Thierry VALLAUD

# Estimation du potentiel individuel de chiffre d'affaires

en utilisant des données issues
d'une base de données clients

SOCIO logiciels
l'intelligence des données...

Thierry Vallaud

# Estimation du potentiel individuel de chiffre d'affaires en utilisant des données issues d'une base de données clients

Utilisation d'une technique de classification automatique
pour déterminer la valeur client

## Auteur

THIERRY VALLAUD est responsable du data mining
et de la modélisation de SOCIO Logiciels www.socio.fr
Il a publié plusieurs ouvrages et articles sur le data mining, la fidélisation,
les analyses de bases de données, la Business Intelligence (BI)

L'auteur tient à remercier
– Daniel Larose de la CCSU pour la relecture de la version anglaise de ce document ;
– Jean-Paul Aimetti du CNAM, professeur titulaire de la chaire marketing pour ses conseils et la relecture de la version française de ce document.

## Résumé

Cette étude met en lumière une méthode pour déterminer le potentiel client individuel de chiffres d'affaires, en se basant uniquement sur des données présentes dans la base de données clients de l'entreprise : des informations descriptives sur ces clients et des enregistrements de leurs transactions.

Nous définissons le potentiel client comme le chiffre d'affaires supplémentaire qu'une société donnée peut atteindre avec ses clients actuels.

Dans le but de calculer avec succès ce potentiel dans une base de données de grande taille avec des multiples variables, nous proposons de regrouper ensemble les clients « qui se ressemblent » (que nous appelons « clones ») en utilisant une technique appropriée de classification : les réseaux de Kohonen.

Puis nous divisons chaque groupe de clones par une approche de césurage particulière (déciles et médiane des déciles) qui nous permet d'obtenir un potentiel réaliste par client.

C'est l'association de ces deux approches, classification de Kohonen puis césurage du chiffre d'affaires des groupes de clones qui fait l'originalité et la validité de la méthode des clones.

Cette méthode est appliquée à un ensemble de données réelles, et plusieurs techniques sont utilisées pour vérifier la stabilité des groupes obtenus. Le potentiel est démontré de manière empirique : une application pratique à une base de données de 5 millions de clients d'un des principaux distributeurs alimentaires français.

## Sommaire

## Le contexte

La plupart des sociétés, quel que soit leur secteur d'activité en BtoB ou en BtoC aimeraient connaitre le potentiel de leurs clients en termes de chiffres d'affaires. Déterminer « un potentiel client » signifie identifier le chiffre d'affaires incrémental qu'une entreprise donnée pourrait générer sur ses clients actuels.

Les modèles de potentiel clients existent et sont principalement basés sur la détermination de la valeur client pendant la durée de vie de ce dernier : LTV (LTV = Life Time Value) *(Bénavent et Crié ; Berger et Nasr 1998; Dwyer 1997; Venkasten, Rajkumar et Kumar 2004).*
Au-delà de ces modèles, d'autres approches existent qui estiment la part de dépense du client *(Cooil et al. 2007 ; Yuxing Du et al. ; Keimingham et al. 2007)*.
Enfin d'autres modèles économétriques de potentiel ont été crées mais ils sont basés sur des données externes à la base de données clients *(Plastria 2001, Huff 2003, Reilly 1931*).

La consommation du client (sa valeur totale) représente la consommation de tous les achats d'un produit donné, sur sa durée de vie, par un consommateur donné. Elle se nomme la Valeur Totale du Client (VTC). Par exemple, au cours de sa vie la valeur totale d'un client pour un distributeur, est la somme de tous les achats qu'il aura fait dans le magasin de ce distributeur durant sa vie.

Il est possible d'estimer la consommation d'un consommateur sur un marché pour une marque et un produit donnés dans une catégorie de produits. Au cours de sa vie, le consommateur va consommer *n* marques $m_1, m_2 ..... m_n$. Le rapport entre la consommation totale d'une de ces marques $m_n$ et la consommation totale des marques, est le taux de nourriture de la marque $TnVTC_{mn}$ sur la durée de vie du client (graphique 1).

$$VTC_m = VTC_{m1} + VTC_{m2} + ....VTC_{mn} \qquad \text{Taux de nourriture de } TnVTC_{mn} = \frac{VTC_{mn}}{VTC}$$

La différence, notée delta dans les graphiques, entre la consommation totale du consommateur sur la catégorie de produit de la marque étudiée $VTC$ et les consommations totales de la marque $m_1$, $VTC_{m1}$ correspond à la consommation totale aux marques concurrentes $VTCC$ (Graphique 2).

$$VTCC = VTC - VTC_{m1} \text{ donc } VTCC = VTC_{m2} + \ldots . VTC_{mn}$$

Selon les stimuli marketing de la marque $m_1$, le client va prendre une part $TauxC_{m1}$ de ce delta à la concurrence et/ou accroitre sa consommation sur le total du marché :

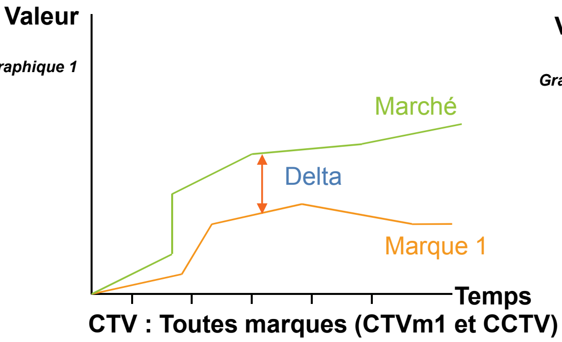
L'accroissement de sa consommation totale marché

Soit : $VA_{m1} =$ La valeur actuelle pour la marque 1

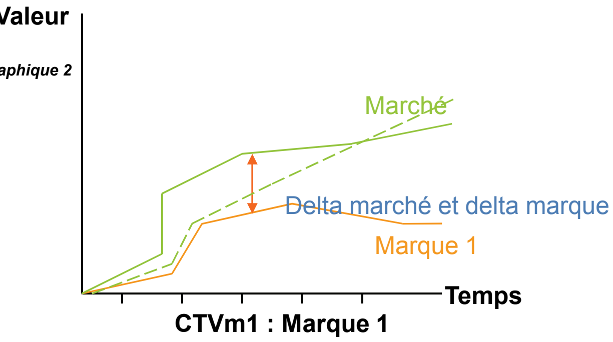
$$VTC_{m1} = VA_{m1} + .TauxC_{m1} \times VTCC + \text{Taux d'accroissement de la consommation}$$

Les clients d'un distributeur vont consommer dans certains magasins concurrents et peuvent accroitre leur consommation totale aux différents distributeurs.

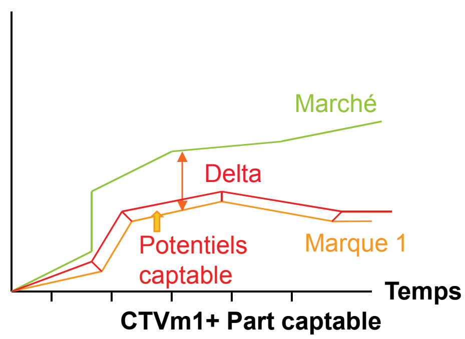
*Graphique 1*



*Graphique 2*



*Graphique 3*



Donc, le potentiel théorique d'un client est la consommation totale sur sa durée de vie $CTV_{m1}$ qui peut être atteinte par ce client ; il peut être estimé par les moyens du modèle économétrique ci-dessous.

$$VTC_{m1} = VA_{m1} + .TauxC_{m1} \times VTCC + \text{Taux d'accroissement de la consommation}$$

Où :

$TauxC_{m1} =$ Part de la consommation prise aux concurrents (Graphique 3).
C'est la part de marché « captable » sur les concurrents
$\text{Taux d'accroissement de la consommation} =$ L'accroissement de sa consommation totale

Le potentiel captable du client correspond à ce que la marque a déjà atteint, augmenté de ce que le client pourra consommer en plus et/ou ce qui sera pris à concurrence. Ce potentiel atteignable peut être estimé de deux façons : en utilisant un modèle économétrique, ce qui nécessite des données exogènes à la base de données ou d'une autre façon, en utilisant uniquement des données de la base de données clients, dans notre exemple la base des porteurs de la carte de fidélité de l'enseigne : la méthode des « clones ».

Une marque donnée ne peut capter qu'un pourcentage donné du potentiel théorique ; selon plusieurs travaux de recherche dont ceux de Berend Wierenga et Gerrit en 2000. Le potentiel théorique est la consommation totale du consommateur.

Certains chercheurs en marketing ont montré qu'une marque peut accroitre son taux de nourriture actuel à une marque de 30%, l'écart type moyen constaté des taux de nourriture. Au-dessus de ce taux, le consommateur perçoit un changement de son comportement de consommation et essaie alors d'y résister.

Au-dessus de 30% d'accroissement il y a trop de modifications de l'ensemble de choix du client[1] (Bremer et Joyce, 1988). Ce sujet a déjà été traité dans une de nos précédentes recherches (Vallaud, 2003).

Les approches les plus avancées de détermination des potentiels essaient de déterminer la proportion de chiffre d'affaires incrémental qui peut être atteinte par une société, en se basant uniquement sur les seules données issues de la base de données clients de la société. Ces approches calculent un potentiel client par client mais évidemment doivent être en phase avec les macros données agrégées du marché.

[1] L'ensemble de choix est un ensemble fini pour une catégorie de produits donnée qu'un client a l'esprit avant de faire un achat.

## Notre sujet de thèse

L'objectif est de travailler avec des algorithmes de classification[2] [*Lerman 1970, Dorofeyuk 1971, Borko et al. Bernick 1963,* Two Steps *(Tan et al. 1997),* K means *(Hartigan et al. 1979, Fang et al. 1982),* SOM *(Teuvo Kohonen 1988, Vesanto 1997, Kaski 1997), etc.*], sur de bases de données de plusieurs millions d'enregistrements de société commerciales (grands distributeurs, opérateurs téléphoniques, fournisseurs d'accès à Internet, sociétés de marketing direct, etc.).

Nous utilisons ces modèles dans le but de déterminer un potentiel client en utilisant une méthode que nous appelons « méthodes des clones » dans laquelle les clients qui se ressemblent le plus, sont considérés comme des clones, et sont supposés avoir le même potentiel.

Nous avons accès à diverses bases de données pour notre processus méthodologique. Dans ce document, nous allons faire un test empirique de notre méthode sur des données clients provenant uniquement de la base de données d'un grand distributeur français.

Au-delà de notre brève présentation du contexte, nous avons à aborder deux sujets majeurs :

- Le calcul du potentiel sur la valeur client en marketing et ces différents indicateurs associés : LTV, taux de nourriture, part de marché captable, etc.
- Les modèles mathématiques qui permettent à des individus similaires d'être regroupés dans des groupes homogènes : les techniques de classification.

Le champ d'investigation est multidisciplinaire avec, en mineur, les études marketing et, en majeur, la discipline des statistiques, du data mining et de la classification.

[2] Les SOM (Sef Organizing Maps, cartes auto organisatrices) appartiennent aux méthodes de classification, souvent appelées « typologies ». Ces méthodes sont dites automatiques non supervisées. Non supervisée veut dire que l'on ne définit pas a priori une variable cible à prédire.

## L'application précise du modèle

Une part importante des objectifs de notre recherche, est de tester séparément différentes techniques, et, si possible, conjointement, pour s'assurer que les groupes formés sont des groupes de clones homogènes. À delà du choix des modèles, une part importante de notre étude implique de définir les variables les plus informatives et donc une topologie des données en entrée du modèle, qui soit bien adaptée à celui-ci. Le but ici, est d'obtenir, des résultats les plus convergents et les plus pertinents possibles.

Une autre partie importante de notre étude est de sélectionner les méthodes mentionnées ci-dessus et de valider ces choix. L'objectif est de trouver une méthode qui converge suffisamment et qui puisse être validée par l'ensemble des approches ci-dessus. La modélisation devient alors une association de plusieurs modèles.

- La réduction des dimensions pour choisir les variables avec un très grand nombre de groupes et des écarts de valeurs important ;
- La projection des variables actives et passives[3] dans les groupes,
- La réallocation des groupes par un algorithme supervisé,
- La validation de la connectivité de « super classes » par d'autres méthodes de classification non supervisées,
- Une vérification empirique via un panel externe comme celui géré par Nielsen ou TNS Sofres[4] qui représente la « réalité marché » du potentiel.

Le modèle définitif est créé en utilisant une plateforme logicielle standard du marché : la version française de PASW Modeler de SPSS/IBM (ex. Clémentine).

La contribution scientifique est :

- L'apport méthodologique pour sélectionner les modèles de classification et valider ces choix ;
- Une application sur des données réelles, validée par un business cas réel : calculer un vrai potentiel captable.

[3] Les variables actives sont utilisées pour construire les groupes eux-mêmes ; les variables passives sont des variables descriptives pour expliquer les groupes.

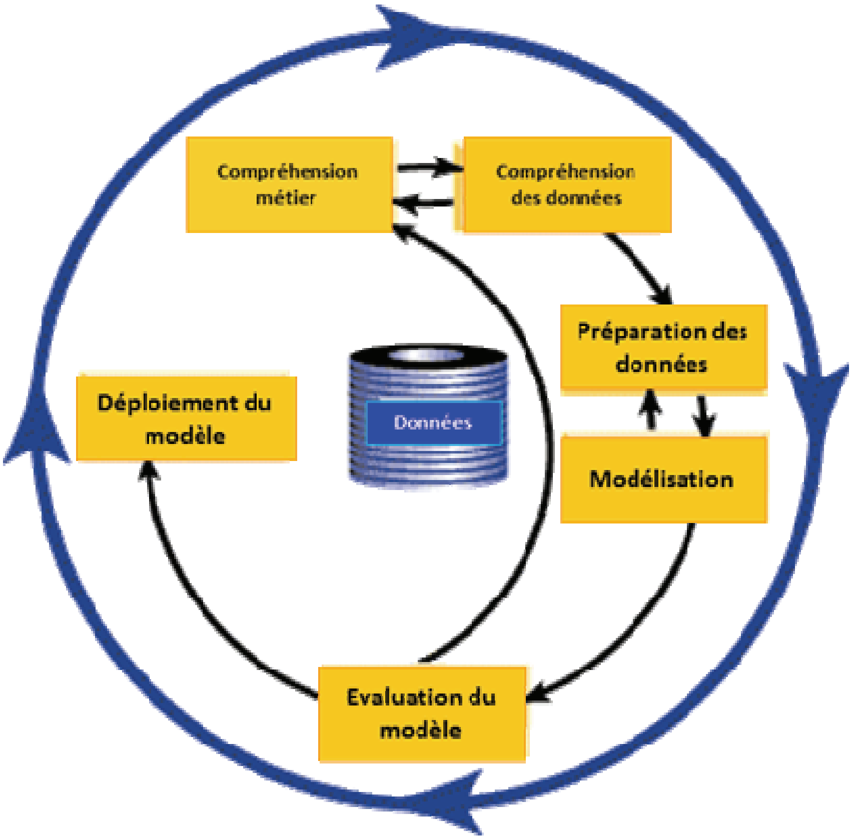
[4] Nielsen et TNS sont des sociétés d'études de marché qui fournissent des panels dans lesquels leurs membres scannent leurs achats. Ces panels peuvent être croisés avec des bases de données clients pour mesurer les effets du marketing mix.

## Les questions de recherche

- Pouvons-nous utiliser une technique de classification pour déterminer les consommateurs qui sont similaires entre eux et donc définir un potentiel réaliste en termes de chiffre d'affaires pour ces clients ?
- Pouvons-nous développer une méthode pérenne, que l'on puisse répliquer ?
- Comment pouvons-nous valider la stabilité des classes obtenues ?

## Le processus de data mining utilisé

Nous utilisons le processus projet standard Cross-Industry Standard Process pour le Data Mining (CRISP)[5] qui nous permet de guider notre approche d'analyse des données. Le processus standard CRISP est constitué des étapes suivantes :



### *La compréhension des données*

Nous allons travailler sur 5 373 026 individus issus de la base de données d'un grand distributeur français. Nous avons le détail de tous les tickets de caisses sur une période de 12 mois, de janvier 2006 à décembre 2006.
Pour le processus de validation externe, nous utilisons aussi une étude de marché disponible sur le marché Français : *Référenseigne 2006* de *TNS Sofres*[6]. Cette étude nous donne le taux de nourriture des principaux distributeurs Français[7].

[5] http://www.crisp-dm.org/
[6] Référenseigne est une étude de marché monographique qui est réalisée sur la marche de la distribution alimentaire en France depuis 10 ans par TNS Sofres intégré en 2009 au groupe Kantar.
[7] http://www.secodip.fr/worldpanel/htm/dossier_presse/tns-plusloin.asp/

### *La préparation des données*

Cette étape consiste à nous familiariser avec les données de la base des membres du programme de fidélité de l'enseigne, dans la but de bien préciser la structure de la base en se basant sur le dessin d'enregistrement, le niveau de complétude des champs contenus dans les fichiers, et aussi l'origine et la nature des données. Chaque champ est vérifié pour s'assurer qu'il n'affecte pas la stabilité du modèle. Nous avons fait un audit et une AED (Analyse Exploratoire des Données) en deux fois, seule la seconde AED est présentée dans ce document.

Sont vérifiés :

- La structure de la base de données
- L'origine et la nature des données (sociodémographiques / consommation)
- La possibilité de réaliser des analyses croisées des données (par marque/ rayon/famille de produits, etc.…)
- La périodicité des données
- L'historique des données
- La complétude des données

Les principaux processus appliqués sur les données de la base incluent les étapes suivantes :

- Contrôler et valider le format des variables
- Recoder et corriger certaines variables appelées « variables aberrantes », parce que contenant des valeurs inattendues ou extrêmes par rapport à la moyenne (plus de 3 écarts types)
- Créer des agrégats spécifiques et utiles pour notre détermination des potentiels (chiffre d'affaire total par famille de produits, fréquence annuelle de visites, panier moyen d'achat…)
- Analyser les corrélations entre la variable cible (le chiffre d'affaire) et d'autres variables (des critères socio démographiques, la fréquence d'achat…) dans le but de vérifier s'il existe des relations de dépendance entre ces variables.
- « Géocoder », étape utile pour enrichir les profils de certaines variables socio démographiques de l'INSEE[8] via les IRIS[9] (unités françaises géographiques spécifiques)
- Calculer les distances entre le client et le point de vente (zone de chalandise)

[8] L'INSEE en France (Institut national de la statistique et des études économiques). Il collecte et publie des informations sur l'économie et la société française en menant régulièrement un recensement national. Situé depuis peu à Strasbourg il est la branche française de Euro Stat, le système des statistiques européen. L'INSEE a été crée en 1946.

[9] L'IRIS est une unité géographique liée aux données du recensement

L'analyse a été réalisée sur un chiffre d'affaire sur 12 mois glissants sur l'ensemble de l'historique, pour s'assurer que le modèle soit le plus fiable possible. Dans ce type d'analyse basée sur la compréhension des comportements d'achat des clients, plus l'historique des données est important et homogène, plus le modèle est stable et prédictif.

Dans ce document, nous n'avons inséré que quelques exemples de la préparation des données, car notre démonstration est centrée sur le modèle et les résultats. Des détails de la seconde AED sont présentés en annexe 2 (p.39).

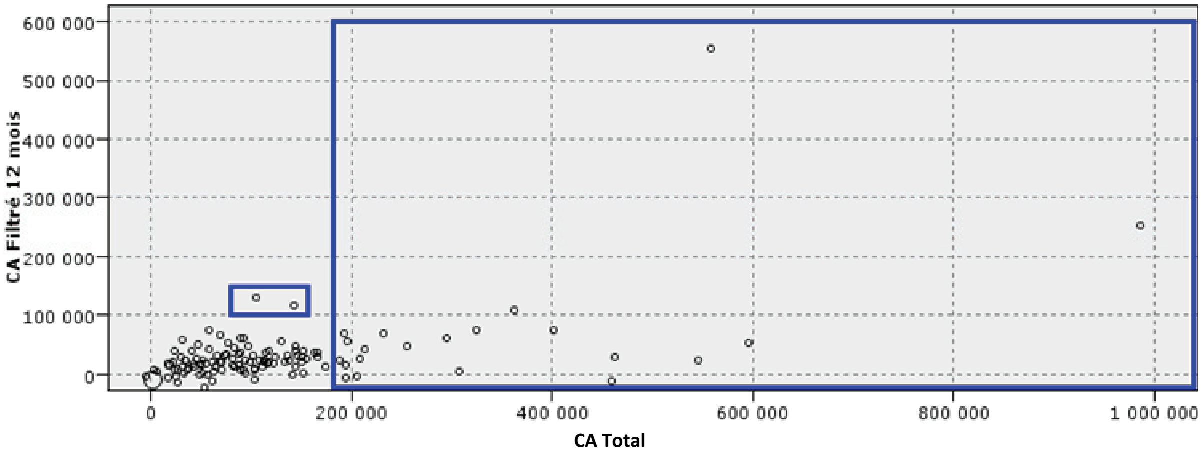
**Les variables en entrée sont présentées dans le tableau ci dessous :**

| **Variables** | **Description** |
|---|---|
| Identifiant client | Pour lier les tables entre elles sur un même client |
| Ratio autre | " |
| Ratio bazar | " |
| Ratio BOF/APF | " |
| Ratio charcuterie LS | " |
| Ratio clients animaux | " |
| Ratio client bébés | " |
| Ratio clients boucherie | " |
| Ratio boulangerie | " |
| Ratio charcuterie | " |
| Ratio diététiques bio | " |
| Ratio fromage | " |
| Ratio fruits et légumes | " |
| Ratio poisson | " |
| Ratio surgelés | " |
| Ration vins | " |
| Ratio DPH | " |
| Ratio épicerie | " |
| Ratio liquide | " |
| Ratio textile | " |
| Ratio ultra frais | " |
| Ratio volaille | " |
| Ratio premier prix | " |
| Ratio marque distributeur 1 | " |
| Ratio marque distributeur 2 | " |
| Nombre d'enfants au foyer | |

| | |
|---|---|
| CA filtré 12 mois | Chiffre d'affaires diminué de quelques dépenses hors mag |
| CA total | Chiffre d'affaires total |
| CA promo annuel | Chiffre d'affaires annuel réalisé sous promotion |
| Nb de points transformés sur 12 mois | Nombre de points « promotionnels » utilisés sur une année |
| CM transformé sur 12 mois | Part du chiffre d'affaires où le client a utilisé des points promotionnels |
| Cumul nb BA pris | Cumul des bons d'achat que le client a utilisé |
| PMG 12 mois[10] | Segmentation |
| RFM 3 mois[11] | Segmentation |

**Identification des clients aberrants :**

Nous avons identifié et éliminé de notre étude quelques clients avec des comportements anormaux sur deux variables portant sur le chiffre d'affaire : le « CA filtré 12 mois[12] » et le « CA Total [13]».



Nous utilisons uniquement ces deux variables pour la détection des clients aberrants, parce qu'elles sont très constitutives du potentiel lui-même.

[10] Segmentation PMG (Petit, Moyen, Gros) divise les consommateurs en fonction de leurs chiffres d'affaires.

[11] Segmentation RFM (Récence, Fréquence, Montant) est une segmentation classique en marketing.

[12] Le chiffre d'affaires filtré (CA filtré) c'est le chiffre d'affaires cumulé sur les 12 derniers mois moins les réductions de chiffre d'affaires faites par les achats sous promotion.

[13] Le chiffre d'affaires total c'est le chiffre d'affaires cumulé sur les 12 mois sans enlever les réductions de chiffre d'affaires faites par les achats sous promotion.

**Discrétisation :**

Nous avons discrétisé quelques variables importantes et étudié leur dispersion. Ci-dessous un exemple en quintiles sur la variable CA Total.

| | Effectif | % |
|---|---|---|
| 0 | 3 985 | 4,0% |
| >0 < 280,52€ | 16 014 | 16,0% |
| >=280,52 < 1 538,50€ | 19 999 | 20,0% |
| >=1 538,50€ < 4 4648,2€ | 19 999 | 20,0% |
| >= 4 4648,2€ < 12 155,82€ | 19 999 | 20,0% |
| >= 12 155,82€ <=805 990,68€ | 20 000 | 20,0% |
| NR | 4 | 0,004% |
| Total | 100 000 | 100% |

| CA total en cumul | | | | | |
|---|---|---|---|---|---|
| Moyen | Minimum | Maximum | Ecart type | Médiane | Somme |
| 7 031,88 € | 0 € | 806 991 € | 10 377,17 € | 2 757,18 € | 703 187 673 € |

Nous fournissons une AED complète dans l'annexe 2 (p.39) avec une analyse descriptive avec des tableaux et des graphiques, estimation des corrélations, etc.

## Les modèles de classification et la détermination du potentiel client

**Le processus de modélisation est divisé entre trois grandes phases :**

(1) La méthode de classification elle-même, (2) le calcul des niveaux d'évolution et (3) le calcul du potentiel individuel :

**1° La méthode de classification** : puisque ce modèle doit être appliqué sur de très grandes bases de données avec un grand nombre de variables et d'enregistrements, les SOM (Self Organizing Map ou « carte auto organisatrice ») semblent particulièrement bien adaptés (Kohonen, 1988) :

- Les réseaux de Kohonen permettent d'obtenir des groupes très homogènes et très stables avec de multiples individus et variables,
- Les réseaux de Kohonen autorisent des relations non linéaires complexes entre beaucoup de variables sur beaucoup d'individus,
- Les réseaux de Kohonen prennent bien en compte les valeurs manquantes

### *Méthode des réseaux de Kohonen*

Les réseaux de Kohonen sont une forme de réseaux de neurones qui utilisent un type de carte auto organisatrice qui elle-même représente une catégorie particulière de réseaux de neurones.

La méthode d'analyse est une méthode de classification. Son principal avantage est de convertir un signal d'entrée avec de nombreuses dimensions en une simple carte discrète, avec peu de dimensions. Les réseaux de Kohonen sont une méthode non supervisée, aucune variable cible ne doit être définie.

