



# Création et utilisation de la base de données

Arnaud Golliot



Laplace Immo

# Contexte du projet

Notre Direction Générale est sensible depuis quelque temps à l'importance des données, et pense que l'agence doit se démarquer de la concurrence en créant un modèle pour mieux prévoir le prix de vente des biens immobiliers.

Ce projet stratégique est appelé en interne le projet "DATAImmo". Clara Daucourt, notre CTO, est en charge de le piloter. Il nécessite des compétences en data (collecte, transformation, combinaison et agrégation des données, génération de rapport de données d'aide factuelle et éclairée à la décision)

Dans ce cadre, Clara m'a confié la modification de la base de données permettant de collecter les transactions immobilières et foncières en France. A cette fin, j'utiliserais cette base de données pour analyser le marché et aider les différentes agences régionales à mieux accompagner leurs clients.



# La stratégie de sauvegarde et la conformité RGPD

## Stratégie de sauvegarde :

- **Stockage** : cloud protégé
- **Fréquence de sauvegarde** : Tous les jours, en dehors de heures ouvrées (02h00 am par exemple)
- **Mode de sauvegarde (totale, différentielle)** : totale
- **Archivage** : time stamp AAA-MM-JJ\_HH-MM-SS (exemple : Laplace\_Immo\_BDD\_2024-10-07\_16h23m10s)
- **Routine de sauvegarde** : soit au moyen d'un script powershell encapsulant un ordre de sauvegarde de la BDD (dump), ou au moyen d'un ETL (type Talend ou Knime)
- **Mode d'exécution** : automatique

## Conformité RGPD :

- **Données sensibles concernées** : le nom de l'acquéreur dans le fichier des ventes
- **Solution** : crypter le champ correspondant dans le fichier CSV Ventes.csv avant chargement
- **Mise en œuvre** : développer un script python faisant appel soit à un algorithme de cryptage disponible dans une bibliothèque prévue à cet effet, soit un algorithme spécifique d'anonymisation de la donnée
- **Résultat attendu** : le nom de l'acquéreur est crypté dans la base de données

# Les données initiales et transformées (normalisation)

- **3 fichiers CSV initiaux**

- ✓ donnees\_communes.xlsx
- ✓ fr-esr-referentiel-geographique.xlsx
- ✓ Valeurs-foncières.xlsx

- **Transformés en 6 fichiers CSV (normalisation en 3<sup>ième</sup> Forme Normale 3FN)**

- ✓ Régions.csv
- ✓ Départements.csv
- ✓ Communes.csv
- ✓ Adresses.csv
- ✓ Biens\_immobiliers.csv
- ✓ Ventes.csv

- **Pourquoi cette normalisation en 3FN des données**

- ✓ **Unicité** : toute donnée est présente en un et un seul exemplaire dans la BDD (aucun doublon), ainsi toute opération sur les données (suppression, insertion, modification) se fait en un et un seul endroit de la base de données
- ✓ **Intégrité** : dans chacune des table de données (à l'image de chaque fichier CSV), chaque ligne contient une combinaison unique des valeurs de données présentes dans la table
- ✓ **Volumétrie** : la normalisation permet de réduire le volume de la base de données

