

MÉTHODOLOGIES DE
CRÉATION ET D'UTILISATION D'UNE
BASE DE DONNÉES

SOMMAIRE

S2 : Sommaire

S3 : Rédiger un dictionnaire de données

S4 : Concevoir un schéma relationnel normalisé

S5 : Créer des tables dans la base de données

S6 : Importer et vérifier des données

S7 à II : Utiliser la BDD

I) REDIGER UN DICTIONNAIRE DE DONNEES

Le dictionnaire de données permet de détailler les fichiers CSV servant de base à la future BDD.
C'est un document de référence dans lequel on trouve les informations suivantes:

Nom de la table

Nom de colonnes

Type de champs

Contrainte

Présence et type de Clé

Description du contenu des colonnes

	Nom des colonnes	Type de données	Taille	Clé	Description
CONTRAT.CSV	contrat_ID	INT		Clé primaire	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
	type_de_voie	VARCHAR	4		Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ...
	voie	VARCHAR	40		Libellé de la voie pour l'adresse du logement assuré
	code_dep_code_commune	VARCHAR	5	Clé étrangère	Concaténation du code département et code commune pour avoir une clé unique
	code_postal	INT			Code postal pour l'adresse du logement assuré
	surface	INT			Surface en M2 du logement assuré
	type_local	VARCHAR	11		Type de logement assuré
	occupation	VARCHAR	12		Qualité des occupants du logement assuré
	type_contrat	VARCHAR	20		Type de contrat en fonction de l'utilisation du logement assuré
	formule	VARCHAR	9		Type de formule du contrat du logement assuré
	valeur_declaree_bien	VARCHAR	15		Fourchette dans laquelle se situe la valeur du logement assuré
REGION.CSV	prix_cotisation_mensuelle	INT			Valeur mensuelle de la cotisation du contrat client pour le logement assuré
	code_dep_code_commune	VARCHAR	5	Clé primaire	Concaténation du code département et code commune pour avoir une clé unique
	reg_code	INT			Numéro de code de la région
	reg_nom	VARCHAR	26		Nom de la région
	aca_nom	VARCHAR	24		Nom de l'académie
	dep_nom	VARCHAR	43		Nom du département
	com_nom_maj_court	VARCHAR	40		Nom de la commune
	dep_code	VARCHAR	3		Numéro du département
	dep_nom_num	VARCHAR	49		Nom et numéro du département

II) CONCEVOIR UN SCHEMA RELATIONNEL NORMALISE

Le schéma relationnel représente les données issues du dictionnaire de données.

On y retrouve les différentes caractéristiques des tables:

Nom de la table

Nom de colonnes

Type de champs

Contrainte

Contrat

contrat_id: INTEGER NOT NULL [PK]

no_voie: INTEGER NOT NULL

b_t_q: VARCHAR(1) NOT NULL

type_de_voie: VARCHAR(4) NOT NULL

voie: VARCHAR(40) NOT NULL

code_dep_code_commune: VARCHAR(5) NOT NULL [FK]

code_postal: INTEGER NOT NULL

surface: INTEGER NOT NULL

type_local: VARCHAR(11) NOT NULL

occupation: VARCHAR(12) NOT NULL

type_contrat: VARCHAR(20) NOT NULL

formule: VARCHAR(9) NOT NULL

valeur_declaree_bien: VARCHAR(12) NOT NULL

prix_cotisation_mensuelle: INTEGER NOT NULL

Region

code_dep_code_commune: VARCHAR(5) NOT NULL [PK]

reg_code: INTEGER NOT NULL

reg_nom: VARCHAR(26) NOT NULL

aca_nom: VARCHAR(24) NOT NULL

dep_nom: VARCHAR(43) NOT NULL

com_nom_maj_court: VARCHAR(40) NOT NULL

dep_code: VARCHAR(3) NOT NULL

dep_nom_num: VARCHAR(49) NOT NULL

Clé primaire [PK]

La relation entre tables

Clé étrangère [FK]

Le code SQL issue du schéma permet la création de la structure de la BDD .

