

# Analyse des ventes du site internet

Jérémy GUINAULT  
Lapage

# Sommaire

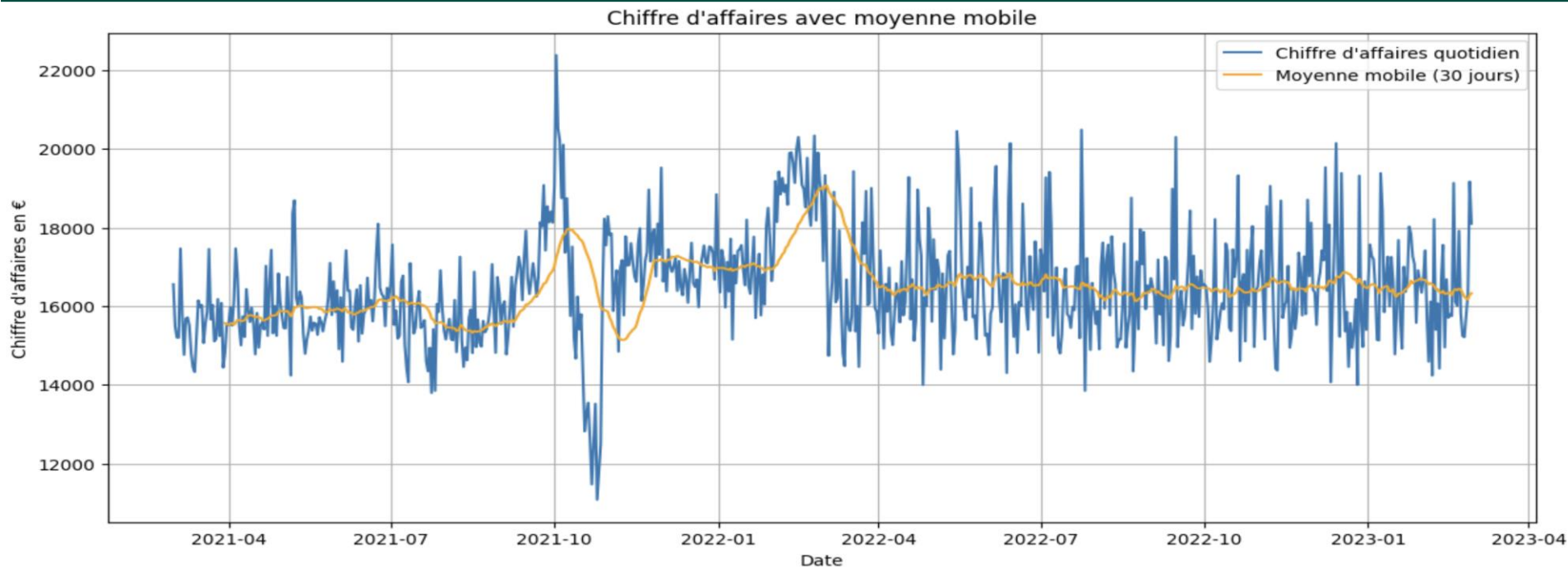
- *Analyse des données fournies,*
- *Chiffre d'affaires du site internet,*
- *Analyse des catégories,*
- *Nos meilleurs produits / nos moins bons produits,*
- *Comparaison entre les ventes, transactions et clients,*
- *Nos clients BtoC / BtoB,*
- *Les corrélations.*

# Analyses exploratoires des données

- *Trois datasets :*
  - *Fichier customers,*
  - *Fichier products,*
  - *Fichier transactions.*
- *Traitement réalisés :*
  - Analyse des datasets (dimension dataset, nature des données),*
  - Verification si présence de valeur nulle ou en doublon.*

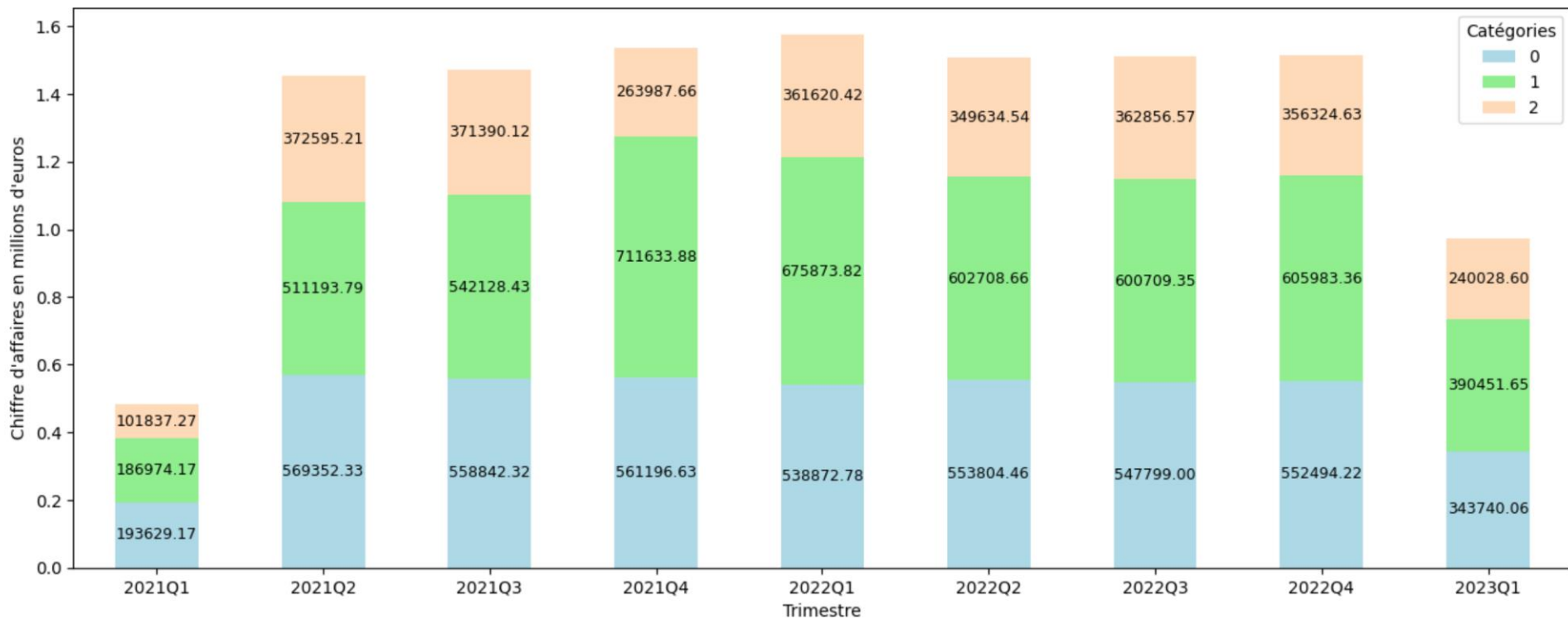
*Jointure de df\_products et df\_transactions via 'id\_prod', puis jointure de ce nouveau df avec df\_customers via 'client\_id' pour obtenir **df\_total regroupant les trois dataframes.***

