



Création et utilisation d'une base de données immobilières



Laplace Immo

Laurent EBEDE OBAMA

Contexte du projet

- Les données sur les ventes immobilières sont très surchargées et incomplètes.
- Création d'une base de données structurée sur les ventes des biens immobiliers pour le 1^{er} semestre 2020.
- **Objectif** : améliorer la prediction des prix des biens immobiliers du réseau national d'agence immobilières.



La stratégie de sauvegarde et la conformité RGPD

- Définir la finalité de la collecte des données.
- **Ne pas collecter les informations inutiles**: suppression du nom de l'acquéreur et des autres variables non utiles pour le projet.
- Tenir un registre des activités de traitement.
- **Mettre en place des mesures de sécurité d'accès aux données** : les données sont stockées dans un espace sécurisé accessible via un mot de passe.
- **Garantir la transparence sur le traitement des données** : mise à disposition de l'ensemble des traitements effectués sur les données.

Les données initiales

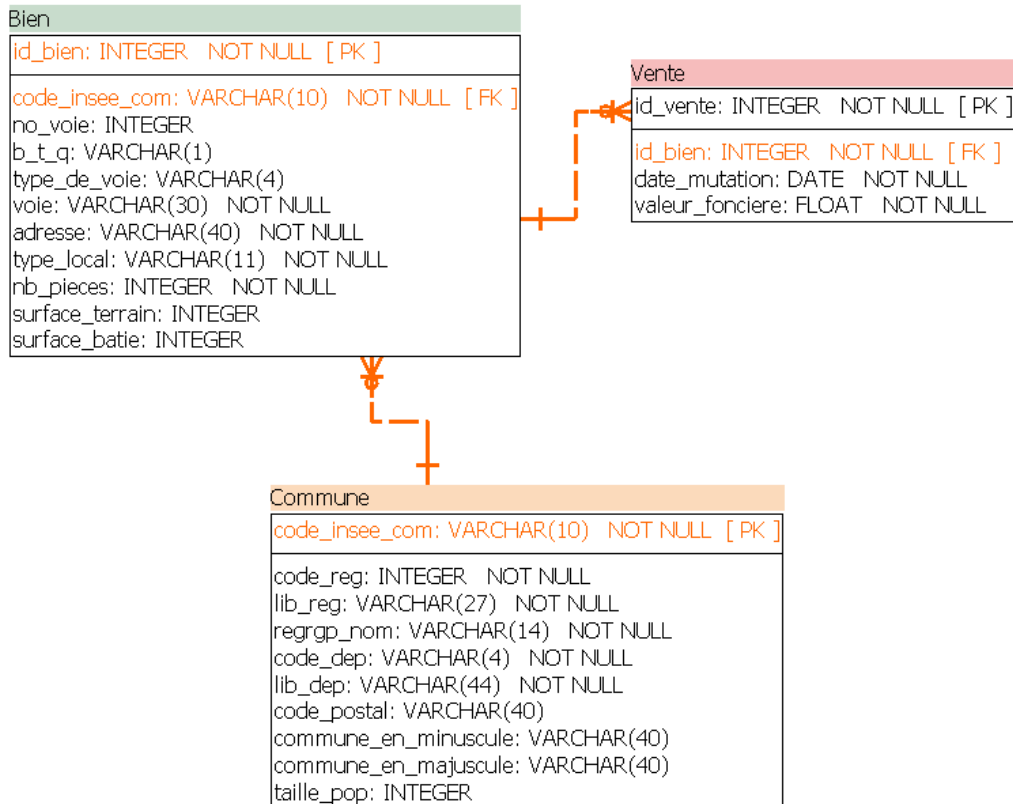
- **Les données sur les valeurs foncières** du 1^{er} semestre 2020 (**BNDP**): vente des biens immobiliers, les caractéristiques (prix, date de mutation, surface, commune, etc) du bien et les personnes concernées.
- **Les données sur les communes (INSEE)** : région, département et population.
- **Le référentiel géographique français (data.gouv)** : communes, unités urbaines, aires urbaines, départements, académies, régions, etc.

