








projet data mining

 Class	 data mining
 Date	@April 16, 2025
 Days Remaining	D-5
 done	<input type="checkbox"/>

Rapport : Analyse et segmentation des comportements d'achat des consommateurs pour l'optimisation des stratégies de marketing

1. Introduction

Dans un environnement commercial dynamique et compétitif, comprendre les comportements d'achat des consommateurs est essentiel pour optimiser les stratégies de marketing. Ce projet vise à exploiter les données transactionnelles d'un détaillant en ligne basé au Royaume-Uni pour identifier des segments de clients et découvrir des règles d'association pouvant améliorer les actions marketing de l'entreprise.

2. Jeu de données

Le jeu de données utilisé provient du **UCI Machine Learning Repository**, intitulé **Online Retail Dataset**, qui contient environ 500 000 transactions effectuées entre le 01/12/2010 et le 09/12/2011. L'entreprise vend principalement des produits de cadeaux pour diverses occasions.

Variables principales :

- **InvoiceNo** : Numéro de facture
- **StockCode** : Code produit
- **Description** : Description du produit

- **Quantity** : Quantité de produits achetés
- **InvoiceDate** : Date de la facture
- **UnitPrice** : Prix unitaire
- **CustomerID** : Identifiant client
- **Country** : Pays de l'acheteur

3. Objectifs

L'objectif principal de ce projet est d'exploiter les techniques de data mining pour segmenter les clients et découvrir des règles d'achat fréquentes afin d'optimiser les stratégies marketing. Plus précisément, nous avons abordé les étapes suivantes :

1. **Segmentation des clients** à l'aide des techniques de clustering : K-means et Clustering Hiérarchique (CAH).
2. **Extraction de règles d'association** pour identifier des produits fréquemment achetés ensemble.
3. **Recommandations stratégiques** pour améliorer les actions marketing de l'entreprise.

4. Méthodologie

4.1. Préparation des données

Avant d'appliquer les algorithmes de segmentation et d'association, nous avons effectué plusieurs étapes de préparation des données :

1. **Nettoyage des données** :
 - Suppression des lignes où **CustomerID** est manquant.
 - Filtrage des transactions avec des **quantités ou prix négatifs**.
 - Création de la colonne **TotalPrice** (quantité × prix unitaire).
2. **Analyse exploratoire des données** :
 - Statistiques descriptives sur les variables numériques.
 - Exploration des ventes par pays (hors Royaume-Uni).

4.2. Segmentation des clients

1. Calcul des variables RFM :

- **Récence (R)** : Nombre de jours depuis le dernier achat.
- **Fréquence (F)** : Nombre de transactions effectuées par le client.
- **Montant (M)** : Somme des dépenses par client.

2. Normalisation des données RFM :

- Afin de garantir que les variables ont la même échelle, nous avons utilisé **StandardScaler** pour normaliser les données.

3. Clustering des clients :

- Application de l'algorithme **K-means** pour identifier les segments de clients.
- Utilisation de la **méthode du coude** pour déterminer le nombre optimal de clusters.
- Application du **clustering hiérarchique (CAH)** et visualisation avec un dendrogramme.

4.3. Extraction des règles d'association

1. Création du panier binaire :

- Les transactions sont converties en une matrice où chaque ligne représente une facture, chaque colonne un produit, et la valeur indique si le produit a été acheté (True/False).

2. Application de FP-growth :

- Nous avons utilisé l'algorithme **FP-growth** pour découvrir des ensembles d'articles fréquemment achetés ensemble, en fixant un **min_support** de 0,01.

3. Extraction des règles d'association :

- Des règles d'association ont été extraites à partir des itemsets fréquents, avec une **lift** minimale de 1,0.

5. Résultats

5.1. Segmentation des clients (RFM)

- Les clients ont été segmentés en **4 groupes** à l'aide de l'algorithme **K-means** (basé sur la méthode du coude).
- Un dendrogramme a été utilisé pour valider ces clusters avec l'approche **hiérarchique (CAH)**.

Les principaux segments sont :

1. **Segment 1** : Clients réguliers, avec des achats fréquents et récents.
2. **Segment 2** : Clients ayant un faible montant mais une fréquence élevée.
3. **Segment 3** : Clients avec une faible fréquence mais des achats élevés.
4. **Segment 4** : Clients inactifs, avec peu d'achats récents et une faible fréquence.

