

Analysez les ventes d'une librairie

Par OpenClassRooms

Contexte et spécification des données



L'entreprise Lapage est une librairie originellement physique

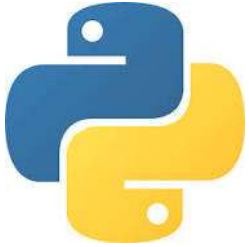
Analysez les ventes d'une librairie

Reprendre la base de données, les outils étant artisanaux

- Faire une jointure entre les database
- Analyse autour du Chiffre d'affaires
- Analyse sur les références, tops et flops
- Quelques informations sur les profils de nos clients
- Faire une corrélation entre certaines données quantitatives

Méthodologie de l'analyse

Outils utilisés



Langage de programmation, conçu pour les analyses de données



Jupyter est un environnement de développement interactif basé sur le web pour les notebooks, le code et les données.

Ressources



customers.csv



products.csv



Transactions. csv

Consolidations des données

Jonction du fichier df_customers et df_transactions et products

```
#Fusion des fichiers df_customers et df_transactions
df_merge = pd.merge(df_customers, df_transactions, on='client_id', how='outer')
#Fusion des fichiers df_merge et df_products
df_merge = pd.merge(df_merge, df_products, on='id_prod', how='outer')
```

Clé utilisé : df_customers
df_products

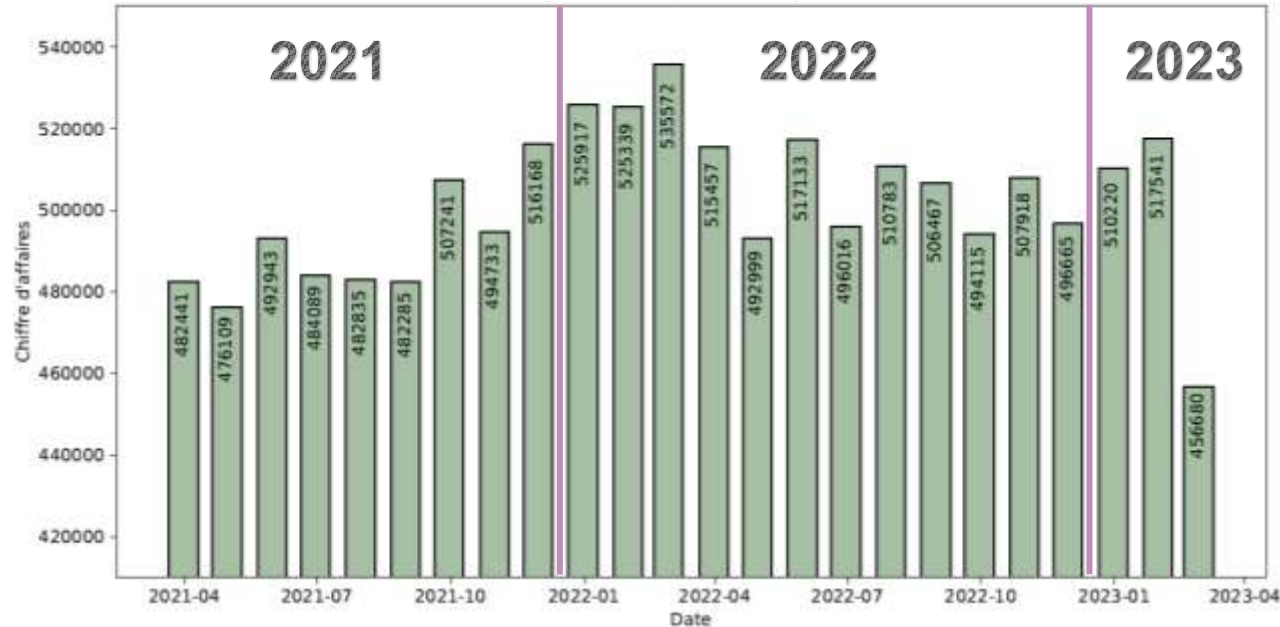
	client_id	sex	birth	id_prod	date	session_id	price	categ
676330	c_8253	f	2001.0	NaN	NaT	NaN	NaN	NaN
676331	c_3789	f	1997.0	NaN	NaT	NaN	NaN	NaN
676332	c_4406	f	1998.0	NaN	NaT	NaN	NaN	NaN
687555	NaN	NaN	NaN	0_1016	NaT	NaN	35.06	0.0
687556	NaN	NaN	NaN	0_1780	NaT	NaN	1.67	0.0
687557	NaN	NaN	NaN	0_1062	NaT	NaN	20.08	0.0

- 21 clients qui n'ont pas fait d'achats.

