

**INSTITUT SUPERIEUR DE STATISTIQUE, D'ECONOMETRIE
ET DE DATASCIENCE**

MASTER 2

STATISTIQUE ECONOMETRIE DATA SCIENCE

**MINI PROJET ANALYSE DES DONNEES
MULTIDIMENSIONNELLES**

**ANALYSE DES COMPORTEMENTS ET
PRÉFÉRENCES POUR L'OPTIMISATION DES
RESSOURCES**

ANNEE UNIVERSITAIRE : 2024-2025

ETUDIANT

N'DRI ONESIME

ENSEIGNANT

AKPOSSO DIDIER MARTIAL

AVANT-PROPOS

L'association du théorique au pratique, des connaissances aux compétences et des savoir-faire aux savoirs est la principale tendance récente dans le secteur technique. Dans ce contexte, l'INSSEDS (Institut Supérieur de la Statistique, d'Econométrie et de la Data Science), dans sa formation en master professionnel en statistique, économie et science des données, impose que les divers crédits soient validés en effectuant un mini-projet à la fin de chaque module. Le projet est donc structuré et supervisé de cette manière, visant principalement à faire de chaque élève un participant dynamique, engagé et libre dans la vie active.

Ce document est un rapport de projet Muni axé sur l'économétrie des séries temporelles ANALYSE DES DONNEES MULTIDIMENSIONNELLES. Il se divise principalement en trois parties : PRETRAITEMENT DES DONNEES, SEGMENTATION et CLUSTERING

En règle générale, toutes les analyses et conclusions présentées dans ce rapport relèvent de la responsabilité de l'auteur, qui ne sollicite ni autrui ni l'INSSEDS (Institut Supérieur de Statistique d'Econométrie et de Data Science).

Table des matières

Introduction générale.....	5
Contexte et justification	5
Problématique.....	5
Principaux résultats attendus.....	5
Methodologie.....	6
Dictionnaire du jeu de Données	6
PARTIE 1 : PRETRAITEMENT DES DONNEES.....	9
a) Importation, restructuration et description du jeu de Données	9
<input type="checkbox"/> Renommage des variables	9
<input type="checkbox"/> Conversion de la structure des variables	10
b) Transformation des variables binaires et recodage des noms des modalités des variables “education” et “situation_mat”	11
<input type="checkbox"/> Variables binaires	11
<input type="checkbox"/> Recodage des modalités des variables “éducation” et “situation_mat” ..	11
c) Traitement des doublons.....	12
d) Traitement de valeurs manquantes	12
<input type="checkbox"/> Visualisation des données manquantes avant traitement	12
<input type="checkbox"/> Visualisation des valeurs manquantes après traitement.....	13
e) Traitement des valeurs abberantes ou extrêmes	13
PARTIE 2 : Analyse des composantes principales (ACP).....	15
1) Importation des variables quantitatives.....	15
2) Choisir les variables et les individus actifs.....	16
Variables Actives.....	16
Variables Supplémentaires	17
3) Choix du nombre des axes	17
4) Analyse des variables	18
Résumés numériques	20
Composante principale 1 (Dim 1).....	21
Composante principale 2 (Dim 2).....	21

5) Analyse du graphe des individus.....	22
DIM 1 :	22
DIM 2 :	23
6) Conclusion de l'Analyse ACP	24
PARTIE 3 : Analyse des correspondances multiple (ACM)	27
1) Importations des variables qualitatives.....	27
Résumés numériques	27
2) Choix du nombre d'axes	28
3) Analyse des résultats	30
Graphe des individus	30
Résumés numériques de ces dix individus :	31
Graphe des modalités actives.....	31
Graphe des variables	32
Lien entre les variables et les modalités.....	33
4) Conclusion	35
PARTIE IV: Clustering	37
1) Classification sur ACP	37
Arbre hiérarchique.....	37
Projection sur le plan factoriel	39
2) Classification sur ACM.....	39
Arbre hiérarchique.....	39
Projection sur le plan factoriel	41
Conclusion générale	42
Recommandations	43

Introduction générale

Contexte et justification

Dans le paysage concurrentiel actuel, les entreprises cherchent constamment à améliorer leur compréhension des clients pour mieux répondre à leurs besoins. L'analyse de la personnalité des clients est devenue une méthode essentielle pour obtenir des informations précieuses sur les différents profils de clients. Elle permet non seulement d'améliorer la connaissance des préférences et des comportements des clients, mais aussi d'adapter les produits et services de manière plus ciblée.

Problématique

L'un des principaux défis pour les entreprises est de segmenter efficacement leurs clients en fonction de leurs comportements et préférences. Cette segmentation est crucial pour:

- Cibler les campagnes marketing de manière plus précise ;
- Optimiser l'allocation des ressources;
- Personnaliser les offres de produits pour maximiser la satisfaction client.

La difficulté réside dans l'identification des segments de clientèle les plus pertinents et dans l'optimisation des stratégies marketing pour ces segments spécifiques.

Principaux résultats attendus

Les résultats attendus de cette analyse incluent : - Segments de clients clairement définis : identification de groupes homogènes de clients en fonction de leurs comportements d'achat et de leurs préférences.

- Amélioration des campagnes marketing : ciblage plus précis des campagnes marketing, menant à une meilleure adoption des produits et services ;
- Optimisation des ressources : allocation efficace des ressources marketing pour maximiser le retour sur investissement.
- Personnalisation accrue des offres : offres de produits et services mieux adaptées aux besoins spécifiques des différents segments de clientèle ;

Methodologie

1- Prétraitement des données : nettoyage et normalisation des données pour garantir leur qualité et leur fiabilité.

2- Segmentation :

Analyse en Composantes Principales (ACP). Pour réduire la dimensionnalité des données quantitatives et comprendre les structures sous-jacentes. Analyse des Correspondances Multiples (ACM) : pour les données qualitatives, afin de visualiser les relations entre les catégories.

3- Clustering : utilisation de techniques de clustering, comme le K-means, pour identifier des segments homogènes.

Dictionnaire du jeu de Données

- Id : identifiant unique pour chaque individu dans l'ensemble de données.
- Année_Naissance : L'année de naissance de l'individu.
- Éducation : Le niveau d'éducation le plus élevé atteint par l'individu.
- Marital_Status : L'état matrimonial de l'individu.
- Revenu : Le revenu annuel de l'individu.
- Kidhome : Le nombre de jeunes enfants dans le ménage.

- Teenhome : Le nombre d'adolescents dans le foyer.
- Dt_Customer : la date à laquelle le client a été inscrit pour la première fois ou est devenu une partie de la base de données de l'entreprise.
- Récence : Le nombre de jours depuis le dernier achat ou la dernière interaction.
- MntWines : Le montant dépensé en vins.
- MntFruits : Le montant dépensé en fruits.
- MntMeatProducts : Le montant dépensé en produits carnés.
- MntFishProducts : Le montant dépensé en produits de la pêche.
- MntSweetProducts : Le montant dépensé en produits sucrés.
- MntGoldProds : Le montant dépensé en produits aurifères.
- NumDealsPurchases : le nombre d'achats effectués avec une remise ou dans le cadre d'une offre.
- NumWebPurchases : Le nombre d'achats effectués via le site Web de l'entreprise.
- NumCatalogPurchases : Le nombre d'achats effectués via des catalogues.
- NumStorePurchases : Le nombre d'achats effectués dans les magasins physiques.
- NumWebVisitsMonth : Le nombre de visites sur le site Web de l'entreprise en un mois.
- AcceptedCmp3 : Indicateur binaire (1 ou 0) indiquant si l'individu a accepté la troisième campagne marketing.
- AcceptedCmp4 : Indicateur binaire (1 ou 0) indiquant si l'individu a accepté la quatrième campagne marketing.
- AcceptedCmp5 : Indicateur binaire (1 ou 0) indiquant si l'individu a accepté la cinquième campagne marketing.

- AcceptedCmp1 : Indicateur binaire (1 ou 0) indiquant si l'individu a accepté la première campagne marketing.
- AcceptedCmp2 : Indicateur binaire (1 ou 0) indiquant si l'individu a accepté la deuxième campagne marketing.
- Plainte : Indicateur binaire (1 ou 0) indiquant si la personne a déposé une plainte.
- Z_CostContact : un coût constant associé à la prise de contact avec un client.
- Z_Revenue : un revenu constant associé à une réponse de campagne réussie.
- Réponse : Indicateur binaire (1 ou 0) indiquant si l'individu a répondu à la campagne marketing

Notre analyse sera divisée comme suit :

I- Prétraitement des Données

II- Analyse des composantes principales

III- Analyse des composantes multiples

IV- Clustering

V- Recommandations

PARTIE 1 : PRETRAITEMENT DES DONNEES

a) Importation, restructuration et description du jeu de Données

```
tibble [2,240 × 29] (S3: tbl_df/tbl/data.frame)
  ID: int [1:2240] 5524 2174 4141 6182 5324 7446 965 6177 4855 5899 ...
  Year_Birth: int [1:2240] 1957 1954 1965 1984 1981 1967 1971 1985 1974
1950 ...
  Education: Factor w/ 5 levels "2n Cycle","Basic",...: 3 3 3 3 5 4 3 5
5 5 ...
  Marital_Status: Factor w/ 8 levels "Absurd","Alone",...: 5 5 6 6 4 6 3
4 6 6 ...
  Income: int [1:2240] 58138 46344 71613 26646 58293 62513 55635 33454
30351 5648 ...
  Kidhome: int [1:2240] 0 1 0 1 1 0 0 1 1 1 ...
  Teenhome: int [1:2240] 0 1 0 0 0 1 1 0 0 1 ...
  Dt_Customer: Factor w/ 663 levels "01-01-2013","01-01-2014",...: 81
158 445 200 391 190 281 161 120 268 ...
  Recency: int [1:2240] 58 38 26 26 94 16 34 32 19 68 ...
  MntWines: int [1:2240] 635 11 426 11 173 520 235 76 14 28 ...
  MntFruits: int [1:2240] 88 1 49 4 43 42 65 10 0 0 ...
  MntMeatProducts: int [1:2240] 546 6 127 20 118 98 164 56 24 6 ...
  MntFishProducts: int [1:2240] 172 2 111 10 46 0 50 3 3 1 ...
  MntSweetProducts: int [1:2240] 88 1 21 3 27 42 49 1 3 1 ...
  MntGoldProds: int [1:2240] 88 6 42 5 15 14 27 23 2 13 ...
  NumDealsPurchases: int [1:2240] 3 2 1 2 5 2 4 2 1 1 ...
  NumWebPurchases: int [1:2240] 8 1 8 2 5 6 7 4 3 1 ...
  NumCatalogPurchases: int [1:2240] 10 1 2 0 3 4 3 0 0 0 ...
  NumStorePurchases: int [1:2240] 4 2 10 4 6 10 7 4 2 0 ...
  NumWebVisitsMonth: int [1:2240] 7 5 4 6 5 6 6 8 9 20 ...
  AcceptedCmp3: int [1:2240] 0 0 0 0 0 0 0 0 0 1 ...
  AcceptedCmp4: int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
  AcceptedCmp5: int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
  AcceptedCmp1: int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
  AcceptedCmp2: int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
  Complain: int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
  Z_CostContact: int [1:2240] 3 3 3 3 3 3 3 3 3 3 ...
  Z_Revenue: int [1:2240] 11 11 11 11 11 11 11 11 11 11 ...
  Response: int [1:2240] 1 0 0 0 0 0 0 0 1 0 ...
```

Ce jeu de données est contient 2240 observations et 28 variables.

