

Rester Livre

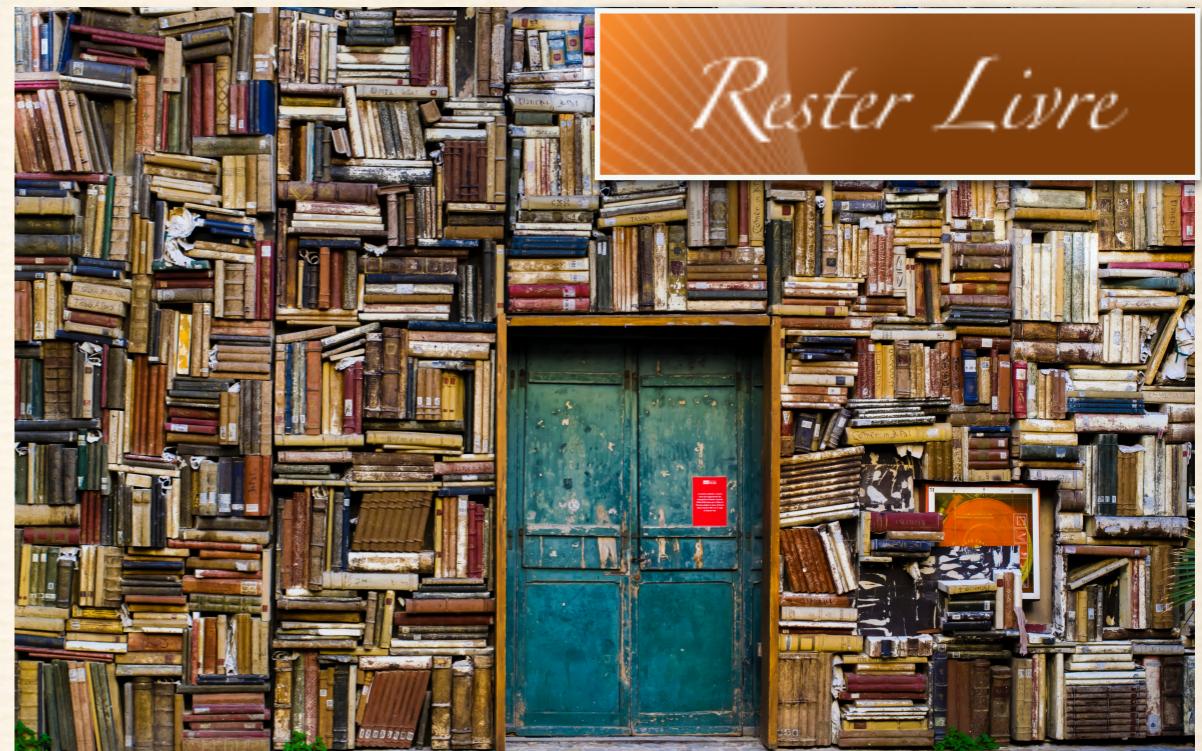


Analyse des ventes

Thibault SCHMIT

Rester livre

- ❖ **Nom** : Rester livre
- ❖ **Forme** : SARL
- ❖ **Date de création** : 2004
- ❖ **Effectifs** : 467 salariés
- ❖ **Clients** : Particuliers et collectivités territoriales.
- ❖ **Canaux de distribution** : Boutiques & Site
- ❖ **Activités** : Ventes de produits culturels
- ❖ **Concurrents** : Amazon, Réseaux de libraires indépendants



Missions

1. Retraitemet des données
2. Analyse descriptive des ventes
3. Analyse des corrélations

Retraitemen~~t~~ des données

- ❖ Traitements des valeurs manquantes
- ❖ Traitements des outliers (Valeurs aberrantes ou valeurs atypiques)
- ❖ Traitements des doublons

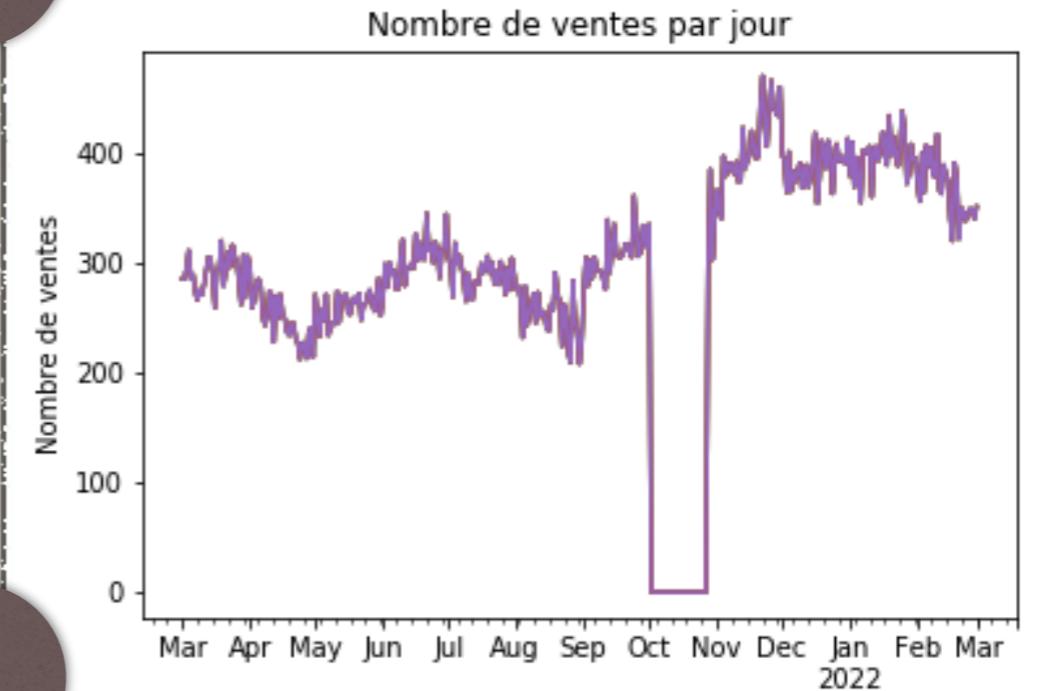
Traitements de valeurs manquantes

Oubli d'individus (Ex : Mise à l'écart du mois d'octobre)

Imputation de valeurs
(Imputation par la moyenne du produit 1_372)

```
df_products_retraite[df_products_retraite['id_prod']=='1_372'] # Ce produit n'existe pas dans df_product mais il existe dans df_add_product_1_372 = pd.DataFrame([{'id_prod':'1_372','price': 25 , 'categ':1}])#On impute le prix de ce produit de 25e si
```

```
df_products_retraite_2 =df_products_retraite.append(df_add_product_1_372, ignore_index=True)
df_products_retraite_2.head(10)
```



#Filtrer par date

```
a = df_transactions_retraite.sort_values(by='date')
df_transactions_retraite_2 = a[(a['date'] < '2021-09-30')|(a['date'] > '2021-11-01')]
df_transactions_retraite_2
```

			date	session_id	client_id
id_prod					
327370	T_0	test_2021-03-01 02:30:02.237442		s_0	ct_1
328618	T_0	test_2021-03-01 02:30:02.237416		s_0	ct_1
328706	T_0	test_2021-03-01 02:30:02.237418		s_0	ct_0
328852	T_0	test_2021-03-01 02:30:02.237439		s_0	ct_1
332736	T_0	test_2021-03-01 02:30:02.237431		s_0	ct_0
336158	T_0	test_2021-03-01 02:30:02.237430		s_0	ct_0
336650	T_0	test_2021-03-01 02:30:02.237447		s_0	ct_0
338550	T_0	test_2021-03-01 02:30:02.237421		s_0	ct_1
339325	T_0	test_2021-03-01 02:30:02.237446		s_0	ct_1
340613	T_0	test_2021-03-01 02:30:02.237447		s_0	ct_0
id_prod		object			
date		object			
session_id		object			
client_id		object			

