

USE CASE :

**FAÇONNER L'EXPÉRIENCE NUMÉRIQUE
VIA LA SEGMENTATION DES CLIENTS**

**PUGET STEPHAN
VIAL SIMON
BOYER LÉA
SATHYAPALAN JONATHAN**

BUSINESS UNDERSTANDING

CONTEXTE

- COMPORTEMENTS ET ATTENTES VARIÉS DES UTILISATEURS
- NÉCESSITÉ D'UN MOTEUR DE RECHERCHE EFFICACE ET ADAPTATIF
- UNE EXPÉRIENCE DE RECHERCHE INEFFICACE PEUT CRÉER UN MANQUE À GAGNER

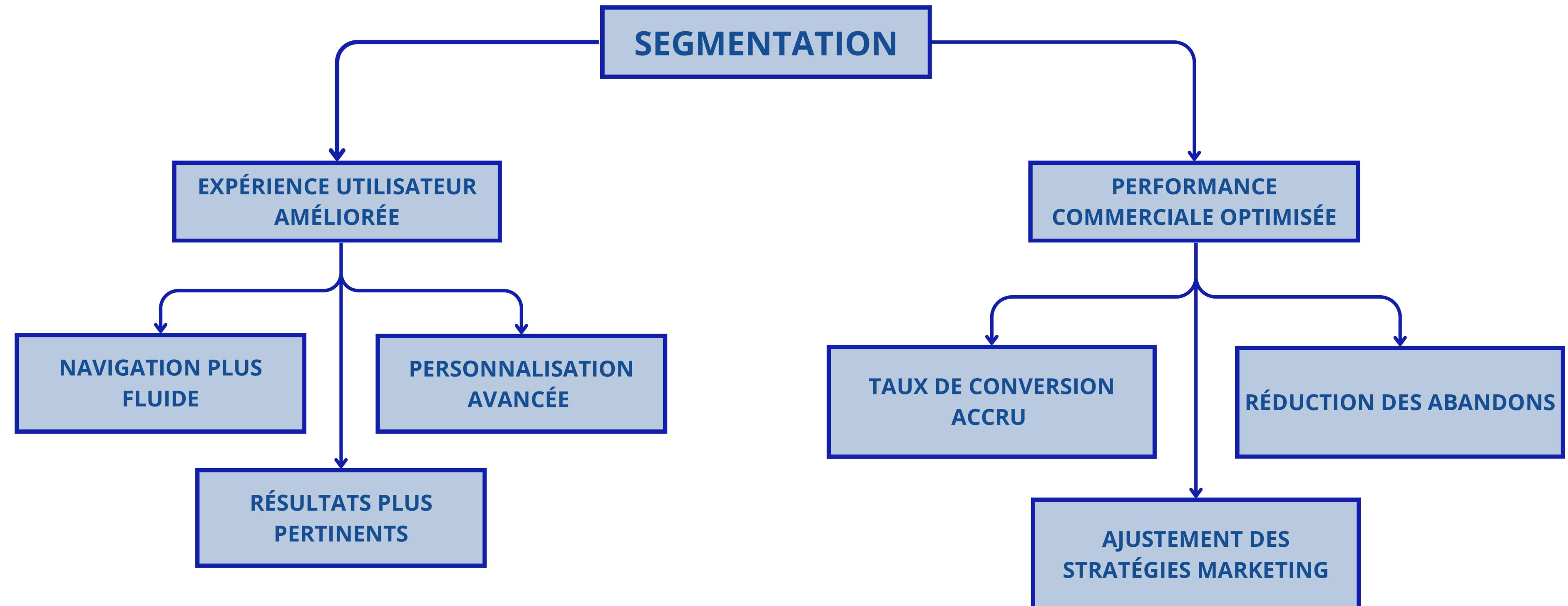
PROBLÉMATIQUE

COMMENT SEGMENTER EFFICACEMENT LES UTILISATEURS POUR
OPTIMISER LEUR EXPÉRIENCE DE RECHERCHE ET AMÉLIORER LA
PERTINENCE DES RÉSULTATS ?

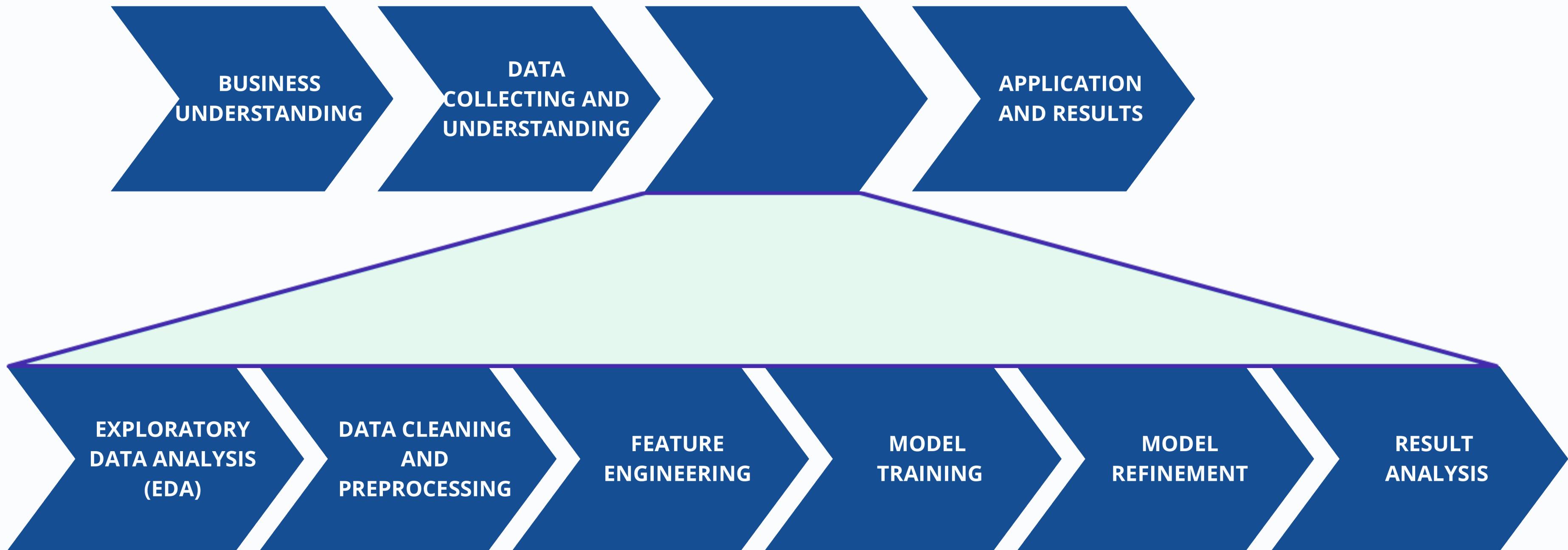
OBJECTIFS

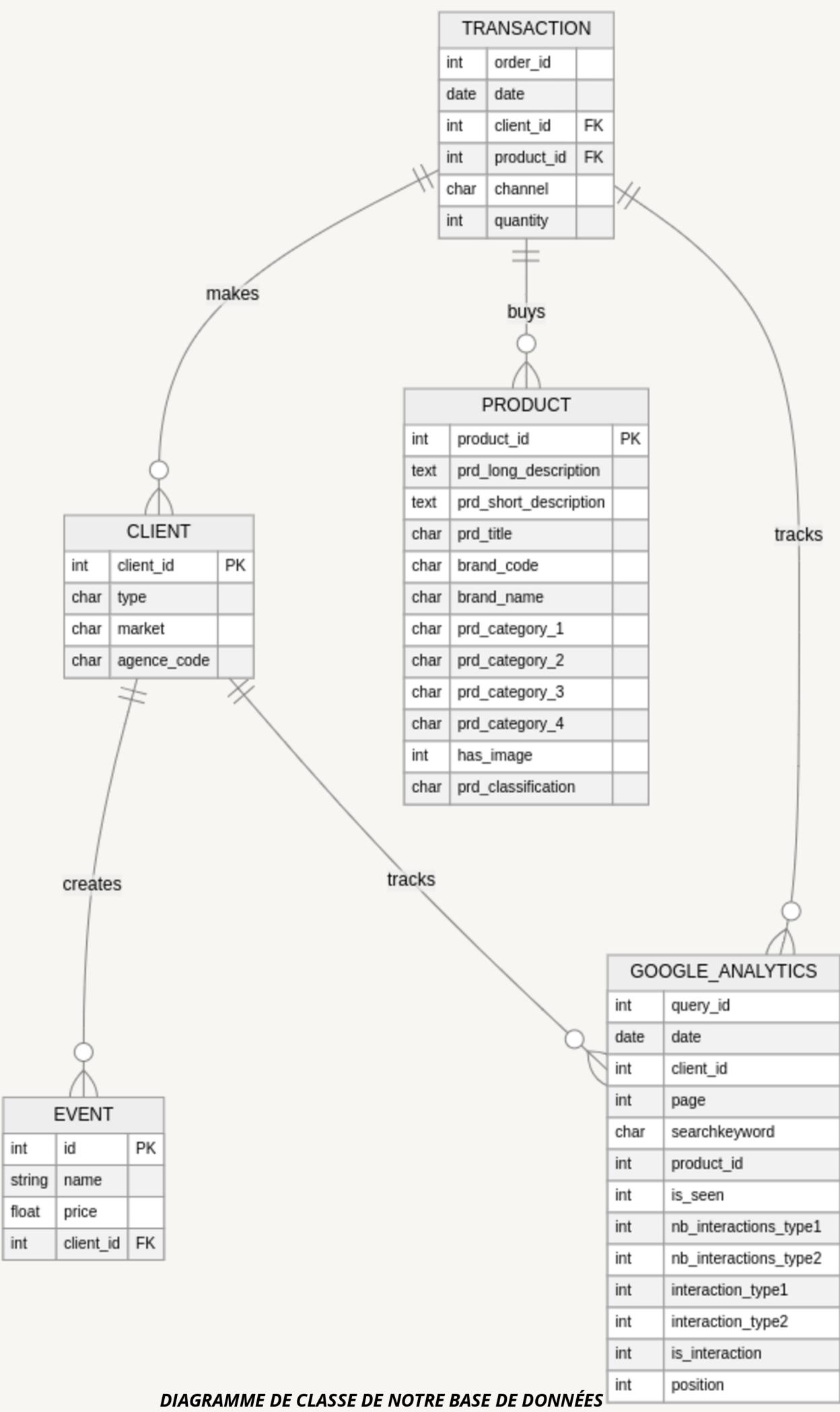
1. IDENTIFIER LES FACTEURS CLÉS
2. PROPOSER UN MODÈLE DE SEGMENTATION ET L'ÉVALUER
3. DÉTERMINER LES PROCHAINES ÉTAPES

