

# Analyse des ventes Lapage

Nicolas Pautet

Mai 2025

# Contexte

---

Analyser des données de ventes et de comportements clients

Sources de données :

- données clients : sexe, année de naissance
- données produits : prix, catégorie de produit
- données transactions : 1 produit, 1 client, 1 session, 1 date

**Objectifs - améliorer nos ventes :**

- Mettre en avant les produits *best-sellers*
- Identifier les tendances clients convertibles en actes d'achats

# Analyses exploratoires des données

---

3 Datasets (format .csv) :

- **Clients** : 8621 entrées
- **Produits** : 3286 entrées réparties dans 3 catégories
- **Transactions** : 687 534 entrées (~ 345 k sessions d'achat) sur la période du 01/03/2021 à 28/02/2023

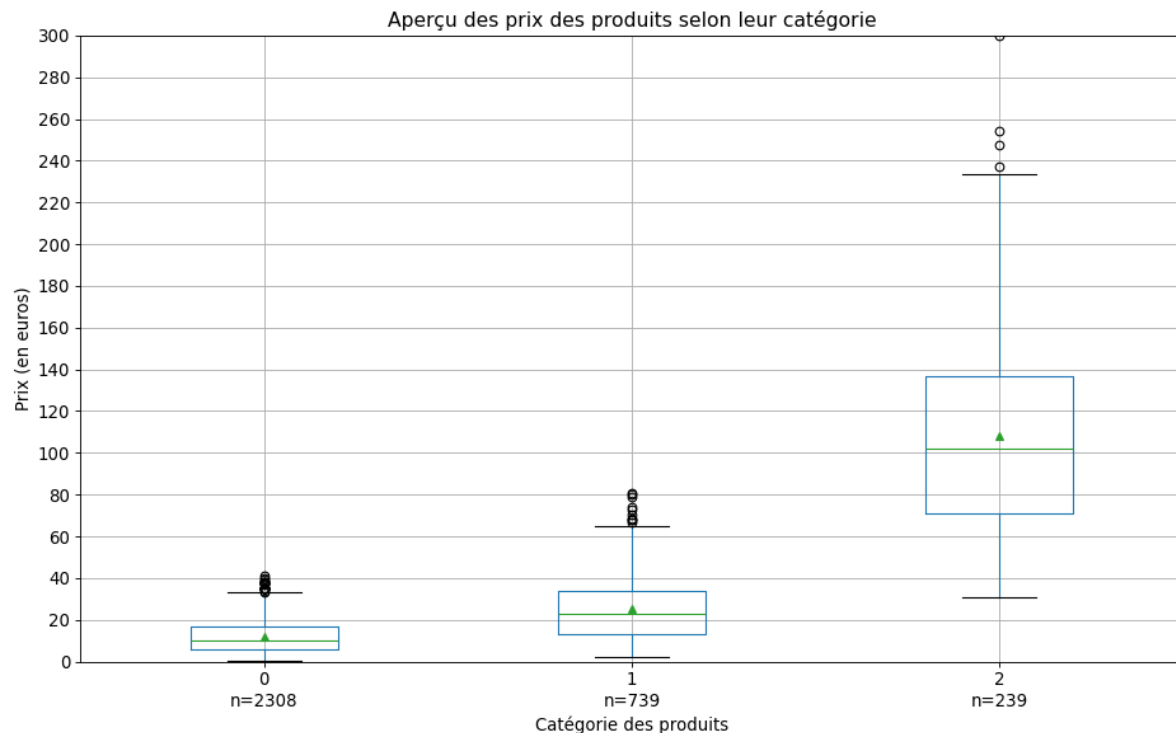
Remarques - Sur cette période :

- 21 clients de notre base de données n'ont pas fait d'achat
- 21 références de notre catalogue n'ont pas été vendues
- 99% des clients ont acheté au moins 2 fois une même référence

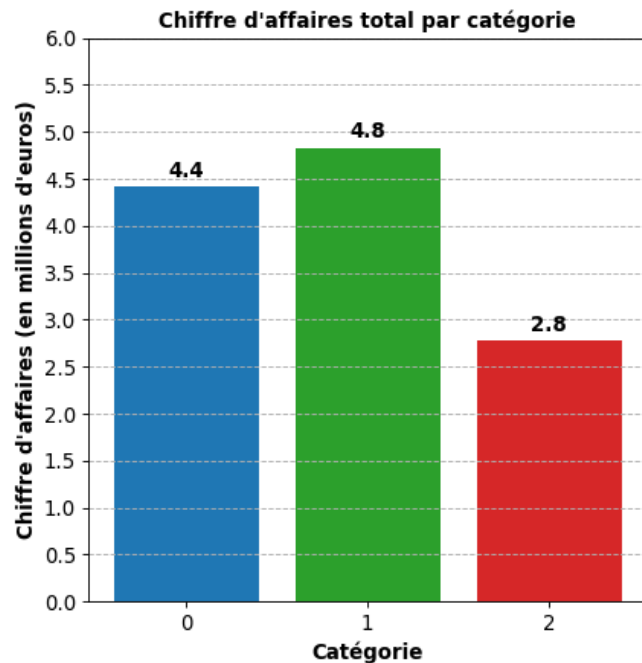
# Zoom sur notre catalogue de produits

**La catégorie**  
d'appartenance  
d'un produit a un  
impact significatif  
sur **son prix**

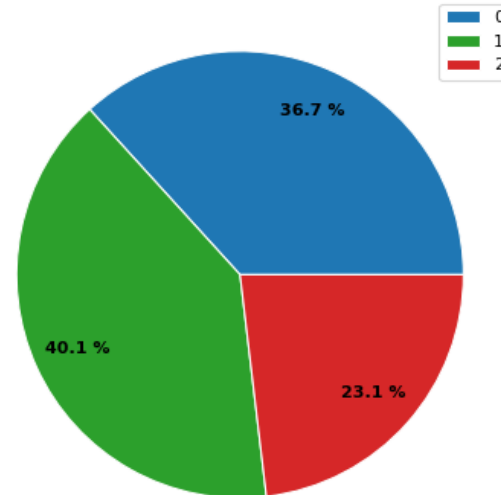
Notre catalogue :  
3286 références



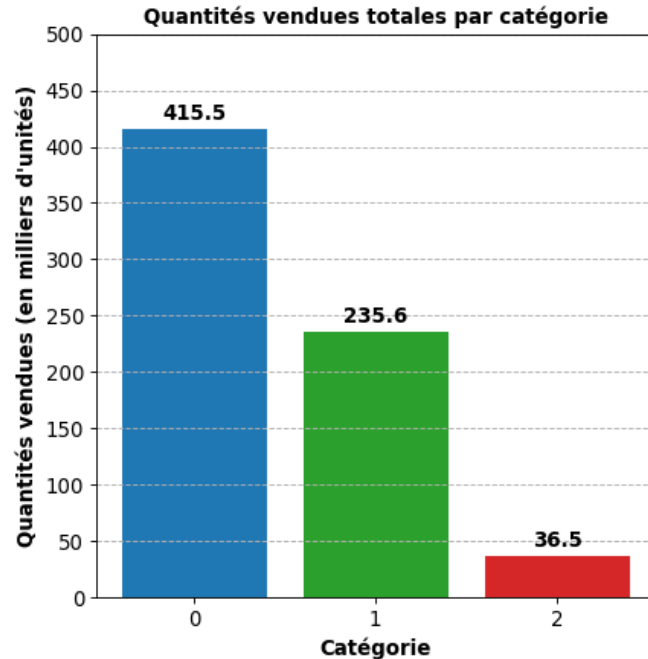
# Chiffre d'affaires sur la période - catégories



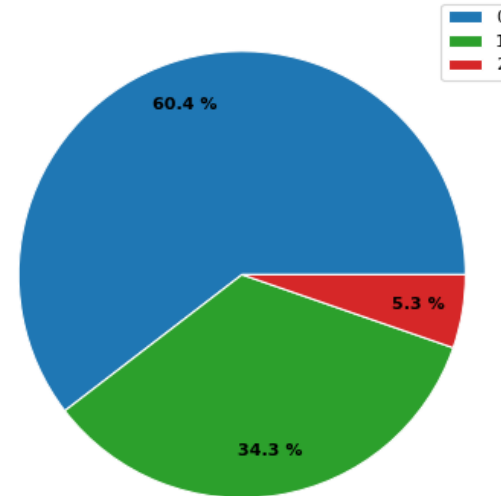
**Répartition du CA total (en %) par catégorie**



