

## Introduction à la Science des données

## QCM n°1

Cet examen a pour but d'évaluer votre compréhension des 5 premières séances de l'UE *Introduction à la Science des données*. Ce questionnaire à choix multiples est sur 10 points.

Aucun document n'est autorisé.

Durée: 15 min (20 min si 1/3 temps)

<input type="checkbox"/> 0							
<input type="checkbox"/> 1							
<input type="checkbox"/> 2							
<input type="checkbox"/> 3							
<input type="checkbox"/> 4							
<input type="checkbox"/> 5							
<input type="checkbox"/> 6							
<input type="checkbox"/> 7							
<input type="checkbox"/> 8							
<input type="checkbox"/> 9							

← codez votre numéro d'étudiant ci-contre, et écrivez votre nom et prénom ci-dessous.

Nom et prénom :

.....

Les questions faisant apparaître le symbole ♣ peuvent présenter zéro, une ou plusieurs bonnes réponses. Les autres ont une unique bonne réponse.

**Question [code]** Sur quoi porte le Projet 1 ?

- La classification automatique de 2 catégories d'images
- Les pommes et les bananes
- Les rescapés du Titanic
- Je ne sais pas*

**Question [taches]** Quel langage utilise-t-on dans ce cours d'Introduction à la Science des Données ?

- c++
- pandas
- html
- Aucune de ces réponses n'est correcte*

**Question [norma11]** Soit un tableau *df*. Normalisez-le comme vu en TP.

- $df - \text{mean}(df) / \text{std}(df)$
- $(df - df.\text{mean}()) / df.\text{std}()$
- $(df - \text{mean}(df)) / \text{std}(df)$
- $df - df.\text{mean}() / df.\text{std}()$

**Question [norma12]** Soit un tableau *df*. Normalisez-le comme vu en TP.

- $df - \text{mean}.df / \text{std}.df$
- $(df - df.\text{mean}()) / df.\text{std}()$
- $(df - \text{mean}(df)) / \text{std}(df)$
- $(df - df.\text{mean}) / df.\text{std}$

**Question [read] ♣** Soit le fichier 'qcm.csv'. Quelles sont les commandes nécessaires pour le charger ?

- import pandas as pd  
 df = pd.read\_csv("qcm.csv")  
 df = np.read\_table("qcm.csv")  
 import numpy as np
- import matplotlib as mpl  
 df = mpl.read\_file("qcm.csv")  
 Aucune de ces réponses n'est correcte.

**Question [nan]** Comment apparait une donnée absente dans un tableau pandas ?

- NaN  
 NA
- False

**Question [assert] ♣** Soit un tableau *df* contenant 5 lignes et 38 colonnes, quelles assert vont renvoyer une erreur ?

- assert df.shape == 43  
 assert len(df) = 42  
 assert df.shape == (5, 38)
- assert len(df) == 5  
 Aucune de ces réponses n'est correcte.

**Question [true]** Soit  $a = [1, 2, 3]$  et  $b = 2$ . Que renvoie la commande  $a == b$  ?

- [FALSE, TRUE, FALSE]  
 FALSE

**Question [flake1] ♣** Soit une erreur de convention de codage identifiée par l'outil Flake8 (cf. TP3 et TP4). Quelles sont les phrases correctes?

```
[3]: from intro_sciences import code_checker
code_checker("flake8 utilities.py")
```

```

1 from intro_sciences import code_checker
----> 2 code_checker("flake8 utilities.py")

File ~/src/miniconda3/envs/jupyter/envs/intro-sciences/lib/python3.
39 if result.returncode:
40     print()
----> 41     raise RuntimeError(
42         f"La commande {command} a émis des avertissements:\n\n"
43         + result.stdout.decode()
44     )

RuntimeError: La commande flake8 utilities.py a émis des avertissements:
utilities.py:143:52: E202 whitespace before ')'
```

- Il y a un espace en trop  
 L'erreur est ligne 143  
 L'erreur est ligne E202  
 L'erreur est ligne 52
- L'erreur est ligne 55  
 Il y a une parenthèse ')' en trop  
 Aucune de ces réponses n'est correcte.

