



KPI et analyse comportementale des clients eshop

Pauline Felten

CONTEXTE

- Lancement de la boutique en ligne il y a deux ans
- Faire un point sur les différents indicateurs afin d'ajuster nos offres ou notre stratégie marketing
- Analyser le profil et le comportement de nos clients

METHODOLOGIE DE L'ANALYSE

- Nettoyer et joindre les différentes bases
- Calcul des KPI
- Analyse des profils clients
- Analyse des comportements clients au moyen de tests statistiques.



ETAPE 1

NETTOYAGE ET

JOINTURE

Nettoyage

Article avec un prix négatif
(T_0)

Suppression :

- Lignes de test (session s_0)
- Clients test (ct_0 et ct_1)
- Article test (T_0)



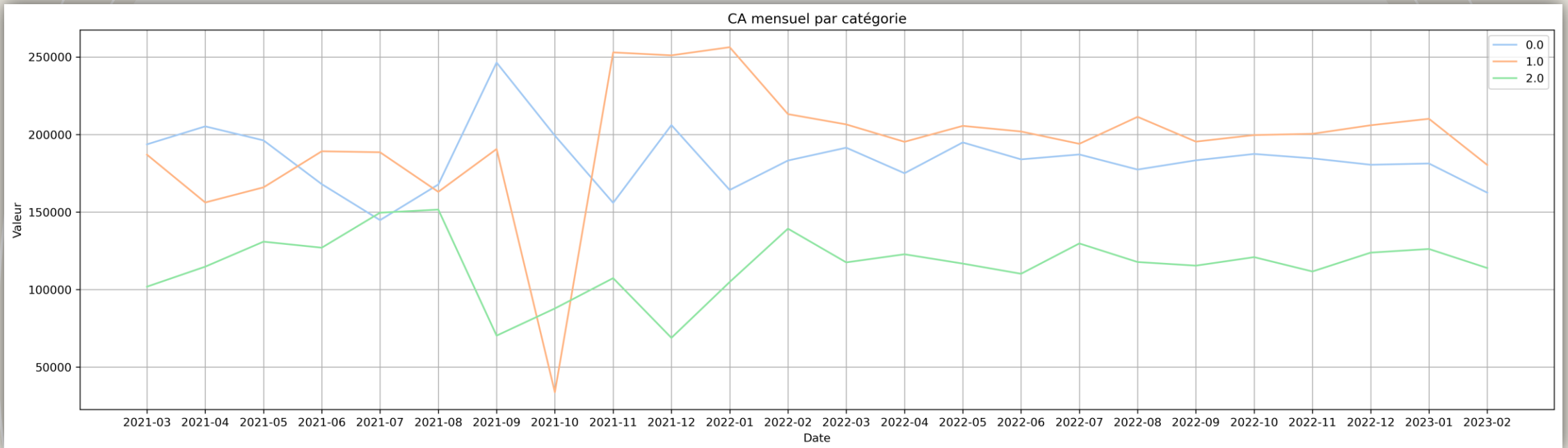
Jointure

21 clients de la BDD customers n'ont jamais passé commande

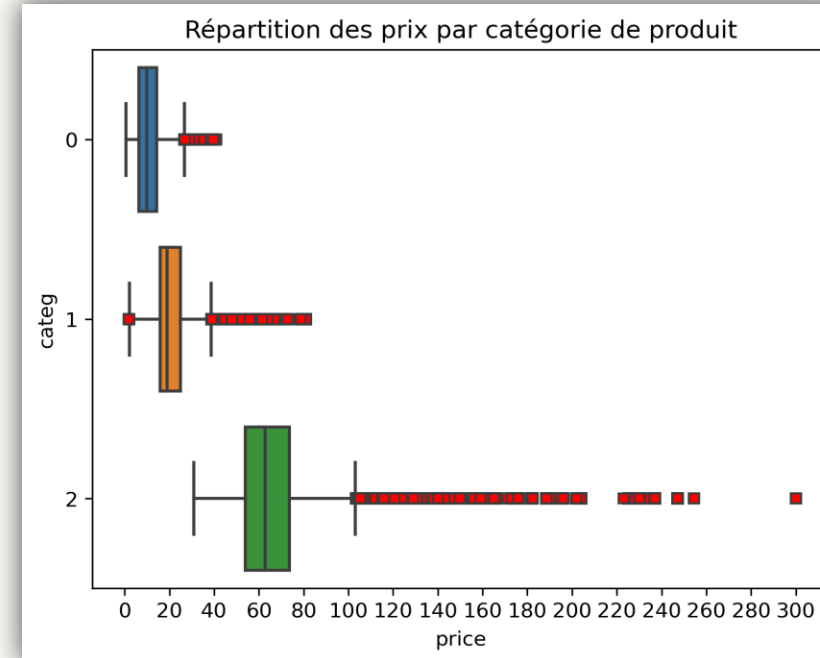
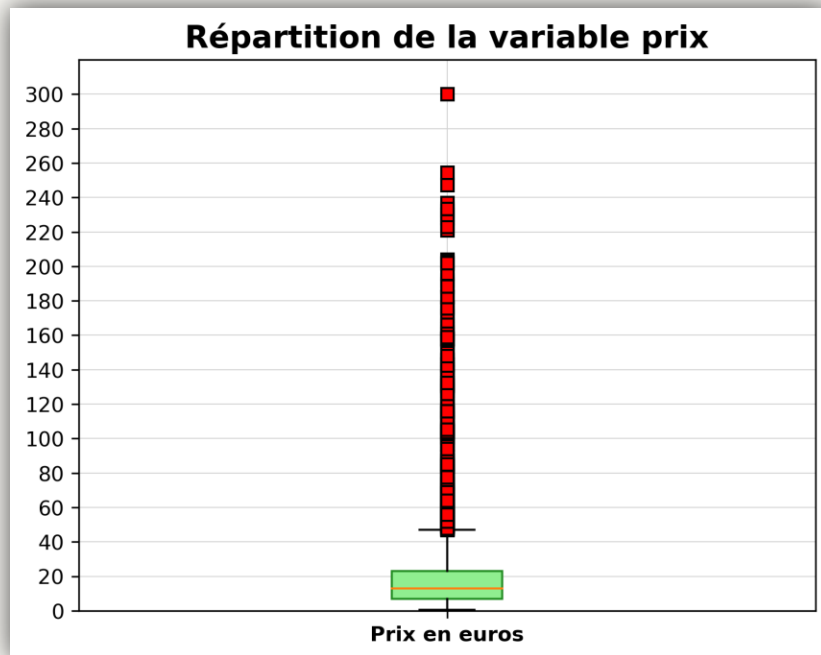
Un article de la BDD transactions ne fait pas parti du fichier products

21 articles de la BDD products n'ont jamais été vendus

Données manquantes



- Les données d'Octobre 2021 sont incomplètes.
- On réalise une imputation par la moyenne pour définir le CA d'octobre 2021
- Pour la partie analyse, on utilise les données nettoyées



Distribution de la variable prix

Le prix d'un article dépend de sa catégorie



ETAPE 2

KPI



Chiffres clés

12M € en 2 ans

501K € par mois (moy)

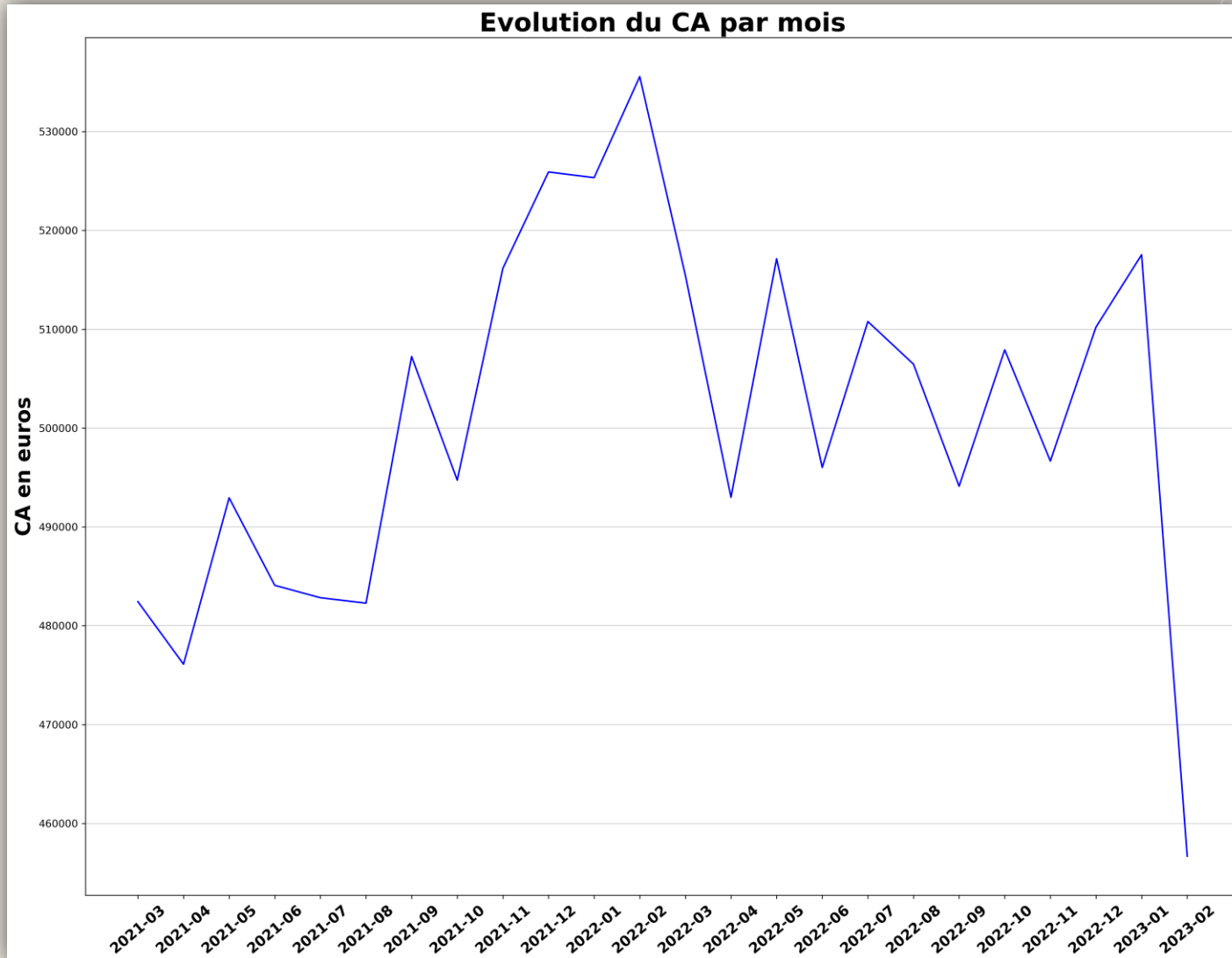
16,5K € par jour (moy)

