



Création et utilisation de la base de données

Daniela MENGUI



Laplace Immo

Contexte du projet

Projet DATAImmO

- Création et modification de BDD dans un réseau d'agence immobilières
- Analyse du marché
- SQL



La stratégie de sauvegarde et la conformité RGPD

- Site open data des Demandes de valeurs foncières (DVF)
- INSEE
- data.gouv

Les données initiales

- Demandes de valeurs foncières (DVF)
- Résultats des recensements de la population
- Référentiel géographique français

L'extrait du dictionnaire des données

DICTIONNAIRE DES DONNÉES - Biens

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	EGLE DE GESTIO
Id_bien	ID dans la base de données	INTEGER	10	Elémentaire	CLE PRIMAIRE, NOT N
Code_dep_code_commun	Concaténation du code département et code commune pour avoir une clé unique	VARCHAF	10	Concaténer	CIE ETRANGERE, NOT
No_voie	Numéro des rues	VARCHAR	10	Elémentaire	
Type_de_voie	Plusieurs valeurs (rue, avenue, chemin, etc.)	VARCHAR	10	Elémentaire	
Voie	Nom de la rue	VARCHAR	100	Elémentaire	NOT NULL
Total_piece	Nombre de pièces	INTEGER	5	Elémentaire	
Surface_carrez	Superficie du logement d'après la loi du 18/12/1996	FLOAT	20	Elémentaire	NOT NULL
Surface_local	Superficie du logement qui correspond à la somme de la surface réelle du local et de la surface des dépendances	INTEGER	10	Elémentaire	NOT NULL
Type_local	Type du local entre appartement et maison	VARCHAF	50	Elémentaire	NOT NULL

L'extrait du dictionnaire des données

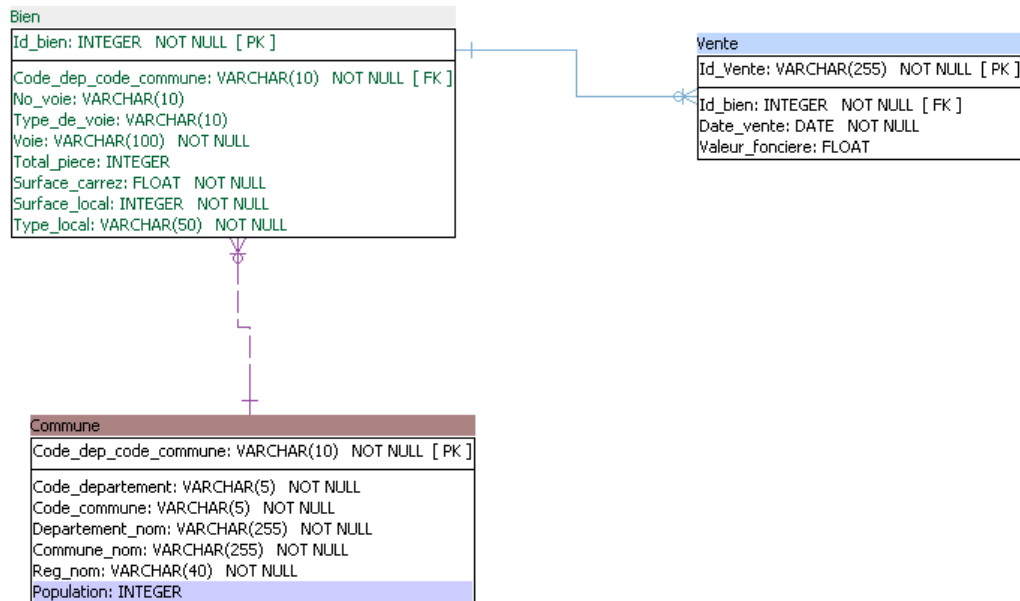
DICTIONNAIRE DES DONNÉES - Commune

CODE	SIGNIFICATION	TYPE	ONGUEU	NATURE	REGLE DE GESTION
Code_dep_code_commune	Concaténation du code département et code commune pour avoir une clé unique	VARCHAR	10	Concaténer	CLE PRIMAIRE, NOT NULL
Code_département	Code département	VARCHAR	5	Elémentaire	NOT NULL
Code_commune	Code commune	VARCHAR	5	Elémentaire	NOT NULL
Département_nom	Libellé du département	VARCHAR	255	Elémentaire	NOT NULL
Commune_nom	Libellé de la commune	VARCHAR	255	Elémentaire	NOT NULL
Reg_nom	Libellé région	VARCHAR	40	Elémentaire	NOT NULL
Population	Nombre de résidents	INTEGER	50	Elémentaire	

