

Sommaire

- Nettoyage de données
- Analyse
- Statistique(corrélation)
- Différence avec et sans le mois d'octobre

Nettoyage des données

Import de mes 3 fichier csv, (customers, products, transactions)

- Vérification de null().
- Vérification de doublons.
- Supprimer les 'test ' devant les date, puis mise a jour du format.
- Créations de colonnes année, mois, jour.
- Suppression client 'ct_0' 'ct_1' car doublons est valeurs négatif.
- Catégories et prix manquants pour le produits '0_2245'.
- Remplacements du prix par moyenne et catégorie par '99' pour le produits '0_2245'.

Analyse de ma dataframe

11 colonnes pour 679 332 ligne

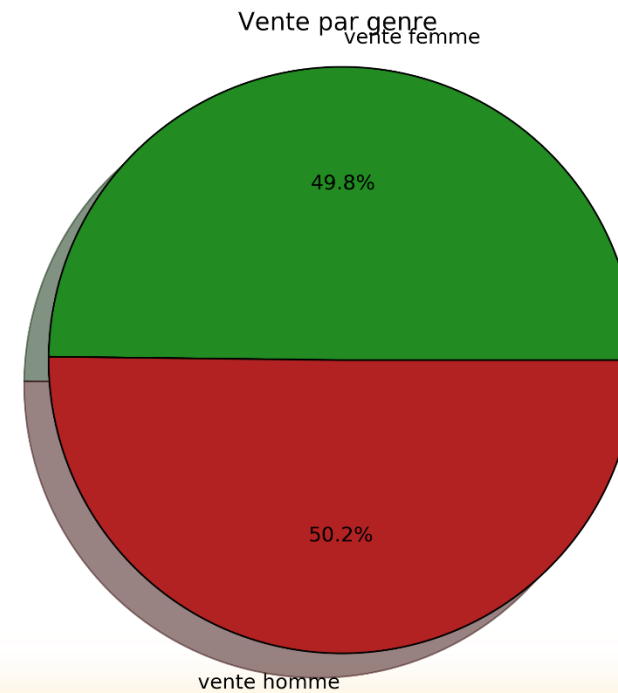
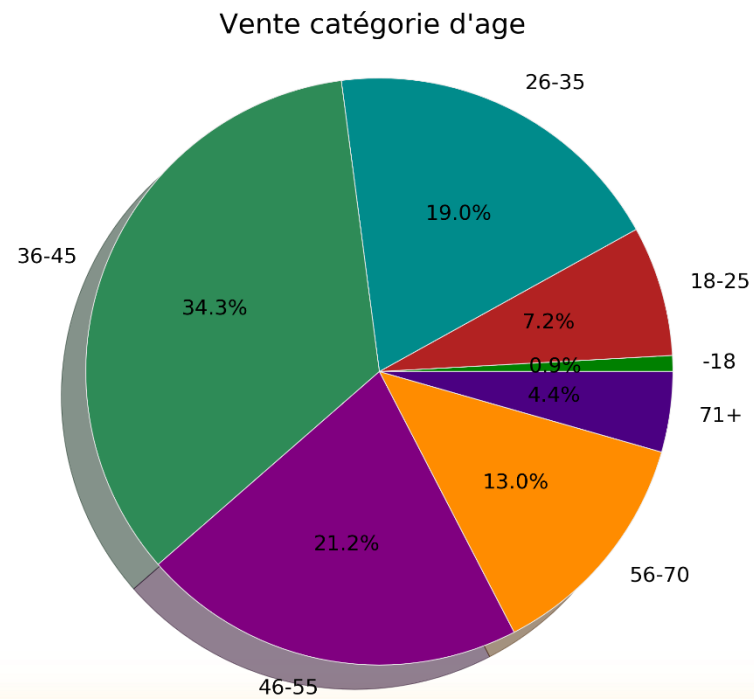
	id_prod	date	session_id	client_id	year_month_day	year_buy	Ymonth	sex	birth	price	categ
0	0_1518	2022-05-20 13:21:29	s_211425	c_103	2022-05-20	2022	2022-05	f	1986	4.18	0.0
1	1_251	2022-02-02 07:55:19	s_158752	c_8534	2022-02-02	2022	2022-02	m	1988	15.99	1.0
2	0_1277	2022-06-18 15:44:33	s_225667	c_6714	2022-06-18	2022	2022-06	f	1968	7.99	0.0
3	2_209	2021-06-24 04:19:29	s_52962	c_6941	2021-06-24	2021	2021-06	m	2000	69.99	2.0
4	0_1509	2023-01-11 08:22:08	s_325227	c_4232	2023-01-11	2023	2023-01	m	1980	4.99	0.0

- Analyse chiffre d'affaires
- Analyse produits
- Analyse clients
- Corrélations
- Zoom mois d'octobre 2021

Analyse chiffre d'affaires

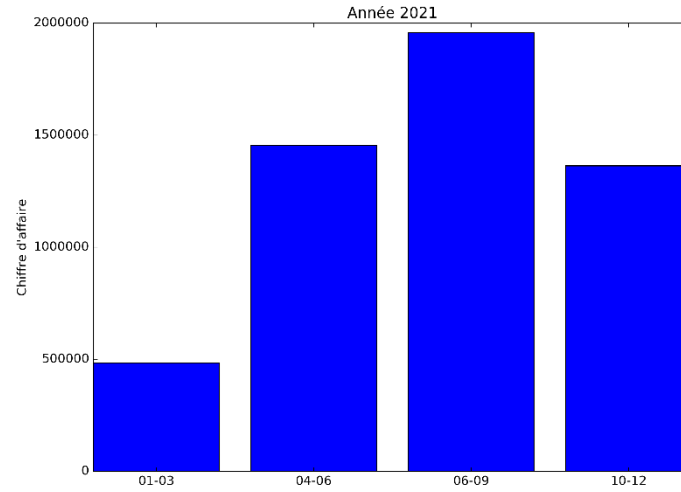
- 4,772,345€ en 2021
- 6,110,776€ en 2022
- 974,465€ en 2023 (année en cours sur le projets)

Chiffre d'affaires total 11,857,586€.

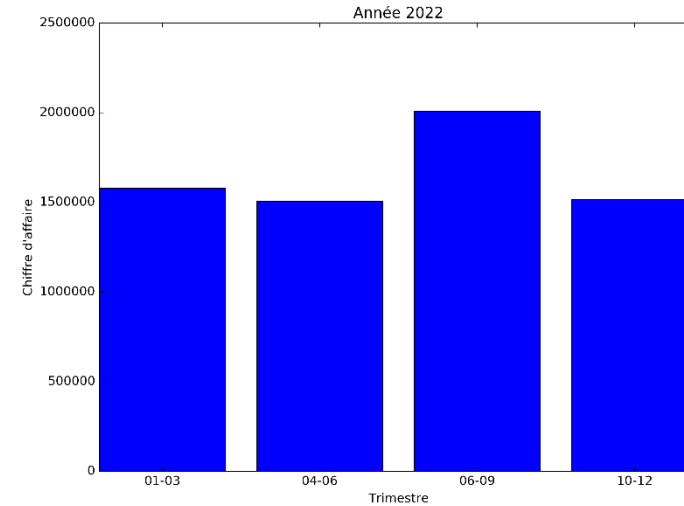


Chiffre d'affaires par ans et trimestre

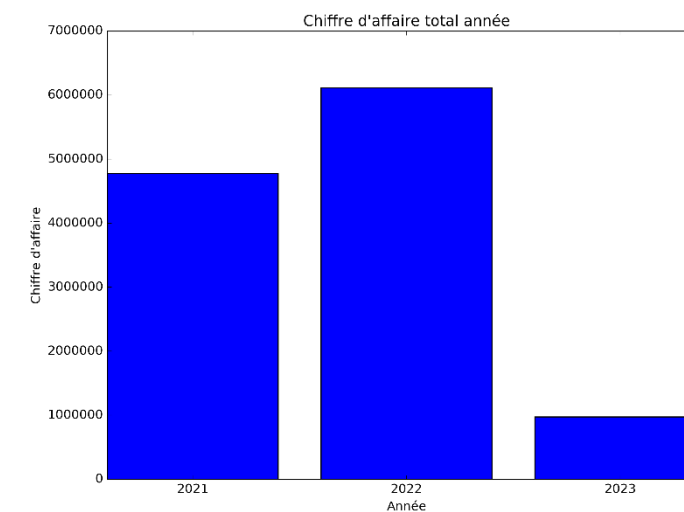
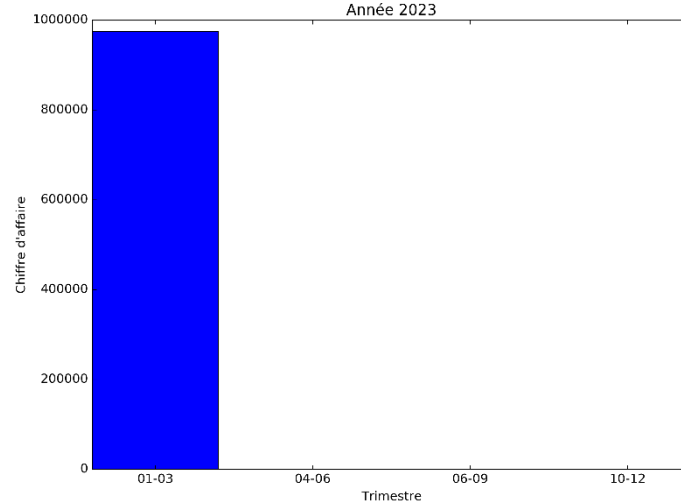
- 4,772,345€ en 2021.



- 6,110,776€ en 2022

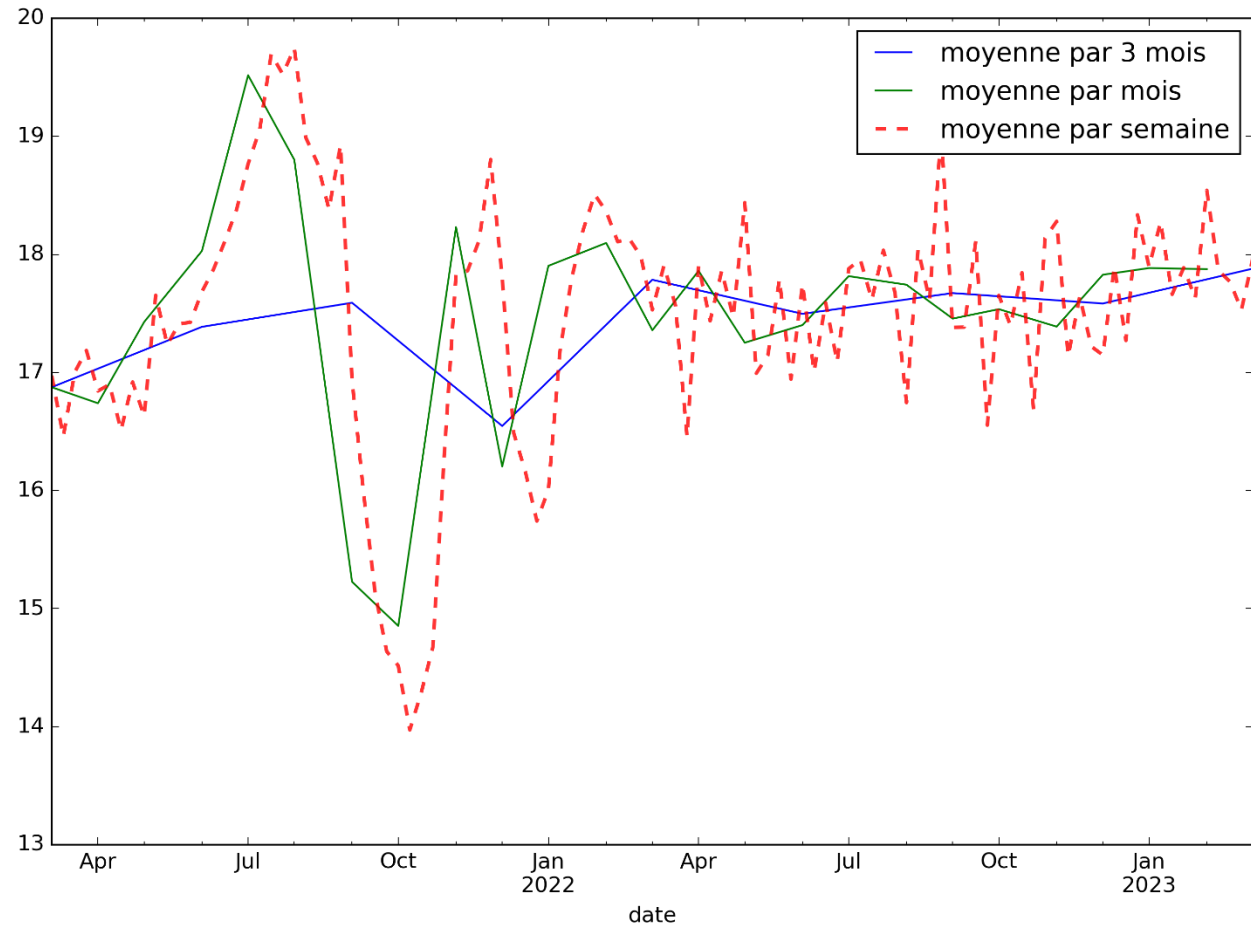


- 974,465€ en 2023
(année en cours sur le projets)