# Chiffre d'affaires du site internet

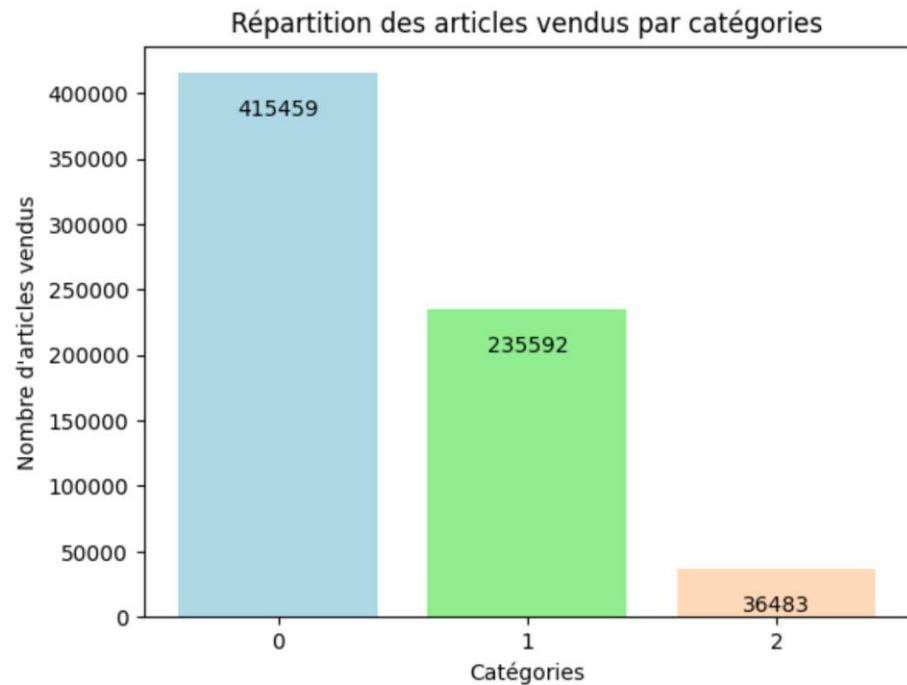
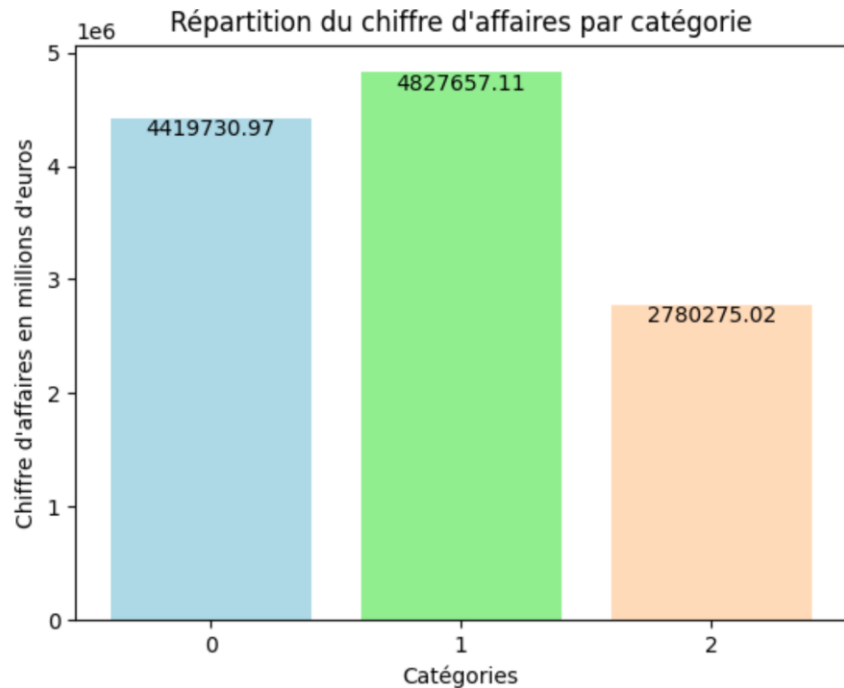


Depuis son lancement, **12 027 663,10€** de chiffre d'affaires ont été généré via le site internet.

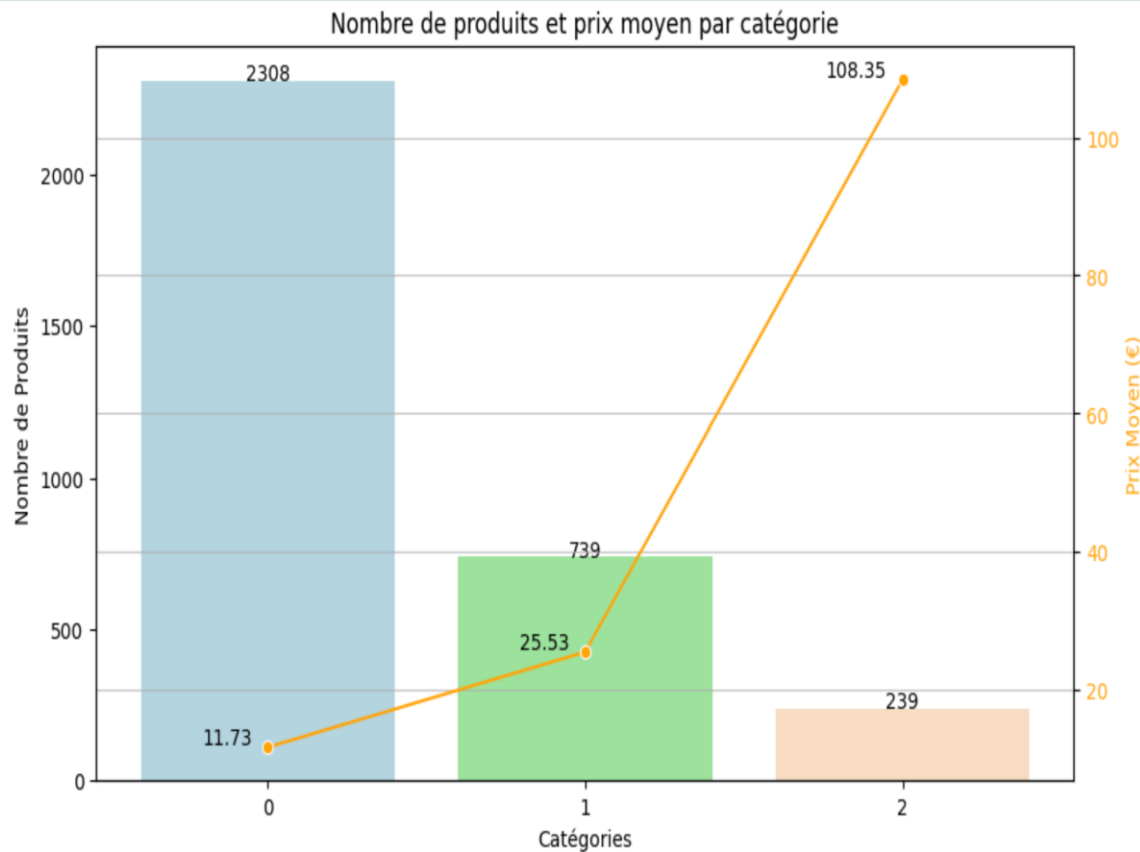
# Répartition du chiffre d'affaires par catégories et trimestres



