

LIBRAIRIE LAPAGE

ANALYSE DES VENTES 2021 - 2023

SOMMAIRE

Préparation des données	Nettoyage Analyses descriptive
Evolution des indicateurs de vente	Chiffre d'affaires et moyenne mobile Répartition du chiffre d'affaires par catégorie de produits Evolution du nombre de clients, de transactions et de produits vendus
Focus sur les références	Les Tops et les Flops Répartition des ventes par catégorie
Profils clients	Répartition du chiffre d'affaires pour les clients B2B Courbe de Lorenz
Analyse comportementale	Lien entre le genre des clients et les catégories de livres achetés Lien entre l'âge des clients et le montant total des achats Lien entre l'âge des clients et la fréquence des achats Lien entre l'âge des clients et la taille du panier moyen Lien entre l'âge des clients et la catégorie des livres achetés

Préparation des données



Customers

8621 lignes et 3 colonnes

Contient les 'client_id', le genre et l'année de naissance des clients

La clé 'client_id' est unique et sans doublons



Products

3286 lignes et 3 colonnes

Contient les 'id_prod', le prix et la catégorie de chaque référence

La clé 'id_prod' est unique est sans doublons



Transactions

1048575 lignes et 4 colonnes

Contient l'historique des transactions entre le 01/03/2021 et le 28/02/2023

Aucun mois manquant sur la période

361041 lignes contenant uniquement des valeurs Nan ont été supprimées

Analyse descriptive

3 625 produits différents
vendus sur la période
répartie en 3 catégories.

Catégorie 0 :

min: 0,62 €

max: 40,99 €

moy: 11,73 €

Catégorie 1:

min: 2 €

max: 80,99€

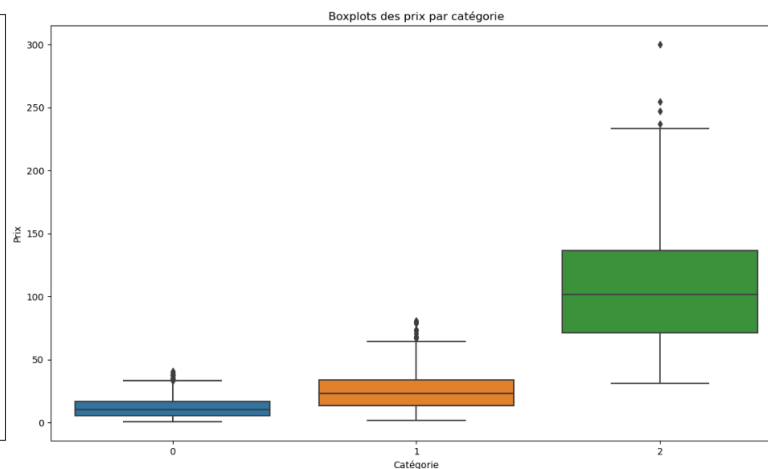
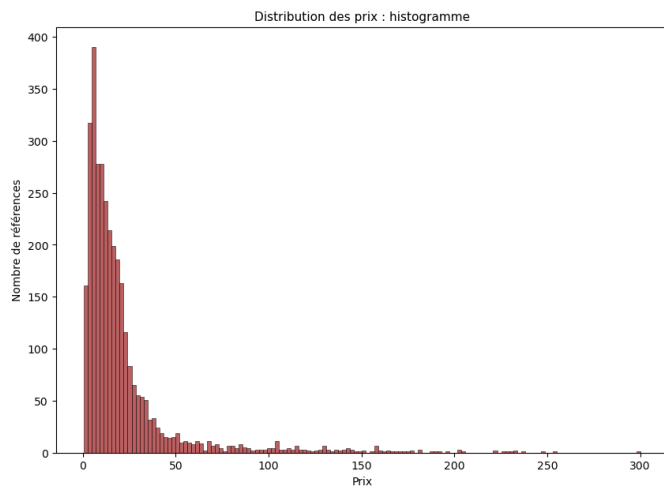
moy: 25.5€

Catégorie 2 :

