

*Analyse des ventes*

***Rester livres***

---

# Sommaire

## **Données de ventes :**

Présentation

Nettoyage

Analyses

Corrélations

# Sommaire

## **Données de ventes :**

**Présentation**

Nettoyage

Analyses

Corrélations

# Source : Base de donnée des ventes

Clients

	client_id	sex	birth
0	c_4410	f	1967
1	c_7839	f	1975
2	c_1699	f	1984
3	c_5961	f	1962
4	c_5320	m	1943

Produits

	id_prod	price	categ
0	0_1421	19.99	0
1	0_1368	5.13	0
2	0_731	17.99	0
3	1_587	4.99	1
4	0_1507	3.99	0

Transactions

	id_prod	date	session_id	client_id
0	0_1483	2021-04-10 18:37:28.723910	s_18746	c_4450
1	2_226	2022-02-03 01:55:53.276402	s_159142	c_277
2	1_374	2021-09-23 15:13:46.938559	s_94290	c_4270
3	0_2186	2021-10-17 03:27:18.783634	s_105936	c_4597
4	0_1351	2021-07-17 20:34:25.800563	s_63642	c_1242

# Sommaire

## **Données de ventes :**

Présentation

**Nettoyage**

Analyses

Corrélations

# Jointure

id_prod	price	categ	id_prod	date	session_id	client_id	client_id	sex	birth		
0	0_1421	19.99	0	0_1483	2021-04-10 18:37:28.723910	s_18746	c_4450	0	c_4410	f	1967
1	0_1368	5.13	0	2_226	2022-02-03 01:55:53.276402	s_159142	c_277	1	c_7839	f	1975
2	0_731	17.99	0	1_374	2021-09-23 15:13:46.938559	s_94290	c_4270	2	c_1699	f	1984
3	1_587	4.99	1	0_2186	2021-10-17 03:27:18.783634	s_105936	c_4597	3	c_5961	f	1962
4	0_1507	3.99	0	0_1351	2021-07-17 20:34:25.800563	s_63642	c_1242	4	c_5320	m	1943

# Nettoyage - suppression et ajout

Suppression des tests

	c_id	c_sex	c_birth	p_id	p_price	p_cat	t_date	t_sess_id
108206	ct_0	f	2001.0	T_0	-1.0	0.0 test	2021-03-01 02:30:02.237446	s_0
108284	ct_0	f	2001.0	T_0	-1.0	0.0 test	2021-03-01 02:30:02.237419	s_0

Produit *0\_2245* manquant

	c_id	c_sex	c_birth	p_id	p_price	p_cat	t_date	t_sess_id
2678	c_4505	m	1976.0	0_2245	10.64	0	2022-01-09 09:23:31.000720	s_147220

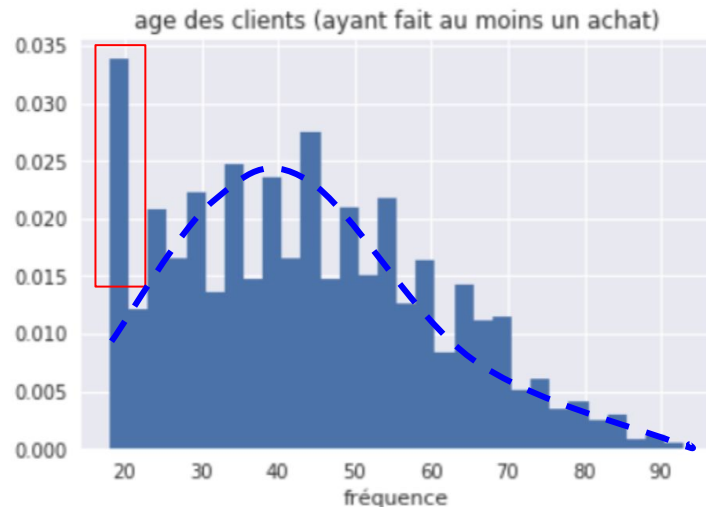
# Nettoyage - Âges

Calcul de l'âge

	c_id	c_birth	c_age	c_age_range	t_date	t_buyer_age	t_buyer_age_range
0	c_4410	1967.0	55.0	51-70	2021-03-22 01:40:22.782925	54.0	51-70


18 ans - pertinent ?

	c_id	c_birth	c_age	c_age_range	data_info
1316	c_1046	2004.0	18.0	30-	age_irrelevant





# Nettoyage - Dates



	<b>t_date</b>	<b>t_month</b>	<b>t_month_part</b>	<b>t_weekday</b>	<b>t_hour</b>
0	2021-03-22 01:40:22.782925	3.0	2	0.0	1.0
1	2021-11-04 16:28:30.169021	11.0	0	3.0	16.0
2	2021-03-22 01:27:49.480137	3.0	2	0.0	1.0
3	2021-08-11 08:40:47.495793	8.0	1	2.0	8.0
4	2022-01-18 17:05:07.468131	1.0	2	1.0	17.0
5	2021-11-12 18:11:43.280574	11.0	1	4.0	18.0
6	2021-09-25 00:17:38.676453	9.0	2	5.0	0.0
7	2021-12-01 07:31:51.359660	12.0	0	2.0	7.0
8	2021-09-25 00:11:19.292740	9.0	2	5.0	0.0

# Enregistrement

**3 fichiers csv :**

- Les produits **invendus**
- les clients **sans achats**
- les **transactions** avec les informations produits / clients

