

# ANALYSE ENTREPRISE LA PAGE

M2 Data Analytics et Manager Marketing  
Thi Mai Hien VU

COURS DE STATISTIQUES DÉCISIONNELLES

INSEEC - MSC2 Data Analytics & Marketing Manager

# SOMMAIRE

- Contexte
- Analyses
  - Répartition des clients par sexe
  - Répartition des clients par tranche d'âge
  - Panier moyen par tranche d'âge
  - Répartition des ventes par tranche d'âge
  - Fréquence d'achat par tranche d'âge
  - Évolution du chiffre d'affaires entre 2021 et 2023
  - Évolution du nombre de commandes par mois
  - Top 20 des produits les plus vendus
  - Répartition du chiffre d'affaires par tranche d'âge et catégorie
  - Répartition du nombre de ventes par sexe et catégories
  - Courbe de Lorenz
- Scoring RFM
- Clustering (K-means)
- Conclusion

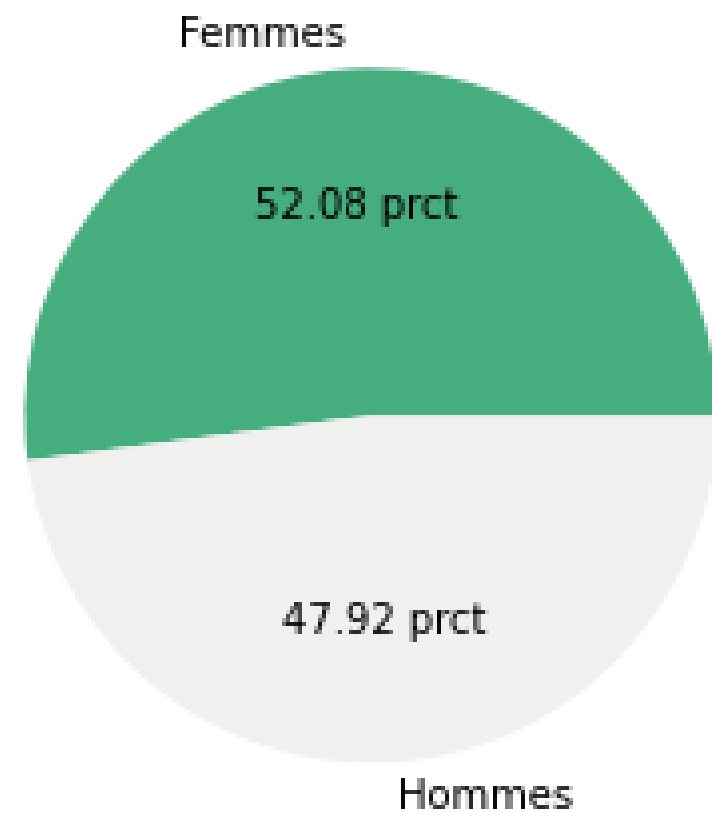
# Contexte

Nous avons été mandatés par l'entreprise Lapage pour réaliser une analyse complète de ses différentes bases de données.

L'entreprise souhaite :

- Le nettoyage des données
- Une série d'indicateurs statistiques de tendances centrales
- La réalisation d'analyses (univariées bivariées)
  - Genre du client et catégorie de produits achetés.
  - Âge des clients et montant total des achats.
  - Âge des clients et fréquence d'achat.
  - Âge des clients et panier moyen.
  - Âge des clients et catégorie de produits achetés.
  - Une courbe de Lorenz
- La segmentation et le clustering de la clientèle

# Répartition des clients par sexe



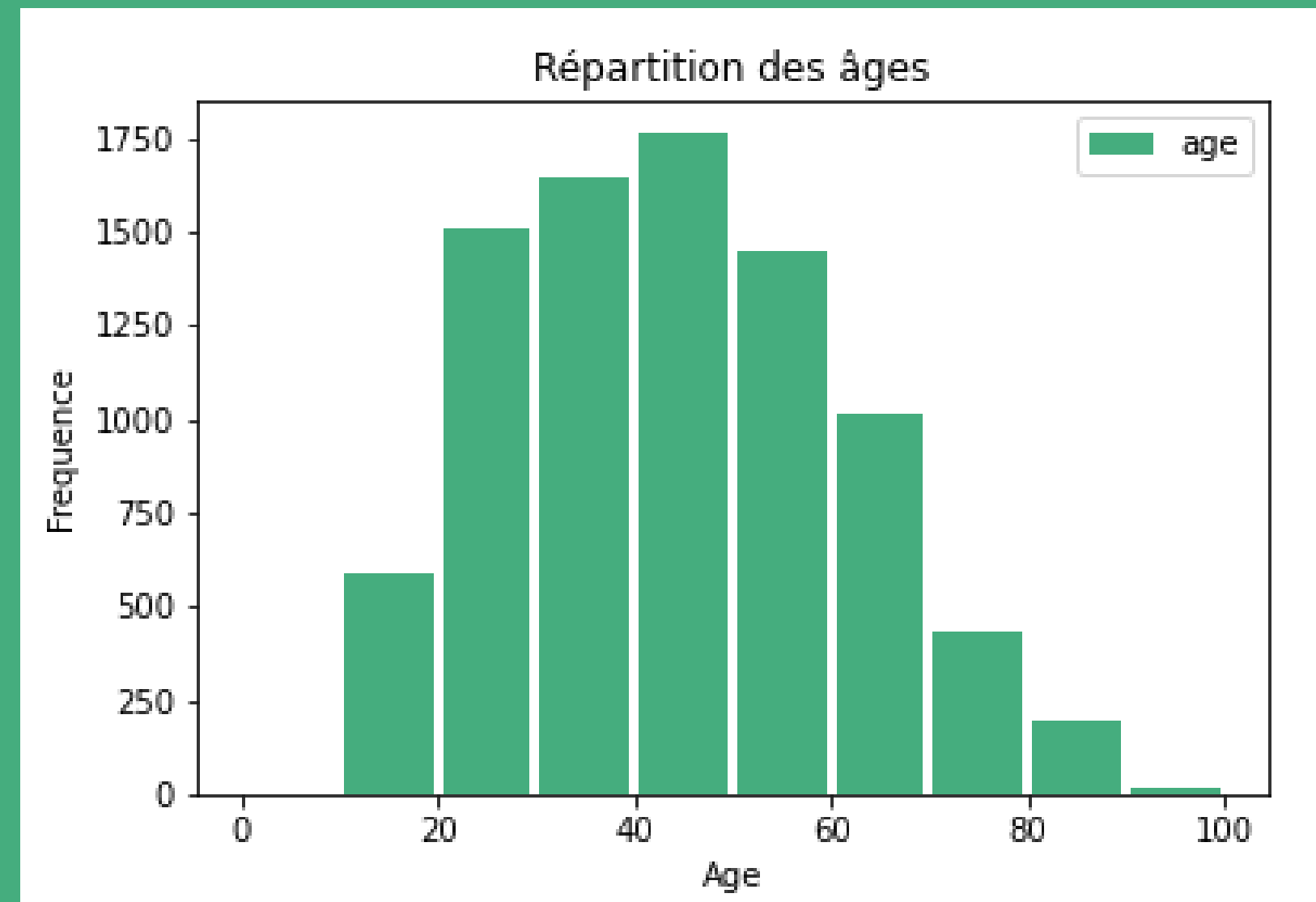
Nombre total de clients : 8 623

Nombre de femmes : 4 491 (soit 52%)

Nombre d'hommes : 4 132 (soit 48%)

La répartition des sexes est plutôt égale mais les femmes sont plus présentes dans la base.