• Renommage des variables

```
tibble [2,240 × 29] (S3: tbl_df/tbl/data.frame)
  ID      : int [1:2240] 5524 2174 4141 6182 5324 7446 965 6177
```

```

4855 5899 ...
Année_naissance: int [1:2240] 1957 1954 1965 1984 1981 1967 1971 1985
1974 1950 ...
education      : Factor w/ 5 levels "2n Cycle","Basic",...: 3 3 3 3 5
4 3 5 5 5 ...
situation_mat  : Factor w/ 8 levels "Absurd","Alone",...: 5 5 6 6 4 6
3 4 6 6 ...
Revenu         : int [1:2240] 58138 46344 71613 26646 58293 62513
55635 33454 30351 5648 ...
Jeune_menage   : int [1:2240] 0 1 0 1 1 0 0 1 1 1 ...
ado_menage     : int [1:2240] 0 1 0 0 0 1 1 0 0 1 ...
date_client    : Factor w/ 663 levels "01-01-2013","01-01-2014",...:
81 158 445 200 391 190 281 161 120 268 ...
dern_achat     : int [1:2240] 58 38 26 26 94 16 34 32 19 68 ...
mtvins         : int [1:2240] 635 11 426 11 173 520 235 76 14 28 ...
mtfruit        : int [1:2240] 88 1 49 4 43 42 65 10 0 0 ...
mtviande       : int [1:2240] 546 6 127 20 118 98 164 56 24 6 ...
mtpoisson      : int [1:2240] 172 2 111 10 46 0 50 3 3 1 ...
mtsucré        : int [1:2240] 88 1 21 3 27 42 49 1 3 1 ...
mtor           : int [1:2240] 88 6 42 5 15 14 27 23 2 13 ...
nbachatremi    : int [1:2240] 3 2 1 2 5 2 4 2 1 1 ...
nbachatweb     : int [1:2240] 8 1 8 2 5 6 7 4 3 1 ...
nbachatcat     : int [1:2240] 10 1 2 0 3 4 3 0 0 0 ...
nbachatmag     : int [1:2240] 4 2 10 4 6 10 7 4 2 0 ...
nbvisitweb     : int [1:2240] 7 5 4 6 5 6 6 8 9 20 ...
ACmp3          : int [1:2240] 0 0 0 0 0 0 0 0 0 1 ...
ACmp4          : int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
ACmp5          : int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
ACmp1          : int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
ACmp2          : int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
Plainte        : int [1:2240] 0 0 0 0 0 0 0 0 0 0 ...
constantclient : int [1:2240] 3 3 3 3 3 3 3 3 3 3 ...
constantrevenu : int [1:2240] 11 11 11 11 11 11 11 11 11 11 ...
Reponse        : int [1:2240] 1 0 0 0 0 0 0 0 1 0 ...

```

• Conversion de la structure des variables

Avant restructuration

ID	Année_naissance	education	situation_mat	Revenu
"integer"	"integer"	"factor"	"factor"	"integer"
Jeune_menage	ado_menage	date_client	dern_achat	mtvins
"integer"	"integer"	"factor"	"integer"	"integer"
mtfruit	mtviande	mtpoisson	mtsucré	mtor
"integer"	"integer"	"integer"	"integer"	"integer"
nbachatremi	nbachatweb	nbachatcat	nbachatmag	nbvisitweb
"integer"	"integer"	"integer"	"integer"	"integer"
ACmp3	ACmp4	ACmp5	ACmp1	ACmp2
"integer"	"integer"	"integer"	"integer"	"integer"
Plainte	constantclient	constantrevenu	Reponse	
"integer"	"integer"	"integer"	"integer"	

Après restructuration

ID	Année_naissance	education	situation_mat	Revenu
"factor"	"numeric"	"factor"	"factor"	"integer"
Jeune_menage	ado_menage	date_client	dern_achat	mtvins
"integer"	"integer"	"numeric"	"integer"	"integer"
mtfruit	mtviande	mtpoisson	mtsucre	mtor
"integer"	"integer"	"integer"	"integer"	"integer"
nbachatremi	nbachatweb	nbachatcat	nbachatmag	nbvisitweb
"integer"	"integer"	"integer"	"integer"	"integer"
ACmp3	ACmp4	ACmp5	ACmp1	ACmp2
"factor"	"factor"	"factor"	"factor"	"factor"
Plainte	constantclient	constantrevenu	Reponse	
"factor"	"integer"	"integer"	"factor"	

b) Transformation des variables binaires et recodage des noms des modalités des variables “education” et “situation_mat”

• Variables binaires

"Non1"	"Oui1"
"Non2"	"Oui2"
"Non3"	"Oui3"
"Non4"	"Oui4"
"Non5"	"Oui5"
"Non plainte"	"Oui plainte"
"Non"	"Oui"

• Recodage des modalités des variables “éducation” et “situation_mat”

modalités variable “education”:

"2n Cycle"	"Basic"	"Graduation"	"Master"	"PhD"
------------	---------	--------------	----------	-------

Recodage “education” :

"2e Cycle"	"De base"	"Licence"	Master"	"Doctorat"
------------	-----------	-----------	---------	------------

Modalités variable “situation_mat”:

"Absurd"	"Alone"	"Divorced"	"Married"	"Single"	"Together"	"Widow"
"YOLO"						

Recodage “situation_mat” :

"Absurde"	"Seul"	"Divorcé"	"Marié"	"Célibataire"	"En couple"	"veuf"
-----------	--------	-----------	---------	---------------	-------------	--------

"YOLO"

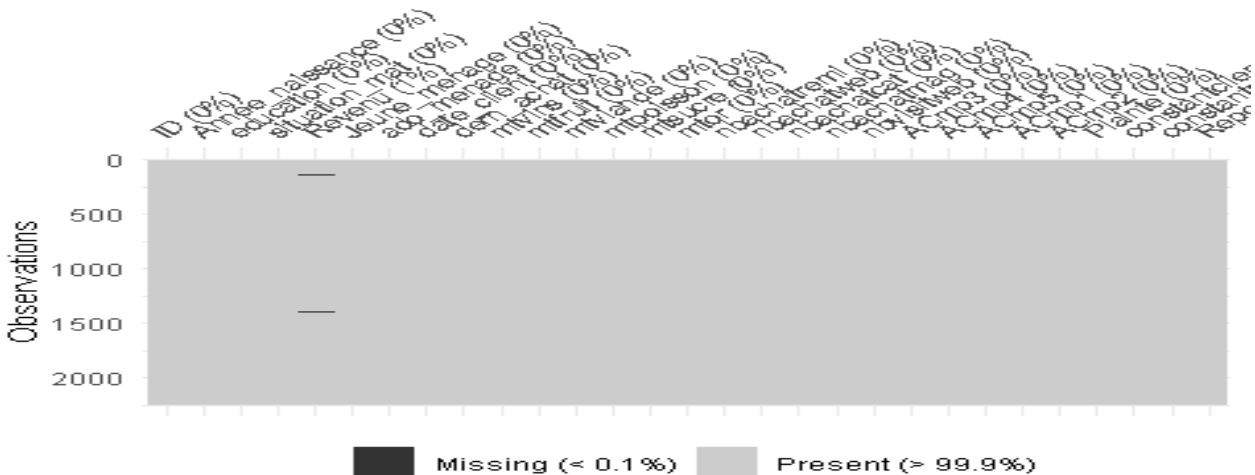
c) Traitement des doublons

Nombre de doublons trouvés et supprimés : 0

d) Traitement de valeurs manquantes

- **Visualisation des données manquantes avant traitement**

ID	Année_naissance	education	situation_mat	Revenu
0	0	0	0	24
Jeune_menage	ado_menage	date_client	dern_achat	mtvins
0	0	0	0	0
mtfruit	mtviande	mtpoisson	mtsucré	mtor
0	0	0	0	0
nbachatremi	nbachatweb	nbachatcat	nbachatmag	nbvisitweb
0	0	0	0	0
ACmp3	ACmp4	ACmp5	ACmp1	ACmp2
0	0	0	0	0
Plainte	constantclient	constantrevenu	Reponse	
0	0	0	0	

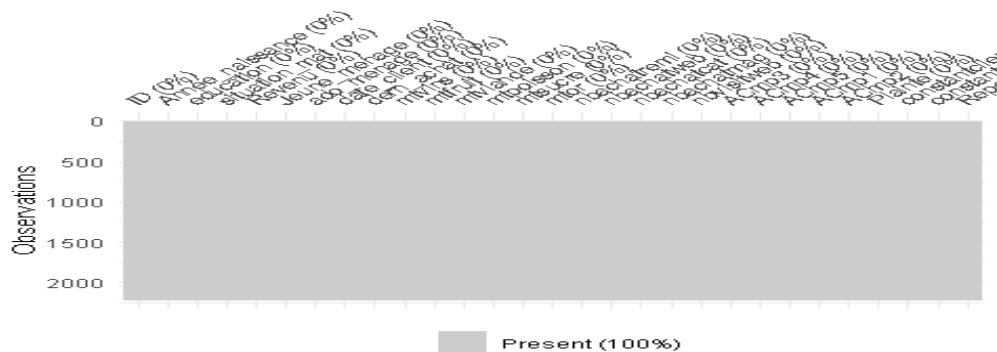


On observe que la variable « Revenu » se distingue par la présence de 24 valeurs manquantes. Par conséquent, 0,1 % des personnes ont des informations manquantes, il est donc nécessaire de les éliminer.

Résumé numérique de la variable “Revenu” avant suppression des données manquantes :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1730	35303	51382	52247	68522	666666	24

- **Visualisation des valeurs manquantes après traitement**



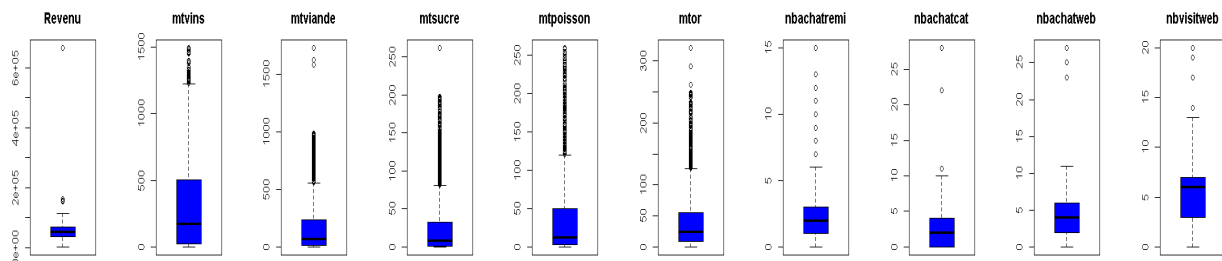
On peut voir que 100% des valeurs sont présentes

Résumé numérique de la variable “Revenu” après suppression des données manquantes :

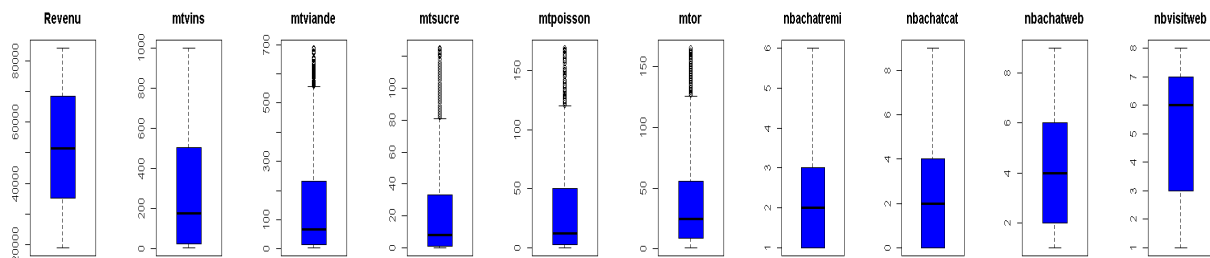
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1730	35303	51382	52247	68522	666666

e) Traitement des valeurs abberantes ou extrêmes

- **Visualisation des valeurs aberrantes avant traitement**



- **Visualisation des valeurs aberrantes après traitement**



Grâce aux informations prétraitées, nous sommes désormais prêts à démarrer les différentes analyses. Cette étape fondamentale permet de faciliter la compréhension et l'interprétation des informations collectées. De cette manière, nous obtenons des résultats exacts et appropriés. La fiabilité de nos analyses est garantie par le prétraitement des données. Nous allons maintenant passer à la prochaine étape de notre étude.

UNSSSED

PARTIE 2 : Analyse des composantes principales (ACP)

L'analyse en composantes principales (ACP) permet de synthétiser et de présenter un tableau de données en fonction des individus et des variables. Elle permet d'analyser les similitudes entre individus en considérant toutes les variables et de déterminer des profils de personnes. Elle permet également d'effectuer un bilan des relations linéaires entre les variables à l'aide des coefficients de corrélation. Étant donné que ces recherches se déroulent dans un contexte cohérent, elles sont interconnectées. Cela permet de caractériser les individus ou groupes d'individus par le biais de variables et de mettre en lumière les relations entre celles-ci à travers des individus spécifiques.

