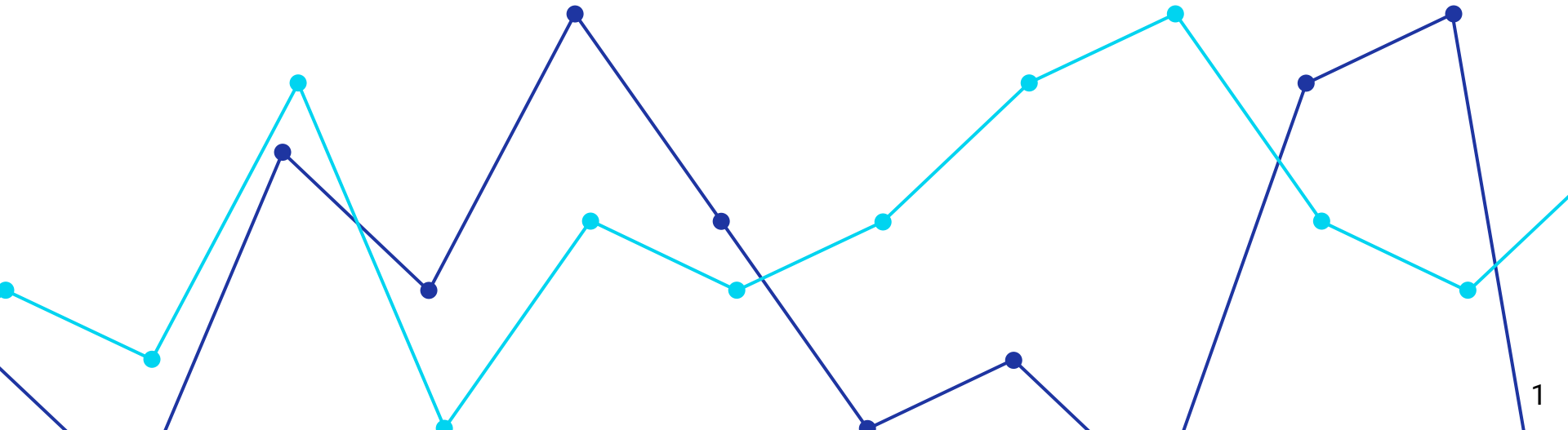


Data Analyst



Problématique

**Comment établir un état des lieux
via les chiffres d'affaires d'une
entreprise ?**

Index



- I. Librairie
 - A. Chiffres
 - B. Corrélation des variables
- II. Traitement des données
 - A. Clear data
- III. Résumé
- IV. Questions