8600 clients actifs

3286 réf produits

3 catégories de produits

Evolution du CA depuis le lancement



A fin février 2023,
nous avons déjà réalisé **16
% du CA 2022** (moy
487K/mois)

CA restant pour atteinte
CA 2022 : 4.8M€ env soit
480K€/mois

(CA 2022 : 5.7M€
moy mois année 2022 : 472K€)

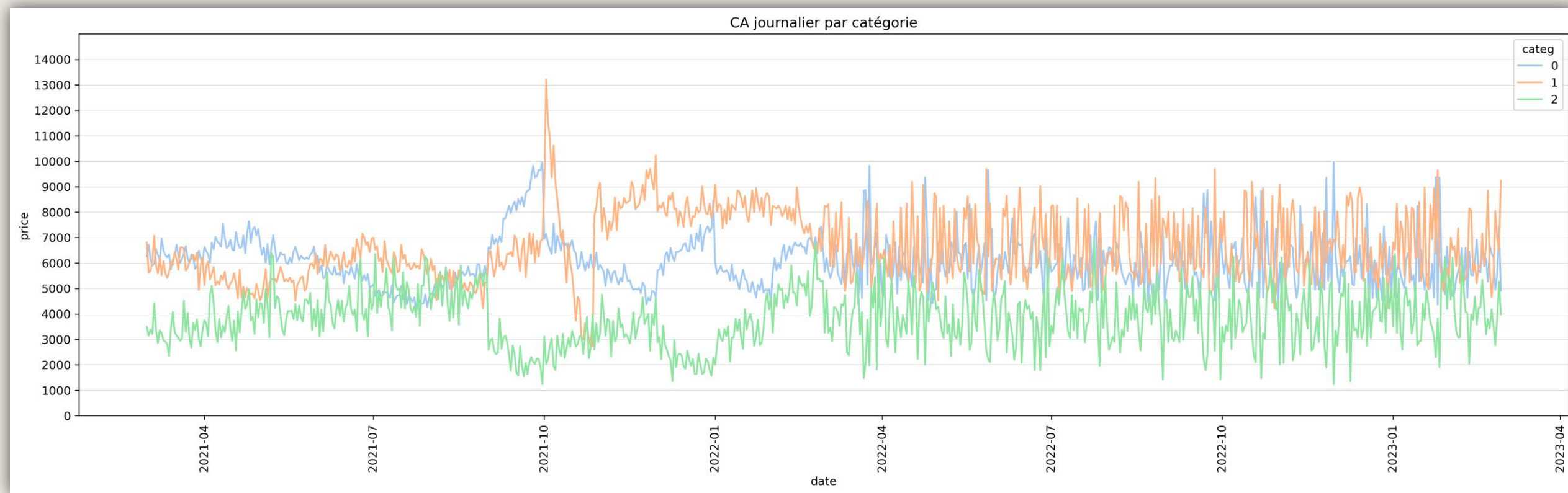
CA mensuel par catégorie



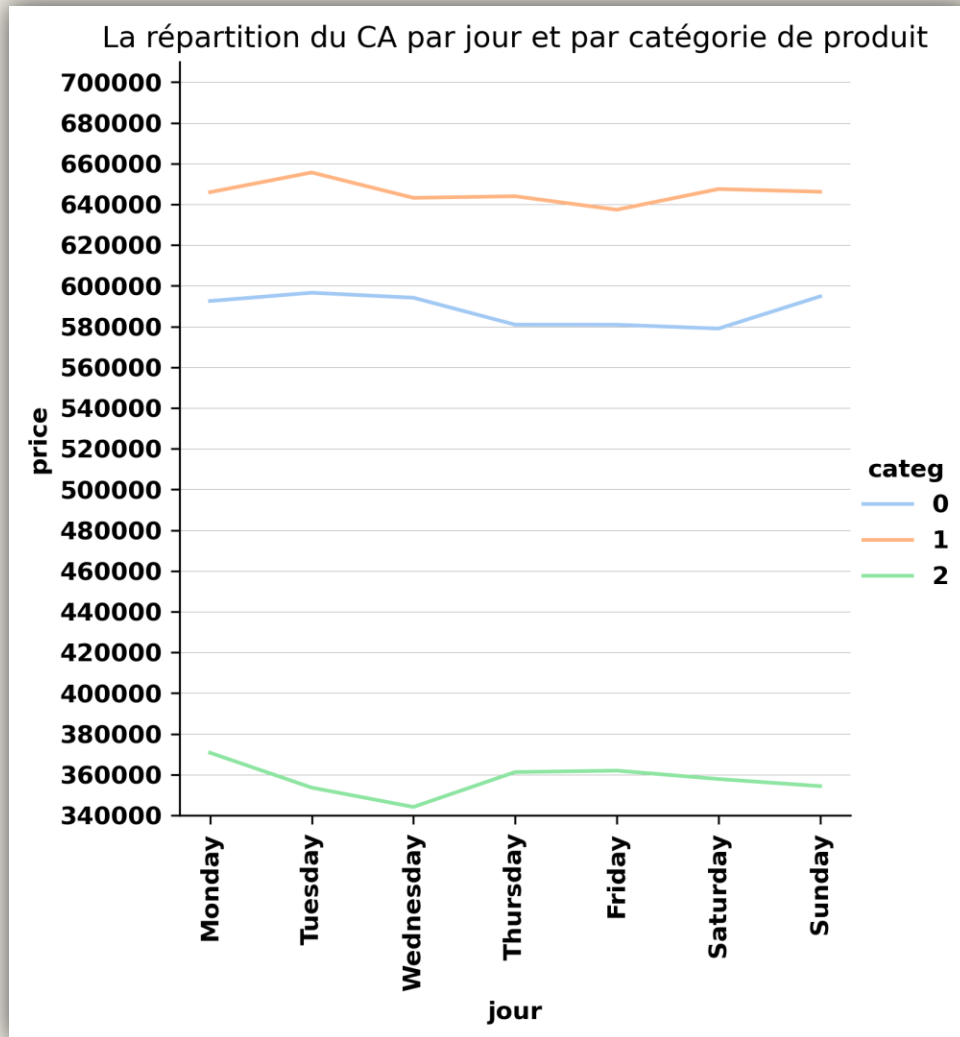
CA mensuel par catégorie de produits

- Pics et creux selon les périodes et selon les catégories

Evolution du CA quotidien par catégories



Les ventes par catégories sont stabilisées depuis mars 2022



Evolution du CA par nom de jour et par catégorie

Chaque catégorie
connait une période plus
faible sur une semaine de
ventes

