



Rester Livre

# Les Données

	client_id	sex	birth	id_prod	date	session_id	price	categ
0	c_4410	f	1967	0_1455	2021-03-22 14:29:25.189266	s_9942	8.99	0
1	c_4389	m	1984	0_1455	2021-07-09 11:16:18.579726	s_59967	8.99	0
2	c_5019	f	1977	0_1455	2022-01-15 00:01:53.456196	s_149928	8.99	0
3	c_7049	f	1987	0_1455	2021-03-04 14:01:38.698752	s_1637	8.99	0
4	c_5110	f	1982	0_1455	2021-09-05 11:48:41.065009	s_85364	8.99	0
...	...	...	...	...	...	...	...	...
336908	c_7135	m	1996	2_99	2021-07-11 20:56:49.820935	s_61009	84.99	2
336909	c_5828	f	1998	2_99	2021-11-21 01:53:46.967570	s_122697	84.99	2
336910	c_8260	m	1991	0_833	2021-09-27 23:22:40.394509	s_96558	2.99	0
336911	c_8138	f	1984	0_394	2021-11-09 09:02:38.299240	s_116986	2.14	0
336912	c_8327	m	1972	0_394	2021-12-28 22:44:11.200205	s_141516	2.14	0

336913 rows × 8 columns

# Formatage des données

	client	age	sex	session	id_prod	price	categ	horaire	jour	mois	annee	date
0	4410	55	f	9942	0_1455	8.99	0	14:29:25	22	03	2021	03-2021
1	4389	38	m	59967	0_1455	8.99	0	11:16:18	09	07	2021	07-2021
2	5019	45	f	149928	0_1455	8.99	0	00:01:53	15	01	2022	01-2022
3	7049	35	f	1637	0_1455	8.99	0	14:01:38	04	03	2021	03-2021
4	5110	40	f	85364	0_1455	8.99	0	11:48:41	05	09	2021	09-2021
...	...	...	...	...	...	...	...	...	...	...	...	...
336908	7135	26	m	61009	2_99	84.99	2	20:56:49	11	07	2021	07-2021
336909	5828	24	f	122697	2_99	84.99	2	01:53:46	21	11	2021	11-2021
336910	8260	31	m	96558	0_833	2.99	0	23:22:40	27	09	2021	09-2021
336911	8138	38	f	116986	0_394	2.14	0	09:02:38	09	11	2021	11-2021
336912	8327	50	m	141516	0_394	2.14	0	22:44:11	28	12	2021	12-2021

336913 rows × 12 columns

# Recherche des erreurs

---

- ☑ Nous avons recherché des erreurs dans le dataframes afin de le nettoyer pour les analyses.

Dans les slides suivantes nous allons voir celles que nous avons trouvées et traitées:

1

Des prix négatifs dû à un test effectué lors du lancement du site. Nous avons donc décidé de supprimer toutes les valeurs négatives ou nulles.

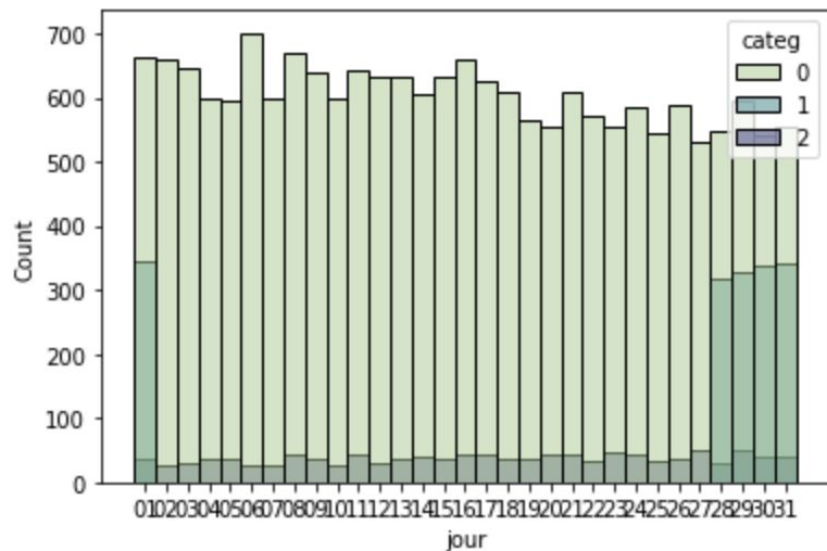
2

Une erreur pour le mois d'Octobre 2021, du 2 au 28 du mois aucune vente de livre de catégorie 1 n'ont été enregistrer dans la base de données. Nous avons donc décidé d'écarter ce mois pour la suite de l'analyse.

