



LIBRAIRIE EN LIGNE LAPAGE

Point sur l'activité



Lapage

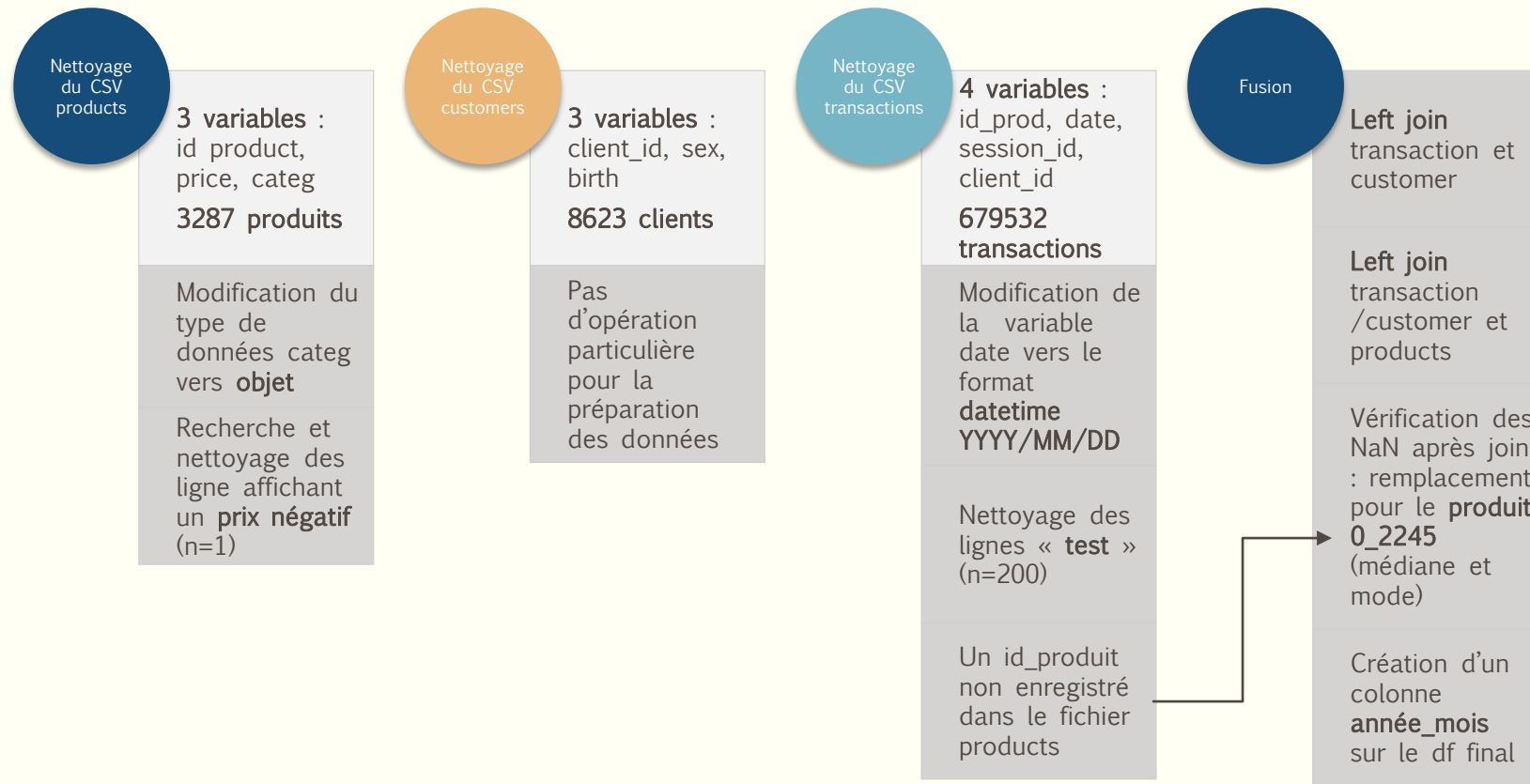
Ordre du jour

- Obtenir un jeu de données exploitable : étapes et méthodologie
- Analyse des principaux indicateurs de ventes
- Connaître nos clients : analyse des caractéristiques, des achats et recherche de corrélations
- Lien entre les références 0-525 et 2-159
- Conclusion



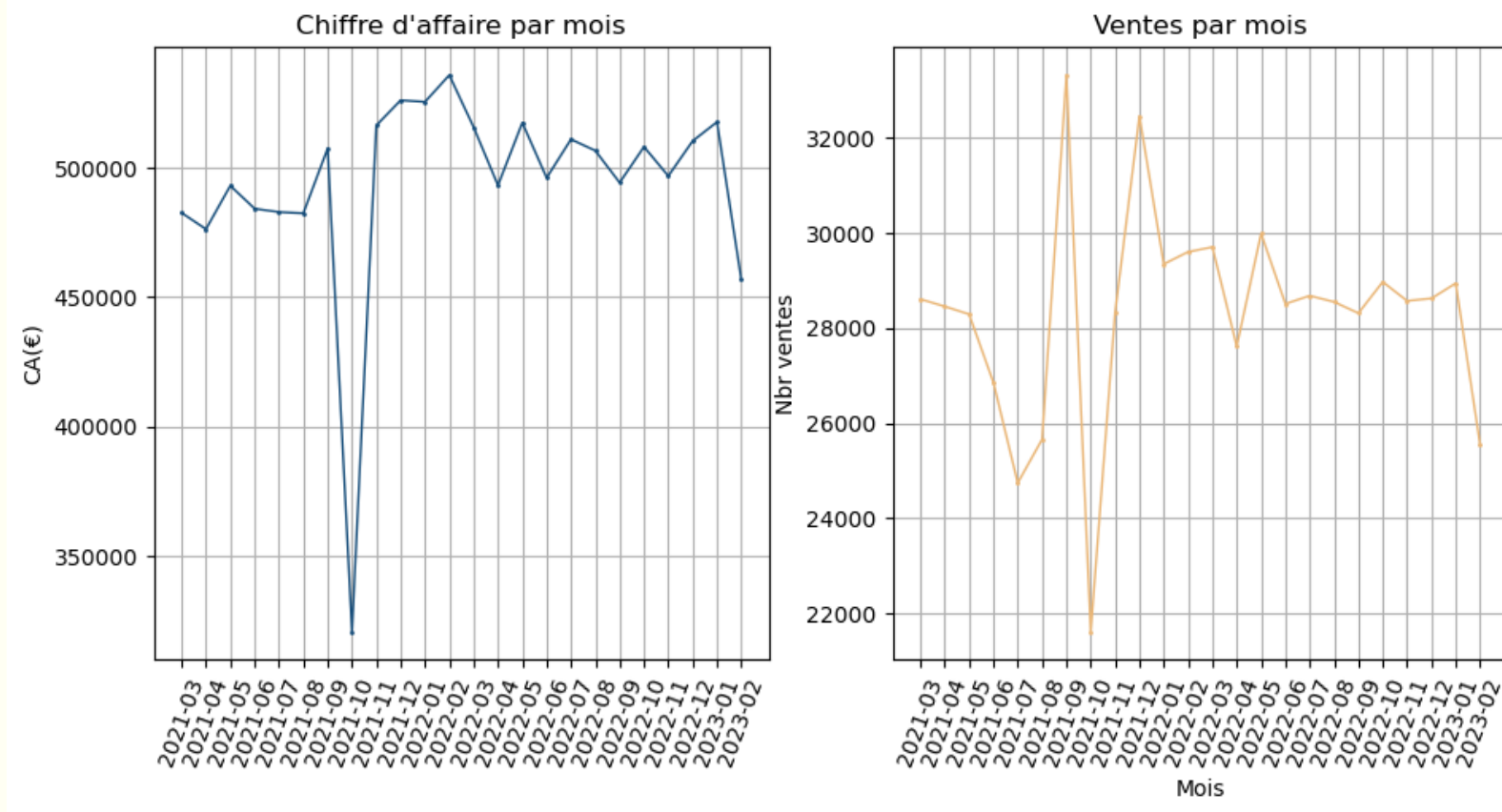
Obtenir un jeu de données exploitable

→ Pour chacun des trois fichiers : étude des **variables**, de la **dimension** du dataframe, des **types** de données, de l'**unicité des clés**, des **valeurs nulles**, des **doublons** et réalisation des modifications si besoin



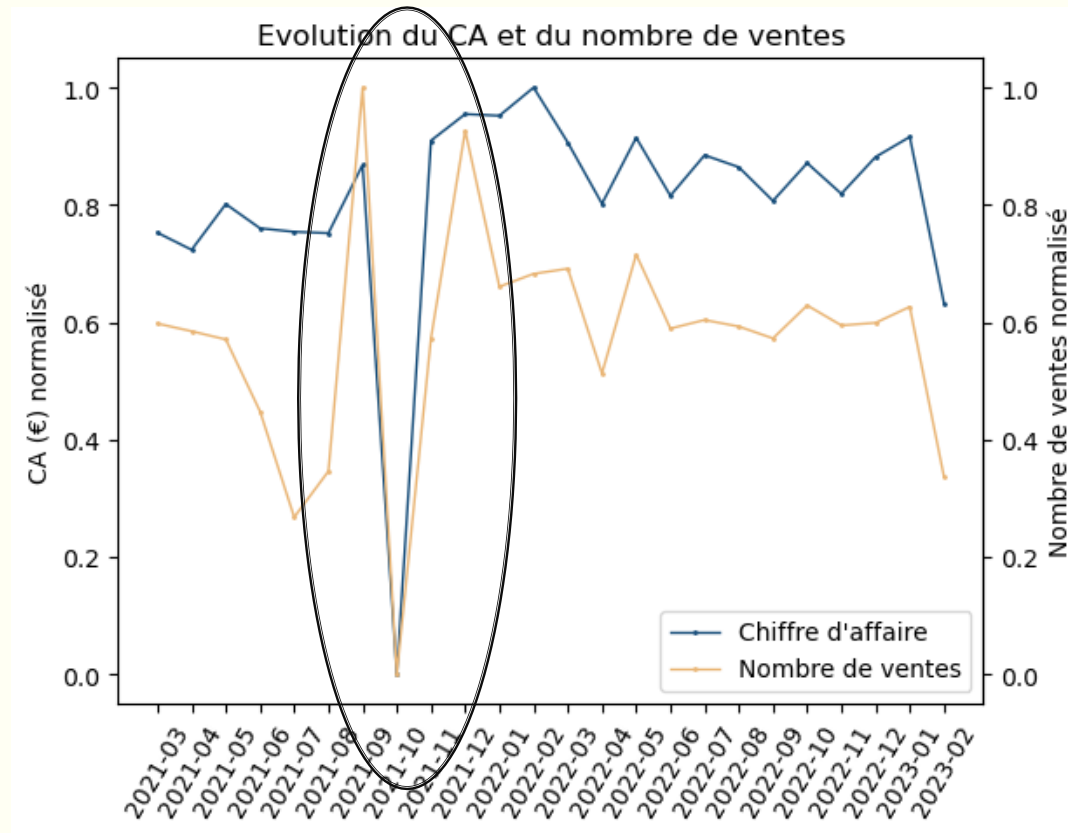
Analyse des principaux indicateurs de ventes : au global

