

« **RESTER LIVRES** »

Réalisation d'une analyse des données pour l'entreprise « *Rester livres* »
sur la période 03/21 – 02/2022



SOMMAIRE

Partie I

Nettoyage des données

Partie II

Analyse des données

Partie III

Analyse des corrélations

Partie IV

Conclusion et pistes d'amélioration

DONNÉES RÉCOLTÉES

Nettoyage des données

Récapitulatif des données récoltées et explication du procédé de nettoyage du jeu de données

NETTOYAGE 01

VALEURS ABERRANTES ET MANQUANTES

Avant de procéder à l'analyse de données, nous procédons au nettoyage.

```
# Données sur les transactions
transactions.describe()
```

	id_prod	date	session_id	client_id
count	337016	337016	337016	337016
unique	3266	336855	169195	8602
top	1_369	test_2021-03-01 02:30:02.237413	s_0	c_1609
freq	1081	13	200	12855

On constate qu'une ligne est apparue 200 fois

```
# On observe plus en détails la session s_0 qui est suspecte
transactions.loc[transactions["session_id"]=="s_0"]
```

id_prod	date	session_id	client_id
1431	T_0 test_2021-03-01 02:30:02.237420	s_0	ct_1
2365	T_0 test_2021-03-01 02:30:02.237446	s_0	ct_1
2895	T_0 test_2021-03-01 02:30:02.237414	s_0	ct_1
5955	T_0 test_2021-03-01 02:30:02.237441	s_0	ct_0
7283	T_0 test_2021-03-01 02:30:02.237434	s_0	ct_1
...
332594	T_0 test_2021-03-01 02:30:02.237445	s_0	ct_0
332705	T_0 test_2021-03-01 02:30:02.237423	s_0	ct_1
332730	T_0 test_2021-03-01 02:30:02.237421	s_0	ct_1
333442	T_0 test_2021-03-01 02:30:02.237431	s_0	ct_1
335279	T_0 test_2021-03-01 02:30:02.237430	s_0	ct_0

Ces lignes correspondent à des tests

Valeurs aberrantes

```
# Données sur les prix
products["price"].describe()
```

count	3287.000000
mean	21.856641
std	29.847908
min	-1.000000
25%	6.990000
50%	13.060000
75%	22.990000
max	300.000000
Name: price, dtype: float64	

Un prix négatif apparaît

Il convient de les supprimer

Valeurs manquantes

```
# Jointure entre les df produits et transactions afin de vérifier s'il n'y a pas d'autres données suspectes
transacts_produits = pd.merge(transacts, produits, on="id_prod", how="left")
transacts_produits.info()
```

#	Column	Non-Null Count	Dtype
0	id_prod	336816 non-null	object
1	date	336816 non-null	object
2	session_id	336816 non-null	object
3	client_id	336816 non-null	object
4	price	336713 non-null	float64
5	categ	336713 non-null	float64

dtypes: float64(2), object(4)
memory usage: 18.0+ MB

On constate qu'il y a 103 lignes nulles

```
# On affiche les données "False"
id_prod_false = id_produits[id_produits['id_prod_produits'] == False]
id_prod_false = id_prod_false.groupby('id_prod').mean()
print(id_prod_false)
```

id_prod	price	categ	id_prod_produits	client_id_transactions
0_2245	NaN	NaN	False	True

Les données du produit 0_2245 sont manquantes dans la table produits

```
# Imputation par la moyenne
transacts_produits["categ"] = transacts_produits[["categ"]].fillna(value=0)
transacts_produits["price"] = transacts_produits[["price"]].fillna(value=10.646828)
transacts_produits.info() # on constate qu'il n'y a plus de lignes nulles
```

On procède à une imputation par la moyenne

NETTOYAGE
01

VALEURS ABERRANTES ET MANQUANTES

Avant de procéder à l'analyse de données, nous procédons au nettoyage.

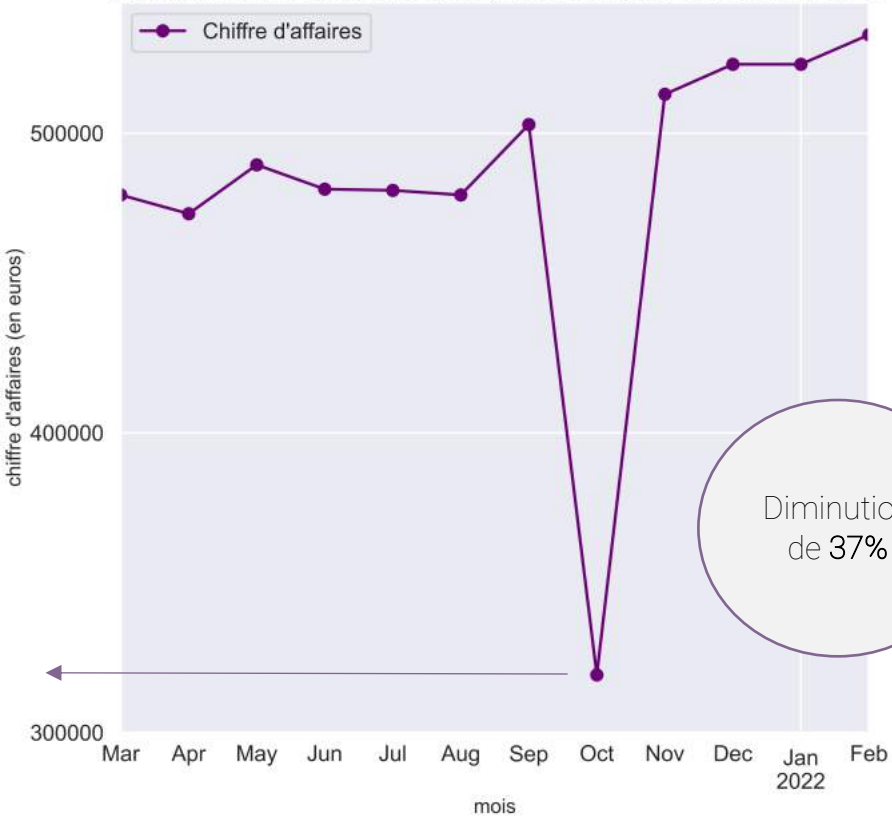
Evolution du volume des ventes
sur la période de mars 2021 à février 2022



Diminution
de 35%

On constate une forte
diminution des ventes et du
chiffre d'affaires au mois
d'octobre

