

ANALYSE DES VENTES

 **Librairie Lapage**



CONTEXTE ET OBJECTIFS

Librairie Lapage : points de ventes physiques

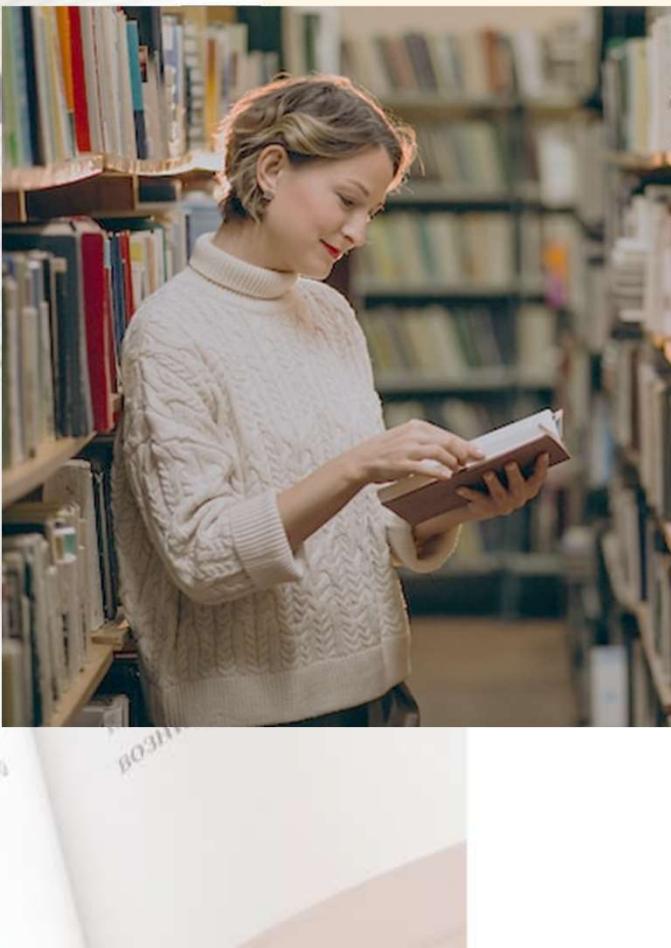
Il y a 2 ans : **Ouverture de la boutique en ligne**



Retour et analyses de cette période de 2 ans :

- **Analyses des indicateurs de ventes** : chiffres clés, KPIs, tendances...
- **Analyse des corrélations et des comportements clients**





SPECIFICATIONS ET NETTOYAGE DES DONNEES



Bases de données internes de Lapage



Période du 1er mars 2021 au 28 février 2023.

Trois ensembles de données :

- ✓ **Fichier Clients** : Informations sur le genre et la date de naissance des clients
- ✓ **Fichier Produits** : Détails sur les articles vendus, incluant leur prix et leur catégorie.
- ✓ **Fichier Transactions** : Historique des achats en ligne (id_clients, date, produit...)

Nettoyage des données :

- ✓ Vérification de l'intégrité des données
- ✓ Suppression des données tests

METHODOLOGIE

Approche basée sur l'exploration des données et la visualisation, afin de dégager des tendances claires et exploitables.



Intégration des données dans un **notebook Jupyter**.

Manipulation et visualisation des données en temps réel.



Analyse des données avec **Python et ses bibliothèques**

- Bibliothèques Matplotlib et Seaborn : pour la création de visualisations.
- Bibliothèque Scipy Stats : pour les analyses statistiques



Cette approche permet :

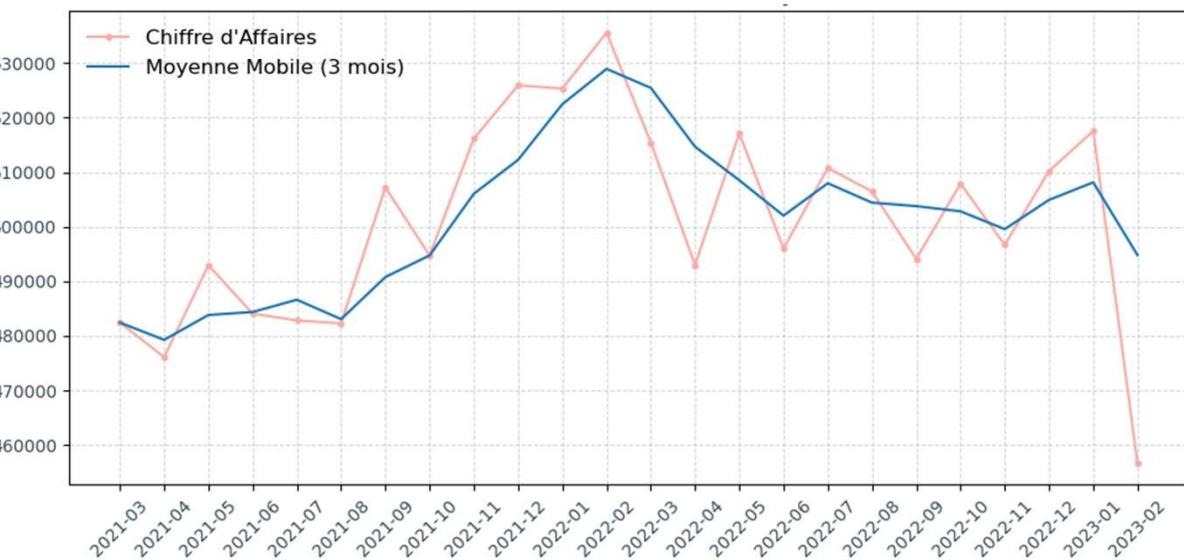
- ❖ de tester différentes hypothèses
- ❖ d'ajuster les analyses au fur et à mesure
- ❖ de documenter chaque étape de manière claire et concise.



ANALYSE DES DIFFERENTS INDICATEURS DE VENTE

EVOLUTION DU CHIFFRE D'AFFAIRES ET MOYENNE MOBILE

Evolution du chiffre d'affaires avec moyenne mobile



Chiffre d'affaires Total
sur la période :
12 027 663€

CHIFFRE D'AFFAIRES PAR CATEGORIE

Evolution du chiffre d'affaires par catégories



Mars 2021 → Janvier 2022

Phase d'adaptation

Février 2022 → Janvier 2023 :

