

Intervenants : Caroline Bérard (caroline.berard@univ-rouen.fr)

Membres du Groupe : Victor Jacob (victor.jacob.24@neoma-bs.com)
Jeremy Peterson (jeremy.petersondutoya.24@neoma-bs.com)

PROJET : Segmentation de clients d'une boutique en ligne

Dans un contexte de forte concurrence du commerce en ligne, il est essentiel pour les entreprises d'optimiser leur connaissance client afin d'améliorer leurs stratégies marketing et commerciales. Ce projet vise à analyser un jeu de données clients d'une boutique en ligne afin d'identifier des tendances et de segmenter les clients selon leurs comportements d'achat.

Ce rapport suit une approche méthodique en plusieurs étapes :

- Analyse exploratoire des données
- Études univariées et bivariées
- Analyse comportementale des clients
- Segmentation via des techniques de clustering

L'objectif final est d'obtenir une typologie claire des clients et d'identifier des leviers stratégiques pour maximiser les ventes en ligne.

1. PRESENTATION DES DONNEES

Ce rapport présente l'analyse exploratoire et la segmentation de la clientèle d'une boutique en ligne. L'objectif est de comprendre en profondeur le comportement des clients à travers l'exploration des données, la visualisation de leurs caractéristiques et l'application d'une méthode de clustering (KMeans) pour créer des segments distincts. Les analyses se fondent sur les variables suivantes : **Age**, **AnnualIncome**, **SpendingScore**, **OnlinePurchases**, **Website-Visits**, **Gender** et

Le jeu de données étudié contient des informations précieuses permettant d'analyser le comportement des clients d'une boutique en ligne. Il comprend **20000** observations et **8** variables principales :

- **CustomerID** : Identifiant unique du client
- **Age** : Âge du client (en années)
- **Gender** : Genre du client (Male/Female)
- **AnnualIncome** : Revenu annuel du client (en milliers d'euros)
- **SpendingScore** : Score de propension à dépenser (échelle de 0 à 100)
- **OnlinePurchases** : Nombre d'achats en ligne effectués au cours des 12 derniers mois
- **WebsiteVisits** : Nombre total de visites sur le site de la boutique sur la même période
- **CategoryPreference** : Catégorie de produits préférée du client (Clothing, Electronics, Books)

1.1 Qualité / Intégrité des données

Avant toute analyse approfondie, il est essentiel d'évaluer la qualité des données. Nous avons examiné les points suivants :

- **Valeurs manquantes** : Vérification des entrées vides ou nulles
- **Valeurs aberrantes** : Identification des données incohérentes ou extrêmes
- **Distribution des variables** : Analyse des statistiques descriptives et des tendances globales

Informations Générales :

`<class 'pandas.core.frame.DataFrame'>`

RangeIndex: 20000 entries, 0 to 19999

Data columns (total 8 columns):

#	COLUMN	NON-NULL COUNT	DTYPE
0	CustomerID	20000 non-null	int64
1	Age	19800 non-null	float64
2	Gender	20000 non-null	object
3	AnnualIncome	19800 non-null	float64
4	SpendingIncome	19800 non-null	float64
5	OnlinePurchases	19800 non-null	float64
6	WebsiteVisits	19800 non-null	float64
7	CategoryPreference	20000 non-null	object

dtypes: float64(5), int64(1), object(2)

memory usage: 1.2+ MB

Aperçu des Données :

CustomerID	Age	Gender	AnnualIncome	SpendingScore	OnlinePurchases	WebsiteVisits	CategoryPreference
1	50.0	Female	58.973825	39.430462	14.0	26.0	Clothing
2	43.0	Female	49.754264	47.801850	9.0	17.0	Electronics
3	37.0	Male	45.072539	25.052909	8.0	7.0	Books
4	38.0	Female	45.474369	43.367459	13.0	23.0	Clothing
5	50.0	Female	60.628914	62.271990	14.0	27.0	Electronics

Statistiques Descriptives :

	CustomerID	Age	AnnualIncome	SpendingScore	OnlinePurchases	WebsiteVisits
count	20000.000000	19800.000000	19800.000000	19800.000000	19800.000000	19800.000000
mean	10000.500000	42.929495	51.463545	50.533765	9.642273	18.942424
std	5773.647028	11.854050	6.670718	15.441311	1.888011	4.130665
min	1.000000	18.000000	32.396365	0.000000	5.000000	6.000000
25%	5000.750000	32.000000	46.324083	40.314082	8.000000	16.000000
50%	10000.500000	44.000000	51.777003	50.456715	9.000000	19.000000
75%	15000.250000	52.000000	56.411454	61.011978	11.000000	22.000000
max	20000.000000	70.000000	72.720464	100.000000	18.000000	37.000000

- Nettoyage des Données :

On peut constater qu'il y a beaucoup de valeur manquante. On va essayer de nettoyer ces données en donnant des valeurs (moyenne, médiane) selon le contexte :

Valeurs manquantes par colonne:	Pourcentage de valeurs manquantes:
CustomerID 0	CustomerID 0.0
Age 200	Age 1.0
Gender 0	Gender 0.0
AnnualIncome 200	AnnualIncome 1.0
SpendingScore 200	SpendingScore 1.0
OnlinePurchases 200	OnlinePurchases 1.0
WebsiteVisits 200	WebsiteVisits 1.0
CategoryPreference 0	CategoryPreference 0.0
dtype: int64	dtype: float64 %

Lignes avec des valeurs NA:								
	CustomerID	Age	Gender	AnnualIncome	SpendingScore	OnlinePurchases	WebsiteVisits	CategoryPreference
18	19	50.0	Male	53.627904	37.311743	NaN	17.0	Clothing
27	28	56.0	Female	54.379013	56.117743	NaN	18.0	Beauty
28	29	50.0	Male	NaN	43.113045	9.0	22.0	Books
29	30	45.0	Female	50.955480	68.529960	9.0	NaN	Sports
110	111	49.0	Female	57.028383	30.916267	NaN	20.0	Clothing

Voici maintenant le nombre de valeurs manquantes après avoir fini de nettoyer notre dataset :

Remplacement Check:	
CustomerID	0
Age	0
Gender	0
AnnualIncome	0
SpendingScore	0
OnlinePurchases	0
WebsiteVisits	0
CategoryPreference	0
dtype:	int64