Les réseaux de Kohonen utilisent la méthode d'apprentissage de Kohonen. Soit un ensemble de m variables pour le énième enregistrement qui servent d'entrée à un vecteur $x_n = x_{n1,} x_{n2,} \dots x_{nm}$ dans l'ensemble des poids en cours m pour un nœud sortie donnée j devant être un vecteur de poids $w_j - w_{1J,} w_{2j,} \dots w_{mj}$. Dans l'apprentissage de Kohonen, les nœuds dans le voisinage des nœuds gagnant ajustent leur poids en utilisant une combinaison linéaire des vecteurs d'entrée dans le vecteur des poids en cours (current) :

$$w_{ij} = w_{ijcurrent} + \eta(x_{ni} - w_{ijcurrent})$$

où $\eta, O < \eta > 1$, représente le taux d’apprentissage. Kohonen précise que le taux d’apprentissage doit être une fonction décroissante des cycles d’apprentissage (mené sur l’ensemble des données).

Après chaque itération, il vérifie la fiabilité de ses regroupements antérieurs.

- C’est évidemment un processus assez long quand on l’applique sur un grand nombre d’individus avec beaucoup de variables ; il faut en tenir compte dans le choix de la puissance de calcul de l’ordinateur et éviter de relancer plusieurs fois le calcul.

- Chaque client est alloué à un groupe déterminé dont il est le plus proche sur toutes les variables qui le caractérisent.

Les réseaux Kohonen montrent trois processus caractéristiques :

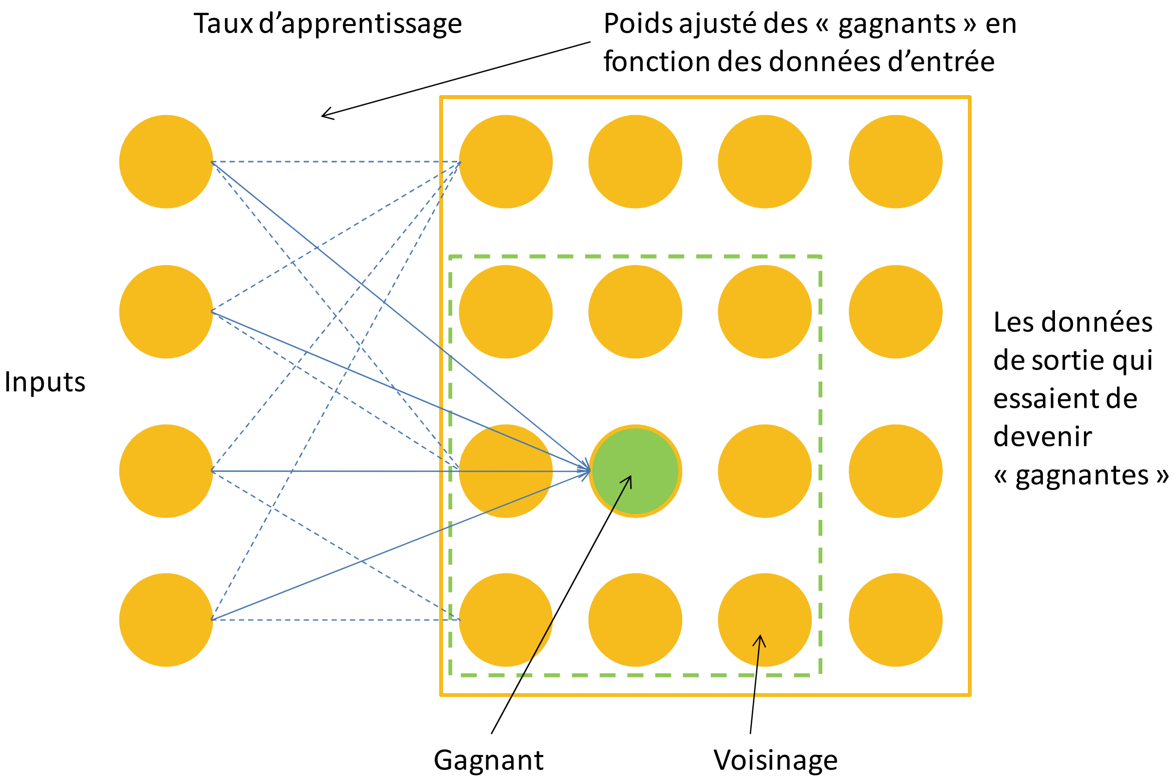
1° Compétition : Les nœuds en sortie sont en concurrence entre eux pour produire la meilleure valeur pour une fonction de score donnée, en général la minimisation des distances euclidiennes.

2° Coopération : Les nœuds gagnants deviennent alors les centres des voisinages des neurones « excités ».

3° Adaptation : Les nœuds dans le voisinage du nœud gagnant, participent à l’adaptation, c’est-à-dire l’apprentissage. Les poids de ces nœuds sont ajustés dans le but d’améliorer la fonction de score.

**L’architecture du réseau :**

Chaque neurone de la carte est lié à tous les autres neurones. Chacun d’eux reçoit une copie complète d’un vecteur en entrée.

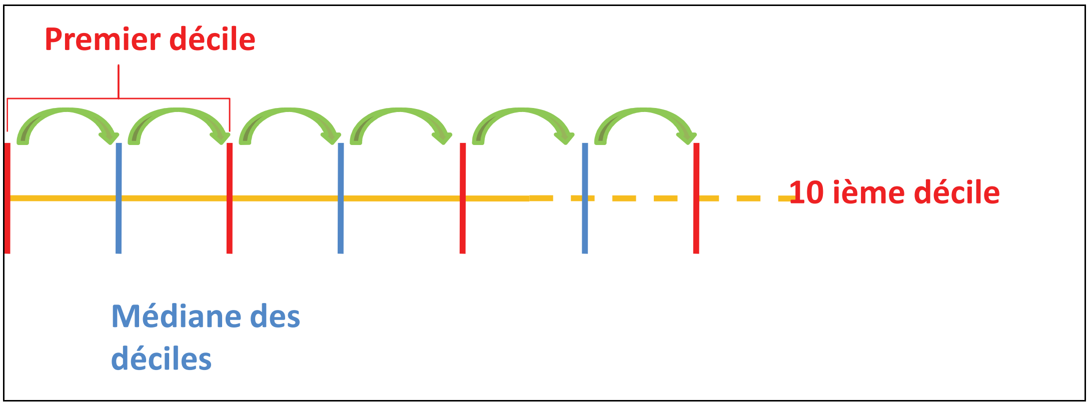

Taux d'apprentissage
Poids ajusté des « gagnants » en fonction des données d'entrée
Inputs
Les données de sortie qui essaient de devenir « gagnantes »
Gagnant
Voisinage

En représentant cartographiquement l'analyse nous pouvons évaluer la similarité entre les groupes. Deux groupes qui sont proches sur le graphique ont des caractéristiques similaires.

Le but est de trouver une méthode :

- Qui représente le meilleur compromis entre beaucoup de groupes, mais qui s'assure aussi que les groupes sont les plus précis possible avec des clients bien homogènes à l'intérieur, et que les groupes sont les plus différents possibles entre eux.
- Qui nous permet d'obtenir un potentiel client réaliste avec des groupes qui sont stables intrinsèquement.

**2° Le calcul des taux « d'évolution » :** « L'évolution » est le petit saut de chiffre d'affaire qu'un consommateur doit atteindre pour être regroupé avec les clients qui lui ressemblent le plus, sur toutes les variables sélectionnées par le modèle mais qui ont un chiffre d'affaire plus élevé que lui. Cela nécessite une méthode de calcul basée sur la division de chaque décile d'une classe de « clones » par la médiane du décile.



Nous retenons la méthode des déciles qui permet de prendre en compte des variations assez significatives de chiffres d'affaires.

Les individus dans un groupe ne doivent pas avoir un écart important à franchir dans le but d'obtenir une estimation réaliste du potentiel[14] : l'accroissement du chiffre d'affaire potentiel qui pourrait être réalisé après la mise en place des actions marketing appropriées. Nous allons essayer de justifier ce calcul par des moyens méthodologiques. Cette étape va nous donner les taux d'évolution dans les classes de Kohonen.

**3° Le calcul du potentiel client individuel :** une fois que les taux sont calculés, nous allons calculer, pour chaque client, un potentiel client individuel à atteindre. Ce calcul nécessite des ajustements spécifiques : pour tous les clients avec un taux d'évolution de potentiel supérieur à 100 %, leur potentiel représente plus du double de leur CA actuel. Ils se voient allouer le taux d'évolution de potentiel moyen de tous les groupes, sans inclure évidement ceux à plus de 100 %.

[14] Exemple: Le client A a un chiffre d'affaires actuel de 1 000 €. Le client A appartient au premier décile pour un groupe dans lequel tous les clients se ressemblent le plus. Le chiffre d'affaire maximum du client de la limite supérieure de ce décile est de 1 200 €. Donc le potentiel est la différence entre les 1 200 € du consommateur max et les 1 000 € du consommateur A : 200 € soit 20 %

# Le développement du modèle

### *1. Les objectifs et la méthodologie*

Pour compléter les segmentations basées sur le chiffre d'affaire futur, la segmentation PMG *(Brusset, 2005)* et la segmentation RFM *(McCartya et Hastak 2007, Chen et al., 2008*), nous calculons un chiffre d'affaire potentiel pour chaque client dans la base de données du programme de fidélisation du distributeur.

Ce score est basé sur une approche itérative nous permettant de prédire la propension à consommer des clients, dans le but de déterminer le chiffre d'affaire potentiel futur.
L'approche consiste à regrouper ensemble les clients qui se ressemblent le plus entre eux, selon les variables présentes en base : variable socio démographiques variables de consommation…

Pour le calcul, nous allons utiliser les données de consommation enregistrées sur une période de 12 mois (de janvier 2006 à décembre 2006).

Les variables utilisées dans le modèle sont celles que nous avons décidé de conserver à l'issue de l'étape de préparation.

| Statut familiale et données de consommation |
|---|
| Nombre d'enfants au foyer |
| CA filtré 12 mois |
| CA total |
| CA promo annuel |
| Nbr de points transformés sur 12 mois |
| CM transformé sur 12 mois |
| Nb de bons de réduction utilisés |
| PMG 12 mois |
| RFM 3 mois |
| Situation familiale |

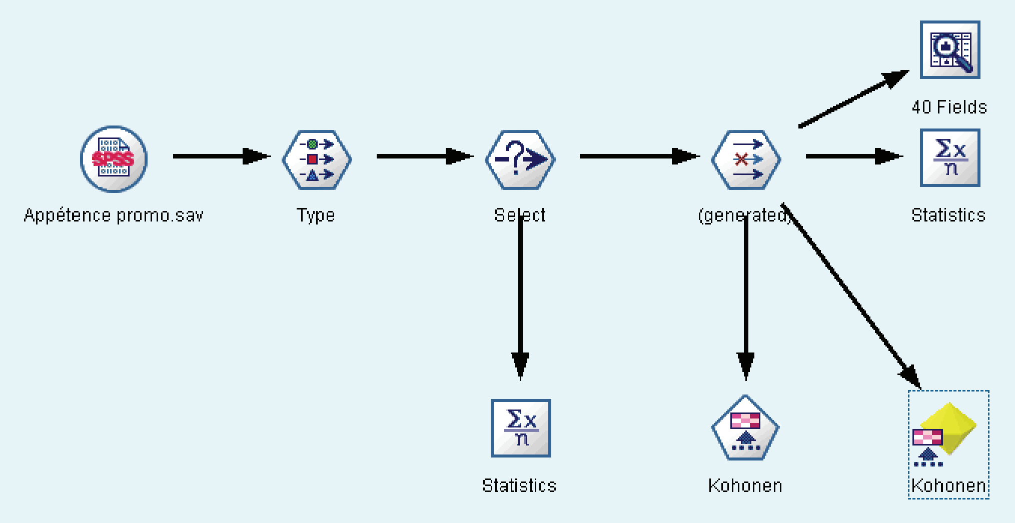
| Part de chiffres d'affaires selon les familles de produits |
|---|
| Ratio clients autre |
| Ratio clients bazar |
| Ratio clients BOF/APF |
| Ratio clients charcuterie LS |
| Ratio clients animaux |
| Ratio clients bébés |
| Ratio clients boucherie |
| Ratio clients boulangerie |
| Ratio clients charcuterie |
| Ratio clients diététique bio |
| Ratio clients fromage |

| |
|---|
| Ratio clients fruits et légumes |
| Ratio clients poisson |
| Ratio clients surgelés |
| Ratio clients Vin AOC |
| Ratio clients DPH |
| Ratio clients épicerie |
| Ratio clients liquide |
| Ratio clients textile |
| Ratio clients ultra frais |
| Ratio clients volaille |
| Ratio clients premier prix |
| Ratio clients marque distributeur 1 |
| Ratio clients marque distributeur 2 |

Les variables non retenues sont éliminées après l'analyse des corrélations pour les variables quantitatives (le chiffre d'affaires et le nombre d'actes d'achat par exemple) ou par des matrices de proximité pour les variables qualitatives. Nous n'utilisons pas d'ACP (analyse en composante principale), de réduction des variables quantitatives, parce que nous voulons garder les informations au niveau le plus désagrégé possible, des variables originales de la base de données.

Les clients inactifs (les clients sans aucune transaction sur la période) sont éliminés.

Le flux PASW Modeler :



*Cette illustration est ici pour montrer la façon dont un modèle est construit sur PASW Modeler de SPSS/IBM (ex. : Clementine). C'est un logiciel statistique qui utilise un langage objet pour construire les modèles, dans ce logiciel les objets sont appelés des « nœuds ». Le nœud de départ sur la gauche est noté SPSS car ce sont les données en entrée, les autres nœuds suivants les flèches sont des opérations sur les données. Le nœud noté « Kohonen » contient le modèle statistique, le diamant jaune en sortie en bas à droite contient les résultats du modèle.*

Nous utilisons une méthode de classification pour créer les groupes de « clones » qui les rend fortement homogènes intrinsèquement mais différents entre eux.

Une fois que les familles de clones ont été obtenues et que les valeurs du potentiel ont été calculées, les principales familles de potentiel sont déterminées :

- « Or » : taux d'évolution supérieur à 20%
- « Argent » : taux d'évolution supérieur ou égale à 15% et inférieur à 20%
- « Bronze » : taux d'évolution inférieur à 15%

Il doit être souligné ici que le potentiel porte sur le potentiel absolu sur douze mois consécutifs.

Ce potentiel est exprimé sous la forme d'un taux. Pour des objectifs opérationnels, les valeurs de potentiels doivent être reclassées en valeur absolue :

P1: Gros potentiel
P2: Moyen potentiel
P3: Petit potentiel

## *2. La fiabilité de la méthode de Kohonen*

Nous avons testé plusieurs méthodes pour déterminer les convergences entre les classes de Kohonen.

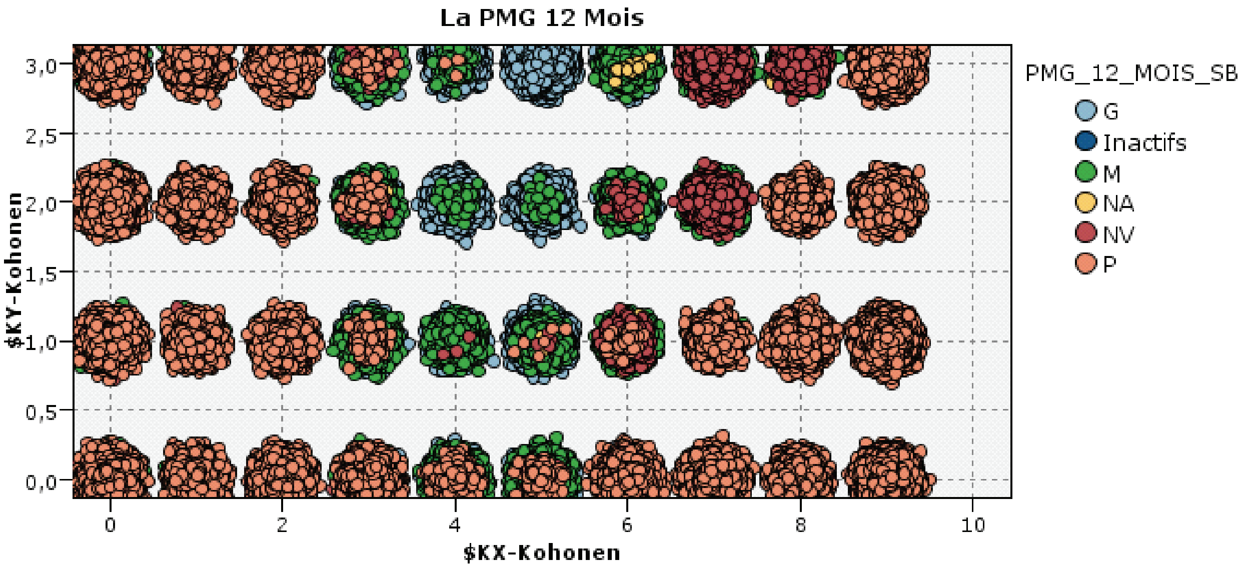
### *2.1. La visualisation des convergences*

Nous avons obtenu 40 classes, numérotées de 00 à 93 (les classes ne sont pas numérotées de manière continue).

Nous voulions obtenir un nombre assez important de classes pour minimiser au maximum l'écart type intra classes.

La cartographie : 00 est la classe de coordonnée 0 sur l'axe des X et 0 sur l'axe des Y et 93 est la classes de coordonnées 9 sur l'axe des X et 3 sur l'axe des Y.

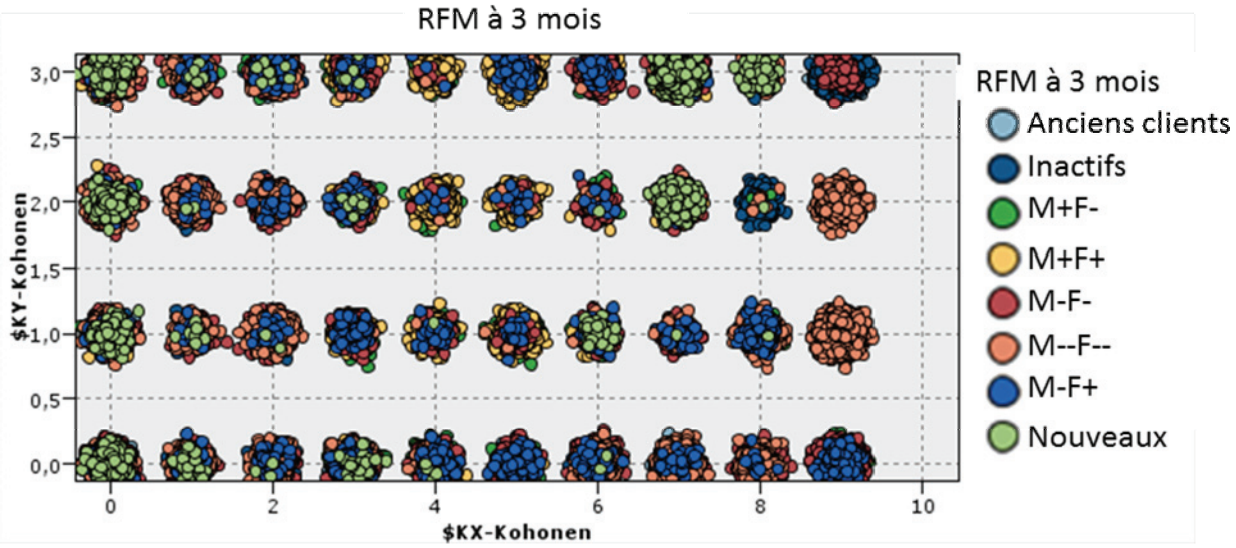
Les classes de Kohonen x la segmentation PMG (12 mois)



Les couleurs représentent les segments PMG, Nouveaux (NV), Inactifs et Non Affectés à un segment (NA)

Les couleurs qui représentent les segments PMG de clients sont en général bien regroupées, les clients appartenant aux mêmes segments PMG étant ensemble. Visuellement, la position des PMG dans les classes montre de la stabilité.

Les classes de Kohonen x la segmentation RFM (3 mois)



Les couleurs représentent les segments RFM, les Anciens clients sans achat depuis plus de 48 mois, les Inactifs sans achat depuis plus de 24 mois, les Nouveaux apparus dans les 6 derniers mois, les M-F-, les M-F-, les M+F-, les M-F+, les M+F+ qui sont les segments basés sur le montant et le fréquence du moins actif au plus actif.

Les couleurs sont généralement assez bien regroupées, les clients appartenant aux mêmes segments RFM (les segments vont de M+F+ à M--F--) se retrouvant majoritairement dans les mêmes classes de Kohonen.

Dans la RFM le mélange de couleurs est plus important à l'intérieur de chaque classe, mais les consommateurs appartenant aux mêmes segments RFM se retrouvent principalement dans les mêmes classes de Kohonen ; l'homogénéité des classes est néanmoins moins évidente que dans le mapping de la PMG.

### *2.2. La robustesse de la classification de Kohonen*

La distribution de la population est-elle stable ? Nous répondons à cette question de quatre façons différentes :
A. Y a-t-il une convergence entre le poids des classes entre l'échantillon des observations actives et les observations passives ?
B. Le regroupement peut-il être reproduit par un réseau Bayésien (*Pourret et al*, Jensen, Stephenson, 2000) ?
C. Peux-t-on reproduire la classification par une segmentation telle que C5.0[15] *(Quinlan 1993, 1996, 2004)* ?
D. Y a-t-il connectivité des super classes[16] ?

#### A. LA CONVERGENCE DE LA MÉTHODE

Nous pouvons vérifier l'allocation des clients sur deux échantillons aléatoires de notre corpus de données. Les classes de Kohonen se retrouvent aux mêmes poids entre les deux échantillons.

[15] C5.0 est un algorithme qui permet de faire des scores discriminant entre des groupes d'individus. Ici nous l'utilisons pour confirmer la classification de Kohonen. C5.0 devant parvenir avec les variables les plus discriminantes de chaque type à réaffecter les individus dans les mêmes classes que celles ou les a affecté l'algorithme de Kohonen.

[16] Connectivité des super classes : si on regroupe les classes de Kohonen (en groupe plus important : super classes). Les regroupements sont-ils pertinents, les classes rapprochées sont-elles bien les plus proches ?

| Clones | Total | | Echantillon Apprentissage | | Echantillon Test | |
|---|---|---|---|---|---|---|
| | Effectif | % | Effectif | % | Effectif | % |
| KH01 | 271 944 | 5,06% | 10 949 | 5,13% | 260 995 | 5,06% |
| KH02 | 171 396 | 3,19% | 6 983 | 3,27% | 164 413 | 3,19% |
| KH03 | 261 136 | 4,86% | 10 498 | 4,92% | 250 638 | 4,86% |
| KH04 | 289 912 | 5,40% | 11 508 | 5,39% | 278 404 | 5,40% |
| KH05 | 80 239 | 1,49% | 3 214 | 1,50% | 77 025 | 1,49% |
| KH06 | 40 698 | 0,76% | 1 596 | 0,75% | 39 102 | 0,76% |
| KH07 | 64 515 | 1,20% | 2 550 | 1,19% | 61 965 | 1,20% |
| KH08 | 93 685 | 1,74% | 3 768 | 1,76% | 89 917 | 1,74% |
| KH09 | 95 415 | 1,78% | 3 757 | 1,76% | 91 658 | 1,78% |
| KH10 | 91 169 | 1,70% | 3 681 | 1,72% | 87 488 | 1,70% |
| KH11 | 57 384 | 1,07% | 2 235 | 1,05% | 55 149 | 1,07% |
| KH12 | 181 691 | 3,38% | 7 224 | 3,38% | 174 467 | 3,38% |
| KH13 | 142 728 | 2,66% | 5 624 | 2,63% | 137 104 | 2,66% |
| KH14 | 83 298 | 1,55% | 3 260 | 1,53% | 80 038 | 1,55% |
| KH15 | 65 365 | 1,22% | 2 597 | 1,22% | 62 768 | 1,22% |
| KH16 | 152 665 | 2,84% | 6 153 | 2,88% | 146 512 | 2,84% |
| KH17 | 119 559 | 2,23% | 4 797 | 2,25% | 114 762 | 2,22% |
| KH18 | 45 360 | 0,84% | 1 794 | 0,84% | 43 566 | 0,84% |
| KH19 | 73 151 | 1,36% | 2 783 | 1,30% | 70 368 | 1,36% |
| KH20 | 35 914 | 0,67% | 1 378 | 0,65% | 34 536 | 0,67% |
| KH21 | 120 165 | 2,24% | 4 688 | 2,20% | 115 477 | 2,24% |
| KH22 | 137 752 | 2,56% | 5 462 | 2,56% | 132 290 | 2,56% |
| KH23 | 36 215 | 0,67% | 1 417 | 0,66% | 34 798 | 0,67% |
| KH24 | 267 939 | 4,99% | 10 739 | 5,03% | 257 200 | 4,99% |
| KH25 | 193 624 | 3,60% | 7 581 | 3,55% | 186 043 | 3,61% |
| KH26 | 50 454 | 0,94% | 2 019 | 0,95% | 48 435 | 0,94% |
| KH27 | 26 271 | 0,49% | 1 036 | 0,49% | 25 235 | 0,49% |
| KH28 | 76 724 | 1,43% | 3 082 | 1,44% | 73 642 | 1,43% |
| KH29 | 199 372 | 3,71% | 7 810 | 3,66% | 191 562 | 3,71% |
| KH30 | 28 913 | 0,54% | 1 102 | 0,52% | 27 811 | 0,54% |
| KH31 | 124 878 | 2,32% | 4 922 | 2,30% | 119 956 | 2,32% |
| KH32 | 347 565 | 6,47% | 13 963 | 6,54% | 333 602 | 6,47% |
| KH33 | 75 304 | 1,40% | 2 998 | 1,40% | 72 306 | 1,40% |
| KH34 | 103 656 | 1,93% | 4 107 | 1,92% | 99 549 | 1,93% |
| KH35 | 24 658 | 0,46% | 989 | 0,46% | 23 669 | 0,46% |
| KH36 | 31 206 | 0,58% | 1 272 | 0,60% | 29 934 | 0,58% |
| KH37 | 301 456 | 5,61% | 11 863 | 5,55% | 289 593 | 5,61% |
| KH38 | 252 820 | 4,71% | 10 042 | 4,70% | 242 778 | 4,71% |
| KH39 | 130 904 | 2,44% | 5 193 | 2,43% | 125 711 | 2,44% |
| KH40 | 425 926 | 7,93% | 16 942 | 7,93% | 408 984 | 7,93% |
| **Total** | **5 373 026** | **100,00%** | **213 576** | **100,00%** | **5 159 450** | **100,00%** |

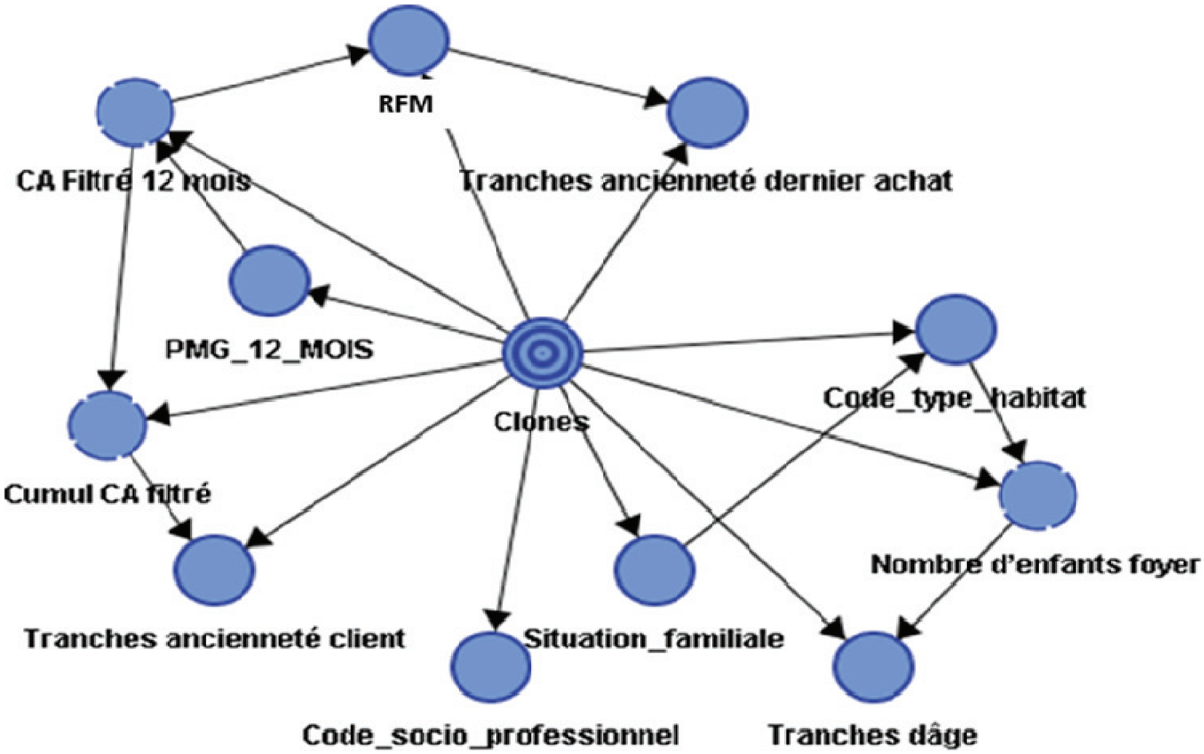
## B. La réallocation en utilisant un réseau Bayésien

Le tableau ci-dessus confirme que l'algorithme est capable de reproduire la distribution sur un ensemble de données plus important (Ensemble d'apprentissage vs Ensemble de test).

