

Analysez les ventes de votre entreprise

Projet 4 – Formation DATA Analyst
d'OpenClassRooms



Sommaire

Nettoyage de données

- Présentations des données
- Analyse des différents DataFrames et des solutions apportées

Analyse des données

- Évolution du chiffre d'affaires
- Prix pratiqués
- Secteur d'activités
- Analyse des ventes
- Typologie clients
- Fidélisation de nos clients

Les corrélations

- Y a-t-il une corrélation entre l'âge des clients et :
 - Le montant total des achats
 - La fréquence d'achat
 - La taille du panier moyen (en nombre d'articles)
 - Les catégories de produits achetés
- Y a-t-il une corrélation entre le sexe des clients et les catégories de produits achetés ?



Nettoyage des données

- *Présentation des données*

Jeu de données composé de 3 dataframes :

Df_customers : client_id, sex, birth

Df_products : id_prod, price, categ

Df_transactions : id_prod, date, session_id, client_id



• Analyse des différents DataFrames et des solutions apportées

Anomalies détectées

- *Détection de valeurs manquantes pour le produit '0_2245'*
- *Détection de valeurs "Test" dans les 3 DataFrames*
- *Anomalie durant le mois d'Octobre 2021*

Solutions apportées

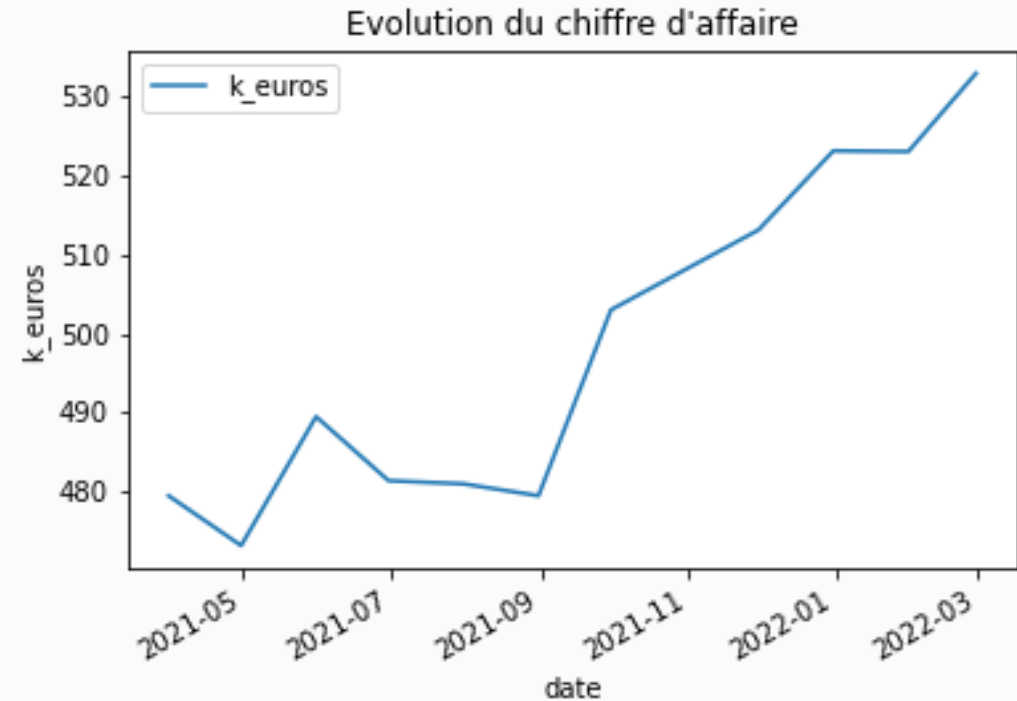
- Ajout de la moyenne des valeurs
- Ajout de la catégorie du produit
- Mise à l'écart des valeurs "Test" dans nos DataFrames durant l'analyse
- Suppression du mois d'Octobre durant notre analyse

The background is a dark blue gradient featuring several abstract data visualization elements. There are multiple overlapping line graphs with white and grey lines and circular markers. Some markers are colored in shades of pink, purple, and blue. A prominent orange bar chart is visible on the left side. On the right, a line graph has a peak with the value '245.57' labeled next to it. In the center, there are some faint, partially obscured numbers like '154.178' and '102'.

