

# ANALYSE POUR LA SEGMENTATION DE LA CLIENTELE

Présenté par

RAKOTOZAFY Andriniaina Fanomezana

Parcours DATA ANALYSTE



# PLAN

- <sup>1.</sup> MISE EN SITUATION
- <sup>2.</sup> PROBLEMES / OPPORTUNITES
- <sup>3.</sup> QUESTIONS/HYPOTHESES
- <sup>4.</sup> REPONSES ET SOLUTIONS (résultats et recommandations)

# CONTEXTE

La segmentation est une méthode fondamentale et largement employée dans de nombreux domaines professionnels. Son objectif principal est de simplifier la compréhension des systèmes complexes en identifiant des sous-groupes homogènes.

## **Domaine des Études de Marché et du Marketing :**

- . Segmentation de marché**
- . Segmentation selon les préférences clients**
- . Segmentation RFM**
- . Segmentation selon la valeur client**

## **Sciences Sociales et la Recherche Académique :**

- . Segmentation géographique**
- . Segmentation démographique**
- . Segmentation comportementale**
- . Segmentation psychographique**

## **Domaine de la Médecine :**

- . Segmentation des patients selon le type de maladie**
- . Segmentation en fonction des caractéristiques génétiques**
- . Segmentation basée sur les facteurs de risque**
- . Segmentation en fonction des préférences de traitement**

Dans notre cas, on est présence d'une étude de segmentation de clients afin d'optimiser les approches, et campagnes marketing.

## Collecte de données

Rassemblements des informations sur les clients : comportements, identités, coordonnées



## Traitements et analyses

Utilisation des méthodes statistiques pour identifier des groupes de clients similaires, phase de segmentation



## Résultats et recommandations

A partir des segments identifiés, personnalisation des stratégies de marketing



# Objectif

Comprendre et analyser les données client sur : les besoins uniques des clients, leurs motivations et des intérêts afin de segmenter notre base de clients et d'identifier des groupes distincts. En comprenant mieux les différents segments de clients, nous pourrons adapter nos stratégies marketing pour répondre plus efficacement à leurs besoins et à leurs attentes.

À travers cette étude, nous visons à définir un cadre stratégique pour la segmentation des clients, jetant ainsi les bases d'une personnalisation ciblée des communications et des offres pour chaque valeur des clients.

# COMPREHENSION ET EXPLORATION DE DONNEES

## Description des variables

### Démographie / Contexte client:

YearBirth, Education, Marital\_Status, Kidhome, Teenhome, Income, ...

### Habitudes de dépenses:

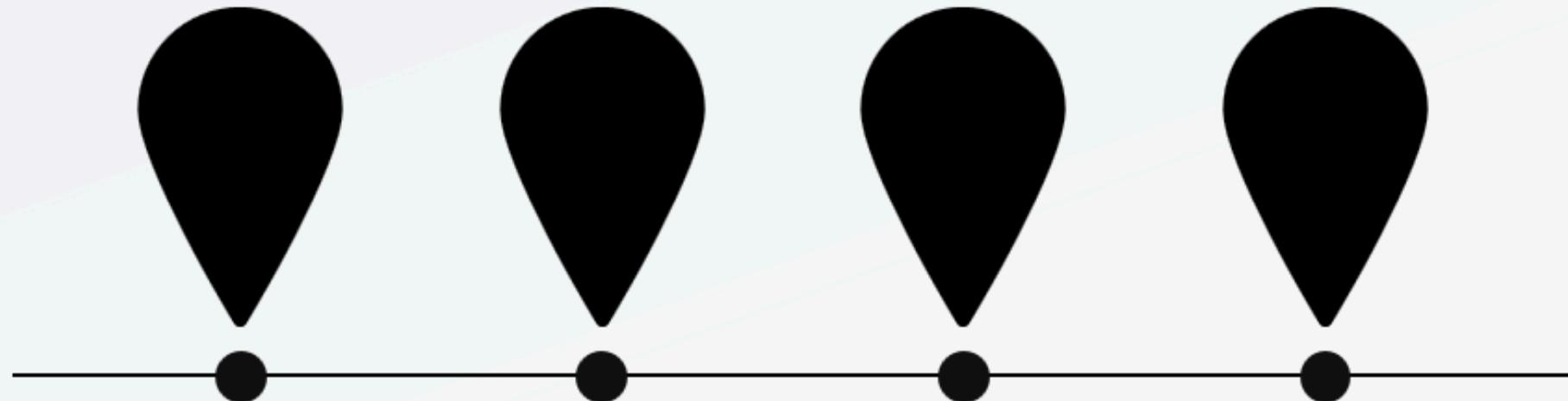
MntFishProducts, MntMeatProducts, MntFruits, MntSweetProducts, MntWines,  
MntGoldProds, NumDealsPurchases, NumCatalogPurchases,  
NumStorePurchases, NumWebPurchases

### Interactions avec les campagnes:

AcceptedCmp1 à 5, Response, Complain, NumWebVisitsMonth

# QUESTIONS

Comment pouvons-nous segmenter nos clients pour mieux cibler nos offres promotionnelles et améliorer leur expérience d'achat ?



Y a-t-il des problèmes avec les données que nous devons résoudre avant l'analyse ?

Quelles variables devrions-nous utiliser pour le regroupement ?

Quelles caractéristiques sont communes parmi les clients dans l'ensemble de données pour créer des groupes de clients ?

Quelle méthode de segmentation utilisée est optimale ?

## Description bref des variables

	count	mean	std	min	25%	50%	75%	max
ID	2240.0	5592.159821	3246.662198	0.0	2828.25	5458.5	8427.75	11191.0
Year_Birth	2240.0	1968.805804	11.984069	1893.0	1959.00	1970.0	1977.00	1996.0
Income	2216.0	52247.251354	25173.076661	1730.0	35303.00	51381.5	68522.00	666666.0
Kidhome	2240.0	0.444196	0.538398	0.0	0.00	0.0	1.00	2.0
Teenhome	2240.0	0.506250	0.544538	0.0	0.00	0.0	1.00	2.0
Recency	2240.0	49.109375	28.962453	0.0	24.00	49.0	74.00	99.0
MntWines	2240.0	303.935714	336.597393	0.0	23.75	173.5	504.25	1493.0
MntFruits	2240.0	26.302232	39.773434	0.0	1.00	8.0	33.00	199.0
MntMeatProducts	2240.0	166.950000	225.715373	0.0	16.00	67.0	232.00	1725.0
MntFishProducts	2240.0	37.525446	54.628979	0.0	3.00	12.0	50.00	259.0
MntSweetProducts	2240.0	27.062946	41.280498	0.0	1.00	8.0	33.00	263.0
MntGoldProds	2240.0	44.021875	52.167439	0.0	9.00	24.0	56.00	362.0
NumDealsPurchases	2240.0	2.325000	1.932238	0.0	1.00	2.0	3.00	15.0
NumWebPurchases	2240.0	4.084821	2.778714	0.0	2.00	4.0	6.00	27.0
NumCatalogPurchases	2240.0	2.662054	2.923101	0.0	0.00	2.0	4.00	28.0
NumStorePurchases	2240.0	5.790179	3.250958	0.0	3.00	5.0	8.00	13.0
NumWebVisitsMonth	2240.0	5.316518	2.426645	0.0	3.00	6.0	7.00	20.0
AcceptedCmp3	2240.0	0.072768	0.259813	0.0	0.00	0.0	0.00	1.0
AcceptedCmp4	2240.0	0.074554	0.262728	0.0	0.00	0.0	0.00	1.0
AcceptedCmp5	2240.0	0.072768	0.259813	0.0	0.00	0.0	0.00	1.0
AcceptedCmp1	2240.0	0.064286	0.245316	0.0	0.00	0.0	0.00	1.0
AcceptedCmp2	2240.0	0.012946	0.113069	0.0	0.00	0.0	0.00	1.0



# Y a-t-il des problèmes avec les données que nous devons résoudre avant l'analyse ?

```
[5]: df.isnull().sum() * 100 / len(df)
```

```
[5]:
```

ID	0.000000
Year_Birth	0.000000
Education	0.000000
Marital_Status	0.000000
Income	1.071429
Kidhome	0.000000
Teenhome	0.000000
Dt_Customer	0.000000
Recency	0.000000
MntWines	0.000000
MntFruits	0.000000
MntMeatProducts	0.000000
MntFishProducts	0.000000
MntSweetProducts	0.000000
MntGoldProds	0.000000
NumDealsPurchases	0.000000
NumWebPurchases	0.000000
NumCatalogPurchases	0.000000
NumStorePurchases	0.000000
NumWebVisitsMonth	0.000000
AcceptedCmp3	0.000000
AcceptedCmp4	0.000000
AcceptedCmp5	0.000000
AcceptedCmp1	0.000000
AcceptedCmp2	0.000000
Complain	0.000000
Response	0.000000

```
dtype: float64
```

```
[9]: # Analyse approfondie Education
```

```
Education_counts = df['Education'].value_counts()
```

```
# Afficher le nombre de personnes pour chacune des valeurs uniques
```

```
print(Education_counts)
```

Graduation	1127
PhD	486
Master	370
2n Cycle	203
Basic	54

```
Name: Education, dtype: int64
```

```
[11]: # Analyse approfondie Marital_Status
```

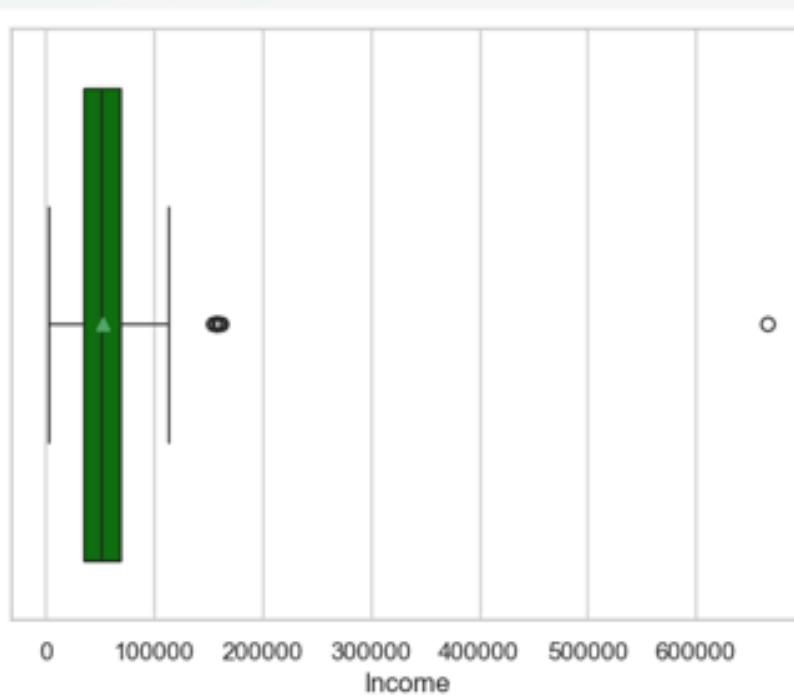
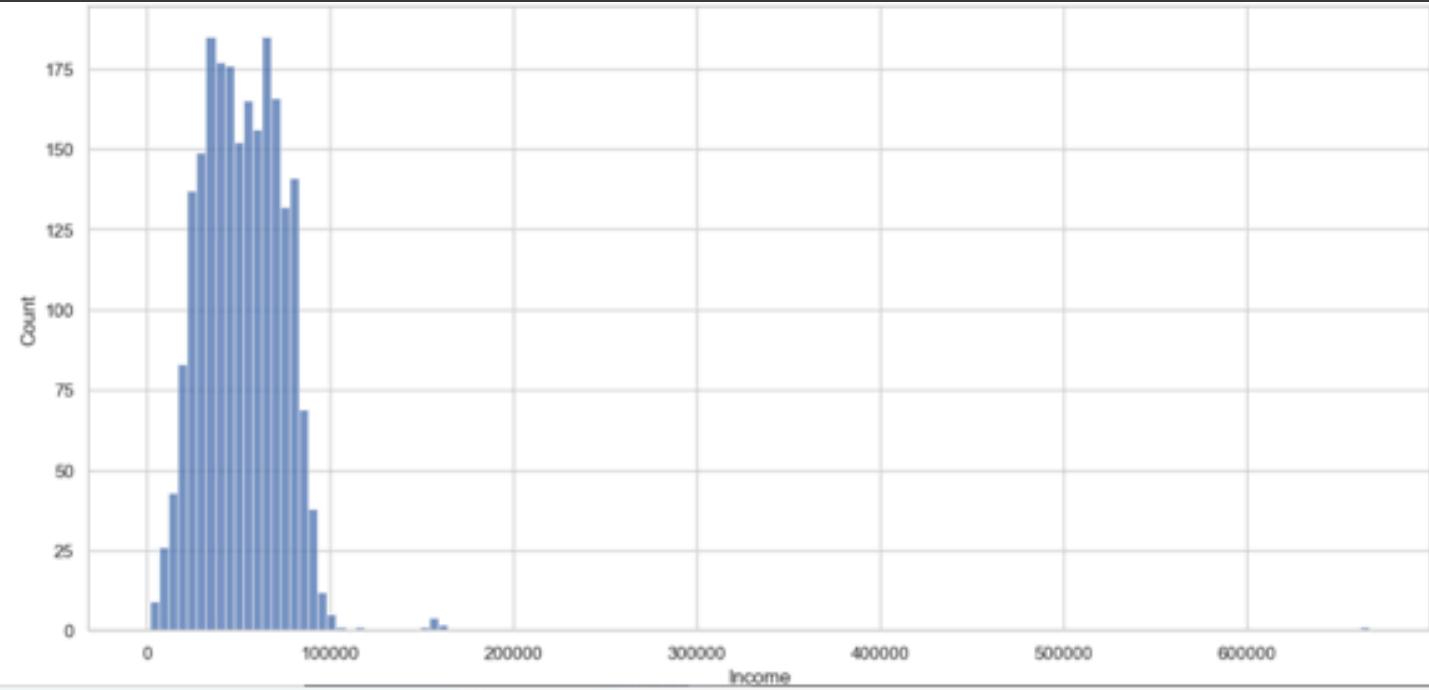
```
Marital_Status_counts = df['Marital_Status'].value_counts()
```

```
# Afficher le nombre de personnes pour chacune des valeurs uniques
```

```
print(Marital_Status_counts)
```

Married	864
Together	580
Single	480
Divorced	232
Widow	77
Alone	3
Absurd	2
YOLO	2

```
Name: Marital_Status, dtype: int64
```





Quelles caractéristiques sont communes parmi les clients dans l'ensemble de données pour créer des groupes de clients ?



l'âge ?

