

**PROJET-6 Analysez les ventes d'une librairie avec Python**

P.Moty

Nous avons ouvert il y a deux ans de cela notre librairie en ligne, et nous aimerions faire le point à présent. Cela nous permettra de décider de la marche à suivre, par exemple décider si nous devons créer certaines offres, adapter certains prix, etc.

Cette analyse va être découpée en deux parties :

* Une analyse des différents indicateurs de vente ;
* Une analyse plus ciblée sur les clients : l’objectif serait cette fois-ci de comprendre le comportement de nos clients en ligne, pour pouvoir ensuite comparer avec la connaissance acquise via nos librairies physiques.

 SOMMAIRE

[ Les données 2](#_Toc102232921)

[ Synthèse 2](#_Toc102232922)

[ Analyse macro 3](#_Toc102232923)

[I. Les différents indicateurs et graphiques autour du chiffre d’affaire 3](#_Toc102232924)

[II. L’évolution dans le temps 3](#_Toc102232925)

[III. Mettre en place une décomposition en moyenne mobile pour évaluer la tendance globale 4](#_Toc102232926)

[IV. Zoom sur les références 5](#_Toc102232927)

[V. La répartition du Chiffre d’affaires par catégorie 5](#_Toc102232928)

[VI. Les profils de nos clients 5](#_Toc102232929)

[VII. La répartition du chiffre d'affaires entre eux, via par exemple une courbe de Lorenz 6](#_Toc102232930)

[ Analyse des comportements des clients en ligne 7](#_Toc102232931)

[I. Le lien entre le genre d’un client et les catégories des livres achetés 7](#_Toc102232932)

[II. le lien entre l’âge des clients et le montant total des achats 8](#_Toc102232933)

[III. le lien entre l’âge des clients et la fréquence d’achat 8](#_Toc102232934)

[IV. le lien entre l’âge des clients et la taille du panier moyen 8](#_Toc102232935)

[V. le lien entre l’âge des clients et les catégories des livres achetés 9](#_Toc102232936)

# Les données

3 tableaux :

* Customers
* Products
* Transactions

Customers :

* 8600 clients
* Le plus jeune client a 18 ans, le plus vieux a 93 ans

Products :

* 3286 produits
* 3 catégories

Transactions :

* Nombre de produits vendus : 679 332 ventes

*Voir le pdf du notebook ci-joint.*

# Synthèse

|  |
| --- |
| * **Le chiffre d’affaire est de 11 856 731.75 € entre le 01/03/2021 et le 28/02/2023**  (à noter : une erreur dans l’acquisition des données au mois d’octobre 2021 concernant la catégorie 1). * **Le genre des clients n’a que peu d’impact sur nos données tandis que l’âge joue un rôle significatif dans les comportements des clients** par rapport à nos références et nos gammes de produits. Les moins de 40 ans dépensent le plus. * **Les catégories 0 et 1 représentent 76.6% du chiffre d’affaire total.** * **4 clients se distinguent** dans notre clientèle puisqu’ils représentent 7.43 % du chiffre d'affaire total. |

PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

# Analyse macro

## Les différents indicateurs et graphiques autour du chiffre d’affaire

**Ventes annuelles en € par an**

|  |  |  |
| --- | --- | --- |
| ANNEES | Chiffre d’affaire | Nombre de ventes |
| 2021 | 4 771 695,69 € | 278 335 |
| 2022 | 6 109 880.61 € | 346 500 |
| 2023 (2 mois) | 974 360.17 € | 54 497 |

* Entre 2021 et 2022 le chiffre d'affaire a connu une progression de 28.04 %
* Entre 2021 et 2022 le nombre de vente a connu une progression de 24,49 %
* Chiffre d’affaire global : 11 856 731.75 €

