

Analysez les ventes d'une librairie avec Python

Rémi Bardey - Projet 9 - septembre 2025

Contexte

Lapage : librairie historique, activité en ligne lancée depuis 2 ans

Demande de la Direction commerciale : faire un bilan de l'activité e-commerce

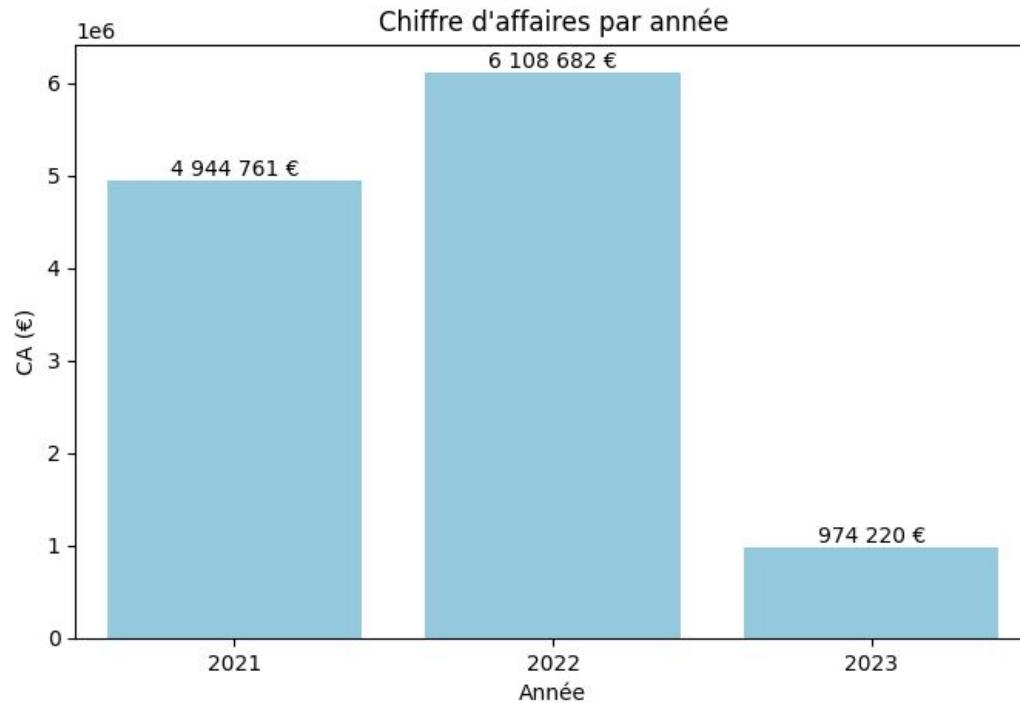
Objectifs de l'analyse :

- Évaluer la performance des ventes
- Comprendre le profil et le comportement des clients
- Identifier des leviers d'action pour la croissance

Destinataire : CODIR – présentation synthétique et orientée décision

Data : 3 fichiers(Transactions, produits et customers)

Chiffre d'affaires (2021-2023)



CA total (3 ans) : 12,0 M€

Évolution annuelle :

- 2021 : 4,9 M€ (mars à décembre)
- 2022 : 6,1 M€ (+24% vs 2021)
- 2023 : 0,97 M€ (année incomplète)

Évolution du CA (2021-2023)

Tendance générale :

2021 (mars-décembre) :

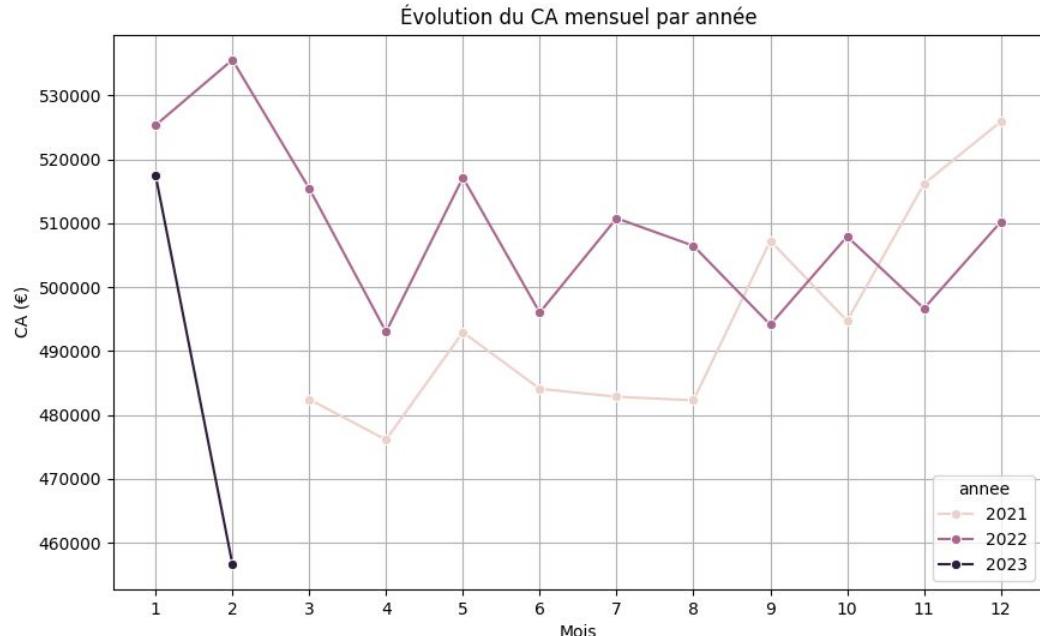
- Pic en mai, puis plateau jusqu'en août
- Rebond en septembre
- Progression continue de septembre à décembre

