Création et utilisation de la base de données

Aurélie RIVIERE

Laplace Immo

Contexte du projet DATA Immo

Modélisation d'une base de donner collectant les transactions immobilières et foncières en France.

- → Analyse du marché global
- → Analyse du marché local pour aider les agences régionales

1. Structuration et création de la base de données

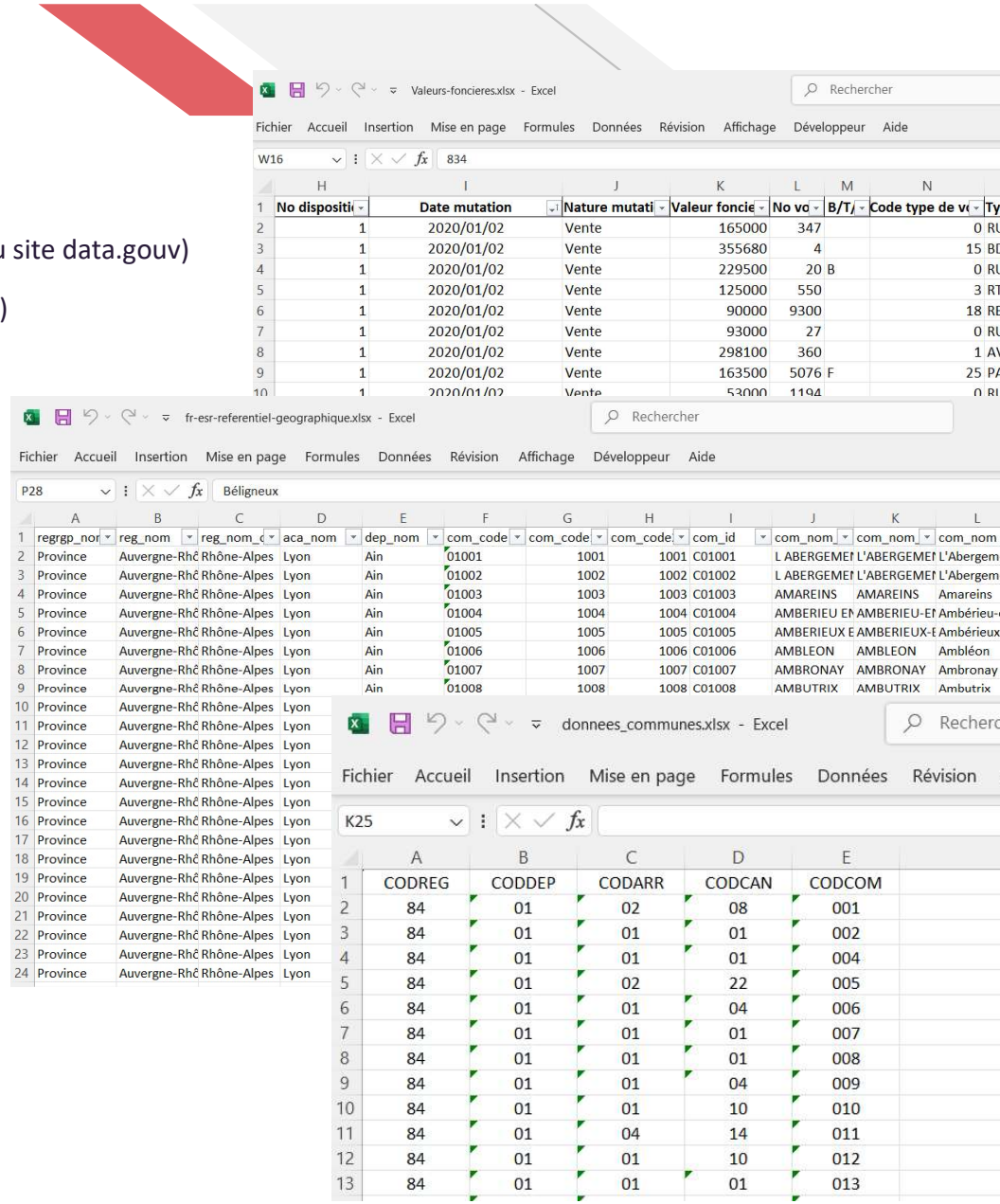
2. Implémentation de données sur une période test et analyse des résultats (1er semestre 2020)

Les données initiales

- Données du référentiel géographique français (extraites du site data.gouv)
- Résultats des recensements de la population (source INSEE)
- Données des Valeurs Foncières (extraites du site open data)

Analyse initiale des fichiers ressources pour comprendre la nature et la structure des informations.

- Conversion des données CSV en Tableau Excel
- Utilisation des tri et des filtres pour déterminer la nature des données.
- Analyse des redondances d'informations et des données absentes. Analyse de la nature des liens entre les données.
- Utilisation de formules Excel, tel que NBCAR et CONCAT afin de déterminer la longueur des données.



La stratégie de sauvegarde et la conformité RGPD

Mise en place d'une stratégie de sauvegarde

- Données issus de sources externes récupérables
- Données pouvant être régulièrement mise à jour depuis leur source
- Données ayant besoin d'être restructurées avant utilisation

→ Prévoir une sauvegarde des données restructurées à chaque mise à jour.