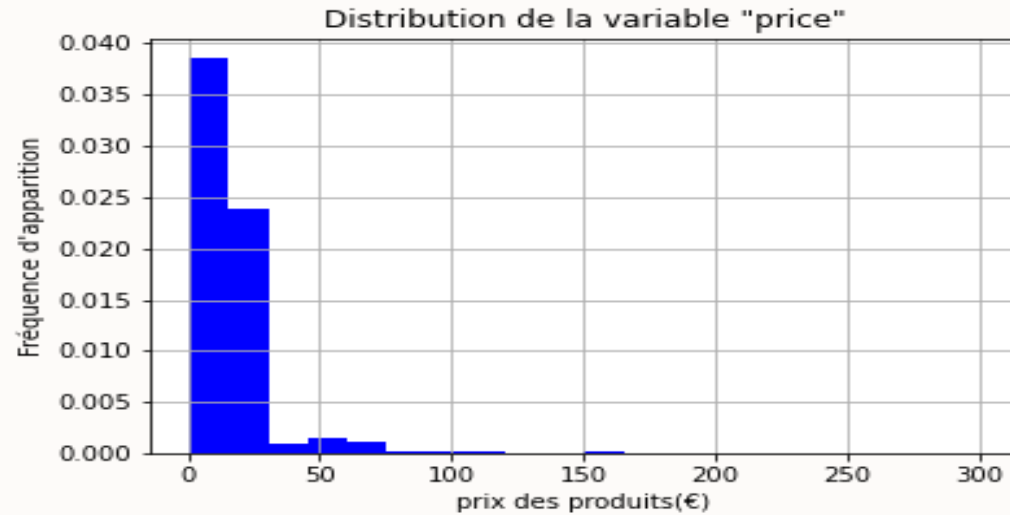
Analyse des données

Évolution du chiffre d'affaire

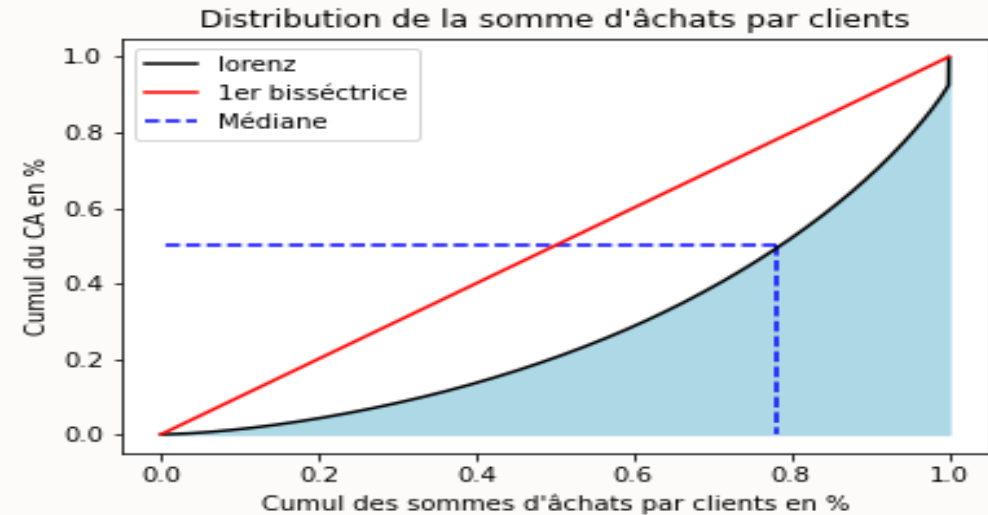
- Un chiffre d'affaire en hausse
- 11.2% de CA en plus par rapport à mars 2021
- Taux de croissance moyen de 1.02% par mois



