

# ANALYSE DES VENTES « LAPAGE »

---

KHADIJA MORCHID ALAOUI

---

# LAPAGE EN QUELQUES MOTS :

---

- Spécialiste de la vente de livres
- Plusieurs points de vente
- Site de vente en ligne.
- Souhaite faire le point et analyser ses points forts, ses points faibles, les comportements clients, etc...

# LAPAGE EN QUELQUES CHIFFRES SUR L'ANNÉE FISCALE 2021 :

---

- Plus que 4 millions d'euros
- Plus que 200 milles ventes

# DONNÉES NÉCESSAIRES À L'ANALYSE DES VENTES :

---

- Liste des transactions
- Liste des clients
- Liste des produits

# MÉTHODOLOGIE D'ANALYSE ET NETTOYAGE DES DONNÉES :

---

- Valeurs manquantes : pas de valeurs manquantes pour les trois fichiers.
- Valeurs doublons : aucune valeur en double( chaque produit/client est représenté par un identifiant unique)
- Valeurs aberrantes : suppression de 21 enregistrements (clients non acheteurs/produits non vendus),suppression de 4 clients pros
- Date : suppressions des lignes dont les dates de transactions n'est pas au bon format puis conversion des valeurs dans un format date
- Autres modification : ajout de colonne( âge, année, mois, jour), cohérence des données (des âges, prix, genre..)



# QUESTIONS PRINCIPALEMENT TRAITÉES AU COURS DE L'ANALYSE :

---

- analyse des différents indicateurs de vente : ca total, évolution du ca dans le temps, son évolution par catégorie ..
- Zoom sur les produits : tops, flops, répartition ca par produits
- Zoom sur les clients : meilleurs clients, répartition du ca par clients
- Analyse des liens possible entre les caractéristiques client/ produit

# 1- DIFFÉRENTS INDICATEURS ET GRAPHIQUES AUTOUR DU CA :

le chiffre d'affaires total réalisé depuis l'ouverture du site de notre librairie est de 12022673.6 € pour un total de 687298 transactions.

Chiffres d'affaires annuels:

year

2021 4939771.43

2022 6108681.81

2023 974220.31

Name: price, dtype: float64

Nombre de ventes annuelles:

year

2021 286435

2022 346380

2023 54483

Le chiffre d'affaires moyen par année :