# Ventes (quantités) sur la période - catégories



**Répartition des ventes (en %) par catégorie**

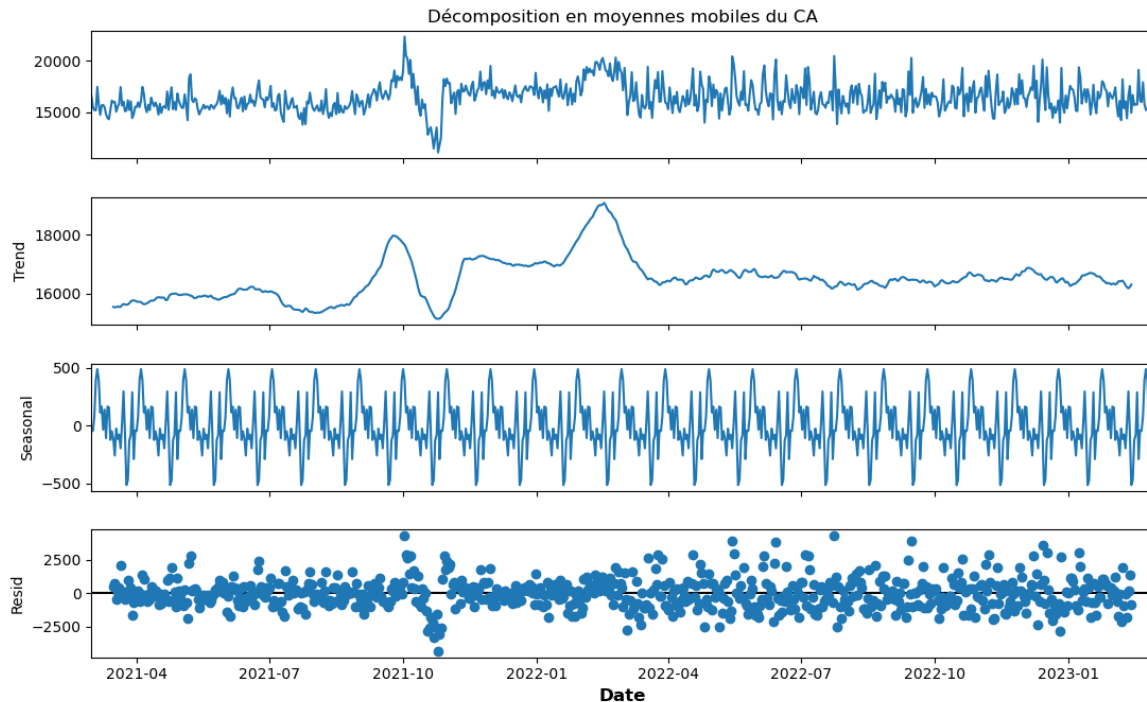


# Évolution du chiffre d'affaires journalier

Plutôt stable  
(15,5 k€ → 16,2 k€)

Composante  
prépondérante :  
**tendance globale**

Composante  
saisonnière faible

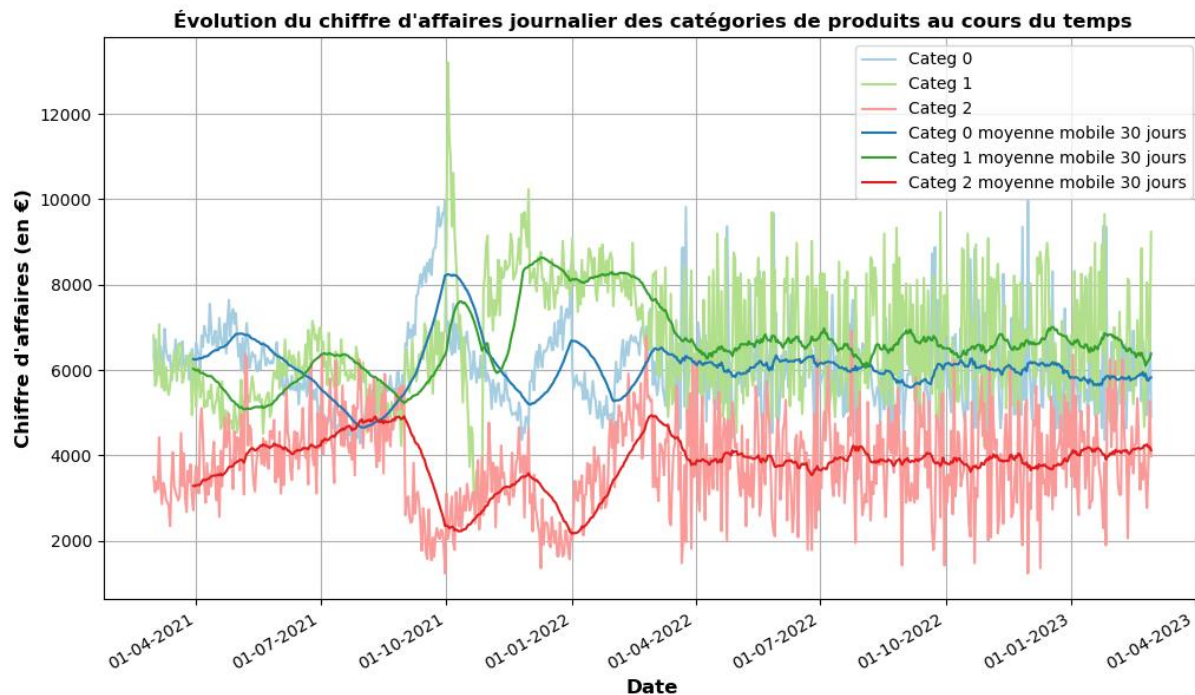


# Chiffre d'affaires journalier - par catégorie

Complémentarité  
des catégories de  
produits

CA moyen /jour :

