



Analyse des ventes

3 datasets

Clients

- Identifiant client (exemple : c_4410)
- Genre (m, f)
- Date de naissance

Livres

- Identifiant livre (exemples : 0_1421, 1_587, 2_226)
- Prix
- Catégorie (0, 1, 2)

Ventes

- Identifiant session (exemple : s_18746)
- Date session (exemple : 2021-04-10 18:37:28.723910)
- Identifiant livre
- Identifiant client

Préparation des données

1 - Recherche des
duplicatas et des
valeurs
manquantes

2 - Suppression
d'une valeur
aberrante (prix
négatif)

3 - Suppression
des sessions de
test

4 - Merge des 3
datasets

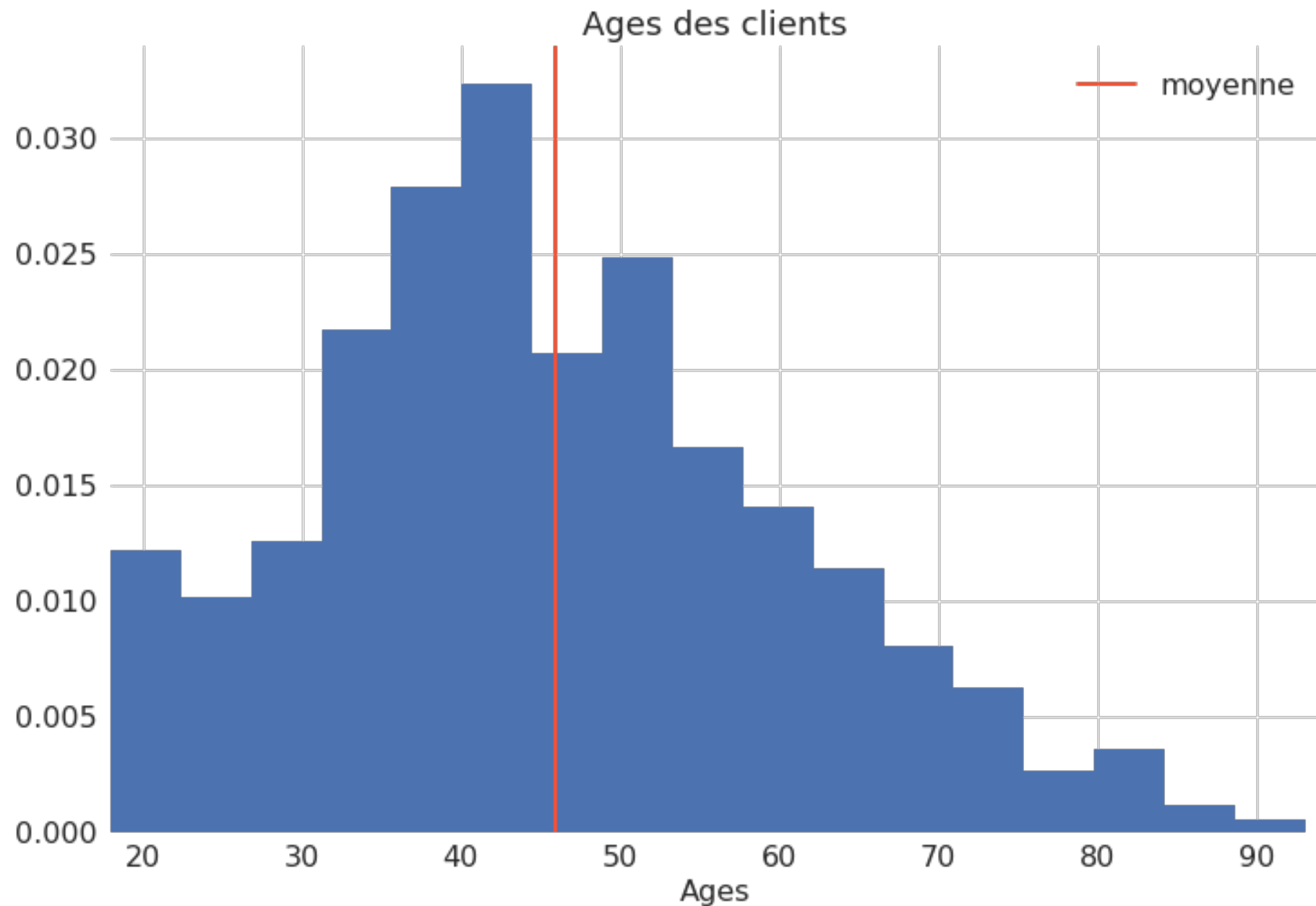
5 – Imputation prix
et catégorie
manquants du
livre n° 0_224