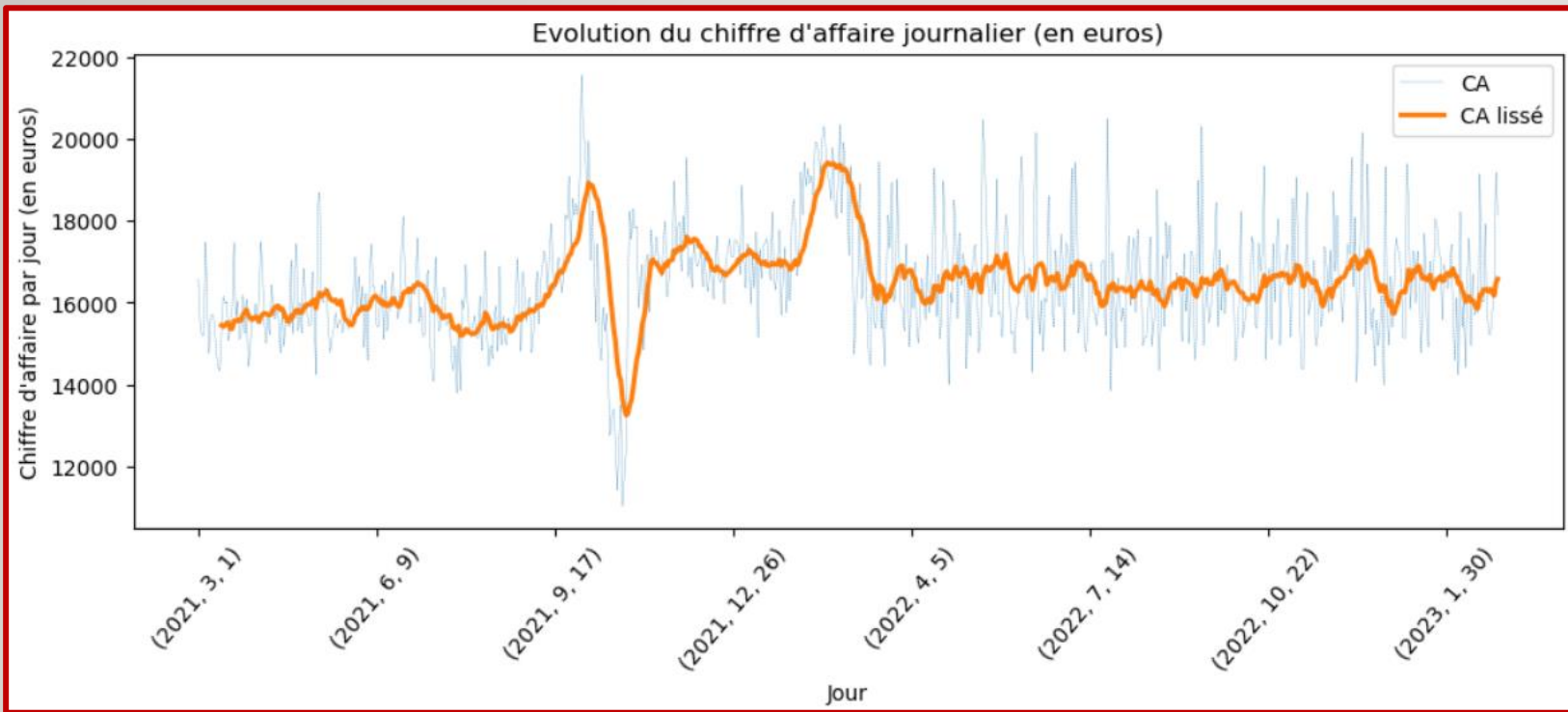
year

2021 17.245698

2022 17.635781

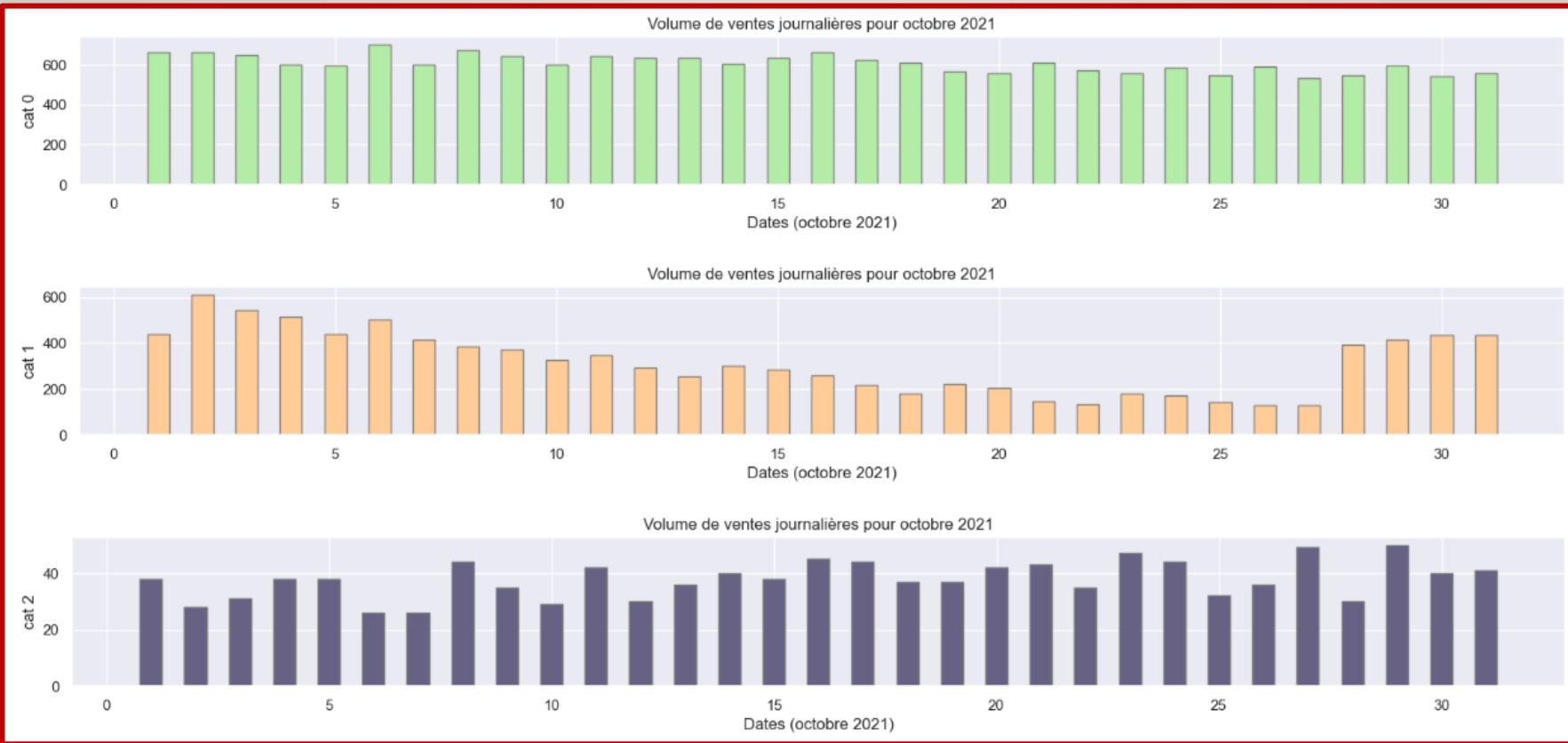
2023 17.881180

- 2022 est la meilleure année
- Baisse du ca en 2023 (les données ne sont pas complètes)
- Le ca moyen est le plus élevé en 2023



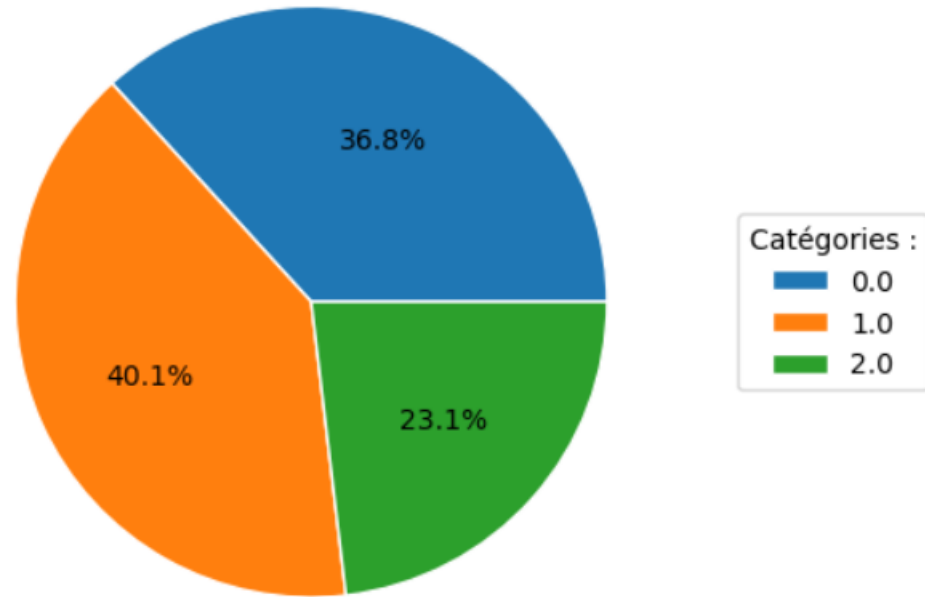
- Une évolution stable en début 2021 Et 2022
- Une chute du CA durant le mois d'octobre 2021
- Nécessaire investigation sur le mois d'octobre