L'extrait du dictionnaire des données

DICTIONNAIRE DES DONNÉES - Valeurs foncières

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
Id_bien	ID du bien vendu dans la base de données	Integer	NC	Calculer	Ne doit pas être nul	Ordre d'apparition du local dans la base
code_departement	Code géographique du département	Varchar	4	Elémentaire	Ne doit pas être nul	
code_postal	Code postal de localisation du local	Varchar	40	Elémentaire	Ne doit pas être nul	
code_id_commune	ID de la commune	Integer	NC	Elémentaire	Ne doit pas être nul	
commune	Libellé de la commune en majuscule	Varchar	35	Elémentaire	Ne doit pas être nul	
code_insee_com	Code Insee de la commune	Varchar	10	Calculer	Ne doit pas être nul	Concatenation des colonnes code département et code commune
no_voie	Numéro dans la voie	Integer	NC	Elémentaire		
B_T_Q	Indice de répétition	Varchar	1	Elémentaire		
type_de_voie	Plusieurs valeurs (rue, avenue, chemin, etc.)	Varchar	4	Elémentaire		Concatenation des colonnes no_voie, B_T_Q, Type_de_voie et voie
voie	Libellé de la voie	Varchar	30	Elémentaire	Ne doit pas être nul	
adresse	Numéro et nom de la voie de localisation du bien	Varchar	40	Calculer	Ne doit pas être nul	
type_local	Type du local	Varchar	11	Elémentaire	Ne doit pas être nul	Transformation de la variable au format aaaa-mm-jj
nb_pieces	Nombre de pieces principales	Integer	NC	Elémentaire	Ne doit pas être nul	
surface_terrain	Surface totale du terrain	Integer	NC	Elémentaire		
surface_batie	Surface réelle du local + surface des dépendances	Integer	NC	Elémentaire		
valeur_fonciere	Prix de vente du local	Float	NC	Elémentaire		
date_vente	Date de vente du local	Date	NC	Changement de format	Format de la date : aaaa-mm-jj	

Le schéma relationnel normalisé



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

- i [16:26:25] Imported data to the table 'Commune' successfully. Number of imported rows: 38916
- i [16:26:56] Imported data to the table 'Bien' successfully. Number of imported rows: 34169
- i [16:27:32] Imported data to the table 'Vente' successfully. Number of imported rows: 34169

Les Screenshot qui permettent de démontrer le bon chargement des données

Commune

Grid view		Form view								
				Total rows loaded: 38916						
	code inse	code req	lib req	reqrap nom	code dep	lib dep	code postal	commune en minuscule	commune en majuscule	taille pop
1	65199	76	Occitanie	Province	65	Hautes-Pyrénées	65240	Germ	GERM	36
2	66154	76	Occitanie	Province	66	Pyrénées-Orientales	66210	Puyvalador	PUYVALADOR	63
3	2A021	94	Corse	Province	2A	Corse-du-Sud	20140	Argiusta-Moriccio	ARGIUSTA-MORICCIO	77
4	79124	75	Nouvelle-Aquitaine	Province	79	Deux-Sèvres	79340	Les Forges	LES FORGES	112
5	88347	44	Grand Est	Province	88	Vosges	88270	Pierrefitte	PIERREFITTE	118
6	2A119	94	Corse	Province	2A	Corse-du-Sud	20157	Frasseto	FRASSETO	137
7	50033	28	Normandie	Province	50	Manche	50270	Baubigny	BAUBIGNY	141
8	38040	84	Auvergne-Rhône-Alpes	Province	38	Isère	38142	Besse	BESSE	141
9	2A200	94	Corse	Province	2A	Corse-du-Sud	20134	Palneca	PALNECA	143
10	65481	76	Occitanie	Province	65	Hautes-Pyrénées	65120	Barèges	BAREGES	159
11	2A018	94	Corse	Province	2A	Corse-du-Sud	20110	Arbellara	ARBELLARA	160
12	2A035	94	Corse	Province	2A	Corse-du-Sud	20110	Belvédère-Campomoro	BELVEDERE-CAMPOMORO	161
13	65304	76	Occitanie	Province	65	Hautes-Pyrénées	65170	Caillan	CAILLAN	172

