



BestMarket

Analyse des Données Retours Client

Thomas

BARAU

22/10/2024

1) Contexte et expression du besoin

BestMarket	Objectif	Base de Donnée	Analyser les données
<ul style="list-style-type: none">• <i>Entreprise de Grande Distribution</i>• <i>Missionner pour le Service Client</i>	<ul style="list-style-type: none">• <i>Améliorer l'expérience Client</i>• <i>Améliorer leur fidélité</i>	<ul style="list-style-type: none">• <i>Sauvegarde et Stockage</i>• <i>Méthodologie</i>	<ul style="list-style-type: none">• <i>Requêtes SQL</i>• <i>Net Promoter Score</i>• <i>Cohérence des données</i>

2) Sauvegarde et stockage de la BDD

Stratégie de sauvegarde

- Régulières (Incrémentale quotidienne / complète intervalle régulier)
- Isolées (Maintenir des sauvegardes hors ligne et déconnecté du réseau)
- Tests Fréquents (Tester régulièrement l'intégrité des sauvegardes et la capacité de restauration)

Stockage des données

- Séparation géographique (Stocker copie sauvegarde dans un lieu géographique distinct)
- Sécurisation des sauvegardes (Chiffrer / Stocker dans un environnement protégé)

Accès aux sauvegardes

- Chiffrement des canaux de transmission (Chiffrer les transferts de données sauvegardées)
- Accès restreint (Limiter l'accès aux seuls utilisateurs autorisés)

3) Méthodologie suivie

➤ Analyse des données

	A	B	C	D	E	F	G	H
1	ref_magasin	departement	departement	libelle_de_cc	population	geo_point_2d		
2	1	95	95355	Magny-en-Ve	5752	49.15861756403842	1.810393221286614	
3	2	92	92012	Boulogne-Bil	114205	48.836584313752894	2.239135990584618	
4	3	75	75105	Paris 5e Arro	60938	48.844508659617546	2.349859385560182	
5	4	94	94028	Créteil	89985	48.78374018359733	2.454635304151828	
6	5	94	94075	Villecresnes	9598	48.72086924859528	2.531577414388262	
7	6	95	95476	Osny	16236	49.067658992344654	2.063116784506375	
8	7	75	75120	Paris 20e Arr	196880	48.86318677744551	2.400819826729021	
9	8	93	93001	Aubervilliers	76087	48.91217226264496	2.384455137684569	
10	9	93	93070	Saint-Ouen	47189	48.90980657500511	2.332570422050525	
11	10	77	77067	Cesson	8572	48.56473465883997	2.595244766625485	
12	11	78	78361	Mantes-la-Jc	42969	48.99816653916364	1.693378068205732	
13	12	95	95280	Goussainvill	30996	49.024343017008746	2.466198488664081	
14	13	91	91044	Ballainvillier	3795	48.67132408019351	2.289138745020021	
15	14	75	75117	Paris 17e Arr	169325	48.88733716648682	2.307485559493426	
16	15	75	75116	Paris 16e Arr	171124	48.86039876035177	2.262099559395783	

