

Création et utilisation de la base de données

GILTIEN Jordan



Prévision des prix de ventes

- Modification de la base de donnée
- Création du schéma relation
- Import des données
- Requêtes demandées



La stratégie de sauvegarde et la conformité RGPD

- Données ciblées seulement
- Durée limitée
- PC personnel
- Vérification et suppression de données personnelles

Les données initiales

- Données de l'INSEE
- Données de data.gouv
- Données extraites du site des Demande de valeurs foncières

Création de CSV à partir de ces tableaux de données

- En filtrant
- En concatenant
- En incrémentant

L'extrait du dictionnaire des données

DICTIONNAIRE DES DONNÉES - Valeurs foncières

| CODE | SIGNIFICATION | 15 A 4 C C | CONTRACTOR CONTRACTOR | MATURE | DECLE DE CECTION | BECLE DE CALCUI |
|---------------------------|--|------------|-----------------------|-------------|----------------------|--------------------------------|
| | | TYPE | LONGUEUR | NATURE | REGLE DE GESTION | REGLE DE CALCUL |
| Id_bien | ID dans la base de données (PK) | Integer | NC | | Ne doit pas être nul | |
| No_voie | Numéro des rues | Integer | NC | Elémentaire | | |
| B/T/Q | Bis, Ter, | Varchar | 1 | Elémentaire | | |
| Type_de_voie | Plusieurs valeurs (rue, avenue, chemin, etc.) | Varchar | 4 | Elémentaire | | |
| Voie | Nom de la rue | Varchar | 50 | Elémentaire | | |
| Code departement | Le code du département (01,02,03,etc) | Integer | | Elémentaire | Ne doit pas être nul | |
| Code commune | Code de la commune | Integer | | Elémentaire | Ne doit pas être nul | |
| Code insee | Code insee = Code dep suivi du code commune (FK) | Varchar | | Concatener | Ne doit pas être nul | Code dep suivi du code commune |
| Code postal | Code postal de la commune | Varchar | 5 | Elémentaire | Ne doit pas être nul | |
| Commune | Nom de la commune du bien | Varchar | 50 | Elémentaire | Ne doit pas être nul | |
| Valeur fonciere | Valeur fonciere du bien | Float | | Elémentaire | | |
| Type local | Appartement ou maison | Varchar | 11 | Elémentaire | Ne doit pas être nul | |
| Surface reelle bati | Surface du bien | Integer | | Elémentaire | | |
| Surface Carrez du 1er lot | Surface du bien loi carrez | Float | | Elémentaire | | |
| Nombre pieces principales | Nombre pieces principales du bien | Integer | | Elémentaire | | |
| Surface terrain | Surface du terrain en m² | Integer | | Elémentaire | | |
| Date mutation | Date de la vente du bien la plus récente | Date | | Elémentaire | aaaa/mm/jj | |

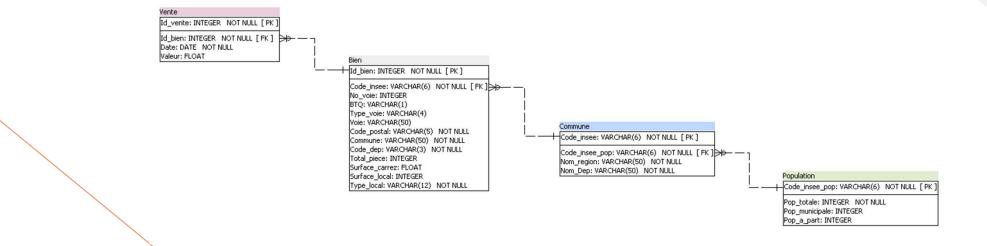
DICTIONNAIRE DES DONNÉES - Référentiel géographique

