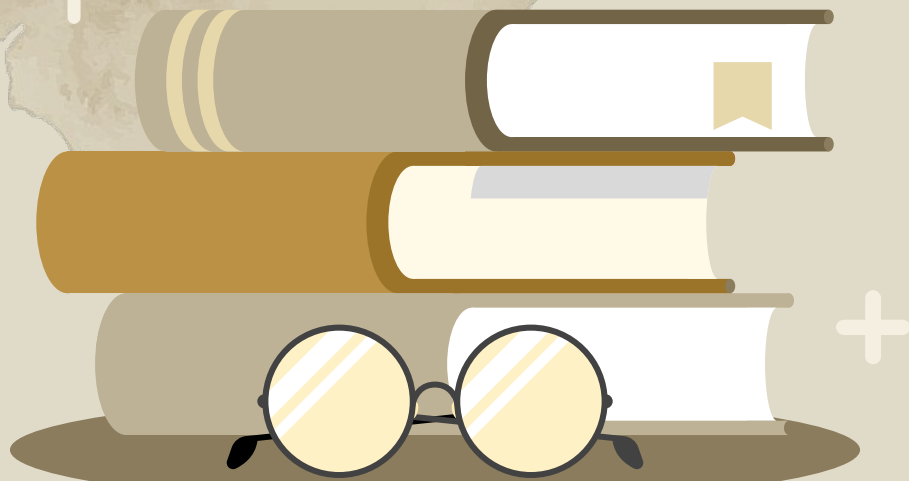


Analyse des ventes Lapage

Par Thomas Fruchard,
Data analyst



Objectif de la mission

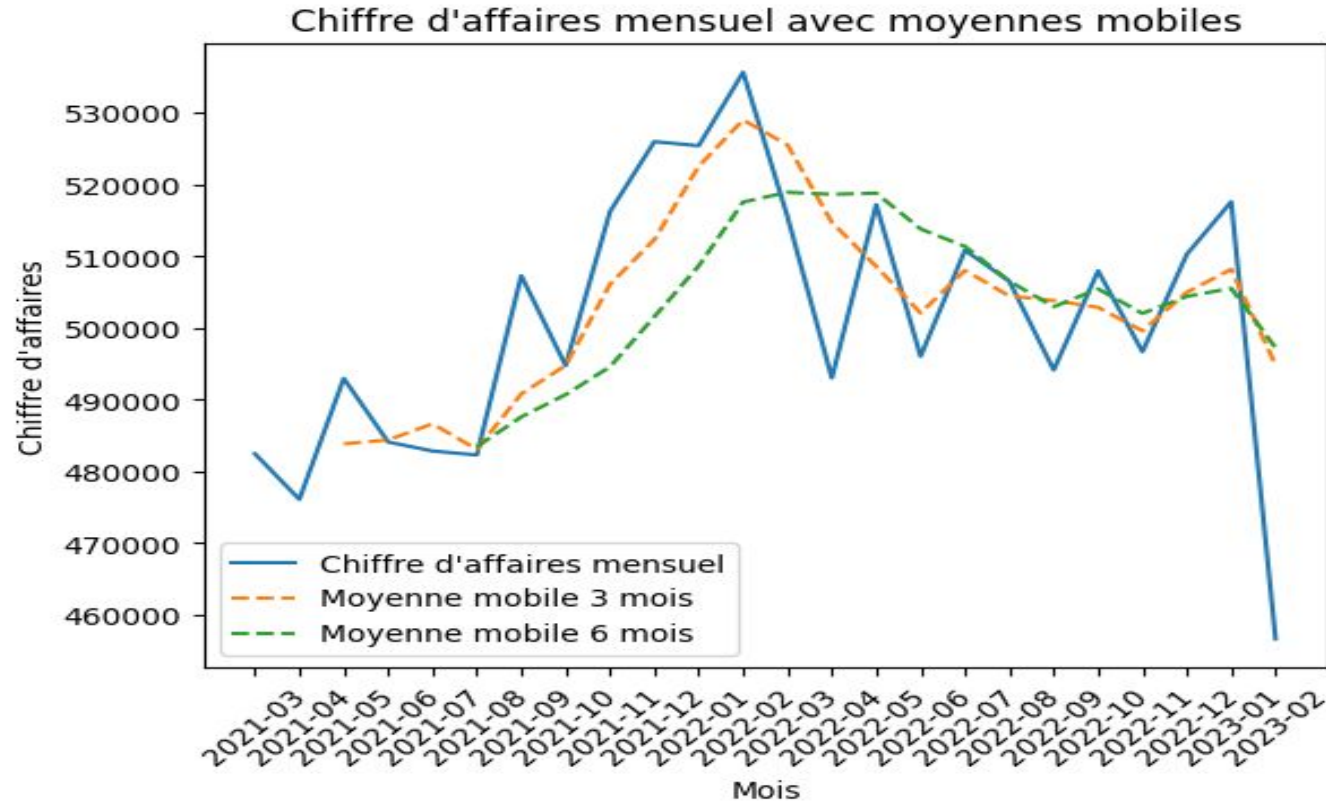
- Une analyse des différents indicateurs de vente.
- Une analyse plus ciblée sur les clients : l'objectif est de comprendre le comportement de nos clients en ligne, pour pouvoir ensuite comparer avec la connaissance acquise via nos librairies physiques.
 - a. le lien entre le genre d'un client et les catégories des livres achetés
 - b. le lien entre l'âge des clients et le montant total des achats,
 - c. le lien entre l'âge des clients et la fréquence d'achat,
 - d. le lien entre l'âge des clients et la taille du panier moyen,
 - e. le lien entre l'âge des clients et la catégorie des livres achetés.