2022 (année complète) :

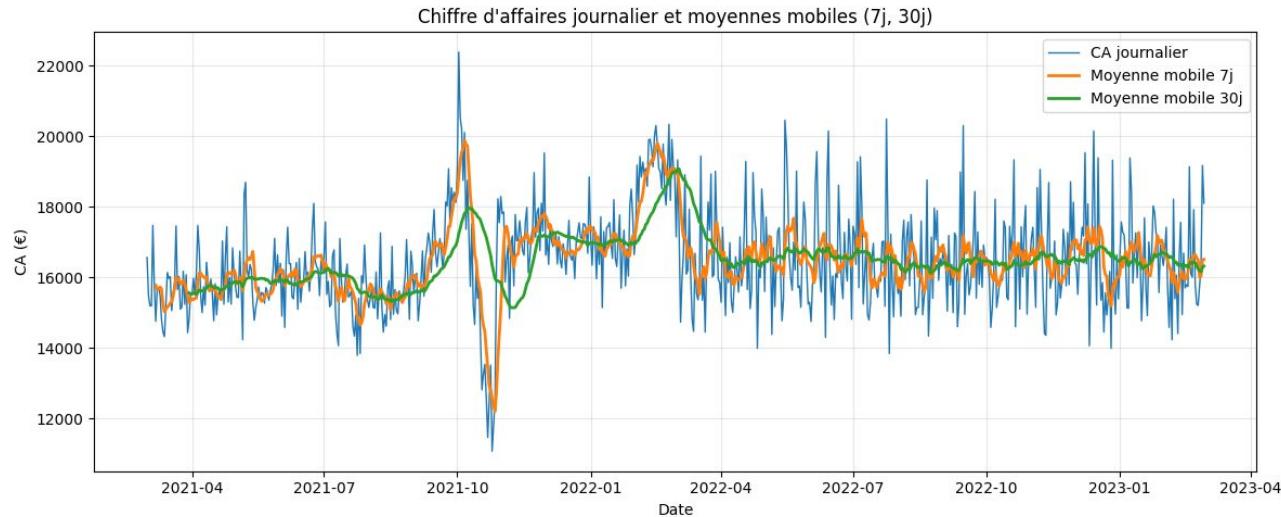
- Fort démarrage en janvier-février
- Chute jusqu'en avril
- Alternance de pics et creux sur le reste de l'année
- Décembre en retrait par rapport à 2021

2023 (janvier-février) :

- Bon démarrage en janvier
- Forte chute en février → tendance à surveiller



CA journalier et moyenne glissante 7 & 30 jours



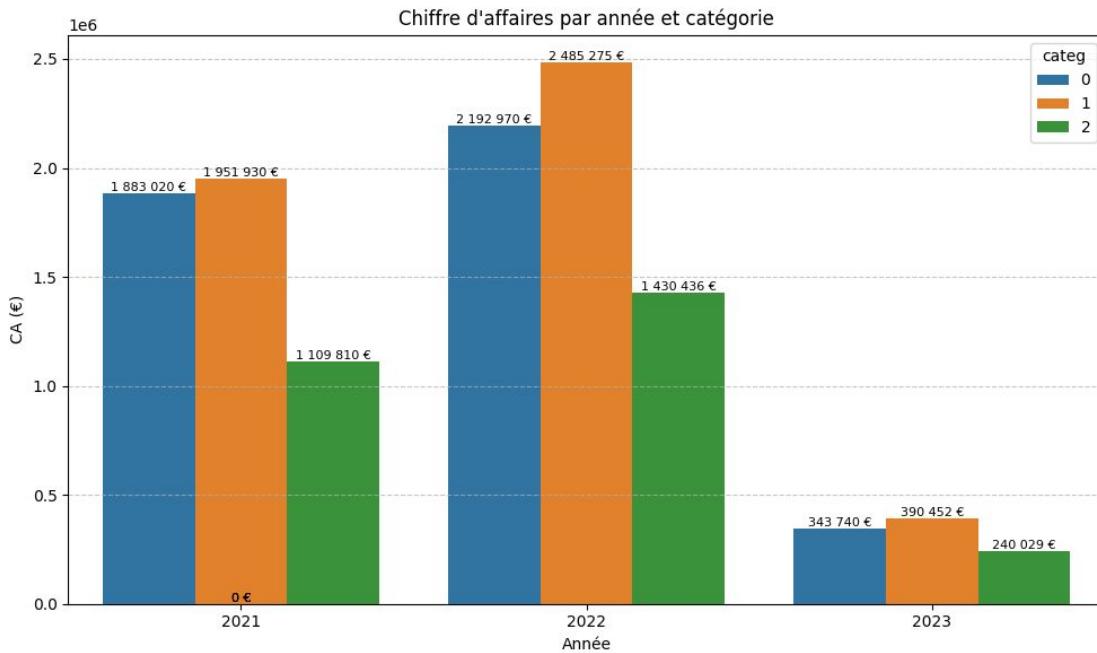
Deux pics majeurs :

- Fin 2021 : forte hausse soudaine du CA
- Début 2022 : deuxième pic significatif
- Un creux marqué : fin 2021, immédiatement après le premier pic

Plateau plus stable ensuite :

- Le CA se stabilise autour d'un niveau régulier
- Variations toujours présentes, mais moins fortes que les pics initiaux
- Indique une activité plus régulière et moins dépendante d'événements exceptionnels

CA par catégorie (2021-2023)



Hiérarchie stable :

- Catégorie 1 en tête chaque année
- Catégorie 0 systématiquement en 2^e position
- Catégorie 2 toujours derrière, mais en progression régulière

Croissance 2021 → 2022 : les trois catégories augmentent, confirmant une dynamique positive avant le repli de 2023.

