Projet n°4 : « Analyse des ventes de votre entreprise »

Sommaire

- A/ Détails du nettoyage
- B/ Présentation de l'analyse demandée
- C/ Interprétation des corrélations

Nettoyage

Structure des fichiers

- 3 fichiers CSV :
  - **« Customers »** : informations à propos des clients
  - **« Products »** : informations à propos des produits
  - **« Transactions »** : informations à propos des ventes

Détails des fichiers

« Customers »

- 8623 entrées
- 202,2KB
- 3 attributs :
  - 'client_id' - type: objet
  - 'sex' – type: objet
  - 'birth' – type: int64
- 4491 femmes / 4132 hommes

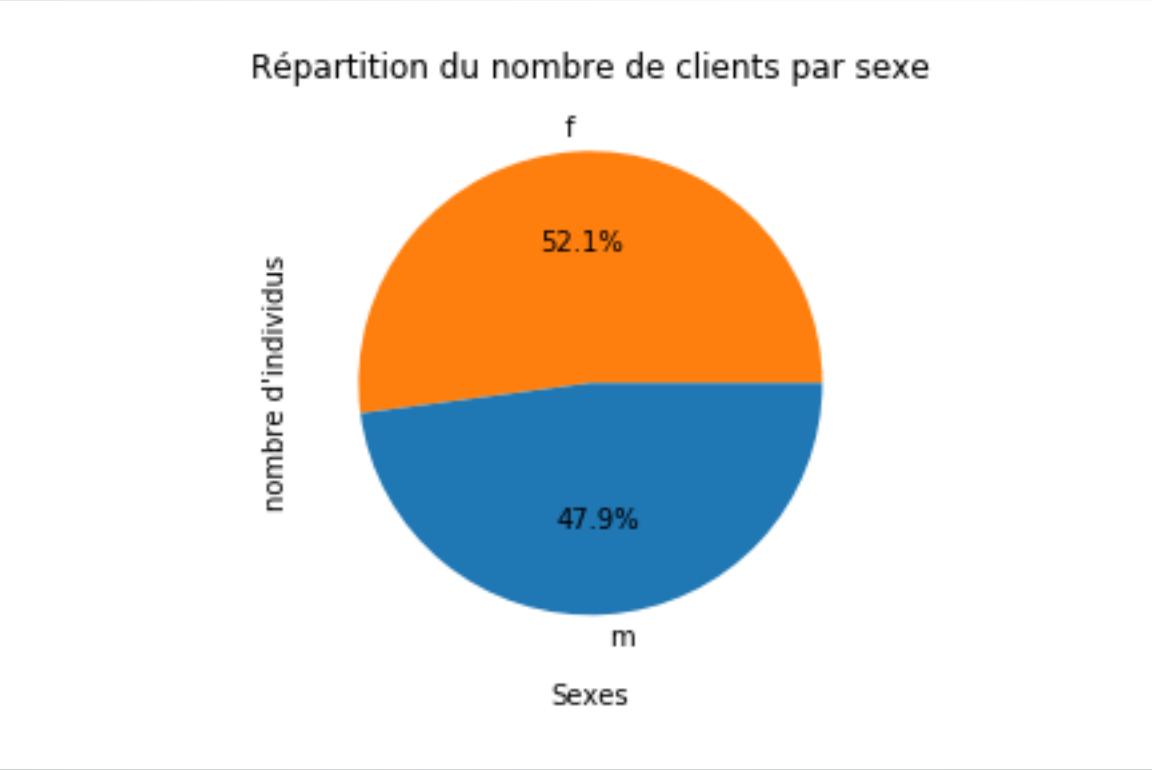
« Transactions »

- 337016 transactions
- 10,3 MB
- 4 attributs :
  - 'id_prod' – type object
  - 'date' – type object
  - 'session_id' – type object
  - 'client_id' – type object

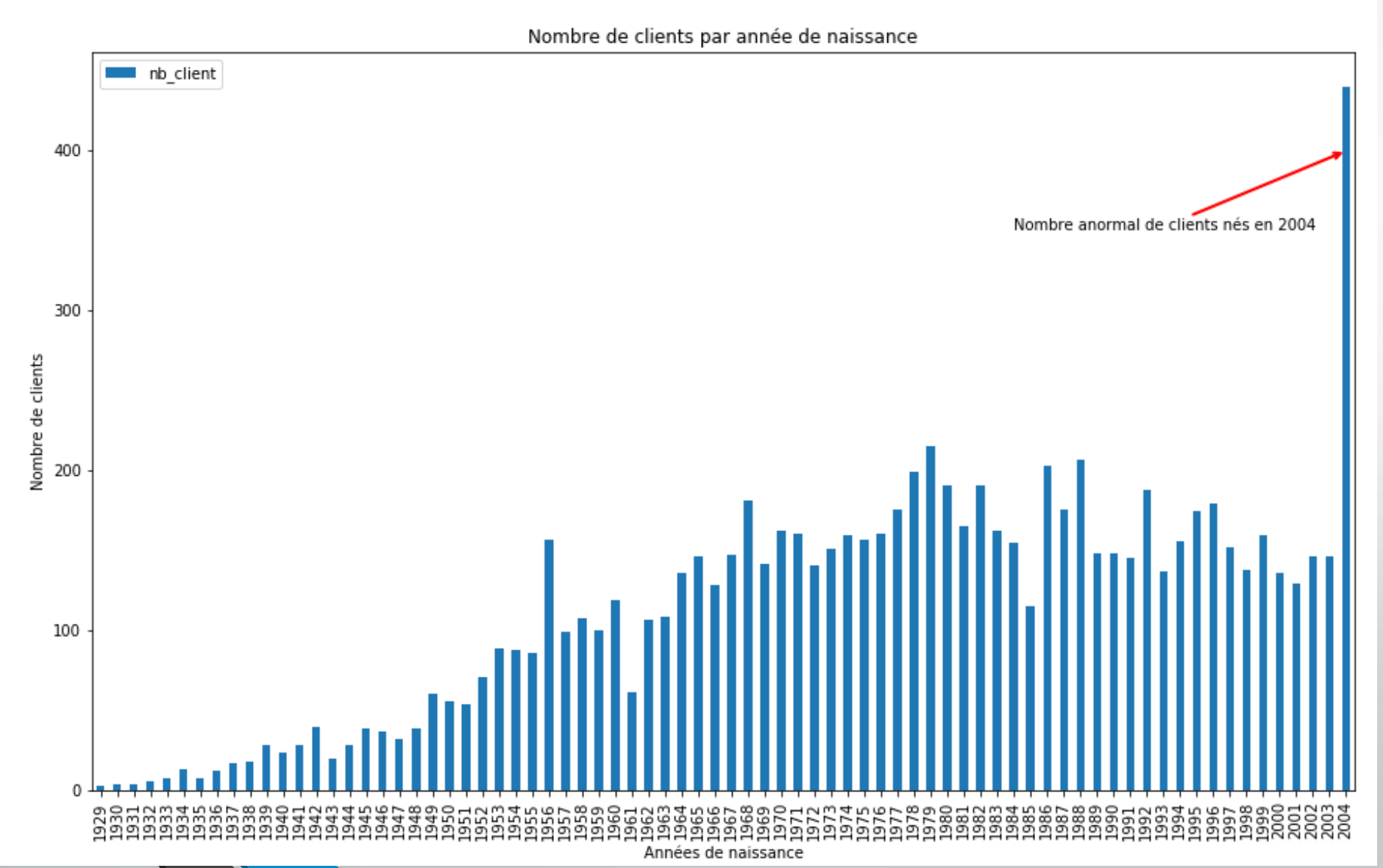
« Products »

- 3287 produits
- 77,2 KB
- 3 attributs :
  - 'id_prod' – type object
  - 'price' – type float64
  - 'categ' – type int64
- Catégories :
  - '0' : 2309 produits
  - '1' : 739 produits
  - '2' : 239 produits

Représentation du nombre de clients par sexe Table « **Customers** »



Nombre de clients par année de naissance



- Interprétation :
  - De nombreux mineurs renseignent l'année de naissance minimum, correspondant à l'âge de 18 ans pour procéder à des achats.
- J'ai créé une version des données sans prendre en compte l'année 2004 pour pouvoir analyser le lien entre la variable âge et d'autres variables

Changement de type de données

'Object' → 'Category'

« Customers »

- Le changement de type de données de la colonne 'sex' a permis de réduire la taille du fichier, de 200kb à 143kb

« Products »

- Le changement de type de données de la colonne 'categ' a permis de réduire la taille du fichier, de 77kb à 54,8kb

Enregistrements « test » (1)
Valeurs atypiques

« Customers » :
Deux entrées avec une structure différente (c_xxxx) des autres numéros clients

	client_id	sex	birth
2735	ct_0	f	2001
8494	ct_1	m	2001

« Products » :
Un produit avec un prix négatif

	id_prod	price	categ
731	T_0	-1.0	0

Enregistrements « test » (2)

« **Transactions** »
Dates commençants par 'test_'

	id_prod	date	session_id	client_id
1431	T_0	test_2021-03-01 02:30:02.237420	s_0	ct_1
2365	T_0	test_2021-03-01 02:30:02.237446	s_0	ct_1
2895	T_0	test_2021-03-01 02:30:02.237414	s_0	ct_1
5955	T_0	test_2021-03-01 02:30:02.237441	s_0	ct_0
7283	T_0	test_2021-03-01 02:30:02.237434	s_0	ct_1
...	...	...	...	...
332594	T_0	test_2021-03-01 02:30:02.237445	s_0	ct_0
332705	T_0	test_2021-03-01 02:30:02.237423	s_0	ct_1
332730	T_0	test_2021-03-01 02:30:02.237421	s_0	ct_1
333442	T_0	test_2021-03-01 02:30:02.237431	s_0	ct_1
335279	T_0	test_2021-03-01 02:30:02.237430	s_0	ct_0

200 rows × 4 columns

On retrouve les éléments identifiés précédemment :

- 'id_prod' : T_0
- 'client_id' : ct_1 ou ct_0

Ainsi que :

- 'session_id' : s_0

- La suppression de ces enregistrements permettraient de traiter l'attribut « date » en tant que tel. Notamment en changeant son type de données d'objet à « datetime64[ns] »

Valeur manquante

- Identification d'un produit dans « **transactions** » mais pas dans « **products** »
- C'est-à-dire qu'il existe un produit '0_2245' qui a été vendu alors qu'il n'est pas dans la base de données des produits
- Ce produit a été rajouté à la base de données « **products** » en indiquant que son prix vaut la médiane du prix de sa catégorie. J'ai choisi cette méthode car l'écart type est relativement faible.

