

Analysez les ventes d'une librairie avec R ou Python

Thomas Hébert

Data Analyst

22/11/2024

Contexte

Lapage : librairie physique + site web en activité depuis mars 2021

Intégration de Thomas dans l'équipe Marketing

Faire le point global sur l'activité du site web :

- les indicateurs et chiffres clés de l'entreprise
- le comportement des clients

Sommaire

Préparation des données

- Nettoyage et jonction des données
- Transformation des données

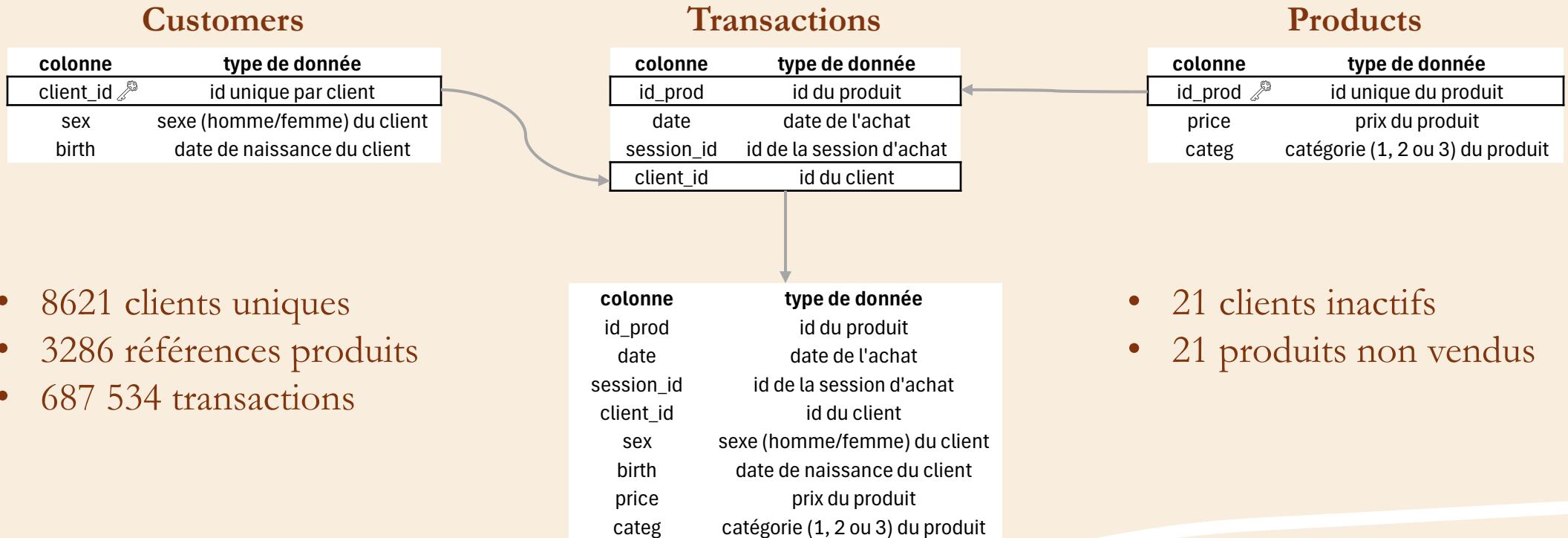
Analyses des indicateurs

- Chiffres d'affaires
- Ventes
- Produits
- Clients

Analyses du comportement clients

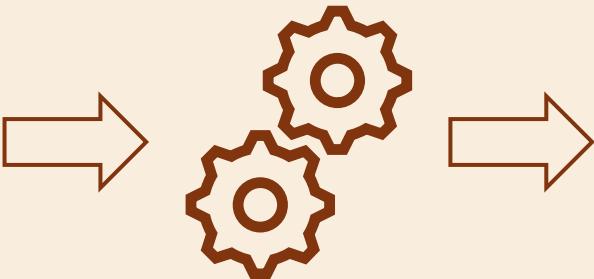
Préparation des données

Nettoyage et jonction des données



Transformation des données

colonne	type de donnée
id_prod	id du produit
date	date de l'achat
session_id	id de la session d'achat
client_id	id du client
sex	sexe (homme/femme) du client
birth	date de naissance du client
price	prix du produit
categ	catégorie (1, 2 ou 3) du produit



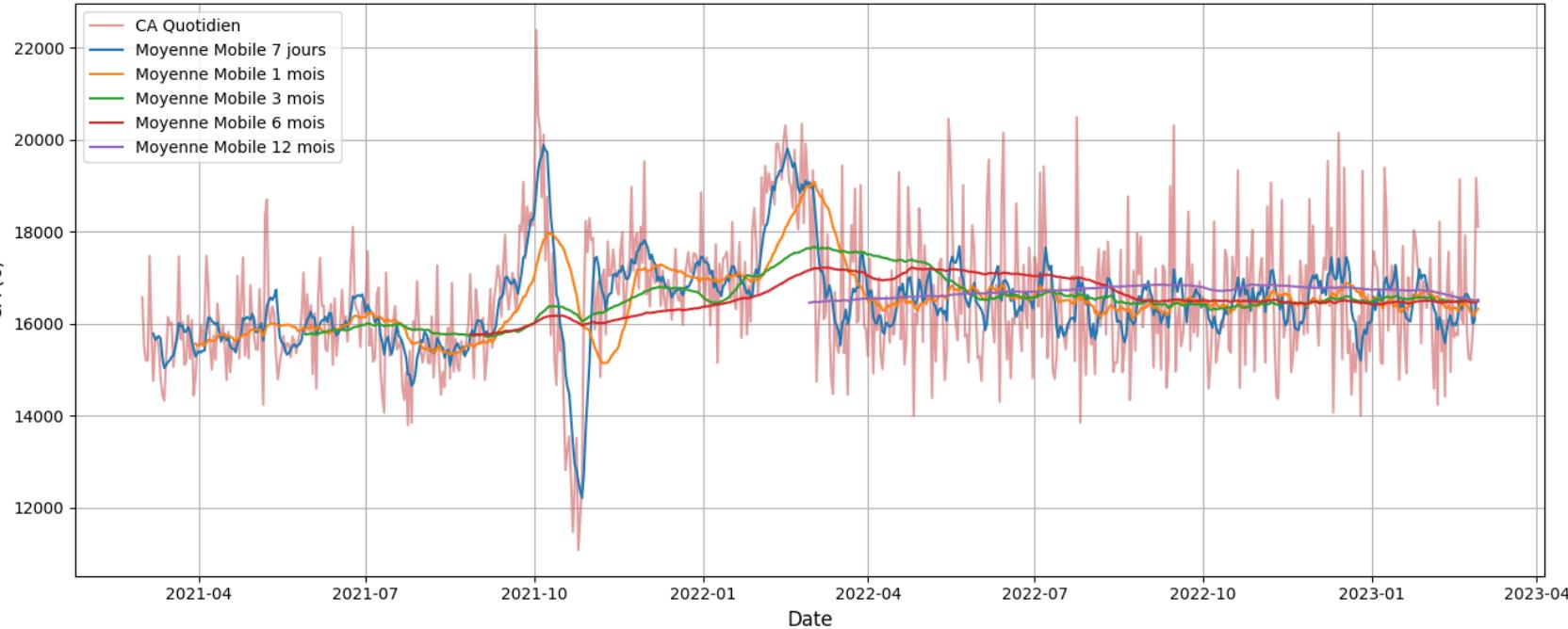
df_lapage

colonne	type de donnée
date_heure	date de l'achat
date	date de l'achat sans l'heure
année	extraction de l'année de 'date'
mois	extraction du mois de 'date'
jour	extraction du jour de 'date'
id_prod	id du produit
price	prix du produit
categ	catégorie (1, 2 ou 3) du produit
session_id	id de la session d'achat
client_id	id du client
sex	sexe (homme/femme) du client
age	calcul de l'âge du client selon 'birth'
tranche_age	calcul de la tranche d'âge selon 'age'