Structure équilibrée : bien que la catégorie 1 domine, les trois segments contribuent de façon significative au CA.

Distribution des prix par catégorie

Catégorie 0 (entrée de gamme)

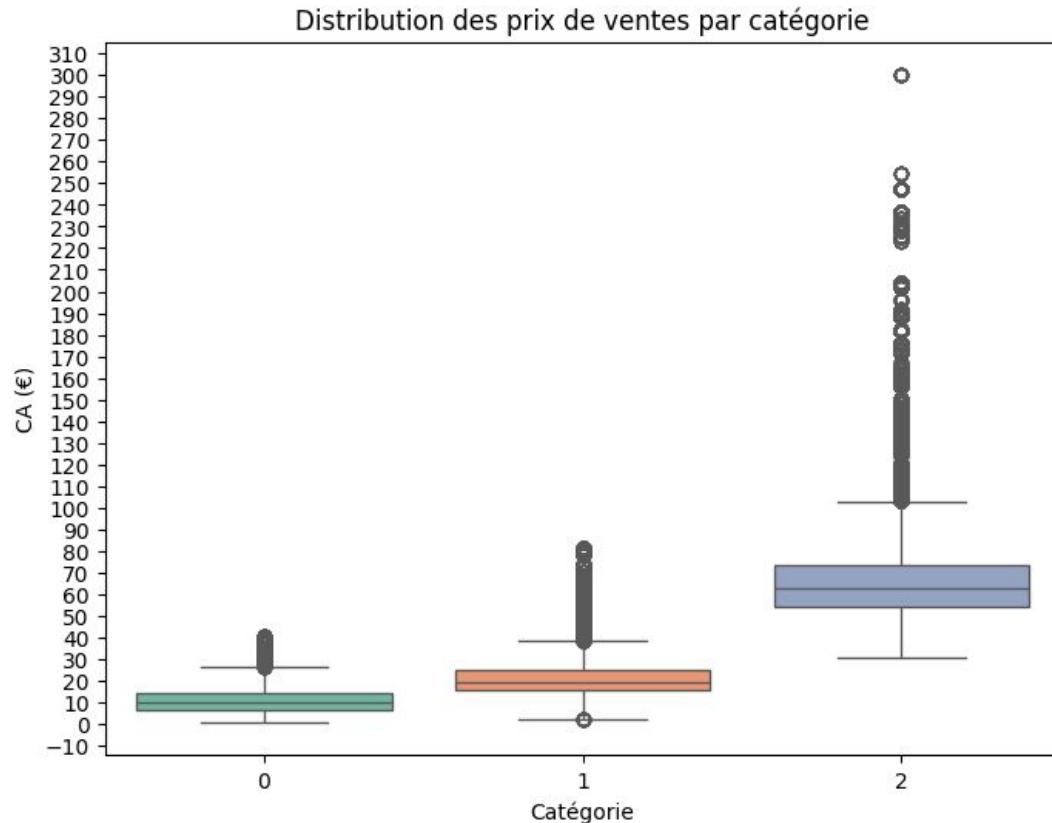
- Prix majoritairement entre 5 et 15 €
- Dispersion faible, produits homogènes

Catégorie 1 (milieu de gamme)

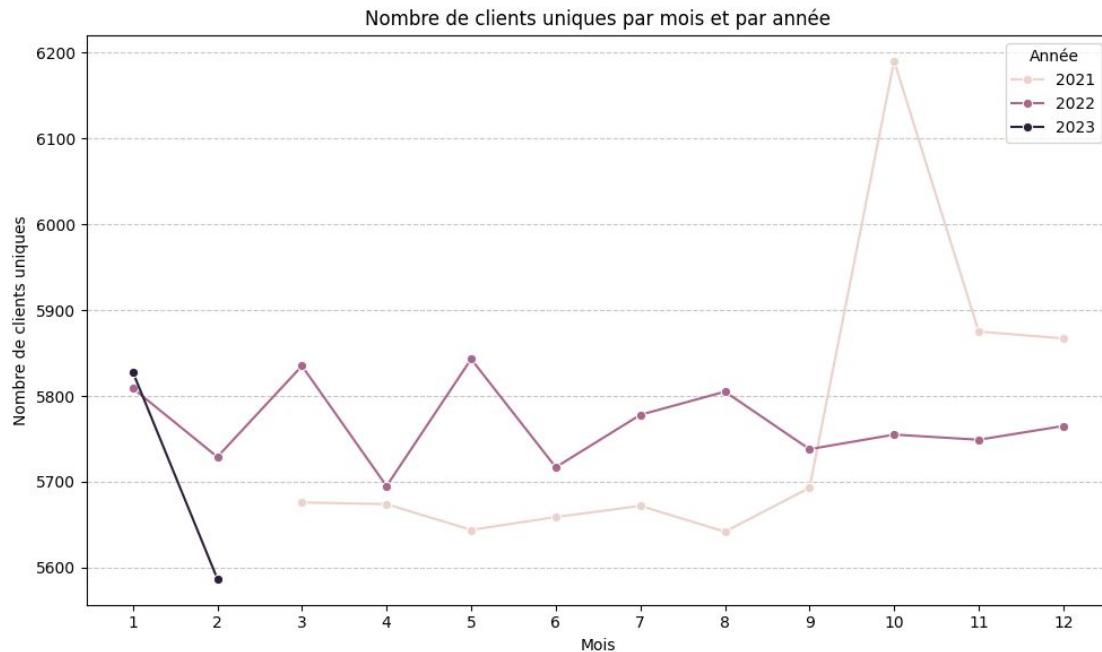
- Prix centrés autour de 20 €
- Dispersion plus large (\approx 10 à 30 €)
- Quelques produits plus chers (jusqu'à 80 €)

