

# 1 - Document technique



*Parcours Data Analyst - Marie G.*  
*Projet 3 : Requêtez une base de données avec SQL*  
*19/02/2025*

1. Exploration des données

2. Schéma relationnel de la base de données

3. Code SQL générant les tables

4. Création des tables

5. Chargement des données

## 1. Exploration des données 1/2

2 fichiers csv sont fournis dans le cadre du projet pour construire notre base de données :

- Contrat+(4).csv : Liste des contrats d'assurance logement de l'entreprise (30 335 lignes de contrats)

Extraction des 10 premières lignes :

Contrat_ID	No_voie	B_T_Q	Type_de_voie	Voie	Code_dep_code_com_mune	Code_post_al	Surface	Type_local	Occupation	Type_contrat	Formule	Valeur_declaree_biens	Prix_cotisation_mensuel
100601	190	A	RUE	CENTRALE	1350	1370	50	Appartement	Locataire	Residence principale	Integral	0-25000	25
100602	347		RUE	DU CHATEAU	1103	1170	48	Appartement	Locataire	Residence principale	Classique	0-25000	30
100603	58		AV	DU MONT BLANC	1143	1220	131	Appartement	Proprietaire	Residence principale	Integral	25000-50000	57
100604	140		RUE	DE L'ABBE JOLIVET	1288	1630	109	Maison	Locataire	Residence principale	Integral	25000-50000	43
100605	39		RUE	BUFFON	1033	1200	109	Appartement	Locataire	Residence principale	Classique	0-25000	33
100606	8		RUE	DE GENEVE	1354	1630	53	Appartement	Proprietaire	Residence principale	Classique	0-25000	19
100607	2		RUE	DU RECULET	1354	1630	59	Appartement	Proprietaire	Residence principale	Integral	0-25000	15
100608	1403		RUE	JEAN DE GINGINS	1143	1220	93	Maison	Proprietaire	Mise en location	Integral	25000-50000	34
100609	226		ALL	DES CAPUCINES	1354	1630	117	Maison	Proprietaire	Residence principale	Classique	25000-50000	32
100610	276		RTE	DE POUIGNY	1288	1630	36	Appartement	Proprietaire	Residence principale	Integral	25000-50000	22

- Region+(7).csv : Liste des communes de France, dont les données sont issues d'un site gouvernemental (38 916 lignes de communes) Certains codes communes issus d'anciens découpages administratifs sont obsolètes, comme ceux de l'ancien département de la Seine comprenant la ville de Paris et les communes de petite couronne.

Extraction des 10 premières lignes :

Code_dep_code_com_mune	reg_co_de	reg_nom	aca_nom	dep_no_m	com_nom_maj_court	dep_c_ode	dep_nom_n_om
1001	84	Auvergne-Rhône-Alpes	Lyon	Ain	LABERGEMENT CLEMENCIAT	1	Ain (01)
1002	84	Auvergne-Rhône-Alpes	Lyon	Ain	LABERGEMENT DE VAREY	1	Ain (01)
1003	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMAREINS	1	Ain (01)
1004	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBERIEU EN BUGEY	1	Ain (01)
1005	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBERIEUX EN DOMBES	1	Ain (01)
1006	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBLEON	1	Ain (01)
1007	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBRONAY	1	Ain (01)
1008	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBUTRIX	1	Ain (01)
1009	84	Auvergne-Rhône-Alpes	Lyon	Ain	ANDERT ET CONDON	1	Ain (01)
1010	84	Auvergne-Rhône-Alpes	Lyon	Ain	ANGLEFORT	1	Ain (01)

## 1. Exploration des données 2/2

Le dictionnaire de données suivant est construit en analysant sous Excel les données des deux fichiers, après création de tableaux pour une meilleure vision de l'ensemble des données pouvant être stockée dans chaque colonne.

