



Lapage

Etude sur les  
ventes de  
notre librairie  
en ligne

Limites & Conclusions

6

Analyses complémentaires

5

Corrélations

4

1

Contexte

2

Méthodologie

3

Analyses globales



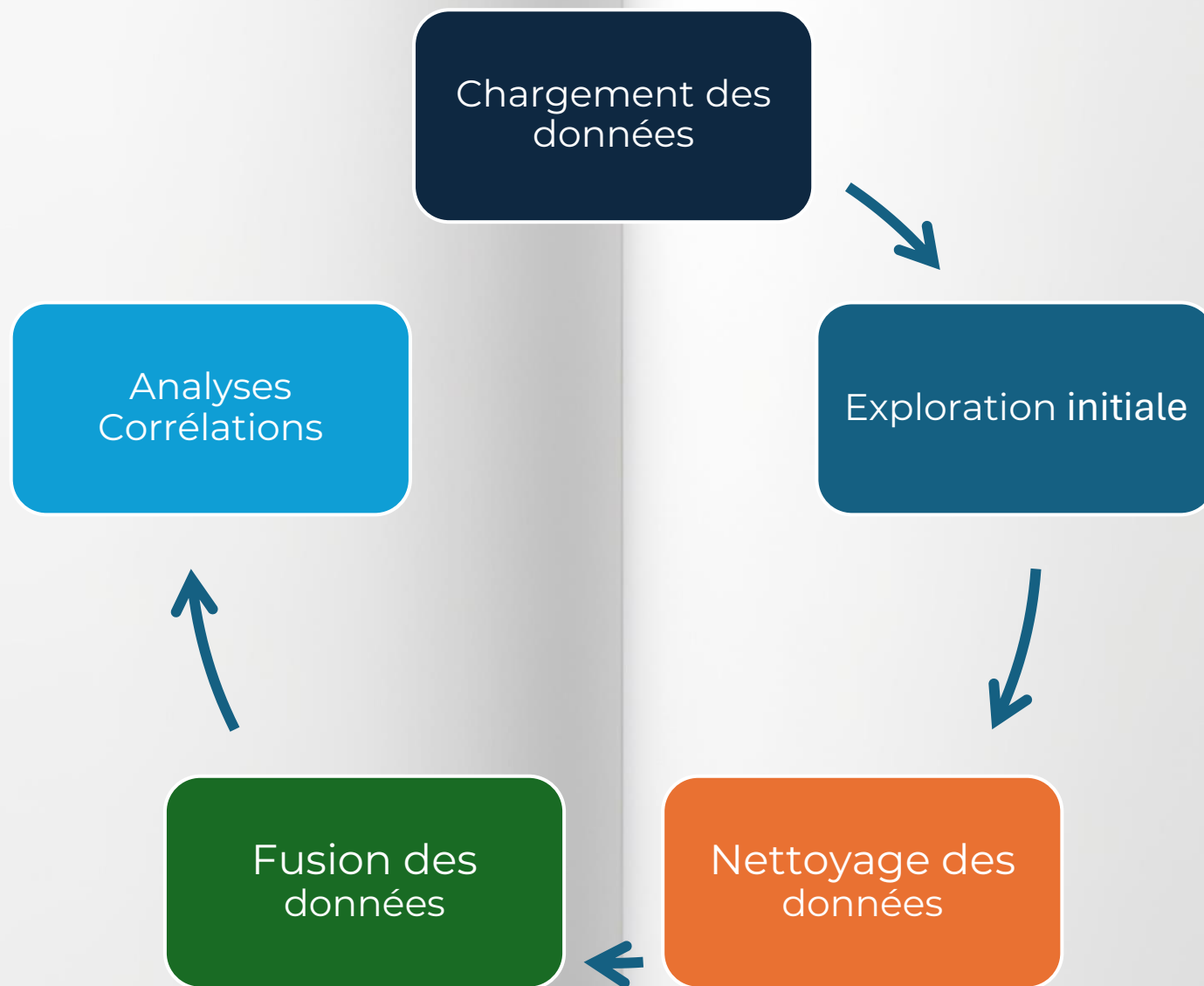
Lapage

### Objectifs :

- Appui à la prise de décisions stratégiques en matière d'offres, de prix produits, de type de client etc...
  - Analyse des différents indicateurs de vente
  - Analyse des comportements clients
  - Analyse de l'influence de l'âge des clients

### Sources :

- Données clients (id, genre, année de naissance)
- Données produits (id, prix, catégorie)
- Données transactions (id, date, id\_client, id\_produit)



Chargement des  
données

Exploration initiale

Nettoyage des  
données

- customers.csv
- products.csv
- transactions.csv

Notebook



- 8623 clients uniques H/F nés entre 1929 et 2004
- 3287 produits uniques divisés en 3 cat. avec leur prix
- 679532 transactions uniques entre le 01/03/2021 et le 28/02/2023

- 1 produit Test supprimé
- 200 transactions Test supprimées
- 2 clients Test supprimés

Fusion des  
données

Analyses  
Corrélations

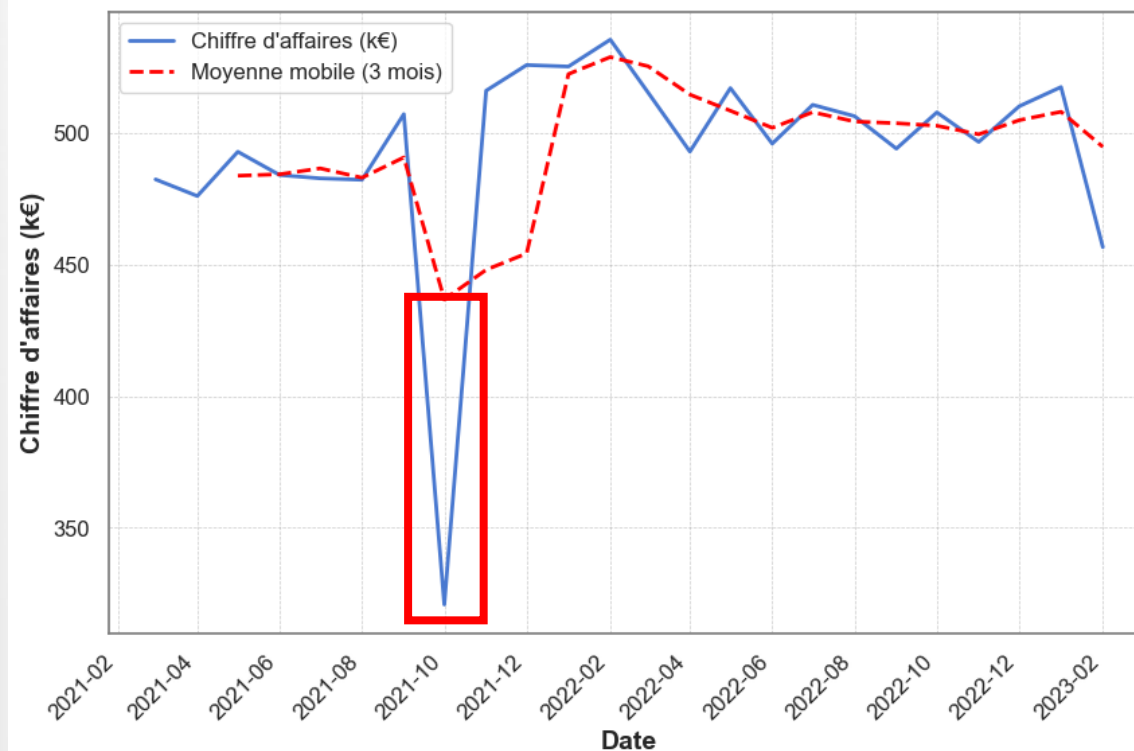
- 1 transactions avec customers (21 clients inactifs)
- 1 fusionné avec products (242 produits pas achetés)
- 356 outliers possibles (conservés)