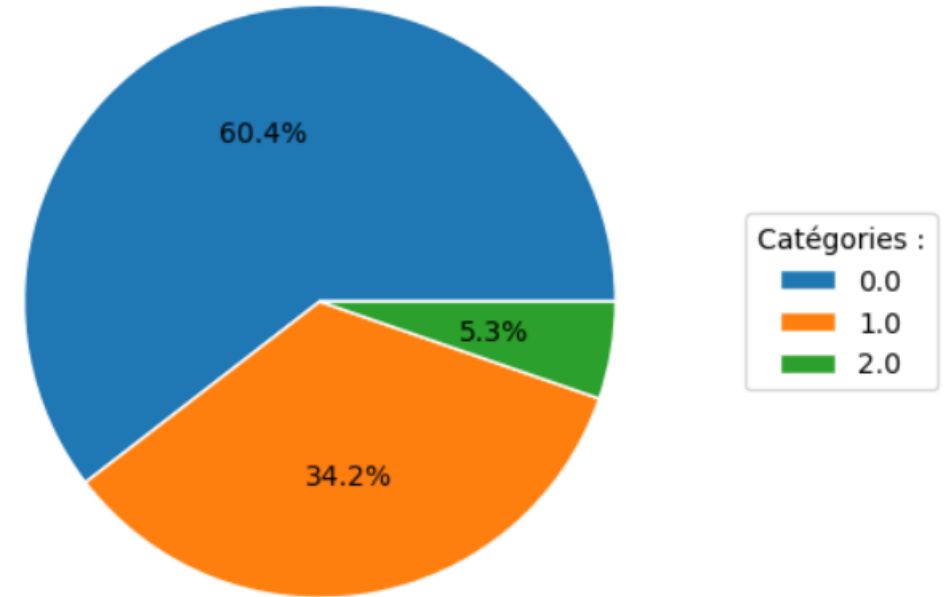


- Les ventes des deux categ 0 et 2 sont stables
- Une baisse de nombre de ventes pour la catégorie 1 en octobre 2021