Habitude d'achat?

La fréquence d'achat?

La réactivité aux campagnes marketing?

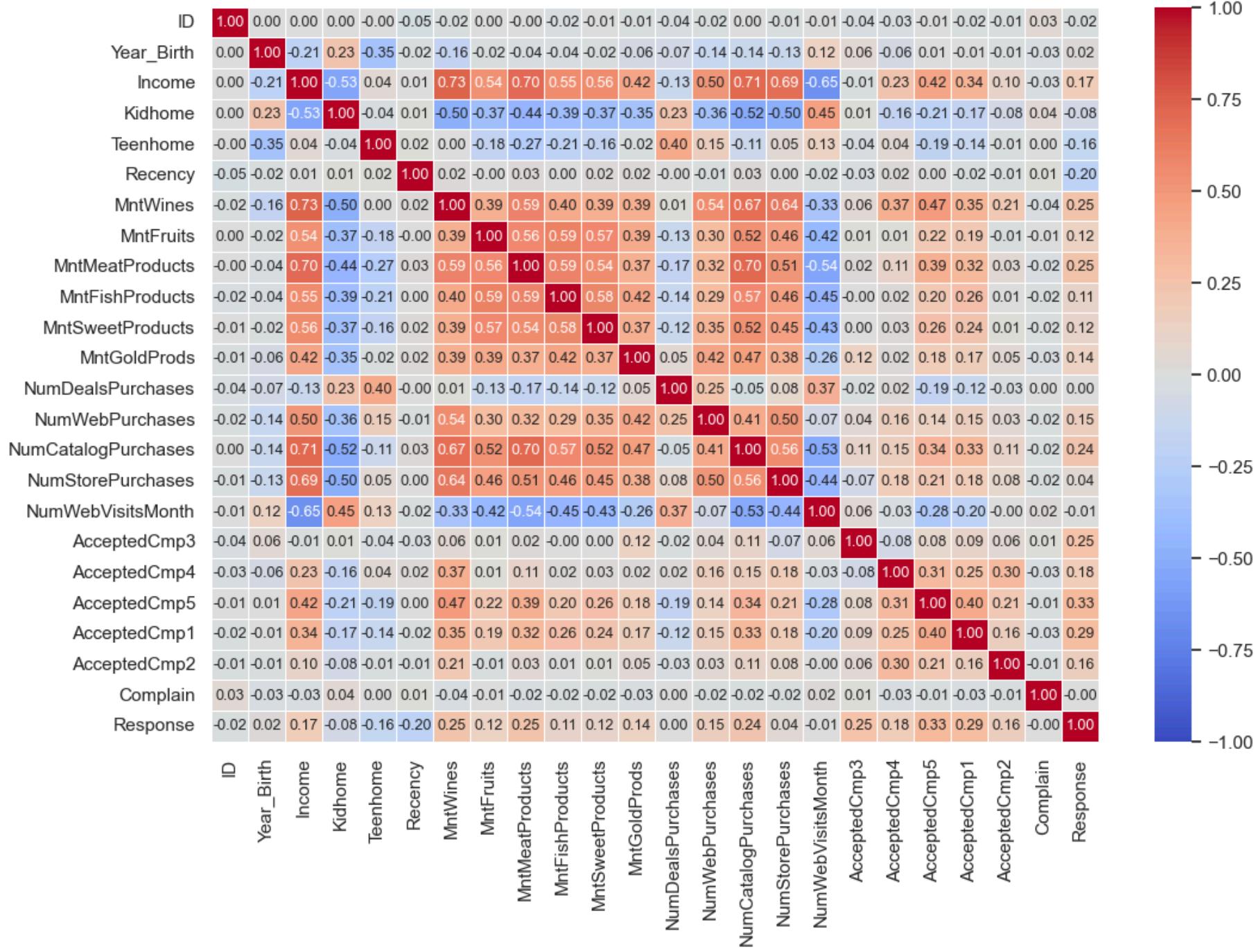
La période Cohorte?

Le Revenue?

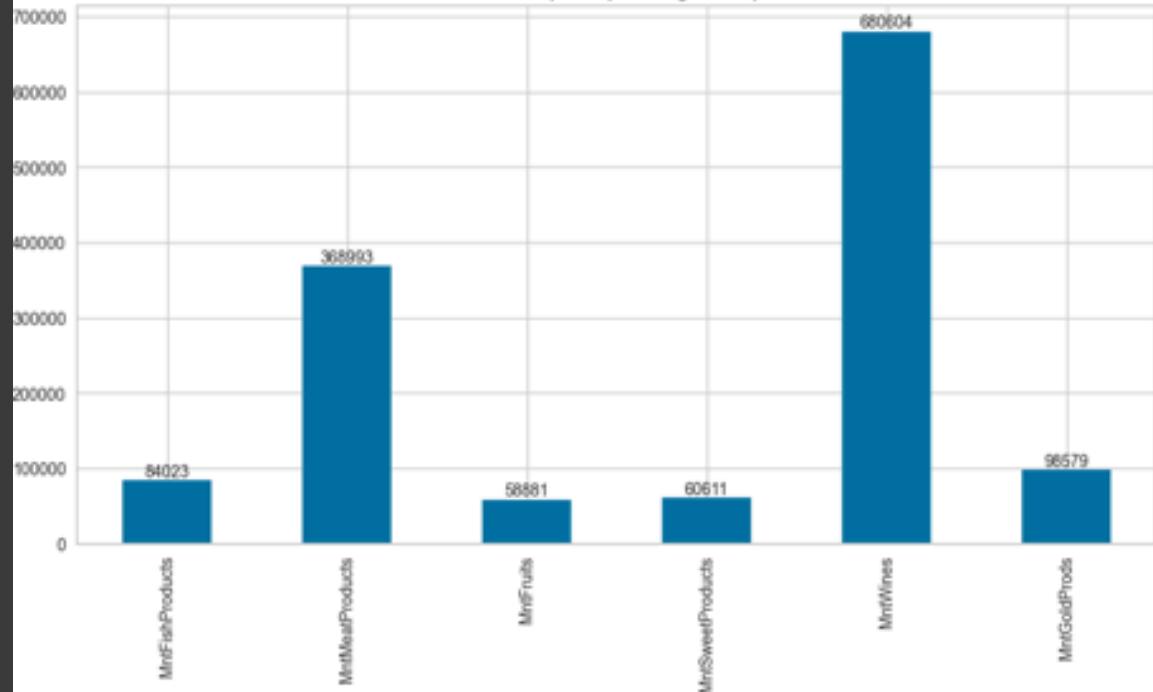


Quelles variables devrions-nous utiliser pour le regroupement ?

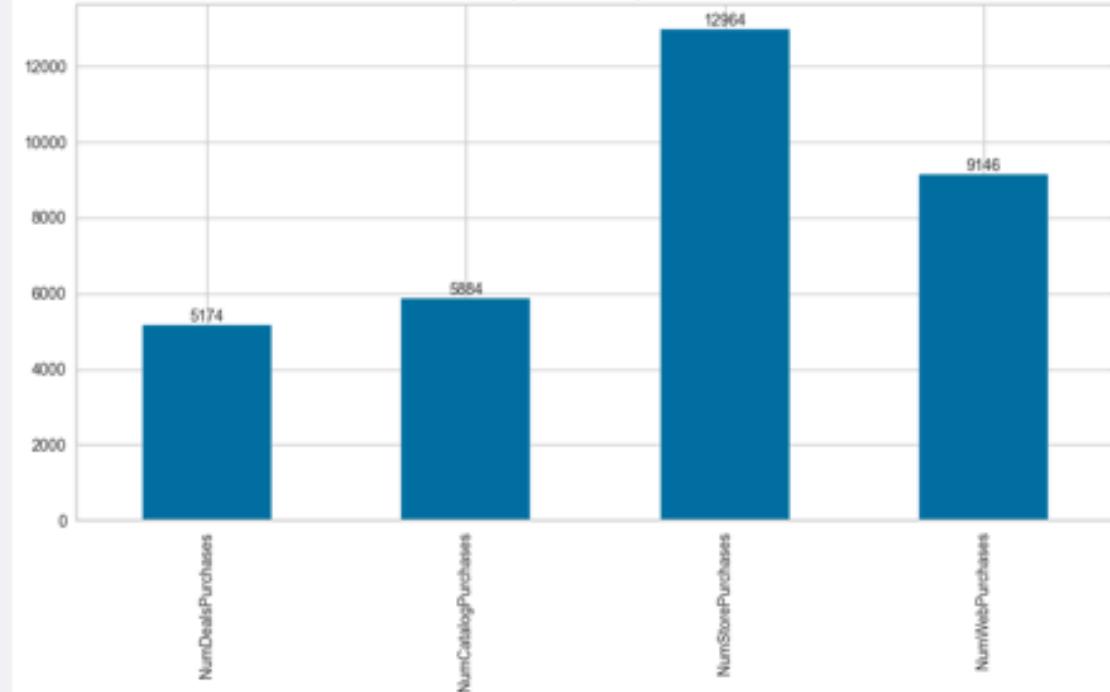
Matrice de Corrélation



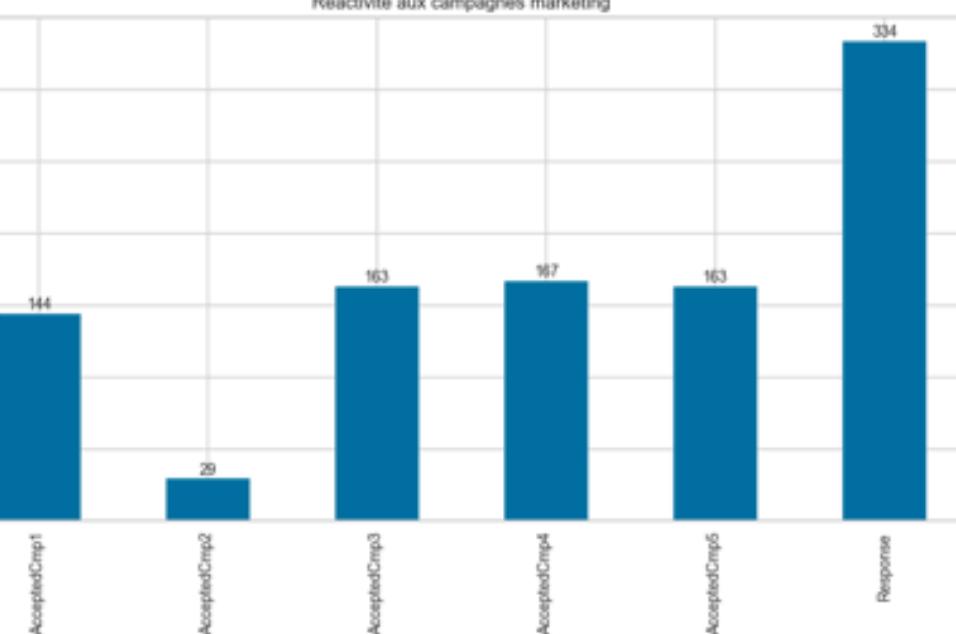
Montant total dépensé par catégorie de produits



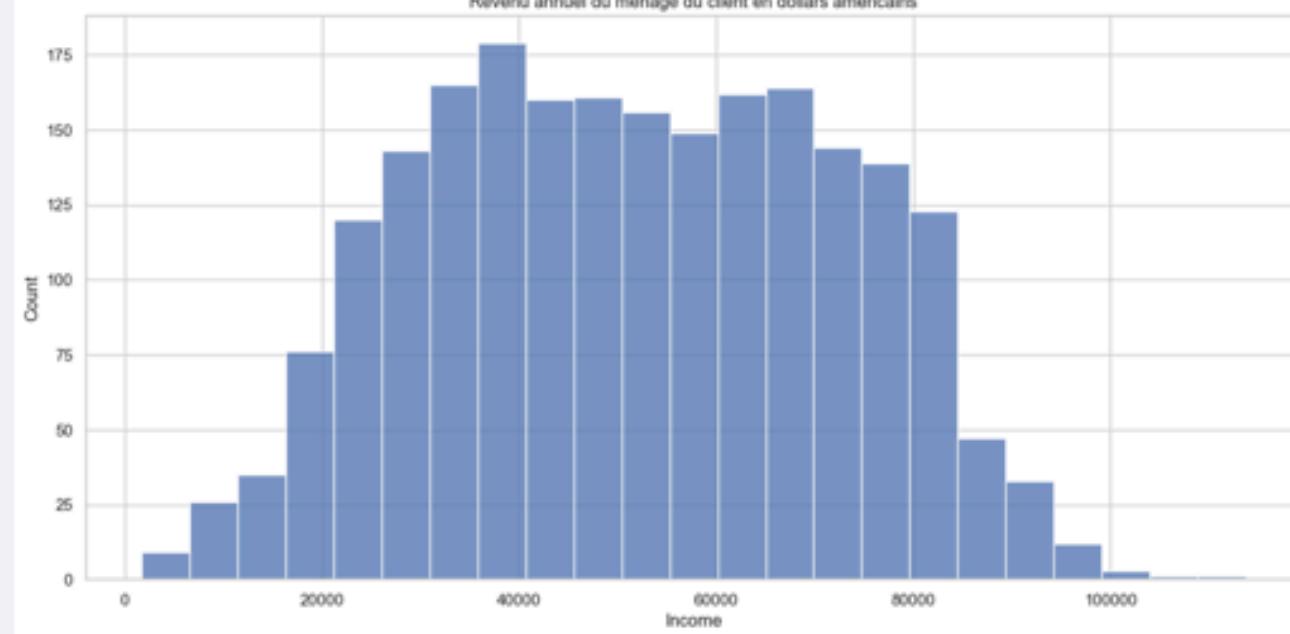
Fréquence d'achat par canal



Réactivité aux campagnes marketing



Revenu annuel du ménage du client en dollars américains





# Quelle méthode de segmentation utilisée est optimale?

Dans notre étude on peut faire différente sorte de segmentations:

## **Segmentation RFM:**

à partir de:

- . Récence du dernier achat
- . Fréquence d'achat
- . Montant total dépensé
- . Valeur moyenne des achats
- . Durée de vie du client

## **Segmentation selon la valeur client:**

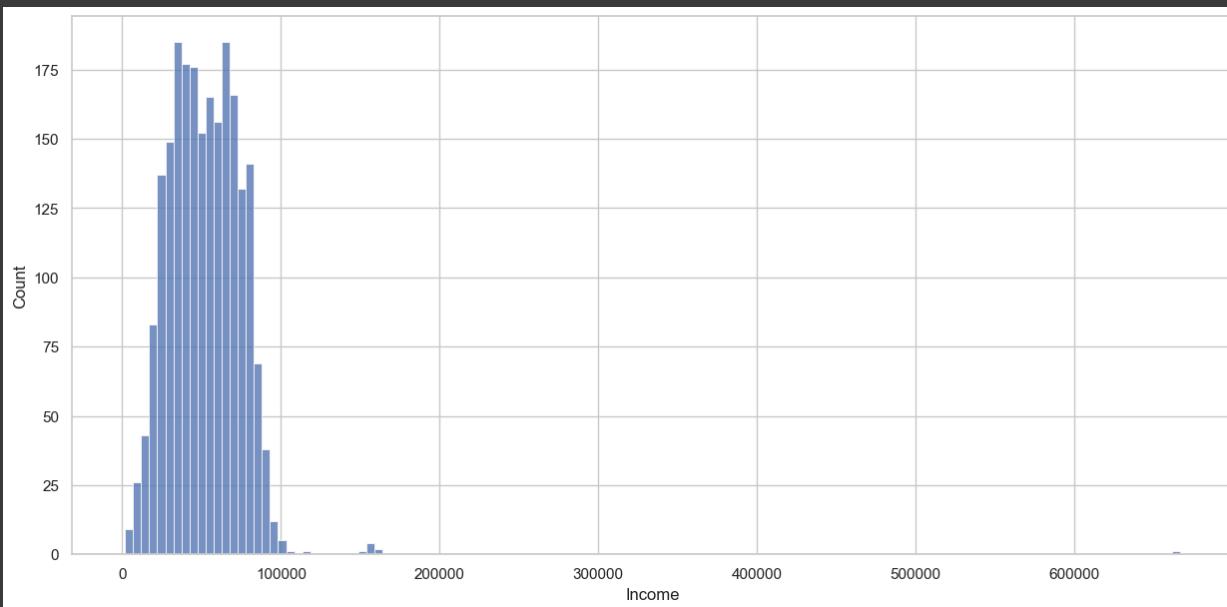
à partir de:

- . Valeur client à long terme
- . Potentiel de revenus futurs
- . Fréquence d'achat
- . Taux de fidélisation
- . Part de portefeuille

Nombre d'entrée avant nettoyage

2240

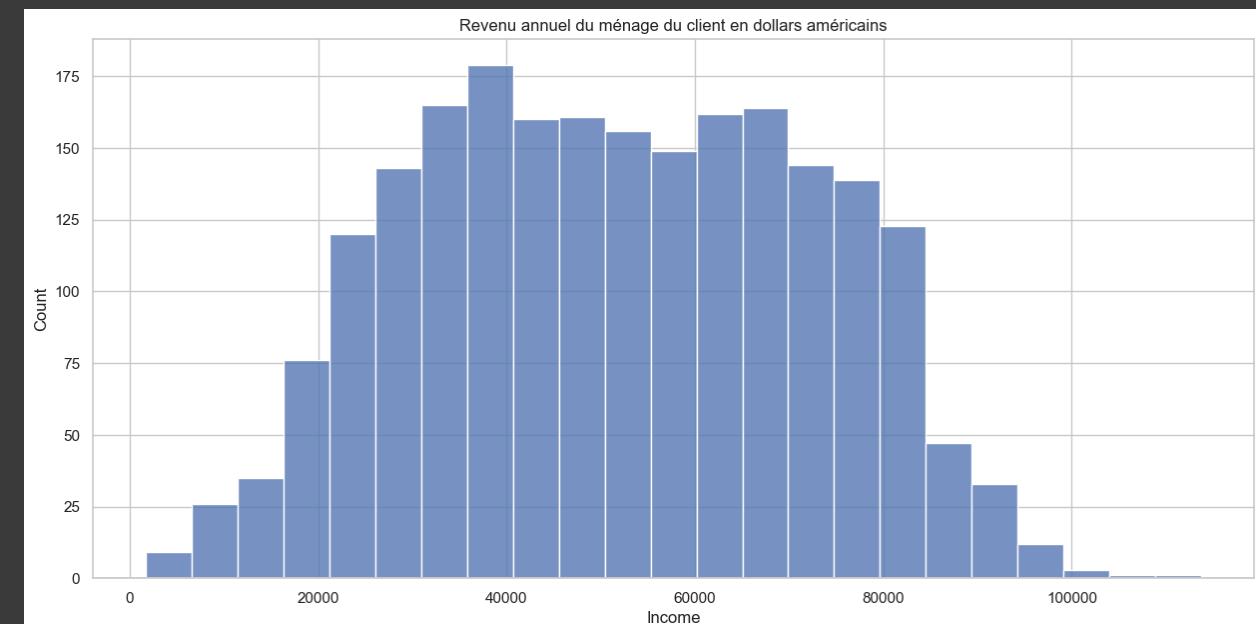
Revenu avant nettoyage



Nombre d'entrées après nettoyage

2232

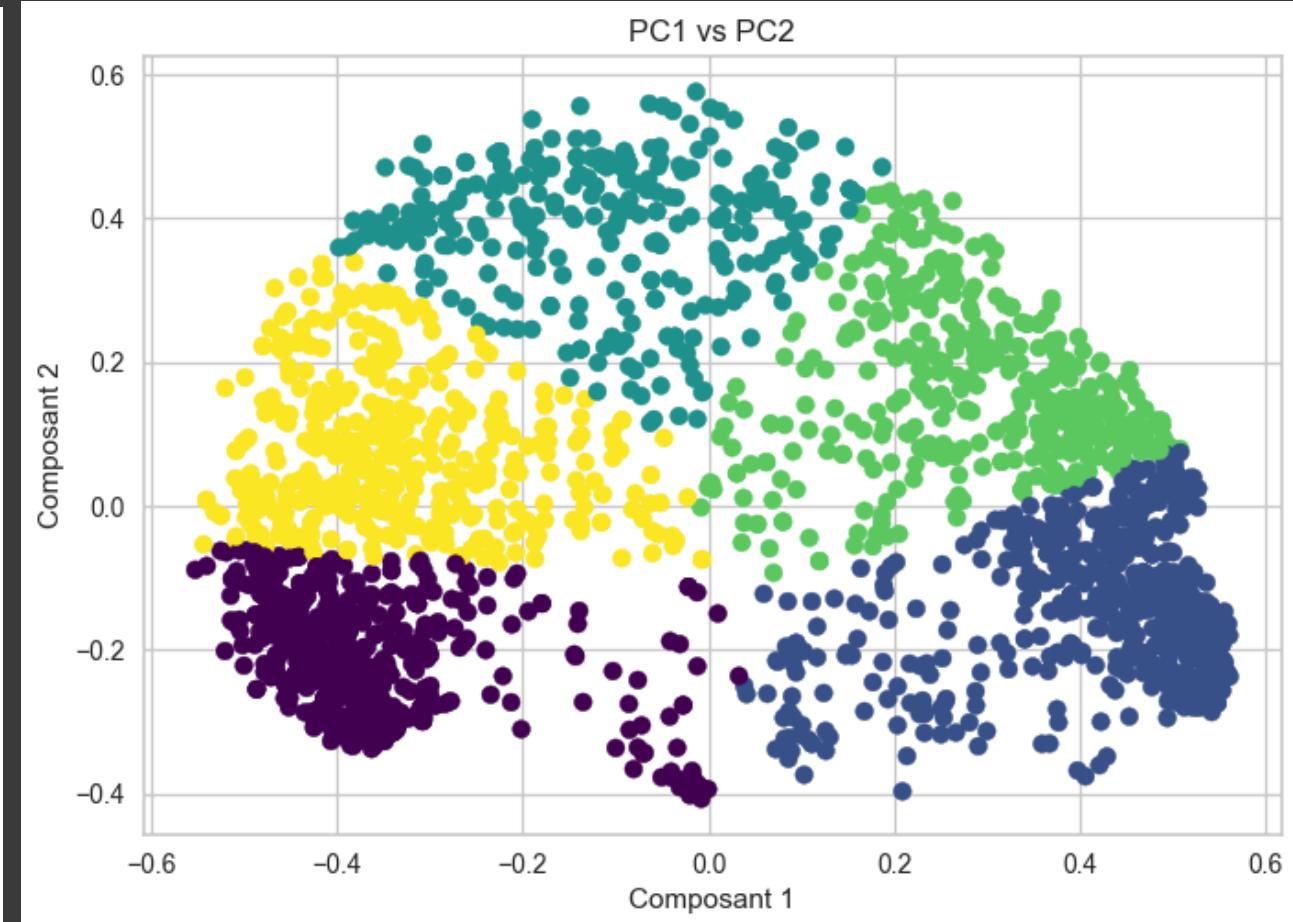
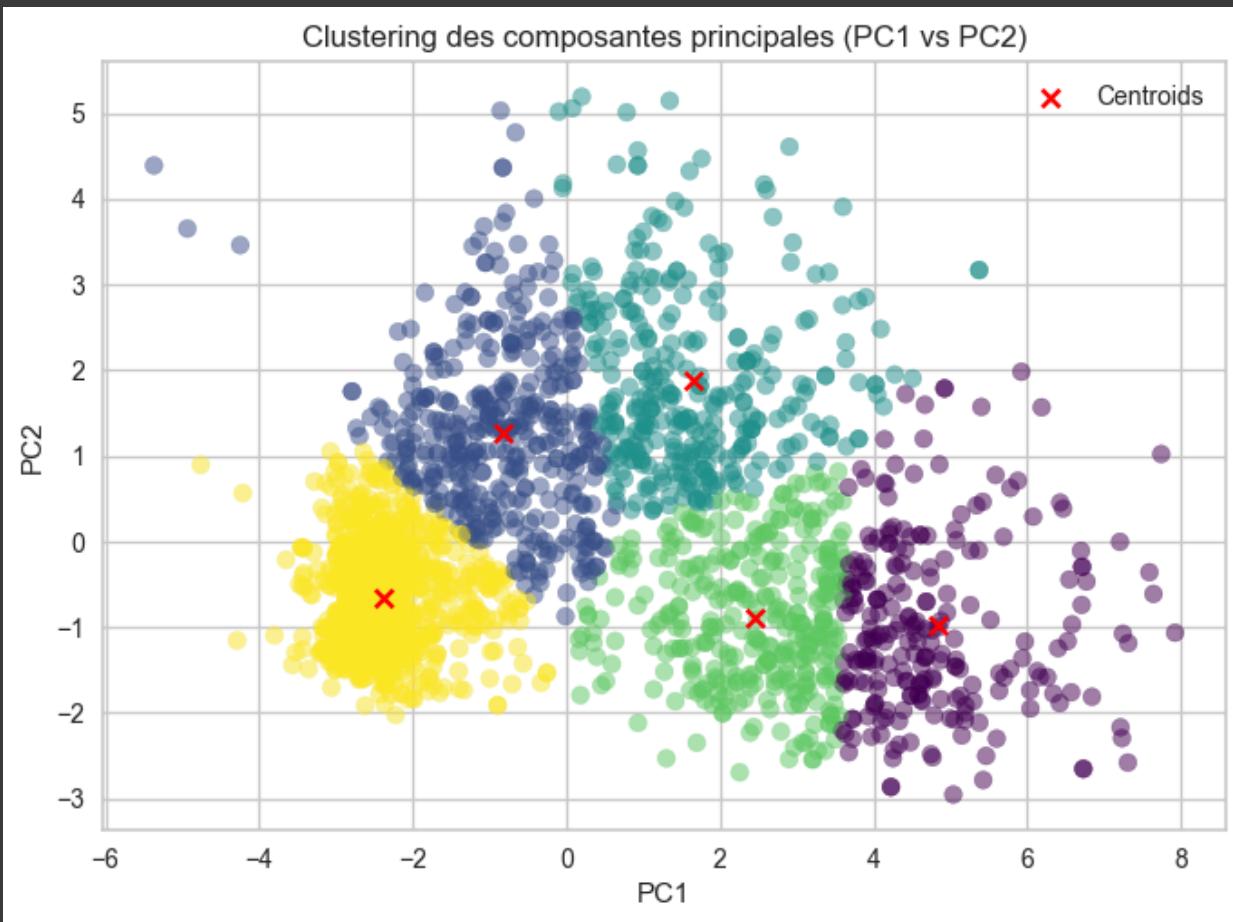
Revenu après nettoyage