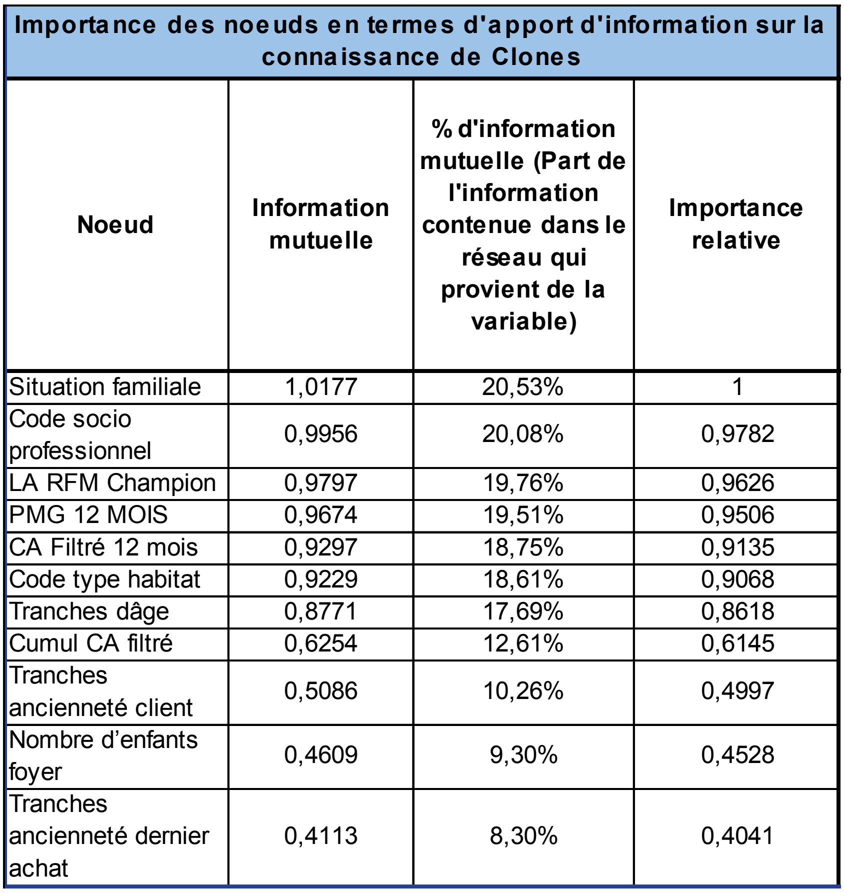
Quoi qu'il en soit, c'est en utilisant un autre algorithme que nous pouvons déterminer si oui ou non, la classification peut être reproduite et si elle est stable ou non.

Une fois encore, l'échantillon d'apprentissage est divisé en deux sous échantillons indépendants : l'échantillon d'apprentissage qui inclut 70 % des observations, l'échantillon test qui inclut les 30 % restant.

Nous utilisons un réseau Bayésien car, pour réallouer 40 groupes, une analyse discriminante n'est pas bien adaptée pour réallouer un si grand nombre de groupe avec autant de variables.
Le réseau Baysien permet une approche pas à pas ; nous pouvons également fixer le niveau de probabilité des liens que nous retenons entre les variables. Si nous fixons une probabilité de 0,9, les résultats sont présentés sous la forme du graphique ci-dessous.

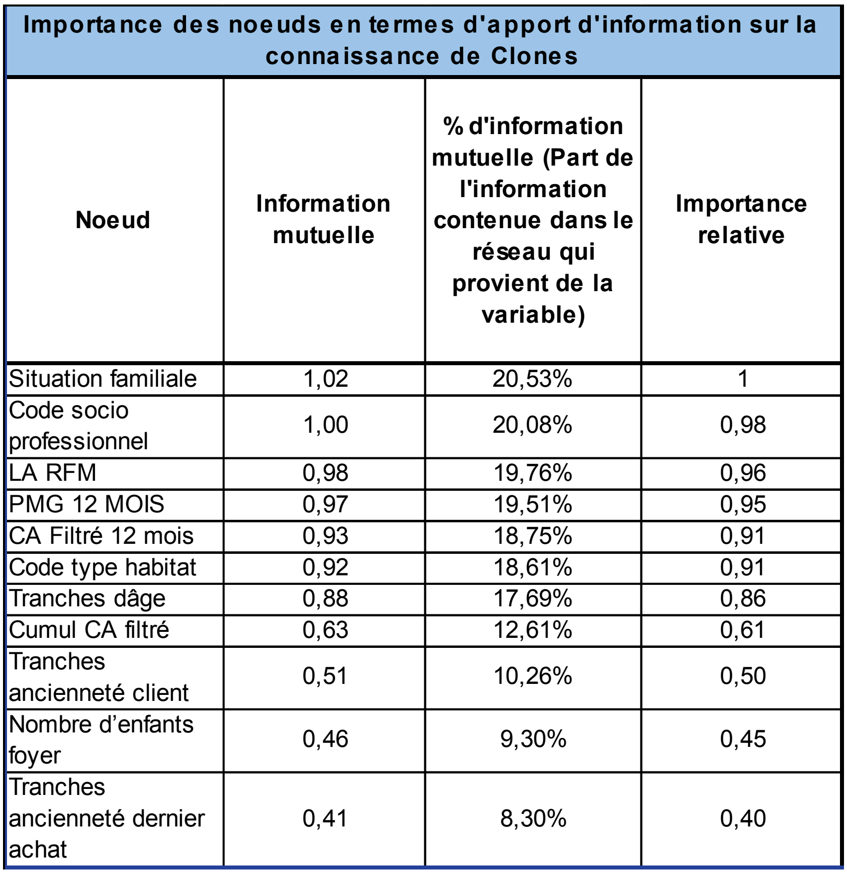


Le réseau utilise 11 variables : le chiffre d'affaires et aussi des variables socio démographiques. On peut remarquer que la PMG et la RFM sont retenues par le modèle. Ces résultats valident la représentativité des densités. Ci-dessous le poids des variables dans le modèle.



**Importance des noeuds en termes d'apport d'information sur la connaissance de Clones**

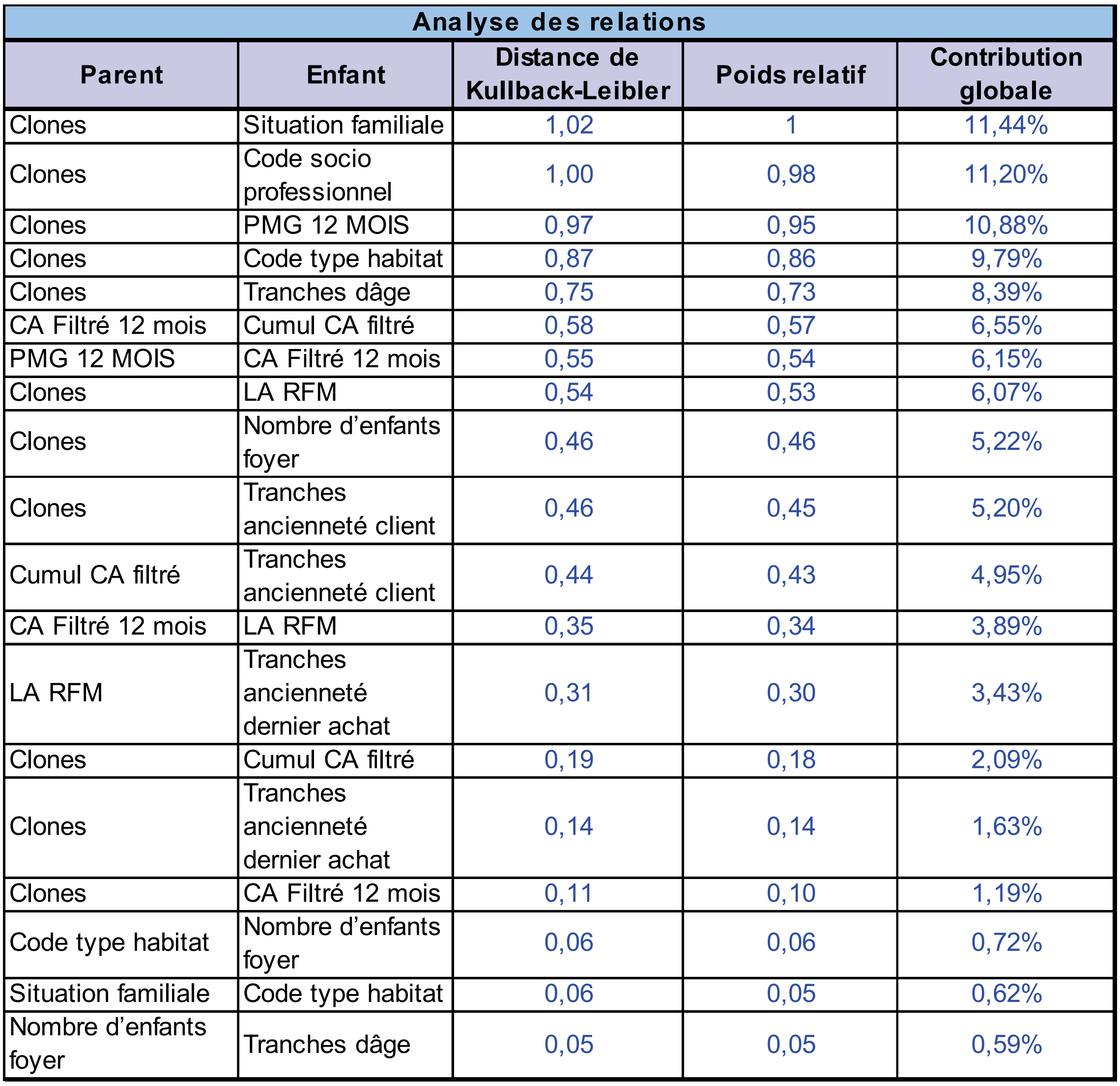
| Noeud | Information mutuelle | % d'information mutuelle (Part de l'information contenue dans le réseau qui provient de la variable) | Importance relative |
|---|---|---|---|
| Situation familiale | 1,0177 | 20,53% | 1 |
| Code socio professionnel | 0,9956 | 20,08% | 0,9782 |
| LA RFM Champion | 0,9797 | 19,76% | 0,9626 |
| PMG 12 MOIS | 0,9674 | 19,51% | 0,9506 |
| CA Filtré 12 mois | 0,9297 | 18,75% | 0,9135 |
| Code type habitat | 0,9229 | 18,61% | 0,9068 |
| Tranches dâge | 0,8771 | 17,69% | 0,8618 |
| Cumul CA filtré | 0,6254 | 12,61% | 0,6145 |
| Tranches ancienneté client | 0,5086 | 10,26% | 0,4997 |
| Nombre d’enfants foyer | 0,4609 | 9,30% | 0,4528 |
| Tranches ancienneté dernier achat | 0,4113 | 8,30% | 0,4041 |



**Importance des noeuds en termes d'apport d'information sur la connaissance de Clones**

| Noeud | Information mutuelle | % d'information mutuelle (Part de l'information contenue dans le réseau qui provient de la variable) | Importance relative |
|---|---|---|---|
| Situation familiale | 1,02 | 20,53% | 1 |
| Code socio professionnel | 1,00 | 20,08% | 0,98 |
| LA RFM | 0,98 | 19,76% | 0,96 |
| PMG 12 MOIS | 0,97 | 19,51% | 0,95 |
| CA Filtré 12 mois | 0,93 | 18,75% | 0,91 |
| Code type habitat | 0,92 | 18,61% | 0,91 |
| Tranches dâge | 0,88 | 17,69% | 0,86 |
| Cumul CA filtré | 0,63 | 12,61% | 0,61 |
| Tranches ancienneté client | 0,51 | 10,26% | 0,50 |
| Nombre d’enfants foyer | 0,46 | 9,30% | 0,45 |
| Tranches ancienneté dernier achat | 0,41 | 8,30% | 0,40 |

La mesure de Kullback-Leibler www.it-innovations.ae/iit005/proceedings/articles/E_6_IIT05_Khalid.pdf vient de la théorie de l'information. C'est une mesure de la convergence entre deux séries après quelles aient été recodées dans un format binaire (0 ou1). Plus la valeur est élevée, plus grande est la probabilité que ces valeurs aient une distribution jointe.



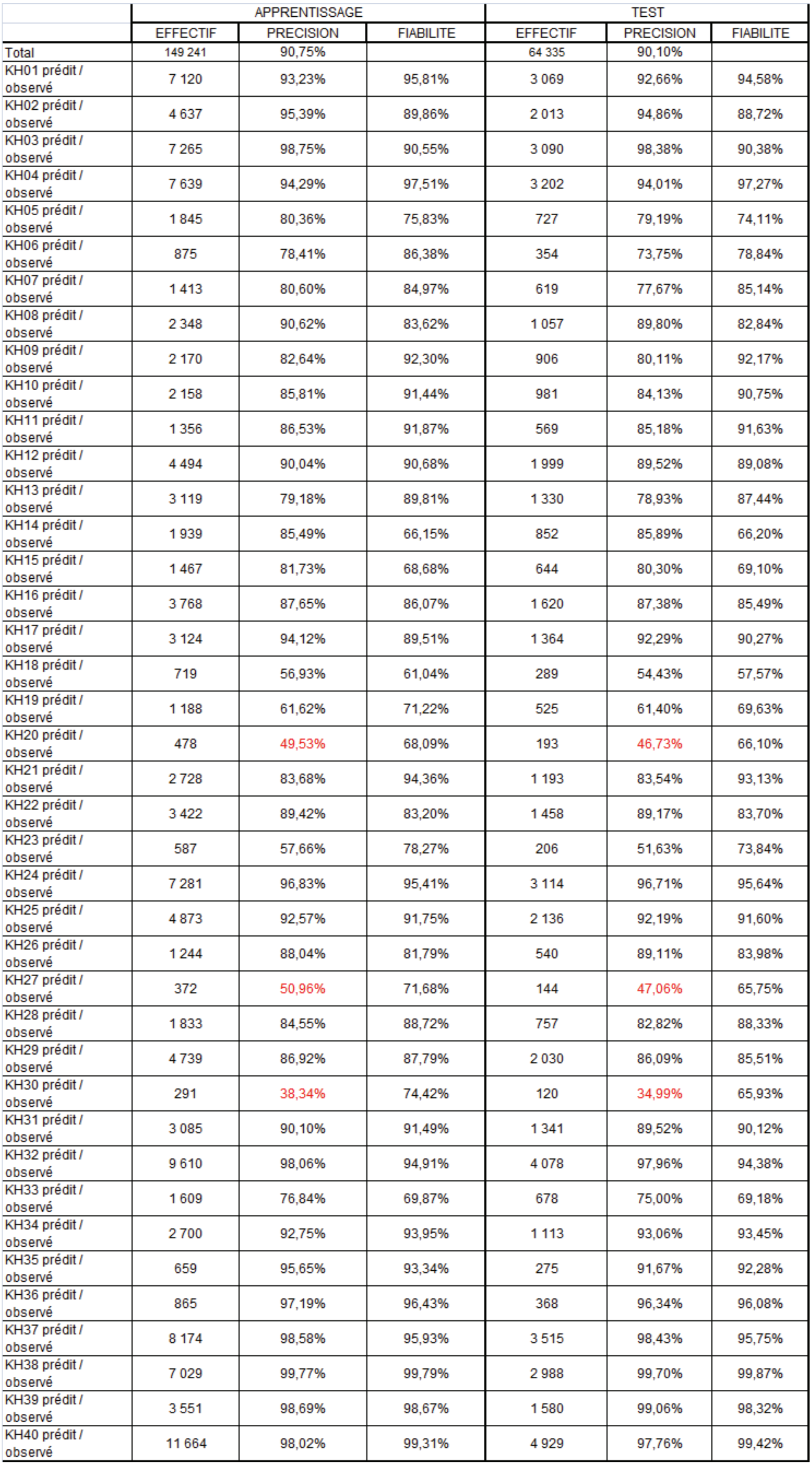
| Analyse des relations | | | | |
|---|---|---|---|---|
| Parent | Enfant | Distance de Kullback-Leibler | Poids relatif | Contribution globale |
| Clones | Situation familiale | 1,02 | 1 | 11,44% |
| Clones | Code socio professionnel | 1,00 | 0,98 | 11,20% |
| Clones | PMG 12 MOIS | 0,97 | 0,95 | 10,88% |
| Clones | Code type habitat | 0,87 | 0,86 | 9,79% |
| Clones | Tranches dâge | 0,75 | 0,73 | 8,39% |
| CA Filtré 12 mois | Cumul CA filtré | 0,58 | 0,57 | 6,55% |
| PMG 12 MOIS | CA Filtré 12 mois | 0,55 | 0,54 | 6,15% |
| Clones | LA RFM | 0,54 | 0,53 | 6,07% |
| Clones | Nombre d'enfants foyer | 0,46 | 0,46 | 5,22% |
| Clones | Tranches ancienneté client | 0,46 | 0,45 | 5,20% |
| Cumul CA filtré | Tranches ancienneté client | 0,44 | 0,43 | 4,95% |
| CA Filtré 12 mois | LA RFM | 0,35 | 0,34 | 3,89% |
| LA RFM | Tranches ancienneté dernier achat | 0,31 | 0,30 | 3,43% |
| Clones | Cumul CA filtré | 0,19 | 0,18 | 2,09% |
| Clones | Tranches ancienneté dernier achat | 0,14 | 0,14 | 1,63% |
| Clones | CA Filtré 12 mois | 0,11 | 0,10 | 1,19% |
| Code type habitat | Nombre d'enfants foyer | 0,06 | 0,06 | 0,72% |
| Situation familiale | Code type habitat | 0,06 | 0,05 | 0,62% |
| Nombre d'enfants foyer | Tranches dâge | 0,05 | 0,05 | 0,59% |

Le résultat du scoring au niveau individuel : sur l'ensemble d'apprentissage, 90,7 % des individus sont correctement réaffectés.

Sur l'échantillon test, le pourcentage est de 90,1 %.

Ci-dessous on trouve les taux en % d'individus correctement réaffectés par les réseaux bayésiens pour chacune des 40 classes.

Les classes de Kohonen peuvent être reproduites.