# Analyse des catégories



# Analyse des catégories



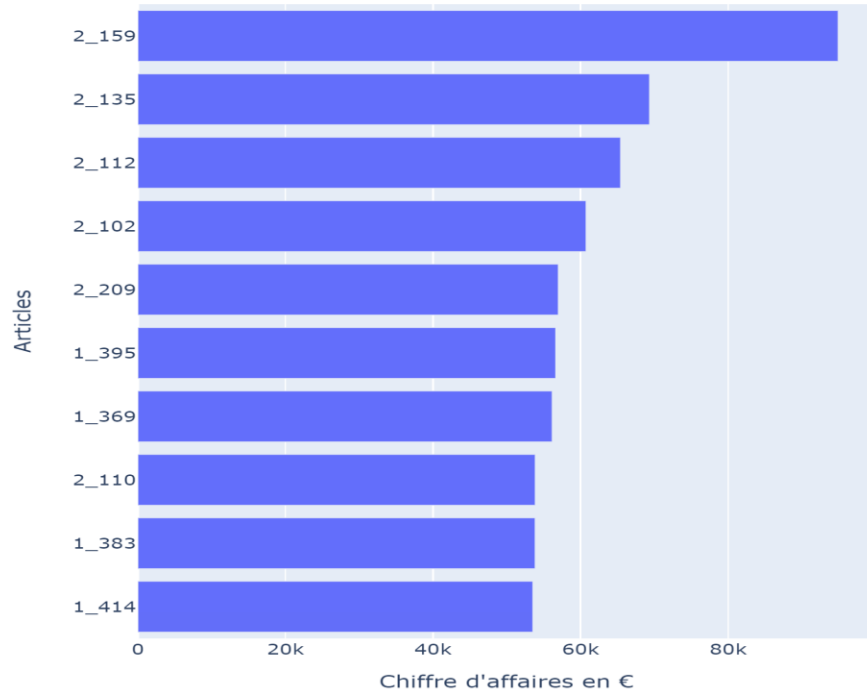
**Catégorie 0** : une offre large et abordable.

**Catégorie 1** : une gamme intermédiaire en volume et en prix.

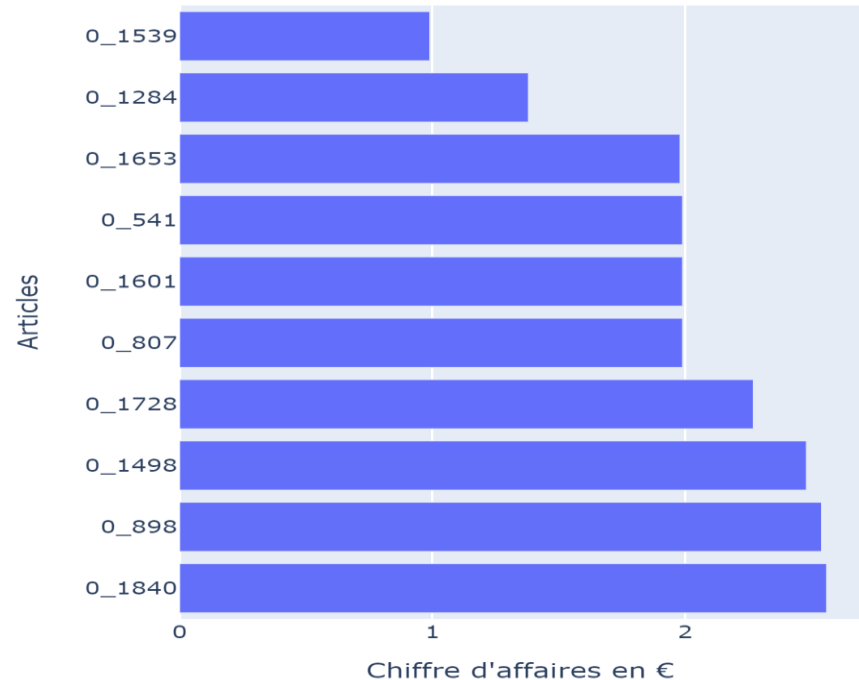
**Catégorie 2** : limitée, mais se distingue par un prix moyen élevé, indiquant une offre premium.