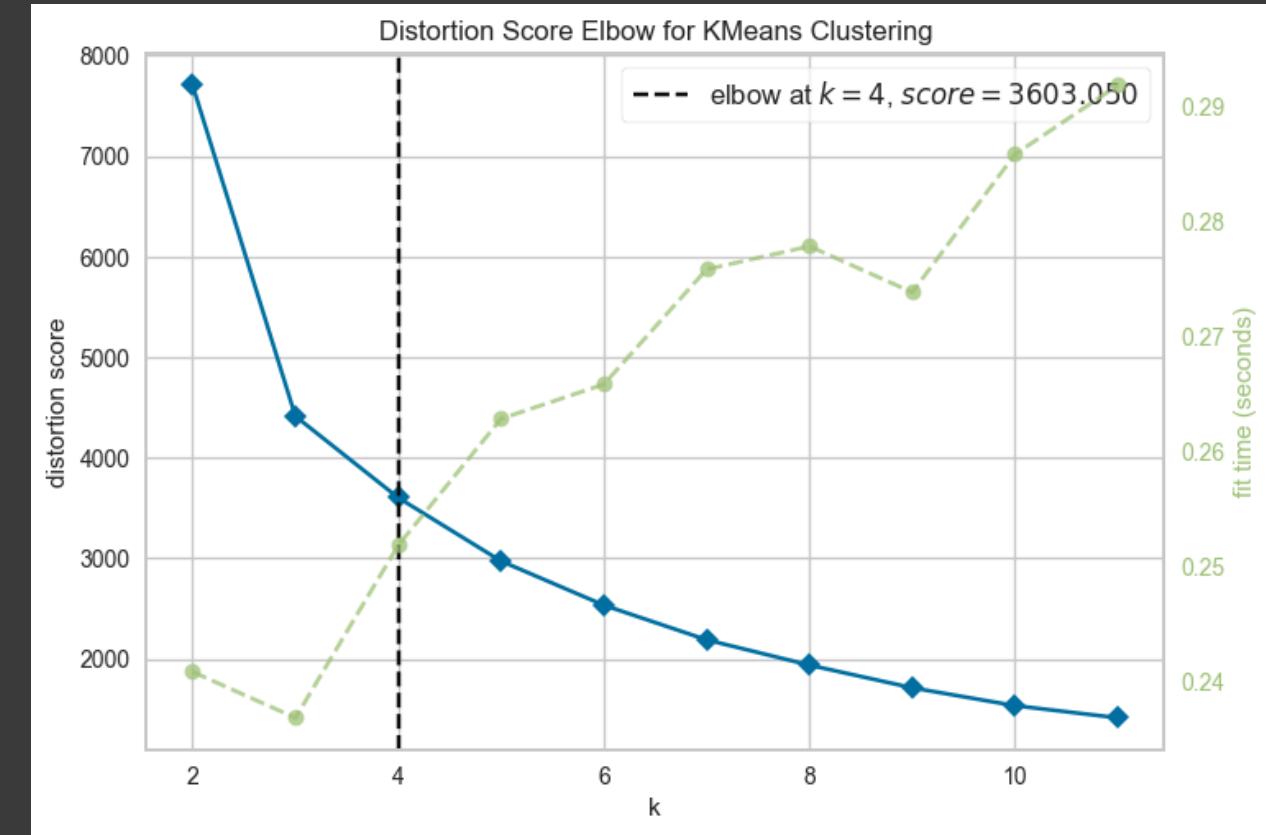
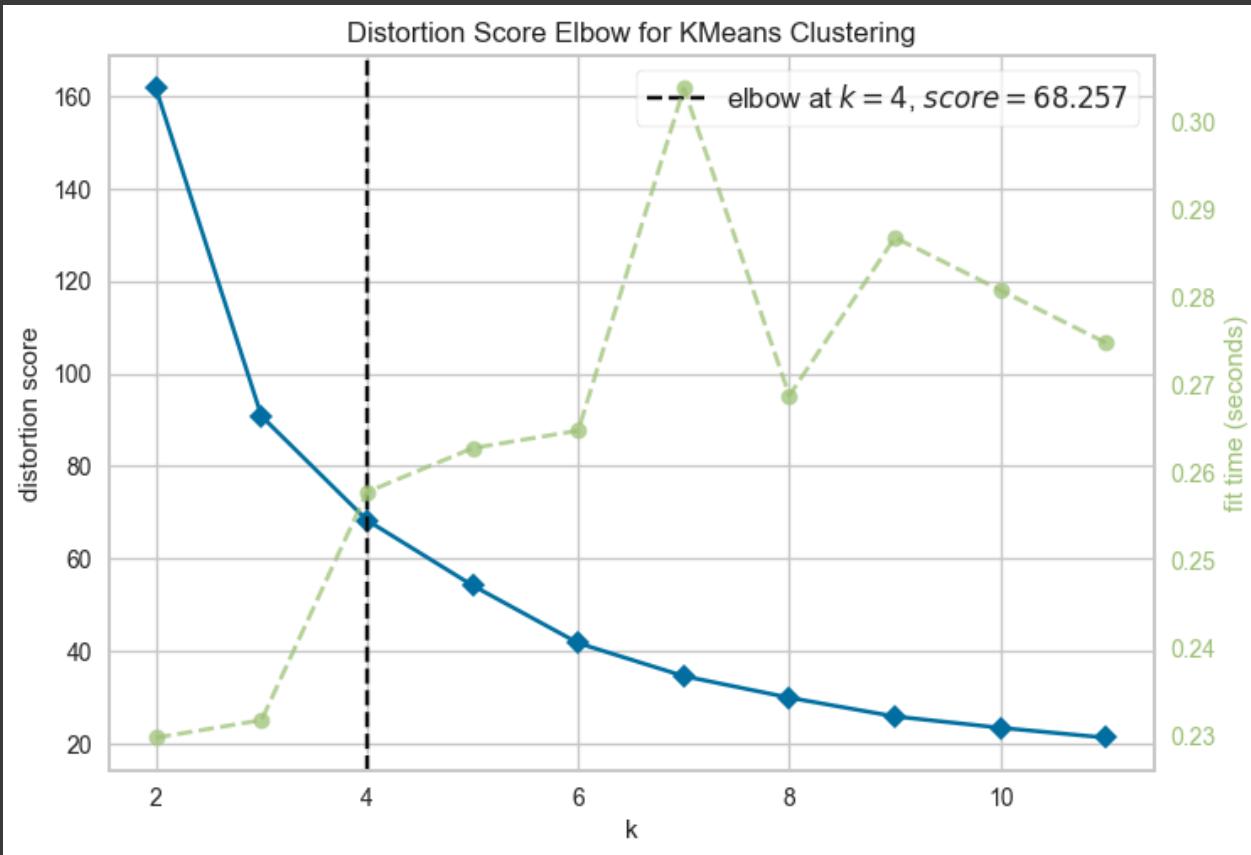
# TECHNIQUES ET ALGORITHMES DE SEGMENTATION

- .PCA (Principal Component Analysis)
- .KPCA (Kernel Principal Component Analysis)
- .K-means
- .Méthode de coude (ou elbow method)
- .DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- .Gaussian Mixture Models (GMM)
- .Hierarchical Clustering

# PCA & KPCA



# Méthode par coude



# RESULTATS RFM

Recency, Frequency, Monetary

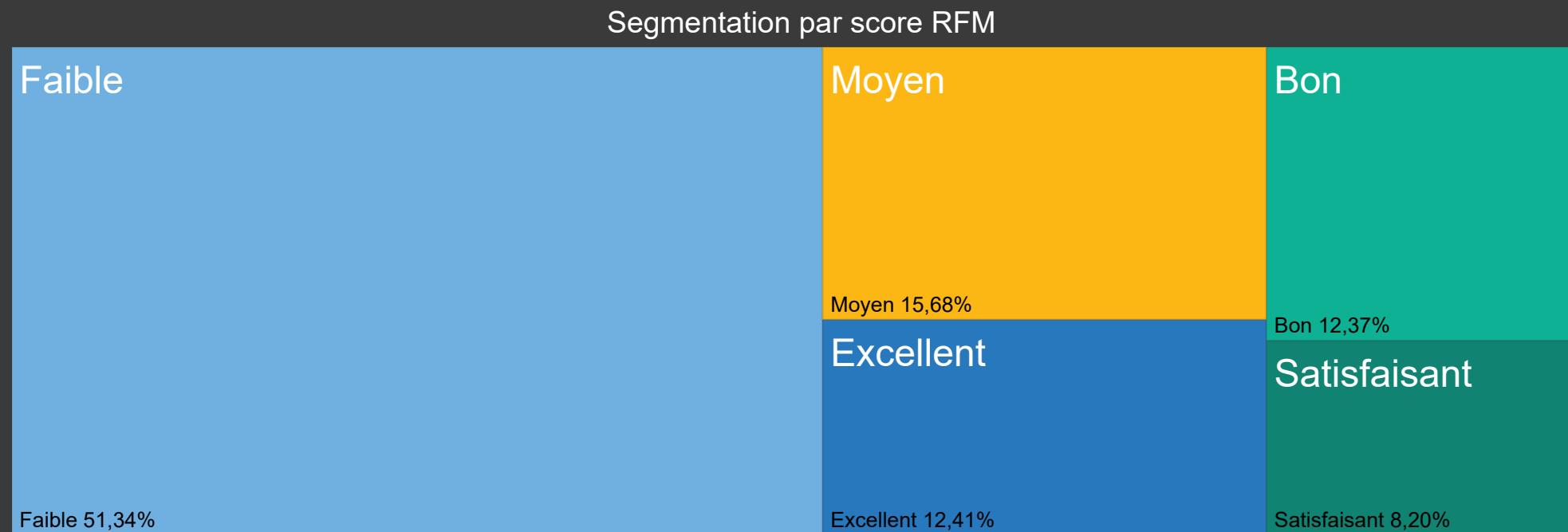
Echelle couper en 5 quartiles et grouper.

Paramètres: High = 0.8 (soit 4/5), Medium = 0.6(soit 3/5), Low = 0.4(soit 2/5)

## Résultats:

si les scores du client sur Recency, Frequency, Monetary sont de 4 et 5 donc il est de Haut catégorie car c'est meilleur, suivi de 3 pour moyen et enfin 1, 2 pour faible.

Nom de segment: Excellent, Bon, Satisfaisant, Moyen, Faible



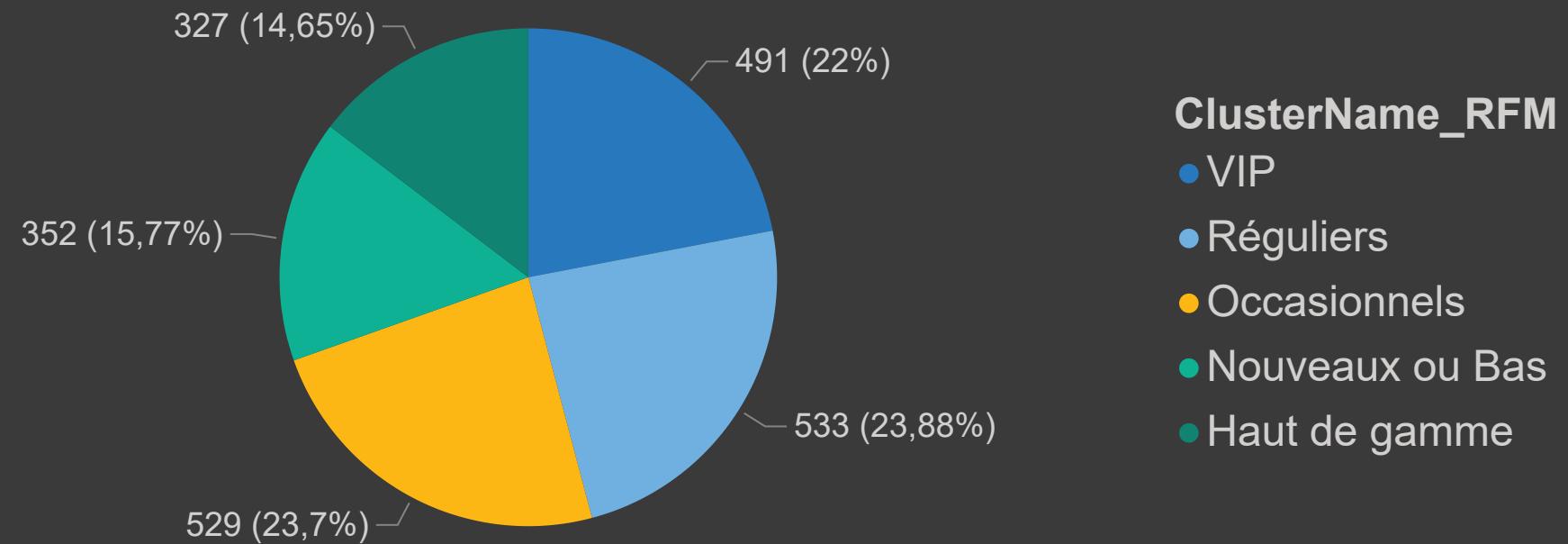
# RESULTATS K-MEANS (segmentation de valeur)