- Le chiffre d'affaire généré depuis le 2021-03-01 est de **11856820.47€** pour un total de **679332 ventes**
- CA et nombre de ventes moyens **mensuels** : **494034.19 €** pour **28305.5 ventes**



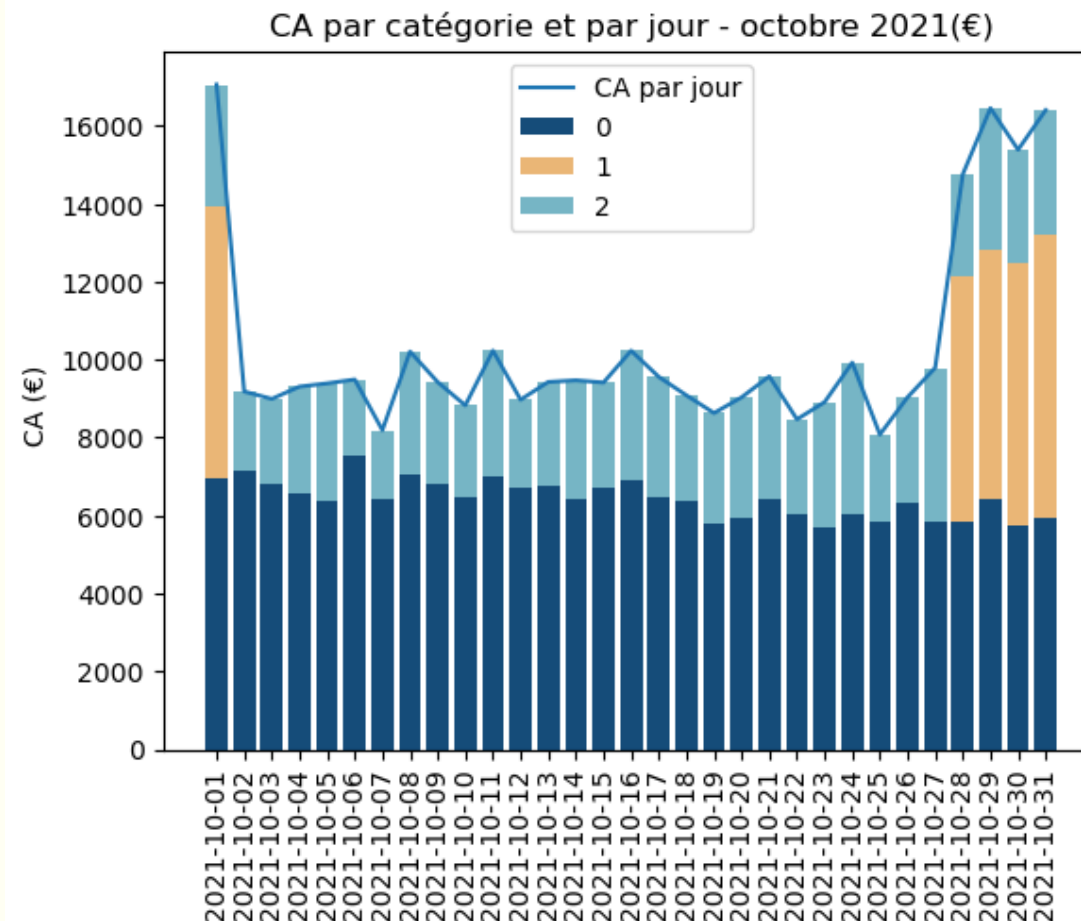
Analyse des principaux indicateurs de ventes : au global

- CA et nombre de ventes moyens mensuels : 494034.19 € pour 28305.5 ventes



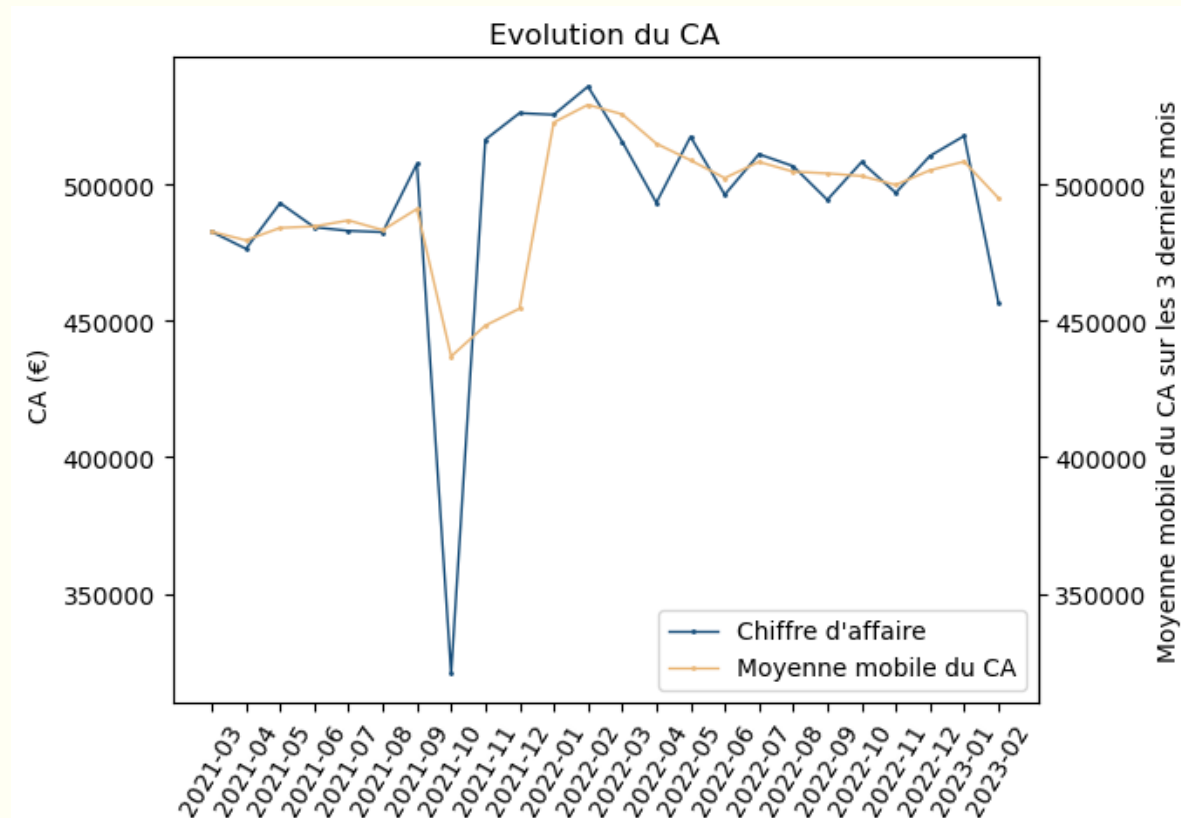
Analyse des principaux indicateurs de ventes : au global

- Le mois d'**octobre 2021** est marqué par une **chute** du CA généré par la **catégorie 1** (aucune vente entre le 02 et le 27 octobre sur cette catégorie)

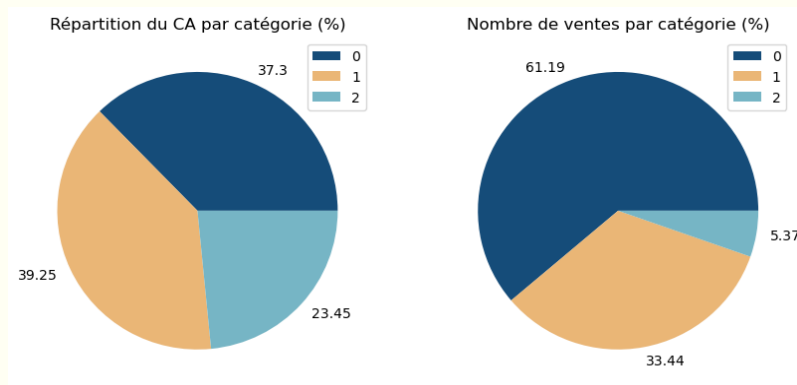
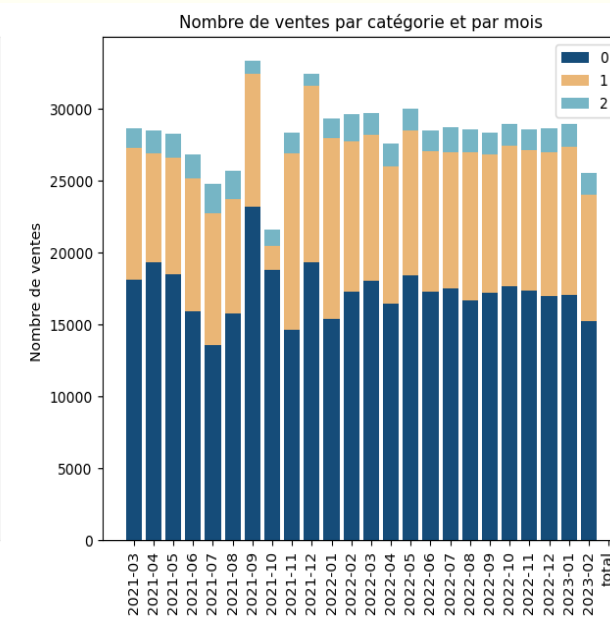
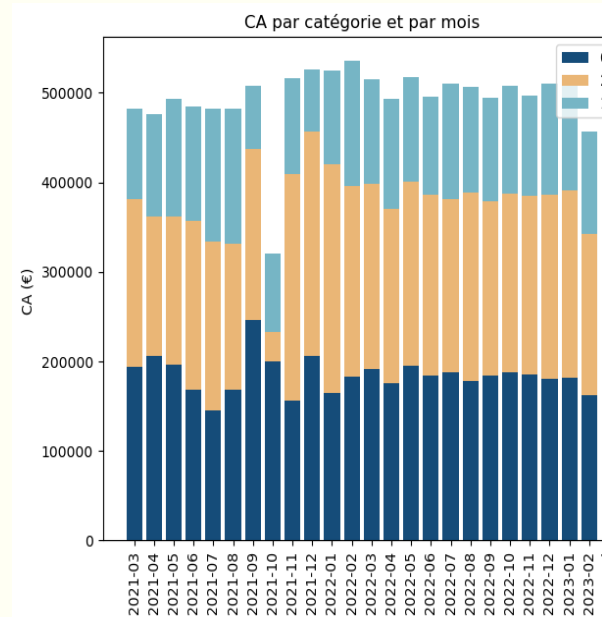
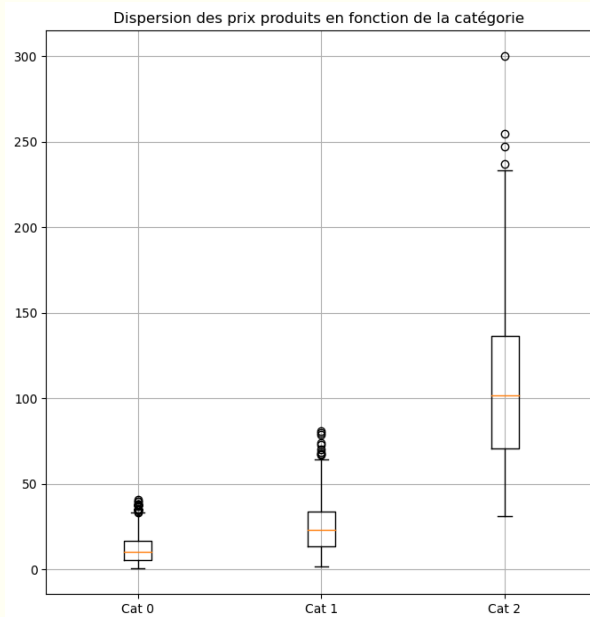


Analyse des principaux indicateurs de ventes : au global

- La **moyenne mobile** calculée sur les **trois derniers mois** montre une relative stabilité du CA mensuel aux alentours de 500000€, marqué tout de même par une période de croissance entre mars et septembre 2021, un chute liée à la perte de CA du mois d'octobre suivie d'un regain immédiat d'activité, puis une période de faible décroissance depuis un an.