# Extrait du dictionnaire 3FN des données (1/3)

		DENOMINATION	SIGNIFICATION / DESCRIPTION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
REGION	001	Code région	Code région de localisation du bien immobilier	INTEGER	-	E	Clé primaire, jamais nulle	Aucune
	002	Nom région	Nom de la région de localisation du bien immobilier	VARCHAR	100	E	jamais nulle	Aucune
	003	Nom de regroupement des régions	Nom de regroupement des régions qui inclut celle du bien immobilier	VARCHAR	100	E	Jamais nulle	Aucune
DEPARTEMENT	001	Code département	Code département de localisation du bien immobilier  Attention un code département peut contenir des caractères alphabétiques	VARCHAR	3	E	Clé primaire, jamais non nulle	Aucune
	002	Nom département	Nom du département de localisation du bien immobilier	VARCHAR	100	E	jamais nulle	Aucune
	003	Code région	Code région de localisation du bien immobilier	INTEGER	-	E	Clé étrangère, potentiellement nulle en cas de suppression d'une région	Aucune
COMMUNE	001	Identifiant de commune	Identifiant unique de la commune de localisation du bien immobilier  Attention le code commune est relatif à chaque département, et deux départements différents peuvent comporter le même code commune pour désigner deux communes différentes.	VARCHAR	10	CO	Clé primaire, jamais nulle	Identifiant de commune = code département + code commune
	002	Code commune	Code relatif (dans son département) de la commune de localisation du bien immobilier	INTEGER	-	E	Non nulle	Aucune
	003	Nom commune	Nom de la commune de localisation du bien immobilier	VARCHAR	100	E	Non nulle	Aucune
	004	Population municipale	Nombre d'habitants de la commune de localisation du bien immobilier, qui y sont rattachés administrativement (à vérifier, absence de lexique)	INTEGER	-	E	Non nulle	Aucune
	006	Population comptée à part	Nombre d'habitants de la commune de localisation du bien immobilier qui n'y sont pas rattachés administrativement (à vérifier, absence de lexique)	INTEGER	-	E	Non nulle	Aucune
	007	Population totale	Somme des deux populations précédentes	INTEGER	-	E	Non nulle	Aucune
	008	Latitude de géolocalisation	Latitude de géolocalisation du bien immobilier A rajouter dans la seconde version du PoC	FLOAT	-	Ca	A rajouter dans seconde version du PoC	Extraction du premier champ
	009	Longitude de géolocalisation	Longitude de géolocalisation du bien immobilier A rajouter dans la seconde version du PoC	FLOAT	-	Ca	A rajouter dans seconde version du PoC	Extraction du second champ
	010	Code département	Code du département de localisation du bien immobilier	VARCHAR	3	E	Clé étrangère, potentiellement nulle en cas de suppression d'un département	Aucune

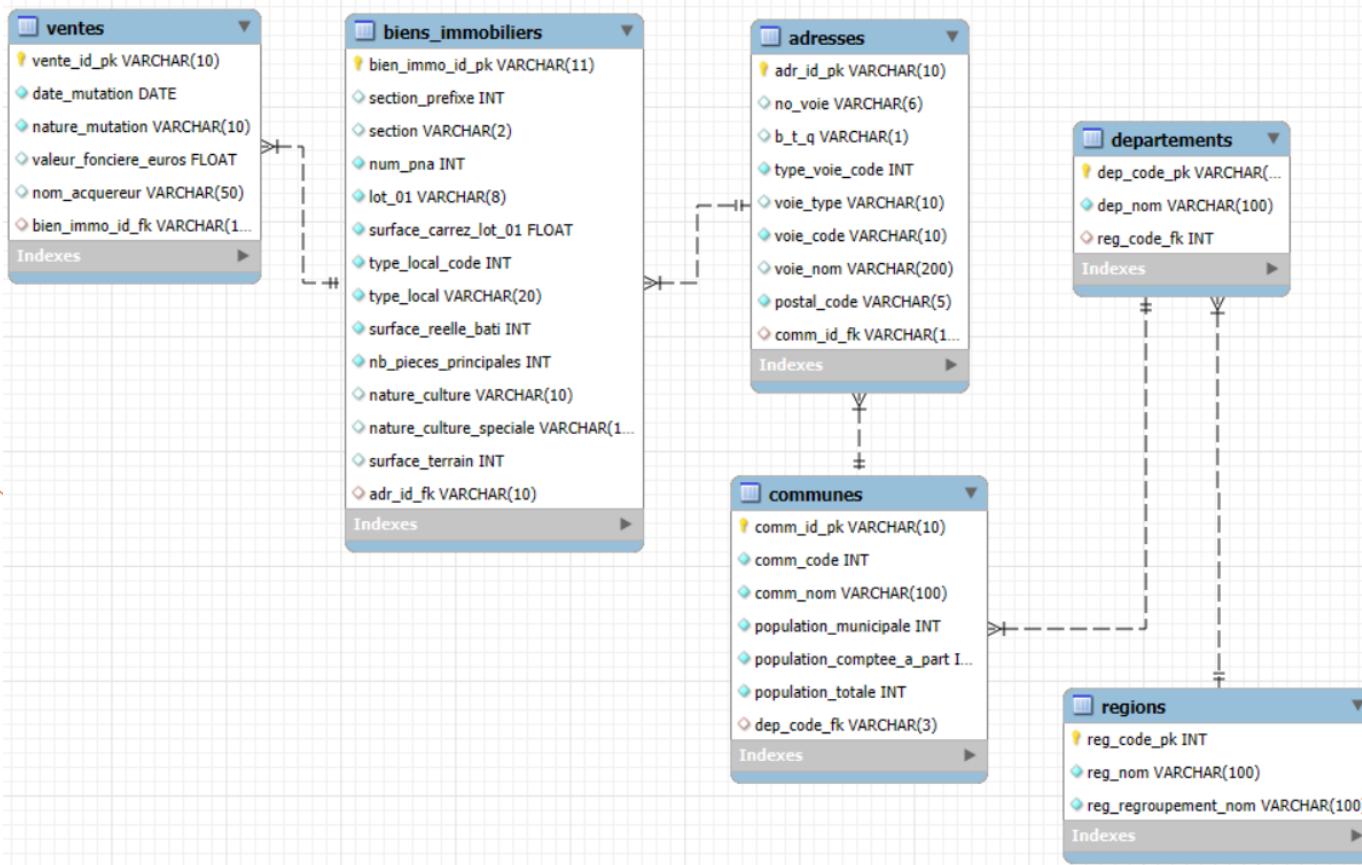
# Extrait du dictionnaire 3FN des données (2/3)

