

Analyse des ventes “RESTER LIVRE”

*Projet 4 - Nalron Octobre 2019
OpenClassrooms ENSAE - ENSAI
Formation Continue*

RESTER LIVRE en quelques mots :

- Spécialiste de la vente de livres
- Plusieurs magasins
- Boutique en ligne
- Algorithmes de recommandation
- Fort développement...

RESTER LIVRE en quelques chiffres sur l'A. fiscale 2021 :

- CA +5 millions d'euros
- Commandes +31 000 transactions
- Produits vendus +3200 unités

Données* nécessaires à l'analyse des ventes :

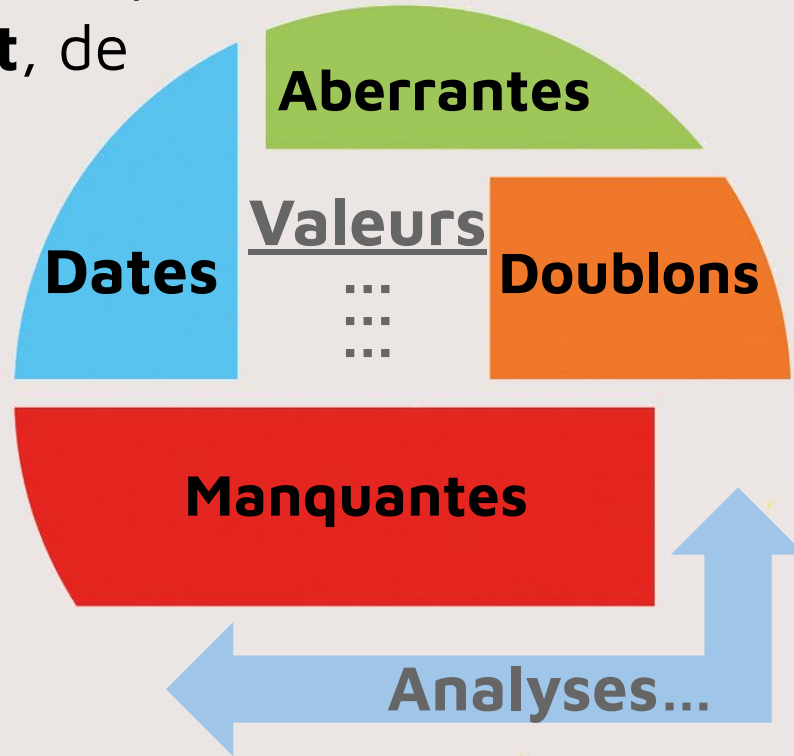
- Liste des **transactions**
- Liste des **produits**
- Liste des **clients**

**extraites de la base de données sous format .csv*

Questions principalement traitées au cours de l'analyse :

- Evolution des ventes, prix pratiqués
- Répartition démographique, sectorielle
- Ressemblance entre les individus
- Lien possible entre les caractéristiques clients/produits

Cycle de **traitement**,
de **retraitement**, de
nettoyage des
données.



Valeurs aberrantes : des valeurs "test" identifiées.

→ Suppression des lignes en correspondance aux tests.

Valeurs manquantes : transactions produit id 0_22450 et transactions catégorie 1 en octobre.

→ Intégration du prix moyen et intégration de la catégorie.

→ Suppression du mois d'octobre, analyse faite sur 11 mois.

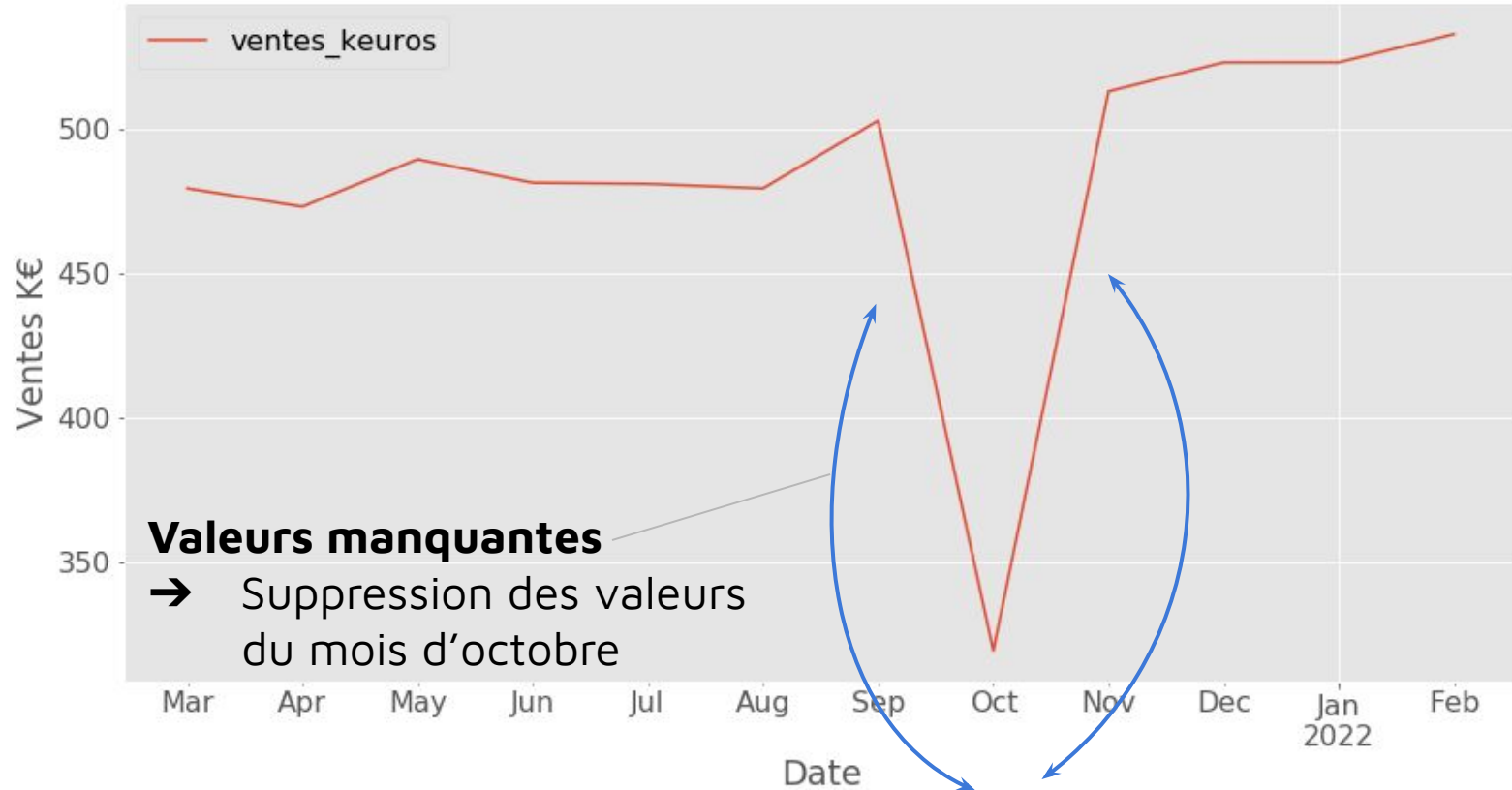
Valeurs doublons : aucune valeur en double.

→ Individus représentés par un identifiant unique.

Dates : conversion des valeurs dans un format date.

Autres : cohérence des prix, des âges, des sexes, etc.

