

# Analyse de clients



e-commerce  
centre commercial  
campagne marketing

# Introduction

## **Projet e-commerce :**

Analyse de données pour comprendre le comportement de clients.

A faire : segmentation RFM et une analyse par cohortes.

Dataset à nettoyer : db\_customers.csv

## **Projet centre commercial :**

Connaître la typologie de client qui achète

Dataset propre : Mall\_Customers.csv

## **Projet campagnes marketing :**

Analyse de clients basé sur les données récoltées des dernières campagnes de marketing.

Dataset à nettoyer : marketing\_campaign.csv



# E-commerce



# Dataset

order_id	customer_id	order_status	order_purchase_t
78b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02
78b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02
78b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02
0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef	delivered	2018-07-24
0c44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089	delivered	2018-08-08
...	...	...	...
b2549525ed64a053c	39bd1228ee8140590ac3aca26f2dfe00	delivered	2017-03-08
6b46f01ca7ac2f7bd8	1fca14ff2861355f6e5f14306ff977a7	delivered	2018-02-06
13d02aae0204711ab	1aa71eb042121263aafbe80c1b562c9c	delivered	2017-08-27
1b2631073c19f07b62	b331b74b18dc79bcd6532d51e1637c1	delivered	2018-01-08
4dee7af250b4be1a5	edb027a75a1449115f6b43211ae02a24	delivered	2018-03-08

1

15 colonnes

2

105 383 lignes

3

10 colonnes object

4

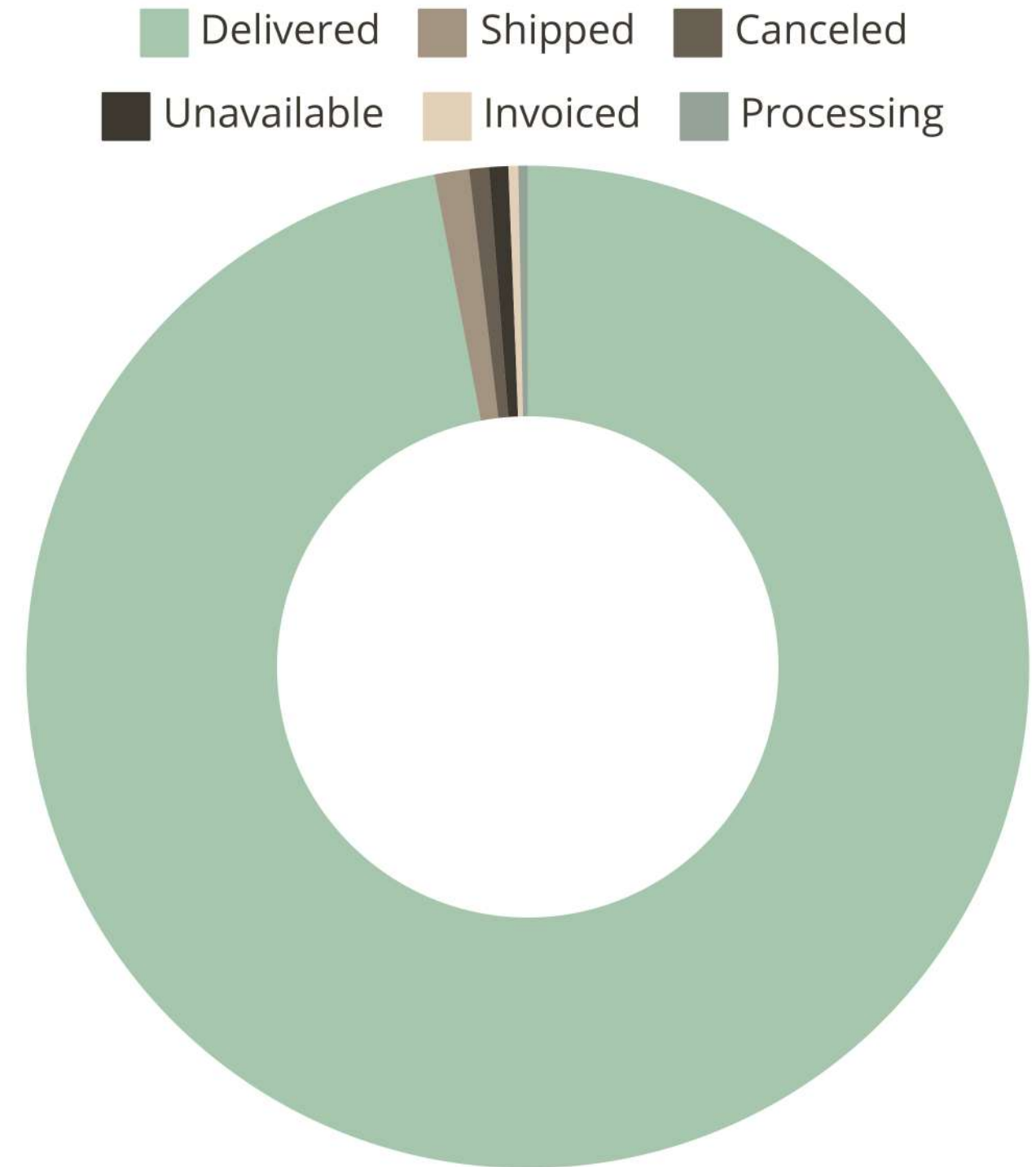
5 colonnes numerical  
(float64 & int 64)

5

Dataset à nettoyer

# Proportion des valeurs uniques dans “order\_status”

- On constate que “Delivered” est majoritaire.
- Nous avons donc divisé le dataset en deux : un dataset delivered et un canceled
- De cette manière, nous pouvons étudier les incidences différentes en fonction de la situation.



**Suppression des NaN  
dans le dataset  
“Delivered” (1%)**

**Remplacement des  
Nan dans le dataset  
“Canceled”**

**Suppression des lignes :  
'invoiced', 'processing',  
'created', 'approved',  
'canceled', 'unavailable' pour  
le dataset “Delivered”**

**Conservation de  
'canceled',  
'unavailable' pour le  
dataset “Canceled”**

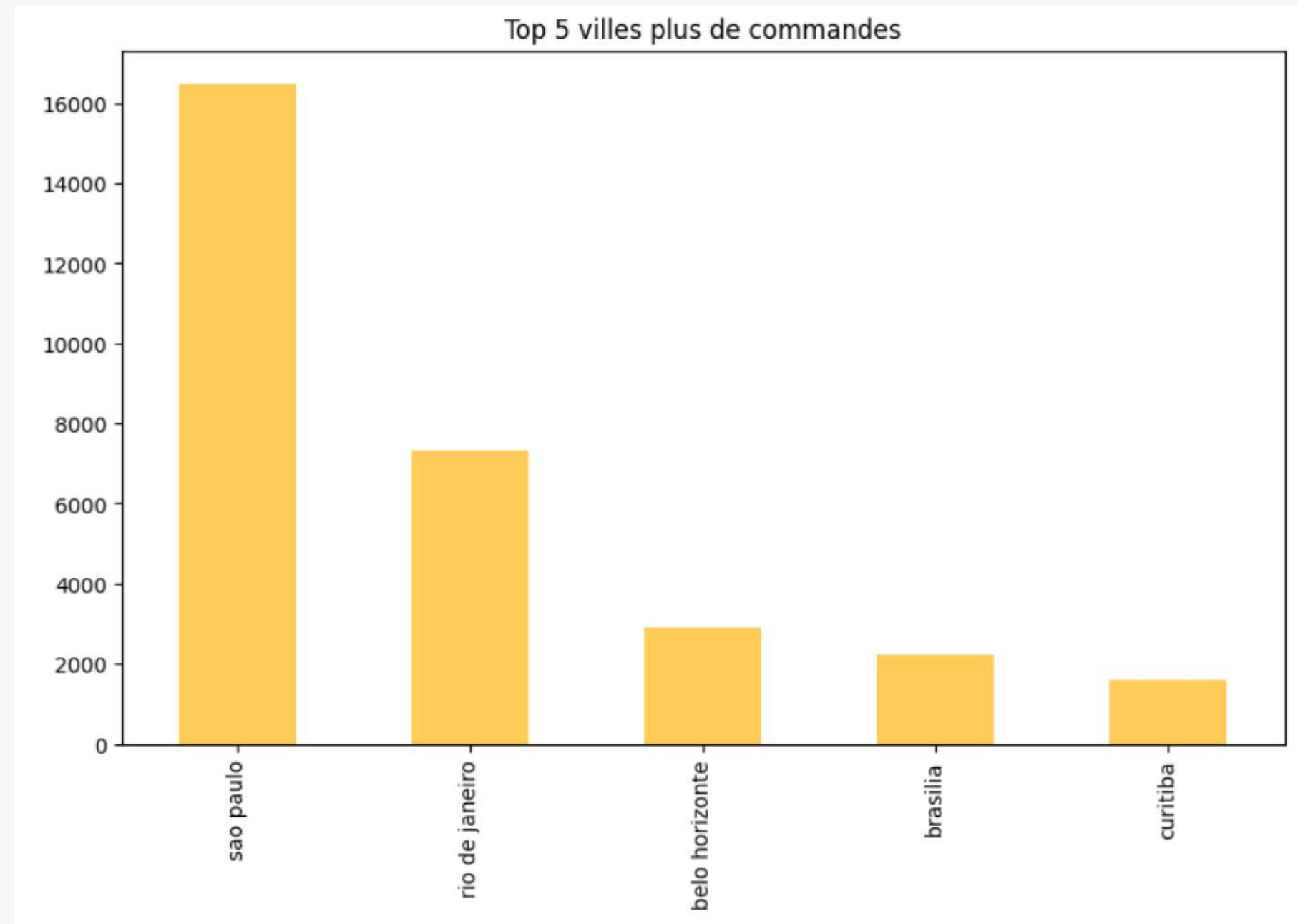


# Une ligne par client

	customer_unique_id	Recency	Frequency	Monetary	proc
0	0000366f3b9a7992bf8c76cfd3221e2	2205	1	141.90	bed_
1	0000b849f77a49e4a4ce2b2a4ca5be3f	2208	1	27.19	hea
2	0000f46a3911fa3c0805444483337064	2630	1	86.22	
3	0000f6ccb0745a6a4b88665a16c9f078	2414	1	43.62	
4	0004aac84e0df4da2b147fca70cf8255	2381	1	196.89	
...	...	...	...	...	
92065	ffcf5a5ff07b0908bd4e2dbc735a684	2540	1	2067.42	hea
92066	fffea47cd6d3cc0a88bd621562a9d061	2355	1	84.58	
92067	ffff371b4d645b6ecea244b27531430a	2662	1	112.46	
92068	ffff5962728ec6157033ef9805bacc48	2213	1	133.69	wa
92069	ffffd2657e2aad2907e67c3e9daecbeb	2577	1	71.56	

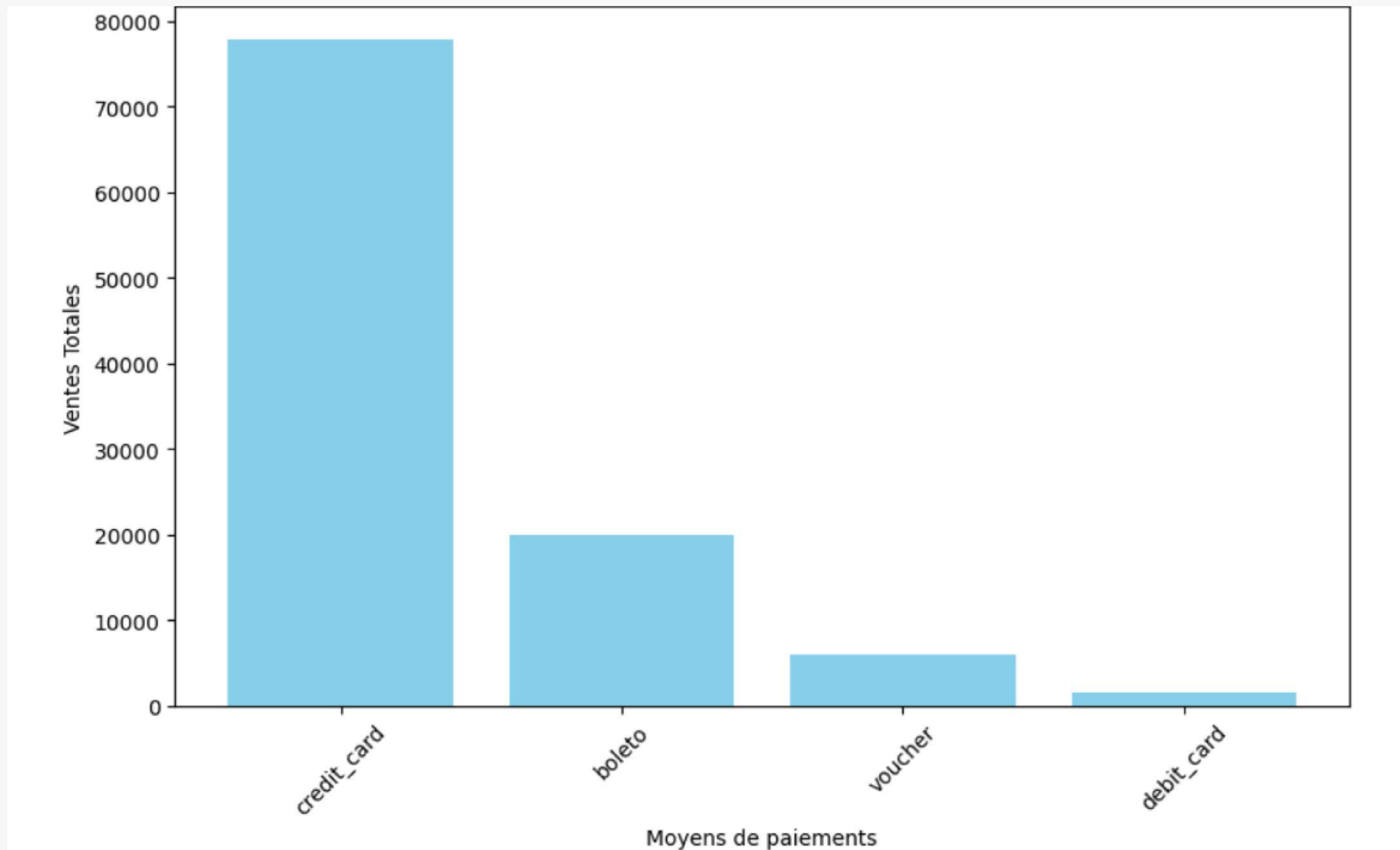
- Agrégation avec création des colonnes  
Délais de livraison, Frequency et Recency
- Rename des colonnes pour plus de  
clarté, on en profite pour passer le  
“payment\_value” en Monetary
- Réorganisation des colonnes
- On passe à 92 070 lignes et 13 colonnes

