

## A. Segmentation des clients à l'aide de l'IA

Les attentes des consommateurs évoluent rapidement et la capacité à comprendre et à répondre à leurs besoins est un avantage concurrentiel majeur. La **segmentation**, pierre angulaire d'une stratégie marketing efficace, permet de diviser un marché en groupes homogènes afin de personnaliser les messages, les offres et les expériences client. L'essor des technologies d'intelligence artificielle permet de franchir un nouveau cap, favorisant une analyse plus aisée et plus fine des données. Ce chapitre explore comment l'IA peut être mise à profit pour améliorer la segmentation et personnaliser les campagnes marketing, offrant ainsi des opportunités nouvelles pour optimiser les performances et renforcer l'engagement des clients.

### 1. Critères de segmentation

La segmentation repose sur des critères variés qui permettent d'identifier des groupes homogènes. Voici un décryptage des principaux types de critères :

**Socio-démographique** : ce critère inclut des éléments tels que l'âge, le sexe, le revenu, le niveau d'éducation ou la situation familiale. Bien qu'ils soient facilement accessibles et souvent utilisés, ces critères risquent de réduire les clients à des stéréotypes et de conduire à des stratégies marketing moins inclusives. Il est donc essentiel de les compléter par d'autres approches pour éviter toute discrimination.

**Géographique** : la localisation des clients permet de cibler des besoins spécifiques en fonction de leur environnement. Par exemple, une entreprise peut adapter son offre selon que les clients vivent en milieu urbain ou rural, ou encore en fonction du climat, de l'ensoleillement ou de la pluviométrie. Ce critère est utile mais ne doit pas être isolé, car il peut manquer de profondeur en termes de comportement ou de motivations.

**Psychographique** : ce type de segmentation se concentre sur les valeurs, les croyances, les styles de vie et les personnalités des consommateurs. Il offre une compréhension plus nuancée des motivations et des aspirations des clients, permettant de créer des messages et des offres qui résonnent avec eux sur le plan émotionnel.

**Comportementale** : ce critère repose sur les actions des clients, telles que leurs habitudes d'achat, la fréquence d'utilisation d'un produit ou leur réaction à des campagnes marketing précédentes. En comprenant ce que les clients font, une entreprise peut prédire leurs futurs comportements et adapter ses stratégies en conséquence. Cette approche, centrée sur les faits, est particulièrement pertinente dans un monde axé sur les données.

**Par avantages recherchés** : ici, l'accent est mis sur les besoins ou les problèmes que les clients souhaitent résoudre. Par exemple, certains consommateurs choisissent un produit pour son prix, tandis que d'autres le préfèrent pour sa qualité ou son aspect innovant. Ce critère est particulièrement puissant, car il permet de comprendre directement les attentes des clients sans les réduire à leurs caractéristiques sociodémographiques, évitant ainsi les biais discriminatoires. Cela en fait une approche à privilégier dans les stratégies marketing modernes.

## 2. Outils d'IA pour une segmentation plus fine et plus efficace

Les outils d'intelligence artificielle (IA) apportent une valeur ajoutée significative à la segmentation en exploitant des volumes massifs de données pour identifier des patterns complexes. Voici une vue d'ensemble de leur fonctionnement, des technologies impliquées ainsi que des avantages et des limites.

### a. Comment cela fonctionne ?

L'IA utilise des algorithmes d'apprentissage automatique (machine learning) pour analyser des données structurées (démographiques, ventes) et non structurées (commentaires clients, posts sur les réseaux sociaux). Par exemple, des modèles comme les réseaux neuronaux identifient des corrélations complexes entre les comportements d'achat et les préférences des consommateurs.

### b. Technologies clés

- ▶ **Traitement du langage naturel (NLP)** : analyse des textes, comme les avis clients ou les commentaires sur les réseaux sociaux, pour capturer des indices sur les préférences psychographiques.
- ▶ **Clustering non supervisé** : les algorithmes comme K-means regroupent les clients ayant des comportements ou besoins similaires.
- ▶ **Modèles prédictifs** : ces outils prédisent les comportements futurs basés sur les données passées.