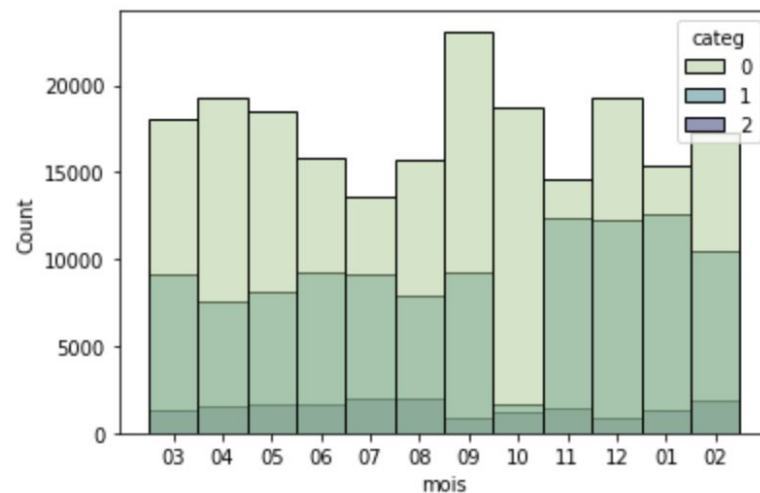
## a / Les valeurs négatives

	client	age	sex	session	id_prod	price	categ	horaire	jour	mois	annee	date
<b>336184</b>	0	21	f	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336234</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336235</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336236</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336237</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336238</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336239</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336240</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336241</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021
<b>336242</b>	1	21	m	0	T_0	-1.0	0	02:30:02	01	03	test_2021	03-test_2021

## b / Le mois d'Octobre



Octobre 2021 (10-2021)

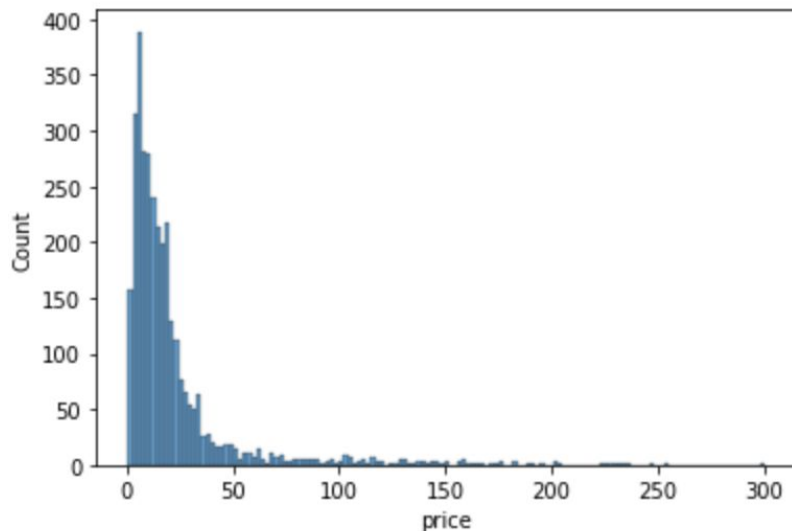


# Les catégories de livres

Dans cette section nous allons étudier la **répartition des livres par catégorie**.

☑ Pour se faire, nous avons déterminé par **catégorie**, le **prix moyen** et **médian**.

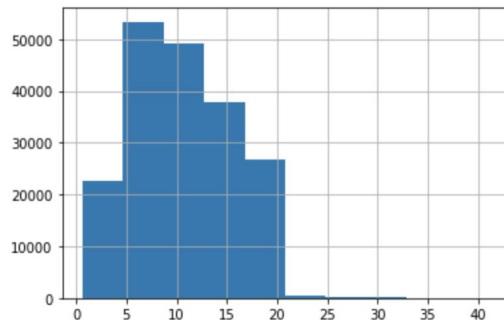
→ On remarque qu'une majorité de livre on un prix entre 0 et 50€.