Dataset final

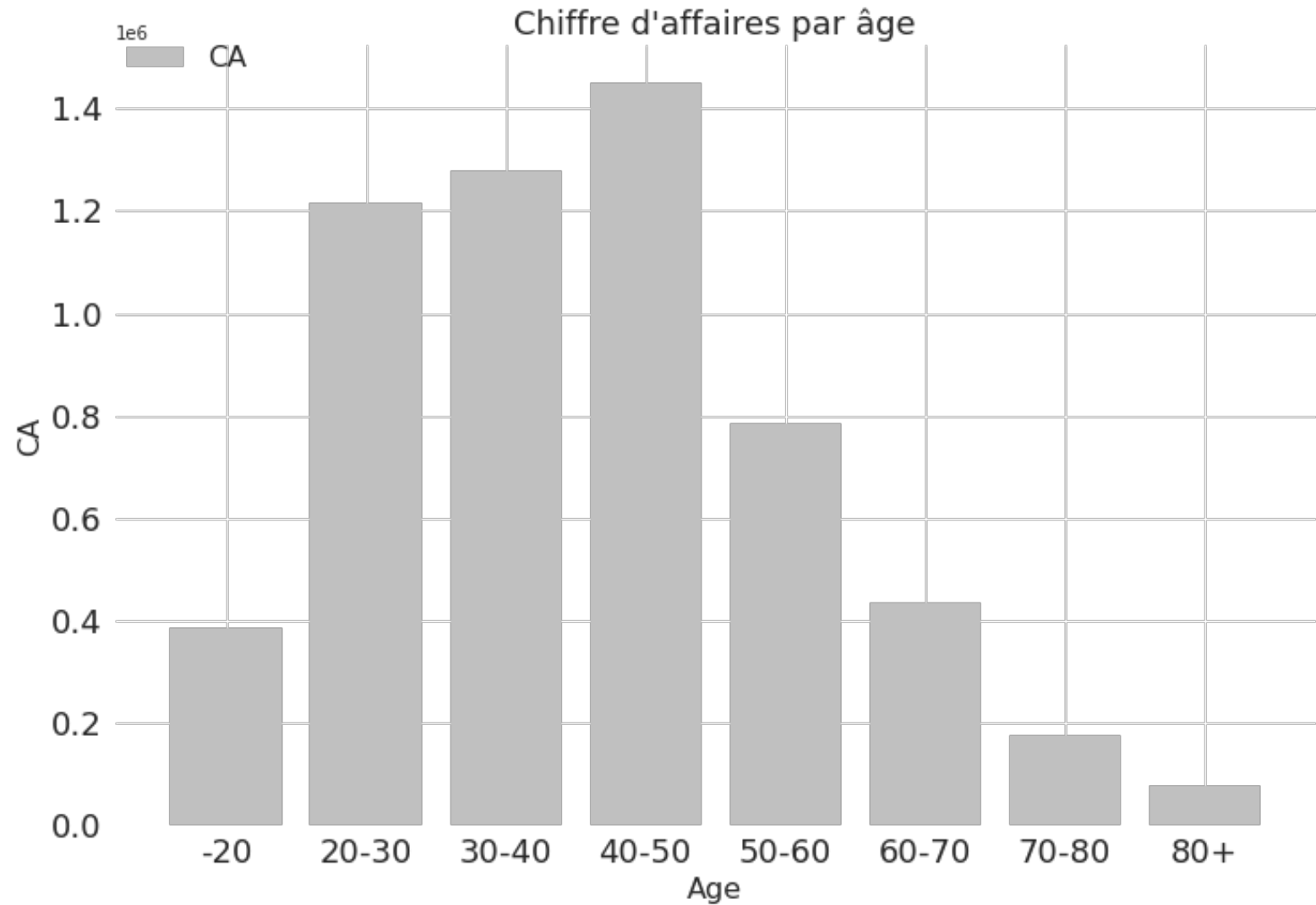
- Nombre de lignes : 336 816
 - Nombre de colonnes : 10
- Date de transaction
 - Identifiant livre
 - Prix (de 0,62 € à 300 €)
 - Catégorie (0, 1, 2)
 - Identifiant de session
 - Identifiant du client
 - Age du client (calculé à partir sa date de naissance)
 - Genre du client (m, f)
 - Date de naissance du client

Distribution des âges des clients

- mode : 42
- âge médian : 44
- âge moyen : 46
- écart type : 15
- skewness : 0.37
- kurtosis : - 0.17