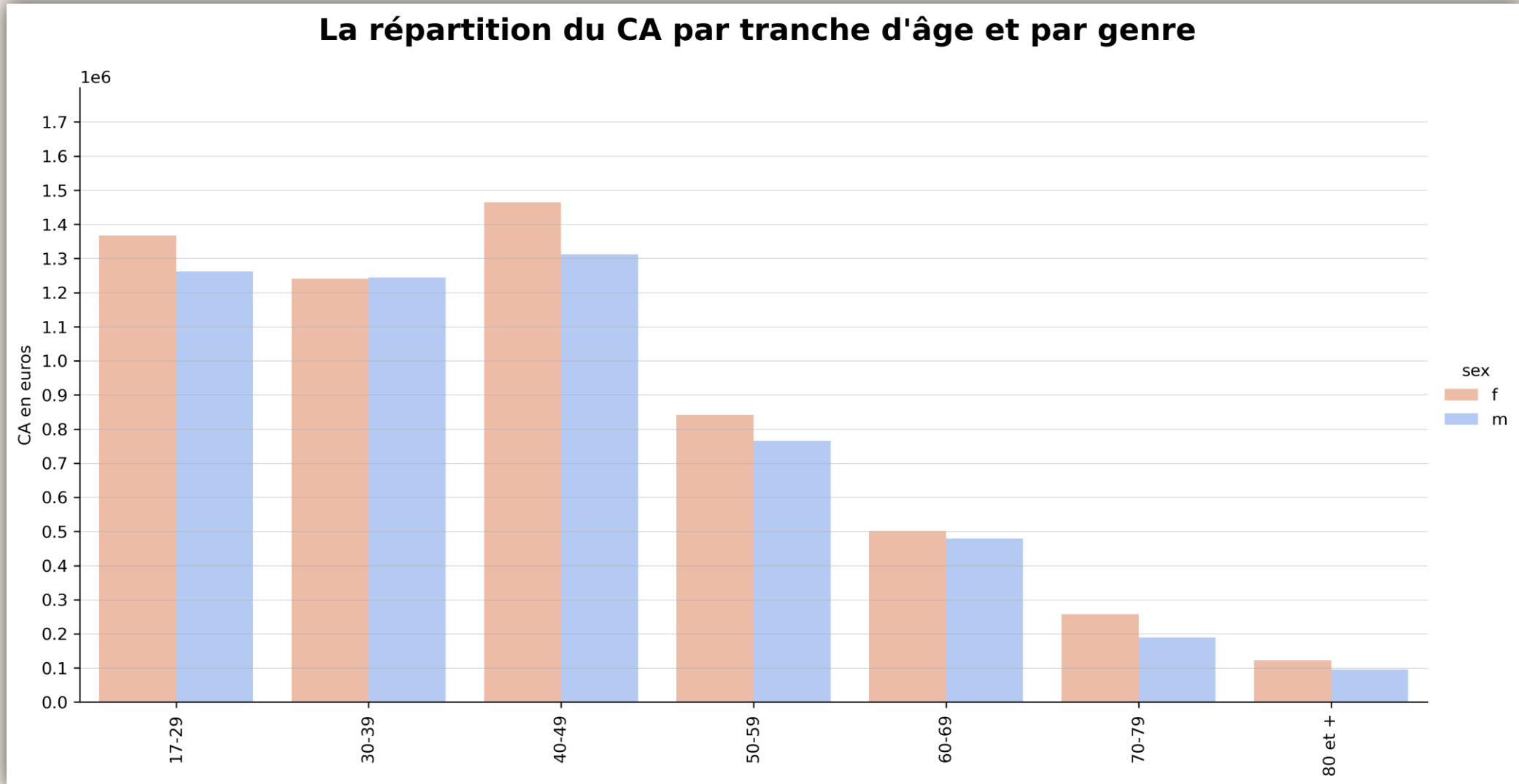
TOP CLIENTS

client_id	total_sales	ca_pourcent
c_1609	326039.89	2.71
c_4958	290227.03	2.41
c_6714	153918.60	1.28
c_3454	114110.57	0.95
c_1570	5285.82	0.04
c_3263	5276.87	0.04
c_2140	5260.18	0.04
c_2899	5214.05	0.04
c_7319	5155.77	0.04
c_7959	5135.75	0.04

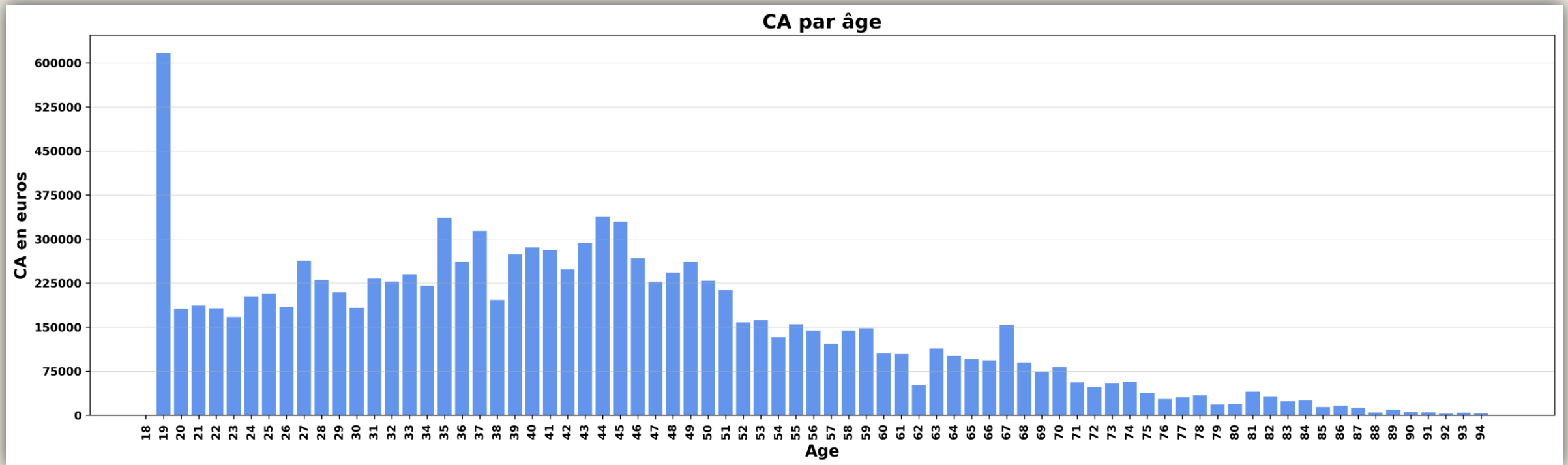


Clients a écarter –
analyse complémentaire ultérieure sur
ces clients

Répartition du CA par tranche d'âge et par genre



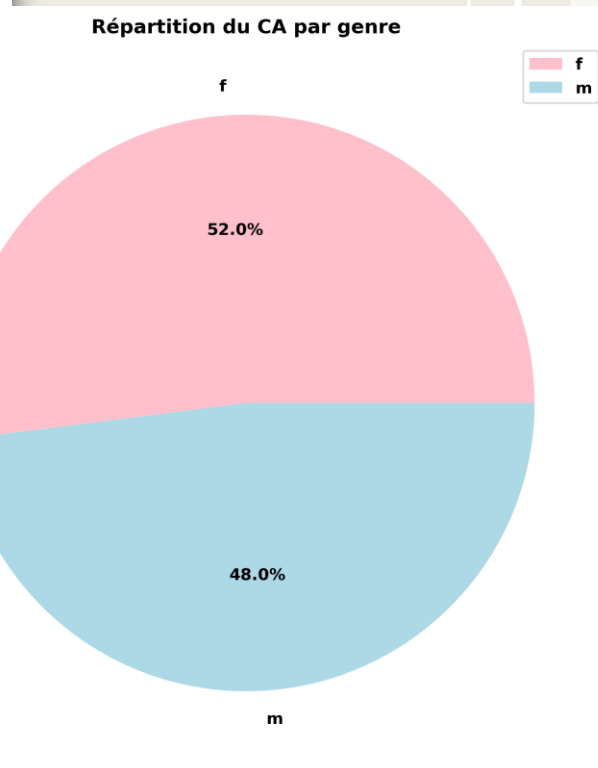
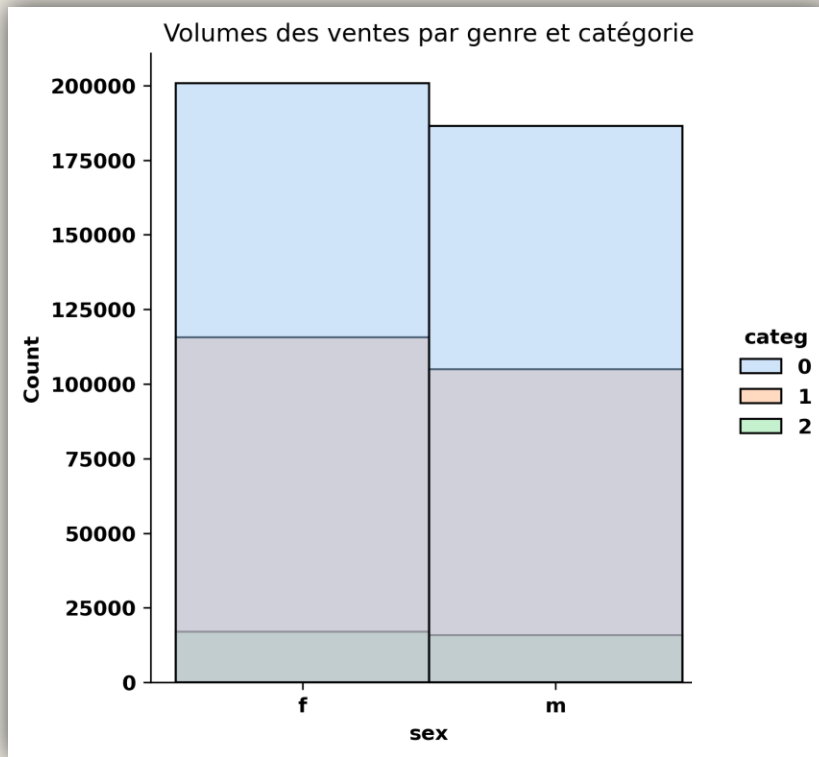
Répartition du CA par âge



437 clients ont une année de naissance renseignée de 2004.

On peut penser qu'il s'agit d'une erreur de la part du client.

On ne cherchera pas à modifier ces données qui semblent être erronées.

