

# PROJET 4

Analyse des ventes de l'entreprise



# Mission 1

# Nettoyage des données

Valeurs aberrantes

Valeurs manquantes

Autres

# Valeurs aberrantes

	id_prod	price	categ
731	T_0	-1.00	0

Produit Test

	client_id	sex	birth
2735	NaN	f	2001
8494	NaN	m	2001

Client Test

	id_prod	date	session_id	client_id
1431	NaN	test_2021-03-01 02:30:02.237420	NaN	NaN
2365	NaN	test_2021-03-01 02:30:02.237446	NaN	NaN
2895	NaN	test_2021-03-01 02:30:02.237414	NaN	NaN
5955	NaN	test_2021-03-01 02:30:02.237441	NaN	NaN
7283	NaN	test_2021-03-01 02:30:02.237434	NaN	NaN
...	...	...	...	...
332594	NaN	test_2021-03-01 02:30:02.237445	NaN	NaN
332705	NaN	test_2021-03-01 02:30:02.237423	NaN	NaN
332730	NaN	test_2021-03-01 02:30:02.237421	NaN	NaN
333442	NaN	test_2021-03-01 02:30:02.237431	NaN	NaN
335279	NaN	test_2021-03-01 02:30:02.237430	NaN	NaN

[200 rows x 4 columns]

## Session Test

```
data_ransac = data_transat.dropna(axis=0)  
print(data_transat.shape, data_ransac.shape)
```

```
(337016, 4) (336816, 4)
```

Suppression des valeurs aberrantes

200 valeurs aberrantes

Jointure des DataFrames après la suppression des valeurs aberrantes

```
data_trancli = pd.merge(data_cust, data_ransac, how='outer', on='client_id')
```

```
data_glob = pd.merge(data_trancli, data_prod, how='outer', on='id_prod')
```

# Valeurs manquantes

	client_id	sex	birth	id_prod	date	session_id
29516	c_8253	f	2001	NaN	NaN	NaN
99355	c_3789	f	1997	NaN	NaN	NaN
108189	c_4406	f	1998	NaN	NaN	NaN
108190	NaN	f	2001	NaN	NaN	NaN
108191	NaN	m	2001	NaN	NaN	NaN
109298	c_2706	f	1967	NaN	NaN	NaN
111920	c_3443	m	1959	NaN	NaN	NaN
123283	c_4447	m	1956	NaN	NaN	NaN
123537	c_3017	f	1992	NaN	NaN	NaN
129494	c_4086	f	1992	NaN	NaN	NaN
155403	c_6930	m	2004	NaN	NaN	NaN
158644	c_4358	m	1999	NaN	NaN	NaN
192279	c_8381	f	1965	NaN	NaN	NaN
194914	c_1223	m	1963	NaN	NaN	NaN
242573	c_6862	f	2002	NaN	NaN	NaN
245943	c_5245	f	2004	NaN	NaN	NaN
250815	c_5223	m	2003	NaN	NaN	NaN
255213	c_6735	m	2004	NaN	NaN	NaN
259513	c_862	f	1956	NaN	NaN	NaN
268949	c_7584	f	1960	NaN	NaN	NaN
303029	c_90	m	2001	NaN	NaN	NaN
303773	c_587	m	1993	NaN	NaN	NaN
333787	c_3526	m	1956	NaN	NaN	NaN

Anciens clients n'ayant plus d'achat à son nom

```
print(data_glob[data_glob['date'].isnull() == True].count())
print('-'*50)
print(data_glob[data_glob['client_id'].isnull() == True].count())
data_glob_nan = data_glob.dropna(axis=0)
print(data_glob.shape, data_glob_nan.shape)
```

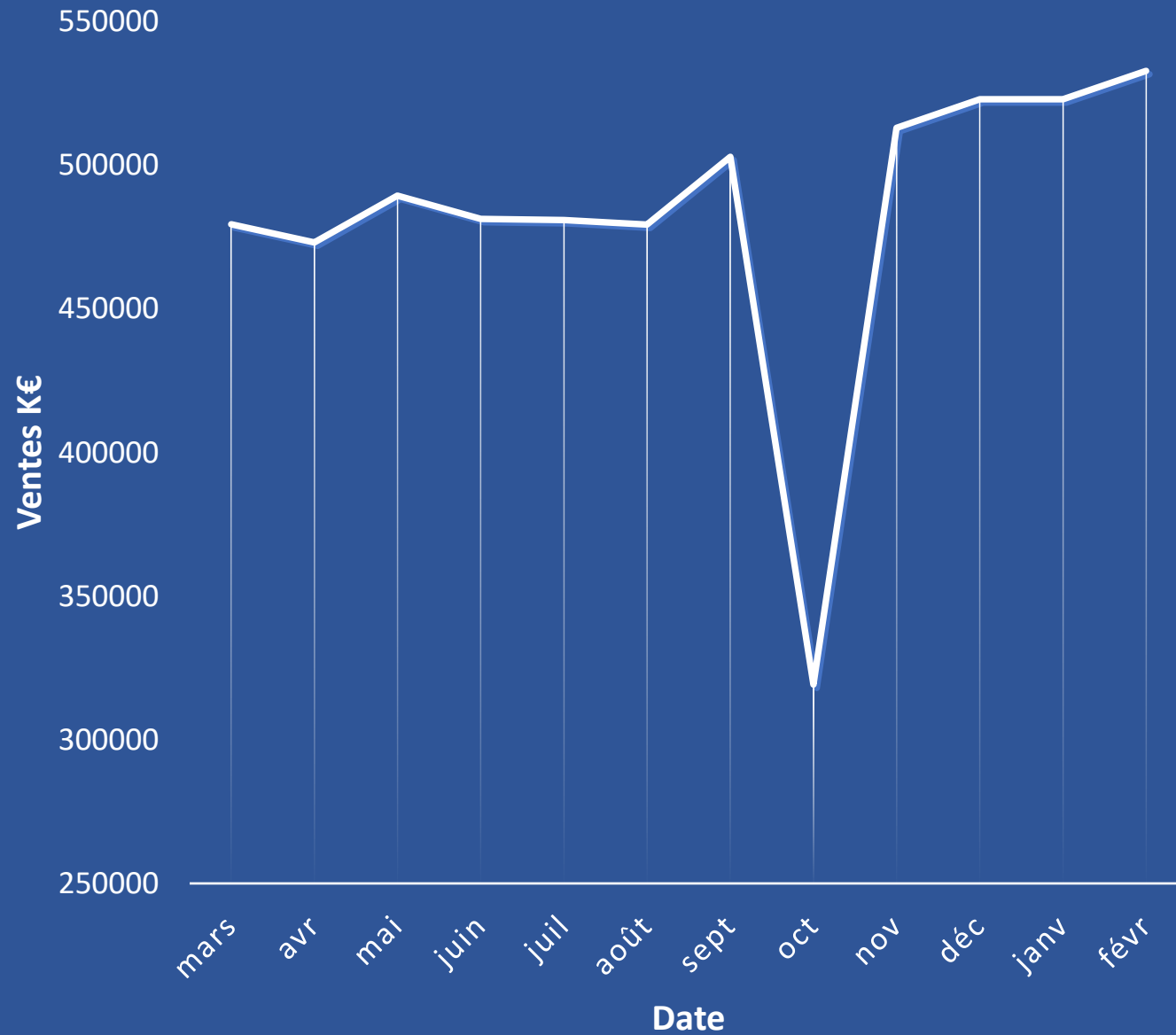
client_id	21
sex	21
birth	21
id_prod	22
date	0
session_id	0
price	43
categ	43
dtype: int64	
-----	
client_id	0
sex	0
birth	0
id_prod	22
date	0
session_id	0
price	22
categ	22
dtype: int64	
(336859, 8) (336816, 8)	

