

*Institut Universitaire Technologique d'Avignon et des pays du Vaucluse
Département Statistique et Traitement Informatique des Données
Promotion 2008 - 2010*

Rapport de stage :

Détecter les appétences multicanaux: les courriers commerciaux

Rédigé par Quentin GUIGUE



Tuteur: M. Pierre Michel BOUSQUET

Maitre de stage: Mlle. Marie-Laure PAUTU et M. René Pierre DABON

Durée: 12 Avril – 18 Juin 2010

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de cette mission.

Un grand merci à René Pierre DABON, responsable de la connaissance client du service marketing de CAP, pour la confiance qu'il m'a accordé tout au long de ce stage.

Ensuite, j'aimerais remercier Marie Laure PAUTU pour nos échanges constructifs, son écoute, son aide et ses conseils qui ont permis d'améliorer significativement mes compétences.

De plus, merci à Mr BOUSQUET pour sa disponibilité et son implication tout au long des dix semaines de stage.

Enfin, j'aimerais remercier les personnes du service marketing pour leur accueil et leur sympathie ainsi que toute l'équipe pédagogique qui m'a donné les outils nécessaires afin de réussir au mieux ma mission.

Résumé

J'ai effectué mon stage au sein de la connaissance client du service marketing du Crédit Agricole Alpes Provence. Le sujet de la mission qui m'a été confiée était de détecter les clients appétents au canal mailing.

Pour réussir ce stage, l'objectif était de mettre en place un score ou un indicateur de "clients appétents au mailing". Ce rapport présente toutes les étapes réalisées, du diagnostic initial à la mise en œuvre pratique des résultats, dans le but d'atteindre l'objectif fixé et ainsi réussir la mission.

Abstract

I have done my internship within the knowledge of the client of the marketing department of Credit Agricole Alpes Provence. The theme of my mission was to detect the client attracted by the mailing channel.

In order to make it, the goal was to create an indicator oh the clients attracted by mailing. This report presents all the steps that have been realised, from the initial diagnostic to the practice implementation of the results with the goal to achieve the objective sets and thus, succeed my mission.

SOMMAIRE

I.	<u>Introduction</u>	page 6
II.	<u>Présentation de l'entreprise</u>	page 7
	II.1. Le groupe Crédit Agricole	page 7
	II.1.1 Son histoire	
	II.1.2 3 échelons pour une meilleure organisation	
	II.1.3 La situation du groupe résumée en quelques chiffres	
	II.2. CAP : le Crédit Agricole Alpes Provence	page 8
	II.2.1 Son histoire	
	II.2.2 Le service marketing	
III.	<u>Périmètre de l'étude</u>	page 10
	III.1. Présentation de la mission	page 10
	III.1.1 Contexte et problématique	
	III.1.2 Données : bases Marketing	
	III.2. Organisation du travail	page 12
	III.2.1 Les différentes étapes	
	III.2.2 Planning	
IV.	<u>Diagnostic et réalisation de la base pour l'étude</u>	page 14
	IV.1. Approche du multi canal : état des lieux	page 14
	IV.2. Détermination de l'appétence (variable cible)	page 15
	IV.3. Variables prises en compte dans l'étude	page 16
V.	<u>Analyses descriptives et prédictives</u>	page 19
	V.1. Une première approche statistique : analyses univariées	page 19
	V.2. Typologie des clients appétents	page 21
	V.3. Analyse discriminante linéaire	page 23
	V.4. Analyse discriminante quadratique	page 24
	V.5. Régression logistique	page 25
VI.	<u>Mapping</u>	page 26
	VI.1. Principe	page 26
	VI.2. Classification et taux de pénétration	page 26
	VI.3. Résultats	page 28
VII.	<u>Préconisations</u>	page 29
	VII.1. Choix du modèle	page 29
	VII.2. Phase de test	page 30
	VII.3. Mise en œuvre	page 31
VIII.	<u>Conclusion</u>	page 33

I. INTRODUCTION

J'ai réalisé ce stage de fin d'étude au service marketing du Crédit Agricole Alpes Provence afin de mettre en application et valider mes connaissances acquises tout au long de mes deux années de DUT STID.

Ce rapport présente les résultats obtenus dans le cadre de l'analyse de l'appétence des clients du Crédit Agricole Alpes Provence aux courriers commerciaux.

Jusqu'à présent, le client d'une banque était contacté par son agence uniquement par téléphone et par courrier. Aujourd'hui, ce n'est plus le cas. En effet, avec l'avancée d'internet et des nouveaux téléphones portables, le client peut être contacté par divers moyens tel que le SMS, le MMS ou le courrier électronique en plus des canaux déjà existants. Ma mission est donc de détecter si le client est plus susceptible d'adhérer à l'offre commerciale suite à la réception d'un courrier ou bien s'il faut privilégier un autre canal.

Les enjeux de la mission sont de réduire les coûts d'envoi d'au moins 30%, d'obtenir une efficacité au moins égale ($\text{efficacité} = \text{nombre de souscription} / \text{nombre d'envoi}$) et d'appliquer systématiquement les résultats de la mission défini à chaque envoi de mailing. De plus, un autre enjeu est de permettre aux autres Caisses Régionales de profiter des résultats.

Ce rapport présentera dans un premier temps l'entreprise puis le plan de travail mis en place afin de réussir la mission. Ensuite, nous évoquerons le déroulement de la mission, ses étapes et ses résultats. Enfin, d'un point de vue pratique, nous aborderons la mise en place des résultats de la mission pour les prochaines campagnes de mailing.

II. Présentation de l'entreprise

II.1. Le groupe Crédit Agricole

II.1.1. son histoire

Le crédit Agricole est fondé à la fin du XIX^{ème} siècle. En effet, La première caisse locale est créée en 1885 dans le Jura, à Salins les Bains. En 1894, les caisses locales sont réglementées par une loi qui ne donne aucun avantage financier et celles-ci se trouvent confrontées à des problèmes financiers. Cependant l'Etat prend des mesures partielles et impose à la Banque de France d'accepter à l'escompte les effets des syndicats agricoles et d'apporter des ressources au Crédit Agricole sous la forme d'une dotation de 40 millions de francs or et d'une redevance annuelle de 2 millions.

Pour répartir ces avances, la loi du 31 mars 1899 institue les **Caisses Régionales de Crédit Agricole Mutuel** (au nombre de 9 la première année). Ces caisses coopératives, deuxième niveau dans la pyramide institutionnelle du Crédit agricole, fédèrent les caisses locales de leur ressort et encouragent la création de nouvelles Caisses Locales.

Dans les premières années, l'activité est d'abord exclusivement composée de prêts court-terme. Il s'agit d'avances sur récoltes qui permettent aux agriculteurs de mieux vivre. Viendront ensuite les prêts à moyen puis long terme qui leur permettront de s'équiper, d'acheter du bétail.

Ce n'est qu'en 1920 qu'apparaît, sous l'égide de Louis TARDY, l'Office National du Crédit Agricole, devenu en 1926 Caisse Nationale du Crédit Agricole. Le troisième échelon de l'institution de la banque est donc créé.

Une dernière étape, le 20 février 1991, donnera le pouvoir au Crédit Agricole de financer toutes les entreprises et professions dans le monde entier.

I.1.2 Le crédit agricole: 3 échelons pour une meilleure organisation

Echelon national: La Fédération Nationale du Crédit Agricole (FNCA) est l'instance de réflexion, d'expression et de représentation des Caisses régionales. La FNCA assure la synthèse des débats nés du terrain, ce qui fait d'elle une force de proposition et d'orientation. D'autre part, elle représente les Caisses régionales et le Groupe auprès des pouvoirs publics, des organisations professionnelles agricoles et des instances du monde coopératif et mutualiste.

Echelon régional: Sociétés coopératives et banques régionales de plein exercice, elles offrent une gamme complète de produits et services financiers à leurs clients.

Echelon local: Socle du Crédit Agricole, les caisses locales détiennent l'essentiel du capital des Caisses régionales. Acteurs de la vie locale et à l'écoute de ses besoins économiques, les administrateurs des Caisses locales permettent au Crédit Agricole d'adapter son offre de produits et services aux attentes de ses clients.

I.1.3 La situation du groupe résumée en quelques chiffres

Aujourd'hui, le groupe Crédit Agricole c'est:

- 36 filiales dans 70 pays du monde
- 11 500 agences bancaires
- 59 millions de clients
- le leader de la banque de détail en Europe
- le premier acteur bancaire en France avec 25% du marché des ménages
- 39 caisses régionales
- 2544 caisses locales

II.2 CAP: le Crédit Agricole Alpes Provence

II.2.1. son histoire

Avant 1988, les caisses régionales étaient au nombre de 94. Pour renforcer leur compétitivité et leur solidité, leur nombre est passé à 34. Dans la région, les caisses des Hautes Alpes, des Bouches du Rhône et du Vaucluse ont été réunies pour créer CAP: Crédit Agricole Alpes Provence. (*Annexe 1: organigramme général du Crédit Agricole Alpes Provence*)

Cette caisse est le premier employeur de sa circonscription avec 13 000 agents commerciaux répartis sur 200 points de ventes et 700 agents administratifs répartis sur 5 sites:

- Arles: organisation logistique, plate forme de maintenance informatique, direction financière et des risques, direction des entreprises.
- Aix en Provence (siège social de CAP): Direction générale, audit, contrôle de gestion, communication et ressources humaines
- Avignon: Services bancaires, marché des particuliers et des professionnels, **marketing stratégique et opérationnel**
- Gap: services bancaires
- Marseille centre de traitement des chèques

Cependant, un site unique sur Aix en Provence est en cours de construction et sera opérationnel pour l'année prochaine (2011) afin de réunir toute l'administration et ainsi faciliter les communications entre les services.

II.2.2. Le service marketing

De manière commune, le marketing cherche à déterminer les offres de biens, de services ou d'idées en fonction des attitudes et de la motivation des consommateurs, du public ou de la société en général. Celui de CAP ne déroge pas à la règle. Le but de ce service est de trouver les moyens pour intensifier la relation avec les clients particuliers et professionnels afin que la relation client/banque s'installe dans la durée et la stabilité. (*Annexe 2 : organigramme du service marketing du Crédit Agricole Alpes Provence*)

Le service se compose de deux groupes (figure 1) qui travaillent en étroite collaboration:

- La Connaissance Client s'appuie sur le traitement et la gestion de données multiples afin de proposer une exploitation immédiatement opérationnelle au service du Développement de la Relation Client : proposer le bon produit, au bon moment, à la bonne personne, selon le mode relationnel qui lui convient.