- Chiffre d'affaires
- Clients
- Produits
- Transactions
- Comportement clients

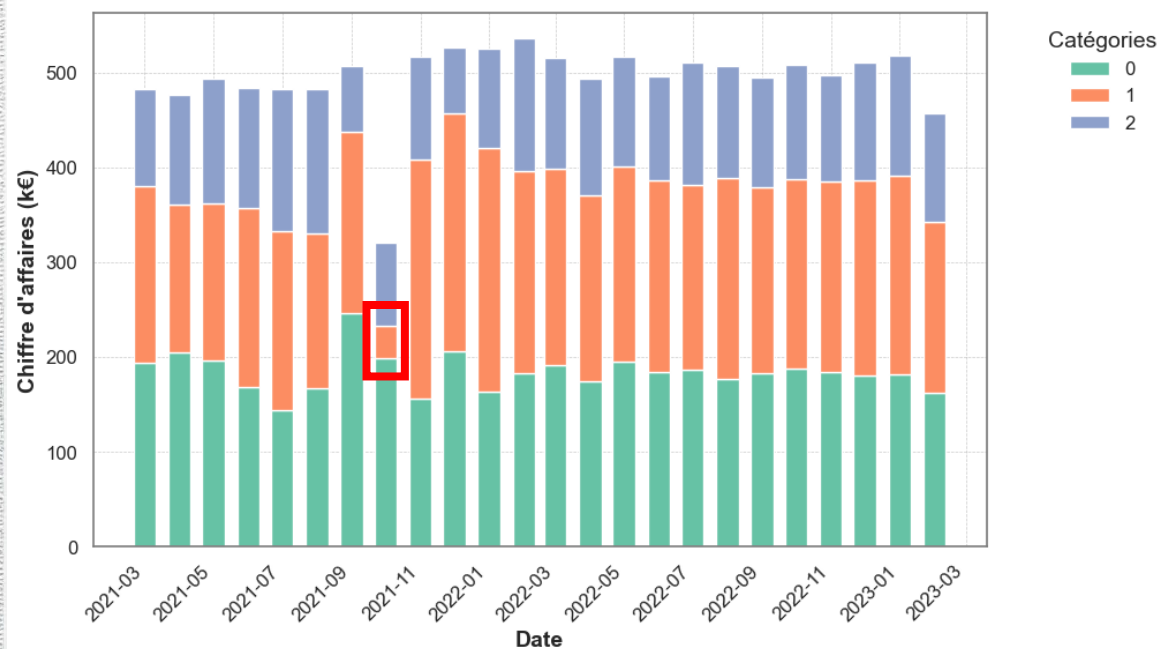


Année	Chiffre d'affaires (k€)
2021	4770.83
2022	6108.68
2023	974.22
Total	11853.73

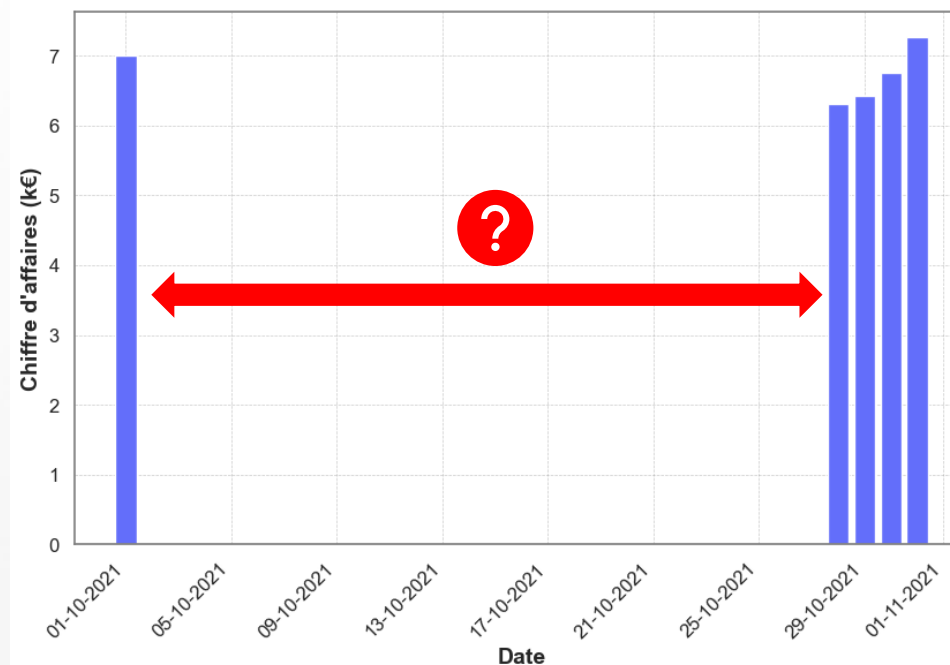
Évolution mensuelle du chiffre d'affaires avec moyenne mobile (3 mois)