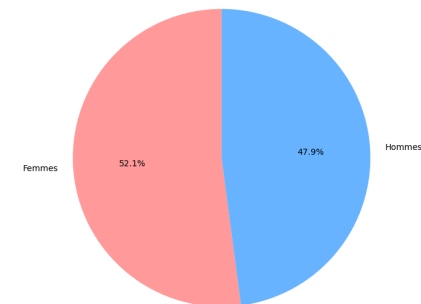
min: 30,99€

max: 300€

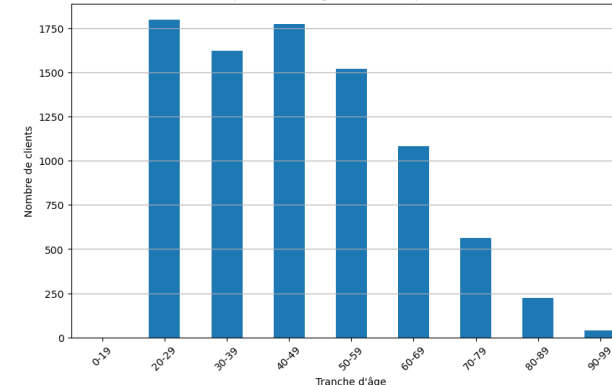
moy: 108,35€



Répartition des sexes des clients

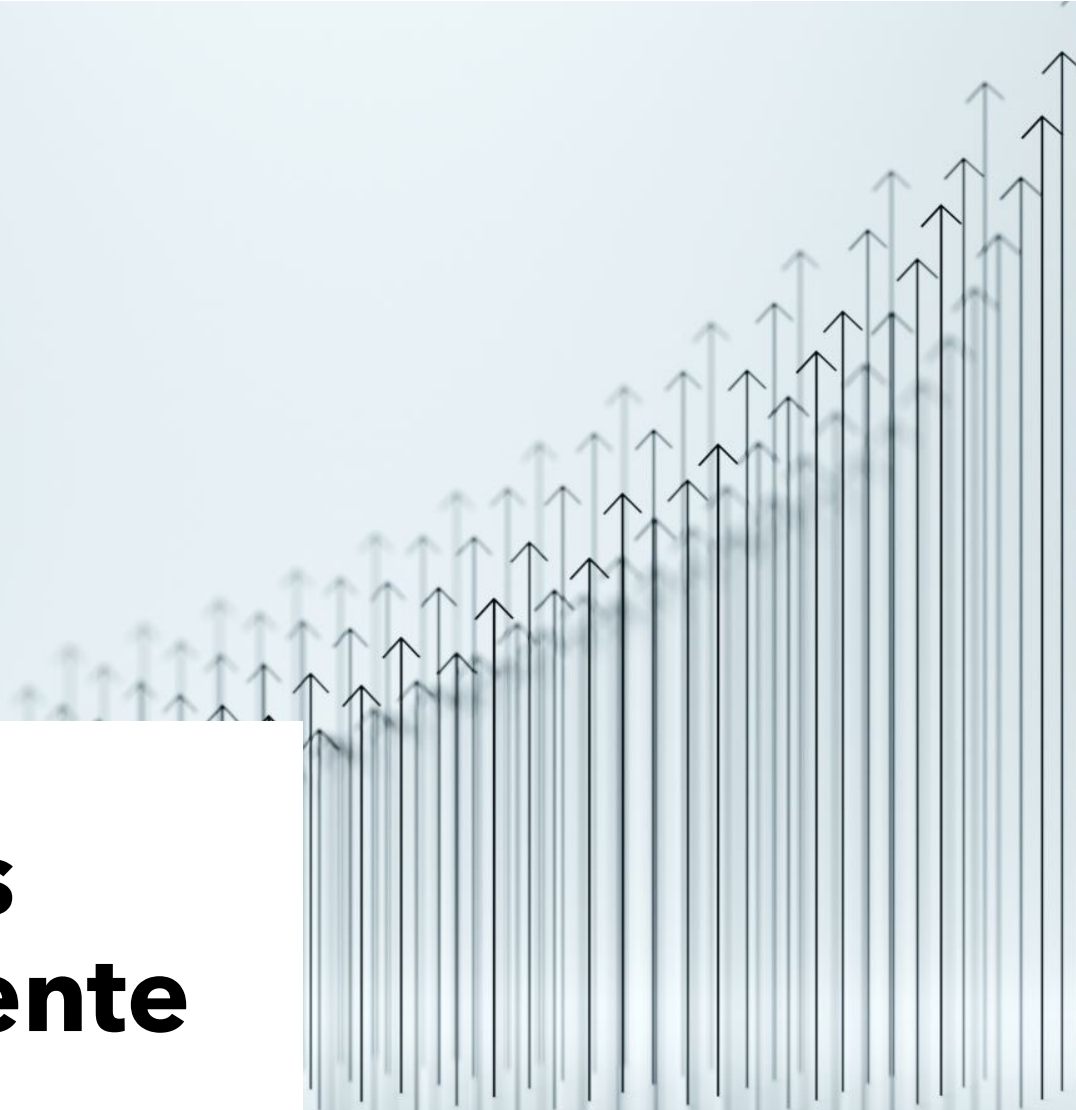


Répartition des âges des clients par tranche



8600 clients différents ont effectué au
moins une transaction sur la période
Age min = 20 ans
Age max = 95 ans
Age moyen = 46 ans

Evolution des indicateurs de vente

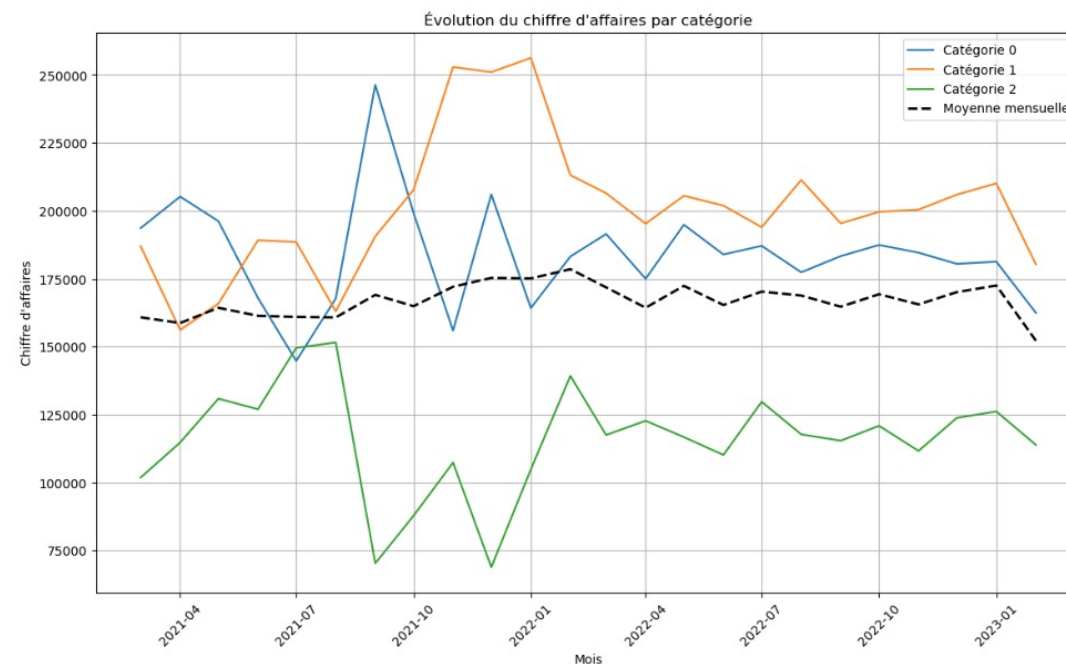
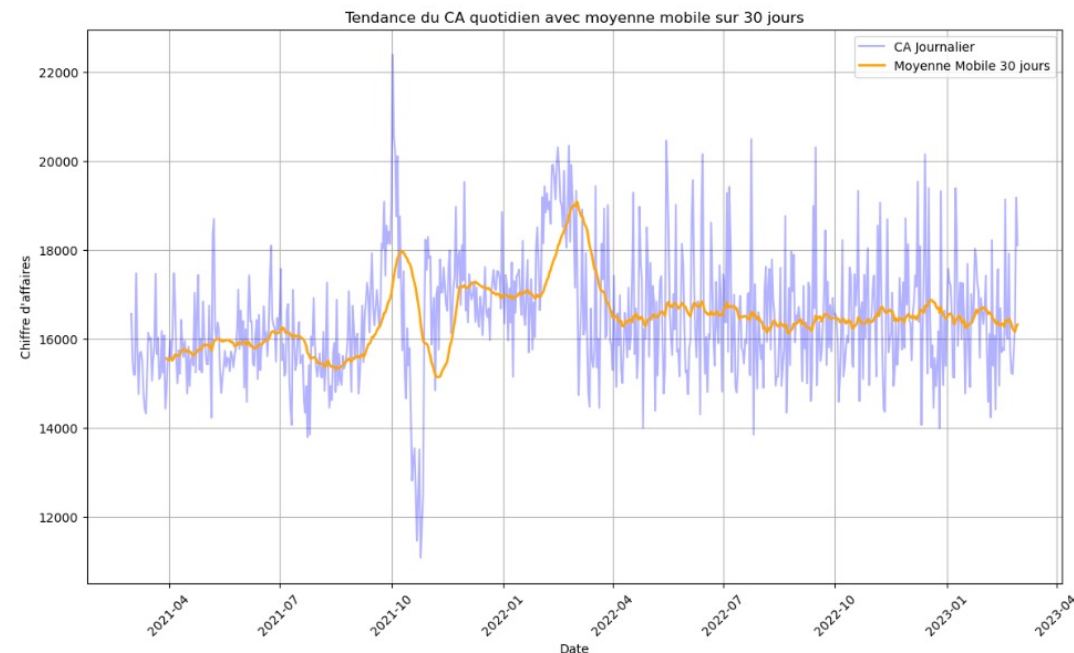


Chiffre d'affaires et moyenne mobile

CA de l'exercice 2021 = 5 986 813 €
CA mensuel moyen 2021 = 498 901 €

CA de l'année 2022 = 6 003 886 €
CA mensuel moyen 2022 = 500 323 €

Soit une variation entre 2021 et 2022 de: 0.29%



Chiffre d'affaires par catégorie de produits.

