



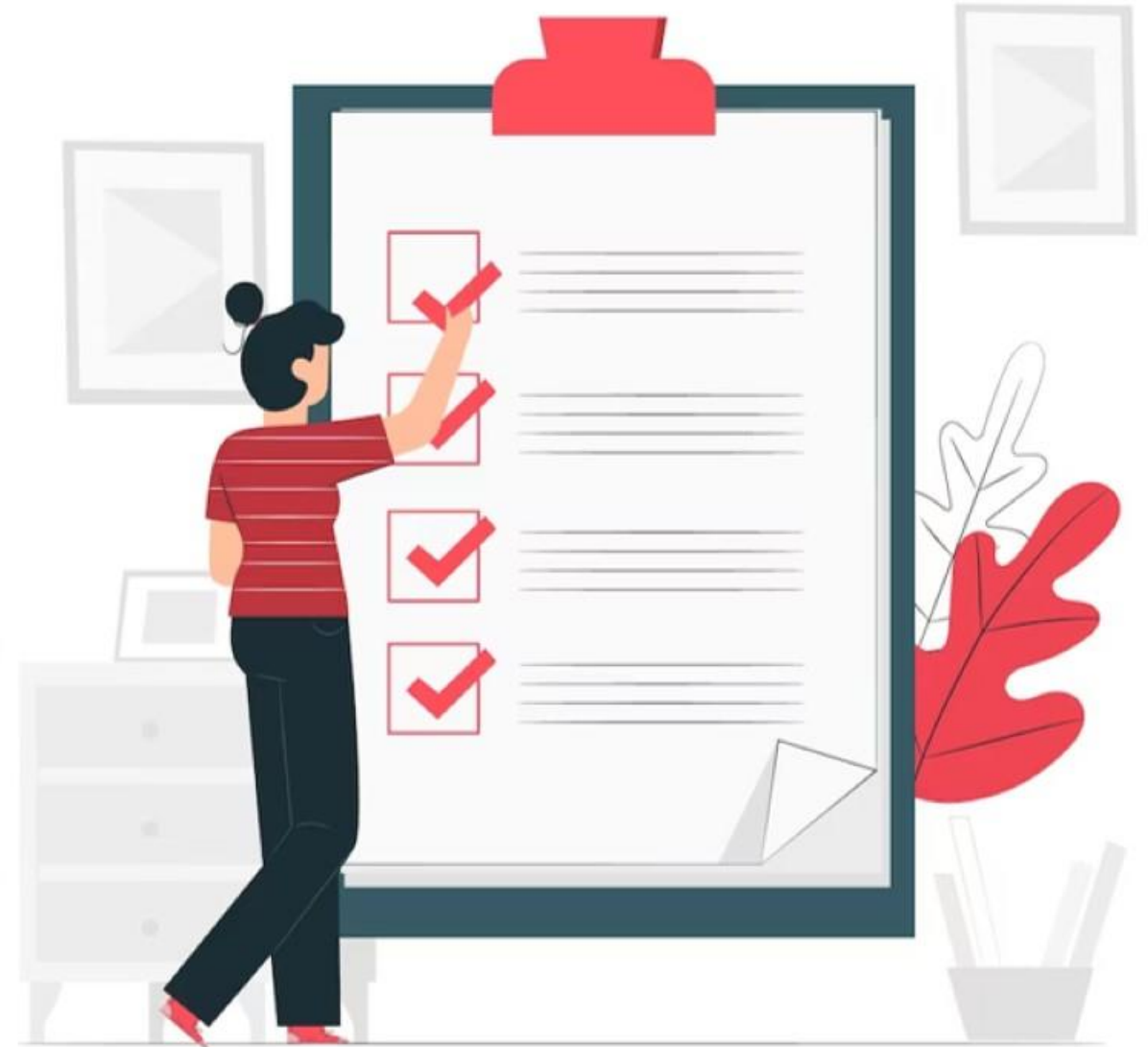
**OBJECTIFS DE LA RÉUNION :**

**LIBRAIRIE EN LIGNE :**

- LES CHIFFRES CLÉS**
- ANALYSE DES CLIENTS**

# SOMMAIRE

1. CONTEXTE (PROBLÉMATIQUE)
2. MÉTHODOLOGIE
3. LE CHIFFRE D'AFFAIRE :  
ÉVOLUTION ET RÉPARTITION
4. LES PRODUITS :  
TOPS/FLOPS
5. ANALYSE DES CLIENTS :  
PROFILS ET LIENS
6. CONCLUSION





# 1 - CONTEXTE

LA LIBRAIRIE EN LIGNE A OUVERT DEPUIS MAINTENANT 2 ANS.  
QUELS SONT LES RÉSULTATS ?

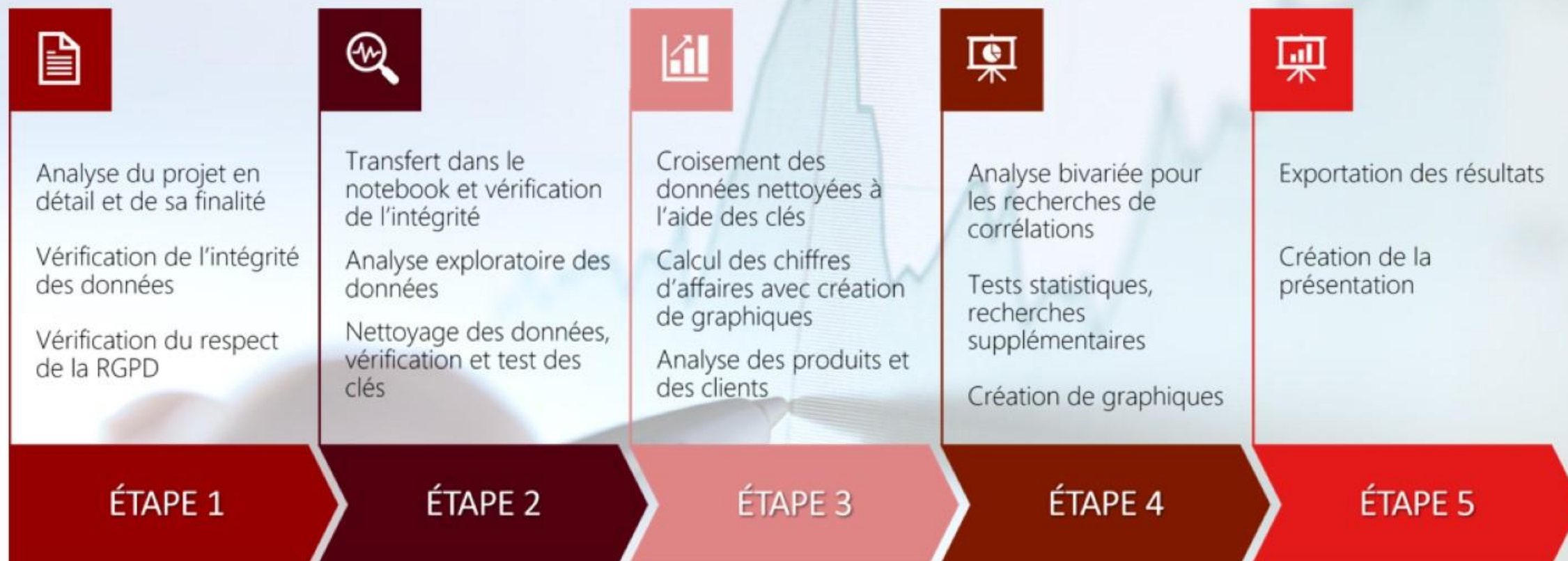
POUR LE SAVOIR, NOUS ALLONS FAIRE LE POINT SUR :

- LES INDICATEURS DE VENTE
- LE COMPORTEMENT ET LES PROFILS DES CLIENTS

CELA NOUS PERMETTRA DE MIEUX CONNAÎTRE NOTRE CLIENTÈLE, DE RECTIFIER CERTAINS POINTS SI BESOIN, ET D'ADAPTER NOTRE OFFRE POUR FIDÉLISER NOS CLIENTS.



## 2.1 - MÉTHODOLOGIE



## 2.2 - ANALYSE EXPLORATOIRE ET NETTOYAGE DES 3 FICHIERS



- VÉRIFICATION DOUBLONS / VALEURS NULLES / ABERRANTES
- PAS DE DONNÉES TESTS
- TEST DE L'UNICITÉ DES CLÉS
- AJOUT D'UNE CLÉ POUR LA TABLE TRANSACTIONS
- CALCUL DE L'ÂGE DES CLIENTS
- TRAITEMENT DES ERREURS DE DATE
- CONVERSION DU TYPE DES DATES
- RENOMMAGE / AJOUT / SUPPRESSION COLONNES



**PAS DE PERTES DE DONNÉES**



## 2.3 - CROISEMENT DES DONNÉES (JOINTURES)



JOINTURE INTERNE DU FICHIER  
PRODUITS AU FICHIER VENTES

PUIS 2<sup>E</sup> JOINTURE INTERNE  
AVEC LE FICHIER CLIENTS

=

687 534 LIGNES  
8 600 CLIENTS  
3 265 PRODUITS

9 COLONNES :

- id\_produit
- id\_session
- id\_client
- date\_de\_vente
- id\_vente
- prix
- categorie
- age
- genre

