K-MEANS

CLUSTERING

CLUSTERING

HIÉRARCHIQUE

SEGMENTATION DE LA VALEUR À VIE DU CLIENT

(CLV - CUSTOMER LIFETIME VALUE)

MÉTHODES DE CLUSTERING (K-MEANS, HIÉRARCHIQUE)

Le clustering, ou segmentation non supervisée, est une technique statistique qui permet de regrouper des individus ou des objets ayant des caractéristiques similaires en clusters (ou groupes). Les individus dans un même cluster partagent des similarités, alors que ceux de clusters différents présentent des différences.

- **Objectifs du Clustering :**
  - Identifier des groupes distincts au sein d'une base de données sans connaissance préalable des catégories.
  - Faciliter la prise de décisions stratégiques en fonction des segments identifiés.
  - Dans le secteur du luxe, le clustering permet de mieux comprendre les types de clientèle et leurs attentes spécifiques.
-

MÉTHODES DE CLUSTERING (K-MEANS, HIÉRARCHIQUE)

Applications dans le Luxe :

- Créer des segments de clients pour des expériences personnalisées.
- Optimiser les actions de fidélisation en ciblant des groupes spécifiques.
- Mieux comprendre les comportements d'achat pour ajuster les stratégies marketing.

MÉTHODES DE CLUSTERING (K-MEANS, HIÉRARCHIQUE)

- Le clustering regroupe les clients en segments en utilisant des algorithmes qui analysent plusieurs variables simultanément (pas seulement Récence, Fréquence, et Monétaire).

- Les méthodes de clustering les plus couramment utilisées sont le **K-means** et le **clustering hiérarchique**. Chacune a des avantages et des limites, et le choix de la méthode dépend des caractéristiques des données et des objectifs de segmentation.

K-MEANS

CLUSTERING

K-MEANS CLUSTERING

Le K-means est une méthode de clustering partitionnel qui divise les données en un nombre fixe de clusters définis à l'avance (k clusters). Cette méthode est largement utilisée pour sa simplicité et sa rapidité.

Principe de Fonctionnement :

- Sélection de k points initiaux (centroïdes).
- Affectation de chaque point aux centroïdes les plus proches.
- Mise à jour des centroïdes jusqu'à stabilisation des groupes.

Avantages :
Rapidité et simplicité pour des groupes bien définis.

Inconvénients :
Sensible aux valeurs initiales et nécessite une connaissance préalable du nombre de clusters (k).

K-MEANS CLUSTERING

Paramètres Clés :

- k (nombre de clusters) : Doit être défini par l'utilisateur en amont. Un choix judicieux de k est crucial pour des segments significatifs.
- Centroïdes : Représentent le centre de chaque cluster et influencent la formation des groupes.

K-MEANS CLUSTERING

- **Étapes du K-means :**
  - Choisir un nombre de clusters (k).
  - Initialiser aléatoirement k centroïdes (points centraux pour chaque cluster).
  - Assigner chaque point de données au centroïde le plus proche.
  - Calculer les nouveaux centroïdes en prenant la moyenne de tous les points assignés à chaque cluster.
  - Répéter les étapes 3 et 4 jusqu'à ce que les centroïdes se stabilisent (n'évoluent plus).

K-MEANS CLUSTERING

K-means dans le Luxe

Un magasin de luxe peut utiliser le K-means pour segmenter ses clients en fonction de caractéristiques telles que la fréquence d'achat, le montant dépensé, et les catégories de produits préférées. Cela permettrait de cibler les grands acheteurs ou les clients réguliers avec des offres spéciales.

EXEMPLE PRATIQUE

Imaginons une base de données* contenant des informations de base sur des clients. Les variables sont : Id Client, âge, fréquence d'achat, montant dépensé (€), Type de produit préféré.

