

Document technique

Exploration des données et création de la
base de données



Le dictionnaire de données

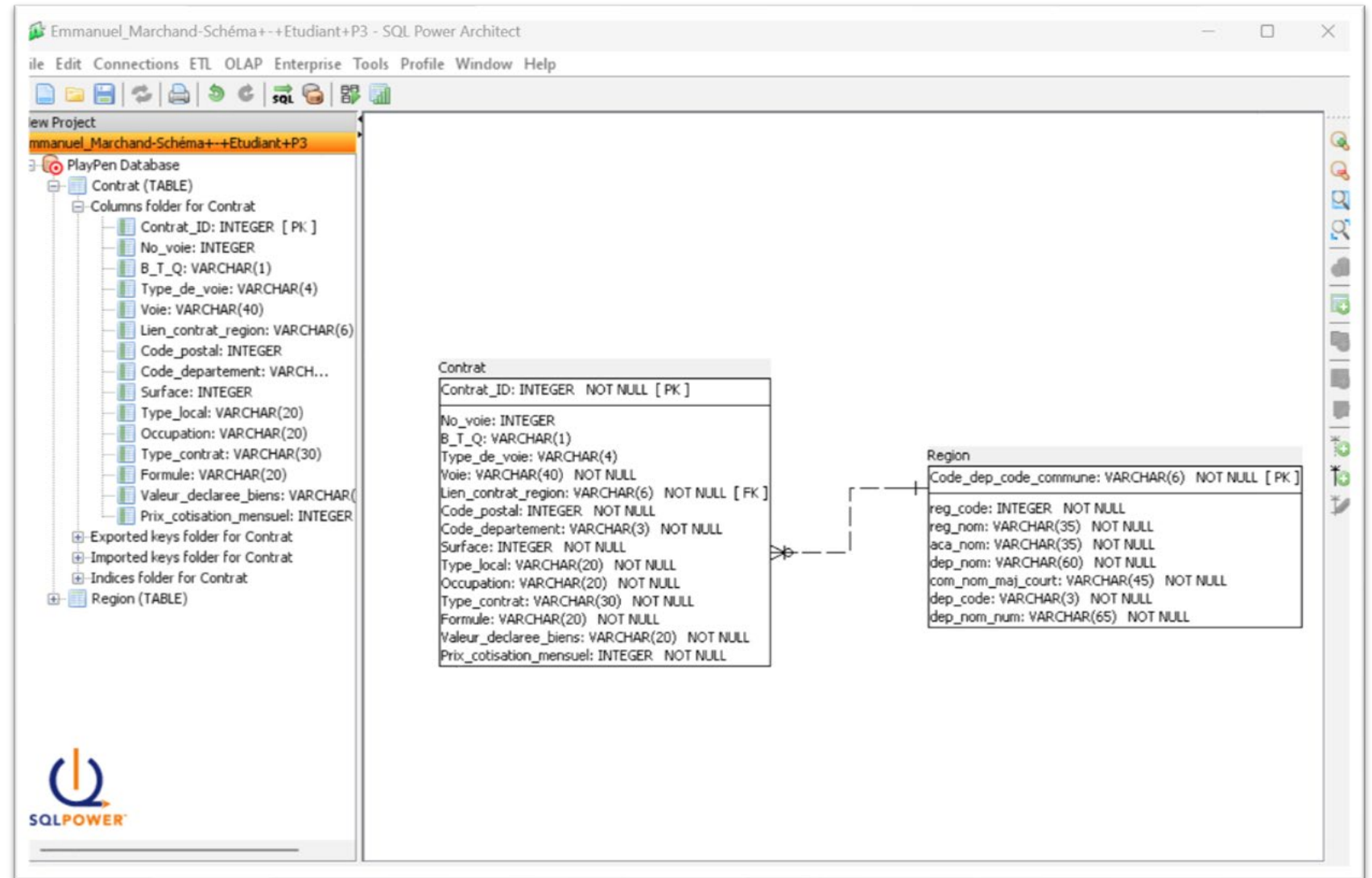
L'observation du contenu des fichiers CSV et de la documentation sur data.gouv.fr, m'a permis de constituer le dictionnaire de données suivant, en appliquant la méthodologie détaillée dans le document 'Marchand_Emanuel_3_methodologie_022024' :

