

EXPLORATION DES DONNEES CONTRATS

BASE DE DONNEES ET REQUETES MYSQL
METHODOLOGIE

IMPORT , ANALYSE ET PREPARATION DES DONNEES

- ▶ Import et analyse du fichier de données concernant les contrats immobiliers
- ▶ Repérage des champs et des données manquantes indispensables : code commune mal récupéré
- ▶ Récupération des codes communes manquants par correspondance avec les codes postaux : recours à un fichier libre ressource adéquat de data-gouv.fr
<https://www.data.gouv.fr/fr/datasets/correspondance-entre-les-codes-postaux-et-codes-insee-des-communes-francaises/>
- ▶ Import et analyse du fichier ressources Régions <https://www.data.gouv.fr/fr/datasets/referentiel-geographique-francais-communes-unites-urbaines-aires-urbaines-departements-academies-regions-1/#resources>
- ▶ Suppression des nombreux champs non pertinents du fichier Régions
- ▶ Mise en forme des deux fichiers Contrats et Régions et conversion au format csv pour import à venir

CREATION DE LA BASE DE DONNEES

MODELISATION ET DICTIONNAIRE DES DONNEES

| | 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 | CVARCHAR | 4 | | Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ... |
| | Voie | CVARCHAR | 26 | | Libellé de la voie pour l'adresse du logement assuré |
| | Code_dep_code_commune | CVARCHAR | 5 | Clé secondaire | Concaténation du code département et code commune pour avoir une clé unique |
| | Code_postal | CVARCHAR | 5 | | Code postal pour l'adresse du logement assuré |
| | Surface | INT | 3 | | Surface du logement |
| | Type_local | CVARCHAR | 11 | | Type du local : appartement, maison, |
| | Occupation | CVARCHAR | 12 | | Nature de l'occupant : propriétaire ou locataire |
| | Type_contrat | CVARCHAR | 20 | | Usage : résidence principale, résidence secondaire, mise en location |
| | Formule | CVARCHAR | 9 | | Intégrale ou classique |
| | Valeur_declaree_biens | CVARCHAR | 12 | | Valeur (tranches) : 0-25000, 25000-50000, 50000-100000 |
| | Prix_cotisation_mensuel | INT | 3 | | Valeur de la cotisation mensuelle |
| REGION.CSV | Code_dep_code_commune | CVARCHAR | 5 | Clé primaire | Concaténation du code département et code commune pour avoir une clé unique |
| | Com_nom | CVARCHAR | 45 | | Nom de la commune |
| | Reg_nom | CVARCHAR | 26 | | Nom de la région |
| | Dep_code | CVARCHAR | 3 | | Numéro du département |
| | Dep_nom | CVARCHAR | 43 | | Nom du département |

Mise en correspondance entre dictionnaire des données, fichiers csv et base de données à créer

CREATION DE LA BASE DE DONNEES

MODELISATION AVEC L'OUTIL SQL POWER ARCHITECT©



- Création des tables et description des champs conformes au dictionnaire des données
- Relation un à plusieurs entre les tables Region et Contrat
- Génération de la requête SQL de création des tables

CREATION DE LA BASE DE DONNEES

MISE EN EXPLOITATION AVEC L'OUTIL MYSQL WORKBENCH 8.0 CE

Par requêtes SQL (cf. document technique) : création du schéma db_contrats et des tables, importation des données dans les deux tables

TABLE CONTRAT

| Field | Type | Null | Key | Default | Extra |
|-------------------------|-------------|------|-----|---------|-------|
| Contrat_ID | int | NO | PRI | | |
| No_voie | int | YES | | | |
| B_T_Q | varchar(1) | YES | | | |
| Type_de_voie | varchar(4) | YES | | | |
| Voie | varchar(26) | NO | | | |
| Code_dep_code_commune | varchar(5) | NO | MUL | | |
| Code_postal | varchar(5) | NO | | | |
| Surface | int | NO | | | |
| Type_local | varchar(11) | NO | | | |
| Occupation | varchar(12) | NO | | | |
| Type_contrat | varchar(20) | NO | | | |
| Formule | varchar(9) | NO | | | |
| Valeur_declaree_biens | varchar(12) | NO | | | |
| Prix_cotisation_mensuel | int | NO | | | |

| Result Grid | Filter Rows |
|-------------------|-------------|
| Nb_lignes_contrat | |
| 30335 | |

TABLE REGION

| Field | Type | Null | Key | Default | Extra |
|-----------------------|-------------|------|-----|---------|-------|
| Code_dep_code_commune | varchar(5) | NO | PRI | | |
| Com_nom | varchar(45) | NO | | | |
| Reg_nom | varchar(26) | NO | | | |
| Dep_code | varchar(3) | NO | | | |
| Dep_nom | varchar(43) | NO | | | |

| Result Grid | Filter Rows |
|------------------|-------------|
| Nb_lignes_region | |
| 38916 | |