ID Client	Âge	Fréquence d'Achat	Montant Dépensé (en €)	Type de Produit Préféré
1	25	10	1500	ACCESSOIRES
2	45	5	3000	VÊTEMENTS
3	35	20	500	SOIN
4	50	7	2000	ACCESSOIRES
5	28	15	800	VÊTEMENTS
6	40	3	4000	SOIN
7	33	8	2500	ACCESSOIRES
8	52	12	1200	VÊTEMENTS
9	29	6	3500	SOIN
10	48	2	4500	ACCESSOIRES

*Base de Données (Exemple Simplifié)

ACCÉS AU CONTENU, EXERCISES ET FICHIERS

https://expertlab.stepsconsulting.fr/course/view.php?id=173

EXEMPLE PRATIQUE

1 **Préparation des Données :**

Les variables numériques (Âge, Fréquence d'Achat, Montant Dépensé) sont directement utilisables pour le clustering.
La variable catégorielle "Type de Produit Préféré" peut être transformée en variables numériques pour le clustering, par exemple :

Accessoires = 0, Vêtements = 1, Soin = 2

2 **Choix du Nombre de Clusters (k) :**

Dans cet exemple, on choisit k=3 pour tenter de diviser les clients en trois segments basés sur leur comportement d'achat.

3 **Exécution du K-means :**

Avec un outil comme Excel, Python ou R, nous appliquons l'algorithme K-means en utilisant les variables normalisées : Âge, Fréquence d'Achat, Montant Dépensé et Type de Produit Préféré.

K-MEANS CLUSTERING

Interprétation des Clusters

- Cluster 1 : Clients plus jeunes, avec une fréquence d'achat moyenne à élevée et un montant dépensé modéré. Ce groupe pourrait représenter des clients intéressés par des produits plus abordables ou avec des achats fréquents mais de montant limité.
- Cluster 2 : Clients d'âge moyen avec une fréquence d'achat modérée et un budget élevé. Ces clients pourraient être intéressés par des produits de haute qualité mais n'achètent pas très fréquemment.
- Cluster 3 : Clients plus âgés, avec une fréquence d'achat faible et un montant dépensé très élevé. Ce groupe pourrait représenter des clients recherchant des produits de luxe haut de gamme et exclusifs.

EXEMPLE PRATIQUE

Analyse des Résultats

Après avoir exécuté l'algorithme K-means, les clients seront assignés à l'un des trois clusters. Voici un exemple de résultat possible :

ID Client	Âge	Fréquence d'Achat	Montant Dépensé (en €)	Type de Produit Préféré	Cluster
1	25	10	1500	ACCESSOIRES	1
2	45	5	3000	VÊTEMENTS	2
3	35	20	500	SOIN	1
4	50	7	2000	ACCESSOIRES	2
5	28	15	800	VÊTEMENTS	1
6	40	3	4000	SOIN	3
7	33	8	2500	ACCESSOIRES	2
8	52	12	1200	VÊTEMENTS	1
9	29	6	3500	SOIN	3
10	48	2	4500	ACCESSOIRES	3

*Base de Données (Exemple Simplifié)

K-MEANS CLUSTERING

EXERCISE

1. **Téléchargez la Base de Données**
2. **Préparez les Données pour le K-means Clustering**
3. **Appliquez le K-means Clustering**
4. **Analysez et Interprétez les Résultats**

EN GROUPES, PREPAREZ UN SLIDE À PRESENTER À TOUTE LA CLASSE.

K-MEANS CLUSTERING

Remplissez le tableau suivant en fonction de vos résultats pour chaque cluster :

CLUSTER	ÂGE MOYEN	DÉPENSE MOYENNE (EN €)	FRÉQUENCE MOYENNE
1			
2			
3			

K-MEANS CLUSTERING

Questions d'Analyse :

- Caractéristiques des Segments : Pour chaque cluster, quels sont les caractéristiques principales ? (ex. : jeunes avec haute fréquence et basse dépense, etc.)
- Implications pour l'Expérience Client :
  - Quel type de service ou d'offre pourrait être pertinent pour chaque segment ?
  - Comment pourriez-vous personnaliser l'expérience client pour répondre aux besoins spécifiques de chaque groupe ?

Rédigez une Brève Conclusion :

- Résumez vos observations sur les segments identifiés et proposez deux actions possibles pour optimiser l'expérience client pour chaque segment.

CLUSTERING

HIÉRARCHIQUE

CLUSTERING HIÉRARCHIQUE

Le clustering hiérarchique est une méthode de segmentation non supervisée qui regroupe les individus ou objets en une hiérarchie de clusters.