**Question [flake2] ♣** Soit une erreur de convention de codage identifiée par l'outil Flake8 (cf. TP3 et TP4). Quelles sont les phrases correctes?

```
[6]: from intro_science_donnees import code_checker
code_checker("flake8 utilities.py")
```

```
1 from intro_science_donnees import code_checker
----> 2 code_checker("flake8 utilities.py")

File ~/src/miniconda3/envs/jupyter/envs/intro-science-donnees/lib/python3.
39 if result.returncode:
40     print()
----> 41     raise RuntimeError(
42         f"La commande {command} a émis des avertissements:\n\n"
43         + result.stdout.decode()
44     )

RuntimeError: La commande flake8 utilities.py a émis des avertissements:
utilities.py:55:49: W291 trailing whitespace
```

- |   |   |
|---|---|
| <input checked="" type="checkbox"/> Il y a un espace en trop à la fin de la ligne | <input checked="" type="checkbox"/> L'erreur est ligne 55                 |
| <input type="checkbox"/> L'erreur est ligne 49                                    | <input type="checkbox"/> Il y a un espace en trop après le mot 'trailing' |
| <input type="checkbox"/> L'erreur est ligne W291                                  | <input type="checkbox"/> Aucune de ces réponses n'est correcte.           |
| <input type="checkbox"/> L'erreur est ligne 52                                    |   |

**Question [class1] ♣** Quelle(s) phrase(s) sont VRAIE(s) ?

- |  |   |
|--|---|
| <input checked="" type="checkbox"/> L'étiquette correspond à la valeur à prédire   | <input type="checkbox"/> Il faut toujours essayer de ramener la représentation des données à seulement deux dimensions pour faciliter la visualisation des données. |
| <input checked="" type="checkbox"/> Si une variable est distribuée normalement, une fois standardisée, environ 95% de ses valeurs sont comprises entre -2 et +2. | <input type="checkbox"/> L'attribut correspond à la valeur à prédire  |
| <input type="checkbox"/> La préparation des données est une des étapes les plus rapides en science des données.  | <input type="checkbox"/> Aucune de ces réponses n'est correcte.   |

**Question [class2] ♣** Soit un jeu de données d'entraînement avec 4 images de pommes et 6 images de bananes. Quelle(s) affirmation(s) sont vraie(s) concernant le classificateur KNN ?

- |  |  |
|--|--|
| <input checked="" type="checkbox"/> Le KNN est le classificateur des k plus proches voisins                                  | <input type="checkbox"/> Nous n'avons utilisé que la bibliothèque numpy, pour appliquer l'algorithme du KNN sur notre jeu de données |
| <input checked="" type="checkbox"/> Si k=10 alors n'importe quelle image sera prédite "Banane" qui est la classe majoritaire | <input type="checkbox"/> Aucune de ces réponses n'est correcte.  |
| <input type="checkbox"/> Le KNN est un classificateur de type réseau de neurones   |  |

**Question [diff1] ♣** Cochez les cases VRAIES concernant les écart-types.

- |   |  |
|---|--|
| <input type="checkbox"/> Un écart type peut être positif ou négatif   | <input checked="" type="checkbox"/> Un écart type correspond à la racine carrée de la variance |
| <input checked="" type="checkbox"/> Un écart type nul indique que toutes les valeurs sont égales à la moyenne | <input type="checkbox"/> Aucune de ces réponses n'est correcte.                                |
| <input type="checkbox"/> Un écart type correspond au carré de la variance                                     |  |

## CATALOGUE

**Question [diff2] ♣** Cochez toutes les affirmations VRAIES concernant les statistiques de la fonction "describe(include = all)":

- "mean" correspond à la valeur moyenne d'une colonne, elle s'applique à toutes les colonnes de la table
- "std" correspond à l'écart type (standard deviation en anglais). C'est la racine carrée de la variance.
- La fonction affiche entre autres l'information de la médiane pour les colonnes numériques
- "count" indique le nombre de lignes du tableau
- Aucune de ces réponses n'est correcte.*