

Fiche client Une entreprise de vente au détail, Sell4All, a ouvert ses portes il y a 6 mois. Leur travail consiste à faire de la vente de vêtements d'occasion sur Internet et souhaitent intégrer leur première fonctionnalité d'IA dans le site Web. Cette fonctionnalité permettra au site Web actuel de suggérer automatiquement certains produits. Les suggestions seront basées sur la similitude des données d'un utilisateur avec d'autres en tenant compte de ses données démographiques, du montant dépensé sur le Web et des produits achetés.

Pour commencer ce projet, les premières étapes nécessaires sont les suivantes :

1. utiliser les données stockées sur les données démographiques et les dépenses des utilisateurs
2. explorer les données disponibles
3. nettoyer les données
4. enregistrer les données

### **Tes missions**

Tu viens d'arriver chez Sell4All en tant que Data Développeur Junior et au cours de ta première semaine, tu es chargé d'assumer cet ensemble de tâches nécessaires pour alimenter ensuite l'algorithme d'IA avec les données explorées. Avant de mettre la main à la pâte, tu dois installer et configurer ton environnement de développement :

1. Installer Python via Miniconda
2. Installer Jupyter Notebook
3. Installez les bibliothèques Python :
  - a. Pandas
  - b. Matplotlib

Ensuite, tu dois créer et exécuter un programme Python sur le Jupyter Notebook qui est capable de :

- Lire les données du fichier CSV 'dataset-sell4all.csv' qui contient les données démographiques et les dépenses des utilisateurs,
- Afficher des informations sur les 5 premières lignes du fichier CSV,
- Afficher un résumé technique des données disponibles dans le fichier CSV avec des informations telles que :
  - nombre de lignes
  - les colonnes du fichier CSV
  - les types de données des champs du fichier CSV
- Expliquer les détails affichés du résumé technique dans une cellule de démarque du bloc-notes Jupyter,
- Calculer la médiane et la moyenne des colonnes :
  - « Age »
  - « Customer spendings »
- recoder l'age en variable qualitative ne faisant des inter
- Question bonus: Calculer la médiane d'âge pour chaque pays
- Créer une visualisation des données du graphique à barres qui montre les dépenses des clients par pays,
- Nettoyer les lignes avec moins de 10 € de dépenses par client : supprimez toutes les lignes d'utilisateurs ayant dépensé moins de 10 € sur le site,
- Nettoyer les doublons : supprimer toutes les lignes qui apparaissent plus d'une fois dans les données,

- Écrire les données nettoyées dans un nouveau fichier CSV avec uniquement les colonnes suivantes :
  - « Country »
  - « Age »
  - « Gender »
  - « Customer spendings »

Le livrable est un fichier Jupyter Notebook contenant le travail effectué