

Projet SEGBO: Restitution finale

30/03/2023

Rencontrez l'Équipe



Victor Guillard

Étudiant à Paris 1 Panthéon Sorbonne et en dernière année d'étude au master M2MO.



Abdoulaye Sakho

Étudiant en dernière année d'étude à l'ENSIIE et au M2MO.



Marta Luraschi

Étudiant Erasmus en dernière année d'étude en CLAMSES à l'Université Bicocca de Milan



Omar El-Natour

Étudiant à Paris 1 Panthéon Sorbonne et en dernière année d'étude au master M2MO.

Contexte et Problématique

Notre but avec cette segmentation est de pouvoir effectuer des actions marketing plus précises à partir des connaissances obtenues sur la base client



Objectifs

On s'intéressera tout particulièrement à l'interprétabilité de la segmentation obtenue et à la création de portraits robots qui représenteront parfaitement la clientèle



Données

Pour effectuer ce travail on s'appuiera sur la base de données fournie comprenant les informations des comportements d'achats des 900 000 clientes allemandes de la marque

Démarche de résolution

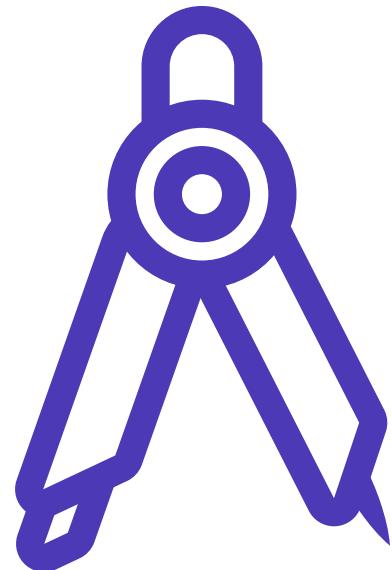
Prétraitement

Segmentation

Affectation

Performance

Prétraitement



DESCRIPTION DE LA BASE DE DONNÉES

NETTOYAGE ET RÉORGANISATION

ANALYSE DES DONNÉES MANQUANTES

GESTION DES DONNEES ABERRANTES

REGROUPEMENT PRODUITS

ANALYSE EXPLORATOIRE

Description de la base de données

Données valeurs (5) :

- Nombre d'achat
- Nombre d'article
- Panier moyen
- CA
- Prix moyen des articles achetées

Données comportementales (72 types de produits):

- Crèmes de jour anti-âge
- Après-shampoings pour cheveux
- Auto-bronzants pour le corps
- Maquillage des yeux
-

La base de données contient de plus 900 297 lignes représentant chacune une cliente.

Analyse des données manquantes

Les données contenaient à l'origine 73 739 données manquantes au total. Ces données se trouvent au niveau des colonnes "**Evasion Crèmes De Nuit**"(8 192 données), "**Evasion Produits De Traitement**"(65 536 données) et "**Prix moyen article**"(11 données manquantes).

Pour les deux types de produits évasion, les données manquantes ont été remplacées par la valeur 0 car ces colonnes sont des montants dépensés dans le type de produits associé; on peut donc supposer que les clientes concernées n'ont rien dépensé dans ces types de produits.

Le "prix moyen article" est une donnée manquante pour les clients ayant remboursé la totalité de leurs achats sur l'année. Nous avons donc remplacé les données manquantes de cette colonne par la valeur 0 qui correspond au prix moyen d'un article d'une personne n'ayant rien acheté.

Gestion des Données Aberrantes

Nous avons remarqué lors de l'audit de la base de données que certaines clientes avaient des CA et des fréquences de visites anormalement élevés.

Nous avons supposé que les clientes qui venaient plus de 100 fois dans l'année, soit presque une fois tous les trois (environ 1054 clientes) n'étaient pas des clientes standard mais sûrement des personnes faisant des achats en gros ou pour des entreprises; en effet la plupart de ces clientes ont acheté plus de 1000 articles pendant l'année.

Nous avons donc fait le choix de retirer ces 1000 clientes de notre base de données ce qui représente environ 0,1% de l'ensemble des clientes.

