



# Analyse des ventes en ligne de la librairie Lapage

Indicateurs de vente et comportements clients

- 
- I. Préparation des données
  - II. Le CA
  - III. Les comportements clients



# I. Préparation des données

# I. Importation, préparation et nettoyage des données

**Langage de programmation :** Python

**Librairies utilisées :** Pandas Numpy Matplotlib Seaborn  
Scipy.stats Statsmodels Researchpy  
Pingouin

## **Préparation des données pour analyse :**

Vérification de l'unicité et de l'absence de données manquantes

Formatage des données (ex: format dates, numériques)

Exploration des données : identification des données aberrantes (ex : dates test non formatables, absence de données d'une catégorie sur un mois ) à explorer

Création des colonnes pour faciliter l'analyse (ex : 'Month', 'classe\_age')



6.3.0



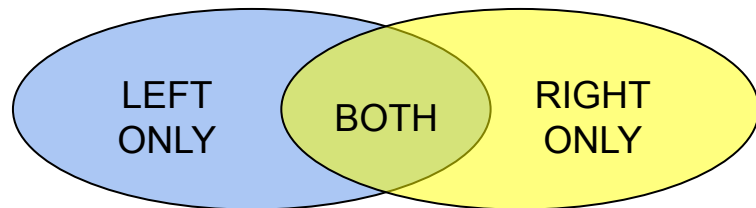
Lapage

# I. Importation, préparation et nettoyage des données

**Les jointures (transactions-produits-clients), création du df global :**

transactions.merge(products,how="outer",on="id\_prod",indicator=True)

```
: general_1['_merge'].unique()  
['both', 'left_only', 'right_only']  
Categories (3, object): ['both', 'left_only', 'right_only']
```



Isolations (écartées de l'analyse)

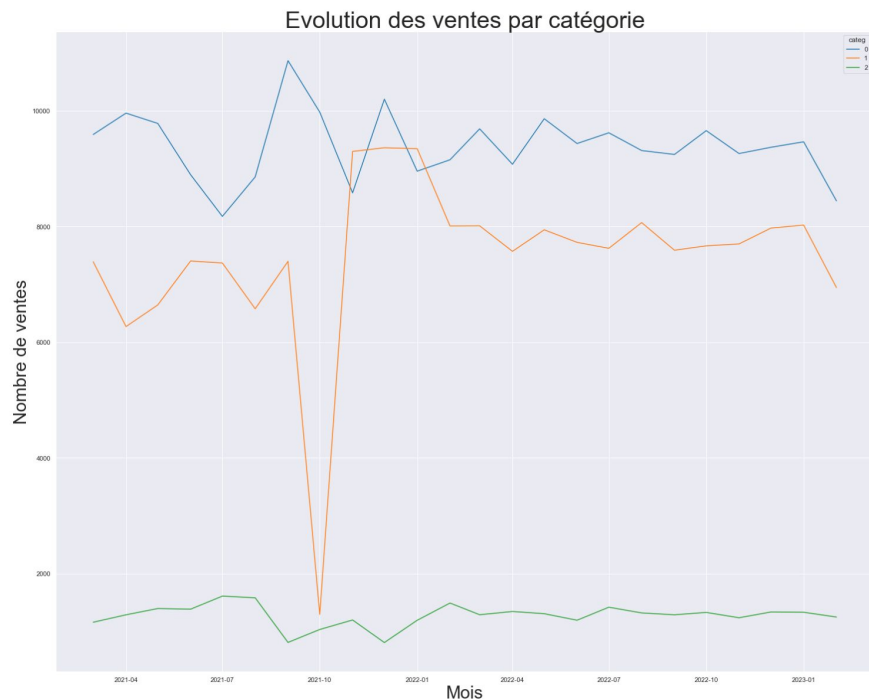
- d'une référence vendue à plusieurs reprises mais non répertoriée : ['0\_2245']
- 22 références non vendues mais répertoriées (dont référence test T\_0)

general\_11.merge(clients,how="outer",on="client\_id",indicator=True)

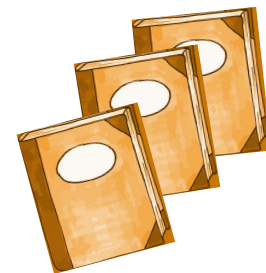
- 23 Clients répertoriés mais non acheteurs (dont client test ct\_0 et ct\_1)



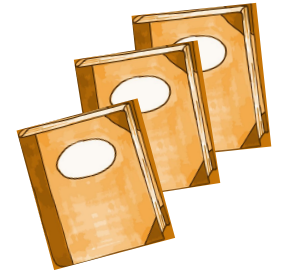
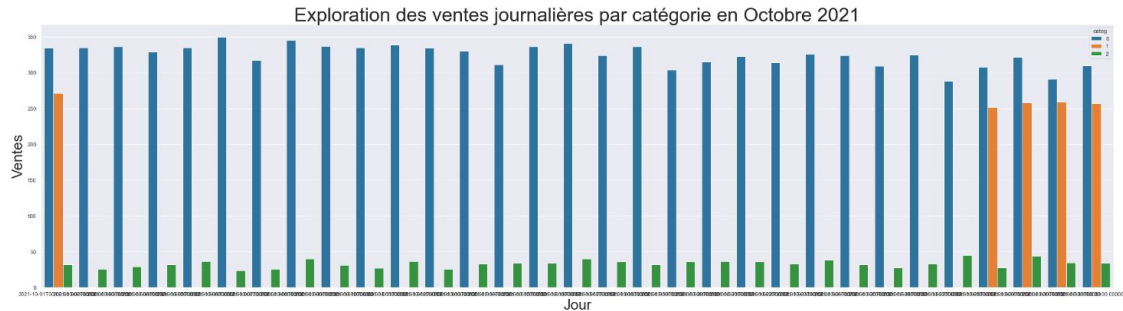
# I. Importation, préparation et nettoyage des données



**Exploration des ventes dans le temps :**  
Soudaine chute des ventes en octobre sur la **catégorie 1**



# I. Importation, préparation et nettoyage des données



ventes du 02 au 27 octobre ?

