

Projet 3 - Requêtez une base de données avec SQL méthodologie et résultats

Candice Plainfossé
Parcours de formation Data analyst

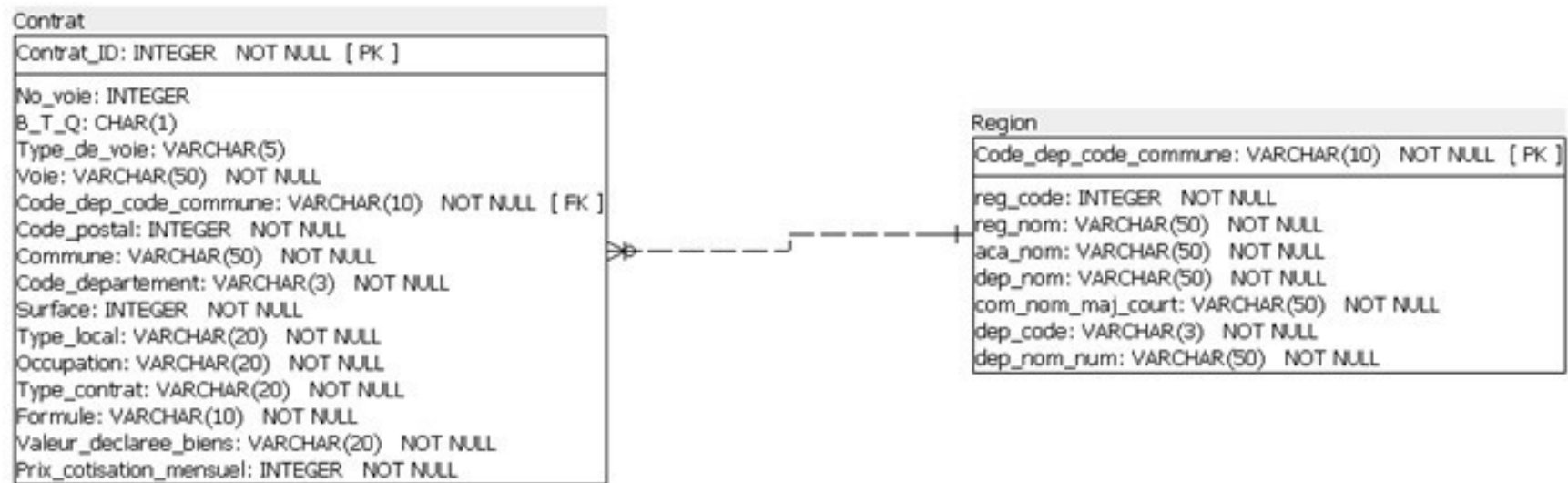
Sommaire

1. Dictionnaire des données explorées
2. Schéma relationnel normalisé
3. Code SQL de création des tables
4. Analyses réalisées avec requêtes SQL associées et résultats

1. Dictionnaire des données explorées

	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	5		Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ...
	Voie	VARCHAR	50		Libellé de la voie pour l'adresse du logement assuré - NOT NULL
	Code_dep_code_commune	VARCHAR	10	Clé secondaire	Concaténation du code département et code commune pour avoir une clé unique
	Code_postal	INT			Code postal de la commune du logement assuré - NOT NULL
	Commune	VARCHAR	50		Nom de la commune de l'adresse du logement assuré - NOT NULL
	Code_departement	VARCHAR	3		Code du département où se situe le logement assuré - NOT NULL
	Surface	INT			Superficie du local assuré en m² - NOT NULL
	Type_local	VARCHAR	20		Type du logement assuré (maison ou appartement) - NOT NULL
	Occupation	VARCHAR	20		Statut de la personne qui souscrit à l'assurance (propriétaire ou locataire) - NOT NULL
	Type_contrat	VARCHAR	20		Statut du bien assuré (résidence principale, secondaire, mise en location) - NOT NULL
	Formule	VARCHAR	10		Type de contrat d'assurance (classique ou intégral) - NOT NULL
	Valeur_declaree_biens	VARCHAR	20		Tranche de valeur déclarée des biens mobiliers à assurer en € - NOT NULL
REGION.CSV 1	Prix_cotisation_mensuel	INT			Prix de la cotisation mensuelle du contrat d'assurance en € - NOT NULL
	Code_dep_code_commune	VARCHAR	10	Clé primaire	Concaténation du code département et code commune pour avoir une clé unique
	reg_code	INT			Code de la région - NOT NULL
	reg_nom	VARCHAR	50		Nom de la région - NOT NULL
	aca_nom	VARCHAR	50		Nom de l'académie (circonscription de l'Education Nationale) - NOT NULL
	dep_nom	VARCHAR	50		Nom du département - NOT NULL
	com_nom_maj_court	VARCHAR	50		Nom de la commune en majuscule dans un format raccourci - NOT NULL
	dep_code	VARCHAR	3		Code du département - NOT NULL
	dep_nom_num	VARCHAR	50		Concaténation du nom du département associé au numéro de département correspondant entre parenthèses - NOT NULL

2. Schéma relationnel normalisé



3. Création des tables dans MySQL (1/2)

→ Code SQL de création de la table *region* et importation des données du fichier *Region+(2)+(1).csv* dans cette table.