Il reprend les éléments qui seront nécessaire pour construire la base de donnée SQL, ainsi qu'une description en documentation annexe

	Nom des colonnes	Type de données	Taille	Clé	Description
CONTRAT.CSV	Contrat_ID	INT	5	Clé primaire	Id unique pour les contrats
	No_voie	INT	4		Numéro dans la voie pour l'adresse du logement assuré
	B_T_Q	CHAR	1		Indicateur éventuel de répétition pour l'adresse du logement assuré sur un caractère
	Type_de_voie	VARCHAR	4		Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ...
	Voie	VARCHAR	100		Libellé de la voie pour l'adresse du logement assuré
	Code_dep_code_commune	VARCHAR	6	Clé secondaire	Concaténation du code département et code commune pour avoir une clé commune unique
	Code_postal	VARCHAR	5		Code postal pour l'adresse du logement assuré
	Surface	INT	4		Surface du logement assuré (m²)
	Type_local	VARCHAR	100		Type de logement (appartement ou maison)
	Occupation	VARCHAR	100		Occupant (locataire ou propriétaire)
	Type_contrat	VARCHAR	100		Occupation (Résidence principale, résidence secondaire ou mise en location)
	Formule	VARCHAR	100		Formule choisie comme assurance (classique ou intégrale)
	Valeur_declaree_biens	VARCHAR	100		Valeur des biens déclarés à l'assurance (€)
REGION.CSV	Prix_cotisation_mensuel	INT	4		Cotisation mensuelle de l'assuré (€)
	Code_dep_code_commune	VARCHAR	6	Clé primaire	Concaténation du code département et code commune pour avoir une clé commune unique
	reg_code	INT	2		Code région
	reg_nom	VARCHAR	100		Nom de la région
	aca_nom	VARCHAR	100		Nom de l'académie
	dep_nom	VARCHAR	100		Nom du département
	com_nom_maj_court	VARCHAR	100		Nom de la commune
	dep_code	VARCHAR	3		Code département
	dep_nom_num	VARCHAR	100		Nom et code du département

Dictionnaire des données de la future base de données

## 2. Schéma relationnel de la base de données 1/2

Le dictionnaire de données présenté à la page précédente permet de modifier le schéma relationnel fourni dans le logiciel **SQL Power Architect**



Points d'attention :

1. SQL Power Architect ne permet pas de donner une taille aux données de type INTEGER. La taille des données renseignée dans le dictionnaire construit devra donc être ajoutée dans la commande de création de tables extraite par SQL Power Architect
2. Les colonnes doivent respecter l'ordre des colonnes fichiers csv prévus pour l'import de données