Évolution mensuelle du chiffre d'affaires par catégorie

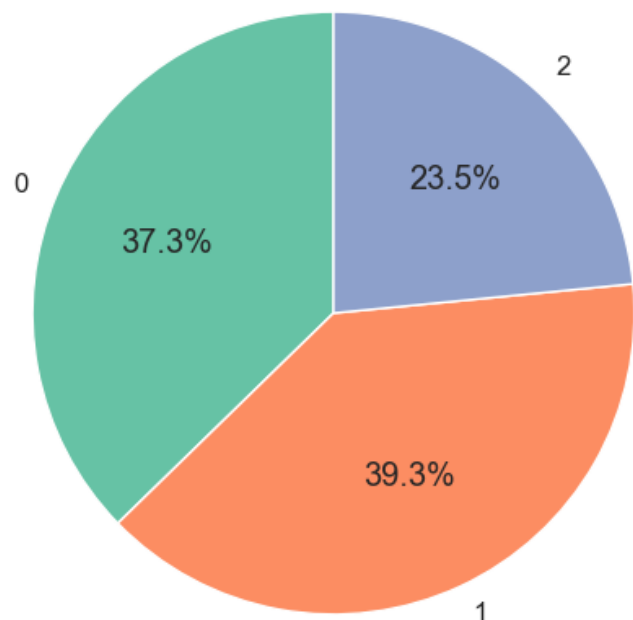


Évolution quotidienne 10/2021 du chiffre d'affaires - Catégorie 1



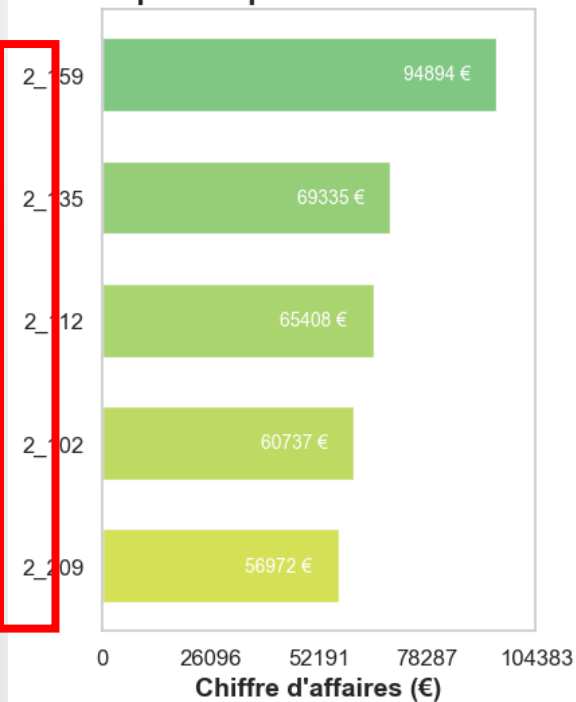


Répartition du chiffre d'affaires  
par catégorie de produits

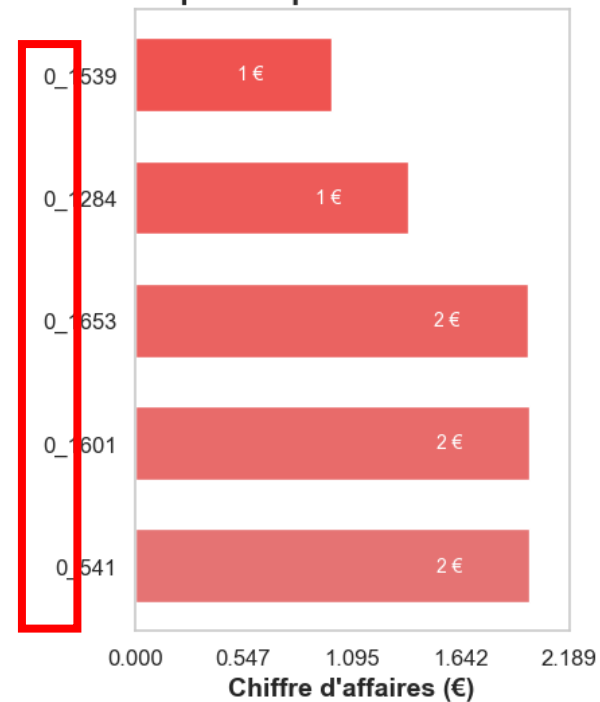


Top et Flop des produits par chiffre d'affaires

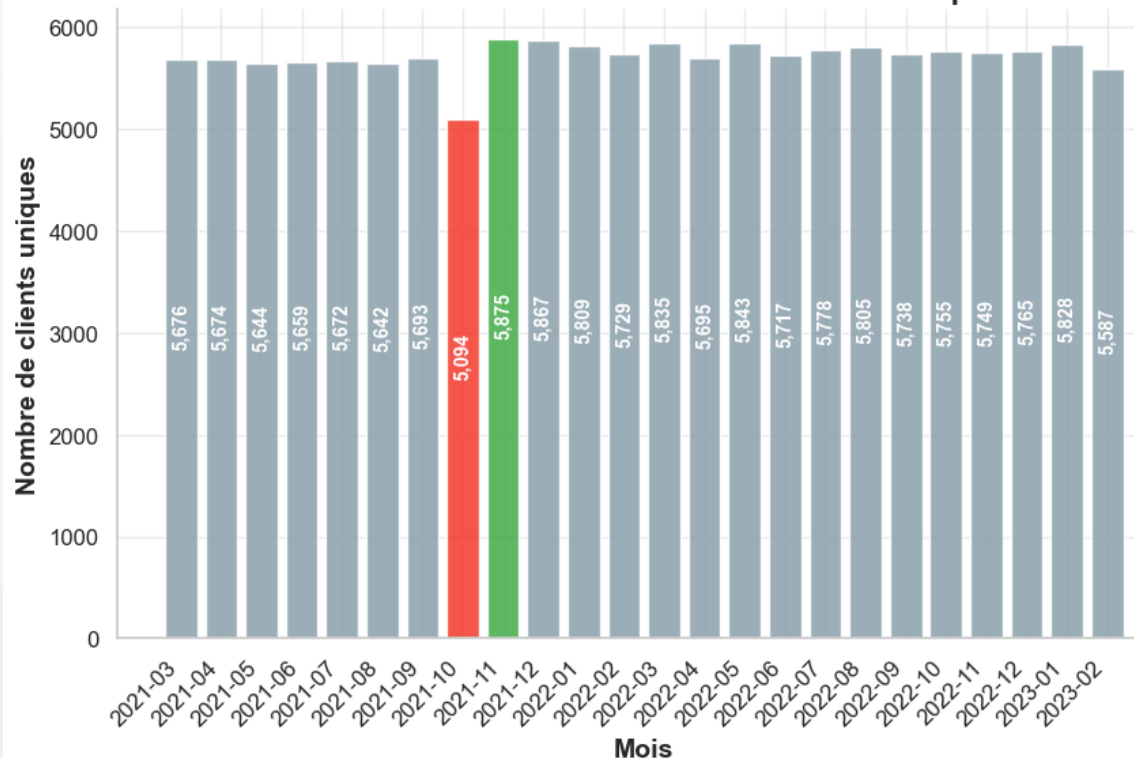
Top 5 des produits



Flop 5 des produits



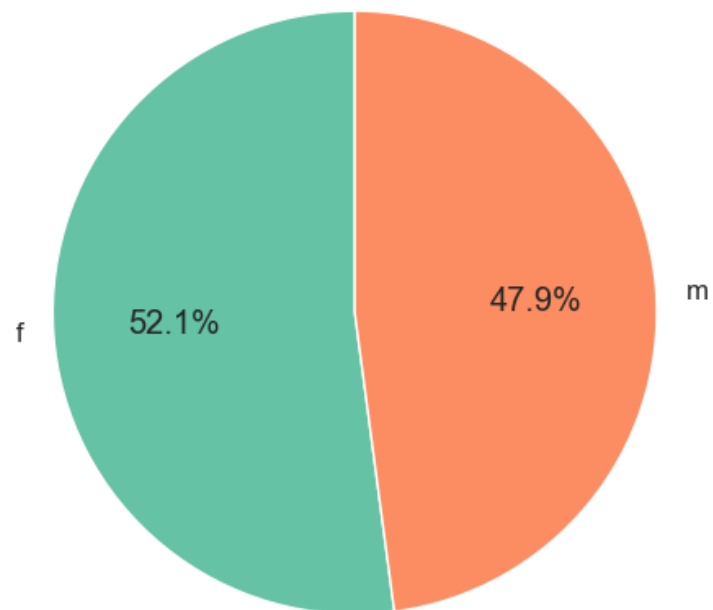
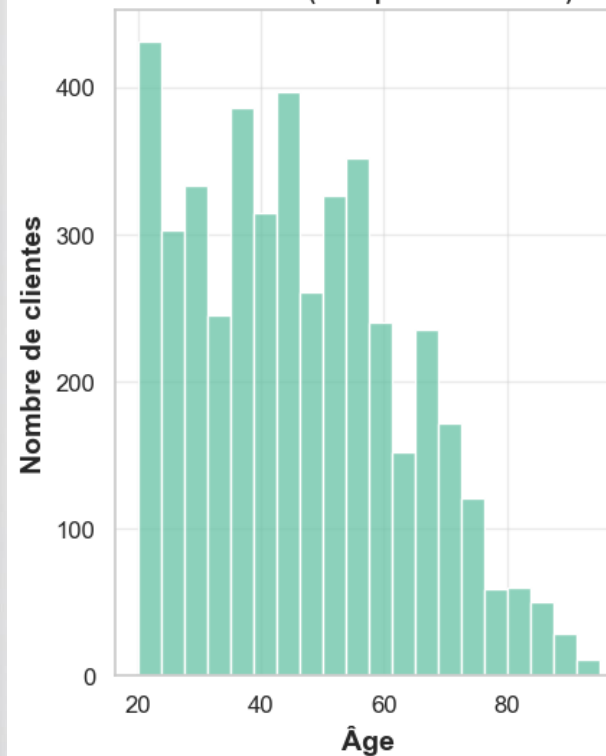
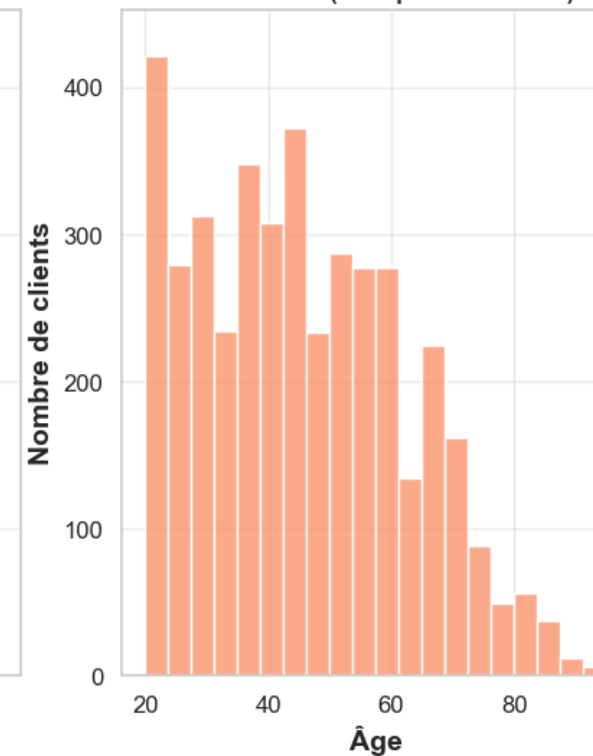
Évolution mensuelle du nombre de clients uniques



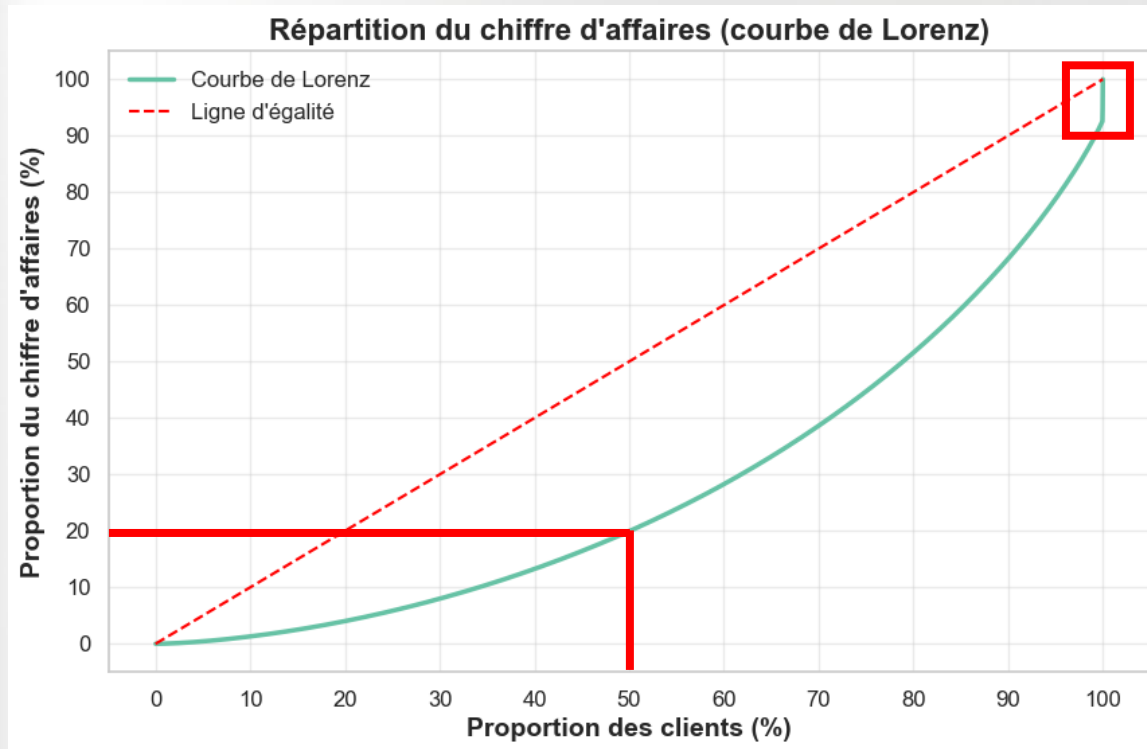
client_id	Chiffre d'affaires (€)	
677	c_1609	324033.35
4388	c_4958	289760.34
6337	c_6714	153598.92
2724	c_3454	113637.93
2513	c_3263	5276.87
634	c_1570	5271.62
2108	c_2899	5214.05
1268	c_2140	5208.82
7006	c_7319	5155.77
7715	c_7959	5085.45