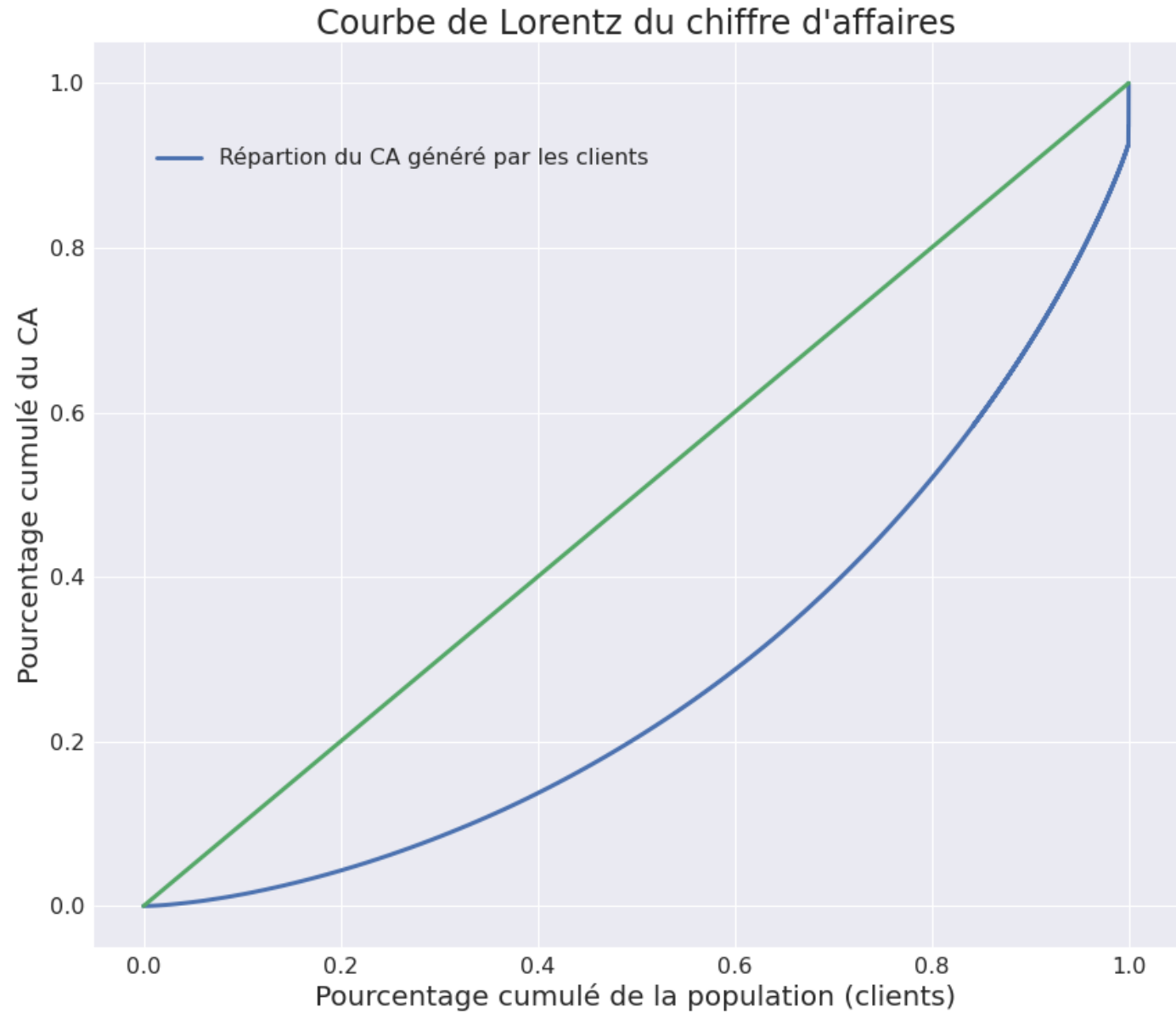


Chiffre
d'affaires
g n r  par
classe d' ge

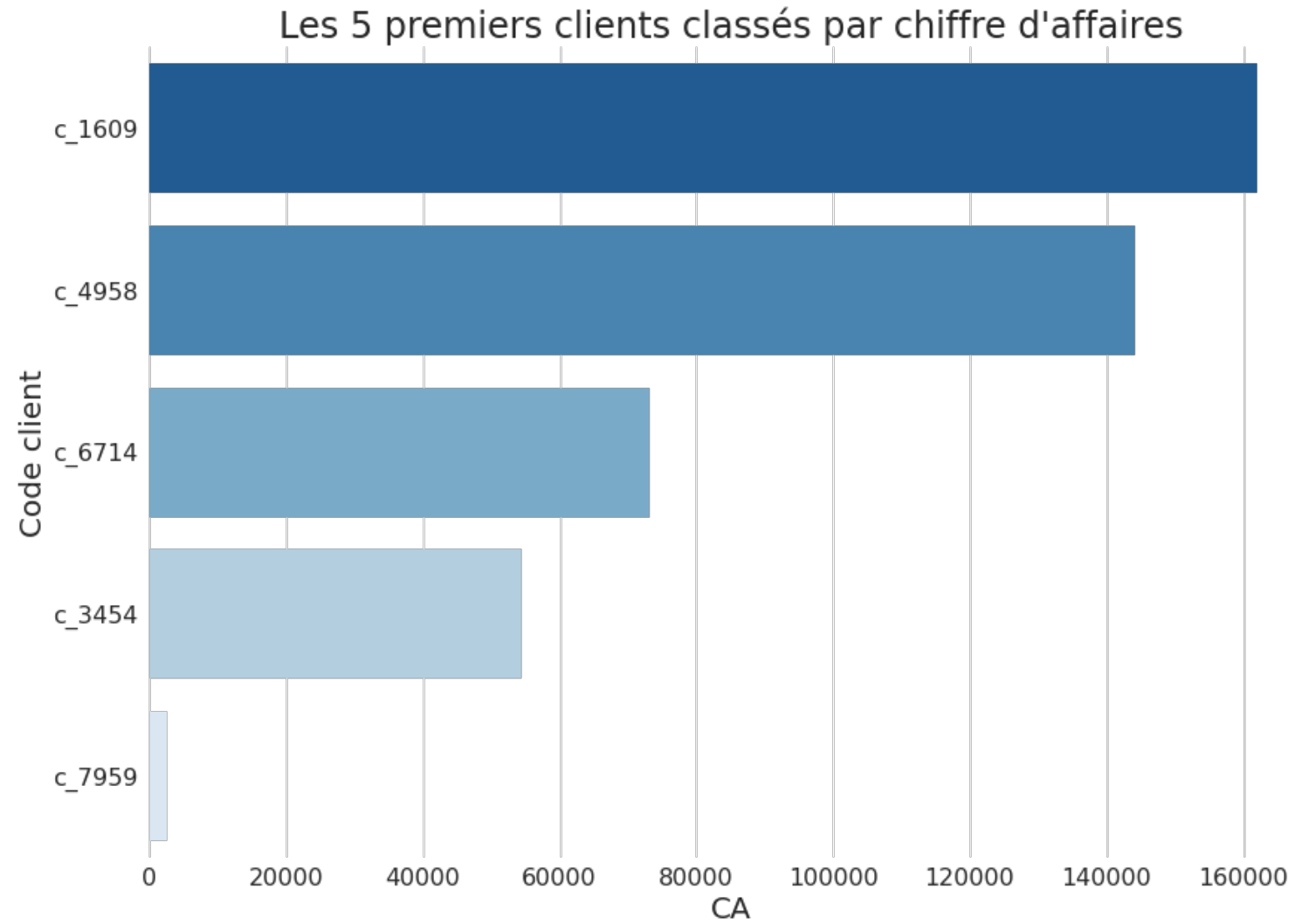


