



# Création et utilisation de la base de données

Julien Séré



Laplace Immo

# Contexte du projet

Création d'un modèle de prédiction des ventes sur le marché immobilier

- Collecte des informations immobilières et foncières en France
- Analyse du marché
- Appui des agences régionales pour l'accompagnement aux clients



# La stratégie de sauvegarde et la conformité RGPD

- Licéité, loyauté et transparence : consentement et confidentialité
- Minimisation des données
- Exactitude : Mise à jour régulière de la base des données
- Sécurité et confidentialité
- Droit des individus
- Anonymisation des données : Suppression des informations personnelles (adresses précises)

# Les données initiales

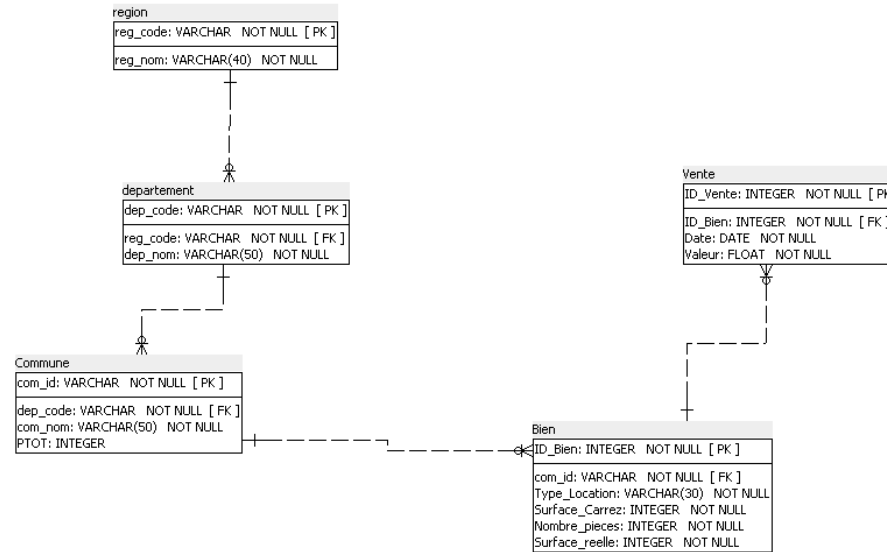
- Demandes de valeurs foncières (Open Data) – 2020 – 1<sup>er</sup> semestre
- Fichier de recensement des populations par commune (INSEE)
- Référentiel géographique français, communes, unités urbaines, aires urbaines, départements, académies, régions (Data.gouv)



# L'extrait du dictionnaire des données

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
<b>Id_bien</b>	ID du bien dans la base de données	Integer	NC	Elémentaire	NOT NULL	A créer à l'enregistrement
<b>ID_vente</b>	ID unique pour chaque vente	Integer	NC	Elementaire	NOT NULL	A créer à l'enregistrement
<b>com_id</b>	ID du département + ID commune	Varchar	10	Concaténer	NOT NULL	Concaténation des colonnes "Code departement" et "code commune"
<b>Date</b>	Date de la vente	DATE	NC	Elementaire	NOT NULL	Sous la forme YYYY/MM/DD
<b>Valeur</b>	Valeur fonciere en euros	FLOAT	NC	Elementaire	NOT NULL	
<b>Type_location</b>	Plusieurs valeurs (maison, appartement)	Varchar	30	Elémentaire	NOT NULL	
<b>Nombre_pieces</b>	Nombre de pièces	Integer	NC	Elementaire	NOT NULL	
<b>Surface_Carrez</b>	Surface du bien mesurée en loi Carrez	Integer	NC	Elementaire	NOT NULL	
<b>Surface_réelle</b>	Surface réelle du bien	Integer	NC	Elementaire		
<b>reg_nom</b>	Nom de la région	VARCHAR	40	Elémentaire	NOT NULL	
<b>com_nom</b>	Nom de la commune en minuscules	VARCHAR	50	Elémentaire	NOT NULL	
<b>dep_code</b>	Code Du département	VARCHAR	10	Concaténer	NOT NULL	Concaténation 0 et dep_code si chaine de caractere = 1
<b>reg_code</b>	Code de la région	VARCHAR	10	Elémentaire	NOT NULL	
<b>dep_nom</b>	Nom du département	VARCHAR	50	Elementaire	NOT NULL	
<b>PTOT</b>	Population Totale	Integer	NC	Elémentaire		

