

# Rester livres



Analyse des ventes

Nettoyage des données

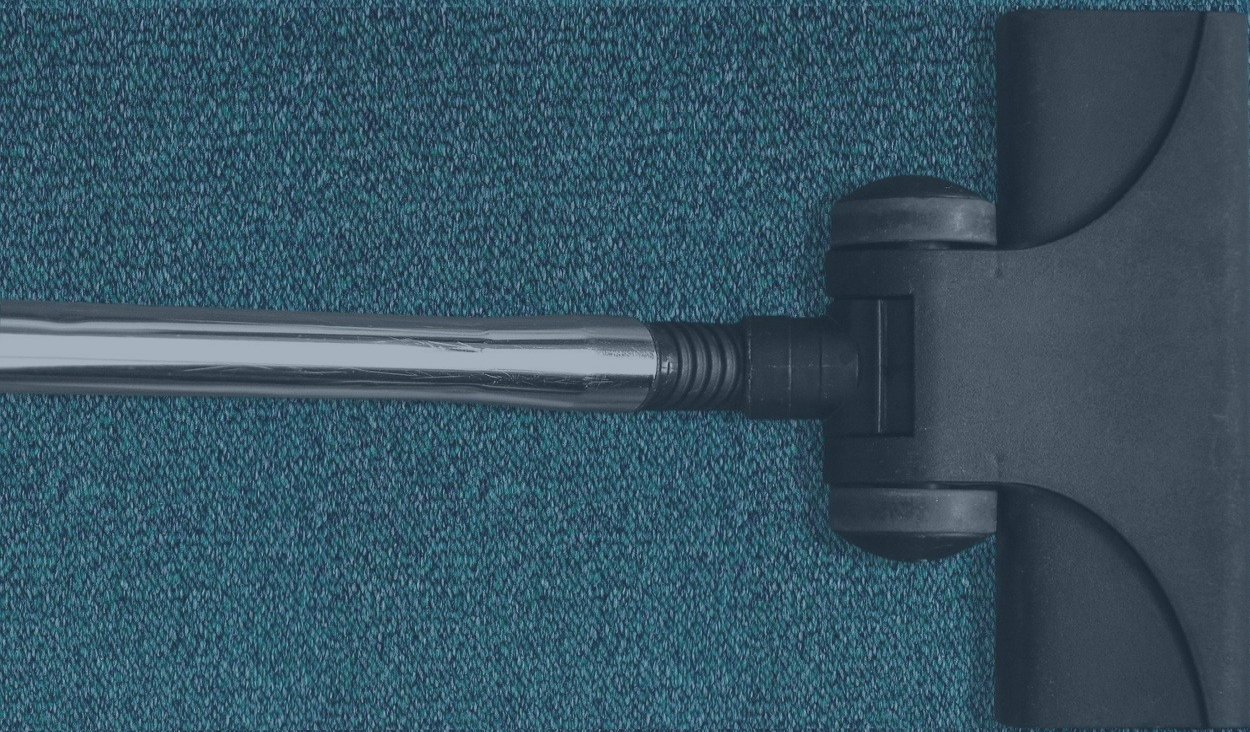
Analyse

Tests de corrélation

Conclusion







Nettoyage  
des  
données



# Les données

3 fichiers : - clients  
- produits  
- transactions

Absence de doublons

	id_prod		date	session_id	client_id
0	0_1483	2021-04-10 18:37:28.723910		s_18746	c_4450
1	2_226	2022-02-03 01:55:53.276402		s_159142	c_277
2	1_374	2021-09-23 15:13:46.938559		s_94290	c_4270
	2	0_731	17.99	0	

337201 priorités

01/03/2021 à 28/12/2022

# Valeurs aberrantes

Données de tests :

- transactions
- clients
- produits



Suppression des données

	id_prod				ion_id	client_id
count	200				200	200
		id_prod	price	categ		
unique	1	731	T_0	-1.0	0	2
top	T_0				s_0	ct_0
freq	200				200	106