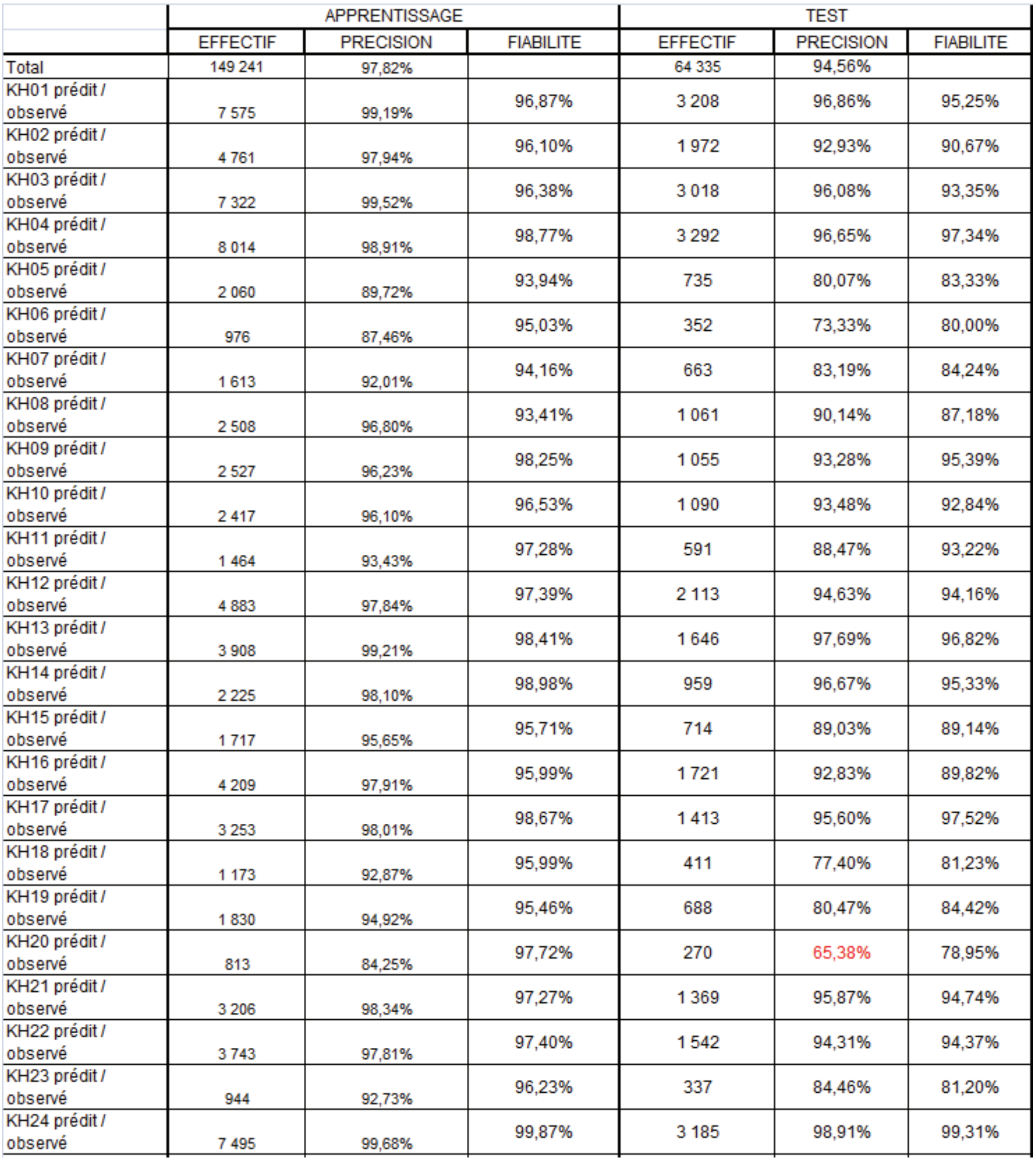


| | APPRENTISSAGE | | | TEST | | |
|---|---|---|---|---|---|---|
| | EFFECTIF | PRECISION | FIABILITE | EFFECTIF | PRECISION | FIABILITE |
| Total | 149 241 | 90,75% | | 64 335 | 90,10% | |
| KH01 prédit / observé | 7 120 | 93,23% | 95,81% | 3 069 | 92,66% | 94,58% |
| KH02 prédit / observé | 4 637 | 95,39% | 89,86% | 2 013 | 94,86% | 88,72% |
| KH03 prédit / observé | 7 265 | 98,75% | 90,55% | 3 090 | 98,38% | 90,38% |
| KH04 prédit / observé | 7 639 | 94,29% | 97,51% | 3 202 | 94,01% | 97,27% |
| KH05 prédit / observé | 1 845 | 80,36% | 75,83% | 727 | 79,19% | 74,11% |
| KH06 prédit / observé | 875 | 78,41% | 86,38% | 354 | 73,75% | 78,84% |
| KH07 prédit / observé | 1 413 | 80,60% | 84,97% | 619 | 77,67% | 85,14% |
| KH08 prédit / observé | 2 348 | 90,62% | 83,62% | 1 057 | 89,80% | 82,84% |
| KH09 prédit / observé | 2 170 | 82,64% | 92,30% | 906 | 80,11% | 92,17% |
| KH10 prédit / observé | 2 158 | 85,81% | 91,44% | 981 | 84,13% | 90,75% |
| KH11 prédit / observé | 1 356 | 86,53% | 91,87% | 569 | 85,18% | 91,63% |
| KH12 prédit / observé | 4 494 | 90,04% | 90,68% | 1 999 | 89,52% | 89,08% |
| KH13 prédit / observé | 3 119 | 79,18% | 89,81% | 1 330 | 78,93% | 87,44% |
| KH14 prédit / observé | 1 939 | 85,49% | 66,15% | 852 | 85,89% | 66,20% |
| KH15 prédit / observé | 1 467 | 81,73% | 68,68% | 644 | 80,30% | 69,10% |
| KH16 prédit / observé | 3 768 | 87,65% | 86,07% | 1 620 | 87,38% | 85,49% |
| KH17 prédit / observé | 3 124 | 94,12% | 89,51% | 1 364 | 92,29% | 90,27% |
| KH18 prédit / observé | 719 | 56,93% | 61,04% | 289 | 54,43% | 57,57% |
| KH19 prédit / observé | 1 188 | 61,62% | 71,22% | 525 | 61,40% | 69,63% |
| KH20 prédit / observé | 478 | 49,53% | 68,09% | 193 | 46,73% | 66,10% |
| KH21 prédit / observé | 2 728 | 83,68% | 94,36% | 1 193 | 83,54% | 93,13% |
| KH22 prédit / observé | 3 422 | 89,42% | 83,20% | 1 458 | 89,17% | 83,70% |
| KH23 prédit / observé | 587 | 57,66% | 78,27% | 206 | 51,63% | 73,84% |
| KH24 prédit / observé | 7 281 | 96,83% | 95,41% | 3 114 | 96,71% | 95,64% |
| KH25 prédit / observé | 4 873 | 92,57% | 91,75% | 2 136 | 92,19% | 91,60% |
| KH26 prédit / observé | 1 244 | 88,04% | 81,79% | 540 | 89,11% | 83,98% |
| KH27 prédit / observé | 372 | 50,96% | 71,68% | 144 | 47,06% | 65,75% |
| KH28 prédit / observé | 1 833 | 84,55% | 88,72% | 757 | 82,82% | 88,33% |
| KH29 prédit / observé | 4 739 | 86,92% | 87,79% | 2 030 | 86,09% | 85,51% |
| KH30 prédit / observé | 291 | 38,34% | 74,42% | 120 | 34,99% | 65,93% |
| KH31 prédit / observé | 3 085 | 90,10% | 91,49% | 1 341 | 89,52% | 90,12% |
| KH32 prédit / observé | 9 610 | 98,06% | 94,91% | 4 078 | 97,96% | 94,38% |
| KH33 prédit / observé | 1 609 | 76,84% | 69,87% | 678 | 75,00% | 69,18% |
| KH34 prédit / observé | 2 700 | 92,75% | 93,95% | 1 113 | 93,06% | 93,45% |
| KH35 prédit / observé | 659 | 95,65% | 93,34% | 275 | 91,67% | 92,28% |
| KH36 prédit / observé | 865 | 97,19% | 96,43% | 368 | 96,34% | 96,08% |
| KH37 prédit / observé | 8 174 | 98,58% | 95,93% | 3 515 | 98,43% | 95,75% |
| KH38 prédit / observé | 7 029 | 99,77% | 99,79% | 2 988 | 99,70% | 99,87% |
| KH39 prédit / observé | 3 551 | 98,69% | 98,67% | 1 580 | 99,06% | 98,32% |
| KH40 prédit / observé | 11 664 | 98,02% | 99,31% | 4 929 | 97,76% | 99,42% |

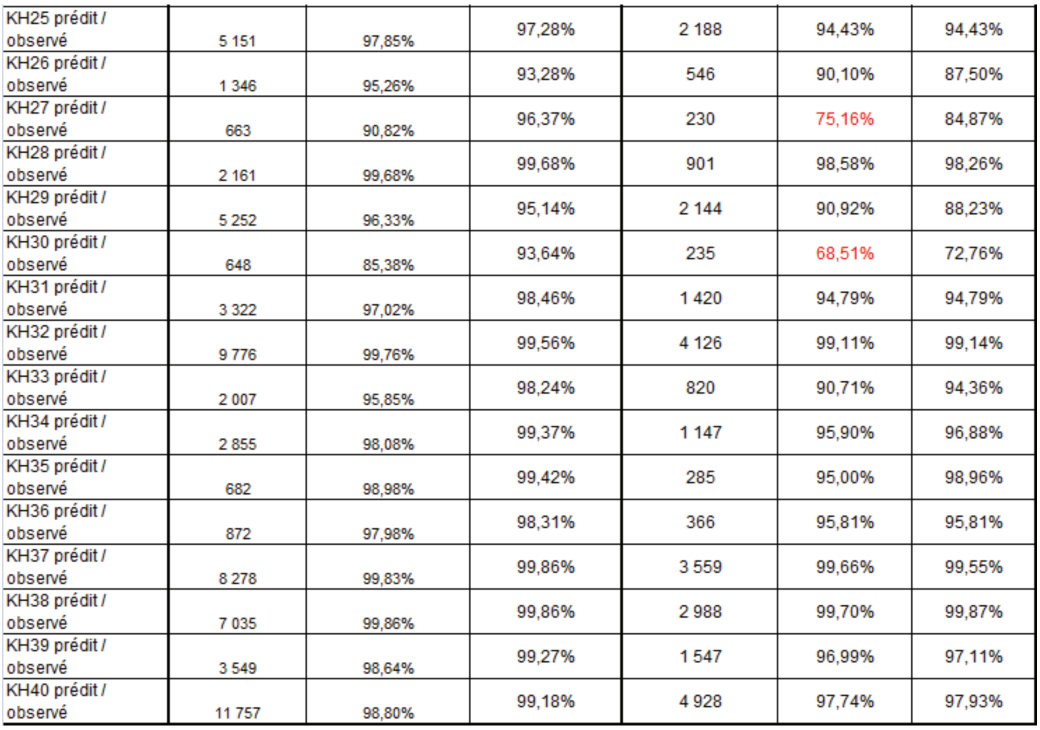
Dans le tableau ci-dessus, les groupes qui sont pauvrement réaffectés, contiennent un petit nombre de consommateurs.
Même pour ces groupes, la fiabilité reste au-dessus des 65 %.

### C. RÉALLOCATION PAR UN ARBRE DE DÉCISION

Le taux de validation croisé est de 94,2 % d'individus correctement affectés aux groupes.
L'échantillon test confirme ce taux.
Il y a une forte convergence entre les deux méthodes d'apprentissage supervisées, les réseaux bayésiens et le C5.0 sont capable de réaffecter correctement les individus dans les 40 classes.
La robustesse de la classification est également validée par cette méthode.



| | APPRENTISSAGE | | | TEST | | |
|---|---|---|---|---|---|---|
| | EFFECTIF | PRECISION | FIABILITE | EFFECTIF | PRECISION | FIABILITE |
| Total | 149 241 | 97,82% | | 64 335 | 94,56% | |
| KH01 prédit / observé | 7 575 | 99,19% | 96,87% | 3 208 | 96,86% | 95,25% |
| KH02 prédit / observé | 4 761 | 97,94% | 96,10% | 1 972 | 92,93% | 90,67% |
| KH03 prédit / observé | 7 322 | 99,52% | 96,38% | 3 018 | 96,08% | 93,35% |
| KH04 prédit / observé | 8 014 | 98,91% | 98,77% | 3 292 | 96,65% | 97,34% |
| KH05 prédit / observé | 2 060 | 89,72% | 93,94% | 735 | 80,07% | 83,33% |
| KH06 prédit / observé | 976 | 87,46% | 95,03% | 352 | 73,33% | 80,00% |
| KH07 prédit / observé | 1 613 | 92,01% | 94,16% | 663 | 83,19% | 84,24% |
| KH08 prédit / observé | 2 508 | 96,80% | 93,41% | 1 061 | 90,14% | 87,18% |
| KH09 prédit / observé | 2 527 | 96,23% | 98,25% | 1 055 | 93,28% | 95,39% |
| KH10 prédit / observé | 2 417 | 96,10% | 96,53% | 1 090 | 93,48% | 92,84% |
| KH11 prédit / observé | 1 464 | 93,43% | 97,28% | 591 | 88,47% | 93,22% |
| KH12 prédit / observé | 4 883 | 97,84% | 97,39% | 2 113 | 94,63% | 94,16% |
| KH13 prédit / observé | 3 908 | 99,21% | 98,41% | 1 646 | 97,69% | 96,82% |
| KH14 prédit / observé | 2 225 | 98,10% | 98,98% | 959 | 96,67% | 95,33% |
| KH15 prédit / observé | 1 717 | 95,65% | 95,71% | 714 | 89,03% | 89,14% |
| KH16 prédit / observé | 4 209 | 97,91% | 95,99% | 1 721 | 92,83% | 89,82% |
| KH17 prédit / observé | 3 253 | 98,01% | 98,67% | 1 413 | 95,60% | 97,52% |
| KH18 prédit / observé | 1 173 | 92,87% | 95,99% | 411 | 77,40% | 81,23% |
| KH19 prédit / observé | 1 830 | 94,92% | 95,46% | 688 | 80,47% | 84,42% |
| KH20 prédit / observé | 813 | 84,25% | 97,72% | 270 | 65,38% | 78,95% |
| KH21 prédit / observé | 3 206 | 98,34% | 97,27% | 1 369 | 95,87% | 94,74% |
| KH22 prédit / observé | 3 743 | 97,81% | 97,40% | 1 542 | 94,31% | 94,37% |
| KH23 prédit / observé | 944 | 92,73% | 96,23% | 337 | 84,46% | 81,20% |
| KH24 prédit / observé | 7 495 | 99,68% | 99,87% | 3 185 | 98,91% | 99,31% |



| | | | | | | |
|---|---|---|---|---|---|---|
| KH25 prédit / observé | 5 151 | 97,85% | 97,28% | 2 188 | 94,43% | 94,43% |
| KH26 prédit / observé | 1 346 | 95,26% | 93,28% | 546 | 90,10% | 87,50% |
| KH27 prédit / observé | 663 | 90,82% | 96,37% | 230 | 75,16% | 84,87% |
| KH28 prédit / observé | 2 161 | 99,68% | 99,68% | 901 | 98,58% | 98,26% |
| KH29 prédit / observé | 5 252 | 96,33% | 95,14% | 2 144 | 90,92% | 88,23% |
| KH30 prédit / observé | 648 | 85,38% | 93,64% | 235 | 68,51% | 72,76% |
| KH31 prédit / observé | 3 322 | 97,02% | 98,46% | 1 420 | 94,79% | 94,79% |
| KH32 prédit / observé | 9 776 | 99,76% | 99,56% | 4 126 | 99,11% | 99,14% |
| KH33 prédit / observé | 2 007 | 95,85% | 98,24% | 820 | 90,71% | 94,36% |
| KH34 prédit / observé | 2 855 | 98,08% | 99,37% | 1 147 | 95,90% | 96,88% |
| KH35 prédit / observé | 682 | 98,98% | 99,42% | 285 | 95,00% | 98,96% |
| KH36 prédit / observé | 872 | 97,98% | 98,31% | 366 | 95,81% | 95,81% |
| KH37 prédit / observé | 8 278 | 99,83% | 99,86% | 3 559 | 99,66% | 99,55% |
| KH38 prédit / observé | 7 035 | 99,86% | 99,86% | 2 988 | 99,70% | 99,87% |
| KH39 prédit / observé | 3 549 | 98,64% | 99,27% | 1 547 | 96,99% | 97,11% |
| KH40 prédit / observé | 11 757 | 98,80% | 99,18% | 4 928 | 97,74% | 97,93% |

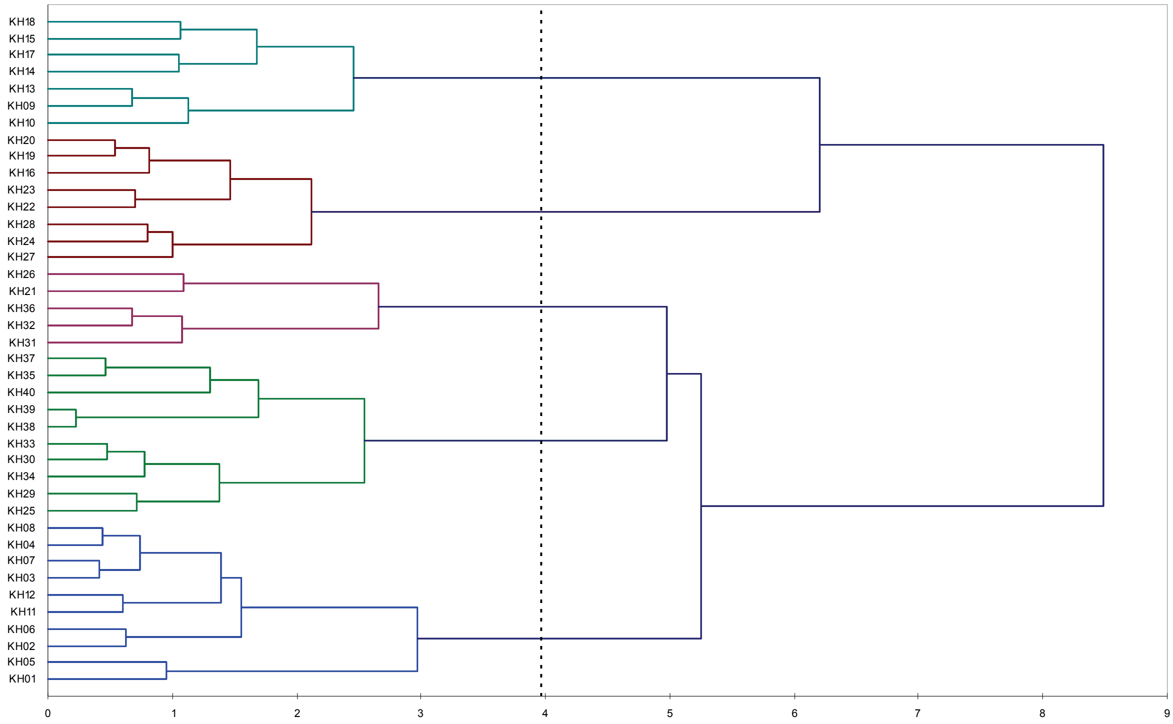
## D. LA CONNECTIVITÉ DES SUPER CLASSES

Nous utilisons une analyse par réseau, qui identifie un petit nombre de variables qui sont les plus importantes pour la classification.
Nous analysons un tableau de contingence entre les 40 groupes et les variables qui contribuent au réseau à plus de 10% des capacités explicatives.

– La situation familiale
– C.S.P.
– R.F.M. à 3 mois
– P.M.G à 3 mois
– Le type d'habitat
– Les classes d'âges
– Le chiffre d'affaire filtré en cumul
– Les classes de séniorité des clients

Sur cette table, l'échelle utilisée et la distance du Khi² *(Ottos, 2007, Meunier et al, Romesburg, 2004)* et une méthode d'agrégation de Ward est appliquée (Clarke and Sun, 1997, Barnier 2008).

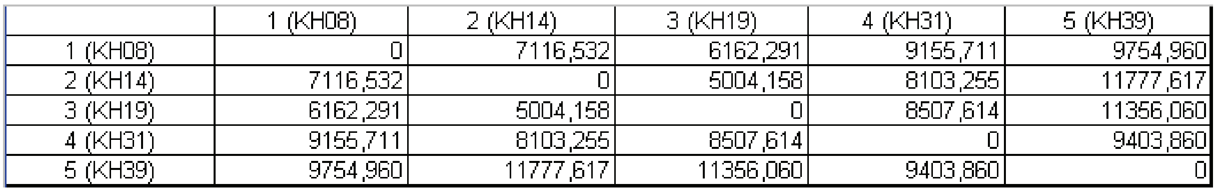
Dendrogramme



Les césures des écarts types pour une classification optimale :

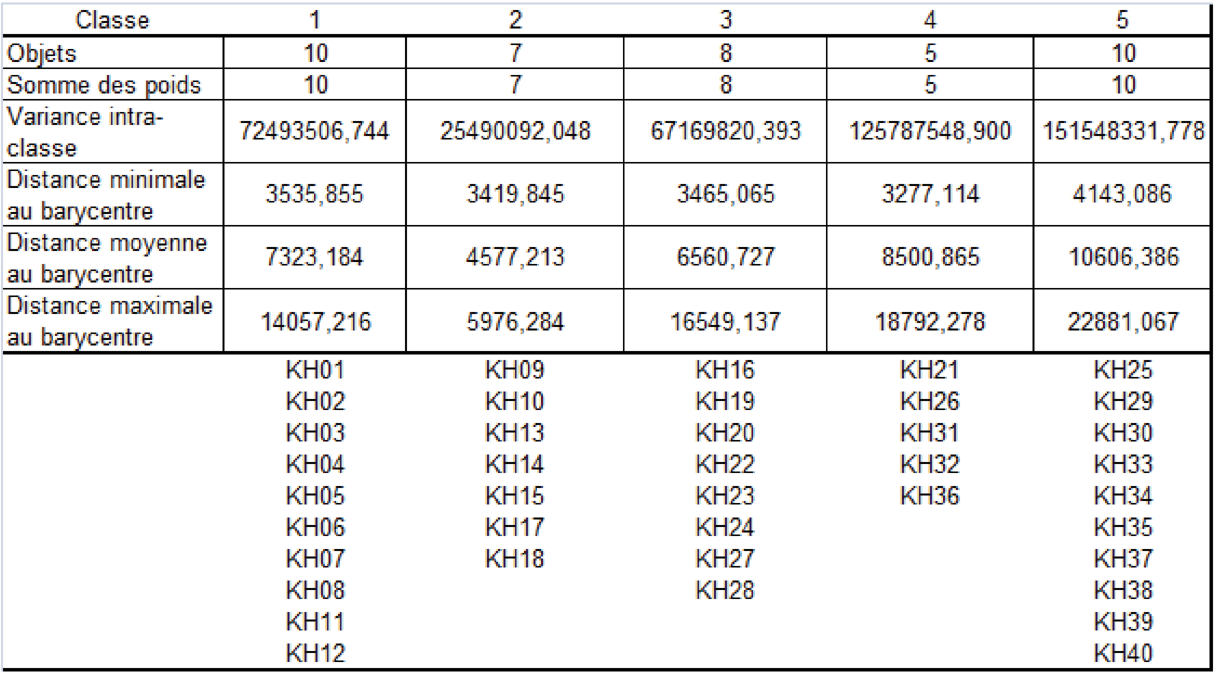
| | |
|---|---|
| Intra-groupes | 89790172,495 |
| Inter-groupes | 25701803,959 |
| Total | 115491976,454 |

Distances entre les centres des classes :



| | 1 (KH08) | 2 (KH14) | 3 (KH19) | 4 (KH31) | 5 (KH39) |
|---|---|---|---|---|---|
| 1 (KH08) | 0 | 7116,532 | 6162,291 | 9155,711 | 9754,960 |
| 2 (KH14) | 7116,532 | 0 | 5004,158 | 8103,255 | 11777,617 |
| 3 (KH19) | 6162,291 | 5004,158 | 0 | 8507,614 | 11356,060 |
| 4 (KH31) | 9155,711 | 8103,255 | 8507,614 | 0 | 9403,860 |
| 5 (KH39) | 9754,960 | 11777,617 | 11356,060 | 9403,860 | 0 |

Résultats par groupe :



| Classe | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Objets | 10 | 7 | 8 | 5 | 10 |
| Somme des poids | 10 | 7 | 8 | 5 | 10 |
| Variance intra-classe | 72493506,744 | 25490092,048 | 67169820,393 | 125787548,900 | 151548331,778 |
| Distance minimale au barycentre | 3535,855 | 3419,845 | 3465,065 | 3277,114 | 4143,086 |
| Distance moyenne au barycentre | 7323,184 | 4577,213 | 6560,727 | 8500,865 | 10606,386 |
| Distance maximale au barycentre | 14057,216 | 5976,284 | 16549,137 | 18792,278 | 22881,067 |
|  | KH01<br>KH02<br>KH03<br>KH04<br>KH05<br>KH06<br>KH07<br>KH08<br>KH11<br>KH12 | KH09<br>KH10<br>KH13<br>KH14<br>KH15<br>KH17<br>KH18 | KH16<br>KH19<br>KH20<br>KH22<br>KH23<br>KH24<br>KH27<br>KH28 | KH21<br>KH26<br>KH31<br>KH32<br>KH36 | KH25<br>KH29<br>KH30<br>KH33<br>KH34<br>KH35<br>KH37<br>KH38<br>KH39<br>KH40 |

Une vérification est réalisée pour s'assurer que le respect des limites inférieures et supérieures de l'ordre des groupes : le clone 40 n'est pas regroupé avec le clone 3. C'est un des critères de « qualité » d'une carte de Kohonen.

En conclusion, la forme de la classification obtenue par l'algorithme de Kohonen, satisfait les critères de stabilité et de reproductibilité qui garantissent un potentiel robuste et durable.

### *3. Calcul des potentiels*

Nous divisons le chiffre d'affaires annuel (le chiffre d'affaires sur 12 mois) en déciles. Pour chacune des classes obtenues par la méthode de Kohonen, nous avons calculé un potentiel basé sur le chiffre d'affaires.

Nous divisons le chiffre d'affaires total de chaque classe de clones en déciles, nous calculons la médiane de chaque décile.

Nous allouons par groupe une valeur de chiffre d'affaires potentiel déduite du calcul des taux de potentiel d'accroissement entre les médianes et les déciles.

Pour chaque groupe de clones, le taux de croissance de la mesure d'accroissement du chiffre d'affaires, va de la limite inférieure du décile, à la limite supérieure.

20 taux d'accroissement par groupe de clones sont déterminés :

– Entre la limite inférieur du premier décile et la médiane du premier décile : T x 01

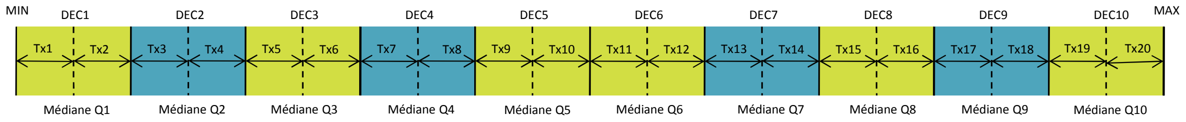
– Entre la médiane du premier décile et limite supérieure du premier décile du second décile : T x 02

– Entre la limite inférieur du premier décile et la médiane du second décile : T x 03...

– Entre la médiane du huitième décile et limite supérieure du huitième décile : T x 18

– Entre la limite inférieure du neuvième décile et la limite médiane du neuvième décile : T x 19

– Entre la médiane du neuvième décile et limite supérieure du dixième décile : T x 20

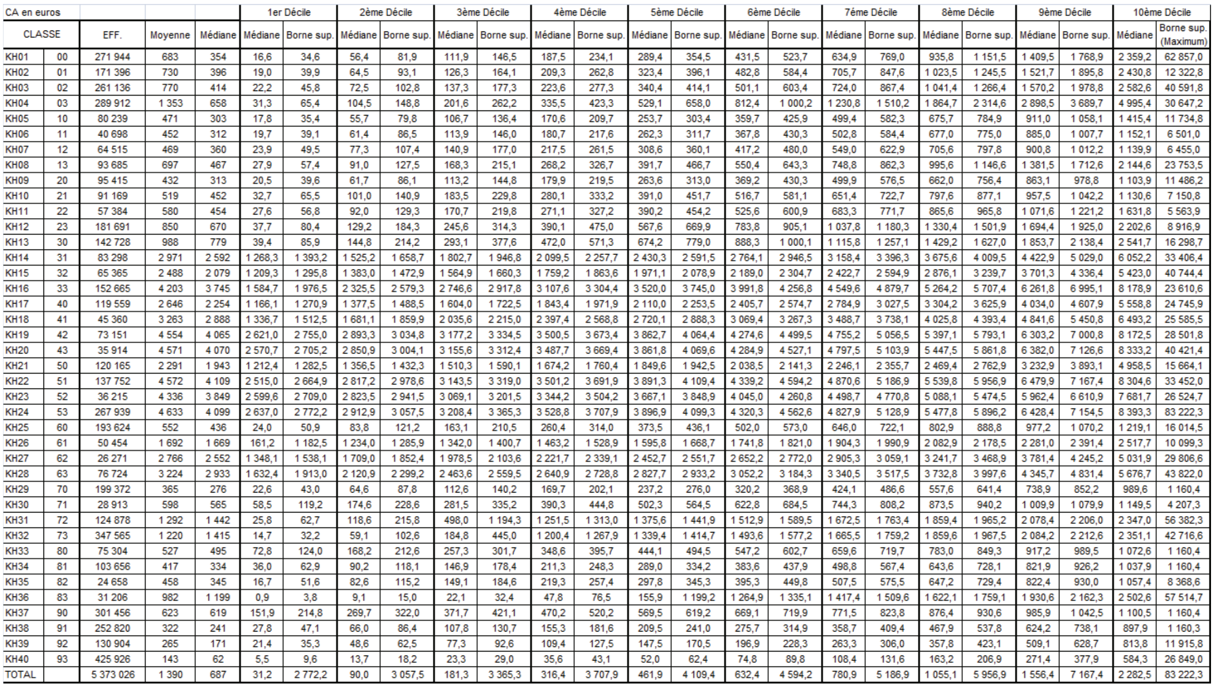


Pour chaque classe de Kohonen, un client pour lequel le chiffre d'affaire filtré, se situe entre le minimum et la médiane du premier décile, se verra attribué un taux d'évolution de 1 (T x 01).