- 21 produits qu'on pas était acheté

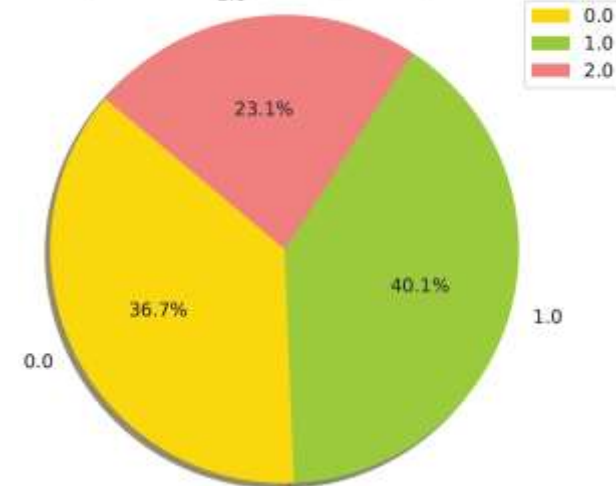
Consolidations des données

Analyse du chiffre d'affaires par mois



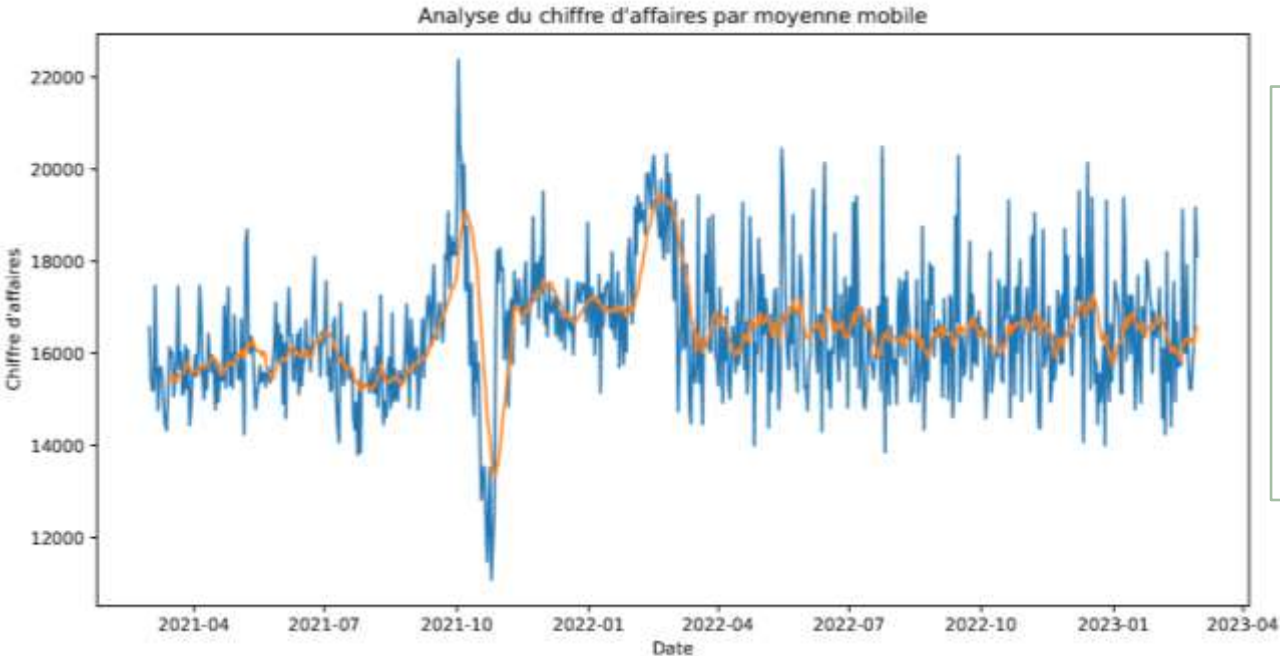
CA total = 12 027 663 €

Répartition des catégories par CA



Analyse du chiffre d'affaires

Moyenne mobile avec fenêtre de 2 semaines



