



# **Analyse des ventes en ligne**



## Lapage

- I) Présentation et nettoyage des données
- II) Analyse du chiffre d'affaires (Les requêtes d'Antoine)
- III) Analyse des comportements des clients (Les requêtes de Julie)

# I) Présentation et nettoyage des données



## Description des tables

- Customers : table des clients (id, sexe, année de naissance)
- Products : table des produits (id, prix, catégorie)
- Transactions : table des transaction (id\_produit, Date, id session, id client)

## Nettoyage des tables

- Vérification des doublons
- Retrait des lignes de test
- Vérification du format (date)
- Retrait de deux ans à toutes les dates (allant de 2021 à 2023)

## Jointure des tables

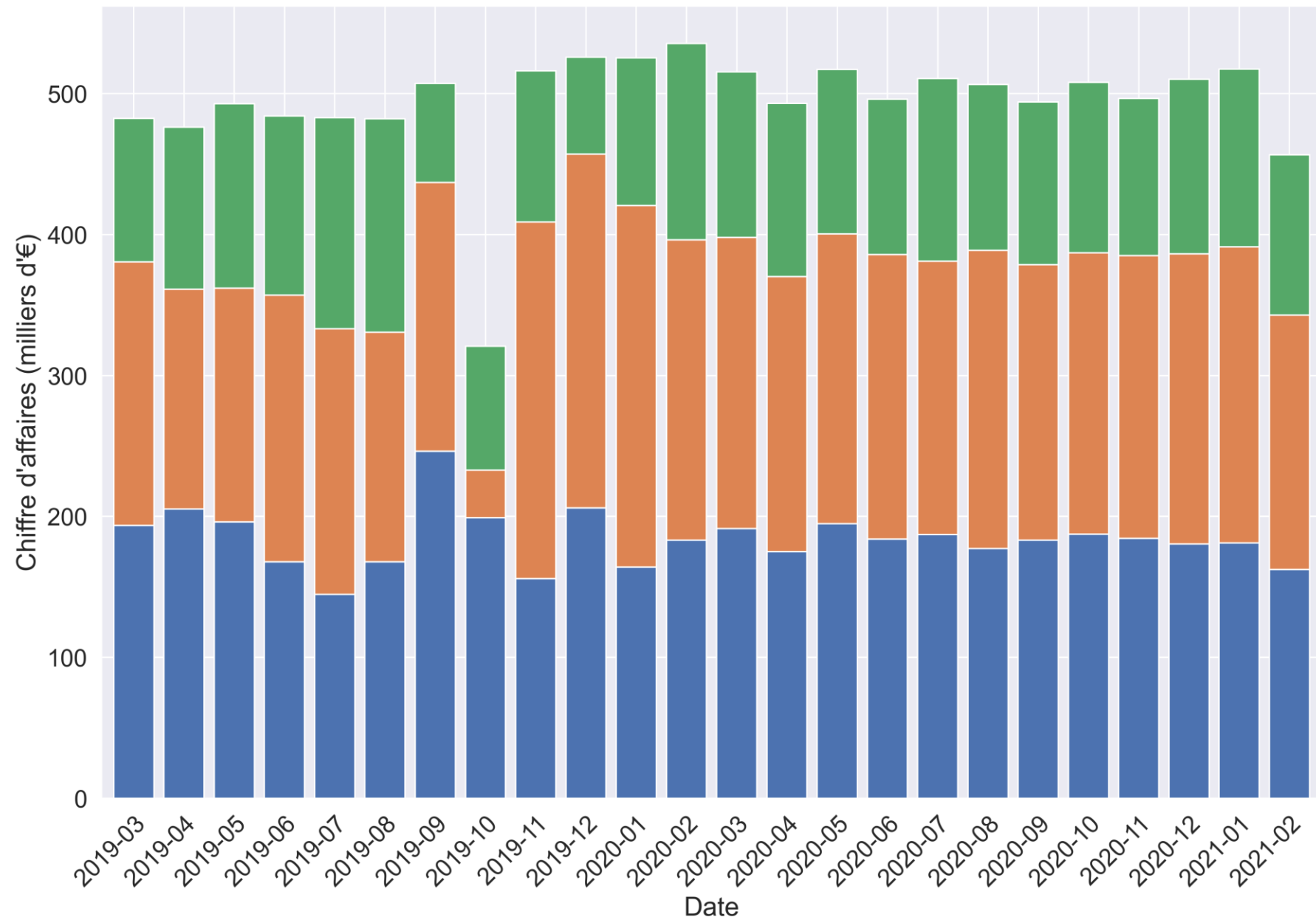
- Jointures des tables (à gauche): transaction < products (id\_prod) < customers (client\_id)
- Vérification des lignes sans valeurs (produit 0\_2245 : ni prix, ni catégorie)
- Quelques modifications (âge, date mois\_année, date jour\_mois\_année)



# Lapage

- 1) Chiffre d'affaire mensuel par catégorie
- 2) Nombre de vente par jour et par catégorie
- 3) Bilan des ventes en lignes
- 4) Répartition du C.A. entre les clients
- 5) Répartition en âge et en genre de la clientèle

# 1) Chiffre d'affaires mensuel et par catégorie

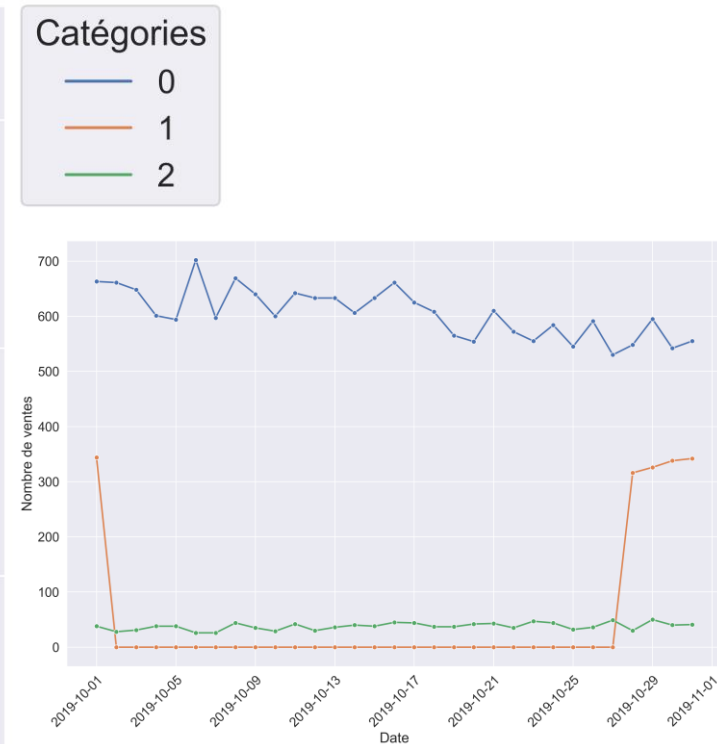
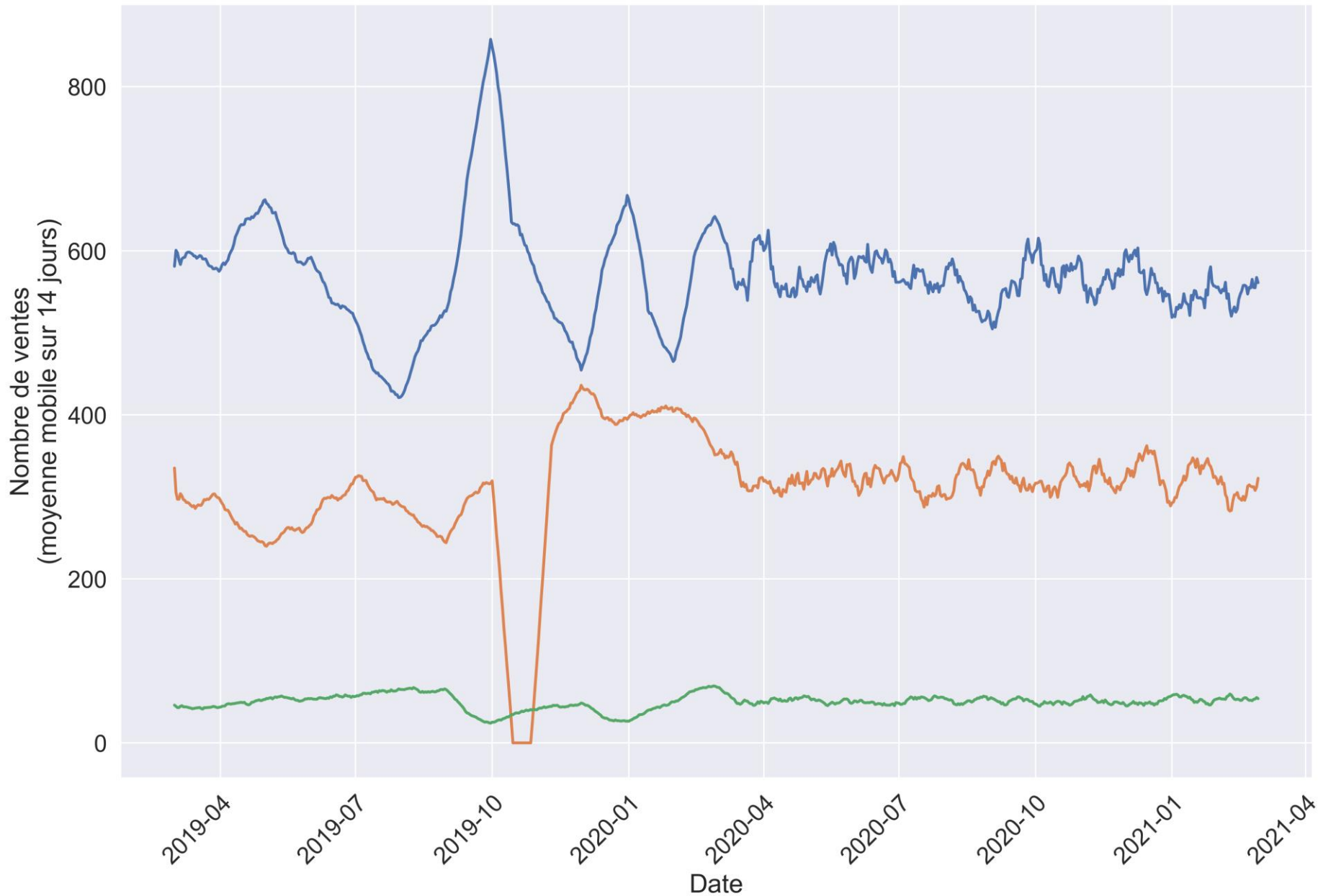


Chiffre d'affaire  
total : 11 853 728.68  
€

Sur les 24 mois de l'analyse :

- C.A. mensuel moyen aux alentours de 500 000€ par mois.
- C.A. par catégorie relativement stable.
- Problème en Octobre 2019 sur les ventes de la catégorie 1.