# Classement des villes les plus acheteuses

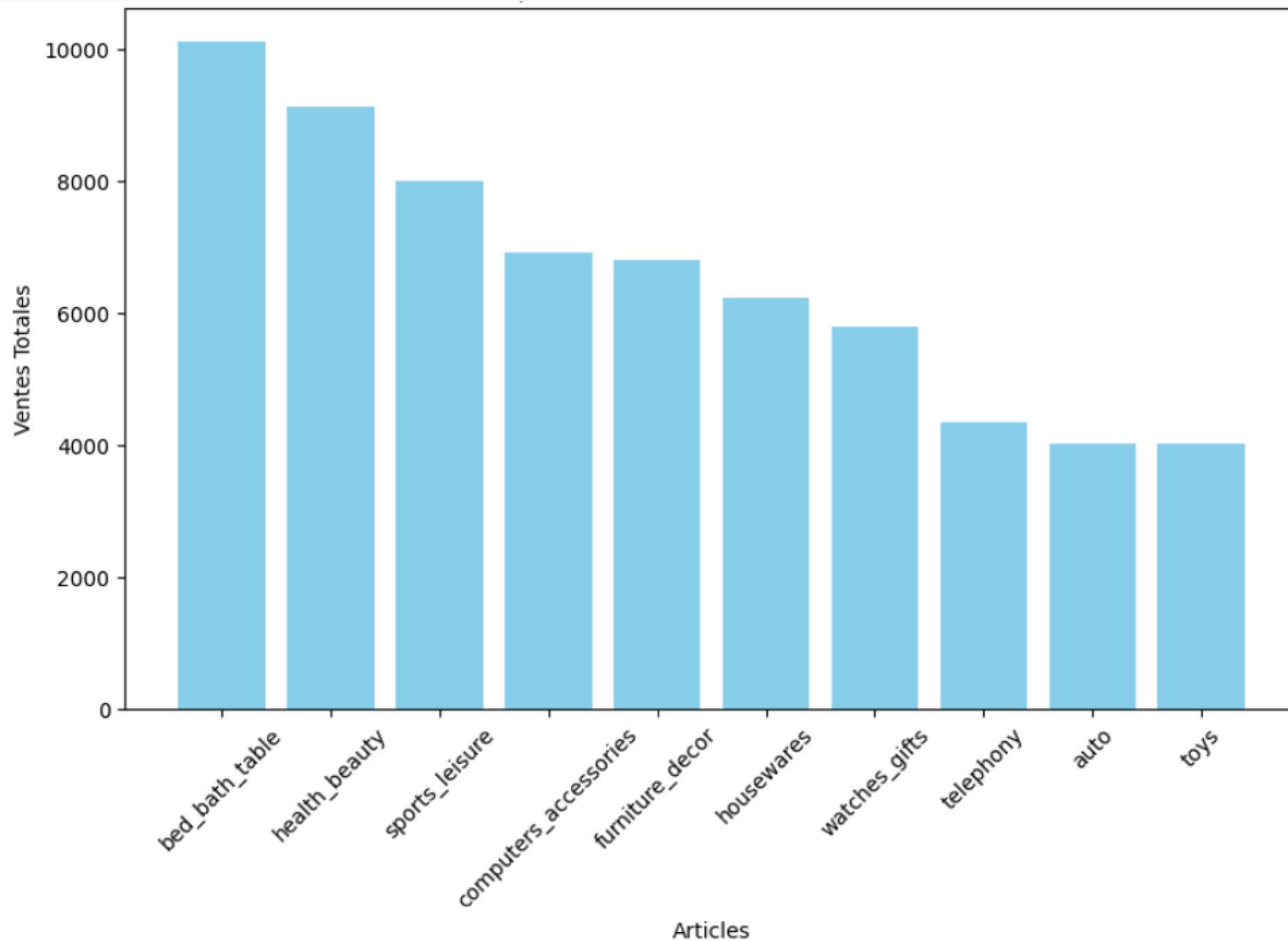




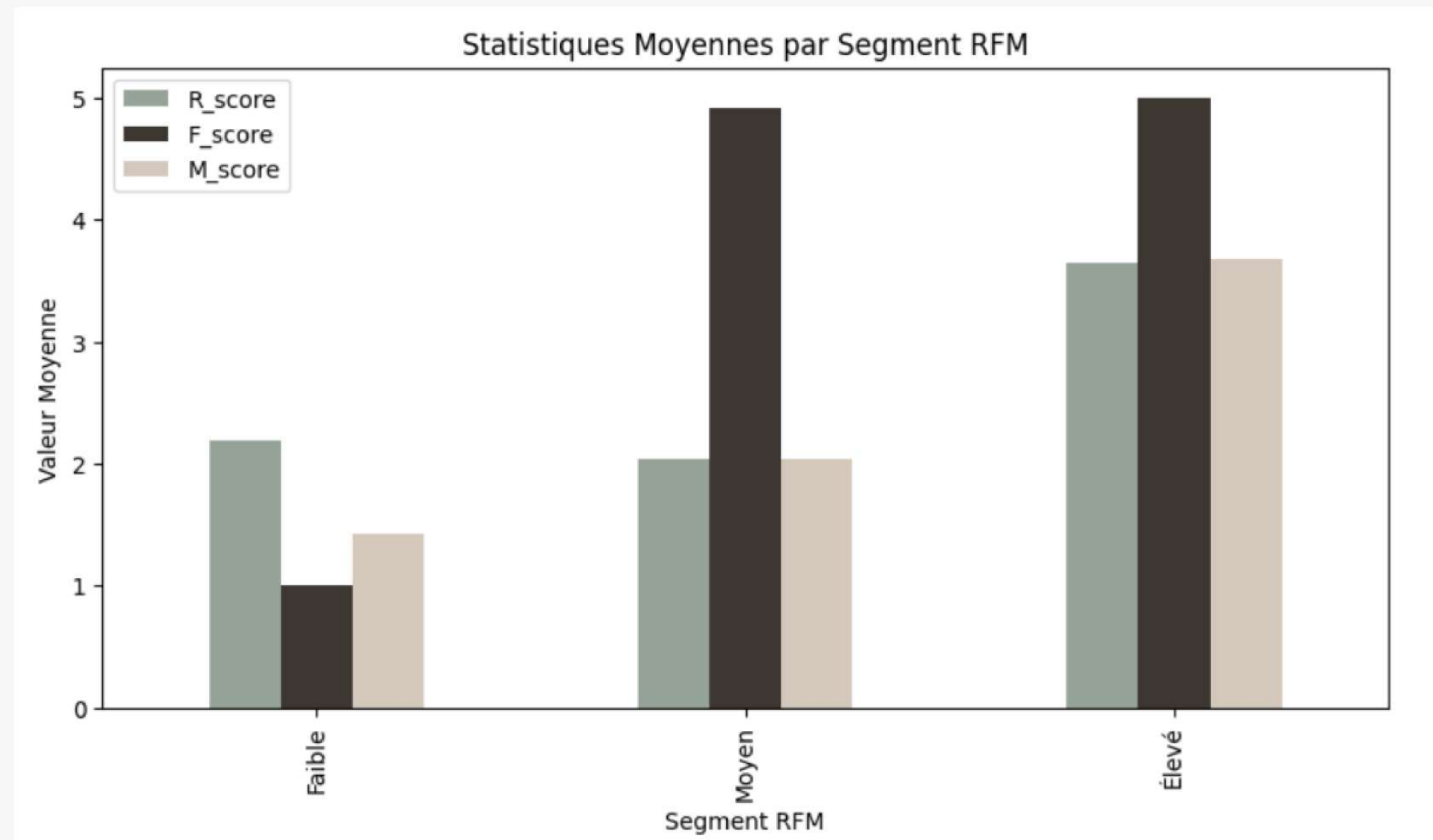
# Répartition des moyens de paiements



# Top 10 des articles les plus vendus

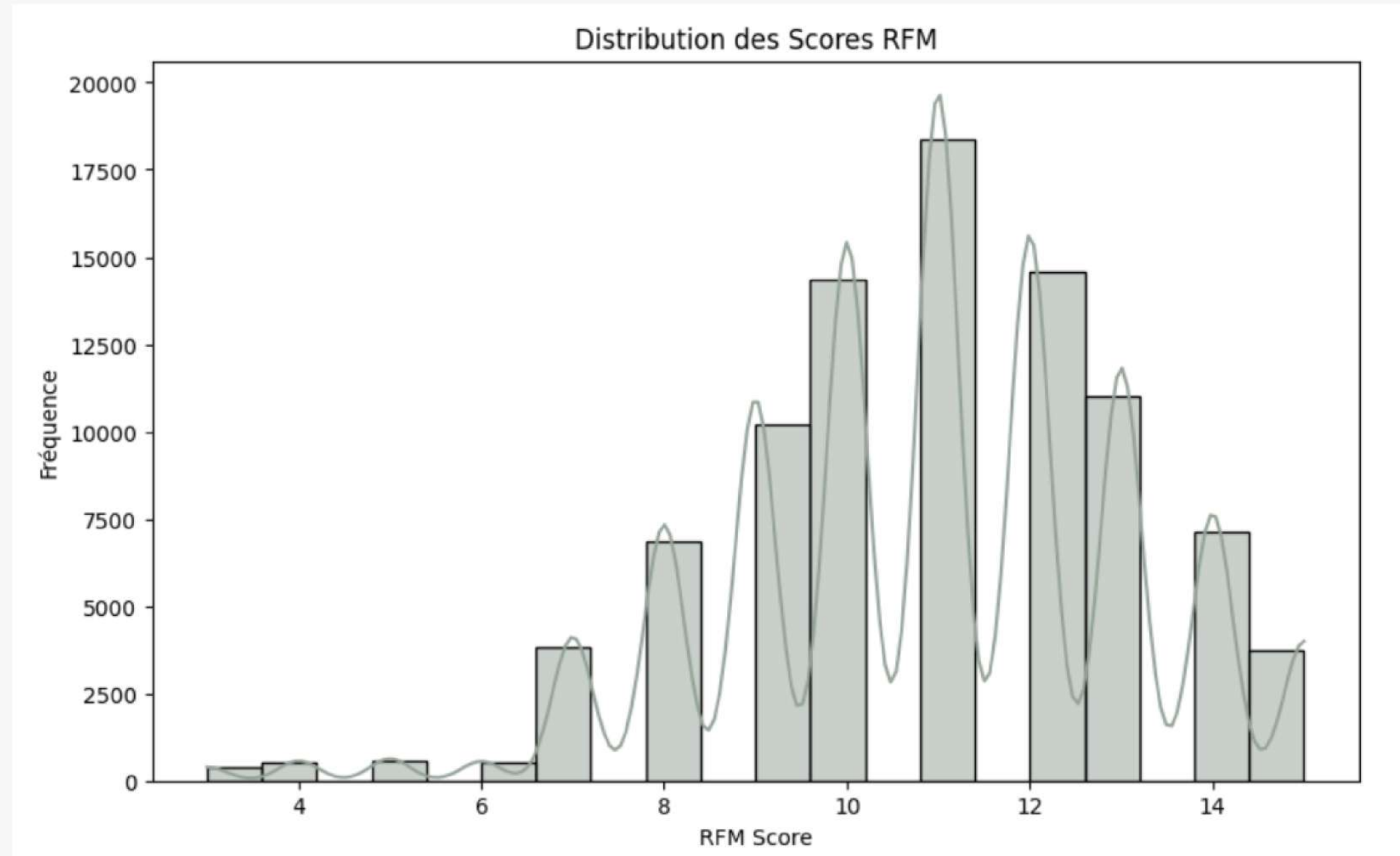


# Distribution score RFM





# Distribution score RFM



# Cohorte de commande



# Centre commercial



# Dataset

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

1

5 colonnes

2

200 lignes

3

1 colonne categorical  
(category)

