

Projet P3- DataImmo

CRÉEZ ET UTILISEZ UNE BASE DE
DONNÉES IMMOBILIÈRE AVEC SQL



PARCOURS DATA ANALYST_V2

ADELIN LE RAY

Sommaire

Slide	
3	Création de la Base de Données : démarche
11	La base de Données « DataImmo » dans MySQL
13	Réponses aux analyses demandées

« DataImmo » : un modèle pour mieux prévoir le prix de vente des biens immobiliers

Création de la base de données

PROJET P3 – BASE DE DONNEES SQL



Données Open Source – INSEE

- Recherche sur le site de l'INSEE des données pertinentes pour région et population
- Ajout des données :
 - Région
 - Population (année 2019)

Téléchargement du fichier d'ensemble des populations légales en 2019

Recensement de la population

CHIFFRES DÉT

Paru le : 29/12/2019

> Imprimer

TÉLÉCHARGEMENT

DOCUMENTATION

Téléchargement du fichier d'ensemble

Les fichiers csv sont encodés en UTF-8, correspondant aux standards d'interopérabilité. Pour ouvrir correctement les fichiers de métadonnées avec le logiciel Microsoft Excel, il est nécessaire d'utiliser la fonctionnalité d'import de données externes accessible depuis le menu Données.

France hors Mayotte

(pdf, 12 Mo)



(xlsx, 2 Mo)



(csv, 971 Ko)



<https://www.insee.fr/fr/statistiques/6011070>

Décès de 2014 à 2021

Bulletin état civil

Résultats pour les communes de France (y compris Mayotte)

BASES DE

Paru le : 2

> Imprimer

Les statistiques annuelles concernent les décès domiciliés ainsi que les jugements déclaratifs de décès. Le lieu retenu pour le décès est le domicile du défunt et non pas le lieu du décès.

Niveau géographique : commune, arrondissement municipal

TÉLÉCHARGEMENT

DOCUMENTATION

DICTIONNAIRE DES VARIABLES

Toutes les statistiques sont proposées dans la géographie communale en vigueur au 1^{er} janvier 2022.

À partir du millésime 2019, le champ a été élargi au département de Mayotte. Dorénavant, les statistiques portent sur les événements survenus en France pour toutes les communes de métropole et les départements d'outre-mer, y compris le département de Mayotte.

Pour les communes de Lyon et Marseille, l'arrondissement municipal du domicile du défunt n'est pas toujours renseigné sur le bulletin de décès. Ces décès ne sont comptabilisés dans les bases téléchargeables qu'au niveau de la commune et non de l'arrondissement municipal.

Décès de 2014 à 2021

(pdf, 2 Mo)



(xlsx, 608 Ko)



<https://www.insee.fr/fr/statistiques/1893253>

donnees_arrondissements
donnees_cantons
donnees_collectivites_outre_mer
donnees_communes
donnees_communes_deleguees
donnees_departements
donnees_fractions_cantonales
donnees_regions
metadonnees_arrondissements
metadonnees_cantons
metadonnees_collectivites_outre_mer
metadonnees_communes
metadonnees_communes_deleguees
metadonnees_departements
metadonnees_fractions_cantonales
metadonnees_regions

Naissances de 2014 à 2021

Bulletin état civil

Résultats pour les communes de France (y compris Mayotte)

Paru le : 2

> Imprimer

Les statistiques annuelles concernent les enfants nés vivants ainsi que les jugements déclaratifs de naissance. Le lieu retenu pour la naissance est le domicile de la mère et non pas le lieu de naissance.

Les statistiques sont dressées à partir des bulletins statistiques de l'état civil établis par les maires, au moment et dans la commune où ont lieu les naissances et des transcriptions des jugements déclaratifs de naissance établis par les tribunaux.

À partir du millésime 2019, le champ a été élargi au département de Mayotte. Dorénavant, les statistiques portent sur les événements survenus en France pour toutes les communes de métropole et les départements d'outre-mer, y compris le département de Mayotte.

Niveau géographique : commune, arrondissement municipal

TÉLÉCHARGEMENT

DOCUMENTATION

DICTIONNAIRE DES VARIABLES

Toutes les statistiques sont proposées dans la géographie communale en vigueur au 1^{er} janvier 2022.

Pour les communes de Lyon et Marseille, l'arrondissement municipal du domicile de la mère n'est pas toujours renseigné sur le bulletin de naissance. Ces naissances ne sont comptabilisées dans les bases téléchargeables qu'au niveau de la commune et non de l'arrondissement municipal.

