# Analysez les ventes d'une librairie

Stéphane Srsa
FAO

# Contexte

Lapage

C’est une librairie physique avec plusieurs points de vente.

Devant le succès de certains produits et l’engouement de ses clients, un site de vente en ligne est en service depuis 2 ans.

## Situation actuelle

Fraîchement recruté en tant que Data Analyst, ma première mission est de faire le point sur l’activité en ligne.

- Préparation et nettoyage des données en vue des analyses.
- Analyser les indicateurs de ventes : les chiffres clés, KPI, représentations graphiques.
- Vérifier certaines corrélations à la demande de Julie.

# Méthodologie de l'analyse

## Information concernant les fichiers utilisés pour l'étude

customers : Identifiant client – genre – année de naissance

products : Identifiant produit en ligne – prix unitaire – catégorie

transactions : Identifiant produit en ligne – date de session –
Identification session - Identifiant client

## Analyse

- Analyse exploratoire des différents fichiers.
- Mise en cohérence des colonnes, nettoyage des données, vérification des types de données et de l'unicité de la clé choisie.
- Création d'une table unique après rapprochement des 3 tables.
- Produit 0_2245 : imputation du prix moyen de la catégorie 0.0
- Octobre 2021 : Ca quotidien de la catégorie 1.0 récupéré à la comptabilité
- Réponses aux demandes d'Antoine et de Julie

# Chiffres d'affaires total et mensuel sur les deux dernières années

## Chiffres d'affaires réalisés

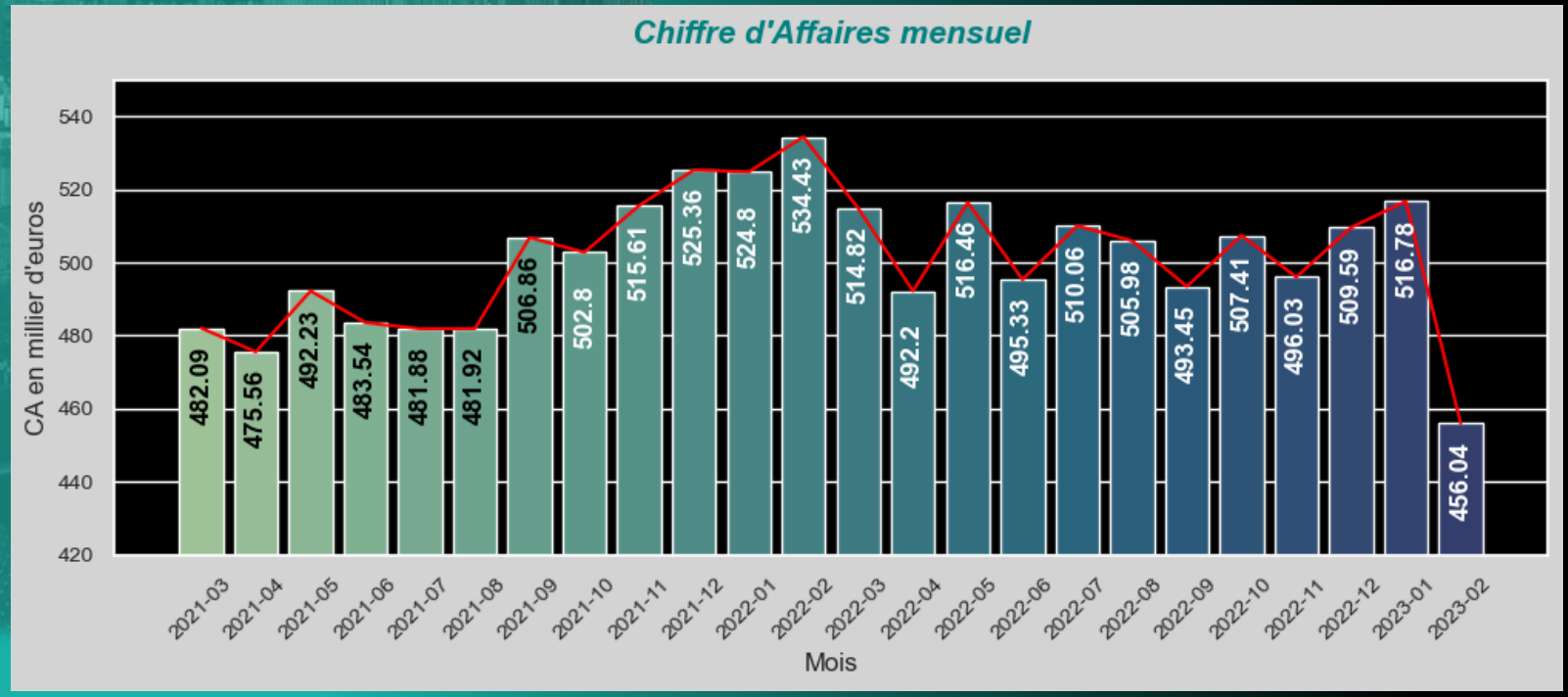
***Chiffre d'affaires total sur 2 ans:*** ***12,021.23 millions d'euros***

***Chiffre d'affaires N :*** ***6,014.15 millions d'euros***

***Chiffre d'affaires N-1 :*** ***6,007.08 millions d'euros***

***Taux de croissance :***

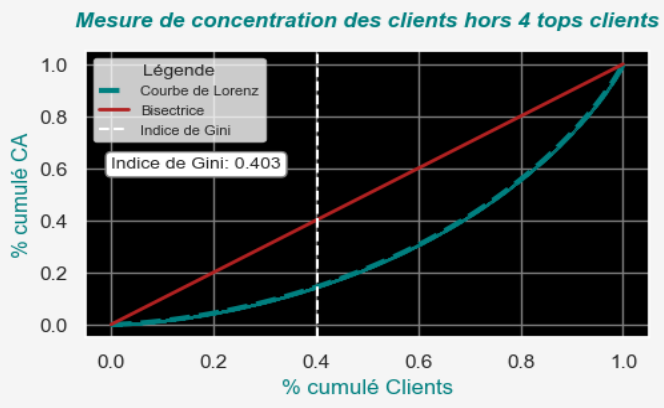
***0,12%***

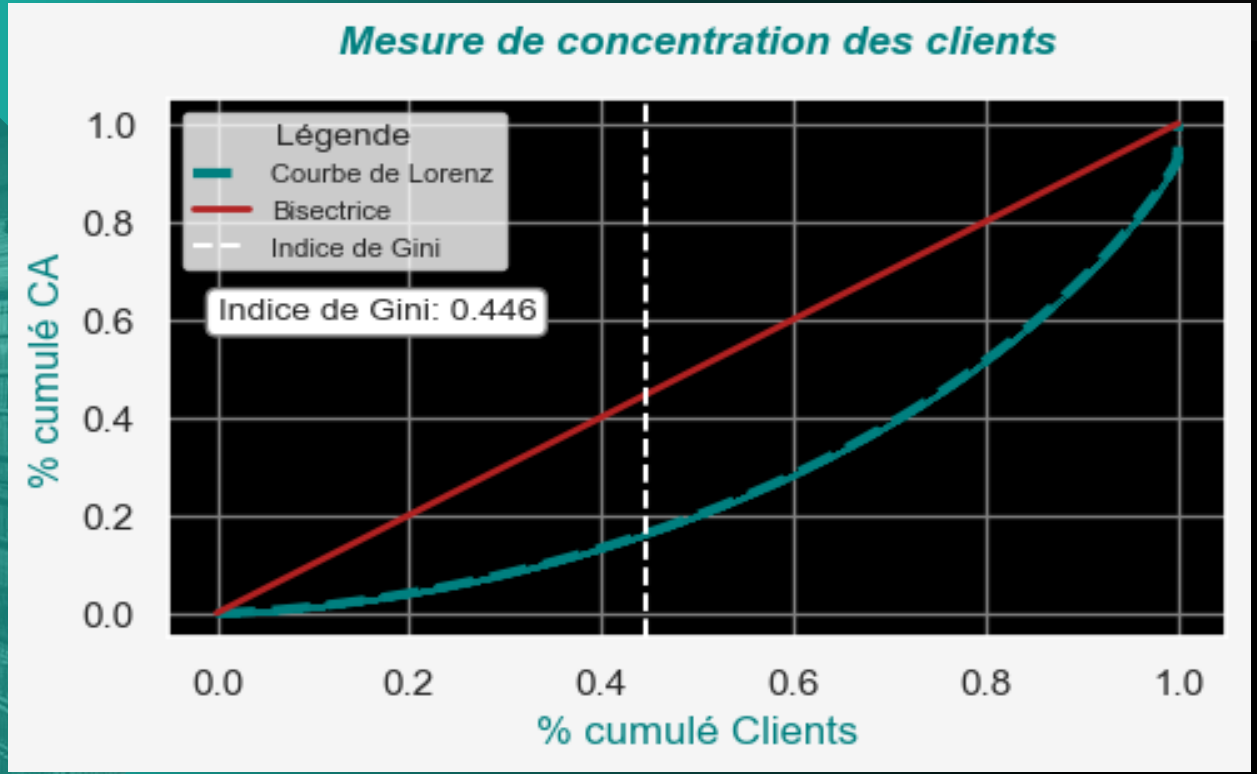


# Répartition du CA – Courbe de Lorenz

## Méthode :

- Visualisation via une courbe de Lorenz
- Calcul de l'indice de Gini




## Observations:

- La distance entre la courbe de Lorenz et la ligne d'égalité montre un degré d'inégalité significatif.
- L'indice de Gini de 0.446 confirme l'inégalité entre les variables.

Tant la courbe de Lorenz que l'indice de Gini ne permettent pas de déterminer une équité ou une inégalité de niveau extrême.