Ce service se compose notamment de la partie datamining et scoring. A partir de données de références unifiées dans la banque (base clients), le datamining permet le rapprochement de la situation personnelle et des besoins du client restant à satisfaire. L'analyse par segment de clientèle, moment de vie et univers de besoin permet notamment la génération de motifs de contact qui alimentent les différents programmes relationnels (motifs client). Des méthodes statistiques valorisent les données clients et permettent la définition de scores pour détecter le comportement ou l'appétence des clients à certains produits.

- Le développement de la relation client permet de développer et fidéliser les différents segments de clientèle en proposant via les différents modes de contact, une gamme de solutions complètes et adaptée à leurs besoins.

Figure 1



III. Périmètre de l'étude

III.1. Présentation de la mission

III.1.1. Contexte et problématique

Cette mission s'inscrit dans le cadre du développement multi canal de CAP. En effet, la pluralité des outils pour contacter le client a considérablement augmenté ces dernières années. Outre les canaux classiques (téléphones et courriers), d'autres se sont développés comme l'e-mail, le SMS ou le MMS à tel point qu'aujourd'hui, contacter le client à bon escient est devenu un véritable enjeu pour la banque. Effectivement, les contacts doivent être suffisamment nombreux pour inciter le client à souscrire des produits mais pas trop afin de ne pas le saturer. Il faut donc, ce qui est déjà le cas, cibler l'offre aux clients qui en ont besoin.

Cependant, La diversité des moyens de contacts n'est pas encore optimisée. C'est dans ce cadre qu'intervient ma mission de stage. J'ai donc pour but de:

- détecter des cœurs de cible appétents au courrier (à qui écrire en priorité)
- détecter des cœurs de cibles non appétents au courrier afin de privilégier l'envoi par un autre canal (e-mail, sms, téléphonie ...)

Le choix de se concentrer sur le canal mailing s'est avéré naturel pour plusieurs raisons: d'abord, c'est celui qui reste le plus développé malgré l'essor des autres types de communications. Ensuite, et surtout, la lettre étant la méthode la plus couteuse pour contacter un client (trois fois plus cher qu'un MMS), l'envoi à un nombre réduit de personnes pour autant de réussite, permettrait de diminuer considérablement les couts de celui ci.

D'autre part, il est important de noter qu'un sondage CSCA du début d'année 2010 montre que 3 consommateurs français sur 4 jugent inconcevable de ne plus recevoir de courrier papier. Ainsi, l'appétence mailing est un vrai enjeu pour CAP.

III.1.2. Données : bases marketing

Les bases marketing sont des bases de données métier dans lesquelles sont répertoriées une multitude de renseignements sur les clients (voir figure 3). Ces datamarts se composent de deux parties: les particuliers et les professionnels. Ces tables sont des concaténations de plusieurs tables provenant d'une banque de données où il y a l'intégralité de la signalétique des clients. Ces bases sont remises à jour mensuellement et donnent l'intégralité des renseignements pour les études marketing. Par exemple, la base DIAGO se compose de 189 variables de type numérique ou alphanumérique et certaines de ces variables sont la précision d'autres variables.

Figure 3

■ **Activités Commerciales** => *Consommation du client*

- Ventes et propositions par canaux
- Liste des canaux de ventes
- Dico. en Pavés par canaux de vente : ex. nb ventes, nb propositions, nb entretiens

■ **Communication** => *Contacts reçus par le client*

- Historiser tous les contacts avec le client
Courrier CR, Courrier externe, E-Mail, SMS, msg BAM, vignette BAM, Motif de Contact, Plateforme tél. ...

■ **Comportement** => *Contacts émis par le client*

- Comportement du client
Il est venu en agence ? Il a appelé ? Quel service ? Il a fait 1 réclamation ? Il est allé sur Internet ?

■ **Équipement** => *Équipement du client*

- Diago et Univers de besoin
A compléter ? Ajout de variables ? Suppression de variables ? Distinguer la signalétique de l'équipement ?

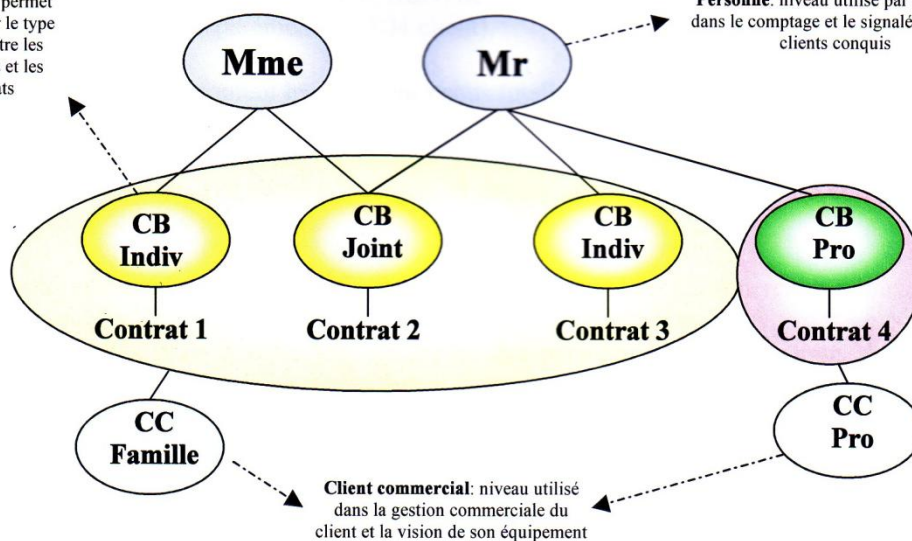
La particularité de ces bases est que l'on réalise les études non pas sur le client lui-même mais plutôt sur le Client Commercial (CC) qui est l'équivalent d'un ménage (figure 4). Par exemple, une famille de parents et de deux enfants ne représentera pas quatre individus mais seulement un CC.

Durant la mission, je trouve toute les données dont j'ai besoin dans ces bases. (*Annexe 3: dictionnaire des données de la base des particuliers*)

Figure 4

Client bancaire:
niveau qui permet
de qualifier le type
de lien entre les
personnes et les
contrats

Personne: niveau utilisé par exemple
dans le comptage et le signalétique des
clients conquis



III.2. Organisation du travail

III.2.1. Les différentes étapes

Afin d'optimiser les chances de réussite de la mission, il a fallu se positionner sur les étapes de celle-ci. Effectivement, un plan élaboré à partir de la problématique permet de cadrer les événements et de donner des rapports à chaque fin d'étape.

Voici le plan de travail développé au préalable:

- Approche générale : état des lieux multi canal en Mars 2010
 - Prendre conscience et connaissance des points forts et faibles de la base pour chaque canal
- Détermination d'une variable d'appétence (cible)
 - Recensement de toutes les campagnes de mailings commerciaux envoyées en 2009 et lors du premier trimestre 2010
 - Choix des courriers pris en compte dans l'étude et détermination du périmètre de l'étude
 - Calcul de l'efficacité de chaque courrier par Client Commerciaux
 - Détermination de l'appétence ou non aux courriers pour chaque CC ayant au moins reçu un courrier commercial
- Première approche statistique : analyse univariée
 - Chercher des préconisations pour les mailings (appétents et non appétents)
 - Permettre la création de filtres afin de redresser le score pour les cas particuliers en fin de mission
- Sélections de variables afin d'expliquer l'appétence
 - Sélection des variables importantes à étudier issue de l'expérience métier
 - Chercher les fortes corrélations entre ces 39 variables afin d'éliminer les doublons
 - Réaliser une régression logistique selon les 3 méthodes (Stepwise, Backward, Forward) afin de sélectionner les variables qui caractérisent le plus la différence entre les appétents et les non appétents
- Profil type des clients appétents : modélisation des cœurs de cibles
 - ACM: afficher une vue qui montre les caractéristiques des points
 - Classification: regrouper les individus semblables pour étudier leur profil
- Analyse prédictive : prédire l'appétence au courrier d'un client commercial:
 - Analyse discriminante : prédire si un client pas encore contacté est appétent ou non appétent au canal courrier

III.2.2. Planning

Toutes ces étapes doivent être réalisées dans une durée de dix semaines. Il a donc fallu appréhender et répartir au mieux le temps pour chaque partie afin de ne pas se trouver en retard à la fin du stage. De ce fait, j'ai dû réaliser un rétro-planning pour permettre d'optimiser l'avancement du projet. Cependant, quantifier la charge de travail de chaque étape est très difficile à réaliser. C'est pourquoi l'aide de Marie Laure PAUTU et René Pierre DABON qui ont l'habitude de développer, gérer et tenir un planning, m'a été très bénéfique car je me suis vraiment rendu compte à quel point il était important de créer des rétro-plannings. La figure 5 présente ce rétro planning à l'état initial.

Figure 5

Jour	Avril	Mai	Juin	Jour
1				1
2				2
3				3
4				4
5				5
6				6
7				7
8				8
9				9
10				10
11				11
12				12
13				13
14				14
15				15
16				16
17				17
18				18
19				19
20				20
21				21
22				22
23				23
24				24
25				25
26				26
27				27
28				28
29				29
30				30
31				31

	I. Etat des lieux
	II. Choix mailing
	III. Efficacité par campagne + Collage base de données DIAGO et campagnes
	IV. Déterminer appétence courrier et créer variable

V. Nettoyage base de données: élimination doublons et sélection variables actives	
VI. Profil types des appétents et non appétents aux courriers	
VII. Scoring sur l'appétence au canal courrier (filtre/test/validation)	
VIII. Autre méthode si aucun résultat	

IV. Diagnostic réalisation de la base pour l'étude

IV.1. Approche du multi canal: état des lieux

En premier lieu, le premier objectif était d'établir un état des lieux des renseignements pour chaque canal à l'heure d'aujourd'hui. Afin d'y parvenir, j'ai dû effectuer sur la base DIAGO de mars 2010 quelques opérations avec ACCESS comme l'utilisation des fonctions ET et OU. Le but était simple: savoir combien de clients peuvent être contactés pour chaque canal. On enlève donc ceux qui sont réfractaires puis ceux dont l'adresse n'est pas renseignée et on trouve de ce fait le nombre de CC que l'on peut contacter par ce canal. Voici les résultats de cette étude:

- Données générales: La base DIAGO des particuliers regroupe 488 778 clients commerciaux, ce qui est le niveau utilisé dans la gestion commerciale du client et la vision de son équipement. 43 892 d'entre eux ont une partie professionnelle (8.98%). Parmi tous ces clients commerciaux, il est intéressant de noter que 95% d'entre eux sont gérés en réseau de proximité (en rapport direct avec l'agence). De plus, lorsqu'un ciblage est effectué, on élimine généralement 48 714 cc (presque 10%) d'office parce qu'ils ont des critères rédhibitoires :

- top_npai : mauvaise adresse
- top_dcd : personne décédée
- top_sseq : cc sans équipement (sans contrat)
- top_min : cc de personnes mineures
- top_inc et/ou top_avec_ps : cc de personnes sous tutelles

- Partie Mailing: Tout d'abord, il est intéressant de noter que seulement 869 cc sont réfractaires au canal courrier (0,18%). Ensuite, plus de 95% des CC (soit 465 474) ont une adresse bien renseignée et sont non réfractaires, ce qui est très intéressant. De plus, seulement 54 CC ont une adresse non renseignée (vide) et 22 448 ont une mauvaise adresse (top_npai). *Ce canal est le mieux renseigné mais le plus couteux.*

- Partie e-Mailing: la partie mail est sûrement la moins bien renseignée. C'est tout d'abord le canal où il y a le plus grand nombre de réfractaires avec 81 618 cc soit 16,70% de la base DIAGO. Ensuite, seulement environ 23%, soit 112 300 adresses mail de cc sont renseignées et potentiellement utilisables. *Ce canal est peu couteux mais c'est le moins bien renseigné et là où il y a le plus de réfractaires.*

- Partie téléphone: le canal téléphone peut être traité en deux sous parties avec les téléphones fixes et les téléphones portables capables de prendre en charge le SMS et le MMS. On sait que 60 991 cc sont réfractaires au téléphone, SMS et MMS, ce qui représente 12,48% de la population de la base DIAGO. A propos des téléphones portables, on compte 202 196 cc renseignés, soit à peu près 41%. En ce qui concerne le téléphone fixe, il est difficile de faire un état des lieux mais on peut en tirer les informations suivantes :

- 2 268 cc qui commencent par 01*
- 1 001 cc qui commencent par 02*
- 972 cc qui commencent par 03*
- 281 668 cc qui commencent par 04*
- 1 404 cc qui commencent par 05*
- 615 cc qui commencent par 08*
- 4 527 cc qui commencent par 09*
- 292 455 numéros de téléphone de cc bien renseignés au minimum (soit 60%)
- 47 489 cc ont un numéro de portable à la place du fixe

Ce canal est moyennement cher et renseigné.

- Partie PAP / PAP personnalisé: le PAP est un petit encart de publicité qui s'affiche lorsque les CC vont consulter leur compte sur Internet. Ce canal, de par sa gratuité, est très intéressant. Il y a 171 452 cc bameurs répertoriés sur la base DIAGO de mars, c'est-à-dire qu'il y a au moins une personne du CC qui est allée vérifier ses comptes sur Internet lors des trois mois précédents. De plus, le fait qu'il n'y ait pas de réfractaire rend ce canal attractif. *Ce canal est attractif et gratuit.*

- A retenir: Seulement 119 cc sur 488 778 sont réfractaires aux 3 canaux (téléphone, mail, courrier). Avec cette étude, on se rend compte que dans tous les cas, on peut contacter le client. Reste à savoir le ou les appétence(s) de chaque type de cc.

IV.2. Détermination de l'appétence (variable cible)

Afin de déterminer l'appétence, il a fallu, en premier lieu, répertorier toutes les campagnes commerciales de mailing de l'année 2009 et du premier semestre 2010. Ce travail de recherche d'informations a été accompli non sans mal car il fallait vérifier chaque information et communiquer dans un langage, celui du monde de la banque, que je ne connaissais pas jusqu'à présent. Ensuite, pour chaque campagne retenue, 16 au total (*Annexe 4*: exemples de lettres envoyées), j'ai recherché la date d'envoi, l'offre d'achat afin de calculer si le client avait souscrit au produit proposé. Ce calcul est intuitif : on regarde la date d'envoi du courrier, le produit à souscrire puis on voit si le client a acheté l'offre dans les 3 mois. On procède avec ACCESS en faisant la soustraction entre le nombre de produit concernant l'offre 3 mois après l'envoi du courrier et le mois précédent l'envoi du courrier. Si c'est supérieur ou égal à 1, le client a souscrit au produit et on peut donc penser que le courrier a eu l'effet escompté. Le tableau suivant (figure 5) résume les campagnes choisies, la volumétrie et spécifie les pourcentages d'efficacité (voir aussi: *Annexe 5* : efficacité des campagnes choisies). Il est intéressant de noter qu'en général, une campagne de mailing très bien réalisée obtient un taux de retour d'environ 2%.

Figure 5

THEME MAILING	DATE D'ENVOI	VARIABLES	EFFICACITE	% D' EFFICACITE
P9	05 Janvier. 2009	02_PREDISSIME9	303 / 8130	3,73%
EPARGNE carre mauve1	05 Février. 2010	NBPEL + 02_PREDISSIME9	177 / 6563	2,70%
MICRO CIBLE PRÊT ATT	05 Janvier. 2009	05_CONSO PRET AUTO	135 / 5089	2,65%
EPARGNE carre mauve2	05 Février. 2010	NBPEL + 02_PREDISSIME9	206 / 9958	2,07%
FLORIANE	05 Janvier. 2009	02_FLORIANE	47 / 2270	2,07%
MICRO CIBLE PRÊT ATT	22 Janvier. 2009	05_AUTRES PRETS CONSO + 05_CONSO PRET PROJET + 05_DIVERS PRETS	227 / 11626	1,95%
IARD_habitation	Octobre. 2009	NBMRH	129 / 7896	1,63%
PREVOYANCE_OBS	30 Avril. 2009	NBGO	131 / 8913	1,47%
IARD_auto	Octobre. 2009	NBAUTO	105 / 11363	0,92%
PREVOYANCE_comp_santé	30 Avril. 2009	07_CPLT SANTE	24 / 3256	0,74%
PREVOYANCE_GAV_OBS	30 Avril. 2009	NBGAV ou NBGO	64 / 8976	0,71%
EPARGNE A RENTES	05 Février. 2010	NBCVCB + 02_PREDISSIME9	50 / 9026	0,55%
EPARGNE carre bleu	05 Février. 2010	NBCVCB + 02_PREDISSIME9	63 / 11503	0,55%
JEUNES	Septembre. 2009	15_CARTE MOZAIC	67 / 17120	0,39%
VERS L'AVENIR	13 Février. 2009	02_VERS LAVENIR	28 / 17401	0,16%
EPARGNE FLORIANE	05 Février. 2010	02_ESPACE LIBERTE + 02_FLORIANE	0 / 481	0,00%
TOTAL (nombre de cc part)	/	/	1756 / 139 571	1,26%

Ensuite, le choix du périmètre de l'étude a été déterminé. Cette étape est primordiale afin de cadrer la mission. On a donc choisi d'étudier les clients avant le premier mailing envoyé (décembre 2008) afin de se concentrer sur leur signalétique et leurs équipements initiaux. De plus, le choix de décembre 2008 évite tout risque de saisonnalité. Enfin, on a décidé de travailler uniquement sur les clients ayant reçu au moins un courrier commercial (voir figure 6).

Figure 6

mailing reçu offre souscrite	0	1	2	3	4	5	6	Total
0	391 682	73 316	23 816	4 894 **	649	54	3	494 414
1	/	1 029	347	138	23	2	1	1 540
2	/	/	32 *	9	2	0	0	43
Total	391 682	74 345	24 195 *	5 041 **	674	56	4	495 997

- Clients considérés appétents : 1 580 cc
- Clients considérés non appétents : 102 734
- Client non pris en compte : 391 682 cc



Exemples:

- * Parmi les 24 195 CC qui ont reçu deux courriers, 32 ont souscrit aux deux
- ** Parmi les 5041 CC qui ont reçu trois courriers, 4894 n'ont souscrit à aucune offre

On affecte donc, aux 104 314 CC ayant reçu au moins un courrier, une variables binaire "Appétence" 0 ou 1 et l'on va à présent chercher des cœurs de cible appétents ou non au mailing.

IV.3. Variables prises en compte dans l'étude

Une fois le périmètre de l'étude défini, il faut choisir les variables que l'on va prendre en compte pour déterminer l'appétence. Ce choix s'est déroulé en 3 étapes. La première était de demander à "l'expérience métier" de sélectionner les variables de bases, utiles pour l'explication de l'appétence. Ainsi, les 39 variables suivantes ont été sélectionnées:

- ETAPE_DE_VIE1
- SEGMENTATION1
- TR_AGE_MAX
- SOC
- TRPCS
- PORT
- CCVUS
- TRVENT
- NBPERSCC
- TRACTI
- NBDAV
- TRSVM
- TRMOYFLUX
- MTFACTU
- NBLIV
- CARTE
- NBCONSO
- NBHAB
- CSCA
- TRPDTS
- NBTIT
- TYP_ES
- PARTIE_PRO
- ELUS
- EMPLOYES
- NB_BAMEUR
- NBOPD
- NBOPC
- PNA_1208
- SCORE_FINAL
- UNIVERS_POND
- EPARGNE_MT
- CREDITS_MT
- ASSURANCES_NB
- SERVICES_NB
- TRANC
- NBPREV
- NBIARD

On ne va bien entendu pas faire l'étude sur ces 39 variables. En effet, certaines sont redondantes et d'autres peuvent être des sous familles de certaines (par exemple, SERVICES_NB est issue des variables CARTE et CSCA). On choisit alors d'étudier les corrélations de ces variables (*Annexe 6*: tableau des corrélations des 39 variables présélectionnées). Ainsi, cela permet d'éliminer les variables qui représentent un trop grand risque de colinéarité avec d'autres et de sélectionner la famille ou le sous ensemble de la famille du produit. Voici les principaux résultats:

Corrélation moyenne $0.5 < \text{corr} < 0.7$		
• <u>SEGMENTATION1</u> - PORT : 0.58 - TRANC2 : 0.56 • <u>TR_AGE_MAX</u> : - TRANC2 : 0.60 - NB_TIT : 0.54 - EPARGNE_MT : 0.53 - SOC : 0.50 • <u>TRANC2</u> : - TR_AGE_MAX : 0.60 • <u>MTFACTU</u> : - PNA_1208 : 0.56 - TRPDTS : 0.50 • <u>NBLIV</u> : - TRUNIV : 0.53 - NB_TIT : 0.53 - EPARGNE_MT : 0.56 • <u>NB_TIT</u> : - SOC : 0.68 - EPARGNE_MT : 0.68 - TRPDTS : 0.60 - PNA_1208 : 0.59 - TR_AGE_MAX : 0.54 - NBLIV : 0.53	• <u>PNA_1208</u> : - EPARGNE_MT : 0.68 - TRSVM : 0.63 - NBTIT : 0.59 - TRPDTS : 0.59 - MTFACU : 0.56 • <u>EPARGNE_MT</u> : - NBTIT : 0.68 - NBOPD : 0.64 - TRSVM : 0.565 - TR_AGE_MAX : 0.57 - SOC : 0.50 - NBLIV : 0.56 - TRMOYFLUX : 0.52 - TRPDTS : 0.51 - PORT : 0.55 • <u>NBOPD</u> : - TRACTI : 0.51 • <u>CREDIT MT</u> : - NBHAB : 0.58 - NBCONSO : 0.23	• <u>SERVICE_NB</u> : - TRPDTS : 0.63 - CARTE : 0.61 - CSCA : 0.57 - TRUNIV : 0.53 • <u>NBOPD</u> : - NBOPC : 0.65 - TRMOYFLUX : 0.55 - TRACTI : 0.54
Corrélation forte $0.7 > \text{corr}$		
		• <u>TR_AGE_MAX</u> : - ETAPE_DE_VIE1 : 0.86 • <u>TRPDTS</u> : - TRUNIV : 0.72 - NBLIV : 0.70 • <u>ASSURANCES_NB</u> : - NBIARD : 0.76 - NBPREV : 0.71
Variables éliminées : -trop grand risque de colinéarité avec d'autres variables -sélection de la famille ou du sous ensemble de la famille de produit		

Cette seconde étape a donc permis de supprimer 7 variables. Il ne reste donc, après les corrélations, que 32 variables.

La troisième et dernière étape consiste à effectuer une régression. Après avoir réalisé une régression linéaire selon les trois méthodes Backward, Forward et Stepwise, mon tuteur m'a préconisé de privilégier la régression logistique car la variable cible est binaire. On effectue donc cette régression logistique afin de sélectionner les variables qui caractérisent le plus la différence entre les appétents et les non appétents. Celle-ci est effectuée selon les 3 méthodes Backward, Forward et Stepwise. Elle est, de plus, réalisée en fonction de trois p-value différentes (Alpha = 0.01, 0.05, 0.10) afin de se laisser une liberté de choix du nombre de variables à sélectionner (*Annexe 7*: tableaux du choix des variables issues des différentes régression logistique). En accord avec la maitre de stage, on prend Alpha = 0.05 et choisit un modèle à 13 variables:

1. Même Résultat selon les 3 méthodes (Stepwise, Backward, Forward)
2. Choix de la p-value:
 - Alpha = 0.01 => 11 variables choisies
 - Alpha = 0.05 => 13 variables choisies
 - Alpha = 0.10 => 13 variables choisies
3. Choix effectué: Alpha = 0.05 => **13 variables retenues** (de la plus influente à la moins influente) :

• 1. TRPDTS	6. NBIARD	11. TRMOYFLUX
• 2. NBPERSCC	7. EPARGNE_MT	12. NBAV
• 3. TRVENT	8. SERVICES_NB	13. TRANC
• 4. CREDITS_MT	9. TRPCS	
• 5. TRSVM	10. NBOPD	

En résumé, on réalise l'étude sur une population de 104 315 CC (1580 appétents et 102 734 non appétents) et 13 variables issues de la base DIAGO de décembre 2008 (TRPDTS – NBPERSCC – TRVENT – CREDITS_MT – TRSVM – NBIARD – EPARGNE_MT – SERVICES_NB – TRPCS – NBOPD – TRMOYFLUX – NBDV – TRANC).

V. Analyse descriptives et prédictives

V.1. Une première approche statistique: l'analyse univariée

En premier lieu, nous avons exécuté une analyse univariée afin d'avoir les premières orientations, les premières tendances des clients appétents et ceux non appétents. Cette étude a été réalisée sur les variables de base de la connaissance clients, c'est-à-dire sur la signalétique générale (âge, fidélité, équipement, activité ...). Les résultats complets de cette étude sont disponibles en annexe (*Annexe 8: résultats complets de l'étude univariée*).

En résumé, ces analyses ont permis de détecter deux cœurs de cible:

- Un cœur de cible appétent :

- ➔ Plus les clients sont **actifs**, plus ils sont **appétents** :
 - TRACTI : 10+
 - TRVENT : 5-9 ; 10+
 - NBOPD : >694
- ➔ Plus les clients sont **équipés**, plus ils sont **appétents**:
 - TRPDTS : 15+
 - TRUNIV : 60-70 ; 70-80 ; 80-90 ; 90-100
- ➔ Plus les cc sont **âgés**, plus ils sont **appétents** :
 - ETAPE_DE_VIE1 : RETRAITES ; SENIORS
 - TR_AGE_MAX : 56-65 ; 65-75 ; 75+
 - TRPCS : RETRAITE ; AGRI
- ➔ Les clients dont le Crédit Agricole est la **banque principale** sont plus **appétents** :
 - SCORE_FINAL : BP
- ➔ Les cc de **personne unique** sont plus **appétentes** :
 - NBPERSCC : 1
- ➔ Les cc qui ont une **tranche de flux moyenne** sont plus **appétentes** :
 - TR_MOY_FLUX : 1500-3000

- ➔ Plus les clients sont **actifs**, plus ils sont **appétents** :
 - opérations mensuelles supérieures à 10
 - ventes supérieures à 5
 - opérations débitrices mensuelles supérieures à 58
- ➔ Plus les clients sont **équipés**, plus ils sont **appétents**:
 - au moins 15 produits
 - % d'univers de besoin supérieur à 60%
- ➔ Plus les cc sont **âgés**, plus ils sont **appétents** :
 - retraités et seniors
 - personne la plus âgée du CC supérieure à 55 ans
 - catégorie socioprofessionnelle: retraités - agriculteurs
- ➔ Les clients **fidèles** (> 20 ans) sont plus **appétents**
- ➔ Les clients dont le Crédit Agricole est la **banque principale** sont plus **appétents**
- ➔ Les cc de **personne unique** sont plus **appétentes**

- Un cœur de cible non appétent :

- Plus les clients sont **inactifs**, plus ils ne sont **pas appétents** :
 - TRMOYFLUX : 0
 - TRUNIV : 0 ; 0-10 ; 10-20 ; 20-30
 - NBOPD : < 94
- **Moins** les clients sont **équipés**, plus ils ne sont **pas appétents** :
 - TRPDTS : <3 ; 3-5
 - SCORE_FINAL : PA
- Plus les cc sont **jeunes**, plus ils ne sont **pas appétents** :
 - ETAPE_DE_VIE1 : MINEURS
 - SEGMENTATION1 : 0-17 ans
- Les cc **exclus** généralement des envois ne sont **pas appétents** :
 - TYPE_ES : AUTRE ; BANQUE PRIVEE ; ENTREPRISE
 - EXCLUSION : 0
 - COULEUR : R

- Plus les clients sont **inactifs**, plus ils ne sont **pas appétents** :
 - flux moyen nul
 - taux d'univers de besoin inférieur à 30 %
 - opérations débitrices mensuelle inférieure à 8
- **Moins** les clients sont **équipés**, plus ils ne sont **pas appétents** :
 - Moins de 5 produits
- Plus les cc sont **jeunes**, plus ils ne sont **pas appétents** :
 - mineurs
 - étudiants
- Les cc **exclus** généralement des envois ne sont **pas appétents** :
 - canaux : autre – banque privée – entreprise
 - exclus
 - risque EMACO rouge

On peut remarquer que les clients exclus généralement (voir section IV.1) ne sont pas appétents. Cette étude vient donc conforter l'idée que ces exclusions avant ciblage sont très judicieuses.

Ces premiers résultats correspondent aux attentes des maitres de stage. En effet, il y a peu de surprises et cela confirme plutôt leurs idées et leurs intuitions d'avant mission. L'individu appétent est plutôt âgé, fidèle, actif et équipé; la notion de "Banque Principale" regroupe ces qualificatifs. A l'inverse, l'individu non appétent est jeune, peu actif et peu équipé; la notion de "Peu Actif" regroupe ces qualificatifs.

Afin de renforcer ces premiers résultats, on cherche à connaître le profil des appétents à l'aide d'une ACM et d'une classification.