4

6 colonnes numerical  
(int32 & int 64)

5

Dataset propre

# Dataset

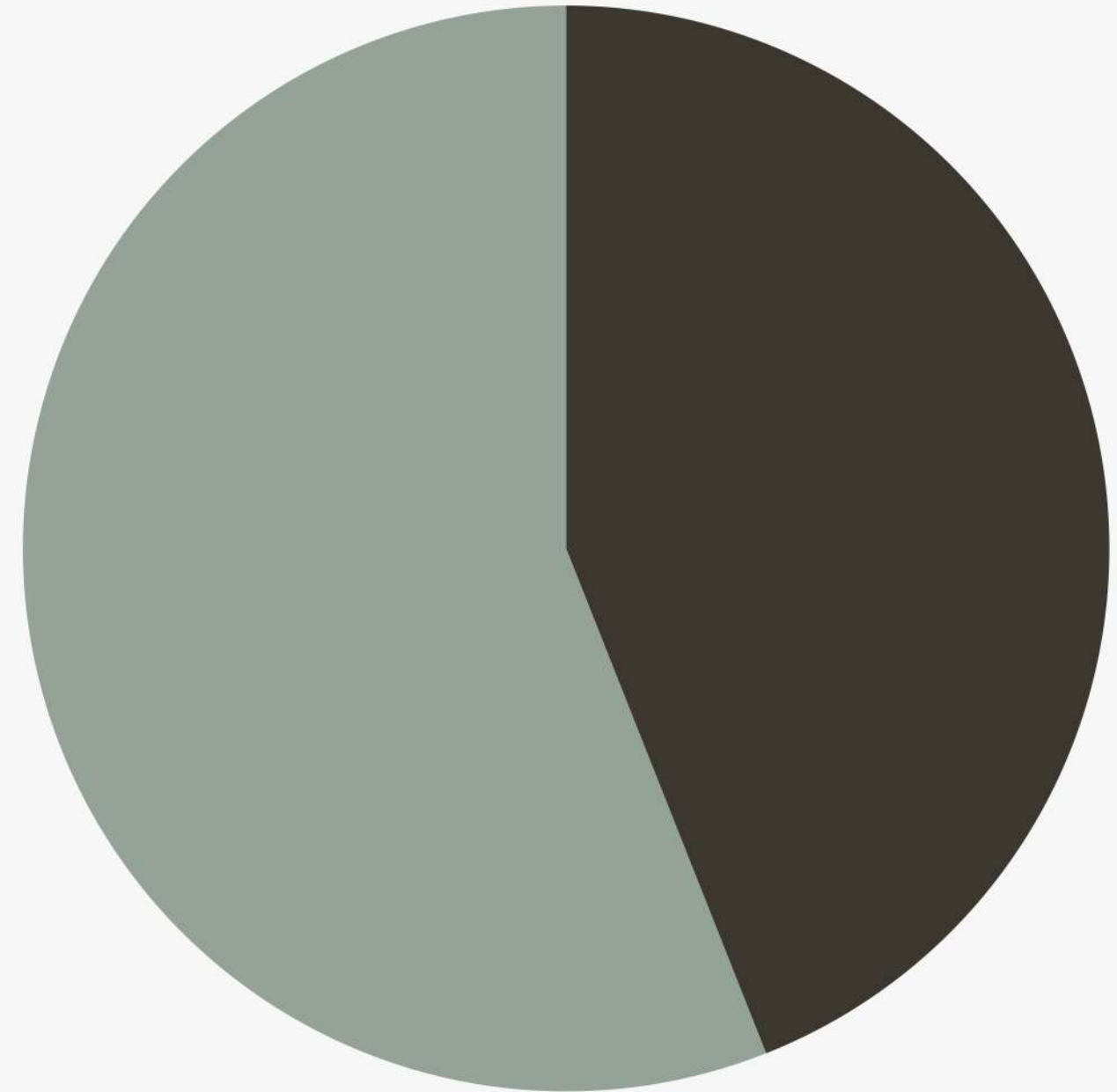
	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	0	19	15	39
1	2	0	21	15	81
2	3	1	20	16	6
3	4	1	23	16	77
4	5	1	31	17	40
...	...	...	...	...	...
195	196	1	35	120	79
196	197	1	45	126	28
197	198	0	32	126	74
198	199	0	32	137	18
199	200	0	30	137	83

1

Encodage de la colonne Gender  
1 pour femmes et 0 pour hommes

# Répartition des clients par sexe

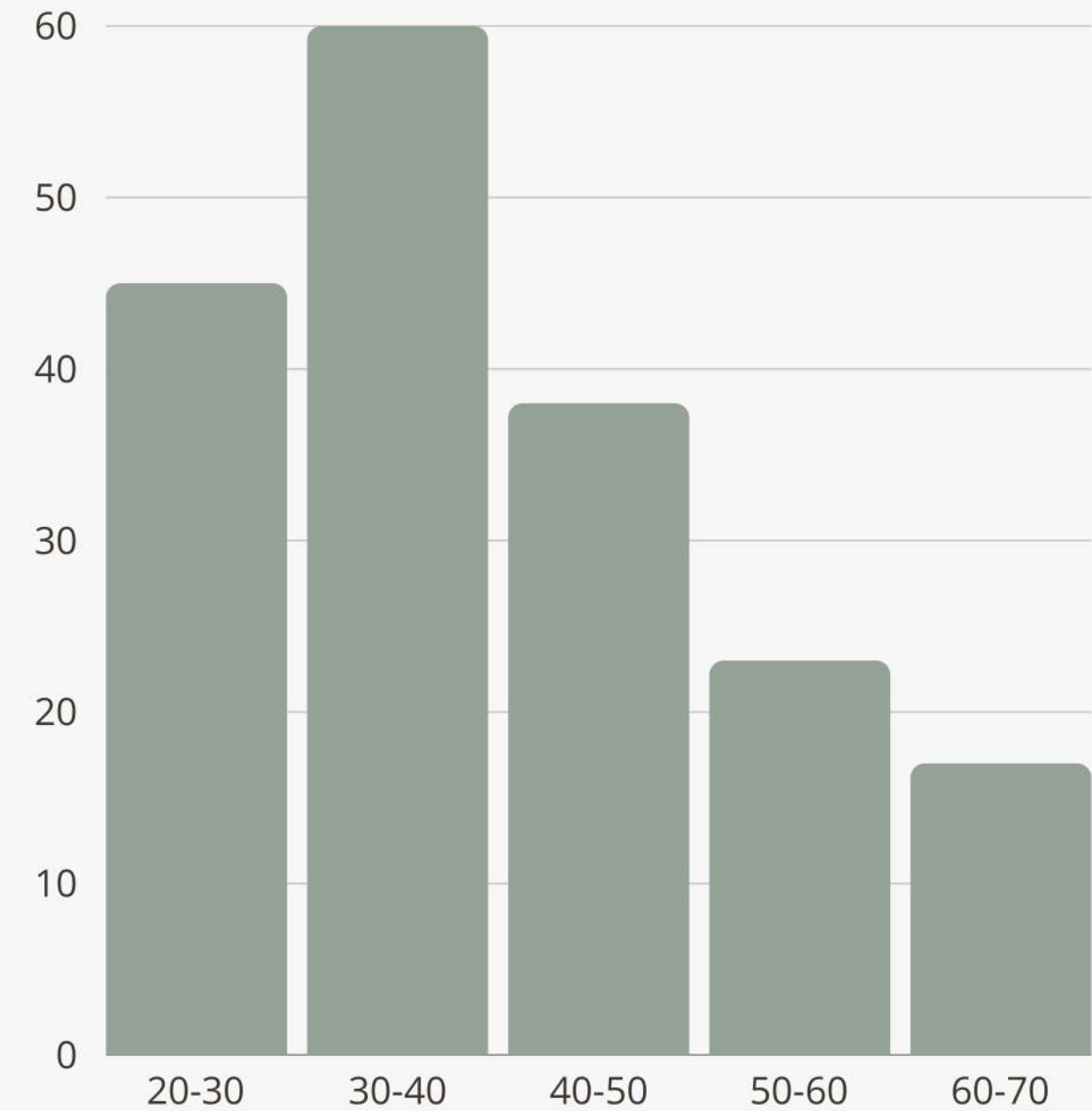
On constate qu'il y a légèrement plus d'acheteuse que d'acheteur (différence de 12%)





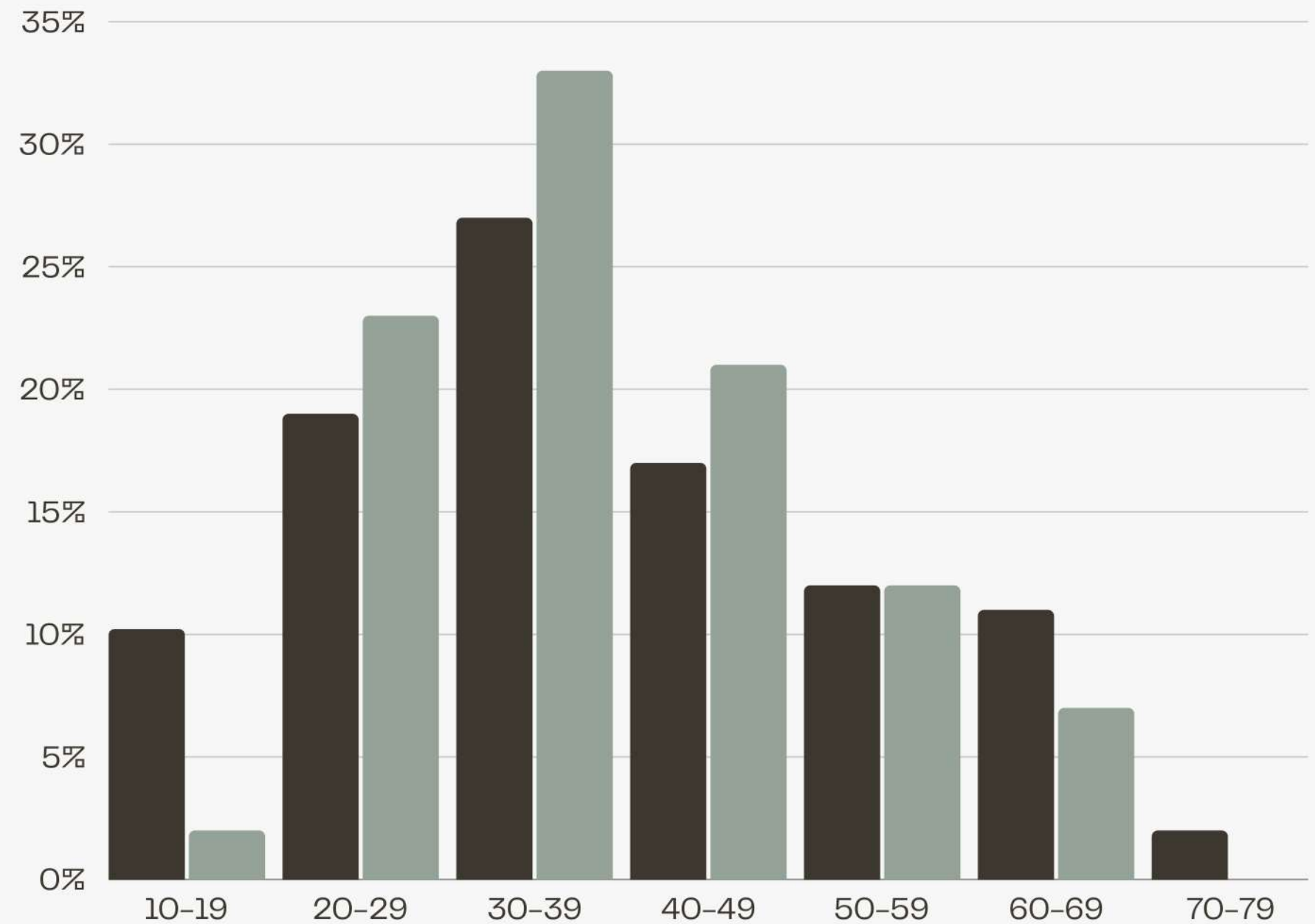
# Nombre de clients par tranche d'âge

- On observe que les acheteurs principaux se situe dans la tranche d'âge 30 – 40 ans.

