



Analyses des ventes de notre site

- **Réaliser un inventaire des données dont nous disposons**
 - **Nettoyer les données**
 - **Analyser les indicateurs de vente**
 - **Analyser les comportements clients**
 - **Quelques recommandations**



3 datasets

- **Customers**
- **Products**
- **Transactions**

Fichier customers

Le tableau comporte 8621 observation(s) ou article(s)
Le tableau comporte 3 colonne(s).

Nature des données dans chacune des colonnes :

```
client_id  object
sex        object
birth      int64
dtype: object
```

Nombre de valeurs présentes dans chacune des colonnes :

```
client_id  8621
sex        8621
birth      8621
dtype: int64
```

Nombre de valeurs uniques dans chacune des colonnes :

```
client_id  8621
sex        2
birth      76
dtype: int64
```

Nombre de doublons présents dans chacune des colonnes :

```
client_id  0
sex        8619
birth      8545
dtype: int64
```

Nombre de valeurs nulles présentes dans chacune des colonnes :

```
client_id  0
sex        0
birth      0
dtype: int64
```

Valeurs uniques dans 'sex' : ['f' 'm']

birth	
count	8621.000000
mean	1978.275606
std	16.917958
min	1929.000000
25%	1966.000000
50%	1979.000000
75%	1992.000000
max	2004.000000

Le fichier clients contient **8 621 client_id** et renseigne le **sexe** et la **date de naissance** de chaque client.

Fichier products

Le tableau comporte 3286 observation(s) ou article(s)
Le tableau comporte 3 colonne(s).

Nature des données dans chacune des colonnes :

```
id_prod    object
price      float64
categ      int64
dtype: object
```

Nombre de valeurs présentes dans chacune des colonnes :

```
id_prod    3286
price      3286
categ      3286
dtype: int64
```

Nombre de valeurs uniques dans chacune des colonnes :

```
id_prod    3286
price      1454
categ      3
dtype: int64
```

Nombre de doublons présents dans chacune des colonnes :

```
id_prod    0
price      1832
categ      3283
dtype: int64
```

Nombre de valeurs nulles présentes dans chacune des colonnes :

```
id_prod    0
price      0
categ      0
dtype: int64
```

Valeurs uniques dans 'categ' : [0 1 2]

	price	categ
count	3286.000000	3286.000000
mean	21.863597	0.370359
std	29.849786	0.615446
min	0.620000	0.000000
25%	6.990000	0.000000
50%	13.075000	0.000000
75%	22.990000	1.000000
max	300.000000	2.000000

Le fichier produits contient **3 286 id_product** et renseigne le **prix** et la **catégorie** de chaque produit.

Fichier transactions

Le tableau comporte 1048575 observation(s) ou article(s)
Le tableau comporte 4 colonne(s).

Nature des données dans chacune des colonnes :

```
id_prod      object
date         object
session_id   object
client_id    object
dtype: object
```

Nombre de valeurs présentes dans chacune des colonnes :

```
id_prod      687534
date         687534
session_id   687534
client_id    687534
dtype: int64
```

Nombre de valeurs uniques dans chacune des colonnes :

```
id_prod      3265
date         687419
session_id   345505
client_id    8600
dtype: int64
```

Nombre de doublons présents dans chacune des colonnes :

```
id_prod      684269
date         115
session_id   342029
client_id    678934
dtype: int64
```

Nombre de valeurs nulles présentes dans chacune des colonnes :

```
id_prod      361041
date         361041
session_id   361041
client_id    361041
dtype: int64
```

```
# Supprimer toutes les lignes entièrement vides
transactions.dropna(how="all", inplace=True)
```

Nombre de valeurs nulles présentes dans chacune des colonnes :

```
id_prod      0
date         0
session_id   0
client_id    0
dtype: int64
```

	année	mois	jour
count	687534.000000	687534.000000	687534.000000
mean	2021.662289	6.547208	15.715258
std	0.618184	3.480693	8.803949
min	2021.000000	1.000000	1.000000
25%	2021.000000	4.000000	8.000000
50%	2022.000000	7.000000	16.000000
75%	2022.000000	10.000000	23.000000
max	2023.000000	12.000000	31.000000

```
id_prod      3265
session_id   345505
client_id    8600
année        3
mois         12
jour         31
heure        1440
dtype: int64
```

Le fichier transactions contient **345 505 session_id**, **3 265 id_prod** et **8 600 client_id**. Notons qu'il y a 21 clients et 21 produits en moins que dans les tables précédentes. Il renseigne la **date** (année, mois, jour) et l'**heure** de chaque commande.

Jonction des fichiers

- Jointure interne
- 21 clients et 21 produits exclus
- Database complète et fiable
- Possibilité d'exporter les données

Chiffre d'affaires

- Le CA total est de 12027663.10 €
- Nombre de commandes : 345505
- Nombre de produits vendus : 687534
- Nombre moyen de produits vendus par commande : 1.99
- Panier moyen : 34.81 €

Première année commerciale (3/2021 - 2/2022):

- CA total : 6005671.47 €
- Nombre de commandes : 172403
- Nombre de produits vendus : 345608
- Nombre moyen de produits vendus par commande : 2.00
- Panier moyen : 34.84 €

