



ANALYSE DES VENTES ET COMPORTEMENTS CLIENTS

Librairie Lapage

Contexte et objectif :

- Vente en ligne ouverte il y a deux ans
- Double objectif :
 - *Analyse des indicateurs de performance*
 - *Analyse du comportement client*

Méthodologie

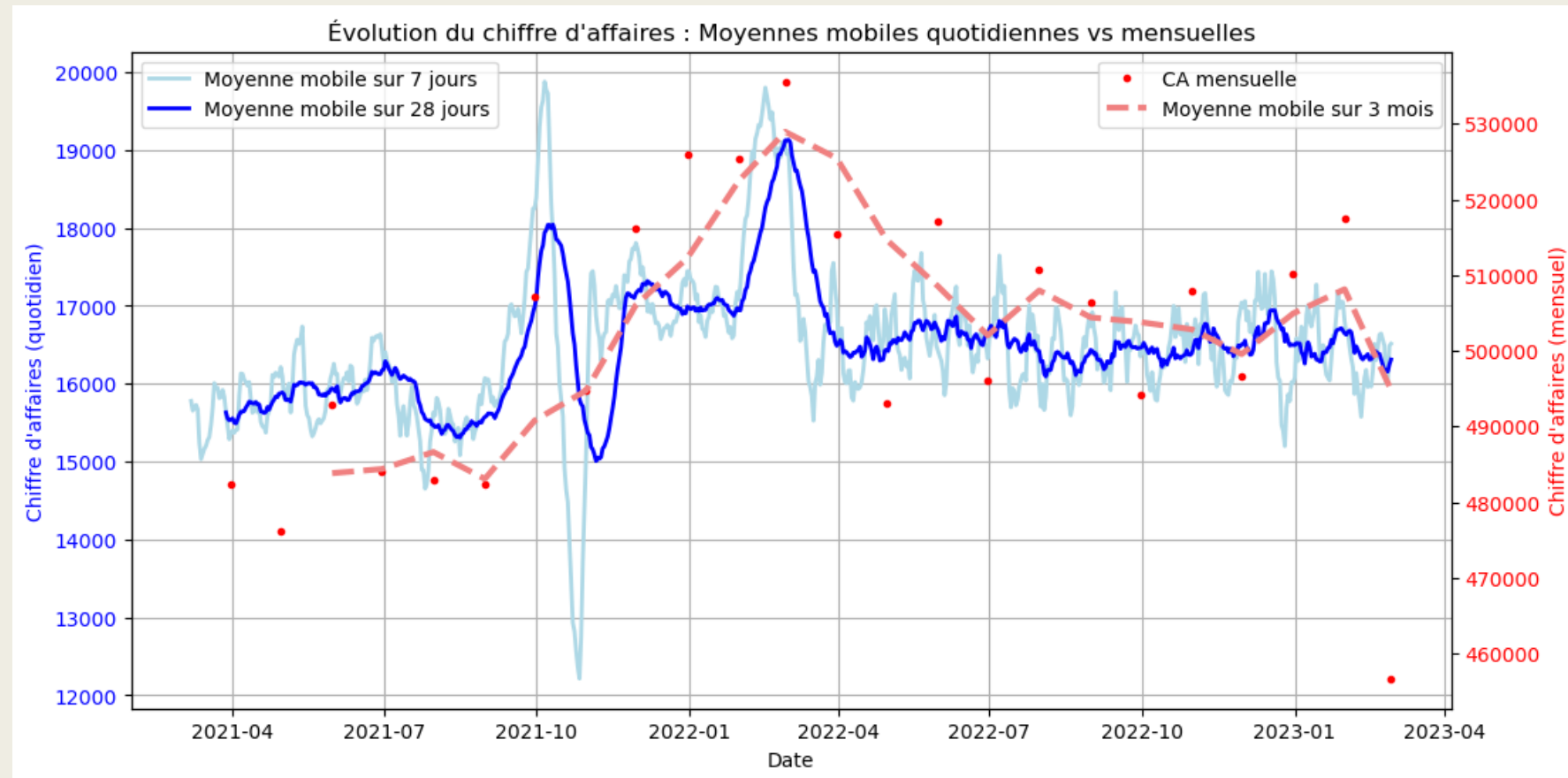
- 3 sources : fichiers csv clients, produits, transactions
- Traitement sur la données :
 - *Nettoyage (suppression de null, vérification doublons, vérification de la cohérence des données...)*
 - *Création de nouvelles variables (ex : âge...)*
 - *Fusion*
 - *Analyse statistique et visualisation*
- Outils : Python (Pandas, matplotlib, seaborn, numpy, scipy...)