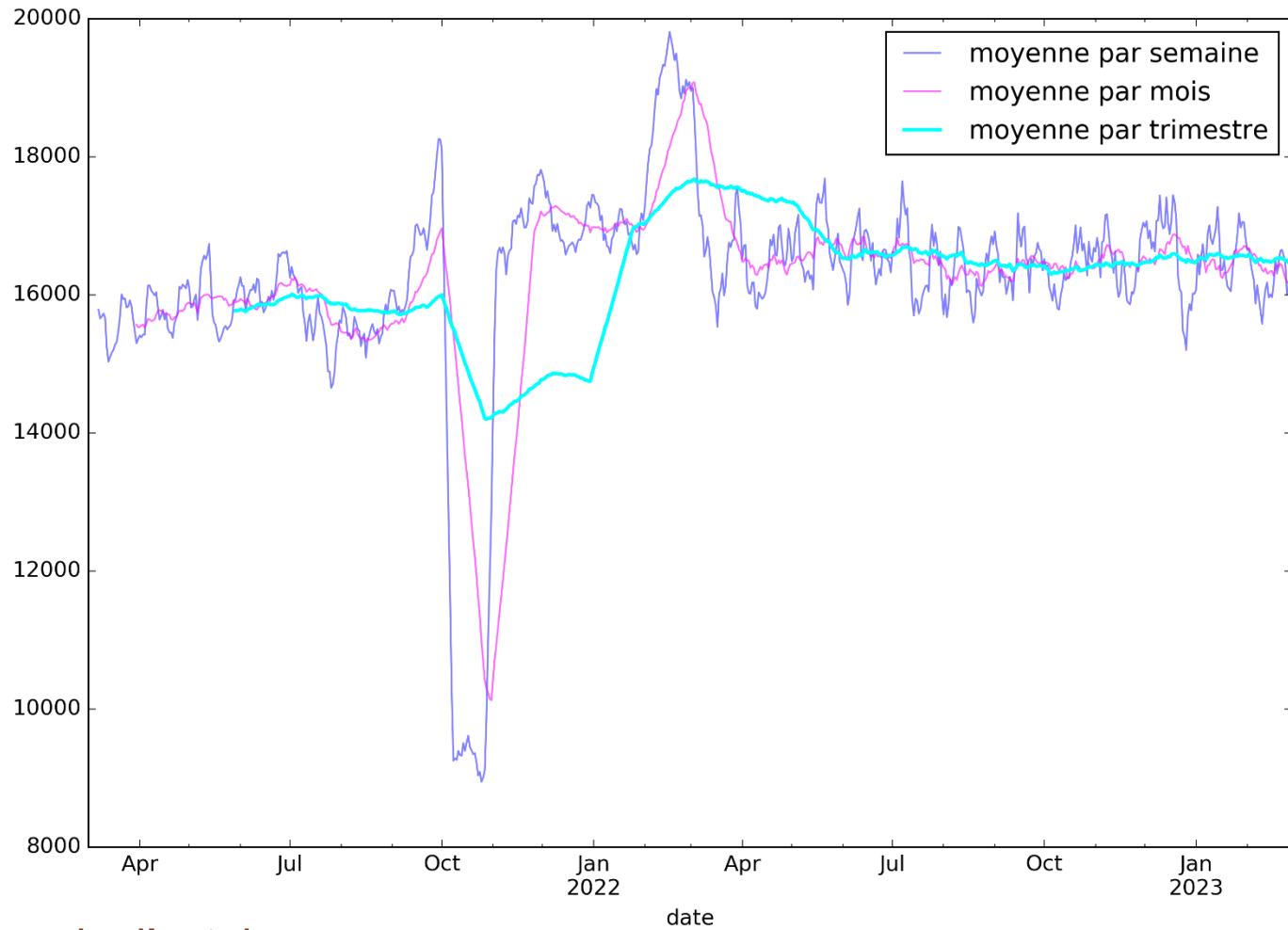


Chiffre
d'affaire total
11,857,586€.

Chiffre d'affaires moyen sur le temps (resample)

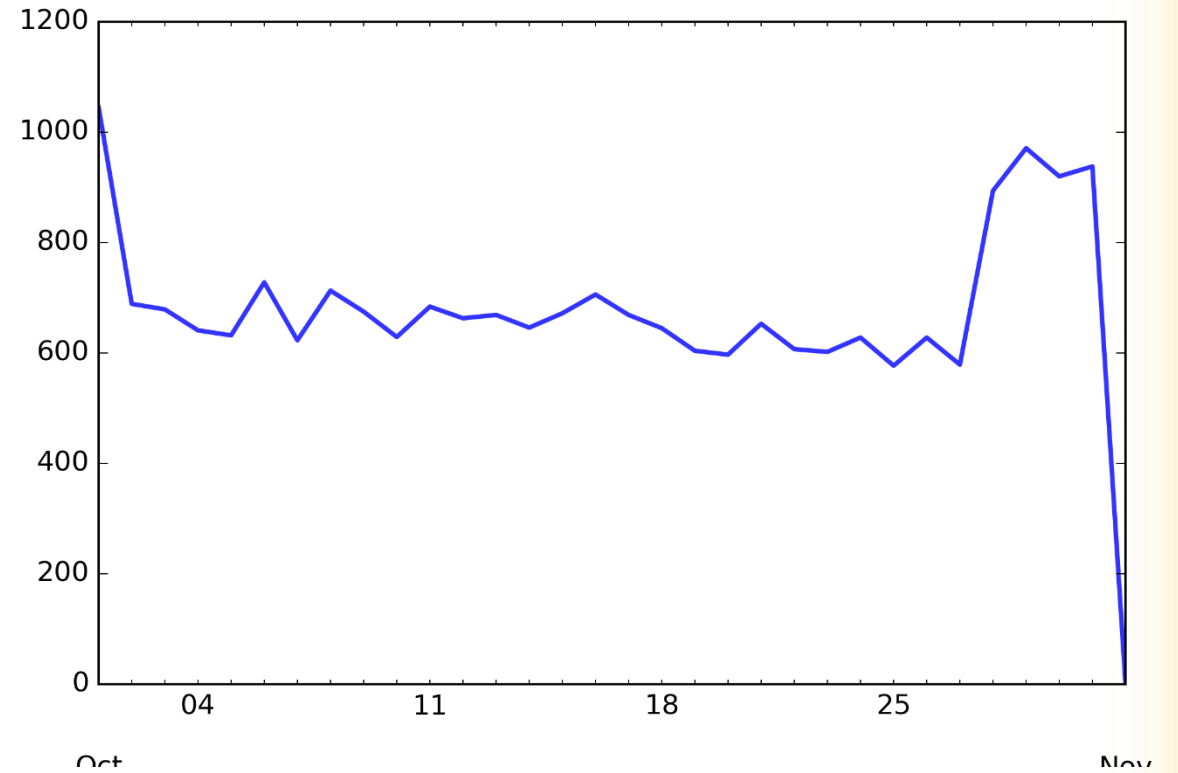
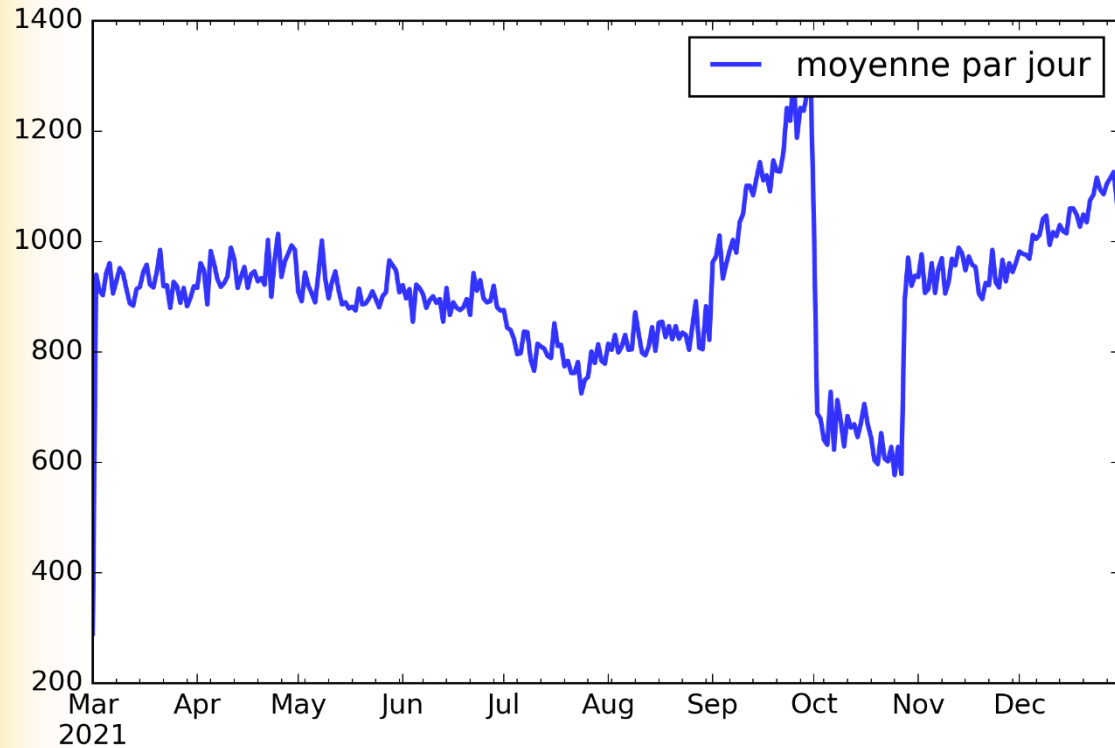


Moyenne mobile (rolling)