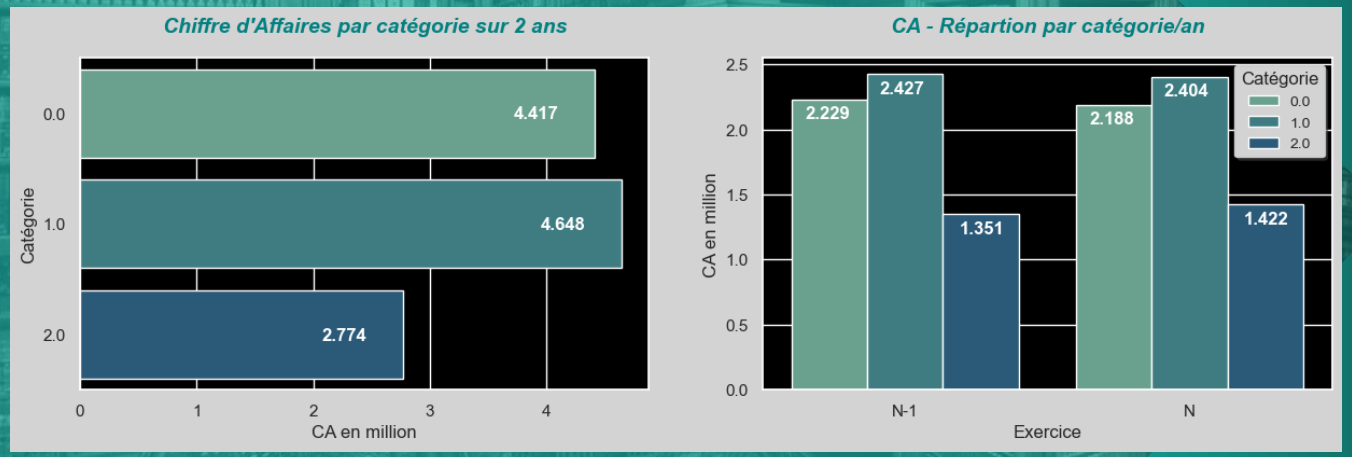
# Chiffre d'affaires par catégorie

## Chiffres d'affaires réalisés par catégorie

***Catégorie 0.0:*** ***4,417 millions d'euros***

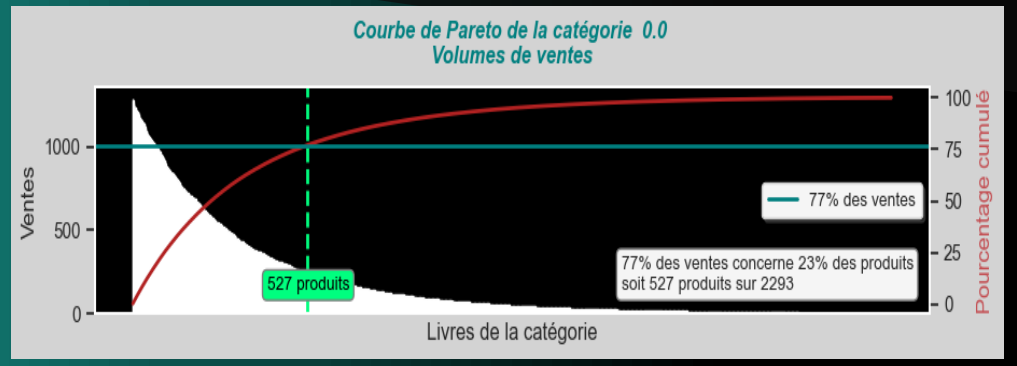
***Catégorie 1.0:*** ***4,648 millions d'euros***

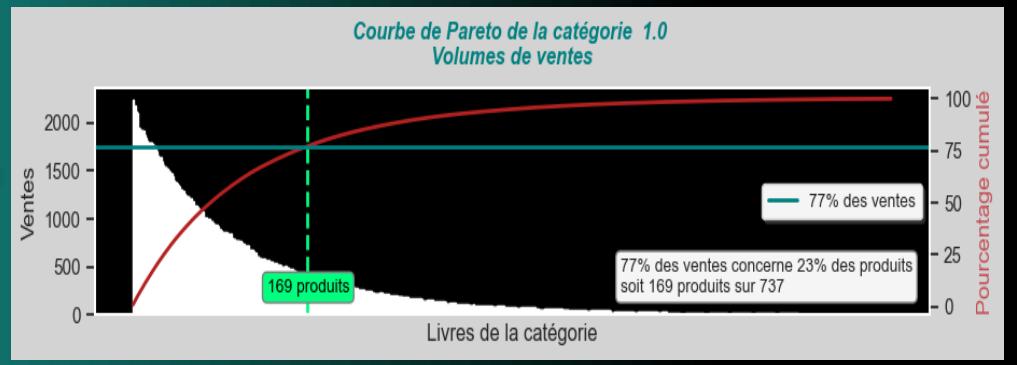
***Catégorie 2.0:*** ***2,774 millions d'euros***

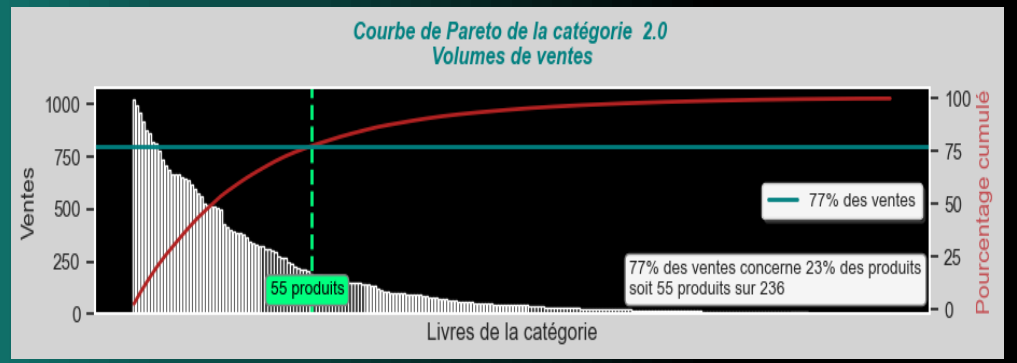



## Observations:

- **Pour toute catégorie, 77% des ventes concernent 23% des produits.**
- **Les catégories 0.0 et 1.0 représentent 76,6% du chiffre d'affaires.**
-



# Analyse des meilleures ventes par catégorie

## Démarche:

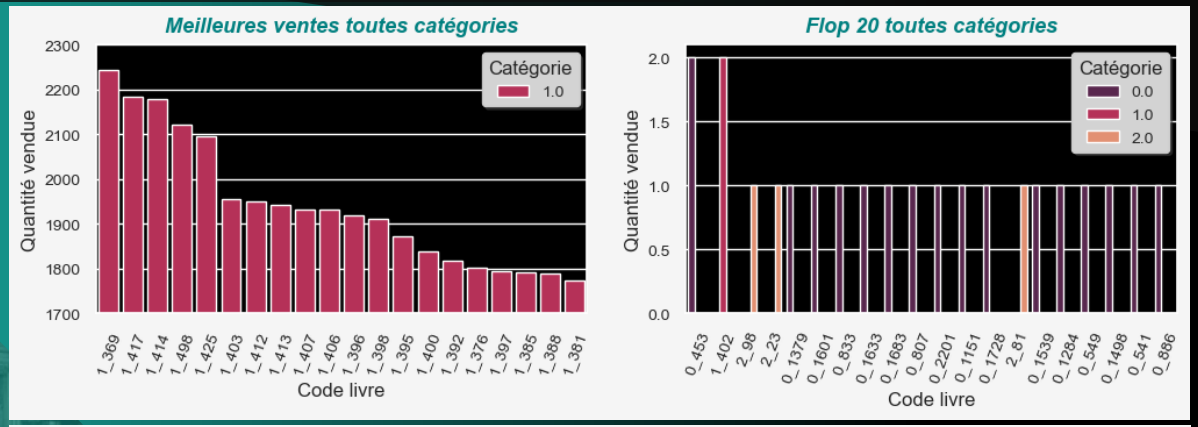
- Calcul du Top et Flop des ventes toutes catégories
- Visualisations basées sur les quantités vendues

- Calcul  Top chiffre d'affaires toutes catégories

## Observations:

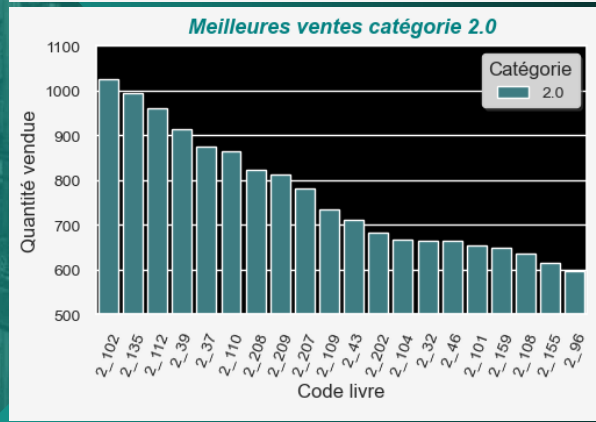
- En nombre d'exemplaires vendus, la catégorie 1.0 se place largement en tête.
- Les Flop concernent surtout la catégorie 0.0

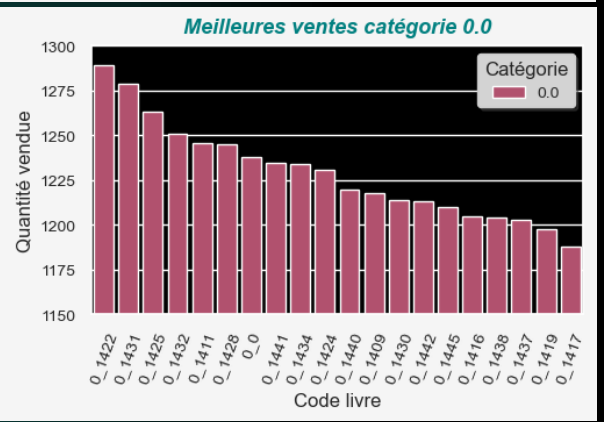
- Au niveau CA, la catégorie 2.0 se distingue et la catégorie 1.0 est présente.