Répartition des ventes par catégories en chiffre d'affaires

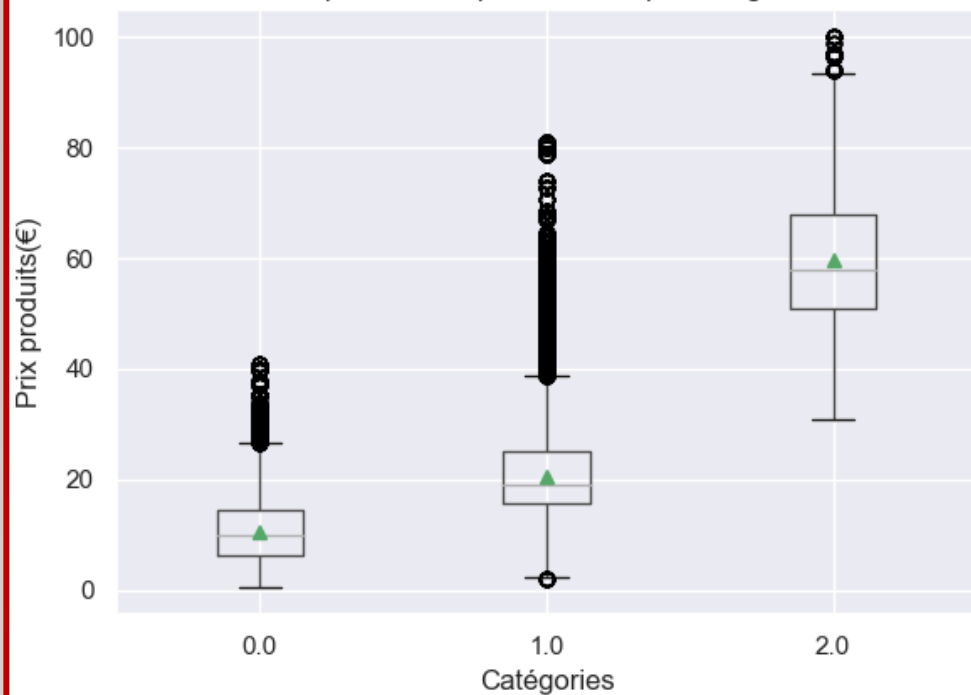


Répartition des ventes par catégories en quantité

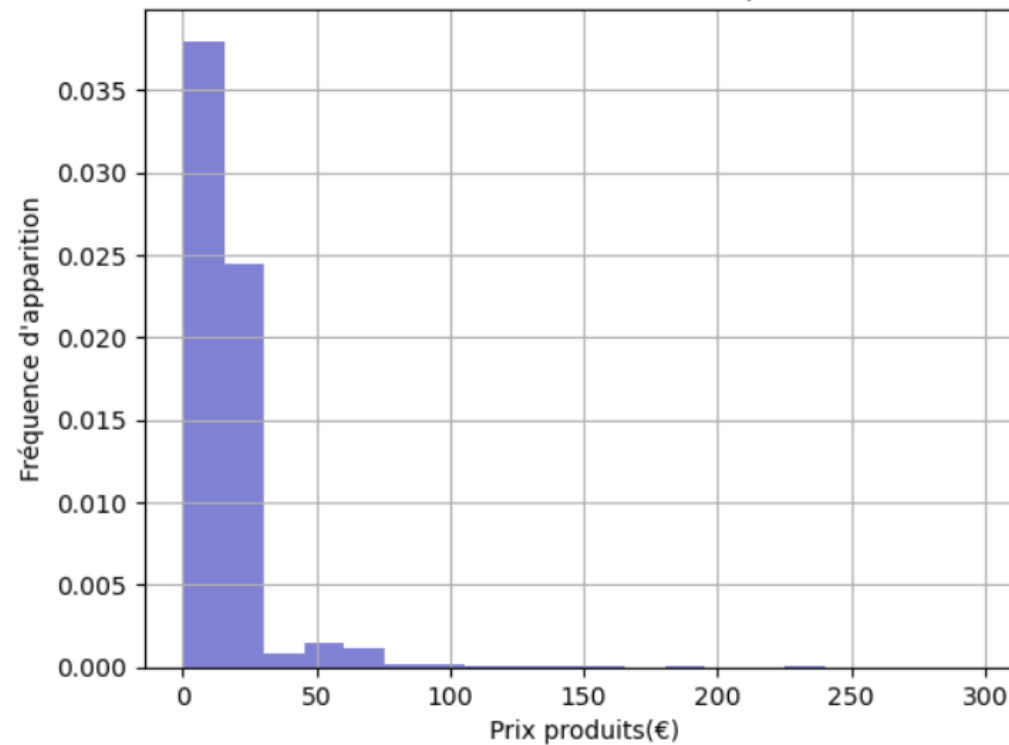


- catégories "0" et "1" génèrent un chiffre d'affaires comparable
- la catégorie "2" génère le moins du chiffre d'affaires
- Plus de 60% des ventes se font sur les produits de la catégorie 0, mais ceux de la catégorie 2 représentent seulement 5% des ventes. Pourquoi ??

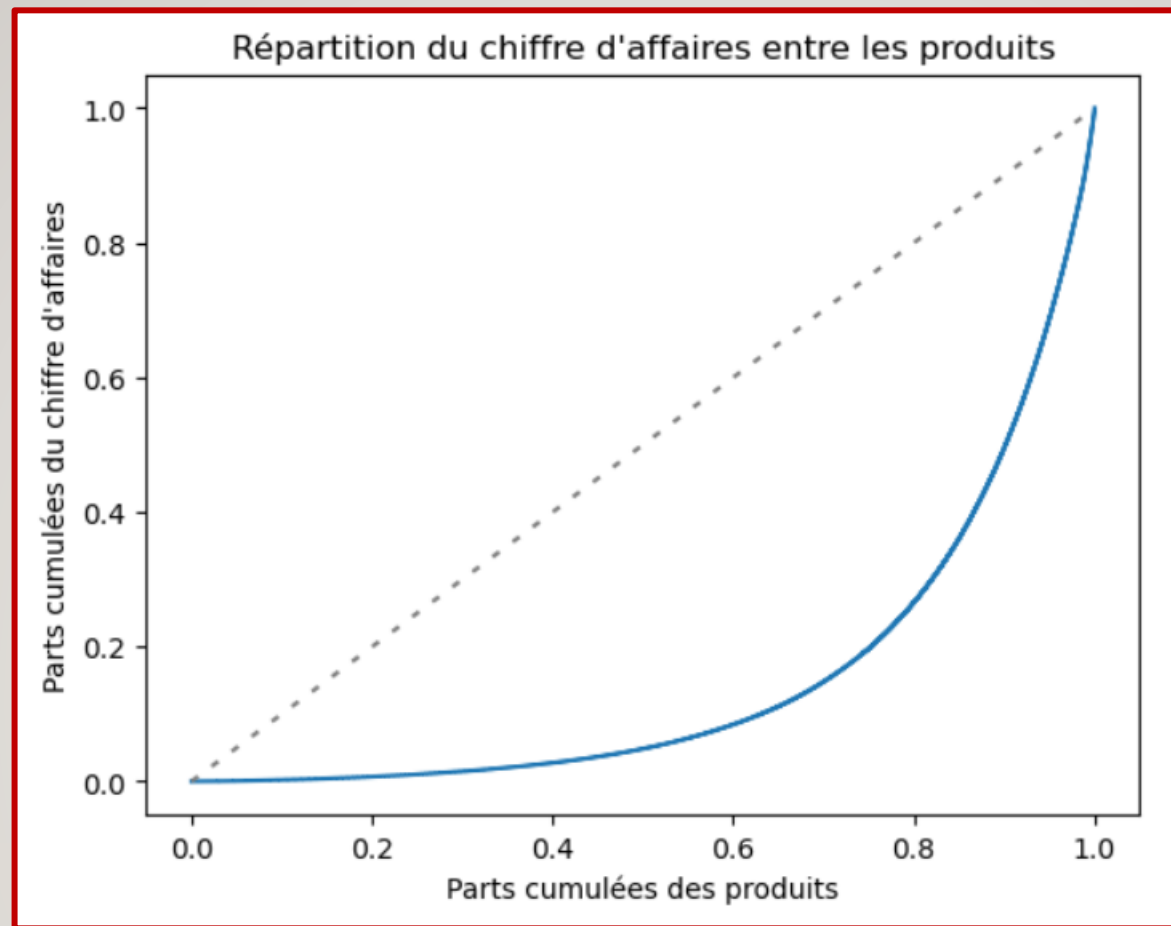
Boxplot grouped by categ  
Dispersion des prix de vente par catégorie



Distribution de la variable "price"



- Prix plus élevé de la categ 2
- Peu de dispersion des prix
- Les produits <30 euros sont les plus populaires.



- 80% des produits ne détient que 30% du CA.
- L'indice de Gini (0,69) s'éloigne de 0, le poids des produits dans le chiffre d'affaires global est donc très inégal.