- **0 : 6 k€**
- **1 : 6,5 k€**
- **2 : 4 k€**



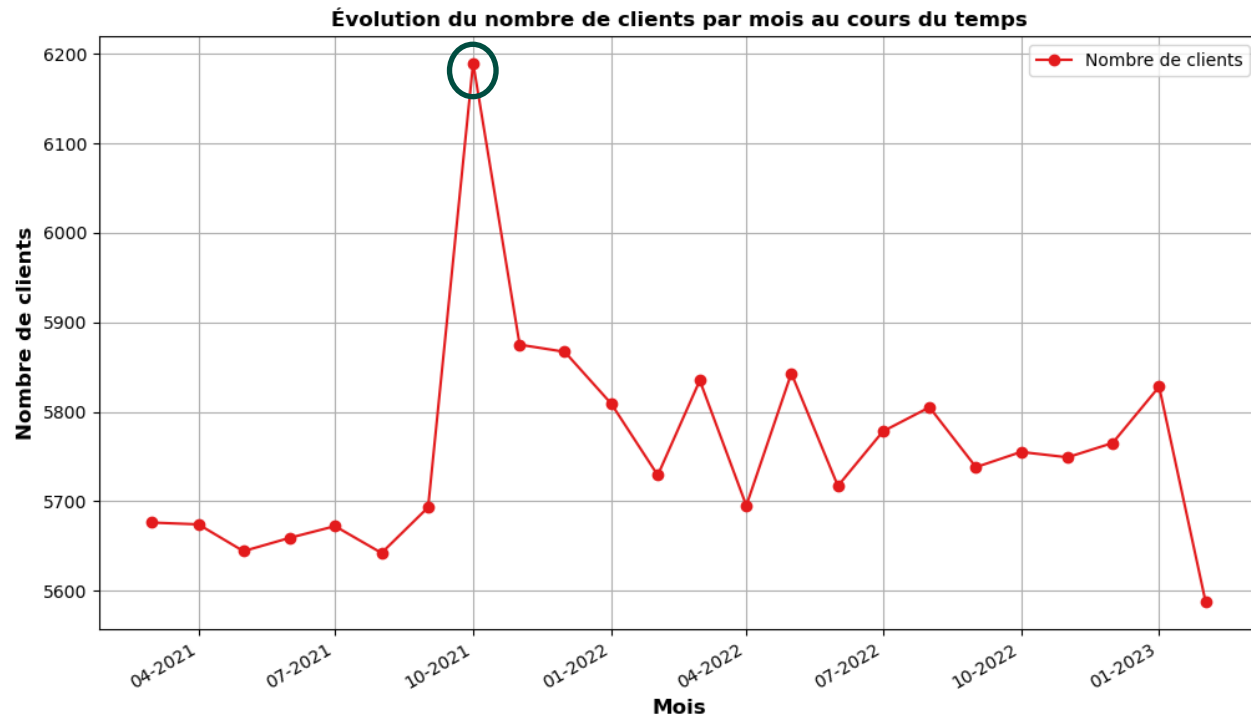


# Nombre de clients mensuels

Plutôt stable  
(~ 5 750) sur la  
période

Pic en 10/2021

En baisse sur le  
dernier mois

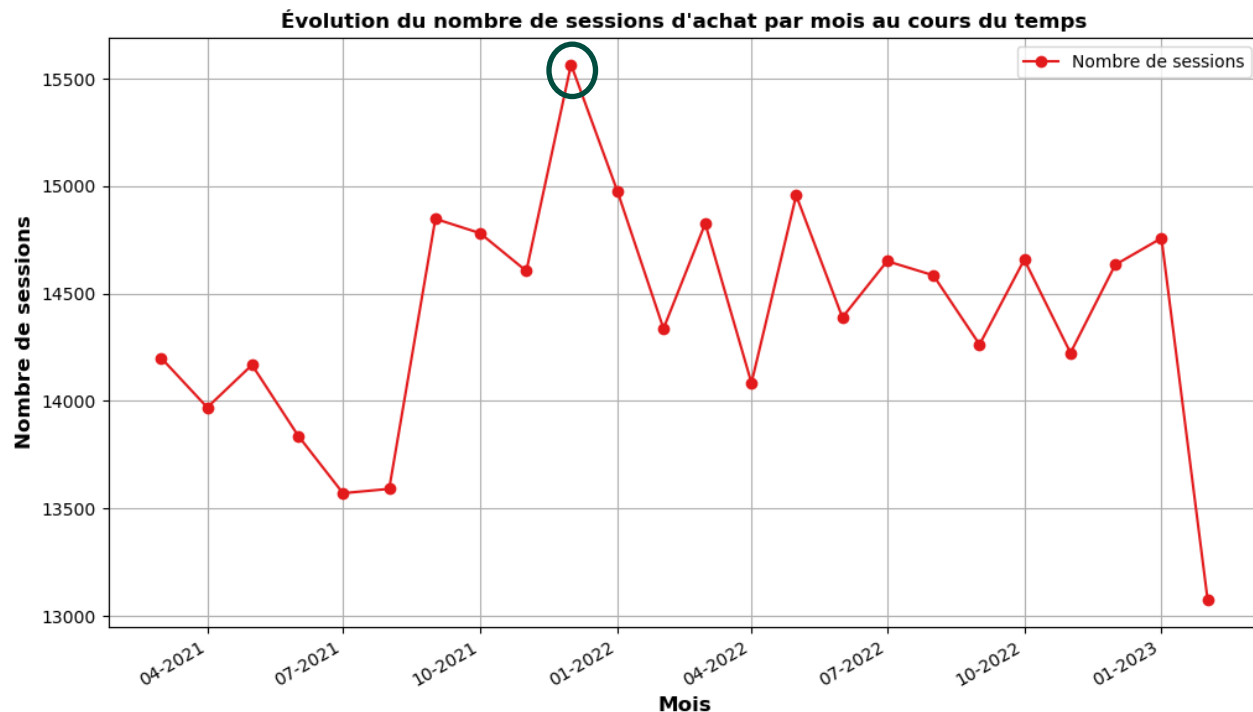


# Nombre de sessions mensuelles sur le site

Plutôt stable  
(~ 14 400) sur  
la période

Pic en 12/2021

Baisse sur le  
dernier mois

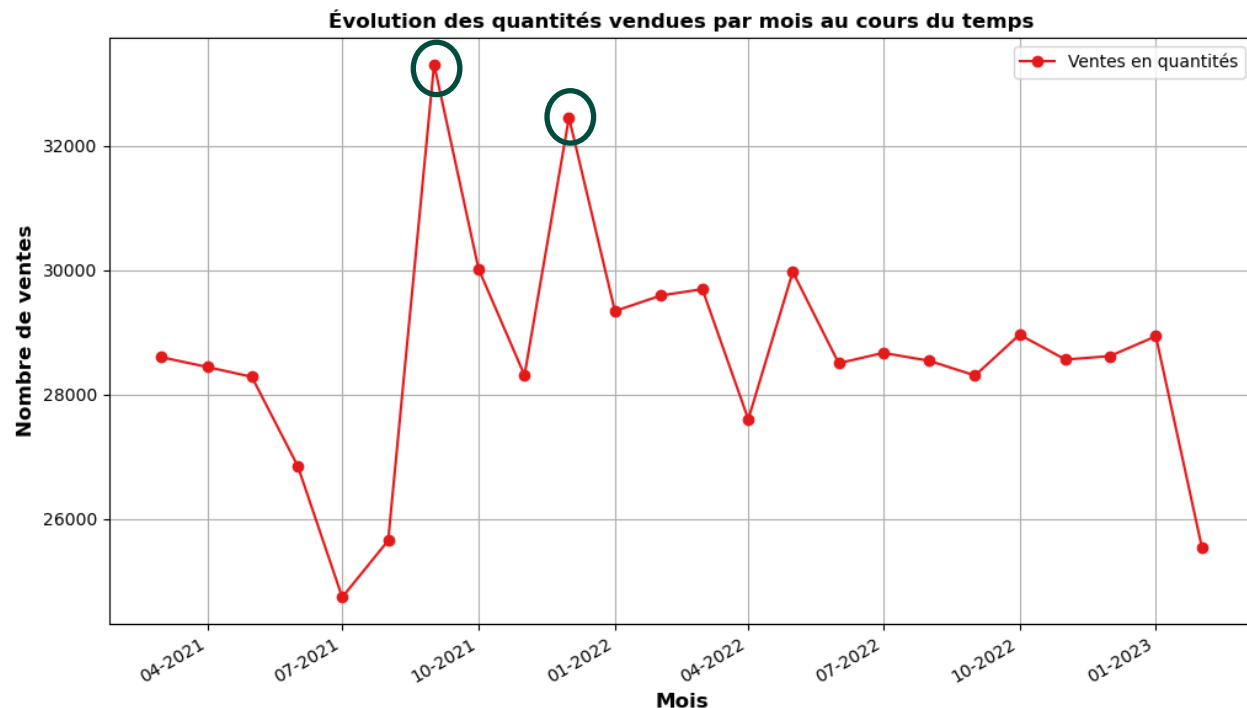


# Ventes (quantités) mensuelles

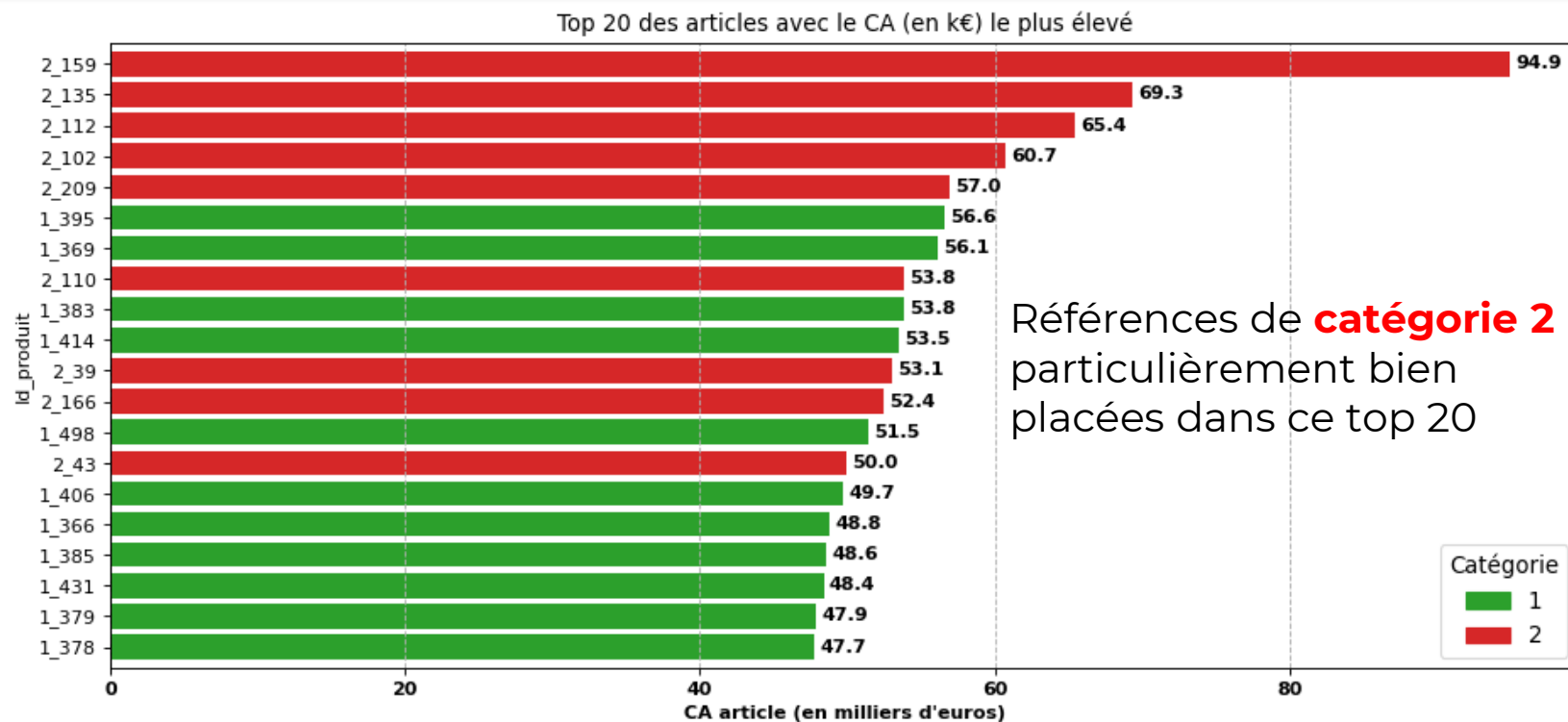
Plutôt stable  
(~ 28 500) sur  
la période