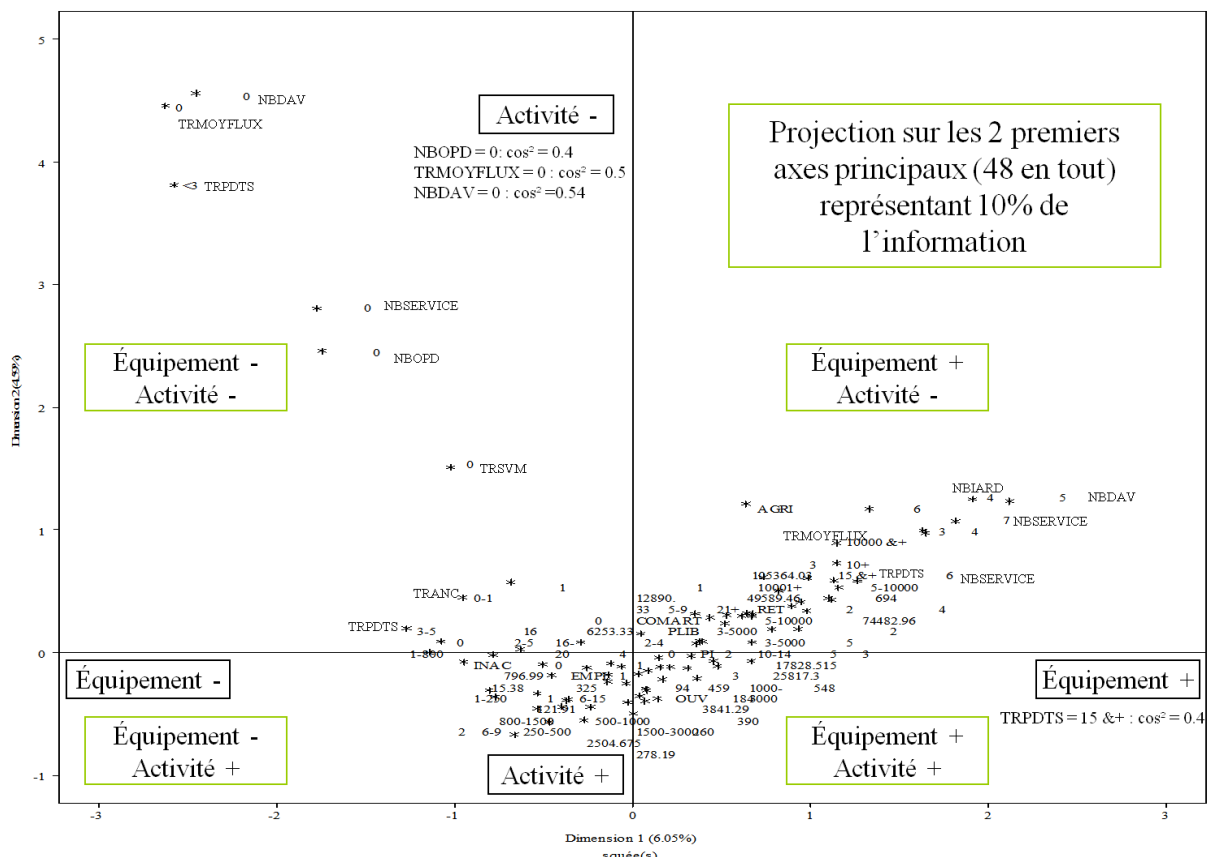
V.2. Typologie des clients appétents

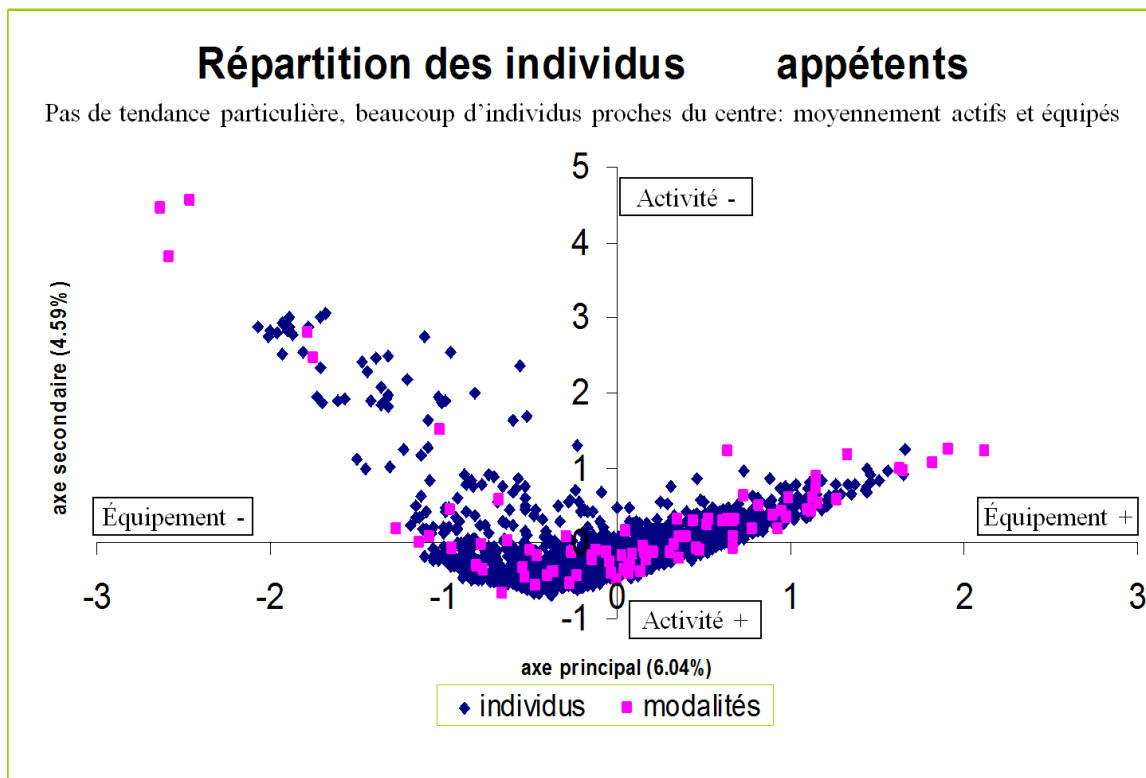
Le but de typer les clients appétents est de savoir qu'elles sont leurs caractéristiques afin de les choisir prioritairement pour les prochaines campagnes de mailing. Pour y parvenir, on réalise une Analyse des Correspondances Multiples (ACM) sur les 13 variables retenues précédemment. Ensuite, on projette une vue 2D des deux premiers axes puis on réalise une classification de ces 1 580 appétents grâce à la méthode de la Classification Ascendante Hiérarchique (CAH).

La première étape consiste à passer le tableau de données en tableau disjonctif complet. Cependant, on s'est rendu compte que certaines variables quantitatives possédaient de nombreuses valeurs presque toutes différentes; cela rendait donc le tableau disjonctif inutilisable. De fait, on a décidé de modifier les données des variables quantitatives en décile afin de réduire considérablement et judicieusement le tableau disjonctif. Après ces modifications, on a pu exécuter l'ACM.

Le tableau de la répartition de l'inertie (*Annexe 9: tableau de répartition de l'inertie*) permet de voir que le "cas d'école", celui où deux ou trois axes suffiraient à restituer l'essentiel de la répartition initiale du nuage de point, n'est pas envisageable ici. Cependant, on remarque que 48 axes restituent 80 % de la population initiale.

Bien que les deux premiers axes ne donnent que 10 % de l'information initiale, on projette sur ceux-ci, les modalités puis les individus:



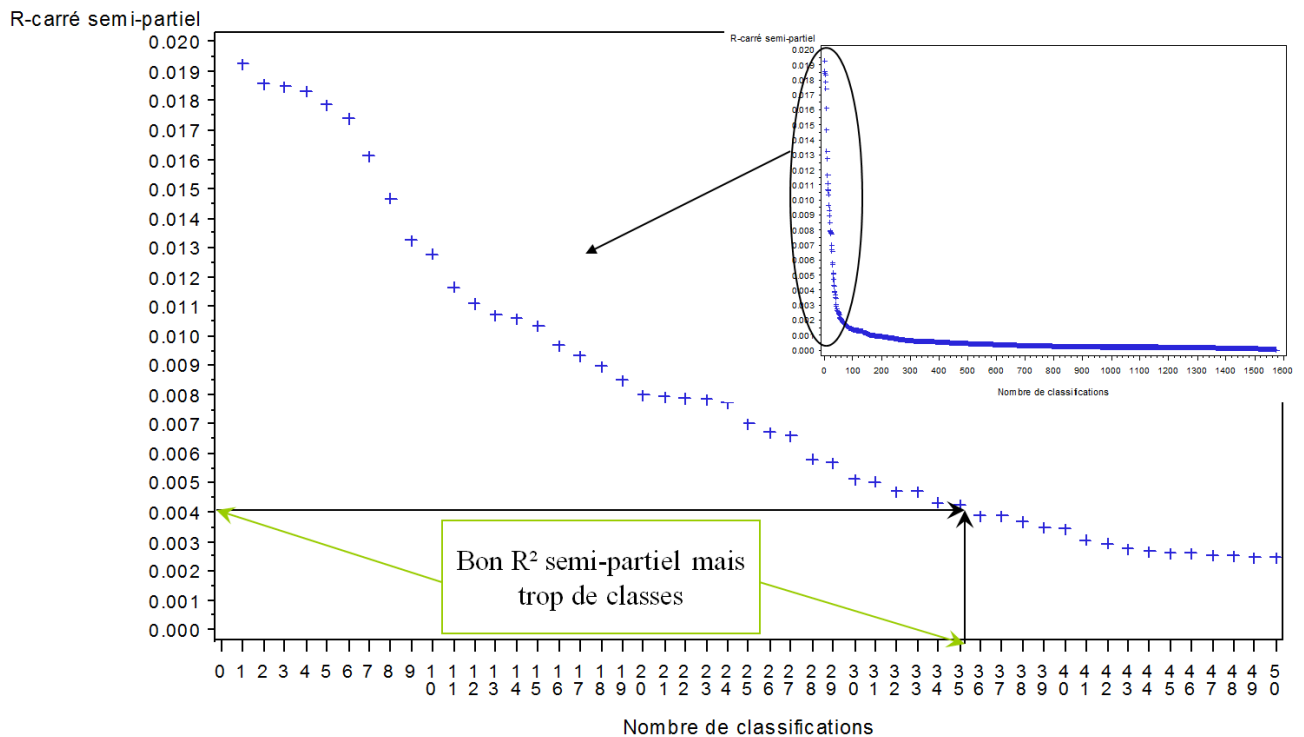


L'étude montre que ces deux composantes principales résument d'une part l'équipement et d'autre part l'activité du client. On ne peut cependant rien conclure en voyant ces graphiques car la dispersion n'est pas réellement celle-là du fait qu'il n'y est que 10% de résumé ici.

On décide alors d'essayer de classer ces appétents en un nombre réduit de classes afin de pouvoir extraire des typologies de clients appétents.

On réalise la classification à l'aide de la méthode des CAH. Celle-ci consiste à organiser les observations en les regroupant de façon hiérarchique. Elle commence par agréger celles qui sont les plus semblables entre elles, puis les observations ou groupes d'observations un peu moins semblables et ainsi de suite jusqu'au regroupement trivial de l'ensemble de l'échantillon. Ces agrégations se font deux à deux. Sa complexité $O(n)^2$ ne gêne absolument pas car, comme nous étudions seulement les appétents, il y a 1580 individus donc le temps d'exécution de cette méthode n'est pas trop long.

Les résultats de cette étape n'apportent malheureusement pas de résultats opérationnels, que la méthode soit lancée à partir du tableau disjonctif ou bien des coordonnées des 48 axes issus de l'ACM. En effet, comme le montre le graphique des CCC qui suit, il faut classer les 1580 appétents en une trentaine de classes pour avoir un résultat satisfaisant.



L'ACM et la classification des appétents n'ayant donné aucun résultat opérationnel, on se propose d'effectuer une analyse prédictive afin d'optimiser l'appétence mailing.

V.3. Analyse discriminante linéaire

À partir de l'historique des informations disponibles sur les clients, l'analyse prédictive détermine, par une analyse statistique des relations entre les données disponibles, si elles sont de nature à prédire, avec la meilleure fiabilité possible, le futur comportement d'un client (*Annexe 10* : principe de l'analyse prédictive).