➤ Fluctuation

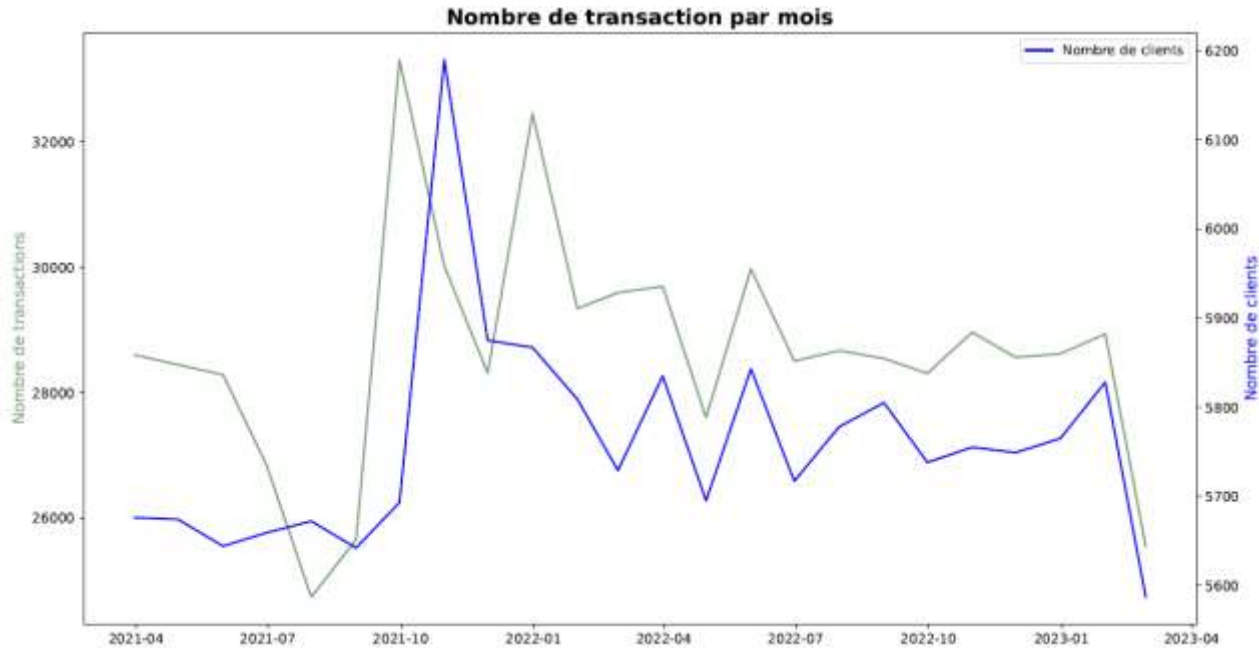
Pic haute et redescente

Hausse très importante de la tendance du CA à partir de septembre 2021, pic en novembre 2021.

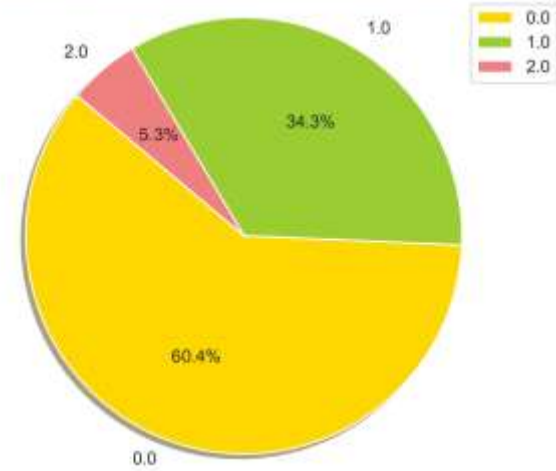
Retournement de tendance et chute du CA jusqu'à la fin de l'année.