1) Importation des variables quantitatives

```
tibble [2,216 × 17] (S3: tbl_df/tbl/data.frame)
  Revenu      : num [1:2216] 58138 46344 71613 26646 58293 ...
  Jeune_menage : int [1:2216] 0 1 0 1 1 0 0 1 1 1 ...
  ado_menage   : int [1:2216] 0 1 0 0 0 1 1 0 0 1 ...
  dern_achat   : int [1:2216] 58 38 26 26 94 16 34 32 19 68 ...
  mtvins       : num [1:2216] 635 11 426 11 173 520 235 76 14 28 ...
  mtfruit      : int [1:2216] 88 1 49 4 43 42 65 10 0 0 ...
  mtviande     : num [1:2216] 546 6 127 20 118 98 164 56 24 6 ...
  mtpoisson    : num [1:2216] 169 2 111 10 46 0 50 3 3 1 ...
  mtsucre      : num [1:2216] 88 1 21 3 27 42 49 1 3 1 ...
  mtor         : num [1:2216] 88 6 42 5 15 14 27 23 2 13 ...
  nbachatremi  : num [1:2216] 3 2 1 2 5 2 4 2 1 1 ...
  nbachatweb   : num [1:2216] 8 1 8 2 5 6 7 4 3 1 ...
  nbachatcat   : num [1:2216] 9 1 2 0 3 4 3 0 0 0 ...
  nbachatmag   : int [1:2216] 4 2 10 4 6 10 7 4 2 0 ...
  nbvisitweb   : num [1:2216] 7 5 4 6 5 6 6 8 8 8 ...
  Année_naissance: num [1:2216] 1957 1954 1965 1984 1981 ...
  date_client  : num [1:2216] 81 158 445 200 391 190 281 161 120 268 ...
```

Cette analyse vise à identifier des profils de clients similaires : y a-t-il des groupes plus susceptibles d'effectuer des achats en ligne, plus attentifs aux offres promotionnelles ou possédant des préférences particulières pour certains genres d'articles ? Nous nous intéresserons également à l'existence d'habitudes d'achat communes : un client enclin à acheter des vins est-il aussi enclin à acheter des articles sucrés ou à profiter d'offres promotionnelles ?

2) Choisir les variables et les individus actifs

Afin d'établir des profils de comportements et de préférences chez nos clients, nous utilisons activement les facteurs liés aux habitudes d'achat et aux réactions face aux stratégies marketing. Il est essentiel de choisir les variables actives, car ce sont elles qui contribuent à la création des axes de l'ACP. En d'autres termes, ce sont elles qui permettent de déterminer les distances entre les individus.

Variables Actives

Revenu : Le revenu annuel de l'individu.

Jeune_menage : Le nombre de jeunes enfants dans le ménage.

ado_menage : Le nombre d'adolescents dans le foyer.

dern_achat : Le nombre de jours depuis le dernier achat ou la dernière Interaction.

Mtvins : Le montant dépensé en vins.

MtFruit : Le montant dépensé en fruits.

Mtviande : Le montant dépensé en produits carnés.

Mtpoisson : Le montant dépensé en produits de la pêche.

Mtsucre : Le montant dépensé en produits sucrés.

Mtor : Le montant dépensé en produits aurifères.

Nbatchatremi : Le nombre d'achats effectués avec une remise ou dans le cadre d'une offre.

Nbatchatweb : Le nombre d'achats effectués via le site Web de l'entreprise.

Nbatchatcat : Le nombre d'achats effectués via des catalogues.

Nbatchatmag : Le nombre d'achats effectués dans les magasins physiques.

Nbvisitweb : Le nombre de visites sur le site Web de l'entreprise en un mois.

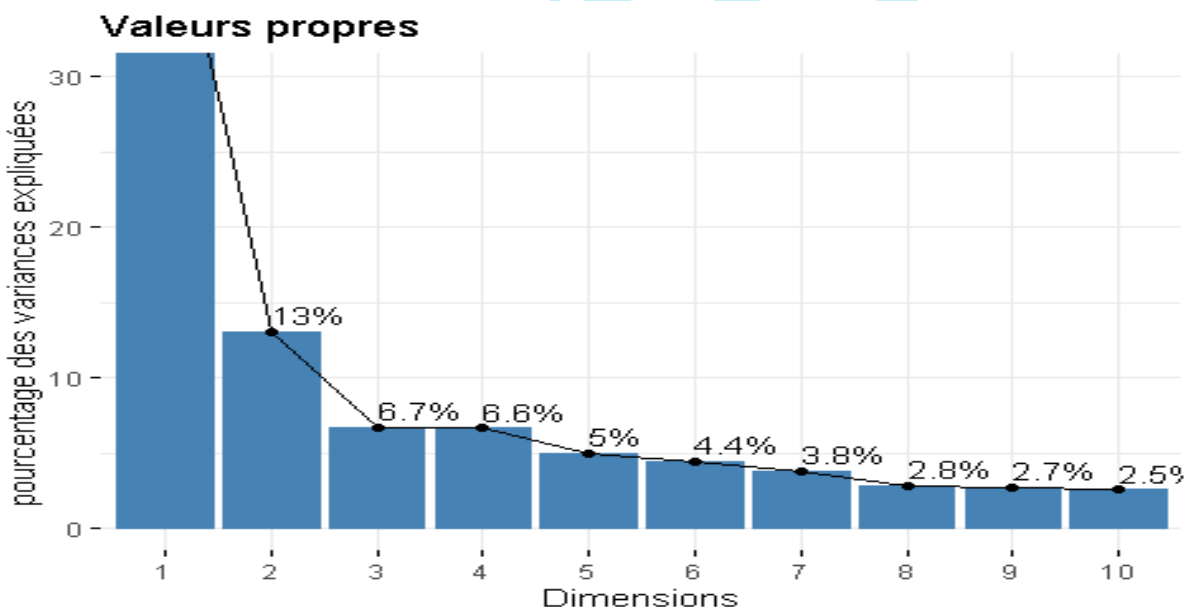
Variables Supplémentaires

Année_Naissance : L'année de naissance de l'individu.

date_client : La date à laquelle le client a été inscrit pour la première fois ou est devenu une partie de la base de données de l'entreprise.

3) Choix du nombre des axes

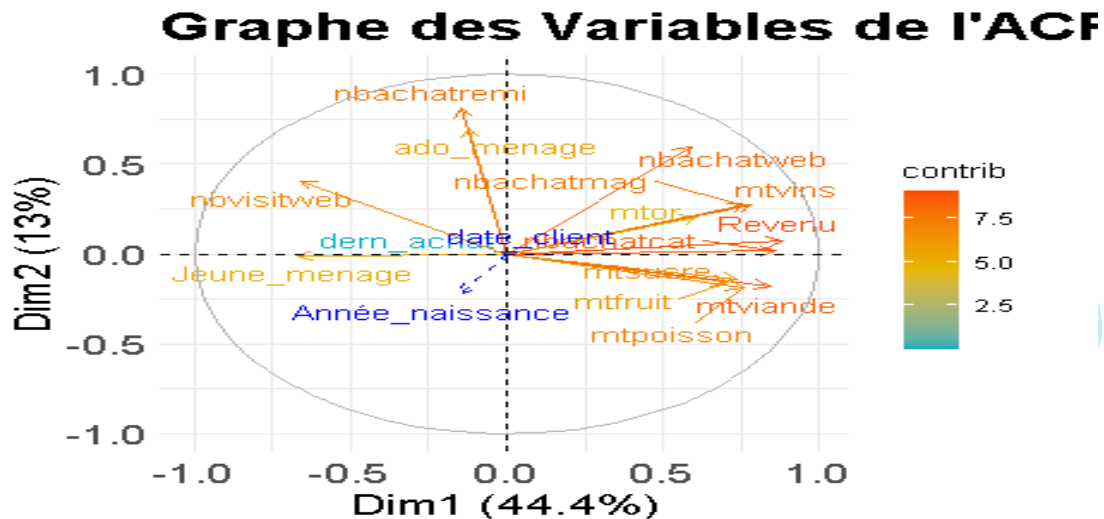
Il existe différentes méthodes pour déterminer le nombre d'axes à examiner en ACP. La méthode la plus répandue consiste à représenter le diagramme sous forme de barres pour les valeurs propres ou les inerties liées à chaque axe :



Les deux premières dimensions représentent 57 % de l'inertie totale, ce qui signifie que 57 % des informations présentes dans le tableau de données se trouvent dans ces deux premiers axes. Cela implique que la variété des profils de performance ne peut se résumer à deux aspects seulement.

4) Analyse des variables

Les variables quantitatives supplémentaires apparaissent en pointillés



Pour la dimension 1 oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage

- de fortes valeurs pour les variables nbachatremi, nbachatweb, ado_menage, nbachatmag, mtvins, mtor, nbvisitweb, Revenu et nbachatcat (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables Année_naissance et Jeune_menage (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée positive sur l'axe) partage

- de fortes valeurs pour les variables mtviande, Revenu, nbachatcat, mtpoisson, mtsucre, mtfruit, mtvins, nbachatmag, mtor et nbachatweb (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables nbvisitweb, Jeune_menage, nbachatremi, ado_menage et Année_naissance (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnée négative sur l'axe) partage

- de fortes valeurs pour les variables Jeune_menage, nbvisitweb et Année_naissance (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que Revenu, mtvins, nbachatmag, nbachatcat, nbachatweb, mtviande, mtpoisson, mtor, mtsucre et mtfruit (de la plus extrême à la moins extrême).

Pour la dimension 2 oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (en haut du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (en bas du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage

- de fortes valeurs pour les variables nbachatremi, nbachatweb, ado_menage, nbachatmag, mtvins, mtor, nbvisitweb, Revenu et nbachatcat (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables Année_naissance et Jeune_menage (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée négative sur l'axe) partage

- de fortes valeurs pour les variables Jeune_menage, nbvisitweb et Année_naissance (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que Revenu, mtvins, nbachatmag, nbachatcat, nbachatweb, mtviande, mtpoisson, mtor, mtsucre et mtfruit (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnée négative sur l'axe) partage

- de fortes valeurs pour les variables mtviande, Revenu, nbachatcat, mtpoisson, mtsucre, mtfruit, mtvins, nbachatmag, mtor et nbachatweb (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables nbvisitweb, Jeune_menage, nbachatremi, ado_menage et Année_naissance (de la plus extrême à la moins extrême).

Résumés numériques

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	6.665	1.952	1.004	0.994	0.747	0.658	0.564
% of var.	44.433	13.014	6.696	6.628	4.978	4.384	3.763
Cumulative % of var.	44.433	57.447	64.143	70.770	75.749	80.133	83.896
	Dim.8	Dim.9	Dim.10	Dim.11	Dim.12	Dim.13	Dim.14
Variance	0.422	0.411	0.380	0.352	0.300	0.235	0.197
% of var.	2.812	2.738	2.534	2.348	1.998	1.568	1.315
Cumulative % of var.	86.708	89.446	91.980	94.328	96.326	97.894	99.209
	Dim.15						
Variance	0.119						
% of var.	0.791						
Cumulative % of var.	100.000						

Individuals (the 15 first)

	Dist	Dim.1	ctr	cos2	Dim.2	ctr	cos2
1	5.480	4.057	0.111	0.548	0.084	0.000	0.000
2	2.961	-2.477	0.042	0.700	-0.410	0.004	0.019
3	3.368	2.095	0.030	0.387	-0.265	0.002	0.006
4	3.035	-2.597	0.046	0.732	-1.006	0.023	0.110
5	2.960	-0.157	0.000	0.003	0.545	0.007	0.034
6	2.774	1.008	0.007	0.132	1.177	0.032	0.180
7	2.549	0.924	0.006	0.132	1.516	0.053	0.354
8	2.823	-2.325	0.037	0.678	-0.302	0.002	0.011
9	3.490	-2.889	0.057	0.685	-1.046	0.025	0.090
10	4.006	-3.516	0.084	0.770	-0.551	0.007	0.019
11	3.389	-2.413	0.039	0.507	-1.176	0.032	0.121
12	4.855	3.245	0.071	0.447	-2.358	0.129	0.236
13	2.313	-1.291	0.011	0.312	1.292	0.039	0.312
14	3.575	-2.602	0.046	0.530	-1.374	0.044	0.148
15	4.463	3.422	0.079	0.588	0.060	0.000	0.000

Variables

	Dim.1	ctr	cos2	Dim.2	ctr	cos2
Revenu	0.881	11.645	0.776	0.071	0.258	0.005
Jeune_menage	-0.667	6.682	0.445	-0.013	0.008	0.000
ado_menage	-0.122	0.223	0.015	0.692	24.550	0.479
dern_achat	0.016	0.004	0.000	0.009	0.004	0.000
mtvins	0.784	9.226	0.615	0.272	3.777	0.074
mtfruit	0.708	7.518	0.501	-0.161	1.322	0.026
mtviande	0.841	10.619	0.708	-0.181	1.678	0.033
mtpoisson	0.756	8.587	0.572	-0.185	1.761	0.034
mtsucré	0.738	8.183	0.545	-0.138	0.974	0.019
mtor	0.606	5.502	0.367	0.204	2.126	0.042
nbachatremi	-0.147	0.323	0.022	0.806	33.261	0.649
nbachatweb	0.588	5.186	0.346	0.595	18.116	0.354
nbachatcat	0.860	11.103	0.740	0.018	0.016	0.000
nbachatmag	0.762	8.703	0.580	0.270	3.724	0.073
nbvisitweb	-0.658	6.497	0.433	0.406	8.425	0.164

Supplementary continuous variables

	Dim.1	cos2	Dim.2	cos2
Année_naissance	-0.144	0.021	-0.221	0.049
date_client	-0.016	0.000	0.009	0.000

Composante principale 1 (Dim 1)

Revenu :

Contribution : 11.663% (élevée) \cos^2 : 0.780 (élevé) Interprétation : “Revenu” est une variable clé pour cette composante, expliquant une grande partie de la variance.

