

Analyse du Comportement d'Achat des Clients

1. Vue d'Ensemble du Projet

Ce projet analyse le comportement d'achat des clients à partir de données transactionnelles couvrant 3 900 achats dans diverses catégories de produits. L'objectif est de dégager des informations sur les habitudes de dépense, les segments de clientèle, les préférences produits et le comportement d'abonnement afin d'orienter les décisions stratégiques de l'entreprise.

2. Résumé du Jeu de Données

- Lignes : 3 900
- Colonnes : 18
- Principales caractéristiques :
 - Données démographiques (âge, sexe, localisation, statut d'abonnement)
 - Détails d'achat (produit acheté, catégorie, montant, saison, taille, couleur)
 - Comportement d'achat (remise appliquée, code promo utilisé, achats précédents, fréquence d'achat, note de satisfaction, type d'expédition)
 - Données manquantes : 37 valeurs manquantes dans la colonne "Note de satisfaction"

3. Analyse Exploratoire avec Python

Les étapes de préparation et de nettoyage des données ont été effectuées en Python :

- Chargement des données : importation via pandas.
- Exploration initiale : utilisation de df.info() et .describe() pour les statistiques descriptives.

	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

- Traitement des données manquantes : imputation de la médiane par catégorie de produit.
- Normalisation des noms de colonnes (snake case).
- Ingénierie de variables : création d'un groupe d'âge et d'un indicateur de fréquence d'achat.
- Vérification de la cohérence et intégration en base de données PostgreSQL.

4. Analyse des Données avec SQL

Les requêtes ont permis de répondre à plusieurs questions commerciales clés :

1. Revenu par sexe – comparaison des revenus hommes/femmes.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

2. Utilisateurs à forte dépense avec remise.

	customer_id	purchase_amount
	bigint	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

3. Top 5 des produits selon la note moyenne.

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

4. Comparaison entre livraison standard et express.

	shipping_type 	round numeric 
1	Standard	58.46
2	Express	60.48

5. Comparaison abonnés vs non-abonnés.

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

6. Produits dépendants des remises.

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

7. Segmentation client : nouveaux, récurrents, fidèles.

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

8. Top 3 des produits par 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

9. Fidélité et abonnement : lien entre fréquence d'achat et souscription.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. Revenu par groupe d'âge.

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

5. Tableau de Bord Power BI

Un tableau de bord interactif a été construit sous Power BI pour visualiser les résultats.



6. Recommandations Stratégiques

- Encourager les abonnements via des avantages exclusifs.
- Fidéliser les clients récurrents avec des programmes de récompense.
- Revoir la politique de remise pour équilibrer ventes et marges.
- Mettre en avant les produits les mieux notés dans les campagnes.
- Cibler les groupes d'âge à fort revenu et les utilisateurs de la livraison express.