### c. Avantages

Les outils d'IA permettent une segmentation plus fine et précise que les approches traditionnelles. En identifiant des micro-segments à travers l'analyse de volumes importants de données, ils ouvrent la voie à des campagnes hyper-personnalisées. L'automatisation facilite une mise à jour en temps réel, permettant aux entreprises de réagir rapidement aux changements dans les comportements des consommateurs. De plus, ces technologies permettent une meilleure compréhension des besoins clients et une optimisation des ressources marketing en se concentrant sur des segments à forte valeur ajoutée.

### d. Inconvénients

Les principaux inconvénients sont :

- ▶ **Coût élevé** : l'implémentation de ces technologies peut demander un investissement initial important.
- ▶ **Dépendance aux données** : une qualité insuffisante des données dégradera la fiabilité des segments produits.
- ▶ **Complexité** : la mise en place de modèles avancés requiert des compétences techniques spécifiques.

Bien que la perception initiale des outils d'IA puisse sembler complexe, des outils tels que ChatGPT et les plateformes d'analyse intuitive ont simplifié leur utilisation. Le coût initial peut constituer un obstacle pour certaines entreprises, mais cet investissement est souvent compensé par l'amélioration des performances marketing.

Par ailleurs, la qualité des données est un facteur déterminant : des données incomplètes ou biaisées peuvent réduire l'efficacité des modèles d'IA. Cependant, avec un accompagnement minimal par un *data analyst* ou une formation de base, ces inconvénients deviennent largement surmontables.

### e. Outils en ligne disponibles

Un certain nombre d'outils en ligne est disponible. Nous pouvons ainsi citer :

- ▶ **Google Analytics** : gratuit pour les fonctionnalités de base. Une version payante (Google Analytics 360) est disponible pour des analyses plus poussées, à partir de 12 500 \$ par an.
- ▶ **Salesforce** : offre des solutions personnalisées selon la taille de l'entreprise, avec des coûts débutant autour de 25 \$ par utilisateur et par mois pour les fonctionnalités de base, et pouvant atteindre plusieurs milliers de dollars pour des solutions avancées intégrant l'IA.
- ▶ **HubSpot** : propose des outils gratuits pour démarrer, mais les plans payants débutent à environ 45 \$ par mois et par utilisateur. Ils permettent de créer des scripts et des stratégies personnalisées basées sur des segments précis.

- **ChatGPT et autres modèles génératifs** : accessible via des abonnements mensuels comme ChatGPT Plus (20 \$/mois), ou via des API facturées à l'utilisation, selon les volumes et la complexité des tâches. Ils permettent de créer des scripts et des stratégies personnalisées basées sur des segments précis.

En exploitant ces outils, les entreprises peuvent réaliser une segmentation avancée, gagner en efficacité et réduire les approximations.

Les outils d'IA permettent d'aller au-delà des méthodes classiques en utilisant des techniques avancées pour analyser des données complexes :

- **Apprentissage automatique** : les modèles d'IA peuvent identifier des segments qui ne seraient pas visibles à l'œil humain.
- **Traitement du langage naturel (NLP)** : analyse des avis clients et des discussions en ligne pour déceler des tendances psychographiques.
- **Plateformes d'analyse avancée** : des outils comme Google Analytics, Salesforce, ou des modèles comme ChatGPT permettent de traiter de grandes quantités de données pour en tirer des segments précis.

### 3. Pas à pas : créer des segments personnalisés avec ChatGPT

Pour illustrer concrètement l'utilisation de ChatGPT dans la segmentation, nous allons travailler sur un exemple basé sur un jeu de données fictif contenant 1 000 lignes (disponible dans le fichier fourni 00 dataset\_ch4.xlsx).