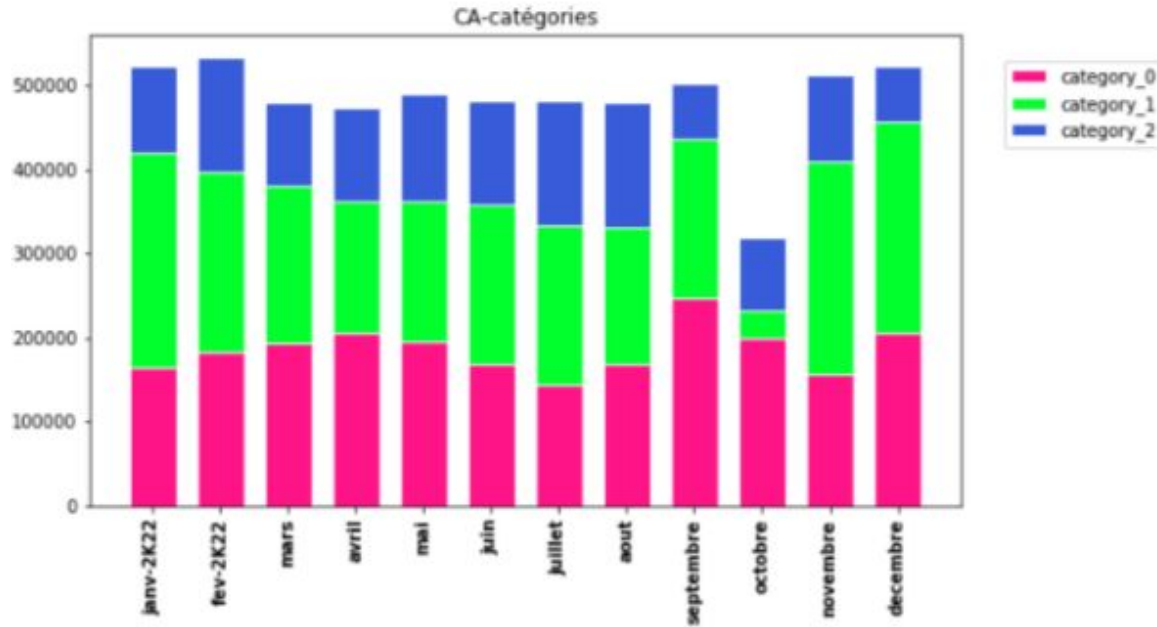
Chiffres

En 2021 : cette librairie réalise un chiffre d'affaire de de plus de **4 millions d'euros en CA.**

Nous pouvons décomposer les ventes en 3 grande catégorie.
La catégorie 0 représente **39% des transactions**

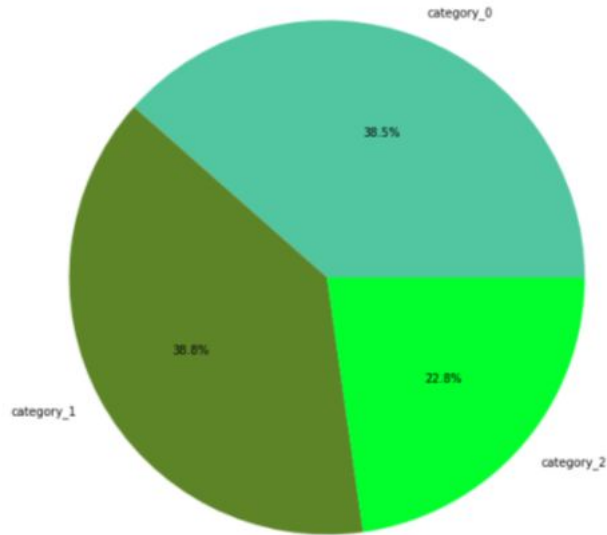
La société à générer plus de **250 000** transactions dans cette année.