01

Analyse de vente



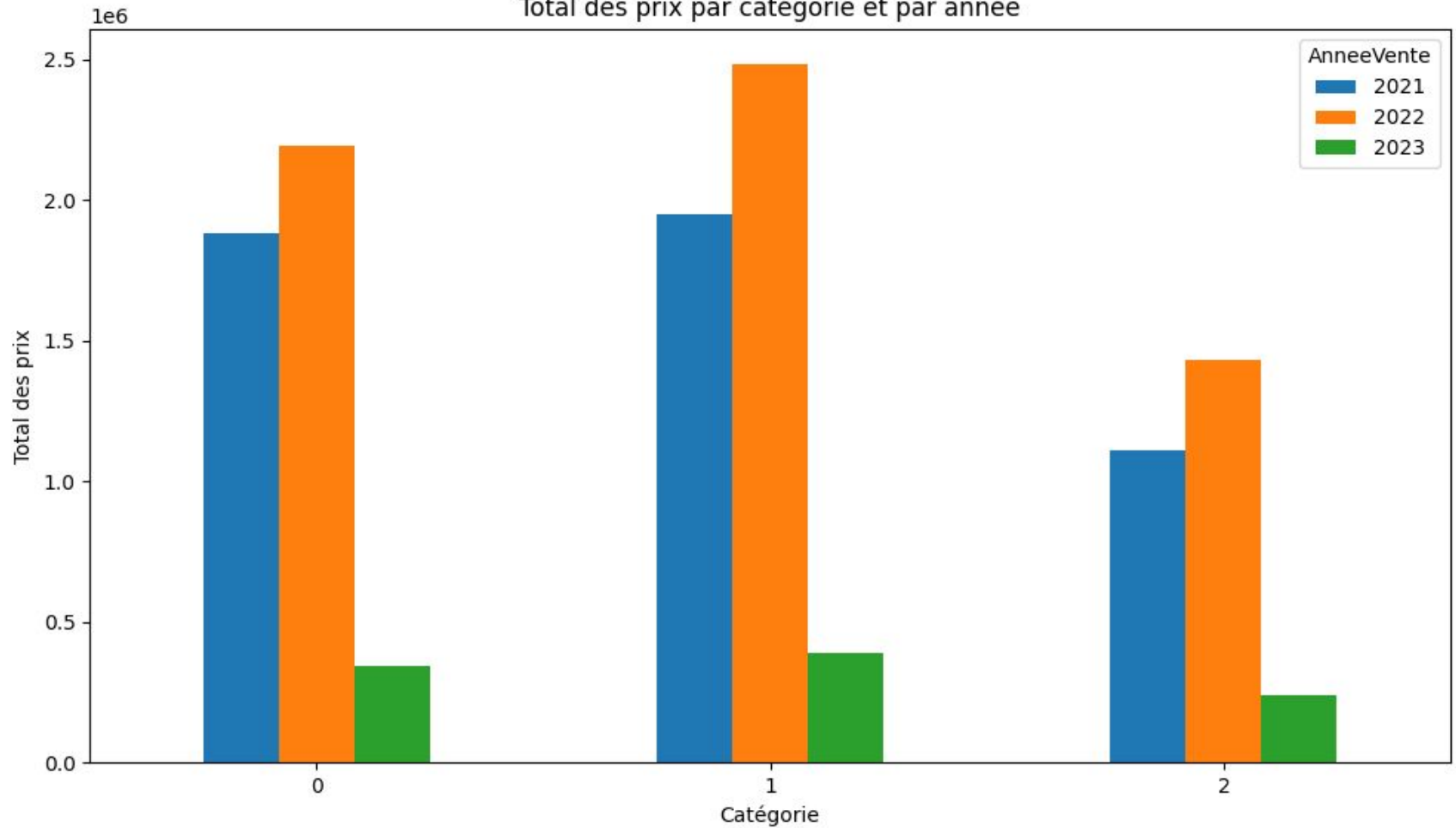
Chiffre d'affaires avec moyenne mobile



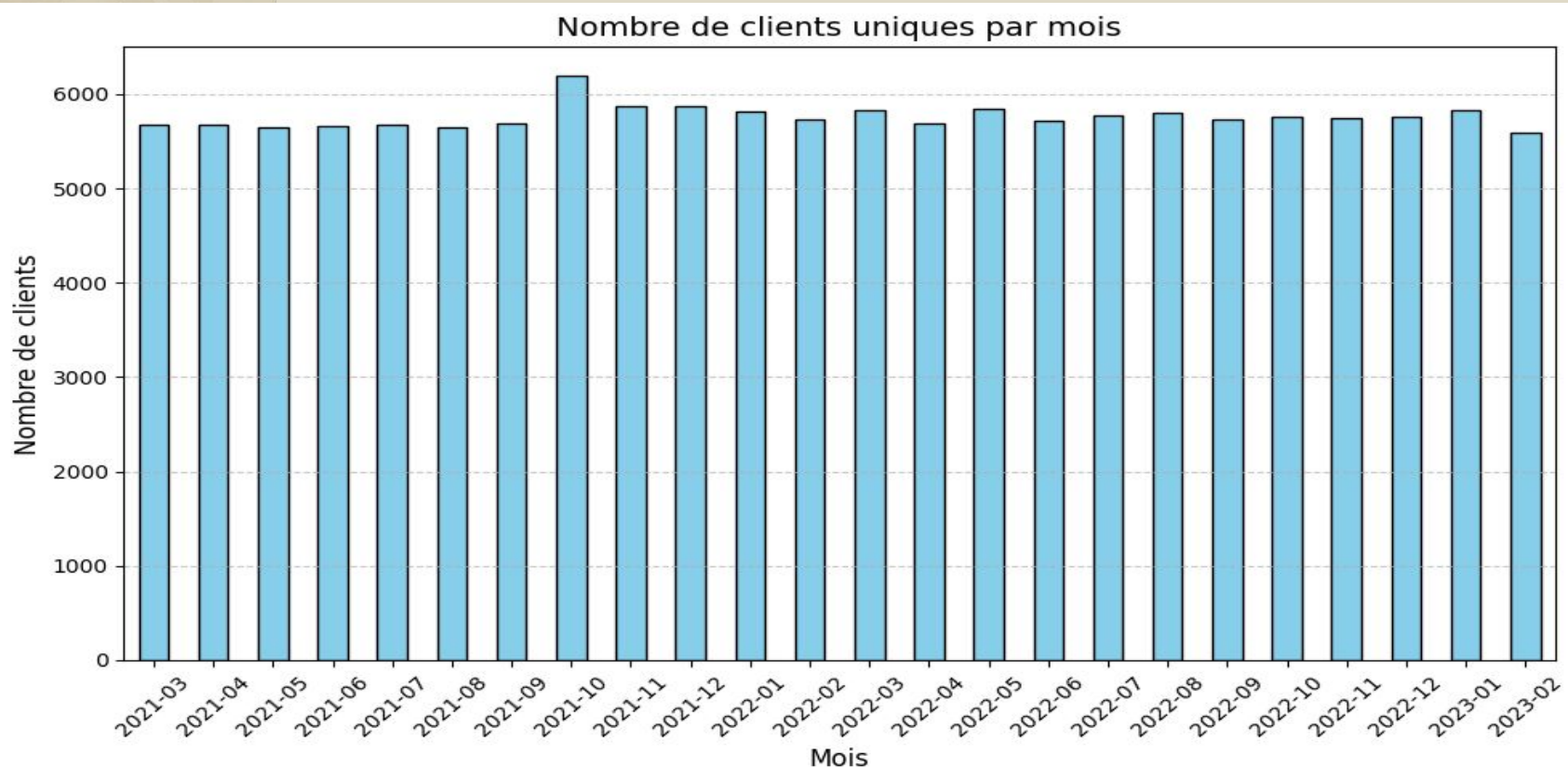
Chiffre d'affaires par catégorie

Catégorie	Année	Chiffre d'affaires
0	2021	1 883 020 €
1	2021	1 951 930 €
2	2021	1 109 810 €
0	2022	2 192 970 €
1	2022	2 485 275 €
2	2022	1 430 436 €
0	2023	343 740 €
1	2023	390 451 €
2	2023	240 028 €

