**Problème 1 :**

1. Veuillez interprétez la règle d’association numéro 3 (support, confiance, lift). D’après vous, comment vous caractériseriez cette règle et est-ce une règle intéressante pour vos futures promotions ? (Expliquez votre réponse).

* Support = 2% :

2% des clients achètent des gans hivernaux et un Manteau avec doublures

* Confiance = 82% :

82% des clients qui achètent des gans hivernaux vont aussi acheter un Manteau avec doublures

* Lift = 3.4 :

Les clients qui achètent des gans hivernaux ont 3.4 fois plus de chance d’acheter un Manteau avec doublures que les clients en général.

La valeur de lift est supérieure à 1.15, donc il s’agit d’une association positive et donc d’une règle intéressante à considérer pour de future promotion.

1. Un de vos partenaires marketing se pose beaucoup de questions sur la règle d’association numéro 2. Est-ce une règle intéressante pour la prochaine promotion ? (Expliquez votre réponse en quelques lignes).

La valeur de lift de la règle 2 est 1.1. Cette valeur est supérieure à 1 et donc il y a une association positive entre l’achat de Bracelet en Or et Chapeau de style MA, mais quand même cette valeur est inferieur a 1.15. Normalement, si la valeur de lift est inférieure à 1.15, l’association n’est pas forte et il ne vaut pas la peins de créer une promotion. Plus la valeur de lift est proche de 1, moins la promotion sera performante. En fait, en comparant le lift de règle 2 avec le lift d’autre règle on remarque que le lift de règle 2 est inferieur aux tous autres lifts, donc c’est le pire choix pour faire une promotion basée sur règle 2.

1. Parmi l’ensemble des règles d’associations calculées, veuillez choisir la meilleure règle pour la prochaine promotion et veuillez décrire en quelques lignes en quoi consisterait cette promotion.

La règle 9 est la meilleure règle car sa valeur de lift est plus grande que 1.15 et en même temps son taux de support est 27% qui est le support les plus élevés des règles. C.-à-d. que déjà 27% des clients achètent des Londres. Selon cette règle, les clients qui achètent des Londres ont 1.9 fois plus de chance d’acheter un parapluie et de faire achat en ligne que les clients en général. La confiance de règle 9 est aussi élevé. En autres termes, 82% des clients qui achètent des Londres vont aussi acheter un parapluie et ils font achat en ligne. Donc, il est un bon investissement de créer une promotion pour ensemble de Londres et parapluies.

1. Vous possédez un petit échantillon sur lequel vous voulez effectuer des règles d’association. Cet échantillon comporte des transactions uniques indiquant la présence d’un article (oui ou non). De plus, il comporte la colonne Montréal indiquant si la personne habite à Montréal. À l’aide du tableau, veuillez déterminer le support, la confiance et le lift des règles d’association suivantes :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| tansaction ID | manteau | chaussures | pantalon | chemises | ceinture | montreal |
| #v25654 | oui | non | non | non | non | non |
| #v54542 | non | oui | oui | oui | non | non |
| #v57545 | non | non | oui | oui | non | oui |
| #v86745 | non | oui | oui | oui | oui | non |
| #v64542 | oui | oui | oui | oui | oui | non |
| #v85163 | oui | non | oui | oui | non | oui |
| #v54258 | oui | non | oui | non | non | non |
| #v85963page2image56518784 | oui | oui | oui | non | non | oui |
| #v54212page2image56510912page2image56509184page2image56514368page2image56509568page2image56518208page2image56513216page2image56503360page2image56503552 | oui | non | oui | oui | non | non |
| #v54265 | oui | non | non | oui | oui | oui |
| #v79745 | oui | oui | oui | oui | oui | oui |
| #v65342 | non | oui | oui | non | non | non |

# de transactions totales = 12

1. Chaussures 🡪 Pantalon

X=2, Y=3

|  |  |  |  |
| --- | --- | --- | --- |
| Support X | Support Y | Confiance (X => Y) | Lift |
| 50% | 83% | 50% | 0.6 |

1. Pantalon 🡪 Chaussures

X=3, Y=2

|  |  |  |  |
| --- | --- | --- | --- |
| Support X | Support Y | Confiance (X => Y) | Lift |
| 83% | 50% | 50% | 1.0 |

1. Montréal & Manteau 🡪 Chemises

X=1&6, Y=4

|  |  |  |  |
| --- | --- | --- | --- |
| Support X | Support Y | Confiance (X => Y) | Lift |
| 33% | 67% | 25% | 0.4 |

**Problème 2 :**

1. Les directeurs vous posent quelques questions afin de mieux comprendre le contexte :
   1. Quel est le produit qui se retrouve le plus au sein des transactions ?

StockCode=85123A, Description= cream/white hanging heart t-light holder

* 1. Quelle est la provenance engendrant le plus de transactions ?

"United Kingdom"

* 1. Quel est le produit le plus rentable au sein des transactions ?

StockCode= 23843, Description= paper craft , little birdie

* 1. Quel est le client ayant fait le plus de visites ?

CustomerID= 12748

1. À l’aide des règles d’association et de l’algorithme Apriori, veuillez déterminer 3 règles que vous pensez intéressantes pour la compagnie. Veuillez donner une interprétation à ces règles (sur toutes les mesures) et justifier en quoi elles sont intéressantes. Enfin veuillez proposer une action avec chacune d’entre elles.

Help :

Pour les variables explicatives (X) quantitatives, ça va être tjrs le Gini. Il y a un calcul supplémentaire qui va dire, pour cette variable-là c quoi le meilleur point de coupure. Et le point de coupure va être aussi calculé à l’intérieur de Gini etc. pour par exemple décider que le meilleur point de coupure est 30. Puis aux finales, on va dire est-ce que la valeur est < 30 ou > 30. Mais pour faciliter (comme dans le devoir) on peut facilement basé sur la moyenne.

Est-ce que la variable cible (Y) doit être tjrs qualitative ? on a 2 concepts : ADC et ADR

* Arbre de décision 🡪 classification 🡪 le fait de classer 🡪 on induit que Y est qualitative. Indice de Gini se fait que sur les variables qualitatives 🡪 évidement quand on parle de l’arbre de décision c’est sur les variables qualitatives nominales.
* Arbre de régression 🡪 Y prendre des variables quantitatives cette fois-ci. Ça c’est l’objectif de devoir. Il n’a rien avec la régression linéaire. C exactement même concept de l’arbre de classification sauf qu’au lien de Gini et entropie et gain informationnel, il y qqch que vous connaissez bien, vous allez devoir travailler avec pour arriver à faire le même concept d’homogénéité etc.

Devoir : Ex2 sur R seulement = méthode Kimen on n’a pas vu. Faire des recherches. k-mean apprentissage non-supervise, kinin est supervisé, utilisé pr prévision. Nettoyage 🡪 stack overflow

Devoir future : Trouver une commande pour passer de data frame a transaction

jusqu’a 20 juin

une semaine avant le finale

20%

4 problems:

2 : règles d’asso - application de réglé d’asso avec R

1: réglé de distance (calcul de distance - un algo)

1: arbre de regression (arbre de classification)