



# ANALYSE DES VENTES

Comprendre le comportement de nos clients en ligne

PRÉSENTÉ PAR FRÉDÉRIC LA ROSA

# Deux années de datas

## DÉCOUVERTE, NETTOYAGE ET ANALYSE DE NOS DATAS

Avant de fusionner l'ensemble de nos jeux de données comprenant les fichiers *customers*, *products* et *transactions*, ces étapes préliminaires nous permettent d'identifier et corriger les données altérées, inexactes ou non pertinentes. Nous pouvons alors optimiser la qualité de notre analyse.



### Clients

8 623 clients enregistrés dans notre base de données. La gente féminine représente 52% de notre clientèle mais la moyenne d'âge à 34-35 ans reste sensiblement la même entre les deux genres.



### Produits

3 287 produits différents référencés sur notre site internet, divisés en trois catégories: 0, 1 et 2. 36 références considérées comme des produits premiums.



### Transactions

Avec une première transaction effectuée le 01 mars 2021 à 00h01m07s, on compte, jusqu'à aujourd'hui, 679 332 transactions réalisées par 8 600 clients sur 3 266 produits différents.

## L'illusionniste

Le produit `O_2245` disparaît de notre fichier `products` après **21** transactions



## A l'aube du huitième ajout

Utilisation des valeurs de la variable `date` pour créer huit colonnes: **Heure, Jour, Mois, Année, Année et mois, Age lors de l'achat** ainsi que **Age lors de l'analyse**.

## Un chiffre fétiche

- **21** clients sans transaction
- **21** produits non vendus

## Testé et supprimé

Toutes données en lien avec les transactions tests ont été supprimées:  
- les clients `ct_0` et `ct_1`  
- la session `s_0`  
- le produit `T_0`

# EVOLUTION DES VENTES PAR CATÉGORIE



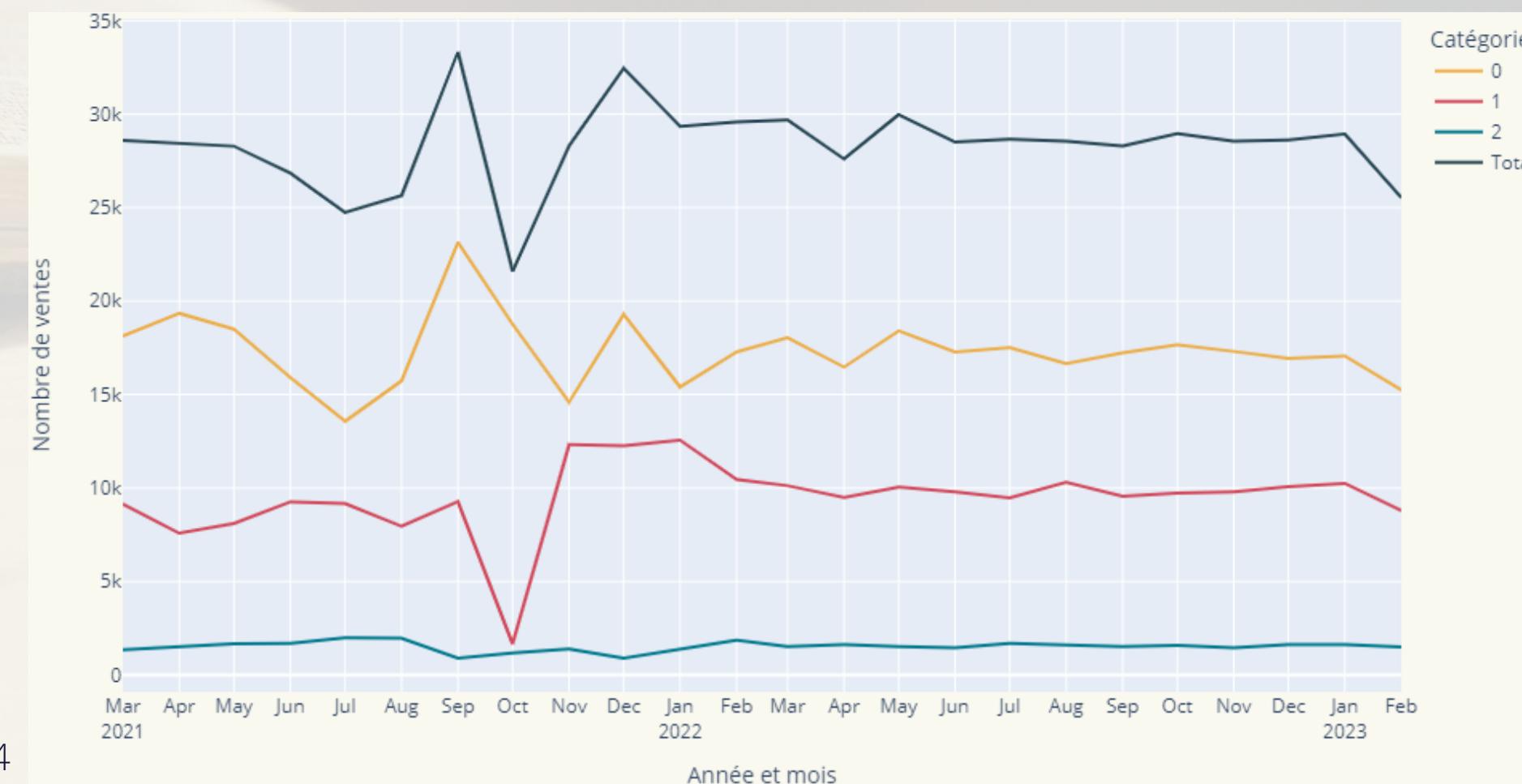
## EVOLUTION DU CHIFFRE D'AFFAIRES PAR CATÉGORIE

On compte en moyenne un chiffre d'affaires s'élevant à 493 905€ par mois.

## EVOLUTION DES VENTES PAR CATÉGORIE

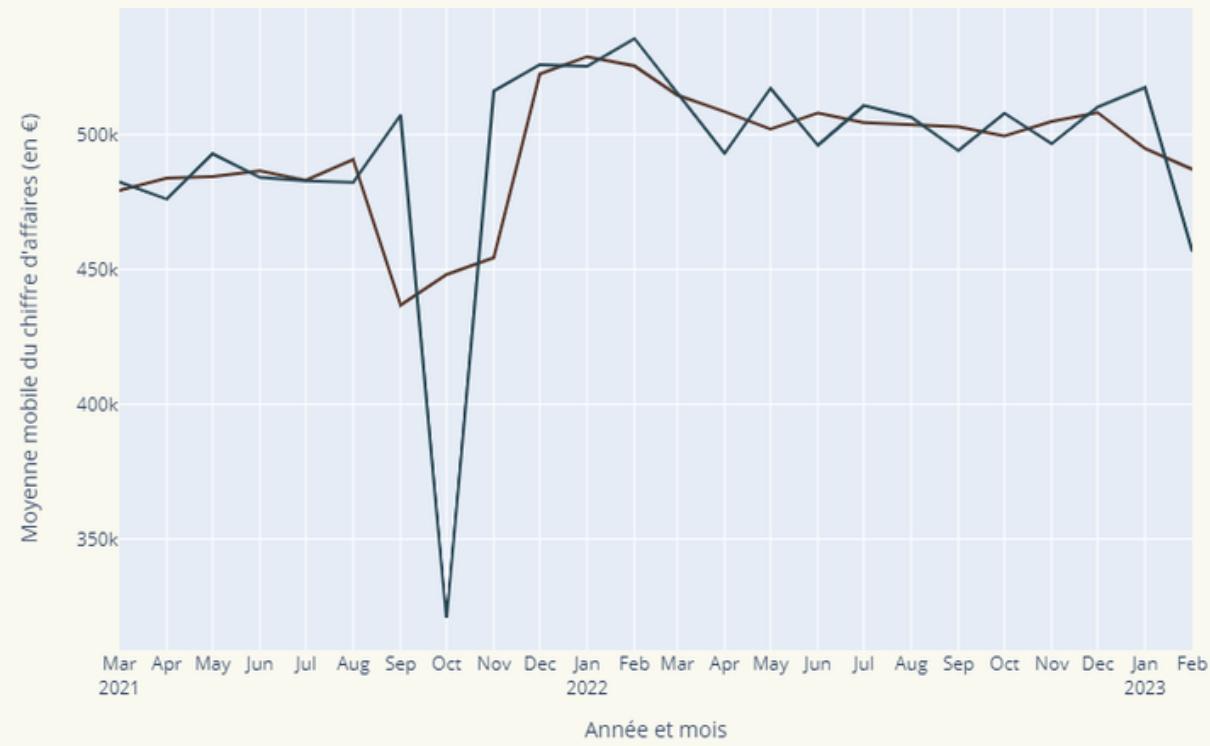
En moyenne 28 295 ventes réalisées par mois, un résultat assez constant dans le temps.

Mois record avec 33 314 ventes en septembre 2021 suivi d'un mois anormalement bas en octobre avec 21 599 ventes.