Analyses des indicateurs

Chiffre D'affaires

Évolution du Chiffre d'Affaires (CA) Quotidien avec Moyennes Mobiles



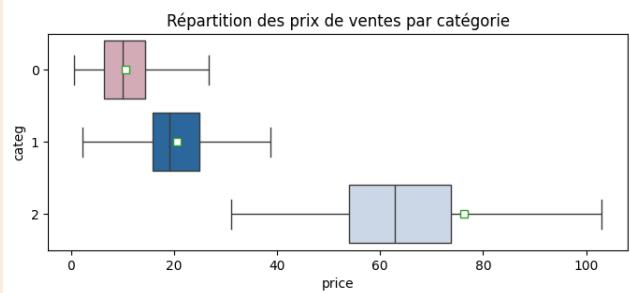
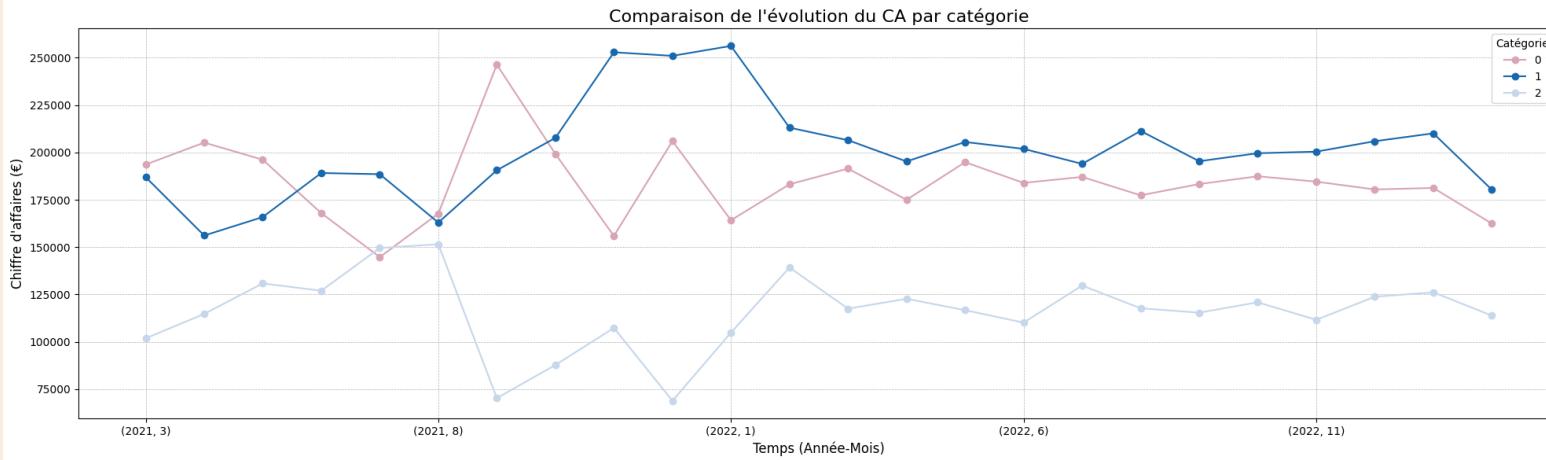
Année	Mois	Chiffre d'Affaires
2021	10	4 944 761 €
2022	12	6 108 682 €
2023	2	974 220 €
TOTAL	24	12 027 663 €

Progression sur 2 exercices
comptables

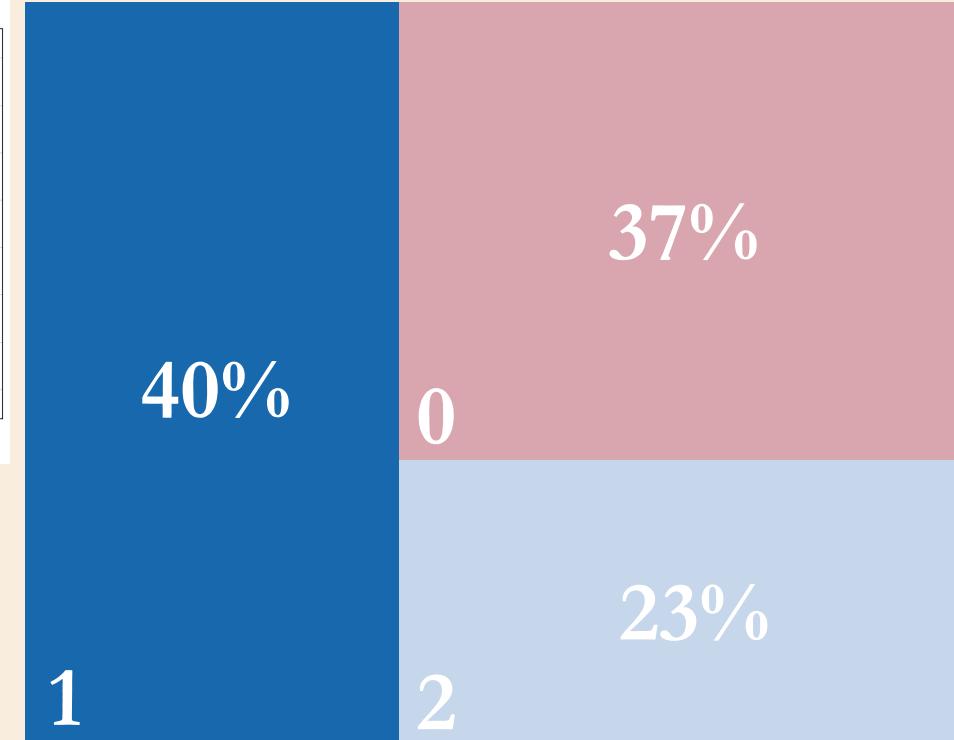
+0,27%

Chiffre D'affaires

Parts du CA



Catégories 1 & 0 dominantes sur le CA, avec des prix de ventes et une variabilité similaire.

