

Analysez les ventes d'une librairie en Python

Parcours Data Analyste

MARCHAND Emmanuel – 08/2024

Agenda

Partie 1 : Analyse des indicateurs ventes

Partie 2 : Etude de corrélations

Conclusion et recommandations

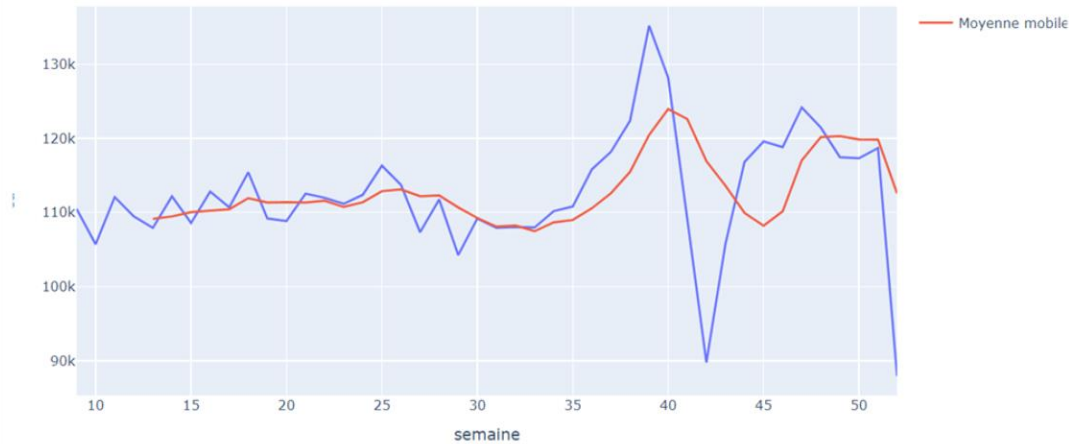


Analyse des indicateurs ventes

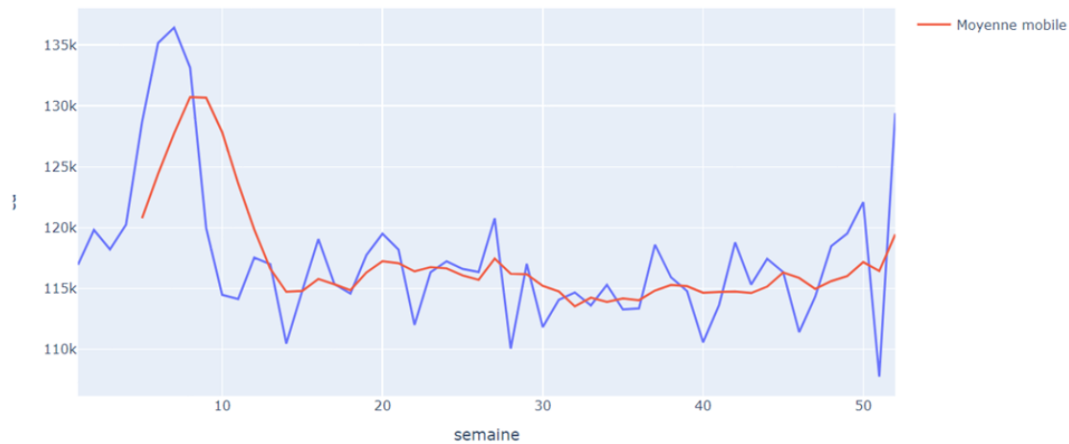


Evolution du Chiffre d'affaires

Chiffre d'affaire 2021

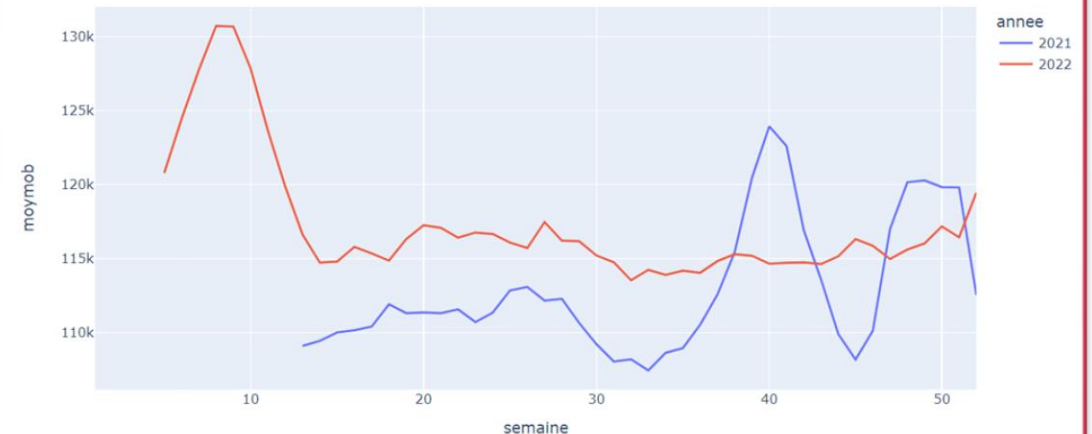


Chiffre d'affaire 2022



- L'année 2023 n'est pas représentative avec seulement deux mois.
- 2021 a démarré lentement avec un pic d'activité pour les achats de la rentrée puis un creux et une reprise pour les achats de Noël
- 2022 a connu un premier trimestre exceptionnel avec une stabilisation du CA sur le reste de l'année qui est meilleure que 2021.
- 2022 pas de pic d'activité pour la rentrée scolaire.

