



Lapage

Analyse des ventes en ligne

03/2021 – 02/2023

Introduction

Contexte : Une librairie physique a décidé depuis 2 ans d'ouvrir un site de vente en ligne. C'est dans ce cadre, que l'entreprise souhaite faire le point après deux ans d'exercice, afin d'analyser les points forts des ventes en ligne, les points faibles et les comportements des clients.

1. Nettoyage des données
 1. Individus « tests »
 2. Jointure et analyse du jeu de données obtenu
 3. Identification et traitement des valeurs nulles
2. Analyses
 1. Chiffres d'affaires
 2. Références
 3. Clients

Nettoyage des données

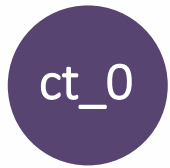
1. Individus « tests »



Session test



Produit test



Clients test



➤ **Suppression des individus « tests »**

2. Jointure et analyse du jeu de données obtenu

Il y a des valeurs manquantes dans la variable price et dans la variable catégorie.

3. Identification et traitement des valeurs nulles

La catégorie et le prix du produit « 0_2245 » sont les valeurs manquantes.

• Traitement de la catégorie manquante

Le premier caractère de l'id_web correspond à la catégorie à laquelle appartient le produit.

➤ **La catégorie manquante de l'id_prod 0_2245 est remplacée par 0.**

• Traitement du prix manquant



Prix minimum



Prix moyen

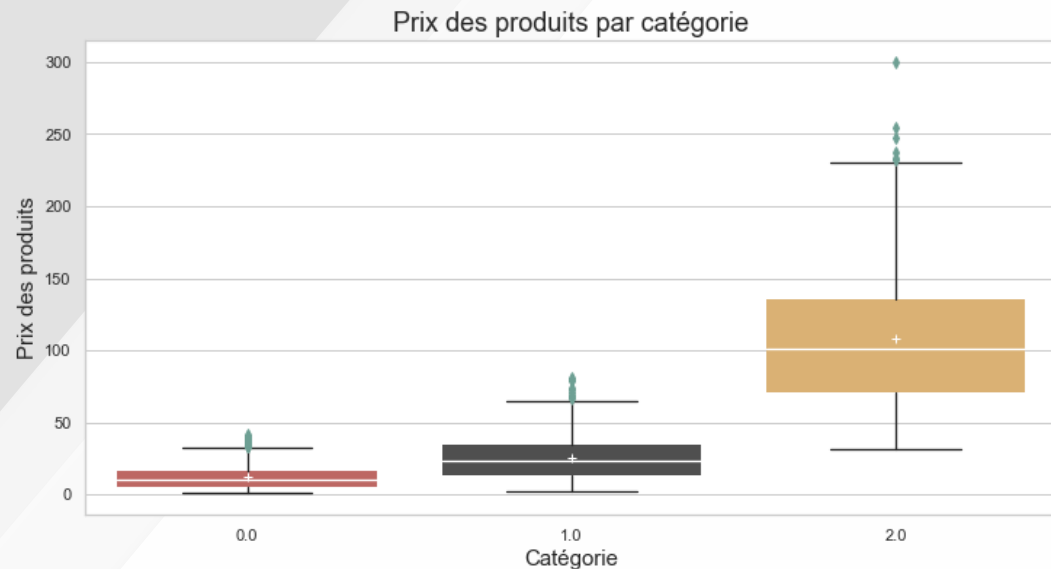


Prix maximum

Il serait intéressant de voir s'il y a une fourchette spécifique de prix pour chaque catégorie

Étude catégorie et prix

- Boîtes à moustaches



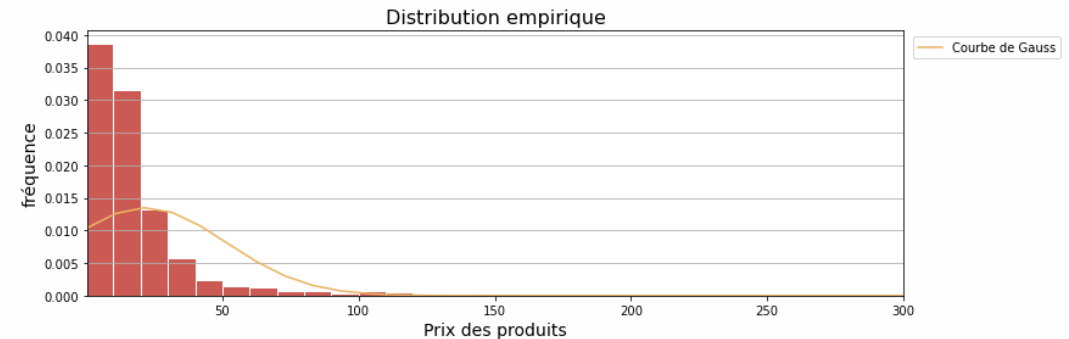
Les catégories semblent avoir des prix différents

- Vérification de cette observation

Mesures de forme

- **Skewness** (Mesure d'asymétrie) : 3.81
- **Kurtosis** (Mesure d'aplatissement) : 18.12
- La variable ne suit pas la loi normale

