Conception et développement d'un écosystème data bancaire moderne

De la collecte à la valorisation des données clients

Mémoire de Master 2 en alternance

Adam Taïeb

TUTRICE UNIVERSITAIRE
Bénédicte Herrmann

MAÎTRE D'APPRENTISSAGE
Jean-Yves Ballot

Juin 2025

Version publique révisée, novembre 2026 : corrigée et resserrée, elle diffère de la version soutenue en 2025.
Données personnelles et identifiants internes retirés.

Note sur cette version

> Ce document reprend le mémoire soutenu en 2025, relu un an plus tard, corrigé et resserré. Les corrections sont les suivantes.
>
> - **Méthode.** La description de la segmentation de potentiel a été alignée sur le code effectivement utilisé : strates d'âge de cinq ans, K-means de 2 à 6 groupes avec une taille minimale de groupe, pas d'analyse en composantes principales, règles de calcul du potentiel épargne et crédit, construction de la segmentation finale.
> - **Statistiques.** L'interprétation des indicateurs a été corrigée : une silhouette négative ne signale pas une structure « acceptable », un seuil de 1,5 MAD correspond à environ un écart-type et non deux, la relation inverse entre PNB et potentiel découle du calcul et ne valide pas la méthode, et la répartition des segments finaux ne suit pas une loi normale.
> - **Chiffres.** Les résultats par groupe ont été relus sur les sorties d'origine, reproduites en figures dans la version soutenue. Cela corrige plusieurs erreurs de recopie : dépense annuelle, ticket moyen, taille des profils remarquables, indices de représentation. Les chiffres qui ne pouvaient pas être vérifiés ont été retirés.
> - **Affirmations.** Les gains qui n'avaient pas été mesurés, comme le temps gagné par entretien, ont été retirés, ainsi que les références bibliographiques introuvables. Les références conservées sont citées dans le texte.
> - **Forme.** Les chapitres consacrés à la gestion de projet, aux compétences et au bilan ont été fusionnés, le glossaire a été réduit aux termes utiles, et le texte a été allégé.

Remerciements

Je remercie Bénédicte Herrmann, ma tutrice universitaire, pour son suivi tout au long de l'année et pour ses exigences de méthode.

Je remercie Jean-Yves Ballot, mon maître d'apprentissage, de m'avoir confié des projets à responsabilité au sein du Crédit Agricole Franche-Comté et de m'avoir accompagné pendant ces deux années.

Merci à l'équipe Data Web Dev et aux équipes du Marketing pour leur accueil et le temps passé à discuter et valider ces travaux, ainsi qu'aux conseillers qui ont testé Parcours 360° : leurs retours ont fait évoluer l'outil.

Enfin, merci à ma famille et à mes proches pour leur soutien pendant cette année d'alternance.

Sommaire

Introduction 4

1. Contexte et infrastructure 5
- 1.1 Le Crédit Agricole Franche-Comté et l'équipe Data Web Dev 5
- 1.2 Une infrastructure hybride 5
- 1.3 Problématique et contraintes 6

2. La segmentation de potentiel 7
- 2.1 Deux segmentations pour deux usages 7
- 2.2 Données et variables 7
- 2.3 Stratification par âge, puis K-means 8
- 2.4 Ce que disent les indicateurs de qualité 8
- 2.5 Le potentiel commercial 9
- 2.6 La segmentation finale 10

3. Des personas enrichis par les paiements 12
- 3.1 Des variables lisibles 12
- 3.2 Le choix de cinq groupes 12
- 3.3 Les paiements par carte comme révélateur 13
- 3.4 Ce que révèlent les dépenses 14
- 3.5 La lecture inverse : des comportements aux groupes 14
- 3.6 Synthèse : six personas 15
- 3.7 Limites, et pourquoi garder deux segmentations 15

4. Parcours 360°, l'application des conseillers 17
- 4.1 Commencer par un irritant quotidien 17
- 4.2 Une architecture dictée par l'existant 17
- 4.3 Les fonctionnalités 18
- 4.4 Déploiement et adoption 19
- 4.5 Les suites prévues 19

5. Bilan 20
- 5.1 Ce qui a été livré 20
- 5.2 La méthode de travail 20
- 5.3 Les limites 20
- 5.4 Ce que je ferais autrement 20
- 5.5 Perspectives et apport personnel 21

Glossaire 22

Bibliographie 23

Résumé 24

Introduction

Le Crédit Agricole Franche-Comté dispose de données riches sur ses clients : contrats, flux, interactions avec la banque, paiements par carte. Pourtant, ces données servent peu au quotidien. Elles sont dispersées entre plusieurs systèmes, exploitées surtout par une petite équipe d'analystes, et rarement accessibles aux conseillers qui rencontrent les clients.

Mon alternance de Master 2, au sein de l'équipe Data Web Dev de la direction Marketing, part de cette situation et d'une question : **comment transformer les données bancaires en outils d'aide à la décision, pour le service Marketing comme pour les conseillers ?**

J'y ai répondu par trois réalisations, qui structurent ce mémoire :

- une **segmentation de potentiel**, qui estime pour chaque client l'écart entre ce qu'il rapporte aujourd'hui à la banque et ce que rapportent des clients comparables, univers par univers (chapitre 2) ;
- une **segmentation en personas**, plus lisible, enrichie par les données de paiement par carte et destinée au Marketing (chapitre 3) ;
- **Parcours 360°**, une application web qui réunit sur une seule page ce qu'un conseiller doit savoir d'un client avant un rendez-vous (chapitre 4).

Le chapitre 1 présente le contexte et l'infrastructure, qui expliquent beaucoup des choix techniques. Le chapitre 5 fait le bilan : ce qui a été livré, les limites, et ce que je ferais autrement.

Un fil conduit l'ensemble : dans une banque régionale, la difficulté tient moins à la sophistication des modèles qu'à leur appropriation. Une segmentation exacte mais incomprise ne sert à rien ; une application techniquement simple mais utilisée tous les jours peut, à terme, porter des modèles plus avancés.

CHAPITRE 1

Contexte et infrastructure

1.1 Le Crédit Agricole Franche-Comté et l'équipe Data Web Dev

Le Crédit Agricole Franche-Comté est l'une des 39 caisses régionales du groupe Crédit Agricole. Il compte plus de 400 000 clients répartis sur le Doubs, le Jura, la Haute-Saône et le Territoire de Belfort. Comme les autres caisses, il s'appuie sur les services informatiques mutualisés de Crédit Agricole Technologies et Services (CA-TS).

L'équipe Data Web Dev compte huit personnes et dépend de la direction Marketing. Elle réunit des compétences de développement web, d'analyse de données et de data mining. Cette position hybride est son intérêt : assez proche des métiers pour comprendre leurs besoins réels, assez technique pour intégrer ses outils au système d'information existant et les maintenir dans la durée.

Le territoire a une particularité qui compte pour la suite : la proximité de la Suisse. Environ 30 000 clients sont frontaliers, ce qui en fait un groupe à part, au potentiel économique élevé, que toute segmentation locale doit savoir distinguer.

1.2 Une infrastructure hybride

L'infrastructure combine des systèmes historiques et une plateforme Big Data plus récente (figure 1.1).