Numéro de catégorie

- Numéro de produit '**0**_2245' donc catégorie **0** ?
- Est-ce que le premier chiffre du numéro de produit est toujours égale à sa catégorie ?

Oui, et c'est vrai pour tous les produits de la base « **products** »

Jointure des dataframes

« Customers »

- 3 attributs :
  - 'client_id' - type: objet
  - 'sex' – type: objet
  - 'birth' – type: int64

« Transactions »

- 4 attributs :
  - 'id_prod' – type object
  - 'date' – type object
  - 'session_id' – type object
  - 'client_id' – type object

« Products »

- 3 attributs :
  - 'id_prod' – type object
  - 'price' – type float64
  - 'categ' – type int64

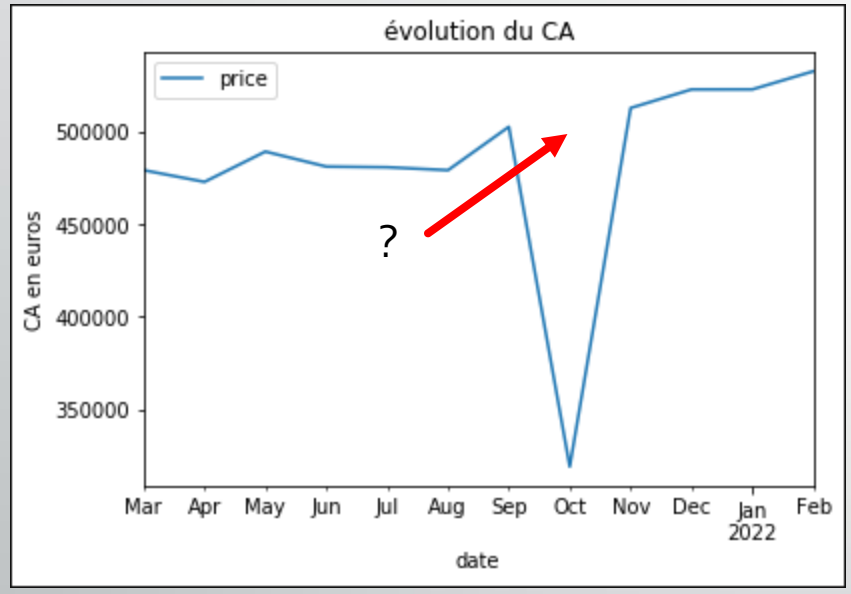
- Les dataframe ont été joint en utilisant leurs relations pour obtenir ce résultat :

date	session_id	client_id	id_prod	sex	birth	price	categ
2021-03-01 00:01:07.843138	s_1	c_329	0_1259	f	1967	11.99	0
2021-03-01 00:02:26.047414	s_2	c_664	0_1390	m	1960	19.37	0
2021-03-01 00:02:38.311413	s_3	c_580	0_1352	m	1988	4.50	0
2021-03-01 00:04:54.559692	s_4	c_7912	0_1458	f	1989	6.55	0
2021-03-01 00:05:18.801198	s_5	c_2033	0_1358	f	1956	16.49	0

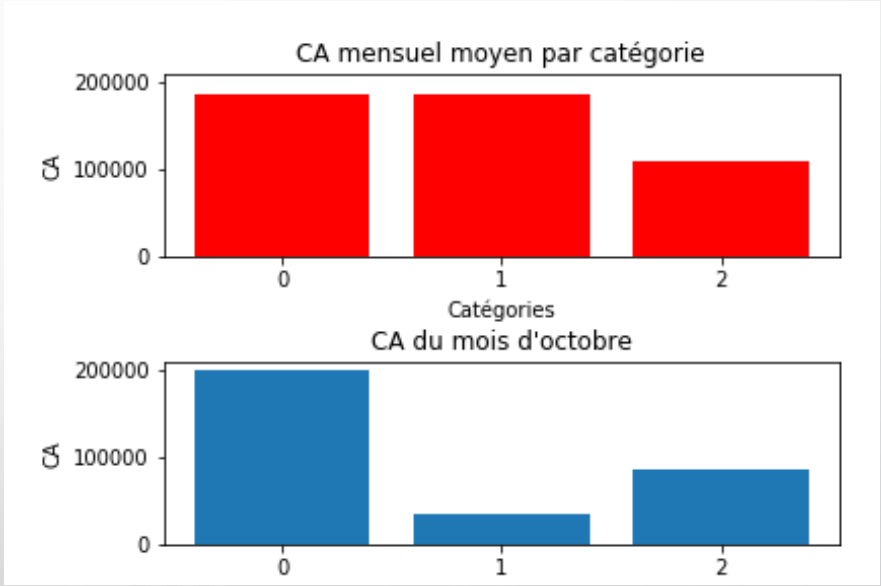
Analyse : le chiffres d'affaires mensuel

Données manquantes

Chiffre d'affaires mensuel



Comparaison CA mensuel moyen et CA du mois d'octobre par catégorie



Baisse de 81% entre le CA mensuel moyen et le CA d'octrobre

Analyse : le chiffres d'affaires mensuel

Données manquantes

- Il manque les ventes de la catégorie n°1 du 2 octobre au 27 octobre inclus.
- Remarque : J'ai comblé ce manque d'informations en pondérant le chiffre d'affaires total de cette catégorie (plus de détails dans les slides suivants)

Vente de la catégorie 1 en octobre

```
date
2021-10-01    7003.79
2021-10-02       0.00
2021-10-03       0.00
2021-10-04       0.00
2021-10-05       0.00
2021-10-06       0.00
2021-10-07       0.00
2021-10-08       0.00
2021-10-09       0.00
2021-10-10       0.00
2021-10-11       0.00
2021-10-12       0.00
2021-10-13       0.00
2021-10-14       0.00
2021-10-15       0.00
2021-10-16       0.00
2021-10-17       0.00
2021-10-18       0.00
2021-10-19       0.00
2021-10-20       0.00
2021-10-21       0.00
2021-10-22       0.00
2021-10-23       0.00
2021-10-24       0.00
2021-10-25       0.00
2021-10-26       0.00
2021-10-27       0.00
2021-10-28    6317.99
2021-10-29    6425.18
2021-10-30    6753.69
2021-10-31    7261.67
Freq: D, Name: price, dtype: float64
```

Analyse

Analyse : principaux clients (1)

par le montant annuel des achats

Top 5 des clients :

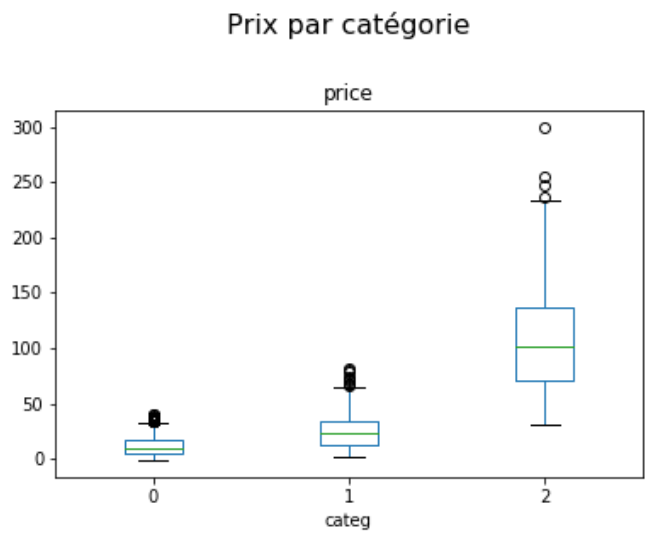
	client_id	price	percent_of_CA
677	c_1609	162007.34	0.027949
4388	c_4958	144257.21	0.024887
6337	c_6714	73197.34	0.012628
2724	c_3454	54442.92	0.009392
7715	c_7959	2564.25	0.000442

- On remarque un décrochage important entre le 4ème client avec un CA annuel supérieur à 50k€ et le 5ème client avec un CA annuel d'environ 2500€.
- On peut supposer que quatre de nos clients sont des grossistes en librairie.
- Il sera judicieux de scinder les clients en différents groupes pour proposer des analyses plus fines. J'ai notamment créé un dataframe sans ces clients

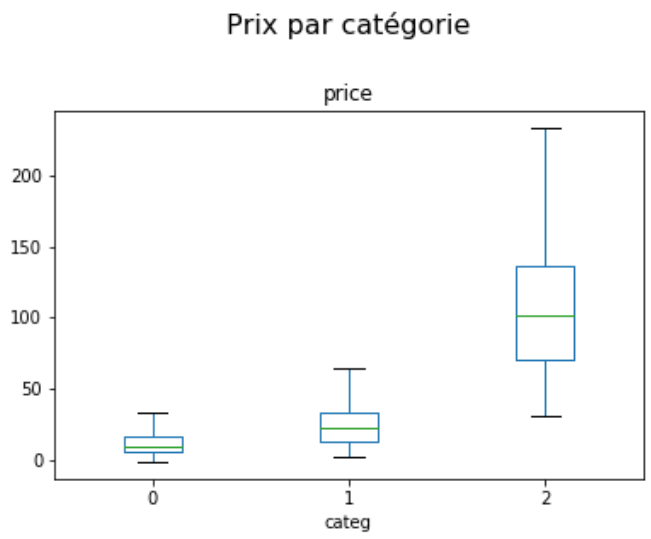
Analyse : distribution des prix par catégorie

Par produit disponible (1)

