
LIBRAIRIE LAPAGE

Romain RAMAT 24/05/2025

SOMMAIRE

Information sur les données :

- Contexte et Présentation des manipulations sur le jeu de données

Indicateurs de ventes :

- Chiffre d'affaires avec la moyenne mobile
- Chiffre d'affaires par catégorie,
- Nombre de clients par mois,
- Nombre de transactions,
- Nombre de produits vendus
- Top, Flop, Répartition par catégorie

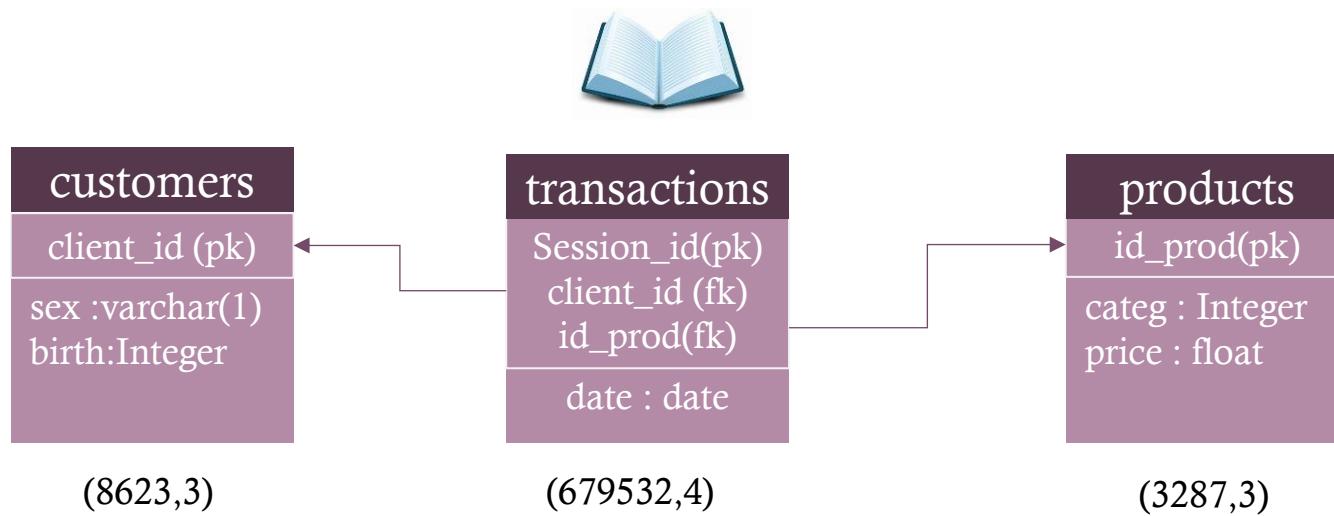
Analyses Statistique des comportements clientèle en ligne

- Lien entre genre d'un client et les catégories des livres achetés
 - Lien entre âge du client et le montant de ses achats
 - Lien entre âge du client et ses fréquences d'achats
 - Lien entre âge du client et la taille du panier moyen
 - Lien entre âge du client et les catégories des livres achetés
-

CONTEXTE

- A l'origine l'entreprise Lapage, était une librairie physique avec des points de vente, elle a eu l'opportunité d'ouvrir un site en ligne il y a deux ans (Mars 2021), et souhaite un bilan de cette activité.
- Ce rapport a pour objectif de présenter les chiffres clés et ainsi déterminer la stratégie à venir.

Celui-ci a été construit sur base des 3 tables « Customers, Transactions et Products » qui reprennent les données brutes (non-nettoyés) des deux dernières années (03/2021 – 02/2023).

