

PRESENTATION DE LA METHODOLOGIE



CONTEXTE

Dans le cadre d'un projet :

Créer une base de données en respect des normes réglementaires et des besoins clients

Gérer une base de données établissant des requêtes et en respect du RGPD

RESSOURCES

- Données de 2 fichiers .csv intitulés « contrats clients » et « référentiel »
- Template d'un dictionnaire des données à compléter
- Ebauche du schéma relationnel à compléter

BESOINS

- Elaboration du dictionnaire des données explorées complété
- Conception du schéma relationnel de base de données
- Création du code SQL générant les tables dans une base de données (BDD)
- Elaboration de requêtes SQL

ORGANISATION

- 1 – Analyse des données et création du dictionnaire
- 2 – Création du schéma relationnel
- 3 – Organisation des données par tables
- 4 – Création de la base de données et insertion des données
- 5 – Analyse des données via des requêtes SQL

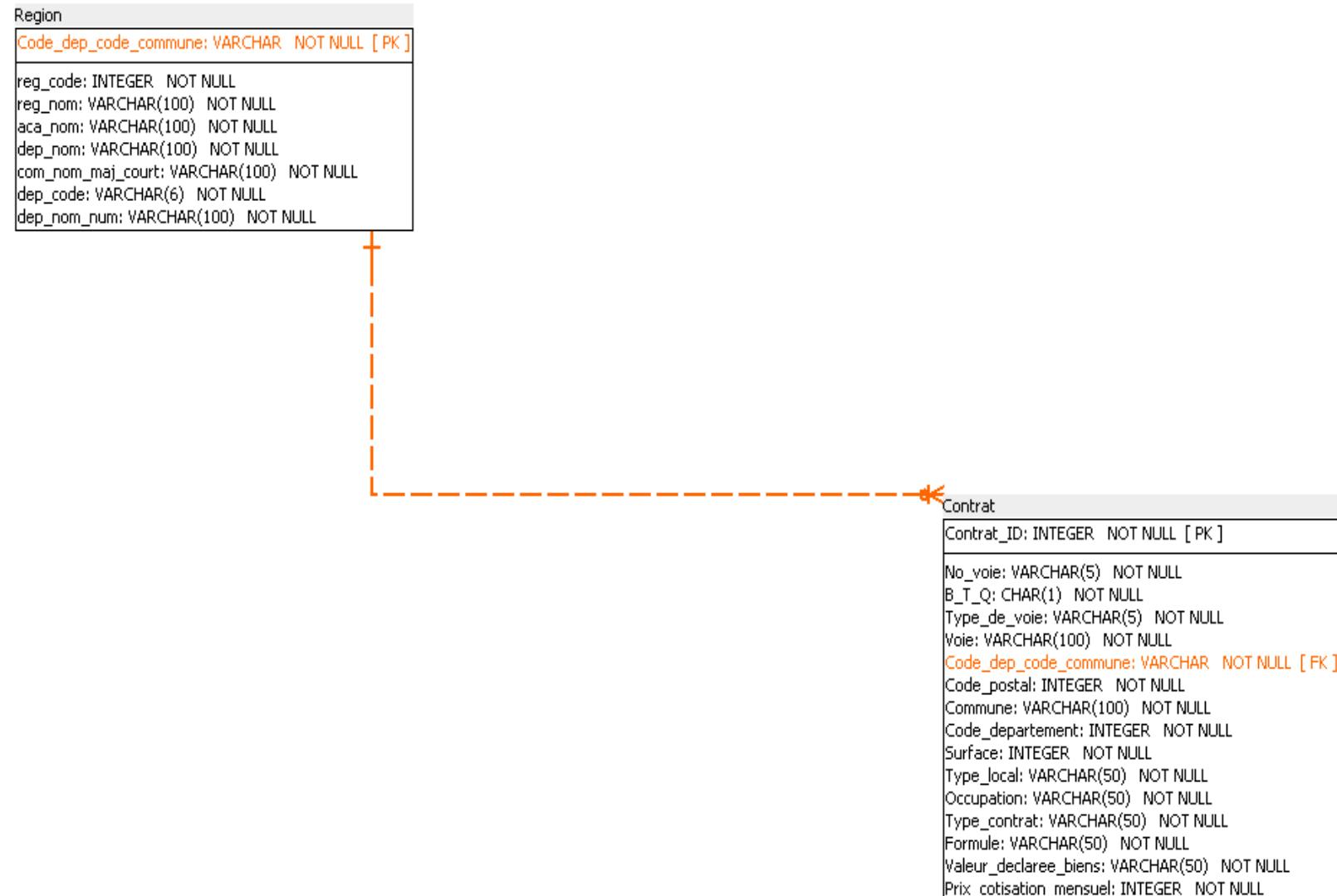
- Report des données dans le dictionnaire
- Regroupement des données par fichiers et par tables

