



Data-Immo

Les données au cœur de l'immobilier



Sommaire

Les fondations de la base de données

Dictionnaire des données	p.4
Définition des tables	p.5
Schéma relationnel	p.6
Configuration des tables	p.7
Schéma conceptuel	p.8

Création de la base de données

Création du script	p.13
Chargement de la base de données	p.14
Base de données opérationnelle	p.15

Préparation des données

Nettoyage des données	p.10
Création du fichier des tables	p.11

Requêtes et résultats

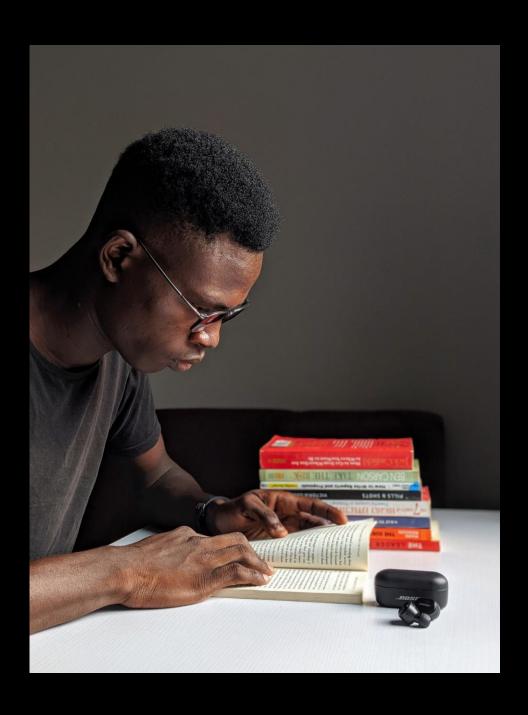
Requêtes et résultats p.17 à 27



Annexes

Dictionnaire des données suite présentation Script SQL de la base de données suite présentation Résultats complet des requêtes fichier .xls externe

Les fondations de la base de données



Dictionnaire des données

Afin de faciliter des modifications et la reprise d'informations par un tiers, toutes les informations de la base sont répertoriées dans un répertoire dédié.

(Dictionnaire complet en annexe p.29)

numero	code propriete	signification	type	observation
9	Date mutation	Date de vente / cession	Date	format ISO-8601 (YYYY-MM-DD)
18	Code ID commune	Code commune INSEE	Numerique	Longueur: 5 caracteres
20	Commune	Nom de la commune	Alphanumerique	Composition lettres et chiffres
43	Nature culture	Type de culture du terrain	Texte	Vegetation presente sur le terrain

Définitiondes tables



Que contiendront les tables?



bien

La table « bien » donnera les caractéristiques technique du ou des biens. Cette dernière permettra d'identifier chaque bien de manière unique malgré les similitudes avec d'autres.

cadastre

La table « cadastre » fournira des informations sur l'environnement de chaque bien mais aussi des caractéristiques liées aux déclarations administratives des biens. Cette partie de la base de données sera pertinente seulement pour des besoins spécifiques ou des recherches plus approfondies.

vente

La table « vente » contiendra des informations liées aux transactions des biens, tel que le prix, la nature ou encore la date.

adresse

La table « adresse » correspondra aux informations de localisation du ou des biens, tel que la rue, la commune, le département ...