Naissances de 2014 à 2021

(pdf, 2 Mo)



(xlsx, 609 Ko)



<https://www.insee.fr/fr/statistiques/1893255>



Le Dictionnaire de données

Répertorie et décrit les données importantes à stocker :

- Code et Signification
- Type, Longueur et Nature
- Règles de gestion et calculs



Parcours Data Analyst V2
Projet P3

DICTIONNAIRE DE DONNEES

CODE	SIGNIFICATION	TYPE A/AN/N	LONGUEUR	NATURE E/Ca/CO	REGLE DE GESTION	REGLE DE CALCUL
ID_BIEN	Identifiant du bien	N	10	E	Ne doit pas être null	
NO_VOIE	Numéro de voie	N	4	E		
ADR_COMPL	Complément d'adresse : Bis, Ter, quarter, etc. ...	AN	1	E		
TYPE_VOIE	Type de voie (Rue, boulevard, ...)	A	4	E		
VOIE	Libellé de la voie	A	50	E	Ne doit pas être null	
SURF_CARREZ	Surface suivant la loi Carrez du 1er lot. Un lot de copropriété = partie privative + quote-part de partie commune (tantièmes)	N	10	E	Ne doit pas être null	
SURF_RB	Surface réelle bâtie (local + dépendances)	N	5	E	Ne doit pas être null	
SURF_TERRAIN	Surface du terrain	N	10	E		
TYPE_LOCAL	Libellé du type de local : appartement, maison, dépendance (isolée), local industriel ou commercial ou assimilé	A	50	E	Ne doit pas être null	
TOT_PIECES	Nombre de pièces principales	N	2	E	Ne doit pas être null	
ID_COM	Identifiant de la commune	N	5	E	Ne doit pas être null	
CODPOS	Code postal de la commune	N	5	E	Ne doit pas être null	
COM	Nom de la commune	A	50	E	Ne doit pas être null	
ID_VENTE	Identifiant de la vente / mutation	N	5	E	Ne doit pas être null	
DATE_VENTE	Date de signature de l'acte de vente / mutation	N	8	E	Ne doit pas être null AAAA-MM-JJ	
VALEUR	Valeur foncière du bien TTC = prix net vendeur, hors frais de notaire	N	7	E	Ne doit pas être null	
ID_REG	Identifiant de la région	N	3	E	Ne doit pas être null	
CODEREG	Code de la région	N	2	E	Ne doit pas être null	
REG	Nom actuel de la région	A	26	E	Ne doit pas être null	
DEP	Nom du département	A	23	E	Ne doit pas être null	
ID_POP	Identifiant de la table population Code commune INSEE	N	5	E	Ne doit pas être null	
ANNEE	Année du recensement population	N	4	E	Ne doit pas être null AAAA	
POP	Population totale de la commune	N	10	E	Ne doit pas être null	
NAIS	Nombre de naissances domiciliées dans la commune	N	5	E	Ne doit pas être null	
DECES	Nombre de décès domiciliés dans la commune	N	5	E	Ne doit pas être null	

A : Alphabétique
N : Numérique
AN : Alphanumérique

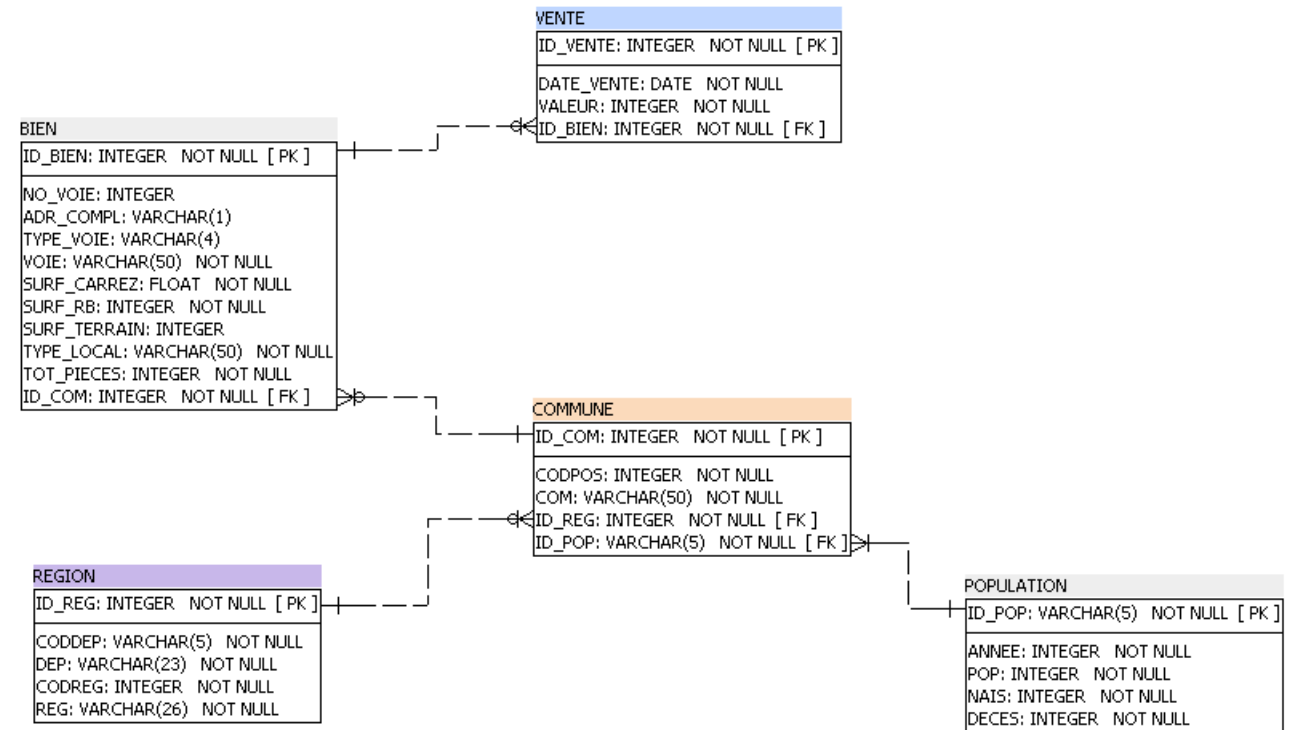
E : Élémentaire
Ca : Calculer
Co : Concatène