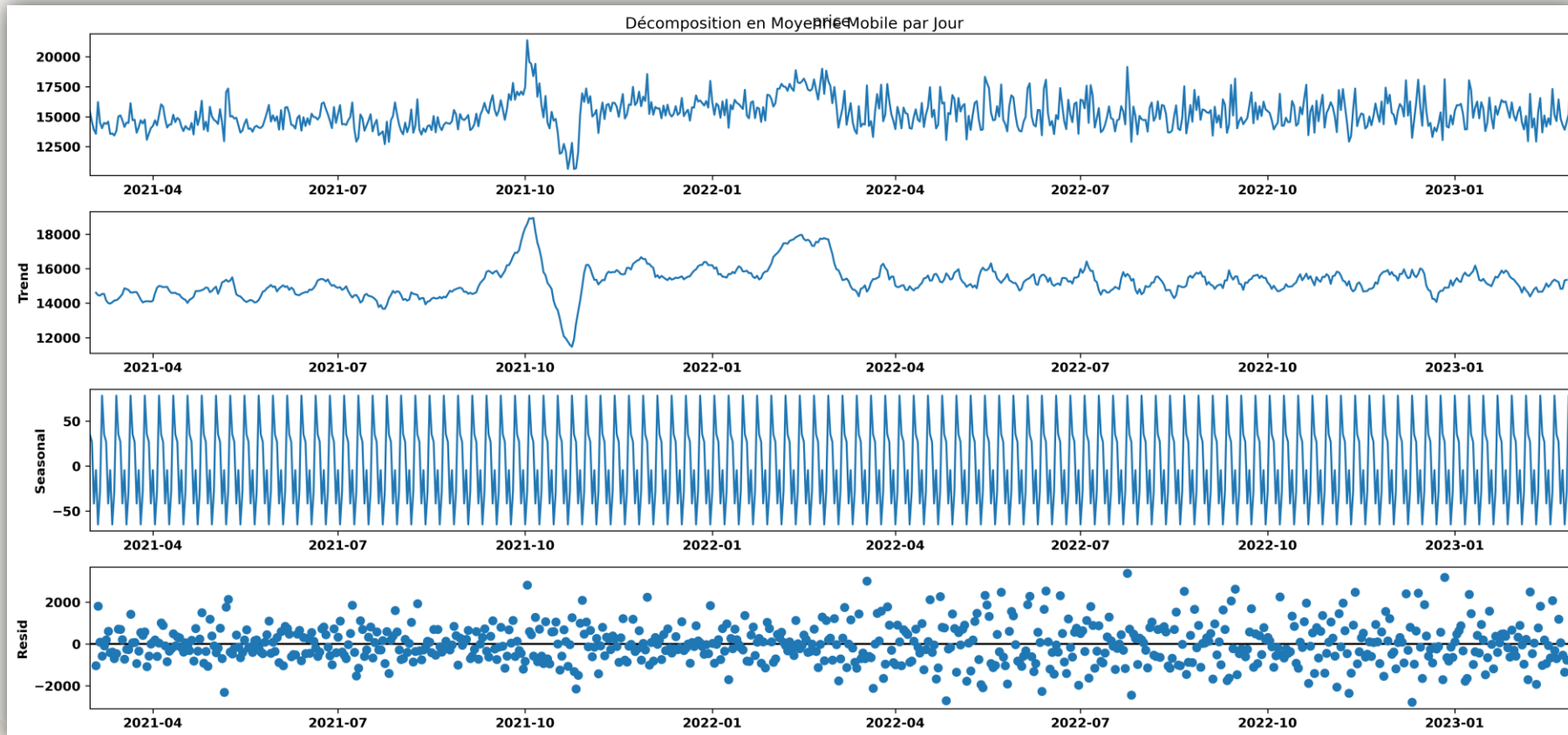


Répartition du CA et du volume des ventes

Répartition clients par genre :

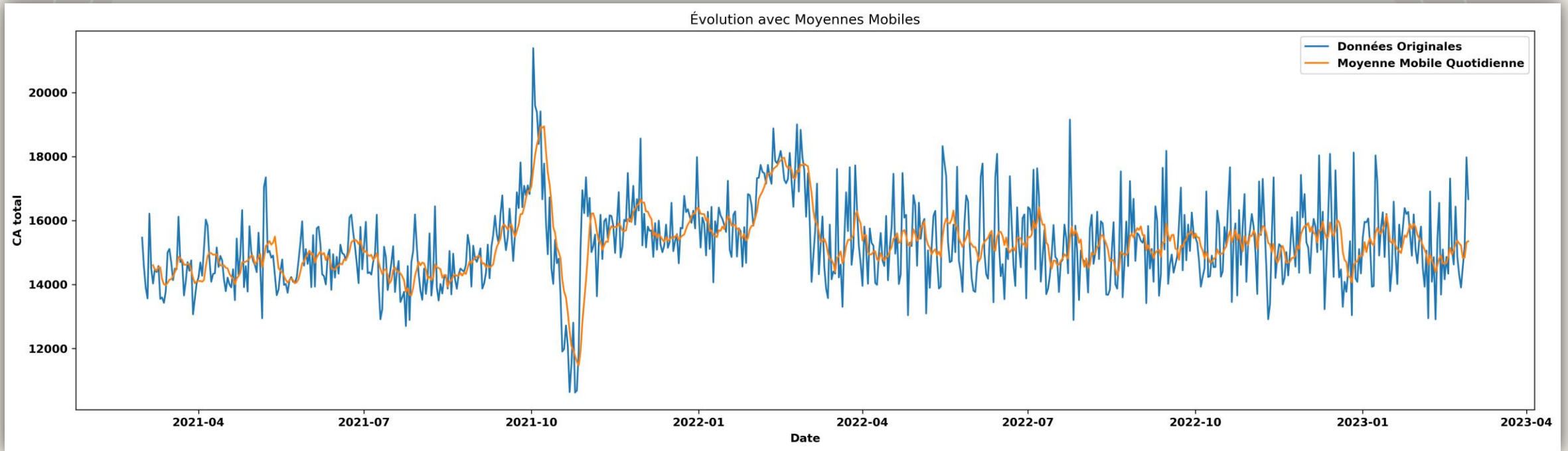
- 4478 femmes
- 4118 hommes

Décomposition en moyenne mobile



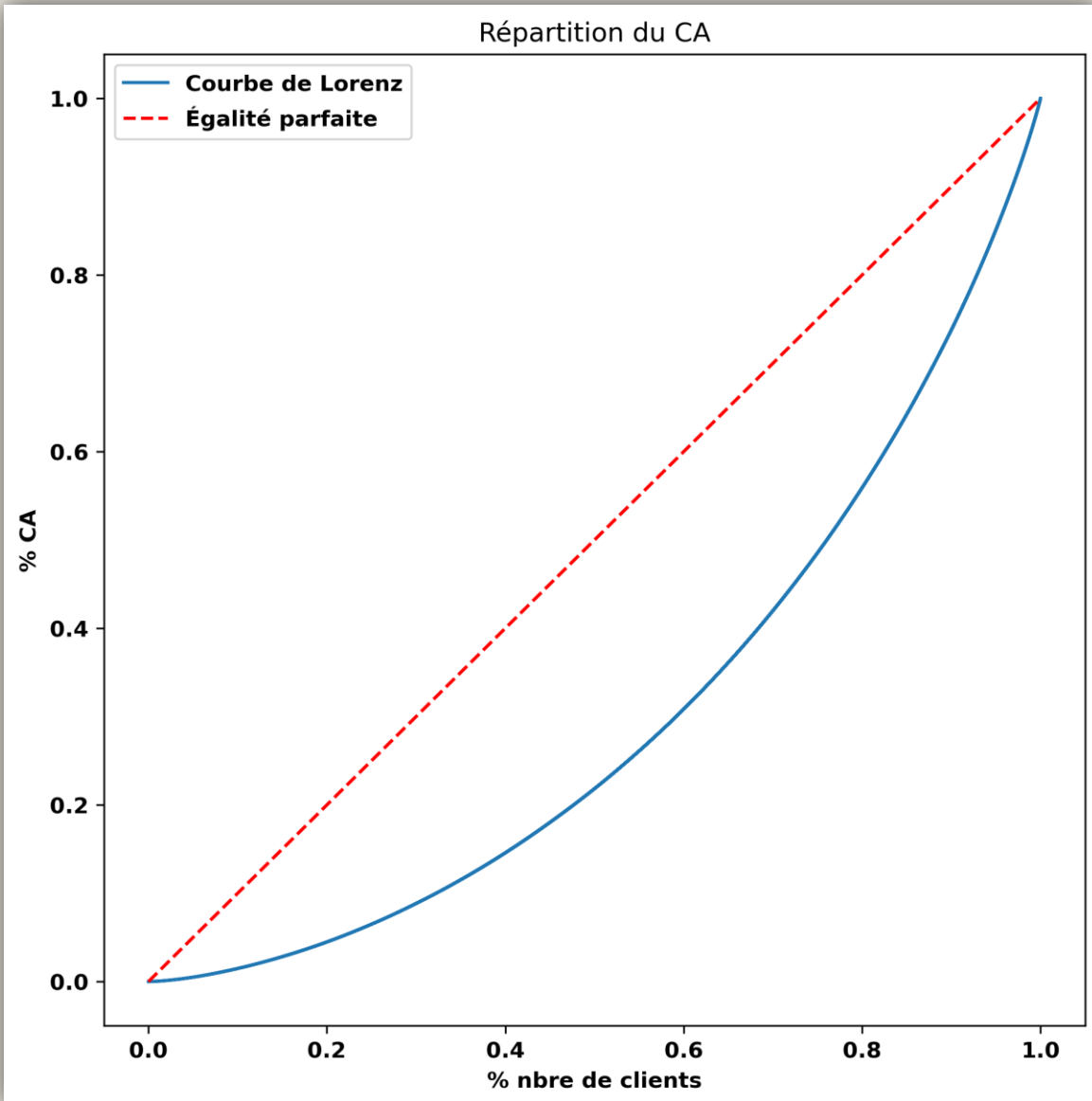
Saisonnalité de 7 jours
 $CA = \text{Tendance} + \text{Saisonnalité} + \text{Résidus}$

Evolution du CA journalier avec moyenne mobile



Moyenne mobile : Lisser une série de valeurs exprimées en fonction du temps. Elle permet d'éliminer les fluctuations les moins significatives. (source INSEE)