Analyse des prix pratiqués



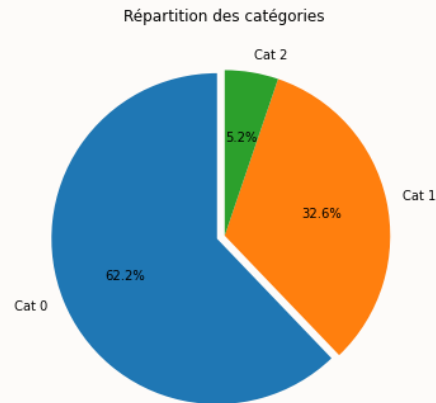
- Plus de 60% de nos transactions ont un article inférieur à 40€



- 78% de nos clients représente 50% du total des prix de ventes(CA)
- Les 4 plus gros acheteurs de notre site ont achetés des produits pour une somme total de 408 681€
- % des 5 plus gros acheteurs

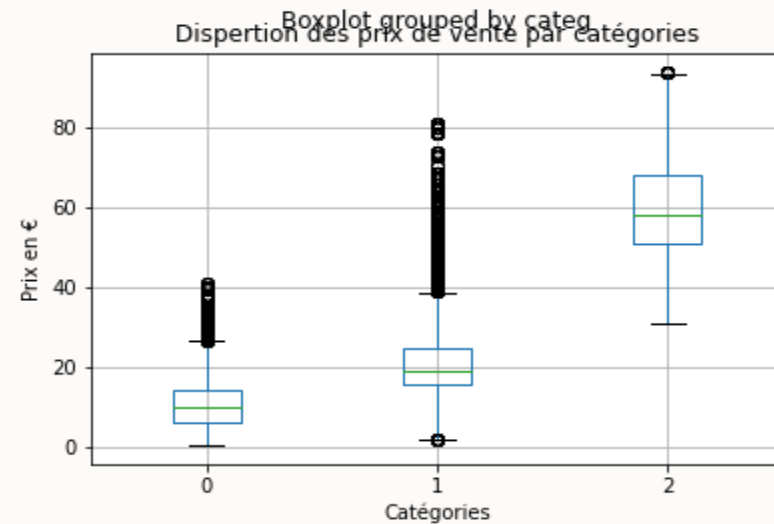
Secteurs d'activités

Analyse sectorielle de l'activité



- La catégorie '0' est la plus représentée avec plus de 60% de nos livres vendus

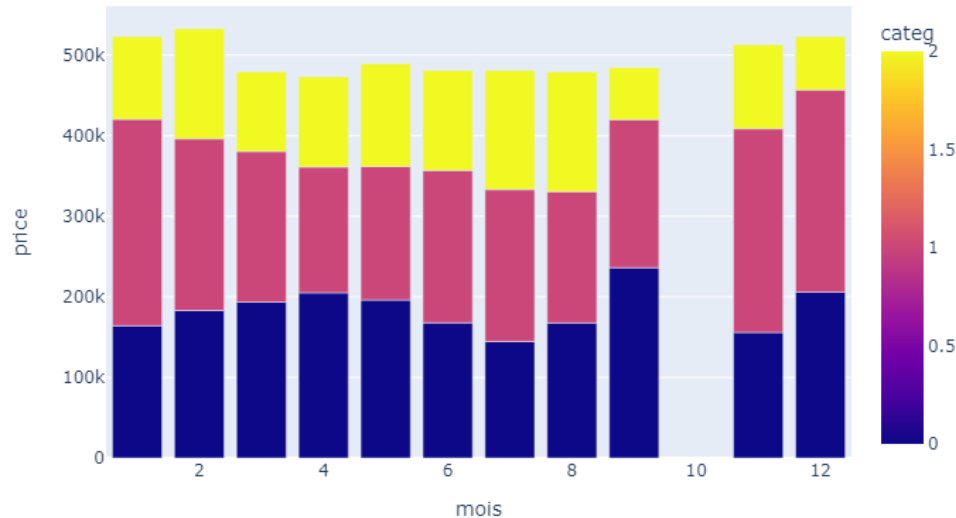
Comparaison des 3 catégories selon leurs prix