Regroupement produits

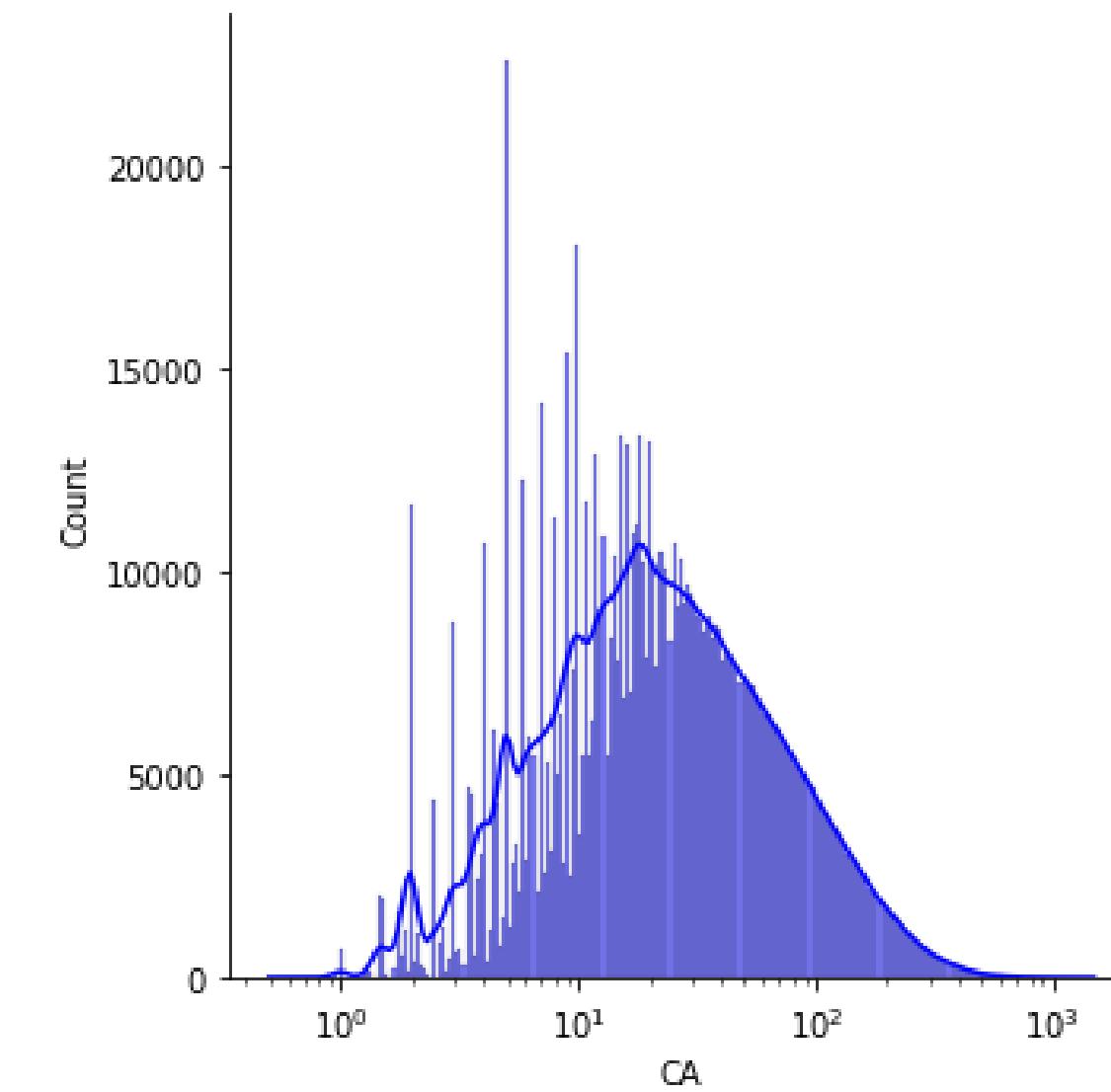
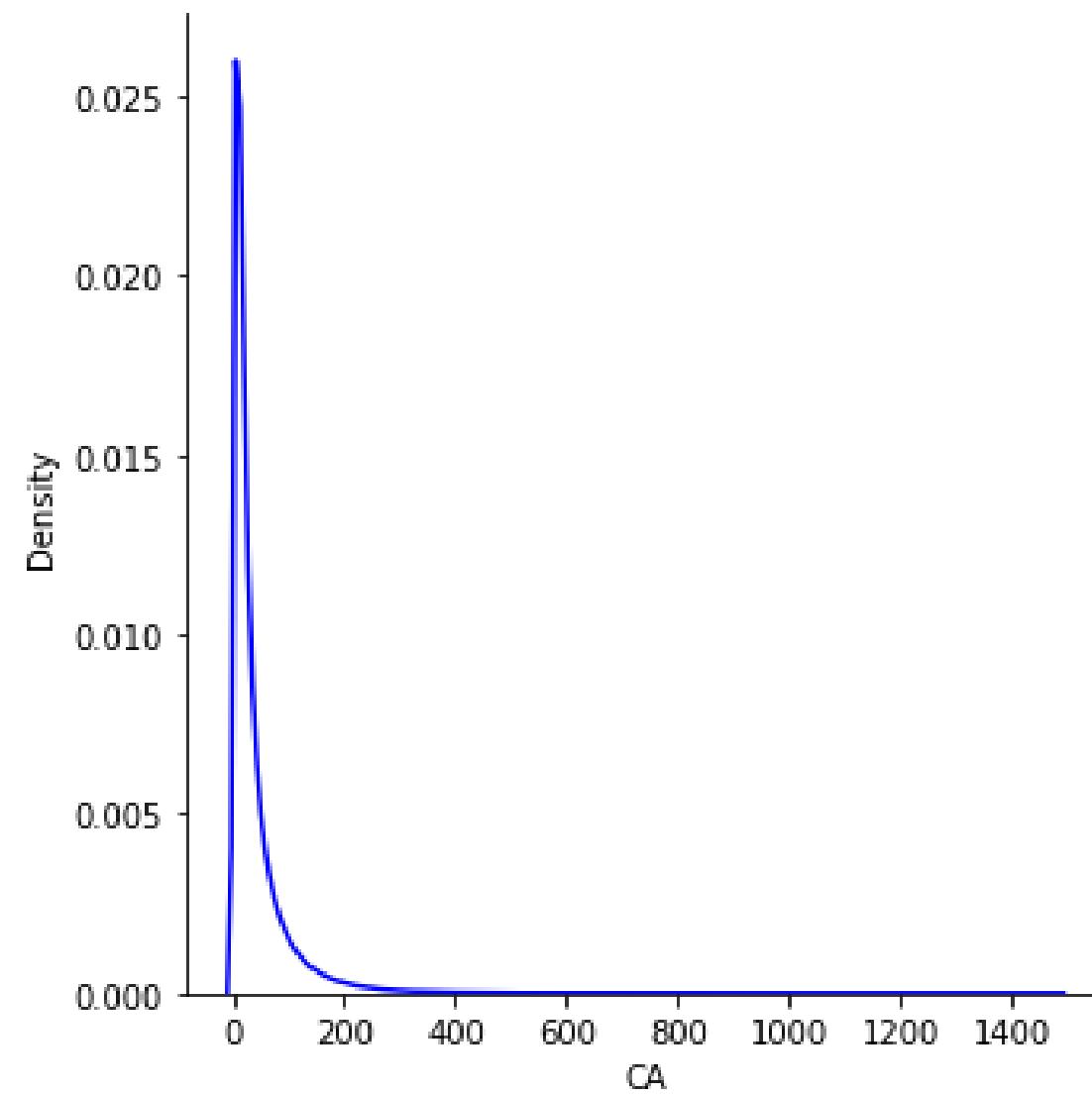
Les 72 types de produits sont classés en 9 catégories différentes:

- Anti-âge (7 types de produits)
- Cheveux (6 types de produits)
- Corps (11 types de produits)
- Evasion (18 types de produits)
- Homme (5 types de produits)
- Maquillage (5 types de produits)
- Nettoyage (7 types de produits)
- Osfa (7 types de produits)
- Parfums (6 types de produits)

Les CA des types de produits associés à une catégorie donnée ont été sommés pour former une nouvelle variable représentant la catégorie en question.

On passe donc de 72 types de produits à 9 catégories de produits, c'est le regroupement produits.

Analyse Exploratoire



Statistiques générales

On observe qu'**une cliente sur 2 n'effectue pas plus de deux achats par an et ne dépense pas plus de 21€ sur l'année.**

Les clients dépensent **en moyenne 39.7€** sur l'année pour un **panier moyen de 13.50€**.

Le prix moyen d'un article est de **5.6€**.