Catégorie 2 (haut de gamme)

- Prix en moyenne autour de 60 €
- Variabilité forte : de 30 € à plus de 100 € pour les produits principaux
- Nombreux produits premium (outliers) jusqu'à 300 €



Nombre de clients (2021-2023)



2021 : progression nette en fin d'année avec un pic marqué en octobre (près de 6 200 clients).

2022 : stabilité autour de 5 700-5 800 clients par mois, sans forte croissance.

2023 (janv-fév) : baisse du nombre de clients uniques par rapport au début 2022.

Points clés :

- La croissance clients s'est ralenti en 2022 → base fidèle mais pas d'expansion marquée.
- Le pic exceptionnel de fin 2021 pourrait être lié à une campagne ou une saisonnalité forte (ex. fêtes de fin d'année).
- 2023 démarre sur une tendance baissière, à surveiller.

Nombre de transactions (2021-2023)

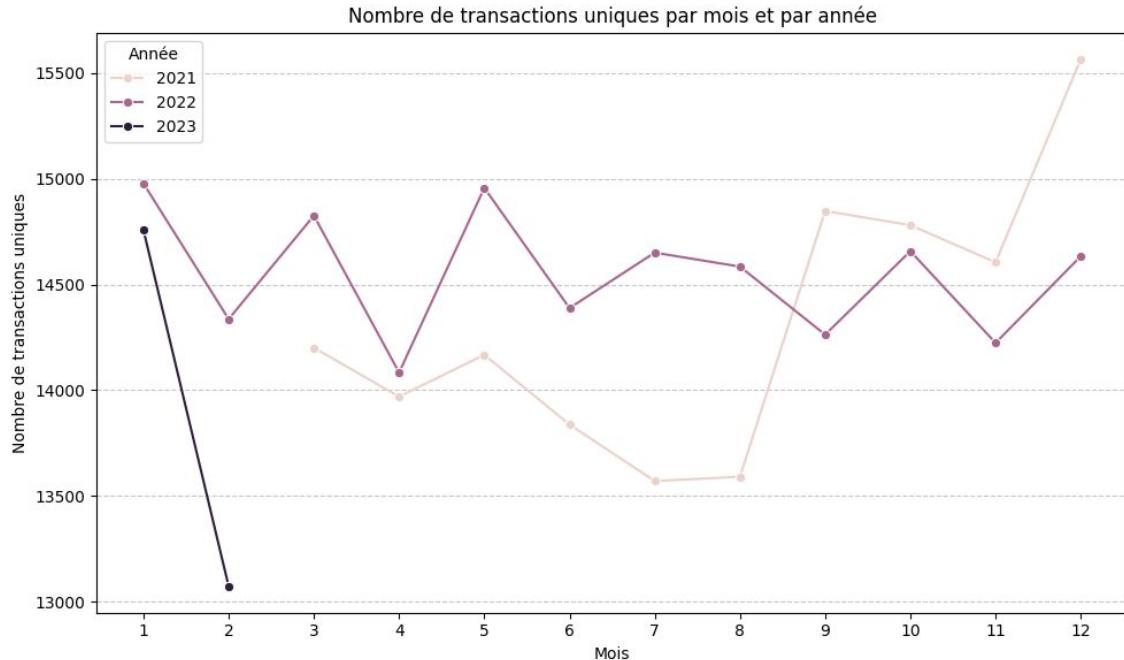
2021 : volume en progression nette en fin d'année (pic en décembre à plus de 15 500 transactions).

2022 : stabilité autour de 14 500 transactions/mois, sans forte variation.