Total des prix par catégorie et par année

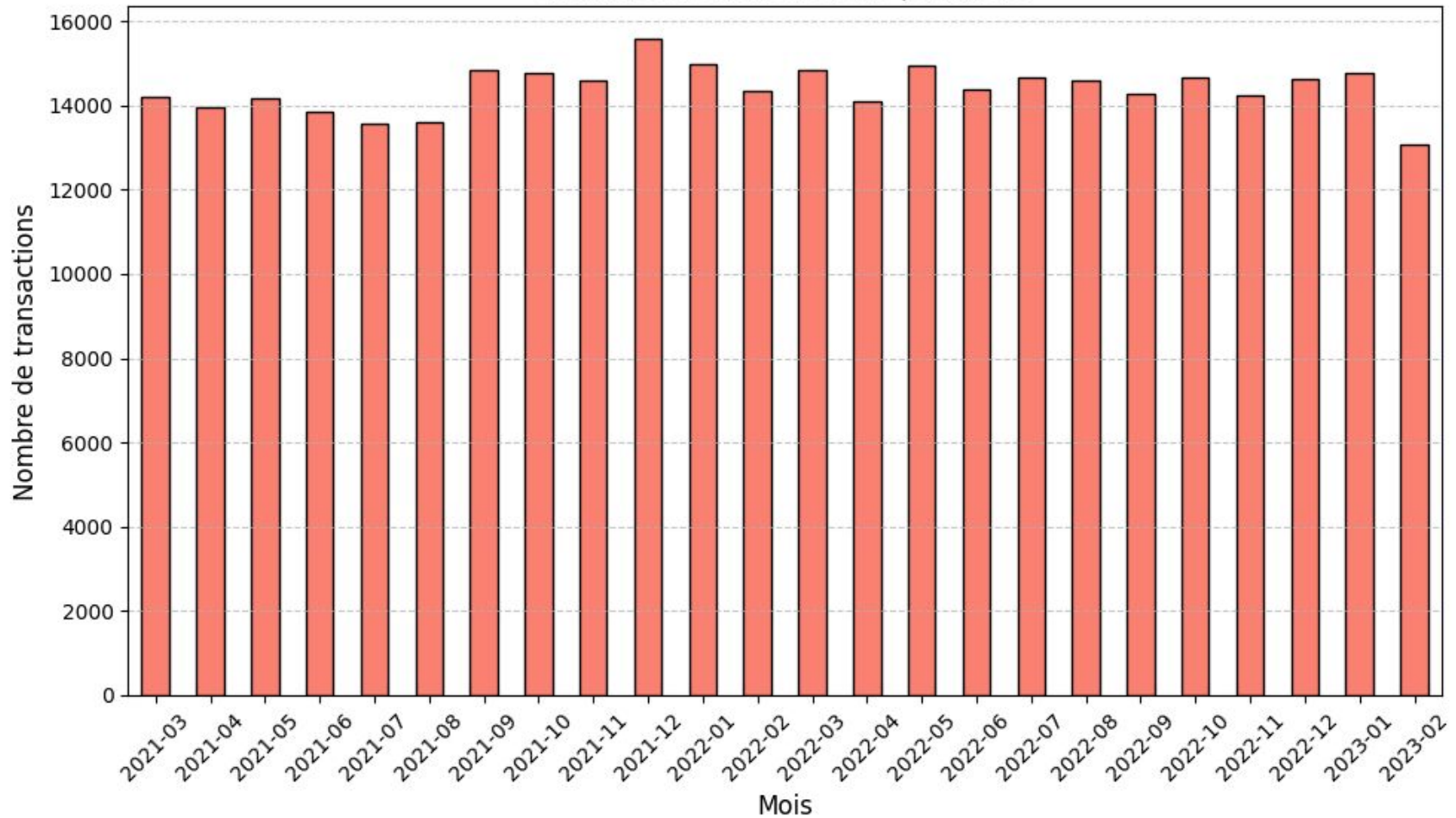


Nombre de client par mois

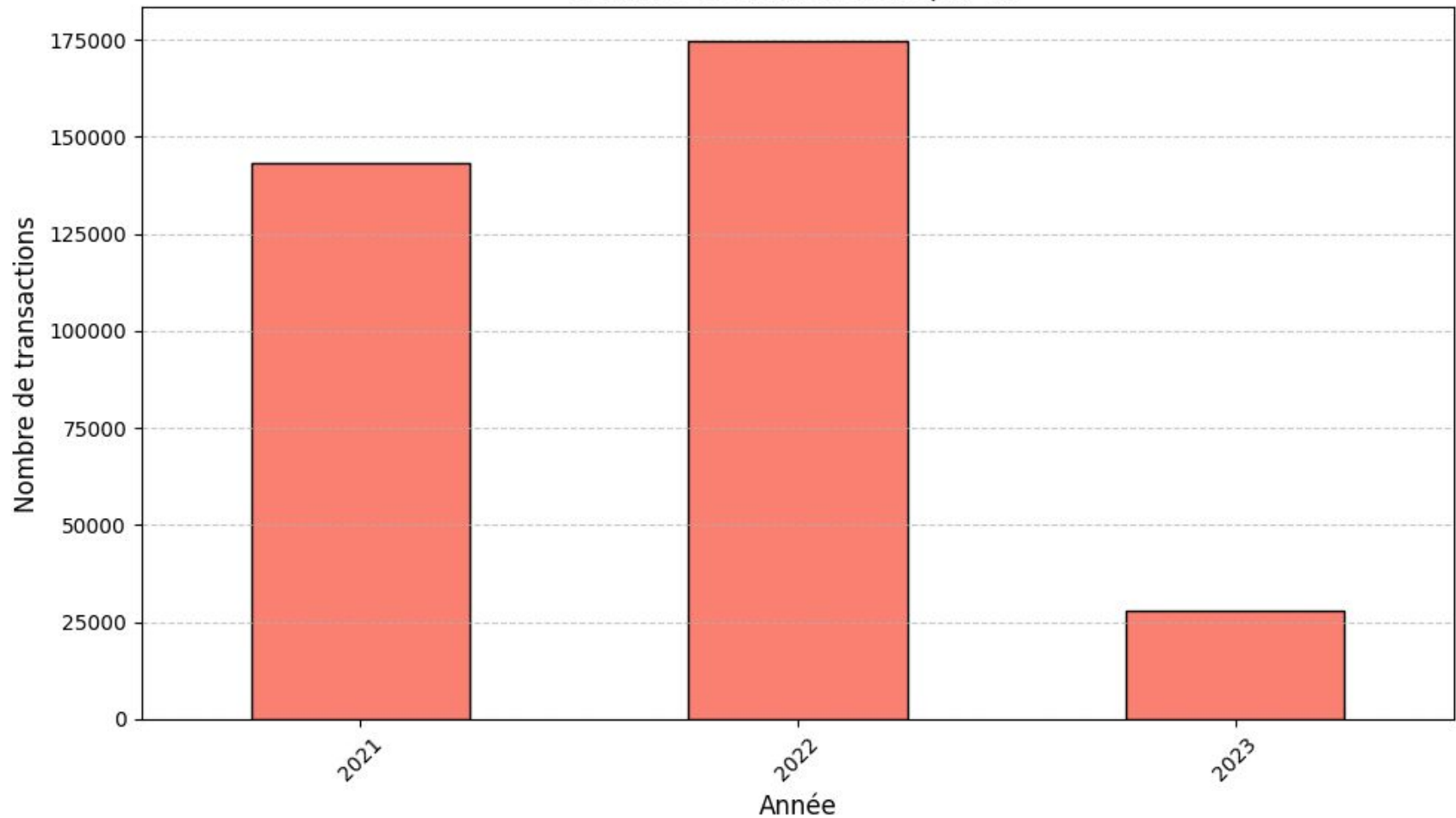


Transaction par mois puis
par année

Nombre de transactions par mois



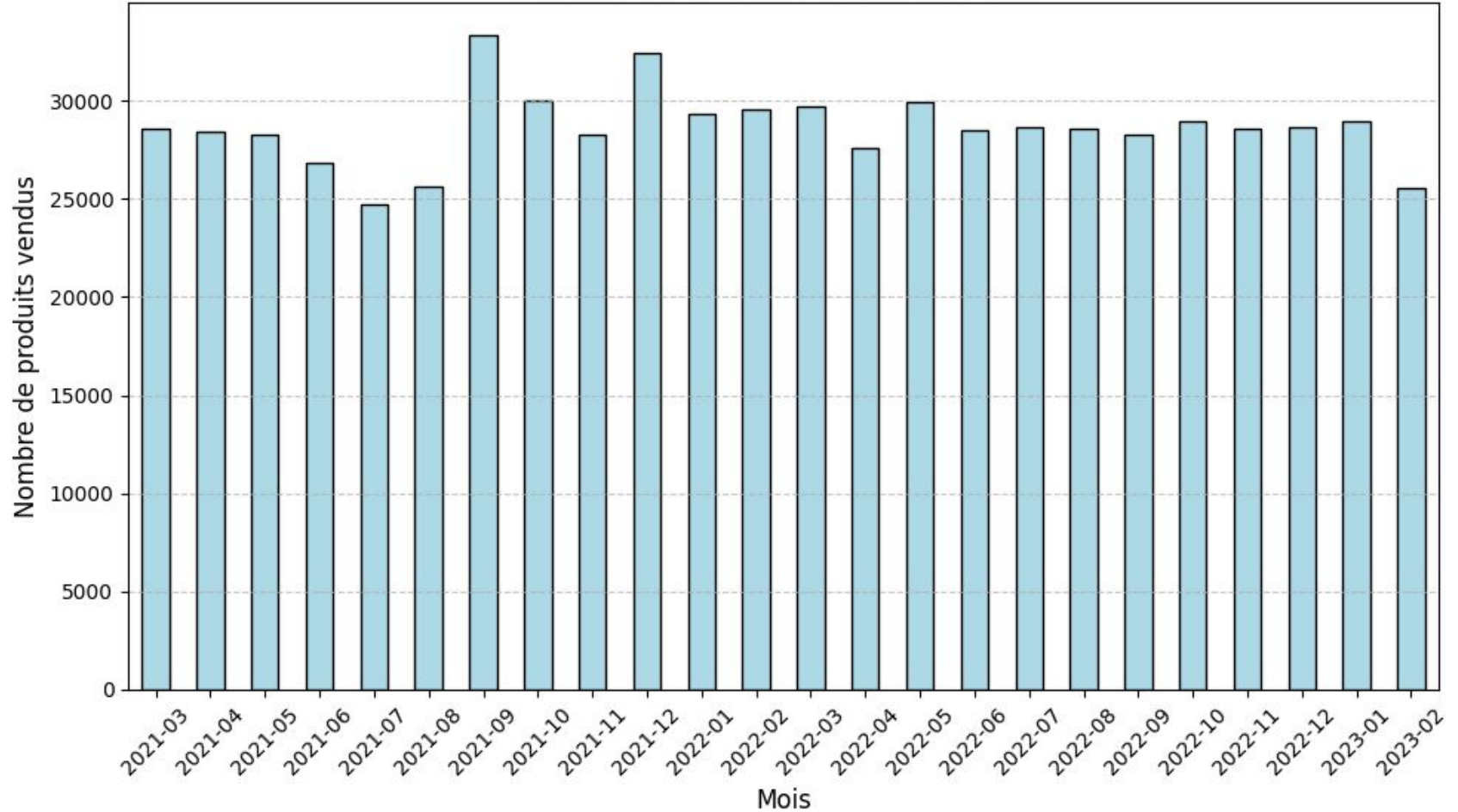
