



Data Mining

Rapport

17/11/2024

Ségmentation de clients à l'aide de techniques de clustering



Réalisé par :
LIROULET Inès
YAKOUBI Oumayma

Encadré par :
Dr. CAZABET Rémy

Table des matières

1	Exploration de Données	1
1.1	Présentation du Dataset	1
1.1.1	Description Générale des Données	1
1.1.2	Objectif	1
1.1.3	Description des Variables	1
1.2	Préparation et Nettoyage des Données	2
1.2.1	Préparation des données	2
1.2.2	Gestion des valeurs manquantes	2
1.2.3	Détection et gestion des doublons	2
1.2.4	Identification et Traitement des Valeurs Aberrantes	2
1.3	Analyse Statistique Descriptive	3
1.3.1	Distribution des variables numériques	3
1.3.2	Distribution des variables Catégorielles	6
1.3.3	Encodage des variables catégorielles	6
1.3.4	Matrice de corrélation	7
2	Clustering	9
2.1	Réduction de dimensionnalité	9
2.1.1	PCA (Principal Components Analysis)	9
2.1.2	t-SNE (t-distributed Stochastic Neighbour Embedding)	10
2.2	Dendrogramme : clustering hiérarchique	10
2.3	Clustering avec K-means	11
2.3.1	Clustering avec K-Means après PCA	11
2.3.2	Clustering avec K-Means après t-SNE	12
2.4	Clustering avec DBSCAN	13
2.5	K-means (k =3, T-SNE) : Interprétation des Clusters	14
2.6	Insights pour la ségmentation des clients	18

Table des figures

1.1	Distribution Income	3
1.2	Distribution des achats et visite du site web	3
1.3	Distribution de l'âge	4
1.4	Distribution de nombre d'enfant	4
1.5	Distribution des dépenses par catégorie	5
1.6	Distribution des dépenses en fonction du revenu et par produit	5
1.7	Distribution du niveau d'éducation	6
1.8	Corrélation entre les variables	7
2.1	Projection du données en 2D après PCA	9
2.2	Projection du données en 2D après T-SNE	10
2.3	Dendrogramme	11
2.4	Clusters K-means après PCA (k=3)	12
2.5	Clusters K-means après T-SNE (k=3)	13
2.6	Relations entre le revenu et les dépenses et entre l'âge et les dépenses selon les clusters	14
2.7	Relation entre le revenu et les dépenses en produits sucrés et entre le revenu et les dépenses en fruits	15
2.8	Relation entre le revenu et les dépenses en poissons et entre le revenu et les dépenses en viande	15
2.9	Relation entre le revenu et les dépenses en gold products et entre le revenu et les dépenses en vin	16
2.10	Moyenne des différentes variables selon les clusters	16
2.11	Revenu médian par cluster	17
2.12	Nombre d'enfants selon les clusters	18

Liste des tableaux

2.1	Silhouette scores par nombre de clusters : K-means après PCA	11
2.2	Silhouette Scores par nombre de clusters : k-means après T-SNE	13

Chapitre 1

Exploration de Données

1.1 Présentation du Dataset

1.1.1 Description Générale des Données

Le dataset utilisé pour ce projet provient de Kaggle et comprend 2240 enregistrements, chaque ligne représentant un client unique identifié par un ID. Il contient un total de 29 variables, incluant des caractéristiques sociodémographiques, des données comportementales (telles que les habitudes d'achat), ainsi que les informations sur les campagnes marketing.

Nous avons choisi de nous concentrer sur 13 caractéristiques. Ce choix vise à simplifier l'analyse tout en capturant les dimensions essentielles du comportement des clients.

1.1.2 Objectif

L'objectif principal de ce dataset est d'effectuer une segmentation des clients afin d'identifier des groupes homogènes basés sur leurs comportements d'achat. Cela permet aux entreprises de cibler plus précisément les segments les plus réceptifs à certaines offres, améliorant ainsi l'efficacité des campagnes marketing.

1.1.3 Description des Variables

Le dataset contient à la fois des variables numériques (e.g., Income) et catégorielle (e.g., Education), ainsi que des indicateurs binaires (e.g., AcceptedCmp1, Complain).

Nous avons travaillé sur les caractéristiques suivantes : Income, MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds, NumWebPurchases, NumStorePurchases, NumWebVisitsMonth, Total_Kids, Age (=Year_Birth).

Caractéristiques sociodémographiques

- Year_Birth : Année de naissance du client
- Education : Niveau d'éducation le plus élevé atteint par le client

Chapitre 1. Exploration de Données

- Income : Revenu annuel du client
- Kidhome et Teenhome : Nombre d'enfants et d'adolescents dans le ménage du client

Comportements d'achat :

- MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds : Montants dépensés sur différents types de produits
- NumWebPurchases, NumStorePurchases : Modes d'achat et fréquence des transactions
- NumWebVisitsMonth : Nombre de visites du site web par mois

1.2 Préparation et Nettoyage des Données

1.2.1 Préparation des données

Nous avons calculé l'âge des clients à partir de leur date de naissance. La variable représentant le nombre total d'enfants a été obtenue en additionnant les colonnes Kidhome et Teenhome. De plus, nous avons supprimé les colonnes non-pertinentes pour notre étude afin de nous concentrer sur les données essentielles.

1.2.2 Gestion des valeurs manquantes

La colonne "Income" contenait des valeurs manquantes. Pour les traiter, nous avons choisi de les remplacer par la médiane des valeurs existantes.

1.2.3 Détection et gestion des doublons

Après une vérification, aucun doublon n'a été détecté dans notre dataset.

1.2.4 Identification et Traitement des Valeurs Aberrantes

Des valeurs aberrantes ont été détectées dans les variables Income et Age. En ce qui concerne le revenu, une seule valeur aberrante a été identifiée, et la ligne correspondante a été supprimée. Pour l'âge, trois clients avaient un âge irréaliste (124, 125 et 131 ans), ce qui représente 0.13% des données. Ces trois enregistrements ont également été supprimés afin de maintenir la cohérence des données.

1.3 Analyse Statistique Descriptive

1.3.1 Distribution des variables numériques

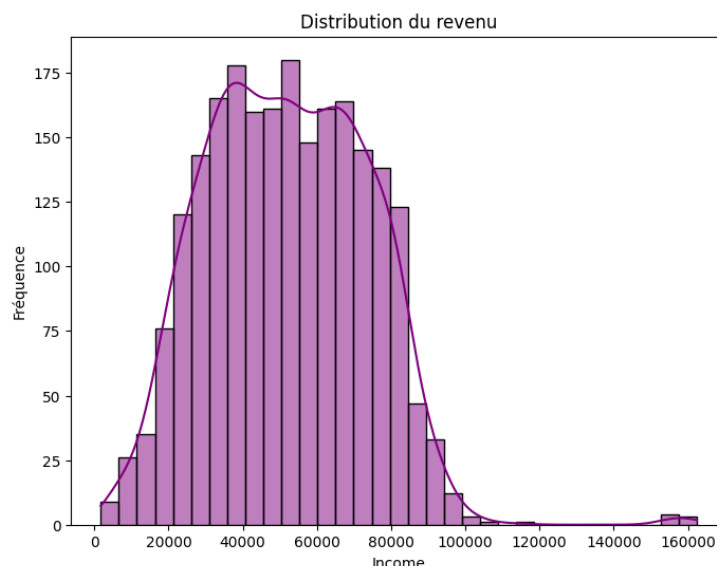


FIG. 1.1 – Distribution Income

La plot de la distribution du revenu montre une répartition légèrement asymétrique, avec une majorité de clients ayant des revenus dans une fourchette moyenne, et quelques clients aux revenus beaucoup plus élevés. Ces clients sont des valeurs extrêmes influentes, justifiant notre choix d'utiliser la médiane pour combler les valeurs manquantes, afin d'atténuer leur impact.

