

# Identifier facilement vos meilleurs clients

*Utilisez le logiciel d'analyse prédictive IBM SPSS pour extraire des informations de votre base de données*



---

## Table des matières :

- 1 Introduction
  - 3 Explorer les données clients  
Où vivent vos clients ?
  - 5 Quel est le revenu moyen  
du ménage de vos clients ?
  - 7 Découvrir des groupes de clients avec  
l'analyse RFM
  - 9 Comment les clients répondent-ils aux  
différentes offres promotionnelles ?
  - 10 La fidélisation de la clientèle  
varie-t-elle selon la région ?
  - 12 La réponse de la clientèle à l'offre 1  
a-t-elle varié selon la zone ?
  - 14 Combien ont dépensé les clients ?
  - 15 Combien vont dépenser les clients ?
  - 18 Agir
  - 19 Conclusion;  
À propos d'IBM Business Analytics
- 

## Introduction

« Qui sont nos meilleurs clients, ceux les plus susceptibles de répondre à nos offres ? » Si vous êtes responsable des ventes, du marketing ou du service clientèle, vous voulez une réponse à cette question. En fait, vous voulez en savoir plus sur tous vos clients, des meilleurs aux pires. Tout simplement parce que planifier et mettre en œuvre des stratégies efficaces et rentables pour chaque segment de clientèle est crucial pour augmenter les bénéfices commerciaux. Ce document explique comment une société peut identifier ses meilleurs clients ou ses clients les plus précieux, mais les mêmes processus pourraient être utilisés pour d'autres segments de clientèle.

Connaître vos meilleurs clients (leurs attitudes, habitudes d'achat et profils démographiques) est crucial pour développer et mettre en œuvre des programmes efficaces de marketing et de gestion de la relation client. De telles connaissances vous aident à cibler efficacement vos campagnes promotionnelles, publicitaires et marketing, ainsi qu'à développer des programmes de vente incitative, de ventes croisées et de fidélisation, rétention et récompense des clients à long terme.

Il est particulièrement important de coordonner ces efforts à une époque où le marketing s'éloigne du marketing de masse pour se rapprocher d'une démarche plus ciblée, s'attachant à positionner des produits ou services particuliers pour des types spécifiques de clients. Des informations fiables et détaillées sur le comportement, les attitudes et autres caractéristiques des clients constituent un réel avantage concurrentiel et aident à améliorer le retour sur investissement de toutes vos interactions client. Les détails obtenus de l'analyse la plus élémentaire des caractéristiques des clients peut avoir de profondes répercussions sur votre entreprise.



---

## Points clés :

Connaître vos meilleurs clients (leurs attitudes, habitudes d'achat et profils démographiques) est crucial pour développer et mettre en œuvre des programmes efficaces de marketing et de gestion de la relation client.

Le logiciel d'analyse IBM SPSS peut vous aider à obtenir des informations fiables et détaillées sur vos clients pour renforcer votre avantage concurrentiel.

---

SPSS a fait partie des pionniers dans le domaine de l'analyse des données ; elle a été la première sur ce créneau et continue à être l'une des applications logicielles les plus populaires et les plus largement utilisées. En tant que nouveau membre de l'organisation IBM, SPSS met ses produits et solutions d'analyse de pointe à la portée d'un nombre encore plus grand d'organisations à travers le monde.

Les offres d'IBM SPSS incluent des produits leaders de l'industrie en matière de data mining et text mining, collecte des données et gestion, ainsi que des logiciels de statistiques vous permettant d'identifier vos meilleurs clients et d'élaborer des programmes de marketing plus efficaces et plus rentables.

Les outils d'IBM SPSS sont basés sur les normes de l'industrie et peuvent facilement s'intégrer à votre infrastructure existante pour améliorer l'exactitude, réduire la main d'œuvre et minimiser les pertes. L'effort conjugué d'IBM et de SPSS vous offre une flexibilité optimale et s'adapte aux types de données que vous explorez et à la façon dont vous déployez les résultats.

Ce livre blanc vous montre comment analyser une base de données clients à l'aide d'un logiciel d'analyse prédictive IBM SPSS. Cette famille de produits intégrés pour l'analyse statistique et la gestion des données vous assiste tout au long du processus d'analyse, que vous effectuiez votre analyse à partir d'un seul ordinateur de bureau ou sur un réseau étendu.

Les données marketing que nous utiliserons comme exemples dans ce document contiennent 2 070 clients et incluent les informations suivantes :

- la date à laquelle le client est devenu l'un de vos clients
- l'historique de ses achats par valeur des commandes en dollars
- sa réponse à différentes offres
- les informations relatives à l'attrition ou à la rotation du client
- le niveau de revenu de son ménage
- sa classification géographique
- son sexe et autres variables démographiques

Notre but est d'identifier des segments ou groupes de clients qui constituent les meilleurs clients de notre société. Nous vous expliquerons également comment utiliser le logiciel IBM SPSS pour exploiter vos données clients, prédire le comportement futur de vos clients et prendre de meilleures décisions opérationnelles.

## Explorer les données clients

Commençons par explorer les différentes variables de notre base de données pour répondre à des questions telles que :

- Où vivent vos clients ?
- Quel est le revenu moyen de leur ménage ?
- Depuis combien de temps vos clients font-ils partie de votre clientèle ?
- Combien d'argent vos clients dépensent-ils chez nous ?

Le logiciel IBM SPSS offre plusieurs méthodes pour obtenir rapidement les réponses à ces questions. Les procédures Fréquences et Descriptifs d'IBM® SPSS® Statistics sont très utiles lorsque vous examinez pour la première fois vos données. Souvent, cet exercice peut suggérer les meilleures méthodes pour analyser vos données.

IBM SPSS Statistics peut vous aider à déterminer si vos clients vivent en zone urbaine, suburbaine ou rurale et vous aider à choisir les bonnes stratégies marketing.

---

### Où vivent vos clients ?

Identifier l'endroit où vivent vos clients (zone urbaine, suburbaine ou rurale) peut vous aider à déterminer la meilleure stratégie marketing pour les atteindre. La procédure Fréquences d'IBM SPSS Statistics offrent un tableau d'effectifs et pourcentages par catégorie, ainsi qu'une représentation visuelle des données sous forme de graphique à barres, d'histogramme ou de diagramme à secteurs, avec des étiquettes de catégorie attribuées à chaque valeur.