200 transactions tests

200 produits tests

# Valeurs extrêmes



Année de naissance  
entre 1929 et 2004



Prix entre 0,62€ et 300€

Données conservées

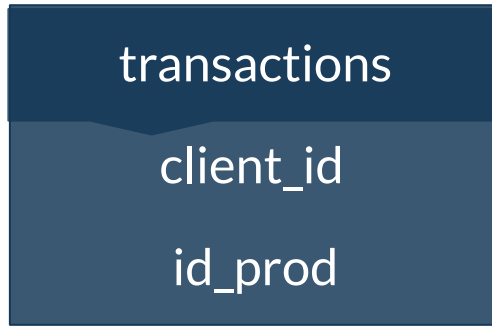
# Top 4

client\_id

c_1609	12855
c_6714	4473
c_3454	3275
c_4958	2562

Données conservées

# Valeurs manquantes



Ajout produit 0\_2245 ➡ categ = 0, price =  $\bar{x}$  prix de la categ



# Sessions manquantes

→ 3230 identifiants manquants

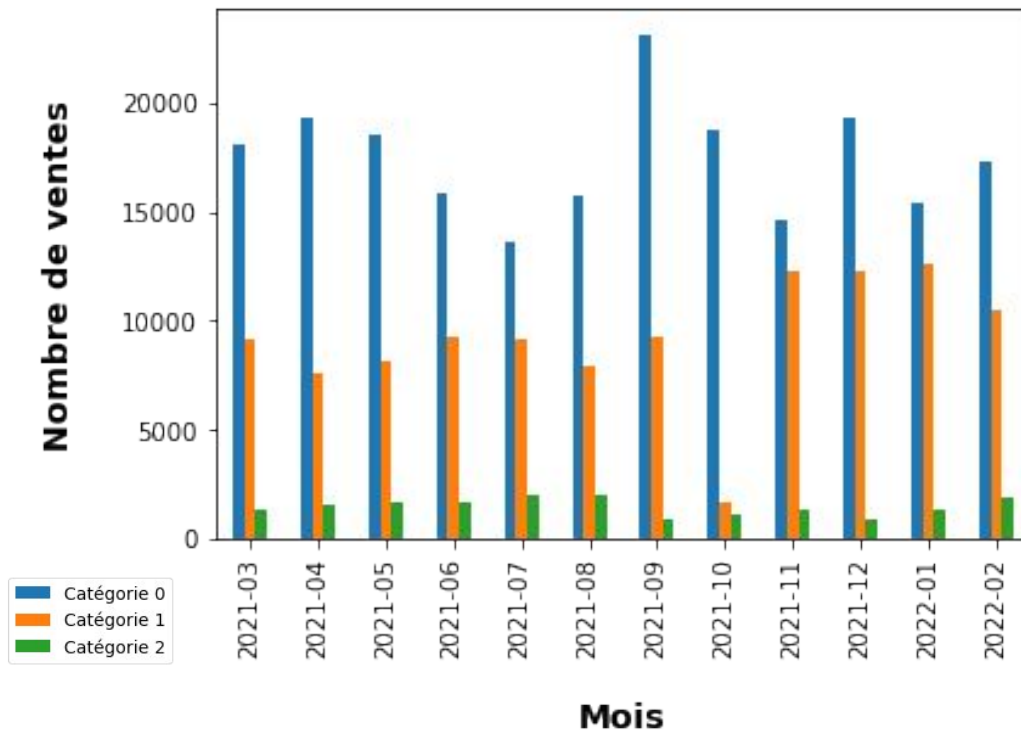
→ octobre 2021

→ catégorie 1

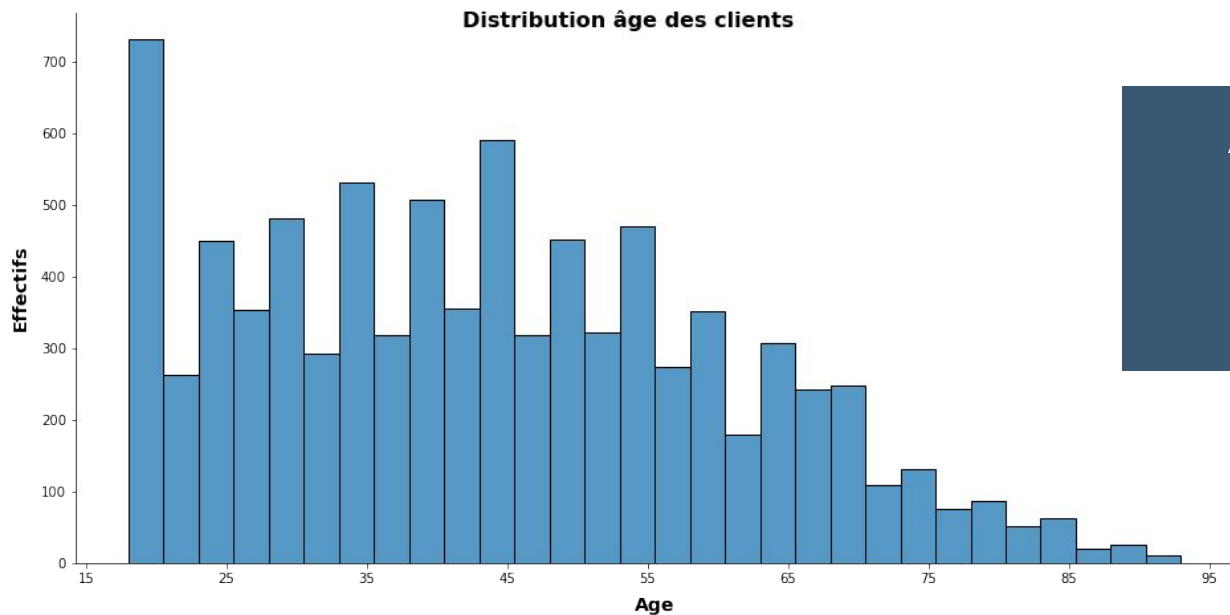


