

Retail Insight 360

**Optimisation de la satisfaction client
grâce à l'analyse des retours produits.**

Giulia Governatori pour



BestMarket



CONTEXTE

BestMarket souhaite comprendre pourquoi les clients retournent des produits pour mieux gérer ces retours et réduire leur impact.

OBJECTIFS

- Identifier les raisons principales des retours.
- Évaluer l'impact sur la satisfaction client.
- Mettre en place des stratégies pour réduire le taux de retour.

BÉNÉFICES ATTENDUS

- Réduction des coûts liés aux retours.
- Fidélisation accrue des clients.
- Renforcement de l'image de marque.



BONNES PRATIQUES RGPD

Sauvegarde et stockage de la base des données

1. CONTRÔLE DES ACCÈS:

Limitier l'accès aux données personnelles aux personnes autorisées et garantir leur usage conforme.

2. STOCKAGE SÉCURISÉ:

Protéger les données avec des systèmes informatiques sûrs et restreindre l'accès physique.

3. ANONYMISATION:

Anonymiser les données pour respecter la vie privée des consommateurs, et faciliter les analyses de marché, en conformité avec la RGPD.



Ce projet utilise uniquement des jeux de données anonymisées.

MÉTHODOLOGIE



ÉTAPE 1

- chargement de la base de données, depuis fichier .sql sur *DBeaver*

ÉTAPE 2

- individualisation des besoins

ÉTAPE 3

- implémentation des données manquantes (*ref_magasin*)

ÉTAPE 4

- implémentation du dictionnaire et du schéma relationnel avec la nouvelle table

ÉTAPE 5

- écriture des requêtes SQL

ÉTAPE 6

- vérification de la cohérence des données et l'ajout de constraints

MÉTHODOLOGIE D'IMPLEMENTATION DES DONNÉES

fichier.SQL

```
... KEY, typologie_produit CHAR (50), titre_produit CHAR (50));  
...  
... VALUES (4, 'Alimentaire', 'Aliments à...'  
... VALUES (5, 'Alimentaire', 'Aliments à...'  
... VALUES (6, 'Alimentaire', 'Aliments à...'  
... VALUES (7, 'Alimentaire', 'Aliments d...'  
... VALUES (8, 'Alimentaire', 'Aliments et...'  
... VALUES (9, 'Alimentaire', 'Amuse-gueul...'  
... VALUES (10, 'Alimentaire', 'Bananes');  
... VALUES (11, 'Alimentaire', 'Beurres de...'  
... VALUES (12, 'Alimentaire', 'Bières');  
... VALUES (13, 'Alimentaire', 'Biscuits e...
```

Statistics 1

Name	Value
Queries	3149
Updated Rows	3145
Execute time	1s
Fetch time	0,000s
Total time	1s
Start time	2025-01-23 09:35:51.829
Finish time	2025-01-23 09:35:54.207

RETOUR_PRODUIT

Record	Test
>	2998
>	2999
>	3000
>	Views

3000

PRODUIT

Record	Test
>	142
>	143
>	144
>	145
>	Views

145

REF_MAGASIN

Record	Testo
>	82
>	83
>	84

84

ref_magasin.csv compte **84 données**

1	ref_magasin
84	83
85	84
86	

```
CREATE TABLE ref_magasin (  
  ref_magasin INT PRIMARY KEY UNIQUE,  
  departement VARCHAR(3),  
  departement_commune INT(6),  
  libelle_de_commune VARCHAR(50),  
  population INT(12),  
  geo_point_2d FLOAT);
```

