

Marketing des bases de données – La segmentation RFM[1]

Introduction

La profitabilité du marketing des bases de données

Un des obstacles majeurs d'un marketing des bases de données efficace est l'absence de stratégies profitables pour l'utilisation des bases de données. L'obstacle n'est pas technique ou technologique car il y a pléthore de sociétés de services qui peuvent s'en charger. Par contre ce que ces intervenants externes normalement capables de faire, c'est de savoir comment rendre la base de données rentable.

Segmentation RFM

Une des plus anciennes et toujours une des meilleures techniques est l'analyse RFM. En utilisant cette méthode, tout homme de marketing qui dispose d'une grande base de données clients peut quasi garantir des promotions profitables envers la base de clients d'une période à une autre

Il y a des alternatives à l'analyse RFM tel la modélisation basée sur des données socio-démographiques, analyse typologique, recherche marketing, interview de groupe ou enquête auprès des consommateurs. L'analyse RFM est préférée car tous les composants sont d'ordre comportemental, c à d quand est-ce que la personne a acheté la dernière fois, combien de fois a-t-elle acheté et combien d'argent a-t-elle dépensé. Ce sont des critères qui sont disponibles et peuvent être utilisés pour prédire le comportement du consommateur. La prévision du comportement n'a besoin d'aucun autre critère. L'analyse dans ce document montrera pourquoi c'est ainsi.

Codification d'une base de données pour la segmentation RFM

La disponibilité des données

Pour effectuer l'analyse RFM, les fichiers client doivent contenir des données d'historique d'achat et utiliser un codage adapté. Cela suppose, que la base de données clients soit capable d'enregistrer en permanence la réponse du client. Cela est vrai pour les sociétés de téléphonie, pétrolières, de distribution en détail qui utilisent des cartes de fidélité propres, les vpc-istes, les assurances, le tourisme et le loisir, les banques et beaucoup d'autres.

Informations pour la RFM

Dans chaque enregistrement client il y a trois catégories d'information à maintenir:

- la date la plus récente quand le client a demandé un changement dans le service, a acheté un produit ou service etc.
- un compteur de la fréquence – le nombre de fois que le client a effectué un achat ou renouvelé un service avec la société. Pour une société de téléphone par exemple, ce serait le nombre de mois de service continu. Pour un magasin se serait le nombre de visites. Ce compteur est incrémenté chaque fois qu'un achat est effectué.
- un compteur pour le montant dépensé – le montant total des sommes dépensées par le client avec la société.

En utilisant cette information des codes RFM seront construits.

Construire un code de récence

Distribution des clients par rapport à la récence

Figure 1 - Simulation d'une BD clients en fonction de la récence – cadre générale

Vue sur la base de données

Figure 2 – Simulation d'une BD clients en fonction de la récence – Extraction d'un échantillon

.....

Tris de la base par rapport à la récence

Pour créer un code de récence, tous les enregistrements de la base sont triés selon la date de la dernière commande, du plus récent au plus ancien.

Figure 3 – Simulation d'une BD clients en fonction de la récence – Tris de l'échantillon

...

Attribution de codes de récence aux clients

L'échantillon est divisé en cinq parties égales (quintiles), le premier quintile reçoit le code 5 et le dernier le code 1.

Figure 4 – Simulation d'une BD clients en fonction de la récence – Codes de récence

...

La réponse des clients par rapport à la récence

Si une offre promotionnelle est envoyée, le résultat des réponses en fonction de la récence peut être résumé de la manière suivante :

Figure 5 – Simulation d'une BD clients en fonction de la récence – Synthèse

Analyse

Que faut-il faire ? On le sait en marketing qu'on trouve la plus grande probabilité de répondre à une campagne parmi les gens qui ont effectué un achat le plus récemment. Un client a plus de chances de réagir à une promotion s'il a eu récemment une transaction satisfaisante avec la société. Cela est vrai dans la distribution; l'informatique, l'automobile, les assurances. C'est un phénomène universel. Il peut être testé à très faible coût en lançant un programme pour insérer de codes de réquences dans la base de données et ensuite lancer une promotion de routine, les réponses par la récence peuvent être vérifiées.