Chute anormale sur le mois d'octobre

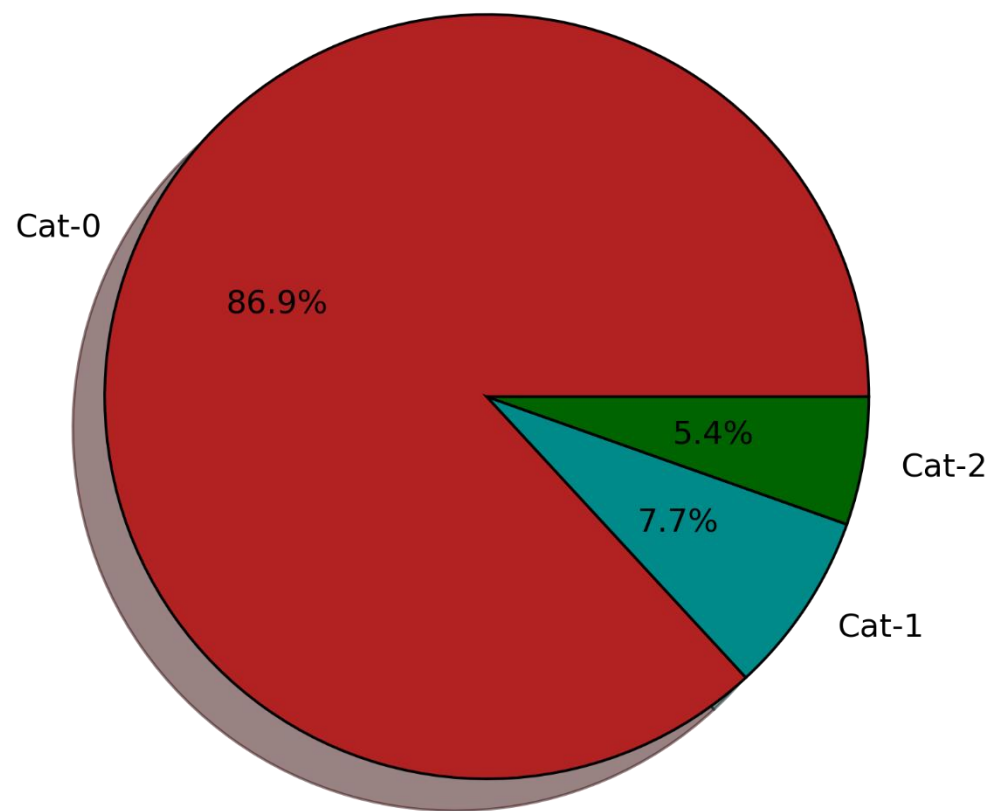
Zoom sur le mois d'octobre



Comparaison du nombre de vente de l'année 2021 et sur octobre

Zoom sur le mois d'octobre

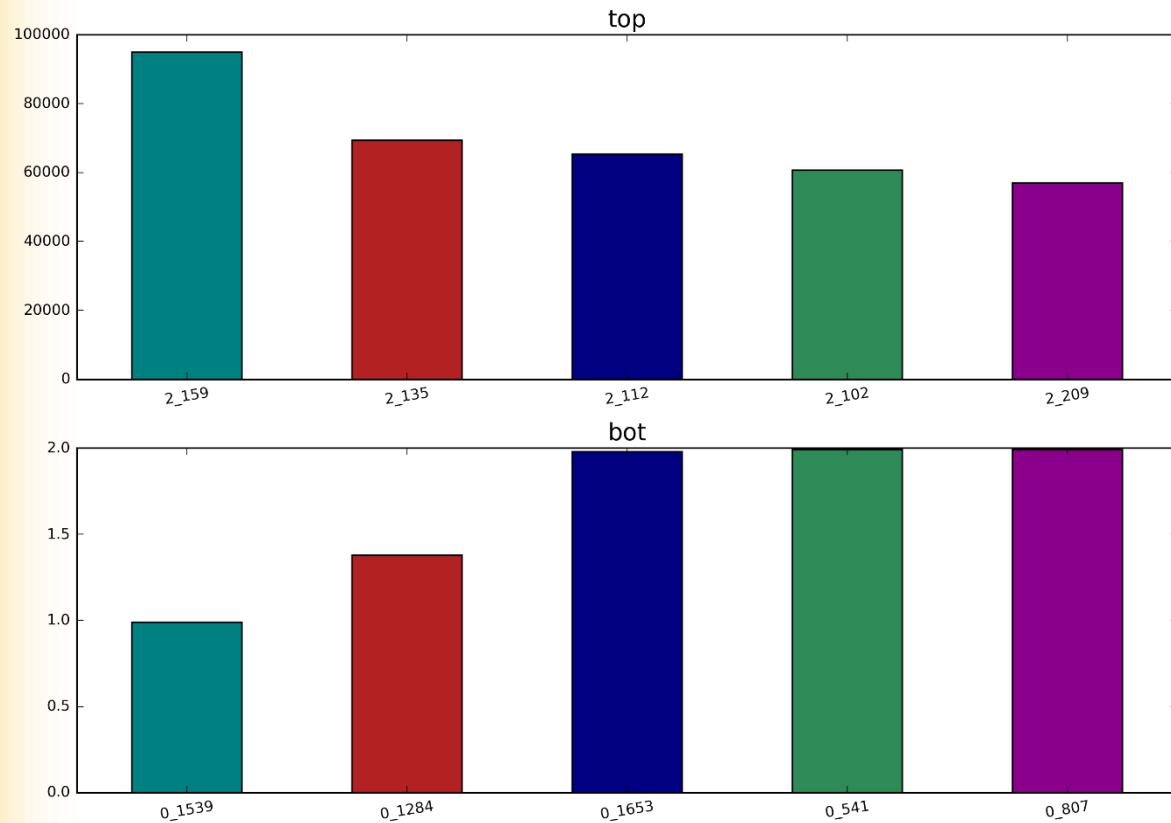
Vente catégorie de produits



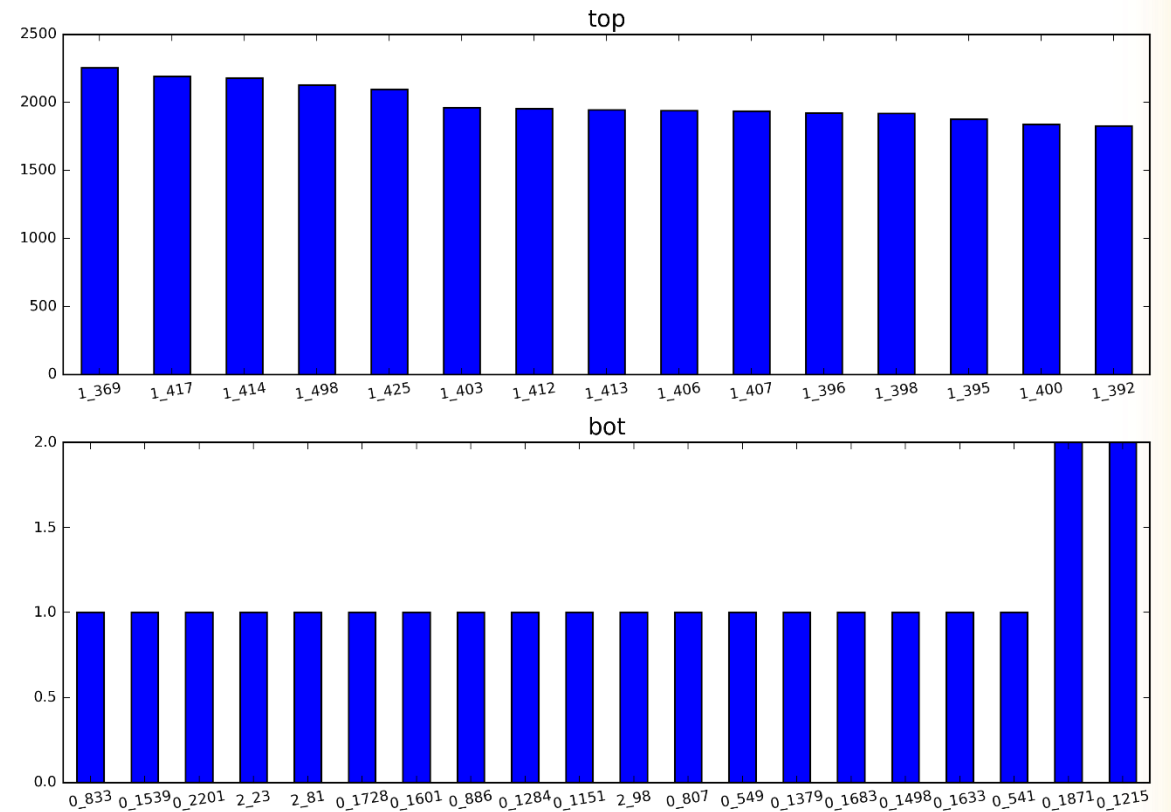
Vente des catégorie du mois d'octobre

Analyse meilleurs produits

Top et bot 5 du chiffre d'affaire produits

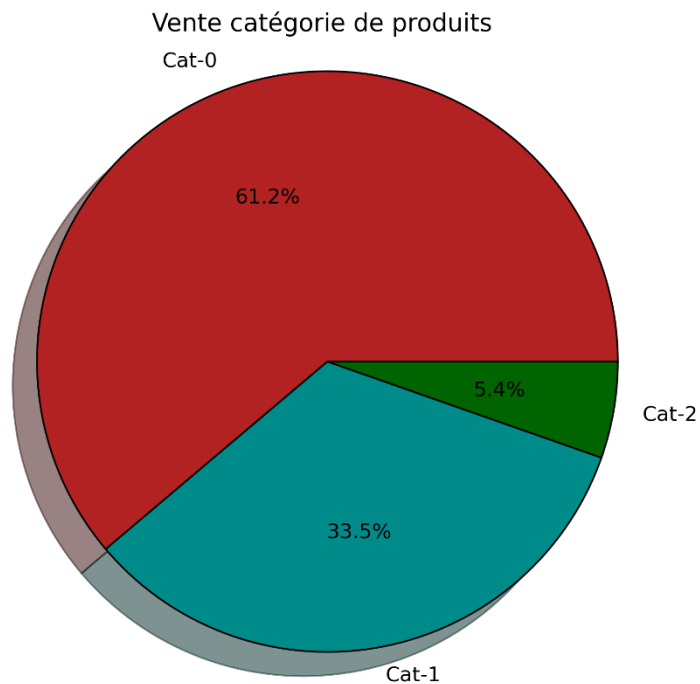


Top et bot du nombre de vente produits



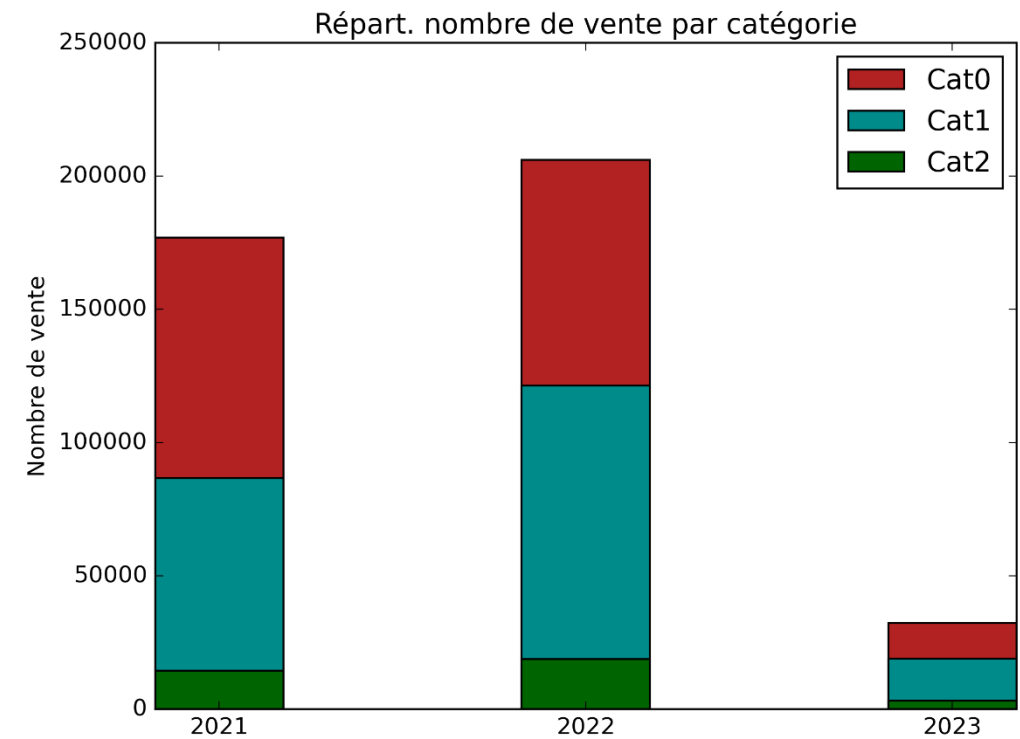