BUSINESS UNDERSTANDING



METHODOLOGY





Jeu de données à notre disposition :

- Table Client
- Table Produit
- Table Event
- Table Google_Analytics
- Table Transaction

TABLE CLIENT :

CLIENT_ID', 'TYPE', 'MARKET', 'AGENCE_CODE'

TABLE PRODUIT :

PRODUCT_ID','PRD_CATEGORY_X'...

TABLE TRANSACTION :

ORDER_ID','CLIENT_ID','PRODUCT_ID

Table Transactions

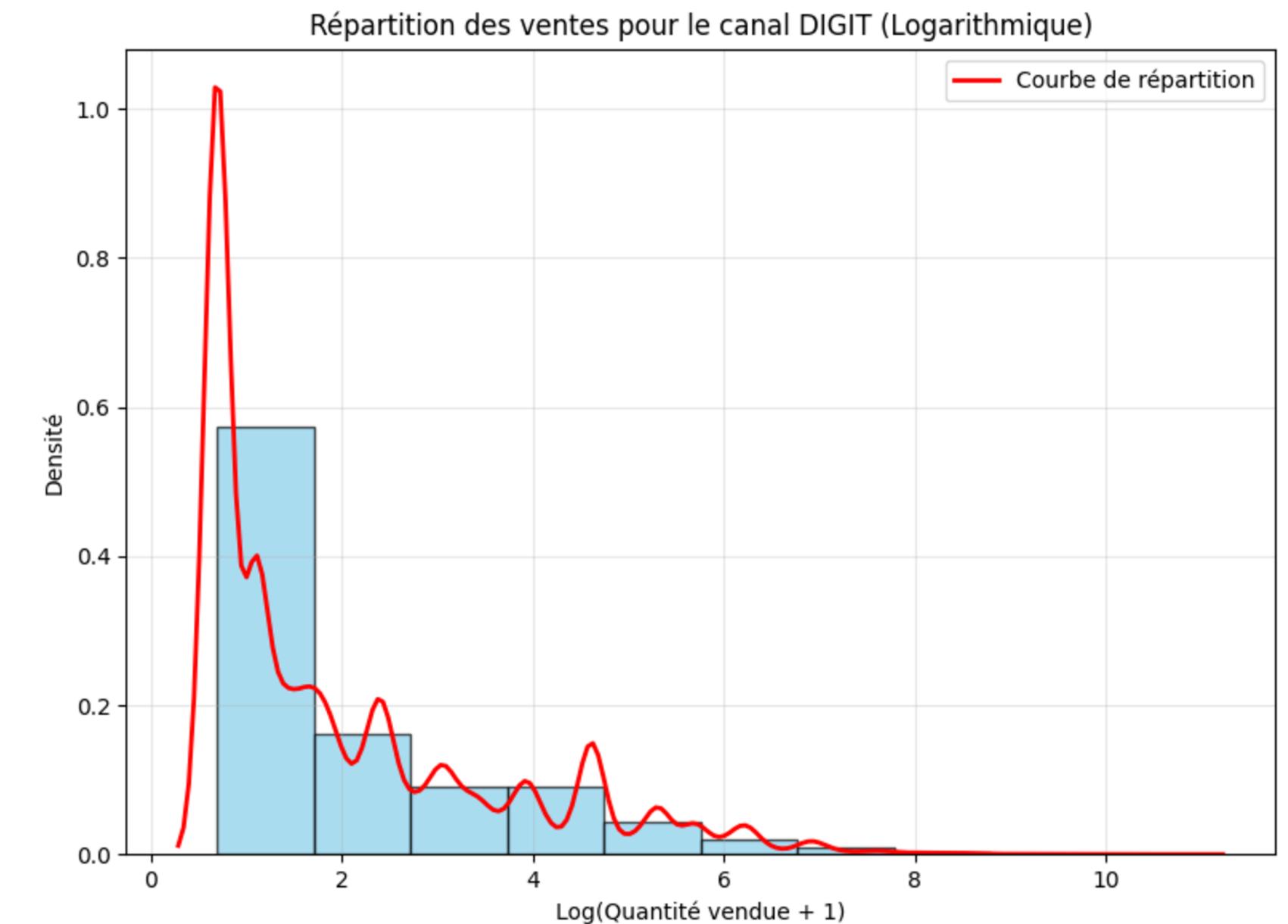
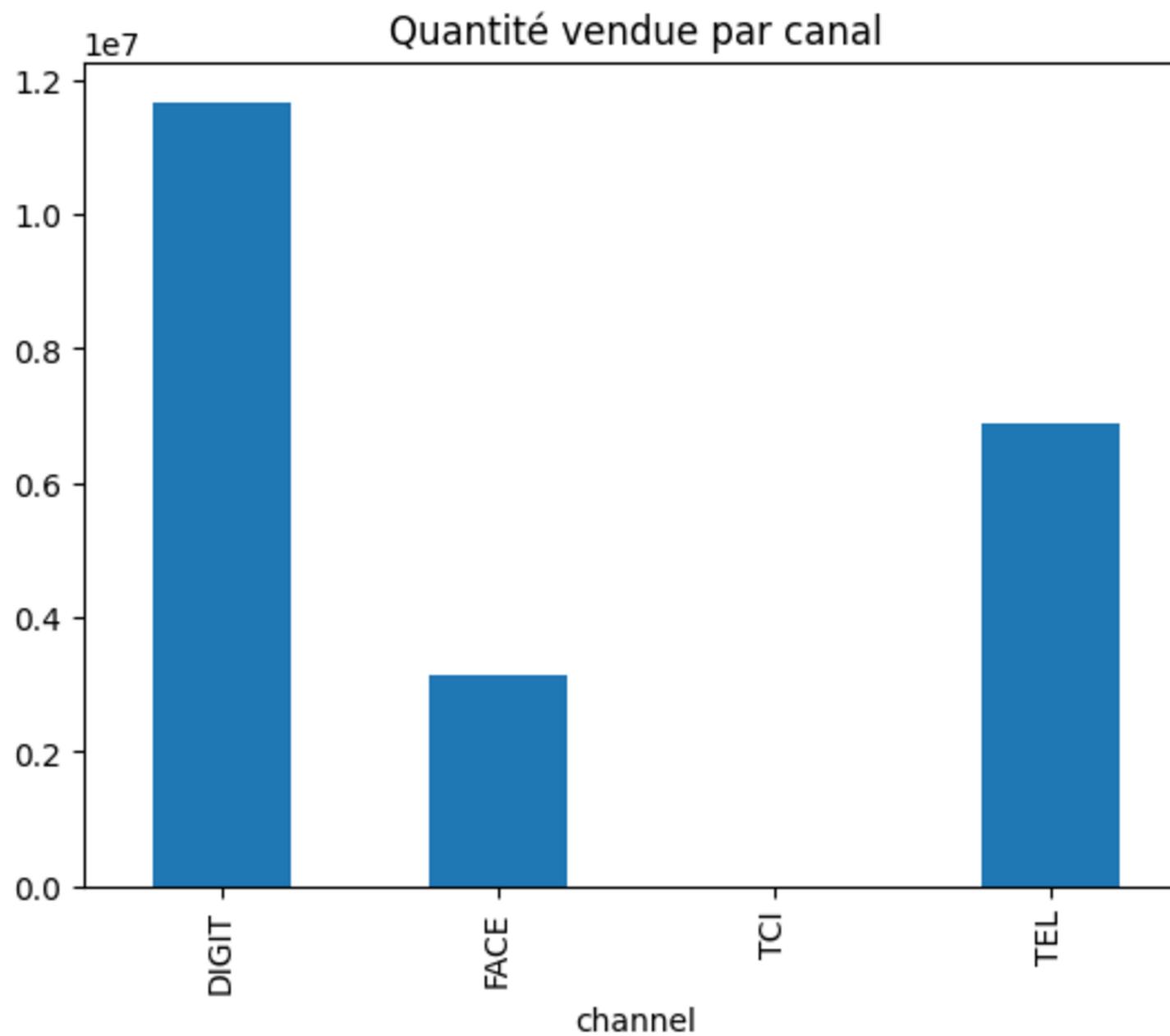