# Le schéma relationnel normalisé



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



Local instance MySQL80

p5

## Schema Details

Default collation: **utf8mb4\_0900\_ai\_ci**







Default character set: **utf8mb4**

Table count: **6**

Database size (rough estimate): **9.1 MiB**

Table	Column	Type
bien	com_id	varchar(10)
bien	ID_Bien	int
bien	Nombre_pieces	int
bien	Surface_Carrez	int
bien	Surface_reelle	int
bien	Type_Location	varchar(30)
commune	com_id	varchar(10)
commune	com_nom	varchar(50)
commune	dep_code	varchar(10)
commune	PTOT	int
departement	dep_code	varchar(10)
departement	departement_reg_code	varchar(10)
nom_dep	dep_code	varchar(10)
nom_dep	dep_nom	varchar(50)
region	reg_code	varchar(10)
region	reg_nom	varchar(40)
vente	Date	date
vente	ID_Vente	int
vente	Valeur	float
vente	Vente_ID_Bien	int

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

Name	Engine	V... ▲	Row Format	Rows
 bien	InnoDB	10	Dynamic	30783
 commune	InnoDB	10	Dynamic	35229
 departement	InnoDB	10	Dynamic	109
 nom_dep	InnoDB	10	Dynamic	109
 region	InnoDB	10	Dynamic	19
 vente	InnoDB	10	Dynamic	34468





## Requêtes SQL et résultats

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

```
SELECT
  bien.Type_location AS 'Type de bien',
  COUNT(bien.type_location) AS 'Nombre de biens vendus',
  CONCAT(' Du ',
    MIN(vente.date),
    ' au ',
    MAX(vente.date)) AS 'Période'
FROM
  bien
  JOIN
    vente ON bien.id_bien= vente.Vente_ID_bien
WHERE
  (vente.date >= '2020-01-01')
  AND (vente.date <= '2020-06-30')
  AND LOWER(bien.type_location) = 'appartement'
GROUP BY bien.type_location;
```

Type de bien	Nombre de biens vendus	Période
Appartement	31 378	Du 2020-01-02 au 2020-06-30

## 2) Le nombre de ventes d'appartement par région pour le 1<sup>er</sup> semestre 2020

```

SELECT
    COUNT(ID_bien) AS 'Nb de biens vendus pour le premier semestre
2020',
    departement.departement_reg_code AS Code_region,
    reg_nom AS 'Nom Région'
FROM
    vente
    LEFT JOIN
    bien on vente.Vente_ID_Bien = bien.ID_Bien
    LEFT JOIN
    commune on bien.com_ID = commune.com_ID
    LEFT JOIN
    departement ON commune.dep_code = departement.dep_code
    LEFT JOIN
    region ON departement.departement_reg_code = region.reg_code
WHERE
    (vente.date >= '2020-01-01')
    AND (vente.Date <= '2020-06-30')
    AND LOWER(bien.type_location) = 'appartement'
GROUP BY departement.departement_reg_code , bien.type_location
ORDER BY COUNT(bien.type_location) DESC;

```

Nb de biens vendus pour le premier semestre 2020			Code_region	Nom Région
13995	11	Ile-de-France		
3649	93	Provence-Alpes-Côte d'Azur		
3253	84	Auvergne-Rhône-Alpes		
1932	75	Nouvelle-Aquitaine		
1640	76	Occitanie		
1357	52	Pays de la Loire		
1254	32	Hauts-de-France		
984	44	Grand Est		
983	53	Bretagne		
862	28	Normandie		
696	24	Centre-Val de Loire		
		Bourgogne-Franche-Comté		
376	27	Comté		
223	94	Corse		
94	2	Martinique		
44	4	La Réunion		
34	3	Guyane		
2	1	Guadeloupe		

### 3) Proportion des ventes d'appartements par le nombre de pièces.

```
SELECT
  bien.Nombre_pieces AS 'Nombre de pièces',
  COUNT(vente.vente_ID_Bien) AS 'Nombre de ventes',
  ROUND(COUNT(vente.vente_ID_Bien) * 100 / (
    SELECT COUNT(*)
    FROM vente
    LEFT JOIN bien ON bien.ID_Bien = vente.vente_ID_Bien
    WHERE LOWER(bien.type_location) = 'appartement'),2) AS
  'Proportion en %'
FROM
  vente
LEFT JOIN bien ON bien.id_bien = vente.vente_ID_Bien
WHERE
  LOWER(bien.type_location) = 'appartement'
GROUP BY
  bien.Nombre_pieces
ORDER BY
  bien.Nombre_pieces;
```