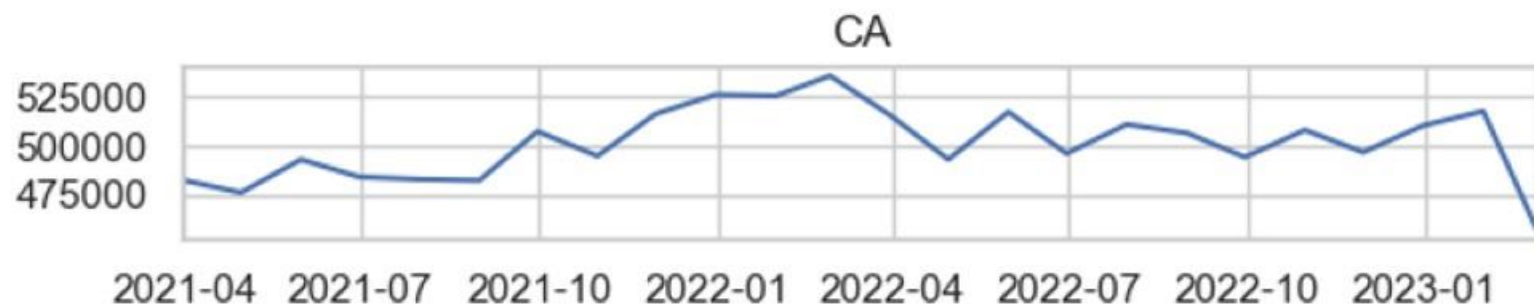
VÉRIFICATION  
À CHAQUE ÉTAPE



# 3.1 – LE CHIFFRE D'AFFAIRE : SON ÉVOLUTION

PÉRIODE : MARS 2021 / FÉVRIER 2023

MOYENNE MOBILE SUR 3 MOIS

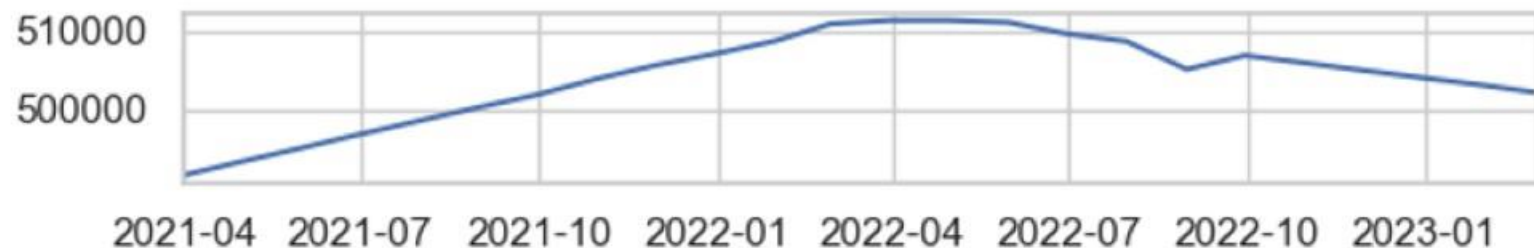


ÉVOLUTION DU CA

CA 2021 : 4 944 761 €

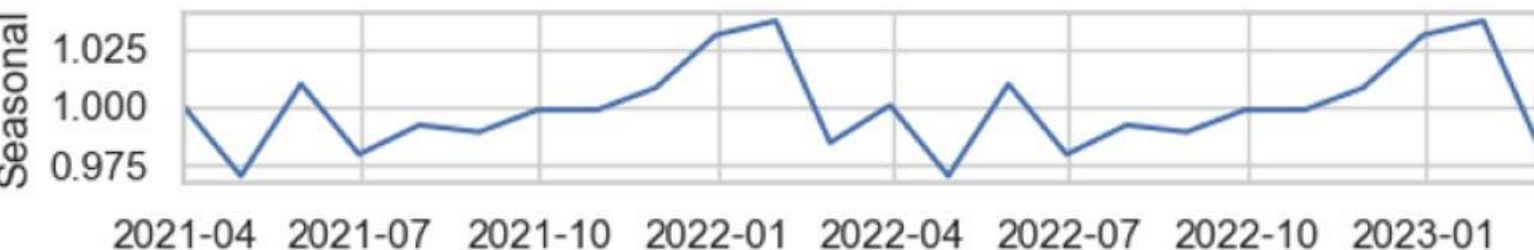
CA 2022 : 6 108 682 €

Trend



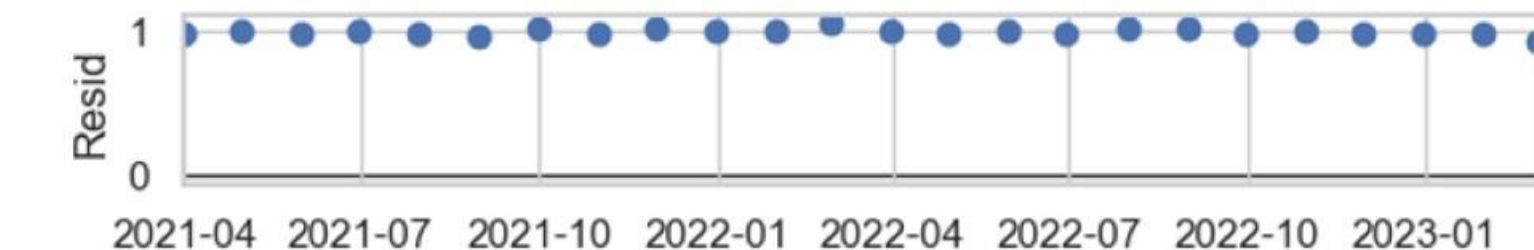
TENDANCE OBTENUE AVEC LA  
MOYENNE MOBILE