Traitement des outliers

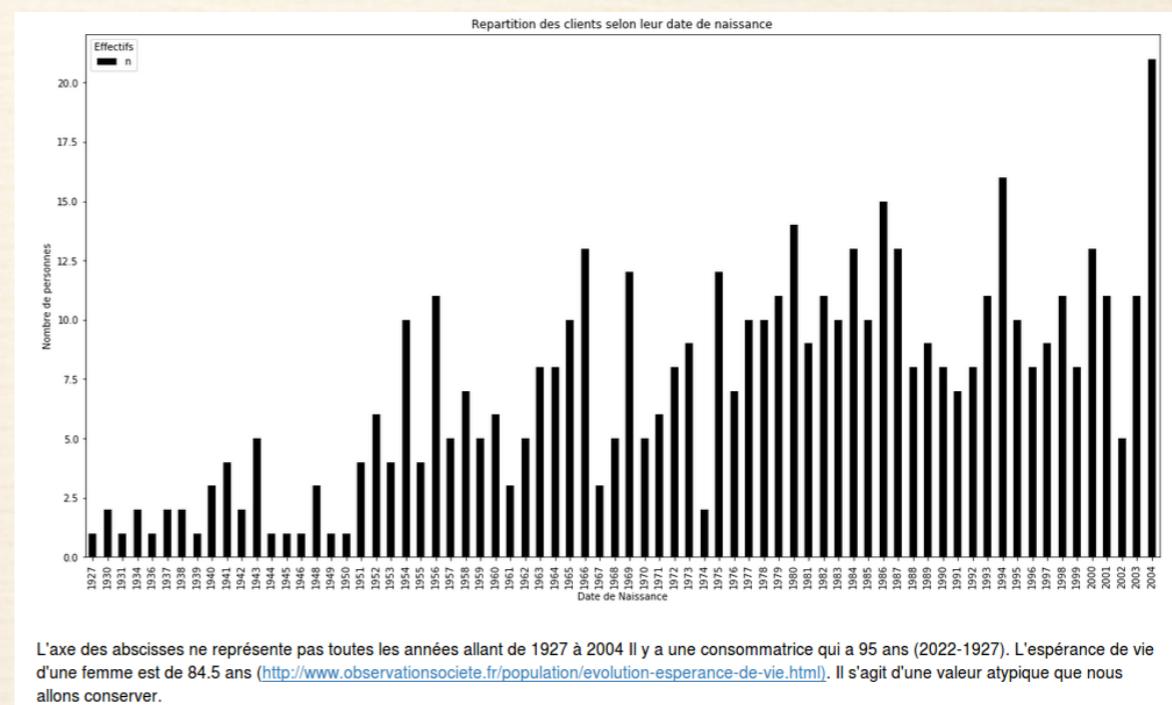
- ❖ Supprimer la valeur aberrante
(Suppression de lignes ‘test’)
- ❖ Supprimer ou conserver la valeur atypique (Conservation d'une date de naissance très ancienne)

```
df_transactions_retraite = df_transactions.reindex(index=order_by_index(df_transactions.index, index_natsorted(df_transactions['date'])), index_natsorted=df_transactions['date'])
print(df_transactions_retraite.tail(10))
df_transactions_retraite = df_transactions_retraite[df_transactions_retraite.id_prod != 'T_0'] #Suppression des lignes test
print(df_transactions_retraite.dtypes)
df_transactions_retraite['date'] = df_transactions_retraite['date'].map(lambda x: str(x)[:-7]) #Suppression des nanosecond
df_transactions_retraite['date'] = pd.to_datetime(df_transactions_retraite['date'], format='%Y-%m-%d %H:%M:%S') #Conversion
print(df_transactions_retraite.dtypes)
df_transactions_retraite = df_transactions_retraite.reset_index(drop=True)
print(df_transactions_retraite.head())

#Filtrer par date
a = df_transactions_retraite.sort_values(by='date')
df_transactions_retraite_2 = a[(a['date'] < '2021-09-30') | (a['date'] > '2021-11-01')]
df_transactions_retraite_2
```

id_prod	date	session_id	client_id
327370	test_2021-03-01 02:30:02.237442	s_0	ct_1
328618	test_2021-03-01 02:30:02.237416	s_0	ct_1
328706	test_2021-03-01 02:30:02.237418	s_0	ct_0
328852	test_2021-03-01 02:30:02.237439	s_0	ct_1
332736	test_2021-03-01 02:30:02.237431	s_0	ct_0
336158	test_2021-03-01 02:30:02.237430	s_0	ct_0
336650	test_2021-03-01 02:30:02.237447	s_0	ct_0
338550	test_2021-03-01 02:30:02.237421	s_0	ct_1
339325	test_2021-03-01 02:30:02.237446	s_0	ct_1
340613	test_2021-03-01 02:30:02.237447	s_0	ct_0

id_prod object
date object
session_id object
client_id object



Traitement des doublons

- Il est ais  de savoir s'il existe des doublons car chacune des tables contient un identifiant de l'individu

V rification de la pr sence de doublons.

```
mfct.verif_doublon(df_customers_retraite, namedf='customers')
Absence de doublon, il n'y a pas de retraitement  faire pour customers

mfct.verif_doublon(df_products_retraite_2, namedf='products')
Absence de doublon, il n'y a pas de retraitement  faire pour products

mfct.verif_doublon(df_transactions_retraite, namedf='transactions')
Absence de doublon, il n'y a pas de retraitement  faire pour transactions
```

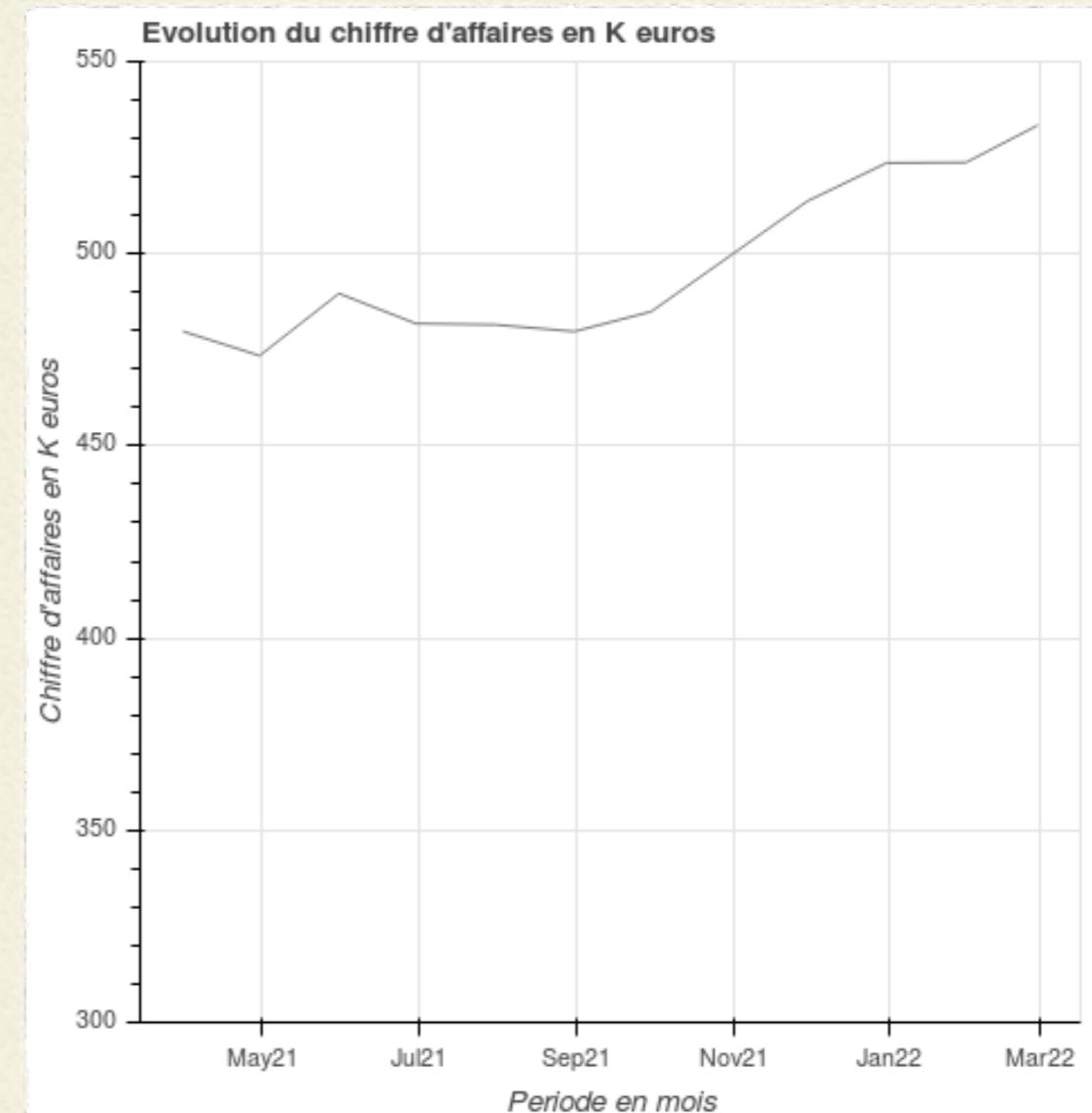
```
def verif_doublon(data, namedf=''):
    """V rifie la pr sence de doublons
    Parameters
    -----
    data : DataFrame
        The first parameter.
    namedf : str
        The second parameter.

    Returns
    -----
    bool
        True if successful, False otherwise.

    """
    data_ss_doublon = data.drop_duplicates()
    if data.shape == data_ss_doublon.shape:
        print("Absence de doublon, il n'y a pas de retraitement  faire pour {}".format(namedf))
    else:
        print('Suppression des doublons ?')
```

Evolution du chiffre d'affaires

- ❖ Moyenne sur la période : 496 K€
- ❖ Le mois d'octobre a été écarté de l'analyse.