5.2. Règles d'association

Voici les **10 meilleures règles d'association** extraites des données :

Antécédents	Conséquents	Support	Confiance	Lift
(WOODLAND CHARLOTTE BAG)	(DOLLY GIRL LUNCH BOX)	0.012	0.50	42.30
(DOLLY GIRL LUNCH BOX)	(WOODLAND CHARLOTTE BAG)	0.012	0.50	42.30
(JUMBO STORAGE BAG SUKI)	(JUMBO BAG BAROQUE BLUE)	0.01	0.45	40.29
(JUMBO BAG PINK POLKADOT)	(JUMBO BAG BAROQUE BLUE)	0.01	0.45	40.29
(PACK OF 60 PINK PAISLEY CAKE CASES)	(JAM MAKING SET)	0.015	0.40	35.25
(60 TEATIME FAIRY CAKE CASES)	(PACK OF 72 RETROSPOT CAKE CASES)	0.018	0.38	35.25

Les règles suggèrent qu'il existe des **produits complémentaires** qui sont souvent achetés ensemble, comme les sacs Woodland Charlotte et les lunch boxes Dolly Girl.

6. Recommandations

1. Stratégies marketing ciblées :

- Pour le **Segment 1**, des offres de fidélisation et de promotions pour maintenir leur fréquence d'achat.
- Pour le **Segment 4**, des campagnes de réactivation via des réductions ciblées.

2. Aménagement du catalogue produit :

- Regrouper les produits fréquemment achetés ensemble dans des **packs promotionnels**.

3. Amélioration de la fidélisation :

- Proposer des **recommandations personnalisées** basées sur les habitudes d'achat précédentes pour maximiser le **cross-selling**.

7. Conclusion

Ce projet a permis d'exploiter les données transactionnelles d'un détaillant en ligne pour segmenter efficacement les clients et extraire des règles d'association utiles. Les segments obtenus permettent de personnaliser les stratégies marketing et d'optimiser les recommandations de produits, contribuant ainsi à améliorer la rentabilité de l'entreprise.

Conclusion du rapport :

Ce rapport couvre toutes les étapes de l'analyse, de la préparation des données à l'extraction des règles d'association, et présente des recommandations pratiques pour l'entreprise.

Slide 1: Title Slide

- **Title:** Analyse et segmentation des comportements d'achat
 - **Subtitle:** Conclusions et recommandations (10 minutes)
-

Slide 2: Introduction

- **Contexte de l'analyse des comportements d'achat**

- Dans un environnement commercial compétitif, comprendre les comportements d'achat des consommateurs est crucial pour développer des stratégies efficaces.
 - L'objectif de ce projet est d'utiliser le data mining pour segmenter les clients et proposer des recommandations marketing basées sur cette segmentation.
-

Slide 3: Préparation des données

- **Nettoyage des données**
 - **Gestion des valeurs manquantes** : Nous avons géré les valeurs manquantes dans des variables critiques comme `CustomerID` et `InvoiceDate` en utilisant l'imputation pour les données essentielles et en éliminant les transactions incomplètes.
 - **Détection des anomalies** : Après avoir analysé les données, nous avons repéré et corrigé des anomalies comme des valeurs de quantités anormalement élevées ou des prix unitaires incorrects.
 - **Analyse exploratoire des données (EDA)**
 - **Statistiques descriptives** : Nous avons calculé des mesures telles que la moyenne, la médiane et l'écart-type des montants dépensés, la fréquence des achats et les dates d'achats pour mieux comprendre le comportement des consommateurs.
 - **Visualisations** : Nous avons créé des graphiques tels que des histogrammes et des box plots pour identifier les tendances, telles que les périodes de forte activité d'achat et les produits les plus populaires.
 - **Feature engineering**
 - Nous avons créé des variables importantes comme :
 - **Récence (R)** : Le temps écoulé depuis le dernier achat d'un client.
 - **Fréquence (F)** : Le nombre total de transactions d'un client.
 - **Montant (M)** : Le montant total dépensé par un client.
-