Analyse catégorie produits

Vente produits all time



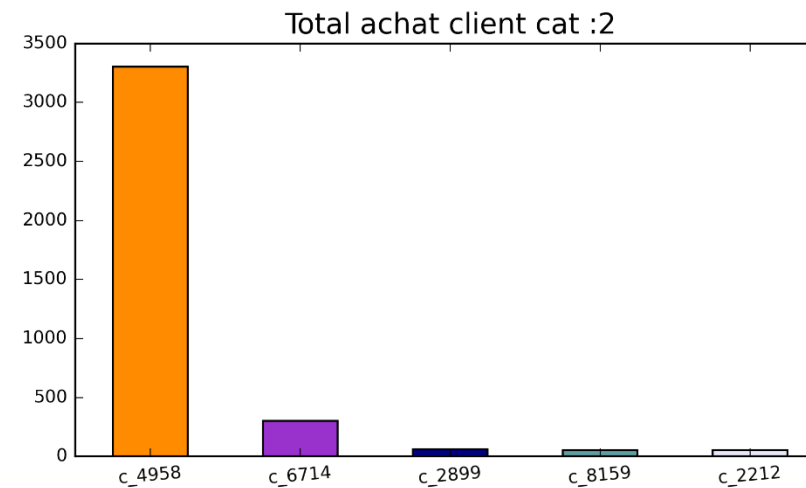
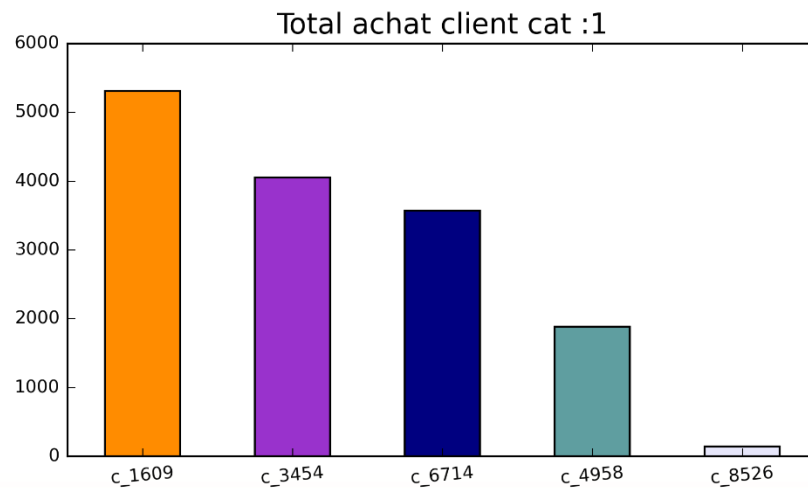
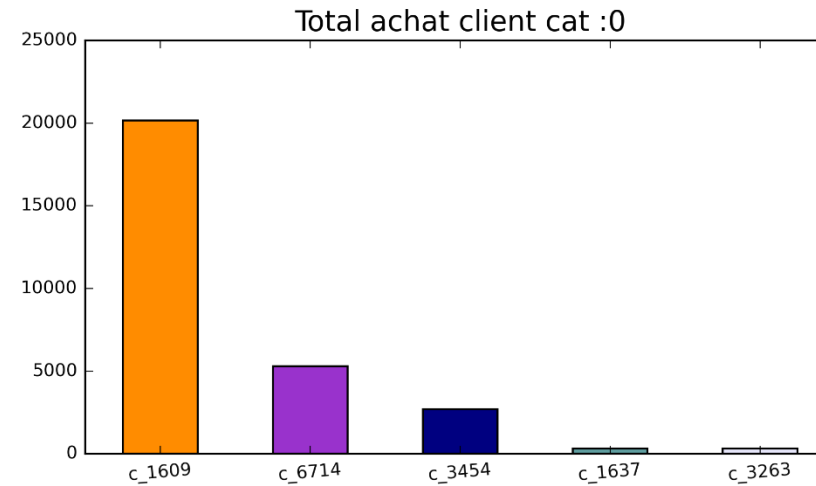
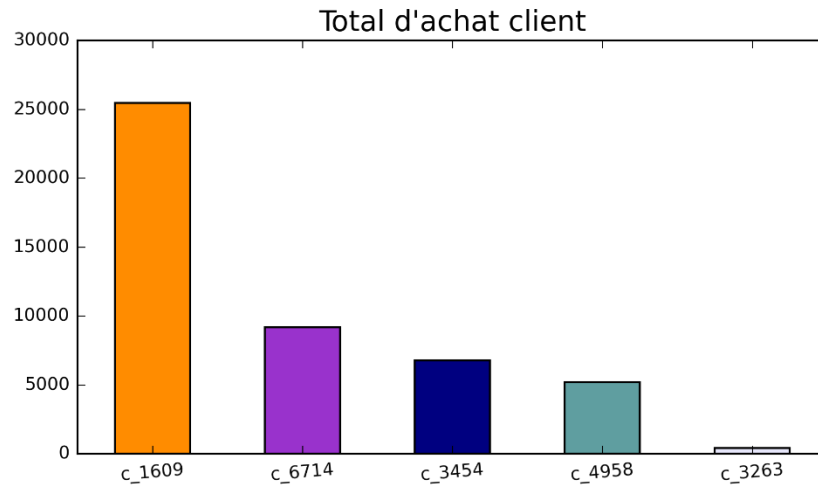
Catégorie 0 : 415,459 | Catégorie 1 : 227,169 | Catégorie 2 : 36,483

Répartition nombre de vente produits



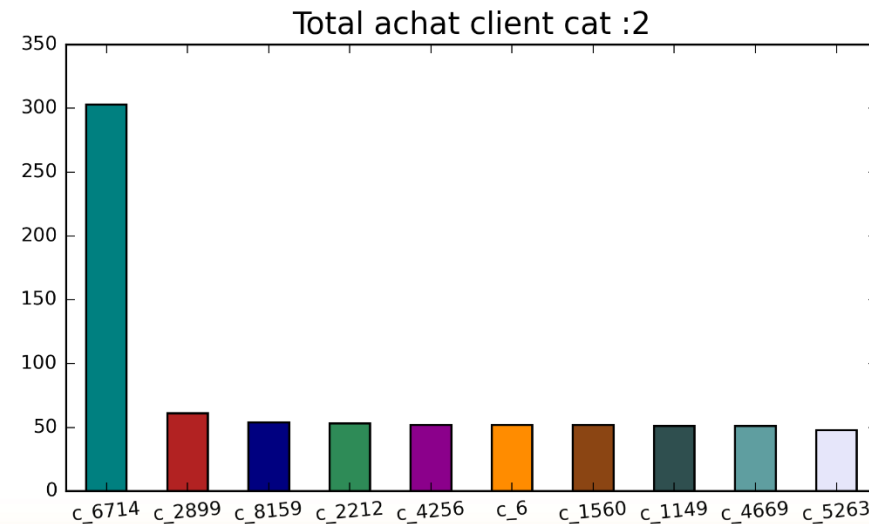
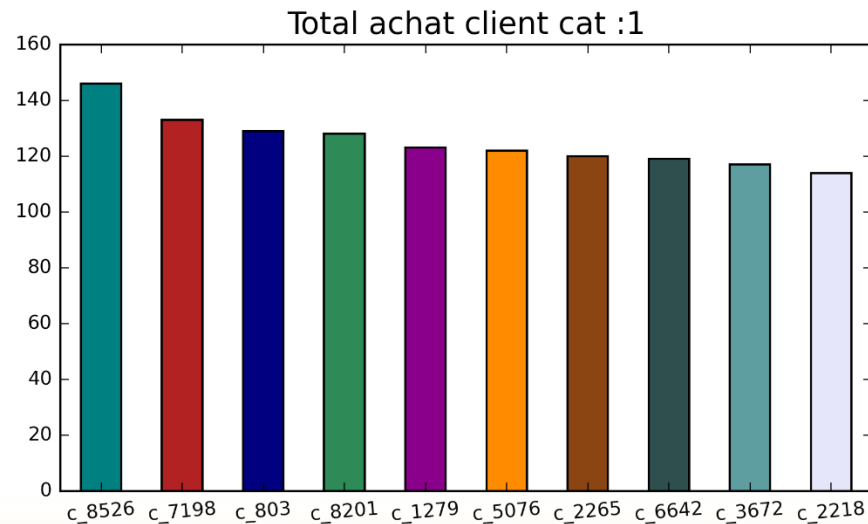
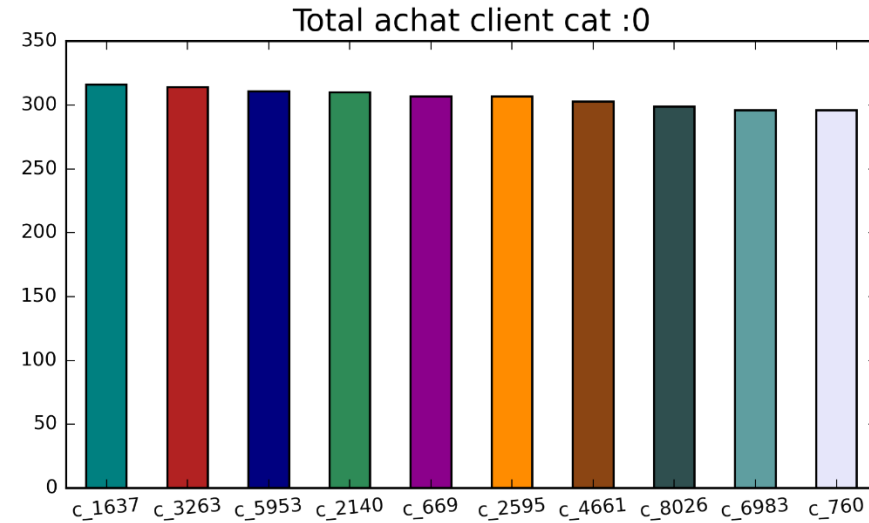
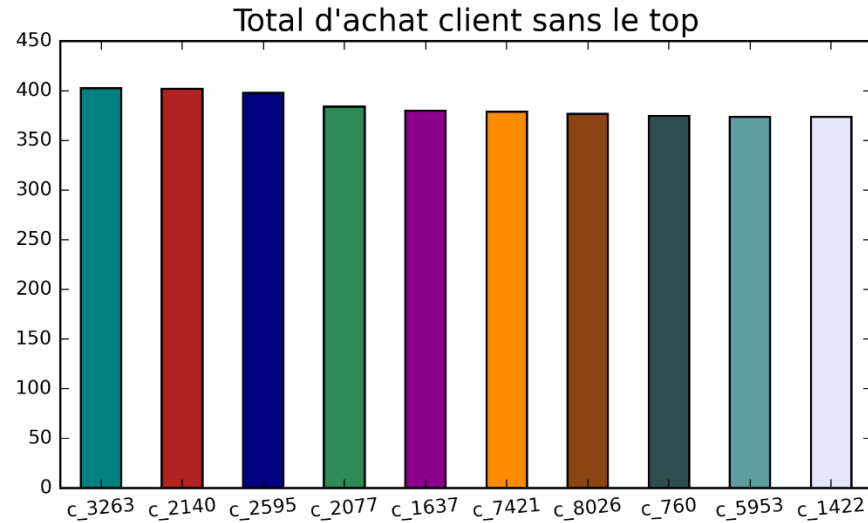
Analyse meilleurs clients