Résultats Préliminaires

Statistiques valeurs et comportementales de l'ensemble des clientes

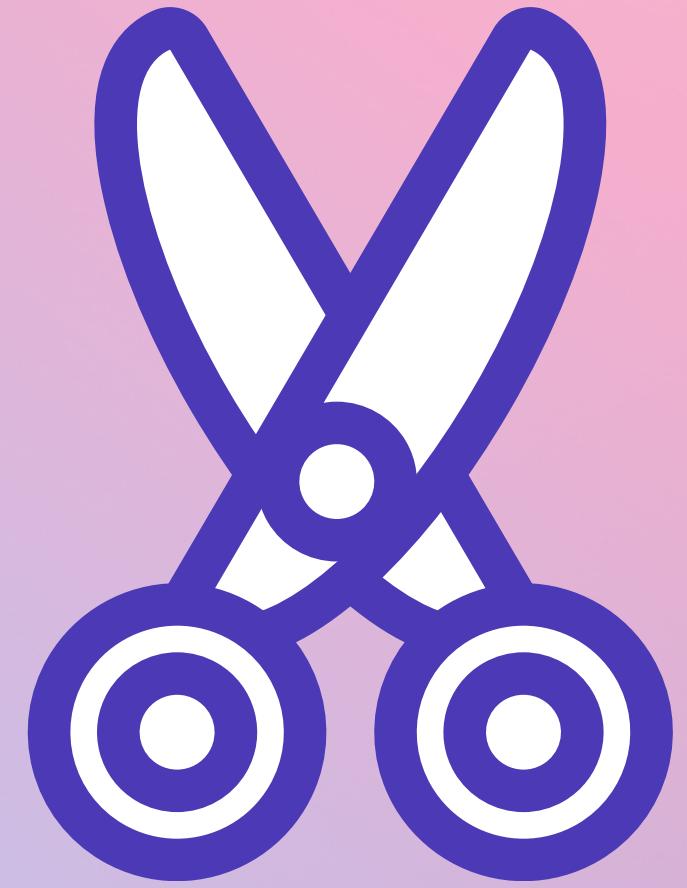
Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Tous	2.85	7.99	13.53	39.68	5.60

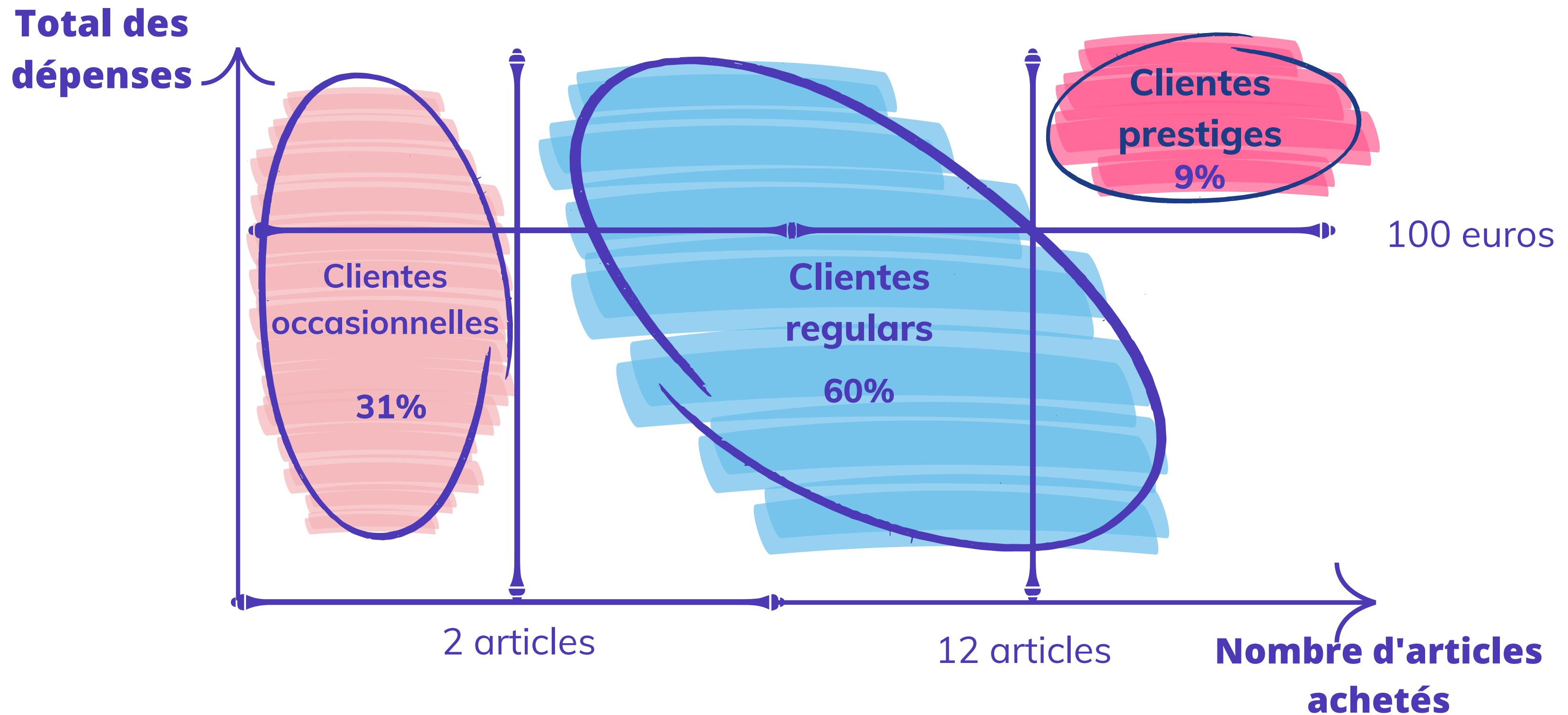
Données comportementales en % du CA par catégorie:

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Segmentation



Règle à priori



Résultats Préliminaires

Cientes prestiges

- 79 735 individus / ≈9%
 - CA moyen : 179€
 - Segment homogène
-

Cientes regulars

- 548 246 individus / ≈60%
 - CA moyen: 35€
 - Segment hétérogène
-

Cientes occasionnelles

- 272 316 individus / ≈31%
- CA moyen : 10€
- Pas plus de 2€ dépensés en moyenne par catégorie produit.

Résultats Préliminaires

Statistiques valeurs et comportementales des clientes regulars

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Regulars	2.89	7.94	14.29	35.28	4.84

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Reguliers	15.4	4.4	8.4	19.0	3.7	16.8	13.0	3.8	15.5