ANALYSE DES VENTES



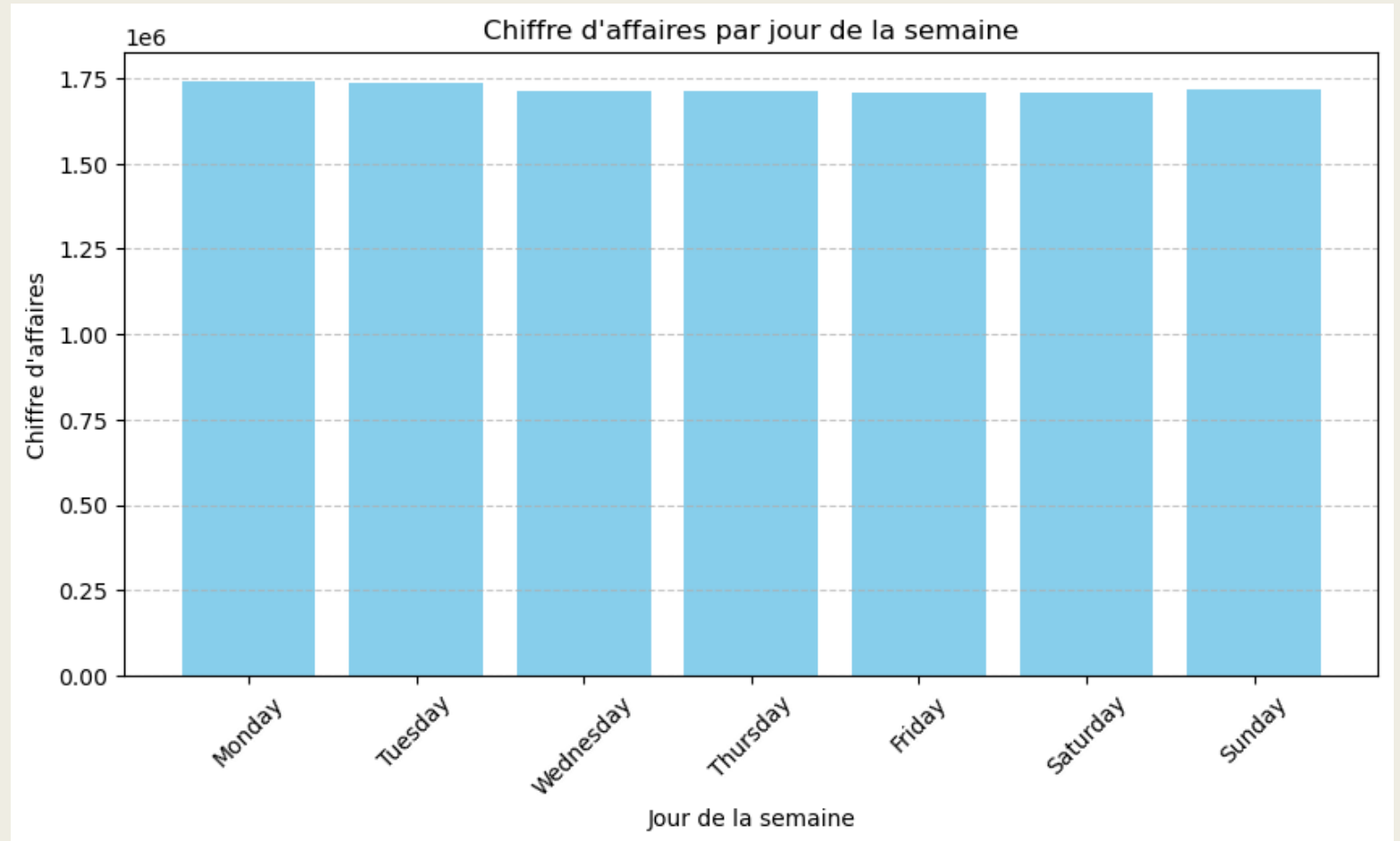
Evolution du CA quotidien et mensuel

- Moyenne mobile du CA quotidien sur 7 et 28 jours utiles pour voir les moments marquant.
- CA quotidien (MM) entre 15K€ et 19K€
- CA mensuel entre 480 et 530k€
- Période notables:
 - Oct 2021
 - Nov 2021
 - Mars 2022