Table Transactions

EXPLORATORY
DATA ANALYSIS
(EDA)

DATA CLEANING
AND
PREPROCESSING

FEATURE
ENGINEERING

MODEL
TRAINING

MODEL
REFINEMENT

RESULT
ANALYSIS

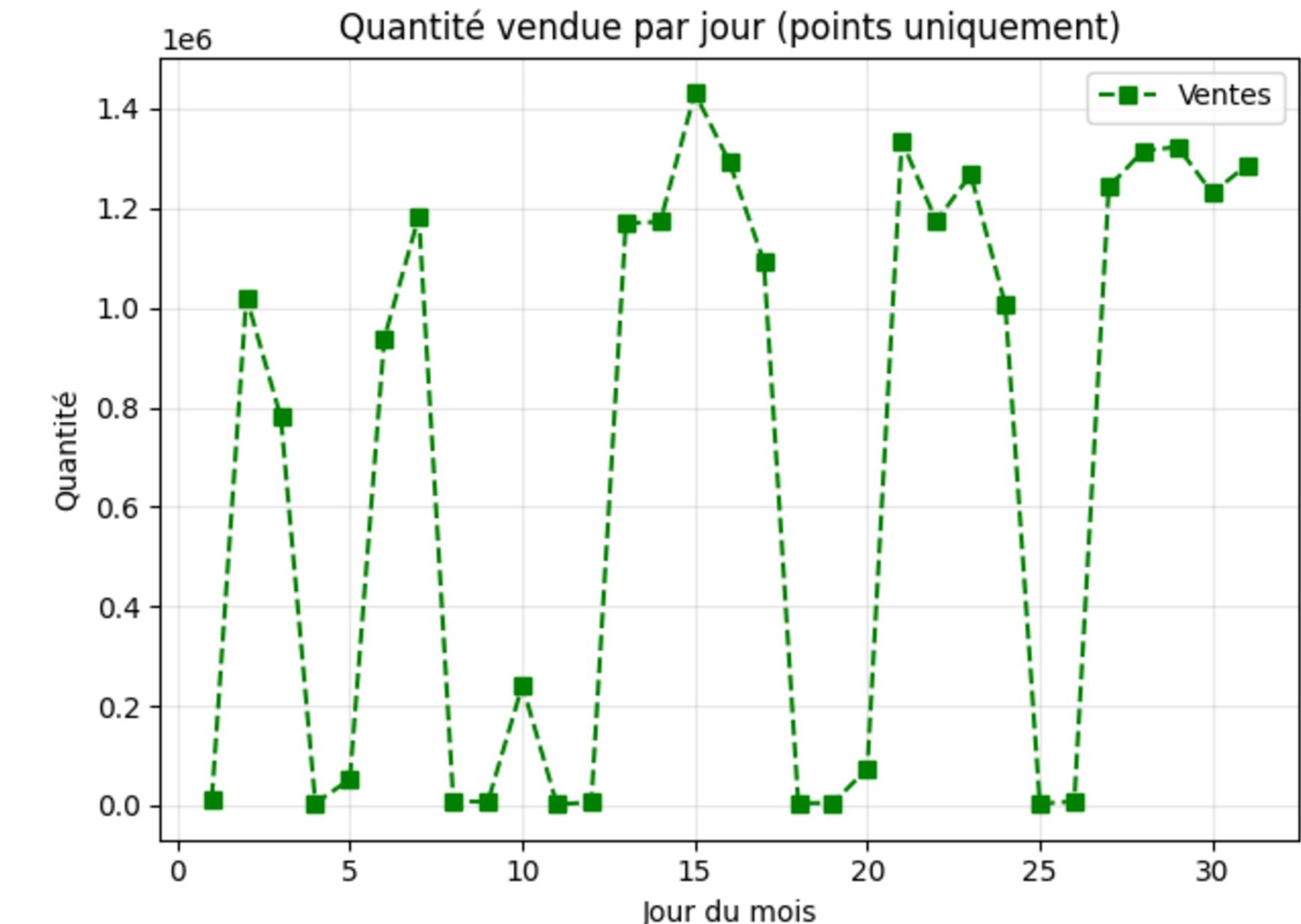
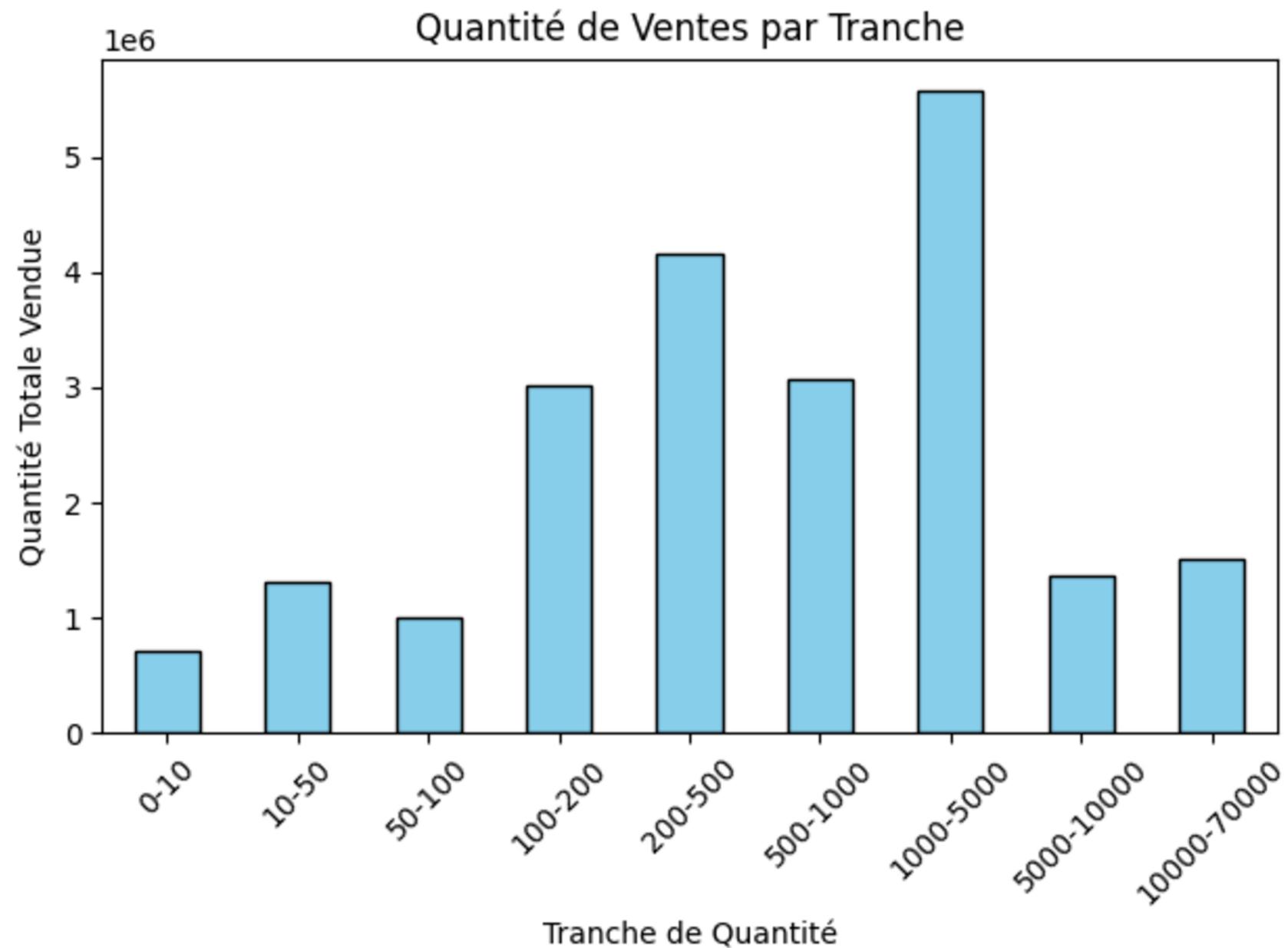