Mise en conformité RGPD

- Données sources contenant des informations nominatives
- Données source non conforme RGPD (nom de l'acquéreur dans les ventes immobilières)

→ Suppression des données nominatives afin d'être en conformité RGPD

L’extrait du dictionnaire des données

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
Table : Bien						
Id_bien	Identifiant unique de chaque bien dans la base	INT	10	Elémentaire (PK)	NOT NULL	AUTO INCREMENT
Id_codedep_codecommune	Concaténation du code département et code commune	VARCHAR	6	(FK)	NOT NULL	Dep_code + Code_commune
No_voie	Numéro des rues	Integer	NC	Elémentaire		
BTQ	Indice de répétition	CHAR	1	Elémentaire		
Type_voie	Plusieurs valeurs (rue, avenue, chemin, etc.)	VARCHAR	4	Elémentaire		
Voie	Nom de la rue	VARCHAR	50	Elémentaire	NOT NULL	
Id_lot	identifiant du 1er lot	VARCHAR	10	Elémentaire	NOT NULL	
Code_postal	Code postal	CHAR	5	Elémentaire	NOT NULL	
Surface_carrez	Surface du bien vendu	DECIMAL	(7, 2)	Elémentaire	NOT NULL	
Surface_local	Surface mesurée au sol	INT	4	Elémentaire	NOT NULL	
Total_piece	Nombre pieces principales	INT	3	Elémentaire	NOT NULL	
Surface_terrain	Surface du terrain	INT	5	Elémentaire		
Type_local	Type de local (maison, appartement,etc..)	VARCHAR	50	Elémentaire	NOT NULL	
Table : Vente						
Id_vente	Identifiant unique de chaque vente	INT	10	PRIMARY KEY	NOT NULL	AUTO INCREMENT
Id_bien	Identifiant unique de chaque bien dans la base	INT	10	FK	NOT NULL	
Date	Date de signature de l’acte	DATE		Elémentaire	JJ/MM/AAAA	
Valeur	Prix de vente du bien	DECIMAL	(13, 2)	Elémentaire		
Table : Commune						
Id_codedep_codecommune	Concaténation du code département et code commune	VARCHAR	6	PRIMARY KEY	NOT NULL	
Dep_code	Code du département	VARCHAR	3	Elémentaire (FK)	NOT NULL	
Code_commune	Code de la commune	VARCHAR	3	Elémentaire	NOT NULL	
Nom_commune	Libellé de la commune	VARCHAR	50	Elémentaire	NOT NULL	
Com_nom_maj_court	Libellé de la commune en majuscule	VARCHAR	50	Elémentaire	NOT NULL	
Population_totale	Population totale (somme de la population municipale et d	INT	50	Elémentaire	NOT NULL	
Table : Departement						
Dep_code	Code départemental	VARCHAR	3	Elémentaire (PK)	NOT NULL	
Dep_nom	Nom du département concerné	VARCHAR	50	Elémentaire	NOT NULL	
Reg_code	Code de la région	VARCHAR	2	FK	NOT NULL	
Table : Region						
Reg_code	Code de la région	VARCHAR	2	Elémentaire (PK)	NOT NULL	
Reg_nom	Nom de la région concernée	VARCHAR	30	Elémentaire	NOT NULL	

Table Bien

Table Vente

Table Commune

Table Departement

Table Region

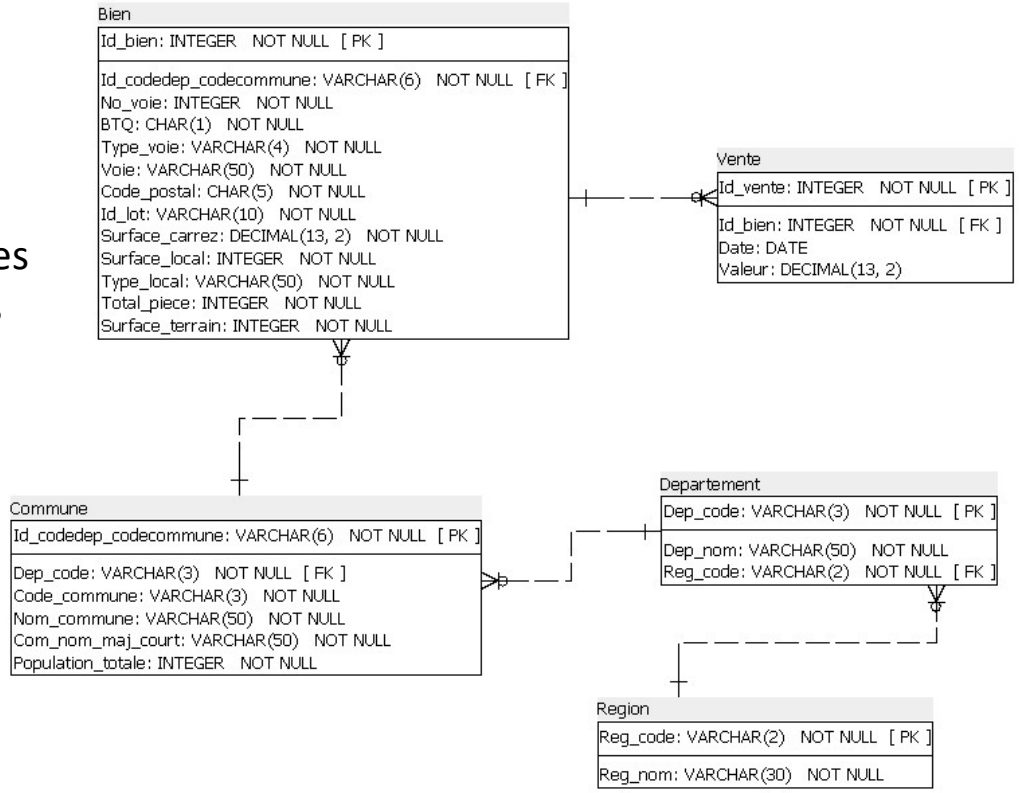
Clés Primaire (PK)