Pour l'analyse, **Isolation du mois d'Octobre 2021**



## II. Le CA



## A- Evolution du CA



### Evolution du CA



le CA total sur la  
période (hors Oct  
2021) est de  
11 532 929.94 euros

# A- Evolution du CA



## CA total par catégorie

categ	CA_par_categorie
-------	------------------



0	4.220480e+06
---	--------------



1	4.619960e+06
---	--------------

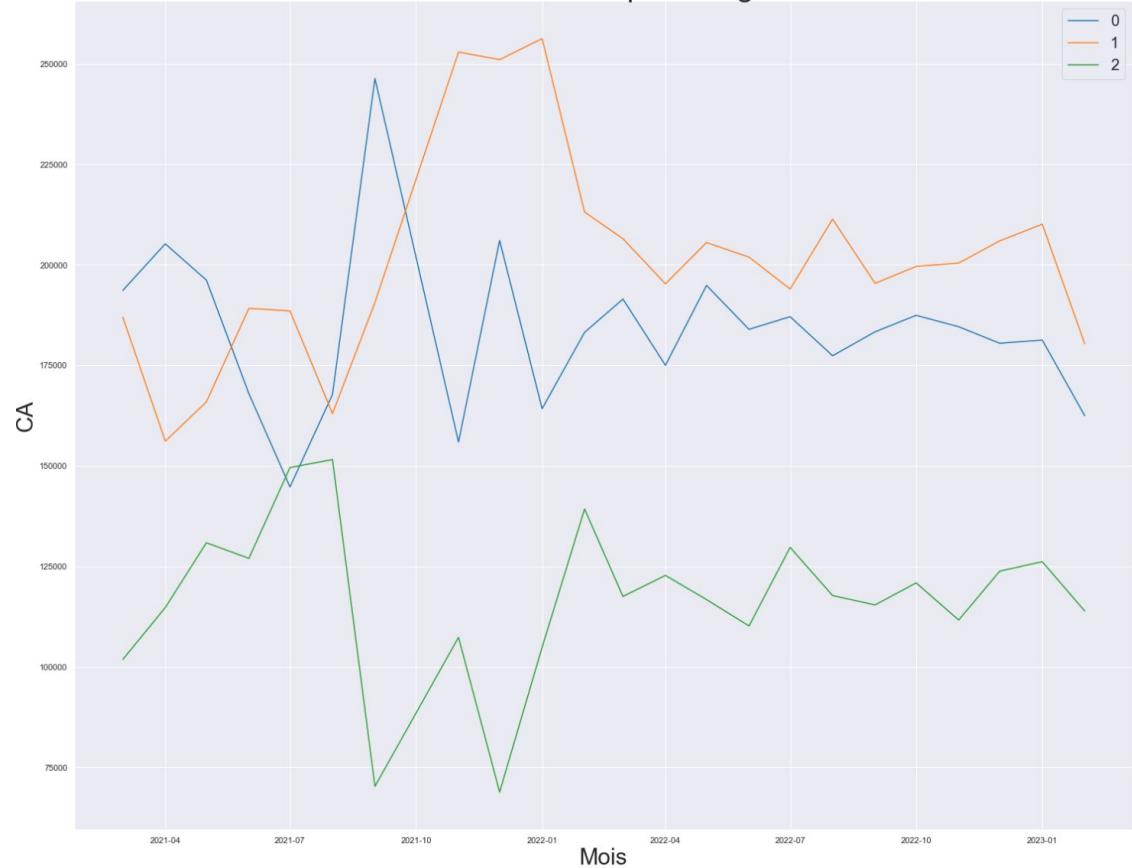


2	2.692489e+06
---	--------------



Lapage

Evolution du CA par catégorie



## B- Zoom sur les références

### Les plus vendus



	id_prod	categ	price	nombre_de_vente
0	1_369	1	23.99	2237
1	1_417	1	20.99	2173
2	1_414	1	23.83	2166
3	1_498	1	23.37	2117
4	1_425	1	16.99	2084
5	1_403	1	17.99	1950
6	1_412	1	16.65	1938
7	1_413	1	17.99	1931
8	1_406	1	24.81	1931
9	1_407	1	15.99	1923