# Sommaire

## **Données de ventes :**

Présentation

Nettoyage

**Analyses**

Corrélations

Cette année

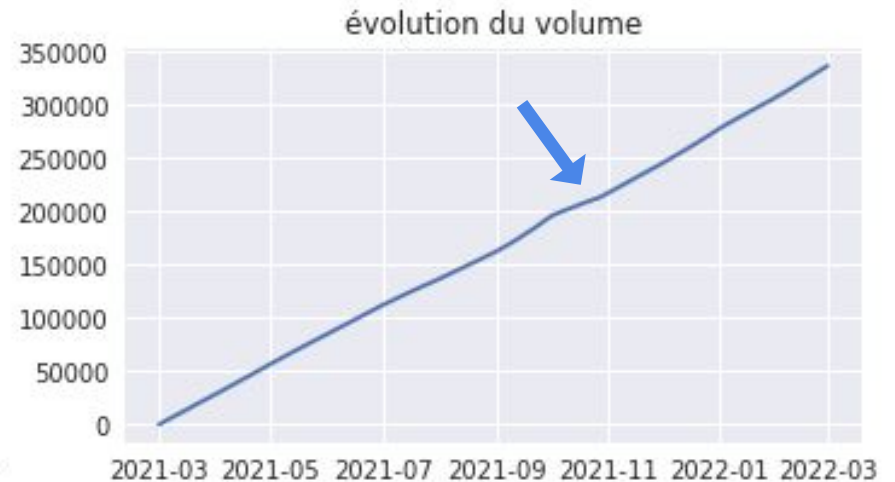
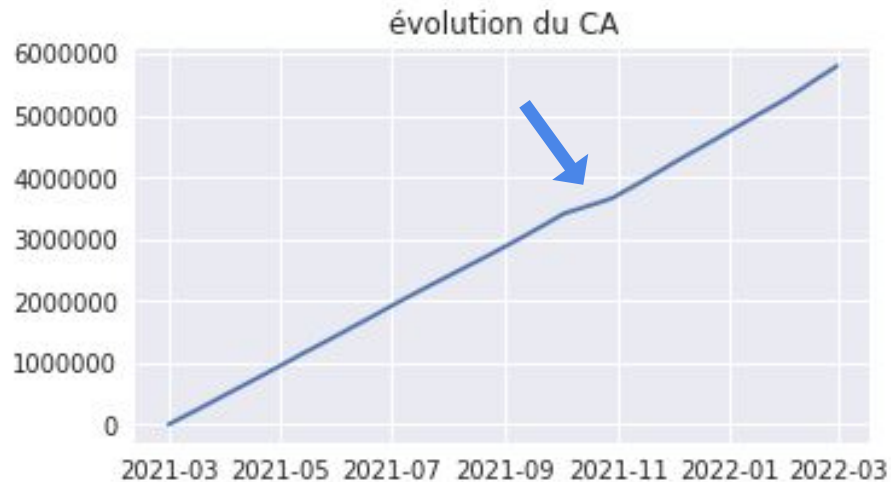
CA

