

TD3

Exercice 1 : Analyse des séquences d'achats d'un site e-commerce

Un site e-commerce a collecté l'historique des sessions d'achats de ses clients. Chaque client représente une séquence, et chaque élément représente un ensemble de produits achetés dans un même panier (même date).

La base séquentielle est donnée ci-dessous :

Client (SID)	Séquence
C1	$\langle (A, B), (C), (D, E) \rangle$
C2	$\langle (A), (C, D), (E) \rangle$
C3	$\langle (B, C), (D) \rangle$
C4	$\langle (A, B), (D, E) \rangle$
C5	$\langle (A), (B), (C, E) \rangle$

On souhaite appliquer **PrefixSpan** pour extraire tous les motifs séquentiels fréquents ayant un support minimum = 2.

Exercice 2 : Exploration d'un dataset (EDA)

On te donne ce DataFrame :

ID	Age	Salaire	Ville	Malade
1	35	2500	Tunis	Non
2	NaN	3000	Sousse	Non
3	52	999999	Tunis	Oui
4	17	1200	NaN	Non
5	41	2700	Sfax	Oui

ID	Age	Salaire	Ville	Malade
6	350	1500	Sfax	Non
7	29	NaN	Tunis	Non

1. Afficher :

- Les 3 premières lignes
- Le type de chaque variable
- Le nombre de valeurs manquantes

2. Que remarques-tu sur les variables *Age* et *Salaire* ?

3. Quelles sont les colonnes quantitatives ? catégorielles ?

4. Pour les données manquantes :

- Remplacer les valeurs manquantes de Age par la médiane
- Remplacer les valeurs manquantes de Salaire par la moyenne
- Remplacer les valeurs manquantes de Ville par le mode (la valeur la plus fréquente)

5. Supposons un dataset avec **200 features** et **50 lignes**.

- Pourquoi ce dataset pose problème ?
- Proposer une méthode pour réduire la dimension

Exercice 3:

Pour chaque scénario suivant, identifier la tâche de data mining la plus appropriée (Classification, Régression, Clustering ou Détection de Déviation), et justifier brièvement :

- Une banque veut prédire le score de crédit (0-850) d'un demandeur de prêt
- Un hôpital veut détecter des motifs inhabituels dans les signes vitaux des patients pouvant indiquer une urgence médicale

- Une plateforme de streaming veut regrouper les utilisateurs selon leurs préférences afin de recommander du contenu