		DENOMINATION	SIGNIFICATION / DESCRIPTION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
ADRESSE	001	Identifiant d'adresse	Identifiant unique de l'adresse du bien immobilier Un adresse est identifié par la concaténation des éléments d'adresse (numéro, voie, nom), du code postal, et de l'identifiant de commune	VARCHAR	10	Ca	Clé primaire, jamais nulle	Identifiant d'adresse = éléments de voie (num voie, btq, nom voie) + code postal + identifiant de commune Nous transcodons cet identifiant en code technique au format ADRnnnnnnn (avec nnnnnnn un entier codé du 8 positions; par exemple ADR00000021)
	002	Numéro de voie	Numéro de voie de l'adresse du bien immobilier	INTEGER	-	E	Peut être nulle	Aucune
	003	Complément de numéro de voie	Bis, Ter, etc.	VARCHAR	1	E	Peut être nulle	Aucune
	004	Code normalisé de type de voie	Code standardisé de type de voie, à préciser (absence de lexique)	INTEGER		E	Jamais nulle	Aucune
	005	Type de voie	Rue, boulevard, avenue, ...	VARCHAR	10	E	Peut être nulle	Aucune
	006	Code normalisé de voie	Code standardisé de voie, à préciser (absence de lexique)	VARCHAR	10	E	Jamais nulle	Aucune
	007	Nom de la voie	Nom de la voie de l'adresse du bien immobilier	VARCHAR	200	E	Peut être nulle	Aucune
	008	Code postal	Code postal de l'adresse du bien immobilier	VARCHAR	5	E	Jamais nulle	Aucune
	009	Identifiant de commune	Identifiant unique de la commune de localisation du bien immobilier  Attention le code commune est relatif à chaque département, et deux départements différents peuvent comporter le même code commune pour désigner deux communes différentes.	VARCHAR	10	CO	Clé étrangère, potentiellement nulle en cas de suppression d'une commune	Identifiant de commune = code département + code commune
BIEN IMMOBILIER	001	Identifiant de bien	Identifiant unique du bien immobilier Un bien est identifié par son identifiant d'adresse et son numéro de PREMIER lot à l'adresse indiquée	VARCHAR	11	Ca	Clé primaire, jamais nulle	identifiant du bien = adresse + numéro du 1er lot Nous transcodons cet identifiant en code technique au format BIMnnnnnnn (avec nnnnnnn un entier codé du 8 positions; par exemple BIM00000021)
	002	Préfixe de section	Non défini (absence de lexique), chargé mais non utilisé	INTEGER	-	E	Peut être nulle	Aucune
	003	Section	Non défini (absence de lexique), chargé mais non utilisé	VARCHAR	2	E	Peut être nulle	Aucune
	004	Numéro de PNA	Non défini (absence de lexique), chargé mais non utilisé	INTEGER		E	Jamais nulle	Aucune
	005	Identifiant du premier lot à l'adresse indiquée	Identifiant à l'adresse indiquée du premier lot du bien immobilier	VARCHAR	8	E	Jamais nulle	Aucune
	006	Surface loi carrez du premier lot	Surface loi Carrez du 1er lot	FLOAT	-	E	La surface du 1er lot correspond en fait à la surface de l'ensemble du bien tous lots	Aucune