	client_id	Chiffre d'affaires (€)
677	c_1609	324033.35
4388	c_4958	289760.34
6337	c_6714	153598.92
2724	c_3454	113637.93
2513	c_3263	5276.87
634	c_1570	5271.62
2108	c_2899	5214.05
1268	c_2140	5208.82
7006	c_7319	5155.77
7715	c_7959	5085.45

Répartition des clients (hors pro.) par genre

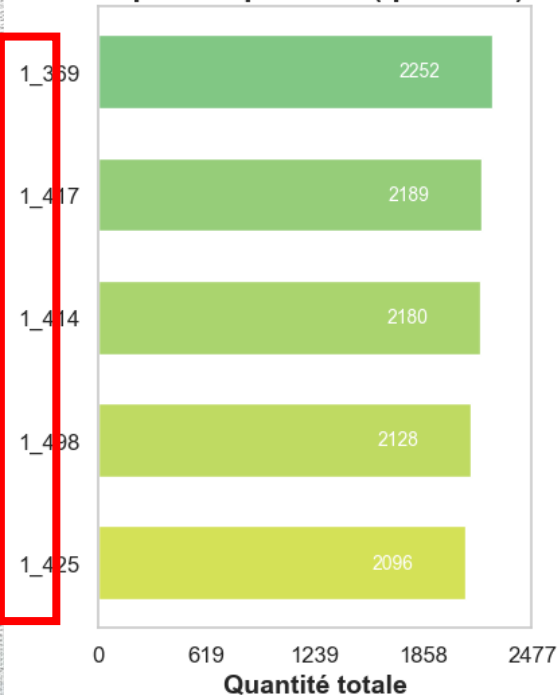
Distribution des âges  
des femmes (hors professionnelles)Distribution des âges  
des hommes (hors professionnels)

gender	f	m
count	4478.000000	4118.000000
mean	46.025011	45.429335
std	17.092603	16.706855
min	20.000000	20.000000
25%	32.000000	32.000000
50%	45.000000	44.000000
75%	58.000000	58.000000
max	95.000000	95.000000



## Top et Flop des produits par quantités totales

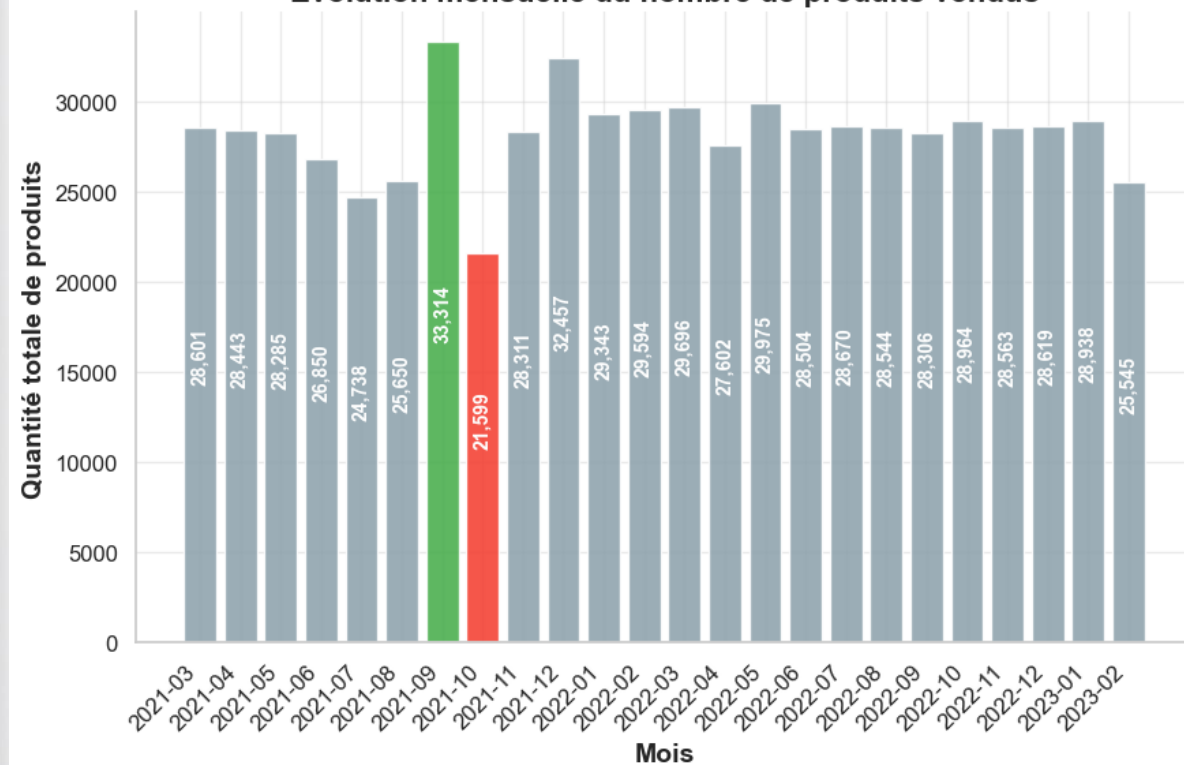
Top 5 des produits (quantités)



Flop 5 des produits (quantités)