Nombre de transactions par an



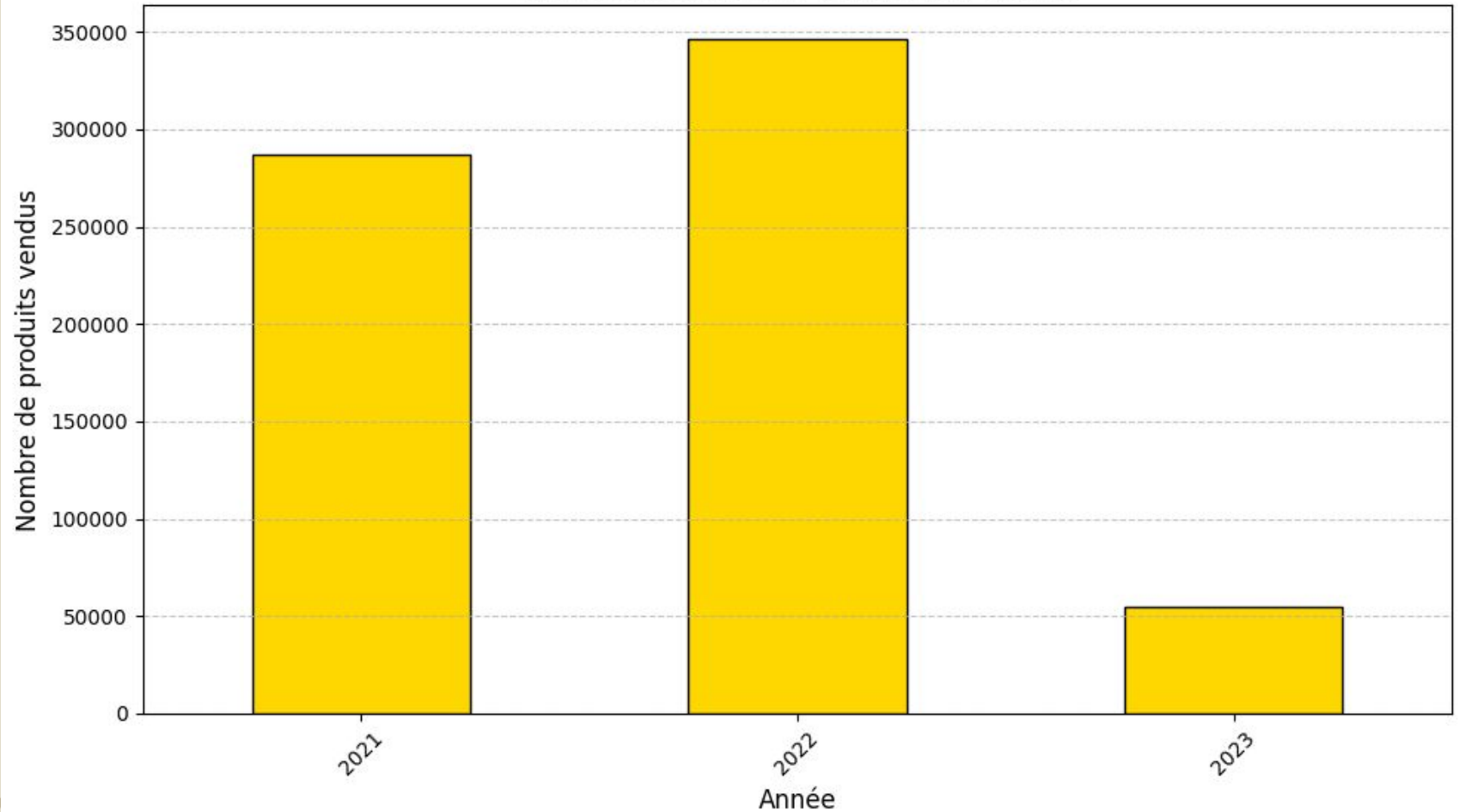


**Nombre de produit vendu par mois
puis par année**

Nombre de produits vendus par mois



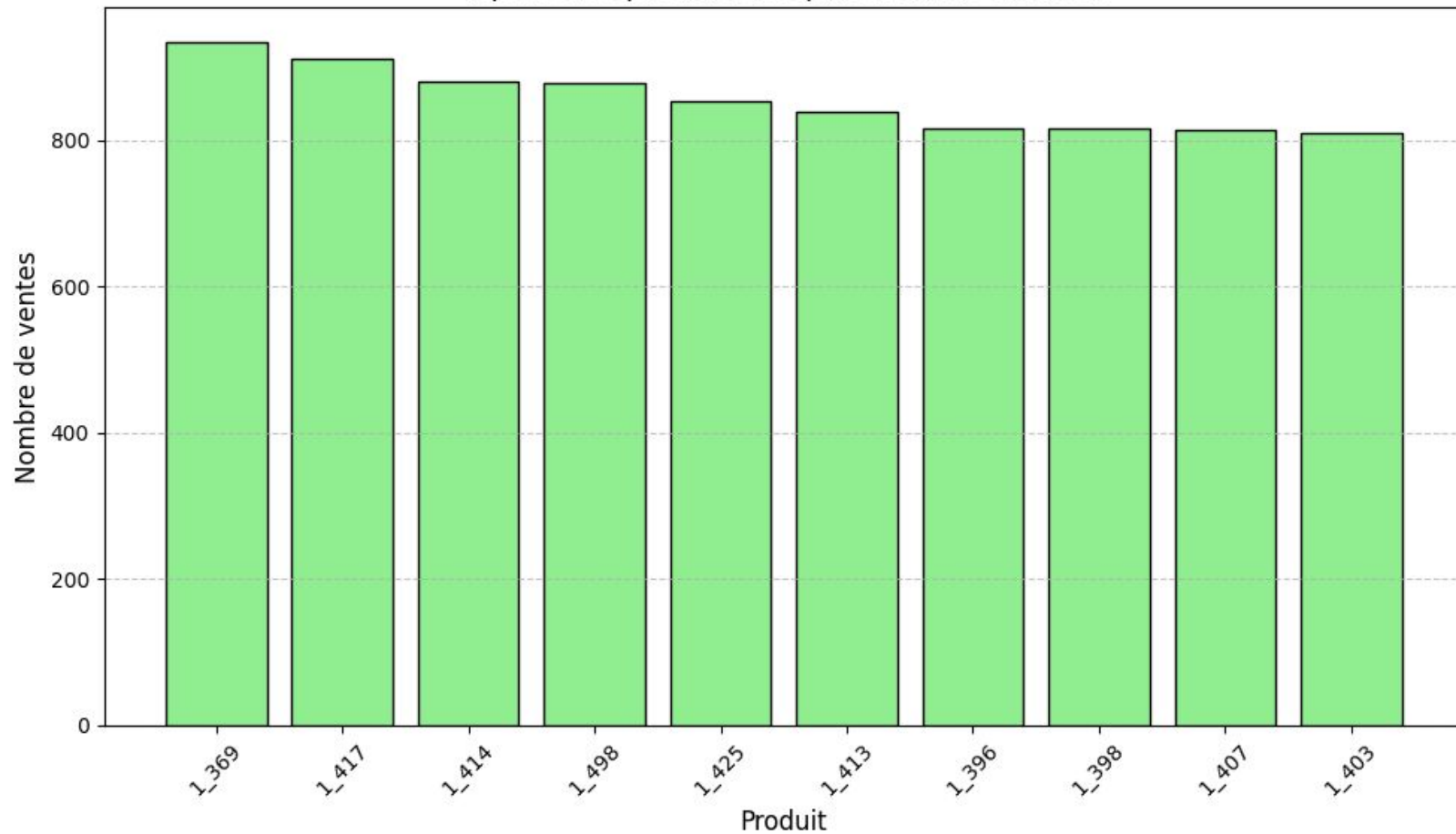
Nombre de produits vendus par année



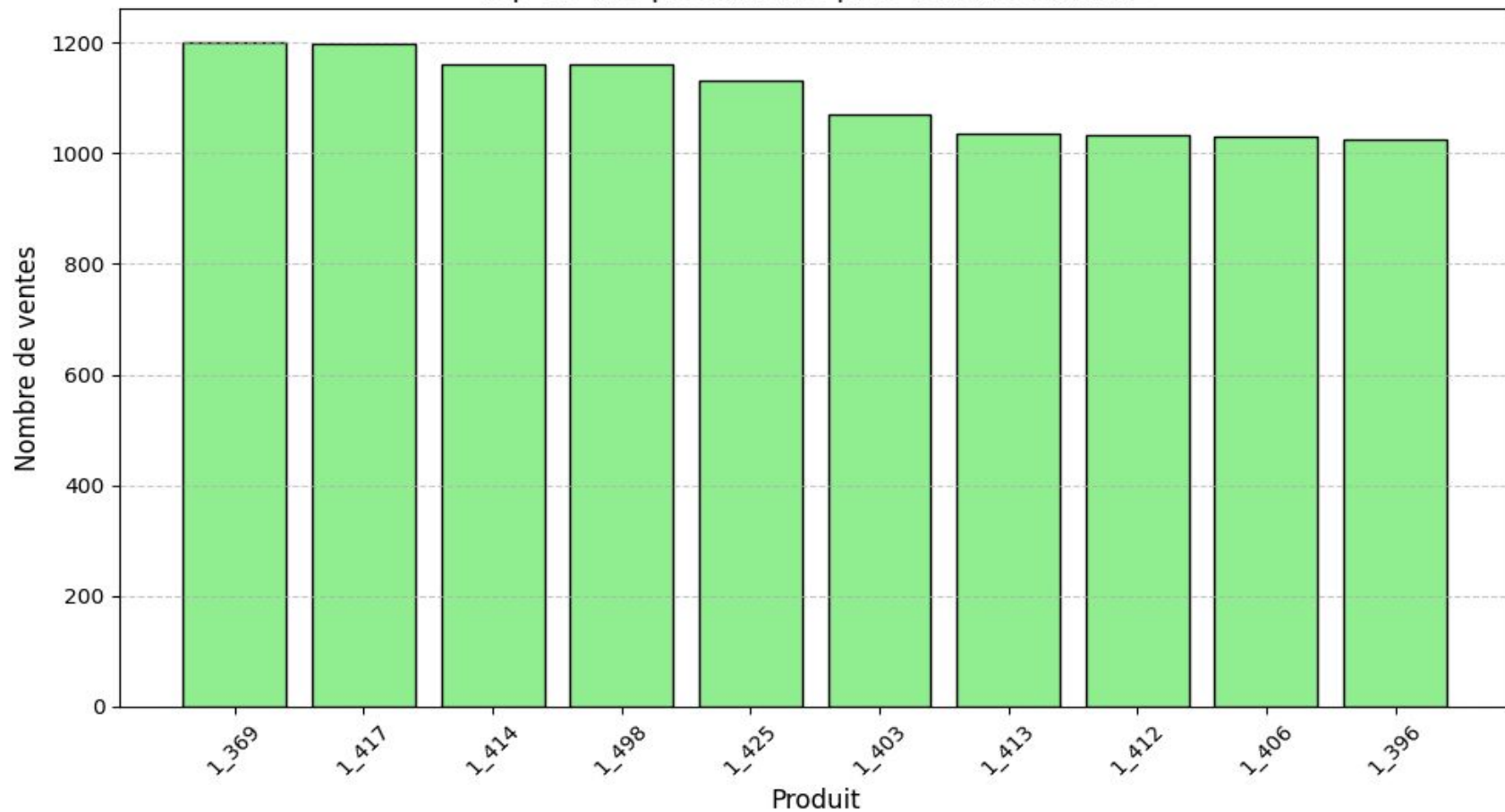