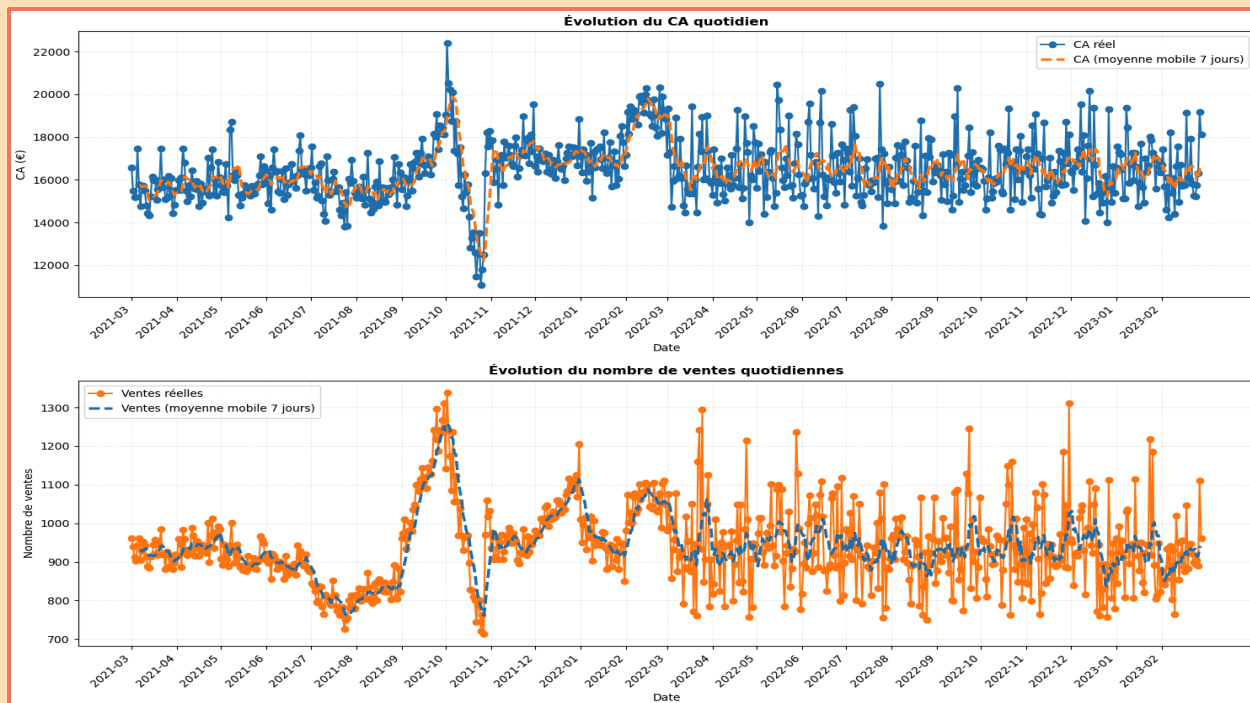
Seconde année commerciale (3/2022 - 2/2023):

- CA total : 6021991.63 €
- Nombre de commandes : 173102
- Nombre de produits vendus : 341926
- Nombre moyen de produits vendus par commande : 1.98
- Panier moyen : 34.79 €

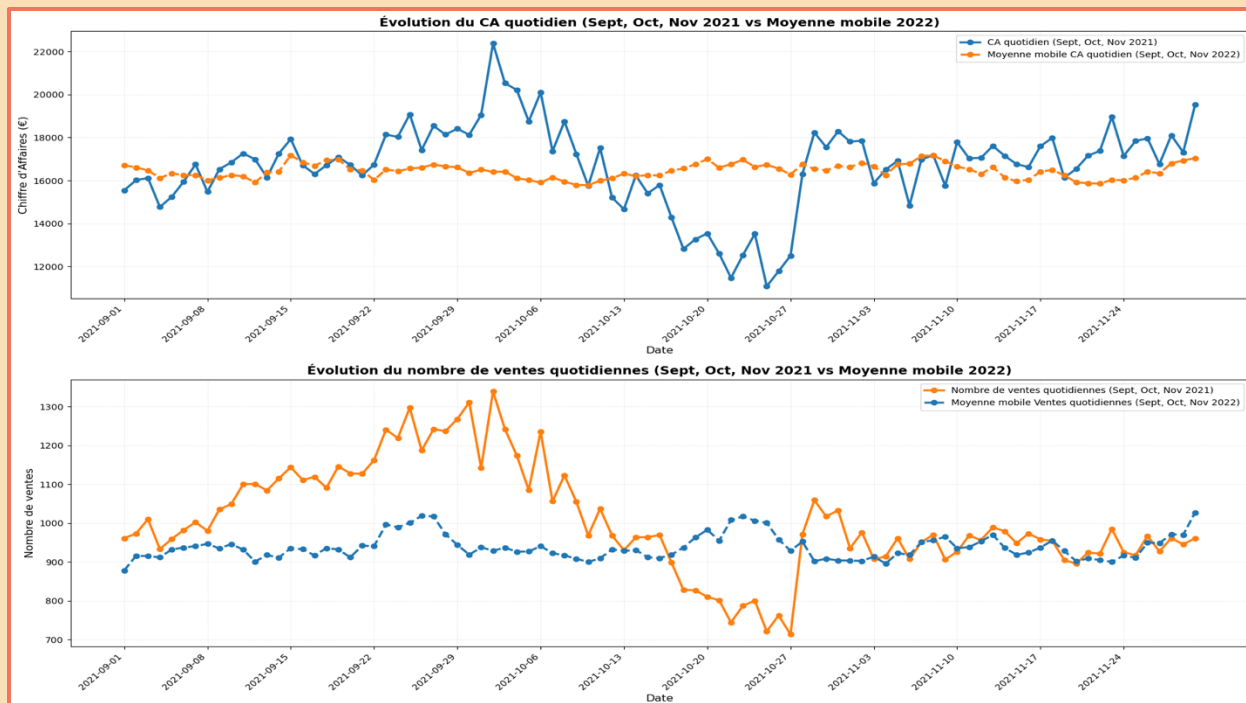
Évolutions entre les deux années :

- Évolution du CA total : 0.27%
- Évolution du nombre de commandes : 0.41%
- Évolution du nombre de produits vendus : -1.07%
- Évolution du nombre moyen de produits vendus par commande : -1.46%
- Évolution du panier moyen : -0.13%