Construire un code de fréquence

Distribution des clients par rapport à la fréquence

Figure 6 - Simulation d'une BD clients en fonction de la fréquence – cadre générale

Vue sur la base de données

Figure 7 – Simulation d'une BD clients en fonction de la fréquence – Extraction d'un échantillon

...

Tris de la base par rapport à la fréquence

Pour créer un code de fréquence, tous les enregistrements de la base sont triés selon la fréquence des commandes, du plus fréquent au moins fréquent.

Figure 8 – Simulation d'une BD clients en fonction de la fréquence – Tris de l'échantillon

...

Attribution de codes de fréquence aux clients

L'échantillon est divisé en cinq parties égales (quintiles), le premier quintile reçoit le code 5 et le dernier le code 1.

Figure 9 – Simulation d'une BD clients en fonction de la fréquence – Codes de fréquence

...

La réponse des clients par rapport à la récence

Si une offre promotionnelle est envoyée, le résultat des réponses en fonction de la fréquence des commandes peut être résumé de la manière suivante :

Figure 10 – Simulation d'une BD clients en fonction de la fréquence – Synthèse

Analyse

Chaque quintile, donne lieu à un code fréquence code de 5, 4, 3, 2 ou 1. Dans l'enregistrement par client, le code de fréquence sera placé juste après le code de récence. Ainsi un code à deux chiffres vient d'être créé pour chaque client qui varie de 55 (le plus récent et le plus fréquent) à 11 (le plus ancien et le moins fréquent). Chaque groupe aura le même nombre d'enregistrements.

Lors de la prochaine promotion un graphique des réponses en fonction de la fréquence pourra être tracé. Il ressemblera au graphique présenté en Figure 10. On remarquera que la différence entre le premier et le second quintile pour la fréquence n'est pas aussi grande que celle des quintiles de récence. C'est parce que la récence est un prédicteur plus fort de la réponse du client que la fréquence.

Construire un code de montant

Distribution des clients par rapport au montant de la commande

Figure 11 - Simulation d'une BD clients en fonction du montant – cadre générale

Vue sur la base de données

Figure 12 – Simulation d'une BD clients en fonction du montant – Extraction d'un échantillon

Tris de la base par rapport au montant

Pour créer un code de montant, tous les enregistrements de la base sont triés selon le montant des commandes, du plus grand au plus petit.

Figure 13 – Simulation d'une BD clients en fonction du montant – Tris de l'échantillon

...

Attribution de codes de fréquence aux clients

L'échantillon est divisé en cinq parties égales (quintiles), le premier quintile reçoit le code 5 et le dernier le code 1.

Figure 14 – Simulation d'une BD clients en fonction du montant – Codes de montant

...

La réponse des clients par rapport à la récurrence

Si une offre promotionnelle est envoyée, le résultat des réponses en fonction du montant des commandes peut être résumé de la manière suivante :

Figure 15 – Simulation d'une BD clients en fonction du montant – Synthèse

Dans l'enregistrement par client, le code de fréquence sera placé juste après le code de fréquence. Ainsi un code à trois chiffres sera créé pour chaque client qui varie de 555 à 111. Il y aura 125 cellules RFM avec le même nombre d'enregistrements chacune.

Ceux-ci devraient être recalculés et révisés à chaque fois que la base de données est actualisée – typiquement une fois par mois.

De l'espace devrait être créé dans le fichier client pour enregistrer les cellules RFM précédentes pour savoir comment les gens ont évolué durant le mois précédent – il est possible de garder la trace des cellules RFM sur plusieurs périodes, six mois par exemple.

La puissance de ces codes peut être démontrée maintenant à l'aide d'un exemple concernant un test.

La pratique de la segmentation RFM

Un mailing test

La puissance des codes RFM peut maintenant être illustrée à l'aide d'un exemple : une action de promotion test envers 40000 clients.

D'abord il faut choisir les 40000 clients de la base de données. On utilise un tirage systématique qui permet de sélectionner les enregistrements de la population de la base avec un ratio N. Pour déterminer un ratio égale à 20, on divise le nombre total de clients dans la base 800000 par la taille de l'échantillon désiré 40000. Chaque 20-ième enregistrement sera sélectionné, le 1-er, le 21-ème, le 41-ème etc.