Pics fin 2021

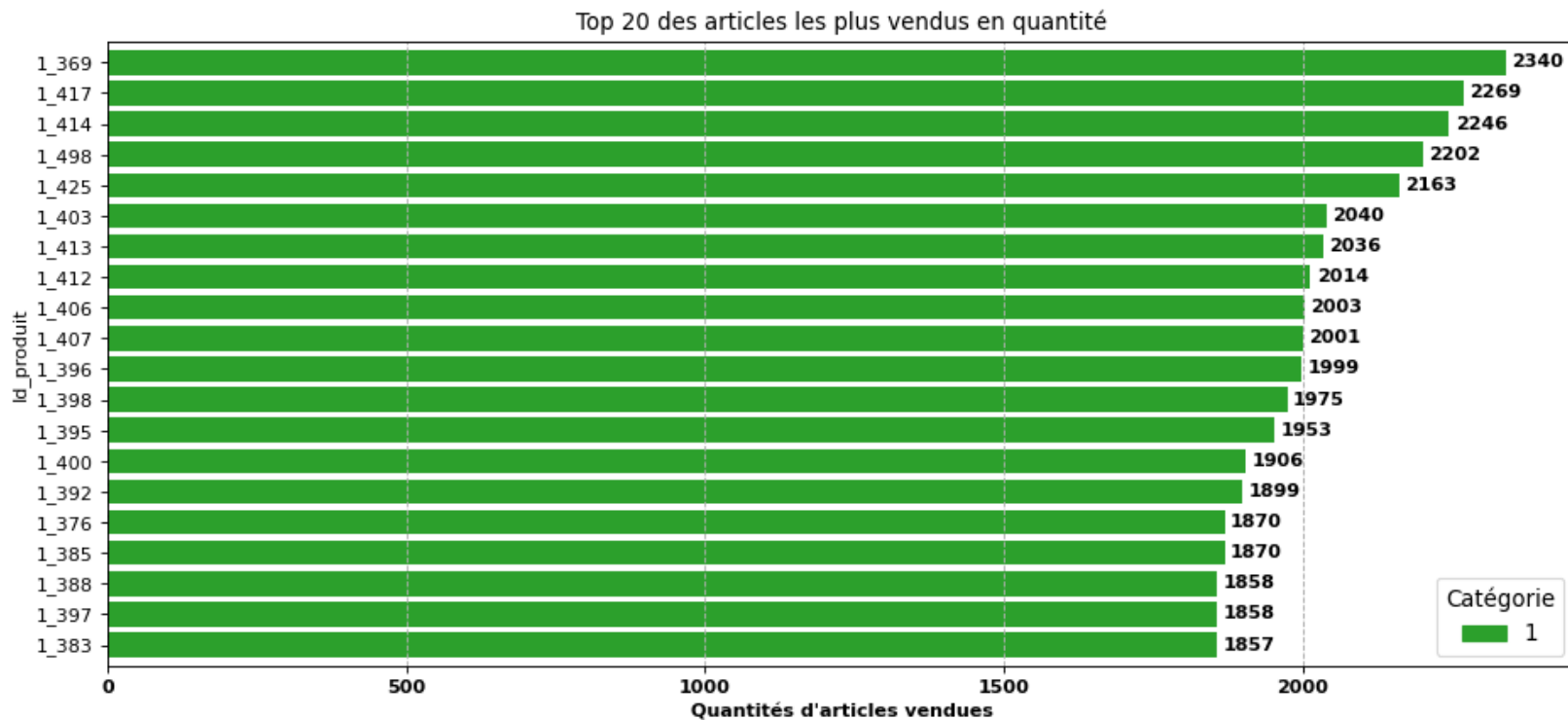
Baisse sur le  
dernier mois



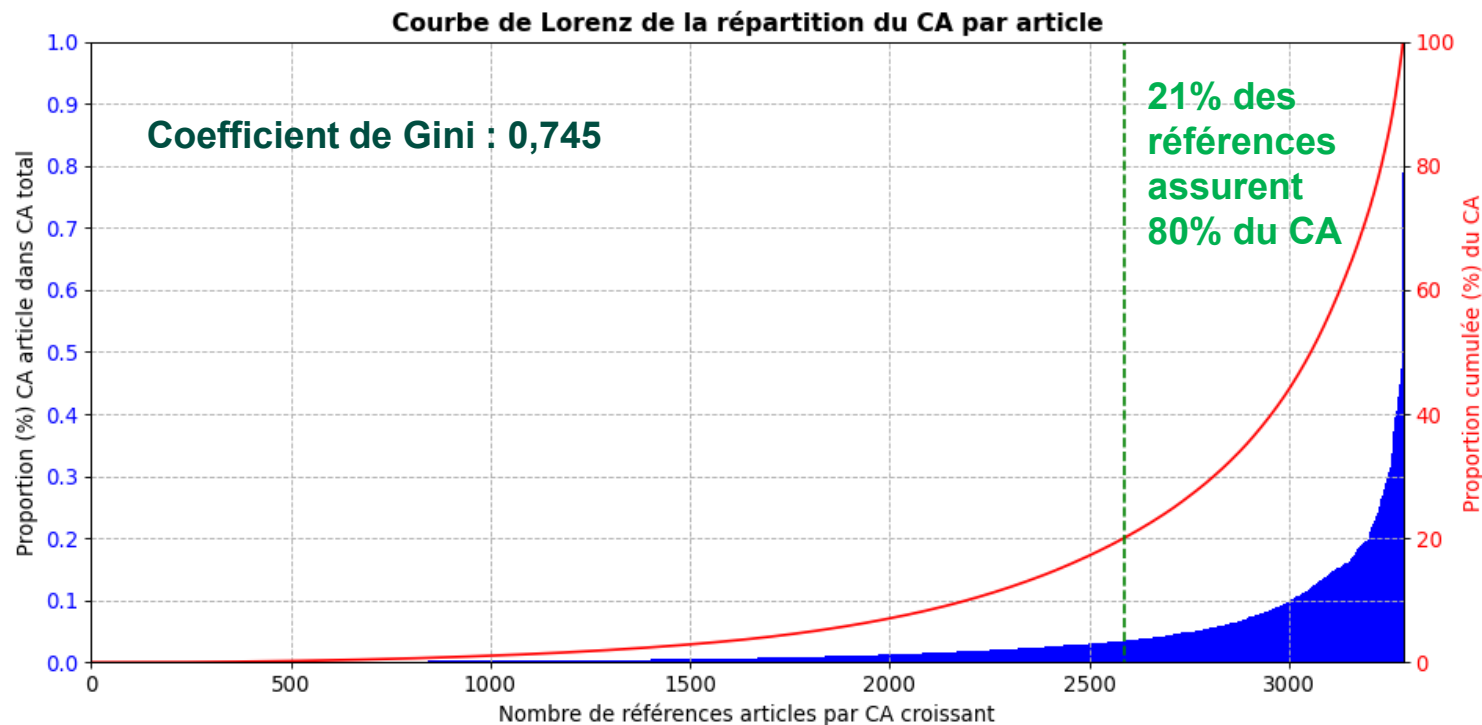
# Top 20 des produits en CA sur la période



