



# Création et utilisation de la base de données

---

Caroline G



Laplace Immo



# Contexte du projet

## - DATA IMMO -



**Mieux prévoir le prix de vente des biens immobiliers :**

- ✓ Modification de la base de données
- ✓ Améliorer l'existant
- ✓ Collecter les transactions immobilières et foncières en France
- ✓ Analyse de marché – aider les agences régionales
- ✓ Respect du RGPD



# Les données initiales

---



- Des données extraites du site open data des Demandes de valeurs foncières (DVF)
- Des données de l'INSEE avec les résultats des recensements de la population
- Des données de data.gouv sur les régions, avec le référentiel géographique français, communes, unités urbaines, aires urbaines, départements, académies, régions



# Extrait du dictionnaire des données



## DICTIONNAIRE DES DONNÉES - Données Valeurs foncières

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_bien	ID dans la base de données. -Clé primaire	INTEGER	NC	Elémentaire	Not NULL	Auto-incrémenté
id_codedep_codecommune	Code département - Code commune - Clé étrangère Table Commune	VARCHAR	NC	CO	Not NULL	Concat (Code departement + code Commune)
no_voie	Numéro des rues	INTEGER	NC	Elémentaire		
B_T_Q	Indice de répétition	VARCHAR	1	Elémentaire		
type_voie	Plusieurs valeurs (rue, avenue, chemin, etc.)	VARCHAR	NC	Elémentaire		
voie	Nom de la rue	VARCHAR	NC	Elémentaire		
date	Date de la vente du bien	DATE		Elémentaire	aaaa/mm/jj - Not NULL	
valeur	Valeur foncière à la vente (€) (18 Valeurs nulles supprimées)	INTEGER	NC	Elémentaire		Arrondir si chiffres à virgule
com_nom	Libellé de la commune	VARCHAR	NC	Elémentaire		
id_code_dep	Code département	VARCHAR	3	Elémentaire		
code_commune	Code commune	VARCHAR	3	Elémentaire		
type_local	Appartement/Maison	VARCHAR	NC	Elémentaire		
nb_pieces	Nombre de pièces dan le logement	INTEGER	NC	Elémentaire		
surface	Surface du local	FLOAT	NC	Elémentaire		

Description des données : <https://s3-eu-west-1.amazonaws.com/static.oc-static.com/prod/courses/files/parcours-data-analyst/DAN-P3-notice-descriptive-du-fichier-dvf.pdf>



Region
id_code_reg: VARCHAR(2) NOT NULL [ PK ]
reg_nom: VARCHAR NOT NULL

Departement
id_code_dep: VARCHAR(3) NOT NULL [ PK ]
id_code_reg: VARCHAR(2) NOT NULL [ FK ]
dep_nom: VARCHAR NOT NULL

Commune
id_codedep_codecommune: VARCHAR(5) NOT NULL [ PK ]
id_code_dep: VARCHAR(3) NOT NULL [ FK ]
code_commune: VARCHAR NOT NULL
com_nom: VARCHAR NOT NULL
population_totale: INTEGER

Bien
id_bien: INTEGER NOT NULL [ PK ]
id_codedep_codecommune: VARCHAR(5) NOT NULL [ FK ]
no_voie: INTEGER NOT NULL
B_T_Q: VARCHAR(1) NOT NULL
type_voie: VARCHAR NOT NULL
voie: VARCHAR NOT NULL
surface: FLOAT NOT NULL
type_local: VARCHAR NOT NULL
nb_pieces: INTEGER NOT NULL

Vente
id_vente: INTEGER NOT NULL [ PK ]
id_bien: INTEGER NOT NULL [ FK ]
date: DATE NOT NULL
valeur: INTEGER NOT NULL

# Le schéma relationnel normalisé



Databases

Filter by name

caro\_gou\_P5\_Etp3\_DATA\_IMMO (SQLite 3)

Tables (5)

Bien

Columns (9)

id\_bien

id\_codedep\_codecommune

no\_voie

B\_T\_Q

type\_voie

voie

surface

type\_local

nb\_pieces

Indexes

Triggers

Commune

Columns (5)

id\_codedep\_codecommune

id\_code\_dep

code\_commune

com\_nom

population\_totale

Indexes

Triggers

Departement

Columns (3)

id\_code\_dep

id\_code\_reg

dep\_nom

Indexes

Triggers

Region

Columns (2)

id\_code\_reg

reg\_nom

Indexes

Triggers

Vente

Columns (4)

id\_vente

id\_bien

date

valeur

Indexes

Triggers

Views

Vente (caro\_gou\_P5\_Etp3\_DATA\_IMMO)