Les Screenshot qui permettent de démontrer le bon chargement des données

Bien

Grid view		Form view									
				1						Total rows loaded: 34169	
	id bien	code insee com	no voie	b t a	type de v	voie	adresse	type local	nb pieces	surface terrain	surface batie
1	1	01004	59		RUE	ALEXANDRE BERARD	59 RUE ALEXANDRE BERARD	Appartement	6		143
2	2	01004	29		RUE	REINE CLOTILDE	29 RUE REINE CLOTILDE	Appartement	2		62
3	3	01004	54		RUE	MARCEL DEMIA	54 RUE MARCEL DEMIA	Appartement	3		68
4	4	01004	25		RUE	DE LA COMMUNE 1871	25 RUE DE LA COMMUNE 1871	Maison	3		68
5	5	01004	60		ALL	DES FRERES CAUDRON	60 ALL DES FRERES CAUDRON	Appartement	2		51
6	6	01004	13		RUE	JEAN MONNET	13 RUE JEAN MONNET	Appartement	2		55
7	7	01004	72		RUE	DE LA REPUBLIQUE	72 RUE DE LA REPUBLIQUE	Appartement	2		33
8	8	01004	2		RUE	DE LA BATISSE	2 RUE DE LA BATISSE	Appartement	3		81
9	9	01004	36		RUE	DES APOTRES	36 RUE DES APOTRES	Appartement	2		79
10	10	01004	54	B	RUE	ARISTIDE BRIAND	54B RUE ARISTIDE BRIAND	Appartement	3		75

Les Screenshot qui permettent de démontrer le bon chargement des données

Vente

Grid view Form view

☑ ☒ ⏪ ⏩ 1 ⏴ ⏵ 🖨 Total rows loaded: 34169

	id vente	id bien	date vente	Valeur fonciere
1	1	4439	2020-04-22	78000
2	2	5808	2020-02-21	71600
3	3	6732	2020-02-28	54459
4	4	6733	2020-02-28	102151
5	5	6735	2020-03-10	102200
6	6	33694	2020-01-06	181230
7	7	33696	2020-02-13	165010
8	8	33697	2020-04-08	181930
9	9	5282	2020-01-30	177450
10	10	5208	2020-06-10	305000












Requêtes SQL et résultats

Requête 1

Nombre total d'appartements vendus au 1er semestre 2020

```
SELECT
    COUNT(*) AS Nb_appartements_vendus
FROM Bien B
LEFT JOIN Vente V
    ON B.id_bien = V.id_bien
WHERE (date_vente BETWEEN '2020-01-01' AND '2020-06-31')
    AND type_local = 'Appartement';
```

Grid view		Form view	
			
			
	Total rows loaded: 1		
Nb appartements vendus			
1	31378		


Requête 2


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


```
SELECT
    code_reg,
    lib_reg,
    COUNT(V.id_vente) AS Nb_appartements_vendus
FROM Bien B
JOIN Vente V
    ON B.id_bien = V.id_bien
JOIN Commune C
    ON C.code_insee_com = B.code_insee_com
WHERE (date_vente BETWEEN '2020-01-01' AND '2020-06-31')
    AND type_local = 'Appartement'
GROUP BY code_reg, lib_reg;
```


Grid view


Form view














1















Total rows loaded: 17

	code req	lib req	Nb appartements vendus
1	1	Guadeloupe	2
2	2	Martinique	94
3	3	Guyane	34
4	4	La Réunion	44
5	11	Ile-de-France	13995
6	24	Centre-Val de Loire	696
7	27	Bourgogne-Franche-Comté	376
8	28	Normandie	862
9	32	Hauts-de-France	1254
10	44	Grand Est	984
11	52	Pays de la Loire	1357
12	53	Bretagne	983
13	75	Nouvelle-Aquitaine	1932
14	76	Occitanie	1640
15	84	Auvergne-Rhône-Alpes	3253
16	93	Provence-Alpes-Côte d'Azur	3649
17	94	Corse	223