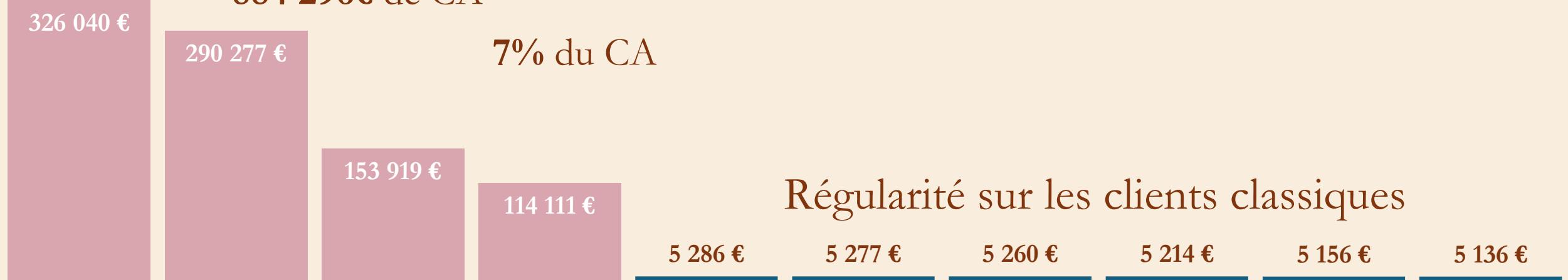


Chiffre D'affaires

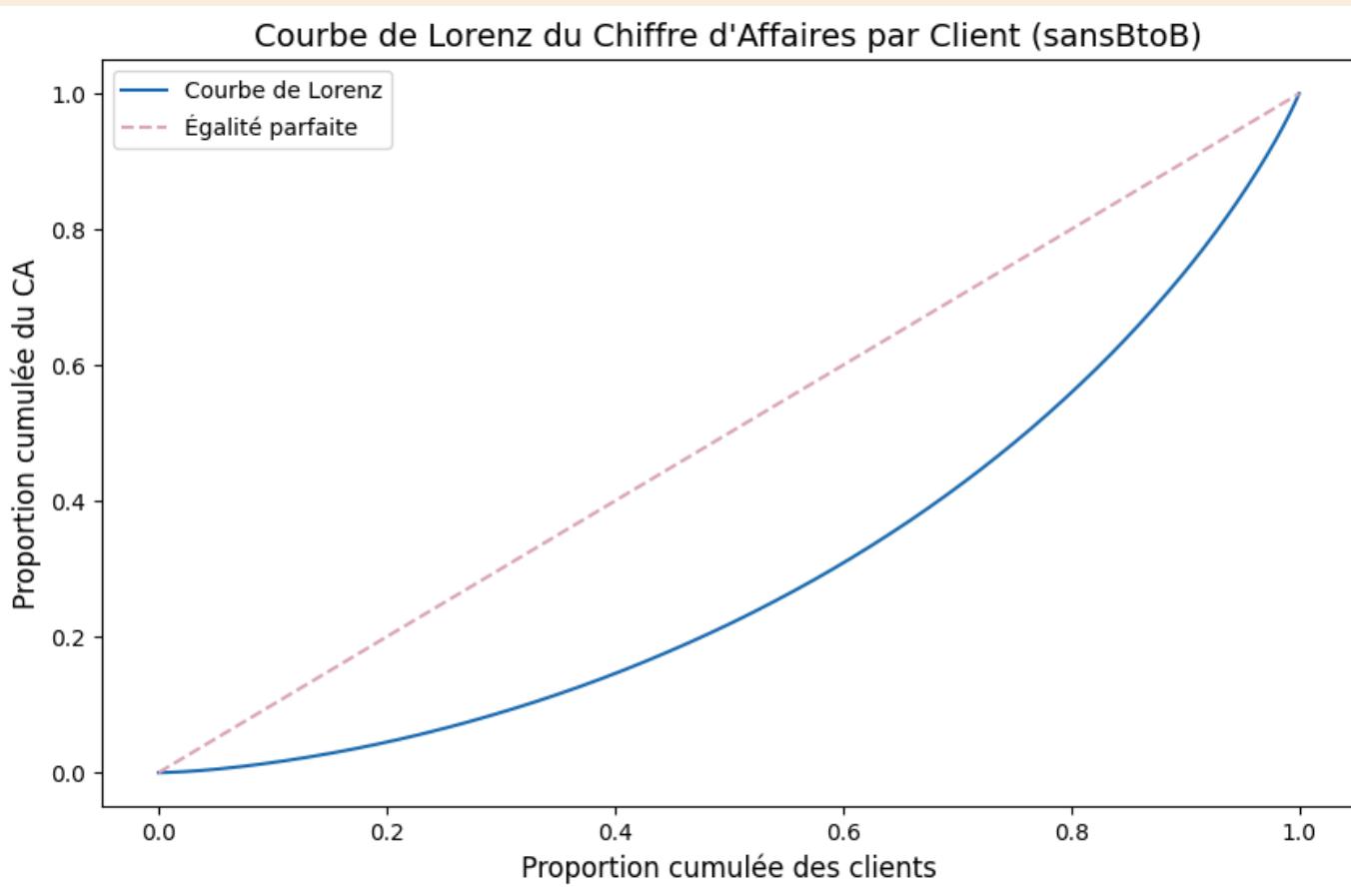
TOP 10 Clients

4 clients BtoB

884 296€ de CA



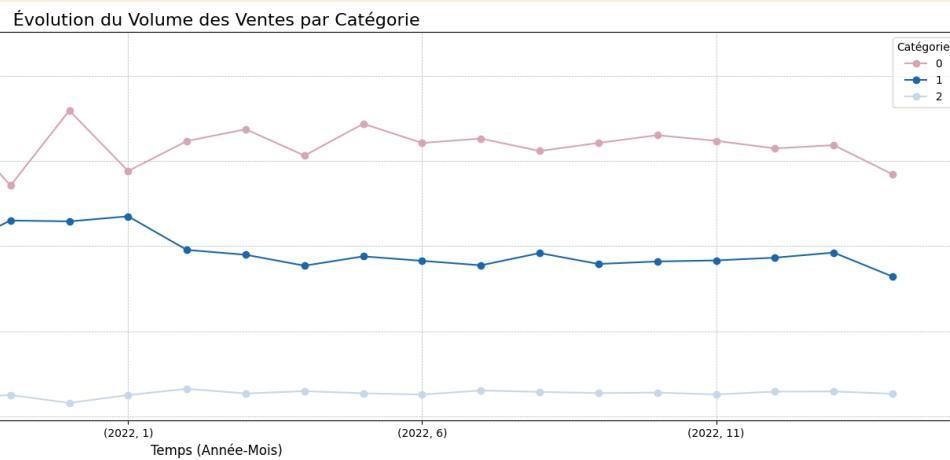
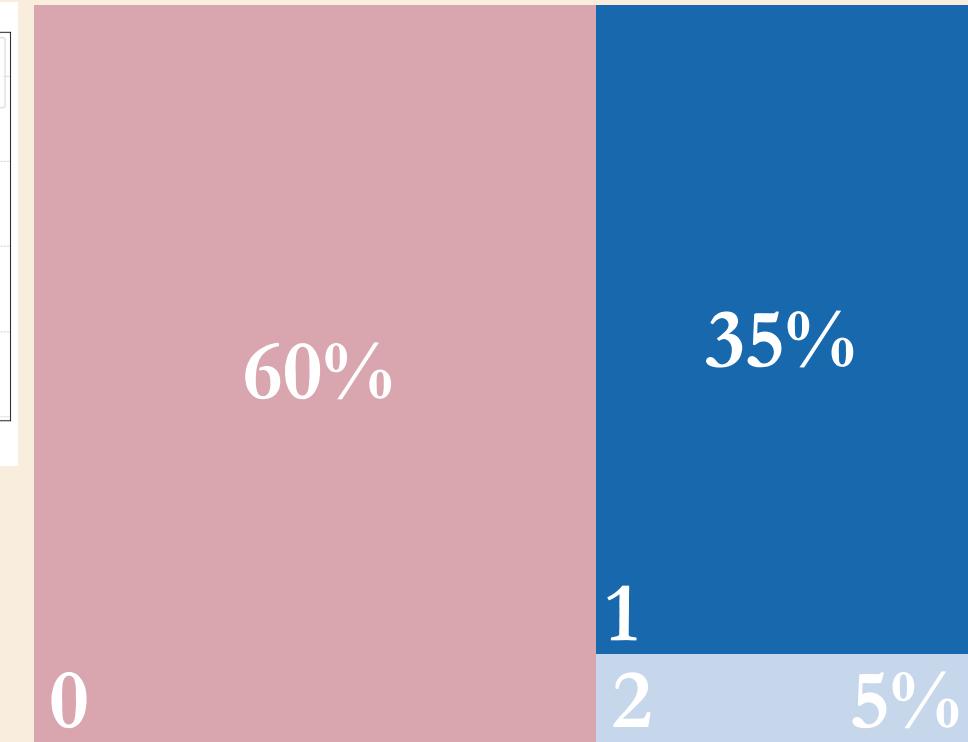
Chiffre D'affaires