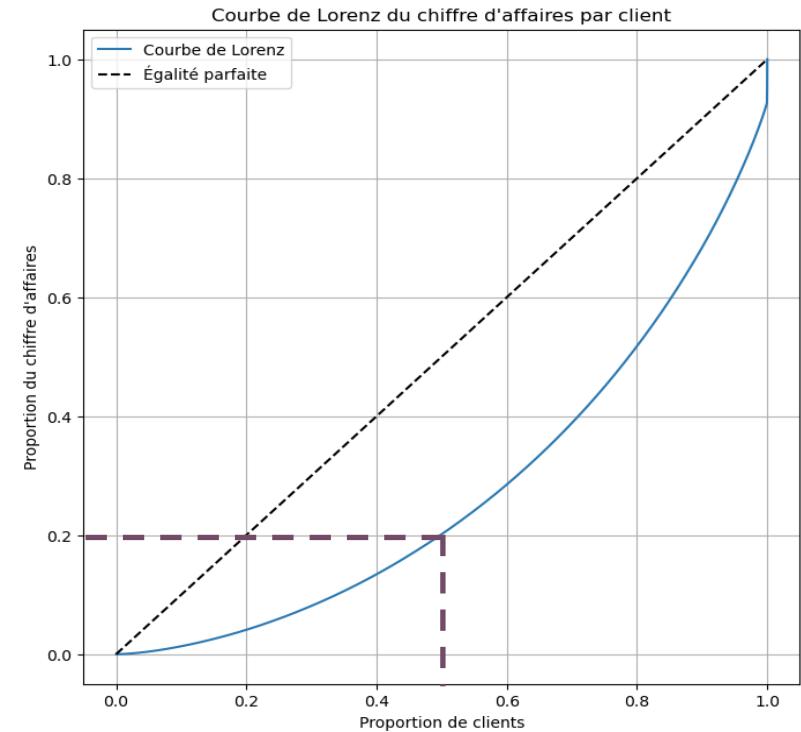
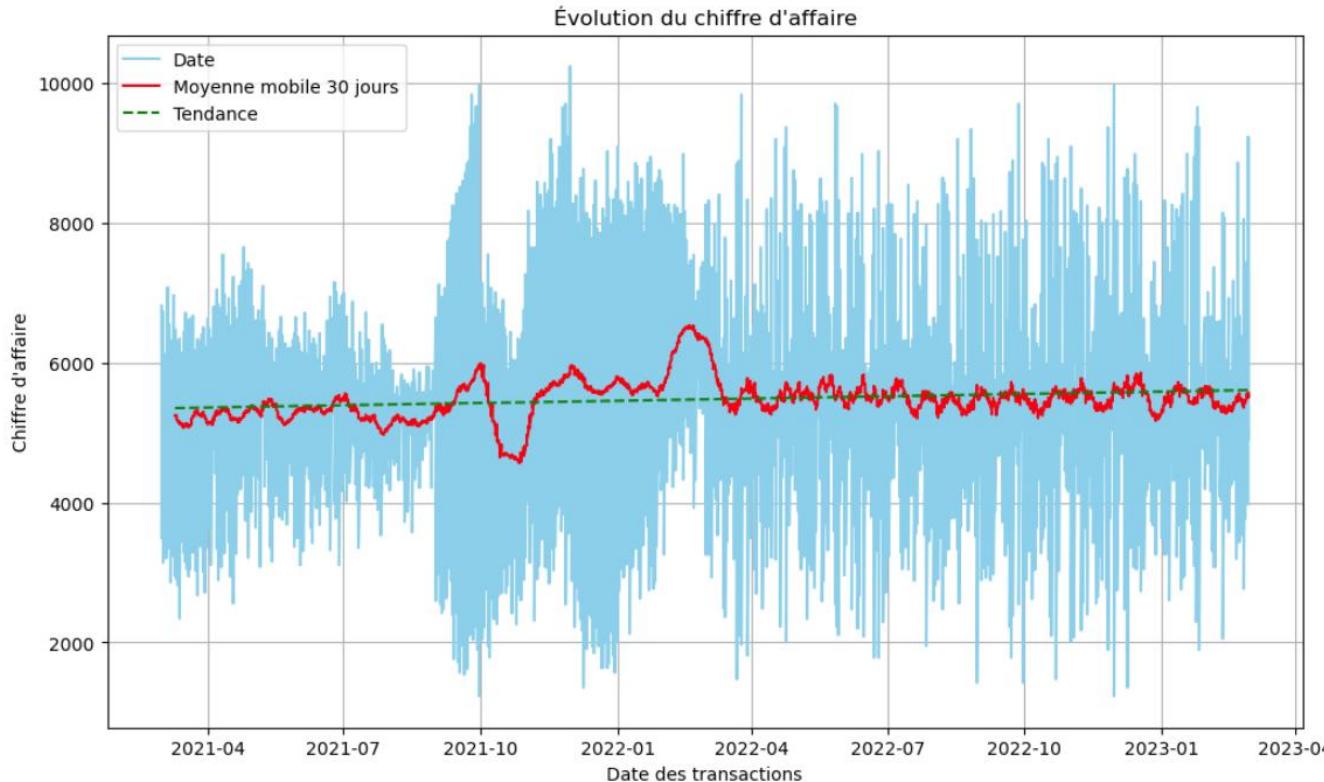


IDENTIFICATION DES ANOMALIES DES DONNÉES

- Retrait dans la table « Customers » :
 - 2 clients tests (ct_0 et ct_1)
- Retrait dans table « Product » :
 - 1 produit à prix négatif « T_0 » de test
- Retrait dans table « Transaction » :
 - 221 valeurs de test sur la journée du 01/03/2021
- Dans le DataFrame mergé :
 - 221 prix de catégorie 0 moyennés à 10,63
 - Retrait des clients sans achats
 - Retrait du mois d'octobre car aucune vente des produits de catégorie 1 entre le 1 au 28/10/2021
 - 4 Clients atypiques : « c_1609 », « c_4958 », « c_6714 », « c_3454 » retirés pour les calculs statistiques