Requête 3

Proportion des ventes d'appartements par le nombre de pièces

```
SELECT
    nb_pieces,
    ROUND((COUNT(id_bien) / CAST(31378 AS FLOAT)) * 100,3) AS proportion_ventes_appartements
FROM bien
WHERE type_local = 'Appartement'
GROUP BY nb_pieces ;
```

Grid view		Form view
		    1    Total rows loaded: 12
	nb pieces	proportion ventes appartements
1	0	0.096
2	1	21.477
3	2	31.178
4	3	28.574
5	4	14.214
6	5	3.55
7	6	0.65
8	7	0.172
9	8	0.054
10	9	0.025
11	10	0.006
12	11	0.003

Requête 4

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

```

SELECT
  code_dep,
  lib_dep,
  MAX(prix_metre_carre) AS max_prix_metre_carre,
  ROUND(AVG(prix_metre_carre)) AS moy_prix_metre_carre
FROM (
  SELECT
    B.id_bien,
    code_dep,
    lib_dep,
    valeur_fonciere,
    surface_batie,
    valeur_fonciere/ surface_batie AS prix_metre_carre
  FROM Bien B
  JOIN Vente V
    ON B.id_bien = V.id_bien
  JOIN Commune C
    ON C.code_insee_com = B.code_insee_com
  WHERE prix_metre_carre IS NOT NULL
)
GROUP BY code_dep, lib_dep
ORDER BY max_prix_metre_carre DESC, moy_prix_metre_carre DESC
LIMIT 10;

```

Grid view		Form view										Total rows loaded: 10	
1	code dep	lib dep	max prix metre carre		moy prix metre carre								
2	75	Paris	900000		12129								
3	94	Val-de-Marne	175000		5398								
4	91	Essonne	138709		3230								
5	77	Seine-et-Marne	91075		3221								
6	59	Nord	80833		2837								
7	92	Hauts-de-Seine	78260		7415								
8	06	Alpes-Maritimes	70200		4685								
9	01	Ain	58250		3384								
10	44	Loire-Atlantique	57142		3434								
	30	Gard	42075		2756								

Requête 5

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

```
SELECT
    code_reg,
    lib_reg,
    ROUND(AVG(prix_metre_carre)) AS prix_moyen_M2_maison
FROM (
    SELECT
        B.id_bien,
        code_reg,
        lib_reg,
        valeur_fonciere,
        surface_batie,
        valeur_fonciere/ surface_batie AS prix_metre_carre
    FROM Bien B
    JOIN Vente V
        ON B.id_bien = V.id_bien
    JOIN Commune C
        ON C.code_insee_com = B.code_insee_com
    WHERE type_local = 'Maison'
)
GROUP BY code_reg, lib_reg
HAVING code_reg == 11;
```

Grid view		Form view	
		1	
		Total rows loaded: 1	
	code req	lib req	prix moyen M2 maison
1	11	Ile-de-France	3997

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









```
SELECT
    B.id_bien,
    lib_reg AS region,
    valeur_fonciere,
    surface_batie
FROM Bien B
JOIN Vente V
    ON B.id_bien = V.id_bien
JOIN Commune C
    ON C.code_insee_com = B.code_insee_com
WHERE type_local = 'Appartement'
ORDER BY valeur_fonciere DESC
LIMIT 10;
```

Grid view		Form view	
	<input checked="" type="checkbox"/>		Total rows loaded: 10
1			
	id bien	region	Valeur fonciere
	surface batie		
1	19408	Ile-de-France	9000000
2	26876	Ile-de-France	8600000
3	16945	Ile-de-France	8577713
4	19569	Ile-de-France	7620000
5	16842	Ile-de-France	7600000
6	16215	Ile-de-France	7535000
7	19073	Ile-de-France	7420000
8	19232	Ile-de-France	7200000
9	16184	Ile-de-France	7050000
10	16219	Ile-de-France	6600000