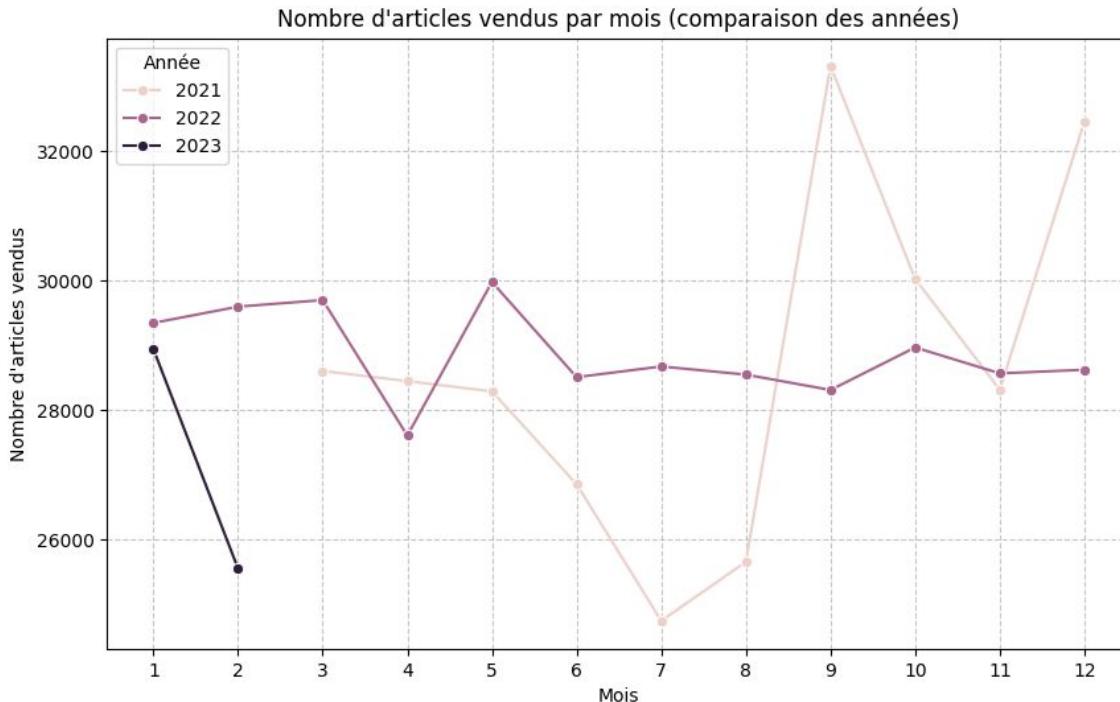
2023 (janv-fév) : baisse importante, en ligne avec la diminution du nombre de clients.

Points clés :

- La dynamique de fin 2021 ne s'est pas reproduite en 2022 → activité plus régulière mais sans croissance marquée.
- Transactions et clients suivent la même tendance : pic fin 2021, stabilité en 2022, baisse en 2023.



Nombre d'articles vendus (2021-2023)



2021 : forte progression en fin d'année avec un pic exceptionnel en septembre (plus de 33 000 articles vendus).

2022 : volume globalement plus régulier (~28 500-30 000 articles/mois), sans pics marqués.

2023 (janv-fév) : recul par rapport aux débuts de 2022.

Points clés :

- Corrélation claire avec le CA et les transactions : hausse en fin 2021, stabilité en 2022, baisse début 2023.
- 2021 atypique (forte poussée de ventes sur quelques mois), alors que 2022 traduit une activité plus régulière.
- Tendance 2023 à surveiller : baisse généralisée sur clients, transactions et articles vendus.

Répartition du volume de vente par catégorie

Catégorie 0

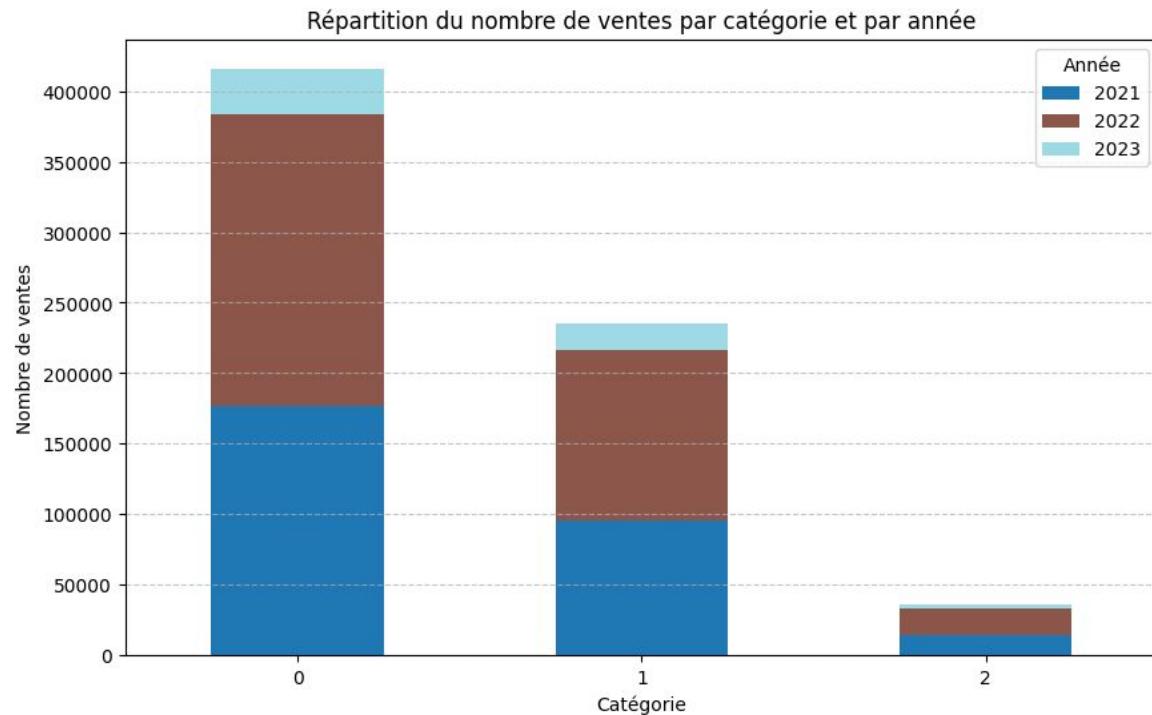
- Première sur l'ensemble de la période
- Forte progression entre 2021 et 2022
- Début 2023 en retrait mais reste largement dominante en volume