Evolution des ventes

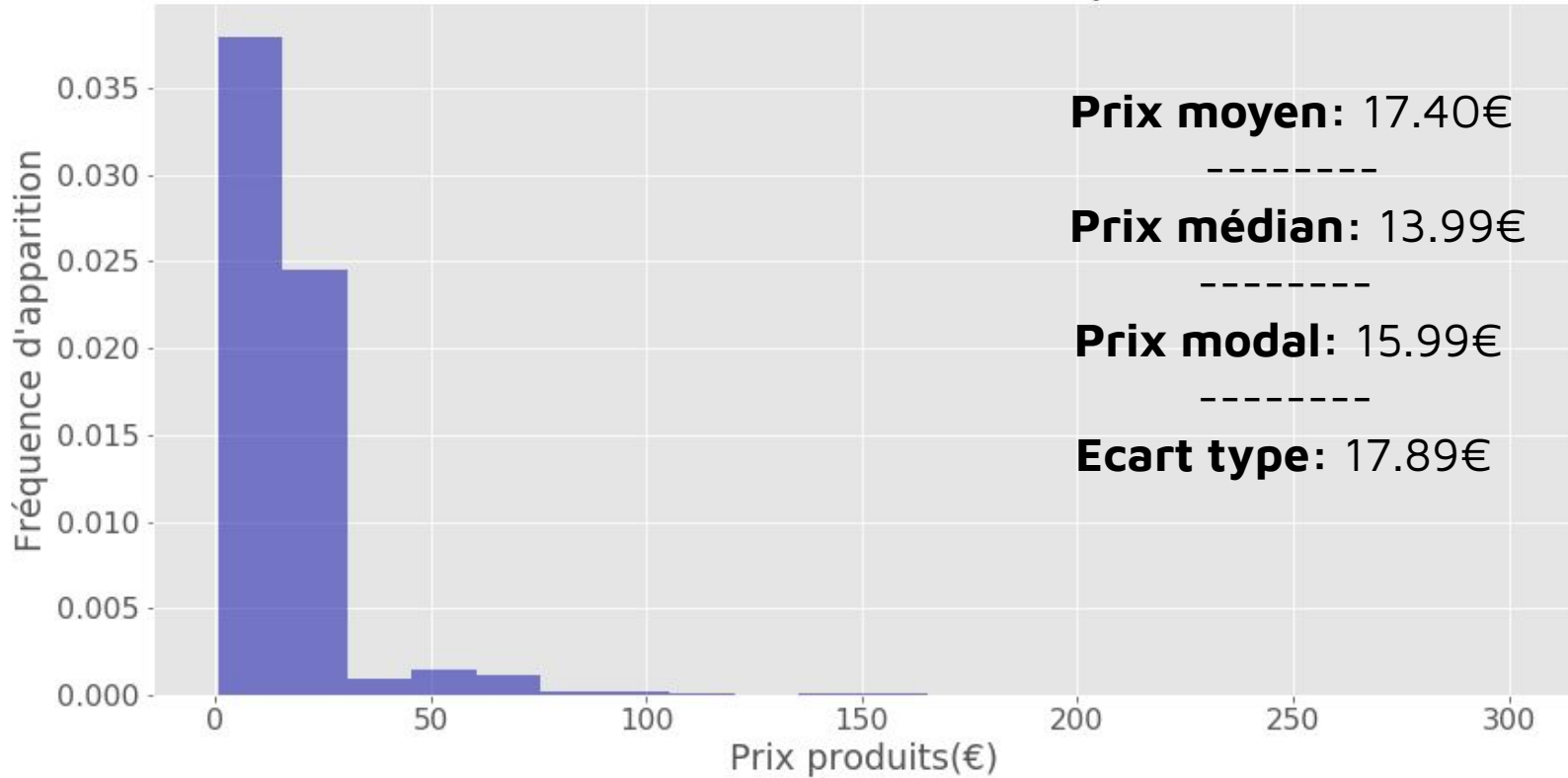


L'analyse exploratoire se fera donc **sur 11 mois** au lieu de 12.



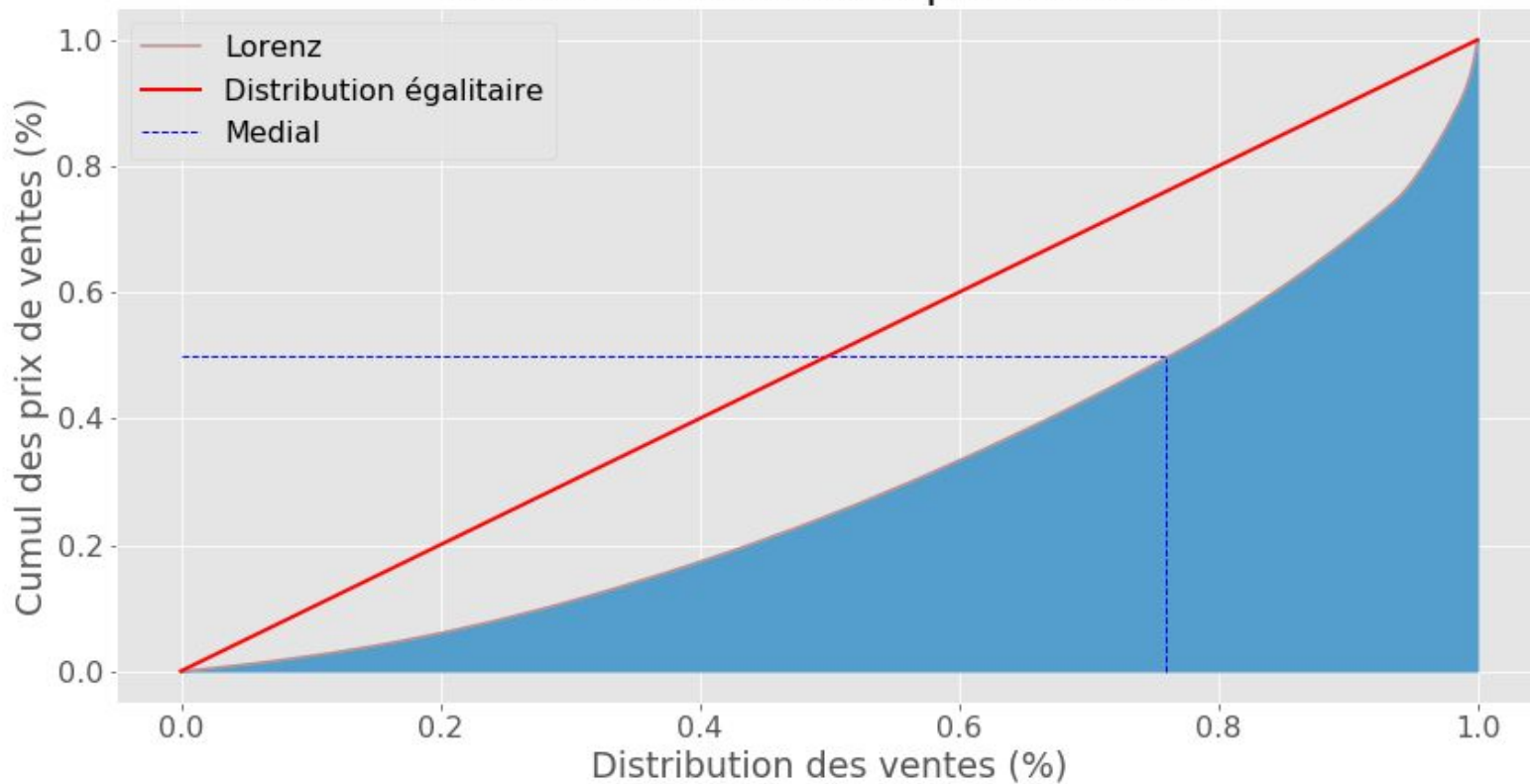
■■■€

Distribution de la variable "price"