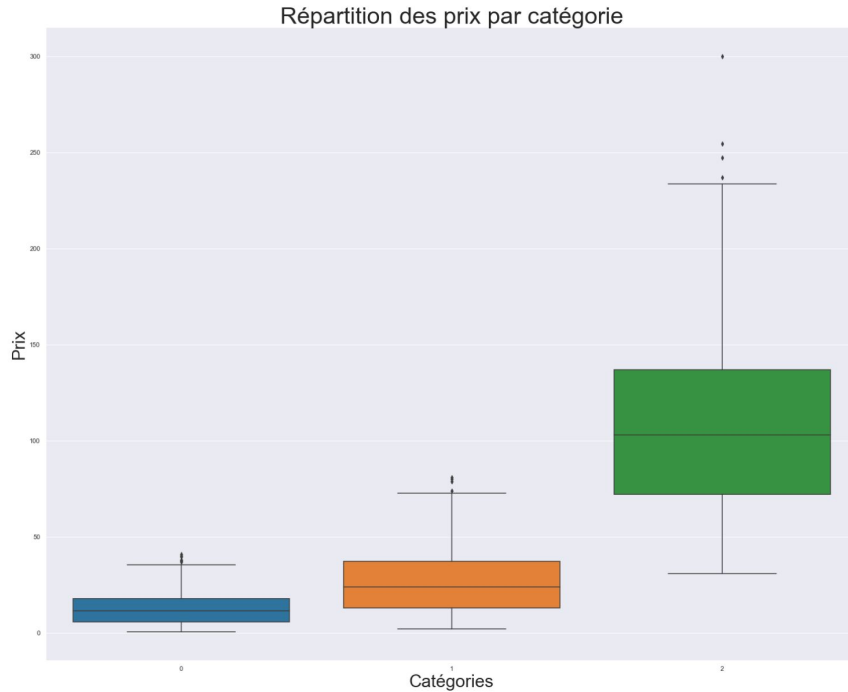
### Les moins vendus



	id_prod	categ	price	nombre_de_vente
0	0_1151	0	2.99	1
1	0_898	0	1.27	1
2	0_886	0	21.82	1
3	0_833	0	2.99	1
4	0_807	0	1.99	1
5	2_23	2	115.99	1
6	0_1498	0	2.48	1
7	0_541	0	1.99	1
8	0_1533	0	27.99	1
9	0_1539	0	0.99	1