Evolution du chiffre d'affaires sur la période de mars 2021 à février 2022

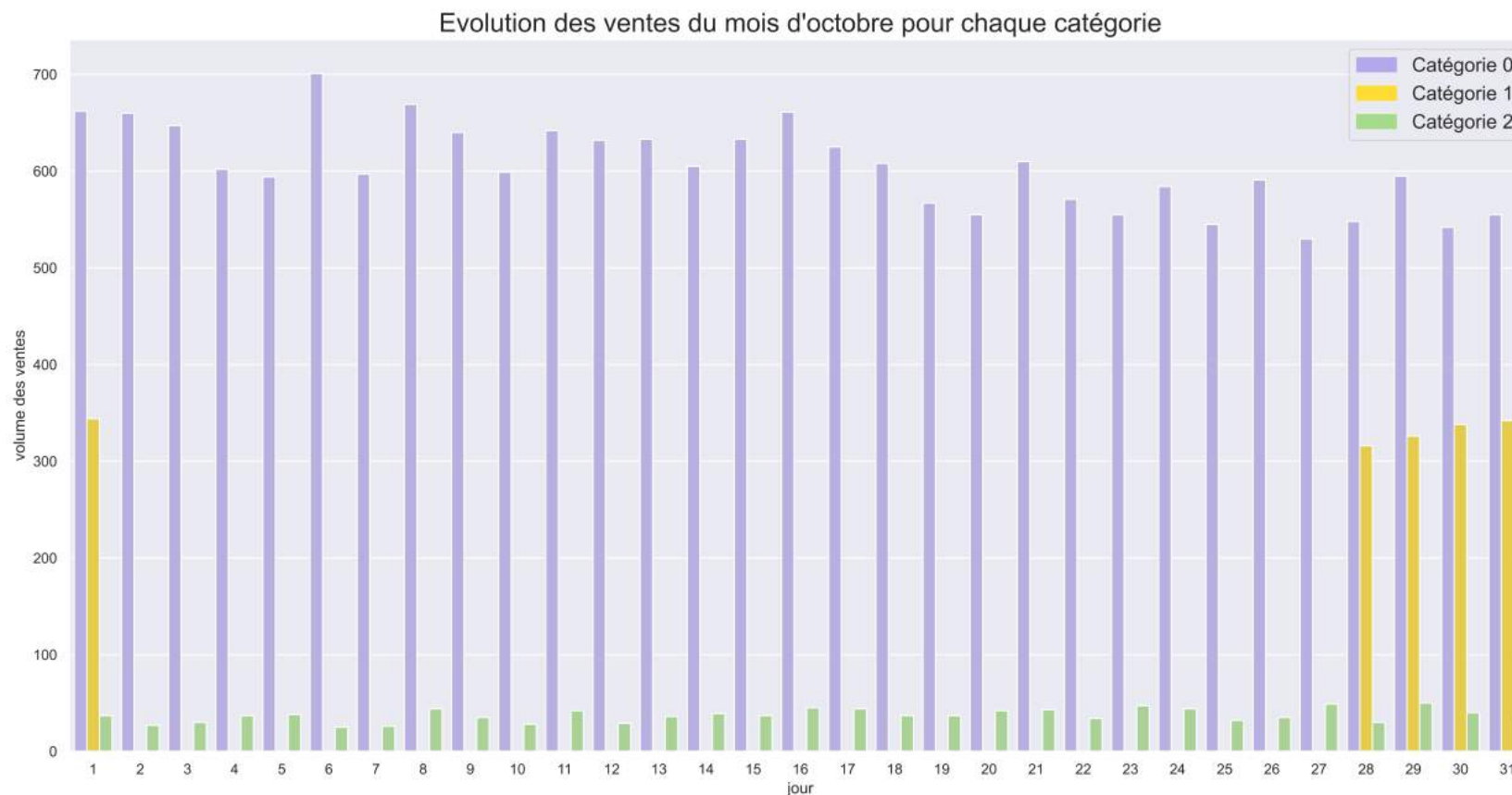


Diminution
de 37%

NETTOYAGE 01

VALEURS ABERRANTES ET MANQUANTES

Le site internet propose à la vente 3 catégories de produits : la catégorie 1, la catégorie 2 et la catégorie 3.



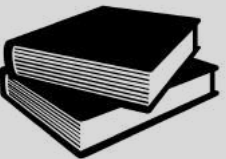
On procède à la **suppression**
des données du mois
d'*octobre* pour ne pas fausser
le reste de l'analyse

Perte de données pour les produits de la catégorie 1 pour les ventes du mois d'*octobre* 2021

« **RESTER LIVRES** »

Analyse des données

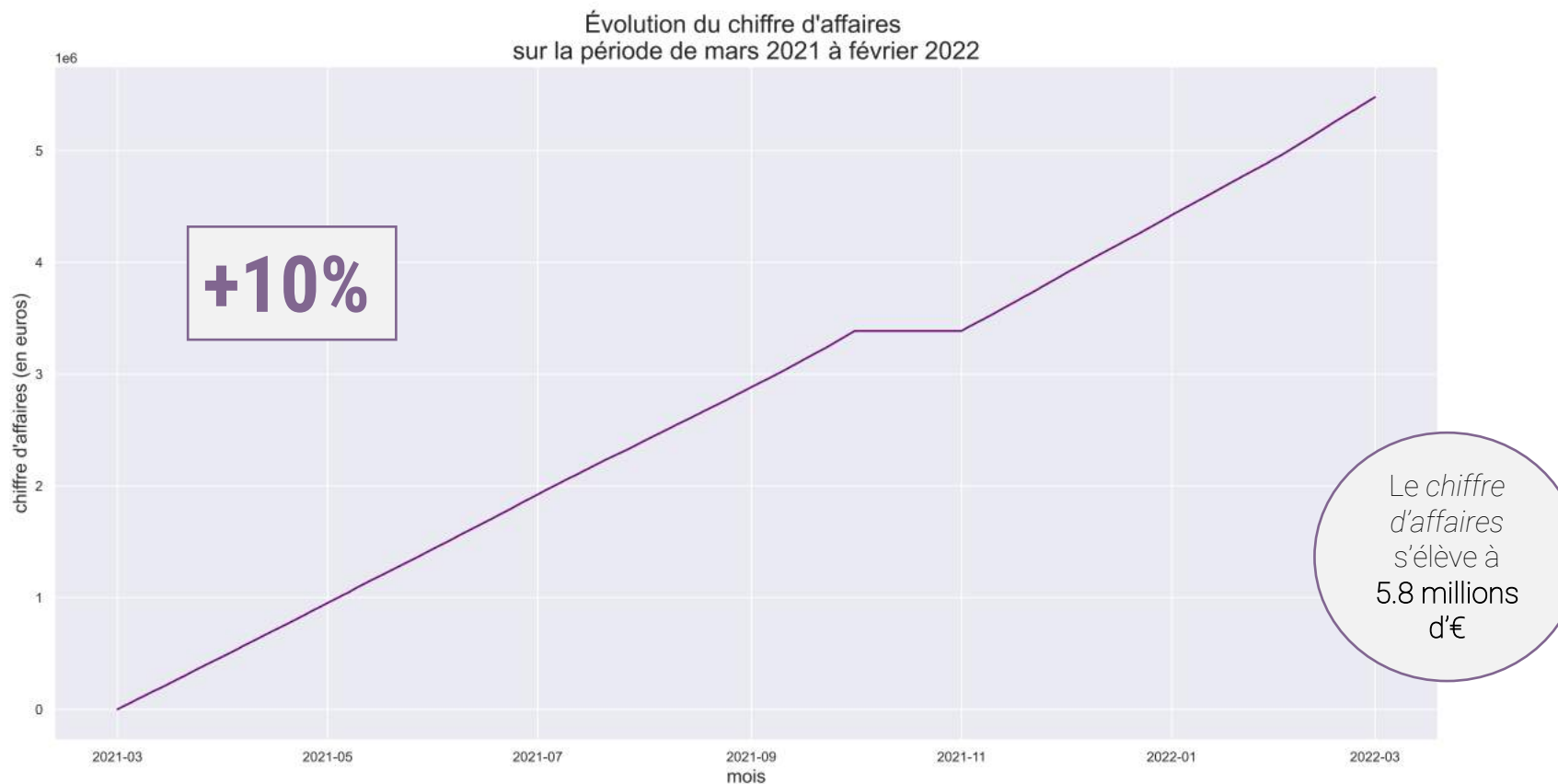
Analyse du comportement des clients et des ventes



ANALYSE 02

EVOLUTION DU CHIFFRE D'AFFAIRES

Analyse de l'évolution de l'évolution du chiffre d'affaires et du volume des ventes sur la période 03/21 – 02/2022.



Le chiffre d'affaires est en augmentation constante sur la période

ANALYSE 02

ÉVOLUTION ANNUELLE DES VENTES

Analyse de l'évolution annuelle des ventes de chaque catégorie.

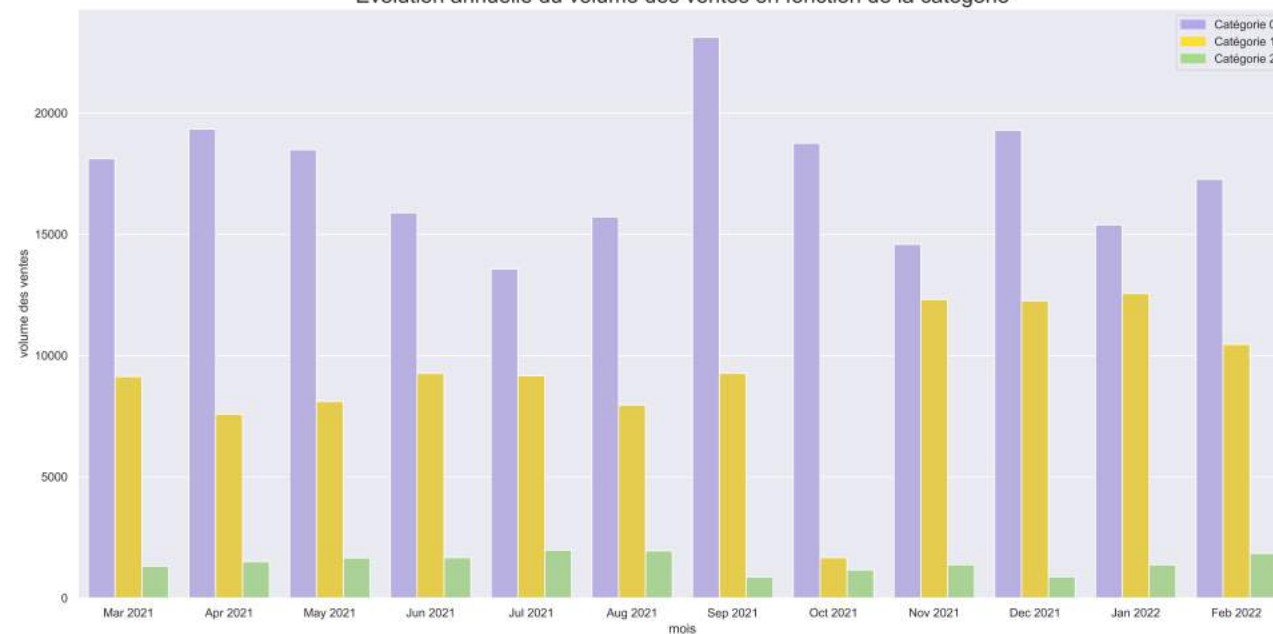
Courbe d'évolution annuelle du volume des ventes de chaque catégorie



Pics pour la rentrée scolaire et les fêtes de fin d'année

On constate que les produits de la **catégorie 0** sont les plus *vendus*.
Vient ensuite la **catégorie 1** et la **catégorie 2**.