Comparaison moyenne mobile 2021 -2022



Chiffre d'affaires par catégories

Chiffre d'affaire par catégorie sur l'année 2021

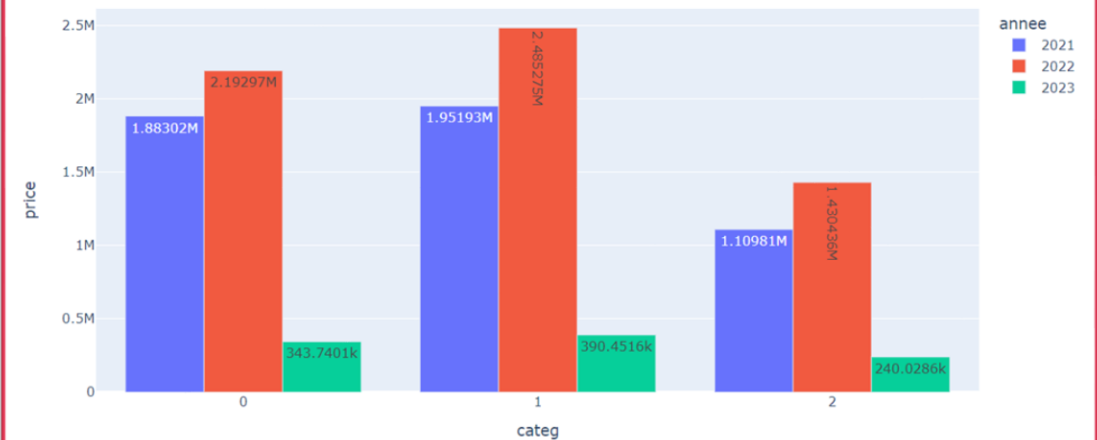


- La catégorie **Cat1** génère le plus de CA suivi par la Cat0 et la Cat2.
- La catégorie **Cat2** est nettement en dessous des deux autres.
- **2021** est une année plus chaotique sur l'ensemble des catégories, alors que 2022 se stabilise

Chiffre d'affaire par catégorie sur l'année 2022

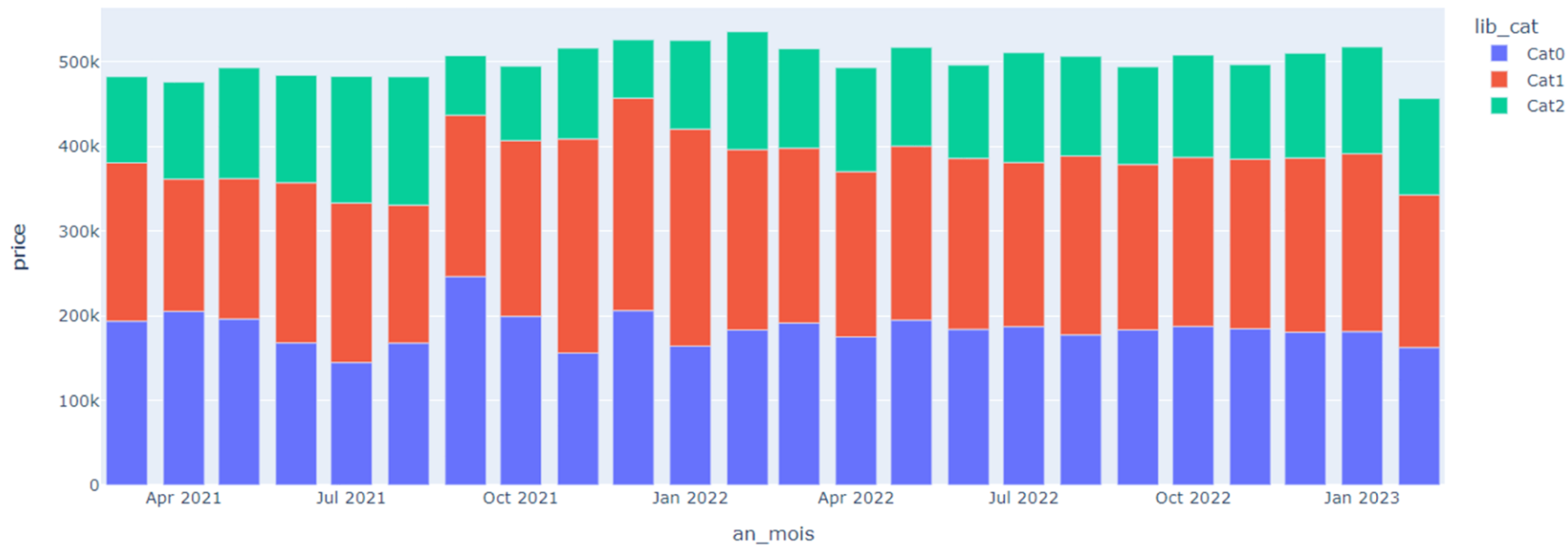


Répartition du CA par catégorie et par année



Répartition du CA.