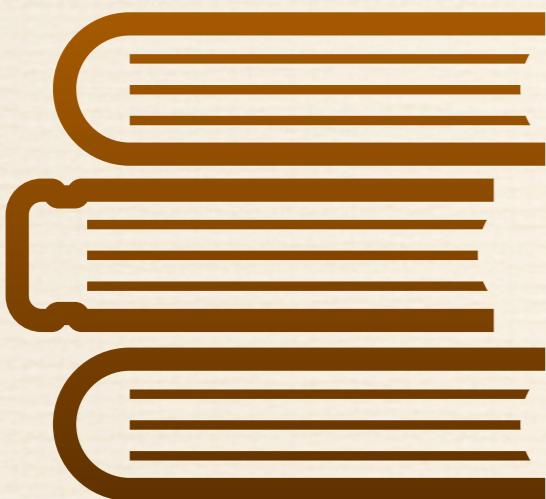
Evolution du nombre de ventes



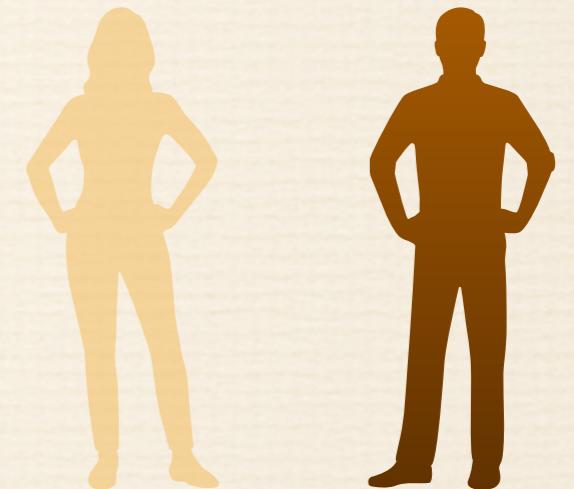
- ❖ Moyenne du nombre de ventes sur la période : 28000
- ❖ Pic à la rentrée et à Noel

Les axes d'analyse des ventes

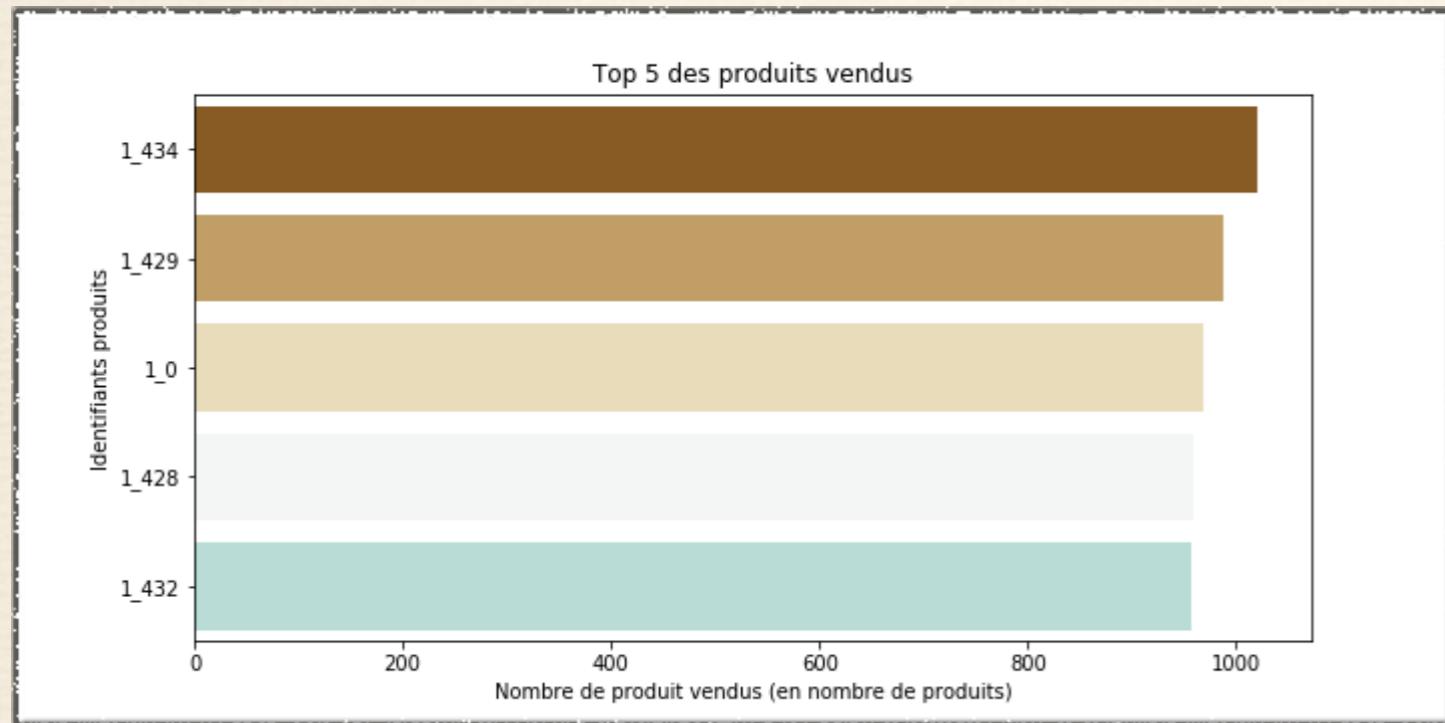
L'axe produit



L'axe client



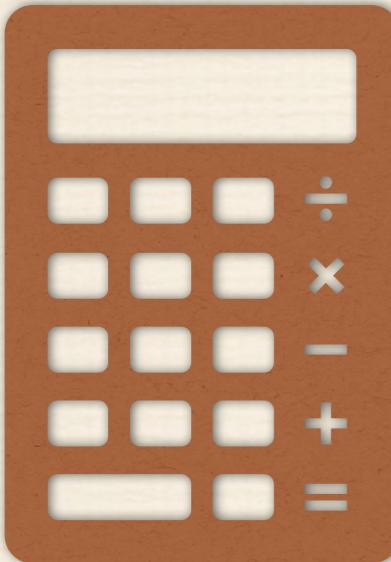
L'axe produit



- ❖ Les deux plus grosses ventes :
 1. Le produit 1_434 : 1020 unités vendues
 2. Le produit 1_429 : 980 unités vendues

