



Création et utilisation de la base de données

Bastien RABANE



Laplace Immo

Contexte du projet

Laplace Immo, réseau national d'agences immobilières, travaille sur un projet stratégique interne appelé "DATAImmo".

Sous la direction de Clara Daucourt, CTO de l'entreprise, notre mission consiste à modifier la base de données afin de collecter et analyser les transactions immobilières et foncières en France.

L'objectif est d'utiliser ces données pour prévoir les prix de vente des biens immobiliers et aider les agences régionales à mieux accompagner leurs clients.



Les données initiales

Les données initiales proviennent de trois sources distinctes :

1. Demandes de valeurs foncières (DVF):

Site web Open Data : <https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/>

Contient des informations sur les transactions immobilières, telles que le prix de vente, la surface du bien, la localisation, etc.

2. INSEE:

Site web de l'INSEE : <https://www.insee.fr/fr/statistiques>

Contient des données sur la population, telles que le nombre d'habitants, la répartition par âge, le niveau de vie, etc.

3. Référentiel géographique français:

Site web data.gouv.fr : <https://www.data.gouv.fr/fr/datasets/referentiel-geographique-francais-communes-unites-urbaines-aires-urbaines-departements-academies-regions/>

Contient des données sur les régions, les communes, les unités urbaines, etc., avec le référentiel géographique français.

L'extrait du dictionnaire des données

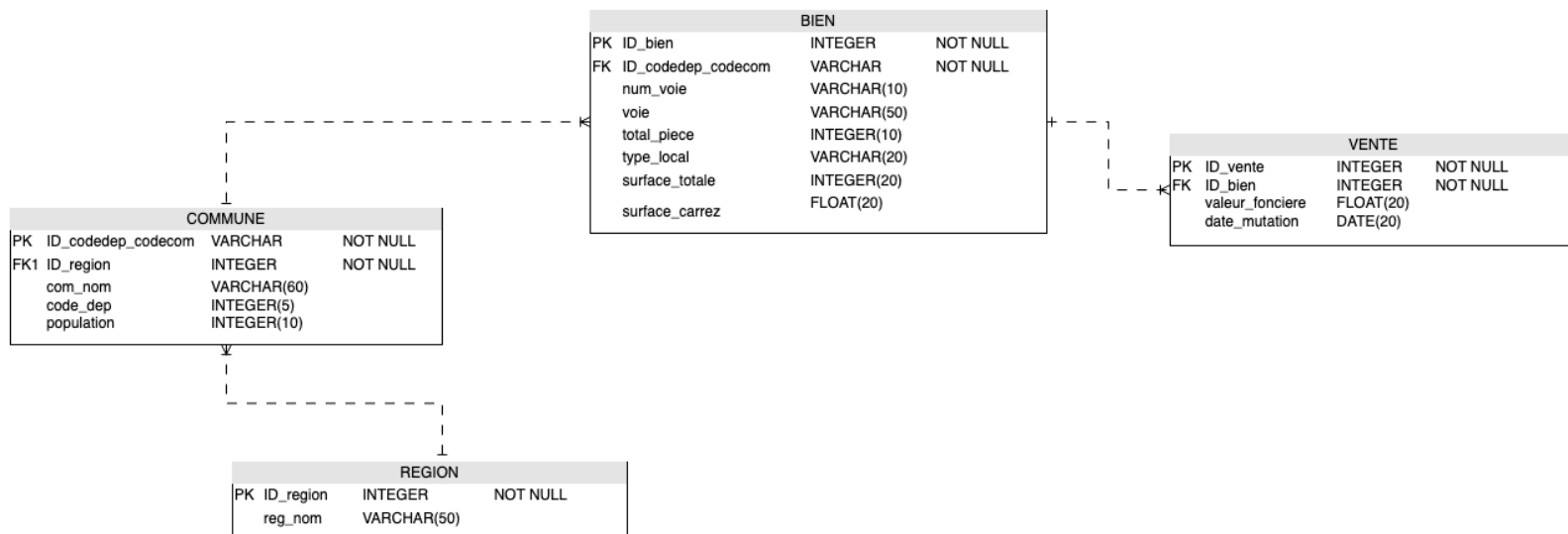
DICTIONNAIRE DES DONNÉES - Bien

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
ID_bien	Identifiant unique du bien	Int		Clé primaire	Not null	
num_voie	Numéro de la voie	Varchar	10	Concatenation		Numero de voie + B/T/Q
voie	Nom de la voie	Varchar	50	Concatenation		Type voie + voie
total_piece	Nombre total de pieces	Int	10	Elementaire		
type_local	Le type de local	Varchar	20	Elementaire		
surface_totale	Surface reelle du bien immobilier	Float	50	Elementaire		
surface_carrez	Superficie habitable loi carrez du bien immobilier	Float	50	Elementaire		