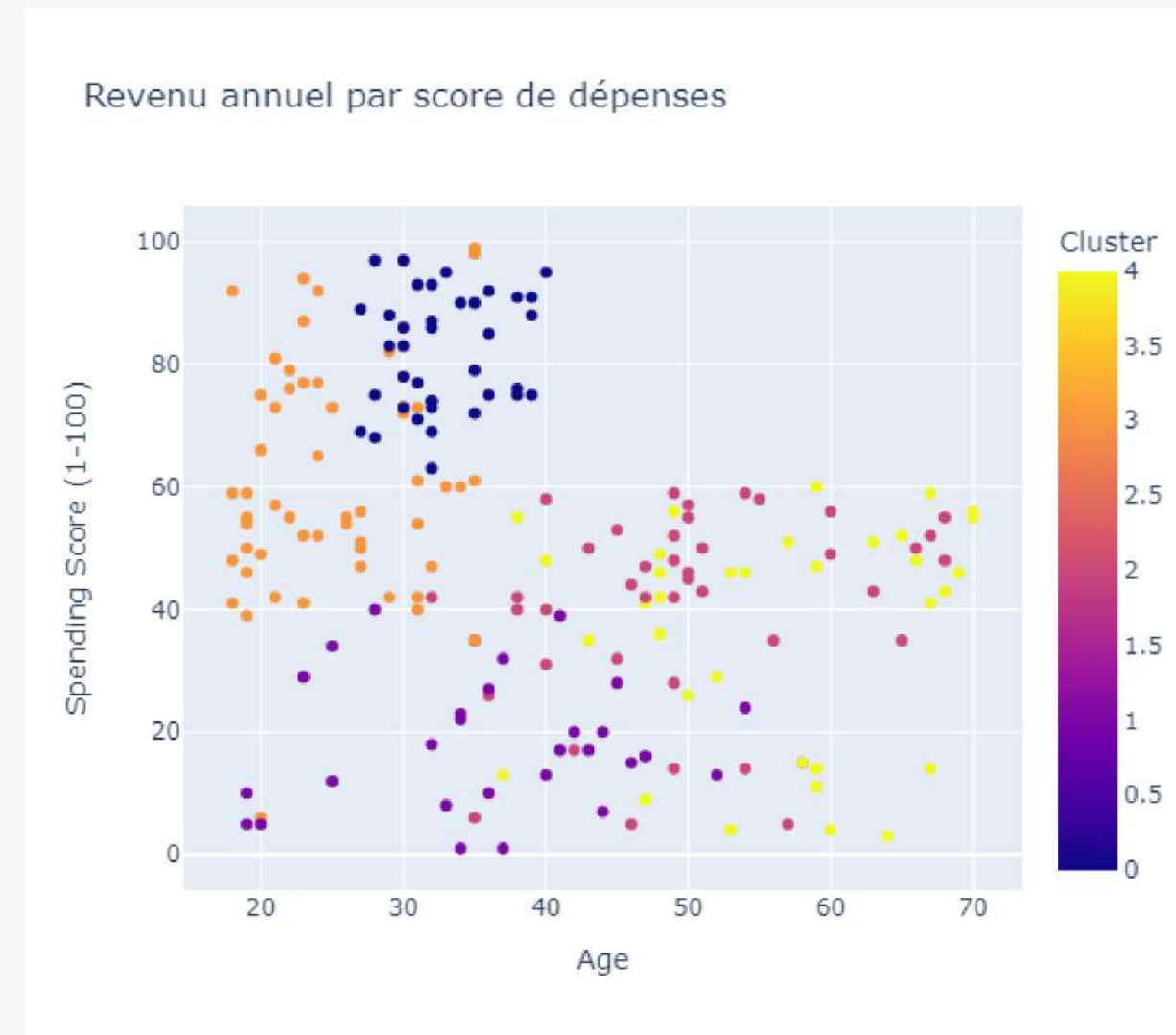
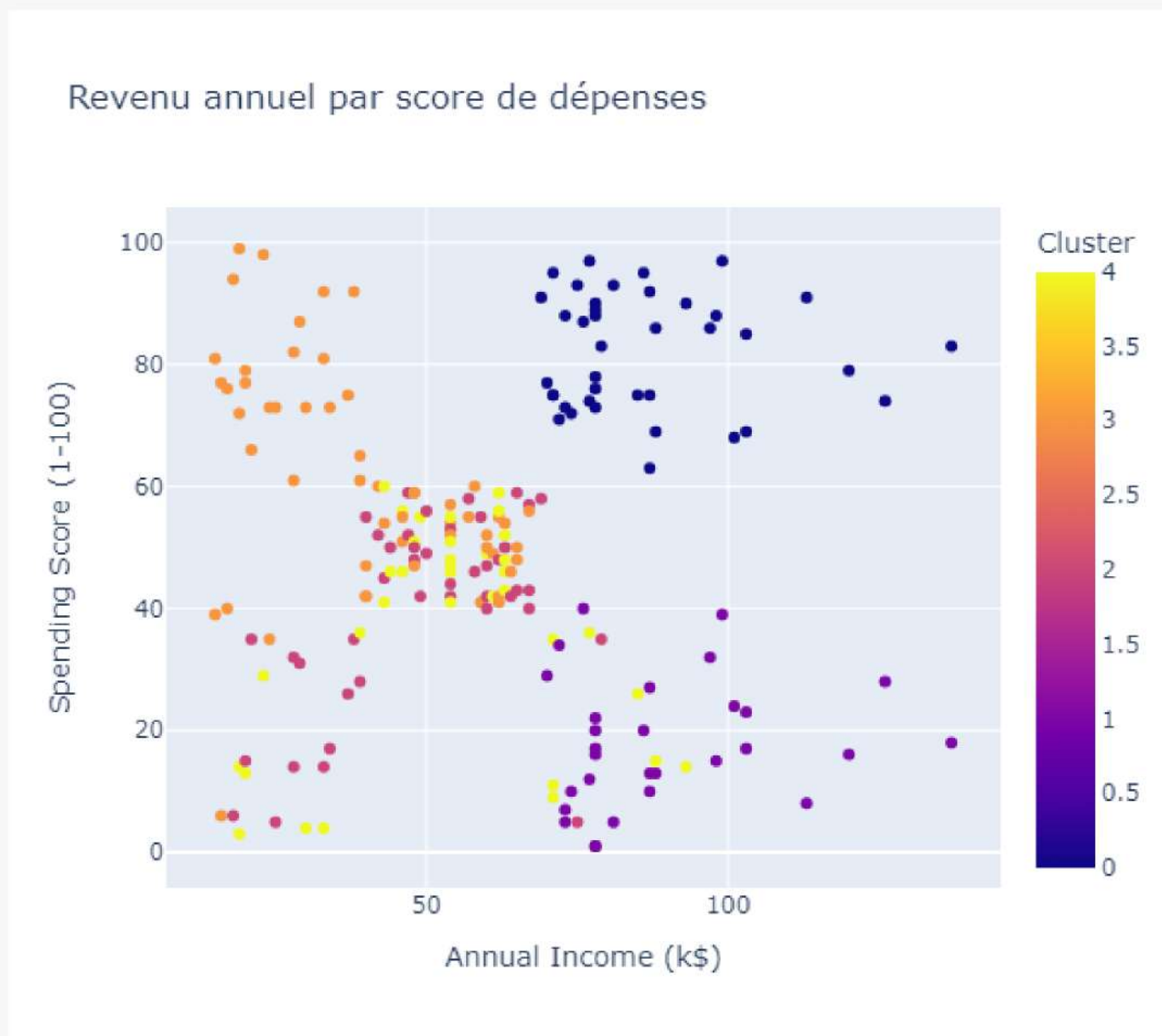


# Distribution des âges des clients en fonction du genre

- La tranche d'âge qui commande le plus est la 30 – 39 ans
- On constate que les femmes achètent globalement plus que les hommes



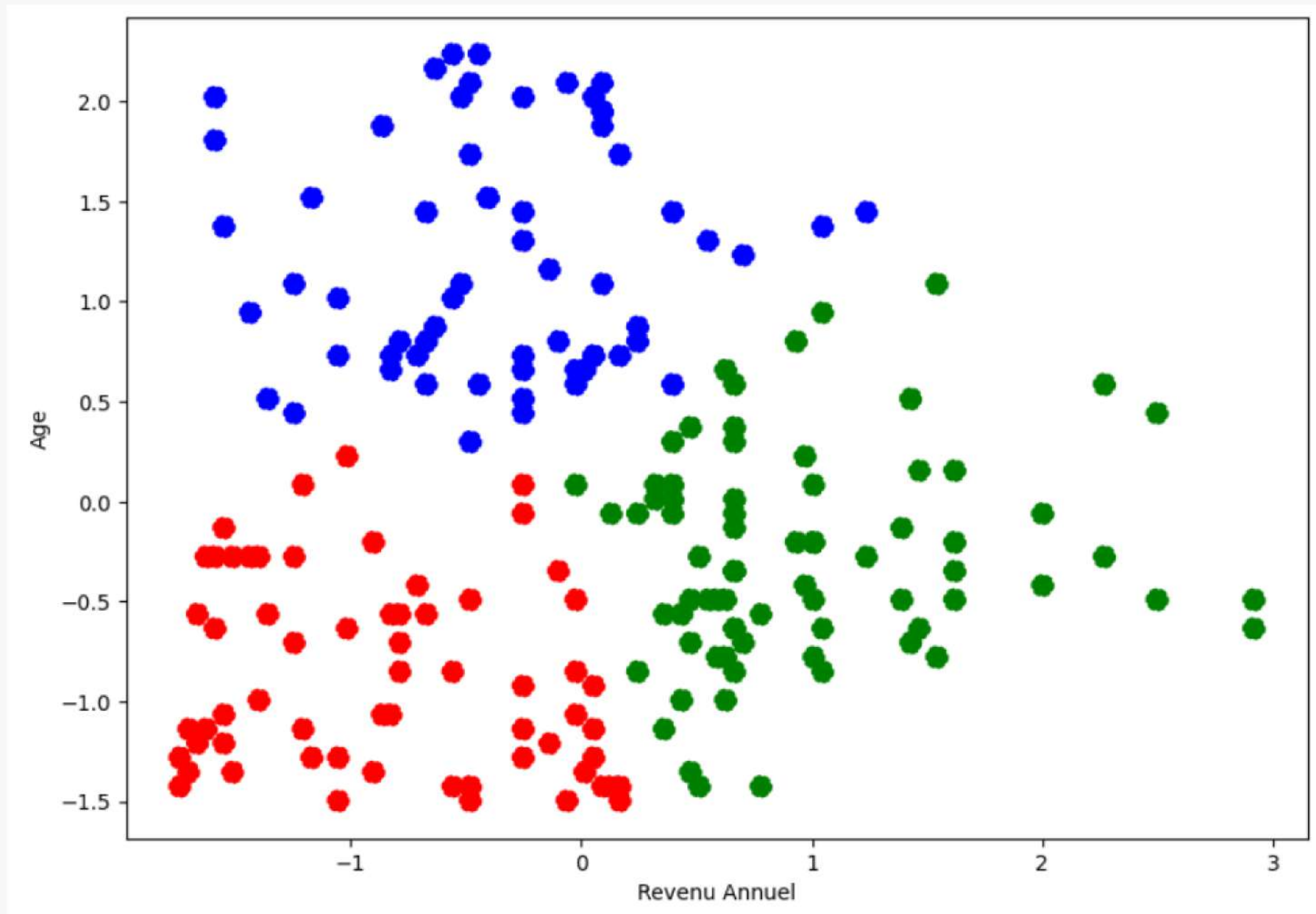
# Constat par visualisation



- On constate **cinq clusters** quand on regarde "**Revenu annuel par score de dépenses**" qui pourraient être notre nombre de clusters idéals.
- On constate également deux clusters légers si on regarde "Âge par score de dépenses" ceci dit, ils sont trop éparpillés pour être réellement utilisable.
- On en déduit donc que nos deux features les plus importantes sont "**Annual Income (k\$)**" & "**Spending Score (1-100)**"



# Constat par visualisation



- On constate **trois clusters** quand on regarde **"Revenu annuel par Age"** qui pourraient être notre nombre de clusters idéals.

