



19/02/2023
Maser 2 MoSEF

CRM Analytics

Segmentation RFM

ABABII Anisoara, KOFFI Eunice, DIAKITE Gaoussou

Suivi par : Jérôme MOLLIER

TABLE DES MATIERES

INTRODUCTION..... 2

PARTIE 1 : PREPARATION DES DONNEES 3

1- Importation des bases de données 3

2- Analyse des valeurs manquantes..... 4

3- Audit des données..... 4

4- Construction de variables 6

PARTIE 2 : SEGMENTATION..... 8

1- Indicateur Récence :..... 8

2- Indicateur Fréquence : 9

3- Indicateur montant : 10

4- Croisements 11

CONCLUSION 13

ANNEXE..... 14

INTRODUCTION

En tant que des étudiants en Data Science, nous savons que les données jouent un rôle clé dans tous les secteurs, et cela est particulièrement vrai dans le domaine du marketing. C'est dans cette optique que nous avons réalisé ce rapport en utilisant les connaissances acquises lors de nos cours de SAS, CRM et Dataviz.

Parmi les différents sujets proposés, notre choix s'est porté sur la segmentation RFM, qui est une méthode de segmentation client largement utilisée en marketing. Cette méthode permet de regrouper les clients en différentes classes en fonction de leur récence, fréquence et montant d'achat. En analysant les données de cette manière, nous pouvons obtenir un aperçu des clients qui ont contribué au chiffre d'affaires de l'entreprise au cours des deux dernières années, et ainsi aider les entreprises à mieux comprendre leur clientèle.

Dans cette étude, nous allons effectuer une analyse approfondie des données, créer des variables pertinentes, puis segmenter les clients en différents groupes pour mieux comprendre leurs comportements d'achat. Nous montrerons ainsi comment l'utilisation de la segmentation RFM peut aider les entreprises à mieux comprendre leurs clients et à développer des stratégies marketing efficaces.

Partie 1 : Préparation des données

1- Importation des bases de données

Nous disposons de deux bases de données en CSV :

- Clients

Elle comprend les informations des clients du magasin. Elle contient 5096 lignes et 7 colonnes.

Nous avons les informations suivantes :

- ID_CLIENT : l'identifiant du client
- GENRE : Femme ou Homme
- DATE_NAISSANCE : Date de naissance du client
- ACTIF : 1 si le compte du client est ouvert, 0 sinon
- DATE_CREATION_COMPTE : Date de création du compte
- INSCRIT_NL : 1 si le client est inscrit à la newsletter et 0 s'il ne l'est pas
- A_ETE_PARRAINE : 1 si le client a été parrainé par un autre 0 sinon

Nous avons donc une ligne par clients.

- Commandes

Elle nous communique des informations sur les commandes passées au cours de la période considérée. Elle contient 11242 lignes et 8 colonnes.

- ID_CLIENT : l'identifiant du client
- NUMERO_COMMANDE : Numéro de commande
- DATE : Date de commandes
- MONTANT_DES_PRODUITS : Montants des produits achetés
- REMISE_SUR_PRODUITS : Remise faite aux clients sur le montant des articles achetés.
- MONTANT_LIVRAISON : Montant de la livraison
- REMISE_SUR_LIVRAISON : Montant total des livraisons
- MONTANT_TOTAL_PAYE : Montant total de la commande. Elle est égale à la somme des montants des produits, des livraisons diminué des remises.

Dans cette table, nous pouvons donner plusieurs lignes par client.

Pour l'importation de la base de données sous SAS, nous optons pour une lecture en mode format. Ce mode était le plus adapté à nos bases de données car il nous permet de spécifier la longueur, le format et le informat de chaque variable pour faciliter la lecture, l'écriture et le bon affichage des données sous SAS.

2- Analyse des valeurs manquantes

Nous avons analysé les valeurs manquantes de notre base de données et les résultats figurent ci-dessous. Nous constatons que nous avons très peu de valeurs manquantes dans notre base de données. Par ailleurs, la majorité des valeurs manquantes peuvent se justifier par plusieurs raisons :

- Le refus du client de renseigner une information (date de naissance, parrainage)
- Le client n'a pas reçu de remise sur sa livraison ou sa commande

Concernant les valeurs manquantes sur les montants de produits, de livraisons et de montant total payé, nous avons mené une analyse à partir de laquelle nous pouvons relever ces informations. Quand le montant total payé est vide, il y'a deux configurations :

- La somme des montants est nulle
- Aucune valeur n'est renseignée

Nous pouvons conclure que pour ces « commandes », cela peut être des retours, des annulations de commandes ou des désistements puisqu'il s'agit de commandes passées en lignes.