Prix par catégorie avec *outliers :*



Prix par catégorie sans *outliers :*

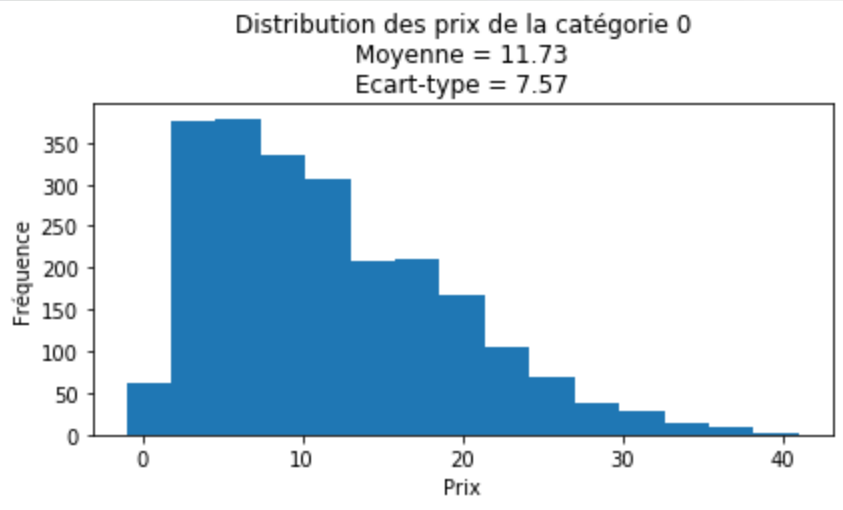


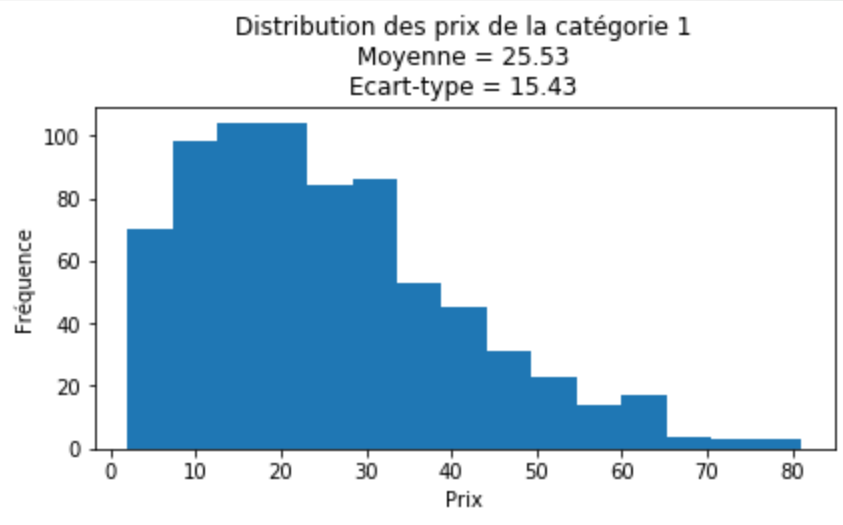
Description des prix par catégorie *:*

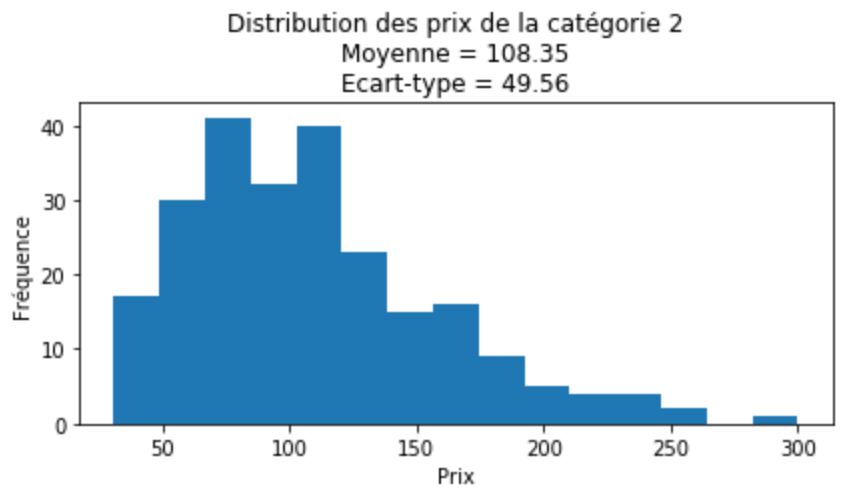
categ	count	mean	std	min	25%	50%	75%	max
0	2309.0	11.727280	7.568756	-1.00	5.580	10.32	16.65	40.99
1	739.0	25.531421	15.425162	2.00	13.390	22.99	33.99	80.99
2	239.0	108.354686	49.561431	30.99	71.065	101.99	136.53	300.00

Analyse : distribution des prix par catégorie

Par produit disponible (2)



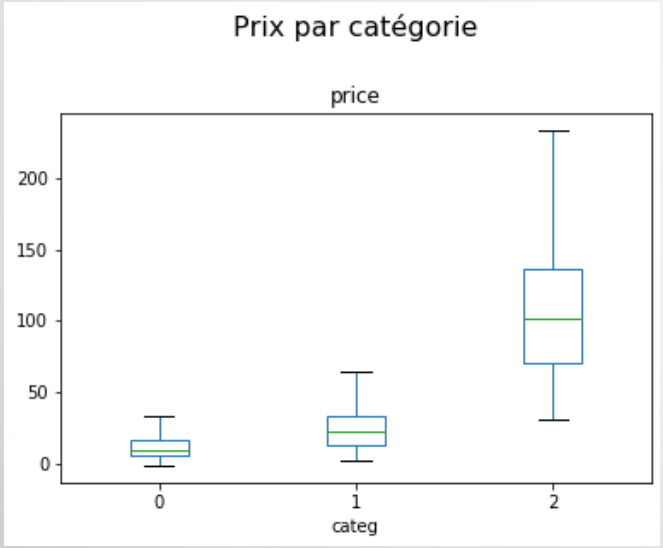




Analyse : distribution des prix par catégorie

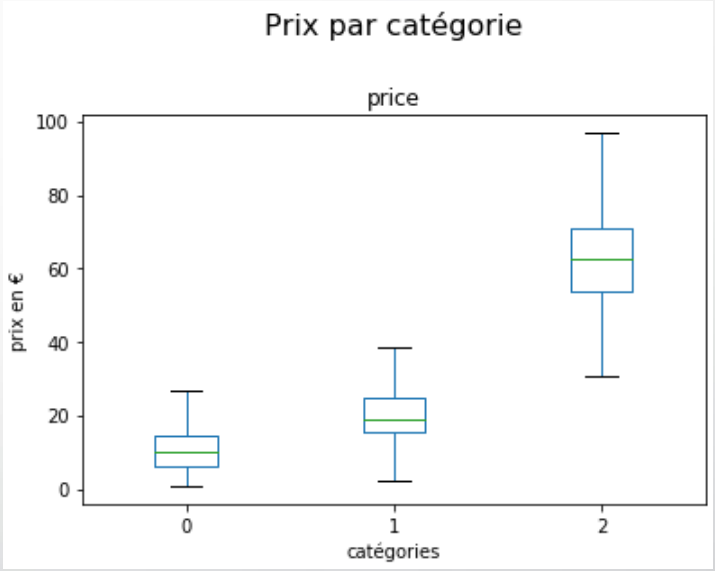
Comparatif

Prix par catégorie (disponible)



categ	count	mean
0	2309.0	11.727280
1	739.0	25.531421
2	239.0	108.354686

Prix par catégorie (ventes)



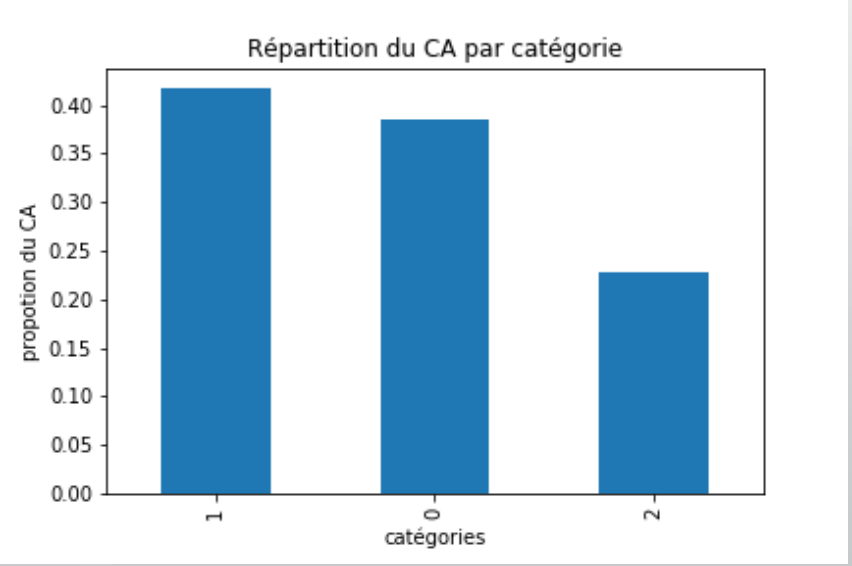
categ	count	mean
0	209426.0	10.646828
1	109735.0	20.480106
2	17552.0	75.174949

Analyse : chiffre d'affaires annuel (corrigé)

- En tenant compte des jours manquants sur les ventes de la catégorie 1, j'ai décidé de pondérer le chiffre d'affaires annuel de cette catégorie.
- J'ai divisé le CA de la catégorie 1 par 339/365
- Le CA est passé de 2 247 384€ à 2 419 750€ soit une augmentation environ 7,67%