Catégorie 1

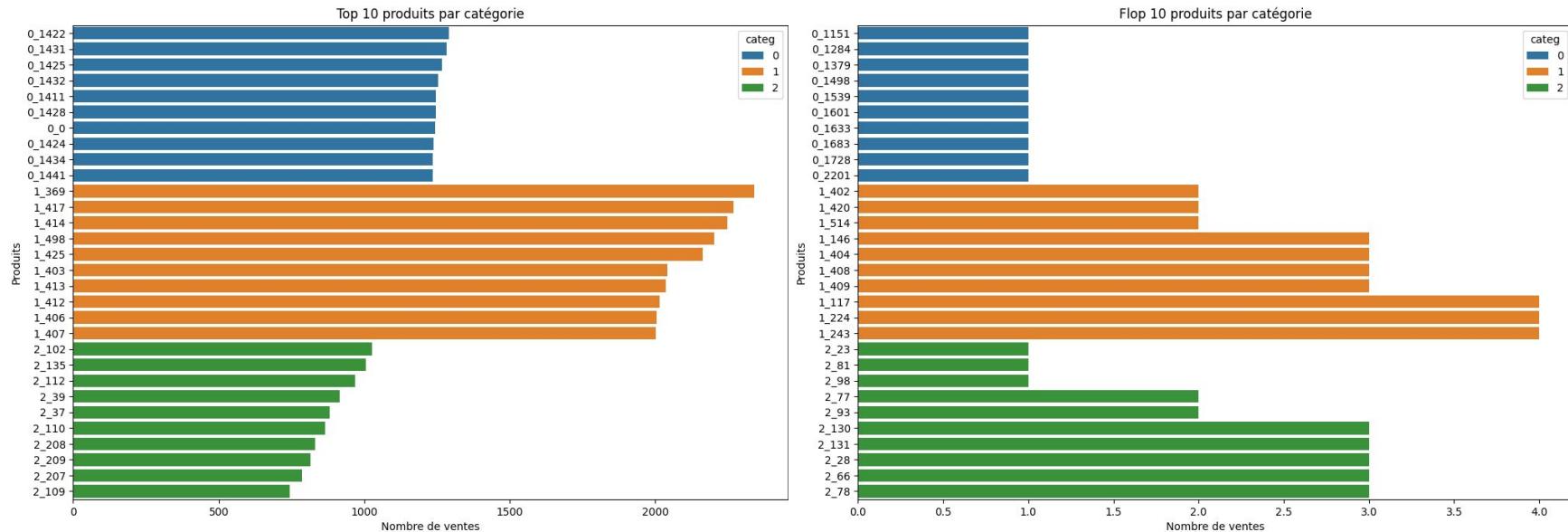
- Deuxième chaque année
- Croissance régulière 2021 → 2022
- Volumes environ deux fois inférieurs à la catégorie 0

Catégorie 2

- Très minoritaire : poids marginal dans le total des ventes
- Légère hausse entre 2021 et 2022, mais contribution encore faible



Top & Flop par catégorie



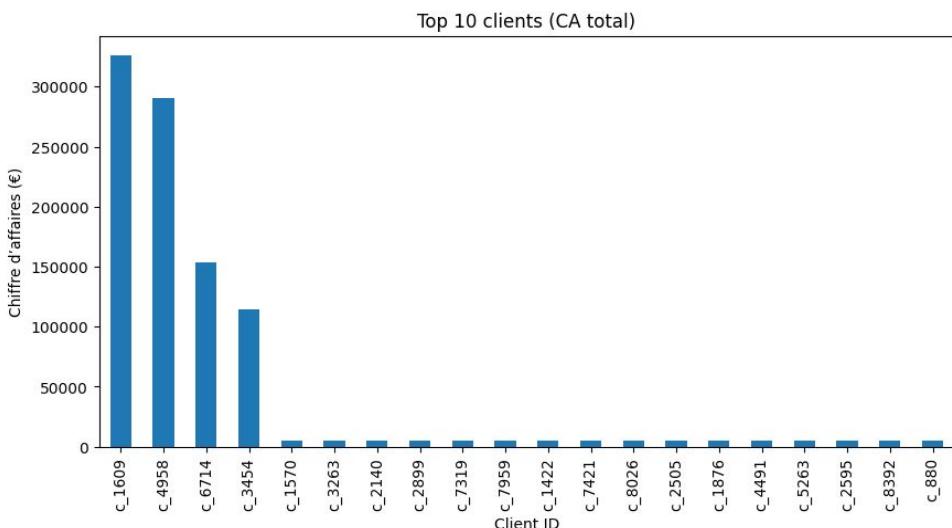
Top 10 produits

- Forte présence de la catégorie 1, qui concentre la majorité des meilleures ventes
- Catégorie 0 bien représentée dans le haut du classement
- Catégorie 2 présente mais avec des volumes plus modestes