# Répartition des clients par tranche d'âge



Age moyen des femmes : 45 ans

Age moyen des hommes : 44 ans

On remarque que la majorité de nos clients ont entre 30 et 50 ans.

# Panier moyen par tranche d'âge

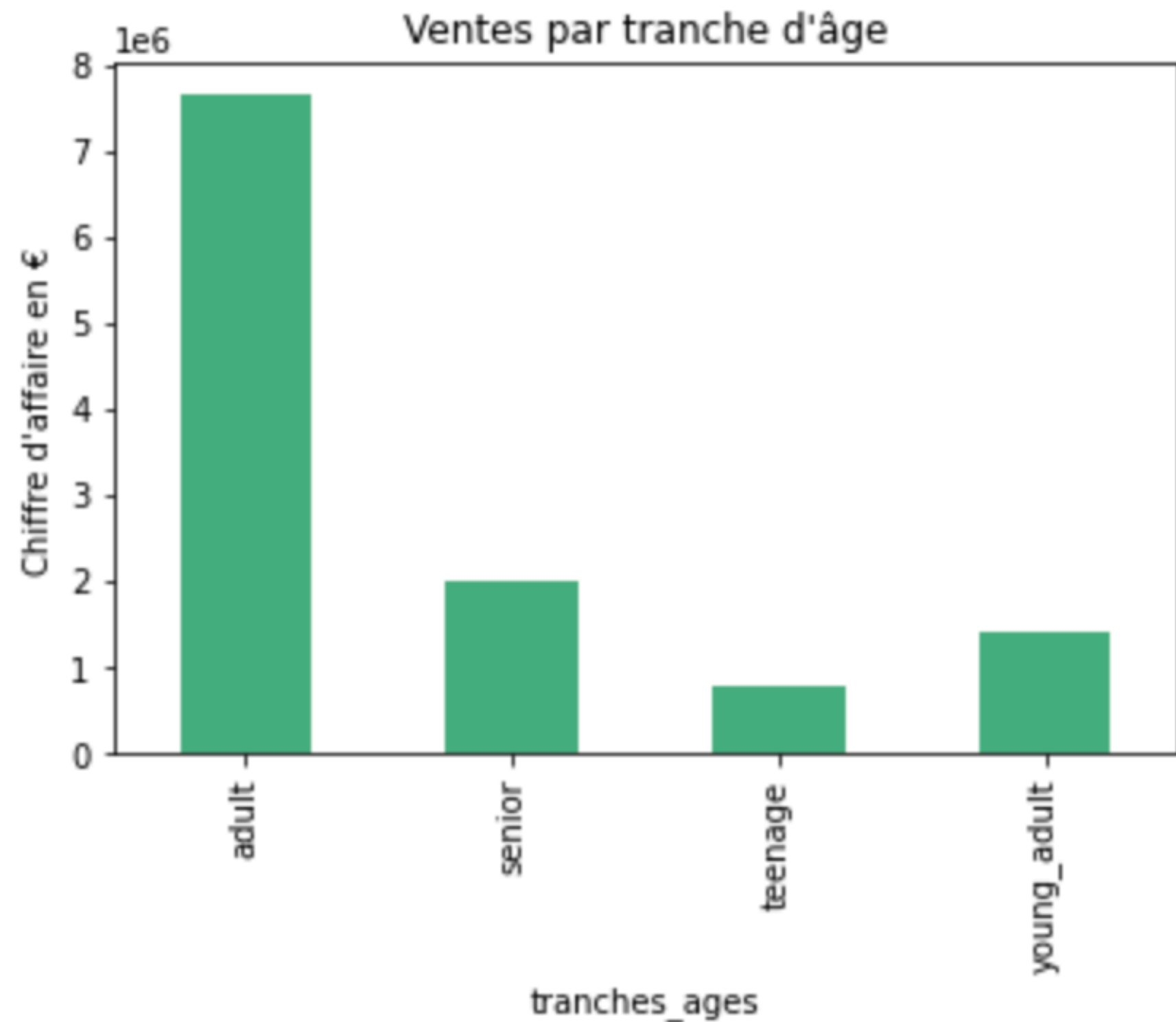


On identifie un top 2 qui se démarque des autres :

- Les **young adult** sont ceux qui dépensent le plus en une seule commande avec un panier moyen à **42,42€**. Leur panier moyen est **3 fois plus élevé que celui des adultes** alors que les adultes représentent 40% du chiffre d'affaires.
- En deuxième position on retrouve les **teenagers** avec un panier moyen de 40.33€.

Les **adultes**, qui sont les premiers clients de la marque, ont le panier moyen le plus bas avec une moyenne de **13.76€**. Ils achètent assez fréquemment dans l'année mais ne dépensent pas assez par commande.