Suppression des données  
d'octobre 2021

Ventes mensuelles par catégorie



# Surreprésentation des 18 ans



Age moyen  
et médian  
=  
43 ans

Création de 4 classes d'âge

→ 15 à 30 ans

→ 31 à 43 ans

→ 44 à 56 ans

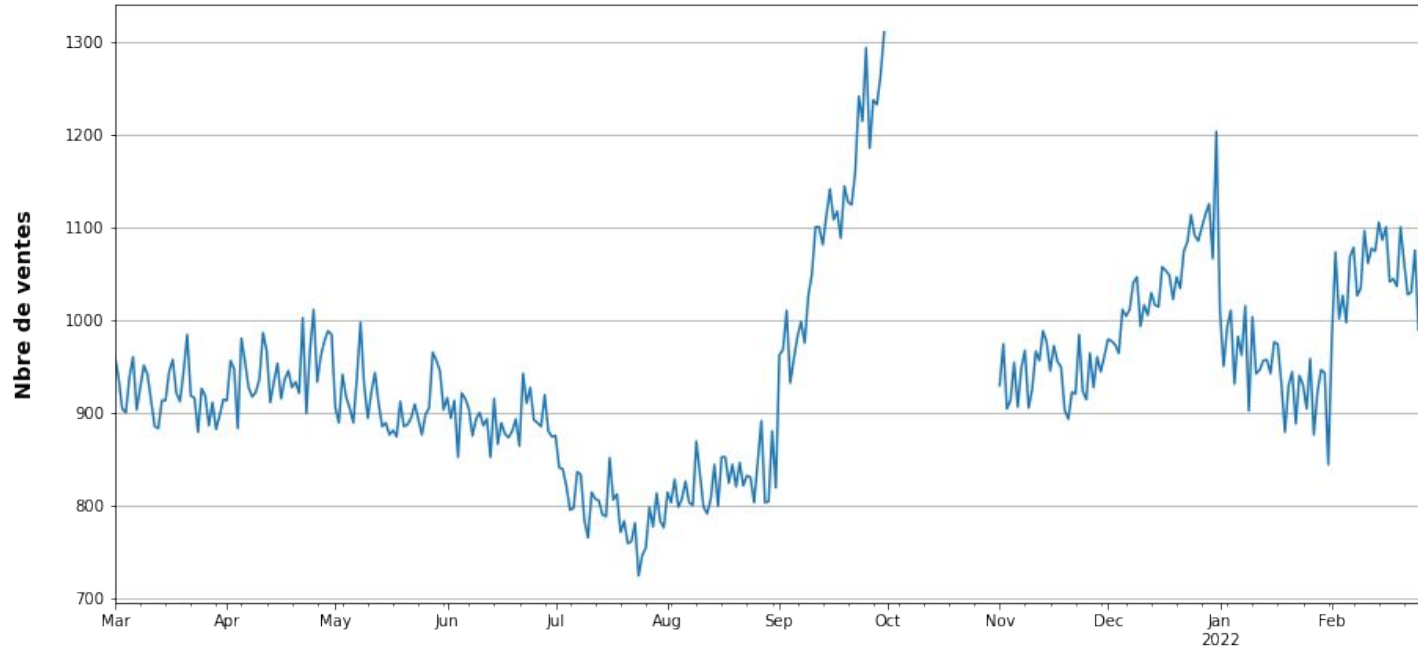
→ 56 à 93 ans



# Analyse

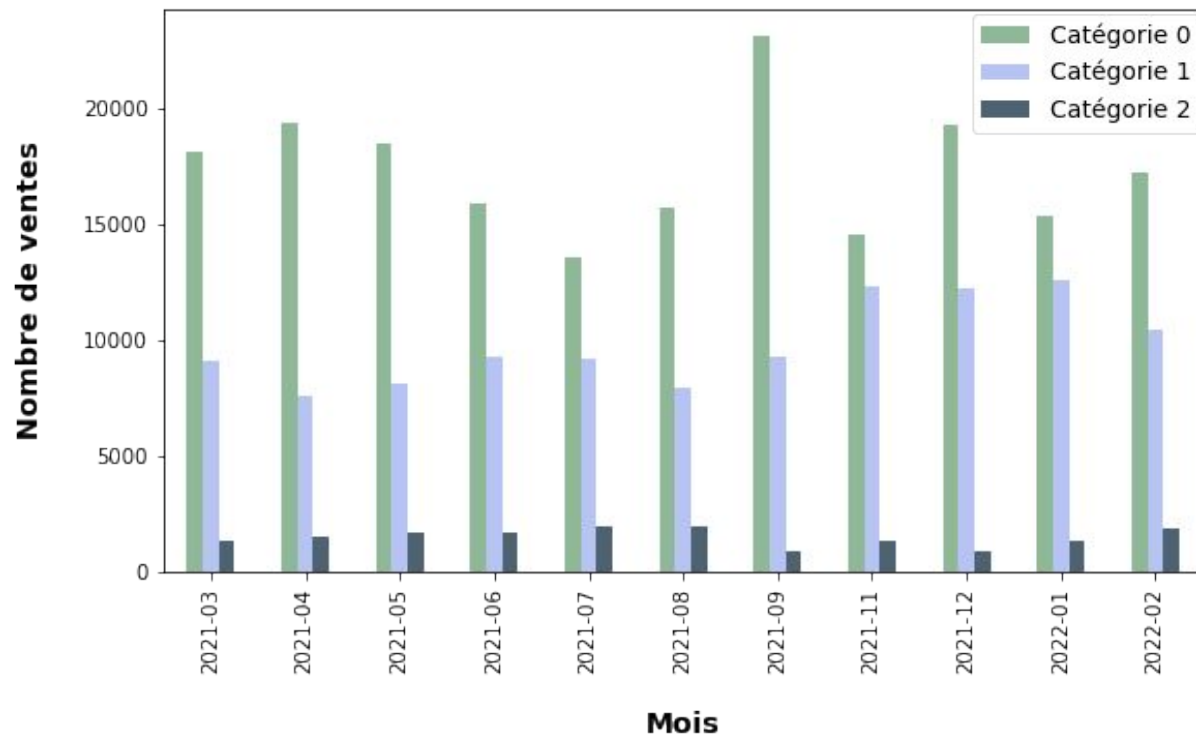
# Saisonnalité des ventes

Suivi des ventes par jour



# Ventes par catégorie

Ventes mensuelles par catégorie



**Cat. 0**  
Hausse  
septembre et  
décembre

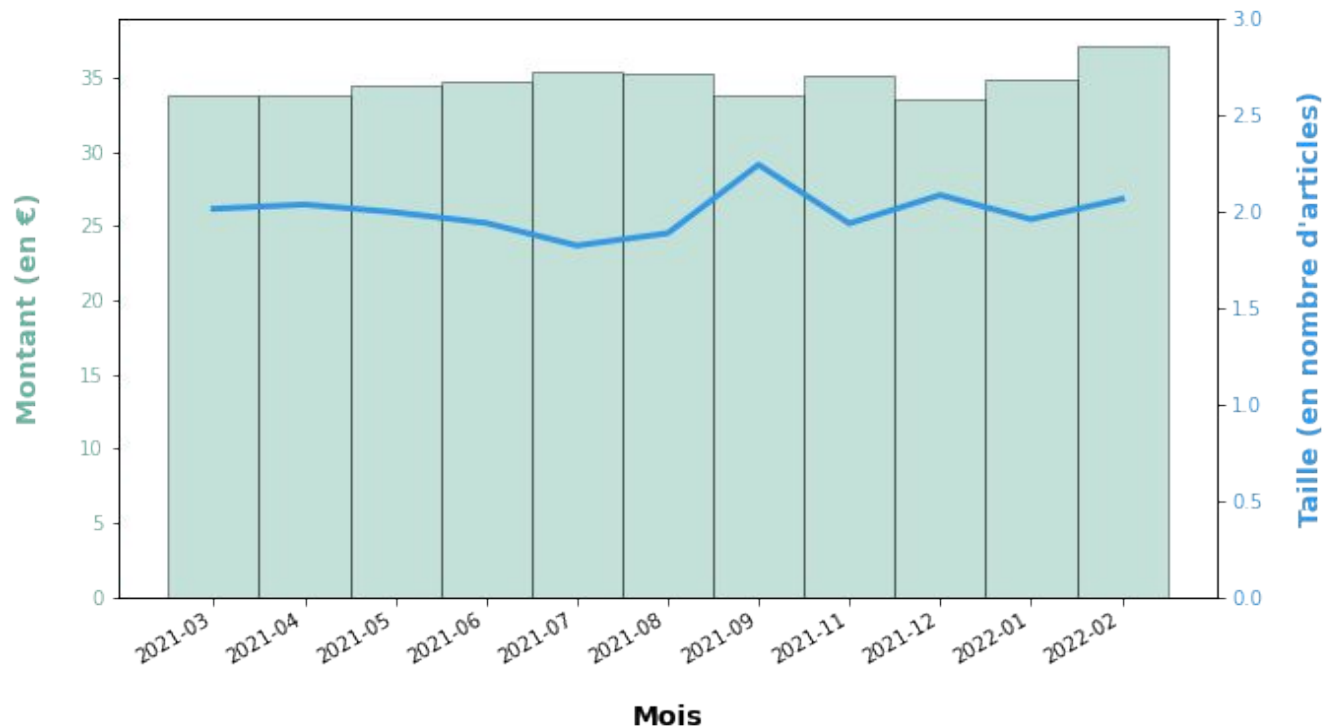
**Cat. 1**  
Hausse  
novembre à janvier