CSV chargé sur DBeaver

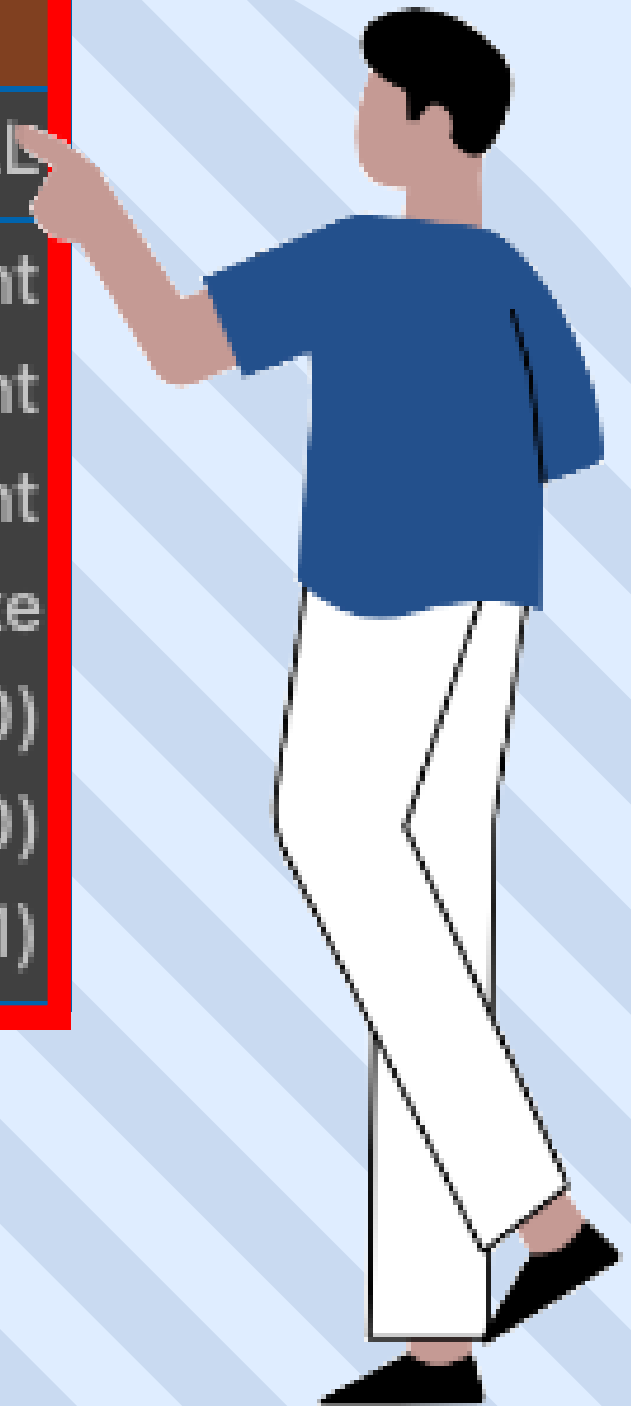
84 row(s) fetched - 0,013s (0,001s fetch), on 2025-01-23 at 09:27:21

SCHÉMA RELATIONNEL

🗄️ produit		
123	cle_produit	int NOT NULL
A-Z	typologie_produit	char(50)
A-Z	titre_produit	char(50)

🗄️ ref_magasin		
123	ref_magasin	int NOT NULL
A-Z	departement	varchar(3)
123	departement_commune	int
A-Z	libelle_de_commune	varchar(50)
123	population	int
123	geo_point_2d	float

🗄️ retour_client		
123	cle_retour_client	int NOT NULL
123	note	int
123	cle_produit	int
123	ref_magasin	int
🕒	date_achat	date
A-Z	libelle_source	char(50)
A-Z	libelle_categorie	char(50)
A-Z	recommandation	char(1)