## Répartition du chiffre d'affaires par tranche d'âge

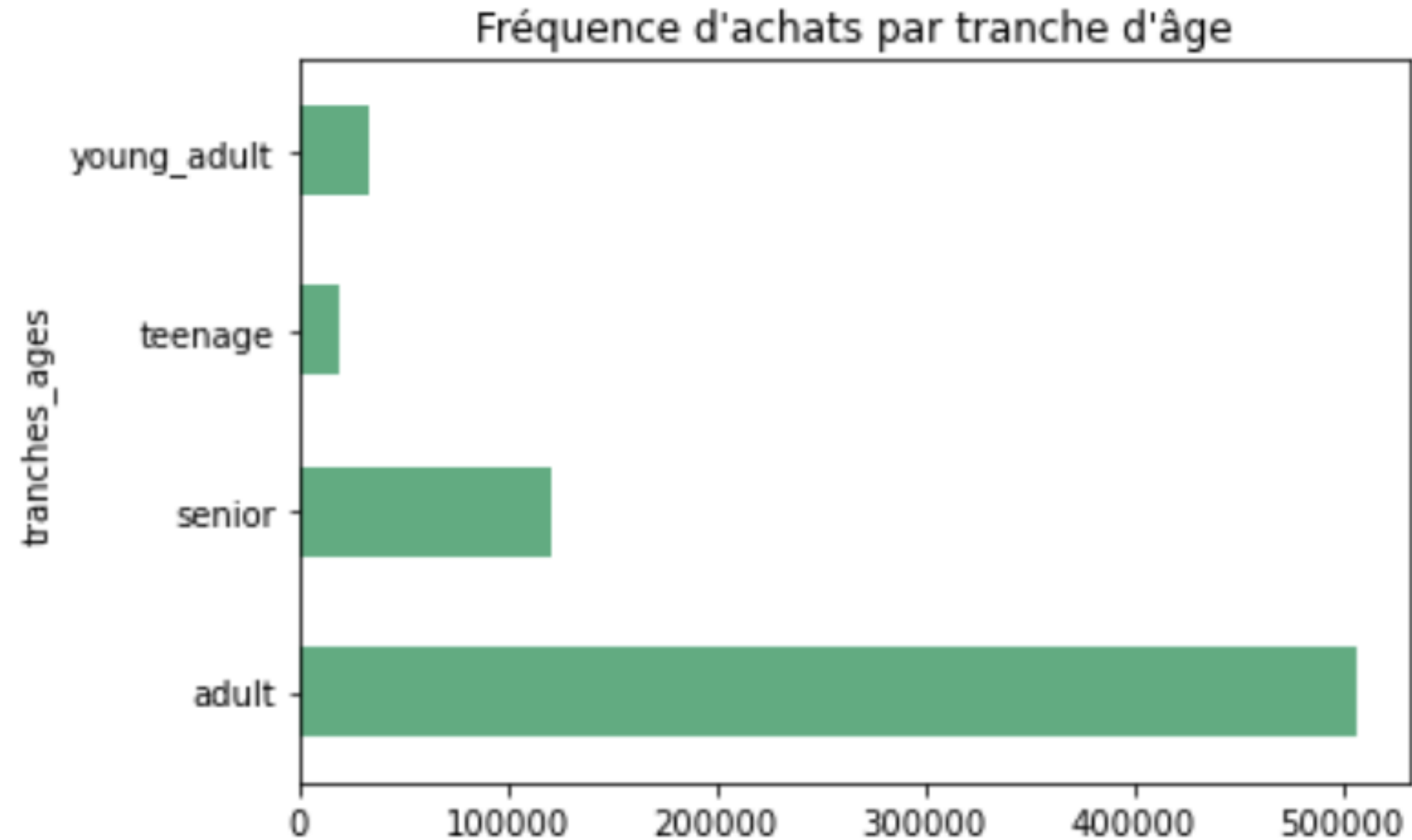


On remarque que les **adultes sont les clients qui génèrent la plus grosse part du chiffre d'affaires** alors qu'ils ont le panier moyen le plus faible à 13,86€. Ils représentent la moitié de notre chiffre d'affaires. C'est notre cible principale.

A l'inverse, les **teenagers représentent le chiffre d'affaires le plus faible** alors qu'ils ont le panier moyen le plus élevé à 40,30€. Donc nous devons adapter nos stratégies de ciblage et de marketing à ce segment de client.



## Fréquence d'achats par tranche d'âge



Ce sont les **adultes** qui achètent le plus souvent, 5 fois plus que les seniors.

Les teenagers achètent le moins souvent mais ont un des plus grands paniers moyens.

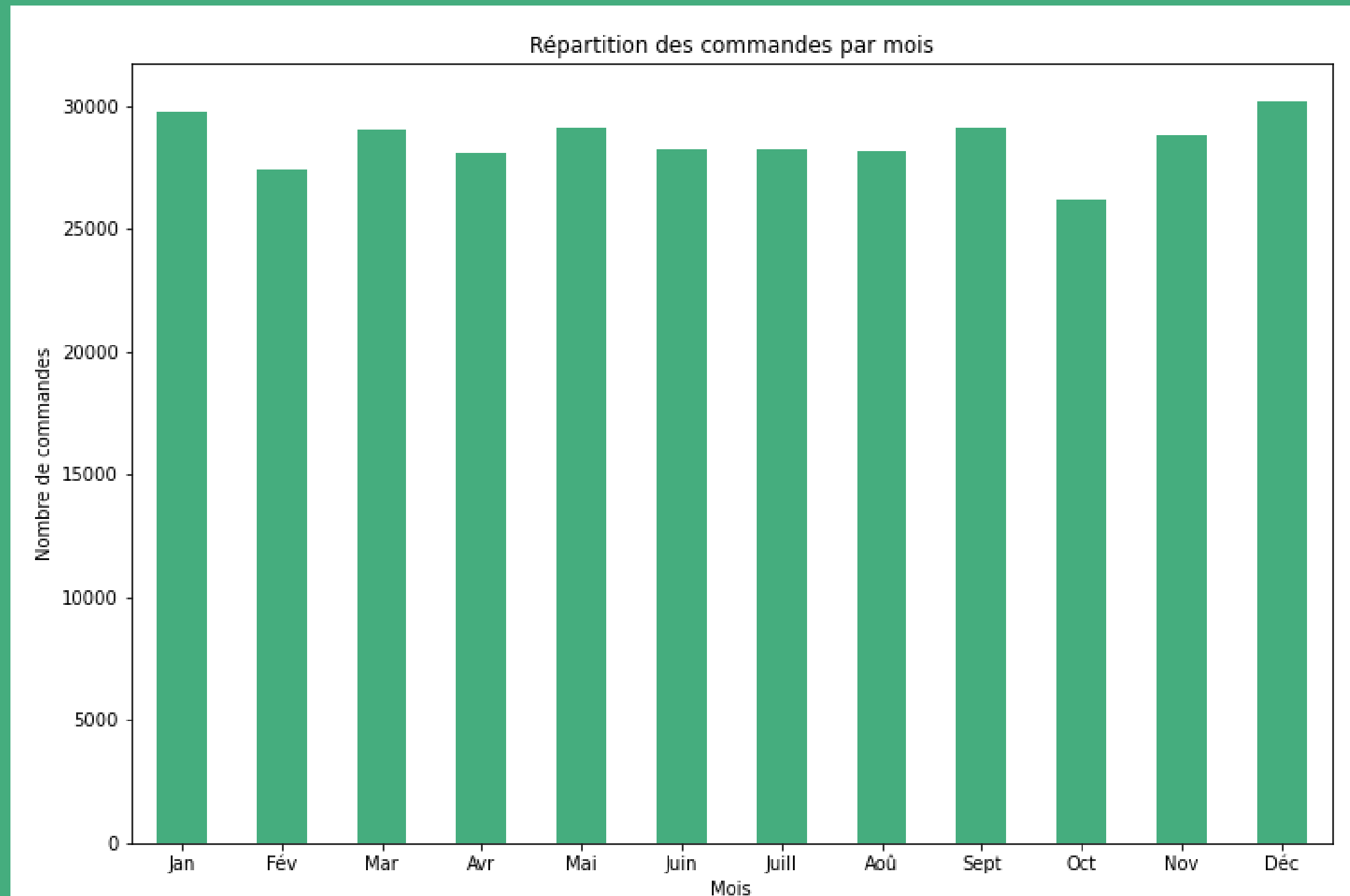
# Evolution du chiffre d'affaires entre 2021 et 2023



Malgré une fin **2021** marquée par un creux durant le mois d'octobre. Le chiffre d'affaires de l'entreprise a pu rebondir et a été le plus haut en **2022** avec un total de plus de **6 M d'euros**. Le début de l'année **2023** est quant à lui très encourageant avec une belle lancée sur les **2 premiers mois** durant lesquels les ventes ont atteint **974 k euros**.