Flop 10 produits

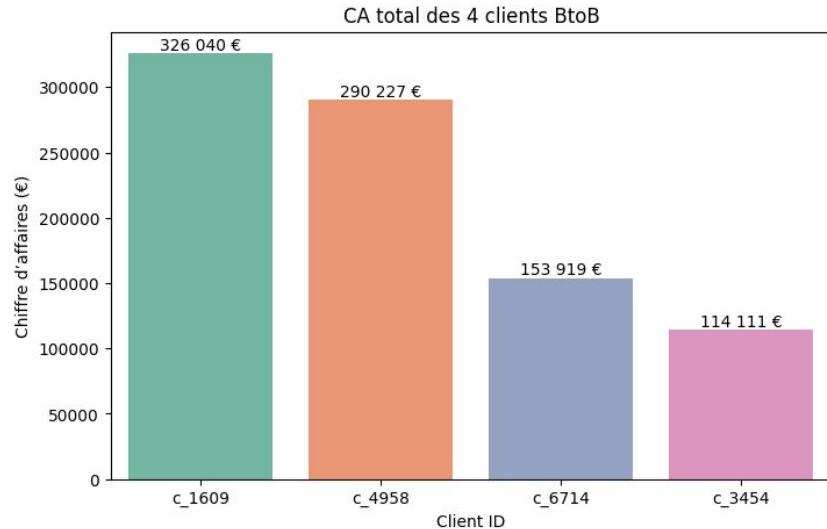
- Produits très dispersés avec seulement quelques ventes
- Catégorie 0 domine les flops, certains produits n'ont pas trouvé leur marché
- Catégorie 1 et 2 également touchée, mais de façon moins marquée

Identification des clients BtoB



Top 10 clients (CA total)

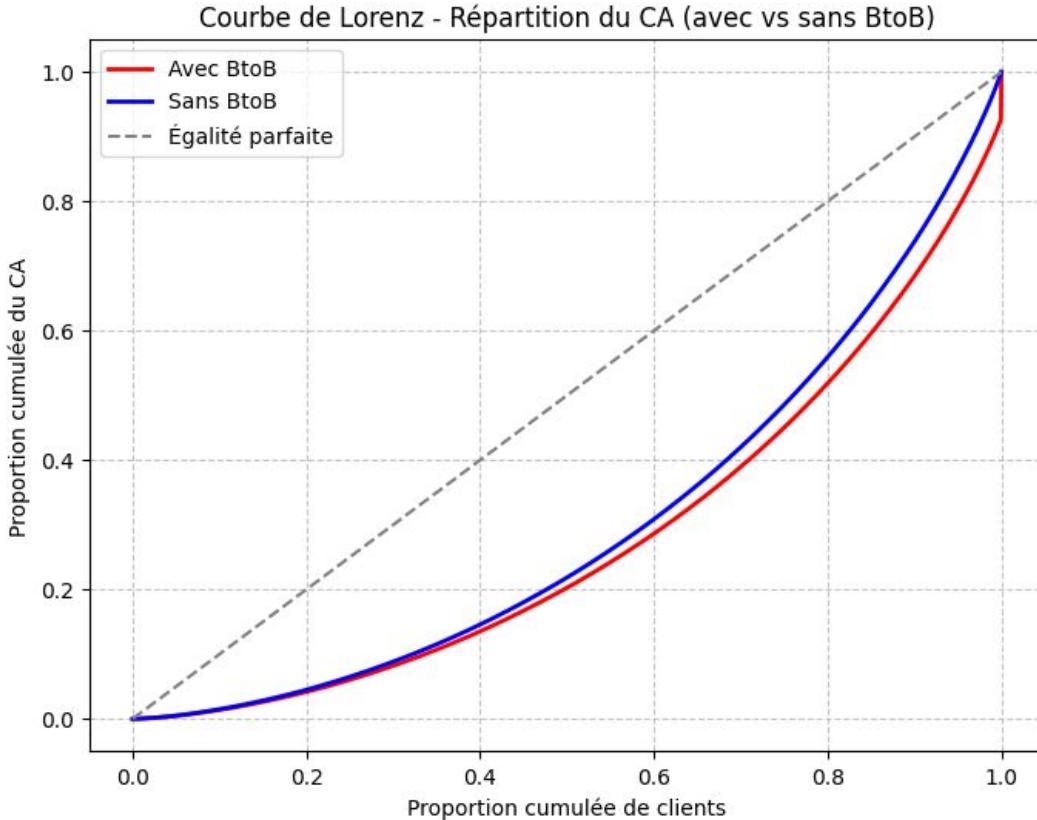
- Classement des clients par chiffre d'affaires décroissant
- Mise en évidence de 4 clients atypiques, générant un CA largement supérieur aux autres
- Ces profils correspondent très probablement à des clients BtoB



Focus sur les 4 BtoB

- CA cumulé : 885 k€, soit 7,4 % du CA total
- Enjeu stratégique : fidélisation & développement de l'offre BtoB

Mesure d'inégalité du CA - (Courbe de Lorenz)