Répartition du CA par catégorie



Relative stabilité du CA pour la catégorie 0.

Les catégories 1 et 2 sont les plus fluctuantes.

Les catégories 0 et 1 portent le CA.

Evolution des transactions et du nombre de client par mois.

Nombre de clients par mois par année

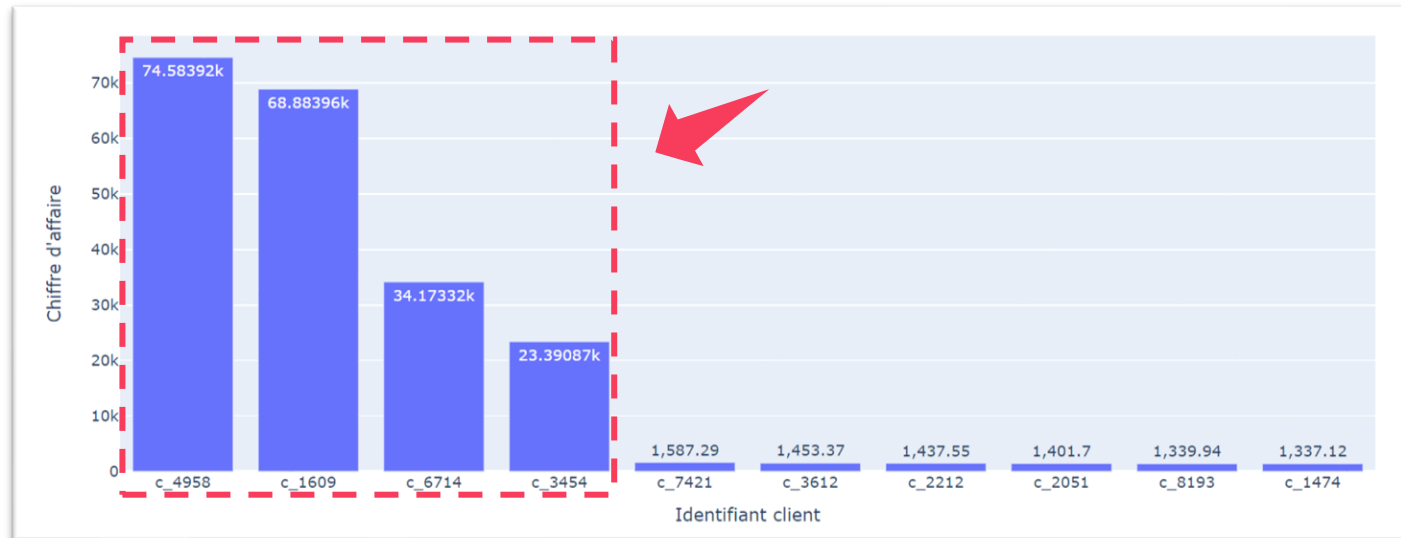


Attention une tendance à la baisse en fin d'année qui se confirme début 2023, que ce soit en session ou en nombre de clients.

Nombre de sessions/transactions par mois par année



TOP de la société



TOP 10 des clients par CA

Quatre clients concentrent le CA, ce sont des clients B2B



TOP 10 des produits par CA

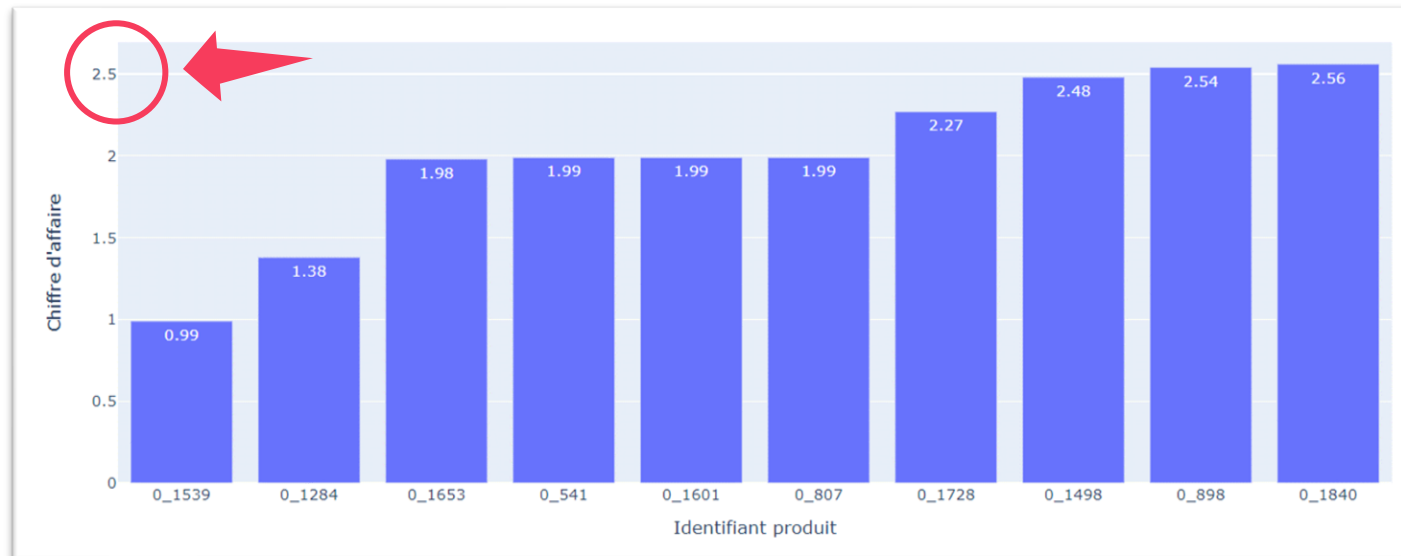
Un produit se distingue nettement des autres.