1 - CRÉATION DU DICTIONNAIRE

	Nom des colonnes	Type de données	Taille	Clé	Description
CONTRAT.CSV	Contrat_ID	INT		Clé primaire	Id unique pour les contrats
	No_voie	VARCHAR	5		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	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 unique
	Code_postal	INT			Code postal pour l'adresse du logement assuré
	Commune	VARCHAR	100		Libellé de la commune de l'adresse du logement
	Code_departement	INT			Numéro du département du logement assuré
	Surface	INT			Surface du logement assuré
	Type_local	VARCHAR	50		Type de logement
	Occupation	VARCHAR	50		Statut de l'occupant du logement
	Type_contrat	VARCHAR	50		Désignation du type de contrat du logement
	Formule	VARCHAR	50		Type de formule du logement
REGION.CSV	Valeur_declaree_biens	VARCHAR	50		Tranche de valeurs déclarées pour le logement
	Prix_cotisation_mensuel	INT			Montant de la cotisation du contrat
	Code_dep_code_commune	VARCHAR	6	Clé primaire	Concaténation du code département et code commune pour avoir une clé unique
	reg_code	INT			Code de la région
	reg_nom	VARCHAR	100		Nom de la région
	aca_nom	VARCHAR	100		Prefecture de la région
	dep_nom	VARCHAR	100		Nom du département

2 – CRÉATION DU SCHÉMA RELATIONNEL

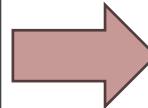
- Utilisation de l'outil SQL Power Schema
- Création des tables et liens grâce au dictionnaire de données



3 – ORGANISATION DES DONNÉES PAR TABLES

Exemple avec la table region

Region	
<code>Code_dep_code_commun: VARCHAR NOT NULL [PK]</code>	
reg_code: INTEGER 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(6) NOT NULL	
dep_nom_num: VARCHAR(100) NOT NULL	



Code_dep_code_commun	reg_code	reg_nom	aca_nom	dep_nom	com_nom_maj_court	dep_code	dep_nom_num
1001	84	Auvergne-Rhône	Lyon	Ain	LABERGEMENT	1	Ain (01)
1002	84	Auvergne-Rhône	Lyon	Ain	LABERGEMENT	1	Ain (01)
1003	84	Auvergne-Rhône	Lyon	Ain	AMAREINS	1	Ain (01)
1004	84	Auvergne-Rhône	Lyon	Ain	AMBERIEU EN	1	Ain (01)
1005	84	Auvergne-Rhône	Lyon	Ain	AMBERIEUX EN	1	Ain (01)
1006	84	Auvergne-Rhône	Lyon	Ain	AMBLEON	1	Ain (01)
1007	84	Auvergne-Rhône	Lyon	Ain	AMBRONAY	1	Ain (01)
1008	84	Auvergne-Rhône	Lyon	Ain	AMBUTRIX	1	Ain (01)
1009	84	Auvergne-Rhône	Lyon	Ain	ANDERT ET CC	1	Ain (01)
1010	84	Auvergne-Rhône	Lyon	Ain	ANGLEFORT	1	Ain (01)
1011	84	Auvergne-Rhône	Lyon	Ain	APREMONT	1	Ain (01)
1012	84	Auvergne-Rhône	Lyon	Ain	ARANC	1	Ain (01)
1013	84	Auvergne-Rhône	Lyon	Ain	ARANDAS	1	Ain (01)
1014	84	Auvergne-Rhône	Lyon	Ain	ARBENT	1	Ain (01)
1015	84	Auvergne-Rhône	Lyon	Ain	ARBIGNIEU	1	Ain (01)
1016	84	Auvergne-Rhône	Lyon	Ain	ARBIGNY	1	Ain (01)
1017	84	Auvergne-Rhône	Lyon	Ain	ARGIS	1	Ain (01)
1018	84	Auvergne-Rhône	Lyon	Ain	ARLOD	1	Ain (01)
1019	84	Auvergne-Rhône	Lyon	Ain	ARMIX	1	Ain (01)

- Création d'un fichier Excel pour chaque table
- Copie des données des fichiers Excel créés
- Ordonnées les données en respectant le MCD