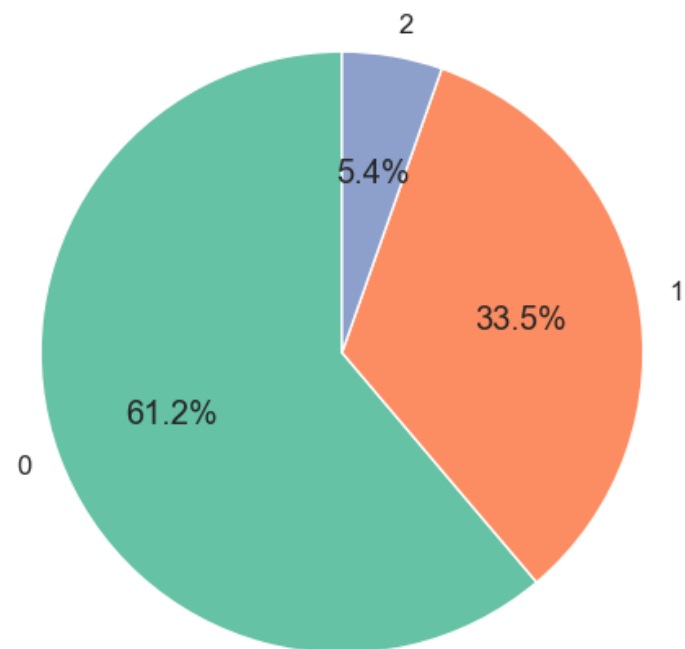


## Évolution mensuelle du nombre de produits vendus



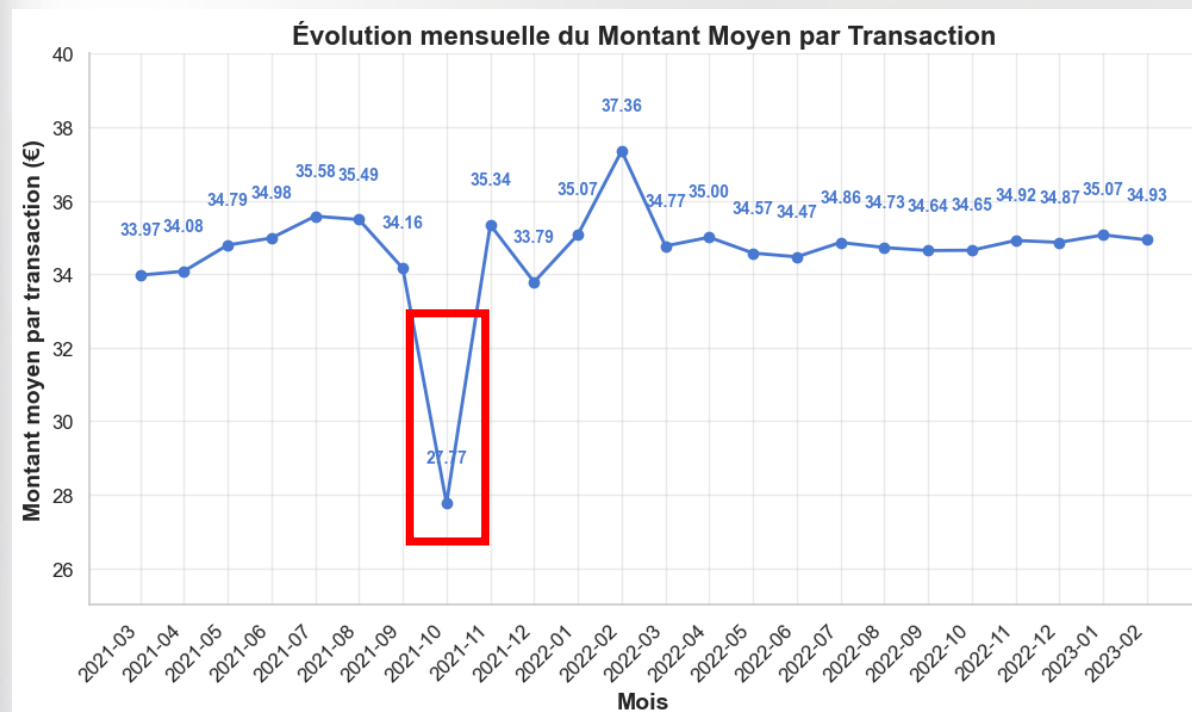
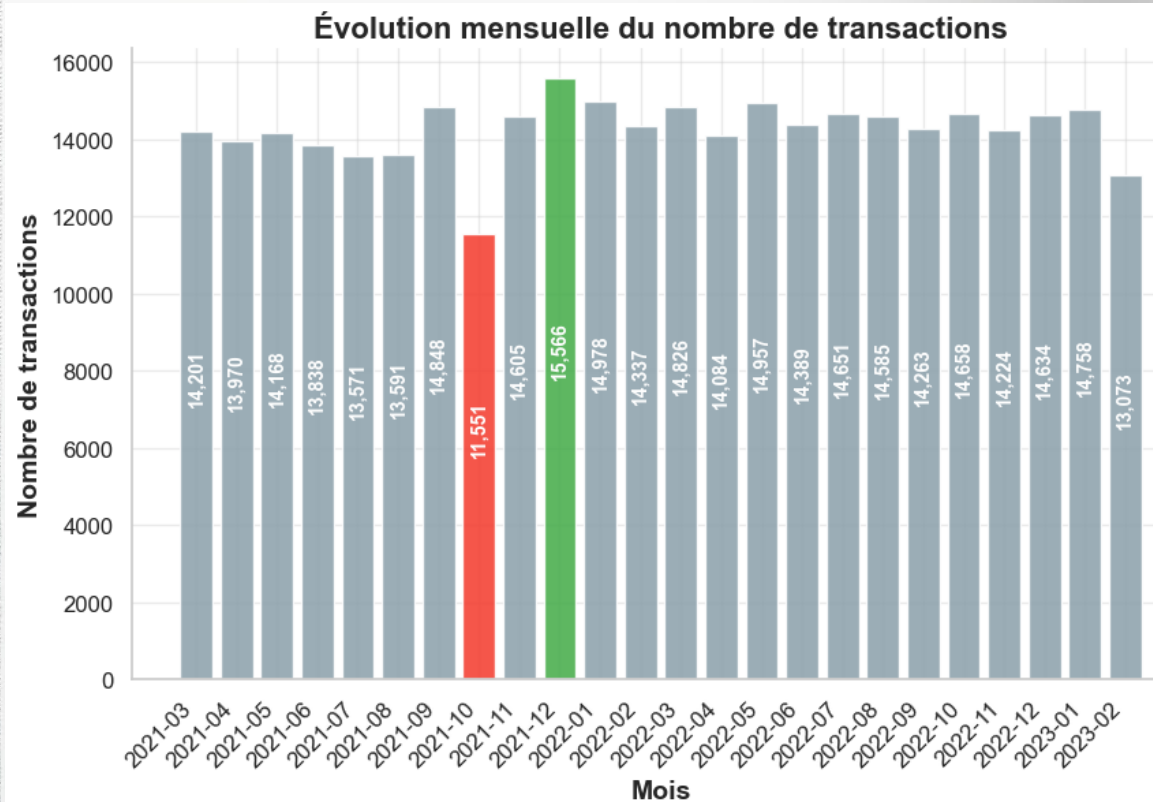


Répartition des quantités de produits  
vendus par catégorie

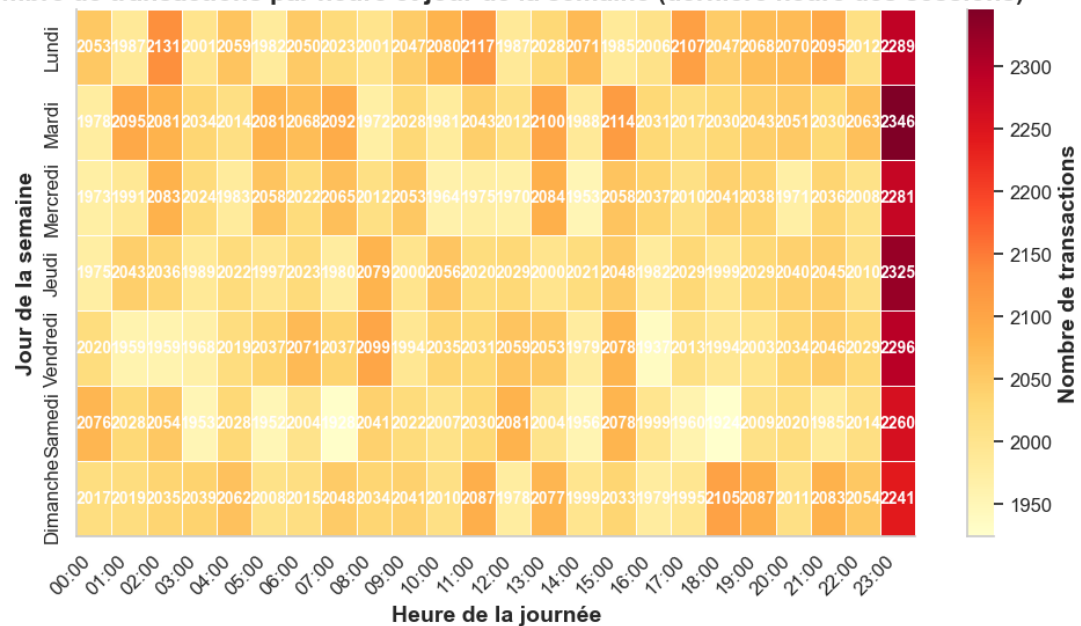


Nombre de produits par catégorie  
category

0	2292
1	737
2	236
Total	3265



Nombre de transactions par heure et jour de la semaine (dernière heure des sessions)



Chi-2 :

teste l'indépendance entre 2 variables qualitatives (catégorielles). Les conditions d'application sont variées :

- Les données doivent être qualitatives
- Les observations doivent être indépendantes
- Les échantillons doivent être suffisamment grands
- Il est utilisé principalement quand les données sont sous forme de tableaux de contingence (2x3, 3x3, etc...)