Les 40000 clients sélectionnés pour le test forment un échantillon représentatif de la base des clients indépendamment de l'ordre dans lequel la base avait été triée (alphabétique, code postal, numéro du client etc.). L'échantillon sera représentatif pour les codes RFM aussi. Il y aura 125 cellules RFM différentes et chacune comportera le même nombre de clients.

Figure 16 – Simulation d'un mailing test

Les réponses par segment RFM

Dans cet exemple l'offre faite aux 40000 clients concerne un produit qui coûte approximativement 3120€ et le profit variable d'une vente réussie est de 335€.

Tableau 1 - Réponses à un mailing test envers 40.000 adresses

Le coût du mailing (qui inclut la création, impression, la personnalisation et le postage) est de \$0.62 par pièce. La réponse à cette offre peut ressembler à ce qui est affiché dans le Tableau 1.

Ce sont seulement les premières 20 cellules. Il y a 125 au total. Les cellules inférieures ont en général des taux de réponse plus faibles. Celles qui sont vers les bas n'ont pas de réponse du tout. (Toutes les cellules ont exactement la même taille à cause de la méthode avec laquelle elles ont été créées. Chaque quintile (R, F, et M) représente exactement 20 pour-cent du fichier. Il résulte que chacune des 125 cellules aura exactement 5 pour-cent de l'ensemble du fichier. (Si le codage a été fait correctement, il n'y aurait pas des cellules plus grandes ou plus petites).

Résultats

Tableau 2: campagne test utilisant un fichier de clients de la distribution en détail

Les résultats de cette promotion test peuvent ressembler à ceux du Tableau 2,

qui sont extraits d'un fichier de clients de magasin en détail ou le produit communiqué en promotion était des vêtements d'homme et de femme vendus à un prix de \$125.

Discussion

La plus part des gens considéreraient cette promotion comme un échec. Approximativement \$25,000 ont été dépensés pour un mailing qui a apporté un profit de seulement \$85 après avoir couvert toutes les dépenses. Cela ne couvre pas le temps dépensé pour planifier l'action de promotion, même si ça couvre les coûts de mailing. Cette promotion sans succès est d'une grande utilité, car c'est une source d'apprentissage des taux de réponse pour chaque cellule RFM. Cela a une grande valeur.

La théorie derrière l'analyse RFM est que le meilleur prédicteur du comportement futur est le comportement passé et présent. Il est certainement meilleur que les données socio-démographiques ou géographiques. Comme les cellules RFM sélectionnées pour le test sont identiques à celles de la base entière, on peut supposer que ces clients pas encore démarchés répondrons de la même manière s'ils reçoivent la même offre. Comme on sait si une cellule donnée répond ou pas, la promotion envers les cellules non profitables peut être évitée et l'attention sera accordée aux cellules profitables.

Determiner les cases RFM profitables

Trouver le point d'équilibre

Les cellules profitables sont celles dont le taux de réponse est supérieur au point mort. Il existe une formule simple pour calculer le point mort:

Point d'équilibre = coût unit. mailing / revenu net d'une vente

Dans l'exemple le coût d'un mailing était de \$0. ,62 et le revenu obtenu d'une vente était \$35.

Par conséquent le taux de réponse d'équilibre est:

$$\$0,62/\$35 = 1,77 \%$$

Figure 17: Profitabilité des cellules RFM suite à un mailing test (40000 adresses)

Eliminer les cases en dessous du point d'équilibre

Toute cellule RFM qui à un taux de réponse prévu de 1,77 pour-cent ou plus devrait être approché. Le reste doit être abandonné. Une représentation graphique des cellules profitables et non-profitables est présentée dans la Figure 16.

Les données pour ce graphique ont été préparées en divisant le taux de réponse par le taux du point mort, en multipliant par 100 et soustrayant 100. Les taux du point mort apparaissent pour cette raison égales à zéro. Il s'agit d'un index de profitabilité. Toute cellule ayant un score supérieur à zéro sera profitable, pendant que celles qui se trouvent en dessous seront non profitables.

En connaissant la profitabilité prévue, il est facile de faire du profit d'une campagne de promotion: il faut envoyer uniquement aux cellules profitables.

Etape préliminaire – déflater les résultats du test

Il est bien connu que le mailing test donne souvent des résultats supérieurs à ce qu'on obtiendra sur l'ensemble de la base de données clients. L'exemple illustre cette situation et propose des calculs de rentabilité comparatifs.