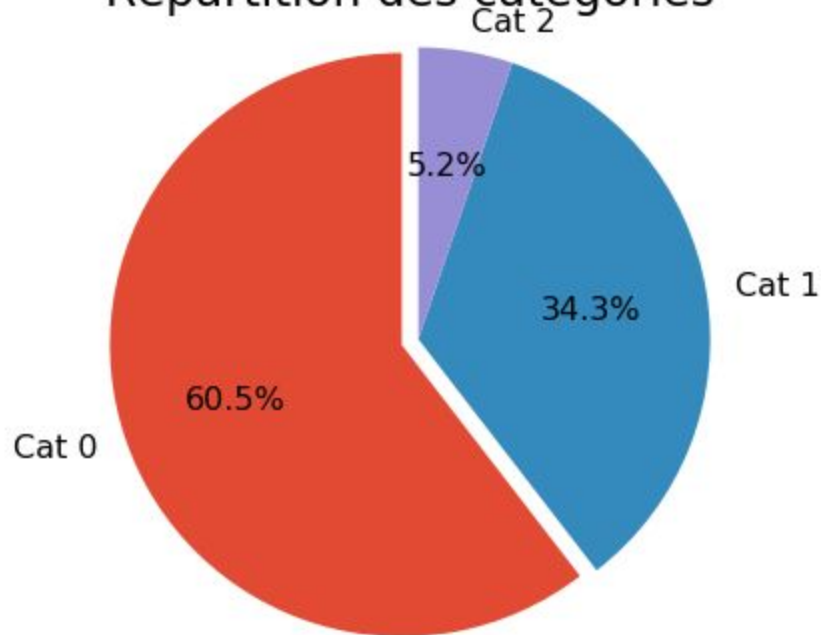
**Les produits < 30€ sont les plus populaires.
Peu de dispersion des prix.**

Courbe de Lorenz des prix de vente



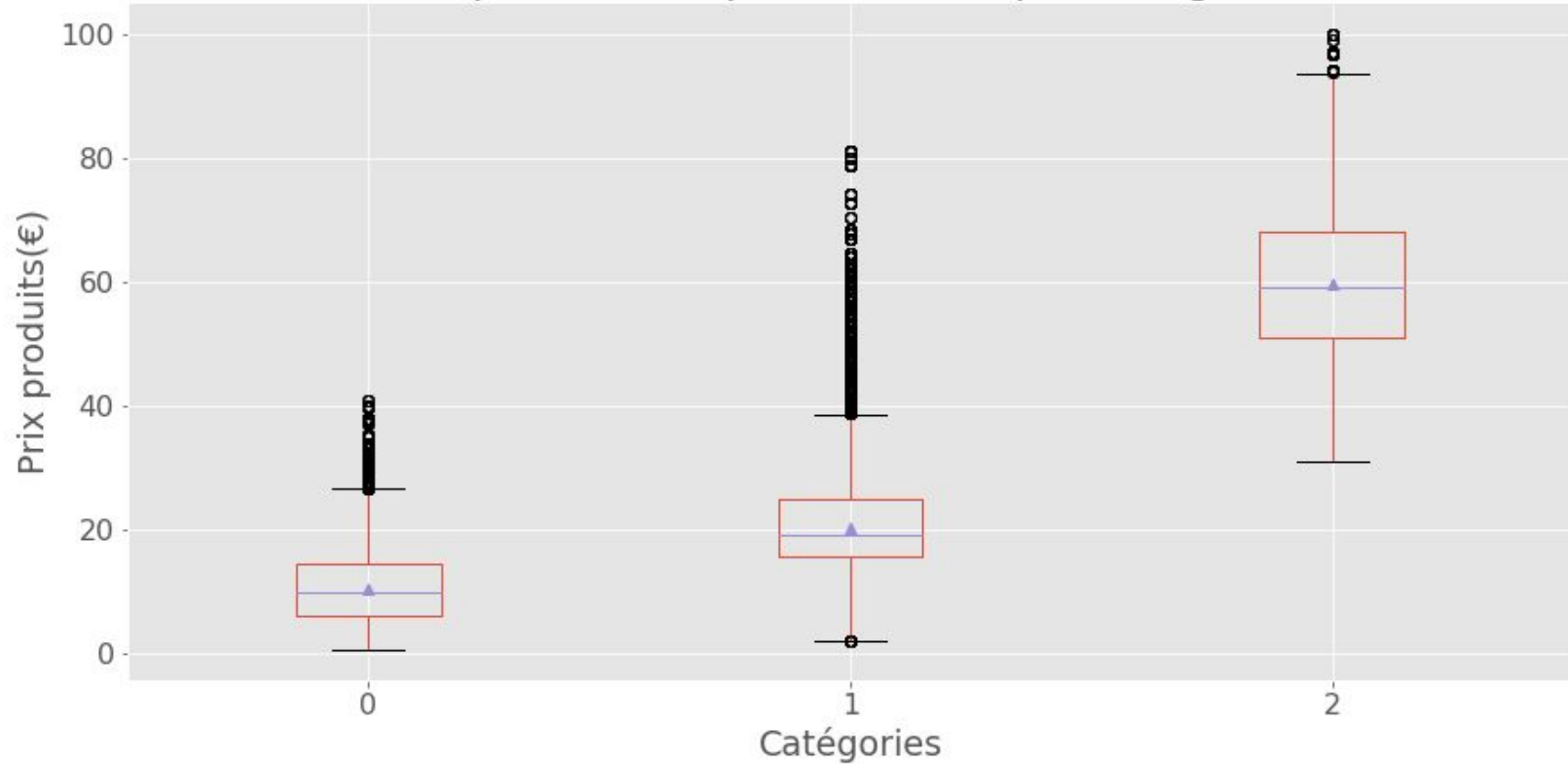
Répartition plutôt égalitaire: 76% des ventes représentent 50% du montant total des prix de vente - Confirmation par **l'indice de Gini 0.39**

Répartition des catégories



Plus de **60% des ventes** se font sur les produits de **catégorie 0**, mais ceux de la **catégorie 2** représentent seulement **5% des ventes**. **Pourquoi ?**

Boxplot grouped by caten
Dispersion des prix de vente par catégorie



Moins de volume en catég. 2 certainement dû aux prix plus élevés.

Des **catégories produits** certainement **complémentaires** mais **non équitables en termes d'offre...**

Catégorie 0 : Plus de 2300 produits.

