

Requêtez une base de données avec SQL

Bootcamp Data Analyst – Zied Belhocine

Objectif du projet

Nous disposons de trois documents relatifs aux contrats d'assurance habitation en France :

- - Un fichier CSV contenant les données des contrats clients
- - Un fichier CSV contenant le référentiel géographique des régions françaises
- - Un dictionnaire de données recensant les champs utilisés, leur type, taille et définition
- But : intégrer ces deux CSV dans une base relationnelle en respectant les contraintes et relations, puis exécuter des requêtes SQL pour extraire des informations utiles.

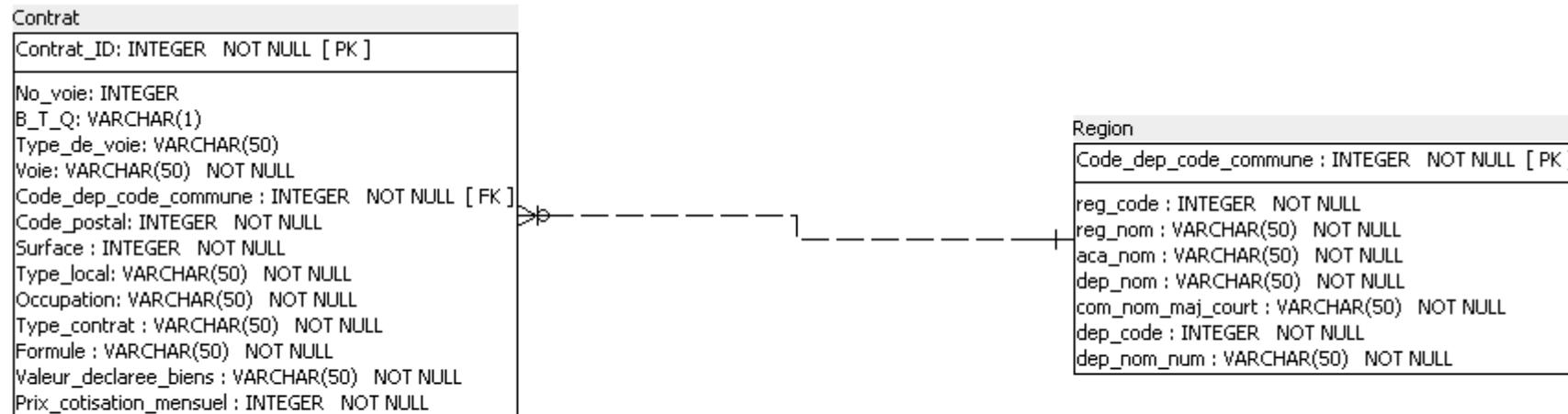
Dictionnaire de données

- À partir des fichiers CSV, on détermine pour chaque table :
- - Noms de colonnes
- - Type et taille des données
- - Clés primaires et étrangères
- - Descriptions des champs
- Tables principales : Contrat / Région

	Nom des colonnes	Type de données	Taille	Clé	Description
CONTRAT.CSV	Contrat_ID	INT	6	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	26		Libellé de la voie pour l'adresse du logement assuré
	Code_dep_code_commune	INT	6	Clé secondaire	Concaténation du code département et code commune pour avoir une clé unique
	Code_postal	INT	5		Code postal pour l'adresse du logement assuré
	Surface	INT	3		Surface du logement assuré
	Type_local	VARCHAR	11		Nature du logement assuré : Appartement, maison, ...
	Occupation	VARCHAR	12		Nature de l'occupant du logement : Locataire, Propriétaire
	Type_contrat	VARCHAR	20		Type d'occupation du bien : résidence principale, mise en location, résidence secondaire.
	Formule	VARCHAR	9		Formule du contrat affectée au logement : Integral, Classique
REGION.CSV	Valeur_declaree_biens	VARCHAR	12		Tranche estimée de la valeur des biens présents dans le logement
	Prix_cotisation_mensuel	INT	3		Prix payé chaque mois par le souscripteur
	Code_dep_code_commune	INT	6	Clé primaire	Concaténation du code département et code commune pour avoir une clé unique
	reg_code	INT	2		Code à un ou deux chiffre pour désigner la région
	reg_nom	VARCHAR	26		Nom de la région où se situe la commune
	aca_nom	VARCHAR	24		Nom de l'académie à laquelle est rattachée la commune
	dep_nom	VARCHAR	43		Nom du département où se situe la commune
	com_nom_maj_court	VARCHAR	32		Nom de la commune en majuscule avec les abréviations (ex : 'Saint' est noté 'ST')
	dep_code	INT	3		Code du département
	dep_nom_num	VARCHAR	50		Concaténation du nom de département et du code département : Nom(code)

