



SOMMAIRE

PARTIE 1. Nettoyage des données

- 1.1 Analyse des tables
- 1.2 Traitement des données

PARTIE 2. Analyse des données

(demandée par Annabelle, responsable Marketing)

- 2.1 Évolution du Chiffre d'Affaires avec Moyenne Mobile
- 2.2 Chiffre d'affaires et répartition par catégories
- 2.3 Les tops et les flops produits
- 2.4 Nombre de clients et nombre de Transactions par mois
- 2.5 Nombre de produits vendus et nombre de transactions par mois
- 2.6 Courbe de Lorenz
- 2.7 Répartition du CA pour les clients BtoB

PARTIE 3. Les corrélations (Clients BtoC)

(demandées par Julie Business Intelligence Analyst)

- 3.1 Lien entre le Genre et les Catégories de Livres Achetés
- 3.2 Lien entre l'âge des Clients et le Montant Total des Achats
- 3.3 Lien entre l'âge des Clients et la Fréquence d'Achat
- 3.4 Lien entre l'âge des Clients et la Taille du Panier Moyen
- 3.5 Lien entre l'âge des Clients et la Catégorie des Livres Achetés

PARTIE 1. Nettoyage des données

1.1 Analyse des tables

3 FICHIERS CSV

customers.csv
products.csv
transactions.csv



Importer dans collaboratory



Nettoyage des valeurs "NaN" + Doublons (361 041 exemplaires)