D'après les résultats présentés dans le diagramme 1, nous voyons que la majorité de la clientèle vit en zone suburbaine (34,2 %) et qu'une minorité vit en zone rurale (19,4 %). Nous voyons également que 16,9 des clients sont répertoriés comme données manquantes, ce qui correspond aux clients dont on ignore le lieu de résidence.

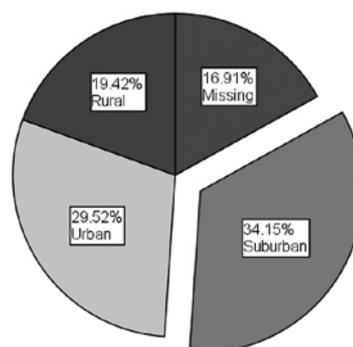


Diagramme 1 : Les zones où vivent vos clients

---

Il est souvent utile de savoir où et pourquoi il manque des informations dans vos données. IBM SPSS Statistics reconnaît une valeur manquante soit comme une valeur entièrement nulle, soit comme une valeur résiduelle telle que « non applicable » ou « sans opinion ». Cela permet aux analystes de faire la distinction entre les données qui sont manquantes parce que la question ou l'observation ne s'applique pas au répondant et les données qui sont manquantes parce qu'aucune réponse n'a été fournie.

Le tableau 1 à la page suivante affiche les résultats d'une distribution de fréquence. Les valeurs affichées dans la colonne « Pourcentage » sont calculées en utilisant tous les cas de l'ensemble de données (N = 2070). La colonne « Pourcentage valide » affiche la distribution proportionnelle des cas avec uniquement des données valides rapportées, ou données non manquantes (N = 1720). Cela permet une comparaison côte-à-côte rapide des distributions ; de grosses différences entre ces deux colonnes peuvent suggérer un biais dans les données. Le tableau 1 montre également que près de 17 % des données sont manquantes, ce qui pourrait indiquer des problèmes potentiels pour les analyses.

---

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Suburban	707	34.2	41.1	41.1
	Urban	611	29.5	35.5	76.6
	Rural	402	19.4	23.4	100.0
	Total	1720	83.1	100.0	
Missing	System	350	16.9		
Total		2070	100.0		

*Tableau 1* : Le tableau révèle que la plupart des clients (41,1 %) vivent en zone suburbaine.

---

IBM SPSS Statistics vous permet de savoir facilement si une information est manquante ou si une valeur est nulle.

---

## Quel est le revenu moyen du ménage de vos clients ?

Il y a plusieurs façons d'obtenir un aperçu plus détaillé de vos clients. Pour obtenir des informations sur le revenu du ménage, par exemple, nous employons des statistiques résumées de base, comme les valeurs moyenne, minimum et maximum, et l'écart type. Le descriptif d'IBM SPSS Statistics offre un premier aperçu informatif de vos données d'échelle ou de niveau d'intervalle, comme le revenu (mesuré en dollars). Le tableau 2 affiche les résumés descriptifs du revenu du ménage.

	N	Minimum	Maximum	Mean	Std. Deviation
HH Income	1984	\$38,552.00	\$85,571.00	\$61,386.39	\$11,006.75
Valid N (listwise)	1984				

Tableau 2 : La procédure Descriptifs dans IBM SPSS Statistics offre un résumé rapide montrant que le revenu de ménage moyen est de 61 386,39 \$

---

Nous voyons dans le tableau 2 que le revenu annuel moyen du ménage des près de 2 000 clients qui ont indiqué leur revenu dans nos données est de 61 386,39 \$. D'après l'écart type qui est d'environ 11 000 \$, nous savons que la majorité des clients (environ 68 %) gagnent entre 50 000 \$ et 72 000 \$.

## Depuis combien de temps vos clients font-ils partie de votre clientèle ?

Pour déterminer depuis combien de temps vos clients font partie de votre clientèle, ajoutez un nouveau champ à vos données pour la date à laquelle le client a été entré dans la base de données. En soustrayant cette date à la date actuelle, ou à la date de sa dernière transaction, vous pouvez déterminer depuis combien de temps un client fait partie de votre clientèle. En utilisant l'une des nombreuses fonctions temporelles disponibles dans IBM SPSS Statistics, vous pouvez facilement transformer la date en nombre d'années.

La base de données contient également les informations d'attrition, ou le statut du client (actuel ou passé). Grâce à ces informations et à la durée depuis laquelle le client est dans la base de données, vous pouvez déterminer sa « survie » ou la durée pendant laquelle il vous est resté fidèle avant de vous abandonner.

Les fonctions de statistiques descriptives d'IBM SPSS Statistics vous permettent d'obtenir des informations détaillées sur le revenu du ménage de vos clients.

---

IBM SPSS Statistics peut déterminer votre taux d'attrition et la durée pendant laquelle chacun de vos clients a fait partie de votre clientèle, vous aidant ainsi à identifier les tendances et à agir de façon proactive pour retenir les clients.

---

L'analyse de survie Kaplan-Meier est particulièrement utile pour mesurer le délai avant qu'un client devienne inactif ou ne fasse plus partie de votre clientèle. L'un des gros avantages de l'analyse de survie Kaplan-Meier est que cette méthode tient compte des clients disparus de la base de données avant que l'attrition soit déterminée : par exemple, si un client reste un client « actuel » mais a été inactif. Le graphique 2 ci-dessous montre que plus les clients sont dans la base de données depuis longtemps moins ils sont actifs. Autrement dit, la proportion cumulée des clients restant dans la base de données diminue régulièrement au fil du temps. Nous pouvons voir que le temps de survie médian, ou le délai au bout duquel 50 pour cent des clients sont partis, est d'environ 11 ans.