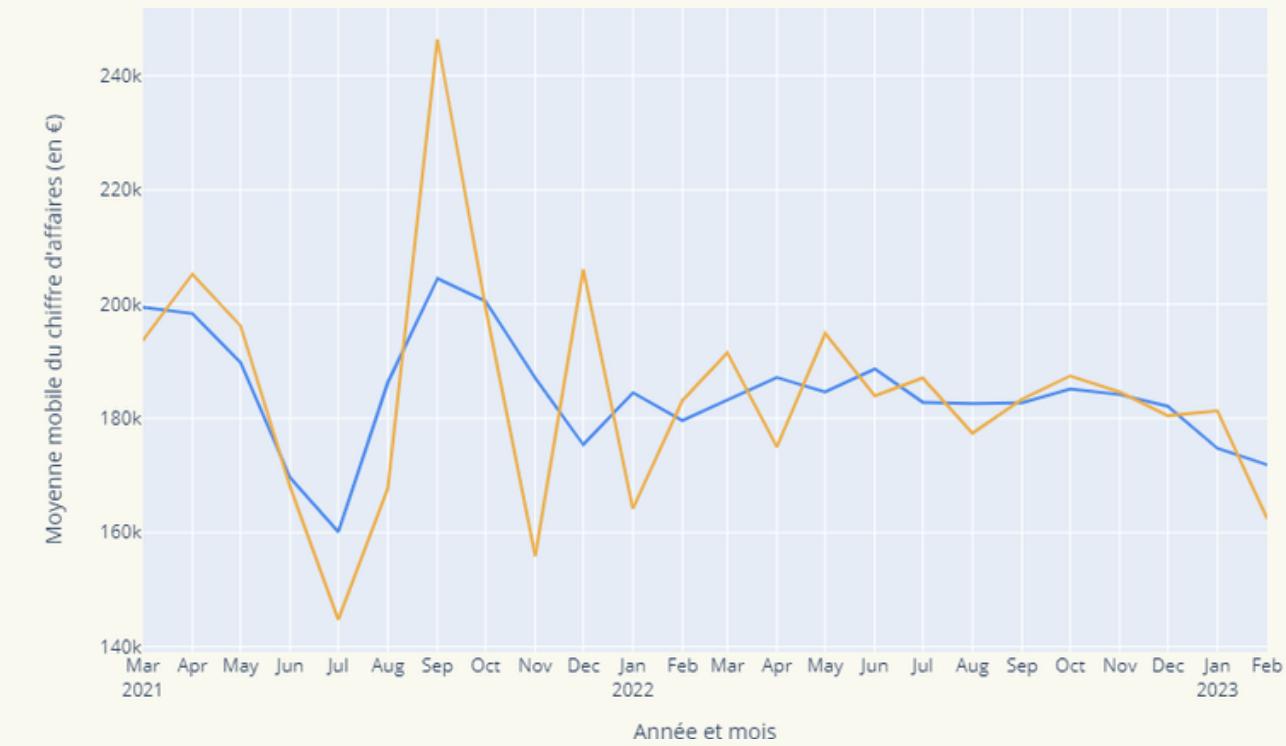


# TENDANCE GLOBALE DU CHIFFRE D'AFFAIRE

## Totale



## Catégorie 0



## Catégorie 1



## Catégorie 2

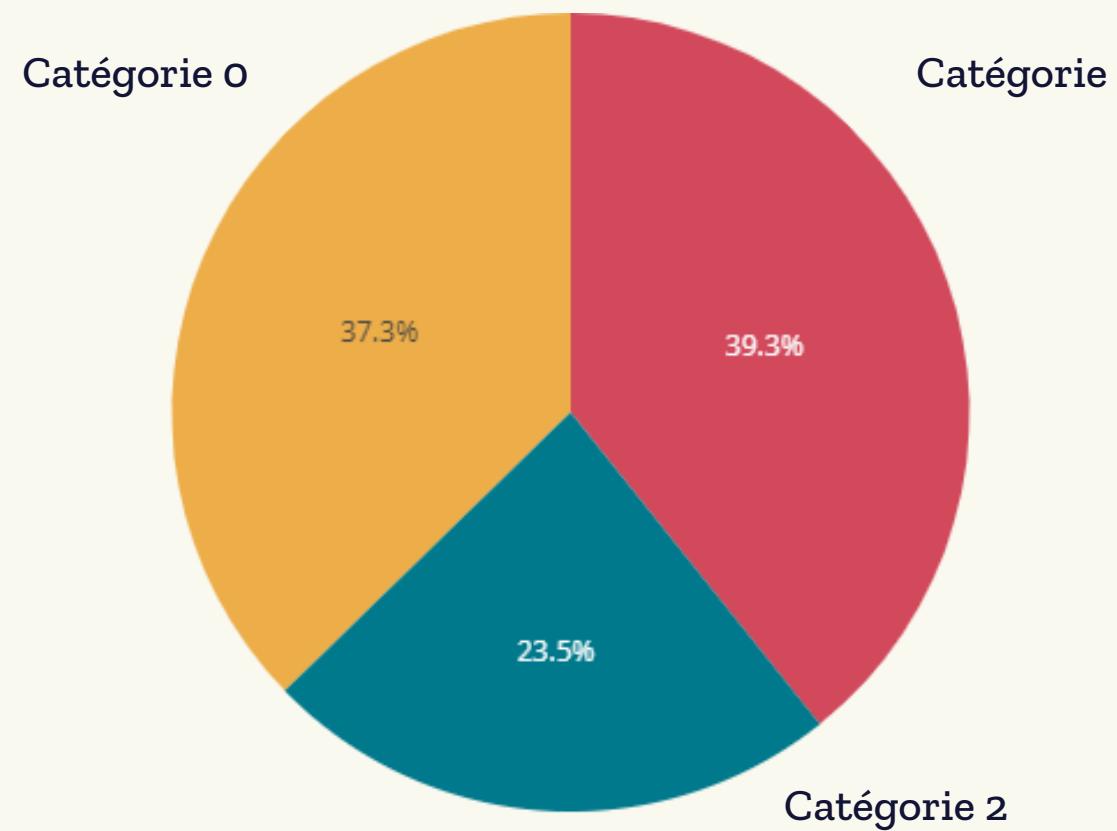


# UN SEUL ACCIDENT TECHNIQUE SURVENU EN DEUX ANS ?

Aucune transaction concernant la catégorie 1 n'a été réalisée du 02 au 27 octobre 2021 inclus. Il est fort probable qu'un accident technique ait eu lieu sur notre site durant cette période. Nous contacterons le service concerné pour obtenir le journal d'erreurs.

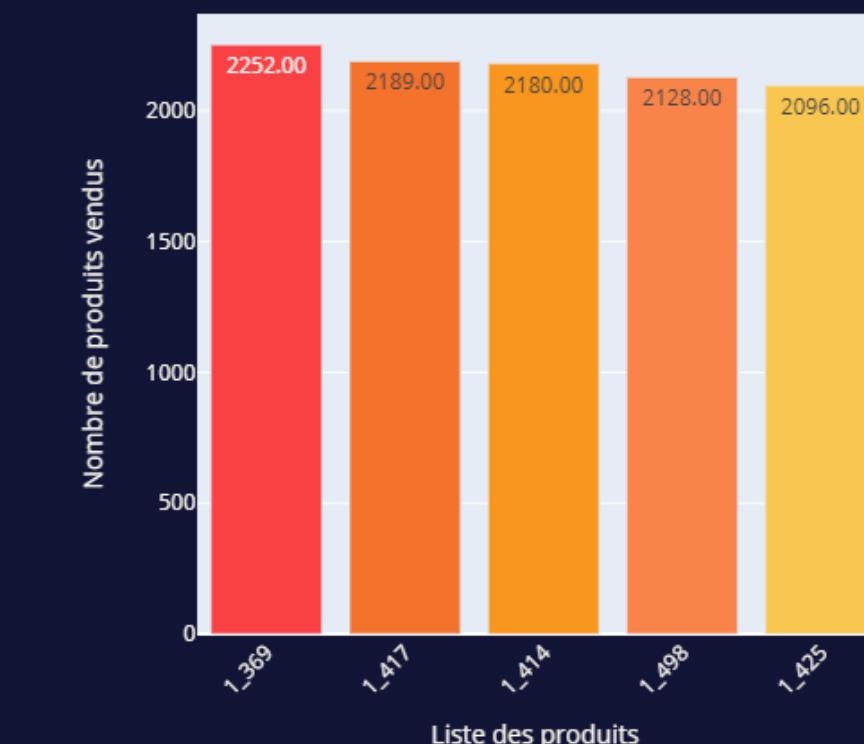


**En deux ans,  
11 853 728.68€ de  
chiffre d'affaires**

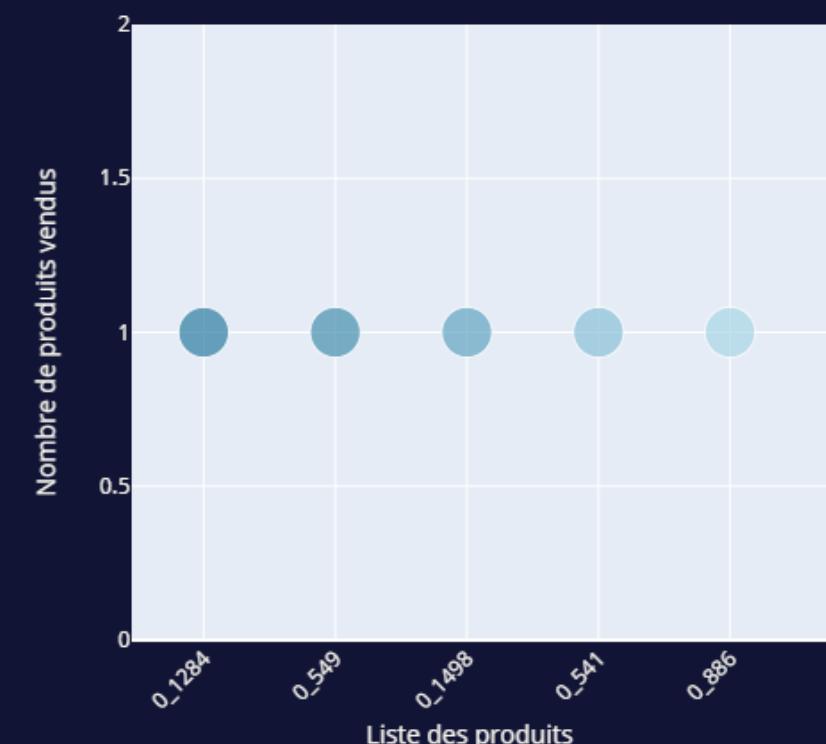


**LES CATÉGORIES DE PRODUITS  
RAPPORTANT LE PLUS SONT LES 0 ET  
LES 1, AVEC EN MOYENNE PAR MOIS  
184 155€ ET 193 905€, CONTRE 115  
845€ POUR LA 2.**