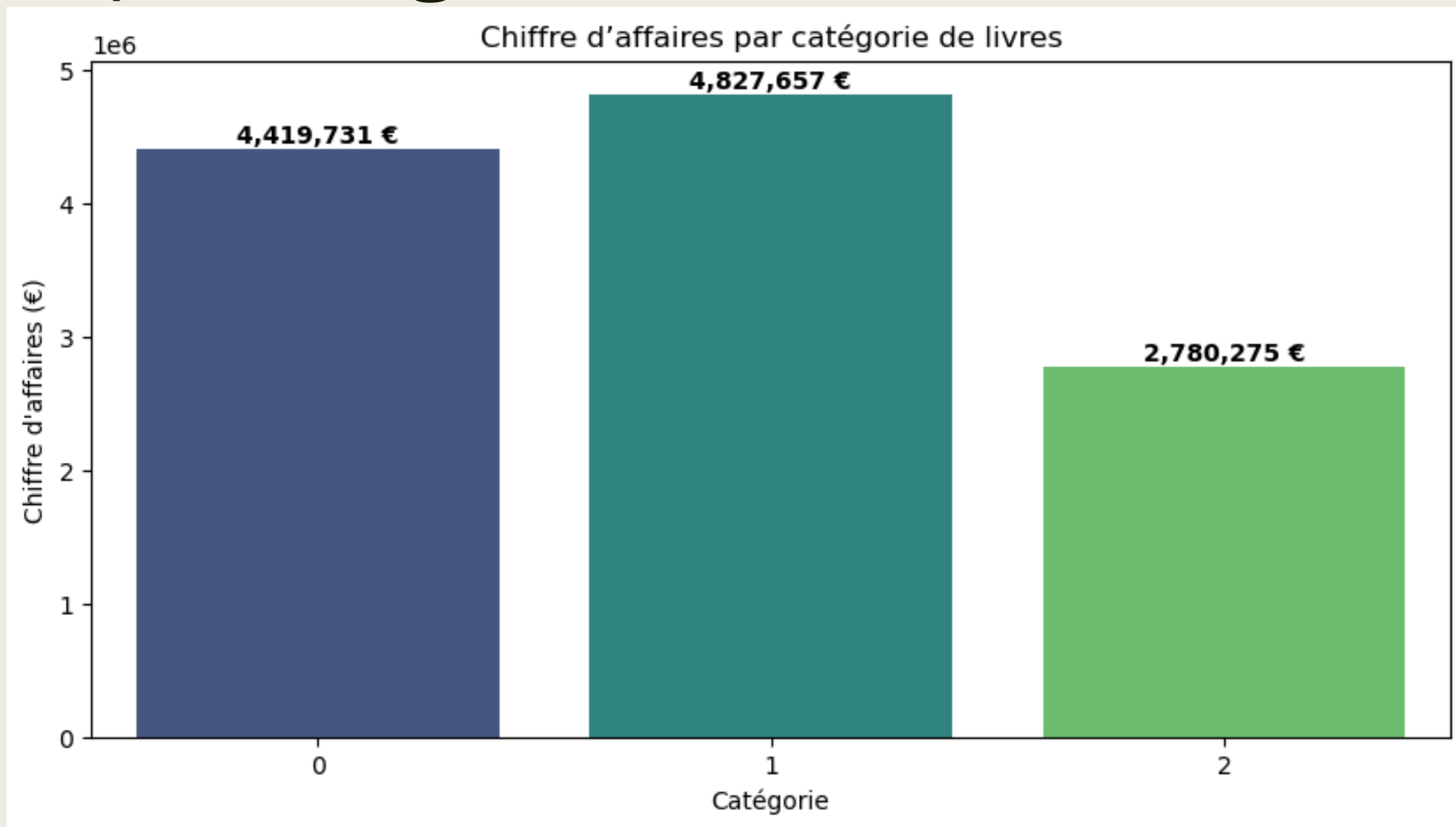


Saisonnalité hebdomadaire

- Il n'y a pas d'influence sur le jour de la semaine et les vente

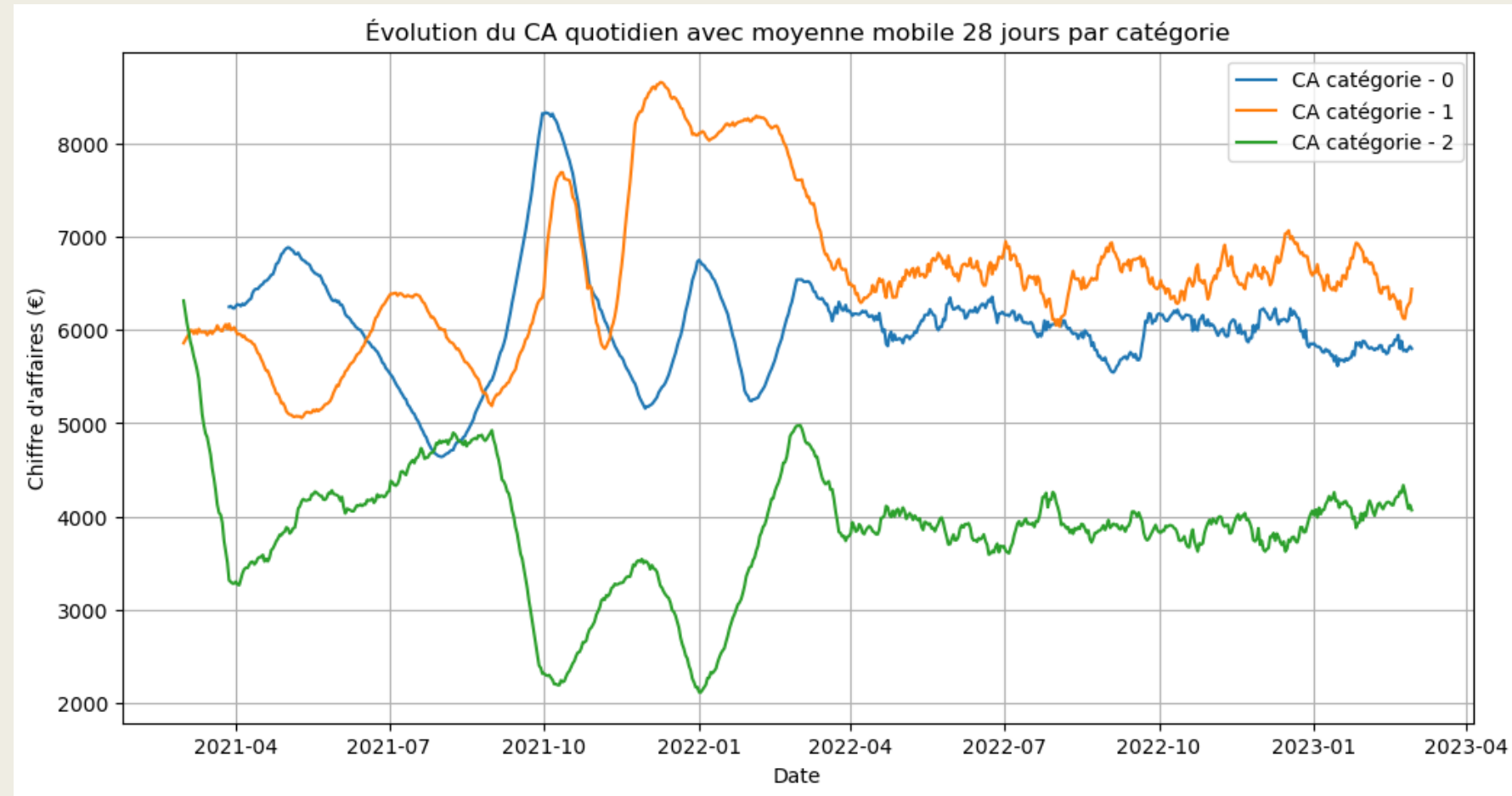


CA par catégorie de livre



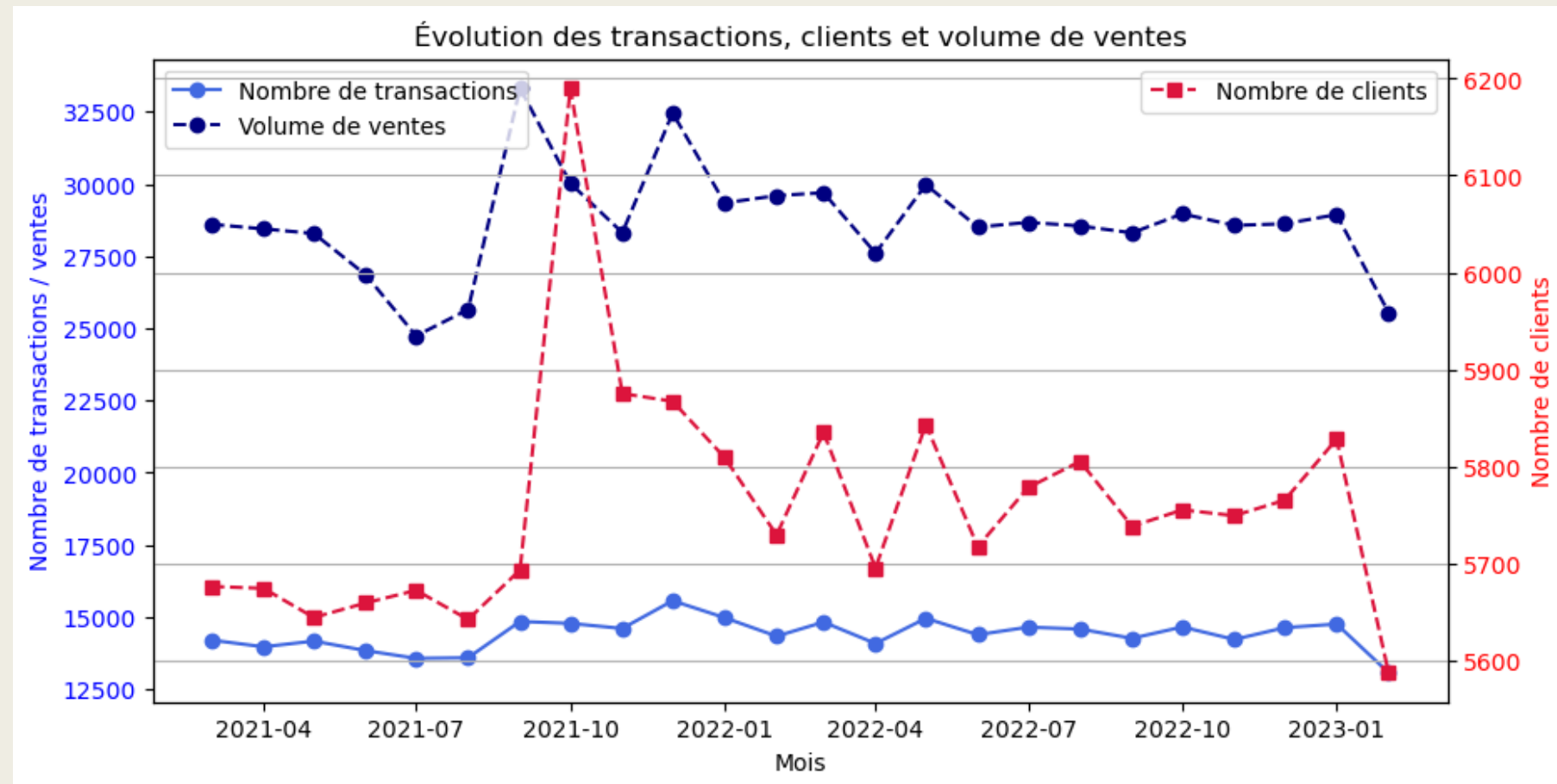
Evolution du CA quotidien par catégorie

- Grande variation du CA par catégorie la première année
- Stabilité sur la deuxième