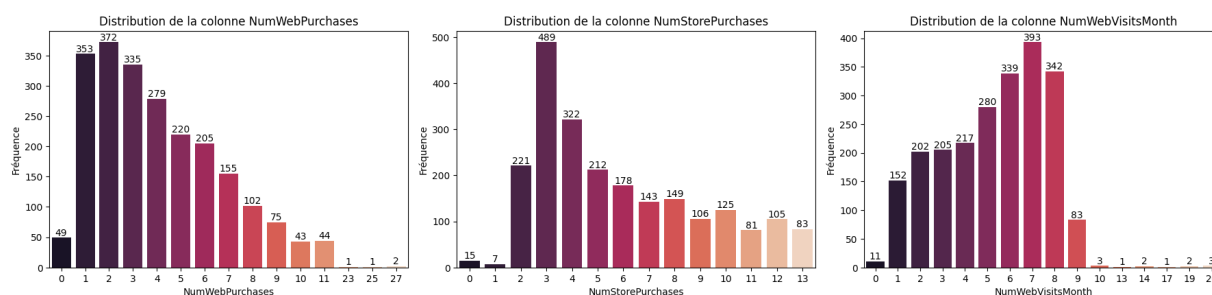


FIG. 1.2 – Distribution des achats et visite du site web

Les distributions du nombre d'achats par internet, d'achats en magasin, et de visites du site web ne suivent pas exactement de loi particulière. On peut voir que le nombre d'achats, que ce soit en web ou en magasin, est généralement peu élevé (entre 1 et 5), surtout sur le web. Le nombre de visites web, lui, a plutôt tendance à être plus élevé, généralement entre 6 et 8 visites par mois. Cette information nous permet déjà de penser à une campagne publicitaire plus ciblée sur la plateforme web, où les clients ont plus de chance de la voir.

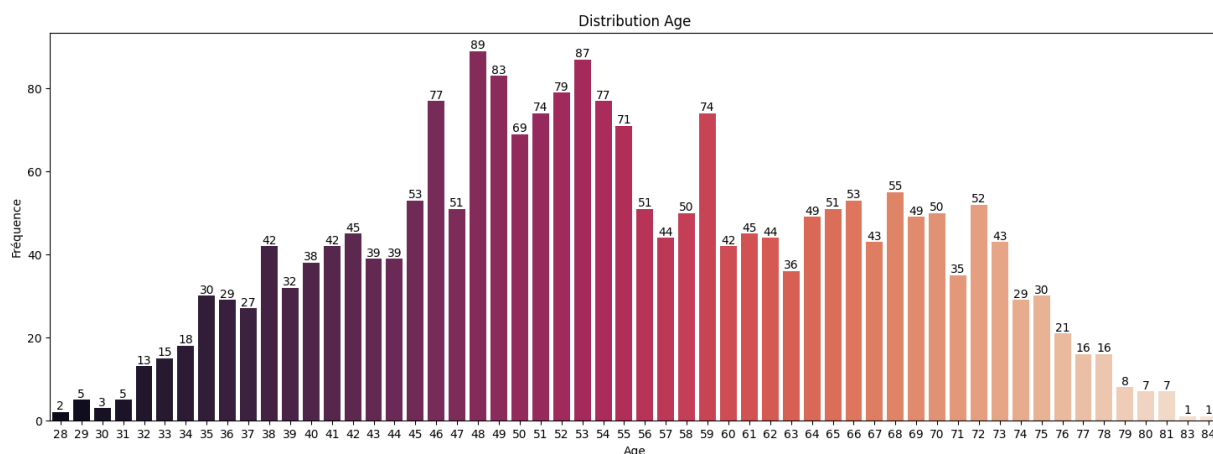


FIG. 1.3 – Distribution de l'âge

La distribution des âges ne suit pas de loi exacte, mais on peut voir que l'on retrouve surtout des personnes d'âge moyen et avancé (entre 35 et 75 ans), avec la majorité des clients ayant entre 48 et 55 ans. Aux extrémités, on retrouve une minorité de clients ayant entre 28 et 35 ans, ou entre 75 et 84 ans. Cela montre que la magasin attire déjà une certaine démographie de clientèle, et le but est de trouver une segmentation au sein de cette démographie.

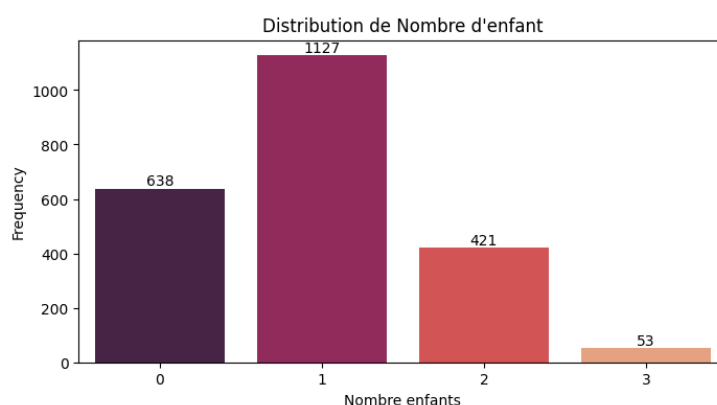


FIG. 1.4 – Distribution de nombre d'enfant

Les clients n'étant pas jeunes pour la plupart, un peu plus de 2/3 d'entre eux ont des enfants (un seul pour la plupart). Cela risque peut-être d'influencer leur comportement d'achat, ou du moins la quantité de produits achetés.