DICTIONNAIRE DE DONNÉES

	Nom du champs	Type	Taille	PK	Description
retour_client	cle_retour_client	INT		PK	ID unique pour les retours clients
	note	INT			Note donnée par le client, comprise entre 0 et 10, la note est la réponse à la question : "Sur une échelle de 0 à 10 quelle est la probabilité que vous recommandiez notre entreprise à votre entourage ?"
	Clé_produit	INT		FK	ID des produits
	ref_magasin	INT		FK	ID des magasins
	date_achat	DATE			Date à laquelle l'achat du client a eu lieu
	libelle_source	VARCHAR	50		Libellé de la source d'où provient le retour client (Réseaux sociaux, téléphone, email)
	libelle_categorie	VARCHAR	50		Libellé de la catégorie du retour client (Drive, service après-vente, qualité produit, expérience en magasin, livraison)
	recommandation	VARCHAR			Recommandation laissée par le client à la question 'Recommandez vous l'entreprise?' True / False
produit	cle_produit	INT		PK	ID unique pour les produits
	titre_produit	VARCHAR	50		Libellé des produits
	typologie_produit	INT			Typologie des produits (Alimentaire, High-tech etc...)
ref_magasin	ref_magasin	INT		PK	ID des magasins
	departement	VARCHAR	3		Les departements des magasins
	departement_commune	INT	6		les code postaux des magasins
	libelle_de_commune	VARCHAR	50		Les communes des magasins
	population	INT	12		la population de la ville du magasin
	geo_point_2d	FLOAT	100		les coordonnées GPS du magasin



Question 1:

Quel est le nombre de retours clients sur la livraison?

```
SELECT COUNT(*) AS nombre_retours_livraison  
FROM retour_client  
WHERE libelle_categorie = 'livraison';
```

123 nombre_retours_livraison

639

Question 2:

Quelle est la liste des notes des clients sur les réseaux sociaux concernant les TV?

```
SELECT rc.note AS note_TV_rs  
FROM retour_client rc  
JOIN produit p  
ON rc.cle_produit = p.cle_produit  
WHERE rc.libelle_source = 'Réseaux sociaux'  
AND p.titre_produit = 'TV'  
GROUP BY rc.note;
```

123 note_TV_rs

10

9

8

Question 3:

Quelle est la note moyenne pour chaque catégorie de produit?

```
SELECT p.typologie_produit,  
ROUND(AVG(rc.note),2) AS note_moyenne  
FROM produit p  
JOIN retour_client rc ON p.cle_produit = rc.cle_produit  
GROUP BY p.typologie_produit  
ORDER BY note_moyenne DESC;
```

A-Z typologie_produit	123 note_moyenne
High-Tech	8,16
Loisirs	8,09
Alimentaire	8,04
Maison	7,85

Question 4:

Quels sont les 5 magasins avec les meilleures notes moyennes?

```
SELECT rm.libelle_de_commune AS Ville,  
rm.departement AS dep_code,  
ROUND(AVG(rc.note),2) AS note_moyenne  
FROM retour_client rc  
JOIN ref_magasin rm  
ON rc.ref_magasin = rm.ref_magasin  
GROUP BY rc.ref_magasin  
ORDER BY note_moyenne DESC  
LIMIT 5;
```

A-Z Ville	A-Z dep_code	123 note_moyenne
Paris 14e Arrondissement	75	8,73
Saint-Pierre-du-Perray	91	8,55
Paris 19e Arrondissement	75	8,5
Paris 11e Arrondissement	75	8,48
Coulommiers	77	8,45

Question 5:

Quels magasins ont plus de 12 feedbacks sur le drive?

```
SELECT rm.libelle_de_commune AS Ville,  
rm.departement AS dep_code,  
COUNT(rc.recommandation) AS Feedback  
FROM retour_client rc JOIN ref_magasin rm  
ON rc.ref_magasin = rm.ref_magasin  
WHERE LOWER(rc.libelle_categorie) = 'drive'  
GROUP BY rm.libelle_de_commune, rm.departement  
HAVING Feedback > 12;
```

A-Z Ville	A-Z dep_code	123 Feedback
Éragny	95	14
Ivry-sur-Seine	94	13
Paris 12e Arrondissement	75	13

Question 6:

Quel est le classement des départements par note ?

```
SELECT rm.departement, ROUND(AVG(rc.note), 2)  
AS note_moyenne  
FROM retour_client rc  
JOIN ref_magasin rm  
ON rc.ref_magasin = rm.ref_magasin  
GROUP BY departement  
ORDER BY note_moyenne DESC;
```