## C- Description des trois catégories



categ	nombre_de_produits
-------	--------------------

0	2290
---	------

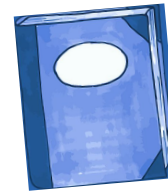
0	2290
---	------

1	737
---	-----

1	737
---	-----

2	236
---	-----

2	236
---	-----

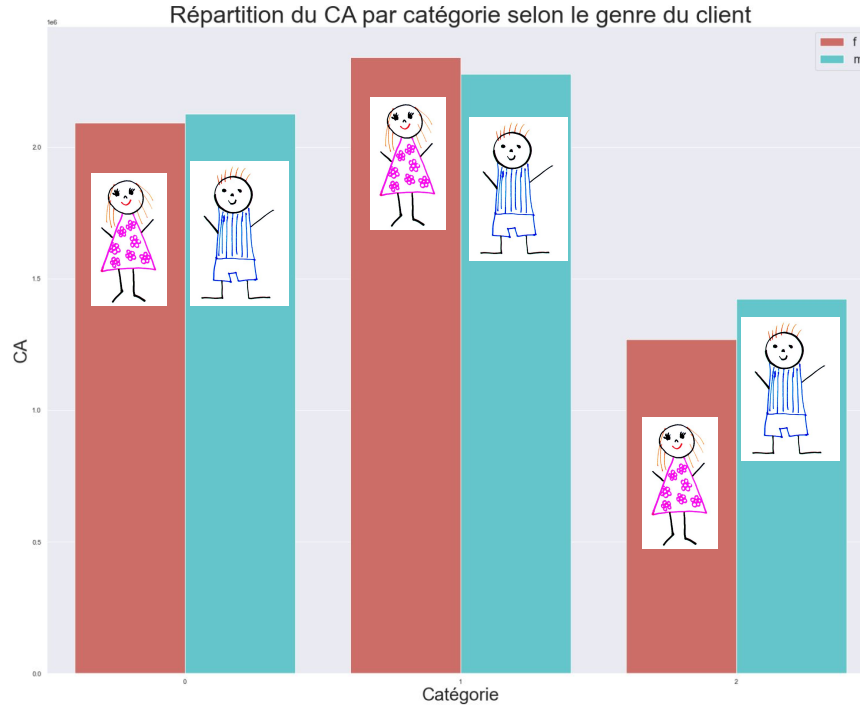


**Beaucoup de référence  
petit prix**



**Peu d'offre et plus cher**

## D- Profils clients



8 600 clients

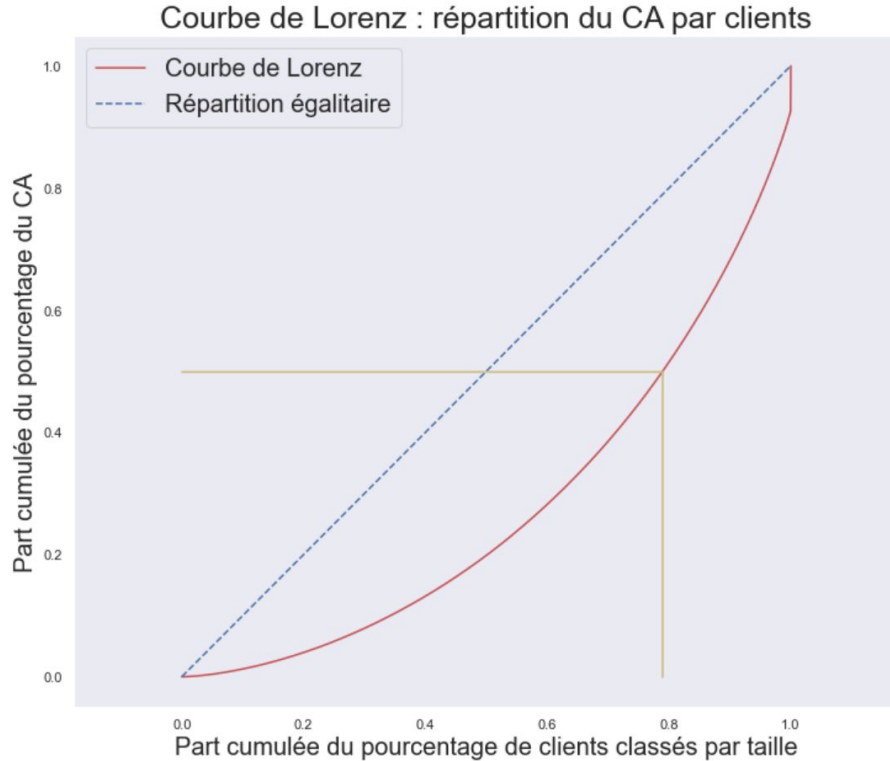
sex

f 4477

m 4121

Les hommes achètent plus  
sur la **catégorie 2** que les  
femmes

## D- Profils clients : gros clients et petits clients ?



**Indice de Gini : 0.45**  
La répartition par client du CA n'est pas égalitaire.

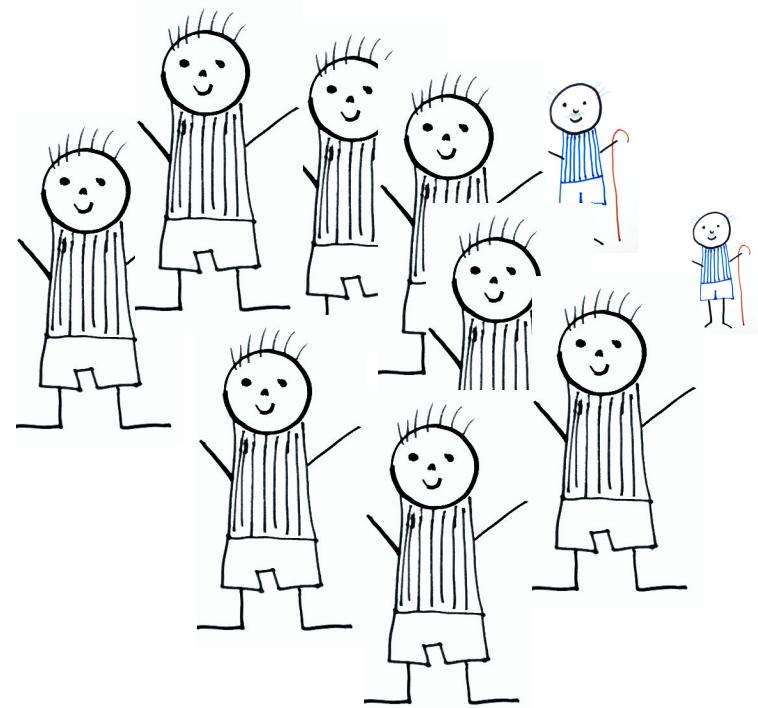
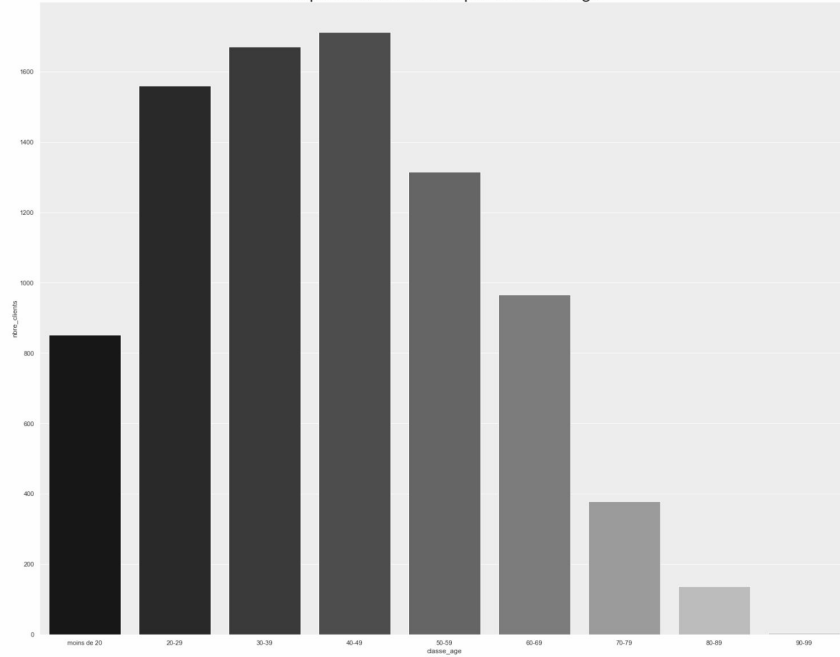
**20% des clients réalisent 50 % du CA**



Lapage

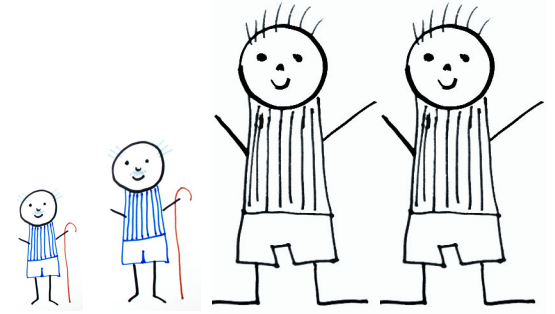
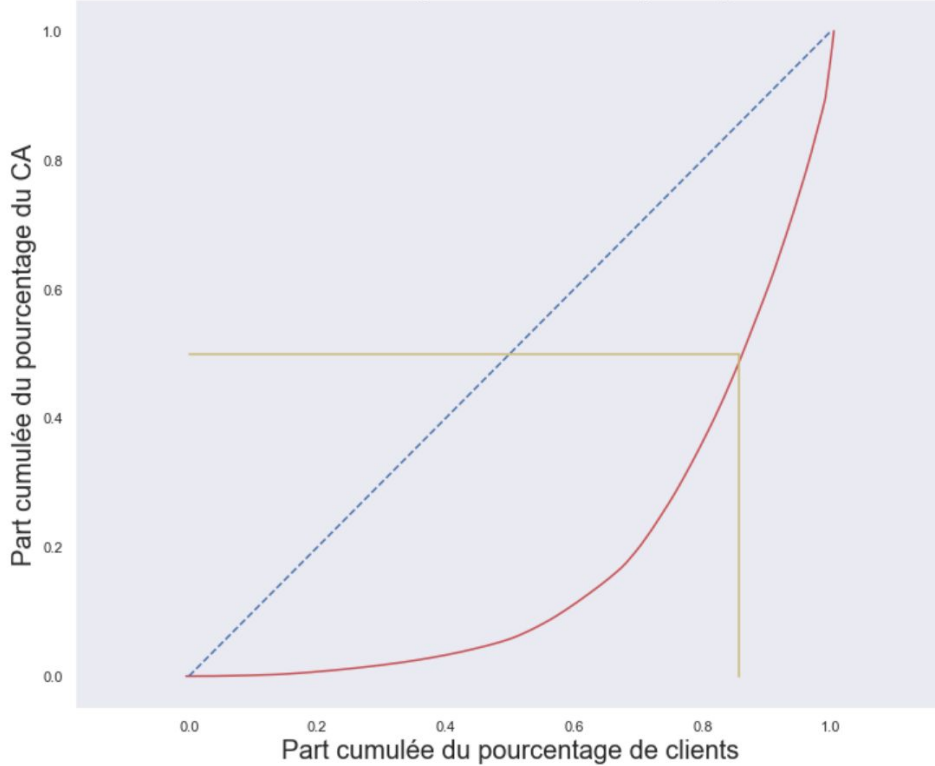
## D- Profils clients : l'âge