Mtvins :

Contribution : 9.231% (élevée) \cos^2 : 0.618 (élevé) Interprétation : La consommation de vin est également significative pour cette composante.

Mtviande :

Contribution : 10.515% (élevée) \cos^2 : 0.703 (élevé) Interprétation : La consommation de viande est une variable importante pour cette dimension.

En analysant les variables et leurs contributions pour cette première composante, il semble que la composante principale 1 (Dim.1) soit fortement influencée par des variables liées au revenu et aux habitudes de consommation, telles que la consommation de vin et de viande. On pourrait donc nommer cette composante comme suit :

“Niveau de vie et habitudes de consommation”.

Composante principale 2 (Dim 2)

ado_menage: La présence d'adolescents dans le ménage est un facteur crucial, contribuant à 24.550% et ayant un \cos^2 élevé de 0.479.

nbachatremi : Le nombre d'achats en remise est également significatif, contribuant à 33.261% avec un \cos^2 élevé de 0.649.

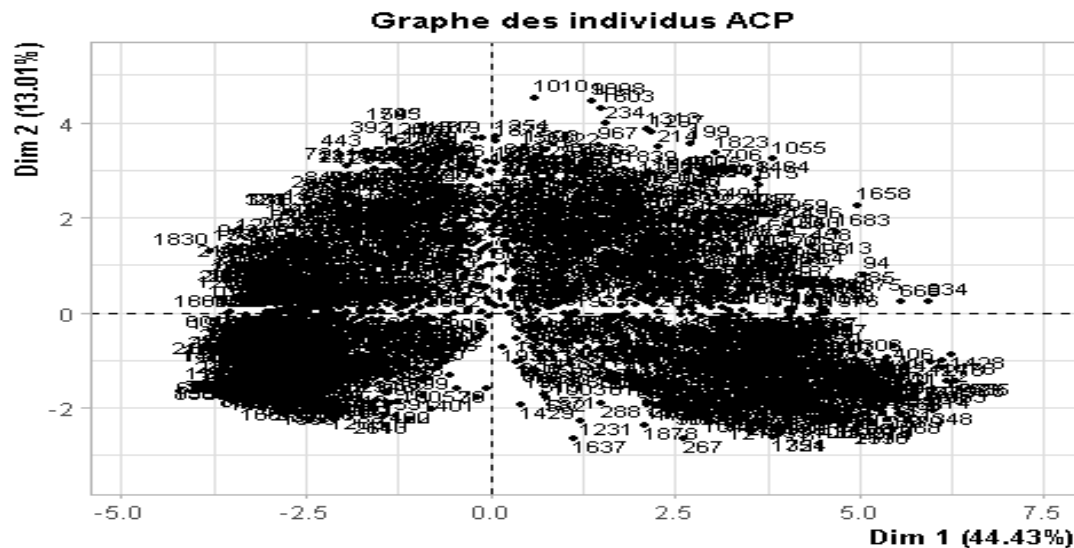
nbachatweb : Le nombre d'achats en ligne contribue aussi, avec 18.116% et un \cos^2 de 0.354.

En analysant les variables et leurs contributions pour la deuxième composante principale (Dim.2), il semble que cette composante soit fortement influencée par des facteurs liés à la présence d'adolescents

dans le ménage et aux pratiques d'achat telles que les achats en remise et en ligne. On pourrait donc nommer cette composante comme suit :

“Mode de vie des familles avec adolescents et leurs comportements d'achat”

5) Analyse du graphe des individus

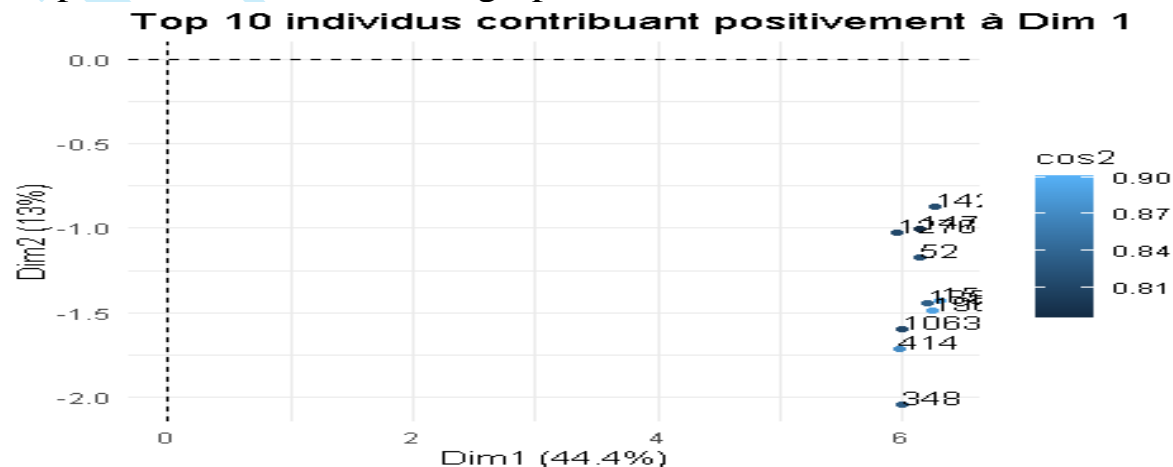


DIM 1 :

Les 10 individus qui contribuent le plus positivement à la construction de l'axe 1 sont :

1585	1428	1966	1853	52	1476	348	1063	414	1276
------	------	------	------	----	------	-----	------	-----	------

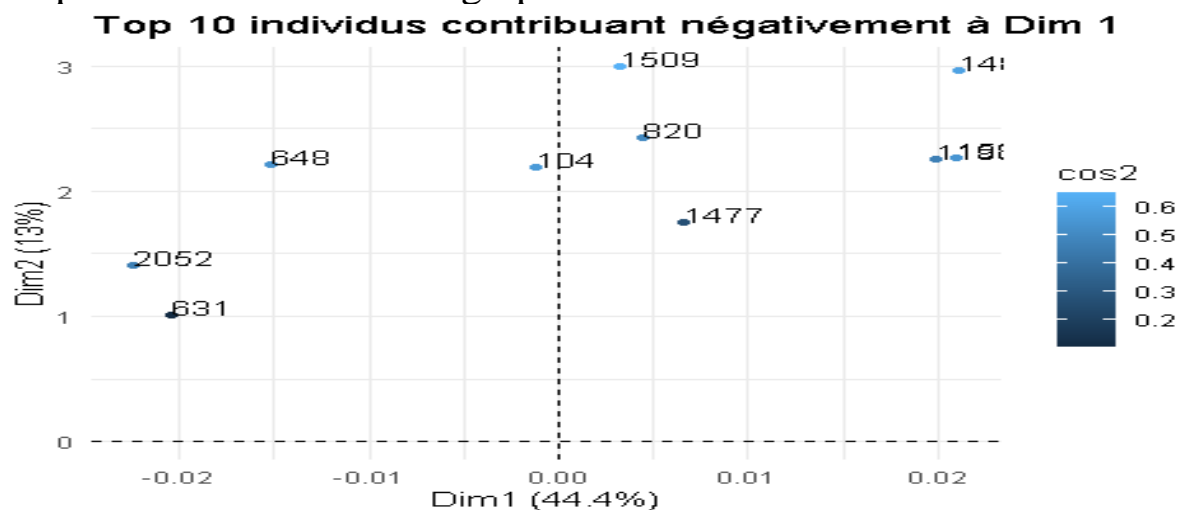
On peut les visualiser sur le graphe suivant :



Les 10 individus qui contribuent le plus négativement à la construction de l'axe 1 sont :

104 1509 820 1477 648 1193 631 1584 1481 2052

On peut les observer sur le graphe suivant :

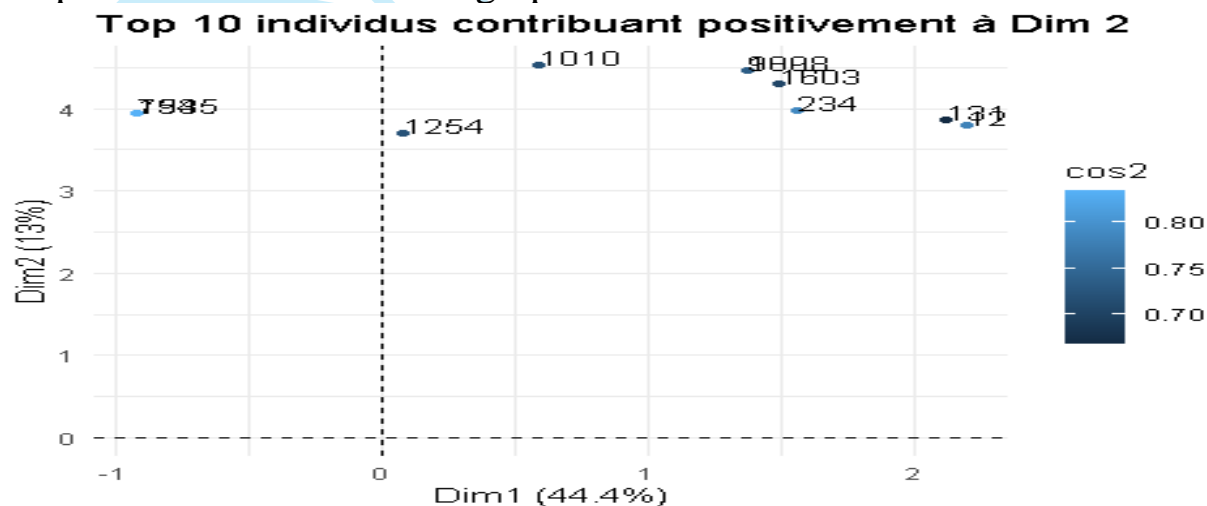


DIM 2 :

Les 10 individus qui contribuent le plus positivement à la construction de l'axe 2 sont :

1010 900 1898 1603 234 793 1545 1313 1297 1254

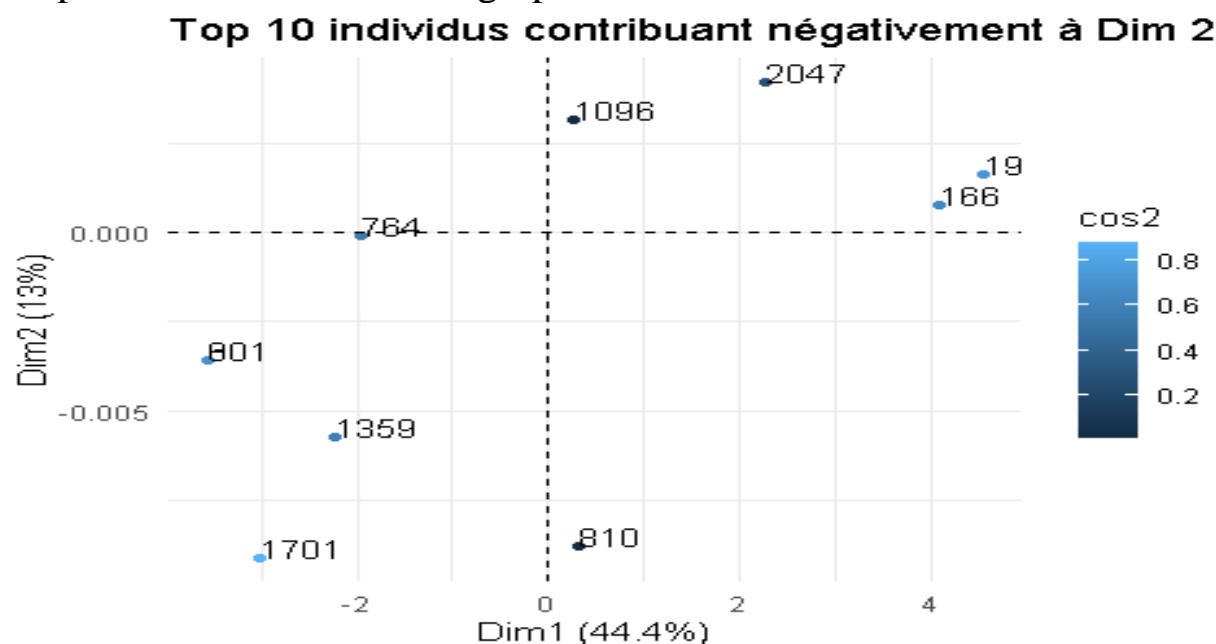
On peut les visualiser sur le graphe suivant :