Nombre de pièces	Nombre de ventes	Proportion en %
0	30	0.10
1	6739	21.48
2	9783	31.18
3	8966	28.57
4	4460	14.21
5	1114	3.55
6	204	0.65
7	54	0.17
8	17	0.05
9	8	0.03
10	2	0.01
11	1	0.00

#### 4) Liste des 10 départements où le prix du mètre carré est le plus élevé.

```
SELECT
  d.dep_code,
  d.dep_nom,
  CONCAT(ROUND(SUM(v.Valeur) / SUM(b.Surface_Carrez), 2), "€") AS
`prix_m2`
FROM
  Vente v
JOIN
  Bien b ON v.Vente_ID_Bien = b.ID_Bien
JOIN
  Commune c ON b.com_id = c.com_id
JOIN
  Departement d ON c.dep_code = d.dep_code
WHERE
  b.Surface_Carrez > 0
GROUP BY
  d.dep_code, d.dep_nom
ORDER BY
  SUM(v.Valeur) / SUM(b.Surface_Carrez) DESC
LIMIT 10;
```

dep_code	dep_nom	prix_m2
75	Paris (75)	11896.61€
92	Hauts-de-Seine (92)	7237.13€
94	Val-de-Marne (94)	4830.75€
06	Alpes-Maritimes (06)	4587.97€
74	Haute-Savoie (74)	4158.07€
93	Seine-Saint-Denis (93)	4065.68€
78	Yvelines (78)	4001.79€
69	Rhône (69)	3892.57€
2A	Corse-du-Sud (2A)	3752.54€
33	Gironde (33)	3567.52€

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

```
SELECT
  bien.type_location AS 'Type de bien',
  CONCAT(ROUND(AVG(vente.valeur / bien.surface_carrez), 2), "€") AS
  "Prix moyen du metre carré",
  region.reg_code AS 'Région',
  region.reg_nom AS 'Nom Région'
FROM
  vente
LEFT JOIN bien ON bien.id_bien = vente.vente_ID_Bien
LEFT JOIN commune ON bien.com_ID = commune.com_ID
LEFT JOIN departement ON commune.dep_code =
departement.dep_code
LEFT JOIN region ON departement.departement_reg_code =
region.reg_code
WHERE
  LOWER(bien.type_location) = 'maison'
  AND region.reg_code = '11'
GROUP BY
  region.reg_code, bien.type_location ;
```

Type de bien	Valeur foncière moyenne	Région	Nom Région
Maison	3744.75€	11	Ile-de-France

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

```
SELECT
Vente.Vente_ID_Bien AS "Numéro d appartement",
CONCAT(vente.valeur, ' €') AS 'Prix',
Region.reg_nom AS "Région",
Bien.Surface_carrez AS "Nombre de m2"
FROM
vente
LEFT JOIN bien ON bien.id_bien = vente.vente_ID_Bien
LEFT JOIN commune ON bien.com_ID = commune.com_ID
LEFT JOIN departement ON commune.dep_code =
departement.dep_code
LEFT JOIN region ON departement.departement_reg_code =
region.reg_code
ORDER BY
'Prix' DESC
LIMIT 10;
```

Numéro d appartement	Prix	Région	Nombre de m2
30603	9000000	Ile-de-France	9
5261	8600000	Ile-de-France	64
3625	8577710	Ile-de-France	21
7602	7620000	Ile-de-France	43
9988	7600000	Ile-de-France	253
17823	7535000	Ile-de-France	140
410	7420000	Ile-de-France	361
16357	7200000	Ile-de-France	595
1924	7050000	Ile-de-France	123
19161	6600000	Ile-de-France	79