FLOP de la société



FLOP 10 des clients par CA

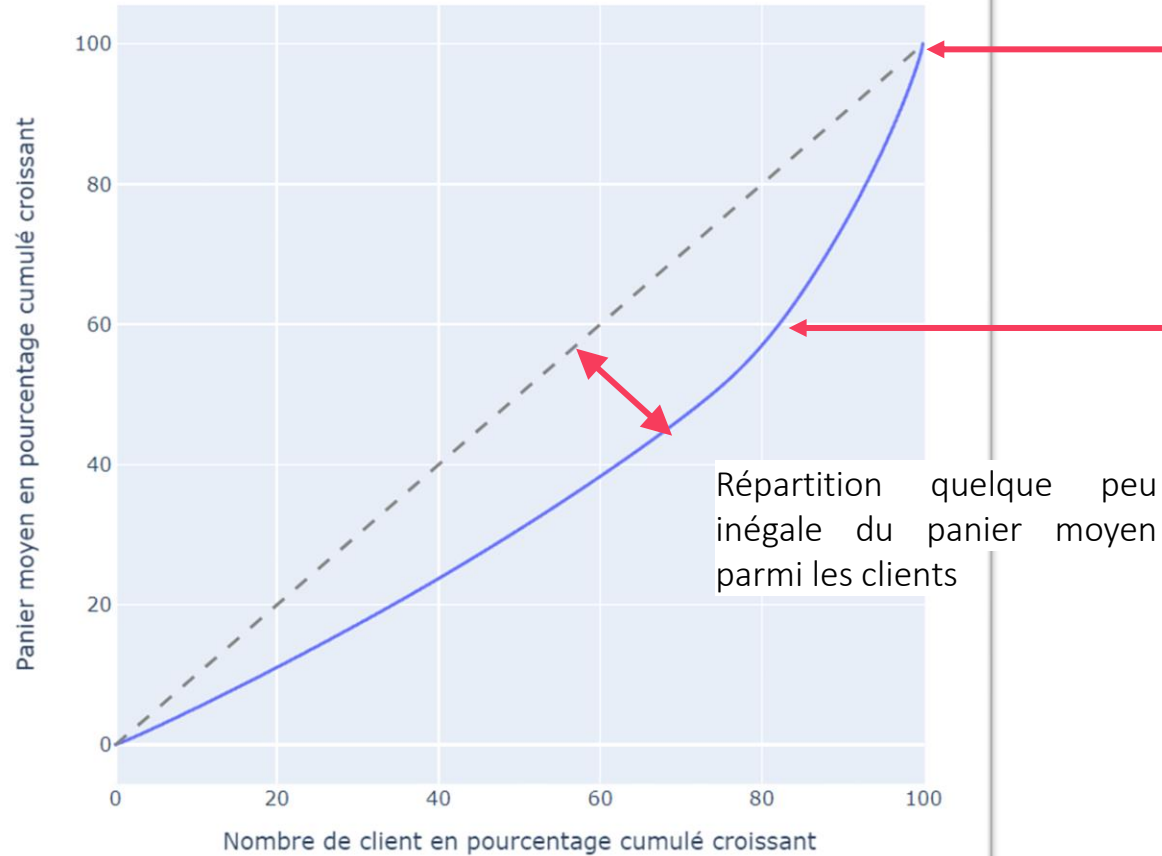
Les cinq plus mauvais client ont un panier inférieur à 15 euros.



FLOP 10 des produits par CA

Le flop 10 a un CA inférieur à 3 euros.