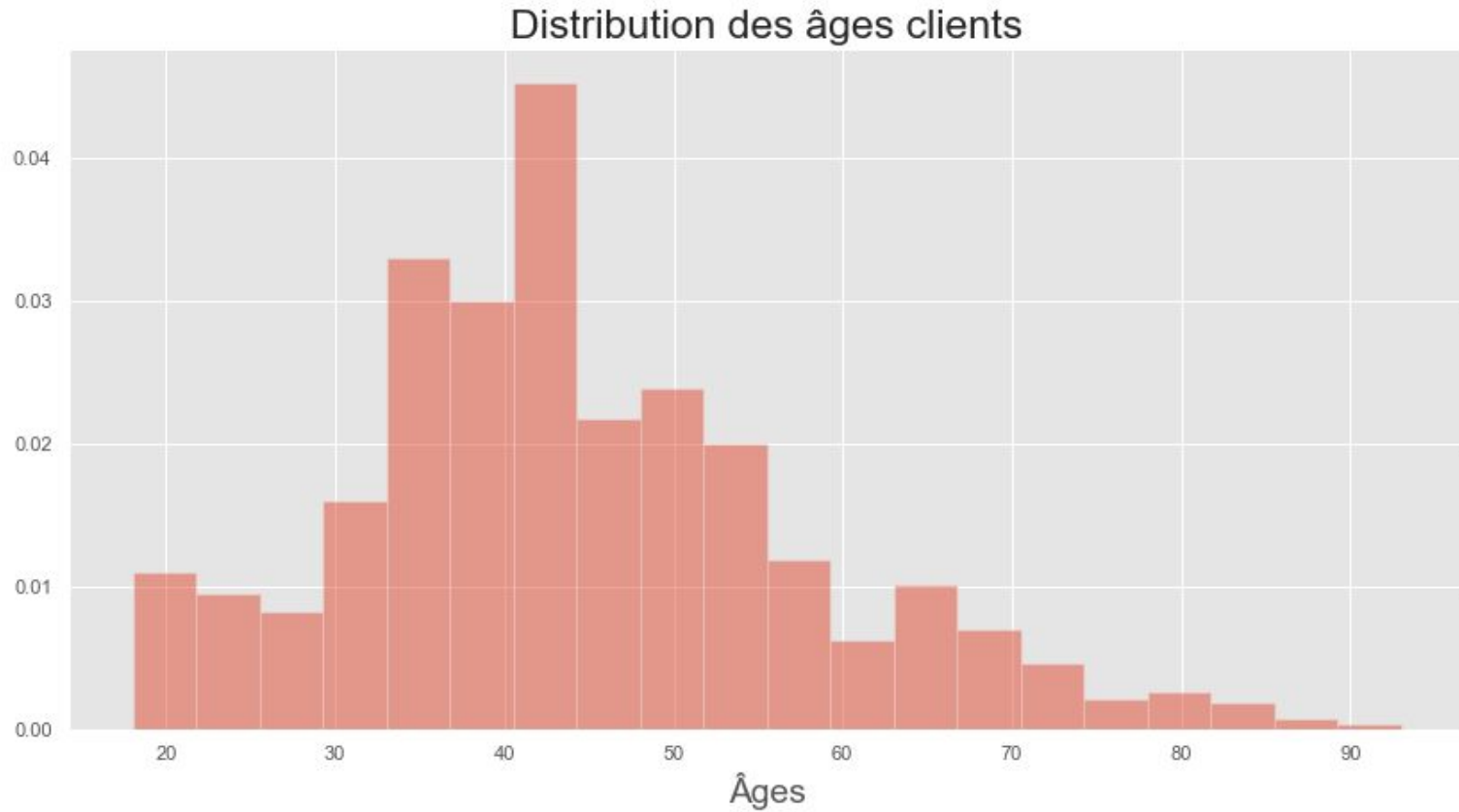
→ CA \approx 2 Millions d'euros

Catégorie 1 : Plus de 739 produits.

→ CA \approx 2.2 Millions d'euros

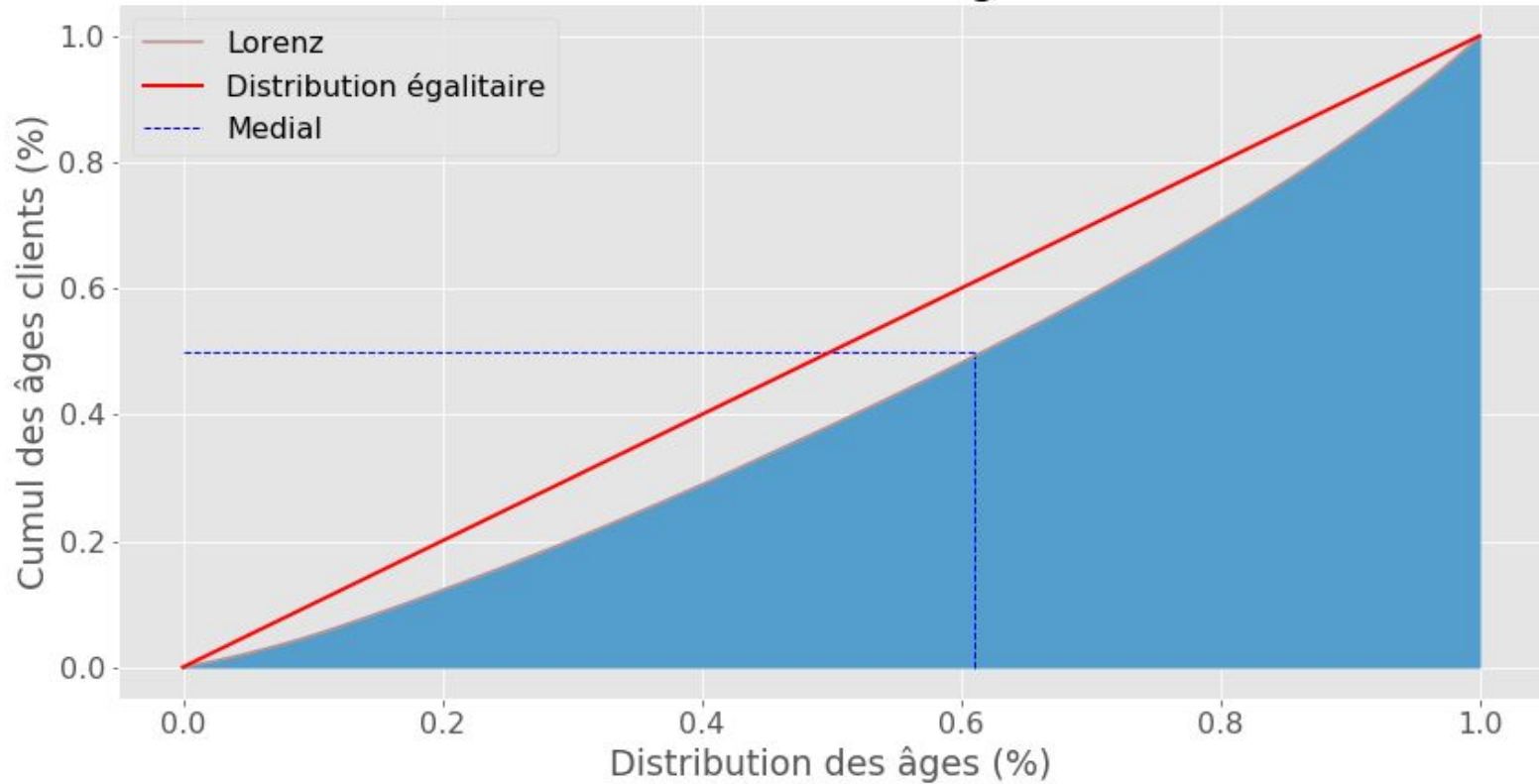
Catégorie 2 : Seulement 239 produits.