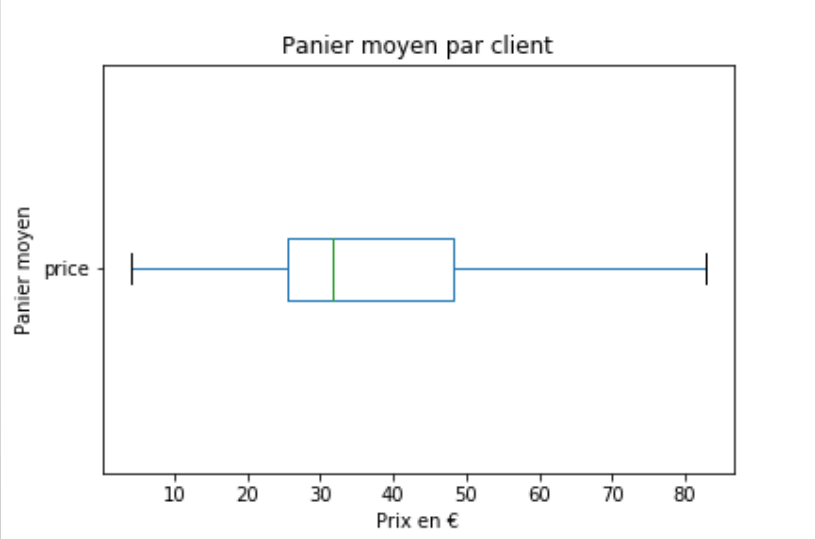
- Résultat :

	categ	price	percentage_of_CA
0	1	2.419750e+06	0.417445
1	0	2.229723e+06	0.384662
2	2	1.319471e+06	0.227629

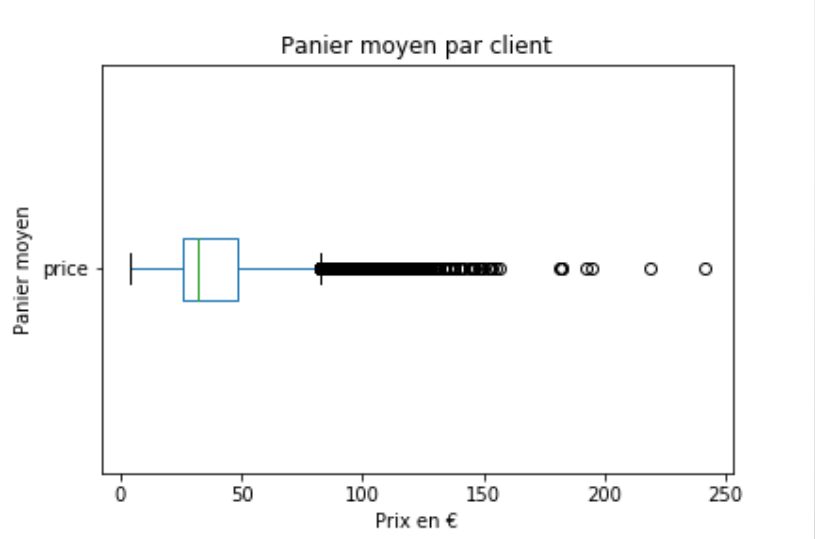


Analyse : panier moyen

Panier moyen sans *outliers :*



Panier moyen avec *outliers :*

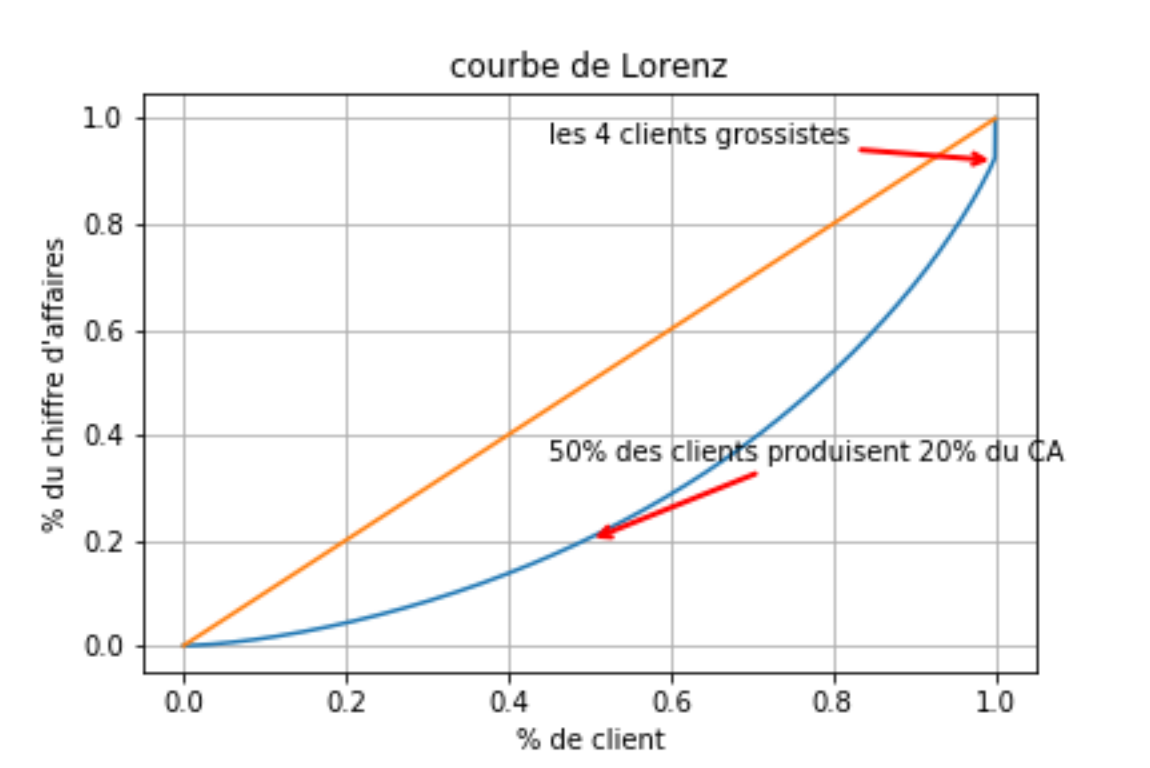


Description du panier moyen :

```
count    8600.000000
mean       40.106253
std        22.643521
min         4.150000
25%        25.396925
50%        31.839702
75%        48.329000
max       241.160000
Name: price, dtype: float64
```

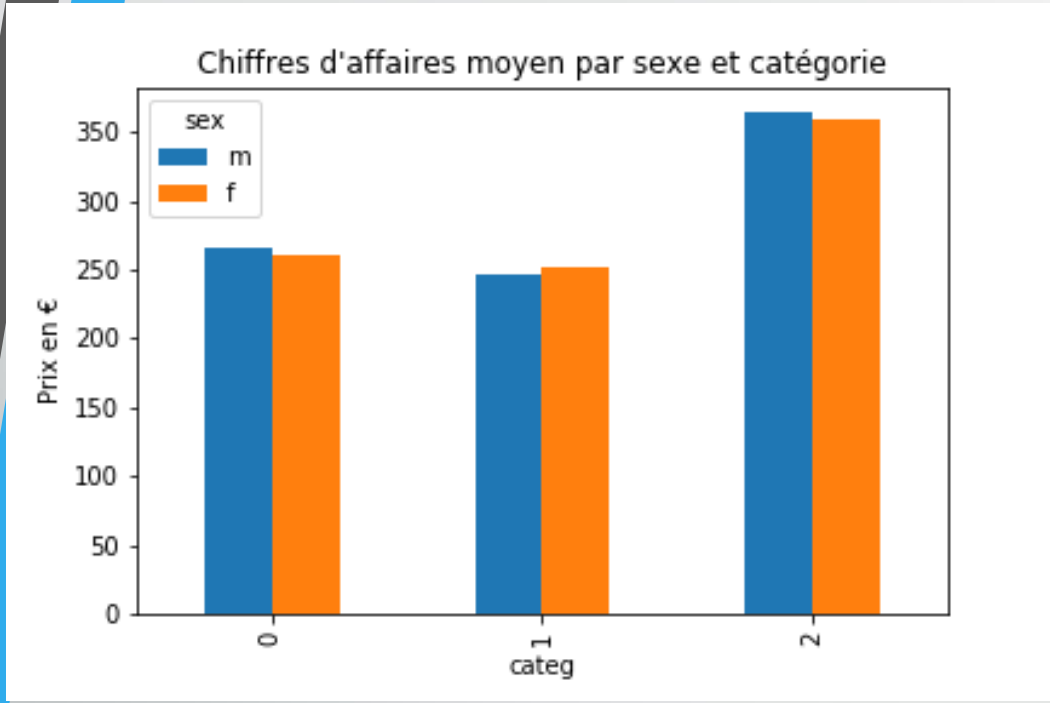
Analyse de la concentration du CA entre les clients

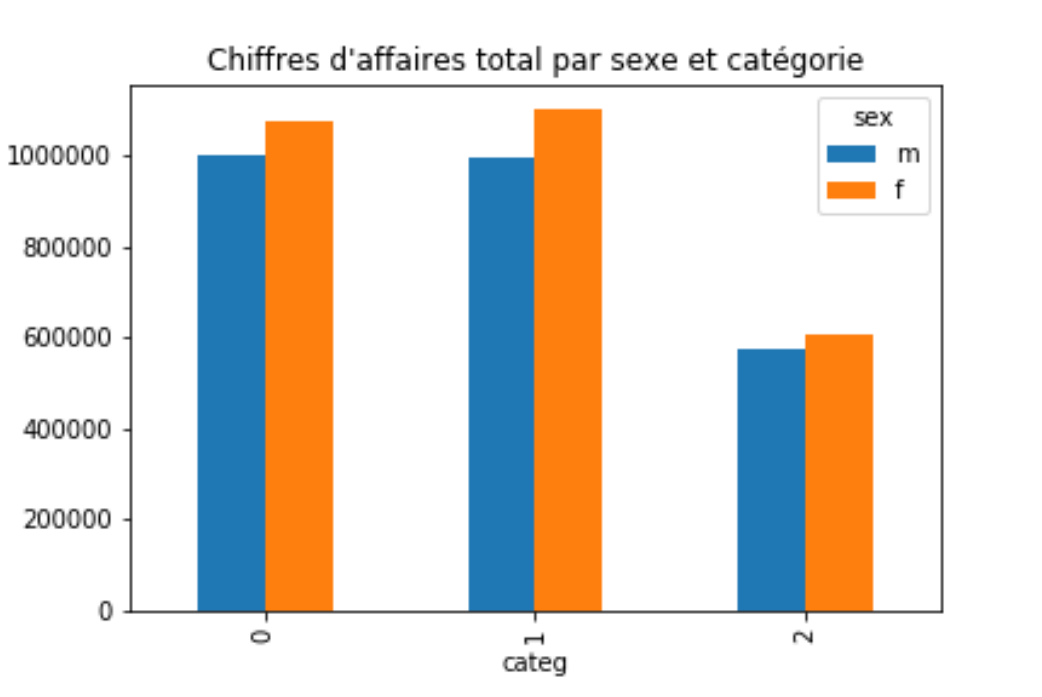
Courbe de Lorenz et indice de Gini



- 4 clients ont produit 7,49% du chiffre d'affaires
- Coefficient de Gini = 0,44
- Le chiffre d'affaires est répartie de façon plutôt égalitaire (0 étant une égalité parfaite et 1 une extrême inégalité)
- Analogie avec les pays :
  - Pays le plus inégalitaire :
    - Namibie (70,7)
  - Pays le plus égalitaire :
    - Suède (23)