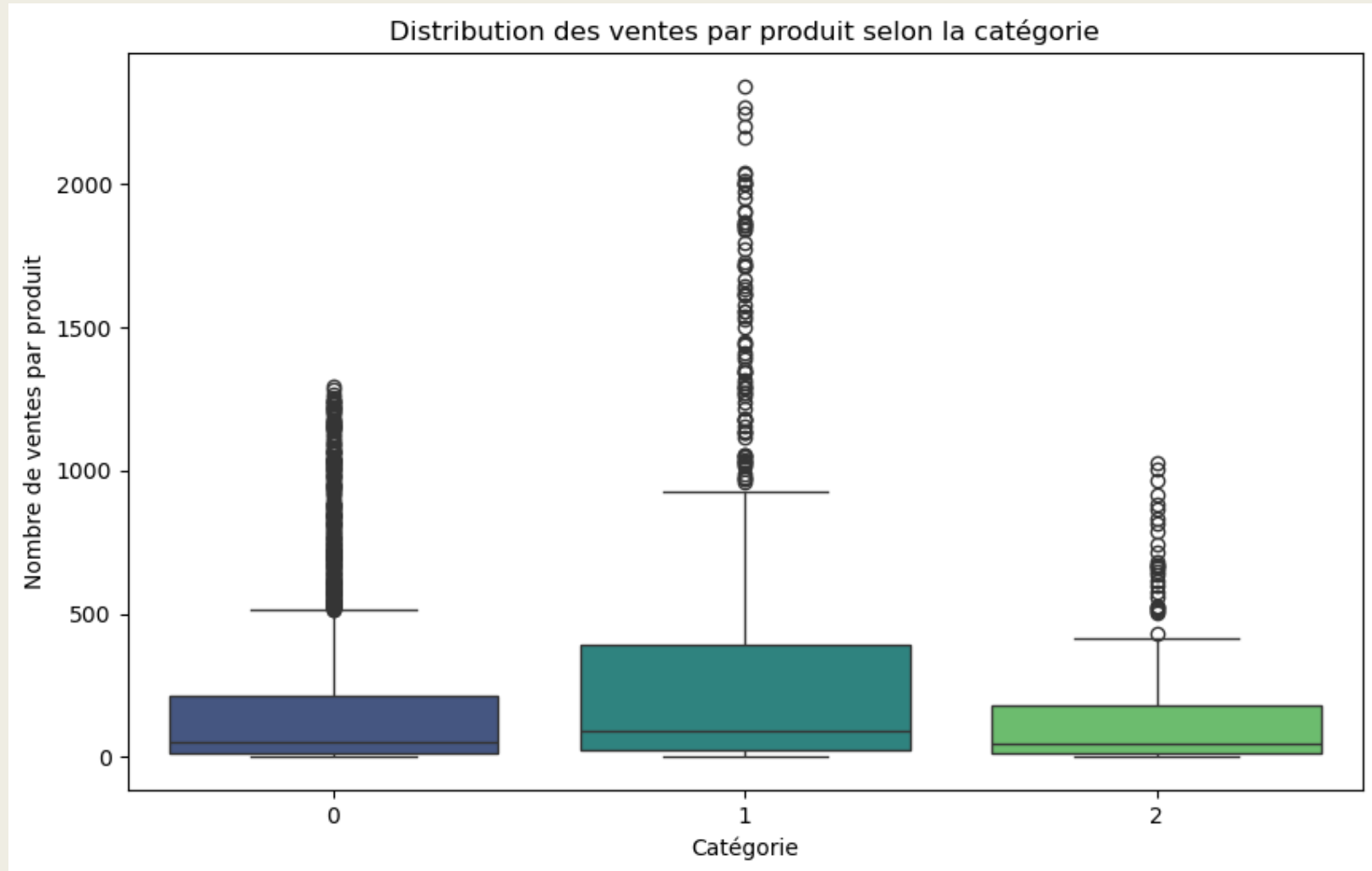
Evolution mensuelles des transactions, clients et volume de ventes

- Nombre de clients ~5800
- Pic bref en oct 2021 +10%
- Nombre de transaction entre 13K et 15K
- Volume de vente plus fluctuant entre 25K et 32K

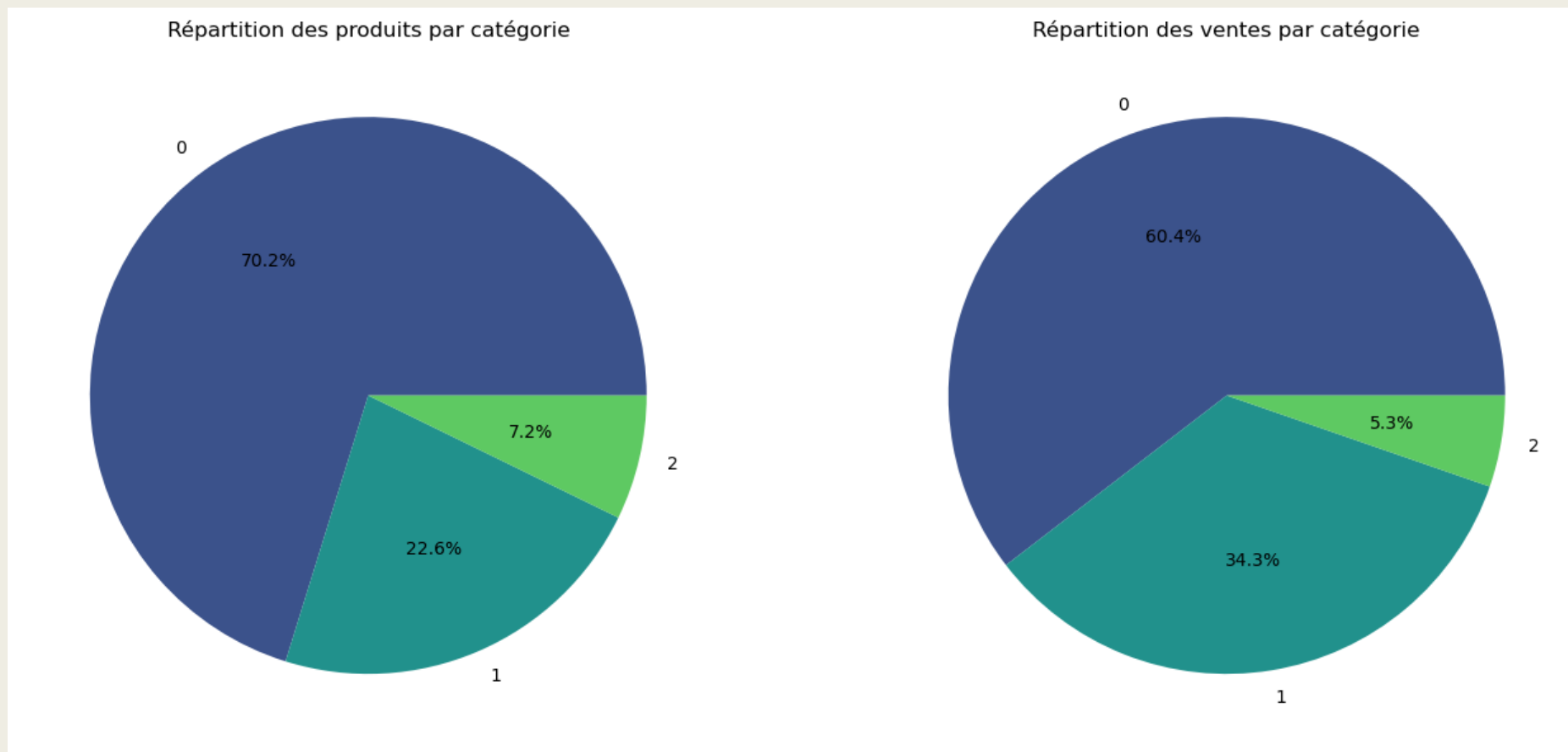


Top / flop vente

- Top vente dans la catégorie 1 avec plus de 2000 unités par référence
- 471 références avec plus de 500 ventes
- Flop nombreux surtout catégorie 0 et 2 avec moins de 3 ventes (44 références)



