

Nom :
Prénom :
Groupe :
Durée : 1 h 00

Copiez le dataset `Ecommerce_Customers.csv` ensuite répondez aux questions ci-dessous en créant un Jupyter Notebook où vous effectuerez les analyses demandées. Écrivez également vos réponses au dos de cette feuille pour la remise finale.

*Veillez nommer votre fichier Jupyter Notebook en respectant la convention suivante : **nom_prenom.ipynb***

Q1 : Lecture des données

- Importez le fichier CSV dans un DataFrame à l'aide de pandas.
- Affichez les premières lignes du dataset pour vérifier la structure des données.

Q2 : Analyse descriptive des données

- Affichez les informations générales sur le dataset : dimensions, noms des colonnes et types de données.
- Fournissez un résumé statistique des colonnes numériques (moyenne, écart-type, etc.).

Q3. Préparation des données

- Vérifiez si le dataset contient des valeurs manquantes. Si oui, identifiez les colonnes qui contiennent des valeurs manquantes et le nombre de valeurs manquantes dans chacune d'elles.
- Quelles stratégies de traitement des valeurs manquantes pourriez-vous appliquer ? Justifiez votre choix pour chaque colonne ?
- Détecter les valeurs aberrantes pour la colonne `Length of Membership`, en utilisant la méthode de l'écart interquartile et la méthode graphique ?

Q4. Visualisation des données

- Créez des graphiques pour:
 1. Une distribution des dépenses annuelles moyennes des clients (`Yearly Amount Spent`).
 2. Un nuage de points pour observer la relation entre `Length of Membership` et `Yearly Amount Spent`. Interpretez le résultat ?
 3. La relation entre `Yearly Amount Spent` et `Time on App`? Que pensez-vous ?
 4. Existe-t-il une corrélation notable entre certaines variables numériques ? Calculez la matrice de corrélation entre les variables numériques et interprétez les résultats.