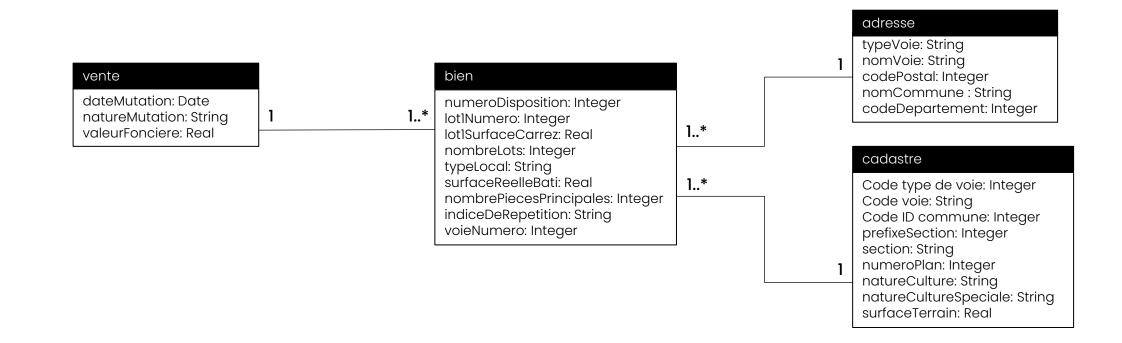
Schéma relationnel

"

La structure de la base de données est représenté par un schéma en langage UML.

Ce dernier met en évidence les différentes tables avec leurs relations ainsi que leurs cardinalités.

Certaines contraintes tel que la norme 3NF consistant à ce qu'il n'y ai pas de dépendance entre attributs clés et non clés, sauf exception, doivent être respectées.



Configuration des tables

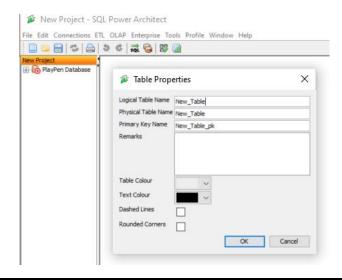


La structure de la base de donnée

Création des tables

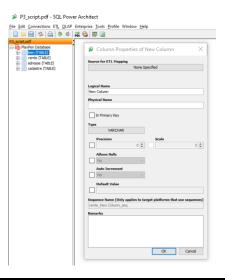


La première étape consiste à créer l'ensemble des tables et de les nomme sur un logiciel spécialisé..



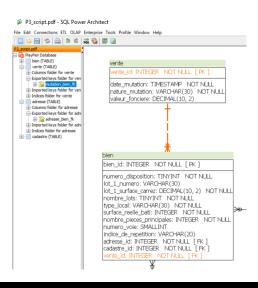
Création des attributs

La seconde étape consiste à créer l'ensembles des attributs de chaque table, en définissant pour chacun le type, la taille et leur fonction (PK, NOT NULL, VARCHAR ...).



Mise en place des relations

La 3º étape consiste à lier les tables entre elles en fonction du schéma relationnel établie via des relations identifiantes ou non.



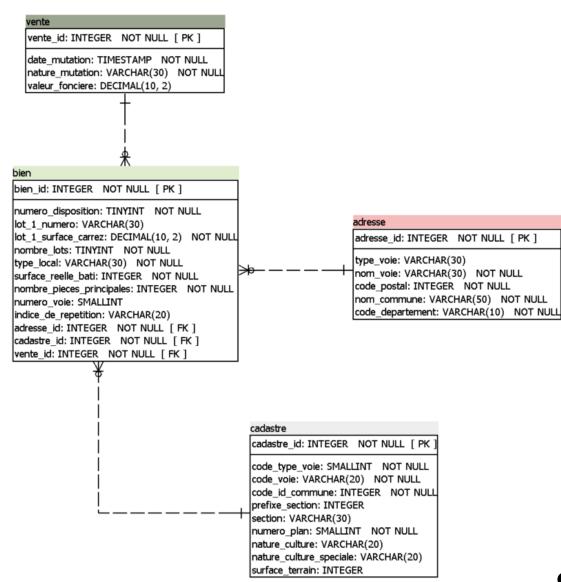


Modèle conceptuel des données

Ce modèle représente l'ensemble des tables de la base de données avec leurs attributs, leurs clés ainsi que leurs relations.