Répartition du chiffre d'affaires

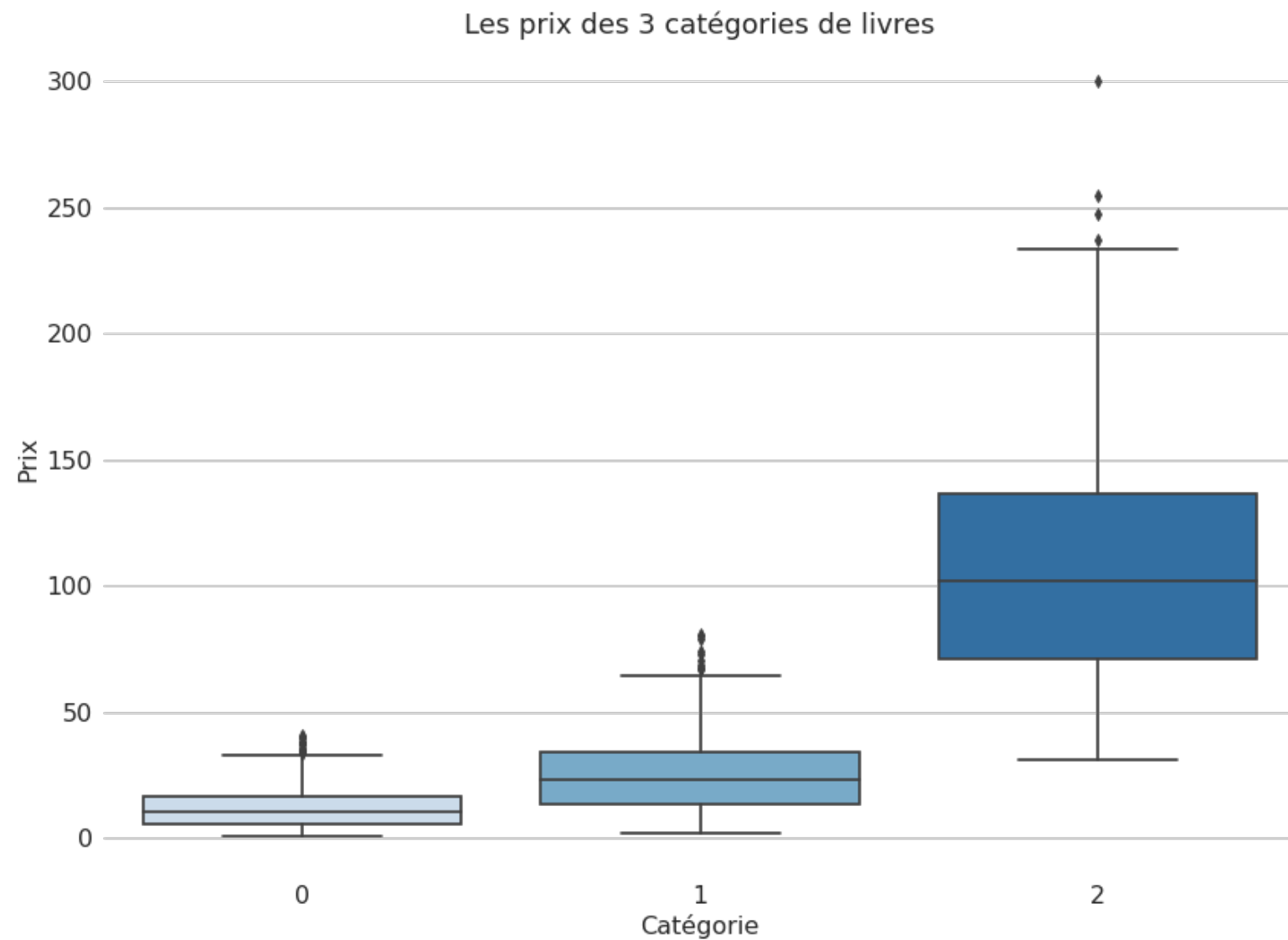
Indice de Gini :
0,40



4 clients sont
des