| SIGNIFICATION | TYPE | LONGUEUR | NATURE | REGLE DE GESTION | REGLE DE CALCUL |
|--|--|--|---|--|---|
| Nom de la région | Varchar | 50 | Elémentaire | Ne doit pas être nul | |
| Nom du département | Varchar | 50 | Elémentaire | Ne doit pas être nul | |
| Code insee, code unique par commune (PK) | Varchar | | Elémentaire | Ne doit pas être nul | |
| duplicatat de la colonne com_code | Varchar | | concatener | Ne doit pas être nul | Colonne com_code |
| Nom de la commune | Varchar | 50 | Elémentaire | Ne doit pas être nul | |
| Coord GPS de la commune | Float | | Elémentaire | | |
| | Nom de la région Nom du département Code insee, code unique par commune (PK) duplicatat de la colonne com_code Nom de la commune | Nom de la région Varchar Nom du département Varchar Code insee, code unique par commune (PK) Varchar duplicatat de la colonne com_code Varchar Nom de la commune Varchar | Nom de la région Varchar 50 Nom du département Varchar 50 Code insee, code unique par commune (PK) Varchar duplicatat de la colonne com_code Varchar Nom de la commune Varchar 50 | Nom de la région Varchar 50 Elémentaire Nom du département Varchar 50 Elémentaire Code insee, code unique par commune (PK) Varchar Elémentaire duplicatat de la colonne com_code Varchar concatener Nom de la commune Varchar 50 Elémentaire | Nom de la régionVarchar50ElémentaireNe doit pas être nulNom du départementVarchar50ElémentaireNe doit pas être nulCode insee, code unique par commune (PK)VarcharElémentaireNe doit pas être nulduplicatat de la colonne com_codeVarcharconcatenerNe doit pas être nulNom de la communeVarchar50ElémentaireNe doit pas être nul |

DICTIONNAIRE DES DONNÉES - Données communes

| CODE | SIGNIFICATION | TYPE | LONGUEUR | NATURE | REGLE DE GESTION | REGLE DE CALCUL |
|-------------|---|---------|----------|-------------|----------------------|------------------------|
| СОМ | Nom de la commune | Varchar | 50 | Elémentaire | Ne doit pas être nul | |
| CODDEP | Code du département de la commune | Integer | | Elémentaire | Ne doit pas être nul | |
| CODCOM | Code de la commune | Integer | | Elémentaire | Ne doit pas être nul | |
| CODINSEEPOP | Code insee = CODDEP suivi du CODDOM (PK) | Varchar | | Concatener | Ne doit pas être nul | CODDEP suivi du CODDOM |
| PMUN | Pop municipale = personnes habitant dans la commune | Integer | | Elémentaire | | |
| PCAP | Pop comptée à part = personnes recensées ailleurs mais qui ont une résidence sur la commune | Integer | | Elémentaire | | |
| РТОТ | Population totale de la commune | Integer | | Elémentaire | Ne doit pas être nul | |

Le schéma relationnel normalisé

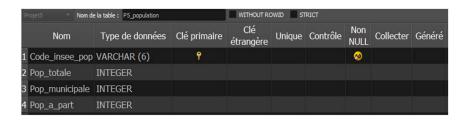


La base de données avec les tables créées et les données chargées





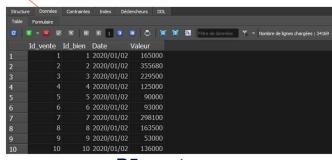




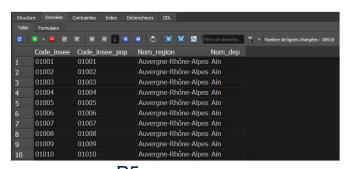
Screenshot qui permettent de démontrer le bon chargement des données

| Table | Formulaire | | | | | | | | | | |
|----------|------------|----------------|---------------|--------|--------------------------------|--------------------|----------------------|----------|-------------|------------------------------|-------------|
| 2 | = | ⊠ × F | e e 1 3 3 | 101 | 🕱 🕱 🛂 Fibre de données 🦙 🕶 | Nombre de lignes c | hargées : 34169 | | | | |
| | Id_bien | Code_inse | e No_voiє BTQ | Type_v | v Voie | Code_pc | ost Commune | Code_dep | Total_piece | Surface_carrez Surface_local | Type_local |
| 1 | | 01103 | 347 | RUE | DU CHATEAU | 1170 | CHEVRY | 01 | 3 | 48 48,22 | Appartement |
| 2 | | 06004 | | BD | EDOUARD BAUDOIN | 6160 | ANTIBES | 06 | 1 | 40 39,11 | Appartement |
| | | 06088 | 20 B | RUE | MARCEAU | 6000 | NICE | 06 | 3 | 82 80,25 | Appartement |
| 8 | | 06123 | 550 | RTE | DES VESPINS RN7 | 6700 | SAINT LAURENT DU VAR | 06 | 1 | 27 27,51 | Appartement |
| | | 13005 | 9300 | RES | LES ARPEGES BD DES ABA | 13400 | AUBAGNE | 13 | 2 | 47 47,33 | Appartement |
| | | 13028 | 27 | RUE | DU GRAND MADIER | 13600 | LA CIOTAT | 13 | 1 | 24 25,31 | Appartement |
| | | 13208 | 360 | AV | DU PRADO | 13008 | MARSEILLE 8EME | 13 | 3 | 70 70,84 | Appartement |
| | 8 | 13212 | 5076 F | PARC | DESSUARD | 13012 | MARSEILLE 12EME | 13 | 3 | 66 67,19 | Appartement |
| | | 14338 | 1194 | RUE | DE NORMANDIE | 14510 | HOULGATE | 14 | 1 | 19 18,89 | Appartement |
| 0 | 10 | 14366 | 30 | ALL | DES NOISETIERS | 14100 | LISIEUX | 14 | 4 | 99 105,37 | Maison |