➤ Dictionnaire de données

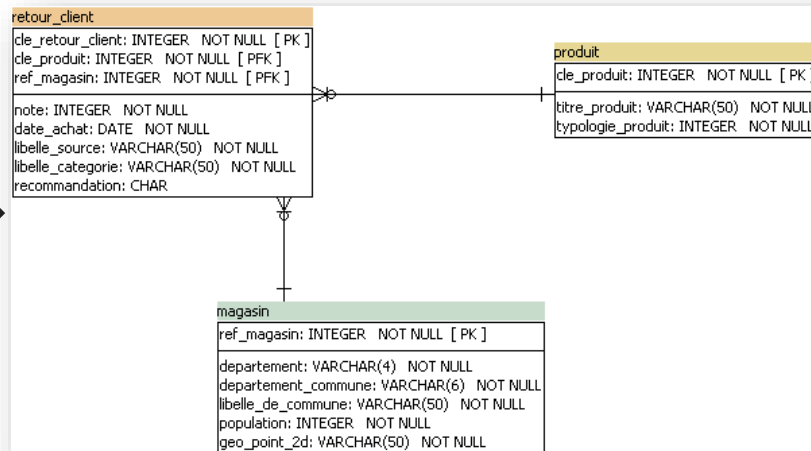
	Nom du champs	Type de données	Taille	Contrainte	
Table Retour client	cle_retour_client	INT		Clé primaire	ID unique pour les retours clients
	note	INT			Note donnée par le client, comprise entre
	cle_produit	INT		Clé étrangère	ID des produits
	ref_magasin	INT		Clé étrangère	ID des magasins
	date_achat	DATE			Date à laquelle l'achat du client a eu lieu
	libelle_source	CHAR	50		Libellé de la source d'où provient le retour
Table Produit	libelle_categorie	CHAR	50		Libellé de la catégorie du retour client (D
	recommandation	CHAR			Recommandation laissée par le client à la
	typologie_produit	CHAR			Typologie des produits (Alimentaire, High
Table Magasin	cle_produit	INT		Clé primaire	ID unique pour les produits
	titre_produit	CHAR	50		Libellé des produits
	ref_magasin	INT		Clé primaire	ID unique pour les magasins
	departement	CHAR	4		identifiant des départements
	departement_commune	CHAR	6		identifiant des communes des départeme
	libelle_de_commune	CHAR	50		Libellé de la commune
	population	INT			Nombre d'habitants de la commune
	geo_point_2d	CHAR	50		Coordonnées du magasin

3) Méthodologie suivie

➤ Dictionnaire de données




















	Nom du champs	Type de données	Taille	Contrainte
Table Retour client	cle_retour_client	INT		Clé primaire ID unique pour les retours clients
	note	INT		Note donnée par le client, comprise entre 1 et 5
	cle_produit	INT		Clé étrangère ID des produits
	ref_magasin	INT		Clé étrangère ID des magasins
	date_achat	DATE		Date à laquelle l'achat du client a eu lieu
	libelle_source	CHAR	50	Libellé de la source d'où provient le retour
	libelle_categorie	CHAR	50	Libellé de la catégorie du retour client (Défaut, Réclamation, etc.)
Table Produit	recommandation	CHAR		Recommandation laissée par le client à la personne qui lui a servi le produit
	cle_produit	INT		Clé primaire ID unique pour les produits
	titre_produit	CHAR	50	Libellé des produits
Table Magasin	typologie_produit	CHAR		Typologie des produits (Alimentaire, High Tech, etc.)
	ref_magasin	INT		Clé primaire ID unique pour les magasins
	departement	CHAR	4	identifiant des départements
	departement_commune	CHAR	6	identifiant des communes des départements
	libelle_de_commune	CHAR	50	Libellé de la commune
	population	INT		Nombre d'habitants de la commune
	geo_point_2d	CHAR	50	Coordonnées du magasin

















➤ Schéma Relationnel






























3) Méthodologie suivie

➤ Base De Données

Structure		Données		Contraintes		Index		Déclencheurs		DDL	
<div></div>											
P5		Nom de la table : magasin						<input type="checkbox"/> WITHOUT ROWID		<input type="checkbox"/> STRICT	
	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL	Collecter			
1	ref_magasin	INT									
2	departement	VARCHAR (4)									
3	departement_commune	VARCHAR (6)									
4	libelle_de_commune	VARCHAR (50)									
5	population	INT									
6	geo_point_2d	VARCHAR (50)									

Structure	Données	Contraintes	Index	Déclencheurs	DDL		
							
							
P5 <input type="text" value="Nom de la table : produit"/> <input type="checkbox"/> WITHOUT ROWID							
	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL
1	cle_produit	INT					
2	typologie_produit	CHAR (50)					
3	titre_produit	CHAR (50)					








Structure		Données	Contraintes	Index	Déclencheurs	DDL	
<div>                </div>							
P5		Nom de la table : retour_client				<input type="checkbox"/> WITHOUT ROWID	
	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL
1	cle_retour_client	INT					
2	note	INT					
3	cle_produit	INT					
4	ref_magasin	INT					
5	date_achat	DATE					
6	libelle_source	CHAR (50)					
7	libelle_categorie	CHAR (50)					
8	recommandation	CHAR					