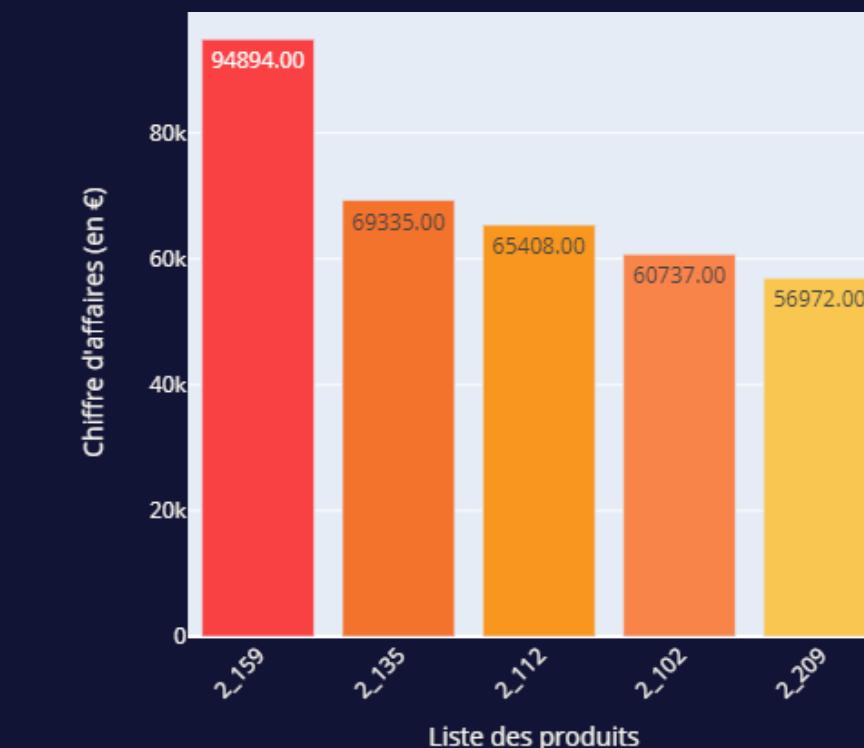
**Top 5 des produits vendus**



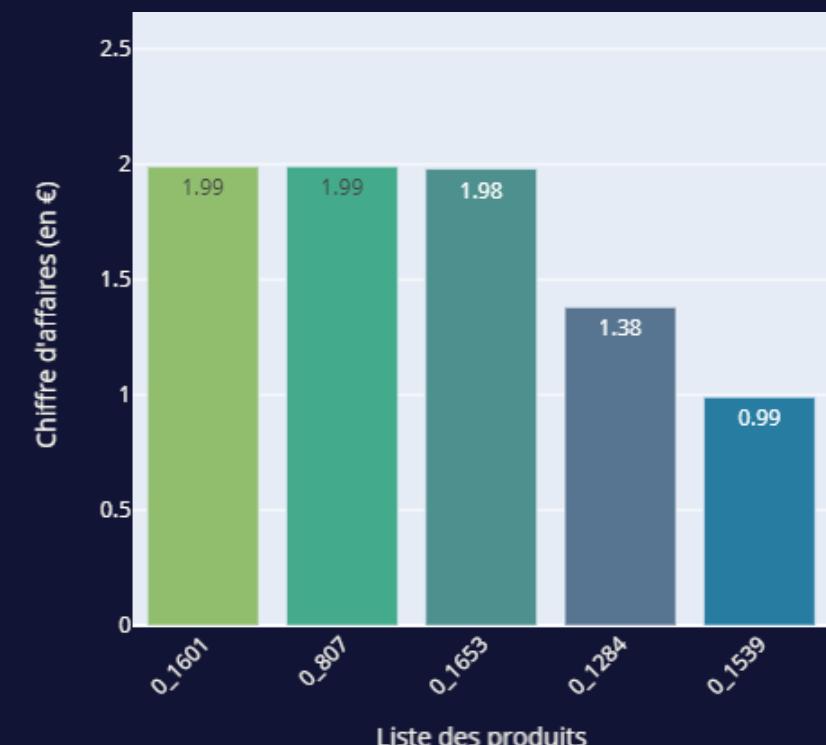
**Flop 5 des produits vendus**



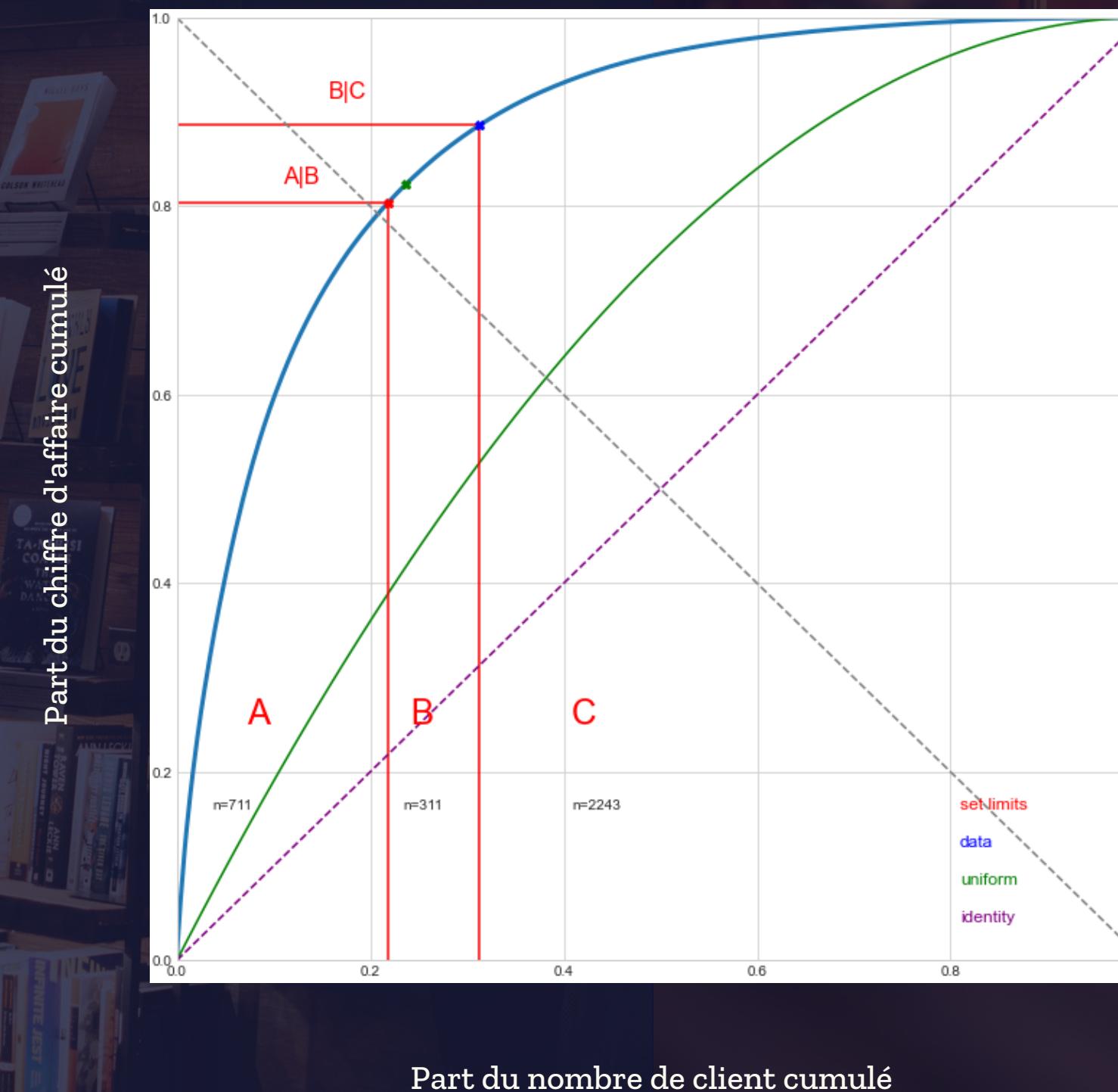
**Top 5 des produits au gros CA**



**Flop 5 des produits au gros CA**



# 20% DE NOS PRODUITS SONT RESPONSABLES DE 80% DE NOTRE CHIFFRE D'AFFAIRES

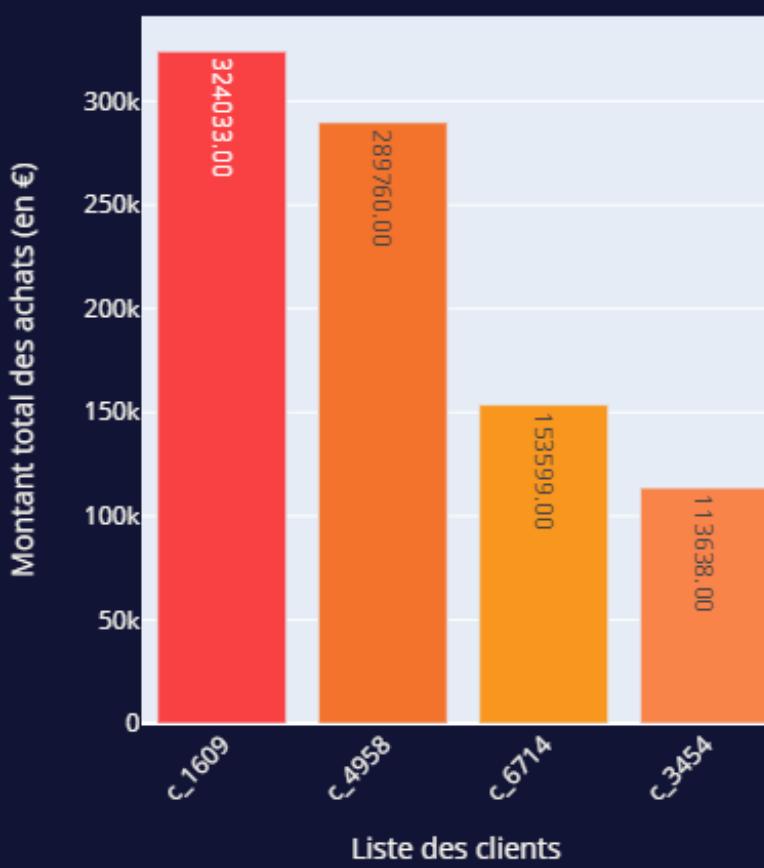




# Des professionnels parmi nos clients

UN VOLUME D'ACHAT EN MOYENNE 5 829 FOIS PLUS GRAND QUE L'ENSEMBLE DE NOTRE CLIENTÈLE

Variant selon le client, entre 6 et 13 exemplaires achetés en moyenne.  
Jusqu'à 25 488 produits achetés par un client en deux ans.



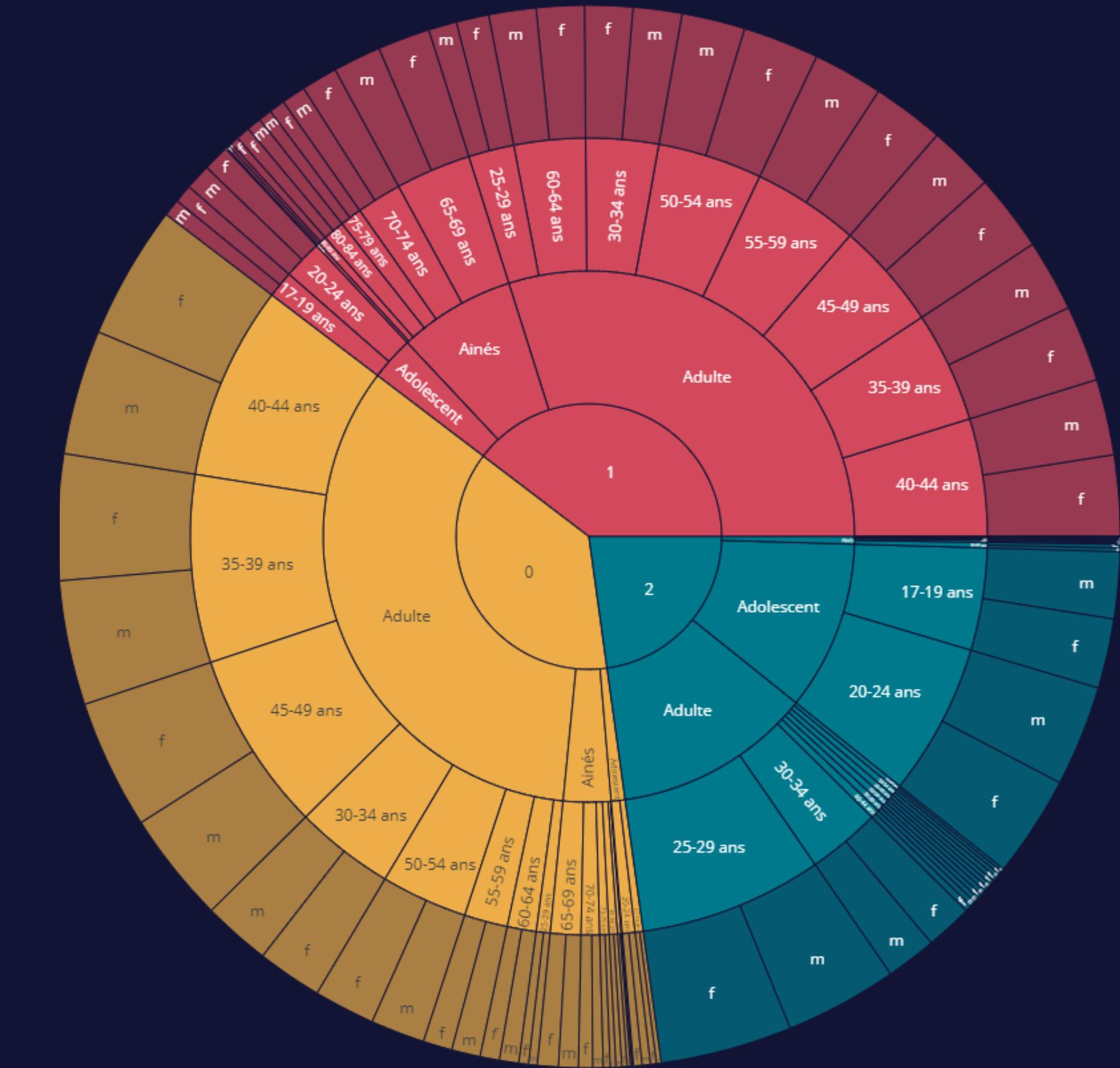
# 50% DE NOTRE CHIFFRE D'AFFAIRES EST RÉALISÉ PAR 21% DE NOTRE CLIENTÈLE



7.5% du chiffre d'affaires est réalisé par seulement 1% de notre clientèle