INDICATEURS DE VENTES

- Evolution du chiffre d'affaires avec la moyenne mobile (sur 30 jours)

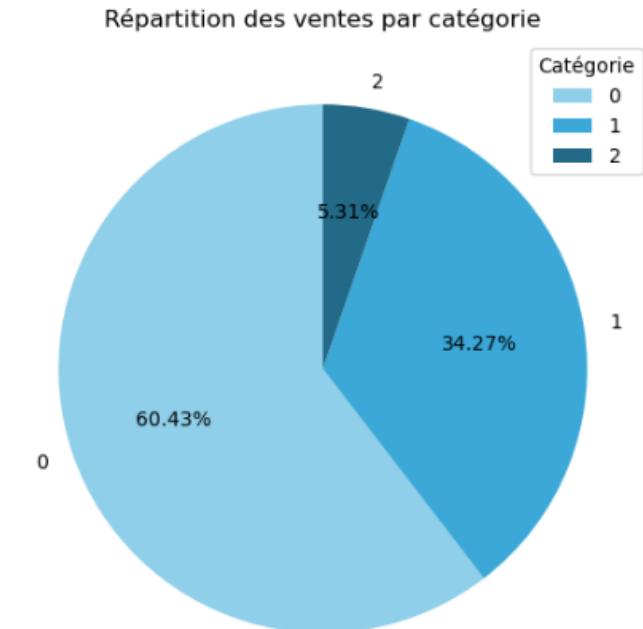
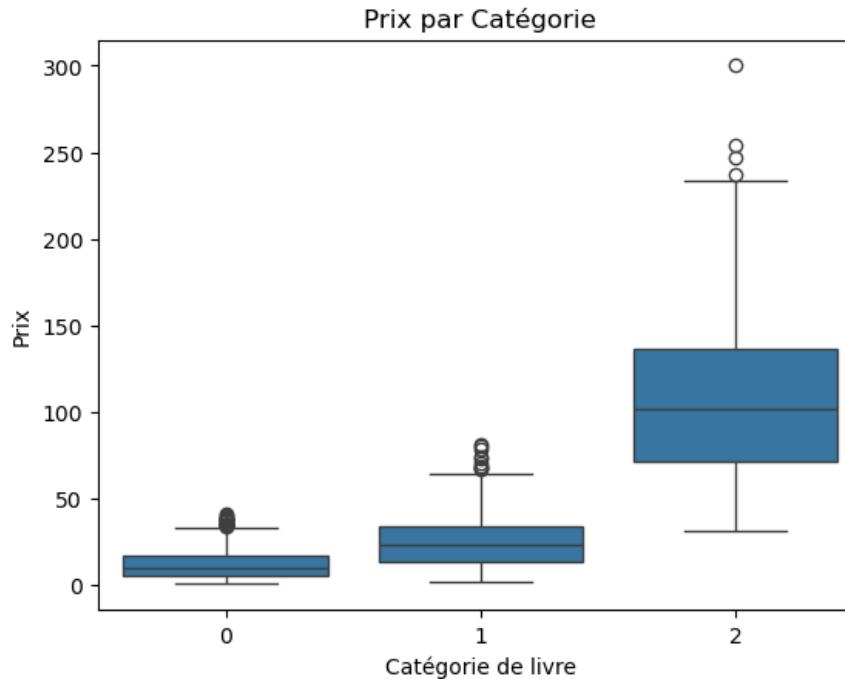
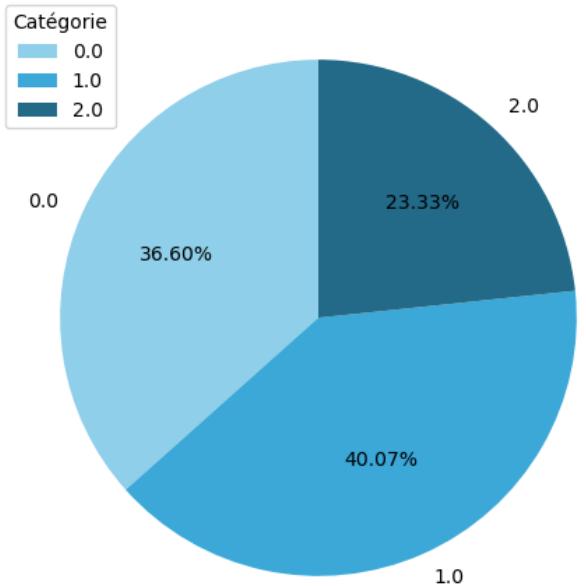