Seasonal



SAISONNALITÉ

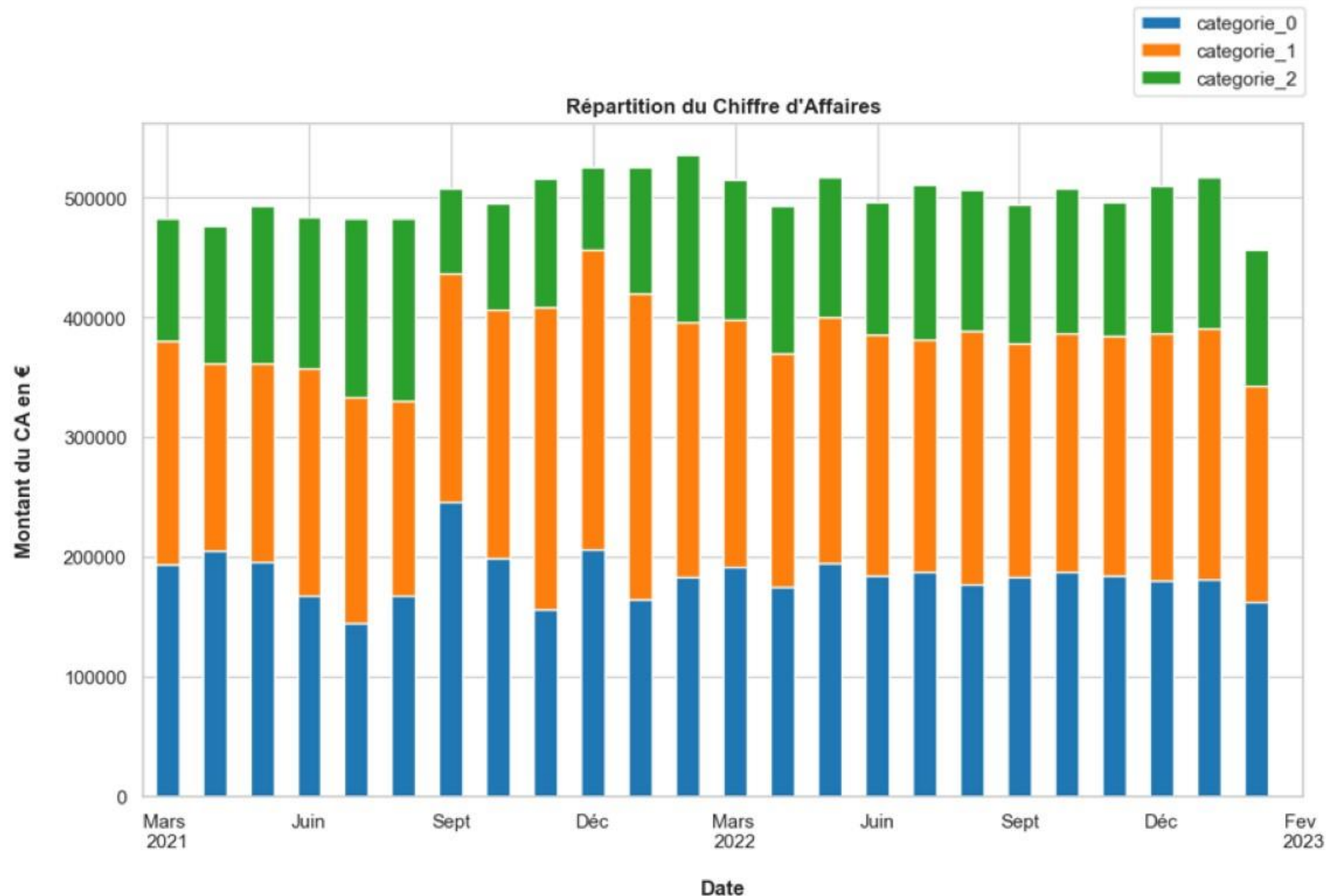
Resid



PHÉNOMÈNES QUI SORTENT DU  
CADRE (ERREURS)  
CA – SAISONNALITÉ

## 3.2 – LE CHIFFRE D'AFFAIRE : SA RÉPARTITION PAR CATÉGORIES

TOTAL CA :  
12 MILLIONS



**CATEGORIE 1 : 40 % DU CA**  
→ PART LA PLUS IMPORTANTE  
DU CA GLOBAL LA PLUPART DU  
TEMPS

**CATEGORIE 0 : 37 % DU CA**  
→ AVEC LA CATEGORIE 1,  
SE PARTAGENT LA PREMIÈRE  
PLACE SELON LES MOIS

**CATEGORIE 2 : 23 % DU CA**  
→ PART PLUS DISCRÈTE DU CA  
GLOBAL, ASSEZ VARIABLE



## 3.3 – LE CHIFFRE D'AFFAIRE : SA RÉPARTITION PAR TYPES DE CLIENTS

**LES CLIENTS PROS :**

**REPRÉSENTENT 7.35% DU CA GLOBAL**

**TOTAL : 884 296 €**

**23 039 COMMANDES SUR TOUTE LA PÉRIODE**



**SERONT RETIRÉS DES ANALYSES DE PROFILS POUR NE PAS FAUSSER LES RÉSULTATS**

**4 PROS DÉTECTÉS DANS LES DONNÉES :**

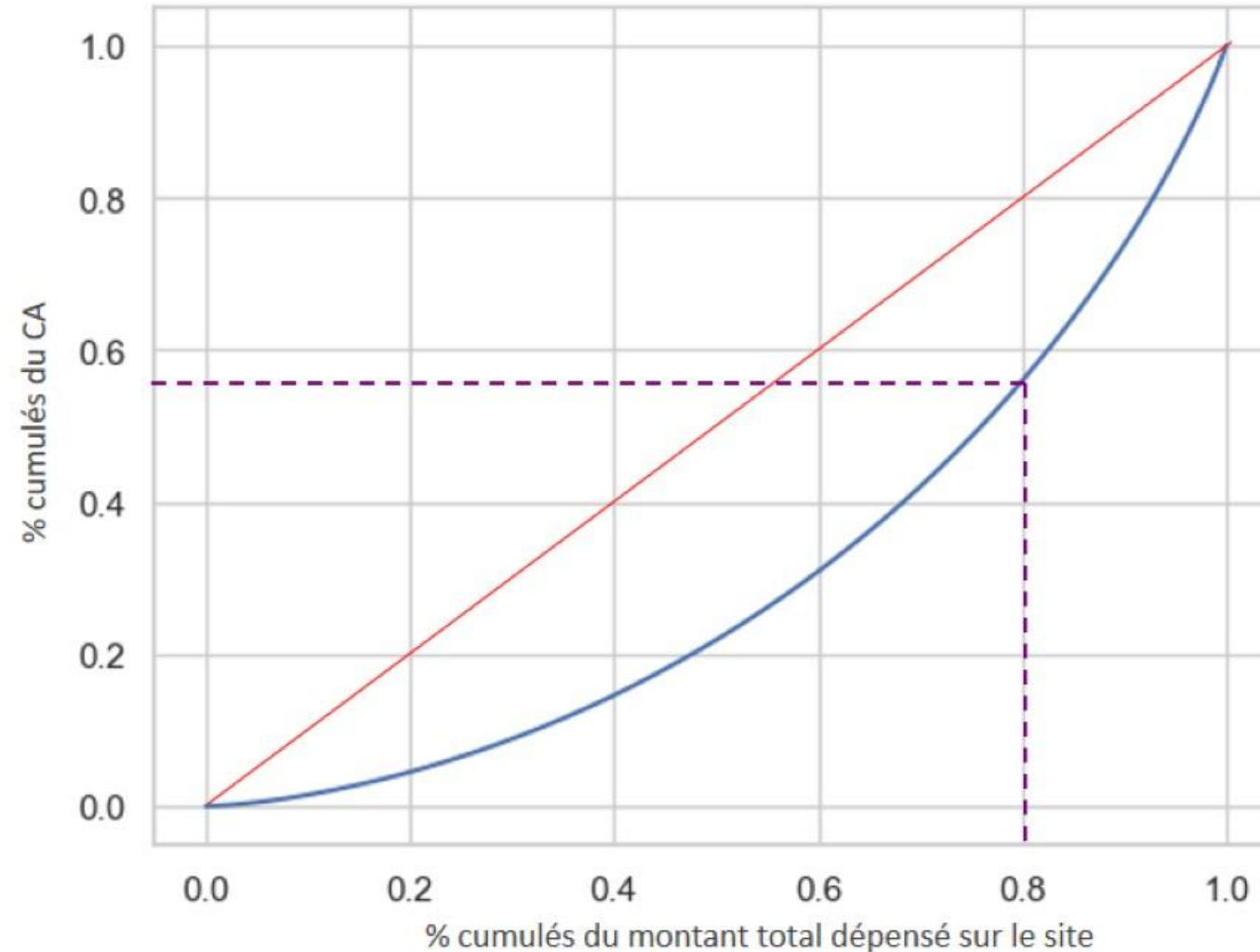
	id_client	age	genre	CA	total_commandes
0	c_1609	43	Homme	326039.89	10997
1	c_4958	24	Homme	290227.03	3851
2	c_6714	55	Femme	153918.60	2620
3	c_3454	54	Homme	114110.57	5571

## 3.4 – LE CHIFFRE D'AFFAIRE : SA RÉPARTITION PAR TYPES DE CLIENTS



### RÉPARTITION NON ÉGALITAIRE SUR LA COURBE DE LORENZ

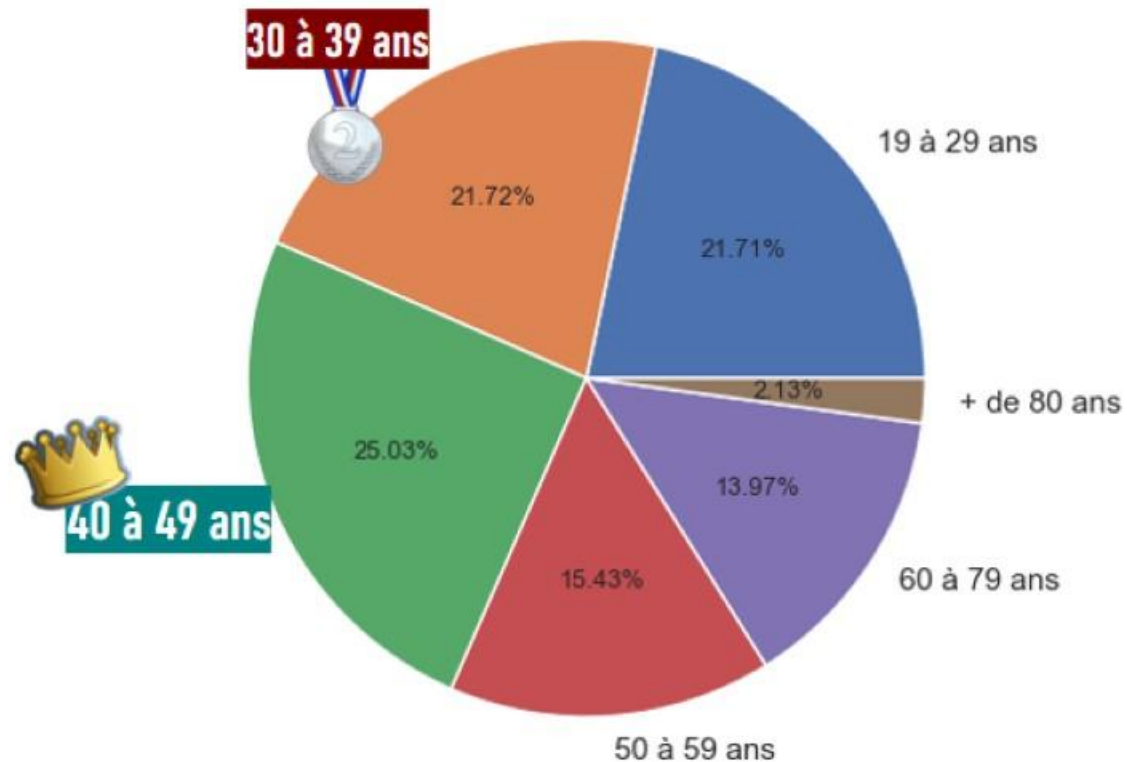
**LES 20% QUI ONT LE PLUS  
DÉPENSÉ SUR LE SITE  
REPRÉSENTENT ENVIRON  
45% DU CA**