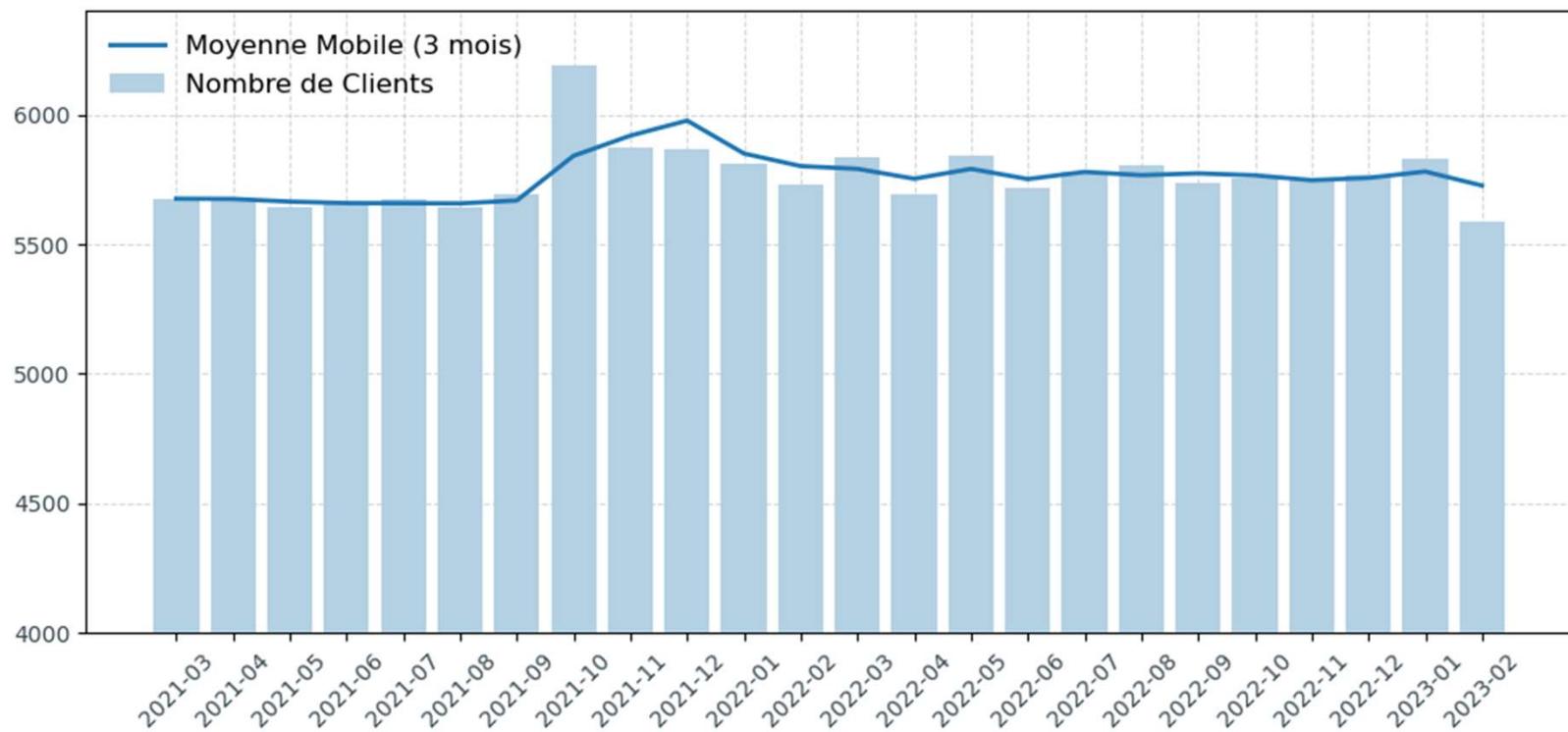
Phase de stabilisation

Février 2023

Décroissance globale importante

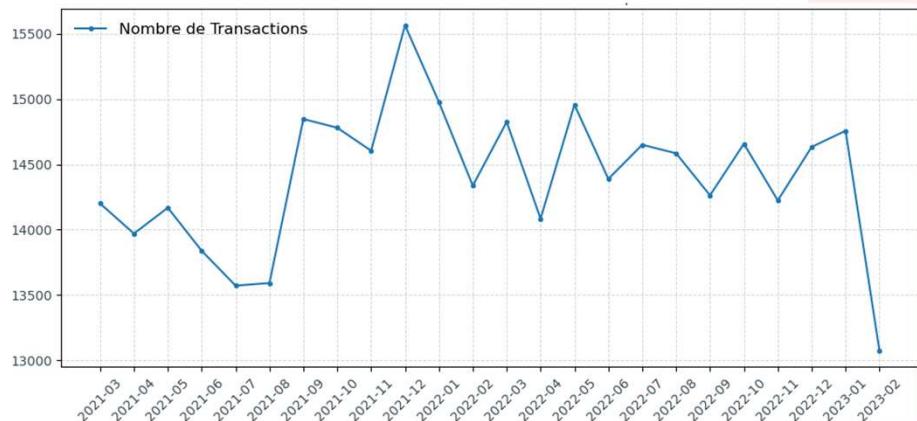
EVOLUTION DU NOMBRE DE CLIENTS PAR MOIS

Nombre de clients uniques par mois (avec moyenne mobile)

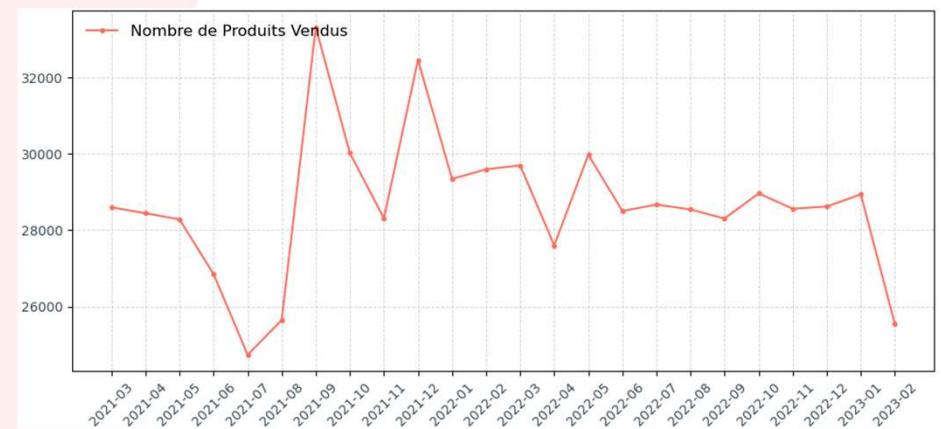


EVOLUTION DU NOMBRE DE TRANSACTIONS MENSUELLES

Nombre de transactions par mois



Nombre de produits vendus par mois

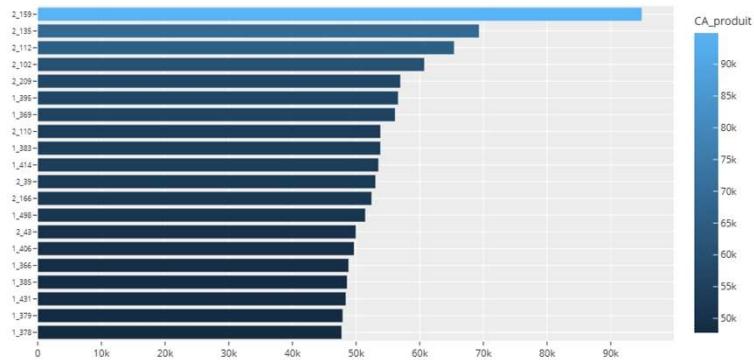


EVOLUTION DU NOMBRE DE PRODUITS VENDUS PAR MOIS

TOP REFERENCES PRODUITS



Top 20 articles (en CA)