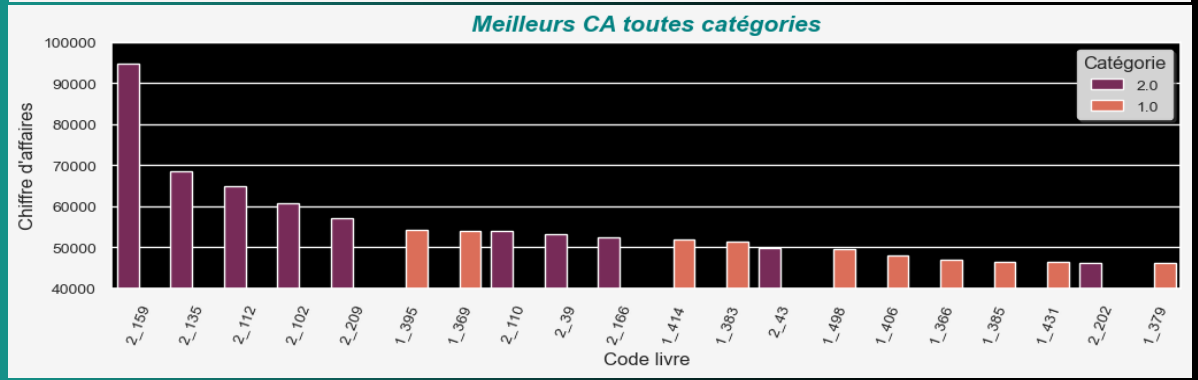








# Analyse de Saisonnalité – par jour et par mois

## Démarche:

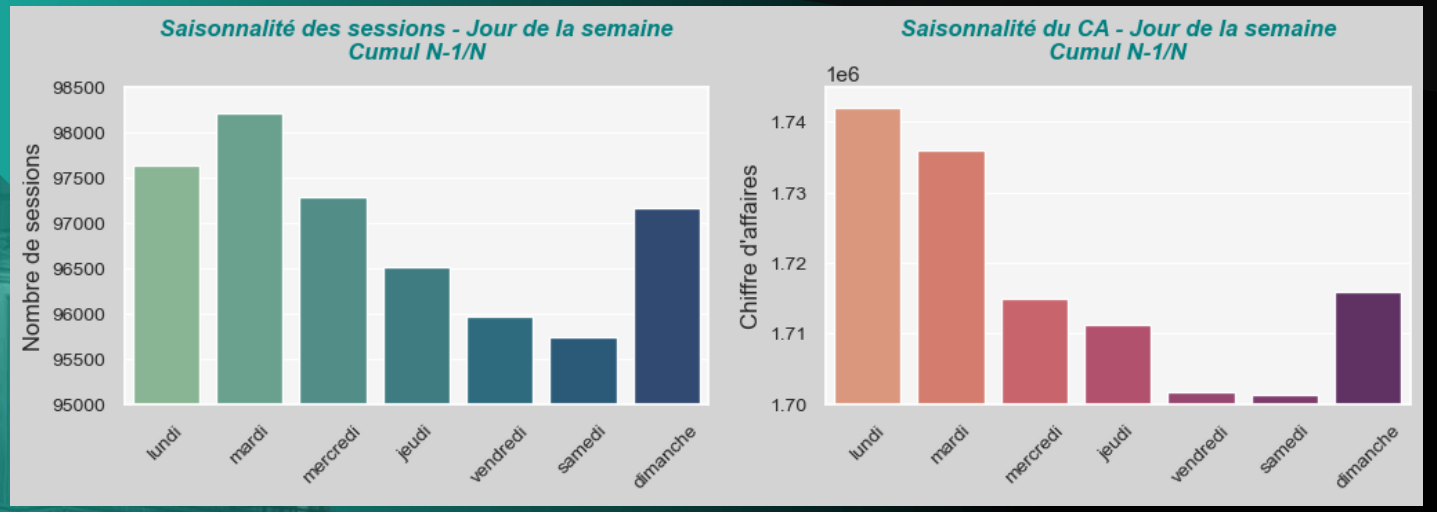
**Pour analyser la saisonnalité, je l'ai décomposé comme suit:**

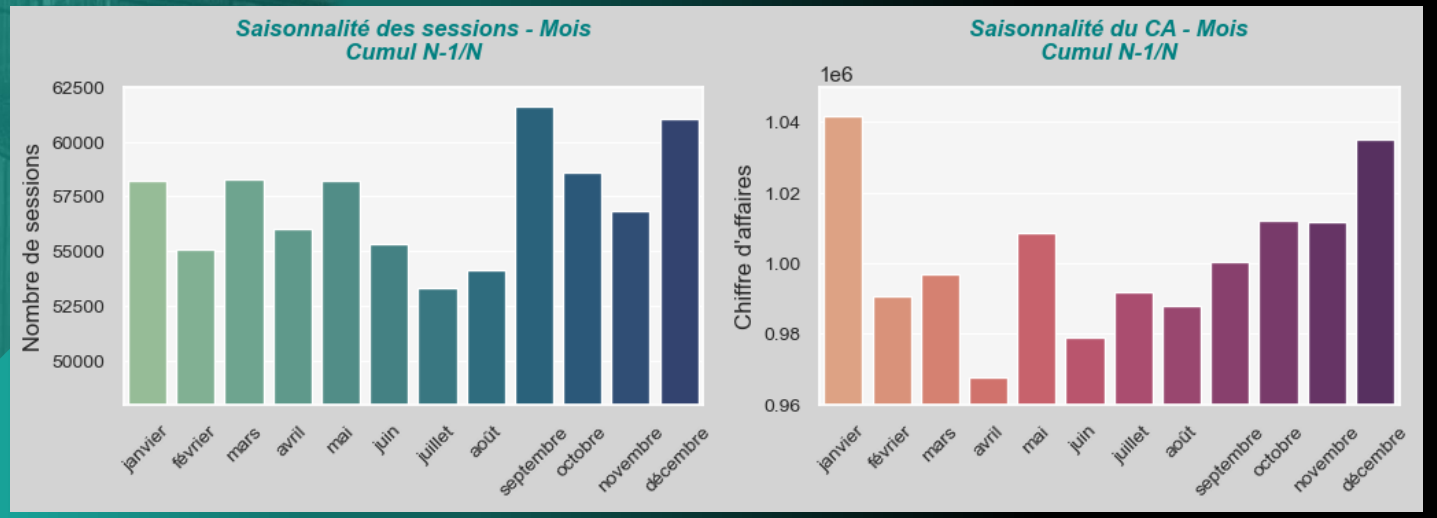
- **Par jour de la semaine**
- **Par mois**

## Observations:

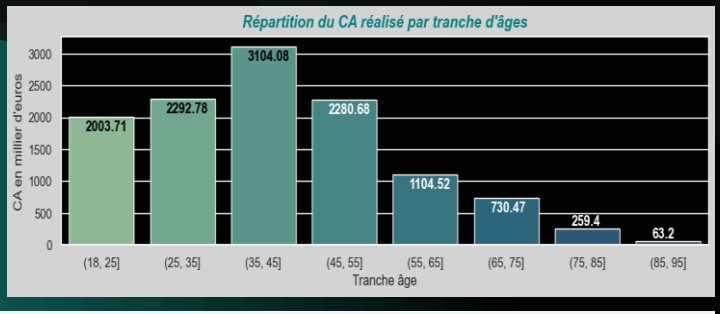
- **L'activité est plus importante les Lundis et Mardis.**
- **Les Dimanches et Mercredis sont aussi notables.**

- **En Septembre et Décembre, le nombre de sessions est plus important.**
- **Le Chiffre d'affaires est plus important en Décembre et en Janvier**



# Analyse du chiffre d'affaires par tranche d'âges

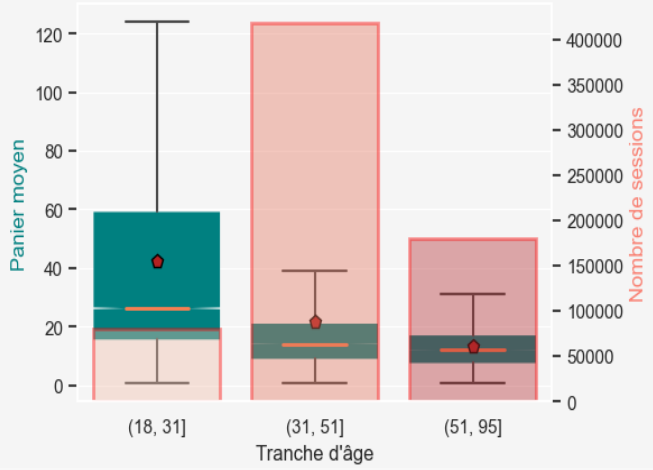



## Démarche:

**Pour obtenir une bonne lecture du comportement clients par tranche d'âges, nous visualisons le volume de ventes ainsi que le panier moyen.**

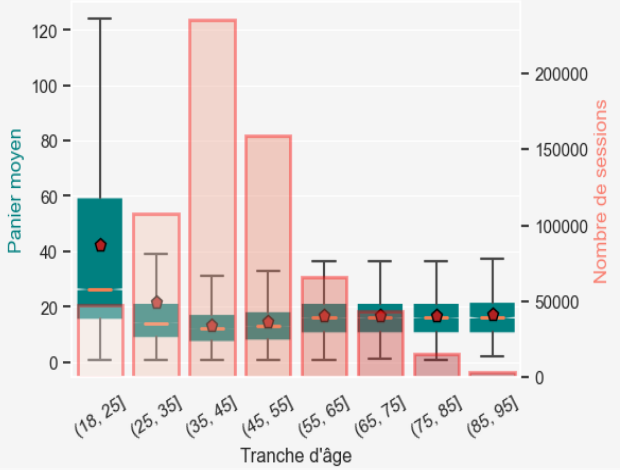
*Panier moyen et nombre d'achats par tranche d'âge*

| Tranche d'âges | Montant d'achat | Nbr de sessions |
|---|---|---|
| (18, 31] | 3287692.25 | 79705 |
| (31, 51] | 5531977.88 | 418693 |
| (51, 95] | 3019178.3 | 180107 |