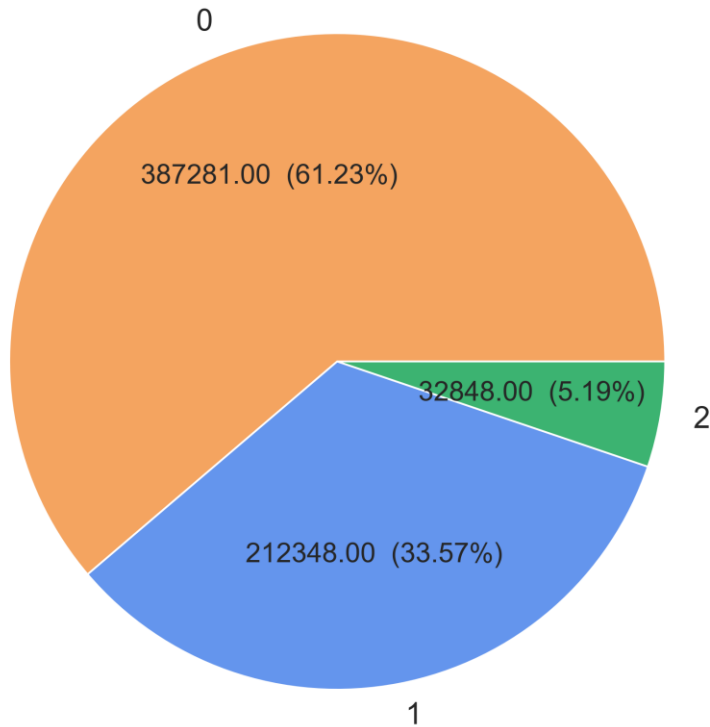
## 2) Nombre de ventes par jour et par catégorie



- Les ventes des produits de catégorie 1 sont nulles entre le 02 et le 27 octobre 2019.

### 3) Bilan de la vente en ligne des produits des trois catégories

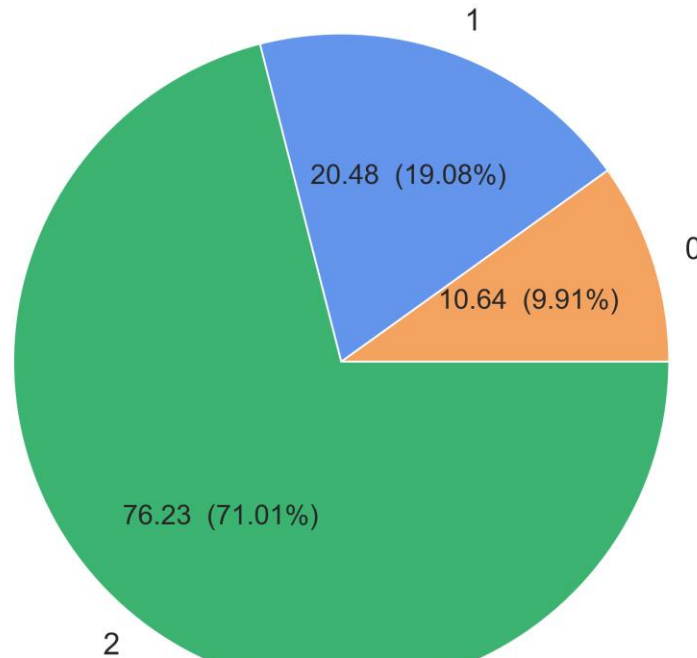
Nombre de ventes



Le nombre de ventes n'est pas identique entre les catégories  
( $\chi^2 = 297946.89$ , p-value=0) :

Catégorie 0 (61%) > Catégorie 1 (34%) > Catégorie 2 (5%).

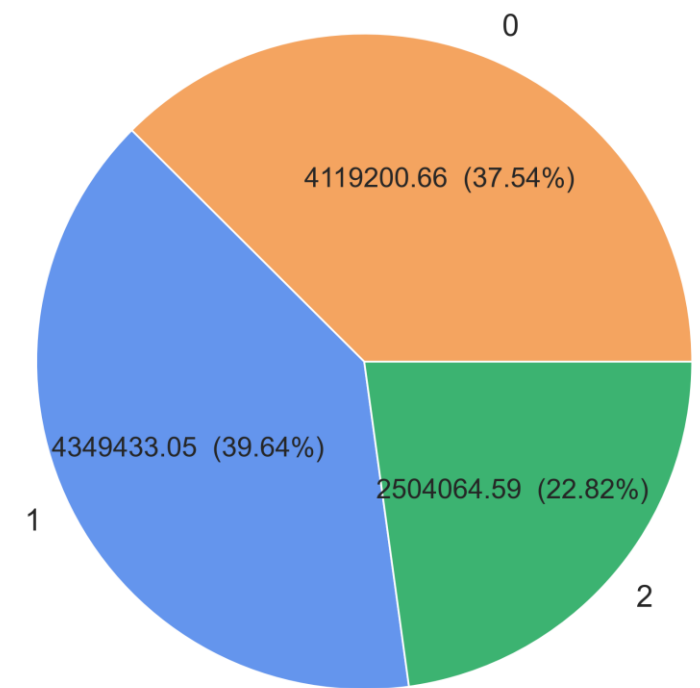
Prix moyen (€)



Le prix moyen des articles est significativement différents entre les catégories (Test de Kruskal-Wallis, Statistique = 324919.9, p-value = 0)

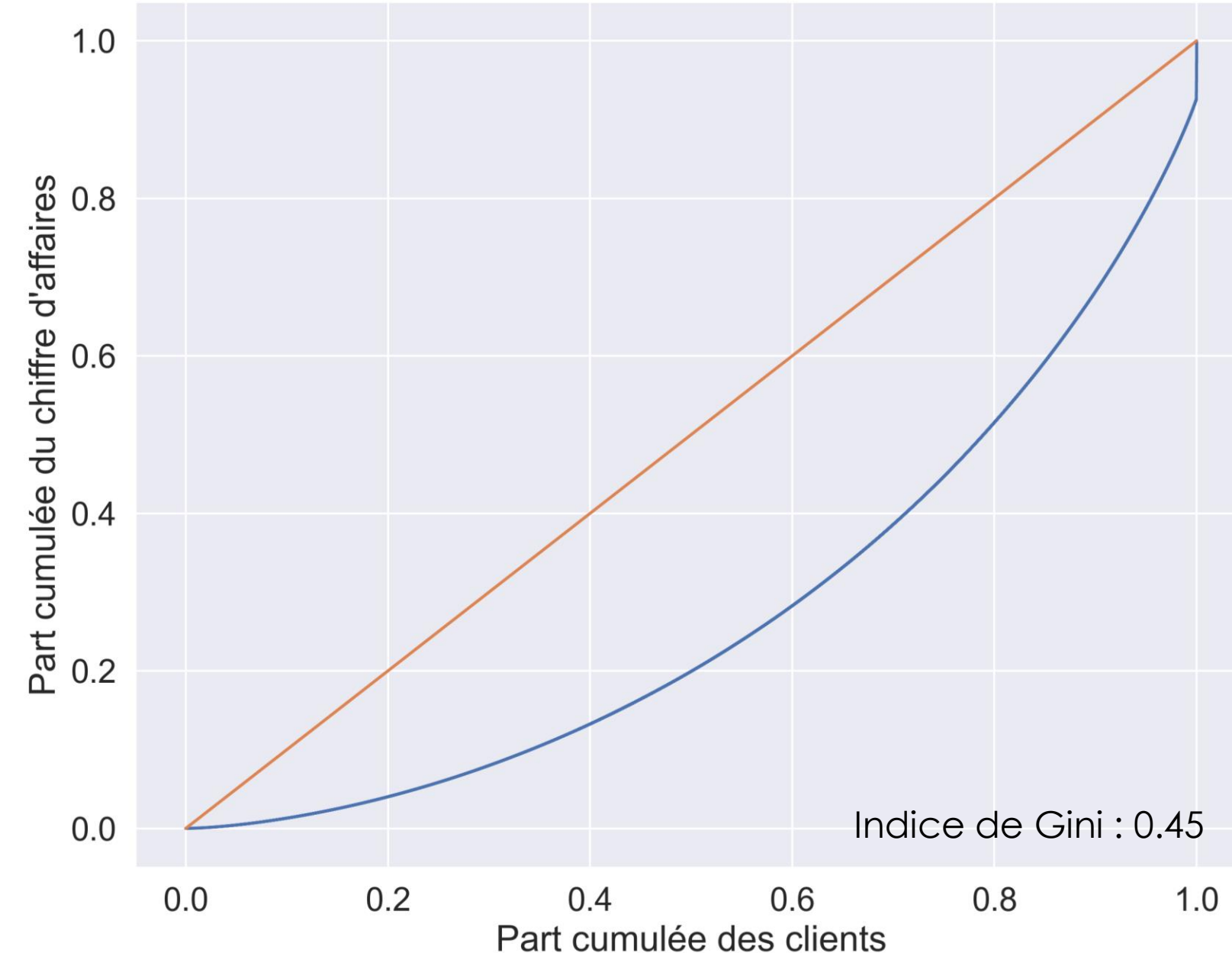
Catégorie 2 > Catégorie 1 > Catégorie 0

Chiffre d'affaires (€)



Au total, le chiffre d'affaires des ventes des produit de catégorie 1 est le plus important, suivi par celui des produit de catégorie 0 puis 2.