Répartition des clients par tranche d'âge



## D- Profils clients : l'âge

Courbe de Lorenz : répartition du CA par âge des clients



**Indice de Gini : 0.65**  
Certains âges réalisent plus de CA.

**15% des clients réalisent 50 % du CA**





### **III. Les comportements clients**

# A - Le sexe du client va-t-il influencer le choix de catégorie d'achat ?

Deux variables qualitatives : Chi 2 suivi d'un Cramer's V



categ	0	1	2
sex			
f	196986	114087	16722
m	199711	111416	18590

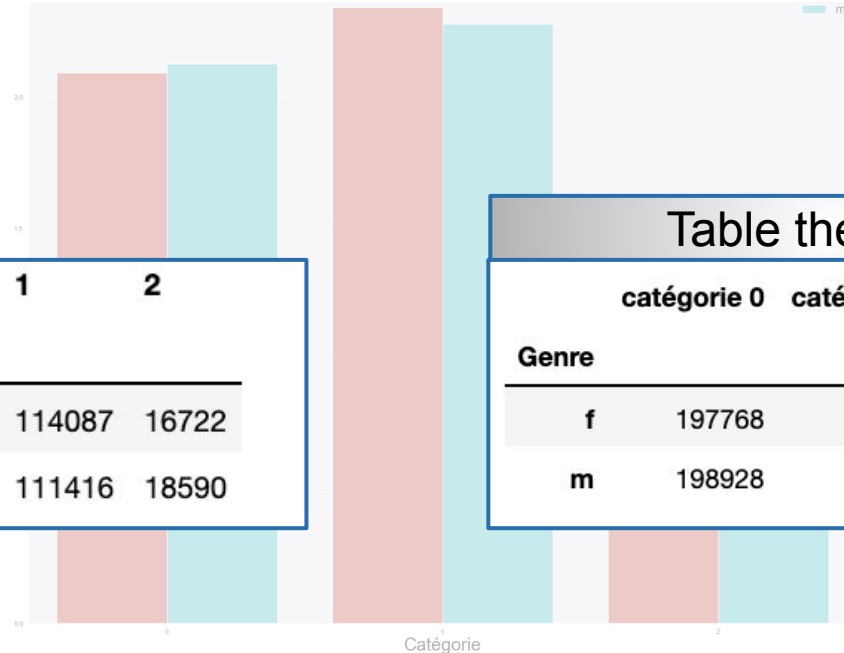
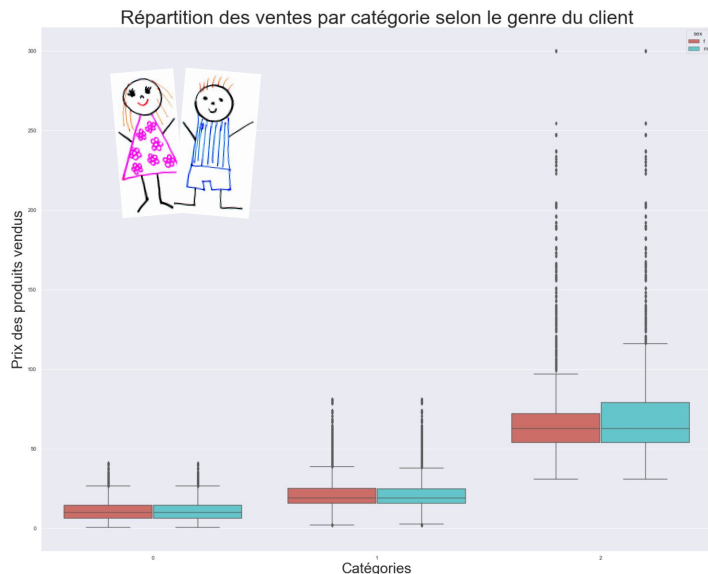


Table théorique			
	catégorie 0	catégorie 1	catégorie 2
Genre			
f	197768	112421	17604
m	198928	113081	17707

# A - Le sexe du client va-t-il influencer le choix de catégorie d'achat ?



Chi-square test		results
0	Pearson Chi-square ( 2.0 ) =	143.5555
1	p-value =	0.0000
2	Cramer's V =	0.0148

Taille de l'effet sur  
chaque catégorie

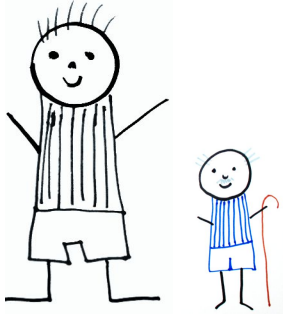


0.019023  
0.025231  
0.063760

Le genre du client va influencer de manière **négligeable** la catégorie de livre acheté.  
Les femmes achètent un peu plus dans la catégorie 1