Répartition références et ventes par catégorie



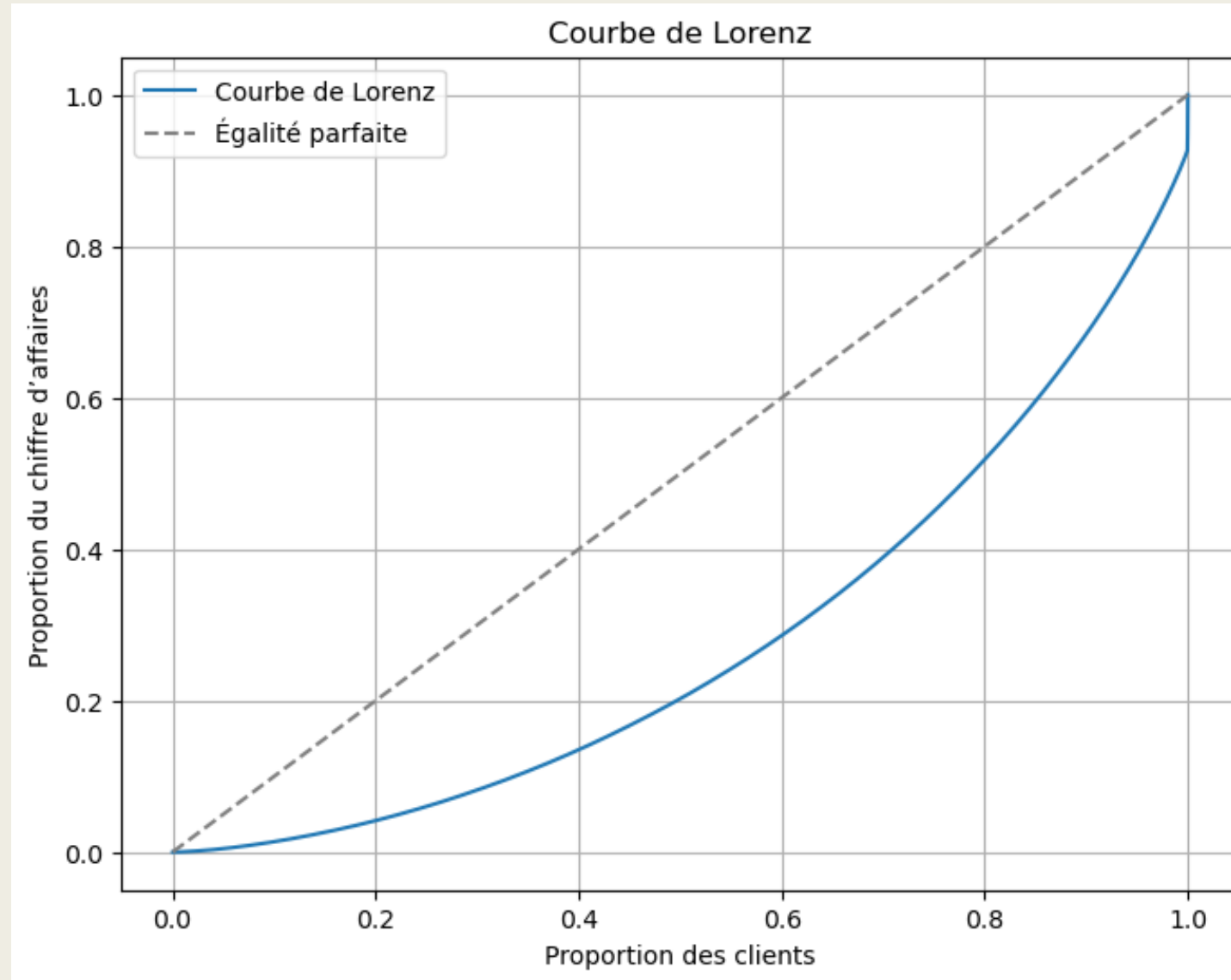
- Catégorie 1 : 34% du volume de vente pour seulement 22% des références
- Catégorie 0 : 60% du volume de vente pour 70% des références

ANALYSE
COMPORTEMENT CLIENTS



Répartition chiffre d'affaire / clients

- Courbe en dessous de l'égalité parfaite
- CA divisé :
 - 80% *clients* => 50% CA
 - 20% *clients* => 50% CA
- Focus sur 4 clients :
 - 7,35% du CA
 - + de 100K€ de dépenses



Genre des clients et catégorie de livre acheté

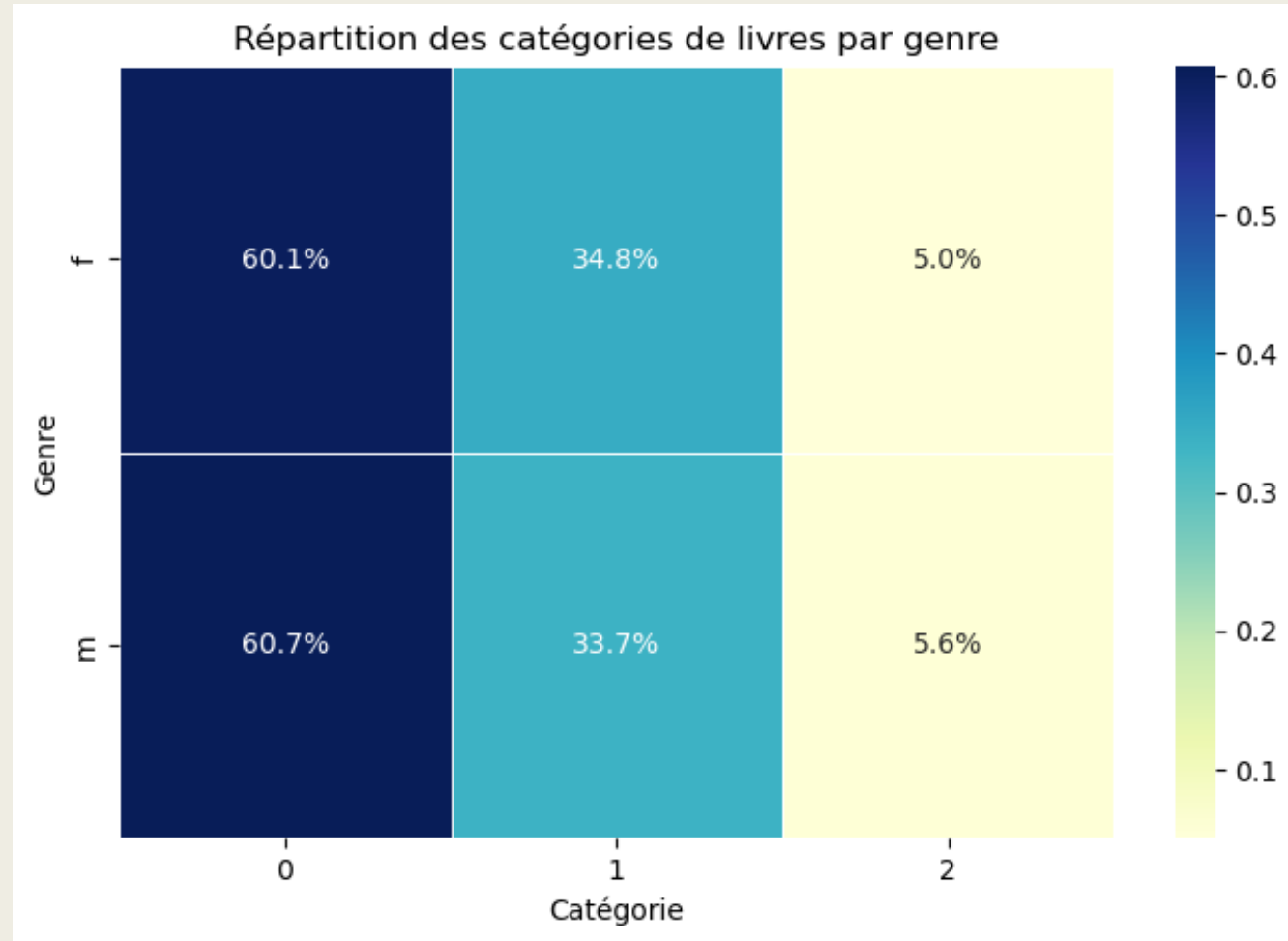
- Type de variables :
 - Sex = catégorielle (qualitatif)
 - Categ = catégorielle (qualitatif)
- Test Khi2 :
 - Score : 158

Score élevé du khi2 augmente les chances de rejeter H_0

- $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard. H_0 rejetée.

Le genre influence significativement les catégories de livres achetées



Age des clients et montant total des achats

Les clients âgés achètent ils plus ou moins ?