DICTIONNAIRE DES DONNÉES - Vente

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION
Id_Vente	ID dans la base de données	VARCHAR	255	Elémentaire	CLE PRIMAIRE, NOT NULL
Date_vente	Date de la vente	DATE	40	Elémentaire	NOT NULL
Valeur_fonciere	Prix net vendeur du bien	FLOAT	40	Elémentaire	
Id_bien	Numéro Identifiant du bien	INTEGER	40	Elémentaire	CLE ETRANGERE NOT NULL

Le schéma relationnel normalisé



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

Structure de la Base de Données Parcourir les données Éditer les Pragma Exécuter le SQL

Table : Bien Filtrer dans n'importe quelle colonne

	Id_bien	Code_dep_code_commune	No_voie	Type_de_voie	Voie	Total_piece	Surface_carrez	Surface_local	Type_local
	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre
1	1 1103	347	RUE	DU CHATEAU		3	48,22	48	Appartement
2	2 6004	4	BD	EDOUARD BAUDOIN		1	39,11	40	Appartement
3	3 6088	20	RUE	MARCEAU		3	80,25	82	Appartement
4	4 6123	550	RTE	DES VESPIANS RN7		1	27,51	27	Appartement
5	5 13005	9300	RES	LES ARPEGES BD DES ABA		2	47,33	47	Appartement
6	6 13028	27	RUE	DU GRAND MADIER		1	25,31	24	Appartement
7	7 13208	360	AV	DU PRADO		3	70,84	70	Appartement
8	8 13212	5076	PARC	DESSUARD		3	67,19	66	Appartement
9	9 14338	1194	RUE	DE NORMANDIE		1	18,89	19	Appartement
10	10 14366	30	ALL	DES NOISETIERS		4	105,37	99	Maison
11	11 17300	11	RUE	ROUGET DE L ISLE		2	31,99	34	Appartement
12	12 25056	13	RUE	BERTHE MORISOT		5	96,21	100	Maison
13	13 29232	1	RUE	DU POHER		1	30,86	31	Appartement
14	14 29260	2	RUE	DES JARDINS		3	66,21	67	Appartement
15	15 31555	5	AV	DU COMMANDANT TAILLANDIER		2	45,58	46	Appartement
16	16 33063	15	RUE	PAUL DENUCE		1	23,2	27	Appartement
17	17 33063	176	RUE	SAINT CATHERINE		1	13,1	12	Appartement
18	18 33097	822	NULL	LA BAYNASSE SUD		2	29,0	29	Appartement

1 - 18 de 34169 Aller à : 1

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

Structure de la Base de Données							
Parcourir les données							
Éditer les Pragma							
Exécuter le SQL							
Table : Commune							
	Code_dep_code_commune	Code_departement	Code_commune	Departement_nom	Commune_nom	Reg_nom	Population
	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre
1	1001	1	1	Ain	L'Abergement-Clémenciat	Auvergne-Rhône-Alpes	779
2	1002	1	2	Ain	L'Abergement-de-Varey	Auvergne-Rhône-Alpes	256
3	1004	1	4	Ain	Ambérieu-en-Bugey	Auvergne-Rhône-Alpes	14134
4	1005	1	5	Ain	Ambérieu-en-Dombes	Auvergne-Rhône-Alpes	1751
5	1006	1	6	Ain	Ambléon	Auvergne-Rhône-Alpes	112
6	1007	1	7	Ain	Ambronay	Auvergne-Rhône-Alpes	2800
7	1008	1	8	Ain	Ambutrix	Auvergne-Rhône-Alpes	762
8	1009	1	9	Ain	Andert-et-Condou	Auvergne-Rhône-Alpes	326
9	1010	1	10	Ain	Anglefort	Auvergne-Rhône-Alpes	1105
10	1011	1	11	Ain	Apremont	Auvergne-Rhône-Alpes	368
11	1012	1	12	Ain	Aranc	Auvergne-Rhône-Alpes	329
12	1013	1	13	Ain	Arandas	Auvergne-Rhône-Alpes	143
13	1014	1	14	Ain	Arbent	Auvergne-Rhône-Alpes	3349
14	1015	1	15	Ain	Arbois en Bugey	Auvergne-Rhône-Alpes	650
15	1016	1	16	Ain	Arbigny	Auvergne-Rhône-Alpes	460
16	1017	1	17	Ain	Argis	Auvergne-Rhône-Alpes	457
17	1019	1	19	Ain	Armix	Auvergne-Rhône-Alpes	27
18	1021	1	21	Ain	Ars-sur-Formans	Auvergne-Rhône-Alpes	1457