4) Requêtes SQL et Analyses

Retour Client

N°16. Quel est le nombre de retour clients par source ?

```
SELECT
    libelle_source,
    COUNT(note) as nb_retour
FROM retour_client
GROUP BY libelle_source
ORDER BY nb_retour DESC
```

Table		Formulaire	
		<input checked="" type="checkbox"/>	<input type="checkbox"/>
			
			
1		Nombre de lignes chargées : 3	
	libelle source	nb retour	
1	email	1032	
2	réseaux sociaux	998	
3	téléphone	970	

N°1. Quel est le nombre de retours clients sur la livraison ?

```
SELECT libelle_categorie,
    COUNT(cle_retour_client) AS nombre_retour_client
FROM retour_client
WHERE libelle_categorie = "livraison"
```

libelle_categorie	nombre_retour_client
livraison	639

4) Requêtes SQL et Analyses

Note Moyenne

N°3. Quelle est la note moyenne pour chaque catégorie de produit ?
(Classé de la meilleure à la moins bonne)

```
SELECT
  p.typologie_produit,
  ROUND(AVG(note),2) as moyenne,
  RANK() OVER (
    ORDER BY ROUND(AVG(note),2) DESC
  ) Classement
FROM retour_client rc
JOIN produit p
  ON rc.cle_produit = p.cle_produit
GROUP BY p.typologie_produit
```

High-Tech



Maison



Table		Formulaire	
		Nombre de lignes chargées : 4	
	typologie produit	moyenne	Classement
1	High-Tech	8.16	1
2	Loisirs	8.09	2
3	Alimentaire	8.04	3
4	Maison	7.85	4

N°2. Quelle est la liste des notes des clients sur les réseaux sociaux sur les TV ?

```
SELECT rc.cle_retour_client,
  rc.note,
  p.titre_produit,
  rc.libelle_source
FROM retour_client rc
JOIN
  produit p ON rc.cle_produit = p.cle_produit
WHERE rc.libelle_source LIKE "%sociaux" AND
  p.titre_produit = "TV"
ORDER BY note DESC
```

Table

Formulaire

</

4) Requêtes SQL et Analyses

Note Moyenne

N°8. Quelle est la note moyenne sur l'ensemble des boissons ?

```
SELECT
  AVG(moyenne) as moyenne_boisson
FROM
  (SELECT
    p.cle_produit,
    p.titre_produit,
    ROUND(AVG(note),2) as moyenne
  FROM retour_client rc
  JOIN produit p
    ON p.cle_produit = rc.cle_produit
  WHERE titre_produit like "boisson%"
  GROUP BY p.cle_produit
  ORDER BY moyenne DESC);
```

Table		Formulaire
		Nombre de lignes chargées : 1
moyenne_boisson		
1	8.342	

N°6. Quel est le classement des départements par note ?

```
SELECT
  m.departement,
  ROUND(AVG(rc.note),2) as moyenne_note
FROM retour_client rc
JOIN magasin m
  ON m.ref_magasin = rc.ref_magasin
GROUP BY m.departement
ORDER BY moyenne_note DESC;
```

Table		Formulaire
		Nombre de lignes chargées : 8
departement	moyenne_note	
1 95	8.14	
2 75	8.11	
3 94	8.06	
4 91	8.05	
5 77	8.04	
6 92	8.03	
7 78	8.02	
8 93	7.94	

4) Requêtes SQL et Analyses

Magasins

N°4. Quels sont les 5 magasins avec les meilleures notes moyennes ?

```
SELECT
    rc.ref_magasin,
    m.libelle_de_commune,
    ROUND(AVG(note),2) as moyenne
FROM retour_client rc
JOIN magasin m
    ON m.ref_magasin = rc.ref_magasin
GROUP BY m.ref_magasin
ORDER BY moyenne DESC
LIMIT 5;
```

Table		Formulaire	
		Nombre de lignes chargées : 5	
	ref magasin	libelle de commune	moyenne
1	75	Paris 14e Arrondissement	8.73
2	78	Saint-Pierre-du-Perray	8.55
3	62	Paris 19e Arrondissement	8.5
4	23	Paris 11e Arrondissement	8.48
5	19	Coulommiers	8.45