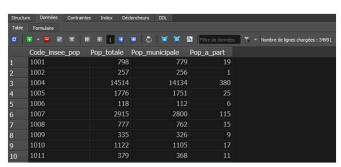
P5_bien



P5_vente



P5_commune



P5_population



Requêtes SQL et résultats

Nombre total d'appartements vendus au 1er semestre 2020

```
SELECT
count(P5_vente.Id_vente) as "Nombre total d'appartements vendus au 1er semestre 2020"
FROM
P5_vente
JOIN
P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien
WHERE
substr(P5_vente.Date, 1, 4) = "2020"
AND
substr(P5_vente.Date, 6, 2) BETWEEN "01" AND "06"
AND
P5_bien.Type_local = "Appartement"
```



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

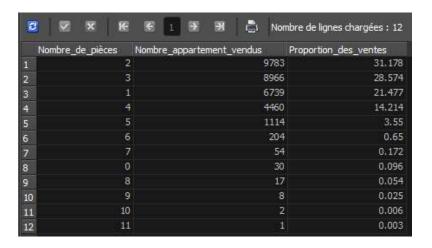
SELECT P5_commune.Nom_region, count(P5_vente.Id_vente) as "Nombre d'appartements vendus au 1er semestre 2020" FROM P5_vente JOIN P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien, P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee WHERE substr(P5 vente.Date, 1, 4) = "2020" substr(P5_vente.Date, 6, 2) BETWEEN "01" AND "06" AND P5_bien.Type_local = "Appartement" P5 commune.Nom region ORDER BY count(P5_vente.Id_vente) DESC

| | Nom_region | Nombre d'appartements vendus au 1er semestre 2020 | |
|----|----------------------------|---|-------|
| 1 | Ile-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 |

Proportion des ventes d'appartements par le nombre de pièces

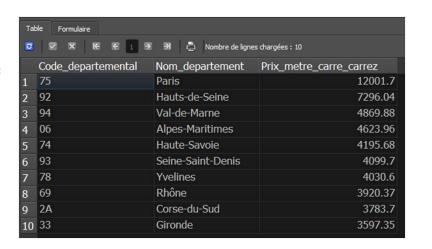
Proportion des ventes DESC

```
SELECT
P5_bien.Total_piece as "Nombre_de_pièces",
count(P5_vente.Id_vente) as "Nombre_appartement_vendus",
ROUND(count(P5_vente.ld_vente) / CAST((SELECT count(P5_vente.ld_vente) FROM P5_vente
JOIN P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien WHERE P5_bien.Type_local =
"Appartement") AS FLOAT) *100 ,3) as Proportion_des_ventes
FROM
P5_vente
JOIN
P5 bien ON P5 vente.ld bien = P5 bien.ld bien
P5_bien.Type_local = "Appartement"
GROUP BY
P5 bien.Total piece
ORDER BY
```



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

SELECT
P5_bien.Code_dep as Code_departemental,
P5_commune.Nom_dep as Nom_departement,
ROUND(SUM(P5_vente.Valeur) / SUM(P5_bien.Surface_carrez),2) as Prix_metre_carrez
FROM
P5_vente
LEFT JOIN
P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien
JOIN
P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee
GROUP BY
P5_commune.Nom_dep
ORDER BY
Prix_metre_carre_carrez DESC
LIMIT 10.



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

SELECT

ROUND(SUM(P5_vente.Valeur) / SUM(P5_bien.Surface_carrez),2) as "Prix moyen du mètre carré
carrez d'une maison en Ile-de-France"

FROM

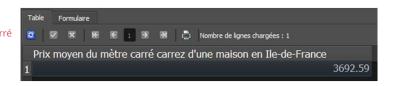
P5_vente
LEFT JOIN

P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien
JOIN

P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee
WHERE

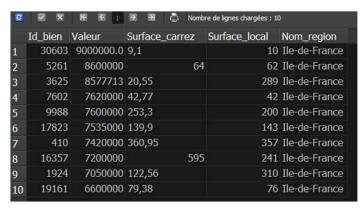
P5_commune.Nom_region = "Ile-de-France"
AND

P5_bien.Type_local = "Maison"



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

SELECT
P5_vente.Id_bien,
P5_vente.Valeur,
P5_bien.Surface_carrez,
P5_bien.Surface_local,
P5_commune.Nom_region
FROM
P5_vente
LEFT JOIN
P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien
JOIN
P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee
ORDER BY
CAST(P5_vente.Valeur AS FLOAT) DESC
LIMIT 10.