# Extrait du dictionnaire 3FN des données (3/3)

	DENOMINATION	SIGNIFICATION / DESCRIPTION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
	007 Code du type de local	Code technique du type de local	INTEGER	-	E	Jamais nulle	Aucune
	008 Type de local	"Appartement" ou "Maison"	VARCHAR	20	E	Jamais nulle	Aucune
	009 Surface réelle du bien	Surface réelle du bien immobilier	INTEGER	-	E	Jamais nulle	Aucune
	010 Nombre de pièce principales	Nombre de pièces principales du biens immobilier (salon, séjour, chambres)	INTEGER	-	E	Jamais nulle	Aucune
	011 Nature de culture	Non défini (absence de lexique), chargé mais non utilisé	VARCHAR	10	E	Non utilisé (à confirmer)	Aucune
	012 Nature de culture spéciale	Non défini (absence de lexique), chargé mais non utilisé	VARCHAR	10	E	Non utilisé (à confirmer)	Aucune
	013 surface du terrain	Surface complète du terrain d'occupation du bien immobilier (jardin, terrasse extérieure, etc.)	INTEGER	-	E	Peut être nulle	Aucune
	014 Identifiant adresse	Identifiant unique d'adresse de localisation du bien immobilier  Un bien est identifié par une et une seule adresse, et le numéro de son premier lot	VARCHAR	10	Ca	Clé étrangère, potentiellement nulle en cas de suppression d'une adresse	Identifiant d'adresse = éléments de voie (num voie, btq, nom voie) + code postal + identifiant de commune  Nous transcodons cet identifiant en code technique au format ADRnnnnnnn (avec nnnnnnn un entier codé du 8 positions; par exemple ADR00000021)
VENTES	001 Identifiant de vente	Identifiant unique de vente d'un bien immobilier  Une vente c'est un et un seul bien immobilier, en revanche un bien peut éventuellement être vendu plusieurs fois	VARCHAR	10	Ca	Une vente est identifiée par la date de vente, l'acquéreur, la valeur de vente, l'identifiant du bien	identifiant de vente = date vente + identifiant du bien  Nous transcodons cet identifiant en code technique au format VTEnnnnnnn (avec nnnnnnn un entier codé du 8 positions; par exemple VTE00000021)
	002 Date de vente	Date de vente du bien immobilier	DATE	-	E	Format anglais "AAAA-MM-JJ" dans le fichier CSV	Aucune
	003 Type transaction	Type de transaction du bien immobilier (Une seule valeur : "Vente")	VARCHAR	10	E	jamais nulle	Aucune
	004 Valeur foncière du bien	Valeur de vente du bien immobilier	FLOAT	-	E	Peut être nulle	Aucune
	005 Acquéreur	Nom de l'acquéreur du bien immobilier	VARCHAR	50	E	jamais nulle	Aucune
	006 Identifiant de bien	Identifiant unique du bien immobilier  Un bien est identifié par son identifiant d'adresse et son numéro de PREMIER lot à l'adresse indiquée	VARCHAR	11	Ca	Clé étrangère, potentiellement nulle en cas de suppression d'un bien	identifiant du bien = adresse + numéro du 1er lot  Nous transcodons cet identifiant en code technique au format BIMnnnnnnn (avec nnnnnnn un entier codé du 8 positions; par exemple BIM00000021)