Chiffre d'affaires sur la période:

Catégorie 0 = 4 419 731€

Catégorie 1 = 4 827 657€

Catégorie 2 = 2 780 275€

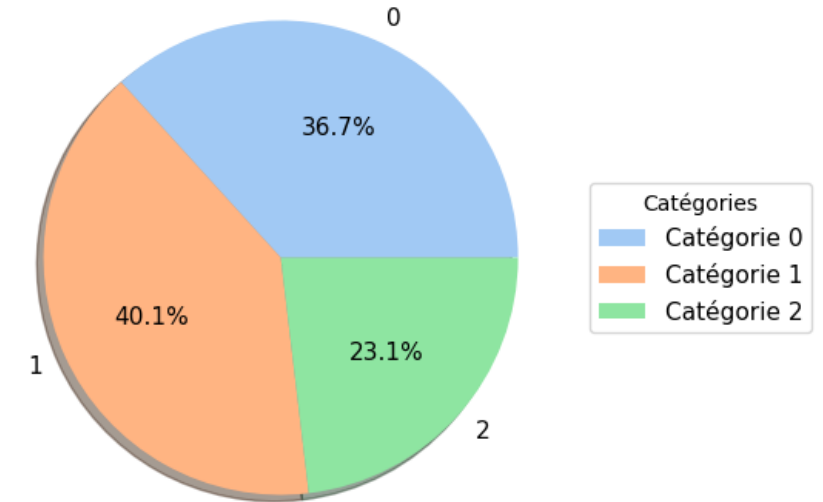
Nombres de ventes:

Catégorie 0 = 225 449

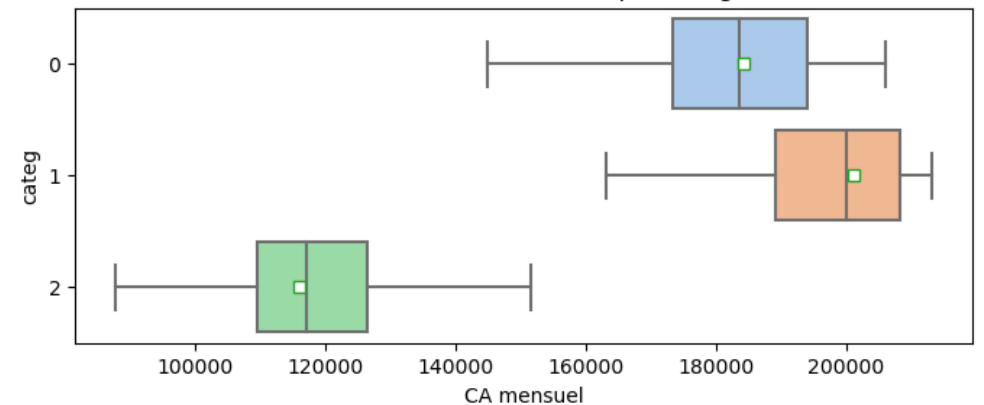
Catégorie 1 = 183 610

Catégorie 2 = 30 574

Répartition du Chiffre d'Affaires par Catégorie de Produits



Distribution des CA mensuel par catégorie



Evolution du nombre de clients, des transactions et des produits vendus

Nombre de clients par mois:

Tendance généralement stable avec une légère augmentation du nombre de clients jusqu'à un pic soudain en octobre 2021.

Après ce pic, les chiffres montrent une certaine volatilité avant de se stabiliser à nouveau autour d'une moyenne légèrement plus élevée qu'au début de la période observée.

Nombre de transactions par mois:

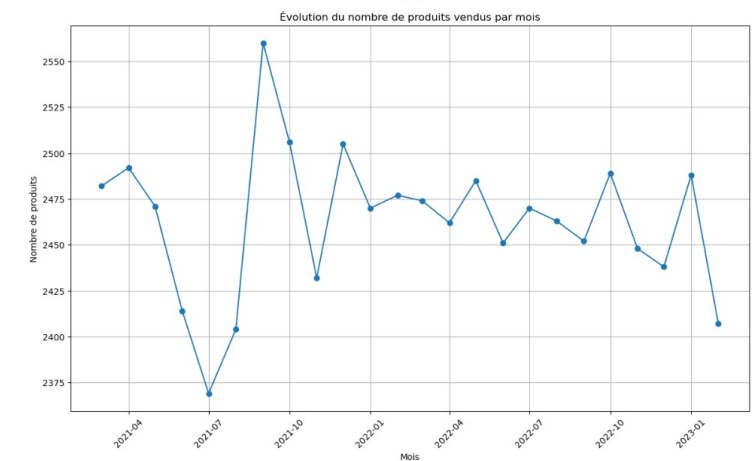
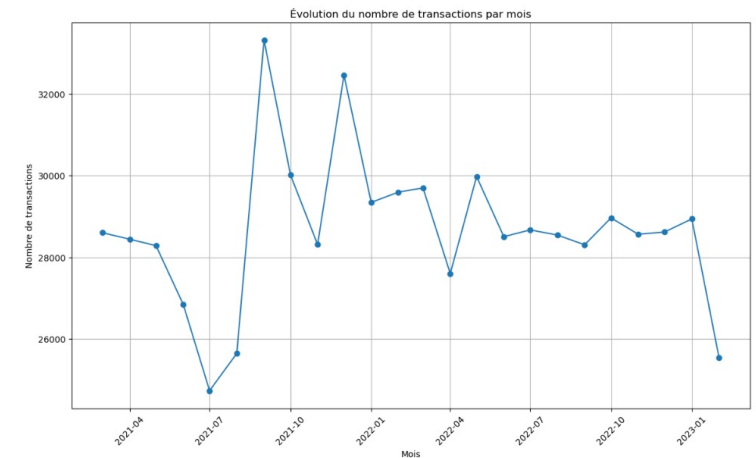
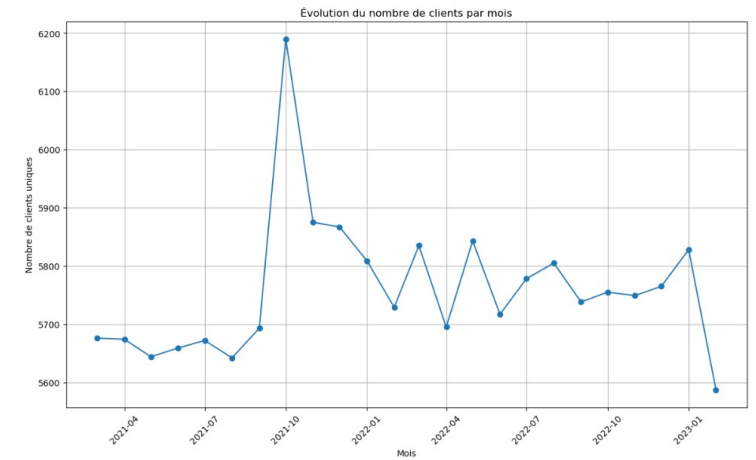
Grande volatilité, avec des pics significatifs, notamment en septembre puis décembre 2021.

La tendance à long terme est relativement stable malgré des fluctuations mensuelles marquées.

Nombre de produits vendus par mois:

Le nombre de produits vendus suit une tendance légèrement décroissante sur la période.

Tout comme les autres graphiques, un pic notable se produit en septembre 2021.