On peut maintenant refaire une Description des Données post-Nettoyage :

```

Informations générales :
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CustomerID            20000 non-null  int64
1   Age                   20000 non-null  float64
2   Gender                20000 non-null  object
3   AnnualIncome          20000 non-null  float64
4   SpendingScore         20000 non-null  float64
5   OnlinePurchases       20000 non-null  float64
6   WebsiteVisits         20000 non-null  float64
7   CategoryPreference    20000 non-null  object
dtypes: float64(5), int64(1), object(2)
memory usage: 1.2+ MB
None

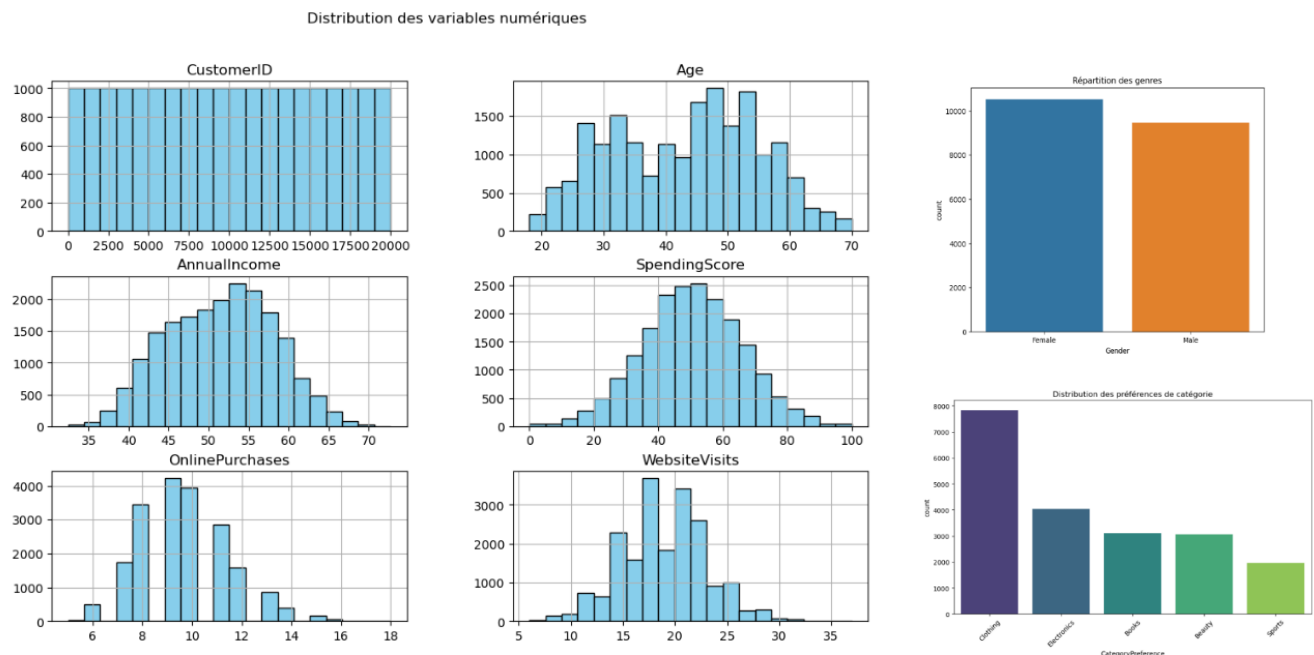
Statistiques descriptives :

```

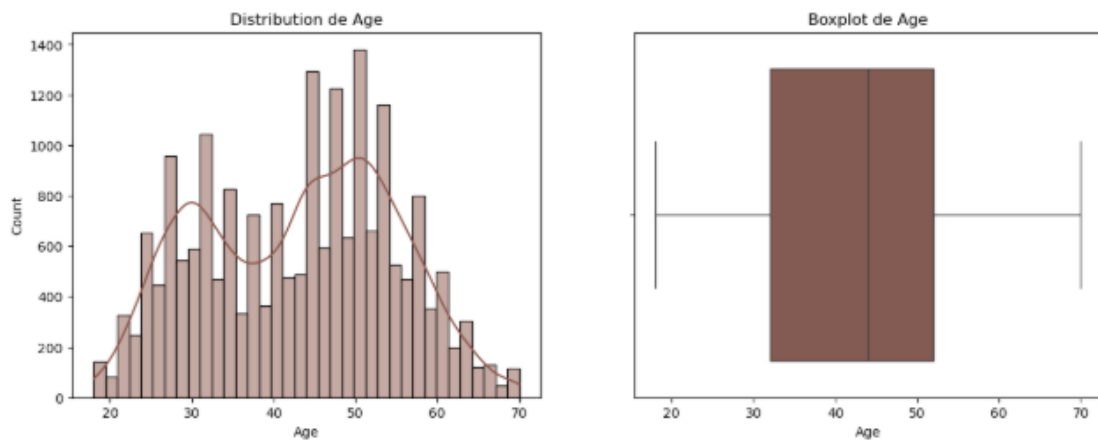
	CustomerID	Age	AnnualIncome	SpendingScore	OnlinePurchases	WebsiteVisits
count	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000
mean	10000.500000	42.940200	51.463545	50.532994	9.635850	18.943000
std	5773.647028	11.795108	6.637279	15.363908	1.879633	4.109963
min	1.000000	18.000000	32.396365	0.000000	5.000000	6.000000
25%	5000.750000	32.000000	46.376150	40.443302	8.000000	16.000000
50%	10000.500000	44.000000	51.671144	50.456715	9.000000	19.000000
75%	15000.250000	52.000000	56.361246	60.904592	11.000000	22.000000
max	20000.000000	70.000000	72.720464	100.000000	18.000000	37.000000

2. ANALYSE UNIVARIEES

Maintenant que notre dataset est complet, on peut passer à l'analyse univariée qui permet d'examiner chaque variable séparément afin d'identifier des tendances, des valeurs atypiques et des insights clés. En étudiant des indicateurs comme l'âge des clients, leur revenu annuel, leurs habitudes de navigation ou encore leurs préférences d'achat, nous pouvons mieux cerner le profil type des consommateurs de la boutique en ligne. Voici une première Overview avant de rentrer plus en détails dans notre analyse :