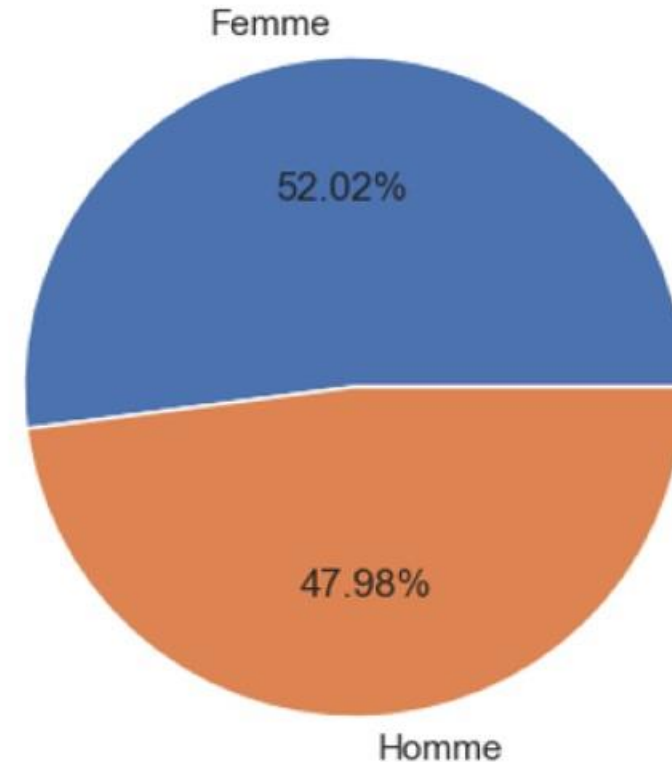
## 3.5 – LE CHIFFRE D'AFFAIRE : SA RÉPARTITION PAR TYPES DE CLIENTS



### EN FONCTION DE L'ÂGE :



### EN FONCTION DU GENRE :

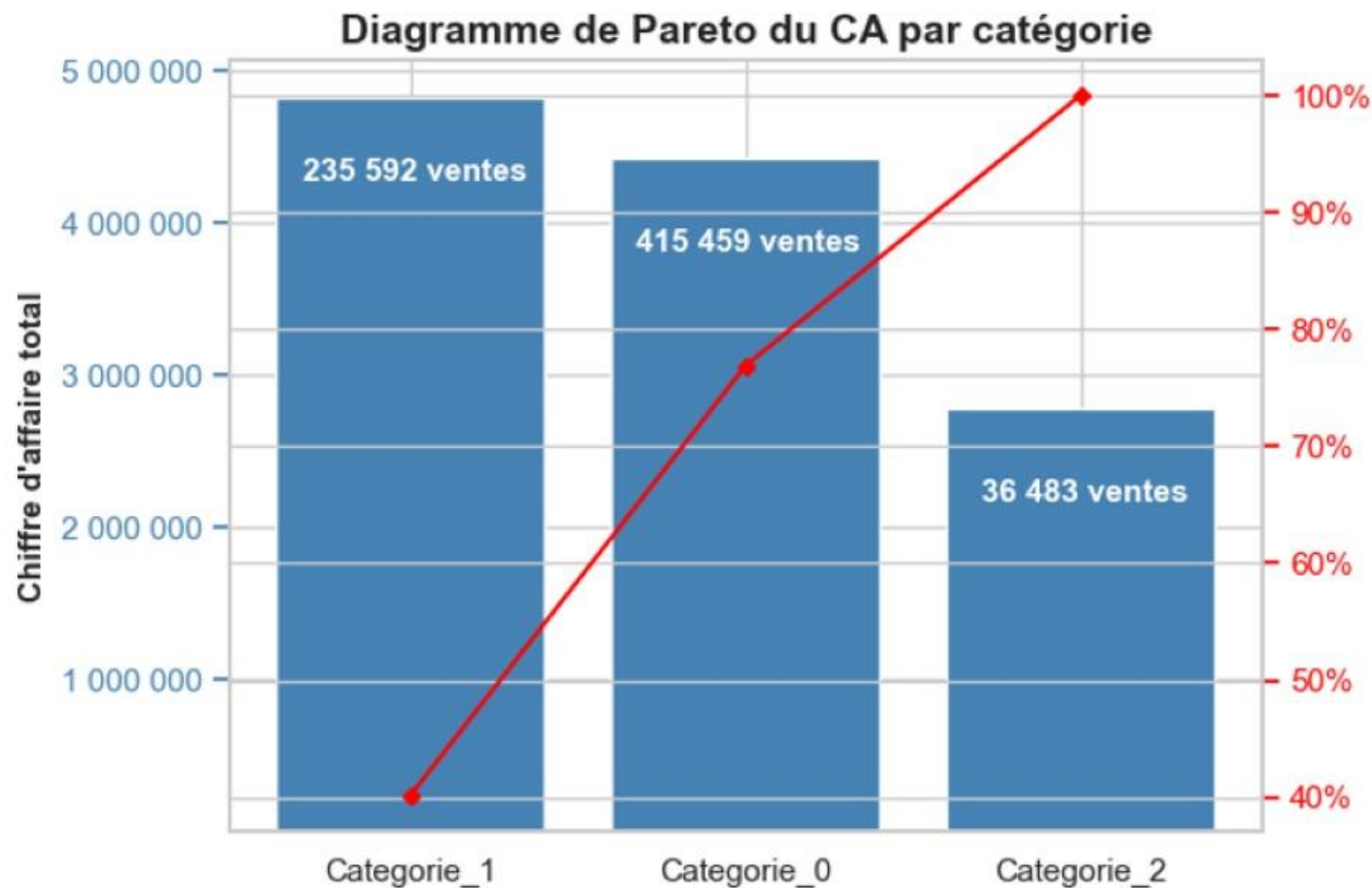


**PRESQUE  
HOMOGÈNE**





## 4.1- LES PRODUITS : DIAGRAMME DE PARETO



LA CATÉGORIE 1 REPRÉSENTE  
40 % DU CA

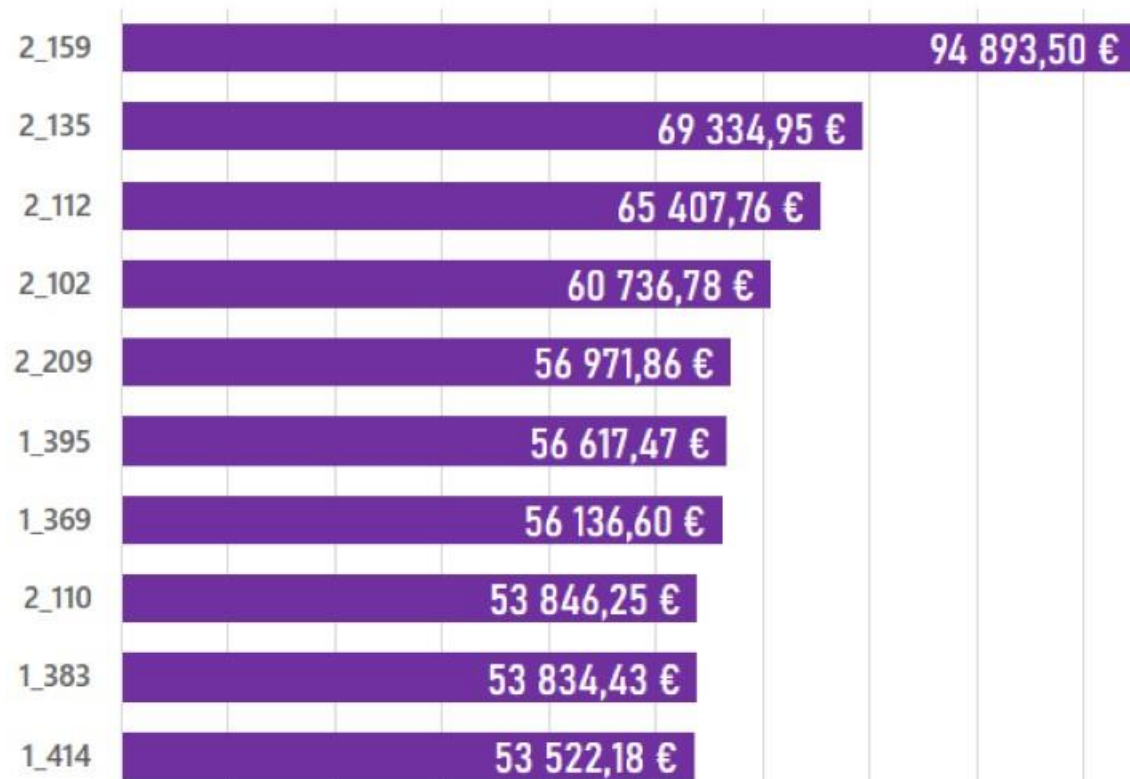
CATÉGORIE 1 + CATÉGORIE 0 =  
ENVIRON 80 % DU CA

LA CATÉGORIE 2 A BEAUCOUP  
MOINS DE VENTES MAIS  
REPRÉSENTE ENVIRON  
20 % DU CA

## 4.2 – LES PRODUITS : TOP 10 DEPUIS L'OUVERTURE



TOP 10 CA



**5% DU CA GÉNÉRAL**  
**TOTAL : 621 302 €**

TOP 10 VENTES



**3,6% DU CA GÉNÉRAL**  
**TOTAL = 434 046 €**

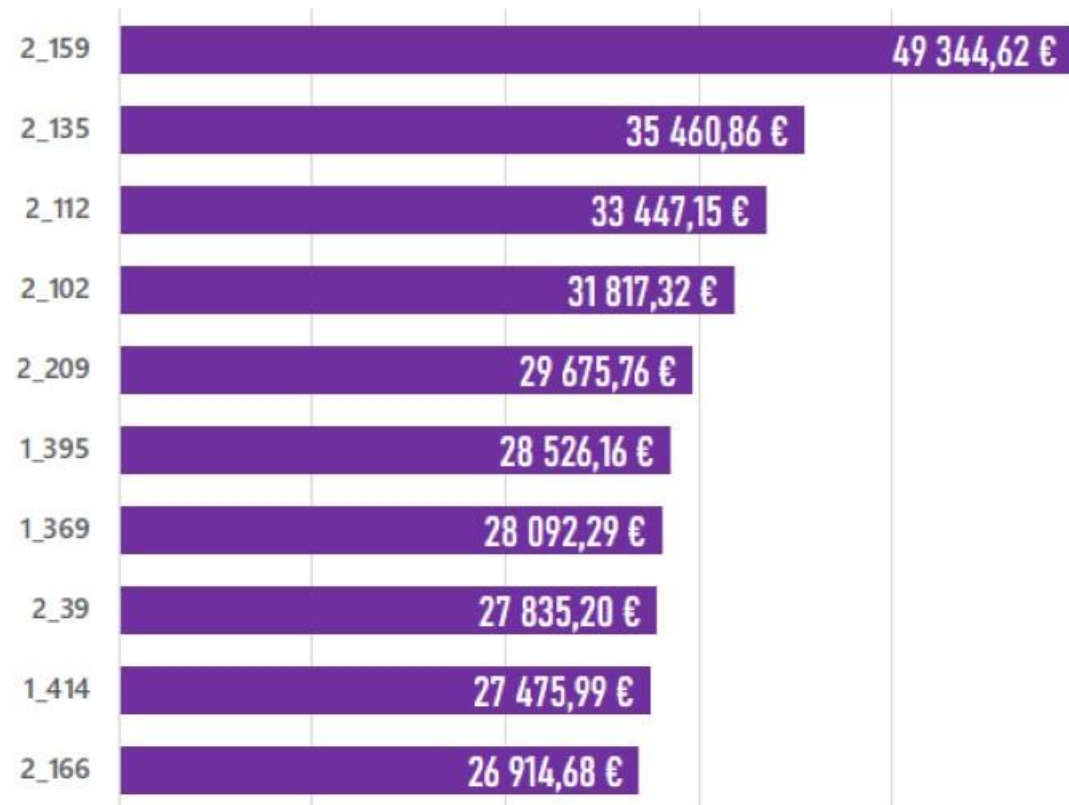
## 4.3 – LES PRODUITS : TOP 10 SUR LES 12 DERNIERS MOIS

PÉRIODE : 03/22 À 02/23

CA DE LA PÉRIODE : 5 967 403 €



### TOP 10 CA



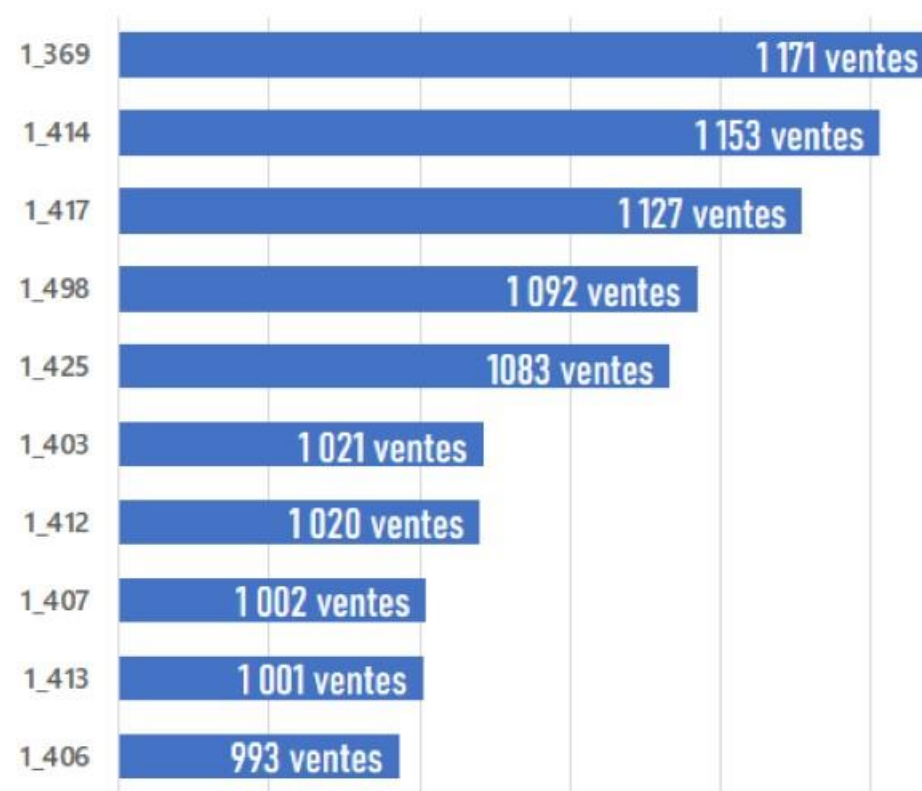
5,3 % DU CA GÉNÉRAL

TOTAL : 318 590 €

7 PREMIERS IDENTIQUES

2 NOUVELLES RÉFÉRENCES

### TOP 10 VENTES



3,6% DU CA GÉNÉRAL

TOTAL = 434 046 €

1<sup>ER</sup> IDENTIQUE

RÉFÉRENCES PRÉSENTES INCHANGÉES



## 4.4 – LES PRODUITS : LES FLOPS (INVENDUS)

id_produit	prix	catégorie
0_1016	35.06	0
0_1780	1.67	0
0_1062	20.08	0
0_1119	2.99	0
0_1014	1.15	0
1_0	31.82	1
0_1318	20.92	0
0_1800	22.05	0
0_1645	2.99	0
0_322	2.99	0
0_1620	0.80	0
0_1025	24.99	0
2_87	220.99	2
1_394	39.73	1
2_72	141.32	2
0_310	1.94	0
0_1624	24.50	0
2_86	132.36	2
0_299	22.99	0
0_510	23.66	0