Slide 4: Méthodologie - Segmentation des clients

- **Choix des algorithmes de segmentation**

- Nous avons utilisé **K-means** et **CAH (Classification Ascendante Hiérarchique)** pour segmenter les clients en fonction de leurs comportements d'achat.
 - **K-means** est un algorithme de clustering basé sur les distances euclidiennes, idéal pour diviser les clients en groupes homogènes.
 - **CAH** est un algorithme hiérarchique qui crée un dendrogramme permettant de visualiser la hiérarchie des groupes de clients. Il aide à déterminer le nombre optimal de clusters.
 - **Modèle RFM**
 - Nous avons appliqué le modèle **RFM** (Récence, Fréquence, Montant) pour identifier des segments de clients en fonction de leur activité d'achat. Par exemple, des clients très actifs avec des dépenses élevées peuvent être vus comme des clients "fidèles".
 - **Normalisation des données**
 - Avant de lancer les algorithmes de segmentation, nous avons normalisé les données pour que chaque variable ait une influence égale sur les résultats des clusters. Cela a permis d'éviter que des variables à plus grande échelle comme le montant total dépensé influencent trop les résultats.
-

Slide 5: Méthodologie - Extraction des règles d'association

- **Objectifs des règles d'association**
 - L'objectif était de découvrir des liens entre les produits achetés ensemble. Par exemple, si un client achète un sac, il pourrait aussi être intéressé par un porte-clés.
 - **Apriori et FP-Growth** sont les deux algorithmes utilisés pour découvrir ces patterns.
- **Transformation des transactions en panier binaire**
 - Les transactions ont été transformées en **panier binaire**, où chaque produit devient une colonne et la présence d'un produit dans une transaction est indiquée par un 1 (présence) ou 0 (absence).
- **Application des algorithmes**

- **Apriori** a permis de générer des itemsets fréquents et de créer des règles d'association entre produits souvent achetés ensemble.
 - **FP-Growth**, une approche plus rapide, a permis de découvrir ces motifs d'achats fréquents sans avoir besoin de générer toutes les combinaisons possibles.
 - **Interprétation des règles**
 - Nous avons utilisé les métriques suivantes pour évaluer les règles :
 - **Support** : La fréquence des items associés dans les transactions.
 - **Confiance** : La probabilité que le produit B soit acheté si A est acheté.
 - **Lift** : Une mesure indiquant la force de l'association, en comparant la probabilité de l'achat de A et B à celle d'un achat aléatoire.
 - **Exemple de règle** : "Si un client achète un sac en toile, il y a 80% de chances qu'il achète également une bouteille d'eau."
-

Slide 6: Résultats et Interprétation

- **Segmentation des clients**
 - Après application des algorithmes de clustering, nous avons obtenu **4 segments distincts** :
 1. **Clients fréquents et fidèles** : Clients ayant un haut nombre d'achats et des dépenses élevées.
 2. **Clients récents** : Clients ayant récemment effectué un achat, mais avec des montants faibles.
 3. **Clients occasionnels** : Clients ayant effectué quelques achats avec des montants plus élevés par achat.
 4. **Clients à faible engagement** : Clients ayant un faible nombre d'achats et des dépenses faibles.
- **Analyse des règles d'association**
 - Nous avons découvert plusieurs règles intéressantes telles que :
 - "Si un client achète un sac à dos, il est probable qu'il achète aussi un porte-clés."

- Ces règles permettent de mieux comprendre les produits complémentaires et ouvrent des opportunités pour **le cross-selling**.
-

Slide 7: Recommandations Business

- **Stratégies marketing ciblées**

- Pour chaque segment de clients, des stratégies personnalisées sont proposées :
 - **Clients fréquents et fidèles** : Offrir des récompenses et des promotions exclusives pour renforcer leur fidélité.
 - **Clients récents** : Proposer des réductions sur leurs achats futurs pour encourager la répétition des achats.
 - **Clients occasionnels** : Suggérer des promotions spéciales ou des bundles pour les inciter à acheter plus.
 - **Clients à faible engagement** : Lancer des campagnes de réactivation avec des offres spéciales pour stimuler leurs achats.

- **Optimisation du catalogue produit**

- Sur la base des règles d'association, nous recommandons de **réorganiser le catalogue produit** pour mettre en avant les produits complémentaires dans des **bundles** ou des **offres spéciales**, augmentant ainsi le taux de conversion et la valeur moyenne des commandes.

- **Amélioration de la fidélisation**

- **Programmes de fidélité** : Mettre en place des programmes qui récompensent les achats fréquents, en offrant des points de fidélité pour chaque achat ou en offrant des remises sur des achats futurs.
 - **Offres personnalisées** : Utiliser les informations de segmentation pour offrir des promotions adaptées à chaque groupe de clients, ce qui maximisera leur satisfaction et leur fidélité.
-