Avec gros clients



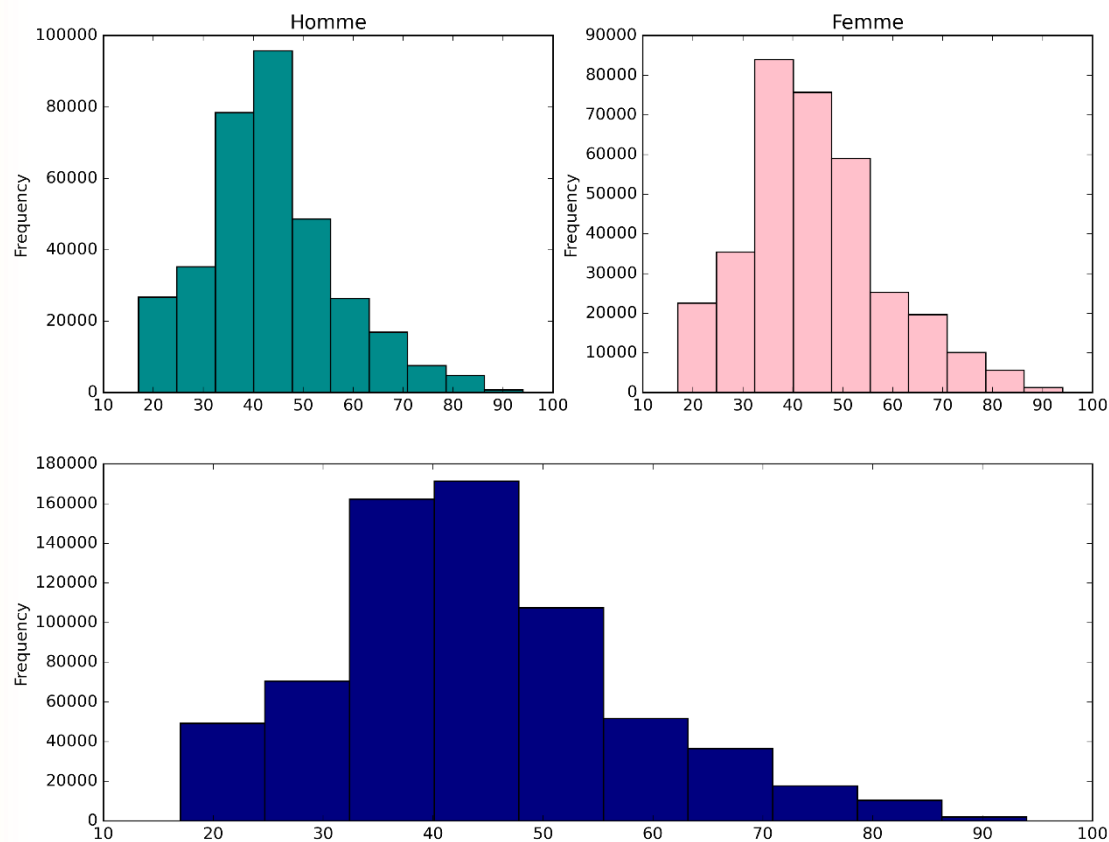
Analyse meilleurs clients

Sans gros clients

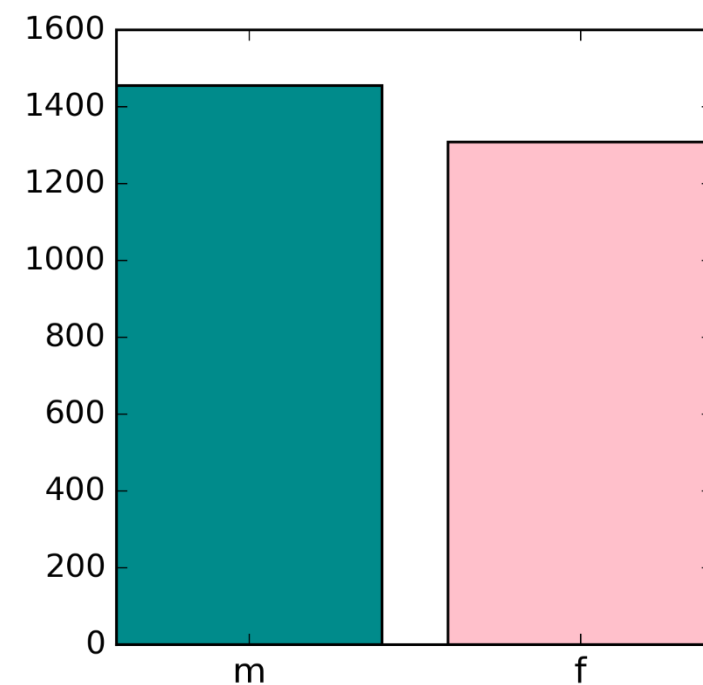


Analyse Clients

Age Femme Homme Total



Panier moyen par genre des clients

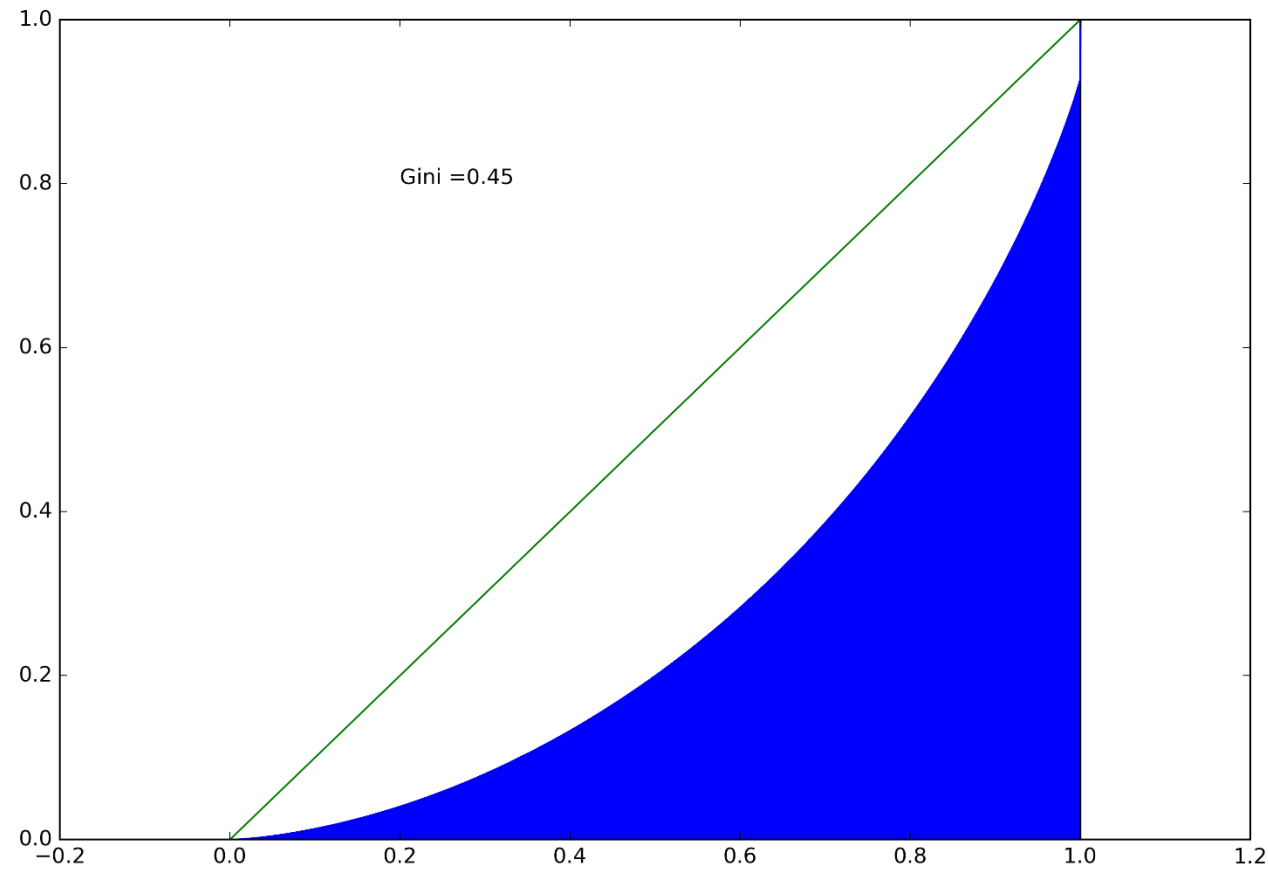


4121 <-- Homme / femme --> 4479

Panier moyen (CA total homme / nb client): 1455

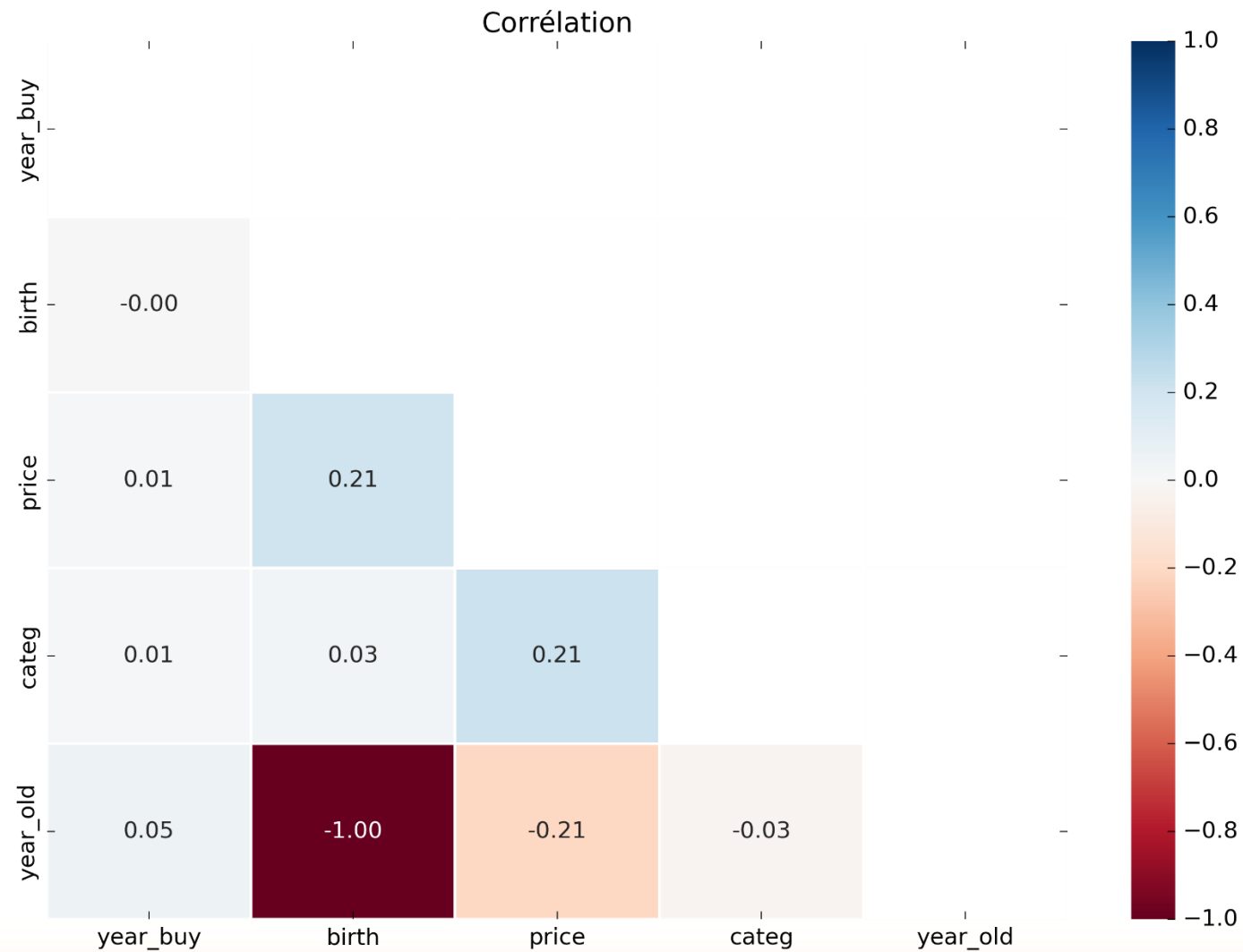
Panier moyen (CA total femme / nb client): 1309

Courbe de Lorenz client chiffre d'affaire



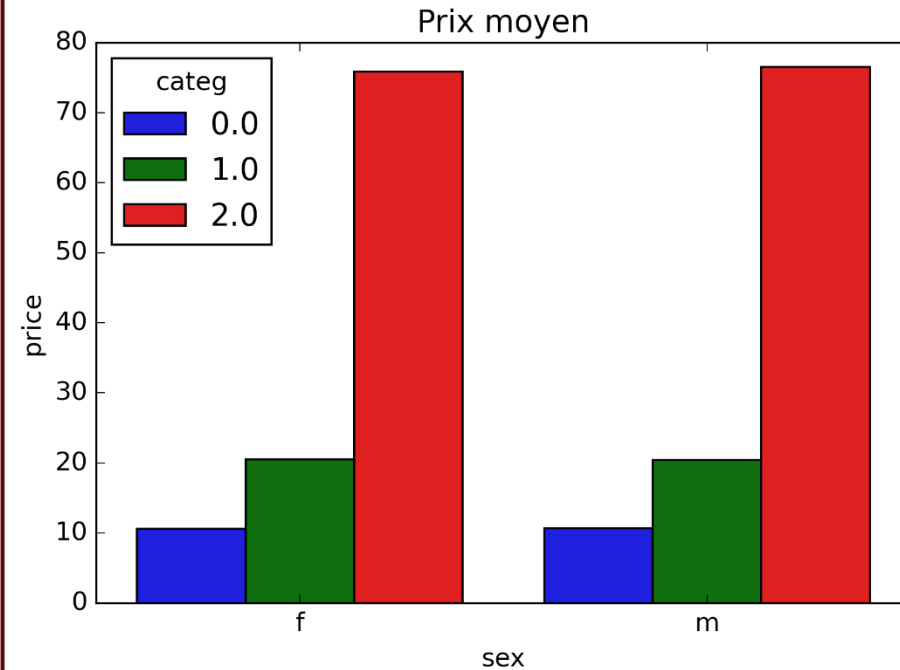
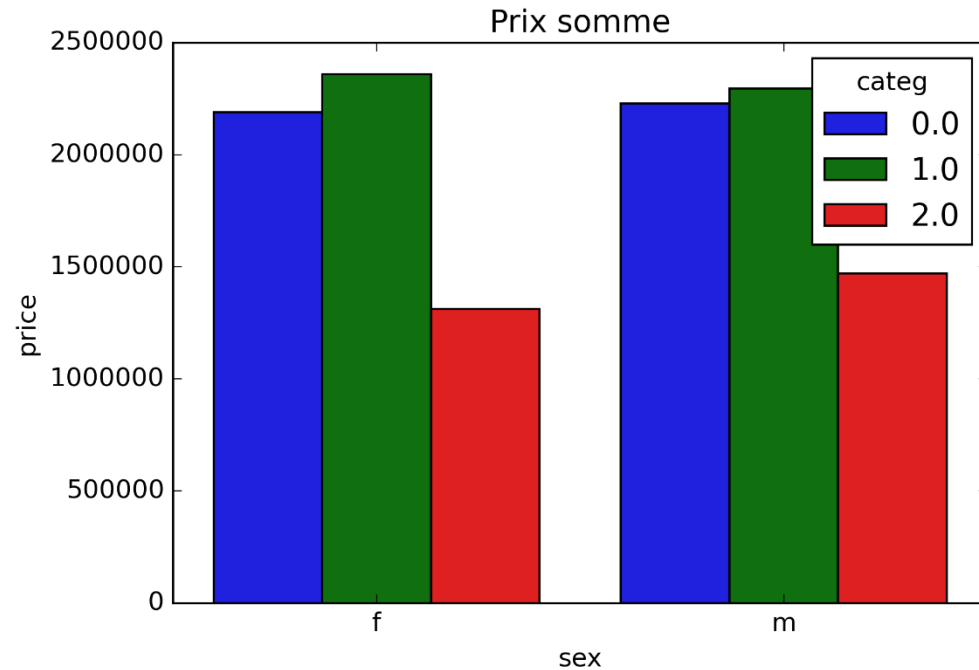
Avec 0,45 il y a une inégalité mais elle n'est pas extrême.