Analyse des principaux indicateurs de ventes : par catégorie



- Catégorie 0 : prix les plus faibles, le plus grand nombre de ventes
- Catégorie 1 : prix moyens, catégorie la plus rémunératrice (CA)
- Catégorie 2 : prix les plus élevés, la plus rémunératrice rapportée au nombre de ventes
- Les proportions CA et nombre de ventes par catégorie sont relativement stables dans le temps (exceptions aux mois de septembre et octobre 2021)

Analyse des principaux indicateurs de ventes : par produit

- Un produit a généré, en moyenne, **3630€ de CA** pour **208 ventes** sur la période

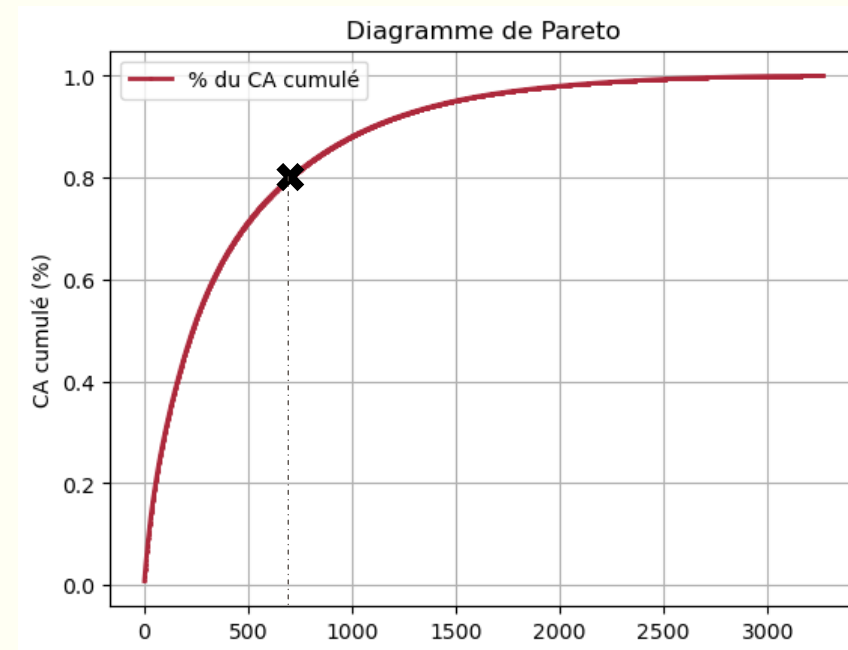
TOPS :

- **80 % du CA** est généré par **702 produits**, soit **22% des produits**
- Le produit le plus populaire a été vendu **2252 fois**, soit **0,3% des ventes**
- Le produit le plus rémunérateur a généré **94893.50€ de CA**, soit **0,8% du CA**

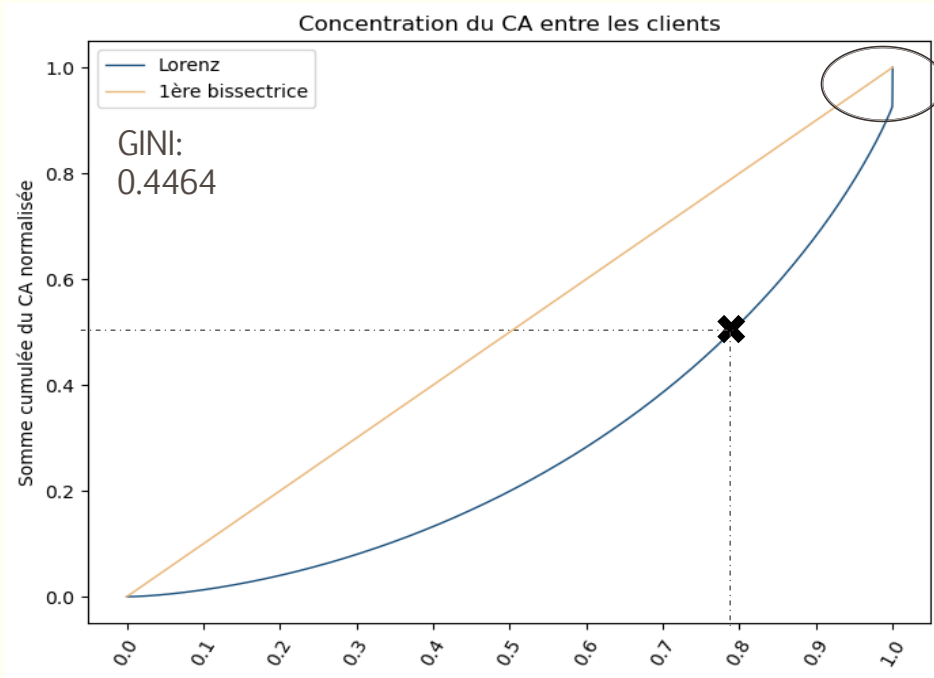
FLOPS :

- **21 produits** n'ont pas été vendus sur la période considérée
- **18 produits** n'ont été vendus qu'une fois

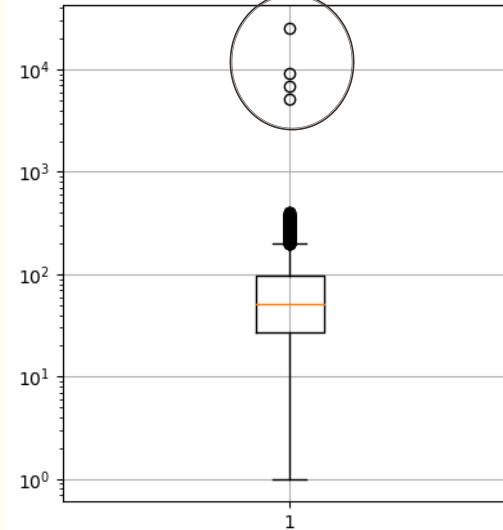
Top 10 des produits les plus vendus / les plus rémunérateurs			
id_prod	Nbr de ventes	id_prod	CA
1_369	2252	2_159	94893.50
1_417	2189	2_135	69334.95
1_414	2180	2_112	65407.76
1_498	2128	2_102	60736.78
1_425	2096	2_209	56971.86
1_403	1960	1_395	54356.25
1_412	1951	1_369	54025.48
1_413	1945	2_110	53846.25
1_406	1939	2_39	53060.85
1_407	1935	2_166	52449.12



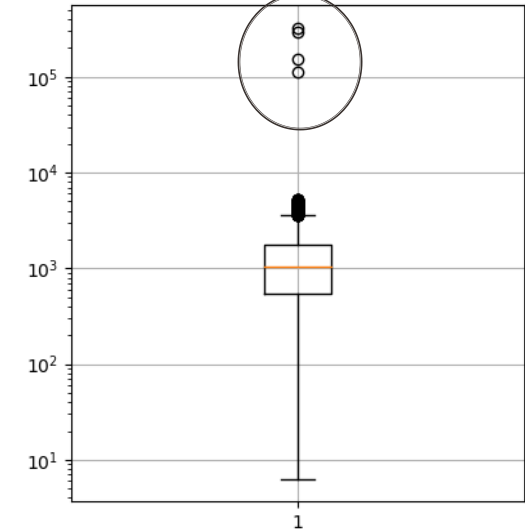
Analyse des principaux indicateurs de ventes : par clients



Dispersion du nombre d'articles achetés par client

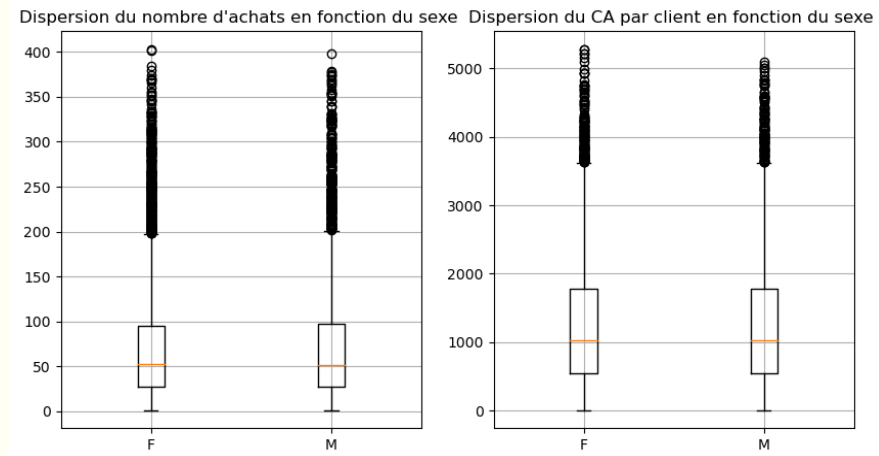
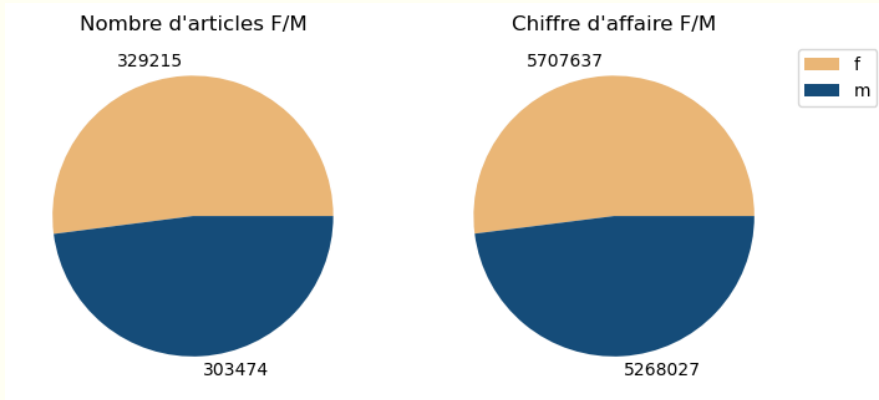