→ CA \approx 1.2 Millions d'euros



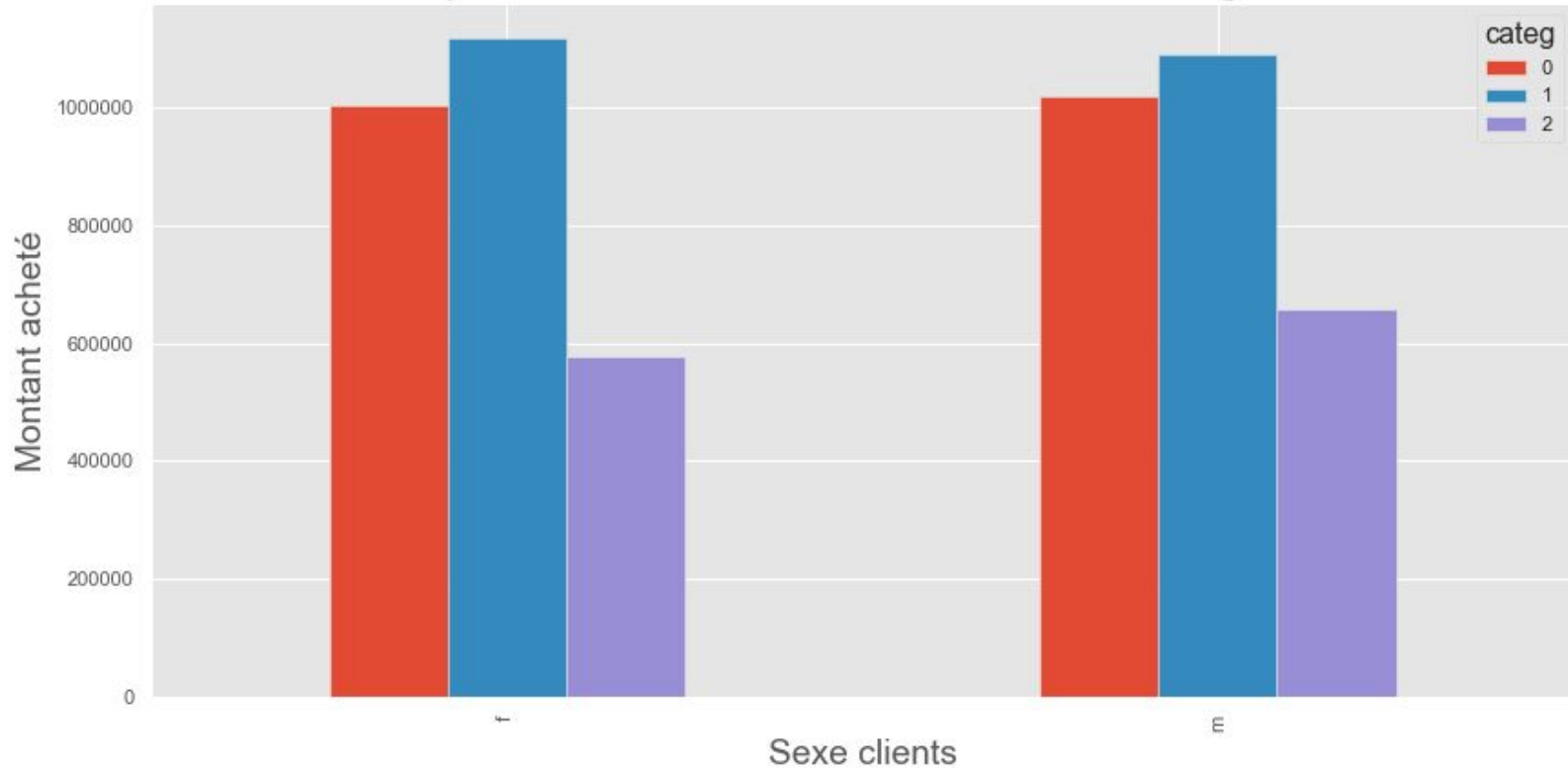
Des clients jeunes avec une plus forte concentration entre 30 et 55 ans.

Courbe de Lorenz des âges clients



Répartition égalitaire: 61% des âges représentent 50% du montant total des âges clients - Confirmation par **l'indice de Gini 0.17**

Répartition Hommes/Femmes selon les catégories



Hommes/Femmes sont équitablement représentés sur les **3 catégories**.

Hypothèses de recherche :

Avant d'analyser les différentes relations entre les dimensions précédemment établies, il est indispensable de poser un cadre hypothétique.

«H0 : Les catégories de produits achetés n'influent pas sur les individus.»

«H1 : Les catégories de produits achetés influent sur l'âge des individus.»

«H2 : Un homme achète plus qu'une femme sur les 3 catégories.»

«H3 : Plus un individu est jeune, plus le montant de ses achats est élevé.»

«H4 : Plus un individu est jeune, plus il achète régulièrement des produits.»

«H5 : Plus un individu est jeune, plus il achète d'articles.»

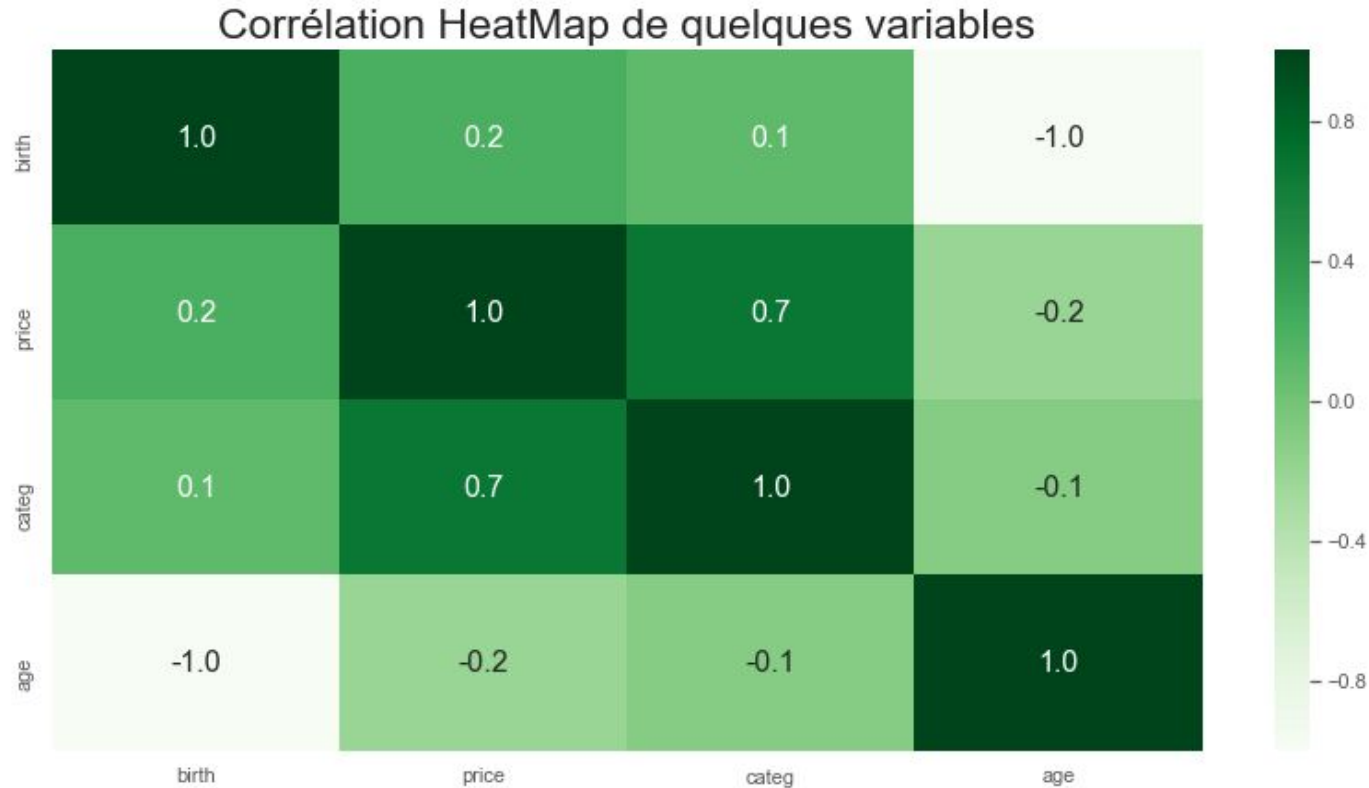
A photograph of a man and a young girl smiling. A diagonal white line runs from the bottom-left to the top-right, bisecting the image. The background is a blurred indoor setting.