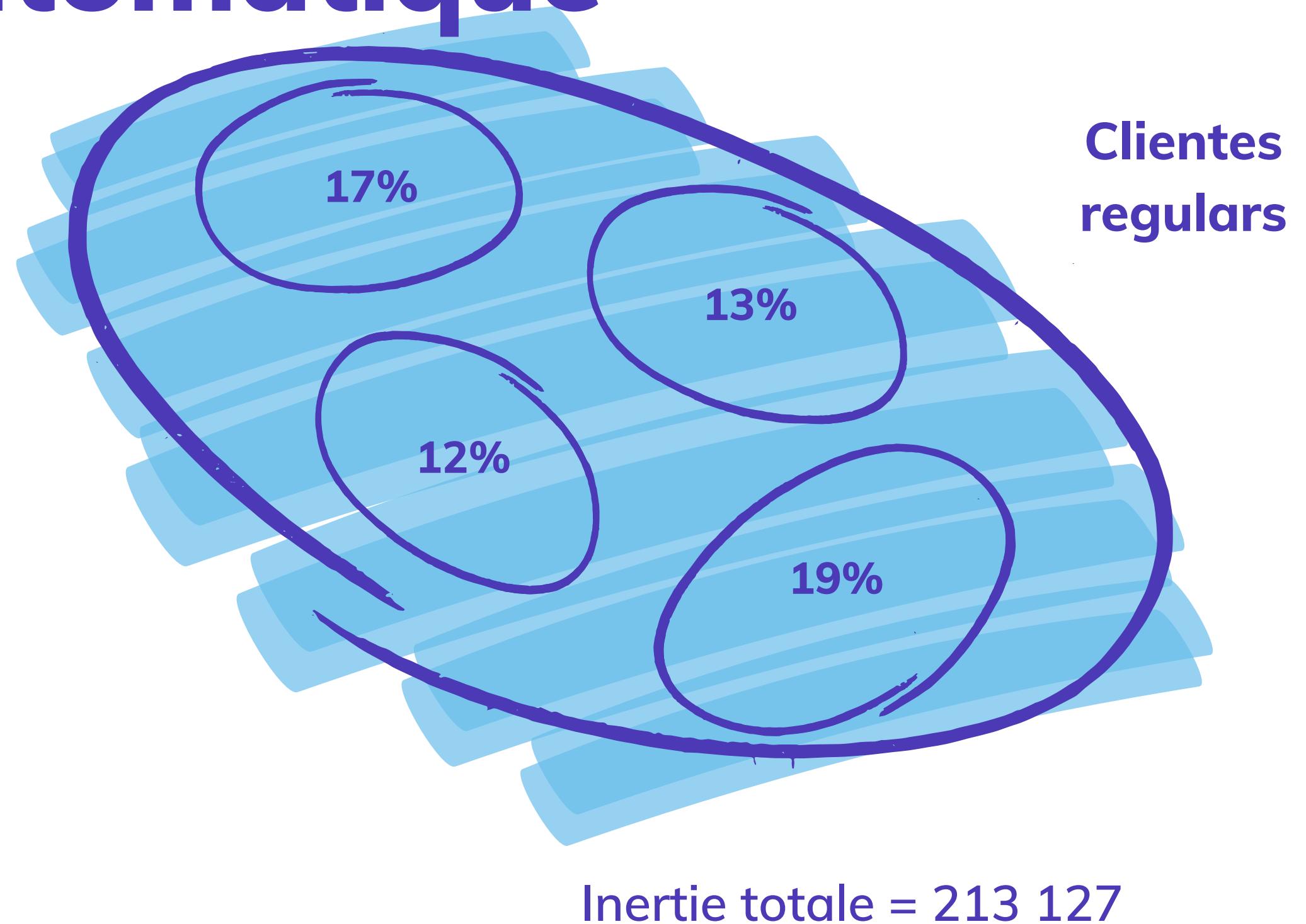
Procédure d'apprentissage automatique

On effectue une procédure d'apprentissage automatique sur le groupe des clientes regulars en utilisant uniquement les données comportementales/produits.

Chaque cercle intérieur représente les groupes obtenus avec l'algorithme des K-means.

Paramètres:

- K=4
- Distance euclidienne
- 10 initialisations
- Initialisation avec kmeans++



Descriptif des personas



La Cliente Prestige

- Effectif : 79 735 ≈ 9 %
- CA moyen : 180 euros
- Clientes dépensant principalement dans les catégories de produits suivantes (dans l'ordre décroissant) : Anti-âge, Evasion, Maquillage, Nettoyage et enfin Parfum



Une cliente Prestige est généralement une personne qui est très attentive à son apparence et qui investit du temps et de l'argent dans ses soins de beauté. Cette personne a des attentes élevées en termes de qualité des produits qu'elle utilise et s'attend à des résultats visibles et durables.

En tant que cliente Prestige, cette personne s'attend à un traitement personnalisé et attentif de la part de la marque, y compris des recommandations personnalisées de produits et des offres exclusives. La cliente Prestige est prête à investir dans des produits haut de gamme qui répondent à ses besoins spécifiques et offre des expériences de beauté exceptionnelles.

La Cliente Prestige

Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Prestige	8.62	30.85	23.71	173.58	6.22
Tous	2.85	7.99	13.53	39.68	5.60

Les clientes prestiges ont un **CA moyen 4 fois supérieur** à celui de toutes les clientes.

La Cliente Prestige

Données comportementales en % du CA par catégorie:

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Prestige	31.7	3.5	8.2	11.7	4.2	13.3	11.0	2.33	14.04
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Les clientes prestiges dépensent principalement dans la catégorie **Anti-âge**.

Cette catégorie représente près d'un tiers de leurs dépenses, **soit 10 points de plus** que le pourcentage d'argent dépensé par toutes les clientes dans cette catégorie de produits.

L'Expert en Soins de la Peau

- Effectif: 110 419 ≈ 12%
- CA moyen : 35 euros
- Clientes dépensant principalement dans les catégories de produits suivantes (dans l'ordre décroissant) : Nettoyage



Ce type de cliente est quelqu'un qui valorise la propreté et l'hygiène de sa peau et est prêt à dépenser de l'argent pour des produits de nettoyage et de soins de qualité. Cette cliente est à la recherche de produits efficaces pour répondre à ses besoins spécifiques en matière de nettoyage et de soins de la peau, et peut être sensible aux ingrédients naturels et non-toxiques. Elle est probablement plus enclin à acheter des marques haut de gamme pour son maquillage, car la qualité des ingrédients et la performance du produit sont importantes. En général, cette client a des attentes élevées en matière de qualité des produits de beauté et est disposé à dépenser de l'argent sur des produits qui répondent à ses besoins en matière de nettoyage et de soins de la peau.

L'Expert en Soins de la Peau

Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Expert Soins de la Peau	3.25	8.60	12.67	35.15	4.26
Regulars	2.89	7.94	14.29	35.28	4.84
Tous	2.85	7.99	13.53	39.68	5.60

Sur les données **valeurs**, les expertes en soins de la beauté ont des **valeurs moyennes similaires** à celles de tous les regulars.