Dispersion du CA par client

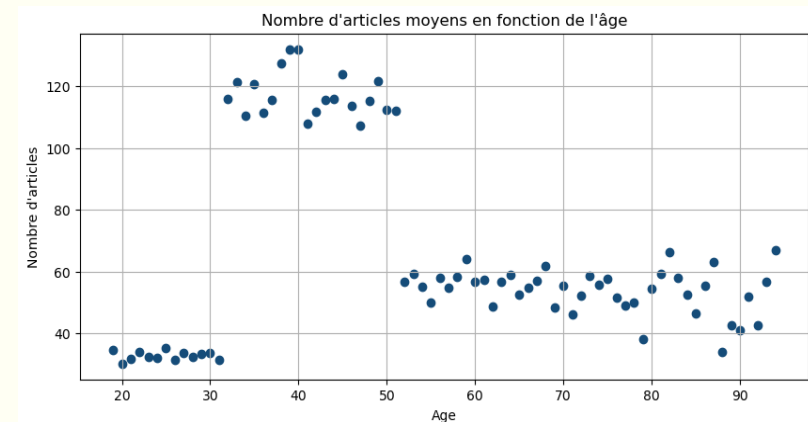
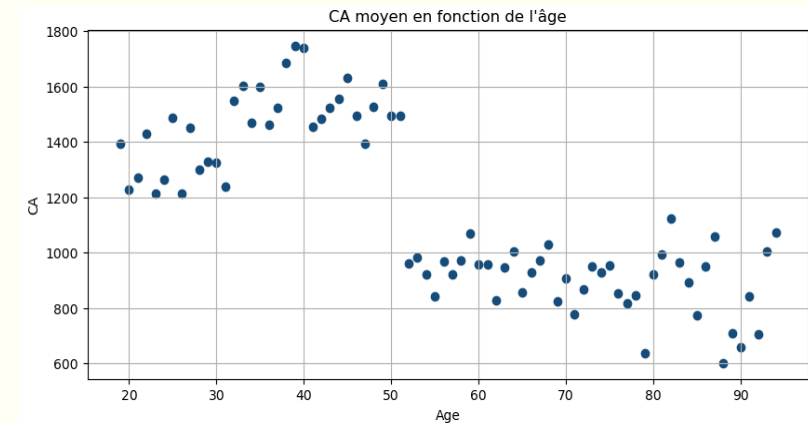


- Un client a généré en moyenne **1378.70€** de CA pour **79** ventes
- **50% du CA est concentré par 21% des plus gros clients**
- Parmi les clients les plus importants :
 - 4 clients ont généré un nombre beaucoup plus élevé de ventes et de CA que la moyenne : clients BtoB?
 - Ces quatre outliers ont été enlevés des données pour le reste des analyses

Analyse des principaux indicateurs de ventes : par clients



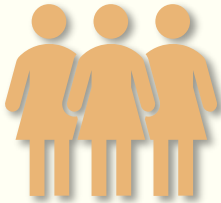
Les graphiques semblent montrer qu'il n'y a pas de différence entre les dépenses effectuées par les hommes et celles effectuées par les femmes



En revanche qu'il semble exister des différences en fonction de l'âge

Qui sont nos clients ?

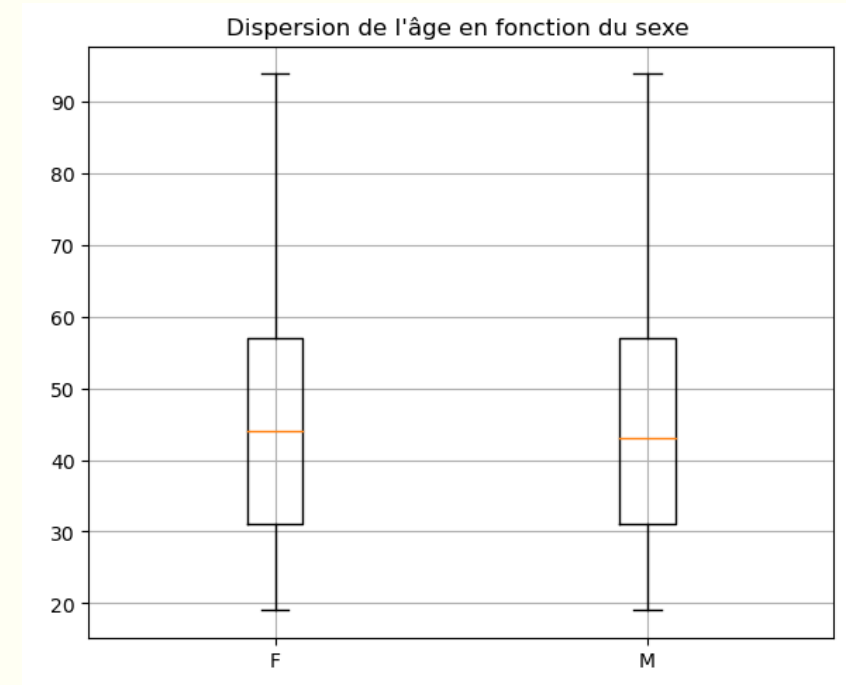
- **8596** clients actifs, moyenne d'âge de **45 ans**
- **Distribution de l'âge identique entre femmes et hommes**
confirmé par test de Mann et Whitney
- **319274** sessions d'achat ont été réalisées en 2 ans, soit **37 par client**
- En moyenne, **13303.08** sessions sont réalisées par mois, soit **1,5 sessions par client**
- Le montant du panier moyen par client est de **40,38** euros, la **distribution des paniers moyens** est identique entre femmes et hommes
confirmé par test de Mann et Whitney



- Parmi les clients : **52 % de femmes**
(hypothèse d'égalité des proportions rejetée par test binomial)
- La **proportion de sessions réalisées par les femmes est supérieure à 52%**
(confirmé par test binomial)

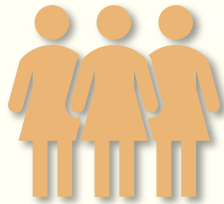


- Parmi les clients : **48 % d'hommes**
(hypothèse d'égalité des proportions rejetée par test binomial)

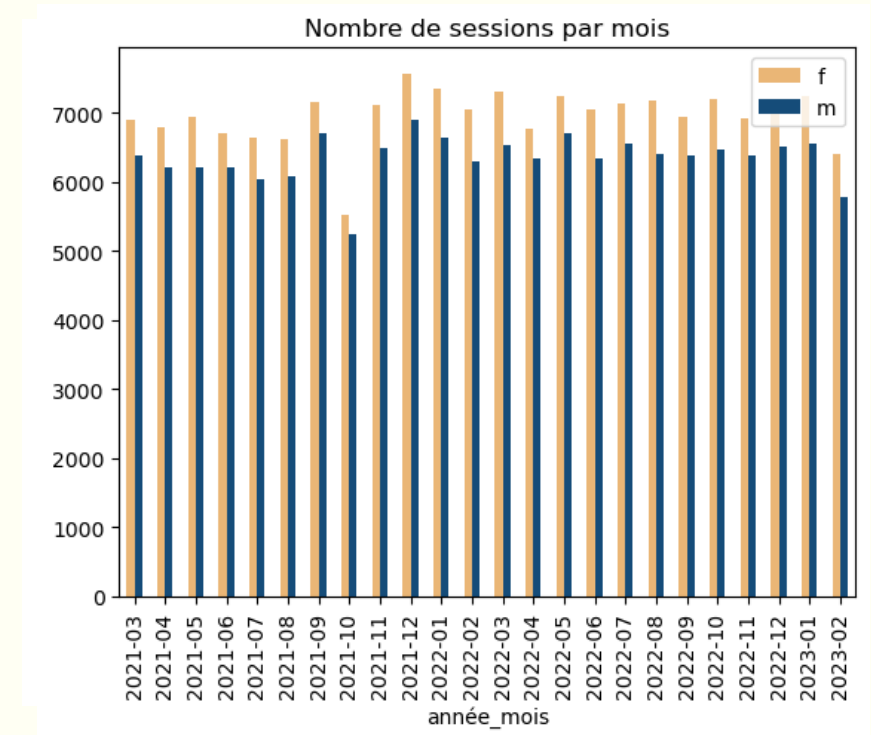


Qui sont nos clients ?

- **8596** clients actifs, moyenne d'âge de **45 ans**
- **Distribution de l'âge identique entre femmes et hommes**
confirmé par test de Mann et Whitney
- **319274** sessions d'achat ont été réalisées en 2 ans, soit **37 par client**
- En moyenne, **13303.08 sessions** sont réalisées par mois, soit **1,5 sessions par client**
- Le montant du panier moyen par client est de **40,38 euros**, la **distribution des paniers moyens** est identique entre femmes et hommes
confirmé par test de Mann et Whitney

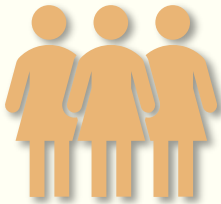


- Parmi les clients : **52 % de femmes**
(hypothèse d'égalité des proportions rejetée par test binomial)
- La **proportion de sessions réalisées par les femmes est supérieure à 52%**
(confirmé par test binomial)
- Parmi les clients : **48 % d'hommes**
(hypothèse d'égalité des proportions rejetée par test binomial)

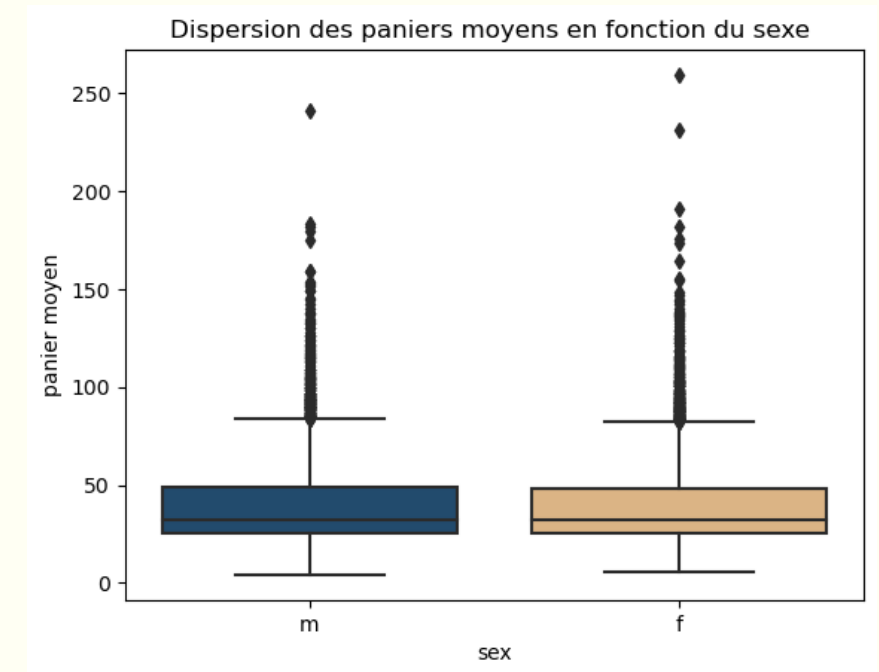


Qui sont nos clients ?