Analyse : par sexe





- En tenant pas compte des clients grossistes on remarque que :
  - Sur le total les femmes représentent une plus grosse part de CA que les hommes
  - Sur la moyenne les différences entre les sexes sont faibles

Mathématiques
Postulats

- Seuil de signification retenu pour rejeter l'hypothèse nulle est $\alpha = 0{,}05$
- C'est le seuil le plus communément appliqué

Interprétation : entre le sexe et la catégorie
Test du khi²

Tableau de contingence des valeurs observées :

sex	categ_0	categ_1	categ_2	total
homme	101148	53774	8122	163044
femme	94023	48851	7634	150508
total	195171	102625	15756	313552

- Chi² = 10,2
- Degrés de liberté = 2
- P-value = 0,006
- Valeur critique = 5,991

Tableau de contingence des valeurs théoriques :

sex	categ_1	categ_2	categ_3	total
homme	101487.027747	53364.005014	8192.96724	163044.0
femme	93683.972253	49260.994986	7563.03276	150508.0
total	195171.000000	102625.000000	15756.00000	313552.0

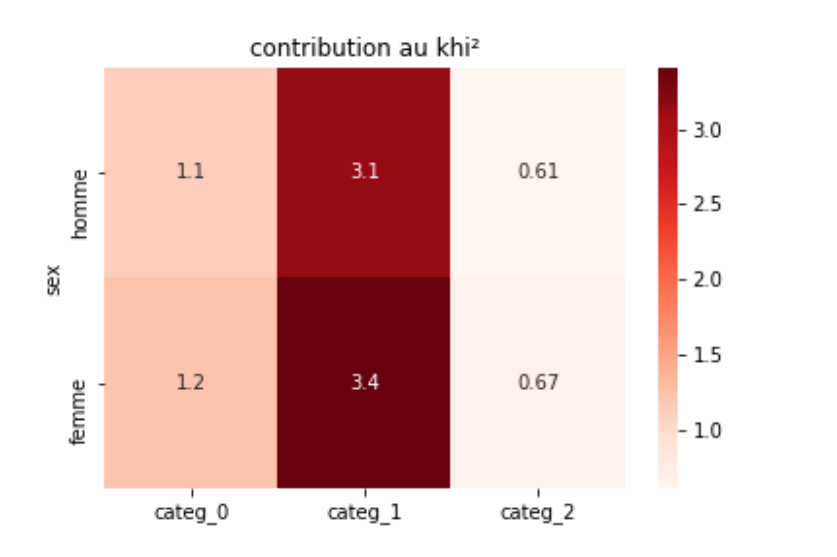
- <u>Hypothèse nulle</u>

  Ho : « Les deux variables sont indépendantes»
- <u>Hypothèse alternative</u>

  H1 : « Les deux variables sont dépendantes »
- P-value < 0,05 et chi² > valeur critique : on rejette l'hypothèse nulle avec une très forte présomption

Le sexe et la catégorie
Test du khi²

Contribution au chi² par attribut :

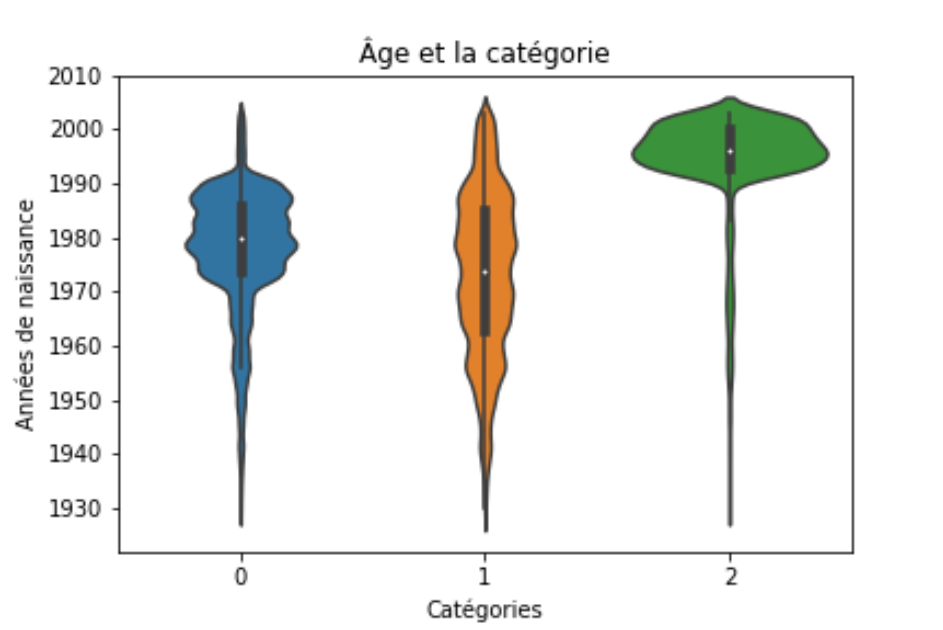


- Conclusion:
  - Le sexe et la catégorie sont deux variables dépendantes.
  - Les hommes achètent plus de livre de la catégorie n°1 que les femmes

Analyse : entre l'âge et la catégorie
ANOVA

Description de l'âge en fonction de la catégorie

categ	count	mean	std	min	25%	50%	75%	max
0	193503.0	1977.978031	11.309139	1929.0	1974.0	1980.0	1986.0	2003.0
1	99866.0	1973.362476	15.211421	1929.0	1963.0	1974.0	1985.0	2003.0
2	12835.0	1994.484768	9.854583	1930.0	1993.0	1996.0	2000.0	2003.0

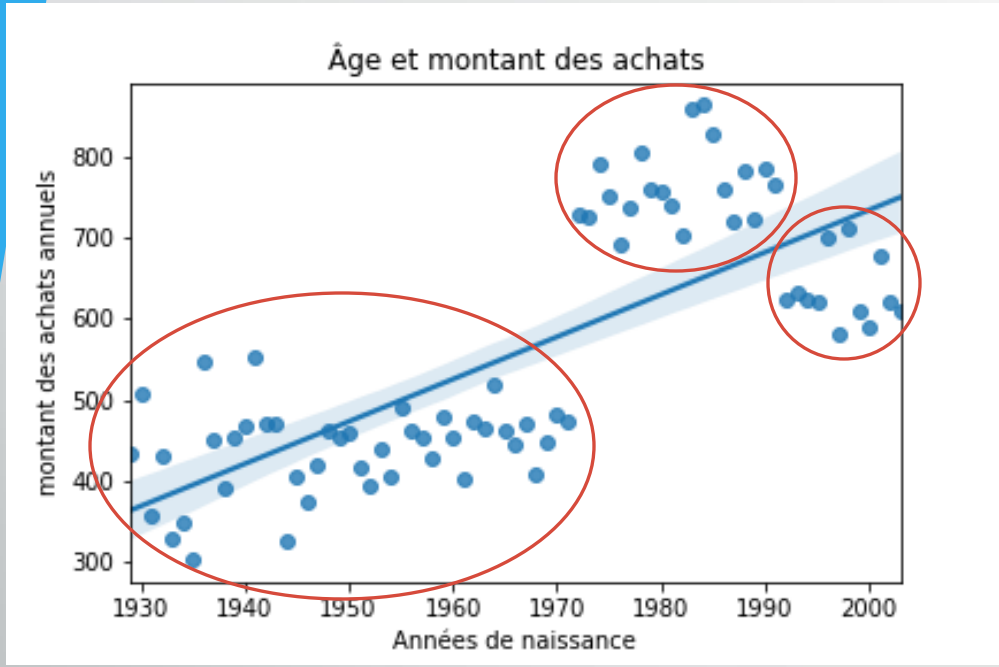


- **Hypothèses :**
  - Ho = "les moyennes des groupes sont égales"
  - H1 = "Au moins, l'une des moyennes de l'un des groupes est différente des moyennes des autres groupes"
- **Résultat du test ANOVA :**
  - F-statistic = 16906,007
  - P-value ≃ 0
  - Valeur critique = 2,996
- **Conclusion :**
  - P-value < 0,05 et F-stat > valeur critique : on rejette l'hypothèse nulle avec une très forte présomption
- **Interprétation :**
  - La catégorie 0 est majoritairement acheté par des clients entre 30 et 50 ans. Tandis que la catégorie 2 est plutôt préféré par les – de 30 ans. Les ventes de la catégorie 1 sont réparties de manière uniforme entre les âges.

Interprétation des corrélations

Identifications des groupes et recommandations

Analyse entre l'âge et le montant des achats



- Graphiquement, on peut distinguer 3 amas
- Ruptures :
  - entre 1971 et 1972
  - Entre 1991 et 1992

- <u>Hypothèse nulle</u>

  Ho : « la pente de la droite de régression est de zéro »
- <u>Hypothèse alternative</u>

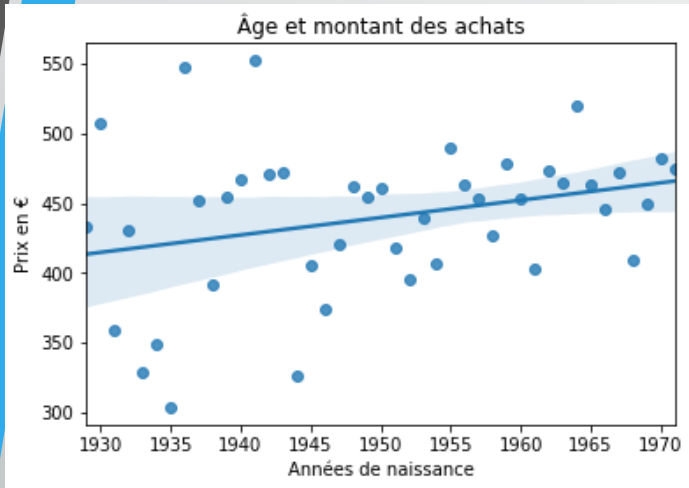
  H1 : « la pente de la droite de régression est différent de zéro »

- $r^2 = 0{,}56$
- La qualité du modèle est modéré

- $p = 1{,}12 \times 10^{-14}$
- $p < \alpha$ On peut rejeter avec une forte présomption l'hypothèse nulle.