---

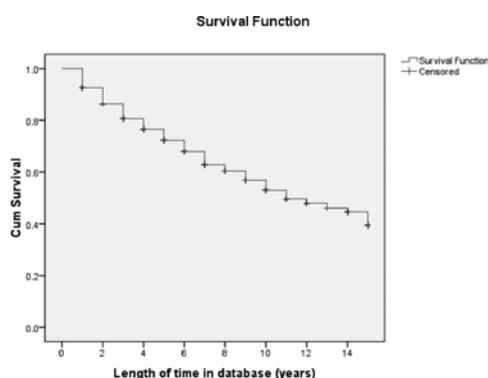


Diagramme 2 : Le diagramme de survie Kaplan-Meier d'IBM SPSS Statistics indique que le temps de survie médian d'un client est de 11 ans.\*

---

\* L'analyse Kaplan-Meier ci-dessus a montré que la durée médiane de présence dans la base de données clients est de 11 ans, et la valeur moyenne (non affichée) est 9,974 ans. Souvenez-vous, toutefois, que l'un des gros avantages de la courbe Kaplan-Meier est que cette méthode peut prendre en compte les données « censurées », les données perdues de l'échantillon avant que le résultat final soit observé (par exemple, si un client est supprimé des données avant que l'attrition soit déterminée).

## Découvrir des groupes de clients avec l'analyse de récence, de fréquence et de valeur monétaire (RFM)

Vous pouvez également découvrir qui sont vos meilleurs clients. Les « meilleurs clients » sont généralement les clients les plus rentables ou ceux qui dépensent le plus d'argent dans votre organisation. Pour obtenir l'aperçu le plus exact de la valeur d'un client tout au long de son cycle de vie, nous utilisons l'analyse de récence, de fréquence et de valeur monétaire (RFM). Par exemple, vous pouvez classer vos clients comme suit :

- Ceux qui ont dépensé le plus – le plus souvent et le plus récemment.
  - Ceux qui ont dépensé le plus – en valeur, mais qui n'ont peut-être rien acheté depuis longtemps.
  - Ceux qui ont dépensé le plus en un nombre minimal de transactions.
  - Ceux qui ont dépensé le moins, ou rarement, et n'ont rien acheté depuis longtemps.
- De cette façon, vous pouvez déterminer quels sont vos meilleurs clients en fonction de la récence et de la fréquence de leurs achats ainsi que de la somme qu'ils ont dépensée.

L'analyse RFM d'IBM SPSS Statistics vous aide à identifier vos « meilleurs clients », ceux qui dépensent le plus d'argent dans votre organisation, ce qui vous permet de cibler spécifiquement et efficacement ce groupe.

Utiliser l'analyse RFM dans IBM SPSS Statistics vous permet de générer une liste des clients qui ont dépensé le plus en intégrant les variables appropriées, le numéro du client, ainsi que la date et le montant des transactions. Dans cet exemple, il est nécessaire d'avoir un numéro client associé à chaque transaction. La date des transactions est importante pour savoir quand ou à quelle fréquence le client achète. Enfin, parce que vous aurez envie de savoir combien un client aura dépensé au cours de son cycle de vie, vous inclurez également le total.

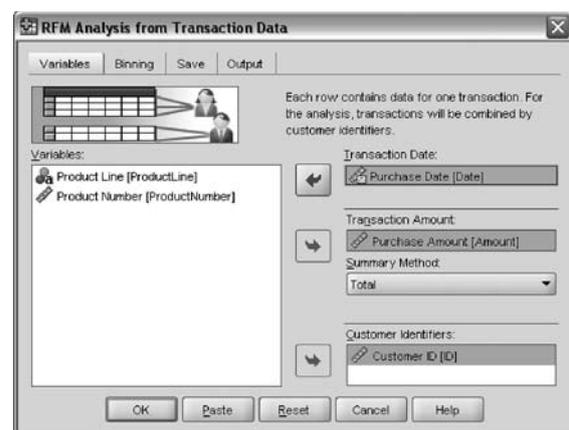


Figure 1 : L'analyse RFM vous permet d'intégrer rapidement les variables nécessaires.

Une fois que vous avez entré ces variables, vous pouvez effectuer l'analyse pour obtenir les scores RFM de vos clients et déterminer ceux que vous voulez cibler. Dans cet exemple, vous pourriez vous concentrer sur les clients qui ont un score RFM de 555, ceux qui ont acheté le plus récemment et le plus souvent et qui ont dépensé le plus d'argent. Une fois le résultat généré, vous pouvez trier les données pour vous concentrer sur vos meilleurs clients (ceux ayant un score RFM de 555).

	Data_entry_date	Transaction_count	Amount	Recency_score	Frequency_score	Monetary_score	RFM_score
1	25-Sep-2005	13	1164.00	5	5	5	555
2	307	11-Oct-2005	10	1249.00	5	5	555
3	672	11-Oct-2005	8	1006.00	5	5	555
4	856	18-Nov-2005	12	1233.00	5	5	555
5	851	25-Nov-2005	10	1062.00	5	5	555
6	871	08-Nov-2005	9	1050.00	5	5	555
7	900	19-Sep-2005	9	1061.00	5	5	555
8	900	08-Oct-2005	10	1118.00	5	5	555
9	31	07-Oct-2005	8	965.00	5	5	554
10	180	26-Nov-2005	8	940.00	5	5	554
11	184	29-Oct-2005	9	960.00	5	4	554
12	223	30-Dec-2005	9	676.00	5	4	554
13	418	19-Oct-2005	10	1036.00	5	4	554
14	420	04-Dec-2005	9	1002.00	5	4	554
15	738	19-Dec-2005	8	892.00	5	4	554
16	774	04-Nov-2005	9	676.00	5	4	554
17	49	19-Dec-2005	8	839.00	5	3	553
18	106	13-Nov-2005	8	819.00	5	3	553
19	264	19-Dec-2005	10	814.00	5	3	553
20	314	29-Nov-2005	10	834.00	5	3	553
21	618	17-Nov-2005	10	888.00	5	3	553
22	838	02-Oct-2005	9	843.00	5	3	553
23	897	26-Oct-2005	9	890.00	5	3	553
24	898	20-Oct-2005	9	876.00	5	3	553
25	401	11-Oct-2005	9	771.00	5	2	552

Figure 2 : L'analyse montre les scores RFM.