L'axe produit

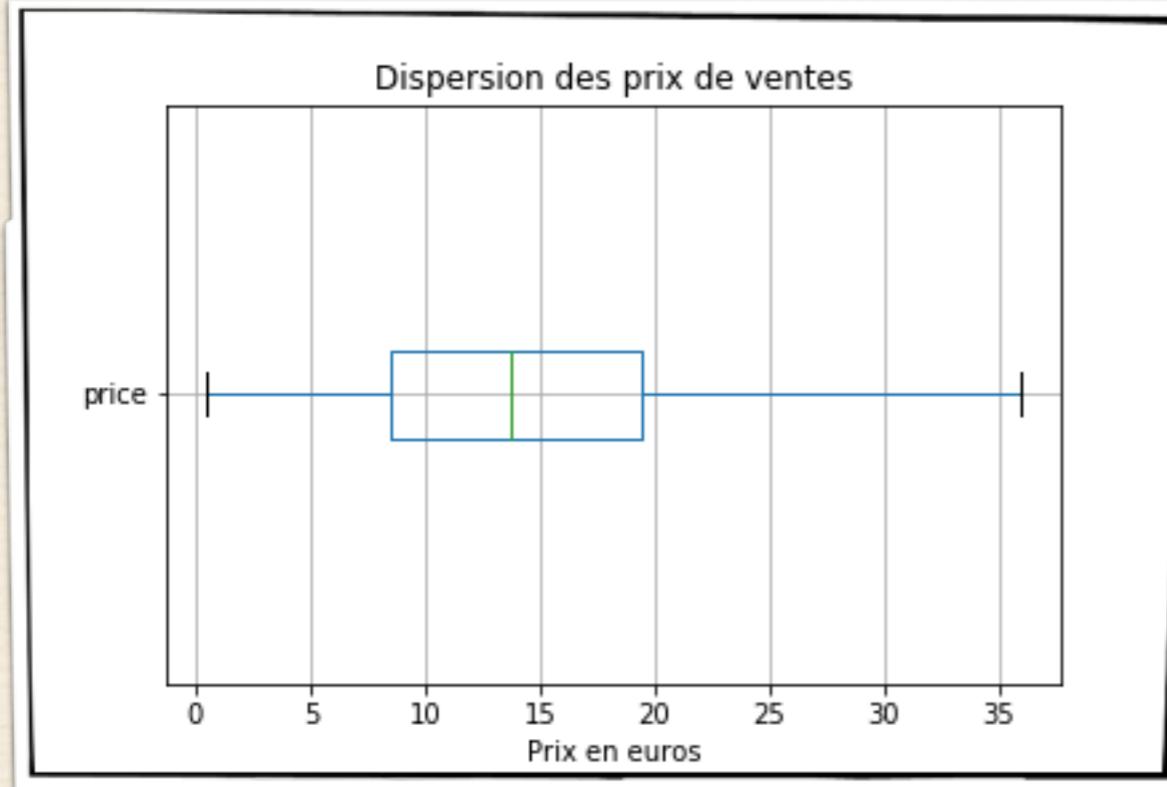
Analyse des prix de ventes



- Prix moyen : 17 €
- Prix modal : 13,86€
- Prix médian : 13,72 €
- Ecart-type : 17,52€ (la plupart des prix sont identiques)

L'axe produit

Analyse des prix de ventes



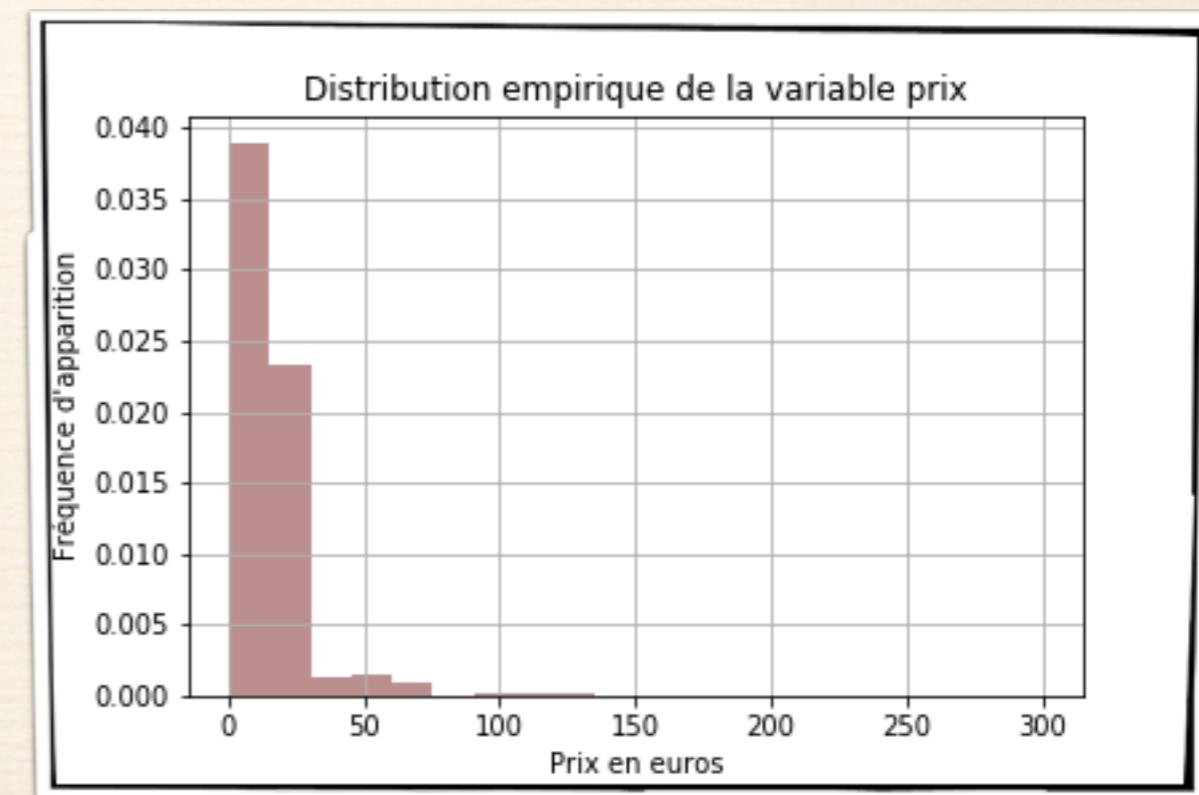
De manière générale, nous vendons des produits ...
... qui ont des prix compris entre 10 et 20 €

Commentaire :

La médiane est 13.72€.

Q1 est égal à 8.54€ et Q3 est égal à 19.48€ .

L'écart inter-quartile est égal à 10.94€ et les bornes sont respectivement de 0.5€ à 35.88€.

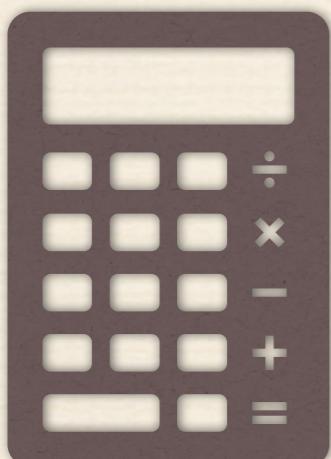


L'axe produit

Analyse des prix de ventes



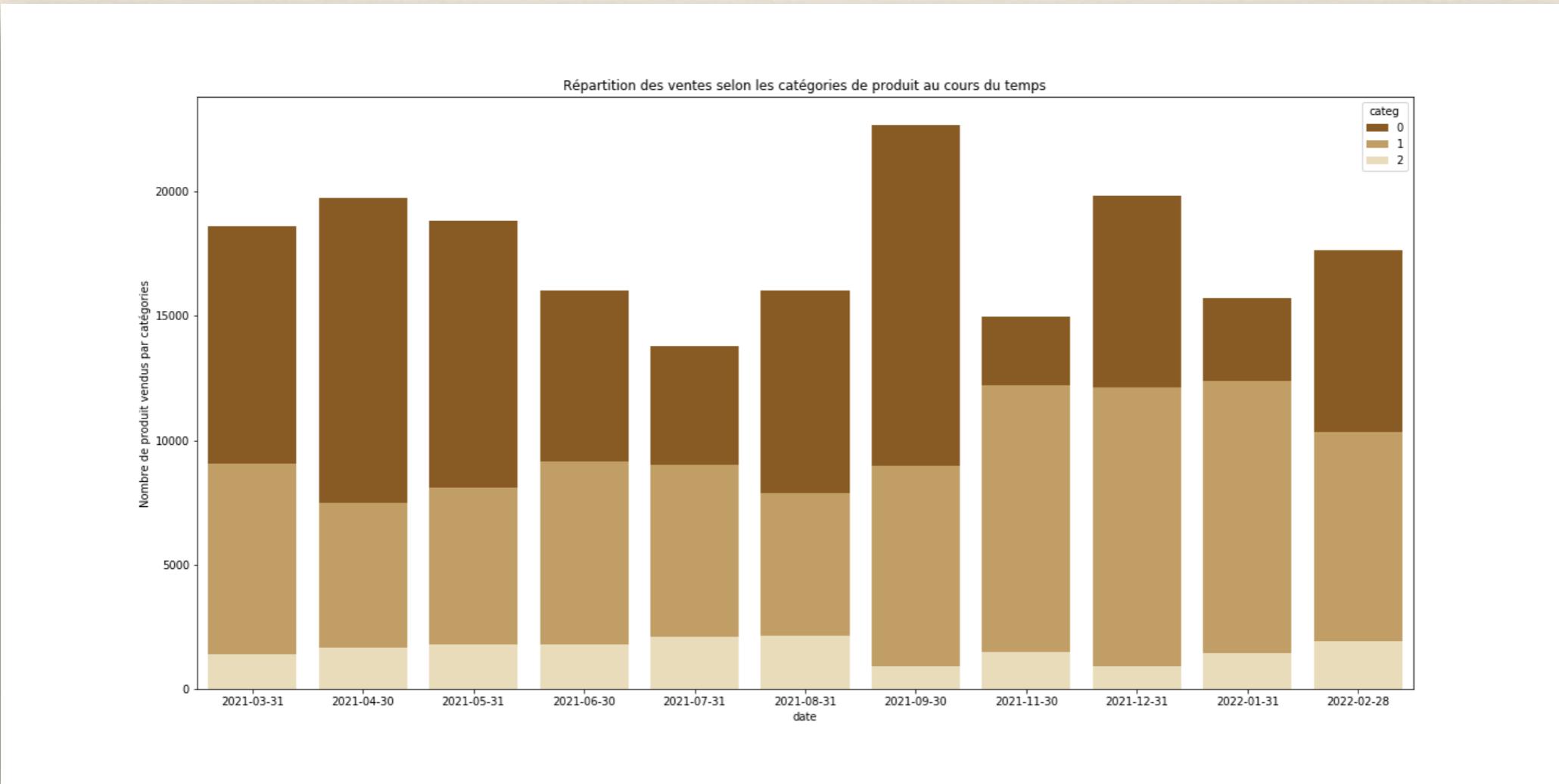
Faible concentration : Les prix sont bien repartis