## 4a) Répartition du chiffre d'affaires entre les clients



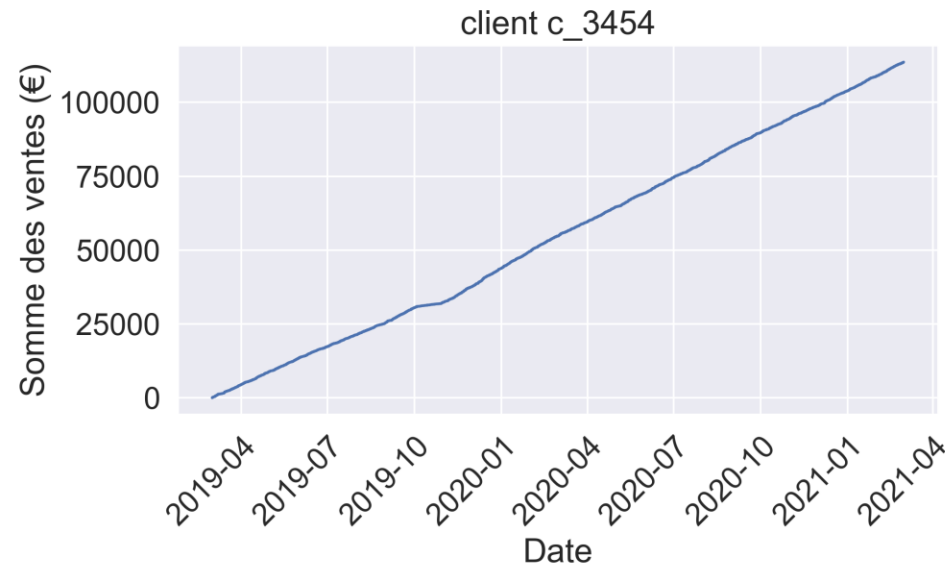
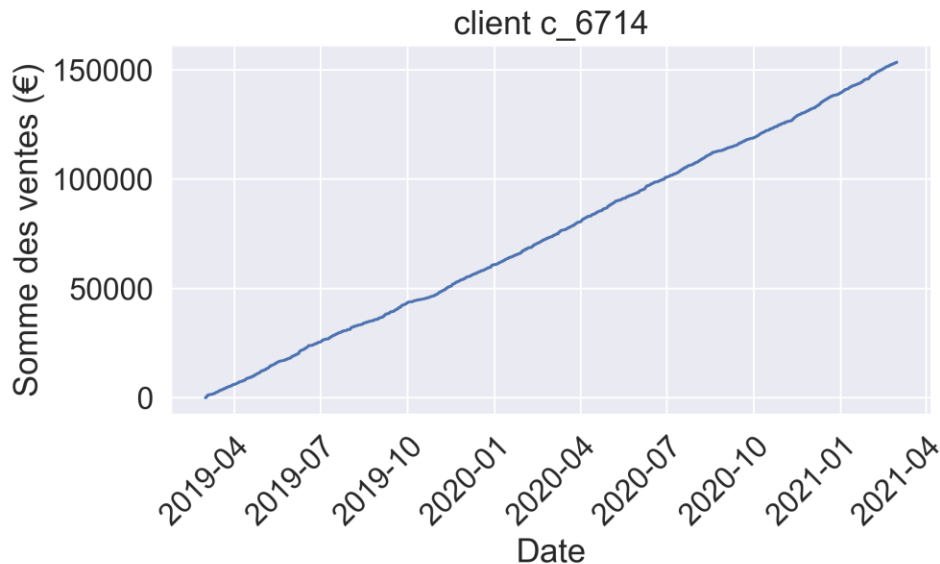
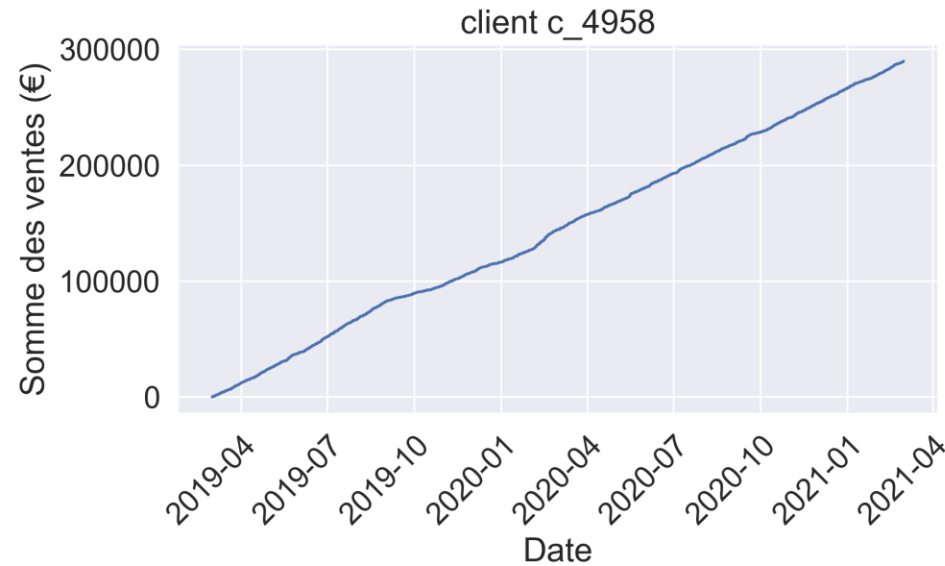
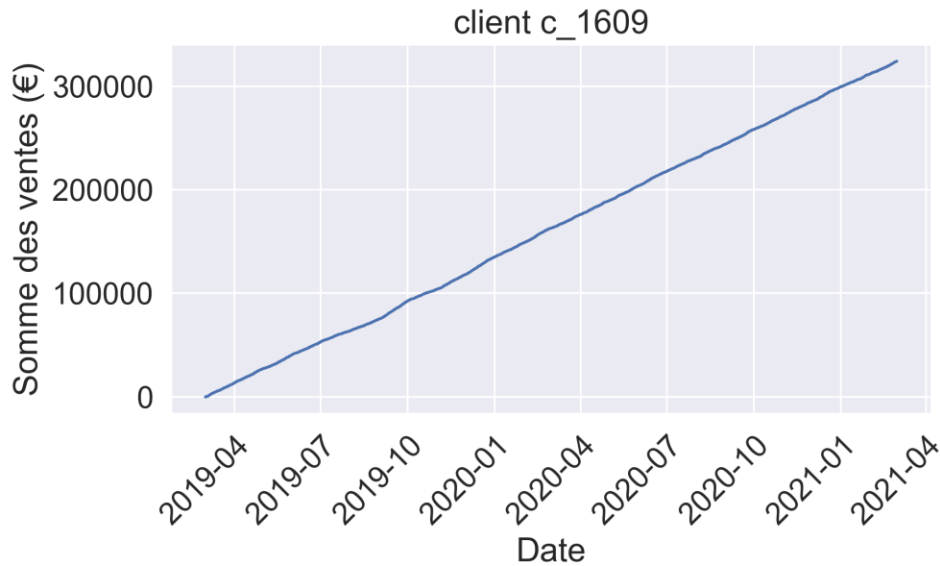
### Courbe de Lorenz:

- Le chiffre d'affaires est globalement bien réparti entre les clients
- Quelques clients ont une part très importante du chiffre d'affaires.