Nombre de dates manquantes

Nombre de client\_id manquant



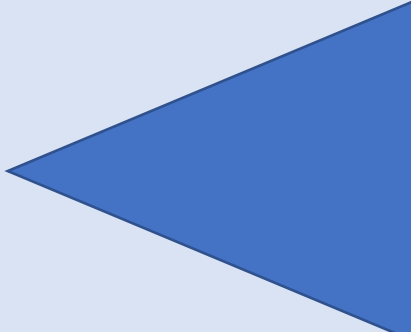
## EVOLUTION DES VENTES



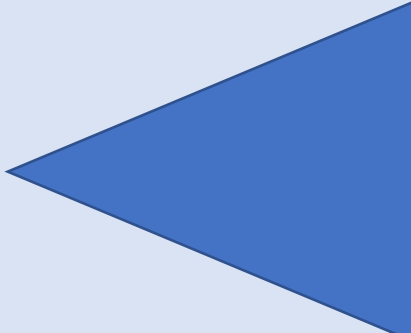
A quoi est due cette chute  
du chiffre d'affaire en  
octobre ?

```
0.0    0.62  
1.0    0.33  
2.0    0.05  
Name: categ, dtype: float64
```

```
0.0    0.94  
2.0    0.06  
Name: categ, dtype: float64
```



Catégories de  
produits vendues  
en général



Catégories de  
produits vendues  
en octobre

## 2 méthodes

### Imputation

Remplacer les  
valeurs erronées par  
d'autres valeurs,  
valeurs moyennes  
des autres données

### Suppression

Supprime les  
valeurs erronées

# Méthode suivie : suppression

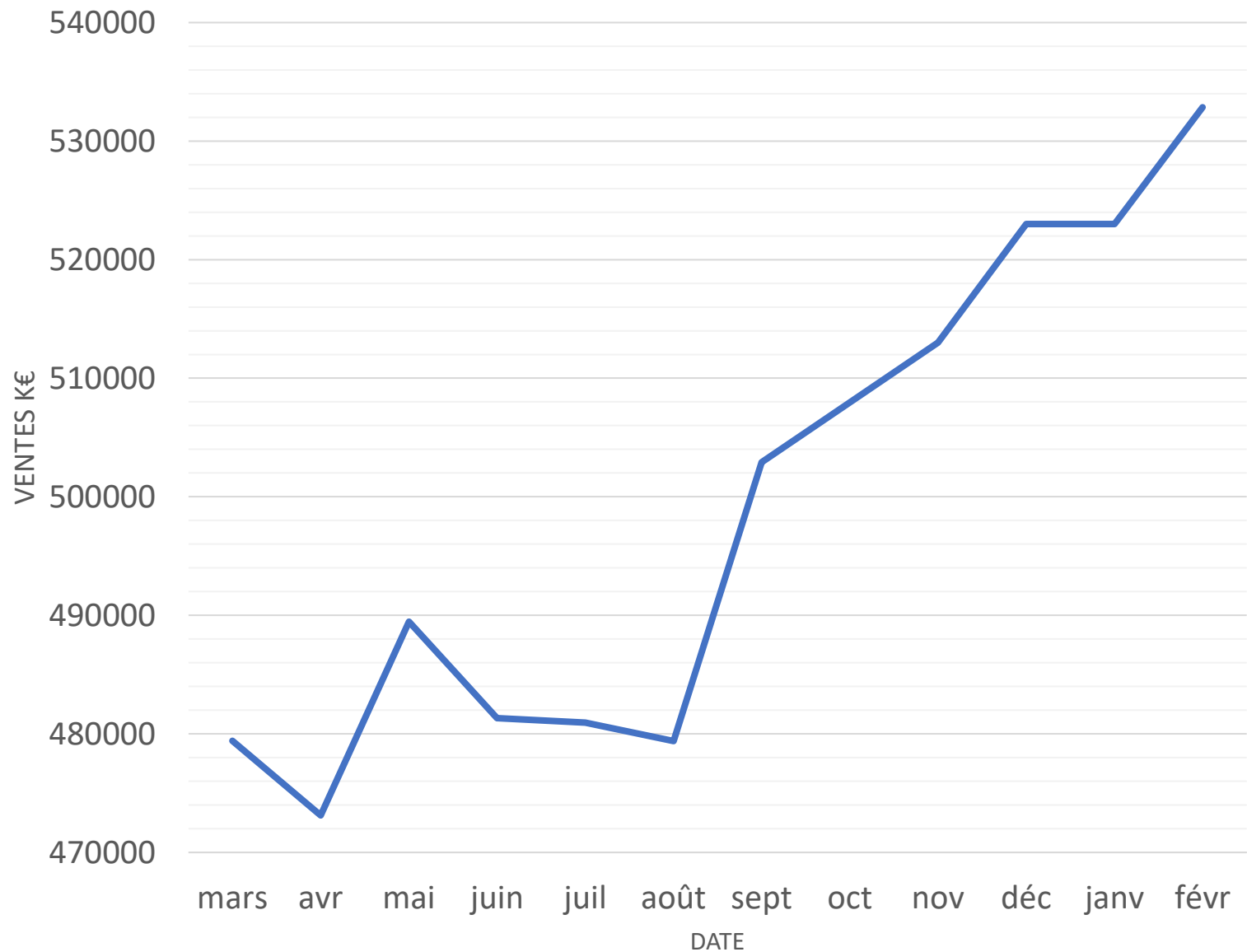
## Fonction de restriction

```
data_final = data_glob_nan[(data_glob_nan['date'] <= '2021-09-30') | (data_glob_nan['date'] >= '2021-11-01')]  
data_final
```