(Script SQL en annexe p.30)





Préparation des données

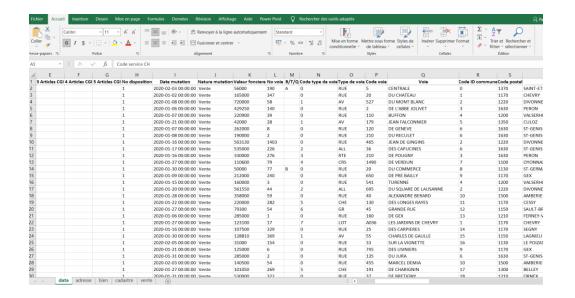
Nettoyagedes données

"

Conserver des données pertinentes

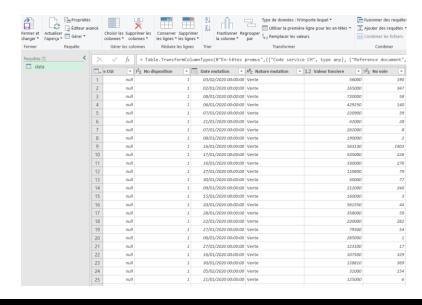
Nettoyage des données

Le nettoyage des données permet de trier et conserver les données de façon pertinente. Dans ce fichier, j'ai volontairement écarté certaines colonnes contenant des données nulles ou non représentatives (-1%).



Manipulations et adaptation des données

A l'aide de l'éditeur Power Query, le fichier est ensuite traité, chaque colonne conservera ou non le type de données fourni par défaut et subira des modifications le cas échéant.



Création du fichier des tables

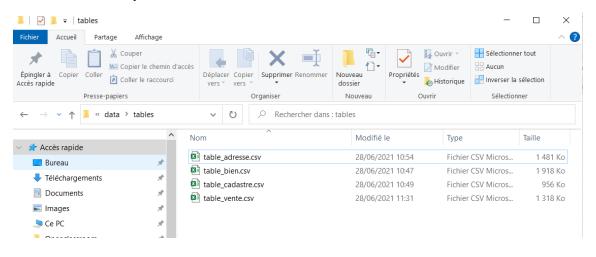
66

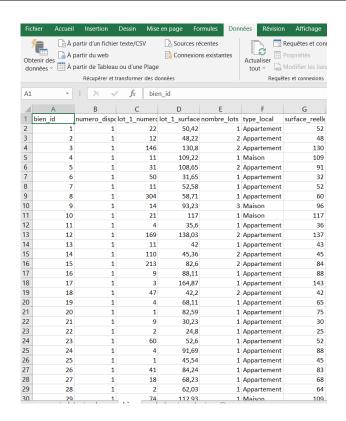
Préparation des tables

Division des données en tables

Une fois le fichier nettoyer et les colonnes configurées, le fichier sera scinder en plusieurs tables, suivant le schéma relationnel qui prendront chacune la forme d'un fichier CSV.

NB : Les clés primaires et étrangères (PK, FK) seront intégrées manuellement car les données sont déjà existantes.





Création de la base de données

Création du script

Le script de la base de données

Création du code des tables

Cette étape consiste à créer la structure de la base de donnée dans le SGBD.

Exemple avec la table VENTE:

-- création de la séquence qui permettra d'incrémenter une série pour la clé primaire (PK)

CREATE SEQUENCE public.vente_vente_id_seq

-- création de la table VENTE avec ses attributs

CREATE TABLE public.vente

-- création de la clé primaire (PK) avec son type INTEGER (nombre entier), un e caractéristique NOT NULL (qui doit obligatoirement contenir une valeur) et une valeur suivante par défaut suivant la séquence.

(vente_id INTEGER NOT NULL DEFAULT nextval('public.vente_vente_id_seg'),

-- ajout des autres attributs

date_mutation TIMESTAMP NOT NULL, nature_mutation VARCHAR(30) NOT NULL valeur_fonciere NUMERIC(10,2),

-- Ajout d'un attribut avec contrainte pour la clé étrangère

CONSTRAINT vente_pk PRIMARY KEY (vente_id)

ALTER SEQUENCE public.vente_vente_id_seq OWNED B\ public.vente.vente_id;

Création du code des clés étrangères

Cette étape consiste à créer les clés qui permettront de relier les tables entre elle et d'indiquer l'action qui sera réalisée en cas de modification de ces dernières.