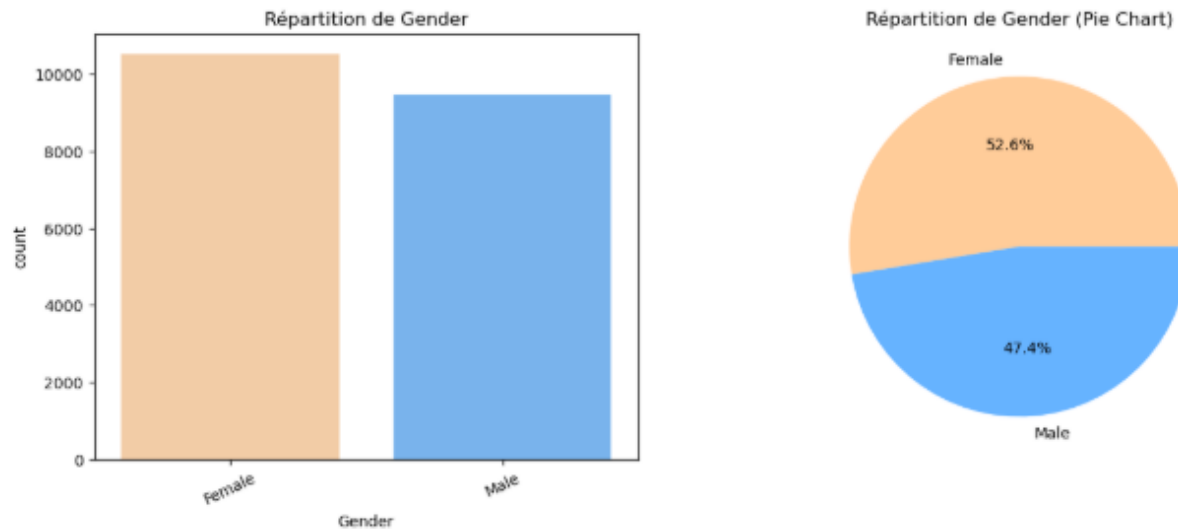
- AGE DES CLIENTS



L'histogramme montre une distribution bimodale (entre 25-33ans et 45-55ans) avec deux pics plus importants à 45 et 52 ans. Si l'étendu va de 18 à 70 ans, l'intervalle interquartile est entre 30 et 52 ans, bien que quelques valeurs extrêmes soient présentes. La concentration est plus forte dans la tranche médiane (40-50 ans), avec peu de jeunes (moins de 25 ans) ou de personnes âgées (plus de 60 ans).

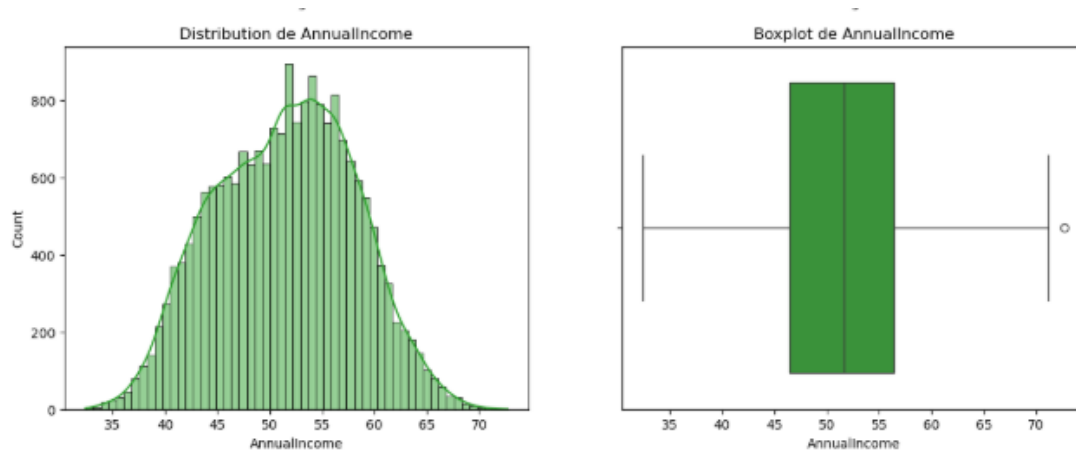
Le online-Shop attire principalement des clients d'âge moyen, potentiellement avec un pouvoir d'achat élevé.

- SEXE DES CLIENTS



On peut constater une répartition presque égale des sexes (47% d'hommes et 53% de femmes) dans la boutique en ligne qui est à notre sens un signe positif. Cela indique que la boutique attire une clientèle diversifiée et qu'elle n'est pas fortement biaisée en faveur d'un sexe particulier. Cette balance suggère que les produits ou services proposés répondent aux besoins et préférences d'un large éventail de clients, ce qui peut être un indicateur de l'efficacité de la stratégie de marketing et de la pertinence de l'offre pour différents segments de la population. Une telle distribution peut également offrir des opportunités d'optimisation, par exemple, en développant des offres ou des campagnes marketing encore plus ciblées pour chaque groupe, tout en maintenant un équilibre qui profite à l'ensemble de la clientèle.

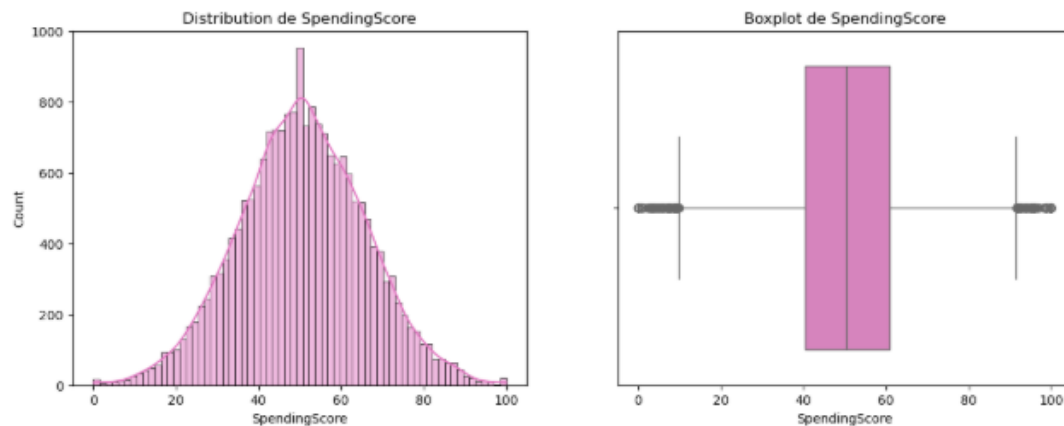
- REVENU ANNUEL



La distribution des revenus annuels suit presque une loi normale, avec un pic autour des 50-55k€. Grace au boxplot on peut voir que l'IQR est entre 45K€ et 57K€, bien qu'une valeur extrême dépasse les 70K€. La dispersion est modérée et la distribution symétrique, avec quelques clients à revenus très élevés, susceptibles d'être intéressés par des produits haut de gamme. On peut voir que le niveau social de ces clients est relativement élevé, surement des cadres en considérant l'âge déjà analysé auparavant.

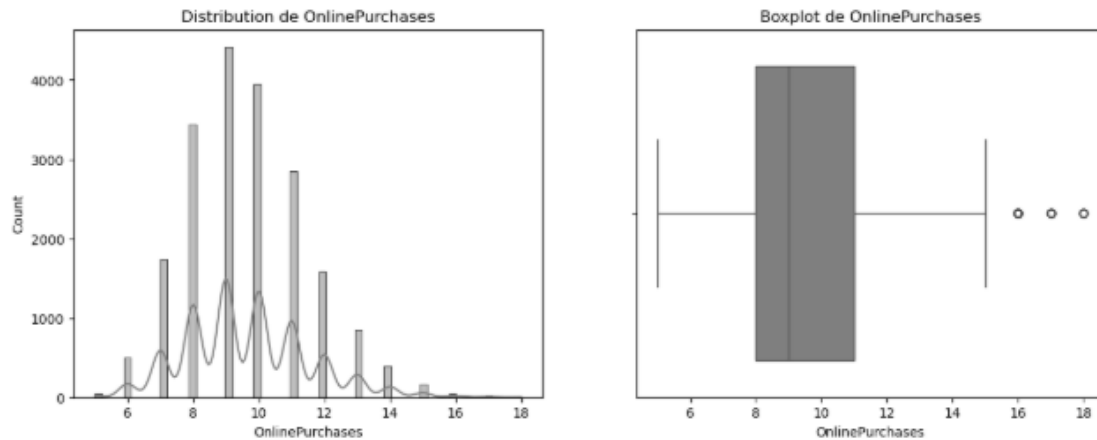
Le public cible semble appartenir à une classe moyenne supérieure, et une segmentation par revenu pourrait être utile pour adapter les offres promotionnelles en fonction du pouvoir d'achat. Le type de produit vendu dans cette boutique doit surement être haut de gamme ou de qualité supérieur, symbolisme d'une appartenance à une classe social middle-aisé.