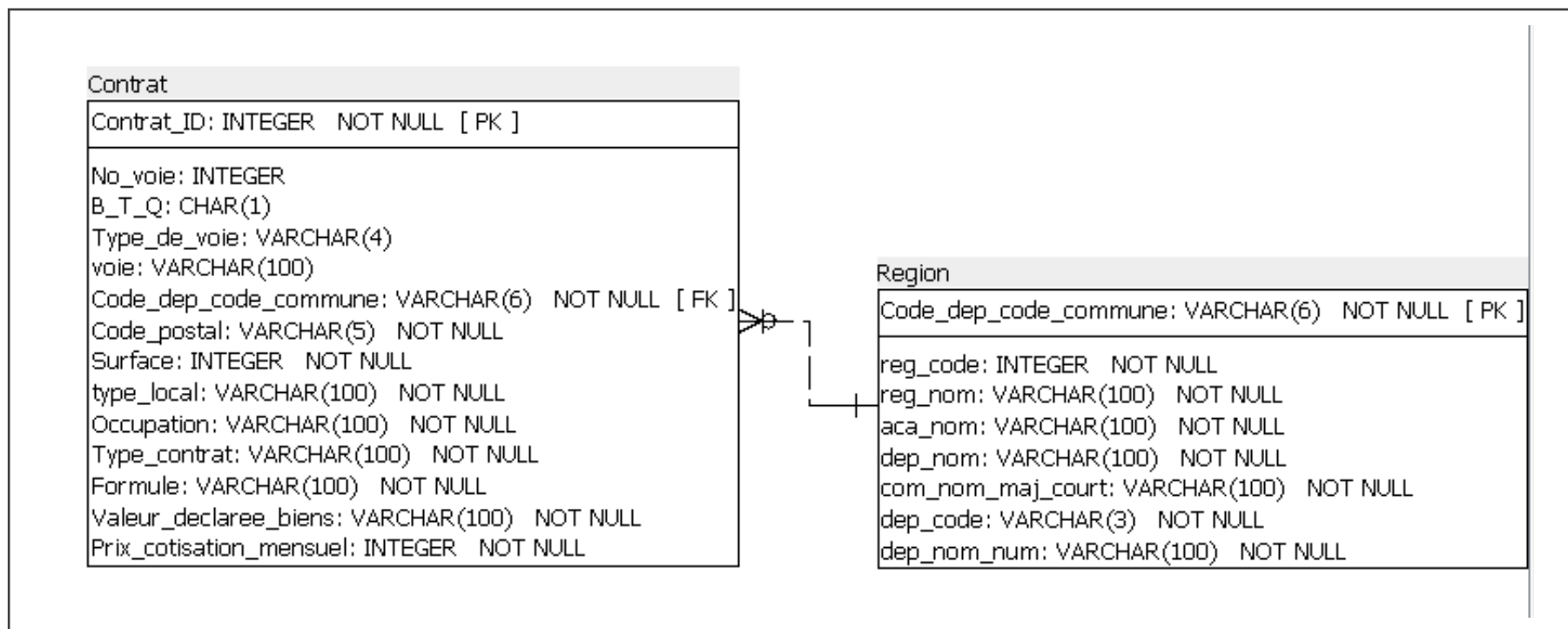


Schéma relationnel de la base de données

## 2. Schéma relationnel de la base de données 2/2

### Principales commandes utilisées pour configurer les tables

#### Ajout des colonnes une par une dans chaque table

Column Properties of New Column

Source for ETL Mapping  
None Specified

Logical Name  
Type\_contrat

Physical Name  
Type\_contrat

☐ In Primary Key

Type  
VARCHAR

Precision 100 Scale 0

Allows Nulls  
☒ No

Auto Increment  
☐ No

Default Value

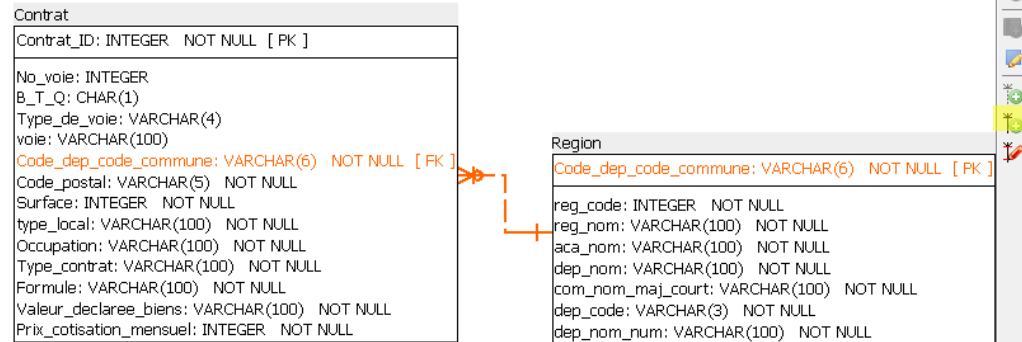
Sequence Name (Only applies to target platforms that use sequences)  
Contrat\_Type\_contrat\_seq

Remarks

OK Cancel

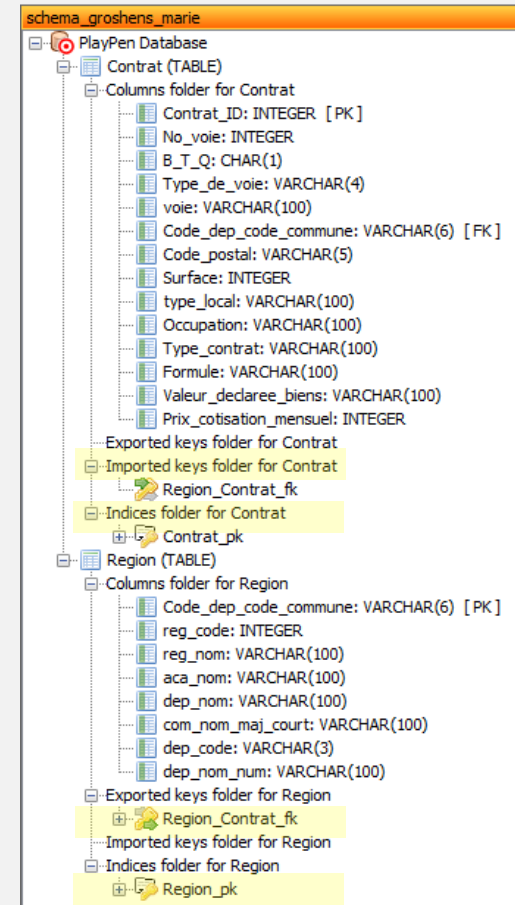
En jaune : les champs utilisés ici

#### Configuration de la clé secondaire (Foreign Key) reliant la table « Contrat » à la table « Région »



Sélection de l'icône sur la droite puis sélection des champs à relier (en jaune)