NB : Les attributs de ces clés ont déjà été incluses dans les tables

Exemple:

-- modification de la table BIEN pour ajouter la contrainte de clé étrangère (FK) cadastre.bien.fk

ALTER TABLE public.bien ADD CONSTRAINT cadastre_bien_fk

-- nom de la clé étrangère

FOREIGN KEY (cadastre_id)

-- provenance de la clé étrangère

REFERENCES public.cadastre (cadastre_id)

-- Ce bloc consiste à spécifier les actions à suivre en cas de suppression ou modification des données.

ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE:

Chargement et exécution du code

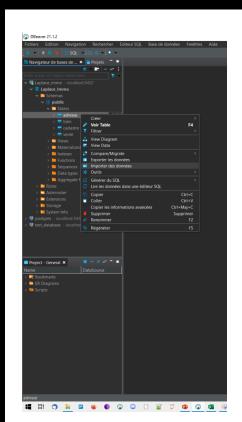
Une fois le code créer, il ne reste plus qu'à l'exécuter afin de créer la base de donnée.

NB: Le script des clés étrangères (FK) devra être exécuté une fois les données importées.

Chargement de la base de données



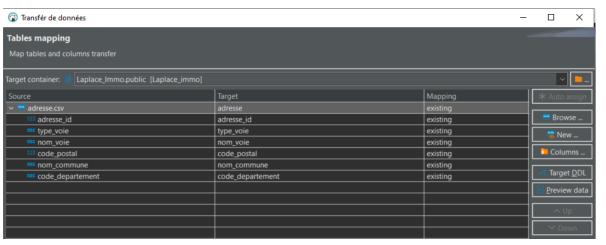
L'import des données dans la base

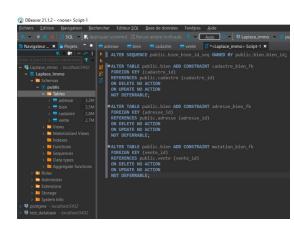


Importation des données dans le Système de Gestion de Base de Données (SGBD)

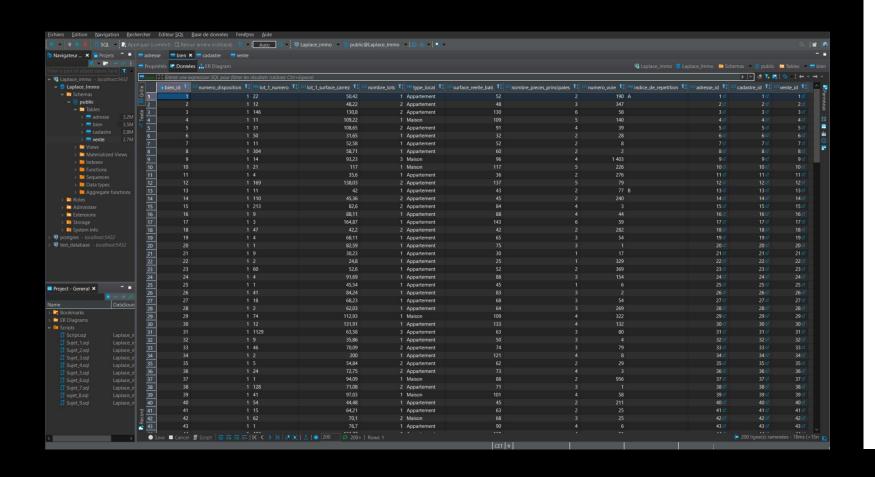
Les données sont importées dans le SGBD via les fichiers CSV de chaque table. Chaque fichier devra faire correspondre ses champs avec les tables de destination du SGBD.

NB: Le script des clés étrangères (FK) peuvent désormais être exécutés afin de créer les liens entre chaque table.





Base de données opérationnelle



Base de données opérationnelle

La base de donnée est désormais opérationnelle.

Toutes les données sont importées dans leurs tables respectives, sont reliées entre elles selon le schéma ralationnel et peuvent désormais faire l'objet de requêtes.