Evolution constante du chiffre d'affaires
Répartition inégale : 50 % des clients ne génèrent que 20 % du chiffre d'affaires.
4 clients atypiques représentent 7,43% du chiffres d'affaires. Indice de Gini

INDICATEURS DE VENTES

- Chiffre d'affaires par catégorie

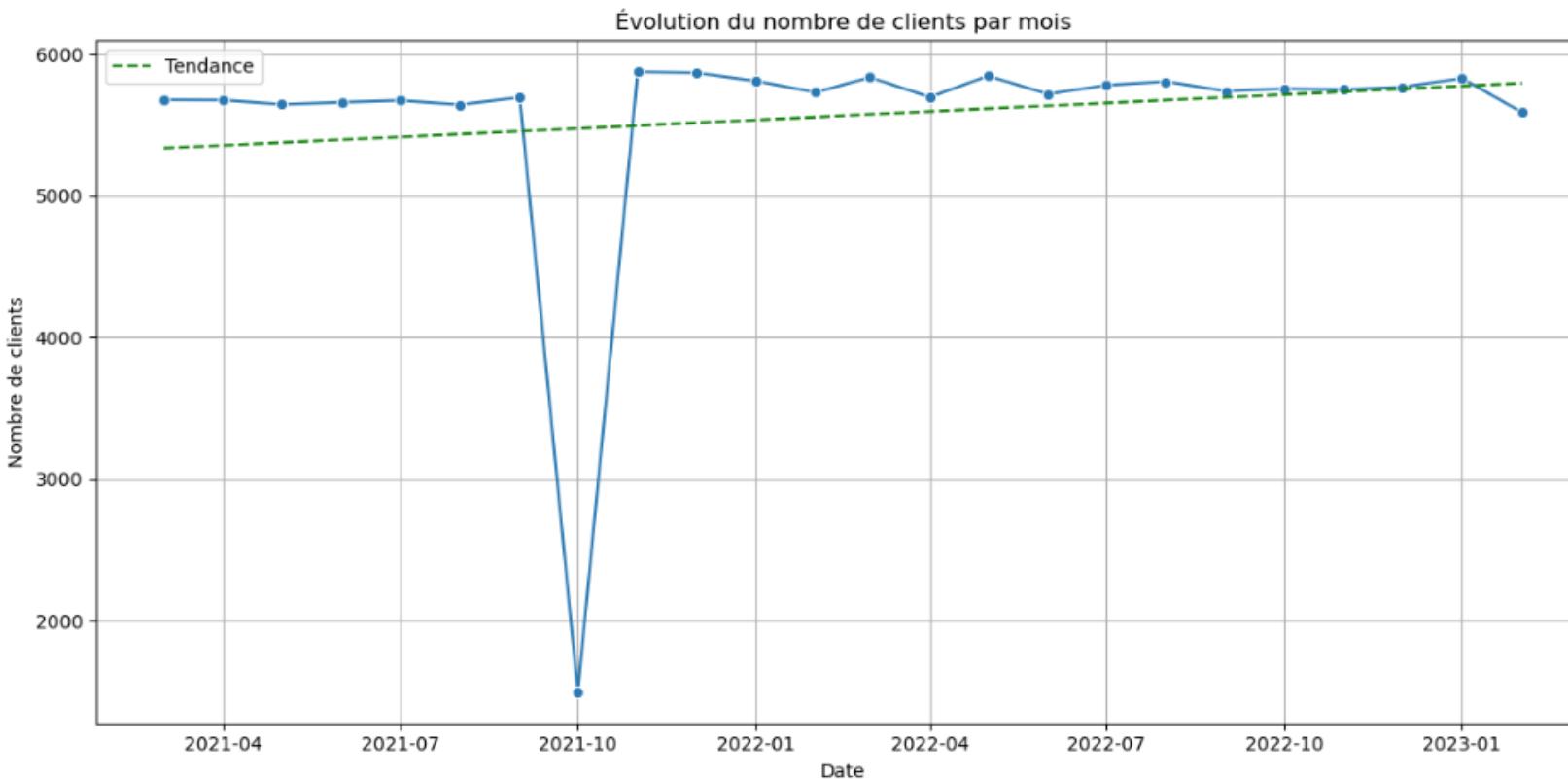
Répartition du chiffre d'affaire par catégorie



60% des ventes de catégorie 0, mais ne représentent que 36% du chiffre d'affaires