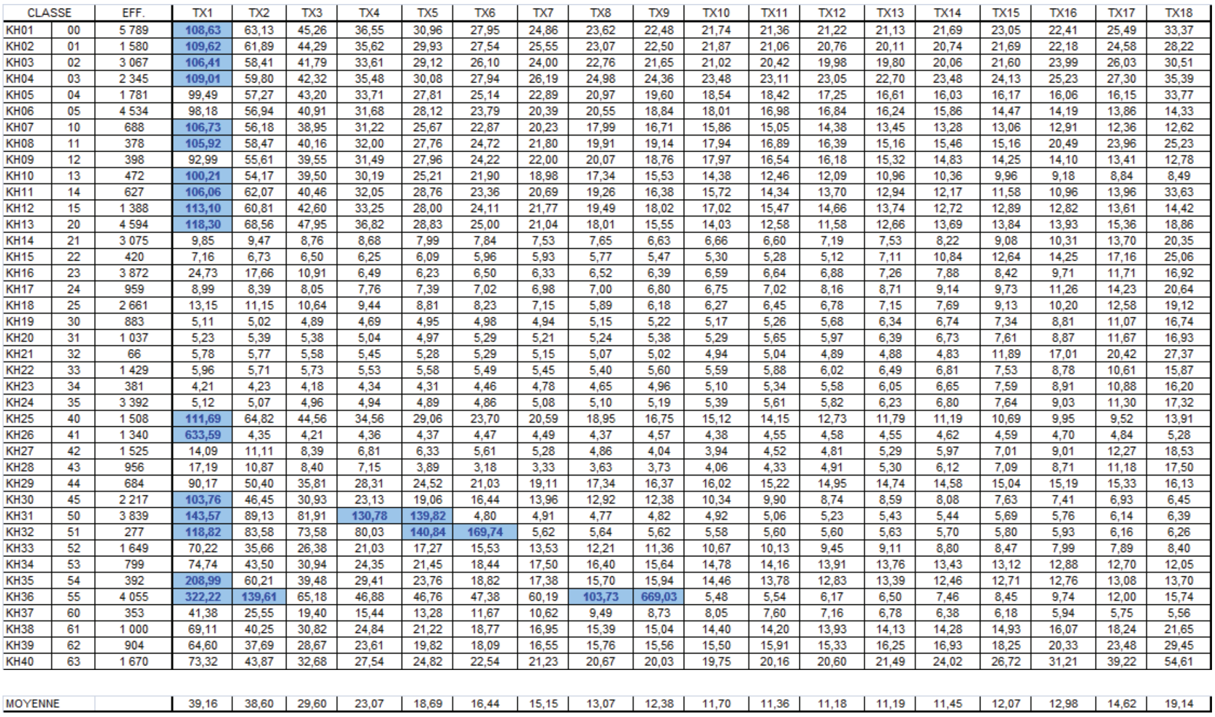
Un client pour lequel le chiffre d'affaires est compris entre la médiane et la limite supérieure de premier décile, se verra attribuer un taux d'accroissement égal au taux de 2 (T x 02) etc.

Chaque client se voit attribuer un taux d'évolution. Le taux multiplié par chiffre d'affaires nous permet d'estimer un potentiel de chiffre d'affaires à chaque client.

Les limites et les médianes des déciles par groupe de Kohonen :



| CA en euros | | | | | 1er Décile | | 2ème Décile | | 3ème Décile | | 4ème Décile | | 5ème Décile | | 6ème Décile | | 7éme Décile | | 8ème Décile | | 9ème Décile | | 10ème Décile | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLASSE | | EFF. | Moyenne | Médiane | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. | Médiane | Borne sup. (Maximum) |
| KH01 | 00 | 271 944 | 683 | 354 | 16,6 | 34,6 | 56,4 | 81,9 | 111,9 | 146,5 | 187,5 | 234,1 | 289,4 | 354,5 | 431,5 | 523,7 | 634,9 | 769,0 | 935,8 | 1 151,5 | 1 409,5 | 1 768,9 | 2 359,2 | 62 857,0 |
| KH02 | 01 | 171 396 | 730 | 396 | 19,0 | 39,9 | 64,5 | 93,1 | 126,3 | 164,1 | 209,3 | 262,8 | 323,4 | 396,1 | 482,8 | 584,4 | 705,7 | 847,6 | 1 023,5 | 1 245,5 | 1 521,7 | 1 895,8 | 2 430,8 | 12 322,8 |
| KH03 | 02 | 261 136 | 770 | 414 | 22,2 | 45,8 | 72,5 | 102,8 | 137,3 | 177,3 | 223,6 | 277,3 | 340,4 | 414,1 | 501,1 | 603,4 | 724,0 | 867,4 | 1 041,4 | 1 266,4 | 1 570,2 | 1 978,8 | 2 582,6 | 40 591,8 |
| KH04 | 03 | 289 912 | 1 353 | 658 | 31,3 | 65,4 | 104,5 | 148,8 | 201,6 | 262,2 | 335,5 | 423,3 | 529,1 | 658,0 | 812,4 | 1 000,2 | 1 230,8 | 1 510,2 | 1 864,7 | 2 314,6 | 2 898,5 | 3 689,7 | 4 995,4 | 30 647,2 |
| KH05 | 10 | 80 239 | 471 | 303 | 17,8 | 35,4 | 55,7 | 79,8 | 106,7 | 136,4 | 170,6 | 209,7 | 253,7 | 303,4 | 359,7 | 425,9 | 499,4 | 582,3 | 675,7 | 784,9 | 911,0 | 1 058,1 | 1 415,4 | 11 734,8 |
| KH06 | 11 | 40 698 | 452 | 312 | 19,7 | 39,1 | 61,4 | 86,5 | 113,9 | 146,0 | 180,7 | 217,6 | 262,3 | 311,7 | 367,8 | 430,3 | 502,8 | 584,4 | 677,0 | 775,0 | 885,0 | 1 007,7 | 1 152,1 | 6 501,0 |
| KH07 | 12 | 64 515 | 469 | 360 | 23,9 | 49,5 | 77,3 | 107,4 | 140,9 | 177,0 | 217,5 | 261,5 | 308,6 | 360,1 | 417,2 | 480,0 | 549,0 | 622,9 | 705,6 | 797,8 | 900,8 | 1 012,2 | 1 139,9 | 6 455,0 |
| KH08 | 13 | 93 685 | 697 | 467 | 27,9 | 57,4 | 91,0 | 127,5 | 168,3 | 215,1 | 268,2 | 326,7 | 391,7 | 466,7 | 550,4 | 643,3 | 748,8 | 862,3 | 995,6 | 1 146,6 | 1 381,5 | 1 712,6 | 2 144,6 | 23 753,5 |
| KH09 | 20 | 95 415 | 432 | 313 | 20,5 | 39,6 | 61,7 | 86,1 | 113,2 | 144,8 | 179,9 | 219,5 | 263,6 | 313,0 | 369,2 | 430,3 | 499,9 | 576,5 | 662,0 | 756,4 | 863,1 | 978,8 | 1 103,9 | 11 486,2 |
| KH10 | 21 | 91 169 | 519 | 452 | 32,7 | 65,5 | 101,0 | 140,9 | 183,5 | 229,8 | 280,1 | 333,2 | 391,0 | 451,7 | 516,7 | 581,1 | 651,4 | 722,7 | 797,6 | 877,1 | 957,5 | 1 042,2 | 1 130,6 | 7 150,8 |
| KH11 | 22 | 57 384 | 580 | 454 | 27,6 | 56,8 | 92,0 | 129,3 | 170,7 | 219,8 | 271,1 | 327,2 | 390,2 | 454,2 | 525,6 | 600,9 | 683,3 | 771,7 | 865,6 | 965,8 | 1 071,6 | 1 221,2 | 1 631,8 | 5 563,9 |
| KH12 | 23 | 181 691 | 850 | 670 | 37,7 | 80,4 | 129,2 | 184,3 | 245,6 | 314,3 | 390,1 | 475,0 | 567,6 | 669,9 | 783,8 | 905,1 | 1 037,8 | 1 180,3 | 1 330,4 | 1 501,9 | 1 694,4 | 1 925,0 | 2 202,6 | 8 916,9 |
| KH13 | 30 | 142 728 | 988 | 779 | 39,4 | 85,9 | 144,8 | 214,2 | 293,1 | 377,6 | 472,0 | 571,3 | 674,2 | 779,0 | 888,3 | 1 000,1 | 1 115,8 | 1 257,1 | 1 429,2 | 1 627,0 | 1 853,7 | 2 138,4 | 2 541,7 | 16 298,7 |
| KH14 | 31 | 83 298 | 2 971 | 2 592 | 1 268,3 | 1 393,2 | 1 525,2 | 1 658,7 | 1 802,7 | 1 946,8 | 2 099,5 | 2 257,7 | 2 430,3 | 2 591,5 | 2 764,1 | 2 946,5 | 3 158,4 | 3 396,3 | 3 675,6 | 4 009,5 | 4 422,9 | 5 029,0 | 6 052,2 | 33 406,4 |
| KH15 | 32 | 65 365 | 2 488 | 2 079 | 1 209,3 | 1 295,8 | 1 383,0 | 1 472,9 | 1 564,9 | 1 660,3 | 1 759,2 | 1 863,6 | 1 971,1 | 2 078,9 | 2 189,0 | 2 304,7 | 2 422,7 | 2 594,9 | 2 876,1 | 3 239,7 | 3 701,3 | 4 336,4 | 5 423,0 | 40 744,4 |
| KH16 | 33 | 152 665 | 4 203 | 3 745 | 1 584,7 | 1 976,5 | 2 325,5 | 2 579,3 | 2 746,6 | 2 917,8 | 3 107,6 | 3 304,4 | 3 520,0 | 3 745,0 | 3 991,8 | 4 256,8 | 4 549,6 | 4 879,7 | 5 264,2 | 5 707,4 | 6 261,8 | 6 995,1 | 8 178,9 | 23 610,6 |
| KH17 | 40 | 119 559 | 2 646 | 2 254 | 1 166,1 | 1 270,9 | 1 377,5 | 1 488,5 | 1 604,0 | 1 722,5 | 1 843,4 | 1 971,9 | 2 110,0 | 2 253,5 | 2 405,7 | 2 574,7 | 2 784,9 | 3 027,5 | 3 304,2 | 3 625,9 | 4 034,0 | 4 607,9 | 5 558,8 | 24 745,9 |
| KH18 | 41 | 45 360 | 3 263 | 2 888 | 1 336,7 | 1 512,5 | 1 681,1 | 1 859,9 | 2 035,6 | 2 215,0 | 2 397,4 | 2 568,8 | 2 720,1 | 2 888,3 | 3 069,4 | 3 267,3 | 3 488,7 | 3 738,1 | 4 025,8 | 4 393,4 | 4 841,6 | 5 450,8 | 6 493,2 | 25 585,5 |
| KH19 | 42 | 73 151 | 4 554 | 4 065 | 2 621,0 | 2 755,0 | 2 893,3 | 3 034,8 | 3 177,2 | 3 334,5 | 3 500,5 | 3 673,4 | 3 862,7 | 4 064,4 | 4 274,6 | 4 499,5 | 4 755,2 | 5 056,5 | 5 397,1 | 5 793,1 | 6 303,2 | 7 000,8 | 8 172,5 | 28 501,8 |
| KH20 | 43 | 35 914 | 4 571 | 4 070 | 2 570,7 | 2 705,2 | 2 850,9 | 3 004,1 | 3 155,6 | 3 312,4 | 3 487,7 | 3 669,4 | 3 861,8 | 4 069,6 | 4 284,9 | 4 527,1 | 4 797,5 | 5 103,9 | 5 447,5 | 5 861,8 | 6 382,0 | 7 126,6 | 8 333,2 | 40 421,4 |
| KH21 | 50 | 120 165 | 2 291 | 1 943 | 1 212,4 | 1 282,5 | 1 356,5 | 1 432,3 | 1 510,3 | 1 590,1 | 1 674,2 | 1 760,4 | 1 849,6 | 1 942,5 | 2 038,5 | 2 141,3 | 2 246,1 | 2 355,7 | 2 469,4 | 2 762,9 | 3 232,9 | 3 893,1 | 4 958,5 | 15 664,1 |
| KH22 | 51 | 137 752 | 4 572 | 4 109 | 2 515,0 | 2 664,9 | 2 817,2 | 2 978,6 | 3 143,5 | 3 319,0 | 3 501,2 | 3 691,9 | 3 891,3 | 4 109,4 | 4 339,2 | 4 594,2 | 4 870,6 | 5 186,9 | 5 539,8 | 5 956,9 | 6 479,9 | 7 167,4 | 8 304,6 | 33 452,0 |
| KH23 | 52 | 36 215 | 4 336 | 3 849 | 2 599,6 | 2 709,0 | 2 823,5 | 2 941,5 | 3 069,1 | 3 201,5 | 3 344,2 | 3 504,2 | 3 667,1 | 3 848,9 | 4 045,0 | 4 260,8 | 4 498,7 | 4 770,8 | 5 088,1 | 5 474,5 | 5 962,4 | 6 610,9 | 7 681,7 | 26 524,7 |
| KH24 | 53 | 267 939 | 4 633 | 4 099 | 2 637,0 | 2 772,2 | 2 912,9 | 3 057,5 | 3 208,4 | 3 365,3 | 3 528,8 | 3 707,9 | 3 896,9 | 4 099,3 | 4 320,3 | 4 562,6 | 4 827,9 | 5 128,9 | 5 477,8 | 5 896,2 | 6 428,4 | 7 154,5 | 8 393,3 | 83 222,3 |
| KH25 | 60 | 193 624 | 552 | 436 | 24,0 | 50,9 | 83,8 | 121,2 | 163,1 | 210,5 | 260,4 | 314,0 | 373,5 | 436,1 | 502,0 | 573,0 | 646,0 | 722,1 | 802,9 | 888,8 | 977,2 | 1 070,2 | 1 219,1 | 16 014,5 |
| KH26 | 61 | 50 454 | 1 692 | 1 669 | 161,2 | 1 182,5 | 1 234,0 | 1 285,9 | 1 342,0 | 1 400,7 | 1 463,2 | 1 528,9 | 1 595,8 | 1 668,7 | 1 741,8 | 1 821,0 | 1 904,3 | 1 990,9 | 2 082,9 | 2 178,5 | 2 281,0 | 2 391,4 | 2 517,7 | 10 099,3 |
| KH27 | 62 | 26 271 | 2 766 | 2 552 | 1 348,1 | 1 538,1 | 1 709,0 | 1 852,4 | 1 978,5 | 2 103,6 | 2 221,7 | 2 339,1 | 2 452,7 | 2 551,7 | 2 652,2 | 2 772,0 | 2 905,3 | 3 059,1 | 3 241,7 | 3 468,9 | 3 781,4 | 4 245,2 | 5 031,9 | 29 806,6 |
| KH28 | 63 | 76 724 | 3 224 | 2 933 | 1 632,4 | 1 913,0 | 2 120,9 | 2 299,2 | 2 463,6 | 2 559,5 | 2 640,9 | 2 728,8 | 2 827,7 | 2 933,2 | 3 052,2 | 3 184,3 | 3 340,5 | 3 517,5 | 3 732,8 | 3 997,6 | 4 345,7 | 4 831,4 | 5 676,7 | 43 822,0 |
| KH29 | 70 | 199 372 | 365 | 276 | 22,6 | 43,0 | 64,6 | 87,8 | 112,6 | 140,2 | 169,7 | 202,1 | 237,2 | 276,0 | 320,2 | 368,9 | 424,1 | 486,6 | 557,6 | 641,4 | 738,9 | 852,2 | 989,6 | 1 160,4 |
| KH30 | 71 | 28 913 | 598 | 565 | 58,5 | 119,2 | 174,6 | 228,6 | 281,5 | 335,2 | 390,3 | 444,8 | 502,3 | 564,5 | 622,8 | 684,5 | 744,3 | 808,2 | 873,5 | 940,2 | 1 009,9 | 1 079,9 | 1 149,5 | 4 207,3 |
| KH31 | 72 | 124 878 | 1 292 | 1 442 | 25,8 | 62,7 | 118,6 | 215,8 | 498,0 | 1 194,3 | 1 251,5 | 1 313,0 | 1 375,6 | 1 441,9 | 1 512,9 | 1 589,5 | 1 672,5 | 1 763,4 | 1 859,4 | 1 965,2 | 2 078,4 | 2 206,0 | 2 347,0 | 56 382,3 |
| KH32 | 73 | 347 565 | 1 220 | 1 415 | 14,7 | 32,2 | 59,1 | 102,6 | 184,8 | 445,0 | 1 200,4 | 1 267,9 | 1 339,4 | 1 414,7 | 1 493,6 | 1 577,2 | 1 665,5 | 1 759,2 | 1 859,6 | 1 967,5 | 2 084,2 | 2 212,6 | 2 351,1 | 42 716,6 |
| KH33 | 80 | 75 304 | 527 | 495 | 72,8 | 124,0 | 168,2 | 212,6 | 257,3 | 301,7 | 348,6 | 395,7 | 444,1 | 494,5 | 547,2 | 602,7 | 659,6 | 719,7 | 783,0 | 849,3 | 917,2 | 989,5 | 1 072,6 | 1 160,4 |
| KH34 | 81 | 103 656 | 417 | 334 | 36,0 | 62,9 | 90,2 | 118,1 | 146,9 | 178,4 | 211,3 | 248,3 | 289,0 | 334,2 | 383,6 | 437,9 | 498,8 | 567,4 | 643,6 | 728,1 | 821,9 | 926,2 | 1 037,9 | 1 160,4 |
| KH35 | 82 | 24 658 | 458 | 345 | 16,7 | 51,6 | 82,6 | 115,2 | 149,1 | 184,6 | 219,3 | 257,4 | 297,8 | 345,3 | 395,3 | 449,8 | 507,5 | 575,5 | 647,2 | 729,4 | 822,4 | 930,0 | 1 057,4 | 8 368,6 |
| KH36 | 83 | 31 206 | 982 | 1 199 | 0,9 | 3,8 | 9,1 | 15,0 | 22,1 | 32,4 | 47,8 | 76,5 | 155,9 | 1 199,2 | 1 264,9 | 1 335,1 | 1 417,4 | 1 509,6 | 1 622,1 | 1 759,1 | 1 930,6 | 2 162,3 | 2 502,6 | 57 514,7 |
| KH37 | 90 | 301 456 | 623 | 619 | 151,9 | 214,8 | 269,7 | 322,0 | 371,7 | 421,1 | 470,2 | 520,2 | 569,5 | 619,2 | 669,1 | 719,9 | 771,5 | 823,8 | 876,4 | 930,6 | 985,9 | 1 042,5 | 1 100,5 | 1 160,4 |
| KH38 | 91 | 252 820 | 322 | 241 | 27,8 | 47,1 | 66,0 | 86,4 | 107,8 | 130,7 | 155,3 | 181,6 | 209,5 | 241,0 | 275,7 | 314,9 | 358,7 | 409,4 | 467,9 | 537,8 | 624,2 | 738,1 | 897,9 | 1 160,3 |
| KH39 | 92 | 130 904 | 265 | 171 | 21,4 | 35,3 | 48,6 | 62,5 | 77,3 | 92,6 | 109,4 | 127,5 | 147,5 | 170,5 | 196,9 | 228,3 | 263,3 | 306,0 | 357,8 | 423,1 | 509,1 | 628,7 | 813,8 | 11 915,8 |
| KH40 | 93 | 425 926 | 143 | 62 | 5,5 | 9,6 | 13,7 | 18,2 | 23,3 | 29,0 | 35,6 | 43,1 | 52,0 | 62,4 | 74,8 | 89,8 | 108,4 | 131,6 | 163,2 | 206,9 | 271,4 | 377,9 | 584,3 | 26 849,0 |
| TOTAL | | 5 373 026 | 1 390 | 687 | 31,2 | 2 772,2 | 90,0 | 3 057,5 | 181,3 | 3 365,3 | 316,4 | 3 707,9 | 461,9 | 4 109,4 | 632,4 | 4 594,2 | 780,9 | 5 186,9 | 1 055,1 | 5 956,9 | 1 556,4 | 7 167,4 | 2 282,5 | 83 222,3 |