Clés Etrangère (FK)

Le schéma relationnel normalisé

Mise en place du schéma relationnel avec le logiciel SQL Power Architect, permettant de visualiser et de créer les relations entre les tables.

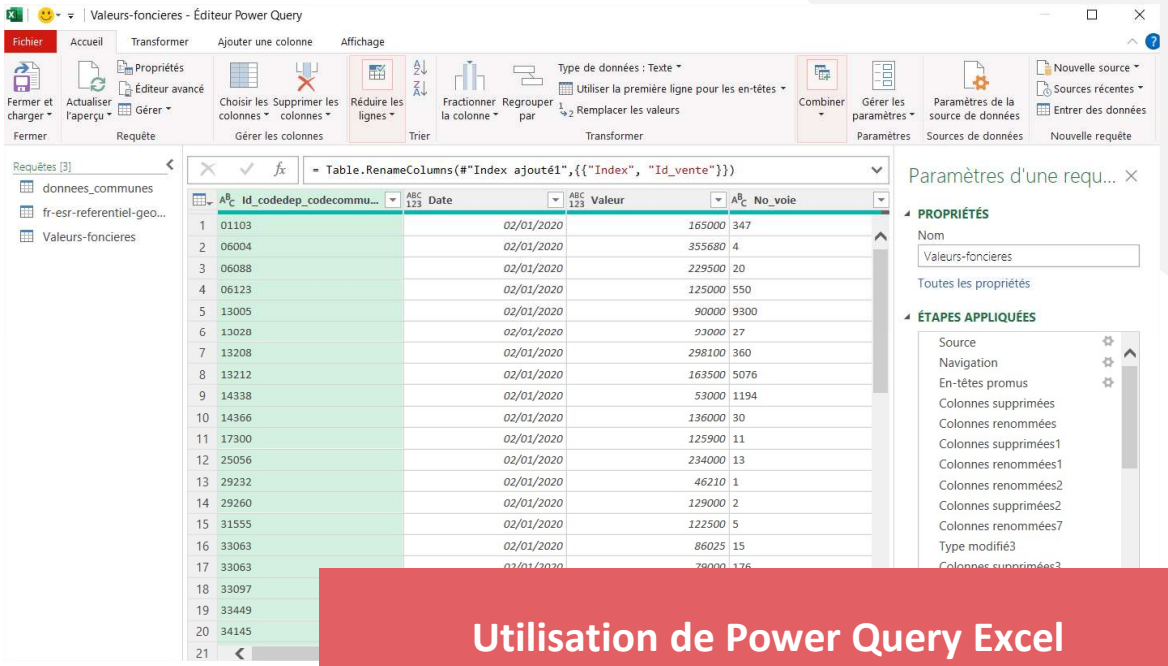
- **Création des tables** Bien, Vente, Commune, Departement et Region et **définition des Clés Primaires**
- **Ajout des attributs** (colonnes) et configuration des données (type, taille) comme prévu par le dictionnaire des données
- **Établissement des relations entre les tables** via l'utilisation **des clés étrangères** pour relier les tables
- **Application des trois premières formes normales** (1NF, 2NF, 3NF) pour garantir que chaque table est correctement organisée.