- Un mode & une médiane complètement différents
- Il y a une corrélation entre le prix et la catégorie

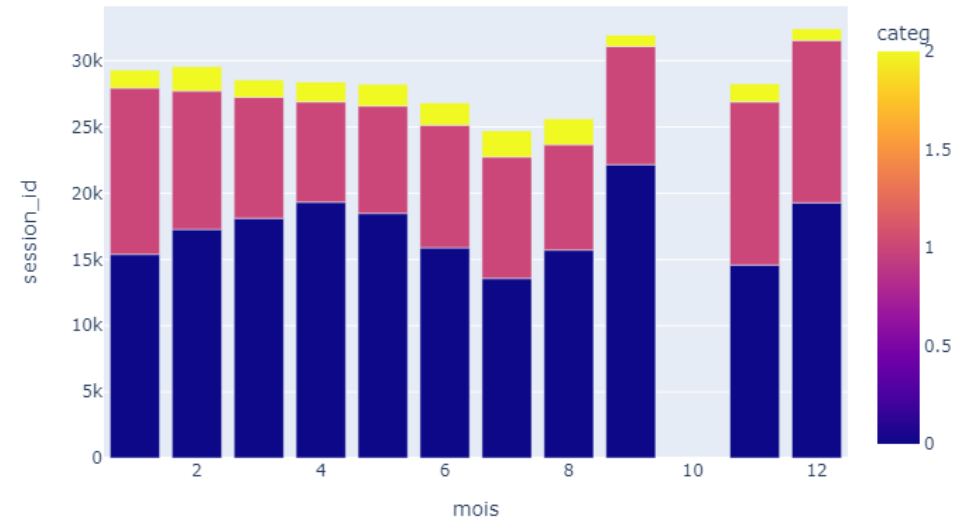
Analyse des ventes

Ventes par mois et par catégories



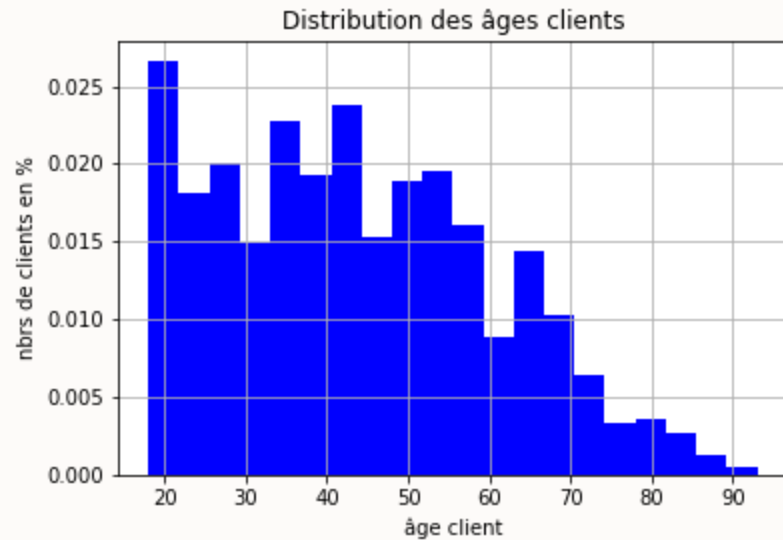
- Un CA proche de 500K durant l'année
- Prix de ventes par catégories assez équitable entre les catégories