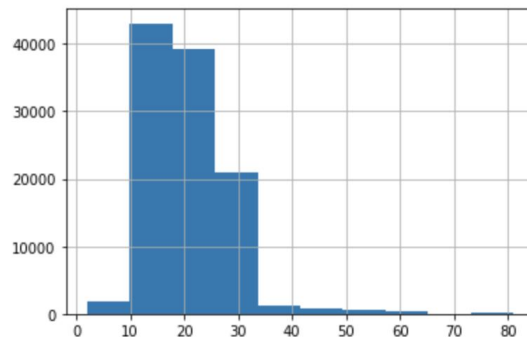
*Représentation du nombre de livres par prix.*

# Descriptif des catégories de livres

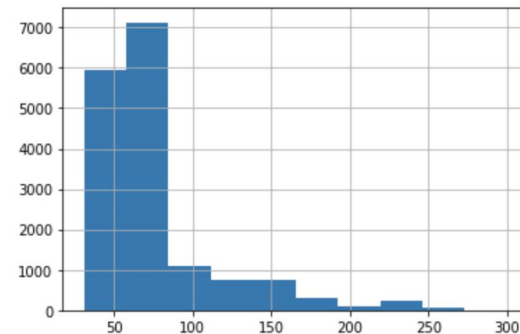
Catégorie : 0  
le prix moyen:  
€ 10.64903479743018  
le prix median:  
€ 9.99



Catégorie : 1  
le prix moyen:  
€ 20.483414207589597  
le prix median:  
€ 19.08

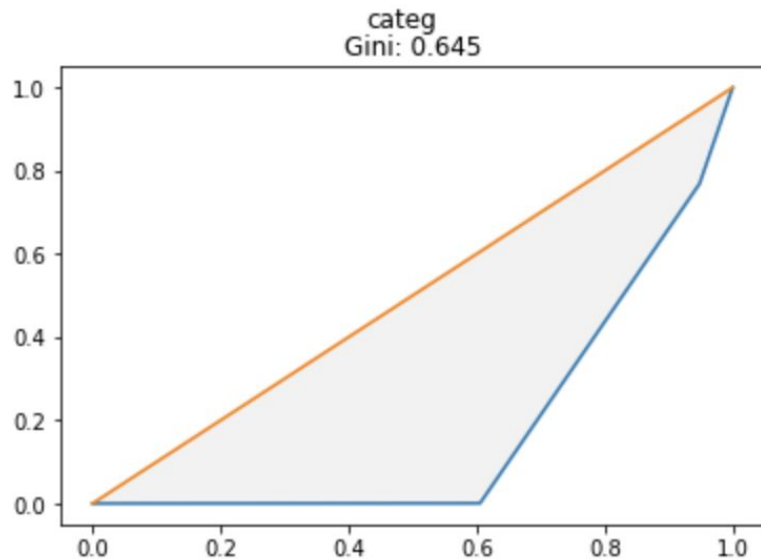


Catégorie : 2  
le prix moyen:  
€ 75.2373724987799  
le prix median:  
€ 62.83





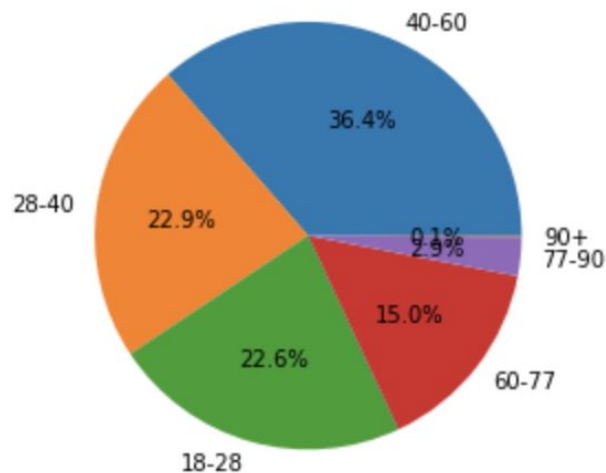
# Courbe de Lorenz et Indice de Gini



→ Nous avons une **répartition inégale** entre les catégories (0,1 et 2) et le prix des livres.

**L'indice de Gini :**  
0,645

# Représentation des catégories d'âge des clients



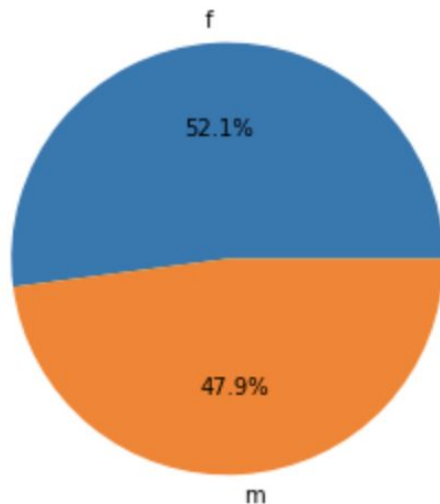
Nos clients ont principalement entre **40 et 60 ans** (36,4%).

La part des 18-28 est pratiquement égale à celle des 28-40. (22,6% et 22,9%).