# ANALYSE DU COMPORTEMENT DE NOS CLIENTS EN LIGNE

Une forte concentration de notre clientèle est âgée entre 30 et 50 ans, cette dernière tendant à décroître avec l'âge. Même si l'âge des premiers achats enregistré est de 17 ans, on observe un fort pic à 18 ans.



# Tests de corrélation

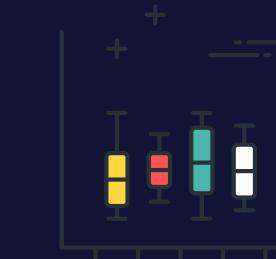
PEUT-ON PRÉDIRE LES ACHATS DE NOS CLIENTS ?

Le test de corrélation est utilisé pour évaluer une association (dépendance) entre deux variables. Le calcul du coefficient de corrélation peut être effectué en utilisant différentes méthodes.



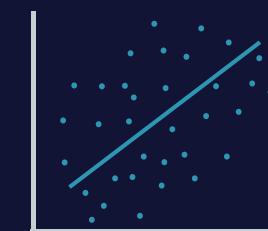
## Chi<sup>2</sup>

Un test d'indépendance vérifie le lien entre deux variables qualitatives (ex: le genre d'un client et les catégories de livres achetés).



## ANOVA

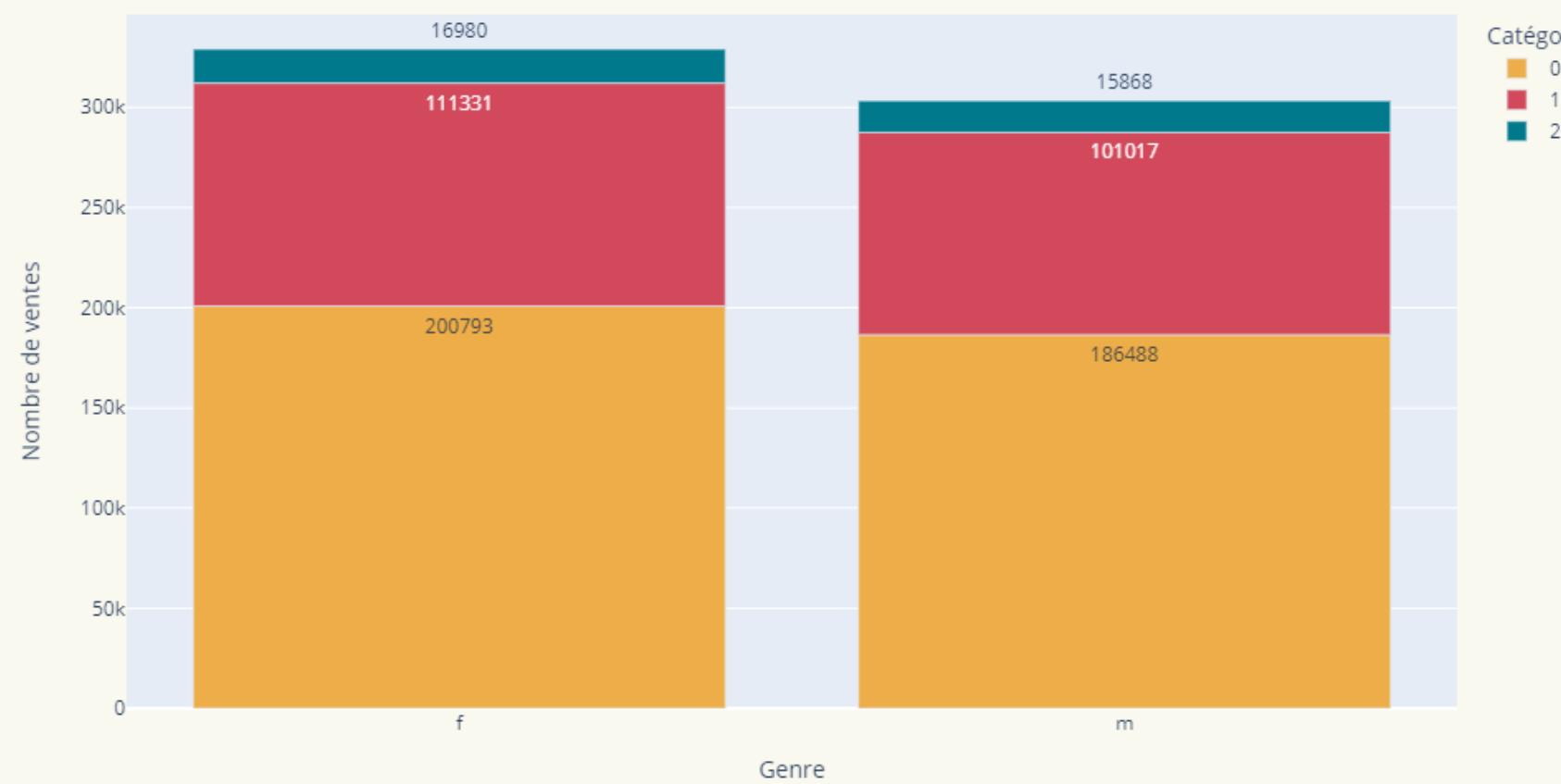
L'analyse de la variance sert à mettre en avant des différences ou des dépendances entre une variable quantitative et une ou deux variables qualitatives (ex: la tranche d'âge des clients et les catégories de livres achetés).