#### Arborescences des tables :



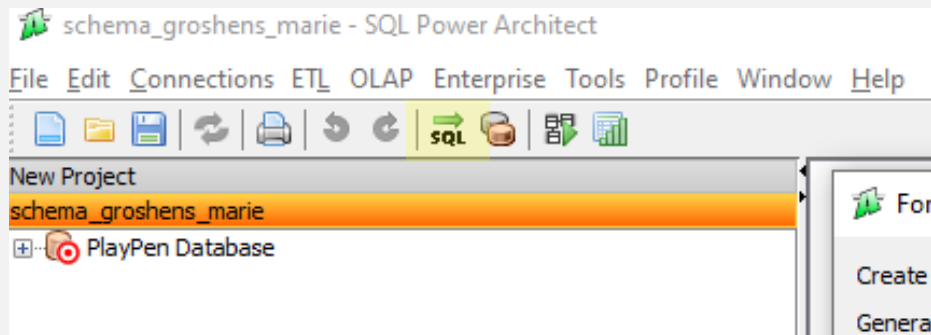
- liste des colonnes pour chaque table
- clés primaires dans le dossier « indices »
- clé secondaire dans les dossiers « imported/exported keys »

### 3. Code SQL générant les tables

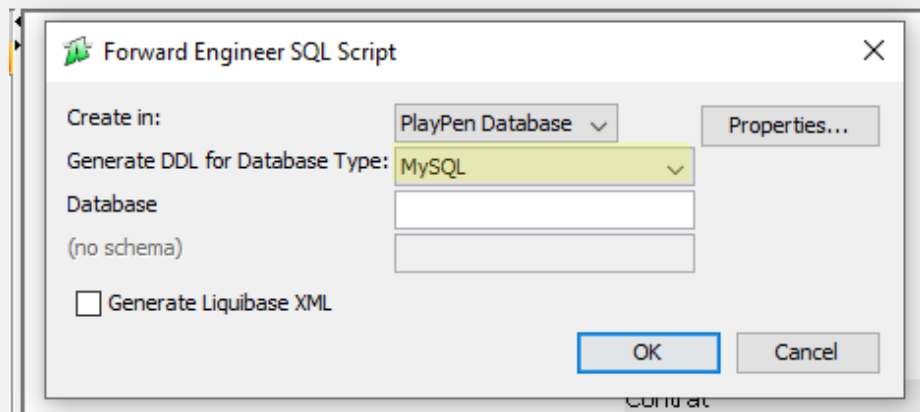
SQL Power Architect permet de générer le code ci-contre sous le choix de format MySQL, après :

- Ajout de la taille des champs de format INTEGER
- Mise en forme sous <https://www.dpriver.com/pp/sqlformat.htm>