4 – CRÉATION DE LA BASE DE DONNÉES ET INSERTION DES DONNÉES

- Création de la base de données sur MySQL Workbench

1. CREATE DATABASE immo_contrat;

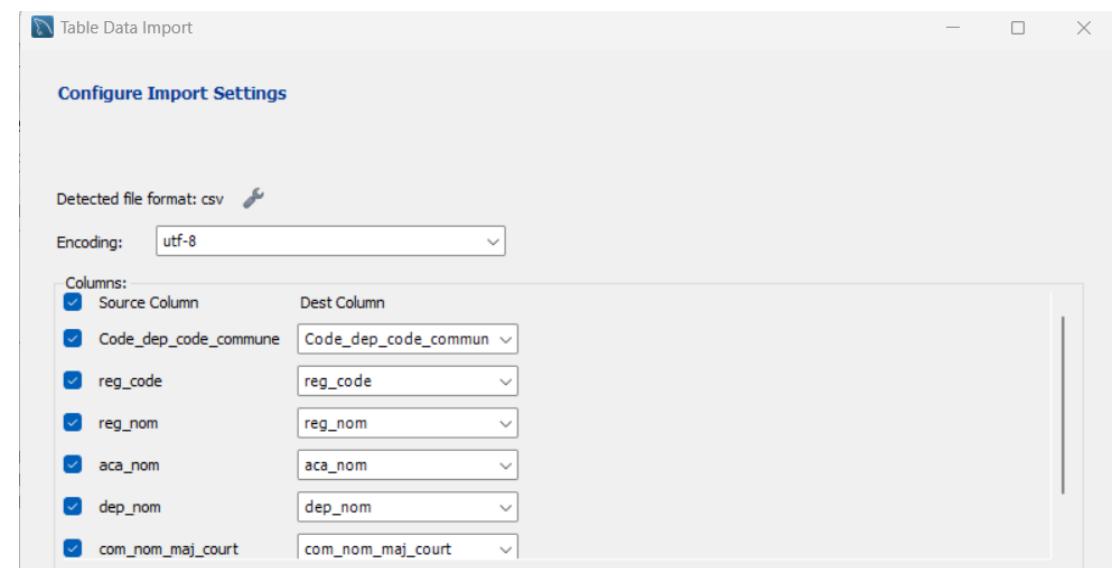
- Création des tables grâce au script SQL

```
CREATE TABLE Region (
    Code_dep_code_commune VARCHAR(6) NOT NULL PRIMARY KEY,
    reg_code INT 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(6) NOT NULL,
    dep_nom_num VARCHAR(100) NOT NULL)

CREATE TABLE Contrat (
    Contrat_ID INTEGER NOT NULL PRIMARY KEY,
    No_voie VARCHAR(5) NOT NULL,
    B_T_Q CHAR(1) NOT NULL,
    Type_de_voie VARCHAR(5) NOT NULL,
    Voie VARCHAR(100) NOT NULL,
    Code_dep_code_commune VARCHAR(6),
    Code_postal INTEGER NOT NULL,
    Commune VARCHAR(100) NOT NULL,
    Code_departement INTEGER NOT NULL,
    Surface INTEGER NOT NULL,
    Type_local VARCHAR(50) NOT NULL,
    Occupation VARCHAR(50) NOT NULL,
    Type_contrat VARCHAR(50) NOT NULL,
    Formule VARCHAR(50) NOT NULL,
    Valeur_declaree_biens VARCHAR(50) NOT NULL,
    Prix_cotisation_mensuel INTEGER NOT NULL,
    FOREIGN KEY (Code_dep_code_commune) REFERENCES region(Code_dep_code_commune),
```

4 – CRÉATION DE LA BASE DE DONNÉES ET INSERTION DES DONNÉES

- Enregistrer les fichiers Excel au format CSV
- Importation des fichiers CSV correspondant aux bonnes tables dans le bon ordre



5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 1:** Lister les numéros de contrats (contrat_ID) avec leur surface pour la commune de Caen.

Requête

```
SELECT contrat_ID, surface  
FROM contrat where commune =  
'caen'
```

Résultat

	contrat_ID	surface
▶	103791	35
	103792	99
	103793	40
	103794	20

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

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

Requête

```
SELECT contrat_ID, Type_contrat,  
Formule  
FROM contrat  
WHERE Code_departement = '71'  
AND Type_local = 'maison'
```

Résultat

	contrat_ID	Type_contrat	Formule
▶	114768	Residence principale	Integral
	114779	Residence principale	Classique
	114782	Residence principale	Classique
	114812	Residence principale	Integral

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 3**: Lister le nom des régions de France.

Requête

```
SELECT distinct reg_nom  
FROM immo_contrat.region
```

Résultat

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

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 4**: Combien existe-t-il de contrats sur les résidences principales ?