En utilisant d'autres analyses descriptives des dépenses des clients, vous pouvez voir que la majorité des clients dépensent 500 \$ ou moins, et qu'à des valeurs en dollars supérieures, le nombre de clients effectuant des achats diminue de façon constante. La somme moyenne dépensée par les clients s'élève à 1 360 \$, et un tout petit nombre de clients a dépensé plus de 7 000 \$.

Pour l'instant, nous savons qu'un client type :

- vit en zone suburbaine
- a un revenu de ménage de 61 000 \$
- dépense 1 360 \$ pour nos produits et services
- a une « durée de vie » moyenne ou survie de 11 ans

## Comment les clients répondent-ils aux différentes offres promotionnelles ?

L'analyse des résultats de promotions marketing spécifiques par IBM SPSS Statistics est une étape importante pour comprendre vos clients. Évaluer les efforts passés aide à identifier ce qui a marché et ce qui n'a pas marché, ce qui vous permet de renouveler vos succès et de tirer des leçons de vos échecs. Ici, nous voulons répondre à deux questions :

- Combien de personnes ont répondu à chacune de nos quatre offres ?
- Quelle est la somme moyenne dépensée en réponse à nos différentes promotions ?

Pour ce faire, nous exécutons la procédure Fréquences sur la réponse à chaque offre et la procédure Descriptifs sur la valeur de la commande pour les quatre offres. Dans le tableau 4, nous voyons que 890 clients, soit 44,5 pour cent de la clientèle, ont répondu à l'offre 1. Une analyse similaire des autres offres montrerait une réponse de 39 pour cent à l'offre 2, une réponse de 37,4 pour cent à l'offre 3 et une réponse de 17,4 pour cent à l'offre 4.

Data	Date	Transaction	Amount	Recency	Frequency	Monetary	RFM score
1	25-Sep-2005	13	1154.00	5	5	5	886
2	307	11-Oct-2005	10	1249.00	5	5	555
3	672	11-Oct-2005	8	1088.00	5	5	555
4	856	18-Nov-2005	12	1223.00	5	5	666
5	851	25-Nov-2005	10	1062.00	5	5	555
6	971	08-Nov-2005	9	1058.00	5	5	554
7	908	19-Sep-2005	9	1061.00	5	5	555
8	900	08-Oct-2005	10	1118.00	5	5	555
9	31	07-Oct-2005	8	955.00	5	5	554
10	180	26-Nov-2005	8	960.00	5	4	554
11	154	25-Oct-2005	9	980.00	5	4	554
12	223	30-Dec-2005	9	876.00	5	4	554
13	418	19-Oct-2005	10	1036.00	5	4	554
14	420	04-Dec-2005	9	1003.00	5	4	554
15	738	18-Dec-2005	8	892.00	5	4	554
16	774	04-Nov-2005	9	870.00	5	4	554
17	48	18-Dec-2005	8	839.00	5	3	553
18	106	13-Nov-2005	9	819.00	5	3	553
19	264	10-Dec-2005	10	814.00	5	3	553
20	314	29-Nov-2005	10	834.00	5	3	553
21	618	17-Nov-2005	10	888.00	5	3	553
22	638	02-Oct-2005	9	843.00	5	3	553
23	107	26-Oct-2005	9	850.00	5	3	553
24	806	25-Oct-2005	8	875.00	5	3	553
25	431	11-Oct-2005	8	771.00	5	2	552

Figure 2 : L'analyse montre les scores RFM.

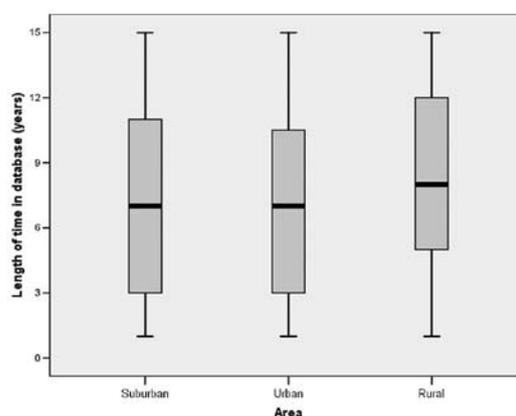
Nous savons donc que l'offre 1 a enregistré le plus fort taux de réponse, mais nous ignorons comment ces réponses se sont traduites en revenus pour la société. Exécuter la procédure Descriptifs sur les offres 1 à 4 révèle que la valeur de l'offre 1, 376,64 \$, a également été la meilleure des quatre, comme l'indique le tableau 4, tandis que l'offre 3, qui avait elle aussi affiché un très fort taux de réponse, a été la pire avec 293,98 \$ par réponse. L'offre 1 a donc été, sur les deux plans, la plus efficace.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not respond	1110	53.6	55.5	55.5
	Responded	890	43.0	44.5	100.0
	Total	2000	96.6	100.0	
Missing	System	70	3.4		
Total		2070	100.0		

Tableau 4 : Près de 45 pour cent de la base de données clients (soit 890 personnes) ont répondu à l'offre 1.

## La fidélisation de la clientèle varie-t-elle selon la région ?

Pour étudier cette question, nous générons un puissant graphique statistique, la boîte à moustaches. Il affiche à la fois la valeur médiane et la distribution des données. D'après la boîte à moustaches du graphique 3, nous pouvons voir que les clients des zones rurales ont une durée médiane de présence dans la base de données supérieure, ce qui suggère qu'ils font partie de la clientèle depuis plus longtemps, en moyenne, que ceux des autres zones.



*Diagramme 3* : La boîte à moustaches affiche à la fois la valeur médiane et la distribution des données. Il est facile de voir que les clients des zones rurales ont une durée médiane de présence supérieure, ce qui suggère qu'ils font partie de la clientèle depuis plus longtemps, en moyenne, que ceux des autres zones.

L'analyse IBM SPSS Statistics des résultats d'une campagne marketing spécifique vous aidera à tirer profit de vos succès et à tirer des leçons de vos échecs.

Une comparaison des moyennes offre des statistiques résumées pour une valeur de mesure par groupe. Le tableau 5 complète les informations affichées dans la boîte à moustaches, mais sous forme de tableau. Il révèle que bien que la durée moyenne globale de présence dans la base de données soit de 7,49 ans, les clients des zones rurales ont fait partie de la clientèle plus longtemps que ceux des zones suburbaines ou urbaines.