professionnels

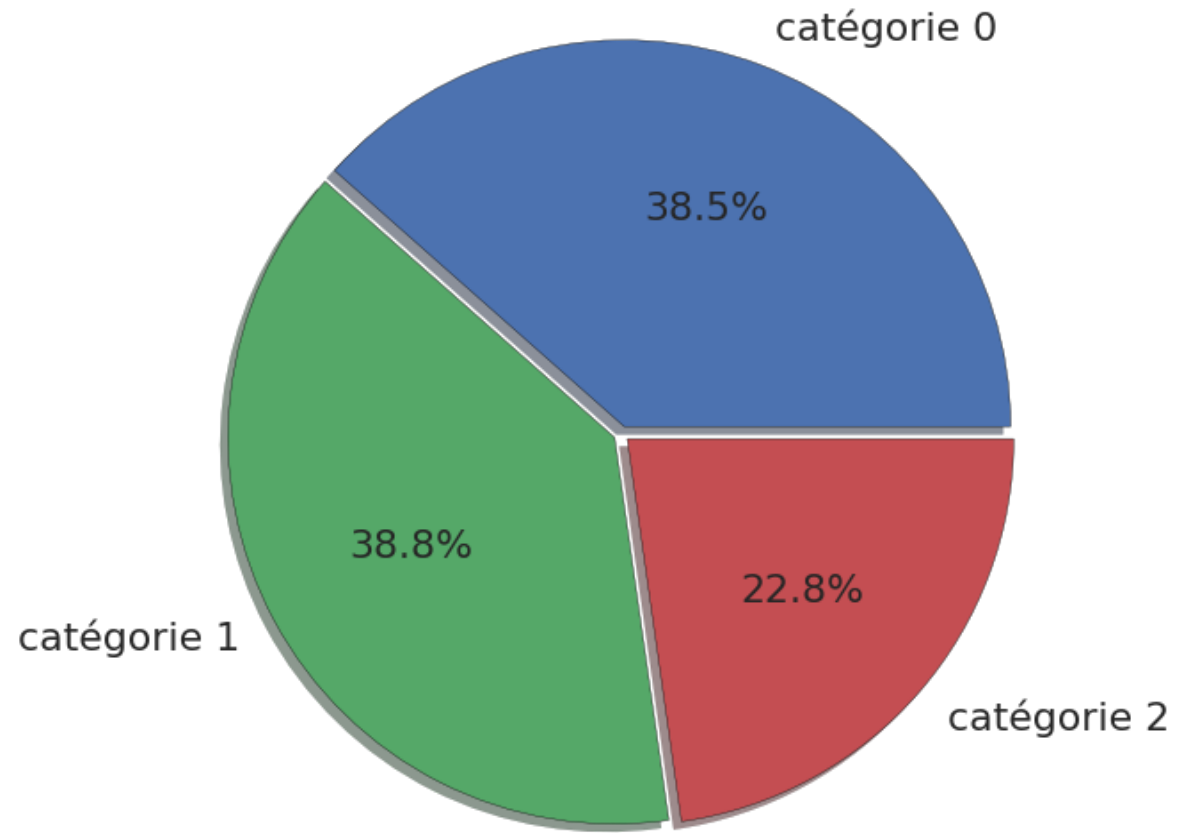


Les prix des 3 catégories de livres

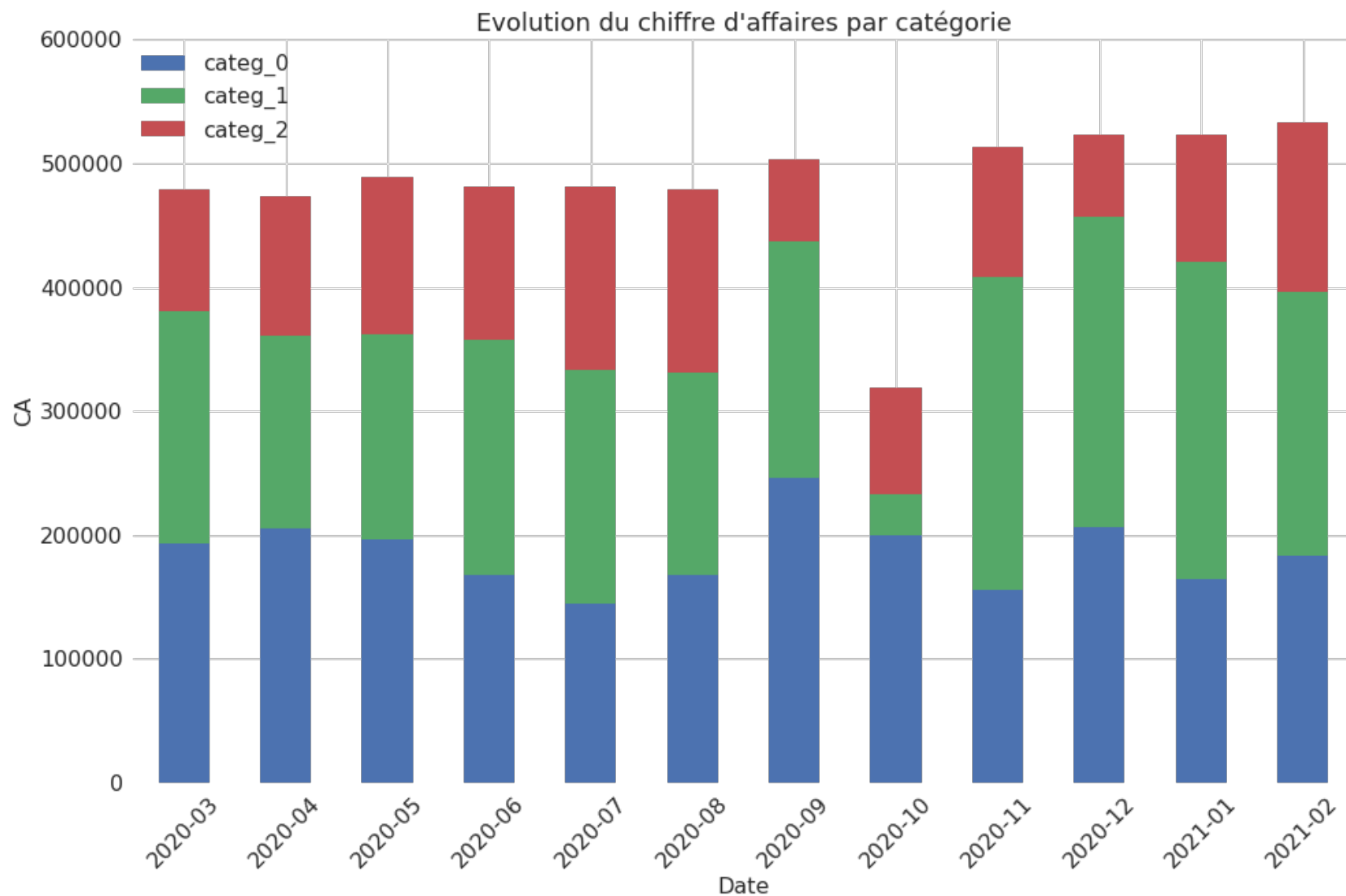