	Nom des colonnes	Type de donnée:	Taille	Clé	Contrainte	Description	longueur Max
CONTRAT.CSV	Contrat_ID	INT		Clé primaire	NOT NULL	Id unique pour les contrats	
	No_voie	INT				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	1
	Type_de_voie	VARCHAR	4			Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ...	4
	Voie	VARCHAR	40		NOT NULL	Libellé de la voie pour l'adresse du logement assuré	26
	Lien_contrat_region	VARCHAR	6	Clé secondaire	NOT NULL	Lien pour la relation PLUSIEURS entre CONTRAT et REGION	6
	Code_postal	INT			NOT NULL	Code postal pour l'adresse du logement assuré	
	Commune	VARCHAR	45		NOT NULL	Libellé de la commune de l'adresse du logement	32
	Code_departement	INT			NOT NULL	Code département pour l'adresse du logement assuré	
	Surface	INT			NOT NULL	Surface du logement assuré	
	Type_local	VARCHAR	20		NOT NULL	Type de logement : Appartement ou Maison	11
	Occupation	VARCHAR	20		NOT NULL	Le statut de l'occupant : Propriétaire ou Locataire	12
	Type_contrat	VARCHAR	30		NOT NULL	Le type de contrat d'assurance : Mise en location, Résidence principale, résidence secondaire	20
	Formule	VARCHAR	20		NOT NULL	La formule souscrite par le client : Classique, Intégral.	9
	Valeur_declaree_biens	VARCHAR	20		NOT NULL	Une fourchette de la valeur des biens déclarés dans le logement.	12
REGION.CSV	Prix_cotisation_mensuel	INT			NOT NULL	Le prix mensuel de la cotisation client.	
	Code_dep_code_commune	VARCHAR	6	Clé primaire	NOT NULL	Concaténation du code département et code commune pour avoir une clé unique	6
	reg_code	INT			NOT NULL	Code région	
	reg_nom	VARCHAR	35		NOT NULL	Libellé région	26
	aca_nom	VARCHAR	35		NOT NULL	Libellé de l'académie	24
	dep_nom	VARCHAR	60		NOT NULL	Libellé du département	43
	com_nom_maj_court	VARCHAR	45		NOT NULL	Libellé court de la commune en majuscule	32
	dep_code	VARCHAR	3		NOT NULL	Code département	3
	dep_nom_num	VARCHAR	65		NOT NULL	Code et libellé du département	49

Le schéma relationnel 1/2

Le respect de la normalisation 3NF implique :

- Que chaque table soit constituée d'une clé primaire
- La suppression des redondances de données. Pour cela j'ai supprimé l'attribut 'Commune' de la table **Contrat**, car cette information se trouve déjà dans la table de référence **Région**, dans l'attribut 'com_nom_maj_court'



Le schéma relationnel 2/2



ATTENTION

J'ai conservé les colonnes 'Code département' et 'dep_code' respectivement dans les tables **Contrat** et **Région** alors qu'elles étaient redondantes. Pourquoi ? : il y a des incohérences de données dans ces tables et le nettoyage des données ne fait pas l'objet de cet exercice. En effet dans la table **Contrat**, il y a des attributs 'Code_département' qui sont erronés, par exemple la ville de CHATENAY MALABRY est notée dans le 75 en lui attachant le Code_dep_code_commune ' égal à 75019 et alors qu'elle devrait être dans le 92 avec le code_dep_code_commune égal à 92019. Cela ne permet donc pas de faire des INNER JOIN fiables entre les deux tables.

1233																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

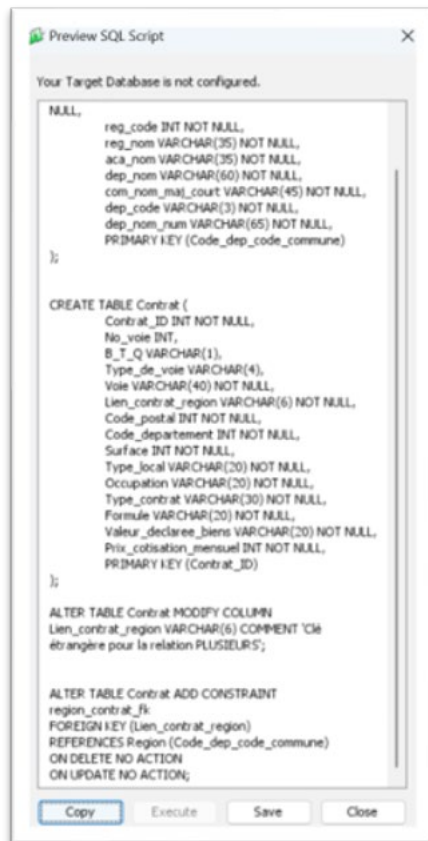
Il faut également noter que dans la table 'Region' on retrouve toujours cette même ville présente dans deux départements différents.