Inégalités de la répartition du CA entre les clients

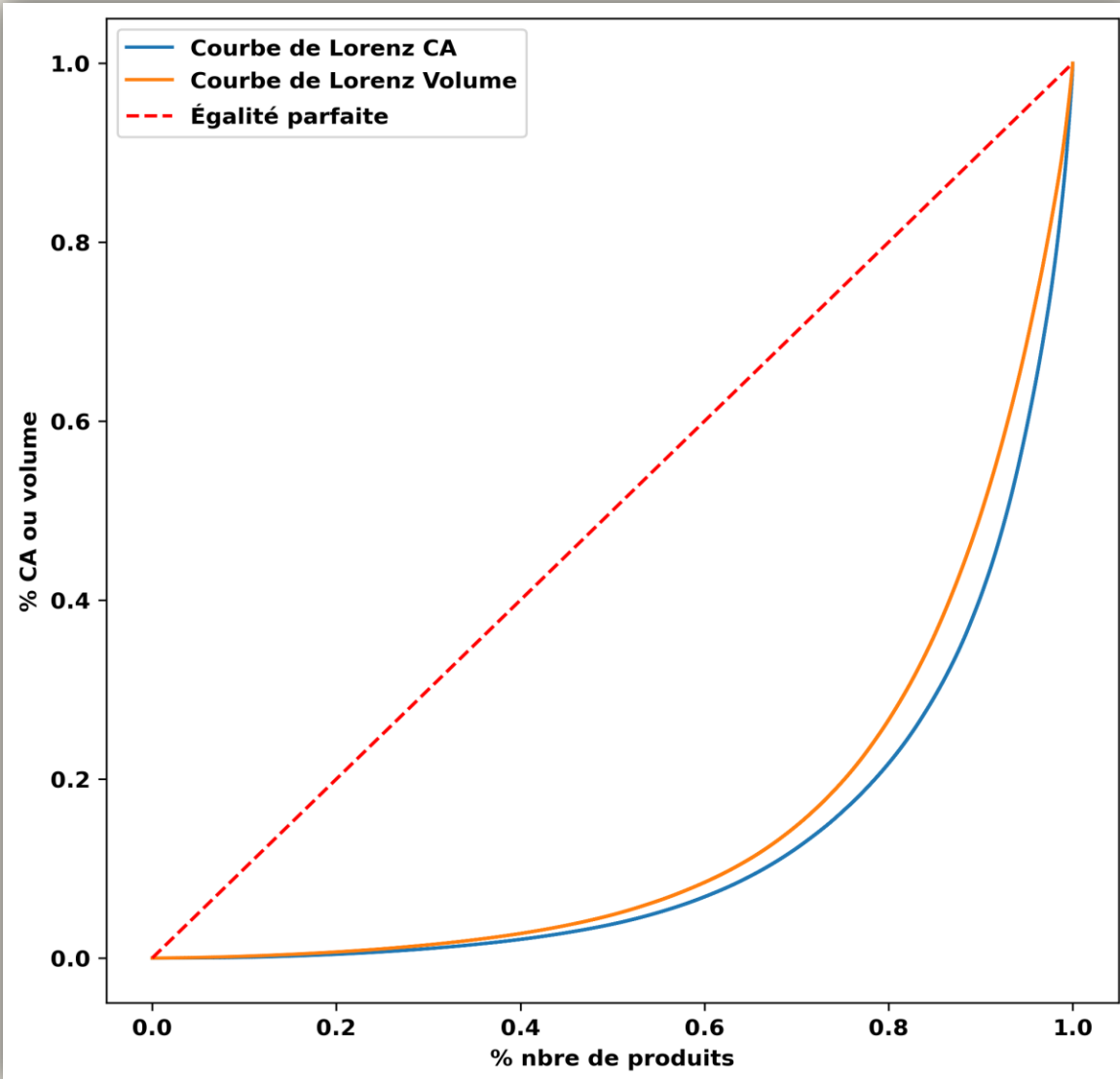


L'indice de Gini est un indicateur synthétique permettant de rendre compte du niveau d'inégalité pour une variable et sur une population donnée. Il varie entre 0 (égalité parfaite) et 1 (inégalité extrême).

Source : INSEE

Indice de Gini : 0.39

Répartition du CA et du volume par rapport de l'offre



80% des produits font
environ 20% du CA ou du
volume vendu.

Cela démontre une certaine
corrélation entre volume et
valeur.

id_prod	total_price	moy_price	nbre_ventes	nbre_clients
2_159	91097.76	145.99	624	526
2_135	63470.80	68.99	920	400
2_112	58785.90	67.57	870	380
2_102	55650.74	59.14	941	400
1_369	52897.95	23.99	2205	994
1_395	52182.00	28.99	1800	829
2_209	50742.75	69.99	725	314
1_414	50376.62	23.83	2114	941
1_383	49746.84	28.99	1716	808
2_166	48308.40	230.04	210	97

id_prod	total_price	moy_price	nbre_ventes	nbre_clients
1_369	52897.95	23.99	2205	994
1_417	44561.77	20.99	2123	983
1_414	50376.62	23.83	2114	941
1_498	47721.54	23.37	2042	950
1_425	34421.74	16.99	2026	923
1_413	34810.65	17.99	1935	915
1_412	31884.75	16.65	1915	877
1_407	30109.17	15.99	1883	872
1_396	34875.00	18.60	1875	877
1_403	33713.26	17.99	1874	854

Top 10 articles par volume des ventes et par CA

Les articles de la **catégorie 2** sont en tête en terme de **CA** tandis que ce sont les articles de la **catégorie 1** qui sont le plus vendus en terme de **volume**

id_prod	total_price	moy_price	nbre_ventes	nbre_clients
0_1539	0.99	0.99	1	1
0_1191	0.99	0.99	1	1
0_1284	1.38	1.38	1	1
0_1726	1.57	1.57	1	1
0_1653	1.98	0.99	2	1
0_807	1.99	1.99	1	1
0_541	1.99	1.99	1	1
0_1601	1.99	1.99	1	1
0_1871	1.99	1.99	1	1
0_1728	2.27	2.27	1	1

Flop 10 articles par CA

Réduire l'offre de la catégorie 0 ? Voir le coût de stockage ?

	jour	categ	total_price	moy_price	nbre_produits	nbre_ventes	nbre_clients
Wednesday	Friday	0	580937.99	10.661565	2044	54489	6220
	Monday	0	592582.45	10.625851	2057	55768	6316
	Tuesday	0	596636.04	10.647750	2054	56034	6331
	Saturday	0	579048.80	10.606649	2038	54593	6275
	Thursday	0	580967.34	10.629525	2064	54656	6355
	Sunday	0	594848.27	10.643387	2039	55889	6348
	Tuesday	1	655685.98	20.511339	704	31967	7005
Wednesday	Thursday	1	644010.20	20.483133	710	31441	6992
	Sunday	1	646223.47	20.439761	701	31616	6920
	Saturday	1	647542.93	20.544526	699	31519	6940
	Monday	1	646006.62	20.464619	694	31567	6984
	Friday	1	637402.41	20.489984	702	31108	6811
	Sunday	2	354375.54	77.021417	203	4601	1866
	Monday	2	370772.71	76.196611	208	4866	1876
Wednesday	Thursday	2	361279.55	76.010846	209	4753	1932
	Friday	2	361986.46	75.856341	210	4772	1912
	Tuesday	2	353636.54	75.773846	204	4667	1880
	Saturday	2	357859.00	76.449263	209	4681	1896
	Sunday	2	344154.66	76.343092	211	4508	1891

CA, volume des ventes et ventes articles par nom de jour et par catégorie

Fluctuations non significatives, l'ensemble des chiffres est plutôt stable sur la semaine