Activité en 2 ans

8621 clients

3286 produits

687 534 Transactions

PARTIE 1. Nettoyage des données

1.2 Traitement des données

Clé primaire

id_prod



Jointure 1 : transactions_df et products_df = merged_df

Clé primaire

client_id



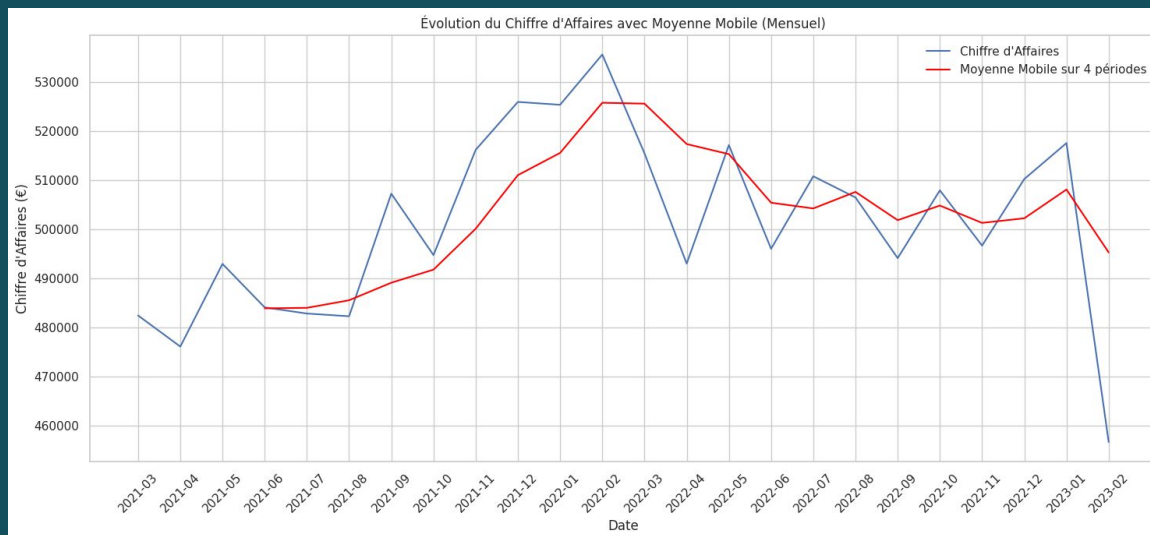
Jointure 2 : customers_df et merged_df



DATAFRAME FINAL : full_df

PARTIE 2. Analyse des données

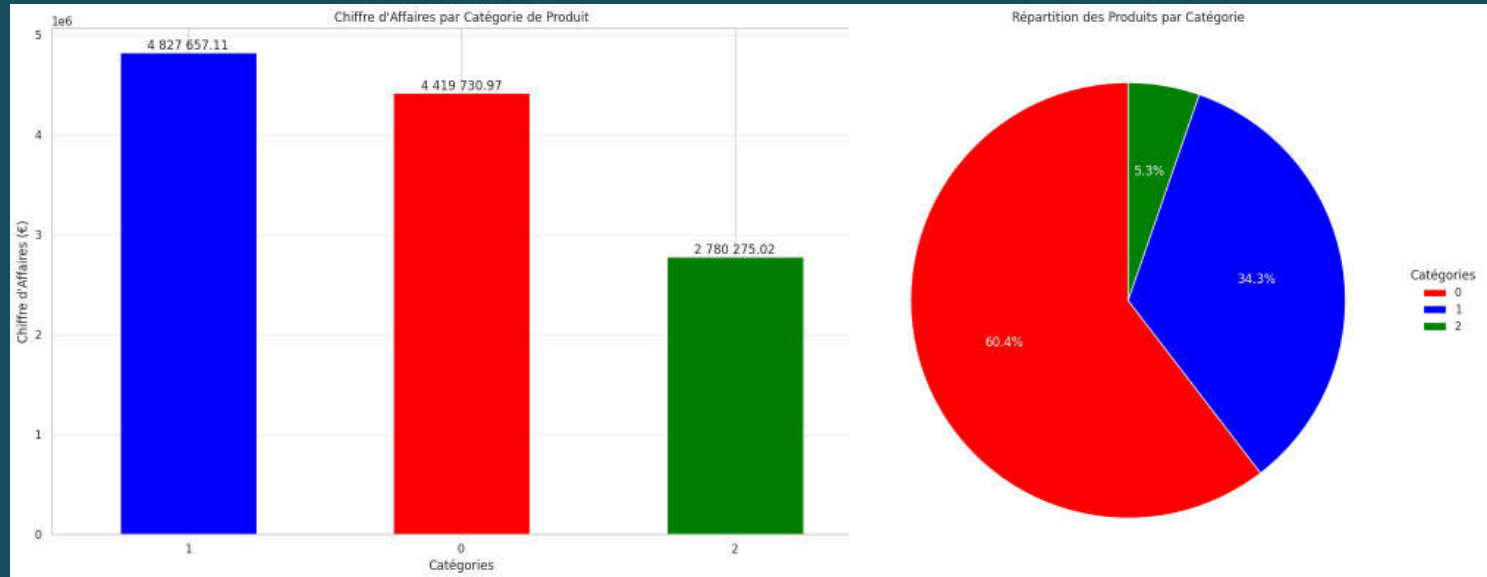
2.1 Évolution du Chiffre d'Affaires avec Moyenne Mobile



Le CA montre une croissance notable entre mars 2021 et son pic en Février 2022 puis une baisse constante durant cette même année 2022, suivie d'une chute abrupte inexplicable début 2023.