■ ■ ■ ■ ■ Modèle Logique de Données (MLD)

Modification du schéma relationnel
existant pour ajouter les données de
région et de population

→ 5 tables identifiées

SQL POWER
ARCHITECT





Préparation des tables (1/2)

- Nettoyage des données de la table DAN_P3 dans Power Query et Excel
- Suivi des modifications dans un fichier de change log (voir ci-contre)



TABLE DAN_P3

ACTIONS	DESCRIPTION
Colonnes vides supprimées	- Code service CH, Reference document, 1 Articles CGI, 2 Articles CGI, 3 Articles CGI, 4 Articles CGI, 5 Articles CGI -No volume - 2eme lot, Surface Carrez du 2eme lot, 3eme lot, Surface Carrez du 3eme lot, 4eme lot, Surface Carrez du 4eme lot, 5eme lot, Surface Carrez du 5eme lot - Identifiant local
Informations redondantes	- code type de voie vs Type de voie - Code voie vs no de voie - code type local vs type de local
Informations non retenues	- Nature mutation (que des ventes), no disposition - Nature culture, Nature culture speciale
Valeurs modifiées	- Code postal : XXXXX.0 -> XXXXX - Valeur foncière : arrondi à la valeur supérieure pour supprimer les décimales - Surface carrez : "," au lieu de "." pour le type décimal (changement des paramètres dans Options avancées) - ID de commune démarre à 1 au lieu de 0 (old_ID +1 = new_ID)
Valeurs NULL	- valeur_fonciere : lignes supprimées (18 enregistrements)
Valeurs aberrantes -> supprimées	- valeur foncière 9millions pour appart de 10m² (1 enregistrement) - surface_carrez < 1m² (3 enregistrements) - Total_piece = 0 (33 enregistrements) dont commune 2078 qui ne comportait qu'un seul bien
Valeurs erronées	Code postal manquant pour commune id 658 "nan" -> complété "20090"
Type de données modifié	- Date de mutation : AAAA-MM-JJ 00:00:00-> AAAA-MM-JJ (paramètres régionaux de DATE en "Texte", "Anglais (Canada)" dans Power Query pour date au format AAAA-MM-JJ) - Code departement : texte

Préparation des Tables (2/2)

- Jointure dans Power Query (Foreign Key)
- Export des fichiers en .csv



TABLE BIEN

ACTIONS	DESCRIPTION
Recherche de doublons sur les biens	Avec les colonnes : NO_VOIE, ADRESS_COMPL, TYPE_VOIE, VOIE, SURF_CARREZ, SURF_RB, SURF_TERRAIN, TYPE_LOCAL, TOT_PIECES, ID_COM


JOINTURES

ACTIONS	DESCRIPTION
Jointure des tables commune et population	<p>Le lien est le code commune INSEE : CODGEO (VARCHAR 5)= CODDEP (VARCHAR 2)+CODCOM(INTEGER 3) Exception : cas des dom-tom, code dep sur 3 chiffres 971.</p> <p>Formule CODDEP : =SI(NBCAR([@CODDEP])=3;GAUCHE([@CODDEP];2);CONCATENER(REPT("0";2-NBCAR([@CODDEP]));[@CODDEP]))</p> <p>Formule CODCOM : =CONCATENER(REPT("0";3-NBCAR([@CODCOM]));[@CODCOM])</p>
Jointure entre les tables	Utilisation de "Fusionner les requêtes " sur Power Query pour s'assurer de la correspondance des tables



Création des tables

SGBDR retenue : My SQL
Workbench

Récupération du code MySQL à
partir du MLD dans SQL Power
Architect 

Preview SQL Script

Your Target Database is not configured.

```
CREATE TABLE POPULATION (  
  ID_POP VARCHAR(5) NOT NULL,  
  ANNEE INT NOT NULL,  
  POP INT NOT NULL,  
  NAIS INT NOT NULL,  
  DECES INT NOT NULL,  
  PRIMARY KEY (ID_POP)  
);  
  
CREATE TABLE REGION (  
  ID_REG INT NOT NULL,  
  CODDEP VARCHAR(5) NOT NULL,  
  DEP VARCHAR(23) NOT NULL,  
  CODREG INT NOT NULL,  
  REG VARCHAR(26) NOT NULL,  
  PRIMARY KEY (ID_REG)  
);  
  
CREATE TABLE COMMUNE (  
  ID_COM INT NOT NULL,  
  CODPOS INT NOT NULL,  
  COM VARCHAR(50) NOT NULL,  
  ID_REG INT NOT NULL,  
  ID_POP VARCHAR(5) NOT NULL,  
  PRIMARY KEY (ID_COM)  
);  
  
CREATE TABLE BIEN (  
  ID_BIEN INT NOT NULL,  
  NO_VOIE INT,  
  ADR_COMPL VARCHAR(1),  
  TYPE_VOIE VARCHAR(4),  
  VOIE VARCHAR(50) NOT NULL,  
  SURF_CARREZ DOUBLE PRECISIONS NOT NULL,  
  SURF_RB INT NOT NULL,  
  SURF_TERRAIN INT,  
  TYPE_LOCAL VARCHAR(50) NOT NULL,  
  TOT_PIECES INT NOT NULL,  
  ID_COM INT NOT NULL,  
  PRIMARY KEY (ID_BIEN)  
);  
  
CREATE TABLE VENTE (  
  ID_VENTE INT NOT NULL,  
  DATE_VENTE DATE NOT NULL,  
  VALEUR INT NOT NULL,  
  ID_BIEN INT NOT NULL,  
  PRIMARY KEY (ID_VENTE)  
);
```