- **8596** clients actifs, moyenne d'âge de **45 ans**
- **Distribution de l'âge identique entre femmes et hommes**
confirmé par test de Mann et Whitney
- **319274** sessions d'achat ont été réalisées en 2 ans, soit **37 par client**
- En moyenne, **13303.08** sessions sont réalisées par mois, soit **1,5 sessions par client**
- Le montant du panier moyen par client est de **40,38** euros, la **distribution des paniers moyens** est identique entre femmes et hommes
confirmé par test de Mann et Whitney

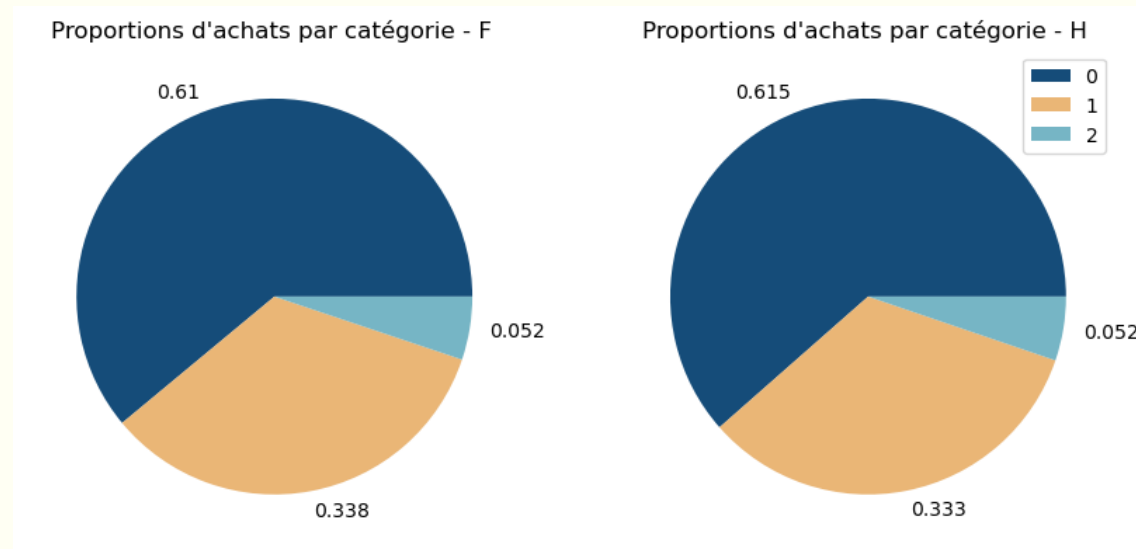


- Parmi les clients : **52 % de femmes**
(hypothèse d'égalité des proportions rejetée par test binomial)
- La **proportion de sessions réalisées par les femmes est supérieure à 52%**
(confirmé par test binomial)
- Parmi les clients : **48 % d'hommes**
(hypothèse d'égalité des proportions rejetée par test binomial)



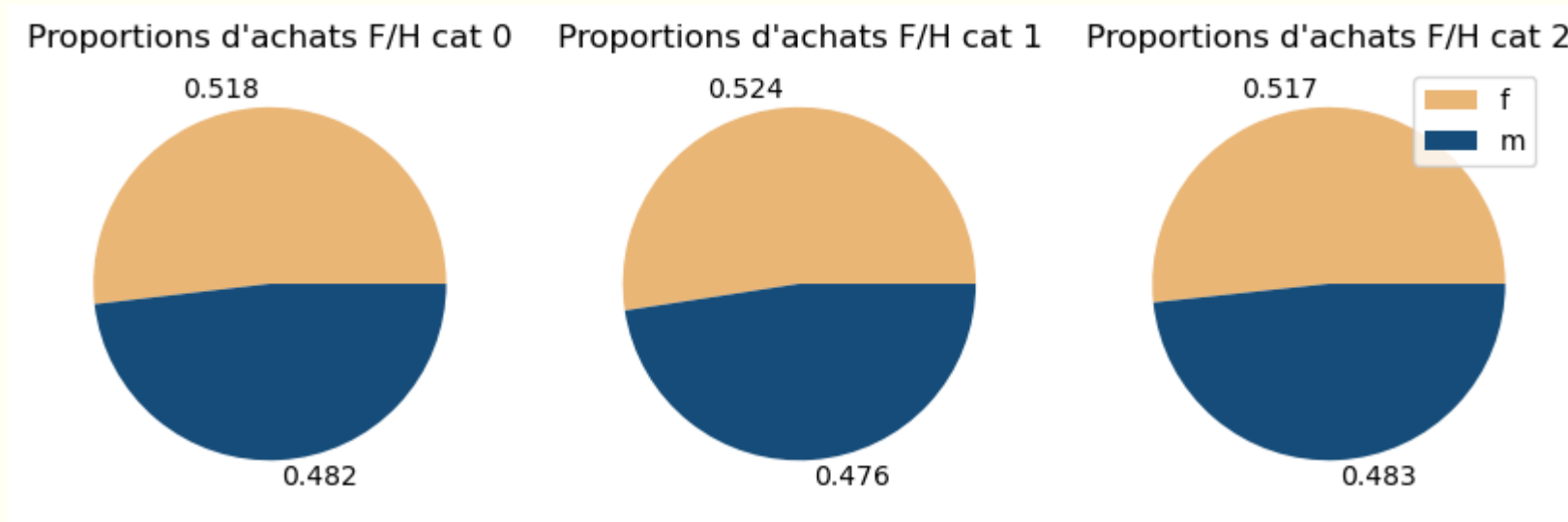
Etude de la corrélation entre genre et catégorie

- Les proportions d'achat par catégorie semblent être les mêmes chez les hommes et les femmes
- Les femmes achètent légèrement plus que les hommes, dans des proportions assez similaires au sein de chaque catégorie,
- D'après les graphiques, il semble ne pas y avoir de lien entre les variables genre et catégorie
 - → vérification par réalisation d'un tableau de contingence et d'un test du Khi2



Etude de la corrélation entre genre et catégorie

- Les proportions d'achat par catégorie semblent être les mêmes chez les hommes et les femmes
- Les femmes achètent légèrement plus que les hommes, dans des proportions assez similaires au sein de chaque catégorie,
- D'après les graphiques, il semble ne pas y avoir de lien entre les variables genre et catégorie
 - → vérification par réalisation d'un tableau de contingence et d'un test du Khi2

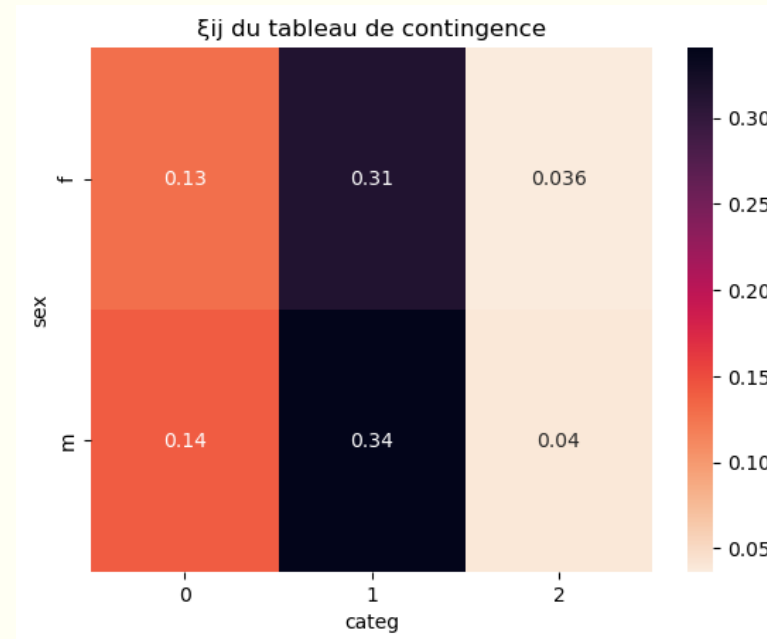


Etude de la corrélation entre genre et catégorie

- Les proportions d'achat par catégorie semblent être les mêmes chez les hommes et les femmes
- Les femmes achètent légèrement plus que les hommes, dans des proportions assez similaires au sein de chaque catégorie,
- D'après les graphiques, il semble ne pas y avoir de lien entre les variables genre et catégorie
 - → vérification par réalisation d'un tableau de contingence et d'un test du Khi2

Tableau de contingence				
	Categ 0	1	2	total
Sex f	200904	111331	16980	329215
m	186589	101017	15868	303474
total	387493	212348	32848	632689

Tableau des effectifs théoriques				
	Categ 0	1	2	total
Sex f	201629.09	110493.70	17092.21	329215.00
m	185863.91	101854.30	15755.79	303474.00
total	387493.00	212348.00	32848.00	632689.00

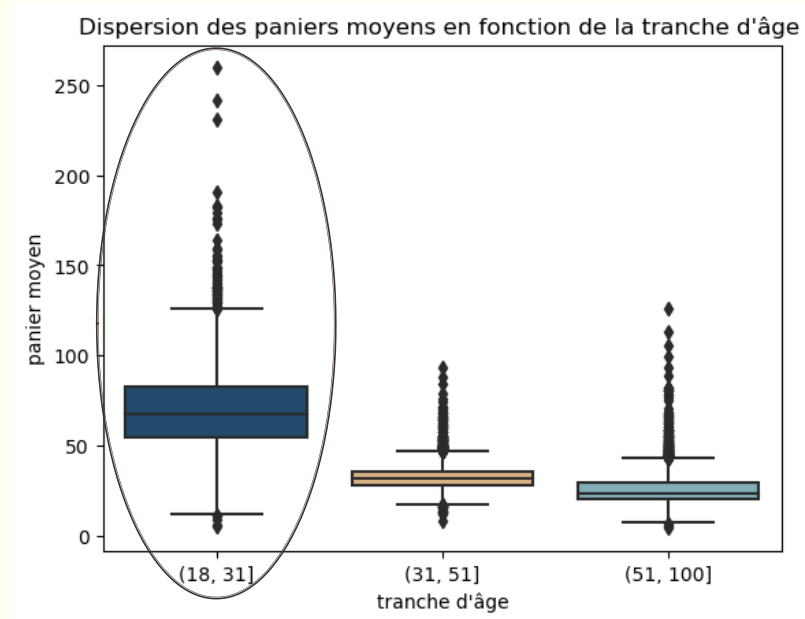
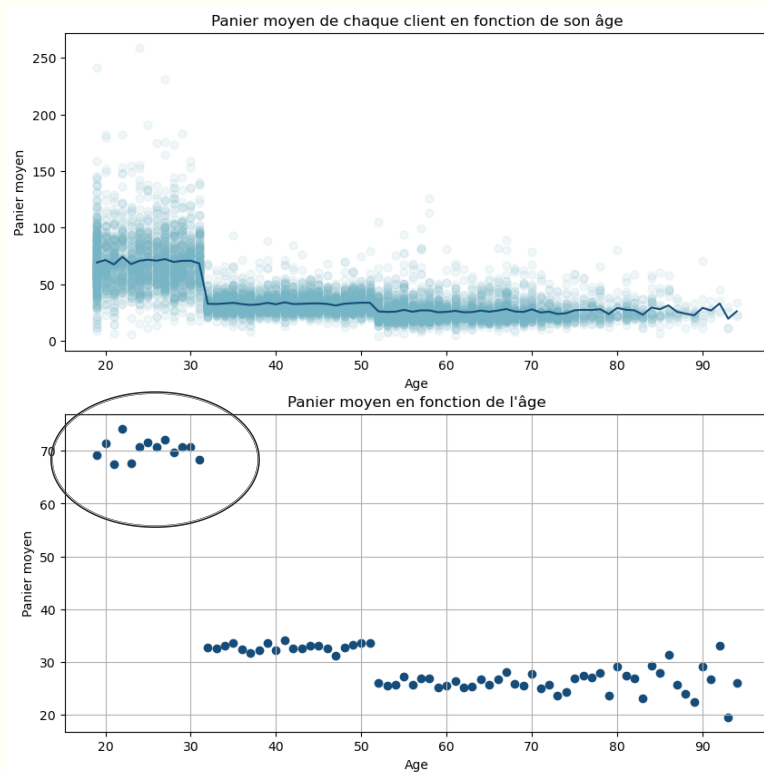