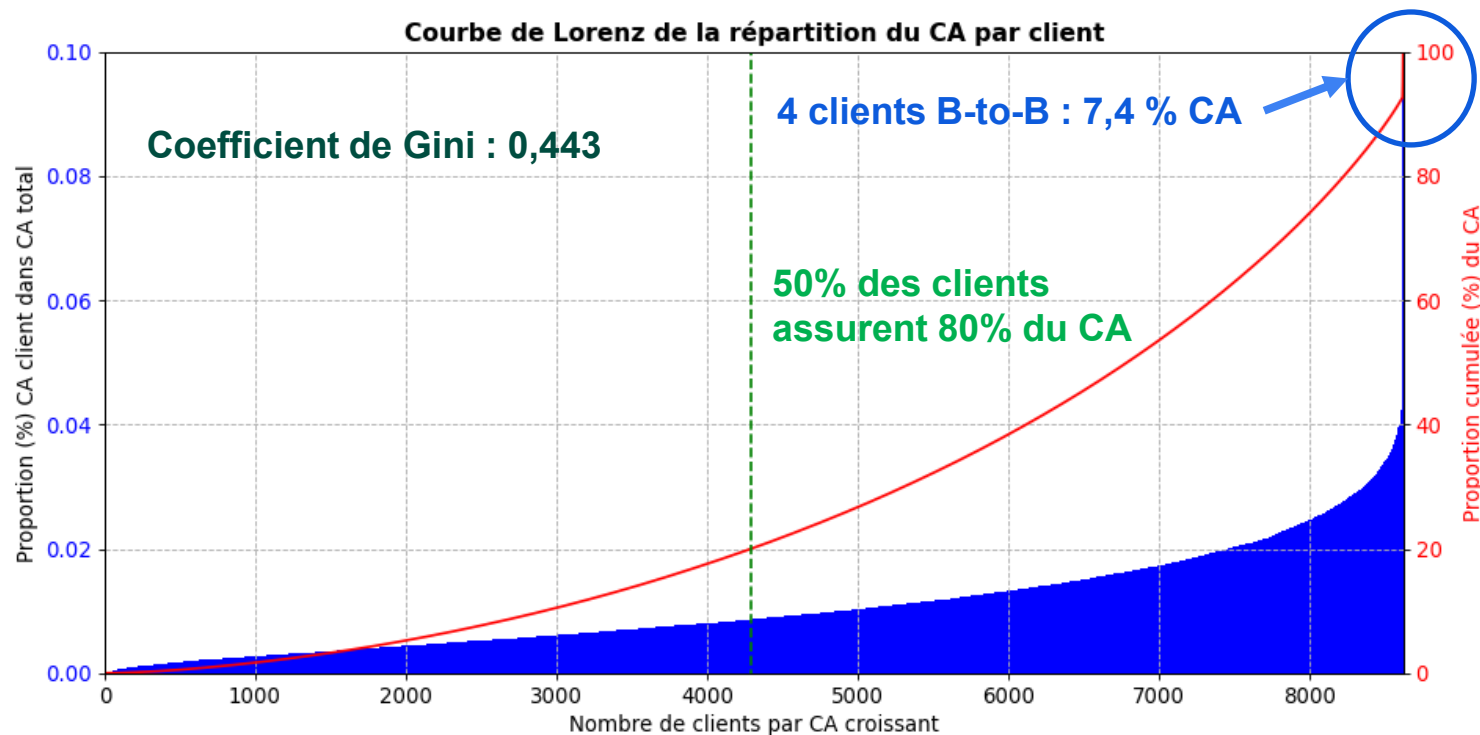
# Best-sellers sur la période



# Courbe de Lorenz du CA par article

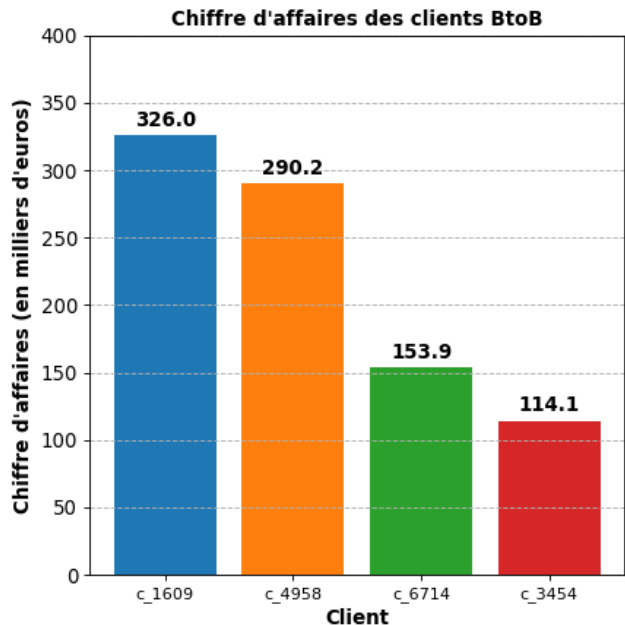


# Courbe de Lorenz du CA par client

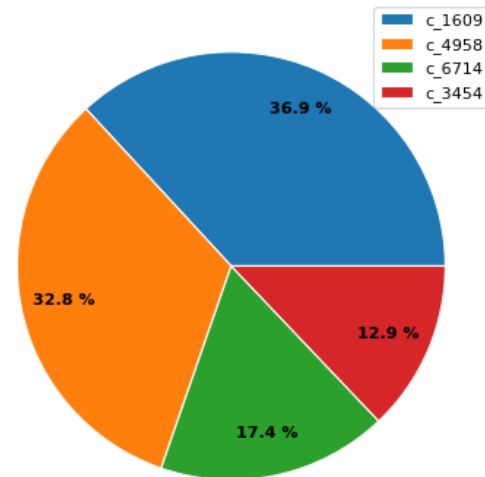


# Zoom sur nos clients B-to-B

**4 clients** à eux-seuls (sur 8621) totalisent **884 k€ de CA**, soit **7,4 % du CA** total sur toute la période



Répartition du CA (en %) des clients BtoB





# Comportements clients

## Étude des corrélations

# Corrélation sexe du client et catégories

---

**Le sexe du client influe-t-il sur la catégorie de produits achetés ?**

Sexe : variable qualitative, 2 valeurs possibles

Catégorie de produits : variable qualitative, 3 valeurs possibles

> 640 k observations indépendantes (transactions), 6 résultats possibles

Effectif théorique d'au moins 5 pour chaque résultat

**Test statistique du  $\chi^2$  (ddl = 2)** - Hypothèse nulle testée :

**$H_0$  : les 2 variables sont indépendantes**

# Corrélation sexe du client et catégories

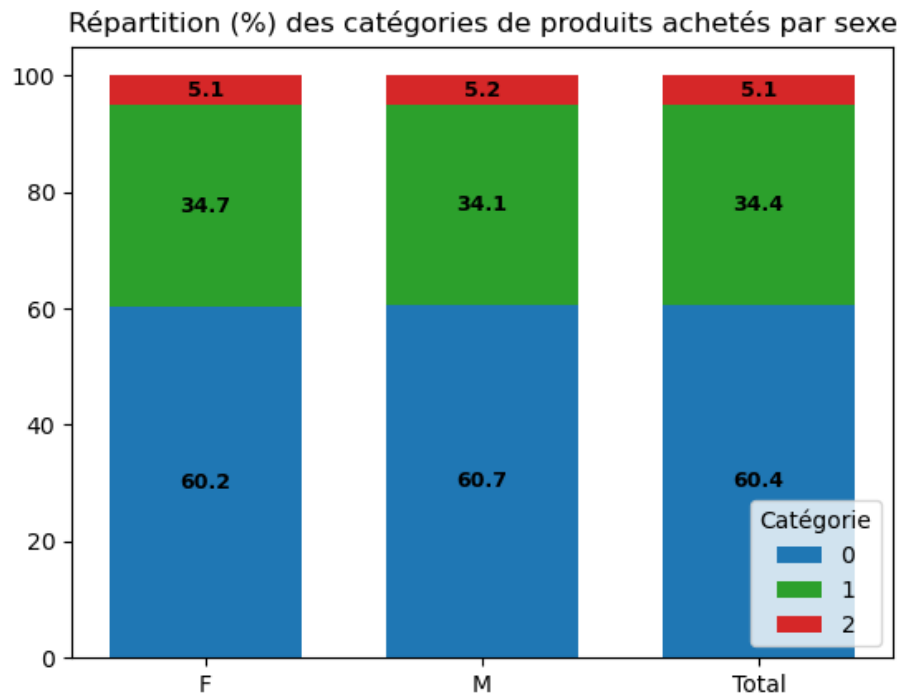
Statistique  $\chi^2 = 22,67$

$p = 1,2 * 10^{-5} < 5\%$

**$H_0$  est rejetée**

L'impact statistique existe,  
mais il est faible :

- M achètent plus de produits de **catégorie 0**
- F achètent plus de produits de **catégorie 1**



# Corrélation âge client et catégorie de produits

---

**L'âge du client influe-t-il sur la catégorie de produits achetés ?**

Âge : variable quantitative

Catégorie de produits : variable qualitative, 3 valeurs possibles

> 640 k observations indépendantes (transactions). 2 tests possibles :