**Cat. 2**  
Baisse  
septembre à janvier



# Panier Moyen

Stabilité du panier moyen

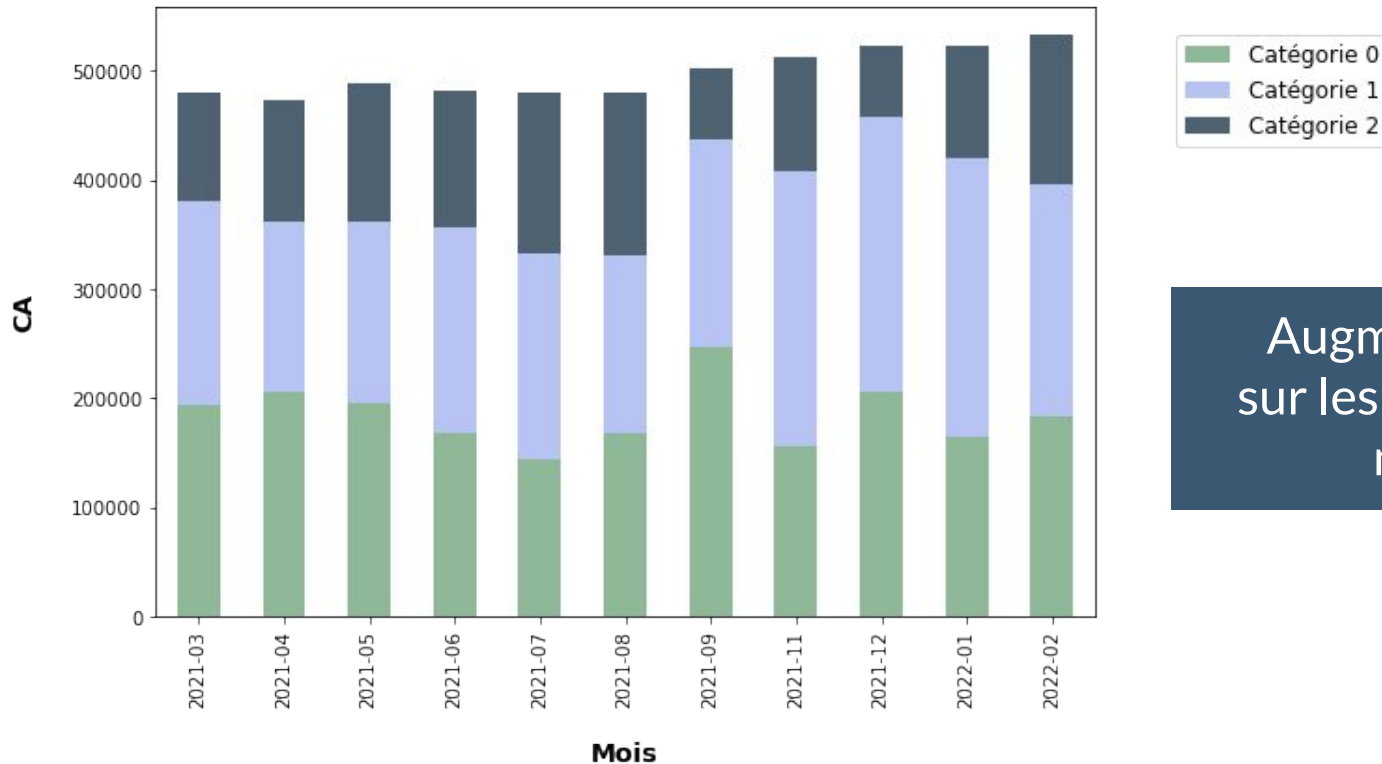


2 articles

Hausse du  
montant en  
février 2022

# Evolution du CA par catégorie

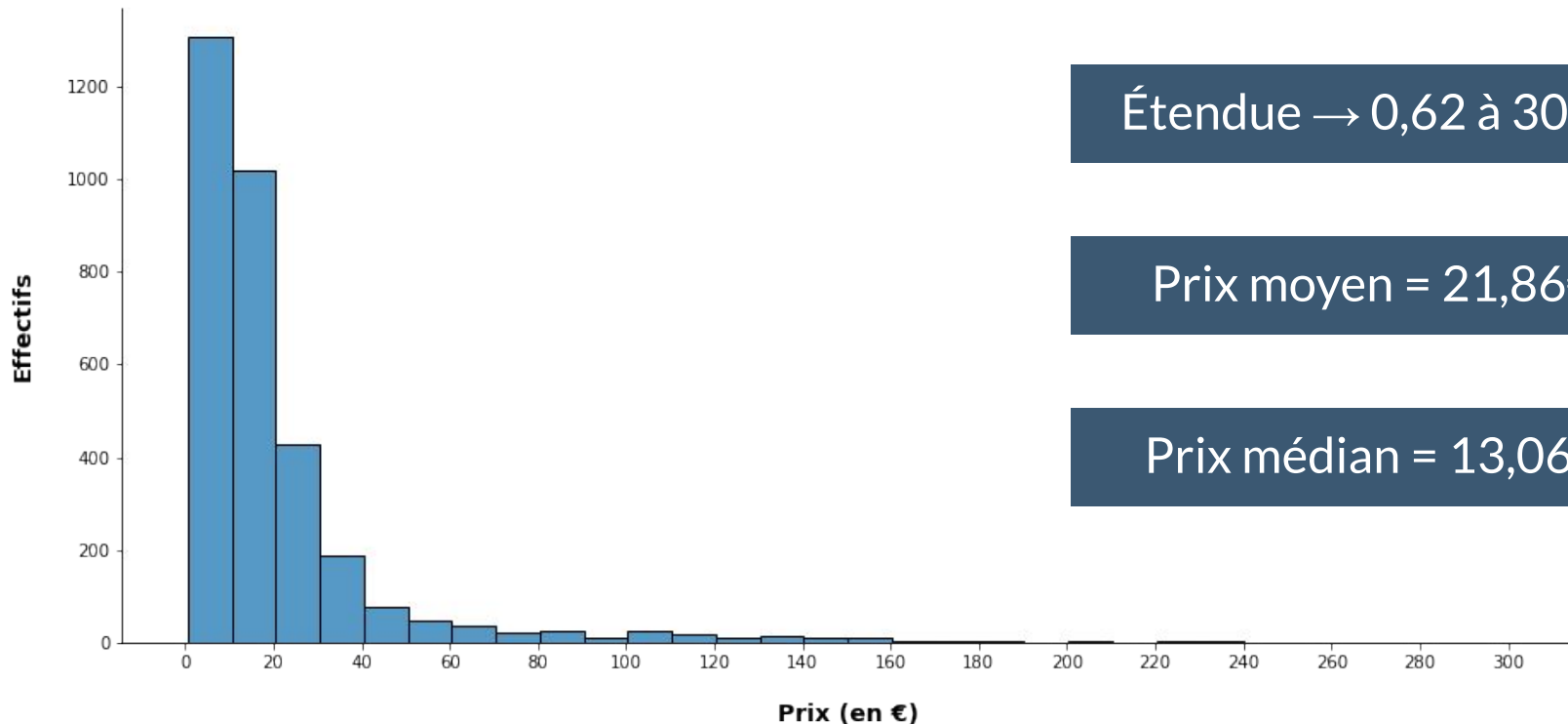
Evolution mensuelle du chiffre d'affaires par catégorie