Chapitre 1. Exploration de Données

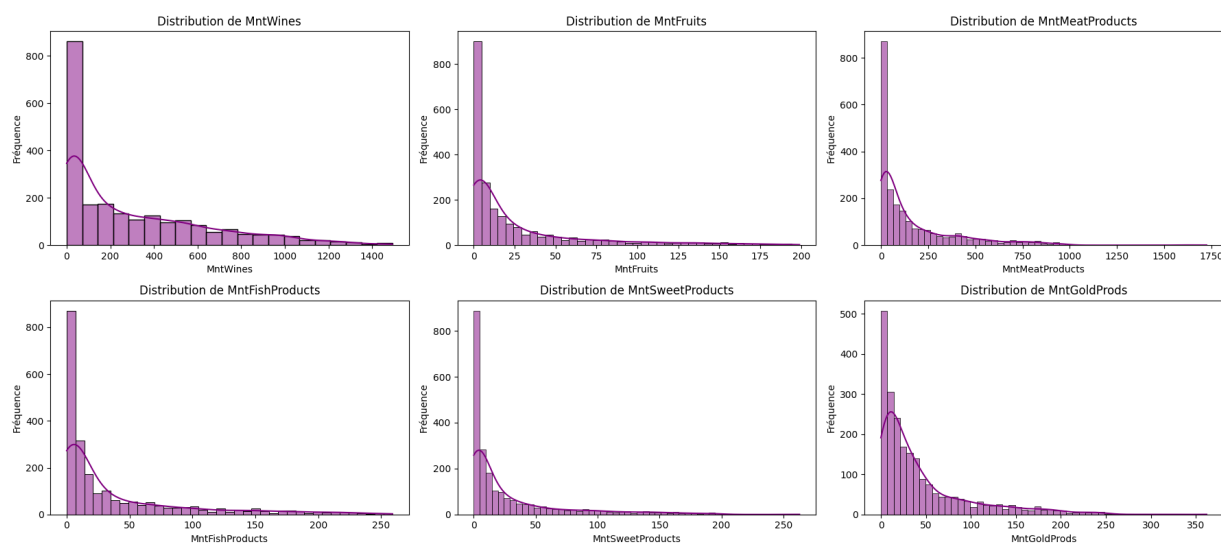


FIG. 1.5 – Distribution des dépenses par catégorie

Les distributions des montants dépensés pour chaque catégorie de produit semblent suivre une loi de puissance : la majorité des achats sont des achats d’une quantité raisonnable de produit, avec quelques clients achetant rarement des quantités très élevées. Cela est particulièrement vrai pour les produits alimentaires qui doivent être rapidement consommés comme les fruits, la viande ou le poisson. Au contraire, pour les produits n’ayant pas de limite de consommation comme le vin, cette différence est moins accentuée (même si toujours présente).

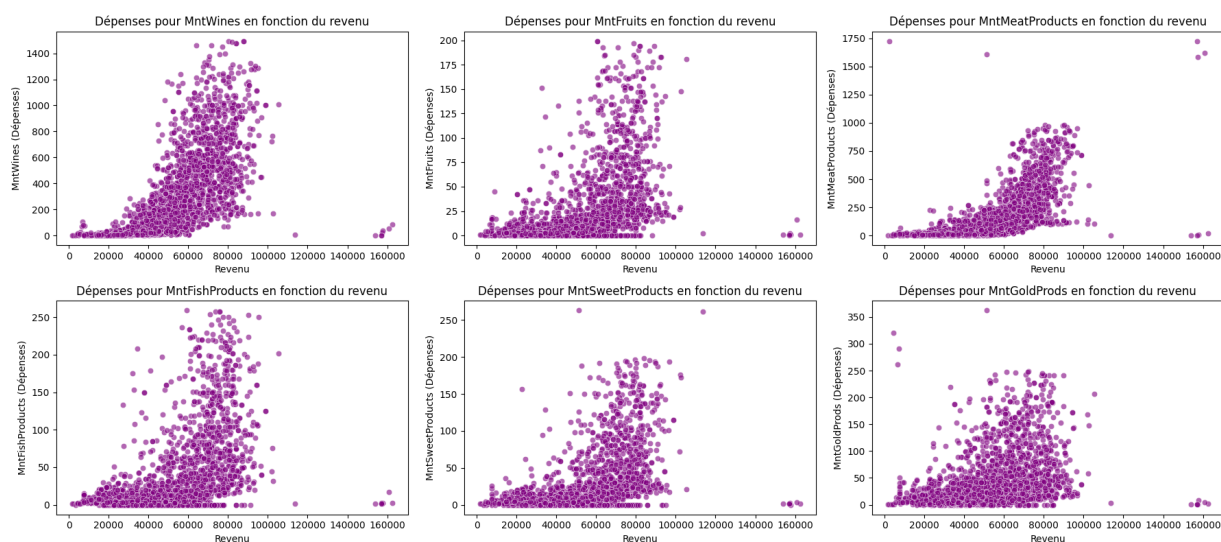


FIG. 1.6 – Distribution des dépenses en fonction du revenu et par produit

On peut voir que naturellement, les dépenses augmentent avec le revenu : les clients avec un revenu faible ne peuvent pas se permettre d’acheter une trop grande quantité de

produits, tandis que les clients aisés peuvent parfois acheter des quantités très grandes (surtout pour le vin et la viande).

Les quelques client très aisés (revenu autour de 160 000) n'achètent étrangement pas beaucoup de produits : peut-être font-ils la plupart de leurs achats dans des magasins plus luxueux. Dans ce cas, il ne faut pas leur donner trop d'importance dans les campagnes publicitaires futures.

1.3.2 Distribution des variables Catégorielles

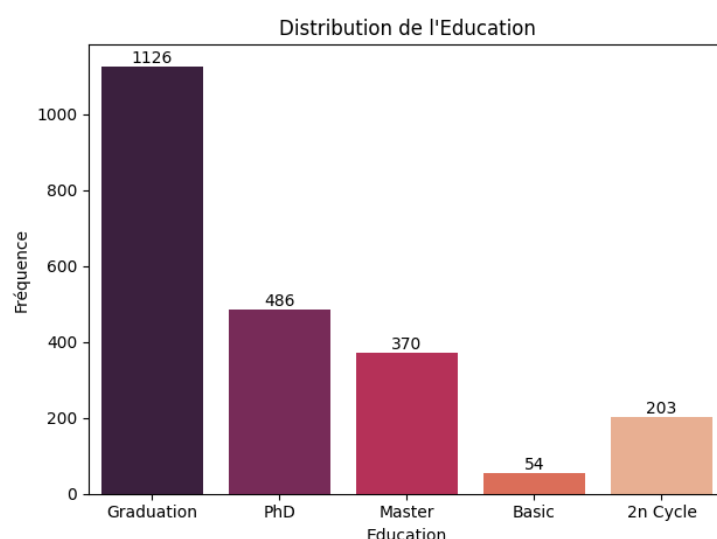


FIG. 1.7 – Distribution du niveau d'éducation

La majorité des clients ont obtenu au minimum un diplôme d'éducation supérieure. Une partie sont allés plus loin et ont obtenu un master ou équivalent (2nd cycle), ou un doctorat. Une minorité n'a pas de diplôme (éducation basique).

1.3.3 Encodage des variables catégorielles

Dans cette étape, nous avons appliqué le label encoding à la variable Education afin de la transformer en variable numérique.