# Le schéma relationnel normalisé



# La base de données avec les tables créées et le typage des données

opc_laplace_immo	
Tables	
► adresses	
► biens_immobilier	
► communes	
► departements	
► regions	
► ventes	

## Table: ventes

Columns:	
<u>vente_id_pk</u>	varchar(10) PK
date_mutation	date
nature_mutation	varchar(10)
valeur_fonciere_euros	float
nom_acquéreur	varchar(50)
bien_immo_id_fk	varchar(11)

### Related Tables:

Target	biens_immobilier (bien_immo_id_fk → bien_immo_id_pk)
On Update	CASCADE
On Delete	SET NULL

## Table: regions

Columns:	
<u>reg_code_pk</u>	int PK
reg_nom	varchar(100)
reg_regroupement_nom	varchar(100)

Table: biens_immobilier	
Columns:	
<u>bien_immo_id_pk</u>	varchar(11) PK
section_prefixe	int
section	varchar(2)
num_pna	int
lot_01	varchar(8)
surface_carrez_lot_01	float
type_local_code	int
type_local	varchar(20)
surface_reelle_bati	int
nb_pieces_principales	int
nature_culture	varchar(10)
nature_culture_speciale	varchar(10)
surfaceTerrain	int
<u>adr_id_fk</u>	varchar(10)

### Related Tables:

Target	adresses (adr_id_fk → adr_id_pk)
On Update	CASCADE
On Delete	SET NULL

## Table: departements

Columns:	
<u>dep_code_pk</u>	varchar(3) PK
dep_nom	varchar(100)
<u>reg_code_fk</u>	int

### Related Tables:

Target	regions (reg_code_fk → reg_code_pk)
On Update	CASCADE
On Delete	SET NULL

## Table: adresses

Columns:	
<u>adr_id_pk</u>	varchar(10) PK
no_voie	varchar(6)
b_t_q	varchar(1)
type_voie_code	int
voie_type	varchar(10)
voie_code	varchar(10)
voie_nom	varchar(200)
postal_code	varchar(5)
<u>comm_id_fk</u>	varchar(10)

### Related Tables:

Target	communes (comm_id_fk → comm_id_pk)
On Update	CASCADE
On Delete	SET NULL

## Table: communes

Columns:	
<u>comm_id_pk</u>	varchar(10) PK
comm_code	int
comm_nom	varchar(100)
population_municipale	int
population_comptee_a_part	int
population_totale	int
<u>dep_code_fk</u>	varchar(3)

### Related Tables:

Target	departements (dep_code_fk → dep_code_pk)
On Update	CASCADE
On Delete	SET NULL

# Screenshot qui permettent de démontrer le bon chargement des données

	reg_code_pk	reg_nom	reg_regroupement_nom
▶	1	Guadeloupe	DROM-COM
	2	Martinique	DROM-COM
	3	Guyane	DROM-COM
	4	La Réunion	DROM-COM
	5	Collectivités d'outre-mer	DROM-COM
	6	Mayotte	DROM-COM
	11	Ile-de-France	Ile-de-France

19 row(s) returned

	dep_code_pk	dep_nom	reg_code_fk
▶	01	Ain	84
	02	Aisne	32
	03	Allier	84
	04	Alpes-de-Haute-Provence	93
	05	Hautes-Alpes	93
	06	Alpes-Maritimes	93
	07	Ardèche	84
	..	..	..