- **Test ANOVA** : suppose que les âges suivent une loi normale pour chaque catégorie, et que ces lois ont une variance égale  
 **$H_0$  : les moyennes des échantillons (par catégorie) sont égales**
- **Test H de Kruskal-Wallis** : version non-paramétrique (rangs)  
 **$H_0$  : les médianes des échantillons (par catégorie) sont égales**

# Corrélation âge client et catégorie de produits

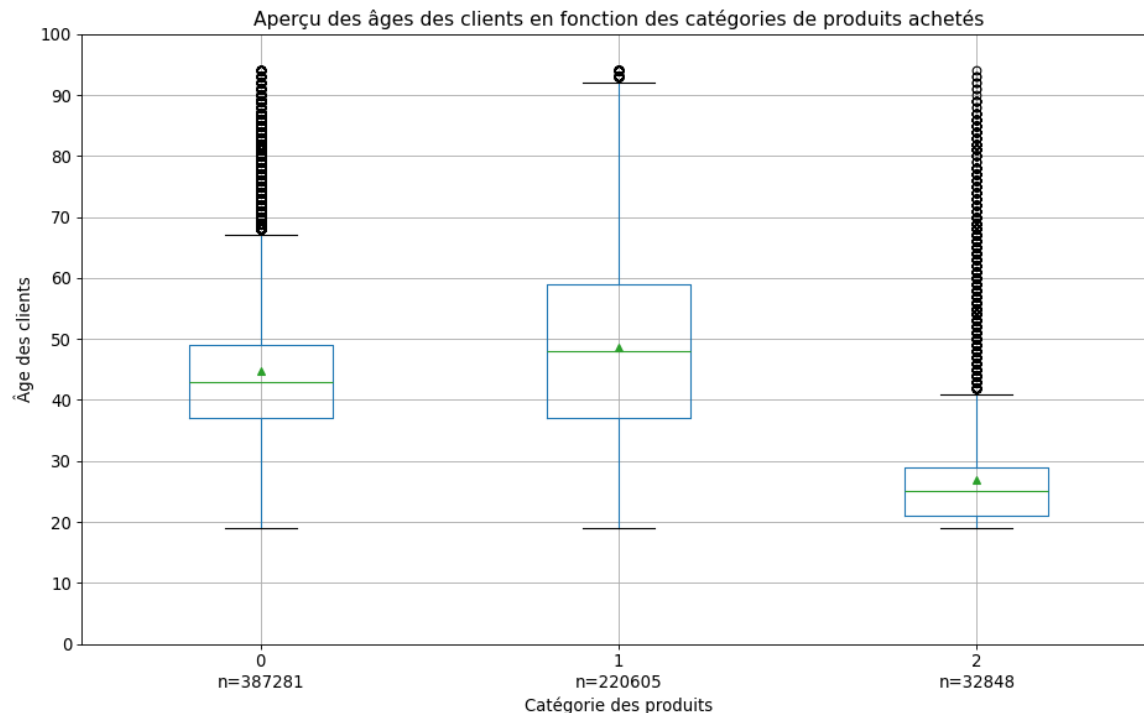
Test ANOVA pas applicable

Stat  $H = 71360$

$p = 0 < 5\%$

$H_0$  est rejetée

Les médianes sont différentes

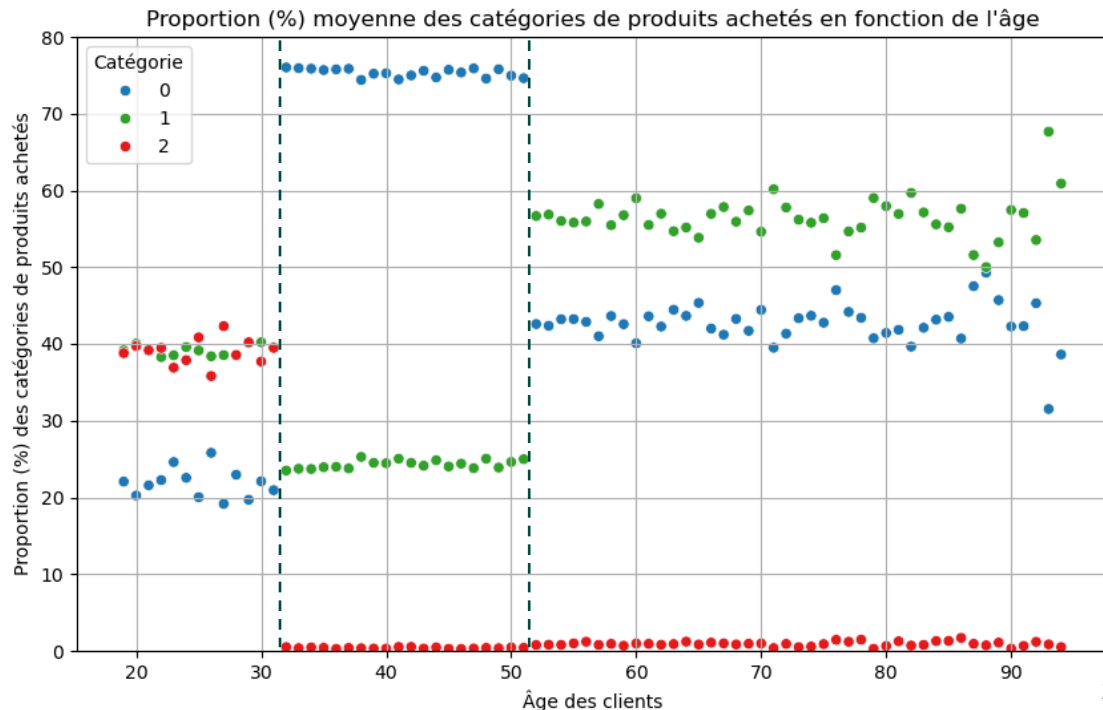


# Corrélation âge client et catégorie de produits

Agrégation  
(moyenne) des  
résultats par âge

3 tranches d'âges :

- 19-31 ans
- 32-51 ans
- 52 ans et plus



# Corrélation âge client et catégorie de produits

---

**L'âge du client influe-t-il sur la catégorie de produits achetés ?**

Âge → Tranche d'âges : variable qualitative, 3 valeurs possibles

Catégorie de produits : variable qualitative, 3 valeurs possibles

> 640 k observations indépendantes (transactions). 9 résultats possibles

Effectif théorique d'au moins 5 pour chaque résultat

**Test statistique du  $\chi^2$  (ddl = 4)** - Hypothèse nulle testée :

***$H_0$  : les 2 variables sont indépendantes***

# Corrélation âge client et catégorie de produits

Statistique  $\chi^2 = 272891$

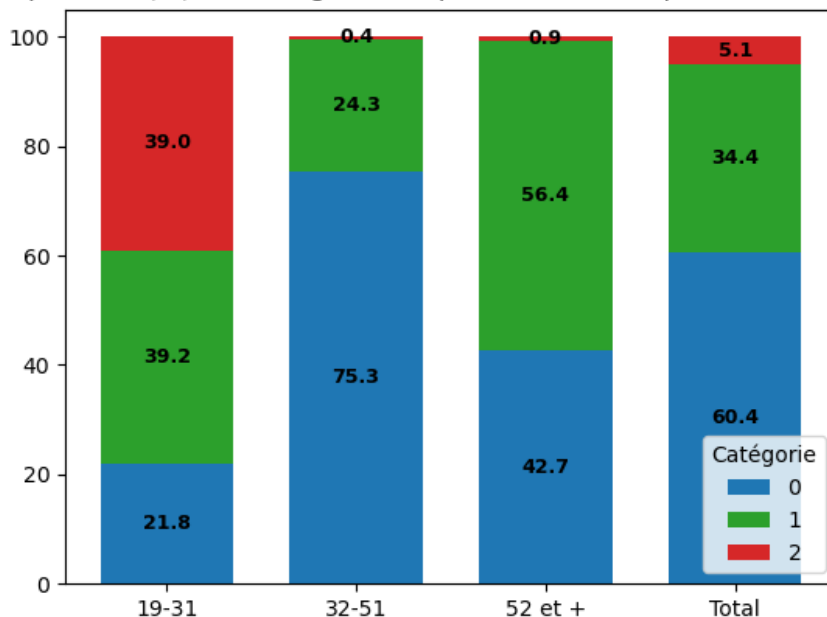
$p = 0 < 5\%$

**$H_0$  est rejetée**

L'impact statistique existe,  
et il est significatif

Chaque tranche d'âge a  
un comportement  
homogène bien distinct

Répartition (%) des catégories de produits achetées par tranche d'âges





# Corrélation âge client et autre variable quantitative

---

***L'âge du client influe-t-il sur l'autre variable quantitative ?***

Âge : variable quantitative

Y : autre variable quantitative

> 8500 observations indépendantes (clients)