Sélection de toutes les valeurs présentes  
avant octobre 2021

Sélection de toutes les valeurs présentes  
après octobre 2021

## Evolution des ventes

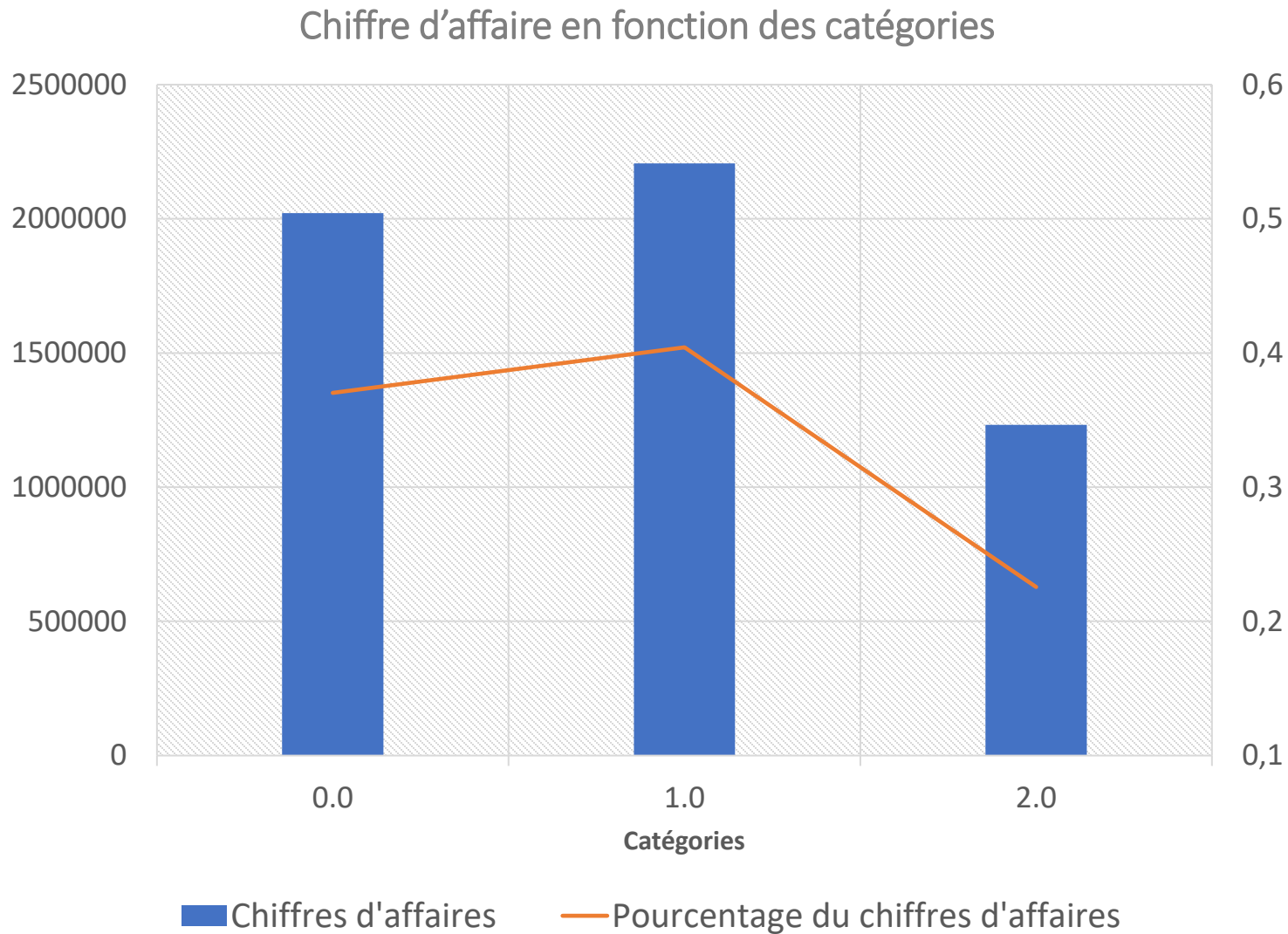


11 %

Pourcentage  
d'augmentation  
du chiffre d'affaire  
entre août et  
février

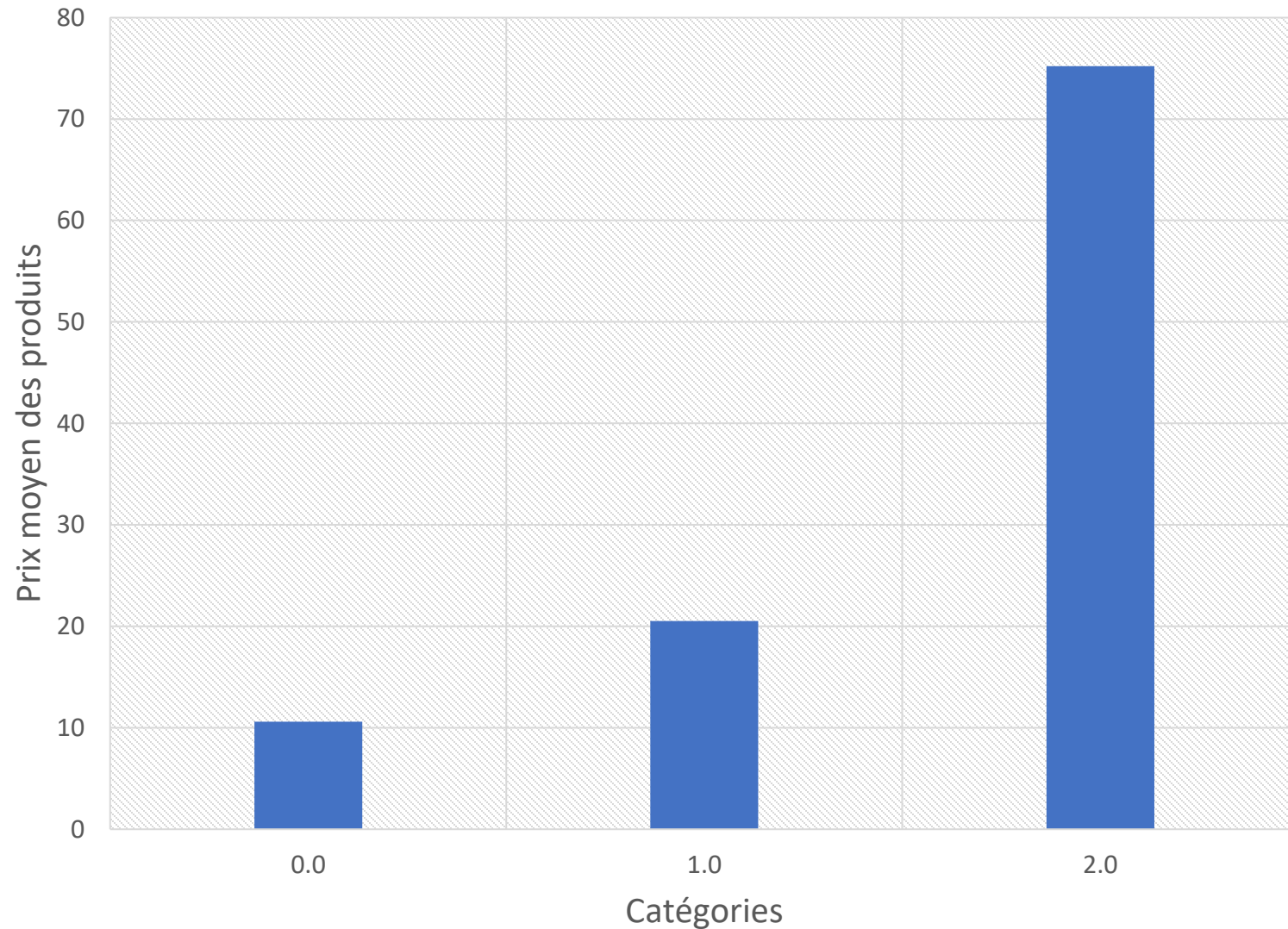