Nbrs de livres vendus par catégorie et par mois

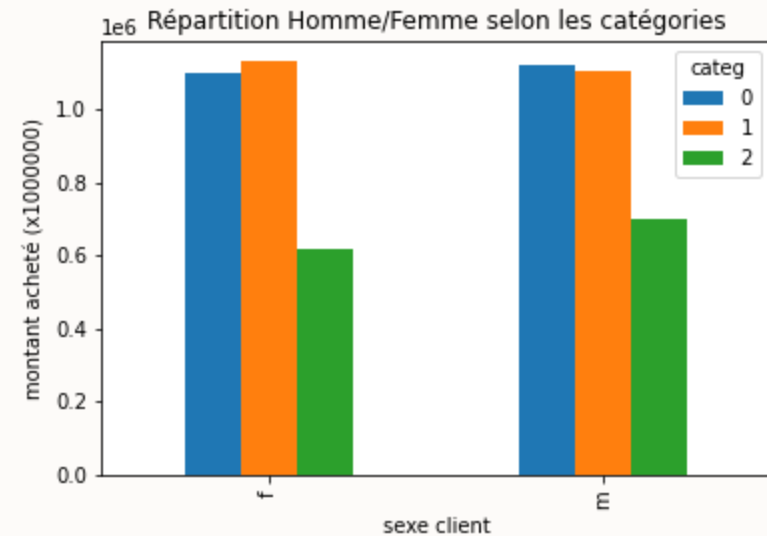


- Des pics de ventes pour le mois de Septembre(rentree) et le mois de decembre(fetes de fin d'annee)
- Nbrs de ventes qui décroît par catégories(0 > 1 > 2)

Typologie clients



- Une clientèle plutôt jeunes
- Une tendance qui décroît



- Répartition équitable entre les Hommes et les Femmes
- Popularité quasi égale de nos produits entre les deux genres
- Autant d'Homme que de Femme dans notre base clients