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

Structure de la Base de Données Parcourir les données Éditer les Pragma's Exécuter le SQL

Table : Vente

	Id_Vente	Date_vente	Valeur_fonciera	Id_bien
	Filtre	Filtre	Filtre	Filtre
1	2020-00...	2020-01-02	165000.0	1
2	2020-00...	2020-01-02	355680.0	2
3	2020-00...	2020-01-02	229500.0	3
4	2020-00...	2020-01-02	125000.0	4
5	2020-00...	2020-01-02	90000.0	5
6	2020-00...	2020-01-02	93000.0	6
7	2020-00...	2020-01-02	298100.0	7
8	2020-00...	2020-01-02	163500.0	8
9	2020-00...	2020-01-02	53000.0	9
10	2020-00...	2020-01-02	136000.0	10
11	2020-00...	2020-01-02	125900.0	11
12	2020-00...	2020-01-02	234000.0	12
13	2020-00...	2020-01-02	46210.0	13
14	2020-00...	2020-01-02	129000.0	14
15	2020-00...	2020-01-02	122500.0	15
16	2020-00...	2020-01-02	86025.0	16
17	2020-00...	2020-01-02	79000.0	17
18	2020-00...	2020-01-02	81795.0	18

1 - 18 de 34169

Le bon chargement des données

```
1  Select count(*)  
2  from Bien;  
3
```

count(*)	
1	34169

```
1  Select count(*)  
2  from Commune;
```

count(*)	
1	34991

```
1  Select count(*)  
2  from Vente;
```

count(*)	
1	34169



Requêtes SQL et résultats

Requête 1

Nombre total d'appartements vendus au 1er semestre 2020

```
SELECT COUNT(*) AS Nombre_total_appartements_vendus  
from Vente  
JOIN Bien on Bien.Id_bien = Vente.Id_bien  
WHERE Type_local = 'Appartement'  
AND Date_vente BETWEEN '2020-01-01' AND '2020-06-30';
```

Nombre_total_appartements_vendus
31378

Requête 2

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

```
SELECT Commune.Reg_nom as Noms_regions, COUNT(*) AS Nombre_ventes_appartements_par_region
from Vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
WHERE Type_local = 'Appartement'
AND Date_vente BETWEEN '2020-01-01' AND '2020-06-30'
group by 1
ORDER by Nombre_ventes_appartements_par_region desc;
```

	Noms_regions	Nombre_ventes_appartements_par_region
1	Île-de-France	13995
2	Provence-Alpes-Côte d'Azur	3649
3	Auvergne-Rhône-Alpes	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-Franche-Comté	376
13	Corse	223
14	Martinique	94
15	La Réunion	44
16	Guyane	34
17	Guadeloupe	2

Requête 3

Proportion des ventes d'appartements par le nombre de pièces

```
select bien.Total_piece, count(id_vente) as nombre_vente, round((cast(count(id_vente) as float)
/(SUM(COUNT(Id_Vente)) OVER())*100),2) AS part_ventes_appart
from vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
WHERE Type_local = 'Appartement'
AND Date_vente BETWEEN '2020-01-01' AND '2020-06-30'
group by Total_piece;
```