PARTIE 2. Analyse des données

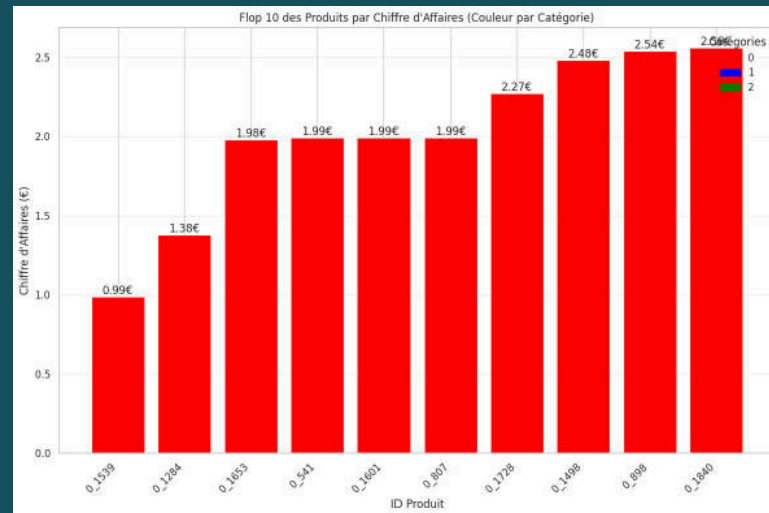
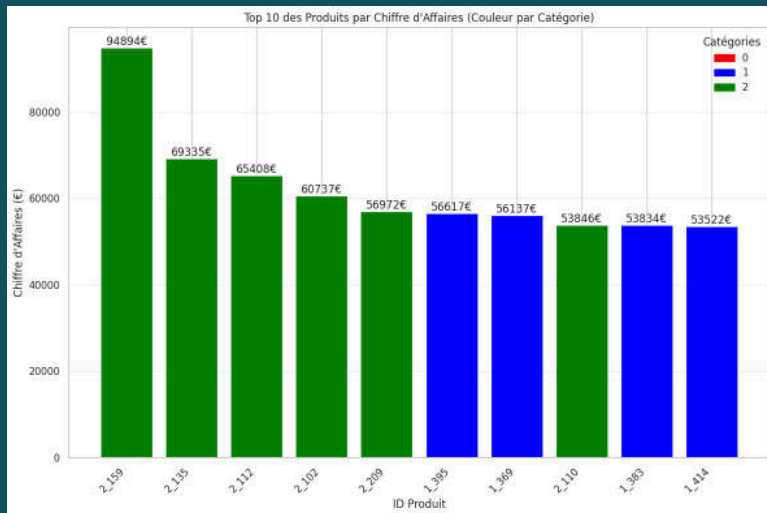
2.2 Chiffre d'affaires et répartition par catégories



Les ventes les plus importantes sont réalisées par la catégorie "1". Arrive ensuite la catégorie 0 et pour finir la catégorie 2. Par contre le stock de livre le plus important n'est pas celui qui se vend le plus.

PARTIE 2. Analyse des données

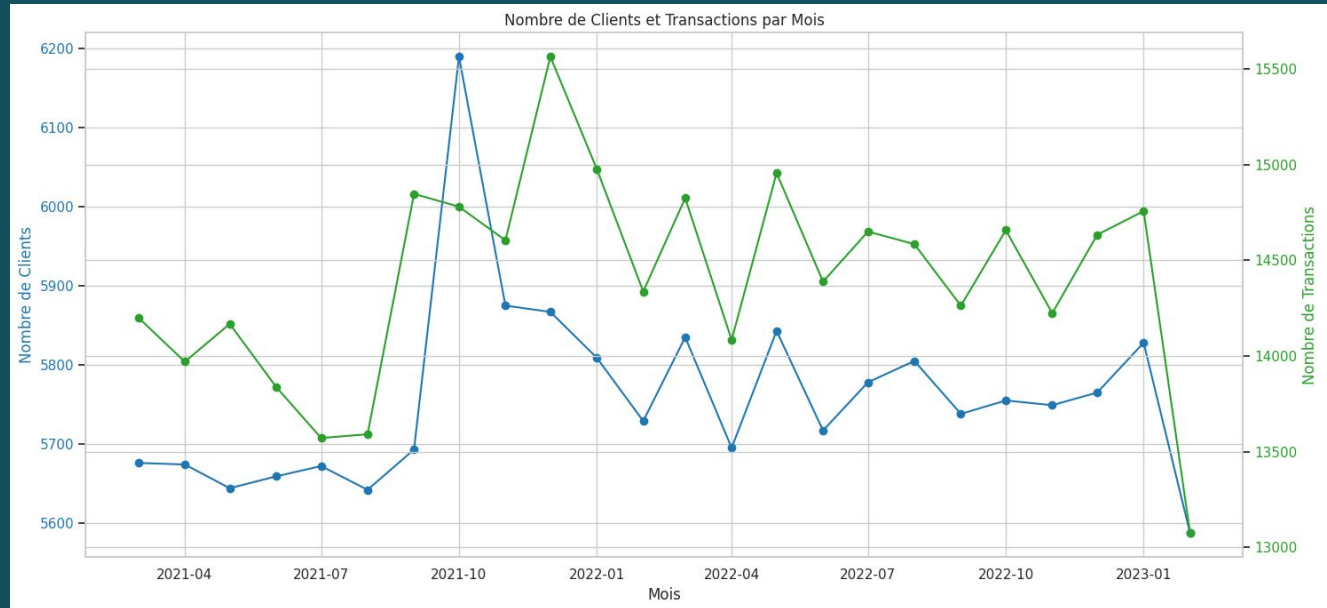
2.3 Les tops et les flops produits



On peut remarquer que le top flop est essentiellement constitué de livre de la catégorie 0 ... La plus importante en stock (à revoir)
Le top 10 est essentiellement constitué de catégorie 1 (40%) et 2 (60%)
(pourtant la catégorie 2 est la moins importante en stock)