**Répartition des prix pour chaque catégorie**

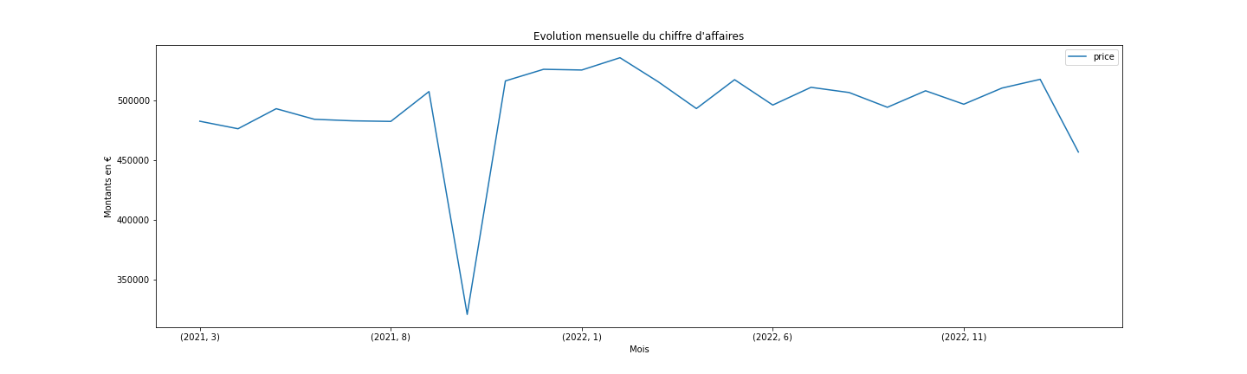
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Catégorie 0 | Catégorie 1 | Catégorie 2 | Total |
| Moyenne | 10.64 € | 20.49 € | 76.21 € | 17.45 € |
| Médiane | 9.99 € | 19.08 € | 62.83 € | 13.99 € |
| Mode | 4.99 € | 15.99 € | 68.99 € | Catégorie 0  Prix 15.99 € |

**L’essentiel des produits vendus toutes années et catégories confondues ont un prix < 50 €**

Analyse générale des mesures de tendances centrales sur l’ensemble des données

* Le prix moyen est : 17.45 €
* Le prix médian est 13.99 €
  + 50% des produits vendus ont un prix inférieur à 13.99 €
  + 50% des produits vendus ont un prix supérieur à 13,99 €
* Le prix le plus souvent vendu appartient à la catégorie 0 et est de 15.99 €

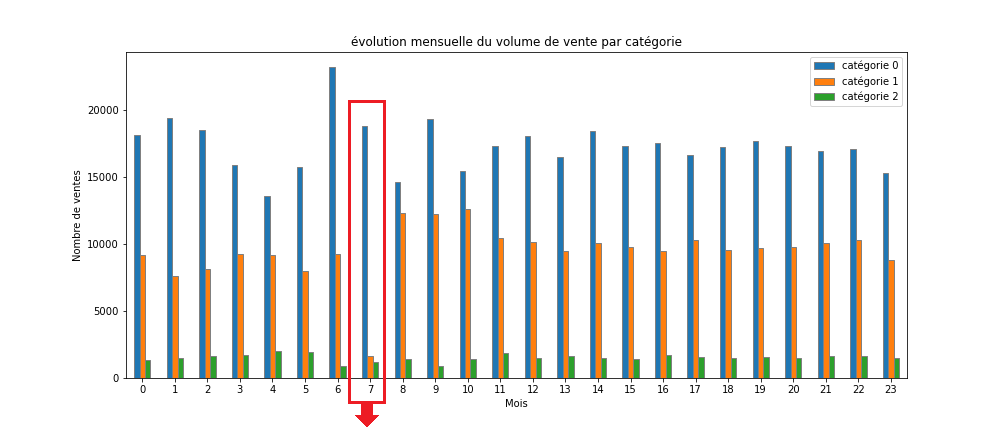
## L’évolution dans le temps



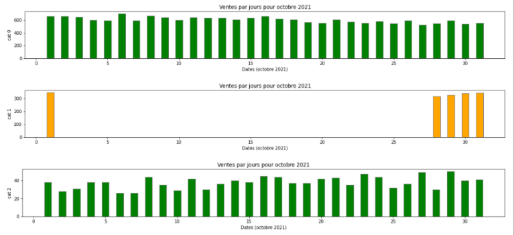
**Mois d’octobre : que s’est-il passé ?**

**Erreur** de collecte de données possible : pas de vente de produits de **catégorie 1 du 2 au 27 octobre** 2021. Cette baisse ne se retrouve pas plus tard.