**6 millions €**

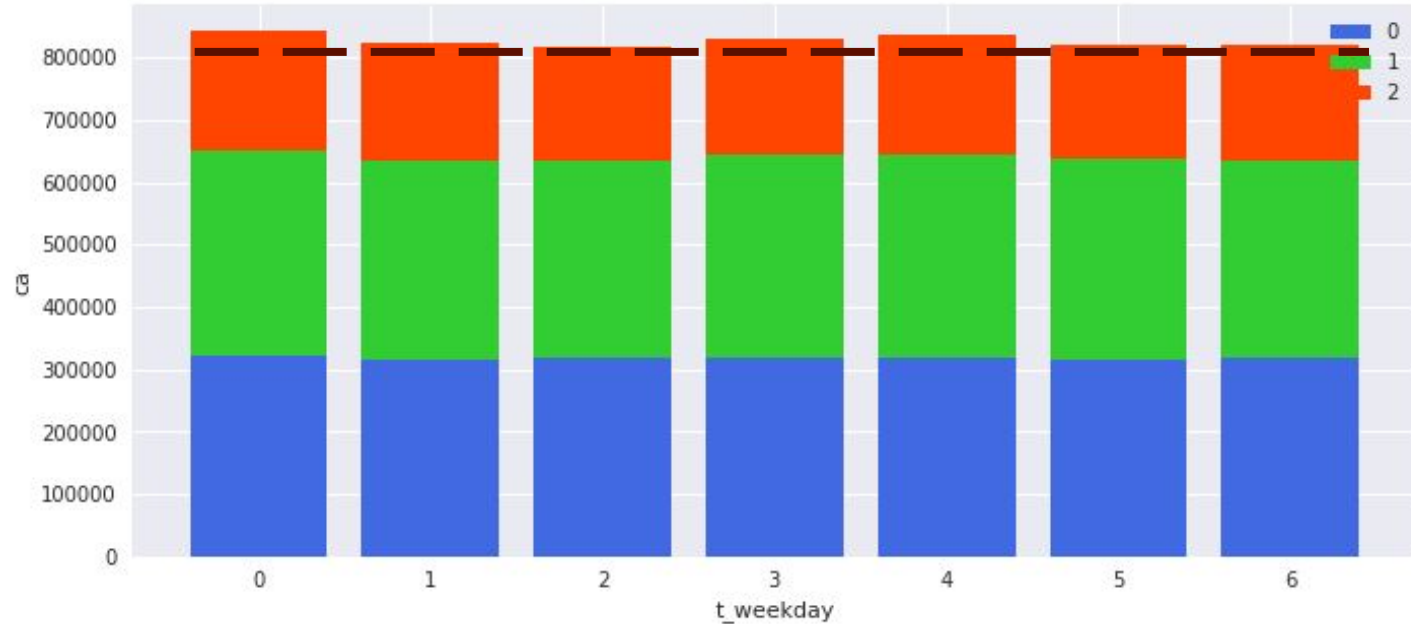
articles vendus

**350 000**

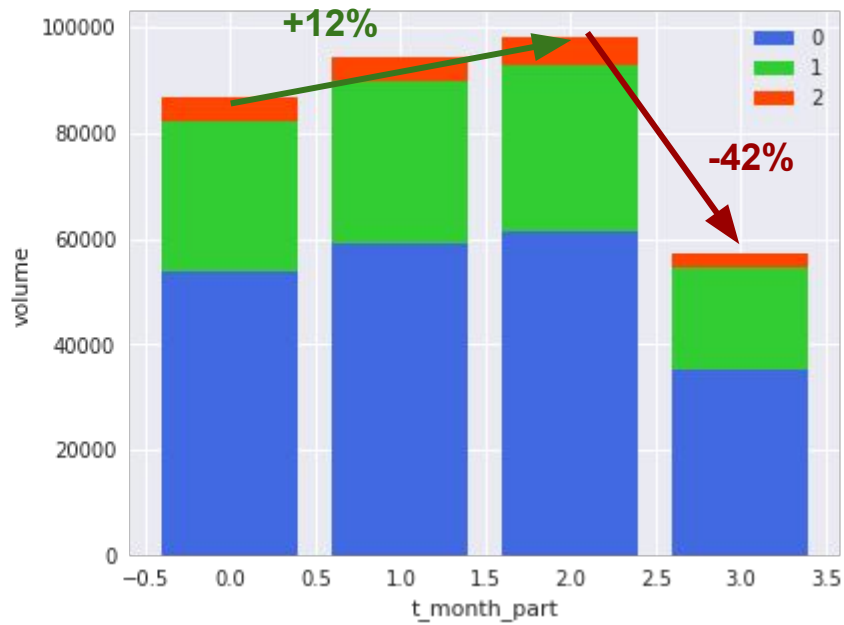
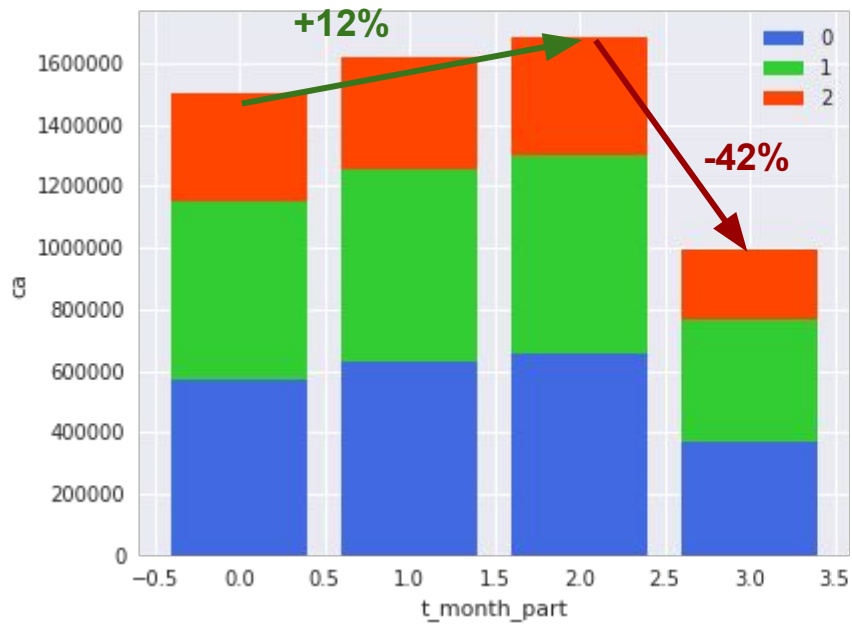
# CA et volume - vue globale