Courbe de LORENZ panier moyen par client



20% des clients concentrent
40% des revenus

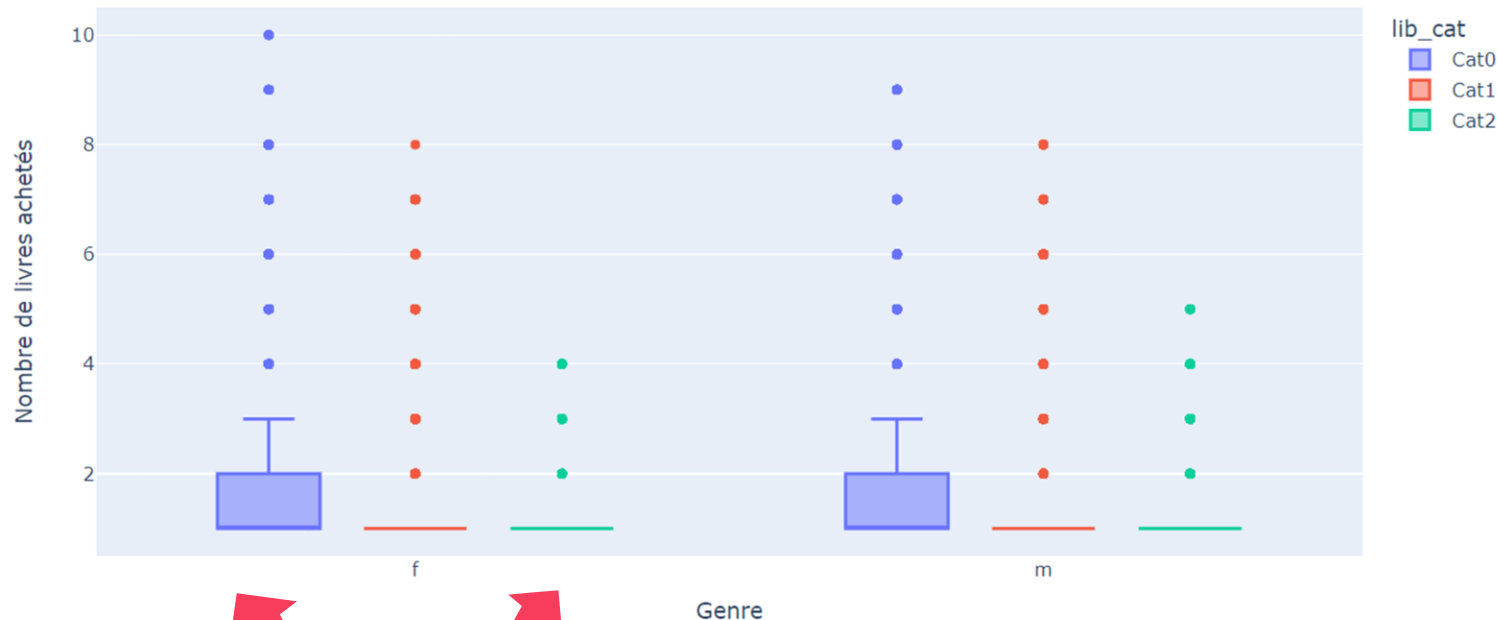
Coefficient de Gini: 0.2996256793075046

La répartition du panier moyen est légèrement inégalitaire.

Etude de corrélations

Corrélation entre le genre des clients et les catégories des livres achetés

Corrélation entre le genre et la catégorie achetée



Ces deux catégories se distinguent selon le genre

Test : Test du KHI-deux

C'est un test d'hypothèse utilisé pour déterminer s'il existe une relation entre deux variables catégorielles.

Hypothèses

H0 : le genre n'influe pas sur la catégorie de livre achetée

Ha : la catégorie de livre achetée est liée au genre

Table de contingence :

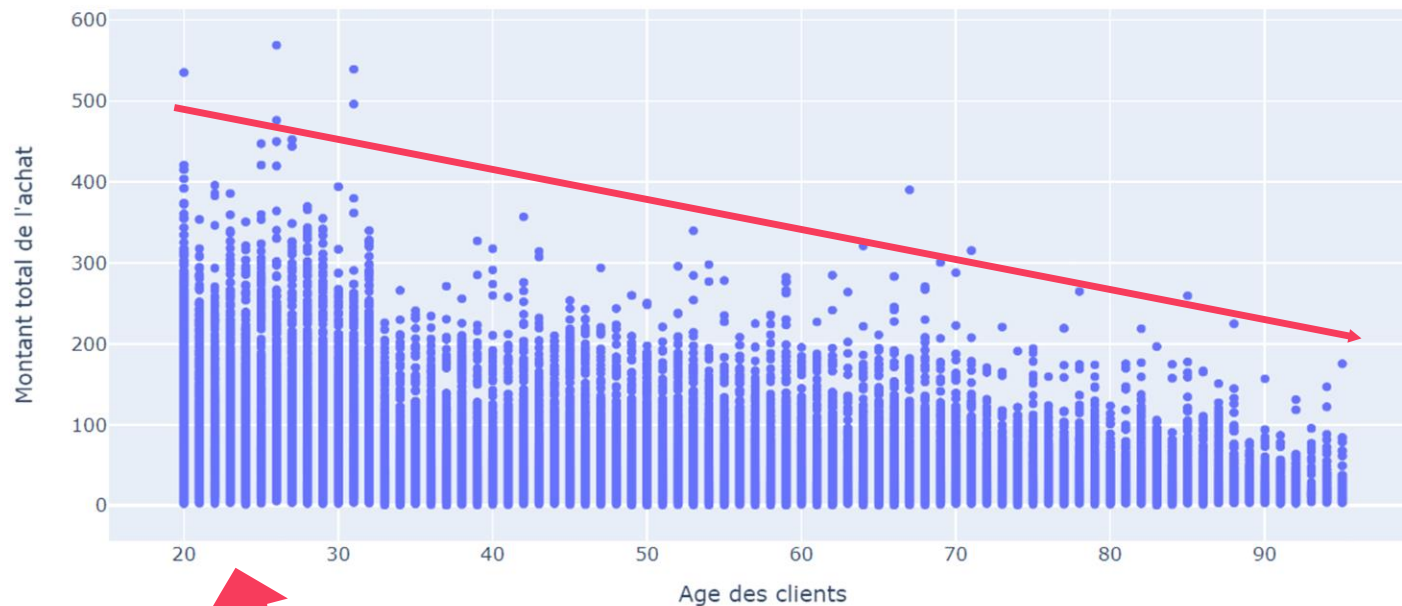
	categ	Femme	Homme
0	0	200793	186488
1	1	115721	104884
2	2	16980	15868