Représentation graphique



Test statistique

Test de Kruskal Wallis (test non-paramétrique)

p-value < 0.05

➤ Il y a des différences significatives entre les catégories

Moyennes des prix par catégorie

11,71€

Catégorie 0

25,50€

Catégorie 1

107,64€

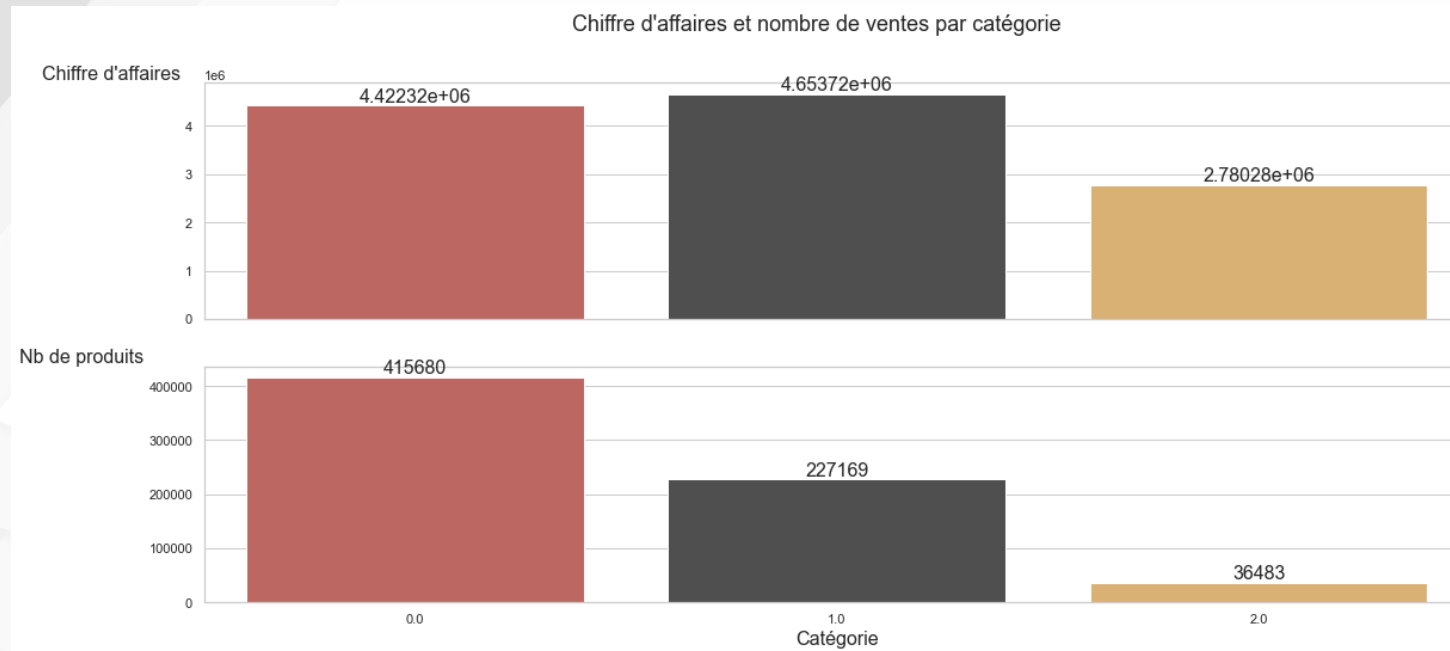
Catégorie 2

➤ **Remplacement des valeurs manquantes prix par la moyenne des prix de la catégorie 0.**

Chiffre d'affaires

Chiffre d'affaires total : **11 856 317€**

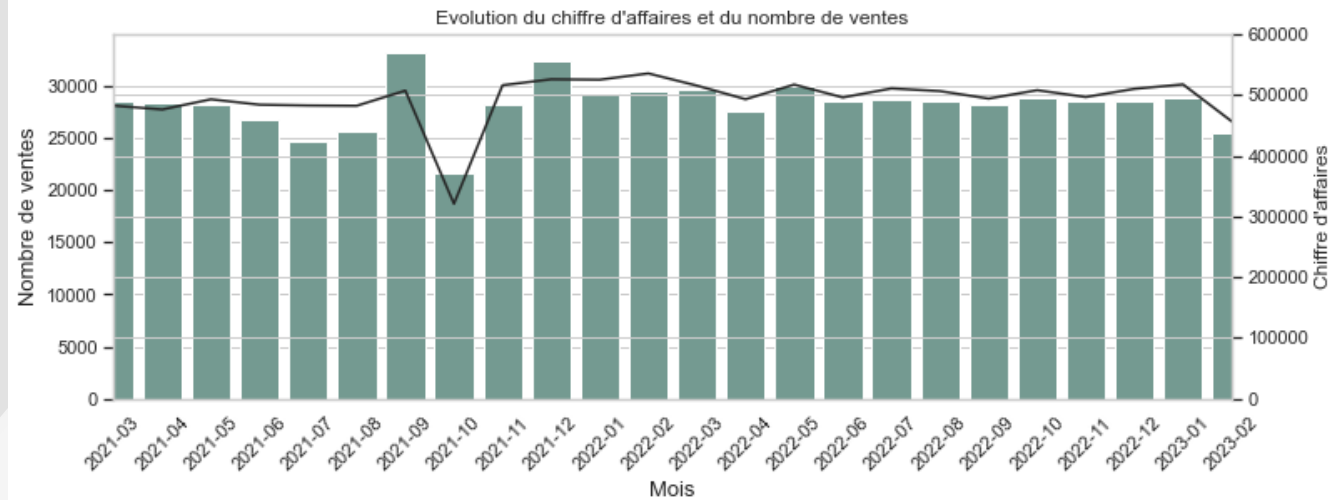
- Chiffre d'affaires et nombre de ventes par catégorie