Requête 7

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

```
SELECT
  nb_vente_1er_trim,
  nb_vente_2e_trim,
  ROUND((nb_vente_2e_trim - nb_vente_1er_trim) / CAST(nb_vente_1er_trim AS FLOAT) * 100, 2) AS taux_evolution
FROM (
  SELECT
    SUM(CASE WHEN date_vente >= '2020-01-01' AND date_vente <= '2020-03-31' THEN 1 ELSE 0 END) AS nb_vente_1er_trim,
    SUM(CASE WHEN date_vente >= '2020-04-01' AND date_vente <= '2020-06-30' THEN 1 ELSE 0 END) AS nb_vente_2e_trim
  FROM vente);
```

Grid view		Form view	
			
			
		Total rows loaded: 1	
nb vente 1er trim		nb vente 2e trim	
1		3.68	

Requête 8

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

```
WITH tab_appart_4p AS(
  SELECT
    code_reg,
    lib_reg,
    ROUND(AVG(prix_metre_carre)) AS prix_moy_M2_appart_4pces
  FROM (
    SELECT
      B.id_bien,
      nb_pieces ,
      code_reg,
      lib_reg,
      valeur_fonciere,
      surface_batie,
      valeur_fonciere/ surface_batie AS prix_metre_carre
    FROM Bien B
    JOIN Vente V
      ON B.id_bien = V.id_bien
    JOIN Commune C
      ON C.code_insee_com = B.code_insee_com
    WHERE type_local = 'Appartement' AND nb_pieces > 4
  )
  GROUP BY code_reg, lib_reg
  ORDER BY prix_moy_M2_appart_4pces DESC)

SELECT
  lib_reg,
  prix_moy_M2_appart_4pces,
  ROW_NUMBER() OVER(ORDER BY prix_moy_M2_appart_4pces DESC) AS classement_region
FROM tab_appart_4p;
```

Grid view Form view			
1 Total rows loaded: 15			
	lib req	prix moy M2 appart 4pces	classement region
1	Ile-de-France	8015	1
2	La Réunion	3659	2
3	Corse	3046	3
4	Provence-Alpes-Côte d'Azur	3005	4
5	Auvergne-Rhône-Alpes	2768	5
6	Nouvelle-Aquitaine	2510	6
7	Bretagne	2271	7
8	Hauts-de-France	2203	8
9	Pays de la Loire	2186	9
10	Occitanie	2096	10
11	Normandie	1994	11
12	Centre-Val de Loire	1428	12
13	Grand Est	1313	13
14	Bourgogne-Franche-Comté	1068	14
15	Martinique	564	15

Requête 9

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

```
SELECT
    commune_en_minuscule AS commune,
    COUNT(V.id_vente) AS nb_vente
FROM Bien B
JOIN Vente V
    ON B.id_bien = V.id_bien
JOIN Commune C
    ON C.code_insee_com = B.code_insee_com
WHERE date_vente >= '2020-01-01' AND date_vente <= '2020-03-31'
GROUP BY commune_en_minuscule
HAVING nb_vente >= 50
ORDER BY nb_vente DESC;
```

Grid view		Form view	
		1	
		Total rows loaded: 48	
	commune	nb vente	
1	Paris 17e	228	
2	Paris 15e	215	
3	Paris 18e	209	
4	Nice	173	
5	Paris 11e	169	
6	Paris 16e	165	
7	Bordeaux	157	
8	Paris 14e	146	
9	Paris 20e	127	
10	Nantes	119	

Requête 10

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

```
WITH tab_appart_3P AS (
    SELECT
        ROUND(AVG(valeur_fonciere/ surface_batie),2) AS prix_moyen_appart_3p
    FROM Bien B
    JOIN Vente V
        ON B.id_bien = V.id_bien
    WHERE type_local = 'Appartement' AND nb_pieces == 3),

tab_appart_2P AS (
    SELECT
        ROUND(AVG(valeur_fonciere/ surface_batie), 2) AS prix_moyen_appart_2p
    FROM Bien B
    JOIN Vente V
        ON B.id_bien = V.id_bien
    WHERE type_local = 'Appartement' AND nb_pieces == 2)

SELECT
    prix_moyen_appart_2p,
    prix_moyen_appart_3p,
    ROUND((prix_moyen_appart_2p - prix_moyen_appart_3p) / prix_moyen_appart_3p * 100, 2) AS diff_prix_2P_3P
FROM tab_appart_3P, tab_appart_2P;
```