# Mission 2

# Analyse des ventes



Produits de catégories 2 en  
retrait par rapport aux  
autres catégories

## Prix moyen des produits en fonction des catégories

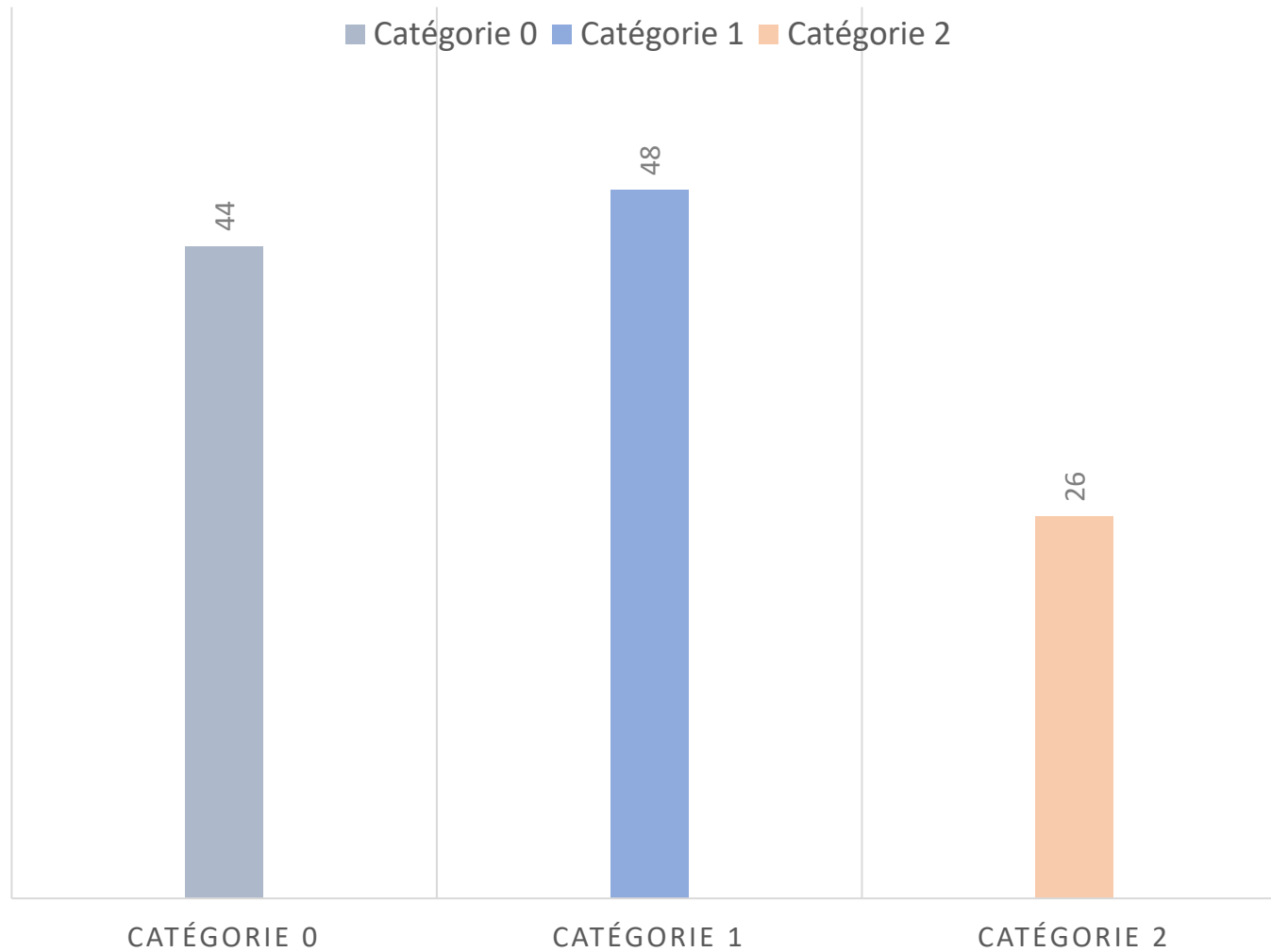


357 %

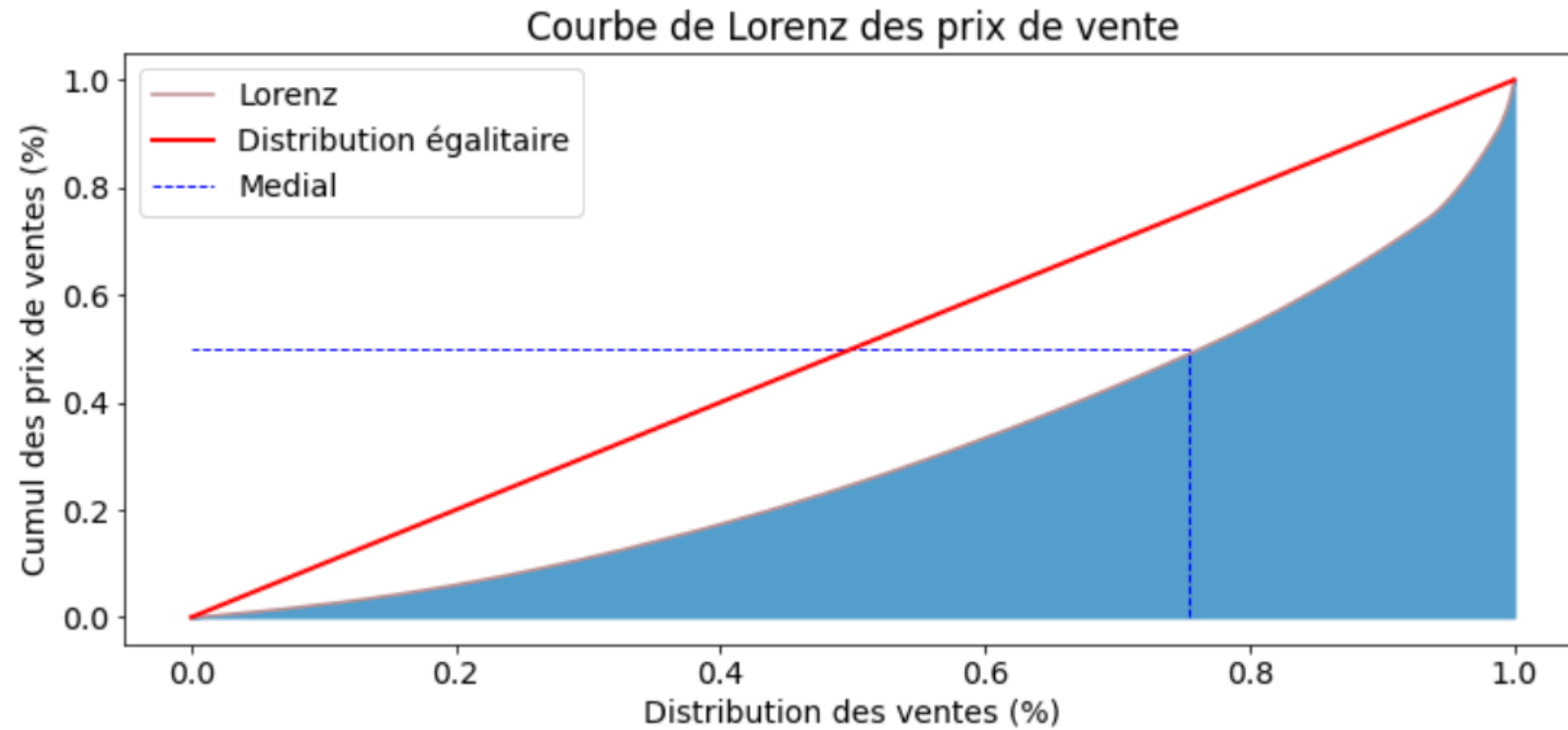
Pourcentage de  
différence de  
prix entre les  
produits de  
catégorie 1 et 2