Code_dep_code_commune	reg_co	reg_nom	aca_no	dep_nom	com_nom_maj_court	dep_code	dep_nom_num
5019	11	Ile-de-France	Paris	Paris	CHATENAY MALABRY	75	Paris (75)
2019	11	Ile-de-France	Versaille	Hauts-de-Sein	CHATENAY MALABRY	92	Hauts-de-Seine (92)

La création des tables dans la base de données

Le code SQL de génération de la base obtenu grâce à **SQL Power Architecte**. La génération du DDL a été faite en choisissant le type de base **MySQL** :



```
CREATE TABLE Region (
    Code_dep_code_commune VARCHAR(6) NOT NULL,
    reg_code INT NOT NULL,
    reg_nom VARCHAR(35) NOT NULL,
    aca_nom VARCHAR(35) NOT NULL,
    dep_nom VARCHAR(60) NOT NULL,
    com_nom_maj_court VARCHAR(45) NOT NULL,
    dep_code VARCHAR(3) NOT NULL,
    dep_nom_num VARCHAR(65) NOT NULL,
    PRIMARY KEY (Code_dep_code_commune)
);
```

```
CREATE TABLE Contrat (
    Contrat_ID INT NOT NULL,
    No_voie INT,
    B_T_Q VARCHAR(1),
    Type_de_voie VARCHAR(4),
    Voie VARCHAR(40) NOT NULL,
    Lien_contrat_region VARCHAR(6) NOT NULL,
    Code_postal INT NOT NULL,
    Code_departement VARCHAR(3) NOT NULL,
    Surface INT NOT NULL,
    Type_local VARCHAR(20) NOT NULL,
    Occupation VARCHAR(20) NOT NULL,
    Type_contrat VARCHAR(30) NOT NULL,
    Formule VARCHAR(20) NOT NULL,
    Valeur_declaree_biens VARCHAR(20) NOT NULL,
    Prix_cotisation_mensuel INT NOT NULL,
    PRIMARY KEY (Contrat_ID)
);
```

```
ALTER TABLE Contrat MODIFY COLUMN Lien_contrat_region VARCHAR(6)
COMMENT 'Clé étrangère pour la relation PLUSIEURS';
```

```
ALTER TABLE Contrat ADD CONSTRAINT region_contrat_fk
FOREIGN KEY (Lien_contrat_region)
REFERENCES Region (Code_dep_code_commune)
ON DELETE NO ACTION
ON UPDATE NO ACTION;
```


La capture d'écran de la table contrat chargée (30 335 lignes)

DBEaver 23.3.5 - contrat

Echiers Edition Navigation Rechercher Editeur SQL Base de données Fenêtres Aide

SQL Appliquer (commit) Retour arrière (rollback) Auto localhost mysql

Navigateur de bases de données Projets

Enter a part of object name here

- AZURE INDUS - pgsindusdatagarden.postgres.database.azure.com:5432
- AZURE PROD - pgsproddatagarden.postgres.database.azure.com:5432
- INNOV INDUS - platform-garden-nonprod-db-cluster-ro-do-user-653277-0.b.d
- INNOV PROD - platform-garden-ro-do-user-653277-0.b.db.ondigitalocean.com
- localhost - localhost:3306
 - Bases de données
 - oclassroom
 - Tables
 - contrat
 - Colonnes
 - Contrat_ID (int)
 - No_voie (int)
 - B_T_Q (varchar(1))
 - Type_de_voie (varchar(4))
 - Voie (varchar(40))
 - Lien_contrat_region (varchar(6))
 - Code_postal (int)
 - Code_departement (int)
 - Surface (int)
 - Type_local (varchar(20))
 - Occupation (varchar(20))
 - Type_contrat (varchar(30))
 - Formule (varchar(20))
 - Valeur_declaree_biens (varchar(20))
 - Prix_cotisation_mensuel (int)
 - Contraintes
 - Clefs étrangères
 - Références
 - Triggers
 - Indexes

contrat Entrez une expression SQL pour filtrer les résultats (utilisez Ctrl+Espace)