Les 10 individus qui contribuent le plus négativement à la construction de l'axe 2 sont :

764	166	1976	1096	301	601	2047	1359	810	1701
-----	-----	------	------	-----	-----	------	------	-----	------

On peut les visualiser sur le graphe suivant :



6) Conclusion de l'Analyse ACP

L'analyse en composantes principales (ACP) met en lumière des lignes d'opposition importantes, facilitant ainsi une segmentation efficace des clients selon leurs niveaux de vie et leurs habitudes de consommation. Les axes principaux révèlent des divergences significatives. La Dimension 1 oppose les profils exubérants ou impulsifs, comme ceux recherchant des expériences uniques, des produits originaux ou des services innovants, aux profils pragmatiques et ordinaires, qui privilégient des achats pratiques, fonctionnels et orientés vers des besoins essentiels. Les coordonnées positives de la Dimension 1 montrent que le Groupe 1 partage des valeurs élevées pour des variables telles que le nombre d'achats en remise, en ligne, les ménages avec adolescents, les achats en magasin, la consommation de vin et autres produits, les visites sur des sites web, le revenu et le nombre d'achats par

catégorie. Le Groupe 2 partage également des valeurs élevées pour des variables liées aux dépenses alimentaires, telles que la viande, le poisson, le sucre et les fruits, ainsi que les achats en magasin et en ligne. En revanche, les coordonnées négatives de la Dimension 1 montrent que le Groupe 3 se distingue par des valeurs élevées pour les variables représentant les jeunes ménages, les visites sur des sites web et l'année de naissance, indiquant peut-être une population plus jeune. Ils ont de faibles valeurs pour le revenu et les dépenses dans divers produits. La Dimension 2 oppose les individus « seuls », préférant des offres individualisées, des solutions simples et rapides, ou des expériences orientées sur le développement personnel, aux individus en couple, qui privilégient des produits partagés, des abonnements familiaux ou des activités adaptées aux familles. Les coordonnées positives de la Dimension 2 montrent que le Groupe 1 partage des valeurs élevées pour des variables relatives aux achats en remise, en ligne, les ménages avec adolescents, les achats en magasin, la consommation de vin et autres produits, ainsi que les visites sur des sites web. En revanche, les coordonnées négatives de la Dimension 2 montrent que le Groupe 2 se distingue par des valeurs élevées pour les jeunes ménages, les visites sur des sites web et l'année de naissance, et que le Groupe 3 partage des valeurs élevées pour des variables liées aux dépenses alimentaires, telles que la viande, le poisson, le sucre et les fruits, ainsi que les achats en magasin et en ligne. Les individus comme 1585, 1428, 1966, etc., se situent du côté positif de la Dimension 1, caractérisés par un revenu élevé et des habitudes de consommation variées, tandis que les individus comme 104, 1509, 820, etc., sont du côté négatif de la Dimension 1, représentant des jeunes ménages avec des comportements de consommation plus modestes. Les individus comme 1010, 900, 1898, etc., sont fortement positifs sur la Dimension 2, représentant des ménages avec adolescents, tandis que les individus comme 764, 166, 1976, etc., sont négatifs sur la Dimension 2, se concentrant sur des comportements d'achat traditionnels. En fin de compte, l'ACP permet de segmenter efficacement les clients en fonction de leurs niveaux de vie et de leurs habitudes de consommation, en identifiant les groupes distincts tels que ceux à revenu élevé avec des habitudes de consommation

sophistiquées, incluant des achats diversifiés en ligne et en magasin, et les jeunes ménages avec un pouvoir d'achat plus limité, mettant davantage l'accent sur les achats en ligne. Les stratégies marketing devraient cibler les ménages à revenu élevé avec des produits haut de gamme, des offres personnalisées et des expériences exclusives, et attirer les jeunes ménages avec des offres promotionnelles attractives, des produits économiques, et des campagnes axées sur les achats en ligne et les valeurs économiques.

PARTIE 3 : Analyse des correspondances multiple (ACM)

L'objectif de l'Analyse des Correspondances Multiples (ACM) est de résumer un tableau de données où les individus sont décrits par des variables qualitatives. L'ACM permet d'étudier les ressemblances entre individus du point de vue de l'ensemble des variables et de dégager des profils d'individus. Elle permet également de faire un bilan des liaisons entre variables et d'étudier les associations de modalités.

1) Importations des variables qualitatives

```
tibble [2,216 × 9] (S3: tbl_df/tbl/data.frame)
  education      : Factor w/ 5 levels "2e Cycle","De base",...: 3 3 3 3 5 4 3 5
5 5 ...
  situation_mat: Factor w/ 8 levels "Absurde","Seul",...: 5 5 6 6 4 6 3 4 6 6
...
  ACmp1         : Factor w/ 2 levels "Non1","Oui1": 1 1 1 1 1 1 1 1 1 1 ...
  ACmp2         : Factor w/ 2 levels "Non2","Oui2": 1 1 1 1 1 1 1 1 1 1 ...
  ACmp3         : Factor w/ 2 levels "Non3","Oui3": 1 1 1 1 1 1 1 1 1 2 ...
  ACmp4         : Factor w/ 2 levels "Non4","Oui4": 1 1 1 1 1 1 1 1 1 1 ...
  ACmp5         : Factor w/ 2 levels "Non5","Oui5": 1 1 1 1 1 1 1 1 1 1 ...
  Plainte       : Factor w/ 2 levels "Non plainte",...: 1 1 1 1 1 1 1 1 1 1
...
  Reponse       : Factor w/ 2 levels "Non","Oui": 2 1 1 1 1 1 1 1 2 1 ...
- attr(, "na.action")= 'omit' Named int [1:24] 11 28 44 49 59 72 91 92 93
129 ...
..- attr(, "names")= chr [1:24] "11" "28" "44" "49" ...
```

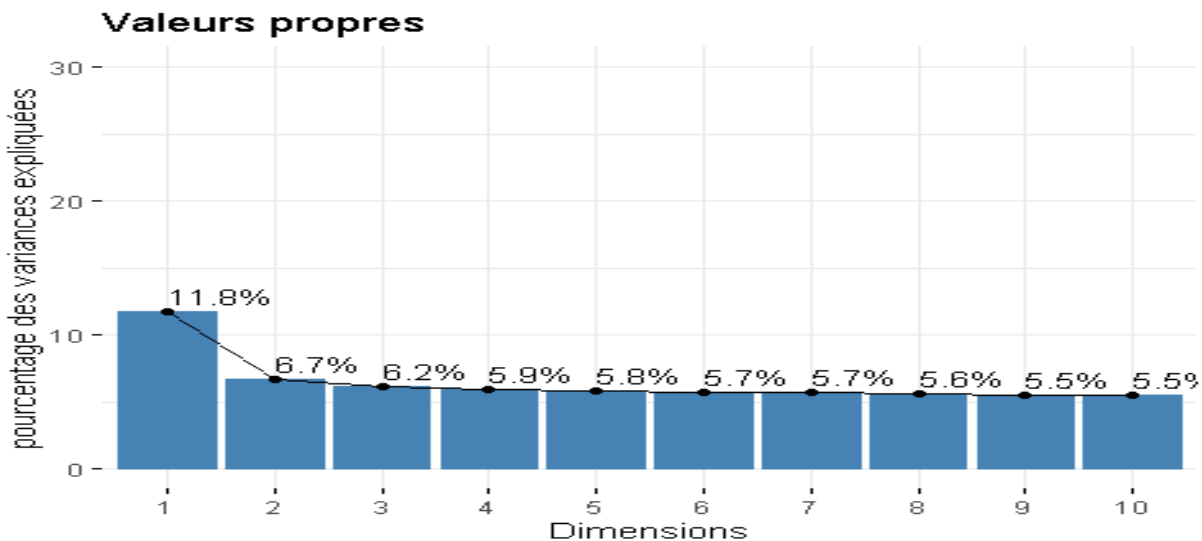
Résumés numériques

education	situation_mat	ACmp1	ACmp2	ACmp3
2e Cycle: 200	Marié :857	Non1:2074	Non2:2186	Non3:2053
De base : 54	En couple :573	Oui1: 142	Oui2: 30	Oui3: 163
Licence :1116	Célibataire:471			
Master : 365	Divorcé :232			
Doctorat: 481	Widow : 76			
	Seul : 3			
	(Other) : 4			
ACmp4	ACmp5	Plainte	Reponse	
Non4:2052	Non5:2054	Non plainte:2195	Non:1883	
Oui4: 164	Oui5: 162	Oui plainte: 21	Oui: 333	

Cette analyse vise à identifier des profils de clients similaires : y a-t-il des groupes avec des niveaux d'éducation spécifiques ou des états matrimoniaux particuliers qui sont plus susceptibles d'accepter les

campagnes marketing, de déposer des plaintes, ou de répondre positivement aux offres promotionnelles ? Nous nous intéresserons également à l'existence de comportements similaires parmi les clients : un client ayant accepté une campagne marketing est-il également susceptible de répondre à d'autres campagnes ou de déposer des plaintes en cas d'insatisfaction ?

2) Choix du nombre d'axes



Ces faibles taux d'inertie et cette décroissance sont très classiques en ACM. Le tableau des pourcentages d'inertie expliquée par chaque axe est donné par :

```
Call :  
MCA(X = segments_quali, graph = FALSE)
```

Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	0.235	0.134	0.123	0.118	0.116	0.114	0.114
% of var.	11.768	6.688	6.151	5.907	5.815	5.723	5.716
Cumulative % of var.	11.768	18.456	24.607	30.514	36.329	42.052	47.768
	Dim.8	Dim.9	Dim.10	Dim.11	Dim.12	Dim.13	Dim.14
Variance	0.112	0.110	0.110	0.108	0.106	0.104	0.102
% of var.	5.598	5.517	5.501	5.406	5.293	5.179	5.086
Cumulative % of var.	53.365	58.882	64.384	69.790	75.082	80.261	85.347
	Dim.15	Dim.16	Dim.17	Dim.18			
Variance	0.092	0.071	0.066	0.063			
% of var.	4.618	3.546	3.314	3.175			
Cumulative % of var.	89.965	93.511	96.825	100.000			

Individuals (the 10 first)									
	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	
1	0.214	0.009	0.038	0.481	0.078	0.195	-0.110	0.004	
2	-0.200	0.008	0.069	0.120	0.005	0.025	-0.164	0.010	
3	-0.232	0.010	0.111	-0.233	0.018	0.112	-0.146	0.008	
4	-0.232	0.010	0.111	-0.233	0.018	0.112	-0.146	0.008	
5	-0.184	0.006	0.053	-0.034	0.000	0.002	0.397	0.058	
6	-0.242	0.011	0.062	-0.254	0.022	0.069	-0.012	0.000	
7	-0.198	0.008	0.035	0.017	0.000	0.000	0.006	0.000	
8	-0.184	0.006	0.053	-0.034	0.000	0.002	0.397	0.058	
9	0.226	0.010	0.037	0.231	0.018	0.038	0.430	0.068	
10	0.030	0.000	0.000	0.681	0.156	0.214	0.246	0.022	
cos2									
1	0.010								
2	0.047								
3	0.044								
4	0.044								
5	0.249								
6	0.000								
7	0.000								
8	0.249								
9	0.134								
10	0.028								
Categories (the 10 first)									
	Dim.1	ctr	cos2	v.test	Dim.2	ctr	cos2		
v.test									
2e Cycle	-0.192	0.156	0.004	-2.840	-0.041	0.013	0.000	-	
0.611									
De base	-0.695	0.555	0.012	-5.168	0.936	1.773	0.022		
6.961									
Licence	-0.002	0.000	0.000	-0.077	-0.091	0.344	0.008	-	
4.300									
Master	-0.044	0.015	0.000	-0.917	-0.160	0.352	0.005	-	
3.352									
Doctorat	0.195	0.389	0.011	4.826	0.244	1.075	0.017		
6.052									
Absurde	2.425	0.251	0.005	3.430	0.115	0.001	0.000		
0.163									
Seul	0.021	0.000	0.000	0.036	5.331	3.196	0.039		
9.238									
Divorcé	0.081	0.033	0.001	1.310	0.344	1.032	0.014		
5.544									
Marié	-0.052	0.049	0.002	-1.938	-0.156	0.785	0.015	-	
5.843									
Célibataire	0.074	0.055	0.001	1.808	0.684	8.250	0.126		
16.715									
	Dim.3	ctr	cos2	v.test					
2e Cycle	-0.778	4.932	0.060	-11.530					
De base	-0.474	0.495	0.006	-3.526					
Licence	-0.391	6.946	0.155	-18.525					
Master	0.033	0.016	0.000	0.693					
Doctorat	1.258	31.034	0.439	31.178					
Absurde	-5.313	2.301	0.026	-7.516					
Seul	2.874	1.010	0.011	4.980					
Divorcé	0.324	0.990	0.012	5.206					
Marié	-0.091	0.291	0.005	-3.410					