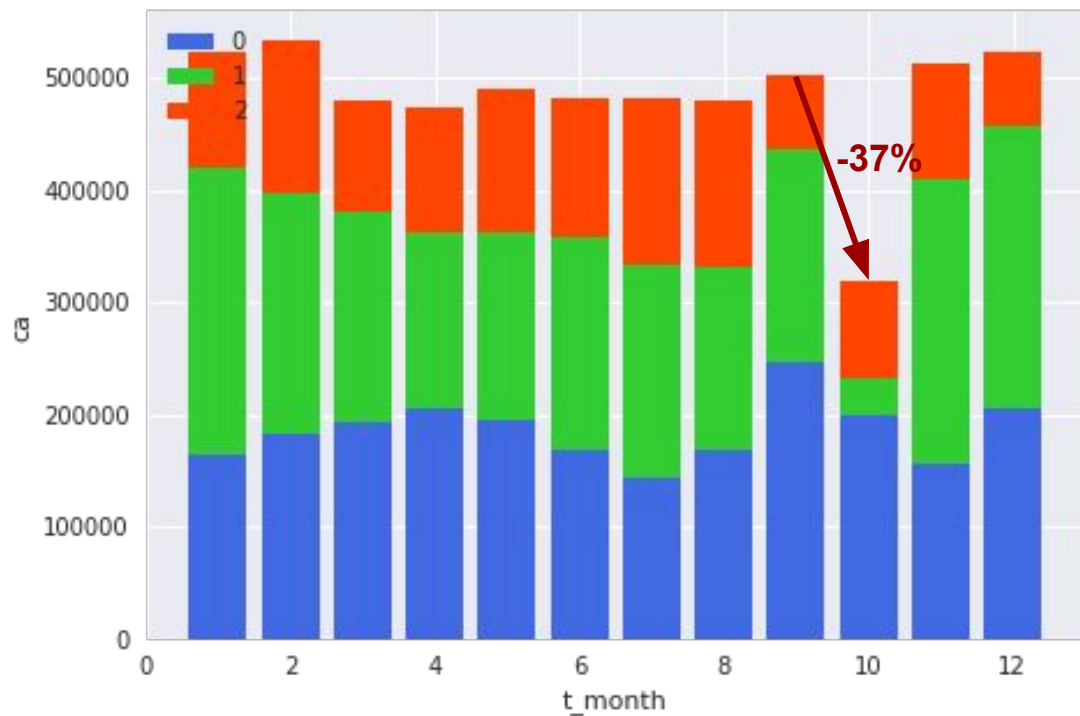
# CA et volume - jour et heure



# CA et volume - 1/4 mois

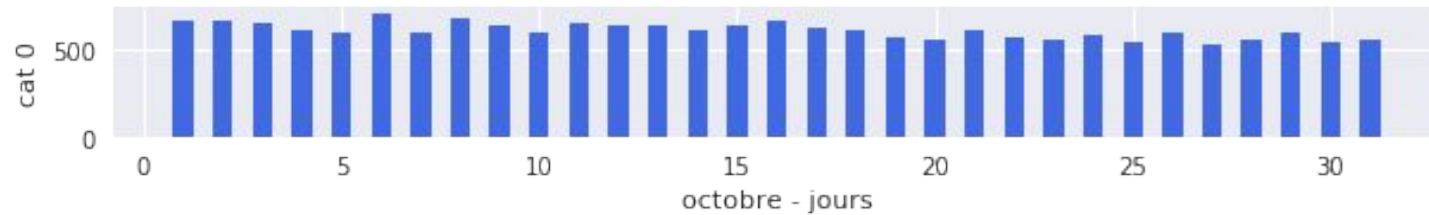


# Chiffre d'affaires - mois

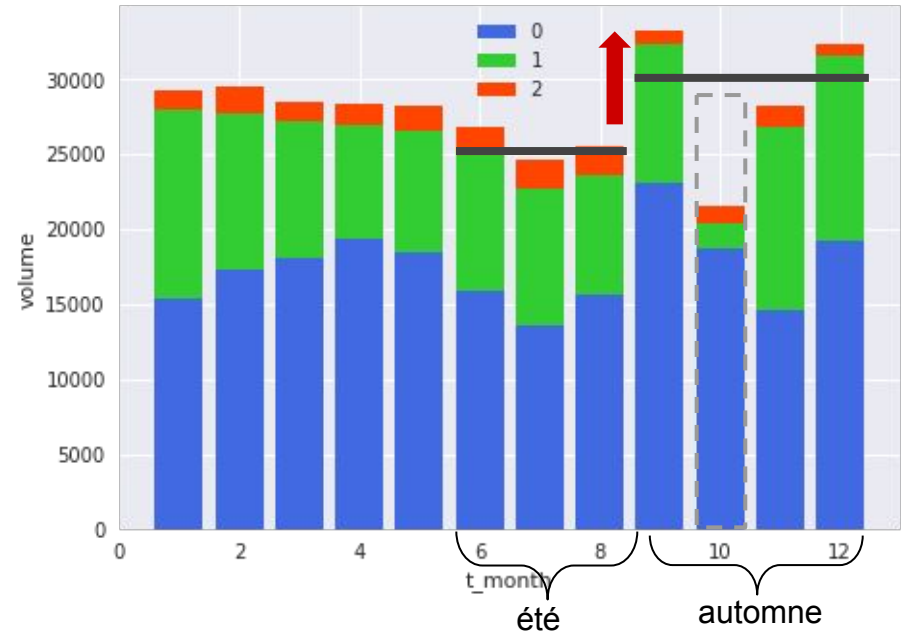
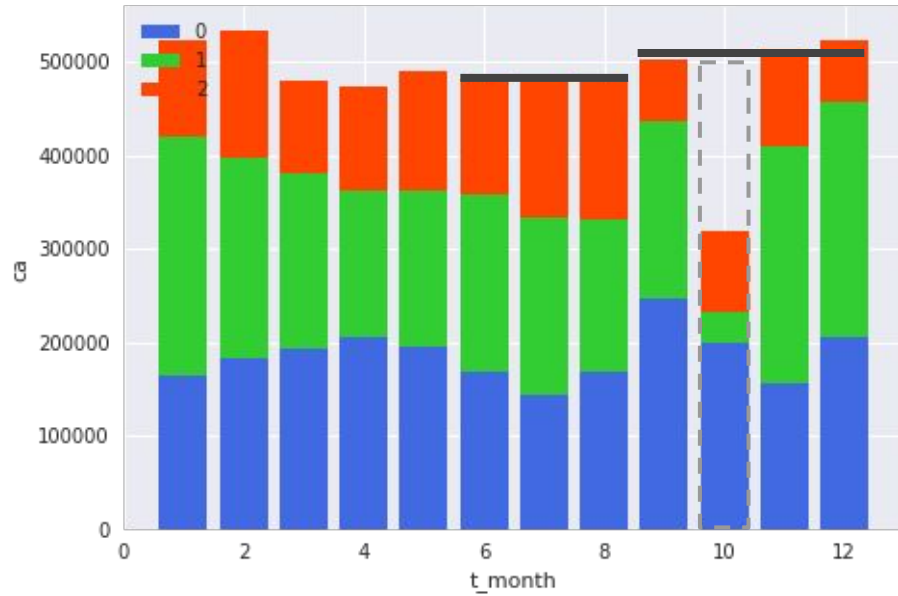