Mise en forme des données

Adaptation des fichiers source à la nouvelle structure des données

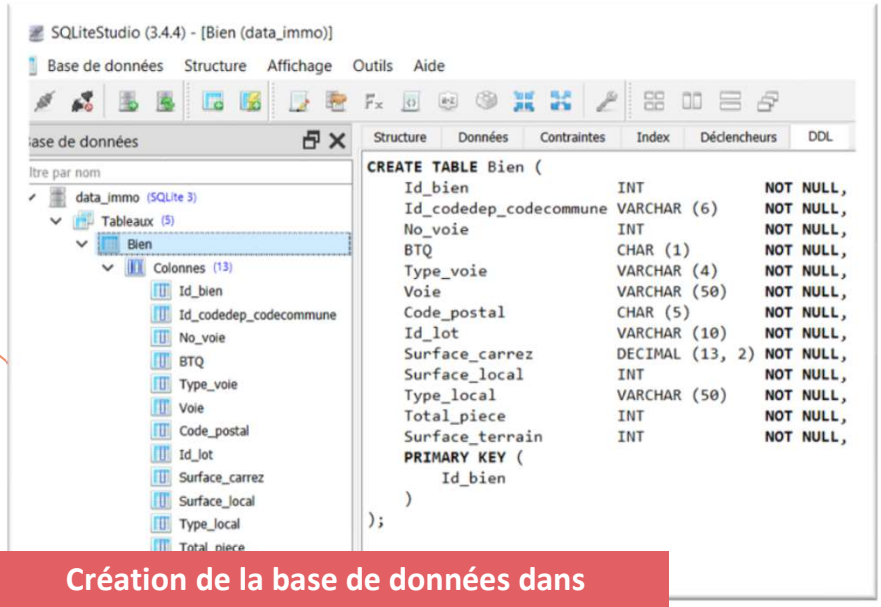
- Suppression des colonnes non pertinentes, renommage et réorganisation des colonnes
- Regroupement de colonnes issus de plusieurs fichiers source
- Modification de la nature des données
- Création d'index et d'identifiant unique



Utilisation de Power Query Excel

Permet la mise en forme des données et la mise en place d'actions réutilisables pour les mises à jour.

La base de données avec les tables créées et les données chargées



Création de la base de données dans le logiciel SGBD SQLite 3.

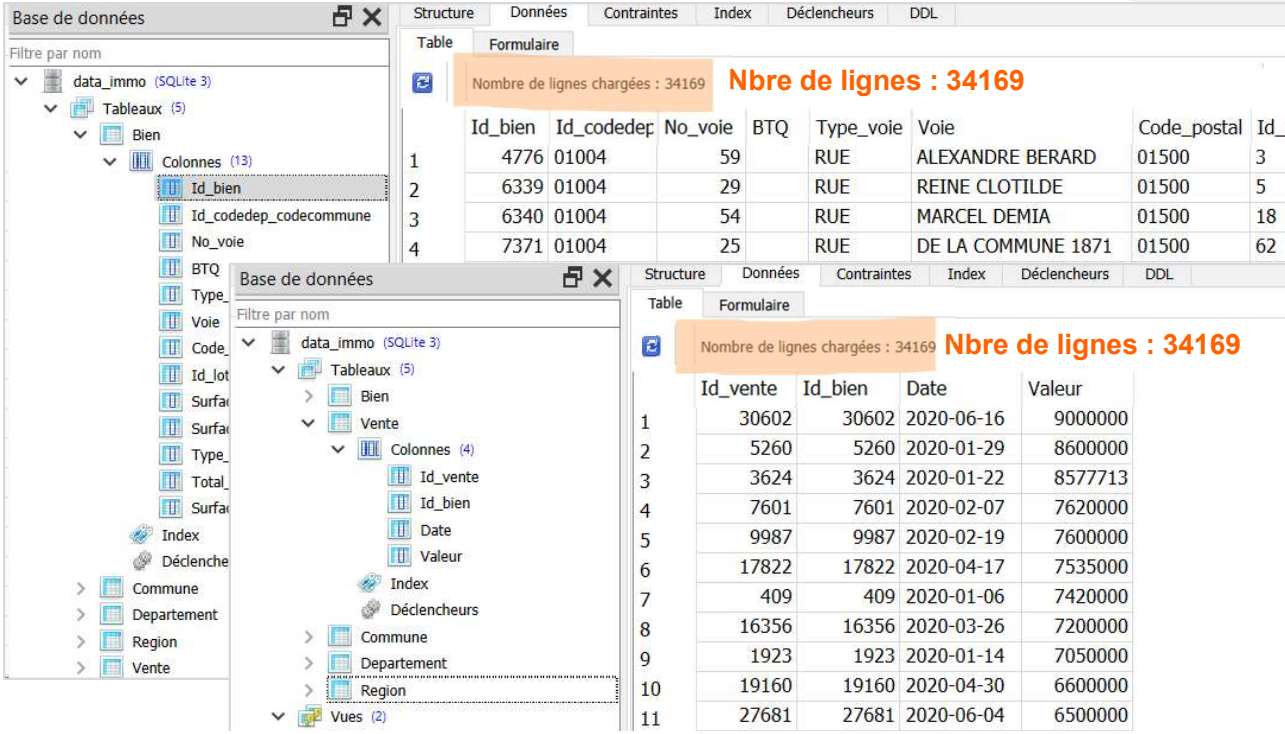
Export SQL de Power Architect Adaptation du code aux spécificités du logiciel

Intégration des données dans les tables

depuis les fichiers csv préparés, via la fonction d'import du logiciel

Maintient de la cohérence entre l'ordre des colonnes du fichier d'origine et de la liste des attributs des tables pour faciliter l'importation