- SPENDING SCORE



Le spending score qui représente la propension du client à acheter suit une parfaite loi normale centrée à 50 et symétrique. Il sera intéressant d'analyser les clients avec un spending score proche des 100 afin d'analyser leur particularité et les caractéristiques qui booste la conversion.

- NOMBRE DE COMMANDE EN LIGNE

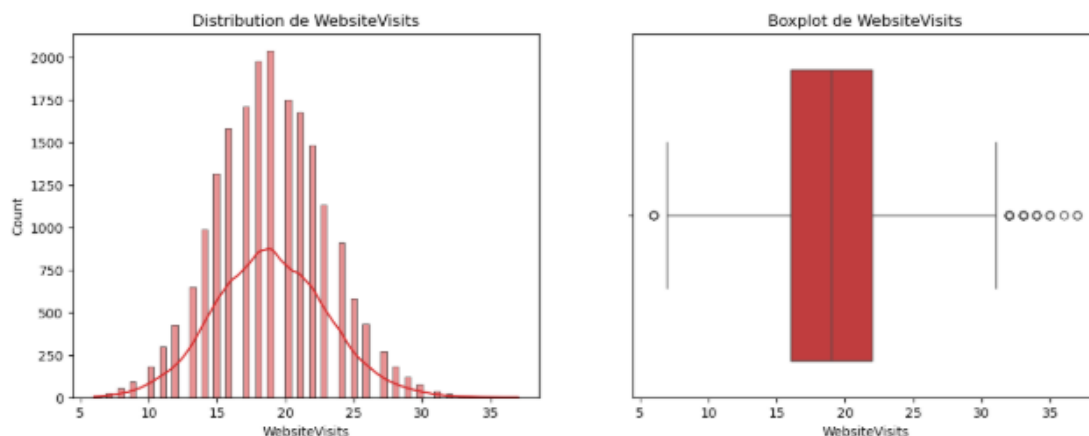


L'histogramme du Nombre de Commandes effectuées couplé aux statistiques descriptives est très intéressant à analyser. Le nombre Moyen de commande effectuée est de 9.6 avec une médiane à 9 et un IQR entre 8 et 11. Cela veut dire que les clients sont satisfait des produits et que la fidélisation est bonne. Nous avons jugé important de s'attarder un peu plus sur ce point :

Un nombre moyen de commandes par client de 9.6 par an, avec une médiane à 9 et un IQR (intervalle interquartile) entre 8 et 11, est un excellent indicateur de la satisfaction client et de la qualité des produits pour plusieurs raisons :

1. Un client qui effectue en moyenne 10 commandes par an montre un intérêt constant pour les produits ou services proposés. Cela suggère une satisfaction générale élevée, car les clients satisfaits sont plus enclins à revenir régulièrement pour effectuer de nouveaux achats.
2. Le fait que la médiane soit proche de la moyenne (9 – 9.6) indique que la majorité des clients effectuent des achats relativement similaires en termes de fréquence. Cela démontre une stabilité dans la fidélité des clients, ce qui est souvent le signe d'une bonne expérience client.
3. L'IQR entre 8 et 11 montre une faible dispersion dans le nombre de commandes, ce qui indique que la plupart des clients commandent dans une plage similaire. Un IQR restreint signifie que la plupart des clients ont des comportements d'achat cohérents, ce qui est typique lorsque la satisfaction et la qualité des produits sont constantes.
4. Une fréquence de commande élevée est souvent le résultat de la satisfaction continue d'un client, qui revient acheter de manière régulière. Cela suggère que les produits répondent aux attentes des clients, ce qui renforce la perception de qualité et ceux peut importe le Spending Score et le Revenu.

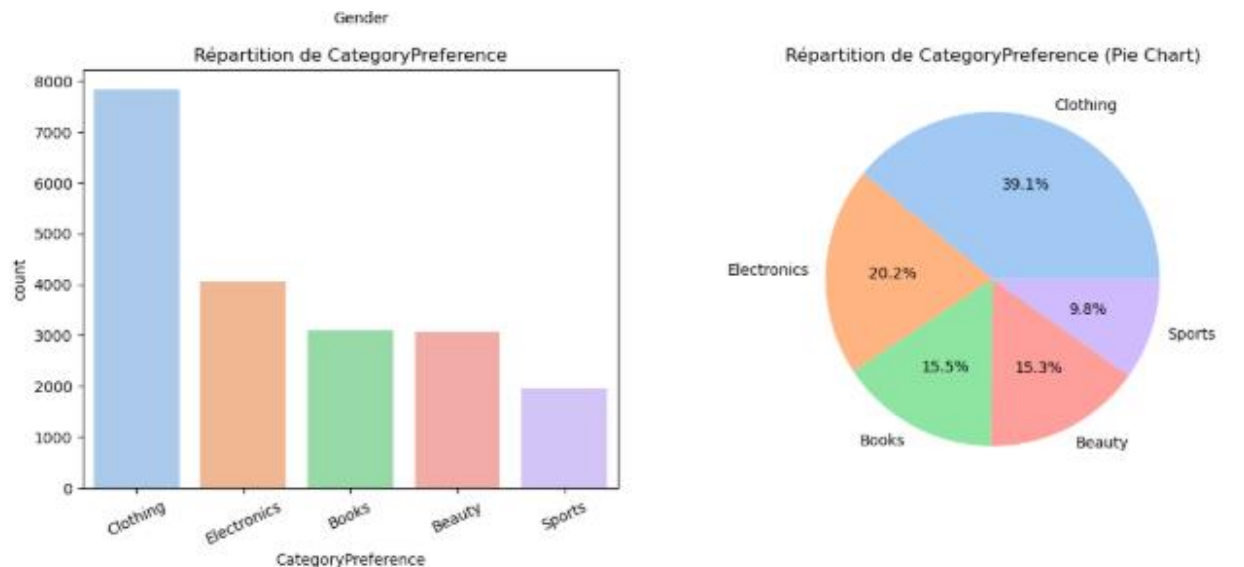
- NOMBRE DE VISITES DU SITE WEB



L'histogramme montre une distribution normale du nombre de visites, centrée autour de 18-22 visites par an, avec quelques valeurs extrêmes au-delà de 30 visites. Certains outsiders visitent bien plus fréquemment le site avec plus de 30 visites, ce qui peut refléter une fidélité ou une indécision. D'autres, avec moins de 10 visites, sont probablement des visiteurs occasionnels ou récents.