PARTIE 2. Analyse des données

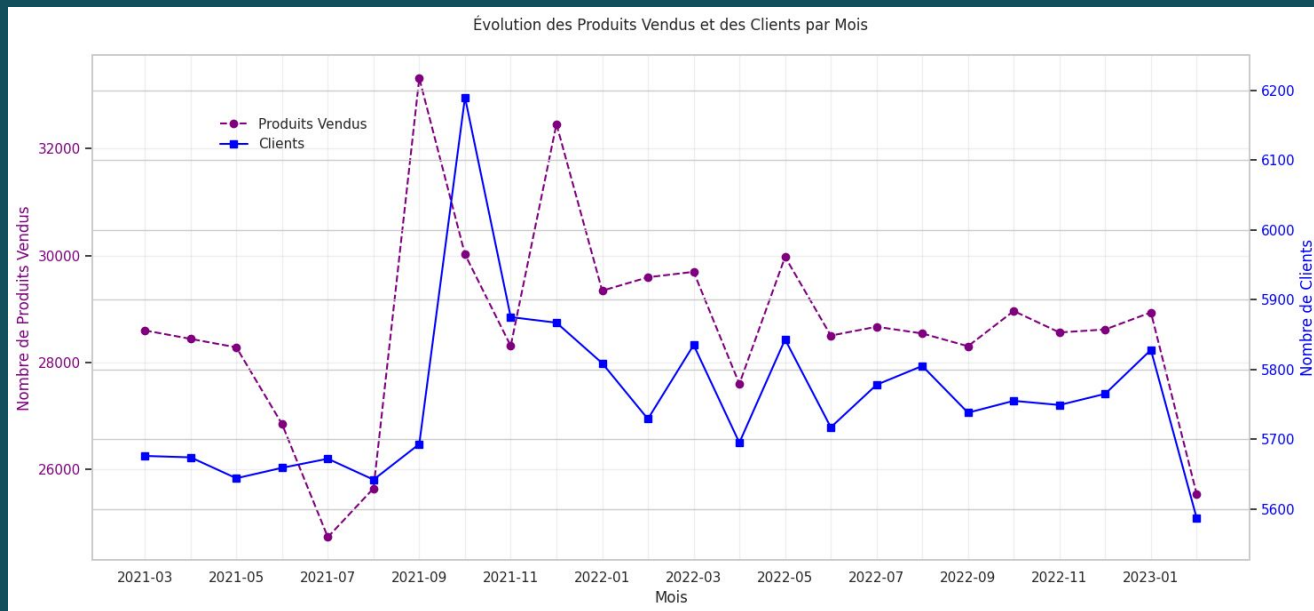
2.4 Nombre de clients et nombre de Transactions par mois



Corrélation évidente : Plus de clients = plus de transactions
Incompréhension en octobre 2021 (campagne sea, soucis ciblage)
et chute inexplicable en janvier 2023 (soucis serveur, fin de campagne SEA, nécessité de renouvellement gamme produit ...)