| CLASSE | | EFF. | TX1 | TX2 | TX3 | TX4 | TX5 | TX6 | TX7 | TX8 | TX9 | TX10 | TX11 | TX12 | TX13 | TX14 | TX15 | TX16 | TX17 | TX18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KH01 | 00 | 5 789 | 108,63 | 63,13 | 45,26 | 36,55 | 30,96 | 27,95 | 24,86 | 23,62 | 22,48 | 21,74 | 21,36 | 21,22 | 21,13 | 21,69 | 23,05 | 22,41 | 25,49 | 33,37 |
| KH02 | 01 | 1 580 | 109,62 | 61,89 | 44,29 | 35,62 | 29,93 | 27,54 | 25,55 | 23,07 | 22,50 | 21,87 | 21,06 | 20,76 | 20,11 | 20,74 | 21,69 | 22,18 | 24,58 | 28,22 |
| KH03 | 02 | 3 067 | 106,41 | 58,41 | 41,79 | 33,61 | 29,12 | 26,10 | 24,00 | 22,76 | 21,65 | 21,02 | 20,42 | 19,98 | 19,80 | 20,06 | 21,60 | 23,99 | 26,03 | 30,51 |
| KH04 | 03 | 2 345 | 109,01 | 59,80 | 42,32 | 35,48 | 30,08 | 27,94 | 26,19 | 24,98 | 24,36 | 23,48 | 23,11 | 23,05 | 22,70 | 23,48 | 24,13 | 25,23 | 27,30 | 35,39 |
| KH05 | 04 | 1 781 | 99,49 | 57,27 | 43,20 | 33,71 | 27,81 | 25,14 | 22,89 | 20,97 | 19,60 | 18,54 | 18,42 | 17,25 | 16,61 | 16,03 | 16,17 | 16,06 | 16,15 | 33,77 |
| KH06 | 05 | 4 534 | 98,18 | 56,94 | 40,91 | 31,68 | 28,12 | 23,79 | 20,39 | 20,55 | 18,84 | 18,01 | 16,98 | 16,84 | 16,24 | 15,86 | 14,47 | 14,19 | 13,86 | 14,33 |
| KH07 | 10 | 688 | 106,73 | 56,18 | 38,95 | 31,22 | 25,67 | 22,87 | 20,23 | 17,99 | 16,71 | 15,86 | 15,05 | 14,38 | 13,45 | 13,28 | 13,06 | 12,91 | 12,36 | 12,62 |
| KH08 | 11 | 378 | 105,92 | 58,47 | 40,16 | 32,00 | 27,76 | 24,72 | 21,80 | 19,91 | 19,14 | 17,94 | 16,89 | 16,39 | 15,16 | 15,46 | 15,16 | 20,49 | 23,96 | 25,23 |
| KH09 | 12 | 398 | 92,99 | 55,61 | 39,55 | 31,49 | 27,96 | 24,22 | 22,00 | 20,07 | 18,76 | 17,97 | 16,54 | 16,18 | 15,32 | 14,83 | 14,25 | 14,10 | 13,41 | 12,78 |
| KH10 | 13 | 472 | 100,21 | 54,17 | 39,50 | 30,19 | 25,21 | 21,90 | 18,98 | 17,34 | 15,53 | 14,38 | 12,46 | 12,09 | 10,96 | 10,36 | 9,96 | 9,18 | 8,84 | 8,49 |
| KH11 | 14 | 627 | 106,06 | 62,07 | 40,46 | 32,05 | 28,76 | 23,36 | 20,69 | 19,26 | 16,38 | 15,72 | 14,34 | 13,70 | 12,94 | 12,17 | 11,58 | 10,96 | 13,96 | 33,63 |
| KH12 | 15 | 1 388 | 113,10 | 60,81 | 42,60 | 33,25 | 28,00 | 24,11 | 21,77 | 19,49 | 18,02 | 17,02 | 15,47 | 14,66 | 13,74 | 12,72 | 12,89 | 12,82 | 13,61 | 14,42 |
| KH13 | 20 | 4 594 | 118,30 | 68,56 | 47,95 | 36,82 | 28,83 | 25,00 | 21,04 | 18,01 | 15,55 | 14,03 | 12,58 | 11,58 | 12,66 | 13,69 | 13,84 | 13,93 | 15,36 | 18,86 |
| KH14 | 21 | 3 075 | 9,85 | 9,47 | 8,76 | 8,68 | 7,99 | 7,84 | 7,53 | 7,65 | 6,63 | 6,66 | 6,60 | 7,19 | 7,53 | 8,22 | 9,08 | 10,31 | 13,70 | 20,35 |
| KH15 | 22 | 420 | 7,16 | 6,73 | 6,50 | 6,25 | 6,09 | 5,96 | 5,93 | 5,77 | 5,47 | 5,30 | 5,28 | 5,12 | 7,11 | 10,84 | 12,64 | 14,25 | 17,16 | 25,06 |
| KH16 | 23 | 3 872 | 24,73 | 17,66 | 10,91 | 6,49 | 6,23 | 6,50 | 6,33 | 6,52 | 6,39 | 6,59 | 6,64 | 6,88 | 7,26 | 7,88 | 8,42 | 9,71 | 11,71 | 16,92 |
| KH17 | 24 | 959 | 8,99 | 8,39 | 8,05 | 7,76 | 7,39 | 7,02 | 6,98 | 7,00 | 6,80 | 6,75 | 7,02 | 8,16 | 8,71 | 9,14 | 9,73 | 11,26 | 14,23 | 20,64 |
| KH18 | 25 | 2 661 | 13,15 | 11,15 | 10,64 | 9,44 | 8,81 | 8,23 | 7,15 | 5,89 | 6,18 | 6,27 | 6,45 | 6,78 | 7,15 | 7,69 | 9,13 | 10,20 | 12,58 | 19,12 |
| KH19 | 30 | 883 | 5,11 | 5,02 | 4,89 | 4,69 | 4,95 | 4,98 | 4,94 | 5,15 | 5,22 | 5,17 | 5,26 | 5,68 | 6,34 | 6,74 | 7,34 | 8,81 | 11,07 | 16,74 |
| KH20 | 31 | 1 037 | 5,23 | 5,39 | 5,38 | 5,04 | 4,97 | 5,29 | 5,21 | 5,24 | 5,38 | 5,29 | 5,65 | 5,97 | 6,39 | 6,73 | 7,61 | 8,87 | 11,67 | 16,93 |
| KH21 | 32 | 66 | 5,78 | 5,77 | 5,58 | 5,45 | 5,28 | 5,29 | 5,15 | 5,07 | 5,02 | 4,94 | 5,04 | 4,89 | 4,88 | 4,83 | 11,89 | 17,01 | 20,42 | 27,37 |
| KH22 | 33 | 1 429 | 5,96 | 5,71 | 5,73 | 5,53 | 5,58 | 5,49 | 5,45 | 5,40 | 5,60 | 5,59 | 5,88 | 6,02 | 6,49 | 6,81 | 7,53 | 8,78 | 10,61 | 15,87 |
| KH23 | 34 | 381 | 4,21 | 4,23 | 4,18 | 4,34 | 4,31 | 4,46 | 4,78 | 4,65 | 4,96 | 5,10 | 5,34 | 5,58 | 6,05 | 6,65 | 7,59 | 8,91 | 10,88 | 16,20 |
| KH24 | 35 | 3 392 | 5,12 | 5,07 | 4,96 | 4,94 | 4,89 | 4,86 | 5,08 | 5,10 | 5,19 | 5,39 | 5,61 | 5,82 | 6,23 | 6,80 | 7,64 | 9,03 | 11,30 | 17,32 |
| KH25 | 40 | 1 508 | 111,69 | 64,82 | 44,56 | 34,56 | 29,06 | 23,70 | 20,59 | 18,95 | 16,75 | 15,12 | 14,15 | 12,73 | 11,79 | 11,19 | 10,69 | 9,95 | 9,52 | 13,91 |
| KH26 | 41 | 1 340 | 633,59 | 4,35 | 4,21 | 4,36 | 4,37 | 4,47 | 4,49 | 4,37 | 4,57 | 4,38 | 4,55 | 4,58 | 4,55 | 4,62 | 4,59 | 4,70 | 4,84 | 5,28 |
| KH27 | 42 | 1 525 | 14,09 | 11,11 | 8,39 | 6,81 | 6,33 | 5,61 | 5,28 | 4,86 | 4,04 | 3,94 | 4,52 | 4,81 | 5,29 | 5,97 | 7,01 | 9,01 | 12,27 | 18,53 |
| KH28 | 43 | 956 | 17,19 | 10,87 | 8,40 | 7,15 | 3,89 | 3,18 | 3,33 | 3,63 | 3,73 | 4,06 | 4,33 | 4,91 | 5,30 | 6,12 | 7,09 | 8,71 | 11,18 | 17,50 |
| KH29 | 44 | 684 | 90,17 | 50,40 | 35,81 | 28,31 | 24,52 | 21,03 | 19,11 | 17,34 | 16,37 | 16,02 | 15,22 | 14,95 | 14,74 | 14,58 | 15,04 | 15,19 | 15,33 | 16,13 |
| KH30 | 45 | 2 217 | 103,76 | 46,45 | 30,93 | 23,13 | 19,06 | 16,44 | 13,96 | 12,92 | 12,38 | 10,34 | 9,90 | 8,74 | 8,59 | 8,08 | 7,63 | 7,41 | 6,93 | 6,45 |
| KH31 | 50 | 3 839 | 143,57 | 89,13 | 81,91 | 130,78 | 139,82 | 4,80 | 4,91 | 4,77 | 4,82 | 4,92 | 5,06 | 5,23 | 5,43 | 5,44 | 5,69 | 5,76 | 6,14 | 6,39 |
| KH32 | 51 | 277 | 118,82 | 83,58 | 73,58 | 80,03 | 140,84 | 169,74 | 5,62 | 5,64 | 5,62 | 5,58 | 5,60 | 5,60 | 5,63 | 5,70 | 5,80 | 5,93 | 6,16 | 6,26 |
| KH33 | 52 | 1 649 | 70,22 | 35,66 | 26,38 | 21,03 | 17,27 | 15,53 | 13,53 | 12,21 | 11,36 | 10,67 | 10,13 | 9,45 | 9,11 | 8,80 | 8,47 | 7,99 | 7,89 | 8,40 |
| KH34 | 53 | 799 | 74,74 | 43,50 | 30,94 | 24,35 | 21,45 | 18,44 | 17,50 | 16,40 | 15,64 | 14,78 | 14,16 | 13,91 | 13,76 | 13,43 | 13,12 | 12,88 | 12,70 | 12,05 |
| KH35 | 54 | 392 | 208,99 | 60,21 | 39,48 | 29,41 | 23,76 | 18,82 | 17,38 | 15,70 | 15,94 | 14,46 | 13,78 | 12,83 | 13,39 | 12,46 | 12,71 | 12,76 | 13,08 | 13,70 |
| KH36 | 55 | 4 055 | 322,22 | 139,61 | 65,18 | 46,88 | 46,76 | 47,38 | 60,19 | 103,73 | 669,03 | 5,48 | 5,54 | 6,17 | 6,50 | 7,46 | 8,45 | 9,74 | 12,00 | 15,74 |
| KH37 | 60 | 353 | 41,38 | 25,55 | 19,40 | 15,44 | 13,28 | 11,67 | 10,62 | 9,49 | 8,73 | 8,05 | 7,60 | 7,16 | 6,78 | 6,38 | 6,18 | 5,94 | 5,75 | 5,56 |
| KH38 | 61 | 1 000 | 69,11 | 40,25 | 30,82 | 24,84 | 21,22 | 18,77 | 16,95 | 15,39 | 15,04 | 14,40 | 14,20 | 13,93 | 14,13 | 14,28 | 14,93 | 16,07 | 18,24 | 21,65 |
| KH39 | 62 | 904 | 64,60 | 37,69 | 28,67 | 23,61 | 19,82 | 18,09 | 16,55 | 15,76 | 15,56 | 15,50 | 15,91 | 15,33 | 16,25 | 16,93 | 18,25 | 20,33 | 23,48 | 29,45 |
| KH40 | 63 | 1 670 | 73,32 | 43,87 | 32,68 | 27,54 | 24,82 | 22,54 | 21,23 | 20,67 | 20,03 | 19,75 | 20,16 | 20,60 | 21,49 | 24,02 | 26,72 | 31,21 | 39,22 | 54,61 |
| MOYENNE | | | 39,16 | 38,60 | 29,60 | 23,07 | 18,69 | 16,44 | 15,15 | 13,07 | 12,38 | 11,70 | 11,36 | 11,18 | 11,19 | 11,45 | 12,07 | 12,98 | 14,62 | 19,14 |

Ce tableau nous donne le taux d’évolution qui est affecté à chaque consommateur en fonction de son groupe de clones et de son chiffre d’affaires actuel.

En bas du tableau, nous pouvons voir la moyenne de ces taux : si nous obtenons un taux d’évolution aberrant (>100%), nous remplaçons cette valeur excessivement élevée par la moyenne du taux calculée sur tous les groupes sauf ceux contenant une valeur extrême. Le tableau ci-dessous monte les corrections effectuées.

Par exemple, si un client appartient au groupe de clones 00 (KH01) et a un chiffre d'affaires en dessous de 21,5 euros, il aura un taux d'évolution égale au taux 1 : soit 89 %.
Si un client appartient à la classe de clones 00 (KH01) et a un chiffre d'affaires entre 21,5 et 40,5 euros, nous aurons alors un taux d'évolution égal au taux 2, soit 54 %.

Les taux d'évolution donnés pour les deux exemples sont très élevés, mais ils concernent très peu de clients.

Chaque client se voit affecter un taux d'évolution. Le taux multiplié par le chiffre d'affaires nous permet d'estimer un chiffre d'affaires potentiel pour chacun des clients.

**Correction des taux aberrants**

| CLASSE | | EFF. | TX1 | TX2 | TX3 | TX4 | TX5 | TX6 | TX7 | TX8 | TX9 | TX10 | TX11 | TX12 | TX13 | TX14 | TX15 | TX16 | TX17 | TX18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KH01 | 0 | 5 789 | 39,16 | 63,13 | 45,26 | 36,55 | 30,96 | 27,95 | 24,86 | 23,62 | 22,48 | 21,74 | 21,36 | 21,22 | 21,13 | 21,69 | 23,05 | 22,41 | 25,49 | 33,37 |
| KH02 | 1 | 1 580 | 39,16 | 61,89 | 44,29 | 35,62 | 29,93 | 27,54 | 25,55 | 23,07 | 22,50 | 21,87 | 21,06 | 20,76 | 20,11 | 20,74 | 21,69 | 22,18 | 24,58 | 28,22 |
| KH03 | 2 | 3 067 | 39,16 | 58,41 | 41,79 | 33,61 | 29,12 | 26,10 | 24,00 | 22,76 | 21,65 | 21,02 | 20,42 | 19,98 | 19,80 | 20,06 | 21,60 | 23,99 | 26,03 | 30,51 |
| KH04 | 3 | 2 345 | 39,16 | 59,80 | 42,32 | 35,48 | 30,08 | 27,94 | 26,19 | 24,98 | 24,36 | 23,48 | 23,11 | 23,05 | 22,70 | 23,48 | 24,13 | 25,23 | 27,30 | 35,39 |
| KH07 | 10 | 688 | 39,16 | 56,18 | 38,95 | 31,22 | 25,67 | 22,87 | 20,23 | 17,99 | 16,71 | 15,86 | 15,05 | 14,38 | 13,45 | 13,28 | 13,06 | 12,91 | 12,36 | 12,62 |
| KH08 | 11 | 378 | 39,16 | 58,47 | 40,16 | 32,00 | 27,76 | 24,72 | 21,80 | 19,91 | 19,14 | 17,94 | 16,89 | 16,39 | 15,16 | 15,46 | 15,16 | 20,49 | 23,96 | 25,23 |
| KH10 | 13 | 472 | 39,16 | 54,17 | 39,50 | 30,19 | 25,21 | 21,90 | 18,98 | 17,34 | 15,53 | 14,38 | 12,46 | 12,09 | 10,96 | 10,36 | 9,96 | 9,18 | 8,84 | 8,49 |
| KH11 | 14 | 627 | 39,16 | 62,07 | 40,46 | 32,05 | 28,76 | 23,36 | 20,69 | 19,26 | 16,38 | 15,72 | 14,34 | 13,70 | 12,94 | 12,17 | 11,58 | 10,96 | 13,96 | 33,63 |
| KH12 | 15 | 1 388 | 39,16 | 60,81 | 42,60 | 33,25 | 28,00 | 24,11 | 21,77 | 19,49 | 18,02 | 17,02 | 15,47 | 14,66 | 13,74 | 12,72 | 12,89 | 12,82 | 13,61 | 14,42 |
| KH13 | 20 | 4 594 | 39,16 | 68,56 | 47,95 | 36,82 | 28,83 | 25,00 | 21,04 | 18,01 | 15,55 | 14,03 | 12,58 | 11,58 | 12,66 | 13,69 | 13,84 | 13,93 | 15,36 | 18,86 |
| KH25 | 40 | 1 508 | 39,16 | 64,82 | 44,56 | 34,56 | 29,06 | 23,70 | 20,59 | 18,95 | 16,75 | 15,12 | 14,15 | 12,73 | 11,79 | 11,19 | 10,69 | 9,95 | 9,52 | 13,91 |
| KH26 | 41 | 1 340 | 39,16 | 4,35 | 4,21 | 4,36 | 4,37 | 4,47 | 4,49 | 4,37 | 4,57 | 4,38 | 4,55 | 4,58 | 4,55 | 4,62 | 4,59 | 4,70 | 4,84 | 5,28 |
| KH30 | 45 | 2 217 | 39,16 | 46,45 | 30,93 | 23,13 | 19,06 | 16,44 | 13,96 | 12,92 | 12,38 | 10,34 | 9,90 | 8,74 | 8,59 | 8,08 | 7,63 | 7,41 | 6,93 | 6,45 |
| KH31 | 50 | 3 839 | 39,16 | 89,13 | 81,91 | 23,07 | 18,69 | 4,80 | 4,91 | 4,77 | 4,82 | 4,92 | 5,06 | 5,23 | 5,43 | 5,44 | 5,69 | 5,76 | 6,14 | 6,39 |
| KH32 | 51 | 277 | 39,16 | 83,58 | 73,58 | 80,03 | 18,69 | 16,44 | 5,62 | 5,64 | 5,62 | 5,58 | 5,60 | 5,60 | 5,63 | 5,70 | 5,80 | 5,93 | 6,16 | 6,26 |
| KH35 | 54 | 392 | 39,16 | 60,21 | 39,48 | 29,41 | 23,76 | 18,82 | 17,38 | 15,70 | 15,94 | 14,46 | 13,78 | 12,83 | 13,39 | 12,46 | 12,71 | 12,76 | 13,08 | 13,70 |
| KH36 | 55 | 4 055 | 39,16 | 38,60 | 65,18 | 46,88 | 46,76 | 47,38 | 60,19 | 13,07 | 12,38 | 5,48 | 5,54 | 6,17 | 6,50 | 7,46 | 8,45 | 9,74 | 12,00 | 15,74 |

La table ci-dessus montre le remplacement des taux aberrants par la moyenne (> 100%).

### *4. Principaux résultats*

| | CA an (CA filtré 12 mois) | Potentiel de CA | Taux d'augmentation |
|---|---|---|---|
| **Total** | 7 467 726 175 € | 953 284 621 € | 12,77 % |
| **Moyenne** | 1 390 € | 177 € | - |

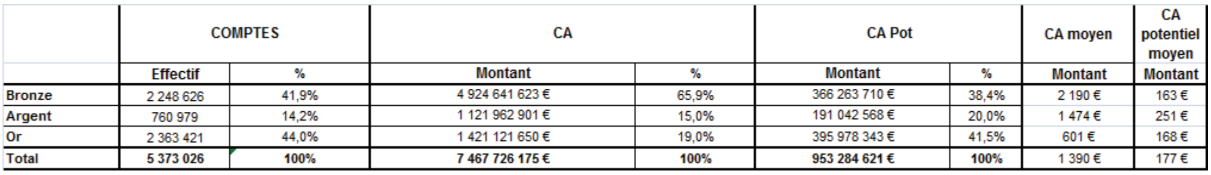
Le taux incrémental du potentiel client est de 12,77 %.
Ce distributeur peut gagner 12,77 % de chiffre d'affaires, sur ses clients.

Les clients affectés d'un chiffre d'affaires <= 0 ne sont pas retenus pour le score de potentiel puisque nous ne sommes concernés que par les clients actifs.

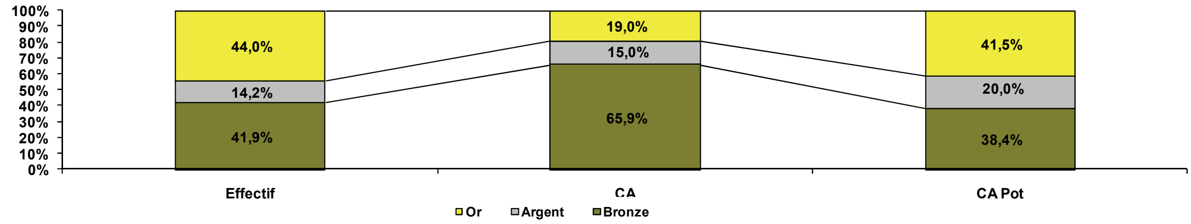
La situation présente vis-à-vis des clients à plus forte valeur semble être plus de la rétention client.

Les classes de potentiel :

On regroupe les taux de potentiel par classe de clones en trois catégories.

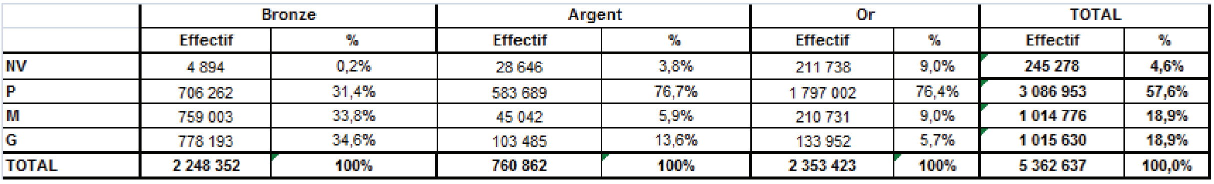


| | COMPTES | | CA | | CA Pot | | CA moyen | CA potentiel moyen |
|---|---|---|---|---|---|---|---|---|
| | Effectif | % | Montant | % | Montant | % | Montant | Montant |
| Bronze | 2 248 626 | 41,9% | 4 924 641 623 € | 65,9% | 366 263 710 € | 38,4% | 2 190 € | 163 € |
| Argent | 760 979 | 14,2% | 1 121 962 901 € | 15,0% | 191 042 568 € | 20,0% | 1 474 € | 251 € |
| Or | 2 363 421 | 44,0% | 1 421 121 650 € | 19,0% | 395 978 343 € | 41,5% | 601 € | 168 € |
| Total | 5 373 026 | 100% | 7 467 726 175 € | 100% | 953 284 621 € | 100% | 1 390 € | 177 € |

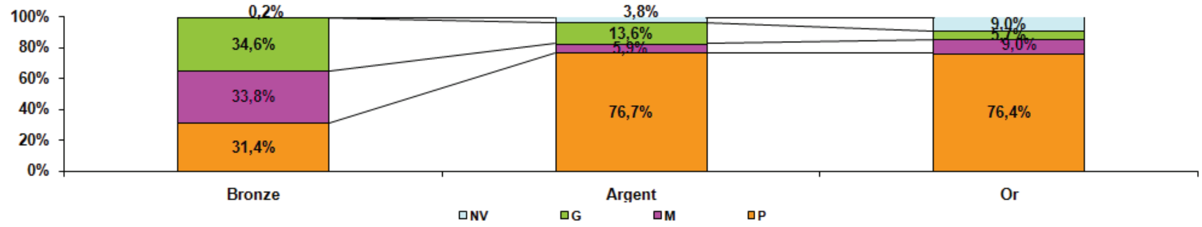


41,9 % des clients du distributeur ont un potentiel « Bronze » générant un chiffre d'affaires annuel de 65,9 % et comptant pour 38,4 % du chiffre d'affaires potentiel. D'un autre coté 44 % des clients sont des potentiels « Or », générant seulement 19 % du chiffre d'affaires mais générant 41,5 % du chiffre d'affaires potentiel.

Le regroupement en segment PMG :



| | Bronze | | Argent | | Or | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | Effectif | % | Effectif | % | Effectif | % | Effectif | % |
| NV | 4 894 | 0,2% | 28 646 | 3,8% | 211 738 | 9,0% | 245 278 | 4,6% |
| P | 706 262 | 31,4% | 583 689 | 76,7% | 1 797 002 | 76,4% | 3 086 953 | 57,6% |
| M | 759 003 | 33,8% | 45 042 | 5,9% | 210 731 | 9,0% | 1 014 776 | 18,9% |
| G | 778 193 | 34,6% | 103 485 | 13,6% | 133 952 | 5,7% | 1 015 630 | 18,9% |
| TOTAL | 2 248 352 | 100% | 760 862 | 100% | 2 353 423 | 100% | 5 362 637 | 100,0% |

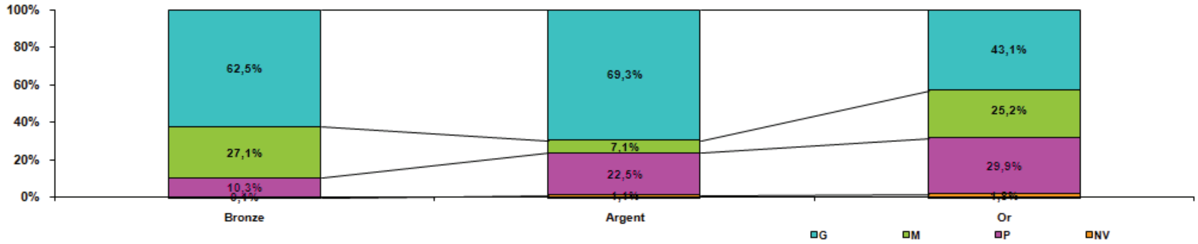


Les consommateurs P comptent pour une proportion élevée des potentiels « Or », basé sur le chiffre d'affaires annuel. Il y a des clients M parmi les potentiels « Bronze », et des clients G parmi les potentiels « Bronze » et « Argent ». La plupart d'entre eux ont une marge de progression intéressante.

En chiffre d'affaires annuel (en €):

| | Bronze | | Argent | | Or | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | Montant | % | Montant | % | Montant | % | Montant | % |
| NV | 2 612 344 € | 0,1% | 12 311 209 € | 1,1% | 25 441 780 € | 1,8% | **40 365 333 €** | **0,5%** |
| P | 507 518 386 € | 10,3% | 252 572 370 € | 22,5% | 425 060 654 € | 29,9% | **1 185 151 410 €** | **15,9%** |
| M | 1 334 543 548 € | 27,1% | 79 117 654 € | 7,1% | 358 381 176 € | 25,2% | **1 772 042 377 €** | **23,7%** |
| G | 3 079 967 346 € | 62,5% | 777 961 668 € | 69,3% | 612 238 041 € | 43,1% | **4 470 167 054 €** | **59,9%** |
| TOTAL | **4 924 641 623 €** | **100%** | **1 121 962 901 €** | **100%** | **1 421 121 650 €** | **100%** | **7 467 726 175 €** | **100,0%** |

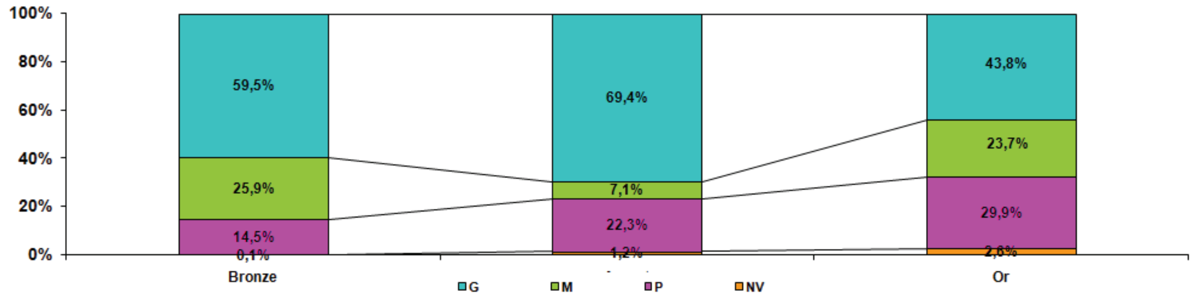
Les clients P sont sur représentés parmi les potentiels « Or », avec 29,9 % du chiffre d'affaires potentiel.



| | **Bronze** | **Argent** | **Or** | **TOTAL** |
|---|---|---|---|---|
| | **Montant moyen** | **Montant moyen** | **Montant moyen** | **Montant moyen** |
| **NV** | 534 € | 430 € | 120 € | 165 € |
| **P** | 719 € | 433 € | 237 € | 384 € |
| **M** | 1 758 € | 1 757 € | 1 701 € | 1 746 € |
| **G** | 3 958 € | 7 518 € | 4 571 € | 4 401 € |
| **TOTAL** | 2 190 € | 1 475 € | 604 € | 1 393 € |

En chiffre d'affaires potentiel (en k€) :

| | Bronze | | Argent | | Or | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | Montant | % | Montant | % | Montant | % | Montant | % |
| NV | 194 573 € | 0,1% | 2 201 431 € | 1,2% | 10 402 290 € | 2,6% | **12 798 293 €** | **1,3%** |
| P | 53 232 096 € | 14,5% | 42 569 788 € | 22,3% | 118 298 788 € | 29,9% | **214 100 672 €** | **22,5%** |
| M | 95 035 722 € | 25,9% | 13 600 554 € | 7,1% | 93 761 586 € | 23,7% | **202 397 862 €** | **21,2%** |
| G | 217 801 319 € | 59,5% | 132 670 795 € | 69,4% | 173 515 680 € | 43,8% | **523 987 794 €** | **55,0%** |
| TOTAL | **366 263 710 k€** | **100%** | **191 042 568 k€** | **100%** | **395 978 343 k€** | **100%** | **953 284 621 €** | **100,0%** |

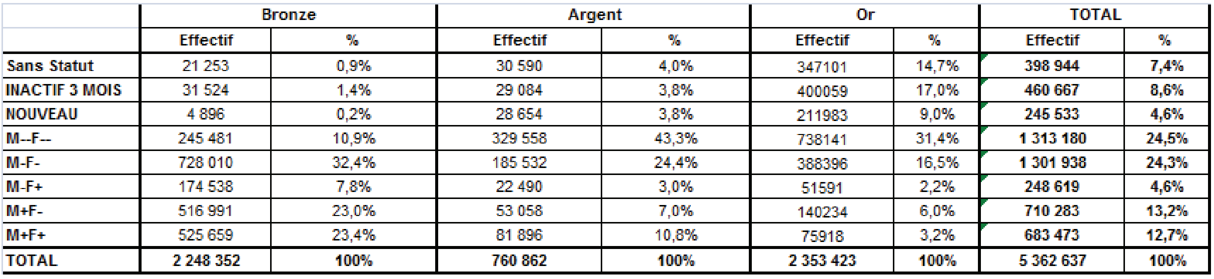


| | Bronze | Argent | Or | TOTAL |
|---|---|---|---|---|
| | Montant moyen | Montant moyen | Montant moyen | Montant moyen |
| **NV** | 40 € | 77 € | 49 € | 52 € |
| **P** | 75 € | 73 € | 66 € | 69 € |
| **M** | 125 € | 302 € | 445 € | 199 € |
| **G** | 280 € | 1 282 € | 1 295 € | 516 € |
| **TOTAL** | 163 € | 251 € | 168 € | 178 € |

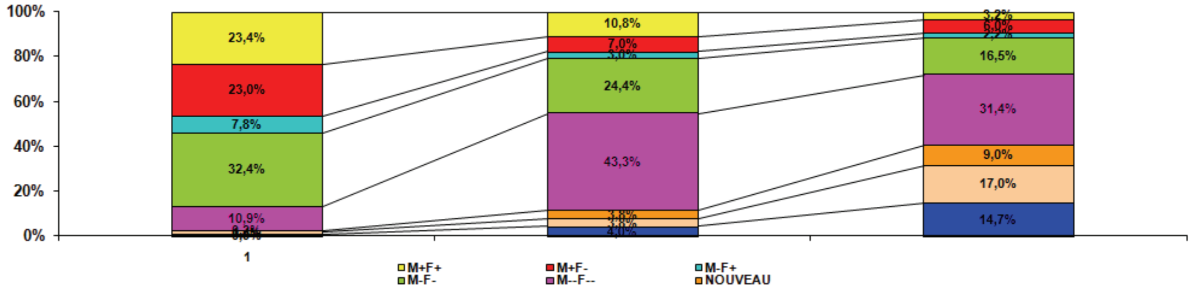
Les clients G ont le plus important potentiel en valeur absolue, bien qu'ils ne soient pas les plus forts taux d'évolution. Ils sont dans cette situation car ils ont un chiffre d'affaires nettement plus significatif que les segments P et M.