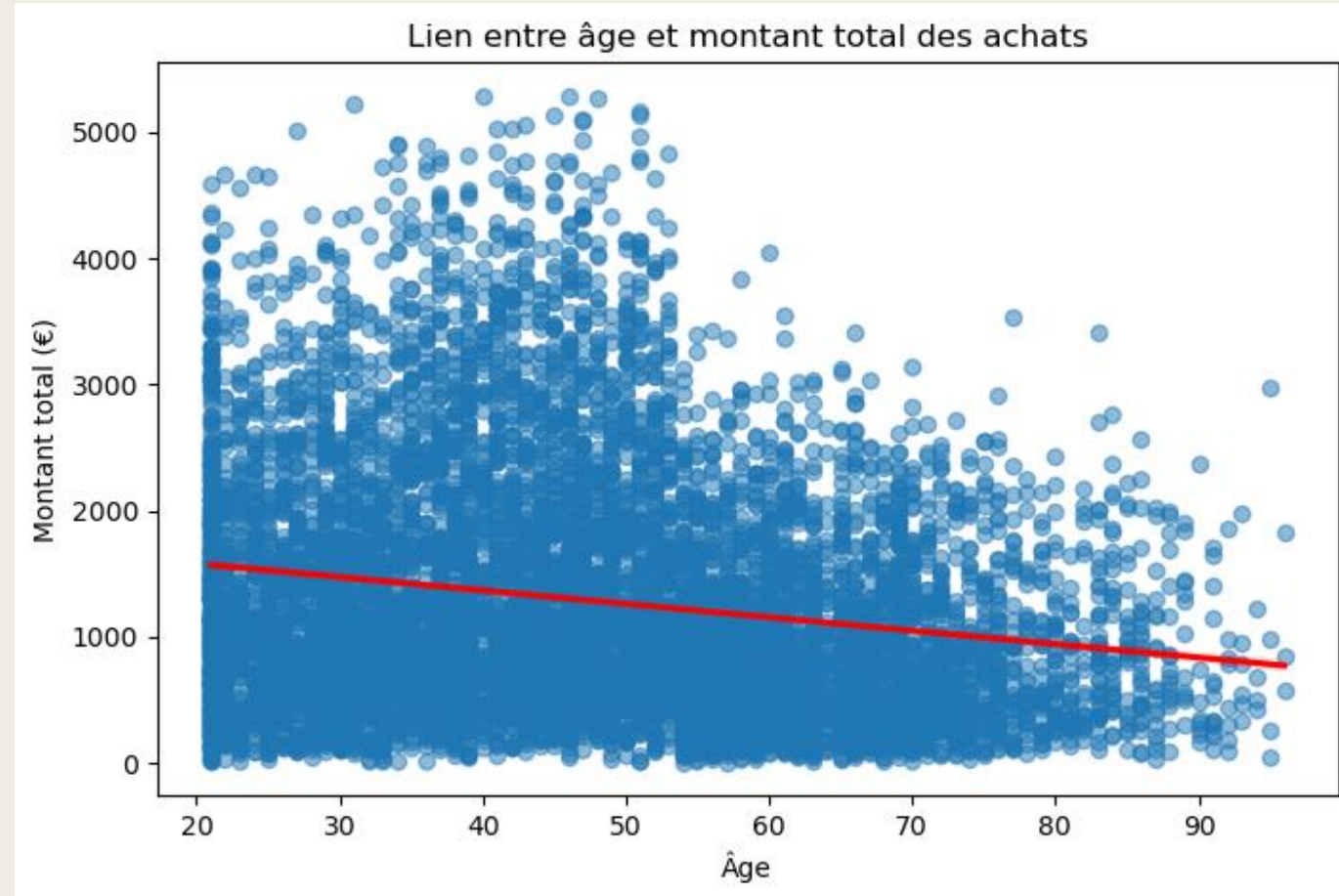
- Type de variables :
 - Age = *quantitatif*
 - Somme achats /client = *quantitatif*
- Test Spearman (car moins sensible aux outliers) :
 - $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard. H0 rejetée

- Score = -0,18

Coefficient très faible, nous sommes dans une situation d'absence de corrélation

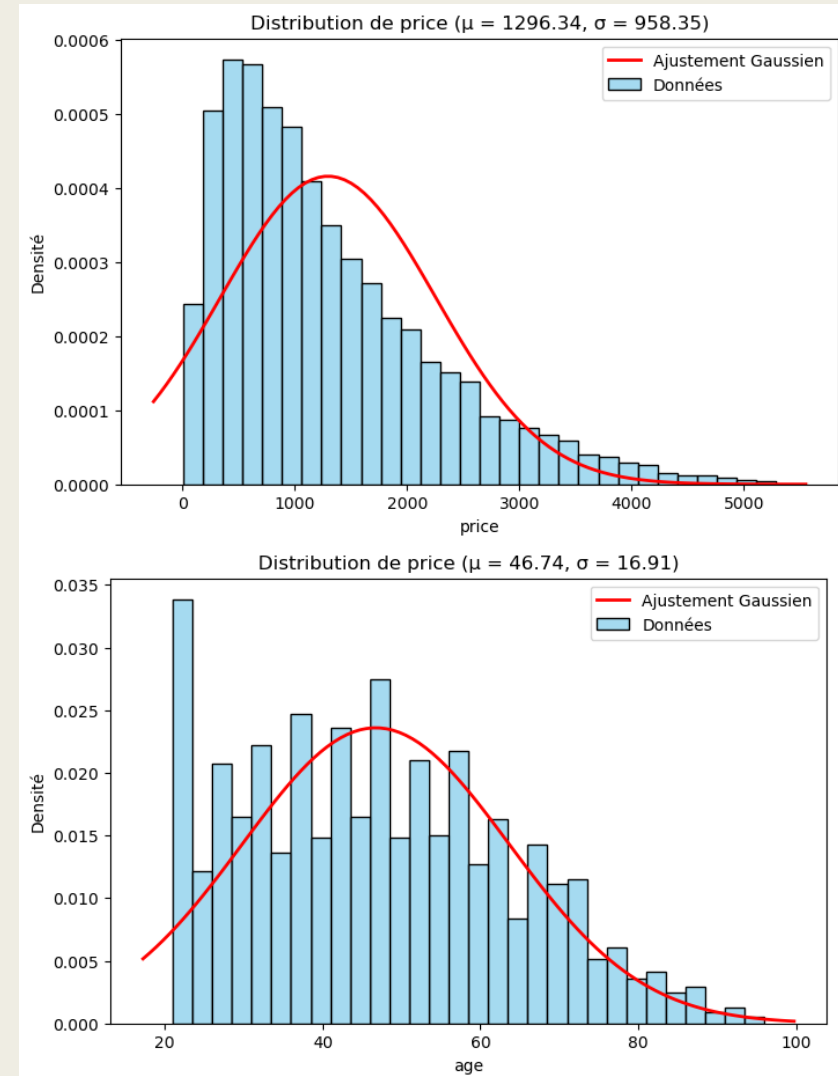
L'âge des clients n'influence pas significativement le montant des achats



Visuel sans les 4 gros clients

Age des clients et montant total des achats - Recherche

- Vérification test Spearman sans les outliers:
 - Score :-0,18 et $p = 0,0$ (même résultat)
- Exploration si test de Pearson possible :
 - Suite à Normaltest : âge et montant total achats ont une valeur p de 0
 - Les deux variables ne suivent pas une loi normale
 - Il n'est pas possible de faire un Pearson



Age des clients et fréquences d'achats

Les clients jeunes ou plus âgés achètent-ils plus souvent ?