*Tranches d'âges affinées*
*Panier moyen et nombre d'achats par tranche d'âge*

| Tranche d'âges | Montant d'achat | Nbr de sessions |
|---|---|---|
| (18, 25] | 2003711.54 | 47770 |
| (25, 35] | 2292783.25 | 107934 |
| (35, 45] | 3104079.78 | 235251 |
| (45, 55] | 2280684.49 | 158846 |
| (55, 65] | 1104519.42 | 65878 |
| (65, 75] | 730472.42 | 43609 |
| (75, 85] | 259397.52 | 15458 |
| (85, 95] | 63200.01 | 3759 |




## Observations:

- les jeunes commandent peu mais achètent des livres chers / collectors / reliés pour offrir par exemple. Nous observons une importante dispersion du montant d'achat.
- les 30-50 ans commandent le plus et achètent des livres correspondant à une consommation de livres de poche plus régulière
- Après 55 ans, les ventes diminuent mais le panier moyen reste stable
- On distingue clairement l'achat plaisir où la moyenne s'éloigne de la médiane et l'achat récurrent où médiane et moyenne sont proches

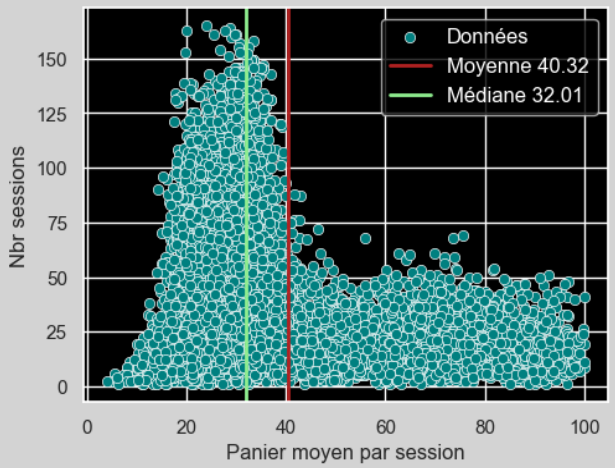
# Analyse panier moyen

## Paniers moyen – Types :

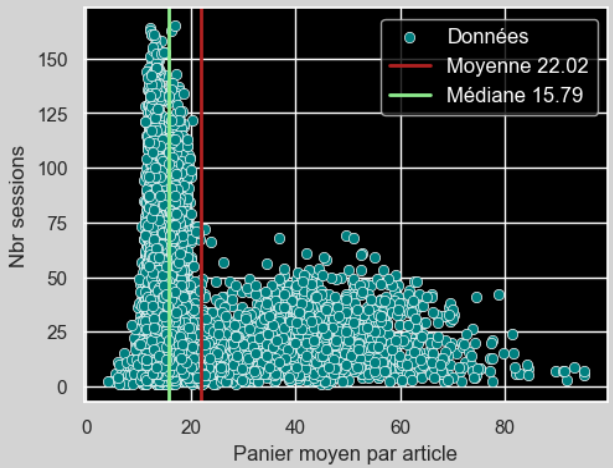
*Le panier moyen est de **34.58 euros** par session*

*Le panier moyen est de **40.32 euros** par client*

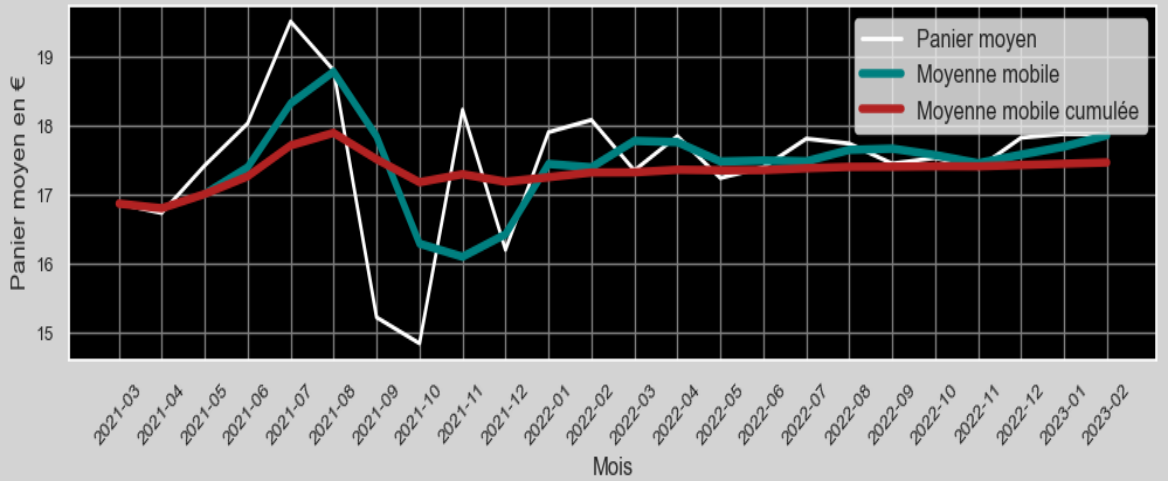
*Panier moyen client par session Hors Top 4 clients*




*Prix moyen par article Hors Top 4 clients*




*Evolution panier moyen (Sessions)*




## Observations:

- **L'évolution du panier moyen par session se stabilise dans le temps.**
- **L'essentiel du montant des sessions se situe entre 15 et 40 €**
- **L'essentiel des produits achetés ont un prix compris entre 10 et 20 €**

# Analyse turnover clients

## Chiffres clés:

*Nouveaux clients*

*8422 clients étaient inscrits au 31 Août 2021*
*178 clients se sont inscrits depuis Septembre 2021*

*Clients perdus*

*8294 clients ont commandé au cours des 6 derniers mois*
*306 clients considérés perdus : aucune commande depuis plus de 6 mois*

*Clients One Shot*

*29 clients n'ont commandé qu'une seule fois*

*Taux d'attrition*

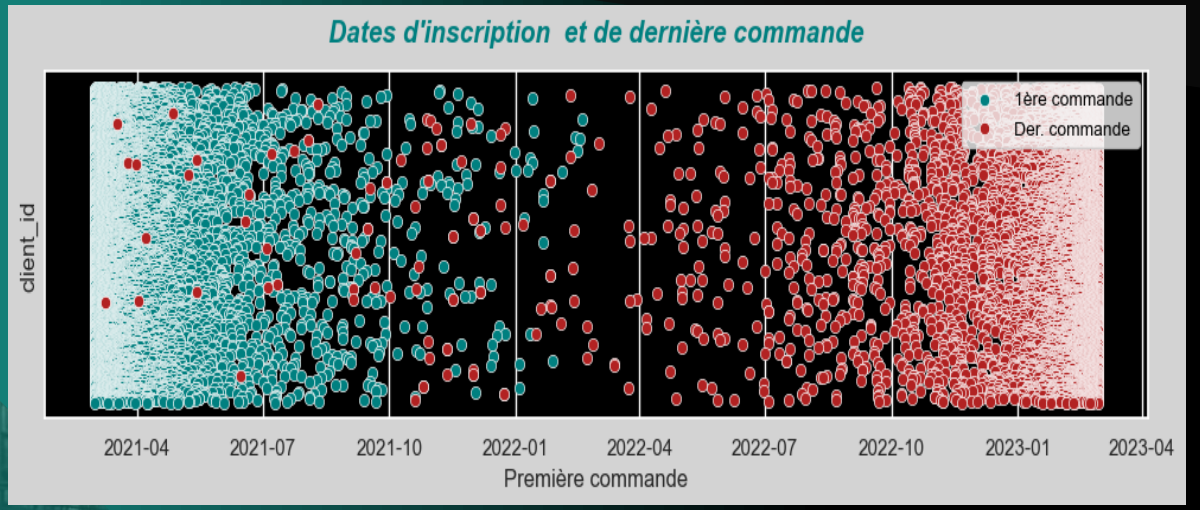
*Le taux d'attrition des clients perdus est de: 3.56%*

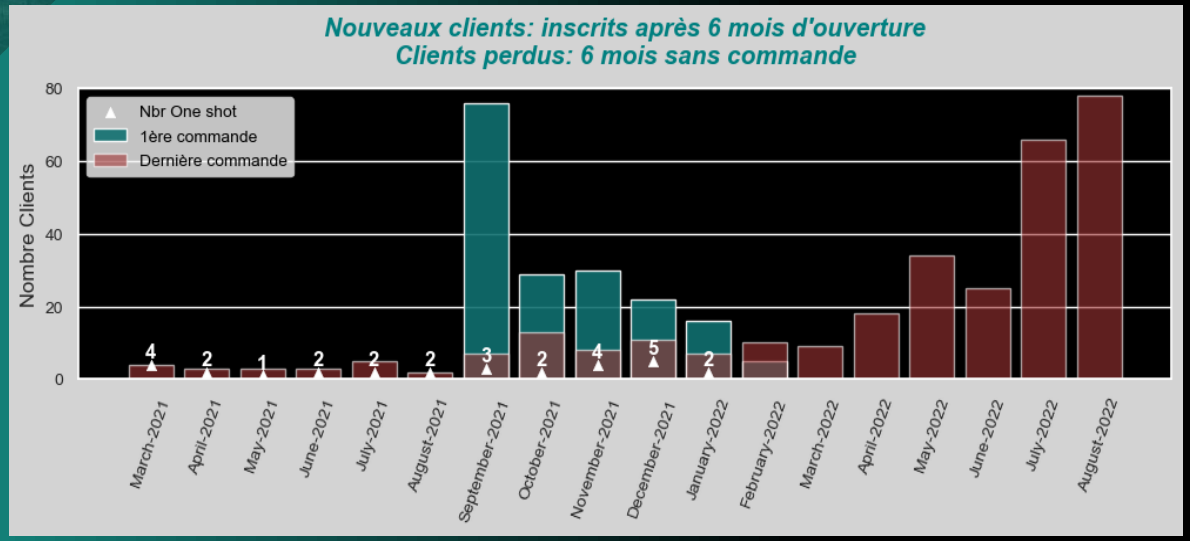
*Stabilité du portefeuille clients*

*Nombre d'inscription sur les deux premiers mois: 7144*
*Nombre de commandes uniques par client sur les 2 derniers mois: 7085*

## Observations:

- Il y a très peu de turnover et très peu de One Shot.
- Sur N (Mars22 - Feb23), il n'y a plus de nouveaux clients
- Avec 7144 inscriptions les deux premiers mois et 7085 dernières commandes les deux derniers mois, le portefeuille clients est statique et sans développement.