Augmentation  
sur les 6 derniers  
mois

# Distribution prix des produits

Les prix des produits sont majoritairement < 20 €



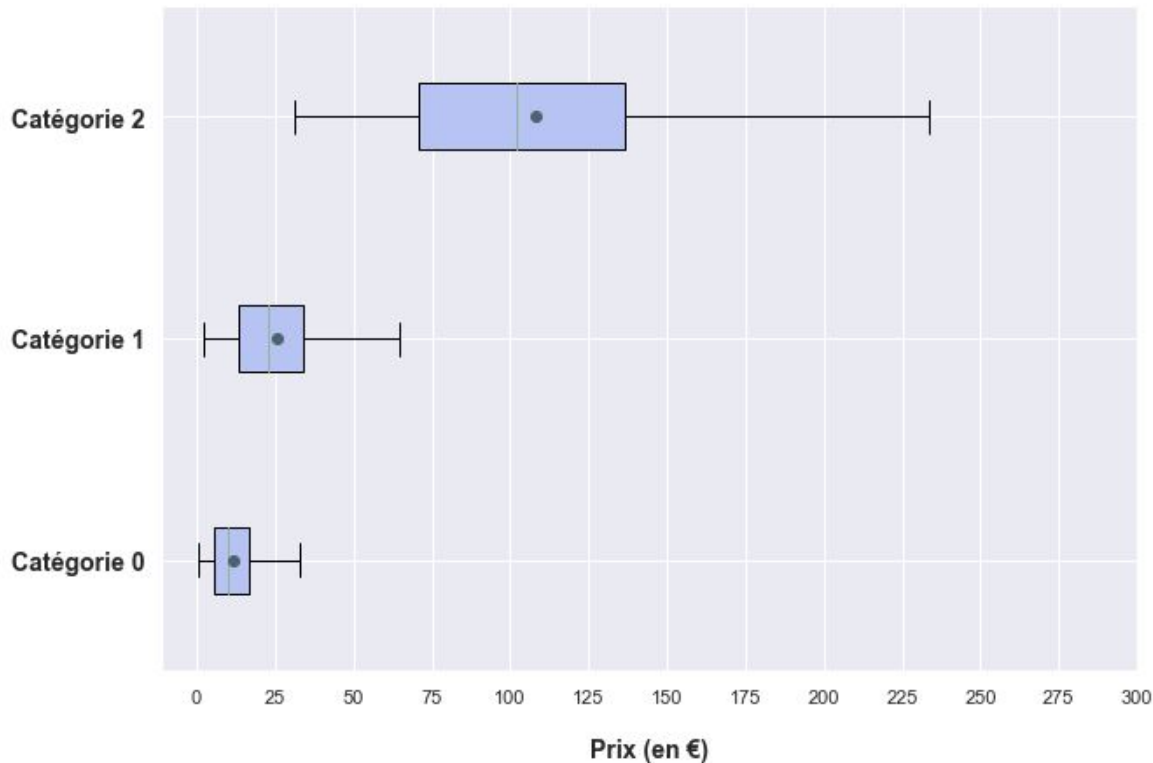
Étendue → 0,62 à 300€

Prix moyen = 21,86€

Prix médian = 13,06€

# Prix par catégorie

Disparité des prix par catégorie



Cat. 2

3 livres sur 4  
prix > 70€

Cat. 1

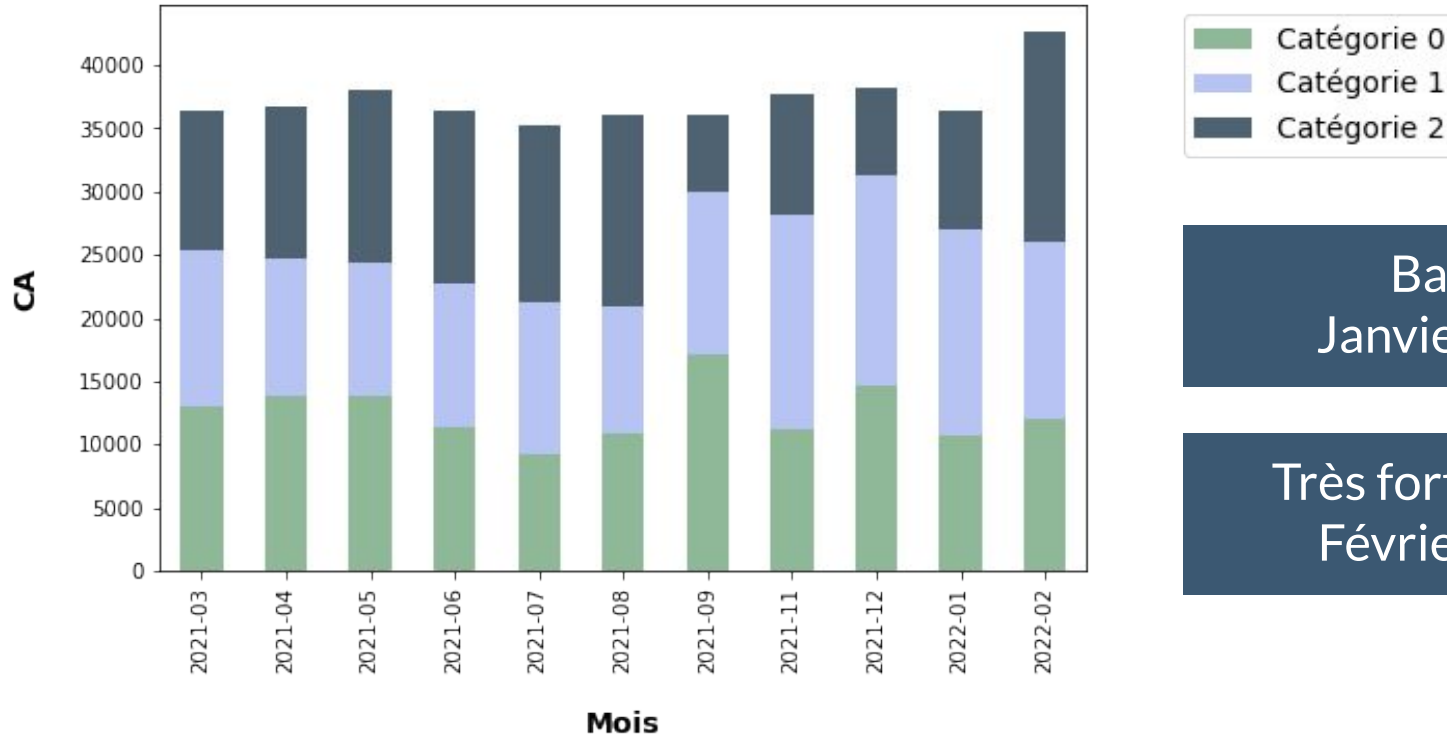
prix moyen  
= 25€

Cat. 0

prix moyen  
= 11€

# Evolution CA top 4 clients

CA par mois des 4 clients effectuant le plus d'achats

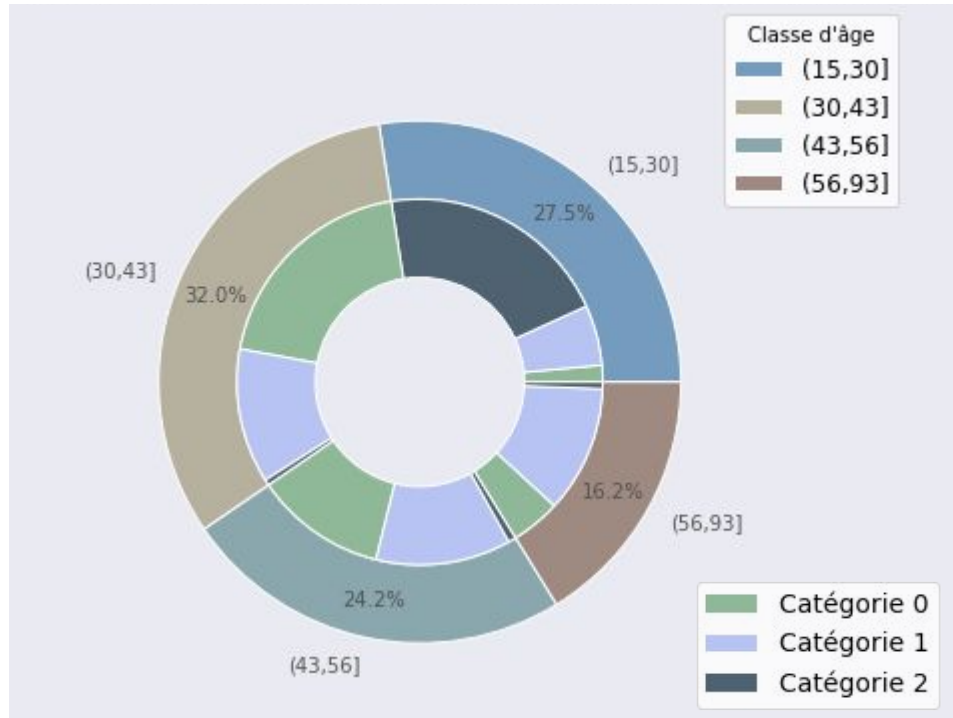


Baisse  
Janvier 2022

Très forte hausse  
Février 2022



# Répartition du CA par classe d'âge et par catégorie

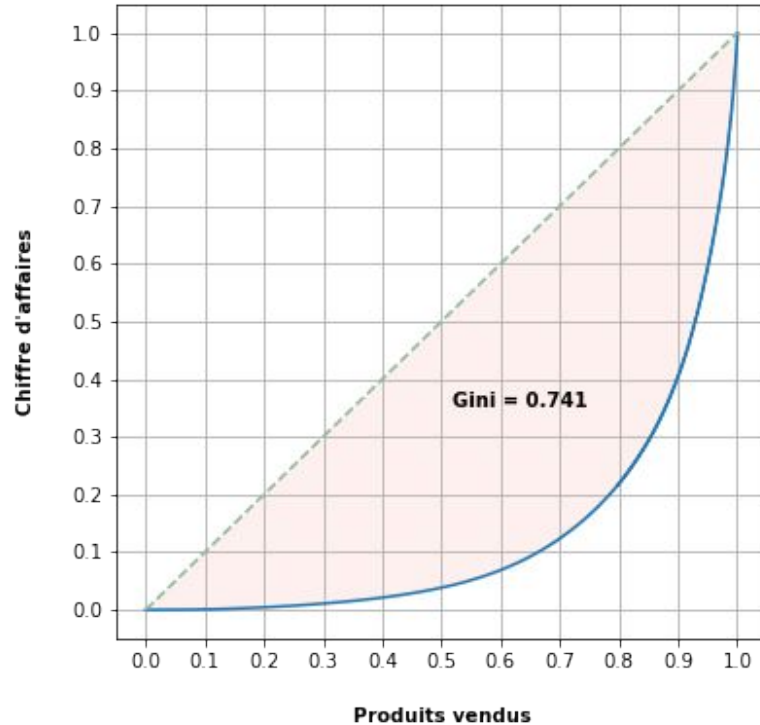


15 à 30 ans :  
Catégorie 2  
prédominante

30 à 43 ans :  
Catégorie 0  
prédominante