La méthode d'analyse discriminante linéaire s'est avérée un premier choix naturel car c'est celle que nous avons étudié durant l'année étudiante. Cependant, mon tuteur m'avait prévenu de la grande difficulté à prédire des phénomènes rares comme c'est le cas ici : les 1580 appétents représentent 1,5 % de la population étudiée (104 315 clients).

Les résultats confortent ce qu'avait pressenti M Bousquet. En effet, comme le montre la table d'erreur suivante, la phase de test (20 % du fichier total), réalisée plusieurs fois, a toujours considéré tous les individus comme non appétents, sans distinction.

Table d'erreur de l'ADL		
APPETENCE (APPETENCE réelle)	_INTO_ (APPETENCE estimée)	
Fréquence Pourcentage Pctage en ligne Pctage en col.	0	Total
0	20720 98.52 100.00 98.52	20720 98.52
1	311 1.48 100.00 1.48	311 1.48
Total	21031 100.00	21031 100.00

Comme cette méthode n'a rien donné, on choisit d'exécuter l'analyse discriminante quadratique en espérant de meilleurs résultats.

V.4. Analyse discriminante quadratique

L'analyse discriminante quadratique est très utilisée en recherche car elle se comporte très bien, en termes de performances malgré qu'elle soit moins répandue auprès des praticiens par rapport aux autres méthodes. Le point négatif de cette méthode est que l'interprétation des résultats est assez périlleuse car il est notamment difficile de distinguer les variables réellement déterminantes.

Table d'erreur de l'ADQ			
APPETENCE (APPETENCE réelle)	_INTO_ (APPETENCE estimée)		
Fréquence Pourcentage Pctage en ligne Pctage en col.	0	1	Total
0	325 1.55 1.57 99.39	20395 96.98 98.43 98.51	20720 98.52
1	2 0.01 0.64 0.61	309 1.47 99.36 1.49	311 1.48
Total	327 1.55	20704 98.45	21031 100.00

Le lancement de cette méthode ne donne aucun résultat opérationnel non plus. Comme précédemment, la phase de test (réalisée plusieurs fois) rend compte d'une mauvaise appréciation lors de l'apprentissage du modèle:

Comme cette méthode n'a rien donné, on choisit d'exécuter la régression logistique.

V.5. Régression logistique

La régression logistique est une technique statistique qui a pour objectif, à partir d'un fichier d'observations, de produire un modèle permettant de prédire les valeurs prises par une variable le plus souvent binaire, à partir d'une série de variables explicatives continues et/ou binaires.

Une nouvelle fois, les résultats confirment que la prédiction de phénomènes rares est très difficile. En effet, de même que l'analyse discriminante linéaire, la table d'erreur montre que la phase de test (20 % du fichier total) a toujours considéré tous les individus comme non appétents, sans distinction.

Table d'erreur de la régression logistique		
APPETENCE (APPETENCE réelle)	I_APPETENCE (Vers : APPETENCE estimée)	
Fréquence Pourcentage Pctage en ligne Pctage en col.	0	Total
0	20373 98.40 100.00 98.40	20373 98.40
1	331 1.60 100.00 1.60	331 1.60
Total	20704 100.00	20704 100.00

A ce stade de la mission, malgré la diversité des méthodes statistiques utilisées, on a toujours aucun résultat opérationnel mis à part l'analyse univariée du début. On demande donc au tuteur de stage une solution. Celle qui est envisagée est le mapping.

VI. Mapping

VI.1. Principe

Le principe du mapping est en soit intuitif. Cette méthode se déroule en trois ou quatre étapes:

- 1) Classifier la population globale de l'étude (104 315 cc) sans tenir compte de la variable cible. Cela permet de regrouper les individus semblables dans le but de trouver des caractéristiques propres à chaque classe.
- 2) Déterminer le taux de pénétration d'appétents pour chaque classe. Pour y parvenir, on calcule la fréquence d'appétents pour chacune de ces classes. Le but de cette étape est de regarder quelles sont les classes où il y a une part importante d'appétents et celles, à l'inverse, où la population d'appétents est quasi nulle.
Par exemple, si la classification d'une population donne deux groupes d'individus et que l'un possède 90% des appétents alors les caractéristiques de cette classe sont des motifs d'appétence.
- 3) établir les caractéristiques de chaque classe (faire une typologie) afin de conclure sur les préconisations à faire.
- 4) Projeter les groupes d'individus et leur taux de pénétrations sur les deux premiers axes de l'ACM de la population globale (*Annexe 11* : exemple de mapping). Cette dernière étape n'est réalisable que si les classes sont compactes et éloignées les unes des autres lorsqu'on les projette sur ces axes. Ce ne sera pas le cas pour cette étude parce que les 2 premiers axes de l'ACM ne résumeront que trop peu d'informations (9,12 %).

VI.2. Classification et taux de pénétrations

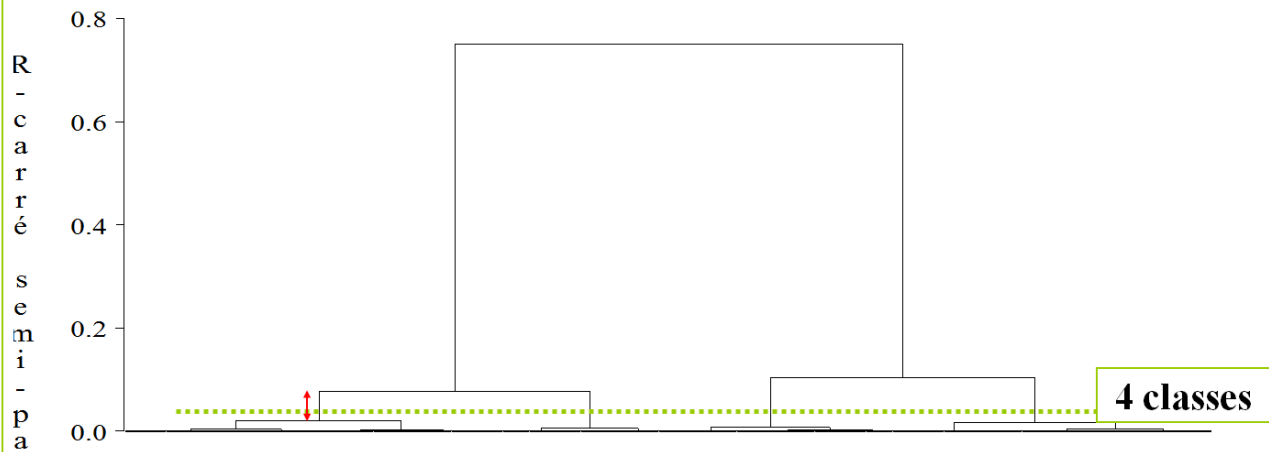
La population globale de l'étude est trop importante pour appliquer la méthode des CAH car sa complexité $O(n)^2$ rendrait le temps d'exécution beaucoup trop long. C'est pourquoi on a du appliquer la méthode de la classification mixte.

La classification mixte consiste à faire une première classification de tout notre nuage de points par la méthode des k-means, puis de procéder à une classification CAH sur les centres de classes précédemment obtenues par les k-means. Enfin, il ne reste plus qu'à réaffecter chaque individu à sa classe.

On décide de prendre un nombre de classe fixé à 1000 (option maxc=1000) car ce nombre de classes doit être significativement plus grand que le nombre de classes attendu au final. Pour plus de robustesse, on décide de supprimer les classes composées de seulement deux individus.

989 centres de classes ont donc été trouvés. Sur ceux-ci, on applique la méthode des CAH qui donne de bons résultats. En effet comme le montre le dendrogramme qui suit, ou l'annexe 12 (Annexe 12 : courbes SPRSQ et RSQ), la population peut être divisée en quatre classes :

- **Résultat** : on détermine le nombre de classe à partir du grand saut $\updownarrow$, soit 4 classes



A partir de ces 4 classes, on trouve pour chacune d'elle le taux de pénétration des appétents :

Fréquence d'appétence par classe			
CLUSTER	APPETENCE(APPETENCE)		
Fréquence Pourcentage Pctage en col.	0	1	Total
1	31365 30.07 30.53	461 0.44 29.18%	31826 30.51
2	21676 20.78 21.10	353 0.34 22.34%	22029 21.12
3	18963 18.18 18.46	258 0.25 16.33%	19221 18.43
4	30731 29.46 29.91	508 0.49 32.15%	31239 29.95
Total	102735 98.49	1580 1.51	104315 100.00

On se rend compte, d'après ce tableau, que la fréquence d'appétents est plus importante pour la classe 1 et 4. A l'inverse, on peut noter une sous population d'appétents pour la classe numéro 3. On se propose maintenant de donner les caractéristiques de chacune des classes.

VI.3. Résultats

On a pu déterminer les caractéristiques de chaque classe en détechant les sureffectifs et les sous-effectifs. Ces éléments ont été récapitulés dans un tableau en annexe (*Annexe 13* : tableau de typologie).