-C'est au sein de la catégorie 1 que l'écart entre les valeurs observées et les valeurs théoriques en cas d'indépendance des deux variables est le plus important
-Les cases m/cat1 et f/cat1 contribuent le plus à l'hypothèse de non indépendance des variables

$\text{Khi2} = 20$ $p_value = 4.1075985e-05$
on rejette l'hypothèse d'indépendance H_0

Etude de la corrélation entre âge et panier moyen

- Les graphiques semblent montrer un lien entre le panier moyen et l'âge : le panier moyen est plus élevé sur la tranche d'âge 19-31 ans
- Etude quantitative/quantitative avec calcul du coefficient de corrélation
 - Coefficient de Spearman : utilisé en cas de rejet de l'hypothèse de normalité des variables, et relation de type monotone entre les variables
- Etude quantitative/qualitative :
 - test de Kruskal-Wallis : en cas de rejet de l'hypothèse de normalité des variables, alternative à l'ANOVA



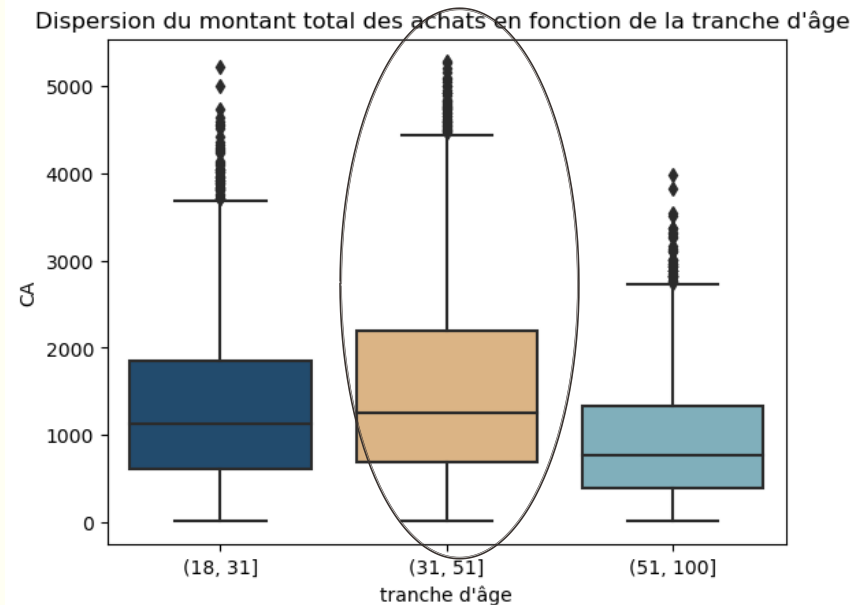
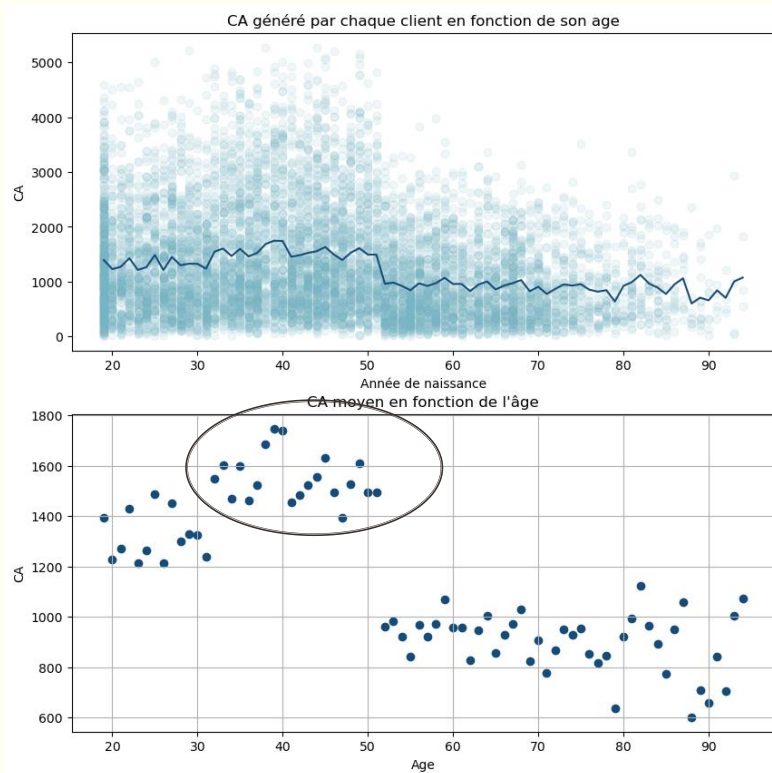
*-Coefficient de Spearman : -0.701
Corrélation négative entre les variables âge
et panier moyen*

*Comparaison des paniers moyens des 3
groupes en fonction de la tranche d'âge
P_value inférieure à 0.05 : 2.66735e-13
Rejet de l'hypothèse d'égalité des médianes
H0*

*D'après le calcul du coefficient de Spearman
et la réalisation du test de Kruskal-Wallis, il
existe une corrélation entre les variables âge
et panier moyen.*

Etude de la corrélation entre âge et montant total des achats

- Les graphiques semblent montrer un lien entre le montant total des achats et l'âge : le CA total est plus élevé sur la tranche d'âge 32-51 ans
- Etude quantitative/quantitative avec calcul du coefficient de corrélation
 - Coefficient de Kendall (Kendall Tau) : utilisé en cas de rejet de l'hypothèse de normalité, et lorsque la relation n'est ni de type linéaire ni de type monotone entre les variables
- Etude quantitative/qualitative :
 - test de Kruskal-Wallis : en cas de rejet de l'hypothèse de normalité des variables, alternative à l'ANOVA



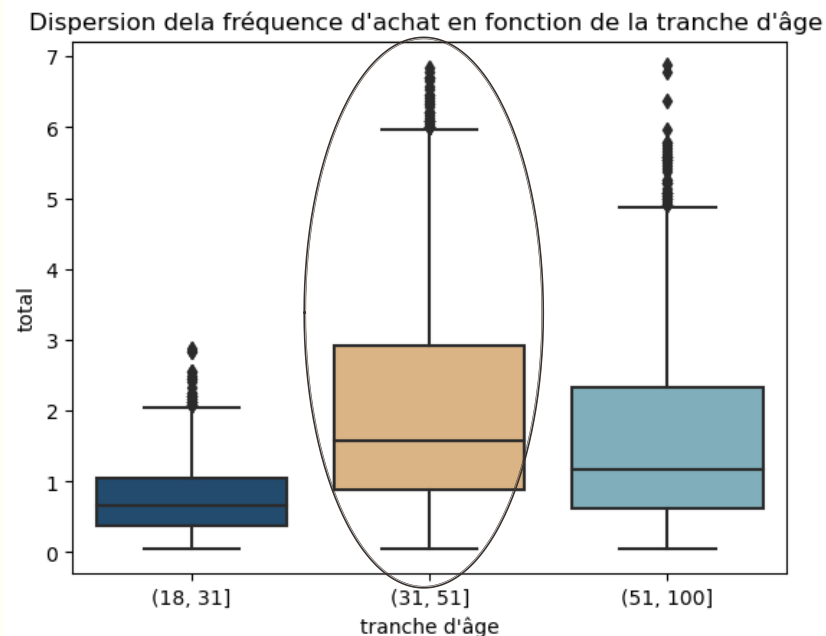
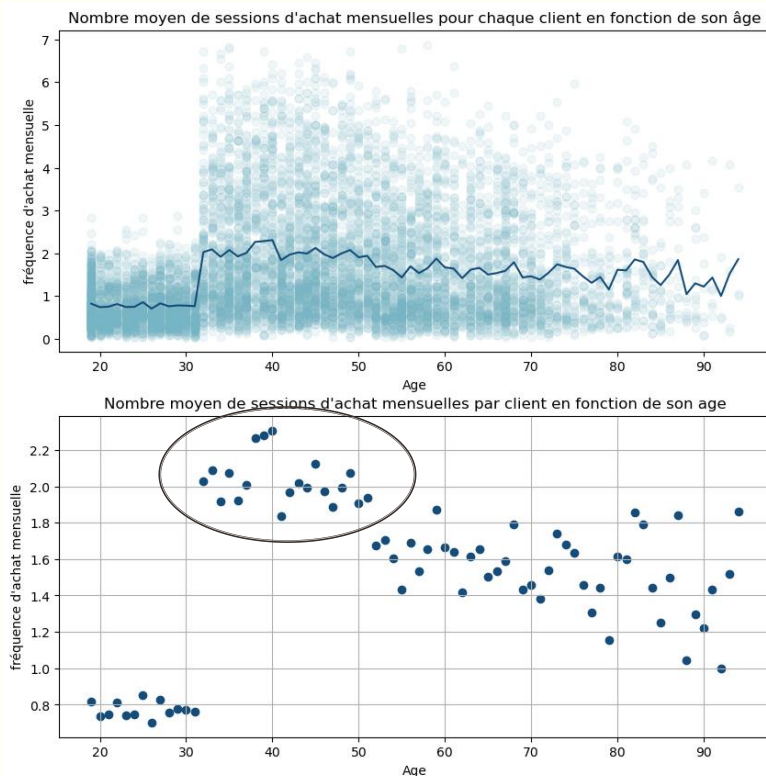
*-Coefficient de Kendall : -0.124
Légère corrélation négative entre les variables âge et montant total des achats*

*Comparaison du montant des achats des 3 groupes en fonction de la tranche d'âge
P_value inférieure à 0.05 : 6.2413e-134,
rejet de l'hypothèse d'égalité des médianes
H0*

D'après le calcul du coefficient de Kendall et la réalisation du test de Kruskal-Wallis, il existe une corrélation entre les variables âge et montant des achats.