Références génèrent le plus de CA :

id\_prod

2\_159 94893.50

2\_135 69334.95

2\_112 65407.76

2\_102 60736.78

2\_209 56971.86

Name: price, dtype: float64

Références génèrent le moins de CA :

id\_prod

0\_1539 0.99

0\_1284 1.38

0\_1653 1.98

0\_1601 1.99

0\_541 1.99

Name: price, dtype: float64

Références les plus vendues :

id\_prod

1\_369 2338

1\_417 2267

1\_414 2245

1\_498 2200

1\_425 2162

Name: session\_id, dtype: int64

Références les moins vendues :

id\_prod

0\_1151 1

0\_1284 1

0\_1379 1

0\_1498 1

0\_1539 1

Name: session\_id, dtype: int64

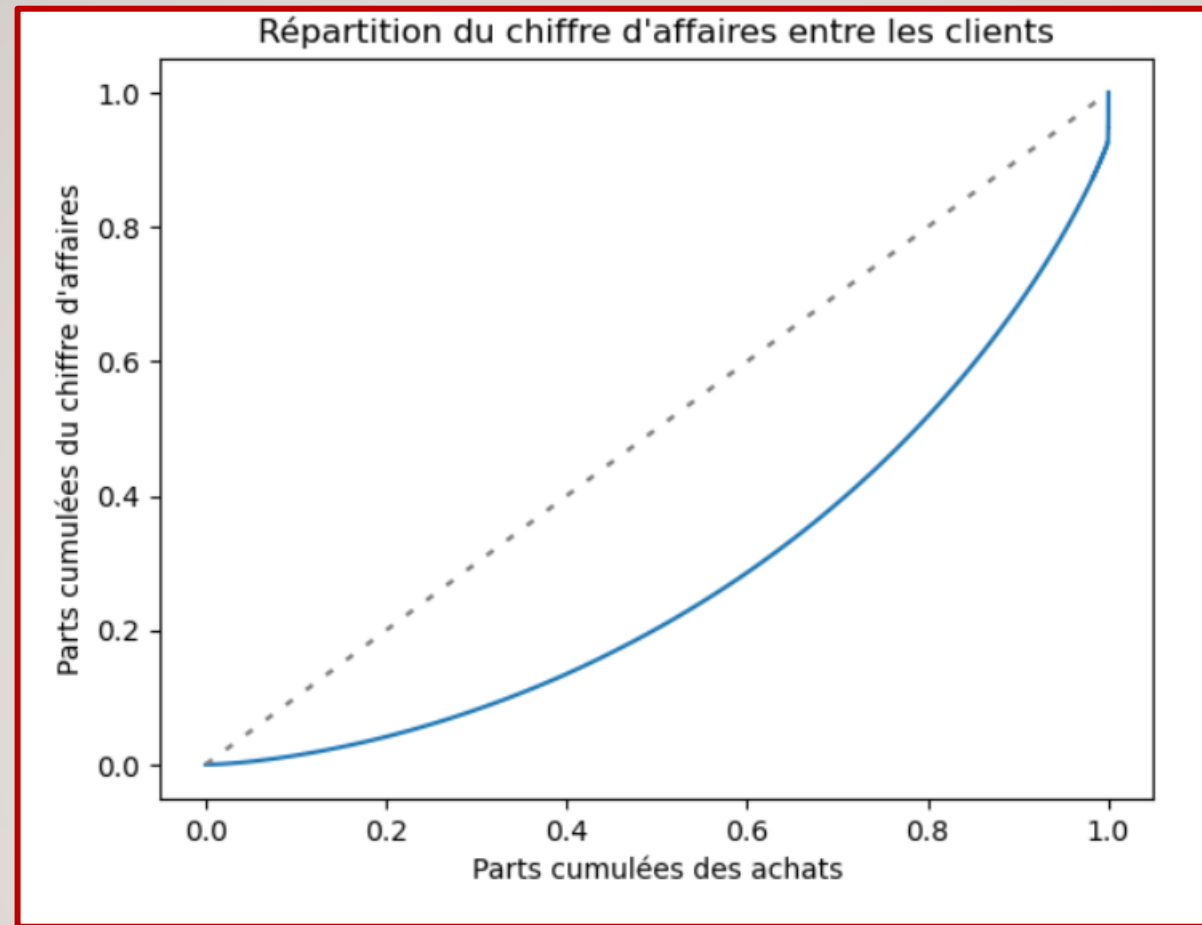
➤ Les produits vendus en grande quantités ne sont donc pas forcément les plus chers.



Clients ayant le plus dépensé  
client\_id

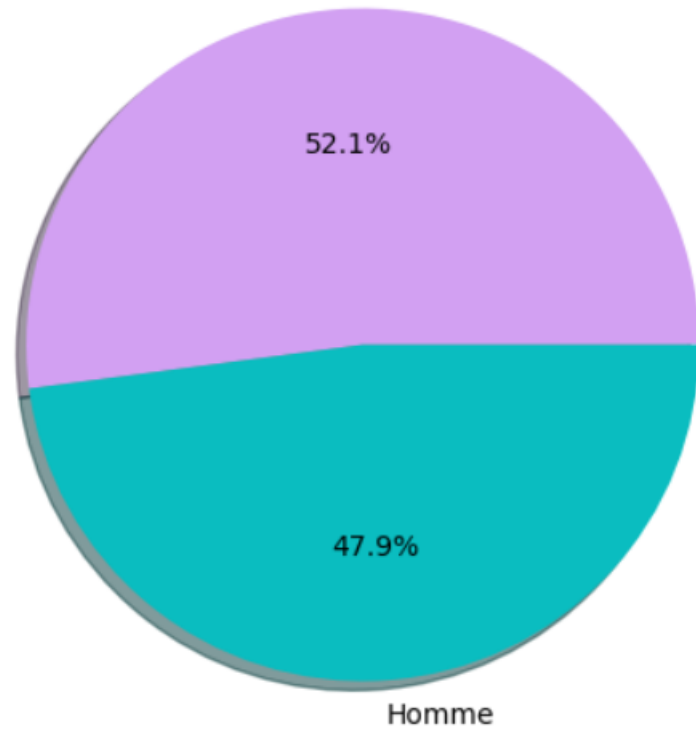
c_1609	326039.89
c_4958	290227.03
c_6714	153918.60
c_3454	114110.57
c_1570	5285.82
c_3263	5276.87
c_2140	5260.18
c_2899	5214.05
c_7319	5155.77
c_7959	5135.75

- Les quatre premiers clients, ils peut s'agir des revendeurs (libraires, bibliothécaires)



- la répartition du CA par client est relativement équitable.
- Indice de Gini ( $0.44 < 0.5$ ) qui témoigne d'une répartition du CA plutôt homogène sans être strictement égalitaire.