A parti des scores RFM on procède à K-Means de 5 clusters avec 42 répétitions

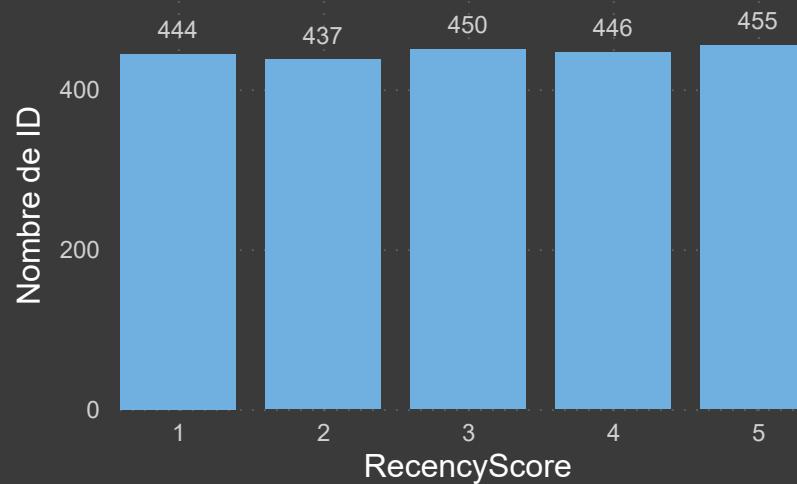
## Résultats:

VIP  
HAUT DE GAMME  
OCCASIONNEL  
REGULIERS  
NOUVEAUX OU BAS

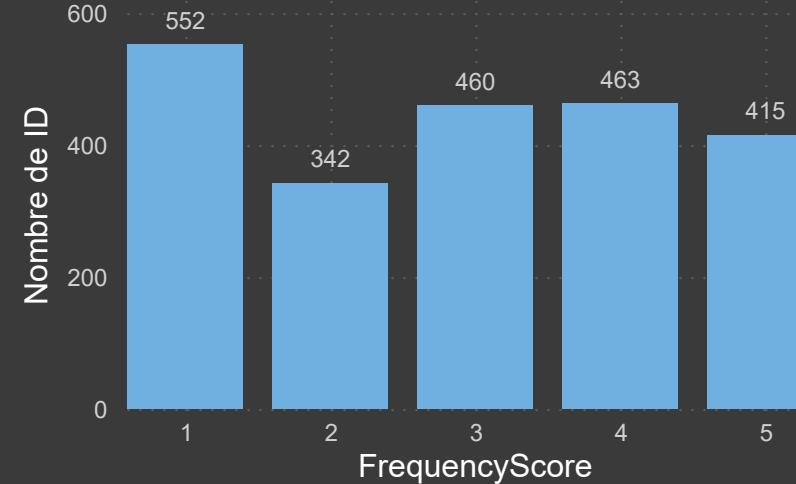
Segmentation par valeur des clients



### Nombre de ID par RecencyScore



### Nombre de ID par FrequencyScore

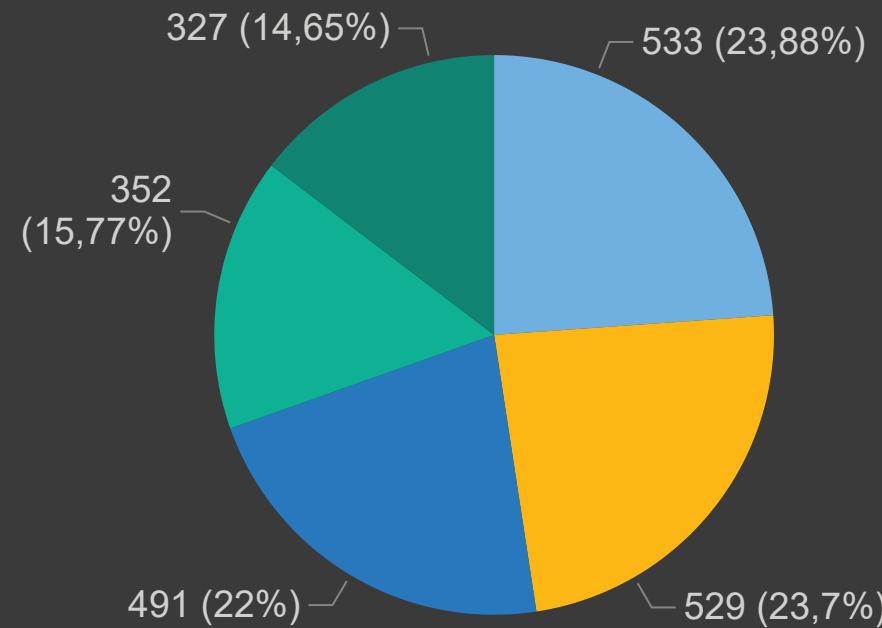
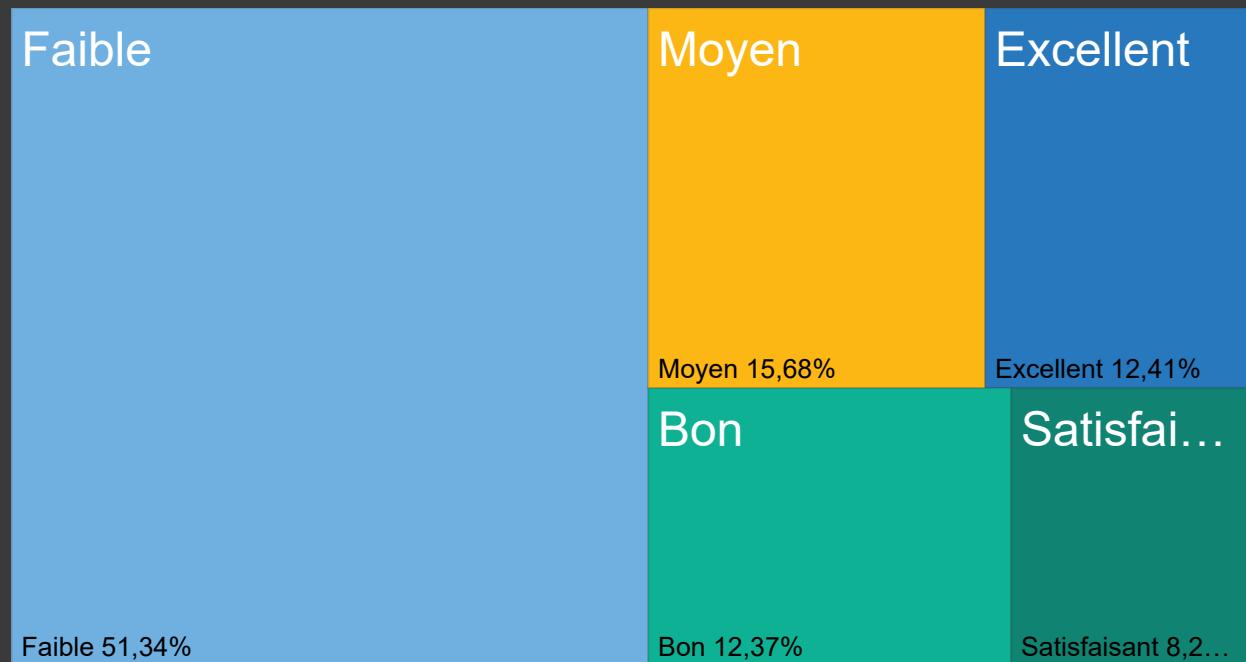


### Nombre de ID par MonetaryScore



## Segmentation par valeur des clients

### Segmentation par potentiel Individuel des clients



- Réguliers
- Occasionnels
- VIP
- Nouveaux ou Bas
- Haut de gamme

# RECOMMANDATION

Après la segmentation **RFM** (Excellent, Bon, Satisfaisant, Moyen, Faible) et la segmentation selon la valeur (VIP, Haut de gamme, Occasionnel, Réguliers, Nouveaux / Bas), voici deux recommandations :

Approche individuelle qui utilise une stratégie personnalisée pour chaque segment, en tenant compte de leurs besoins et préférences spécifiques. Par exemple, offre de promotions ou des produits spéciaux adaptés à chaque niveau de valeur et de fidélité.

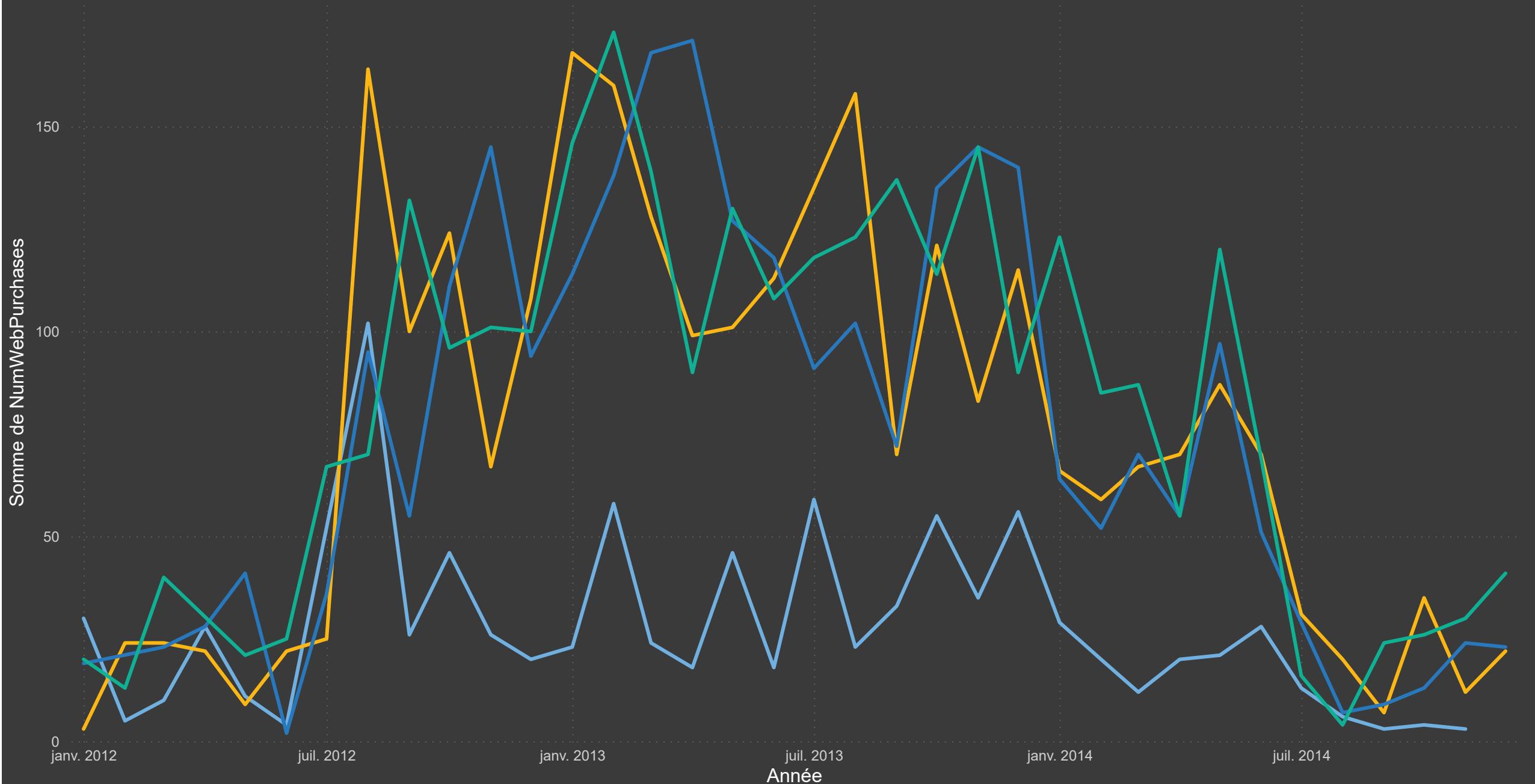
Pack de cible et indice de remerciement pour la fidélisation et la recommandation à d'autres clients : Créez des offres groupées ou des programmes de fidélité spécifiques à chaque segment pour les inciter à rester fidèles et à recommander vos produits ou services à d'autres clients.

Utilisez également des indices de remerciement, tels que des réductions exclusives ou des cadeaux spéciaux, pour montrer votre appréciation envers vos clients les plus fidèles.

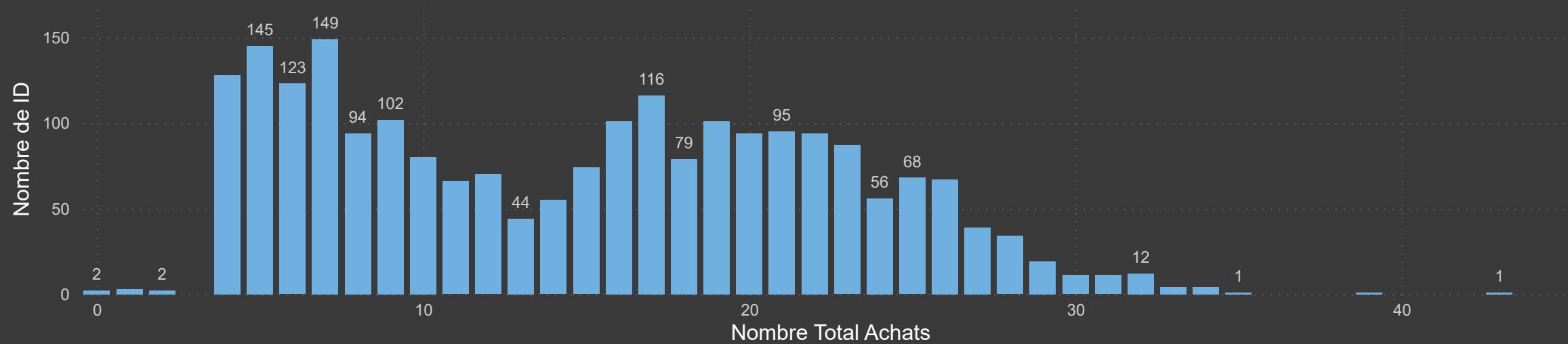
MERCI !!