# Produits générant le plus/moins de chiffre d'affaires

Les 10 articles generant le plus de chiffre d'affaires



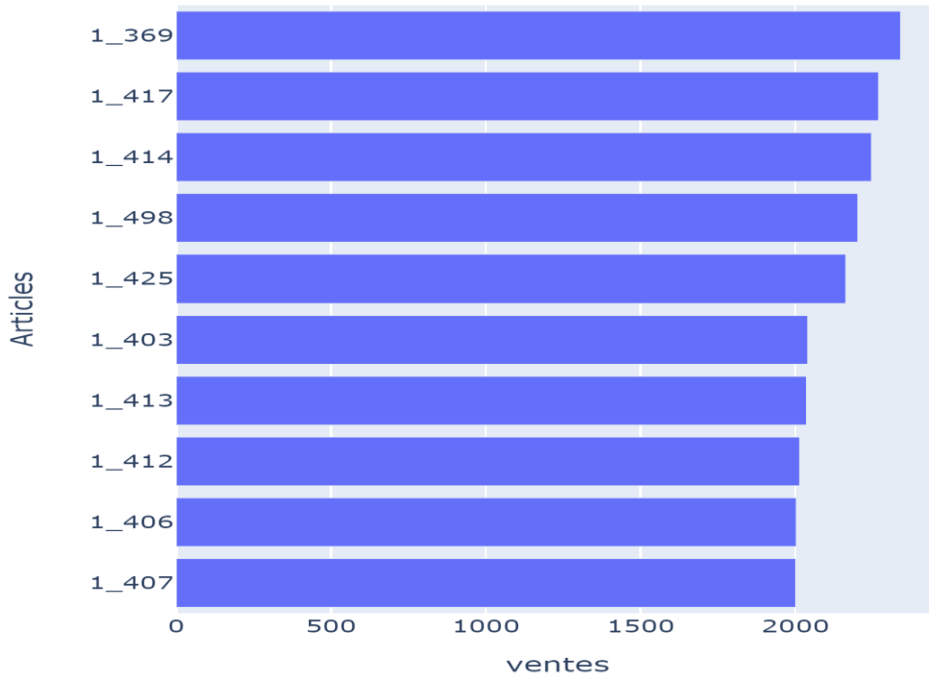
Les 10 articles generant le moins de chiffre d'affaires