	Total_piece	nombre_vente	part_ventes_appart
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	0.0

Requête 4

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

```
SELECT Departement_nom, round(avg(Valeur_fonciere / Surface_carrez),2)
as Prix_au_metre_carre
from VENTE
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
group by 1
order by 2 desc
limit 10;
```

	Departement_nom	Prix_au_metre_carre
1	Paris	12091.58
2	Hauts-de-Seine	7300.66
3	Val-de-Marne	5430.68
4	Haute-Savoie	4781.39
5	Alpes-Maritimes	4759.05
6	Seine-Saint-Denis	4393.91
7	Yvelines	4275.57
8	Rhône	4100.28
9	Corse-du-Sud	4079.5
10	Gironde	3807.23

Requête 5

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

```
SELECT Reg_nom, round(avg(Valeur_fonciere / Surface_carrez), 2)
as Prix_moyen_metre_carre_maison
from Vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
where Type_local = 'Maison'
and Reg_nom = 'Île-de-France';
```

Reg_nom	Prix_moyen_metre_carre_maison
Île-de-France	3764.84

Requête 6

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

```
SELECT bien.Id_bien, commune.Reg_nom, bien.Surface_carrez, vente.Valeur_fonciere
from vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
where Type_local = 'Appartement'
Order by CAST(Valeur_fonciere as float) DESC
LIMIT 10;
```

	Id_bien	Reg_nom	Surface_carrez	Valeur_fonciere
1	30591	Île-de-France	9,1	9000000.0
2	5261	Île-de-France	64.0	8600000.0
3	3625	Île-de-France	20,55	8577713.0
4	7602	Île-de-France	42,77	7620000.0
5	9988	Île-de-France	253,3	7600000.0
5	17821	Île-de-France	139,9	7535000.0
7	410	Île-de-France	360,95	7420000.0
8	16355	Île-de-France	595.0	7200000.0
9	1924	Île-de-France	122,56	7050000.0
10	19158	Île-de-France	79,38	6600000.0

Requête 7

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

```
With Q1(T1) as (select count(Id_Vente) from Vente where Date_vente BETWEEN '2020-01-01' AND '2020-03-31'),  
Q2(T2) as (SELECT count(Id_Vente) from Vente where Date_vente BETWEEN '2020-04-01' AND '2020-06-30')  
SELECT ROUND((cast (T2 - T1 as float)/T1 *100),2) as Taux_évoluation_ventes_entre_T1_et_T2  
FROM Q2, Q1;
```

Taux_évoluation_ventes_entre_T1_et_T2

3.68

Requête 8

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

```
SELECT Reg_nom, round(avg(Valeur_fonciere / Surface_carrez),2) as Prix_au_metre_carre
from venteE
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
where Total_piece > 4
and Type_local = 'Appartement'
group by 1
order by 2 desc;
```

	Reg_nom	Prix_au_metre_carre
1	Île-de-France	8819.54
2	La Réunion	3659.83
3	Provence-Alpes-Côte d'Azur	3616.71
4	Corse	3117.88
5	Auvergne-Rhône-Alpes	2903.86
6	Nouvelle-Aquitaine	2476.5
7	Bretagne	2427.12
8	Pays de la Loire	2329.2
9	Hauts-de-France	2199.92
10	Occitanie	2107.24
11	Normandie	2026.31
12	Grand Est	1560.91
13	Centre-Val de Loire	1459.97
14	Bourgogne-Franche-Comté	1260.73
15	Martinique	574.77

Requête 9

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

```
SELECT commune.Commune_nom, COUNT(*) AS Plus_de_50_ventes
from Vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
where Date_vente BETWEEN '2020-01-01' AND '2020-03-31'
group by 1
HAVING COUNT (*) >= 50
order by 2 desc;
```