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

```
WITH
TRIM1 AS (
  SELECT COUNT(vente.vente_ID_Bien) AS VentesT1
  FROM vente
  WHERE vente.date >= '2020-01-01' AND vente.date <= '2020-03-31'
),
TRIM2 AS (
  SELECT COUNT(vente.vente_ID_Bien) AS VentesT2
  FROM vente
  WHERE vente.date >= '2020-04-01' AND vente.date <= '2020-06-30'
)
SELECT
  T1.VentesT1 AS ventes_T1,
  T2.VentesT2 AS ventes_T2,
  ROUND((T2.VentesT2 - T1.VentesT1) * 100.0 / T1.VentesT1, 2) AS
taux_evolution
FROM
  TRIM1 AS T1,
  TRIM2 AS T2;
```

ventes_T1	ventes_T2	taux_evolution
16776	17393	3.68



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

```
WITH ClassementRegions AS (  
  SELECT  
    ROW_NUMBER() OVER (ORDER BY ROUND(AVG(vente.valeur /  
bien.surface_carrez), 2) DESC) AS Classement,  
    region.reg_nom AS "Région",  
    CONCAT(ROUND(AVG(vente.valeur / bien.surface_carrez), 2), "€")  
  AS "Prix moyen au M2"  
  FROM  
    vente  
  LEFT JOIN bien ON bien.id_bien = vente.vente_ID_Bien  
  LEFT JOIN commune ON bien.com_ID = commune.com_ID  
  LEFT JOIN departement ON commune.dep_code =  
departement.dep_code  
  LEFT JOIN region ON departement.departement_reg_code =  
region.reg_code  
  WHERE  
    LOWER(bien.type_location) = 'appartement'  
    AND bien.nombre_pieces > 4  
    AND bien.surface_carrez > 0 -- Évite les divisions par zéro  
  GROUP BY  
    region.reg_nom  
)  
SELECT *  
FROM ClassementRegions  
WHERE Classement <= 10;
```

Classement	Région	Prix moyen au M2
1	Ile-de-France	8757.86€
2	La Réunion	3636.52€
3	Provence-Alpes-Côte d'Azur	3587.58€
4	Corse	3106.69€
5	Auvergne-Rhône-Alpes	2890.44€
6	Nouvelle-Aquitaine	2465.68€
7	Bretagne	2409.28€
8	Pays de la Loire	2315.2€
9	Hauts-de-France	2189.28€
10	Occitanie	2097.85€

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

```
SELECT
    commune.com_nom AS `Nom de la commune`,
    COUNT(vente.ID_Vente) AS `Nombre de Ventes au 1er trimestre`
FROM
    vente
LEFT JOIN
    bien ON bien.id_bien = vente.vente_ID_Bien
LEFT JOIN
    commune ON bien.com_ID = commune.com_ID
WHERE
    vente.date >= '2020-01-04'
    AND vente.date <= '2020-03-31'
GROUP BY
    commune.com_nom
HAVING
    COUNT(vente.ID_Vente) >= 50
ORDER BY
    `Nombre de Ventes au 1er trimestre` DESC;
```

Nom de la commune	Nombre de Ventes au 1er trimestre
Paris 17e Arrondissement	226
Paris 15e Arrondissement	212
Paris 18e Arrondissement	201
Nice	168
Paris 16e Arrondissement	165
Paris 11e Arrondissement	164
Bordeaux	154
Paris 14e Arrondissement	145
Paris 20e Arrondissement	128
Nantes	118
Paris 19e Arrondissement	116
Paris 10e Arrondissement	107
Paris 12e Arrondissement	106
Paris 9e Arrondissement	106
Grenoble	104
Boulogne-Billancourt	94
Paris 13e Arrondissement	94
Paris 7e Arrondissement	86
Paris 6e Arrondissement	85
Marseille 8e Arrondissement	80
Paris 3e Arrondissement	78
Asnières-sur-Seine	78
Courbevoie	78
Toulouse	77
Paris 5e Arrondissement	77
Antibes	75
Marseille 4e Arrondissement	71
Marseille 1er Arrondissement	70
Lille	67
Marseille 9e Arrondissement	65
Rueil-Malmaison	65
Vincennes	65
Angers	64
Montreuil	64
Nîmes	62
Sète	62
Paris 8e Arrondissement	62
La Clotat	61
Rennes	61
Paris 2e Arrondissement	60
Paris 4e Arrondissement	58
Toulon	58
Levallois-Perret	58
Saint-Maur-des-Fossés	55
Ajaccio	54
Versailles	53
Puteaux	53
Issy-les-Moulineaux	50