Augmentation du CA et récupération en janvier 2022.

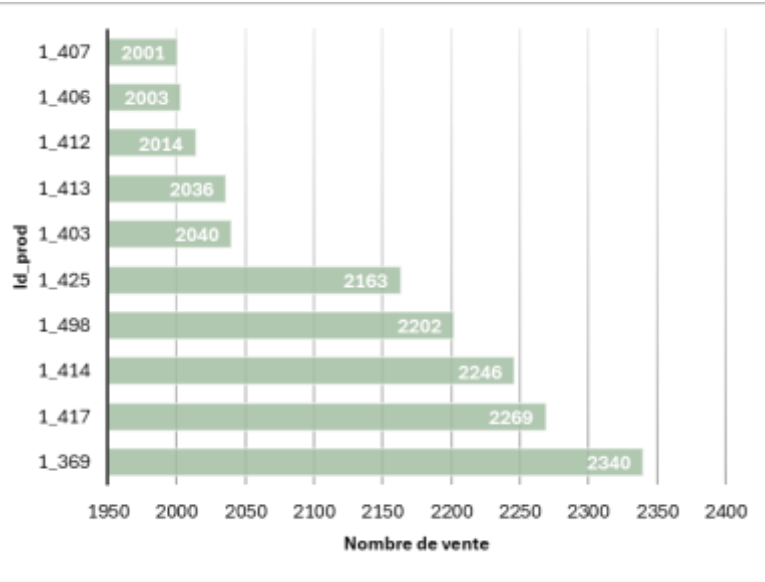
Analyse des transactions et clients



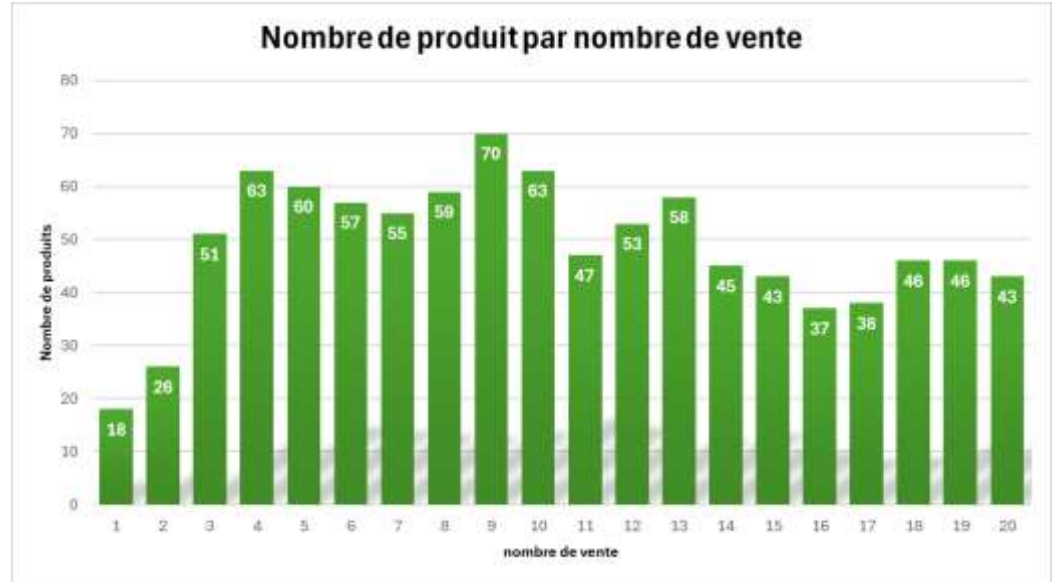
Répartition des catégories par nombre d'article



