# Analyse et traitement de la data client

Cette fiche montre comment une équipe commerciale rassemble, fiabilise et exploite les données de ses clients et de ses prospects. Elle donne les indicateurs de qualité d'un fichier, les grands types d'analyse et la méthode RFM, utilisable tout de suite pour cibler une relance ou une offre.

**Point d'actualité 2026 : data client et intelligence artificielle**

- **Conservation :** les données d'un prospect se gardent trois ans après leur collecte ou le dernier contact émanant de lui ; un clic dans un e-mail compte comme contact, la simple ouverture d'un e-mail, non.
- **Pixels de suivi :** selon la recommandation de la CNIL publiée le 14 avril 2026, mesurer l'ouverture des e-mails personne par personne pour personnaliser, optimiser les envois ou profiler suppose le consentement du destinataire, sauf exceptions étroites (sécurité, fraude, délivrabilité).
- **Règlement européen sur l'IA :** depuis le 2 août 2026, un client doit être clairement informé qu'il échange avec un chatbot ou un agent d'IA ; le règlement (UE) 2026/1744 du 8 juillet 2026 a repoussé à 2027 et 2028 les obligations des IA « à haut risque ».
- **IA générative :** utile pour résumer des avis ou repérer les motifs d'insatisfaction, avec des données pseudonymisées (nom et e-mail remplacés par un code) et des résultats vérifiés.

Sources : CNIL – référentiel « gestion des activités commerciales » (consulté en 2026) ; CNIL – webinaire sur les pixels de suivi (mai 2026) ; Commission européenne – une IA plus transparente (août 2026) ; EUR-Lex – règlement (UE) 2026/1744 (2026)

## La data client, matière première de la vente

La data client désigne l'ensemble des données qu'une entreprise détient sur ses clients et ses prospects. Bien exploitée, elle dit à qui s'adresser, quand relancer et quoi proposer. Deux distinctions servent à la classer.

- **Structurées ou non structurées.** Une donnée structurée se range dans les colonnes d'une base de données (date, montant, code postal). Une donnée non structurée est un texte libre, une image, un message ou un e-mail : elle demande d'autres outils d'analyse.
- **Internes ou externes.** Les données internes viennent des outils de l'entreprise : caisse, ERP (logiciel de gestion intégré), CRM (logiciel de gestion de la relation client), service après-vente, ses propres réseaux sociaux (social CRM). Les données externes viennent d'ailleurs : données ouvertes (INSEE, météo), moteurs de recherche, réseaux sociaux, fichiers achetés.

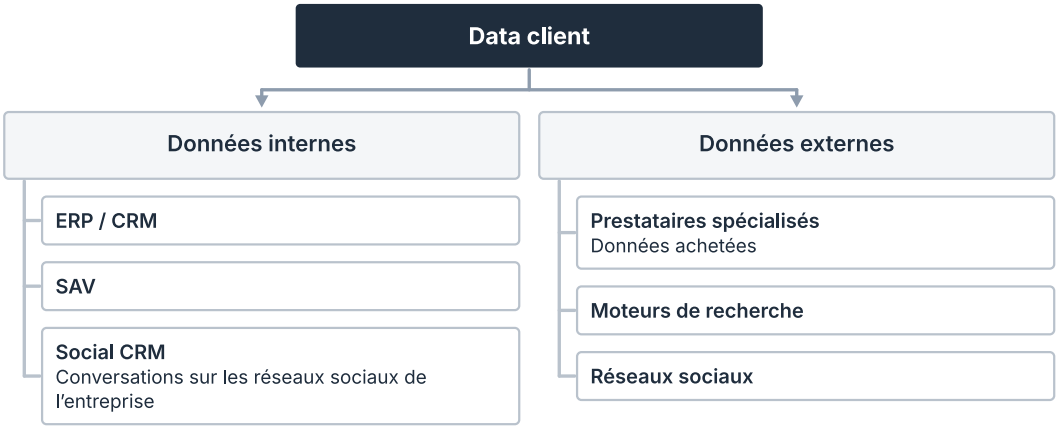



**Schéma 1.** D'où vient la data client : des sources internes et des sources externes.

Le big data (mégadonnées) désigne les outils qui transforment de très grandes quantités de données en informations utiles à la décision. On l'analyse par cinq V :

- **volume** : la quantité de données à stocker et à traiter ;
- **variété** : les types de données, structurées ou non, internes ou externes ;
- **vélocité** : la rapidité avec laquelle elles arrivent et doivent être traitées ;
- **véracité** : leur fiabilité, que dégradent doublons, adresses fausses et fiches incomplètes ;
- **valeur** : l'utilité économique tirée de l'information, seule raison de collecter.