Contrôles cohérence pour vérifier les données importées

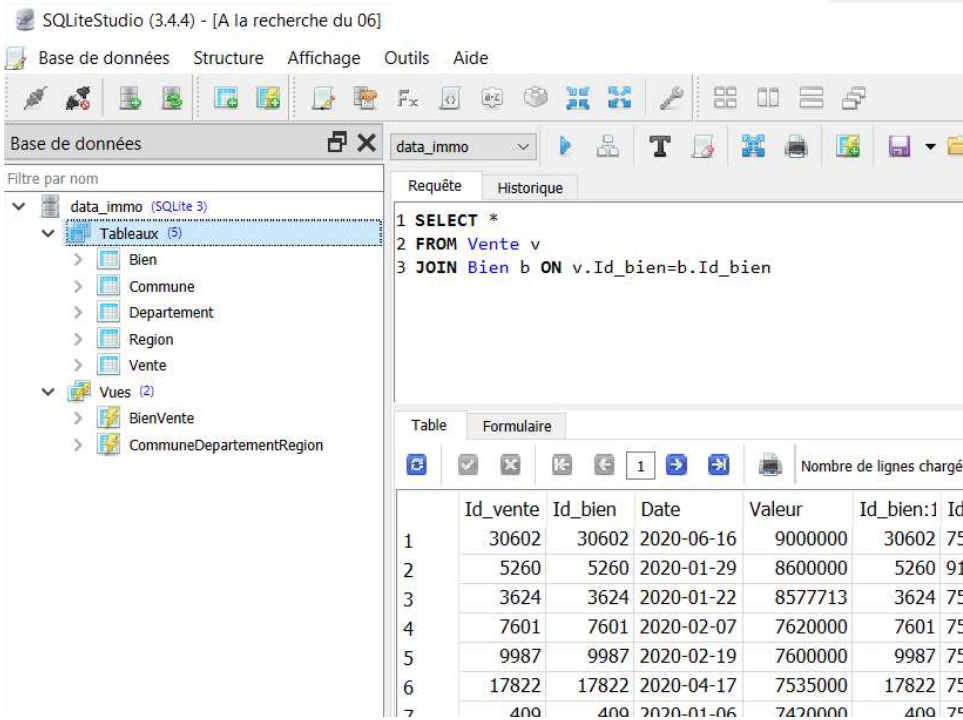


Les requêtes et screenshot qui permettent de démontrer le bon chargement des données

Utilisation de l'éditeur SQL de SQLite3 pour rédiger des requêtes et extraire des données et analyse de notre base de données.

Grâce à cela on va pouvoir interroger notre base sur des questions métier pertinentes (voir requêtent suivantes).

- Utilisation de **SELECT** afin d'extraire nos données ciblées
- Utilisation **WHERE** pour établir des conditions
- Utilisation de **JOIN** afin d'extraire des données liées entre nos deux tables **via la clé étrangère**
- Utilisation des fonctions d'agrégation **COUNT** et **AVG** pour les analyses
- Utilisation des requêtes **ORDER By** et **LIMIT** pour trier et limiter les résultats
- Utilisation de **GROUP By** et **HAVING** pour grouper des résultats



Requêtes SQL et résultats

Requête 1

Nombre total d’appartements vendus au 1er semestre 2020

```
SELECT COUNT(*) AS "Appartements Vendus au 1er Semestre 2020"
FROM Vente v
JOIN Bien b ON v.Id_bien = b.Id_bien
WHERE b.Type_local = "Appartement"
AND v.Date BETWEEN "2020-01-01" AND "2020-06-30";
```

	Appartements Vendus au 1er Semestre 2020
1	31378

Requête 2

Le nombre de ventes d'appartement par région pour le 1er semestre 2020

```
SELECT
        r.Reg_nom AS Region,
        COUNT(*) AS Nombre_Ventes_Appartements
FROM Vente v
JOIN Bien b ON v.Id_bien = b.Id_bien
JOIN Commune c ON b.Id_codedep_codecommune =
    c.Id_codedep_codecommune
JOIN Departement d ON c.Dep_code = d.Dep_code
JOIN Region r ON d.Reg_code = r.Reg_code
WHERE b.Type_local = "Appartement"
AND v.Date BETWEEN "2020-01-01" AND "2020-06-30"
GROUP BY r.Reg_nom
ORDER BY Nombre_Ventes_Appartements DESC;
```

	Region	Nombre_Ventes_Appartements
1	Ile-de-France	13995
2	Provence-Alpes-Cô...	3649
3	Auvergne-Rhône-...	3253
4	Nouvelle-Aquitaine	1932
5	Occitanie	1640
6	Pays de la Loire	1357
7	Hauts-de-France	1254
8	Grand Est	984
9	Bretagne	983
10	Normandie	862
11	Centre-Val de Loire	696
12	Bourgogne-Franch...	376
13	Corse	223
14	Martinique	94
15	La Réunion	44
16	Guyane	34
17	Guadeloupe	2

Création de Vues

Une **vue** est une **table virtuelle** qui stocke une requête.
Elle permet de simplifier l'accès à des données provenant de plusieurs tables en permettant de limiter les jointures.