# CA par livres

**Courbe Lorenz - Chiffre d'affaires par produit**

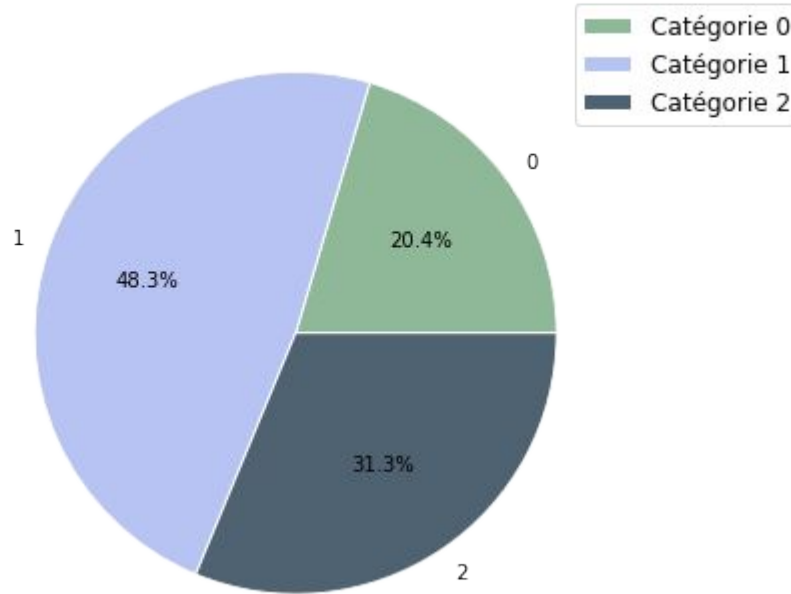


50% des références  
représentent moins de  
5% du CA

60% du CA généré par 10%  
des références

# 328 livres génèrent 60% du CA

Répartition du CA par catégorie des 10% des références les plus vendues



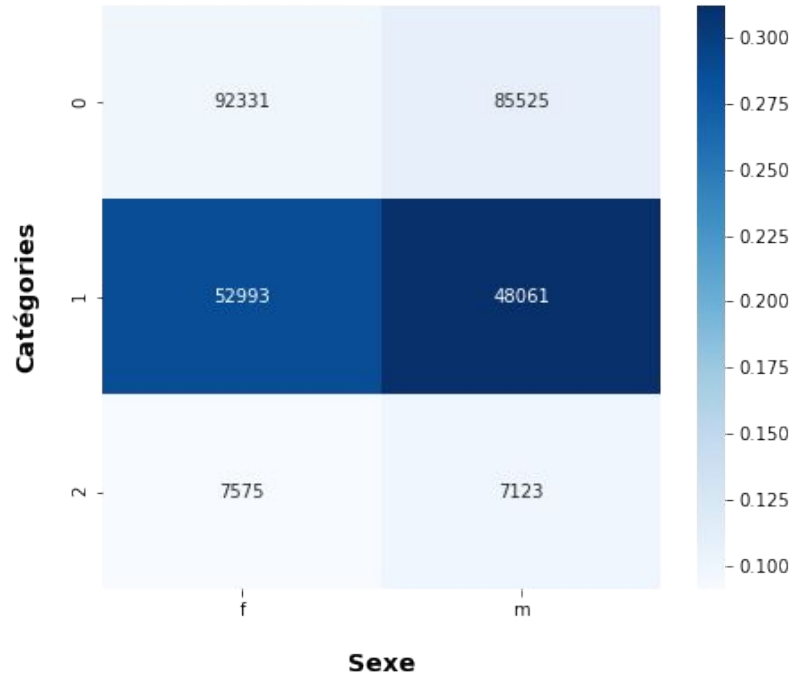
10% des  
références =  
328 livres



Tests de  
corrélation

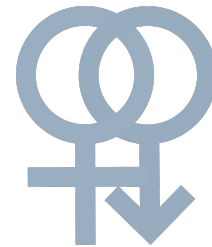
# Corrélation entre le sexe des clients et les catégories de produits achetés ?

Tableau de contingence catégories et sexe des clients



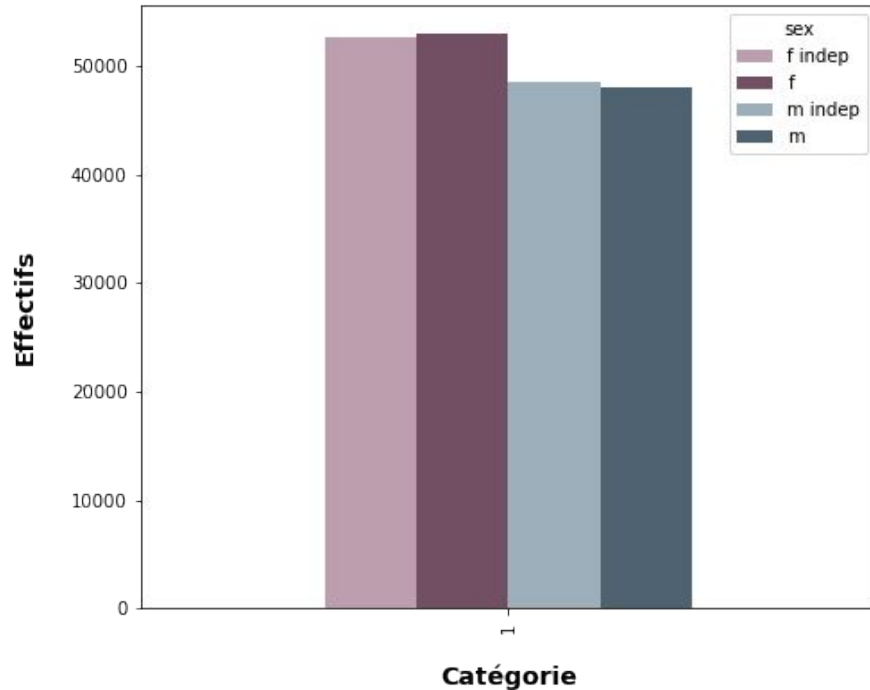
Chi-2 : variables dépendantes

Relation entre la catégorie 1  
et le sexe des clients





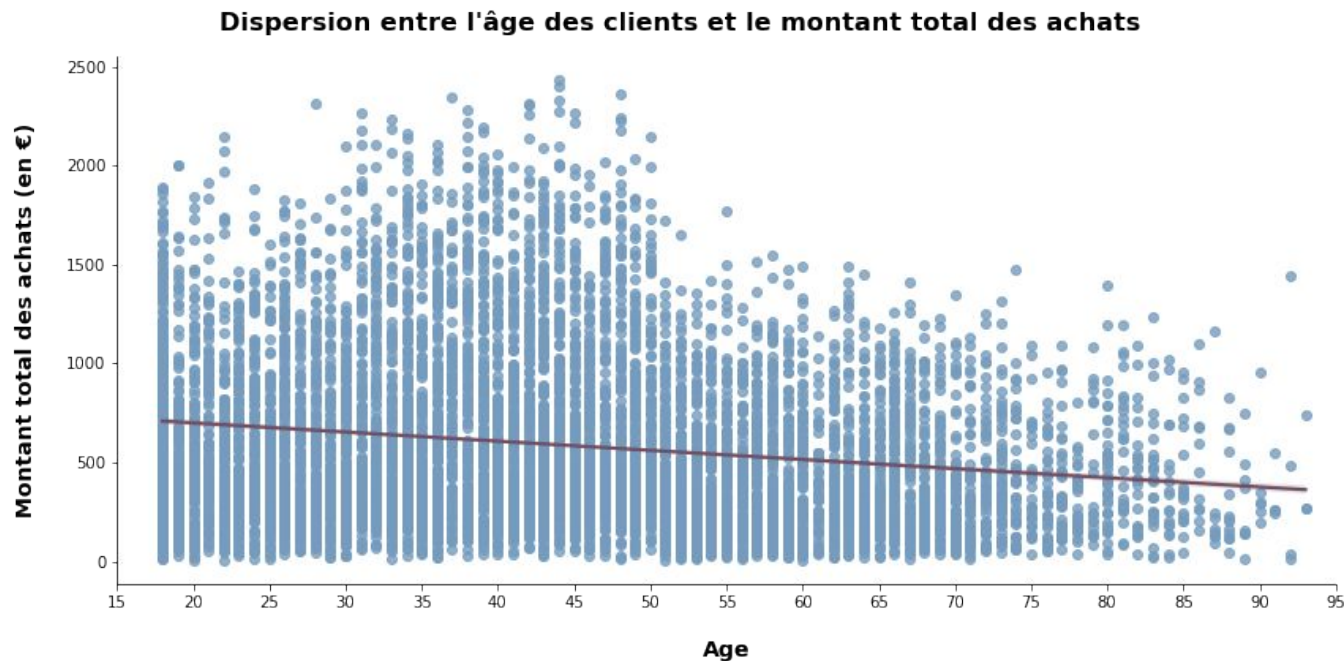
# Focus catégorie 1 - Comparaison avec l'hypothèse d'indépendance entre le sexe et la catégorie de produits achetés