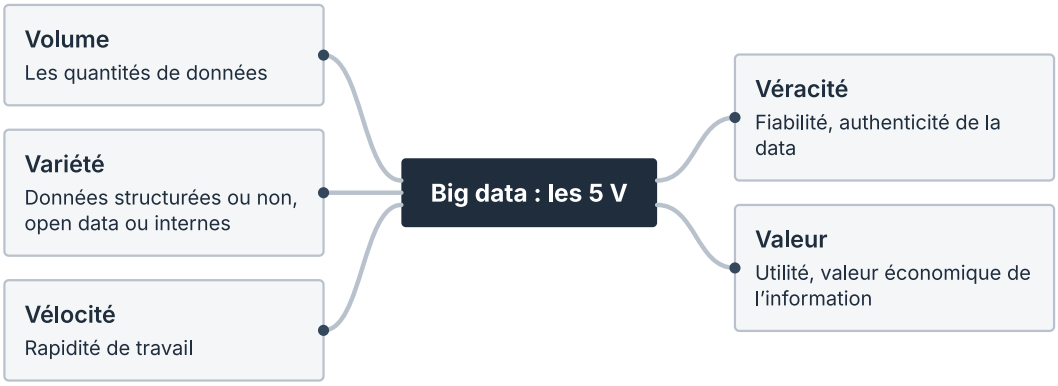



**Schéma 2.** Les 5 V qui servent à analyser le big data.

## Les types de données clients

Chaque type de données répond à une question : qui est le client, ce qu'il achète, comment il se comporte, ce qu'il pense.

| Type de données | Exemples | Comment on les collecte |
|---|---|---|
| Identité (données déclarées) | Nom, adresse, e-mail, téléphone, préférences | Création de compte, formulaire, carte de fidélité |
| Transactions | Achats, dates, montants, produits, retours | Logiciel de caisse, site marchand, ERP |
| Comportement en ligne | Pages vues, paniers abandonnés, clics dans les e-mails | Mesure d'audience, e-mailing, cookies et pixels (avec consentement quand il est requis) |
| Relation et contacts | Appels, réclamations, rendez-vous | CRM, centre de contacts, chat |
| Opinion | Avis, notes, enquêtes de satisfaction, commentaires | Enquêtes, plateformes d'avis, écoute des réseaux sociaux |
| Données externes | Profil des habitants d'une zone, fichiers loués, données ouvertes | Prestataires, INSEE, partenaires (avec une base légale) |

## Rassembler les données dans les bons outils

- **Le CRM** centralise les fiches, l'historique des contacts, les devis et les relances de l'équipe commerciale.
- **La caisse et le site marchand** enregistrent les transactions ; l'**ERP** relie ventes, stocks et facturation ; la **plateforme d'e-mailing** garde abonnements, désinscriptions et clics.
- **La plateforme de données clients** (CDP, *customer data platform*) rassemble toutes ces sources en un profil unique par client et alimente la personnalisation en temps réel : le conseiller voit tout l'historique du client, quel que soit le canal.

Pour les gros volumes, l'**entrepôt de données** (*data warehouse*) range des données structurées, préparées pour l'analyse. Le **lac de données** (*data lake*) accueille aussi les données non structurées, stockées brutes, avec des solutions comme Hadoop ou Spark. Une entreprise dont les données sont surtout des textes et des images a besoin d'un lac de données ; beaucoup combinent les deux.

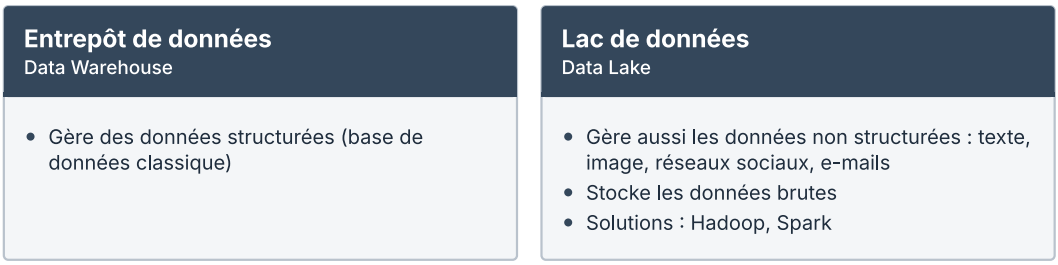



**Schéma 3.** L'entrepôt de données range des données structurées ; le lac de données accueille aussi les données non structurées, stockées brutes.