## Régression linéaire

L'analyse de régression linéaire sert à prévoir la valeur d'une variable quantitative en fonction de la valeur d'une autre variable quantitative (ex: l'âge des clients et le montant des achats).

## DISTRIBUTION DES VENTES PAR CATÉGORIE SELON LE GENRE DE LA CLIENTÈLE



## HEATMAP DU TABLEAU DE CONTINGENCE

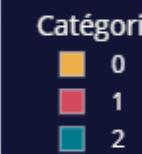
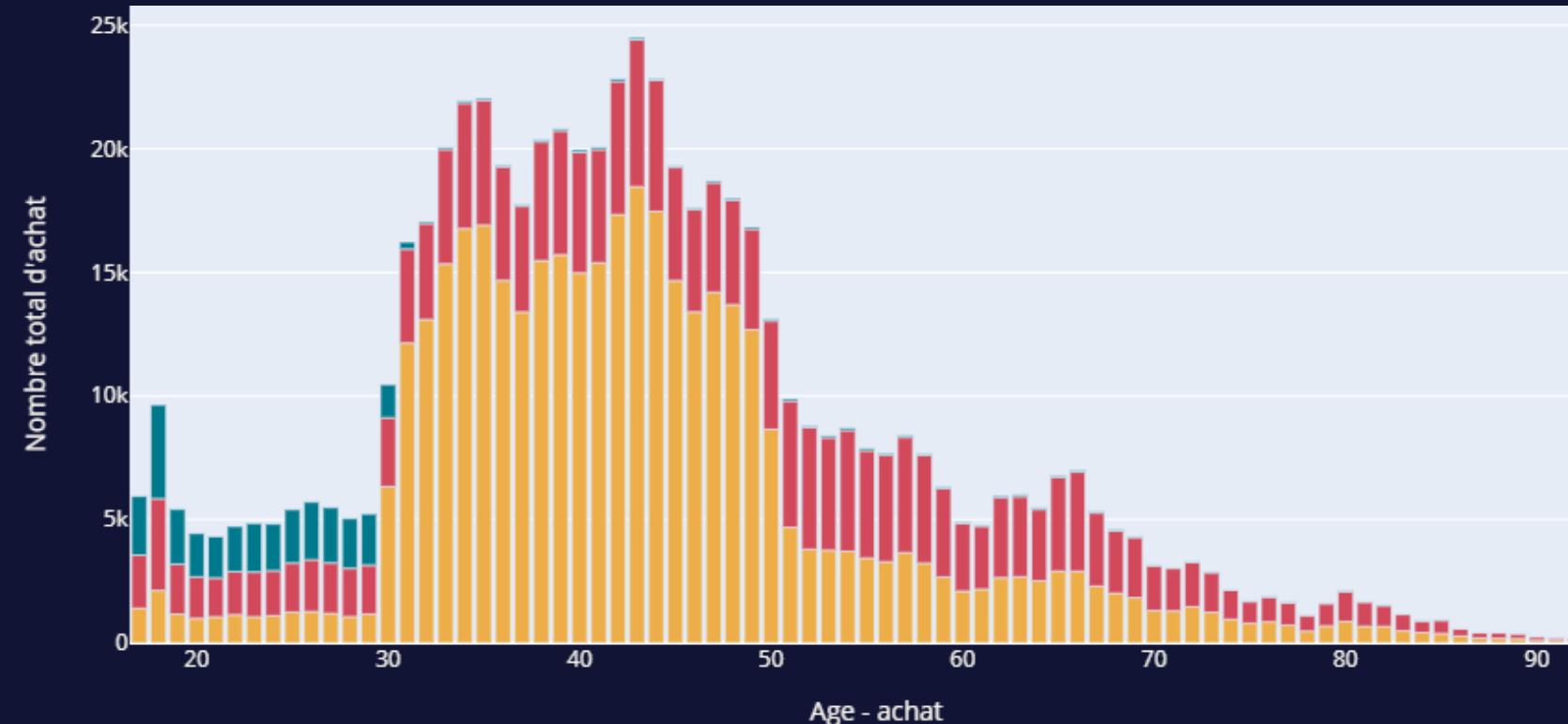


TEST DU CHI2:  
LIEN ENTRE LE GENRE ET  
LES CATÉGORIES DE  
LIVRES ACHETÉS



Un lien est observé entre le genre et la catégorie 1, les hommes semblent être plus hermétiques à cette catégorie que les femmes.

# LIEN ENTRE L'ÂGE ET LES CATÉGORIES DE LIVRES



## DISTRIBUTION DU NOMBRE TOTAL D'ACHAT PAR CATÉGORIE SELON L'ÂGE

Une forte concentration de notre clientèle est âgée entre 30 et 50 ans, cette dernière tendant à décroître avec l'âge. Même si l'âge des premiers achats enregistré est de 17 ans, on observe un fort pic à 18 ans.

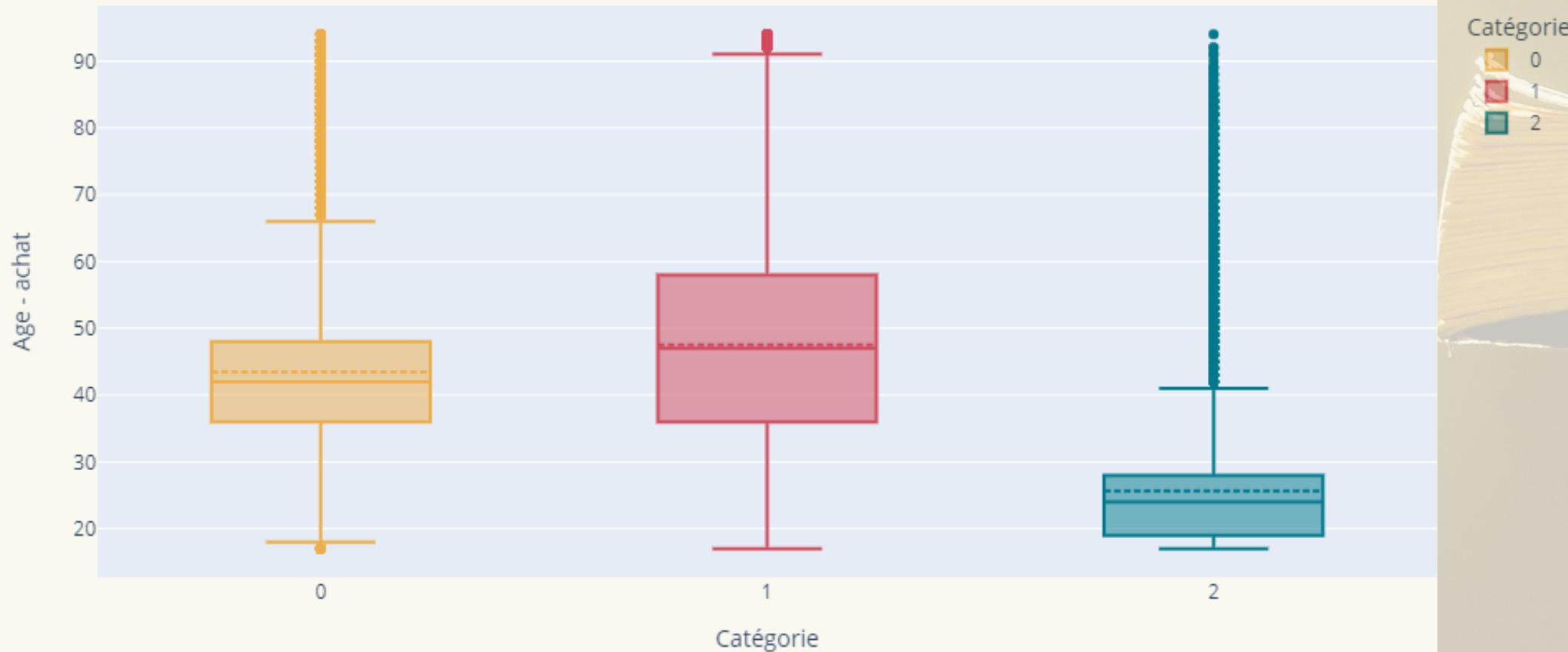


## HEATMAP DU TABLEAU DE CONTINGENCE

Un lien fort existe entre la tranche d'âge des 17-29 ans et l'achat de produits de catégorie 2.

