

Analyse des données avancée

Chapitre 19 : Étude de cas



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Une étude complète

Étapes 1 à 3

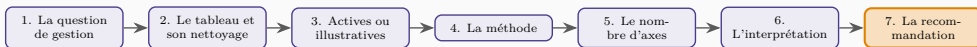
Étapes 4 à 6

Étape 7

Ce qu'il faut retenir

Une étude complète

Les sept étapes

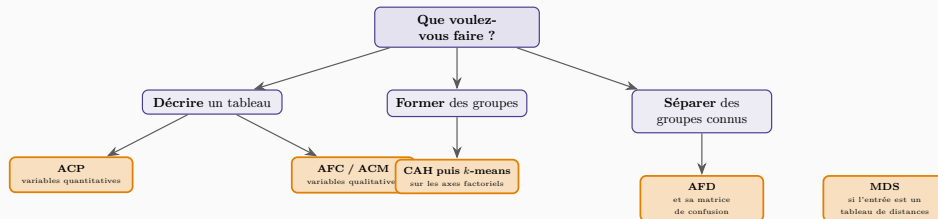


L'étape 1 décide de tout, et c'est la seule qui ne s'apprend pas dans un manuel.

« Segmenter la clientèle » n'est pas une question : c'est une méthode. La question est : *à quoi servira la segmentation — un plan de communication, un programme de fidélité, l'implantation d'un rayon ?* Elle commande le choix des variables actives, donc les axes, donc tout le reste.

L'étape 7 est celle que les mémoires oublient. Un plan factoriel n'est pas un résultat : c'est un support. Le résultat est la phrase qu'un directeur commercial peut lire et appliquer.

Quelle méthode pour quelle question



Les méthodes s'enchaînent plus souvent qu'elles ne se choisissent.

La séquence la plus fréquente en études de marché : **ACP** pour comprendre et débruiter, puis **CAH** sur les premiers axes pour trouver le nombre de segments, puis **k-means** pour consolider, enfin **AFD** pour construire une règle applicable aux clients futurs.

Quatre méthodes, un seul fichier, une seule question de gestion.

Étapes 1 à 3

Étape 1 — la commande

La direction veut lancer un **programme de fidélité** et ne peut l'adresser qu'à une partie de sa clientèle.

Traduction analytique : *quels groupes de clients existent, lesquels sont fidélisables, et à quoi sont-ils sensibles?*

Étape 2 — le tableau

300 clients, 8 notes de satisfaction, 7 variables signalétiques. Aucune valeur manquante, aucune note aberrante, aucun doublon.

Écarts-types des notes : de 2,38 à 2,40 — homogènes.

Étape 3 — la décision qui engage tout

Actives : les huit notes. C'est l'opinion que l'on veut structurer.

Illustratives : sexe, âge, CSP, fréquence, carte, panier, motif.

Pourquoi la fréquence de visite reste illustrative

C'est la variable qui intéresse le plus la direction — donc la tentation est grande de la mettre en actif.

Ce serait une faute. Elle fabriquerait alors l'axe 1, et l'on découvrirait que les clients fidèles sont fidèles. En illustratif, elle devient une **validation externe** : si elle s'aligne sur un axe construit sans elle, le résultat a une valeur.

Étapes 4 à 6

Étape 4 puis étape 5

Valeurs propres	4,535 et 1,645
Inertie retenue	$56,69 + 20,57 = 77,26 \%$
Coude, Kaiser, cumul	deux axes , les trois concordent
Axe 1	satisfaction d'ensemble
Axe 2	offre contre service

La validation externe fonctionne

Projetée en illustratif, la fréquence de visite s'étage sur l'axe 1 : $-1,32, +0,06, +1,41$.

Un axe construit sans elle la retrouve. C'est la meilleure garantie de solidité dont on dispose en analyse des données.

Étape 6 — CAH de Ward, puis consolidation

Coupure	4 classes (saut de 3,62 à 9,74 %)
Inertie interclasse	74,93 %, puis 76,50 %
Clients déplacés	26 sur 300
Confirmation	<i>k</i> -means à quatre départs : 76,54 %

Trois signaux concordants

Un saut d'inertie net, quatre classes qui s'interprètent, deux algorithmes qui convergent.