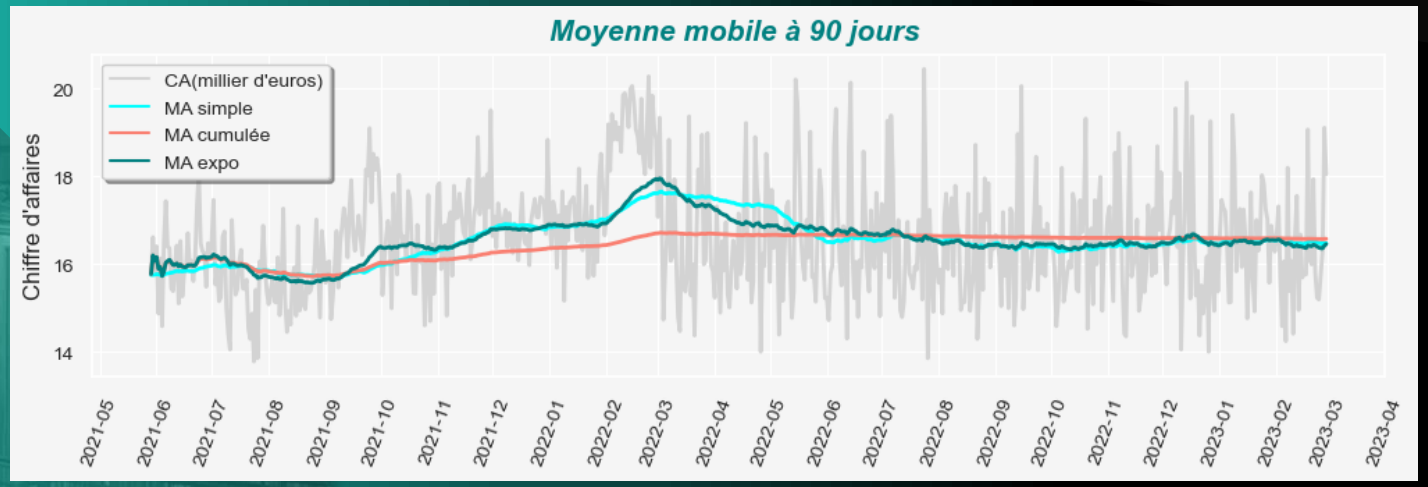
# Tendance globale - Moyenne mobile

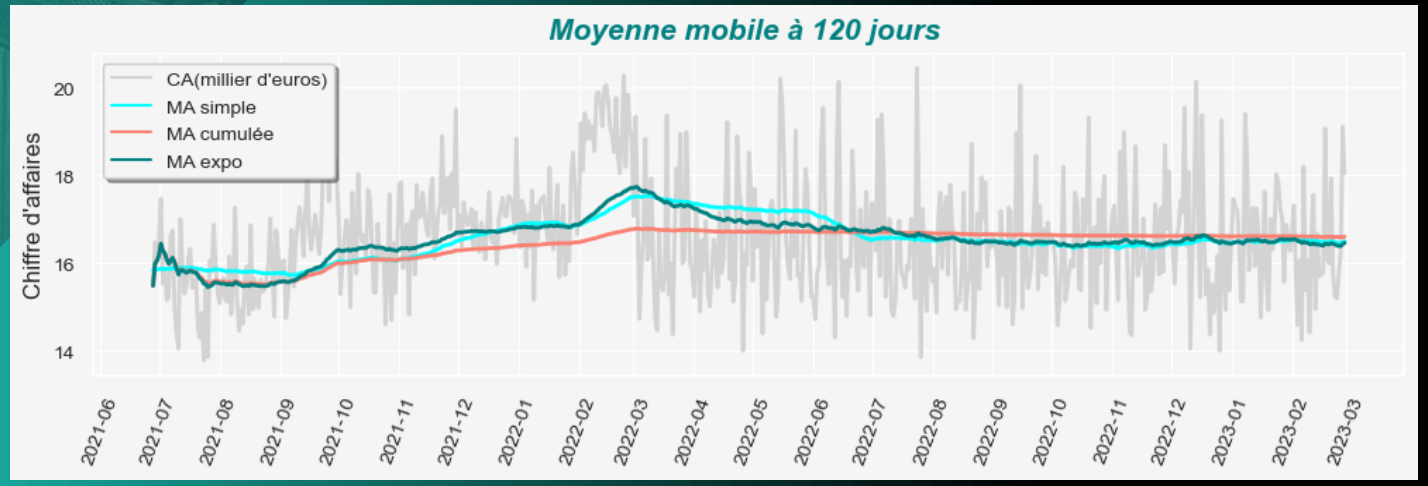
## Méthode :

- Tracé des moyennes simples, cumulées et exponentielles.
- Choix du temps de report à 90 et 120 jours

## Observations:

- Nous observons une croissance jusqu'en mars 2022.
- Ensuite, le chiffre d'affaires stagne en suivant une légère décroissance au fil du temps.



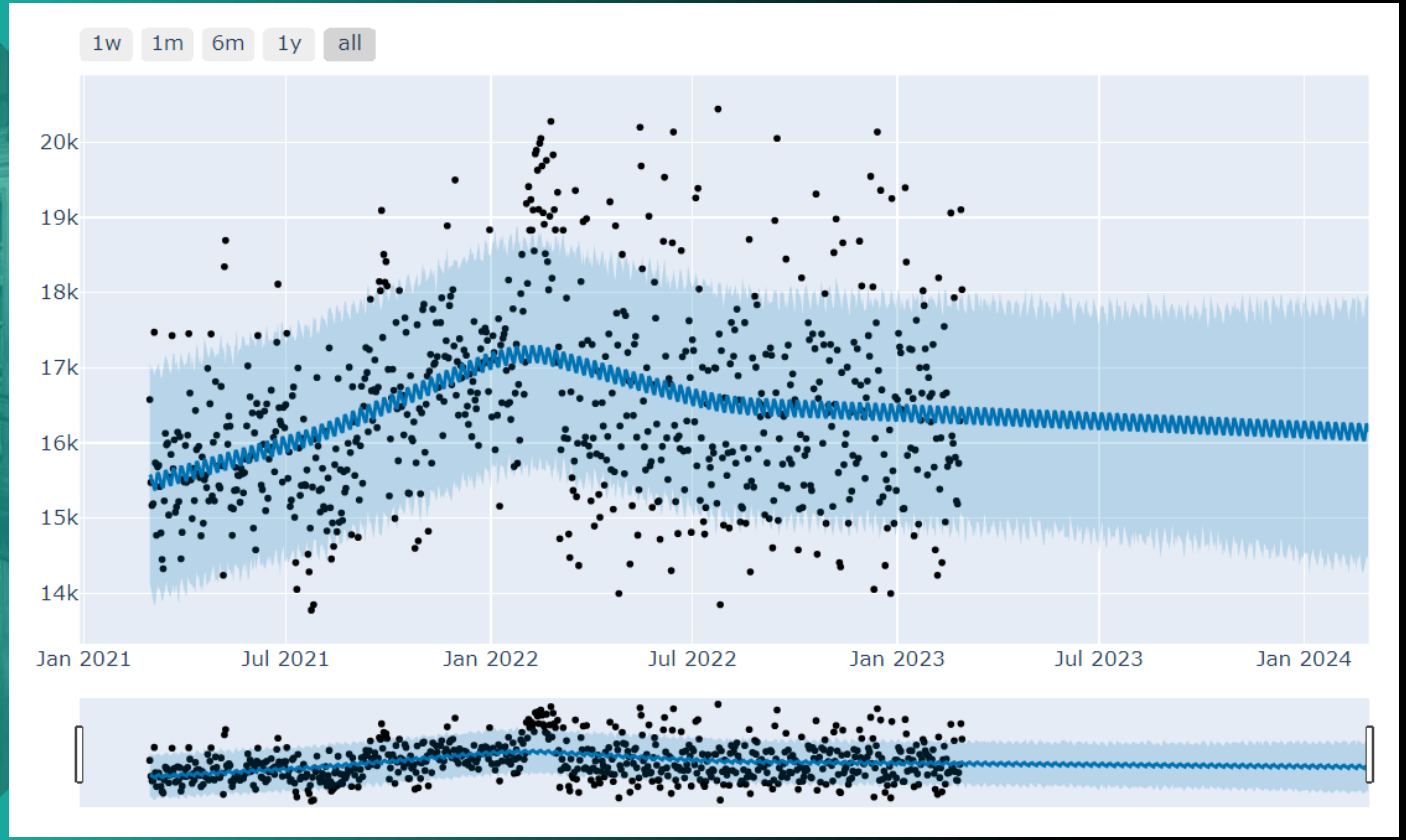
# Projection du chiffre d’affaires

## Méthode :

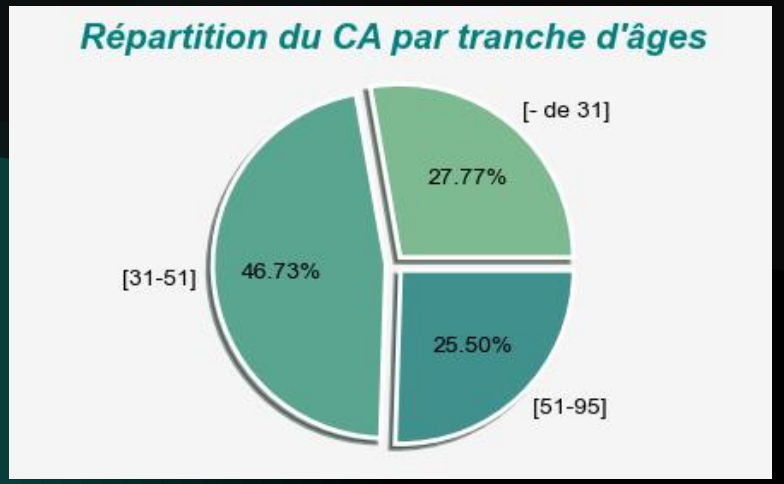
- Projection du CA jusqu’en février 2024
- Utilisation de la librairy ‘Prophet’