Catégorie 1 ➤ *Génère le plus de chiffre d'affaires*

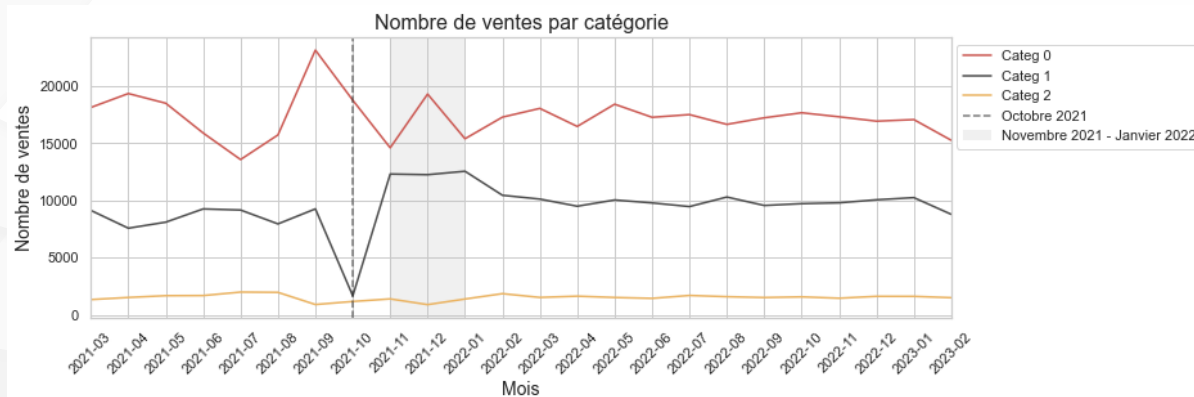
Catégorie 0 ➤ *Catégorie la plus vendue.*

Évolution du chiffre d'affaires



➤ Baisse au mois d'octobre, du nombre de ventes et du chiffre d'affaires

• Nombre de ventes mensuels par catégorie



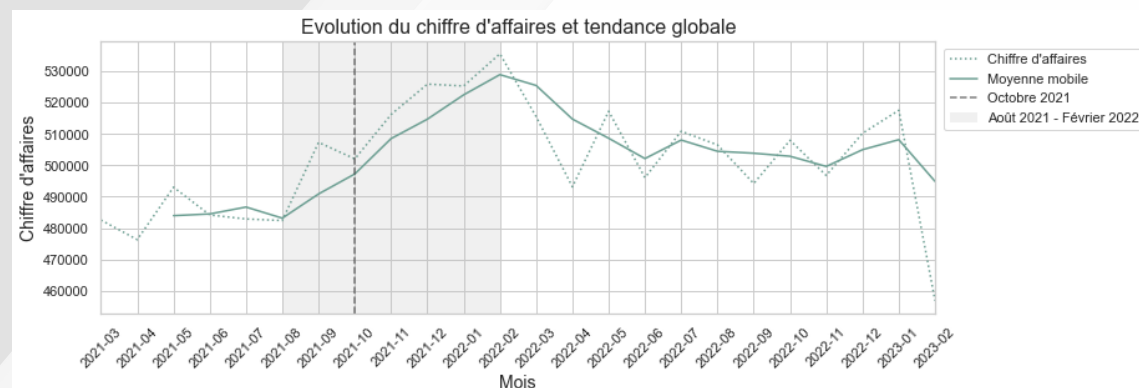
Ventes catégorie 1 au mois d'octobre :

Jour	Chiffre d'affaires	Nombre de ventes
01/10/2021	7003.79	344
28/10/2021	6317.99	316
29/10/2021	6425.18	326
30/10/2021	6753.69	338
31/10/2021	7261.67	342

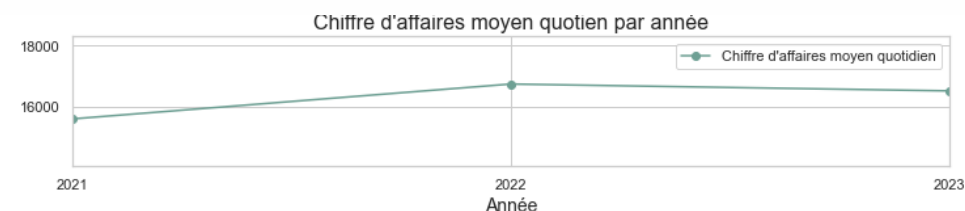
➤ Remplacement des données manquantes en octobre par la moyenne d'août à décembre en tenant compte du quart mois.

Évolution du chiffre d'affaires et moyenne mobile après traitement

- Tendance du chiffre d'affaires



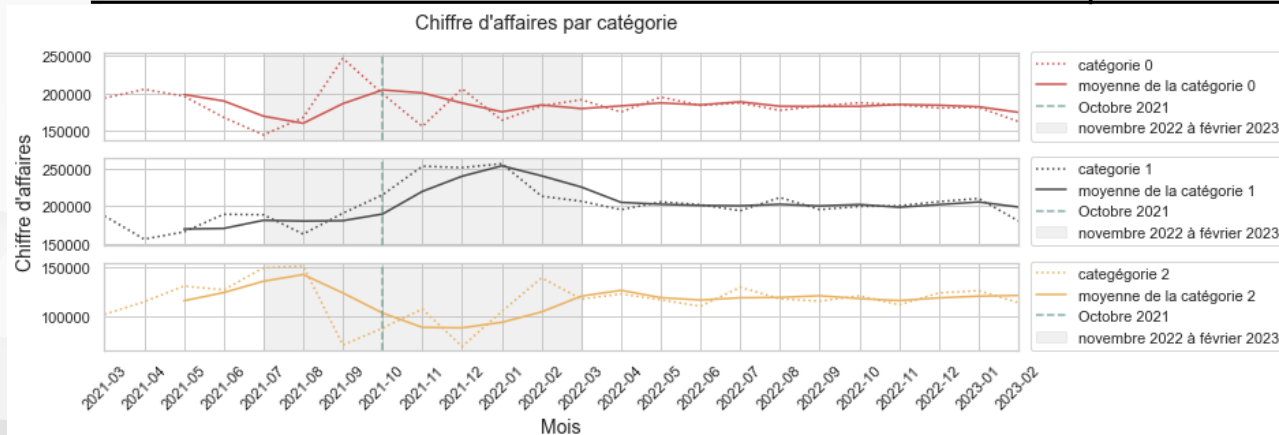
Chiffre d'affaires quotidien moyen



➤ Taux d'augmentation du CA quotidien moyen : 5.9%.