# ANALYSE BIVARIÉE LIEN ENTRE L'ÂGE DES CLIENTS ET LES CATÉGORIES DES LIVRES ACHETÉS

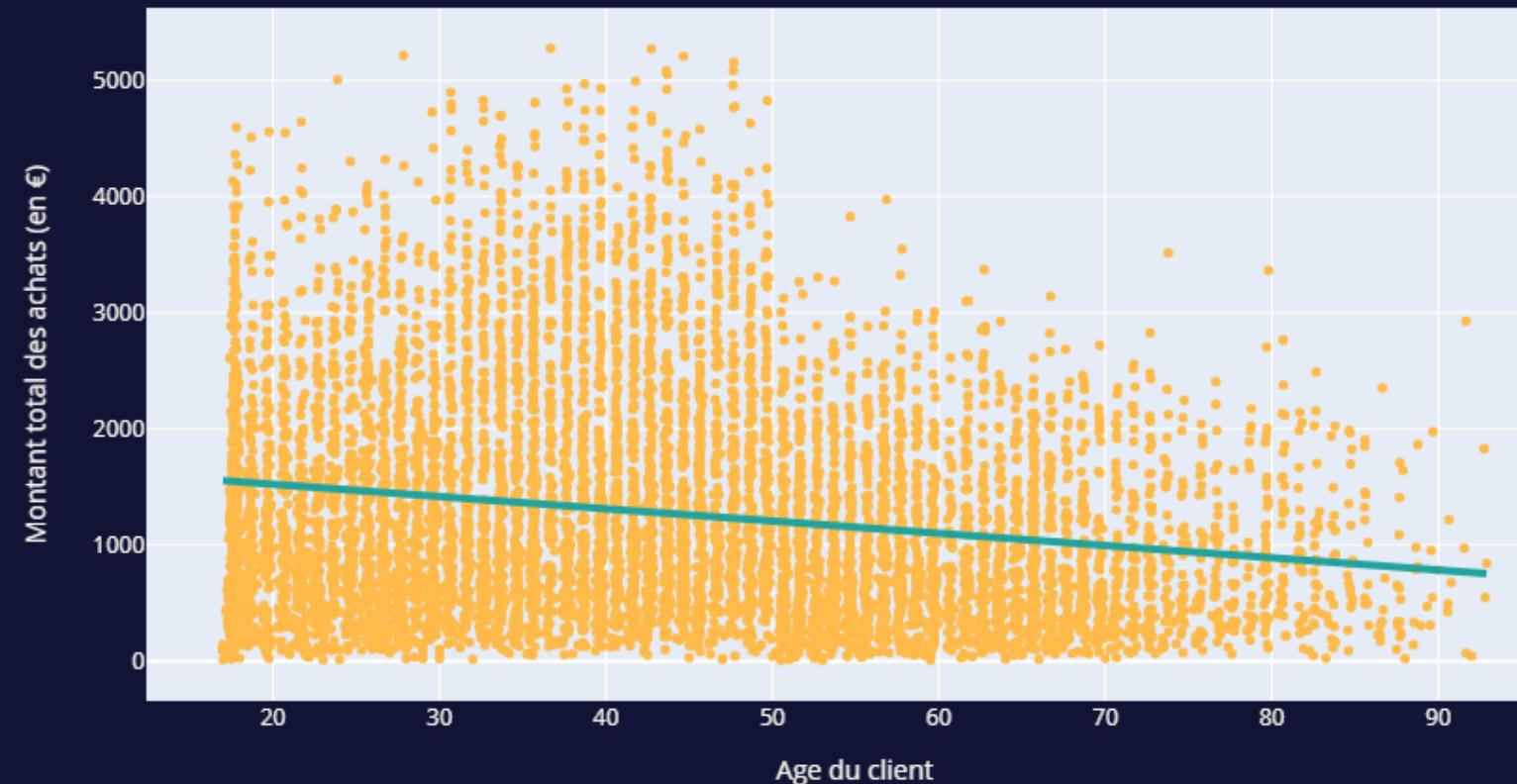
RÉPARTITION DE L'ÂGE DES CLIENTS PAR CATÉGORIE DE LIVRES ACHETÉS



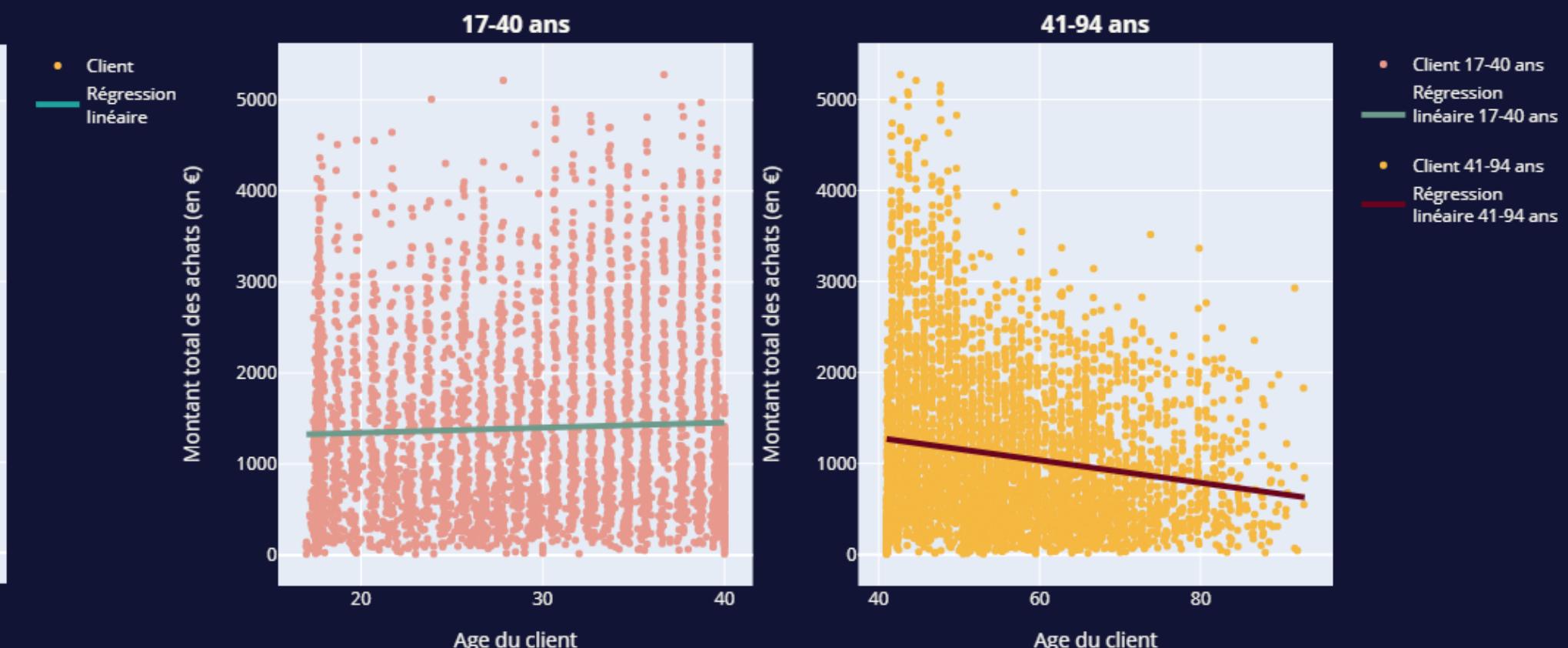
L'analyse de la variance révèle qu'un lien existe mais l'âge nous donne 11% des informations nécessaires pour pouvoir prédire à 100% la dispersion de nos données.