- **Teradata** centralise les données de production dans deux environnements : le SID V1, copie brute des systèmes de production avec leur nomenclature technique, et le SID V2, restructuré avec une nomenclature métier et des modèles dimensionnels.
- **SAS Enterprise** porte les traitements analytiques historiques, connectés à Teradata par SAS/ACCESS.
- **SQL Server** alimente les applications web internes ; il récupère les données de Teradata par OpenQuery. Ces applications sont servies par Apache HTTP Server.
- La plateforme **Cloudera** (CDP) compte 21 nœuds de calcul répartis sur trois sites et trois nœuds maîtres en haute disponibilité. Hive y sert aux requêtes en traitement par lots, Spark (via PySpark) aux traitements distribués et au machine learning, Jupyter Lab au développement.

L'accès est sécurisé à plusieurs niveaux : authentification Kerberos par tickets, qui évite de faire circuler des mots de passe sur le réseau ; chiffrement TLS des échanges ; et stockage des secrets applicatifs (mots de passe, clés) dans des magasins JCEKS plutôt que dans le code.

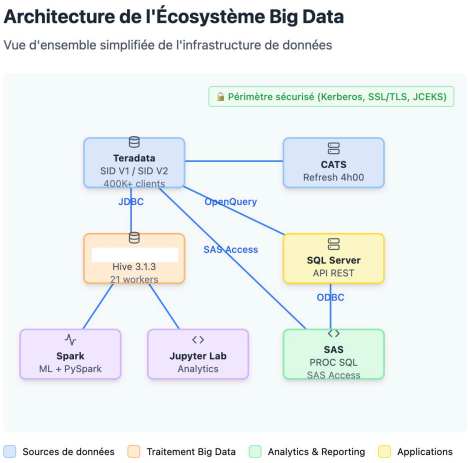

Figure 1.1 : Vue d'ensemble simplifiée de l'infrastructure de données. Le nom interne de la plateforme Big Data a été masqué.

1.3 Problématique et contraintes

La mission se résume à une question : **comment transformer les données bancaires en outils d'aide à la décision pour le service Marketing et pour les conseillers ?** Trois contraintes ont pesé sur toutes les réponses.

- **Le volume.** Plus de 400 000 clients et plus de 5 millions de contrats : les traitements doivent être distribués sur le cluster et conçus pour limiter les échanges de données entre nœuds.
- **L'existant.** Aucune refonte n'était envisageable. Chaque solution devait s'intégrer à Teradata, SQL Server et Apache tels qu'ils sont.
- **L'adoption.** Un outil n'a de valeur que s'il est utilisé. Les segmentations devaient être compréhensibles par le Marketing, et l'application utile aux conseillers dès le premier jour.

Le cadre réglementaire s'y ajoute : le RGPD impose que chaque traitement de données personnelles ait une finalité explicite et que les accès soient maîtrisés (Règlement (UE) 2016/679).

CHAPITRE 2

La segmentation de potentiel

Le Marketing voulait savoir où concentrer l'effort commercial : quels clients pourraient apporter davantage à la banque, et dans quel univers de produits. Ce chapitre présente la segmentation construite pour répondre à cette question.

2.1 Deux segmentations pour deux usages

Une segmentation unique devait servir deux besoins qui tirent en sens contraire. Pour estimer un potentiel commercial, il faut comparer chaque client à des clients très proches de lui : beaucoup de groupes fins, peu importe qu'ils soient faciles à décrire. Pour bâtir une stratégie de communication, il faut au contraire quelques groupes que l'on peut résumer en une phrase (Kumar et Reinartz, 2012).

J'ai donc proposé deux segmentations distinctes, chacune conçue pour son usage (figure 2.1) : une segmentation de potentiel pour le pilotage commercial, présentée dans ce chapitre, et une segmentation en personas pour le Marketing, présentée au chapitre 3.

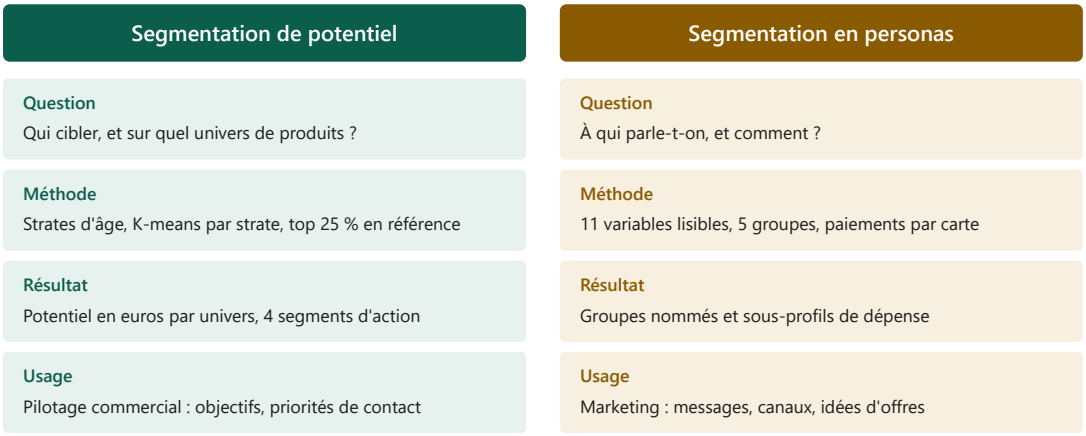

Figure 2.1 : Deux segmentations, chacune conçue pour un usage.

2.2 Données et variables

La segmentation porte sur les clients particuliers majeurs. Chaque client est décrit par quatre caractéristiques socio-démographiques, choisies parce qu'elles pèsent sur les besoins bancaires et restent faciles à interpréter :

- **la situation**, couple ou célibataire : un client est considéré en couple si son foyer compte au moins deux majeurs ou si sa situation familiale l'indique (marié, pacsé, concubin) ;
- **le niveau de revenu**, sur une échelle ordinale à quatre niveaux tirée de la catégorie socio-professionnelle : 0 pour les CSP modestes, 1 pour les CSP intermédiaires, 2 pour les CSP supérieures, et 3 pour les frontaliers, quelle que soit leur CSP ;
- **la présence d'enfants** dans le foyer ;
- **le statut d'occupation du logement**, propriétaire ou non.

Les variables binaires sont encodées en indicatrices, ce qui donne sept variables, puis toutes sont centrées et réduites avant le clustering. L'âge n'en fait pas partie : il sert à découper la population en strates.

2.3 Stratification par âge, puis K-means

Les besoins bancaires dépendent d'abord de l'étape de vie : un étudiant de 20 ans et un retraité de 70 ans ne sont pas comparables, quelles que soient leurs autres caractéristiques. La population est donc d'abord découpée en **strates d'âge de cinq ans** (18-22 ans, 23-27 ans, et ainsi de suite jusqu'à 58-62 ans, puis 63 ans et plus), soit dix strates. Une strate de moins de 10 000 clients n'est pas segmentée.

Dans chaque strate, un K-means (Hastie et al., 2009), implémenté avec scikit-learn (Pedregosa et al., 2011), regroupe les clients selon les sept variables. Le nombre de groupes est choisi automatiquement (listing 2.1) : on teste de 2 à 6 groupes, on écarte toute solution dont le plus petit groupe compte moins de 800 clients, et l'on retient celle dont le coefficient de silhouette est le plus élevé (Rousseeuw, 1987). Le seuil de 800 clients garantit que chaque groupe en contient assez pour servir de base de comparaison.