Table google_analytics_data_anonymized

Triplet {client_id, date, searchkeyword}

	count_non_null_values	%_non_null_v
query_id	5662375	
date	5662375	
client_id	5662375	
page	5662375	
searchkeyword	5662375	
product_id	5662375	
is_seen	5662375	
nb_interactions_type1	5662375	
nb_interactions_type2	5662375	
interaction_type1	5662375	
interaction_type2	5662375	
is_interaction	5662375	100.0
position	5662375	100.0

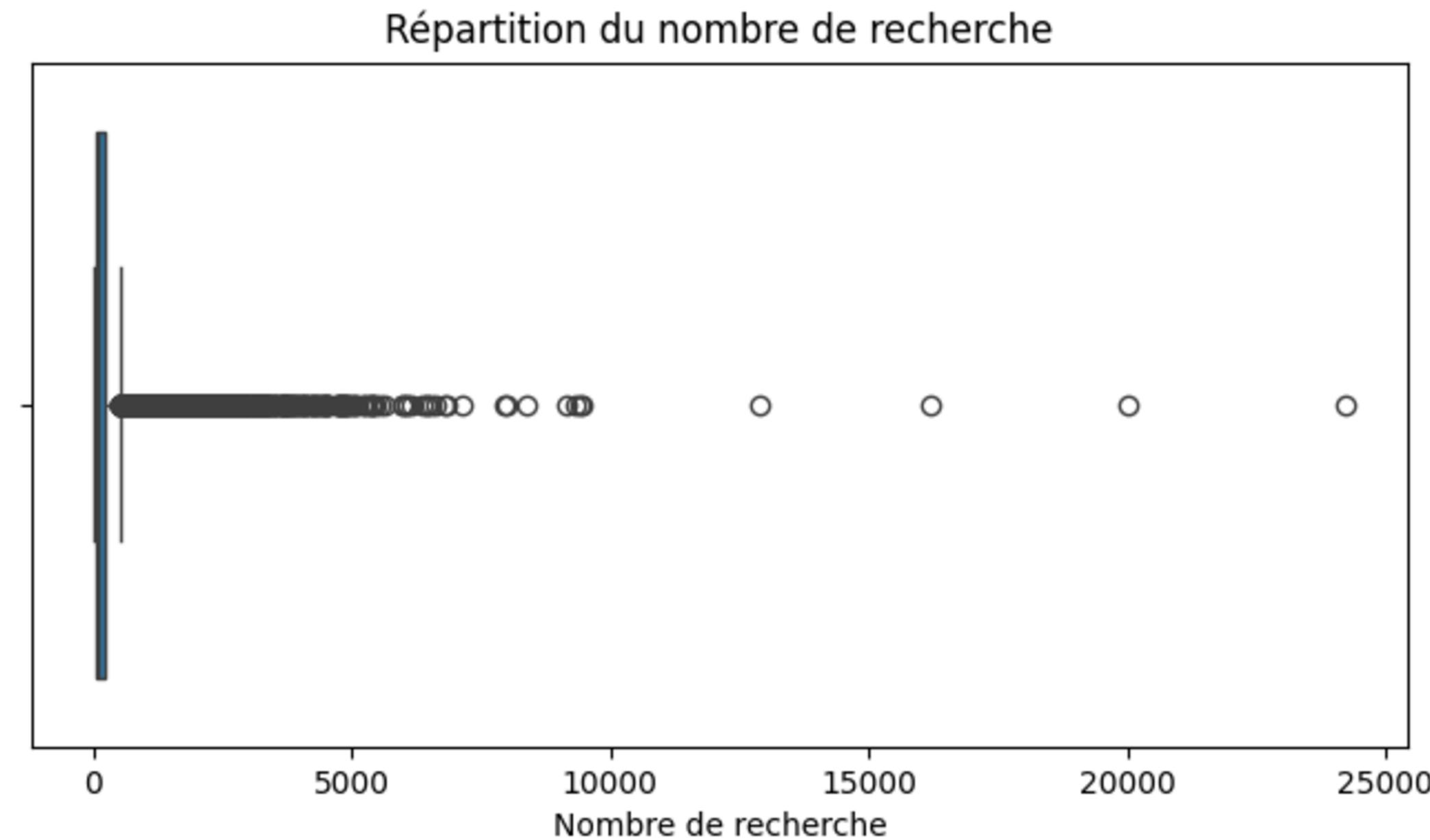




Table google_analytics_data_anonymized

Triplet {client_id, date, searchkeyword}

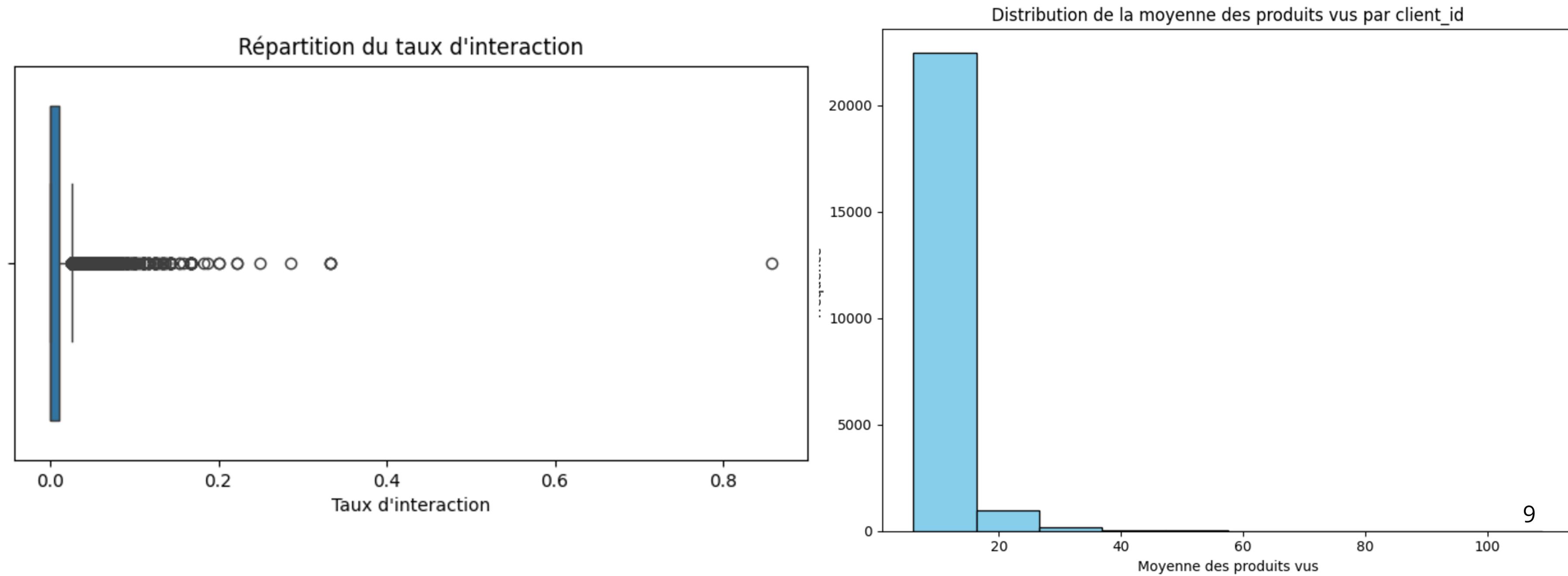


Table Google Analytics



Jeu de données à notre disposition :