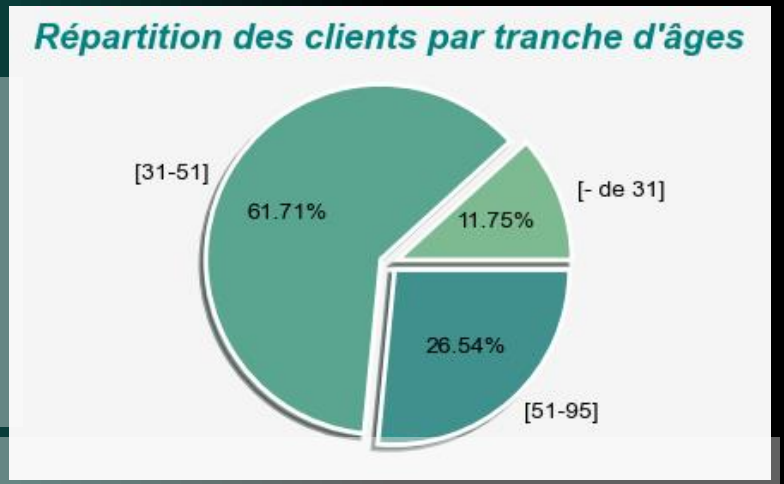
## Observations:

- Nous avons la confirmation d’une légère décroissance du chiffre d’affaires au fil du temps.



# Conclusion – choix orientation du site web




## Cœur de cible:

- Bien que représentant seulement 11,75% des clients, les - de 31 ans engendrent 27,77% du chiffre d'affaires.
- La tranche des 31-51 ans, comprenant plus de 60% de la clientèle, représente quasi 50% du chiffres d'affaires.
- Que ce soit concernant la proportion clients ou du chiffre d'affaires, les 3/4 proviennent des moins de 51 ans.

## Scénario 1 : Statu Quo ...

Le site web n'est qu'un outil mis à disposition des clients associés aux librairies physiques.
Chaque librairie donne un accès web à ses clients

- Dans cette configuration, l'inertie des clients existants suffit à justifier l'existence du site (7085 commandes ces 2 derniers mois pour 8600 clients) malgré une légère décroissance au fil du temps.
- A vérifier : pourquoi n'y a-t-il plus de nouveaux client depuis Mars 2022

## Scénario 2 : ou Dynamisme

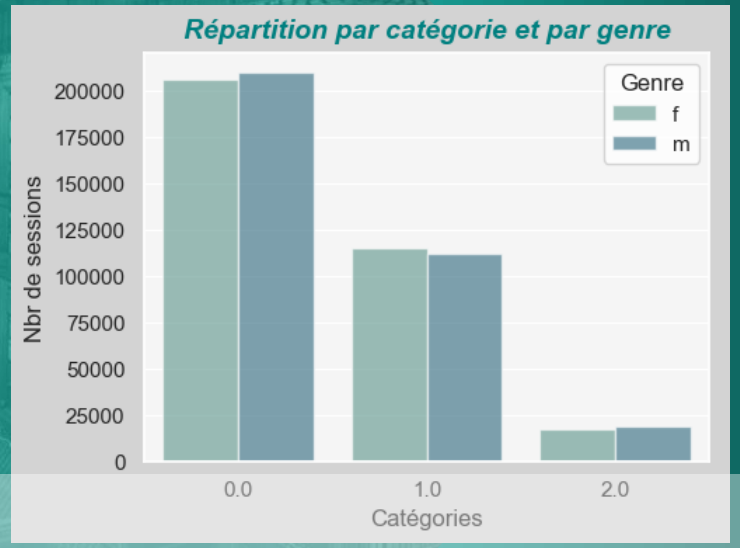
Si, en revanche, LAPAGE souhaite que le site soit un vecteur d'acquisition de nouveaux clients et de développement des clients existants, des actions commerciales et/ou marketing sont indispensables.

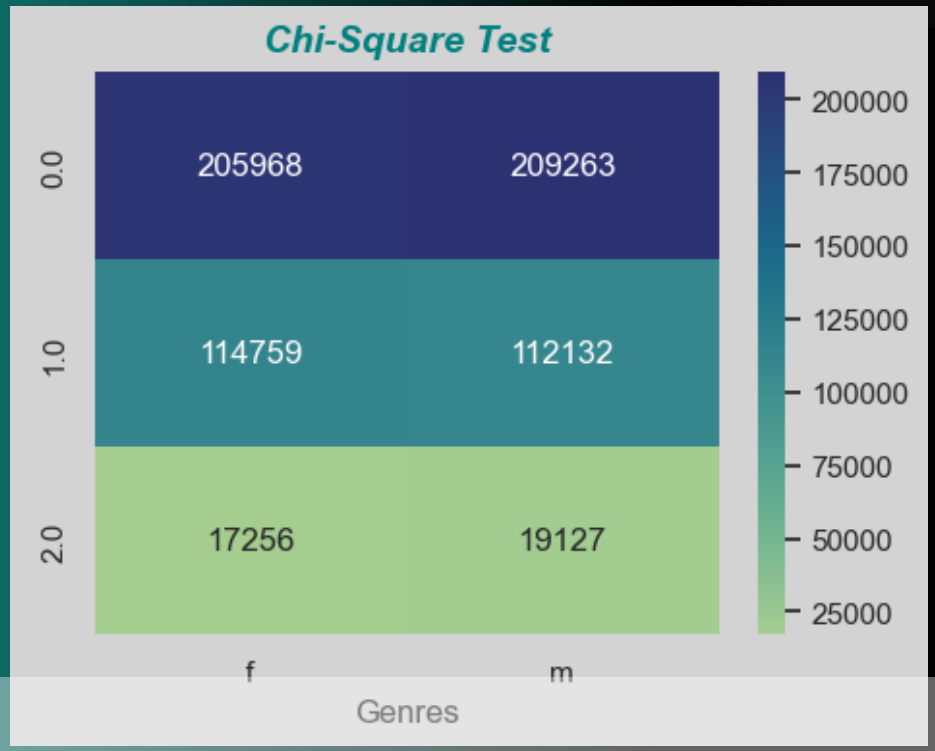
- Dynamiser les clients existants : Fidélisation (ex : par points), promotions (black Friday, soldes…) , proposition d'articles en fonction des précédentes commandes ….
- Acquisition clients: campagnes marketing ciblées, achats de fichiers clients (SMS, e-mails, adresses postales..), publicités médias….
- Recrutement d'un assistant marketing web pour faire « vivre » le site.

# 1 - Corrélation entre le genre client et la catégorie de livre acheté

## Test de Chi carré d'indépendance - Hypothèses :

**H0 (nulle): Les deux variables sont indépendantes**
**H1 (alternative) : Les deux variables sont corrélées**




## Résultat du Test Chi2 :

**Statistic : 143.28**
**pvalue : 7.71e-32**

## Observations :

**Avec le test du Chi2, nous observons une corrélation entre le genre des clients et la catégorie de livre achetée**

**les données fournissent des preuves solides pour rejeter l'hypothèse nulle. Il y a une association significative entre les variables étudiées, les fréquences observées diffèrent considérablement des fréquences attendues selon l'hypothèse nulle.**

**Les catégories 0.0 et 2.0 sont légèrement plus prisées par les hommes.**

**La catégorie 1.0 est légèrement plus prisée par les femmes.**

# 2 - Corrélation entre l'âge des clients et le montant total des achats

## Test d'ANOVA (Analyse de variance)- Hypothèses :

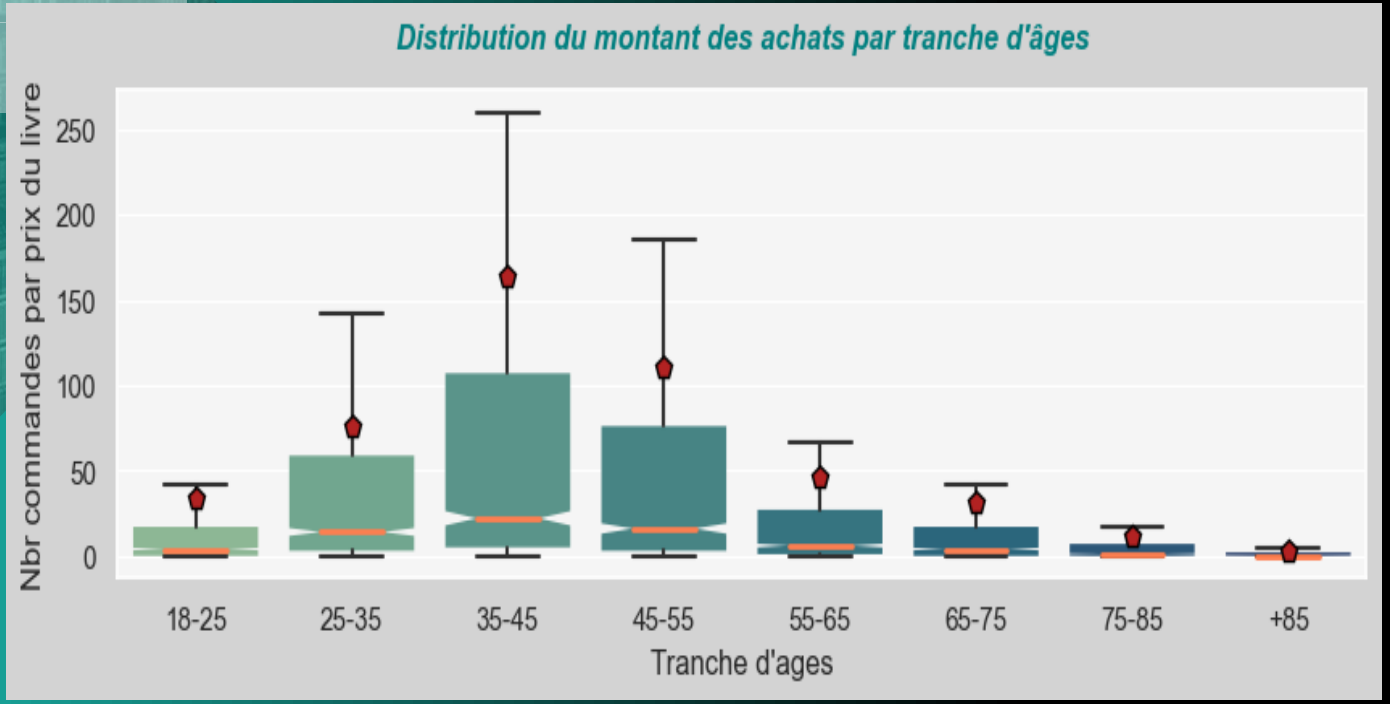
**H0 : Il n'y a aucune différence significative entre les moyennes des groupes.**