III) CRÉER DES TABLES DANS LA BASE DE DONNEES

Le code SQL suivant permet la création des tables dans la base de données.

```
CREATE TABLE Region (  
  code_dep_code_commune VARCHAR(5) NOT NULL,  
  reg_code INT NOT NULL,  
  reg_nom VARCHAR(26) NOT NULL,  
  aca_nom VARCHAR(24) NOT NULL,  
  dep_nom VARCHAR(43) NOT NULL,  
  com_nom_maj_court VARCHAR(40) NOT NULL,  
  dep_code VARCHAR(3) NOT NULL,  
  dep_nom_num VARCHAR(49) NOT NULL,  
  PRIMARY KEY (code_dep_code_commune)  
);
```

```
CREATE TABLE Contrat (  
  contrat_id INT NOT NULL,  
  no_voie INT NOT NULL,  
  b_t_q VARCHAR(1) NOT NULL,  
  type_de_voie VARCHAR(4) NOT NULL,  
  voie VARCHAR(40) NOT NULL,  
  code_dep_code_commune VARCHAR(5) NOT NULL,  
  code_postal INT NOT NULL,  
  surface INT NOT NULL,  
  type_local VARCHAR(11) NOT NULL,  
  occupation VARCHAR(12) NOT NULL,  
  type_contrat VARCHAR(20) NOT NULL,  
  formule VARCHAR(9) NOT NULL,  
  valeur_declaree_bien VARCHAR(12) NOT NULL,  
  prix_cotisation_mensuelle INT NOT NULL,  
  PRIMARY KEY (contrat_id)  
);
```

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

Nom de la BDD

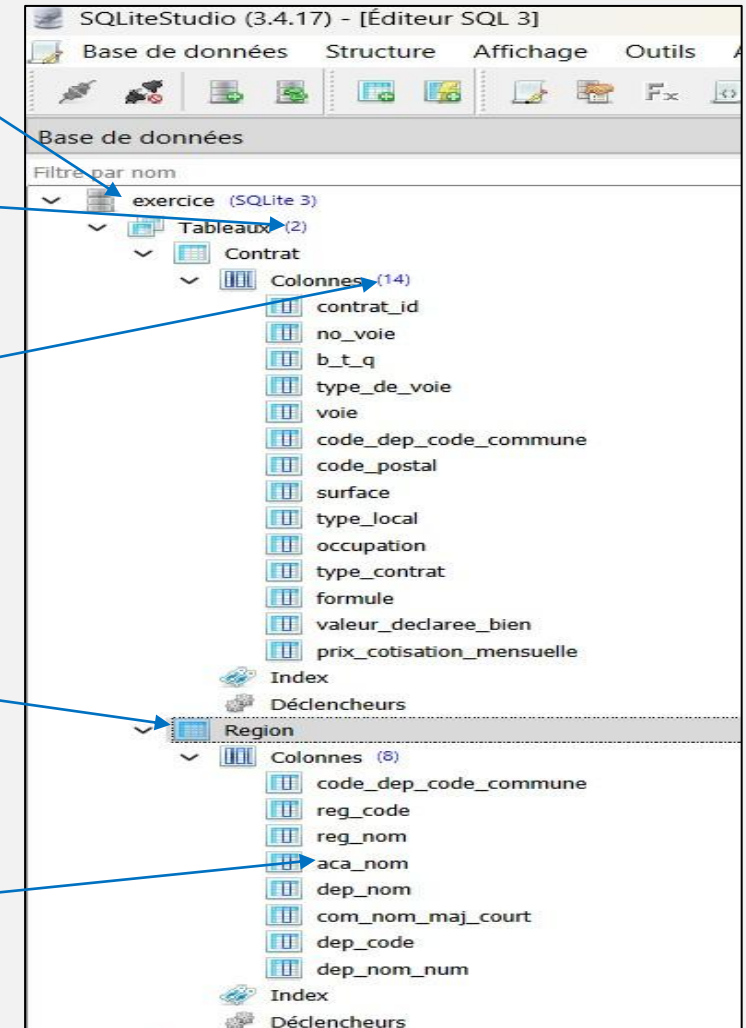
NB de tables

NB de colonnes

Nom de la table

Nom de colonnes

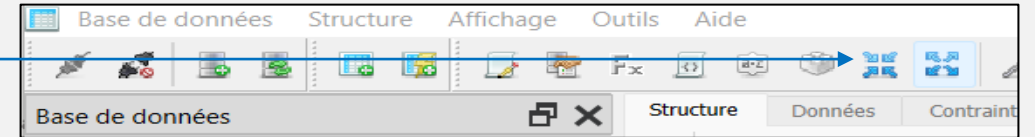
Etat de la BDD après application
du code SQL.