EXTRACTIONS DE LA BASE DE DONNEES

REQUETES SQL ET RESULTATS

```
1 -- Requête 1 : Lister les numéros de contrats (contrat_ID) avec leur surface pour la commune de Caen
2 • SELECT Contrat_ID, Surface
3 FROM contrat
4 JOIN region
5 ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6 WHERE Com_nom = 'CAEN';
```

| Contrat_ID | Surface |
|------------|---------|
| 103791 | 35 |
| 103792 | 99 |
| 103793 | 40 |
| 103794 | 20 |

```
1 -- Requête 2 : Lister les numéros de contrats avec le type de contrat et leur formule pour les maisons du département 71.
2 • SELECT Contrat_ID, Type_contrat, Formule
3 FROM contrat
4 JOIN region
5 ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6 WHERE Dep_code = '71' AND Type_local = 'Maison' ;
```

| Contrat_ID | Type_contrat | Formule |
|------------|----------------------|-----------|
| 114768 | Residence principale | Integral |
| 114779 | Residence principale | Classique |
| 114782 | Residence principale | Classique |
| 114812 | Residence principale | Integral |

```
1 -- Requête 3 : Lister le nom des régions de France.
2 • SELECT DISTINCT Reg_nom AS 'Région'
3 FROM region
4 ORDER BY Reg_nom ASC;
```

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

```
1 -- Requête 4 : Quels sont les 5 contrats qui ont les surfaces les plus élevées ?
2 • SELECT Contrat_ID, Surface
3 FROM contrat
4 ORDER BY Surface DESC
5 LIMIT 5;
```

| Contrat_ID | Surface |
|------------|---------|
| 104211 | 815 |
| 105463 | 742 |
| 130878 | 595 |
| 100822 | 570 |
| 109872 | 559 |
| NULL | NULL |

EXTRACTIONS DE LA BASE DE DONNEES

REQUETES SQL ET RESULTATS

```
1 -- Requête 5 : Quel est le prix moyen de la cotisation mensuelle ?
2 • SELECT AVG(Prix_cotisation_mensuel) AS 'Prix moyen cotisation mensuelle'
3 FROM contrat ;
```

| Result Grid | | Filter Rows: | Export: | Wrap Cell Content: |
|---------------------------------|--|--------------|---------|--------------------|
| Prix moyen cotisation mensuelle | | | | |
| 19.3287 | | | | |

```
1 -- Requête 6 : Quel est le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens ?
2 • SELECT Valeur_declaree_biens, COUNT(Prix_cotisation_mensuel) AS 'Nombre de contrats'
3 FROM contrat
4 GROUP BY Valeur_declaree_biens
5 ORDER BY COUNT(Prix_cotisation_mensuel) DESC;
```

| Result Grid | | Filter Rows: | Export: | Wrap Cell Content: |
|-----------------------|--------------------|--------------|---------|--------------------|
| Valeur_declaree_biens | Nombre de contrats | | | |
| 0-25000 | 22720 | | | |
| 25000-50000 | 6815 | | | |
| 50000-100000 | 696 | | | |
| 100000+ | 104 | | | |

```
1 -- Requête 7 : Quel est le nombre de formule "integral" sur la région Pays de la Loire ?
2 • SELECT COUNT(Formule) AS 'Nb formules intégrales en Pays de la Loire'
3 FROM contrat
4 JOIN region
5 ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6 WHERE Reg_nom = 'Pays de la Loire' AND Formule = 'Integral';
```

| Result Grid | | Filter Rows: | Export: | Wrap Cell Content: |
|--|--|--------------|---------|--------------------|
| Nb formules intégrales en Pays de la Loire | | | | |
| 589 | | | | |

```
1 -- Requête 8 : Lister les numéros de contrats avec le type de contrat et leur formule pour les maisons du département 71.
2 • SELECT Contrat_ID, Type_contrat, Formule
3 FROM contrat
4 JOIN region
5 ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6 WHERE Dep_code = '71' AND Type_local = 'Maison';
```

| Result Grid | | Filter Rows: | Export: | Wrap Cell Content: |
|-------------|----------------------|--------------|---------|--------------------|
| Contrat_ID | Type_contrat | Formule | | |
| 114768 | Residence principale | Integral | | |
| 114779 | Residence principale | Classique | | |
| 114782 | Residence principale | Classique | | |
| 114812 | Residence principale | Integral | | |