Résultat du test : La pvalue est inférieure à 0,05, donc H0 est rejeté et Ha validé

```
Chi2ContingencyResult(statistic=46.64333908608719, pvalue=1.8093433648730353e-09, dof=4,
expected_freq=array([[1.81329157e+00, 2.01573953e+05, 1.85705234e+05],
[1.03290117e+00, 1.14822115e+05, 1.05782852e+05],
[1.53807256e-01, 1.70979324e+04, 1.57519138e+04]]))
```

Corrélation entre l'âge des clients et le total des achats

Corrélation entre l'âge client et le total achat



Plus l'âge est faible plus le montant des achats est élevé

Test : Corrélation de Pearson

Ce test permet de mesurer la relation linéaire entre deux variables , et permet de déterminer l'intensité et la direction.

Hypothèses

H_0 : le montant des achats n'est pas corrélé à l'âge

H_a : le montant des achats est corrélé à l'âge

Force de corrélation :

Quantité de r	Force de la corrélation
$0.0 < 0.1$	pas de corrélation
$0.1 < 0.3$	faible corrélation
$0.3 < 0.5$	corrélation moyenne
$0.5 < 0.7$	corrélation élevée
$0.7 < 1$	corrélation très élevée

Résultat du test : La pvalue est inférieure à 0,05, donc H_0 est rejeté et H_a validé

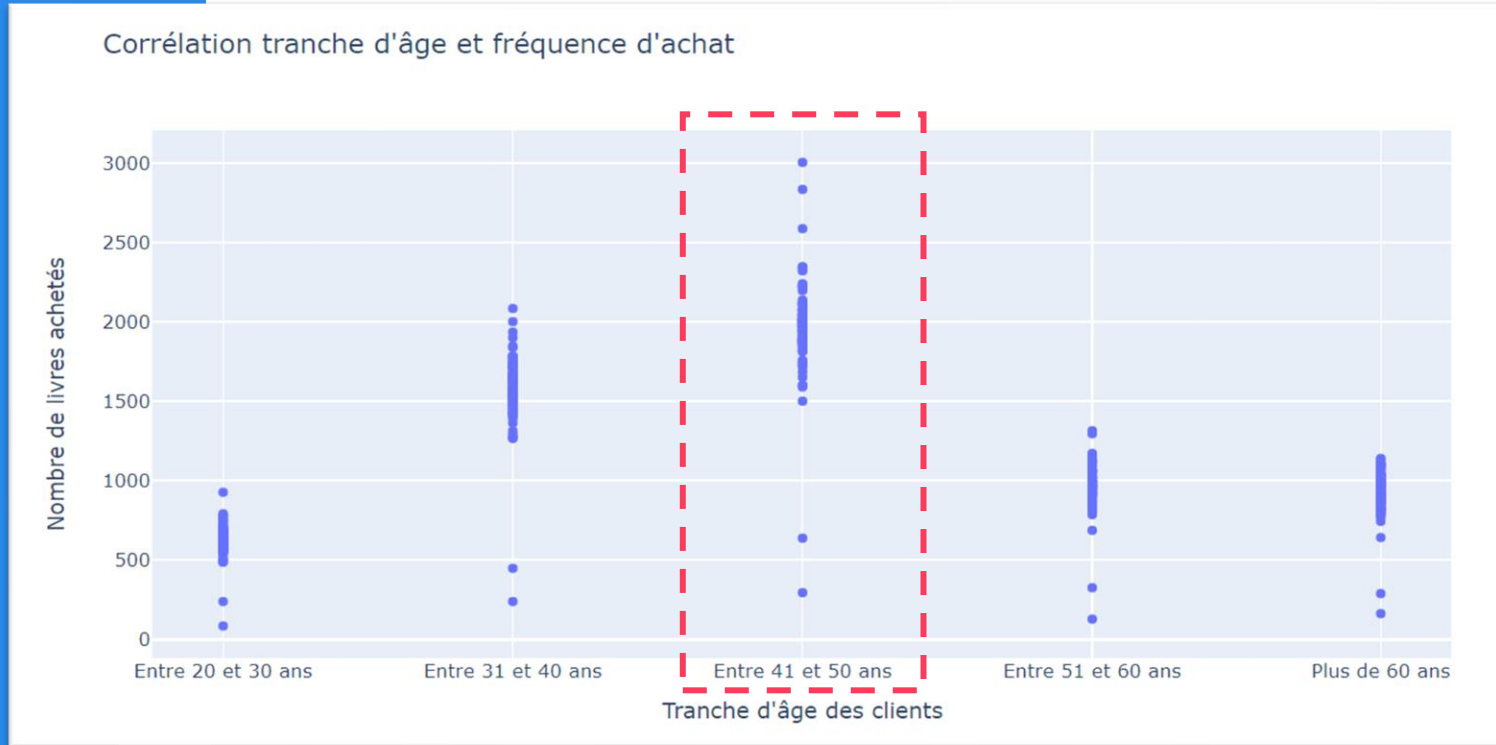
`PearsonRResult(statistic=-0.327090874403226, pvalue=0.0)`

Le coefficient de corrélation de Pearson est **négatif**, on a donc une **corrélation négative**

Le coefficient de corrélation est de force moyenne (-0,32)

Corrélation entre la tranche d'âge et la fréquence d'achat

Le test ANOVA confirme cette corrélation.



