



DOCUMENT TECHNIQUE

1 INTRODUCTION

Le présent document technique a pour objectif de fournir une description détaillée des différentes composantes du projet en vue de sa réalisation dans les meilleurs délais et dans le respect des exigences préalablement définies.

Le projet en question concerne la mise en place d'une stratégie qui permettrait d'améliorer la satisfaction client ainsi que le réseau de magasins de l'entreprise. Pour une bonne croissance de l'activité de l'entreprise, il est essentiel de suivre la satisfaction client afin d'optimiser ses performances opérationnelles.

Ce document s'adresse aux équipes techniques en charge de la réalisation du projet, ainsi qu'au responsable du service client et aux parties prenantes concernées. Il vise à établir une vision claire du projet et à fournir les informations nécessaires pour sa mise en œuvre.

Pour rappel, le système de gestion de base de données relationnelle (SGBDR) utilisé dans le cadre de ce projet est SQLite Studio version 3.4.4.

2 NETTOYAGE DES DONNEES

Avant l'importation des données de la table **ref_magasin** dans le SGBDR, elles ont été préalablement été nettoyées dans Power Query :

- La table a été renommée en « **magasin** » ;
- L'extension du fichier a été modifiée en « **CSV** » et le type d'encodage des caractères est « **UTF-8** » ;
- Les champs ont été fractionnés et renommés ;
- Une recherche des doublons, cellules vides et espaces inutiles a ensuite été faite et cela n'a renvoyé aucun résultat.

Après cette étape, les données ont pu être importées dans le SGBDR pour l'analyse.

3 IMPORTATION DES DONNEES DANS LE SGBDR

La base de données « **BESTMARKET** » a été créée dans le SGBDR afin de pouvoir importer le fichier « **CustomerDataFeedback.sql** » qui constitue les données de l'entreprise à analyser. Une requête a été exécutée dans l'interpréteur de commandes Windows afin de pouvoir importer le jeu de données. Ce fichier comporte deux tables :

- « **Produit** » qui compte les informations sur les types de produits que proposent les magasins ;



- « **Retour_client** » qui renseigne sur les différentes sources d'informations et les notes des clients.

COMMANDE POUR L'IMPORTATION DU CustomerDataFeedback

```
Interpréteur de commandes Windows - sqlite3 BESTMARKET.db
Microsoft Windows [version 10.0.19045.3324]
(c) Microsoft Corporation. Tous droits réservés.

C:\Windows\System32> cd C:\Sqlite

C:\Sqlite> sqlite3 BESTMARKET.db
SQLite version 3.42.0 2023-05-16 12:36:15
Enter ".help" for usage hints.
sqlite> .databases
main: C:\Sqlite\BESTMARKET.db r/w
sqlite> .read CustomerDataFeedback.sql
sqlite>
```

Après cette étape, la table « **magasin** » qui comporte des informations sur la localisation des magasins a été importée à partir du SGBDR en précisant les caractéristiques comme sur la capture d'écran suivante :

- Le type de données source « **CSV** » ;
- Le type d'encodage « **UTF-8** » ;
- La première ligne représente les noms des colonnes ;
- Le séparateur est le **point-virgule**.

IMPORTATION DE LA TABLE « magasin »

← Import données

Source de données à importer de

Type de données source

CSV

Options

Fichier : sktop/Formation BI_Analyst_OP/Projets/Projet_5/Documents/magasin.csv

Texte codé : UTF-8

☒ Ignorer les erreurs

Options de source de données

☒ La première ligne représente les noms de colonnes CSV

Séparateur de colonnes : ; (point virgule)

☐ Valeurs NULL :