La base de données contient :

• 34169 instances

La table ADRESSE contient :

• 5 attributs dont 1 clé primaire

La table **BIEN** contient:

• 13 attributs dont 1 clé primaire (PK) et 3 clés étrangère (FK)

La table CADASTRE contient:

• 10 attributs dont 1 clé primaire (PK

La table **VENTE** contient:

4 attributs dont 1 clé primaire (PK)

Requêtes

Nombre total d'appartements vendus au ler semestre 2020

select from COUNT(bien.type_local) as "Nombre d'appartements S1 2020"

bien

join where and

vente on vente.vente_id = bien.vente_id

bien.type_local = 'Appartement'

date_mutation between '2020-01-01' and '2020-06-30'

Requêtes 1

Contexte

Cette requête va chercher le nombre d'occurrences de l'attribut « bien.type_local » dans la table « bien ».

Fonctions utilisées

SELECT - FROM COUNT JOIN ON WHERE

Jointures

Une jointure entre la table **BIEN** et la table **VENTE** est créé pour ajouter des conditions venant de la table **VENTE**.

Conditions

```
« bien.type_local » = 'Appartement'
```

« date_mutation » between '2020-01-01' and '2020-06-30'

Proportion des ventes d'appartements

par le nombre de pièces

select bien.nombre_pieces_principales as "Nombre de pieces",

count(bien.nombre_pieces_principales)

as "Nombre d'appartements",

round(count(bien.nombre_pieces_principales) *100) /

(select count(bien.nombre_pieces_principales)

from bien

bien.type_local = 'Appartement') where

as "Proportion des ventes"

bien from

bien.type_local = 'Appartement' where

"Nombre de pieces" group by "Nombre de pieces" order by

Requêtes 2

Contexte

Cette requête va chercher le ratio de ventes d'appartements en fonction de son nombre de pièces.

Fonctions utilisées

SELECT - FROM

COUNT

ROUND(COUNT (opération avec sous requête))

WHERE

GROUP BY - ORDER BY

Conditions

« bien.type_local » = 'Appartement'

Groupement

« Nombre de pieces »

Ordre

« Nombre de pieces » - Ascendant

Nombre de pieces		Nombre d'appartements	Proportion des ventes
	0	30	0.0956083880425776
	1	6739	21.476830900631015
	2	9783	31.177895340684557
	3	8966	28.574160239658358
	4	4460	14.213780355663204
	5	1114	3.550258142647715
	6	204	0.6501370386895277
	7	54	0.17209509847663967
	8	17	0.05417808655746064
	9	8	0.02549557014468736
	10	2	0.00637389253617184
	11	1	0.00318694626808592

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

select adresse.code_departement as "Departement",

round(avg (vente.valeur_fonciere /

bien.lot_1_surface_carrez),2) as "Prix moyen du m2"

from vente

join bien on vente.vente_id = bien.vente_id

adresse on adresse.adresse_id=bien.adresse_id

group by "Departement"

order by "Prix moyen du m2" desc

limit

;

join

Requêtes 3

Contexte

Cette requête va chercher le prix moyen du mètre carré de chaque département et limiter le résultat sur 10 d'entres eux.

Fonctions utilisées

SELECT - FROM
ROUND(AVG (opération))
COUNT
JOIN ON
GROUP BY - ORDER BY
LIMIT

Jointures

Une jointure entre les tables **VENTE**, **BIEN** et **ADRESSE** est crée pour ajouter des conditions venant de la table **ADRESSE**.

Groupement

« Departement »

Ordre de tri

« Prix moyen du m2 » - Descendant

Departement		Prix moyen du m2
7	75	12052,89
9	2	7219,39
9	4	5343,28
	6	4700,33
7	4	4667,13
9	3	4344,78
7	78	4225,25
6	9	4059,31
2	Α	4026,97
3	3	3764,14

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

Requêtes 4

Contexte

Cette requête va chercher le prix moyen du mètre carré pour les maisons situées dans les départements de l'Île de France.