1

K-Means

Silhouette score

0.55

2

DBSCAN

Silhouette score

0.11

3

Affinity  
Propagation

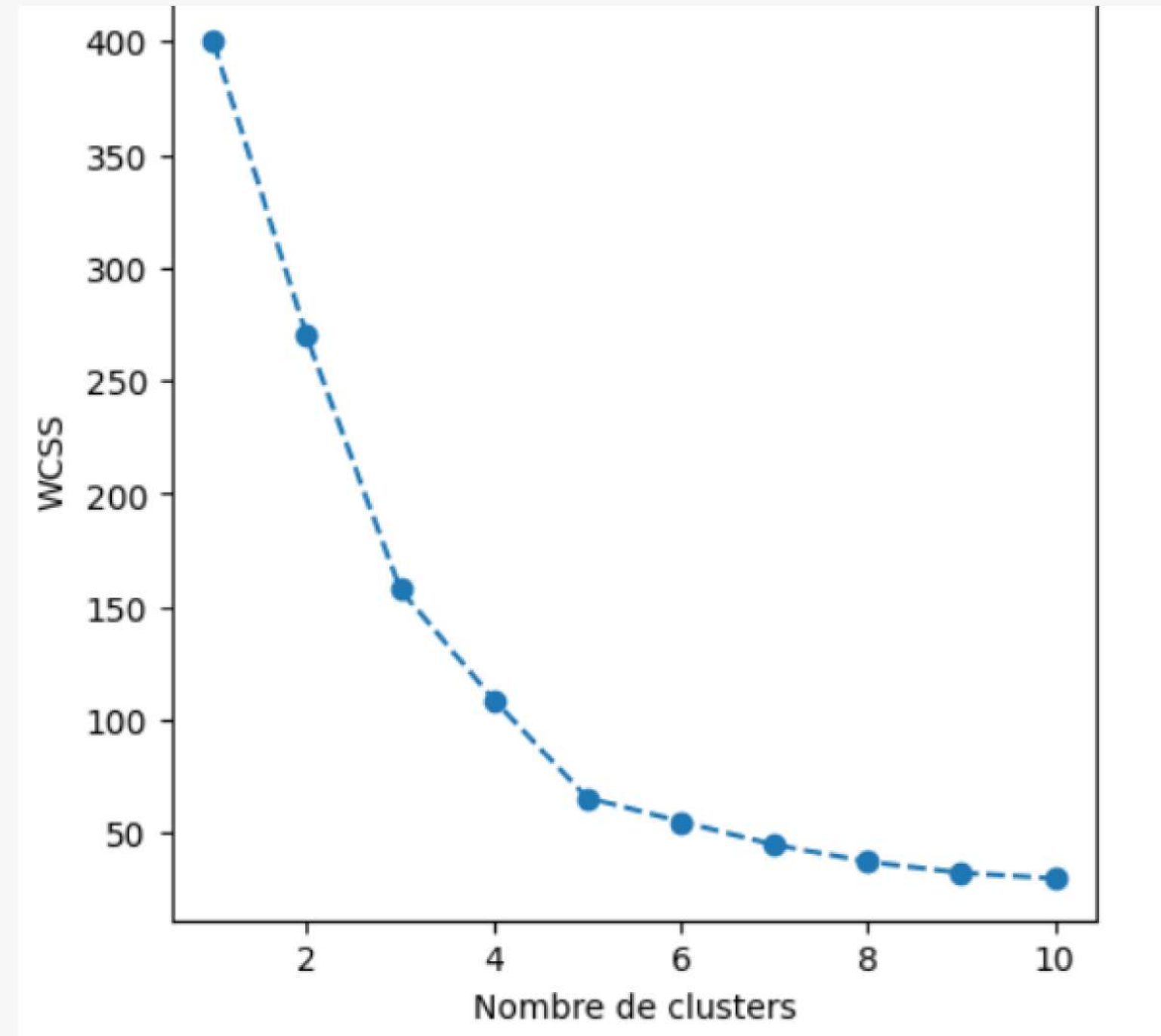
Silhouette score

0.43

“

## Elbow method

- On constate que le nombre idéal de cluster est 5 car on voit une cassure dans la courbe qui amorce une descente plus lente.
- Cela confirme ce qu'on à observé précédemment avec le graphique “Revenu annuel par score de dépenses”



# k-means & cluster

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
Cluster					
0	162.000000	0.538462	32.692308	86.538462	82.128205
1	165.482759	0.551724	36.482759	89.517241	18.000000
2	73.651163	1.000000	49.813953	49.232558	40.069767
3	52.814815	0.592593	24.907407	39.722222	61.203704
4	84.685714	0.000000	55.714286	53.685714	36.771429



Cluster	Proportion de femmes	Âge moyen	Revenu annuel moyen (k\$)	Score de dépense moyen (1-100)	Interprétation
0	53.8%	32.7	86.5	82.1	Ce cluster semble représenter des clients jeunes avec un revenu annuel élevé et un score de dépense très élevé. Ils sont légèrement plus susceptibles d'être des femmes. Ils sont probablement des clients réguliers et dépensiers.
1	55.2%	36.5	89.5	18.0	Ce cluster représente des clients légèrement plus âgés avec un revenu annuel élevé mais avec un score de dépense très faible. Bien qu'ils aient les moyens financiers, ils ne dépensent pas beaucoup dans le centre commercial.
2	100%	49.8	49.2	40.1	Ce cluster est exclusivement composé de femmes plus âgées avec un revenu moyen et un score de dépense modéré. Ce sont des clientes avec des revenus modestes mais qui dépensent de manière raisonnable.
3	59.3%	24.9	39.7	61.2	Ce cluster représente des clients très jeunes avec un revenu faible à moyen mais avec un score de dépense assez élevé. Ce sont principalement des jeunes adultes qui aiment dépenser même avec des revenus limités.
4	0% (uniquement des hommes)	55.7	53.7	36.8	Ce cluster est composé exclusivement d'hommes plus âgés avec un revenu moyen et un score de dépense modéré. Ces clients ont des revenus modestes et une propension à dépenser de façon modérée.



## Cluster 0 : Jeunes avec revenus élevés et gros dépensiers

- Campagne de fidélisation : Offrez des programmes de fidélité exclusifs avec des récompenses attractives pour encourager les visites fréquentes et les dépenses supplémentaires.
- Événements exclusifs : Organisez des événements VIP, des avant-premières ou des ventes privées pour ces clients.
- Produits haut de gamme : Mettez en avant des produits de luxe ou des marques premium qui correspondent à leur pouvoir d'achat élevé.

## Cluster 1 : Adultes avec revenus élevés mais faibles dépensiers

- Incitations à l'achat : Proposez des réductions ciblées ou des offres promotionnelles limitées dans le temps pour inciter à l'achat.
- Communication personnalisée : Utilisez des campagnes d'emailing personnalisées pour mettre en avant des produits ou services qui pourraient les intéresser.
- Expérience client améliorée : Améliorez l'expérience en magasin avec des services premium (conciergerie, livraison à domicile) pour les inciter à dépenser plus.



## Cluster 2 : Femmes plus âgées avec revenus moyens et dépensent modérément

- Offres spéciales : Proposez des offres spéciales pour les produits du quotidien ou des produits pour la maison, qui peuvent être plus attrayants pour ce groupe.
- Ateliers et événements : Organisez des ateliers ou des événements (cuisine, jardinage, bien-être) qui peuvent les intéresser et les attirer dans le centre commercial.
- Programme de récompenses : Mettez en place un programme de récompenses qui offre des remises ou des cadeaux après un certain nombre d'achats.

## Cluster 3 : Jeunes adultes avec revenus faibles à moyens mais gros dépensiers

- Promotions flash : Proposez des promotions flash ou des soldes sur des articles tendance pour attirer leur attention.
- Activités sociales : Créez des événements sociaux ou des rassemblements, comme des soirées à thème, qui attirent les jeunes adultes.
- Partenariats : Collaborez avec des marques populaires parmi les jeunes adultes pour des promotions conjointes.

## Cluster 4 : Hommes plus âgés avec revenus moyens et dépensent modérément

- Offres sur des produits spécifiques : Proposez des réductions sur des produits typiquement masculins, comme les gadgets technologiques, les articles de sport ou les équipements de bricolage.
- Activités et ateliers : Organisez des ateliers ou des démonstrations sur des sujets comme le bricolage, la technologie ou les voitures, qui peuvent attirer ce groupe.
- Marketing ciblé : Utilisez des campagnes publicitaires ciblées via des canaux comme les magazines spécialisés ou les réseaux sociaux pour hommes.



# Stratégie générale

1

**Segmentation des campagnes :**  
Utilisez les données de segmentation pour envoyer des messages marketing personnalisés à chaque cluster via email, SMS ou applications mobiles.

2

**Suivi et ajustement :**  
Suivez les performances de chaque campagne et ajustez-les en fonction des résultats pour maximiser l'efficacité.

3

**Feedback des clients :**  
Sollicitez les retours des clients pour mieux comprendre leurs besoins et ajuster vos offres et services en conséquence.

# Campagne marketing

Pour un super marcher

# Dataset

	ID	Year_Birth	Education	Marital_Status	Income	Kid
0	5524	1957	Graduation	Single	58138.0	
1	2174	1954	Graduation	Single	46344.0	
2	4141	1965	Graduation	Together	71613.0	
3	6182	1984	Graduation	Together	26646.0	
4	5324	1981	PhD	Married	58293.0	
...	...	...	...	...	...	...
2235	10870	1967	Graduation	Married	61223.0	
2236	4001	1946	PhD	Together	64014.0	
2237	7270	1981	Graduation	Divorced	56981.0	
2238	8235	1956	Master	Together	69245.0	
2239	9405	1954	PhD	Married	52869.0	

1

9 colonnes

2

2240 lignes

3

3 colonnes object

4

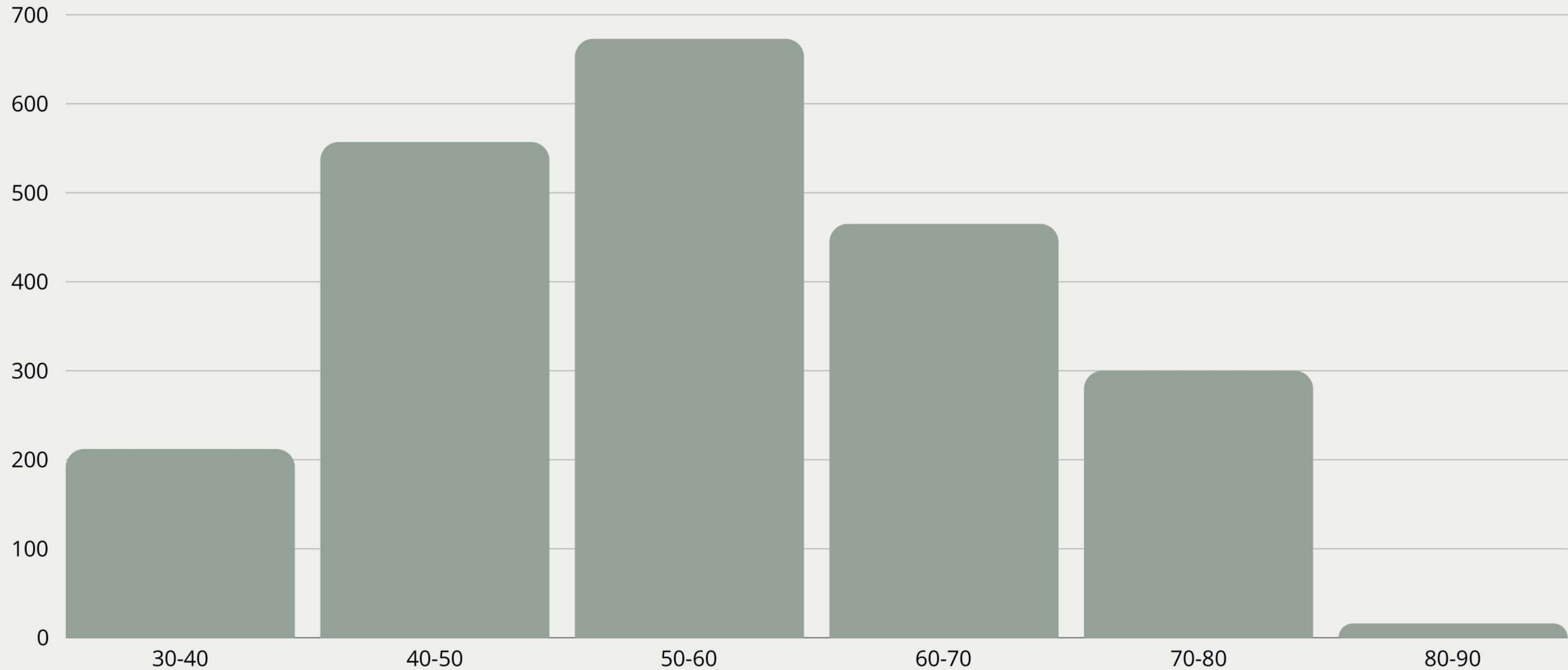
26 colonnes numérique (float64 & int64)