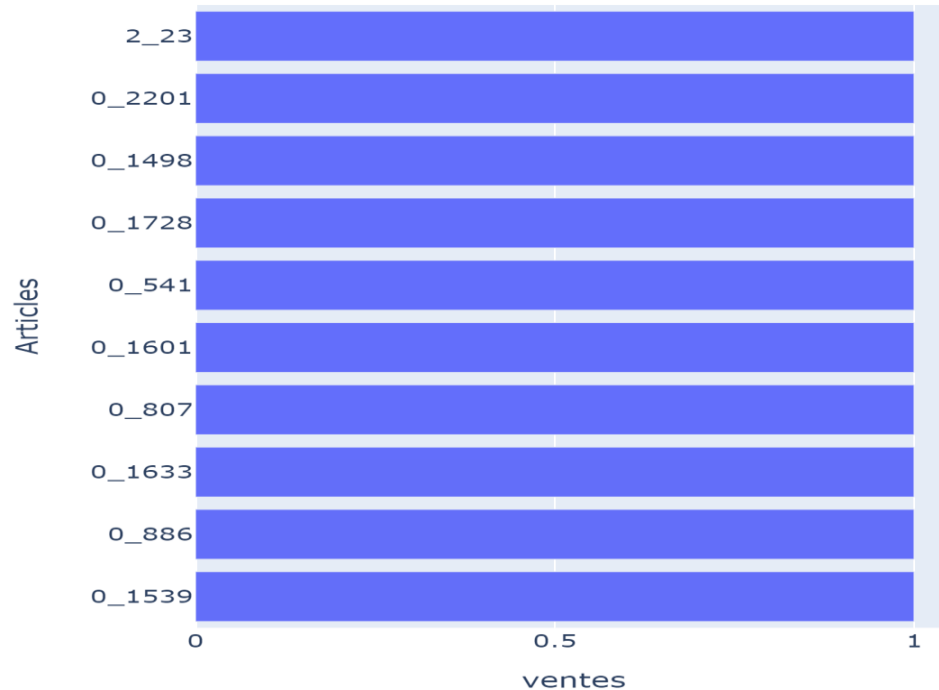


# Produits se vendant le plus/moins

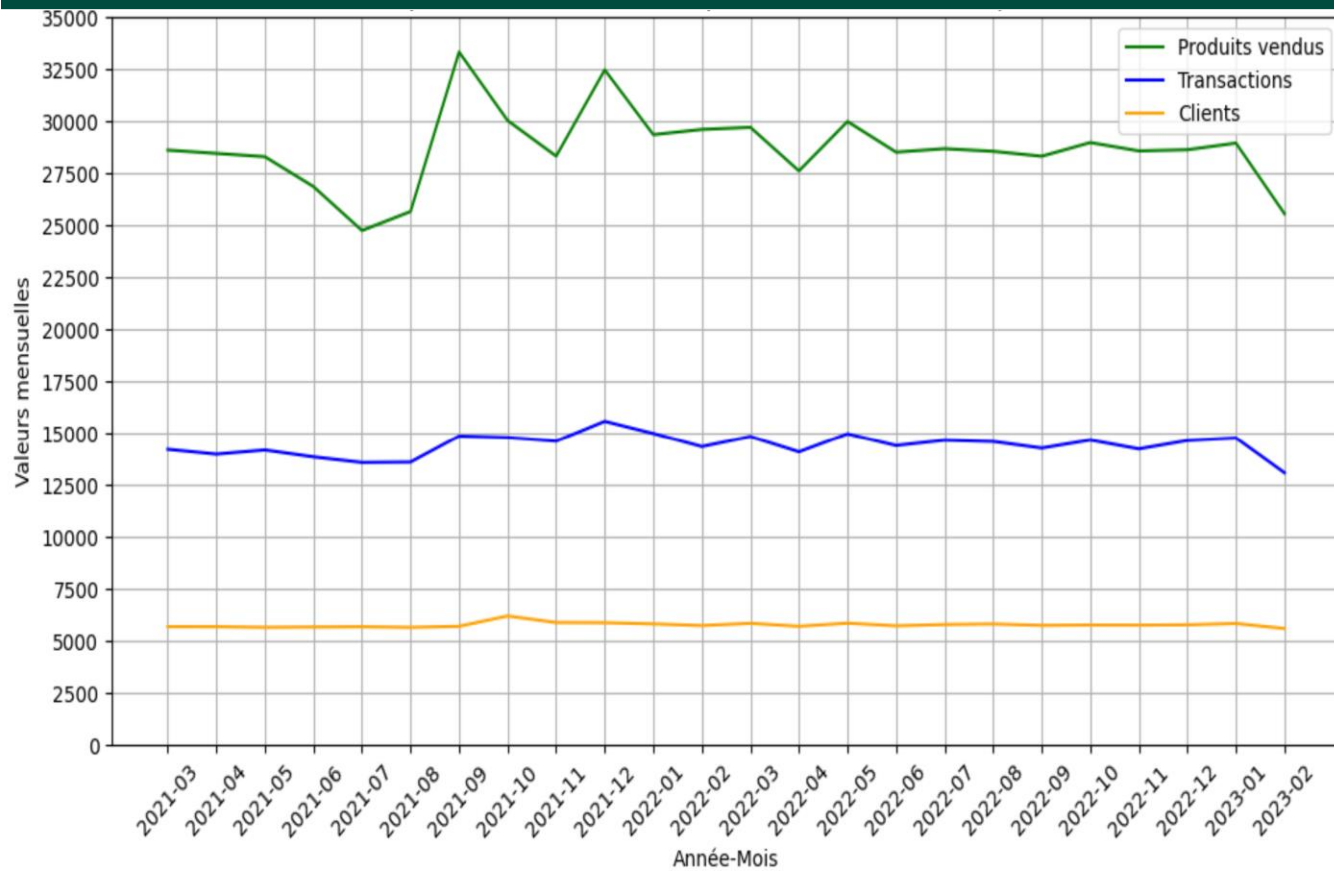
Les 10 articles les plus vendus



Les 10 articles les moins vendus



# Comparaison des ventes, transactions et clients mensuels



**Moyenne ventes : 28647**

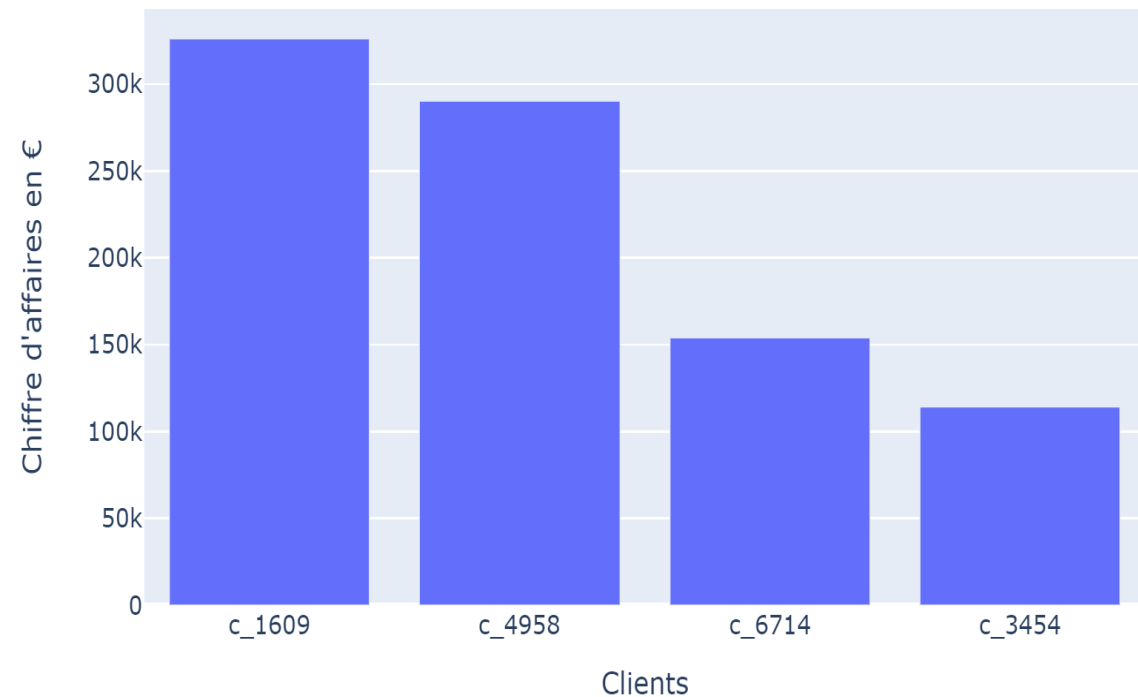
**Moyenne transactions: 14398**

**Moyenne clients: 5759**

Un client achète 5 produits par mois en moyenne.

# Clients BtoB

## Chiffre d'affaires des clients BtoB

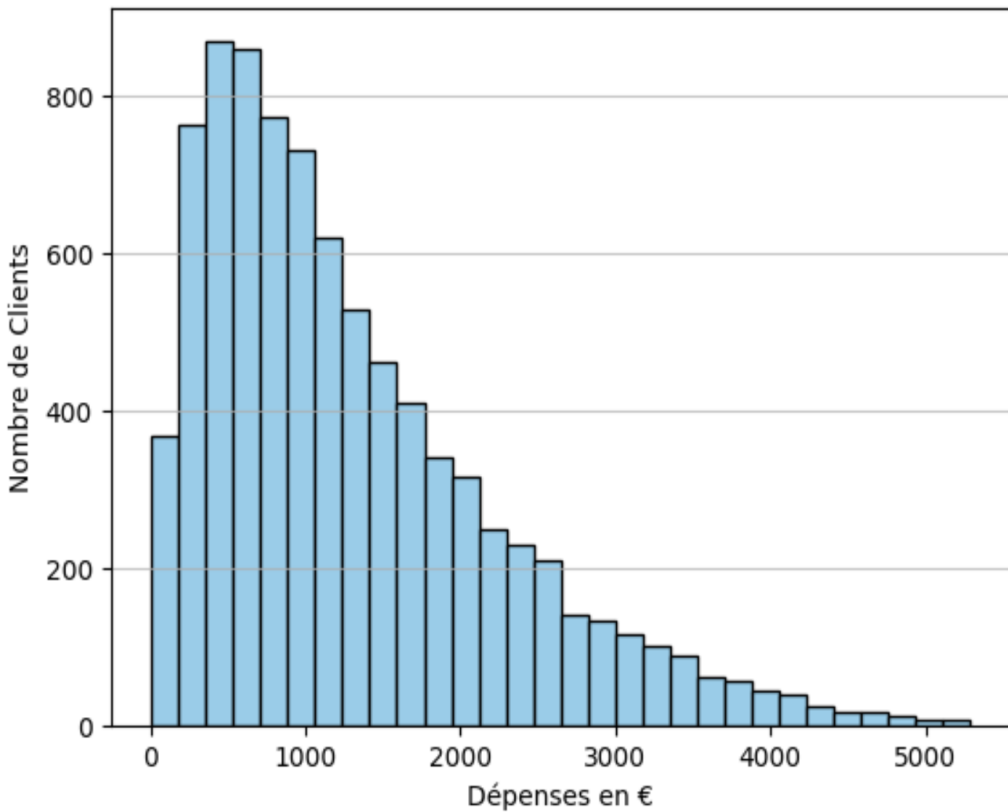


**4 clients BtoB dans notre portefeuille.**

**Ils ont généré 7,35% du chiffre d'affaires du site internet.**

# Clients Btoc

Distribution des dépenses clients



8596 clients BtoC dans notre portefeuille.