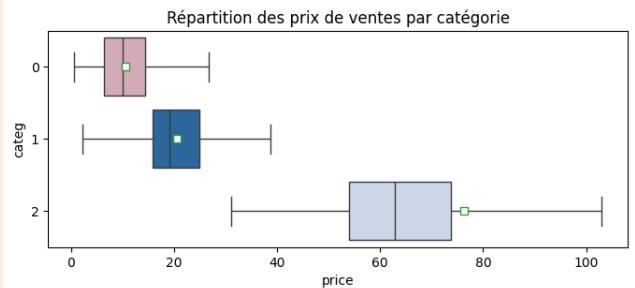
- Indice de Gini : **0,39**
- Répartition inégale du CA
- **50% des clients génèrent 20% du CA**

Ventes

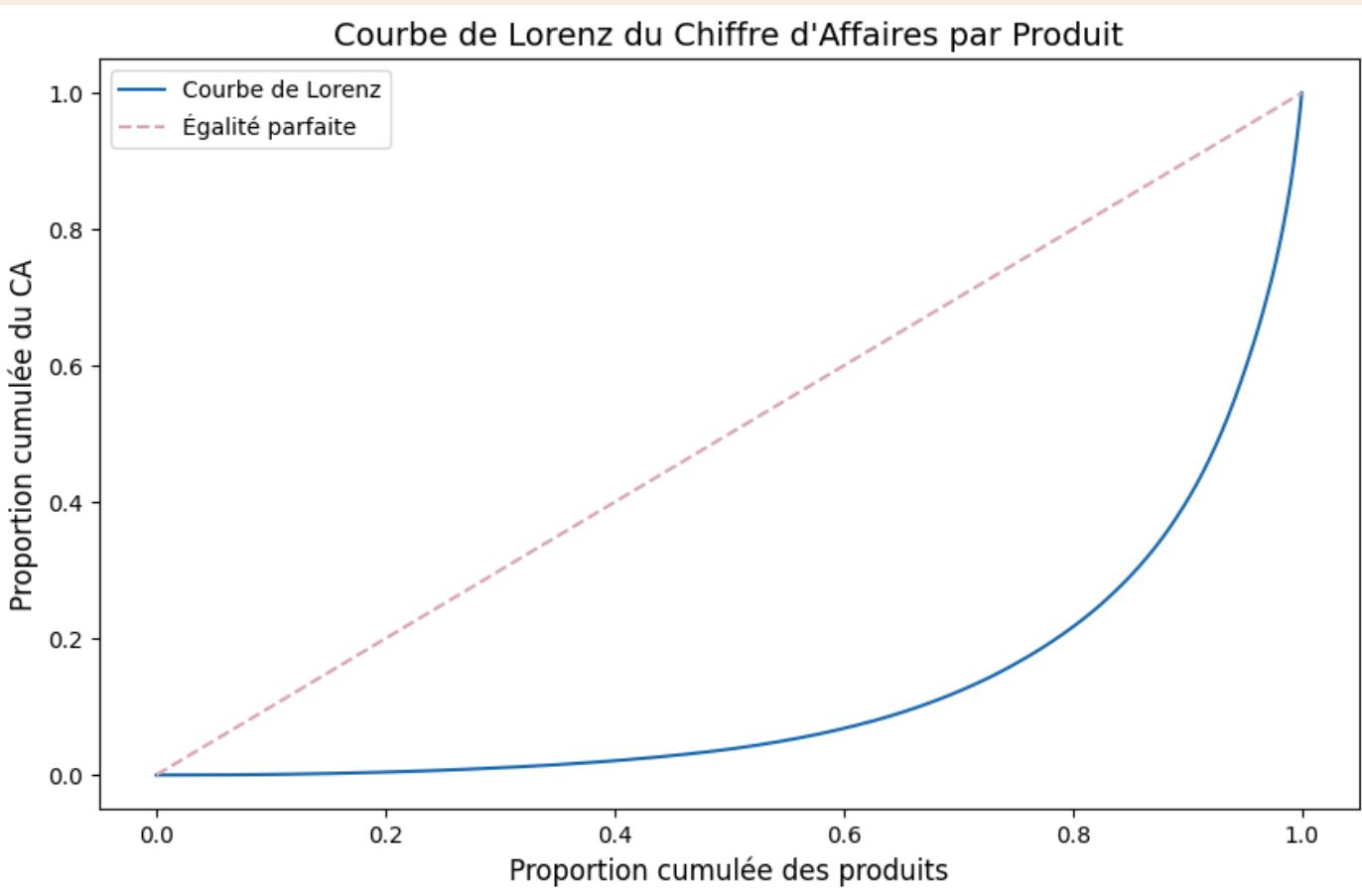
Parts des ventes



Catégorie 0 dominante en termes de volume, son faible prix appuie cette position.