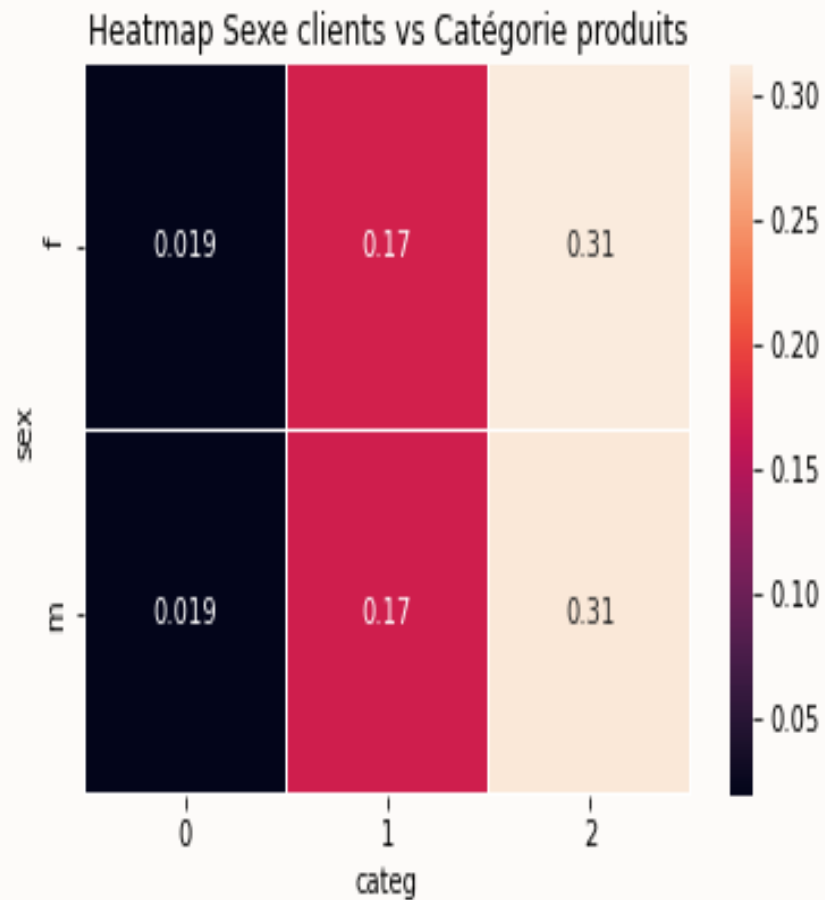
Fidélisation de nos clients

Produits achetés	Nombres de clients sur 8598 au total	Pourcentages
1	46	0.5 %
> 1	8552	99.4%
5	7964	92.6 %
10	6948	80.8 %

- Uniquement 0.5% de notre clientèle n'a acheté 1 seul article
- 99.4 % de notre clientèle a fait plus d'un achat sur notre site
- Plus de 80% ont effectué 10 achats sur notre site cette année

Les Corrélations

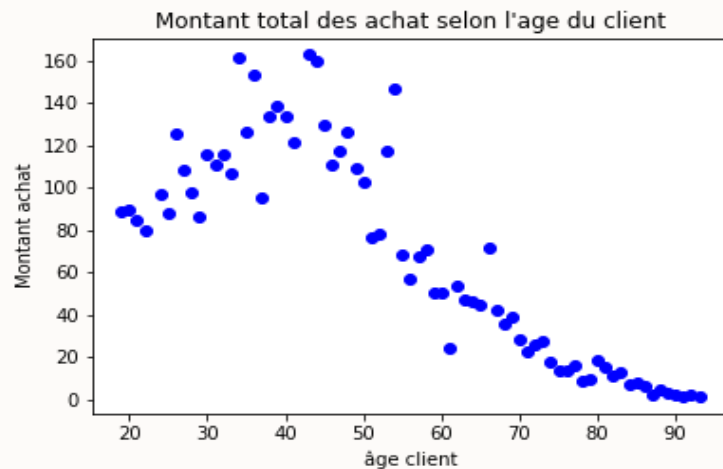
Corrélations sexe vs catégorie



- Aucune corrélation notable entre ces deux variables quantitatives selon le test du khi-2 (khi_n = 0.79)

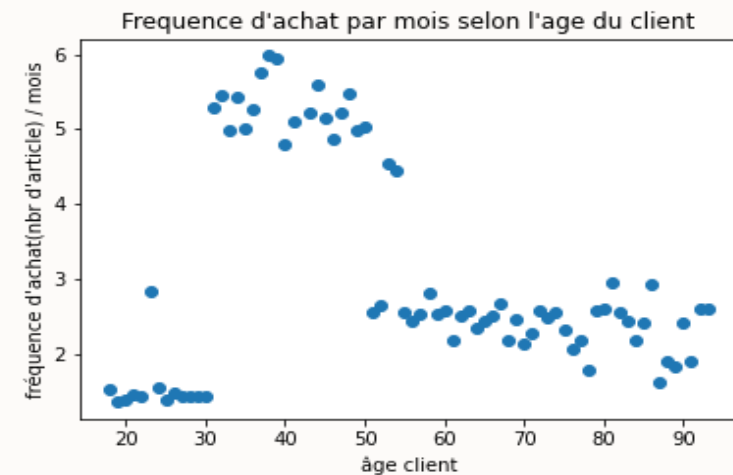
Y a-t-il une corrélation entre l'âge des clients et :