	Contrat_ID	No_voie	B_T_Q	Type_de_voie	Voie	Lien_contrat_region	Cc
1	100601	190	A	RUE	CENTRALE	01350	
2	100602	347		RUE	DU CHATEAU	01103	
3	100603	58		AV	DU MONT BLANC	01143	
4	100604	140		RUE	DE L'ABBE JOLIVET	01288	
5	100605	39		RUE	BUFFON	01033	
6	100606	8		RUE	DE GENEVE	01354	
7	100607	2		RUE	DU RECULET	01354	
8	100608	1403		RUE	JEAN DE GINGINS	01143	
9	100609	226		ALL	DES CAPUCINES	01354	
10	100610	276		RTE	DE POUIGNY	01288	
11	100611	79		CRS	DE VERDUN	01283	
12	100612	240		RUE	DE PRE BAILLY	01173	
13	100613	3		RUE	TURENNE	01033	
14	100614	44		ALL	DU SQUARE DE LAUSANNE	01143	
15	100615	59		RUE	ALEXANDRE BERARD	01004	
16	100616	282		CHE	DES LONGES RAYES	01071	
17	100617	54		GR	GRANDE RUE	01396	
18	100618	1		RUE	DE GEX	01160	
19	100619	17		LOT	LES JARDINS DE CHEVRY	01103	
20	100620	329		RUE	DES CARPIERES	01399	
21	100621	369		AV	CHARLES DE GAULLE	01202	
22	100622	6		RUE	DES USINIERS	01173	
23	100623	2		RUE	DU JURA	01354	
24	100624	54		RUE	MARCEL DEMIA	01004	
25	100625	269		CHE	DE CHARIGNIN	01034	
26	100626	322		RUE	DE BREIGNY	01281	

Project - General x

Name DataSource

Régénérer Save Cancel Exporter les résultats ... 200 30335

200 row(s) fetched - 0,003s (0,001s fetch), on 2024-03-01 at 18:41:46

La capture d'écran de la table region chargée (38 916 lignes)

DBBeaver 23.3.5 - region

Echiers Edition Navigation Rechercher Editeur SQL Base de données Fenêtres Aide

SQL Appliquer (commit) Retour arrière (rollback) Auto localhost mysql

Navigateur de bases de données Projets

Enter a part of object name here

- AZURE INDUS - pgsindusdatagarden.postgres.database.azure.com:5432
- AZURE PROD - pgsproddatagarden.postgres.database.azure.com:5432
- INNOV INDUS - platform-garden-nonprod-db-cluster-ro-do-user-653277-0.b.d
- INNOV PROD - platform-garden-ro-do-user-653277-0.b.db.ondigitalocean.com
- localhost - localhost:3306
 - Bases de données
 - oclassroom
 - Tables
 - contrat
 - region (4,5M)
 - Colonnes
 - Code_dep_code_commune (varchar(6))
 - reg_code (int)
 - reg_nom (varchar(35))
 - aca_nom (varchar(35))
 - dep_nom (varchar(60))
 - com_nom_maj_court (varchar(45))
 - dep_code (varchar(3))
 - dep_nom_num (varchar(65))
 - Contraintes
 - Clefs étrangères
 - Références
 - Triggers
 - Indexes
 - Partitions
 - Vues
 - Indexes
 - Procédures
 - Triggers
 - Evènements

Script-1 region contrat oclassroom

Propriétés Données ER Diagram

localhost Bases de données oclassroom Tables region

Entrez une expression SQL pour filtrer les résultats (utilisez Ctrl+Espace)

	Code_dep_code_commune	reg_code	reg_nom	aca_nom	dep_nom	com_nom_maj_cc
1	01001	84	Auvergne-Rhône-Alpes	Lyon	Ain	L ABERGEMENT CLE
2	01002	84	Auvergne-Rhône-Alpes	Lyon	Ain	L ABERGEMENT DE
3	01003	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMAREINS
4	01004	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBERIEU EN BUGE
5	01005	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBERIEUX EN DOI
6	01006	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBLEON
7	01007	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBRONAY
8	01008	84	Auvergne-Rhône-Alpes	Lyon	Ain	AMBUTRIX
9	01009	84	Auvergne-Rhône-Alpes	Lyon	Ain	ANDERT ET CONDO
10	01010	84	Auvergne-Rhône-Alpes	Lyon	Ain	ANGLEFORT
11	01011	84	Auvergne-Rhône-Alpes	Lyon	Ain	APREMONT
12	01012	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARANC
13	01013	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARANDAS
14	01014	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARBENT
15	01015	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARBIGNIEU
16	01016	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARBIGNY
17	01017	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARGIS
18	01018	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARLOD
19	01019	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARMIX
20	01020	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARNANS
21	01021	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARS SUR FORMANS
22	01022	84	Auvergne-Rhône-Alpes	Lyon	Ain	ARTEMARE
23	01023	84	Auvergne-Rhône-Alpes	Lyon	Ain	ASNIERES SUR SAO
24	01024	84	Auvergne-Rhône-Alpes	Lyon	Ain	ATTIGNAT
25	01025	84	Auvergne-Rhône-Alpes	Lyon	Ain	BAGE LA VILLE
26	01026	84	Auvergne-Rhône-Alpes	Lyon	Ain	BAGE LE CHATEL

Record

Régénérer Save Cancel Exporter les résultats ... 200 38916

200 row(s) fetched - 0,001s, on 2024-03-01 at 18:37:34



MERCI