➤ Hausse globale depuis 2 ans

- Evolution mensuel du chiffre d'affaires par catégorie



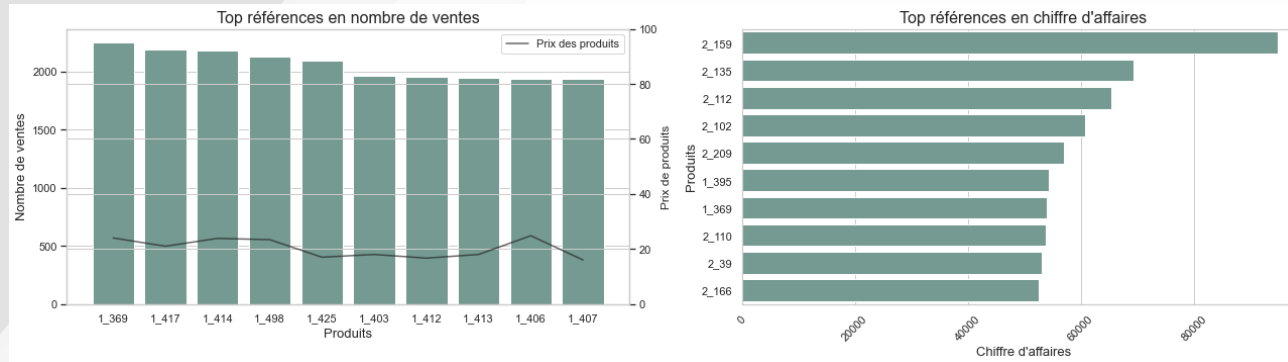
➤ Légère baisse de la **catégorie 0**

➤ Hausse de la **catégorie 1**

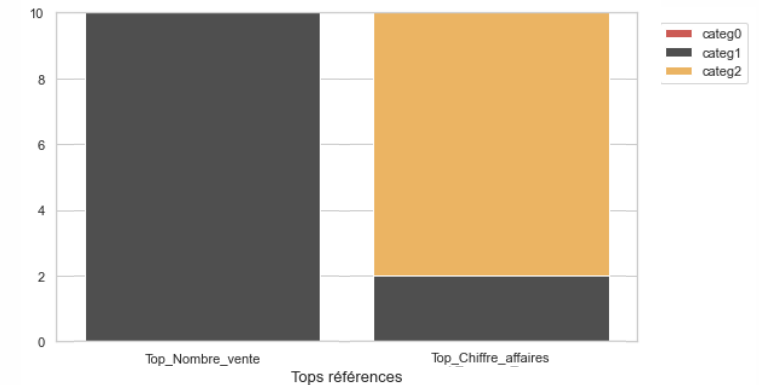
➤ Stagnation de la **catégorie 2**

Références

• Top références

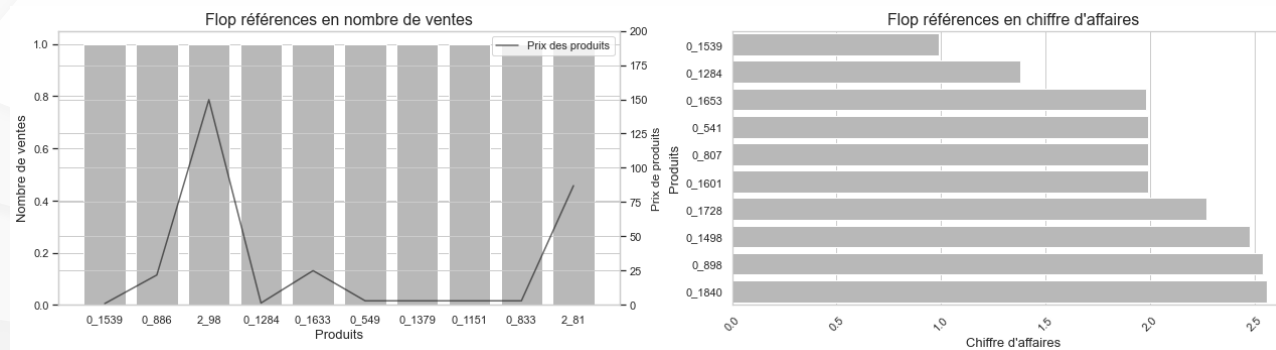


Répartition par catégorie

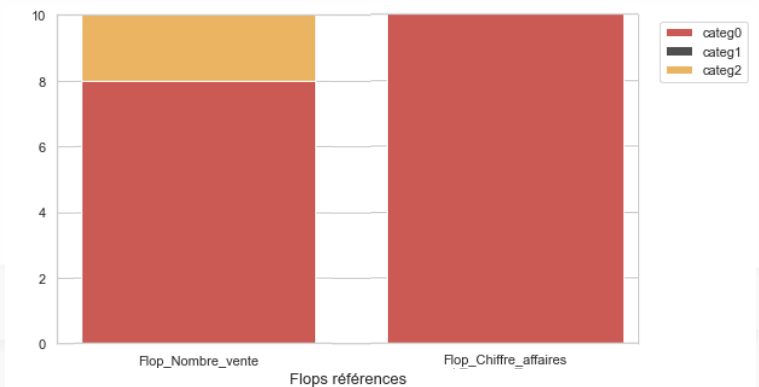


➤ Les produits du top de ventes sont de la catégorie 1 et ceux du chiffres d'affaires sont en majorité de catégorie 2