109 row(s) returned

	comm_id_pk	comm_code	comm_nom	population_municipale	population_comptee_a_part	population_totale	dep_code_fk
▶	01001	1	L'Abergement-Clémenciat	779	19	798	01
	01002	2	L'Abergement-de-Varey	256	1	257	01
	01004	4	Ambérieu-en-Bugey	14134	380	14514	01
	01005	5	Ambérieux-en-Dombes	1751	25	1776	01
	01006	6	Ambléon	112	6	118	01
	01007	7	Ambronay	2800	115	2915	01
	01008	8	Ambutrix	762	15	777	01

34991 row(s) returned

	adr_id_pk	no_voie	b_t_q	type_voie_code	voie_type	voie_code	voie_nom	postal_code	comm_id_fk
▶	ADR00000001	59		0	RUE	40	ALEXANDRE BERARD	1500	01004
	ADR00000002	54	B	0	RUE	120	ARISTIDE BRIAND	1500	01004
	ADR00000003	2		0	RUE	161	DE LA BATISSE	1500	01004
	ADR00000004	25		0	RUE	246	DE LA COMMUNE 1871	1500	01004
	ADR00000005	72		0	RUE	550	DE LA REPUBLIQUE	1500	01004
	ADR00000006	36		0	RUE	100	DES APOTRES	1500	01004
	ADR00000007	60		2	ALL	295	DES FRERES CAUDRON	1500	01004
	ADR00000008	70		4	AV	220	DU GENDARMAT	1500	01004

30793 row(s) returned

# Screenshots qui permettant de démontrer le bon chargement des données

bien_immo_id_pk	section_prefixe	section	num_pna	lot_01	surface_carrez_lot_01	type_local_code	type_local	surface_reelle_bati	nb_pieces_principales	nature_culture	nature_c
BIMO0000001	NULL	AO	539	3	164.87	2	Appartement	143	6	NULL	NULL
BIMO0000002	NULL	AO	1057	10	70.45	2	Appartement	75	3	NULL	NULL
BIMO0000003	NULL	BR	443	26	81.15	2	Appartement	81	3	NULL	NULL
BIMO0000004	NULL	AP	963	62	70.1	1	Maison	68	3	NULL	NULL
BIMO0000005	NULL	AO	739	1	35.59	2	Appartement	33	2	NULL	NULL
BIMO0000006	NULL	AW	333	10	74.72	2	Appartement	79	2	NULL	NULL
BIMO0000007	NULL	PP	212	101	50.64	2	Appartement	54	2	NULL	NULL

34169 row(s) returned

vente_id_pk	date_mutation	nature_mutation	valeur_fonciere_euros	nom_acquereur	bien_immo_id_fk
VTE0000001	2020-02-01	Vente	50000	ABDERRAHMANE	BIMO0004737
VTE0000002	2020-02-01	Vente	139500	BACLE	BIMO0032722
VTE0000003	2020-02-01	Vente	150000	BARBEREAU	BIMO0022506
VTE0000004	2020-02-01	Vente	105000	BEAUCOURT	BIMO0022059
VTE0000005	2020-02-01	Vente	290000	BEAUMONT	BIMO0012013
VTE0000006	2020-02-01	Vente	86000	BOOUNOUA	BIMO0026356
VTE0000007	2020-02-01	Vente	29000	CHATELET	BIMO0024172
VTE0000008	2020-02-01	Vente	25000	CHERIF	BIMO0022215

34169 row(s) returned



# Requêtes SQL et résultats

Sous-titre

# Requête 1

## Nombre total d'appartements vendus au 1<sup>er</sup> semestre 2020

```
/* ===== */
/* Requête 1 : Nombre total d'appartements vendus au 1er semestre 2020. */
/* ===== */
SELECT COUNT(VTE.vente_id_pk) AS 'Nombre total d''appartements vendus (2020/S1)'
FROM Biens_immobiliers BIM
LEFT JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
WHERE (VTE.date_mutation IS NULL OR (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')))
AND BIM.type_local = 'Appartement'
;
```