Célibataire | -0.215 0.890 0.013 -5.265 |

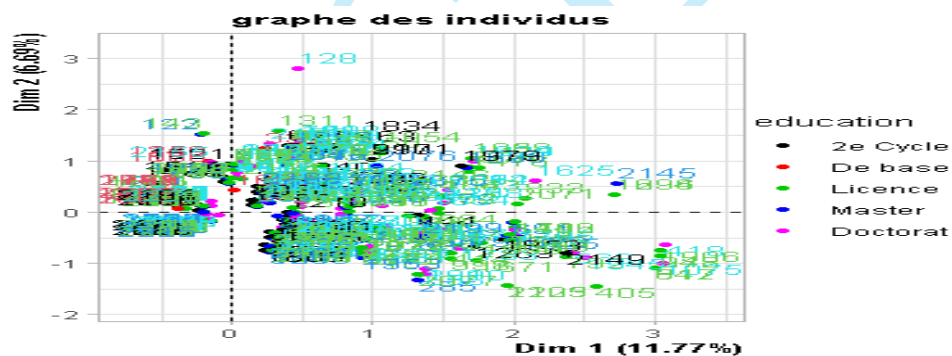
Categorical variables (eta2)

	Dim.1	Dim.2	Dim.3
education	0.024	0.043	0.481
situation_mat	0.013	0.229	0.401
ACmp1	0.456	0.001	0.030
ACmp2	0.275	0.044	0.001
ACmp3	0.061	0.487	0.011
ACmp4	0.342	0.199	0.011
ACmp5	0.529	0.010	0.012
Plainte	0.002	0.009	0.158
Reponse	0.416	0.181	0.004

Les deux premiers axes expriment 18 % de l'inertie totale : autrement dit, 18 % de l'information du tableau de données est résumée par les deux premières dimensions, ce qui est relativement important dans le cadre d'une ACM.

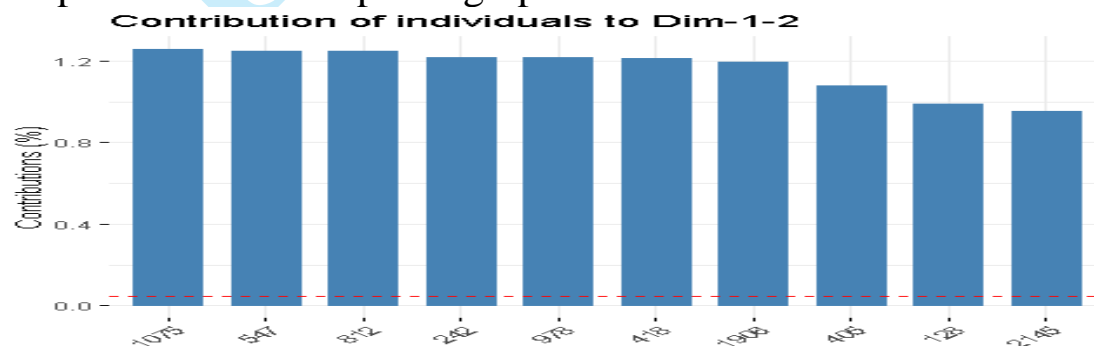
3) Analyse des résultats

Graphe des individus



Les 10 individus qui contribuent le plus aux à la construction des axes sont : 1075 547 812 242 978 418 1906 405 128 2145

On peut le visualiser par le graphe suivant :



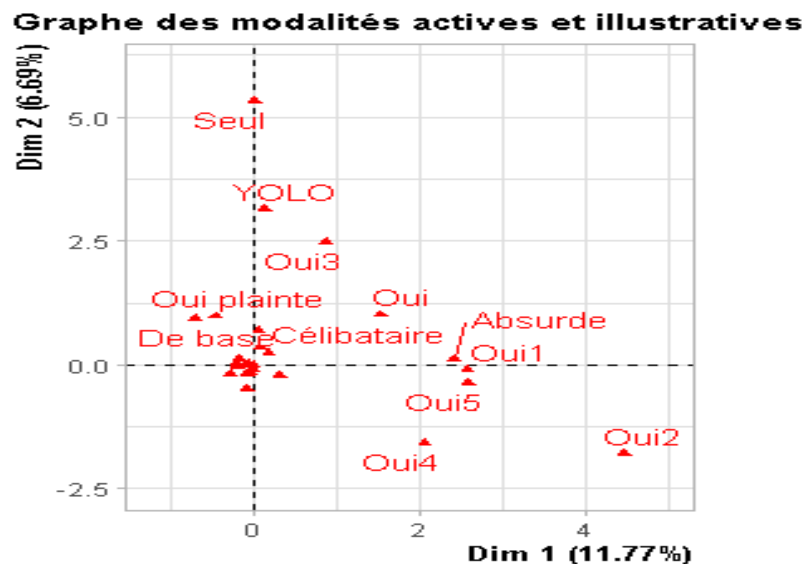
Résumés numériques de ces dix individus :

#	education	situation_mat	ACmp1	ACmp2	ACmp3	ACmp4	ACmp5	Plainte	Reponse
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>
1	Doctorat	En couple	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
2	Licence	En couple	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
3	Licence	En couple	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
4	Licence	Divorcé	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
5	Licence	Divorcé	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
6	Doctorat	Célibataire	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
7	Licence	Célibataire	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Oui
8	Licence	En couple	Oui1	Oui2	Non3	Oui4	Oui5	Non plainte	Non
9	Doctorat	Seul	Non1	Non2	Oui3	Non4	Non5	Non plainte	Oui
10	Master	Célibataire	Oui1	Oui2	Oui3	Non4	Oui5	Non plainte	Oui

On peut s'apercevoir que qu'ils ont à peu près les mêmes comportements face aux campagnes marketing.

Graphe des modalités actives

Le graphique avec l'ensemble des modalités des variables qualitatives actives et illustratives permet d'apprécier les grandes tendances qui se dégagent de l'analyse.

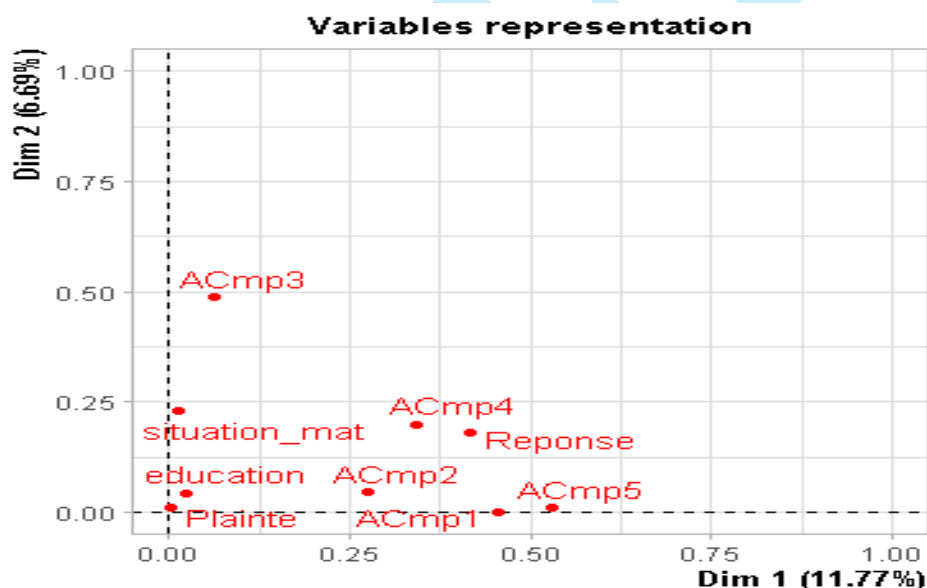


Le premier axe (Dim 1) oppose des individus adoptant une posture « absurde » (à droite) à ceux dans une posture plus « de base » (à gauche). Ce contraste pourrait représenter des comportements liés au mode de vie ou aux choix de consommation, où des profils plus exubérants ou impulsifs (par exemple, des personnes recherchant des expériences

uniques, des produits originaux ou des services innovants) s'opposent à des profils pragmatiques. Et ordinaires (par exemple, des individus privilégiés des achats pratiques, fonctionnels et orientés vers des besoins essentiels)

Le deuxième axe (Dim 2) oppose des individus « seuls » (en bas) à ceux en couple (en haut). Cet axe pourrait être qualifié de lien social ou familial, avec des implications sur les préférences pour des produits ou services destinés aux célibataires (comme des offres individualisées, des solutions simples et rapides ou des expériences orientées sur le développement personnel) versus ceux destinés à des ménages ou familles (comme des produits partagés, des abonnements familiaux ou des activités adoptées aux familles).

Graphes des variables



La variable qui contribue le plus à la création de l'axe 1 est la variable Acmp5, celle qui contribue le plus à la création de l'axe 2 est la variable Acmp3. Cette information résume l'influence globale de toutes les modalités de chacune des variables sur la construction des axes ; ainsi, les modalités de la variable Acmp5 contribuent fortement à la création de l'axe 1 et les modalités de la variable Acmp3 à la création de l'axe 2.

Lien entre les variables et les modalités

`Dim 1`

Link Between the variable and the categorical variable (1-way anova)

=====

	R2	p.value
ACmp5	0.528995670	0.000000e+00
ACmp1	0.456350271	2.498891e-295
Reponse	0.416108713	5.609731e-261
ACmp4	0.342181693	1.273770e-203
ACmp2	0.274876082	9.672913e-157
ACmp3	0.061218848	2.895028e-32
education	0.023626995	8.984839e-11
situation_mat	0.012923102	1.605080e-04
Plainte	0.001967246	3.681851e-02

Link between variable and the categories of the categorical variables

=====

	Estimate	p.value
ACmp5=Oui5	0.67775902	0.000000e+00
ACmp1=Oui1	0.66912550	2.498891e-295
Reponse=Oui	0.43788806	5.609731e-261
ACmp4=Oui4	0.54203179	1.273770e-203
ACmp2=Oui2	1.10049923	9.672913e-157
ACmp3=Oui3	0.22991181	2.895028e-32
education=Doctorat	0.16600114	1.319253e-06
situation_mat=Absurde	0.99824849	5.957440e-04
Plainte=Non plainte	0.11104764	3.681851e-02
Plainte=Oui plainte	-0.11104764	3.681851e-02
education=2e Cycle	-0.02142957	4.484236e-03
situation_mat=Widow	-0.02279614	4.478235e-03
education=De base	-0.26555148	2.196693e-07
ACmp3=Non3	-0.22991181	2.895028e-32
ACmp2=Non2	-1.10049923	9.672913e-157
ACmp4=Non4	-0.54203179	1.273770e-203
Reponse=Non	-0.43788806	5.609731e-261
ACmp1=Non1	-0.66912550	2.498891e-295
ACmp5=Non5	-0.67775902	0.000000e+00

`Dim 2`

Link between the variable and the categorical variable (1-way anova)

=====

	R2	p.value
ACmp3	0.487003032	2.964394e-323
situation_mat	0.229206465	4.688801e-120
ACmp4	0.199154703	6.415865e-109
Reponse	0.181018809	3.924606e-98
ACmp2	0.044417904	1.143102e-23
education	0.042827241	4.668379e-20
ACmp5	0.010440409	1.436377e-06
Plainte	0.009164933	6.348601e-06