• Flop références



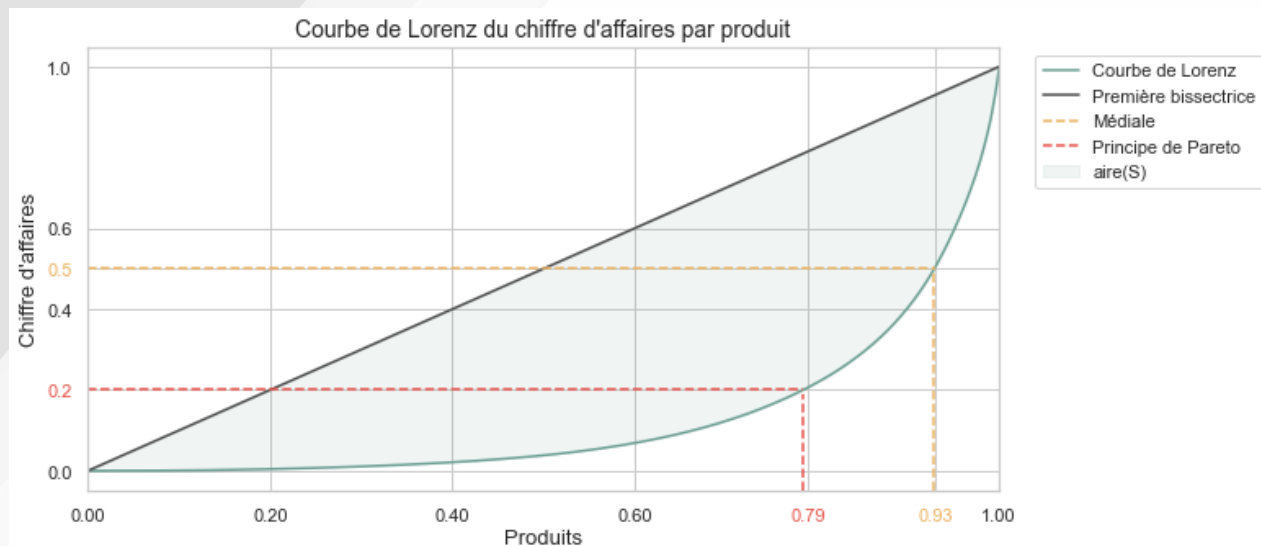
Répartition par catégorie



➤ Les produits du flop ventes sont de la catégorie 0 et 2 et ceux du chiffres d'affaires sont de catégorie 0

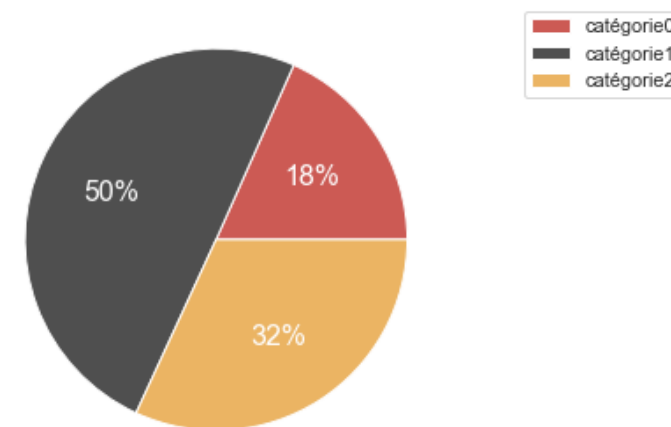
Répartition du chiffre d'affaires par référence

- Courbe de Lorenez



- **Médiale du chiffre d'affaires** : 50% du chiffre d'affaires généré par la vente de 7% des produits
- **Principe de Pareto** : 80% du chiffre d'affaires est généré par 20% des produits vendus
- **Coefficient de Gini** : 0.74
- La répartition du chiffre d'affaires par produit est inégale

Répartition par catégorie des 7% de produits générant le plus de chiffre d'affaires



La moitié des produits sont de la catégorie 1.

Clients

- Profil

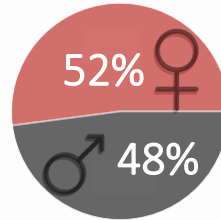
8 600
clients

Age
moyen
45 ans

19 ans – 93 ans

Age
+fréquent
19 ans

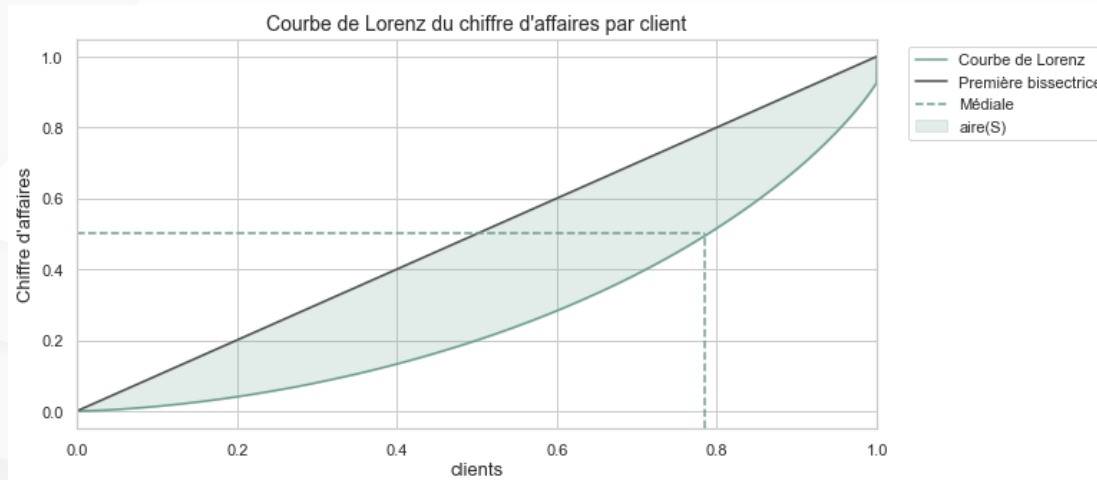
Parité de la clientèle



4
clients
pros

Les achats des 4 clients
"professionnels"
représentent 7.43% du
chiffres d'affaires total

- Répartition du chiffre d'affaires par clients



➤ **Coefficient de Gini : 0.45**

➤ *La répartition du chiffre d'affaires par client reste assez homogène*

➤ Mis en évidence des 4 clients « pros » sur la courbe de Lorenz, que l'on peut identifier par la verticale à la fin de la courbe.