DICTIONNAIRE DES DONNÉES - Commune

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
ID_codedep_codecom	Identifiant unique du code departement et du code commune	Varchar		Clé primaire	Not Null	code departement + code commune
com_nom	Nom de la commune	Varchar	30	Elementaire		
code_dep	Numero de code du departement	Int	5	Elementaire		
population	Population totale de la commune	Int	10	Elementaire		

Le schéma relationnel normalisé



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

PROJET_5_OC_DATAIMMO

Tableaux (4)

BIEN

Colonnes (8)

- ID_bien
- ID_codedep_codecom
- num_voie
- voie
- total_pie
- type_local
- surface_totale
- surface_carrez

Index

Déclencheurs

COMMUNE

Colonnes (5)

- ID_codedep_codecom
- ID_region
- com_nom
- code_dep
- population

Index

Déclencheurs

REGION

Colonnes (2)

- ID_region
- reg_nom

Index

Déclencheurs

VENTE

Colonnes (4)

- ID_vente
- ID_bien
- date_mutation
- valeur_foncier

Index

Déclencheurs

	ID_bien	ID_code	num_voi	voie	total_pie	type_local	surface_	surface_
1	1	01103	347	RUE DU CHATEAU	3	Appartement	48	48,22
2	2	06004	4	BD EDOUARD BAUDOUIN	1	Appartement	40	39,11
3	3	06088	20B	RUE MARCEAU	3	Appartement	82	80,25
4	4	06123	550	RTE DES VESPINS RN7	1	Appartement	27	27,51
5	5	13005	9300	RES LES ARPEGES BD DES ABA	2	Appartement	47	47,33




	ID_codedep_codecom	ID_region	com_nom	code_dep	population
1	1001	17	L'Abergement-Clémenciat	1	798
2	1002	17	L'Abergement-de-Varey	1	257
3	1003	17	Amareins	1	
4	1004	17	Ambérieu-en-Bugey	1	14514
5	1005	17	Ambérieux-en-Dombes	1	1776

	ID_region	reg_nom
1	1	Collectivités d'outre-mer
2	2	Guadeloupe
3	3	Martinique
4	4	Guyane
5	5	La Réunion

	ID_vente	ID_bien	date_mutation	valeur_foncier
1	1	17294	2020-01-02	165000
2	2	22116	2020-01-02	355680
3	3	10446	2020-01-02	229500
4	4	24323	2020-01-02	125000
5	5	32824	2020-01-02	90000

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

```
1 SELECT 'BIEN' AS Table_name, COUNT(*) AS Row_count FROM BIEN
2 UNION ALL
3 SELECT 'COMMUNE' AS Table_name, COUNT(*) AS Row_count FROM COMMUNE
4 UNION ALL
5 SELECT 'REGION' AS Table_name, COUNT(*) AS Row_count FROM REGION
6 UNION ALL
7 SELECT 'VENTE' AS Table_name, COUNT(*) AS Row_count FROM VENTE;
8 |
```

     1    Nombre de lignes chargées : 4

	Table_name	Row_count
1	BIEN	34169
2	COMMUNE	38916
3	REGION	19
4	VENTE	34169











Requêtes SQL et résultats

Requête 1

Nombre total d'appartements vendus au 1er semestre 2020

```
1 SELECT COUNT(DISTINCT B.ID_bien) AS Nombre_appartements_vendus
2 FROM VENTE V
3 JOIN BIEN B ON V.ID_bien = B.ID_bien
4 WHERE V.date_mutation >= '2020-01-01' AND V.date_mutation <= '2020-06-30'
5 AND B.type_local = 'Appartement';
6 |
```

     1   		Nombre de lignes chargées : 1
Appartements vendus 1er semestre 2020		
1	31378	

Requête 2

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

```
1 SELECT
2     r.reg_nom AS "Région",
3     COUNT(v.ID_vente) AS "Nombre de ventes"
4 FROM VENTE v
5 INNER JOIN BIEN b ON b.ID_bien = v.ID_bien
6 INNER JOIN COMMUNE c ON c.ID_codedep_codecom = b.ID_codedep_codecom
7 INNER JOIN REGION r ON r.ID_region = c.ID_region
8 WHERE b.type_local = 'Appartement'
9 AND v.date_mutation BETWEEN '2020-01-01' AND '2020-06-30'
10 GROUP BY r.reg_nom
11 ORDER BY "Nombre de ventes" DESC;
```