**20 RÉFÉRENCES VENDUES 0 FOIS**

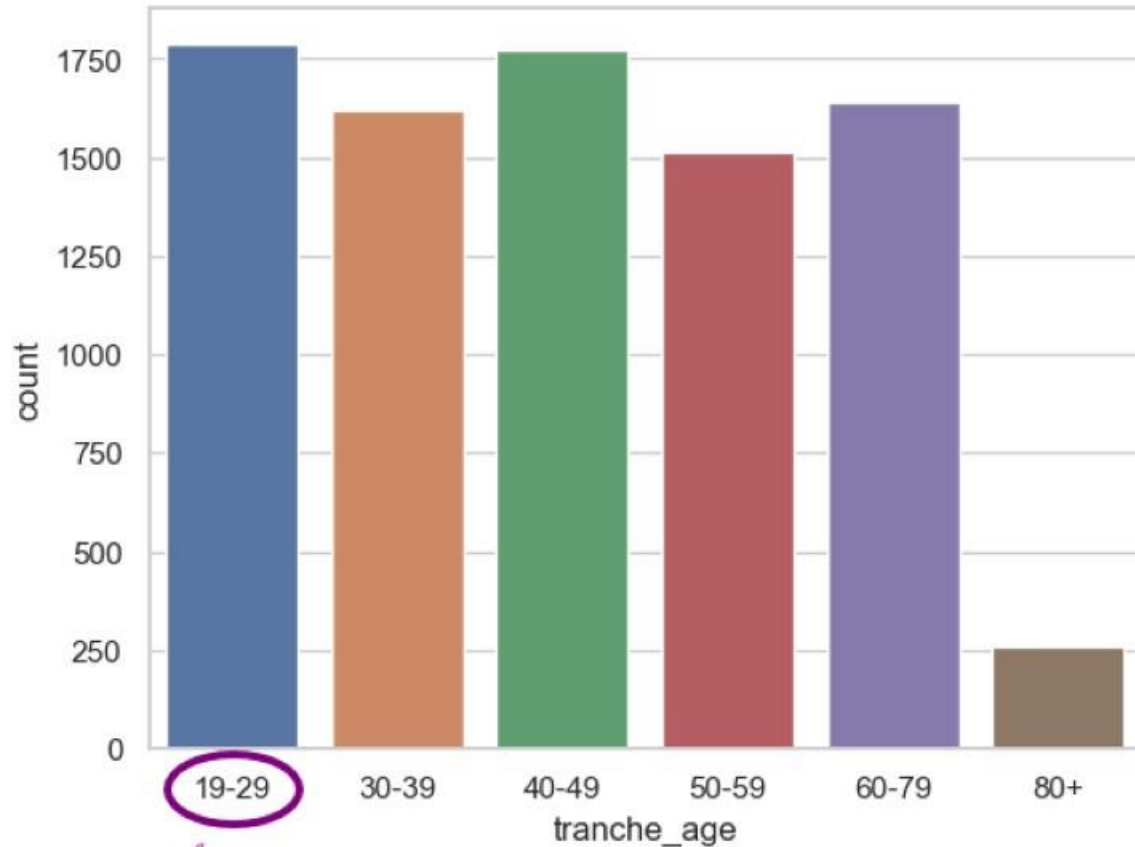
**DEPUIS L'OUVERTURE**

**LISTE À FAIRE VÉRIFIER  
(PRIX, VISIBILITÉ SUR LE SITE... )**

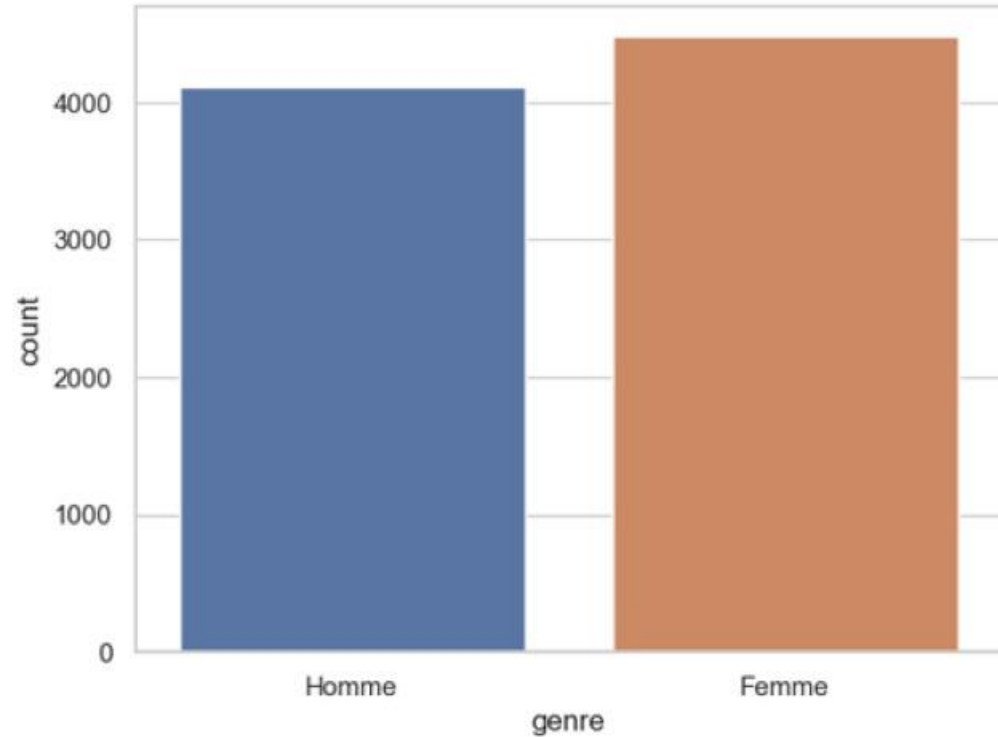
## 5.1 – ANALYSE DES CLIENTS : PROFILS



**8596 CLIENTS PARTICULIERS**



**BEAUCOUP DE CLIENTS DE « 19 » ANS**



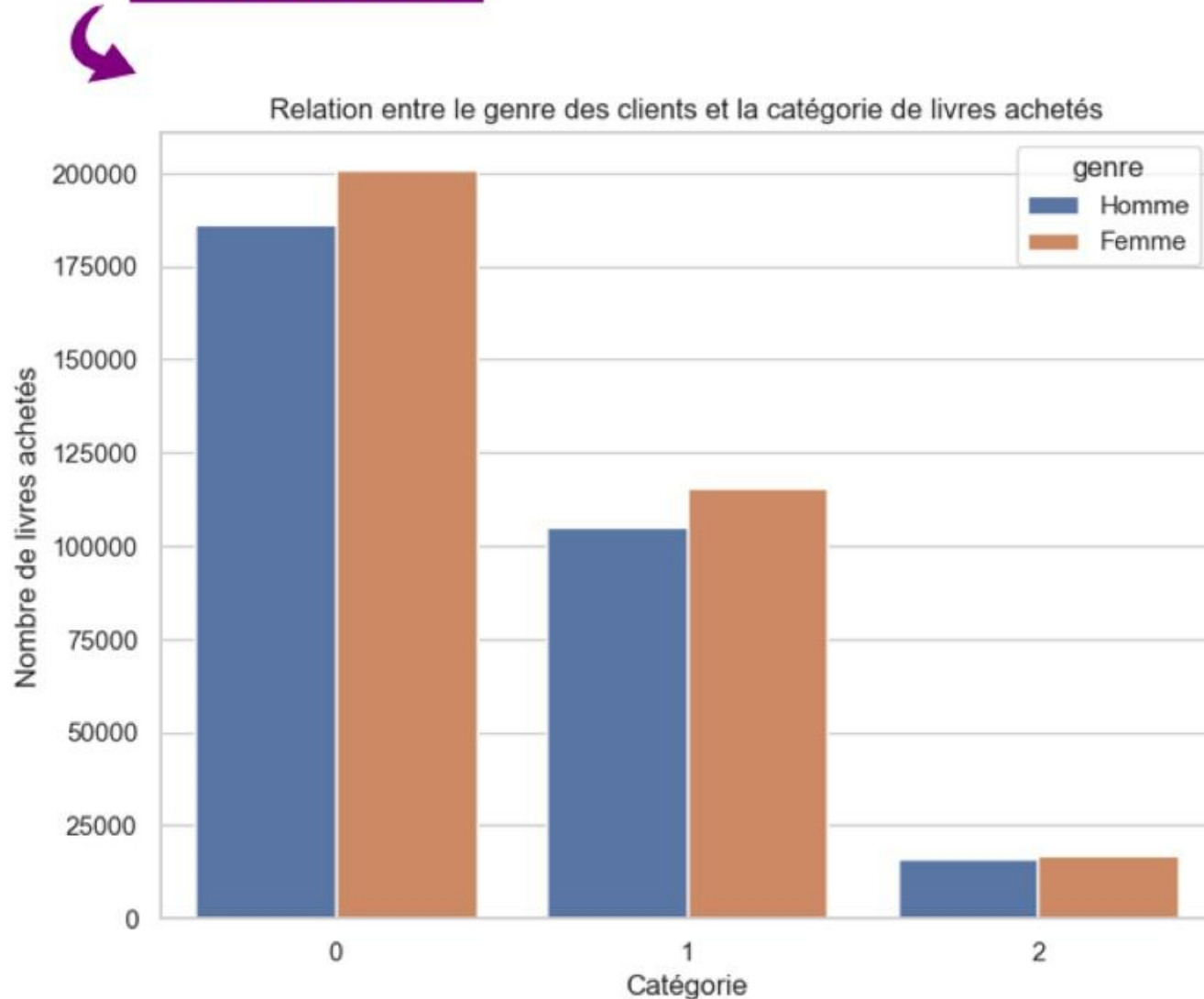
**+ 4 CLIENTS PROS**

**+ 21 INSCRITS SANS ACHATS**

## 5.2 – ANALYSE DES CLIENTS : LIEN GENRE VS CATÉGORIE



### ANALYSE GRAPHIQUE



### 2 VARIABLES QUALITATIVES



#### TABLEAU DE CONTIGENCE



#### TEST DE CHI-2 :

Stat : 22,67  
P-value < 0,05

#### INDIQUE UNE ASSOCIATION

LA P-VALUE REJETTE L'HYPOTHÈSE NULLE  
(PAS D'ASSOCIATION)