☒ Interpréter * comme un guillemet

Cancel Finish

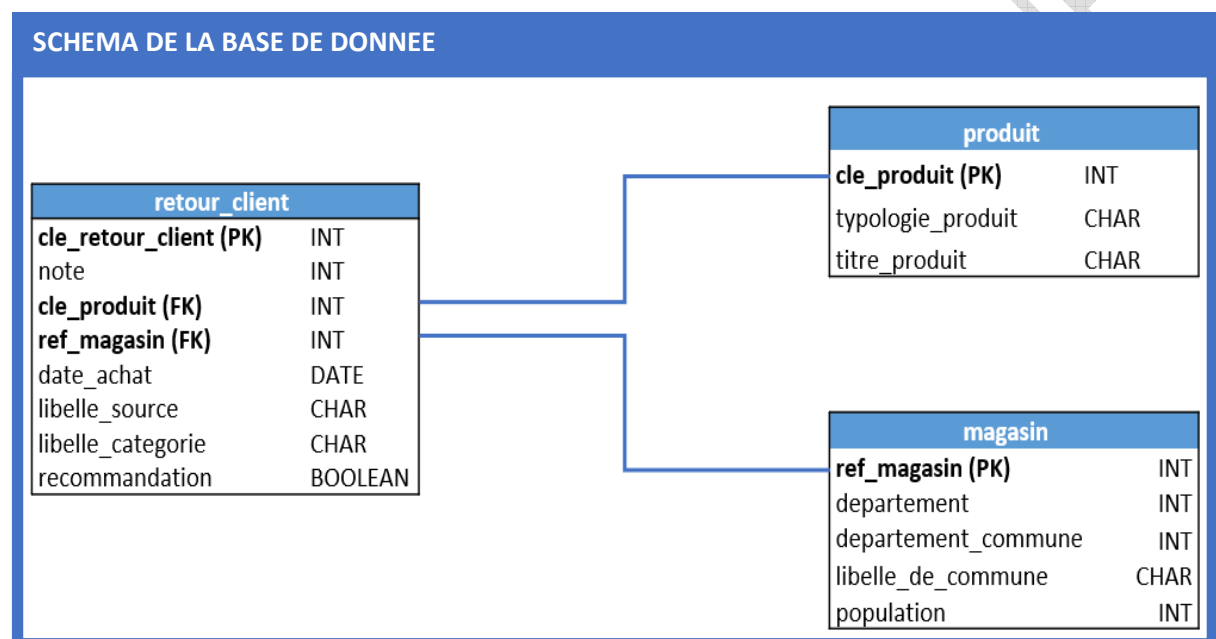


4 SCHEMA DE LA BASE DE DONNEES

Le jeu de données comporte 03 tables :

- ✚ La table « produit » a pour clé primaire la caractéristique « **cle_produit** » ;
- ✚ La table « magasin » a pour clé primaire la caractéristique « **ref_magasin** » ;
- ✚ La table « retour_magasin » a pour clé primaire la caractéristique « **cle_retour_client** » et **clés étrangères** « **cle_produit** » qui référence la table produit et « **ref_magasin** » qui référence la table magasin.

Le schéma de la base de données des différentes tables est le suivant :



5 DICTIONNAIRE DE DONNEES

Les types de données présents dans notre BD sont les suivants :

- **INT (INTEGER)** : utilisé pour stocker des nombres entiers (non décimaux) qu'ils soient positifs ou négatifs ;
- **CHAR (CHARACTER)** : utilisé pour stocker des chaînes de caractères fixes (dont le nombre de caractères ne varie pas) ;
- **DATE** : utilisé pour stocker les données de type date/heure
- **BOOLEAN** : utilisé pour les données qui prennent deux valeurs possibles : « OUI » ou « NON », « VRAI » ou « FAUX », 0 ou 1.

Ces différents champs sont présentés dans le tableau ci-dessous :



DICTIONNAIRE DES DONNEES

Nom du champs	Type de données	Taille	Contrainte	Description	Table
cle_retour_client	INT		Clé primaire (PK)	Id unique pour les retours clients	retour_client
note	INT		NOT NULL	Note donnée par le client, comprise entre 0 et 10, à la question : "Sur une échelle de 0 à 10 quelle est la probabilité que vous recommandiez notre entreprise à votre entourage ?"	
cle_produit	INT		Clé étrangère (FK)	Id unique pour les produits, référence la table produit	
ref_magasin	INT		Clé étrangère (FK)	Id unique pour chaque magasin, référence la table magasin	
date_achat	DATE		NOT NULL	Date à laquelle l'achat du client a eu lieu	
libelle_source	CHAR	50	NOT NULL	Libellé de la source d'où provient le retour client (Réseaux sociaux, téléphone, email)	
libelle_categorie	CHAR	50	NOT NULL	Libellé de la catégorie du retour client (Drive, service après-vente, qualité produit, expérience en magasin, livraison)	
recommandation	BOOLEAN			Recommandation laissée par le client à la question 'Recommandez vous l'entreprise?' True / False	
cle_produit	INT		Clé primaire (PK)	Id unique pour les produits	produit
titre_produit	CHAR	50	NOT NULL	Libellé des produits	
typologie_produit	CHAR	50	NOT NULL	Typologie des produits (Alimentaire, High-tech etc...)	
ref_magasin	INT		Clé primaire (PK)	Id unique pour chaque magasin	magasin
departement	INT		NOT NULL	Code du departement de l'adresse du magasin	
departement_commune	INT		NOT NULL	Code de la commune du magasin	
libelle_commune	CHAR	50	NOT NULL	Libellé de la commune de l'adresse du magasin	
population	INT		NOT NULL	Population du département	

