**Exercice 2 : Datamining**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Action** | **Situation\_famille** | **Revenu** | **Catégorie\_emission** | **Action** | **Heure** | **Durée\*** |
| A1 | Couple | 60000 | Meteo | Passif | 20h30 | 5 |
| A2 | Célibataire | 20000 | Musique | Zap | 7h30 | 15 |
| A3 | Famille | 35000 | Meteo | Eteint | 7h45 | 5 |
| A4 | Famille | 40000 | Emission\_pognon | Demande d’info | 18h | 20 |
| A5 | Célibataire | 35000 | Musique | Zap | 19h | 6 |
| A6 | Couple | 50000 | Emission\_pognon | Demande d’info | 18h | 15 |

Ce tableau donne la description des actions, le revenu indique le revenu annuel du foyer en Euros, la durée indique la durée écoulée entre le début de l’émission et l’action (en minute).

**Q1.** On se positionne dans l’espace tri-dimensionnel défini par les attributs «Catégorie\_emission», «Revenu» et «Durée».

* Lesquelles des visites parmi A1, A2 et A3 sont les plus similaires.
* Donner la description, dans ce même espace, du nuplet A123 centre de A1, A2 et A3.
* Quel est le problème rencontré ?. Comment y remédier ?

**Q2.** On se positionne dans l’espace défini par les dimensions « Revenus », « Catégorie\_emission » et « Durée ». On souhaite partitionner l’ensemble des visites en trois groupes.

Utiliser la méthode appropriée afin de répondre à cet objectif.