#### Bouton permettant l'ouverture d'une fenêtre de dialogue pour l'export SQL



#### Sélection des champs permettant la génération du code sous format texte




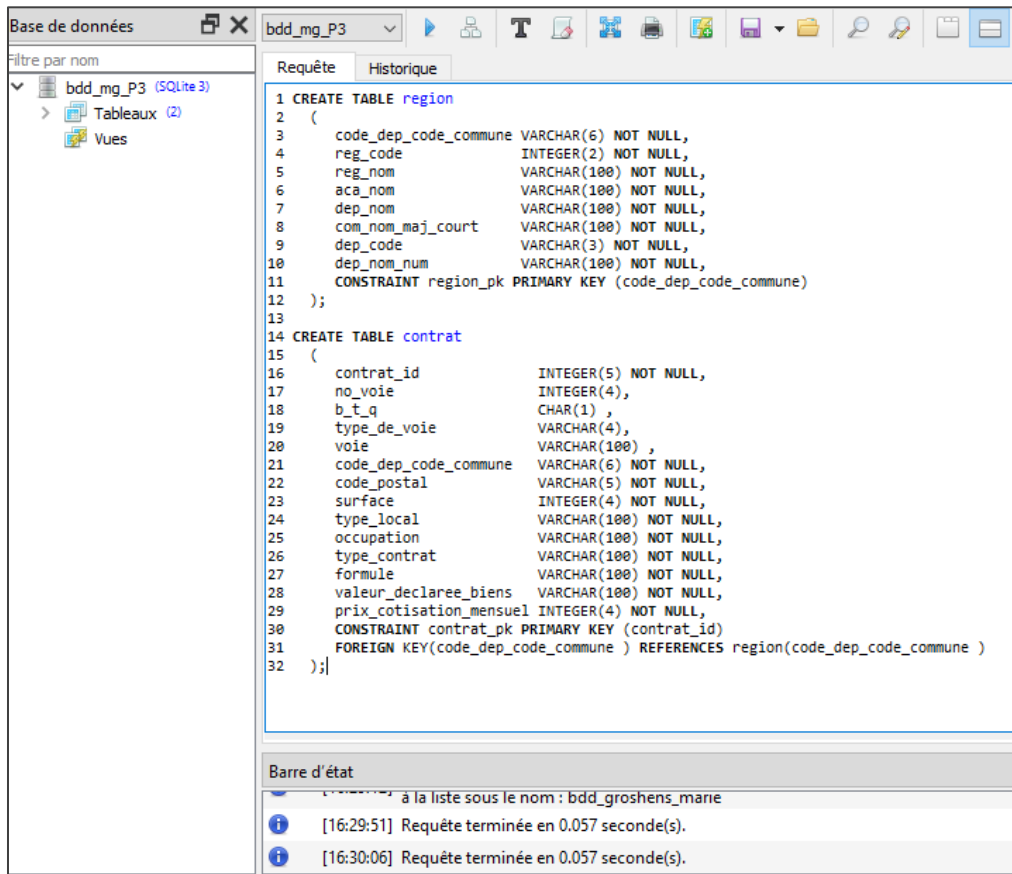
```
CREATE TABLE region
(
  code_dep_code_commune VARCHAR(6) NOT NULL,
  reg_code               INTEGER(2) NOT NULL,
  reg_nom                 VARCHAR(100) NOT NULL,
  aca_nom                 VARCHAR(100) NOT NULL,
  dep_nom                 VARCHAR(100) NOT NULL,
  com_nom_maj_court       VARCHAR(100) NOT NULL,
  dep_code                VARCHAR(3) NOT NULL,
  dep_nom_num             VARCHAR(100) NOT NULL,
  CONSTRAINT region_pk PRIMARY KEY (code_dep_code_commune)
);

CREATE TABLE contrat
(
  contrat_id              INTEGER(5) NOT NULL,
  no_voie                 INTEGER(4),
  b_t_q                   CHAR(1),
  type_de_voie            VARCHAR(4),
  voie                   VARCHAR(100),
  code_dep_code_commune   VARCHAR(6) NOT NULL,
  code_postal              VARCHAR(5) NOT NULL,
  surface                 INTEGER(4) NOT NULL,
  type_local              VARCHAR(100) NOT NULL,
  occupation              VARCHAR(100) NOT NULL,
  type_contrat            VARCHAR(100) NOT NULL,
  formule                 VARCHAR(100) NOT NULL,
  valeur_declaree_biens   VARCHAR(100) NOT NULL,
  prix_cotisation_mensuel INTEGER(4) NOT NULL,
  CONSTRAINT contrat_pk PRIMARY KEY (contrat_id)
);

ALTER TABLE contrat
ADD CONSTRAINT region_contrat_fk FOREIGN KEY (code_dep_code_commune)
REFERENCES region (code_dep_code_commune)
```