Figure 18 – Simulation et profitabilité d'une opération de marketing direct

Calculs

Tableau 3 – Calculs comparatifs de profitabilité

Il y a quant même un pas préliminaire, qui va assurer le succès – les résultats du test doivent être déflatés. C'est une chose bien connue en marketing direct marketing que le test donne de résultats meilleurs que la vraie campagne. Ce ci signifie que si le taux de réponse de 2 pour-cent au test, il y a sûrement 2 pour-cent de moins quand le fichier en entier sera envoyé. Une explication plausible est que les marketeurs ne supportent pas d'avoir des tests sans succès. Ils feront toujours quelque chose pour que le test ait un bon taux de réponse. Ils supposent que si leur test ne fonctionne pas, leur réputation peut en souffrir, donc ils sélectionnent les meilleures listes clients, les meilleurs moments dans l'année, les meilleures offres – dans beaucoup de cas il s'agit de choses qu'on ne peut pas répliquer durant une vraie campagne de promotion.

Dans ce cas, la liste ne peut pas être trafiquée même si un N-ième à été utilise. En tous cas il y a des choses subtiles qui déforment les résultats du test. Pour cette raison, il est recommandé de déflater les résultats du test de 15% pour

déterminer les réponses de cellules RFM dans la campagne. 25% ou un autre nombre peut être utilisé. Cela signifie que s'il y a eu un taux de réponse de 2 pour-cent dans une cellule RFM on devrait considérer qu'en réalité il y aura seulement 1,7 pour-cent ($2 \times 0,85$) taux de réponse pour la même cellule à l'occasion d'un mailing plus ample.

Utiliser les taux de réponses déflatés

En utilisant ces taux de réponse déflatés, seulement les cellules qui dépassent le point mort devraient être prises en compte. Les résultats peuvent être étonnants, comme dans le tableau 3. Le tableau 3 montre que si la campagne est menée sur l'ensemble de la base, le taux de réponse aura été de 1.15 pour-cent (c à d 85 pour-cent du taux du test 1.78 pour-cent). 12.104 commandes auraient été passées avec un profit net \$35 par commande. Cela aurait apporté un revenu de \$423,640. Mais la campagne totale aurait donné une perte de 572,360, après avoir déduit les coûts de mailing de 8496,000.

En utilisant les résultats du test déflaté, il y a seulement 211,200 clients qui se trouvent dans des cellules RFM profitables dont le taux de réponse déflaté est supérieur à 1.77 pour-cent ou supérieur. Le taux de réponse déflaté est obtenu en baissant les résultats du test pour chaque cellule RFM de 15 pour-cent. En adressant seulement ces 211200, on obtient seulement 7293 commande pour un revenu net de \$255,255. Comme les coûts de mailing sont de seulement \$130,944, le profit net du projet est de \$124,311.

Ce n'est pas une coïncidence mais de l'analyse RFM normale. Elle va toujours fonctionner si des bonnes données existent pour construire de cellules RFM valables et si les méthodes sont suivies avec rigueur. Presque tout test valable peut être transformé dans une campagne de marketing direct profitable.

Compléments

Dimensions de l'échantillon test

Une question qui se pose est quelle doit être la taille de l'échantillon de test pour être valable. Une règle de base est que le nombre de réponses attendues soit supérieur ou égal à 500. Pour connaître cette taille il est nécessaire d'avoir une idée sur le taux de réponse attendu. S'il devait y avoir normalement un taux de 5 pour-cent de la part des clients, alors pour produire 500 réponses, seulement 10000 clients doivent être testés.

Le nombre de réponses attendu devrait être de 500 ou plus car il y a 125 cellules RFM; 500 signifie que la cellule RFM moyenne aura quatre répondants ($500/125$). Quand la cellule moyenne est inférieure à quatre, la chance devient

plus importante que le comportement du consommateur. Bien-sur, plus le test est grand, plus les résultats sont précis. Un test de 40000 sera plus précis qu'un test de 30000. Le coût sera évidemment plus grand. Est-ce que le coût supplémentaire de 10000 est justifié pour un plus de précision dans la sélection des adresses pour la campagne? Cette décision devra être faite en fonction des situations particulières de chaque société et les expériences doivent être répétées pour trouver la taille idéale du test pour la base de données clients.