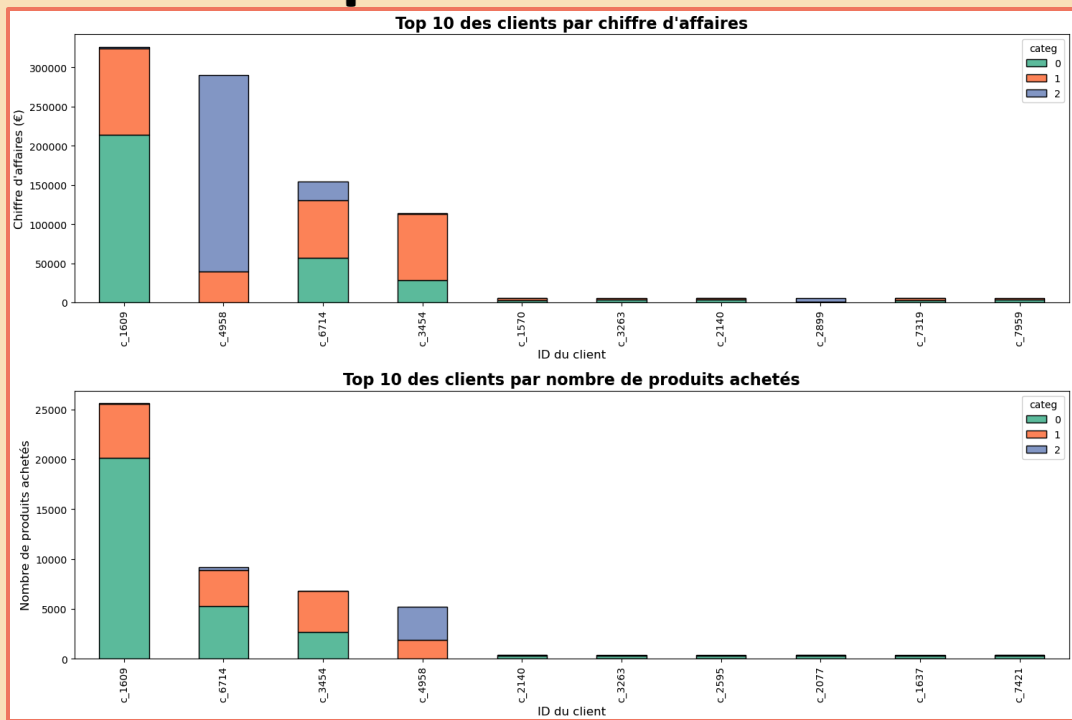
Evolutions quotidiennes



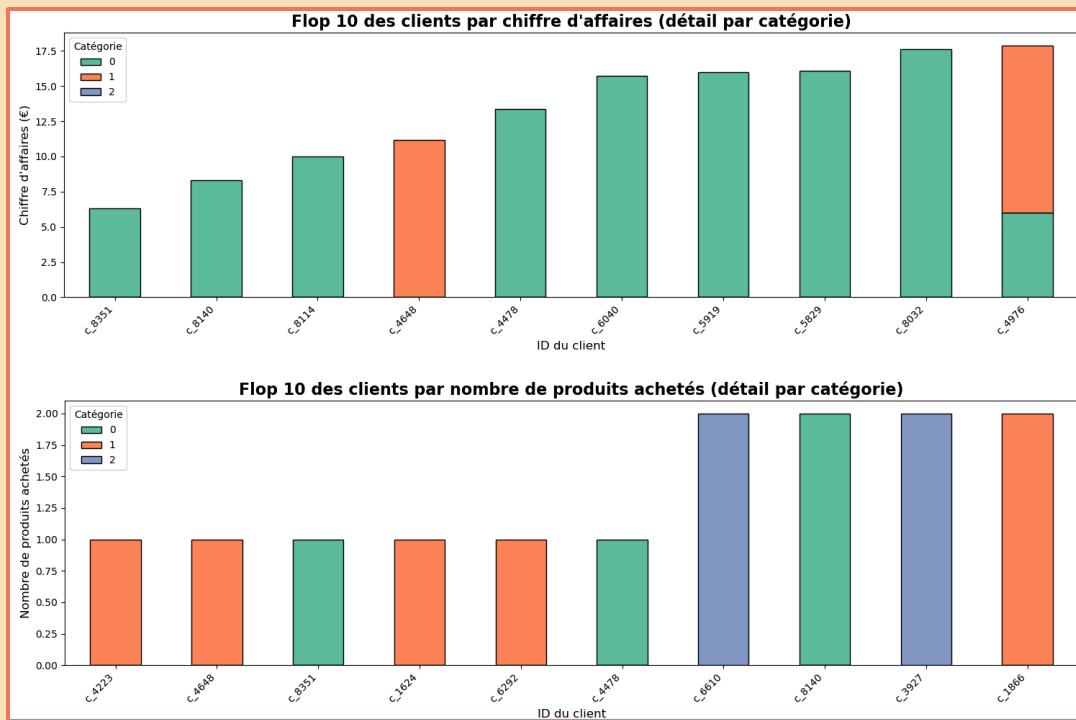
Zoom sur octobre 2021



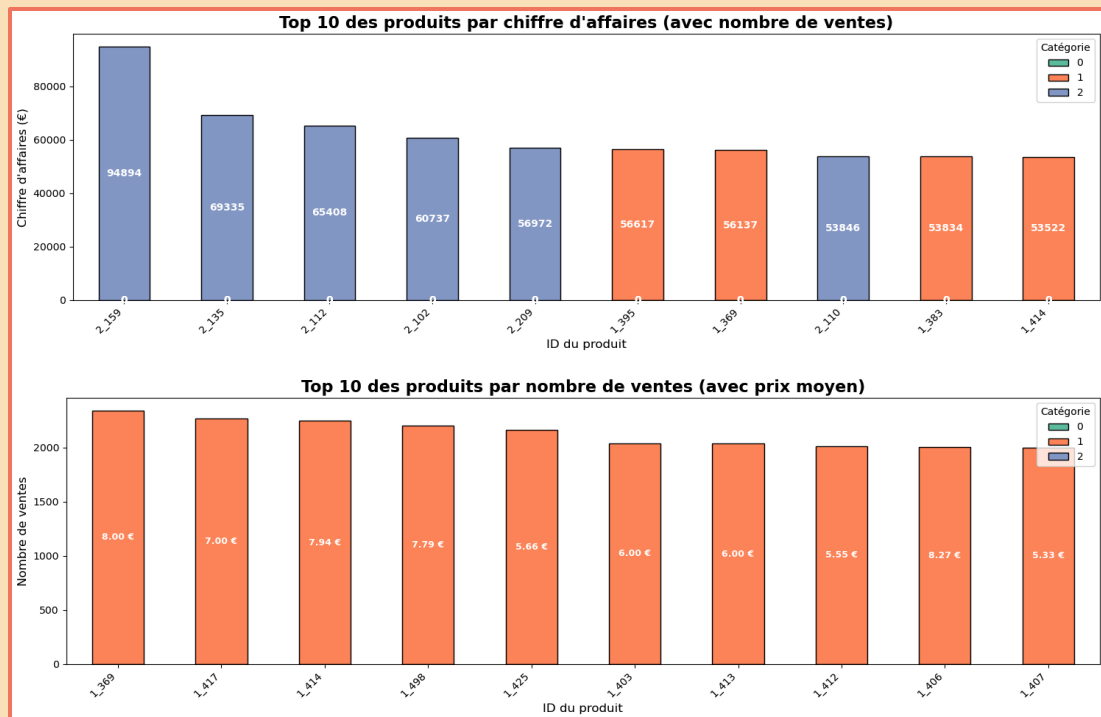
Top 10 des clients



Flop 10 des clients



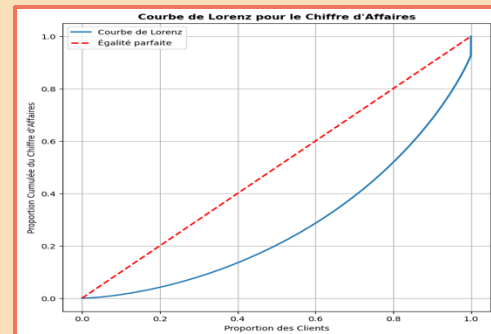