5

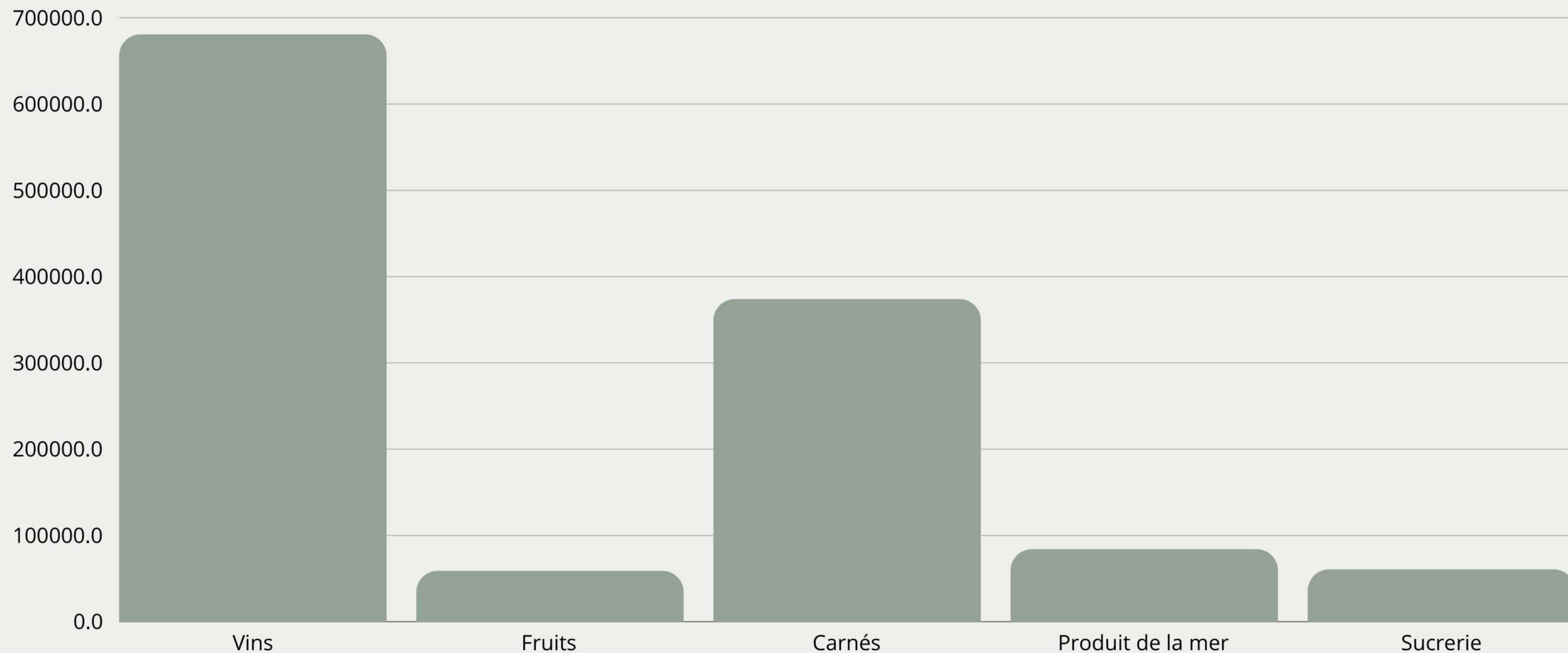
Dataset à optimiser



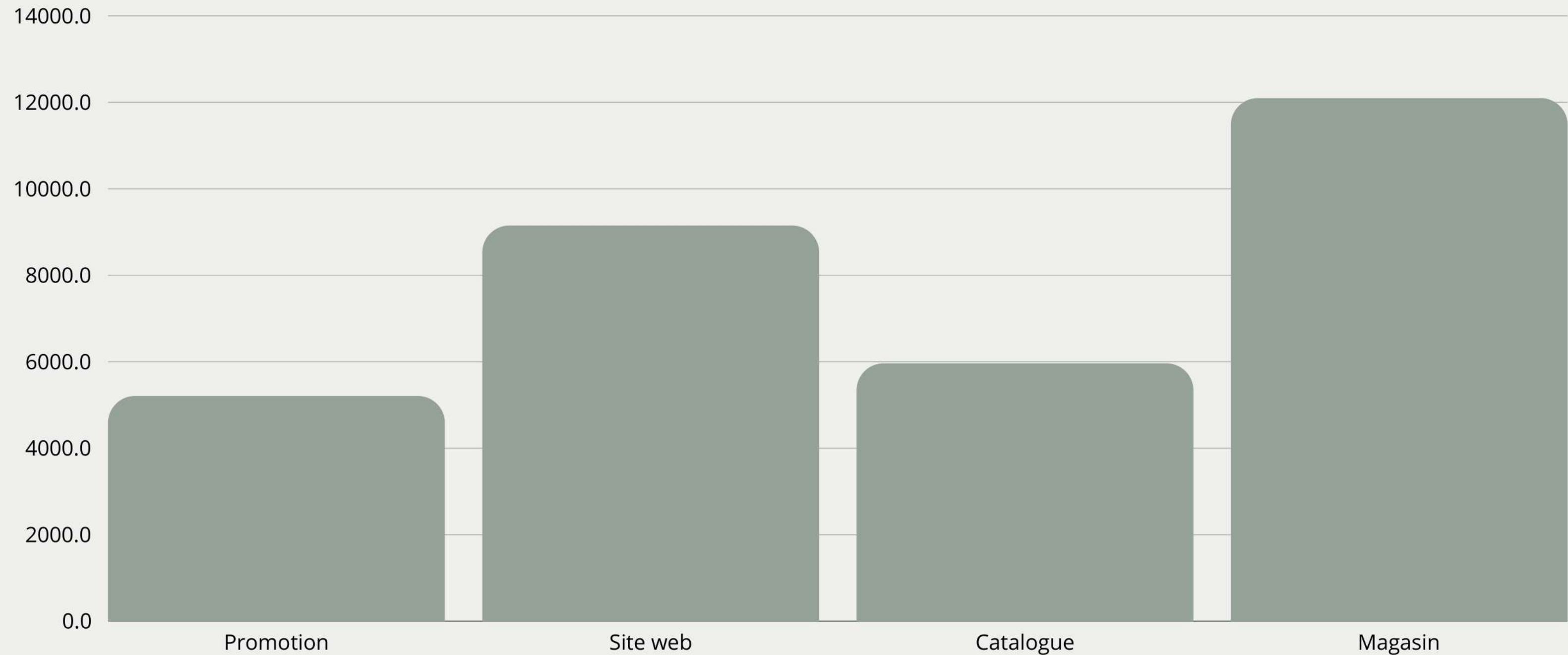
# Répartition des âges



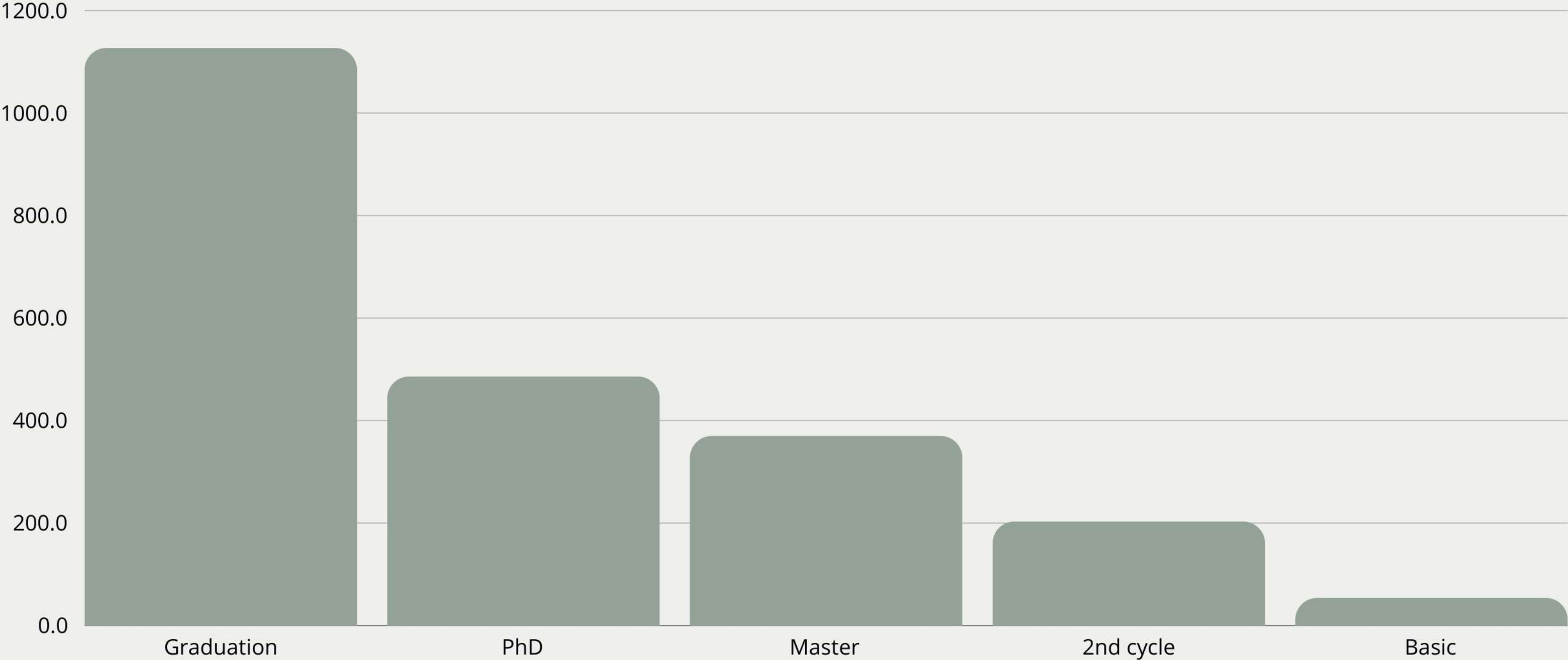
# Montant dépensé par catégorie de produit



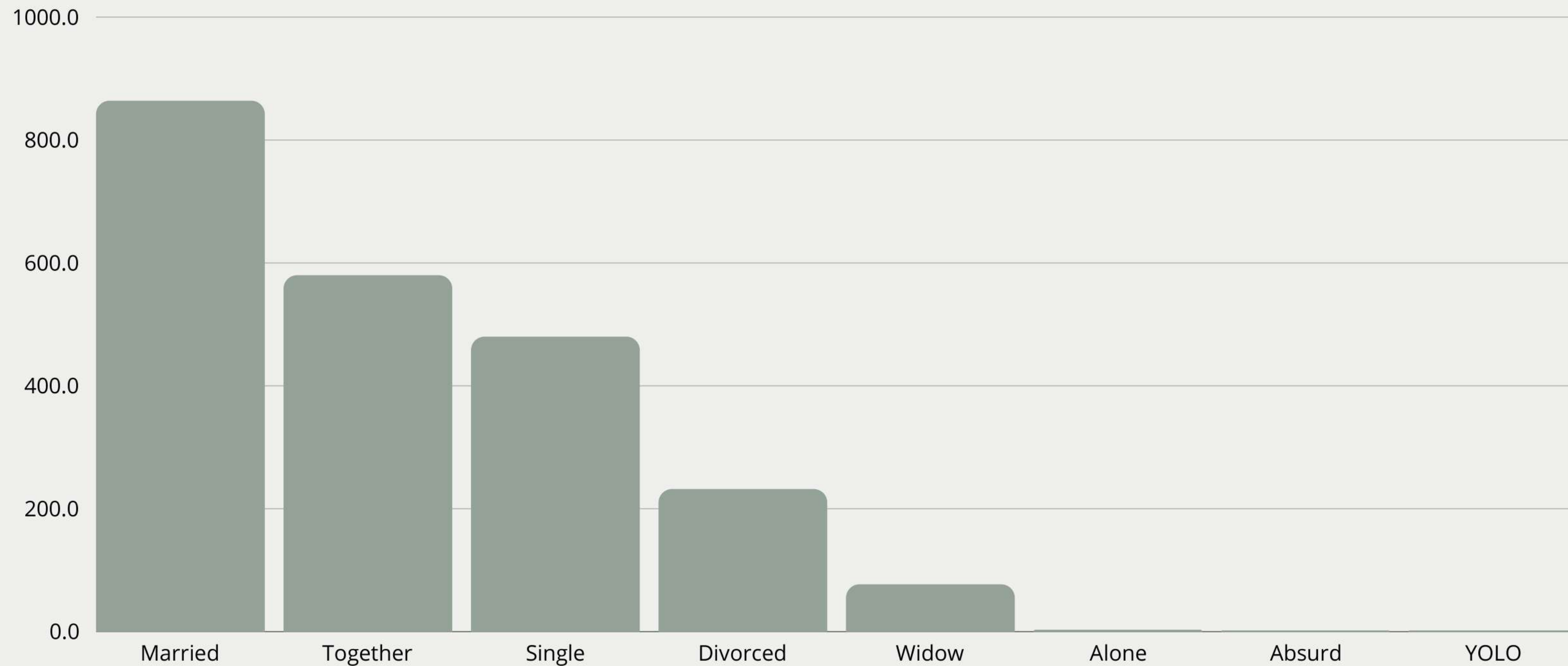
# Répartition des ventes par canal de vente



