredi 07 avril 2023 de 16h à 18h***

**Thématiques abordées : Visualisation des données dans python**

**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.

Vous devez utiliser

**Vous devez utiliser matplotlib ou seaborn pour la visualisation des données**

**Barème : chaque question est notée sur 5pts**

**Partie1 : Visualisation d’une fonction**

1. Définir dans python la fonction f (a) = [exp](https://numpy.org/doc/stable/reference/generated/numpy.ufunc.html#numpy.ufunc)(-[a](https://numpy.org/doc/stable/reference/generated/numpy.ndarray.html#numpy.ndarray)) x [cos](https://numpy.org/doc/stable/reference/generated/numpy.ufunc.html#numpy.ufunc)(2π x [a](https://numpy.org/doc/stable/reference/generated/numpy.ndarray.html#numpy.ndarray)) [x est le signe de la multiplication]
2. Tracer la fonction y = f(a) dans l’intervalle [0, 5].
3. L’équation f(a) = 0 a-t-elle des solutions entières dans l’intervalle [0,100] ?

**Partie 2 : Visualisation des données d’un dataset**

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 dernières lignes de dataset.
2. Rendre conforme la typologie des variables de dataset.
3. En utilisant la programmation orientée objet de matplotlib, représenter le nuage de points décrivant fev en fonction de l’âge.
4. Représenter graphiquement la variable « ethnicity ».
5. Tracer à l’aide du package « seaborn » l’histogramme d’âge des sujets de sexe masculin, la largeur des classes étant de 3 et indiquant la courbe de densité sur la même figure.
6. Faire deux figures superposées montrant pour l’une la relation entre le fvc et l’âge chez les sujets de sexe masculin et pour l’autre la relation entre le fvc et l’âge chez les sujets de sexe féminin.
7. Ecrire une fonction permettant de réaliser une figure permettant de faire l’histogramme de toutes les données quantitatives.

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

**Good luck**