Les 60-90+ représentent une faible part de nos clients (< 20%)

# Représentation par sexe

---

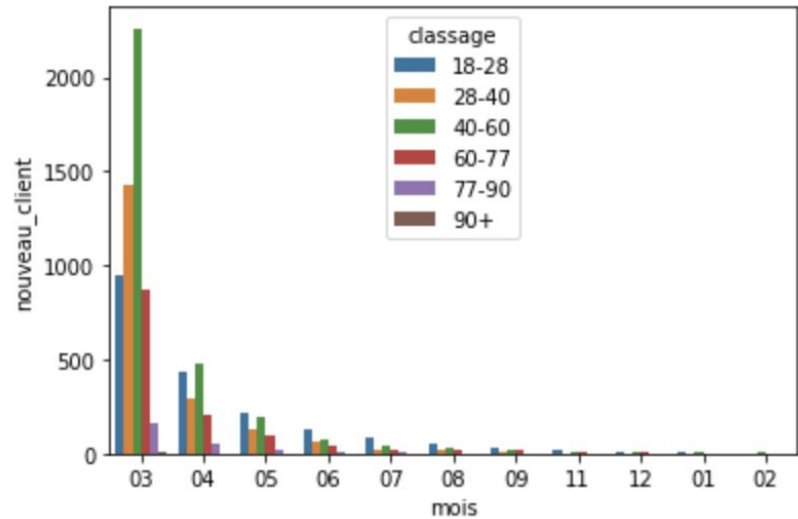


Bien que parmi nos clients, nous avons plus de femmes que d'hommes, cela reste très marginale.

La représentation reste très **égalitaire**.

# Évolution des nouveaux clients

	annee	mois	18-28	28-40	40-60	60-77	77-90	90+	Total
0	2021	03	944.0	1426.0	2261.0	876.0	160.0	9.0	5676
1	2021	04	435.0	297.0	477.0	206.0	52.0	1.0	1468
2	2021	05	221.0	132.0	193.0	95.0	17.0	0.0	658
3	2021	06	133.0	59.0	77.0	44.0	9.0	1.0	323
4	2021	07	81.0	23.0	45.0	21.0	5.0	0.0	175
5	2021	08	55.0	17.0	31.0	17.0	2.0	0.0	122
6	2021	09	28.0	7.0	22.0	16.0	3.0	0.0	76
7	2021	11	22.0	3.0	7.0	8.0	2.0	0.0	42
8	2021	12	12.0	3.0	10.0	4.0	0.0	0.0	29
9	2022	01	11.0	3.0	4.0	1.0	0.0	0.0	19
10	2022	02	3.0	1.0	4.0	2.0	0.0	0.0	10
11	Total		1945.0	1971.0	3131.0	1290.0	250.0	11.0	8598



# Produit qui a attiré le plus de nouveaux clients

	id_prod	18-28	28-40	40-60	60-77	77-90	90+	Total
<b>2158</b>	Total	3457.0	4667.0	6466.0	2034.0	396.0	24.0	17044
<b>1659</b>	1_406	12.0	17.0	13.0	17.0	1.0	0.0	60
<b>1630</b>	1_366	12.0	8.0	25.0	10.0	2.0	0.0	57
<b>1720</b>	1_498	17.0	9.0	17.0	11.0	3.0	0.0	57
<b>1648</b>	1_389	13.0	14.0	19.0	10.0	0.0	0.0	56
<b>1666</b>	1_417	12.0	10.0	21.0	10.0	2.0	0.0	55
<b>1658</b>	1_403	8.0	13.0	21.0	12.0	1.0	0.0	55
<b>1653</b>	1_397	15.0	9.0	18.0	9.0	3.0	0.0	54
<b>1654</b>	1_398	10.0	11.0	20.0	10.0	2.0	1.0	54
<b>1664</b>	1_413	10.0	8.0	27.0	6.0	2.0	0.0	53
<b>1663</b>	1_412	13.0	10.0	21.0	7.0	1.0	0.0	52