Copy Execute Save Close

DAN_V2-P3

Preview SQL Script

Your Target Database is not configured.

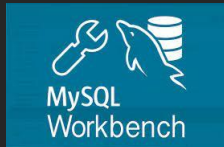
```
CREATE TABLE VENTE (  
  ID_VENTE INT NOT NULL,  
  DATE_VENTE DATE NOT NULL,  
  VALEUR INT NOT NULL,  
  ID_BIEN INT NOT NULL,  
  PRIMARY KEY (ID_VENTE)  
);  
  
ALTER TABLE COMMUNE ADD CONSTRAINT population_commune_fk  
FOREIGN KEY (ID_POP)  
REFERENCES POPULATION (ID_POP)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION;  
  
ALTER TABLE COMMUNE ADD CONSTRAINT region_commune_fk  
FOREIGN KEY (ID_REG)  
REFERENCES REGION (ID_REG)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION;  
  
ALTER TABLE BIEN ADD CONSTRAINT commune_bien_fk  
FOREIGN KEY (ID_COM)  
REFERENCES COMMUNE (ID_COM)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION;  
  
ALTER TABLE VENTE ADD CONSTRAINT bien_vente_fk  
FOREIGN KEY (ID_BIEN)  
REFERENCES BIEN (ID_BIEN)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION;
```

Copy Execute Save Close



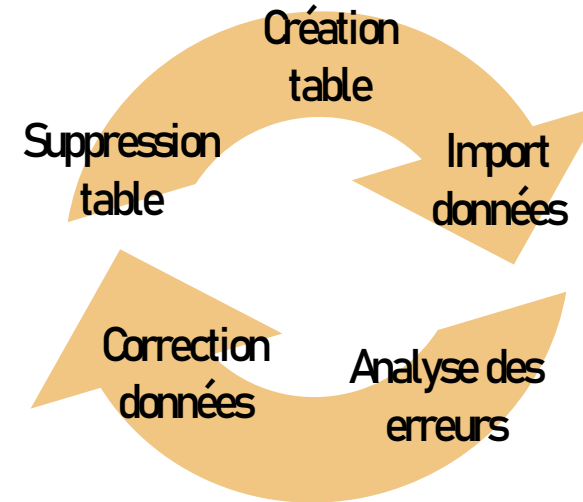
! Importation des tables

Exécution du code de création des tables par étapes pour permettre le chargement des données sans les contraintes de Foreign key



- ETAPE 1 :
- Création de la table puis chargement des données
 - Ordre : 1-region, 2-population, 3-commune, 4-bien, 5-vente
 - Vérification nombre de lignes importées / tables du fichiers .csv

➡ Processus par itérations



FICHIER	ERREURS	ERREURS	SOLUTIONS
1-TABLE_REGIONS	Row import failed with error: ("Data too long for column 'REG' at row 1", 1406) Row import failed with error: ("Data too long for column 'DEP' at row 1", 1406)	1406	Enregistrement du fichier csv en Export .csv et non .csv utf8
2-TABLE_POPULATION	RAS	-	-
3-TABLE_COMMUNES	Row import failed with error: ("Cannot add or update a child row: a foreign key constraint fails ('dan_p3'.commune', CONSTRAINT 'region_commune_fk' FOREIGN KEY ('ID_REG') REFERENCES 'region' ('ID_REG'))", 1452)	1452	ok avec corrections erreurs table_region
4-TABLE_BIENS			NO VOIE et SHIRE TERRAIN : cellules vides

ETAPE 2 : Ajout des contraintes des clés étrangères

#	Time	Action	Message
✓ 1	14:04:20	ALTER TABLE COMMUNE ADD CONSTRAINT population_commune_fk FOREIGN KEY (ID_POP) REFERENCES ...	3213 row(s) affected Records: 3213 Duplicates: 0 Warnings: 0
✓ 2	14:04:20	ALTER TABLE COMMUNE ADD CONSTRAINT region_commune_fk FOREIGN KEY (ID_REG) REFERENCES ...	3213 row(s) affected Records: 3213 Duplicates: 0 Warnings: 0
✓ 3	14:04:20	ALTER TABLE BIEN ADD CONSTRAINT commune_bien_fk FOREIGN KEY (ID_COM) REFERENCES COMMU...	34088 row(s) affected Records: 34088 Duplicates: 0 Warnings: 0
✓ 4	14:04:20	ALTER TABLE VENTE ADD CONSTRAINT bien_vente_fk FOREIGN KEY (ID_BIEN) REFERENCES BIEN (ID_...	34112 row(s) affected Records: 34112 Duplicates: 0 Warnings: 0