Focus sur les références

Les Tops et les Flops

809 références responsables de 80% des ventes soit 24.78% des références vendus

Nombre de produits responsables de 80% du CA : 699 soit 21.41% des produits

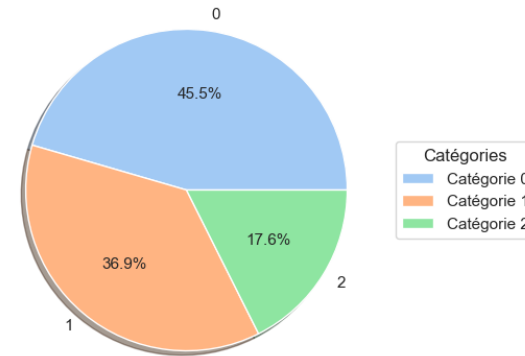
Nombre de références totalisant moins de 5 ventes:

Catégorie 0 = 131 soit 5,71% de la catégorie 0

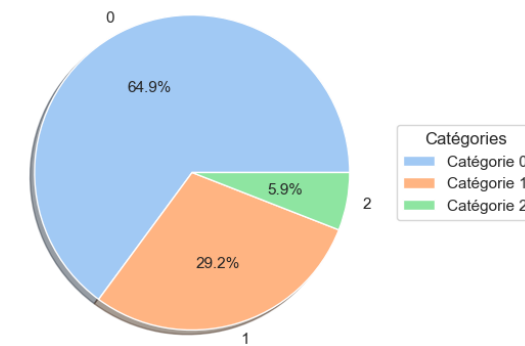
Catégorie 1 = 14 soit 1,89% de la catégorie 1

Catégorie 2 = 13 soit 5,50% de la catégorie 2

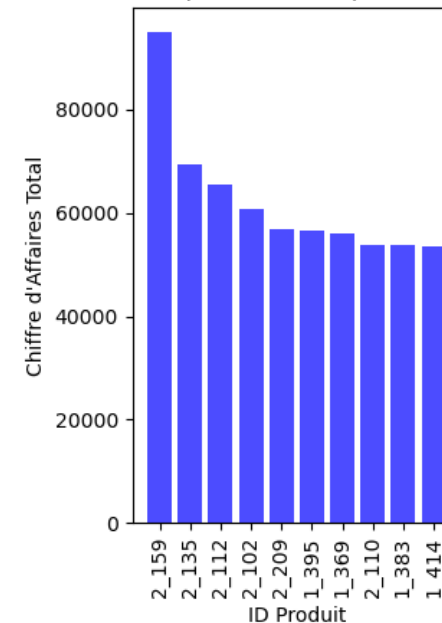
Répartition des produits responsables de 80% du CA par catégorie



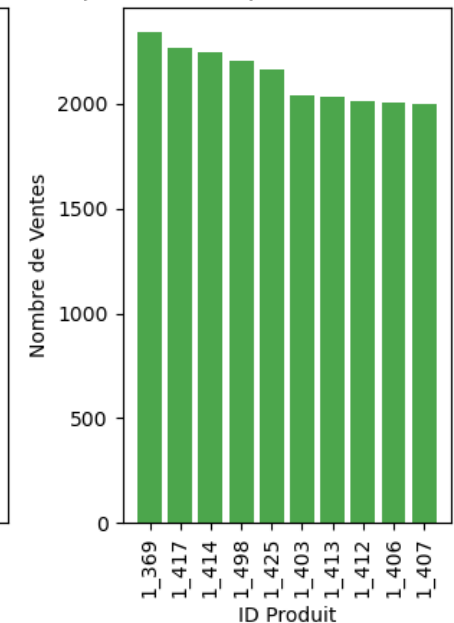
Répartition des produits responsables de 80% des ventes par catégorie



Top 10 Produits par CA

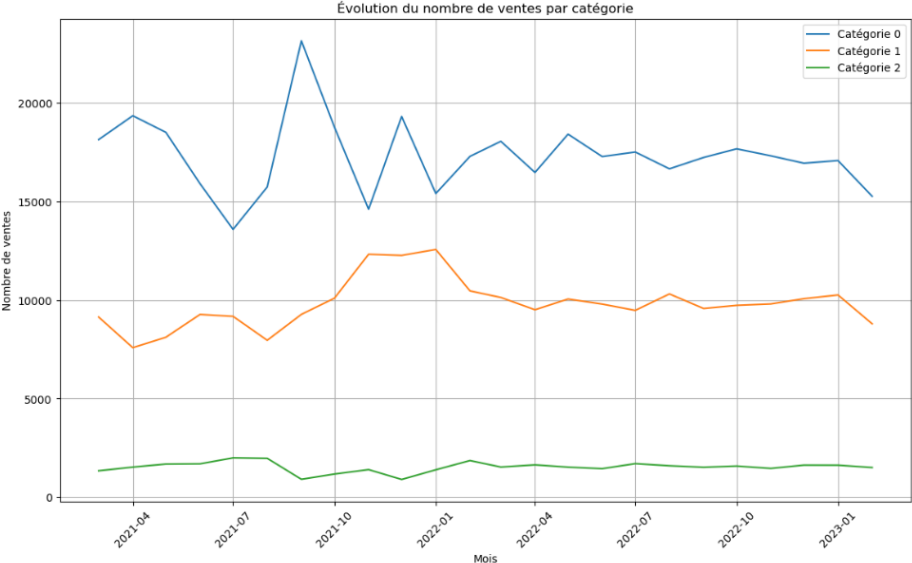


Top 10 Produits par Nombre de Ventes

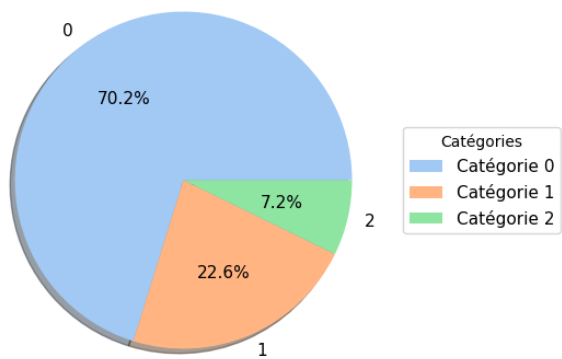


2 produits figurant dans les deux listes : 1_369 et 1_414

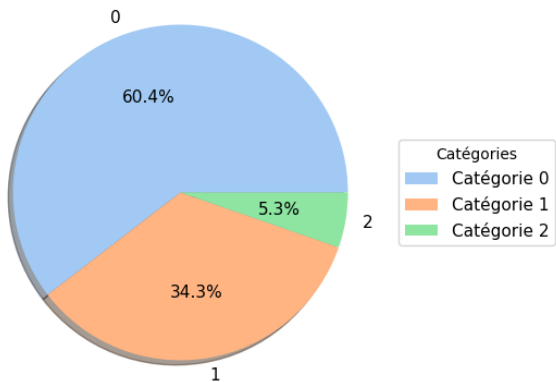
Répartition des ventes par catégorie



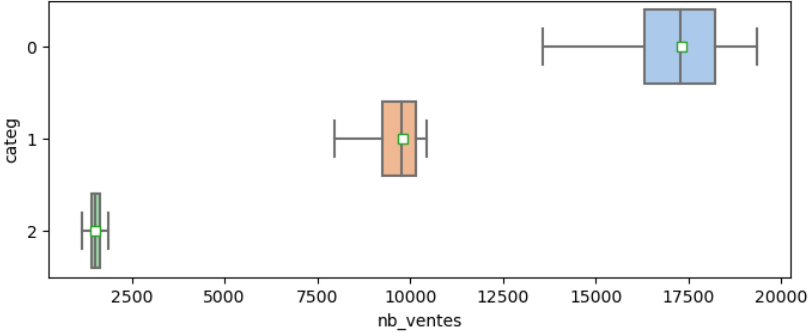
Répartition du Nombre de Produits par Catégorie