6 BASE DE DONNEES CHARGEE

Les différentes tables de la base de données « BESTMARKET » et les données qu'elles contiennent sont présentées dans les figures ci-dessous :

- Présentation de la table « produit » dans le SGBDR :

CHARGEMENT DE LA TABLE « produit »			
<div> <div> <div>Base de données</div> <div> <div>Tableaux (3)</div> <div> <div>magasin</div> <div>produit</div> <div>Colonnes (3)</div> <div> <div>cle_produit</div> <div>typologie_produit</div> <div>titre_produit</div> </div> <div>Index</div> <div>Déclencheurs</div> <div>retour_client</div> <div>Vues</div> </div> </div> </div> <div> <div>Données</div> <div>Structure</div> <div>Contraintes</div> <div>Index</div> <div>Déclencheurs</div> <div>DDL</div> </div> <div> <div>Table</div> <div>Formulaire</div> </div> <div> <div>de_produit</div> <div>typologie_produit</div> <div>titre_produit</div> </div> <div> <div>1</div><div>1</div><div>Alimentaire</div><div>Additifs alimentaires</div> </div> <div> <div>2</div><div>2</div><div>Alimentaire</div><div>Aliments à base de fruits et de légumes</div> </div> <div> <div>3</div><div>3</div><div>Alimentaire</div><div>Aliments à base de plantes en conserve</div> </div> <div> <div>4</div><div>4</div><div>Alimentaire</div><div>Aliments à base de plantes frais</div> </div> <div> <div>5</div><div>5</div><div>Alimentaire</div><div>Aliments à base de plantes séchées</div> </div> <div> <div>6</div><div>6</div><div>Alimentaire</div><div>Aliments à base de plantes surgelés</div> </div> <div> <div>7</div><div>7</div><div>Alimentaire</div><div>Aliments d'origine végétale</div> </div> <div> <div>8</div><div>8</div><div>Alimentaire</div><div>Aliments et boissons à base de végétaux</div> </div> <div> <div>9</div><div>9</div><div>Alimentaire</div><div>Amuse-gueules</div> </div> <div> <div>10</div><div>10</div><div>Alimentaire</div><div>Bananes</div> </div> <div> <div>11</div><div>11</div><div>Alimentaire</div><div>Beurres de cacahuètes</div> </div> <div> <div>12</div><div>12</div><div>Alimentaire</div><div>Bières</div> </div> <div> <div>13</div><div>13</div><div>Alimentaire</div><div>Biscuits et gâteaux</div> </div> <div> <div>14</div><div>14</div><div>Alimentaire</div><div>Boissons</div> </div> <div> <div>15</div><div>15</div><div>Alimentaire</div><div>Boissons alcoolisées</div> </div> </div>			



STRUCTURE DE LA TABLE « produit »

Base de données: BESTMARKET (SQLite 3)

Tableaux (3): magasin, produit, retour_client

Colonnes (3): cle_produit, typologie_produit, titre_produit

Index, Déclencheurs, retour_client, Vues

Données Structure Contraintes Index Déclencheurs DDL

Nom de la table : produit

	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL	Collecter	Généré	Valeur par défaut
1	cle_produit	INT	🔑				🚫			NULL
2	typologie_produit	CHAR (50)					🚫			NULL
3	titre_produit	CHAR (50)					🚫			NULL

- Présentation de la table « retour_client » dans le SGBDR :

CHARGEMENT DE LA TABLE « retour_client »

Base de données: BESTMARKET (SQLite 3)

Tableaux (3): magasin, produit, retour_client

Colonnes (8): cle_retour_client, note, cle_produit, ref_magasin, date_achat, libelle_source, libelle_categorie, recommandation