INDICATEURS DE VENTES

- Nombre de clients par mois

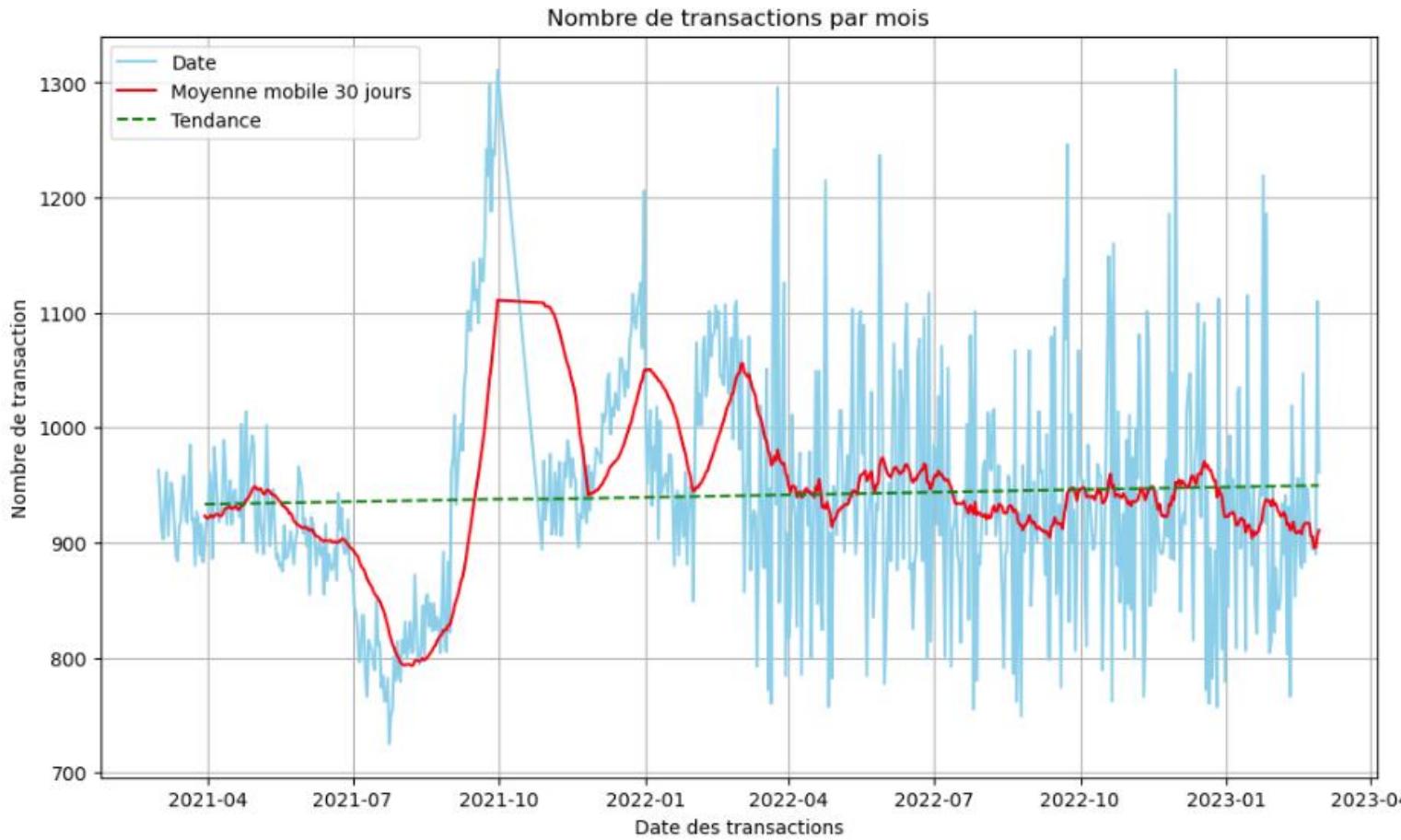


Progression constante du nombre de client par mois, malgré une chute du nombre de client sur le mois d'Octobre 2021.

Cette anomalie est très certainement occasionnée par un bug des valeurs remontés sur octobre

INDICATEURS DE VENTES

- Nombre de transactions



Progression constante au
globale du nombre de
transaction sur les 3 dernières
années

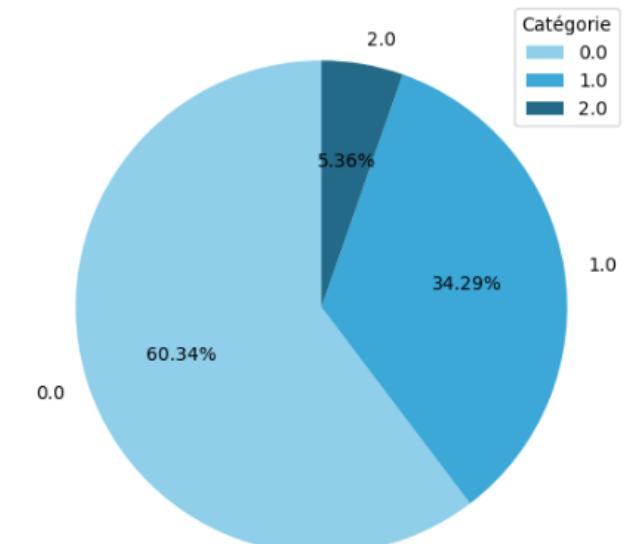
INDICATEURS DE VENTES

- Nombre de produits vendus par mois (TOP / FLOP / Répartition par Catégories)