ETAPE 3

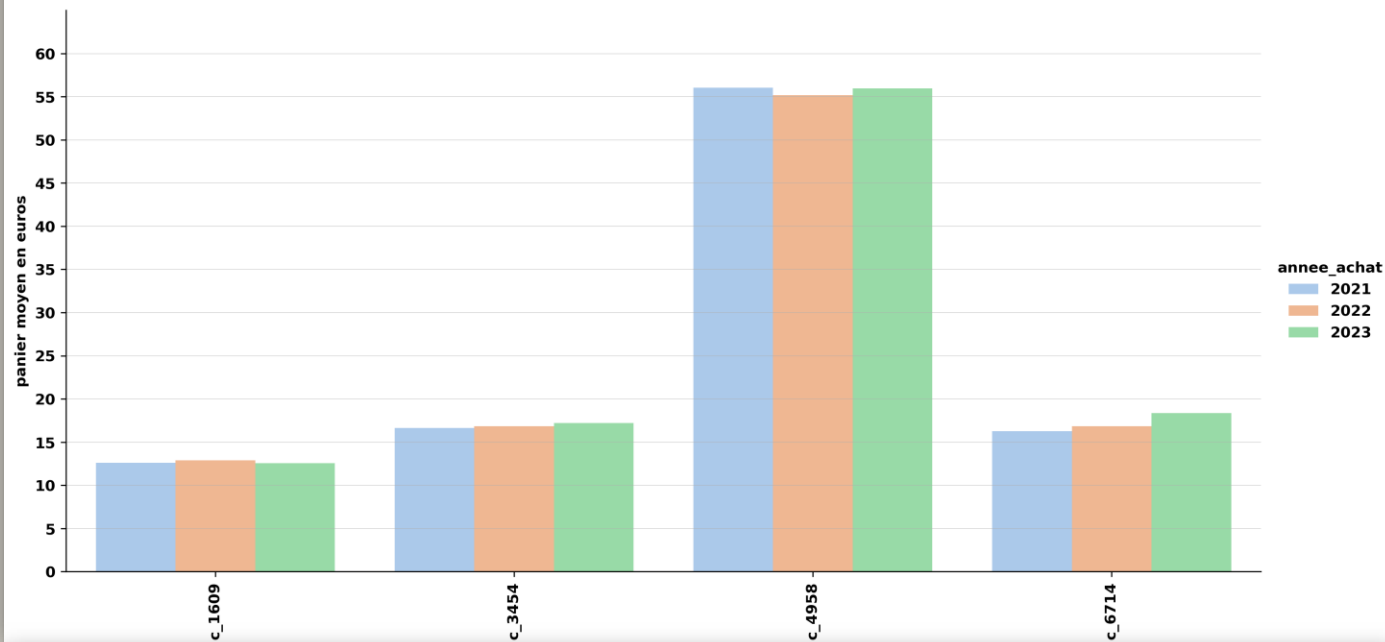
TOP CLIENTS

Quelques chiffres

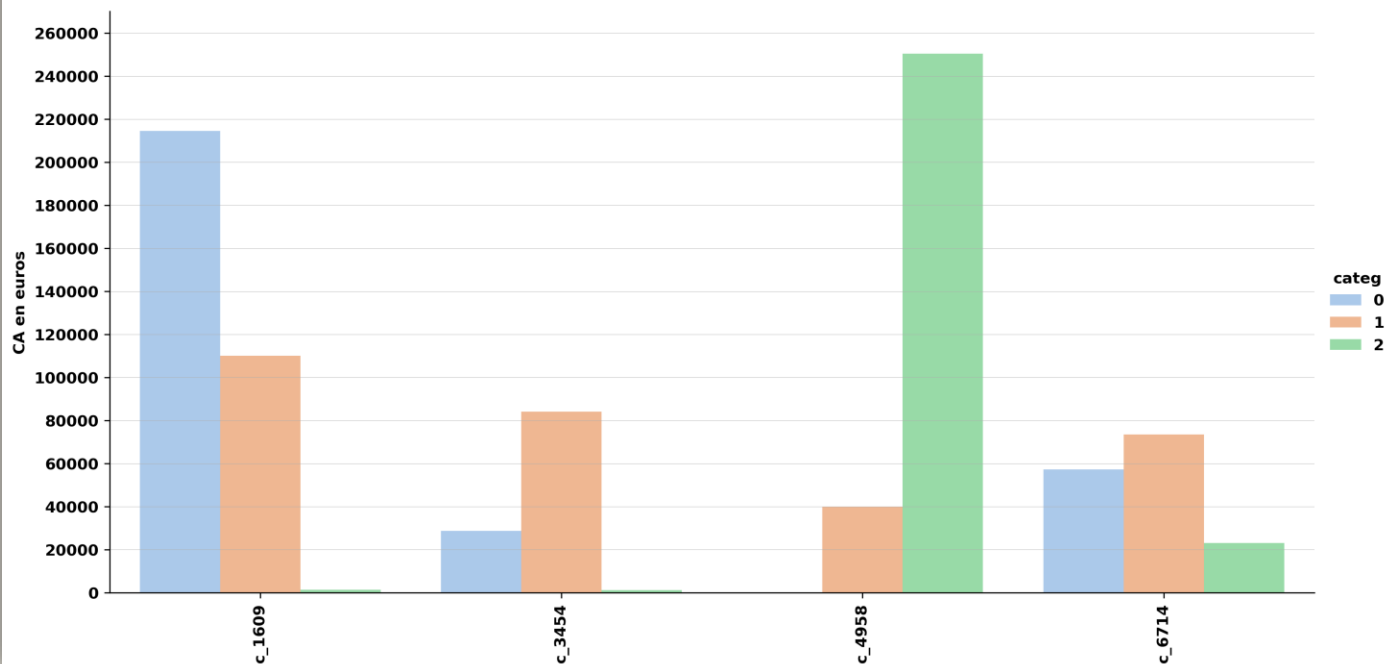
Ces clients représentent **7.35% du CA** depuis le lancement

client_id	sex	age	total_price	moy_price	nbre_produits	nbre_ventes	nbre_visites	frequence_achat	panier_moyen
c_1609	m	43	326039.89	12.742902	1922	25586	10997	453.0	12.608559
c_4958	m	24	290227.03	55.577754	470	5222	3851	87.0	56.028405
c_6714	f	55	153918.60	16.732101	1381	9199	2620	157.0	16.269453
c_3454	m	54	114110.57	16.798259	1093	6793	5571	111.0	16.637134

Panier moyen par top clients et par an



CA top clients par catégories de produits



Informations
supplémentaires top
clients



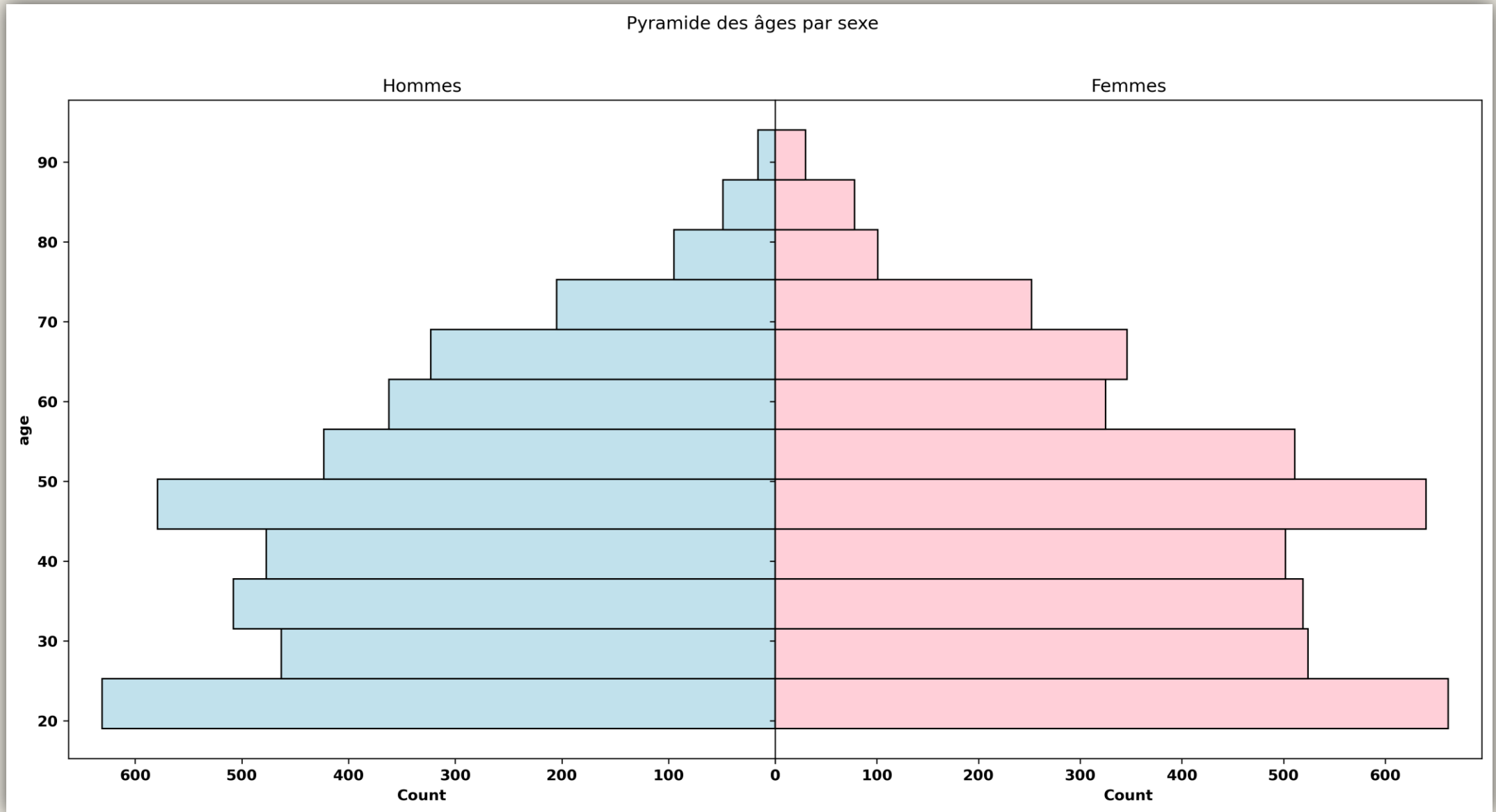
ETAPE 4

PROFILS CLIENTS

Informations générales et statistiques sur nos clients

	client_id	age	total_price	moy_price	nbre_produits	nbre_ventes	nbre_clients	nbre_visites	panier_moyen	frequence_achat
count	8596	8596.000000	8596.000000	8596.000000	8596.000000	8596.000000	8596.0	8596.000000	8596.000000	8596.000000
unique	8596	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	c_1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	44.739646	1296.343300	21.926199	36.046417	74.538623	1.0	37.513495	21.926199	3.100745
std	NaN	16.910542	958.404544	14.634385	31.618068	68.074309	0.0	30.898817	14.634385	2.860226
min	NaN	19.000000	6.310000	4.150000	1.000000	1.000000	1.0	1.000000	4.150000	0.000000
25%	NaN	31.000000	562.645000	13.472771	14.000000	28.000000	1.0	15.000000	13.472771	1.000000
50%	NaN	44.000000	1045.585000	15.928454	26.000000	52.000000	1.0	27.000000	15.928454	2.000000
75%	NaN	57.000000	1796.015000	21.961065	47.000000	97.000000	1.0	50.000000	21.961065	4.000000
max	NaN	94.000000	5285.820000	181.990000	180.000000	405.000000	1.0	167.000000	181.990000	17.000000