1

VUE CommuneDepartementRegion

Table virtuelle liant les informations entre la commune le département et la région.

```
CREATE VIEW CommuneDepartementRegion AS
SELECT *
FROM Commune c
JOIN Departement d ON c.Dep_code = d.Dep_code
JOIN Region r ON d.Reg_code = r.Reg_code;
```

2

VUE BienVente

Table virtuelle liant les informations entre les ventes et le bien associé.

```
CREATE VIEW BienVente AS
SELECT *
FROM Bien b
JOIN Vente v ON b.Id_bien = v.Id_bien;
```

Requête 3

Proportion des ventes d'appartements par le nombre de pièces

```
SELECT
        b.Total_piece AS Nombre_de_pieces,
        COUNT(*) AS Nombre_Appartements_Vendus,
        ROUND(
        COUNT(*) * 100.0 /
        (
           SELECT
           COUNT(*)
           FROM Vente v
           JOIN Bien b ON v.Id_bien = b.Id_bien
           WHERE b.Type_local = "Appartement"
        )
        ,2) AS Proportion
FROM Vente v
JOIN Bien b ON v.Id_bien = b.Id_bien
WHERE b.Type_local = "Appartement"
GROUP BY b.Total_piece;
```

	Nombre_de_pieces	Nombre_Appartements_Ve	Proportion
1	0	30	0.1
2	1	6739	21.48
3	2	9783	31.18
4	3	8966	28.57
5	4	4460	14.21
6	5	1114	3.55
7	6	204	0.65
8	7	54	0.17
9	8	17	0.05
10	9	8	0.03
11	10	2	0.01
12	11	1	0

Requête 4

Liste des 10 départements où le prix du mètre carré est le plus élevé

```
SELECT cdr.Dep_nom AS Departement,
  cdr.Dep_code AS Code_département,
  ROUND(AVG(v.Valeur / b.Surface_carrez), 2) AS Prix_moyen_au_m²
FROM Vente v
JOIN Bien b ON v.Id_bien = b.Id_bien
JOIN CommuneDepartementRegion cdr
ON b.Id_codedep_codecommune = cdr.Id_codedep_codecommune
GROUP BY cdr.Dep_nom
ORDER BY Prix_moyen_au_m² DESC
LIMIT 10;
```

	Departement	Code_département	Prix_moyen_au_m²
1	Paris	75	12091.16
2	Hauts-de-Seine	92	7300.21
3	Val-de-Marne	94	5430.25
4	Haute-Savoie	74	4780.97
5	Alpes-Maritimes	06	4758.62
6	Seine-Saint-Denis	93	4393.48
7	Yvelines	78	4275.13
8	Rhône	69	4099.83
9	Corse-du-Sud	2A	4079.06
10	Gironde	33	3806.8

Requête 5

Prix moyen du mètre carré d'une maison en région Île-de-France

```
SELECT
        cdr.Reg_nom AS Region,
        ROUND(AVG(bv.Valeur / bv.Surface_carrez), 2) AS Prix_m2_moyen
FROM BienVente bv
JOIN CommuneDepartementRegion cdr
    ON b.Id_codedep_codecommune = cdr.Id_codedep_codecommune
WHERE cdr.Reg_nom = "Ile-de-France"
AND b.Type_local = "Maison"
GROUP BY cdr.Reg_nom;
```

	Region	Prix_m2_moyen
1	Ile-de-France	3764.39

Requête 6

Liste des 10 appartements les plus chers avec la région et le nombre de mètres carrés

```
SELECT ROUND(bv.Surface_carrez,0) As Surface,
        bv.Valeur AS Valeur,bv.No_voie,bv.BTQ,bv.Type_voie,bv.Voie,bv.Id_lot AS Lot,bv.Code_postal,
        cdr.Com_nom_maj_court AS Ville,
        cdr.Reg_nom AS Region
FROM BienVente bv
JOIN CommuneDepartementRegion cdr ON bv.Id_codedep_codecommune = cdr.Id_codedep_codecommune
WHERE bv.Type_local = "Appartement"
ORDER BY bv.Valeur DESC
LIMIT 10;
```

	Surface	Valeur	No_voie	BTQ	Type_voi	Voie	Lot	Code_pos	Ville	Region
1	9	9000000	6		BD	SUCHET	853	75016	PARIS 16	Ile-de-France
2	64	8600000	16		CHE	DE LA CAVIGNON	322	91100	CORBEIL ESSONNES	Ile-de-France
3	20	8577713	104		RUE	DU BAC	100	75007	PARIS 7	Ile-de-France
4	42	7620000	33		RUE	LEMERCIER	15	75017	PARIS 17	Ile-de-France
5	253	7600000	72		RUE	D ASSAS	104	75006	PARIS 6	Ile-de-France
6	139	7535000	8		RUE	SAINT HYACINTHE	79	75001	PARIS 1	Ile-de-France
7	360	7420000	36		AV	GEORGES MANDEL	5	75016	PARIS 16	Ile-de-France
8	595	7200000	23		BD	DE BEAUSEJOUR	11	75016	PARIS 16	Ile-de-France
9	122	7050000	26		RUE	CAMBON	2	75001	PARIS 1	Ile-de-France
10	79	6600000	108		RUE	SAINT HONORE	181	75001	PARIS 1	Ile-de-France