Fonctions utilisées

SELECT - FROM ROUND(AVG (opération)) JOIN ON WHERE - AND

Jointures

Une jointure entre les tables **ADRESSE**, **BIEN** et **VENTE** est crée pour ajouter des conditions venant de la table **VENTE**.

Conditions

```
«adresse.code_departement» in ('75','77','78','91','92','93','94','95')
« bien.type_local » = 'Maison'
```

Prix moyen du m2 en lle de France

3745.01

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

select
bien.bien_id as "ID appartements",
adresse.code_departement as "Departement",
vente.valeur_fonciere as "Valeur fonciere",
bien.lot_1_surface_carrez as "Surface Carrez"

from bien

join vente on bien.vente_id = vente.vente_id

join adresse on bien.adresse_id =adresse.adresse_id

where bien.type_local = 'Appartement' and vente.valeur_fonciere is not null

group by "ID appartements",

"Departement",

"Valeur fonciere",

"Surface Carrez"

order by "Valeur fonciere" desc

limit 10

;

ID appartements	Departement	Valeur fonciere	Surface Carrez
32275	75	9000000.00	9.10
21835	91	8600000.00	64.00
29799	75	8577713.00	20.55
32433	75	7620000.00	42.77
29850	75	7600000.00	253.30
29522	2 75	7535000.00	139.90
31973	75	7420000.00	360.95
32135	75	7200000.00	595.00
29353	75	7050000.00	122.56
29513	75	6600000.00	79.38

Requêtes 5

Contexte

Cette requête va chercher la liste des 10 appartements les plus chers et afficher l'ID, le Departement, la Valeur Foncière et la Surface Carrez.

Fonctions utilisées

SELECT - FROM JOIN WHERE - AND GROUP BY - ORDER BY LIMIT

Jointures

Une jointure entre les tables **BIEN**, **VENTE** et **ADRESSE** est crée pour ajouter des conditions venant de la table **VENTE**.

Conditions

« bien.type_local » = 'Appartement'
« Valeur fonciere » is not null

Groupement

ID appartements, Departement, Valeur Fonciere, Surface Carrez

Ordre

« Valeur fonciere »

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

```
with
               as (
table1
               round(count(vente.vente_id),2) as nb_ventes_trimestre_1
select
from
               vente.date_mutation between '2020-01-01' and '2020-03-31'),
where
               as (
table2
               round(count(vente.vente_id),2) as nb_ventes_trimestre_2
select
from
               vente
               vente.date_mutation between '2020-04-01' and '2020-06-30')
where
               round(((table2.nb_ventes_trimestre_2 -
select
               table1.nb_ventes_trimestre_1)/table1.nb_ventes_trimestre_1*100),2)
                              as "Taux d'evolution"
               table1, table2
from
```

Requêtes 6

Contexte

Cette requête va chercher la ratio de ventes entre le 1er et le 2nd trimestre

Fonctions utilisées

WITH
SELECT - FROM
ROUND(COUNT (opération))
WHERE - AND

Tables temporaires

2 tables intermédiaires sont crées avec chacune leur conditions respectives afin de pouvoir faire l'opération qui conclura au ratio des ventes dans un nouveau SELECT.

Conditions

vente.date_mutation between '2020-01-01' and '2020-03-31" vente.date_mutation between '2020-04-01' and '2020-06-30'