Proportion du
chiffre
d'affaires par
catégorie de
livre

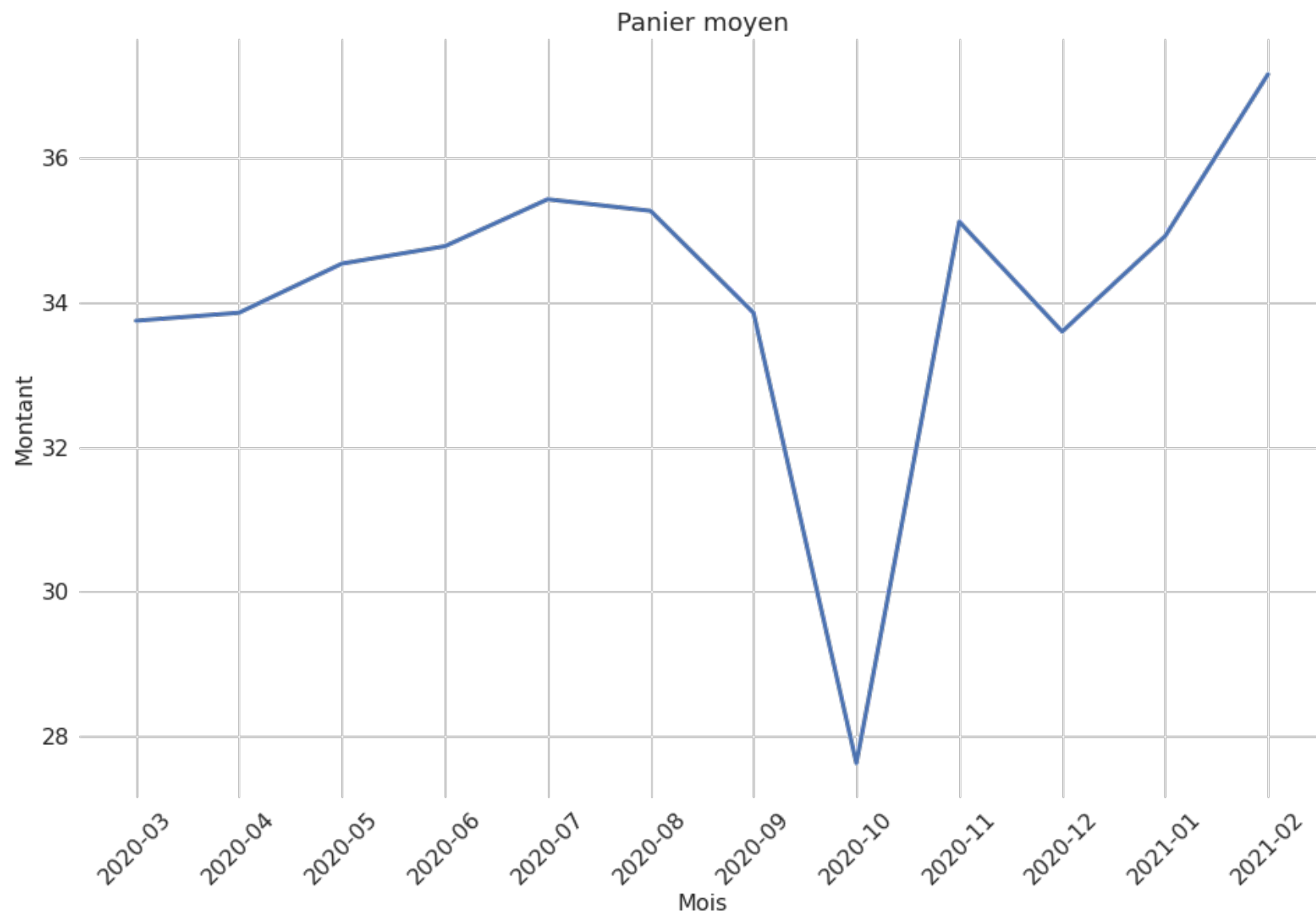