## B- Le montant total des achats varie selon l'âge

Deux variables quantitatives : test de PEARSON et test de SPEARMAN



```
3]: pg.corr(CA_par_age[ 'age' ],CA_par_age[ 'CA' ])
```

	n	r	CI95%	p-val	BF10	power
pearson	228	-0.486246	[-0.58, -0.38]	6.211433e-15	1.101e+12	1.0



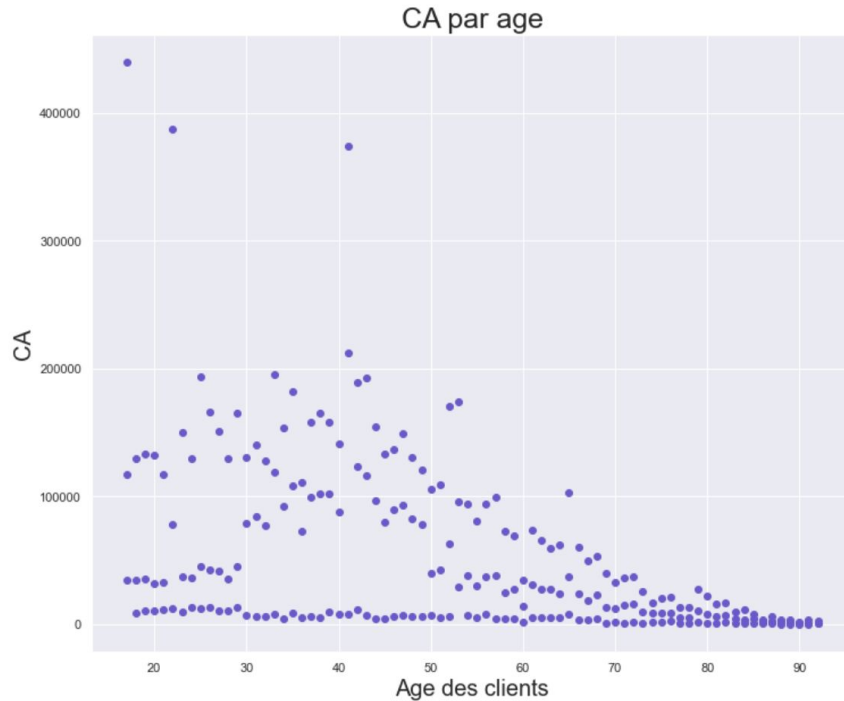
```
4]: pg.corr(CA_par_age[ 'age' ],CA_par_age[ 'CA' ],method='spearman')
```

	n	r	CI95%	p-val	power
spearman	228	-0.6192	[-0.69, -0.53]	1.582073e-25	1.0

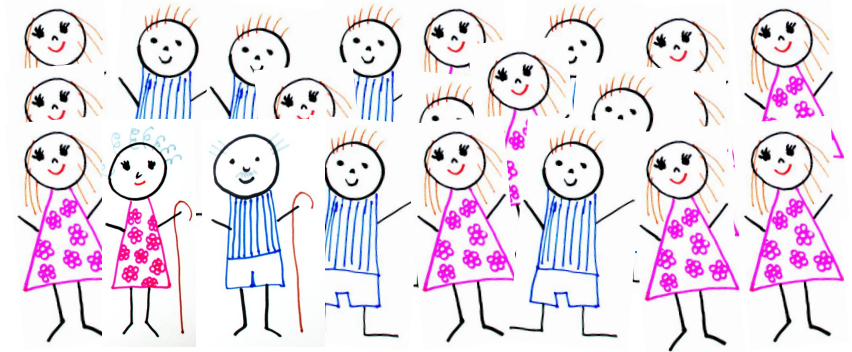


Spearman's  $r$  > Pearson's  $r$  : suggère une relation plutôt non-linéaire entre les deux variables.

## B- Le montant total des achats varie selon l'âge



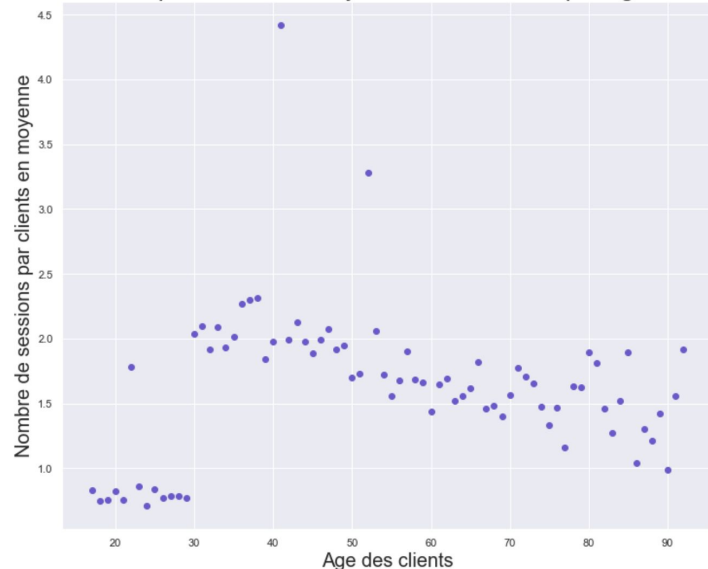
Les clients les plus jeunes contribuent le plus au CA