	N	Minimum	Maximum	Mean	Std. Deviation
Order value (offer 1)	2000	\$ .00	\$3,280.23	\$376.64	\$602.32
Order value (offer 2)	2000	\$ .00	\$3,477.31	\$312.36	\$569.28
Order value (offer 3)	2000	\$ .00	\$3,377.18	\$293.98	\$536.49
Order value (offer 4)	2000	\$ .00	\$3,339.58	\$375.63	\$595.73
Valid N (listwise)	2000				

*Tableau 5* : L'analyse de l'historique des achats révèle que la valeur de l'offre 3, 293,98 \$, est inférieure à la valeur moyenne des autres offres.

Est-ce une découverte significative ? La signification statistique nous indique si les différences observées dans nos données sont dues au hasard ou si elles tendent à refléter des tendances dans la population au sens large et méritent une plus grande attention.

Le rapport ANOVA du tableau 6 montre que les différences entre zones en ce qui concerne la durée de vie d'un client sont statistiquement significatives. Généralement, nous considérons un fait comme statistiquement significatif lorsque la probabilité qu'il soit dû au hasard est inférieure à 5 pour cent (fait survenant moins de 5 fois sur 100). Cela est indiqué par un niveau de signification de 0,05 ou moins. Puisque le niveau de signification indiqué dans ce tableau est 0,000 bien en dessous du seuil de 0,05, nous pouvons conclure que les différences de moyennes observées dans les données ne sont sûrement pas dues au hasard. La distribution globale de la fidélisation moyenne de la clientèle et des zones n'est probablement pas due à des causes aléatoires, mais à autre chose.

---

Length of time in database (years)			
Area	Mean	N	Std. Deviation
Suburban	7.38	703	4.398
Urban	6.96	607	4.266
Rural	8.48	398	4.306
Total	7.49	1708	4.366

*Tableau 6* : Ce rapport de comparaison des moyennes révèle que bien que la durée moyenne globale de présence dans la base de données soit de 7,49 ans, les clients des zones rurales ont fait partie de la clientèle plus longtemps, en moyenne, que ceux des zones suburbaines ou urbaines.

---

Exemples de causes possibles :

- Le premier bureau a ouvert dans une zone rurale.
- On a davantage besoin du produit dans une zone que dans une autre.
- Une certaine caractéristique du produit a été introduite avec succès dans une zone.

D'autres causes peuvent exister et nécessiter une investigation. C'est pourquoi il est également important de connaître votre entreprise, afin de rassembler les bonnes données pour tester vos théories sur les relations.

IBM SPSS Statistics offre de puissantes fonctions de création de graphiques statistiques pour visualiser les données de fidélisation de la clientèle par région.

## La réponse de la clientèle à l'offre 1 a-t-elle varié selon la zone ?

Continuons à analyser la réponse à nos offres. IBM SPSS Statistics offre un moyen facile de présenter graphiquement les informations sur les quatre offres, en utilisant un graphique à barres groupées. Le graphique 4 offre un résumé des schémas de réponse par zone. Nous voyons que les clients des zones urbaines ont tendance à moins commander que les clients des deux autres zones, notamment de la zone rurale. C'est une chose que nous n'aurions pas pu deviner au vu de la distribution de fréquence par zone, qui montrait que les zones rurales comptaient moins de clients.

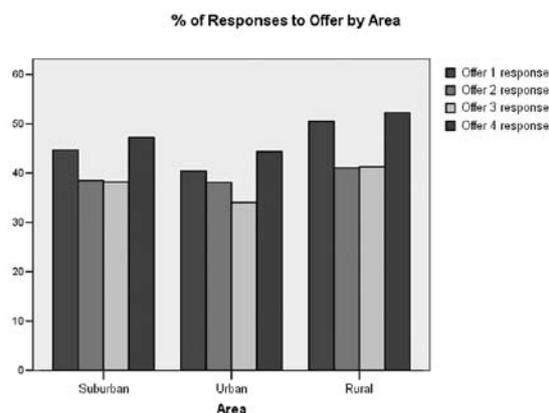


Diagramme 4 : Le graphique à barres groupées IBM SPSS Statistics offre un moyen rapide et clair de présenter les schémas de réponse par zone.

Pour savoir si cela est significatif, nous pouvons explorer plus en détail les résultats de chaque offre par zone. Pour répondre à la question « Comment les personnes de chaque zone ont-elles répondu à l'offre 1 ? », nous réalisons un tableau croisé IBM SPSS Statistics sur l'offre 1 par zone. Le tableau 7 montre que 41,3 pour cent des personnes qui ont répondu à l'offre 1 se trouvaient en zone suburbaine. Même si seulement 26,5 pour cent des personnes qui ont répondu à l'offre 1 se trouvaient en zone rurale, plus de la moitié (50,5 pour cent) des clients de cette zone ont répondu à l'offre.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	569.499	2	284.749	15.188	.000
Within Groups	31965.164	1705	18.748		
Total	32534.663	1707			

Tableau 7 : Le rapport ANOVA montre que les différences observées sont statistiquement significatives, ce qui peut mériter un approfondissement.

Pour comprendre si une zone est associée ou pas à la réponse à l'offre 1, nous comparons les pourcentages des lignes « % de zone » et découvrons que 45 pour cent des personnes en zone suburbaine et 40 pour cent des personnes en zone urbaine ont répondu à cette offre. D'après ces informations, nous concluons que les zones rurales sont de bonnes zones pour une offre telle que l'offre 1.

**Area \* Offer 1 response Crosstabulation**

		Offer 1 response		Total	
		Did not respond	Responded		
Area	Suburban	Count	391	316	707
		% within Area	55.3%	44.7%	100.0%
		% within Offer 1 response	41.0%	41.3%	41.1%
Urban	Count	364	247	611	
		% within Area	59.6%	40.4%	100.0%
		% within Offer 1 response	38.2%	32.2%	35.5%
Rural	Count	199	203	402	
		% within Area	49.5%	50.5%	100.0%
		% within Offer 1 response	20.9%	26.5%	23.4%
Total	Count	954	786	1720	
		% within Area	55.5%	44.5%	100.0%
		% within Offer 1 response	100.0%	100.0%	100.0%

*Tableau 8* : Même si seulement 26,5 pour cent des personnes qui ont répondu à l'offre 1 se trouvaient en zone rurale, plus de la moitié (50,5 pour cent) des clients de cette zone ont répondu à l'offre.