L'Expert en Soins de la Peau

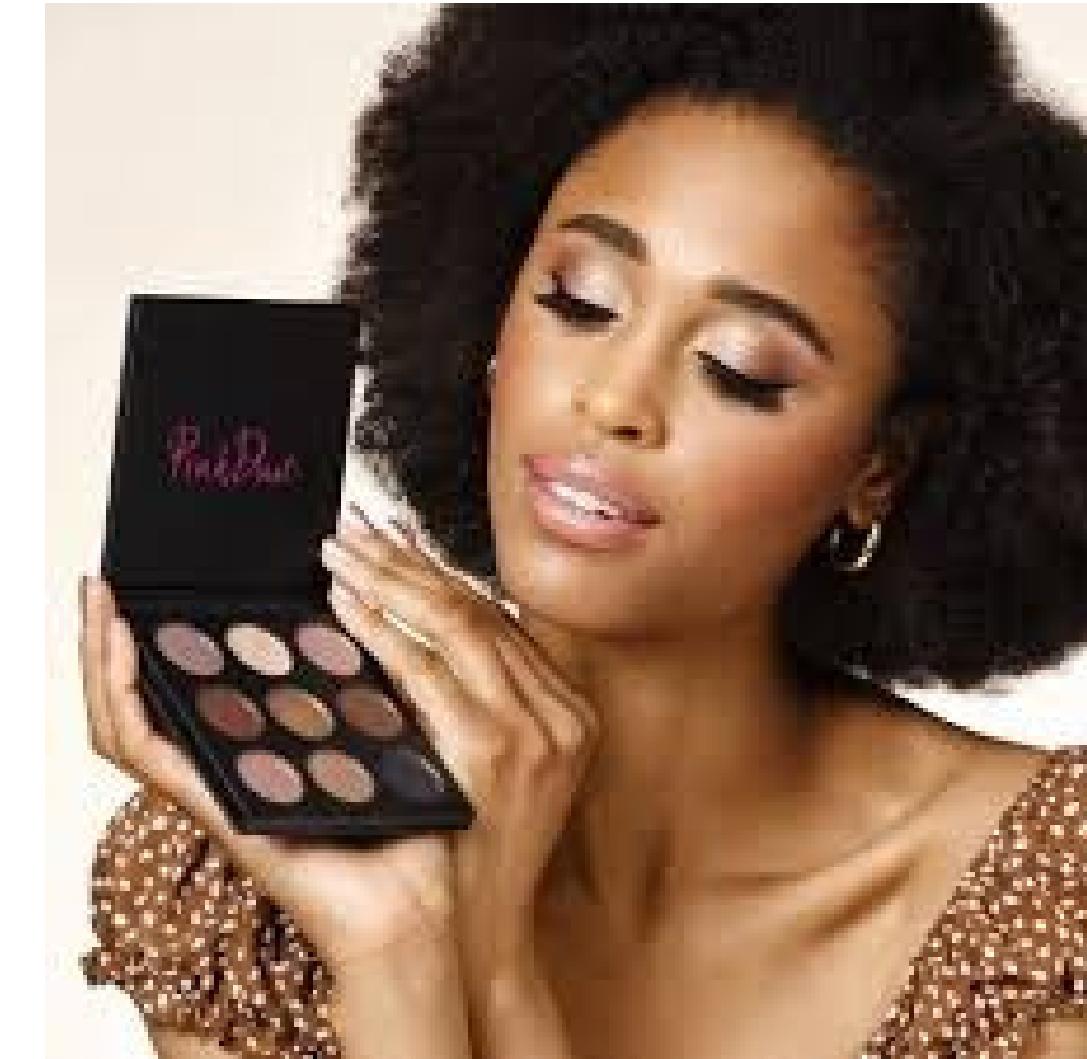
Données comportementales en % du CA par catégorie:

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Expert Soins de la Peau	7.7	4.0	7.7	11.2	2.5	8.6	47.7	2.2	8.4
Reguliers	15.4	4.4	8.4	19.0	3.7	16.8	13.0	3.8	15.5
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Sur les données **comportementales**, on observe que ces clientes dépensent **près de la moitié (47.7%) de leur CA** en produit de la catégorie **nettoyage**.

La Passionnée du Maquillage

- Effectif: 117 652 ≈ 13%
- CA moyen : 37 euros
- Clientes dépensant principalement dans les catégories de produits suivantes (dans l'ordre décroissant) : Maquillage



Ce profil de cliente est une jeune adulte qui est soucieuse de son apparence et achète régulièrement des produits de maquillage de la marque. Bien qu'elle utilise des produits de nettoyage moins chers pour retirer son maquillage, elle est prête à investir dans des produits de qualité pour obtenir l'effet désiré. Elle suit les tendances et est prête à essayer de nouveaux produits de la marque pour voir s'ils correspondent à ses besoins. Cette cliente cherche des produits qui lui permettent d'obtenir un look naturel et impeccable tout en étant sensible aux prix.

La Passionnée du Maquillage

Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Passionnée Maquillage	2.94	7.52	15.09	37.25	5.32
Reguliers	2.89	7.94	14.29	35.28	4.84
Tous	2.85	7.99	13.53	39.68	5.60

Sur les données valeurs, les passionnées du maquillage ont des **valeurs moyennes similaires** à celles de tous les regulars.

La Passionnée du Maquillage

Données comportementales en % du CA par catégorie:

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Passionnée Maquillage	7.2	2.7	5.7	9.0	2.1	55.4	6.2	1.8	9.6
Reguliers	15.4	4.4	8.4	19.0	3.7	16.8	13.0	3.8	15.5
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Sur les données **comportementales**, on observe que ces clientes dépensent **plus de la moitié de leur CA** en produit de la catégorie **maquillage**.

L'Expert Élégance

- Effectif: 169 980 ≈ 19%
- CA moyen : 41 euros
- Clientes dépensant principalement dans les catégories de produits suivantes (dans l'ordre décroissant) : Parfumerie, Anti-Age et Corps



Il s'agit d'une cliente féminine plus âgée, ayant une situation financière confortable et qui accorde une grande importance à la beauté et à la santé de sa peau. Elle cherche avant tout des produits de qualité pour lutter contre les signes de l'âge et préserver sa jeunesse. Elle est également susceptible d'acheter des produits pour hommes, que ce soit pour son compagnon ou son fils, et accorde une grande importance aux soins masculins, tout en appréciant aussi des parfums raffinés. Elle apprécie les marques qui offrent des produits efficaces et abordables, tout en offrant une expérience de soin agréable.

L'Expertise Élégance

Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Expertise Élégance	2.81	7.38	17.13	41.20	6.19
Reguliers	2.89	7.94	14.29	35.28	4.84
Tous	2.85	7.99	13.53	39.68	5.60

Sur les données valeurs, les expertes élégance ont un CA et un panier moyen quelque peu supérieurs à celui des régulars et de l'ensemble des clientes.

L'Expertise Élégance

Données comportementales en % du CA par catégorie:

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Expertise Élégance	31.6	5.8	10.9	6.6	5.7	3.9	3.0	6.3	26.22
Reguliers	15.4	4.4	8.4	19.0	3.7	16.8	13.0	3.8	15.5
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Nous observons sur les données **comportementales** que les clientes de ce segment ont des tendances d'achat plus **diversifiées** que les autres segments regulars, avec notamment plus de dépenses dans les catégories anti-âge, hommes et parfums.