Index, Déclencheurs, Vues

Données Structure Contraintes Index Déclencheurs DDL

Table Formulaire

clé de données

	cle_retour_client	note	cle_produit	ref_magasin	date_achat	libelle_source	libelle_categorie	recommandation
1	157263	0	115	5	2021-03-24	email	drive	0
2	157288	0	119	20	2021-06-08	email	drive	0
3	157310	0	80	36	2021-08-21	email	livraison	0
4	157364	0	47	32	2021-06-04	email	drive	0
5	157907	0	63	57	2021-11-16	email	service après-vente	0
6	157919	0	135	55	2021-04-30	email	drive	0
7	157947	0	39	80	2021-07-11	email	qualité produit	0
8	158792	0	129	46	2021-05-19	email	service après-vente	0
9	158847	0	2	82	2021-05-31	email	livraison	0
10	158909	0	79	64	2021-10-10	email	service après-vente	1
11	158925	0	108	51	2021-03-22	email	service après-vente	0
12	158947	0	36	58	2021-03-13	email	livraison	0
13	158981	0	93	57	2021-01-06	email	livraison	1
14	159342	0	142	59	2021-05-12	email	qualité produit	0
15	159694	0	144	44	2021-09-29	email	qualité produit	0

STRUCTURE DE LA TABLE « retour_client »

Base de données: BESTMARKET (SQLite 3)

Tableaux (3): magasin, produit, retour_client

Colonnes (8): cle_retour_client, note, cle_produit, ref_magasin, date_achat, libelle_source, libelle_categorie, recommandation

Index, Déclencheurs, Vues

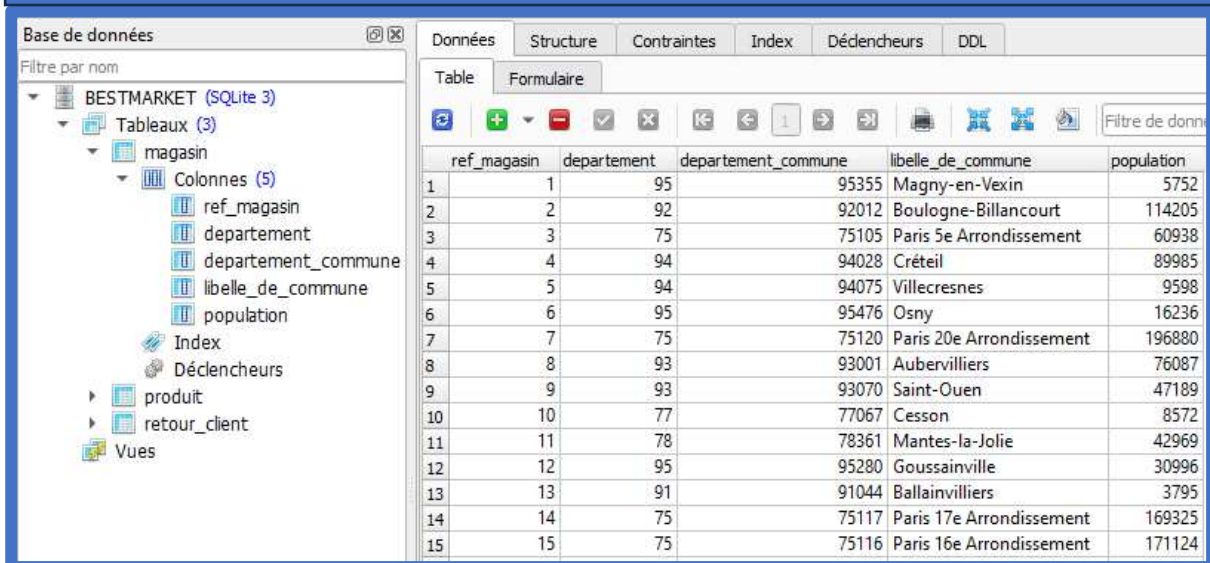
Données Structure Contraintes Index Déclencheurs DDL

Nom de la table : retour_client

	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL	Collecter	Généré	Valeur par défaut
1	cle_retour_client	INT	🔑				🚫			NULL
2	note	INT					🚫			NULL
3	cle_produit	INT		🔗			🚫			NULL
4	ref_magasin	INT		🔗			🚫			NULL
5	date_achat	DATE					🚫			NULL
6	libelle_source	CHAR (50)					🚫			NULL
7	libelle_categorie	CHAR (50)					🚫			NULL
8	recommandation	CHAR					🚫			NULL

- Présentation de la table « magasin » dans le SGBDR :