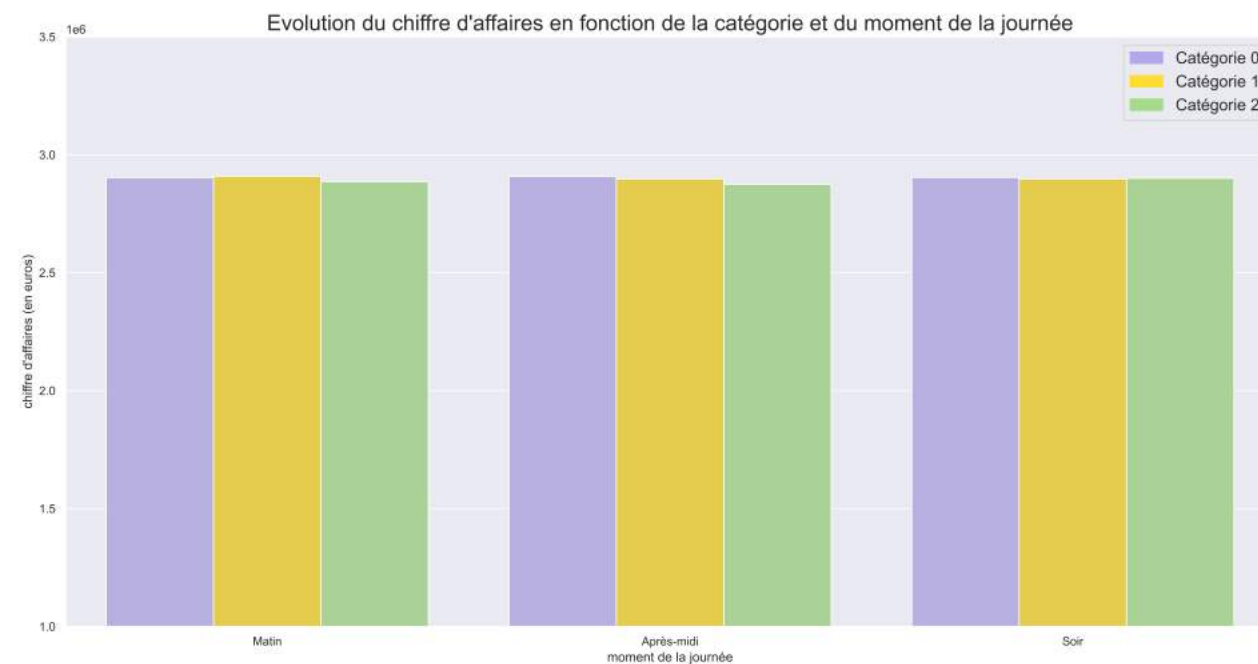
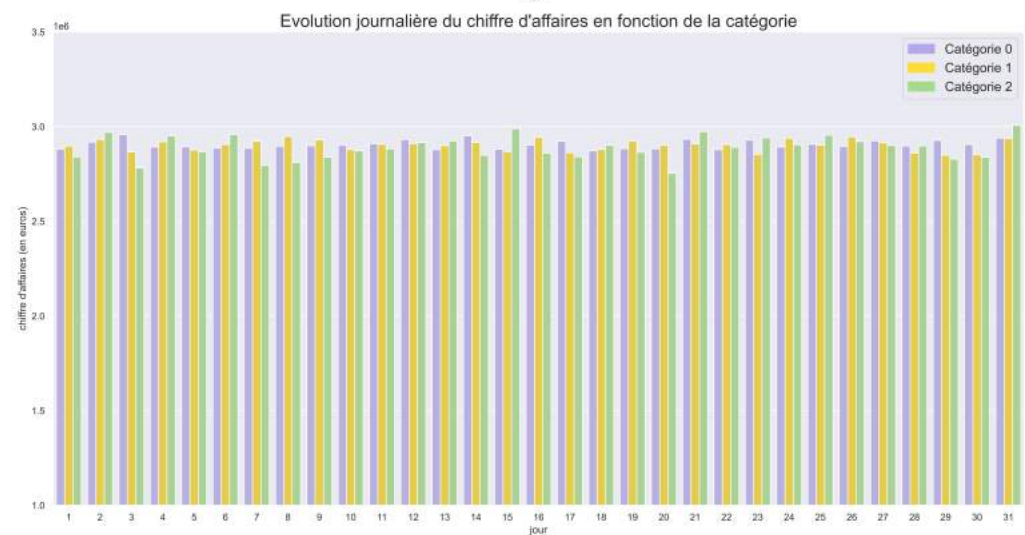
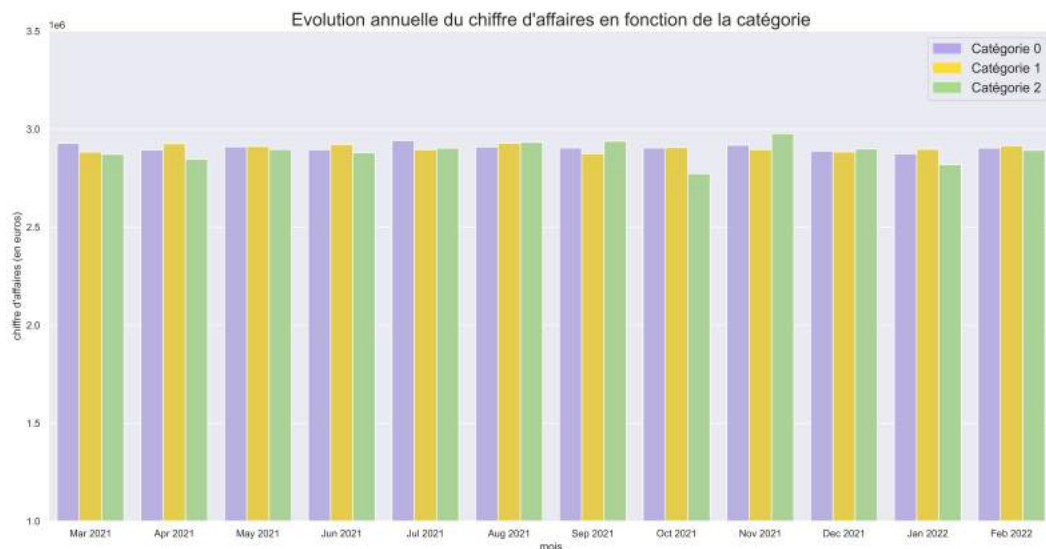
Evolution annuelle du volume des ventes en fonction de la catégorie



ANALYSE 02

EVOLUTION DU CHIFFRE D'AFFAIRES DE CHAQUE CATÉGORIE

Analyse du chiffre d'affaires de chaque catégorie en fonction : évolution journalière, évolution en fonction du moment de la journée et évolution annuelle.

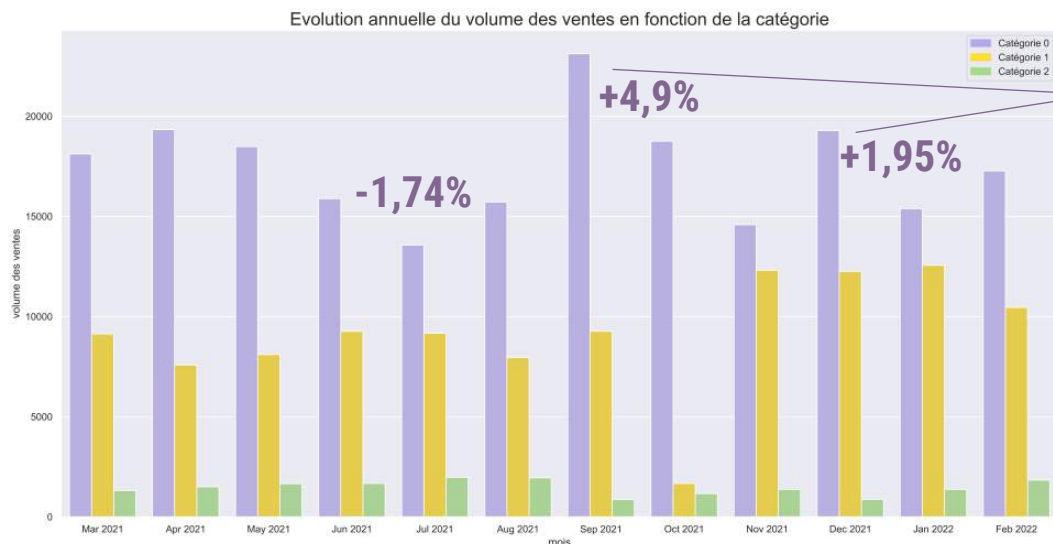


Le chiffre d'affaires est **constants** quels que soient la catégorie, le mois, le jour et le moment de la journée.