Nombre de lignes chargées : 17		
	Région	Nombre de ventes
1	Ile-de-France	13995
2	Auvergne-Rhône-Alpes	2977
3	Provence-Alpes-Côte d'Azur	2153
4	Nouvelle-Aquitaine	1932
5	Occitanie	1628
6	Pays de la Loire	1357
7	Hauts-de-France	1183
8	Bretagne	983
9	Grand Est	933
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

```
1 SELECT
2     b.total_piece AS "Nombre de pièces",
3     ROUND((COUNT(v.ID_vente) * 100.0) / (SELECT COUNT(*) FROM BIEN WHERE type_local = 'Appartement'), 2) AS "Proportion des ventes"
4 FROM BIEN b
5 JOIN VENTE v ON v.ID_bien = b.ID_bien
6 WHERE b.type_local = 'Appartement'
7 GROUP BY b.total_piece
8 ORDER BY b.total_piece ASC;
```

Nombre de lignes chargées : 12			
	Nombre de pièces	Proportion des ventes	
1	0	0.1	
2	1	21.48	
3	2	31.18	
4	3	28.57	
5	4	14.21	
6	5	3.55	
7	6	0.65	
8	7	0.17	
9	8	0.05	
10	9	0.03	
11	10	0.01	
12	11	0	

Requête 4

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






```
1 SELECT C.code_dep AS 'Departement',
2 ROUND(AVG(V.valeur_fonciere / B.surface_carrez), 2) AS 'Prix_au_metre_carre'
3 FROM VENTE V
4 JOIN BIEN B ON B.ID_bien = V.ID_bien
5 JOIN COMMUNE C ON C.ID_codedep_codecom = B.ID_codedep_codecom
6 WHERE B.surface_carrez != 0
7 GROUP BY C.code_dep
8 ORDER BY Prix_au_metre_carre DESC
9 LIMIT 10;
10 |
```

Nombre de lignes chargées : 10			
	Departement	Prix_au_metre_carre	
1	75	12091.17	
2	92	7300.22	
3	94	5430.26	
4	74	4780.99	
5	6	4758.63	
6	93	4393.49	
7	78	4275.15	
8	69	4099.84	
9	2A	4079.07	
10	33	3806.82	

Requête 5

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

```
1 SELECT round(avg(V.valeur_fonciere/B.surface_carrez), 2) as 'Prix_moyen_au_metre_carre',  
2 R.reg_nom  
3 FROM VENTE V  
4 JOIN BIEN B on B.ID_bien=V.ID_bien  
5 JOIN COMMUNE C on C.ID_codedep_codecom=B.ID_codedep_codecom  
6 JOIN REGION R on R.ID_region=C.ID_region  
7 WHERE B.type_local= 'Maison'  
8 AND R.reg_nom= 'Ile-de-France'
```

					1				Nombre de lignes chargées : 1
Prix_moyen_au_metre_carre		reg_nom							
1	3764.42	Ile-de-France							

Requête 6

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

```
1 SELECT
2     B.ID_bien,
3     V.valeur_fonciere AS Prix_de_vente,
4     R.reg_nom AS Region,
5     B.surface_carrez
6 FROM BIEN B
7 JOIN VENTE V ON V.ID_bien = B.ID_bien
8 JOIN COMMUNE C ON B.ID_codedep_codecom = C.ID_codedep_codecom
9 JOIN REGION R ON C.ID_region = R.ID_region
10 WHERE B.type_local = 'Appartement'
11 ORDER BY V.valeur_fonciere DESC
12 LIMIT 10;
```

Nombre de lignes chargées : 10					
	ID_bien	Prix_de_vente	Region	surface_carrez	
1	30603	9000000	Ile-de-France	9,1	
2	5261	8600000	Ile-de-France	64	
3	3625	8577713	Ile-de-France	20,55	
4	7602	7620000	Ile-de-France	42,77	
5	9988	7600000	Ile-de-France	253,3	
6	17823	7535000	Ile-de-France	139,9	
7	410	7420000	Ile-de-France	360,95	
8	16357	7200000	Ile-de-France	595	
9	1924	7050000	Ile-de-France	122,56	
10	19161	6600000	Ile-de-France	79,38	