Schéma relationnel

- Le dictionnaire de données permet de construire un diagramme UML illustrant la structure et les liens entre les tables.



Création des tables

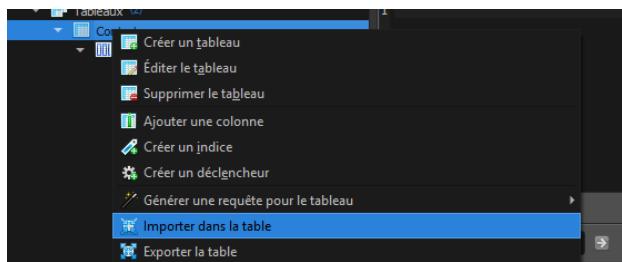
```
CREATE TABLE Region (
    Code_dep_code_commune INTEGER PRIMARY KEY,
    reg_code      INTEGER,
    reg_nom       VARCHAR(50),
    aca_nom       VARCHAR(50),
    dep_nom       VARCHAR(50),
    com_nom_maj_court  VARCHAR(50),
    dep_code      INTEGER,
    dep_nom_num   VARCHAR(50)
);
```

```
CREATE TABLE Contrat (
    Contrat_ID      INTEGER PRIMARY KEY,
    No_voie         INTEGER,
    B_T_Q           VARCHAR(50),
    Type_de_voie   VARCHAR(50),
    Voie            VARCHAR(50),
    Code_dep_code_commune  INTEGER,
    Code_postal     INTEGER,
    Surface          INTEGER,
    Type_local      VARCHAR(50),
    Occupation      VARCHAR(50),
    Type_contrat    VARCHAR(50),
    Formule          VARCHAR(50),
    Valeur_declaree_biens  VARCHAR(50),
    Prix_cotisation_mensuel  INTEGER,
    FOREIGN KEY (Code_dep_code_commune) REFERENCES
    Region(Code_dep_code_commune) ON DELETE RESTRICT
    ON UPDATE CASCADE
);
```