Kolmogorov-Smirnov :

teste si les données suivent une distribution normale pour des échantillons > 2000

Spearman :

teste s'il existe une corrélation entre 2 variables qualitatives et lorsque la distribution ne suit pas une loi normale

Levene :

teste si les variances entre différentes variables sont égales et lorsque les distributions ne suivent pas une loi normale

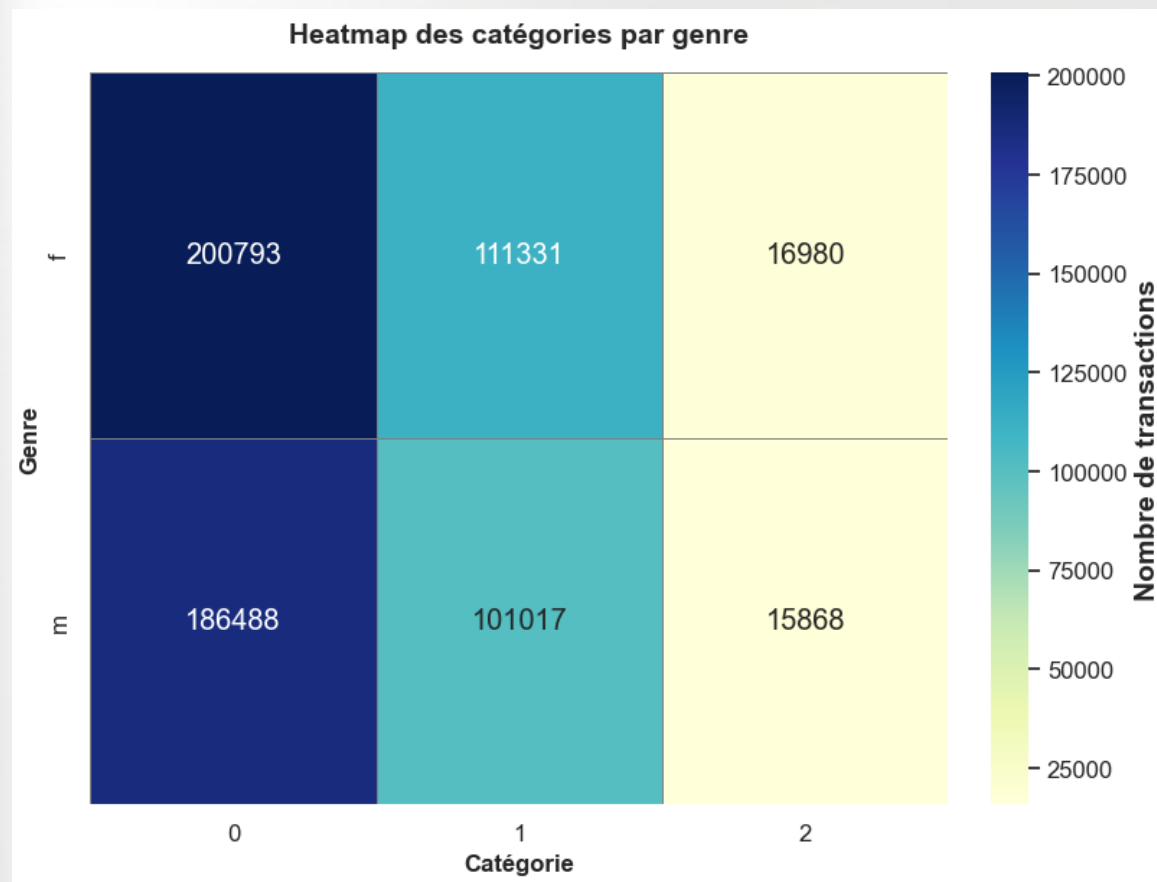
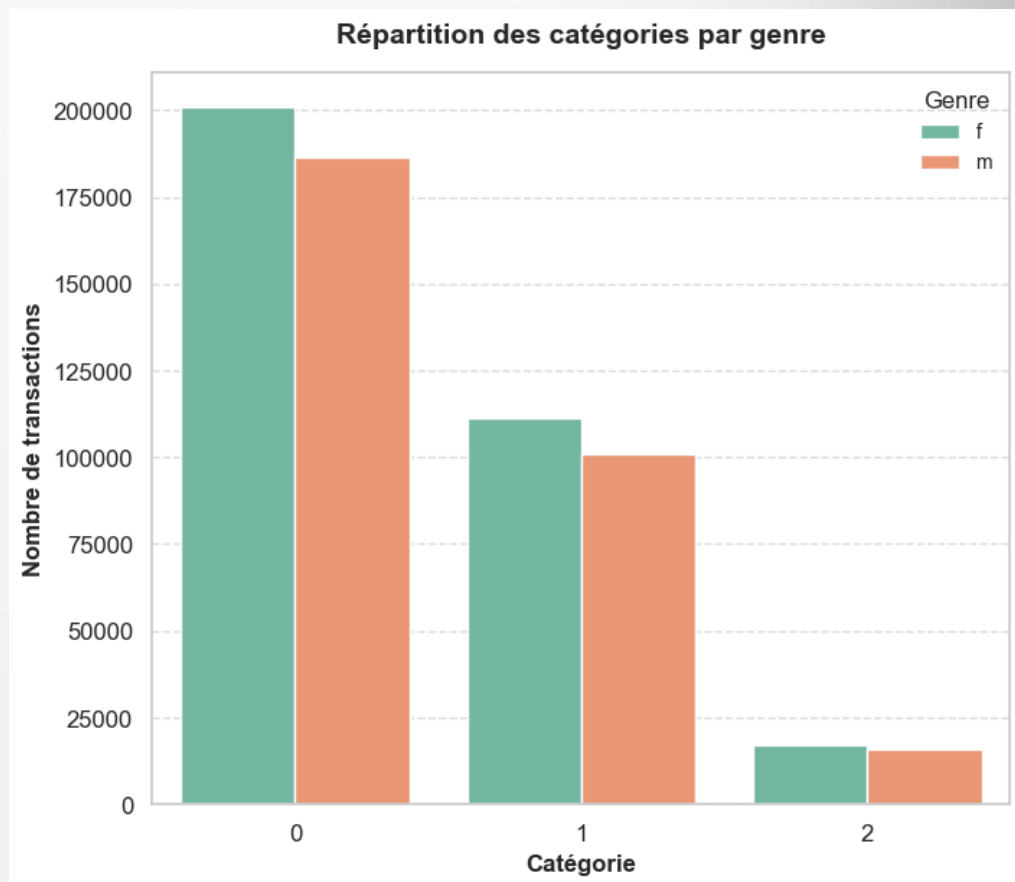
Kruskal-Wallis :

teste s'il existe une relation significative entre 1 variable quantitative et 1 variable qualitative. Les conditions d'application sont variées :

- Les observations doivent être indépendantes
- Elles ne respectent pas une distribution normale
- Les variances ne sont pas égales

ANOVA :

idem que le Kruskal-Wallis mais à condition que les variables respectent une distribution normale et que les variances soient égales





Chi2-statistique: 20.21346069

Valeur p: 0.00004080

Degrés de liberté: 2

Fréquences attendues :

[[201518.35754344 110493.46646914 17092.17598743]

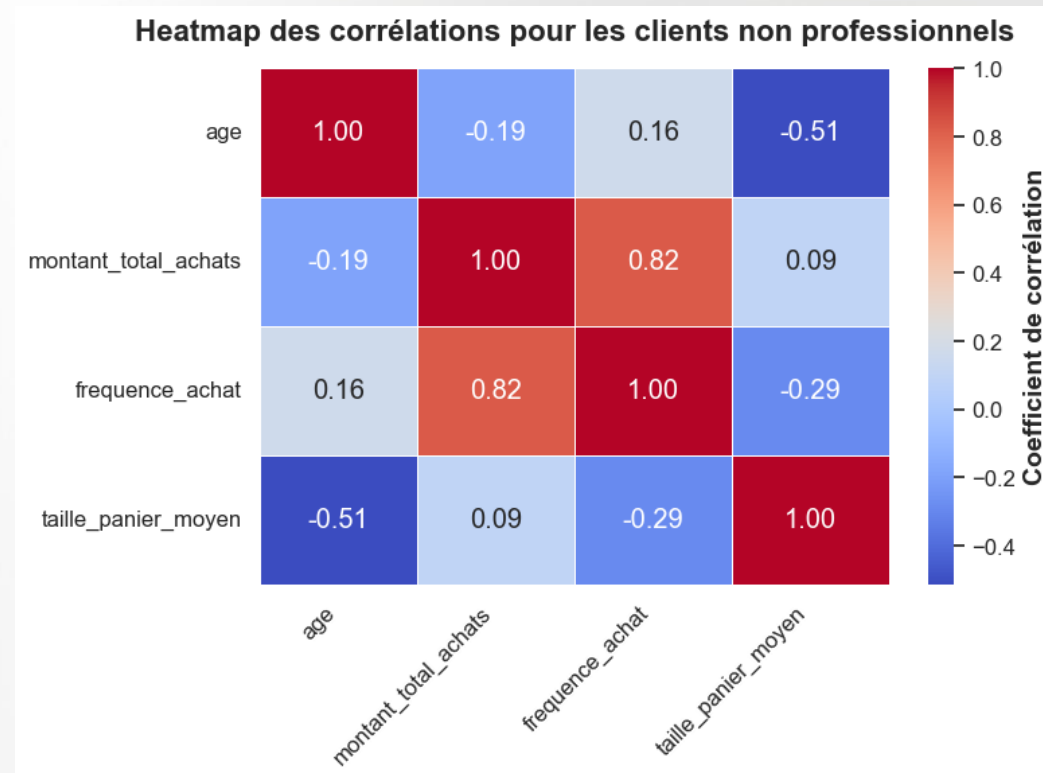
[185762.64245656 101854.53353086 15755.82401257]]