Product ID	Nb Ventes
0_1016	0
0_1620	0
0_1624	0
0_322	0
1_0	0
0_1025	0
0_310	0
0_299	0
0_1318	0
0_1645	0
1_394	0
0_1062	0
2_72	0
0_2308	0
0_1014	0
0_1800	0
0_510	0
2_87	0
0_1780	0
2_86	0
0_1119	0

Répartition des ventes par catégorie



Le nombre de ventes varie entre 2252 et 0 selon les produits.
60% des produits vendus appartiennent à la catégorie 0
21 produits sans acheteurs

IDENTIFICATION DES TESTS À RÉALISER

Formulation	Type de données	Test	Graphique
Lien entre genre d'un client et les catégories des livres achetés	Association de valeurs qualitative Tableau de contingence	Chi-2	Heatmap
Lien entre âge du client et le montant total des achats	Quantitative / Quantitative	Pearson / Spearman	Scatter plot
Lien entre âge du client et ses fréquences d'achats	Quantitative / Quantitative	Pearson / Spearman	Scatter plot
Lien entre âge du client et la taille du panier moyen	Quantitative / Moyenne quantitative	Spearman	Scatter plot
Lien entre âge du client et les catégories des livres achetés	Quantitative / Qualitative	Anova /Kruskal-Wallis	Boxplot

ANALYSE DES COMPORTEMENTS EN LIGNE : TEST 1

Question : Existe-t-il un lien entre le genre d'un client et les catégories des livres achetés

Hypothèses :

H0 : Il n'existe pas de lien entre le genre d'un client et les catégories de livres achetés

H1 : Il peut exister un lien entre le genre d'un client et les catégories de livres achetés

Choix du test statistique :

Sélection du test de Chi² car il s'agit de valeurs qualitatives (Genre et Catégorie),

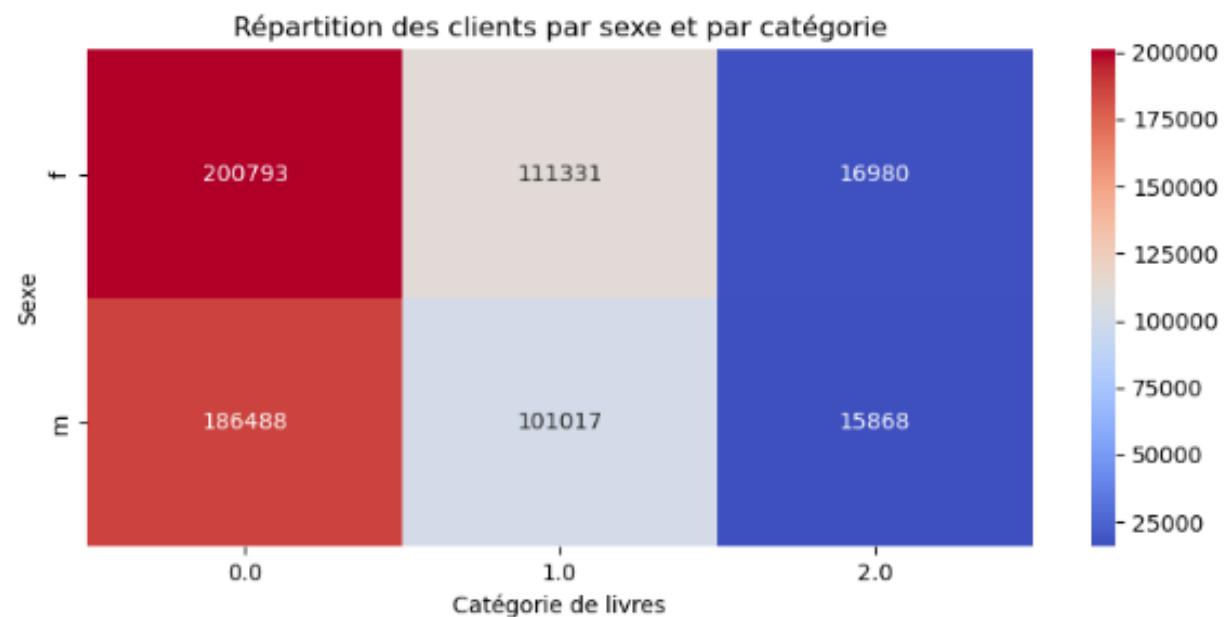
et construction d'un test de contingence de (2 x 3.)