Chiffre d'affaires par catégories



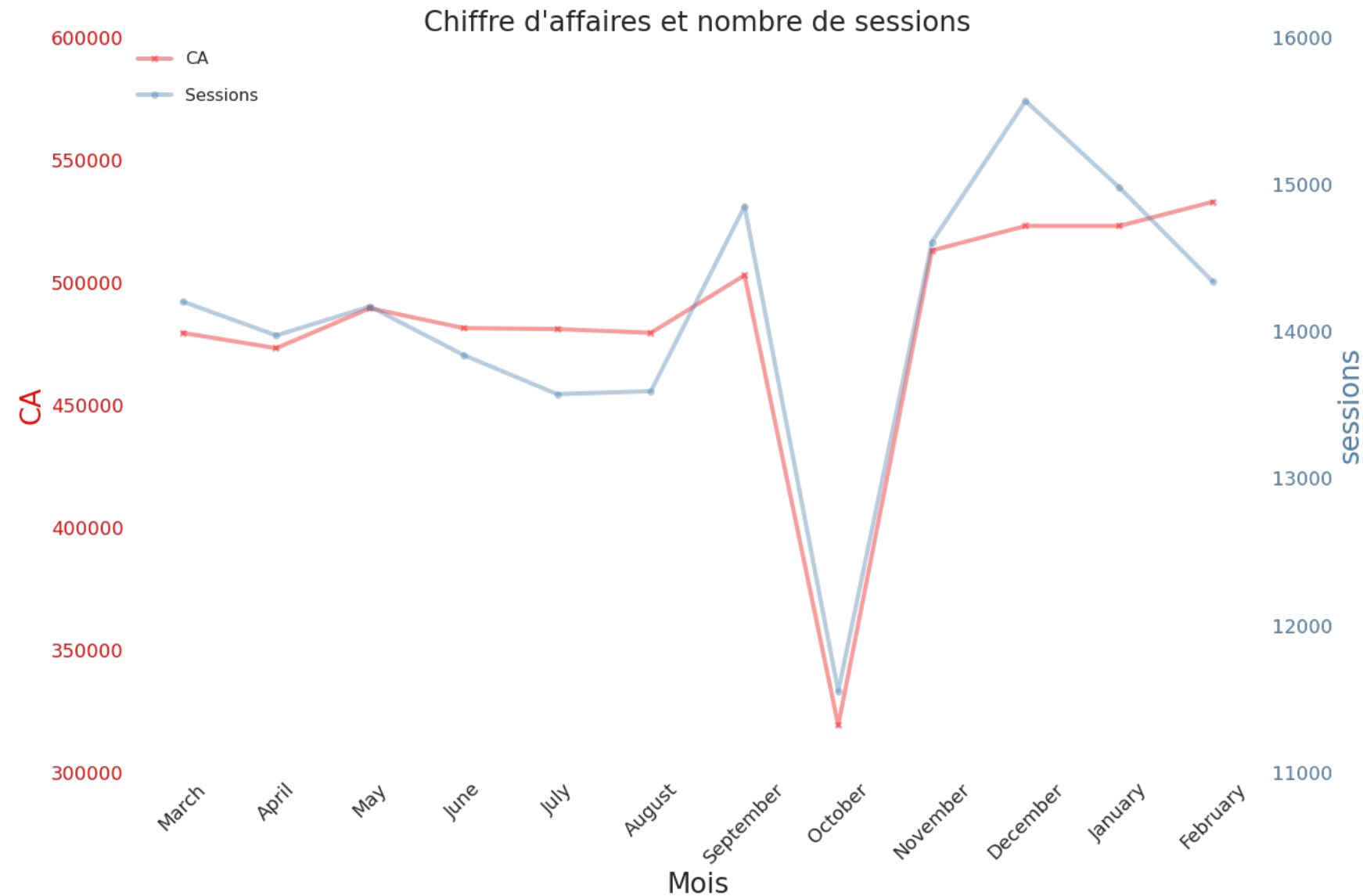
Evolution du
chiffre
d'affaires par
catégorie de
livre



Evolution du panier moyen



Evolution du
chiffre
d'affaires et
du nombre
de sessions



Chute des
visites en
octobre 2020