Tableau des valeurs manquantes

Table de référence	Variable	Nombre	Pourcentage
CLIENTS	A_ETE_PARRAINE	106	2%
CLIENTS	DATE_NAISSANCE	62	1%
COMMANDES	MONTANT_DES_PRODUITS	135	1%
COMMANDES	MONTANT_LIVRAISON	132	1%
COMMANDES	REMISE_SUR_LIVRAISON	10282	91%
COMMANDES	MONTANT_TOTAL_PAYE	117	1%

3- Audit des données

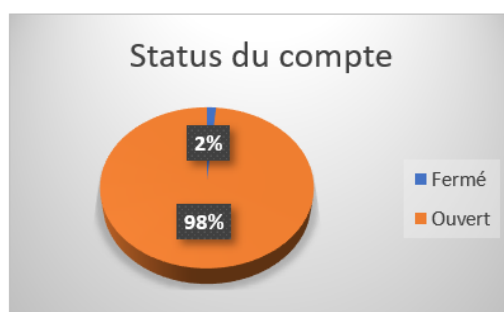
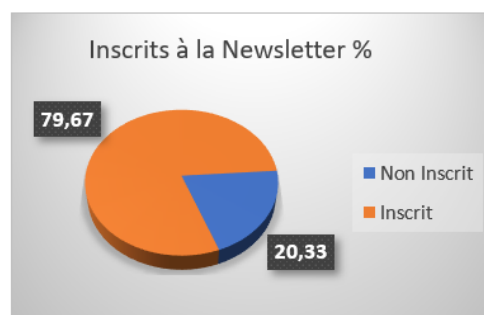
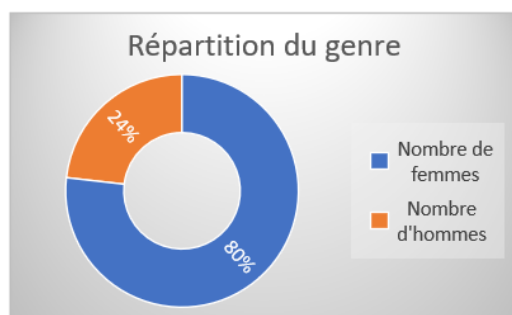
Nous effectuons, à présent quelques statistiques descriptives sur la base de données afin de nous approprier son contenu. Cette étape nous a permis de relever plusieurs informations sur la base de données.

Toutes nos variables sont au bon format. Nous avons par ailleurs noté que la variable **A_ETE_PARRAINE** qui contient trois modalités : « OUI », « NON » et « ? ». Pour une meilleure analyse dans la suite, nous imputons les valeurs manquantes de la variable en attribuant la modalité « INC » aux valeurs autres que « OUI » et « NON ». Par ailleurs, la plupart des clients ont été parrainés et sont également inscrits à la newsletter (76% des clients).

La plupart des clients qui ont passé commandes sur ces deux dernières années sont des clients anciens (ouvert leurs comptes plus de 5 ans avant la date d'arrêt des données).

Certains clients de la base de données n'ont pas fait de commandes sur la période considérée. En effet, nous avons dans la base de données clients 5096 clients mais seulement 4196 ont passé commandes. Pour la suite, on travaillera uniquement avec les clients qui ont passé au moins une commande.

Dans la base de données nous avons beaucoup de femmes (plus de 80% de la population).



Nous avons quelques comptes fermés sur la période considérée.