1.3.4 Matrice de corrélation

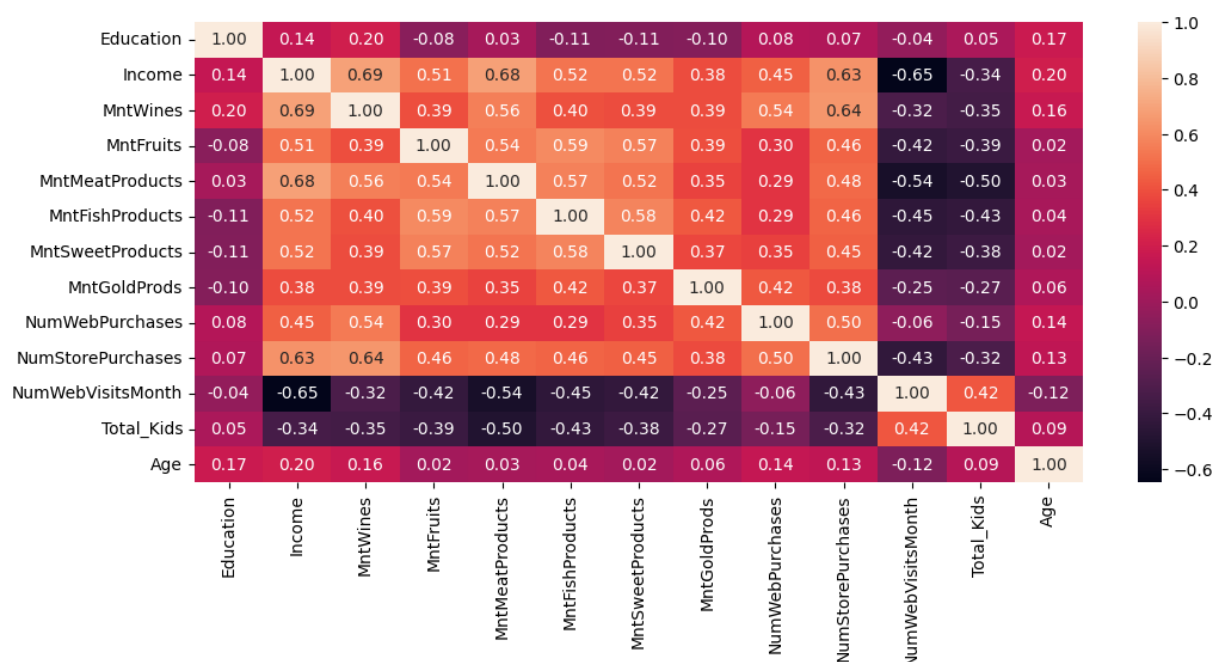


FIG. 1.8 – Corrélation entre les variables

Cette figure présente une évaluation des relations entre les variables numériques du dataset. Elle met en évidence les niveaux de corrélation, permettant de repérer des associations qui pourraient influencer les comportements d'achat et guider la segmentation.

La majorité des variables montrent des corrélations nulles ou faibles. Toutefois, quelques relations notables émergent :

- Les variables de produits présentent des corrélations positives significatives entre elles, notamment entre la variable des vins et la variable de la viande. Cela indique que les clients qui dépensent beaucoup dans une catégorie de produits tendent également à avoir des dépenses élevées dans les autres catégories, ce qui est probablement dû à un budget plus élevé.
- Le revenu est moyennement corrélé avec plusieurs variables de dépenses, notamment le vin et la viande. Cela laisse penser que, bien que le revenu influence les dépenses, d'autres facteurs influencent aussi le comportement d'achat des clients.
- Une corrélation négative assez forte est observée entre le nombre de visites web et le revenu. Il semblerait que les clients au revenu élevé n'aillent pas souvent sur le site. Pourtant, on observe une corrélation positive entre le revenu et l'achat de produits en magasin ET en ligne. Les clients aisés vont donc sur le site pour acheter, mais n'y vont qu'une fois, tandis que les clients moins aisés vont aller sur le site plusieurs fois sans forcément acheter, peut-être pour repérer des promotions intéressantes. On peut donc cibler ces clients en affichant bien les promotions sur le site.

- Enfin, on voit une corrélation moyenne entre le nombre d'enfants des clients et le nombre de visites sur internet. Une explication possible serait que tout simplement, une partie des clients étant assez âgée, ils n'ont pas le réflexe d'aller sur internet pour acheter des produits. Mais s'ils ont des enfants, ces enfants pourraient eux avoir ce réflexe, et inciter leurs parents à le faire.

Conclusion

Ainsi, l'exploration des données a pu permettre d'observer quelques comportements des clients et de supposer quelques tendances, mais sans pouvoir faire de réelles observations concrètes et une vraie ségmentation. Il ne faut donc pas se fier seulement aux chiffres observés jusqu'alors, mais implémenter une stratégie de clustering.

Chapitre 2

Clustering

2.1 Réduction de dimensionnalité

Afin de pouvoir mieux visualiser nos données et les clusters obtenus par la suite, nous devons réduire la dimensionnalité du dataset : nous allons donc passer de 13 à 2 dimensions.

2.1.1 PCA (Principal Components Analysis)

Nous avons d'abord appliqué la PCA, qui permet de passer d'un espace à 13 dimensions à un espace bidimensionnel en fonction de la variance des données : les dimensions avec peu de variance seront considérées comme porteuses de moins d'information.

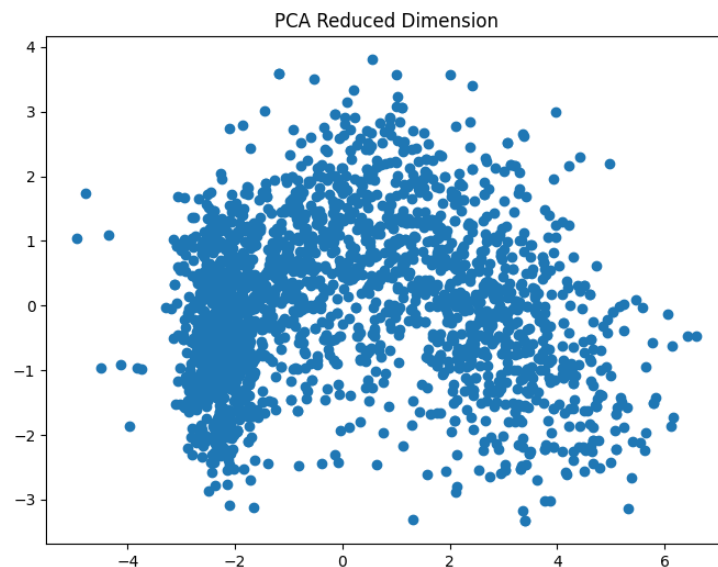


FIG. 2.1 – Projection du données en 2D après PCA

On peut voir que la distribution des données présente une concentration dans certaines zones (notamment à gauche), tandis que quelques valeurs sont un peu plus éloignées du

groupe principal. Cependant, les clusters ne sont pas clairement distinguables dans cet espace réduit.

2.1.2 t-SNE (t-distributed Stochastic Neighbour Embedding)

Nous avons également appliqué l'algorithme t-SNE sur nos données. Contrairement à la PCA, qui est une méthode linéaire, t-SNE est une technique non linéaire qui cherche à préserver la similarité entre les données. Cela permet de mieux capturer des structures complexes.

L'objectif de l'utilisation de t-SNE est de comparer la projection des données obtenue par cette méthode à celle obtenue par PCA.

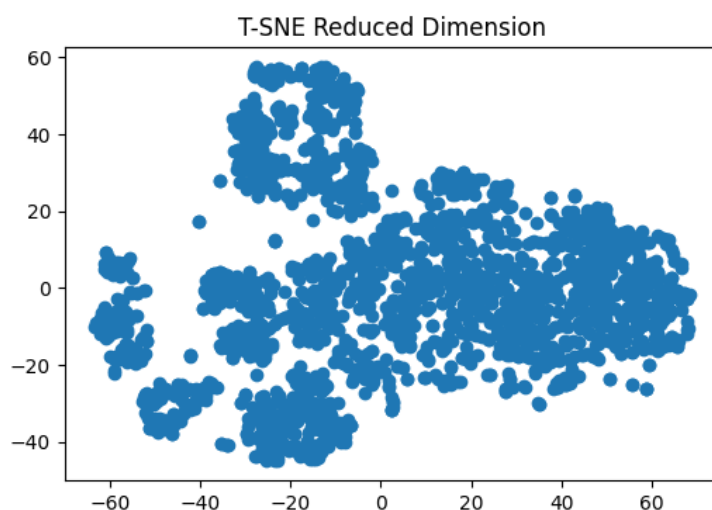


FIG. 2.2 – Projection du données en 2D après T-SNE

Comme illustré par la figure, t-SNE met en évidence des structures et des regroupements dans les données qui étaient moins perceptible avec la projection PCA. En particulier, on observe des regroupements légèrement distincts, séparés par des distances relativement importantes. Cette méthode semble donc plus appropriée pour nos données, qui de plus présentent certainement des relations non-linéaires.