#### a. Présentation de la base de données

La base de données contient les informations suivantes :

**ID\_Client** : un identifiant unique pour chaque client.

**Âge** : l'âge des clients (18 à 65 ans).

**Genre** : le genre des clients, comprenant les valeurs :

- Homme
- Femme
- Non-binaire

**Revenu** : le revenu annuel des clients en euros, allant de 20 000 à 120 000 €.

**Localisation** : le type de lieu où réside le client :

- Urbain
- Périurbain
- Rural

**Style\_de\_vie** : les préférences ou modes de vie des clients, comme :

- ▶ Actif
- ▶ Détendu
- ▶ Aventurier
- ▶ Technophile

**Fréquence\_d'achat** : la régularité avec laquelle les clients effectuent des achats :

- ▶ Rarement
- ▶ Occasionnellement
- ▶ Fréquemment

**Canal\_préféré** : le canal d'achat privilégié par les clients :

- ▶ En ligne
- ▶ En magasin
- ▶ Hybride (combinaison des deux)

**Avantage\_recherché** : les bénéfices recherchés par les clients lors de leurs achats :

- ▶ Prix
- ▶ Qualité
- ▶ Commodité
- ▶ Durabilité

ID_Client	Âge	Genre	Revenu	Localisation	Style_de_vie	Fréquence_d'achat	Canal_préféré	Avantage_recherché
1	34	Femme	45 000	Urbain	Actif	Fréquemment	En ligne	Durabilité

*Exemple de ligne de la base de données*

**b. Présentation de la démarche**

Dans la section suivante, nous allons guider le lecteur à travers une démarche complète de segmentation des clients à l'aide d'une approche quantitative mêlant Classification ascendante hiérarchique (CAH), K-means et analyses statistiques. Cette méthode permettra de regrouper les clients selon leurs caractéristiques comportementales, géographiques, psychographiques et les avantages recherchés, tout en caractérisant ces segments à l'aide de variables telles que l'âge et le genre.

L'intérêt de cette démarche réside dans sa capacité à fournir des segments précis et exploitables pour des actions marketing ciblées. En suivant ce processus, les entreprises peuvent mieux comprendre les besoins de leurs clients, adapter leurs offres et messages, et maximiser l'efficacité de leurs campagnes.

### c. Étape 1 : préparer les données

#### Objectif

Créer une base de données prête à être utilisée pour des analyses avancées. Cela inclut la sélection des variables pertinentes, la gestion des valeurs manquantes et la normalisation.

#### Méthode

Si vous n'êtes pas sûr de ce qu'il est nécessaire de préparer dans votre base de données, voici un prompt pour ChatGPT :

« Voici un extrait d'une base de données :

*[Insérer la base de données en pièce jointe].*

*Analyse cette base et identifie :*

- Les types de variables (catégorielles, numériques, texte, etc.).
- Les éventuelles valeurs manquantes ou anomalies.
- Les prétraitements nécessaires pour effectuer une segmentation ou des analyses statistiques, comme le clustering.
- Si des transformations spécifiques (comme l'encodage ou la normalisation) sont recommandées. Explique pourquoi chaque étape de préparation est nécessaire et comment la réaliser. »

Une fois la réponse générée, demandez-lui d'appliquer les préparations nécessaires à la base de données et de la générer à nouveau. S'il commet des erreurs ou des oublis, demandez-lui de générer une nouvelle base de données pour chaque étape. Ainsi, vous pourrez mieux contrôler ses modifications et il risque moins de s'emmêler les pinceaux.

Dans notre exemple, trois étapes sont nécessaires :

- Nettoyage des données : supprimer les doublons et vérifier qu'il n'y a pas de valeurs manquantes critiques.
- Encodage des variables catégorielles : transformer les textes (ex. : « Urbain ») en valeurs numériques pour faciliter les calculs.
- Normalisation : standardiser les variables pour leur donner un poids équivalent dans les calculs.

Une nouvelle base de données disponible en téléchargement, **01 dataset\_prepared**, est générée.

#### d. Étape 2 : définir le bon nombre de segments (clusters)

##### Objectif

La Classification ascendante hiérarchique (CAH) est une méthode exploratoire utilisée pour déterminer le **nombre optimal de clusters** dans un ensemble de données. Elle regroupe les individus ou les objets similaires en fonction des variables sélectionnées, selon une approche hiérarchique.