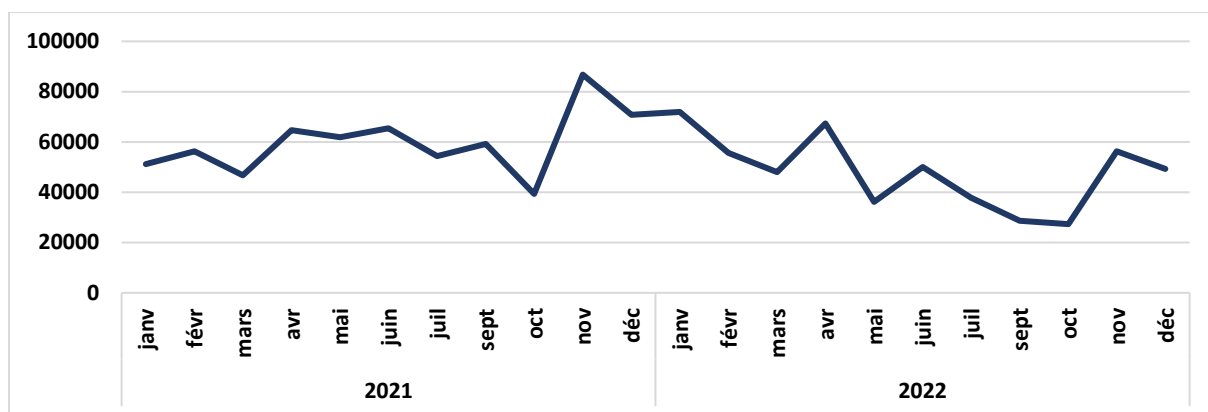
Au cours des deux années couvertes par les données de commandes, le montant des ventes a oscillé entre 3,50 € et 3631 €. Le montant total des ventes, y compris les frais de livraison et les remises, a, quant à lui, varié de -3,90 € à 3716 €. Malgré cette amplitude de valeurs, la moyenne du montant total des ventes s'élevait à 107 €, offrant ainsi une première impression de la diversité des consommateurs présents dans notre base de données.

Variable	Libellé	N	Moyenne	Ec-type	Minimum	Maximum
MONTANT_DES_PRODUITS	Montants des achats	11107	99.39	153.70	3.50	3631.00
REMISE_SUR_PRODUITS	Montant des remises sur le montant des achats	1548	-23.52	56.99	-1396.00	7.54
MONTANT_LIVRAISON	Montant de la livraison	11110	11.96	11.62	0.09	130.00
REMISE_SUR_LIVRAISON	Remise sur la livraison	960	-15.28	16.98	-130.00	35.00
MONTANT_TOTAL_PAYE	Montant total payé	11125	106.57	154.43	-3.99	3716.00

Au cours de la période d'analyse de deux ans, la base de données a enregistré un total de 11 237 commandes passées par 4 201 clients différents. Lors de l'élaboration des indicateurs RFM, seuls les clients ayant effectué des commandes ont été pris en compte, tandis que ceux sans montant total ont été exclus pour éviter de fausser l'analyse.

Le montant moyen par commande était de 107€ tandis que le montant moyen par produit (hors livraison et remise) était de 99€. En outre, la société a réalisé un chiffre d'affaires total de 1 185 606 € sur la période d'analyse, avec 656 836 € en 2021 et 528 770 € en 2022. L'analyse de l'évolution des ventes révèle une baisse des ventes chaque année en octobre, suivie d'une forte augmentation en novembre. Toutefois, pour l'année 2022, une tendance décroissante est observée.

	En montant	En pourcentage
Chiffres d'affaires total	1 185 606 €	100%
Chiffres d'affaires en 2021	656 836 €	55%
Chiffres d'affaires en 2022	528 770 €	45%



Sur les 11 242 commandes passées, la société a appliqué une remise sur 1 548 commandes, avec un montant moyen de remise de 24 €, et une remise sur livraison sur 960 commandes, avec un montant moyen de remise sur livraison de 15€.

Ces informations sont pertinentes pour l'étude de la réglementation RFM.

Pour approfondir l'analyse, il serait intéressant de calculer d'autres indicateurs tels que la fréquence d'achat des clients qu'on vous présentera dans la section suivante, ou encore de comparer les performances de l'entreprise par rapport à ses concurrents. De plus, il pourrait être utile de creuser les raisons de la baisse des ventes en octobre et de la tendance décroissante en 2022, afin de proposer des solutions pour inverser cette tendance.