Chiffre d'affaires / Cat



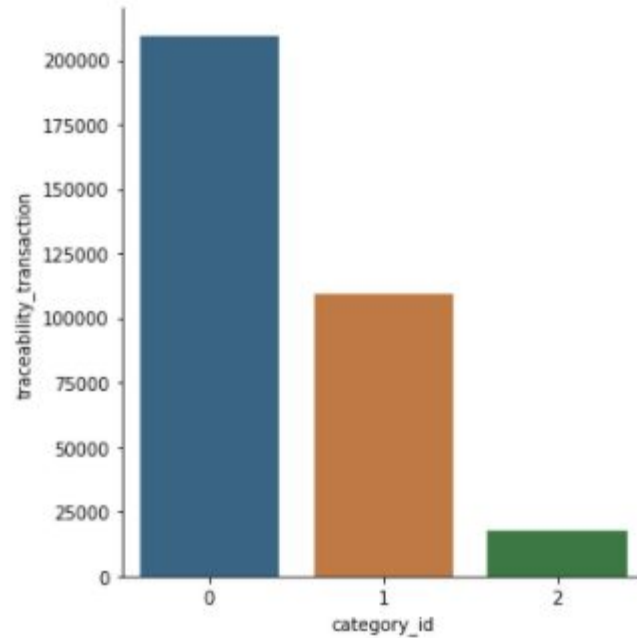
Ca / Catégories par Mois

Les catégories



Répartition des transactions par catégories

Les transactions

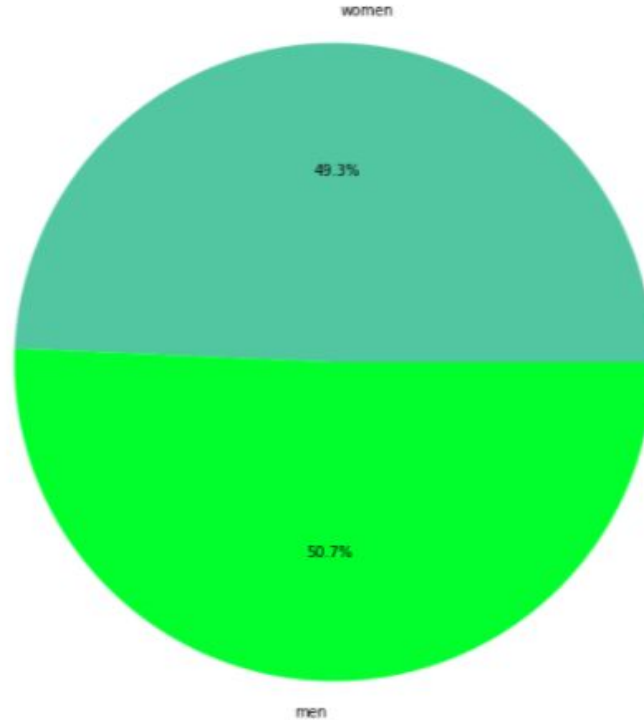


Transactions / Categorie

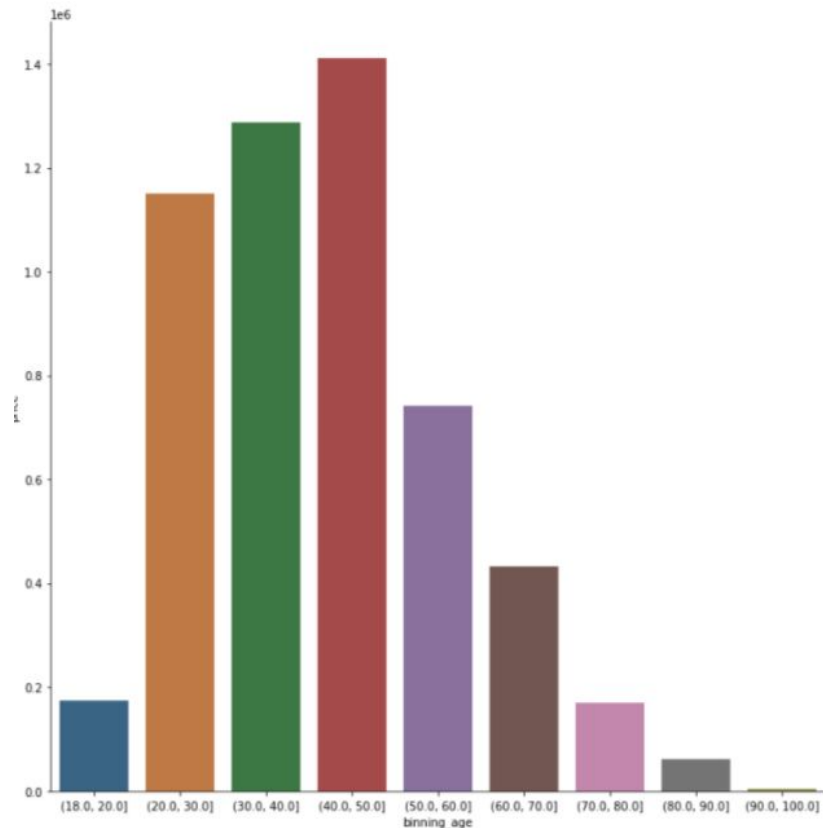
Corrélation des variables

Le but d'avoir des grosses données l'optimisation du CA dans ce cas précis en extrayant différentes variable on peut voir le comportement des consommateurs