2.2 Dendrogramme : clustering hiérarchique

L'objectif de la création du dendrogramme dans cette analyse était de regrouper les objets (clients) en fonction de leurs similarités, afin d'identifier des clusters naturels au sein des données.

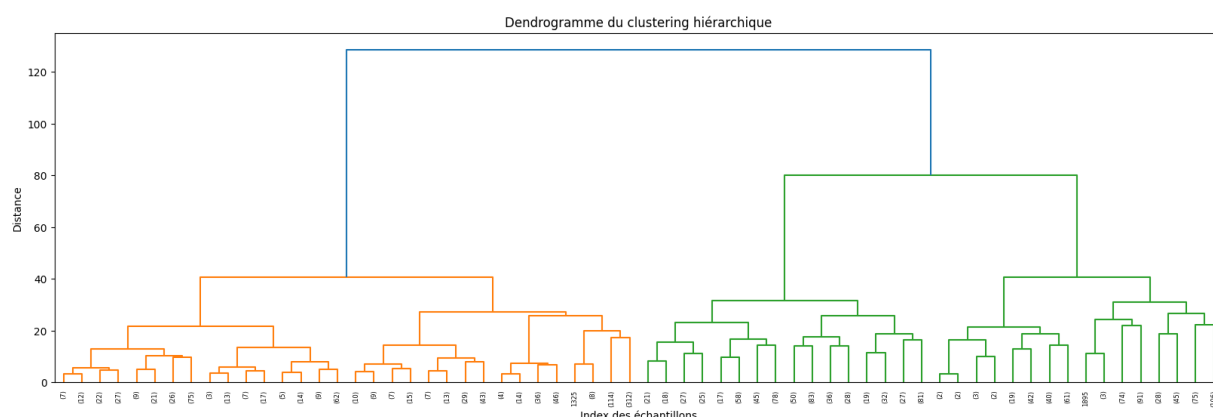


FIG. 2.3 – Dendrogramme

Ce processus de regroupement hiérarchique facilite la visualisation des relations structurelles entre les clients de manière claire et intuitive. En coupant le dendrogramme à un niveau précis, nous pouvons délimiter plusieurs groupes homogènes, où les éléments au sein de chaque groupe représentent des caractéristique, tels que des groupes de clients partageant des comportements d'achat similaires. Par ailleurs, l'examen des branches plus longues indique la présence d'éléments distincts, suggérant qu'ils diffèrent significativement des autres groupes.

2.3 Clustering avec K-means

2.3.1 Clustering avec K-Means après PCA

Choix du Paramètre K

Nous avons testé plusieurs valeurs pour le nombre de clusters allant de 2 à 6, afin de déterminer la partition la plus pertinente de données. Pour chaque nombre de clusters, nous avons calculé le score de silhouette. Ce dernier permet de juger de la qualité du clustering : un score proche de 1 indique des clusters bien séparés et cohérents.

Nombre de Clusters	Average Silhouette Score
2	0.5344
3	0.4668
4	0.3885
5	0.3597
6	0.3732

TAB. 2.1 – Silhouette scores par nombre de clusters : K-means après PCA

Après l'analyse des résultats, nous avons constaté que le nombre optimal de clusters

basé sur la maximisation du score de silhouette est de 2. Cependant, étant donné le contexte de segmentation des clients, nous avons opté pour 3 clusters. Cette segmentation permet de mieux capturer la diversité des comportements des clients tout en assurant une distinction claire entre les groupes. En choisissant 3 clusters, nous pouvons identifier des segments plus spécifiques qui permettront de cibler des stratégies marketing adaptées à chaque groupe.

Résultats

Les résultats du clustering K-means avec $k=3$, illustrés dans la figure ci-dessous, révèlent qu'un des clusters est plus condensé par rapport aux autres. Les clusters sont adjacents, sans séparation nette entre eux, et présentent des tailles globalement similaires, bien que le premier cluster se distingue par sa densité plus élevée.

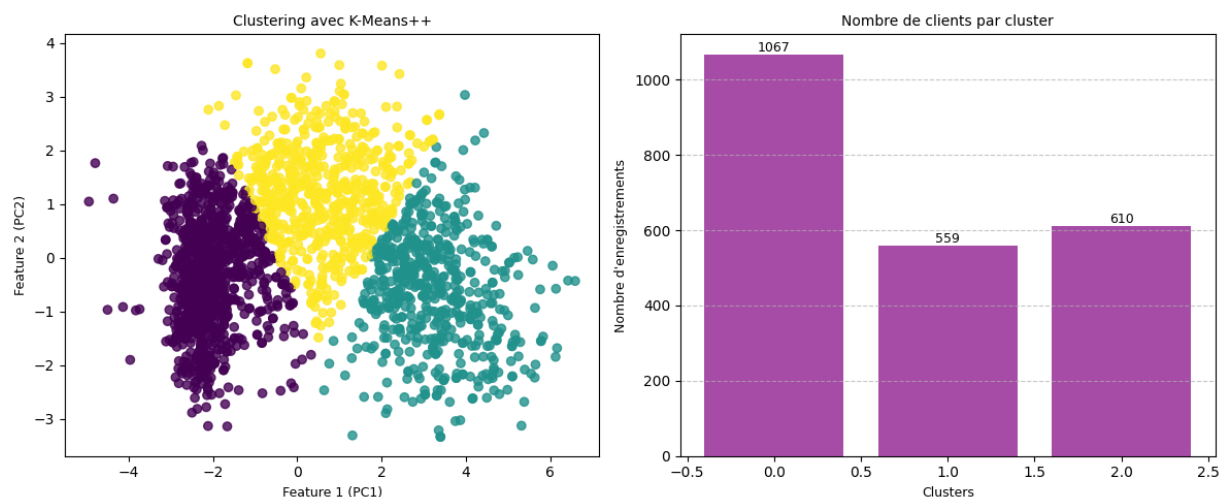


FIG. 2.4 – Clusters K-means après PCA ($k=3$)

2.3.2 Clustering avec K-Means après t-SNE

Choix du Paramètre K

Nous avons testé plusieurs valeurs pour le nombre de clusters afin de déterminer la partition la plus pertinente des données après réduction de dimension par t-SNE. Comme pour l'analyse précédente, nous avons calculé le score de silhouette pour chaque nombre de clusters.

Nombre de Clusters	Average Silhouette Score
2	0.4214
3	0.4881
4	0.4470
5	0.4497
6	0.4283

TAB. 2.2 – Silhouette Scores par nombre de clusters : k-means après T-SNE

En nous basant sur le score de silhouette, nous avons constaté que le nombre optimal de clusters basé sur la maximisation du score de silhouette est de 3.

Résultats

Les résultats du clustering K-means avec $k=3$, appliqué après la réduction de dimension par t-SNE, sont présentés dans la figure ci-dessous.