Les données ne suivent pas nécessairement une loi normale

**Test de corrélation de Spearman** - Hypothèse nulle testée :

***$H_0$  : âge et Y ne sont pas corrélées***

# Corrélation âge client et montant total achats

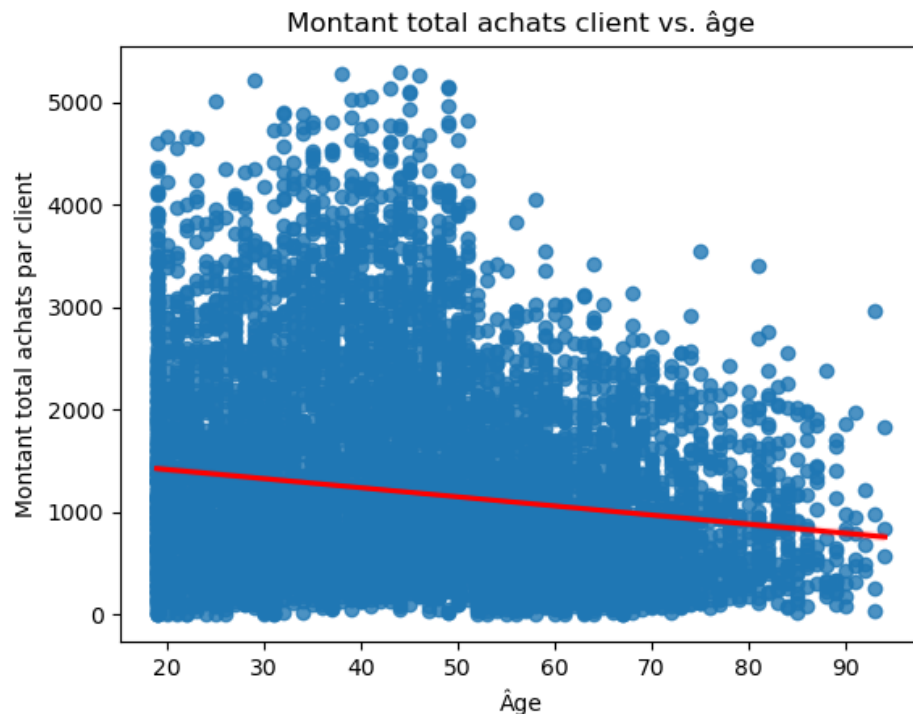
$$\rho = -0,182$$

$$p = 6 * 10^{-65} < 5\%$$

**$H_0$  est rejetée**

Il y a une corrélation  
entre ces 2 variables

Forte dispersion des  
observations



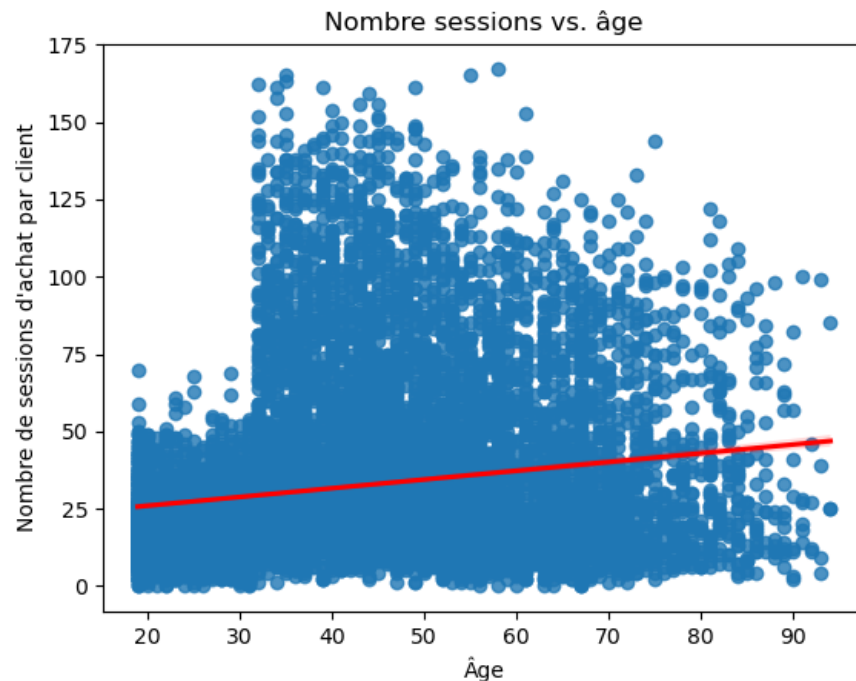
# Corrélation âge client et nombre de sessions

$$\rho = 0,212$$
$$p = 1,7 * 10^{-88} < 5\%$$

**$H_0$  est rejetée**

Il y a une corrélation  
entre ces 2 variables

Forte dispersion des  
observations



# Corrélation âge client et taille panier moyen

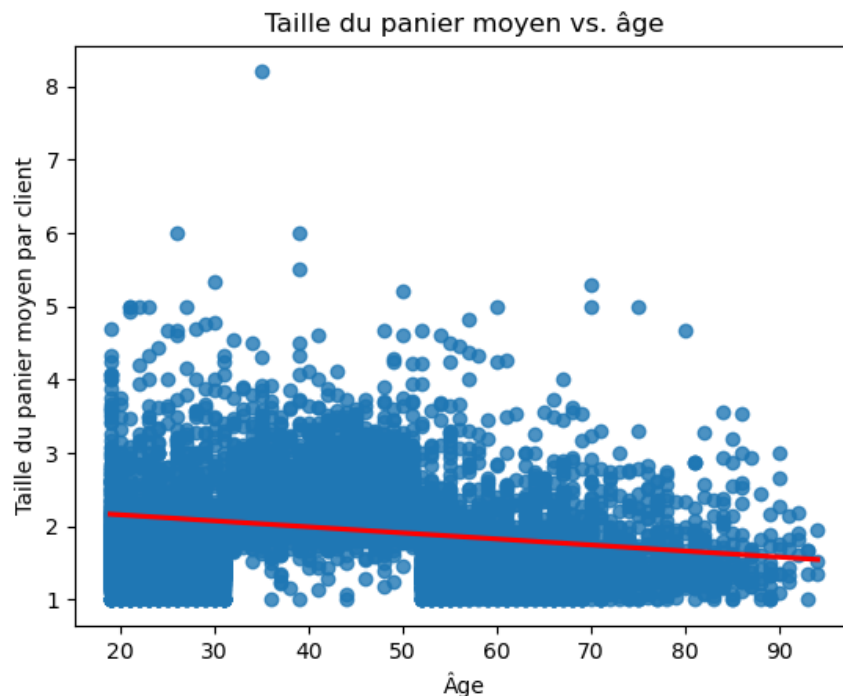
$$\rho = -0,208$$

$$p = 2 * 10^{-84} < 5\%$$

**$H_0$  est rejetée**

Il y a une corrélation  
entre ces 2 variables

Forte dispersion des  
observations



# Corrélation âge client et autre variable quantitative

---

***L'âge du client influe-t-il sur l'autre variable quantitative ?***

Âge → Tranche d'âges : variable qualitative, 3 valeurs possibles

Les tranches d'âges sont : 19-31 ans, 32-51 ans, 52 ans et plus

Y : autre variable quantitative

> 8500 observations indépendantes (clients)