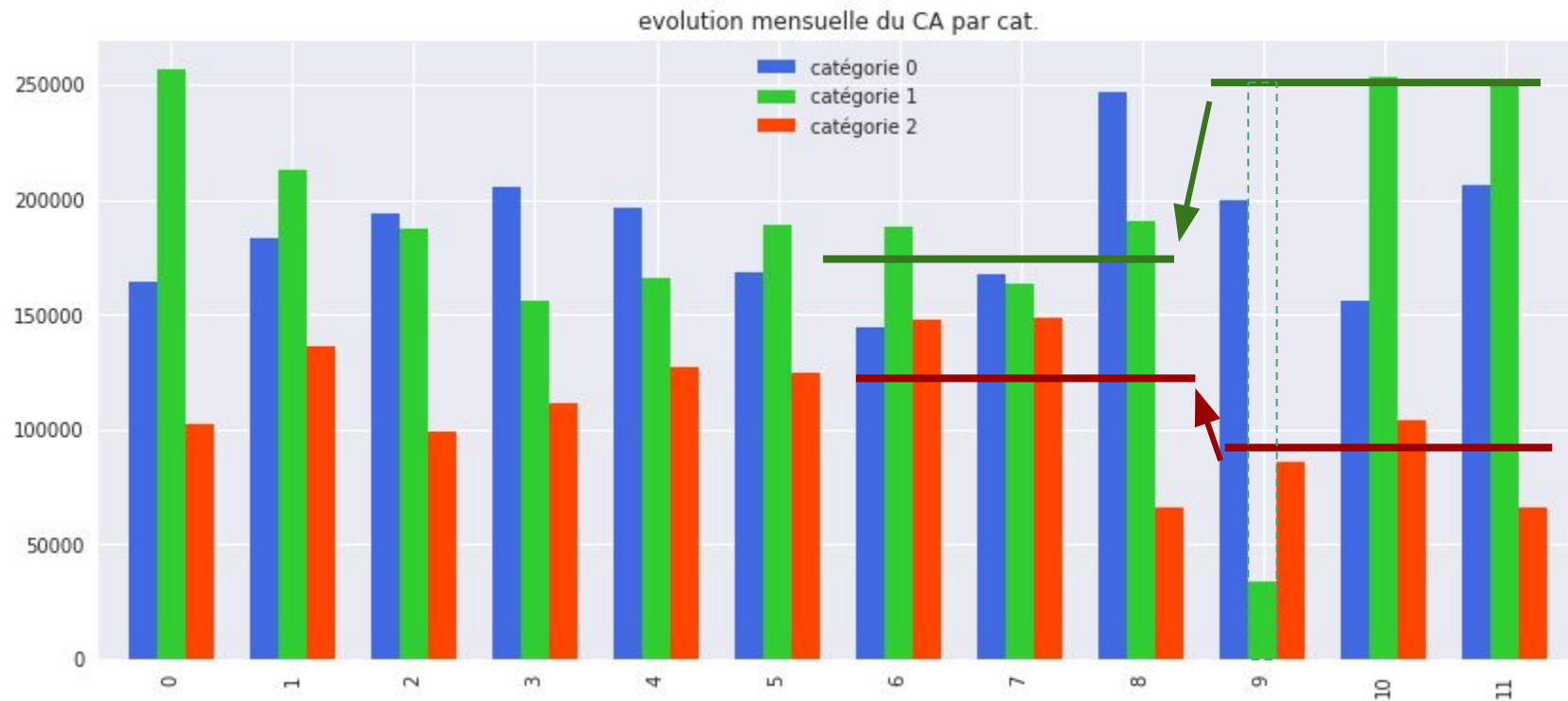
# CA et volume - volume des ventes en Octobre