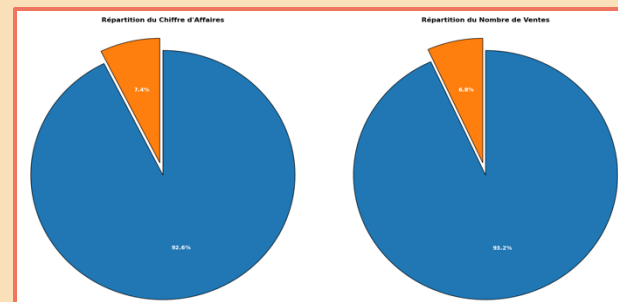
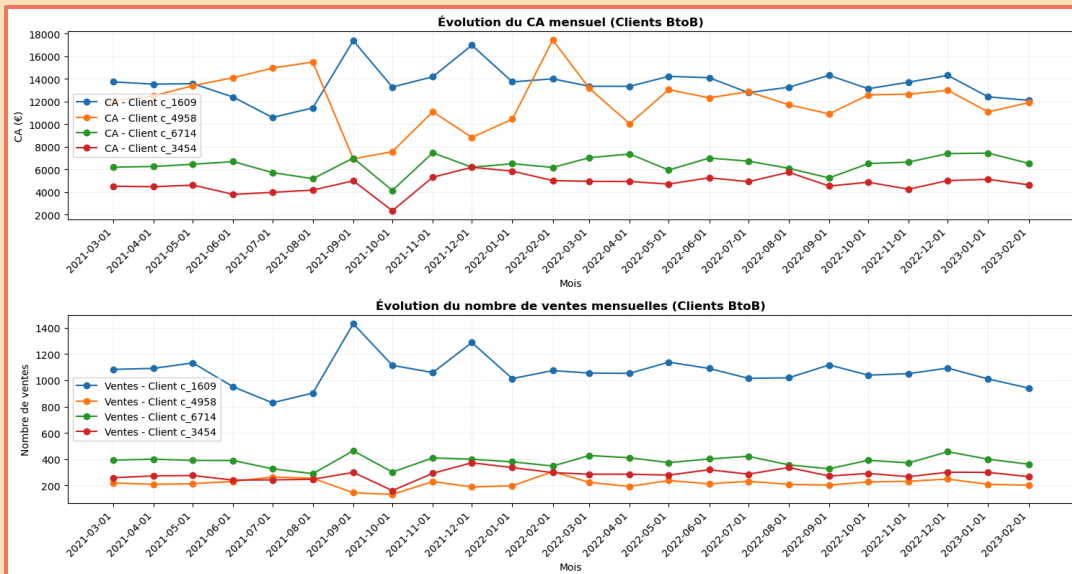
Top 10 des produits



Indicateurs de vente



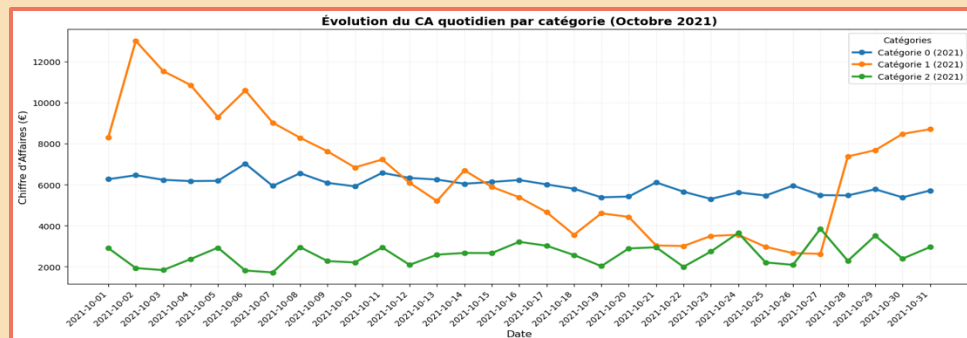
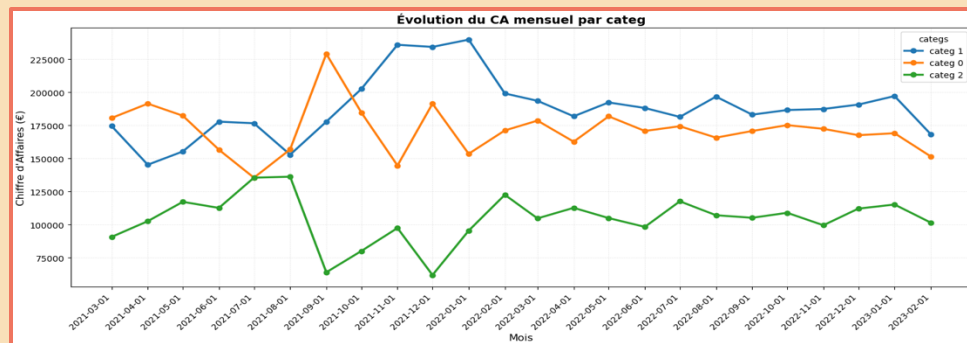
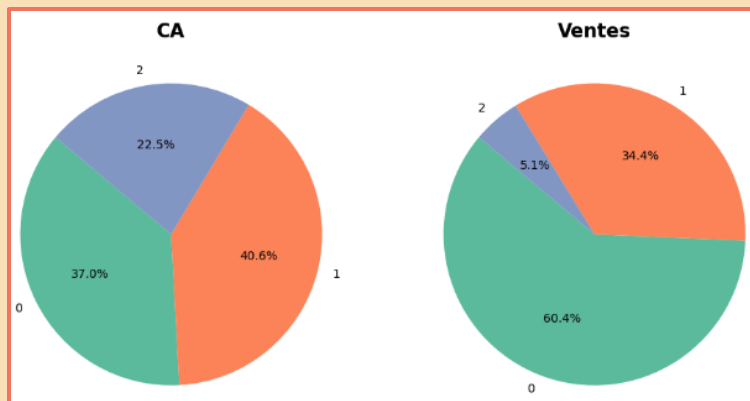
Clients BtoB