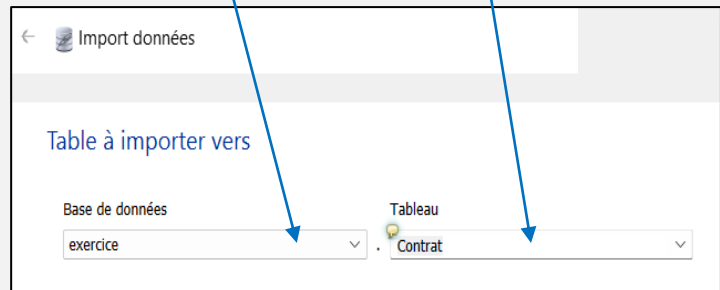
IV)IMPORTER ET VERIFIER DES DONNEES

IMPORTATIONS DES DONNEES:

-Sélection de la commande importer dans le bandeau du logiciel.



-Choix de la BDD et de la table vers laquelle on veut importer les données, puis sélection du fichier comportant les données



L'importation est achevée lorsque tous les fichiers CSV ont été importés.

VERIFICATIONS DES DONNEES:

-Contrôle de la cohérence des données importées en procédant à des vérifications entre les données de la BDD et les fichiers source.

8 -- Données dans la table Contrat
9 SELECT * FROM Contrat;

Tableau

Nombre de lignes chargées : 30335

	contrat_id	no_voie	b_t_q	type_de_voi	voie	code_dep_c	code_postal	surface	type_local	occupation	type_contr	deciaree_	prix_cotisati
1	100601	190	A	RUE	CENTRALE	1350	1370	50	Appartement	Locataire	Residen	00	25
2	100602	347		RUE	DU CHATEAU	1103	1170	48	Appartement	Locataire	Residen	00	30
3	100603	58		AV	DU MONT BLANC	1143	1220	131	Appartement	Propriétaire	Residence principale	Integral	25000-50000
4	100604	140		RUE	DE L'ABBE JOLIVET	1288	1630	109	Maison	Locataire	Residence principale	Integral	25000-50000
5	100605	39		RUE	BUFFON	1033	1200	109	Appartement	Locataire	Residence principale	Classique	0-25000
6	100606	8		RUE	DE GENEVE	1354	1630	53	Appartement	Propriétaire	Residence principale	Classique	0-25000
7	100607	2		RUE	DU RECULET	1354	1630	59	Appartement	Propriétaire	Residence principale	Integral	0-25000
8	100608	1403		RUE	JEAN DE GINGINS	1143	1220	93	Maison	Propriétaire	Mise en location	Integral	25000-50000
9	100609	226		ALL	DES CAPUCINES	1354	1630	117	Maison	Propriétaire	Residence principale	Classique	25000-50000
10	100610	276		RTE	DE POUIGNY	1288	1630	36	Appartement	Propriétaire	Residence principale	Integral	25000-50000
11	100611	79		CRS	DE VERDUN	1283	1100	138	Appartement	Propriétaire	Residence secondaire	Classique	0-25000
12	100612	240		RUE	DE PRE BAILLY	1173	1170	45	Appartement	Locataire	Residence principale	Classique	0-25000
13	100613	3		RUE	TURENNE	1033	1200	83	Appartement	Locataire	Residence principale	Classique	0-25000
14	100614	44		ALL	DU SQUARE DE LAUSANNE	1143	1220	88	Appartement	Locataire	Residence principale	Integral	25000-50000
15	100615	59		RUE	ALEXANDRE BERARD	1004	1500	165	Appartement	Locataire	Residence principale	Classique	25000-50000
16	100616	282		CHE	DES LONGES RAYES	1071	1170	42	Appartement	Propriétaire	Residence principale	Classique	0-25000
17	100617	54		GR	GRANDE RUE	1306	1150	68	Annartement	Propriétaire	Residence secondaire	Classique	0-25000

SELECT * FROM Contrat;

Nb de
lignes
Chargées:
30335

9 -- Données dans la table Region
10 SELECT * FROM Region;