The screenshot displays the MySQL Workbench interface. On the left, the 'Navigator' pane shows the 'project3' database with tables 'contrat' and 'region'. The 'region' table is selected, and its columns are listed: 'Code dep code commune' (varchar, PK), 'reg_code' (int), 'reg_nom' (varchar), 'aca_nom' (varchar), 'dep_nom' (varchar), 'com_nom_maj_court' (varchar), 'dep_code' (varchar), and 'dep_nom_num' (varchar). The main editor shows the SQL code for creating the 'region' table and a query to count its rows.

```
1 • CREATE TABLE region (  
2     Code_dep_code_commune VARCHAR(10) NOT NULL,  
3     reg_code INT NOT NULL,  
4     reg_nom VARCHAR(50) NOT NULL,  
5     aca_nom VARCHAR(50) NOT NULL,  
6     dep_nom VARCHAR(50) NOT NULL,  
7     com_nom_maj_court VARCHAR(50) NOT NULL,  
8     dep_code VARCHAR(3) NOT NULL,  
9     dep_nom_num VARCHAR(50) NOT NULL,  
10    PRIMARY KEY (Code_dep_code_commune)  
11 );  
12 • SELECT COUNT(*) FROM region ;
```

The 'Result Grid' at the bottom shows the result of the query:

COUNT(*)
38916

3. Création des tables dans MySQL (2/2)

→ Code SQL de création de la table *contrat* et importation des données du fichier *Contrat+(2)+(1).csv* dans cette table.

The screenshot displays the MySQL Workbench interface. On the left, the 'SCHEMAS' pane shows the 'projet3' database with tables 'contrat' and 'region'. Below it, the 'Table: contrat' section lists columns with their data types and constraints:

Columns:	
<u>Contrat_ID</u>	int PK
No_voie	int
B_T_Q	char(1)
Type_de_voie	varchar
Voie	varchar
Code_dep_code_commune	varchar
Code_postal	int
Commune	varchar
Code_departement	varchar
Surface	int
Type_local	varchar
Occupation	varchar
Type_contrat	varchar
Formule	varchar
Valeur_declaree_biens	varchar
Prix_cotisation_mensuel	int

The central SQL editor shows the following code:

```
12 SELECT COUNT(*) FROM region ;
13
14 CREATE TABLE contrat (
15     Contrat_ID INT NOT NULL,
16     No_voie INT,
17     B_T_Q CHAR(1),
18     Type_de_voie VARCHAR(5),
19     Voie VARCHAR(50) NOT NULL,
20     Code_dep_code_commune VARCHAR(10) NOT NULL,
21     Code_postal INT NOT NULL,
22     Commune VARCHAR(50) NOT NULL,
23     Code_departement VARCHAR(3) NOT NULL,
24     Surface INT NOT NULL,
25     Type_local VARCHAR(20) NOT NULL,
26     Occupation VARCHAR(20) NOT NULL,
27     Type_contrat VARCHAR(20) NOT NULL,
28     Formule VARCHAR(10) NOT NULL,
29     Valeur_declaree_biens VARCHAR(20) NOT NULL,
30     Prix_cotisation_mensuel INT NOT NULL,
31     PRIMARY KEY (Contrat_ID),
32     FOREIGN KEY (Code_dep_code_commune) REFERENCES region (Code_dep_code_commune)
33 );
34 SELECT COUNT(*) FROM contrat ;
```

The bottom right pane shows the 'Result Grid' with the following data:

COUNT(*)
30335

4. Analyses avec requêtes SQL associées et résultats

Requête 1 - liste des numéros de contrats avec leur surface pour la commune de Caen.

```
SELECT Contrat_ID, Surface, Commune
FROM contrat WHERE Commune =
"Caen";
```

	Contrat_ID	Surface	Commune
▶	103791	35	CAEN
	103792	99	CAEN
	103793	40	CAEN
	103794	20	CAEN