N°5. Quels sont les magasins qui ont plus de 12 feedbacks sur le drive ?

```
SELECT
    m.libelle_de_commune,
    COUNT(rc.cle_retour_client) as nb_retour
FROM retour_client rc
JOIN magasin m
    ON m.ref_magasin = rc.ref_magasin
WHERE libelle_categorie = "drive"
GROUP BY m.ref_magasin
HAVING nb_retour > 12
ORDER BY nb_retour DESC;
```

Table		Formulaire	
		Nombre de lignes chargées : 3	
	libelle_de_commune	nb_retour	
1	Éragny	14	
2	Ivry-sur-Seine	13	
3	Paris 12e Arrondissement	13	

4) Requêtes SQL et Analyses

Magasins

N°12. Quels sont les magasins qui ont une note inférieure à la moyenne ?

```
SELECT
  rc.ref_magasin,
  m.libelle_de_commune,
  ROUND(AVG(rc.note),2) as note,
  (SELECT ROUND(AVG(note),2) FROM retour_client) as moyenne
FROM retour_client rc
JOIN magasin m
ON rc.ref_magasin = m.ref_magasin
GROUP BY m.ref_magasin
HAVING ROUND(AVG(rc.note),2) < moyenne
ORDER BY note DESC;
```

Nombre de lignes chargées : 37				
	ref magasin	libelle de commune	note	moyenne
1	3	Paris 5e Arrondissement	8.04	8.05
2	47	Freneuse	8.03	8.05
3	59	Paris 1er Arrondissement	8	8.05
4	7	Paris 20e Arrondissement	8	8.05
5	73	Viry-Châtillon	7.97	8.05
6	53	Puteaux	7.97	8.05
7	38	Rueil-Malmaison	7.97	8.05
8	79	Livry-Gargan	7.95	8.05
9	64	Ozoir-la-Ferrière	7.93	8.05
10	45	Paris 12e Arrondissement	7.93	8.05
11	13	Ballainvilliers	7.92	8.05
12	58	Coignières	7.9	8.05
13	14	Paris 17e Arrondissement	7.9	8.05
14	6	Osny	7.9	8.05
15	34	Asnières-sur-Seine	7.89	8.05
16	33	Paris 18e Arrondissement	7.89	8.05
17	20	Provins	7.87	8.05
18	63	Ivry-sur-Seine	7.86	8.05

19	65	Saint-Cyr-l'École	7.85	8.05
20	54	Paris 2e Arrondissement	7.84	8.05
21	51	Saint-Denis	7.84	8.05
22	25	Conflans-Sainte-Honorine	7.83	8.05
23	18	Saint-Germain-lès-Arpajon	7.83	8.05
24	66	Les Ulis	7.82	8.05
25	68	Fleury-Mérogis	7.79	8.05
26	76	Carrières-sur-Seine	7.74	8.05
27	57	Herblay	7.73	8.05
28	74	Chelles	7.7	8.05
29	44	Fontenay-sous-Bois	7.67	8.05
30	8	Aubervilliers	7.66	8.05
31	80	Lognes	7.62	8.05
32	24	Levallois-Perret	7.62	8.05
33	55	Rosny-sous-Bois	7.59	8.05
34	46	Paris 15e Arrondissement	7.56	8.05
35	82	Montgeron	7.53	8.05
36	81	Nanterre	7.44	8.05
37	60	Buchelay	7.38	8.05

4) Requêtes SQL et Analyses

Magasins

N°17. Quels sont les 5 magasins avec le plus de feedbacks

```
SELECT m.ref_magasin,  
       m.departement_commune,  
       m.libelle_de_commune,  
       COUNT(rc.note) AS nb_retour,  
       RANK() OVER (ORDER BY COUNT(rc.note) DESC) as classement  
FROM retour_client rc  
JOIN magasin m USING (ref_magasin)  
GROUP BY ref_magasin  
LIMIT 10
```

	ref_magasin	departement_commune	libelle_de_commune	nb_retour	classement
1	29	77276	Mareuil-lès-Meaux	55	1
2	6	95476	Osny	49	2
3	80	77258	Lognes	47	3
4	5	94075	Villecresnes	45	4
5	83	77294	Mitry-Mory	44	5
6	63	94041	Ivry-sur-Seine	44	5
7	59	75101	Paris 1er Arrondissement	42	7
8	24	92044	Levallois-Perret	42	7
9	22	95539	Saint-Brice-sous-Forêt	42	7
10	19	77131	Coulommiers	42	7