# 4b) Présentation des quatre meilleurs clients

Somme des ventes des quatres meilleurs clients



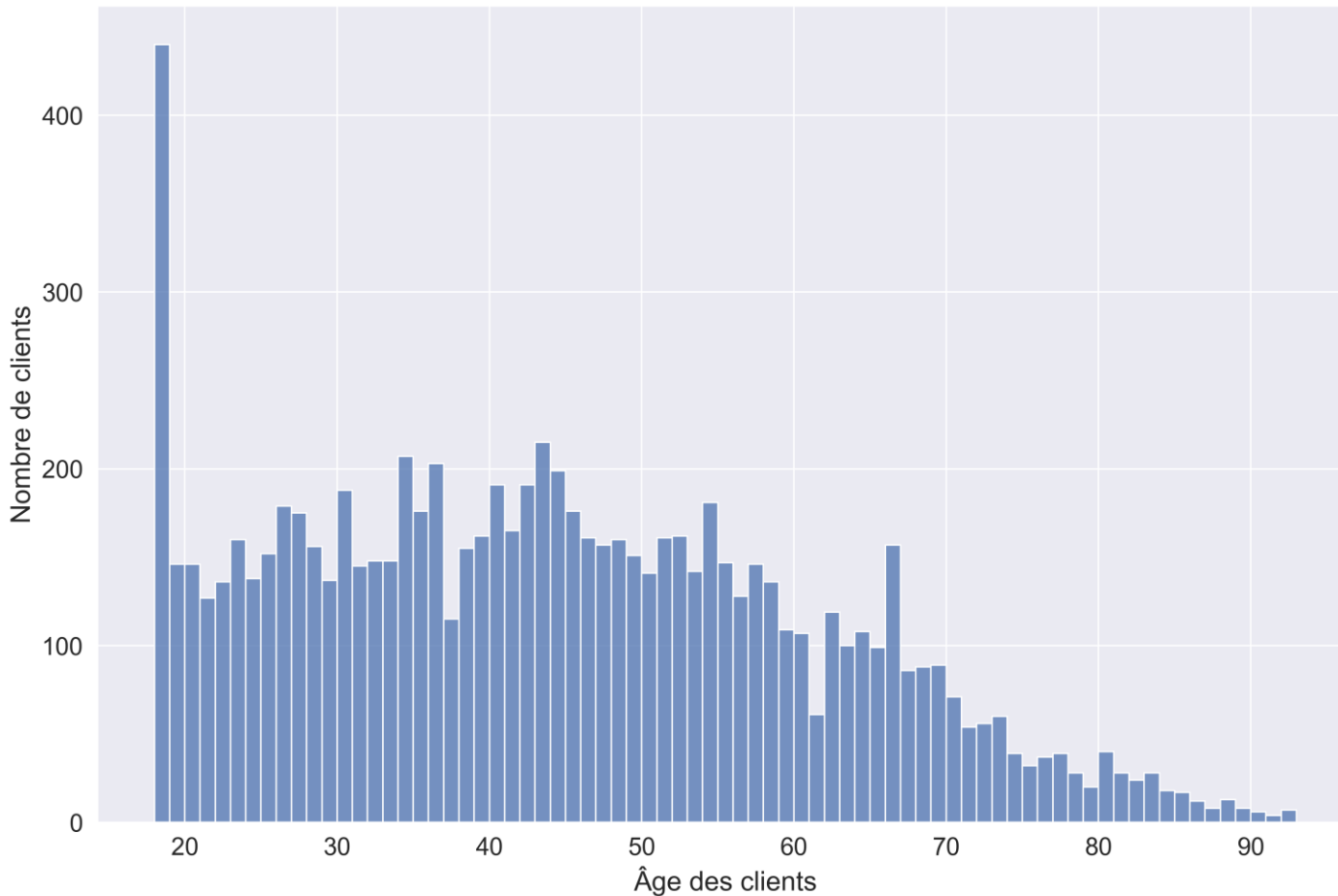
➤ Ces quatre clients représentent plus de 7% du chiffre d'affaires total.

➤ Leurs achats sont relativement constant tout au long des deux ans.

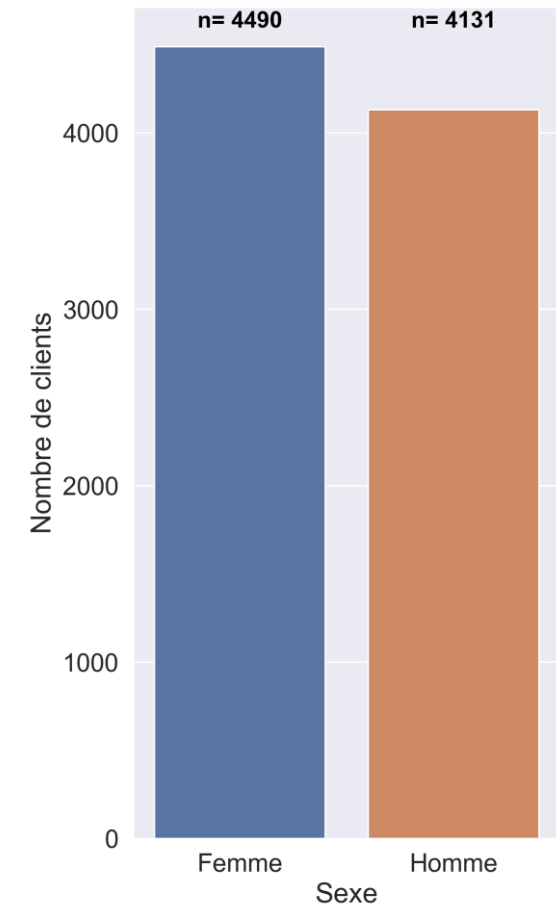
➤ La somme des ventes du client c\_3454 (4<sup>ème</sup>) est 21 fois supérieure à celle du client c\_1570 (5<sup>ème</sup>).

Pour la suite des requêtes, ces quatre individus seront mis à

## 5) Répartition en âge et en genre de la clientèle



- L'âge des clients est inégalement réparti.
- Les clients de 18 ans représentent à eux seuls 5,10 % de notre clientèle totale.
- Pour la suite des analyses, les requêtes de Julie mettent en lumière trois groupes d'âge : 18-30, 31-50, 51+.



- Le nombre de clients *femmes* est significativement supérieur au nombre de clients *hommes* ( $\chi^2= 14,95$ , p-value=  $1.10^{-4}$ ).
- Elles représentent 52% de notre clientèle contre 48% d'*hommes*.



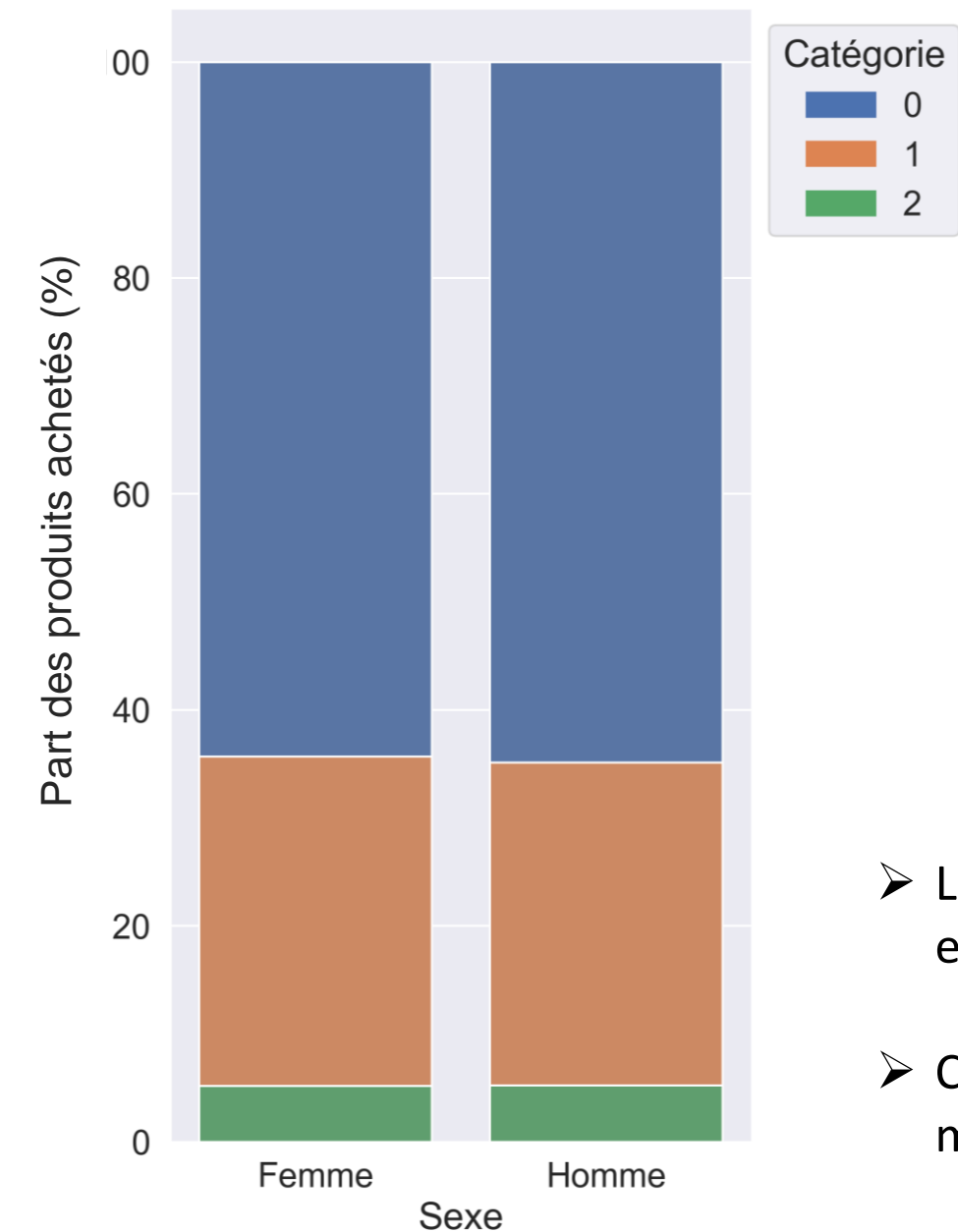
# Lapage

1) Relation entre le genre et les catégories de livres achetés

Relation entre l'âge et :

- 2) La taille du panier moyen
- 3) Le montant du panier moyen
- 4) La fréquence d'achat
- 5) Le montant total des achats
- 6) Les catégories des livres achetés