Requête 7

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

```
1 WITH ventes_premier_trimestre AS (  
2     SELECT COUNT(*) AS ventes_premier_trimestre  
3     FROM VENTE  
4     WHERE date_mutation >= '2020-01-01' AND date_mutation <= '2020-03-31'  
5 ),  
6 ventes_deuxieme_trimestre AS (  
7     SELECT COUNT(*) AS ventes_deuxieme_trimestre  
8     FROM VENTE  
9     WHERE date_mutation >= '2020-04-01' AND date_mutation <= '2020-06-30'  
10 )  
11 SELECT  
12     round(((vd.ventes_deuxieme_trimestre - vp.ventes_premier_trimestre) * 100.0 / vp.ventes_premier_trimestre), 2) AS evolution_des_ventes  
13 FROM  
14     ventes_premier_trimestre vp,  
15     ventes_deuxieme_trimestre vd;  
16
```

Nombre de lignes chargées : 1	
evolution_des_ventes	3.68

Requête 8

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

```
1 SELECT
2 R.reg_nom AS Region,
3 ROUND(AVG(V.valeur_fonciere / B.surface_carrez), 2) AS Prix_m2_moyen
4 FROM BIEN B
5 JOIN VENTE V ON V.ID_bien = B.ID_bien
6 JOIN COMMUNE C ON B.ID_codedep_codecom = C.ID_codedep_codecom
7 JOIN REGION R ON C.ID_region = R.ID_region
8 WHERE B.type_local = 'Appartement'
9 AND B.total_piece > 4
10 AND B.surface_carrez != 0
11 GROUP BY R.reg_nom
12 ORDER BY Prix_m2_moyen DESC;
13
```

Nombre de lignes chargées : 15		
	Region	Prix_m2_moyen
1	Ile-de-France	8819.09
2	La Réunion	3659
3	Provence-Alpes-Côte d'Azur	3616.22
4	Corse	3117.2
5	Auvergne-Rhône-Alpes	2903.38
6	Nouvelle-Aquitaine	2476.02
7	Bretagne	2426.73
8	Pays de la Loire	2328.76
9	Hauts-de-France	2199.47
10	Occitanie	2106.79
11	Normandie	2025.89
12	Grand Est	1560.46
13	Centre-Val de Loire	1459.57
14	Bourgogne-Franche-Comté	1260.52
15	Martinique	574

Requête 9

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

```
1 WITH Ventes_par_Commune AS (  
2     SELECT  
3         B.ID_codedep_codecom,  
4         COUNT(*) AS Nombre_ventes  
5     FROM VENTE V  
6     JOIN BIEN B ON V.ID_bien = B.ID_bien  
7     WHERE V.date_mutation >= '2020-01-01' AND V.date_mutation <= '2020-03-31'  
8     GROUP BY B.ID_codedep_codecom  
9     HAVING COUNT(*) >= 50  
10 )  
11 SELECT  
12     C.com_nom AS Commune,  
13     VPC.Nombre_ventes AS Nombre_ventes  
14 FROM  
15     Ventes_par_Commune VPC  
16 JOIN  
17     COMMUNE C ON VPC.ID_codedep_codecom = C.ID_codedep_codecom;  
18
```

Nombre de lignes chargées : 48		
	Commune	Nombre_ventes
1	Antibes	77
2	Nice	173
3	La Ciotat	62
4	Marseille 1er	71
5	Marseille 4e	72
6	Marseille 8e	81
7	Marseille 9e	66
8	Ajaccio	54
9	Nîmes	63
10	Toulouse	78
11	Bordeaux	157
12	Sète	62
13	Rennes	61
14	Grenoble	106
15	Nantes	119
16	Angers	64
17	Lille	67
18	Paris 2e	61
19	Paris 3e	79
20	Paris 4e	60
21	Paris 5e	79
22	Paris 6e	86
23	Paris 7e	87
24	Paris 8e	62
25	Paris 9e	106
26	Paris 10e	109
27	Paris 11e	169
28	Paris 12e	110
29	Paris 13e	94
30	Paris 14e	146

Requête 10

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

```
1 WITH prix_m2_2_pieces AS (  
2     SELECT round(AVG(V.valeur_fonciere / B.surface_carrez), 2) AS prix_m2_2_pieces  
3     FROM BIEN B  
4     JOIN VENTE V ON V.ID_bien = B.ID_bien  
5     WHERE B.type_local = 'Appartement'  
6           AND B.total_piece = 2  
7 ),  
8 prix_m2_3_pieces AS (  
9     SELECT round(AVG(V.valeur_fonciere / B.surface_carrez), 2) AS prix_m2_3_pieces  
10    FROM BIEN B  
11    JOIN VENTE V ON V.ID_bien = B.ID_bien  
12    WHERE B.type_local = 'Appartement'  
13           AND B.total_piece = 3  
14 )  
15 SELECT prix_m2_2_pieces, prix_m2_3_pieces, round(((pm3.prix_m2_3_pieces - pm2.prix_m2_2_pieces) * 100.0 / pm2.prix_m2_2_pieces), 2) AS difference_pourcentage  
16 FROM prix_m2_2_pieces pm2, prix_m2_3_pieces pm3;  
17 |
```