4) Requêtes SQL et Analyses

Temporalité

N°9. Quel est le classement des jours de la semaine où l'expérience client est la meilleure expérience en magasin ?

```
SELECT
  CASE WHEN strftime('%w',date_achat)+1 = 1
    THEN "lundi"
  WHEN  strftime('%w',date_achat)+1 = 2
    THEN "mardi"
  WHEN  strftime('%w',date_achat)+1 = 3
    THEN "mercredi"
  WHEN  strftime('%w',date_achat)+1 = 4
    THEN "jeudi"
  WHEN  strftime('%w',date_achat)+1 = 5
    THEN "vendredi"
  WHEN  strftime('%w',date_achat)+1 = 6
    THEN "samedi"
  WHEN  strftime('%w',date_achat)+1 = 7
    THEN "dimanche"
  END
  as jour,
  ROUND(AVG(note),2) as moyenne_note
FROM retour_client
WHERE libelle_categorie = "expérience en magasin"
GROUP BY strftime('%w',date_achat)
ORDER BY moyenne_note DESC
```

	jour	moyenne_note
1	dimanche	8.34
2	lundi	8.18
3	samedi	8.07
4	vendredi	8.04
5	jeudi	7.99
6	mercredi	7.95
7	mardi	7.74

Dimanche



Mardi



4) Requêtes SQL et Analyses

Temporalité

N°10. Sur quel mois a-t-on le plus de retour sur le service après-vente ?

```
SELECT
CASE WHEN strftime('%m',date_achat) = "01"
      THEN "Janvier"
      WHEN strftime('%m',date_achat) = "02"
      THEN "Février"
      WHEN strftime('%m',date_achat) = "03"
      THEN "Mars"
      WHEN strftime('%m',date_achat) = "04"
      THEN "Avril"
      WHEN strftime('%m',date_achat) = "05"
      THEN "Mai"
      WHEN strftime('%m',date_achat) = "06"
      THEN "Juin"
      WHEN strftime('%m',date_achat) = "07"
      THEN "Juillet"
      WHEN strftime('%m',date_achat) = "08"
      THEN "Août"
      WHEN strftime('%m',date_achat) = "09"
      THEN "Septembre"
      WHEN strftime('%m',date_achat) = "10"
      THEN "Octobre"
      WHEN strftime('%m',date_achat) = "11"
      THEN "Novembre"
      WHEN strftime('%m',date_achat) = "12"
      THEN "Décembre"
      END as mois,
COUNT(cle_retour_client) as nb_retour
FROM retour_client
WHERE libelle_categorie like "service%"
GROUP BY strftime('%m',date_achat)
ORDER BY nb_retour DESC
```

	mois	nb retour
1	Octobre	55
2	Septembre	53
3	Juin	53
4	Novembre	52
5	Août	52
6	Mai	52
7	Avril	52
8	Mars	52
9	Janvier	52
10	Juillet	48
11	Février	44
12	Décembre	38

Octobre



Décembre



4) Requêtes SQL et Analyses

Typologie de Produit

N°7. Quelle est la typologie de produit qui apporte le meilleur service après-vente ?

```
SELECT
  rc.libelle_categorie,
  p.typologie_produit,
  ROUND(AVG(note),2) as moyenne,
  RANK() OVER(
    ORDER BY ROUND(AVG(note),2) DESC
  ) Classement
FROM retour_client rc
JOIN produit p
  ON p.cle_produit = rc.cle_produit
WHERE rc.libelle_categorie like "%service%"
GROUP BY p.typologie_produit
```

Table		Formulaire		
		Nombre de lignes chargées : 4		
	libelle_categorie	typologie_produit	moyenne	Classement
1	service après-vente	Loisirs	8.51	1
2	service après-vente	High-Tech	8.12	2
3	service après-vente	Alimentaire	8.03	3
4	service après-vente	Maison	7.88	4