En moyenne, un client a dépensé 1296,34€ sur notre site.

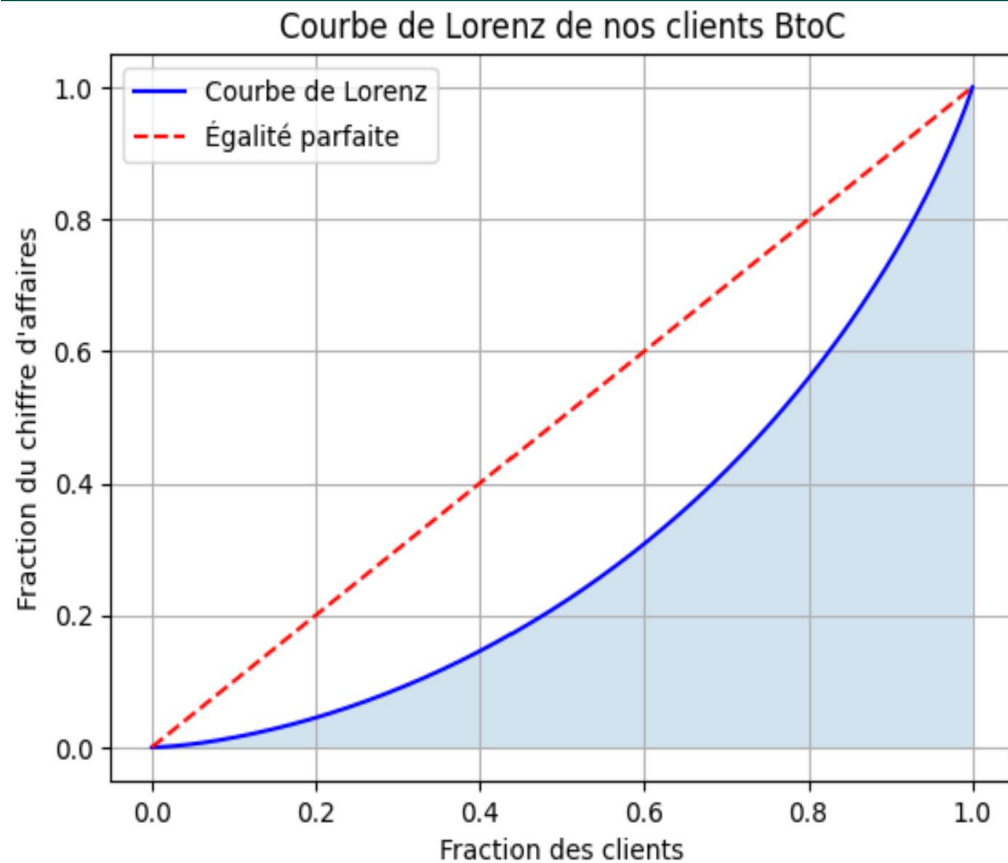
50% de nos clients ont dépensé plus de 1045,85€.

Notre « meilleur » client a dépensé 5285,82€.

Notre « moins bon » client a dépensé 6,31€.

Le niveau de dépense est très varié.

# Clients Btoc : courbe de Lorenz



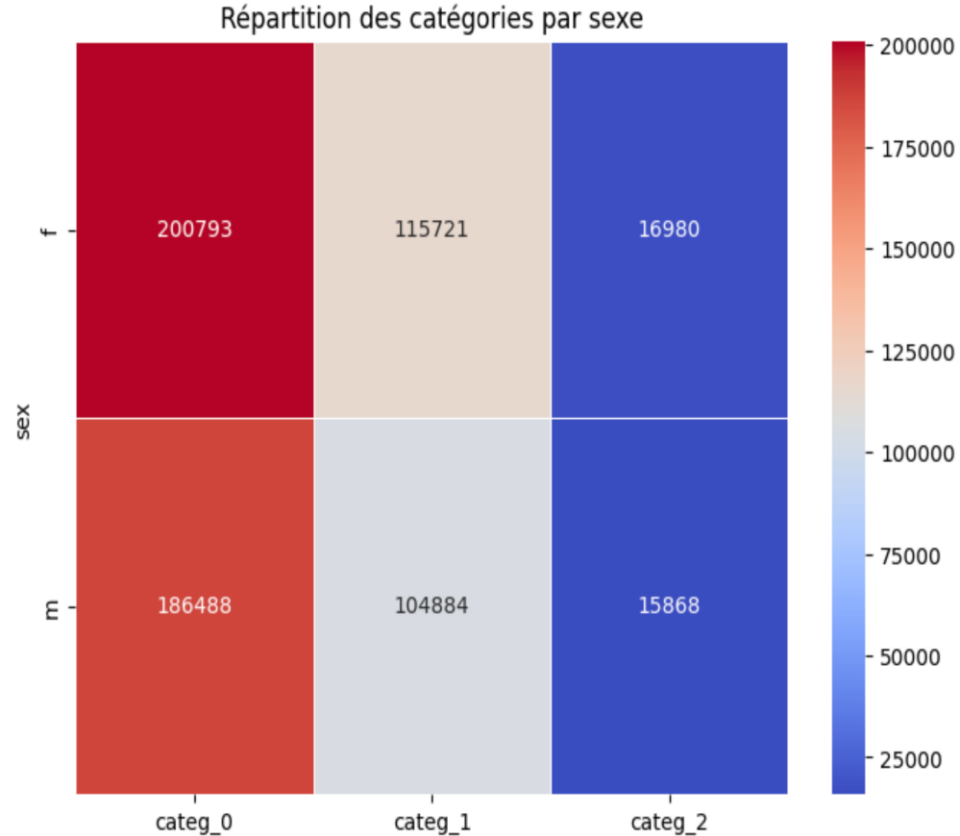
## Concentration du chiffre d'affaires :

Une petite proportion de clients génère la majorité du chiffre d'affaires, indiquant une forte concentration de celui-ci.

Indice de Gini  $\approx 0,398$ . Traduction d'une inégalité modérée, mais d'une dépendance vis-à-vis de certains clients.

➡ Nécessité de fidéliser ces clients, et développer davantage les autres.