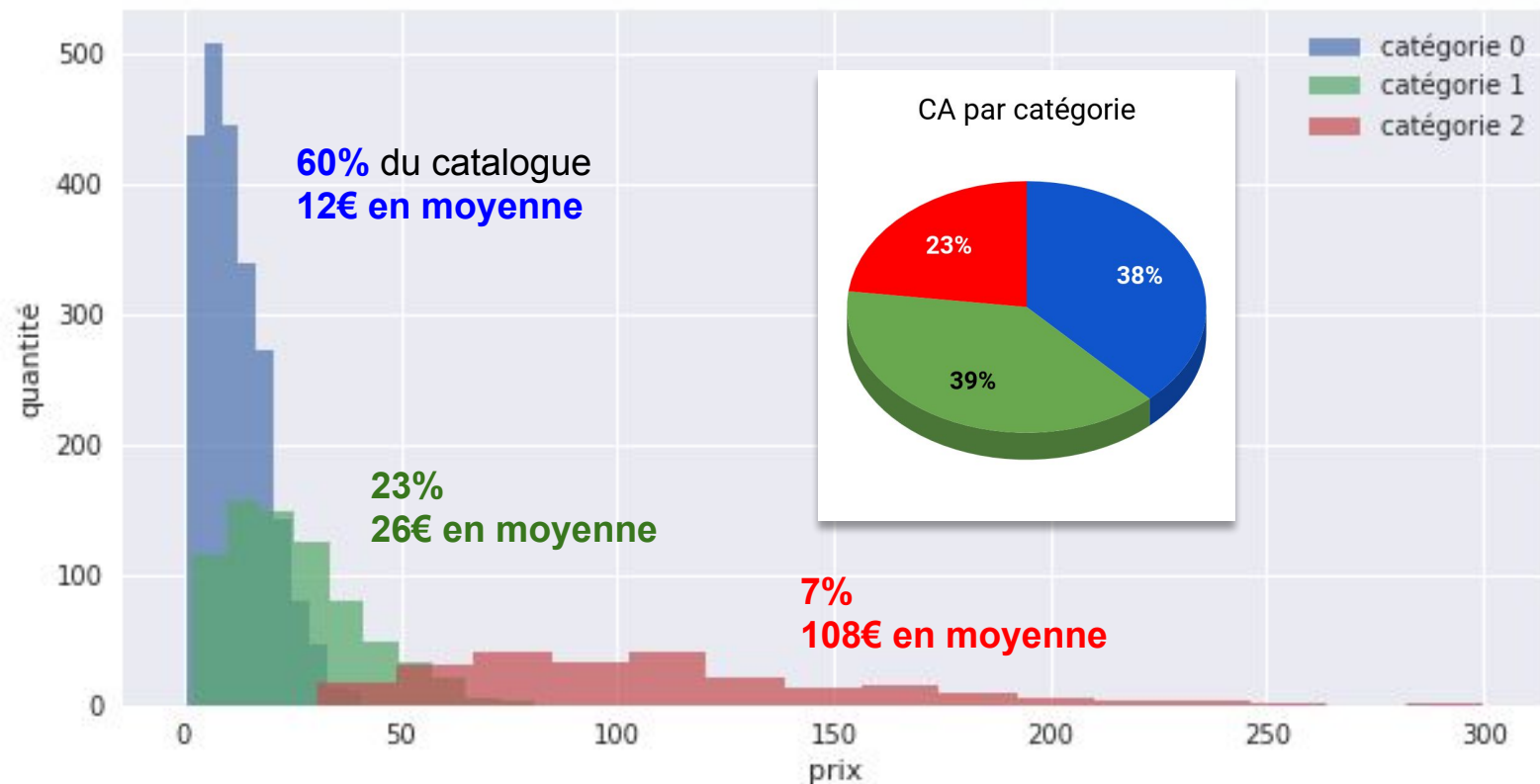
# CA et volume - mois



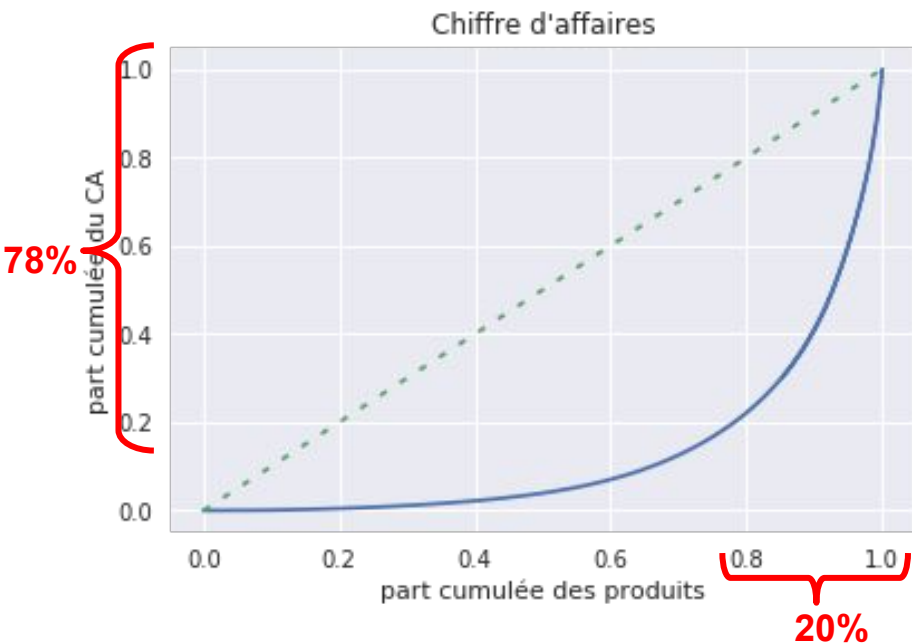
# CA et volume - produits régulateurs



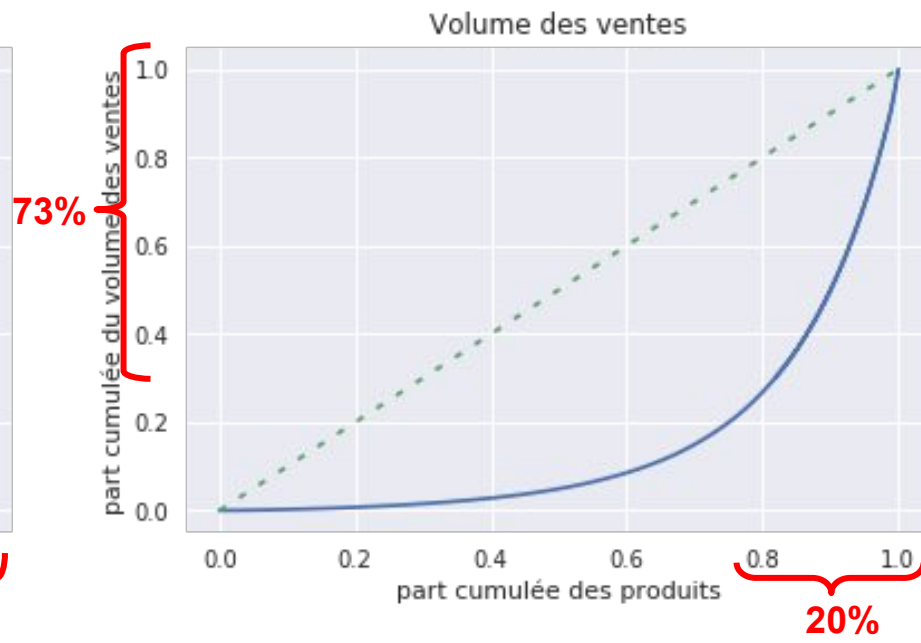
# Produits - distribution des prix par catégories



# Produits - concentration des ventes

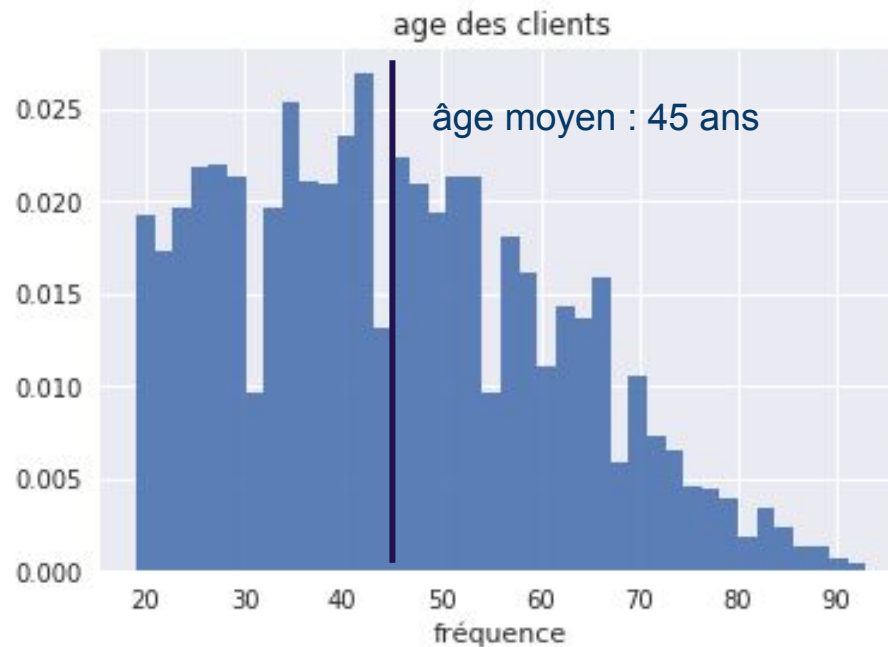
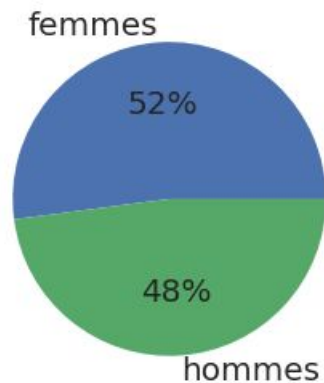