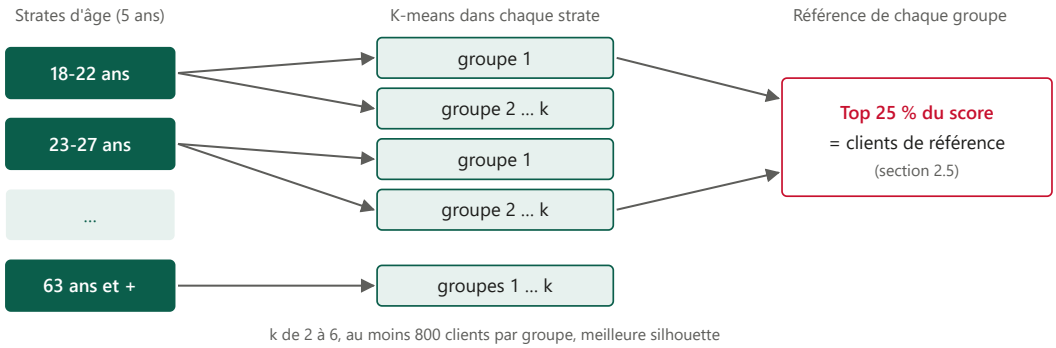

Figure 2.2 : La population est découpée en strates d'âge, puis chaque strate en groupes de clients comparables.

```python
def choisir_k(X, k_max=6, taille_min=800):
    """Retient le k dont la silhouette est la meilleure, parmi ceux
    dont chaque groupe compte au moins taille_min clients."""
    k_max = max(2, min(k_max, len(X) // taille_min))
    meilleur_k, meilleure_silhouette = None, -1
    for k in range(2, k_max + 1):
        labels = KMeans(n_clusters=k, n_init=10, random_state=42).fit_predict(X)
        if np.bincount(labels).min() < taille_min:
            continue                        # un groupe trop petit : solution écartée
        s = silhouette_score(X, labels)
        if s > meilleure_silhouette:
            meilleur_k, meilleure_silhouette = k, s
    return meilleur_k or 1              # aucune solution valide : un seul groupe
```

Listing 2.1 : Choix du nombre de groupes dans une strate (version simplifiée du code).

2.4 Ce que disent les indicateurs de qualité

Indicateur	Valeur	Lecture	Interprétation
Silhouette	−0,037	de −1 à 1, plus haut = mieux	Proche de zéro et négative : en moyenne, un client n'est pas plus proche de son groupe que du groupe voisin.
Davies-Bouldin	19,86	à partir de 0, plus bas = mieux	Très élevé : groupes peu compacts et peu séparés.
Calinski-Harabasz	437,8	plus haut = mieux	Ne s'interprète qu'en comparant plusieurs solutions entre elles ; seul, il ne prouve pas l'existence de groupes.

Tableau 2.1 : Indicateurs de qualité de la segmentation de potentiel (Rousseeuw, 1987 ; Davies et Bouldin, 1979 ; Caliński et Harabasz, 1974).

Ces indicateurs sont sans ambiguïté : les clients ne forment pas de groupes naturels, mais un continuum de situations. Le K-means découpe ce continuum ; il ne révèle pas des « types » de clients qui existeraient indépendamment de lui.

Ce constat ne remet pas en cause l'usage prévu, à condition de bien nommer ce que sont ces groupes. Ce ne sont pas des segments à décrire, mais des **groupes de comparaison** : des clients d'âge proche et de situation semblable, entre lesquels comparer les revenus a un sens. Pour cet usage, ce qui compte est la ressemblance des clients au sein d'un groupe et une taille suffisante, que garantissent la stratification et le seuil de 800 clients, bien plus que la séparation entre groupes. En revanche, ces groupes ne doivent pas être présentés au Marketing comme des segments de clientèle : c'est l'une des raisons d'être de la segmentation du chapitre 3.

2.5 Le potentiel commercial

Le potentiel d'un client est l'écart entre ce qu'il rapporte aujourd'hui à la banque et ce que rapportent les meilleurs clients de son groupe de comparaison. La mesure utilisée est le **PNB** (produit net bancaire) par client sur douze mois glissants, décomposé en quatre univers : épargne, crédit, assurance et services. On note $P_{u,i}$ le PNB du client i dans l'univers u, et $c(i)$ son groupe.

Étape 1 : identifier les clients de référence

Chaque univers reçoit, dans chaque groupe, un poids égal à sa part dans le PNB moyen du groupe :

$$w_{u,c} = \frac{\left|\bar{P}_{u,c}\right|}{\sum_{u'} \left|\bar{P}_{u',c}\right|} \tag{2.1}$$

où $\bar{P}_{u,c}$ est le PNB moyen de l'univers u dans le groupe c. Le PNB de chaque univers est ensuite ramené entre 0 et 1 au sein du groupe, pour que les univers soient comparables : sans cela, un univers où le PNB se compte en centaines d'euros écraserait un univers où il se compte en dizaines.

$$\tilde{P}_{u,i} = \frac{P_{u,i} - \min\limits_{j \in c} P_{u,j}}{\max\limits_{j \in c} P_{u,j} - \min\limits_{j \in c} P_{u,j}} \tag{2.2}$$

Le score de benchmark agrège ces valeurs pondérées ; il est compris entre 0 et 1 :

$$S_i = \sum_{u} w_{u,c(i)}\, \tilde{P}_{u,i} \tag{2.3}$$

Dans chaque groupe, les 25 % de clients au score le plus élevé forment les **clients de référence**. Ce seuil est un choix : assez large pour que les médianes calculées sur ces clients soient stables, assez exigeant pour représenter un objectif ambitieux mais atteignable. On note $m_{u,c}$ la médiane du PNB de l'univers u chez les clients de référence du groupe c.

Étape 2 : calculer le potentiel par univers

Assurance et services. Ces univers se cumulent (assurance auto, habitation, santé…) : le potentiel est l'écart à la médiane des références, jamais négatif.

$$\mathrm{Pot}_{u,i} = \max\left(0,\ m_{u,c(i)} - \max\left(0, P_{u,i}\right)\right) \tag{2.4}$$

Crédit. Même écart, mais seulement pour les clients qui ne détiennent aucun crédit ; il est nul pour les autres. On ne cherche pas à proposer un crédit supplémentaire à un client déjà endetté. L'éligibilité réelle de chaque client relève ensuite des règles de risque de la banque, qui n'entrent pas dans ce calcul.

$$\mathrm{Pot}_{\mathrm{créd},i} = \begin{cases} \max\left(0,\ m_{\mathrm{créd},c(i)} - \max\left(0, P_{\mathrm{créd},i}\right)\right) & \text{si le client ne détient aucun crédit} \\ 0 & \text{sinon} \end{cases} \tag{2.5}$$

Épargne. Le potentiel d'épargne repose sur la part de l'argent qui reste sur le compte courant (dépôts à vue, DAV) plutôt que placée. Pour chaque client dont l'épargne dépasse 200 €, on calcule :

$$r_i = \frac{\text{DAV}_i}{E_i^{\text{liquide}} + E_i^{\text{stable}}} \quad \text{borné entre 0 et 1} \tag{2.6}$$

On note $r_c^{\text{réf}}$ la médiane de ce ratio chez les clients de référence du groupe. Un client qui laisse proportionnellement plus d'argent sur son compte courant que les références a un potentiel d'épargne ; un client qui épargne déjà autant qu'elles, ou davantage, a un potentiel nul :

$$\text{Pot}_{\text{ép},\,i} = \begin{cases} m_{\text{ép},\,c(i)} \times \dfrac{r_i - r_c^{\text{réf}}}{1 - r_c^{\text{réf}}} & \text{si } r_i > r_c^{\text{réf}} \\ 0 & \text{sinon} \end{cases} \tag{2.7}$$

La fraction mesure l'excès du client par rapport aux références, rapporté à la marge maximale possible : elle vaut 0 au niveau des références et 1 si tout l'argent du client dort sur son compte courant.