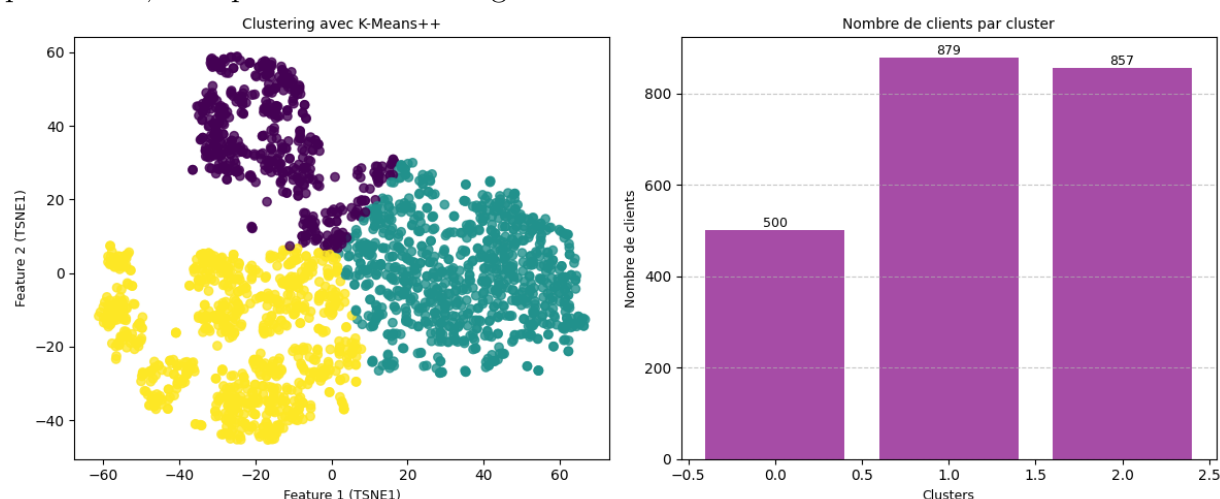


FIG. 2.5 – Clusters K-means après T-SNE ($k=3$)

Les clusters ont des tailles relativement similaires. Visuellement, les clusters sont plus séparés que précédemment, bien que la séparation ne soit pas parfaitement nette. Cela semble confirmer que la réduction par t-SNE est plus adaptée que par PCA : on obtient des données plus cohérentes visuellement, avec des clusters mieux séparés, sans compter que le k optimal est ici à 3 et non à 2 comme pour la PCA (ce qui est moins cohérent par rapport aux données).

2.4 Clustering avec DBSCAN

Nous avons également appliqué l'algorithme DBSCAN pour effectuer un clustering sur les données réduites avec PCA et avec t-SNE, en optimisant le paramètre epsilon à l'aide de

la méthode du knee point. Cependant, l'algorithme a retourné un seul cluster dans les deux cas ce qui ne correspond pas à l'objectif de notre étude, qui vise principalement à segmenter les clients en plusieurs groupes distincts pour mieux cibler les stratégies marketing.

Nous avons supposé que la réduction de dimension avait peut-être altéré la structure des données, ce qui aurait empêché DBSCAN d'identifier des clusters. Pour vérifier cette hypothèse, nous avons appliqué DBSCAN sur les données d'origine standardisées. Toutefois, le résultat est resté identique : l'algorithme n'a pas pu segmenter les données en plusieurs clusters.

2.5 K-means ($k=3$, T-SNE) : Interprétation des Clusters

Afin de comprendre à quels genres de clients chaque cluster correspond, nous pouvons revisualiser certaines des relations que nous avons examiné au début, en ajoutant cette fois-ci l'information des clusters trouvés.

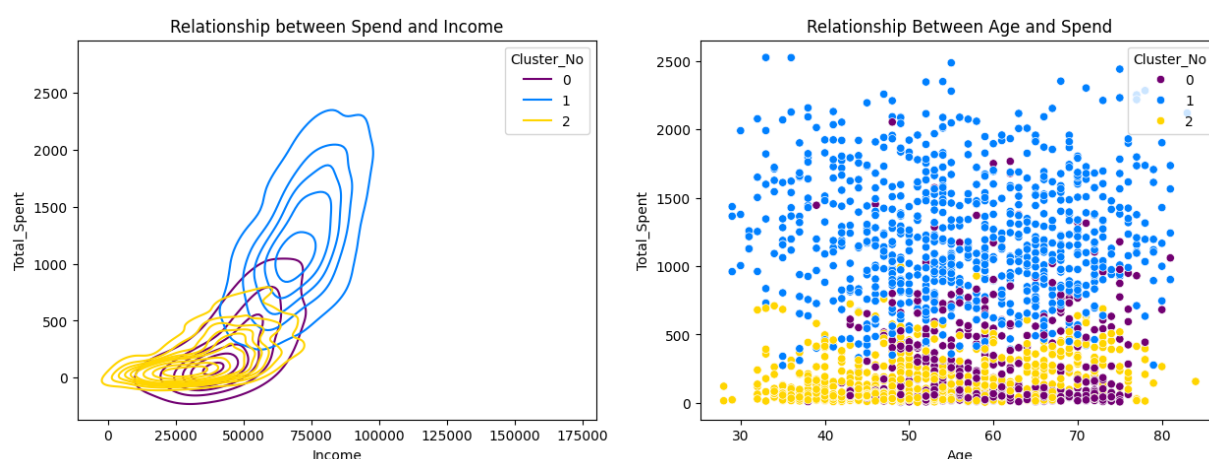


FIG. 2.6 – Relations entre le revenu et les dépenses et entre l'âge et les dépenses selon les clusters

On peut déjà voir sur la figure de gauche ci-dessus que le cluster 1 (cluster bleu) correspond à une partie de la clientèle qui a un plus grand revenu et qui dépense significativement plus que les deux autres. On peut voir grâce à la figure de droite que l'âge ne rentre pas en compte dans cette distinction, le cluster bleu comprend des clients de tout âge.

Les clusters 0 et 2 (violet et jaune), eux, correspondent à une partie de la population avec un revenu moins élevé, et donc qui dépense moins. On peut cependant voir que le cluster violet comprend des personnes à revenu légèrement plus élevé et légèrement plus dépensières que le cluster jaune. Les clients appartenant au cluster violet sont aussi généralement plus âgés, là où le cluster jaune comprend des personnes de tout âge.

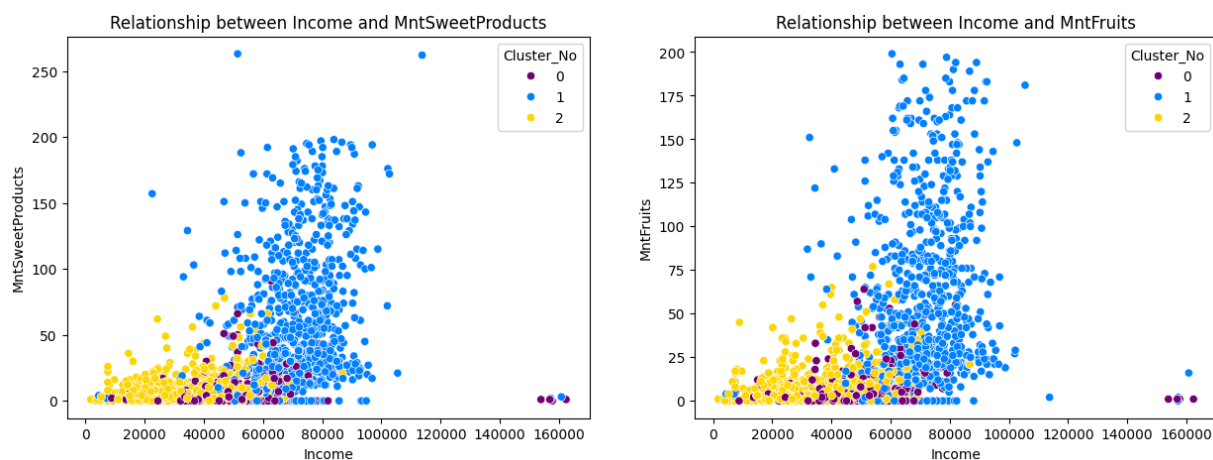


FIG. 2.7 – Relation entre le revenu et les dépenses en produits sucrés et entre le revenu et les dépenses en fruits