Il existe deux principales approches pour le clustering hiérarchique :

Agglomératif

Divisif

CLUSTERING HIÉRARCHIQUE

Le clustering hiérarchique est une méthode de segmentation non supervisée qui regroupe les individus ou objets en une hiérarchie de clusters. Contrairement au K-means, il ne nécessite pas de définir le nombre de clusters à l'avance et crée un regroupement visuel, sous forme d'arbre, appelé dendrogramme.

Il existe deux principales approches pour le clustering hiérarchique :

Agglomératif (Ascendant) : Chaque donnée commence dans son propre cluster, puis des clusters sont fusionnés les uns avec les autres jusqu'à ce qu'il n'y ait plus qu'un seul cluster.

Divisif (Descendant) : L'ensemble des données commence dans un seul cluster qui est progressivement divisé en sous-clusters jusqu'à ce que chaque point de donnée soit dans un cluster unique.

CLUSTERING HIÉRARCHIQUE

Calcul des Distances

Fusion de Clusters

Répétition du Processus

Dendrogramme

CLUSTERING HIÉRARCHIQUE

Le clustering hiérarchique fonctionne en suivant ces étapes :

1. **Calcul des Distances** : Une mesure de distance est calculée entre chaque paire de données (ex. : distance euclidienne). Ces distances servent à déterminer la similarité entre les points.
2. **Fusion de Clusters** : Les deux clusters les plus proches (ayant la plus petite distance) sont fusionnés pour former un nouveau cluster.
3. **Répétition du Processus** : Cette fusion se poursuit de manière répétée jusqu'à ce que toutes les données soient regroupées dans un seul cluster global.
4. **Dendrogramme** : Le processus est représenté visuellement dans un dendrogramme, où chaque branche représente une fusion de clusters.

CLUSTERING HIÉRARCHIQUE

Interprétation du Dendrogramme

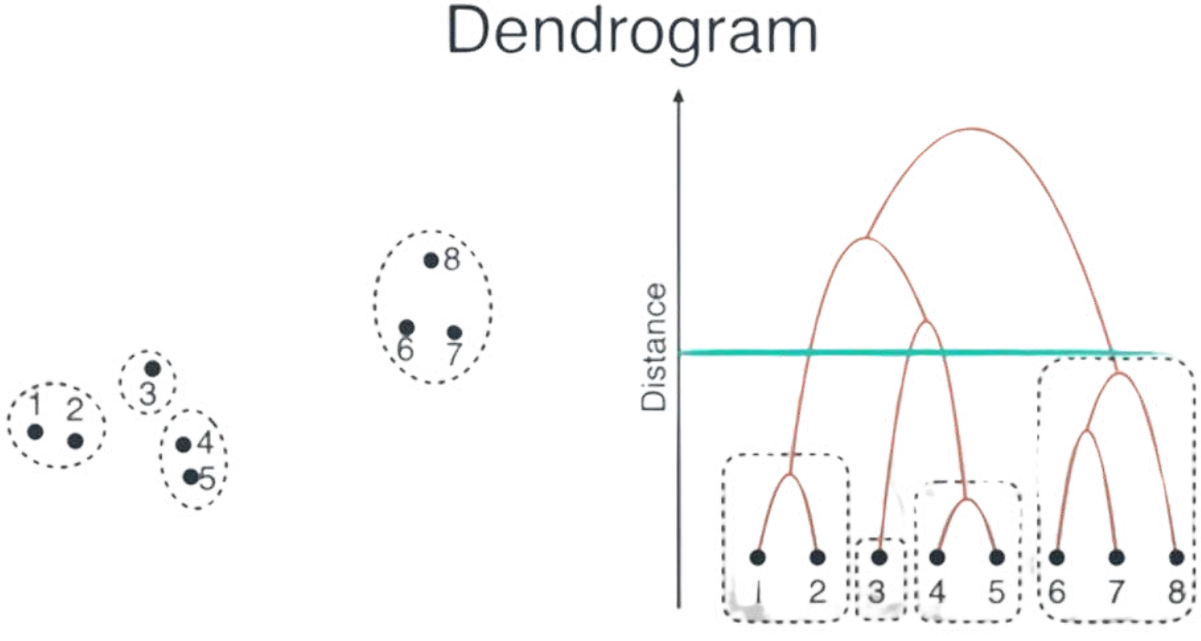
Le dendrogramme est un diagramme arborescent qui montre la manière dont les clusters sont formés. Voici quelques éléments pour bien l'interpréter :