Test : ANOVA

Une analyse de la variance (ANOVA) permet de vérifier s'il existe des différences statistiquement significatives entre plus de deux échantillons. Ici nous avons cinq groupes, caractérisés par des tranches d'âge.

Hypothèses

H0 : La fréquence d'achat n'est pas corrélée à la tranche d'âge

Ha : La fréquence d'achat est corrélée à la tranche d'âge

Hypothèses de l'analyse de variance

H0 : la valeur moyenne de tous les groupes est la même.

Ha : il existe des différences entre les valeurs moyennes des groupes.

L'analyse des cinq group (cette tranche d'âge se distingue nettement sur la fréquence d'achat. moyennes.

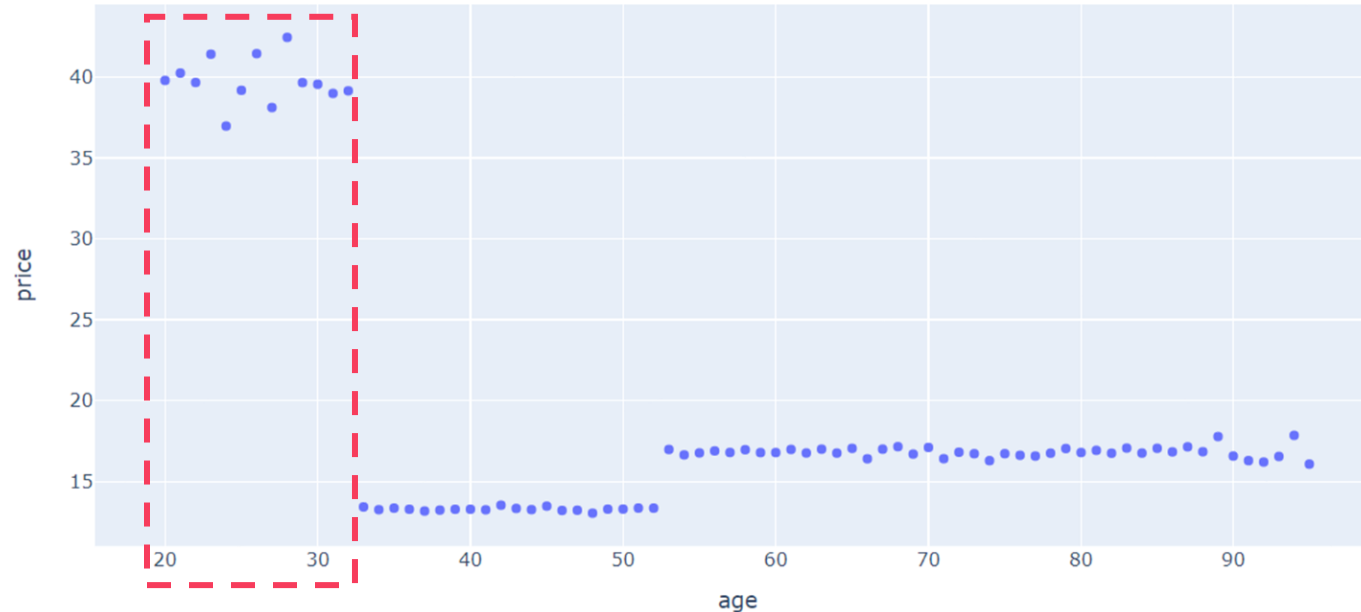
Résultat du test : La pvalue est inférieure à 0,05, donc H0 est rejeté et Ha validé

`F_onewayResult(statistic=834.4711781228216, pvalue=6.840422978100761e-226)`

Au moins une des moyennes diffère, comme il n'y a pas égalité des moyennes on peut affirmer que la fréquence d'achat est corrélée à la tranche d'âge.

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

Corrélation entre l'âge et la taille du panier moyen



Cette tranche d'âge se distingue nettement sur la taille du panier moyen.

Test : Corrélation de Pearson

Ce test permet de mesurer la relation linéaire entre deux variables , et permet de déterminer l'intensité et la direction.