Y Ne suit pas nécessairement une loi normale

**Test H de Kruskal-Wallis** : version non-paramétrique du test ANOVA

***$H_0$  : les médianes des échantillons (par tranche) sont égales***

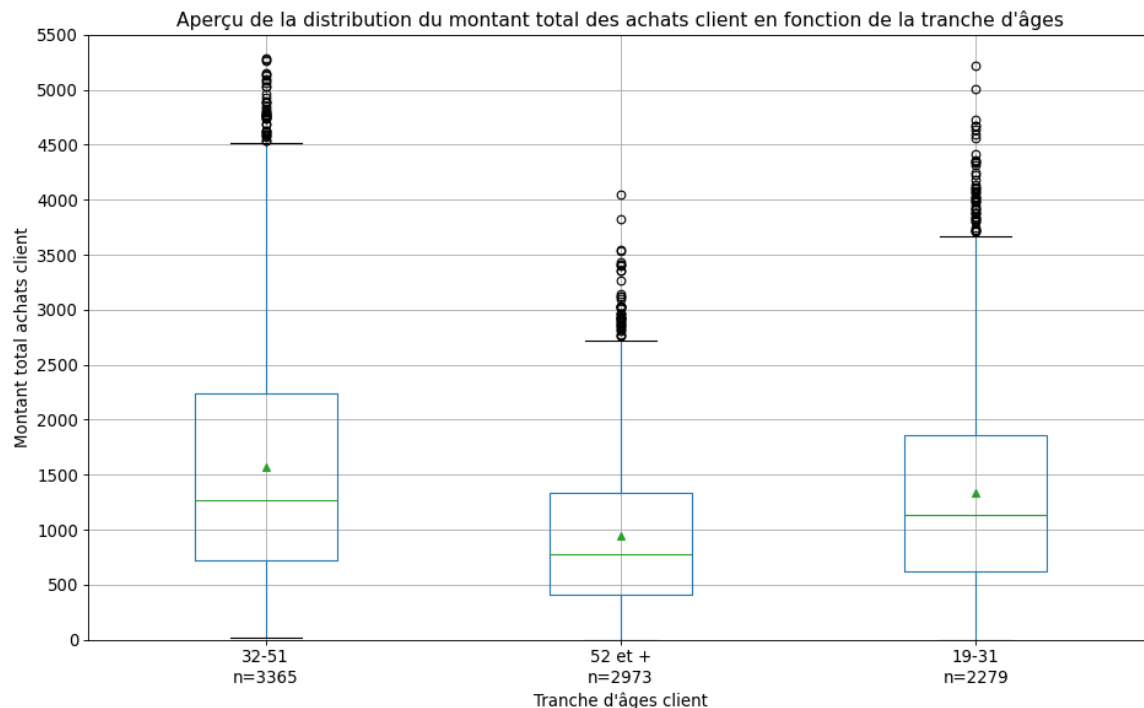
# Corrélation âge client et montant total achats

$$H = 615$$

$$p = 0 < 5\%$$

**$H_0$  est rejetée**

La tranche d'âges  
a un impact  
statistique sur le  
montant total des  
achats



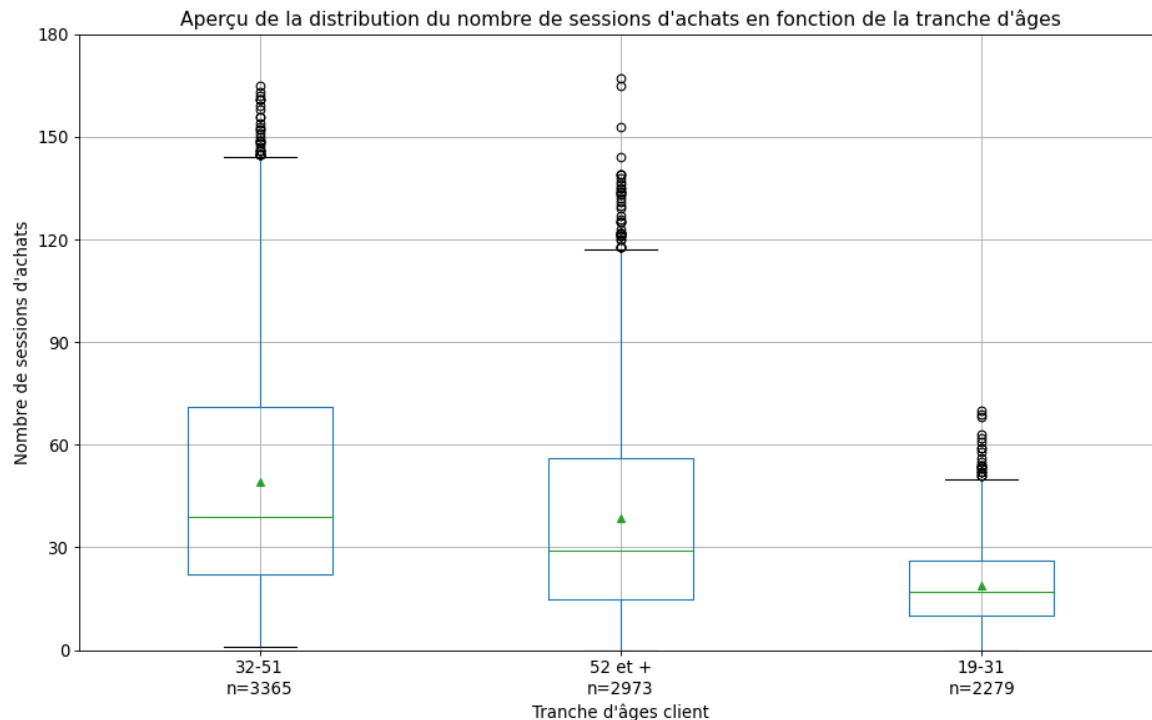
# Corrélation âge client et nombre de sessions

$$H = 1435$$

$$p = 0 < 5\%$$

**$H_0$  est rejetée**

La tranche d'âges  
a un impact  
statistique sur le  
nombre de  
sessions d'achat



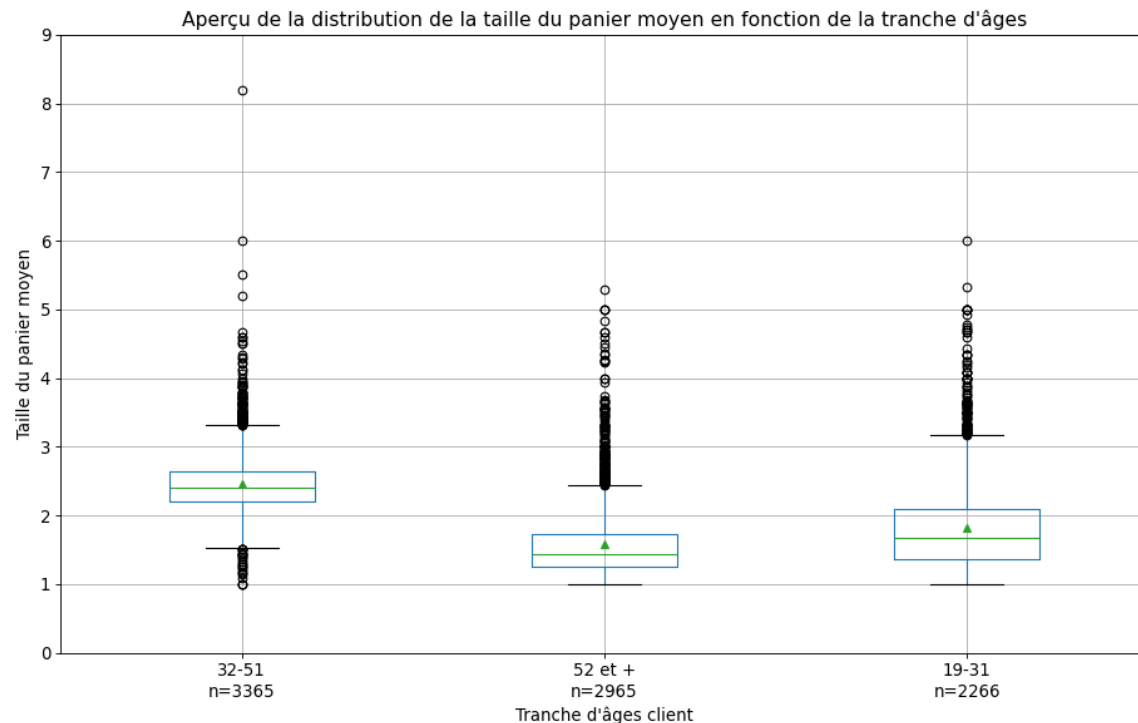
# Corrélation âge client et taille panier moyen

$$H = 3746$$

$$p = 0 < 5\%$$

**$H_0$  est rejetée**

La tranche d'âges  
a un impact  
statistique sur la  
taille du panier  
moyen





# Principales observations

---

Plusieurs indicateurs clés sont en baisse sur le dernier mois  
→ Être vigilant - cause conjoncturelle ?

Même si les 4 clients B-to-B (7,4 % du CA) requièrent un suivi spécifique, la **structure du CA vis-à-vis des produits est plus inégalitaire que vis-à-vis des clients**

Comportements individuels des clients difficiles à cerner, mais **3 tranches d'âges avec des pratiques homogènes émergent**

# Préconisations

---

- Suivre attentivement les produits qui se vendent le mieux, limiter les achats (en quantité) des produits moins prisés
- Mettre en avant les produits de **catégorie 2** auprès des **19-31** qui font peu de sessions avec des paniers de petite taille
- Entretenir la fidélité des **32-51** qui assurent une bonne part de notre part de CA avec les produits de **catégorie 0**
- Promouvoir les produits de **catégorie 1** qui se vendent le mieux auprès des **52 et +**

Merci pour votre attention

Avez-vous des questions ?