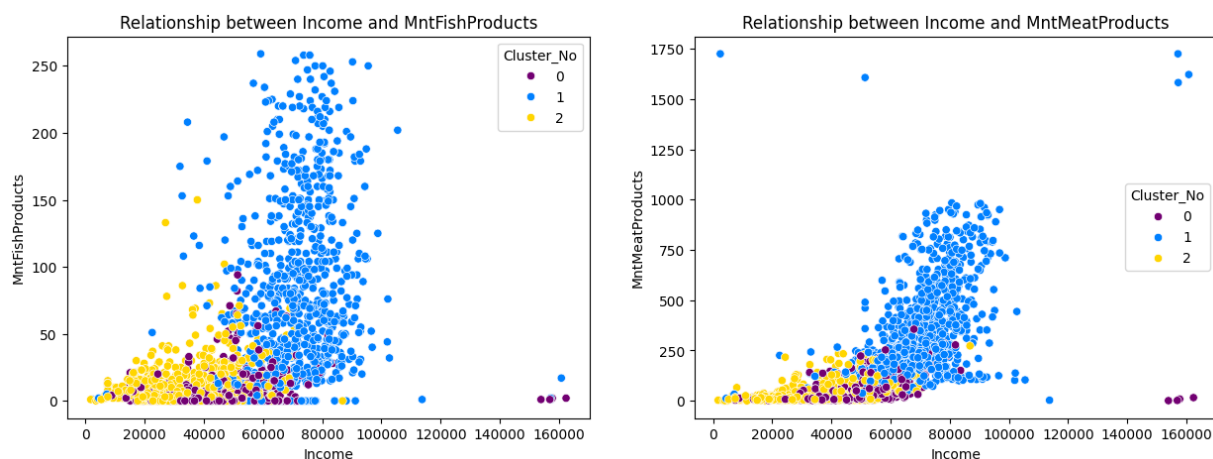


FIG. 2.8 – Relation entre le revenu et les dépenses en poissons et entre le revenu et les dépenses en viande

On peut continuer à observer cette différence de revenu entre les clusters au niveau de leurs habitudes d'achat : le cluster bleu a plus de revenus et achète donc plus, que cela soit pour les produits sucrés, les fruits, le poisson ou la viande, tandis que les clusters violet et jaune achètent moins. On peut cependant bien observer une très légère distinction entre ces deux derniers clusters : le cluster violet a en général un revenu plus élevé que le jaune, même s'il n'est pas forcément plus dépensier.

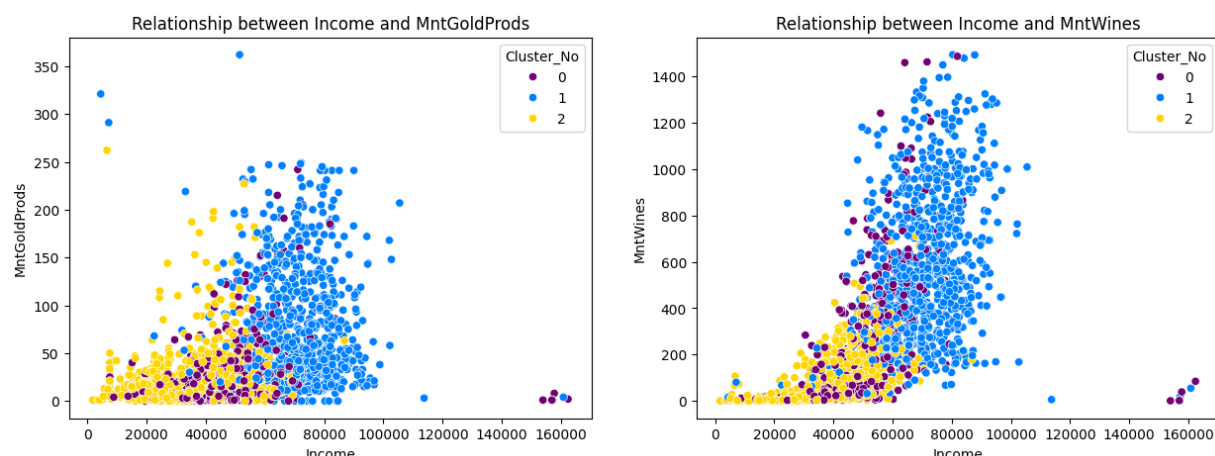


FIG. 2.9 – Relation entre le revenu et les dépenses en gold products et entre le revenu et les dépenses en vin

Pour les gold products, cette différence est moins accentuée, car ce sont des produits qui sont moins souvent achetés dans tous les cas.

Pour le vin, on retrouve l'attitude plus dépensière pour le cluster bleu, seulement il y a cette fois-ci une plus forte distinction entre les clusters violet et jaune : le cluster violet achète significativement plus de vin que le cluster jaune.

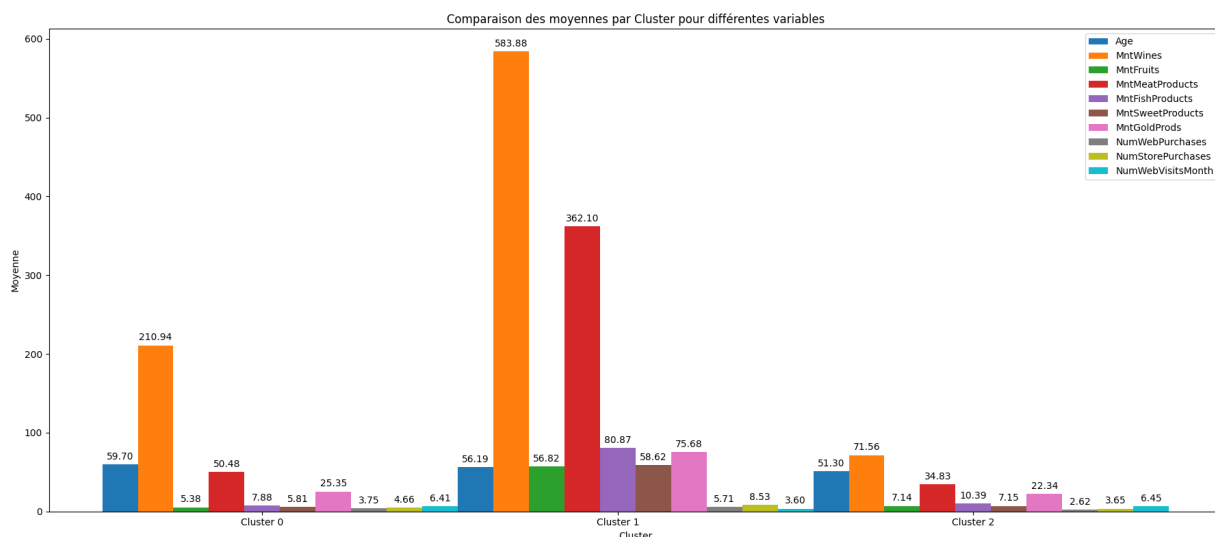


FIG. 2.10 – Moyenne des différentes variables selon les clusters

En regardant les moyennes des variables en fonction des clusters, on peut bien mieux visualiser ces différences de comportement d'achat. Dans le cluster 1 (bleu), qui contient des clients aisés et dépensiers, on peut bien voir que les dépenses sont plus élevées : ils

achètent plus de fruits, de poisson, de produits sucrés et de gold products, mais surtout beaucoup plus de vin et de viande. Les deux autres clusters achètent de manière beaucoup plus modérée, mais on peut quand même voir des différences : le cluster 0 (violet) achète effectivement plus de vin, et aussi un peu plus de viande que le cluster 2 (jaune).