EXTRACTIONS DE LA BASE DE DONNEES

REQUETES SQL ET RESULTATS

```
1  -- Requête 9 : Quelle est la surface moyenne des contrats à Paris ?
2  • SELECT AVG(Surface) AS 'Surface moyenne des contrats à Paris'
3    FROM contrat
4      JOIN region
5      ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6    WHERE Com_nom LIKE 'PARIS%';
```

| Result Grid | | Filter Rows: | Export: | Wrap Cell Content: |
|-------------|--------------------------------------|--------------|---------|--------------------|
| | Surface moyenne des contrats à Paris | | | |
| ▶ | 51.7695 | | | |

```
1  -- Requête 10 : Classements des 10 départements où le prix moyen de la cotisation est le plus élevé.
2  • SELECT Dep_nom AS 'Département', AVG(Prix_cotisation_mensuel) AS 'Prix moyen de la cotisation mensuelle'
3    FROM contrat
4      JOIN region
5      ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6    GROUP BY Dep_nom
7    ORDER BY AVG(Prix_cotisation_mensuel) DESC
8    LIMIT 10 ;
```

| Result Grid | | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows: |
|-------------|-------------------|---------------------------------------|---------|--------------------|-------------|
| | Département | Prix moyen de la cotisation mensuelle | | | |
| ▶ | Paris | 36.3991 | | | |
| | Hauts-de-Seine | 26.2694 | | | |
| | Val-de-Marne | 19.8242 | | | |
| | Yvelines | 18.8881 | | | |
| | Rhône | 18.4855 | | | |
| | Ain | 18.2389 | | | |
| | Alpes-Maritimes | 18.1376 | | | |
| | Charente-Maritime | 17.3173 | | | |
| | Haute-Savoie | 17.1487 | | | |
| | Corse-du-Sud | 17.0744 | | | |

EXTRACTIONS DE LA BASE DE DONNEES

REQUETES SQL ET RESULTATS

```
1  -- Requête 11 : Liste des communes ayant eu au moins 150 contrats
2  • SELECT Com_nom AS 'Commune', COUNT(Contrat_ID) AS 'Nombre de contrats'
3  FROM contrat
4      JOIN region
5      ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6  GROUP BY Com_nom
7  HAVING COUNT(Contrat_ID) >= 150
8  ORDER BY COUNT(Contrat_ID) DESC ;
```

| Result Grid | | | Filter Rows: | Export: | Wrap Cell Content: |
|--------------------------|--------------------|--|--------------|---------|--------------------|
| Commune | Nombre de contrats | | | | |
| PARIS 18E ARRONDISSEMENT | 515 | | | | |
| PARIS 17E ARRONDISSEMENT | 468 | | | | |
| PARIS 15E ARRONDISSEMENT | 407 | | | | |
| PARIS 16E ARRONDISSEMENT | 394 | | | | |
| NICE | 387 | | | | |
| PARIS 11E ARRONDISSEMENT | 381 | | | | |
| BORDEAUX | 302 | | | | |
| PARIS 20E ARRONDISSEMENT | 302 | | | | |
| NANTES | 291 | | | | |
| PARIS 19E ARRONDISSEMENT | 266 | | | | |
| PARIS 10E ARRONDISSEMENT | 263 | | | | |
| PARIS 12E ARRONDISSEMENT | 252 | | | | |
| PARIS 14E ARRONDISSEMENT | 222 | | | | |
| GRENOBLE | 220 | | | | |
| PARIS 9E ARRONDISSEMENT | 204 | | | | |
| TOULOUSE | 187 | | | | |
| TOULON | 170 | | | | |
| COURBEVOIE | 163 | | | | |
| LILLE | 161 | | | | |
| PARIS 3E ARRONDISSEMENT | 159 | | | | |

```
1  -- Requête 12 : Quel est le nombre de contrats pour chaque région ?
2  • SELECT Reg_nom AS 'Région', COUNT(Contrat_ID) AS 'Nombre de contrats'
3  FROM contrat
4      JOIN region
5      ON contrat.Code_dep_code_commune = region.Code_dep_code_commune
6  GROUP BY Reg_nom
7  ORDER BY COUNT(Contrat_ID) DESC ;
```

| Result Grid | | | Filter Rows: | Export: | Wrap Cell Content: |
|----------------------------|--------------------|--|--------------|---------|--------------------|
| Région | Nombre de contrats | | | | |
| Île-de-France | 14177 | | | | |
| Provence-Alpes-Côte d'Azur | 3279 | | | | |
| Auvergne-Rhône-Alpes | 3042 | | | | |
| Nouvelle-Aquitaine | 2038 | | | | |
| Occitanie | 1609 | | | | |
| Pays de la Loire | 1196 | | | | |
| Hauts-de-France | 1189 | | | | |
| Bretagne | 947 | | | | |
| Normandie | 824 | | | | |
| Grand Est | 769 | | | | |
| Centre-Val de Loire | 598 | | | | |
| Bourgogne-Franche-Comté | 293 | | | | |
| Corse | 247 | | | | |
| Martinique | 73 | | | | |
| Guyane | 37 | | | | |
| La Réunion | 17 | | | | |