- Table google_analytics_events_data_anonymized

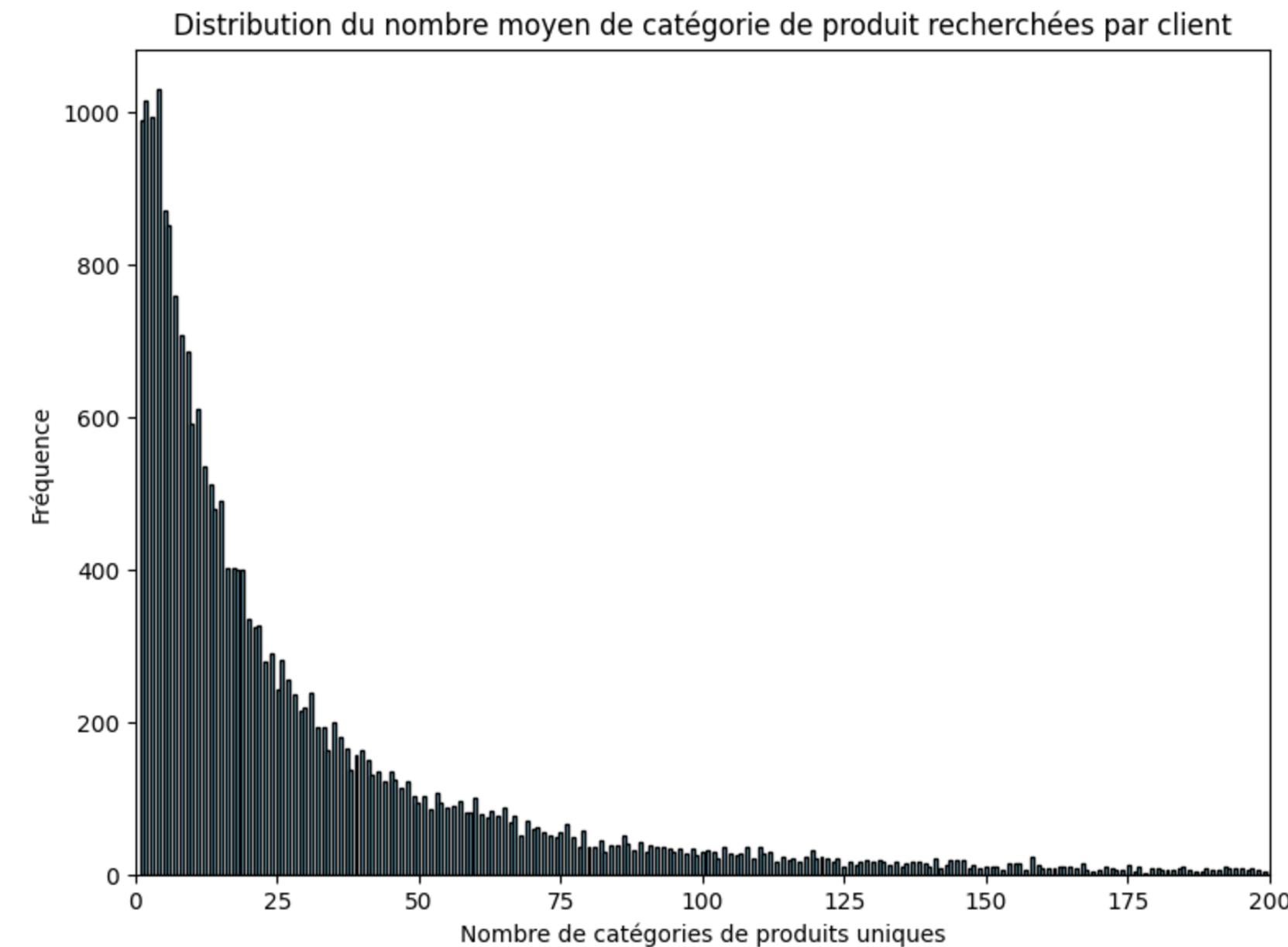
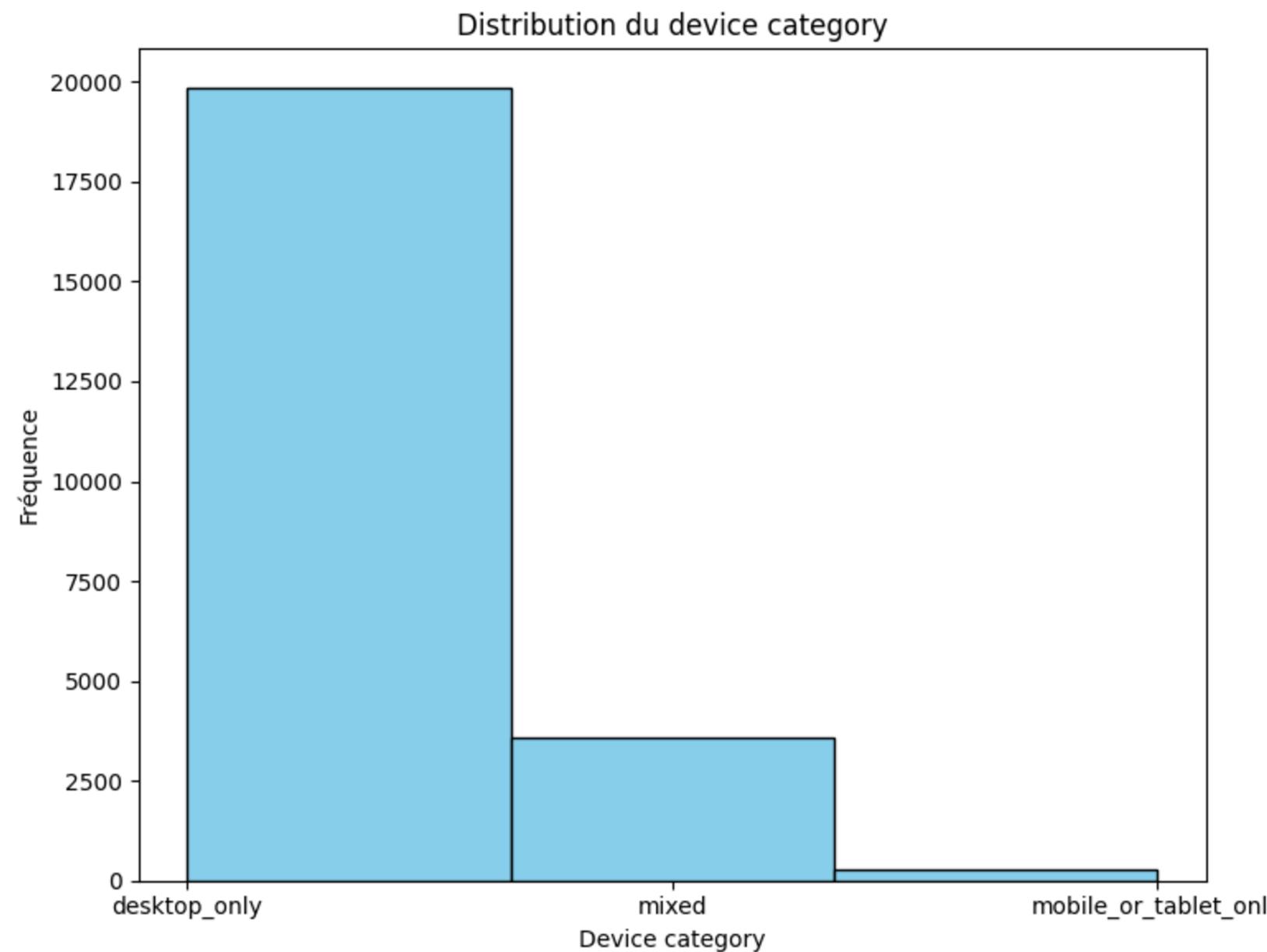