Heatmap corrélations



Statistique (Corrélations)

Corrélations Sexe et Catégorie



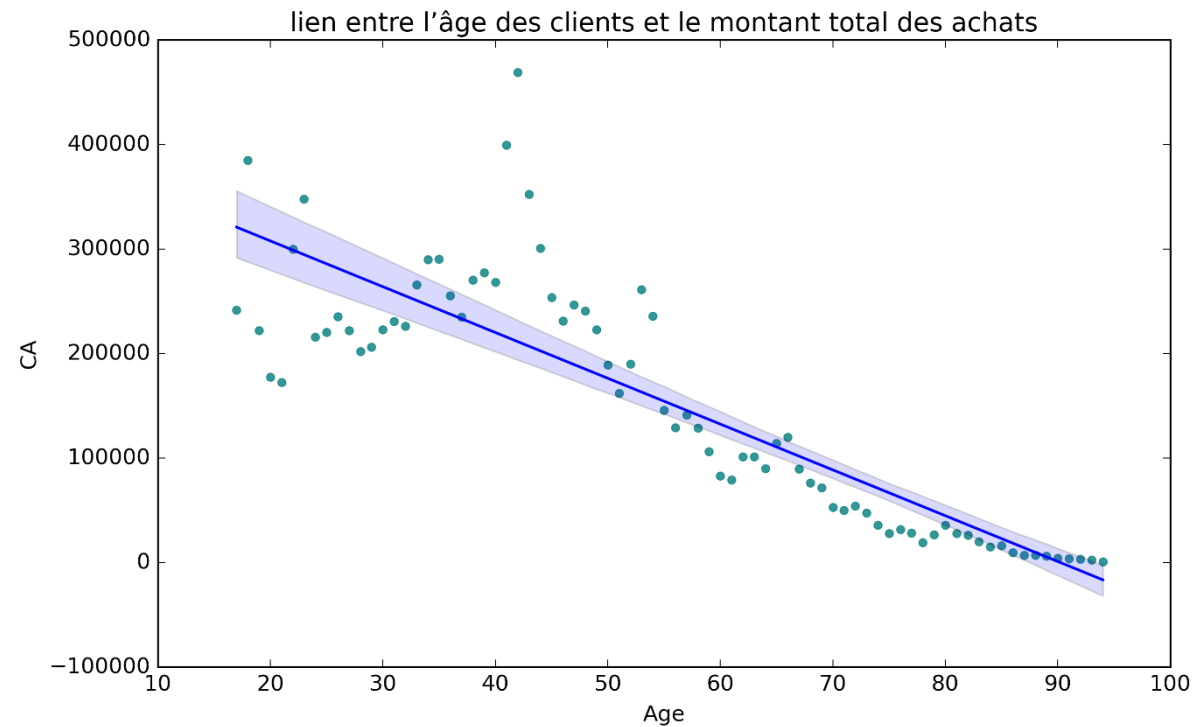
- H_0 : Il y a un liens entre la catégorie de livre acheté et le sexe du clients
- H_1 : Il n'y a pas de un liens entre la catégorie de livre acheté et le sexe du client

Avec une p-value de 0,99 je rejette H_0 .

Résultat test khi-2

```
(0.0002166347889539628,  
0.9998916884716402,  
2,
```

Corrélations âge et montant total achat

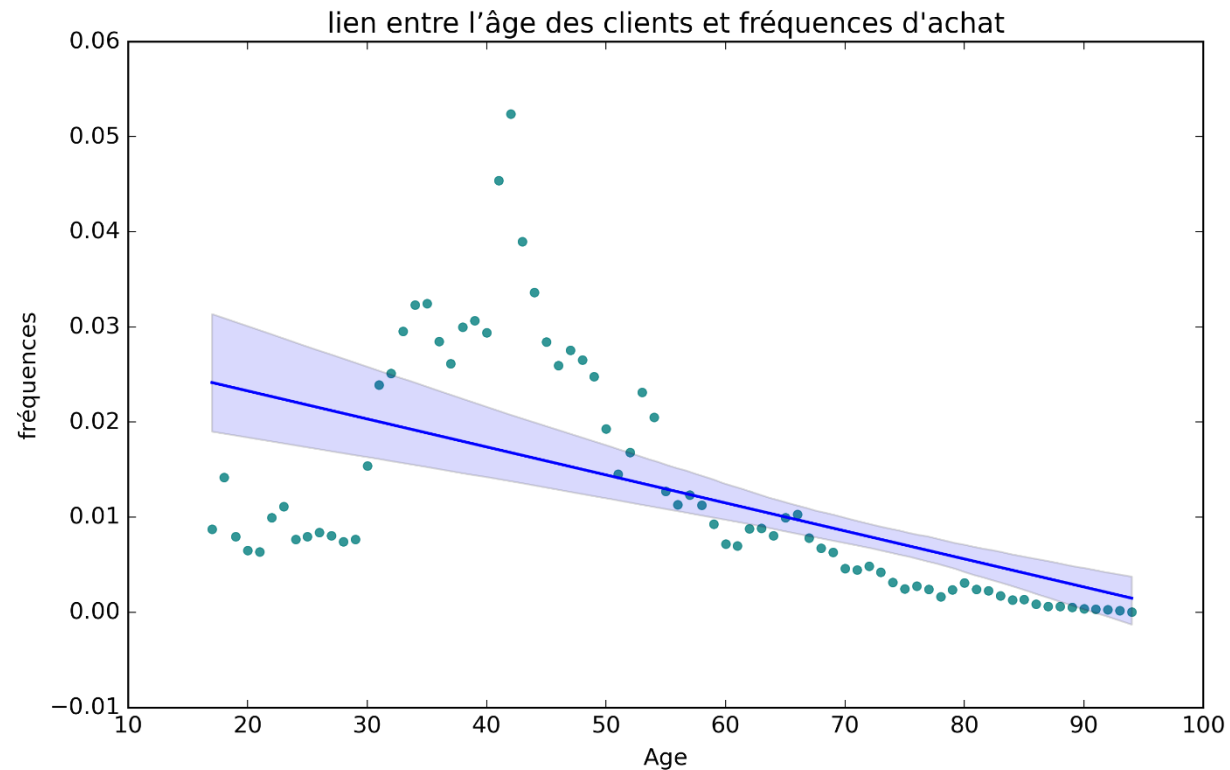


- H_0 : Le montant total des achat n'a pas de liens avec l'âge des clients
- H_1 : Le montant total des achat a un liens avec l'âge des clients

Avec une p-value de -0,85 j'ai une forte corrélation dites inversée.

Je rejette donc H_0 et valide l' H_1 .

Corrélations âge et fréquences d'achat

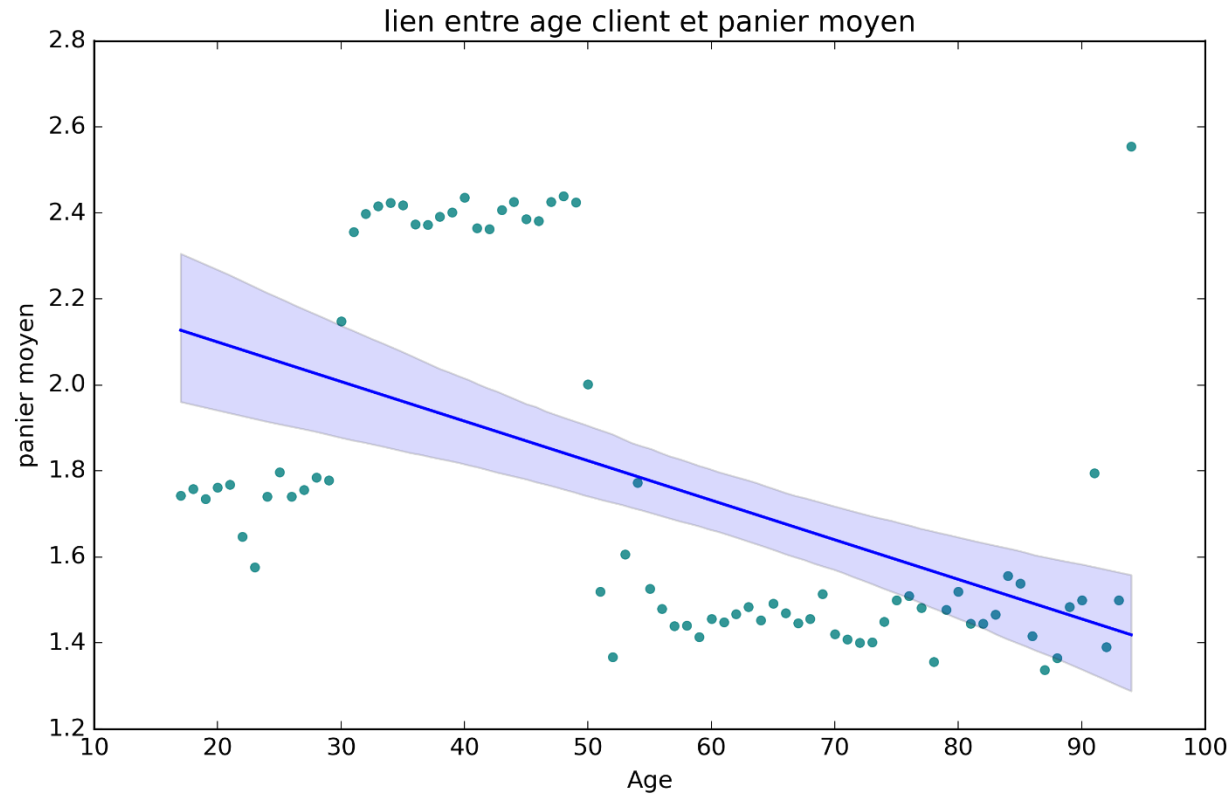


- H_0 : Il n'y a pas de liens entre l'âge et la fréquence d'achat
- H_1 : Il y a un liens entre l'âge et la fréquence d'achat

Avec une p-value de -0,55 j'ai une corrélation dites inversée.

Je rejette donc H_0 et valide l' H_1 .

Corrélations âge et panier moyen

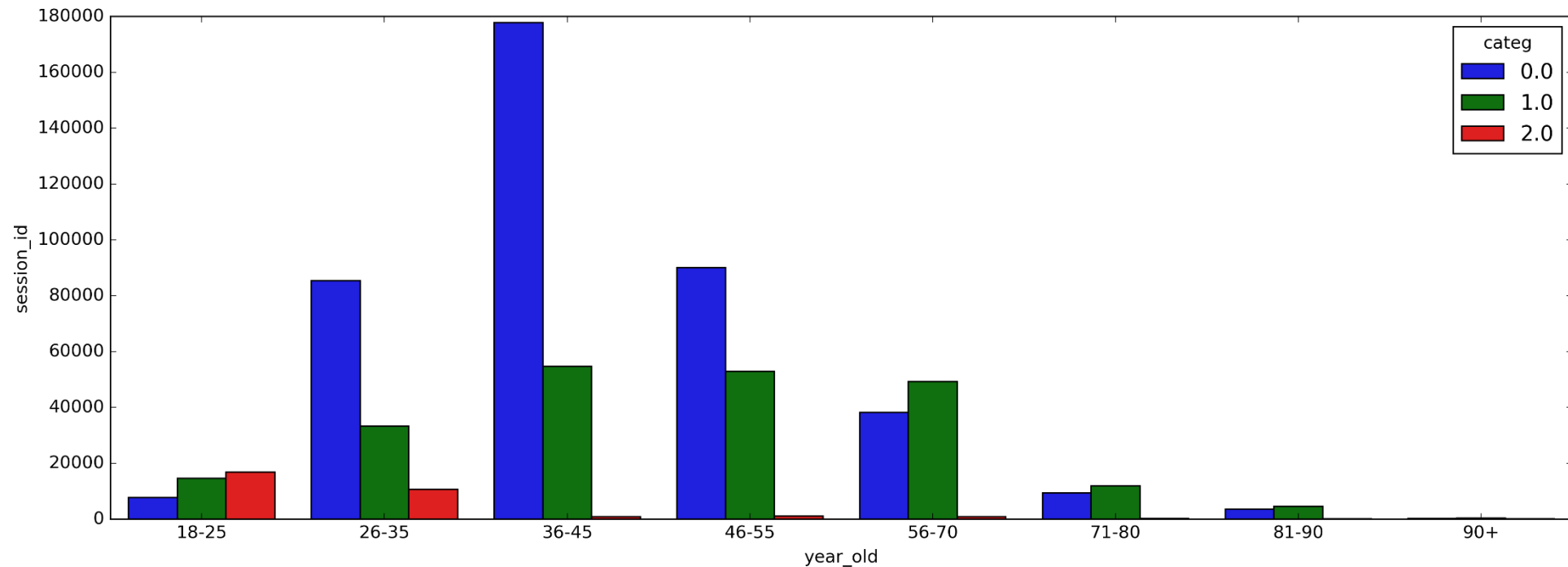


- H_0 : Il n'y a pas de liens entre l'âge et le panier moyen
- H_1 : Il y a un liens entre l'âge et le panier moyen

Avec une p-value de -0,51 j'ai une corrélation dites inversée.