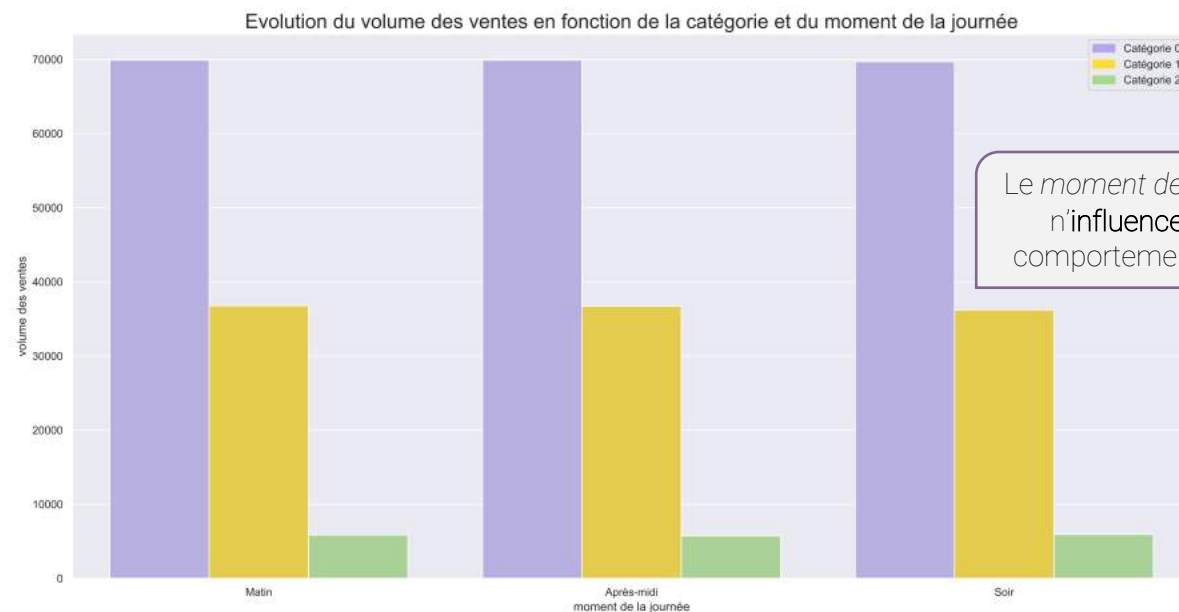
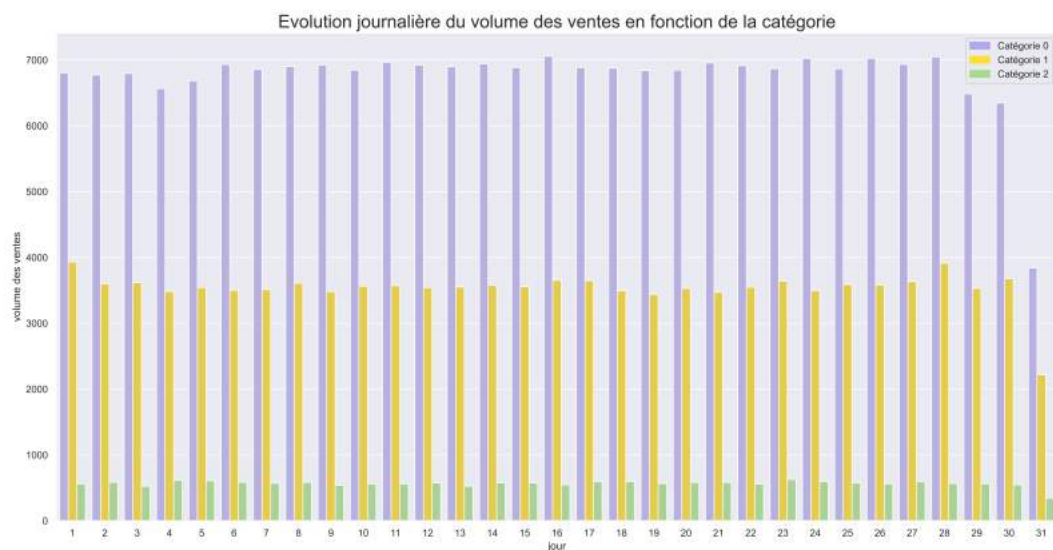
ANALYSE 02

EVOLUTION DES VENTES DE CHAQUE CATÉGORIE

Analyse des ventes de chaque catégorie en fonction : évolution journalière, évolution en fonction du moment de la journée et évolution annuelle.



Pics pour la rentrée scolaire et les fêtes de fin d'année



Le moment de la journée n'influence pas le comportement d'achat

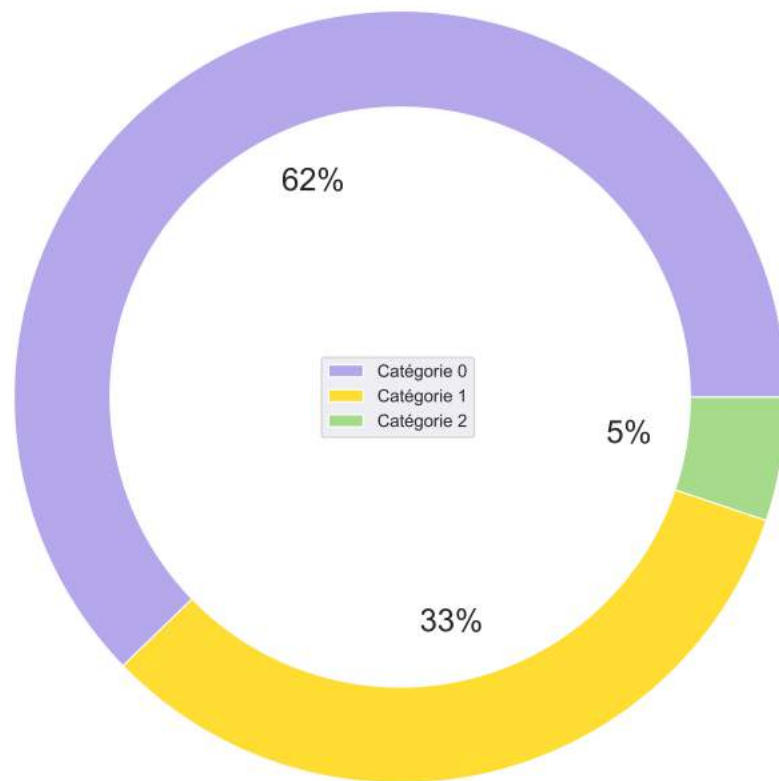
Le **mois** nous permet de nous donner une idée des comportements d'achat : On observe un pic aux mois de *septembre* et *décembre* ainsi qu'une baisse sur les mois de *juin* et *juillet*.

ANALYSE 02

CONTRIBUTION DE LA CATÉGORIE SUR LE CHIFFRE D'AFFAIRES

Analyse de la proportion de chaque catégorie sur le chiffre d'affaires.

Contribution de la catégorie sur le chiffre d'affaires



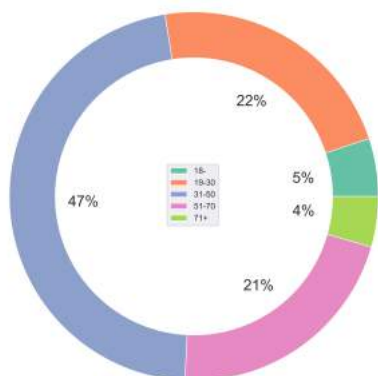
La **catégorie 0** est la catégorie *la plus achetée* par les clients. Elle représente une part de **62%** du *chiffre d'affaires*.

ANALYSE 02

SATISFACTION CLIENT ET FIDÉLITÉ

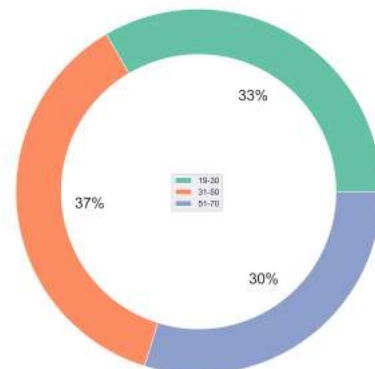
Contribution de chaque tranche sur le chiffre d'affaires, taux de rachat et taux de churn.

Contribution de chaque tranche d'âge sur le chiffre d'affaires



Les achats de la tranche d'âge 31-50 représentent 47% du *chiffre d'affaires*.
Ceux de la tranche 18- et 71+ représentent le pourcentage *le moins élevé*.