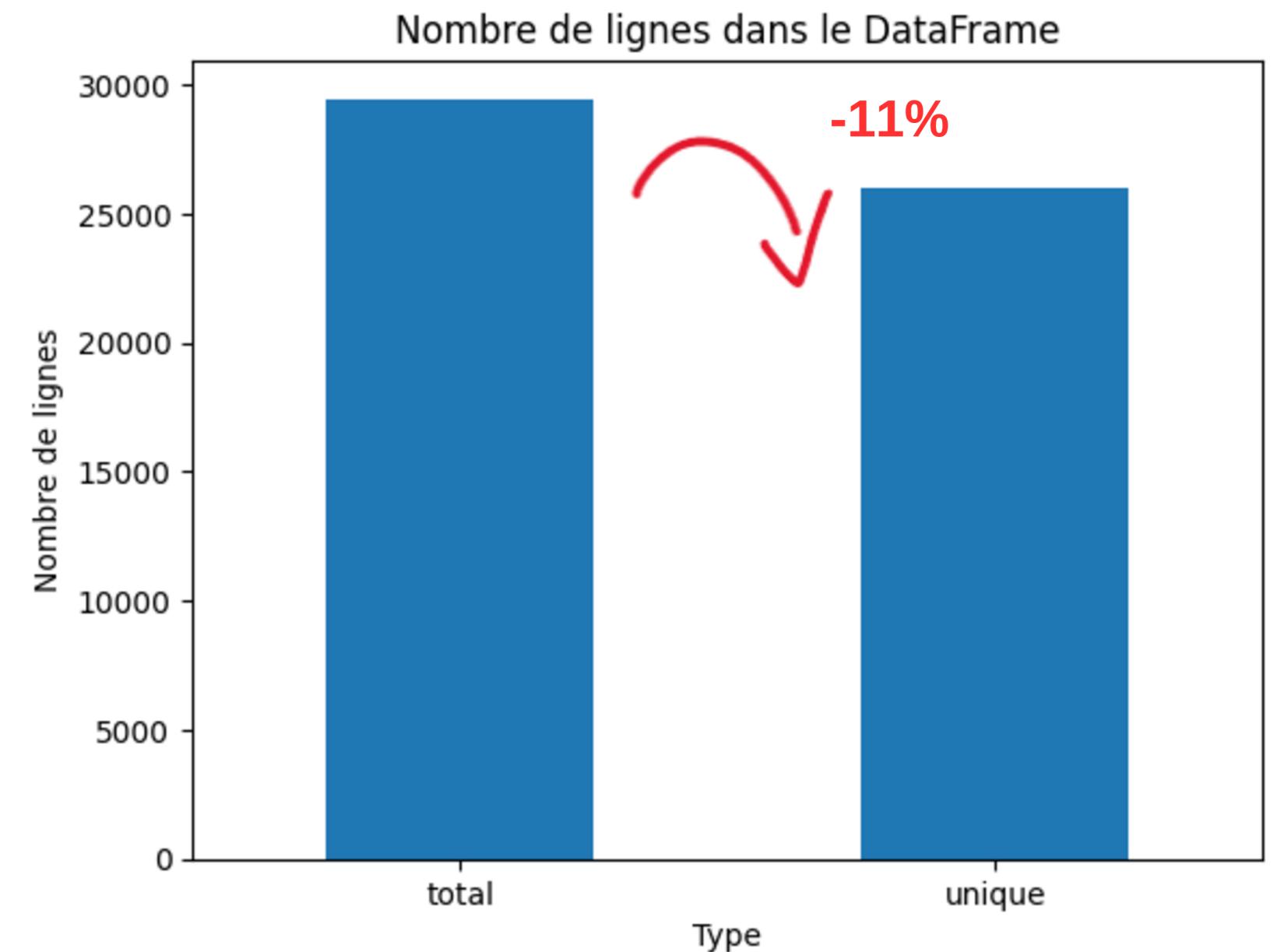
Table Client



Cleaning réalisé sur la table Client :

- Suppression des doublons
 - 29 447 -> 26 024 rows
- Suppression de la colonne agence_code

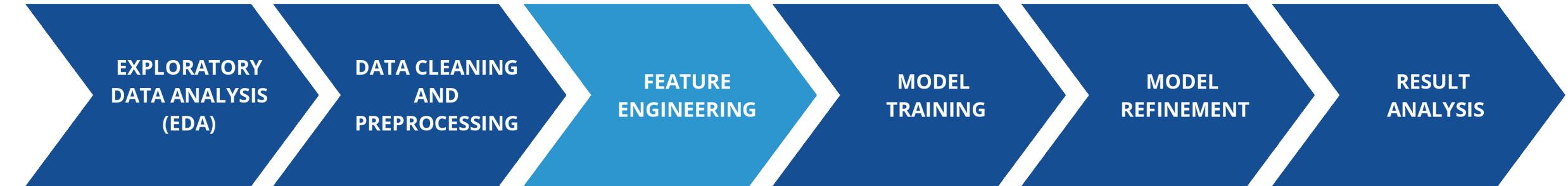
	client_id	type	market	agence_code	count_id
19978	817c3c	Installateurs	Habitat / Petit Tertiaire	7c98	2
28682	817c3c	Installateurs	Habitat / Petit Tertiaire	2b65	2





FEATURES COMPORTEMENT D'ACHAT

- Les 3 produits achetés le plus souvent par client (fréquence)
- Canal le plus utilisé par client
- Nombre de canaux différents utilisés par un client
- Taux d'interaction par recherche
- Nombre moyen de produits vus par recherche
- Temps moyen entre deux recherches
- Nombre de mots moyen par recherche



FEATURES VALEUR CLIENT

- **Nombre de commandes différentes par client**
- **Nombre de produits différents commandés par client**
- **Quantité moyenne par commande pour chaque client**
- **Taux d'achat par recherche**



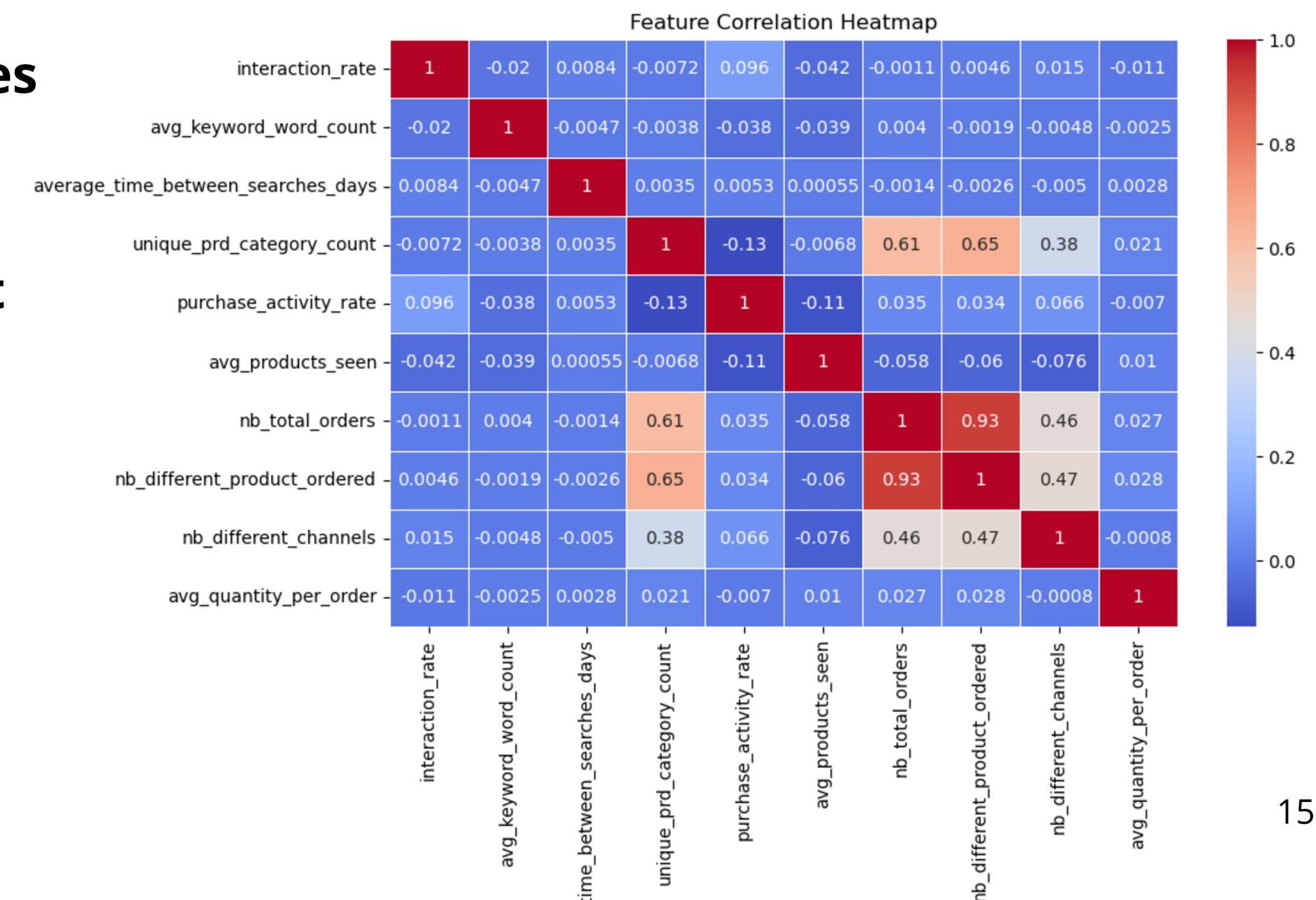
FEATURES PRÉFÉRENCES ET AFFINITÉS

- Marque préférée du client
- Le produit acheté en plus grande quantité par client
- **Catégorie produit 3 préférée du client**
- Nombre de catégorie différentes de produit recherchées
- Appareil de recherche favori