Répartition des genres au sein de notre clientèle



- Homme/Femme sont équitablement représentés au sein de la clientèle
- L'entreprises parvient à attirer et à fidéliser une **clientèle diversifiée**

# Corrélations possibles

---

**Catégorie ?**

**PRODUITS ?**

?

**GENRE ?**

?

**Age ?**

# CORRÉLATION ENTRE LE GENRE ET LA CATÉGORIE DES PRODUITS ACHETÉS

---

## Les hypothèses du test chi2 :

- H0 : l'hypothèse nulle d'indépendance des variables
- H1 : l'hypothèse alternative de dépendances des variables
- Pour un risque de 5% :



Les fréquences théoriques sont:

```
[[207073.63745566 117306.4562737 18183.90627064]  
 [208385.36254434 118049.5437263 18299.09372936]]
```

Ce qui donne un  $X^2$  de 157.590 à 2 ddl

soit une p-value de 6.023430519385375e-35

Le V de Cramer vaut: 0.015142280789843969

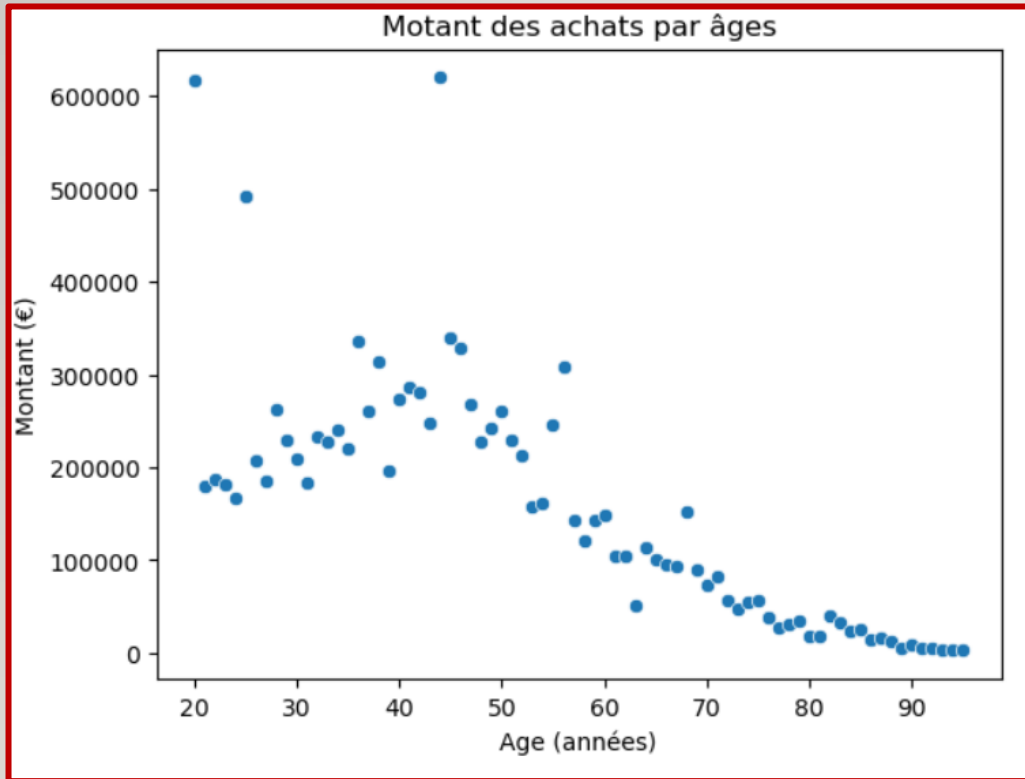


- Une p\_value aussi faible(<0.05)
- Le test chi2 confirme la corrélation
- Le coefficient de Cramer proche de zéro indique une faible association