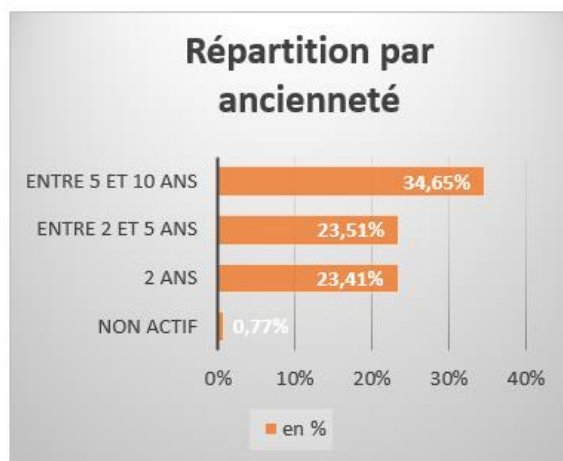
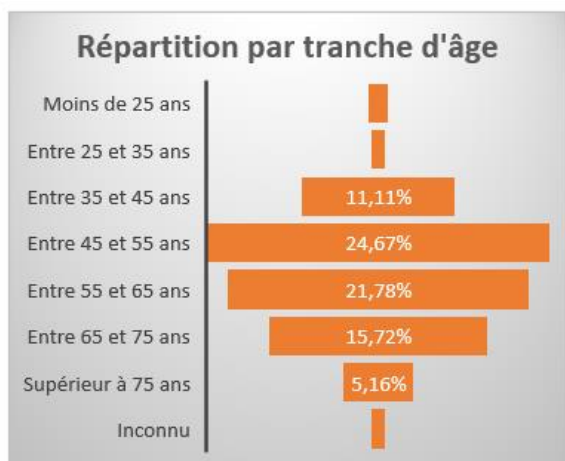
4- Construction de variables

Pour poursuivre notre analyse, nous construisons quelques variables supplémentaires.

- Age du client : Nous calculons l'âge du client à la date d'arrêt des données (31/12/2023).
- Ancienneté du client : Pour les comptes encore ouverts sur la période nous calculons la période qui s'est écoulé depuis la création de compte du client.

Avec ces différentes variables nous effectuons des études statistiques pour approfondir notre analyse.

Dans notre base de données clients, l'âge varie de 8 à 92 ans. Toutefois, plus de 50% des clients appartiennent à la tranche d'âge de 45 à 65 ans, suivie de près par celle de 65 à 75 ans. Ces chiffres indiquent clairement que notre population d'étude est composée de clients relativement âgés.



En ce qui concerne l'ancienneté de nos clients, celle-ci varie de 1 à 10 ans. Cependant, on constate que 20% de nos clients ont une ancienneté de 6 ans. La tranche d'ancienneté la plus représentée est celle de 5 à 10 ans, représentant 46% de nos clients, suivie de la tranche d'ancienneté de 2 à 5 ans, qui représente 30% de nos clients.

PARTIE 2 : Segmentation

Cette section est destinée à nos calculs d'indicateurs RFM, qui nous fournissent des informations sur les comportements d'achat de nos clients suivant trois indicateurs :

- La **récence** qui correspond au temps écoulé entre la dernière commande et la période d'analyse en mois. Les clients qui ont acheté récemment sont plus précieux que ceux qui n'ont pas acheté depuis longtemps ;
- La **fréquence**, qui indique le nombre de commandes passées au cours des deux dernières années : Les clients qui achètent plus souvent sont plus précieux que ceux qui n'achètent pas souvent ;
- Le **montant**, qui représente le montant moyen payé par client pour l'ensemble de ces commandes : Les clients qui dépensent plus d'argent sont plus précieux que ceux qui dépensent moins.

En utilisant ces indicateurs, nous sommes en mesure de différencier nos clients selon leur comportement d'achat.

1- Indicateur Récence :

	RECENCE	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
R1	1	463	11,03	463	11,03
	2	302	7,20	765	18,23
	3	204	4,86	969	23,09
	4	141	3,36	1110	26,45
	5	239	5,70	1349	32,15
R2	6	256	6,10	1605	38,25
	7	136	3,24	1741	41,49
	8	212	5,05	1953	46,54
	9	145	3,46	2098	50,00
	10	196	4,67	2294	54,67
	11	265	6,32	2559	60,99
	12	303	7,22	2862	68,21
R3	13	299	7,13	3161	75,33
	14	86	2,05	3247	77,38
	15	133	3,17	3380	80,55
	16	111	2,65	3491	83,20
	17	119	2,84	3610	86,03
	18	133	3,17	3743	89,20
	19	120	2,86	3863	92,06
	20	99	2,36	3962	94,42
	21	125	2,98	4087	97,40
	22	109	2,60	4196	100,00
	23				
	24				