Conclusion :

Le test donne un Chi² de 20,21 et une « p_value » de 0.

L'hypothèse 0 est rejetée, avec un indice de confiance élevé en raison
d'un p_value à 0.

Il existe un lien entre le genre du client et la
catégorie de livres achetée



ANALYSE DES COMPORTEMENTS EN LIGNE : TEST 2

Question : Existe-t-il un lien entre **l'âge du client** et le **montant total des achats**

Hypothèses :

H0 : Il n'existe pas de lien entre l'âge du client et le montant total des achats

H1 : Il peut exister un lien entre l'âge du client et le montant total des achats

Choix du test statistique :

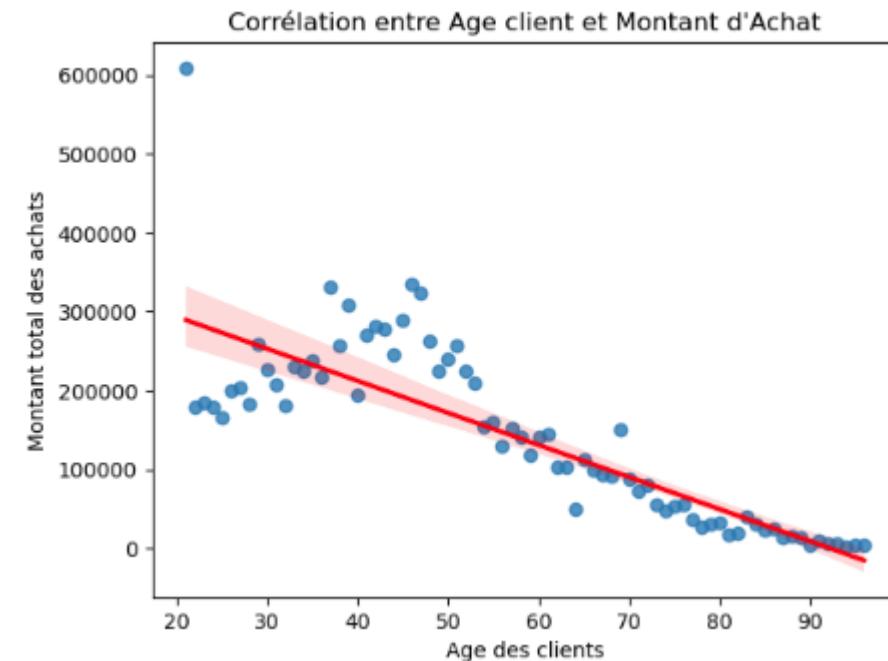
Les deux variables (âge du client et Montant total d'achats) sont quantitatives, les données ne suivent pas une distribution normale et la relation entre les variables est linéaire, ce qui correspond **aux conditions d'application du test de Spearman**.

Conclusion :

Le test de Spearman donne un coefficient négatif et élevé de -0,87 et une « p_value » de 0.

L'hypothèse 0 est rejetée, avec un indice de confiance élevé en raison d'un p_value à 0.

Il existe une forte **corrélation négative** entre l'âge du client et le montant des achats. Comme le montre le graphique, les clients d'un âge avancé consomme moins que leurs cadets



ANALYSE DES COMPORTEMENTS EN LIGNE : TEST 3

Question : Existe-t-il un lien entre l'âge du client et ses fréquences d'achats

Hypothèses :

H0 : Il n'existe pas de lien entre l'âge du client et ses fréquences d'achats

H1 : Il peut exister un lien entre l'âge du client et ses fréquences d'achats

Choix du test statistique :

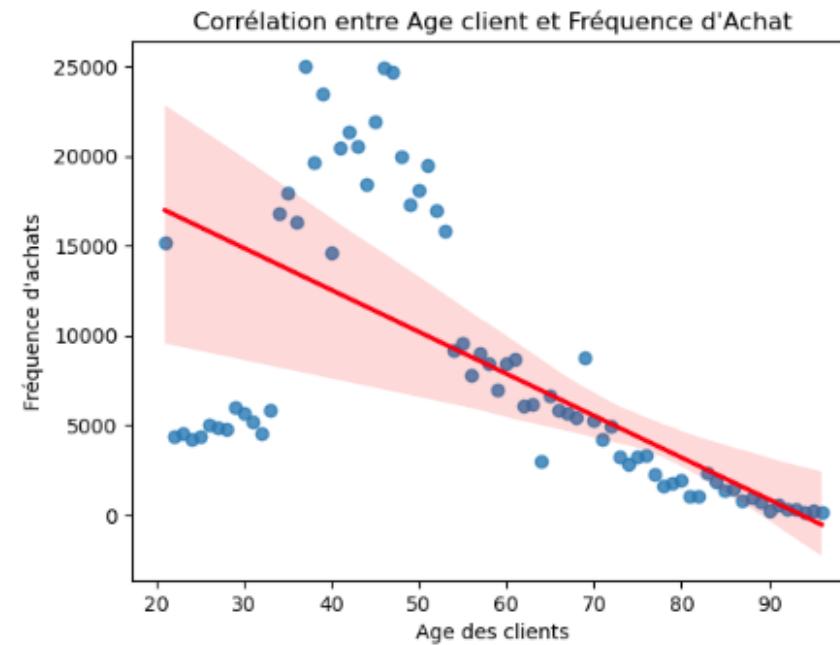
Les deux variables (age du client et fréquences d'achats) sont quantitatives, les données ne suivent pas une distribution normale et la relation entre les variables est linéaire, ce qui correspond aux conditions d'application du test de **Spearman**.