La distribution par la RFM du distributeur et par les catégories de potentiel.

En effectifs :

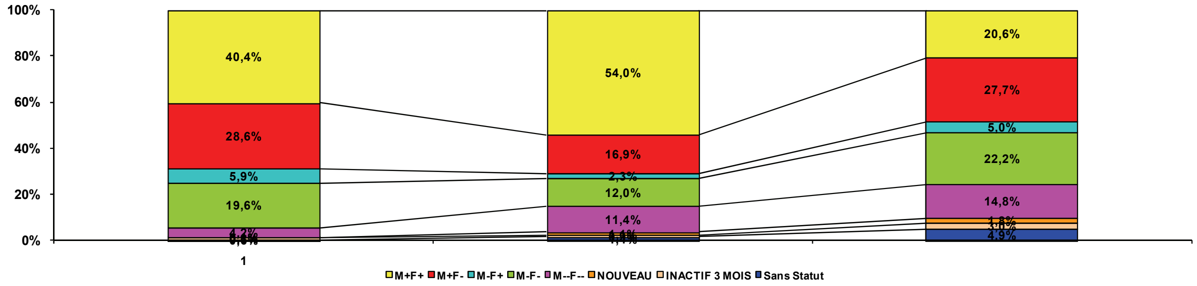


| | Bronze | | Argent | | Or | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | Effectif | % | Effectif | % | Effectif | % | Effectif | % |
| Sans Statut | 21 253 | 0,9% | 30 590 | 4,0% | 347101 | 14,7% | 398 944 | 7,4% |
| INACTIF 3 MOIS | 31 524 | 1,4% | 29 084 | 3,8% | 400059 | 17,0% | 460 667 | 8,6% |
| NOUVEAU | 4 896 | 0,2% | 28 654 | 3,8% | 211983 | 9,0% | 245 533 | 4,6% |
| M--F-- | 245 481 | 10,9% | 329 558 | 43,3% | 738141 | 31,4% | 1 313 180 | 24,5% |
| M-F- | 728 010 | 32,4% | 185 532 | 24,4% | 388396 | 16,5% | 1 301 938 | 24,3% |
| M-F+ | 174 538 | 7,8% | 22 490 | 3,0% | 51591 | 2,2% | 248 619 | 4,6% |
| M+F- | 516 991 | 23,0% | 53 058 | 7,0% | 140234 | 6,0% | 710 283 | 13,2% |
| M+F+ | 525 659 | 23,4% | 81 896 | 10,8% | 75918 | 3,2% | 683 473 | 12,7% |
| TOTAL | 2 248 352 | 100% | 760 862 | 100% | 2 353 423 | 100% | 5 362 637 | 100% |



En CA annuel (en €) :

| | Bronze | | Argent | | Or | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | CA | % | CA | % | CA | % | CA | % |
| Sans Statut | 28 482 104 € | 0,6% | 15 908 974 € | 1,4% | 69 877 281 € | 4,9% | 114 268 358 € | 1,5% |
| INACTIF 3 MOIS | 38 084 339 € | 0,8% | 10 476 055 € | 0,9% | 42 374 499 € | 3,0% | 90 934 893 € | 1,2% |
| NOUVEAU | 2 613 184 € | 0,1% | 12 313 384 € | 1,1% | 25 462 408 € | 1,8% | 40 388 976 € | 0,5% |
| M--F-- | 206 804 797 € | 4,2% | 128 061 467 € | 11,4% | 210 286 948 € | 14,8% | 545 153 212 € | 7,3% |
| M-F- | 964 638 682 € | 19,6% | 134 255 494 € | 12,0% | 316 010 822 € | 22,2% | 1 414 904 998 € | 18,9% |
| M-F+ | 288 843 514 € | 5,9% | 25 721 163 € | 2,3% | 70 733 841 € | 5,0% | 385 298 518 € | 5,2% |
| M+F- | 1 407 713 946 € | 28,6% | 189 736 462 € | 16,9% | 393 166 652 € | 27,7% | 1 990 617 060 € | 26,7% |
| M+F+ | 1 987 461 057 € | 40,4% | 605 489 903 € | 54,0% | 293 209 200 € | 20,6% | 2 886 160 159 € | 38,6% |
| TOTAL | 4 924 641 623 k€ | 100% | 1 121 962 901 k€ | 100% | 1 421 121 650 k€ | 100% | 7 467 726 175 € | 100,0% |



| | Bronze | Argent | Or | TOTAL |
|---|---|---|---|---|
| | Montant moyen | Montant moyen | Montant moyen | Montant moyen |
| Sans Statut | 1 340 € | 520 € | 201 € | 286 € |
| INACTIF 3 MOIS | 1 208 € | 360 € | 106 € | 197 € |
| NOUVEAU | 534 € | 430 € | 120 € | 164 € |
| M--F-- | 842 € | 389 € | 285 € | 415 € |
| M-F- | 1 325 € | 724 € | 814 € | 1 087 € |
| M-F+ | 1 655 € | 1 144 € | 1 371 € | 1 550 € |
| M+F- | 2 723 € | 3 576 € | 2 804 € | 2 803 € |
| M+F+ | 3 781 € | 7 393 € | 3 862 € | 4 223 € |
| TOTAL | 2 190 € | 1 475 € | 604 € | 1 393 € |

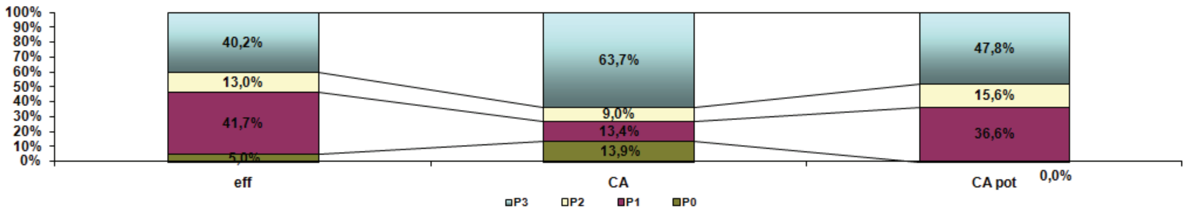
Logiquement les potentiels les plus importants sont les plus présents en valeur absolue dans les segments RFM +.

Les catégories de potentiel :

Les taux de potentiel par classe de clones sont regroupés en quatre catégories :
– P0 : Pas de CA potentiel
– P1 : Potentiel supérieur à 20 %

– P2: Potentiel supérieur ou égale à 15 et inférieur à 20 %
– P3: Potentiel inférieur à 15%

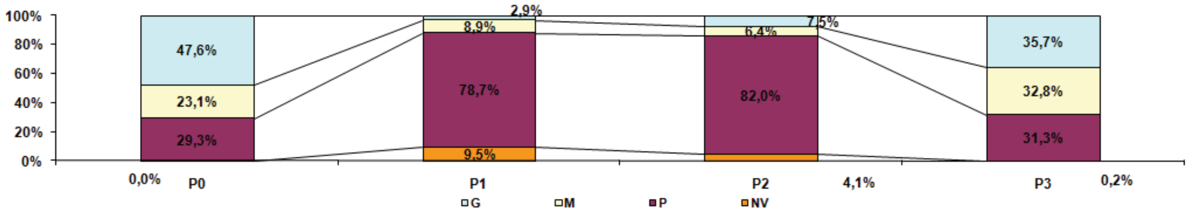
| | COMPTES | | CA | | CA Pot | | CA moyen | CA potentiel moyen |
|---|---|---|---|---|---|---|---|---|
| | Effectif | % | Montant | % | Montant | % | Montant | Montant |
| P0 | 268 675 | 5,0% | 1 035 180 496 € | 13,9% | 0 € | 0,0% | 3 853 € | 0 € |
| P1 | 2 242 226 | 41,7% | 999 789 812 € | 13,4% | 268 786 859 € | 36,6% | 446 € | 120 € |
| P2 | 699 709 | 13,0% | 672 185 375 € | 9,0% | 114 265 871 € | 15,6% | 961 € | 163 € |
| P3 | 2 162 416 | 40,2% | 4 760 570 492 € | 63,7% | 350 848 889 € | 47,8% | 2 202 € | 162 € |
| Total | **5 373 026** | **100%** | **7 467 726 175 €** | **100%** | **733 901 619 €** | **100%** | 1 390 € | 137 € |



40 % des clients créent 63 % du chiffre d'affaires et 47,8 % du CA potentiel. En moyenne, ils réalisent un chiffre d'affaires de 2 202 € pour un potentiel moyen de 162 €. Ces clients qui ont déjà une contribution substantielle, sont les plus à même (pour le moindre effort perçu) d'atteindre leur potentiel.

Le regroupement des segments PMG :

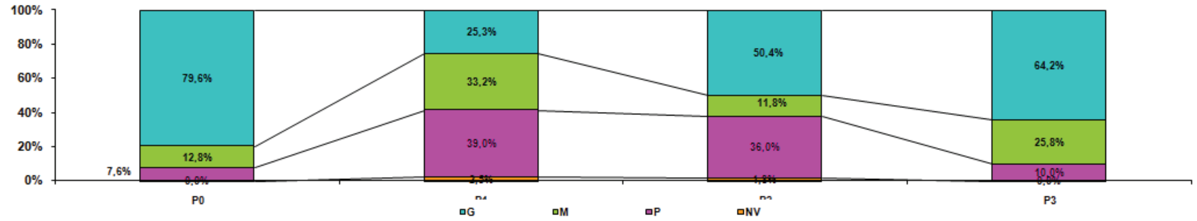
| | P0 | | P1 | | P2 | | P3 | | TOTAL | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Effectif | % | Effectif | % | Effectif | % | Effectif | % | Effectif | % |
| NV | 117 | 0,0% | 211 720 | 9,5% | 28 646 | 4,1% | 4 795 | 0,2% | 245 278 | 4,6% |
| P | 78 682 | 29,3% | 1 757 427 | 78,7% | 573 720 | 82,0% | 677 124 | 31,3% | 3 086 953 | 57,6% |
| M | 61 962 | 23,1% | 198 393 | 8,9% | 45 032 | 6,4% | 709 389 | 32,8% | 1 014 776 | 18,9% |
| G | 127 914 | 47,6% | 64 688 | 2,9% | 52 194 | 7,5% | 770 834 | 35,7% | 1 015 630 | 18,9% |
| TOTAL | **268 675** | **100%** | **2 232 228** | **100%** | **699 592** | **100%** | **2 162 142** | **100,0%** | **5 362 637** | **100,0%** |



Les clients P représentent une part élevée de P1 et de P2 basée sur le CA annuel généré par les potentiels P1. Nous trouvons des clients M principalement par les potentiels P3, alors que les clients G pour leur part se retrouvent dans P0 mais aussi dans P3.

En chiffre d’affaires annuel (en €) :

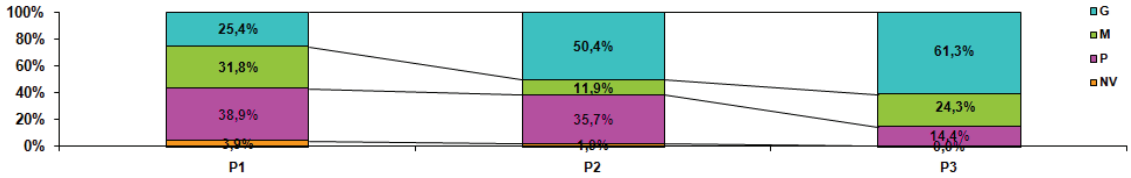
| | P0 | | P1 | | P2 | | P3 | | TOTAL | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Montant | % | Montant | % | Montant | % | Montant | % | Montant | % |
| NV | 489 934 € | 0,0% | 25 376 536 € | 2,5% | 12 311 209 € | 1,8% | 2 187 654 € | 0,0% | 40 365 333 € | 0,5% |
| P | 78 985 251 € | 7,6% | 389 503 691 € | 39,0% | 241 887 660 € | 36,0% | 474 774 808 € | 10,0% | 1 185 151 410 € | 15,9% |
| M | 132 044 286 € | 12,8% | 331 934 951 € | 33,2% | 79 092 619 € | 11,8% | 1 228 970 521 € | 25,8% | 1 772 042 377 € | 23,7% |
| G | 823 661 025 € | 79,6% | 252 974 633 € | 25,3% | 338 893 887 € | 50,4% | 3 054 637 509 € | 64,2% | 4 470 167 054 € | 59,9% |
| **TOTAL** | **1 035 180 496 €** | **100%** | **999 789 812 €** | **100%** | **672 185 375 €** | **100%** | **4 760 570 492 €** | **100,0%** | **7 467 726 175 €** | **100,0%** |



| | P0 | P1 | P2 | P3 | TOTAL |
|---|---|---|---|---|---|
| | **Montant moyen** | **Montant moyen** | **Montant moyen** | **Montant moyen** | **Montant moyen** |
| **NV** | 4 187 € | 120 € | 430 € | 456 € | 165 € |
| **P** | 1 004 € | 222 € | 422 € | 701 € | 384 € |
| **M** | 2 131 € | 1 673 € | 1 756 € | 1 732 € | 1 746 € |
| **G** | 6 439 € | 3 911 € | 6 493 € | 3 963 € | 4 401 € |
| **TOTAL** | 3 853 € | 448 € | 961 € | 2 202 € | 1 393 € |

En CA potentiel (en €) :

| | P1 | | P2 | | P3 | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | Montant | % | Montant | % | Montant | % | Montant | % |
| NV | 10 381 965 € | 3,9% | 2 201 431 € | 1,9% | 165 329 € | 0,0% | 12 748 726 € | 1,7% |
| P | 104 608 676 € | 38,9% | 40 846 282 € | 35,7% | 50 583 326 € | 14,4% | 196 038 284 € | 26,7% |
| M | 85 568 776 € | 31,8% | 13 596 614 € | 11,9% | 85 102 427 € | 24,3% | 184 267 817 € | 25,1% |
| G | 68 227 442 € | 25,4% | 57 621 544 € | 50,4% | 214 997 807 € | 61,3% | 340 846 793 € | 46,4% |
| **TOTAL** | **268 786 859 €** | **100%** | **114 265 871 €** | **100%** | **350 848 889 €** | **100%** | **733 901 619 €** | **100,0%** |



Les clients P sont sur représentés dans la catégorie « P1 » avec un chiffre d’affaires potentiel de 38,9% pour ce segment. En terme absolu, ce sont vraiment les clients G qui ont le potentiel le plus élevé. C’est avec le « gros » client que nous pouvons accroitre le chiffre d’affaire et que celui-ci a le plus de chance d’être atteint, que dans n'importe quel autre segment. Le budget marketing et l'intensité des offres marketing peuvent donc être alloués sur la base d'un chiffre d'affaire moyen par potentiel. Les deux concepts sont complémentaires dans la définition des mécaniques de fidélisation/rétention.

### *5. Résumé des résultats*

Le réseau de Kohonen nous permet de regrouper les clients en 50 classes de clones. La matrice 4X10 n'a pas de groupe vide, donc nous la retenons.

Les clients dans un même groupe se ressemblent entre eux sur des caractéristiques socio démographiques et de consommation.

En utilisant la méthode des déciles, nous affectons un taux d'évolution de chiffre d'affaires à chaque client de la base.

Nous créons les scores de potentiels suivants :
– « Or » : taux d'évolution supérieur à 20 %
– « Argent » : taux d'évolution supérieur ou égale à 15 % inférieur à 20 %
– « Bronze » : taux d'évolution inférieur à 15 %

Nous calculons le chiffre d'affaires potentiel à partir de ce taux et du chiffre d'affaires réel.

Notre base est composée de 5 373 026 clients générant un chiffre d'affaires annuel de 7,46 milliards d'euros, et représentant un chiffre d'affaires potentiel de 953,2 millions d'euros.

Le distributeur peut gagner presque 12,77 % de chiffre d'affaires sur ses clients.

En termes de taux, la base est composée de 41,9 % de clients « Bronze », de 14,1 % clients « Argent » et de 44 % de clients « OR ». Dans la réalité, il doit être admis que les meilleurs clients sont ceux avec les plus fortes valeurs absolues de potentiel.
76,6 % sont des consommateurs « Or » et P et ils génèrent 29,9 % du chiffre d'affaire annuel et 38,4 % du chiffre d'affaire potentiel.
65 % sont des clients « Or », inactifs 3 mois, et RFM – et RFM – et ils génèrent 40 % du chiffre d'affaires annuel et 38,4 % du chiffre d'affaires potentiel.

Logiquement les clients avec les taux d'évolution les plus élevés, se retrouvent parmi les clients avec les plus petits chiffres d'affaires en valeur.

À la fin de l'échelle, les clients avec le chiffre d'affaires le plus élevé, ont le potentiel le plus fort en valeur absolue.

| | P1 | P2 | P3 | TOTAL |
|---|---|---|---|---|
| | Montant moyen | Montant moyen | Montant moyen | Montant moyen |
| NV | 49 € | 77 € | 34 € | 52 € |
| P | 60 € | 71 € | 75 € | 64 € |
| M | 431 € | 302 € | 120 € | 182 € |
| G | 1 055 € | 1 104 € | 279 € | 336 € |
| TOTAL | 120 € | 163 € | 162 € | 137 € |

## Les procédures de validation de nos modèles

**La validité interne :**

Nous avons mené plusieurs tests sur notre modèle :

- La division de notre population en sous populations pour vérifier la cohérence d'allocation des classes de clones
- Le benchmark entre différentes techniques de classification
- La réaffectation des groupes par des modèles supervisés (C5.0, Réseaux Bayésiens)
- La connectivité des super classes

Les méthodes de validation devront bien sûre être complétées.

**La validé externe :**

Le taux de nourriture par client selon TNS est de 24 %. Sur, l'ensemble de la consommation, le potentiel atteignable sur le taux de nourriture la fait croitre de 28 %. Un différentiel de 2 % de taux de nourriture entre le réel et le prédit semble tout à fait admissible.

Nous aurions pu dédupliquer notre base avec le panel Home Scan de Nielsen pour vérifier si les ventes s'accroissent réellement, mais cela n'était pas possible dans ce contexte précis.

## Conclusions

### *Discussion des résultats de l'étude*

Les résultats de notre recherche doivent être replacés dans le contexte de la détermination du potentiel client d'une entreprise donnée : déterminer le potentiel client va impacter fortement le choix d'investissement en marketing direct ou en promotion d'une société. La plupart des grands programmes de fidélisation sont basés sur cette notion. Nous allons regarder comment notre approche comparée à d'autres méthodes, nous permet de déterminer des résultats convergents pour nos questions de recherches :

- La technique de classification de Kohonen (SOM) est utilisée pour identifier des clients qui sont similaires et pour définir un potentiel réaliste ;
- Notre méthode de césurage en déciles des classes de clones donne un potentiel par client, plus fins que d'autres méthodes : un potentiel atteignable ;
- Nous pouvons estimer la stabilité des groupes de différentes manières et elles montrent une stabilité interne de la méthode ;
- Nous avons développé une approche pragmatique qui est une méthode de détermination des potentiels : « la méthode des clones ».

### *Les limites et la contribution de notre étude*

Nous n'avons utilisé que certaines techniques de classifications particulières dans le but de valider notre méthode. Nous avons montré les éventuelles limites statistiques de notre approche en termes de complexité ou de fiabilité des modèles utilisés.
Pour des raisons de faisabilité, nous n'avons travaillé que sur un seul secteur industriel : la grande distribution alimentaire en France, et nous n'avons pas utilisé d'autres données venant d'autres secteurs.
Nous n'avons pas eu accès pour le moment à des données de distributeurs étrangers.

Le calcul du CA potentiel dans les groupes est très empirique et devrait être plus scientifiquement justifié.

***Pistes de recherche***

Il y a plusieurs façons d'améliorer notre recherche :

- Affiner le choix de nos variables ;
- Déterminer une méthode plus empirique que la méthode des déciles/ médianes pour estimer le potentiel par groupe ;
- Faire plus de rotations du modèle sur d'autres secteurs industriels ; nous avons déjà fait cela, et la méthode s'applique plutôt bien, mais il est important que d'autres le testent ;
- Valider les résultats dans le temps, en observant la réalité des valeurs de potentiels prédites sur les ventes réelles.

Nous espérons que par son impact stratégique sur les résultats des entreprises, et le fait que son calcul ne soit basé que sur des données clients internes déjà en possession de l'entreprise, cette « méthode de clones » trouvera un usage important.

## Bibliographie

1. Aguilera, P. A., Frenich, A. G., Torres, J. A., Castro, H., Vidal, J. L. M., and Canton, M., 2001, *Application of the cohune neural network in coastal water management: Methodological development for the assessment and prediction of water quality*, Water Research, 35(17):4053-4062.
2. Anderson, B., 1999, *Kohonen neural networks and language. Brain and Language*, 70(1):86-94
3. B. Meunier, E. Dumas, I. Piec, D. Bechet, M. Hebraud, J. Proteome Res, 2007, *Assessment of hierarchical clustering methodologies for proteomic data mining,* les 4 versions : www.aseanbiotechnology.info
4. Baran, Stanley J. *Theories of Mass Communication*.
5. Benavent and Crie, http://christophe.benavent.free.fr/publications/ltv1.pdf
6. Beran, R., 1986, Discussion of Wu, C.F.J.: Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). Ann. Statist., 14:1295-1298.
7. Berend Wierenga and Gerrit Harm van Bruggen, 2000, *Marketing Management, Springer Support Systems: Principles, Tools, and Implementation, Springer*.
8. Berger, Paul D. and Nada I. Nasr (1998), « Customer lifetime value: Marketing models and applications » *Journal of Interactive Marketing*, 12 (1), pp. 17-30.
9. Bertrand Clarke et Dongchu Sun, *Reference priors under the Chi-Squared distance: The Indian Journal of Statistics 1997*, volume 59, Series A, Pt. 2, 215-231.
10. Boos, D.D., 2003, « Introduction to the bootstrap world. Statist », *Science*, 18:168-174.
11. Borko, H. and Bernick, M., « Automatic document classification », *Journal of the ACM*, 10, 151-162, 1963.
12. Bremer and Joyce, 1988, *Human Judgment,The SJT View*, North-Holand.
13. Bruce Cooil, Timothy L. Keiningham, Lerzan Aksoy, Michael Hsu, 2007, « A Longitudinal Analysis of Customer Satisfaction and Wallet share: Investigating the Moderating Effect of Customer Characteristics » *Journal of Marketing* 71:1, 67-83.
14. Charles Romesburg, « Cluster Analysis for Researchers » (2004) *Lulu press*, p. 135.
15. Ching-Hsue Cheng and You-Shyang Chen, *Classifying the segmentation of customer value via RFM model and RS theory Expert Systems with Applications*, In Press, Corrected Proof, Available online 16 April 2008, Collectif, « Recherche sur la distribution modern », p. 64, editions l'Univers du Livre.
16. Ciampi, A. and Lechevallier, Y., 2000, « Clustering large, multi-level data sets: an approach based on Kohonen self-organizing maps ». In *Principles of Data Mining and Knowledge Discovery*. 4th European Conference, PKDD 2000.

Proceedings (Lecture Notes in Artificial Intelligence Vol.1910). Springer-Verlag, Berlin, Germany, pp. 353-358.

17. Ciampi, A. and Lechevallier, Y., 2000, « Clustering large, multi-level data sets: an approach based on Kohonen self-organizing maps ». In *Principles of Data Mining and Knowledge Discovery*. 4th European Conference, PKDD 2000. Proceedings (Lecture Notes in Artificial Intelligence Vol. 1910). Springer-Verlag, Berlin, Germany, pp. 353-358

18. Dahbur, K. and Muscarello, T., 2001. « Hybrid Kohonen neural network in data mining ». In *Proceedings of the IASTED International Conference. Artificial Intelligence and Applications*. ACTA Press, Anaheim, CA, USA, pp. 30-33.

19. David Huff, 18 Jun 2003, University of Texas Austin, « A Retrospective View of the Huff Model and its Application to Spatial Interaction Analysis », University of Redlands/ESRI Colloquium Series.

20. Dorofeyuk, A. A., « Automatic Classification Algorithms (Review) », Automation and Remote Control, 32, 1928-1958, 1971.

21. Dwyer, R.F., 1997, « Customer lifetime valuation to support marketing decision making », *Journal of Direct Marketing*, vol. 11, n° 4, pp. 6-13.

22. Efron B., 1981, « Non parametric estimates of standard error: the jackknife, the bootstrap and other methods ». *Biometrika*, 68. pp. 589-599.

23. Eric Chen-Kuo Tsao, James C. Bezdek and Nikhil R. Pal, « Fuzzy Kohonen clustering networks », 1994, Published by Elsevier Science B.V.

24. F. V. Jensen, *Introduction to Bayesian Networks*, 1st edition, 1996, Springer-Verlag New York, Inc.

25. Fang K., He S., *The problem of selecting a given number of representative points in a normal population and a generalized mill's ratio*. Technical report, Department of Statistics, Stanford University, 1982. MacQueen J., « Some methods for classification and analysis of multivariate observations. Proceedings 5th Berkeley Symposium on Mathematics », *Statistics and Probability*, 1967;3:281-297.