Etude corrélation entre l'âge et la fréquence mensuelle des achats

- Les graphiques semblent montrer un lien entre la fréquence des achats et l'âge : le panier moyen est plus élevé sur la tranche d'âge 32-51 ans
- Etude quantitative/quantitative avec calcul du coefficient de corrélation
 - Coefficient de Kendall (Kendall Tau) : utilisé en cas de rejet de l'hypothèse de normalité, et lorsque la relation n'est ni de type linéaire ni de type monotone entre les variables
- Etude quantitative/qualitative :
 - test de Kruskal-Wallis : en cas de rejet de l'hypothèse de normalité des variables, alternative à l'ANOVA



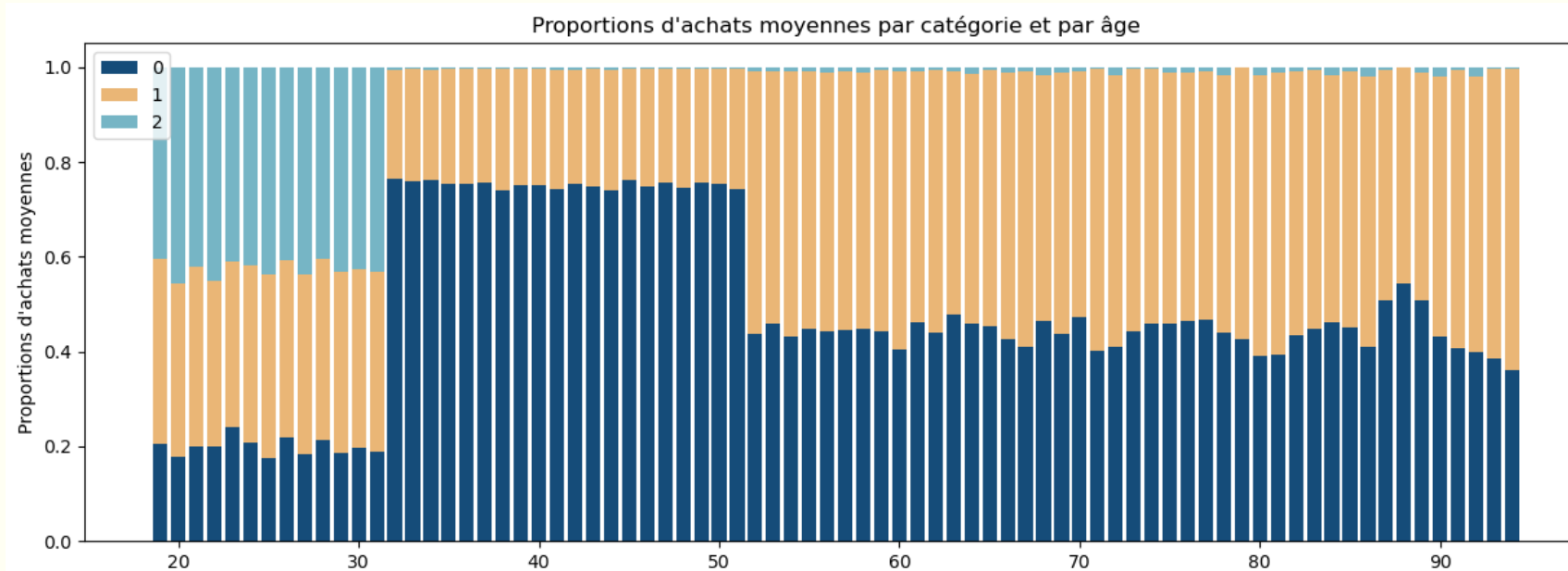
*-Coefficient de Kendall : 0.143
Légère corrélation entre les variables âge et fréquence d'achats*

*Comparaison des fréquences d'achats des 3 groupes en fonction de la tranche d'âge
P_value inférieure à 0.05 : $2.4051e-310$,
rejet de l'hypothèse d'égalité des médianes H_0*

D'après le calcul du coefficient de Kendall et la réalisation du test de Kruskal-Wallis, il existe une corrélation entre les variables âge et fréquence des achats.

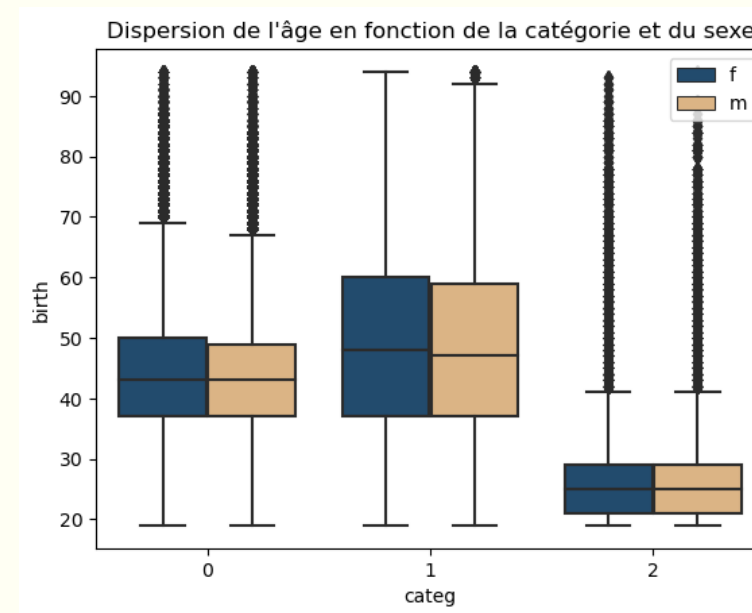
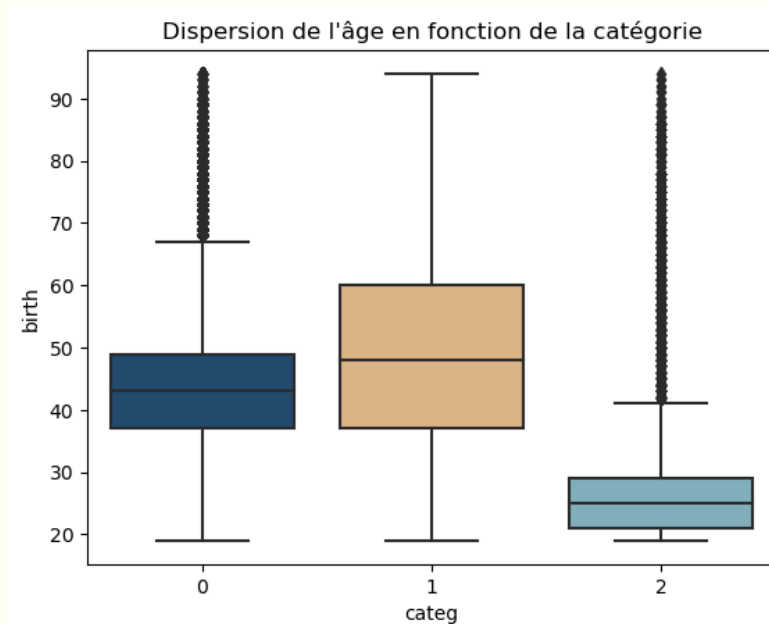
Etude de la corrélation entre âge et catégorie

- Les proportions d'achat par catégorie diffèrent grandement selon l'âge :
 - de 19 à 31 ans : intérêt pour toutes les catégories de produits
 - de 32 à 51 ans : intérêt majeur pour les produits de catégorie 0, pas d'intérêt pour la catégorie 2
 - de 52 à 95 ans : intérêt équivalent pour les produits de catégorie 0 et 1, pas d'intérêt pour la catégorie 2
- Les dispersions des âges sont différentes selon la catégorie :
 - catégorie 0 : 50% des achats effectués par des clients ayant entre 35 et 50 ans, peu d'achats après 65 ans
 - catégorie 1 : catégorie grand public, intéresse tous les âges
 - catégorie 2 : 50 % des achats effectués par des clients ayant entre 20 et 30 ans, peu d'achats après 40 ans
 - pas de différence flagrante si l'on compare la dispersion des âges par catégorie et en fonction du genre
- D'après les graphiques, il semble exister un lien entre les variables âge et catégorie
 - Etude qualitative /qualitative : réalisation d'un tableau de contingence



Etude de la corrélation entre âge et catégorie

- Les proportions d'achat par catégorie diffèrent grandement selon l'âge :
 - de 19 à 31 ans : intérêt pour toutes les catégories de produits
 - de 32 à 51 ans : intérêt majeur pour les produits de catégorie 0, pas d'intérêt pour la catégorie 2
 - de 52 à 95 ans : intérêt équivalent pour les produits de catégorie 0 et 1, pas d'intérêt pour la catégorie 2
- Les dispersions des âges sont différentes selon la catégorie :
 - catégorie 0 : 50% des achats effectués par des clients ayant entre 35 et 50 ans, peu d'achats après 65 ans
 - catégorie 1 : catégorie grand public, intéresse tous les âges
 - catégorie 2 : 50 % des achats effectués par des clients ayant entre 20 et 30 ans, peu d'achats après 40 ans
 - pas de différence flagrante si l'on compare la dispersion des âges par catégorie et en fonction du genre
- D'après les graphiques, il semble exister un lien entre les variables âge et catégorie
 - Etude qualitative /qualitative : réalisation d'un tableau de contingence

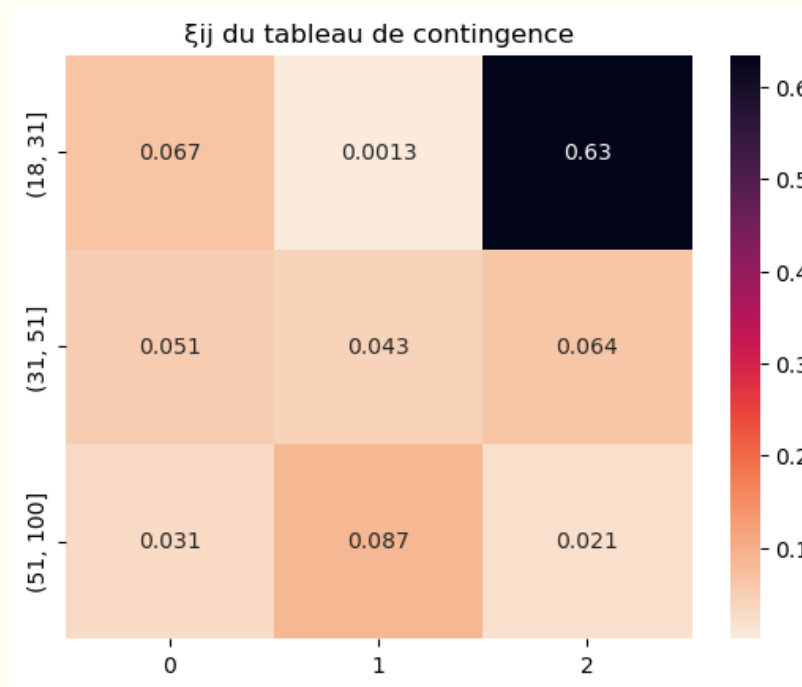


Etude de la corrélation entre âge et catégorie

- Les proportions d'achat par catégorie diffèrent grandement selon l'âge :
 - de 19 à 31 ans : intérêt pour toutes les catégories de produits
 - de 32 à 51 ans : intérêt majeur pour les produits de catégorie 0, pas d'intérêt pour la catégorie 2
 - de 52 à 95 ans : intérêt équivalent pour les produits de catégorie 0 et 1, pas d'intérêt pour la catégorie 2
- Les dispersions des âges sont différentes selon la catégorie :
 - catégorie 0 : 50% des achats effectués par des clients ayant entre 35 et 50 ans, peu d'achats après 65 ans
 - catégorie 1 : catégorie grand public, intéresse tous les âges
 - catégorie 2 : 50 % des achats effectués par des clients ayant entre 20 et 30 ans, peu d'achats après 40 ans
 - pas de différence flagrante si l'on compare la dispersion des âges par catégorie et en fonction du genre
- D'après les graphiques, il semble exister un lien entre les variables âge et catégorie
 - Etude qualitative / qualitative : réalisation d'un tableau de contingence