PRÉPARATION DES DONNÉES POUR L'ENTRAÎNEMENT

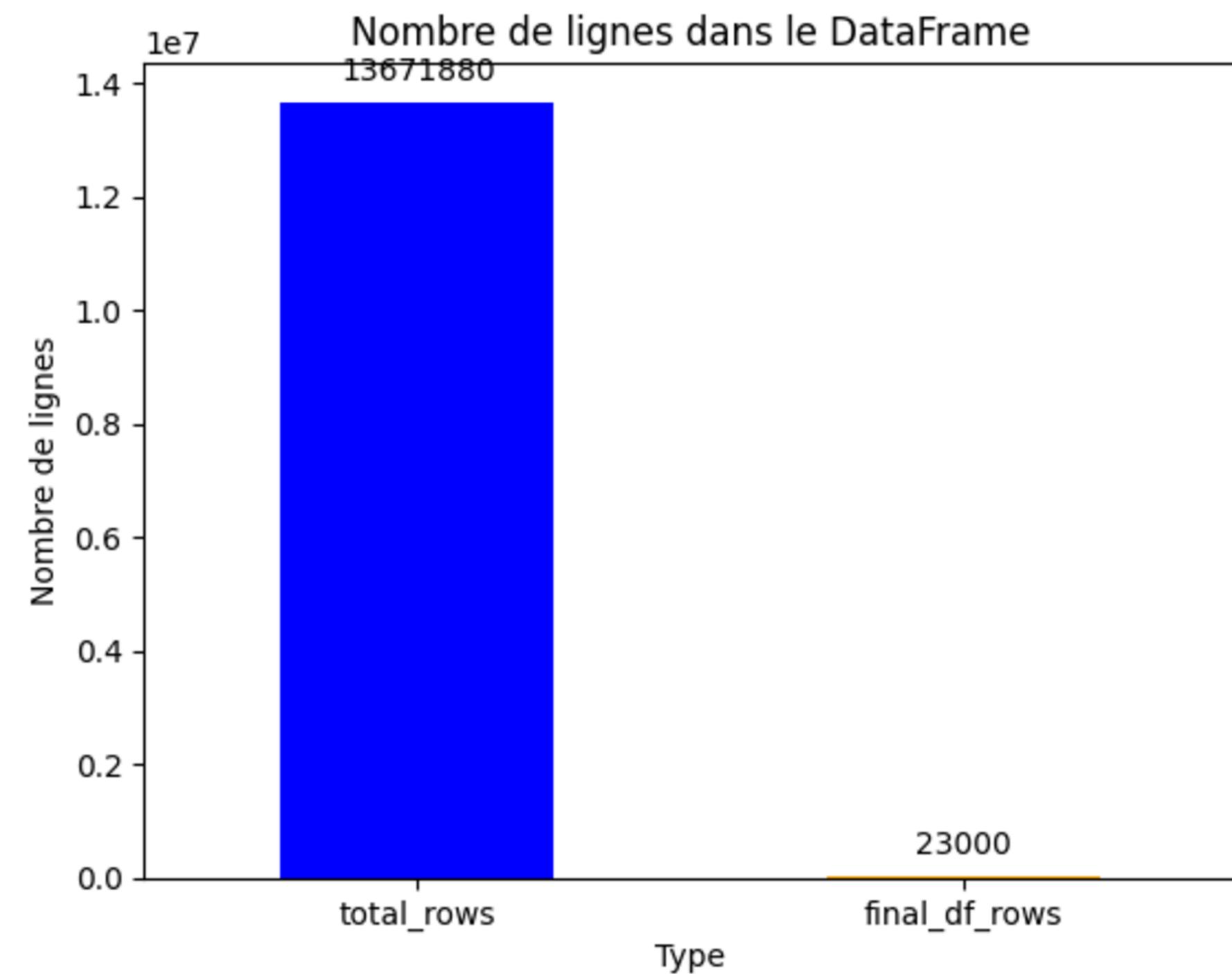
- Sélection des features pertinentes
- Transformation logarithmique et Winsorization





PRÉPARATION DES DONNÉES POUR L'ENTRAÎNEMENT

- **Encodage des données catégorielles**
- **Normalisation des données**
- **Dataframe final :**
 - **23 000 lignes**
 - **22 colonnes**





METRICS UTILISÉES

- **Silhouette score:** Mesure si les données sont bien labelisées ($[-1, 1]$)
- **Davies-Bouldin index:** Mesure de similitude entre les clusters
- **Calinski-Harabasz index:** Mesure la densité et la séparation des clusters