PARTIE 2. Analyse des données

2.5 Nombre de produits vendus et nombre de transactions par mois



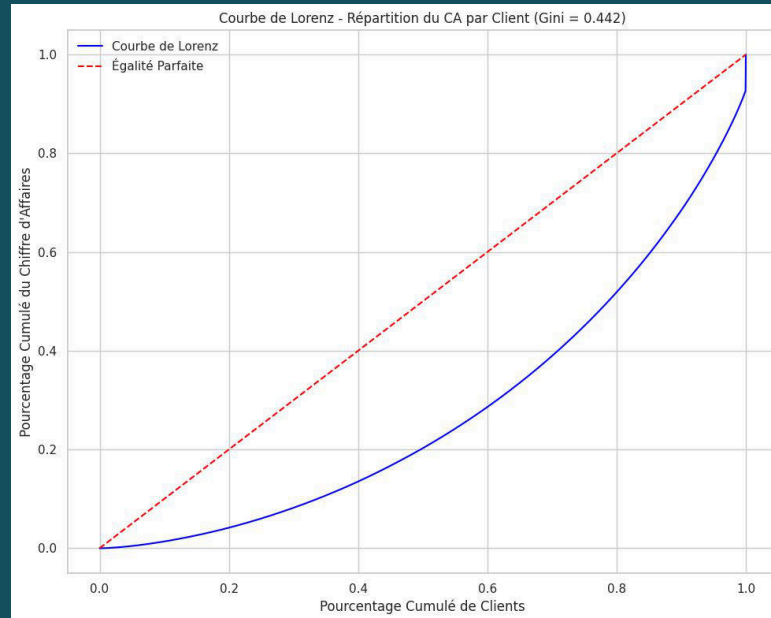
Corrélation évidente : Plus de clients = plus de produits

Septembre 2021 : opération de liquidation de stock ou de promotion (moins de transactions, plus de produits, moins de clients)

Chute inexplicable en janvier 2023 (soudis serveur, fin de campagne SEA, nécessité de renouvellement gamme produit ...)

PARTIE 2. Analyse des données

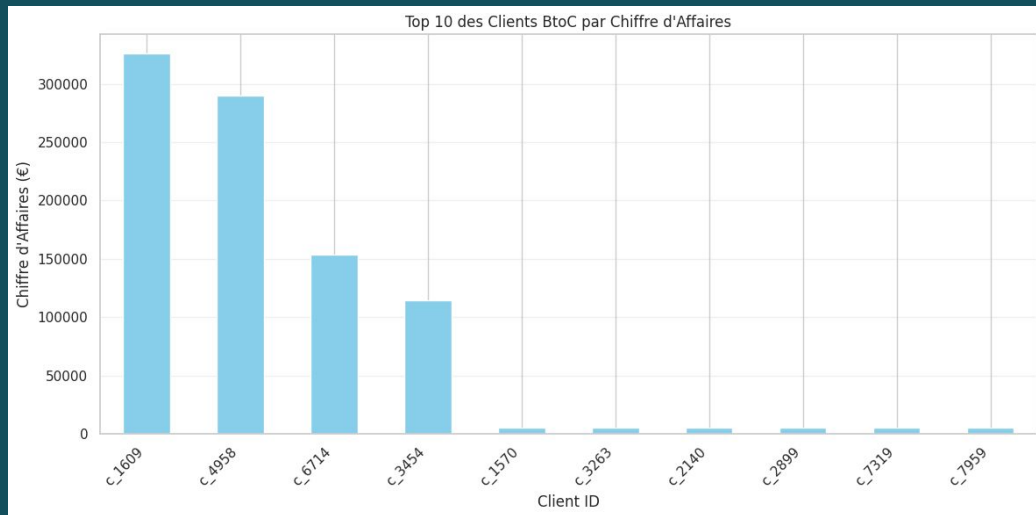
2.6 Courbe de Lorenz



La courbe de Lorenz montre que le CA du site web est dépendant d'un petit nombre d'acheteur qui représentent une part importante du CA.
50 % du CA est apporté par 80 % des clients