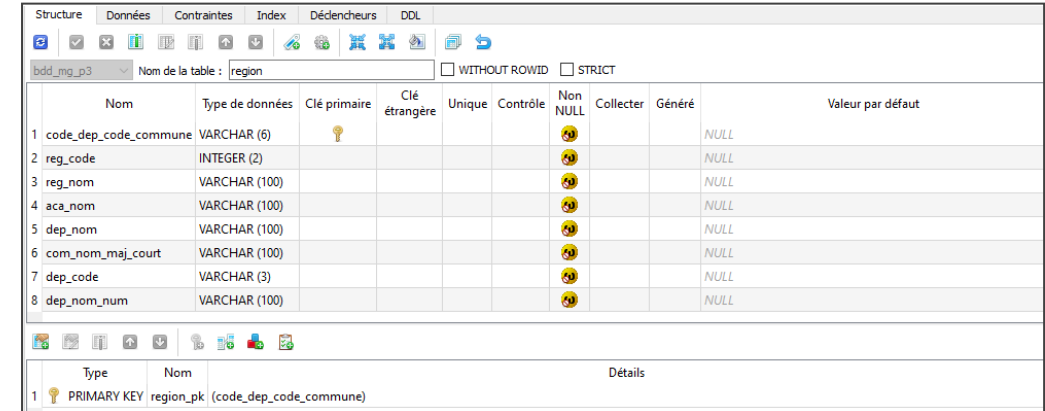
## 4. Création des tables

1. Installation sous Bash du système de gestion de bases de données relationnelles choisi **SQLite3** :  
`sudo apt-get install sqlite3`
2. Création de la base données sous Bash : `sqlite3 bdd_mg_P3.db`
3. Installation de **SQLiteStudio** pour manipuler plus facilement la base de données depuis   
<https://sqlitestudio.pl/>
4. Ouverture de la base de données « bdd\_mg\_P3.db » sous SQLiteStudio
5. Exécution du code généré par SQL Power Architect - avec un ajustement car l'instruction « ALTER TABLE [...] FOREIGN KEY [...] » ne fonctionne pas avec SQLite



```
1 CREATE TABLE region
2 (
3     code_dep_code_commune VARCHAR(6) NOT NULL,
4     reg_code INTEGER(2) NOT NULL,
5     reg_nom VARCHAR(100) NOT NULL,
6     aca_nom VARCHAR(100) NOT NULL,
7     dep_nom VARCHAR(100) NOT NULL,
8     com_nom_maj_court VARCHAR(100) NOT NULL,
9     dep_code VARCHAR(3) NOT NULL,
10    dep_nom_num VARCHAR(100) NOT NULL,
11    CONSTRAINT region_pk PRIMARY KEY (code_dep_code_commune)
12 );
13
14 CREATE TABLE contrat
15 (
16    contrat_id INTEGER(5) NOT NULL,
17    no_voie INTEGER(4),
18    b_t_q CHAR(1),
19    type_de_voie VARCHAR(4),
20    voie VARCHAR(100),
21    code_dep_code_commune VARCHAR(6) NOT NULL,
22    code_postal VARCHAR(5) NOT NULL,
23    surface INTEGER(4) NOT NULL,
24    type_local VARCHAR(100) NOT NULL,
25    occupation VARCHAR(100) NOT NULL,
26    type_contrat VARCHAR(100) NOT NULL,
27    formule VARCHAR(100) NOT NULL,
28    valeur_declaree_biens VARCHAR(100) NOT NULL,
29    prix_cotisation_mensuel INTEGER(4) NOT NULL,
30    CONSTRAINT contrat_pk PRIMARY KEY (contrat_id)
31    FOREIGN KEY (code_dep_code_commune ) REFERENCES region (code_dep_code_commune )
32 );
```

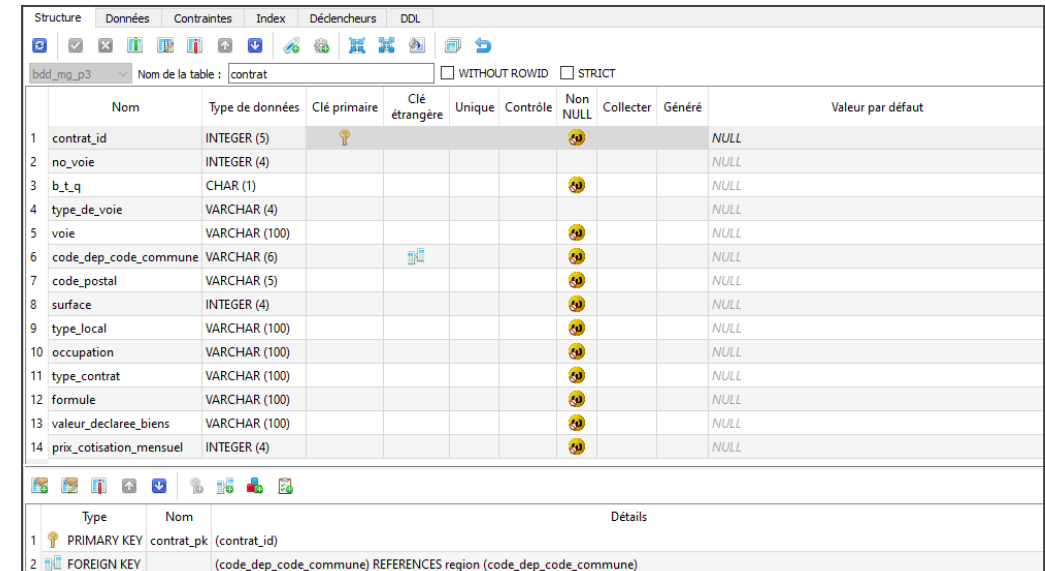
6. Les deux tables sont bien créées avec les clés voulues :