Indicateurs de vente



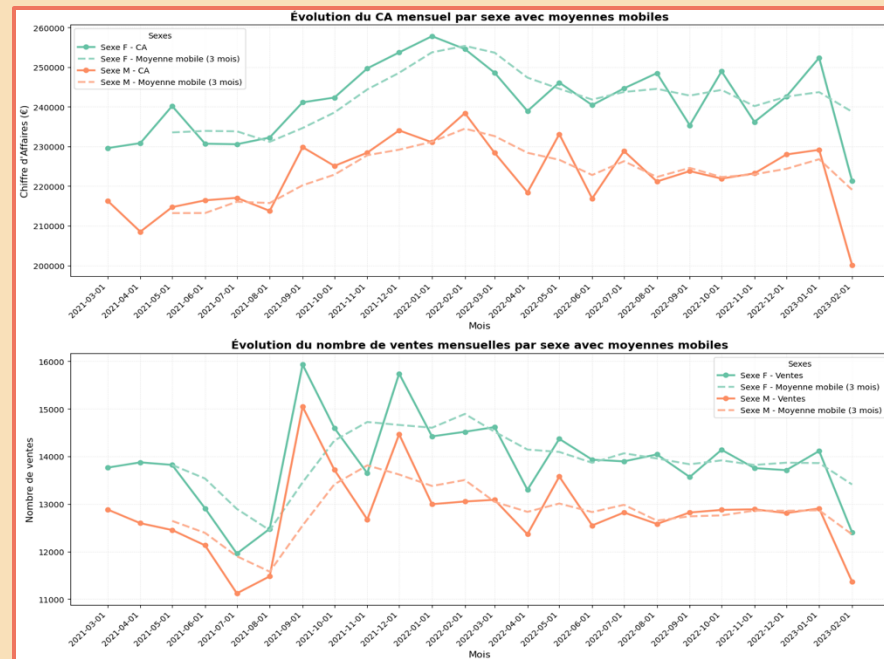
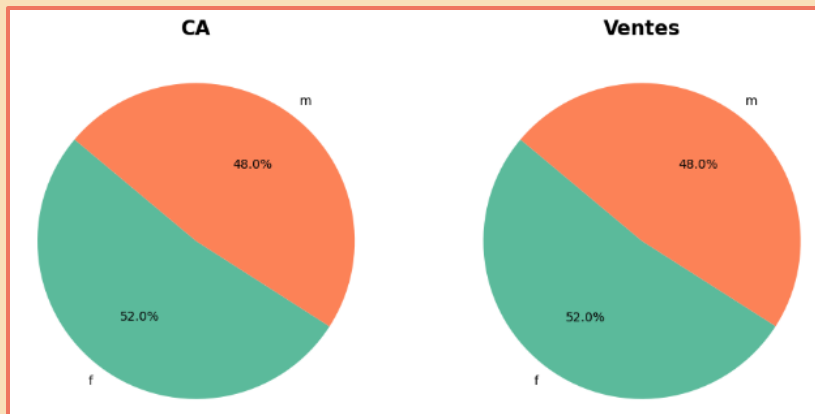
Répartition par catégorie



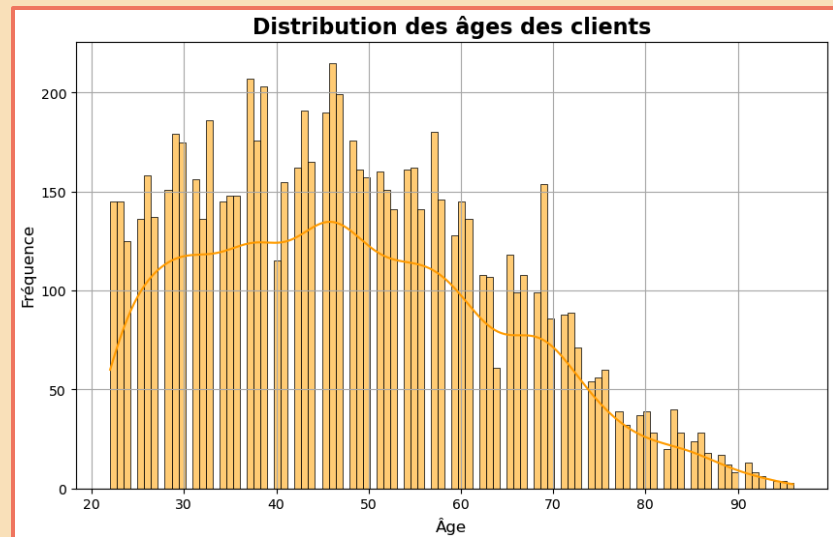
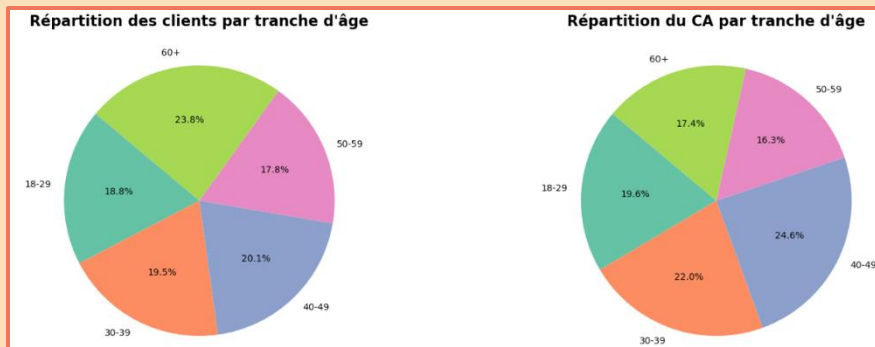
Indicateurs de vente