## C - L'âge et fréquence d'achat

### Deux variables quantitatives : test de PEARSON et test de SPEARMAN

Répartition de la moyenne des sessions par âge



```
: pg.corr(sessions_par_age['age'],sessions_par_age['frequence_achat'])
```

	n	r	CI95%	p-val	BF10	power
pearson	76	0.06247	[-0.17, 0.28]	0.591889	0.165	0.083473



```
: pg.corr(sessions_par_age['age'],sessions_par_age['frequence_achat'],method='spearman')
```

	n	r	CI95%	p-val	power
spearman	76	-0.056187	[-0.28, 0.17]	0.629742	0.076938

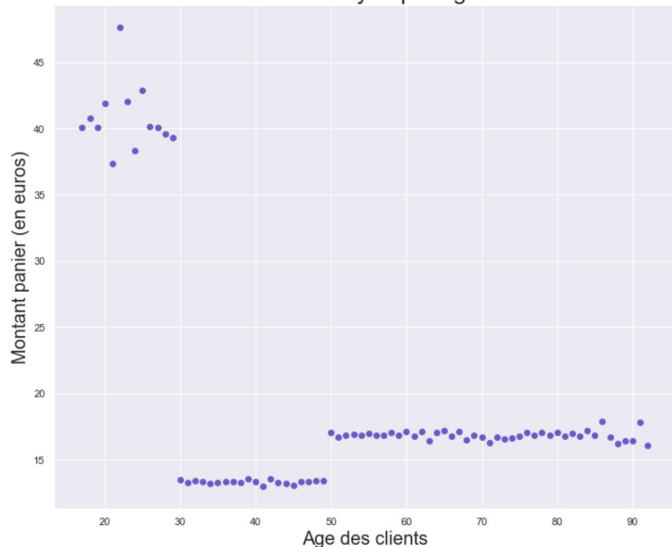


L'âge n'influence pas significativement la fréquence d'achat

## D - La taille du panier moyen dépend de l'âge

### Deux variables quantitatives : test de PEARSON et test de SPEARMAN

Panier moyen par age



```
] : pg.corr(panier_par_age['age'], panier_par_age['panier_moyen'])
```

	n	r	CI95%	p-val	BF10	power
pearson	76	-0.547412	[-0.69, -0.37]	3.100878e-07	5.404e+04	0.999552



```
] : pg.corr(panier_par_age['age'], panier_par_age['panier_moyen'], method='spearman')
```

	n	r	CI95%	p-val	power
spearman	76	-0.090062	[-0.31, 0.14]	0.439108	0.121086

Le panier moyen est corrélé par paliers à l'âge du client



## E - La catégorie de livre achetée dépend de l'âge

### Une variable qualitative et une quantitative : ANOVA, Tukey et eta-squared

```
pg.anova(data=general2, dv='age', between='categ', detailed=True)
```

	Source	SS	DF	MS	F	p-unc	np2
0	categ	1.449148e+07	2	7.245742e+06	44359.518228	0.0	0.11889
1	Within	1.073984e+08	657509	1.633413e+02	NaN	NaN	NaN

```
pg.pairwise_tukey(dv='age', between='categ', data=general2 ,effsize='eta-square')
```



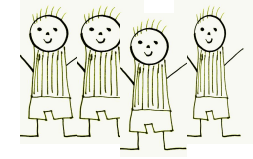
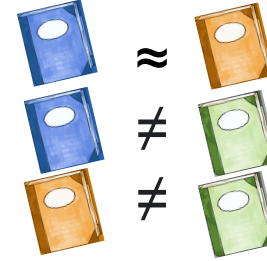
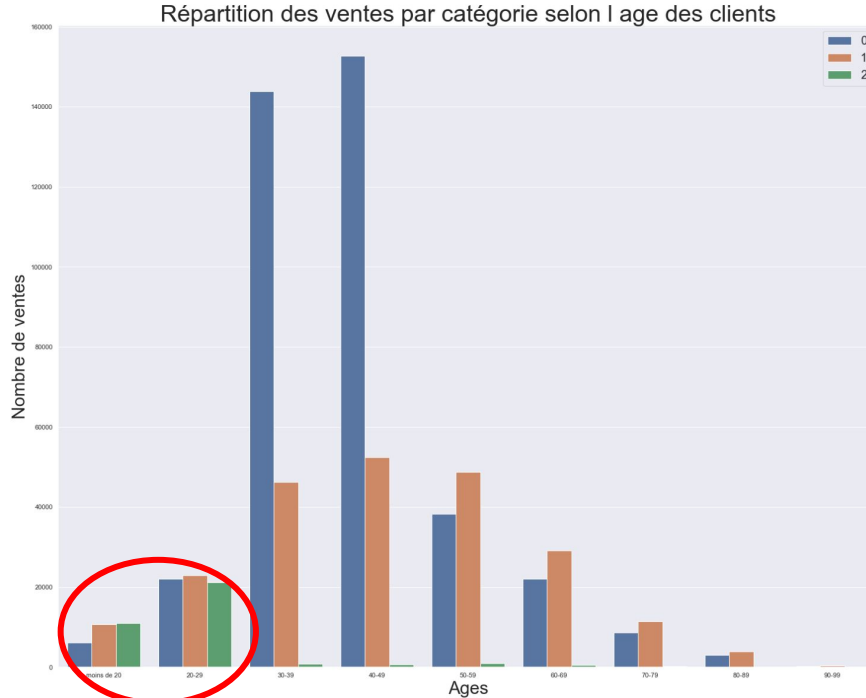
A	B	mean(A)	mean(B)	diff	se	T	p-tukey	eta-square
0	1	42.902760	46.650364	-3.747604	0.033706	-111.185034	0.001	0.021043
0	2	42.902760	24.941408	17.961352	0.070975	253.066826	0.001	0.330552
1	2	46.650364	24.941408	21.708956	0.073144	296.798677	0.001	0.419047