PARTIE 2. Analyse des données

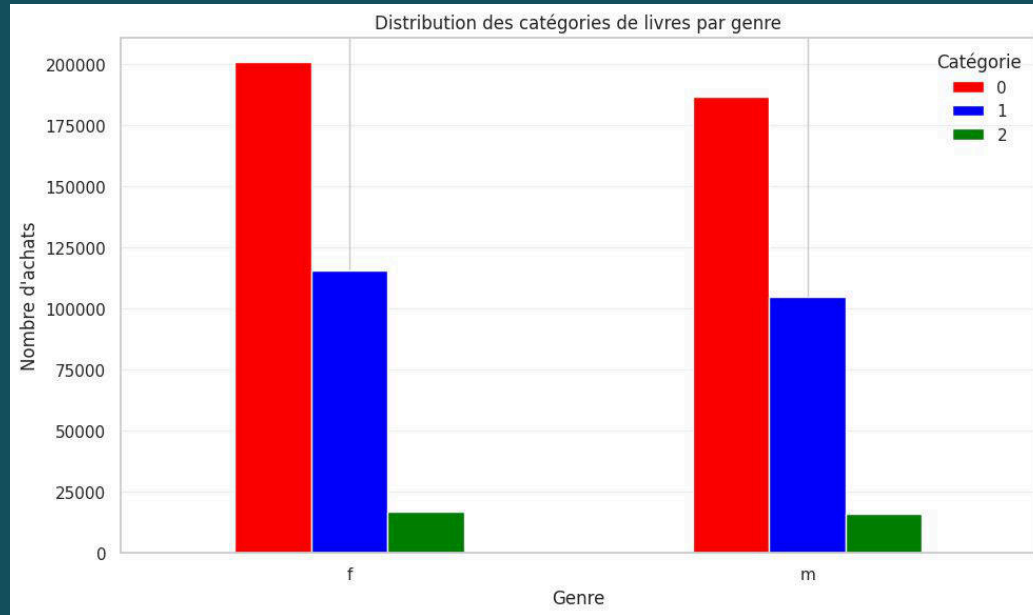
2.7 Répartition du CA pour les clients BtoB



4 CLIENTS BtoB = 7,35 % du chiffre d'affaires

PARTIE 3. Les corrélations (Clients BtoC)

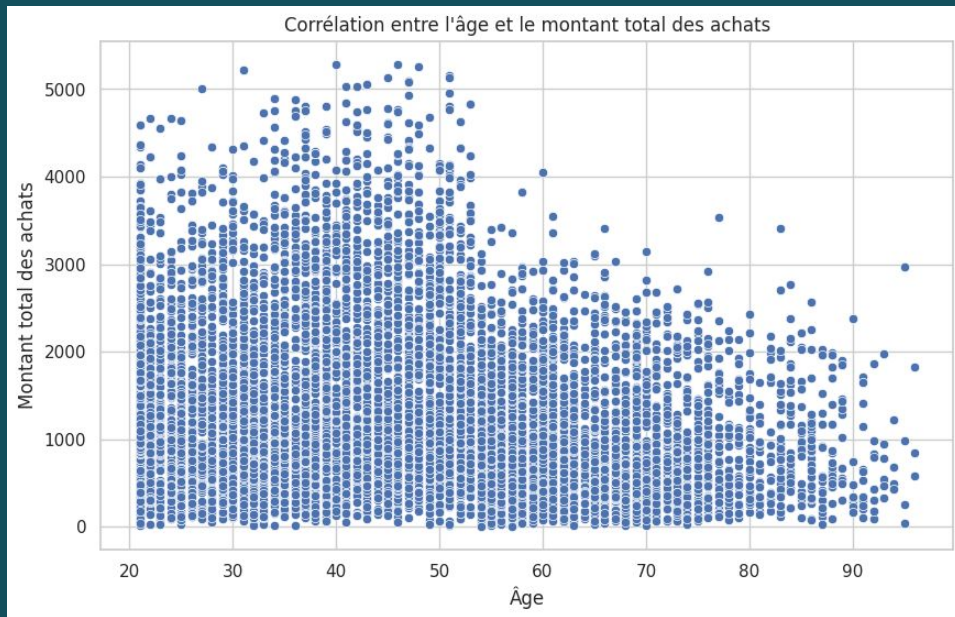
3.1 Lien entre le Genre et les Catégories de Livres Achetés



Test khi-deux d'indépendance : entre les variables qualitatives "genre" et catégorie de livre
Valeur : 22,6 (La valeur élevée du Chi-deux indique que le genre a une grande importance sur la catégorie de livre achetée)

PARTIE 3. Les corrélations (Clients BtoC)