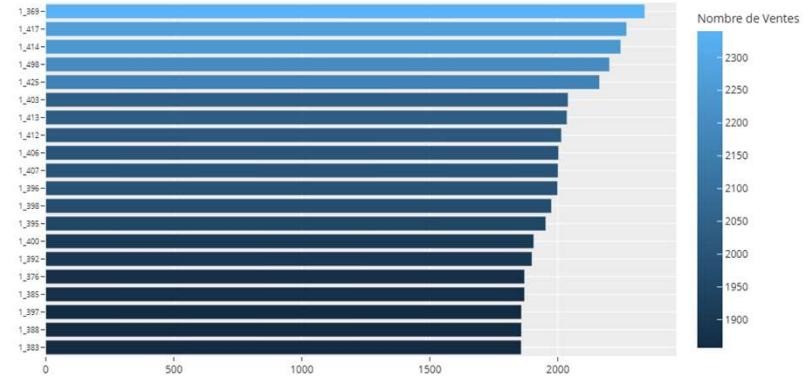


Nbre d'articles représentant 80% du CA :
699 produits

Proportion des articles représentant 80% du CA :
21,27 %

Les 20 produits avec le plus gros CA sont de catégories 1 et 2.
Le Top 5 sont des produits de catégorie 2

Top 20 articles (en nombre de ventes)



Nbre d'articles représentant 80% du nbre de ventes :
809 produits

Proportion des articles représentant 80% du nbre de ventes :
24,62 %

Les 20 produits avec le plus de ventes
sont de catégorie 1

PRODUITS PERFORMANTS MOINS

- **21 produits n'ont réalisé aucune vente.**
(principalement de catégorie 0)

25 % des produits ayant les plus faibles performances :

Calcul interquartile et sélection du premier quart

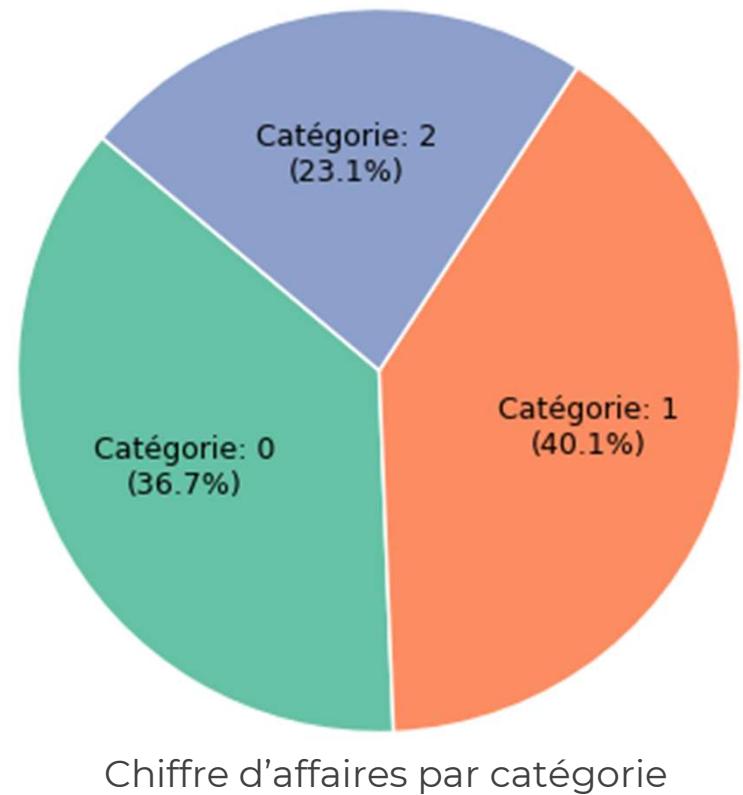
- **805 produits** ont enregistré moins de 16 ventes
- **816 produits** ont réalisé un chiffre d'affaires inférieur à 234 €



REPARTITION PAR CATEGORIES

CATEGORIES	NBRE PRODUITS	NBRE VENTES	CA
Catégorie 0	2308	415 459	4 419 730 €
Catégorie 1	739	235 592	4 827 657 €
Catégorie 2	239	36 483	2 780 275 €

- **Catégorie 0 :** CA important basé sur un grand nombre de ventes. Catégorie avec de nombreux produits.
- **Catégorie 1 :** CA important basé sur moins de ventes. Nbre de produits limités mais prix élevé.
- **Catégorie 2 :** Bon CA basé sur un nombre de ventes très faible. Nbre de produits limité mais plutôt de prix élevé.



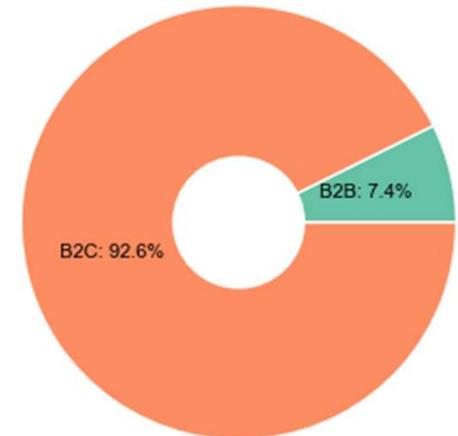
PROFILS CLIENTS & CLIENTS B2B

Clients BtoB : **clients avec volumes de ventes largement supérieur à la moyenne.**

- Calcul du Z-Score pour définir les outliers (z-score > 3)
- **4 clients ont été identifiés comme B2B.**



Chiffre d'affaires B2B
884 296 €
soit **7,4 %** du chiffre d'affaires total.