A-Z departement	123 note_moyenne
95	8,14
75	8,11
94	8,06
91	8,05
77	8,04
92	8,03
78	8,02
93	7,94

Question 7:

Quelle typologie de produit apporte le meilleur service après-vente?

```
SELECT p.typologie_produit,  
ROUND(AVG(rc.note), 2) AS note_moyenne  
FROM retour_client rc  
JOIN produit p  
ON rc.cle_produit = p.cle_produit  
WHERE rc.libelle_categorie = 'service après-vente'  
GROUP BY p.typologie_produit  
ORDER BY note_moyenne DESC  
LIMIT 1;
```

A-Z typologie_produit	123 note_moyenne
Loisirs	8,51

```
SELECT p.typologie_produit,  
COUNT(rc.recommandation) AS nb_recommandations  
FROM produit p  
JOIN retour_client rc  
ON p.cle_produit = rc.cle_produit  
WHERE rc.recommandation = TRUE  
AND rc.libelle_categorie = 'service après-vente'  
GROUP BY p.typologie_produit  
ORDER BY nb_recommandations DESC  
LIMIT 1;
```

A-Z typologie_produit	123 nb_recommandations
Alimentaire	333

...mais sur quelle pourcentage???
(réf. "Question 20")



Question 8:

Quelle est la note moyenne sur l'ensemble des boissons?

```
SELECT ROUND(AVG(rc.note), 2) as note_boissons  
FROM retour_client rc  
JOIN produit p ON rc.cle_produit = p.cle_produit  
WHERE LOWER(p.titre_produit) LIKE 'boisson%';
```

123 note_boissons
8,32

Question 9:

Quel est le classement des jours de la semaine offrant la meilleure expérience client?

```
SELECT DAYNAME(date_achat) AS jour,  
ROUND(AVG(note),2) AS note_moyenne  
FROM retour_client  
WHERE libelle_categorie = 'expérience en magasin'  
GROUP BY jour ORDER BY note_moyenne DESC;
```

A-Z jour	123 note_moyenne
Saturday	8,34
Sunday	8,18
Friday	8,07
Thursday	8,04
Wednesday	7,99
Tuesday	7,95
Monday	7,74

Question 10:

Sur quel mois a-t-on le plus de retours après-vente?

```
SELECT MONTH(date_achat) AS mois,  
COUNT(*) AS nombre_retours  
FROM retour_client  
WHERE libelle_categorie = 'service après-vente'  
GROUP BY mois  
ORDER BY nombre_retours DESC  
LIMIT 1;
```

123 mois	123 nombre_retours
10	55

Question 11:

Quel est le pourcentage de recommandations client?

```
SELECT  
CONCAT(ROUND((COUNT(CASE WHEN recommandation = TRUE THEN 1 END)  
* 100 / COUNT(*)),1), '%') AS "Positive",  
CONCAT(ROUND((COUNT(CASE WHEN recommandation = 0 THEN 1 END)  
* 100 / COUNT(*)),1), '%') AS "Negative"  
FROM  
retour_client;
```

A-Z Positive	A-Z Negative
70.5 %	29.5 %

Calcul 11: Pourcentage

= recommandations positives / total de retours x 100

Question 12:

Quels magasins ont une note inférieure à la moyenne?

```
SELECT rm.libelle_de_commune, rm.departement AS dep,
ROUND(AVG(rc.note),2) AS note_magasin
FROM retour_client rc
JOIN ref_magasin rm
ON rc.ref_magasin = rm.ref_magasin
GROUP BY rc.ref_magasin
HAVING note_magasin < (SELECT AVG(rc.note) FROM retour_client rc)
ORDER BY note_magasin ASC;
```