Répartition par sexe

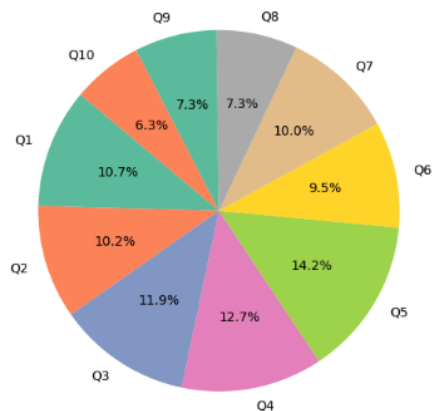


Répartition par âge 1/2

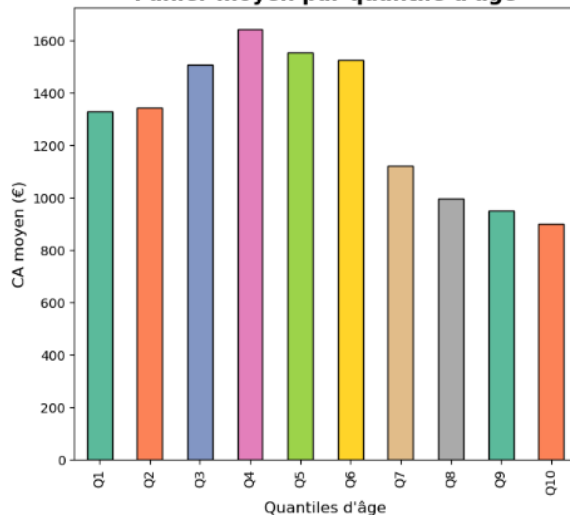


Répartition par âge 2/2

Répartition du CA par quantile d'âge



Panier moyen par quantile d'âge



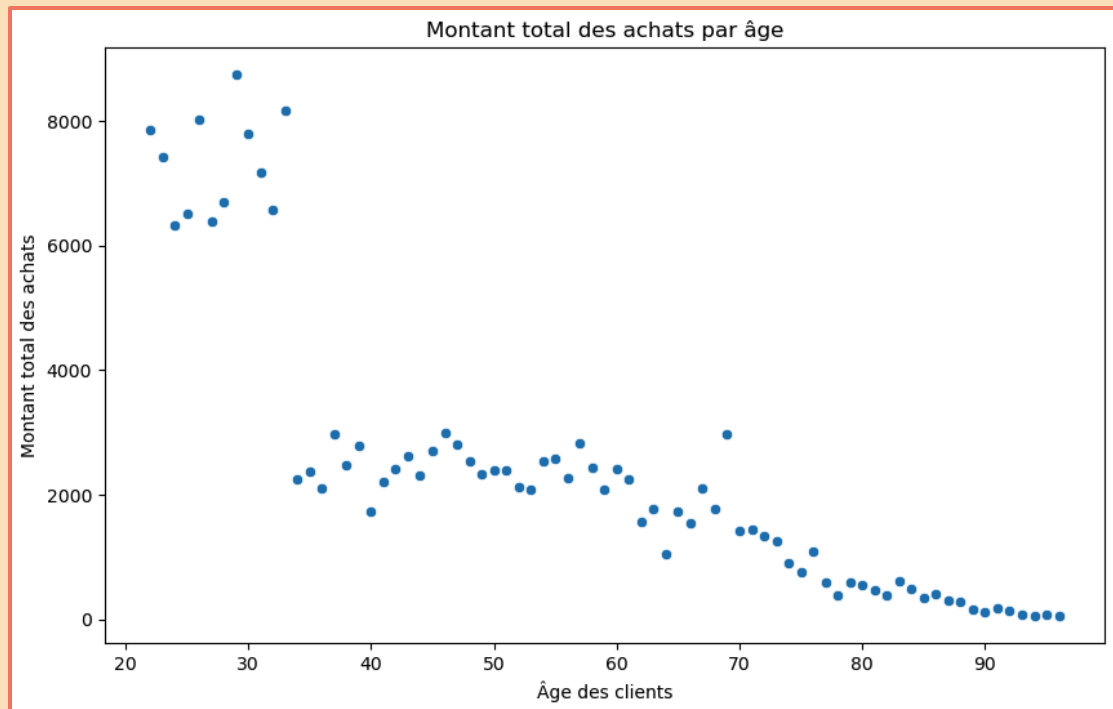
Limites d'âge pour chaque quantile :

	min	max
groupe_quantile_age		
Q1	22	27
Q2	28	32
Q3	33	37
Q4	38	42
Q5	43	47
Q6	48	51
Q7	52	57
Q8	58	63
Q9	64	71
Q10	72	96