Une fréquence élevée de visites suggère un certain intérêt pour les produits. Il serait pertinent d'analyser le taux de conversion des visiteurs fréquents et de proposer des programmes de fidélisation ou des offres exclusives pour encourager l'achat.

- PREFERENCES EN TERMES DE CATEGORIES DE PRODUITS



En analysant le bar chart et le pie chart, il est clair que les préférences des clients en termes de catégories de produits sont bien définies. Le Clothing domine largement avec 39%, suivi de l'Electronics à 20%, puis les Books à 15,5%, la Beauty à 15,3% et enfin le Sports à 9,8%. Cette répartition explique la diversité de la clientèle en termes d'âge, car les produits couvrent une large gamme d'intérêts et peuvent attirer différentes tranches d'âge et les deux sexes.

L'équilibre des sexes et la diversité des préférences d'âge laissent entendre qu'une analyse plus approfondie de l'influence de l'âge ou du sexe sur les choix de catégories serait intéressante. Cela pourrait fournir des insights précieux pour personnaliser l'expérience d'achat, comme adapter le marketing ou les recommandations de produits.

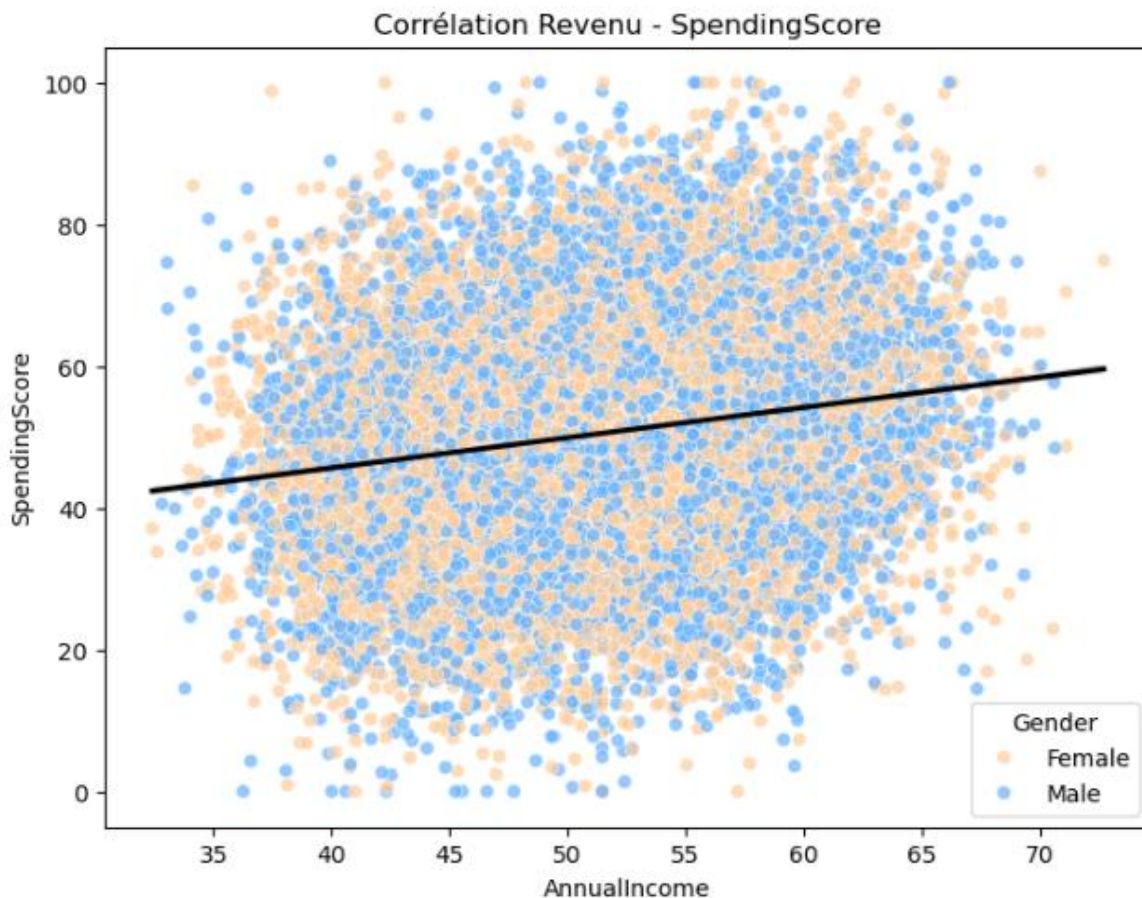
Cette boutique en ligne rappelle une sorte de Walmart virtuel ou une marketplace, offrant une large variété de produits pour satisfaire une clientèle diversifiée, ce qui permet de répondre à des besoins très variés tout en maintenant une plateforme accessible à tous.

3. ANALYSE BIVARIEES

Pour donner suite aux analyses univariés, il est crucial d'étudier les interactions entre les variables pour identifier des corrélations et tendances significatives. L'analyse bivariée permet d'explorer les relations entre deux variables, offrant des insights précieux pour comprendre le comportement des clients et ajuster les stratégies marketing.

3.1- Relation entre le Revenu Annuel du client et sa propension à Dépenser

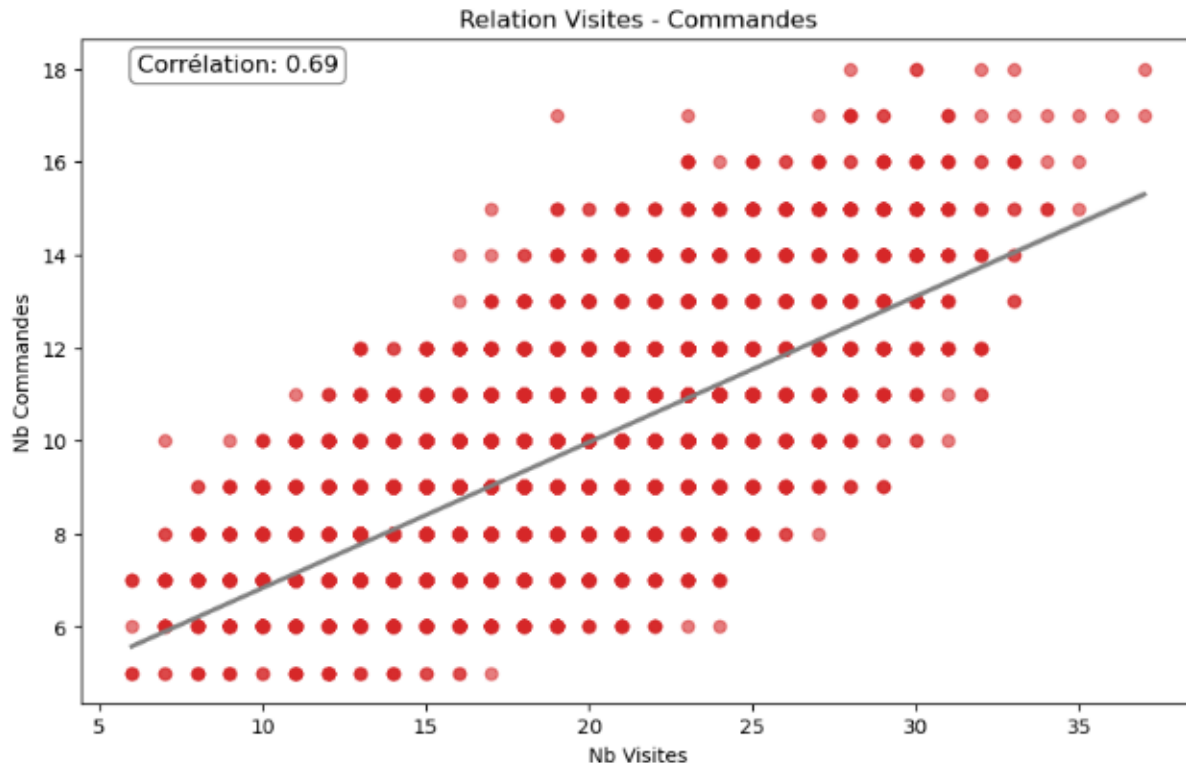
Corrélation entre AnnualIncome et SpendingScore: 0.18