COEFFICIENT DE CRAMER-V INDIQUE UNE TRÈS  
TRÈS FAIBLE INTENSITÉ (0,005)



COEFFICIENT DE TSCHUPROW  
DONNE LE MÊME RÉSULTAT

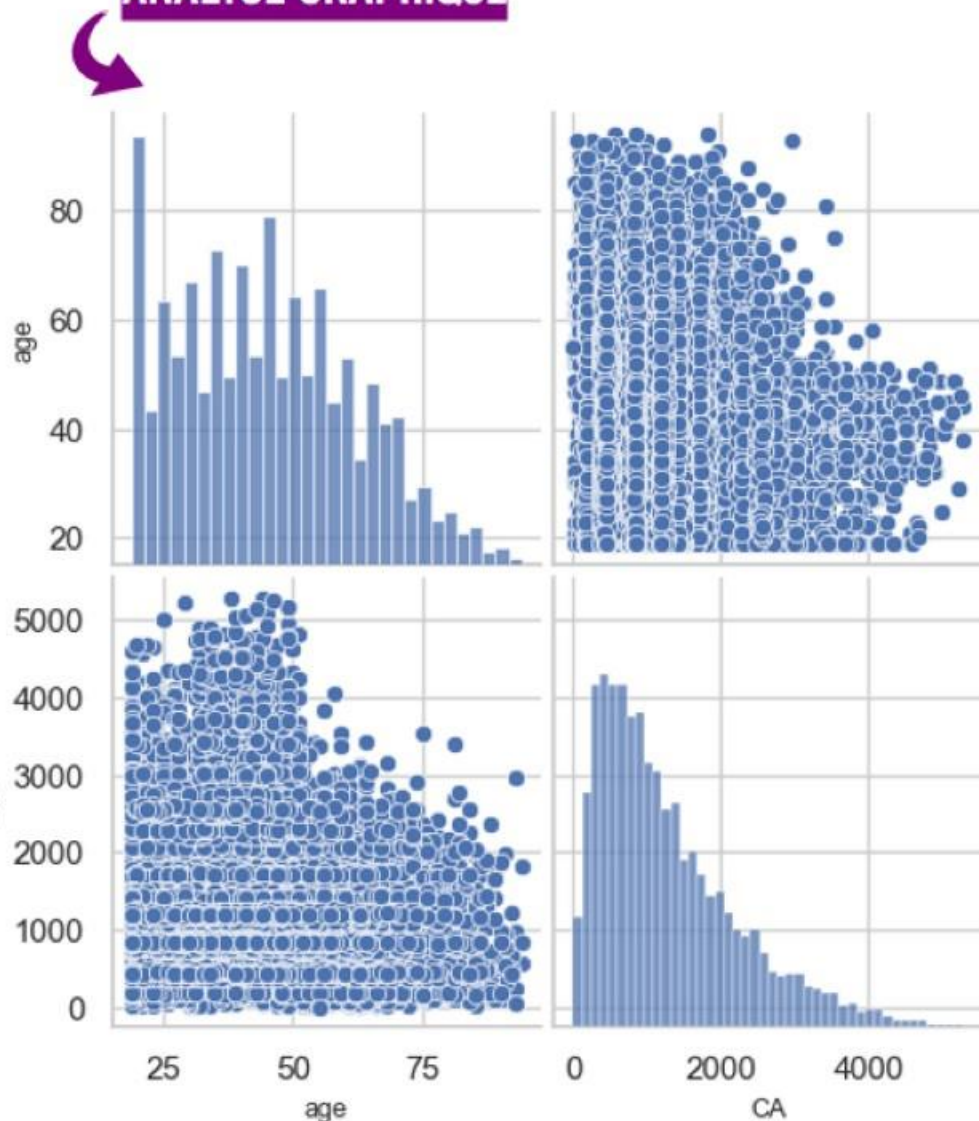
**DONC ASSOCIATION DE TRÈS FAIBLE INTENSITÉ**



## 5.3 – LIEN ÂGE VS MONTANT TOTAL DES ACHATS



### ANALYSE GRAPHIQUE



### 2 VARIABLES QUANTITATIVES

DISTRIBUTIONS NON NORMALES

CORRÉLATION DE SPEARMAN

Stat : - 0,18  
P-value < 0,05

INDIQUE UNE CORRÉLATION, LÉGÈRE, NÉGATIVE ET MONOTONE  
LA P-VALUE REJETTE L'HYPOTHÈSE NULLE (PAS DE CORRÉLATION)