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

```
WITH vente2 AS (
SELECT
count(P5_vente.Id_vente) AS ventes_2em_trim
FROM
P5 vente
WHERE
substr(P5 vente.Date, 1, 4) = "2020"
substr(P5_vente.Date, 6, 2) BETWEEN "04" AND "06"
SELECT
count(P5_vente.Id_vente) AS "ventes_1er_trim",
ROUND(CAST(ventes_2em_trim AS FLOAT)/ CAST(count(P5_vente.Id_vente) AS FLOAT),3) AS
"Evolution(%)"
FROM
P5_vente, vente2
WHERE
substr(P5_vente.Date, 1, 4) = "2020"
substr(P5_vente.Date, 6, 2) BETWEEN "01" AND "03"
```



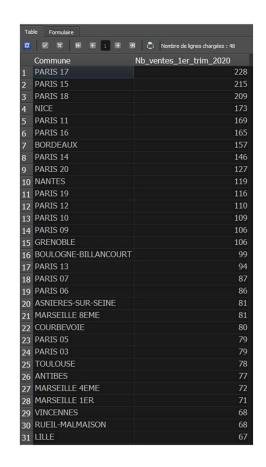
Le classement des régions par rapport au prix au mètre carré des appartement de plus de 4 pièces

```
SELECT
P5_commune.Nom_region,
ROUND(SUM(P5_vente.Valeur) / SUM(P5_bien.Surface_carrez),2) as
Prix_mètre_carré_appartement
FROM
P5_vente
LEFT JOIN
P5_bien ON P5_vente.ld_bien = P5_bien.ld_bien
P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee
WHERE
P5_bien.Type_local = "Appartement"
AND
P5_bien.Total_piece > "4"
GROUP BY
P5_commune.Nom_region
ORDER BY
Prix_mètre_carré_appartement DESC
```

| Tat | le Formulaire | |
|-----|----------------------------|--------------------------------|
| C | | Nombre de lignes chargées : 15 |
| | Nom_region | Prix_mètre_carré_appartement |
| 1 | Ile-de-France | 9205.05 |
| 2 | La Réunion | 3484.52 |
| 3 | Corse | 3458.45 |
| 4 | Provence-Alpes-Côte d'Azur | 3179.85 |
| 5 | Auvergne-Rhône-Alpes | 2943.1 |
| 6 | Nouvelle-Aquitaine | 2556 |
| 7 | Pays de la Loire | 2506.48 |
| 8 | Bretagne | 2336.68 |
| 9 | Hauts-de-France | 2289.74 |
| 10 | Occitanie | 2192.33 |
| 11 | Normandie | 2129.83 |
| 12 | Grand Est | 1481.82 |
| 13 | Centre-Val de Loire | 1466.03 |
| 14 | Bourgogne-Franche-Comté | 1131.03 |
| 15 | Martinique | 574.77 |

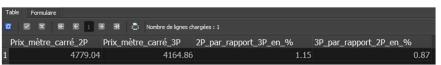
Liste des communes ayant eu au moins 50 ventes au 1er trimestre

SELECT
P5_bien.Commune,
COUNT(P5_vente.Id_vente) as Nb_ventes_1er_trim_2020
FROM
P5_vente
JOIN
P5_bien ON P5_vente.Id_bien = P5_bien.Id_bien
WHERE
substr(P5_vente.Date, 1, 4) = "2020"
AND
substr(P5_vente.Date, 6, 2) BETWEEN "01" AND "03"
GROUP BY
P5_bien.Code_insee
HAVING
COUNT(*) >= 50
ORDER BY
Nb_ventes_1er_trim_2020 DESC