Genre des clients et catégories achetées

- Tableau de contingence

	Catégorie 0	Catégorie 1	Catégorie 2	Total
Femme	4 120	4 392	1 840	10 352
Homme	3 776	4 061	1 757	9 594
Total	7 896	8 453	3 597	19 946

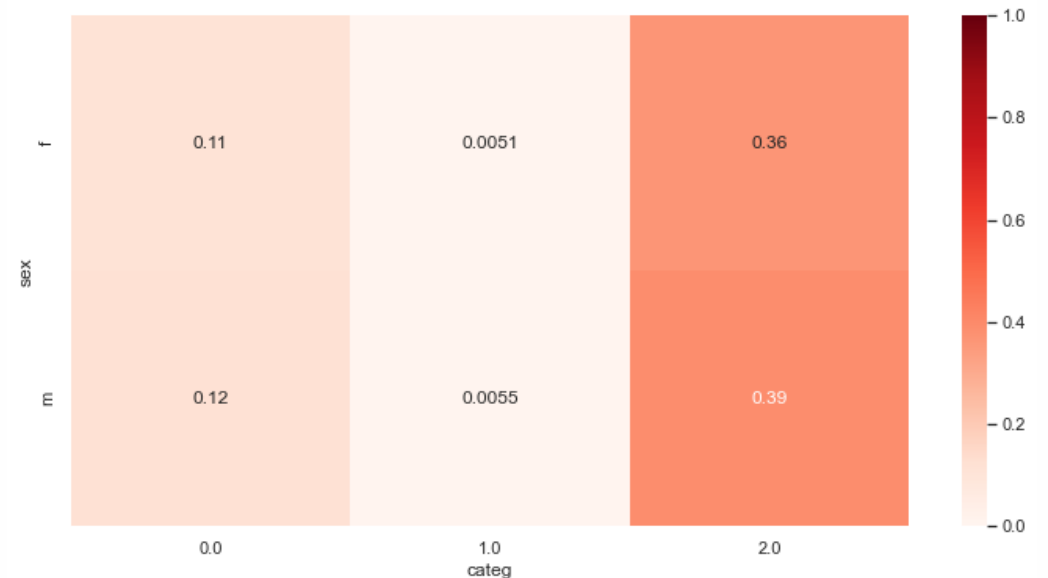
- Test d'indépendance du χ^2

Le χ^2 est de 1.06

Le p-value est de 0.589

Le degré de liberté est de 2

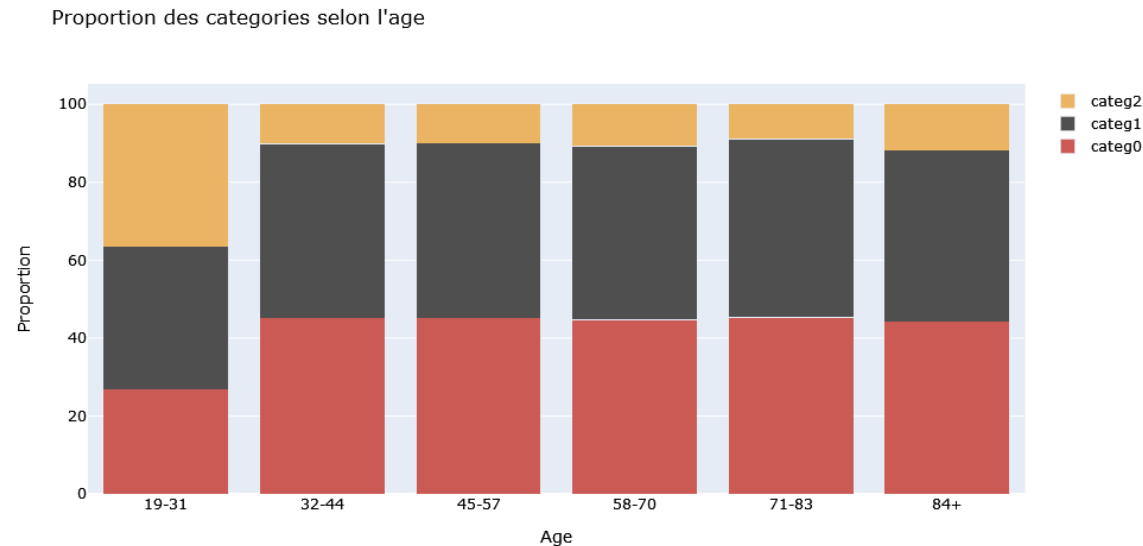
- Heatmap



- *La catégorie du livre acheté et le genre du client sont indépendants*
- *Visualisation de l'indépendance des 2 variables via le heatmap*