A-Z libelle_de_commune	A-Z dep	123 note_magasin
Buchelay	78	7,38
Nanterre	92	7,44
Montgeron	91	7,53
Paris 15e Arrondissement	75	7,56
Rosny-sous-Bois	93	7,59
Lognes	77	7,62
Levallois-Perret	92	7,62
Aubervilliers	93	7,66
Fontenay-sous-Bois	94	7,67
Chelles	77	7,7
Herblay	95	7,73
Carrières-sur-Seine	78	7,74
Fleury-Mérogis	91	7,79
Les Ulis	91	7,82
Saint-Germain-lès-Arpajon	91	7,83
Conflans-Sainte-Honorine	78	7,83
Paris 2e Arrondissement	75	7,84
Saint-Denis	93	7,84

Calcul 13: Taux d'évolution
 = (moyenne T2 - moyenne T1)
 / moyenne T1 x 100

Question 13:

Quelles sont les typologies produits qui ont amélioré leur moyenne entre le 1er et le 2ème trimestre 2021?

```
WITH medie AS (
SELECT p.typologie_produit AS categorie,
AVG(CASE WHEN MONTH(rc.date_achat) BETWEEN 1 AND 3 THEN rc.note ELSE NULL END) AS mT1,
AVG(CASE WHEN MONTH(rc.date_achat) BETWEEN 4 AND 6 THEN rc.note ELSE NULL END) AS mT2
FROM produit p JOIN retour_client rc ON p.cle_produit = rc.cle_produit
GROUP BY p.typologie_produit)
SELECT categorie, ROUND(mT1, 2) AS moyenneT1, ROUND(mT2, 2) AS moyenneT2,
-- taux d'evolution = (moyenne T2 - moyenne T1) / moyenne T1 x 100
CONCAT(ROUND(((mT2 - mT1) / mT1 *100), 2), ' %') AS tauxEvolution
FROM medie WHERE mT2 > mT1;
```

A-Z categorie	123 moyenneT1	123 moyenneT2	A-Z tauxEvolution
Alimentaire	7,99	8,06	0.81 %
Loisirs	8	8,34	4.24 %

Question 14:

Quel est le NPS global?

```
SELECT ROUND((promoteurs * 100.0 / tot) - (detracteurs * 100.0 / tot))  
AS NPS_global  
FROM (SELECT  
    SUM(CASE WHEN note > 8 THEN 1 ELSE 0 END) AS promoteurs,  
    SUM(CASE WHEN note <= 6 THEN 1 ELSE 0 END) AS detracteurs,  
    COUNT(note) AS tot  
FROM retour_client) AS stats;
```

123 NPS_global

31

Question 15:

Quel est le NPS par source
(réseaux sociaux, téléphone, etc.)?

```
SELECT source, ROUND((promoteurs * 100.0 / tot) - (detracteurs * 100.0 / tot))  
AS NPS_par_source  
FROM (SELECT  
    SUM(CASE WHEN note > 8 THEN 1 ELSE 0 END) AS promoteurs,  
    SUM(CASE WHEN note <= 6 THEN 1 ELSE 0 END) AS detracteurs,  
    COUNT(note) AS tot,  
    libelle_source AS source  
FROM retour_client  
GROUP BY libelle_source  
) AS stats;
```

A-Z source

123 NPS_par_source

téléphone

34

email

30

réseaux sociaux

30

Calcul 14-15: Net Promoter Score

%promoteurs - %détracteurs

avec promoteurs: note >= 9

et détracteurs: note <= 6

Question 16:

Quel est le nombre de retours clients par source?

```
SELECT libelle_source, COUNT(*) AS nombre_retours
FROM retour_client
GROUP BY libelle_source;
```

A-Z libelle_source	123 nombre_retours
téléphone	970
email	1.032
réseaux sociaux	998

Question 18:

Quelle est la différence entre les feedbacks reçus sur le drive et en magasin?

```
SELECT libelle_categorie,
COUNT(CASE WHEN recommandation = TRUE THEN 1 END) AS fb_positive
FROM retour_client
WHERE libelle_categorie = 'Drive'
OR libelle_categorie LIKE '%magasin'
GROUP BY libelle_categorie;
```

A-Z libelle_categorie	123 fb_positive
expérience en magasin	393
drive	423

Question 17:

Quels sont les 5 magasins avec les plus de feedbacks?

```
SELECT rm.libelle_de_commune, rc.ref_magasin,
COUNT(rc.recommandation) AS reviews
FROM retour_client rc
JOIN ref_magasin rm
ON rc.ref_magasin = rm.ref_magasin
GROUP BY rm.libelle_de_commune, rc.ref_magasin
ORDER BY reviews DESC LIMIT 5;
```

A-Z libelle_de_commune	123 ref_magasin	123 reviews
Mareuil-lès-Meaux	29	55
Osny	6	49
Lognes	80	47
Villecresnes	5	45
Ivry-sur-Seine	63	44

Question 19:

Quel est le classement des notes pour chaque catégorie?

```

WITH Classement AS (
  SELECT
    p.typologie_produit AS TypologieProduit,
    p.titre_produit AS Produit,
    ROUND(AVG(rc.note),2) AS NoteMoyenne,
    ROW_NUMBER() OVER (
      PARTITION BY p.typologie_produit
      ORDER BY AVG(rc.note) DESC
    ) AS Classement
  FROM
    produit p
  JOIN
    retour_client rc
  ON
    p.cle_produit = rc.cle_produit
  GROUP BY
    p.typologie_produit, p.titre_produit
)
SELECT
  TypologieProduit,
  Produit,
  NoteMoyenne,
  Classement
FROM
  Classement
WHERE
  Classement <= 5
ORDER BY
  TypologieProduit, Classement;
  
```

A-Z TypologieProduit	A-Z Produit	123 NoteMoyenne	123 Classement
Alimentaire	Plantes aromatiques surgelées	9,2	1
Alimentaire	Sodas	9	2
Alimentaire	Boissons alcoolisées	8,79	3
Alimentaire	Sauces au soja	8,77	4
Alimentaire	Aliments à base de plantes frais	8,71	5
High-Tech	TV	8,71	1
High-Tech	Smartphones	8,55	2
High-Tech	Tablettes	8,29	3
High-Tech	Son	8,28	4
High-Tech	Objets connectés	8,21	5
Loisirs	DVD, Blu-Ray	8,56	1
Loisirs	Films	8,44	2
Loisirs	Beauté	8,25	3
Loisirs	Jeux	8,08	4
Loisirs	Series TV	7,94	5
Maison	Electroménager	8,52	1
Maison	Puericulture	7,86	2
Maison	Cuisine	7,06	3

Calcul 21: Proportion

= nombre de retours négatifs / population de la ville x 100

Question 20:

Quel est le pourcentage de retour pour chaque typologie de produit?

```
SELECT p.typologie_produit AS Produit,  
COUNT(rc.cle_retour_client) AS Retours,  
CONCAT(ROUND(COUNT(rc.cle_retour_client) /  
(SELECT COUNT(*) FROM retour_client) * 100),  
" %")  
AS "%retours"  
FROM produit p  
JOIN retour_client rc  
ON p.cle_produit = rc.cle_produit  
GROUP BY p.typologie_produit  
ORDER BY "%retours" DESC;
```

A-Z Produit	A-Z %retours
Alimentaire	81 %
High-Tech	10 %
Loisirs	6 %
Maison	2 %

Calcul 20: Pourcentage
= nombre de retours typologie / nombre de retours total x 100

Question 21:

Quel est le nombre de retours négatifs par ville, rapporté à la population?

```
SELECT rm.libelle_de_commune AS Ville, rm.population AS Population_totale,  
COUNT(CASE WHEN rc.recommandation = FALSE THEN 1 END) AS Retours_negatifs,  
ROUND((COUNT(CASE WHEN rc.recommandation = FALSE THEN 1 END) / rm.population) * 100, 2)  
AS Proportion  
FROM retour_client rc  
JOIN ref_magasin rm ON rc.ref_magasin = rm.ref_magasin  
GROUP BY rm.libelle_de_commune, rm.population  
HAVING Retours_negatifs > 0  
ORDER BY Proportion DESC;
```