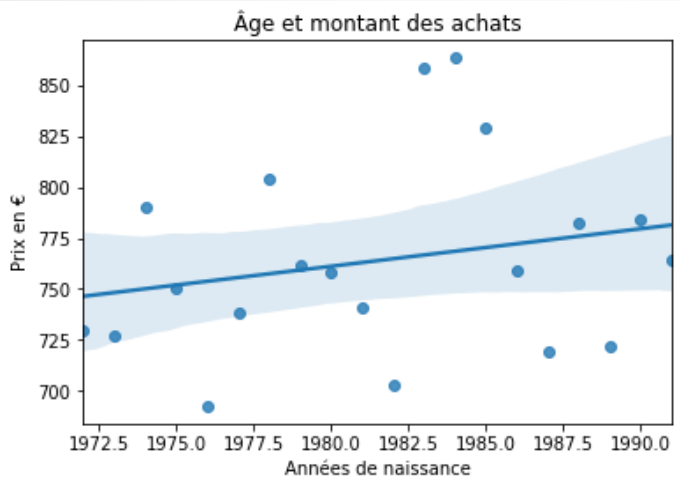
Analyse entre l'âge et le montant des achats
Analyse des groupes

Groupe 1 : les + de 50 ans



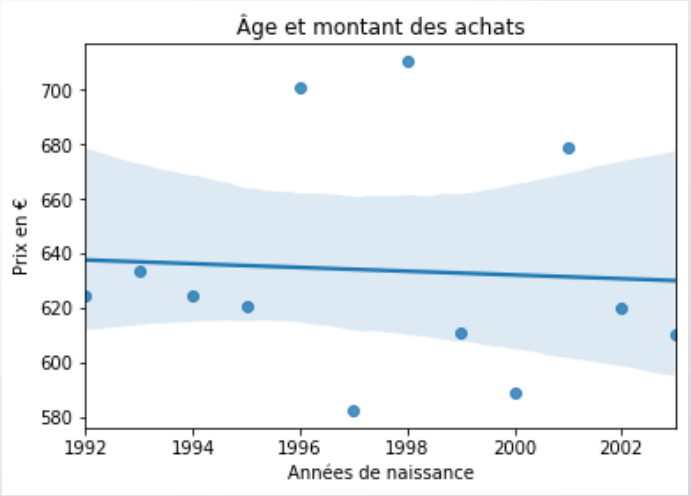
R^2 = 0,083
P = 0,06
Mean = 439,63
Std = 54,38
CV = 0,12

Groupe 2 : les 31-50 ans



R^2 = 0,053
P = 0,329
Mean = 763,91
Std = 47,41
CV = 0,06

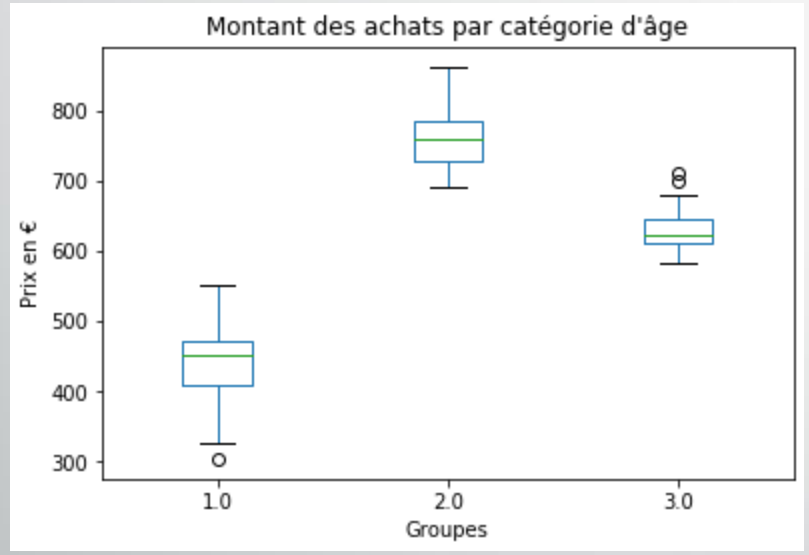
Groupe 3 : les – de 31 ans



R^2 = 0,053
P = 0,329
Mean = 633,82
Std = 41,25
CV = 0,07

- Les indicateurs statistiques nous montrent que les modèles intra-groupe ne sont pas significatifs. C'est-à-dire qu'au sein d'un groupe les comportements sont similaires. Cependant le CV intra-groupe a été largement réduit par rapport au CV général qui était de 0,27.

Analyse entre l'âge et le montant des achats groupes



- Groupe 1 : les + de 50 ans
- Groupe 2 : les 31-50 ans
- Groupe 3 : les – de 31 ans

- Dispersion des données significativement différente en fonction des groupes d'âge

Analyse entre l'âge et le montant des achats Groupes – ANOVA

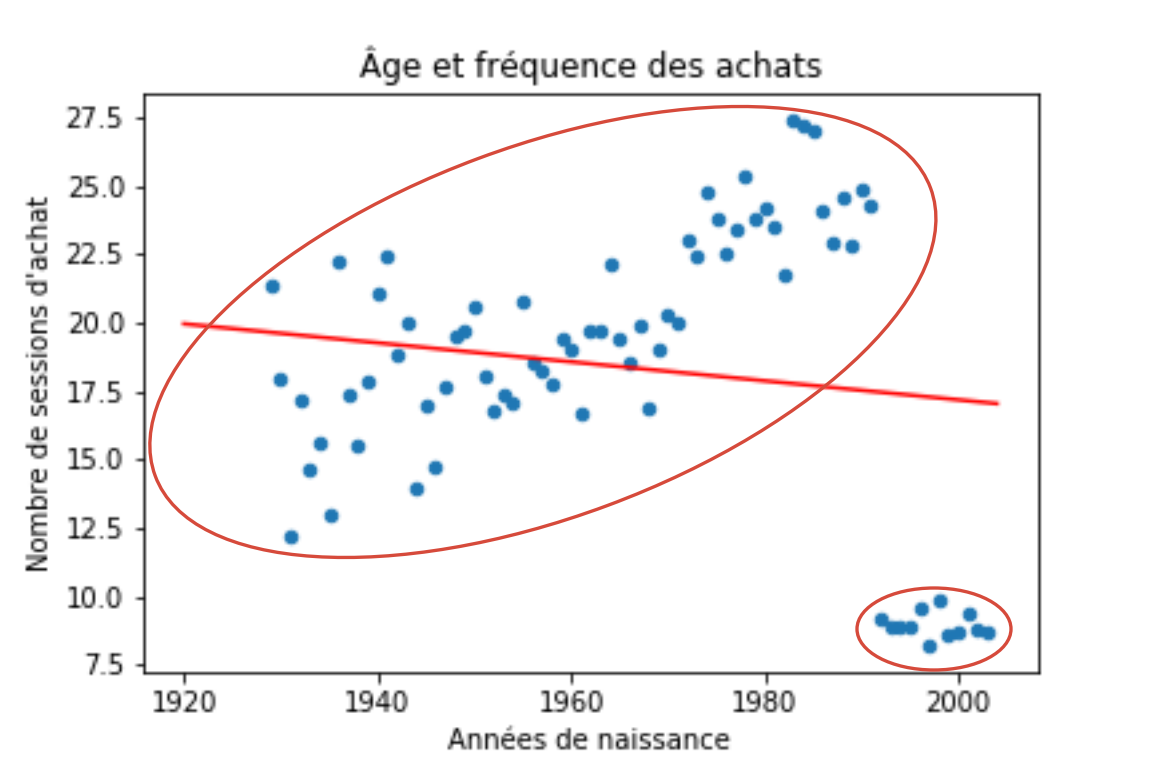
- **Hypothèses :**
  - Ho = "les moyennes des groupes sont égales"
  - H1 = "Au moins, l'une des moyennes de l'un des groupes est différente des moyennes des autres groupes"
- **Résultat du test ANOVA :**
  - F-statistic = 294,68
  - P-value= 2,129 * 10^-35
  - Valeur critique = 3,12
- **Conclusion :**
  - P-value < 0,05 et F-stat > valeur critique : on rejette l'hypothèse nulle avec une très forte présomption. De plus l'analyse post-hoc nous montre que chaque paire de groupe ont des moyennes différentes
- **Interprétation :**
  - En fonction de leur âge les clients achètent pour un montant différent

Analyse Post-hoc

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1.0	2.0	324.2768	0.001	291.3893	357.1643	True
1.0	3.0	194.1805	0.001	154.5102	233.8509	True
2.0	3.0	-130.0962	0.001	-174.4652	-85.7273	True

Analyse entre l'âge et la fréquence des achats



- Graphiquement, on peut distinguer 2 amas
- Rupture :
  - Entre 1991 et 1992

- <u>Hypothèse nulle</u>

  Ho : « la pente de la droite de régression est de zéro »