Pyramide des âges



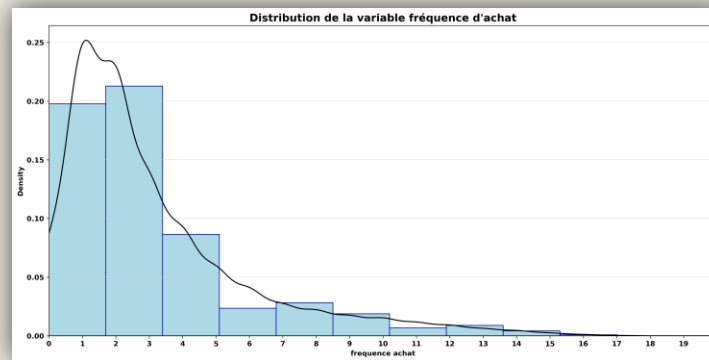


ETAPE 4

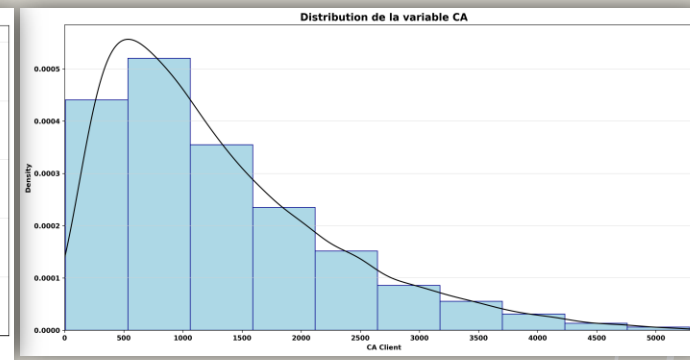
ANALYSE COMPORTEMENTALE

TESTS DE NORMALITE

Fréquence d'achat



Achats clients



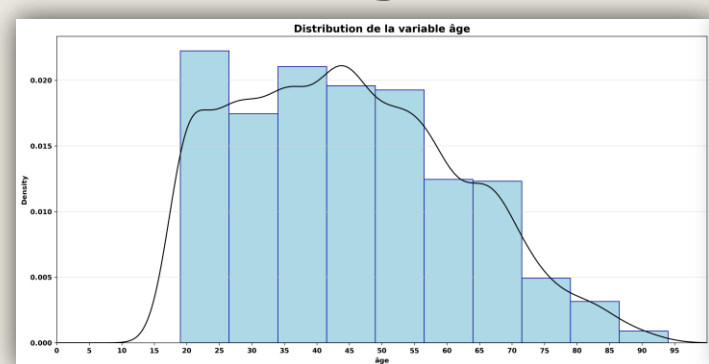
Tests de Kolmogorov et de Shapiro

Les hypothèses émises sont les suivantes :

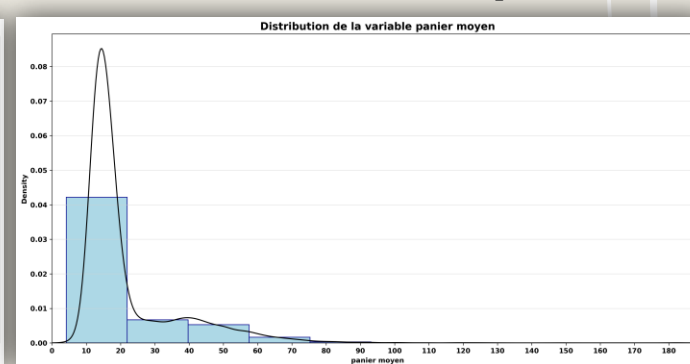
H0 : la variable suit la loi normale

H1 : la variable ne suit pas la loi normale

Age



Panier moyen



Résultat :

Les hypothèses nulles ont toutes été rejetées => **aucune variable ne suit la loi normale**

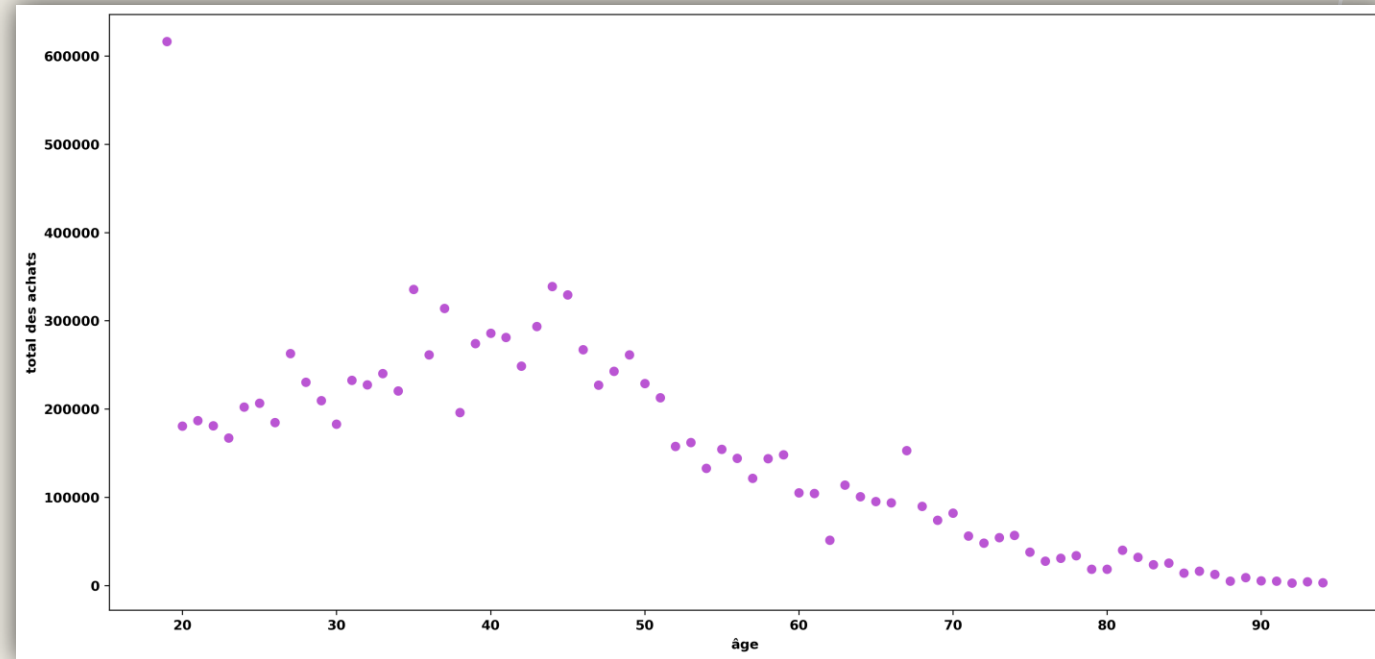
$\alpha = 5\%$

TESTS STATISTIQUES

Lien entre l'âge et le montant total des achats

Deux variables quantitatives

Tests utilisés : Spearman



$\alpha = 5\%$

Test paramétrique

Conditions de validité PEARSON :

- Chaque variable suit une distribution normale
- Au moins 20 individus dans l'échantillon

Les conditions de validité ne sont pas validées

Test non paramétrique

Hypothèses

H0 : Il n'y a pas de corrélation significative

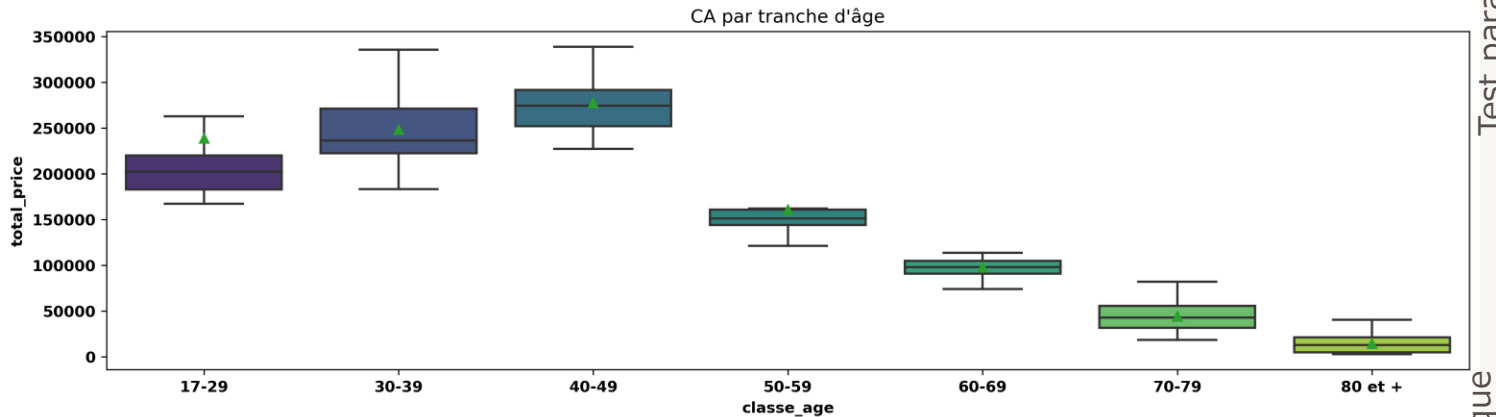
H1 : Il y a une corrélation significative

Résultat Spearman
L'hypothèse nulle est rejetée

Lien entre l'âge et le montant total des achats

(une variable qualitative, une variable quantitative)

Tests utilisés : ANOVA et Kruskal Wallis



Test paramétrique

Test non paramétrique

Hypothèses

H0 : Les moyennes de chaque groupe sont égales
H1 : Les moyennes de chaque groupe ne sont pas toutes égales

Résultat ANOVA

L'hypothèse nulle est rejetée

Conditions de validité :

- Les mesures sont indépendantes
- Les échantillons ont des variances égales
- Les résidus suivent une distribution normale
- Au moins 20 individus dans l'échantillon

Hypothèses

H0 : Les médianes de chaque groupe sont égales
H1 : Les médianes de chaque groupe ne sont pas toutes égales.