Produit?

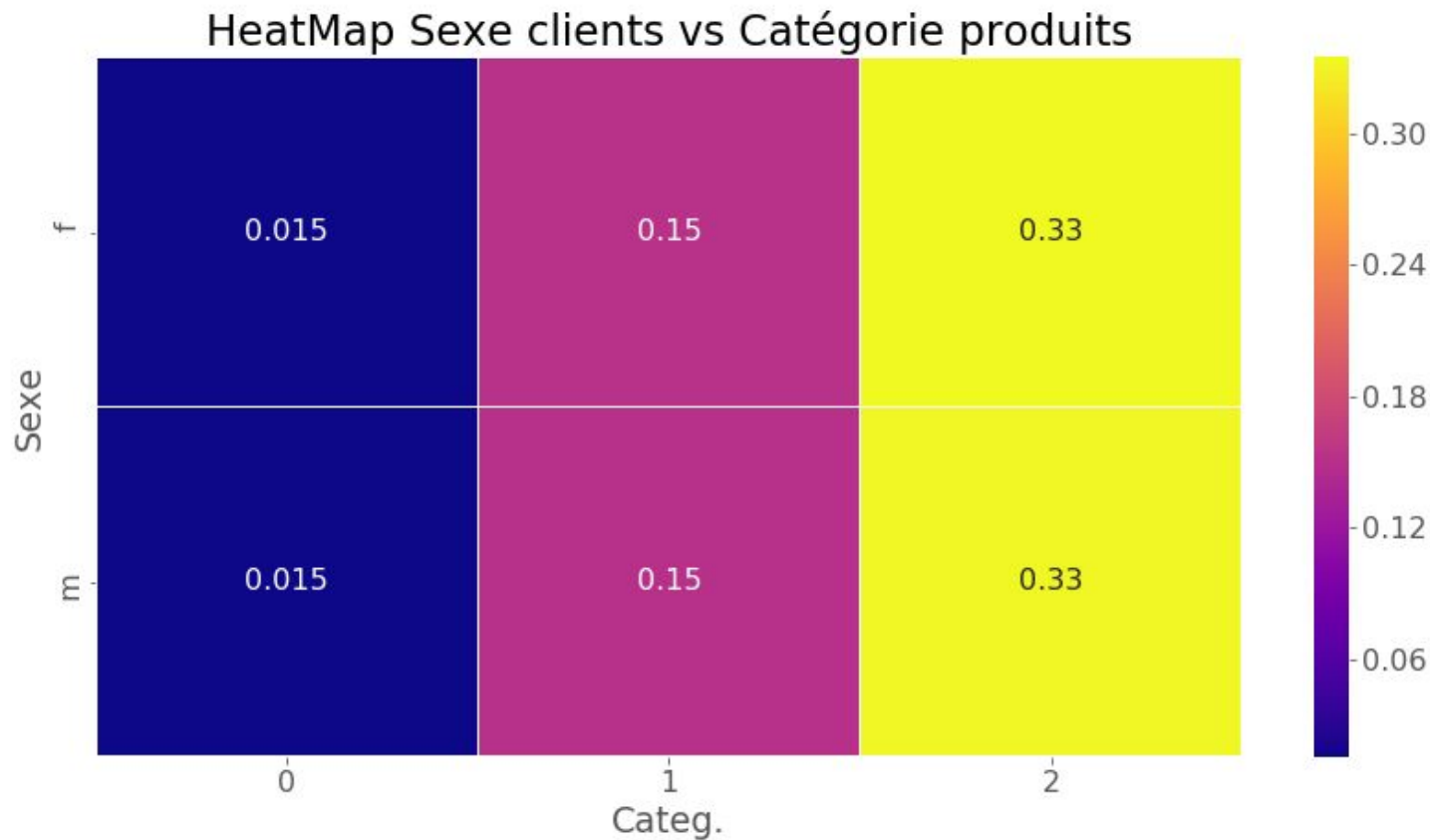
Corrélations possibles ?

Prix?

Âge?
Sexe?