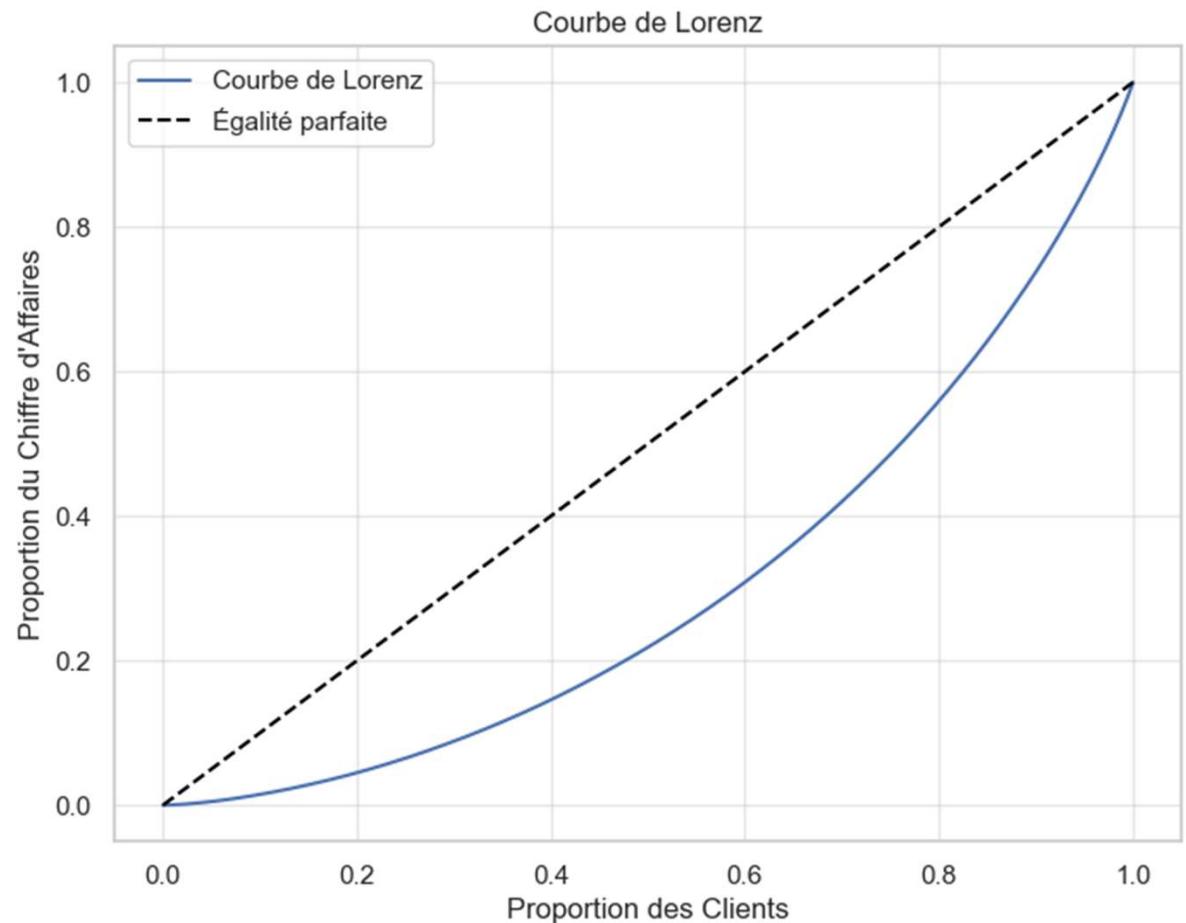
Chiffre d'affaires B2B vs B2C

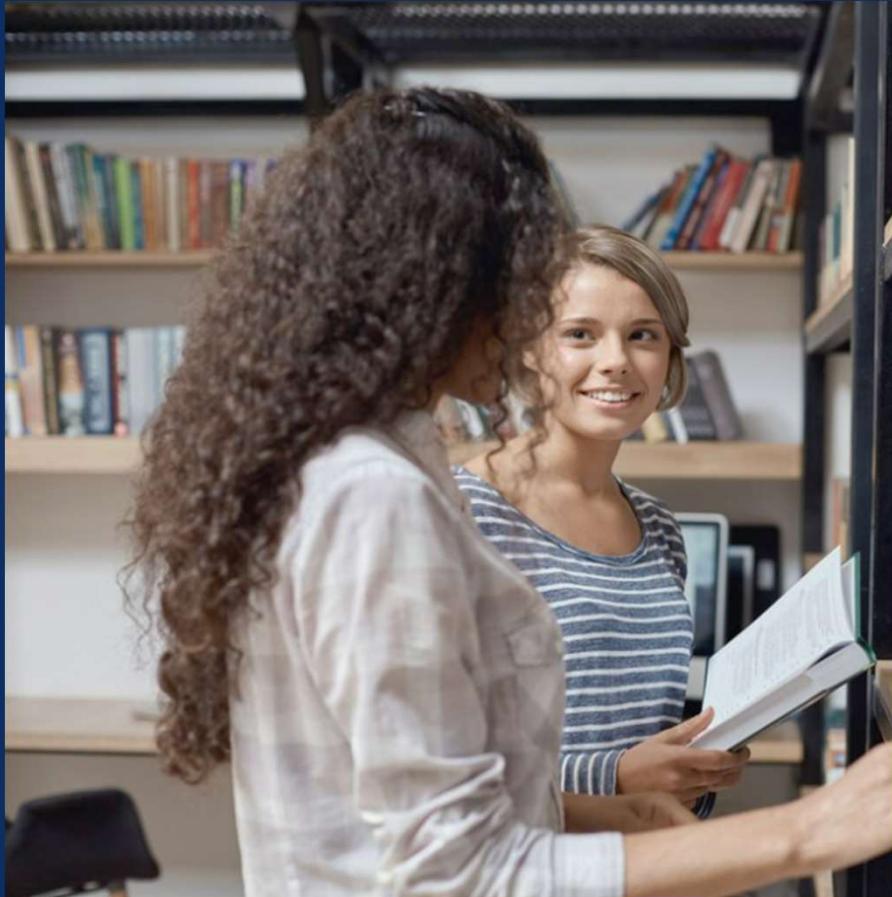
COURBE DE LORENZ

La courbe de Lorenz montre une répartition inégale du chiffre d'affaires entre les clients

- une petite proportion de clients génère une grande part du chiffre d'affaires
- la majorité des clients contribue peu au CA.

Plus la courbe s'éloigne de la diagonale, plus l'inégalité est forte.