Loisirs 

Maison 

4) Requêtes SQL et Analyses

Typologie de Produit

N°13. Quelles sont les typologies produits qui ont amélioré leur moyenne entre le 1^{er} et le 2^{ème} trimestre 2021 ?

```
WITH
Trimestre1 as (SELECT p.typologie_produit,
ROUND(AVG(note),2) as moyenne
FROM retour_client rc
JOIN produit p
    USING(cle_produit)
WHERE strftime('%m',rc.date_achat) BETWEEN "01" AND "03"
GROUP BY p.typologie_produit),

Trimestre2 as (SELECT p.typologie_produit,
ROUND(AVG(note),2) as moyenne
FROM retour_client rc
JOIN produit p
    USING(cle_produit)
WHERE strftime('%m',rc.date_achat) BETWEEN "04" AND "06"
GROUP BY p.typologie_produit)

SELECT typologie_produit,
t1.moyenne as note_T1,
t2.moyenne as note_T2,
round(((t2.moyenne -t1.moyenne)/t1.moyenne)*100,2) AS "% progression"
FROM Trimestre1 t1
JOIN Trimestre2 t2
    USING (typologie_produit)
WHERE t2.moyenne > t1.moyenne
```

Table		Formulaire		
		Nombre de lignes chargées : 2		
typologie produit	note T1	note T2	% progression	
1 Alimentaire	7.99	8.06	0.88	
2 Loisirs	8	8.34	4.25	

Alimentaire / Loisirs

Maison/ High-Tech

4) Requêtes SQL et Analyses

Recommandation / NPS

N°11. Quel est le pourcentage de recommandations client ? (Comptabiliser le nombre de retours client qui ont répondu “Oui” divisé par le nombre de retours total)

```
SELECT  
ROUND(  
SUM(rc.recommandation)*100/  
SUM (IIF(rc.recommandation >=0, 1, 0)),2) as pourcentage_reco  
FROM retour_client rc  
ORDER BY pourcentage_reco DESC
```

pourcentage_reco
90.93

Pourcentage de recommandation élevé

4) Requêtes SQL et Analyses

Recommandation / NPS


N°14. Net Promoter Score


NPS = % de Promoteur (note 9-10) – % Détracteur (note 0-6)


```
SELECT  
  
SUM(CASE WHEN note >=9 THEN 1  
ELSE 0  
END)*100/COUNT(note) as promoteur,  
  
SUM(CASE WHEN note <=6 THEN 1  
ELSE 0  
END)*100/COUNT(note) as detracteur,  
  
SUM(CASE WHEN note >=9 THEN 1  
ELSE 0  
END)*100/COUNT(note) -  
SUM(CASE WHEN note <=6 THEN 1  
ELSE 0  
END)*100/COUNT(note) as NPS  
FROM retour_client
```


Table


Formulaire














1







Nombre de lignes chargées : 1

	promoteur	detracteur	NPS
1	40	9	31

NPS = 31



4) Requêtes SQL et Analyses

Recommandation / NPS


N°15. NPS par source


NPS = % de Promoteur (note 9-10) – % Détracteur (note 0-6)


```
SELECT libelle_source,  
       SUM(CASE WHEN note >= 9 THEN 1 ELSE 0 END) * 100 / COUNT(note) AS promoteur,  
       SUM(CASE WHEN note <= 6 THEN 1 ELSE 0 END) * 100 / COUNT(note) AS detracteur,  
       SUM(CASE WHEN note >= 9 THEN 1 ELSE 0 END) * 100 / COUNT(note) - SUM(CASE WHEN note <= 6 THEN 1 ELSE 0 END) * 100 / COUNT(note) AS NPS  
FROM retour_client  
GROUP BY libelle_source  
ORDER BY NPS DESC
```


Table


Formulaire















1







Filtre de données



Nombre de lignes chargées : 3

	libelle_source	promoteur	detracteur	NPS
1	téléphone	41	7	34
2	réseaux sociaux	41	11	30
3	email	37	7	30

Réseaux Sociaux / Email < NPS moyen

5) Cohérence des données



- Pas de retour clients :
 - Magasin : Paris 13^e Arrondissement
 - Produit : Forme / Typo Loisir
- Note 9-10 avec recommandation 0
- Note 0-6 avec recommandation 1
- Nombre de retour équivalent pour chaque magasin alors que la population des communes varie de 1 à 100