Conclusion :

Le test de Spearman donne un coefficient négatif de -0,67 et une « p_value » de 0.

L'hypothèse **0 est rejetée**, avec un indice de confiance élevé en raison d'un p_value à 0.

Il existe bien une **corrélation négative** entre l'âge du client et la fréquence d'achats



ANALYSE DES COMPORTEMENTS EN LIGNE : TEST 4

Question : Existe-t-il un lien entre l'âge du client et la taille du panier moyen

Hypothèses :

H0 : Il n'existe pas de lien entre l'âge du client et la taille du panier moyen

H1 : Il peut exister un lien entre l'âge du client et la taille du panier moyen

Choix du test statistique :

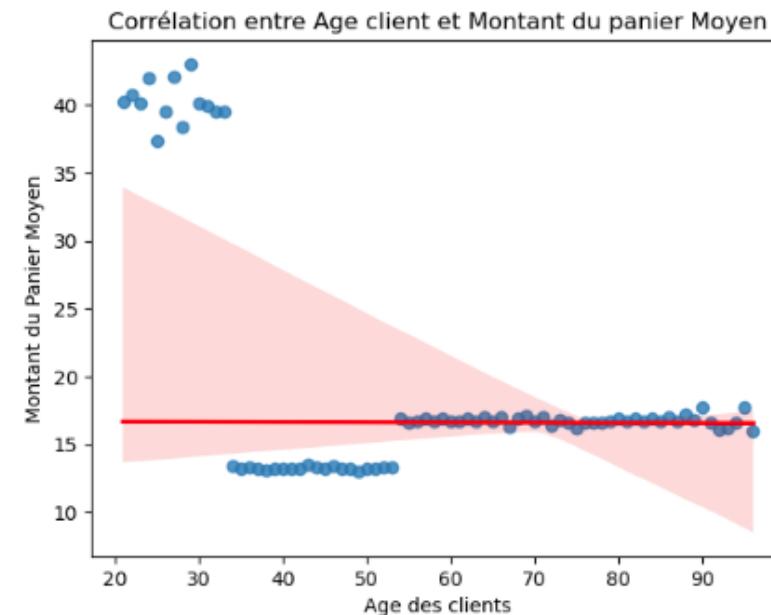
Les deux variables (âge du client et taille du panier moyens sont quantitatives, les données ne suivent pas une distribution normale et la relation entre les variables est linéaire, ce qui correspond aux conditions d'application du test de **Spearman**.

Conclusion :

Le test de Spearman donne un coefficient négatif de -0,09 et une « p_value » de 0,47.

L'hypothèse **0** ne peut pas être rejetée.

Il n'existe pas de lien statistique entre l'âge du client et la taille du panier moyen



ANALYSE DES COMPORTEMENTS EN LIGNE : TEST 5

Question : Existe-t-il un lien entre l'âge du client et les catégories des livres achetés

Hypothèses :

H0 : Il n'existe pas de lien entre l'âge du client et les catégories des livres achetés

H1 : Il peut exister un lien entre l'âge du client et les catégories des livres achetés

Choix du test statistique :

Nous avons deux variables (âge du client) quantitative et la catégorie de livre qui est qualitative. De plus, nous 3 groupes à comparer. Le test Anova est donc parfaitement adapté au test à réaliser.

Conclusion :

Le test de ANOVA donne un coefficient de 39266,29 et une « p_value » de 0.

L'hypothèse nulle est rejetée, il y a donc bien un lien entre âge du client et la catégorie de livre.

Il existe un lien entre l'âge du client et la catégorie de livre acheté.

Les 20/30 ans vont préférés des livres de catégorie 2, les 40/50 ans des livres de catégorie 0 alors que les autres préféreront des livres de catégorie 1.