Répartition du Nombre de Produits Vendus par Catégorie



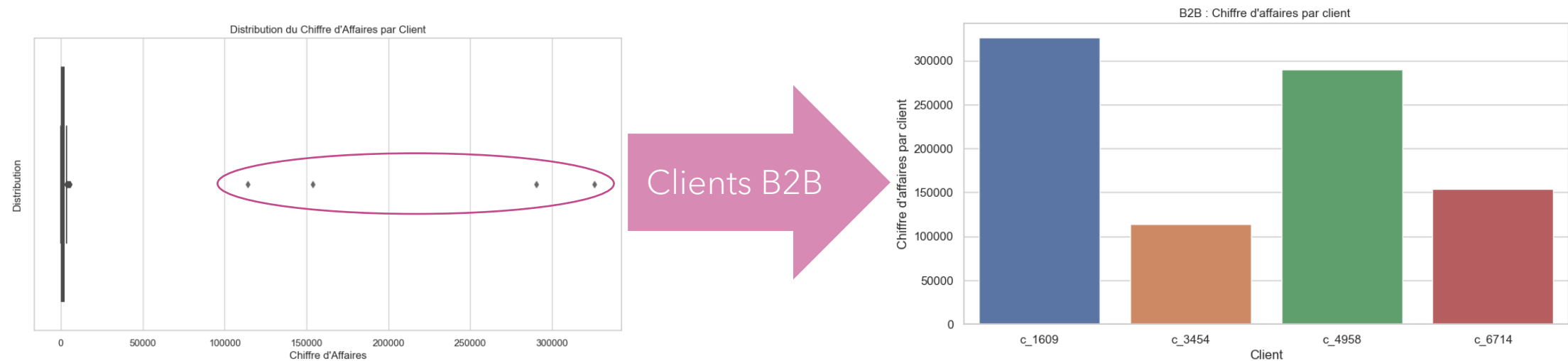
Distribution des ventes mensuelles par catégorie



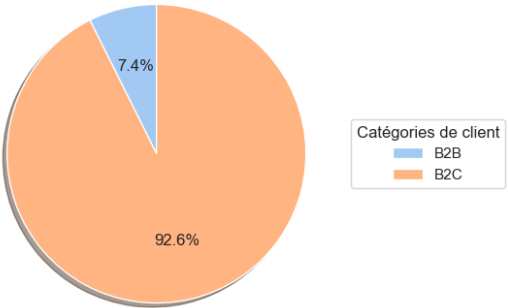


Profils clients

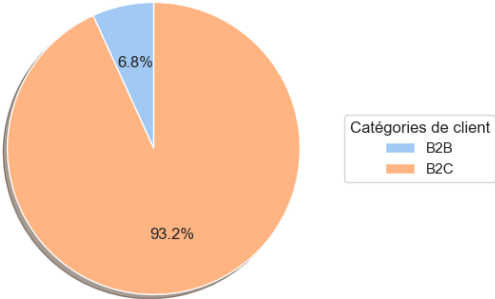
Répartition du chiffre d'affaires pour les clients B2B



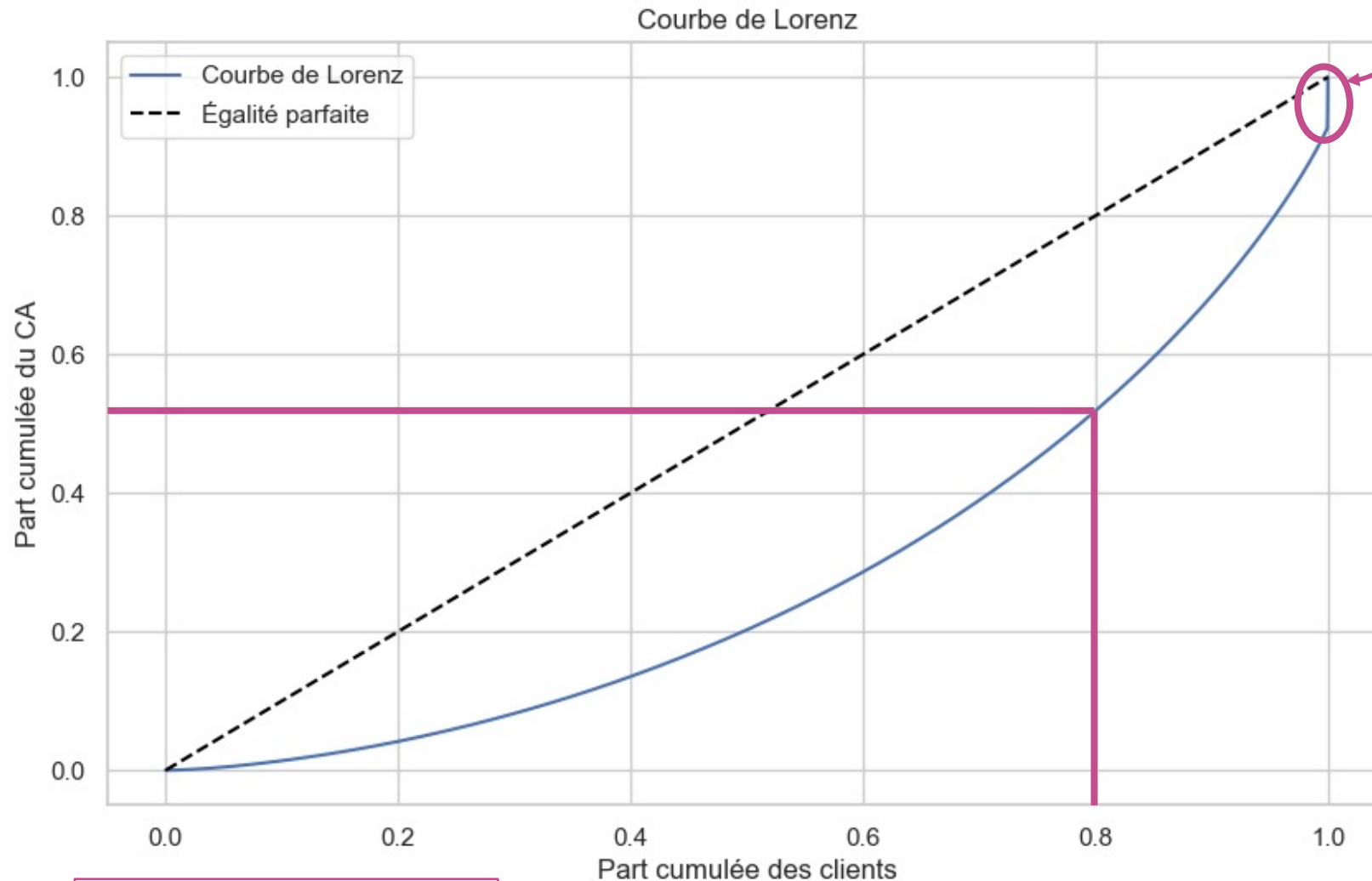
Répartition du Chiffre d'Affaires par Catégorie de Client