## ÂGE MOYEN DES CLIENTS EN FONCTION DES CATEGORIES DE PRODUITS ACHETÉS



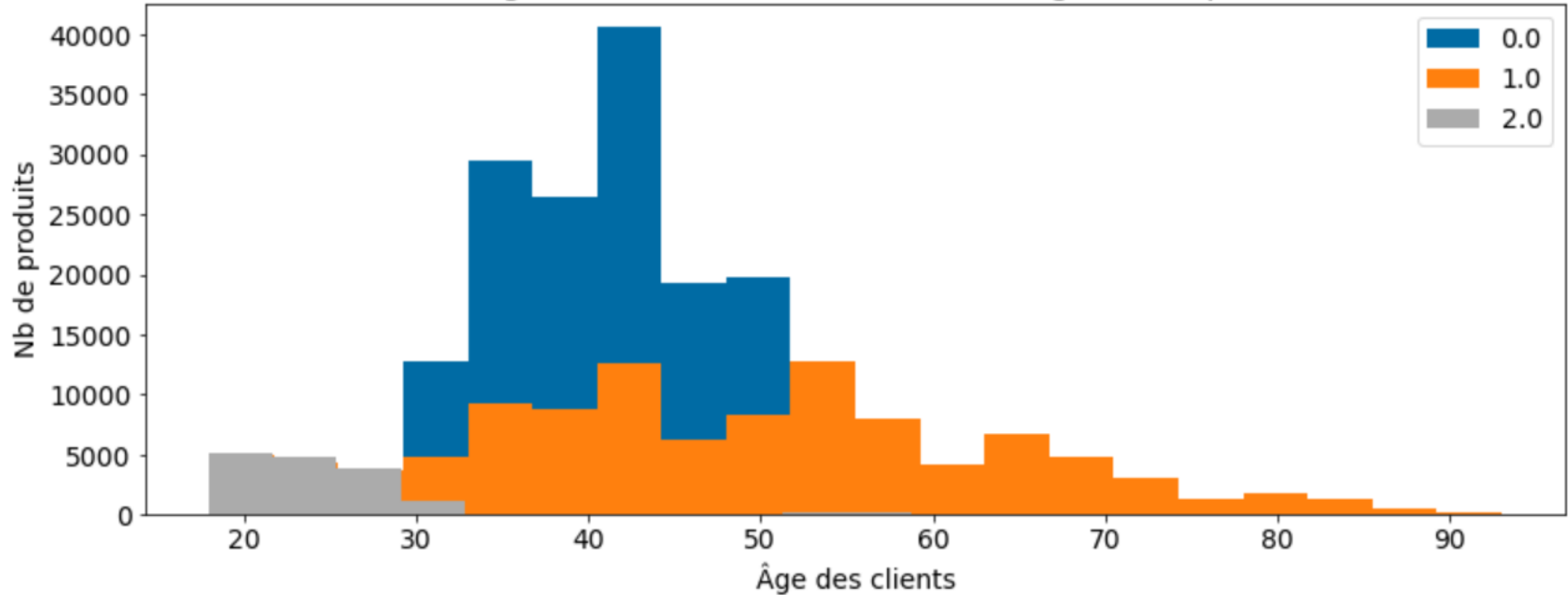
Les produits se  
différencient des autres  
produits en s'adressant à  
une clientèle beaucoup  
plus jeune