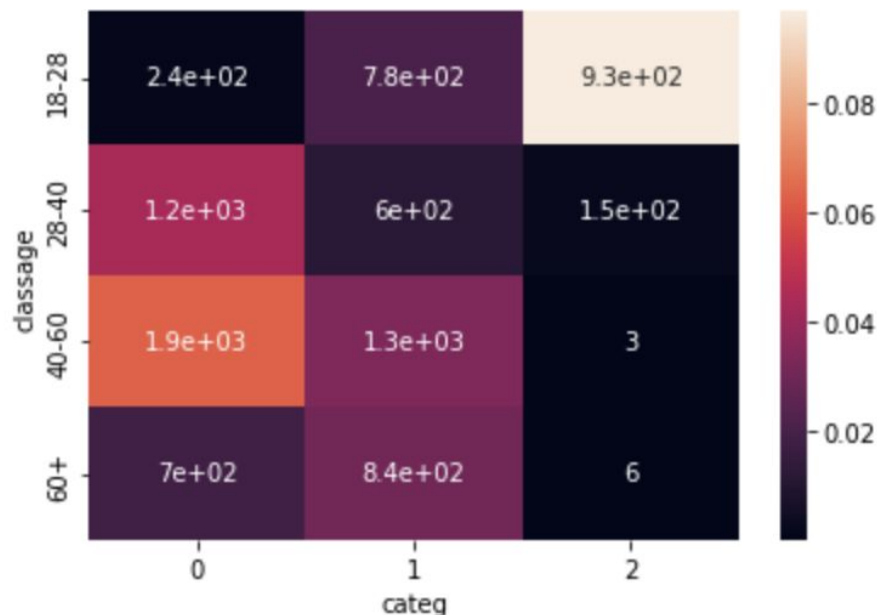
Le produit qui a attiré le plus de client est le livre : '**1\_406**' avec 60 ventes lors du premier achat, suivi de près par les livres '**1\_366**' et '**1\_498**' avec 57.

On remarque, l'absence des livres des catégories '0' et '2' dans le classement des 10 premiers.

# Analyse KHI-2 / Qualitative

categ	0	1	2	Total
classage				
18-28	236	778	931	1945
28-40	1228	596	147	1971
40-60	1862	1266	3	3131
60+	701	844	6	1551
Total	4027	3484	1087	8598

En plus des couleurs, proche de 1 chez les 18-28 dans la catégorie '2' et chez les 40-60 pour la catégorie '0' on obtient une p-valeur de 0.0, inférieur à 0.1% ce qui signifie que les variables présentent une association significative entre elles.



# BEST-SELLER

classage	18-28	28-40	40-60	60-77	77-90	90+	Total
id_prod							
Total	32013.0	95526.0	148211.0	32793.0	6320.0	273.0	315136
1_369	109.0	224.0	505.0	188.0	39.0	1.0	1066
1_417	117.0	265.0	444.0	174.0	43.0	3.0	1046
1_498	115.0	235.0	484.0	153.0	36.0	2.0	1025
1_414	107.0	216.0	463.0	185.0	39.0	3.0	1013
1_425	111.0	238.0	426.0	194.0	31.0	1.0	1001
1_406	109.0	211.0	428.0	163.0	26.0	1.0	938
1_398	110.0	197.0	445.0	153.0	29.0	2.0	936
1_413	101.0	194.0	437.0	156.0	41.0	1.0	930
1_403	91.0	208.0	441.0	165.0	23.0	1.0	929
1_407	109.0	201.0	431.0	147.0	32.0	1.0	921

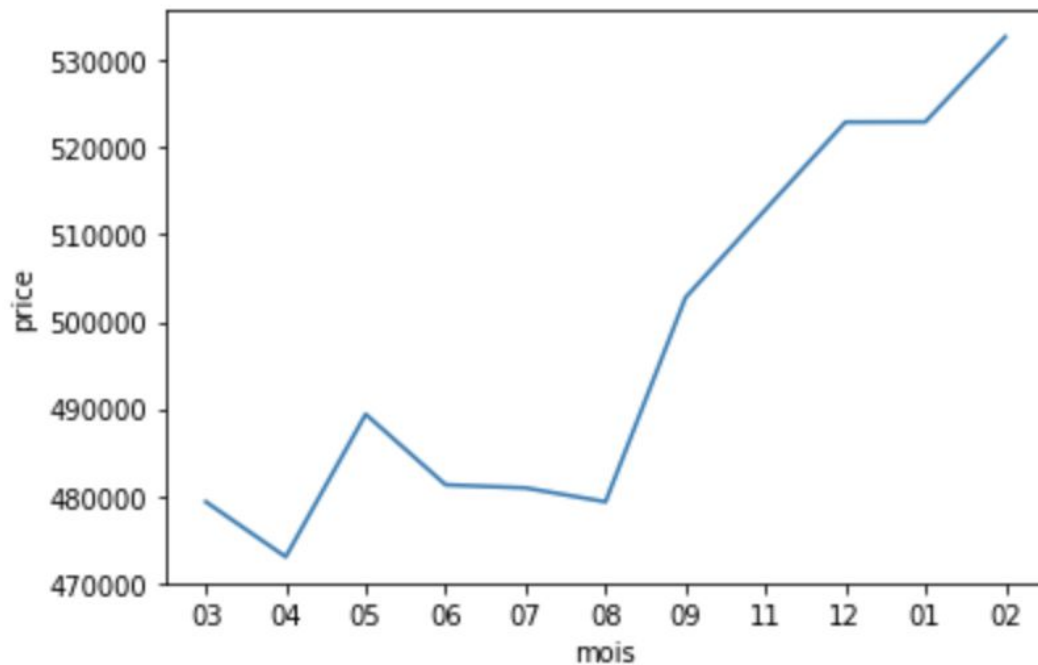
Nous avons un champion, le livre : **'1\_369'** avec 1066 ventes sur la période, dont 505 chez les 40-60 (premier des ventes dans cette catégorie d'âge).