# Corrélations : genre des clients / catégories livres achetés



## Test Chi2 entre genre et catégories :

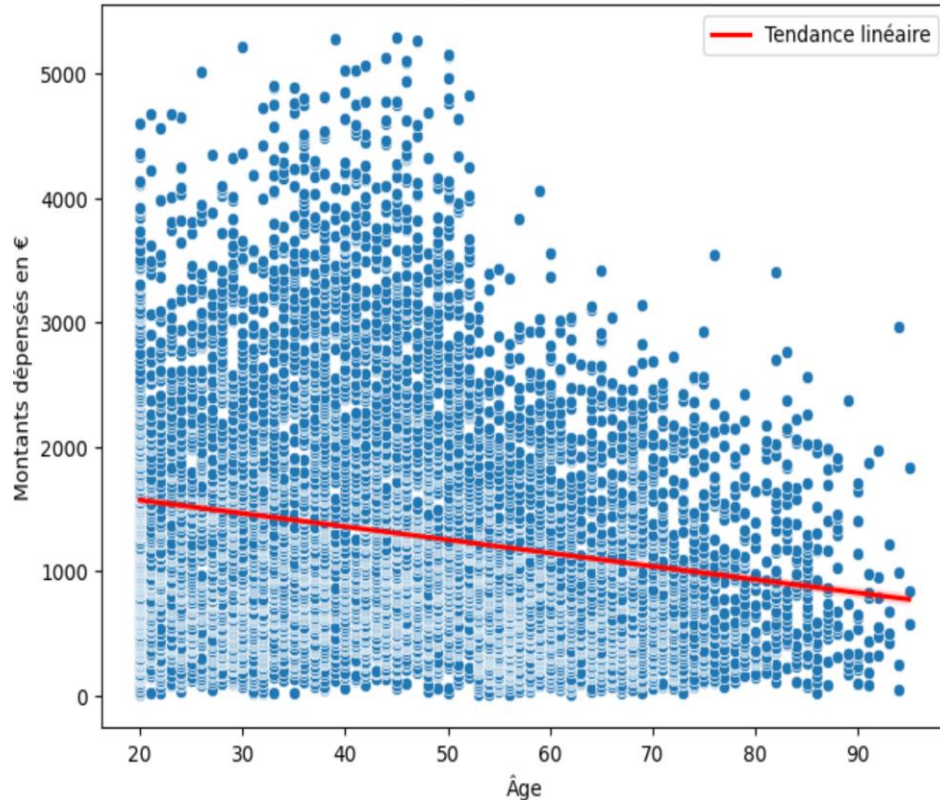
Chi2: 6396.186414612592

p-value: 0.35882927857471575

**Conclusion :** Aucune association statistiquement significative n'a été trouvée entre le sexe des clients et les catégories de livres (p-value  $\geq 0.05$ ). Par conséquent, nous ne pouvons pas rejeter l'hypothèse nulle, ce qui suggère que les deux variables sont indépendantes dans cet échantillon.

# Corrélations : age des clients / montants dépensés

Relation entre l'âge et les montants dépensés



Test de Shapiro-Wilk pour l'âge: p-valeur =  $4.628808323714737e-39$

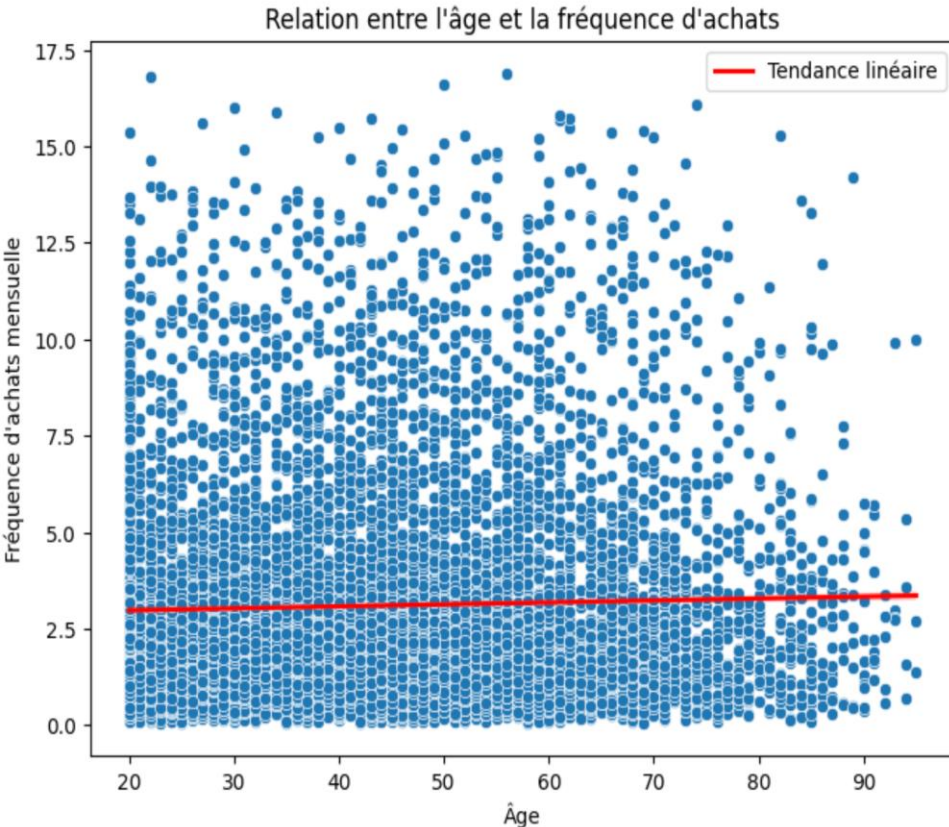
Test de Shapiro-Wilk pour le prix: p-valeur = 0.0

Test de dépendance utilisé : Spearman

Résultat de la dépendance de Spearman: -0.18 (p-valeur =  $1.0214524504733596e-66$ )

**Conclusion :** Il existe une dépendance significative entre l'âge et les montants dépensés, indiquant une relation négative.