Tranche d'âge	0	1	2	Total
(18, 31]	16697	28087	29879	74663
(31, 51]	299424	92811	1475	393710
(51, 100]	71372	91450	1494	164316
total	387493	212348	32848	632689

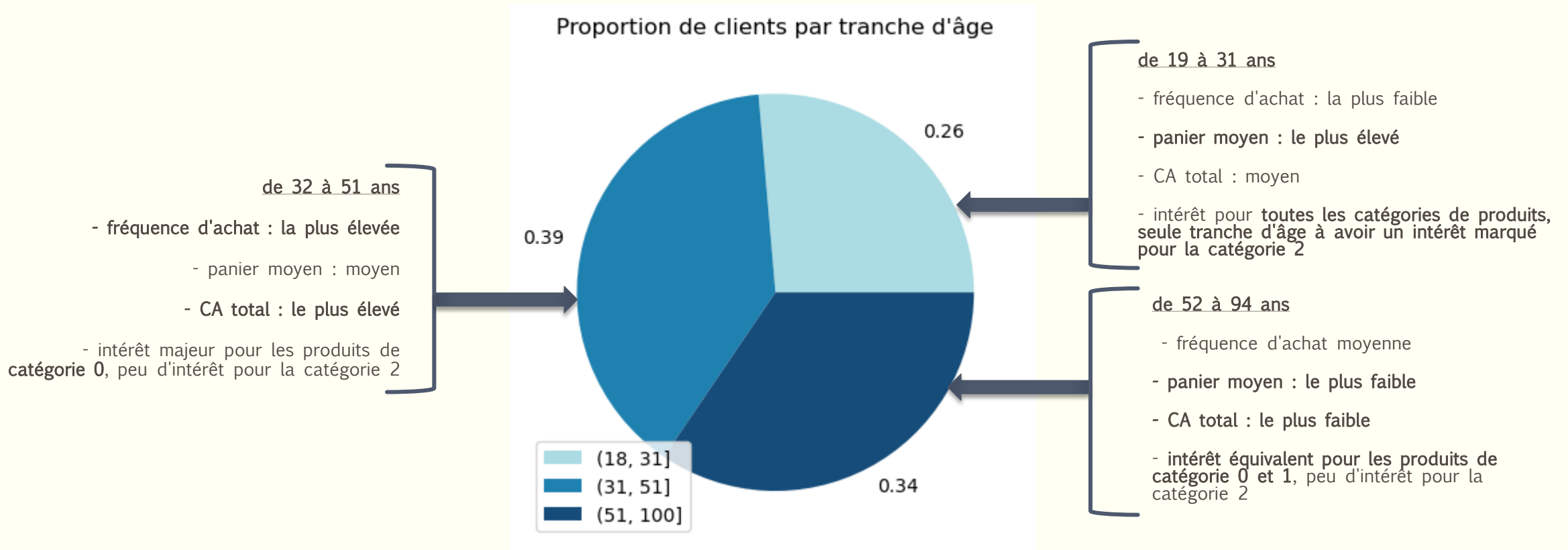
Tranche d'âge	0	1	2	Total
(18, 31]	16697	28087	29879	74663
(31, 51]	299424	92811	1475	393710
(51, 100]	71372	91450	1494	164316
total	387493	212348	32848	632689



- C'est au sein de la catégorie 2 pour la tranche d'âge 18-31 que l'écart entre les valeurs observées et les valeurs théoriques en cas d'indépendance des deux variables est le plus important

- La case 18-31/cat2 contribue le plus à l'hypothèse de non indépendance des variables

Synthèse : segmentation clients selon l'âge



Lien entre les références 0-525 et 2-159

- La fréquence empirique est ici utilisée comme estimateur de la probabilité
 - La probabilité qu'un client achète la référence 2_159 est estimée par la proportion de clients ayant acheté 2_159,*
 - La probabilité qu'un client achète 2_159 et 0_525 est estimée par la proportion de clients ayant acheté les deux références.*
- 526 clients uniques ont acheté la référence 2_159, 455 clients uniques ont acheté la référence 0_525, tous les clients ayant acheté 0_525 ont aussi acheté 2_159 :
 - l'évènement acheter 0_525 est donc inclus dans l'évènement acheter 2_159*
 - la probabilité d'acheter 0_525 et 2_159 est donc égale à la probabilité d'acheter 0_525*

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

2^e évènement

combinaison du 2^e évènement et de la référence

événement de référence

La probabilité qu'un client achète l'article '0_525' sachant qu'il a acheté l'article '2_159' est de 0.865

Recommandations

Concernant nos données :

- Intégrer la référence **0_2245** dans la base products

Concernant nos produits :

- Compléter l'analyse des produits par le CA réalisée par **une analyse de la rentabilité et de l'âge des produits** (*50% du CA réalisé est généré par 7 % des produits, certains produits ne se sont pas vendus depuis le lancement du site*)

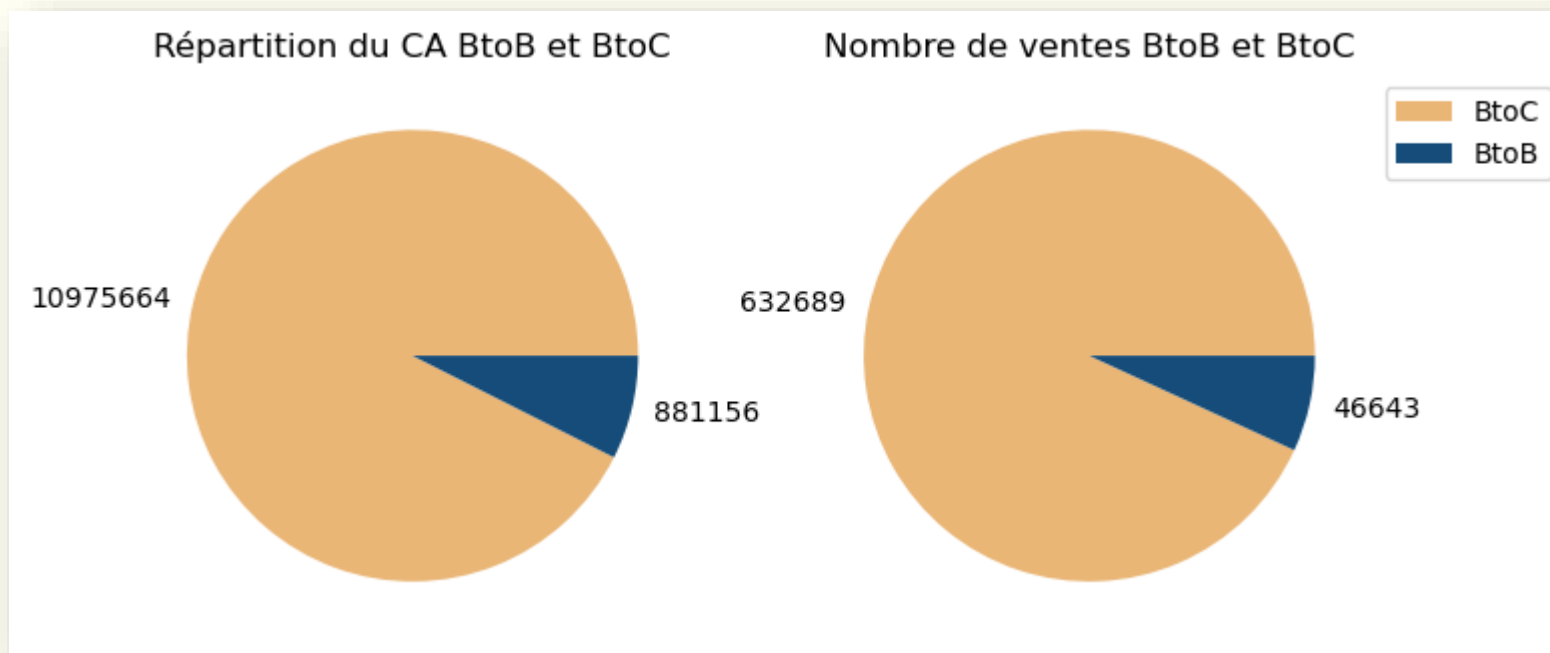
Concernant les analyses :

- Une meilleure identification des clients **BtoB** et de leurs caractéristiques pour une analyse plus précise du segment
 - *exemples de caractéristiques organisation : attributs de l'organisation, localisation, niveau de chiffre d'affaire, taille, ancienneté*
- Identification d'autres caractéristiques clients **BtoC** pour une segmentation plus fine
 - *exemples d'autres caractéristiques : profession, centres d'intérêt*

Concernant l'objectif de croissance du CA par client :

- Les leviers d'action pourraient être différents en fonction du segment :
 - pour le segment 19 à 31 ans : augmenter la **fréquence** d'achat
 - pour le segment 32 à 51 ans : augmenter la **valeur du panier moyen**
 - pour le segment 52 à 94 ans : augmenter le **panier moyen et la fréquence d'achat**
- Exemples d'actions visant à augmenter le panier moyen :
 - proposer des **produits complémentaires**/alternatifs/up-selling
 - ex : lien entre références 0_525 et 0_259
 - mettre en avant les promotions sur le site, créer des ventes flash
 - définir un seuil au-delà duquel les frais de livraison sont offerts
- Exemples d'actions visant à augmenter la fréquence d'achat :
 - lancer un programme de **fidélisation**
 - relancer les **abandonnistes, les clients inactifs (calcul de la récence)**
 - **cibler les campagnes marketing :**
 - **en fonction de la tranche d'âge** par exemple
 - en fonction de la récence/fréquence de achats
 - pas de ciblage particulier en fonction du genre
 - améliorer l'expérience utilisateur sur le site internet :
 - améliorer les fiches produit, ajouter une fonction ciblée recherche de produit, simplifier le tunnel d'achat (retravailler les boutons call-to-action, moyens de paiement divers...)

Annexes : BtoB BtoC



Annexes : CA et nombre de ventes par jour février 2023