- <u>Hypothèse alternative</u>

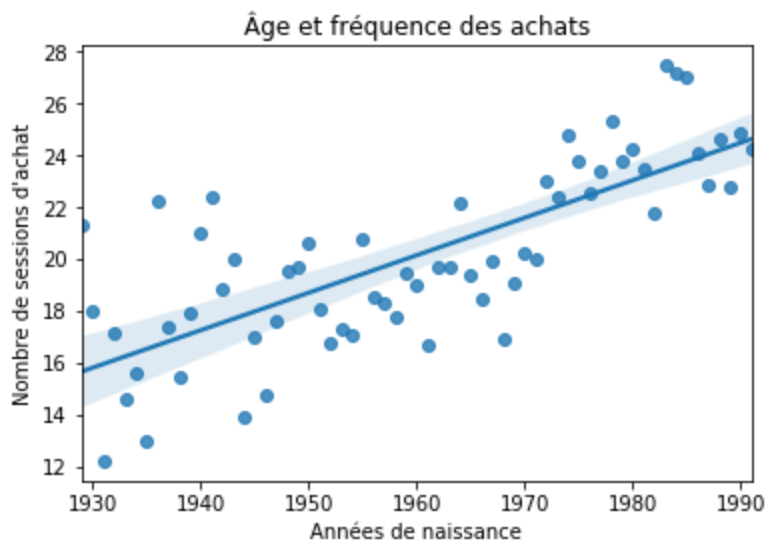
  H1 : « la pente de la droite de régression est différent de zéro »

- $r^2 = 0,02$
- La qualité du modèle est très faible

- $p = 0,216$
- La p-value est trop élevé pour rejeter l'hypothèse nulle. (droite de régression nulle)

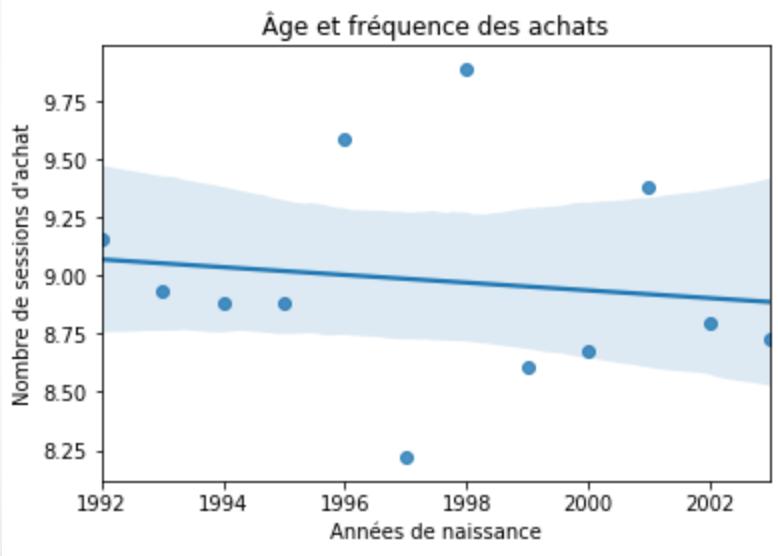
Analyse entre l'âge et la fréquence des achats
Analyse par groupe

Groupe 1 : les + de 30 ans



R^2 = 0,569
P = 9,16 * 10^-13
Mean = 20,15
Std = 3,51
CV = 0,17

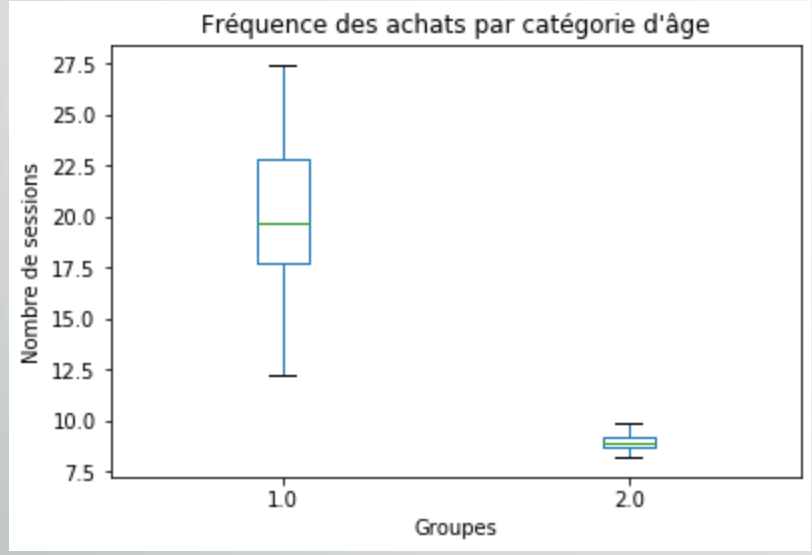
Groupe 2 : les – de 31 ans



R^2 = 0,017
P = 0,689
Mean = 8,98
Std = 0,46
CV = 0,05

- Au sein du groupe des + de 30 ans, on peut remarquer une corrélation importante entre l'âge et la fréquence des achats. Pour le groupe 2, le modèle est moins efficace mais l'amplitude du nombre de session est faible, on peut constater un coefficient de variation de 0,05, sachant qu'avant la dissection de la population en groupe il était de 0,2847

Analyse entre l'âge et la fréquence des achats
Distribution des données par groupe



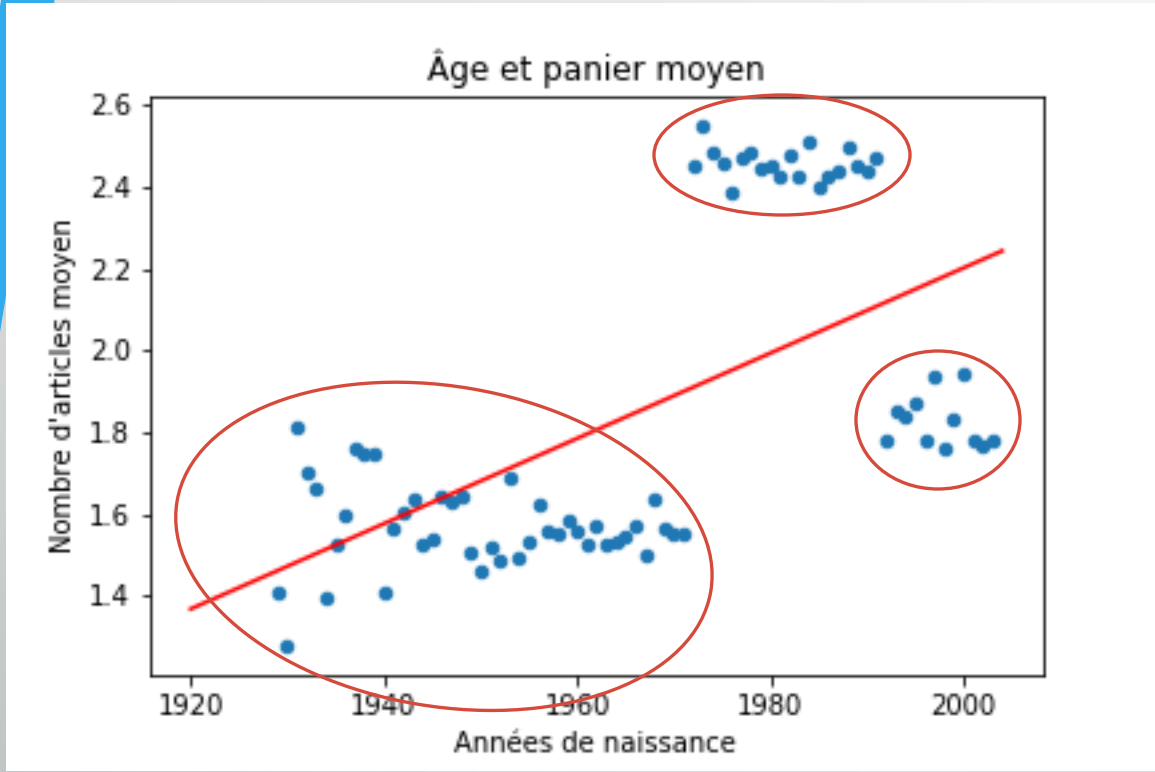
- Groupe 1 : les + de 30 ans
- Groupe 2 : les – de 31 ans

- Dispersion des données significativement différente en fonction des groupes d'âge

Analyse entre l'âge et la fréquence des achats Groupes – ANOVA

- **Hypothèses :**
  - Ho = "les moyennes des groupes sont égales"
  - H1 = "Au moins, l'une des moyennes de l'un des groupes est différente des moyennes des autres groupes"
- **Résultat du test ANOVA :**
  - F-statistic = 120,18
  - P-value= 4,386 * 10^-17
  - Valeur critique = 3,972
- **Conclusion :**
  - P-value < 0,05 et F-stat > valeur critique : on rejette l'hypothèse nulle avec une très forte présomption
- **Interprétation :**
  - En fonction de leur âge les clients achètent à des fréquences différentes. Notamment les +30 achètent environ 20 fois par an quand les – de 31 ans achètent environ 9 fois par an.

Analyse entre l'âge et le panier moyen



- Graphiquement : on peut distinguer 3 amas
- Ruptures :
  - entre 1971 et 1972
  - Entre 1991 et 1992

- Hypothèse nulle

  Ho : « la pente de la droite de régression est de zéro »

- Hypothèse alternative

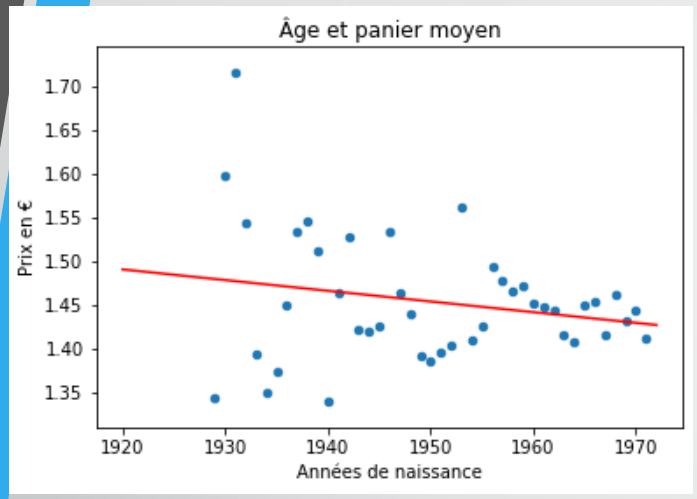
  H1 : « la pente de la droite de régression est différent de zéro »

- $r^2 = 0{,}3385$
- La qualité du modèle est médiocre

- $p = 4{,}38 * 10^{-8}$
- On peut rejeter avec une forte présomption l'hypothèse nulle.