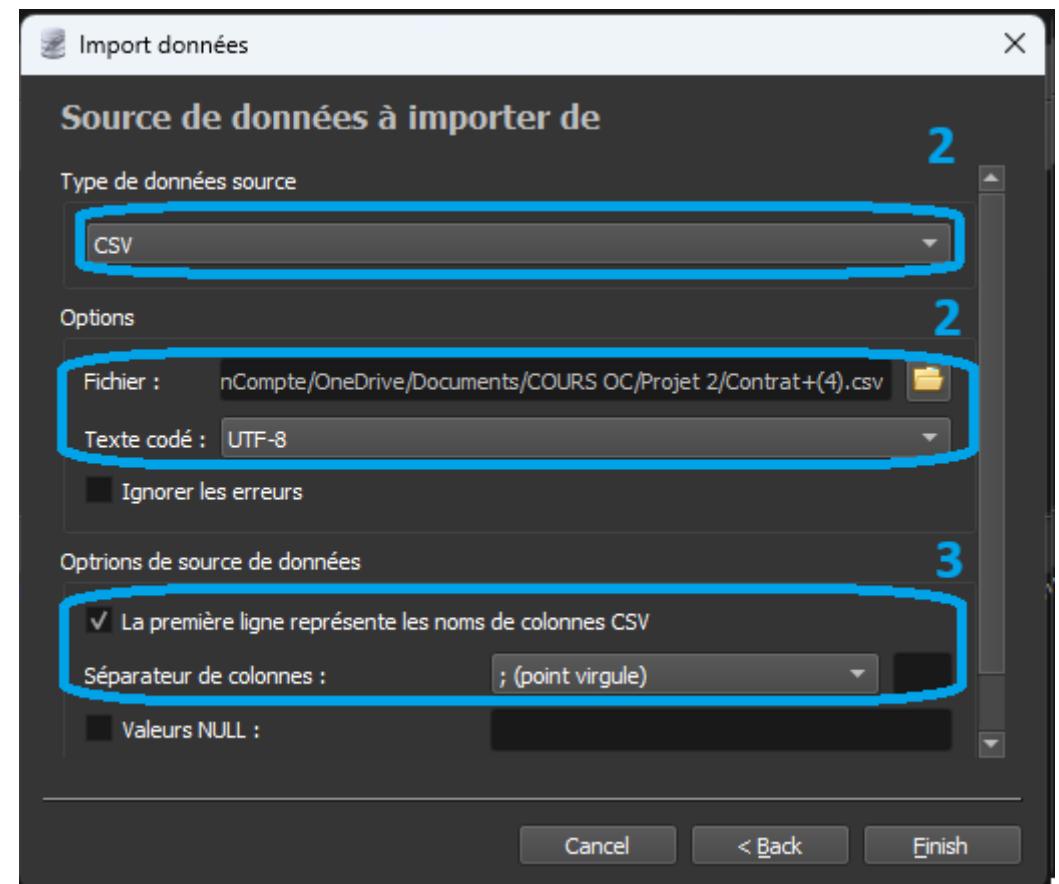
- 💡 Remarque : la table Region est créée avant Contrat car cette dernière contient une clé étrangère la référant.

Importation des données

- Les fichiers CSV peuvent être importés directement dans SQLite :
- 1. Clic droit sur la base → Importer



- 2. Sélectionner le fichier CSV
- 3. Cocher :
 - La première ligne contient les noms de colonnes
 - Séparateur : ;



Premières requêtes SQL

- I - Lister les numéros de contrats (contrat_ID) avec leur surface pour la commune de Caen.

```
1 SELECT
2     c.contrat_ID,
3     c.surface,
4     r.com_nom_maj_court
5 FROM contrat c
6 JOIN region r ON c.Code_dep_code_commune = r.Code_dep_code_commune
7 WHERE lower(r.com_nom_maj_court) LIKE 'caen' ;
```

	Contrat ID	Surface	com nom maj court
1	103791	35	CAEN
2	103792	99	CAEN
3	103793	40	CAEN
4	103794	20	CAEN

- II - Lister les numéros de contrats avec le type de contrat et leur formule pour les maisons du département 71.

```
1 SELECT
2 contrat_ID,
3 type_contrat,
4 formule,
5 type_local,
6 r.dep_code
7 FROM
8 contrat c
9 JOIN Region r ON c.Code_dep_code_commune = r.Code_dep_code_commune
10 WHERE
11 dep_code = 71 AND lower(Type_local) LIKE 'maison' ;
```

	Contrat ID	Type contrat	Formule	Type local	dep code
1	114768	Residence principale	Integral	Maison	71
2	114779	Residence principale	Classique	Maison	71
3	114782	Residence principale	Classique	Maison	71
4	114812	Residence principale	Integral	Maison	71

- III - Lister le nom des régions de France

```
1 SELECT DISTINCT reg_nom FROM region ;
```

reg_nom
1 Auvergne-Rhône-Alpes
2 Haute-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

Analyses complémentaires

- IV - Quels sont les 5 contrats qui ont les surfaces les plus élevées ?

```
1 SELECT
2     Contrat_ID,
3     surface
4 FROM
5     contrat
6 ORDER BY
7     surface DESC
8 LIMIT 5 ;
```

Contrat ID	Surface
1 104211	815
2 105463	742
3 130878	595
4 100822	570
5 109872	559

- V - Quel est le prix moyen de la cotisation mensuelle ?

```
SELECT
    ROUND(AVG(Prix_cotisation_mensuel)) AS moyenne_Prix_cotisation_mensuel
FROM
    contrat ;
```

moyenne Prix cotisation mensuel
1 19

- VI - Quel est le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens ?

```
1 SELECT
2     Valeur_declaree_biens,
3     count(Contrat_ID) AS total_contrats
4 FROM
5     contrat
6 GROUP BY
7     Valeur_declaree_biens ;
```