Présentation de la base de données dans MySQL Workbench

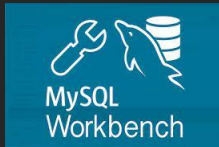
PROJET P3 – BASE DE DONNEES SQL



Présentation de la base de données



Go to



MySQL Workbench

db x

File Edit View Query Database Server Tools Scripting Help

Navigator

SQL File 3*

1 • SELECT * FROM bien

2 ;

Limit to 1000 rows

Result Grid

	ID_BIEN	NO_VOIE	ADR_COMPL	TYPE_VOIE	VOIE	SURF_CARREZ	SURF_RB	SURF_TERRAIN	TYPE_LOCAL	TOT_PIECES	ID_COM
1		NULL		RES	L OURS BLANC	24.8	26	NULL	Appartement	2	1098
2		NULL		VGE	ARTS ET VIE	51.89	56	NULL	Appartement	4	878
3		NULL		RES	LES CHARMILLES	33.49	32	NULL	Appartement	1	346
4		NULL		RES	LA TOVIERE	30	31	NULL	Appartement	1	99
5		NULL		RUE	DU VIEUX VILLAGE CARNON	46.27	43	NULL	Appartement	3	891
6		NULL			LES PARQUETS	68.8	54	NULL	Appartement	2	3074
7		NULL		AV	BIOCHAUD	32.77	32	NULL	Appartement	2	2729
8		NULL		ALL	DES JACINTHES	65.08	65	NULL	Appartement	3	1674
9		NULL			LOUSTEAUNEUF	33.09	33	NULL	Appartement	2	881
10		NULL		QUAI	DE LA PECHERIE	19.65	20	NULL	Appartement	1	1175
11		NULL		RES	LES ARGONAUTES	22.63	23	NULL	Appartement	1	370
12		NULL		RES	LES COLLINES DU RICANTO	43.87	44	NULL	Appartement	2	619
13		NULL		AV	DES BALEINIERS	67.37	67	NULL	Maison	3	368
14		NULL		AV	ALFRED ANATOLE THEVENET	67.5	67	NULL	Appartement	3	1501
15		NULL			LA HAIE GRISELLE	67.48	67	NULL	Appartement	3	3045
16		NULL		RES	LES COLLINES DU RICANTO	50.73	51	NULL	Appartement	2	619

bien 3 x

Administration Schemas Information

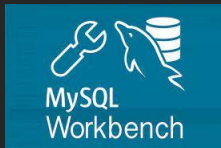
Analyse des données : Requêtes et résultats

PROJET P3 – BASE DE DONNEES SQL



Requêtes

11 analyses des données
demandées lors de la réunion de
validation du schéma relationnel



1. Nombre total d'appartements vendus au 1er semestre 2020.
2. Le nombre de ventes d'appartement par région pour le 1er semestre 2020.
3. Proportion des ventes d'appartements par le nombre de pièces.
4. Liste des 10 départements où le prix du mètre carré est le plus élevé.
5. Prix moyen du mètre carré d'une maison en Île-de-France.
6. Liste des 10 appartements les plus chers avec la région et le nombre de mètres carrés.
7. Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020
8. Le classement des régions par rapport au prix au mètre carré des appartement de plus de 4 pièces.
9. Liste des communes ayant eu au moins 50 ventes au 1er trimestre
10. Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces.
11. Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69

REQUETE 1 : Nombre total d'appartements vendus au 1er semestre 2020



Requête : Jointure JOIN, filtrage WHERE et Agrégation GROUP BY

```
1 • SELECT TYPE_LOCAL, count(*) as nb_vente_1er_semestre
2 FROM dan_p3.bien
3 JOIN dan_p3.vente USING (ID_BIEN)
4 WHERE
5     DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'
6     AND TYPE_LOCAL = 'appartement'
7 GROUP BY TYPE_LOCAL
8 ;
```



Résultat :

TYPE_LOCAL	nb_ventes_1er_semestre
Appartement	31326

REQUETE 2 : Nombre de ventes d'appartement par région pour le 1er semestre 2020



Requête : Jointure et filtrage avec WHERE, agrégation
Group BY et tri ORDER BY

```
1 • SELECT b.TYPE_LOCAL as Type_de_local, r.REG as Region, count(*) as Nb_vente_1er_semestre
2 FROM dan_p3.bien b, dan_p3.commune c, dan_p3.region r, dan_p3.vente v
3 WHERE
4     v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'
5     AND b.ID_BIEN = v.ID_BIEN
6     AND b.ID_COM = c.ID_COM
7     AND c.ID_REG = r.ID_REG
8     AND b.TYPE_LOCAL = 'appartement'
9 GROUP BY b.TYPE_LOCAL, r.REG
10 ORDER BY Nb_vente_1er_semestre DESC
11 ;
```