Je rejette donc H_0 et valide l' H_1 .

Corrélations âge et catégorie



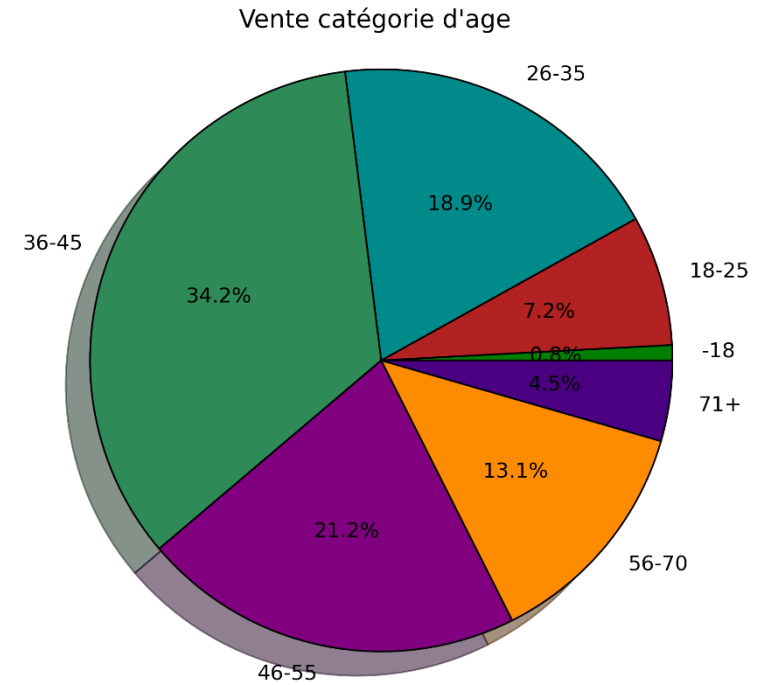
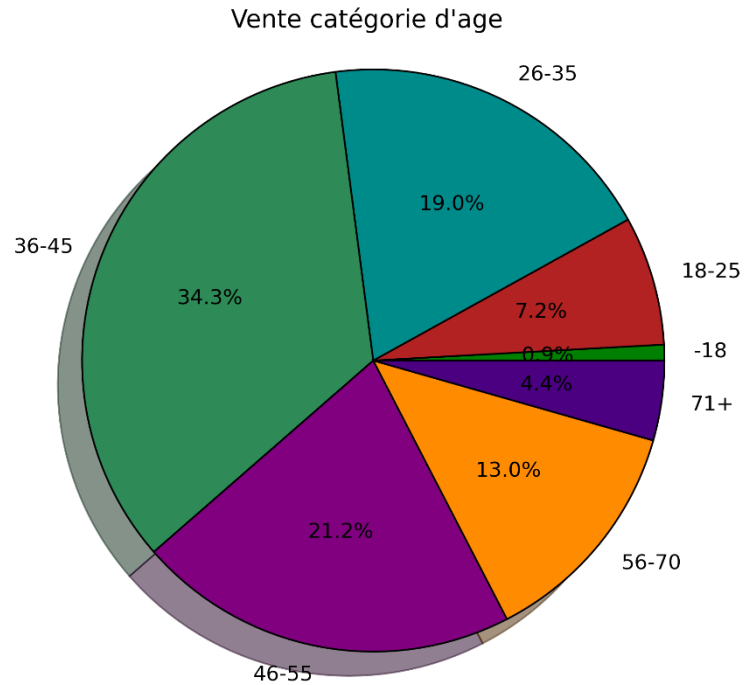
- H_0 : Il n'y a pas de liens entre l'âge et les catégorie
- H_1 : Il y a un liens entre l'âge et les catégorie

Avec une p-value de 0,38 j'ai une corrélation.

Je rejette donc H_0 et valide l' H_1 .

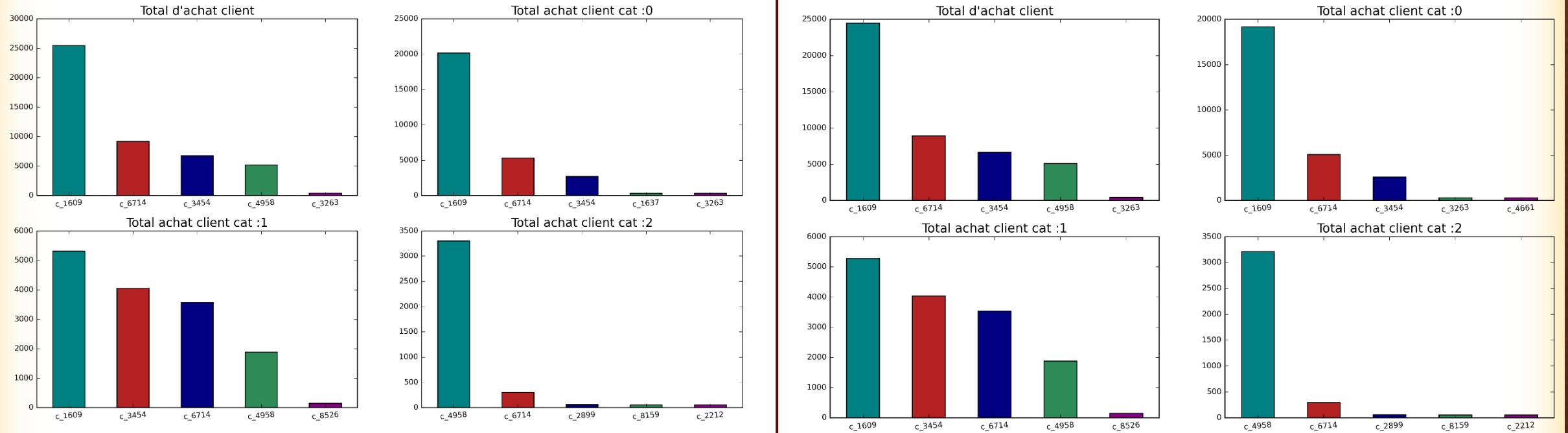
Différence avec le mois d'octobre

Différence avec ou sans mois octobre



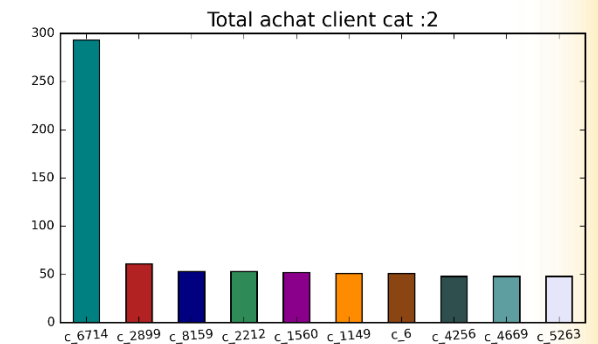
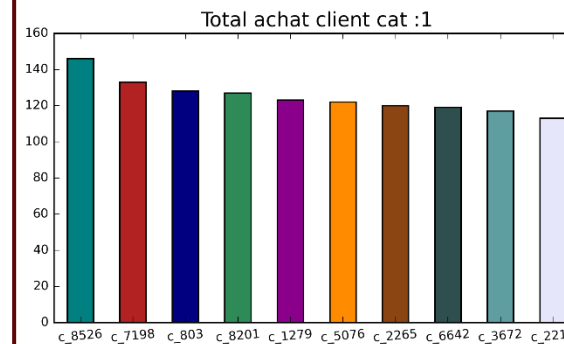
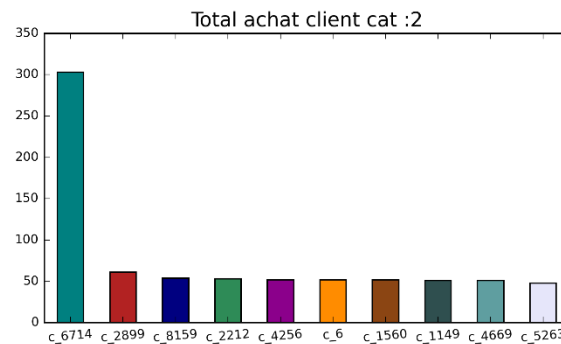
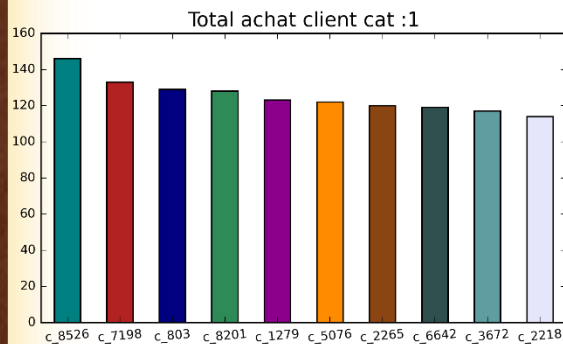
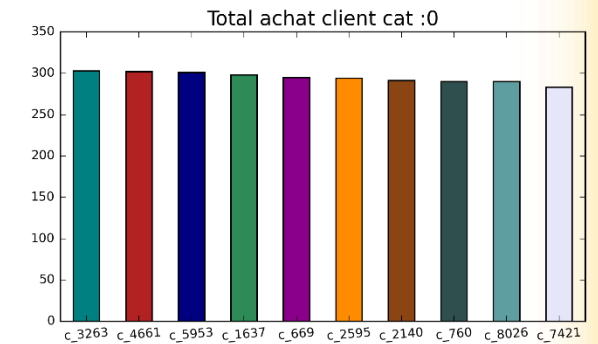
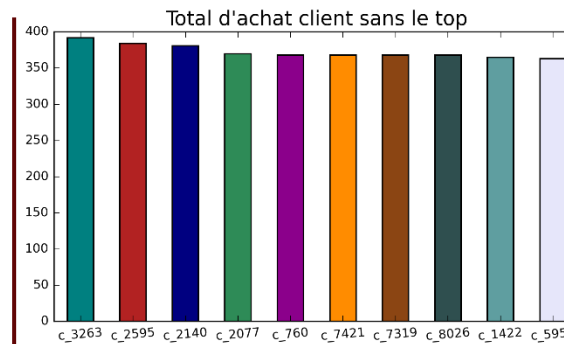
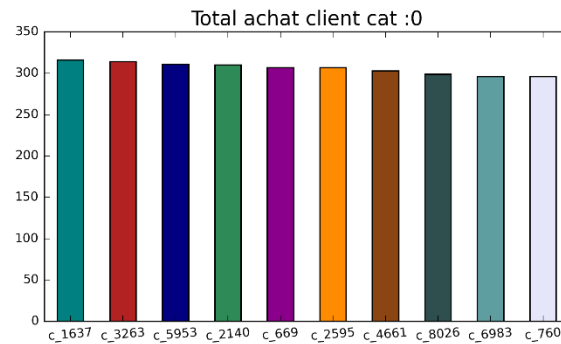
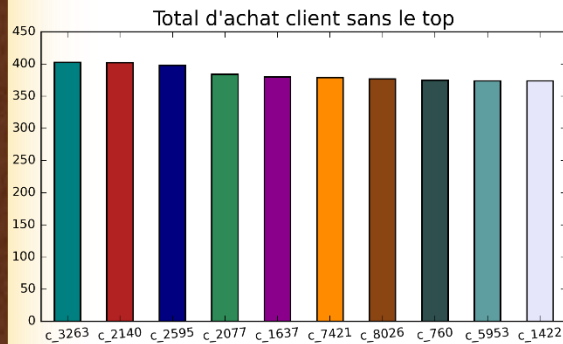
Comparaison du nombre d'achat catégorie avec et sans le mois d'octobre 2021