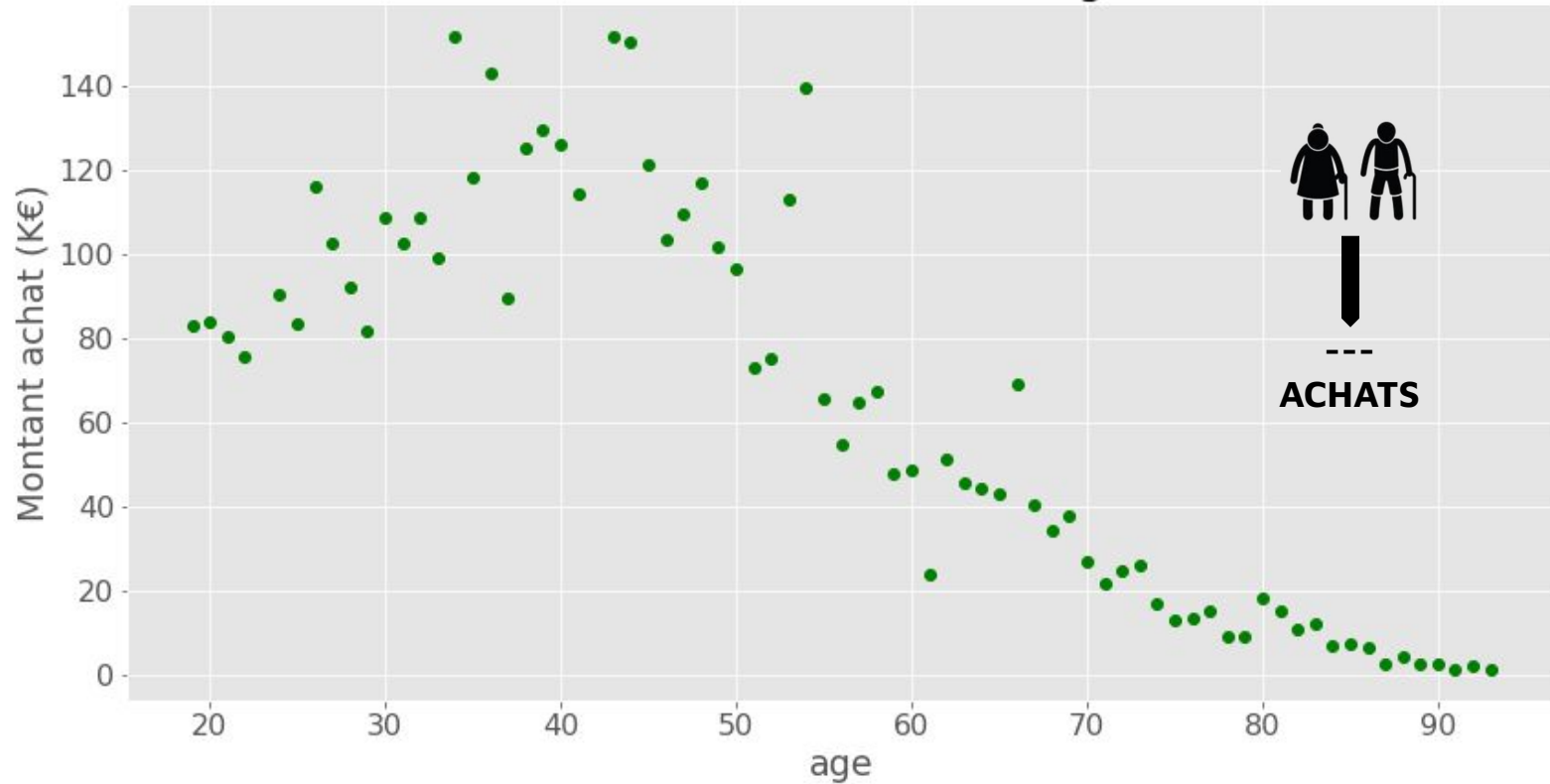


Premier aperçu des **corrélations** possibles entre les **variables numériques**.
On peut distinguer un lien entre **les prix et les catégories**.



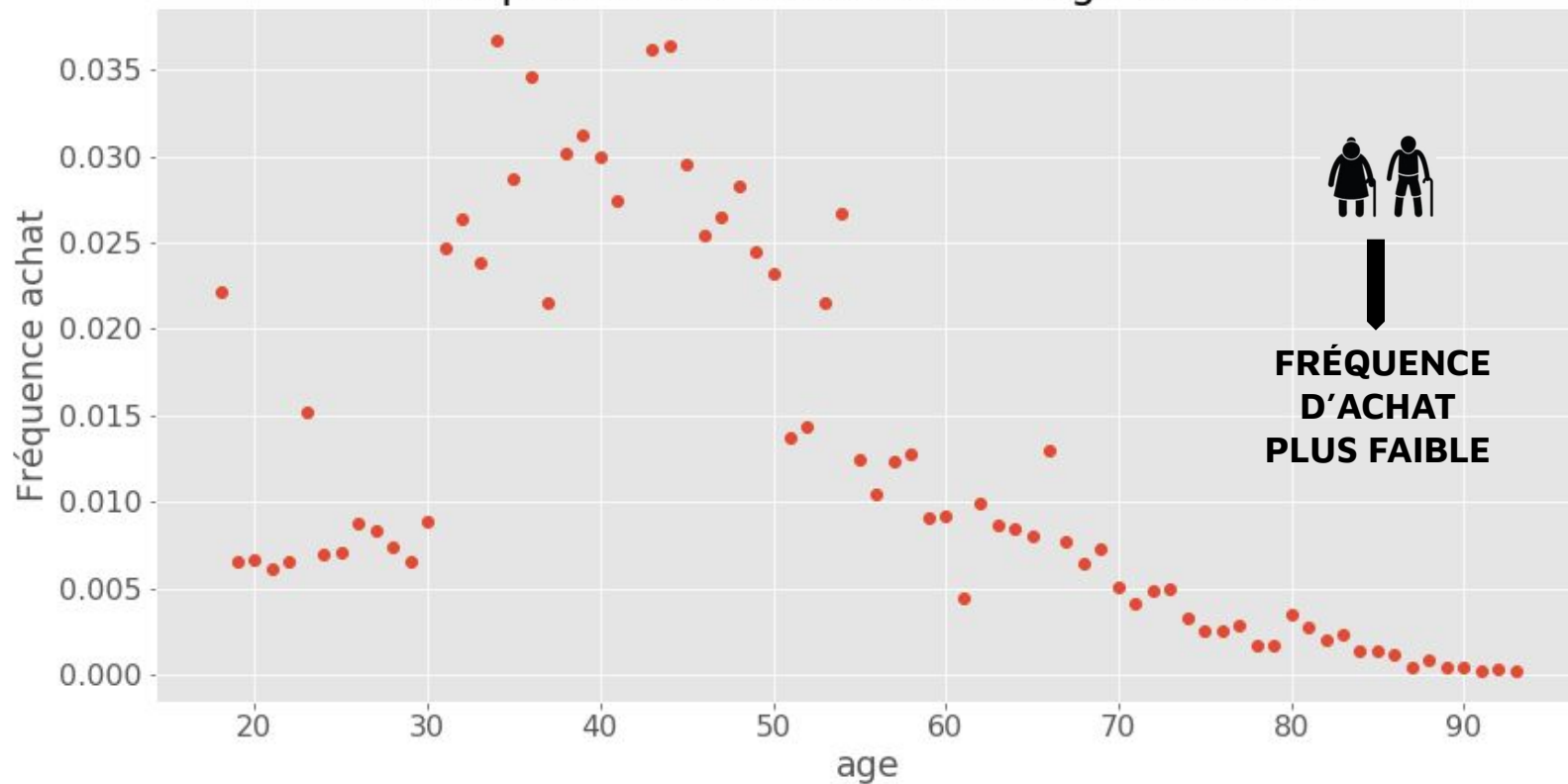
Corrélation entre les **catégories** et le **sexe des clients**, avec plus d'intensité sur la catégorie 2 ici en jaune. Le **Test statistique du Chi2** confirme la corrélation.

Montant Total des achats selon l'âge du client



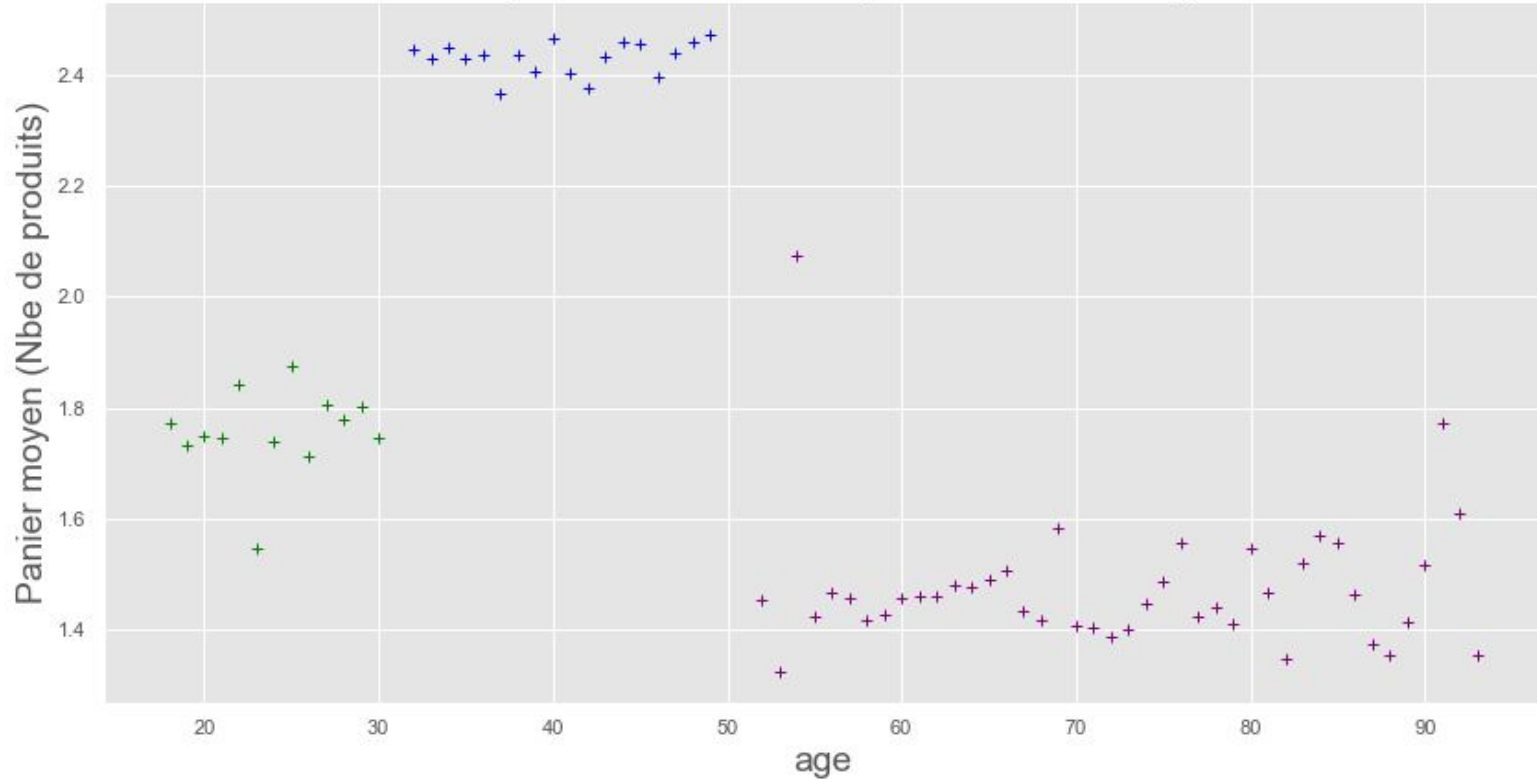
Corrélation linéaire appuyée également par le **coefficient de Pearson -0.77**
Liaison selon les tranches d'âge et le montant des achats.