Résultat : 45% des ventes réalisées en IdF

Type_de_local	Region	Nb_vente_1er_semestre
Appartement	Île-de-France	13974
Appartement	Provence-Alpes-Côte d'Azur	3639
Appartement	Auvergne-Rhône-Alpes	3248
Appartement	Nouvelle-Aquitaine	1931
Appartement	Occitanie	1640
Appartement	Pays de la Loire	1356
Appartement	Hauts-de-France	1251
Appartement	Grand Est	982
Appartement	Bretagne	982
Appartement	Normandie	860
Appartement	Centre-Val de Loire	693
Appartement	Bourgogne-Franche-Comté	376
Appartement	Corse	222
Appartement	Martinique	94
Appartement	La Réunion	44
Appartement	Guyane	34

REQUETE 3 : Proportion des ventes d'appartements par le nombre de pièces



Requête : Jointure JOIN, filtrage WHERE, agrégation Group BY et tri ORDER BY

```
1 • SELECT b.TYPE_LOCAL as Type_de_local, b.TOT_PIECES as Nb_pièces,  
2       CONCAT(ROUND(count(*)*100/(select count(*)from dan_p3.vente),2),'%') as Pourcentage_vente  
3 FROM dan_p3.bien b  
4 JOIN dan_p3.commune c USING (ID_COM)  
5 JOIN dan_p3.vente v USING (ID_BIEN)  
6 WHERE  
7     v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'  
8     AND b.TYPE_LOCAL = 'appartement'  
9 GROUP BY b.TYPE_LOCAL, b.TOT_PIECES  
10 ORDER BY b.TOT_PIECES ASC  
11 ;
```



Résultat : 55% des ventes d'appartements sont des 2 et 3 pièces

	Type_de_local	Nb_pièces	Pourcentage_vente
▶	Appartement	1	19.73%
	Appartement	2	28.65%
	Appartement	3	26.28%
	Appartement	4	13.07%
	Appartement	5	3.27%
	Appartement	6	0.60%
	Appartement	7	0.16%
	Appartement	8	0.05%
	Appartement	9	0.02%
	Appartement	10	0.01%
	Appartement	11	0.00%

Nota : Nombre de pièces principales = 0 considéré comme une valeur aberrante et non retenue dans la préparation des données.

REQUETE 4 : Liste des 10 départements où le prix du mètre carré est le plus élevé



Requête : Jointure JOIN, Filtrage WHERE, Agrégation GROUP BY, Tri ORDER BY et LIMIT

```
Requete4b x
SELECT r.CODDEP as Code_departement, r.DEP as Departement,
       round(avg(v.VALEUR/b.SURF_CARREZ)) as Prix_m²_carrez,
       round(avg(v.VALEUR/b.SURF_RB)) as Prix_m²_reel_bati
FROM dan_p3.bien b
JOIN dan_p3.commune c USING (ID_COM)
JOIN dan_p3.vente v USING (ID_BIEN)
JOIN dan_p3.region r USING (ID_REG)
WHERE
    v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'
GROUP BY r.CODDEP, r.DEP
ORDER BY Prix_m²_carrez DESC
LIMIT 10
;
```



Résultat : Prix m² à Paris 62% plus cher que le 2^{ème} département.

Top 3 des départements = IDF

Code_departement	Departement	Prix_m²_carrez	Prix_m²_reel_bati
75	Paris	11716	11947
92	Hauts-de-Seine	7219	7415
94	Val-de-Marne	5343	5399
6	Alpes-Maritimes	4701	4686
74	Haute-Savoie	4667	4150
93	Seine-Saint-Denis	4343	4369
78	Yvelines	4221	4122
69	Rhône	4059	4063
2A	Corse-du-Sud	4027	3922
33	Gironde	3764	3839

REQUETE 5 : Prix moyen du mètre carré d'une maison en Île-de-France



Requête : Jointure JOIN, Filtrage WHERE, Agrégation GROUP BY

```
1 • SELECT REG as Region, TYPE_LOCAL as Type_de_local,  
2         round(avg(VALEUR/SURF_CARREZ)) as Prix_m²_carrez,  
3         round(avg(VALEUR/SURF_RB)) as Prix_m²_reel_bati  
4 FROM bien  
5 JOIN vente USING (ID_BIEN)  
6 JOIN commune USING (ID_COM)  
7 JOIN region USING (ID_REG)  
8 WHERE  
9     DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'  
10    AND TYPE_LOCAL='maison'  
11    AND CODDEP IN ('75', '77', '78', '91', '92', '93', '94', '95')  
12 GROUP BY REG, TYPE_LOCAL  
13 ;
```



Résultat : Prix moyen m² en IDF ~ 4000 €/m²

	Region	Type_de_local	Prix_m²_carrez	Prix_m²_reel_bati
►	Île-de-France	Maison	3737	3999

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



Requête : Jointure et filtrage WHERE, Tri ORDER BY et LIMIT

```
1 • SELECT r.REG as Region, c.COM as Commune, b.TYPE_LOCAL as Type_de_local, b.SURF_CARREZ as Surface_carrez,  
2         b.SURF_RB as Surface_reel_bati, v.VALEUR as Valeur_fonciere  
3 FROM dan_p3.bien b, dan_p3.commune c, dan_p3.vente v, dan_p3.region r  
4 WHERE  
5     v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'  
6     AND b.ID_BIEN = v.ID_BIEN  
7     AND b.ID_COM = c.ID_COM  
8     AND c.ID_REG = r.ID_REG  
9     AND b.TYPE_LOCAL = 'appartement'  
10 ORDER BY v.VALEUR DESC  
11 LIMIT 10  
12 ;
```