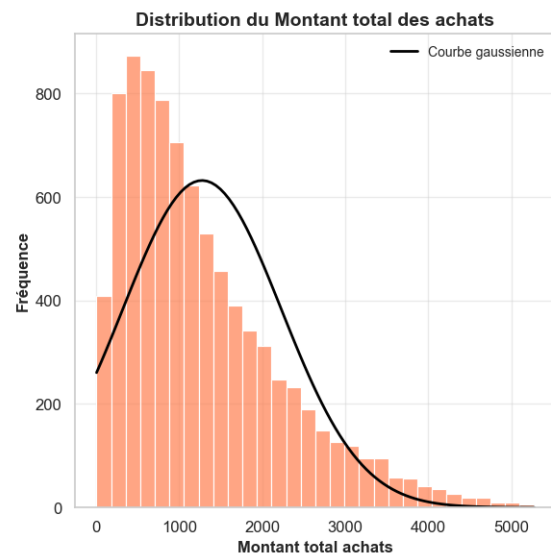
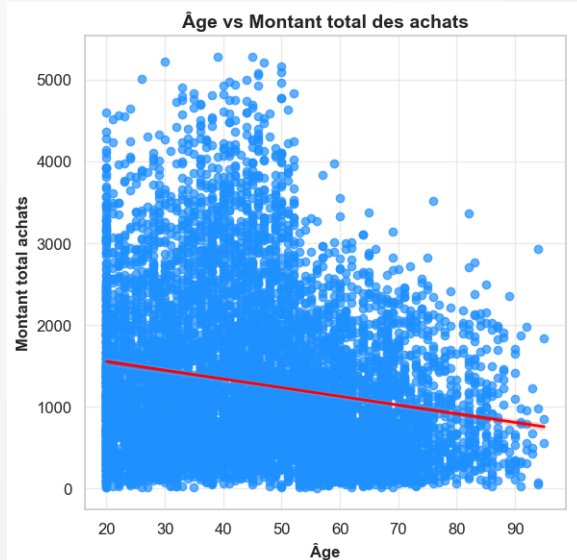
H1 : la catégorie de livres achetés dépend du genre du client,  
il y a une corrélation

## Genre d'un client et catégories des livres achetés

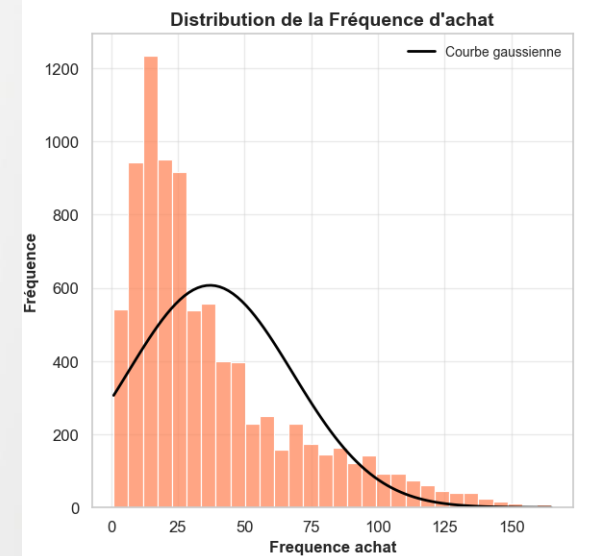
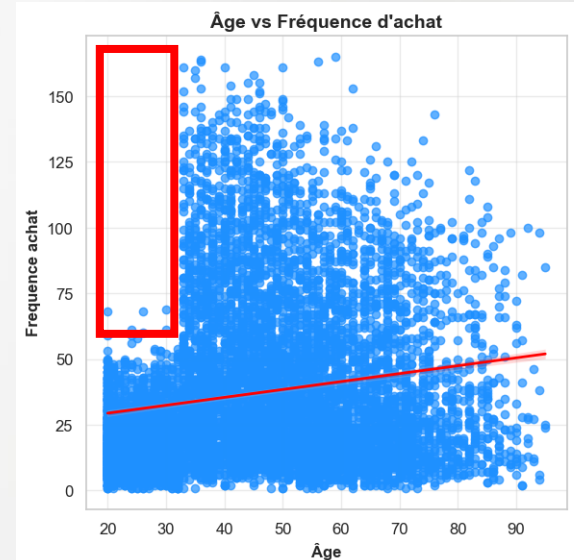
	client_id	age	montant_total_achats	frequence_achat	taille_panier_moyen
0	c_1	69.0	558.18	33	14.312308
1	c_10	68.0	1353.60	34	23.337931
2	c_100	32.0	254.85	5	31.856250
3	c_1000	58.0	2261.89	93	18.095120
4	c_1001	42.0	1812.86	47	17.773137

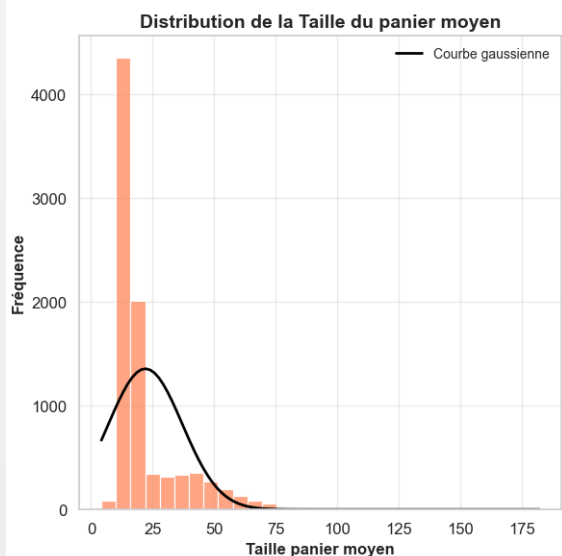
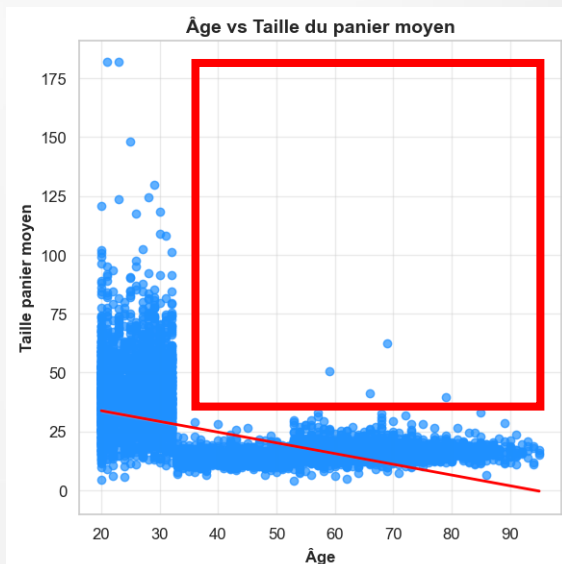
Âge des clients et Montant total des achats,  
 Âge des clients et Fréquence d'achat,  
 Âge des clients et Taille du panier moyen





Âge des clients et Montant total des achats,  
 Âge des clients et Fréquence d'achat,  
 Âge des clients et Taille du panier moyen





## Âge des clients et Montant total des achats, Âge des clients et Fréquence d'achat, Âge des clients et Taille du panier moyen

Résultats du test de Kolmogorov-Smirnov :

Variable : montant\_total\_achats

- Statistique KS : 0.1064

- p-valeur : 3.0398e-85

-> La distribution n'est PAS normale (rejette l'hypothèse nulle).

Variable : frequence\_achat

- Statistique KS : 0.1528

- p-valeur : 1.3515e-175

-> La distribution n'est PAS normale (rejette l'hypothèse nulle).