Les femmes achètent plus

Les hommes achètent moins

# Corrélation entre l'âge des clients et le montant total des achats ?

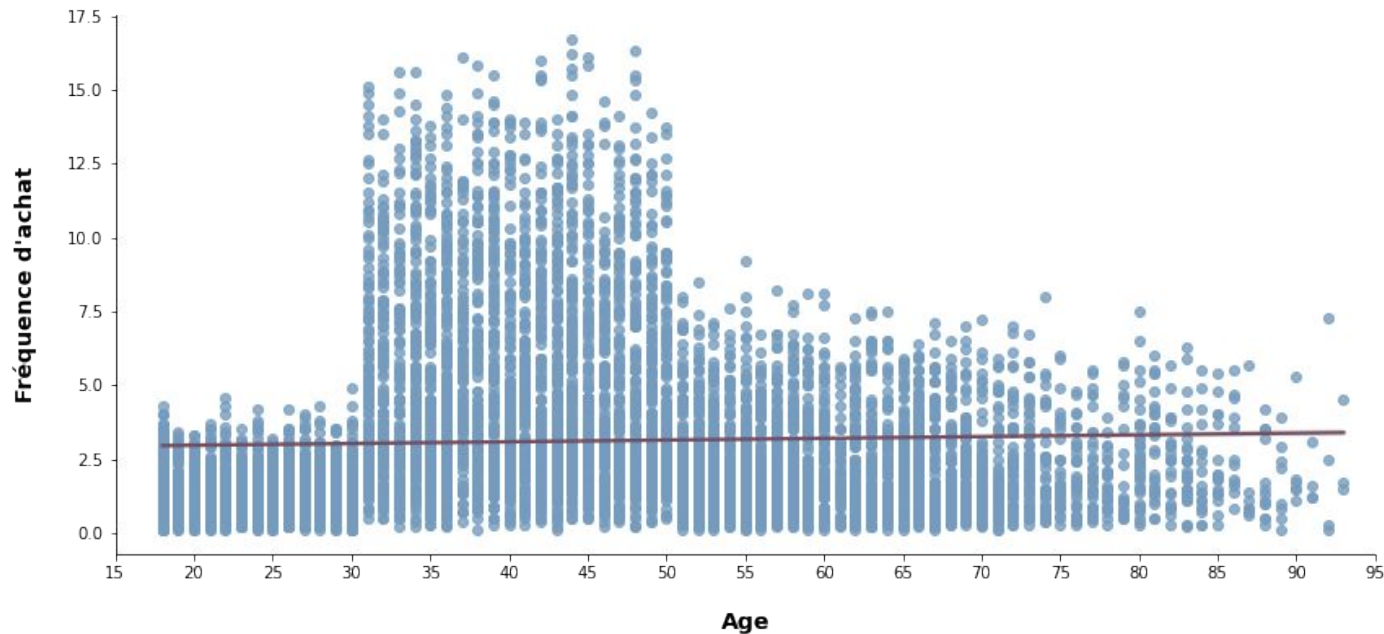


Montant plus  
dispersé pour  
les 18 à 50 ans

$$r^2 = 0,03$$

# Corrélation entre l'âge des clients et la fréquence d'achat ?

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

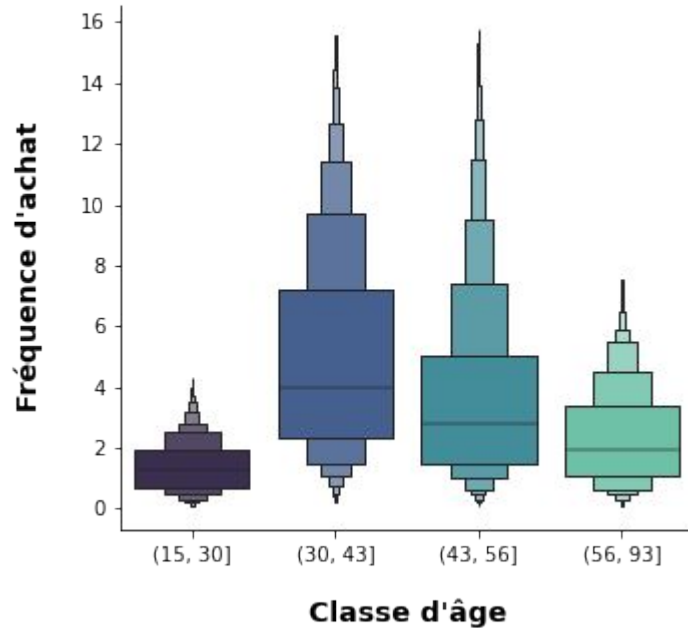


$$r^2 = 0$$

Les 30-50 ans ont une étendue de la fréquence d'achat allant jusqu'à 17

# Corrélation entre la fréquence d'achat et les classes d'âge ?

Fréquence d'achat par classe d'âge

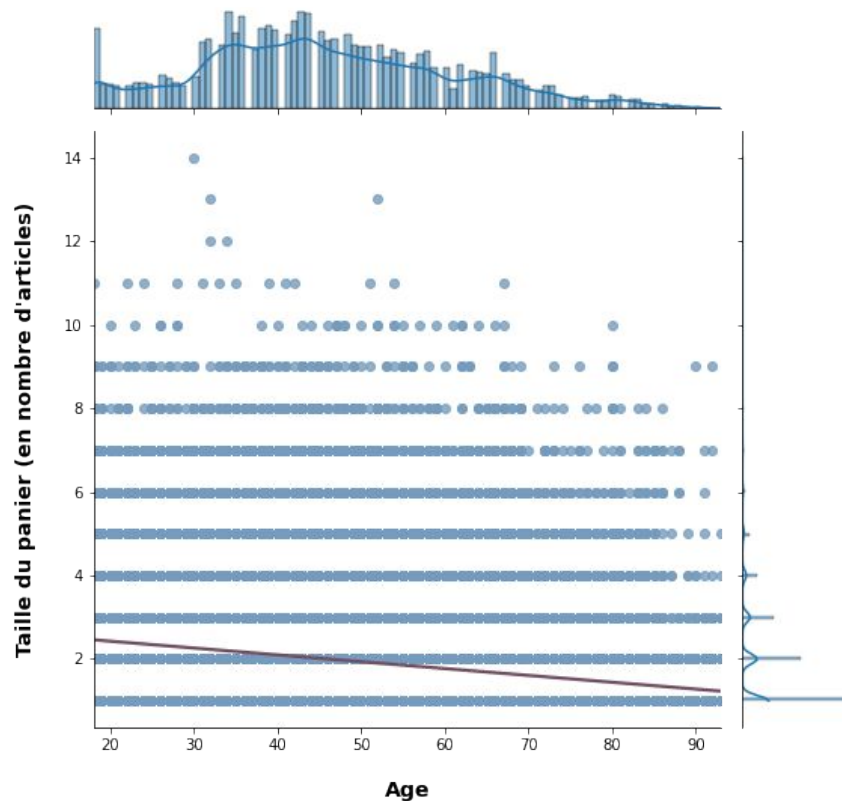


Chi-2 : variables dépendantes

30-43 ans et 43-56 ans :  
fréquence d'achats plus  
importantes

Moins de 2 achats par mois  
pour les jeunes

# Corrélation entre l'âge des clients et la taille du panier moyen?



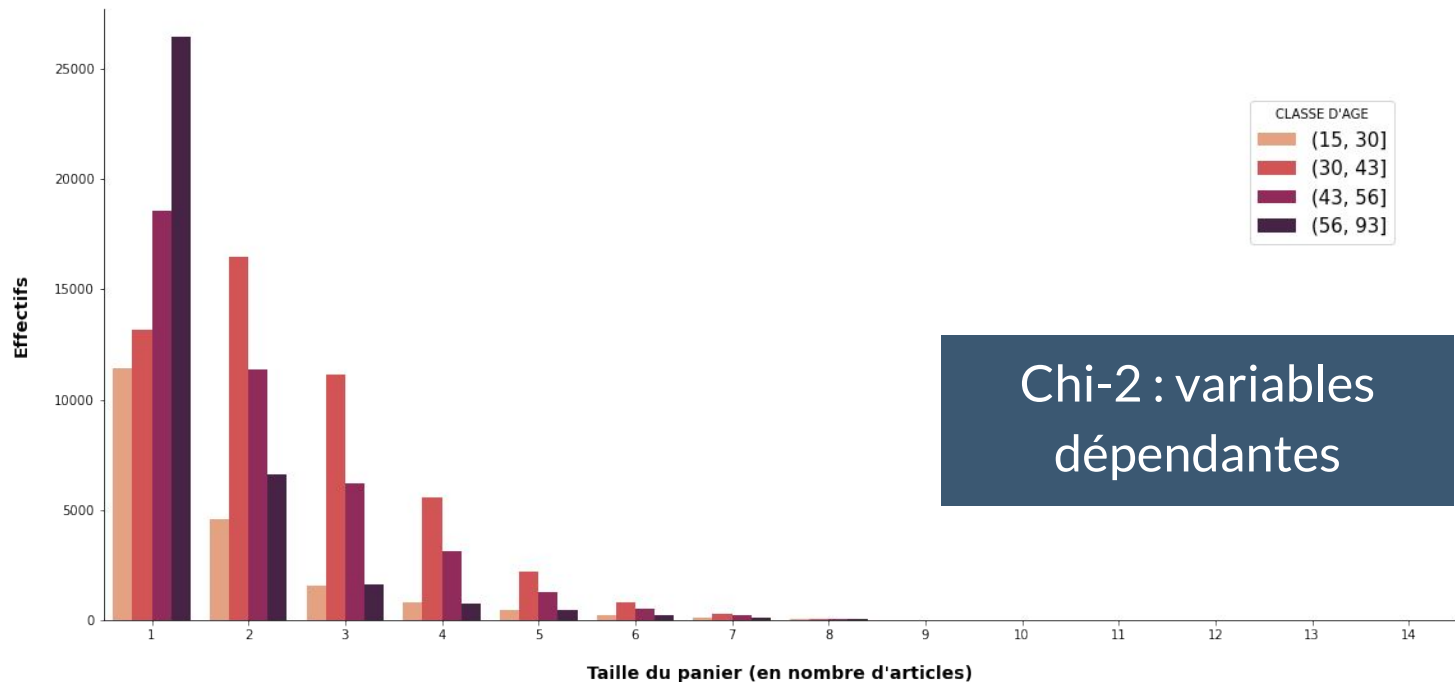
La taille du panier est majoritairement de 1 article

Coefficient de pearson :  
 $r^2 = 0,03$



# Corrélation avec les classes d'âge

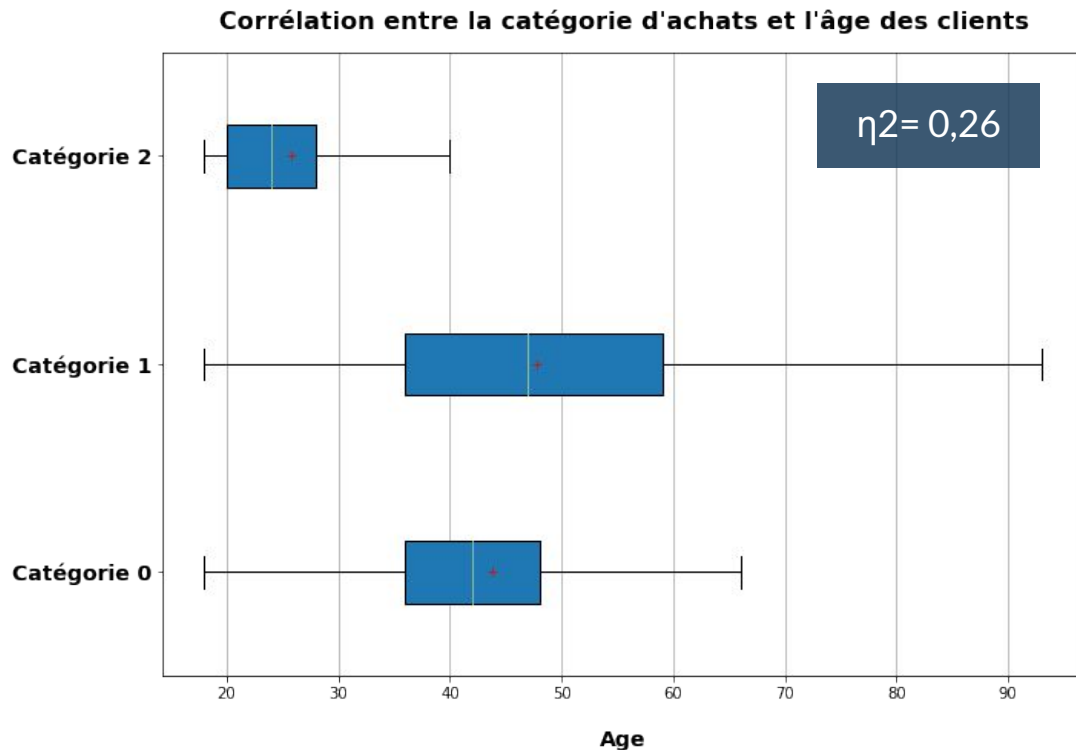