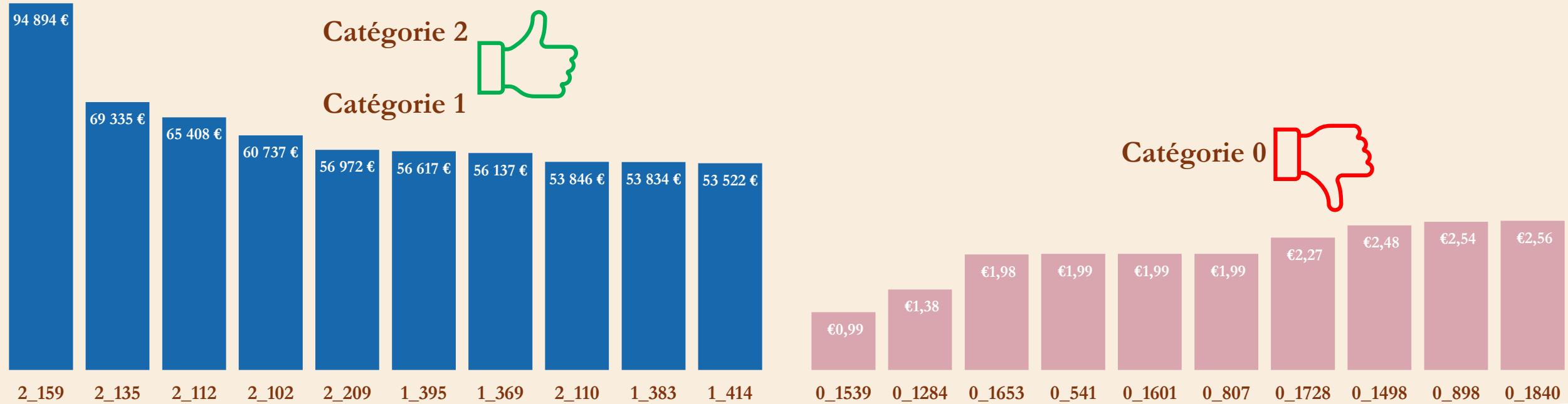
Produits



- Indice de Gini → **0,74**
- Répartition du CA très inégale
- Configuration en **80/20** ou seulement **20% des produits** génèrent **80% du Chiffre d'Affaires**.

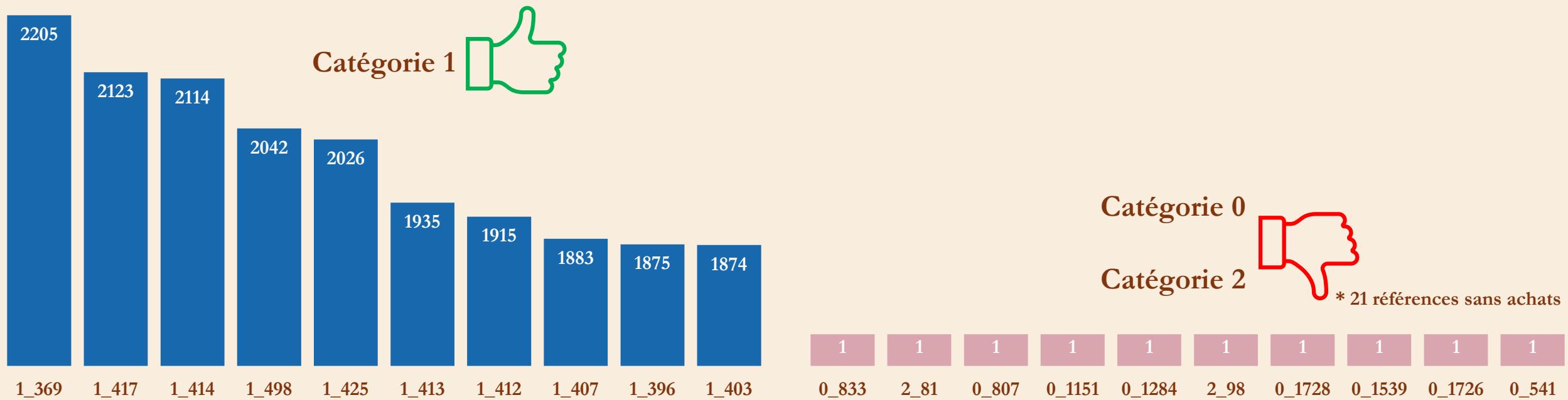
Produits

TOP et FLOP Références (CA)



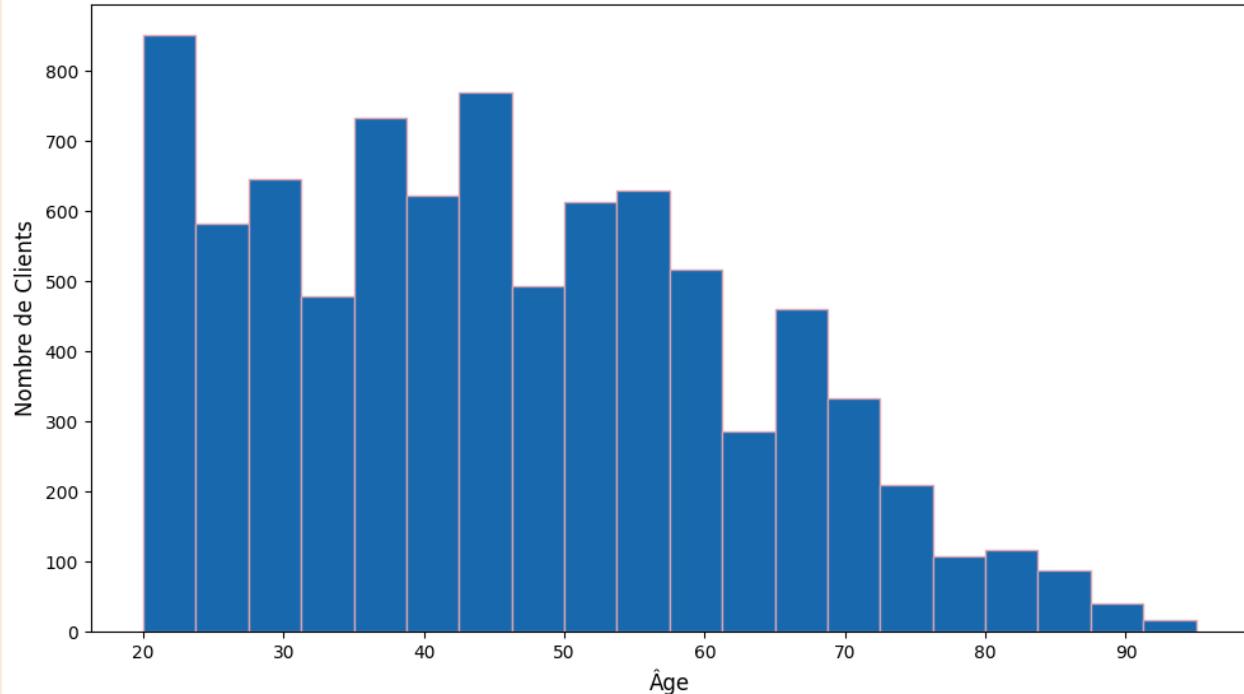
Produits

TOP et FLOP Références (Ventes)