En règle générale, un client qui a effectué une transaction plus récemment est considéré comme plus actif et engagé. En fonction de la durée depuis la dernière transaction, les clients sont répartis en trois

SEG_RECENCE	Fréquence	Pourcentage
R1	1637	39,01
R2	1210	28,84
R3	1349	32,15

groupes : les "nouveaux clients (R3)" qui ont effectué leur dernière transaction récemment, les "clients récents (R2)" qui ont effectué leur dernière transaction il y a quelque temps et les "clients inactifs ou perdus (R1)" qui ont effectué leur dernière transaction il y a plus d'un an.

Ce tableau fait référence qu'après le regroupement des clients selon ces critères, les résultats ont montré que 39% de nos clients sont dans la catégorie R1, ayant effectué leur dernière transaction il y a plus de 12 mois, ces clients peuvent donc être considérés comme inactifs. Près de 29% des clients appartiennent à la catégorie R2, ayant effectué leur dernière transaction il y a plus de 6 mois mais moins de 12 mois, tandis que 30% appartiennent à la catégorie R3, ayant effectué leur dernière transaction il y a moins de 6 mois. Les clients appartenant à la catégorie R3 sont considérés comme plus précieux et susceptibles de faire une autre commande.

En général, les clients ayant une valeur de récence élevée sont considérés comme étant à risque d'être perdus et nécessitant une réactivation, tandis que les clients ayant une valeur de récence faible sont considérés comme étant plus engagés et plus susceptibles d'effectuer des transactions futures.

2- Indicateur Fréquence :

Il est normal que l'ancienneté de la dernière commande soit comprise entre 1 et 24 mois, étant donné que nous disposons de deux années d'analyses d'études. Pour ce qui est de regrouper cet indicateur, nous veillerons à avoir un nombre de modalités comparable d'une classe à l'autre.

	FREQUENCE	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
F1	1	1624	38,70	1624	38,70
F2	2	908	21,64	2532	60,34
	3	581	13,85	3113	74,19
	4	459	10,94	3572	85,13
	5	263	6,27	3835	91,40
	6	125	2,98	3960	94,38
	7	96	2,29	4056	96,66
	8	48	1,14	4104	97,81
	9	33	0,79	4137	98,59
F3	10	25	0,60	4162	99,19
	11	13	0,31	4175	99,50
	12	3	0,07	4178	99,57
	13	5	0,12	4183	99,69
	14	8	0,19	4191	99,88
	15	2	0,05	4193	99,93
	16	1	0,02	4194	99,95
	18	2	0,05	4196	100,00

En ce qui concerne la fréquence des commandes, sur les deux dernières années, le nombre de commandes varie de 1 à 18. Cependant, plus de 85% des commandes ont été comprises entre 1 et 4, dont plus de 35% ne représentent qu'une seule commande. Ce dernier sera donc seul dans sa catégorie.

SEG_FREQUENCE	Fréquence	Pourcentage
F1	1624	38,70
F2	1489	35,49
F3	1083	25,81

Le deuxième indicateur que nous avons calculé mesure la fréquence d'achat des clients auprès de l'entreprise au cours d'une période spécifique (24 derniers mois dans notre cas). Les clients sont classés en fonction de cette fréquence et ceux qui achètent le plus souvent sont considérés comme étant plus fidèles, générant ainsi une valeur plus élevée pour l'entreprise. En effet, plus un client achète souvent, plus il est susceptible d'être fidèle à l'entreprise et de générer des revenus réguliers à long terme. Après avoir regroupé les clients en fonction du nombre de commandes passées au cours des 24 derniers mois, nous avons constaté que 38,70% des clients achètent rarement (commande unique), 35,49% achètent souvent (entre 2 et 4 commandes), et 25,81% achètent très souvent (plus de 3 commandes) sur cette période. Ces résultats proviennent de notre analyse de la base de données client sur les deux dernières années.