Taux d'evolution

3.68

Liste des communes où le nombre de ventes a augmenté d'au moins 20% entre le premier et le second trimestre de 2020

```
----- Tables temporaires
with
table1 as (
select
                 adresse.nom_commune,
                round(count(vente.vente_id),2) as nb_ventes_1
from
                 adresse
                 bien on bien.adresse_id =adresse.adresse_id
join
join
                 vente on bien.vente_id = vente.vente_id
                 vente.date_mutation between '2020-01-01' and '2020-03-31'
where
group by
                 adresse.nom_commune),
table2 as (
select
                 adresse.nom_commune,
                round(count(vente.vente_id),2) as nb_ventes_2
from
                 adresse
                 bien on bien.adresse_id =adresse.adresse_id
join
join
                 vente on bien.vente_id = vente.vente_id
                 vente.date_mutation between '2020-04-01' and '2020-06-30'
where
                 adresse.nom_commune)
group by
----- Requête select
                 table1.nom_commune as "Communes",
select
                table1.nb_ventes_las "Nombre de ventes trimestre 1",
                table2.nb_ventes_2as "Nombre de ventes trimestre 2",
                 ROUND((table2.nb_ventes_2 - table1.nb_ventes_1)/table1.nb_ventes_1*100,2) as "Evolution"
from
                 table1
join
                 table2 on table1.nom_commune = table2.nom_commune
                round(((table2.nb_ventes_2 - table1.nb_ventes_1)/table1.nb_ventes_1)*100) >= 20
where
order by
```

Requêtes 7

Contexte

Cette requête va chercher la différence en pourcentage des ventes sur chaque commune entre le 1^{er} et le 2nd trimestre 2020. Les résultats seront limités aux communes excédant 20%.

Fonctions utilisées

WITH
SELECT - FROM
ROUND(COUNT (opération))
JOIN ON
WHERE
GROUP BY
ORDER BY

Tables temporaires

2 tables intermédiaires sont crées avec chacune leur conditions respectives afin de pouvoir effectuer l'opération qui calculera la différence des ventes dans un nouveau SELECT.

Jointures

Une jointure entre les tables **ADRESSE**, **BIEN** et **VENTE** est créé pour chaque table temporaire pour ajouter des conditions venant de la table **VENTE**.

Conditions

```
« date_mutation » between '2020-01-01' and '2020-03-31' 
« date_mutation » between '2020-04-01' and '2020-06-30'
```

Groupement

« adresse.nom_commune »

Ordre

« Evolution » - Ascendant

Communes	Nombre de ventes trimestre 1	Nombre de ventes trimestre 2	Evolution
BESSANCOURT	5.00	6.00	20.00
LE BEAUSSET	10.00	12.00	20.00
SCHOELCHER	5.00	6.00	20.00
SAINT-HILAIRE-DE-RIEZ	15.00	18.00	20.00
CHESSY	5.00	6.00	20.00
GOLBEY	5.00	6.00	20.00
LA QUEUE-EN-BRIE	10.00	12.00	20.00
SAINTE FOY LES LYON	5.00	6.00	20.00
DIVONNE-LES-BAINS	5.00	6.00	20.00
GIVORS	5.00	6.00	20.00
BRIANCON	5.00	6.00	20.00
ARCACHON	25.00	30.00	20.00
LEUCATE	29.00	35.00	20.69
ETAMPES	24.00	29.00	20.83
LE CHESNAY-ROCQUENCOURT	14.00	17.00	21.43
BOISSY-SAINT-LEGER	9.00	11.00	22.22
ORLY	9.00	11.00	22.22
PARIS 08	62.00	77.00	24.19
RENNES	61.00	76.00	24.59
L'ILE-SAINT-DENIS	4.00	5.00	25.00
MERY-SUR-OISE	4.00	5.00	25.00
BONNEVILLE	4.00	5.00	25.00
VAL DE BRIEY	4.00	5.00	25.00
VALENCE	28.00	35.00	25.00
LONGPONT-SUR-ORGE	4.00	5.00	25.00
PONTOISE	24.00	30.00	25.00
BATZ-SUR-MER	4.00	5.00	25.00
SOUSTONS	8.00	10.00	25.00
CHAMONIX MONT BLANC	12.00	15.00	25.00
SAINT-GENIS-LAVAL	8.00	10.00	25.00
NOGENT SUR OISE	8.00	10.00	25.00
CADAUJAC	4.00	5.00	25.00
PARIS 11	169.00	214.00	26.63
VILLEMOMBLE	15.00	19.00	26.67
SARCELLES	15.00	19.00	26.67
NICE	173.00	220.00	27.17
1404	44.00	44.00	07.07