Clients

Distribution des Âges (sans BtoB)



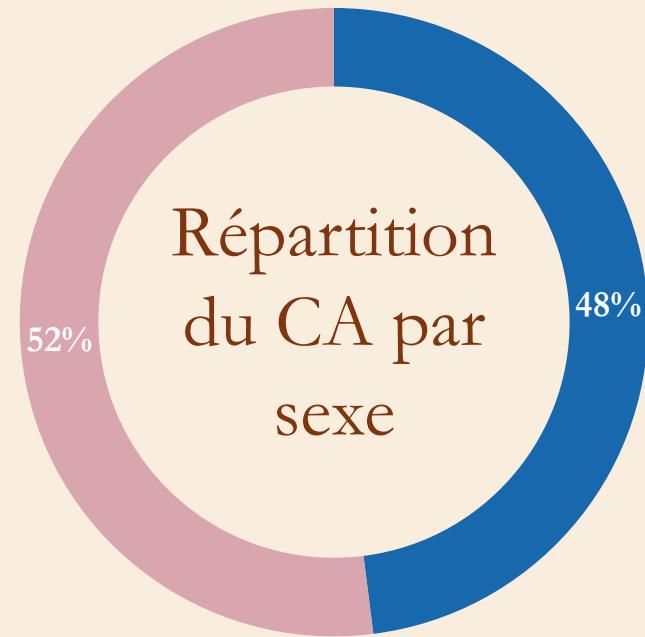
Répartition des clients par tranches d'âges