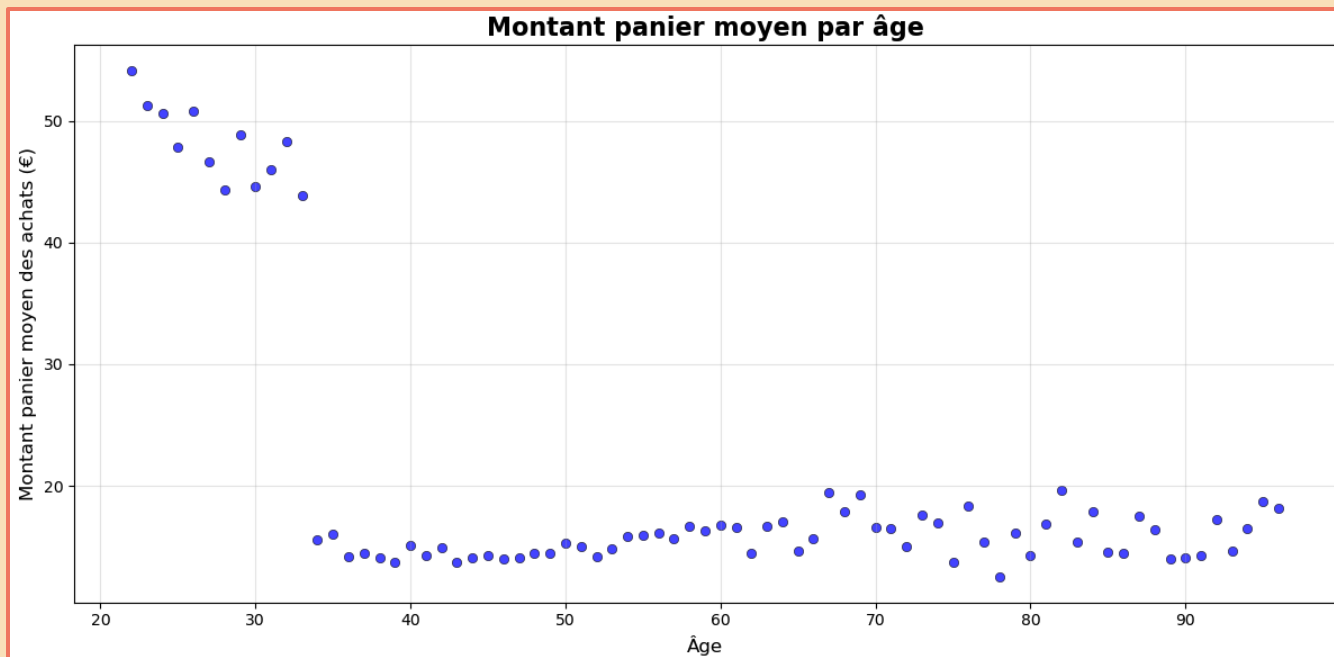
Lien entre genre et catégorie d'achat

categ		0	1	2	Total
sex					
f		60.21%	34.70%	5.09%	100.00%
m		60.70%	34.14%	5.16%	100.00%

Lien entre âge et montant des achats



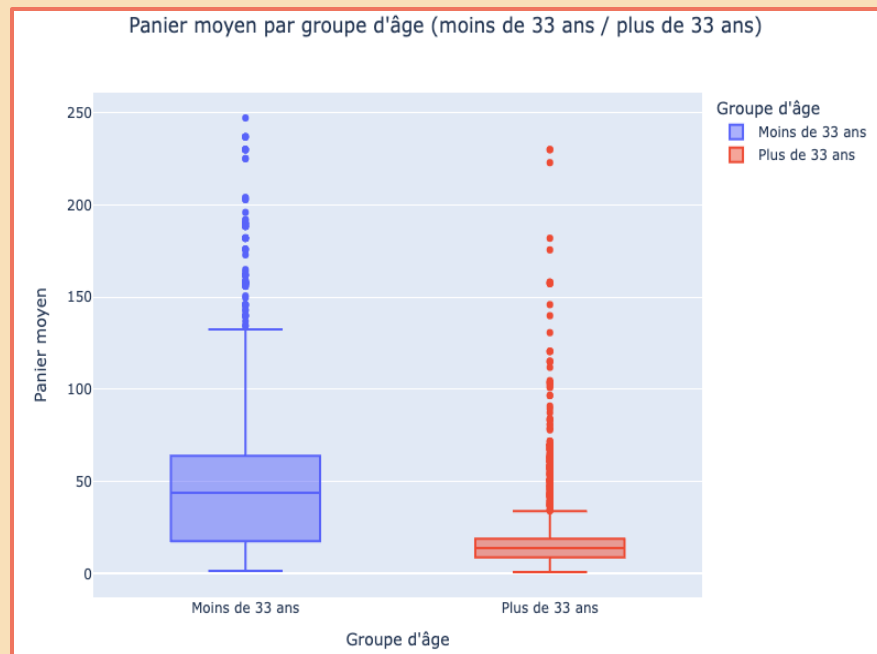
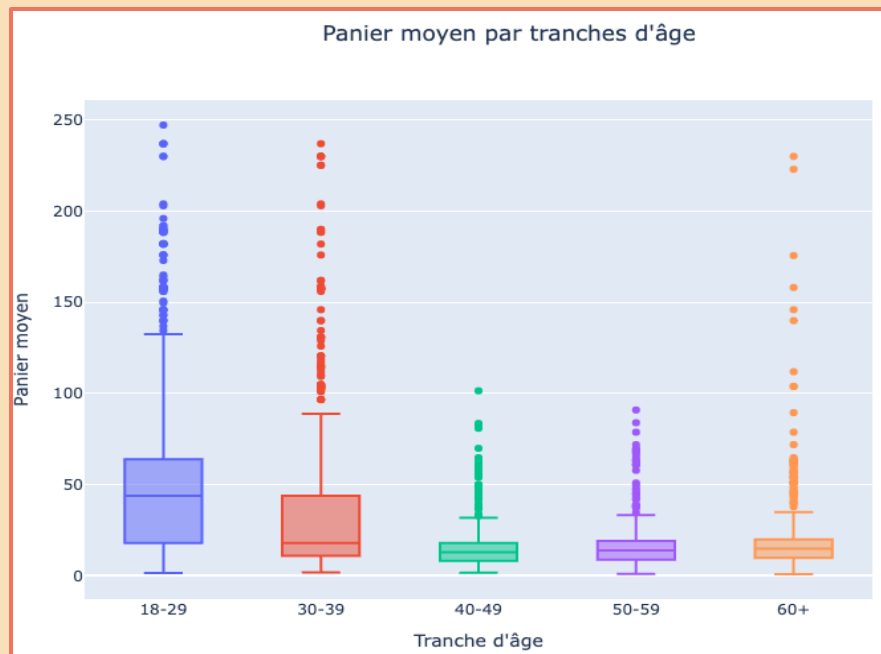
Lien entre âge et panier moyen 1/2