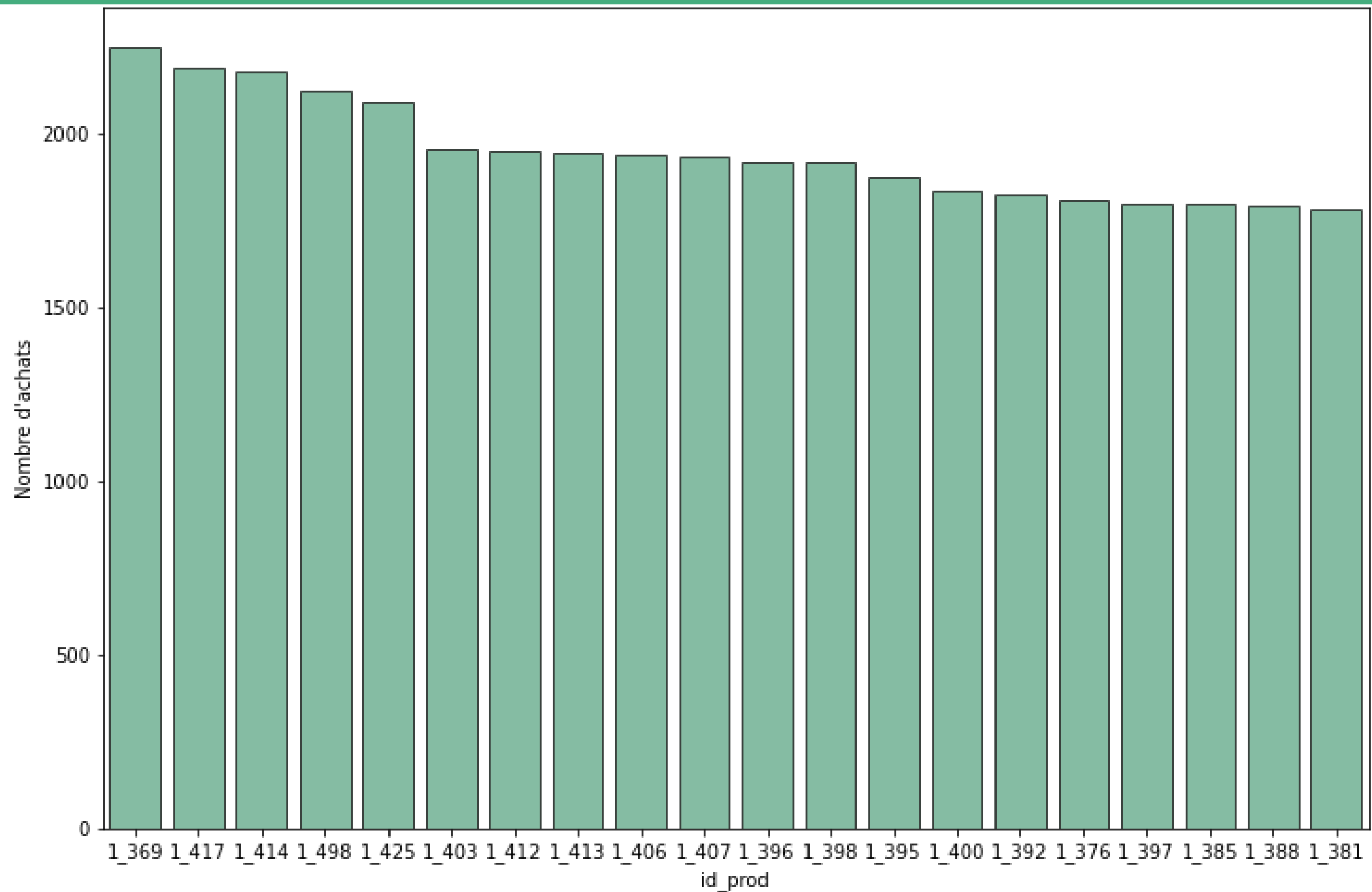


# Evolution du nombre de commandes par mois



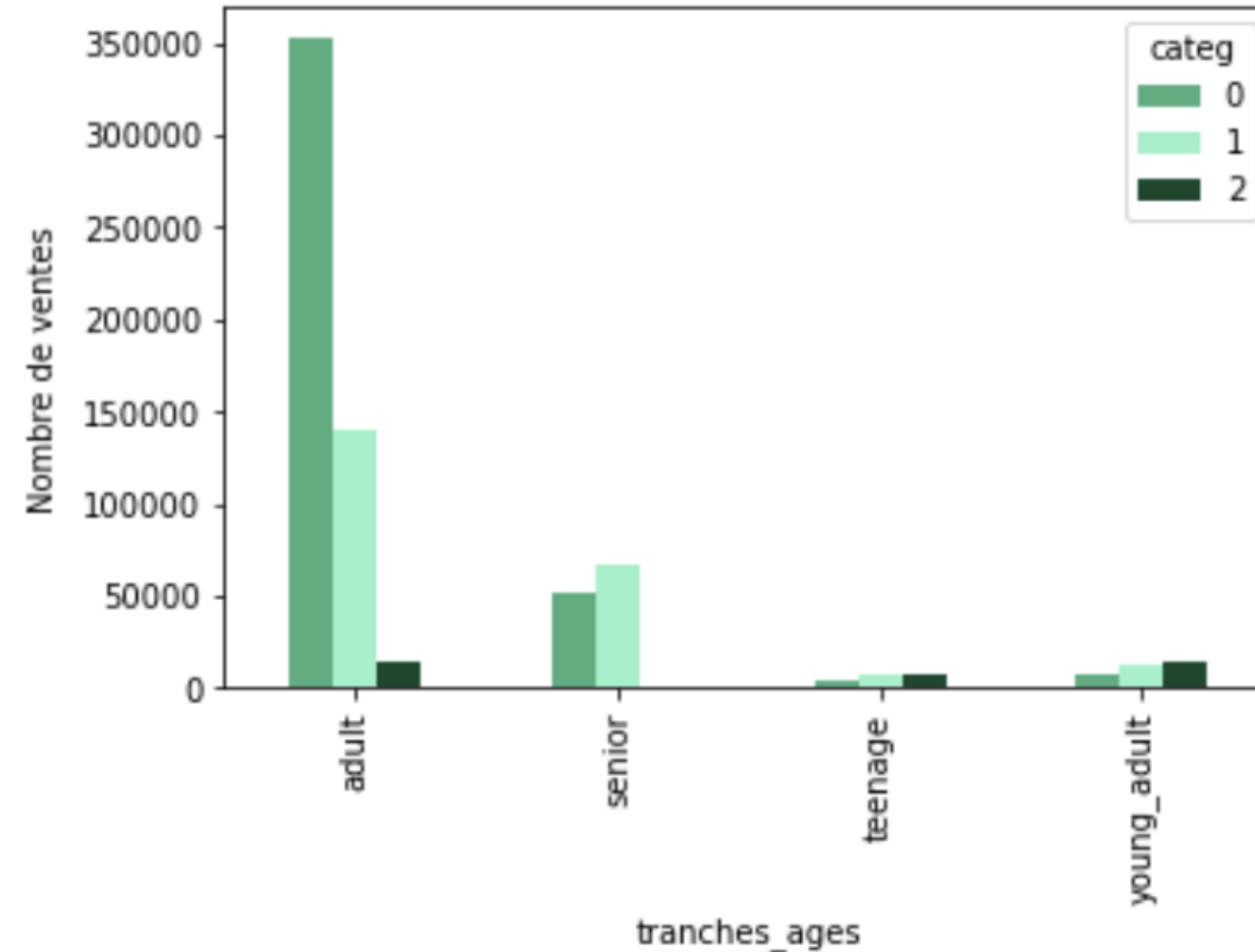
Janvier et Décembre sont donc les mois avec le plus de commandes, ce qui peut-être expliqué par les fêtes de fin d'années et les soldes d'hiver.  
On remarque aussi la saisonnalité des vacances d'été (juin, juillet, août) pour lesquelles on enregistre moins de commandes.