Différence avec ou sans mois octobre



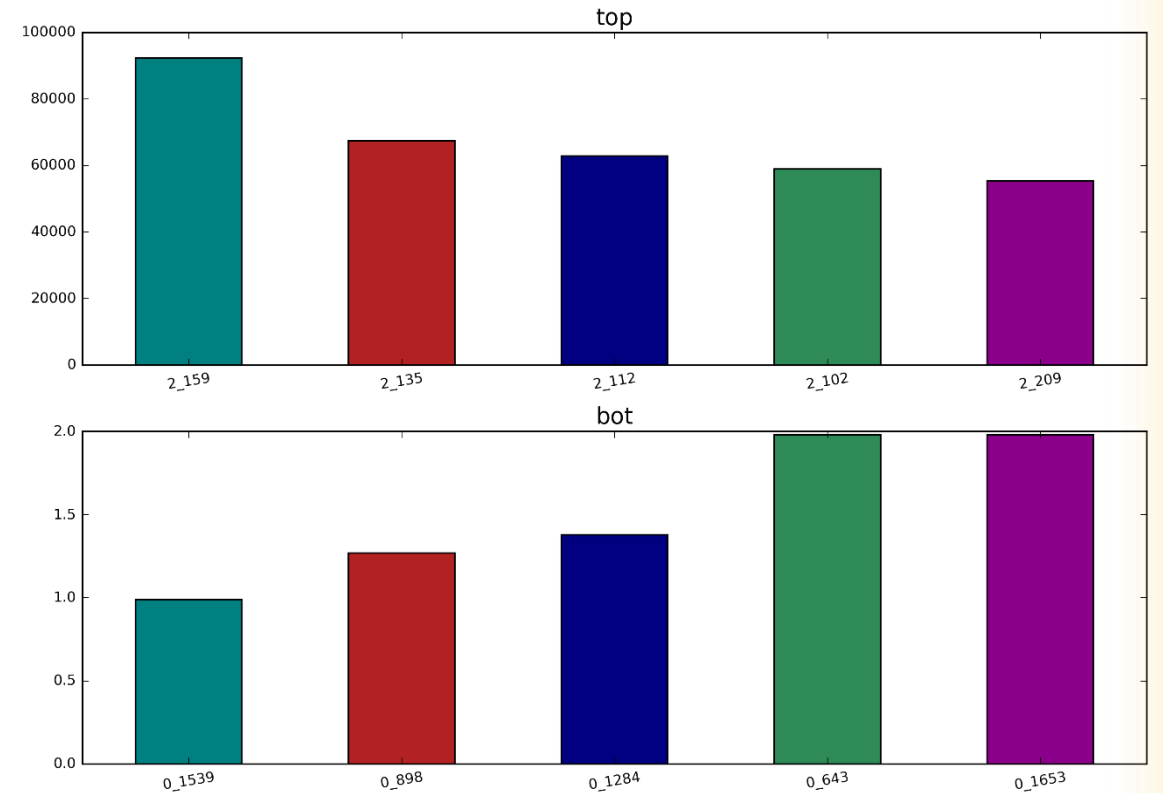
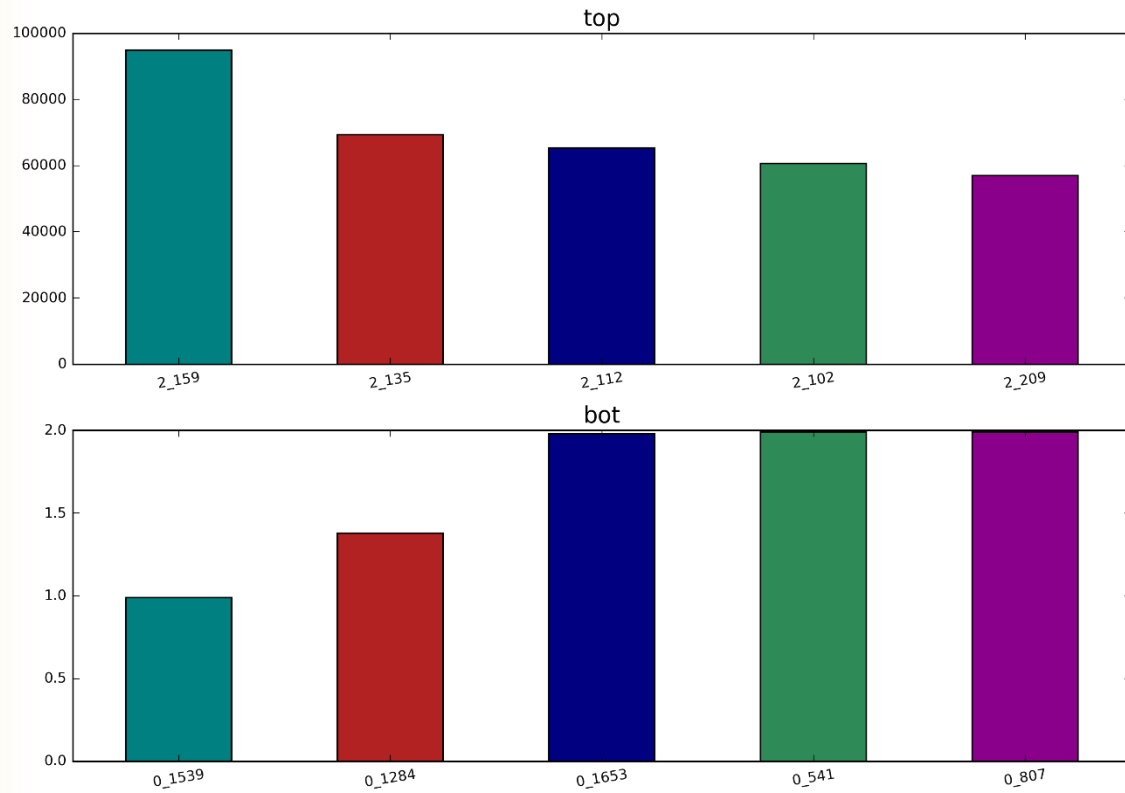
Comparaison du nombre d'achat client avec et sans le mois d'octobre 2021

Différence avec ou sans mois octobre



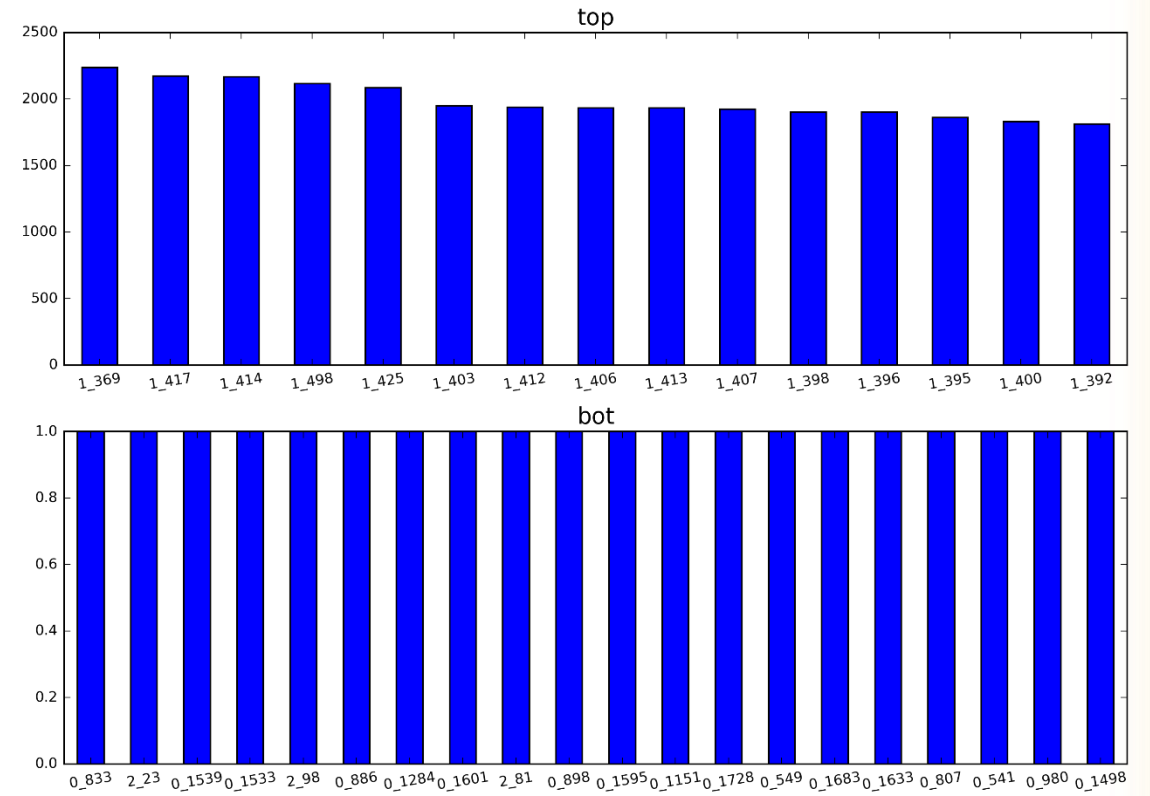
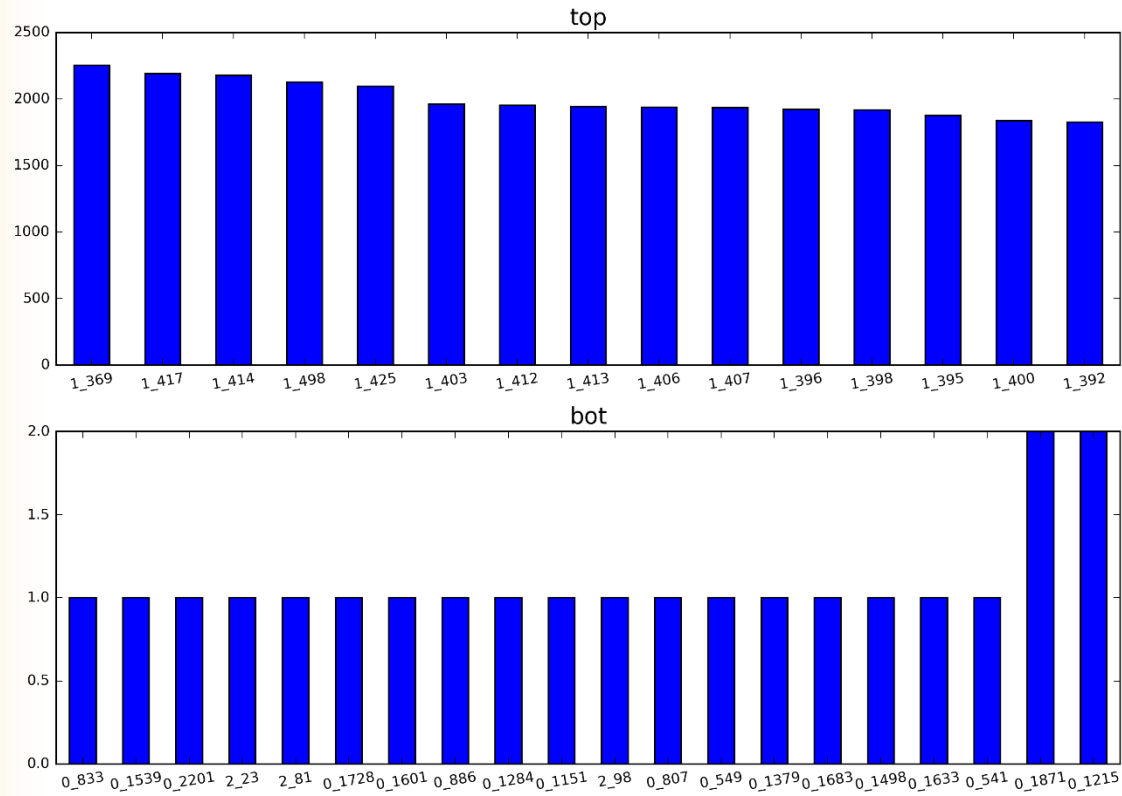
Comparaison du nombre d'achat client(sans les top clients) avec et sans le mois d'octobre 2021

Différence avec ou sans mois octobre



Comparaison du CA des produits avec et sans le mois d'octobre 2021

Différence avec ou sans mois octobre



Comparaison des produits en nombre de vente avec et sans le mois d'octobre 2021

Récapitulatif

1. **Différences CA total** avec et sans le mois d'octobre : **11,857,586€ / 11,536,665€ (320 921 €)**
2. pertes de **0,1%** dans les catégorie d'âge : **18, 26-35, 36-45, 56-70, 71+**
3. Changement des meilleurs acheteur dans total d'achat client par catégorie 0 : **c_3263 et c_4661 remplace c_1637 et c_3263**
(Et une baisse d'achat total.)
4. **Même constat** pour la **total d'achat sans les gros clients**
5. Sur les produits changement du bot 5 avec : **0_898, 0_1284, 0_643, 0_1653** qui s'ajoute an **bottom d'octobre**
6. Quelque **changement de place** dans le nombre de **ventes top**, et **ajout de produits en baisse pour le bot**