# 1) Relation entre le genre et les catégories de livres achetés

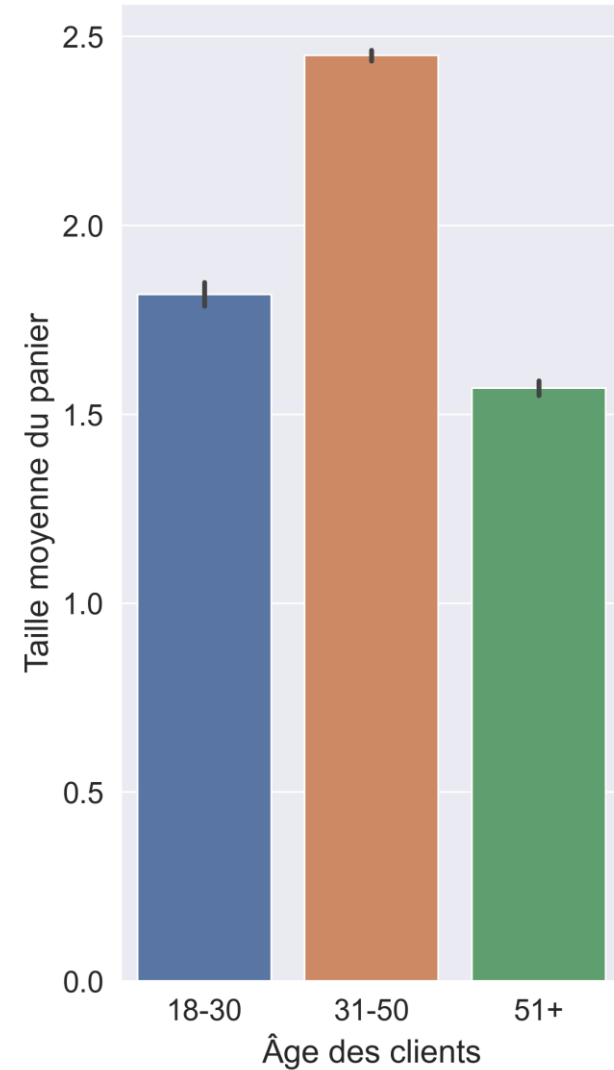
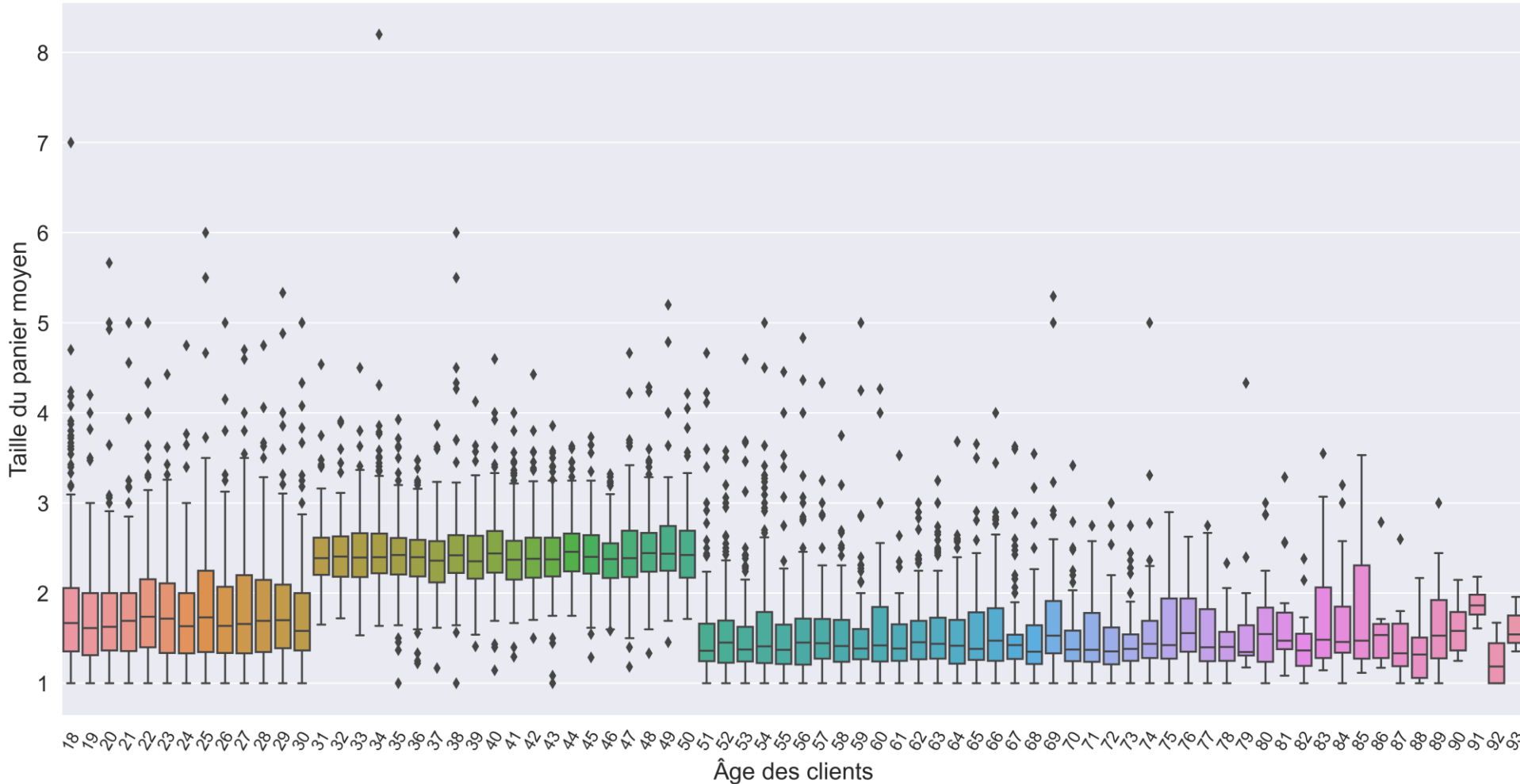


		Âge des clients	
		Femme	Homme
Catégorie	0	200793 (61,01 %)	186488 (61,47 %)
	1	111331 (33,83 %)	101017 (33,30 %)
	2	16980 (5,16 %)	15868 (5,23 %)

*Nombre de ventes par genre et catégorie*

- Les catégories des livres achetés sont statistiquement différentes entre les *hommes* et les *femmes* ( $\chi^2= 20,21$ , p-value=  $4.10^{-5}$ )
- Cependant, lorsque l'on regarde les chiffres, la différence est minime.

## 2) Relation entre age des clients et taille du panier moyen



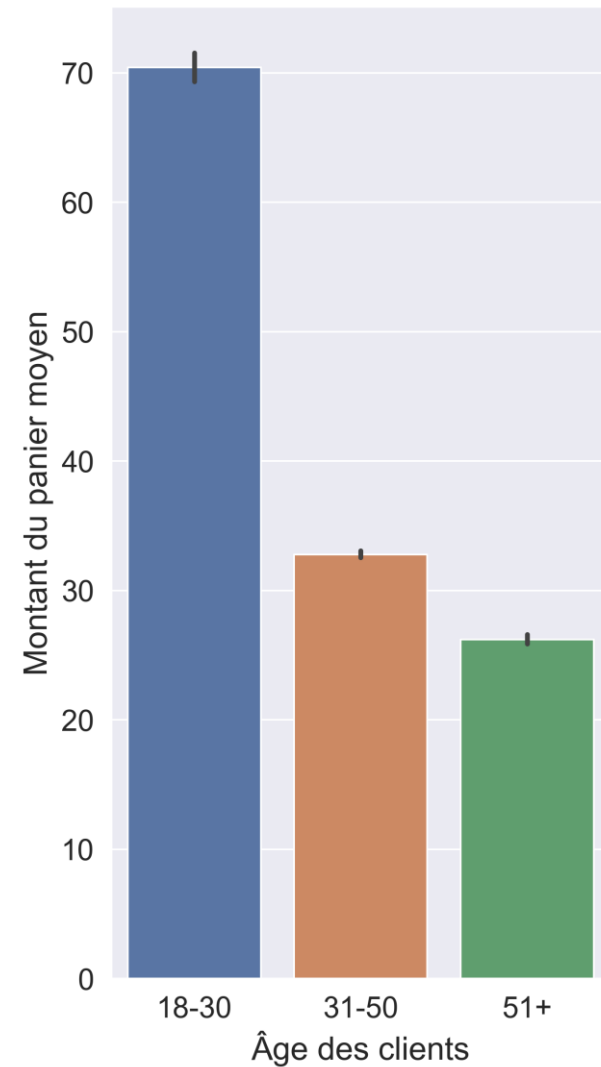
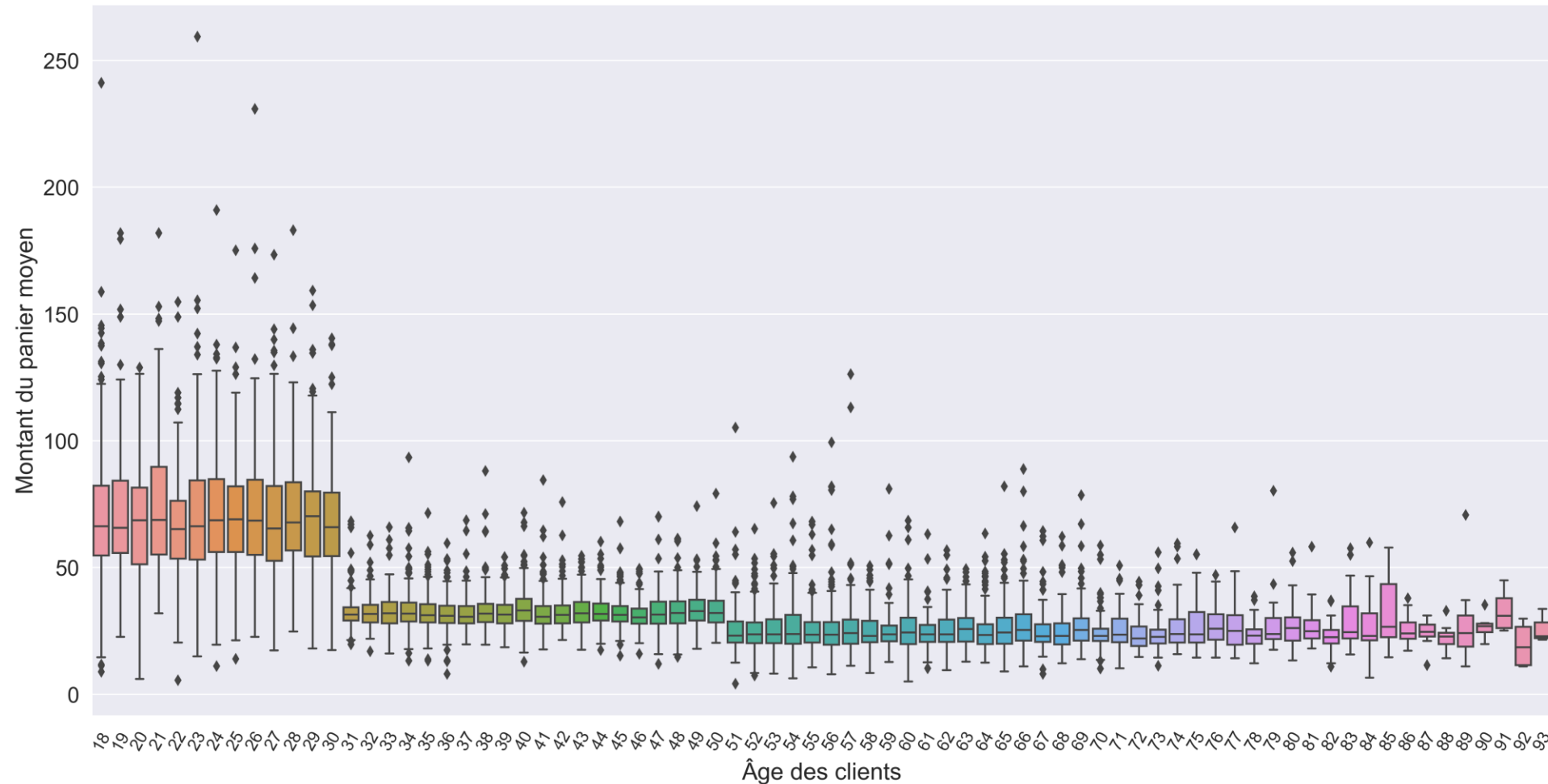
La taille du panier moyen (nombre d'articles) suit une Loi de Poisson. Nous ne pouvons donc pas réaliser un modèle linéaire (ANOVA). Pour tester l'effet des paramètres, nous pouvons comparer l'AIC de notre modèle d'intérêt à l'AIC d'un modèle nul.