Toutefois, même s'il apparaît que les pourcentages sont différents, c'est une raison insuffisante pour commencer à renouveler l'offre 1 dans les zones rurales. Nous devons d'abord déterminer si la zone et la réponse à l'offre 1 sont indépendantes l'une de l'autre ou pas. Ici, la variable khi-deux est utile pour déterminer si les distributions observées dans les données reflètent des tendances de la population au sens large.

Le tableau 9 contient les informations khi-deux pour la zone et l'offre 1. Dans ce cas, le khi-deux est significatif ( $p = 0,007$ ) et indique que les schémas du tableau ne sont sûrement pas dus au hasard. Il pourrait y avoir une raison spécifique et identifiable expliquant le plus grand succès de l'offre 1 dans les zones rurales. Peut-être que le message répondait plus directement à leurs besoins, ou que le type de support était plus approprié pour attirer et retenir leur attention.

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9.970 <sup>a</sup>	2	.007
Likelihood Ratio	9.964	2	.007
Linear-by-Linear Association	1.999	1	.157
N of Valid Cases	1720		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 179.03.

*Tableau 9* : Un khi-deux de 0,007 pour la zone et l'offre 1 indique que les différences entre les zones sont significatives.

En identifiant les clés du succès de la campagne dans les zones rurales, nous pouvons les exploiter dans les futures offres proposées à cette zone. Nous pouvons aussi choisir d'explorer d'autres relations par rapport à cette zone.

Avec IBM SPSS Statistics, il est facile de présenter simultanément les résultats de plusieurs offres sous forme graphique.

---

## Combien ont dépensé les clients ?

Une fois encore, grâce à l'analyse RFM, nous pouvons intégrer les variables pour déterminer ce que nous voulons savoir. Dans ce cas, nous voulons savoir quels clients ont dépensé le plus d'argent. Une fois l'analyse effectuée, nous pouvons trier les transactions totales par valeur en dollar.

Une autre façon d'examiner l'historique des achats consiste à évaluer la somme totale dépensée, plutôt que juste l'argent dépensé à chaque commande. Peut-être qu'une relation entre la somme totale dépensée et la zone révélera des informations précieuses. Une ANOVA à un facteur offre des informations spécifiques sur la signification des différences dans les valeurs moyennes que vous pouvez observer.

La première chose que fournit une ANOVA à un facteur est un tableau des statistiques descriptives. Le tableau 9 montre que la somme totale moyenne dépensée par zone en réponse à chacune des quatre offres varie considérablement. La somme moyenne dépensée était de 1 206,01 \$ dans les zones urbaines, 1 391,70 \$ dans les zones suburbaines et 1 618,27 \$ dans les zones rurales.

Le rapport montre également que la différence moyenne affichée entre les niveaux de dépense dans les zones suburbaines et rurales n'est pas statistiquement significative. En revanche, il montre que la différence entre les zones rurales et urbaines est significative.

Vous pouvez utiliser ces informations pour mieux comprendre en quoi et pourquoi ces zones diffèrent et pour développer des plans marketing ciblés afin d'exploiter ces différences. Par exemple, il se peut qu'une combinaison différente de marketing et ventes, une offre différente ou une offre groupée spéciale de produits et services fonctionne mieux dans les zones urbaines. Les programmes marketing des zones rurales doivent y être répétés pour un succès renouvelé.

## Combien vont dépenser les clients ?

Les modèles prédictifs sont de puissants outils pour aider à cibler les prospects et optimiser les ressources marketing. Ils aident à répondre à des questions telles que « Combien vont dépenser les clients, vu leur niveau de revenu ? »

Dans de nombreuses études statistiques, l'objectif est d'établir une relation, exprimée sous la forme d'une équation, pour prédire les valeurs types d'une variable en fonction de la valeur d'une autre. IBM SPSS Statistics offre plusieurs procédures pour établir des relations et définir des modèles prédictifs. Ces procédures incluent les diagrammes de dispersion et les corrélations, l'analyse de régression linéaire et logistique et les arbres de classification. Grâce aux instructions pas-à-pas et aux fonctions d'aide intégrées de la famille de produits IBM SPSS Statistics, vous pouvez exécuter ces procédures avec succès, même si vous n'êtes pas statisticien.

Avec l'analyse de récurrence, de fréquence et de valeur monétaire (RFM) d'IBM SPSS Statistics, vous pouvez analyser la somme totale dépensée en dollars ainsi que les informations relatives à chaque commande pour mieux cibler vos projets marketing et offres spéciales.

---

Le graphique 5 montre la forme de la relation entre ces deux variables. Le diagramme de dispersion est idéal pour représenter la distribution conjointe de deux variables continues ou d'intervalle. Le coefficient de corrélation de 0,608 affiché dans le tableau 10, indique une relation forte et indéniable entre le revenu du ménage et la somme totale dépensée. L'analyse de régression affine la relation avec un modèle, comme le montre le tableau 11. Cette relation montre que la somme totale dépensée pour des produits augmente proportionnellement au revenu du ménage. C'est une information précieuse qui, combinée à d'autres renseignements sur vos clients, peut vous aider à prédire combien chaque client est susceptible de dépenser.

---



Diagramme 5 : Le diagramme de dispersion montre la forme de la relation entre ces deux variables. Plus les clients gagnent d'argent, plus ils dépensent pour nos produits.