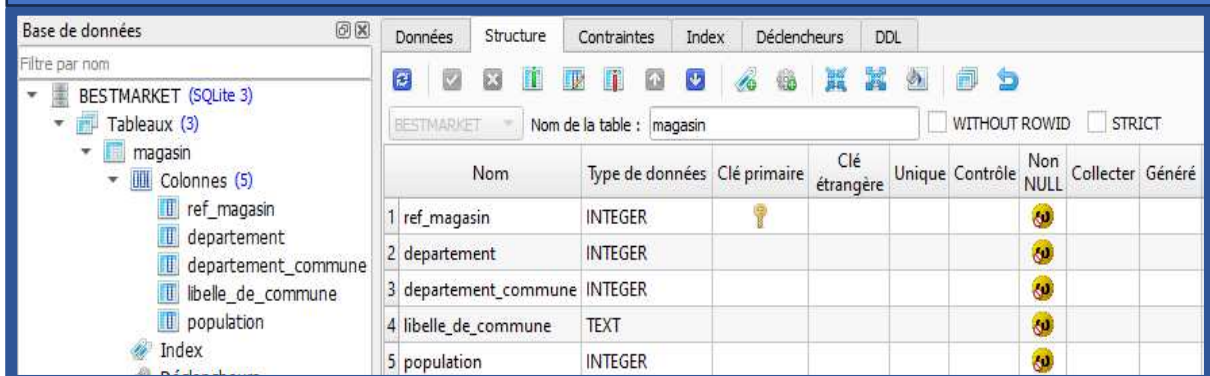
CHARGEMENT DE LA TABLE « magasin »



The screenshot shows the 'magasin' table in the SGBDR interface. The table has 5 columns: ref_magasin, departement, departement_commune, libelle_de_commune, and population. The data is displayed in a grid with 15 rows.

	ref_magasin	departement	departement_commune	libelle_de_commune	population
1	1	95	95355	Magny-en-Vexin	5752
2	2	92	92012	Boulogne-Billancourt	114205
3	3	75	75105	Paris 5e Arrondissement	60938
4	4	94	94028	Créteil	89985
5	5	94	94075	Villecresnes	9598
6	6	95	95476	Osny	16236
7	7	75	75120	Paris 20e Arrondissement	196880
8	8	93	93001	Aubervilliers	76087
9	9	93	93070	Saint-Ouen	47189
10	10	77	77067	Cesson	8572
11	11	78	78361	Mantes-la-Jolie	42969
12	12	95	95280	Goussainville	30996
13	13	91	91044	Ballainvilliers	3795
14	14	75	75117	Paris 17e Arrondissement	169325
15	15	75	75116	Paris 16e Arrondissement	171124

STRUCTURE DE LA TABLE « magasin »



The screenshot shows the 'Structure' tab of the 'magasin' table in the SGBDR interface. It displays the table's schema with columns, data types, and constraints.

Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL	Collecter	Généré
1 ref_magasin	INTEGER	✓				✓		
2 departement	INTEGER					✓		
3 departement_commune	INTEGER					✓		
4 libelle_de_commune	TEXT					✓		
5 population	INTEGER					✓		

7 COHERENCE DES DONNEES

Afin de garantir la cohérence des données au cours de ce projet, les actions suivantes ont été menées :

- ✚ **Respecter les principes du RGPD** tels que la **minimisation des données**, la **protection des données personnelles** et l'**anonymisation** : en effet seules les données strictement nécessaires à la finalité de l'analyse ont été retenues. Les données portant sur les coordonnées géographiques (latitude et longitude) par exemple ont été supprimées. Aucune donnée personnelle (nom, prénom, sexe, etc.) des clients n'a été collectée et les données ont été traitées de sorte qu'on ne puisse pas identifier un client en croisant les informations par exemple en ne gardant que le libellé de la commune et le département au lieu de l'adresse complète.



- ✚ **Rechercher les doublons et les valeurs aberrantes** : identifiez et supprimez les enregistrements en double ou les valeurs aberrantes qui pourraient fausser les résultats de l'analyse ;
- ✚ **Utiliser des contraintes d'intégrité** : Afin de garantir que les relations entre les tables sont respectées, nous avons défini une contrainte de clé étrangère pour assurer qu'une valeur dans une table existe également dans une autre table ;
- ✚ **S'assurer de l'exactitude de la typologie des données** : afin d'optimiser la qualité des données et interpréter correctement les résultats ;
- ✚ **Vérifier que les résultats obtenus correspondent aux résultats attendus** : par exemple se rassurer que les pourcentages soit bien compris entre 0 et 100, le NPS compris entre -100 et +100.