Analyse du nombre de vente par article : tops & flops

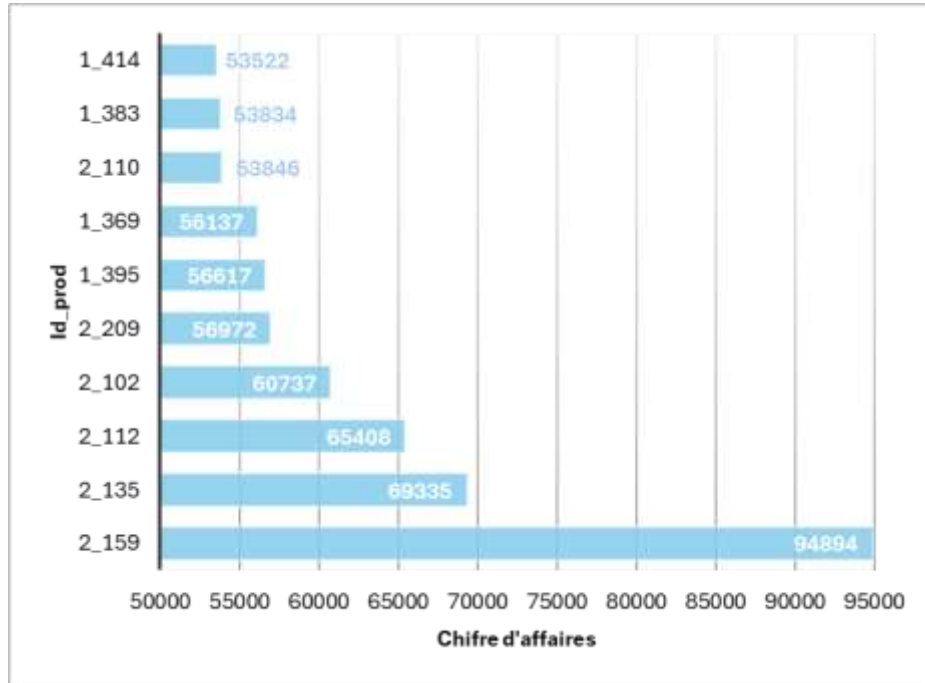


- Moyenne : 210 ventes

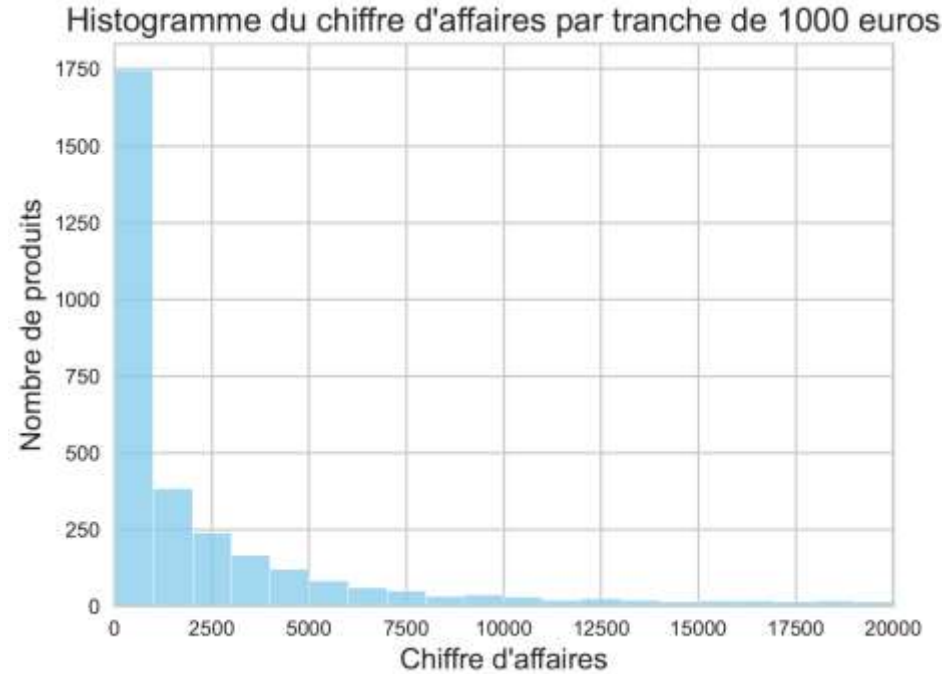


- Médiane : 57 ventes

Analyse du CA par article: tops & flops



- Moyenne : 3683 € CA



- Médiane : 810 € CA

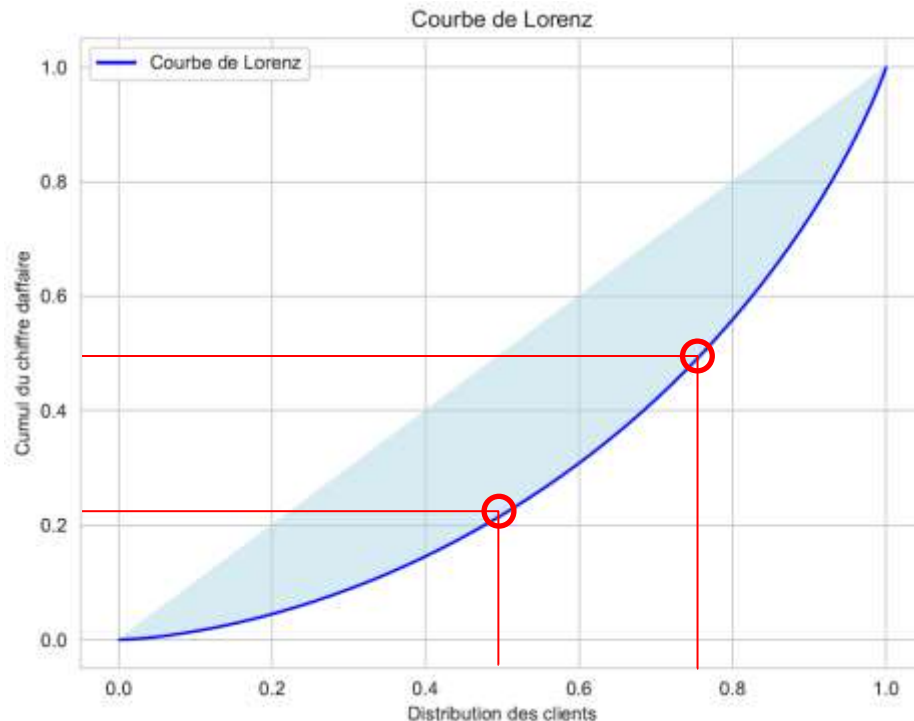
Informations sur les profils de nos clients

Client BtoB

Chiffre d'affaires: 884 296,09 €

7.35 % du CA total

ClientID	CA_par_client
c_1609	326 039,89 €
c_4958	290 227,03 €