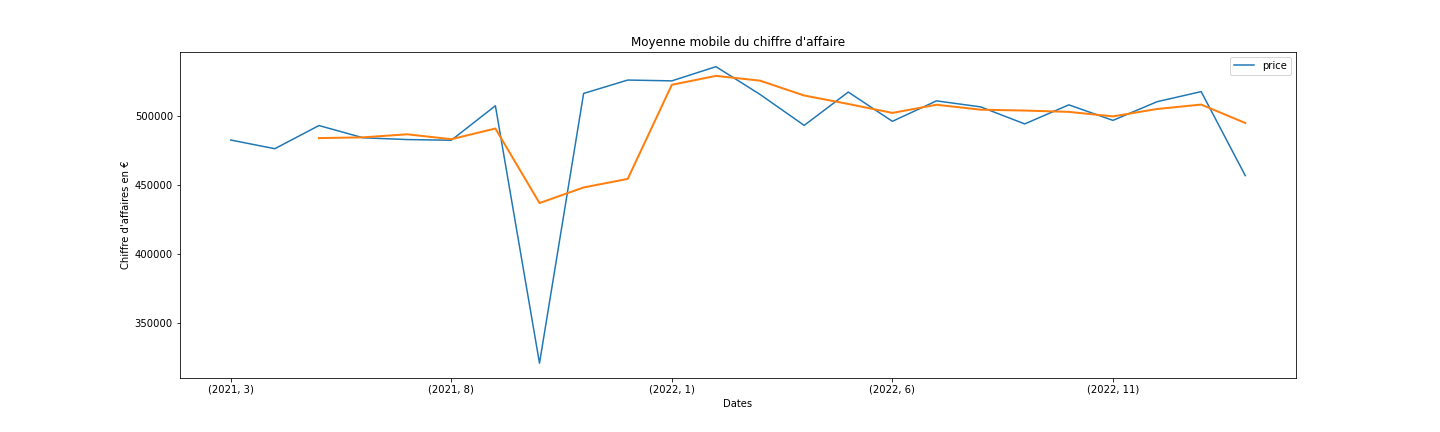
PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty



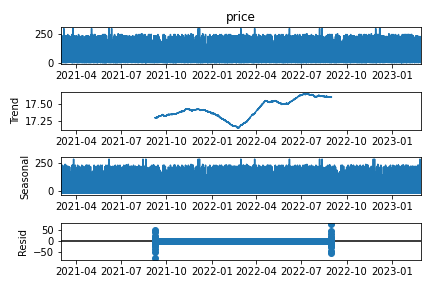
Ventes par jour et par catégorie au mois d’octobre 2021



## ***Mettre en place une décomposition en moyenne mobile pour évaluer la*** ***tendance globale***



**▾Après une croissance entre 2021 et 2022, le chiffre d’affaires se maintient en plateau. Pas de saisonnalité marquée.**



▸ Panier moyen : pas de saisonnalité mais une tendance croissante qui connaît une légère hausse de 17.25 € à 17.50 €

PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

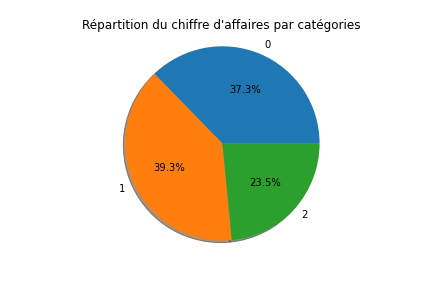
## Zoom sur les références

|  |  |  |  |
| --- | --- | --- | --- |
| Les 10 produits les plus vendus | | Les 10 produits les moins vendus | |
| LES TOPS | | **LES FLOPS** | |
| Références | Nombre de ventes | Références | Nombre de ventes |
| 1\_369 | 2252 | 0\_1151 | 1 |
| 1\_417 | 2189 | 0\_1284 | 1 |
| 1\_414 | 2180 | 0\_1379 | 1 |
| 1\_498 | 2128 | 0\_1498 | 1 |
| 1\_425 | 2096 | 0\_1539 | 1 |
| 1\_403 | 1960 | 0\_1601 | 1 |
| 1\_412 | 1951 | 0\_1633 | 1 |
| 1\_413 | 1945 | 0\_1683 | 1 |
| 1\_406 | 1939 | 0\_1728 | 1 |
| 1\_407 | 1935 | 0\_2201 | 1 |