	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL	Collecter	Généré	Valeur par défaut
1	code_dep_code_commune	VARCHAR (6)	🔑				🚫			NULL
2	reg_code	INTEGER (2)					🚫			NULL
3	reg_nom	VARCHAR (100)					🚫			NULL
4	aca_nom	VARCHAR (100)					🚫			NULL
5	dep_nom	VARCHAR (100)					🚫			NULL
6	com_nom_maj_court	VARCHAR (100)					🚫			NULL
7	dep_code	VARCHAR (3)					🚫			NULL
8	dep_nom_num	VARCHAR (100)					🚫			NULL

Type Nom Détails

1 🔑 PRIMARY KEY region\_pk (code\_dep\_code\_commune)




	Nom	Type de données	Clé primaire	Clé étrangère	Unique	Contrôle	Non NULL	Collecter	Généré	Valeur par défaut
1	contrat_id	INTEGER (5)	🔑				🚫			NULL
2	no_voie	INTEGER (4)					🚫			NULL
3	b_t_q	CHAR (1)					🚫			NULL
4	type_de_voie	VARCHAR (4)					🚫			NULL
5	voie	VARCHAR (100)					🚫			NULL
6	code_dep_code_commune	VARCHAR (6)		🔗			🚫			NULL
7	code_postal	VARCHAR (5)					🚫			NULL
8	surface	INTEGER (4)					🚫			NULL
9	type_local	VARCHAR (100)					🚫			NULL
10	occupation	VARCHAR (100)					🚫			NULL
11	type_contrat	VARCHAR (100)					🚫			NULL
12	formule	VARCHAR (100)					🚫			NULL
13	valeur_declaree_biens	VARCHAR (100)					🚫			NULL
14	prix_cotisation_mensuel	INTEGER (4)					🚫			NULL

Type Nom Détails

1 🔑 PRIMARY KEY contrat\_pk (contrat\_id)

2 🔗 FOREIGN KEY (code\_dep\_code\_commune) REFERENCES region (code\_dep\_code\_commune)

## 5. Chargement des données 1/3

1. On charge en premier lieu les données de la table « régions » car elle contient la donnée « code\_dep\_code\_commune », qui est une clé secondaire de la table « contrats »  
On sélectionne l'outil « importer »  qui mène vers les fenêtres de dialogue suivantes :


**Table à importer vers**

Base de données :  Tableau :

**Source de données à importer de**

Type de données source :

Options

Fichier :  

Texte codé :

☐ Ignorer les erreurs

Options de source de données

☒ La première ligne représente les noms de colonnes CSV

Séparateur de colonnes :

☒ Valeurs NULL :


☐ Interpréter " comme un guillemet

2. En chargeant la table « contrats », on constate que la clé secondaire n'est pas connue pour certains contrats :

 [16:31:22] Erreur lors de l'import des données : FOREIGN KEY constraint failed

En effectuant une recherche sous Excel (par exemple en ajoutant dans la table « contrat » une colonne qui vérifie si la valeur « code\_dep\_code\_commune » figure dans la table « région », on repère les lignes dont le **code commune** a mal été renseigné (ici, le code postal a été renseigné au lieu du code commune sur des contrats situés à la Réunion), puis on enregistre le fichier csv avec les valeurs rectifiées grâce à une recherche sur internet.

