



Test : Manipulation de données avec python

But : Analyser les données du jeu de données iris pour répondre à des QCM

Les données iris ont été collectées afin de quantifier les variations de morphologie des fleurs d'iris d'un certain nombre d'espèces. 4 caractéristiques ont été mesurées à partir de chaque échantillon : la longueur et la largeur des sépales et des pétales, en centimètres.

Charger les données avec :

```
pip install seaborn    # installer si besoin
```

```
import seaborn as sns
```

```
iris = sns.load_dataset('iris')
```

```
iris.head()    # Afficher les 5 premières lignes
```

- Le test contient 10 questions

1) Quelle est la taille du dataset iris ?

Répondez en cochant les bonnes réponses

☐ 100

☐ 120

☐ 150

☐ 200

2) Quelle est la moyenne de la colonne "sepal_length" (réponse à 2 chiffres après la virgule)

Répondez en cochant les bonnes réponses

☐ 5,84

☐ 5,33

☐ 0,00

☐ 9,84

3)

Ranger dans l'ordre décroissant (du plus grand au plus petit) les espèces (species) selon la longueur maximale de sépales (sepal_length). L'espèce ayant la plus grande longueur de sépales reste en haut et celle ayant la plus faible en bas

Répondez en numérotant (1 - 3) chaque option

☐ setosa

☐ versicolor

☐ virginica

4) Quelle est l'espèce ayant la plus grande longueur moyenne des sépales ?

Répondez en cochant les bonnes réponses

☐ virginica

☐ versicolor

☐ setosa

5) Comment trouver les espèces les plus fréquentes dans le dataset : iris.species.[1]()

Répondez en écrivant le numéro de la position (1-1) sur l'option appropriée

☐ mod

☐ freq

☐ mode

☐ lot

6)

Combien d'espèces de fleurs présentent une longueur de sépales (sepal_length) de plus de 6 cm

Répondez en cochant les bonnes réponses

☐ 5

☐ 47

☐ 60

☐ 61

7) Quelle relation est correcte entre sepal_length et petal_length

Répondez en cochant les bonnes réponses

- ☐ Il n'y a aucune corrélation
- ☐ Plus `sepal_length` est grand, plus `petal_length` est grand
- ☐ Plus `sepal_length` est grand, plus `petal_length` est petit
- ☐ On ne peut pas savoir

8)

Quelle est la pente de la droite de régression entre les variables longueur de sépales (sepal_length) et longueur de pétales (petal_length) ? (arrondi à 2 chiffres après la virgule)

Répondez en cochant les bonnes réponses

- ☐ 0
- ☐ 0,52
- ☐ 1,86
- ☐ 2,01

9)

Une nouvelle fleur a été observée avec les caractéristiques suivantes. Ranger dans l'ordre les étapes suivantes pour prédire son espèce.

sepal_length = 5.8sepal_width = 3.0petal_length = 4.2petal_width = 1.2

Répondez en numérotant (1 - 7) chaque option

- ☐ Séparer en jeu d'entraînement et de test
- ☐ Charger les données (df = sns.load_dataset('iris'))
- ☐ Evaluer sur les données de test
- ☐ Entraîner la régression logistique (ou un autre modèle de classification)
- ☐ Prédire l'espèce en ayant défini le nouvel individu
- ☐ Importer les librairies (LogisticRegression, Seaborn, Numpy, sklearn, etc.)
- ☐ Encoder les labels (transformer les espèces en nombres), en ayant bien préparer les variables explicatives et la cible

10)

On veut obtenir le boxplot de sepal_length par espèce (ajouter titre, axes, grille

Écrivez votre réponse ci-dessous