# LIEN ENTRE L'ÂGE DES CLIENTS ET LE MONTANT DES ACHATS

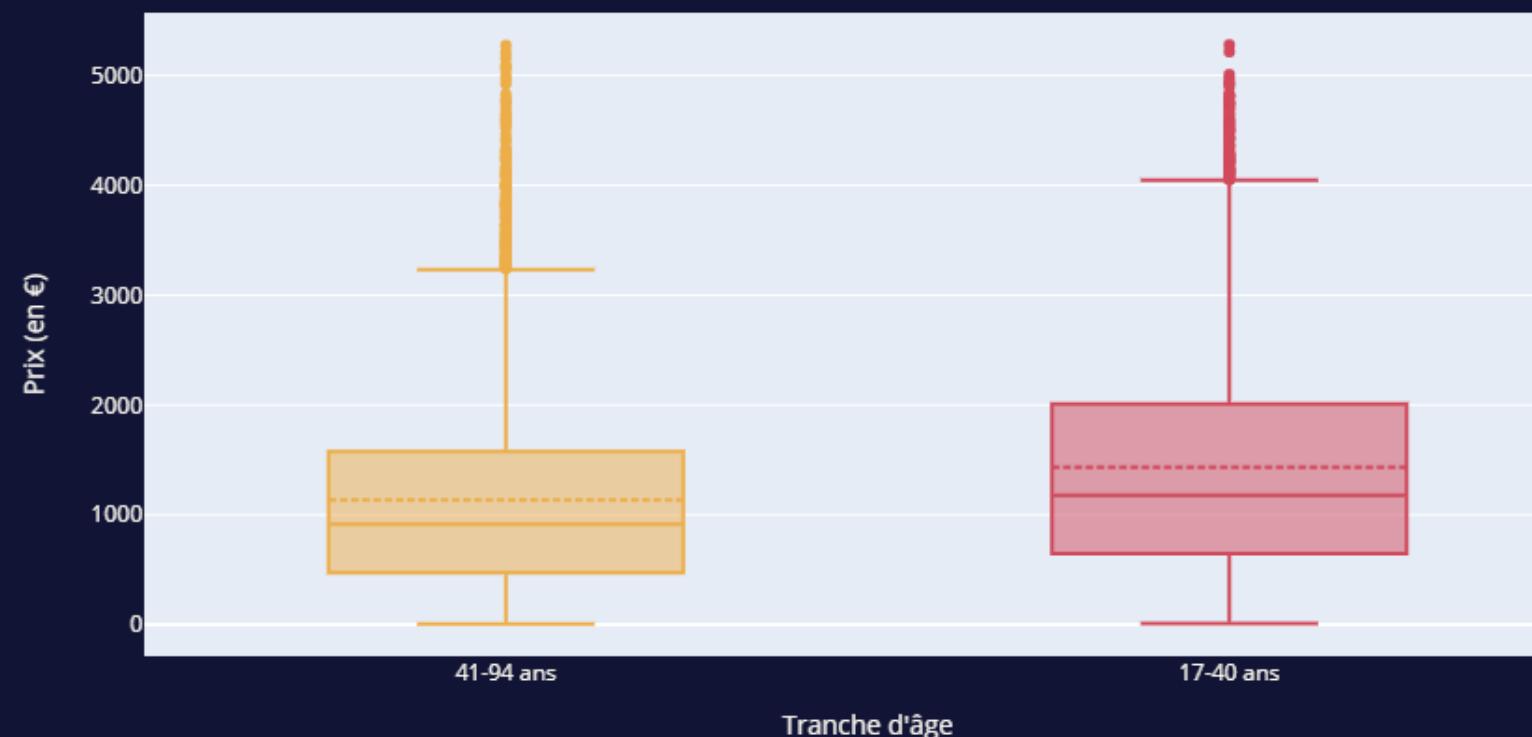
Dispersion et prédition du montant total des achats en fonction de l'âge des clients



Dispersion et prédition du montant total des achats par tranche d'âge



Répartition du montant des achats par la tranche d'âge



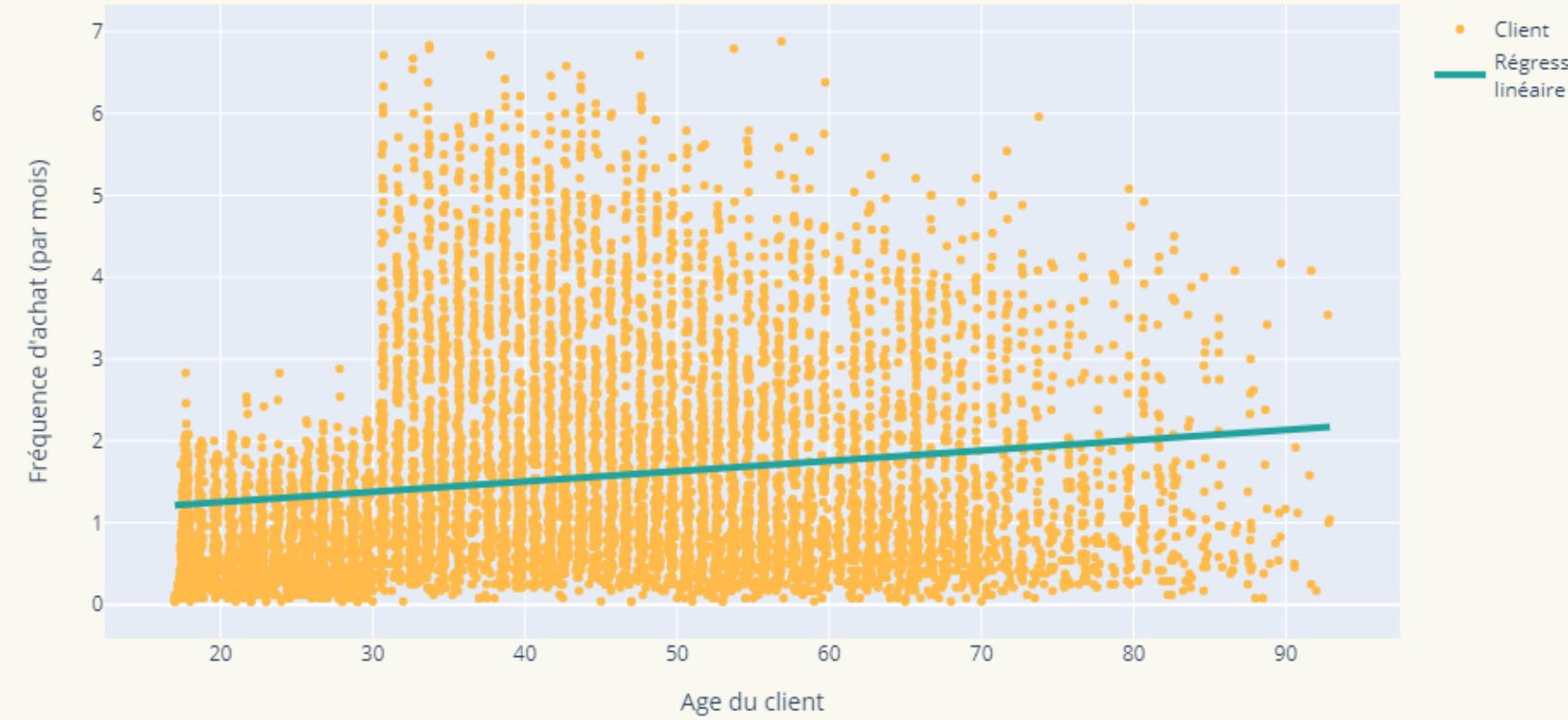
Tranche d'âge  
41-94 ans  
17-40 ans



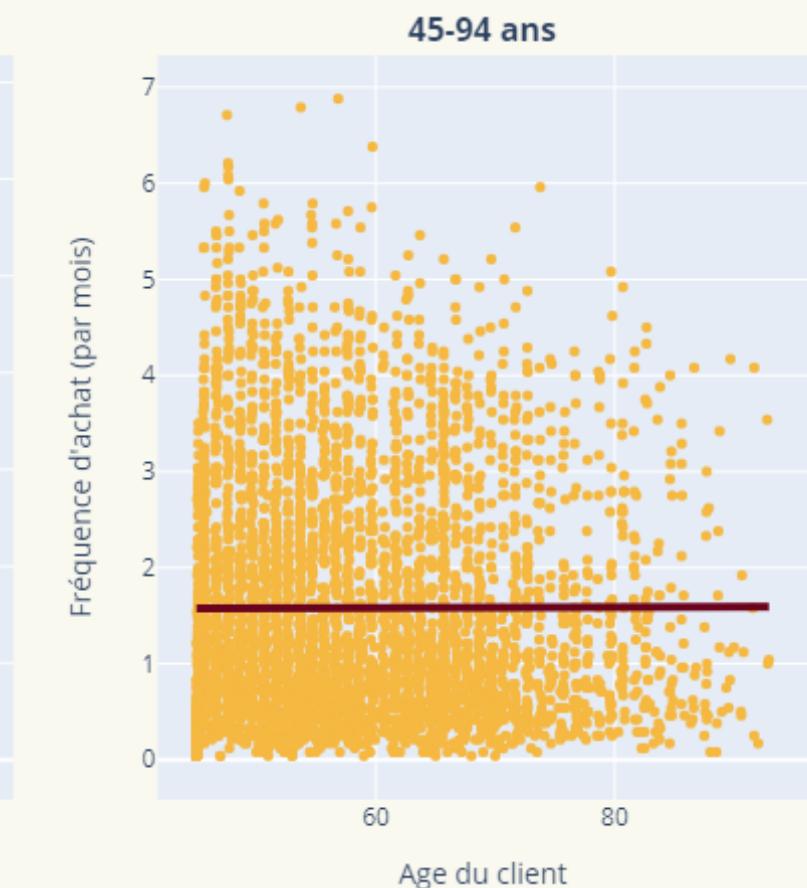
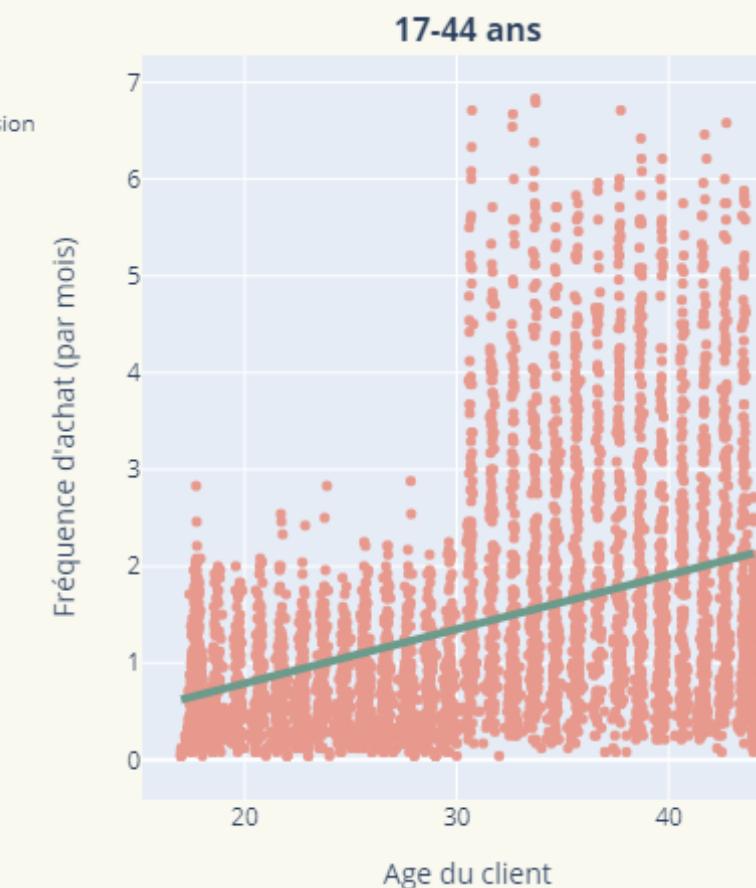
Même si un lien existe bien, le coefficient de détermination nous indique que la catégorie d'âge nous donne 0.02% des informations nécessaires pour pouvoir prédire à 100% la dispersion de nos données.