Les Voyageuses de la Beauté

- Effectif: 150 275 ≈ 17%
- CA moyen : 27 euros
- Clientes dépensant principalement dans les catégories de produits suivantes (dans l'ordre décroissant) : Evasion, Parfumerie



Les clientes qui achètent des produits d'évasion sont souvent des personnes à la recherche d'une expérience sensorielle unique. Ces produits exotiques et diversifiés sont souvent offerts en cadeau, ce qui en fait une option populaire pour les occasions spéciales. Bien que ces clientes dépensent légèrement moins que la moyenne, elles sont très actives et sont souvent à la recherche de nouveaux produits à essayer. Avec leur jeunesse et leur soif d'aventure, ces clientes peuvent être une cible clé pour une marque qui souhaite proposer des produits innovants et exotiques. Les marques peuvent utiliser des descriptions attrayantes et des noms de produits évocateurs pour attirer cette clientèle et leur offrir une expérience de beauté inoubliable.

Les Voyageuses de la Beauté

Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Voyageuses	2.67	8.44	11.64	27.14	3.35
Reguliers	2.89	7.94	14.29	35.28	4.84
Tous	2.85	7.99	13.53	39.68	5.60

Sur les données valeurs, les voyageuses de la beauté ont un CA et un panier moyen inférieurs à celui des régulars et de l'ensemble des clientes.

Les Voyageuses de la Beauté

Données comportementales en % du CA par catégorie:

(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Voyageuses	3.7	4.0	7.7	58.2	3.2	5.2	4.7	3.6	9.9
Reguliers	15.4	4.4	8.4	19.0	3.7	16.8	13.0	3.8	15.5
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Sur les données **comportementales**, on observe que ces clientes dépensent **plus de la moitié de leur CA** en produit de la catégorie **evasion**.

Les Clientes Occasionnelles

- Effectif : 272 316 ≈ 30 %
- CA moyen : 10 euros



Ce type de clientèle représente une partie importante des clients d'une marque de produit de beauté milieu de gamme. Bien que peu actives, ces clientes apprécient les produits de qualité et ne sont pas nécessairement à la recherche des dernières tendances. Elles préfèrent des produits qui sont faciles à utiliser et qui conviennent à leur style de vie. Souvent fidèles à une marque qu'elles connaissent et en laquelle elles ont confiance, ces clientes recherchent avant tout des produits qui leur conviennent parfaitement, qu'elles peuvent utiliser au quotidien et qui leur apportent un sentiment de bien-être. Bien que peu enclines à dépenser beaucoup d'argent sur des produits de beauté, ces clientes apprécient le rapport qualité-prix et sont prêtes à dépenser un peu plus pour des produits qui répondent à leurs besoins spécifiques.

Les Clientes Occasionnelles

Données valeurs en euros:

(moyennes en €)	Nb achat	Nb art	Panier moyen	CA	Prix moyen des articles
Occas	1.10	1.47	9.08	9.85	6.96
Tous	2.85	7.99	13.53	39.68	5.60

Les clientes occasionnelles ont un **CA moyen 4 fois inférieur** à celui de toutes les clientes.

Les Clientes Occasionnelles

Données comportementales en % du CA par catégorie:

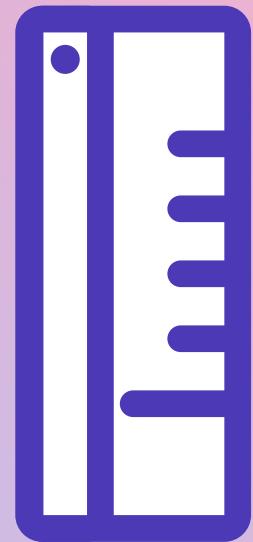
(%CA)	Aage	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Occas	14.5	2.2	5.3	14.0	3.9	20.3	9.0	1.5	29.0
Tous	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Les clientes occasionnelles dépensent principalement dans les catégories **Maquillage et Parfums**.

Nous remarquons notamment que ces clientes dépensent en moyenne **deux fois plus** en parfums que l'ensemble des clientes.

	Nb achat	Nb art	Panier moyen	CA	Prix moy art	Age	Cheveux	Corps	Evasion	Hommes	Maquillage	Nage	Osfa	Parfums
Prestige	8.6	30.9	23.7	173.6	6.2	31.7	3.5	8.2	11.7	4.2	13.3	11.0	2.3	14.0
Expertise Soins	3.2	8.6	12.7	35.2	4.3	7.7	4.0	7.7	11.2	2.5	8.6	47.7	2.2	8.4
Expertise Maqu	3.0	7.5	15.1	37.3	5.3	7.2	2.7	5.7	9.0	2.1	55.4	6.2	1.8	9.6
Expertise élégance	2.8	7.4	17.1	41.2	6.2	31.6	5.8	10.9	6.6	5.7	3.9	3.0	6.3	26.2
Voyageuse	2.7	8.4	11.6	27.1	3.3	3.7	4.0	7.7	58.2	3.2	5.2	4.7	3.6	9.9
Regulars	2.9	7.9	14.3	35.3	4.8	15.4	4.4	8.4	19.0	3.7	16.8	13.0	3.8	15.5
Tous	2.8	8.0	13.5	39.7	5.6	21.5	3.8	8.1	15.8	3.9	15.7	11.9	3.1	15.9

Affection



Présentation de la procédure

L'objectif de cette partie est de mettre en place une solution permettant d'affecter de nouveaux clientes aux 6 segments précédemment formés.

Les segments des clientes prestiges et occasionnelles ayant été engendrée par des règles à priori, il sera aisément d'affecter des clientes à ces segments.