# Top 20 des produits en fonction du chiffre d'affaires



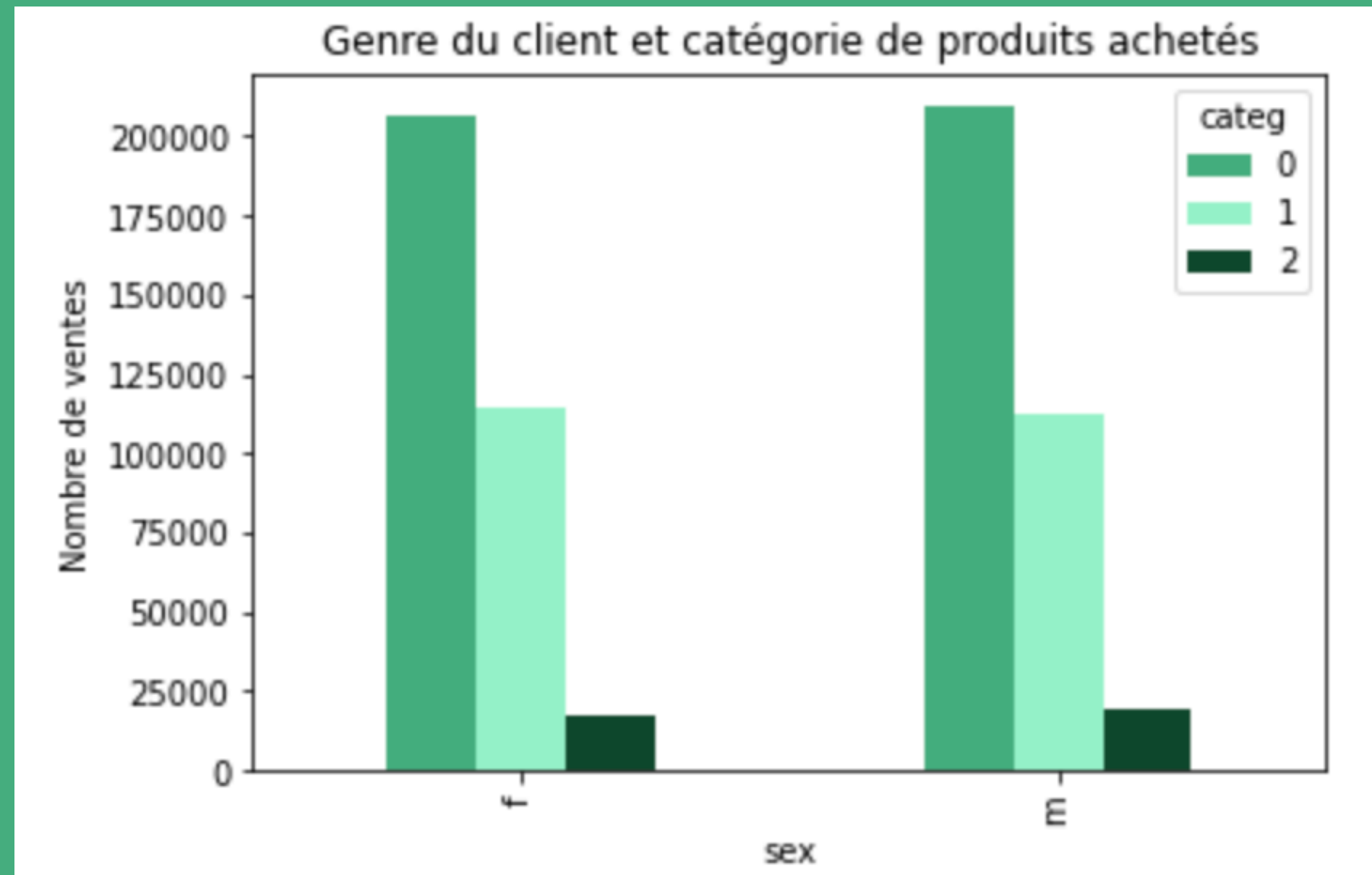
Nous constatons que les 20 produits les plus vendus sont **tous** de la catégorie 1.

# Répartition du nombre de ventes par tranche d'âge et catégories



Nous observons ici un phénomène assez intéressant. La fréquence des achats par catégorie de produits **fluctue selon la tranche d'âge**. Il pourrait donc y avoir donc **une corrélation** entre ces deux paramètres. La catégorie 0 est la catégorie la plus achetée par les adultes mais pas chez les seniors et les jeunes chez qui elle passe 2ème.

# Répartition du nombre de ventes par sexe et catégories



On remarque que les produits ont la même popularité que ce soit chez les hommes ou chez les femmes. **La catégorie 0 rapporte presque 2 fois plus que la catégorie 1.**