**H1 : Il y a une différence significative entre au moins deux moyennes de groupes**

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(tranchages) | 7.0 | 3.013e+07 | 4.3054e+06 | 44.258 | 3.516e-62 |
| Residual | 11528 | 1.121e+09 | 9.7279e+04 | NaN | |

## Observations :

**On rejette l'hypothèse nulle, il existe une différence significative entre les échantillons**



# 3 - Corrélation entre l'âge des clients et la fréquence d'achat

## Test d'ANOVA (Analyse de variance)- Hypothèses :

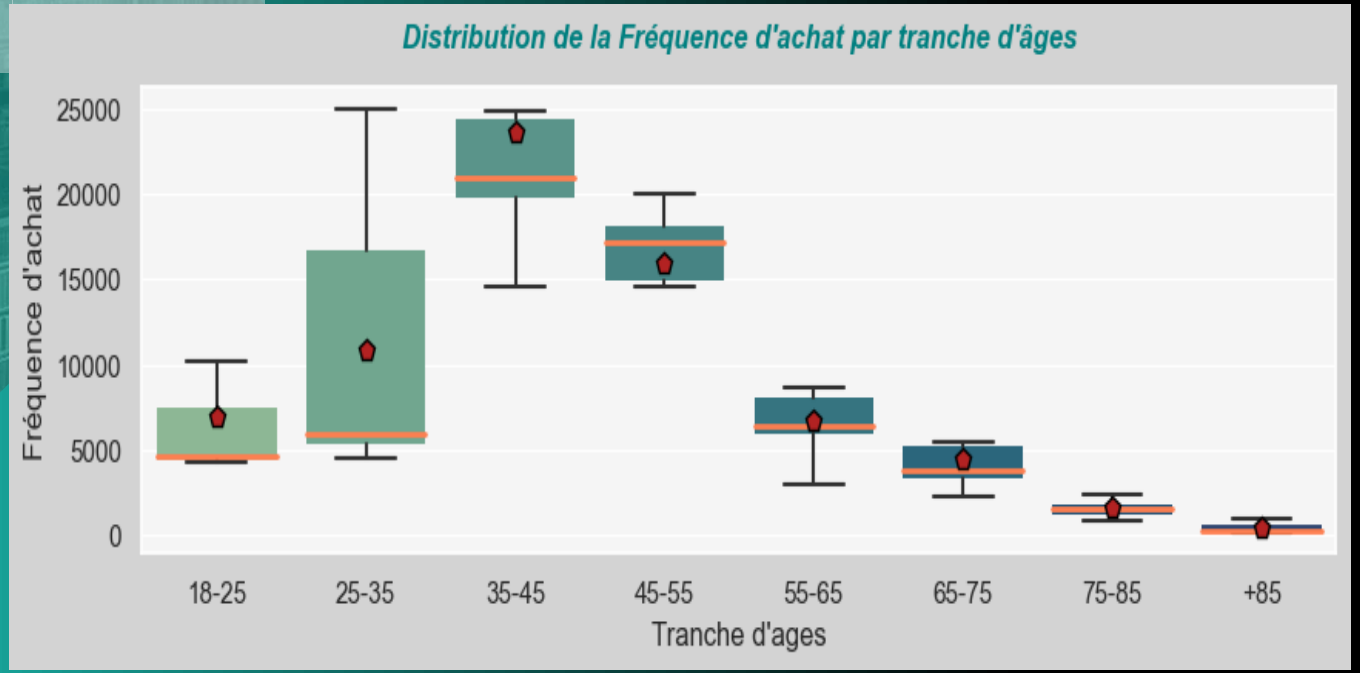
**H0 : Il n'y a aucune différence significative entre les moyennes des groupes.**

**H1 : Il y a une différence significative entre au moins deux moyennes de groupes (H1 : Au moins une paire de moyennes de groupes est différente).**

| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(tranchages) | 7.0 | 4.140e+09 | 5.915e+08 | 26.579 | 3.737e-17 |
| Residual | 68.0 | 1.513e+09 | 2.225e+07 | NaN | |

## Observations :

**On rejette l'hypothèse nulle, il existe une différence significative entre les échantillons**



# 4 - Corrélation entre l'âge des clients et la taille du panier moyen

## Test d'ANOVA (Analyse de variance)- Hypothèses :

**H0 : Il n'y a aucune différence significative entre les moyennes des groupes.**

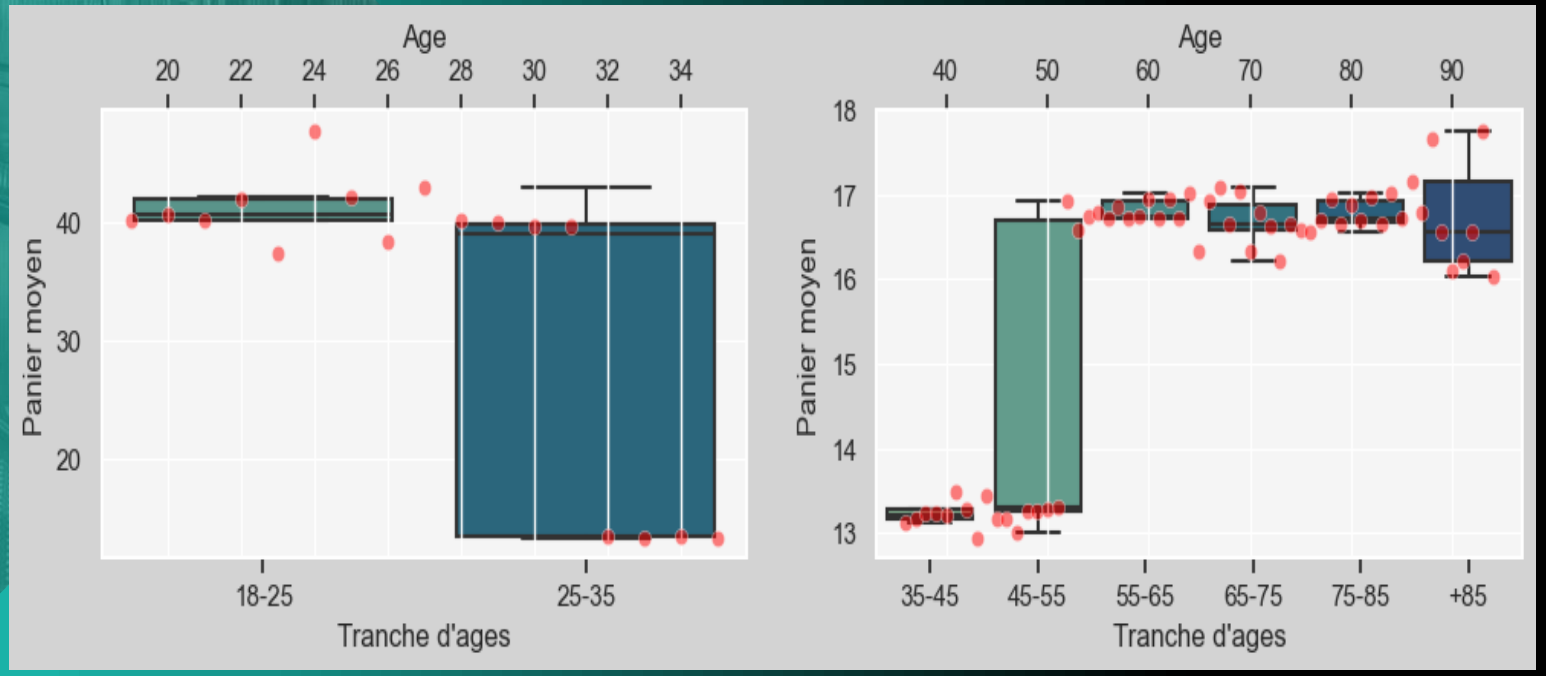
**H1 : Il y a une différence significative entre au moins deux moyennes de groupes (H1 : Au moins une paire de moyennes de groupes est différente).**

## Résultat de l'Anova:

**Statistique = 0.9896**
**p-valeur = 0.4377**

## Observations :

**on ne peut pas rejeter l'hypothèse nulle, il n'y a pas suffisamment de preuves pour affirmer qu'il y a des différences significatives entre l'âge et le montant du panier moyen**