A-Z Ville	123 Population_totale	123 Retours_negatifs	123 Proportion
Mareuil-lès-Meaux	2.194	14	0,64
Buchelay	2.346	11	0,47
Freneuse	3.930	12	0,31
Ballainvilliers	3.795	9	0,24
Étréchy	6.295	13	0,21
Coignières	4.527	9	0,2
Villecresnes	9.598	16	0,17
Longpont-sur-Orge	6.561	11	0,17
Orgeval	5.855	10	0,17
Chambourcy	5.872	8	0,14
Magny-en-Vexin	5.752	8	0,14
Saint-Pierre-du-Perray	8.350	11	0,13
Lognes	14.560	18	0,12
Cesson	8.572	9	0,1
Paris 1er Arrondissement	17.308	17	0,1
Osny	16.236	14	0,09
Fleury-Mérogis	9.121	8	0,09

VÉRIFICATION DE LA COHÉRENCE DES DONNÉES

1. Cohérence recommandations et notes:

Des incohérences ont été détectées entre les notes, les recommandations et la catégorie "qualité produit".

```
SELECT
  COUNT(*) AS Retours,
  CASE
    WHEN (note < 7 AND recommandation = TRUE)
    THEN 'Incohérence: note basse mais recommandation positive'
    WHEN (note >= 7 AND recommandation = FALSE)
    THEN 'Incohérence: note élevée mais recommandation négative'
    WHEN libelle_categorie = 'qualité produit'
    AND (note < 7 AND recommandation = FALSE)
    THEN 'Incohérence: note basse, recommandation négative mais marqué de bonne qualité'
    ELSE 'Cohérent'
  END AS coherence_status
FROM retour_client
GROUP BY coherence_status
ORDER BY coherence_status
```

123 Retours	A-Z coherence_status
2.150	Cohérent
105	Incohérence: note basse mais recommandation positive
26	Incohérence: note basse, recommandation négative mais marqué de bonne qualité
719	Incohérence: note élevée mais recommandation négative

VÉRIFICATION DE LA COHÉRENCE DES DONNÉES

2. Unicité des clés primaires:

Aucune anomalie n'a été détectée à ce niveau.

```
SELECT cle_produit, COUNT(*) AS total
FROM produit
GROUP BY cle_produit
HAVING total > 1;
```

Constraint PK UNIQUE:

```
ALTER TABLE retour_client
ADD UNIQUE (cle_retour_client);
```

3. Correspondance des clés étrangères:

Aucune clé orpheline n'a été trouvée.

```
SELECT rc.cle_produit
FROM retour_client rc
LEFT JOIN produit p
ON rc.cle_produit = p.cle_produit
WHERE rc.cle_produit IS NULL;
```

Constraint FK et relation UNIQUE:

```
ALTER TABLE retour_client ADD CONSTRAINT fk_cle_produit
FOREIGN KEY (cle_produit) REFERENCES produit(cle_produit);
```

```
ALTER TABLE retour_client
ADD UNIQUE (cle_retour_client, cle_produit);
```

4. Vérification des valeurs des champs:

- a. *recommandation* contient des valeurs vides et de 0.
- b. Les notes des clients sont bien comprises entre 0 et 10.

```
SELECT recommandation, COUNT(*)
FROM retour_client
GROUP BY recommandation
ORDER BY recommandation;
```

A-Z recommandation	123 COUNT(*)
	674
0	211
1	2.115

5. Contrôle des doublons:

Aucun doublon n'a été trouvé.

```
SELECT cle_retour_client, ref_magasin
FROM retour_client
WHERE ref_magasin IS NULL
OR ref_magasin = '';
```

```
SELECT cle_retour_client, cle_produit
FROM retour_client
WHERE cle_produit IS NULL
OR cle_produit = '';
```

6. Relations entre retours, produits et magasins:

chaque retour client est correctement associé à son produit et à son magasin correspondant. Les relations sont cohérentes.

MERCI
pour votre attention,
Giulia Governatori