Requête 2 - Liste des numéros de contrats avec le type de contrat et leur formule pour les maisons du département 71.

```
SELECT Contrat_ID, Type_contrat, Formule,
Code_departement, Type_local
FROM contrat
WHERE Type_local = "Maison" AND Code_departement = "71";
```

Result Grid					
	Contrat_ID	Type_contrat	Formule	Code_departement	Type_local
▶	114768	Residence principale	Integral	71	Maison
	114779	Residence principale	Classique	71	Maison
	114782	Residence principale	Classique	71	Maison
	114812	Residence principale	Integral	71	Maison

Requête 3 - Liste des noms des régions de France.

```
SELECT DISTINCT reg_nom
FROM region
ORDER BY reg_nom ASC ;
```

reg_nom
▶ Auvergne-Rhône-Alpes
Bourgogne-Franche-Comté
Bretagne
Centre-Val de Loire
Collectivités d'outre-mer
Corse
Grand Est
Guadeloupe
Guyane
Hauts-de-France
Ile-de-France
La Réunion
Martinique
Mayotte
Normandie
Nouvelle-Aquitaine
Occitanie
Pays de la Loire
Provence-Alpes-Côte d'Azur

4. Analyses avec requêtes SQL associées et résultats

Requête 4 - Combien existe-t-il de contrats sur les résidences principales ?	Requête 5 - Surface moyenne des logements avec un contrat à Paris	Requête 6 - Les 5 contrats qui ont les surfaces les plus élevées																						
<pre>SELECT count(Contrat_ID) AS nb_contrats, Type_contrat FROM contrat WHERE Type_contrat= "Residence principale" ;</pre>	<pre>SELECT Code_departement, avg(Surface) AS moy_surface FROM contrat WHERE Code_departement=75 ;</pre>	<pre>SELECT Contrat_ID, Surface FROM contrat ORDER BY Surface DESC LIMIT 5 ;</pre>																						
<p>Il y a 25 620 contrats pour des résidences principales.</p> <div><div>Result Grid</div><table><thead><tr><th>nb_contrats</th><th>Type_contrat</th></tr></thead><tbody><tr><td>25620</td><td>Residence principale</td></tr></tbody></table></div>	nb_contrats	Type_contrat	25620	Residence principale	<p>La surface moyenne des contrats de Paris est de 51,7695 m²</p> <div><table><thead><tr><th>Code_departement</th><th>moy_surface</th></tr></thead><tbody><tr><td>75</td><td>51.7695</td></tr></tbody></table></div>	Code_departement	moy_surface	75	51.7695	<div><table><thead><tr><th>Contrat_ID</th><th>Surface</th></tr></thead><tbody><tr><td>104211</td><td>815</td></tr><tr><td>105463</td><td>742</td></tr><tr><td>130878</td><td>595</td></tr><tr><td>100822</td><td>570</td></tr><tr><td>109872</td><td>559</td></tr><tr><td>NULL</td><td>NULL</td></tr></tbody></table></div>	Contrat_ID	Surface	104211	815	105463	742	130878	595	100822	570	109872	559	NULL	NULL
nb_contrats	Type_contrat																							
25620	Residence principale																							
Code_departement	moy_surface																							
75	51.7695																							
Contrat_ID	Surface																							
104211	815																							
105463	742																							
130878	595																							
100822	570																							
109872	559																							
NULL	NULL																							