Résultat requêtes 7





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

```
with
                as (
table1
                round(avg(vente.valeur_fonciere/bien.lot_1_surface_carrez),2)
select
                                  as valfonc_moy_2
from
                 adresse
join
                 bien on bien.vente_id =adresse.adresse_id
join
                 vente onvente.vente_id =bien.adresse_id
                 bien.nombre_pieces_principales = 2
where
                bien.type_local ='Appartement'),
and
                as (
table2
                round(avg(vente.valeur_fonciere/bien.lot_l_surface_carrez),2)
select
                                  as valfonc_moy_3
                 adresse
from
                 bien on bien.vente_id =adresse.adresse_id
join
join
                 vente onvente.vente_id =bien.adresse_id
                bien.nombre_pieces_principales = 3
where
and
                 bien.type_local ='Appartement')
                 round(((table2.valfonc_moy_3 - table1.valfonc_moy_2)/table1.valfonc_moy_2)*100,2)
select
                                  as "Taux d'evolution"
                table1, table2
from
```

Requêtes 8

Contexte

Cette requête va chercher la ratio de ventes entre le 1er et le 2nd trimestre

Fonctions utilisées

WITH
SELECT - FROM
ROUND(AVG (opération))
WHERE - AND

Jointures

2 tables intermédiaires sont crées avec chacune leur conditions respectives afin de pouvoir faire l'opération qui conclura au ratio des ventes dans un nouveau SELECT.

Conditions

bien.type_local = 'Appartement' bien.nombre_pieces_principales = 3 / 2

Taux d'evolution

-12.40

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

```
with table1 as (
               adresse.code_departement as "Departement",
select
               adresse.nom_commune as "Commune",
               round(avg(vente.valeur_fonciere),2) as "Valeur fonciere",
               rank () over(partition by adresse.code_departement
                              order by round(avg(vente.valeur_fonciere),2)desc)
                                                             as "Classement"
from
               adresse
               bien on bien.adresse_id= adresse.adresse_id
join
ioin
               vente on vente.vente_id=bien.vente_id
               adresse.code_departement in ('6','13','33','59','69')
where
               vente.valeur_fonciere is not null
and
group by
               "Departement", "Commune"
order by
               "Valeur fonciere" desc)
select *
               table1
from
                "Classement" <= 3
where
               "Departement","Valeur fonciere" desc
order by
```

Requêtes 9

Contexte

Cette requête va chercher la moyenne des valeurs foncières, pour les 3 communes les plus chers des départements 6, 13, 33, 59 et 69.

Fonctions utilisées

WITH
SELECT - FROM
ROUND(AVG (opération))
RANK() OVER(PARTITION BY)
JOIN ON
WHERE
GROUP BY
ORDER BY

Tables temporaires

1 table temporaire est créée pour créer un classement des communes avec les valeurs foncières les plus élevées.

Jointures

Une jointure entre les tables **ADRESSE**, **BIEN** et **VENTE** est créé pour chaque table temporaire pour ajouter des conditions venant de la table **VENTE**.

Conditions

```
« adresse.code_departement » in ('6','13','33','59','69')
"Classement" <= 3</pre>
```

Groupement

"Departement", "Commune"

Ordre

"Departement","Valeur fonciere" – Descendant

Résultat requêtes 9

Departement	Commune	Valeur fonciere	Classement
13	GIGNAC-LA-NERTHE	330000.00	1
13	SAINT SAVOURNIN	314425.00	2
13	CASSIS	313416.88	3
33	LEGE-CAP-FERRET	549500.64	1
33	VAYRES	335000.00	2
33	ARCACHON	307435.93	3
59	BERSEE	433202.00	1
59	CYSOING	408550.00	2
59	HALLUIN	322250.00	3
6	SAINT-JEAN-CAP-FERRAT	968750.00	1
6	EZE	655000.00	2
6	MOUANS-SARTOUX	476898.10	3
69	VILLE SUR JARNIOUX	485300.00	1
69	LYON 2EME	455217.27	2
69	LYON 6EME	426968.25	3,



Merci

de votre attention