C'est le **'1\_417'** qui prend la seconde place avec 1046 ventes, 1er des ventes chez les 28-40 ainsi que chez les 77-90.

Et la médaille de bronze revient au **'1\_498'** avec 1025 ventes.

Chez les 18-28, c'est le **'2\_102'** fait un carton avec 361 ventes dans leur classe d'âge.

# Evolution du chiffre d'affaire



→ Le chiffre d'affaire moyen de Rester Livre sur la période est de : **497 947€**.



# Part du chiffre d'affaire

mois	Total	03	04	05	06	07	08	09	11	12	01	02
annee		2021	2021	2021	2021	2021	2021	2021	2021	2021	2022	2022
classage												
18-28	265.184792	22.006769	24.302183	25.917817	26.421663	30.553181	30.368530	16.105077	23.567225	16.382419	22.367934	27.191994
28-40	279.333772	26.942149	26.312119	25.344482	25.205367	23.813741	24.589844	27.172739	24.474109	26.478752	25.105692	23.894778
40-60	423.392638	39.086983	38.145768	37.860880	36.634382	34.164063	34.655099	44.166397	38.969285	43.296081	39.209201	37.204500
60-77	109.972443	10.047655	9.415589	8.999778	10.008695	9.553749	8.662454	10.412952	10.795122	11.477215	10.930020	9.669215
77-90	21.203265	1.840528	1.734644	1.820671	1.621964	1.853580	1.646821	2.055818	2.114634	2.278547	2.296022	1.940037
90+	0.913091	0.075917	0.089697	0.056373	0.107930	0.061687	0.077251	0.087018	0.079625	0.086987	0.091130	0.099476
Total	1100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

Au mois de février 2022, les **40-60 représentent 37%** du chiffre d'affaire de Rester Livre, soit une baisse de 2 point, par rapport à l'année précédente. En revanche **les 18-28 gagnent environ 5 points pour atteindre 27%** et les **28-40 perdent 3 points pour tomber à 23%**.

# Part du chiffre d'affaire

Chiffre d'Affaire par classe d'âge



# Les 10 principaux clients

	client	session	age	sex	annee	mois	classage	panier_moyen	nb_session_client	ca_client_mois	ca_client	ca_date
<b>5458</b>	1609	132578	42	m	2021	12	40-60	12.731571	5042	16986.79	150729.07	522937.68
<b>38845</b>	4958	64539	23	m	2021	07	18-28	55.820708	1801	14962.60	137151.48	480898.66
<b>55927</b>	6714	53148	54	f	2021	06	40-60	16.570855	1177	6389.57	69382.17	481257.94
<b>24870</b>	3454	94839	53	m	2021	09	40-60	16.806082	2580	4993.11	52720.68	502771.28
<b>67664</b>	8026	125366	44	m	2021	11	40-60	13.248634	70	159.18	2424.50	512890.17
<b>62428</b>	7421	138715	44	m	2021	12	40-60	13.517809	70	339.71	2406.17	522937.68
<b>61514</b>	7319	169663	48	f	2022	02	40-60	14.084524	63	150.81	2366.20	532740.88
<b>21234</b>	3263	160591	37	f	2022	02	28-40	13.256158	62	255.03	2346.34	532740.88
<b>70908</b>	8392	139827	44	f	2021	12	40-60	13.637895	70	459.89	2332.08	522937.68
<b>18064</b>	2899	77322	28	f	2021	08	18-28	49.224255	31	335.93	2313.54	479305.19
<b>63903</b>	760	11265	42	m	2021	03	40-60	13.134716	70	325.31	2311.71	479312.03

On peut remarquer que 4 clients (clients '1609','4958','6714' et '3454') se démarquent des autres, ils ont dépensé sur la période entre 50 000€ et 150 000€ chacun..

On peut remarquer que certains font des petits achats fréquents alors que d'autres font des achats moins fréquents mais avec un panier moyen plus élevé.