Le **potentiel total** est la somme des quatre univers, en euros par an. Il sert à classer les clients pour prioriser les contacts, et son univers principal indique la proposition à travailler (par exemple : « potentiel crédit = 72 % du total »).

2.6 La segmentation finale

Le potentiel seul ne suffit pas à choisir une action : un client très engagé avec la banque et un client presque absent n'appellent pas le même discours, même à potentiel égal. La segmentation finale croise donc la position du client dans son groupe avec sa **note MIRE** (mesure de l'intensité relationnelle), une note interne existante de 0 à 100 qui tient compte du nombre de produits détenus dans chaque univers, de la diversité des univers et de l'activité transactionnelle. Je l'utilise telle quelle, sans la recalculer.

Chaque client reçoit deux notes de 1 à 4 :

- une **note de performance** : le quartile de son score de benchmark au sein de son groupe, de 1 (quart inférieur) à 4 (quart supérieur) ;
- une **note relationnelle** : 1 si la note MIRE est inférieure à 25, 2 de 25 à 49, 3 de 50 à 74, 4 à partir de 75.

La somme des deux notes, de 2 à 8, détermine le segment (figure 2.3).

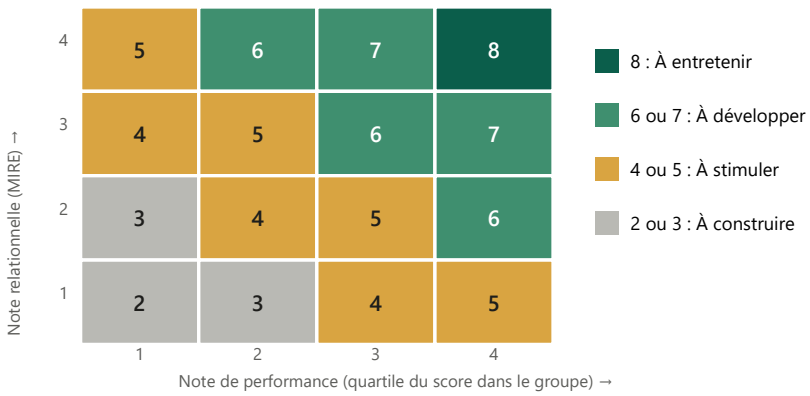

Figure 2.3 : Le segment est déterminé par la somme de la note de performance et de la note relationnelle.

Segment	Part des clients	PNB moyen	Potentiel moyen	Note MIRE moyenne
À développer	42,9 %	666 €	362 €	62,6
À stimuler	39,1 %	208 €	506 €	37,0
À construire	9,1 %	28 €	464 €	18,9
À entretenir	8,9 %	1 458 €	112 €	84,1

Tableau 2.2 : Caractéristiques des segments finaux (262 929 clients actifs). PNB et potentiel annuels.

Les quatre segments se lisent simplement :

- **À entretenir** : le PNB le plus élevé, le potentiel le plus faible et une relation très forte. Ces clients sont à fidéliser plus qu'à équiper.
- **À développer** : un PNB déjà élevé et un potentiel encore important. C'est le cœur de cible pour compléter l'équipement.
- **À stimuler** : un PNB intermédiaire, mais le potentiel le plus élevé ; une relation à activer.
- **À construire** : un PNB presque nul et une relation faible, mais un potentiel réel. Il s'agit probablement de relations récentes ou de clients servis surtout par une autre banque.

> **Deux précautions de lecture.** La concentration de 82 % des clients dans les deux segments centraux ne traduit pas une loi statistique : elle vient du barème. La somme de deux notes de 1 à 4 prend plus souvent des valeurs centrales que des valeurs extrêmes, et « À entretenir » exige que les deux notes soient au maximum. L'effet est voulu, puisque la majorité des clients se retrouve dans des segments pilotables à grande échelle, mais c'est une conséquence de la construction, pas une découverte.
>
> De même, la relation inverse entre PNB actuel et potentiel est attendue : le potentiel étant un écart à la médiane des meilleurs clients, il diminue mécaniquement quand le PNB augmente. Elle confirme que le calcul se comporte comme prévu ; elle ne prouve pas que le potentiel annonce des revenus futurs. Seul un suivi dans le temps, qui mesurerait la part du potentiel effectivement transformée en PNB, permettrait de le valider.

CHAPITRE 3

Des personas enrichis par les paiements

La segmentation de potentiel répond à la question « qui cibler, et sur quel univers ? ». Elle ne répond pas à celle du Marketing : « à qui parle-t-on, et comment ? ». Un segment défini par « quart supérieur du score et note MIRE d'au moins 75 » ne permet pas d'écrire une campagne. J'ai donc construit une seconde segmentation, conçue d'abord pour être comprise.

3.1 Des variables lisibles

Cette segmentation reprend les variables du chapitre 2 (situation, revenu, enfants, logement), en y ajoutant cette fois l'âge, découpé en quatre grandes étapes de vie : installation (18-29 ans), constitution du patrimoine (30-49 ans), préparation de la retraite (50-64 ans) et retraite (65 ans et plus). Sur plus de 50 variables disponibles, on n'en garde ainsi que onze dimensions : deux pour la situation, quatre pour l'âge, une pour le revenu, deux pour les enfants et deux pour le logement.

3.2 Le choix de cinq groupes

Un K-means regroupe les clients selon ces onze dimensions. Le nombre de groupes a été choisi en comparant les solutions de 5 à 9 groupes (figure 3.1).