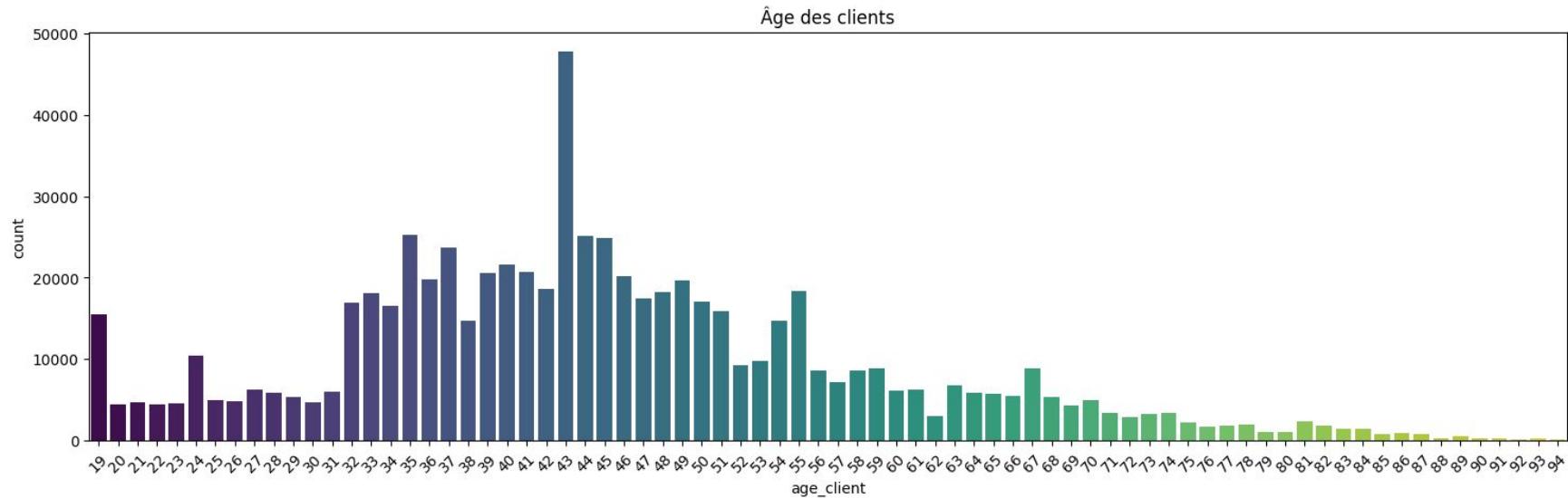
Principe de lecture

- La courbe de Lorenz compare la répartition réelle du CA avec une répartition égalitaire parfaite (droite grise).
- Plus la courbe s'éloigne de la diagonale, plus la concentration du CA est forte sur un petit nombre de clients.
- Ici, deux scénarios :
 - Avec BtoB (rouge)
 - Sans BtoB (bleu)

Résultat & interprétation

- Avec BtoB : inégalités plus marquées → une poignée de clients génèrent une part très importante du CA.
- Sans BtoB : répartition un peu plus équilibrée, mais concentration toujours présente.
- Conclusion : la dépendance aux grands comptes BtoB accentue l'inégalité, mais même en BtoC, une minorité de clients reste prédominante.

Répartition des âges des clients particuliers



Structure démographique

- Large éventail de clients, de 19 à 90 ans
- Concentration forte entre 30 et 50 ans
- Pic autour de 42 ans

Focus démographique

- Les classes d'âges centrales (30-50 ans) représentent la majorité des acheteurs.
- Présence de clients jeunes et plus agés, mais en volume réduit

Tests Statistiques - Résultats & Analyses

Genre et Catégories achetées (Chi²)

- Le test est significatif : le genre influence les catégories achetées
- Les hommes et les femmes n'achètent pas les mêmes proportions selon les catégories

Age et montant total d'achats (Spearman)

- Corrélation positive modérée, plus l'âge augmente, plus le montant cumulé est élevé
- Cela suggère une fidélité plus forte et/ou un pouvoir d'achat plus élevé

Age et fréquence d'achat (Spearman)

- Corrélation faible, mais significative. En moyenne, les clients plus âgés réalisent légèrement plus de transactions
- Cela montre qu'ils ne se contentent pas d'achats ponctuels à panier élevé, mais qu'ils reviennent aussi régulièrement

Age et panier moyen (Spearman)

- Forte corrélation négative significative
- Plus l'âge augmente, plus le panier moyen diminue
- Les jeunes dépensent plus par commande, mais achètent moins souvent
- Les plus âgés achètent plus régulièrement, mais avec des paniers plus petits

Age et catégories achetées (Kruskal-Wallis)

- Relation significative, les catégories varient selon l'âge
- Cela confirme que les préférences produits sont liées à l'âge

Conclusion

- Les tests statistiques confirment que certaines variables clés (âge, genre) influencent réellement le comportement d'achat.
- En revanche, la fréquence d'achat semble moins dépendre des caractéristiques démographiques

Recommandations stratégiques

Stimuler la fréquence chez les jeunes

- Mettre en place des abonnements (ex. livraison mensuelle, pack découverte) pour créer un rendez-vous régulier.
- Proposer des offres récurrentes (ex. “1 achat/semaine = remise”, “5 commandes = 1 produit offert”).
- Utiliser des notifications ciblées (email, SMS, push) pour rappeler les nouveautés ou promotions adaptées à leur profil.

Optimiser la fidélité des seniors

- Créer des programmes de points simples et clairs (chaque achat = points cumulés = avantages).
- Offrir des avantages récurrents (ex. réductions anniversaires, frais de port offerts après X commandes).
- Valoriser la stabilité de leur engagement (ex. statut “client premium” pour récompenser la régularité).

Cibler par genre et age

- Adapter les campagnes marketing : visuels, messages et offres correspondant aux attentes spécifiques des hommes/femmes, jeunes/seniors.
- Mettre en place un moteur de recommandations personnalisées : produits suggérés selon l’âge et le genre, basés sur les corrélations observées.
- Développer des promotions ciblées (ex. jeunes = nouveautés tendance, seniors = packs économiques ou confort).

Renforcer la relation BtoB

- Développer une offre dédiée aux grands comptes : tarifs préférentiels, conditions de paiement avantageuses.
- Mettre en place une gestion grands comptes avec un interlocuteur unique (account manager).
- Explorer des contrats cadre (volume garanti, livraisons régulières) pour sécuriser et stabiliser cette part du CA.