c_6714	153 918,60 €
c_3454	114 110,57 €



L'indice de Gini est de 0.40

- 50 % des clients contribuent à hauteur de 20 % du chiffre d'affaires.
- 50 % du chiffre d'affaires est généré par seulement 23 % des clients

Analyse des corrélations

Corrélation entre le genre de client et catégories de livres achetés

Tableau de contingence

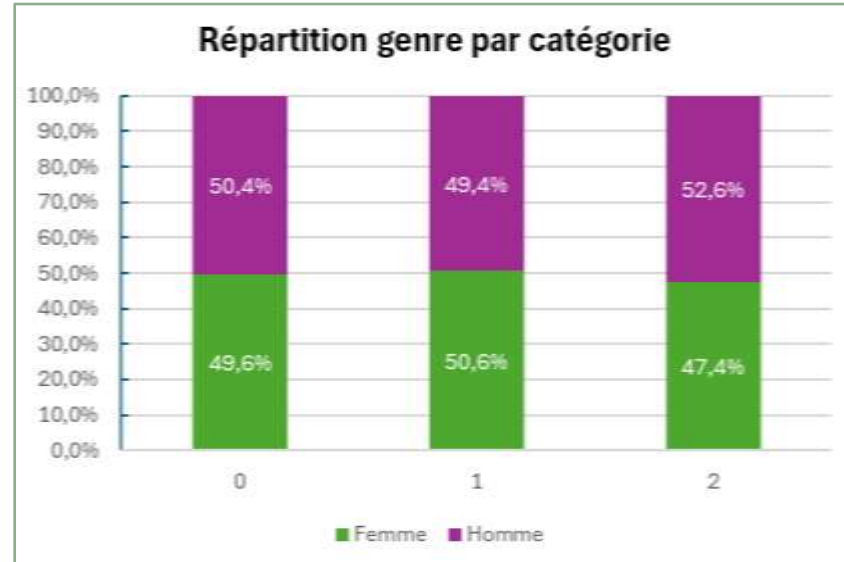
categ	0.0	1.0	2.0	Total
sex				
f	206103	119307	17283	342693
m	209356	116285	19200	344841
Total	415459	235592	36483	687534