CORRÉLATION DE KENDALL

Stat : - 0,12  
P-value < 0,05

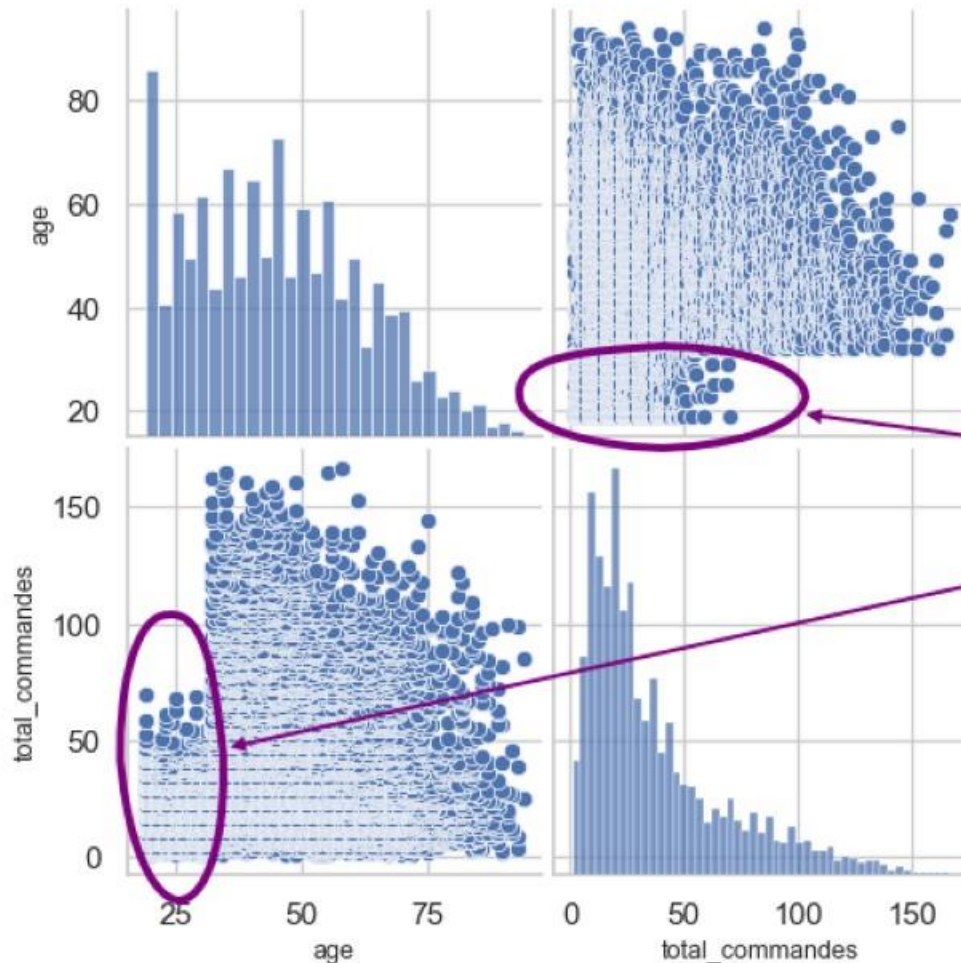
MÊMES RÉSULTATS

DONC CORRÉLATION DE FAIBLE INTENSITÉ,  
NÉGATIVE ET MONOTONE

## 5.4 – LIEN ÂGE VS FRÉQUENCE DES ACHATS (TOUTE LA PÉRIODE)



### ANALYSE GRAPHIQUE



### 2 VARIABLES QUANTITATIVES

DISTRIBUTIONS NON NORMALES

CORRÉLATION DE SPEARMAN

Stat : 0,21  
P-value < 0,05

CORRÉLATION DE KENDALL

Stat : 0,14  
P-value < 0,05

RÉSULTATS ILLOGIQUES,  
FAUSSÉ PAR LE GROUPE DES JEUNES

VISUALISATION PAR TRANCHES D'ÂGE



## 5.4 – LIEN ÂGE VS FRÉQUENCE DES ACHATS (TOUTE LA PÉRIODE)



**CE GROUPE SE DÉTACHE NETTEMENT  
DES AUTRES**



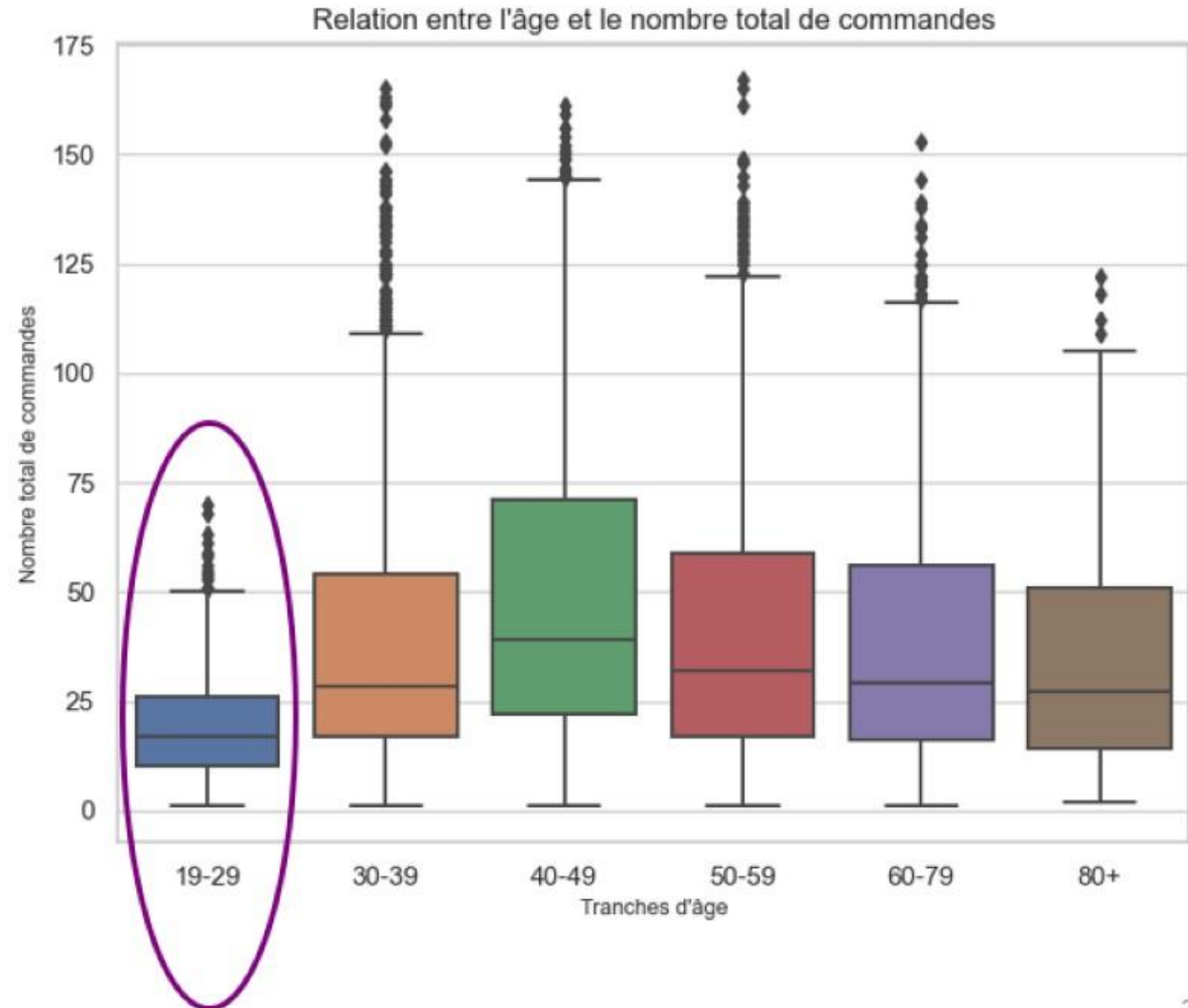
**ÂGE RENSEIGNÉ SUR INTERNET,  
NON VÉRIFIABLE**



**RETRAIT DE CE GROUPE POUR NE PAS  
FAUSSER LES ANALYSES**



**LIMITE FIXÉE : -31 ANS**

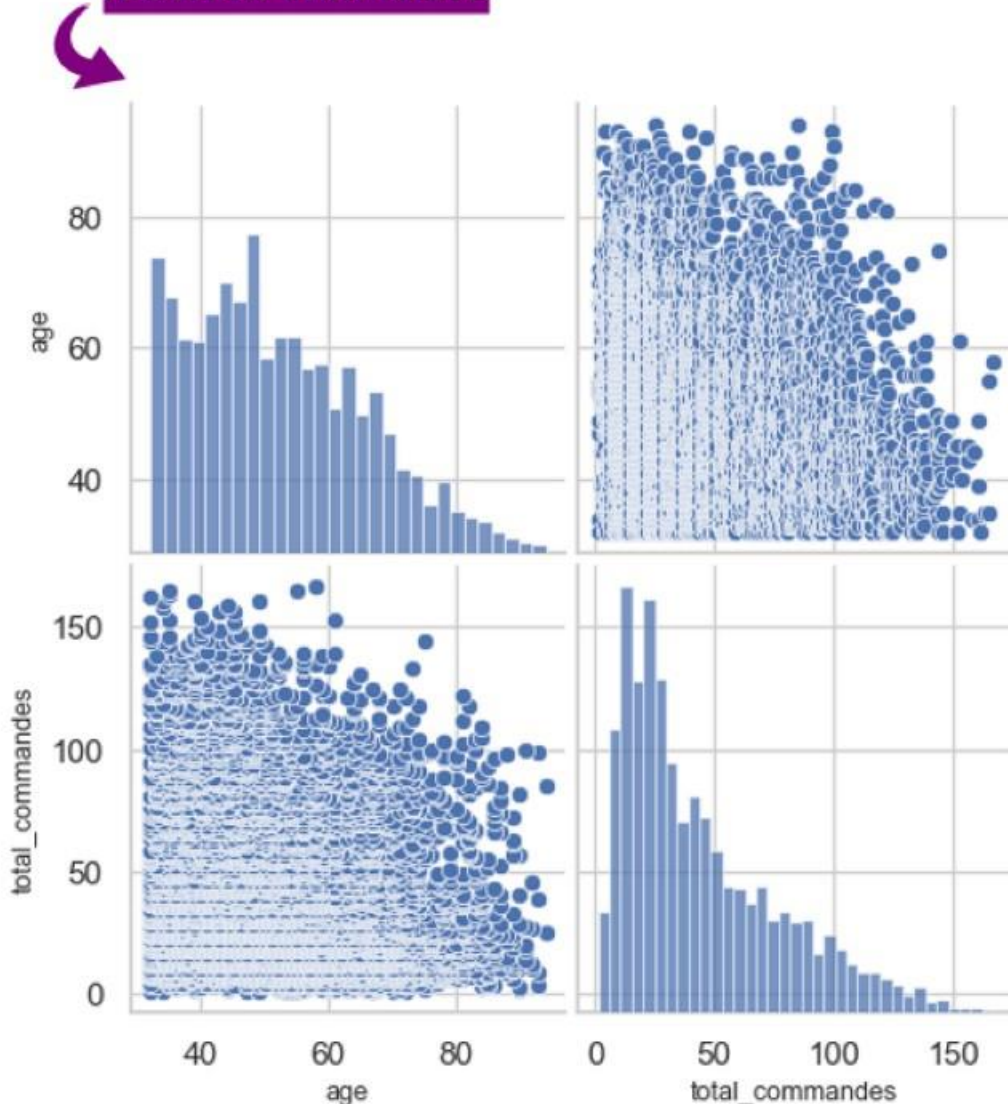




## 5.4 – LIEN ÂGE VS FRÉQUENCE DES ACHATS (SANS LES « JEUNES »)



### ANALYSE GRAPHIQUE



### 2 VARIABLES QUANTITATIVES

DISTRIBUTIONS NON NORMALES

CORRÉLATION DE SPEARMAN

Stat : - 0,14  
P-value < 0,05

INDIQUE UNE CORRÉLATION, LÉGÈRE, NÉGATIVE ET MONOTONE  
LA P-VALUE REJETTE L'HYPOTHÈSE NULLE (PAS DE CORRÉLATION)