Fréquence des achats selon l'âge du client

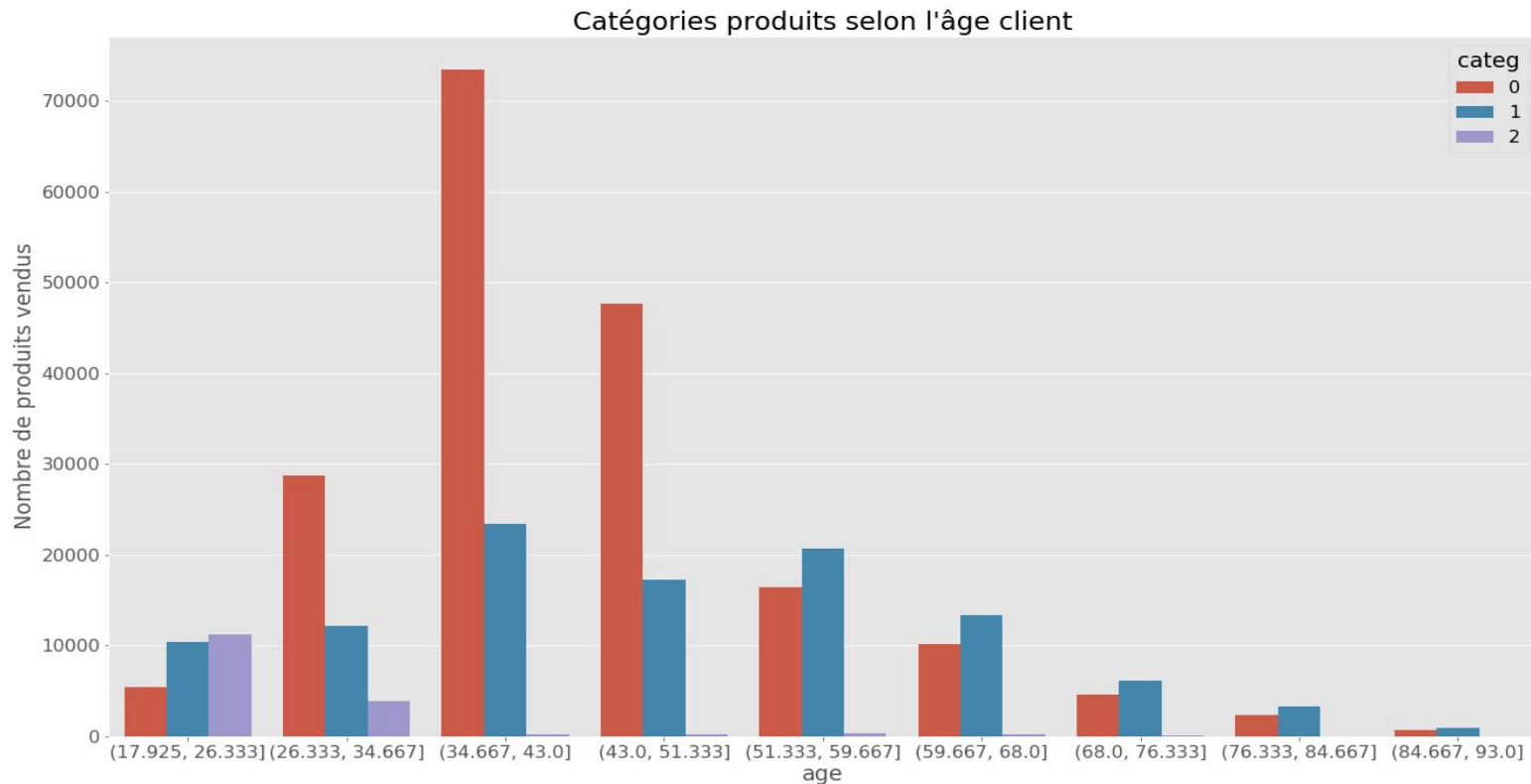


Corrélation linéaire appuyée également par le **coefficient de Pearson -0.53**
Liaison selon les tranches d'âge, on peut deviner certains groupes d'individus.

Panier moyen en nombre de produits selon l'age client



Corrélation linéaire appuyée également par le **coefficient de Pearson -0.57**
On distingue **3 clusters d'individus** qui impactent directement le panier moyen.



Corrélation entre les **catégories** et l'**âge des clients**.

Le **coefficient de corrélation eta carré 0.35** précise un lien de **faible intensité**.

Conclusions des différents tests de corrélation :

L'hypothèse nulle est rejetée en faveur de l'hypothèse alternative H1, il semble se créer des directions distinctes d'achat d'un individu à l'autre, suivant sa position démographique, suivant son âge...

L'hypothèse H2 est rejetée, les femmes comme les hommes achètent en proportion quasi égale sur les 3 catégories.

L'hypothèse H3 n'est pas rejetée, il semblerait que les individus plus âgés dépensent moins.

L'hypothèse H4 n'est pas rejetée, il y a une consommation plus régulière, une plus grande fidélité à l'enseigne par les individus plus jeunes.

L'hypothèse H5 n'est pas rejetée, les individus plus jeunes ont tendance à acheter plus d'articles.

Axes d'amélioration :

Perte de compétitivité sur les axes suivants :

- Catégorie 2 n'enregistre que très peu de ventes
- Clients plus âgés achètent moins que les plus jeunes

Propositions stratégiques :

- **Possibilité de développer l'offre produit en catégorie 2**
- **Adaptation de l'ergonomie du site**
- **Adaptation de l'offre produit**
- **Adaptation de services personnalisés**
- **Développement de points de vente de proximité**