Link between variable and the categories of the categorical variables

=====

Estimate	p.value
----------	---------

ACmp3=Oui3	0.48885985	2.964394e-323
ACmp4=Non4	0.31173926	6.415865e-109
Reponse=Oui	0.21773156	3.924606e-98
ACmp2=Non2	0.33350331	1.143102e-23
situation_mat=Seul	1.54937559	1.097071e-20
education=De base	0.27735887	2.604151e-12
education=Doctorat	0.02438099	1.237318e-09
ACmp5=Non5	0.07178073	1.436377e-06
Plainte=Oui plainte	0.18069434	6.348601e-06
situation_mat=YOLO	0.74899455	8.414525e-06
situation_mat=Widow	-0.48226776	4.740212e-02
education=Master	-0.12360902	7.922198e-04
education=Licence	-0.09811585	1.652596e-05
Plainte=Non plainte	-0.18069434	6.348601e-06
ACmp5=Oui5	-0.07178073	1.436377e-06
situation_mat=Divorcé	-0.27452596	2.672551e-08
situation_mat=Marié	-0.45769207	4.538104e-09
ACmp2=Oui2	-0.33350331	1.143102e-23
situation_mat=En couple	-0.57505263	9.803933e-42
situation_mat=Célibataire	-0.15050221	7.214795e-67
Reponse=Non	-0.21773156	3.924606e-98
ACmp4=Oui4	-0.31173926	6.415865e-109
ACmp3=Non3	-0.48885985	2.964394e-323

`Dim 3`

Link between the variable and the categorical variable (1-way anova)

=====

	R2	p.value
education	0.480755329	1.175163e-312
situation_mat	0.400609172	4.872833e-240
Plainte	0.157613300	1.481483e-84
ACmp1	0.029615795	3.420089e-16
ACmp5	0.012196613	1.869842e-07
ACmp3	0.011466117	4.362613e-07
ACmp4	0.010597171	1.197010e-06
Reponse	0.003735845	3.997876e-03

Link between variable and the categories of the categorical variables

=====

	Estimate	p.value
education=Doctorat	0.46592674	4.383354e-280
situation_mat=Widow	0.21569342	1.085160e-96
situation_mat=YOLO	4.11399731	1.366621e-86
Plainte=Non plainte	0.71860569	1.481483e-84
ACmp1=Non1	0.12323506	3.420089e-16
ACmp5=Non5	0.07440173	1.869842e-07
ACmp3=Non3	0.07193513	4.362613e-07
situation_mat=Seul	0.43723088	5.962849e-07
ACmp4=Oui4	0.06896131	1.197010e-06
Reponse=Oui	0.02999634	3.997876e-03
Reponse=Non	-0.02999634	3.997876e-03
situation_mat=Marié	-0.60277296	6.414069e-04
education=De base	-0.14162191	4.159041e-04
situation_mat=En couple	-0.62620850	1.076316e-05
ACmp4=Non4	-0.06896131	1.197010e-06
ACmp3=Oui3	-0.07193513	4.362613e-07

ACmp5=Oui5	-0.07440173	1.869842e-07
situation_mat=Divorcé	-0.45730573	1.782635e-07
situation_mat=Célibataire	-0.64628845	1.295758e-07
situation_mat=Absurde	-2.43434596	3.972460e-14
ACmp1=Oui1	-0.12323506	3.420089e-16
education=2e Cycle	-0.24816264	1.201076e-31
education=Licence	-0.11241328	5.068583e-83
Plainte=Oui plainte	-0.71860569	1.481483e-84

Ce tableau est très utile lorsqu'il y a beaucoup de modalités. On retrouve ici que le premier axe est surtout lié aux modalités "Oui1", "Non1", "Oui5" et "Non5" des variables Acmp1(Indicateur binaire indiquant si l'individu a accepté la première campagne marketing) et Acmp5(Indicateur binaire indiquant si l'individu a accepté la cinquième campagne marketing.). Ensuite le deuxième axe est lié à la modalité "YOLO" de la variable situation_mat. Aussi, le troisième axe est lié à la modalité "Ouiplainte" de la variable Plainte.

4) Conclusion

La réalisation de l'ACM met en lumière des lignes d'opposition importantes qui facilitent une segmentation efficace des clients selon leurs comportements et leurs goûts. Elle apporte ainsi des indications utiles pour le ciblage des stratégies marketing, l'amélioration des ressources et la personnalisation des propositions de produits. Les axes majeurs mettent en évidence des divergences significatives, en particulier entre les individus adoptant un comportement « absurde » et ceux adoptant un comportement « de base » (Dim 1), ainsi qu'entre les individus « seuls » et ceux en couple (Dim 2). Ces axes mettent en lumière divers comportements associés au style de vie, aux décisions d'achat et aux relations familiales ou sociales. Les axes 1 et 2 sont fortement influencés par les variables Acmp5 et Acmp3, tandis que les modalités « Oui1 », « Non1 », « Oui5 » et « Non5 » des variables Acmp1 et Acmp5 jouent un rôle crucial dans le premier axe, et que la modalité « YOLO » de la variable situation_mat est essentielle pour le second. Les dix personnes qui ont le plus contribué au développement des axes sont : 1075, 547, 812, 242, 978, 418, 1906, 405, 128 et 2145. En saisissant ces segments, les entreprises peuvent élaborer des stratégies marketing plus précises, optimiser la distribution des ressources et

adapter les propositions de produits afin de mieux répondre aux attentes des diverses catégories de clients. En définitive, cette ACM offre une base solide pour des stratégies marketing ciblées et efficaces, renforçant ainsi le lien avec les clients tout en optimisant les ressources de la société.

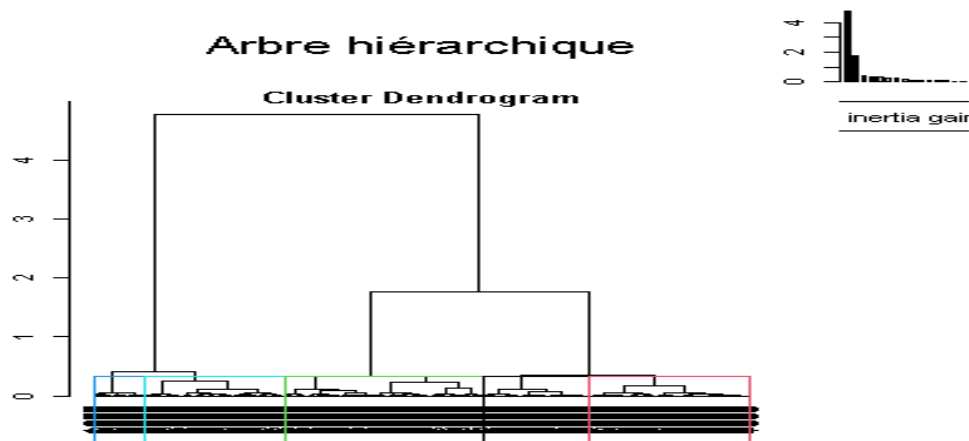
UNSSSED\$

PARTIE IV: Clustering

La Classification Ascendante Hiérarchique (CAH) a pour objectif de construire une hiérarchie sur les individus et se présente sous la forme d'un dendrogramme. Cette classification permet de regrouper des individus dans des classes les plus homogènes possibles à partir d'un jeu de données individus×variables. Cette méthode nécessite de choisir : une distance entre individus, ou plus généralement une dissimilarité et un indice d'agrégation entre les classes.

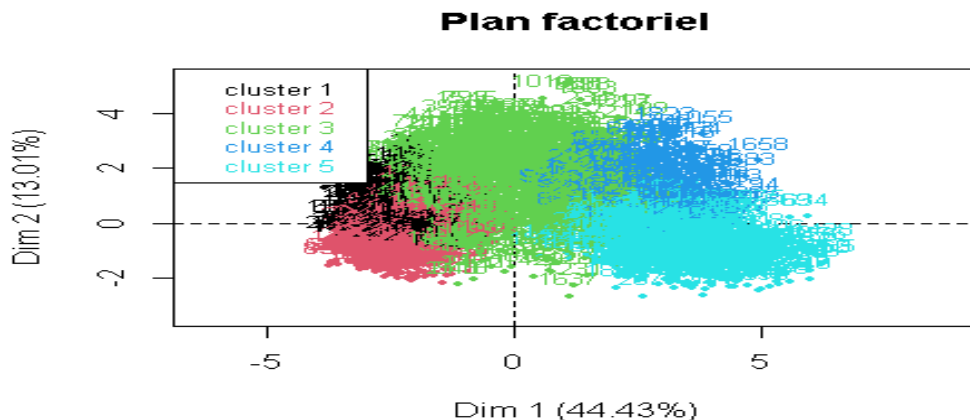
1) Classification sur ACP

Arbre hiérarchique



La classification réalisée sur les individus fait apparaître 5 classes.

- **Classification hiérarchique des individus**



La classe 1 est composé d'individus partageant :

- de fortes valeurs pour les variables classe, Jeune_menage, Année_naissance et nbvisitweb (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que Revenu, ado_menage, nbachatmag, mtvins, nbachatcat, nbachatweb, mtviande, mtor, mtpoisson et mtsucre (de la plus extrême à la moins extrême).

La classe 2 est composé d'individus partageant :

- de fortes valeurs pour les variables ado_menage, Jeune_menage, dern_achat et nbvisitweb (de la plus extrême à la moins extrême).
- de faibles valeurs pour des variables telles que nbachatweb, mtvins, nbachatcat, mtviande, nbachatmag, mtor, mtpoisson, mtsucre, mtfruit et Revenu (de la plus extrême à la moins extrême).

La classe 3 est composé d'individus partageant :

- de fortes valeurs pour les variables nbachatremi, nbachatweb, ado_menage, nbvisitweb, mtvins, nbachatmag, classe, mtor et Revenu (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables mtsucre, mtpoisson, mtfruit, Année_naissance, mtviande, Jeune_menage et dern_achat (de la plus extrême à la moins extrême).

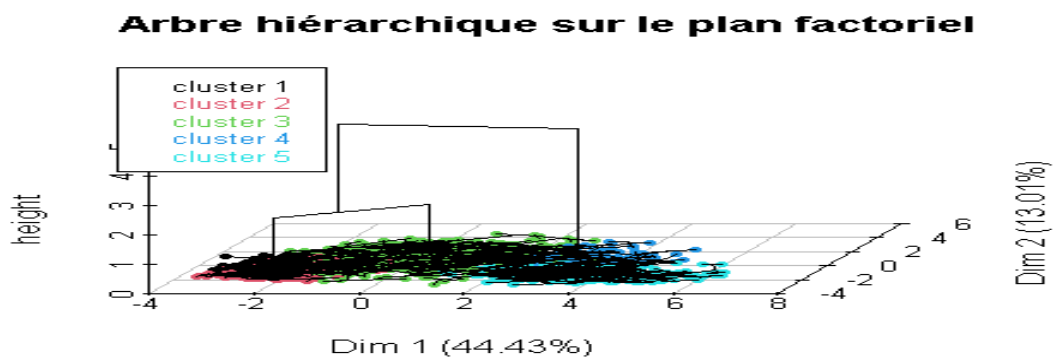
La classe 4 est composé d'individus partageant :

- de fortes valeurs pour des variables telles que nbachatweb, mtsucre, nbachatmag, mtpoisson, mtor, mtfruit, nbachatcat, Revenu, mtvins et nbachatremi (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables Jeune_menage et Année_naissance (de la plus extrême à la moins extrême).

La classe 5 est composé d'individus partageant :

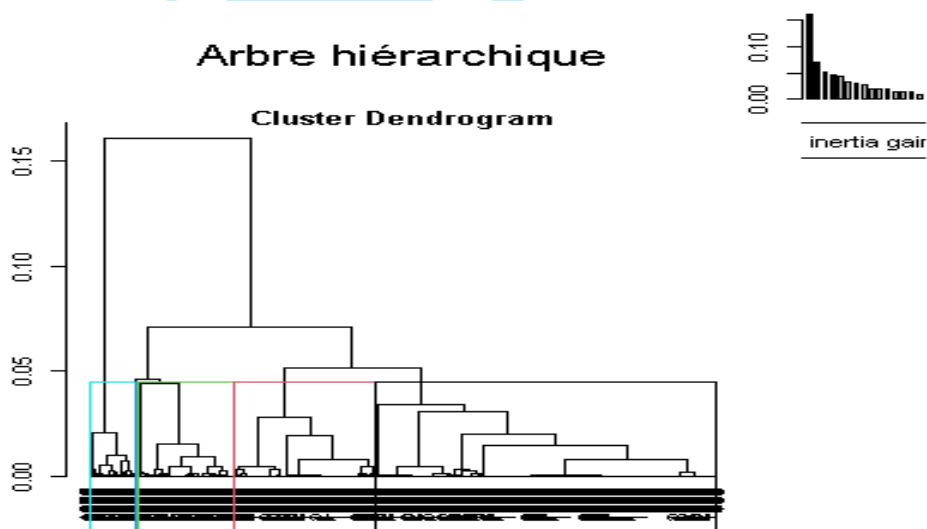
- de fortes valeurs pour les variables mtviande, nbachatcat, Revenu, mtpoisson, mtsucre, mtfruit, mtvins, nbachatmag, mtor et nbachatweb (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables classe, nbvisitweb, Jeune_menage, ado_menage et nbachatremi (de la plus extrême à la moins extrême).