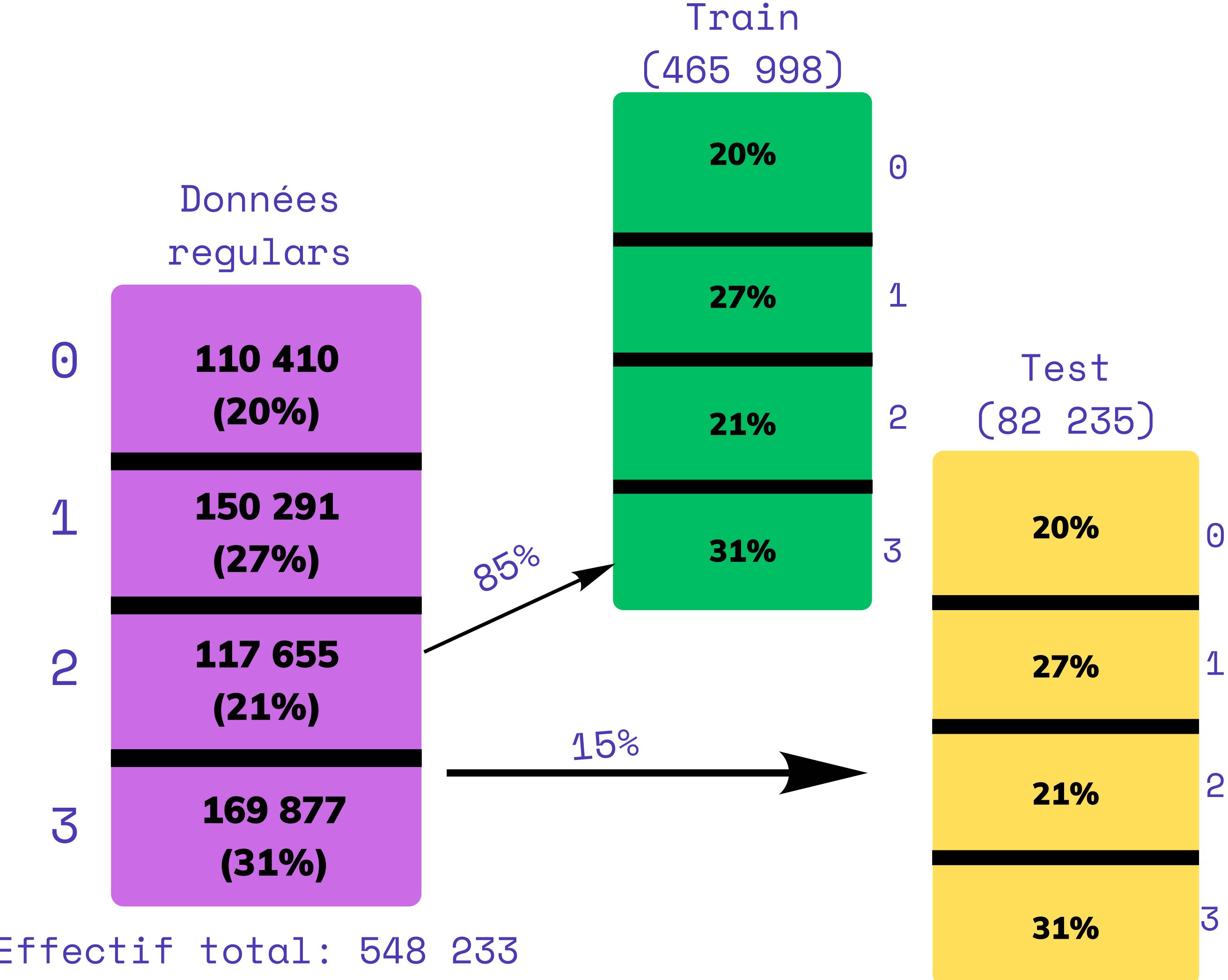
La difficulté réside chez les clients regulars. Nous avons mis en place un algorithme capable de réaffacter des clientes regulars dans les segments associés. C'est ce point que nous allons développer.

Remarque: Une fonction finale contenant les règles à priori et l'algorithme d'affectation des regulars a été implémentée afin de tout synthétiser. Nous pouvons ainsi appeler une seule fonction qui applique la règle à priori et/ou l'algorithme d'affectation.

Train/Test Split

On s'applique ici à séparer les données des regulars en deux ensembles tout en **respectant les proportions** de chaque classe.

L'un servira à entraîner notre modèle et l'autre à tester ses performances.

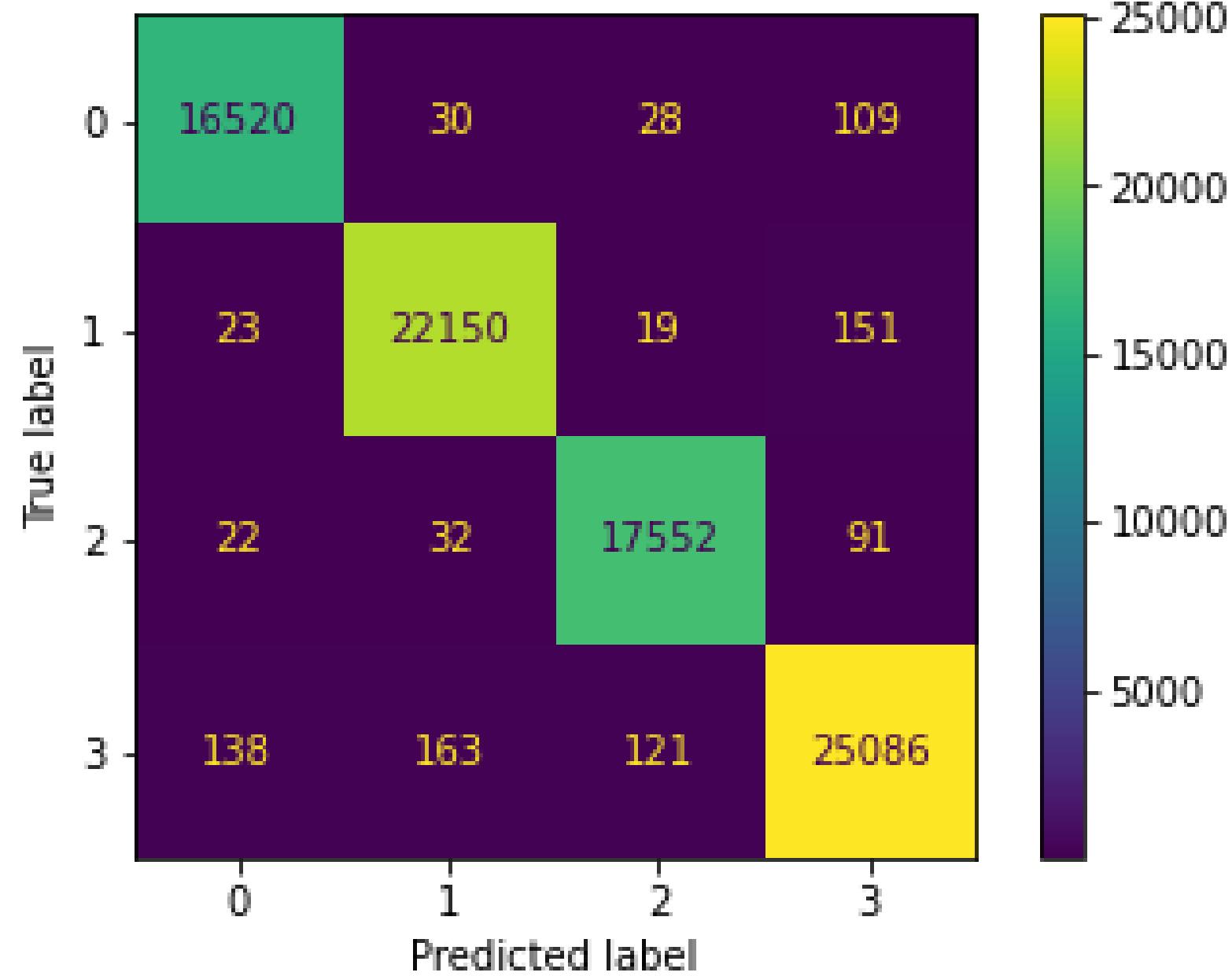


Procédure d'apprentissage automatique

C'est un algorithme de **Random Forest** qui a été retenu pour effectuer l'affection. Voici ces paramètres :