## E - La catégorie de livre achetée dépend de l'âge

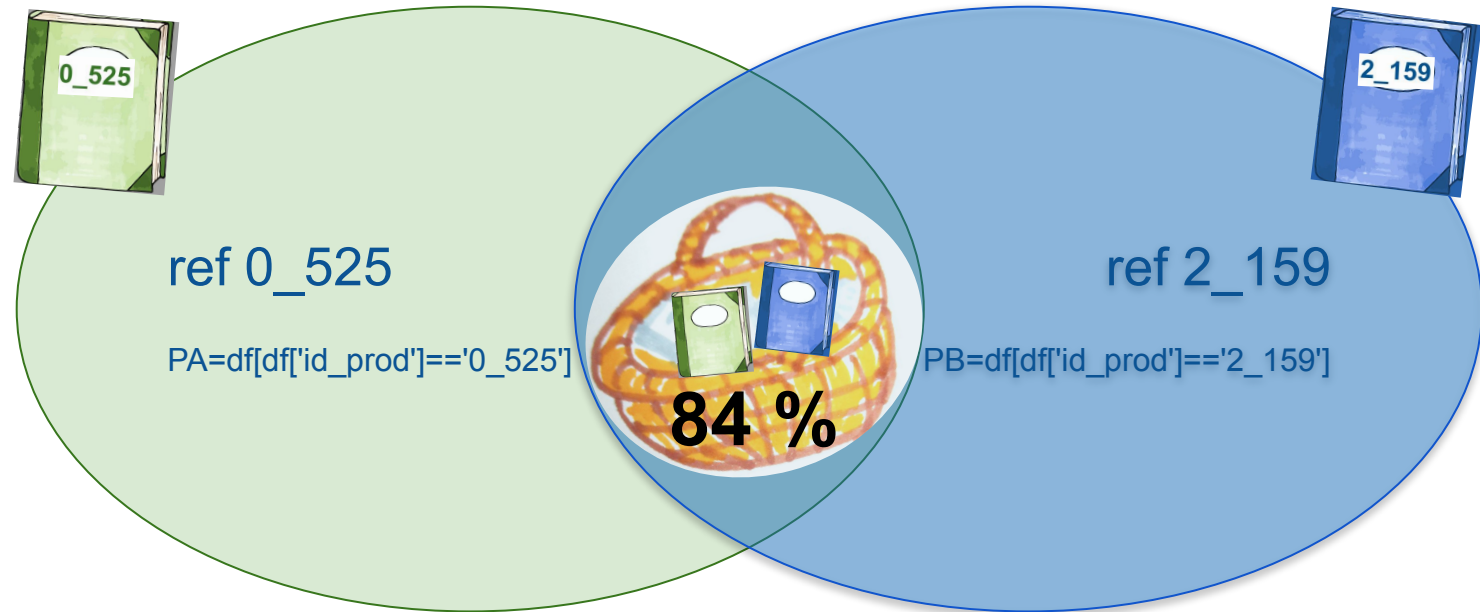
Une variable qualitative et une quantitative : ANOVA, Tukey et eta-squared



La **catégorie 2** a des acheteurs jeunes uniquement

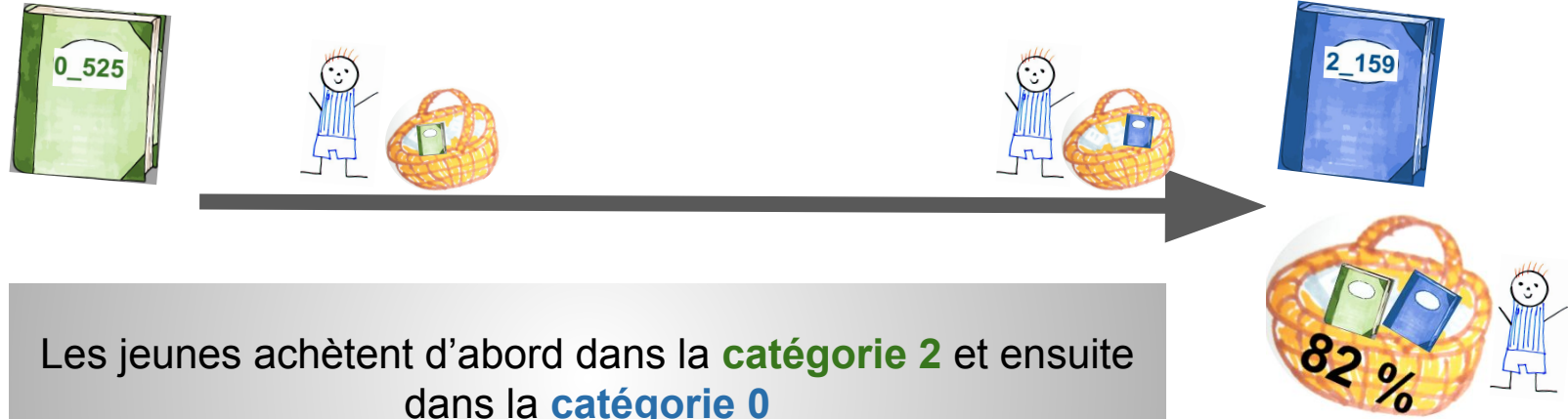
les jeunes achètent dans les trois catégories

## F - Probabilité qu'un client achète la **ref 0\_525** sachant qu'il a acheté **2\_159** ?

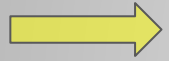


$$\text{PA.count()} / ( \text{PA.count()} + \text{PB.count()} ) * 100$$

F - Probabilité qu'un client achète la **ref 0\_525** sachant qu'il a acheté **2\_159** ?



## Recommandations :



Fort potentiel sur la clientèle jeune  
Proposer plus de références dans la **catégorie 2**  
(stratégie marketing pour augmenter leur fréquence d'achat)



va très probablement booster les ventes de la **catégorie 0**  
Epurer la **catégorie 0** des références en échec



Différencier le B2B



Lapage

**Merci !**