Table

Nombre de lignes chargées : 38916

	code_dep_code_c	reg_code	aca_nom	dep_nom	com_nom_maj_court	dep_code	dep_nom_n
1	1001	84	pes Lyon	Ain	L ABERGEMENT CLEMENCIAT	1	Ain (01)
2	1002	84	pes Lyon	Ain	L ABERGEMENT DE VAREY	1	Ain (01)
3	1003	84	Auvergne-Rhône-Alpes	Lyon	AMAREINS	1	Ain (01)
4	1004	84	Auvergne-Rhône-Alpes	Lyon	AMBERIEU EN BUGEY	1	Ain (01)
5	1005	84	Auvergne-Rhône-Alpes	Lyon	AMBERIEUX EN DOMBES	1	Ain (01)
6	1006	84	Auvergne-Rhône-Alpes	Lyon	AMBLEON	1	Ain (01)
7	1007	84	Auvergne-Rhône-Alpes	Lyon	AMBRONAY	1	Ain (01)
8	1008	84	Auvergne-Rhône-Alpes	Lyon	AMBUTRIX	1	Ain (01)
9	1009	84	Auvergne-Rhône-Alpes	Lyon	ANDERT ET CONDON	1	Ain (01)
10	1010	84	Auvergne-Rhône-Alpes	Lyon	ANGLEFORT	1	Ain (01)
11	1011	84	Auvergne-Rhône-Alpes	Lyon	APREMONT	1	Ain (01)
12	1012	84	Auvergne-Rhône-Alpes	Lyon	ARANC	1	Ain (01)
13	1013	84	Auvergne-Rhône-Alpes	Lyon	ARANDAS	1	Ain (01)
14	1014	84	Auvergne-Rhône-Alpes	Lyon	ARBENT	1	Ain (01)
15	1015	84	Auvergne-Rhône-Alpes	Lyon	ARBIGNIEU	1	Ain (01)
16	1016	84	Auvergne-Rhône-Alpes	Lyon	ARBIGNY	1	Ain (01)
17	1017	84	Auvergne-Rhône-Alpes	Lyon	ARGIS	1	Ain (01)

Nb de
lignes
Chargées:
38916

V)UTILISER LA BBD

Pour analyser les éléments présents dans une base de données on utilise des requêtes en langage SQL.

```
1 --Requête 1: Lister Les numéros de contrats (contrat_ID) avec Leur surface pour La commune de Caen.  
2  
3 SELECT contrat_id, Contrat.code_dep_code_commune as cdcc, com_nom_maj_court as commune, surface  
4 FROM Contrat INNER JOIN Region ON Contrat.code_dep_code_commune = Region.code_dep_code_commune  
5 WHERE com_nom_maj_court = 'CAEN';  
6
```

Table	Formulaire
Nombre de lignes chargées : 4	

	contrat_id	cdcc	commune	surface
1	103791	14118	CAEN	35
2	103792	14118	CAEN	99
3	103793	14118	CAEN	40
4	103794	14118	CAEN	20

FROM
Sélection
de la Table
(ici Contrat)

SELECT Sélection
de colonne
(ici contrat_id...)

as
Permet de
renommer une
colonne
(ici com_nom_maj_court en
commune)

Il y a 4 contrats à Caen avec des surfaces allant de 20 à 99 m².

```
38 --Requêtes 2 et 8 : Lister Les numéros de contrats avec Le type de contrat et Leur formule pour Les maisons du département 71  
39  
40 SELECT contrat_id, type_contrat, formule, type_local  
41 FROM Contrat INNER JOIN Region ON Contrat.code_dep_code_commune = Region.code_dep_code_commune  
42 WHERE type_local = 'Maison' AND dep_code = '71';
```

Table	Formulaire
Nombre de lignes chargées : 4	

	contrat_id	type_contrat	formule	type_local
1	114768	Residence principale	Integral	Maison
2	114779	Residence principale	Classique	Maison
3	114782	Residence principale	Classique	Maison
4	114812	Residence principale	Integral	Maison

WHERE
Indique une ou plusieurs (**AND**)
colonnes dans lesquelles vont
s'effectuer les recherches (ici type_local
et dep_code) et la ou les conditions
recherchées (ici Maison et 71)

INNER JOIN et ON
Réalise une jointure entre les tables
(ici Contrat et Region) via la colonne
commune aux 2 tables (ici
code_dep_code_commune)

Il y a 4 contrats concernant des maisons du département 71, toutes des résidences principales. Deux contrats ont une formule Classique tandis que les deux autres sont de type Integral.