## Fiabiliser la base avant de l'exploiter

Une base peu fiable coûte cher : messages perdus, clients comptés deux fois, données gardées sans droit. Avant une campagne ou un changement d'outil, on la nettoie dans cet ordre.

1. **Inventorier** les fichiers : sources, nombre de fiches, champs.
2. **Normaliser** les champs : même format pour les noms, téléphones, codes postaux et e-mails ; listes de valeurs plutôt que texte libre.
3. **Dédoublonner** : rapprocher les fiches d'une même personne grâce à une clé (e-mail, téléphone, nom et code postal) et les fusionner.
4. **Supprimer** ce qui doit l'être : prospects dont la collecte et le dernier contact émanant d'eux remontent à plus de trois ans, demandes d'effacement, données inutiles.
5. **Vérifier et compléter** : contrôle des adresses, souvent confié à un prestataire, puis complément des champs manquants au prochain contact.

6. **Maintenir** : contrôles de saisie, mise à jour à chaque échange, tableau de bord de qualité mensuel.

Supprimer avant de vérifier évite de payer le contrôle de données que l'on n'a plus le droit de garder. Les règles de collecte et de conservation sont détaillées dans la fiche Collecte, organisation et protection des données clients.

**Les indicateurs de qualité d'un fichier**

| Indicateur | Formule | Ce qu'il révèle |
|---|---|---|
| Taux de doublons | Fiches en double ÷ fiches avant fusion × 100 | Fiches qui décrivent une personne déjà connue |
| Profils uniques | Fiches avant fusion – fiches en double | Nombre réel de personnes connues |
| Taux de complétude | Profils complets ÷ profils uniques × 100 | Profils utilisables pour toutes les actions |
| Part des données non structurées | Volume non structuré ÷ volume total × 100 | Besoin d'un lac de données |

**Exemple : le fichier de Thés de l'Aiguebrun**

Thés de l'Aiguebrun, PME fictive de 11 salariés installée à Apt (Vaucluse), vend des thés et des infusions en boutique et en ligne. Avant d'installer une plateforme de données clients, la responsable commerciale fait le point le 2 novembre 2026 (chiffres fictifs).

- Fiches avant fusion : 3 150 (caisse) + 5 480 (site) + 2 370 (lettre d'information) = 11 000.
- Doublons : 2 640, soit 2 640 ÷ 11 000 × 100 = 24,0 %.
- Profils uniques : 11 000 – 2 640 = 8 360, dont 6 270 complets : complétude = 6 270 ÷ 8 360 × 100 = 75,0 %.
- Prospects collectés avant le 2 novembre 2023 et sans contact émanant d'eux depuis cette date : 410 profils à supprimer ; il en reste 8 360 – 410 = 7 950.
- Vérification des adresses à 0,04 € HT par profil : 7 950 × 0,04 = 318,00 € HT, soit 318,00 × 1,2 = 381,60 € TTC.
- Volume stocké : 3,2 Go de données structurées et 12,8 Go de photos, d'e-mails et de messages, soit 12,8 ÷ 16 × 100 = 80,0 % de non structuré.

Une fiche sur quatre était un doublon : l'entreprise fusionne, supprime, fait vérifier les 7 950 profils restants et choisit un outil adapté aux textes et aux images.

## Analyser les données

| Type d'analyse | Question commerciale | Exemple d'usage |
|---|---|---|
| Descriptive (tableau de bord) | Que s'est-il passé ? | Ventes par canal, évolution du panier moyen ; elle résume le passé sans le prolonger |
| Segmentation | Quels clients se ressemblent ? | Méthode RFM, segments de valeur |
| Prédictive (modèle supervisé) | Que va-t-il se passer ? | Prévision des ventes, probabilité qu'un client achète ou parte |
| Exploratoire (non supervisée) et datamining | Quelles relations n'avait-on pas vues ? | Groupes de clients proches, produits achetés l'un après l'autre |

Un modèle **prédictif** apprend sur un historique dont on connaît le résultat (ventes passées selon la météo, clients partis ou restés) pour estimer un résultat futur. Un modèle **exploratoire** cherche des structures sans résultat défini à l'avance. Le **datamining** recherche automatiquement les relations entre les variables : si les acheteurs de thé vert achètent souvent une théière à leur visite suivante, on leur envoie, quelques semaines après, un conseil de préparation avec une offre sur les théières.

Le data scientist, qui maîtrise l'informatique, les statistiques et le marketing, construit les modèles ; le commercial traduit les résultats en actions.