26. Frank Plastria, *Static competitive facility location: An overview of optimisation approaches European Journal of Operational Research*, Volume 129, Issue 3, 16 March 2001, pp. 461-470.

27. Gehrlein W. V., « General mathematical programming formulations for the statistical classification problem », *Operations research letters* , ISSN : 0167-6377, CODEN ORLED5.

28. Harris, M. J. and N. Blisard, 1995, « Characteristics of the Nielsen Homescan Data », *Working paper*, Washington, DC: U.S. Department of Agriculture, Economic Research Service.

29. Hartigan J. A, Wong M. A., *A k-means clustering algorithm*, Applied Statistics, 1979;28:100–108.

30. http://en.wikipedia.org/wiki/Lifetime_value

31. J. R. Quinlan, « Improved use of continuous attributes », in *c4.5. Journal of Artificial Intelligence Research*, 4:77-90, 1996.
32. Jajuga K., *Classification, Clustering and Data Analysis : Recent Advances and Applications*, 2002, Lavoisier.
33. John A., McCartya and Manoj Hastak, « Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression », *Journal of Business Research*, volume 60, Issue 6, June 2007, pp. 656-662.
34. Juha Vesanto, 1997, *The SOM in data mining: analysis of world pulp and paper technology*.
35. Julien Barnier, *Tout ce que vous n'avez jamais voulu savoir sur le Chi2 sans jamais avoir eu envie de le demander*, Groupe de Recherche sur la Socialisation CNRS – UMR 5040, 15 avril 2008.
36. Kaski, S., « Data exploration using self-organizing maps. Acta Polytechnica Scandinavica, Mathematics, Computing and Management », in *Engineering Series* n° 82, Espoo, 1997.
37. Kohonen T., *Self-Organization and Associative Memory* , New York, Springer-Verlag, 1988.
38. Lerman I.C., *Les Bases de la classification automatique*, Gauthier-Villars, Paris, 1970.
39. M. Roux, 1985, *Algorithmes de classification*, Éditions Masson, Paris.
40. Mattias Otto, « Chemometrics Statistics and Computer Application », in *Analytical Chemistry,* 2007, Wiley-VCH.
41. Nielsen, Inc., May 2006, « Understanding the Homescan Advantage. » Presentation by Liz Crews and Ed Groves, Nielsen at RTI International, Research Triangle Park, NC.
42. O. Pourret, P. Naim and B. Marcot, 2008, *Bayesian Networks: A Practical Guide to Applications*, Chichester, UK, Wiley. ISBN 978-0-470-06030-8.
43. Olivier Brusset « Segmentation Cibler, scorer, analyser, une seule limite, les rendements », *Marketing direct* n° 92, 01/04/2005, p. 2.
44. Pena M., Vanegas A., Valencia, « Digital Hardware Architectures of Kohonen's Self Organizing Feature Maps with Exponential Neighboring Function », 2006, IEEE International Conference on Reconfigurable Computing and FPGA's J. (*ReConFig*, 2006), pp. 1-8.
45. Quinlan J. R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.
46. Quinlan R., 2004, *Data mining tools see5 and c5.0*.
47. Rajanee Ranjan, *Encyclopaedia of Marketing Research*, 2002, Anmol Publications PVT. LTD., p. 585.
48. Reilly, W. J., 1931, *The law of retail gravitation*, New York.
49. S. Kaski, J. Nikkila, and T. Kohonen, « Methods for Exploratory Cluster Analysis » *Intelligent Exploration of the Web*, De Piotr S., Szczepaniak, 2003 Springer.

50. « Size and Share of Customer Wallet », Rex Yuxing Du, Wagner A., Kamakura, Carl F. Mela*. Journal of Marketing*, volume 71, Issue: 2, pp. 94-113.
51. Tan, Peter J., Dowe David L., Dix Tevor I, *Building classification model in two steps*, 1997.
52. Teuvo Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 3rd edition, 1989.
53. Teuvo Kohonen, *Self-Organizing Maps*, 3rd edition, Springer, 20.
54. *The Useful Words from a Decisional Corpus*. « Contribution of Correspondence Analysis », Springer Berlin, Heidelberg, volume 185, 2005, pp. 159-179.
55. Timothy L. Keiningham, Bruce Cooil, Lerzan Aksoy, Tor W. Andreassen, Jay Weiner, 2007, « The value of different customer satisfaction and loyalty metrics » in *Predicting customer retention, recommendation, and share-of-wallet*, Managing Service Quality 17:4, 361-384.
56. Todd A. Stephenson, *An Introduction to Bayesian Network Theory and Usage* IDIAP-RR 00-03, 2000.
57. Vallaud Thierry, 2003, *La fidélisation rentable : la proposition du modèle composite*, www.numilog.com
58. Venkatesan, Rajkumar and V. Kumar, 2004, « A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy », *Journal of Marketing*, n° 68, October, pp. 106-125.

# Annexes

## *Annexe 1 : Traduction des noms des champs*

| Variables | Description |
|---|---|
| ID Client | Pour lier les tables entre elles sur un même client |
| Ratio Autre | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Bazar | % du chiffre d'affaire dans la catégorie de produit |
| Ratio BOF/APF | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Charcuterie LS | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Clients Animaux | % du chiffre d'affaire dans la catégorie de produit |
| Ratio client bébés | % du chiffre d'affaire dans la catégorie de produit |
| Ratio clients boucherie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio boulangerie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio charcuterie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio dietetiques bio | % du chiffre d'affaire dans la catégorie de produit |
| Ratio fromage | % du chiffre d'affaire dans la catégorie de produit |
| Ratio fruits et légumes | % du chiffre d'affaire dans la catégorie de produit |
| Ratio poisson | % du chiffre d'affaire dans la catégorie de produit |
| Ratio surgelés | % du chiffre d'affaire dans la catégorie de produit |
| Ration vins | % du chiffre d'affaire dans la catégorie de produit |
| Ratio DPH | % du chiffre d'affaire dans la catégorie de produit |
| Ratio épicerie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio liquide | % du chiffre d'affaire dans la catégorie de produit |
| Ratio textile | % du chiffre d'affaire dans la catégorie de produit |
| Ratio ultra frais | % du chiffre d'affaire dans la catégorie de produit |
| Ratio volaile | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Premier Prix | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Marque Distributeur 1 | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Marque Distributeur 2 | % du chiffre d'affaire dans la catégorie de produit |
| Nombre d'enfants au foyer | |
| CA Filtré 12 mois | Chiffre d'affaire diminué de quelques dépenses hors mag |
| CA Total | Chiffre d'Affaire total |
| CA Promo Annuel | CA annuel réalisé sous promo |
| Nb de points transformés sur 12 mois | Nb de point "promotionnels" utilisés sur une année |
| CM transformé sur 12 mois | Part CA ou le client a utilisé des points promotionnels |
| Cumul nb BA pris | Cumul des bons d'achat que le client a utilisé |
| PMG 12 mois | Segmentation |
| RFM 3 mois | Segmentation |

## *Annexe 2 : Détail du premier audit des données*

L'ensemble des données a été audité en deux étapes : une première étape pour déterminer toutes les données utiles à l'analyse dans la base de données originale, et une seconde étape pour déterminer les données disponibles pour calculer un potentiel. Dans cette annexe, seule la seconde étape est présentée.

### Analyse des tables « Potential_Ratio » et « Potentiel_SOCIO »

La table Potentiel_Ratio contient 5 373 048 observations (comptes client).
Elle est composée de 26 champs.

La table Potentiel_Socio contient 5 373 056 observations (comptes client).
Elle est composée de 18 champs.

Cet audit est basé sur la combinaison des deux tables soit 5 373 048 observations.

### Le format des données

C'est le format original des données reçues. Nous avons du changer certains formats pour mieux atteindre nos objectifs de modélisation.

| Variables | Description |
|---|---|
| Identifiant Client | Pour lier les tables entre elles sur un même client |
| Ratio Autre | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Bazar | % du chiffre d'affaire dans la catégorie de produit |
| Ratio BOF/APF | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Charcuterie LS | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Clients Animaux | % du chiffre d'affaire dans la catégorie de produit |
| Ratio client bébés | % du chiffre d'affaire dans la catégorie de produit |
| Ratio clients boucherie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio boulangerie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio charcuterie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio dietetiques bio | % du chiffre d'affaire dans la catégorie de produit |
| Ratio fromage | % du chiffre d'affaire dans la catégorie de produit |
| Ratio fruits et légumes | % du chiffre d'affaire dans la catégorie de produit |
| Ratio poisson | % du chiffre d'affaire dans la catégorie de produit |
| Ratio surgelés | % du chiffre d'affaire dans la catégorie de produit |
| Ration vins | % du chiffre d'affaire dans la catégorie de produit |
| Ratio DPH | % du chiffre d'affaire dans la catégorie de produit |

| | |
|---|---|
| Ratio épicerie | % du chiffre d'affaire dans la catégorie de produit |
| Ratio liquide | % du chiffre d'affaire dans la catégorie de produit |
| Ratio textile | % du chiffre d'affaire dans la catégorie de produit |
| Ratio ultra frais | % du chiffre d'affaire dans la catégorie de produit |
| Ratio volaile | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Premier Prix | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Marque Distributeur 1 | % du chiffre d'affaire dans la catégorie de produit |
| Ratio Marque Distributeur 2 | % du chiffre d'affaire dans la catégorie de produit |
| Nombre d'enfants au foyer | |
| CA Filtré 12 mois | Chiffre d'affaire diminué de quelques dépenses hors mag |
| CA Total | Chiffre d'Affaire total |
| CA Promo Annuel | CA annuel réalisé sous promo |
| Nb de points transformés sur 12 mois | Nb de point "promotionnels" utilisés sur une année |
| CM transformé sur 12 mois | Part CA ou le client a utilisé des points promotionnels |
| Cumul nb BA pris | Cumul des bons d'achat que le client a utilisé |
| PMG 12 mois | Segmentation |
| RFM 3 mois | Segmentation |

**RFM 3 mois la variable est vide, pour cette raison elle est écartée**

Analyse variable par variable.

Les variables discrètes

| RFM 3 mois | Effectif | % |
|---|---|---|
| Nouveaux | 247 326 | 4,60% |
| Anciens clients | 400 236 | 7,45% |
| Inactifs | 465 107 | 8,66% |
| M--F-- | 1 315 873 | 24,49% |
| M-F- | 1 302 089 | 24,23% |
| M-F+ | 248 619 | 4,63% |
| M+F- | 710 311 | 13,22% |
| M+F+ | 683 487 | 12,72% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1$^{er}$ audit |
|---|
| 3,74% |
| |
| 19,15% |
| 21,00% |
| 26,00% |
| 6,06% |
| 11,65% |
| 12,40% |
| 100,00% |

| Statut familial | Effectif | % |
|---|---|---|
| Couple | 1 557 871 | 28,99% |
| Célibataire | 642 374 | 11,96% |
| Vide | 3 172 803 | 59,05% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1er audit |
|---|
| 26,19% |
| 10,81% |
| 62,99% |
| 100,00% |

| PMG sur 12 mois | Effectif | % |
|---|---|---|
| NA | 7 816 | 0,15% |
| Nouveau | 245 278 | 4,56% |
| I nconnu | 2 573 | 0,05% |
| P | 3 086 955 | 57,45% |
| M | 1 014 777 | 18,89% |
| G | 1 015 649 | 18,90% |
| Total | 5 373 048 | 100,00% |

| Type d'habitat | Effectif | % |
|---|---|---|
| Appartement | 880 722 | 16,39% |
| Maison et appartement | 1 300 | 0,02% |
| Maison | 1 576 829 | 29,35% |
| Vide | 2 914 197 | 54,24% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1er audit |
|---|
| 18,76% |
| 0,00% |
| 34,98% |
| 65,02% |
| 100,00% |

| Nombre d'enfants dans le foyer | Effectif | % |
|---|---|---|
| 0 | 4 088 282 | 76,09% |
| 1 | 510 448 | 9,50% |
| 2 | 508 967 | 9,47% |
| 3 | 196 520 | 3,66% |
| 4 | 48 644 | 0,91% |
| 5 | 11 482 | 0,21% |
| > 5 | 8 705 | 0,16% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1er audit |
|---|
| 75,72% |
| 9,57% |
| 9,54% |
| 3,79% |
| 1,01% |
| 0,22% |
| 0,14% |
| 100,00% |

| Catégories sociales | Effectif | % |
|---|---|---|
| Fermier | 49 892 | 0,93% |
| Artisan | 86 807 | 1,62% |
| Autre | 84 696 | 1,58% |
| Directeur | 188 017 | 3,50% |
| Employé | 737 469 | 13,73% |
| Étudiant | 85 928 | 1,60% |
| Femme au foyer | 211 003 | 3,93% |
| Fonctionnaire | 233 386 | 4,34% |
| Travailleur indépendant | 42 913 | 0,80% |
| Ouvrier | 138 099 | 2,57% |
| Retraité | 664 304 | 12,36% |
| Sans emploi | 147 700 | 2,75% |
| Technicien | 91 956 | 1,71% |
| Vide | 2 610 877 | 48,59% |
| 24 | 1 | 0,00% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1er audit |
|---|
| 0,96% |
| 1,79% |
| 1,39% |
| 3,50% |
| 14,42% |
| 1,28% |
| 4,38% |
| 3,70% |
| 0,72% |
| 2,73% |
| 14,05% |
| 3,28% |
| 2,08% |
| 45,71% |
| 0,00% |
| 100,00% |

La valeur 24 est une erreur, nous l'avons éliminé.

| Age | Effectif | % |
|---|---|---|
| 0 à 18 ans | 8 604 | 0,16% |
| 19 à 29 ans | 317 371 | 5,91% |
| 30 à 39 ans | 537 010 | 9,99% |
| 40 à 49 ans | 652 497 | 12,14% |
| 50 à 59 ans | 649 038 | 12,08% |
| 60 à 69 ans | 458 669 | 8,54% |
| 70 ans et plus | 592 227 | 11,02% |
| Vide | 2 157 632 | 40,16% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1er audit |
|---|
| 0,21% |
| 6,00% |
| 10,76% |
| 12,59% |
| 12,30% |
| 8,07% |
| 10,92% |
| 39,15% |
| 100,00% |

| Historique client | Effectif | % |
|---|---|---|
| 0 à 2 mois | 194 269 | 3,62% |
| 3 à 5 mois | 238 267 | 4,43% |
| 6à 8 mois | 182 733 | 3,40% |
| 9 à 11 mois | 221 313 | 4,12% |
| 12 à 17 mois | 354 680 | 6,60% |
| 18 à 23 mois | 231 513 | 4,31% |
| 24 à 35 mois | 517 972 | 9,64% |
| 36 à 47 mois | 396 706 | 7,38% |
| 48 à 59 mois | 403 389 | 7,51% |
| 60 mois et plus | 2 631 717 | 48,98% |
| Vide | 489 | 0,01% |
| Total | 5 373 048 | 100,00% |

| Comparaison avec le 1er audit |
|---|
| 2,76% |
| 2,54% |
| 3,20% |
| 3,26% |
| 6,39% |
| 6,17% |
| 9,58% |
| 7,96% |
| 10,01% |
| 44,60% |
| 3,53% |
| 100,00% |

| Temps depuis le dernier achat | Effectif | % |
|---|---|---|
| 0 ta 2 mois | 4 213 358 | 78.42% |
| 3 a 5 mois | 562 908 | 10.48% |
| 6 a 8 mois | 312 721 | 5.82% |
| 9 a 11 mois | 241 093 | 4.49% |
| 12 a 17 mois | 42 968 | 0.80% |
| Total | 5 373 048 | 100.00% |

Les variables numériques

| | Ratio autre | Ratio bazar | Ratio BOF / APF | Ratio charcuterie allégée | Ratio Pet | Ratio beauté/maquillage | Ratio bébé | Ratio boucherie |
|---|---|---|---|---|---|---|---|---|
| Effectif | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 |
| Moyenne | 77,1 | 7,0 | 9,0 | 6,5 | 1,2 | 0,7 | 1,4 | 6,5 |
| Min | -34 035,0 | -27 666,7 | -195,2 | -1 580,0 | -15,3 | -61,2 | -23,5 | -125,2 |
| Max | 294,9 | 5 328,9 | 9 685,6 | 3 186,7 | 1 206,5 | 574,5 | 3 614,4 | 4 193,0 |
| SD | 22,2 | 16,3 | 8,4 | 6,7 | 3,7 | 2,4 | 5,8 | 8,6 |

| | Ratio boulangerie | Ratio charcuterie | Ratio nourriture diététique | Ratio fromage | Ratio fruits et légume | Ratio poisson | Ratio surgelé | Ratio vin |
|---|---|---|---|---|---|---|---|---|
| Effectif | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 |
| Moyenne | 2,5 | 2,6 | 0,4 | 1,5 | 8,4 | 1,9 | 3,1 | 2,1 |
| Min | -138,5 | -74,5 | -11,3 | -29,4 | -105,5 | -200,0 | -308,3 | -530,3 |
| Max | 1 247,6 | 1 373,1 | 2 428,6 | 262,7 | 17 364,3 | 1 542,9 | 11 528,6 | 610,2 |
| SD | 4.9 | 4,7 | 2,3 | 2,9 | 12,0 | 4,5 | 7,6 | 5,4 |

| | Ratio produits d'entretien | Ratio épicerie | Ratio liquide | Ratio textile | Ratio ultra frais | Ratio volaille | Ratio premier prix | Ratio marque distributeur 1 | Ratio marque distributeur 2 |
|---|---|---|---|---|---|---|---|---|---|
| Effectif | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 | 5 373 055 |
| Moyenne | 8,9 | 16,8 | 10,8 | 2,0 | 4,7 | 1,7 | 6,2 | 14,6 | 2,0 |
| Min | -1 472,8 | -1 463,6 | -2 843,2 | -1 250,0 | -372,3 | -37,5 | -393,8 | -194,9 | -30,9 |
| Max | 14 007,1 | 16 122,2 | 6 680,4 | 2 713,2 | 5 814,3 | 2 822,8 | 26 242,0 | 7 501,1 | 2 031,8 |
| SD | 12,4 | 14,9 | 13,8 | 5,4 | 6,4 | 3,8 | 15,0 | 11,5 | 3,4 |

Les champs monétaires n'ont pas de décimal. Nous avons divisé le CA par 100.

| PMG 12 mois | Effectif | % de clients | Chiffre d'affaire filtré sur 12 mois | % de chiffre d'affaire filtré sur 12 mois | Chiffre d'affaire filtré sur 12 mois : Moyenne par client |
|---|---|---|---|---|---|
| NA | 7 816 | 0,15% | 0 € | 0,00% | 0,0 |
| New | 245 278 | 4,56% | 40 365 377 € | 0,54% | 164,6 |
| I | 2 573 | 0,05% | 0 € | 0,00% | 0,0 |
| S | 3 086 955 | 57,45% | 1 185 154 005 € | 15,87% | 383,9 |
| M | 1 014 777 | 18,89% | 1 772 044 857 € | 23,72% | 1 746,2 |
| L | 1 015 649 | 18,90% | 4 471 932 029 € | 59,87% | 4 403,0 |
| Total | 5 373 048 | 100,00% | 7 469 496 268 € | 100,00% | 1 390,2 |

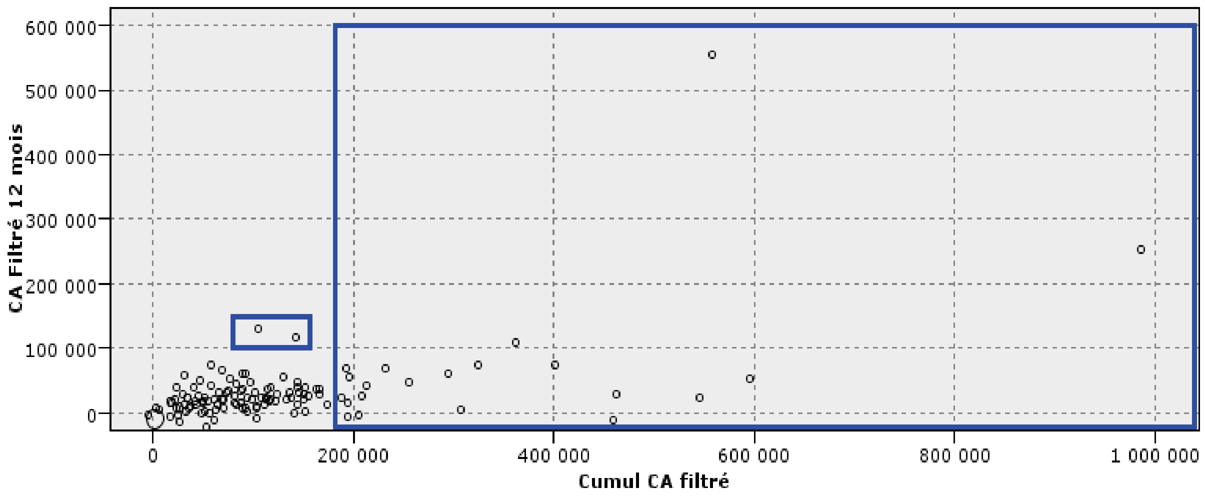
| PMG<br>12 mois | Chiffre d'affaire total | % du chiffre d'affaire total | Chiffre d'affaire total : Moyenne par client | Chiffre d'affaire filtré cumulé | % du chiffre d'affaire filtré | Chiffre d'affaire filtré : Moyenne par client |
|---|---|---|---|---|---|---|
| NA | 29 189 799 € | 0,06% | 3 734,6 € | 16 060 558 € | 0,05% | 2 054,8 € |
| New | 118 781 921 € | 0,25% | 484,3 € | 91 122 432 € | 0,31% | 371,5 € |
| I | 19 120 403 € | 0,04% | 7 431,2 € | 9 701 975 € | 0,03% | 3 770,7 € |
| S | 11 506 545 207 € | 23,89% | 3 727,5 € | 6 400 520 698 € | 21,51% | 2 073,4 € |
| M | 10 946 321 611 € | 22,72% | 10 786,9 € | 6 841 635 615 € | 22,99% | 6 742,0 € |
| L | 25 549 213 252 € | 53,04% | 25 155,6 € | 16 398 620 825 € | 55,11% | 16 146,0 € |
| Total | 48 169 172 193 € | 100,00% | 8 965,0 € | 29 757 662 104 € | 100,00% | 5 538,3 € |

| PMG 12 mois | Chiffre d'affaire annuelle sous promo | % du chiffre d'affaire annuelle sous promo | Chiffre d'affaire annuelle sous promo : Moyenne par client | Total des bons de réduction pris (BA) | %des bons de réduction pris (BA) | Total des bons de réduction pris (BA): Moyenne par client |
|---|---|---|---|---|---|---|
| NA | 173 429 € | 0,04% | 22,2 | 1 529 | 0,07% | 0,20 |
| New | 4 749 065 € | 1,02% | 19,4 | 4 156 | 0,18% | 0,02 |
| I | 16 358 € | 0,00% | 6,4 | 727 | 0,03% | 0,28 |
| S | 84 017 089 € | 18,12% | 27,2 | 235 034 | 10,11% | 0,08 |
| M | 109 248 453 € | 23,56% | 107,7 | 469 769 | 20,21% | 0,46 |
| L | 265 468 817 € | 57,25% | 261,4 | 1 613 749 | 69,41% | 1,59 |
| Total | 463 673 210 € | 100,00% | 86,3 | 2 324 964 | 100,00% | 0,43 |

| PMG 12 mois | Nb. de points transformés sur les 12 moins | % de points transformés sur les 12 moins | Nb. de points transformés sur les 12 moins : Moyenne par client | Nb. de CM transformé dans les 12 mois | % de CM transformé dans les 12 mois | Nb. de CM transformé dans les 12 mois : Moyenne par client |
|---|---|---|---|---|---|---|
| NA | 5 916 575 | 0,08% | 756,98 | 7 870,7 | 0,03% | 1,01 |
| Nouveau | 13 551 700 | 0,18% | 55,25 | 127 290,4 | 0,54% | 0,52 |
| Inconnu | 4 396 785 | 0,06% | 1 708,82 | 501,4 | 0,00% | 0,19 |
| P | 850 847 714 | 11,31% | 275,63 | 6 024 886,1 | 25,34% | 1,95 |
| M | 1 628 377 733 | 21,64% | 1 604,67 | 5 079 909,9 | 21,37% | 5,01 |
| G | 5 020 142 644 | 66,73% | 4 942,79 | 12 531 310,3 | 52,72% | 12,34 |
| Total | 7 523 233 151 | 100,00% | 1 400,18 | 23 771 769,0 | 100,00% | 4,42 |

| PMG 12 mois | Nb de point bonus utilisés | % de point bonus utilisés | Nb de point bonus utilisés : Moyenne par client |
|---|---|---|---|
| NA | 516 210 | 0,03% | 66,05 |
| Nouveau | 3 972 503 | 0,23% | 16,20 |
| Inconnu | 50 615 | 0,00% | 19,67 |
| Petit | 115 501 391 | 6,80% | 37,42 |
| Moyen | 376 564 221 | 22,18% | 371,08 |
| Gros | 1 201 136 053 | 70,75% | 1 182,63 |
| Total | 1 697 740 993 | 100,00% | 315,97 |

## La recherche des clients aberrants



Dans ce nouveau fichier, obtenu après le premier audit nous n'avons pas le nombre d'achats (corrélé avec le chiffre d'affaires) ; dans le but d'identifier les individus aberrants, nous avons utilisé les deux versions du chiffre d'affaires. Ci-dessous les ID des clients aberrants.

| ID Client |
|---|
| 2012027201917 |
| 2012030213679 |
| 2012031991910 |
| 2012033265293 |
| 012036595809 |
| 2012043773283 |
| 2012047323217 |
| 2012049821087 |
| 2012059641187 |
| 3228461599988 |
| 3228465186504 |
| 3228465989082 |
| 3547800078848 |
| 3572567264395 |
| 3594036148395 |
| 3594037801794 |
| 3597040504980 |
| 3597045375189 |
| 3597055994080 |
| 3597059399003 |
| 3600060213265 |
| 3600067130046 |