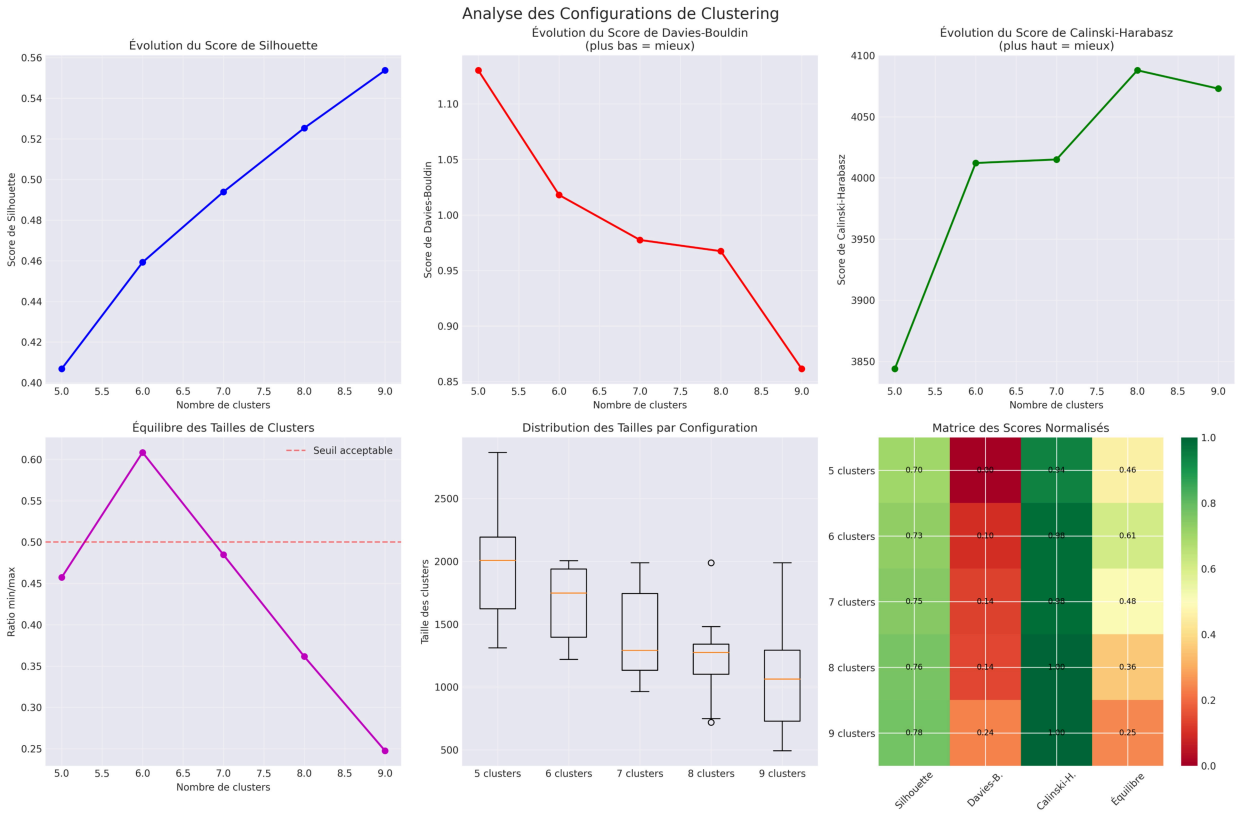

Figure 3.1 : Comparaison des solutions de 5 à 9 groupes : silhouette, Davies-Bouldin, Calinski-Harabasz, équilibre et distribution des tailles, synthèse normalisée.

Les indicateurs techniques s'améliorent presque tous avec le nombre de groupes : la silhouette passe de 0,41 avec cinq groupes à 0,55 avec neuf, l'indice de Davies-Bouldin baisse de 1,13 à 0,86, et l'indice de Calinski-Harabasz progresse d'environ 3 840 à près de 4 090. L'équilibre des tailles, mesuré par le rapport entre le plus petit et le plus

grand groupe, est le meilleur avec six groupes (0,61) et reste correct avec cinq (0,46).

J'ai pourtant retenu cinq groupes, pour une raison métier. Avec cinq groupes, la segmentation suit une logique claire : d'abord la présence d'enfants, puis l'âge pour les foyers sans enfants. Au-delà, les groupes supplémentaires redécoupent les tranches d'âge selon la situation familiale ou le logement, des variables très liées à l'âge dans une population régionale assez homogène. Les indicateurs progressent parce que ces découpages sont plus nets géométriquement, pas parce qu'ils décrivent des comportements différents. Pour un usage marketing, cinq groupes que l'on sait nommer valent mieux que neuf que l'on ne sait pas distinguer.

Les cinq groupes sont donc : les **Jeunes**, les **Actifs**, les **Pré-retraités**, les **Retraités** (foyers sans enfants, par étape de vie) et les **Familles** (foyers avec enfants, tous âges confondus).

3.3 Les paiements par carte comme révélateur

Les paiements par carte, classés par code commerçant (MCC, *Merchant Category Code*), montrent comment les clients dépensent réellement. Sur douze mois glissants, les centaines de codes MCC ont été regroupés, avec le Marketing, en cinq catégories :

- **Essentiel** : alimentation, santé, énergie, télécommunications, logement, éducation, assurances ;
- **Mobilité** : transports, carburant, automobile, voyages, hôtels ;
- **Lifestyle** : restaurants, mode, décoration, loisirs, commerces spécialisés ;
- **Services** : banques, immobilier, services professionnels, impôts et taxes ;
- **Commerce** : grande distribution, bricolage, commerce de gros.

Pour comparer des clients aux budgets très différents, on raisonne en parts. Pour le client i et la catégorie j, de dépense D_{ij} :

$$R_{ij} = \frac{D_{ij}}{\sum_{k=1}^{5} D_{ik}} \tag{3.1}$$

Les habitudes varient naturellement avec l'âge et la situation familiale. Pour repérer ce qui est atypique *au sein* d'un groupe, chaque part est donc comparée à la médiane du groupe, en unités de déviation absolue médiane (MAD), un indicateur de dispersion peu sensible aux valeurs extrêmes (Leys et al., 2013) :

$$z_{ij} = \frac{R_{ij} - \text{médiane}_{g}(R_{\cdot j})}{\text{MAD}_{g}(R_{\cdot j})} \tag{3.2}$$

où g est le groupe du client. Un client sur-consomme une catégorie si $z_{ij} > 1{,}5$, et la sous-consomme si $z_{ij} < -1{,}5$. Ce seuil est volontairement bas. Pour une distribution normale, la MAD vaut environ 0,67 écart-type : 1,5 MAD correspondent donc à environ un écart-type, bien en deçà des seuils usuels de détection de valeurs aberrantes (Iglewicz et Hoaglin, 1993 ; Leys et al., 2013). L'objectif n'est pas de repérer des anomalies, mais des sous-groupes aux habitudes marquées.

Un tel sous-groupe n'est retenu comme **profil remarquable** que s'il remplit deux conditions : réunir au moins 20 % des clients du groupe, et s'écarter d'au moins 15 points de la médiane du groupe dans la catégorie concernée.

3.4 Ce que révèlent les dépenses

Groupe	Clients	Dépense annuelle	Paiements par an	Montant moyen	Essentiel	Mobilité	Lifestyle	Services	Commerce
Jeunes (18-30)	51 117	11 374 €	370	30,70 €	18,0 %	17,0 %	35,3 %	15,9 %	13,7 %
Actifs (30-50)	47 200	15 411 €	348	44,35 €	21,6 %	16,0 %	28,1 %	17,2 %	17,2 %
Pré-retraités (50-65)	65 468	15 826 €	304	51,99 €	24,3 %	15,7 %	23,4 %	16,8 %	19,8 %
Retraités (65+)	74 807	11 534 €	206	55,92 €	27,0 %	12,4 %	17,2 %	23,4 %	20,0 %
Familles	70 175	16 892 €	357	47,37 €	24,8 %	14,8 %	27,3 %	14,4 %	18,8 %

Tableau 3.1 : Dépenses par carte des cinq groupes sur douze mois. Les cinq dernières colonnes donnent la part de chaque catégorie dans la dépense moyenne du groupe.

Plusieurs tendances se dégagent nettement :

- avec l'âge, la part des loisirs (Lifestyle) baisse régulièrement, de 35 % du budget carte chez les 18-30 ans à 17 % après 65 ans, tandis que celle des dépenses essentielles monte de 18 % à 27 % ;
- les services ne prennent du poids qu'après 65 ans : 23 % du budget, contre 16 à 17 % avant ;
- les clients plus âgés paient moins souvent par carte, mais pour des montants plus élevés : 370 paiements par an à 31 € en moyenne chez les 18-30 ans, 206 paiements à 56 € après 65 ans.

Groupe	Profil remarquable	Part du groupe	Part du budget dans la catégorie	Médiane du groupe	Écart
Jeunes (18-30)	Sur-consommation Lifestyle	27,0 %	47,8 %	31,2 %	+16,6 pts
Jeunes (18-30)	Sur-consommation Essentiel	24,8 %	35,4 %	19,4 %	+16,1 pts
Actifs (30-50)	Sur-consommation Essentiel	28,0 %	38,4 %	21,6 %	+16,7 pts
Actifs (30-50)	Sur-consommation Lifestyle	25,3 %	42,2 %	25,7 %	+16,5 pts
Pré-retraités (50-65)	Sur-consommation Essentiel	25,0 %	51,7 %	25,8 %	+25,8 pts
Familles	Sur-consommation Essentiel	28,0 %	39,1 %	24,4 %	+14,7 pts*

Tableau 3.2 : Profils remarquables par groupe. * Écart légèrement inférieur au seuil de 15 points ; le profil avait été retenu dans la version d'origine et il est conservé ici à titre indicatif. Pour les retraités, le profil remarquable de la sortie d'origine reproduit celui des familles (mêmes libellés, mêmes valeurs) : il n'est pas exploitable et n'est pas repris.