Indice de Gini : 0,39

L'axe produit

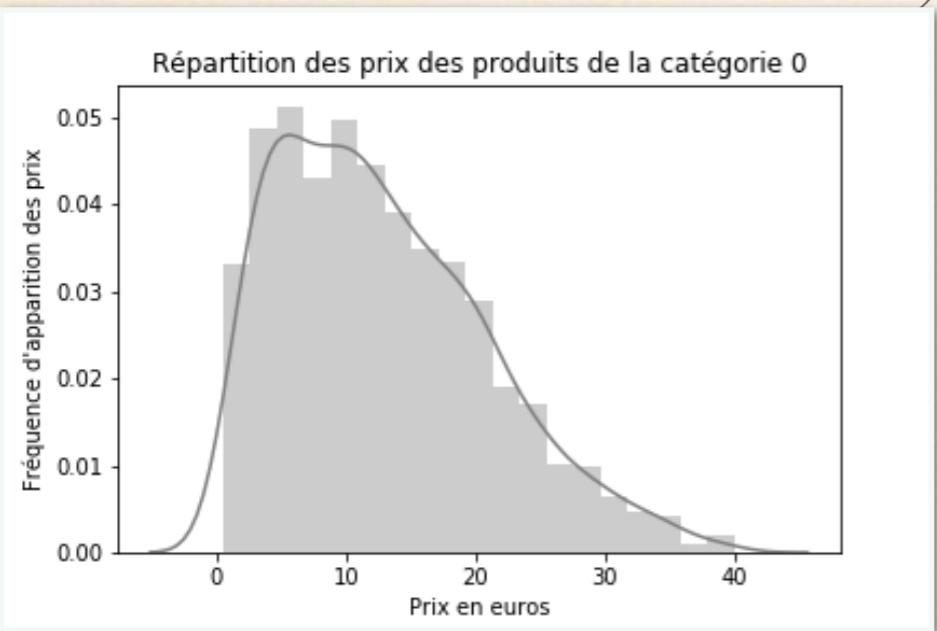
Analyse de la structure des ventes



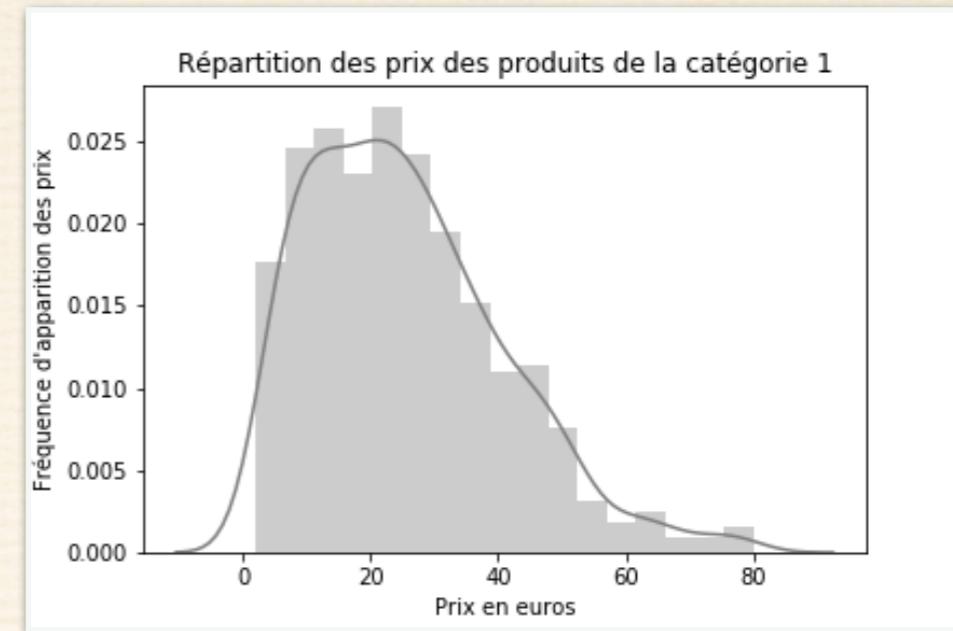
On vend davantage de produits 0 & 1

L'axe produit

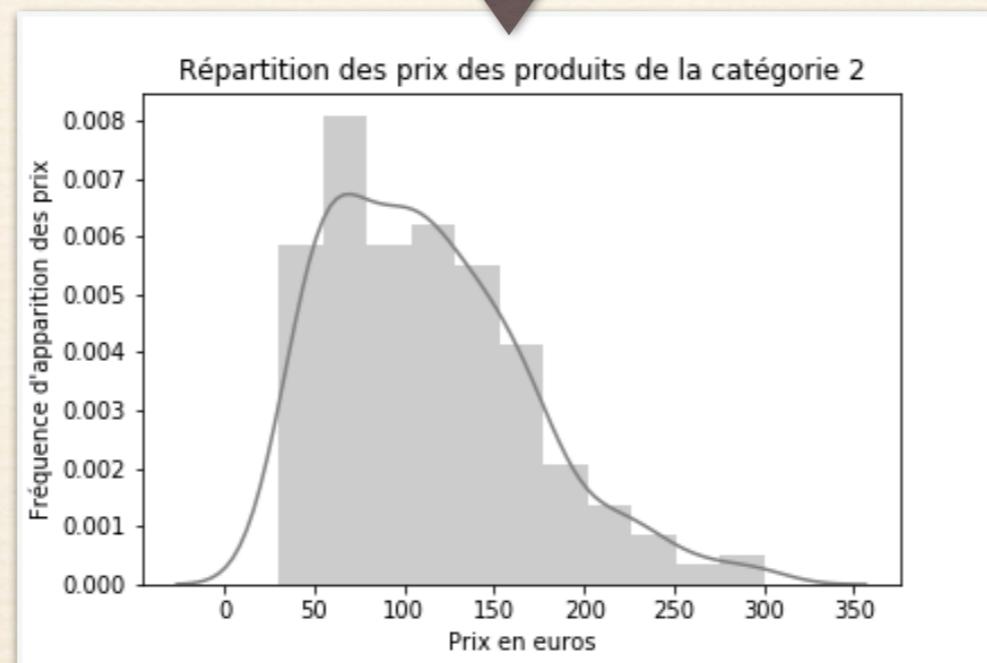
Analyse de la structure des ventes