3- Indicateur montant :

	RANG	_FREQ_	MONTAN_MIN	MONTANT_MAX
M1	0	419	6.000	26.39
	1	420	26.400	35.25
	2	419	35.392	44.20
M2	3	420	44.200	55.74
	4	420	55.745	67.16
	5	420	67.170	84.00
M3	6	419	84.211	106.00
	7	420	106.082	142.20
	8	420	142.250	217.00
	9	419	217.100	3377.00

Le dernier indicateur que nous avons calculé est celui du montant, qui permet de mettre en évidence les clients qui génèrent le plus de revenus pour l'entreprise, ainsi que ceux qui ont un potentiel de dépenses élevé et qui pourraient bénéficier d'actions marketing ciblées pour augmenter leur valeur pour l'entreprise.

SEG_MONTANT	Fréquence	Pourcentage
M1	1503	35,82
M2	1319	31,43
M3	1374	32,75

En analysant le montant total dépensé par chaque client, nous avons décidé de regrouper les clients en fonction de trois classes : ceux qui dépensent moins de 50 € sont classés en (M1) (faible valeur d'achat), ceux qui dépensent entre 50 € et 100 € sont dans la classe des valeurs moyennes d'achat (M2), et ceux qui dépensent plus de 100 € sont classés en tant que clients à forte valeur d'achat (M3). Cette segmentation nous permettra de mettre en place des actions de fidélisation et de développement commercial adaptées à chaque groupe de clients, afin de maximiser leur valeur pour l'entreprise.

4- Croisements

➤ Indicateurs de récence et de fréquence

Le croisement de l'indicateur de récence et de fréquence dans la segmentation RFM permet de classer les clients en fonction de leur comportement d'achat récent et de leur fréquence d'achat globale.

Le critère de récence (R) évalue la date du dernier achat effectué par le client. En le croisant avec le critère de fréquence (F), qui mesure le nombre total d'achats effectués par le client sur une période donnée, nous pouvons établir un profil de comportement d'achat pour chaque client.

SEG_RECENCE	SEG_FREQUENCY		
	F1	F2	F3
R1	21,26	13,25	4,50
R2	9,80	10,84	8,20
R3	7,65	11,39	13,11

Ainsi en croisant ces deux indicateurs, on peut classifier les clients en fonction de leur comportement d'achat récent et de leur fréquence d'achat globale, ce qui nous permet d'établir un profil de comportement d'achat pour chaque client. Cette segmentation permet de distinguer les clients les plus fidèles et rentables, qui ont acheté récemment et qui achètent souvent, des clients moins fidèles et moins rentables, qui ont acheté il y a longtemps et qui n'achètent pas souvent.

En utilisant cette méthode, l'entreprise peut adapter ses stratégies de marketing et de fidélisation en fonction des différents profils de clients, en proposant des offres spéciales et des programmes de fidélisation pour encourager les clients les plus rentables à continuer d'acheter, et en cherchant à stimuler l'activité d'achat des clients moins fidèles. Pour notre cas, en fonction de la proportion et de la signification de chaque croisement, nous avons décidé de regrouper les clients de manière que les indicateurs RF soient pertinents et interprétables, en créant la classe RF1 pour les clients qui ont passé 3 commandes ou moins il y a plus de 12 mois, et la classe RF3 pour les clients qui ont commandé 2 fois ou plus il y a moins de 12 mois.

➤ Indicateurs de Récence - Fréquence et Montant

RECENCE_FREQUENCY	SEG_MONTANT		
	M1	M2	M3
RF1	14,39	9,68	10,44
RF2	12,94	9,84	10,01
RF3	8,48	11,92	12,30

A chaque croisement de recence-frequence et seg_Montant correspond un segment .

➤ Interprétation

Après avoir segmenté nos clients en 9 groupes différents selon leur comportement d'achat, nous avons pu obtenir un score allant de RFM1 à RFM9. Plus le score est élevé, plus le comportement d'achat du client est favorable. Le RFM9 représente les clients les plus fidèles et les plus rentables, qui ont acheté récemment, fréquemment et pour des montants élevés. Nous avons identifié 516 clients dans cette catégorie, avec une récence moyenne de 4,35 mois entre la dernière commande et la période d'analyse, la fréquence la plus élevée en moyenne de 4,7 commandes passées au cours des deux dernières années et un montant moyen d'achat de 453 845,87 €. Il est crucial de maintenir une bonne relation avec ces clients et de les fidéliser davantage.