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

```
WITH troispieces AS (
SELECT
ROUND(SUM(P5 vente.Valeur) / SUM(P5 bien.Surface carrez),2) as Prix mètre carré 3P
FROM
P5 vente
LEFT JOIN
P5_bien ON P5_vente.ld_bien = P5_bien.ld_bien
P5_bien.Type_local = "Appartement"
AND
P5_bien.Total_piece = "3"
SELECT
ROUND(SUM(P5_vente.Valeur) / SUM(P5_bien.Surface_carrez),2) as "Prix_mètre_carré_2P",
Prix_mètre_carré_3P,
ROUND((SUM(P5_vente.Valeur) / SUM(P5_bien.Surface_carrez)) / Prix_mètre_carré_3P,2) AS
"2P_par_rapport_3P_en_%",
ROUND(Prix_mètre_carré_3P / (SUM(P5_vente.Valeur) / SUM(P5_bien.Surface_carrez)),2) AS
"3P_par_rapport_2P_en_%"
FROM
P5_vente, troispieces
LEFT JOIN
P5_bien ON P5_vente.ld_bien = P5_bien.ld_bien
P5_bien.Type_local = "Appartement"
P5 bien.Total piece = "2"
```



WHERE Top_commune <= 3

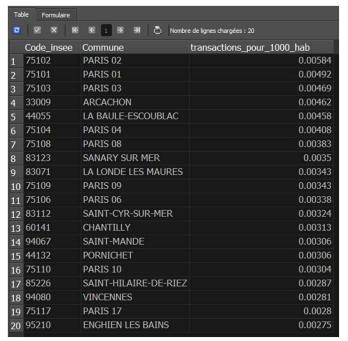
Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69

```
WITH rang AS (
SELECT
P5_bien.Code_dep,
P5_commune.Nom_dep,
P5_bien.Commune,
ROUND(AVG(P5_vente.Valeur), 2) AS moy_val_fonciere,
ROW_NUMBER() OVER (PARTITION BY P5_bien.Code_dep ORDER BY AVG(P5_vente.Valeur) DESC)
AS Top_commune
FROM
P5_vente
LEFT JOIN
P5_bien ON P5_vente.ld_bien = P5_bien.ld_bien
JOIN
P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee
P5_bien.Code_dep IN ("06", "13", "33", "69", "59")
GROUP BY
P5_bien.Code_dep, P5_bien.Code_insee
SELECT *
FROM rang
```

| Tab | Table Formulaire | | | | | | | |
|-----|------------------|------------------|--------------------------------|------------------|-------------|--|--|--|
| 2 | ⊠ × | E E 1 9 91 | Nombre de lignes chargées : 15 | | | | | |
| | Code_dep | Nom_dep | Commune | moy_val_fonciere | Top_commune | | | |
| 1 | 06 | Alpes-Maritimes | SAINT-JEAN-CAP-FERRAT | 968750 | 1 | | | |
| 2 | 06 | Alpes-Maritimes | EZE | 655000 | 2 | | | |
| 3 | 06 | Alpes-Maritimes | MOUANS-SARTOUX | 476898 | 3 | | | |
| 4 | 13 | Bouches-du-Rhône | GIGNAC-LA-NERTHE | 330000 | 1 | | | |
| 5 | 13 | Bouches-du-Rhône | SAINT SAVOURNIN | 314425 | 2 | | | |
| 6 | 13 | Bouches-du-Rhône | CASSIS | 313416.88 | 3 | | | |
| 7 | 33 | Gironde | LEGE-CAP-FERRET | 549500.64 | 1 | | | |
| 8 | 33 | Gironde | VAYRES | 335000 | 2 | | | |
| 9 | 33 | Gironde | ARCACHON | 307435.93 | 3 | | | |
| 10 | 59 | Nord | BERSEE | 433202 | 1 | | | |
| 11 | 59 | Nord | CYSOING | 408550 | 2 | | | |
| 12 | 59 | Nord | HALLUIN | 322250 | 3 | | | |
| 13 | 69 | Rhône | VILLE SUR JARNIOUX | 485300 | 1 | | | |
| 14 | 69 | Rhône | LYON 2EME | 455217.26 | 2 | | | |
| 15 | 69 | Rhône | LYON 6EME | 426968.25 | 3 | | | |

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

```
SELECT
P5 bien.Code insee,
P5 bien.Commune,
Round(count(P5_vente.ld_vente) / CAST(P5_population.Pop_totale AS FLOAT), 5) as
transactions pour 1000 hab
FROM
P5 vente
LEFT JOIN P5_bien ON P5_vente.ld_bien = P5_bien.ld_bien
JOIN P5_commune ON P5_bien.Code_insee = P5_commune.Code_insee,
P5_population ON P5_commune.Code_insee_pop = P5_population.Code_insee_pop
WHERE
P5_population.Pop_totale > "10000"
GROUP BY
P5 bien.Code insee
ORDER BY
transactions_pour_1000_hab DESC
LIMIT 20
```





Merci!