Chaque groupe cache ainsi des sous-groupes aux habitudes très différentes. Chez les 18-30 ans, un quart des clients consacre près de la moitié de son budget carte aux loisirs, quand un autre quart en consacre plus d'un tiers aux dépenses essentielles : deux situations qui n'appellent pas le même discours, et qu'une segmentation par âge seule confondrait. Chez les 50-65 ans, un quart des clients consacre plus de la moitié de son budget carte aux dépenses essentielles, ce qui peut signaler une contrainte budgétaire.

3.5 La lecture inverse : des comportements aux groupes

L'analyse peut aussi partir des comportements. Les clients porteurs de carte sont regroupés selon leur catégorie de dépense dominante, puis on regarde quels groupes composent chaque comportement. L'**indice de représentation** rapporte la part d'un groupe dans un comportement à sa part dans l'ensemble : au-dessus de 1, le groupe est surreprésenté ; en dessous, sous-représenté.

Comportement dominant	Part des clients	Jeunes	Actifs	Pré-retraités	Retraités	Familles
Lifestyle	25,9 %	**2,18**	**1,24**	0,68	0,28	1,07
Essentiel	25,0 %	0,31	0,76	1,09	**1,45**	1,11
Commerce	19,0 %	0,36	0,80	**1,20**	**1,35**	1,04
Mobilité	18,3 %	1,19	**1,20**	1,16	0,64	0,99
Services	11,9 %	0,60	0,98	0,93	**1,60**	0,58

Tableau 3.3 : Indices de représentation des groupes dans chaque comportement dominant (209 018 clients porteurs de carte). Dans ce périmètre, les groupes pèsent respectivement 17,7 % (Jeunes, 18-29 ans), 16,8 % (Actifs, 30-49 ans), 20,2 % (Pré-retraités, 50-64 ans), 27,4 % (Retraités) et 17,9 % (Familles). En couleur : indice d'au moins 1,20 ou inférieur à 0,80.

Le Lifestyle est très marqué par l'âge : les 18-29 ans y sont deux fois plus présents que dans l'ensemble (indice 2,18), les retraités presque quatre fois moins (0,28). Les services attirent au contraire les retraités (1,60). Le comportement Essentiel n'est pas neutre non plus : les retraités y sont surreprésentés (1,45) et les jeunes très sous-représentés (0,31). Seule la Mobilité se répartit à peu près également entre les âges actifs, avant un net recul après 65 ans (0,64).

3.6 Synthèse : six personas

Les deux lectures se recoupent et dessinent six personas (tableau 3.4).

Persona	Lecture directe (groupe → dépenses)	Lecture inverse (comportement → groupes)
Jeunes hédonistes	Un quart des 18-30 ans consacre près de la moitié de son budget carte aux loisirs.	Jeunes deux fois plus présents dans le comportement Lifestyle.
Jeunes sous contrainte	Un quart des 18-30 ans consacre plus d'un tiers de son budget aux dépenses essentielles.	Profil minoritaire : les jeunes restent sous-représentés dans le comportement Essentiel.
Actifs partagés	Deux sous-groupes de taille comparable, l'un tourné vers l'essentiel, l'autre vers les loisirs.	Surreprésentés dans le Lifestyle et la Mobilité.
Pré-retraités sous contrainte	Un quart consacre plus de la moitié de son budget aux dépenses essentielles.	Surreprésentés dans le Commerce ; présence légèrement supérieure à la moyenne dans l'Essentiel.
Retraités tournés vers les services	Les services pèsent 23 % du budget carte, nettement plus qu'aux autres âges.	Surreprésentés dans les Services et l'Essentiel.
Familles	Plus d'un quart des familles concentre près de 40 % de son budget sur l'essentiel.	Présence proche de la moyenne dans tous les comportements, sauf les Services.

Tableau 3.4 : Personas issus du croisement des deux lectures.

3.7 Limites, et pourquoi garder deux segmentations

Un biais de couverture. Seuls les clients qui paient par carte au Crédit Agricole apparaissent dans l'analyse des dépenses. En sont absents ou sous-représentés les clients qui utilisent peu la carte, comme certains clients patrimoniaux ou des personnes qui privilégient les espèces, ceux qui paient surtout avec la carte d'une autre banque, et les comptes peu actifs. Les personas décrivent donc les clients actifs par carte, pas l'ensemble du portefeuille.

Deux périmètres différents. La lecture directe porte sur 308 767 clients, la lecture inverse sur 209 018 porteurs de carte, avec des bornes d'âge légèrement différentes (18-30 et 18-29 ans, par exemple). La version d'origine ne documentait pas ces deux périmètres, et je n'ai plus accès aux données pour les réconcilier : les deux lectures se comparent en tendance, pas chiffre à chiffre.

Le biais de couverture justifie de garder deux segmentations séparées plutôt que de les fusionner :

- la **segmentation de potentiel** couvre tous les clients actifs, évolue lentement et sert au pilotage commercial : objectifs, priorités de contact, univers à travailler ;

- les **personas** décrivent les modes de vie des clients actifs par carte et servent à la communication : messages, canaux, idées d'offres.

Les deux peuvent se croiser ponctuellement, sur un périmètre restreint, par exemple pour tester une offre auprès d'un segment pilote.

CHAPITRE 4

Parcours 360°, l'application des conseillers

Une segmentation n'a d'effet que si elle arrive jusqu'aux personnes qui rencontrent les clients. Or, pour préparer un rendez-vous, un conseiller pouvait avoir à consulter jusqu'à sept applications. Parcours 360° part de ce constat : une page unique qui réunit ce qu'il faut savoir d'un client, et qui pourra ensuite accueillir les scores des chapitres précédents.

4.1 Commencer par un irritant quotidien

La première fonctionnalité n'a pas été un score, mais la « ligne de vie » du client : une chronologie de ses contrats, de ses contacts avec la banque, des propositions commerciales et des réclamations. Ce n'est pas la fonctionnalité la plus ambitieuse, mais c'est celle qui répondait à une gêne reconnue par tous : la dispersion des informations. L'idée était de faire de l'outil un réflexe avant d'y ajouter des éléments plus avancés.

4.2 Une architecture dictée par l'existant

Le serveur Apache de la caisse régionale ne peut pas exécuter Node.js : impossible, donc, d'utiliser le rendu côté serveur de Next.js. L'application, écrite avec React 18 et Next.js 13, est exportée en fichiers statiques (HTML, CSS, JavaScript) servis directement par Apache. Avec l'option `output: 'export'` (listing 4.1), la commande de construction produit directement ce site statique. L'état partagé entre les pages passe par le Context de React, sans bibliothèque supplémentaire.

```
// next.config.js
module.exports = {
  output: 'export',             // site entièrement statique, sans serveur Node.js
  images: { unoptimized: true },   // pas d'optimisation d'images côté serveur
  basePath: '/parcours360',       // sous-répertoire du serveur Apache
  assetPrefix: '/parcours360/',
  trailingSlash: true,
};
```

Listing 4.1 : Configuration Next.js pour l'export statique.