V)UTILISER LA BBD

```
32 --Lister Le nom des régions de France.  
33  
34  
35 SELECT DISTINCT reg_nom FROM Region;  
36
```

Table		Formulaire
		Nombre de lignes chargées : 19
	reg_nom	
1	Auvergne-Rhône-Alpes	
2	Hauts-de-France	
3	Provence-Alpes-Côte d'Azur	
4	Grand Est	
5	Occitanie	
6	Normandie	
7	Nouvelle-Aquitaine	
8	Centre-Val de Loire	
9	Bourgogne-Franche-Comté	
10	Bretagne	
11	Pays de la Loire	
12	Ile-de-France	
13	Guadeloupe	
14	Martinique	
15	Guyane	
16	La Réunion	
17	Collectivités d'outre-mer	
18	Mayotte	
19	Corse	

SELECT DISTINCT
Sélection de chaque donnée présente au moins une fois dans la colonne (*ici reg_nom*)

Il y a 19 régions répertoriées en France.

```
20 --Requête 4 : Quels sont les 5 contrats qui ont les surfaces les plus élevées ?  
21  
22 SELECT contrat_id, surface  
23 FROM Contrat  
24 ORDER BY surface DESC  
25 LIMIT 5;  
26  
27
```

Table		Formulaire
		Nombre de lignes chargées : 5
	contrat_id	surface
1	104211	815
2	105463	742
3	130878	595
4	100822	570
5	109872	559

ORDER BY et DESC
classent les données de la colonne sélectionnée par ordre décroissant (*ici surface*)

Les 5 contrats ayant les surfaces les plus élevées sont les contrats 104211, 105463, 130878, 100822, et 109872.

V)UTILISER LA BBD

9 --Requête 5 : Quel est Le prix moyen de La cotisation mensuelle ?

```
1 SELECT ROUND(avg(prix_cotisation_mensuelle),1) AS prix_moyen
2 FROM Contrat;
```

3

Table Formulaire

Nombre de lignes chargées : 1

	prix_moyen
1	19,3

ROUND
Permet d'arrondir un
résultat (ici au nombre
décimal souhaité)

avg()
Permet de calculer la
moyenne des données (ici des
prix_cotisation_mensuelle)

Le prix moyen de la cotisation mensuelle est de 19,3€.

16 --Requête 6 : Quel est Le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens ?

```
17
18
19 SELECT valeur_declaree_bien, COUNT(*) AS nb_contrat
20 FROM Contrat
21 GROUP BY valeur_declaree_bien
22 ORDER BY nb_contrat DESC;
23
```

Table Formulaire

Nombre de lignes chargées : 4

	valeur_declaree_bien	nb_contrat
1	0-25000	22720
2	25000-50000	6815
3	50000-100000	696
4	100000+	104

GROUP BY
Indique la colonne par
laquelle on veut
grouper les données
(ici par valeur_declaree_bien)

Il y a 22 720 contrats dont la valeur est comprise entre 0 et 25 000€, 6 815 entre 25 000 et 50 000€, 696 entre 50 000 et 100 000€ et 104 supérieurs à 100 000 €.

V)UTILISER LA BBD

```
24
25 --Requête 7 : Quel est le nombre de formules "Integral" sur la région Pays de La Loire ?
26
27 SELECT formule, COUNT(*) AS nb_formule
28 FROM Contrat INNER JOIN Region ON Contrat.code_dep_code_commune = Region.code_dep_code_commune
29 WHERE formule = 'Integral' AND reg_nom = 'Pays de La Loire';
30
31
```

Table		Formulaire
		Nombre de lignes chargées : 1
formule	nb_formule	
1 Integral	589	

COUNT(*)
Comptabilise le nombre de lignes présentes dans la colonne (ici formule)

Il y a **589** contrats avec des formules intégral dans la région Pays de la Loire.

```
53 --Requête 9 : Quelle est la surface moyenne des contrats à Paris ?
54
55 SELECT ROUND(avg(surface),1) as surface_moyenne
56 FROM Contrat INNER JOIN Region ON Contrat.code_dep_code_commune = Region.code_dep_code_commune
57 WHERE com_nom_maj_court LIKE 'Paris%';
58
59
```