Top 10 des produits vendues par an

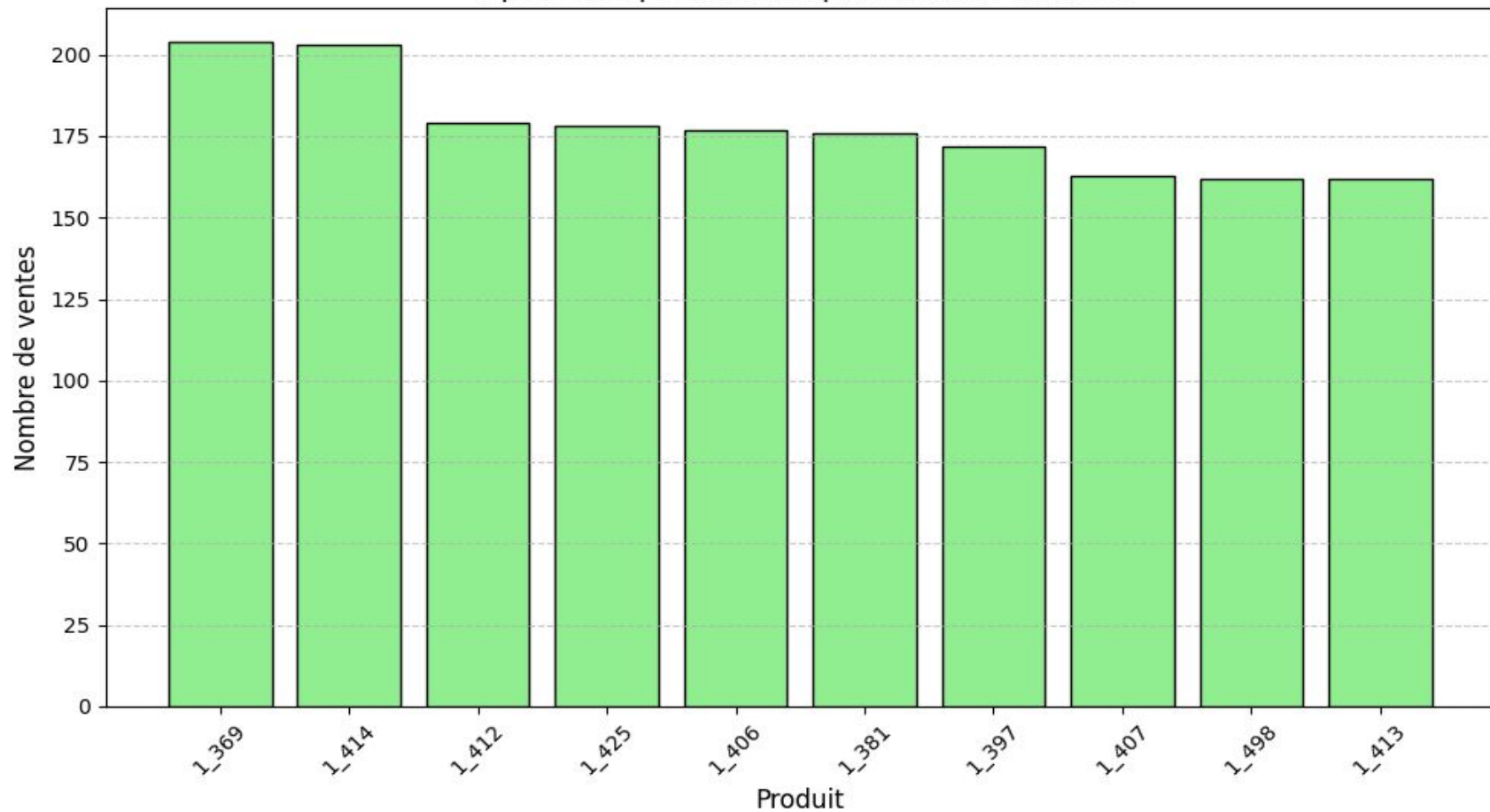
Top 10 des produits les plus vendus en 2021



Top 10 des produits les plus vendus en 2022



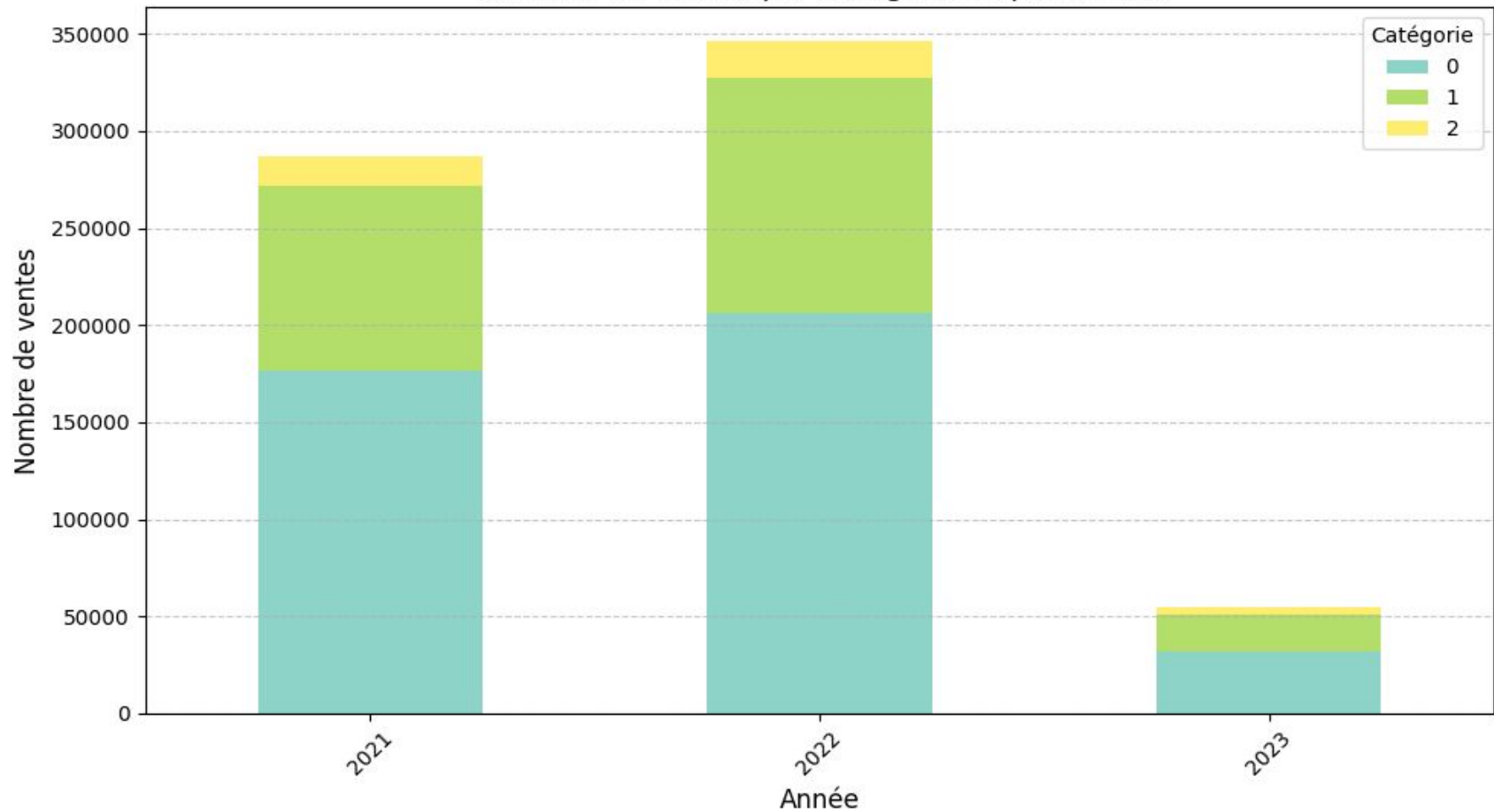
Top 10 des produits les plus vendus en 2023