Grid view		Form view	
		Total rows loaded: 1	
	prix moyen appart 2p	prix moyen appart 3p	diff prix 2P 3P
1	4932.42	4284.92	15.11

Requête 11

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

```
WITH tab_moy_commune AS(
  SELECT
    code_dep,
    commune_en_minuscule AS commune,
    ROUND(AVG(valeur_fonciere),2) AS moy_val_fonciere
  FROM Bien B
  JOIN Vente V
    ON B.id_bien = V.id_bien
  JOIN Commune C
    ON C.code_insee_com = B.code_insee_com
  WHERE code_dep IN ('6', '13', '33', '59', '69')
  GROUP BY commune_en_minuscule
  ORDER BY code_dep, commune_en_minuscule, moy_val_fonciere DESC),

tab_rang_moy_comm AS(
  SELECT *,
    ROW_NUMBER() OVER(
      PARTITION BY code_dep
      ORDER BY moy_val_fonciere DESC) AS rank_moy_val_fonc
  FROM tab_moy_commune)

SELECT
  code_dep,
  commune,
  moy_val_fonciere,
  rank_moy_val_fonc
FROM tab_rang_moy_comm
WHERE rank_moy_val_fonc <= 3;
```

Grid view Form view				
1 Total rows loaded: 12				
	code dep	commune	moy val fonciere	rank moy val fonc
1	13	Gignac-la-Nerthe	330000	1
2	13	Saint-Savournin	314425	2
3	13	Cassis	313416.88	3
4	33	Lège-Cap-Ferret	549500.64	1
5	33	Vayres	335000	2
6	33	Arcachon	307435.93	3
7	59	Bersée	433202	1
8	59	Cysoing	408550	2
9	59	Halluin	322250	3
10	69	Ville-sur-Jarnioux	485300	1
11	69	Lyon 2e	455217.26	2
12	69	Lyon 6e	426968.25	3

Requête 12

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

```
SELECT
  commune_en_minuscule AS commune,
  COUNT(V.id_vente) AS nb_vente,
  taille_pop,
  ROUND((COUNT(V.id_vente) / CAST(taille_pop AS FLOAT)) * 1000, 2) AS nb_vente_pour_1000hbs
FROM Bien B
JOIN Vente V
  ON B.id_bien = V.id_bien
JOIN Commune C
  ON C.code_insee_com = B.code_insee_com
WHERE taille_pop > 10000
GROUP BY commune_en_minuscule
ORDER BY nb_vente_pour_1000hbs DESC
LIMIT 20;
```

Grid view Form view				
1 Total rows loaded: 20				
	commune	nb vente	taille pop	nb vente pour 1000hbs
1	Paris 2e	127	21595	5.88
2	Paris 1er	79	15917	4.96
3	La Baule-Escoublac	77	16255	4.74
4	Paris 3e	161	34025	4.73
5	Arcachon	55	11630	4.73
6	Paris 4e	120	29131	4.12
7	Paris 8e	139	35655	3.9
8	Sanary-sur-Mer	60	16889	3.55
9	La Londe-les-Maures	37	10641	3.48
10	Paris 9e	208	60026	3.47
11	Paris 6e	139	40303	3.45
12	Saint-Cyr-sur-Mer	38	11484	3.31
13	Chantilly	35	10950	3.2
14	Pornichet	35	11141	3.14
15	Saint-Mandé	69	22377	3.08
16	Paris 10e	264	86472	3.05
17	Saint-Hilaire-de-Riez	33	11297	2.92
18	Vincennes	141	49788	2.83
19	Paris 17e	470	166543	2.82
20	Enghien-les-Bains	32	11470	2.79



Merci !