# 4.1 - Corrélation entre l'âge des clients et la taille du panier moyen

## Test d'ANOVA  sur un échantillon modifié:

**H0 : Il n'y a aucune différence significative entre les moyennes des groupes.**

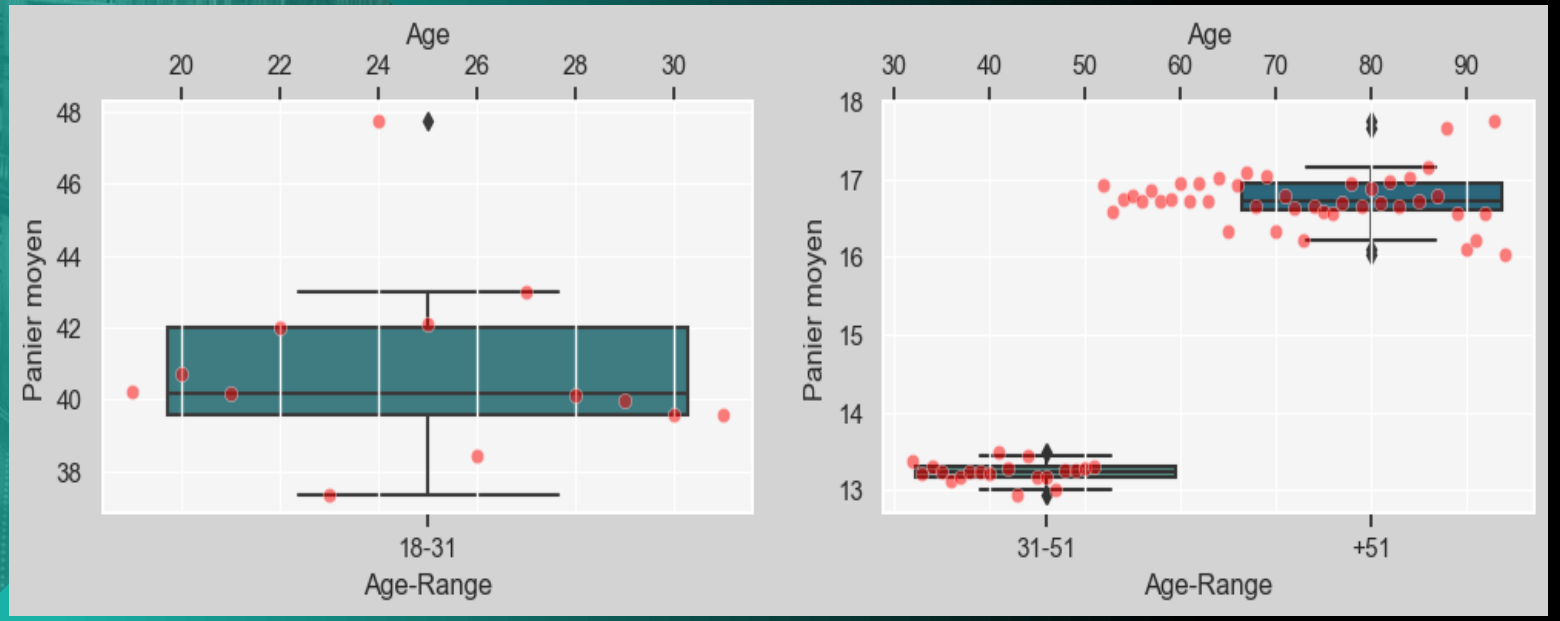
**H1 : Il y a une différence significative entre au moins deux moyennes de groupes (H1 : Au moins une paire de moyennes de groupes est différente).**

## Résultat de l'Anova:

**Statistique = 5.9963**
**p-valeur = 0.0029**

## Observations :

**On rejette l'hypothèse nulle, il y a des différences significatives entre l'âge et le montant du panier moyen.**



# 5 - Corrélation entre l'âge des clients et la catégorie des livres achetés

## Test de Kruskal-Wallis :

***H0 (hypothèse nulle) : Les médianes des groupes sont égales, il n'y a pas de différence statistiquement significative entre les groupes.***

***H1 (hypothèse alternative) : Au moins une médiane de groupe est différente, il y a une différence statistiquement significative entre au moins deux groupes.***

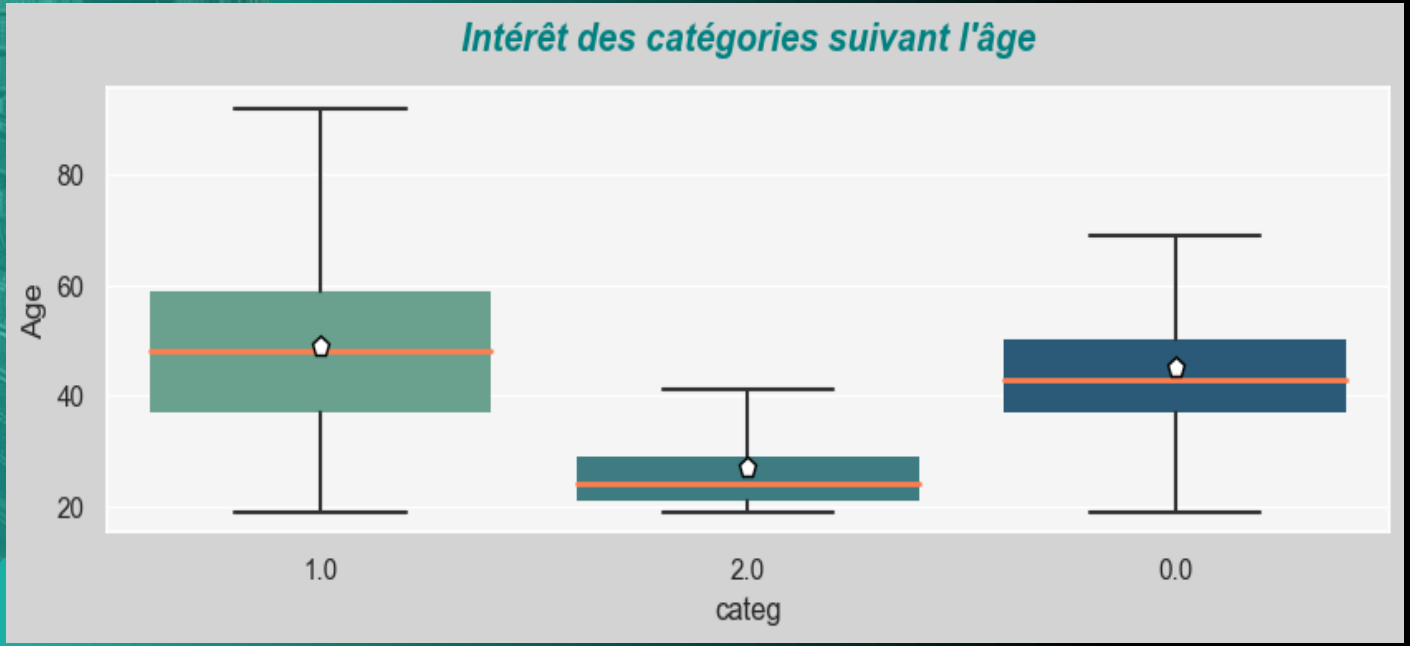
## Résultat du test de Kruskal-Wallis :

***Statistique de test = 79095.0292***

***p-valeur = 0.0***

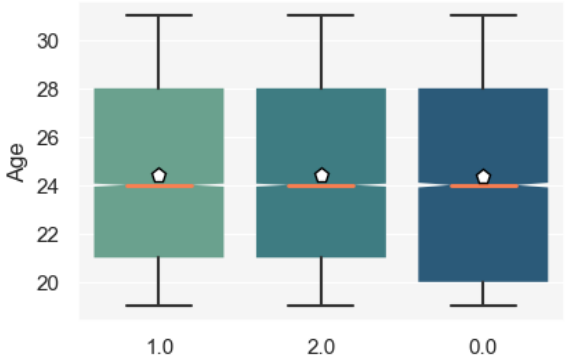
## Observations :

***Au moins une catégorie a une distribution médiane différente (hypothèse nulle rejetée)***



# 5.1 - Corrélation entre l'âge des clients et la catégorie des livres achetés

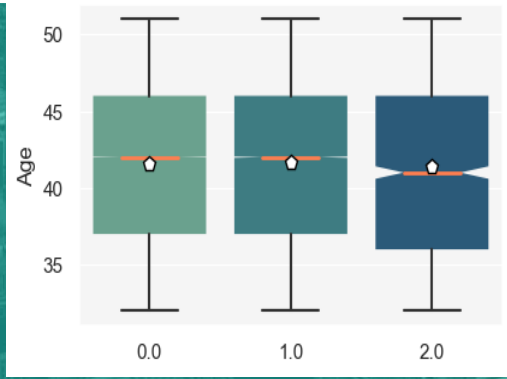
***Intérêt des catégories pour les 18-31 ans***




Test de Kruskal-Wallis - Tranche 18-31 ans
Statistique de test = 0.9720943975532833
p-valeur = 0.6150527768897947
Les catégories ont des distributions médianes égales (hypothèse nulle acceptée).

Test ANOVA (Tranche 18-31 ans):
Statistique de test = 0.33303562913768114
p-valeur = 0.7167456541160713
Les catégories ont des moyennes égales (hypothèse nulle acceptée).

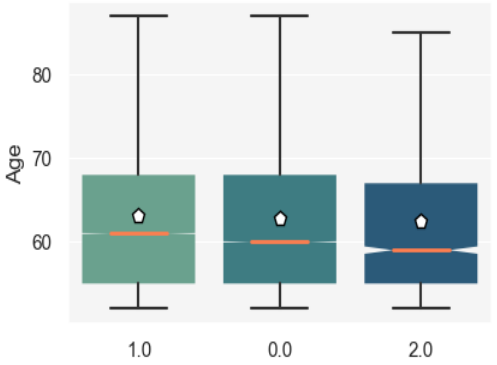
***Intérêt des catégories pour les 31-51 ans***




Test de Kruskal-Wallis - Tranche 31-51 ans
Statistique de test = 10.84757408724901
p-valeur = 0.004410412556710242
Au moins une catégorie a une distribution médiane différente (hypothèse nulle rejetée).

Test ANOVA - Tranche 31-51 ans
Statistique de test = 5.633582835630971
p-valeur = 0.003576012108452224
Il existe une différence significative entre les moyennes des catégories (hypothèse nulle rejetée).

***Intérêt des catégories pour les +51 ans***




Test de Kruskal-Wallis - Tranche +51 ans
Statistique de test = 26.6874920637383
p-valeur = 1.6028199170218643e-06
Au moins une catégorie a une distribution médiane différente (hypothèse nulle rejetée).

Test ANOVA - Tranche +51 ans
Statistique de test = 13.78469223756778
p-valeur = 1.0323866494416385e-06
Il existe une différence significative entre les moyennes des catégories (hypothèse nulle rejetée).