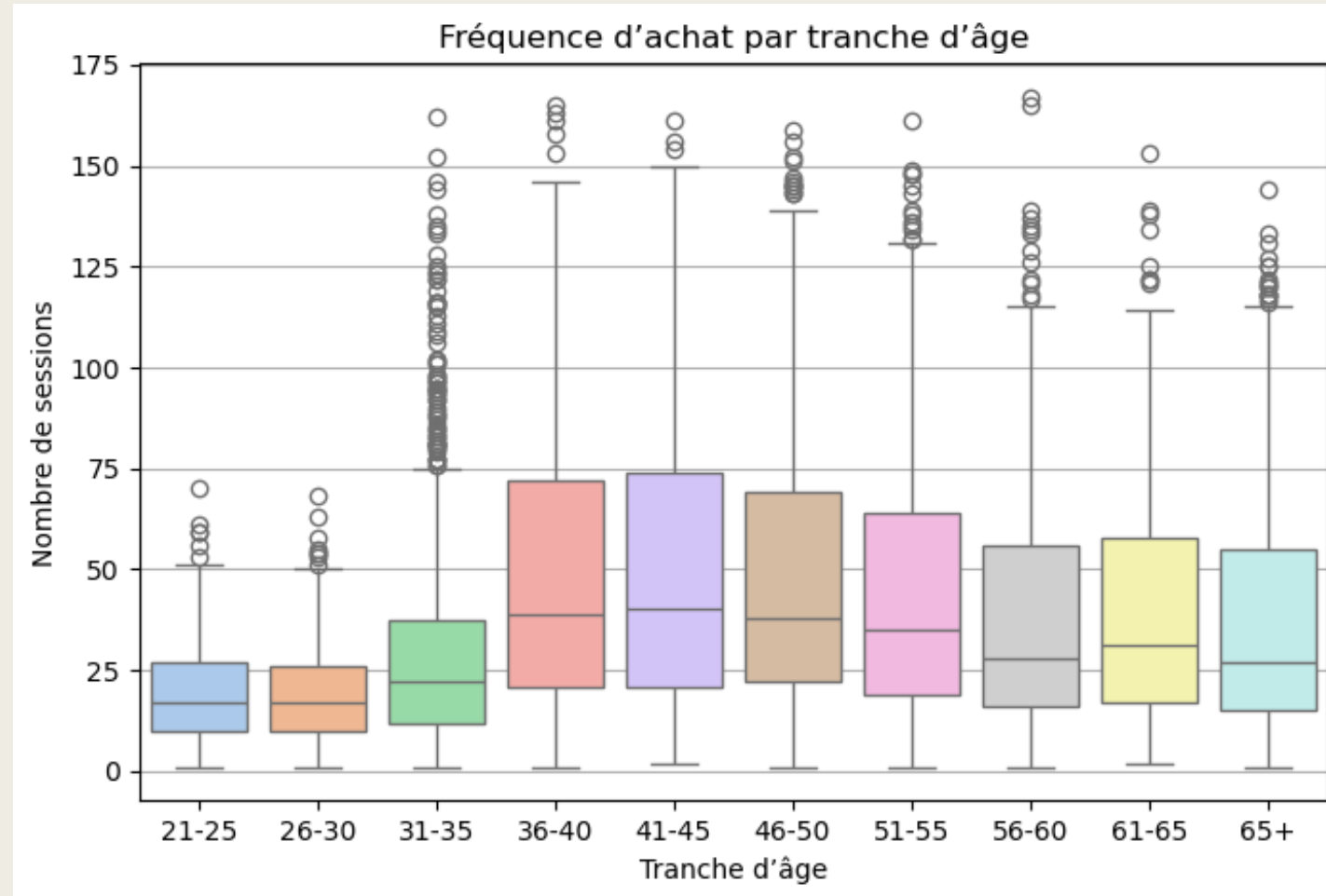
- Type de variables :
 - Age = *quantitatif*
 - Fréquence d'achat = *quantitatif*
- Test de Spearman :
 - $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard. H0 rejetée

- Score = 0,21

Coefficient faible, il y a une faible corrélation positive.

L'âge des client influence faiblement la fréquence des achats.



Visuel sans les 4 gros clients

Age des clients et fréquences d'achats

test par tranches d'âge

- Type de variables :
 - *Tranche d'âge = qualitative*
 - *Fréquence d'achat = quantitatif*

- Test de Kruskal-Wallis
 - $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard. H_0 rejetée

- *Score : 1204*

Score élevé il y a une grande différence entre les groupes

Le groupe d'âge influence significativement et fortement la fréquences d'achats

- Test ANOVA
 - *Il y a plus de deux groupes*
 - *Chaque groupe est normal (test Shapiro)*
 - *Les variances ne sont pas égales (test de Levene avec $p = 0,0$)*

Les conditions en sont pas remplis pour faire un ANOVA

Age des clients et taille du panier moyen

Le panier moyen varie-t-il avec l'âge ?

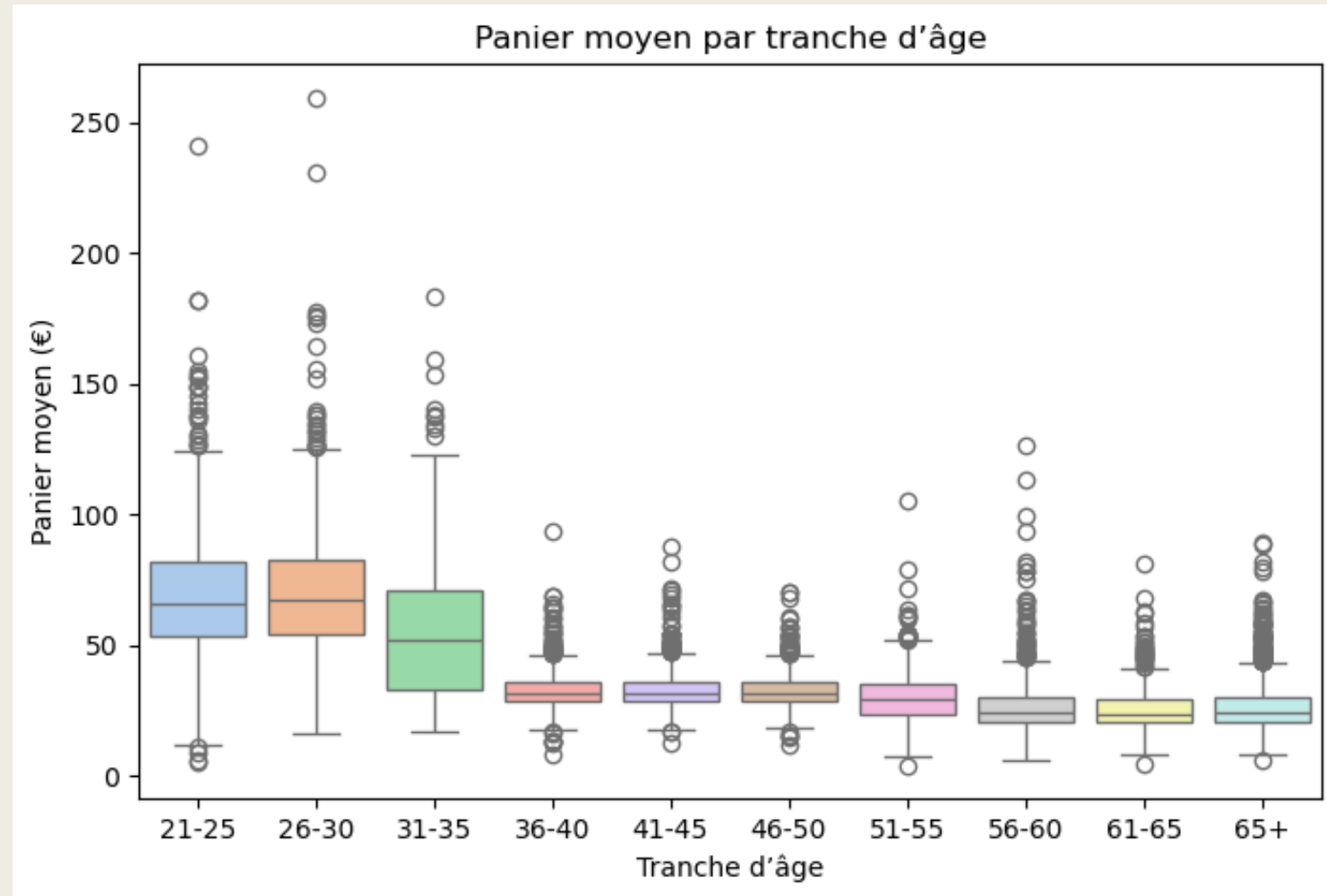
- Type de variables :
 - Age = *quantitatif*
 - Panier moyen = *quantitatif*
- Test de Spearman :
 - $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard.
H0 rejetée