- 100 arbres
- Gini index comme mesure d'impureté à minimiser
- Arbres non élagués
- 4 variables explicatives par split d'un noeud d'arbre

Performances



L'accuracy est défini comme le rapport suivant: $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$
Accuracy obtenu: 0.98

La matrice de confusion ici présente les résultats obtenus sur les données de test avec le Random Forest. Les individus bien classées se trouvent uniquement sur la diagonale de la matrice.

On observe que la grande majorité des individus sont bien classés, en effet les individus sont principalement présents sur la diagonale de la matrice.

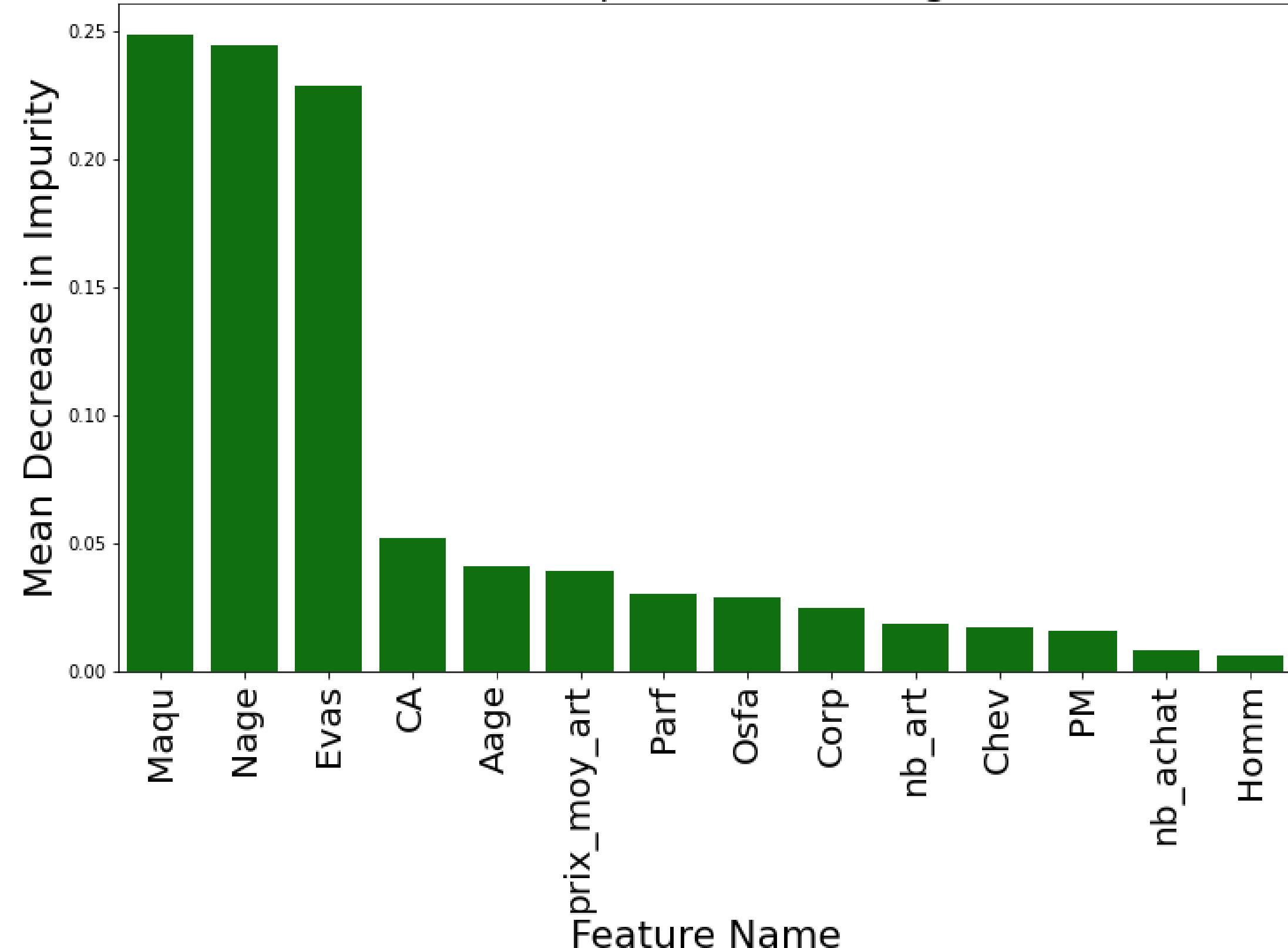
Nos très bonnes performances nous permettent de valider notre segmentation.

Résultats

On présente ici l'importance des variables **normalisées** basées sur la Mean Decrease Impurity (MDI).

Les variables les plus importantes sont **3 variables comportementales**. Cela met en évidence la pertinence de la **segmentation comportementale** effectuée avec le K-means auparavant.

Feature importances using MDI



Cette importance des variables ne nous donne pas directement d'information quant à l'influence de chaque variable dans l'affectation mais uniquement qu'elles apportent plus ou moins d'information.

Conclusion

En résumé, à partir de la base de données brute, nous avons tout d'abord procédé au prétraitement des données (nettoyage, données manquantes et aberrantes, regroupement produits et analyse exploratoire). Nous avons ensuite effectué la segmentation avec en premier une règle à priori puis une procédure d'apprentissage automatique; ceci nous a permis de définir six profils types de clientes cohérents et homogènes. Enfin nous avons implémenté un algorithme de réaffectation pour assigner les nouvelles clientes aux profils types qui nous a donné de très bons résultats, confirmant ainsi la pertinence de notre segmentation.

Pour aller plus loin

En plus de cette segmentation et de notre procédure de réaffectation, nous pourrions mettre en place un algorithme de recommandations personnalisées permettant de cibler plus finement les clientes et leurs besoins.

Contactez-nous

+ 33 01 47 12 23 07

étudiants@m2mo.fr

www.étudiants.m2mo.fr