##### Prompt pour ChatGPT

Donnez le prompt suivant à ChatGPT :

*« Je souhaite réaliser une Classification ascendante hiérarchique (CAH) sur une base de données client pour identifier le nombre optimal de clusters avant d'appliquer un K-means. Voici les variables que je veux utiliser :*

*Comportementales : Canal\_préféré, Fréquence\_d'achat.*

*Avantages recherchés : Avantage\_recherché.*

*Géographiques : Localisation.*

*Psychographiques : Style\_de\_vie.*

*Calcule une matrice de distances en utilisant la méthode de Ward et affiche un dendrogramme pour visualiser les regroupements hiérarchiques. Explique comment choisir le nombre de clusters optimal à partir du dendrogramme. »*

##### Réponse de ChatGPT

##### Classification ascendante hiérarchique (CAH)

Voici l'application du prompt et la réponse correspondante :

##### Données analysées

##### ► Variables utilisées :

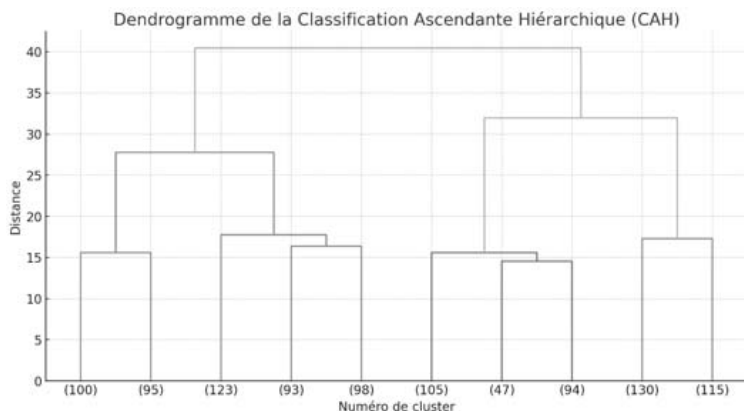
- Comportementales : Canal\_préféré, Fréquence\_d'achat.
- Avantages recherchés : Avantage\_recherché.
- Géographiques : localisation.
- Psychographiques : Style\_de\_vie.

##### ► Méthode utilisée :

- Calcul d'une matrice de distances selon la méthode de Ward, qui minimise la variance intra-cluster.

### Résultat : Dendrogramme

J'ai calculé et affiché le dendrogramme ci-dessous à partir des données fournies :



### Interprétation du dendrogramme

Pour déterminer le nombre optimal de clusters :

- ▶ Repérez la plus grande coupure verticale dans le dendrogramme : cette coupure indique les regroupements les plus significatifs.
- ▶ Nombre de clusters suggérés :
  - Si une large coupure est visible entre 3 et 4 branches, cela suggère 3 clusters comme nombre optimal.
  - Vous pouvez choisir un autre nombre si la logique métier le justifie.

### Proposition pour la suite

Avec cette interprétation, le nombre de clusters optimal peut être utilisé pour appliquer un algorithme de K-means afin d'attribuer chaque client à un groupe. Souhaitez-vous :

- ▶ Poursuivre avec un K-means en utilisant les 3 clusters suggérés ?
- ▶ Modifier les variables ou les paramètres pour affiner l'analyse ?

Nous préférons créer 4 clusters (c'est-à-dire 4 segments) pour être un peu plus précis. Nous gardons donc cette information pour l'étape suivante.



### e. Étape 3 : segmentation du portefeuille clients

#### Objectif

Le **K-means** est un algorithme de clustering qui regroupe les individus en clusters en minimisant la variance intra-cluster. L'objectif ici est d'attribuer chaque client à l'un des clusters identifiés grâce à la CAH, en se basant sur les variables comportementales, psychographiques, géographiques et avantages recherchés.

**Pourquoi utiliser le K-means après la CAH ?**

**Validation des clusters** : le K-means permet d'affiner les clusters obtenus avec la CAH et de garantir une séparation optimale des groupes.

**Assignation précise** : chaque individu est associé à un cluster spécifique, ce qui est utile pour les analyses et les actions marketing.

#### Prompt pour ChatGPT

« Utilise l'algorithme de K-means pour regrouper les clients en clusters. Utilise les variables suivantes :

- ▶ Comportementales : Canal\_préfééré, Fréquence\_d'achat.
- ▶ Avantages recherchés : Avantage\_recherché.
- ▶ Géographiques : Localisation.
- ▶ Psychographiques : Style\_de\_vie.

*Le nombre de clusters optimal est 4. Fournis les résultats du regroupement et des caractéristiques descriptives de chaque cluster (taille, moyenne des variables utilisées). »*