## Nombre d'achat sur le Web en 2 ans

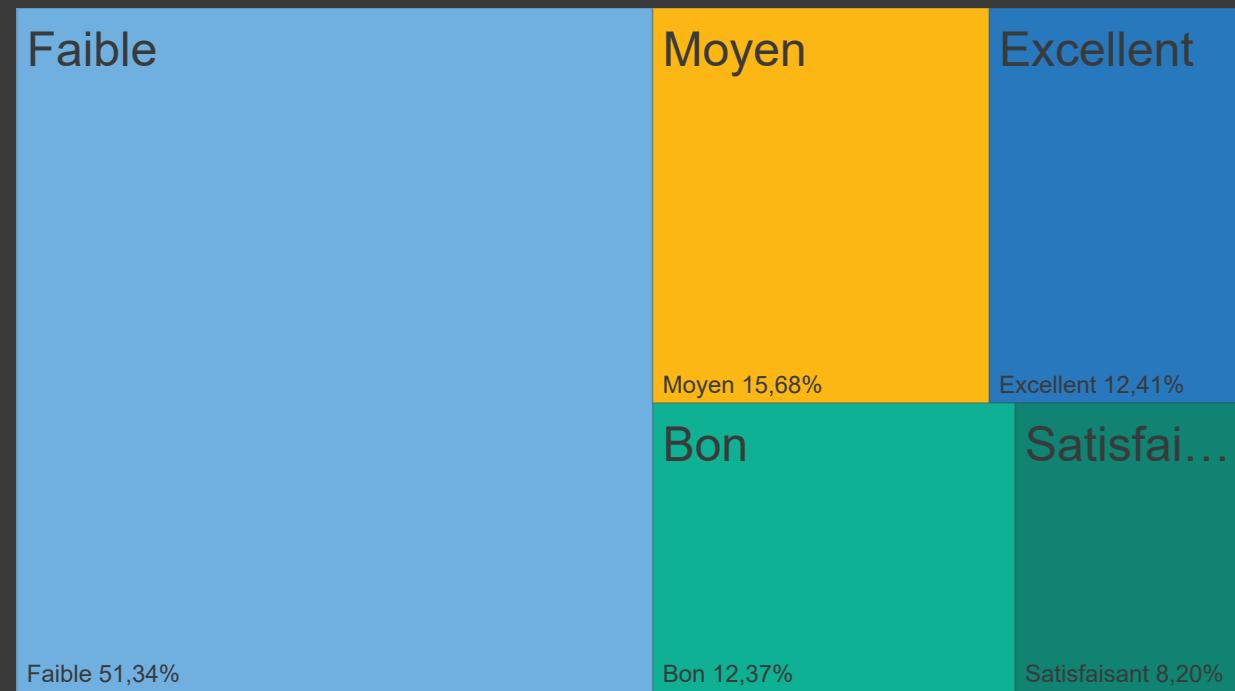
RecencySegment ● 90 jours et plus ● Entre 30 et 59 jours ● Entre 60 et 89 jours ● Moins de 30 jours



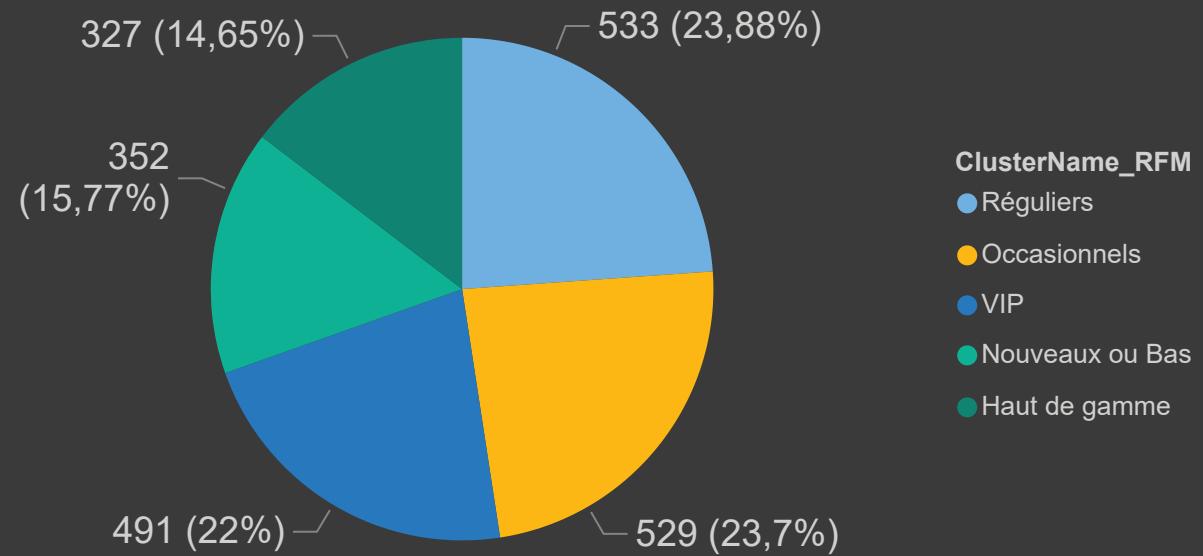
## Nombre de ID par Nombre Total Achats



## Segmentation par potentiel Individuel des clients

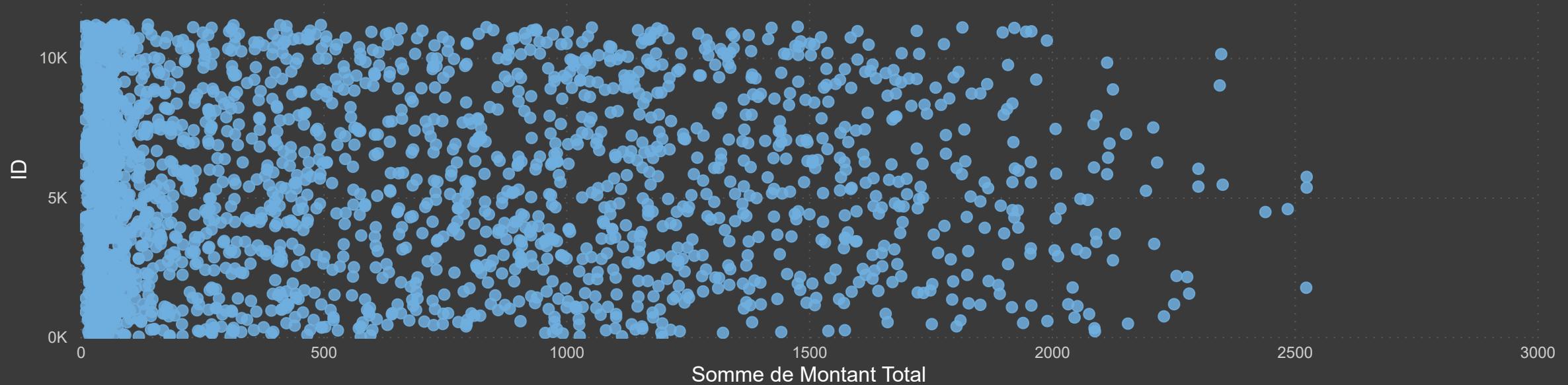


## Segmentation par valeur des clients



- ClusterName\_RFМ**
- Réguliers
  - Occasionnels
  - VIP
  - Nouveaux ou Bas
  - Haut de gamme

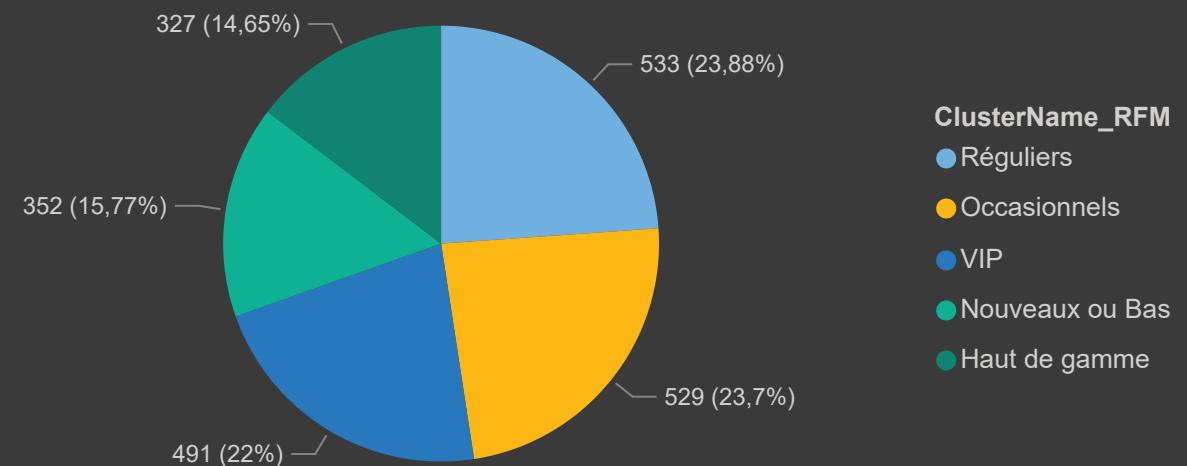
## Somme de Montant Total par ID



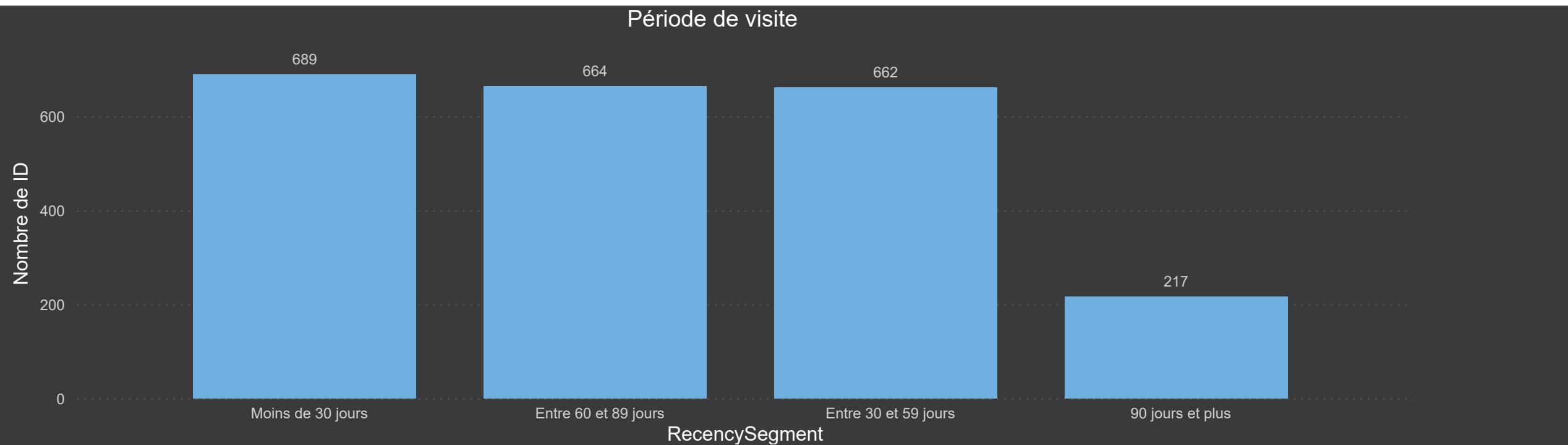
## Segmentation par potentiel Individuel des clients