ANALYSES CIBLEES ET CORRELATIONS

LIEN ENTRE GENRE ET CATEGORIES DE LIVRES ACHETES

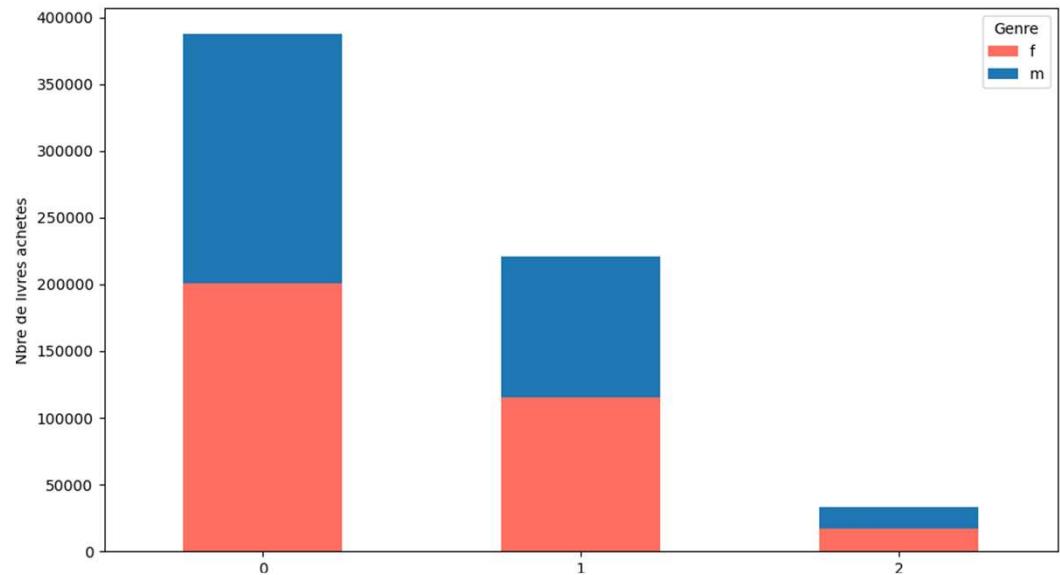
Répartition des clients par genre

Sexe	Nombre	Pourcentage
f	4478	52.09%
m	4118	47.91%

Tableau de contingence
Nbre de livres achetés par genre
et par catégorie

sex	0	1	2
f	200793	115721	16980
m	186488	104884	15868

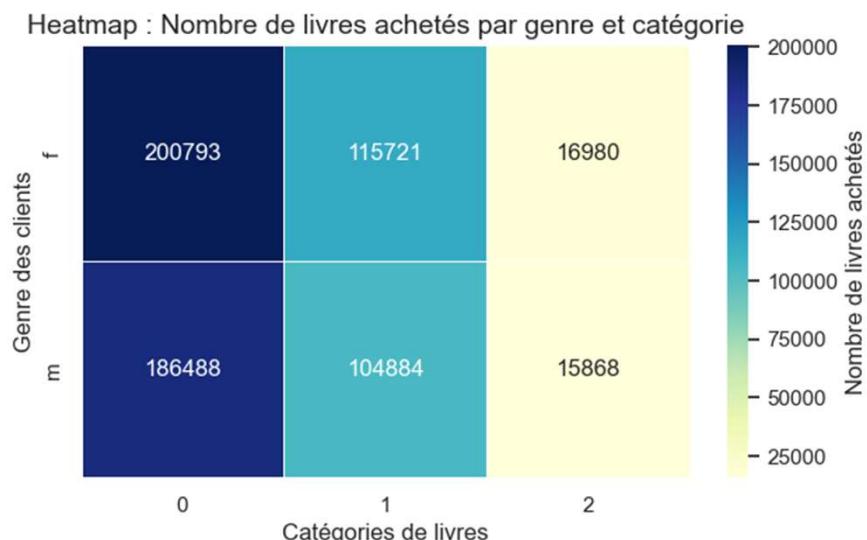
Répartition des ventes par catégories de livres par genre



TEST DE CHI-2

Hypothèse nulle (H_0) : Il n'y a aucune relation entre le genre d'un client et les catégories de livres achetés.

Hypothèse alternative (H_1) : Il existe une relation entre le genre d'un client et les catégories de livres achetés.



Choix du test :

- Variables de type qualitatif
- Tableau de contingence 2*3

Niveau de signification

- Alpha de 5%

Test du Chi-2

- Statistique du Chi-2 : 22.66857
- p-valeur : 0.00001
- Degrés de liberté : 2

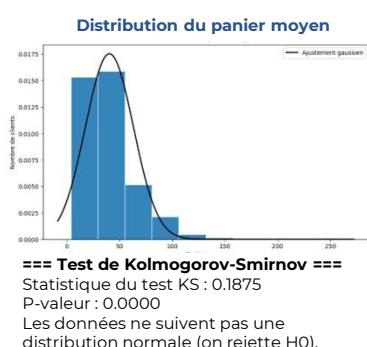
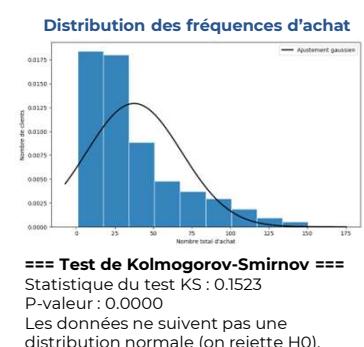
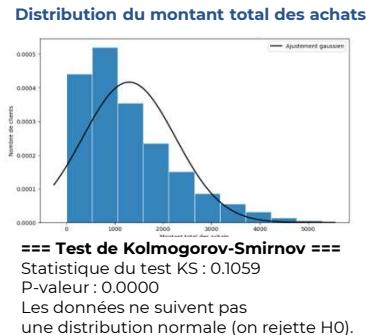
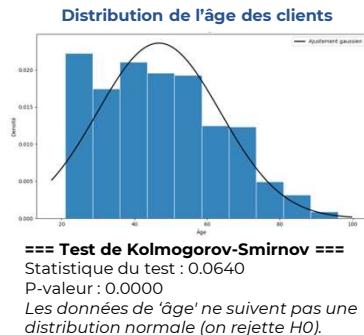
p.value < 0,005 → on rejette l'hypothèse nulle

Il existe bien une association significative statistiquement entre le genre et les catégories de livres.

ANALYSES DES VARIABLES

Hypothèse nulle (H_0) : Les données suivent une **distribution normale**.

Hypothèse alternative (H_1) : Les données **ne suivent pas une distribution normale**.



Visualisations des distributions sous forme d'histogramme, avec courbe ajustement gaussien pour identifier si les distributions se rapprochent d'une courbe normale

Résultats Test Kolmogorov-Smirnov

Les données ne suivent pas une distribution normale. On rejette l'hypothèse nulle.

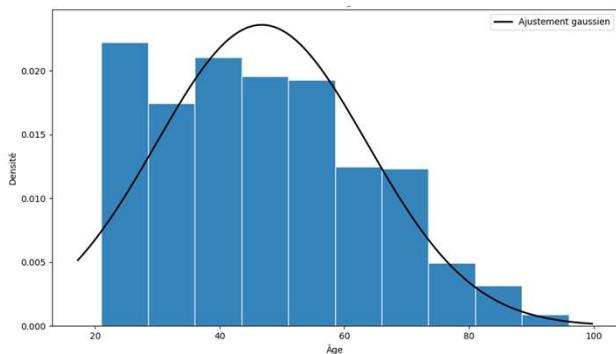
Données :