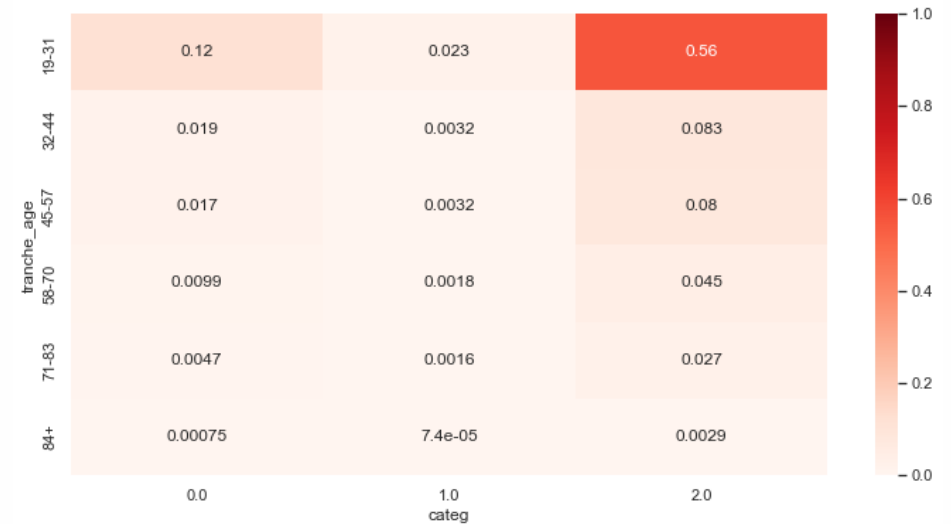
Âge des clients et catégories

- Proportion des catégories par tranche d'âge



- La proportion de produits de catégorie 2 est plus importante chez les clients de moins de 31 ans avec 36%, que dans les autres tranches d'âge, où elle est à environ 10%.

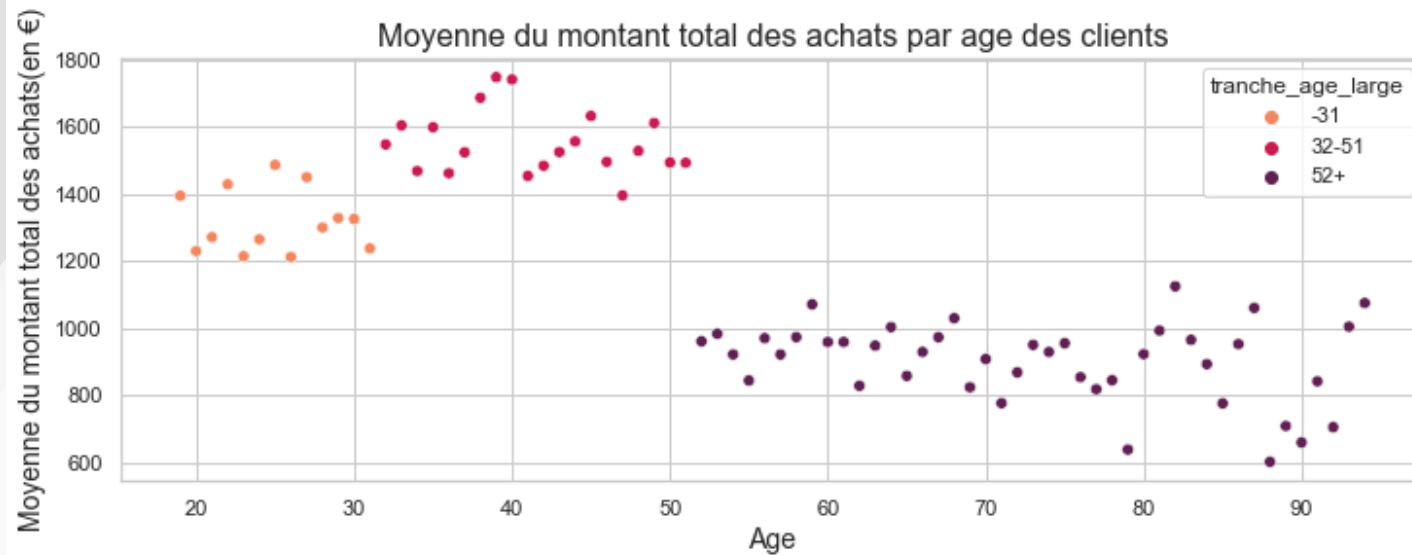
- Heatmap du χ^2



- Les catégories 0 et 1 sont indépendantes des tranches d'âge des clients
- Il y a une non-indépendance de 0,56 entre la catégorie 2 et les clients âgés de 19 à 31 ans

Montant des achats selon l'âge des clients (ss pros)

- Moyenne du montant total par âge



Test de la normalité (Kolmogorov-Smirnov)

- **Pour les 2 variables :**

p-value < 0.05

- Les 2 variables ne suivent pas une loi normale

Test de corrélation (Coefficient de Spearman)

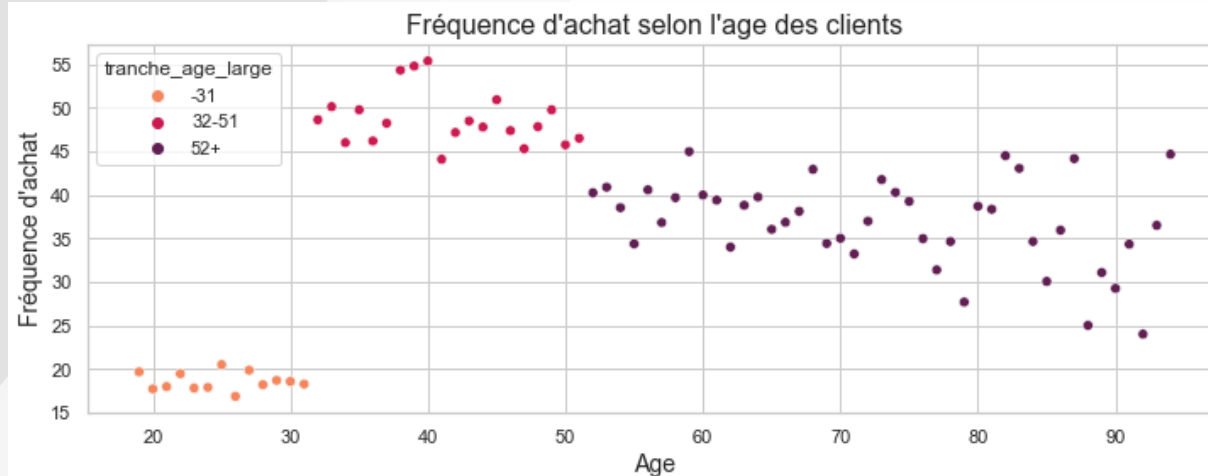
coefficient de Spearman : -0.72 / p-value < 0.05