Taille du panier moyen  $\sim 1 > \text{AIC} = 22654$

Taille du panier moyen  $\sim \text{âge}(\text{groupe}) > \text{AIC} = 22013$

➤ L'âge des clients affecte significativement la taille moyenne de leur panier.

### 3) Relation entre age des clients et montant du panier moyen



Pour analyser l'effet de l'âge sur le montant du panier moyen, nous pouvons réaliser un modèle linéaire (ANOVA), mais les conditions d'applications ne sont pas remplies. > Modèle linéaire généralisé (Family = Gaussian)

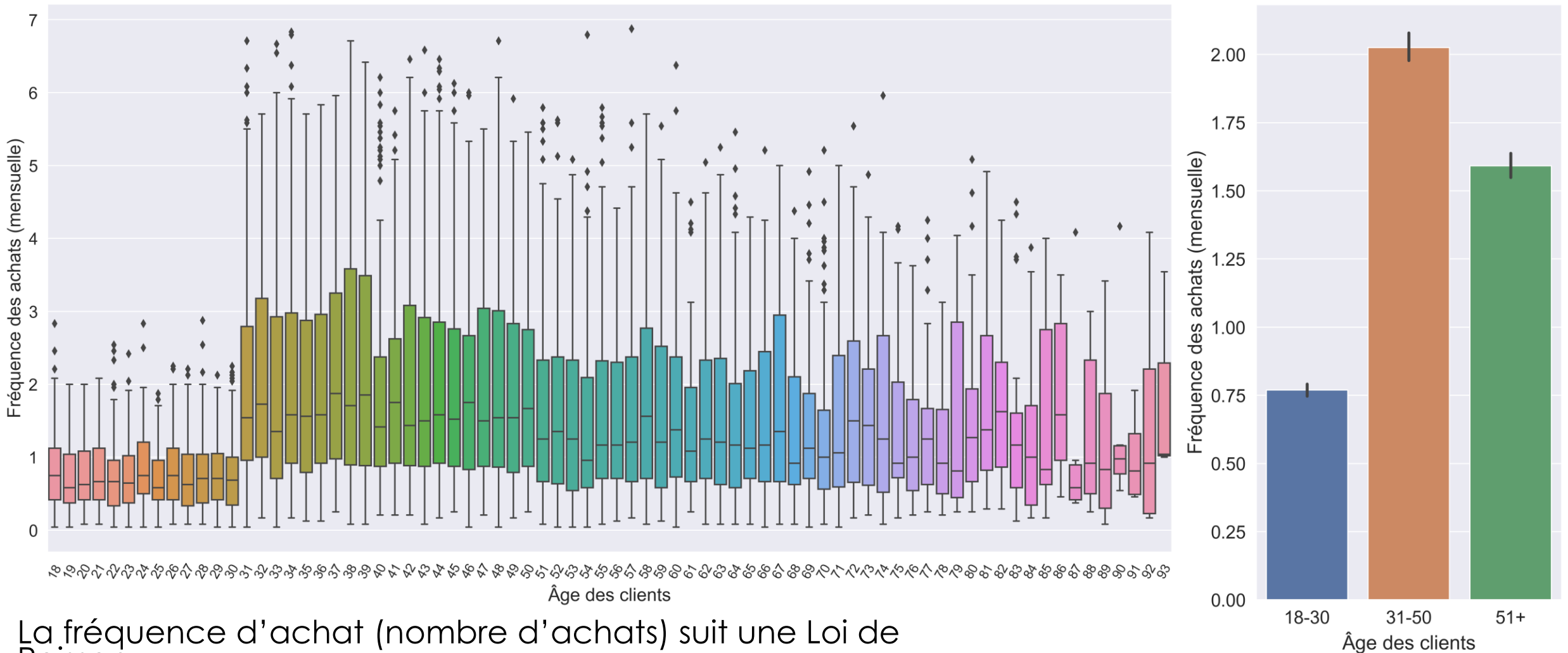
Nombre de produits achetés ~ 1 > AIC = 73727

Nombre de produits achetés ~ âge(groupe) > AIC = 66200

➤ L'âge des clients affecte significativement le montant du panier moyen.