Répartition des ventes par catégories

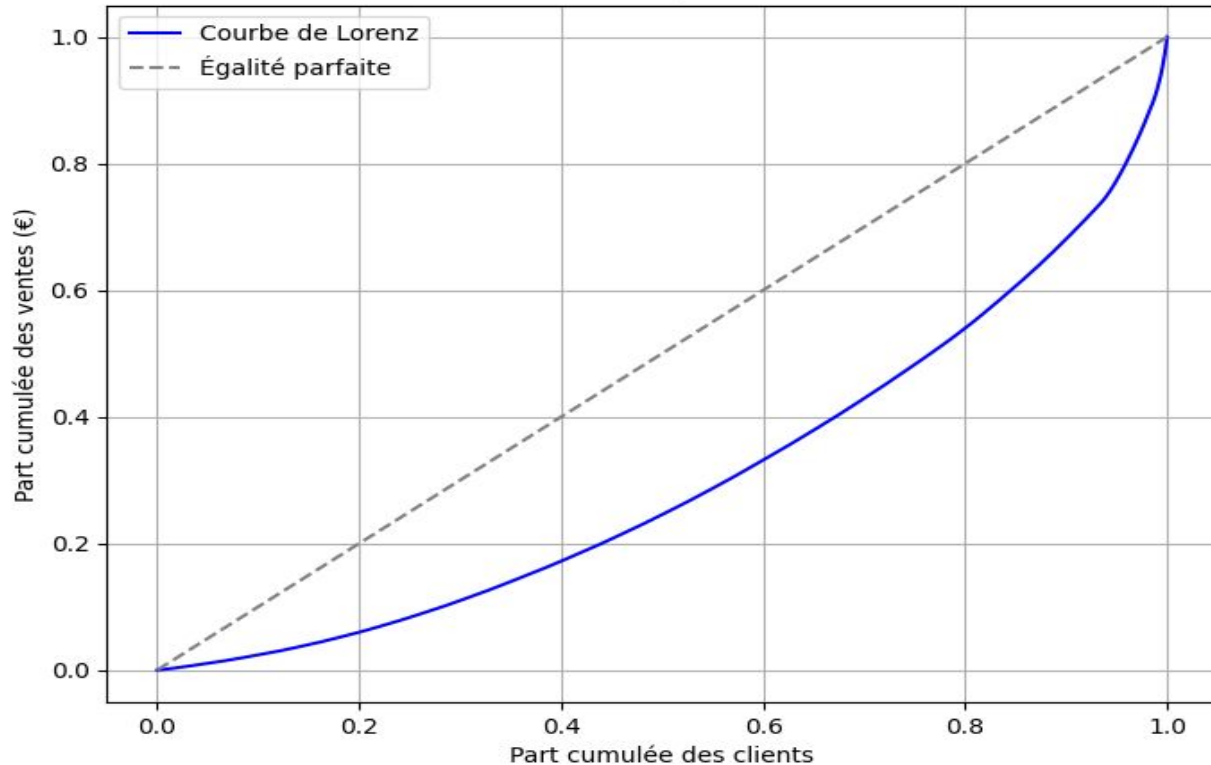
Année / Catégorie	0	1	2
2021	176 984	95 139	14 548
2022	206 152	121 412	18 816
2023	32 323	19 041	3 119

Nombre de ventes par catégorie et par année



Courbe de Lorenz

Courbe de Lorenz - Répartition des ventes



La courbe est très incurvée, ce qui indique que la majorité des clients achètent très peu, tandis qu'une minorité achète beaucoup (phénomène typique en commerce et marketing)

02

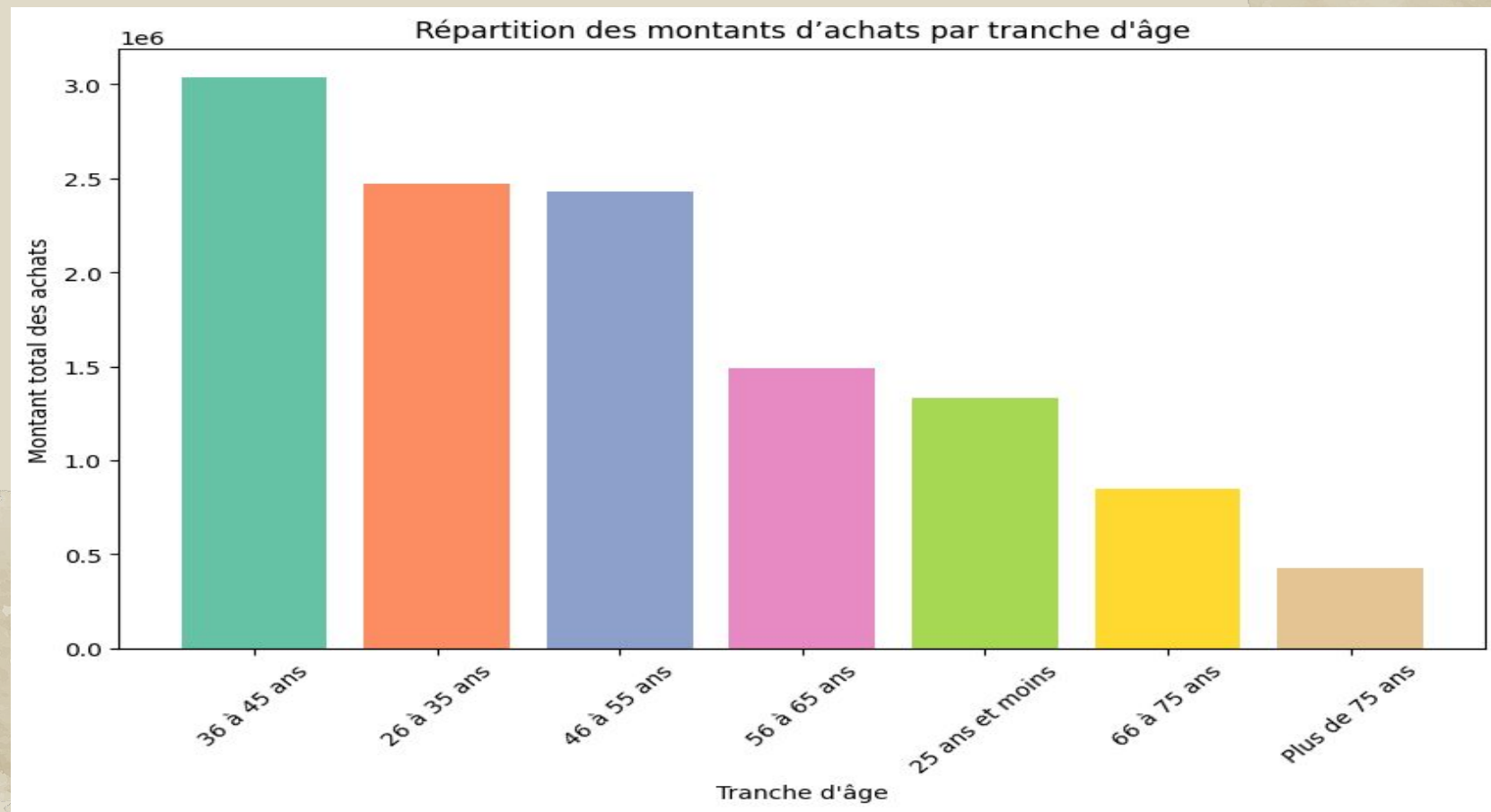
Corrélation des ventes



Lien entre le genre et la catégorie des livres

Sexe / Catégorie	0	1	2
Feminin	206 103	119 307	17 283
Masculin	209 356	116 285	19 200

Lien entre age et montants

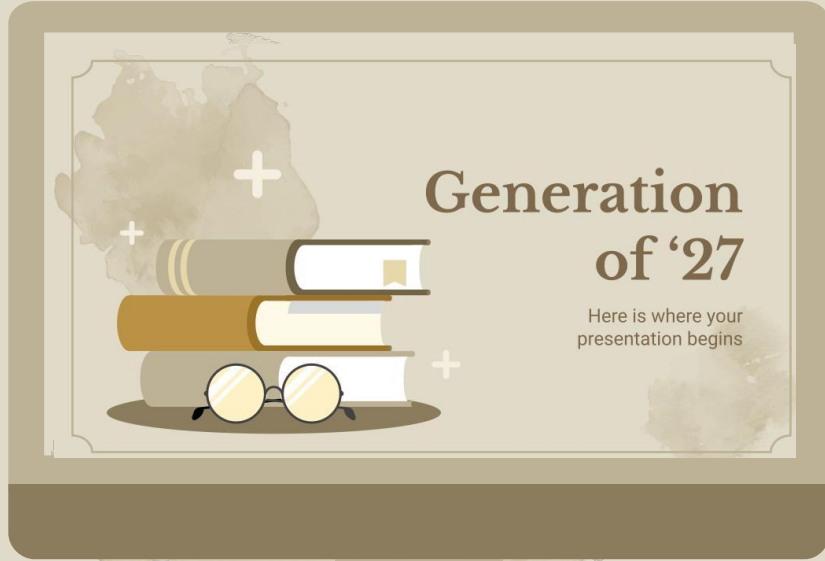


Lien entre interval d'achat et age

client_id	Intervale entre achat	AgeClient
c_1609	0.066297	45
c_3454	0.130880	56
c_4958	0.189351	26
c_6714	0.278351	57
c_8526	4.337349	60
c_1637	4.347561	37
c_669	4.401235	37
c_2265	4.445122	57
c_682	4.450000	51
c_5602	4.481250	36