 20 ans à 95 ans / Moyenne et Médiane → 45 ans

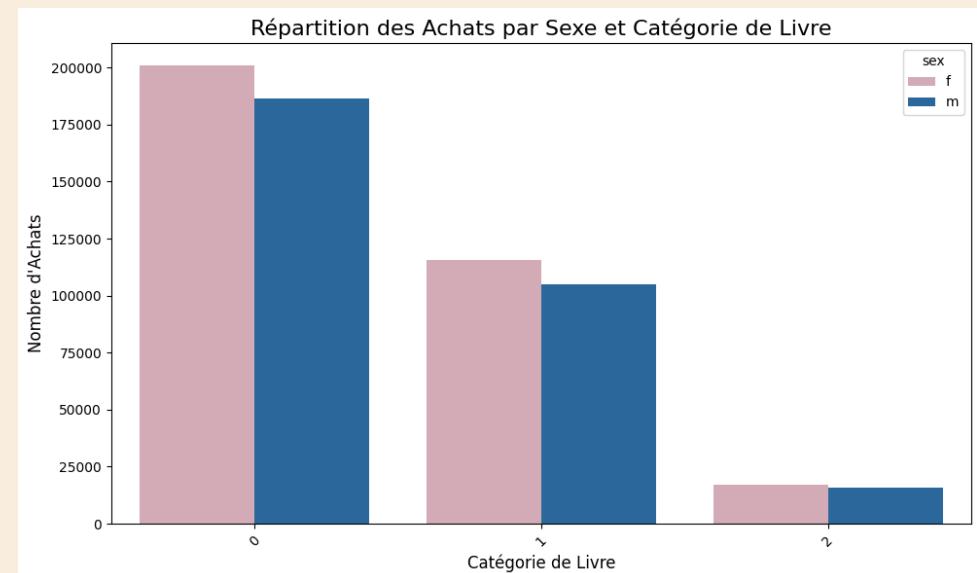
Clients

Répartition du CA par tranches d'âges



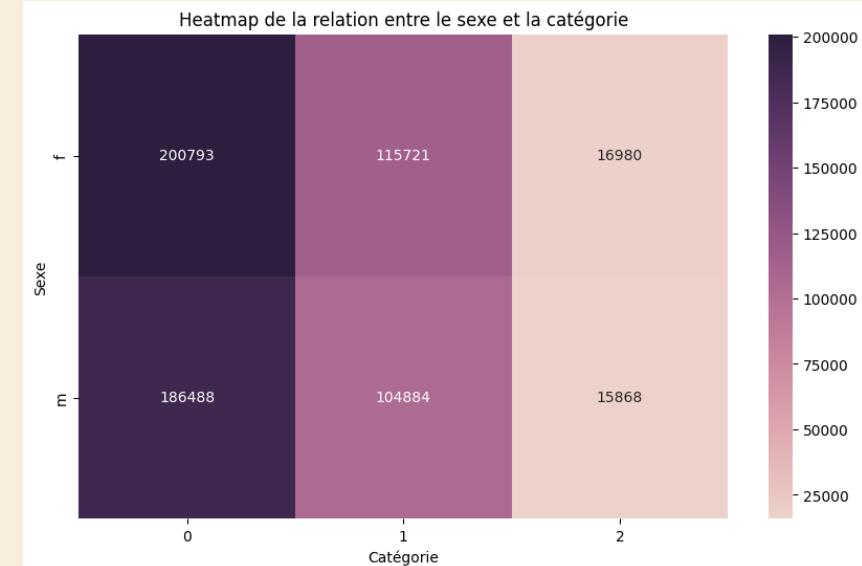
Analyses du comportement clients

Corrélation entre le sexe et la catégorie de produits



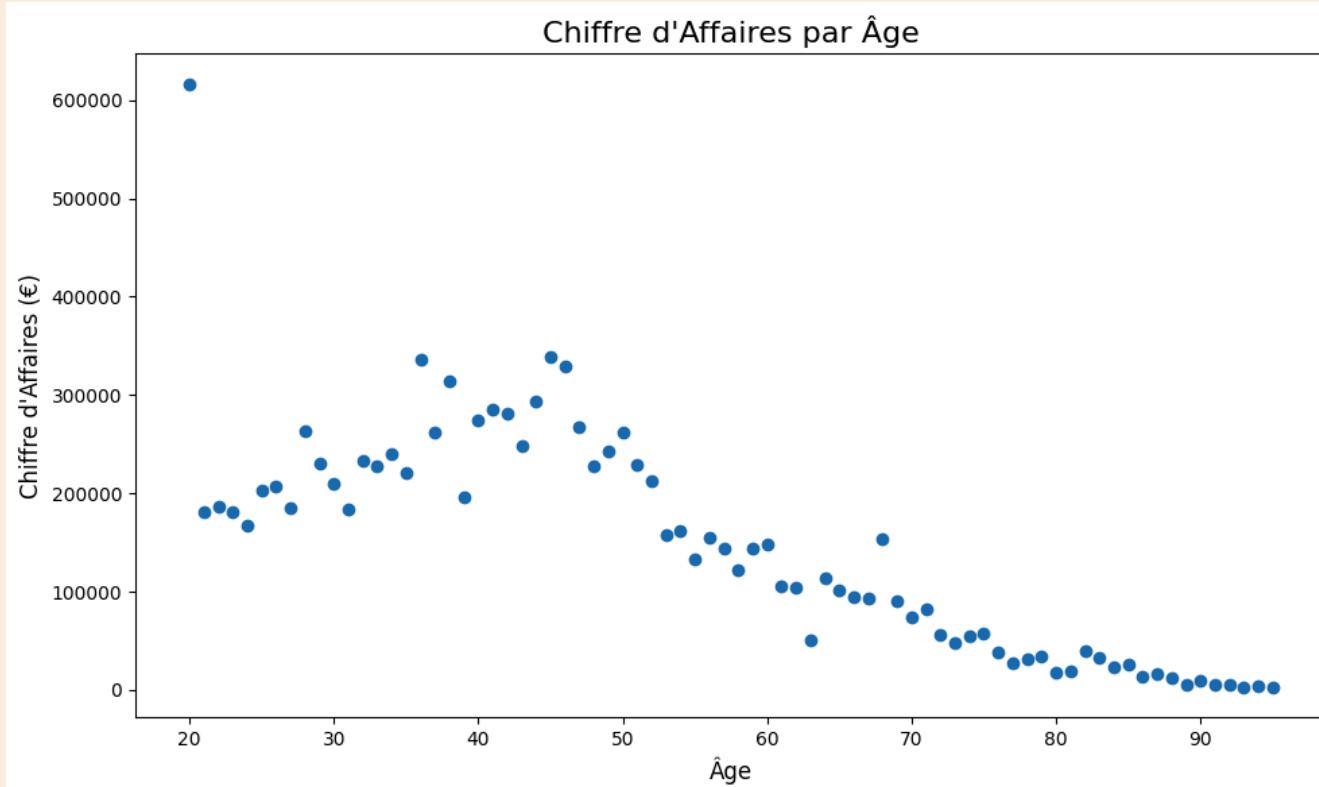
Chi-2 → 22,67
P-value → $1,20 \times 10^{-5}$
V de Cramer → 0,006

La répartition des genres est quasi identique entre les catégories avec une supériorité des femmes dans toutes les catégories.



Il y a bien une corrélation mais celle-ci est très faible.

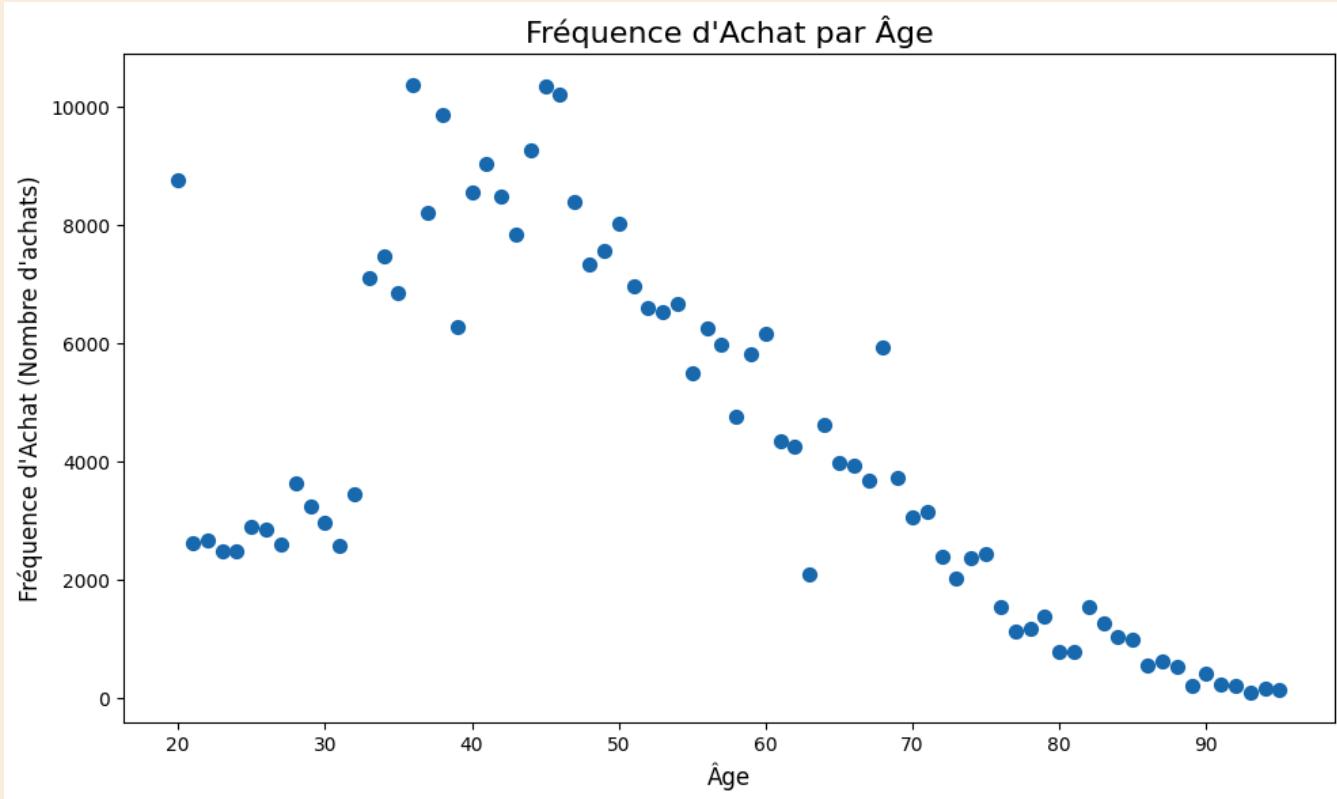
Corrélation entre le chiffre d'affaires et l'âge des clients



Spearman → -0,87
P-value → 5,95^e-25

Corrélation négative :
Plus les clients sont âgés, moins ils génèrent de chiffre d'affaires.