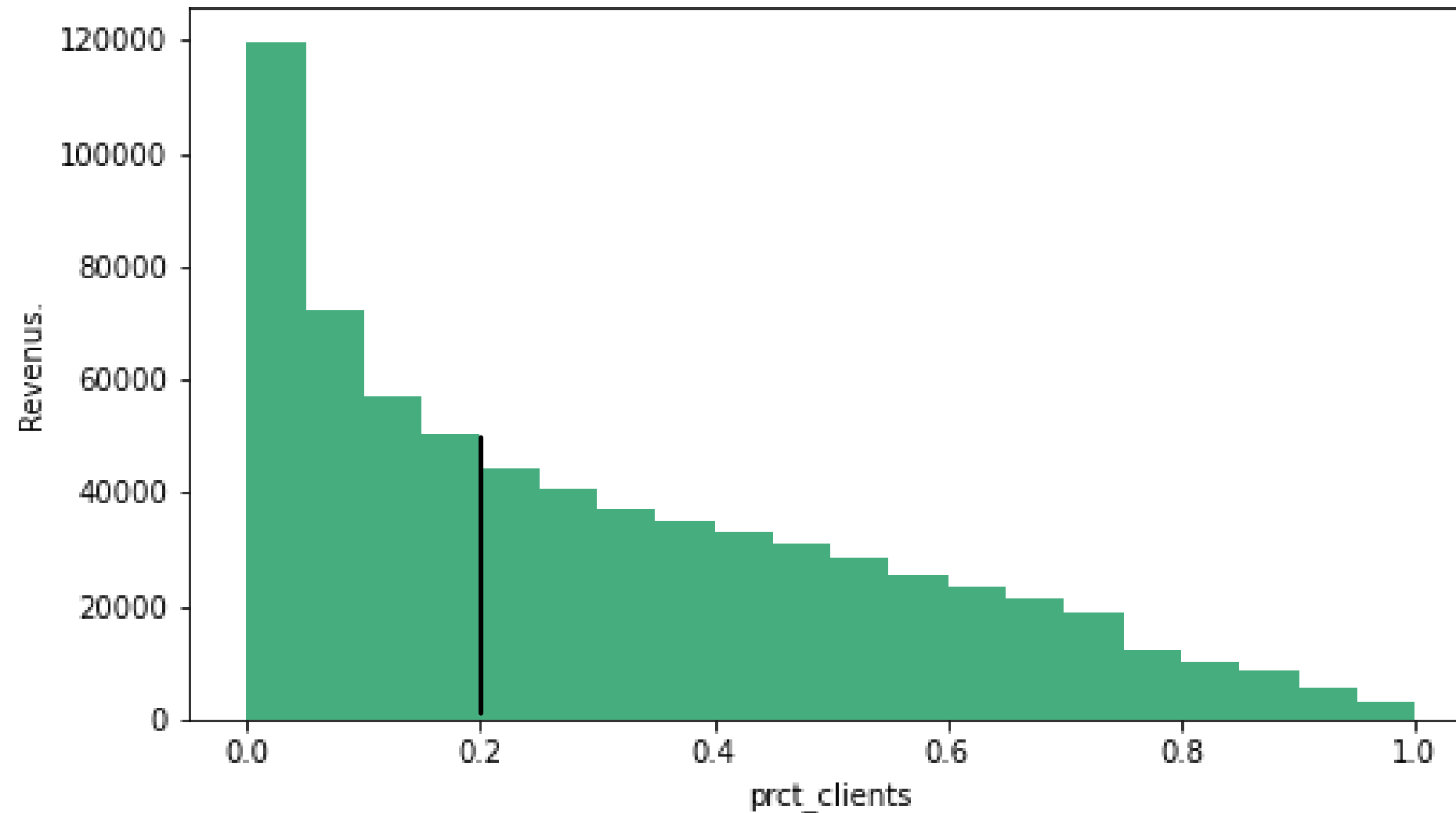
Les prix moyens ne reflètent pas le chiffre d'affaires. En effet, même si les produits de catégorie 2 sont en moyenne plus chers, ce sont les catégories 0 et 1 qui rapportent le plus à l'entreprise.

**Prix moyen d'un produit :**

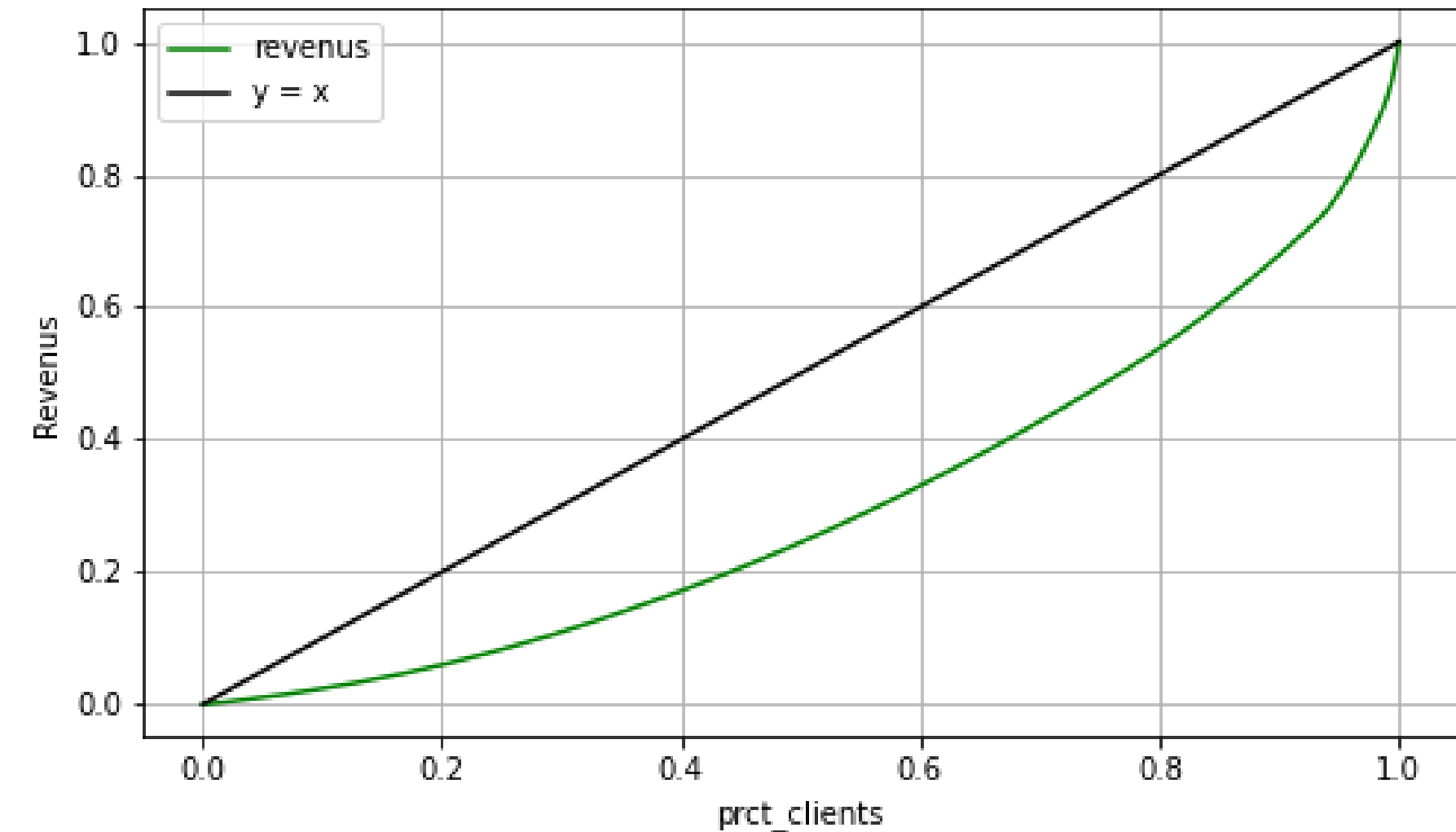
- Catégorie 0 : 10,64€
- Catégorie 1 : 20,48€
- Catégorie 2 : 76,21€

# Effet Pareto

Diagramme de la loi de Pareto



Courbe de Lorenz



**La courbe de Lorenz nous montre que 55% des dépenses sont réalisées par seulement 20% des clients de la base.**

# Scoring et Clustering





# Scoring RFM avec la méthode des quantiles

Grâce au scoring RFM (récence, fréquence, montant) on peut identifier les catégories de clients suivantes :