# Corrélations : age des clients / fréquence d'achats



Test de Shapiro-Wilk pour l'âge: p-valeur =  $4.628808323714737e-39$

Test de Shapiro-Wilk pour la fréquence d'achats: p-valeur = 0.0

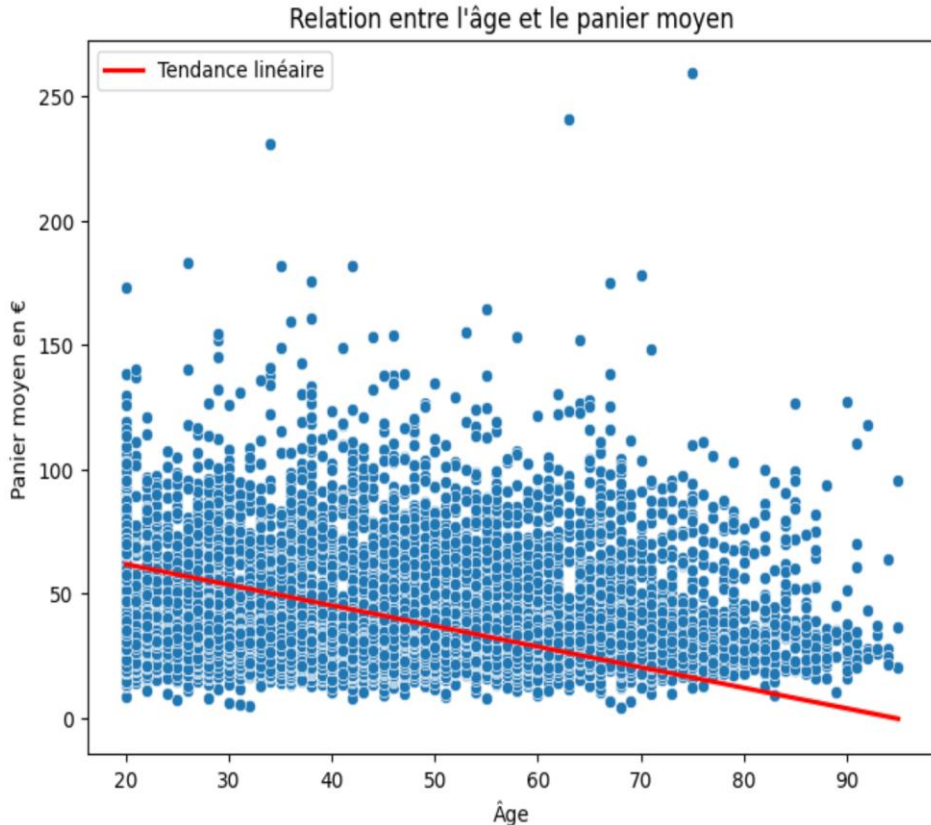
Test de corrélation utilisé : Spearman

Résultat de la corrélation de Spearman: 0.13 (p-valeur =  $1.4497470156818763e-32$ )

**Conclusion :** Il existe une dépendance significative entre l'âge et la fréquence des achats, indiquant que lorsque l'âge augmente, la fréquence des achats a tendance à augmenter.



# Corrélations : age des clients / panier moyen



Test de Shapiro-Wilk pour l'âge: p-valeur =  $4.628808323714737e-39$

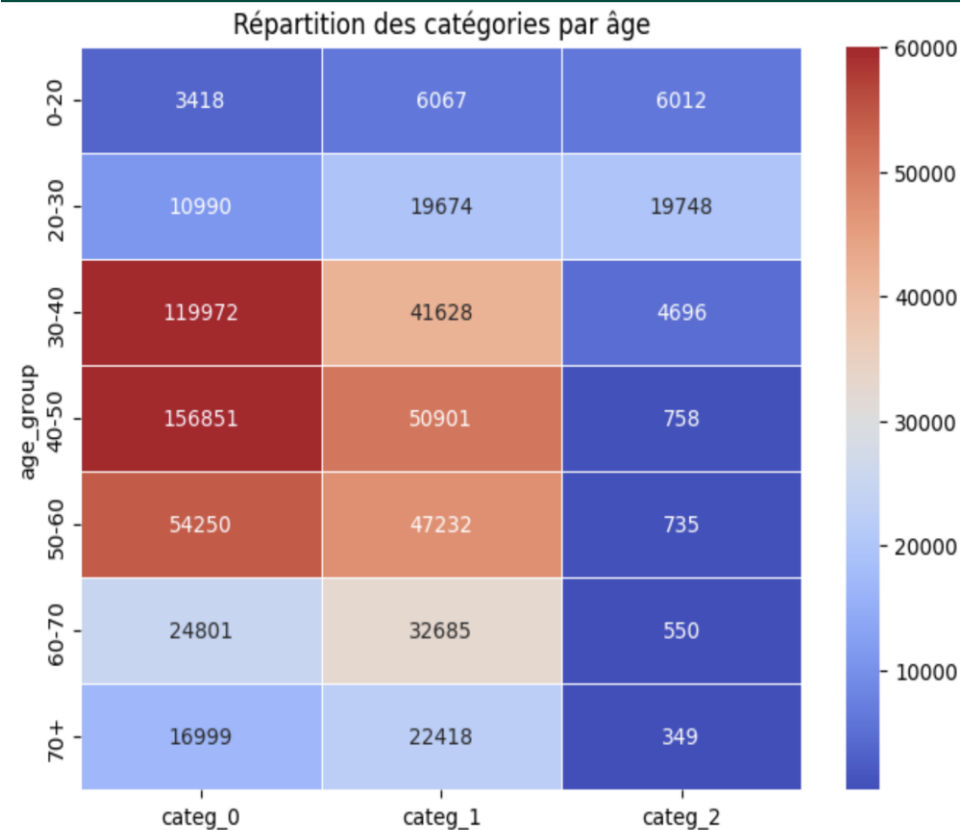
Test de Shapiro-Wilk pour le panier moyen: p-valeur = 0.0

Test de corrélation utilisé : Spearman

Résultat de la corrélation de Spearman: -0.70 (p-valeur = 0.0)

**Conclusion :** Il existe une dépendance significative entre l'âge et le panier moyen, suggérant qu'à mesure que l'âge augmente, le panier moyen tend à diminuer.

# Corrélations : age des clients / catégories de livres achetés



Test de Shapiro-Wilk pour les résidus:  
p-valeur = 5.4567066399613395e-05

Test de Levene pour l'homogénéité des variances:  
p-valeur = 0.24434284481874202

Les conditions de normalité et d'homogénéité des variances ne sont pas remplies pour un test ANOVA. Un test de Kruskal-Wallis est plus approprié.

Résultats du test de Kruskal-Wallis:  
Statistique de Kruskal-Wallis: 9.8182  
p-valeur: 0.0074

**Conclusion :** Il y a une différence significative entre les groupes (rejeter l'hypothèse nulle).