Contribution sur le chiffre d'affaires des grands comptes en fonction de leur âge



La tranche 31-50 ans contribue à hauteur de 37% au chiffre d'affaires

Clients les plus fidèles : **grands comptes** (key customers)

7,07% Proportion des achats des gros comptes sur le CA

7% Proportion des achats des gros comptes sur le **volume des ventes**

entre **2457** et **11839** fréquence d'achat

65% des produits achetés par un client proviennent de la **même catégorie**

96% **taux de rachat** (8285 clients ont de nouveau effectué un achat)

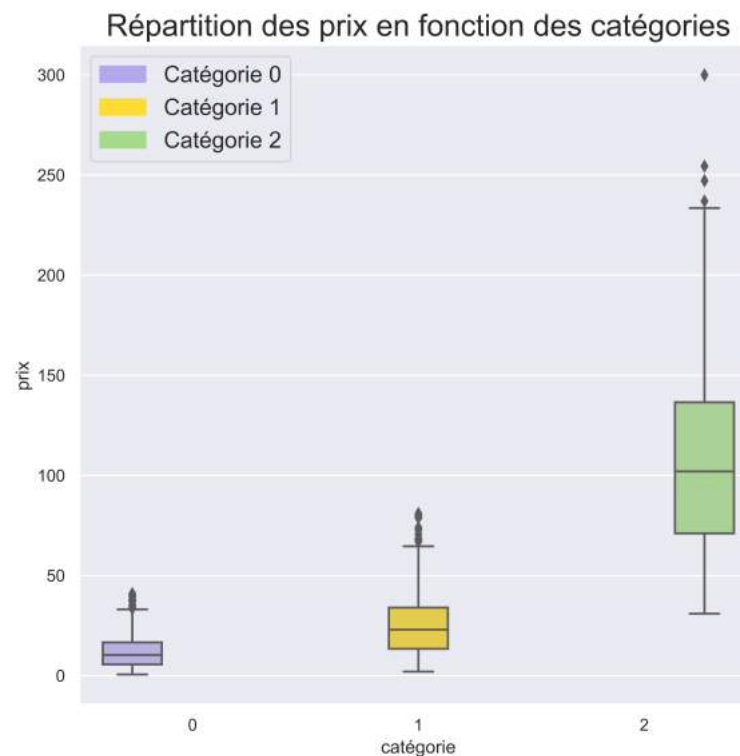
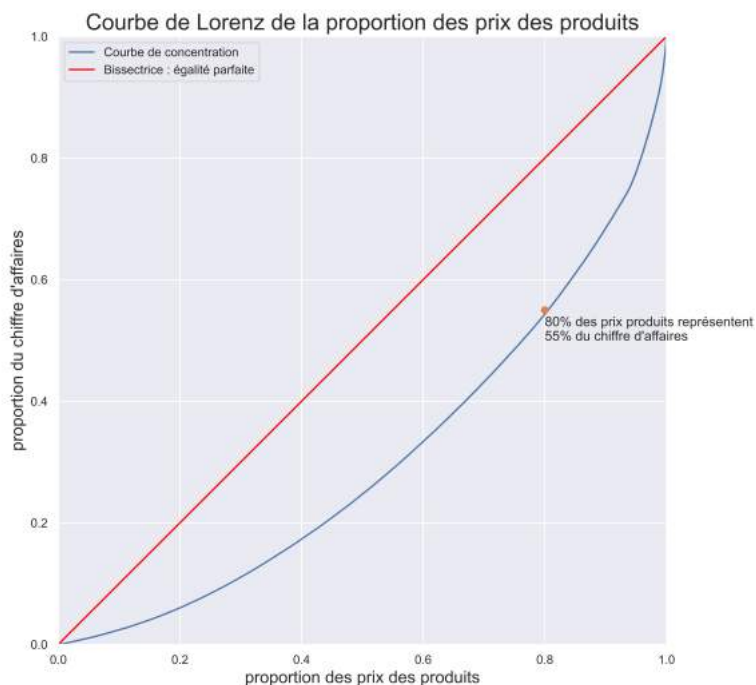
3,9% Churn rate / **taux d'attrition** (266 clients n'ont pas effectué de nouvel achat)

4 fréquence d'achat

ANALYSE 02

PRIX ET PROPORTION

Analyse de la répartition des prix, des valeurs moyennes de chaque catégorie et proportion des prix sur le chiffre d'affaires.



Les prix de la catégorie 2 sont **plus dispersés** que ceux des *autres catégories*.

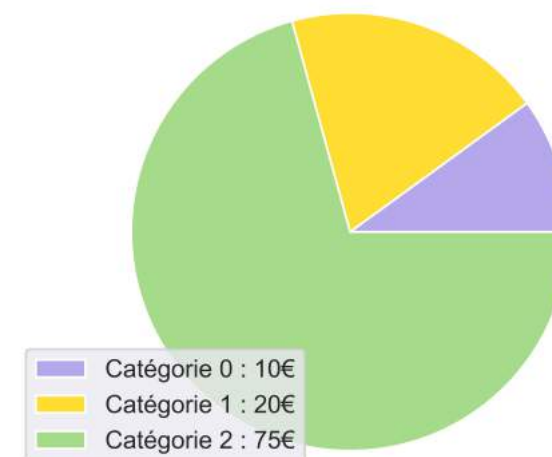
Écart-type de la catégorie 2 : 39€

Écart-type de la catégorie 1 : 7€

Écart-type de la catégorie 0 : 4€

La **catégorie 0** est la catégorie de produits *la moins chère* tandis que la **catégorie 2** correspond aux produits *les plus chers*.

Répartition des valeurs moyennes des transactions par catégorie

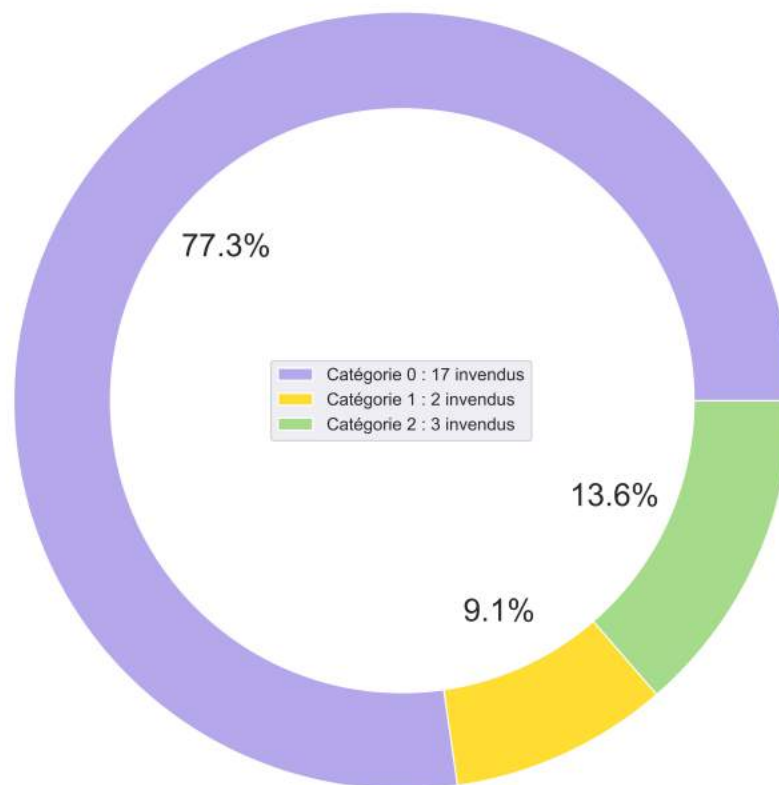


ANALYSE 02

PRODUITS INVENDUS

Analyse des produits invendus en fonction de la catégorie.

Produits invendus en fonction de la catégorie



22 produits invendus.
Majoritairement des produits de la **catégorie 0** (77.3%)

→ Le *montant total* des produits invendus s'élève à 798€

ANALYSE 02

PANIER MOYEN

Analyse du panier moyen et comparatif entre les clients les plus fidèles et les clients lambda.

Clients lambda

18,77€

1 à 3 articles achetés par session (moyenne 1,9)

Clients les plus fidèles : **grands comptes** (key customers)

21,29€

2 articles achetés par session (écart-type 1.3)

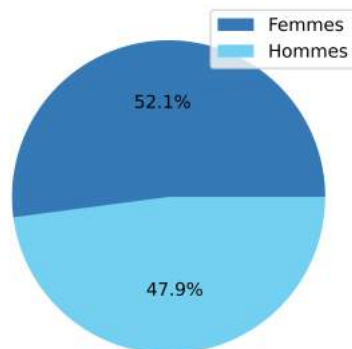
Pas de grande différence au niveau *du nombre d'articles achetés*. Ce qui les distingue c'est la fréquence d'achat

ANALYSE 02

CLIENTS ET VISITEURS

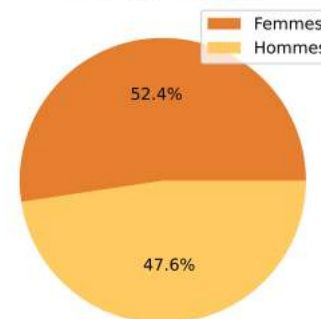
Les clients ont effectué au moins un achat tandis que les visiteurs n'en n'ont effectué aucun.

Genre des clients



Pas grande différence.
Les femmes un peu plus représentées.