# LIEN ENTRE L'ÂGE DES CLIENTS ET LA FRÉQUENCE D'ACHAT

Dispersion et prédition de la fréquence d'achat en fonction de l'âge des clients



Dispersion et prédition de la fréquence d'achat par tranche d'âge

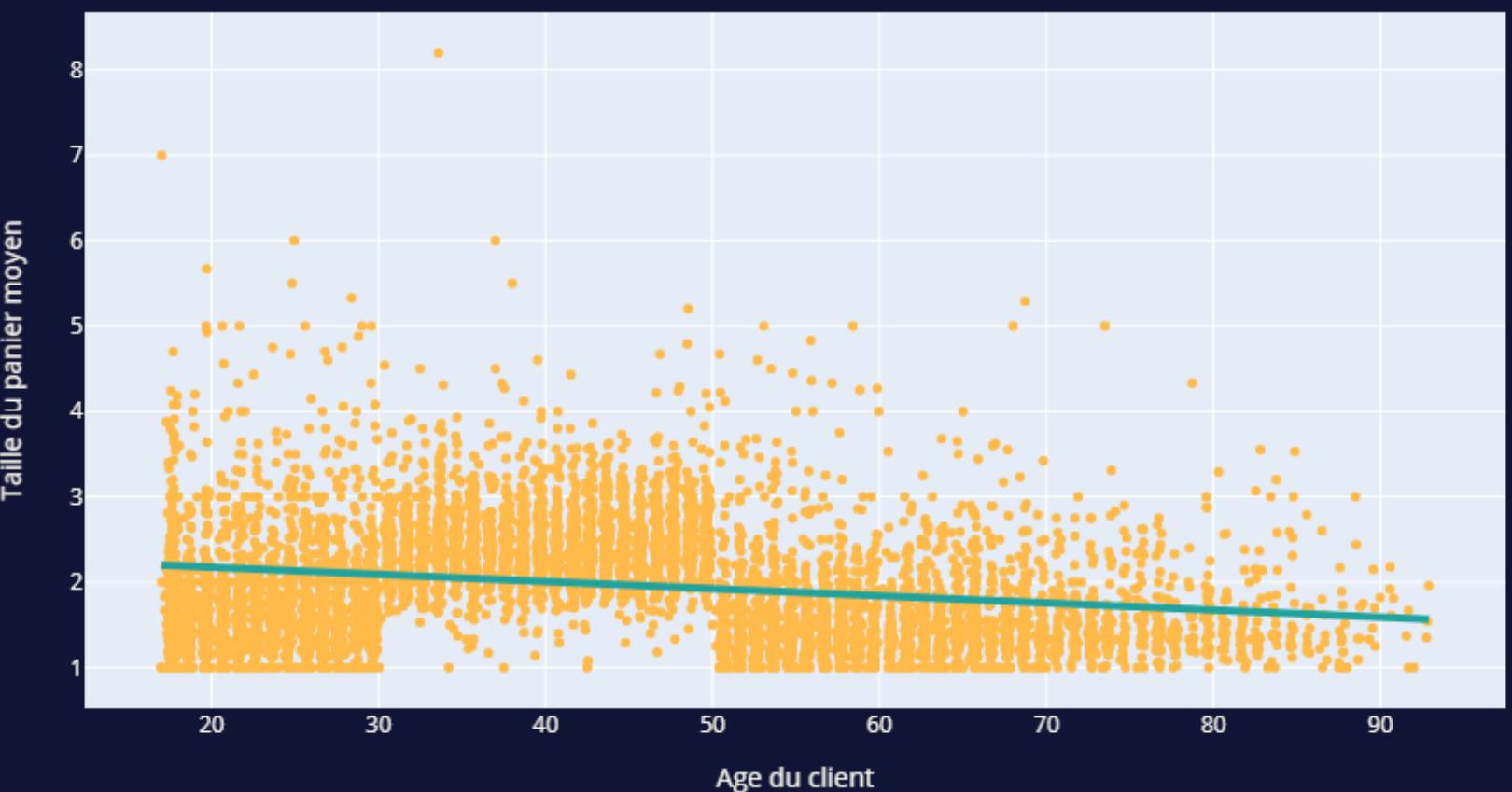


Répartition de la fréquence d'achat par la tranche d'âge

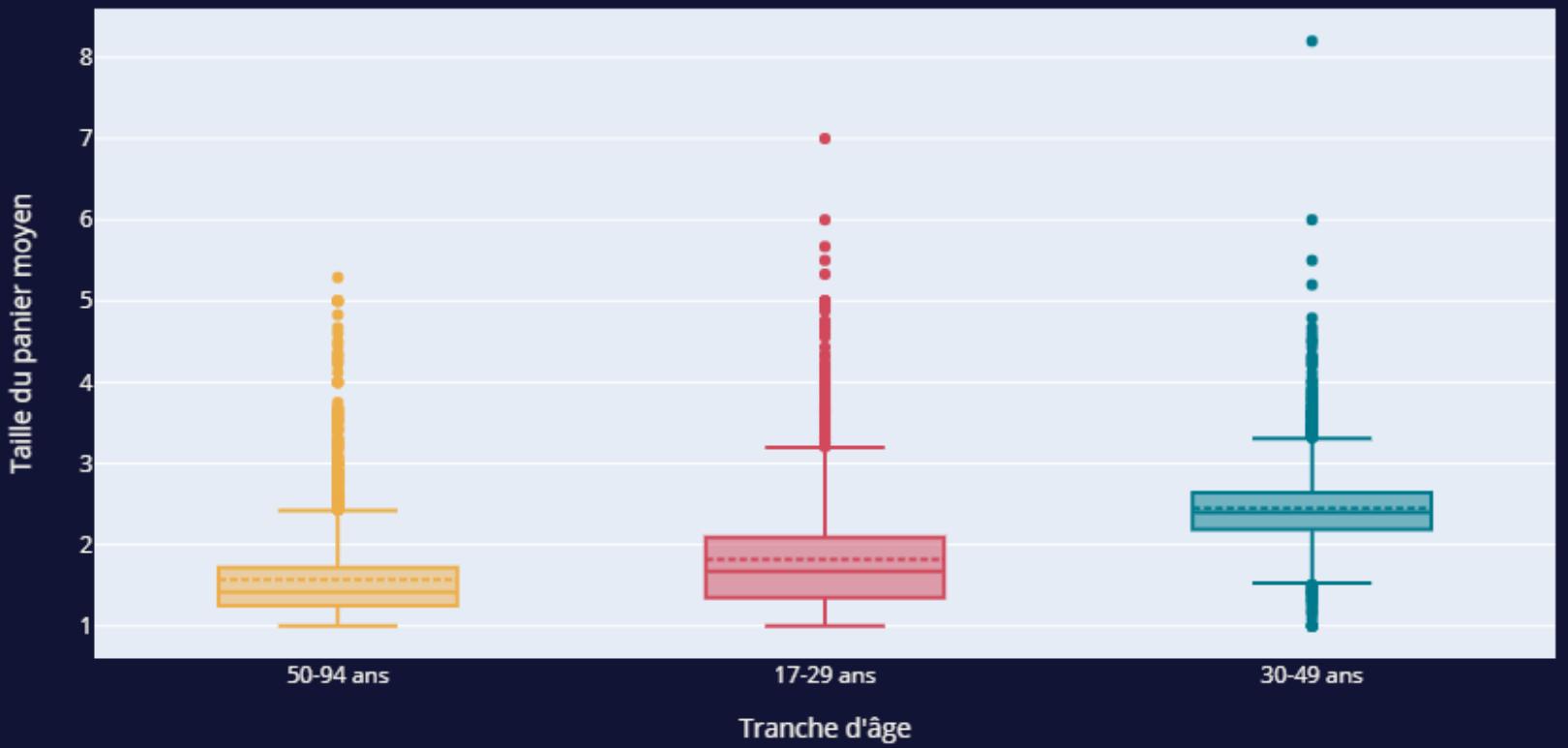


Un lien existe entre la fréquence d'achat et l'âge de nos clients mais il ne permet aucune prédition. La fréquence d'achat est ascendante entre 17 et 44 ans avec une net augmentation à partir de 32 ans. La fréquence d'achat moyenne est légèrement plus forte chez les 45-94 ans mais semble être stagnante dans le temps.

## DISPERSION ET PRÉDICTION DE LA TAILLE DU PANIER MOYEN EN FONCTION DE L'ÂGE DES CLIENTS



## RÉPARTITION DE LA TAILLE DU PANIER MOYEN PAR LA TRANCHE D'ÂGE



La catégorie d'âge des clients nous donne 34% de l'information nécessaire pour prédire à 100% la taille du panier.

La corrélation est confirmée par l'ANOVA en divisant la variable âge en trois tranches: les 17-29 ans, les 30-49 ans qui en moyenne possède un peu plus de deux articles dans leurs paniers et les 50-94 ans qui possède le plus petit panier moyen.