StructureDataConstraintsIndexesTriggersDDL

Grid viewForm view

	id_vente	id_bien	date	valeur
1	0	0	2020/01/13	85000
2	1	1	2020/01/31	54000
3	2	2	2020/01/03	43000
4	3	3	2020/01/10	265000
5	4	4	2020/01/14	137500
6	5	5	2020/01/16	70000
7	6	6	2020/01/18	147000
8	7	7	2020/01/20	154470
9	8	8	2020/01/23	109500
10	9	9	2020/01/24	94000
11	10	10	2020/01/27	57800

Bien (caro\_gou\_P5\_Etp3\_DATA\_IMMO)

StructureDataConstraintsIndexesTriggersDDL

Grid viewForm view

	id_bien	id_codedep	no_voie	B_T_Q	type_voi	voie
1	0	971127	58		LOT	MONT VERNON I
2	1	971127	9094		IMP	SIMONE
3	2	93001	55		RUE	DU MOUTIER
4	3	94001	12		RUE	SAINT GEORGES
5	4	94001	11		AV	AUGUSTE DURU
6	5	94001	16		RUE	DU GAL DE GAULLE
7	6	93001	16		RUE	HENRI BARBUSSE
8	7	93001	154		RUE	LEOPOLD RECHOSSIER
9	8	93001	114		BD	EDOUARD VAILLANT
10	9	93001	12		RUE	LEOPOLD RECHOSSIER
11	10	47001	44		RUE	RICHARD COEUR DE LI

Commune (caro\_gou\_P5\_Etp3\_DATA\_IMMO)

StructureDataConstraintsIndexesTriggersDDL

Grid viewForm view

	id_codedep_cod	id_code_dep	code_commune	com_nom
1	01001	01	001	L'ABERGEMENT-CLEN
2	01002	01	002	L'ABERGEMENT-DE-V
3	01004	01	004	AMBERIEU-EN-BUGE
4	01005	01	005	AMBERIEUX-EN-DOM
5	01006	01	006	AMBLEON
6	01007	01	007	AMBERIEU-EN-BUGE
7	01008	01	008	AMBERIEU-EN-BUGE
8	01009	01	009	ANDERT-ET-CONDON
9	01010	01	010	ANGLEFORT
10	01011	01	011	APREMONT
11	01012	01	012	ARANC

Region (caro\_gou\_P5\_Etp3\_DATA\_IMMO)

StructureDataConstraintsIndexesTriggersDDL

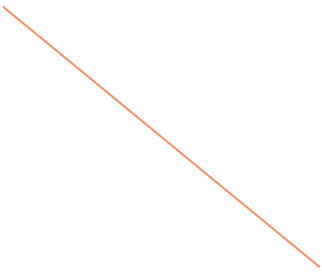
Grid viewForm view

	id_code_reg	reg_nom
1	00	Collectivités d'outre-mer
2	01	Guadeloupe
3	02	Martinique
4	03	Guyane
5	04	La Réunion
6	05	Mayotte
7	06	Alsace
8	24	Centre-Val de Loire
9	27	Bourgogne-Franche-Comté
10	28	Normandie
11	32	Hauts-de-France

La base de données chargée



# Requêtes SQL & Résultats





## Nombre total d'appartements vendus au 1er semestre 2020



Query

```
1 SELECT COUNT( * ) AS Nb_Appart
2 FROM Bien b
3 INNER JOIN Vente v
4 ON v.id_bien = b.id_bien
5 WHERE (1 = 1 |
6         AND type_local = "Appartement"
7         AND date >= "2020/01/01"
8         AND date <= "2020/06/30");
```

Grid view

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Total rows

	Nb_Appart
1	31362





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

```
1 SELECT reg_nom AS Region,  
2         COUNT(id_vente) AS Nb_ventes  
3 FROM Region r  
4 JOIN Departement d  
5 ON d.id_code_reg = r.id_code_reg  
6 JOIN Commune c  
7 ON c.id_code_dep = d.id_code_dep  
8 JOIN Bien b  
9 ON b.id_codedep_codecommune = c.id_codedep_codecommune  
10 JOIN Vente v  
11 ON v.id_bien = b.id_bien  
12 WHERE (1 = 1 AND  
13         type_local = "Appartement" AND  
14         date >= "2020/01/01" AND  
15         date <= "2020/06/30")  
16 GROUP BY Region  
17 ORDER BY Nb_ventes DESC;
```