Corrélation entre âge et fréquence d'achat



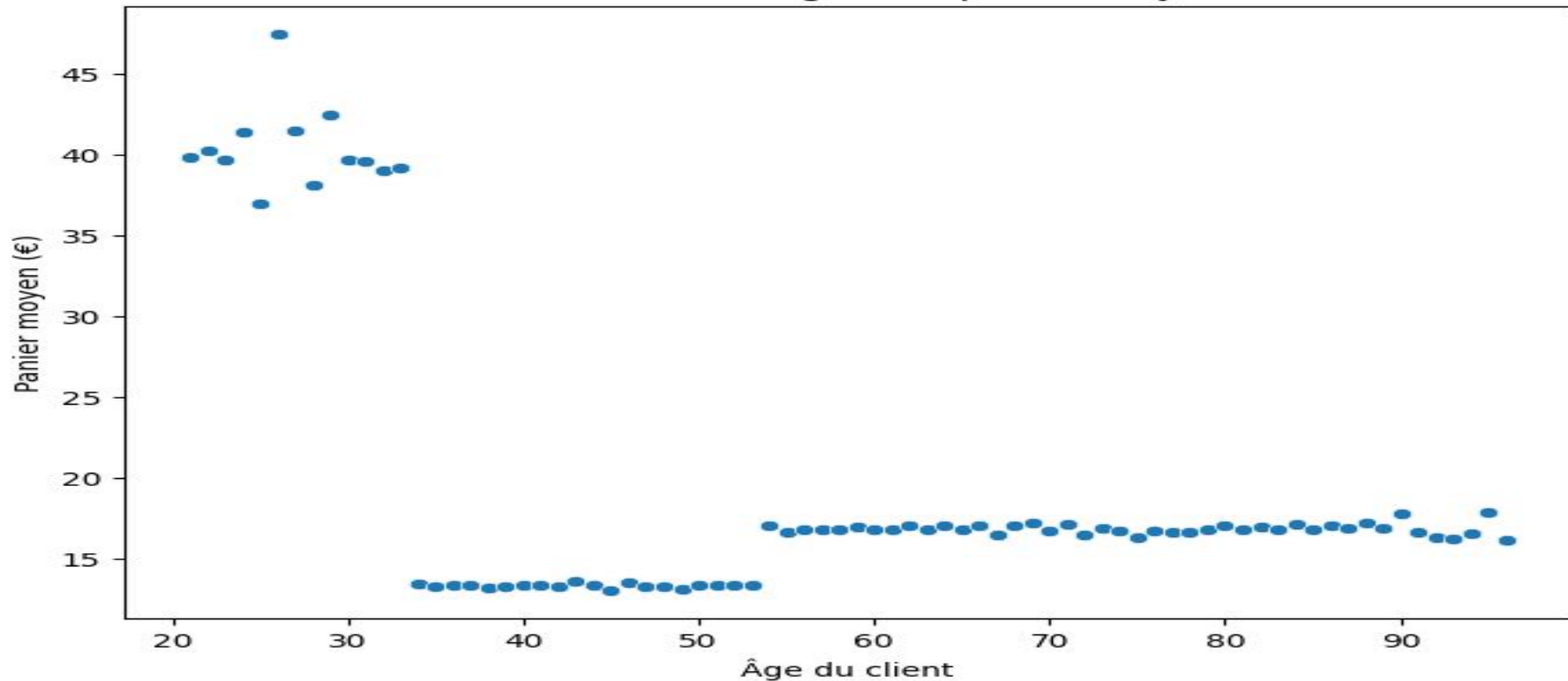
La corrélation de -0.149 indique une relation faible et négative entre l'âge des clients et l'intervalle moyen entre leurs sessions d'achat.

Cela signifie que, dans l'ensemble, à mesure que l'âge des clients augmente, l'intervalle entre leurs achats a tendance à augmenter légèrement.

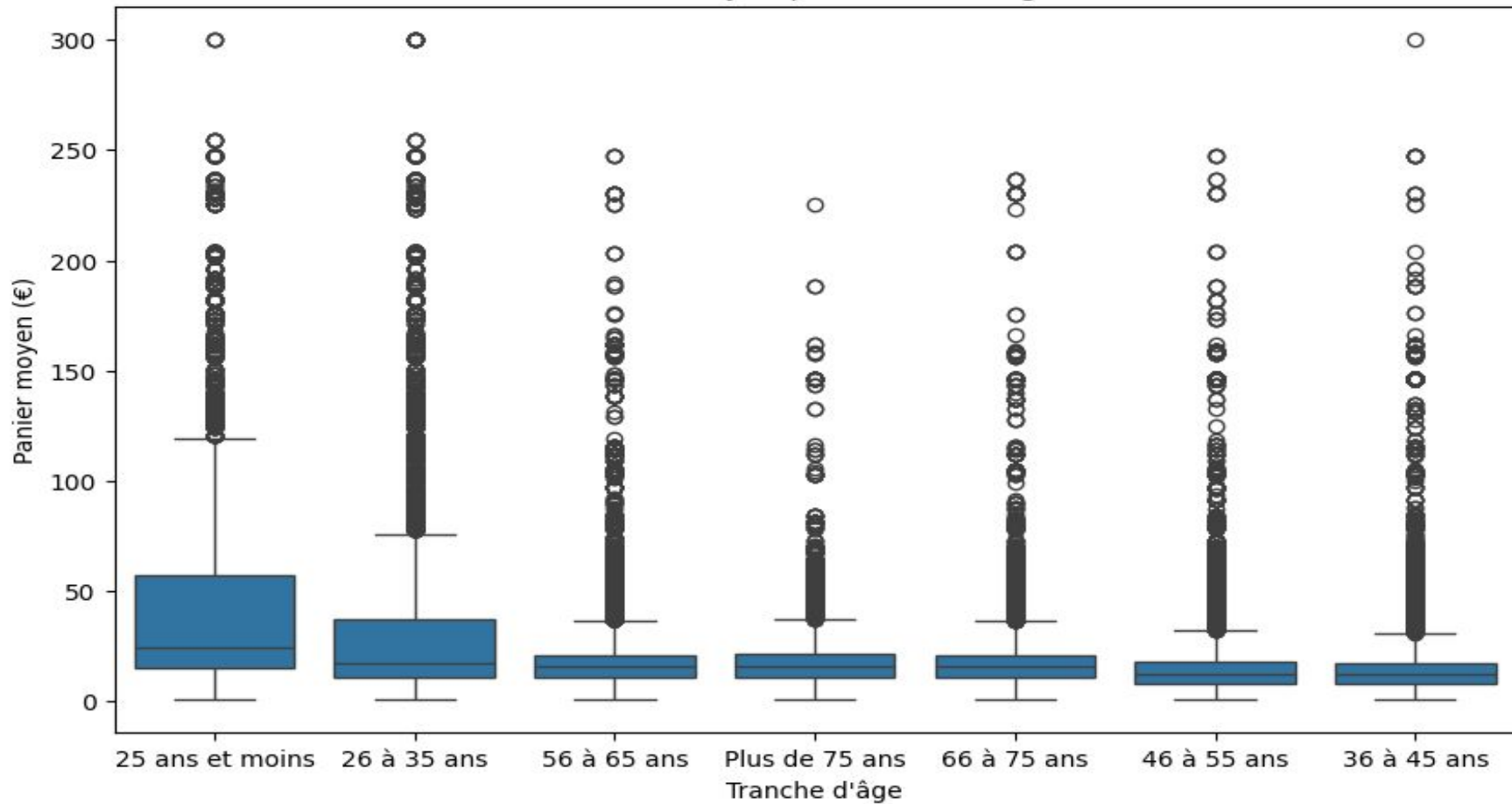
Cependant, la corrélation étant relativement faible, cela suggère que l'âge n'est qu'un facteur peu influent dans la détermination de la fréquence des achats.