Nombre de lignes chargées : 1				
	prix_m2_2_pieces	prix_m2_3_pieces	difference_pourcentage	
1	4969.97	4335.15	-12.77	

Requête 11

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

```
1 WITH top_communes AS (  
2 SELECT  
3 C.com_nom AS Commune,  
4 AVG(V.valeur_fonciere) AS Moyenne_VF,  
5 RANK() OVER(PARTITION BY C.code_dep ORDER BY AVG(V.valeur_fonciere) DESC) AS classement,  
6 C.code_dep  
7 FROM VENTE V  
8 JOIN BIEN B ON V.ID_bien = B.ID_bien  
9 JOIN COMMUNE C ON B.ID_codedep_codecom = C.ID_codedep_codecom  
10 WHERE C.code_dep IN (6, 13, 33, 59, 69)  
11 GROUP BY C.code_dep, C.com_nom  
12 )  
13 SELECT  
14 Commune,  
15 Moyenne_VF,  
16 code_dep,  
17 classement  
18 FROM top_communes  
19 WHERE classement <= 3  
20 ORDER BY code_dep, classement;  
21
```

Nombre de lignes chargées : 15				
	Commune	Moyenne_VF	code_dep	classement
1	Gignac-la-Nerthe	330000	13	1
2	Saint-Savournin	314425	13	2
3	Cassis	313416.875	13	3
4	Lège-Cap-Ferret	549500.6363636364	33	1
5	Vayres	335000	33	2
6	Arcachon	307435.92727272725	33	3
7	Bersée	433202	59	1
8	Cysoing	408550	59	2
9	Halluin	322250	59	3
10	Saint-Jean-Cap-Ferrat	968750	6	1
11	Eze	655000	6	2
12	Mouans-Sartoux	476898.1	6	3
13	Ville-sur-Jarnioux	485300	69	1
14	Lyon 2e	455217.26666666666	69	2
15	Lyon 6e	426968.25	69	3

Requête 12

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

```
1 WITH transactions_par_commune AS (  
2 SELECT  
3 C.com_nom AS Commune,  
4 COUNT(*) AS Nombre_transactions,  
5 C.population,  
6 (C.population / 1000.0) AS Habitant_1000  
7 FROM VENTE V  
8 JOIN BIEN B ON V.ID_bien = B.ID_bien  
9 JOIN COMMUNE C ON B.ID_codedep_codecom = C.ID_codedep_codecom  
10 GROUP BY C.com_nom, C.population  
11 HAVING C.population > 10000  
12 )  
13 SELECT  
14 Commune,  
15 round((Nombre_transactions / Habitant_1000), 2) AS Transactions_pour_1000_habitants,  
16 population,  
17 Nombre_transactions  
18 FROM transactions_par_commune  
19 ORDER BY Transactions_pour_1000_habitants DESC  
20 LIMIT 20;  
21 |
```

	Commune	Transactions_pour_1000_habitants	population	Nombre_transactions
1	Paris 2e	5.84	21735	127
2	Paris 1er	4.92	16055	79
3	Paris 3e	4.69	34306	161
4	Arcachon	4.62	11898	55
5	La Baule-Escoublac	4.58	16797	77
6	Paris 4e	4.08	29390	120
7	Roquebrune-Cap-Martin	3.99	13041	52
8	Paris 8e	3.83	36250	139
9	Sanary-sur-Mer	3.5	17160	60
10	La Londe-les-Maures	3.43	10776	37
11	Paris 9e	3.43	60563	208
12	Paris 6e	3.38	41171	139
13	Saint-Cyr-sur-Mer	3.24	11725	38
14	Chantilly	3.13	11178	35
15	Pornichet	3.06	11440	35
16	Saint-Mandé	3.06	22576	69
17	Paris 10e	3.04	86863	264
18	Menton	2.94	30981	91
19	Saint-Hilaire-de-Riez	2.87	11501	33
20	Vincennes	2.81	50230	141



Merci !