Projection sur le plan factoriel



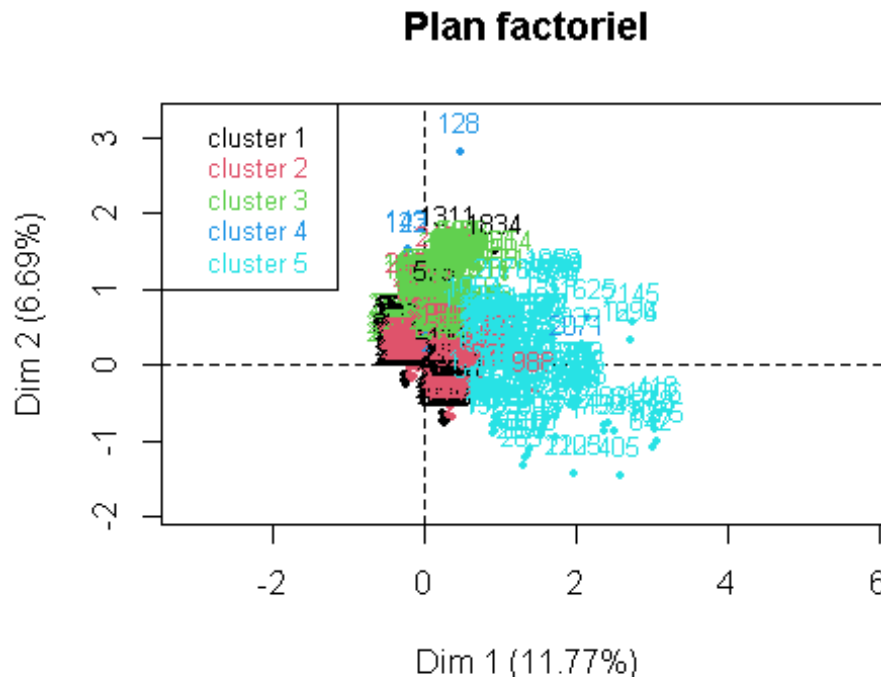
2) Classification sur ACM

Arbre hiérarchique



La classification fait apparaitre 5 classes.

- **Classification hiérarchique des individus**



Classe 1:

Composée d'individus adoptant une éducation « de base », ces profils pragmatiques et ordinaires privilégient des achats pratiques, fonctionnels et orientés vers des besoins essentiels. Ils ne recherchent pas forcément des expériences ou produits innovants, mais plutôt ce qui est utile au quotidien.

Classe 2 :

Composée d'individus « seuls », ces profils tendent à préférer des offres individualisées, des solutions simples et rapides, et des expériences orientées sur le développement personnel. Leurs choix de consommation sont influencés par leur statut relationnel.

Classe 3 :

Composée d'individus adoptant une posture « absurde », ces profils exubérants et impulsifs recherchent des expériences uniques, des

produits originaux et des services innovants. Ils se démarquent par leurs choix de consommation plus audacieux et créatifs.

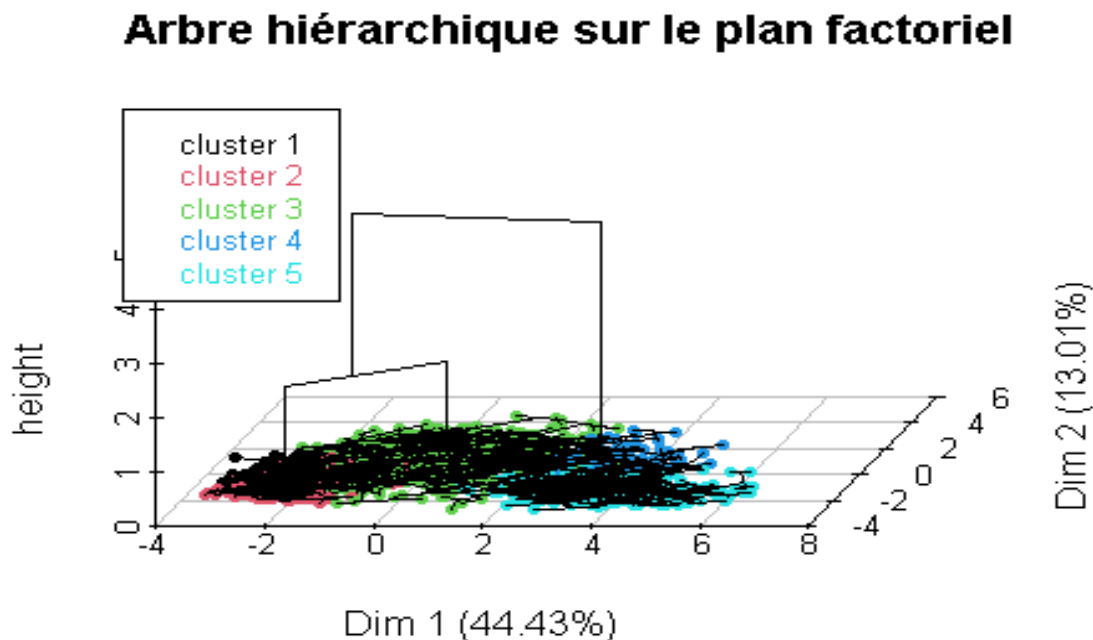
Classe 4 :

Composée d'individus en couple, cette classe privilégie des produits partagés, des abonnements familiaux et des activités adaptées aux familles. Leurs décisions d'achat sont influencées par les besoins collectifs et la dynamique familiale.

Classe 5 :

Composée d'individus réactifs aux campagnes marketing, cette classe regroupe des individus ayant répondu de manière significative aux campagnes marketing (variables Acmp5 et Acmp3). Ils montrent une forte interaction avec les initiatives marketing, influençant ainsi les stratégies à adopter pour toucher ce segment.

Projection sur le plan factoriel



Conclusion générale

En associant l'analyse en composantes principales (ACP) et l'analyse des correspondances multiples (ACM), la recherche révèle des points de divergence notables qui permettent une segmentation efficace des clients en fonction de leur mode de vie, de leurs habitudes et de leurs préférences. Selon l'ACP, la Dimension 1 distingue les profils exubérants et impulsifs qui aspirent à des expériences singulières et des articles uniques ou novateurs, de ceux pragmatiques et conventionnels qui favorisent des acquisitions fonctionnelles et utiles orientées vers des exigences cruciales. La Dimension 2 met en opposition les personnes « seules », qui privilégient des services personnalisés et des réponses rapides, aux couples, qui favorisent des biens communs et des activités conçues pour la vie familiale. Cette analyse est complétée par l'ACM qui met en évidence des différences notables entre les individus adoptant un comportement « absurde » et ceux qui adoptent une conduite « de base », aussi bien entre les individus considérés comme « seuls » que ceux en couple. Ces oppositions impliquent des variables telles que Acmp5, Acmp3, « Oui1 », « Non1 », « Oui5 » et « Non5 ».

En repérant des catégories de clients spécifiques, comme les foyers à haut revenu qui adoptent des modes de consommation avancés, comprenant une variété d'achats sur internet et en magasin, ou les jeunes foyers qui ont un pouvoir d'achat plus restreint, privilégiant principalement les acquisitions en ligne, cette analyse contribue à une meilleure compréhension du comportement et des préférences des clients. Les foyers comprenant des adolescents et des couples ont une préférence pour les biens communs et les moments familiaux, tandis que les jeunes foyers et les individus « seuls » privilégient des propositions sur mesure et des réponses promptes.

Il est donc possible d'affiner les stratégies de marketing afin de toucher plus spécifiquement ces groupes : des articles de qualité supérieure, des promotions sur mesure et des expériences uniques pourraient séduire les foyers fortunés et les individus « seuls » grâce à des promotions attrayantes, des articles abordables et des campagnes orientées vers les

achats en ligne. En améliorant la répartition des moyens et en adaptant les propositions de produits, les entreprises peuvent mieux satisfaire les exigences des diverses catégories de clients, consolidant par conséquent le rapport avec ces derniers tous en maximisant la répartition des ressources au sein de l'entreprise. Cette méthode globale crée une fondation robuste pour des tactiques marketing spécifiques et performantes, augmentant ainsi le lien avec les clients et la performance globale de l'entreprise.

Recommandations

Pour mieux cibler les clients et améliorer les offres de produits, en tenant compte des acceptations des campagnes marketing (campagnes 1 et 5) et des plaintes possibles des clients, voici les suggestions pour chaque classe :

Classe 1 : Profils pragmatiques et ordinaires

- Variables importantes : fortes valeurs pour classe, Jeune_menage, Année_naissance, nbvisitweb ; acceptations et plaintes à prendre en compte : plaintes concernant des produits fonctionnels, acceptations pour des offres pratiques.
- Offres de produits : Produits pratiques et fonctionnels, articles de première nécessité, promotions sur les produits de consommation courante. Veiller à la qualité des produits de base pour minimiser les plaintes. Proposer des programmes de fidélité pour les achats réguliers et des canaux de retour simplifiés pour les clients insatisfaits. Mettre en avant les campagnes marketing 1 et 5 pour ces produits pratiques afin d'augmenter l'acceptation.

Classe 2 : Individus « seuls »

- Variables importantes : fortes valeurs pour ado_menage, Jeune_menage, dern_achat, nbvisitweb ; acceptations et plaintes à prendre en compte : acceptations pour des solutions rapides, plaintes concernant des offres non adaptées à un mode de vie individuel.

- Offres de produits : Offres individualisées, solutions rapides et simples, expériences orientées sur le développement personnel, produits économiques et promotionnels. Veiller à proposer des produits et des services adaptés à une vie individuelle. Des options de personnalisation et un support client réactif peuvent réduire les plaintes. Utiliser les campagnes marketing 1 et 5 pour promouvoir des solutions rapides et personnalisées.

Classe 3 : Profils exubérants et impulsifs

- Variables importantes : fortes valeurs pour nbachatremi, nbachatweb, ado_menage, nbvisitweb, mtvins, nbachatmag, classe, mtor, Revenu ; acceptations et plaintes à prendre en compte : acceptations pour des produits innovants, plaintes concernant des expériences non satisfaisantes.

- Offres de produits : Expériences uniques, produits originaux et innovants, éditions limitées, services exclusifs, offres promotionnelles sur les achats impulsifs. S'assurer que les produits innovants sont de haute qualité pour éviter les plaintes. Proposer des essais gratuits ou des démonstrations pour limiter les retours. Utiliser les campagnes marketing 1 et 5 pour attirer l'attention sur les produits uniques et innovants.

Classe 4 : Individus en couple

- Variables importantes : fortes valeurs pour nbachatweb, mtsucre, nbachatmag, mtpoisson, mtor, mtfruit, nbachatcat, Revenu, mtvins, nbachatremi ; acceptations et plaintes à prendre en compte : acceptations pour des offres familiales, plaintes concernant des produits non adaptés aux besoins collectifs.

- Offres de produits : Produits partagés, abonnements familiaux, activités et services adaptés aux familles, packs promotionnels pour les ménages, offres groupées pour les couples. Assurer la qualité des produits destinés aux familles pour minimiser les plaintes. Offrir des services après-vente efficaces pour résoudre les problèmes rapidement. Promouvoir ces offres familiales à travers les campagnes marketing 1 et 5.

Classe 5 : Réactifs aux campagnes marketing

- Variables importantes : fortes valeurs pour mtviande, nbachatcat, Revenu, mtpoisson, mtsucre, mtfruit, mtvins, nbachatmag, mtor, nbachatweb ; acceptations et plaintes à prendre en compte : acceptations pour des campagnes ciblées, plaintes liées à la saturation des offres marketing.
- Offres de produits : Campagnes marketing ciblées, offres personnalisées, programmes de fidélité exclusifs, newsletters avec offres spéciales, récompenses pour l'engagement avec les initiatives marketing, promotions ciblées basées sur les historiques d'achat. Éviter de surcharger ces clients avec des offres marketing excessives pour réduire les plaintes. Proposer des options de désinscription faciles des listes de diffusion. Utiliser les campagnes marketing 1 et 5 pour offrir des promotions spéciales et des récompenses de fidélité.