- Données quantitatives
- Données non paramétriques

➡ **Sélection du Test de Spearman** pour les analyses de corrélations

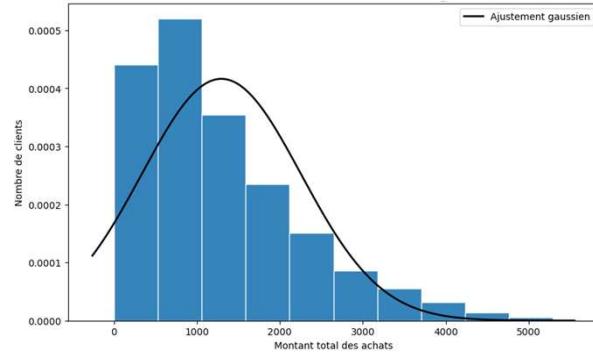
LIEN ENTRE AGE DES CLIENTS ET MONTANT TOTAL DES ACHATS

Distribution de l'âge des clients



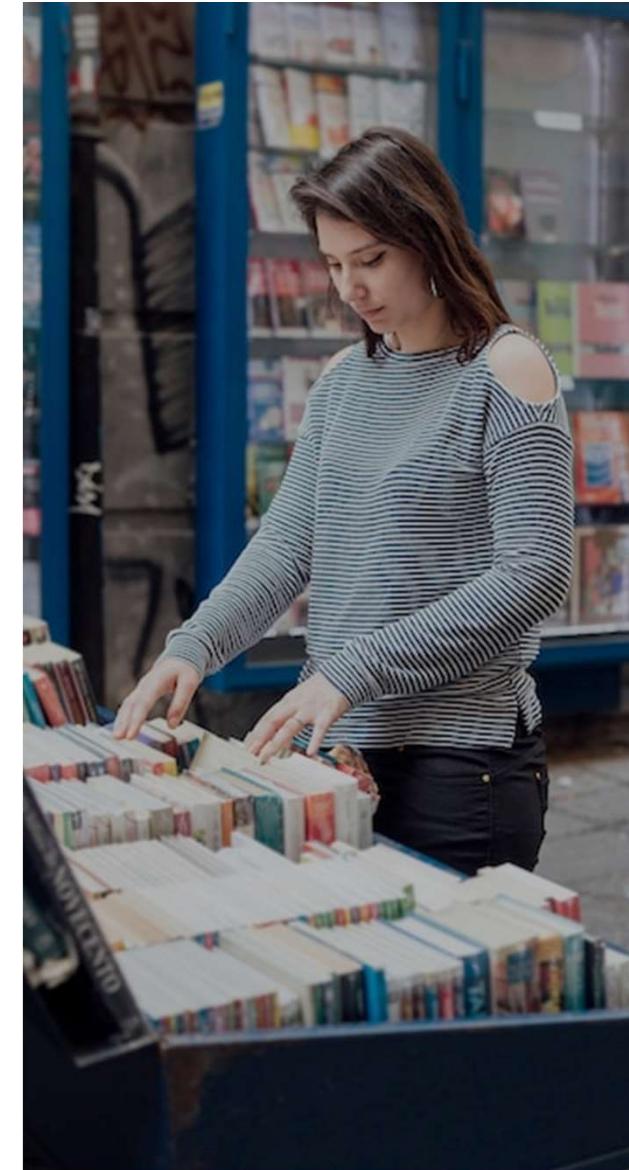
- Moyenne de l'âge : 46.74 ans
- Médiane de l'âge : 46.00 ans
- Écart-type de l'âge : 16.91 ans
- Variance de l'âge : 285.97

Distribution du montant total des achats



- Moyenne du CA : 1 296.34 €
- Médiane du CA : 1045.59 €
- Écart-type du CA : 958.40 €
- Variance du CA : 918539.27

la plupart des clients ont un montant total d'achats relativement faible, mais quelques clients ont des montants beaucoup plus élevés

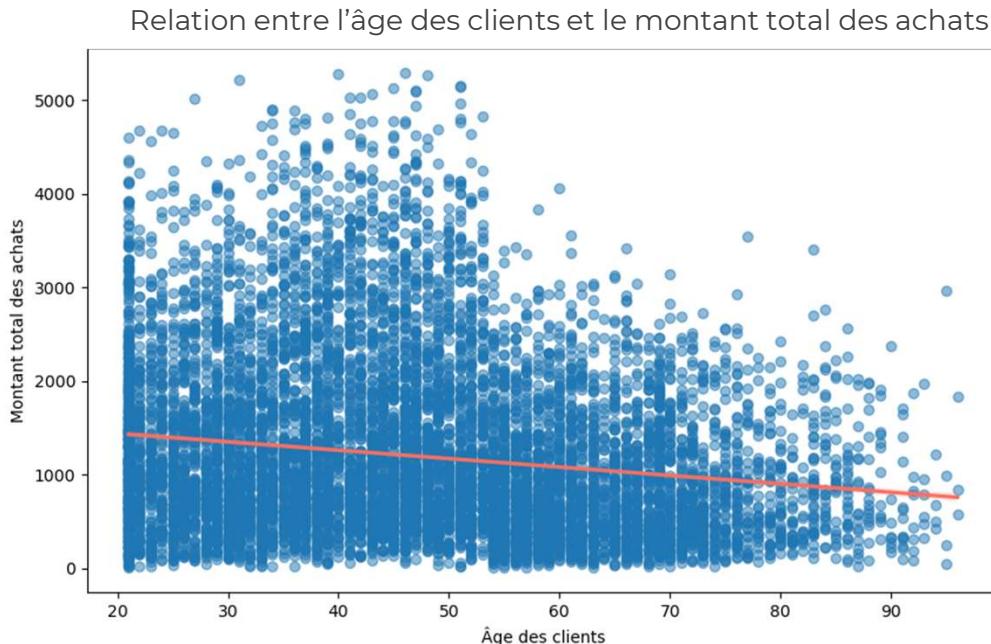


TEST DE CORRELATION DE SPEARMAN

LIEN ENTRE AGE DES CLIENTS ET MONTANT TOTAL DES ACHATS

Hypothèse nulle (H_0) : Il n'y a aucune relation entre l'âge des clients et le montant total des achats

Hypothèse alternative (H_1) : Il existe une relation entre l'âge des clients et le montant total des achats



Choix du test de Spearman

- Données quantitatives
- Données non paramétriques

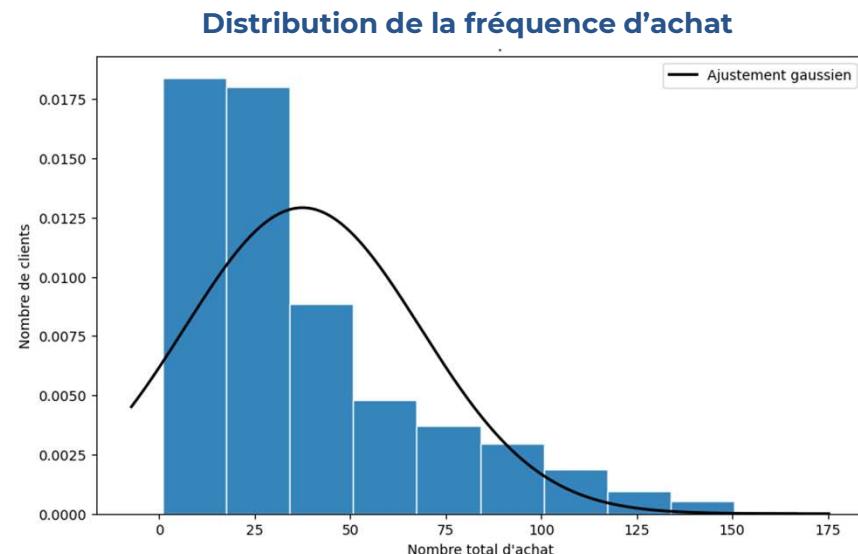
Test de Spearman

- Coef.t de corrélation : -0.1845
- Valeur p: 1.0212910436382683e-66

p.value < 0,005 → on rejette l'hypothèse nulle

lorsque l'âge augmente, le montant total des achats tend à diminuer.
Mais le coefficient est proche de 0, la relation est faible.