Le nuage de points montre une faible corrélation positive de 0,18 entre le revenu annuel et le score de dépenses, sans différence marquée entre les sexes (Nous avons un message d'erreur pour notre graphique lorsque nous ne spécifions pas le 'hue' du Scatter Plot alors nous avons trouvé judicieux d'ajouter un niveau d'information supplémentaire avec le Sexe)

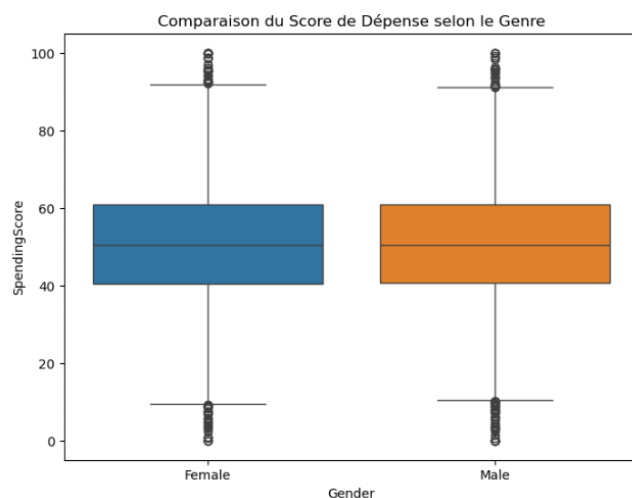
La dispersion des points indique que le revenu seul n'explique pas de manière significative la propension à dépenser. Bien qu'un revenu plus élevé soit légèrement associé à un score de dépenses plus élevé, cette relation reste faible. Cette faible corrélation suggère que d'autres facteurs, tels que les préférences d'achat, les habitudes de consommation ou la fréquence des visites, influencent probablement davantage le comportement d'achat. Il serait donc pertinent d'explorer ces variables pour mieux comprendre les comportements des clients.

3.2- Relation entre le nombre de visite et le nombre d'achats en ligne



Le scatter plot montre une corrélation de 0,69 entre les visites sur le site et les achats en ligne, indiquant une relation positive significative. Plus un client visite fréquemment le site, plus il effectue des achats. On remarque qu'un client effectuant plus de 25 visites par an commande entre 8 et 16 fois par an. Toutefois, certains clients visitent fréquemment jusqu'à 24 fois en achetant aussi peu que certains qui n'ont visité le site que 6 fois. Il est toutefois intéressant de maximiser le nombre de visites par an, en utilisant des publicités ciblées, des emails, des activités sur les réseaux sociaux ou des campagnes Paid Media (Google Ads), afin de stimuler les ventes en ligne.

3.2- Différences d'aptitude à dépenser entre les sexes



Il n'y a aucune différence notable entre la propension à acheter et le sexe du client. Les comportements d'achat semblent similaires chez les hommes et les femmes. C'est une bonne chose pour la boutique en ligne car cela montre que l'offre de produits répond à un large éventail de consommateurs, sans être biaisée vers un sexe particulier. Cela signifie que la boutique attire une clientèle diversifiée et que les stratégies marketing sont efficaces pour atteindre tous les segments de la population.

Cependant il serait intéressant de voir comment cela varie en fonction des catégories de produit :

Analyse des préférences de catégorie par genre :

CategoryPreference	Beauty	Books	Clothing	Electronics	Sports
Female	1599	1646	4119	2119	1042
Male	1465	1454	3703	1926	927

Test du χ^2 :

χ^2 : 0.68

p-value : 0.954

Degrés de liberté : 4

V de Cramer : 0.006

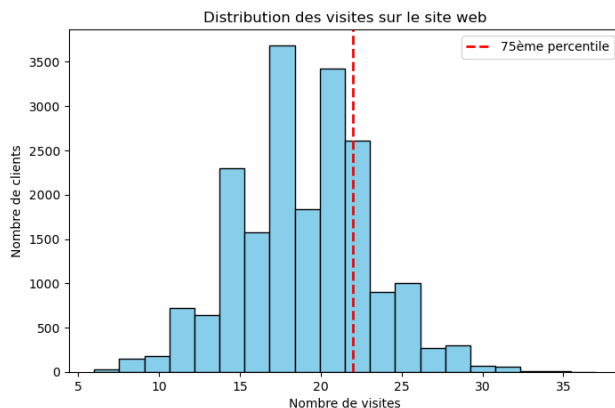
ANALYSE COMPORTEMENTALE

Analyse des clients très actifs

Objectif : Dans cette analyse, nous avons exploré les comportements des clients pour identifier ceux qui visitent très souvent le site web. L'objectif est de distinguer un segment de clients « hyper-actifs » pouvant avoir un impact significatif sur les stratégies marketing.

Méthodologie : Nous avons choisi d'utiliser le 75^e percentile du nombre de visites pour définir un seuil pertinent. Une fois ce seuil déterminé, nous avons extrait un sous-ensemble de clients dépassant cette valeur et comparé leurs caractéristiques moyennes (âge, revenu annuel, SpendingScore et nombre d'achats en ligne) avec celles de l'ensemble des clients. La visualisation de la distribution via un histogramme, avec une ligne verticale marquant le seuil retenu, a été utilisée pour faciliter l'interprétation des données.

Résultats et Interprétation : Cette démarche a révélé que les clients très actifs semblent avoir un SpendingScore et une fréquence d'achats en ligne supérieurs à la moyenne, indiquant un engagement plus important envers la boutique en ligne. La visualisation renforce l'existence d'un groupe distinct de clients dont le comportement pourrait justifier des actions marketing ciblées.