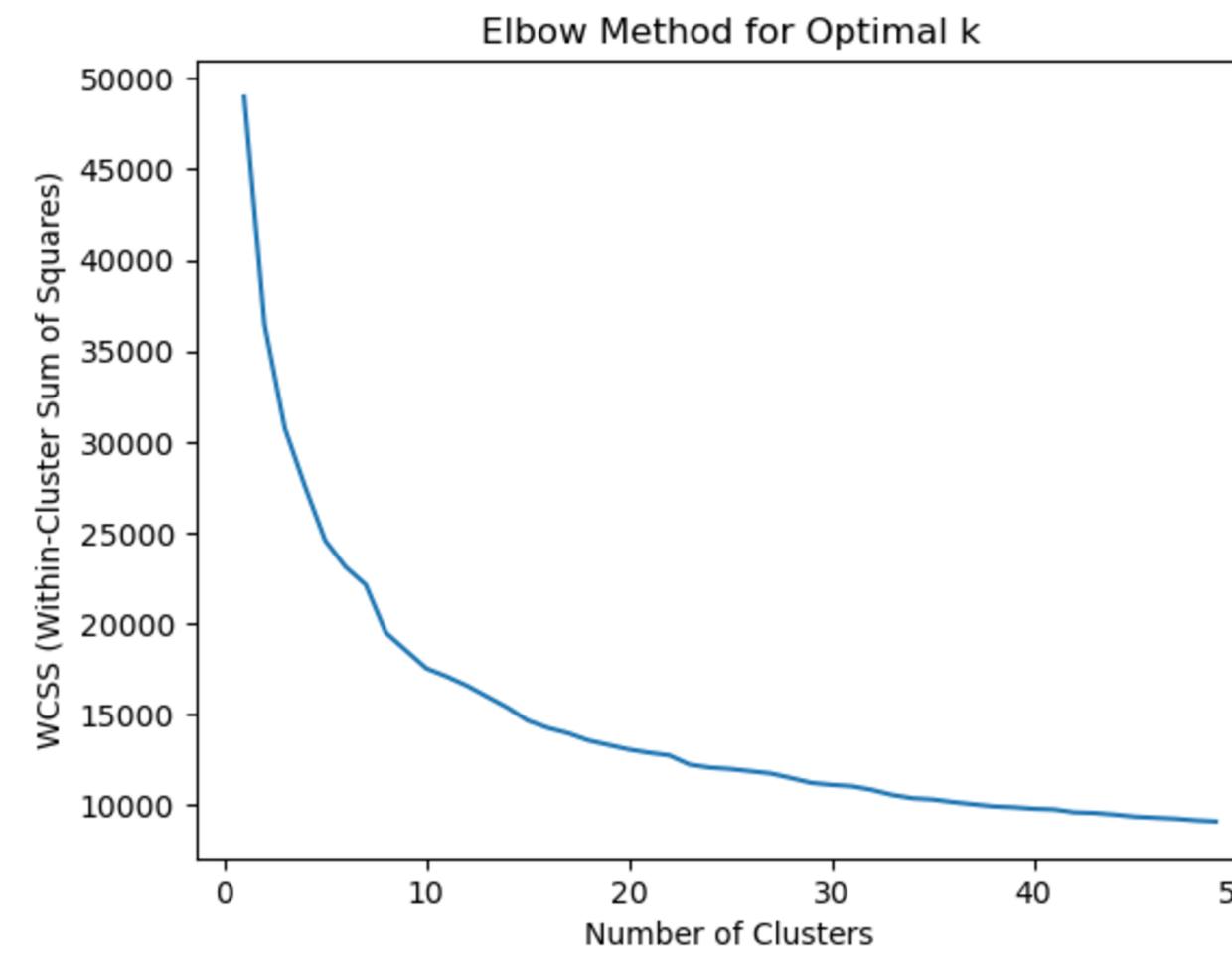
MODÈLES UTILISÉES

- **KMeans**
- **DBSCAN**
- **Clustering Agglomératif**



EXPLORATION D'HYPÉRPARAMÈTRES

- **KMeans:** Nombre de cluster



- **DBSCAN:** eps
- **Agglo:** Nombre de clusters



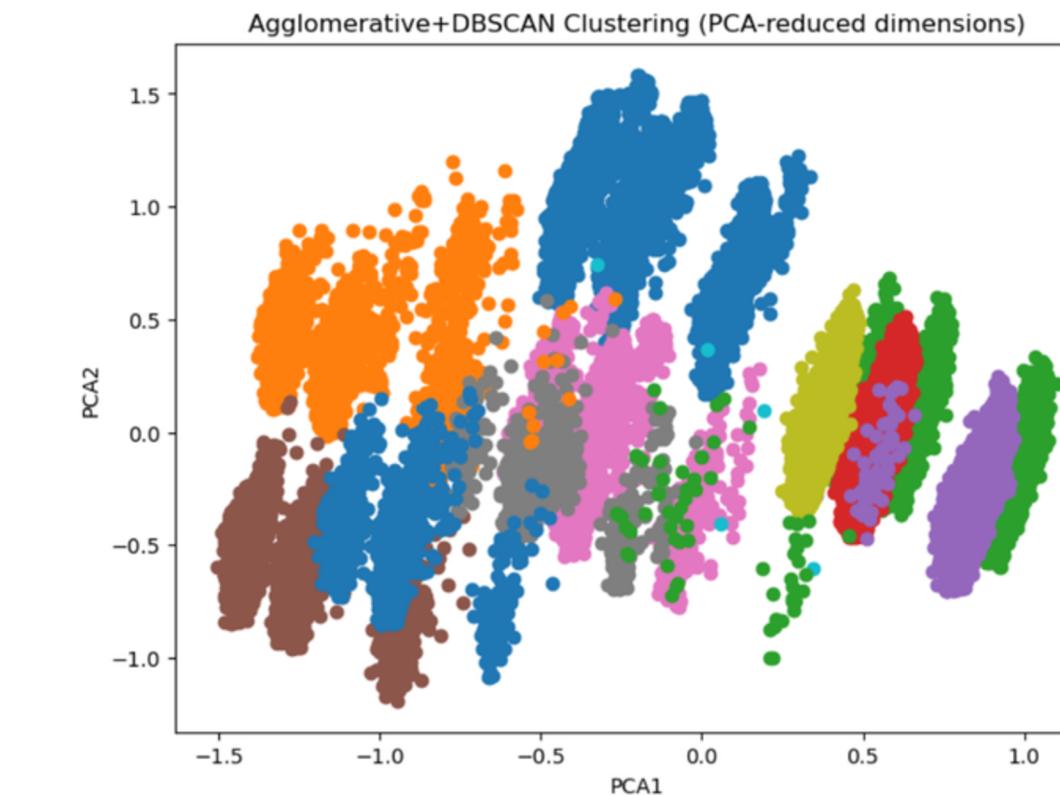
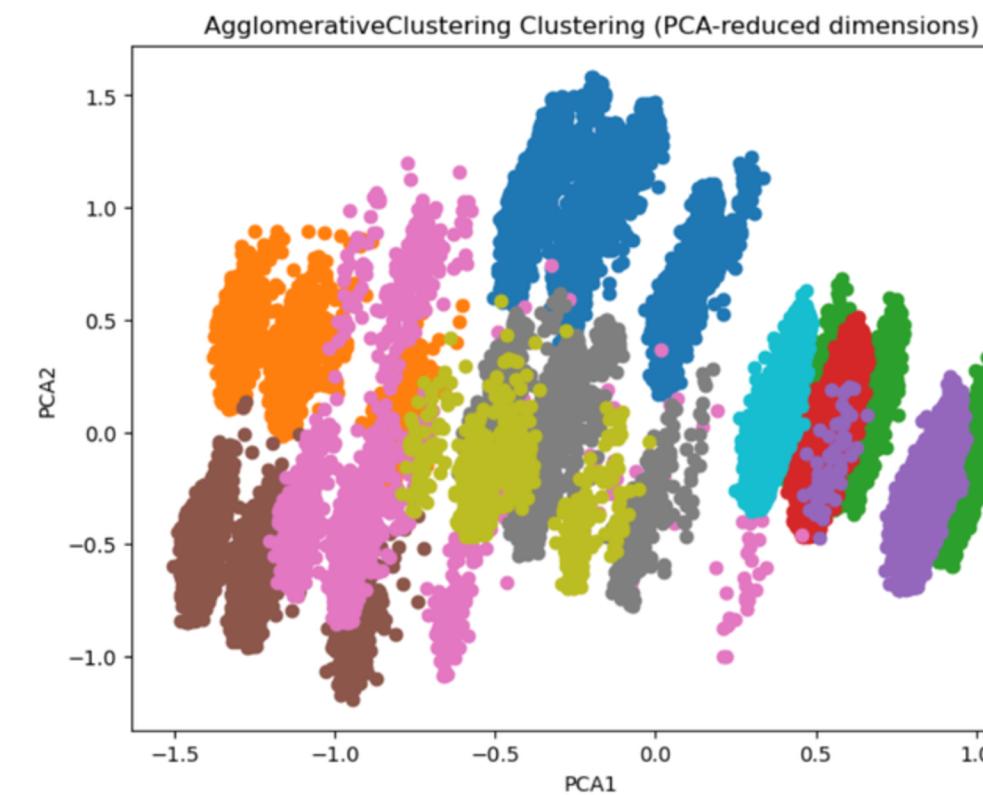
SCORES

	KMeans	DBSCAN	Agglo
Silhouette	0.24	0.31	0.28
DB	1.47	1.60	1.36
CH	4.2k	2.8k	4.7k



AMÉLIORATION DU CLUSTERING

- **Rafinement du modèle**
 - Clustering agglomératif > DBSCAN pour trouver sous-clusters



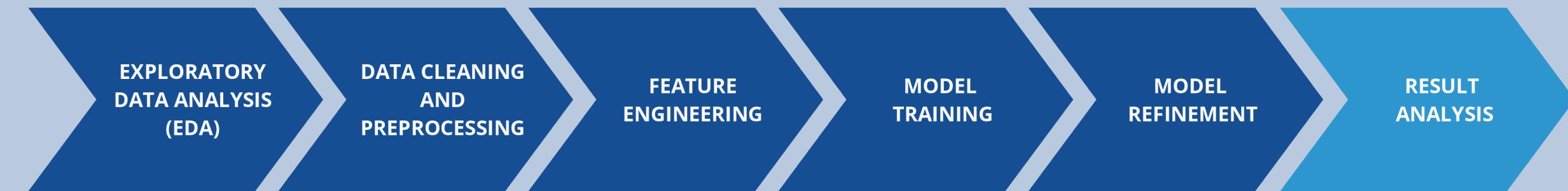
--> score non améliorée



AMÉLIORATION DU CLUSTERING

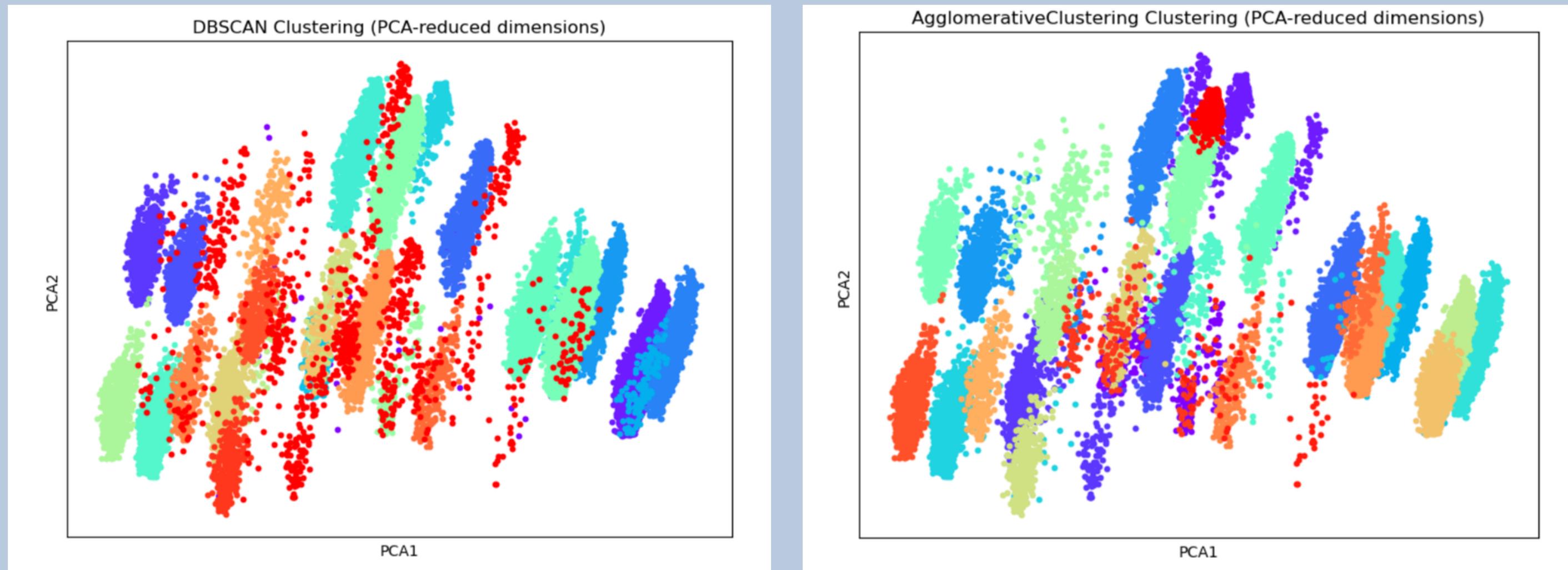
- **Ajustement des poids des données**
 - Plus de poids aux données catégorielles et au nombres de commandes

Agglo	Sans ajustement de poids	Avec ajustement de poids
Silhouette Score	0.28	0.47
DB	1.36	0.82
CH	4.7k	9.1k



RÉSULTATS FINAUX

- **Aggro:** 30 clusters, 2291 à 69 clients par cluster
- **DBSCAN:** 59 clusters, 4485 à 5 clients par cluster



NEXT STEPS

DONNÉES SUPPLÉMENTAIRES NECESSAIRES

- **Prix des commandes**
- **Historique plus important**

CHOISIR LES RÈGLES DE RECOMMANDATION SELON LES CLUSTERS

- **Identifier les caractéristiques fortes de chaque cluster**

**THANK YOU
FOR YOUR ATTENTION**