# CORRÉLATION ENTRE ÂGE CLIENT ET MONTANT TOTAL DES ACHATS

---



- Tendence de baisse du montant des achats avec l'âge
- Présence des outliers

# CORRÉLATION ENTRE ÂGE CLIENT ET MONTANT TOTAL DES ACHATS

---

## Les hypothèses du test Spearman :

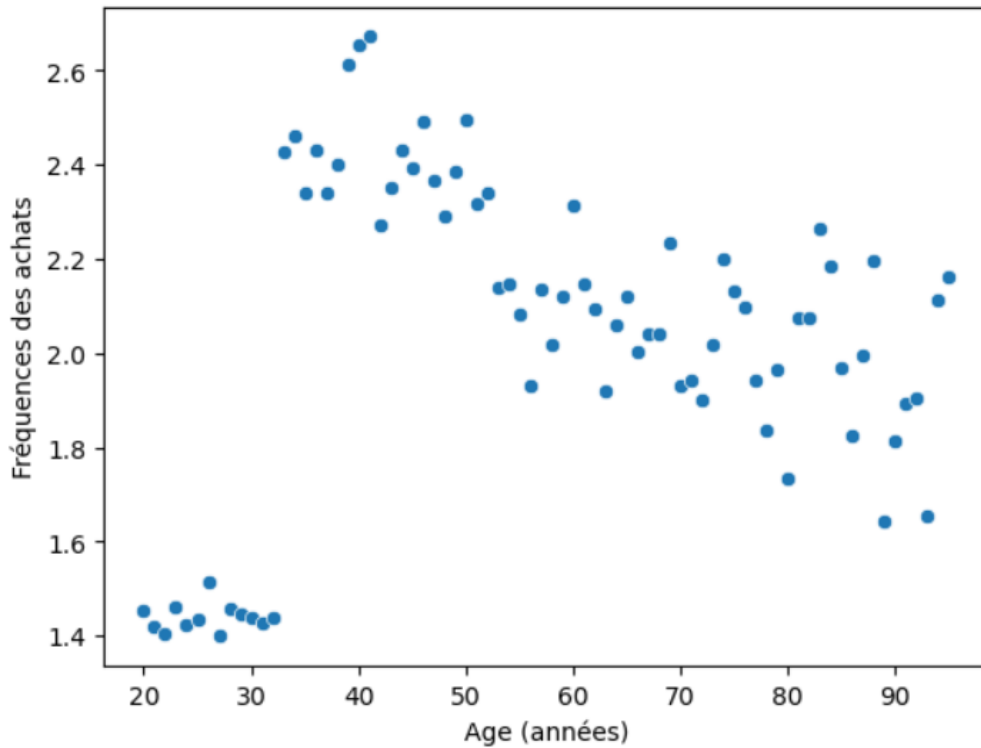
- H0 : Il n'y a pas de corrélation entre les deux variables
- H1 : Il y a une corrélation entre les deux variables.



- Le coefficient de corrélation est négatif (-0,85)
- Le test Spearman confirme une forte corrélation négative
- Plus la tranche d'âge augmente plus le chiffre d'affaires baisse

# CORRÉLATION ENTRE ÂGE CLIENT ET LA FRÉQUENCE D'ACHAT

Fréquence mensuelle d'achat par âge



- Une tendance à la baisse du fréquence d'achats avec l'âge
- Des clients dont l'âge entre 20 et 30 ans qui se démarquent du reste des clients (un nombre d'achats faible entre 1.4 et 1.5)

# CORRÉLATION ENTRE ÂGE CLIENT ET LA FRÉQUENCE D' ACHAT

---

## Les hypothèses du test de Spearman :

- H0 : les variables sont indépendantes et qu'il n'y a pas de corrélation entre elles.
- H1 : les deux variables ne sont pas indépendantes et qu'il existe une corrélation entre elles.

statistic=0.2041015650722228, pvalue=1.6541857614222247e-81

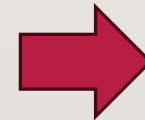
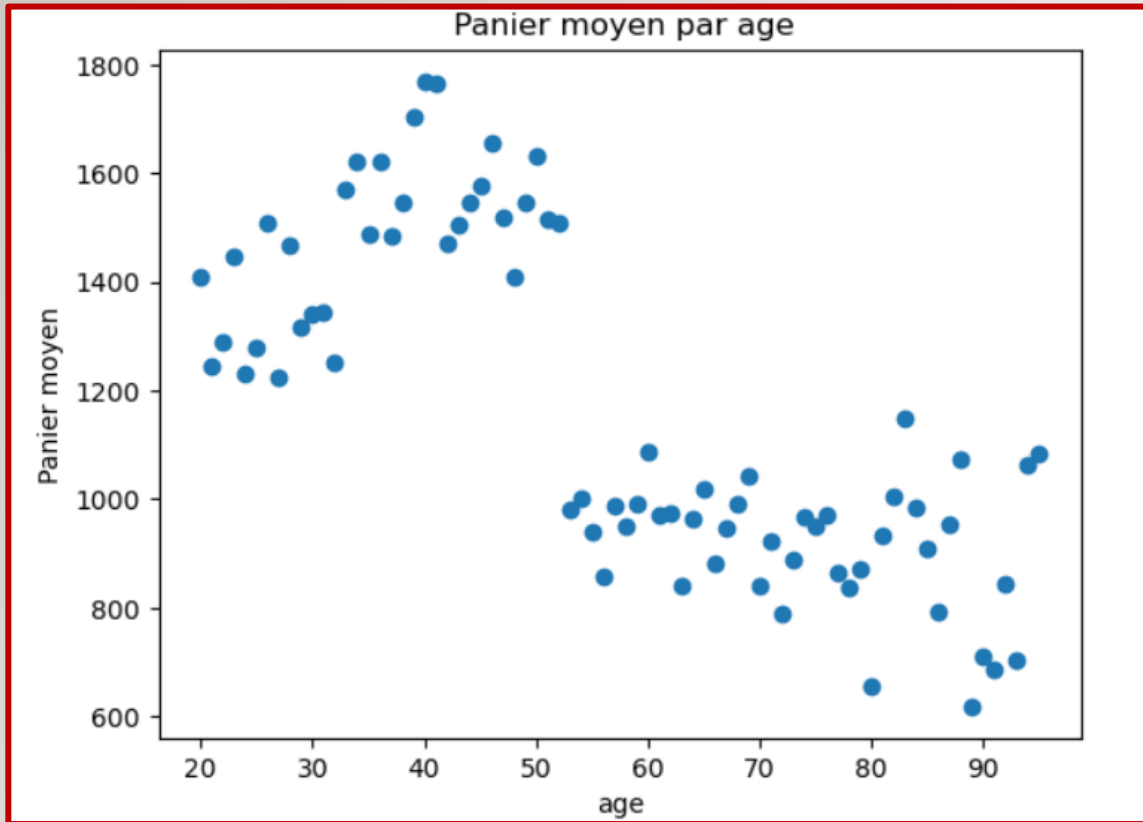


➤ Le résultat de Spearman induit une légère corrélation entre l'âge et la fréquence d'achat.



# CORRÉLATION ENTRE ÂGE CLIENT ET PANIER MOYEN

---



- La distribution des variables ne semble pas suivre une loi normale donc on passe au test de Spearman.