Variable : taille\_panier\_moyen

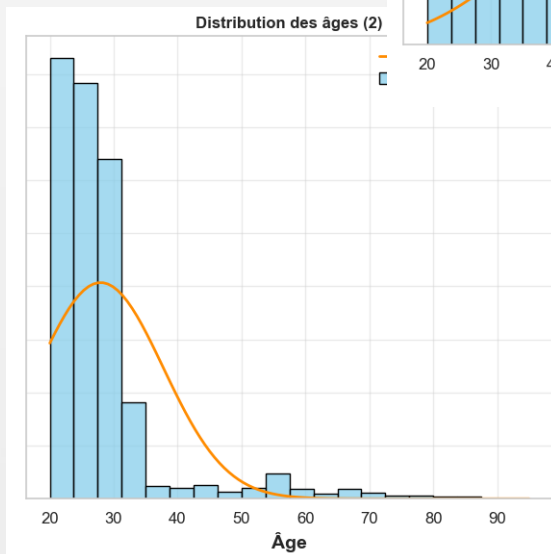
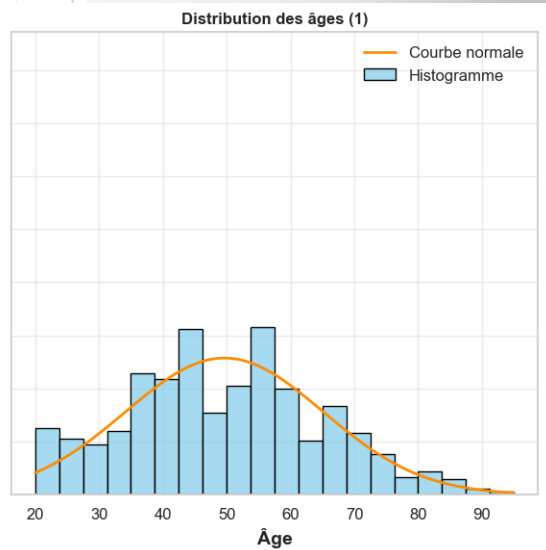
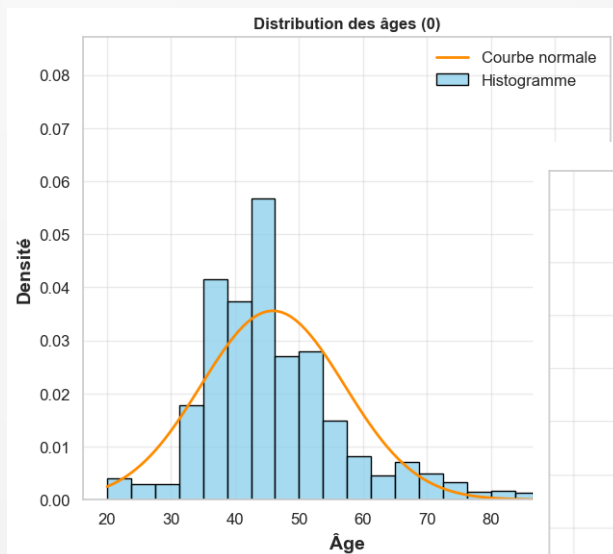
- Statistique KS : 0.2782

- p-valeur : 0.0000e+00

-> La distribution n'est PAS normale (rejette l'hypothèse nulle).

Résultats du test de Spearman :

	Variable	Correlation	P-value	Significatif
0	montant_total_achats	-0.185139	3.788575e-67	Oui
1	frequence_achat	0.212297	3.504908e-88	Oui
2	taille_panier_moyen	-0.326166	3.280394e-212	Oui



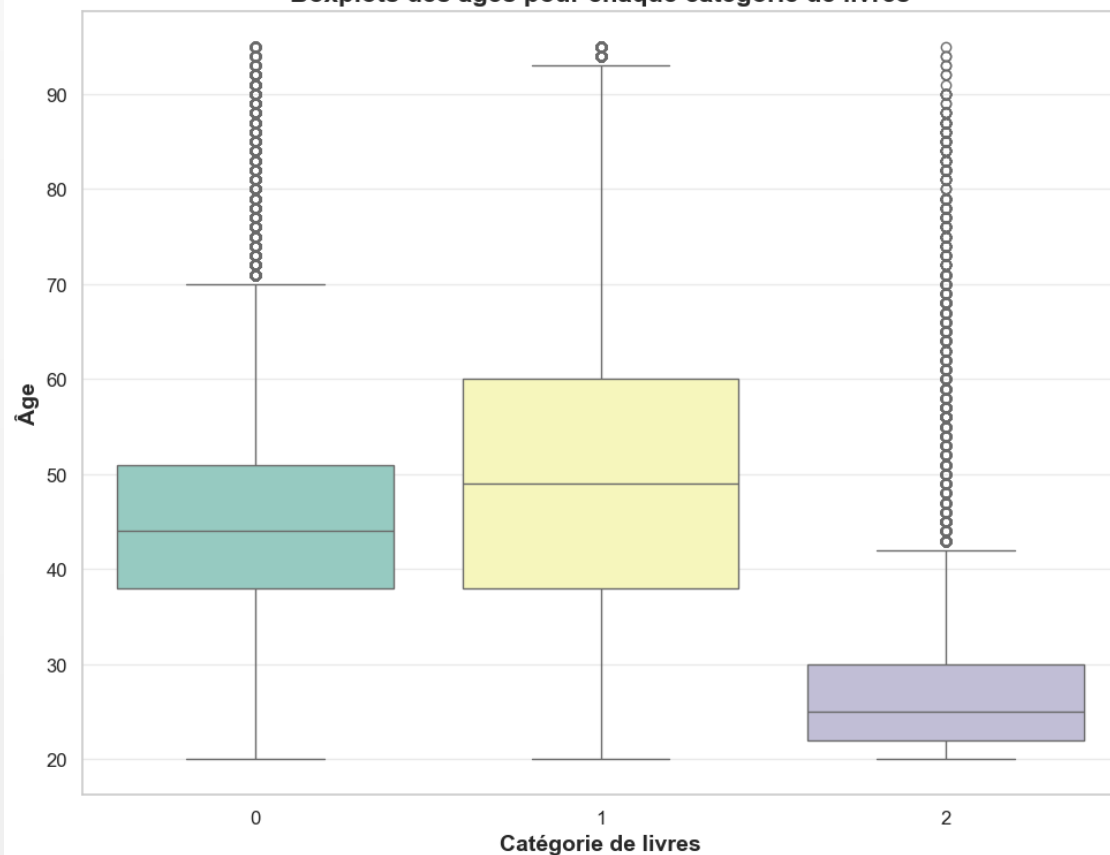
## Âge des clients et Catégories des livres achetés

Résultats du test de Kolmogorov-Smirnov :

Category	KS Statistic	p-value	Normality
0	0.110027	0.0	Reject H0 (Non-Normal)
1	0.044560	0.0	Reject H0 (Non-Normal)
2	0.249121	0.0	Reject H0 (Non-Normal)



Boxplots des âges pour chaque catégorie de livres



## Âge des clients et Catégories des livres achetés

Statistique de Levene : 26085.920

Valeur p : 0.000

Les variances ne sont pas égales (rejeter l'hypothèse nulle).

Nous savons maintenant que :

- Les variables sont indépendantes
- Elles ne respectent pas une distribution normale
- Les variances ne sont pas égales

## Âge des clients et Catégories des livres achetés

Statistique de Kruskal-Wallis : 79346.099

Valeur p : 0.000

Les âges varient en fonction de la catégorie des livres (rejeter l'hypothèse nulle).

Statistique ANOVA : 45879.682

Valeur p : 0.000

Les âges varient en fonction de la catégorie des livres (rejeter l'hypothèse nulle).

### Segmentation des clients basés sur le comportement d'achat :

- But : Identifier des groupes de clients avec des comportements similaires
- Résultat attendu : Identifier des profils clients comme gros acheteurs fréquents, acheteurs occasionnels, etc...

### Analyse de la fidélité des clients :

- But : Etudier si les clients reviennent souvent et pour quelles catégories
- Résultat attendu : Identifier les catégories qui favorisent la fidélité

### Prévisions des ventes :

- But : Prévoir les ventes futures en fonction des tendances passées
- Résultat attendu : Identifier les périodes potentielles de forte demande

### Identification des clients à haut potentiel :

- But : Détecter les clients les plus susceptibles de générer des revenus élevés
- Résultat attendu : Cibler les clients à fort potentiel pour des campagnes marketing

### Impact des remises sur les ventes :

- But : Analyser comment les promotions influencent les ventes
- Résultat attendu : Mesurer l'efficacité des promotions par catégorie

Limites de l'étude :

- Durée des données
- Valeurs manquantes
- Outliers

Implications des limites :

- Fiabilité et validité des conclusions

Conclusions :

- Données manquantes 10/2021
- Tendance à la baisse du chiffre d'affaires
- 4 clients professionnels
- Influence de l'âge des clients
- Alignement des comparaisons avec librairies physiques
- Analyses complémentaires