
Examen - Session normale-
Python pour les sciences de données
17/01/2024. **Durée : 2 heures**

Exercice 1

1. Quelle est la différence entre une liste et un tuple en Python?
2. Expliquez la différence entre `append()` et `extend()` dans le contexte des listes en Python.
3. Écrivez une liste de compréhension en Python pour générer une matrice $n \times n$ où les éléments sur la diagonale principale sont les carrés des nombres de 0 à n , et les éléments hors de la diagonale sont le produit des indices de ligne et de colonne.

Exemple de résultat attendu pour une matrice de 3×3 :

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix}$$

4. Etant donnée une matrice A générée par list de comprehension, donnez le code Python permettant d'afficher toute une colonne de A en utilisant l'indigage (sans utiliser une boucle).

Exercice 2

Considérant un fichier `data.csv` contenant des données représentant les ventes de produits dont voici un extrait :

Produit	Ventes	Stock	Prix Unitaire
A	100	60	12MAD
B	17	141	47MAD
C	214	20	50MAD
D	86	122	85MAD
E	163	35	34MAD
F	45	121	16MAD

Donner la ligne de code permettant de :

1. Charger "data.csv" dans un DataFrame pandas appelé `df` et d'afficher les cinq premières lignes du DataFrame.
2. Sélectionner uniquement les lignes du DataFrame où la valeur de la colonne "Stock" est inférieur à 100.
3. Sélectionner les lignes du DataFrame où la valeur de Ventes > 80 et du Stock < 60 .
4. Calculer la moyenne de la colonne "Ventes" dans le DataFrame `df`.
5. Rajouter un nouveau produit ["A1", 81, 23, "10MAD"] au dataFrame.
6. Ajouter une nouvelle colonne "Prix Total" au `df` rempli par **Ventes*Prix Unitaire**.
7. Remplacer les valeurs manquantes dans la colonne "Ventes" par la moyenne.
8. Sélectionner les données des produits B, C, D, E.
9. Sélectionner les données des ventes des produits A, D, E.
10. Modifier les stocks et les prix des produits A et E par des valeurs de votre choix.

Exercice 3

Etant donné le code html suivant :

```
<body>

  <h1>Contactez-nous</h1>

  <p>Pour répondre à vos questions, envoyez-nous un e-mail à l'adresse suivante : <a href="mailto:support@example.com">support@example.com</a>. </p>
  <p>N'hésitez pas à nous écrire si vous avez des préoccupations, des suggestions ou simplement pour dire bonjour!</p>

  <p>Pour des demandes spécifiques à nos professeurs :</p>

  <ul>
    <li class="professeur" id="prof1"><strong>Professeur 1:</strong> Contactez le professeur 1 à l'adresse <a href="mailto:prof1@example.com">prof1@example.com</a>.</li>
    <li class="professeur" id="prof2"><strong>Professeur 2:</strong> Envoyez un e-mail à <a href="mailto:prof2@example.com">prof2@example.com</a> pour toute question liée au professeur 2.</li>
    <li class="professeur" id="prof3"><strong>Professeur 3:</strong> L'adresse e-mail de contact du professeur 3 est <a href="mailto:prof3@example.com">prof3@example.com</a>.</li>
  </ul>

</body>
```

1. On souhaite utiliser web scraping pour extraire des informations pertinentes de ce document.
 - a. Donner le code Python permettant d'extraire et d'afficher le texte de tous les paragraphes *du document*.
 - b. Utilisez la méthode `find_previous_sibling()` pour extraire le contenu la troisième paragraphe *p* du document.
2. Utilisez une expression régulière pour extraire toutes les adresses email contenu dans un texte(**contenu_texte**).

Exercice 4

Pour créer un système de classification des fleurs, un grand ensemble d'images de différentes catégories a été rassemblé et stocké dans des dossiers spécifiques à chaque catégorie. Supposons que ces dossiers soient situés dans un répertoire appelé "dataset", à côté d'un autre répertoire appelé "Scripts" où divers scripts seront placés. Il est important d'utiliser des chemins relatifs pour accéder à ces dossiers, c'est-à-dire spécifier le chemin en partant du même endroit que les scripts.

1. Donner le code Python permettant de déplacer toutes les images directement dans le dossier "dataset" en les renommant par "i.j.jpg" avec $0 \leq i < \text{nbr_of_image}$ représentant la i-ème image et j représentant la j-ème classe.
2. Donner le code Python permettant de lire toutes les images(image par image) en utilisant `imread` de OpenCV et de les stocker après resizing dans une variable X de type numpy :
 - La variable X doit avoir la forme suivante : $(k, 225, 255, 3) = X.\text{shape}$, où k est le nombre des images, 255 représente la taille (le nombre de lignes et de colonnes) de chaque image après resizing, 3 représente la couleur.
 - Le code doit **également créer une variable Y (représentant les étiquètes des images)** qui contiendra l'indice de la classe où chaque image appartient, en commençant par l'indice zéro.
3. Donner le code python permettant d'encoder la variable Y. Vous pouvez utiliser **OneHotEncoder()** de `sklearn.preprocessing` ou bien de procéder par des boucles.
4. Fournissez le code permettant de sélectionner de 80% des données de X et Y, puis de les affecter respectivement aux variables `X_train` et `y_train`.