	Commune_nom	Plus_de_50_ventes
1	Paris 17e Arrondissement	228
2	Paris 15e Arrondissement	215
3	Paris 18e Arrondissement	209
4	Nice	173
5	Paris 11e Arrondissement	169
6	Paris 16e Arrondissement	165
7	Bordeaux	157
8	Paris 14e Arrondissement	146
9	Paris 20e Arrondissement	127
10	Nantes	119
11	Paris 19e Arrondissement	116
12	Paris 12e Arrondissement	110
13	Paris 10e Arrondissement	109
14	Paris 9e Arrondissement	106
15	Grenoble	106
40	Paris 2e Arrondissement	61
41	Paris 4e Arrondissement	60
42	Toulon	59
43	Levallois-Perret	59
44	Saint-Maur-des-Fossés	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

```
WITH P(A) as (SELECT ROUND (avg(Valeur_fonciere / Surface_carrez))
from Vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
where Total_piece = 2
and Type_local = 'Appartement'),
S(B) as (SELECT ROUND (avg(Valeur_fonciere / Surface_carrez))
from Vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
where Total_piece = 3
and Type_local = 'Appartement')
SELECT ROUND((cast(B - A as float)/A *100),2)
as Difference_Prix_m2_entre_appt_2pieces_et_3pieces
FROM P,S
;
```

Difference_Prix_m2_entre_appt_2pieces_et_3pieces
-12.76

Requête 11

Moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69

```
WITH T as (  
  SELECT Commune.Code_departement, Commune.Commune_nom, round(((avg(cast(Valeur_fonciere as float)))) as Valeur_fonciere_moyenne  
  from Vente  
  JOIN Bien on Bien.Id_bien = Vente.Id_bien  
  JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune  
  where Commune.Code_departement in (6,13,33,59,69)  
  GROUP BY 1,2)  
select Code_departement, Commune_nom, Valeur_fonciere_moyenne  
from (select Code_departement,Commune_nom, Valeur_fonciere_moyenne, rank () over (PARTITION BY Code_departement order by Valeur_fonciere_moyenne Desc) as Rank FROM T)  
where Rank <=3;
```

	Code_departement	Commune_nom	Valeur_fonciere_moyenne
1	6	Saint-Jean-Cap-Ferrat	968750.0
2	6	Eze	655000.0
3	6	Mouans-Sartoux	476898.0
4	13	Gignac-la-Nerthe	330000.0
5	13	Saint-Savournin	314425.0
6	13	Cassis	313417.0
7	33	Lège-Cap-Ferret	549501.0
8	33	Vayres	335000.0
9	33	Arcachon	307436.0
10	59	Bersée	433202.0
11	59	Cysoing	408550.0
12	59	Halluin	322250.0
13	69	Ville-sur-Jarniou	485300.0
14	69	Lyon 2e Arrondissement	455217.0
15	69	Lyon 6e Arrondissement	426968.0

Requête 12

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

```
select Commune_nom, round((cast (Nombre_transactions as float) / Population*1000),4)
as top20_nombre_transactions_pour_1000_habitants
from (select Commune_nom, Population, count(Id_Vente) AS Nombre_transactions
from vente
JOIN Bien on Bien.Id_bien = Vente.Id_bien
JOIN Commune on Commune.Code_dep_code_commune = Bien.Code_dep_code_commune
where Population > 10000
GROUP BY Commune_nom
ORDER BY 3 desc)
GROUP BY 1
ORDER BY 2 DESC
LIMIT 20;
```

	Commune_nom	top20_nombre_transactions_pour_1000_habitants
1	Paris 2e Arrondissement	5.881
2	Paris 1er Arrondissement	4.963
3	La Baule-Escoublac	4.737
4	Paris 3e Arrondissement	4.732
5	Arcachon	4.729
6	Paris 4e Arrondissement	4.119
7	Roquebrune-Cap-Martin	4.05
8	Paris 8e Arrondissement	3.898
9	Sanary-sur-Mer	3.553
10	La Londe-les-Maures	3.477
11	Paris 9e Arrondissement	3.465
12	Paris 6e Arrondissement	3.449
13	Saint-Cyr-sur-Mer	3.309
14	Chantilly	3.196
15	Pornichet	3.142
16	Saint-Mandé	3.084
17	Paris 10e Arrondissement	3.053
18	Menton	2.981
19	Saint-Hilaire-de-Riez	2.921
20	Vincennes	2.832



Merci !