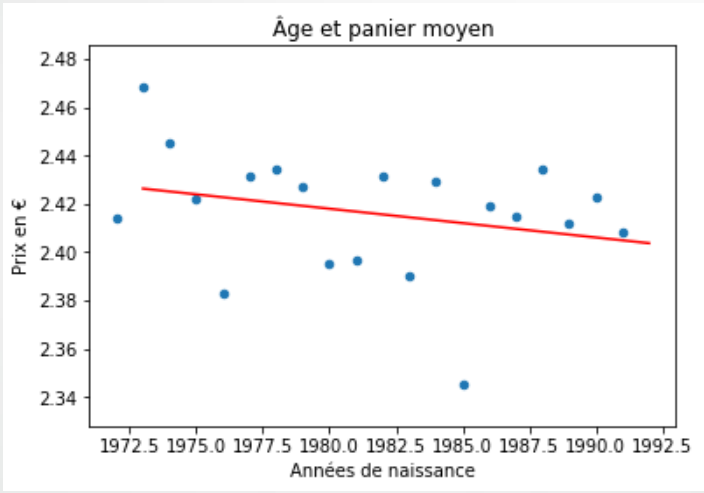
Analyse entre l'âge et le panier moyen

Groupe 1 : les + de 50 ans



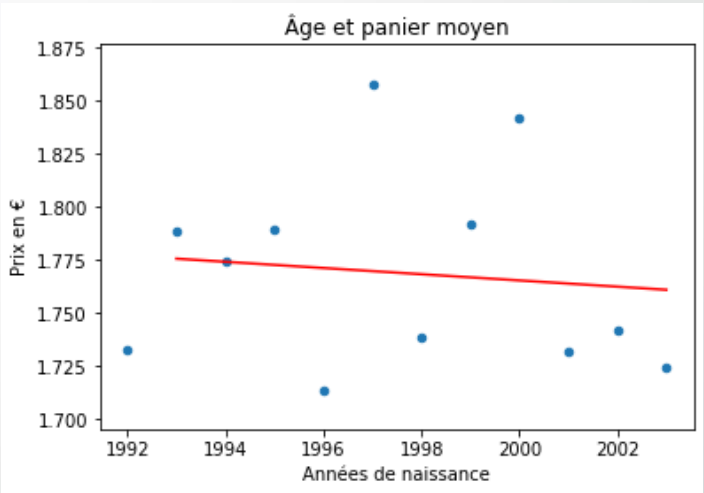
R^2 = 0,0085
P = 0,56
Mean = 1,57
Std = 0,103
CV = 0,065

Groupe 2 : les 31-50 ans



R^2 = 0,034
P = 0,4371
Mean = 2,46
Std = 0,037
CV = 0,015

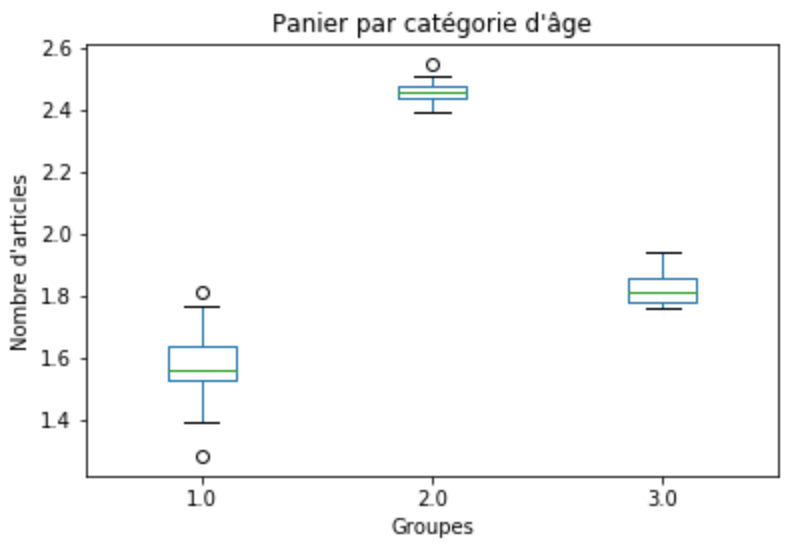
Groupe 3 : les – de 31 ans



R^2 = 0,027
P = 0,612
Mean = 1,83
Std = 0,063
CV = 0,035

- Les indicateurs statistiques nous montrent que les modèles intra-groupe ne sont pas significatifs. C'est-à-dire qu'au sein d'un groupe les comportements sont similaires. Cependant le coefficient de variation a été largement réduit (CV général = 0,211)

Analyse entre l'âge et le panier moyen
Distribution des données par groupe



- Groupe 1 : les + de 50 ans
- Groupe 2 : les 31-50 ans
- Groupe 3 : les – de 31 ans

- Dispersion des données significativement différente en fonction des groupes d'âge. Particulièrement entre le groupe 2 et les autres groupes.

Analyse entre l'âge et le panier moyen
Groupes – ANOVA

- **Hypothèses :**
  - Ho = "les moyennes des groupes sont égales"
  - H1 = "Au moins, l'une des moyennes de l'un des groupes est différente des moyennes des autres groupes"
- **Résultat du test ANOVA :**
  - F-statistic = 754,20817
  - P-value= 5,107 * 10^-49
  - Valeur critique = 3,13
- **Conclusion :**
  - P-value < 0,05 et F-stat > valeur critique : on rejette l'hypothèse nulle avec une très forte présomption
- **Interprétation :**
  - En fonction de leur âge les clients achètent pour un montant différent

Analyse Post-hoc

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1.0	2.0	0.8878	0.001	0.8331	0.9425	True
1.0	3.0	0.2575	0.001	0.1915	0.3235	True
2.0	3.0	-0.6303	0.001	-0.7041	-0.5565	True

Résumé des groupes (1)

Paramètres	Groupe 1 : les + de 50 ans	Groupe 2 : les 31-50 ans	Groupe 3 : les – de 31 ans
Fréquence d'achat	Moyenne : 20 Médiane : 20		Moyenne : 9 Médiane : 9
Montant moyen des achats	Moyenne : 439€ Médiane : 452€	Moyenne : 763€ Médiane : 759€	Moyenne : 633€ Médiane : 622€
Panier moyen	Moyenne : 1,45 Médiane : 1,44	Moyenne : 2,42 Médiane : 2,42	Moyenne : 1,77 Médiane : 1,76
Catégorie préférentielle		Catégorie n°0	Catégorie n°2
Poids dans CA total	26,46 %	50,67 %	22,87 %

- Les 31-50 ans sont notre cœur de cible, ils arrivent en tête en :
  - Fréquence d'achat
  - Montant des achats
  - Panier moyen
  - Poids dans le CA total

Résumé des groupes (2)

Paramètres	Groupe 1 : les + de 50 ans	Groupe 2 : les 31-50 ans	Groupe 3 : les – de 31 ans
Fréquence d'achat	Moyenne : 20 Médiane : 20		Moyenne : 9 Médiane : 9
Montant moyen des achats	Moyenne : 439€ Médiane : 452€	Moyenne : 763€ Médiane : 759€	Moyenne : 633€ Médiane : 622€
Panier moyen	Moyenne : 1,45 Médiane : 1,44	Moyenne : 2,42 Médiane : 2,42	Moyenne : 1,77 Médiane : 1,76
Catégorie préférentielle		Catégorie n°0	Catégorie n°2
Poids dans CA total	26,46 %	50,67 %	22,87 %

- Les - 31 ans achètent peu souvent et avec un panier moyen modéré mais ils achètent pour un montant assez élevé. En effet, leur catégorie d'achat préférentielle est la catégorie n°2. C'est la catégorie qui a les prix de ventes les plus élevées.

Résumé des groupes (3)

Paramètres	Groupe 1 : les + de 50 ans	Groupe 2 : les 31-50 ans	Groupe 3 : les – de 31 ans
Fréquence d'achat	Moyenne : 20 Médiane : 20		Moyenne : 9 Médiane : 9
Montant moyen des achats	Moyenne : 439€ Médiane : 452€	Moyenne : 763€ Médiane : 759€	Moyenne : 633€ Médiane : 622€
Panier moyen	Moyenne : 1,45 Médiane : 1,44	Moyenne : 2,42 Médiane : 2,42	Moyenne : 1,77 Médiane : 1,76
Catégorie préférentielle		Catégorie n°0	Catégorie n°2
Poids dans CA total	26,46 %	50,67 %	22,87 %

- Les + de 50 ans achètent régulièrement mais pour un panier et un montant faible.

Interprétation - Recommandations

- On remarque une rupture dans les comportements d'achats des clients quand ils arrivent à la trentaine. Ils achètent plus souvent pour un montant plus important.

Algorithme de recommandation

- L'algorithme de recommandation est particulièrement utile pour les personnes de + de 31 ans car ils achètent fréquemment et pour un panier moyen assez important.
- Il est moins performant pour les clients de - de 31 ans car ils achètent moins régulièrement.
- Pour que l'algorithme de recommandation soit performant il a besoin d'un volume de données par personne important
- La catégorie de produit n°0 est la préféré des 31-50ans, qui sont notre cœur de cible. Je recommande donc un perfectionnement de l'algorithme sur cette catégorie en priorité

Interprétation - Recommandations

- Promotions :
  - Les plus de 30 ans pourraient être réceptifs aux promotions sur la catégorie 0
  - Les moins de 31 ans pourraient être réceptifs aux promotions sur la catégorie 2.
    - La catégorie n°2 contient les produits les plus chers.
    - On peut supposer qu'à leur âge ils ont un pouvoir d'achat modéré.
    - Les moins de 31 ans ont la fréquence d'achat la plus faible.
  - Ce qui m'amène à conclure qu'il faudrait proposer des promotions au – de 31 ans sur la catégorie n°2 à des périodes charnières comme les fêtes de Noël et la rentrée scolaire, où les moins de 31 ans pourraient bénéficier d'une hausse temporaire de leur pouvoir d'achat.
- Cross-selling :
  - Les + de 50 ans achètent régulièrement mais pour un panier moyen bien inférieur aux 31-50 ans. Il serait très intéressant de leur proposer des articles complémentaires ou associées, basés sur l'algorithme de recommandation, notamment sous la forme de « bundle ».