Répartition du Nombre de Ventes par Catégorie de Client



Courbe de Lorenz



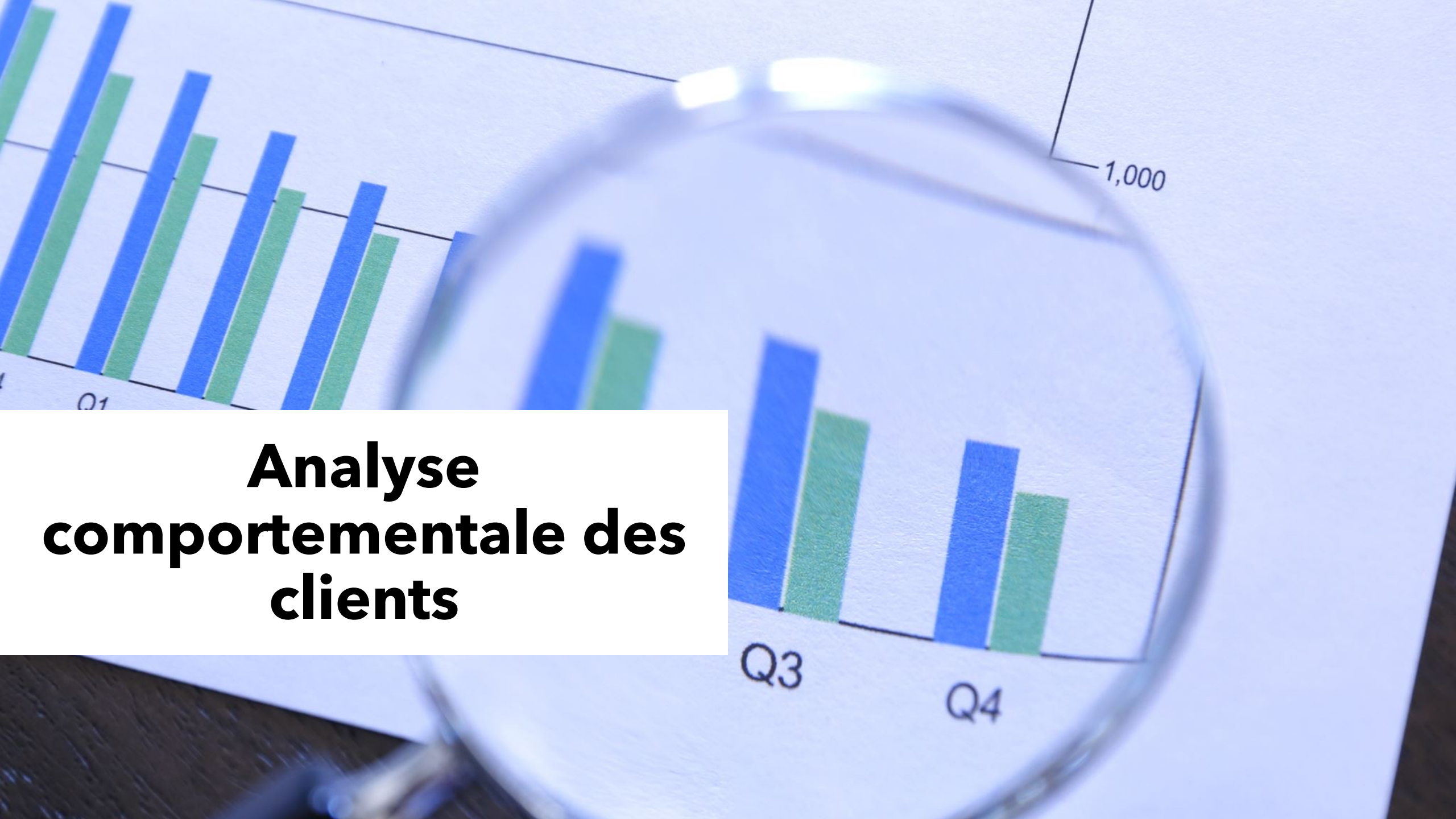
Clients B2B

Inégalité significative dans la répartition du chiffre d'affaires (CA) parmi les clients.

Une petite proportion de clients (B2B) est responsable d'une grande part du CA total.

Un indice de Gini de 0,44 confirme cette inégalité modérée dans la distribution du CA par clients.

Indice de Gini = 0,44



**Analyse
comportementale des
clients**

Lien entre le genre des clients et les catégories de livres achetés

Hypothèse:

Hypothèse nulle (H0) : Il n'y a pas de relation significative entre le genre des clients et les catégories de livres achetés.

Hypothèse alternative (H1) : Il existe une relation significative entre le genre des clients et les catégories de livres achetés.

Nous choisissons un niveau de confiance de 95 %, c'est-à-dire que nous rejetterons l'hypothèse nulle en faveur de l'alternative si la p-value est inférieure à 0,05.

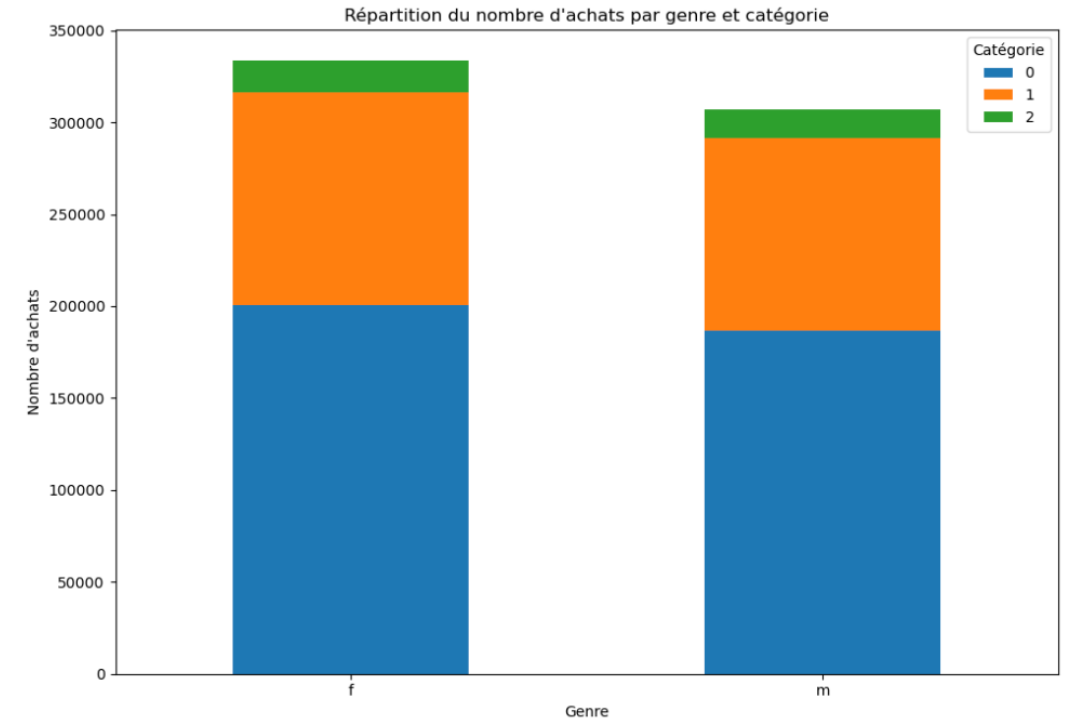
Il s'agit ici d'étudier la relation entre la variable "genre du client" et "catégorie des livres achetés" qui sont toutes deux des variables qualitatives. Nous effectuerons donc un test d'indépendance du Chi2.