On peut aussi observer une légère différence d'âge, même s'il semble que l'âge n'influence pas vraiment le choix des clusters. Les niveaux d'achats en ligne ou en magasin ne sont pas significativement différents entre les clusters, et l'on observe pour tous les trois plus d'achats en magasin qu'en ligne. Cependant, on peut voir que le cluster 1, plus aisé, visite très peu le site par rapport aux autres : nous avons déjà fait l'hypothèse plus tôt qu'ils n'avaient pas besoin de souvent le visiter car pas de restriction budgétaire, donc pas besoin de régulièrement visiter le site pour repérer d'éventuelles promotions. Cela est à garder en tête par rapport à la stratégie publicitaire du commerce sur le site web : il faut viser les démographies des deux autres clusters, avec des promotions les ciblant spécifiquement.

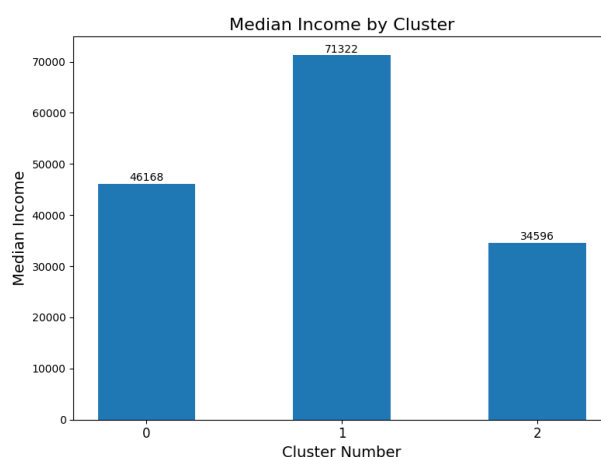


FIG. 2.11 – Revenu médian par cluster

On peut ici voir une autre confirmation des caractéristiques des clusters observées précédemment : les clients du cluster 1 sont significativement plus aisés par rapport aux deux autres clusters. On peut aussi confirmer la différence supposée entre les clusters 0 et 2 : malgré qu'ils soient tous les deux bien moins aisés que le cluster 1, le cluster 0 est plus aisé que le cluster 2.

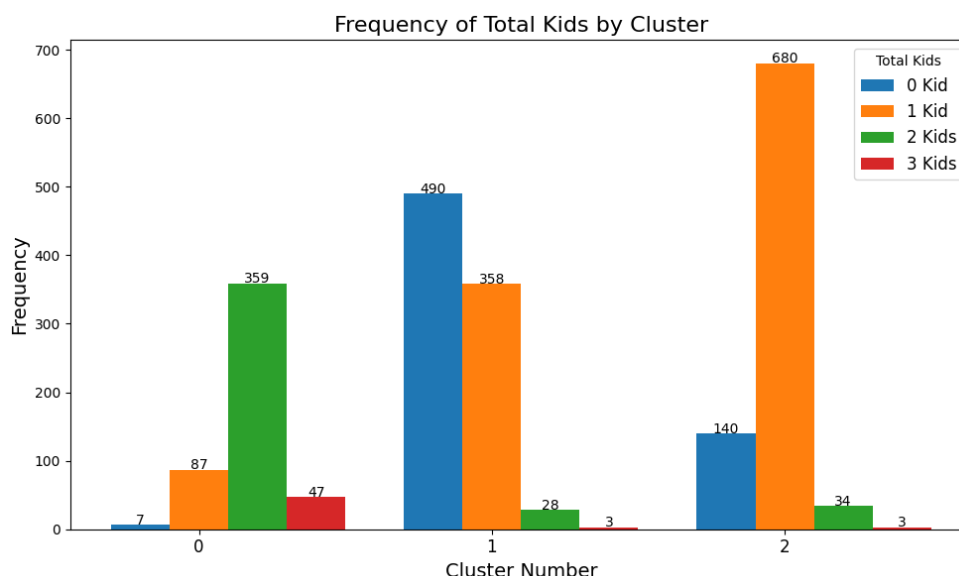


FIG. 2.12 – Nombre d'enfants selon les clusters

Cette dernière figure nous permet de repérer une nouvelle caractéristique différenciatrice de nos clusters : le cluster 1, aisé et dépensier, a généralement peu ou pas d'enfants (1 ou 0). Les clusters 0 et 2, moins aisés, ont eux beaucoup plus tendance à avoir des enfants, mais on peut là aussi observer une différence entre eux : le cluster 0, un peu plus aisé et plus dépensier en matière de viande et de vin, a tendance à avoir plus deux enfants, là où les clients du cluster 0 ont en majorité un seul enfant. Peut-être peuvent-ils avoir deux enfants en raison de leur revenu un peu plus élevé ?

Ces informations sur la propension des clients à avoir des enfants sont à garder en tête lors de l'élaboration de campagnes publicitaires, cela pourrait influencer leur choix de produits.

2.6 Insights pour la segmentation des clients

La segmentation de ce dataset nous a permis d'identifier trois démographies distinctes dans la clientèle de ce commerce.

Avec les observations que nous avons pu faire pour chacune d'entre elles, nous pouvons déjà penser à plusieurs possibilités de stratégie publicitaire :

- Nous pouvons cibler les clusters 0 et 2 dans des campagnes publicitaires en ligne : en effet, ces clients visitent plus souvent le site, certainement en attente de promotions intéressantes. On peut donc répondre à ces attentes et augmenter les achats, donc le chiffre d'affaires du commerce.
- Le cluster 0 est un cluster plus aisé que le cluster 2, mais qui n'achète pas significativement plus pour ce qui est des produits autres que le vin et la viande. Il faut donc cibler cette clientèle spécifiquement lors des prochaines campagnes, pour les pousser à acheter plus des autres produits. Cette clientèle étant généralement plus âgée avec

deux enfants, ce sont des facteurs à prendre en compte lors de l'élaboration de cette campagne.

- Les gold products étant peu achetés par rapport aux autres produits, il y a également une opportunité de faire une campagne et des promotions sur ces produits spécifiquement, pour pousser toutes les clientèles à en acheter un peu plus.

Conclusion

En conclusion, nous avons pu, à partir d'un dataset à priori difficilement interprétable, distinguer trois clientèles différentes avec des caractéristiques différents et des besoins différents. Cela a permis d'élaborer plusieurs idées préliminaires de campagnes publicitaires afin d'augmenter l'efficacité de ce commerce.

Un point d'amélioration que nous aurions pu intégrer dans notre étude est le statut marital, en plus de la présence d'enfants : les clients auront très certainement des fréquences ou des habitudes d'achat différents selon s'ils sont célibataires, ou vivent en couple avec un partenaire.