---

Avec IBM® SPSS® Decision Trees, nous pouvons identifier des segments à l'intérieur de notre base de données, basés sur la probabilité que chaque client ait une caractéristique ou un comportement spécifique que nous souhaitons prédire. Ceci est illustré dans le diagramme 6 ci-dessous :

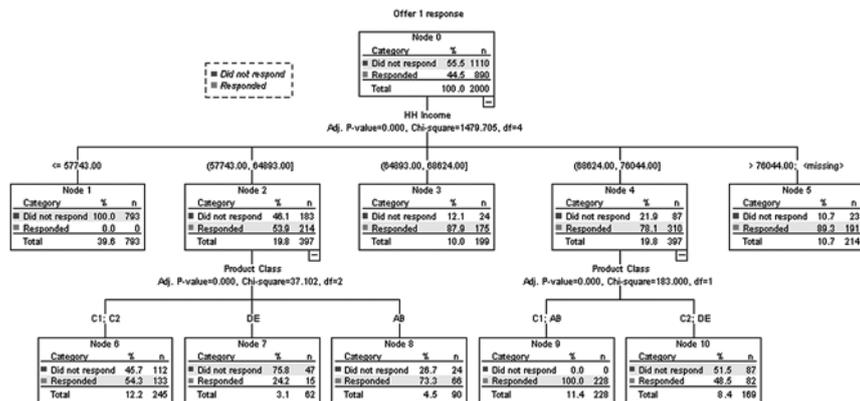


Diagramme 6 : IBM SPSS Decision Trees présente un modèle montrant que les clients affichant certaines combinaisons de caractéristiques sont les plus susceptibles de répondre à l'offre 1.

Pour commencer l'analyse, nous entrons des informations sur la zone, la catégorie de produits et le revenu du ménage dans un modèle afin de découvrir quels clients sont les plus susceptibles de répondre à l'offre 1. IBM SPSS Decision Trees peut utiliser l'un des quatre algorithmes établis de construction d'arbre pour créer une arborescence des résultats, comme illustré dans le diagramme 6.

Descriptives								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Suburban	707	\$1,391.70	\$1,640.44	\$61.70	\$1,270.57	\$1,512.82	\$0.00	\$9517.75
Urban	611	\$1,206.01	\$1,442.51	\$58.36	\$1,091.40	\$1,320.61	\$0.00	\$8402.60
Rural	402	\$1,618.27	\$1,772.04	\$88.38	\$1,444.52	\$1,792.02	\$0.00	\$8996.70
Total	1720	\$1,378.69	\$1,612.52	\$38.88	\$1,302.43	\$1,454.95	\$0.00	\$9517.75

Multiple Comparisons						
(I) Area	(J) Area	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Suburban	Urban	\$185.69	\$88.71	.112	-\$31.64	\$403.01
	Rural	-\$226.58	\$100.32	.078	-\$472.34	\$19.19
Urban	Suburban	-\$185.69	\$88.71	.112	-\$403.01	\$31.64
	Rural	-\$412.26*	\$103.14	.000	-\$664.93	-\$159.59
Rural	Suburban	\$226.58	\$100.32	.078	-\$19.19	\$472.34
	Urban	\$412.26*	\$103.14	.000	\$159.59	\$664.93

\*. The mean difference is significant at the .05 level.

Tableau 10 : La somme moyenne dépensée par les clients en réponse à nos quatre offres s'élevait à 1 378,69 \$, mais cela varie par région.

Le revenu s'avère être le plus fort prédicteur, ce qui correspond aux résultats de la régression antérieure. Si seul le revenu du ménage est pris en compte, le groupe de clients dont le revenu se situe entre 57 743 \$ et 64 893 \$, avec un taux de réponse de 53,9 pour cent, ne semble pas être une aussi bonne cible que ceux ayant des revenus supérieurs. Mais IBM SPSS Decision Trees peut aller au-delà de la simple régression linéaire pour explorer d'autres interactions entre les

caractéristiques des clients, permettant aux interactions entre les prédicteurs de se définir d'elles-mêmes, directement à partir des données au lieu de devoir être définies par l'analyste.

		HH Income	Total Orders
HH Income	Pearson Correlation	1	.608**
	Sig. (2-tailed)		.000
	N	1984	1984
Total Orders	Pearson Correlation	.608**	1
	Sig. (2-tailed)	.000	
	N	1984	2000

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Tableau 11 : Le coefficient de corrélation de 0,608 indique une corrélation forte et indéniable entre le revenu et la commande, signe que la somme totale dépensée pour nos produits augmente proportionnellement au revenu du ménage.

Lorsque l'on utilise aussi les détails du niveau de branche suivant pour comparer les segments, nous découvrons que les ménages dont le revenu se situe entre 57 743 \$ et 64 893 \$ qui ont également effectué des achats dans la catégorie de produits « AB » (nœud 8 dans les diagrammes 6 et 7) sont 21,8 pour cent plus susceptibles de répondre à l'offre 1 que les ménages du nœud 10, qui ont un revenu plus élevé mais ont effectué des achats dans les catégories de produits « C2 » et « DE ».

Category	%	n
■ Did not respond	26.7	24
■ Responded	73.3	66
Total	4.5	90

Diagramme 7 : Un aperçu détaillé d'un nœud de l'arbre de classification montre que 73 pour cent des clients dont le revenu se situe entre 57 743 \$ et 64 893 \$ et qui effectuent des achats dans la catégorie de produits « AB » sont susceptibles de répondre à l'offre 1.

IBM SPSS Decision Trees nous donne un aperçu beaucoup plus clair des sous-segments et de l'ensemble des critères qui constituent nos « meilleurs clients » que les types précédents d'analyse. Nous pourrions utiliser cet aperçu plus détaillé pour prévoir les ventes avec plus d'exactitude et améliorer nos efforts marketing.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-4075.349	162.252		-25.117	.000
	HH Income	.099	.003	.608	34.093	.000

a. Dependent Variable: Total Orders

Tableau 12 : Une régression linéaire définit la relation entre le revenu du ménage et la somme dépensée par les clients.

Les modèles prédictifs sont de puissants outils qui vous aident à répondre à des questions comme la somme que dépenseront les clients en fonction de leurs revenus.

---

## Agir

À travers les analyses décrites ici, IBM SPSS Statistics nous a permis d'analyser rapidement nos données afin d'apprendre des choses importantes sur nos clients types. Nous avons appris que ce sont plutôt des clients de longue date vivant dans des zones suburbaines. Ils ont aussi tendance à avoir un revenu de ménage supérieur à la moyenne, et n'ont pas répondu favorablement, dans l'ensemble, à l'offre 3.