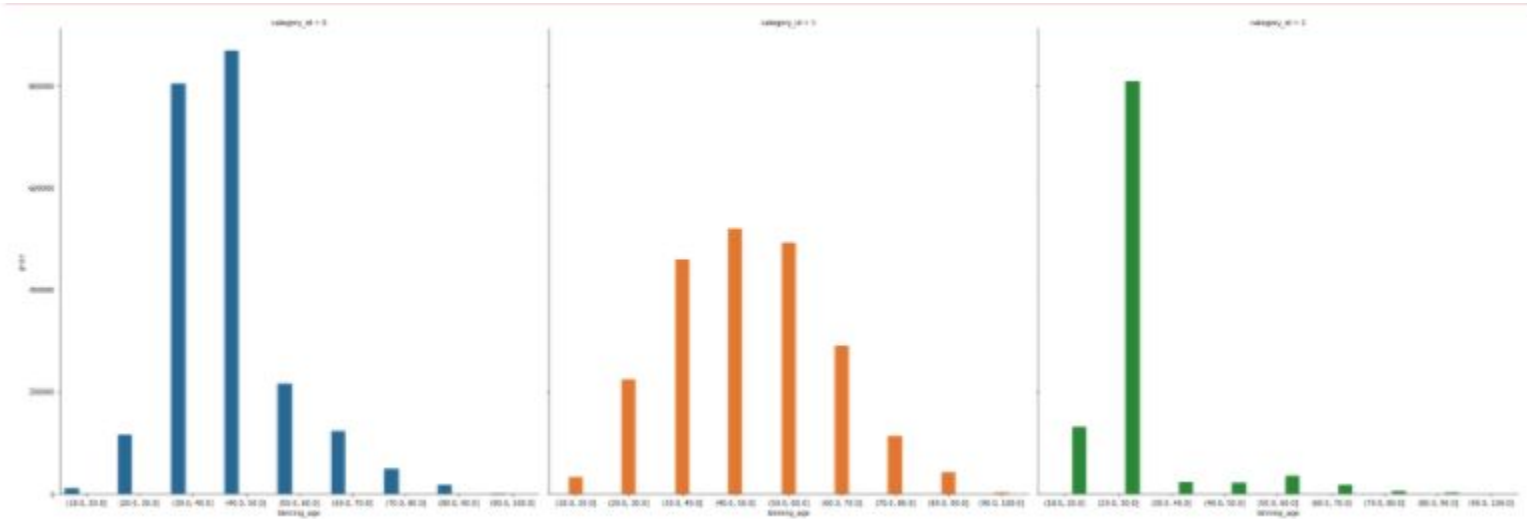
Impact sur le genre ?



Les dépenses par tranche d'âge



Analyse bi-varié CA = Catégorie / tranche d'âge



Traitement des données

Pour obtenir les précédents résultats, il nous faut nettoyer et fusionner les données en amont.

Ceci est un processus assez long et peut être complexe

Clear data

Le nétoyage des données est décomposer en 3 partie.

1. Soustraire les valeurs négatives
2. Sélectionner les valeurs null
3. Une fois traité fusionner les données et comparer par les valeurs de bases

Exemple

```
customersTransaction = customers.merge(transactions, on="client_id")

productCustomerTransac = customersTransaction.merge(products, on="id_product")

outer_join_df = pd.merge(customersTransaction, products, on='id_product', how='outer')
originePrice = productCustomerTransac['price'].sum()

outer_join_df['price'] = outer_join_df['price'].fillna((outer_join_df['price'].mean()))

df_na = outer_join_df.loc[outer_join_df.isnull().any(axis=1)]
```

Conclusion



1. Une augmentation du chiffre d'affaire à part au mois d'octobre grosse baisse
2. Le sexe n'interfère pas avec le CA
3. Les 30-40 sont ce qui dépense le plus
4. Il y a une niche commerciale pour les plus de 40 ans car ce groupe achète les livres les plus chers

Questions ?