# Les plus important client/mois

	client	session	age	sex	annee	mois	classage	panier_moyen	nb_session_client	ca_client_mois	ca_client	ca_date
<b>5458</b>	1609	132578	42	m	2021	12	40-60	12.731571	5042	16986.79	150729.07	522937.68
<b>38845</b>	4958	64539	23	m	2021	07	18-28	55.820708	1801	14962.60	137151.48	480898.66
<b>55927</b>	6714	53148	54	f	2021	06	40-60	16.570855	1177	6389.57	69382.17	481257.94
<b>24870</b>	3454	94839	53	m	2021	09	40-60	16.806082	2580	4993.11	52720.68	502771.28
<b>12316</b>	2234	165643	21	m	2022	02	18-28	39.363889	12	665.40	1417.10	532740.88
<b>65469</b>	7773	81021	21	m	2021	08	18-28	46.610244	22	626.17	1911.02	479305.19
<b>47507</b>	5825	81572	30	f	2021	08	28-40	36.564667	16	541.77	1096.94	479305.19
<b>216</b>	1023	168083	23	m	2022	02	18-28	49.737308	17	529.32	1293.17	532740.88
<b>35072</b>	4658	170703	24	m	2022	02	18-28	44.034375	15	520.18	1409.10	532740.88
<b>62260</b>	740	28214	27	m	2021	05	18-28	73.957619	14	515.75	1553.11	489367.95
<b>41381</b>	5186	125875	25	m	2021	11	18-28	46.917500	19	493.66	1501.36	512890.17

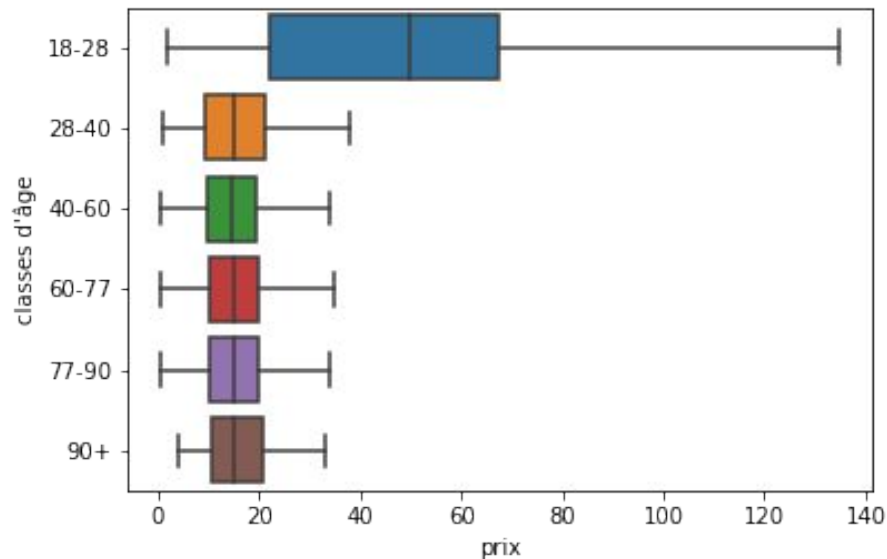
On remarque que les clients 18/28 sont plus présent dans ce classement. Notamment avec des achats lors du mois de février 2022.

## Les principaux clients (02/2022)

	client	session	age	sex	annee	mois	classage	panier_moyen	nb_session_client	ca_client_mois	ca_client	ca_date
<b>12316</b>	2234	165643	21	m	2022	02	18-28	39.363889	12	665.40	1417.10	532740.88
<b>216</b>	1023	168083	23	m	2022	02	18-28	49.737308	17	529.32	1293.17	532740.88
<b>35072</b>	4658	170703	24	m	2022	02	18-28	44.034375	15	520.18	1409.10	532740.88
<b>59765</b>	7135	159546	26	m	2022	02	18-28	45.819333	15	487.00	1374.58	532740.88
<b>68824</b>	8159	166929	18	m	2022	02	18-28	47.076111	23	427.14	1694.74	532740.88
<b>59178</b>	707	160962	36	m	2022	02	28-40	13.334710	59	393.96	2066.88	532740.88
<b>31051</b>	4172	161853	27	f	2022	02	18-28	34.670682	22	364.30	1525.51	532740.88
<b>28947</b>	3918	170214	18	f	2022	02	18-28	67.589600	14	340.25	1689.74	532740.88
<b>6812</b>	1634	169712	18	m	2022	02	18-28	61.898261	14	330.54	1423.66	532740.88
<b>17598</b>	283	167032	26	m	2022	02	18-28	48.122632	22	321.39	1828.66	532740.88
<b>67493</b>	801	171417	24	f	2022	02	18-28	41.804063	20	315.45	1337.73	532740.88

En regardant précisément le mois de février, on constate que les 18-28 deviennent les principaux clients de notre boutique avec des paniers élevés.