- Gros dépensiers : 2 150
- Loyaux : 2 134
- Champions : 860
- Presque perdus : 87
- Petits dépensiers perdus : 958
- Perdus : 323

## Points positifs

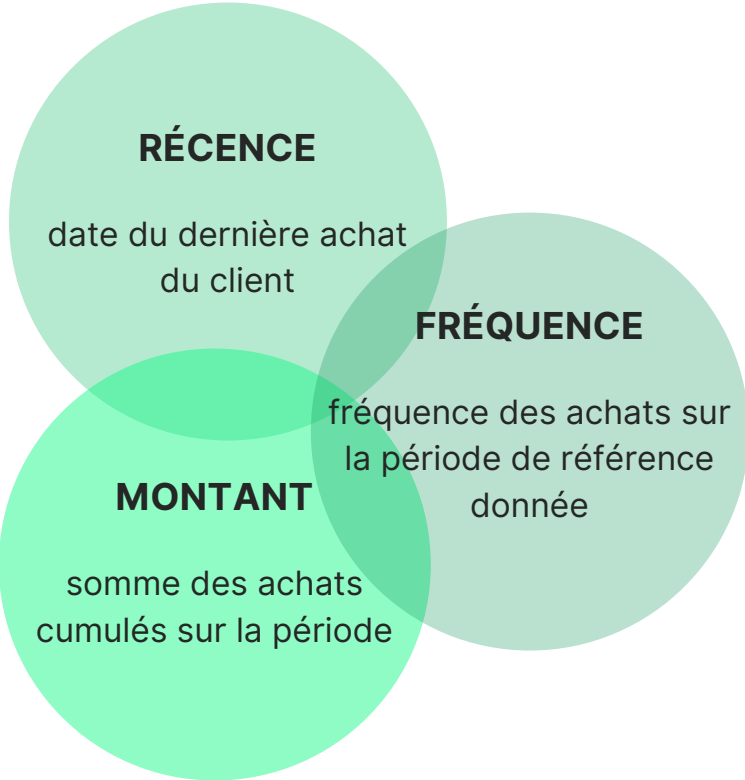


- Majorité de loyaux & gros dépensiers
- Peu de presque perdus

## Points négatifs



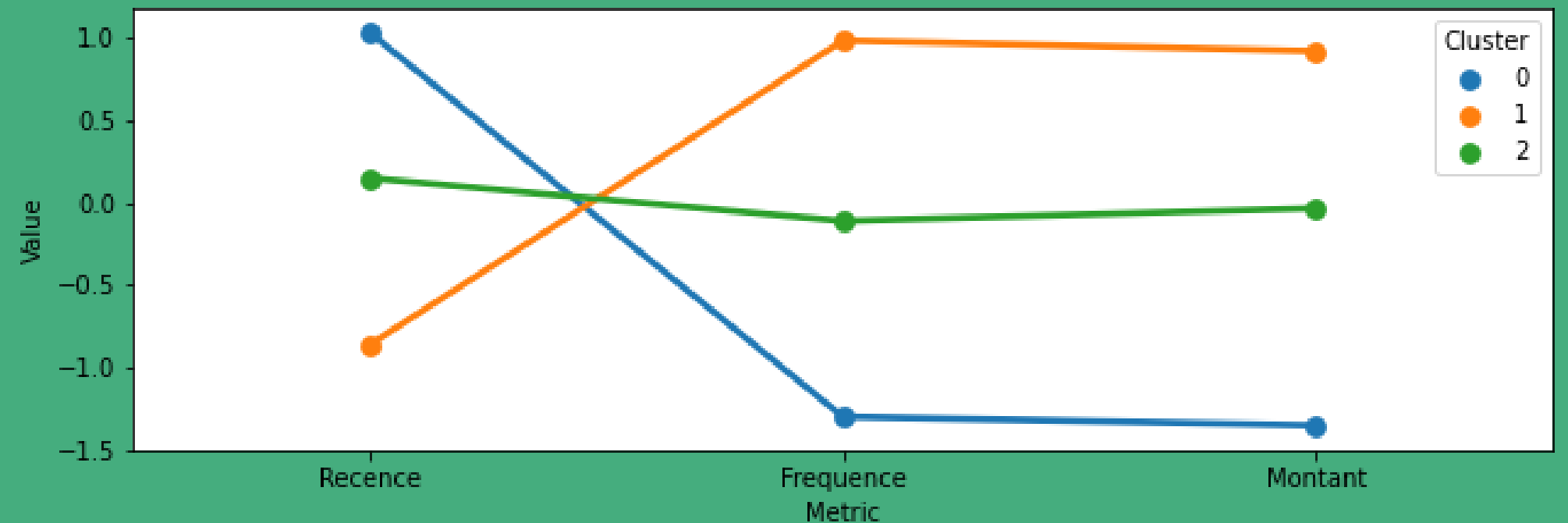
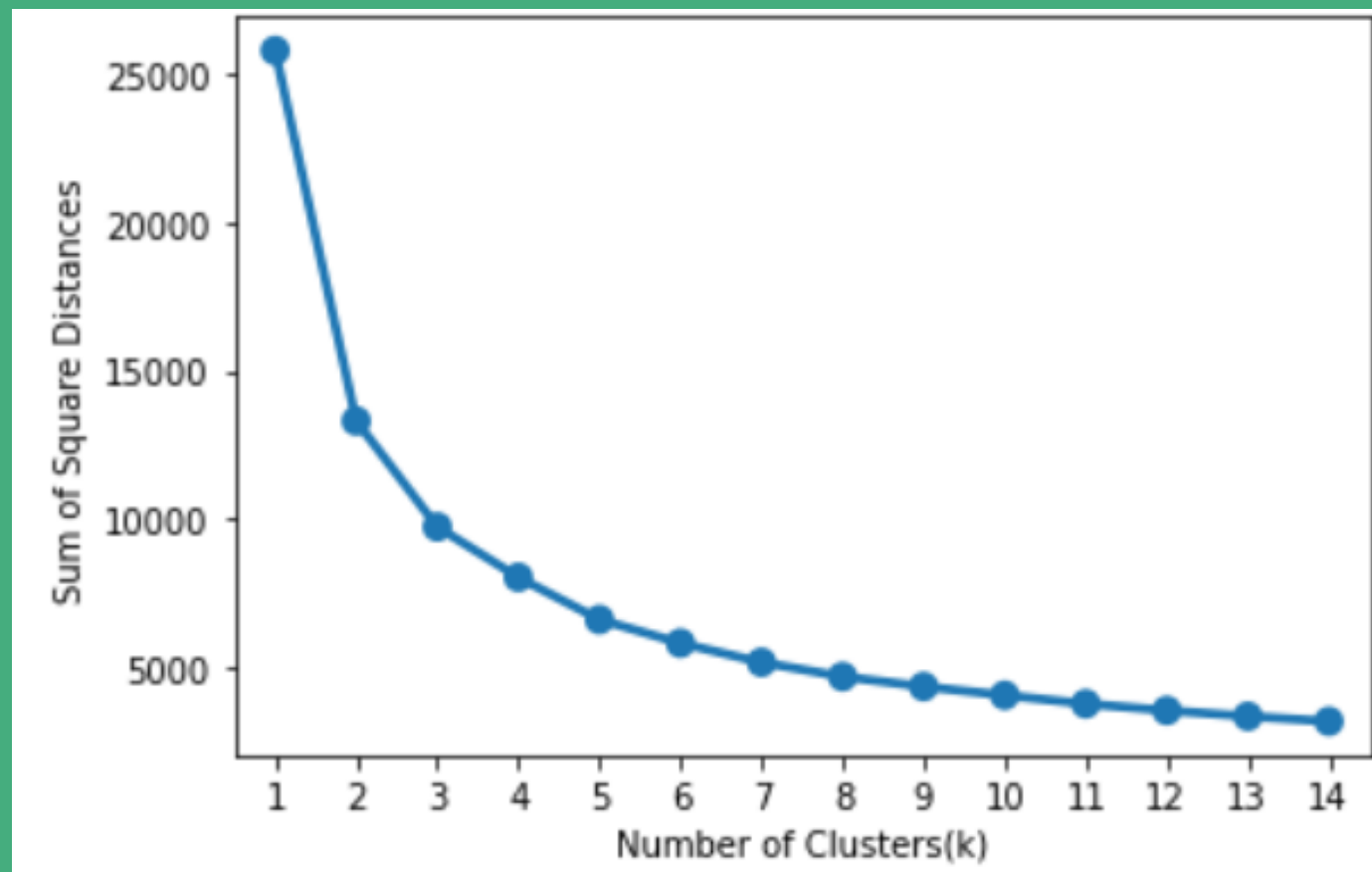
- Beaucoup de perdus & petits dépensiers perdus