A l'aide de ce tableau, on remarque que l'on peut différencier les caractéristiques intéressantes selon quatre items:

Marqueurs de l'appétence pour les deux classes où la population d'appétents est la plus forte :	TRPCS : RET NBOPD : >694 TRANC : 20 +
Marqueurs de l'appétence pour l'une des deux classes où la population d'appétents est la plus forte :	TRPDTS : 15 &+ TRSVM : 10001 +
Marqueur de l'appétence divergeant pour les deux classes où la population d'appétents est la plus forte :	EPARGNE_MT : <2500 / >17500
Marqueurs de la non appétence :	TRPDTS : <3 TRVENT : 0 TRSVM : 0 SERVICES_NB : 0 TRMOYFLUX : 0 NBDAV : 0 NBOPD : <15

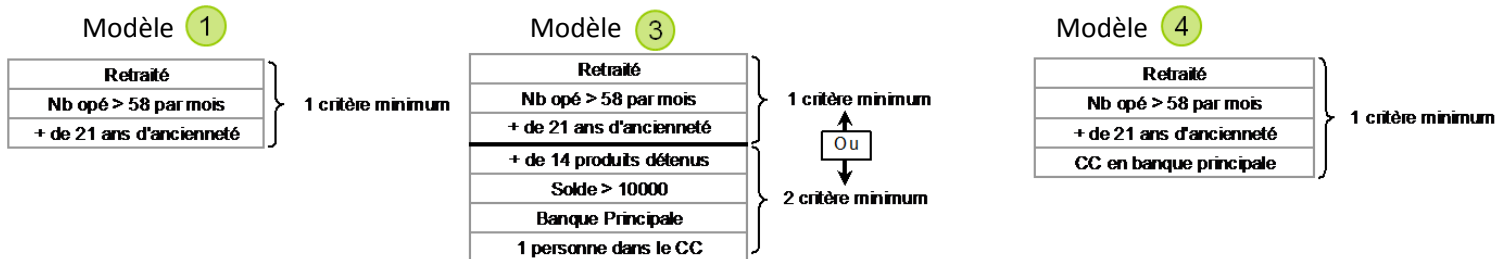
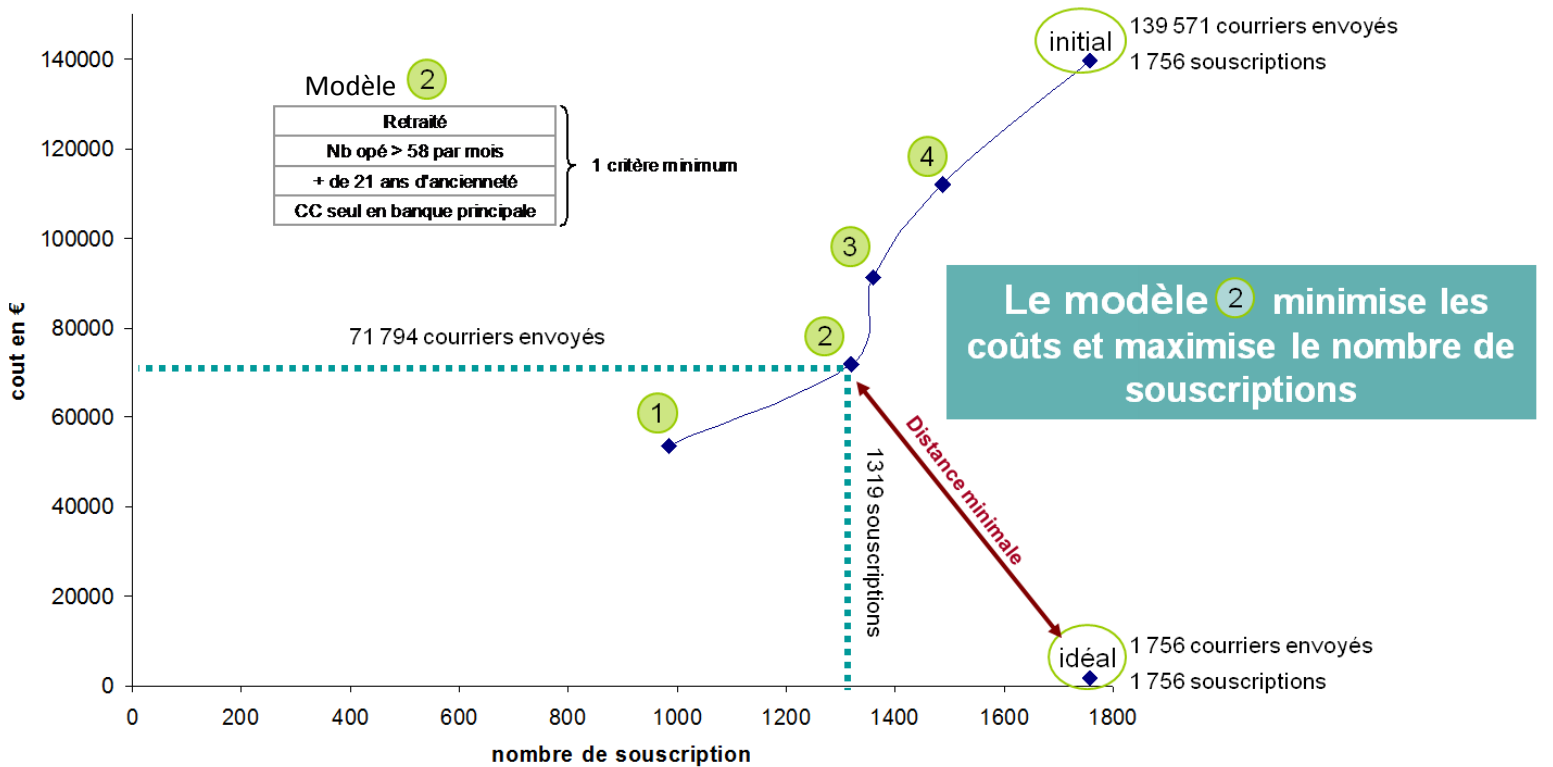
Marqueurs de l'appétence pour les deux classes où la population d'appétents est la plus forte (61,33 % des appétents) :	- catégorie socioprofessionnelle: retraités - opérations débitrices mensuelles supérieures à 58 -Les clients fidèles d'au moins 21ans
Marqueurs de l'appétence pour l'une des deux classes où la population d'appétents est la plus forte (29,18% des appétents) :	-au moins 15 produits -Solde en valeur moyen supérieur à 10 000
Marqueur de l'appétence divergeant pour les deux classes où la population d'appétents est la plus forte :	-montant d'épargne inférieur à 2500 ou supérieur à 17 500
Marqueurs de la non appétence :	-moins de trois produits -aucune vente -solde en valeur moyen journalier nul -aucun produit de type service -tranche moyenne de flux nulle -aucun DAV -nombre d'opération débitrice mensuelle inférieure à 15

Nous avons alors en possession deux types de préconisations : celles issues de l'analyse univariée et celles issues du mapping. Le but est donc maintenant de croiser ses préconisations entre elles afin de trouver un modèle satisfaisant. Cela permettra de créer un indicateur d'appétence mailing.

VII. Préconisations finales

VII.1. Choix du modèle

Maintenant que l'on possède des préconisations issues du mapping et de l'analyse univariée, on va les croiser afin de trouver le modèle qui minimise les coûts et maximise le nombre de souscriptions. Voici un graphique qui synthétise les résultats :



Le 1^{er} modèle à regarder sur le graphique est : "initial". En effet, durant l'année 2009 et le premier trimestre 2010, 139 571 courriers ont été envoyés pour 1756 souscriptions. L'idéal aurait été de n'envoyer que 1756 courriers pour autant de réponses positives (100% de retour). On cherche donc à se rapprocher le plus du modèle : "idéal".

On choisit, avec Marie Laure PAUTU et René Pierre DABON, quatre modèles à étudier. Pour les quatre modèles (1 2 3 4), on prend les marqueurs de l'appétence qui se retrouvent dans les préconisations à la fois du mapping et de l'analyse univariée. Ensuite, on complète selon les modèles par des préconisations uniquement présentes soit dans le mapping, soit d'ans l'analyse univariée.

Ainsi, on remarque que le modèle **2** est celui qui se rapproche le plus du modèle idéal. C'est donc le modèle que l'on applique : **tout client ayant au moins un des critères suivants sera considéré comme appétents** :

- **TRPCS = "RET"** (catégories socio professionnelles : retraité)
- **NBOPD >694** (nombre d'opérations débitrices annuelles > 698 / nombre d'opérations débitrices mensuelles >58)
- **TRANC = "21+"** (ancienneté d'au moins 21 ans)
- **NBPERSCC = 1 + SCORE_FINAL="BP"** (CC seul en banque principale)

Il faut à présent tester la fiabilité de ce modèle. Pour cela, on reprend l'efficacité initiale des 16 campagnes, on calcule leur efficacité par rapport à l'application des préconisations ci-dessus, et on compare.

VII.2. Phase de test

Cette étape permet de tester la robustesse du modèle. On a repris pour chaque campagne, grâce à l'historisation des requêtes ACCESS, la volumétrie d'envoi, le nombre de souscription et l'efficacité. On a ensuite réalisé la même étude en ne sélectionnant que les clients appétents.

Voici le tableau de comparaison entre l'efficacité avant préconisation (modèle initial) et après préconisation (modèle **2**) :

THEME MAILING	EFFICACITE AVANT PRECO	EFFICACITE APRES PRECO	% EFFICACITE AVANT PRECO	% EFFICACITE APRES PRECO	ECART
PREVOYANCE_OBS	131 / 8913	126 / 4353	1,47 %	2,89 %	1,42 points
MICRO CIBLE PRÊT ATT	227 / 11626	191 / 5846	1,95 %	3,27 %	1,32 points
EPARGNE carre mauve2+P9	206 / 9958	146 / 4354	2,07 %	3,35 %	1,28 points
P9	303 / 8130	270 / 556	3,73 %	4,85 %	1,12 points
EPARGNE carre mauve1+P9	717 / 6563	691 / 2457	2,70 %	3,70 %	1,00 point
FLORIANE	47 / 2270	47 / 1775	2,07 %	2,65 %	0,58 point
MICRO CIBLE PRÊT ATT	135 / 5089	117 / 3848	2,65 %	3,04 %	0,39 point
IARD_auto	105 / 11363	69 / 5940	0,92 %	1,16 %	0,24 point
EPARGNE A RENTES + P9	50 / 9026	47 / 6115	0,55 %	0,77 %	0,22 point
PREVOYANCE_GAV_OBS	64 / 8976	37 / 4395	0,71 %	0,84 %	0,13 point
JEUNES	67 / 17120	38 / 7359	0,39 %	0,52 %	0,13 point
VERS L'AVENIR	28 / 17401	15 / 5654	0,16 %	0,27 %	0,11 point
PREVOYANCE_comp_santé	24 / 3256	16 / 2028	0,74 %	0,79 %	0,05 point
EPARGNE carre bleu + P9	63 / 11503	42 / 7177	0,55 %	0,59 %	0,04 point
EPARGNE FLORIANE	0 / 481	0 / 393	0,00 %	0,00 %	0,00 point
IARD_habitation	129 / 7896	67 / 4534	1,63 %	1,48 %	-0,15 point
TOTAL (nombre CC Part)	1756 / 139 571	1319 / 71 794	1,26 %	1,84 %	0,58 point