Valeur declaree biens	total contrats
1 0-25000	22720
2 100000+	104
3 25000-50000	6815
4 50000-100000	696

Analyses complémentaires

- VII - Quel est le nombre de formules “integral” sur la région Pays de la Loire ?

```
1 SELECT
2   r.reg_nom,
3   count(*) AS quantite_formule_integral_pays_de_la_loire
4 FROM
5   contrat c
6 JOIN region r ON c.Code_dep_code_communne = r.Code_dep_code_communne
7 WHERE
8   c.formule LIKE 'Integral'
9   AND lower(r.reg_nom) LIKE '%pays%Loire%';
```

req nom	quantite formule integral pays de la loire
Pays de la Loire	589

- VIII - Lister les numéros de contrats avec le type de contrat et leur formule pour les maisons du département 71.

```
1 SELECT
2   Contrat_ID,
3   Type_contrat,
4   formule,
5   r.dep_code
6 FROM
7   contrat c
8 JOIN Region r ON c.Code_dep_code_communne = r.Code_dep_code_communne
9 WHERE
10  dep_code = 71 AND lower(Type_local) LIKE 'maison'
```

Contrat ID	Type contrat	Formule	dep code
114768	Residence principale	Integral	71
114779	Residence principale	Classique	71
114782	Residence principale	Classique	71
114812	Residence principale	Integral	71

- IX - Quelle est la surface moyenne des contrats à Paris ?

```
SELECT
  r.com_nom_maj_court AS commune,
  ROUND(AVG(c.surface)) AS surface_moyenne
FROM
  contrat c
JOIN
  Region r ON c.Code_dep_code_communne = r.Code_dep_code_communne
WHERE
  r.dep_code = 75 ;
```

commune	surface moyenne
PARIS 3	52

Analyses complémentaires

- X - Classement des 10 départements où le prix moyen de la cotisation est le plus élevé.

```
1 SELECT
2     r.dep_nom AS departement,
3     ROUND( AVG(c.Prix_cotisation_mensuel)) AS Prix_cotisation_mensuel_moyenne
4 FROM
5     contrat c
6 JOIN
7     Region r ON c.Code_dep_code_commune = r.Code_dep_code_commune
8 GROUP BY
9     r.dep_nom
10 ORDER BY
11     Prix_cotisation_mensuel_moyenne DESC
12 LIMIT 10 ;
```

	departement	Prix cotisation mensuel moyenne
1	Paris	36
2	Hauts-de-Seine	26
3	Val-de-Marne	20
4	Yvelines	19
5	Rhône	18
6	Alpes-Maritimes	18
7	Ain	18
8	Seine-Saint-Denis	17
9	Haute-Savoie	17
10	Corse-du-Sud	17

- XI - Liste des communes ayant eu au moins 150 contrats.

```
1 SELECT
2     r.com_nom_maj_court AS commune,
3     count(contrat_id) AS nombre_contrats
4 FROM
5     Contrat c
6 JOIN
7     Region r ON c.Code_dep_code_commune = r.Code_dep_code_commune
8 GROUP BY
9     r.com_nom_maj_court
10 HAVING
11     nombre_contrats > 150
12 ORDER BY
13     nombre_contrats DESC ;
```

	commune	nombre contrats
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

- XII - Quel est le nombre de contrats pour chaque région ?

```
1 SELECT
2     r.reg_nom AS region,
3     count(contrat_id) AS nombre_contrats
4 FROM
5     Contrat c
6 JOIN
7     Region r ON c.Code_dep_code_commune = r.Code_dep_code_commune
8 GROUP BY
9     r.reg_nom
10 ORDER BY
11     nombre_contrats DESC ;
```

	region	nombre contrats
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	8

Conclusion

- Ce projet illustre la chaîne complète d'un mini-projet SQL :
- - Création du modèle de données
- - Intégration de fichiers bruts
- - Analyse et extraction d'informations via des requêtes structurées
- Merci pour votre attention.