3.2 Lien entre l'âge des Clients et le Montant Total des Achats

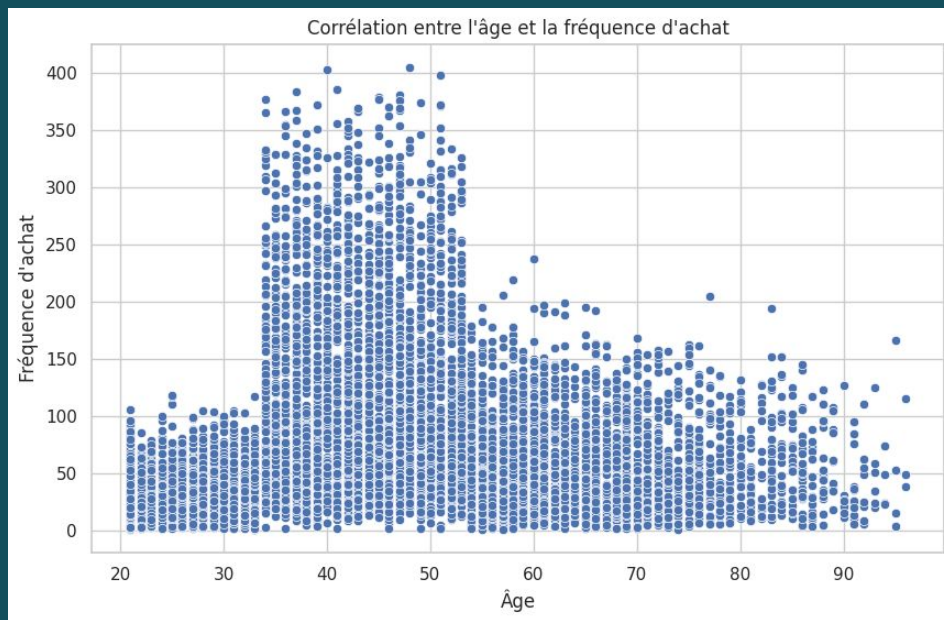


TEST DE CORRELATION DE PEARSON : entre les variables quantitatives "age" et "montant total des achats"

Valeur : -0.18 (La corrélation négative de "-0,18" suggère qu'il existe une légère tendance pour les clients plus âgés à dépenser un peu moins au total que les clients plus jeunes, et vice versa)

PARTIE 3. Les corrélations (Clients BtoC)

3.3 Lien entre l'âge des Clients et la Fréquence d'Achat

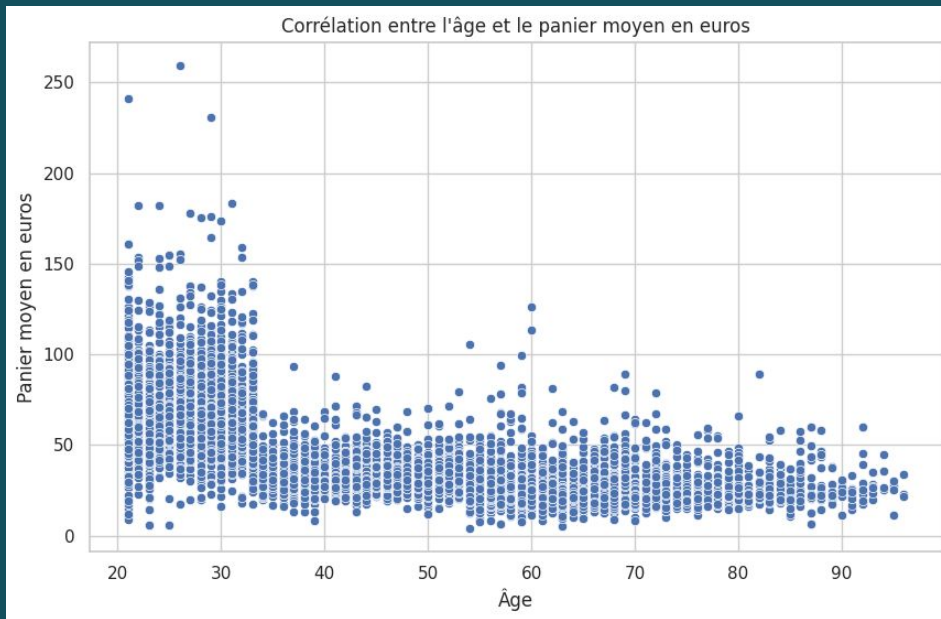


TEST DE CORRELATION DE PEARSON : entre les variables quantitatives "age" et "fréquence d'achats"

Valeur : +0,16 (La corrélation est de " + 0,1645" : Positive, cela signifie que lorsque l'âge tend à augmenter, la fréquence d'achat tend légèrement à augmenter également, et inversement. Mais elle est aussi faible est donc cela signifie que la relation linéaire entre l'âge et la fréquence d'achat n'est pas très forte.)