# Répartition de Education



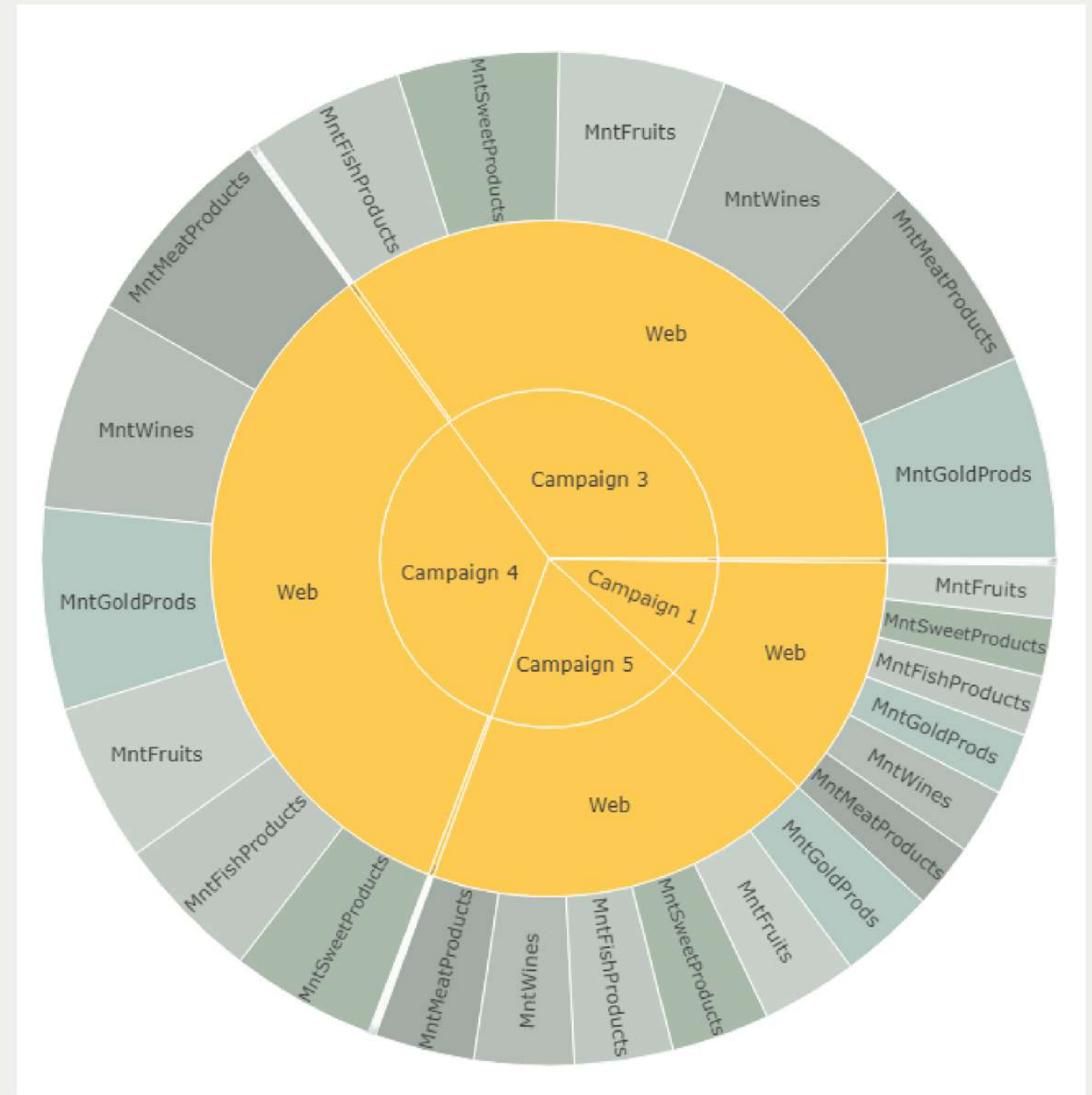
# Répartition des status matrimoniaux





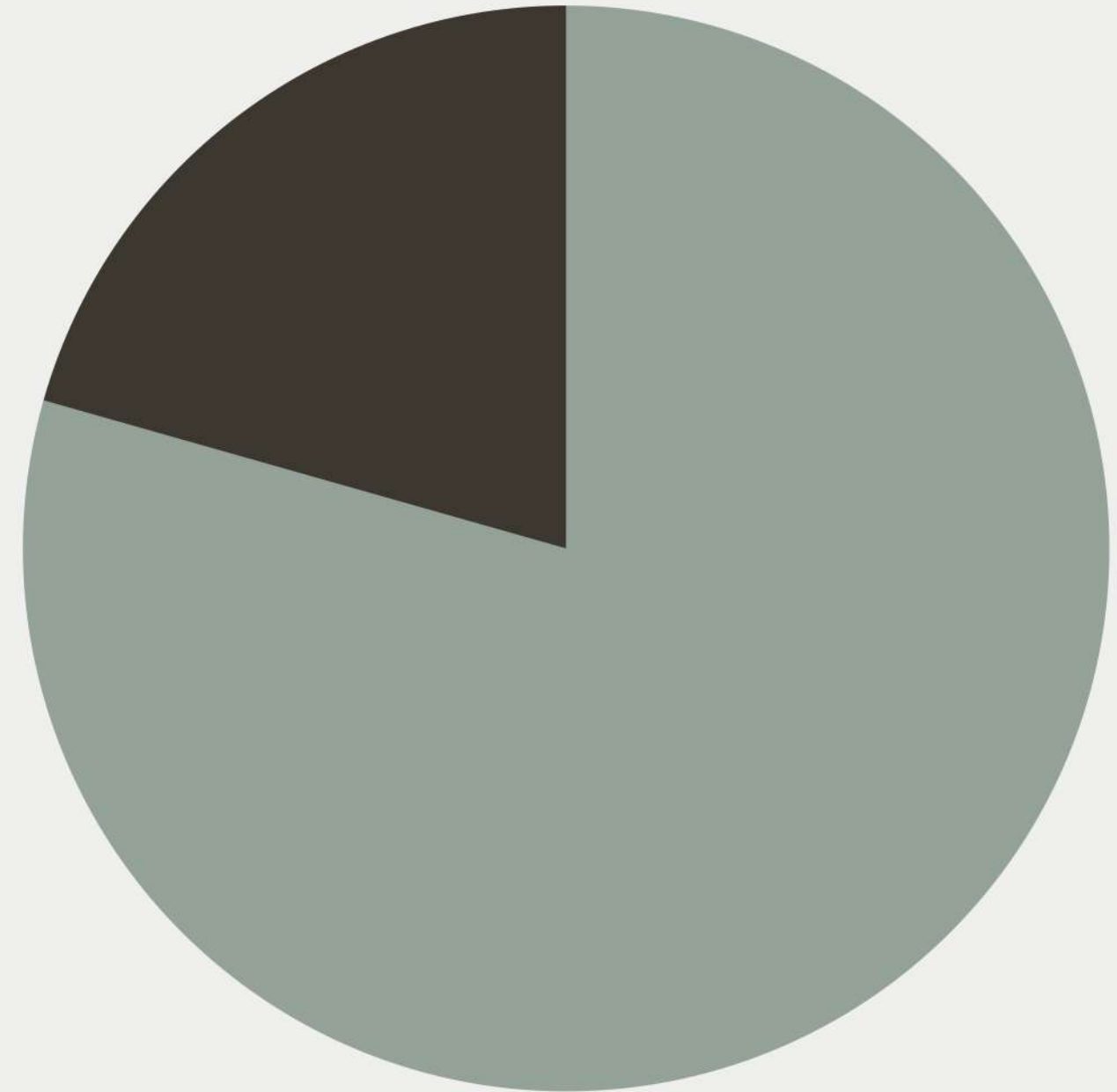
# Répartition des ventes de produits en fonction des campagnes

- On observe qu'à chaque campagne, la répartition des ventes est assez homogène
- On observe également que les campagnes ont excellé via le web
- Cependant, la campagne 2 n'a pas performé puisqu'aucun ventes ne s'est produite lors de cette campagne