# 4) Relation entre age des clients et frequence d'achat

Fréquence d'achats = nombre de commandes (sessions) / client / 24



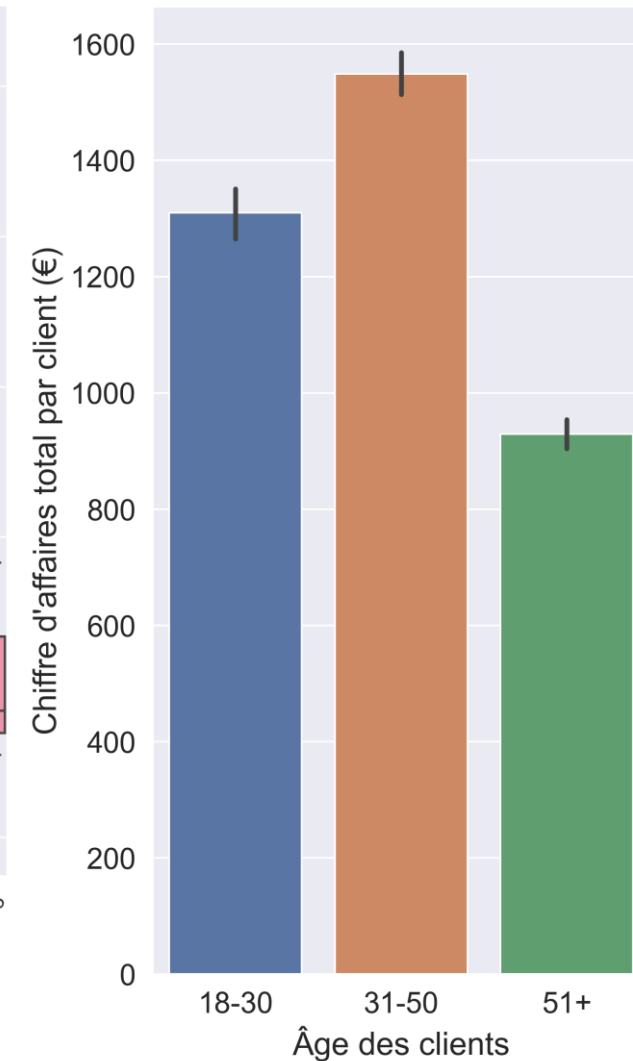
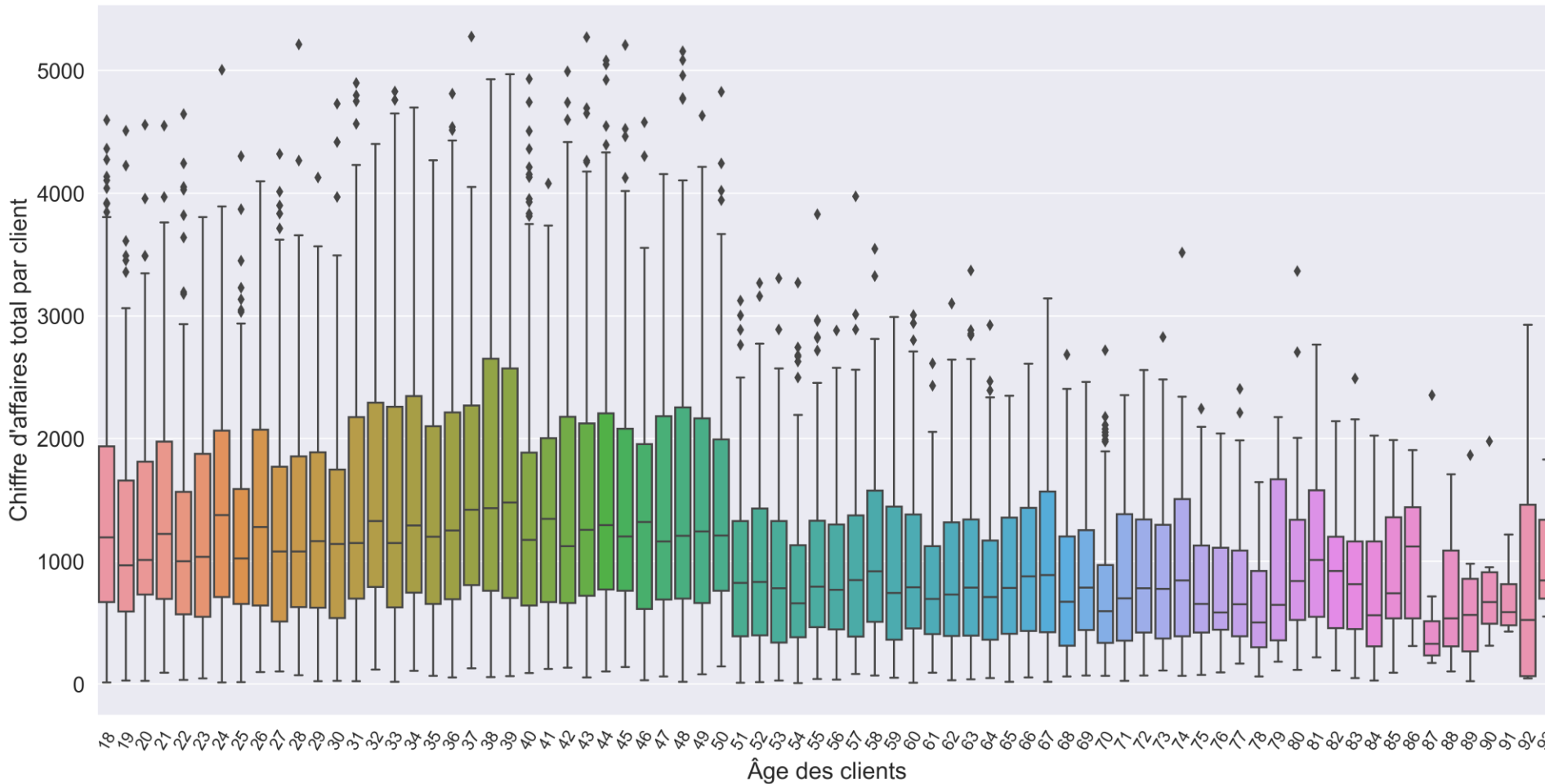
La fréquence d'achat (nombre d'achats) suit une Loi de Poisson.

Fréquence d'achat  $\sim 1 > AIC = 25380$

Fréquence d'achat  $\sim \text{âge(groupe)} > AIC = 24056$

➤ L'âge des clients affecte significativement leur fréquence d'achat.

# 5) Relation entre age des clients et montant total des achats



Pour analyser l'effet de l'âge sur C.A. total > Modèle linéaire généralisé (Family = Gaussian)

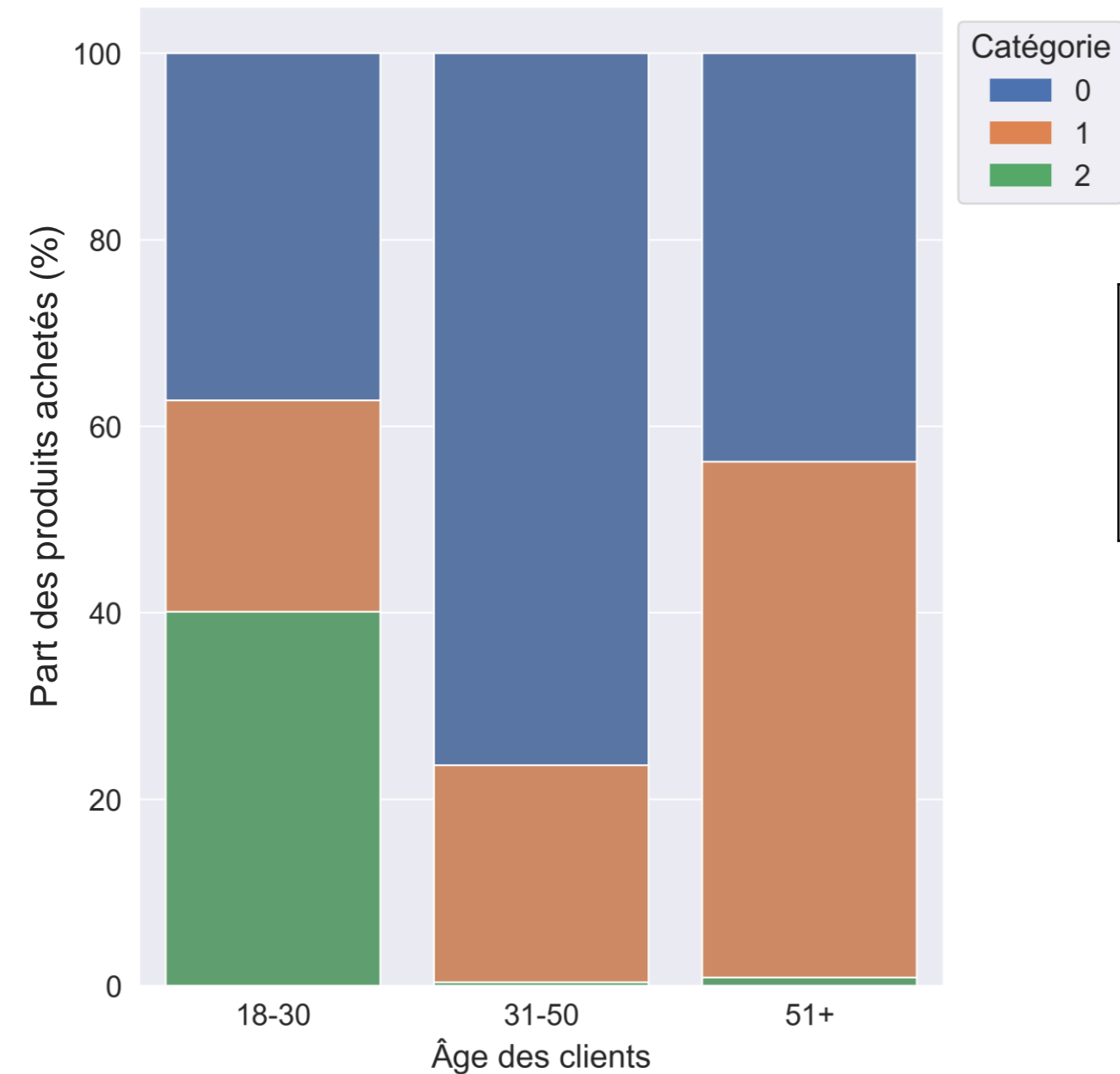
Nombre de produits achetés ~ 1 > AIC = 135102

Nombre de produits achetés ~ âge(groupe) > AIC = 134406

➤ L'âge des clients affecte significativement le montant total des achats.



## 6) Relation entre l'âge des clients et catégories des livres achetés



		Âge des clients		
		18-30	31-50	51+
Catégorie	0	13273 (22%)	299245 (76%)	71345 (43%)
	1	22368 (38%)	92811 (24%)	91450 (56%)
	2	23867 (40%)	1475 (~0%)	1494 (1%)

*Nombre de ventes par âge et catégorie*

- L'âge des clients influence significativement les catégories de produits achetés ( $\chi^2 = 266441,7$ ; p-value = 0).