Résultat : Les appartements les + chers sont en IDF et majoritairement sur Paris (9/10)

Region	Commune	Type_de_local	Surface_carrez	Surface_reel_bati	Valeur_fonciere
Île-de-France	CORBEIL ESSONNES	Appartement	64	62	8600000
Île-de-France	PARIS 07	Appartement	20.55	289	8577713
Île-de-France	PARIS 17	Appartement	42.77	42	7620000
Île-de-France	PARIS 06	Appartement	253.3	200	7600000
Île-de-France	PARIS 01	Appartement	139.9	143	7535000
Île-de-France	PARIS 16	Appartement	360.95	357	7420000
Île-de-France	PARIS 16	Appartement	595	241	7200000
Île-de-France	PARIS 01	Appartement	122.56	310	7050000
Île-de-France	PARIS 01	Appartement	79.38	76	6600000
Île-de-France	PARIS 16	Appartement	317.05	287	6500000

Nota: Appartement ayant une valeur foncière 9millions pour une surface carrez de 10m² considéré comme valeur aberrante et supprimé lors de la préparation des données.

REQUETE 7 : Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020



Requête : Sous-requête WITH et filtrage WHERE

```
1 • WITH ventes_1er_trim_table AS (  
2     SELECT count(*) as ventes_1er_trim  
3     FROM dan_p3.vente  
4     WHERE DATE_VENTE BETWEEN '2020-01-01' AND '2020-03-31'),  
5  
6     ventes_2nd_trim_table AS (  
7     SELECT count(*) as ventes_2nd_trim  
8     FROM dan_p3.vente  
9     WHERE DATE_VENTE BETWEEN '2020-04-01' AND '2020-06-30')  
10  
11     SELECT CONCAT(round((ventes_2nd_trim -ventes_1er_trim)*100/ventes_1er_trim,2), '%') AS Taux_evolution_ventes_entre_1er_et_2nd_trim  
12     FROM ventes_1er_trim_table, ventes_2nd_trim_table  
13 ;
```



Résultat : Augmentation des ventes

	Taux_evolution_ventes_entre_1er_et_2nd_trim
▶	3.65%

REQUETE 8 : Le classement des régions par rapport au prix au mètre carré des appartement de plus de 4 pièces



Requête : Jointure JOIN, filtrage WHERE, agrégation GROUP BY, tri ORDER BY

```
1 • SELECT REG as Region, TYPE_LOCAL as Type_de_local,
2         round(avg(VALEUR/SURF_CARREZ)) as Prix_m²_carrez,
3         round(avg(VALEUR/SURF_RB)) as Prix_m²_reel_bati
4 FROM bien
5 JOIN vente USING (ID_BIEN)
6 JOIN commune USING (ID_COM)
7 JOIN region USING (ID_REG)
8 WHERE
9     DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'
10    AND TYPE_LOCAL='appartement'
11    AND TOT_PIECES > 4
12 GROUP BY REG, TYPE_LOCAL
13 ORDER BY Prix_m²_carrez DESC ;
```



Résultat : Prix/m² (loi Carrez) le plus élevé en IDF

Region	Type_de_local	Prix_m²_carrez	Prix_m²_reel_bati
Île-de-France	Appartement	8770	8015
La Réunion	Appartement	3642	3660
Provence-Alpes-Côte d'Azur	Appartement	3588	3005
Corse	Appartement	3105	3046
Auvergne-Rhône-Alpes	Appartement	2891	2769
Nouvelle-Aquitaine	Appartement	2465	2510
Bretagne	Appartement	2412	2272
Pays de la Loire	Appartement	2316	2187
Hauts-de-France	Appartement	2190	2204
Occitanie	Appartement	2097	2096
Normandie	Appartement	2016	1994
Grand Est	Appartement	1541	1313
Centre-Val de Loire	Appartement	1453	1429
Bourgogne-Franche-Comté	Appartement	1251	1069
Martinique	Appartement	573	564

REQUETE 9 : Liste des communes ayant eu au moins 50 ventes au 1er trimestre



Requête : Jointure JOIN, filtrage WHERE, agrégation GROUP BY, filtrage HAVING, tri ORDER BY

```
1 • SELECT CODPOS as Code_Postal, COM as Commune, count(*) AS nb_ventes_1er_trim
2 FROM dan_p3.bien
3 JOIN dan_p3.vente USING (ID_BIEN)
4 JOIN dan_p3.commune USING (ID_COM)
5 WHERE
6     DATE_VENTE BETWEEN '2020-01-01' AND '2020-03-31'
7 GROUP BY CODPOS, COM
8 HAVING nb_ventes_1er_trim >=50
9 ORDER BY nb_ventes_1er_trim DESC
10 ;
```