	Region	Nb_ventes
1	Ile-de-France	13989
2	Provence-Alpes-Côte d'Azur	3645
3	Auvergne-Rhône-Alpes	3253
4	Nouvelle-Aquitaine	1931
5	Occitanie	1640
6	Pays de la Loire	1357
7	Hauts-de-France	1252
8	Grand Est	984
9	Bretagne	983
10	Normandie	861
11	Centre-Val de Loire	695
12	Bourgogne-Franche-Comté	376
13	Corse	222
14	Martinique	94
15	La Réunion	44
16	Guyane	34
17	Guadeloupe	2



```
1 SELECT nb_pieces,  
2         ROUND( (COUNT(id_vente)*100/31362.00),2) AS Repartition_appart_vendus,  
3         |"%"  
4 FROM Bien b  
5 JOIN Vente v  
6 ON v.id_bien = b.id_bien  
7 WHERE (type_local = "Appartement" )  
8 GROUP BY nb_pieces  
9 ORDER BY Repartition_appart_vendus DESC;
```

Proportion des ventes  
d'appartements par le nombre  
de pièces

---

	nb_pieces	Repartition_appart_vendus	%
1	2	31.16	%
2	3	28.59	%
3	1	21.48	%
4	4	14.21	%
5	5	3.55	%
6	6	0.65	%
7	7	0.17	%
8	0	0.1	%
9	8	0.05	%
10	9	0.03	%
11	10	0.01	%
12	11	0	%



```
SELECT id_code_dep nO_departement,  
       dep_nom nom_departement,  
       ROUND(AVG(ROUND(valeur/surface))) prix_m2_moy  
FROM Departement d  
JOIN Commune c  
ON c.id_code_dep = d.id_code_dep  
JOIN Bien b  
ON b.id_codedep_codecommune = c.id_codedep_codecommune  
JOIN Vente v  
ON v.id_bien = b.id_bien  
GROUP BY dep_nom  
ORDER BY prix_m2_moy DESC  
LIMIT 10
```

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

---



	nO_departement	nom_departement	prix_m2_moy
1	75	Paris	12129
2	92	Hauts-de-Seine	7415
3	94	Val-de-Marne	5399
4	06	Alpes-Maritimes	4685
5	93	Seine-Saint-Denis	4371
6	74	Haute-Savoie	4150
7	78	Yvelines	4126
8	69	Rhône	4064
9	2A	Corse-du-Sud	3922
10	33	Gironde	3839



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



```
SELECT reg_nom,  
       ROUND(AVG(ROUND(valeur/surface))) prix_m2_moy, "€/m2"  
FROM   Region r  
JOIN   Departement d  
       ON d.id_code_reg = r.id_code_reg  
JOIN   Commune c  
       ON c.id_code_dep = d.id_code_dep  
JOIN   Bien b  
       ON b.id_codedep_codecommune = c.id_codedep_codecommune  
JOIN   Vente v  
       ON v.id_bien = b.id_bien  
WHERE  (1 = 1  
       AND type_local = "Maison"  
       AND reg_nom = "Ile-de-France"  
       )
```

Grid view

Form view

✓ ✗ ⏪ ⏩ 1 ⏴ ⏵ 🖨 Total rows loaded: 1

reg_nom	prix_m2_moy	€/m2
Ile-de-France	3998	€/m2



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

---



```
SELECT valeur prix_appartement,  
       surface m2,  
       reg_nom region  
FROM Region r  
  JOIN Departement d  
    ON d.id_code_reg = r.id_code_reg  
  JOIN Commune c  
    ON c.id_code_dep = d.id_code_dep  
  JOIN Bien b  
    ON b.id_codedep_codecommune = c.id_codedep_codecommune  
  JOIN Vente v  
    ON v.id_bien = b.id_bien  
WHERE (1 = 1  
       AND type_local = "Appartement"  
      )  
ORDER BY prix_appartement DESC  
LIMIT 10
```

prix_appartement	m2	region
9000000	10	Ile-de-France
8600000	62	Ile-de-France
8577713	289	Ile-de-France
7620000	42	Ile-de-France
7600000	200	Ile-de-France
7535000	143	Ile-de-France
7420000	357	Ile-de-France
7200000	241	Ile-de-France
7050000	310	Ile-de-France
6600000	76	Ile-de-France