# Bilan des requêtes de Julie sur le comportement des clients

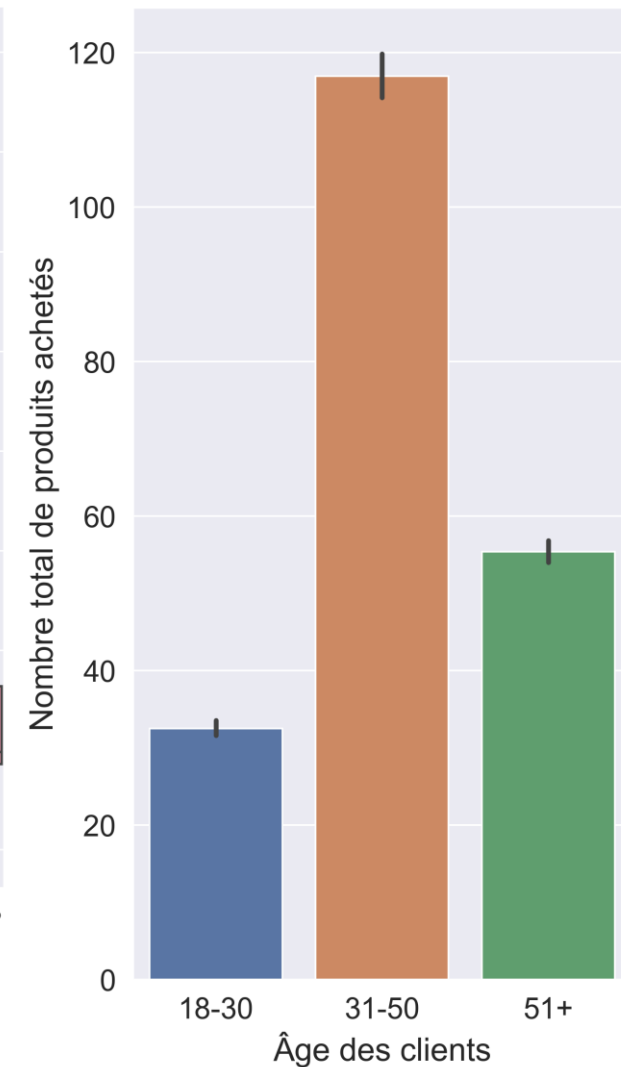
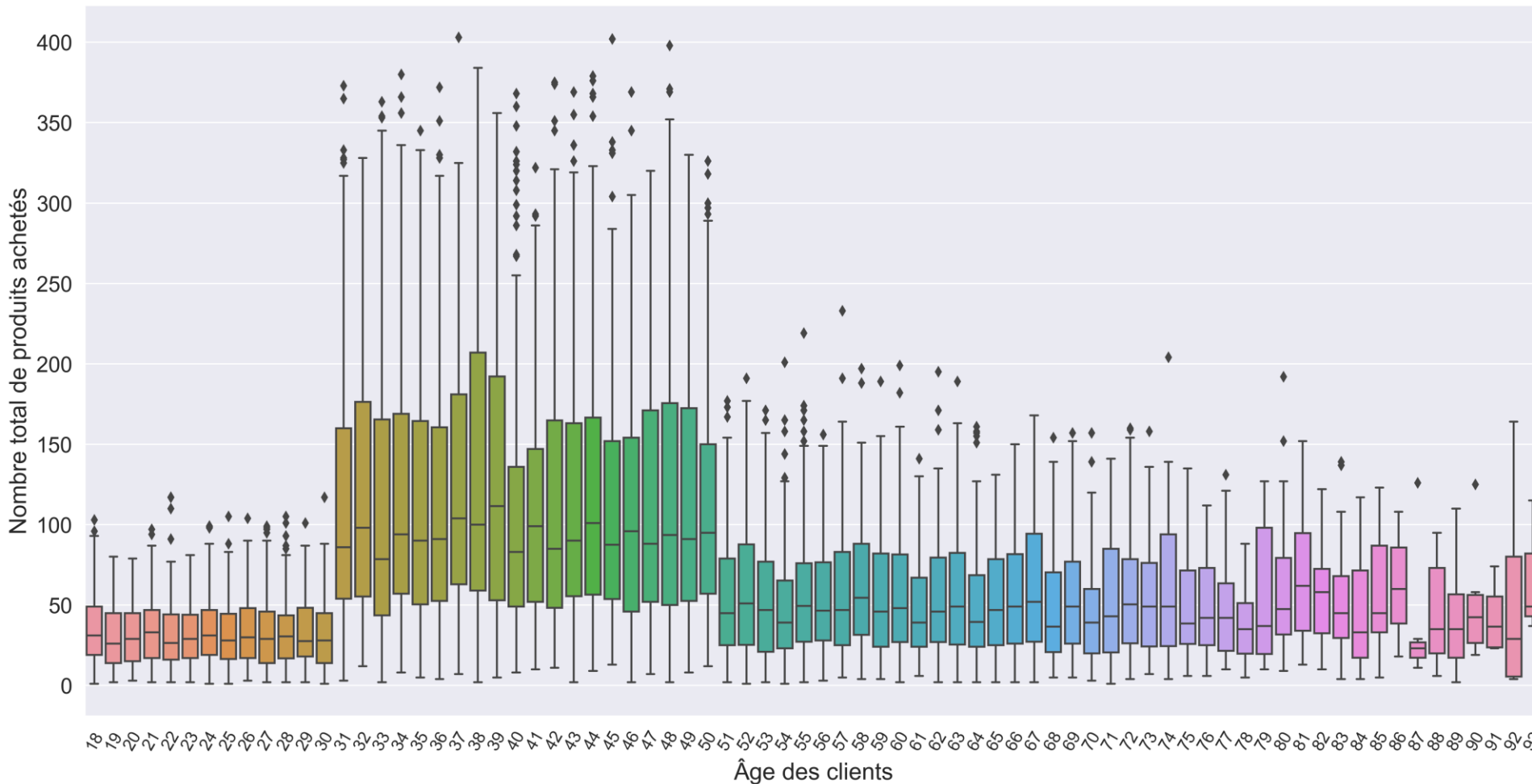
## Les trois groupes d'âges ont des comportements différents :

- Le groupe 18-30 : Achats peu fréquents de produits relativement coûteux, plus spécialement de catégorie 2.
- Le groupe 31-51 : Achats fréquents de nombreux articles à bas tarif, majoritairement de catégorie 0.
- Le groupe 51+ : Achats fréquents de peu d'articles à bas tarif, avec un panier moyen peu élevé, majoritairement des produits de catégorie 1.

## Le genre affecte statistiquement mais faiblement la catégorie de livre acheté.



# Lien entre l'âge des clients et le nombre total de produits achetés

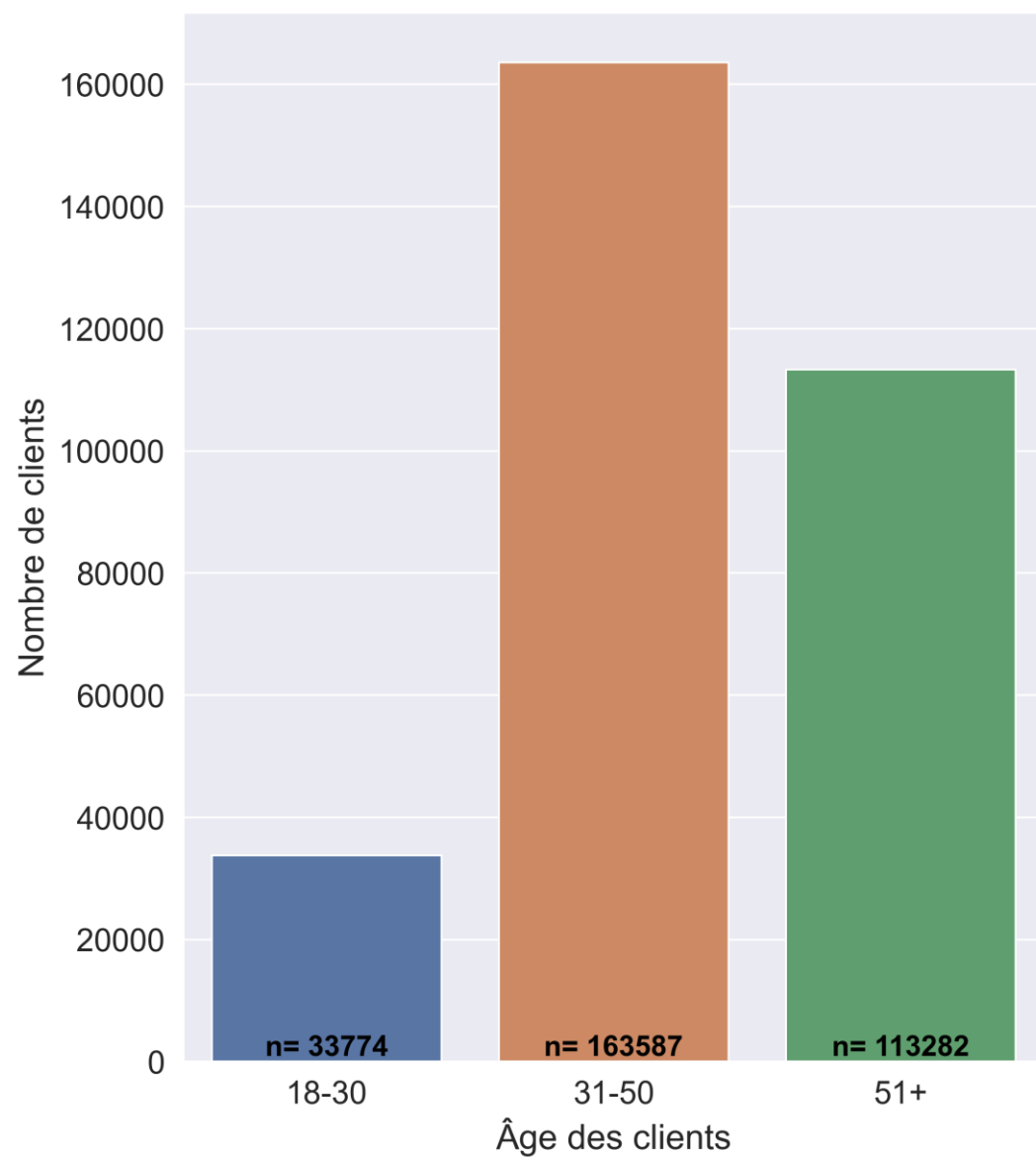


le nombre total de produits achetés suit une Loi de Poisson.

Nombre de produits achetés  $\sim 1 > AIC = 478521$

Nombre de produits achetés  $\sim \hat{\text{âge}}(\text{groupe}) > AIC =$

➤ L'âge des clients affecte significativement leur nombre total de produits achetés



➤ Les groupes d'âges ne sont pas équitablement répartis ( $\chi^2 = 82742.99$ ; p-value = 0).