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

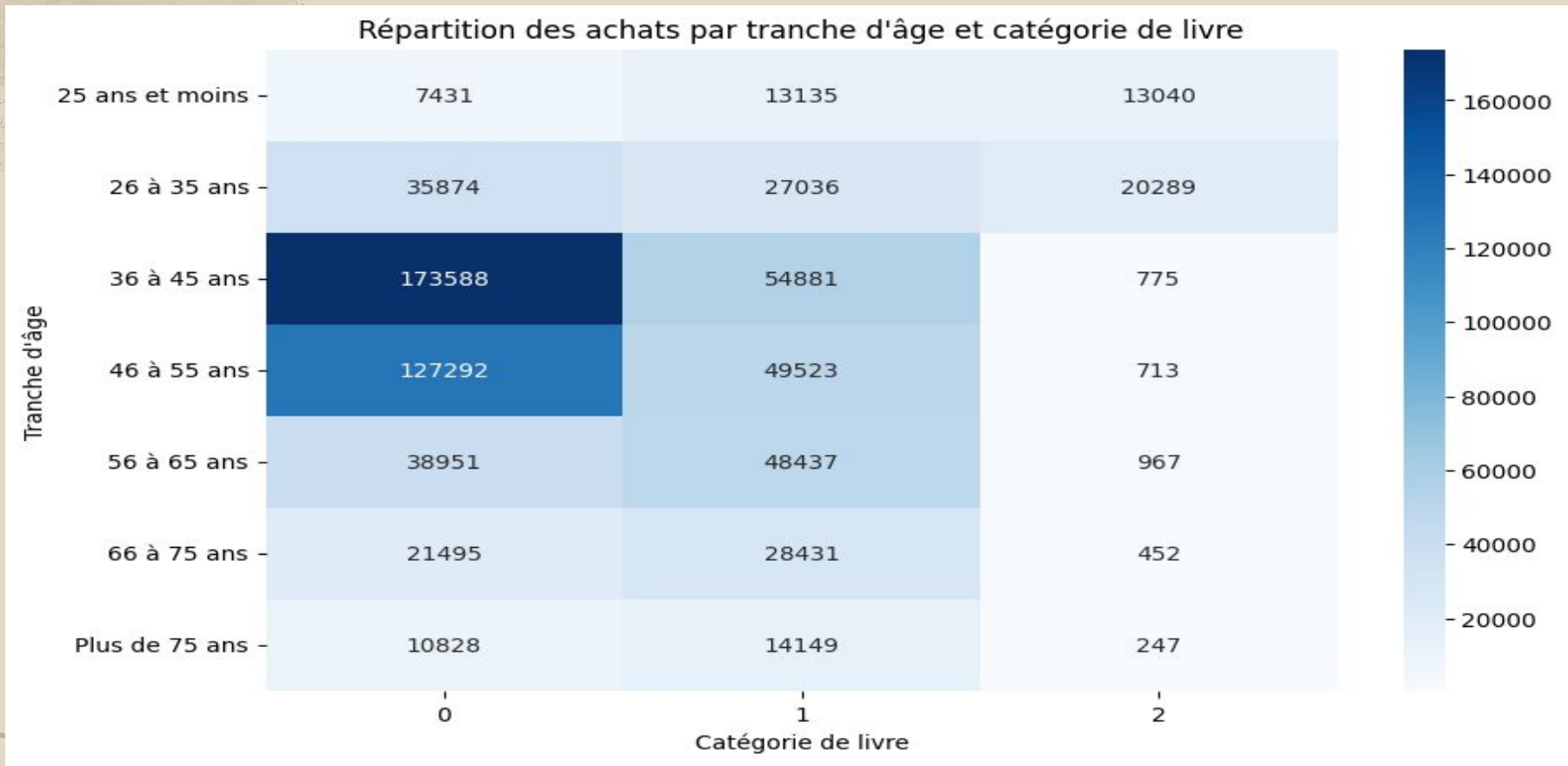
Relation entre l'âge et le panier moyen



Panier moyen par tranche d'âge



Lien entre l'âge des clients et les catégories de livres



Top 10 des clients

client_id	sex	AgeClient	trancheAge	birth	Dépenses
c_1609	m	45	36 à 45 ans	1980	326 039.89
c_4958	m	26	26 à 35 ans	1999	290 227.03
c_6714	f	57	56 à 65 ans	1968	153 918.60
c_3454	m	56	56 à 65 ans	1969	114 110.57
c_1570	f	46	46 à 55 ans	1979	5 285.82
c_3263	f	40	36 à 45 ans	1985	5 276.87
c_2140	f	48	46 à 55 ans	1977	5 260.18
c_2899	f	31	26 à 35 ans	1994	5 214.05
c_7319	f	51	46 à 55 ans	1974	5 155.77
c_7959	f	51	46 à 55 ans	1974	5 135.75

Corrélation entre sexe et catégorie

Calcul du Khi2

Sexe / catégorie	0	1	2
f	206103	119307	17283
m	209356	116285	19200

Statistique du khi² : 158.25417617304882

Plus la statistique khi² est élevée, plus l'écart entre les valeurs attendues et observées est grand → les variables sont probablement liées.

V de Cramer

V de Cramer : 0.0152

Avec V de Cramer = 0.0152, on est très proche de 0, ce qui signifie que la corrélation est quasi inexistante.

Le test du χ^2 donne une valeur élevée, mais c'est uniquement dû à la grande taille de l'échantillon.

V de Cramer est extrêmement faible → Les variables 'sex' et 'categ' sont quasiment indépendantes.

Conclusion : Il n'y a pas de véritable relation significative entre le sexe et la catégorie de produit acheté.

Corrélation entre prix et âge

Coefficient de corrélation de Pearson: -0.21

Le coefficient de corrélation de Pearson de -0.21 indique une corrélation faible et négative entre les dépenses et l'âge des clients.

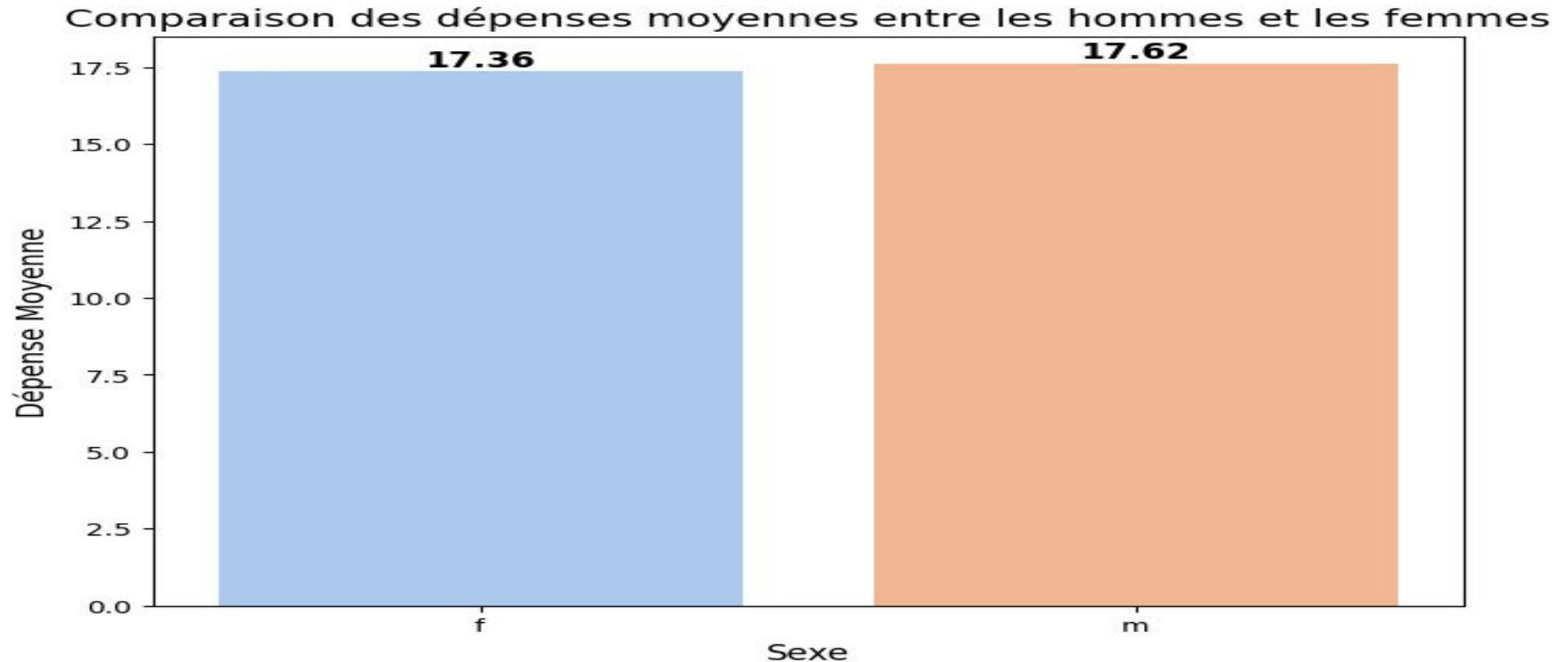
Test de student

Homme -> 344 841

Femme -> 342 693

Les résultats montrent qu'il y a une vraie différence entre les prix des hommes et des femmes. La valeur p est très proche de zéro, ce qui signifie que cette différence n'est pas due au hasard. En clair, les hommes et les femmes n'ont pas le même panier moyen, et c'est statistiquement prouvé.

Résultats analytiques



Résultats analytiques

Dépenses hommes : 6 076 819.42 €

Dépenses femmes : 5 950 843.68 €

👉 Ce que ça signifie ?

Oui, statistiquement, c'est "significatif" ($p < 0.05$).

Mais en pratique, la différence est insignifiante (0,26 € sur 17 €).

L'effet réel est minuscule → pas de vraie inégalité ou de tendance notable.

Axe d'amélioration pour 2023

1. Promotions ciblées

- Offres sur les produits peu vendus.
- Réductions personnalisées selon le profil client.

2. Fidélisation des clients

- Programme de fidélité pour les acheteurs réguliers.
- Relance des clients inactifs avec des offres spéciales.

3. Amélioration de la data

- Récupération des ventes physiques
- Collecte des avis / note client (produits ou des ventes)

Conclusions

- Les ventes sont en augmentations d'années en années.
- Il n'y a pas de réel lien entre les ventes et les personnes à qui vous vendez. Cependant des améliorations sont possibles sur la tranche d'âge des 25 ans et moins. En effet leur panier moyen est plus important que celui des autres. Il faudrait penser à augmenter le type d'offres pour cette catégorie d'âge.
- Il semble aussi que vous ayez des clients très importants , pourquoi ne pas les rencontrer et comprendre leur besoin pour avoir des offres spécifiques à eux.
- Il faudrait mettre en place un système de récupération de données sur les agences physiques pour comprendre la différence entre les ventes en lignes et celles-ci.



MERCI