En outre, en utilisant les puissantes techniques de modélisation prédictive et de segmentation pour identifier les relations, nous avons développé un modèle qui décrit la relation entre le revenu et la somme totale dépensée pour aider à prévoir les ventes futures. Nous avons également identifié des segments de clientèle selon leur susceptibilité de répondre à l'offre 1.

En comparant plusieurs caractéristiques et groupes, IBM SPSS Statistics nous a aidé à en savoir plus sur les tendances sous-jacentes. Non seulement l'offre 3 a été la moins lucrative pour nous, mais elle a été particulièrement improductive dans les zones urbaines, qui ont eu tendance à répondre avec moins d'enthousiasme à nos offres que les deux autres zones. Le fait que les clients des zones urbaines aient le revenu moyen le plus bas contribue à expliquer leur réponse relativement faible à nos offres. En identifiant de tels groupes de clients, nous pouvons mieux cibler nos programmes de marketing et de fidélisation de la clientèle.

Par exemple, comme les ménages à revenu supérieur montrent un potentiel de recettes plus élevé, nous pourrions leur proposer des produits et services supplémentaires, ou développer des programmes de fidélisation de la clientèle contribuant à en faire des clients satisfaits à long terme. Par ailleurs, nous pourrions découvrir que, bien que les clients des zones urbaines n'aient pas de façon générale répondu favorablement à nos offres, les femmes de ces zones ayant un niveau de revenu spécifique y ont répondu favorablement, ce qui laisse penser qu'il pourrait être judicieux de les cibler dans un certain type de campagne.

Suite aux analyses que nous avons effectuées, nous pourrions projeter de :

- Créer un nouveau programme de rétention de la clientèle pour nos meilleurs clients, définis comme des clients de longue date, au revenu élevé, vivant en zone suburbaine et effectuant des achats dans la catégorie de produits « AB ».
- Développer et tester une nouvelle offre groupée de produits et services pour mieux cibler les besoins des clients et prospects des zones urbaines à faible revenu.
- Reproduire le programme de développement des ventes de la zone rurale dans les zones urbaines et suburbaine pour fidéliser les clients.
- Renouveler l'offre 1 aux prospects des zones rurales.
- Répartir les fonds des futures campagnes marketing en fonction de la rentabilité prédite des segments (basée initialement sur le revenu du ménage).

## Conclusion

Ce document ne décrit que quelques-unes des façons dont vous pouvez utiliser l'analyse prédictive pour mieux comprendre vos clients. En voyant vos clients sous un certain nombre de perspectives différentes, vous pouvez prévoir des programmes plus efficaces et en mesurer systématiquement les résultats. Ainsi, vous établirez des relations plus solides avec vos clients les plus précieux et réduirez les coûts consacrés aux segments de clientèle moins précieux.

D'autres produits IBM SPSS vous permettent d'anticiper les changements dans les préférences et le comportement de vos clients. Les solutions d'analyse prédictive vous permettent de préparer de façon proactive vos stratégies commerciales et vous offrent un sérieux avantage concurrentiel dans n'importe quel secteur.

Pour les besoins de ce document, toutefois, nous avons montré que la famille de produits IBM SPSS Statistics offre un large choix d'options d'analyse, intégrées en une seule suite de produits. Même si vous n'êtes pas statisticien, vous pouvez appliquer ces informations pour mettre en place un marketing plus efficace, fidéliser vos clients les plus précieux et augmenter la rentabilité de votre entreprise.

## À propos d'IBM Business Analytics

Le logiciel IBM Business Analytics fournit des informations complètes, cohérentes et précises permettant aux décideurs d'augmenter les performances de leur entreprise. Un portefeuille complet de business intelligence, advanced analytics, financial performance and strategy management et analytic applications fournit un aperçu clair, immédiat et pouvant être mis en pratique des performances actuelles et permet de prévoir les résultats à venir. Associées à des solutions industrielles riches, des pratiques établies et des services professionnels, les organisations de toute taille peuvent dynamiser la productivité des technologies de l'information et améliorer leurs résultats.

Dans le cadre de ce portefeuille, le logiciel IBM SPSS Predictive Analytics (logiciel d'analyse prédictive IBM SPSS) permet aux organisations de prévoir les événements futurs et de se conformer de manière proactive à cet aperçu afin d'augmenter les résultats de leur entreprise. Les clients commerciaux, gouvernementaux et universitaires du monde entier se fient à la technologie SPSS d'IBM, qu'ils considèrent comme un avantage concurrentiel pour attirer, retenir et augmenter la clientèle tout en réduisant la fraude et en minimisant les risques. En intégrant le logiciel SPSS d'IBM à leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives, capables de prendre et d'automatiser des décisions afin de respecter les objectifs de l'entreprise et d'atteindre un avantage concurrentiel significatif. Pour plus de renseignements, visitez le site [www.ibm.com/spss/fr](http://www.ibm.com/spss/fr).



---

© Copyright IBM Corporation 2010

IBM Corporation  
Route 100  
Somers, NY 10589

Droits restreints pour les utilisateurs du gouvernement américain - l'utilisation, la duplication ou la divulgation sont soumises aux restrictions visées dans le contrat GSA ADP Schedule conclu avec IBM Corp.

Produit aux États-Unis d'Amérique  
Mai 2010  
Tous droits réservés

IBM, le logo d'IBM, ibm.com, WebSphere, InfoSphere et Cognos sont des marques commerciales ou des marques déposées d'International Business Machines Corporation aux États-Unis, dans d'autres pays ou les deux. Si ces termes ou si d'autres termes déposés d'IBM sont représentés pour la première fois dans ce document suivis du symbole de marque déposée (® ou TM), ces symboles indiquent des marques déposées ou de droit commun appartenant à IBM lors de la publication de ce document. Ces marques commerciales doivent également faire l'objet d'un enregistrement ou doivent être des marques de droit commun dans les autres pays. Une liste des marques commerciales actuelles d'IBM est disponible sur Internet sous « Droits d'auteur et marques » à l'adresse [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

SPSS est une marque commerciale de SPSS, Inc., an IBM Company, déposée dans de nombreuses juridictions dans le monde.

Les autres noms d'entreprises, de produits ou de services peuvent être des marques commerciales ou des marques de service d'autres organisations.



Veuillez recycler

---