- Corrélation monotone décroissante

- *Les plus de 52 ans ont une moyenne du montant total des achats moins élevée par rapport aux clients*

fréquence des achats selon l'âge des clients (ss pros)

• Moyenne de la fréquence d'achats par âge



Test de la normalité (Kolmogorov-Smirnov)

• **Sur la fréquence d'achat :**

p-value < 0.05

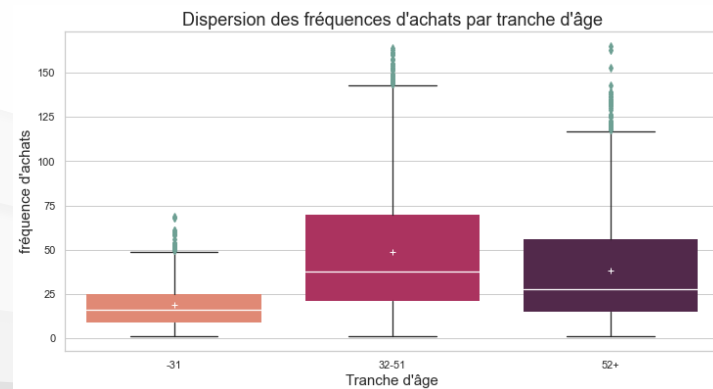
➤ La variable ne suit pas une loi normale

Test de corrélation (Coefficient de Spearman)

p-value > 0.05

➤ Pas de corrélation monotone

• Fréquence d'achats par tranche d'âge



Test de comparaison (Kruskal Wallis)

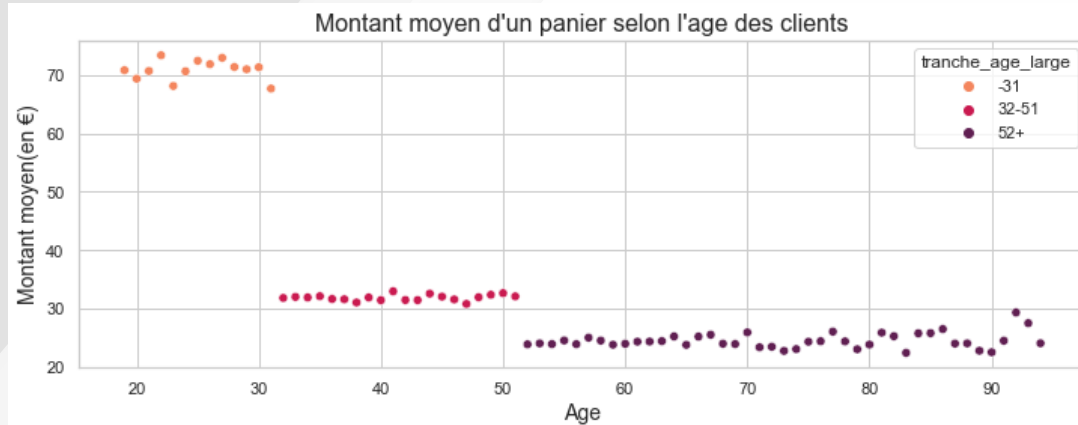
p-value < 0.05

➤ Il existe des différences entre au moins 2 groupes

➤ *Les - de 31 ans ont une fréquence d'achats basse*

Panier et l'âge des clients (ss pros)

- Panier moyen par âge



Test de la normalité (Kolmogorov-Smirnov)

p-value < 0.05

➤ La variable ne suit pas une loi normale

Test de corrélation (Coefficient de Spearman)

coefficient de Spearman : -0.76 / p-value < 0.05

➤ Corrélation monotone décroissante

➤ *Les - de 31 ans ont un panier moyen élevé (X2 par rapport aux autres clients)*

- Nombre moyen d'articles par panier selon l'âge



Test de la normalité (Kolmogorov-Smirnov)

p-value < 0.05

➤ La variable ne suit pas une loi normale

Test de corrélation (Coefficient de Spearman)

coefficient de Spearman : -0.63 / p-value < 0.05

➤ Corrélation monotone décroissante

➤ *Les 32-51 ans ont le plus de produits par panier*

Conclusion

- Le chiffre d'affaires a augmenté en 2 ans mais celui de la catégorie 2 a légèrement baissé.
- La catégorie qui se vend le plus est la catégorie 0 et celles qui génèrent le plus de chiffre d'affaires sont les catégories 1 et 0.
- 4 clients « professionnels » ont été identifiés dont les achats génèrent 7% du chiffre d'affaires.
- Parmi les clients non « professionnels », on a pu distinguer 3 types de comportements selon les tranches d'âge suivantes: (- 31 ans, 32 ans - 51 ans, 52 ans+).

Recommandations pour les prochaines campagnes marketing

- Cibler en priorité, pour les produits de la catégorie 2, les clients âgés de moins de 31 ans
- Créer une offre incitant les plus de 32 ans à commander plus de produits par panier

MERCI



Lapage