4. Analyses avec requêtes SQL associées et résultats

Requête 7 - Le prix moyen de la cotisation mensuelle	Requête 8 - Le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens	Requête 9 - Le classement des 10 départements où le prix moyen de la cotisation est le plus élevé																																																				
<pre>SELECT avg(Prix_cotisation_mensuel) AS cotis_moy FROM Contrat ;</pre>	<pre>SELECT Valeur_declaree_biens, Count(DISTINCT Contrat_ID) AS nb_contrat FROM contrat GROUP BY Valeur_declaree_biens ;</pre>	<pre>SELECT dep_code,dep_nom, avg(Prix_cotisation_mensuel) as cotis_moy FROM contrat JOIN region USING(code_dep_code_commune) GROUP BY dep_code,dep_nom ORDER BY cotis_moy DESC LIMIT 10;</pre>																																																				
<p>Le prix moyen de la cotisation mensuelle est de 19,33€</p> <table><tr><td></td><td>cotis_moy</td></tr><tr><td>►</td><td>19.3287</td></tr></table>		cotis_moy	►	19.3287	<table><tr><th></th><th>Valeur_declaree_biens</th><th>nb_contrat</th></tr><tr><td>►</td><td>0-25000</td><td>22720</td></tr><tr><td></td><td>100000+</td><td>104</td></tr><tr><td></td><td>25000-50000</td><td>6815</td></tr><tr><td></td><td>50000-100000</td><td>696</td></tr></table>		Valeur_declaree_biens	nb_contrat	►	0-25000	22720		100000+	104		25000-50000	6815		50000-100000	696	<table><tr><th>dep_code</th><th>dep_nom</th><th>cotis_moy</th></tr><tr><td>75</td><td>Paris</td><td>35.3428</td></tr><tr><td>92</td><td>Hauts-de-Seine</td><td>26.4882</td></tr><tr><td>3</td><td>Allier</td><td>20.0000</td></tr><tr><td>94</td><td>Val-de-Marne</td><td>19.7327</td></tr><tr><td>69</td><td>Rhône</td><td>18.9572</td></tr><tr><td>78</td><td>Yvelines</td><td>18.1831</td></tr><tr><td>6</td><td>Alpes-Maritimes</td><td>18.1221</td></tr><tr><td>1</td><td>Ain</td><td>17.9526</td></tr><tr><td>17</td><td>Charente-Maritime</td><td>17.2957</td></tr><tr><td>974</td><td>La Réunion</td><td>17.2500</td></tr></table>	dep_code	dep_nom	cotis_moy	75	Paris	35.3428	92	Hauts-de-Seine	26.4882	3	Allier	20.0000	94	Val-de-Marne	19.7327	69	Rhône	18.9572	78	Yvelines	18.1831	6	Alpes-Maritimes	18.1221	1	Ain	17.9526	17	Charente-Maritime	17.2957	974	La Réunion	17.2500
	cotis_moy																																																					
►	19.3287																																																					
	Valeur_declaree_biens	nb_contrat																																																				
►	0-25000	22720																																																				
	100000+	104																																																				
	25000-50000	6815																																																				
	50000-100000	696																																																				
dep_code	dep_nom	cotis_moy																																																				
75	Paris	35.3428																																																				
92	Hauts-de-Seine	26.4882																																																				
3	Allier	20.0000																																																				
94	Val-de-Marne	19.7327																																																				
69	Rhône	18.9572																																																				
78	Yvelines	18.1831																																																				
6	Alpes-Maritimes	18.1221																																																				
1	Ain	17.9526																																																				
17	Charente-Maritime	17.2957																																																				
974	La Réunion	17.2500																																																				

4. Analyses avec requêtes SQL associées et résultats

Requête 10 - Le nombre de contrats avec des formules "Intégral" pour la région Pays de la Loire

```
SELECT reg_nom, count(distinct Contrat_ID) AS nb_contrats,
Formule
FROM contrat JOIN region ON
contrat.code_dep_code_commune =
region.code_dep_code_commune
WHERE Formule = "Integral" AND reg_nom = "Pays de la Loire";
```

reg_nom	nb_contrats	Formule
Pays de la Loire	561	Integral

Requête 11 - La liste des communes ayant au moins 150 contrats

```
SELECT com_nom_maj_court, count(Contrat_ID) AS
nb_contrat
FROM Contrat JOIN region
USING(code_dep_code_commune)
GROUP BY com_nom_maj_court
HAVING count(Contrat_ID) >= 150
ORDER BY nb_contrat DESC;
```

com_nom_maj_court	nb_contrat
PARIS 18	515
PARIS 17	468
PARIS 15	407
PARIS 16	394
NICE	387
PARIS 11	381
BORDEAUX	302
PARIS 20	302
NANTES	291
PARIS 19	266
PARIS 10	263
PARIS 12	252
PARIS 14	222
GRENOBLE	220
PARIS 9	204
TOULOUSE	187
TOULON	170
COURBEVOIE	163
LILLE	161
PARIS 3	159

Requête 12 - Le nombre de contrats pour chaque région

```
SELECT reg_nom, count(Contrat_ID) AS
nb_contrat FROM Contrat JOIN region USING
(code_dep_code_commune) GROUP BY reg_nom
HAVING count(Contrat_ID)
ORDER BY reg_nom ;
```

reg_nom	nb_contrat
Auvergne-Rhône-Alpes	2972
Bourgogne-Franche-Comté	402
Bretagne	945
Centre-Val de Loire	804
Corse	247
Grand Est	806
Guyane	37
Hauts-de-France	1327
Île-de-France	13474
La Réunion	4
Martinique	60
Normandie	898
Nouvelle-Aquitaine	2097
Occitanie	1837
Pays de la Loire	1138
Provence-Alpes-Côte d'Azur	3287