Seuil de visites (75ème percentile) : 22.0

Caractéristiques moyennes des clients hyper-actifs :

Age 46.259041
 AnnualIncome 53.541627
 SpendingScore 51.267747
 OnlinePurchases 11.325657
 dtype: float64

Caractéristiques moyennes de l'ensemble des clients :

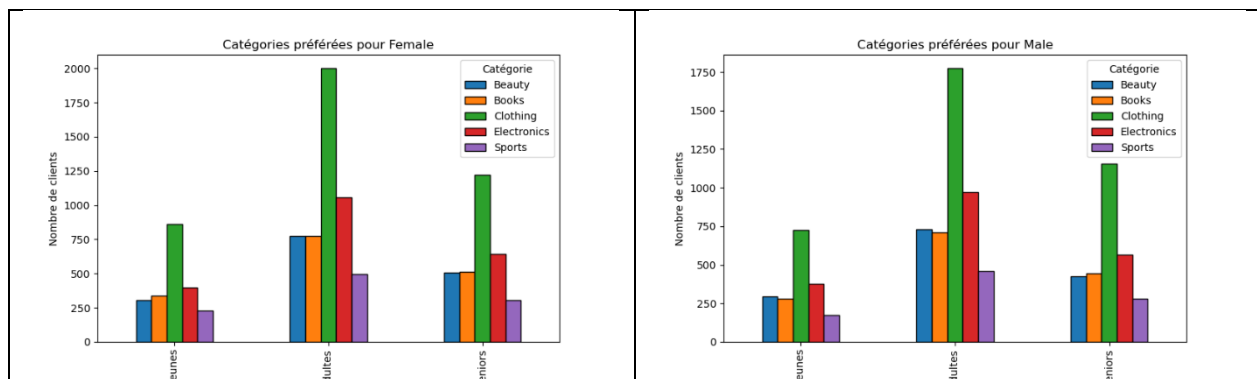
Age 42.940200
 AnnualIncome 51.463545
 SpendingScore 50.532994
 OnlinePurchases 9.635850
 dtype: float64

Analyse des Catégories Préférées par Sexe et par Âge

Objectif : Cette partie vise à analyser comment les préférences de catégories de produits varient en fonction du sexe et de l'âge des clients, afin d'identifier des tendances spécifiques.

Méthodologie : Nous avons créé une nouvelle variable « AgeGroup » en segmentant l'âge des clients en trois catégories : « Jeunes » (moins de 30 ans), « Adultes » (entre 30 et 50 ans) et « Seniors » (plus de 50 ans). Un tableau croisé regroupant les préférences de catégories par sexe et par tranche d'âge a été construit, et des graphiques en barres pour chaque sexe ont été générés pour visualiser la répartition des préférences selon les tranches d'âge.

Résultats et Interprétation : L'analyse a permis de visualiser des tendances distinctes selon les segments, suggérant que certains groupes, par exemple les jeunes hommes ou les adultes femmes, peuvent présenter des préférences marquées pour certaines catégories. Ces informations sont cruciales pour orienter des campagnes marketing personnalisées en fonction du profil des clients.



Classification des Clients par Aptitude à Dépenser

Objectif :

L'objectif ici est de segmenter les clients en trois groupes selon leur aptitude à dépenser, afin d'identifier des profils distincts susceptibles d'être ciblés par des offres spécifiques.

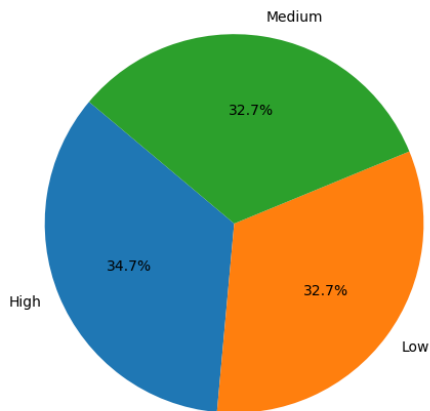
Méthodologie :

En se basant sur le SpendingScore, nous avons défini des seuils aux 33^e et 66^e percentiles pour classer les clients en trois catégories : Low (faible dépense), Medium (dépense moyenne) et High (forte dépense). Une nouvelle colonne « SpendingCategory » a été ajoutée au jeu de données, la répartition a été visualisée via un diagramme circulaire et nous avons calculé les moyennes de variables clés (âge, revenu annuel, nombre d'achats en ligne et nombre de visites) pour chaque groupe.

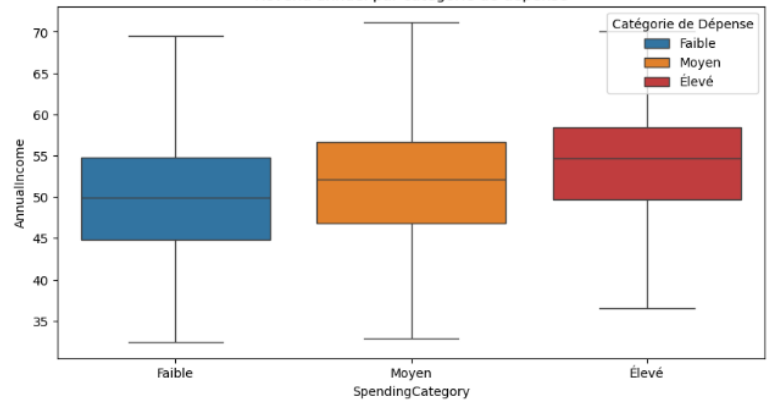
Résultats et Interprétation :

La distribution obtenue révèle des différences marquées entre les groupes, avec notamment un revenu et une fréquence d'achats plus élevés pour le groupe High. Ces insights fournissent une base solide pour adapter des actions marketing en fonction du profil de dépense de chaque segment.

Répartition des clients par aptitude à dépenser



Revenu annuel par catégorie de dépense



Statistiques moyennes par catégorie de dépense :

	Age	AnnualIncome	OnlinePurchases	WebsiteVisits
SpendingCategory				
High	44.956618	52.781768	9.767059	19.159118
Low	40.849091	50.113528	9.520909	18.781061
Medium	42.953788	51.455394	9.615606	18.882273

SEGMENTATION & CLUSTERING

Après imputation des valeurs manquantes et standardisation, un algorithme KMeans a été appliqué pour segmenter la clientèle en trois clusters distincts.

Clusters identifiés

Cluster 0 :

- **Revenu modéré :**

Ces clients disposent d'un pouvoir d'achat moyen. Ils ne possèdent pas de revenus très élevés, ce qui signifie qu'ils sont susceptibles de surveiller leurs dépenses et d'être sensibles aux prix. Leur revenu modéré suggère qu'ils n'ont pas forcément les moyens d'acheter systématiquement des produits hauts de gamme, mais ils restent une clientèle précieuse pour des offres adaptées.

- **SpendingScore faible à moyen :**

Le SpendingScore reflète la propension à dépenser. Un score faible à moyen indique que ces clients ne sont pas très enclins à dépenser de grosses sommes lors de leurs achats. Cela peut être dû à plusieurs facteurs : une attitude prudente vis-à-vis des dépenses, des habitudes d'achat conservatrices ou encore une réticence à investir dans des produits considérés comme non essentiels.