Hypothèses

H_0 : le panier moyen n'est pas corrélé à l'âge

H_a : le panier moyen est corrélé à l'âge

Force de corrélation :

Quantité de r	Force de la corrélation
$0.0 < 0.1$	pas de corrélation
$0.1 < 0.3$	faible corrélation
$0.3 < 0.5$	corrélation moyenne
$0.5 < 0.7$	corrélation élevée
$0.7 < 1$	corrélation très élevée

Résultat du test : La pvalue est inférieure à 0,05, donc H_0 est rejeté et H_a validé

PearsonRResult(statistic=-0.5443989912175686, pvalue=3.7068907776337503e-07)

Le coefficient de corrélation de Pearson est **négatif**, on a donc une **corrélation négative**
Le coefficient de corrélation est de force **élevée** (-0,544)

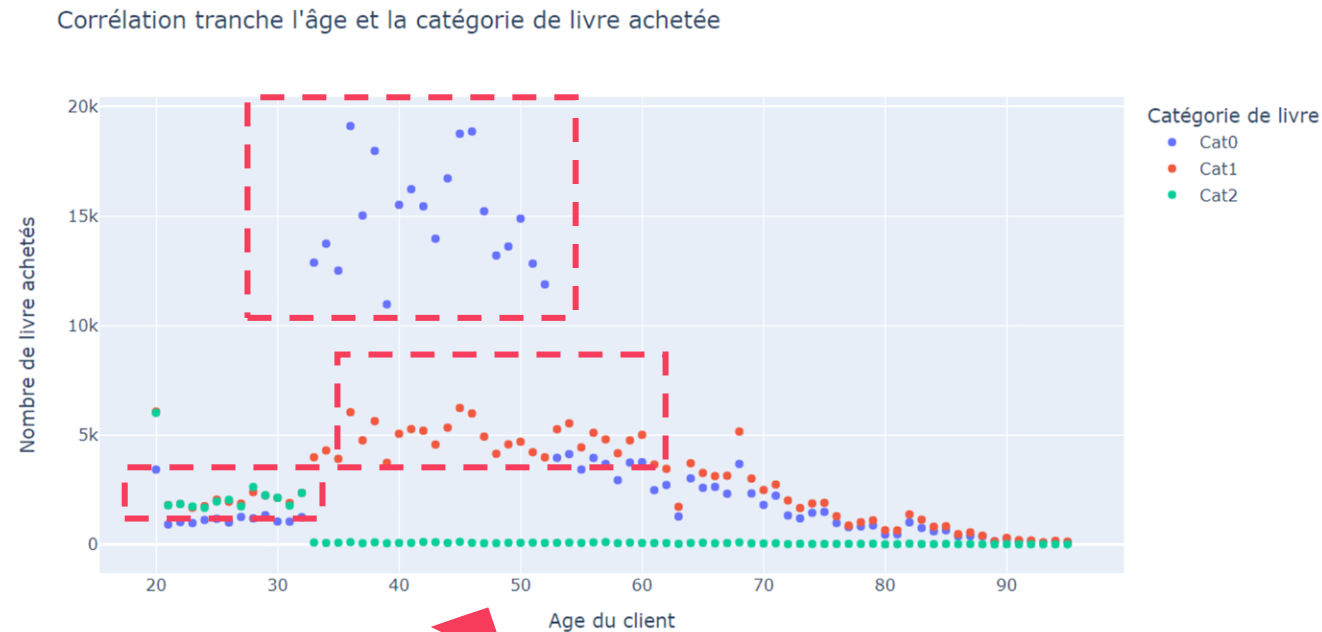
Corrélation entre l'âge des clients et la catégorie des livres achetés

Test : Test du KHI-deux

Hypothèses

H_0 : l'âge n'influe pas sur la catégorie de livre achetée

H_a : la catégorie de livre achetée est liée à l'âge



index	age	Cat0	Cat1	Cat2
0	20	3418	6067	6012
1	21	909	1798	1784
2	22	1016	1850	1845
3	23	974	1676	1728
4	24	1114	1744	1671
...
71	91	142	193	1
72	92	132	178	2
73	93	77	91	2
74	94	75	161	2
75	95	78	123	1

Table de
contingence :

On identifie trois tendances majeures :

- Cat0 est le plus achetée par les 30-55 ans
- Cat1 est plutôt achetée par les 35-60 ans
- Cat2 achetée par les 20-30 ans

Résultat du test : La pvalue est inférieure à 0,05, donc H_0 est rejeté et H_a validé

```
Chi2ContingencyResult(statistic=306412.61732033, pvalue=0.0, dof=225, expected_freq=array([[1.05113734e+02,
9.31545809e+03, 5.30631927e+03, 7.90108906e+02],
[3.05647461e+01, 2.70872894e+03, 1.54296014e+03, 2.29746174e+02],
[3.20618226e+01, 2.84140382e+03, 1.61853510e+03, 2.40999256e+02],
[2.98128209e+01, 2.64209132e+03, 1.50500168e+03, 2.24094174e+02],
[3.08424843e+01, 2.73334283e+03, 1.55698084e+03, 2.31833850e+02],
```




Conclusions

Conclusions

- La catégorie 0 est la plus achetée et par une clientèle féminine.
- La catégorie 0 a les plus de livres achetés, et cela par les 30 / 55ans
- La tranche des 20 / 30 ans rapportent le plus de revenus
- La tranche des 40 / 50 ans achètent en plus grande quantité.
- La tranche des 20 / 30 ans a le panier moyen le plus élevé.



Recommandations

- Augmenter l'effort marketing sur la catégorie 0 pour en faire le cœur d'activité de la marque
- Fidéliser la tranche des 20 / 30 ans : programme VIP, journée privilèges, goodies.
- Travailler sur la tranche des 40 / 50 ans pour augmenter leur panier moyen.
- Supprimer du catalogue des produits très peu vendus ou à faible revenu.



MERCI