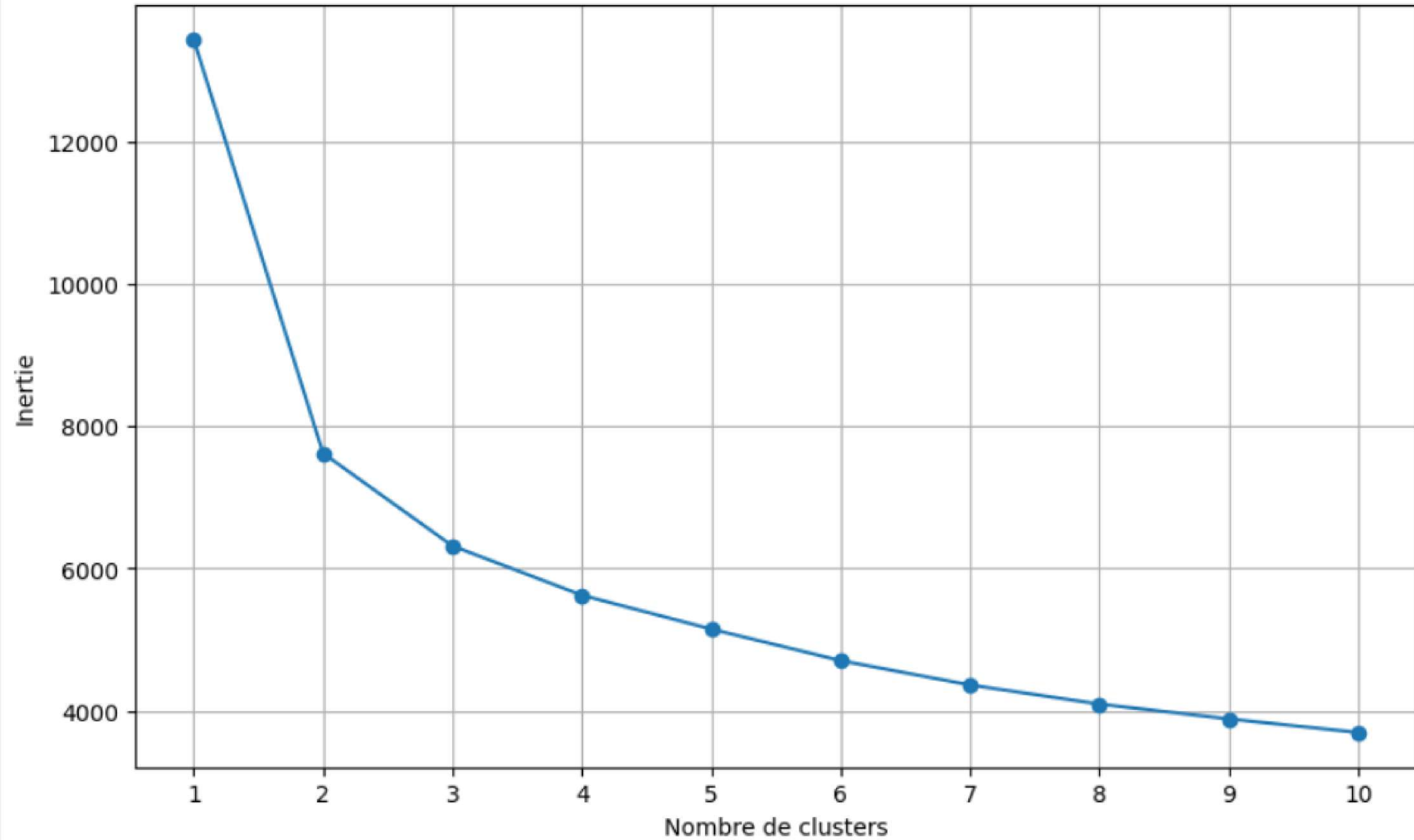


# Receptif ou non aux campagnes

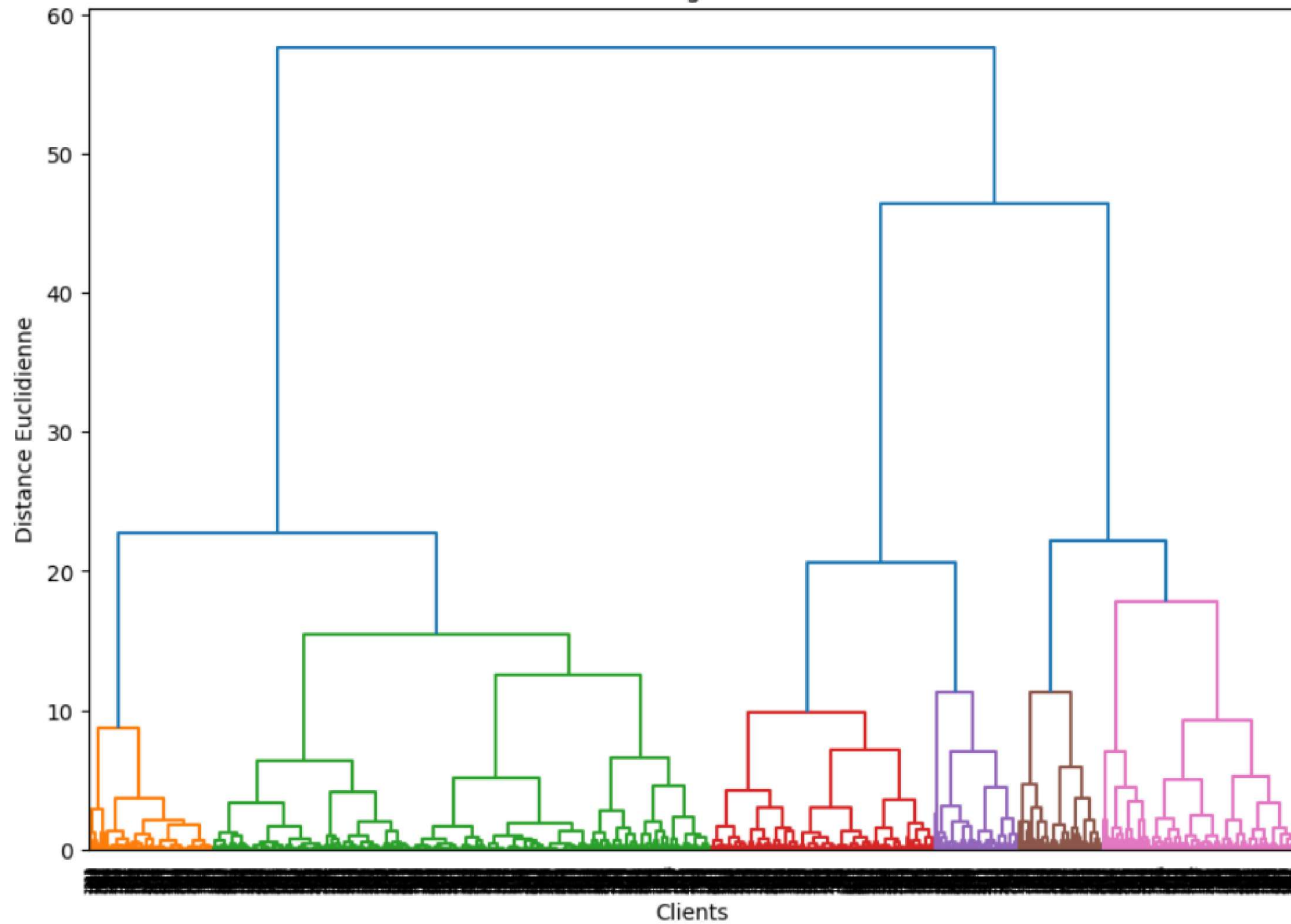
Globalement, plus des trois quarts des clients sont réceptifs aux campagnes marketing.



Méthode du Coude

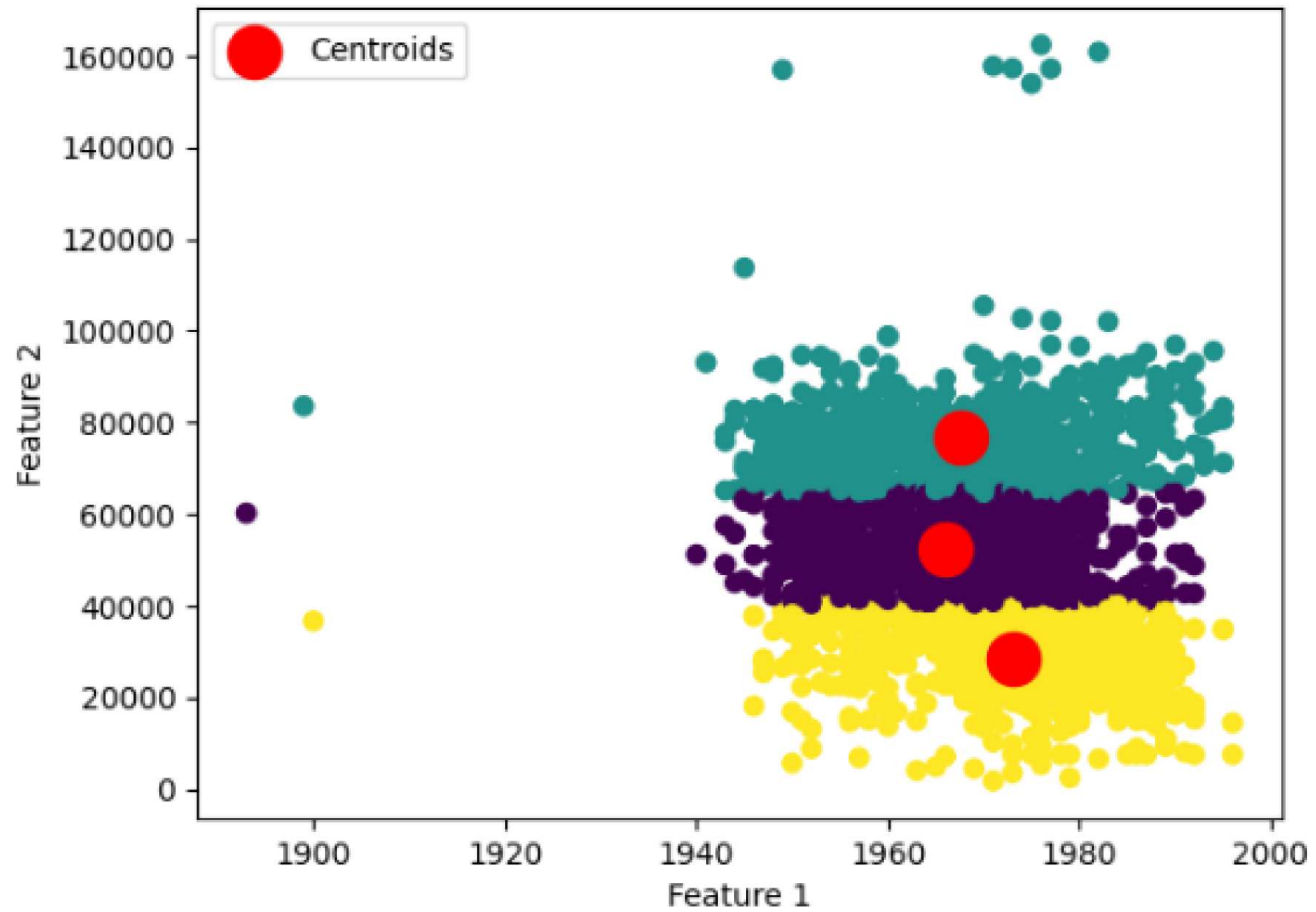


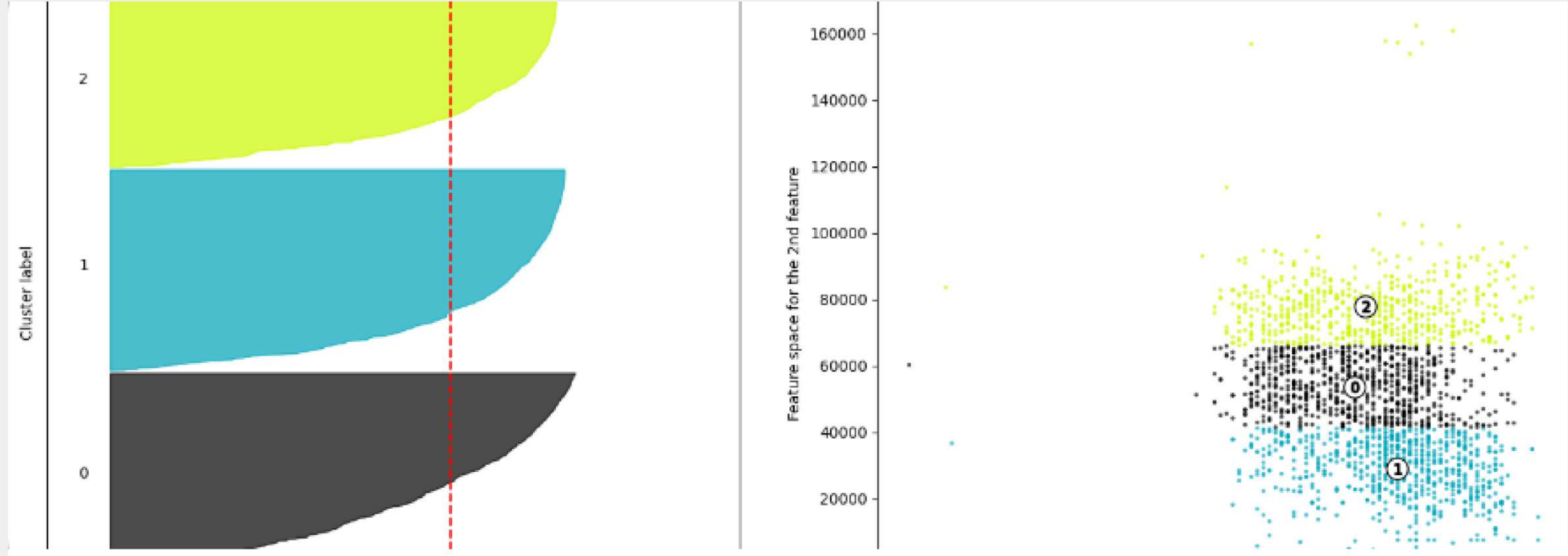
Dendrogramme





Clusters et Centroids





	Age	MntWines	Cluster
0	67	635	2
1	70	11	2
2	59	426	1
3	40	11	0
4	43	173	0

1

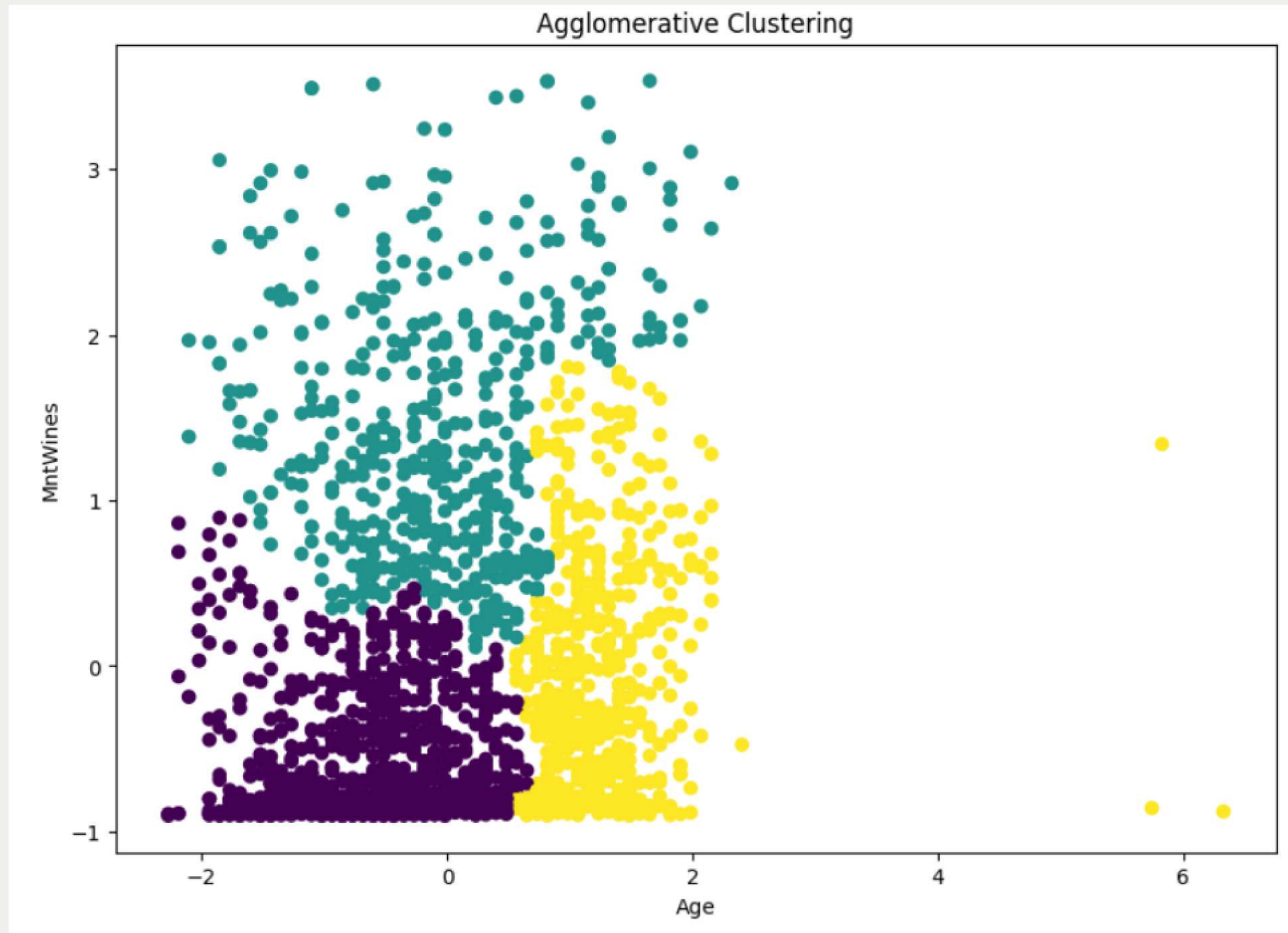
## Affinity Propagation

Silhouette score pour 3 clusters

0.43

Inertia

3696.7





Moyennes des montants de produits par cluster  
MntMeatProducts