LIEN ENTRE AGE DES CLIENTS ET FREQUENCE D'ACHAT

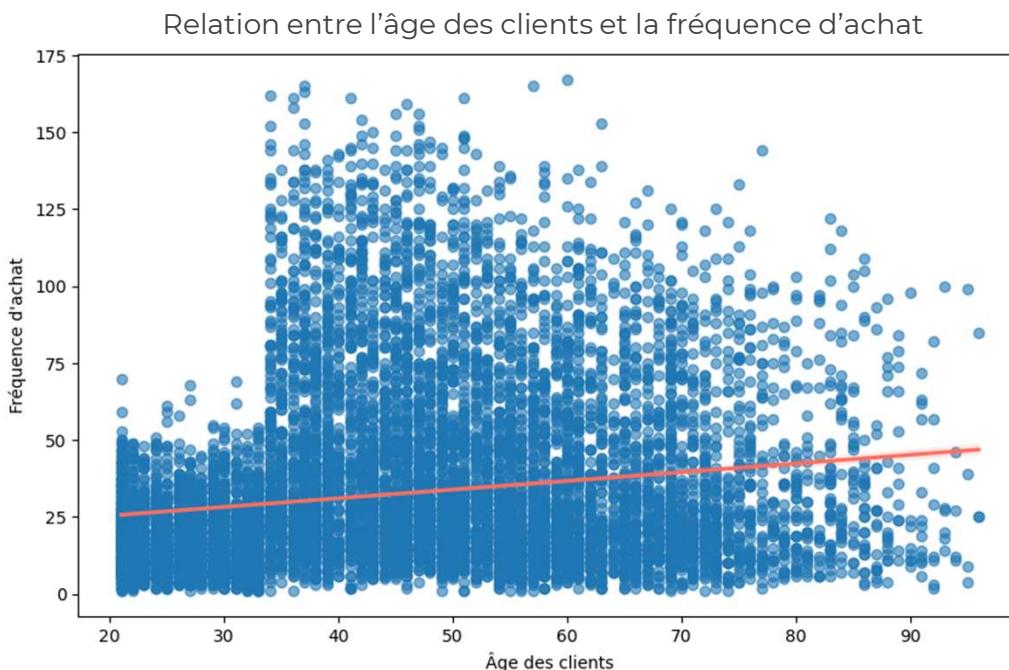


- **Moyenne de la fréquence d'achat : 37.51 fois**
- Médiane de la fréquence d'achat : 27.00 fois
- Écart-type de la fréquence d'achat : 30.90 fois
- Variance de la fréquence d'achat : 954.74

distribution asymétrique vers la droite.
*la majorité des clients ont une fréquence d'achat faible
Un petit nombre de clients effectuent des achats très fréquents*

TEST DE CORRELATION DE SPEARMAN

LIEN ENTRE AGE DES CLIENTS ET FREQUENCE D'ACHAT



Hypothèses :

- (H_0) : Il n'y a aucune relation entre l'âge des clients et la fréquence d'achat
- (H_1) : Il existe une relation entre l'âge des clients et la fréquence d'achat

Observations

- Relation monotone légèrement croissante

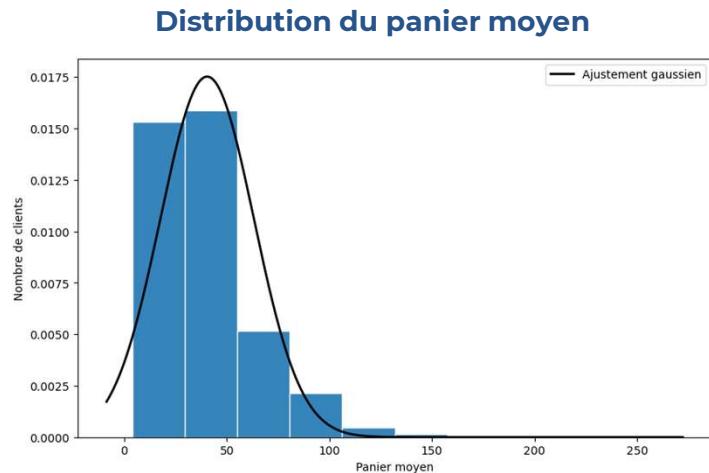
Test de Spearman

- Coef.t de corrélation : 0.211
- Valeur p: 6.629168433162815e-88

p.value < 0,005 → on rejette l'hypothèse nulle

On observe une légère augmentation de la fréquence d'achat avec l'âge des clients

LIEN ENTRE AGE DES CLIENTS ET PANIER MOYEN



- Moyenne du panier moyen : 40.40 €
- Médiane du panier moyen : 32.34 €
- Écart-type du panier moyen : 22.76 €
- Variance du panier moyen : 518.20

Distribution asymétrique vers la droite avec une longue queue

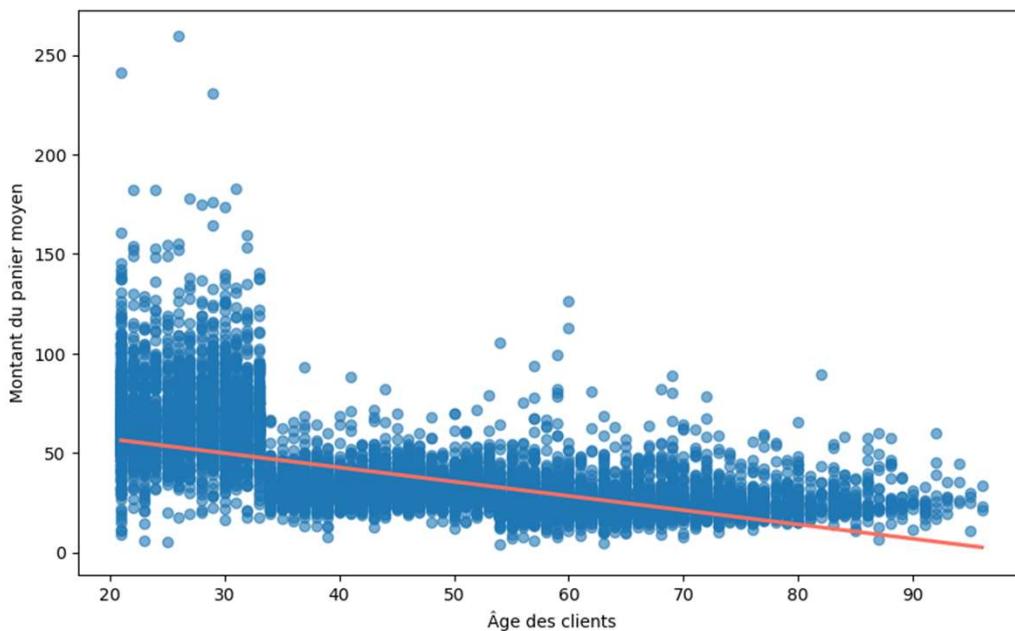
La plupart des clients ont un panier moyen faible, tandis que quelques-uns ont un panier moyen élevé



TEST DE CORRELATION DE SPEARMAN

LIEN ENTRE AGE DES CLIENTS ET TAILLE DU PANIER MOYEN

Relation entre l'âge des clients et le montant du panier moyen



Hypothèses :

- (H_0) : Il n'y a aucune relation entre l'âge des clients et le montant du panier moyen
- (H_1) : Il existe une relation entre l'âge des clients et le montant du panier moyen