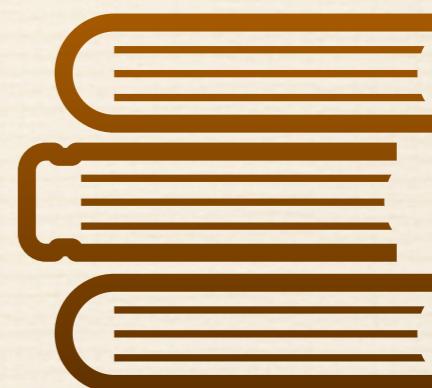
Prix faibles : moyenne 11€



Prix moyens : moyenne 25 €



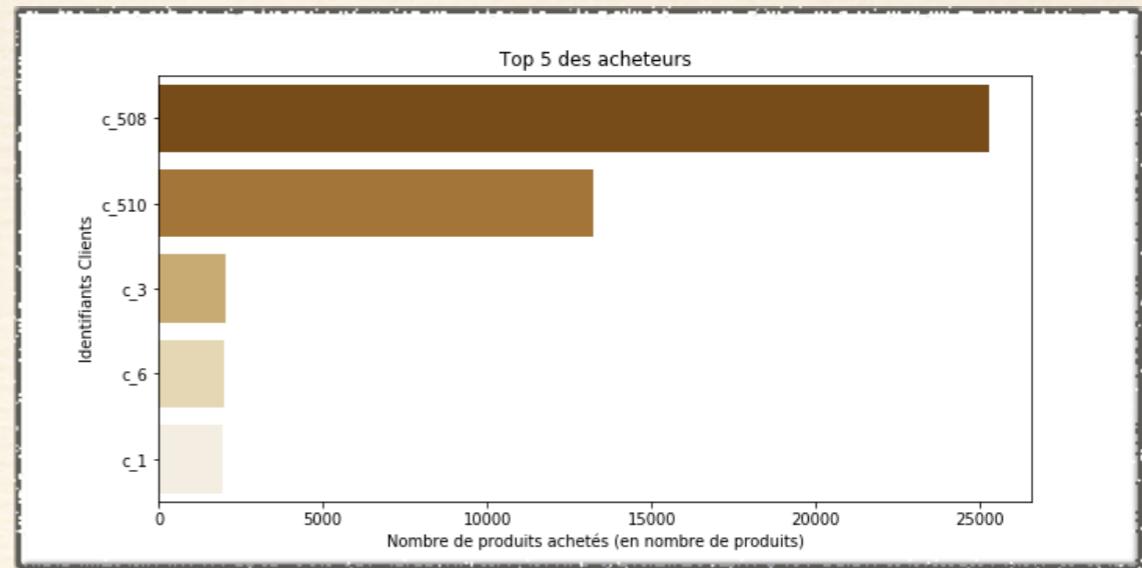
Prix élevés : moyenne 113€



L'axe client

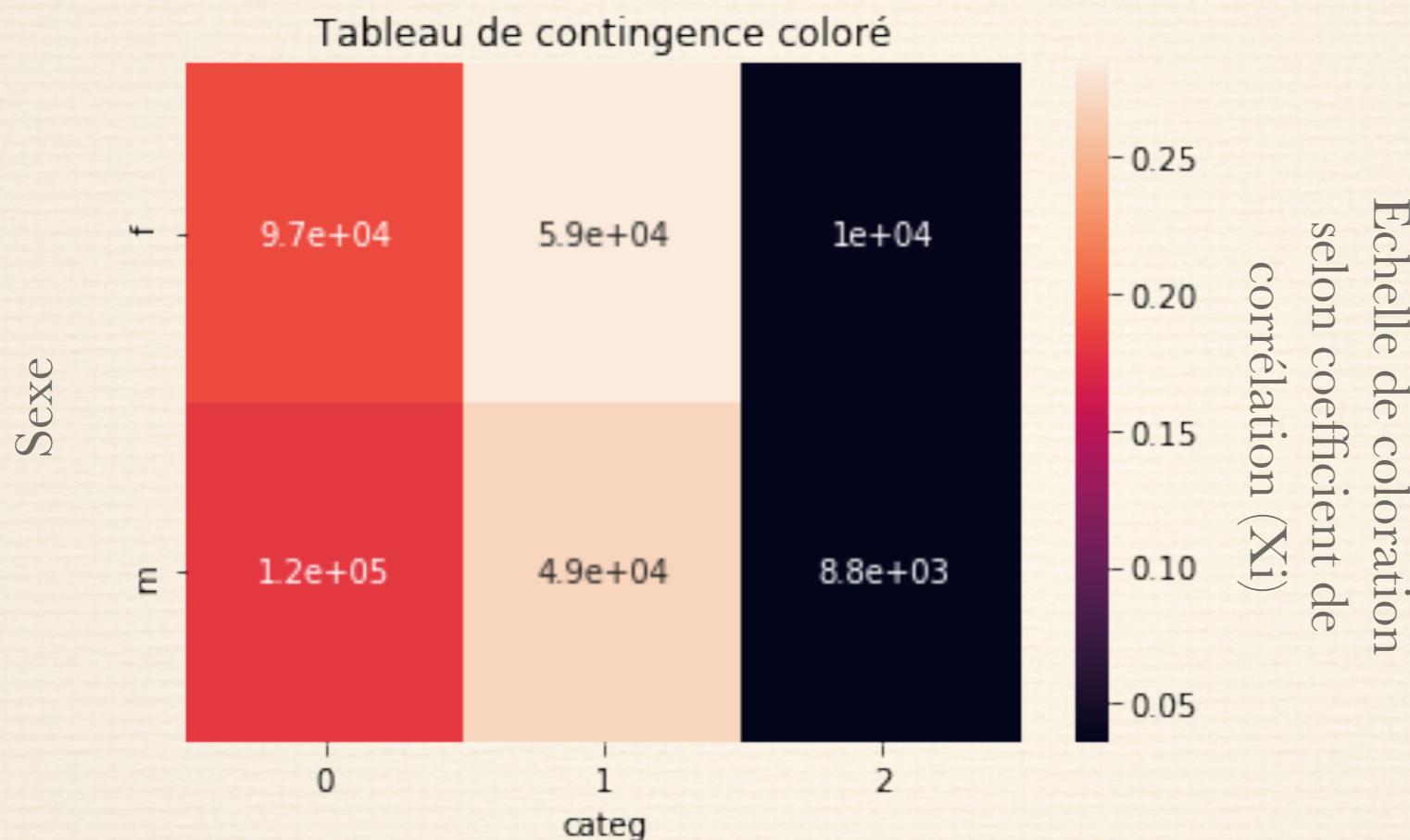
- ❖ Les deux plus gros acheteurs sont :