On parvient ensuite à charger la table « contrats » rectifiée :

 [16:31:43] Données importées dans le tableau 'contrat' avec succès. Nombre de lignes importées : 30335



## 5. Chargement des données 2/3

Table « région » chargée

Base de données

Filtre par nom

bdd\_mg\_P3 (SQLite 3)

Tableaux (2)

contrat

region

Vues

Structure

Données

Contraintes

Index

Déclencheurs

DDL

Table

Formulaire

## 5. Chargement des données 3/3

Table « contrat » chargée

Base de données														
Structure														
Données														
Contraintes														
Index														
Déclencheurs														
DDL														
Table														
Formulaire														
Filtre de données														
Nombre de lignes chargées : 30335														
	contrat_id	no_voie	b_t_q	type_de_voie	voie	code_dep_code_commune	code_postal	surface	type_local	occupation	type_contrat	formule	valeur_declare	prix_cotisation_mensuel
1	100601	190	A	RUE	CENTRALE	1350	1370	50	Appartement	Locataire	Residence principale	Integral	0-25000	25
2	100602	347	NULL	RUE	DU CHATEAU	1103	1170	48	Appartement	Locataire	Residence principale	Classique	0-25000	30
3	100603	58	NULL	AV	DU MONT BLANC	1143	1220	131	Appartement	Proprietaire	Residence principale	Integral	25000-50000	57
4	100604	140	NULL	RUE	DE L'ABBE JOLIVET	1288	1630	109	Maison	Locataire	Residence principale	Integral	25000-50000	43
5	100605	39	NULL	RUE	BUFFON	1033	1200	109	Appartement	Locataire	Residence principale	Classique	0-25000	33
6	100606	8	NULL	RUE	DE GENEVE	1354	1630	53	Appartement	Proprietaire	Residence principale	Classique	0-25000	19
7	100607	2	NULL	RUE	DU RECULET	1354	1630	59	Appartement	Proprietaire	Residence principale	Integral	0-25000	15
8	100608	1403	NULL	RUE	JEAN DE GINGINS	1143	1220	93	Maison	Proprietaire	Mise en location	Integral	25000-50000	34
9	100609	226	NULL	ALL	DES CAPUCINES	1354	1630	117	Maison	Proprietaire	Residence principale	Classique	25000-50000	32
10	100610	276	NULL	RTE	DE POUIGNY	1288	1630	36	Appartement	Proprietaire	Residence principale	Integral	25000-50000	22
11	100611	79	NULL	CRS	DE VERDUN	1283	1100	138	Appartement	Proprietaire	Residence secondaire	Classique	0-25000	11
12	100612	240	NULL	RUE	DE PRE BAILLY	1173	1170	45	Appartement	Locataire	Residence principale	Classique	0-25000	16
13	100613	3	NULL	RUE	TURENNE	1033	1200	83	Appartement	Locataire	Residence principale	Classique	0-25000	14
14	100614	44	NULL	ALL	DU SQUARE DE LAUSANNE	1143	1220	88	Appartement	Locataire	Residence principale	Integral	25000-50000	34
15	100615	59	NULL	RUE	ALEXANDRE BERARD	1004	1500	165	Appartement	Locataire	Residence principale	Classique	25000-50000	24
16	100616	282	NULL	CHE	DES LONGES RAYES	1071	1170	42	Appartement	Proprietaire	Residence principale	Classique	0-25000	17
17	100617	54	NULL	GR	GRANDE RUE	1396	1150	68	Appartement	Proprietaire	Residence secondaire	Classique	0-25000	10
18	100618	1	NULL	RUE	DE GEX	1160	1210	83	Appartement	Proprietaire	Residence principale	Classique	0-25000	20
19	100619	17	NULL	LOT	LES JARDINS DE CHEVRY	1103	1170	30	Appartement	Proprietaire	Residence secondaire	Classique	0-25000	12
20	100620	329	NULL	RUE	DES CARPIERES	1399	1170	25	Appartement	Locataire	Residence principale	Integral	0-25000	11
21	100621	369	NULL	AV	CHARLES DE GAULLE	1202	1150	53	Appartement	Locataire	Residence principale	Integral	0-25000	12
22	100622	6	NULL	RUE	DES USINIERS	1173	1170	46	Appartement	Proprietaire	Residence secondaire	Classique	0-25000	12
23	100623	2	NULL	RUE	DU JURA	1354	1630	84	Appartement	Proprietaire	Residence principale	Classique	0-25000	20
24	100624	54	NULL	RUE	MARCEL DEMIA	1004	1500	68	Appartement	Locataire	Residence principale	Integral	0-25000	13
25	100625	269	NULL	CHE	DE CHARIGNIN	1034	1300	62	Appartement	Proprietaire	Residence principale	Classique	0-25000	11