Observations

- Relation monotone décroissante forte

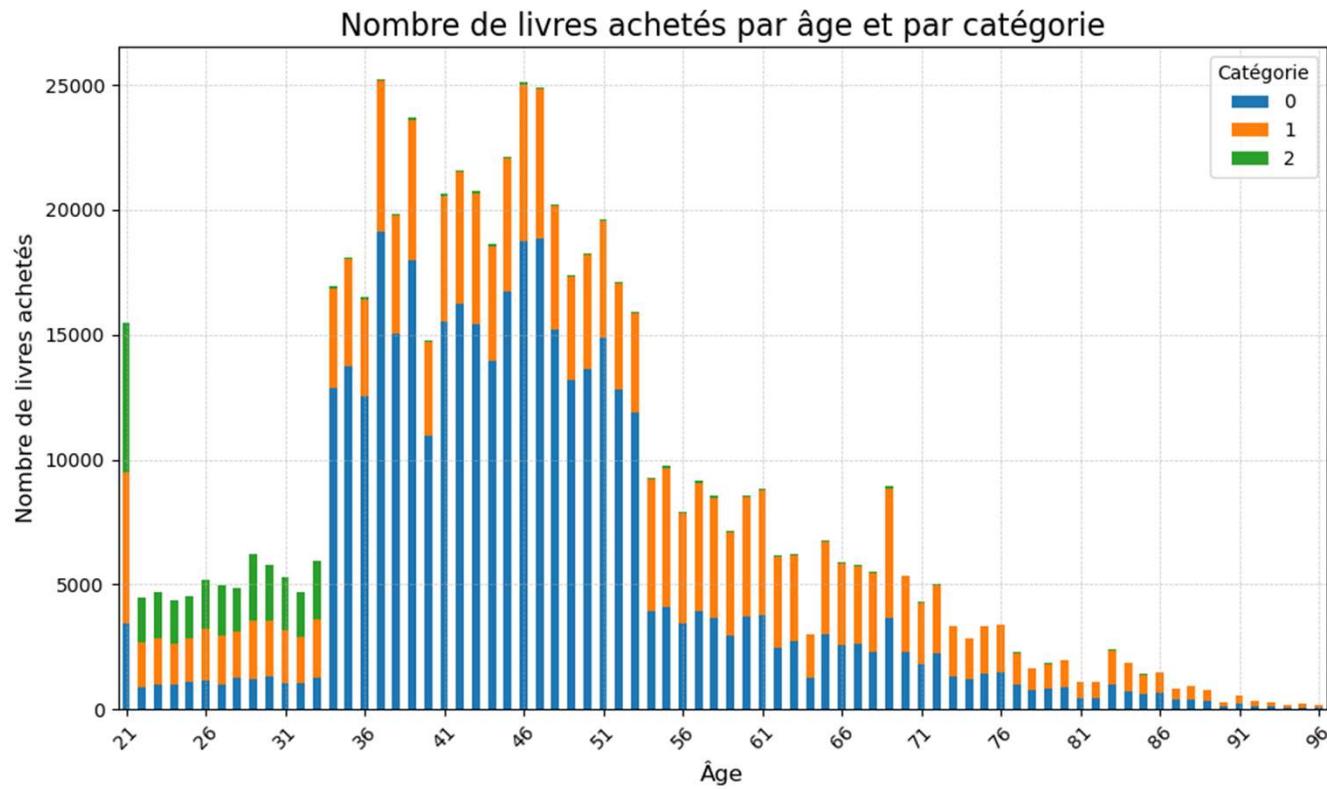
Test de Spearman

- Coef. de corrélation : -0,0700
- Valeur p: 0.0

p.value < 0,005 → on rejette l'hypothèse nulle

plus l'âge des clients augmente,
plus le montant moyen de leur panier diminue.

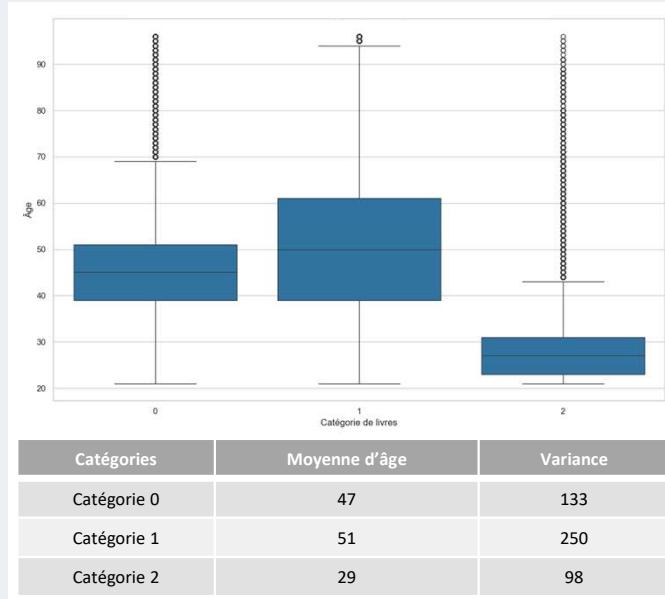
LIEN ENTRE AGE DES CLIENTS ET CATEGORIE DES LIVRES ACHETES



TEST DE KRUSKAL-WALLIS

LIEN ENTRE AGE DES CLIENTS ET CATEGORIE DES LIVRES ACHETES

Distribution des âges par catégories de livres



Test de Levene pour déterminer la constance des variances

Statistique de Levene: 24651.780367284307

p-valeur: 0.0

Conclusion : Les variances entre les groupes sont significativement différentes ($p < 0.05$).

Hypothèses :

- **(H₀)** : Il n'y a aucune relation entre l'âge des clients et les catégories de livres achetés
- **(H₁)** : Il existe une relation entre l'âge des clients et les catégories de livres achetés

Observations

- Relation monotone légèrement croissante

Test Kruskal-Wallis (Anova non paramétrique) :

- Statistique de Kruskal-Wallis : 71359.73412
- P-valeur : 0.00000

p.value < 0,005 → on rejette l'hypothèse nulle

Conclusion : Il existe un lien significatif sur le plan statistique entre les âges des clients et les catégories de livres.

RESULTATS ET LIMITES DES TESTS STATISTIQUES

Résultats clés

- **Genre et Catégories de Livres :**

Une association significative existe entre genre et catégories de livres acheté (*Test du Chi-2*).

- **Âge et Montant Total des Achats :**

Une corrélation montre que les clients plus âgés dépensent en moyenne moins (*Test de Spearman*).

- **Âge et Fréquence d'Achat :**

Une corrélation faible indique que la fréquence des achats augmente légèrement avec l'âge (*Test de Spearman*).

- **Âge et Taille du Panier Moyen :**

Une corrélation forte révèle que la taille du panier moyen diminue avec l'âge (*Test de Spearman*).

- **Âge et Catégories de Livres :**

L'âge influence les préférences d'achat de catégories de livres (*Test de Kruskal-Wallis*).

Limites des analyses

- **Taille Échantillon Temporel :**

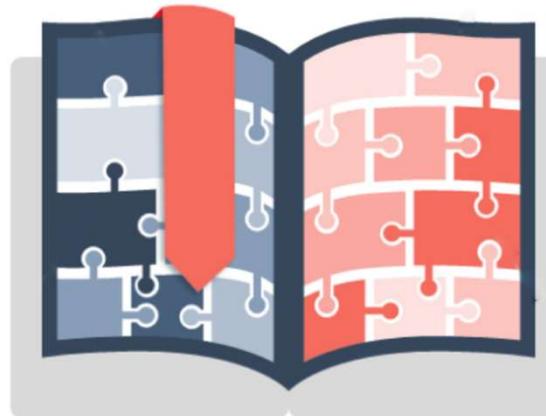
Les données couvrent une période de deux ans

- **Corrélation vs Causalité :**

Les relations identifiées sont descriptives et ne permettent pas de conclure à des liens causaux.

- **Erreurs Statistiques :**

Des erreurs de type I ou II pourraient survenir, influençant la robustesse des conclusions.



MERCI
POUR VOTRE
ATTENTION