- **Interprétation globale :**

Ces clients représentent un segment stable mais qui n'est pas à forte croissance en termes de dépenses. Ils font leurs achats, mais restent dans une zone de confort budgétaire limitée. Il est donc essentiel de trouver des moyens pour les inciter à augmenter leur engagement et leurs dépenses, sans pour autant les pousser au dépassement de leur budget habituel.

Cluster 1 :

- **Revenu élevé, Spending Score modéré, activité en ligne régulière :**

Les clients ciblés affichent un revenu élevé, un SpendingScore modéré et une activité en ligne régulière, indiquant une gestion équilibrée de leurs dépenses et un intérêt pour des offres de qualité.

- **Interprétation globale :**

La stratégie consiste à leur proposer des programmes VIP et des offres personnalisées, adaptées à leur pouvoir d'achat, afin de renforcer leur fidélité par une expérience client exclusive.

Cluster 2 :

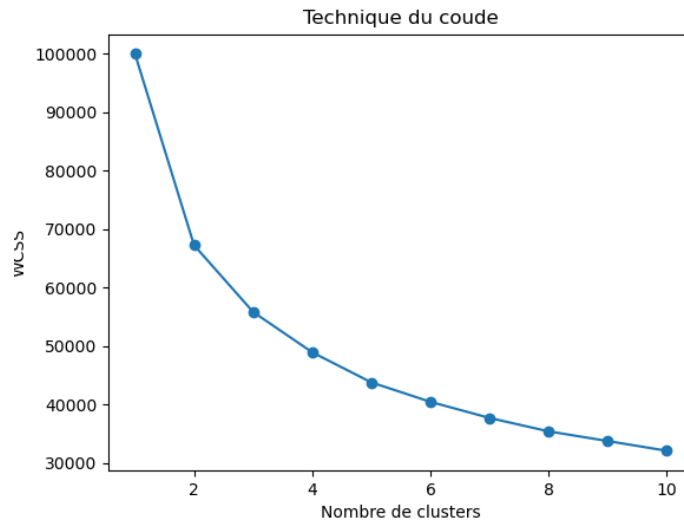
- **Revenu moyen, Spending Score élevé, activité en ligne élevé :**

Les clients présentant un SpendingScore élevé, bien qu'ils puissent parfois avoir un revenu moins élevé, se distinguent par une fréquence élevée de visites et d'achats,

reflétant un comportement d'achat très dynamique et une recherche continue d'opportunités.

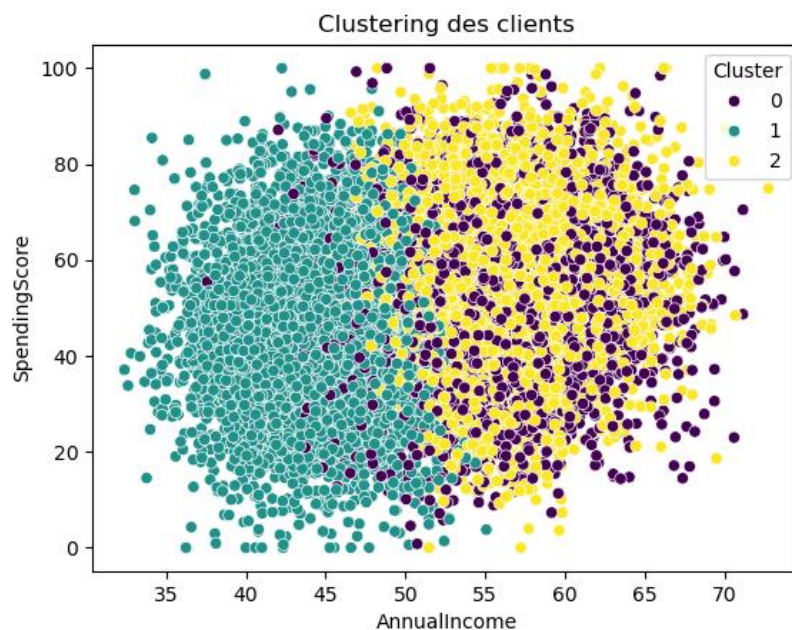
- **Interprétation globale :**

La stratégie repose sur la mise en place d'offres exclusives, d'avant-premières et de programmes de récompenses afin de valoriser leur engagement. Ces initiatives visent à renforcer leur fidélisation en leur offrant des avantages sur-mesure qui maximisent leur expérience et encouragent la récurrence de leurs achats.



Visualisation des clusters

Le graphique ci-dessous illustre la segmentation en utilisant AnnualIncome et SpendingScore.



Conclusion

1. Typologie des clients :

- Trois groupes principaux émergent du clustering :
 - **Faible revenu et faible dépense** : Visites fréquentes mais peu d'achats.
 - **Revenu moyen et dépense élevée** : Clients engagés avec une forte propension à acheter.
 - **Revenu élevé et dépense variable** : Certains sont des acheteurs fréquents, d'autres non.

2. Facteurs influençant la consommation :

- Une relation positive entre le **revenu annuel et le score de dépense**, bien qu'elle ne soit pas linéaire.
- Une corrélation claire entre le **nombre de visites sur le site et les achats en ligne**.
- Les **femmes semblent dépenser davantage en moyenne** que les hommes.
- Certaines catégories de produits attirent davantage selon l'âge et le sexe.

3. Comportements des clients :

- Les clients qui visitent fréquemment le site sans acheter peuvent être intéressés par des offres ou du contenu mais hésitent à finaliser leurs achats.
- Les préférences de catégories varient : certains groupes sont plus enclins à acheter des vêtements, d'autres des produits électroniques.

Plan d'action pour augmenter les ventes en ligne

1. Personnalisation & Segmentation :

- Adapter les recommandations et promotions selon les **profils identifiés** par le clustering.
- Mettre en place des **campagnes ciblées** en fonction des catégories de produits préférées.

2. Amélioration de l'Expérience Utilisateur :

- Optimiser le **parcours d'achat en ligne** pour réduire l'abandon de panier.
- Proposer des **suggestions de produits basées sur le comportement** des clients.

3. Incitation à l'Achat :

- Offrir des **réductions exclusives** pour les visiteurs fréquents sans achat.
- Utiliser le **marketing par email et SMS** pour relancer ces clients avec des offres adaptées.

4. Programme de Fidélisation :

- Mettre en place un **programme de fidélité** récompensant les achats répétés.
- Proposer des **avantages VIP** pour les clients à forte valeur ajoutée.

5. Optimisation du Tunnel de Conversion :

- Réduire les **frictions dans le processus d'achat** (ex : options de paiement simplifiées).
- Mettre en place un **chatbot ou assistance en ligne** pour répondre aux questions des clients en temps réel.

En appliquant ces stratégies, la boutique en ligne pourra **maximiser l'engagement et les conversions**, augmentant ainsi le nombre de ventes en ligne.