Quant au RFM8, ce segment regroupe des clients très fidèles mais qui n'achètent pas aussi fréquemment que les clients RFM9. Ces clients ont peut-être réduit leur fréquence d'achat récemment ou ont dépensé un peu moins. Dans ce segment, nous avons identifié 500 clients sur 4196 clients, avec une récence plus faible, c'est-à-dire qu'ils commandent très fréquemment en moyenne tous les 4,35 mois entre la date de la dernière commande et la période d'analyse, un montant moyen d'achat de 170 265,63 € et un nombre de commandes moyen de 4,7 sur les deux dernières années. Il est donc important d'offrir des offres personnalisées à ces clients et de les encourager à acheter à nouveau.

Les clients du segment RFM7 ont acheté récemment et ont dépensé des montants élevés, mais n'ont pas acheté autant de fois que les clients des segments RFM9 et RFM8. Les clients du segment RFM6 ont également acheté récemment, mais ont moins acheté que les clients des segments RFM9, RFM8 et RFM7. Les clients des segments RFM5 et RFM4 ont acheté il y a un certain temps, mais ont dépensé des montants élevés et ont acheté assez souvent. Les clients du segment RFM3 ont une valeur d'achat un peu plus élevée, mais ont une faible fréquence d'achat, ce qui rend difficile leur rentabilisation. Les clients des segments RFM2 et RFM1 ont acheté il y a un certain temps et n'ont pas dépensé beaucoup d'argent ou acheté très souvent.

En utilisant cette segmentation, l'entreprise peut mieux comprendre les habitudes d'achat de ses clients et adapter ses stratégies marketing et de fidélisation en conséquence. Par exemple, elle peut proposer des offres promotionnelles attractives pour les clients des segments RFM6 et RFM3 afin de les inciter à acheter plus souvent, ou élaborer une stratégie de réactivation pour les clients du segment RFM1 qui n'ont pas acheté depuis longtemps. De même, elle peut mettre en place des actions de fidélisation pour les clients des segments RFM9 et RFM8 qui sont les plus rentables.

Segments	Effectifs	en %	Recence	moye Montant	Montant moyen par client	Nombre de commandes	Nombre de commandes par clients	Age moyen
RFM1	604	14%	17,62	27 030,73 €	44,75 €	861	1,4	54,87
RFM2	406	9,68%	17,76	44 919,00 €	110,64 €	632	1,6	54,86
RFM3	438	10,44%	17,91	144 790,91 €	330,57 €	694	1,6	53,65
RFM4	543	12,94%	8,36	31 791,81 €	58,55 €	959	1,8	56,52
RFM5	413	9,84%	9,08	65 150,48 €	157,75 €	902	2,2	56,13
RFM6	420	10,01%	9,46	193 345,27 €	460,35 €	913	2,2	54,64
RFM7	356	8,48%	4,13	54 466,15 €	152,99 €	1486	4,2	58,55
RFM8	500	11,92%	4,45	170 265,63 €	340,53 €	2371	4,7	57,32
RFM9	516	12,30%	4,35	453 845,87 €	879,55 €	2419	4,7	56,12

CONCLUSION

En conclusion, ce projet a été une véritable opportunité pour nous de développer nos compétences en informatique, en statistiques et en datavisualisation, ainsi que de mieux comprendre le métier de CRM analytics. Nous avons su tirer parti de nos différentes expériences et connaissances en travaillant en groupe pour aboutir à un résultat de qualité.

En effet, en mutualisant nos compétences, nous avons pu produire un reporting clair et compréhensible pour l'ensemble des parties prenantes, avec des résultats pertinents et facilement interprétables. Nous sommes ainsi parvenus à identifier neuf segments équilibrés en termes d'effectifs, permettant de mieux cibler les actions de marketing et de vente.

Nous sommes convaincus que ce travail est non seulement bénéfique pour notre formation, mais aussi pour l'entreprise, qui pourra utiliser ces résultats pour améliorer ses performances et sa compréhension de sa clientèle. Nous sommes fiers du travail accompli et avons hâte de pouvoir vous présenter en vive voix le jour de la soutenance.

ANNEXE

1. Le rapport écrit en PDF
2. La maquette Excel
3. Les programmes
4. Le fichier excel des sorties SAS