Un backend PHP assure trois fonctions : l'authentification unique (SSO) via deux frameworks internes, l'un pour identifier le collaborateur, l'autre pour vérifier ses droits (listing 4.2) ; une API REST qui expose les données aux seuls utilisateurs authentifiés ; et la journalisation des consultations.

```
class SSOAuthMiddleware {
    public function authenticate($request) {
        // Identification du collaborateur par le premier framework
        $matricule = $this->frameworkSSO->getMatricule();

        // Vérification des droits par le second framework
        $droits = $this->frameworkDroits->validateAccess($matricule, 'PARCOURS_360');

        if (!$droits->hasAccess()) {
            throw new UnauthorizedException();
        }
        return ['matricule' => $matricule, 'droits' => $droits];
    }
}
```

Listing 4.2 : Middleware d'authentification (simplifié).

Les données viennent de SQL Server, interrogé par des procédures stockées. La base est alimentée chaque jour depuis Teradata par OpenQuery : l'application n'interroge jamais Teradata directement, et affiche la date de fraîcheur des données (figure 4.1).

4.3 Les fonctionnalités

La recherche est le point d'entrée de l'application. Elle accepte un numéro de compte ou un nom, dans un ordre quelconque, et gère les noms composés (listing 4.3, figure 4.1).

```
const searchClient = (query) => {
  const q = query.trim().toLowerCase();

  // Numéro de compte : 14 chiffres
  if (/^\d{14}$/.test(q)) return searchByAccountNumber(q);

  // Nom et prénom : ordre libre, noms composés avec ou sans tiret
  return searchByName([
    q,
    q.split(' ').reverse().join(' '),
    ...q.split(' ').flatMap((part) =>
      part.includes('-') ? [part, part.replaceAll('-', ' ')] : [part]),
  ]);
};
```

Listing 4.3 : Recherche tolérante aux formats de saisie.

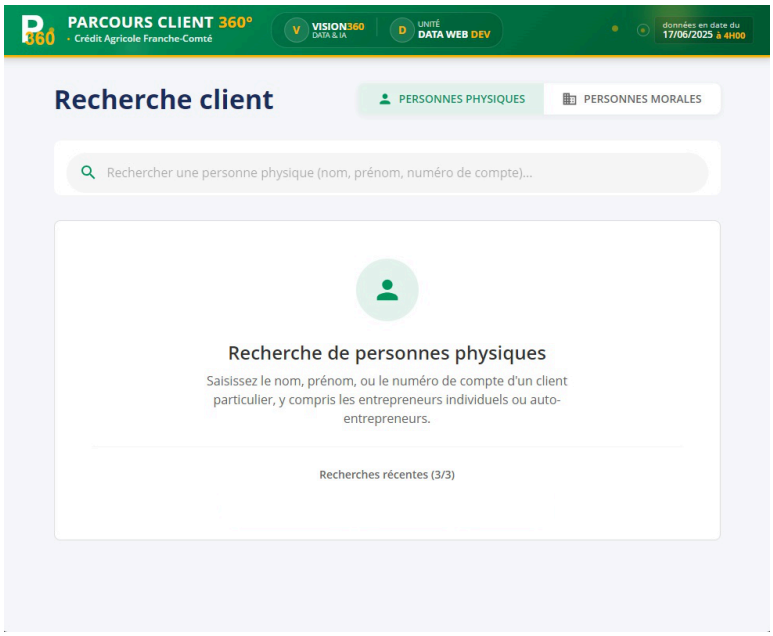

Figure 4.1 : Parcours 360°, page de recherche. La date de fraîcheur des données apparaît en haut à droite.

La fiche client regroupe les informations en modules : synthèse (données démographiques, agence, conseiller attitré), contrats ouverts et clos sur l'année, historique relationnel (contacts, réclamations, avis clients), usage des canaux digitaux et préférences de contact, actions commerciales en cours. Toutes ses captures contiennent des données clients ; aucune n'est reproduite ici.

4.4 Déploiement et adoption

Le déploiement s'est fait en deux temps : une phase de test auprès d'un groupe de conseillers, avec un questionnaire envoyé par e-mail pour recueillir leurs retours, puis l'ouverture à l'ensemble du réseau. Chaque consultation est journalisée (horodatage, matricule du conseiller, identifiant du client consulté), ce qui permet de mesurer l'usage réel ; ces journaux contiennent des identifiants de conseillers et de clients et doivent être protégés comme les données clients.

Phase	Jours ouvrés	Utilisateurs	Consultations	Consultations par jour
Test (mai-juin 2025)	38	113	441	11,6
Ouverture au réseau (juillet 2025)	11	165	385	35,0

Tableau 4.1 : Usage de Parcours 360°, mesuré dans les journaux de consultation. Pendant le test, 345 clients différents ont été consultés.

Pendant le test, l'outil a été peu utilisé : sur 113 conseillers qui l'ont ouvert, seuls 9 (8 %) y revenaient régulièrement, alors même que les retours oraux étaient positifs. Les échanges ont fait ressortir trois causes : l'outil ne figurait pas dans les favoris des postes, sa valeur n'avait pas été expliquée, et les habitudes allaient aux applications connues. Cet écart entre l'intention déclarée et l'usage est classique : l'adoption dépend autant des conditions d'accès et de l'accompagnement que de l'utilité perçue (Venkatesh et al., 2003 ; Rogers, 2003).

Avant l'ouverture au réseau, nous avons donc travaillé l'accompagnement : une vidéo de 45 secondes montrant ce que l'outil apporte, diffusée sur les canaux internes, et des sessions « Micro ouvert », des démonstrations informelles d'un quart d'heure, qui se sont révélées plus efficaces que des formations classiques.

Après l'ouverture, l'usage quotidien a été multiplié par trois : 35 consultations par jour ouvré contre 11,6 pendant le test. Le résultat est encourageant mais repose sur une période courte, onze jours ouvrés, dans la foulée de la communication de lancement. Seul un suivi sur plusieurs mois dira si l'usage se maintient.

4.5 Les suites prévues

La suite logique est d'afficher dans la fiche client le segment et le potentiel par univers du chapitre 2, accompagnés d'une explication lisible : pourquoi ce client a-t-il ce potentiel, et par rapport à quels clients ? Deux autres évolutions ont été identifiées avec les conseillers : des suggestions de produits fondées sur le potentiel et l'historique, et une version adaptée à la tablette pour les rendez-vous hors de l'agence.

CHAPITRE 5

Bilan

5.1 Ce qui a été livré

- Une **segmentation de potentiel** calculée pour 262 929 clients actifs, avec un potentiel en euros par univers et un segment d'action, mise à la disposition du Marketing pour le ciblage des campagnes.
- Des **personas** fondés sur les dépenses par carte, présentés aux équipes du Marketing.
- **Parcours 360°**, ouvert à l'ensemble du réseau, avec un usage quotidien multiplié par trois après l'accompagnement du lancement.

Les retours des conseillers évoquent un gain de temps à la préparation des rendez-vous. Ce gain n'a pas été mesuré, et je ne l'avance pas comme un résultat.

5.2 La méthode de travail

Le travail s'est organisé en cycles de deux à trois semaines, chacun conclu par une démonstration aux équipes métier. Ce rythme a orienté les développements : la toute première présentation, une simple répartition des clients par âge, a suscité les questions qui ont guidé la suite (« pourquoi ne pas croiser avec les revenus ? », « comment intégrer les frontaliers ? »). Des points hebdomadaires avec mon maître d'apprentissage et des revues régulières avec ma tutrice universitaire complétaient ce suivi.