Requête 7

Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020

```
SELECT "Premier Trimestre" AS Trimestre,
            COUNT(*) AS Nombre_Ventes
FROM Vente
WHERE Date BETWEEN '2020-01-01' AND '2020-03-31'
 UNION ALL
 SELECT "Second Trimestre" AS Trimestre,
            COUNT(*) AS Nombre_Ventes
FROM Vente
WHERE Date BETWEEN '2020-04-01' AND '2020-06-30'
 UNION ALL
 SELECT "Taux évolution (en%)" AS Trimestre,
    ROUND((
          (SELECT COUNT(*) FROM Vente WHERE Date BETWEEN '2020-04-01' AND '2020-06-30')
          - (SELECT COUNT(*) FROM Vente WHERE Date BETWEEN '2020-01-01' AND '2020-03-31')
          ) * 100.0
          / (SELECT COUNT(*) FROM Vente WHERE Date BETWEEN '2020-01-01' AND '2020-03-31')
    ,2)  AS Nombre_Ventes;
```

	Trimestre	Nombre_Ventes
1	Premier Trimestre	16776
2	Second Trimestre	17393
3	Taux évolution (en%)	3.68

$$\text{Taux d'évolution (en \%)} = \frac{\text{Ventes du 2e trimestre} - \text{Ventes du 1er trimestre}}{\text{Ventes du 1er trimestre}} \times 100$$

Requête 8

Le classement des régions par rapport au prix au mètre carré des appartement de plus de 4 pièces

```
SELECT
        cdr.Reg_nom As Région,
        ROUND(AVG(v.Valeur / b.Surface_carrez),2) AS Prix_moyen_au_m²
FROM Vente v
JOIN Bien b ON v.Id_bien = b.Id_bien
JOIN CommuneDepartementRegion cdr
    ON b.Id_codedep_codecommune = cdr.Id_codedep_codecommune
WHERE b.Type_local = "Appartement"
AND b.Total_piece > 4
GROUP BY cdr.Reg_nom
ORDER BY Prix_moyen_au_m² DESC;
```

	Région	Prix_moyen_au_m²
1	Ile-de-France	8819.07
2	La Réunion	3659
3	Provence-Alpes-Côte d'Azur	3616.21
4	Corse	3117.2
5	Auvergne-Rhône-Alpes	2903.36
6	Nouvelle-Aquitaine	2476.01
7	Bretagne	2426.72
8	Pays de la Loire	2328.71
9	Hauts-de-France	2199.43
10	Occitanie	2106.77
11	Normandie	2025.88
12	Grand Est	1560.43
13	Centre-Val de Loire	1459.53
14	Bourgogne-Franche-Comté	1260.42
15	Martinique	574

Requête 9

Liste des communes ayant eu au moins 50 ventes au 1er trimestre

```
SELECT
   c.Com_nom_maj_court AS Commune,
   COUNT(v.Id_vente) AS Nombre_de_ventes
FROM Vente v
JOIN Bien b
ON v.Id_bien = b.Id_bien
JOIN Commune c
ON b.Id_codedep_codecommune =
c.Id_codedep_codecommune
WHERE  v.Date
BETWEEN "2020-01-01" AND "2020-03-31"
GROUP BY c.Com_nom_maj_court
HAVING COUNT(v.Id_vente) >= 50
ORDER BY Nombre_de_ventes DESC;
```

	Nom de la Commune	Nombre_de_vente:
1	PARIS 17	228
2	PARIS 15	215
3	PARIS 18	209
4	NICE	173
5	PARIS 11	169
6	PARIS 16	165
7	BORDEAUX	157
8	PARIS 14	146
9	PARIS 20	127
10	NANTES	119
11	PARIS 19	116
12	PARIS 12	110
13	PARIS 10	109
14	PARIS 9	106
15	GRENOBLE	106
16	BOULOGNE BILLANCOURT	99
17	PARIS 13	94
18	PARIS 7	87
19	PARIS 6	86
20	MARSEILLE 8	81
21	ASNIERES SUR SEINE	81
22	COURBEVOIE	80
23	PARIS 5	79
24	PARIS 3	79
25	TOULOUSE	78
26	ANTIBES	77
27	MARSEILLE 4	72
28	MARSEILLE 1	71
29	VINCENNES	68
30	RUEIL MALMAISON	68
31	LILLE	67
32	MARSEILLE 9	66
33	MONTREUIL	65
34	ANGERS	64
35	NIMES	63
36	SETE	62
37	PARIS 8	62
38	LA CIOTAT	62
39	RENNES	61
40	PARIS 2	61
41	PARIS 4	60
42	TOULON	59
43	LEVALLOIS PERRET	59
44	ST MAUR DES FOSSES	56
45	VERSAILLES	54
46	AJACCIO	54
47	PUTEAUX	53
48	ISSY LES MOULINEAUX	50