Aucun des trois ne prouve quoi que ce soit isolément. Ensemble, ils rendent la segmentation défendable — c'est le mieux que la méthode permette.

Les quatre segments

	Les conquis	Sensibles à l'offre	Attachés au service	Les déçus
effectif	61 (20,3 %)	82 (27,3 %)	59 (19,7 %)	98 (32,7 %)
prix	12,5	13,7	9,8	9,8
choix	15,2	15,0	11,9	11,2
qualité	16,0	16,1	13,0	11,9
accueil	16,4	13,2	15,9	11,7
conseil	15,1	12,4	14,7	10,6
panier moyen	454 dh	394 dh	439 dh	370 dh
carte de fidélité	75 %	49 %	71 %	40 %
visite hebdomadaire	69 %	18 %	64 %	6 %

Le tableau qui transforme une partition en décision.

Les **conquis** et les **attachés au service** reviennent chaque semaine et dépensent le plus — pourtant les seconds notent le prix 9,8 sur 20. Ils reviennent malgré l'offre, pour le personnel.

Les **sensibles à l'offre** jugent la marchandise excellente et ne viennent que 18 % du temps chaque semaine. **Un client satisfait du produit mais pas du service ne devient pas fidèle.**

C'est le résultat le plus actionnable de l'étude, et aucune moyenne générale ne l'aurait montré : sur l'ensemble des 300 clients, prix et accueil ne sont pas corrélés du tout.

Étape 7

Le classement des critères

Corrélations avec l'axe discriminant de la fréquence de visite : conseil 0,907, accueil 0,899, rapidité 0,867 — et prix 0,110.

Taux de bon classement : 67,0 %, 65,3 % en validation croisée, contre 33,3 % au hasard.

Le fait central de l'étude

Les *attachés au service* notent le prix 9,8 sur 20 et reviennent chaque semaine à 64 %.

Les *sensibles à l'offre* notent le prix 13,7 et ne reviennent qu'à 18 %.

Deux méthodes indépendantes — segmentation et discrimination — aboutissent à la même conclusion : la fidélité se joue sur le service.

Trois phrases pour un comité de direction

1. Le programme de fidélité doit viser en priorité les **sensibles à l'offre** — 82 clients, satisfaits de la marchandise, déçus par le service, et qui ne reviennent pas.
2. Le levier n'est **pas le prix** : il n'a aucun pouvoir discriminant. Ce sont le conseil, l'accueil et la rapidité en caisse.
3. Une segmentation par âge ou par CSP serait sans fondement : ces variables **ne structurent rien** dans ce fichier.

Ce qui rend ces phrases utilisables

Chacune nomme un groupe, chiffre son effectif, et désigne une action.

Aucune ne comporte de plan factoriel. Le plan a servi à trouver; il n'a pas à figurer dans la conclusion.

La partie que les mémoires oublient

1. **Aucune causalité.** Le sens de la relation entre service et fidélité n'est pas établi.
2. **Une seule enseigne, une seule date.** Rien ne se généralise.
3. **Trois cents clients interrogés** — pas trois cents clients tirés au sort.
4. **Aucun test.** Ni les axes, ni les segments ne sont « significatifs » : le mot n'a pas de sens ici.

Pourquoi cette section est la plus importante

Un lecteur compétent la lit en premier : elle indique si l'auteur a compris ce qu'il a fait.

Une segmentation présentée sans ses limites n'est pas une segmentation sans limites : c'est une segmentation dont on ignore les limites.

Ce qu'il faut retenir

Six points

1. **Sept étapes**, de la question de gestion à la recommandation.
2. Mettre la variable cible en **illustratif** : elle devient une validation externe.
3. Enchaînement standard : **ACP**, puis **CAH**, puis **k-means**, puis **AFD**.
4. Trois signaux de solidité : saut d'inertie net, classes interprétables, méthodes concordantes.
5. Le résultat : **la fidélité tient au service** — le prix ne discrimine pas ($r = 0,110$).
6. La conclusion se rédige **sans plan factoriel**, et avec ses limites.

La suite

Tout a été calculé pas à pas, pour que chaque nombre soit vérifiable.

Le chapitre 20 refait l'ensemble en quelques lignes de code — et montre qu'aucune sortie de logiciel ne contient rien que nous n'ayons établi.