

Analyse du comportement des clients

1. But du projet

Ce projet analyse le comportement d'achat des clients à l'aide de données transactionnelles provenant de 3 900 achats dans différentes catégories de produits. L'objectif est de découvrir des informations sur les habitudes de dépenses, les segments de clientèle, les préférences produits et le comportement des abonnements afin d'orienter les décisions stratégiques d'affaires.

2. Dataset

Lignes : 3 900 - Colonnes : 18 - Caractéristiques clés :

Démographie des clients (Âge, sexe, localisation, statut de l'abonnement)

Détails de l'achat (article acheté, catégorie, montant d'achat, saison, taille, couleur)

Comportement d'achat (Remise appliquée, code promo utilisé, achats précédents, fréquence d'achats, évaluation, type d'expédition)

Données manquantes : 37 valeurs dans la colonne de note des avis

3. Analyse exploratoire avec Python

Nous commençons par la préparation et le nettoyage des données en Python :

Chargement des données : Importer le jeu de données à l'aide `de pandas`.

Exploration initiale : Utilisation `de df.info()` pour vérifier la structure et `.describe()` pour des statistiques résumées.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	39
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	22
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN

Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
3900	3900	3900.000000	3900	3900
2	2	NaN	6	7
No	No	NaN	PayPal	Every 3 Months
2223	2223	NaN	677	584
NaN	NaN	25.351538	NaN	NaN
NaN	NaN	14.447125	NaN	NaN
NaN	NaN	1.000000	NaN	NaN
NaN	NaN	13.000000	NaN	NaN
NaN	NaN	25.000000	NaN	NaN
NaN	NaN	38.000000	NaN	NaN
NaN	NaN	50.000000	NaN	NaN

Gestion des données manquantes : Vérification des valeurs nulles et imputées valeurs manquantes dans la **colonne Note d'avis** en utilisant la note médiane de chaque catégorie de produit.

Standardisation des colonnes : Renommé les colonnes en **snake case** pour une meilleure lisibilité et documentation.

Ingénierie des fonctionnalités :

- Crée **age_group** colonne en triant les âges des clients.
- Création **purchase_frequency_days** colonne à partir des données d'achat.

Vérification de la cohérence des données : Vérification si **discount_applied** et **promo_code_used** étaient redondants ; **promo_code_used** supprimés.

Intégration de la base de données : Connexion d'un script Python à PostgreSQL et chargé le DataFrame nettoyé dans la base de données pour l'analyse SQL.

4. Analyse des données via SQL

Nous avons réalisé une analyse structurée dans PostgreSQL pour répondre aux questions clés du métier :

Revenus par genre – Comparaison du chiffre d'affaires total généré par des clients masculins et féminins.

	gender text 	revenue numeric 
1	Female	75191
2	Male	157890

Utilisateurs dépensant beaucoup de rabais – Identifié les clients qui ont utilisé des réductions mais ont tout de même dépensé au-dessus du montant moyen des achats.

	customer_id bigint 	purchase_amount bigint 
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	88

Total rows: 839 Query complete 00:00:00

Top 5 des produits par note – Produits trouvés avec les meilleures notes moyennes d'avis.

	item_purchased text	Average Product Rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

Comparaison des types d'expédition – Comparaison des montants moyens d'achat entre l'expédition Standard et Express.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

Abonnés vs. Non-abonnés – Comparaison des dépenses moyennes et du chiffre d'affaires totaux selon le statut d'abonnement.

	subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

Produits dépendants des remises – Identifié 5 produits ayant le pourcentage le plus élevé d'achats à réduction.

	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

Segmentation client – Classé les clients en segments Nouveau, Retour et Fidèle selon l'historique d'achats.

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

Top 3 des produits par catégorie – Listé les produits les plus achetés dans chaque catégorie.

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

Acheteurs réguliers et abonnements – Vérification des clients ayant >5 achats sont plus susceptibles de s'abonner.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

Revenus par groupe d'âge – Contribution totale calculée des recettes de chaque groupe d'âge.

	age_group	total_revenue
	text	numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763

5. Tableau de bord dans Power BI

Enfin, nous avons construit un tableau de bord interactif dans **Power BI** pour présenter visuellement des informations.



6. Recommandations

Boostez les abonnements – Promouvoir des avantages exclusifs pour les abonnés.

Programmes de fidélité client – Récompenser les acheteurs réguliers pour les faire passer dans le segment « Fidèle ».

Révisez la politique de réduction – Équilibrer les gains de ventes avec le contrôle des marges.

Positionnement produit – Mettre en avant les produits les mieux notés et les plus vendus lors des campagnes.

Marketing ciblé – Concentrer vos efforts sur les groupes d'âge à forte croissance et les utilisateurs de l'expédition express.