* **Les produit les plus vendus appartiennent à la catégorie 1**
* **Les produits les moins vendus appartienne à la catégorie 0**

## La répartition du Chiffre d’affaires par catégorie

|  |  |  |  |
| --- | --- | --- | --- |
| Catégorie | Répartition des références | Prix moyen d’un livre | Apport au chiffre d’affaire |
| 0 | 2308 livres  (70.24% du stock) | 11.73 € | 4 421 938.76 €  (37.3%) |
| 1 | 739  (22.49%) | 25.53 € | 4 653 722.69 €  (39.3%) |
| 2 | 239  (7.27%) | 108.35 € | 2 780 275.02 €  (23.5%) |



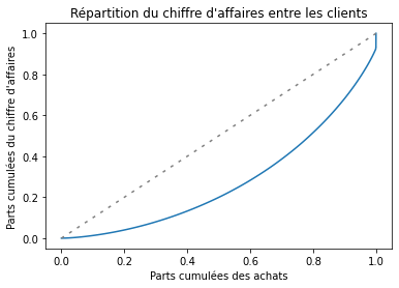
* **L’apport des catégories au chiffre d’affaires est équilibré** avec néanmoins une plus faible représentation de la catégorie 2 (23.3%) et un avantage de la catégorie 1 (39.3%)
* Bien que la catégorie 1 comporte moins de références que la catégorie 0 (22.49% vs 70.20%), elle apporte la plus grande part de chiffre d’affaires (39.3%)
* Le prix moyen du livre de catégorie 2 est le plus élevé (108.35 €). Bien que la catégorie 2 ne représente que 7.27% du stock elle apporte 1/5e du chiffre d’affaires

## Les profils de nos clients

* Le client avec l'id : c\_1609 est un homme né en 1980 âgé de 42 ans qui a passé 25488 commandes et a dépensé au total 324033.35 euros. Il peut s'agir d'un revendeur, d’un libraire ...
* **Lui comme les 4 meilleurs clients ont dépensé un montant supérieur aux autres acheteurs sur 3 ans qui permet de voir en eux des professionnels :**
  + c\_1609 🡺 324033.35 euros
  + c\_3454 🡺 113637.93 euros
  + c\_4958 🡺 289760.34 euros
  + c\_6714 🡺 153598.92 euros

PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

## La répartition du chiffre d'affaires entre eux, via par exemple une courbe de Lorenz

**Répartition du chiffre d’affaires entre les clients**

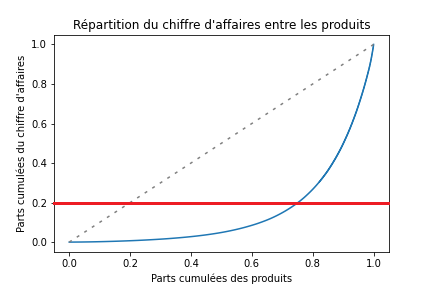
La loi de Pareto des 80/20 ou 20/80 s'applique dans ce cas confirmé par un indice de Gini à  **0.447**:

* 20% des clients font 50% du chiffre d’affaires
* 80% des clients font 50% du chiffre d’affaires

**CONCLUSION : Il y a un petit nombre de gros acheteurs. Parmi eux, Les 4 meilleurs clients représentent : 7.43 % du Chiffre d'affaire total ce qui tire les moyennes vers le haut, peu représentatives des particuliers.**

* Le nombre moyen d'achats par client : 78.99209302325582
* Le montant moyen par client : 1378.597263953488

**Répartition du Chiffre d’affaire entre les produits**

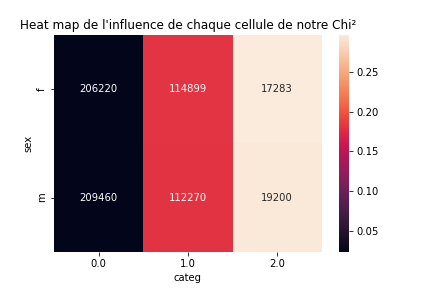


L'indice de Gini (0.694) est loin de 0, le poids des produits dans le chiffre d'affaires total est donc très inégal. Loi des 80/20 : moins de 30 % des produits fournissent 80% du Chiffre d'affaires

PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

# Analyse des comportements des clients en ligne

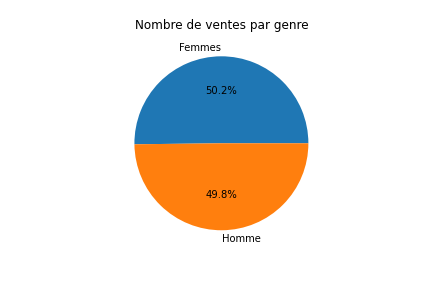
## Le lien entre le genre d’un client et les catégories des livres achetés

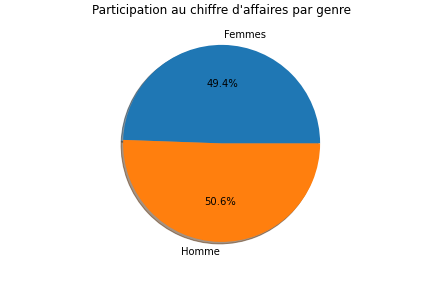


Le X² est de 147.003 à 6 ddl

soit une p-value de 3.327978654785752e-29

* **Selon l'hypothèse nulle (indépendance), il n'existe pas de lien entre les catégories achetées et le genre.**
  + Que l'on prenne un seuil à 5% (0,05) ou 1% (0,01), la p-value sera ici toujours très inférieure au seuil de significativité,
  + **L’hypothèse nulle est donc rejetée ici.**
* **L'hypothèse alternative HA est retenue : les catégories des livres achetées dépendent du genre de l'acheteur.** Il y a bien une dépendance entre le genre et les catégories de livre. Cependant, elle n’impacte pas significativement les comportements.

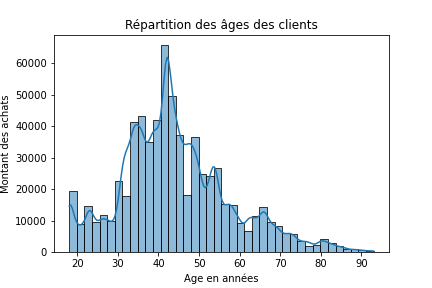
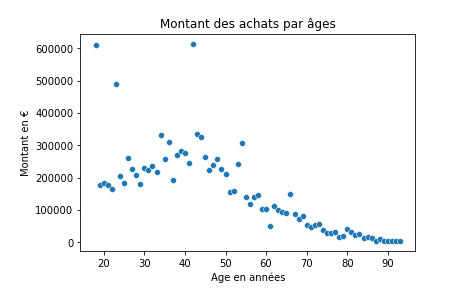
**Répartition des ventes et du chiffre d’affaires par genre**



**Les hommes et les femmes achètent autant de livres les uns que les autres et ont le même apport au chiffre d’affaires.**

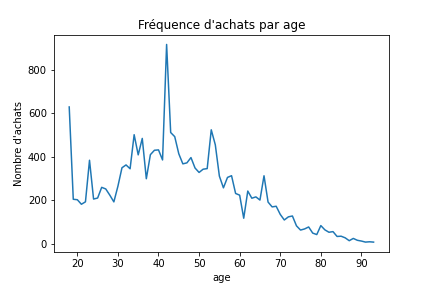
PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

## le lien entre l’âge des clients et le montant total des achats

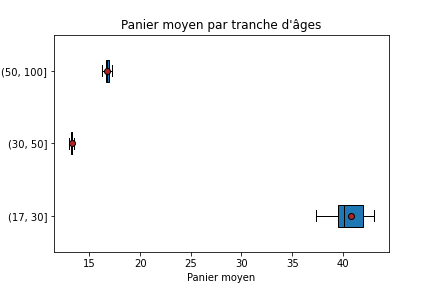
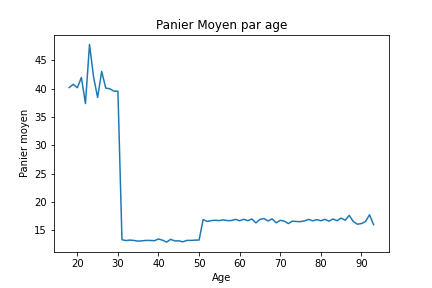
* **Les clients qui dépensent le plus sont entre 35 et 50 ans**
* Dans l’ensemble il existe une corrélation négative entre le montant des achats et l’âge des clients. **Plus un client est âgé plus ces dépenses diminuent.**