CORRÉLATION DE KENDALL

Stat : - 0,09  
P-value < 0,05

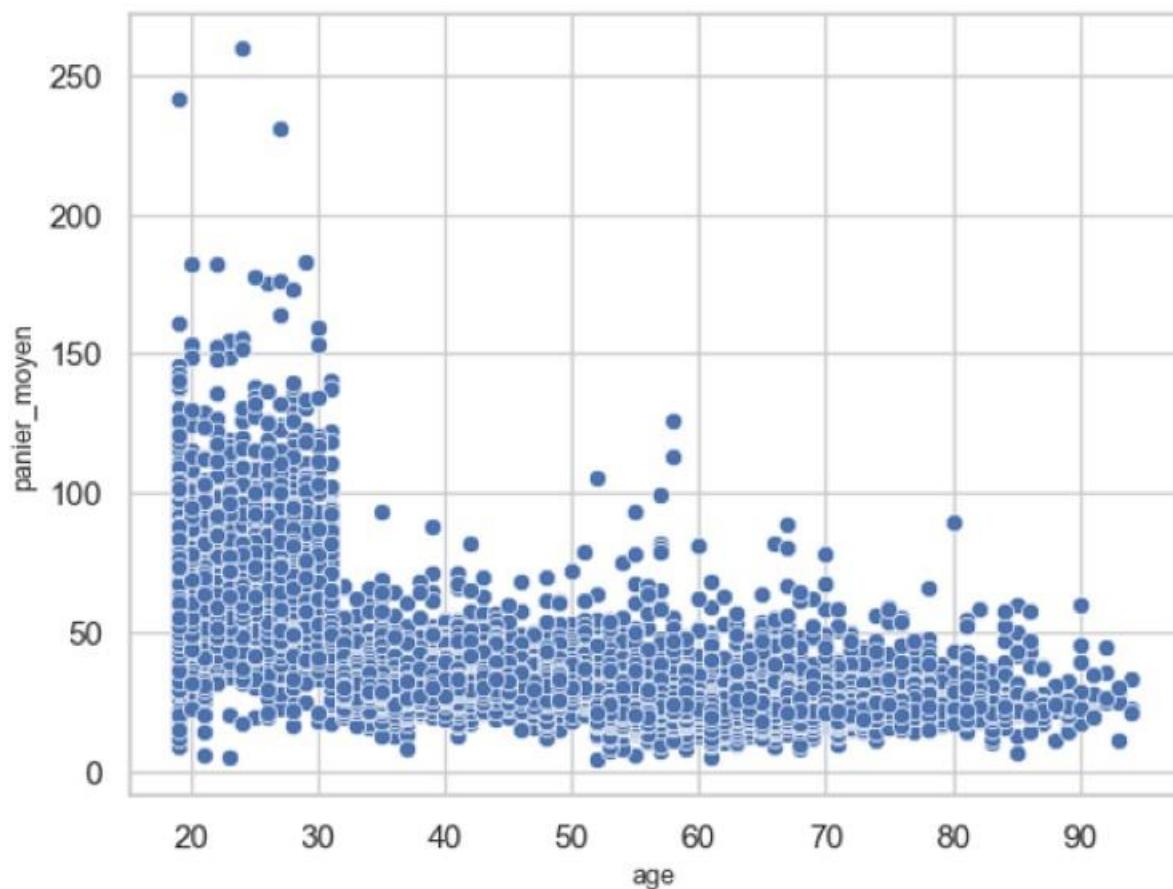
MÊMES RÉSULTATS

**DONC CORRÉLATION DE FAIBLE INTENSITÉ,  
NÉGATIVE ET MONOTONE**

## 5.5 – LIEN ÂGE VS TAILLE DU PANIER MOYEN



### ANALYSE GRAPHIQUE



### 2 VARIABLES QUANTITATIVES

DISTRIBUTIONS NON NORMALES

CORRÉLATION DE SPEARMAN

Stat : - 0,70  
P-value = 0

INDIQUE UNE CORRÉLATION, MOYENNEMENT FORTE,  
NÉGATIVE ET MONOTONE  
LA P-VALUE REJETTE L'HYPOTHÈSE NULLE

CORRÉLATION DE KENDALL

Stat : - 0,50  
P-value = 0

MÊMES RÉSULTATS

**DONC CORRÉLATION MOYENNEMENT FORTE,  
NÉGATIVE ET MONOTONE**

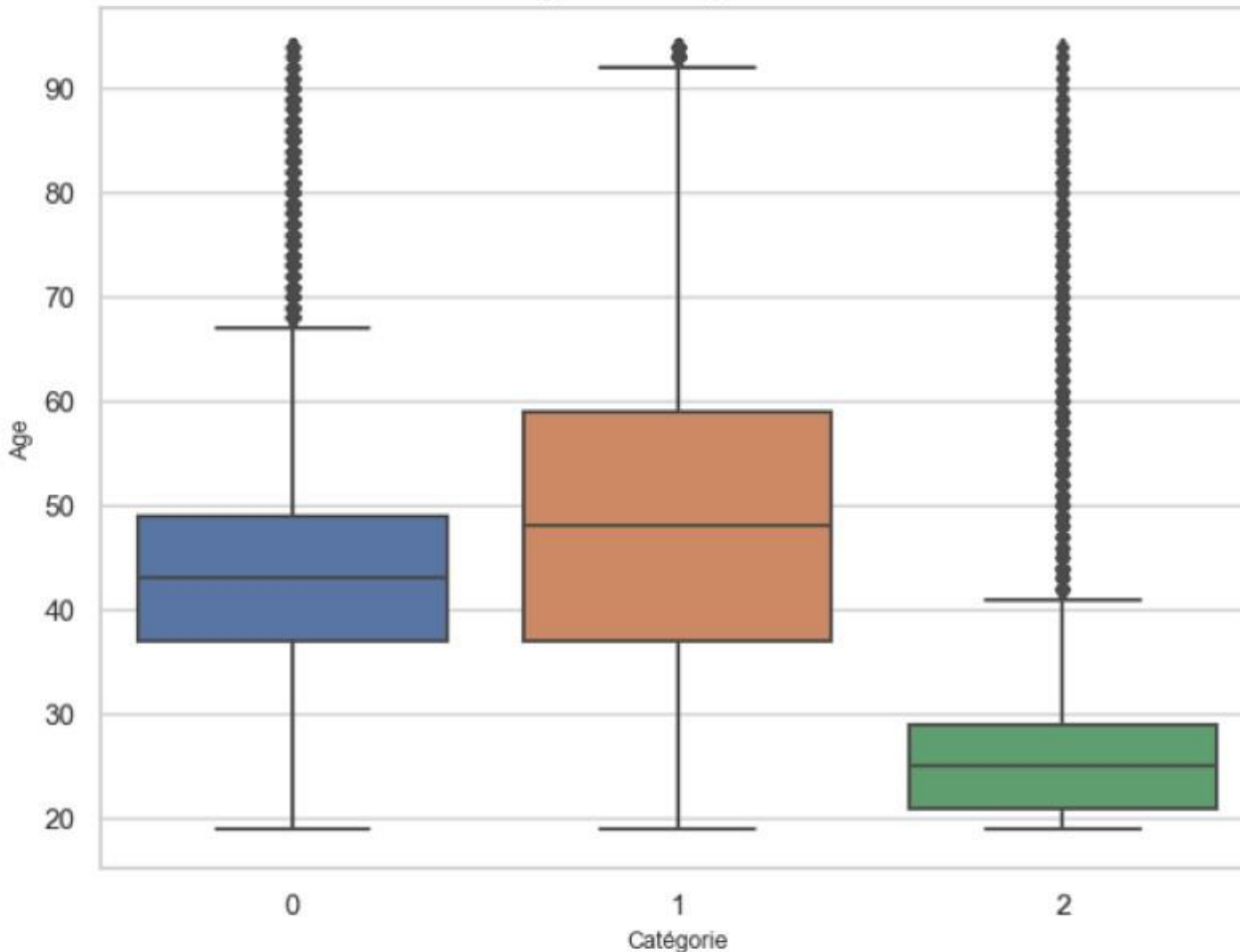


## 5.6 – LIEN ÂGE VS CATÉGORIE



### ANALYSE GRAPHIQUE

Relation entre l'âge et la catégories des livres achetés



### 1 VARIABLE QUANTITATIVE ET 1 QUALITATIVE

DISTRIBUTIONS NON NORMALES

ANOVA NON RECEVABLE  
(PAS D'ÉGALITÉ DES VARIANCES)

TEST DE KRUSKAL-WALLIS

HYPOTHÈSE NULLE REJETÉE, IL Y A  
AU MOINS UNE DISTRIBUTION DIFFÉRENTE

TEST DE MANN-WHITNEY

TOUTES LES DISTRIBUTIONS SONT  
STATISTIQUEMENT DIFFÉRENTES

Stat : 71359  
P-value = 0,0

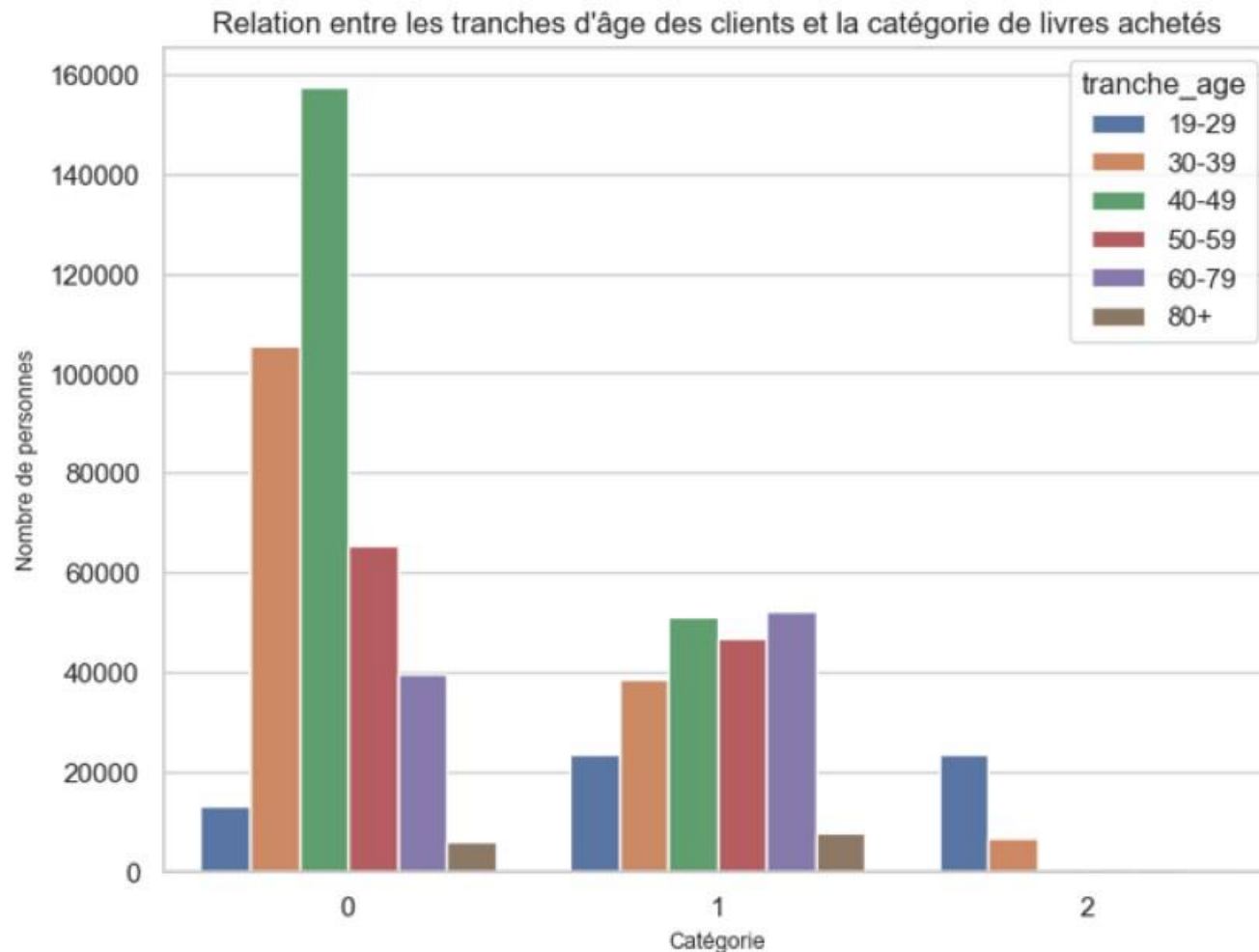
DONC IL Y A CORRÉLATION



## 5.6 – LIEN ÂGE (TRANCHES) VS CATÉGORIE



### ANALYSE GRAPHIQUE



### 2 VARIABLES QUALITATIVES

#### TABLEAU DE CONTIGENCE

#### TEST DE CHI-2 :

Stat : 209618  
P-value = 0,0

#### CONFIRME UNE ASSOCIATION

LA P-VALUE REJETTE L'HYPOTHÈSE NULLE  
(PAS D'ASSOCIATION)

COEFFICIENT DE CRAMER-V INDIQUE UNE  
INTENSITÉ MOYENNE (0,40)

COEFFICIENT DE TSCHUPROW  
DONNE LE MÊME RÉSULTAT (0,40)

**DONC ASSOCIATION MOYENNEMENT FORTE**



## 6 - CONCLUSION

- LES CLIENTS ENTRE 40 ET 49 ANS REPRÉSENTENT LA PLUS GROSSE PART DU CA
- LES 2<sup>èmes</sup> ONT ENTRE 30 ET 39 ANS
- CES 2 GROUPES PORTENT UN GRAND INTÉRÊT À LA CATÉGORIE 0, ET APPRÉCIENT LA CATÉGORIE 1
- LES CLIENTS DE « 19 ANS » FAUSSENT LES RÉSULTATS ET PROFILS DU GROUPE DES JEUNES (ÂGE MINIMUM ?)
- LA CATÉGORIE 2 EST LA MOINS APPRÉCIÉE ET N'EST ACHETÉE QUE PAR LES MOINS DE 30 ANS

### ACTIONS RECOMMANDÉES :

- ACTION MARKETING POUR STABILISER LE CA
- SEGMENTATION PAR TRANCHES D'ÂGE
- > IL SERAIT BON D'ADAPTER L'OFFRE EN LIGNE :  
RECOMMANDATIONS AVEC CIBLAGE DES GROUPES PAR ÂGE  
DIVERSIFICATION ? (LIVRES D'OCCASIONS, LECTURE EN LIGNE...)