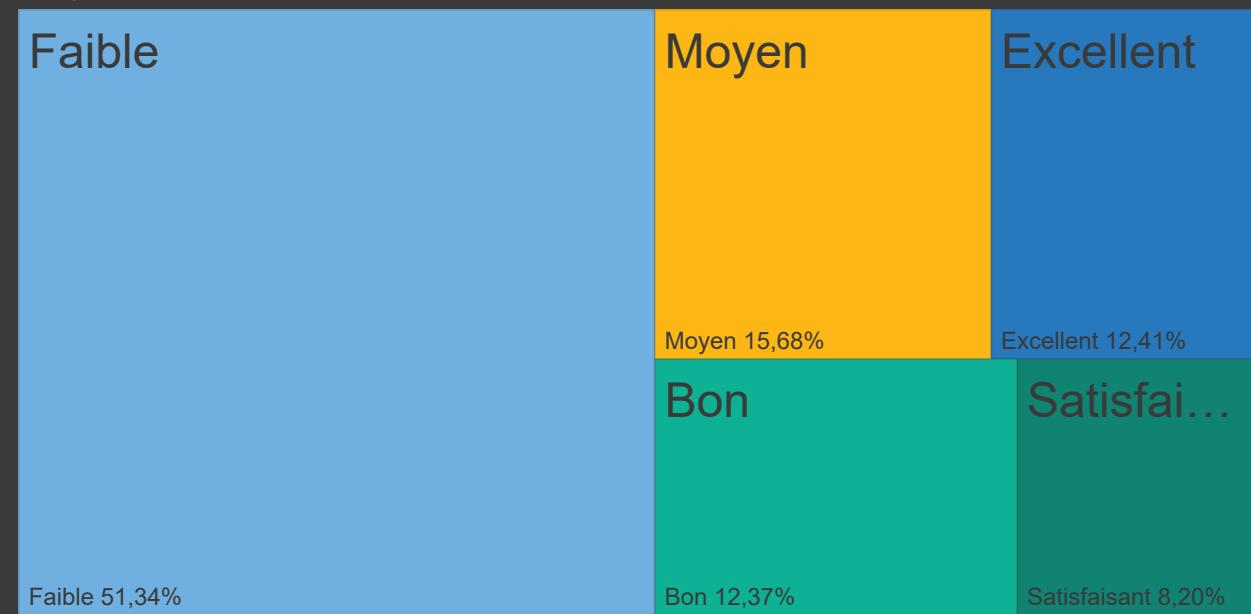
## Segmentation par valeur des clients



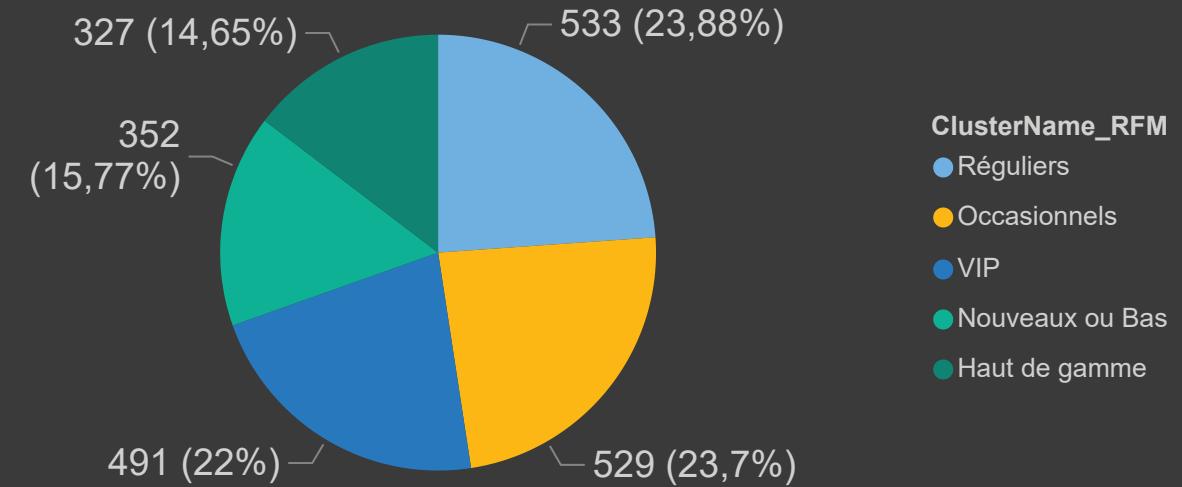
- ClusterName\_RF**
- Réguliers
  - Occasionnels
  - VIP
  - Nouveaux ou Bas
  - Haut de gamme



### Segmentation par potentiel Individuel des clients

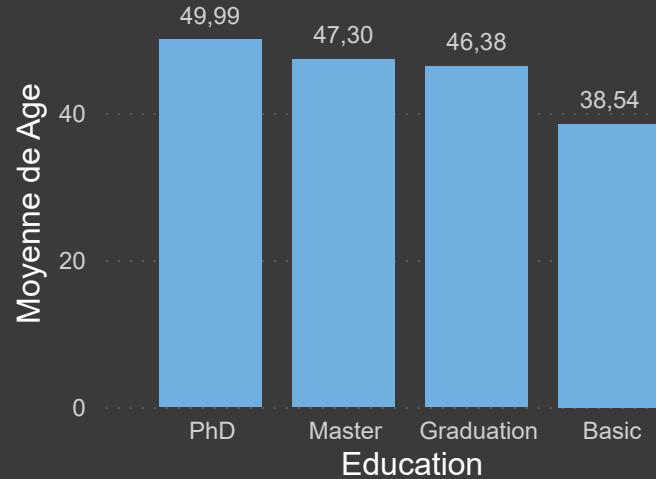


### Segmentation par valeur des clients



- Réguliers
- Occasionnels
- VIP
- Nouveaux ou Bas
- Haut de gamme

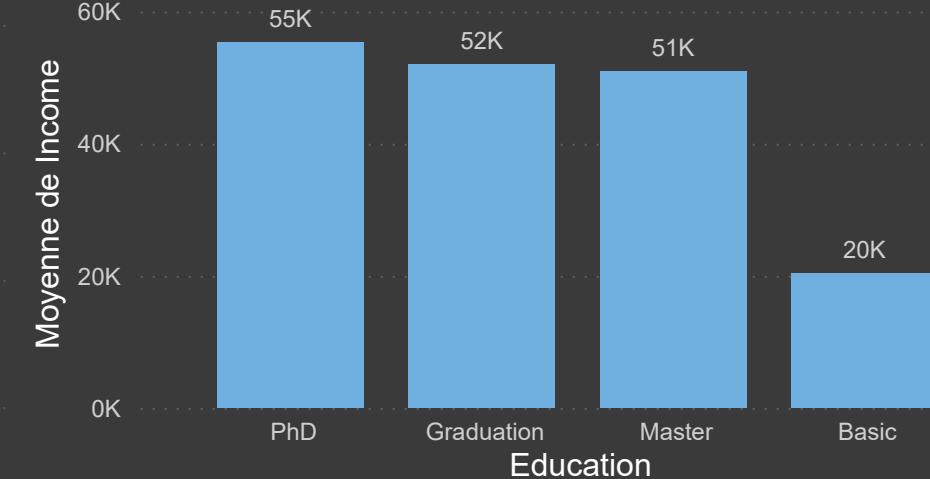
### Moyenne de Age par Education



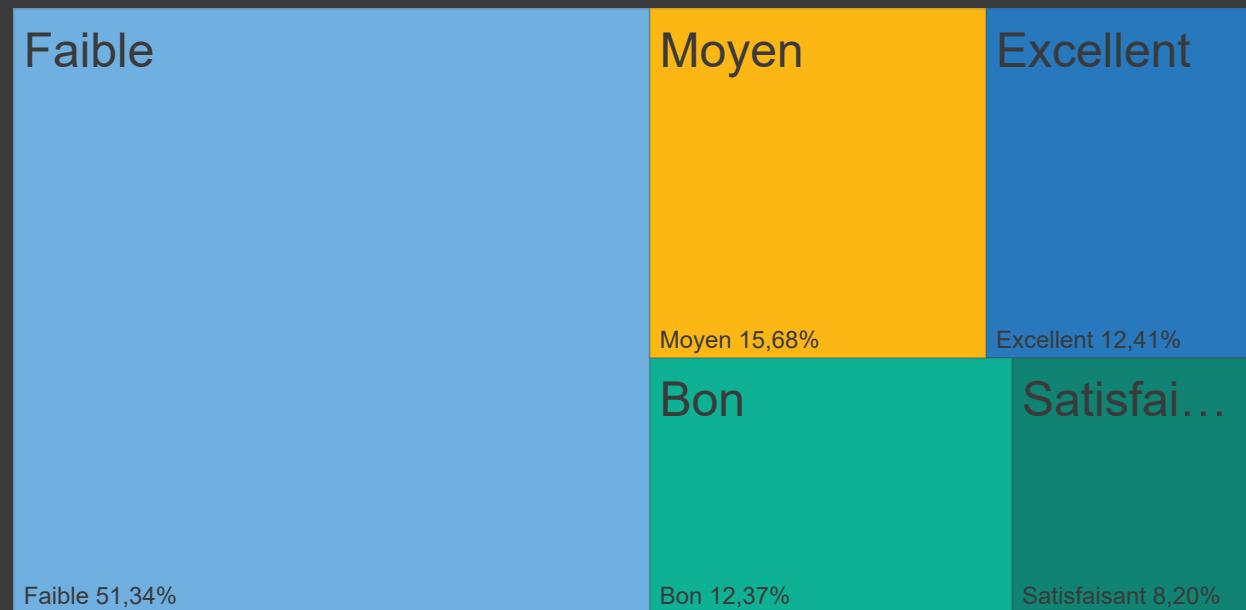
### Moyenne de Age par Marital\_Status



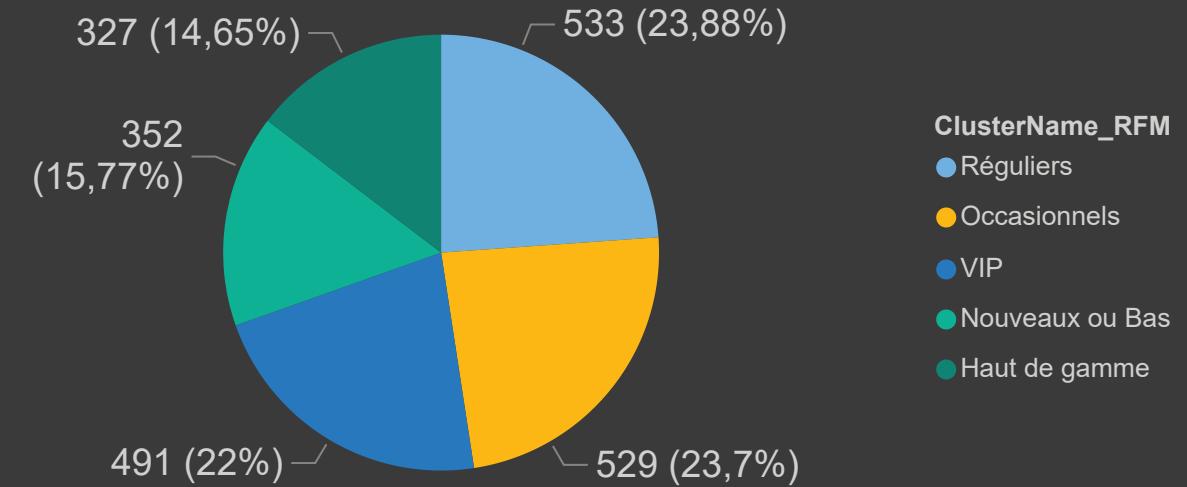
### Moyenne de Income par Education



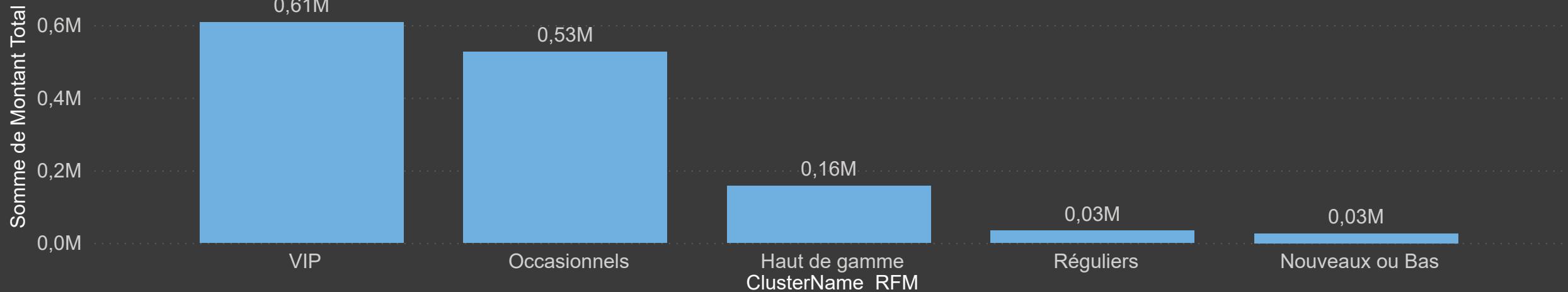
### Segmentation par potentiel Individuel des clients