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



```
SELECT ROUND((t2-t1)*100.0/t1,2) Taux_evo_T1_T2,"%"
FROM

(SELECT COUNT(id_vente) t1
FROM vente v
WHERE (date >= "2020/01/01")
      AND (date <= "2020/03/31")
),

(SELECT COUNT(id_vente) t2
FROM vente v
WHERE (date >= "2020/04/01")
      AND (date <= "2020/06/30")
)
```

Grid view

Form view



1



Total rows loaded: 1

Taux\_evo\_T1\_T2 | %

3.66 %



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



```
SELECT reg_nom Region, ROUND(AVG(ROUND(valeur/surface))) prix_m2_moy
FROM Region r
  JOIN Departement d
    ON d.id_code_reg = r.id_code_reg
  JOIN Commune c
    ON c.id_code_dep = d.id_code_dep
  JOIN Bien b
    ON b.id_codedep_codecommune = c.id_codedep_codecommune
  JOIN Vente v
    ON v.id_bien = b.id_bien
WHERE (1 = 1
      AND type_local = "Appartement"
      AND nb_pieces > 4
      )
GROUP BY reg_nom
ORDER BY prix_m2_moy DESC;
```

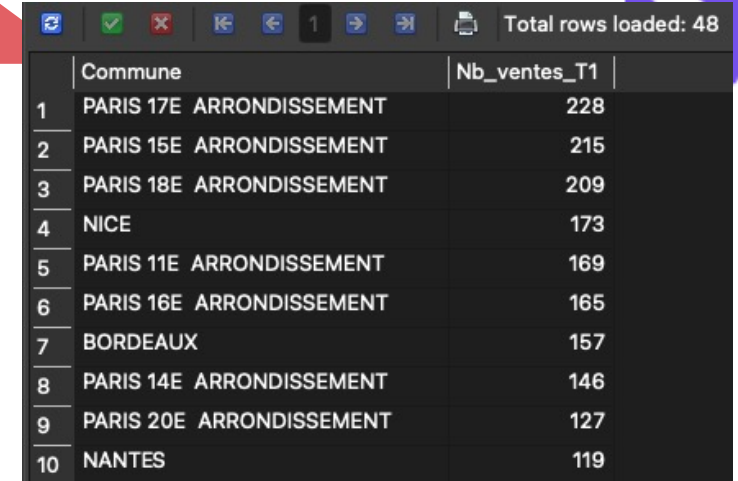
	Region	prix_m2_moy
1	Ile-de-France	8015
2	La Réunion	3660
3	Corse	3047
4	Provence-Alpes-Côte d'Azur	3005
5	Auvergne-Rhône-Alpes	2769
6	Nouvelle-Aquitaine	2510
7	Bretagne	2272
8	Hauts-de-France	2204
9	Pays de la Loire	2187
10	Occitanie	2096
11	Normandie	1994
12	Centre-Val de Loire	1428
13	Grand Est	1313
14	Bourgogne-Franche-Comté	1069
15	Martinique	564



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

---

```
SELECT com_nom Commune,  
       COUNT(id_vente) Nb_ventes_T1  
FROM Commune c  
  JOIN Bien b  
    ON b.id_codedep_codecommune = c.id_codedep_codecommune  
  JOIN Vente v  
    ON v.id_bien = b.id_bien  
WHERE (1 = 1  
       AND date >= "2020/01/01"  
       AND date <= "2020/03/31"  
      )  
GROUP BY c.id_codedep_codecommune  
HAVING Nb_ventes_T1 >= 50  
ORDER BY Nb_ventes_T1 DESC;
```



Total rows loaded: 48

	Commune	Nb_ventes_T1
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

37	SETE	62
38	LA CIOTAT	62
39	PARIS 2E ARRondissement	61
40	RENNES	61
41	LEVALLOIS-PERRET	59
42	TOULON	59
43	PARIS 4E ARRondissement	59
44	SAINT-MAUR-DES-FOSSES	56
45	VERSAILLES	54
46	AJACCIO	54
47	PUTEAUX	53
48	ISSY-LES-MOULINEAUX	50