# CORRÉLATION ENTRE ÂGE CLIENT ET PANIER MOYEN

---

## Les hypothèses du test de Spearman :

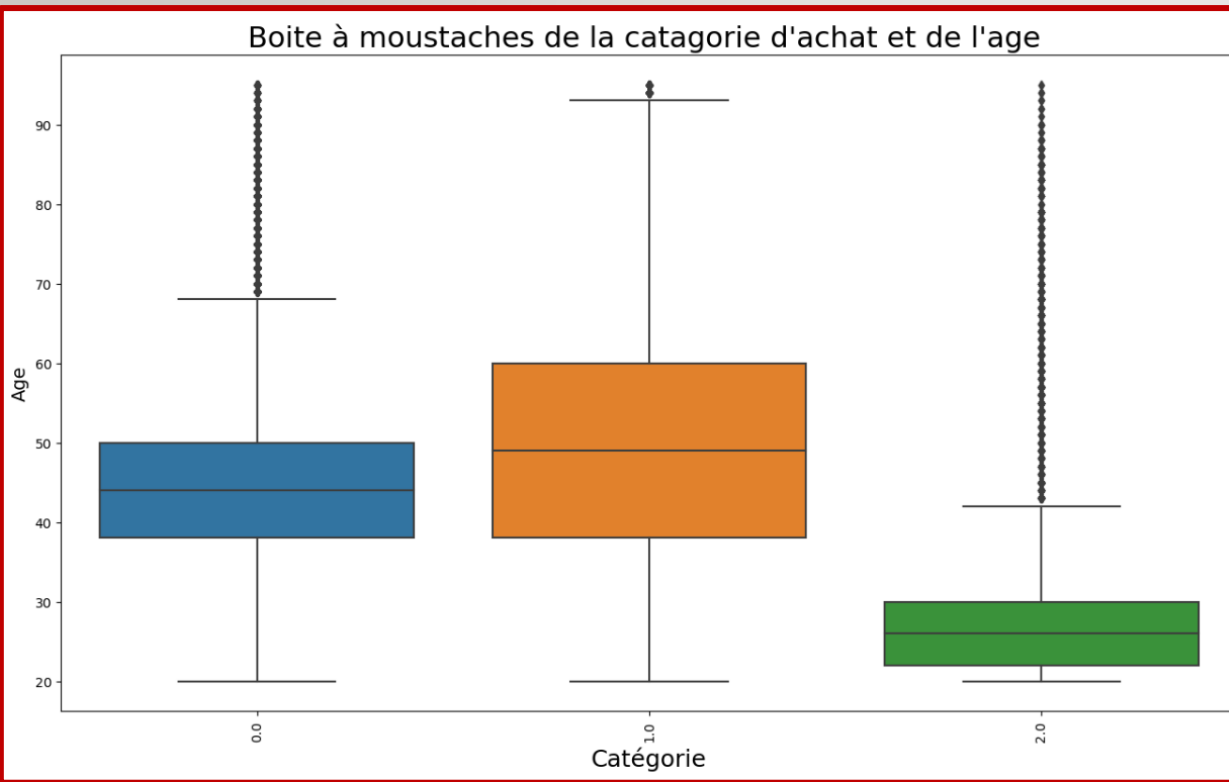
- H0 : les variables sont indépendantes et qu'il n'y a pas de corrélation entre elles.
- H1 : les deux variables ne sont pas indépendantes et qu'il existe une corrélation entre elles.

Corrélation de Spearman: SignificanceResult(statistic=-0.7274367737525631, pvalue=9.891970162274518e-14)



- Coefficient de corrélation négatif -0,72
- Le résultat de Spearman induit une forte corrélation entre l'âge et le panier moyen.

# CORRÉLATION ENTRE ÂGE CLIENT ET LA CATÉGORIE DES PRODUITS



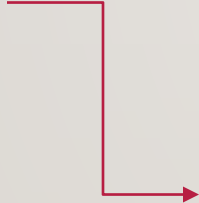
- Les individus qui achètent les produits de la catégorie 0 ont un âge médian de 41 ans
- Les individus qui achètent les produits de la catégorie 1 ont un âge médian de 48 ans
- Les individus qui achètent les produits de la catégorie 2 ont un âge médian de 23 ans

# CORRÉLATION ENTRE ÂGE CLIENT ET LA CATÉGORIE DES PRODUITS

---

## Rappel des conditions de l'ANOVA :

- Indépendance : il n'y a pas de relations cachées entre les observations.
- Normalité : Les données doivent suivre une distribution normale.
- Homogénéité de la variance : les variances des données sont similaires dans chaque groupe.

- 
- Test de Bartlett montre une violation de l'homogénéité des variances
  - Test Shapiro Wilk montre une violation de la normalité

# CORRÉLATION ENTRE ÂGE CLIENT ET LA CATÉGORIE DES PRODUITS

---

## Les hypothèses du test Kruskal Wallis:

- $H_0$  : toutes les catégories sont vendues de manière similaire selon la catégorie d'âge
- $H_1$  : L'une des catégories se vend plus ou moins selon l'âge.



- $p\text{-val} < 0,05$ , on peut ici rejeter aisément  $H_0$
- L'âge influe sur le choix d'acheter une catégorie plutôt qu'une autre.



# AXES D'AMÉLIORATION :

---

## **Perte de compétitivité sur les axes suivants :**

- Catégorie 2 n'enregistre que 5% des ventes totales : étudier les besoins et les préférences des clients dans la catégorie de produits, évaluer la stratégies de tarification et s'assurer que les prix sont compétitifs par rapport aux produits similaires sur le marché, développer des campagnes marketing ciblées pour promouvoir ce produit...
- Clients plus âgés achètent moins que les plus jeunes : offrir une expérience d'achat personnalisée en tenant compte des préférences et des besoins spécifiques des clients plus âgés, proposez des promotions spéciales et des remises pour inciter les clients plus âgés à acheter...



MERCI  
À VOUS

---