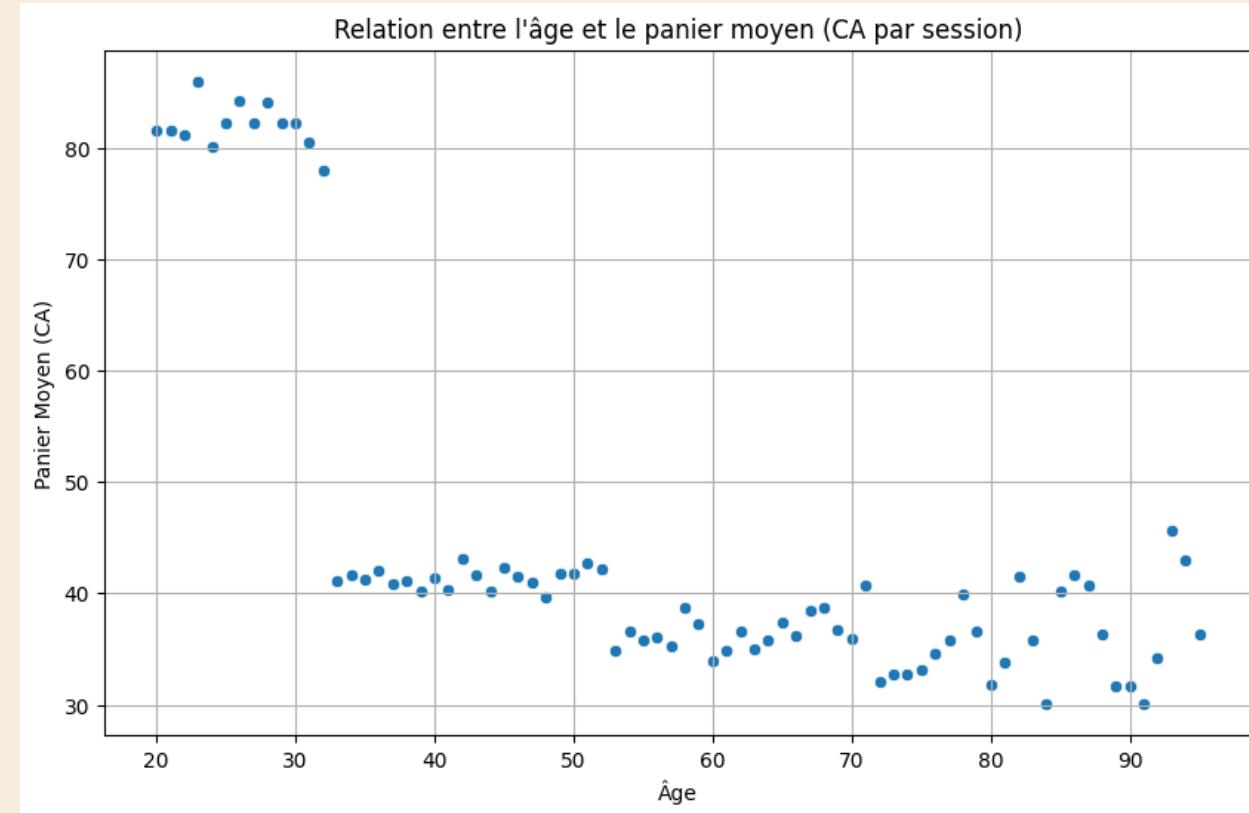
Corrélation entre la fréquence d'achat et l'âge des clients



Spearman → -0,65
P-value → 9,36^e-11

Corrélation négative :
A mesure que l'âge augmente, la
fréquence d'achat diminue.

Corrélation entre le panier moyen et l'âge des clients



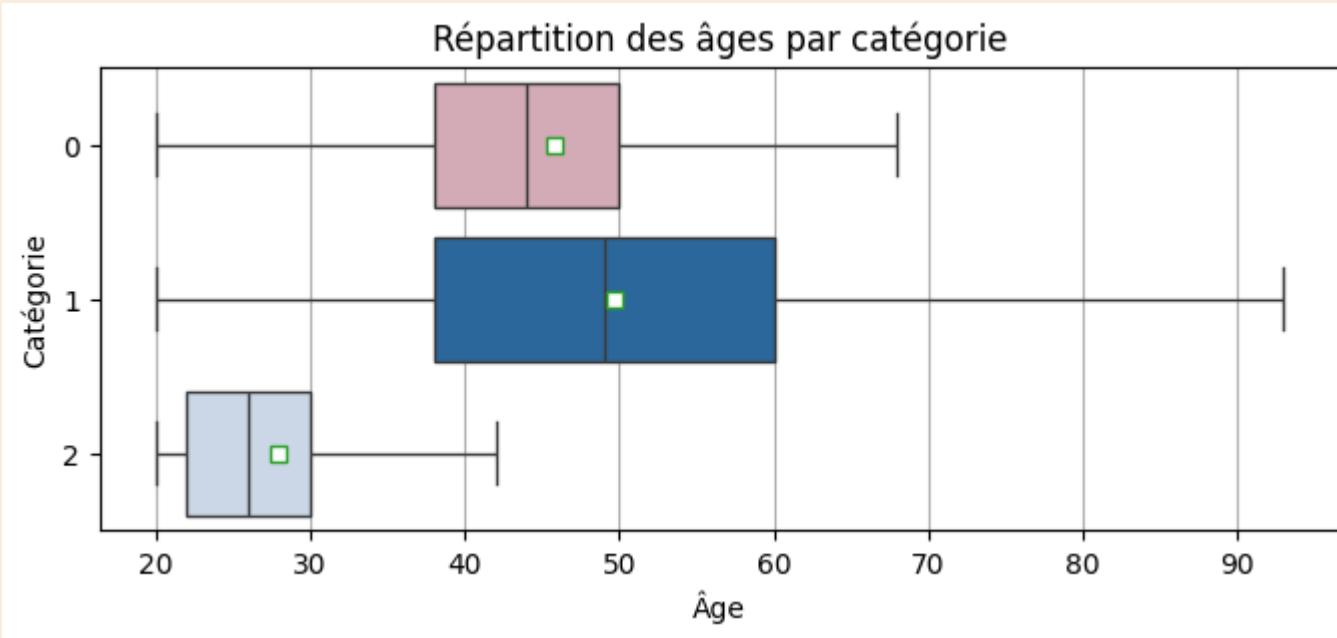
Spearman → -0,69

P-value → 2,69^e-12

Corrélation négative :

Plus les clients sont âgés, moins leur panier moyen est élevé.

Corrélation entre la catégorie de produits et l'âge des clients



Kruskal-Wallis → 71359
P-value → 0,0

L'âge joue un rôle déterminant dans la séparation des groupes, avec des différences d'âge statistiquement significatives entre les catégories.

Conclusions

L'analyse du **chiffre d'affaires** entre 2021 et 2023 montre une **stabilité (+0,27%)**, mais souligne des **dynamiques contrastées** selon les catégories :

- **Catégorie 0** (prix moyen 10€) : 37% du **CA**, mais dépend trop du volume (**60%**) avec une rentabilité limitée.
- **Catégorie 1** (prix moyen 20€) : 40% du **CA** et un équilibre volumes/catalogue, elle est le **pilier de l'offre**.
- **Catégorie 2** (prix moyen 76€) : 23% du **CA** avec seulement 5% des ventes, une forte valeur ajoutée mais faible contribution en volume.

Les **Tops/Flops** ainsi que la courbe de **Lorenz CA/Produits** renforcent cette conclusion.

Différentes stratégies marketing et une offre produits à revoir sur les **catégories 0 et 2** seraient nécessaires.

Les analyses clients révèlent un lien important entre **l'âge des clients** et leurs **comportements d'achat** :

- Une diminution du **chiffre d'affaires**, du **panier moyen** et de la **fréquence d'achat** avec l'âge.
- Des **différences marquées** entre les catégories, suggérant que l'âge est un facteur clé à intégrer dans la segmentation et les actions ciblées.