Chi2 Statistic: 22.66856665178056
P-Value: 1.1955928116587024e-05
Degrees of Freedom: 2

Interprétation :

La p-valeur (1.1955928116587024e-05) est très petite, bien en dessous du seuil de 0.05. Cela signifie que nous rejetons l'hypothèse nulle selon laquelle le sexe des clients et les catégories des livres achetés sont indépendants.

Nous pouvons donc rejeter l'hypothèse nulle, il y a une association statistiquement significative entre le sexe des clients et les catégories des livres qu'ils achètent.



Lien entre l'âge des clients et le montant total des achats

Hypothèses

Hypothèse nulle (H0) : Il n'y a pas de relation significative entre l'âge des clients et le montant total des achats.

Hypothèse alternative (H1) : Il existe une relation significative entre l'âge des clients et le montant total des achats.

Nous choisissons un niveau de confiance de 95 %, c'est-à-dire que nous rejeterons l'hypothèse nulle en faveur de l'alternative si la p-value est inférieure à 0,05.

Nous commençons donc par représenter graphiquement la variable âge sous forme d'histogramme, puis nous utiliserons le test d'adéquation de Kolmogorov-Smirnov.

Test de Kolmogorov-Smirnov pour l'âge:

Statistique = 0.06399118988651292, P-Value = 4.802543174125307e-31

Test de Kolmogorov-Smirnov pour le montant total des achats:

Statistique = 0.10589907400752302, P-Value = 2.137067509720825e-84

Coefficient de corrélation de Spearman: -0.18453804793783096

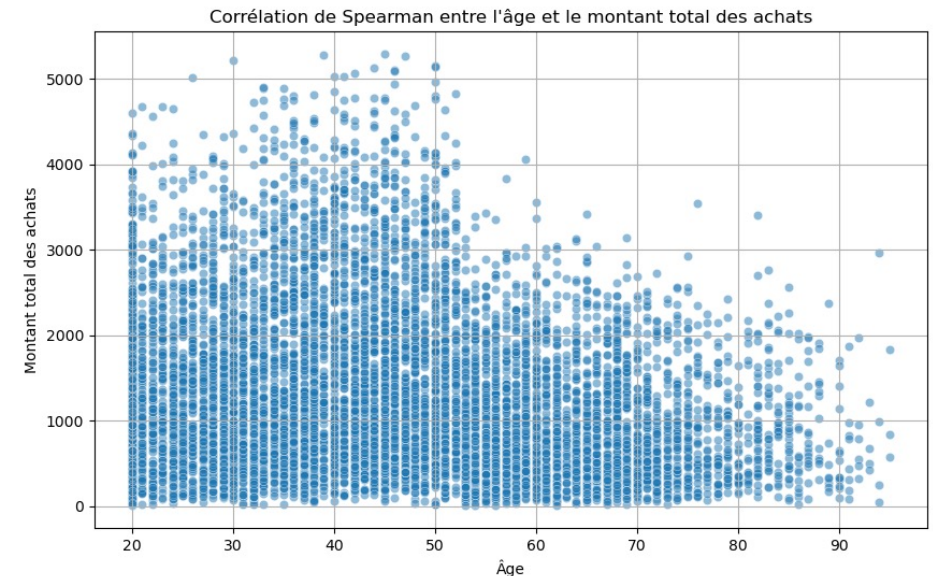
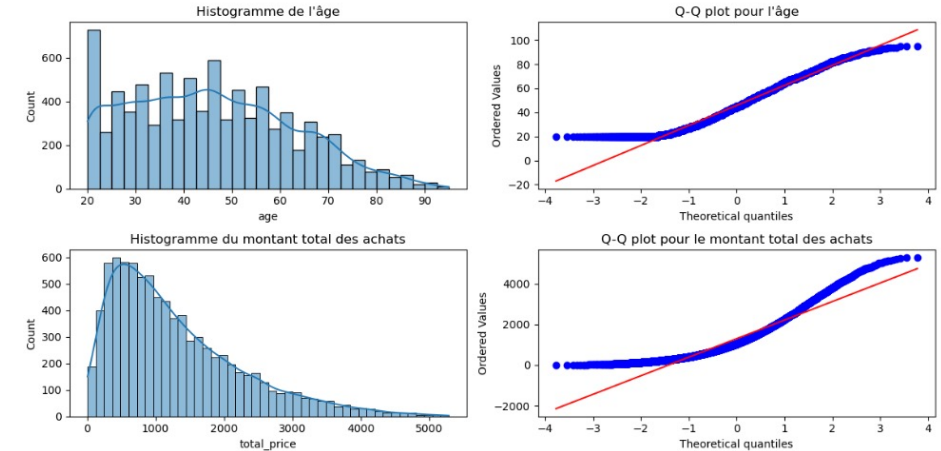
P-Value: 1.0212910436382683e-66

Interprétation:

Coefficient de corrélation de Spearman : La valeur de -0.1845 indique une faible corrélation négative entre l'âge des clients et le montant total de leurs achats. Cela signifie que, en général, à mesure que l'âge des clients augmente, le montant total de leurs achats a tendance à diminuer légèrement.

P-Value : La p-valeur est extrêmement faible (bien en dessous de 0.05), ce qui signifie que cette corrélation est statistiquement significative.

Nous pouvons donc rejeter l'hypothèse nulle et conclure qu'il existe une relation significative entre l'âge des clients et le montant total de leurs achats.



Lien entre l'âge des clients et la fréquence des achats

Hypothèses

Hypothèse nulle (H0) : Il n'y a pas de relation significative entre l'âge des clients et la fréquence d'achat.

Hypothèse alternative (H1) : Il existe une relation significative entre l'âge des clients et la fréquence d'achat.

Nous choisissons un niveau de confiance de 95 %, c'est-à-dire que nous rejeterons l'hypothèse nulle en faveur de l'alternative si la p-value est inférieure à 0,05.

Nous sommes à nouveau dans la situation où nous souhaitons analyser la relation entre 2 variables quantitatives. La distribution de la variable "âge" ne suivant pas une loi normale, nous pouvons à nouveau opter pour un test non paramétrique mais en utilisant la catégorisation par tranche d'âge, il est également possible d'appliquer une ANOVA.