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



```
SELECT ROUND((prix_m2_moy_F3 - prix_m2_moy_F2)*100/prix_m2_moy_F3,2) F2_F3_Pourcentage,"%"
FROM
(
  (SELECT AVG(valeur/surface) prix_m2_moy_F2
   FROM vente v, bien b
   WHERE ( 1 = 1
          AND type_local = "Appartement"
          AND nb_pieces = 2
          AND v.id_bien = b.id_bien
         )
   GROUP BY type_local
  ),
  (SELECT AVG(valeur/surface) prix_m2_moy_F3
   FROM vente v, bien b
   WHERE (1 = 1
          AND type_local = "Appartement"
          AND nb_pieces = 3
          AND v.id_bien = b.id_bien
         )
   GROUP BY type_local
  )
);
```

Grid view

Form view

✓ ✗ ⏪ ⏩ 1 ⏴ ⏵ 📄 Total rows loaded: 1

F2\_F3\_Pourcentage | %

-15.11 %



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

```
SELECT *
FROM
  (SELECT id_code_dep NO_Dpt, com_nom Commune, ROUND(AVG(valeur)) Moy_valeur_fonciere
   FROM vente v, bien b, commune c
   WHERE ( 1= 1
           AND v.id_bien = b.id_bien
           AND b.id_codedep_codecommune = c.id_codedep_codecommune
           AND c.id_code_dep ="06"
          )
   GROUP BY b.id_codedep_codecommune
   ORDER BY Moy_valeur_fonciere DESC
   LIMIT 3
  )
UNION
SELECT *
FROM
  (
    SELECT id_code_dep NO_Dpt, com_nom Commune, ROUND(AVG(valeur)) Moy_valeur_fonciere
    FROM vente v, bien b, commune c
    WHERE ( 1= 1
           AND v.id_bien = b.id_bien
           AND b.id_codedep_codecommune = c.id_codedep_codecommune
           AND c.id_code_dep ="13"
          )
  )
  AND b.id_codedep_codecommune = c.id_codedep_codecommune
  AND c.id_code_dep ="69"
  )
  GROUP BY c.id_code_dep, b.id_codedep_codecommune
  ORDER BY Moy_valeur_fonciere DESC
  LIMIT 3|
)
ORDER BY NO_Dpt, Moy_valeur_fonciere DESC;
```

NO_Dpt	Commune	Moy_valeur_fonciere
06	SAINT-JEAN-CAP-FERRAT	968750
06	EZE	655000
06	MOUANS-SARTOUX	476898
13	GIGNAC-LA-NERTHE	330000
13	SAINT-SAVOURNIN	314425
13	CASSIS	313417
33	LEGE-CAP-FERRET	549501
33	VAYRES	335000
33	ARCACHON	307436
59	BERSEE	433202
59	CYSOING	408550
59	HALLUIN	322250
69	VILLE-SUR-JARNIOUX	485300
69	LYON 2E ARRONDISSEMENT	455217
69	LYON 6E ARRONDISSEMENT	426968

```

SELECT com_nom Commune,
       COUNT(id_vente) nb_transactions,
       ROUND(COUNT(id_vente)*1000.0/population_totale,2) nb_tr_1000
FROM Commune c
  JOIN Bien b
  ON b.id_codedep_codecommune = c.id_codedep_codecommune
  JOIN Vente v
  ON v.id_bien = b.id_bien
WHERE population_totale >10000
GROUP BY Commune
ORDER BY nb_tr_1000 DESC
LIMIT 20;

```



Total rows loaded: 20			
	Commune	nb_transactions	nb_tr_1000
1	PARIS 2E ARRondissement	127	5.84
2	PARIS 1ER ARRondissement	78	4.86
3	PARIS 3E ARRondissement	161	4.69
4	ARCACHON	55	4.62
5	LA BAULE-ESCOUBLAC	77	4.58
6	PARIS 4E ARRondissement	119	4.05
7	ROQUEBRUNE-CAP-MARTIN	52	3.99
8	PARIS 8E ARRondissement	139	3.83
9	SANARY-SUR-MER	60	3.5
10	PARIS 9E ARRondissement	208	3.43
11	LA LONDE-LES-MAURES	37	3.43
12	PARIS 6E ARRondissement	139	3.38
13	SAINT-CYR-SUR-MER	37	3.16
14	CHANTILLY	35	3.13
15	SAINT-MANDE	69	3.06
16	PORNICHET	35	3.06
17	PARIS 10E ARRondissement	264	3.04
18	MENTON	91	2.94
19	SAINT-HILAIRE-DE-RIEZ	33	2.87
20	VINCENNES	141	2.81

**Les 20 communes avec le plus de transactions  
 pour 1000 habitants  
 pour les communes >10 000 habitants**

---



**Merci !**