$\text{gini} = 0.74$



$\text{gini} = 0.69$

# Clients - sexe et âge



Proche de la **population française**

# 65%

des produits acheté par un client  
sont de la **même catégorie**

→ **Suggestions** sur page produits

→ marketing **axé par catégorie**

## Clients spéciaux

**4** clients ont dépensé chacun plus de **70 000 €**

**7%** du chiffre d'affaires

→ accompagnement personnalisé



# Panier moyen

panier moyen : **34 €**

panier article : entre **1 et 3**

suggestions sur la page panier

(en fonction des produits)

# Sommaire

## **Données de ventes :**

Présentation

Nettoyage

Analyses

**Corrélations**

# Corrélation entre le sexe et la catégorie ?



## Test du $\chi^2$

$H_0$  : le sexe du client et la catégorie de produit acheté n'est pas corrélée

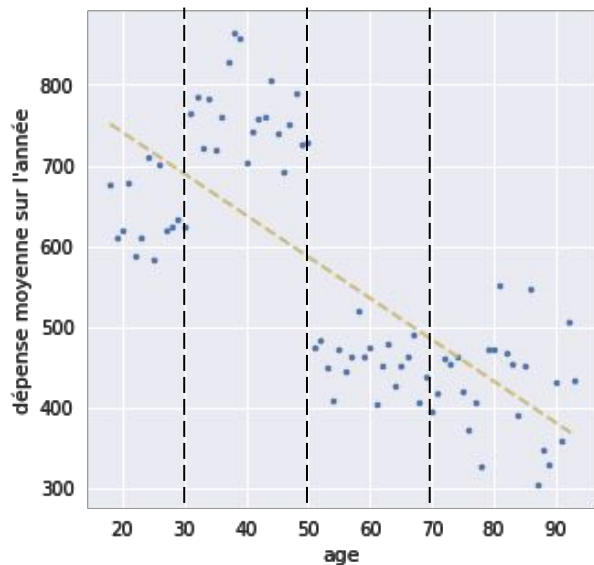
$H_1$  : le sexe du client et la catégorie de produit acheté est corrélée