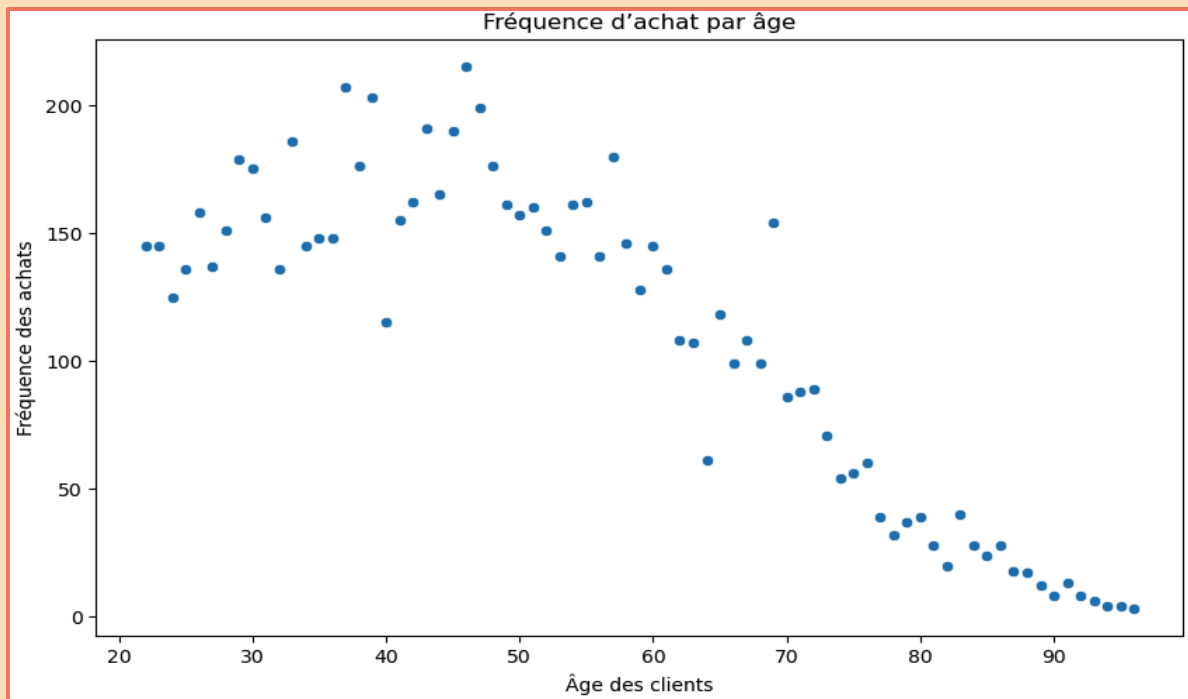
Analyse des corrélations



Lien entre âge et panier moyen 2/2



Lien entre âge et fréquence d'achat

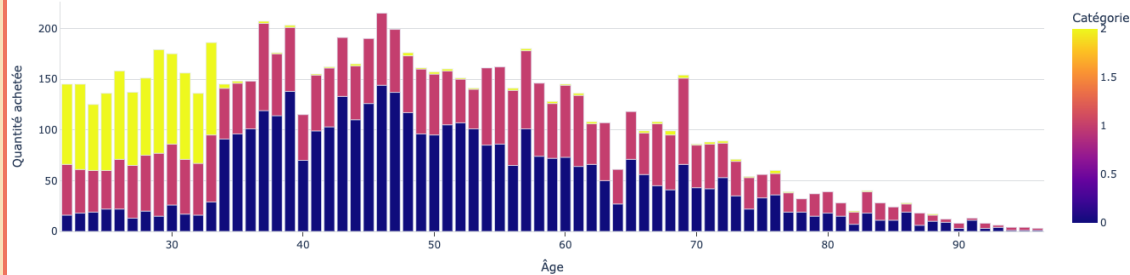


Analyse des corrélations

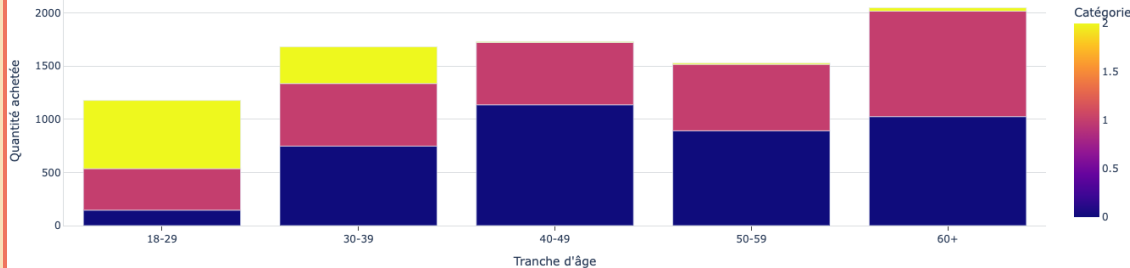


Lien entre âge et catégorie d'achat

Quantité achetée par tranches d'âge et par catégorie



Quantité achetée par âge et par catégorie



	categ 0	categ 1	categ 2	Total
18-29	21.96%	39.01%	39.03%	100.00%
30-39	67.50%	26.20%	6.30%	100.00%
40-49	75.17%	24.48%	0.36%	100.00%
50-59	61.32%	38.08%	0.60%	100.00%
60+	42.79%	56.31%	0.90%	100.00%

Conclusion



- Les **4 clients BtoB** génèrent **7,4 % du CA** et **6,8 % des ventes**
- Les produits de **catégorie 1** sont à l'origine de la **variation des ventes** et du **CA du mois d'octobre 2021**
- La **moitié du CA** est réalisé par les clients **âgés de 33 à 51 ans**
- Les **clients de moins de 33 ans** sont **moins nombreux** mais **dépensent 3x plus** que les autres
- Les produits de **catégorie 2** sont les **plus chers** et sont achetés essentiellement par les **moins de 33 ans**

Recommandations



- ❖ Vérifier et comprendre pourquoi nous avons autant de clients âgés de 21 ans
- ❖ Ajouter aux données les marges de chaque produit et les catégories socio-professionnelles (si possible)
- ❖ Créer une segmentation des clients
- ❖ Etudier les différences entre les ventes physiques et celles en ligne
- ❖ Mettre en place un suivi régulier des indicateurs



Merci pour votre attention