Deux réorientations ont marqué l'année. La découverte tardive du biais de couverture des données carte a conduit à séparer les deux segmentations plutôt qu'à les fusionner. Et les priorités de la direction ont glissé de « mieux comprendre nos clients » à « comment les conseillers peuvent-ils s'en servir ? », ce qui a déplacé du temps de l'optimisation des modèles vers l'application et l'accompagnement des utilisateurs.

5.3 Les limites

- **Des groupes de comparaison peu séparés.** La segmentation de potentiel ne découvre pas de groupes naturels (section 2.4) ; elle fournit des bases de comparaison, et doit être présentée comme telle.
- **Un potentiel non validé dans le temps.** Rien ne montre encore que le potentiel annonce des revenus futurs : cela demande de suivre sa conversion en PNB réalisé.
- **Une couverture partielle des personas**, limités aux clients actifs par carte (section 3.7).
- **Un traitement mensuel.** L'infrastructure ne permet pas de traitement en continu : un changement brutal de situation n'est visible qu'au calcul suivant.
- **Une adoption à confirmer**, mesurée sur une période courte, avec une part encore faible d'utilisateurs réguliers.
- **Une gouvernance de la donnée embryonnaire.** Sans responsable des données attitré, les référentiels présentent des incohérences qui compliquent les analyses entre systèmes.

5.4 Ce que je ferais autrement

La relecture de ce mémoire, un an après, a fait apparaître des erreurs de recopie entre les sorties des traitements et le texte, et des affirmations plus fortes que ce que les chiffres permettaient. J'en tire quatre leçons de méthode.

- **Mesurer avant d'affirmer.** Prévoir la mesure dès la conception : le temps de préparation d'un rendez-vous avant et après Parcours 360°, la conversion du potentiel.
- **Produire les chiffres du rapport depuis le code.** Générer tableaux et figures directement à partir des sorties du traitement, plutôt que de les recopier à la main.

- **Tester le potentiel** en comparant un groupe de conseillers qui en dispose à un groupe témoin.
- **Documenter le périmètre de chaque analyse** : quels clients, quelle date, quels filtres.

5.5 Perspectives et apport personnel

À court terme, l'intégration du potentiel dans Parcours 360° ferait de l'application le canal par lequel les modèles atteignent le terrain. À plus long terme, je recommande de structurer une gouvernance de la donnée à l'échelle de la caisse régionale, avec des référentiels communs et des responsables de données identifiés, pour passer d'initiatives portées par quelques personnes motivées à une démarche durable.

Pour ma part, cette année m'a appris que la valeur d'un modèle tient d'abord à son appropriation par ceux qui doivent s'en servir, et qu'il faut autant de rigueur dans ce que l'on affirme que dans ce que l'on calcule. Ce sont les deux convictions qui orientent la suite de mon parcours, à l'interface entre la technique et les métiers.

Glossaire

CA-TS
Crédit Agricole Technologies et Services, entité qui mutualise les services informatiques des caisses régionales.

Calinski-Harabasz (indice de)
Rapport entre la dispersion entre groupes et la dispersion au sein des groupes ; plus il est élevé, mieux c'est.

CSP
Catégorie socio-professionnelle, selon la nomenclature de l'INSEE.

DAV
Dépôts à vue : l'argent disponible sur le compte courant.

Davies-Bouldin (indice de)
Mesure de la ressemblance moyenne entre chaque groupe et son voisin le plus proche ; plus il est bas, mieux c'est.

Export statique
Construction d'une application web sous forme de fichiers HTML, CSS et JavaScript, servis sans serveur applicatif.

Frontalier
Personne qui réside en France et travaille en Suisse.

JCEKS
Format de magasin de clés chiffré, utilisé pour stocker les secrets applicatifs hors du code.

K-means
Algorithme qui répartit des observations en k groupes en minimisant la distance de chacune au centre de son groupe.

Kerberos
Protocole d'authentification par tickets, qui évite de faire circuler des mots de passe sur le réseau.

MAD
Déviation absolue médiane : médiane des écarts à la médiane, mesure de dispersion peu sensible aux valeurs extrêmes.

MCC
Merchant Category Code : code à quatre chiffres qui classe les commerçants par activité.

MIRE
Note interne de 0 à 100 mesurant l'intensité de la relation d'un client avec la banque.

PNB
Produit net bancaire : pour un client, les revenus qu'il génère pour la banque sur douze mois glissants.

SID
Système d'information décisionnel : entrepôt de données destiné à l'analyse.

Silhouette (coefficient de)
Pour chaque observation, compare la distance à son groupe et la distance au groupe voisin ; de −1 à 1, plus il est élevé, mieux les groupes sont séparés.

SSO
Single Sign-On : authentification unique, qui donne accès à plusieurs applications.

Bibliographie

Caliński, T., et Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27.

Davies, D. L., et Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227.

Hastie, T., Tibshirani, R., et Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2e éd.). Springer.

Iglewicz, B., et Hoaglin, D. C. (1993). *How to Detect and Handle Outliers*. ASQC Quality Press.

Kumar, V., et Reinartz, W. (2012). *Customer Relationship Management: Concept, Strategy, and Tools* (2e éd.). Springer.

Leys, C., Ley, C., Klein, O., Bernard, P., et Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel (règlement général sur la protection des données). *Journal officiel de l'Union européenne*, L 119, 1–88.

Rogers, E. M. (2003). *Diffusion of Innovations* (5e éd.). Free Press.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

Venkatesh, V., Morris, M. G., Davis, G. B., et Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.

Résumé

Ce mémoire présente une année d'alternance au Crédit Agricole Franche-Comté consacrée à une question : comment transformer les données bancaires en outils d'aide à la décision, pour le Marketing et pour les conseillers ? Trois réalisations y répondent. Une segmentation de potentiel compare chaque client aux meilleurs clients d'un groupe de clients comparables, construit par stratification d'âge et K-means, pour estimer en euros ce qu'il pourrait apporter dans chaque univers de produits. Une segmentation en personas, volontairement simple, s'appuie sur les dépenses par carte pour décrire les modes de vie de cinq groupes et de leurs sous-profils. Enfin, l'application web Parcours 360° réunit sur une page les informations utiles à un conseiller ; son usage quotidien a triplé une fois l'accompagnement des utilisateurs mis en place. Le mémoire discute aussi les limites de ces travaux : des groupes de comparaison peu séparés, un potentiel encore non validé dans le temps, une couverture partielle des données carte et une adoption à confirmer.

Mots-clés : segmentation client, potentiel commercial, K-means, données de paiement, application métier, adoption, banque régionale.

Abstract

This thesis reports on a work-study year at Crédit Agricole Franche-Comté, built around one question: how can banking data become decision-support tools for the Marketing department and for client advisors? Three deliverables address it. A potential-based segmentation compares each customer with the best customers of a group of similar customers, built through age stratification and K-means, to estimate in euros what they could bring in each product line. A deliberately simple persona segmentation uses card spending to describe the lifestyles of five groups and their sub-profiles. Finally, the Parcours 360° web application brings together on a single page the information an advisor needs; its daily use tripled once user support was in place. The thesis also discusses the limits of this work: weakly separated comparison groups, a potential not yet validated over time, partial coverage of card data, and adoption still to be confirmed.

Keywords: customer segmentation, commercial potential, K-means, payment data, business application, user adoption, regional bank.