Distribution de la taille du panier par classe d'âge



Chi-2 : variables  
dépendantes

30-43 ans achètent plus fréquemment 2 ouvrages

# Corrélation entre l'âge des clients et les catégories de produits achetés ?



Cat. 2

3 clients sur 4  
âge < 28 ans

Cat. 1

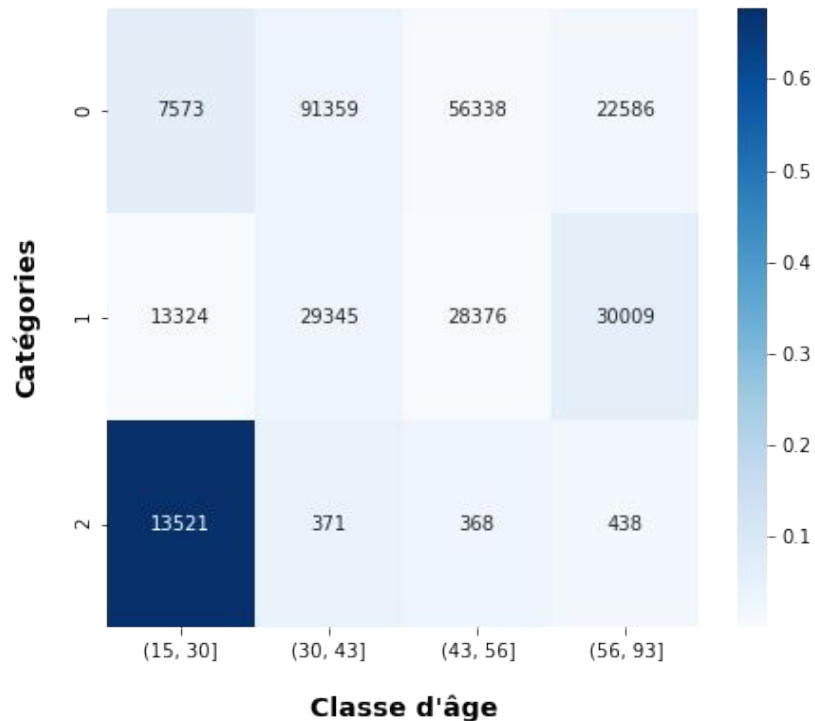
âge médian  
= 47 ans

Cat. 0

3 clients sur 4  
âge < 49 ans

# Corrélation avec les classes d'âge

Tableau de contingence



Chi-2 : variables dépendantes

Lien très significatif entre la  
catégorie 2 et les moins de 30  
ans

# Conclusions

## Ventes

- Catégorie 0 : Forte saisonnalité (septembre et décembre)
- Catégorie 1 : augmentation des ventes de novembre à janvier
- Catégorie 2 : Retour en février au niveau des ventes du mois d'août

## Chiffre d'affaires

- Augmentation depuis septembre (6 mois)
- 10% des références génèrent 60% du CA
- CA catégorie 2 très important chez les moins de 30 ans

## Préconisations

- Réaliser une opération marketing ciblant les 15 à 30 ans avec des produits de catégorie 2
- Faire une étude plus poussée sur les 4 professionnels (clientèles, besoins)