0.39

Indice de Gini

Distribution des âge des clients en fonction des catégories de produits achetées



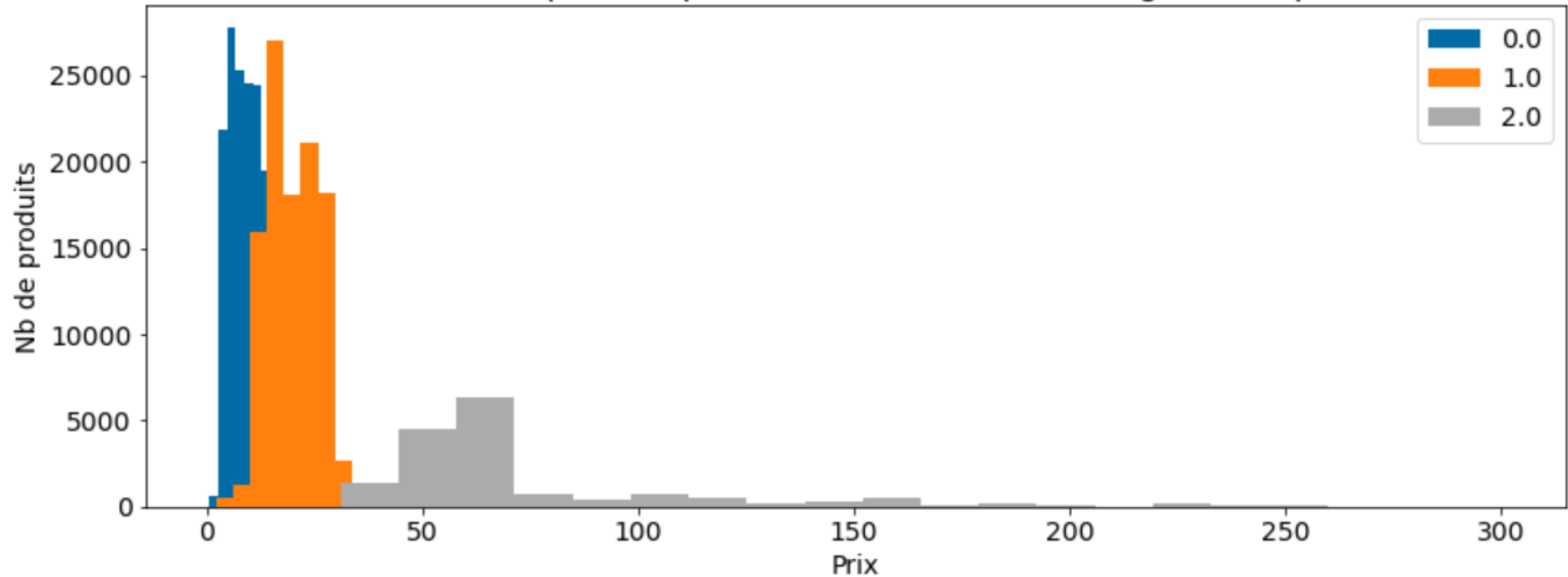
18 - 32

Tranche d'âge  
cible des produits  
de catégorie 2

30 - 51

Tranche d'âge la plus  
touchée des produits  
de catégorie 0

Distribution des prix des produits en fonction des catégories de produits



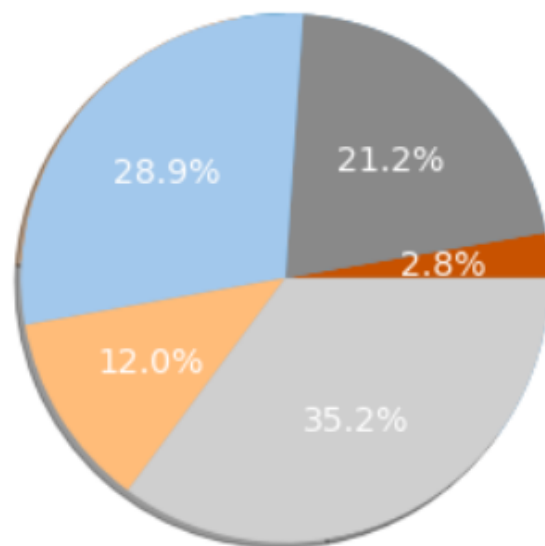
- de 2500  
produits vendus

De catégories 2 entre  
mars 2019 et février 2020

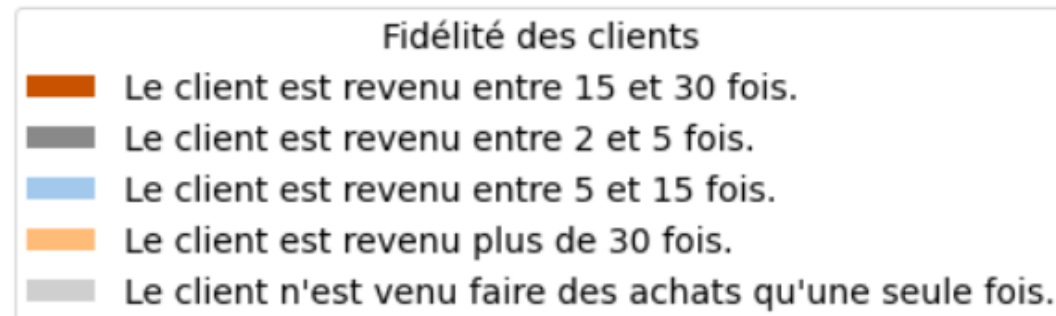
+ de  
40 €

Tranche de prix des  
produits de catégories 2

## Répartition des clients en fonction de leur fidélité



## Pour l'achat de produits de catégories 2 :



35 %

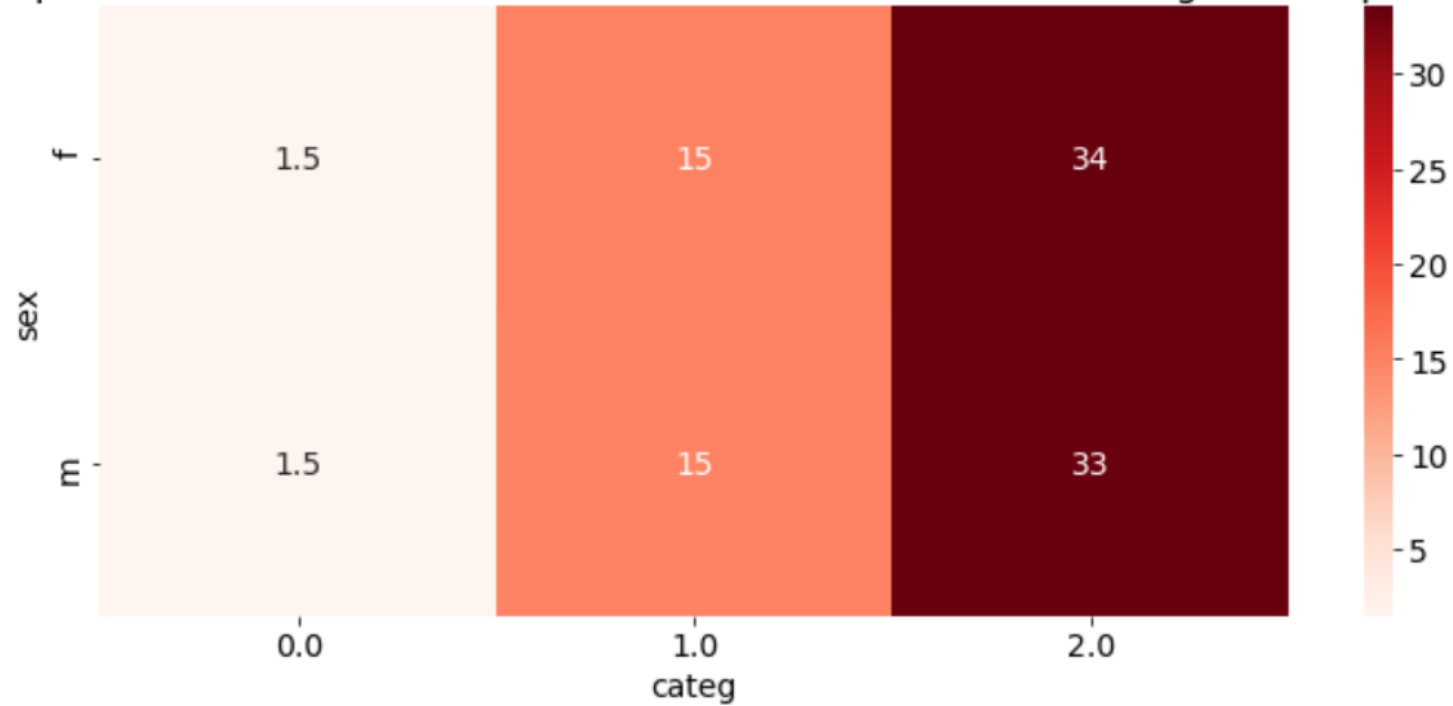
Des clients n'ont acheté qu'une seule fois des produits de catégorie 2

4 %

Environ des clients pour les 2 autres catégories

# Mission 3

Heat map montrant la corrélation entre le sexe des clients et les catégories de produits