- Score = -0,70

Coefficient moyen, il y a une corrélation moyenne négative.

L'âge des client influence moyennement la taille du panier moyen.



Tranche d'âge des clients et taille du panier moyen

- Type de variables :
 - *Tranche d'âge = qualitative*
 - *Panier moyen = quantitatif*
- Test de Kruskal-Wallis
 - $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard. H0 rejetée

- *Score : 4650*

Score élevé il y a une grande différence entre les groupes

Le groupe d'âge influence significativement et fortement la taille du panier moyen.

Age des clients et catégorie des livres achetés

■ Type de variables :

- *Tranche d'âge* = catégorielle (qualitatif)
- *Categ* = catégorielle (qualitatif)

■ Test Khi2 :

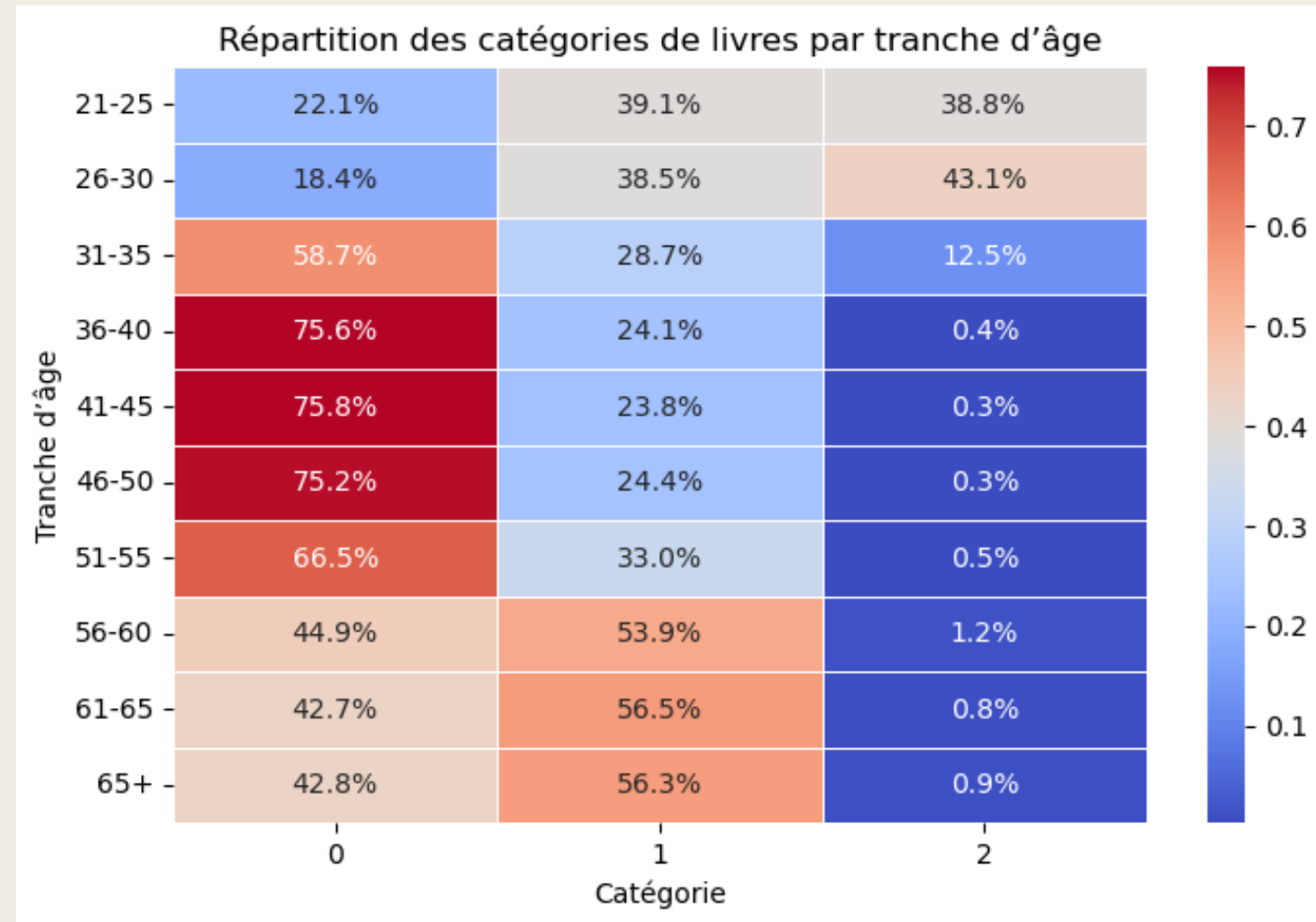
- *Score : 258551*

Score très élevé du khi2 augmente les chances de rejeter H0

- $P = 0,0 < 0,05$

Très peu de chance que les différences soient dues au hasard. H0 rejetée.

La tranche d'âge influence significativement les catégories de livres achetées



Synthèse

- Le CA a été très fluctuant puis s'est stabilisé à partir d'avril 2022.
- Questionnement sur les performances inégales des catégories de livre avant 04/2022 qui se sont stabilisées par la suite.
- Les jours de la semaines n'influencent pas les ventes. (voir différence avec librairie physique)
- 471 références sur 3265 se vendent à plus de 500 exemplaires
- 4 gros clients avec plus de 100K€ de dépenses
- L'âge n'a pas d'influence sur le montant des achats mais en a sur son comportement d'achat. Cette influence n'est pas linéaire mais par tranche d'âge.
- Les 21-35 ans achètent 2x moins souvent mais dépensent 2x plus.
- Les tranches d'âges influencent les catégories de livres achetées.

21-35 ans	36-55 ans	56-65 ans
Catégorie 2 (sont les seuls acheteurs)	Catégorie 0	Catégorie 1

Recommandations

- Optimisation de l'offre produit :
 - *Prioriser les références à fort potentiel*
 - *Créer des offres groupées pour les produits peu performants*
- Personnalisation marketing par âge :
 - *Ciblage des 21-35 ans avec des campagnes à panier élevé*
 - *Catégories mises en avant selon l'âge*
- Suivi et pilotage :
 - *Reporting régulier dès le lancement produit*
 - *Monitoring mensuel des catégories*
- Développement du B2B :
 - *Programme de fidélité ou service VIP dédié aux 4 clients stratégiques*