## le lien entre l’âge des clients et la fréquence d’achat



* **Les 40 ans ont la fréquence d’achat la plus élevée**
* Dans l’ensemble il existe une corrélation négative entre la fréquence des achats et l’âge des clients : **plus un client est âgé plus cette fréquence d’achats diminue**

## le lien entre l’âge des clients et la taille du panier moyen



**Il existe une corrélation entre les paniers moyens et les âges avec 3 paliers :**

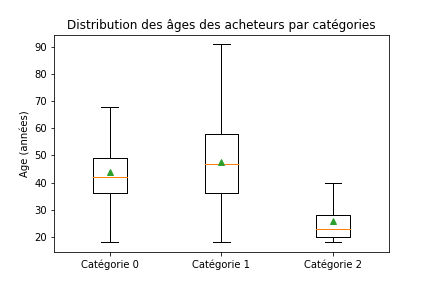
* **les moins de 30 ans ont un panier moyen 40 €**
* **les 30-50 ans ont un panier moyen <15 €**
* **Les plus de 50 ans entre 15 et 20 €**

PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

|  |  |  |  |
| --- | --- | --- | --- |
| Tranches d’âges | Montant des achats par tranche d’âges | Nombre d’achats par tranche d’âges | Montant moyen des achats par tranche d’âge |
| 17-37 | 5 064 750.89 | 213 655 | 23.70 |
| 37-57 | 5 031 521,58 | 360 725 | 13.94 |
| 57-77 | 1 526 947.13 | 91 110 | 16.75 |
| 77-100 | 232 716.13 | 13 842 | 16.81 |

Les 17-37 ans ont les paniers moyens les plus élevés et les 37-57 ans les moins élevés. Ces 2 tranches ceux celles qui rapportent le plus au CA de l’entreprise. En effet les 37-57 achètent plus de produits ce qui contribuent à leur importante représentation dans le CA.

## le lien entre l’âge des clients et les catégories des livres achetés



* **La catégorie 0 concerne en majorité   
  les 20-70**
* **La catégorie 1 touche l'ensemble de la clientèle**
* **La catégorie 2 n'est acheté que par   
  les 20-40 ans**

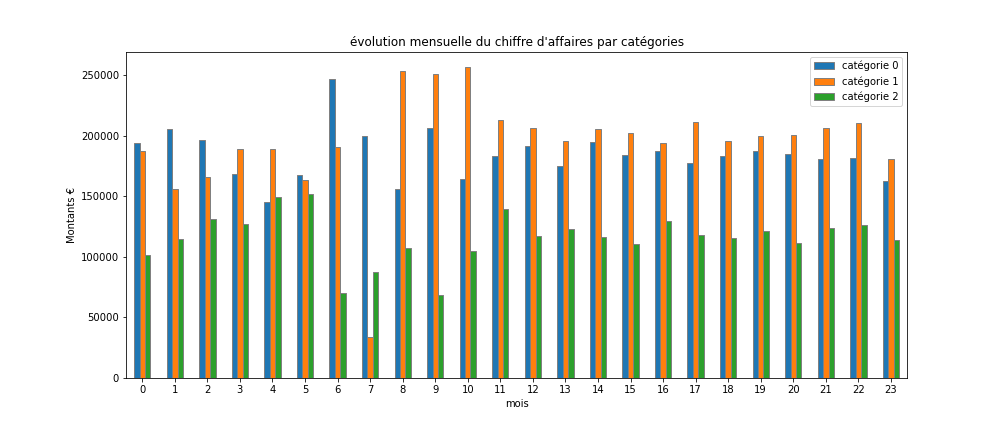
PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty

ANNEXES

Répartition des prix par catégories

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Catégories** | **Moyennes** | **Médianes** | **Modes** | **Graphiques** |
| **Catégorie 0** | 10.64 | 9.99 | 4,99 |  |
| **Catégorie 1** | 20,49 | 19.08 | 15.99 |  |
| **Catégorie 2** | 76,20 | 62,83 | 68,99 |  |

Evolution mensuelle du chiffre d’affaire



PROJET-6 Analysez les ventes d'une librairie avec R ou Python P.Moty