Test du Chi²

Test du chi² : 158
P-valeur : 1.3e-31
Degrés de liberté : 6

P-value < 5% = rejette l'hypothèse nulle



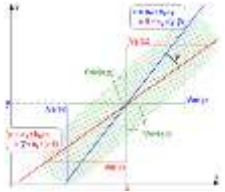
Les variables ne sont probablement pas indépendantes. Il y a une forte relation entre elles.

Analyse des corrélations

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

Coefficient de Pearson

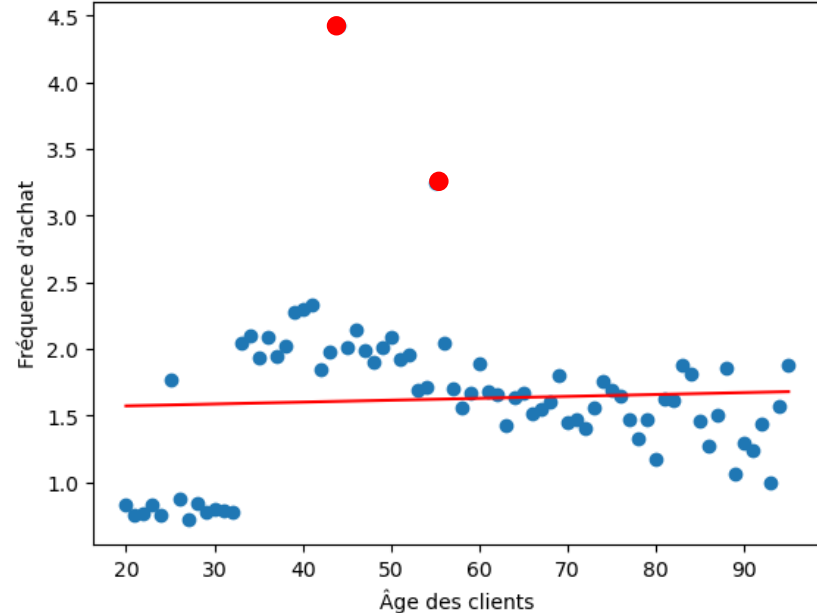
	Age	frequence_achat
Age	1.000000	0.054824
frequence_achat	0.054824	1.000000



Pearson : 0.054
P-valeur : 0.638
Degrés de liberté : 4

P-value > 5% = corrélation observée
n'est pas statistiquement significative

Relation entre l'âge des clients et leur fréquence d'achat



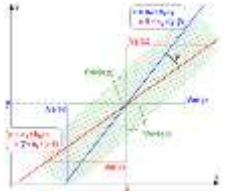
Corrélation entre ces deux variables est minime on
peut la considérer négligeable.