- **Branches** : Chaque branche du dendrogramme représente un cluster.
- **Hauteur** : La hauteur à laquelle deux clusters sont fusionnés dans le dendrogramme indique la distance ou la similarité entre eux. Plus la fusion est haute, moins les clusters sont similaires.
- **Découpage du Dendrogramme** : Pour déterminer le nombre de clusters final, on peut découper le dendrogramme à un certain niveau de hauteur. Ce découpage horizontal définit les clusters finaux.

CLUSTERING HIÉRARCHIQUE

Exemple de Lecture

Imaginons un dendrogramme où, en coupant au niveau d'une hauteur spécifique, nous obtenons 4 clusters principaux. Cela signifie que les individus de chaque groupe ont des similarités significatives entre eux, mais diffèrent des individus des autres groupes.



CLUSTERING HIÉRARCHIQUE

EXERCISE

1. **Téléchargez la Base de Données**
2. **Préparez les Données**
3. **Appliquez le Clustering Hiérarchique**
4. **Analysez le Dendrogramme**

EN GROUPES, PREPAREZ UN SLIDE À PRESENTER À TOUTE LA CLASSE.

CLUSTERING HIÉRARCHIQUE

Interprétation et Analyse des Groupes

Remplissez le tableau suivant pour chaque cluster principal :

Cluster	Âge Moyen	Dépense Moyenne (en €)	Fréquence Moyenne	Satisfaction Moyenne
1				
2				
3				

Description des Segments : Décrivez brièvement chaque segment en fonction des caractéristiques dominantes (âge, dépense, fréquence d'achat, satisfaction, etc.).

CLUSTERING HIÉRARCHIQUE

Questions de Réflexion

Type de Stratégie pour Chaque Segment :

- Comment adapteriez-vous l'expérience client pour chaque segment ? (Ex. : programmes de fidélisation, offres de produits exclusifs, services personnalisés).

Identifiez les Sous-groupes :

- Quels sous-groupes avez-vous repérés dans chaque cluster ?
- Comment ces sous-groupes influenceraient-ils une personnalisation encore plus ciblée ?

Rédigez une Conclusion

- Résumez brièvement les segments identifiés et proposez deux à trois actions stratégiques pour améliorer l'expérience client en fonction des insights obtenus.

COMPARAISON DES MÉTHODES

K-MEANS VS. CLUSTERING HIÉRARCHIQUE

Critère	K-means	Clustering Hiérarchique
Nombre de clusters	Fixé à l'avance	Déterminé selon la hiérarchie
Vitesse de Calcul	Rapide, adapté aux grandes bases	Plus lent, computationnellement intensif
Utilisation	Grands ensembles homogènes	Segmentation détaillée avec hiérarchie
Visualisation	Pas de hiérarchie visuelle	Dendrogramme pour visualiser les clusters
Exigences	Choix préalable du nombre de k	Pas de paramétrage du nombre de clusters

COMPARAISON DES MÉTHODES

K-MEANS VS. CLUSTERING HIÉRARCHIQUE

Critères de Choix entre K-means et Clustering Hiérarchique

Le choix de la méthode dépend de plusieurs facteurs :

- **Connaissance du Nombre de Clusters :** Si vous connaissez le nombre approximatif de segments, le K-means peut être efficace. Sinon, le clustering hiérarchique permet de déterminer ce nombre en analysant le dendrogramme.
- **Volume de Données :** Pour de grandes bases de données, le K-means est recommandé pour sa rapidité. Le clustering hiérarchique est plus adapté pour des bases de taille moyenne où des relations détaillées sont importantes.
- **Nature des Segments :** Le clustering hiérarchique est pertinent pour explorer des sous-groupes et des structures internes complexes, tandis que le K-means convient aux segments bien définis.