	BEST			WORST		
R	1	2	3	4		
F	4	3	2	1		
M	4	3	2	1		

# K-means clustering

Après avoir scorer nos clients nous avons appliqué la méthode d'apprentissage automatique K-means afin de créer des Clusters.



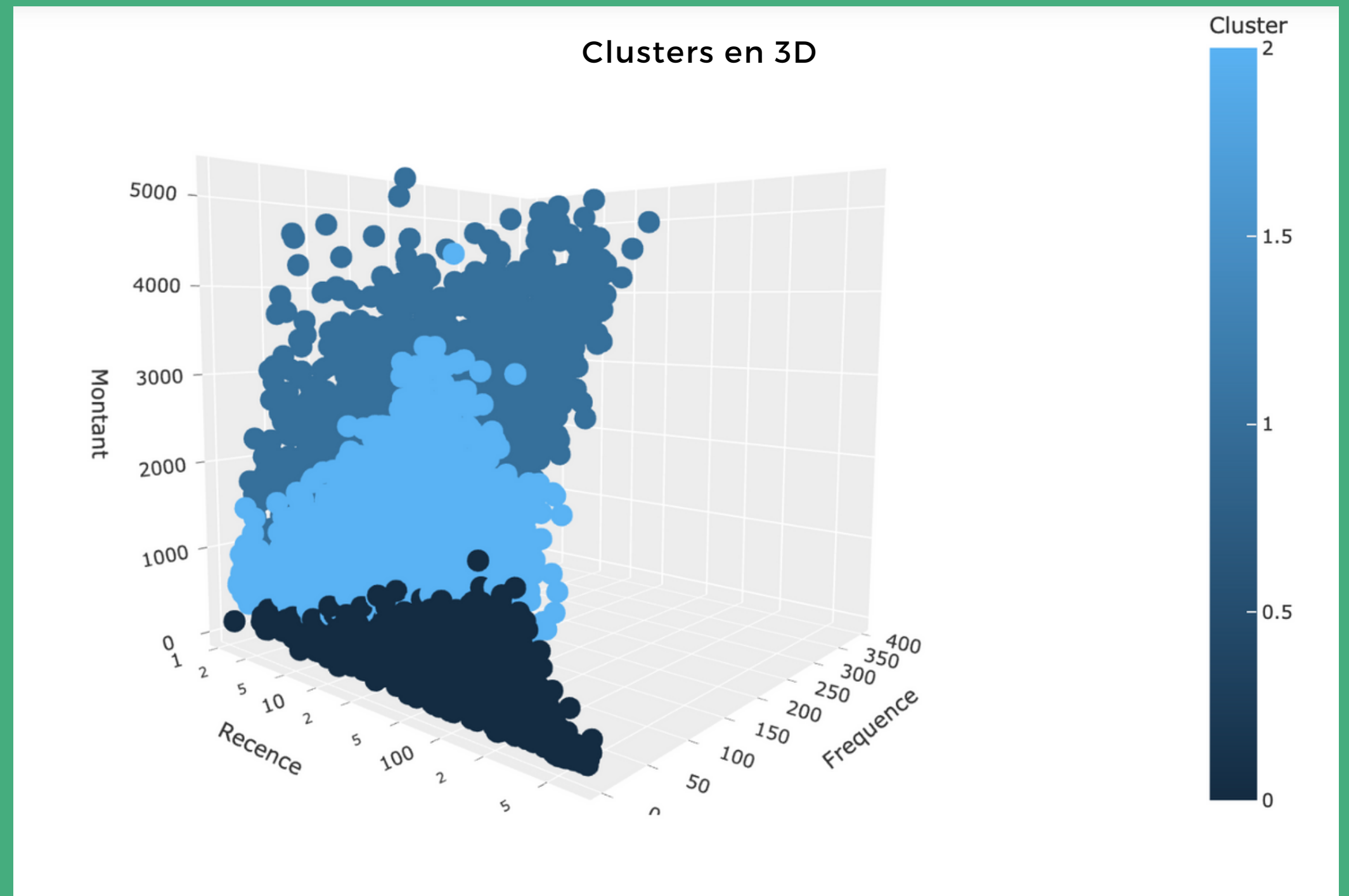
La méthode de distorsion nous permet donc de faire ressortir Clusters suivants :

- Cluster 0 : clients avec le moins de valeur
- Cluster 1 : clients avec le plus de valeur
- Cluster 2 : clients moyens (entre le cluster 0 et 1)

# K-means clustering

Ces éléments nous permettent ainsi de diviser notre bases en **3 clusters** :

- Les clients **Gold** (cluster 1) : **2,9 K**
- Les clients **Silver** (cluster 2) : **3,7 K**
- Les clients **Bronze** (cluster 0) : **1,9 k**



# CONCLUSION

L'entreprise LA PAGE est en croissance. De 2021 à 2023 on enregistre un bon nombre de commande avec une clientèle large qui rajeunit.

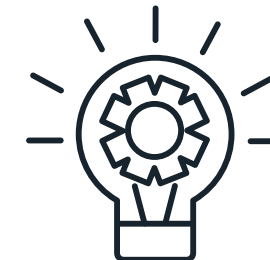
Cette analyse va nous permettre d'adapter les actions marketing et de retargeter nos clients.



**Catégorie de produit  
1 la plus vendue**



**Chiffre d'affaires 2022  
supérieure à 2021**



**Croissance prévue pour  
2023**



**Clients 30-50 ans  
Dépensent le plus**

**Merci !**