1. Le client 508 : 25000 achats
2. Le client 510 : 13000 achats



L'axe client

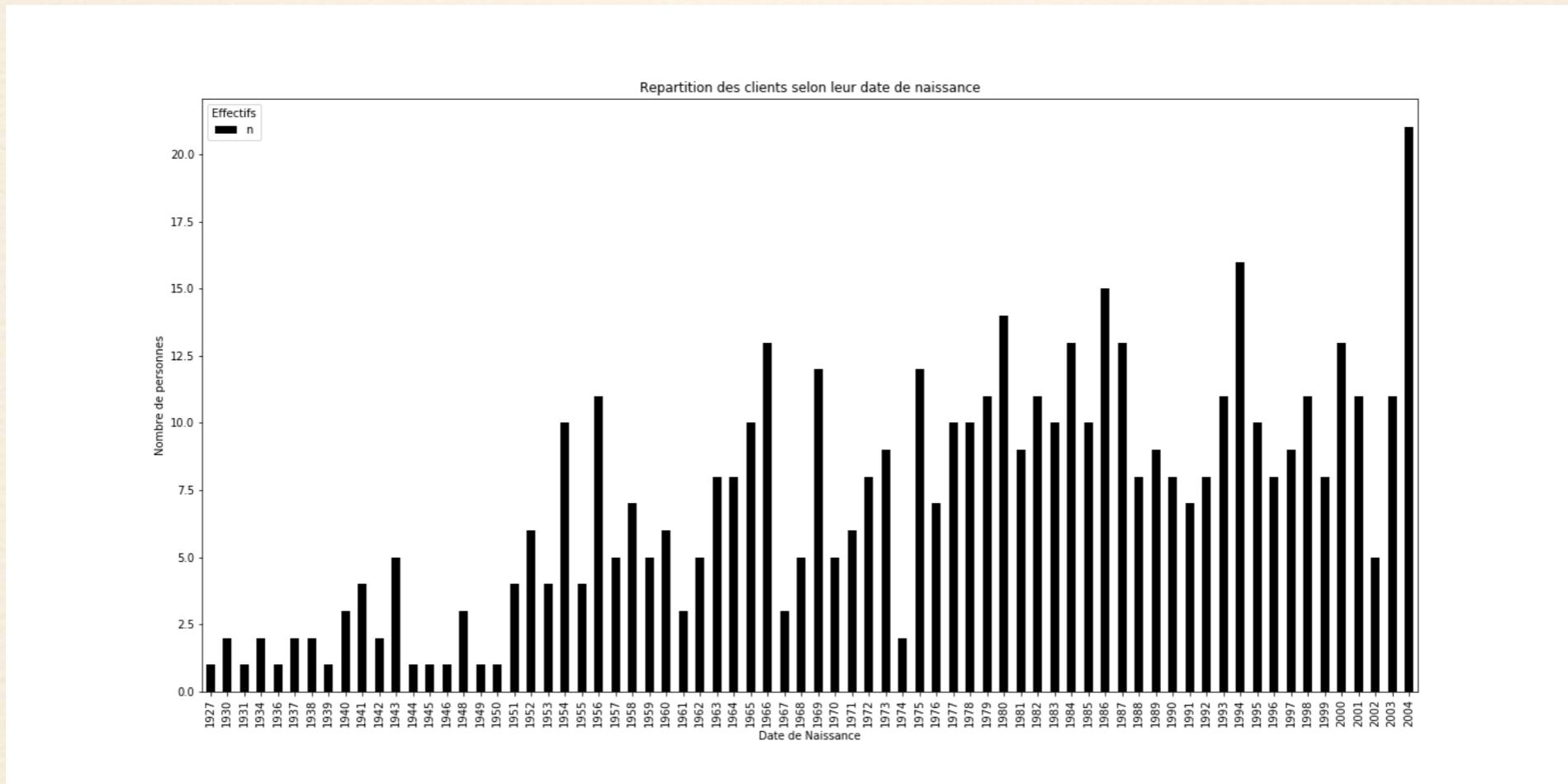
Analyse des ventes selon le sexe des consommateurs



On ne peut pas dire qu'un homme a tendance à acheter plus de produits de catégorie x qu'une femme et inversement.

L'axe client

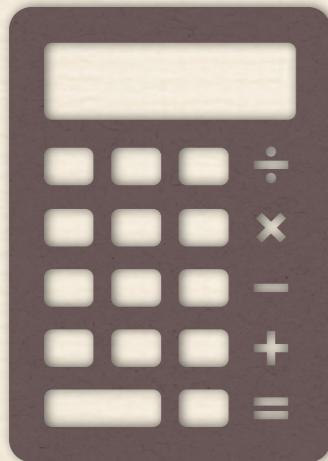
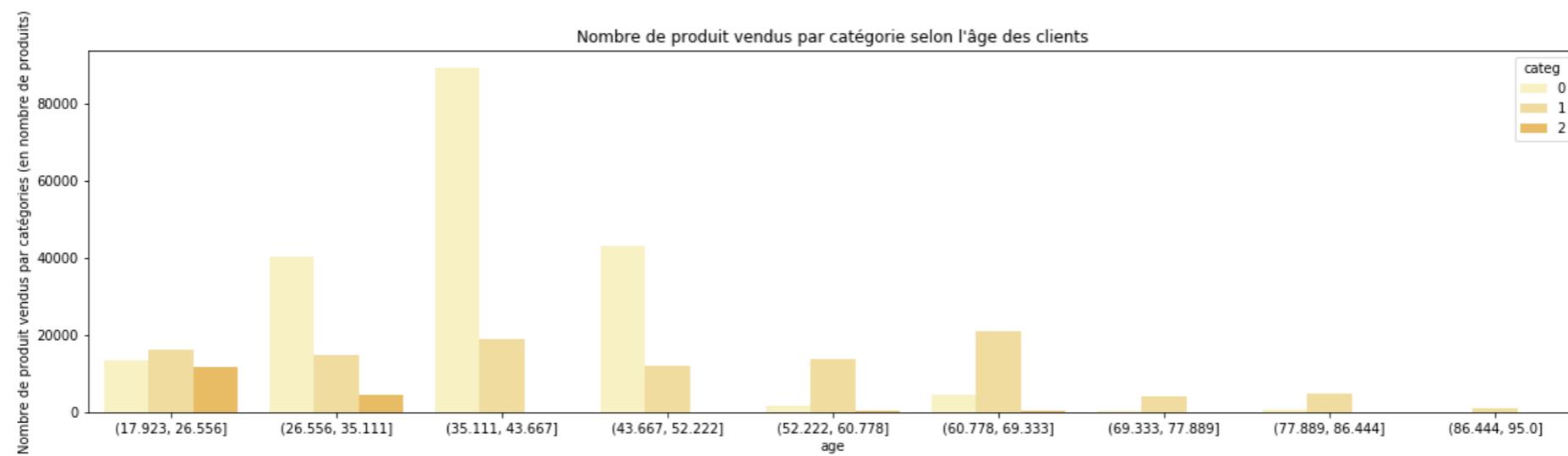
Analyse des ventes selon l'âge des consommateurs



La majorité de nos clients sont jeunes.

L'axe client

Analyse des ventes selon l'âge des consommateurs

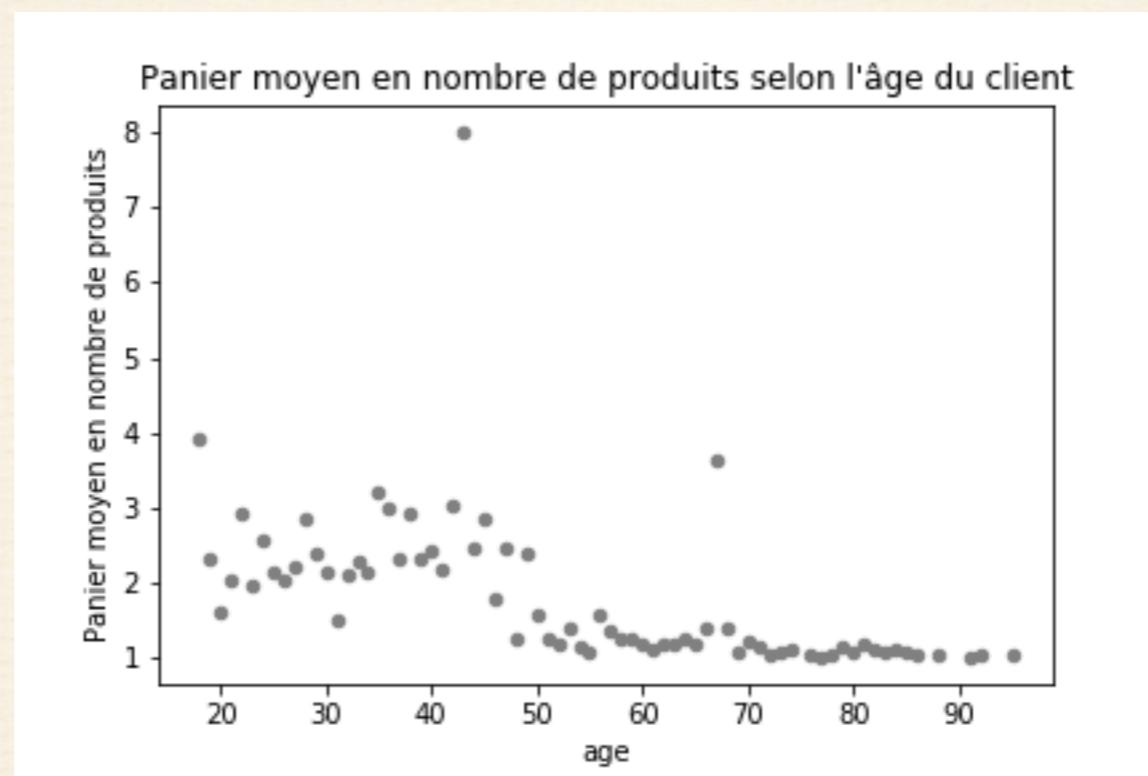


Coefficient de corrélation (eta squared)= 0,35

L'âge des consommateurs et le type de catégorie de produits achetés seraient très peu corrélés.