Table		Formulaire
		Nombre de lignes chargées : 1
surface_moyenne		
1 51.8		

LIKE %
Indique une recherche dans les données commençant par les caractères mentionnés avant % (ici Paris)

La surface moyenne des contrats à Paris est de **51,8 m²**.

```
60 --Requête 10 : Classement des 10 départements où le prix moyen de la cotisation est le plus élevé.
61
62 SELECT dep_nom as departement, ROUND(avg(prix_cotisation_mensuelle),1) as prix_moyen_cotisation
63 FROM Contrat INNER JOIN Region ON contrat.code_dep_code_commune = region.code_dep_code_commune
64 GROUP BY departement
65 ORDER BY prix_moyen_cotisation DESC
66 LIMIT 10;
```

Table		Formulaire
		Nombre de lignes chargées : 10
departement	prix_moyen_cotisation	
1 Paris	36.4	
2 Hauts-de-Seine	26.3	
3 Val-de-Marne	19.8	
4 Yvelines	18.9	
5 Rhône	18.5	
6 Ain	18.2	
7 Alpes-Maritimes	18.1	
8 Charente-Maritime	17.3	
9 Haute-Savoie	17.1	
10 Corse-du-Sud	17.1	

LIMIT
Limite le nombre de lignes de résultats souhaités (ici 10)

Les cotisation mensuelles des départements où celles-ci sont les plus fortes varient de **17,1 à 36,4€**

V)UTILISER LA BBD

```
55 --Requête 11 : Liste des communes ayant eu au moins 150 contrats.
56
57 SELECT com_nom_maj_court as commune, count(contrat_id) as nb_contrat
58 FROM Contrat INNER JOIN Region ON region.code_dep_code_commune = contrat.code_dep_code_commune
59 GROUP BY commune
60 HAVING nb_contrat >= 150
61 ORDER BY nb_contrat DESC;
62
```

Table		Formulaire	
		Nombre de lignes chargées : 20	
	commune	nb_contrat	
1	PARIS 18	515	
2	PARIS 17	468	
3	PARIS 15	407	
4	PARIS 16	394	
5	NICE	387	
6	PARIS 11	381	
7	PARIS 20	302	
8	BORDEAUX	302	
9	NANTES	291	
10	PARIS 19	266	
11	PARIS 10	263	
12	PARIS 12	252	
13	PARIS 14	222	
14	GRENOBLE	220	
15	PARIS 9	204	
16	TOULOUSE	187	
17	TOULON	170	
18	COURBEVOIE	163	
19	LILLE	161	
20	PARIS 3	159	

HAVING
Conditionne le résultat
de la recherche en
fonction du critère
choisit (ici > ou = 150)

Il y a 20 villes ayant au moins 150 contrats.

```
77 --Requête 12 : Quel est Le nombre de contrats pour chaque région ?
78
79 SELECT reg_nom as region, COUNT(contrat_id) as nb_contrat
80 FROM Region LEFT JOIN Contrat ON contrat.code_dep_code_commune = region.code_dep_code_commune
81 GROUP BY region
82 ORDER BY nb_contrat DESC;
```

Table		Formulaire	
		Nombre de lignes chargées : 19	
	region	nb_contrat	
1	Ile-de-France	14177	
2	Provence-Alpes-Côte d'Azur	3279	
3	Auvergne-Rhône-Alpes	3042	
4	Nouvelle-Aquitaine	2038	
5	Occitanie	1609	
6	Pays de la Loire	1196	
7	Hauts-de-France	1189	
8	Bretagne	947	
9	Normandie	824	
10	Grand Est	769	
11	Centre-Val de Loire	598	
12	Bourgogne-Franche-Comté	293	
13	Corse	247	
14	Martinique	73	
15	Guyane	37	
16	La Réunion	17	
17	Mayotte	0	
18	Guadeloupe	0	
19	Collectivités d'outre-mer	0	

LEFT JOIN et ON
Réalise une jointure à gauche entre des
tables (ici Region et Contrat) via la
colonne commune aux 2 tables (ici
code_dep_code_commune) afin de
prendre en compte l'ensemble des
données de la table Region.

La liste nous indique le nombre de contrats par régions. On remarque que trois d'entre elles n'ont pas de contrat répertoriés..