+ 46%

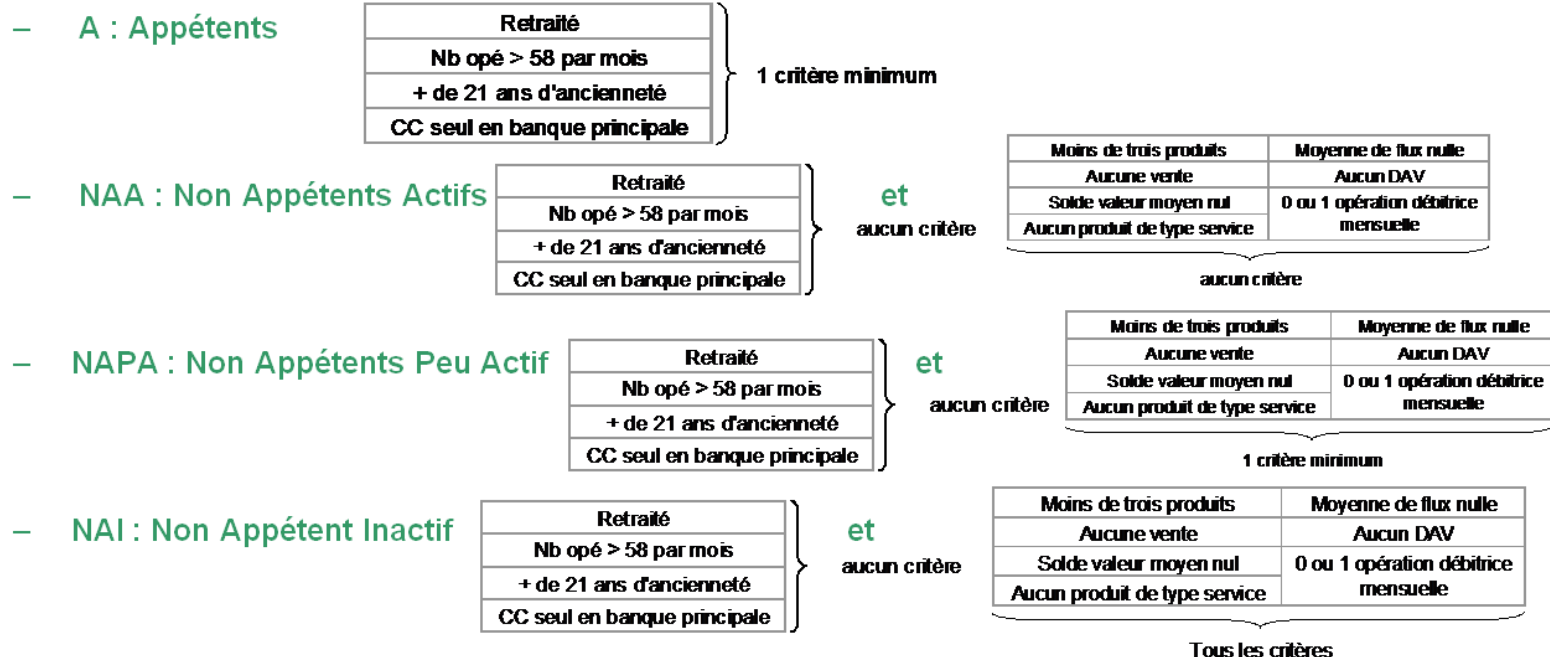
Les résultats s'avèrent réellement positifs. En effet, lors des campagnes de mailing de 2009 et du 1^{er} trimestre 2010, nous aurions pu :

- **Diminuer les charges de 49 %**, soit une économie de 54 222 €/an
 - o 67 777 € sur une période de 15 mois lorsque le courrier coûte 1 €
 - o Envoyer 71 794 courriers au lieu de 139 571
- **Gagner 46 % d'efficacité**

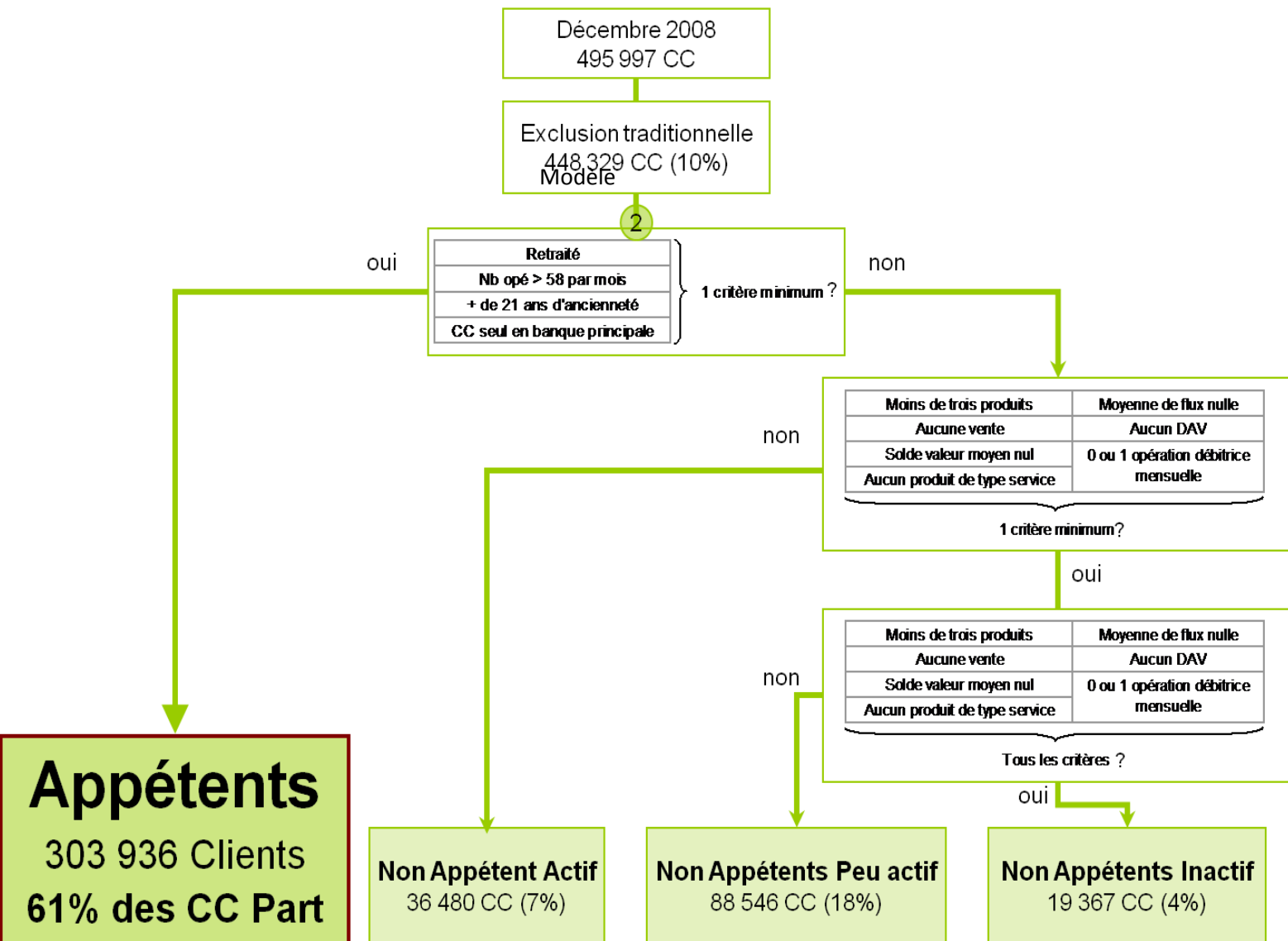
Suite à ces résultats convaincants, l'indicateur d'appétence mailing peut être créé.

VII.3. Mise en œuvre

Nous avons détecté les clients appétents. Cependant, dire que tous les autres clients ne sont pas appétents sans faire de distinction n'est pas totalement juste. C'est pourquoi on réalise une étude en croisant, de la même manière que pour les appétents, les deux types de préconisations (mapping et analyse univariée) afin de diviser les non appétents en 3 catégories. On en déduit alors, à la fin ce processus, une variable "TOP_APPC" comportant 4 modalités:



Ce qui donne, sur la base clients particuliers de décembre 2008, l'arbre de décision suivant:



Pour les prochaines campagnes de mailing, on suivra ces étapes:

- Ciblage traditionnel
- Sélection des appétents (A) en priorité
- Exclure les non appétents inactifs (NAI)
- Selon la volumétrie, cibler les non appétents actifs (NAA) puis les non appétents peu actifs (NAPA)

Nous sommes donc arrivés au terme de la mission avec un résultat nous permettant la création d'un indicateur d'appétence mailing.

On a ensuite présenté les résultats de cette mission à l'ensemble des personnes du service qui n'est pas forcément connaisseur de l'univers des statistiques. Vulgariser le langage statistique a été un très bon exercice pour moi, très constructif et enrichissant.

VIII. Conclusion

Cette mission, malgré les résultats des statistiques prédictives non opérationnels, a permis de créer un indicateur d'appétence mailing afin d'optimiser l'envoi de courriers. Ainsi, nous avons remplis les objectifs initiaux, à savoir:

Objectifs initiaux	Objectifs réalisés
réduire les coûts d'envoi d'au moins 30%	réduction de 49%
obtenir une efficacité au moins égale	gain de 46 % d'efficacité
appliquer systématiquement les résultats de la mission	application réalisée
permettre aux autres Caisses Régionales de profiter des résultats	création d'un document power point

Il serait à présent intéressant de créer les indicateurs pour les autres canaux (e mail, SMS, MMS ...) puis ensuite, de créer des indicateurs sur l'appétence au produit. Enfin le croisement de ces deux indicateurs optimiserait au mieux l'approche multi canal de la banque.

D'autre part, ce stage a été un énorme enrichissement personnel. Durant ces dix semaines, j'ai acquis de nouvelles connaissances telle que la régression logistique et je me suis perfectionné dans les outils de SAS et ACCESS. Cependant, j'ai découvert un environnement de travail à la fois studieux et convivial favorisant les échanges. De plus, J'ai pu appréhender le monde d'une grande entreprise, l'open space et découvrir la vie quotidienne d'un service pendant deux moi et demi.

Bibliographies

- Stéphane TUFFERY - Data mining et statistique décisionnelle - 3ème Edition
- Support de cours SAS de l'Université Lumière Lyon II
- ISEL Cours de logiciel SAS. - E. PREUD'HOMME

Sources internet

- <http://www.chefdentreprise.com/Chef-d-entreprise-Magazine/Article/Le-marketing-direct-l-arme-a-tout-faire-des-PME-20318-1.htm&t=Le-marketing-direct-l-arme-a-tout-faire-des-PME>
- http://www.lentreprise.com/3/3/1/non-le-mailing-postal-n-est-pas-mort_16196.html
- <http://www.e-marketing.fr/Article-A-La-Une/Delphine-Martelli-CSA-Le-courrier-papier-joue-de-ses-atouts-1306.htm>