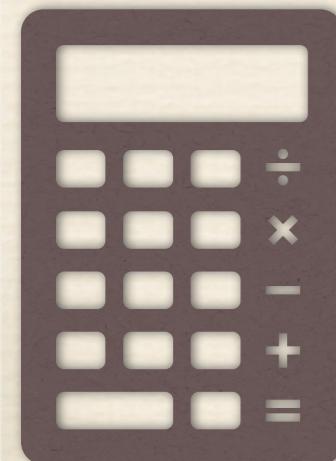
L'axe client

Analyse des ventes selon l'âge des consommateurs



Coefficient de corrélation de Pearson : -0,55

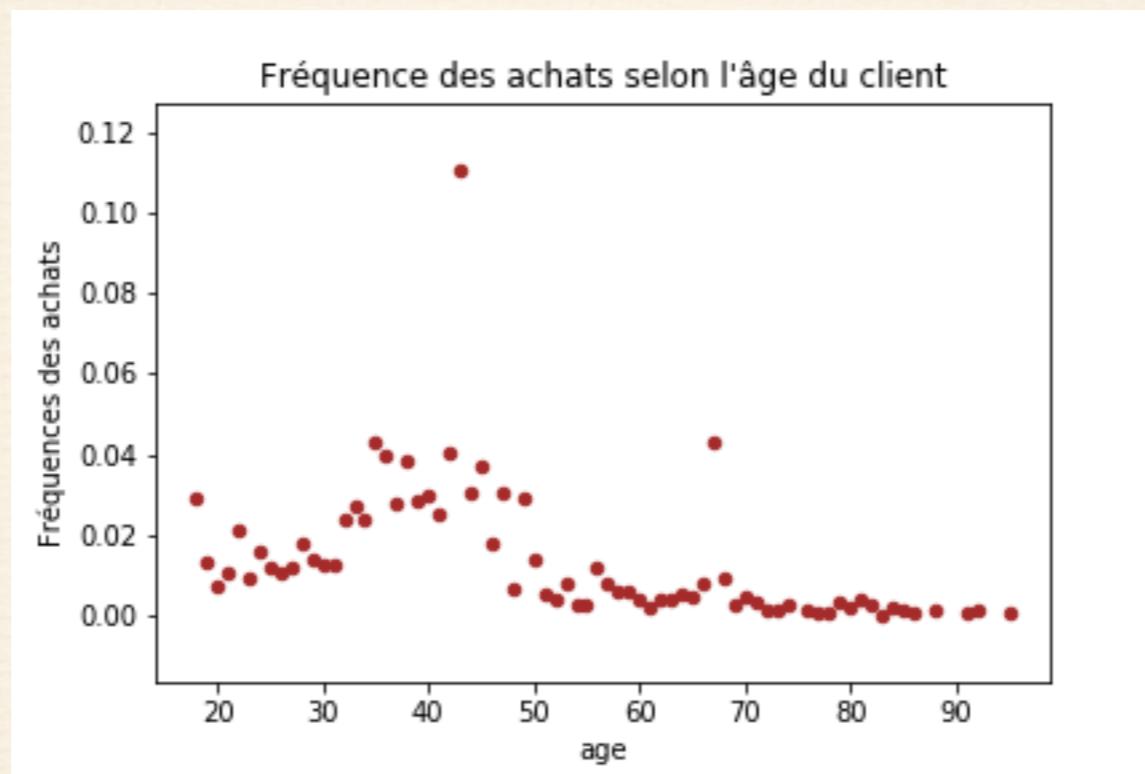
P-valeur : 0,000006%



Il est possible que plus les consommateurs sont âgés plus le panier moyen en nombre de produit soit faible

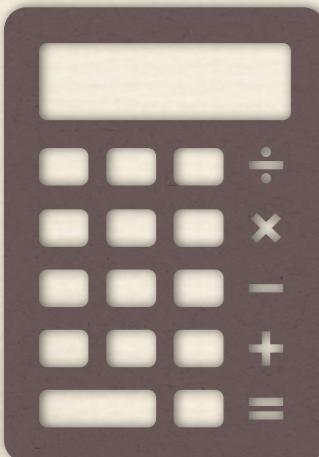
L'axe client

Analyse des ventes selon l'âge des consommateurs



Coefficient de corrélation de Pearson : -0,47

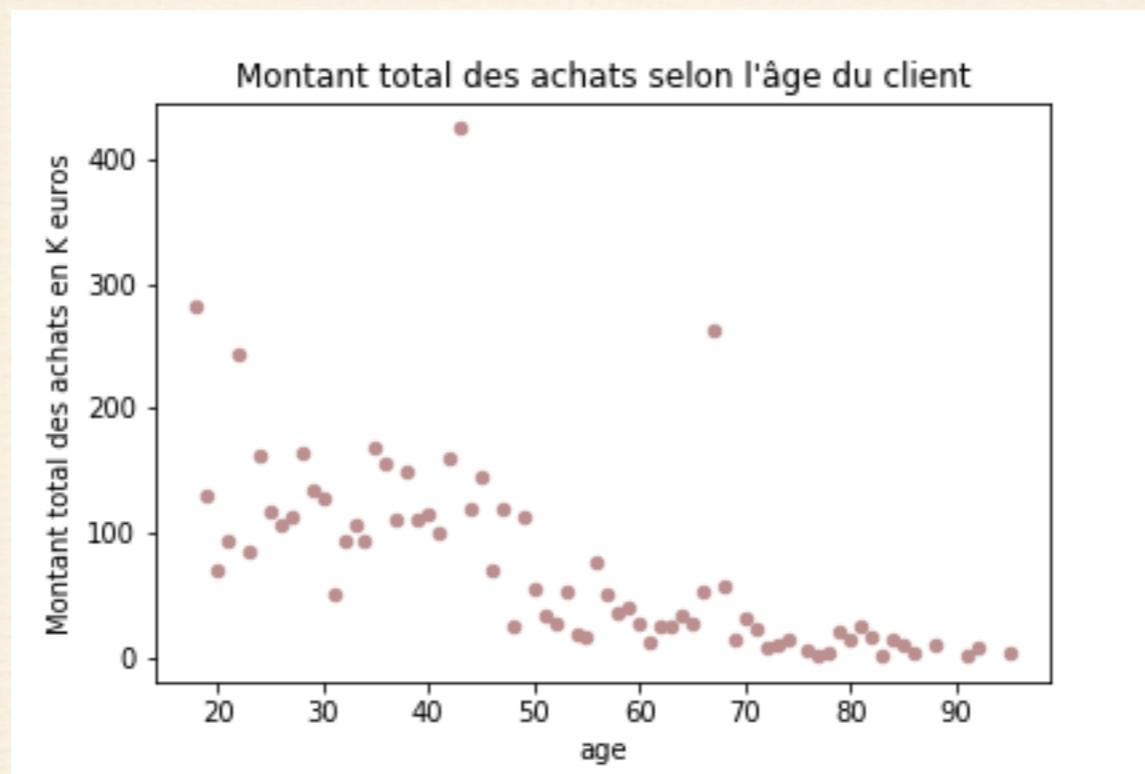
P-valeur : 0,00306%



Les consommateurs plus âgés auraient tendance à avoir une fréquence d'achats plus faible.

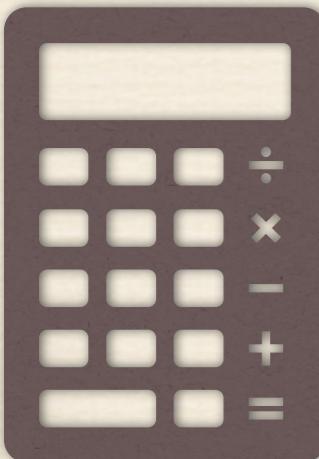
L'axe client

Analyse des ventes selon l'âge des consommateurs



Coefficient de corrélation de Pearson : -0,65

Si les consommateurs sont âgés, alors il est fortement possible que le montant de leurs achats soit faible.



Conclusion

Propositions de stratégies

- ❖ Développer de nouveaux canaux de distribution, notamment physique pour séduire le segment de consommateurs plus âgés.
- ❖ Continuer à développer le site pour faciliter davantage les transactions du segment des personnes plus jeunes.

Autres axes

Propositions de récolte de données

L'axe des canaux de ventes



L'axe représentant des ventes

L'axe des marchés