Quelques exceptions

Comme on peut le voir de l'analyse, RFM fonctionne seulement avec des bases de données clients et seulement avec des bases qui contiennent les données nécessaires. Elle ne peut pas être utilisée avec la base des prospects. Si une base de données clients n'a qu'une partie des données nécessaires (par exemple la récence et la fréquence, mais ne dispose pas de montants dépensés par commande) on peut toujours tirer profit en essayant la technique.

Déterminer la récence peut poser des problèmes. Pour la distribution d'électricité, la société de téléphone, la presse ou les sociétés d'assurance, par exemple, les clients reçoivent souvent des reçus mensuels et ont à payer leurs factures tous les mois. Si la dernière facture payée était utilisée comme récence, chaque client aurait la récence un, ce qui ne serait pas d'une grande utilité. Dans de telles situations, on peut prendre comme la date la plus récente, la dernière fois quand les clients ont changé leur service ou la dernière fois quand ils ont appelé le service clientèle ou une autre date similaire. L'idée est de trouver la dernière date quand ils ont pensé à l'entreprise et ont fait quelque chose. A de tels moments, l'entreprise a été évoquée dans leur mémoire consciente.

Personnalité des cases RFM

La méthode RFM a été illustrée en considérant que la campagne allait se dérouler par voie postale ou par télémarketing. Il y a nombreuses autres utilisations. Si par exemple il est décidé de créer des segments de clients spéciaux, qui recevront un traitement spécial (détenteurs de cartes "gold", par exemple), RFM est une méthode idéale pour trouver ceux qui iront dans la plus haute catégorie.

Chaque cellule RFM a sa propre personnalité. Tous les nouveaux clients entrent comme des 511. Ils sont les plus récents, mais d'habitude ils sont les moins fréquents et ont les plus faibles montants. Après leur entrée ils peuvent évoluer vers le haut ou vers le bas, en fonction de leur comportement ultérieur. Il convient d'enregistrer où se trouvent les gens Durant le dernier mois. On peut fixer des objectifs à atteindre pour chaque cellule, d'amener les nouveaux clients à avancer dans leur deuxième mois de 511 à 512, par exemple, au lieu de tomber vers 411.

Ce du groupe le plus bas peuvent ne pas être des clients du tout. Ceux ci pourraient être archives, par extraction de la base de données, après avoir essayé de les réactiver par un mailing. Les garder dans la base année après année peut simplement gaspiller l'argent de l'entreprise. Les 155s ne devraient pas être purges sans des efforts de réactivation préalable, on doit leur accorder une attention toute particulière.

Maintenir le contact

Si les cellules RFM sont utilisées comme bas pour le contact de l'entreprise avec ses clients, le contact avec certains clients pourra être perdu pour de bon – ceux qui se trouvent en dessous du point mort dans les tests. Cela peut aussi bien être bien ou mal. De toute manière l'objectif de construction d'une relation avec les clients est lié à la possibilité de construire des profits. Quatre-vingt pour-cent des revenus de l'entreprise viennent d'habitude des premiers 20 pour-cent – ceux qui sont dans 5 au niveau du montant. La discrimination est utile pour l'entreprise et en même temps bien acceptée par les clients. Bien venue car certaines personnes ne veulent pas être embêtées par leurs fournisseurs tout le temps. Un 111 essaye peut-être de dire "Laissez moi tranquille, s'il vous plait". Les deux parties gagnent de la séparation.

Influencer le comportement

Est-ce que l'analyse RFM influence le comportement du consommateur? Pas vraiment. RFM est simplement utilisé comme méthode de segmentation de la base des clients à partir du comportement passé et en l'utilisant on arrive à prédire les comportements futurs. Le comportement du client est plus influencé par la communication qui est faite dans les catalogues etc.

Marketing des bases de données – La segmentation RFM... 1

M. Calciu – notes de cours. 1

Introduction. 1

Codification d'une base de données pour la segmentation RFM... 1

Construire un code de récence. 2

Construire un code de fréquence. 4

Construire un code de montant 6

La pratique de la segmentation RFM... 9

Determiner les cases RFM profitables. 11

Compléments. 13

Ressources. 16

Ressources

Exemples sur tableur: Initiation aux bases de données et au marketing direct