# Corrélation classes d'âge et prix ?

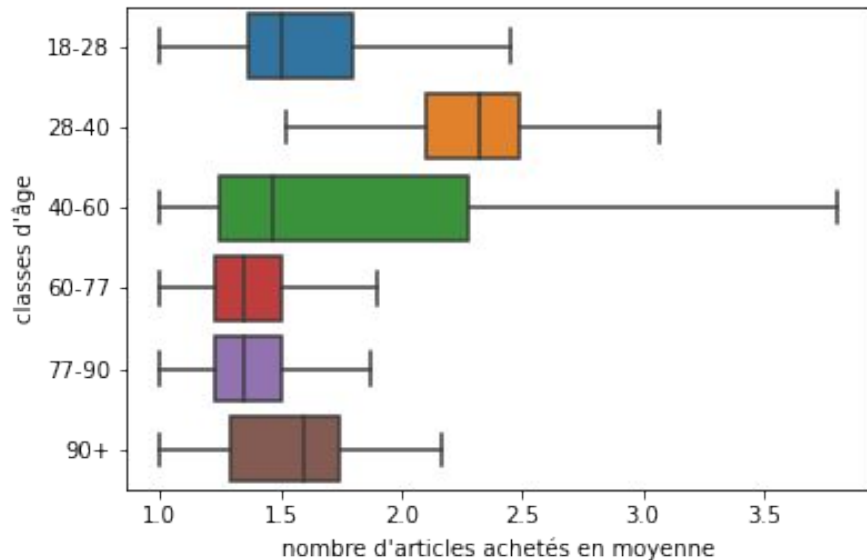


On voit dans le graphique, qu'il semble exister une variance entre le prix et la classe d'âge, notamment chez les 18-28 ans.

Nous vérifions donc ça, avec l'analyse de la variance. On obtient au rapport de variance, un pourcentage de variation proche de 0.3 ce qui veut dire que les variables sont donc corrélées.

$$\text{Eta}^2 = 0,3165$$

# Corrélation classes d'âge et nombre d'articles achetés ?

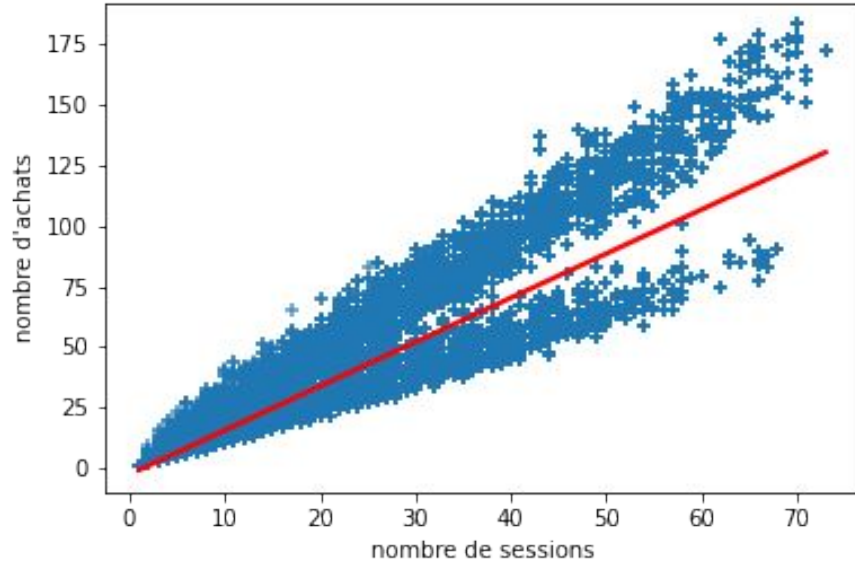


Dans le graphique qu'il semble exister une variance entre le nombre d'article achetés en moyenne et la classe d'âge.

Au rapport de variance, le pourcentage de variation proche de 0.3 ce qui veut dire que les variables sont donc corrélées, il existe une variance interclasse.

$$\text{Eta}^2 = 0,2551$$

# Corrélation nombre d'achats et nombre de sessions ?



Dans le graphique ci-contre, on cherche à savoir s'il existe une relation entre le nombre d'achats du client et son nombre de session.

Afin d'éviter une erreur de calcul, nous supprimons les 4 outliers (> 1000 sessions).

Grâce au graphique, on remarque qu'au delà de 30 sessions, une séparation s'effectue dans le nombre d'achats.

$$y = a \cdot x + b$$

$$a = 1.8$$

x: nombre de sessions

$$b = -2.5$$

Avec 50 sessions, le client fera en moyenne 88.4 achats soit 89 achats.