Analyse des corrélations

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

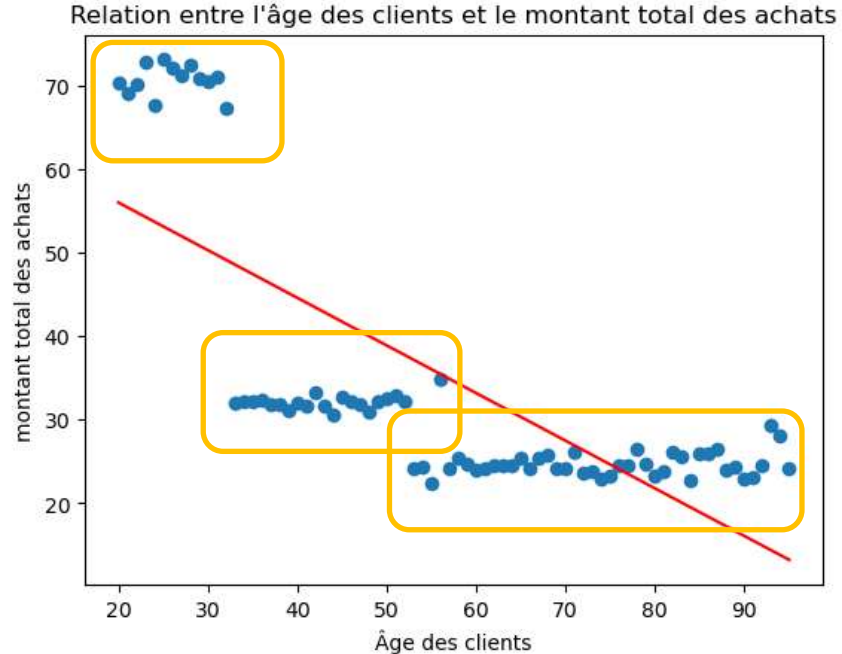
Coefficient de Pearson

	Age	Total_achat
Age	1.000000	-0.747360
Total_achat	-0.747360	1.000000



Pearson : - 0.747
P-valeur : 8.88e-15
Degrés de liberté : 4

P-value < 5% = rejette l'hypothèse nulle



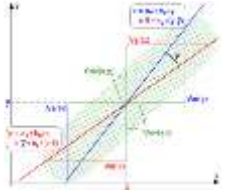
Corrélation négative modérée : Lorsque le prix des produits augmente, les ventes ont tendance à diminuer

Analyse des corrélations

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

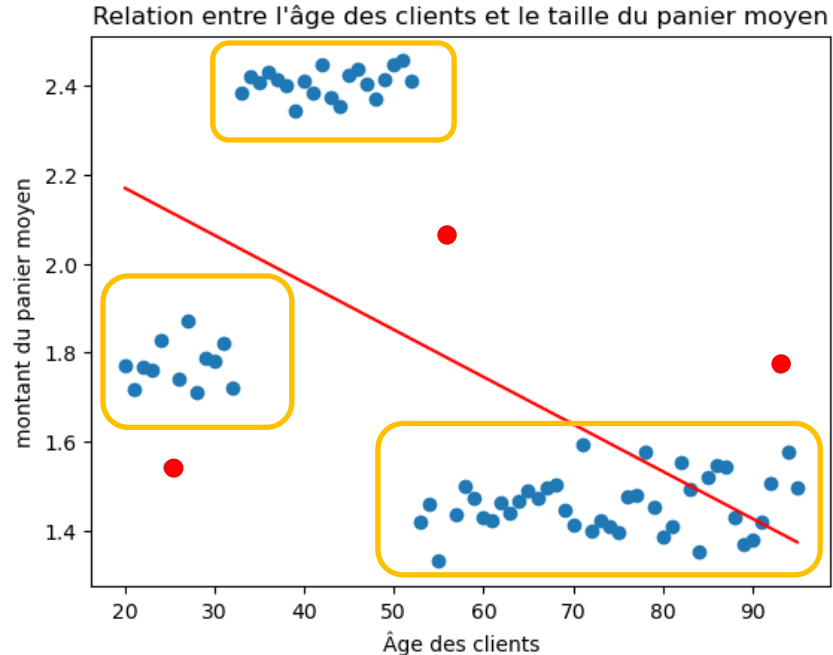
Coefficient de Pearson

	Age	Taille_panier_moyen
Age	1.000000	-0.576468
Taille_panier_moyen	-0.576468	1.000000



Pearson : - 0.576
P-valeur : 5.05e-08
Degrés de liberté : 4

P-value < 5% = rejette l'hypothèse nulle



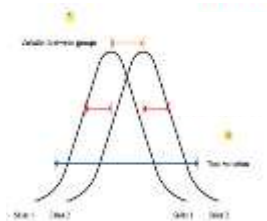
Corrélation entre ces deux variables est minime on peut la considérer négligeable.

Analyse des corrélations

Corrélation entre l'âge des clients et la catégorie de livre achetés

Test ANOVA

Statistique de test F : 45120
Valeur p associée : 0.0



Probabilité d'observer une telle différence de moyennes entre les groupes par hasard est extrêmement faible