$$\chi^2 = 10.20$$

$p$  - valeur : 0.0061

# Corrélation entre l'âge du client et ..

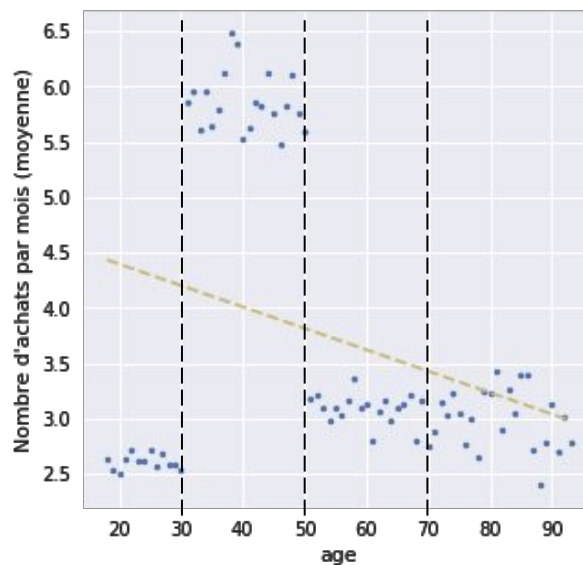
la dépense sur l'année



$$R^2 = 0.56$$

$p - \text{valeur} \approx 0$

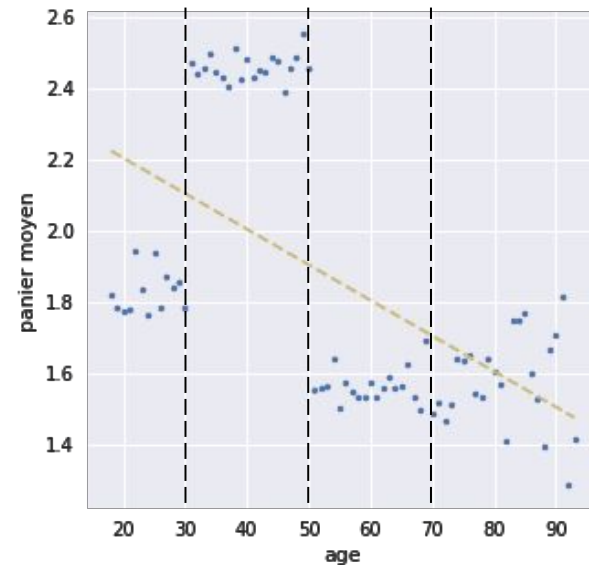
La fréquence d'achat



$$R^2 = 0.10$$

$p - \text{valeur} \approx 0$

Le panier moyen

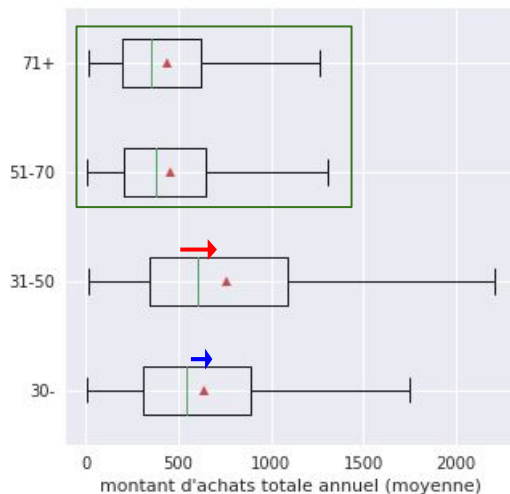


$$R^2 = 0.32$$

$p - \text{valeur} \approx 0$

# Corrélation entre l'âge du client et ..

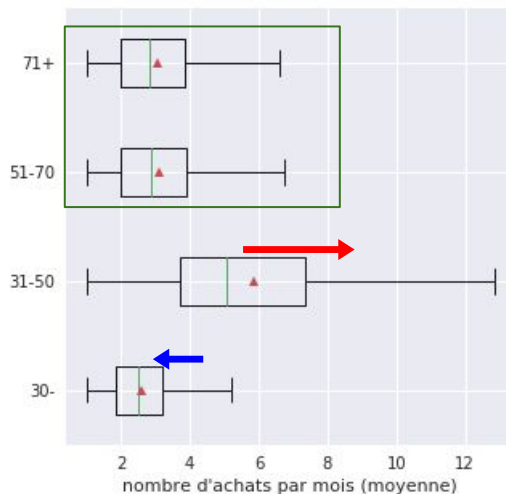
la dépense sur l'année



$F - \text{valeur} = 267$   
 $p - \text{valeur} \approx 0$

$$\eta^2 = 0.085$$

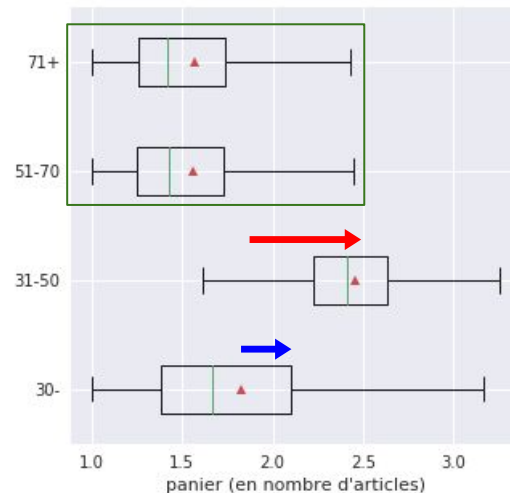
La fréquence d'achat



$F - \text{valeur} = 98.45$   
 $p - \text{valeur} \approx 0$

$$\eta^2 = 0.33$$

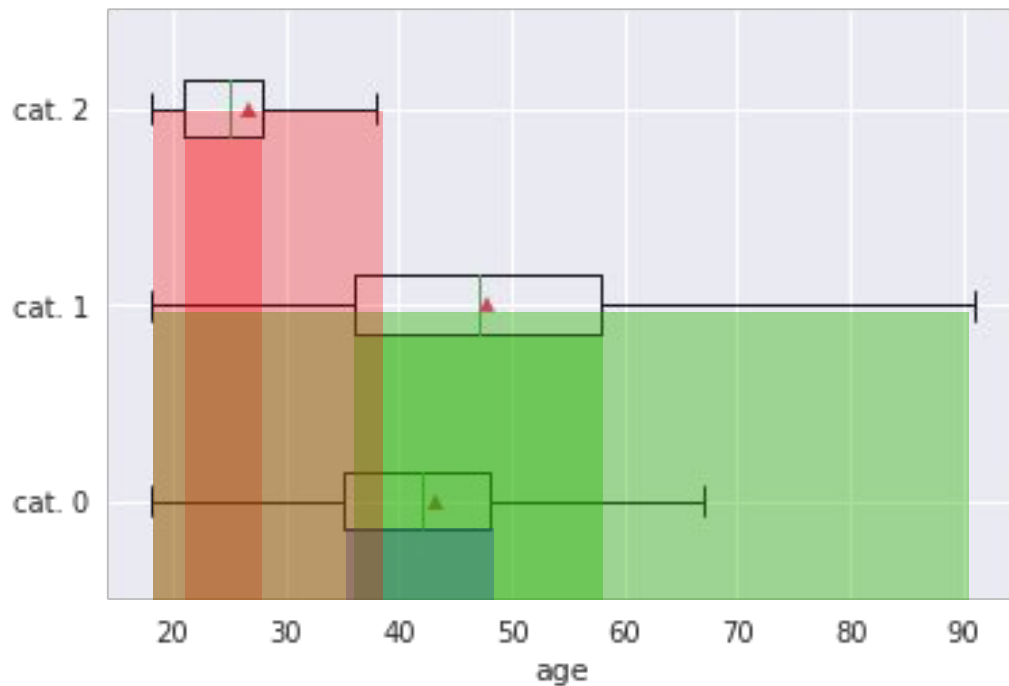
Le panier moyen



$F - \text{valeur} = 267$   
 $p - \text{valeur} \approx 0$

$$\eta^2 = 0.39$$

# Corrélation entre l'âge du client et la catégorie de produits acheté



$F - \text{valeur} = 19992$

$p - \text{valeur} \approx 0$

$$\eta^2 = 0.27$$

# Conclusion : Profils clients en fonction de l'âge

**31-50 ans**

**40%** des clients - **47%** du CA

cette année, ils ont dépensé en moyenne **750 €**

*achètent plus de produits et plus souvent (**6 par mois**)*

des produits de **catégorie 0 puis 1**

# Conclusion : Profils clients en fonction de l'âge

**moins de 30 ans**

**26%** des clients - **30%** du CA

cette année, ils ont dépensé en moyenne **650 €**

*achètent **moins fréquemment** que les 31-50 ans,  
mais ce sont des produits de la **catégorie 2** (les **plus chers**)*



## Conclusion : Profils clients en fonction de l'âge

**plus de 50 ans**

**33%** des clients - **23%** du CA

cette année, ils ont dépensé en moyenne **500 €**

*achètent aussi fréquemment que les 30 ans et moins,*

***exclusivement*** des produits de **catégorie 1**

## *Produits associés* (en plus)

- proximité des produits selon le goût des clients
- Quelle est la chance qu'un client achète **B** sachant qu'il a acheté **A** ?

	pid_1	pid_2	pid_2 sachant pid_1
<b>8</b>	2_40	2_208	0.636364
<b>978</b>	2_200	2_110	0.600000
<b>1177</b>	2_199	2_102	0.583333
<b>793</b>	2_58	2_102	0.545455
<b>1208</b>	2_195	2_209	0.545455