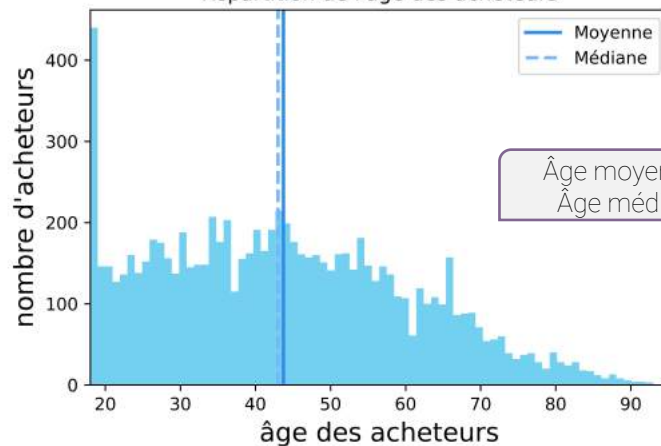
Genre des visiteurs



Taux de conversion : 2.56%

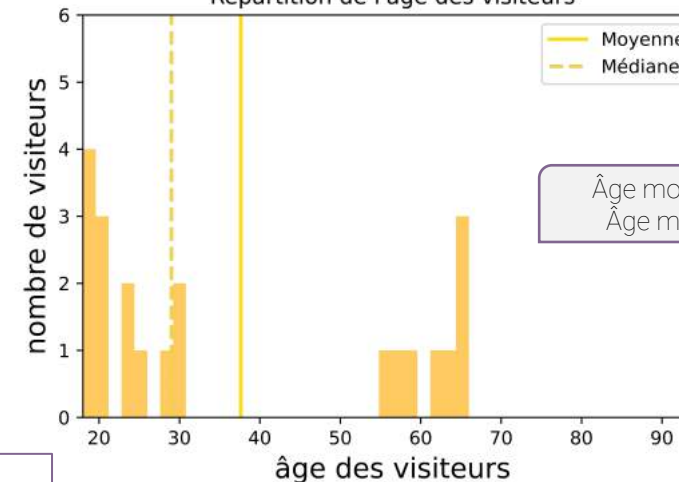
* proportion d'achats sur un site par rapport au nombre de visiteurs ayant parcouru le site

Répartition de l'âge des acheteurs



On constate un pic à 18 ans.
Âge légal pour effectuer un achat ou souhait de confidentialité

Répartition de l'âge des visiteurs



CORRÉLATIONS

Analyse du comportement des clients

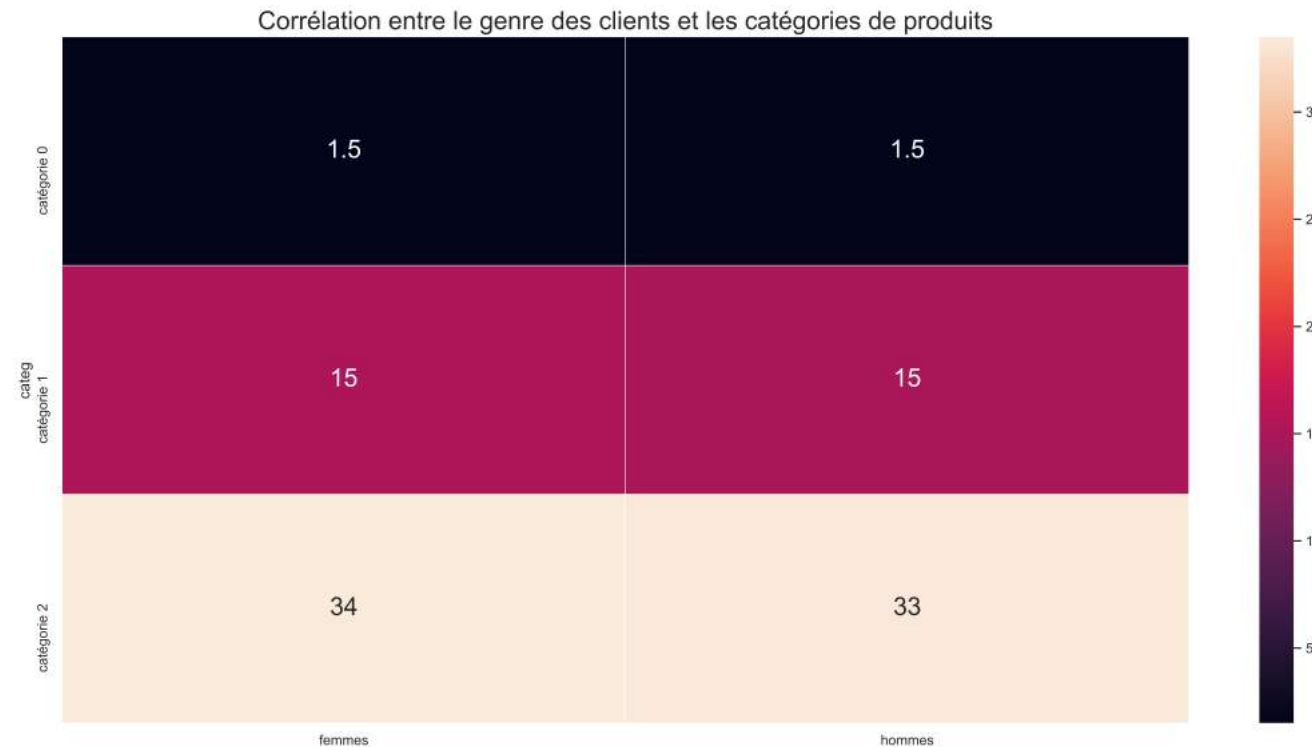
Corrélations entre différentes variables



CORRÉLATION 03

CORRÉLATION ENTRE LE GENRE DES CLIENTS ET LES CATÉGORIES DE PRODUITS

Analyse des deux variables : le genre des clients et les catégories de produits.



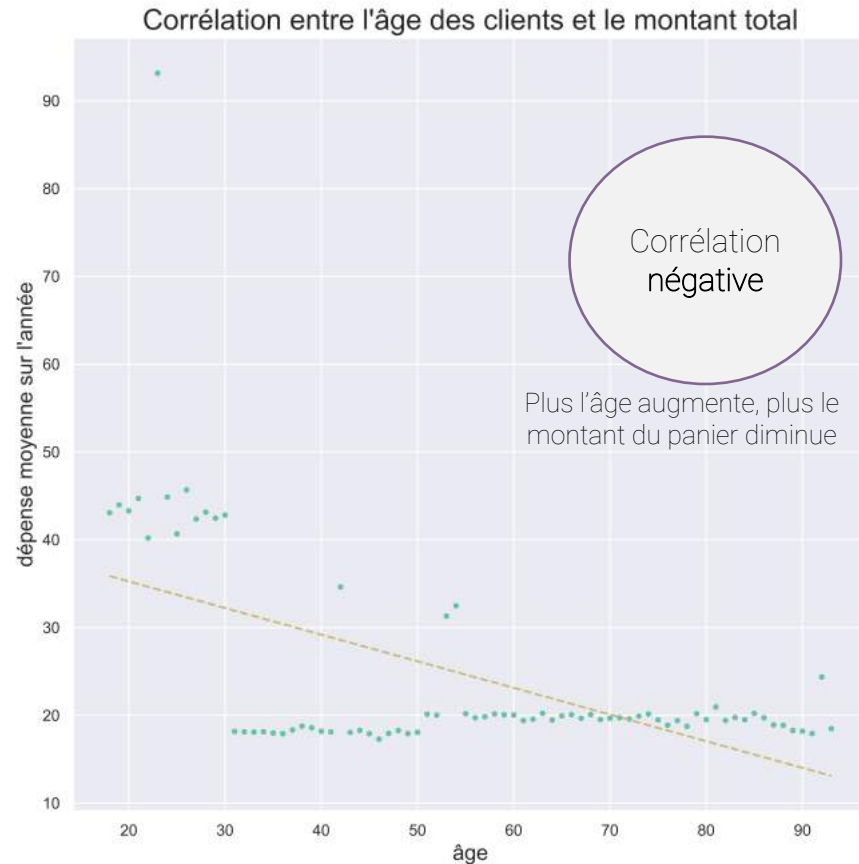
La catégorie 2 est la plus achetée par les femmes