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

```
WITH
2Pieces AS (
    SELECT round(AVG(vente.valeur / bien.Surface_Carrez),2) AS
Prix_2pieces
    FROM vente
    LEFT JOIN bien ON bien.id_bien = vente.vente_ID_Bien
    WHERE
    LOWER(bien.type_location) = 'appartement'
    AND bien.nombre_pieces = 2
),
3Pieces AS (
    SELECT round(AVG(vente.valeur / bien.Surface_Carrez),2) AS
Prix_3pieces
    FROM vente
    LEFT JOIN bien ON bien.id_bien = vente.vente_ID_Bien
    WHERE
    LOWER(bien.type_location) = 'appartement'
    AND bien.nombre_pieces = 3
)
SELECT
    concat(2Pieces.Prix_2pieces, "€") AS "Prix moyen app 3 pièces",
    concat(2Pieces.Prix_2pieces) AS "Prix moyen app 2 pièces",
    ROUND((3Pieces.Prix_3pieces - 2Pieces.Prix_2pieces) * 100.0 /
2Pieces.Prix_2pieces, 2) AS "Différence de prix"
FROM
    3Pieces,
    2Pieces;
```

Prix moyen app 2 pièces	Prix moyen app 3 pièces	Différence de prix
4908.49	4300.81€	-12.38

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

```
WITH CommunesParDept AS(
SELECT departement.dep_code AS 'Département', com_nom AS
Commune, CONCAT(FORMAT(AVG(vente.valeur),2), ' EUR') AS 'Valeur
foncière moyenne',
RANK() OVER(PARTITION BY departement.dep_code ORDER BY
AVG(vente.valeur) DESC) AS 'Classement'
FROM bien
LEFT JOIN vente ON vente.vente_ID_Bien = bien.ID_Bien
LEFT JOIN commune ON commune.com_ID = Bien.com_ID
LEFT JOIN departement on commune.dep_code =
departement.dep_code
GROUP BY bien.com_ID
)
SELECT * FROM CommunesParDept
WHERE ((Classement <=3) AND (Département = '06')) OR ((Classement
<=3) AND (Département = '13')) OR ((Classement <=3) AND
(Département = '33')) OR ((Classement <=3) AND (Département = '59'))
OR ((Classement <=3) AND (Département = '69'));
```

Département	Commune	Valeur foncière moyenne	Classement
6	Saint-Jean-Cap-Ferrat	968,750.00 EUR	1
6	Eze	655,000.00 EUR	2
6	Mouans-Sartoux	476,898.09 EUR	3
13	Gignac-la-Nerthe	330,000.00 EUR	1
13	Saint-Savournin	314,425.00 EUR	2
13	Cassis	313,416.88 EUR	3
33	Lège-Cap-Ferret	549,500.64 EUR	1
33	Vayres	335,000.00 EUR	2
33	Arcachon	307,435.93 EUR	3
59	Bersée	433,202.00 EUR	1
59	Cysoing	408,550.00 EUR	2
59	Halluin	322,250.00 EUR	3
69	Ville-sur-Jarnioux	485,300.00 EUR	1
69	Lyon 2e Arrondissement	455,217.27 EUR	2
69	Lyon 6e Arrondissement	426,968.25 EUR	3

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

```
SELECT
commune.com_nom AS Communes,
(COUNT(vente.vente_id_bien) / commune.PTOT * 1000) AS
`Transactions pour 1000 habitants`
FROM
Bien
LEFT JOIN
Vente ON vente.vente_ID_Bien = bien.ID_Bien
LEFT JOIN
Commune ON bien.com_ID = commune.com_ID
WHERE commune.ptot > 10000
GROUP BY commune.com_ID
ORDER BY
`Transactions pour 1000 habitants` DESC
LIMIT 20;
```

Communes	Transactions pour 1000 habitants
Paris 2e Arrondissement	5.8431
Paris 1er Arrondissement	4.9206
Paris 3e Arrondissement	4.6931
Arcachon	4.6226
La Baule-Escoublac	4.5842
Paris 4e Arrondissement	4.0830
Roquebrune-Cap-Martin	3.9874
Paris 8e Arrondissement	3.8345
Sanary-sur-Mer	3.4965
Paris 9e Arrondissement	3.4344
La Londe-les-Maures	3.4336
Paris 6e Arrondissement	3.3762
Saint-Cyr-sur-Mer	3.2409
Chantilly	3.1312
Pornichet	3.0594
Saint-Mandé	3.0563
Paris 10e Arrondissement	3.0393
Menton	2.9373
Saint-Hilaire-de-Riez	2.8693
Vincennes	2.8071



**Merci !**