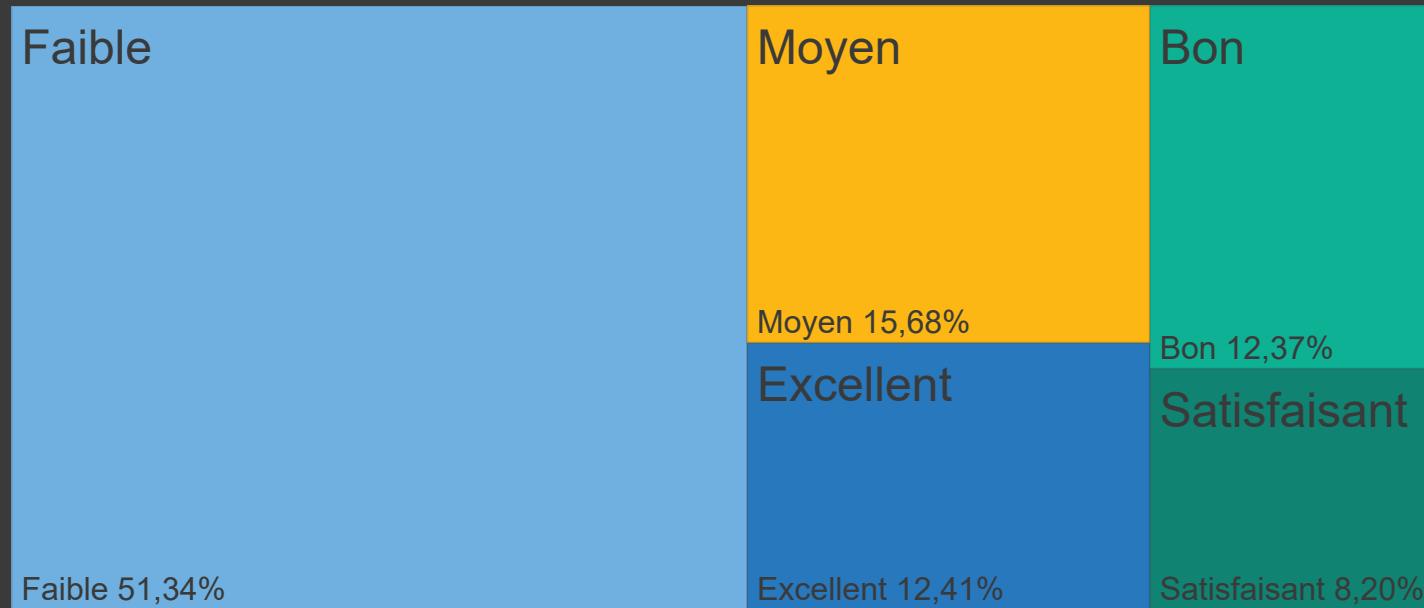
### Segmentation par valeur des clients



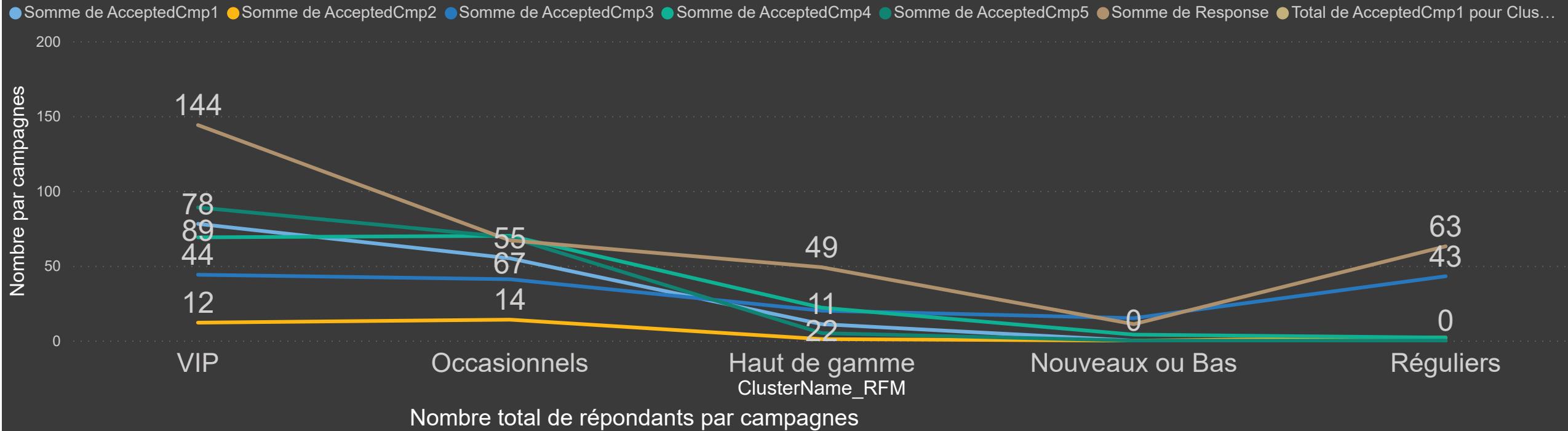
## Montant dépensé par valeur de client



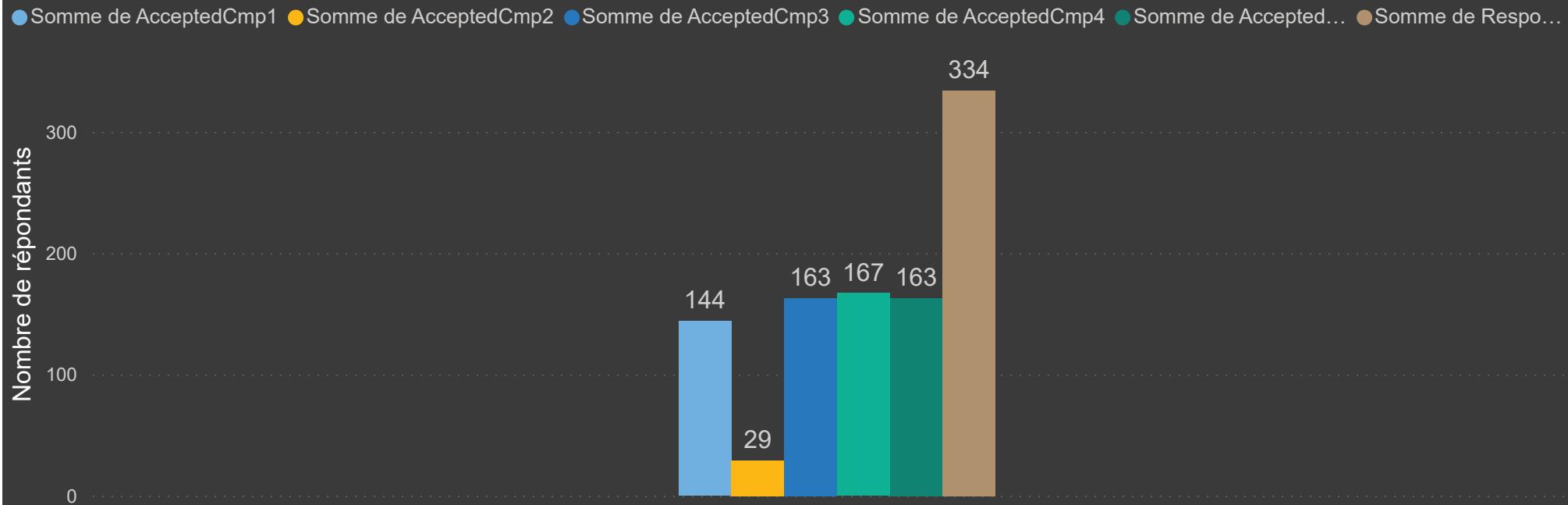
## Segmentation par potentiel Individuel des clients

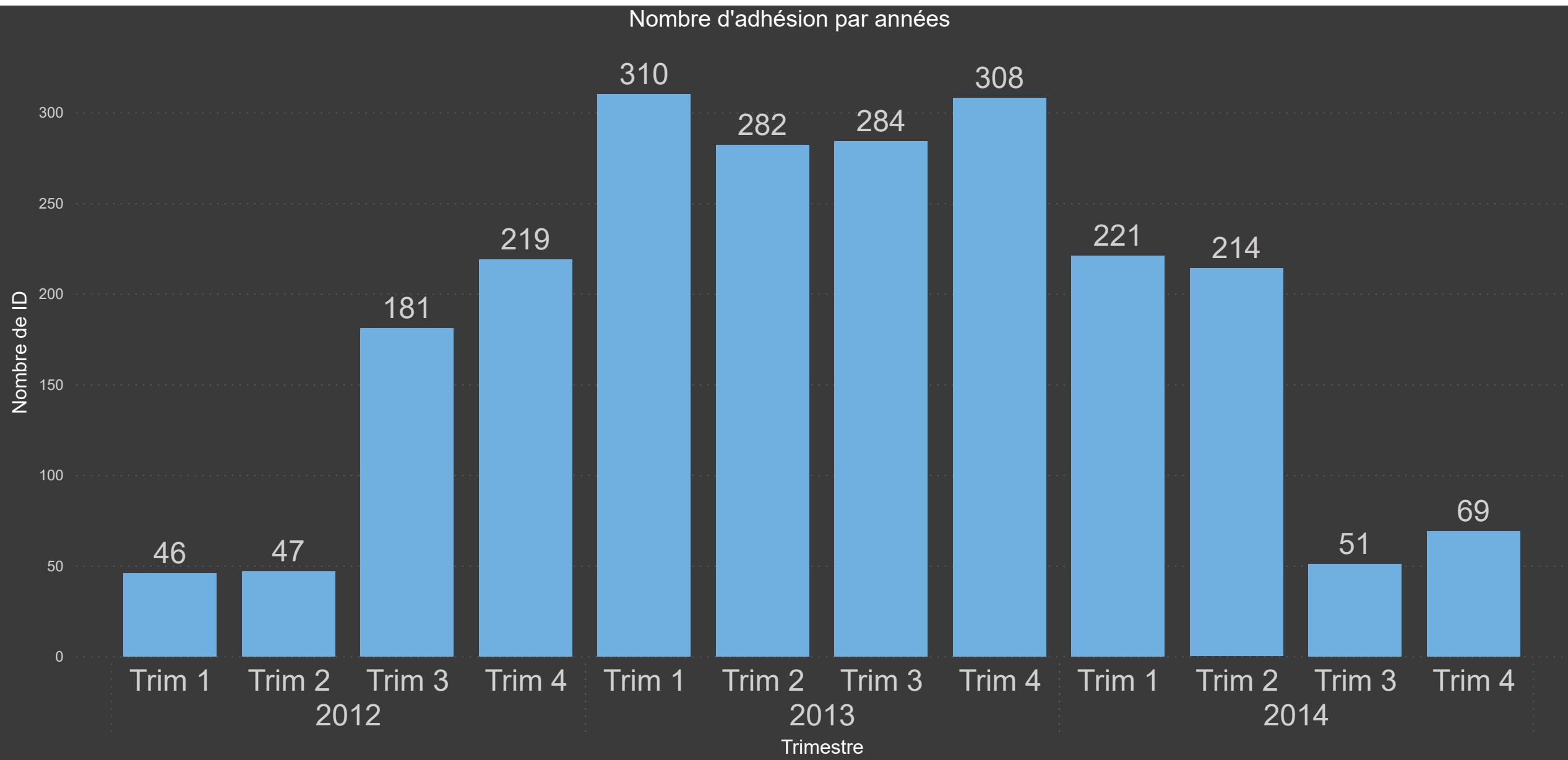


## Nombre de réponses de campagnes par segment valeur des clients

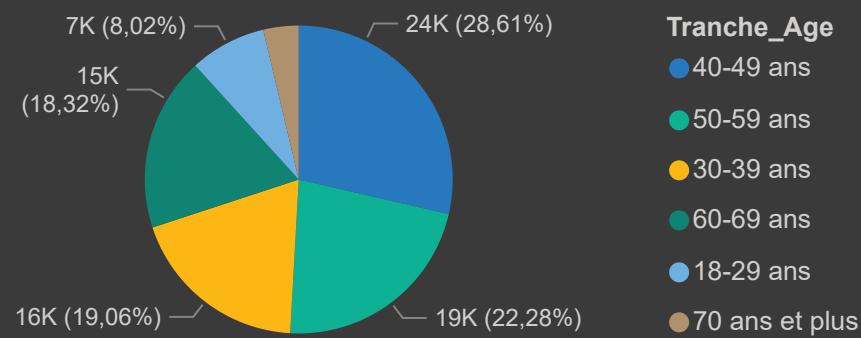


## Nombre total de répondants par campagnes

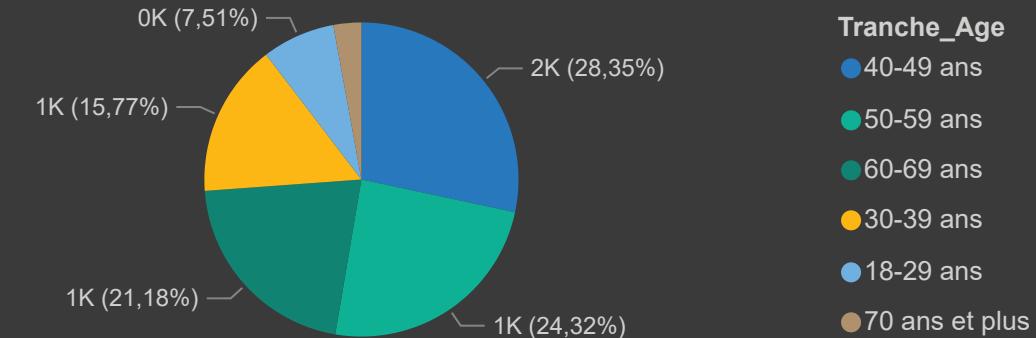




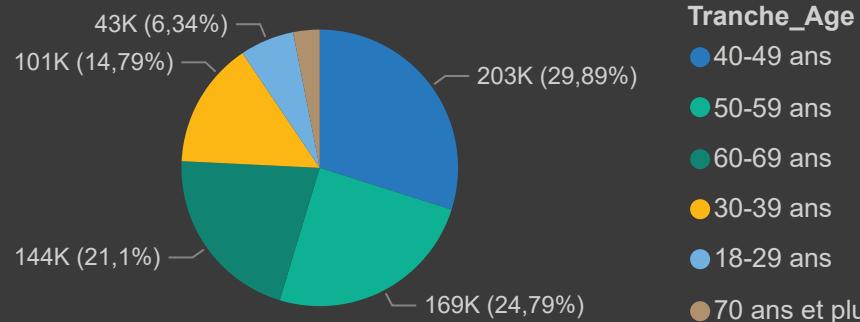
### Total MntFishProducts par Tranche\_Age



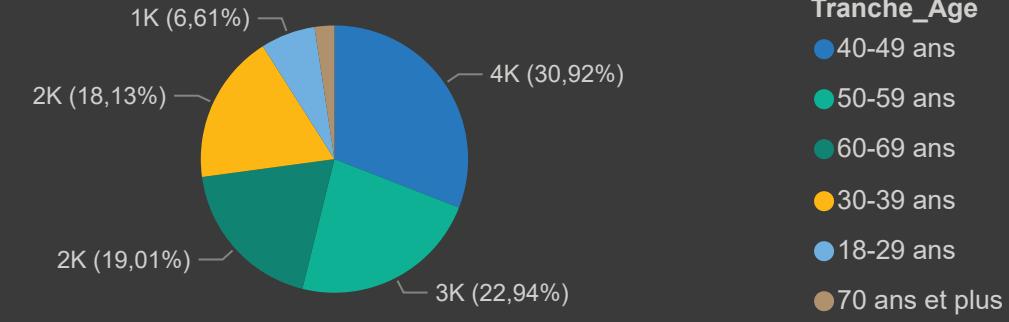
### Total NumCatalogPurchases par Tranche\_Age



### Total MntWines par Tranche\_Age



### Total NumStorePurchases par Tranche\_Age



### Nombre de ClusterName\_RFMs par ClusterName\_RFMs