Résultat Kruskal

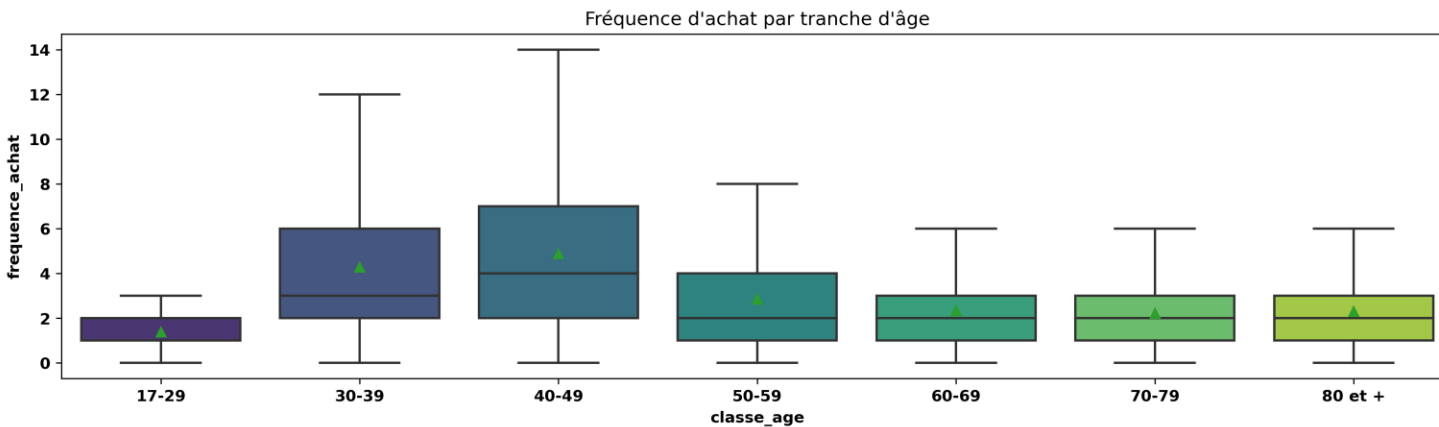
L'hypothèse nulle est rejetée

$\alpha = 5\%$

Lien entre l'âge des clients et la fréquence d'achat

(une variable qualitative, une variable quantitative)

Tests utilisés : Kruskal Wallis



Test paramétrique

Test non paramétrique

Conditions de validité ANOVA :

- Les mesures sont indépendantes
- Les échantillons ont des variances égales
- Les résidus suivent une distribution normale
- Au moins 20 individus dans l'échantillon

Les conditions de validité ne sont pas validées

Hypothèses

H0 : Les médianes de chaque groupe sont égales

H1 : Les médianes de chaque groupe ne sont pas toutes égales.

Résultat Kruskal

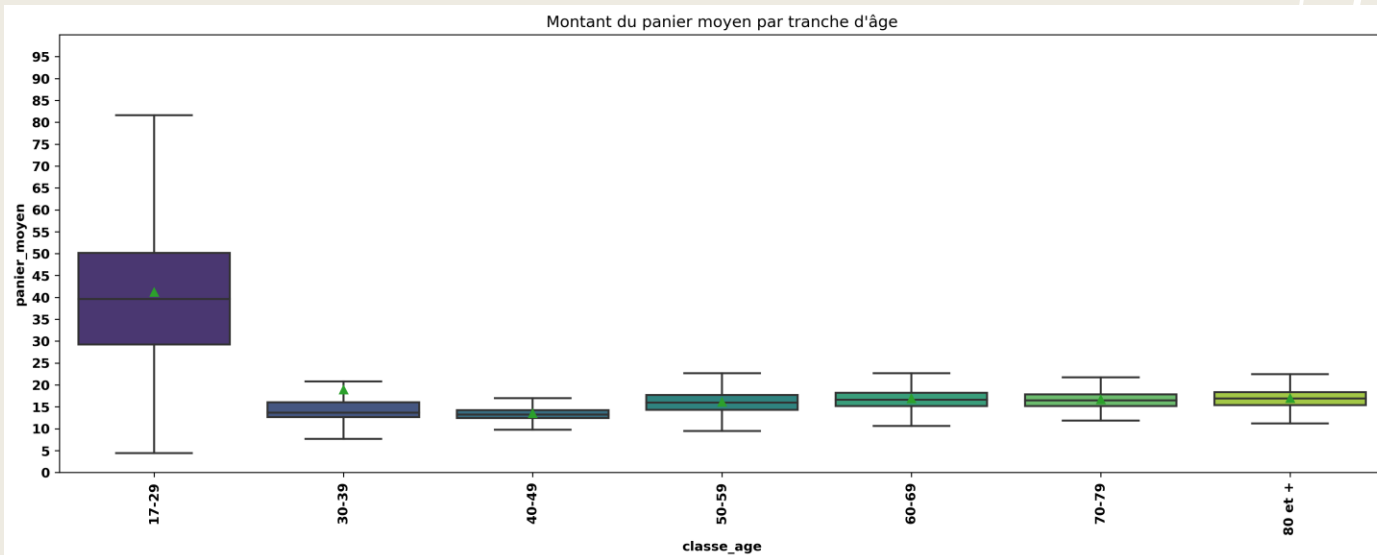
L'hypothèse nulle est rejetée

$\alpha = 5\%$

Lien entre l'âge des clients et panier moyen

(une variable qualitative, une variable quantitative)

Tests utilisés : Kruskal Wallis



$\alpha = 5\%$

Conditions de validité ANOVA :

- Les mesures sont indépendantes
- Les échantillons ont des variances égales
- Les résidus suivent une distribution normale
- Au moins 20 individus dans l'échantillon

Test paramétrique

Les conditions de validité ne sont pas validées

Hypothèses

H0 : Les médianes de chaque groupe sont égales
H1 : Les médianes de chaque groupe ne sont pas toutes égales.

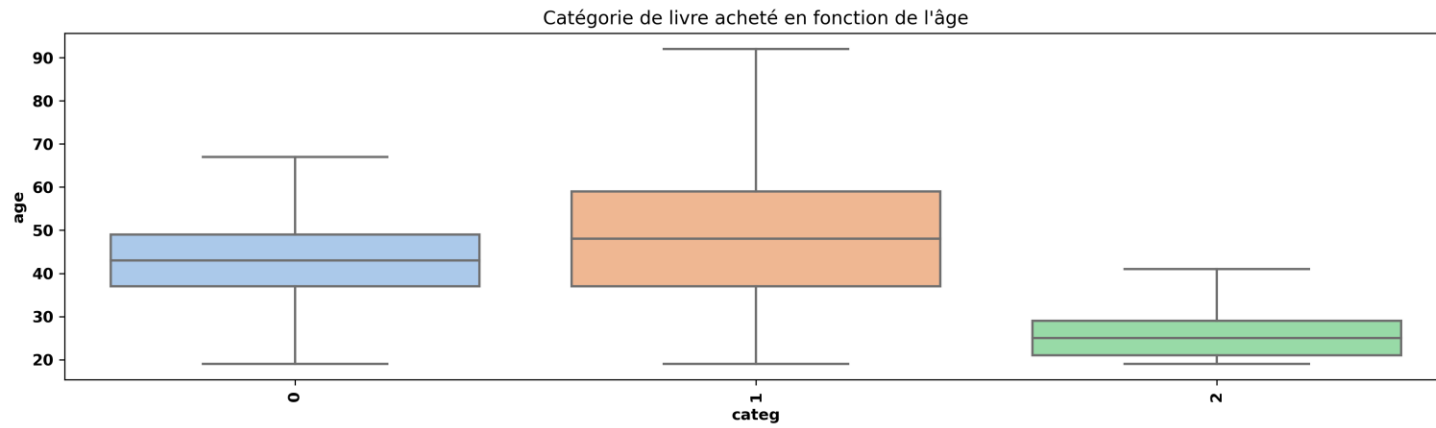
Test non paramétrique

Résultat Kruskal
L'hypothèse nulle est rejetée

Lien entre l'âge et la catégorie

(une variable qualitative, une variable quantitative)

Tests utilisés : Kruskal Wallis



$\alpha = 5\%$

Test paramétrique

Conditions de validité ANOVA :

- Les mesures sont indépendantes
- Les échantillons ont des variances égales
- Les résidus suivent une distribution normale
- Au moins 20 individus dans l'échantillon

Les conditions de validité ne sont pas validées

Test non paramétrique

Hypothèses

H0 : Les médianes de chaque groupe sont égales
H1 : Les médianes de chaque groupe ne sont pas toutes égales.

Résultat Kruskal
L'hypothèse nulle est rejetée

Lien entre genre et catégorie

(deux variables qualitatives)

	f_perc	m_perc	Total_perc
categ			
0	52.192093	47.807907	39.590649
1	51.969249	48.030751	42.414969
2	51.182856	48.817144	18.024481

sex	f	m	Total
categ			
0	4119	3773	7892
1	4394	4061	8455
2	1839	1754	3593

$\alpha = 5\%$

Test paramétrique

Hypothèses

H0 : il n'y a pas de lien entre catégorie de livre acheté et genre d'un client

H1 : il y a un lien entre catégorie de livre acheté et genre d'un client

Résultat Khi-2

L'hypothèse nulle n'est pas rejetée

Conditions de validité :

- Les mesures sont indépendantes
- Pas d'effectifs théoriques inférieurs à 5 dans les cases du tableau

Test non paramétrique

Les conditions de validité sont vérifiées, nous n'appliquons pas de test exact de Fisher

CONCLUSION

L'évolution mensuelle du CA sur ces deux années d'activité nous montre une courbe en dent de scie. Il serait peut être judicieux de mettre en place des actions commerciales sur "les mois creux" pour booster les ventes.

Au fil des ans, on constate aussi une augmentation de la valeur du panier moyen d'un top client, il faut développer ce client et le fidéliser.

Des actions ciblées sur les clients en fonction des corrélations étudiées pourraient aussi augmenter les ventes et fidéliser la clientèle.