CORRÉLATION

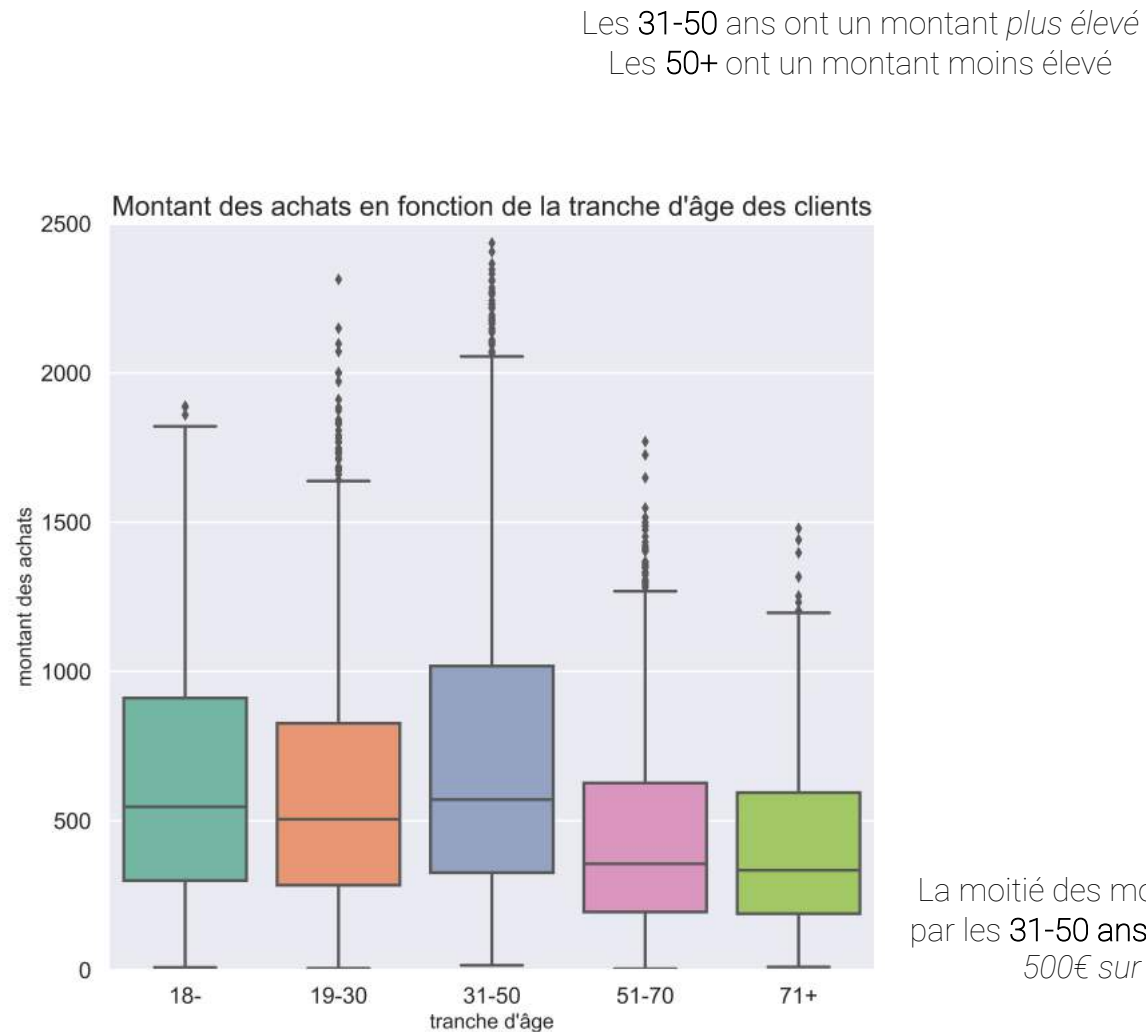
03

CORRÉLATION ENTRE L'ÂGE DES CLIENTS ET LE MONTANT TOTAL

Analyse des deux variables : âge / tranche d'âge et montant total des achats.



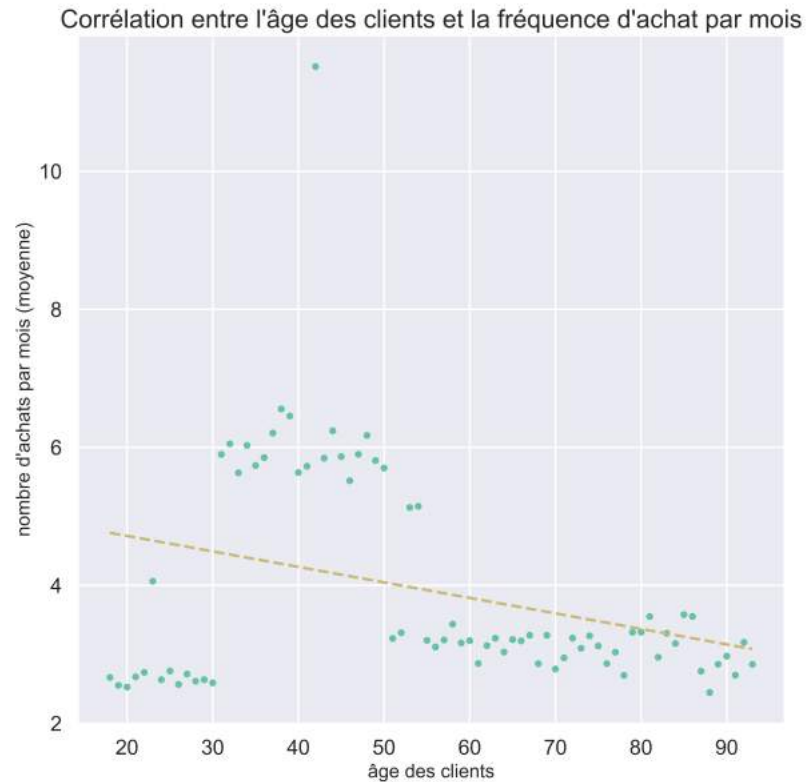
3 tranches d'âge se distinguent et consomment *de la même façon*



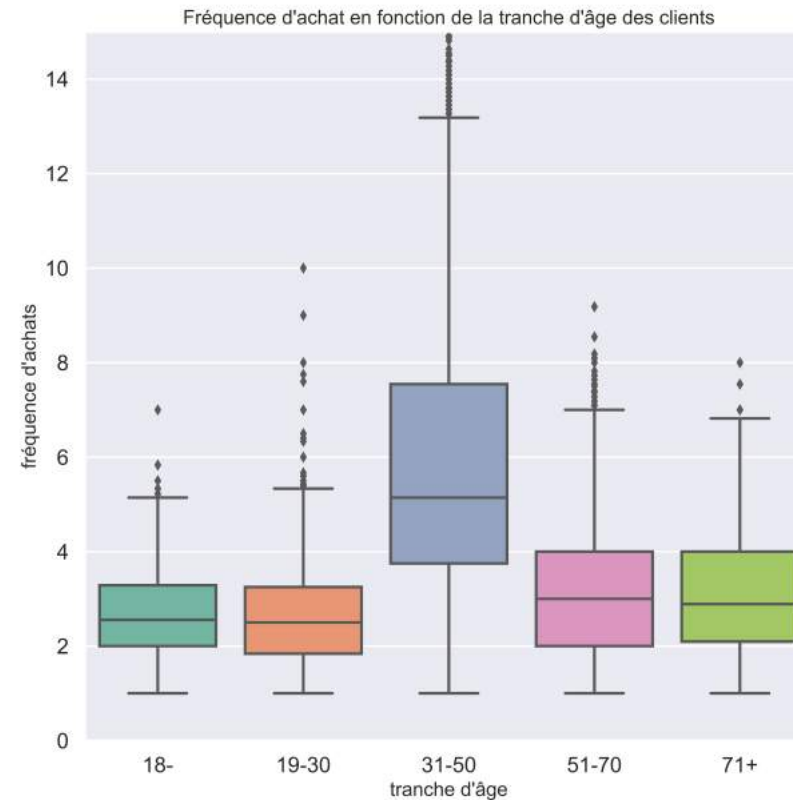
CORRÉLATION 03

CORRÉLATION ENTRE L'ÂGE DES CLIENTS ET LA FRÉQUENCE D'ACHAT

Le site Analyse des deux variables : âge / tranche d'âge des clients et fréquence d'achat.



3 tranches d'âge se distinguent et consomment de la même façon



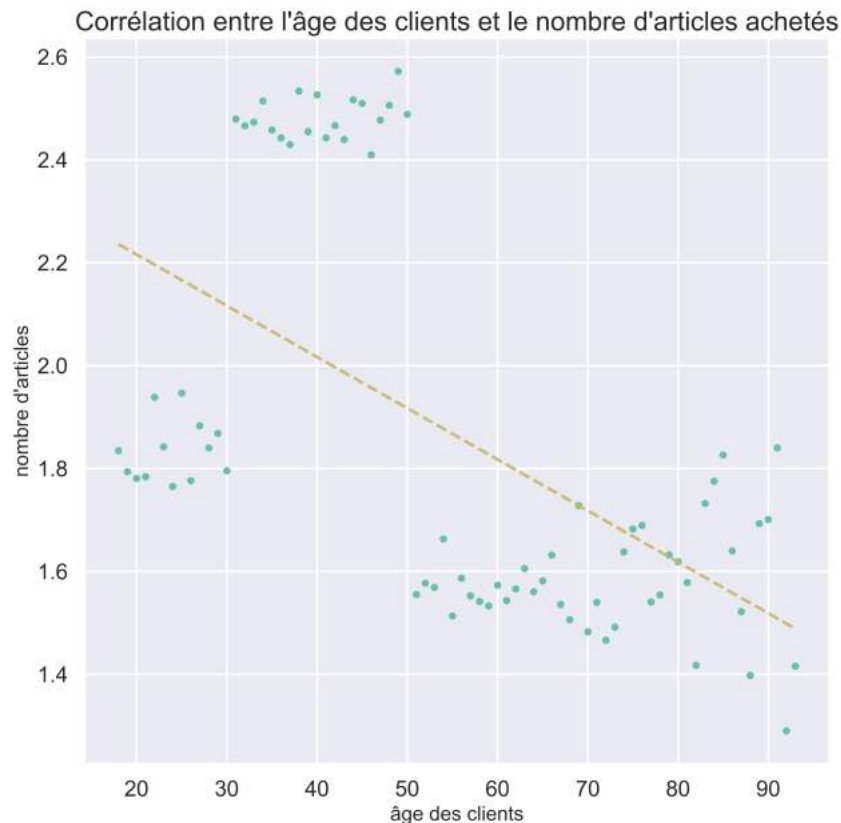
Les 31-50 ans ont une fréquence d'achat plus élevée
Les 30- ans ont une fréquence moins élevée