2

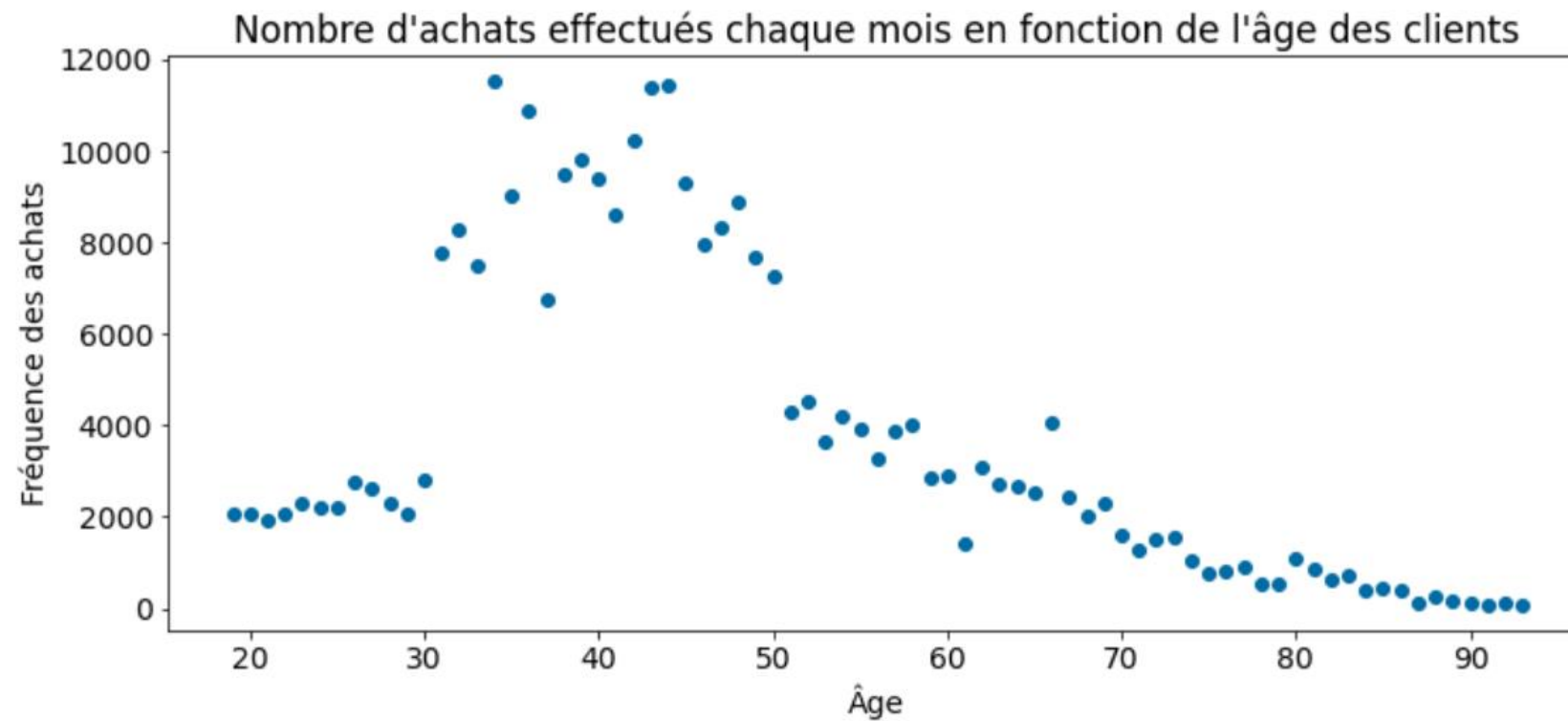
Degré de liberté

79.8

Test du Khi-2

7.4

Valeur critique  
pour un risque  
de 0.025 %



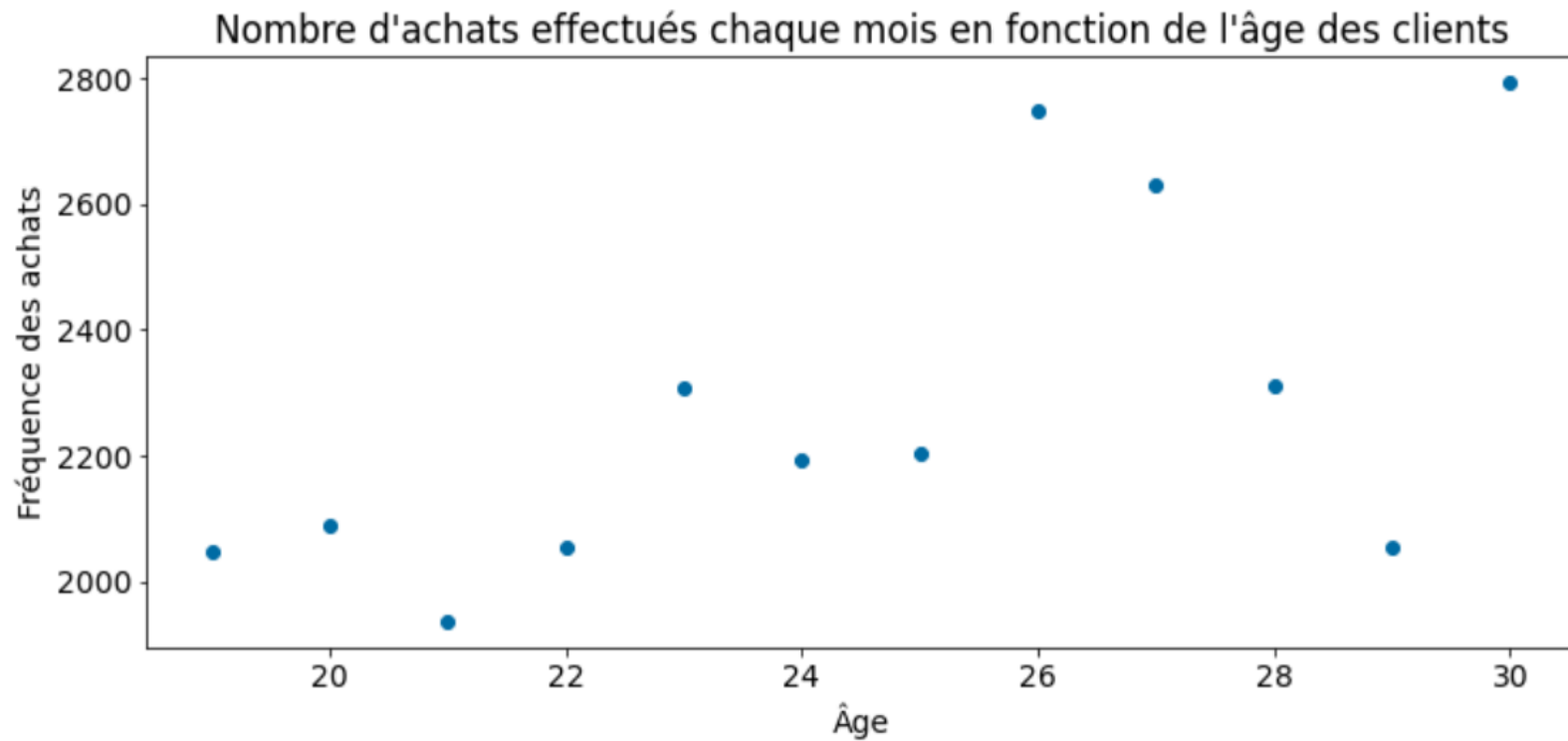
- 42153

Covariance empirique

- 0.6

Coefficient de  
corrélation de Pearson



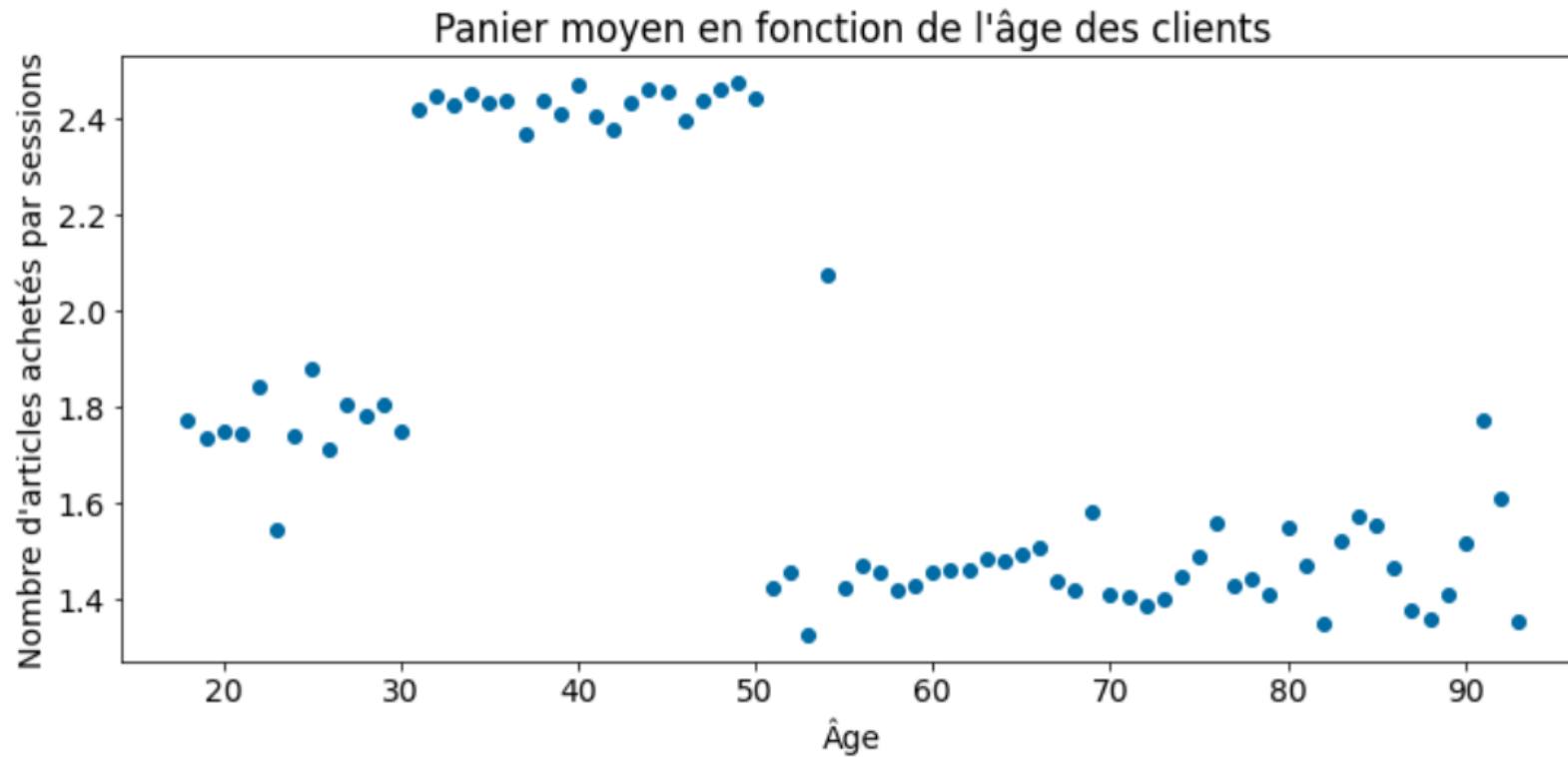


614

Covariance empirique de  
la tranche des 30-50 ans

0.6

Coefficient de corrélation de  
Pearson  
de la tranche des 30-50 ans



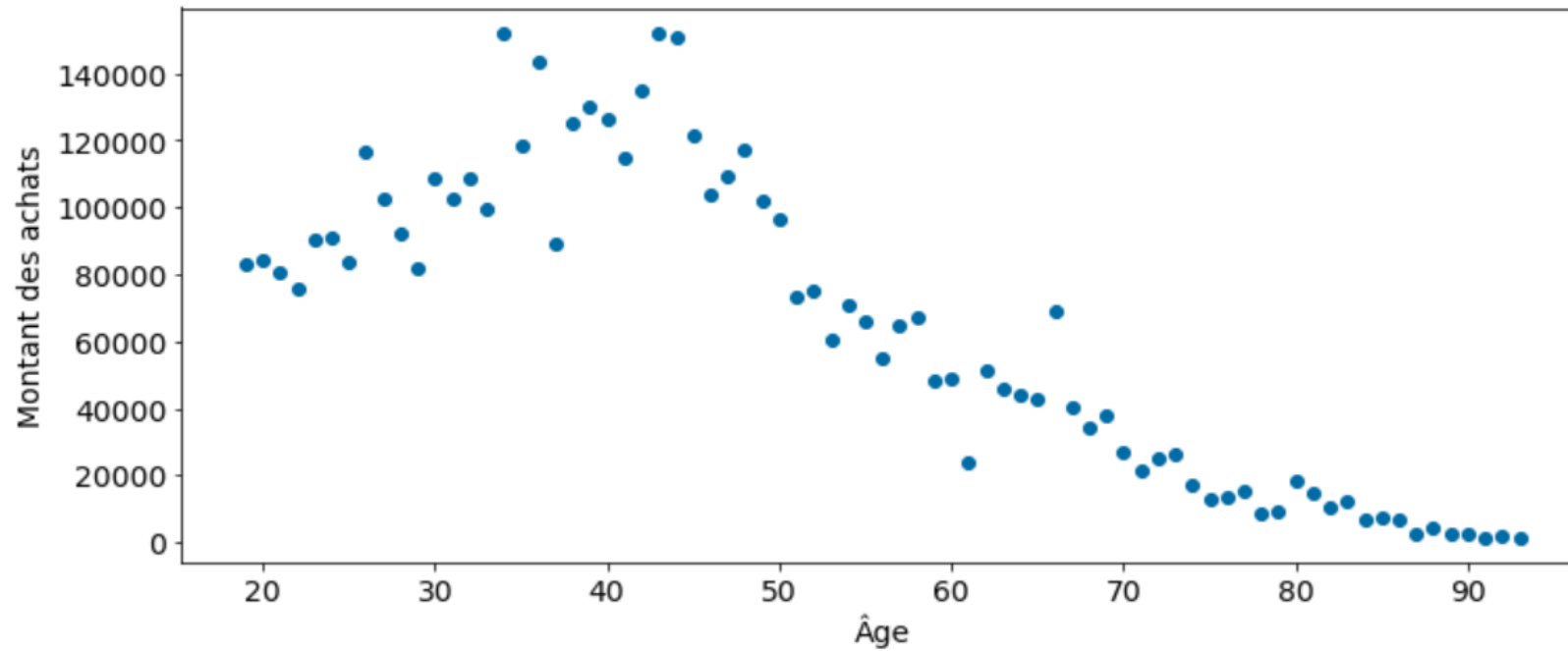
- 5.3

Covariance empirique

- 0.6

Coefficient de corrélation  
de Pearson

Montant global des dépenses en fonction de l'âge du client



- 836846

Covariance empirique