PARTIE 3. Les corrélations (Clients BtoC)

3.4 Lien entre l'âge des Clients et la Taille du Panier Moyen

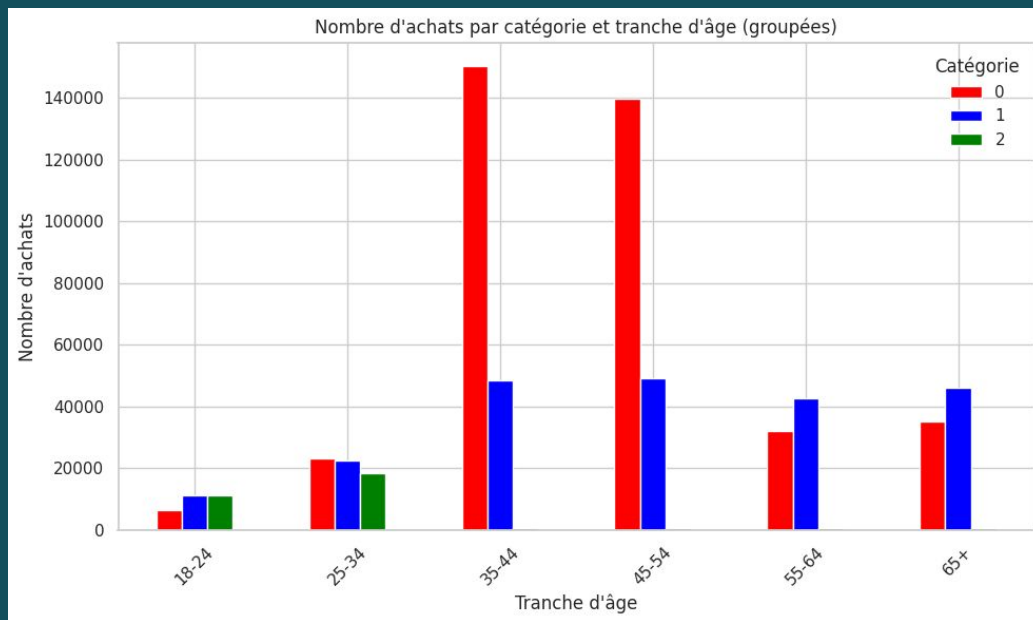


TEST DE CORRELATION DE PEARSON : entre les variables quantitatives "age" et "panier moyen"

Valeur : -0,61 (la corrélation négative indique la tendance que plus l'âge augmente, plus le panier moyen a tendance à diminuer. les clients plus jeunes ont tendance à avoir des paniers moyens plus élevés et plus variables, tandis que les clients plus âgés ont tendance à avoir des paniers moyens plus faibles et moins variables.)

PARTIE 3. Les corrélations (Clients BtoC)

3.5 Lien entre l'âge des Clients et la Catégorie des Livres Achetés



TEST DE CORRELATION DE Kruskal-Wallis : entre la variable quantitative "age" et la variable qualitative "la catégorie des livres"
valeur Statistique de Kruskal-Wallis: 108939.00. Elle indique une différence significative dans la distribution des catégories de livres achetées entre les tranches d'âge. En d'autres termes, l'âge des clients semble être lié à la catégorie des livres qu'ils achètent.



CONCLUSION

CA 12 Millions
2 ANS

7,35 % CA
BtoB (4)

92,65 % CA
BtoC

50 % du CA est apporté par 80 % des clients

Test de corrélation de Pearson, du Khi-deux d'indépendance et de Kruskal-Wallis

- Le genre a une grande importance sur la catégorie de livre achetée.
- L'âge des clients semble être lié à la catégorie des livres qu'ils achètent.
- Plus l'âge augmente, plus le panier moyen a tendance à diminuer.
- Les clients plus âgés dépensent un peu moins au total que les clients plus jeunes.
- Lorsque l'âge tend à augmenter, la fréquence d'achat tend légèrement à augmenter également.