50% des fréquences d'achat sont supérieures à 5 pour la tranche d'âge 31-50 ans

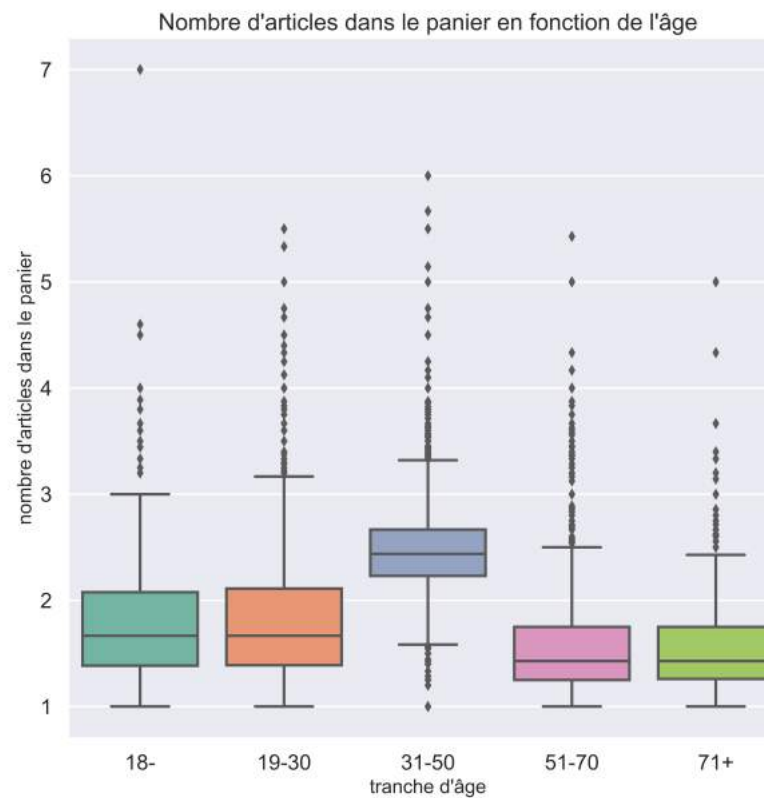
CORRÉLATION 03

CORRÉLATION ENTRE L'ÂGE DES CLIENTS ET LE NOMBRE D'ARTICLES ACHETÉS

Analyse des deux variables : âge / tranche d'âge et nombre d'articles dans le panier moyen.



3 tranches d'âge se distinguent et consomment *de la même façon*



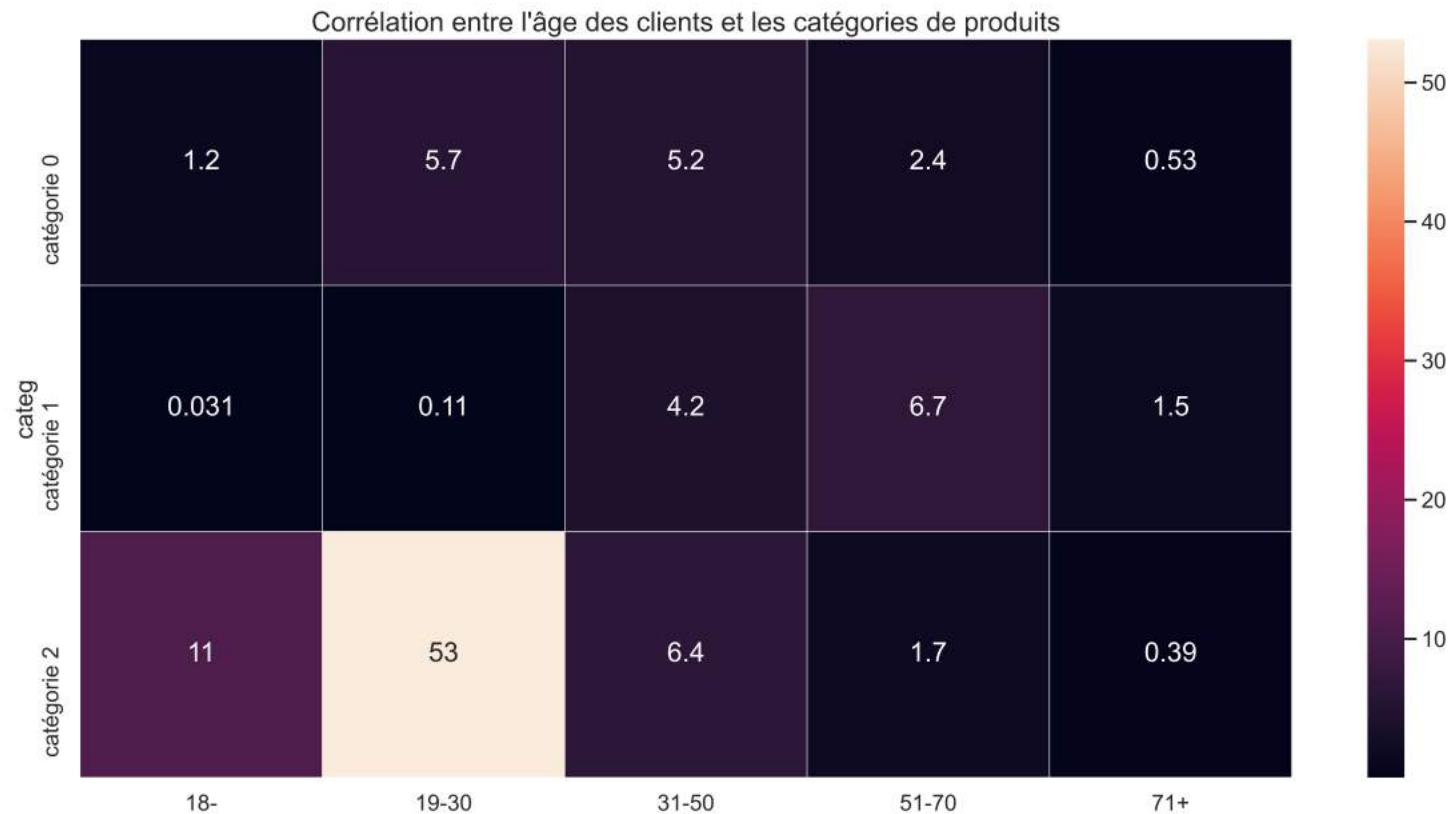
Les 31-50 ans ont *plus d'articles* dans leur panier
Les 50+ ans ont *moins d'articles* dans leur panier

50% des articles *par panier* sont supérieurs à 2,5 pour la tranche 31-50 ans

CORRÉLATION 03

CORRÉLATION ENTRE L'ÂGE DES CLIENTS ET LES CATÉGORIES DE PRODUITS

Analyse de la relation entre les deux variables : âge / tranche d'âge des clients et les catégories de produits.



La catégorie 2 est la plus achetée par les 19-30 ans

CONCLUSION

Bilan des analyses réalisées

Conclusion globale et focus sur les points essentiels



Le **chiffre d'affaires** sur la période s'élève à 5,8 millions d'euros. Ce chiffre est en *augmentation*.

On observe des *pics de ventes* en **septembre** (rentrée scolaire) et en **décembre** (fêtes de fin d'année). Mais une *baisse* pendant l'été.

4 clients effectuent des achats avec une fréquence *très élevée*. Leurs dépenses s'élèvent à 410 025€ ce qui représente 7,07% du CA et 7% du volume des ventes.

3 tranches d'âge se distinguent : leur comportement d'achat est le même.

RÉCAPITULATIF 04

PERSONAS

3 tranches d'âge se distinguent : comparatif entre les comportements des consommateurs (profils types).

Moins de 30 ans

27% contribution au chiffre d'affaires

500€ montant médian des dépenses
sur la période

2,5 fréquence d'achat

1,5 articles par panier

2 catégorie la plus
achetée

31-50 ans

47% contribution au chiffre d'affaires

600€ montant médian des dépenses
sur la période

5 fréquence d'achat

2,5 articles par panier

La tranche d'âge la plus importante

Plus de 50 ans

25% contribution au chiffre d'affaires

400€ montant médian des dépenses
sur la période

3 fréquence d'achat

1,2 articles par panier

1 catégorie la plus
achetée

PISTES D'AMÉLIORATION

Pistes d'amélioration afin d'augmenter le chiffre d'affaires et de continuer dans cette tendance positive.

Pistes d'amélioration :

- Fréquence d'achat : les 30- ans sont ceux qui achètent le moins souvent
- Nombre d'articles dans le panier, ce qui augmentera le montant du panier et donc le chiffre d'affaires
- Fidélité à une catégorie : 65% des achats viennent de la même catégorie
- Diminution des ventes l'été : période où l'entreprise connaît une diminution de son chiffre d'affaires

MERCI

Anissa MANSOUR

parcours Data Analyst (2020/2021)