 Résultat : Ventes majoritairement en IDF et sur Paris

	Code_Postal	Commune	nb_ventes_1er_trim
▶	75017	PARIS 17	228
	75015	PARIS 15	215
	75018	PARIS 18	209
	75011	PARIS 11	169
	75016	PARIS 16	165
	75014	PARIS 14	146
	75020	PARIS 20	127
	75019	PARIS 19	116
	75012	PARIS 12	110
	75010	PARIS 10	108
	75009	PARIS 09	106
	92100	BOULOGNE-BILLANCOURT	97
	75013	PARIS 13	94
	75007	PARIS 07	87
	75006	PARIS 06	86
	92600	ASNIERES-SUR-SEINE	81
	13008	MARSEILLE 8EME	80
	92400	COURBEVOIE	80

Code_Postal	Commune	nb_ventes_1er_trim
75003	PARIS 03	79
75005	PARIS 05	79
6000	NICE	76
38000	GRENOBLE	76
33000	BORDEAUX	73
13001	MARSEILLE 1ER	71
13004	MARSEILLE 4EME	71
92500	RUEIL-MALMAISON	68
94300	VINCENNES	68
13009	MARSEILLE 9EME	66
93100	MONTREUIL	65
34200	SETE	62
75008	PARIS 08	62
13600	LA CIOTAT	62
75002	PARIS 02	61
75004	PARIS 04	59
92300	LEVALLOIS-PERRET	59
6600	ANTIBES	53
78000	VERSAILLES	53
92800	PUTEAUX	53
92130	ISSY-LES-MOULINEAUX	50

REQUETE 10 : Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces



Requête : Sous-requête WITH, jointure et filtrage WHERE, agrégation GROUP BY

```
1 WITH table_2p AS (  
2     SELECT TOT_PIECES, round(avg(v.VALEUR/b.SURF_CARREZ)) as prix_m2_carrez_2  
3     FROM dan_p3.bien b, dan_p3.vente v  
4     WHERE  
5         v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'  
6         AND v.ID_BIEN = b.ID_BIEN  
7         AND b.TYPE_LOCAL = 'appartement'  
8         AND b.TOT_PIECES = 2  
9     GROUP BY b.TOT_PIECES),  
10  
11 table_3p AS (  
12     SELECT TOT_PIECES, round(avg(v.VALEUR/b.SURF_CARREZ)) as prix_m2_carrez_3  
13     FROM dan_p3.bien b, dan_p3.vente v  
14     WHERE  
15         v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'  
16         AND v.ID_BIEN = b.ID_BIEN  
17         AND b.TYPE_LOCAL = 'appartement'  
18         AND b.TOT_PIECES = 3  
19     GROUP BY b.TOT_PIECES)  
20  
21 SELECT CONCAT(ROUND((prix_m2_carrez_3-prix_m2_carrez_2)*100/prix_m2_carrez_2),'%') as difference_prix_m2_2_3_pieces  
22 FROM table_2p, table_3p;  
23
```



Résultat : Appartements 3 pièces
12% moins cher que les 2 pièces
(prix/m²)

	difference_prix_m2_2_3_pieces
▶	-12%

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

 Requête : Sous-requête WITH, classement RANK () OVER, jointure et filtrage WHERE, agrégation GROUP BY

```
1 • WITH top3_tab as (  
2     SELECT r.CODDEP as Code_departement, r.DEP as Departement, c.COM as Commune,  
3           round(avg(v.VALEUR)) as Valeur_fonciere_moyenne,  
4           RANK() OVER (PARTITION BY r.CODDEP ORDER BY round(avg(v.VALEUR)) DESC) Classement  
5     FROM dan_p3.bien b, dan_p3.vente v, dan_p3.commune c, dan_p3.region r  
6     WHERE  
7         v.DATE_VENTE BETWEEN '2020-01-01' AND '2020-06-30'  
8         AND b.ID_BIEN = v.ID_BIEN  
9         AND b.ID_COM = c.ID_COM  
10        AND c.ID_REG = r.ID_REG  
11        AND r.CODDEP IN ('6', '13', '33', '59', '69')  
12    GROUP BY c.ID_COM  
13 )  
14 SELECT * FROM top3_tab WHERE Classement <4  
15 ;
```



Résultat : Biens les + chers en Gironde et Alpes-Maritimes

Code_departement	Departement	Commune	Valeur_fonciere_moyenne	Classement
13	Bouches-du-Rhône	MARSEILLE 12EME	399500	1
13	Bouches-du-Rhône	GIGNAC-LA-NERTHE	330000	2
13	Bouches-du-Rhône	SAINT SAVOURNIN	314425	3
33	Gironde	LA TESTE-DE-BUCH	705000	1
33	Gironde	LEGE-CAP-FERRET	695051	2
33	Gironde	VAYRES	335000	3
59	Nord	BERSEE	433202	1
59	Nord	CYSOING	408550	2
59	Nord	HALLUIN	322250	3
6	Alpes-Maritimes	SAINT-JEAN-CAP-FERRAT	968750	1
6	Alpes-Maritimes	EZE	655000	2
6	Alpes-Maritimes	MOUANS-SARTOUX	476898	3
69	Rhône	VILLE SUR JARNIOUX	485300	1
69	Rhône	LYON 2EME	455217	2
69	Rhône	LYON 6EME	426968	3