Requête

```
SELECT COUNT(Contrat_ID) AS  
Nombre_Contrats_Residence_Princip  
ale  
  
FROM contrat  
WHERE Type_contrat = 'Residence  
principale' ;
```

Résultat

	Nombre_Contrats_Residence_Principale
▶	25620

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 5:** Quelle est la surface moyenne des logements avec un contrat à Paris ?

Requête

```
SELECT avg(Surface) AS  
Surface_moyenne_des_logements_à_  
Paris  
FROM contrat  
WHERE Code_departement = '75' ;
```

Résultat

	Surface_moyenne_des_logements_à_Paris
▶	51.7695

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 6 :** Quels sont les 5 contrats qui ont les surfaces les plus élevées ?

Requête

```
SELECT Contrat_ID, Surface AS
5_Contrat_surface_la_plus_élevée
FROM contrat
ORDER BY Surface DESC limit 5 ;
```

Résultat

	Contrat_ID	5_Contrat_surface_la_plus_élevée
▶	104211	815
	105463	742
	130878	595
	100822	570
	109872	559
*	HULL	NULL

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 7:** Quel est le prix moyen de la cotisation mensuelle ?

Requête

```
SELECT avg(Prix_cotisation_mensuel)
AS
Prix_moyen_de_la_cotisation_mensue
lle
FROM contrat ;
```

Résultat

	Prix_moyen_de_la_cotisation_mensuelle
▶	19.3287

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 8 :** Quel est le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens ?

Requête

```
SELECT Valeur_declaree_biens,  
COUNT(*) AS  
Nombre_de_contrats_pour_chaque_c  
atégorie_de_prix  
FROM contrat  
GROUP BY Valeur_declaree_biens ;
```

Résultat

Valeur_dedaree_biens	Nombre_de_contrats_pour_chaque_catégorie_de_prix
0-25000	22720
25000-50000	6815
100000+	104
50000-100000	696

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 9 :** Quel est le classement des 10 départements où le prix moyen de la cotisation est le plus élevé ?

Requête

```
SELECT Code_departement,  
AVG(Prix_cotisation_mensuel) AS  
Prix_moyen_cotisation  
FROM contrat  
GROUP BY Code_departement  
ORDER BY Prix_moyen_cotisation  
DESC LIMIT 10;
```

Résultat

	Code_departement	Prix_moyen_cotisation
▶	75	36.3991
	92	26.2694
	94	19.8242
	78	18.8802
	69	18.4568
	1	18.2389
	6	18.1376
	17	17.3173
	74	17.1584
	20	17.0283

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 10 :** Quel est le nombre de contrats avec des formules « intégral » pour la région Pays de la Loire ?

Requête

```
SELECT COUNT(*) AS
Nombre_Contrats_Integral
FROM contrat
INNER JOIN region ON
contrat.Code_dep_code_commune =
region.Code_dep_code_commune
WHERE contrat.Formule = 'integral'
AND region.reg_nom = 'Pays de la
Loire';
```

Résultat

	Nombre_Contrats_Integral
▶	561

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 11:** Quelle est la liste des communes ayant au moins 150 contrats ?

Requête

```
SELECT region.Code_dep_code_commune,  
region.com_nom_maj_court AS Ville, COUNT(*) AS  
Nombre_Contrats  
  
FROM contrat  
  
INNER JOIN region ON  
contrat.Code_dep_code_commune =  
region.Code_dep_code_commune  
  
GROUP BY region.Code_dep_code_commune  
  
HAVING COUNT(*) >= 150  
  
ORDER BY Nombre_Contrats DESC;
```

Résultat

	Code_dep_code_commune	Ville	Nombre_Contrats
▶	75118	PARIS 18	515
	75117	PARIS 17	468
	75115	PARIS 15	407
	75116	PARIS 16	394
	6088	NICE	387
	75111	PARIS 11	381
	33063	BORDEAUX	302
	75120	PARIS 20	302
	44109	NANTES	291
	75119	PARIS 19	266
	75110	PARIS 10	263
	75112	PARIS 12	252
	75114	PARIS 14	222
	38185	GRENOBLE	220
	75109	PARIS 9	204
	31555	TOULOUSE	187
	83137	TOULON	170
	92026	COURBEVOIE	163
	59350	LILLE	161
	75103	PARIS 3	159

5 – ANALYSE DES DONNÉES VIA DES REQUÊTES SQL

- **Requête 12 :** Quel est le nombre de contrats pour chaque région ?

Requête

```
SELECT region.reg_nom, COUNT(*) AS  
Nombre_Contrats  
FROM contrat  
INNER JOIN region ON  
contrat.Code_dep_code_commune =  
region.Code_dep_code_commune  
GROUP BY region.reg_nom  
ORDER BY Nombre_Contrats DESC;
```

Résultat

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