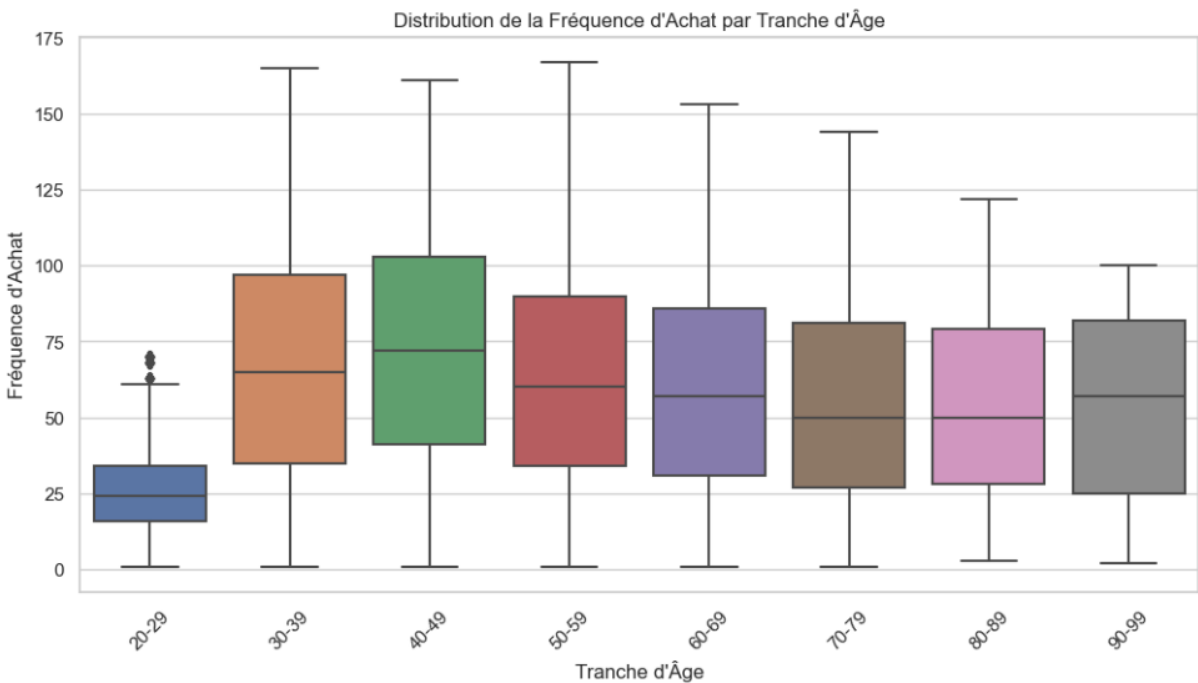
	Sum_sq	F	P-value
C (tranche_age)	1.166600e+08	13402.51851	0.0
Residuel	7.967277e+08	NaN	NaN

Interprétation:

F (Statistique F) : La valeur F de 13402.52 est extrêmement élevée, suggérant que les différences entre les groupes d'âge en termes de fréquence d'achat sont très marquées et statistiquement significatives.

PR(>F) (Valeur-p) : Une valeur-p de 0.0, il existe donc une relation très significative.

Les résultats confirment clairement l'hypothèse alternative (H1), c'est-à-dire qu'il existe une association significative entre les tranches d'âge des clients et la fréquence des achats.



Lien entre l'âge des clients et la taille du panier moyen

Hypothèses

Hypothèse nulle (H0) : Il n'y a pas de différence significative entre les tailles moyennes des paniers des différentes tranches d'âge.

Hypothèse alternative (H1) : Il existe une différence significative entre les tailles moyennes des paniers des différentes tranches d'âge.

Nous choisissons un niveau de confiance de 95 %, c'est-à-dire que nous rejeterons l'hypothèse nulle en faveur de l'alternative si la p-value est inférieure à 0,05.

Nous sommes à nouveau dans la situation où nous souhaitons analyser la relation entre 2 variables quantitatives. La distribution de la variable "âge" ne suivant pas une loi normale, nous optons à nouveau pour un test non paramétrique qui sera toujours le test de corrélation de Spearman.

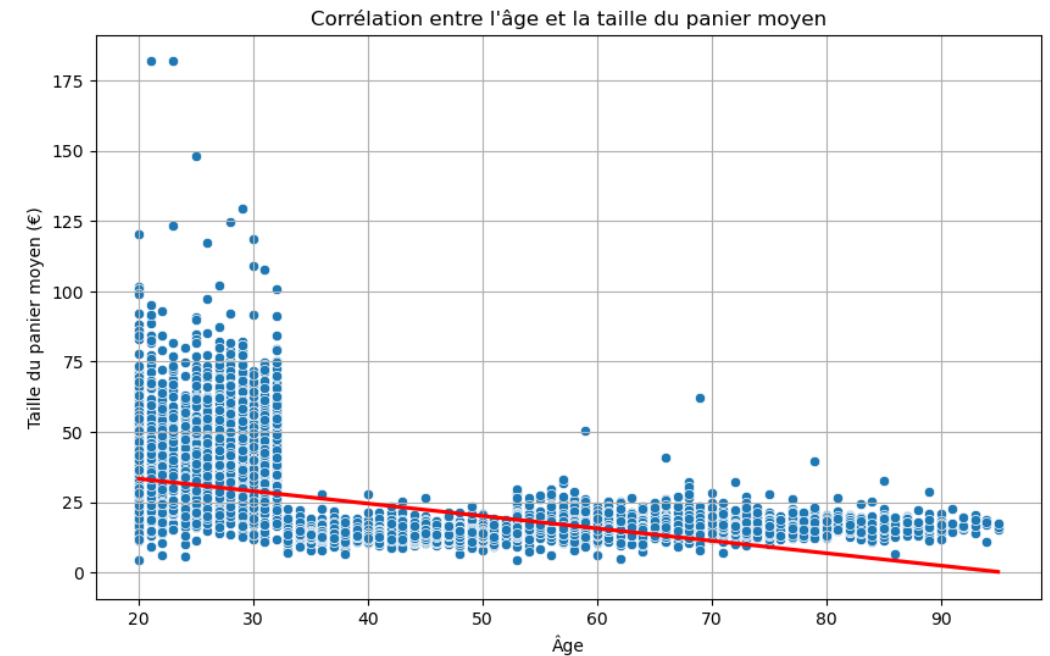
Corrélation de Spearman: -0.32586821258009735
P-value: 8.352869998649884e-212

Interprétation

Coefficient de Corrélation de Spearman : Le coefficient de corrélation de Spearman de -0.326 indique une corrélation négative modérée entre l'âge des clients et la taille moyenne du panier. Cela signifie que, dans l'ensemble, à mesure que l'âge des clients augmente, la taille moyenne de leur panier a tendance à diminuer.

P-Value : La p-value extrêmement faible indique que la corrélation observée est statistiquement significative. En d'autres termes, il est très peu probable que cette corrélation soit due au hasard.

Nous pouvons donc rejeter l'hypothèse nulle, les résultats indiquent que la taille moyenne des paniers varie de manière significative en fonction de la tranche d'âge des clients.



Lien entre l'âge des clients et la catégorie des livres achetés

Hypothèses

Hypothèse nulle (H0) : Il n'y a pas d'association significative entre les tranches d'âge des clients et les catégories de livres achetés.

Hypothèse alternative (H1) : Il existe une association significative entre les tranches d'âge des clients et les catégories de livres achetés.

Nous choisissons un niveau de confiance de 95 %, c'est-à-dire que nous rejeterons l'hypothèse nulle en faveur de l'alternative si la p-value est inférieure à 0,05.

Nous sommes ici dans la situation où nous souhaitons analyser la corrélation entre 1 variable quantitative (âge) et 1 variable qualitative à 3 modalités (catégories 0,1,2). Nous effectuerons donc un test d'indépendance du Chi2.

Chi2 Statistic: 209618.73091797682
P-Value: 0.0

Interprétation

Chi2 Statistic : La valeur élevée de la statistique chi2 indique qu'il existe une variance significative entre les groupes de tranches d'âge par rapport à la variance au sein des groupes.

P-Value : La p-valeur est extrêmement faible (0.0), ce qui signifie que nous pouvons rejeter l'hypothèse nulle. Cela indique qu'il existe une association significative entre les tranches d'âge et les catégories de livres achetés.

Nous pouvons donc rejeter l'hypothèse nulle , les résultats indiquent que les tranches d'âge des clients et les catégories de livres qu'ils achètent sont statistiquement associées.