## La démarche : de la collecte à l'action

1. **Collecter avec une base légale**, comme l'exige le RGPD (règlement européen sur la protection des données) : contrat pour une commande, consentement préalable pour prospecter un particulier par e-mail ou SMS, sauf client pour des produits analogues.
2. **Centraliser** dans un CRM ou une plateforme de données clients : un profil par client.
3. **Fiabiliser** : dédoublonner, corriger, supprimer les données trop anciennes.
4. **Analyser** : segmenter, scorer, calculer les indicateurs de valeur (panier moyen, taux d'attrition, valeur vie client), présentés dans la fiche Méthodes d'analyse des données et création de valeur client.
5. **Agir** : campagne ciblée, offre personnalisée, relance des clients à risque.
6. **Mesurer** : clics, commandes générées, clients réactivés, comparés à l'objectif.

## Exemple : segmenter ses clients avec la méthode RFM

La méthode RFM note chaque client sur trois critères tirés de ses commandes : la **récence** du dernier achat (R), la **fréquence** des achats (F) et le **montant** dépensé (M). Chez Thés de l'Aiguebrun, chaque client reçoit une note de 1 à 3 par critère, calculée sur les douze derniers mois.

**Les seuils de notation RFM de Thés de l'Aiguebrun**

| Critère | Note 3 | Note 2 | Note 1 |
|---|---|---|---|
| Récence : jours depuis le dernier achat | Moins de 30 jours | De 30 à 90 jours | Plus de 90 jours |
| Fréquence : nombre de commandes | 6 ou plus | De 3 à 5 | 1 ou 2 |
| Montant : total dépensé | 300 € ou plus | De 100 € à moins de 300 € | Moins de 100 € |

**Cinq clients notés selon la méthode RFM**

| Client et historique | Notes R-F-M | Segment et action |
|---|---|---|
| n° 1048 : dernier achat il y a 9 jours, 8 commandes, 412 € | 3-3-3 | **Champion** : reconnaissance, avant-premières, parrainage |
| n° 2210 : dernier achat il y a 21 jours, 6 commandes, 186 € | 3-3-2 | **Fidèle** : montée en gamme, programme de fidélité |
| n° 3175 : dernier achat il y a 12 jours, 1 commande, 39 € | 3-1-1 | **Nouveau** : scénario de bienvenue, incitation au deuxième achat |
| n° 0587 : dernier achat il y a 148 jours, 7 commandes, 365 € | 1-3-3 | **À risque** : relance personnalisée, offre de retour |
| n° 0932 : dernier achat il y a 260 jours, 1 commande, 27 € | 1-1-1 | **Perdu** : dernière relance, puis sortie des campagnes courantes |

Le client n° 0587 est la priorité : il a beaucoup acheté (7 commandes, 365 €) mais plus rien depuis près de cinq mois. Une relance personnalisée, avec un rappel de son thé habituel, coûte bien moins cher que de recruter un nouveau client. Le champion n° 1048 n'a pas besoin de remise : on le remercie et on s'appuie sur lui pour recruter.

## Les erreurs fréquentes

- Calculer le taux de complétude sur les fiches d'origine au lieu des profils uniques.
- Faire vérifier des données avant d'avoir supprimé celles que l'on n'a plus le droit de garder.
- Croire qu'un e-mail envoyé ou ouvert prolonge la conservation d'un prospect : seul un contact venant de lui relance le délai de trois ans.

**À l'épreuve E5**

Avec cette fiche, le candidat doit savoir :

- calculer et interpréter taux de doublons, profils uniques, taux de complétude, part de données non structurées et coût TTC d'un traitement ;
- repérer les données à supprimer, classer des données (structurées ou non, internes ou externes) et relier un défaut au bon « V » ;
- choisir l'outil (CRM, plateforme de données clients, entrepôt ou lac de données) et le type d'analyse adaptés ;
- appliquer une grille RFM et justifier, chiffres à l'appui, l'action proposée pour chaque segment.

**L'essentiel à retenir**

- La data client réunit des données structurées ou non, internes ou externes ; le big data s'analyse par cinq V : volume, variété, vélocité, véracité, valeur.
- Une base se fiabilise avant d'être exploitée : dédoublonner, supprimer ce qui doit l'être, vérifier, puis suivre doublons et complétude.
- L'outil dépend des données (CRM, plateforme de données clients, entrepôt ou lac de données) ; l'analyse peut être descriptive, prédictive ou exploratoire.
- La méthode RFM classe les clients selon la récence, la fréquence et le montant de leurs achats pour adapter l'action, dans le respect du RGPD.