Requête 10

Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces

```
SELECT "Appartement 2 pièces" AS Nombre_de_pièces,
        ROUND(AVG(bv.Valeur/bv.Surface_carrez),2) AS Prix_au_m²
FROM BienVente bv
WHERE bv.Type_local="Appartement" AND bv.Total_piece=2
UNION ALL
SELECT "Appartement 3 pièces" AS Nombre_de_pièces,
        ROUND(AVG(bv.Valeur/bv.Surface_carrez),2) AS Prix_au_m²
FROM BienVente bv
WHERE bv.Type_local="Appartement«  AND bv.Total_piece=3
UNION ALL
SELECT "Différence (en %)" AS Nombre_de_pièces,
        ROUND((
          (SELECT AVG(bv.Valeur/bv.Surface_carrez) FROM BienVente bv  WHERE bv.Type_local="Appartement" AND bv.Total_piece=3 )
          -  (SELECT AVG(bv.Valeur/bv.Surface_carrez) FROM BienVente bv WHERE bv.Type_local="Appartement" AND bv.Total_piece=2 )
          ) * 100.0
          /  (SELECT AVG(bv.Valeur/bv.Surface_carrez) FROM BienVente bv  WHERE bv.Type_local="Appartement" AND bv.Total_piece=2 )
    ,2)  AS Prix_au_m²;
```

	Nombre_de_pièces	Prix_au_m²
1	Appartement 2 pièces	4969.96
2	Appartement 3 pièces	4335.13
3	Différence (en %)	-12.77

$$\text{Différence en \%} = \frac{\text{Prix moyen au m² des appartements de 3 pièces} - \text{Prix moyen au m² des appartements de 2 pièces}}{\text{Prix moyen au m² des appartements de 2 pièces}} \times 100$$

Requête 11

Les moyennes de valeurs foncières pour le top 3 des communes des départements 06, 13, 33, 59 et 69

```
SELECT cdr.Dep_Code AS Département,
        cdr.Com_nom_maj_court AS Commune,
ROUND(AVG(v.Valeur), 2) AS Moyenne_des_valeurs_foncières
FROM Vente v
JOIN Bien b ON v.Id_bien = b.Id_bien
JOIN CommuneDepartementRegion cdr
    ON b.Id_codedep_codecommune = cdr.Id_codedep_codecommune
WHERE Dep_code = "06" OR Dep_code = "13" OR Dep_code = "33" OR
Dep_code = "59" OR Dep_code = "69"
GROUP By cdr.Dep_code, cdr.Com_nom_maj_court
HAVING AVG(v.Valeur) >= ( SELECT AVG(v2.Valeur)
        FROM Vente v2
        JOIN Bien b2 ON v2.Id_bien = b2.Id_bien
        JOIN CommuneDepartementRegion cdr2 ON
b2.Id_codedep_codecommune = cdr2.Id_codedep_codecommune
        WHERE cdr2.Dep_code = cdr.Dep_code
        GROUP BY cdr2.Com_nom_maj_court
        ORDER BY AVG(v2.Valeur) DESC
        LIMIT 1 OFFSET 2 )
ORDER BY Dep_code ASC, Moyenne_des_valeurs_foncières DESC;
```

	Département	Commune	Moyenne_valeurs_foncières
1	06	ST JEAN CAP FERRAT	968750
2	06	EZE	655000
3	06	MOUANS SARTOUX	476898
4	13	GIGNAC LA NERTHE	330000
5	13	ST SAVOURNIN	314425
6	13	CASSIS	313416.88
7	33	LEGE CAP FERRET	549500.64
8	33	VAYRES	335000
9	33	ARCACHON	307435.93
10	59	BERSEE	433202
11	59	CYSOING	408550
12	59	HALLUIN	322250
13	69	VILLE SUR JARNIOUX	485300
14	69	LYON 2	455217.26
15	69	LYON 6	426968.25

Requête 12

Les 20 communes avec le plus de transactions pour 1000 habitants pour les communes qui dépassent les 10 000 habitants

```
SELECT
   c.Com_nom_maj_court AS Commune,
   COUNT(bv.Id_vente) AS Nombre_de_ventes,
   ROUND
        (COUNT(bv.Id_vente) * 1000.0) / c.Population_totale
        , 2) AS Transactions_pour_1000_habitants
FROM BienVente bv
JOIN Commune c
   ON bv.Id_codedep_codecommune =
        c.Id_codedep_codecommune
WHERE c.Population_totale >= 10000
GROUP BY c.Com_nom_maj_court, c.Population_totale
ORDER BY Transactions_pour_1000_habitants DESC
LIMIT 20;
```

	Commune	Nombre_de_ventes	Transactions_pour_1000_habitants
1	PARIS 2	127	5.84
2	PARIS 1	79	4.92
3	PARIS 3	161	4.69
4	ARCACHON	55	4.62
5	LA BAULE	77	4.58
6	PARIS 4	120	4.08
7	ROQUEBRUNE CAP MARTIN	52	3.99
8	PARIS 8	139	3.83
9	SANARY SUR MER	60	3.5
10	LA LONDE LES MAURES	37	3.43
11	PARIS 9	208	3.43
12	PARIS 6	139	3.38
13	ST CYR SUR MER	38	3.24
14	CHANTILLY	35	3.13
15	PORNICHET	35	3.06
16	ST MANDE	69	3.06
17	PARIS 10	264	3.04
18	MENTON	91	2.94
19	ST HILAIRE DE RIEZ	33	2.87
20	VINCENNES	141	2.81

Merci !