Le montant total des achats



- Corrélation linéaire négative : de -0.77
- 2 tendances distincte

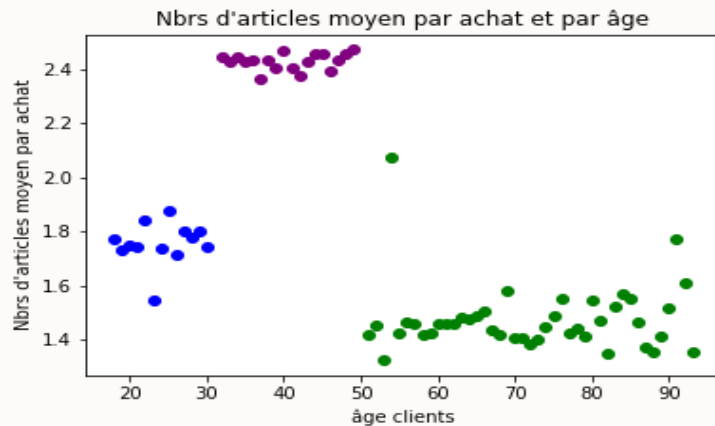
La fréquences d'achat par mois



- Aucune corrélation linéaire distincte : de -0.22
- 3 différents groupes visible
- Fréquence d'achat important entre 30 & 50ans

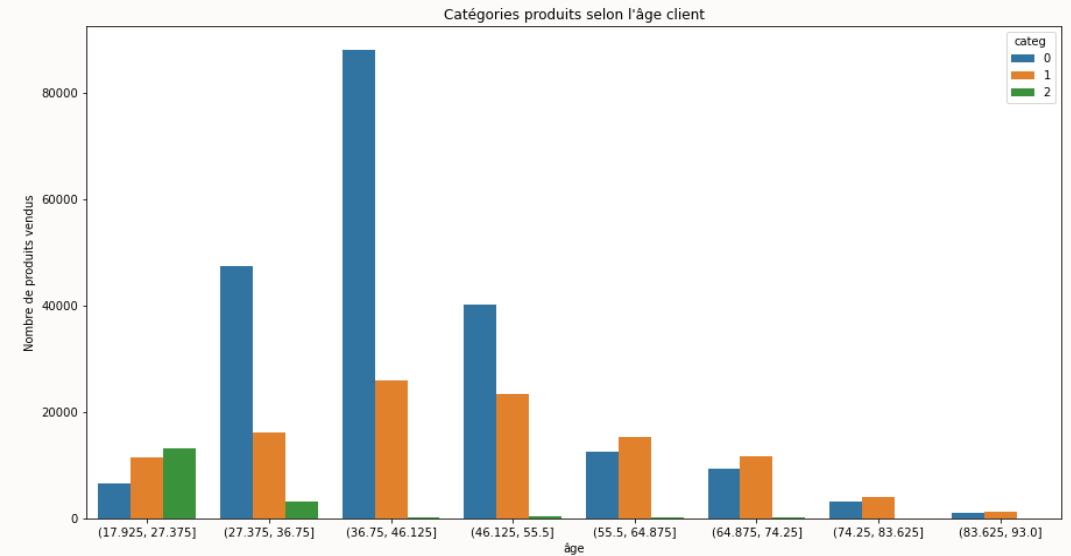
Y a-t-il une corrélation entre l'âge des clients et :

La taille du panier moyen (en nombre d'articles)



- Aucune Corrélation Linéaire :
- 18-30ans = 0.15
- 30-50ans = 0.19
- 50-90ans = -0.0016
- 3 groupes d'âge distinctes

Les catégories de produits achetés



- 3 différents groupes visible
- Catégories apprécié selon la tranche d'âge
- Il y a bien une corrélation entre ces deux variables : 0.36 (éta carré)



Conclusion

Chiffre d'affaires
stable et en
augmentation

Clientèle jeune et
quasi égale selon le
genre

Clientèle fidèle

3 groupes distinct
observable(18-30,
30-50, 50-90)

Données incomplète

Concentration de
nos efforts sur le
groupe de personnes
ayant 30-50 ans

Récompenser nos
clients fidèles



Remarques

- Attention à la syntaxe : Aucune corrélation et Aucune corrélation 'Linéaire' est différents
- Au lieu de 'supprimer' les données on peut en faire une simulation du CA
- Pour vérifier et présenter une ANOVA le graphique de type boxplot est le plus approprié pour représenter la distribution