- 0.9

Coefficient de  
corrélacion de Pearson

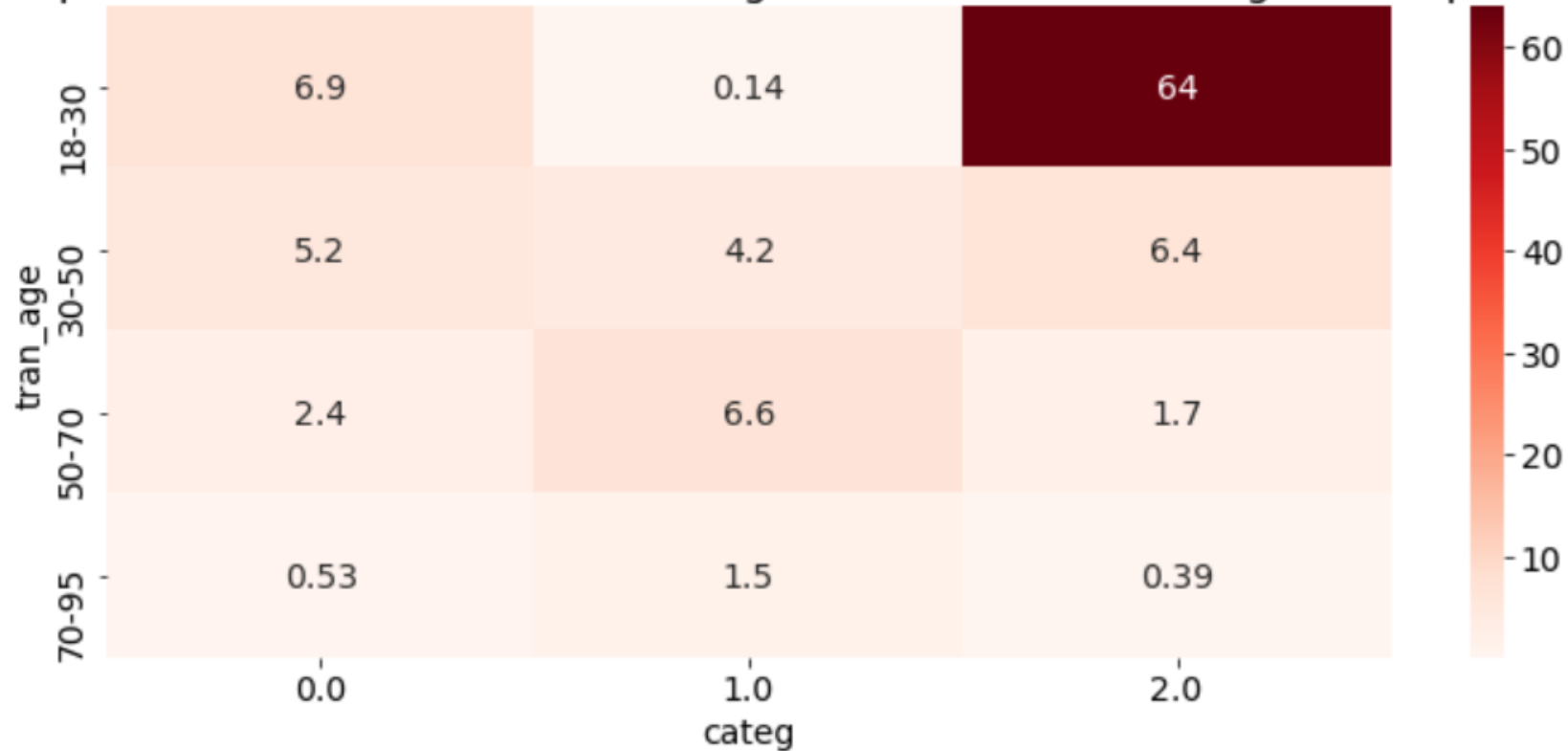
22889

Coefficient de corrélacion  
de Pearson  
de la tranche des 18-30 ans

0.6

Coefficient de corrélacion  
de Pearson  
de la tranche des 18-30 ans

Heat map montrant la corrélation entre l'âge des clients et les catégories de produits



140 722

Test du Khi-2

6

Degré de liberté

14.45

Valeur critique  
pour un risque  
de 0.025 %

# Conclusion

Hypothèse n°1 :

**Attirer davantage  
les jeunes**

Hypothèse n°2 :

**Augmenter nos  
ventes**

Hypothèse n°3 :

**Accroître la  
fidélité des  
clients**

Merci pour votre attention

Au revoir !