	Nombre total d'appartements vendus (2020/S1)
▶	31 378

1 row(s) returned

# Requête 2

## Nombre de ventes d'appartement par région(2020/S1)

```
/* ===== */
/* Requête 2 : Le nombre de ventes d'appartement par région pour le 1er semestre 2020 */
/* ===== */
SELECT REG.reg_nom AS 'Région',
       LPAD(COUNT(VTE.vente_id_pk), 10, '_') AS 'Nb appartements vendus par région (2020/S1)'
FROM Regions REG
LEFT JOIN Départements DEP ON DEP.reg_code_fk = REG.reg_code_pk
LEFT JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
LEFT JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
LEFT JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
LEFT JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
WHERE ( VTE.vente_id_pk IS NULL
      OR ( VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d')
           AND STR_TO_DATE('2020-06-30', '%Y-%m-%d'))
      )
      AND (BIM.bien_immo_id_pk IS NULL OR BIM.type_local = 'Appartement')
GROUP BY REG.reg_code_pk
ORDER BY COUNT(VTE.vente_id_pk) DESC , REG.reg_nom ASC
;
```

Région	Nombre d'appartements vendus par région (2020/S1)
Ile-de-France	13995
Provence-Alpes-Côte d'Azur	3649
Auvergne-Rhône-Alpes	3253
Nouvelle-Aquitaine	1932
Occitanie	1640
Pays de la Loire	1357
Hauts-de-France	1254
Grand Est	984
Bretagne	983
Normandie	862
Centre-Val de Loire	696
Bourgogne-Franche-Comté	376
Corse	223
Martinique	94
La Réunion	44
Guyane	34
Guadeloupe	2
Collectivités d'outre-mer	0
Mayotte	0

19 row(s) returned

# Requête 3

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

```
/*
 * =====*
 * Requête 3 : % d'appartements vendus par nombre de pièces (2020/S1)
 * =====*
 WITH VENTES_CONSOLIDEES_VIEW AS
(
    SELECT COUNT(VTE.vente_id_pk) as 'NB_TOTAL_APPARTEMENTS_VENDUS'
    FROM Biens_immobiliers BIM
    LEFT JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
    WHERE (VTE.date_mutation IS NULL OR (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')))
        AND (BIM.type_local = 'Appartement')
)
SELECT LPAD(BIM.nb_pieces_principales, 10, '_') AS 'Nombre de pièces principales',
    LPAD(COUNT(VTE.vente_id_pk), 10, '_') AS 'Nombre d'appartements vendus par pièce',
    LPAD(REPLACE(FORMAT(VENTES_CONSOLIDEES_VIEW.NB_TOTAL_APPARTEMENTS_VENDUS, 0), ',', ','), 10, '_') AS 'Nombre total d'appartements vendus',
    LPAD((CONCAT(ROUND(COUNT(VTE.vente_id_pk) / VENTES_CONSOLIDEES_VIEW.NB_TOTAL_APPARTEMENTS_VENDUS * 100, 1), "%")), 10, '_') AS '% de ventes par pièce'
FROM Biens_immobiliers BIM LEFT JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk,
/* ATTENTION PRODUIT CARTESIEN ENTRE LA JOINTURE CI-DESSUS ET LA VUE DEFINIE DANS LA CLAUSE WITH ( CETTE VUE NE RENVOIT QU'UNE SEULE LIGNE ) */
VENTES_CONSOLIDEES_VIEW
WHERE (VTE.date_mutation IS NULL OR (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')))
    AND BIM.type_local = 'Appartement'
GROUP BY BIM.type_local, VENTES_CONSOLIDEES_VIEW.NB_TOTAL_APPARTEMENTS_VENDUS, BIM.nb_pieces_principales
ORDER BY BIM.type_local ASC, BIM.nb_pieces_principales
;
```

Nombre de pièces principales	Nombre d'appartements vendus par pièce	Nombre total d'appartements vendus	% de ventes par pièce
0	30	31 378	0.1%
1	6739	31 378	21.5%
2	9783	31 378	31.2%
3	8966	31 378	28.6%
4	4460	31 378	14.2%
5	1114	31 378	3.6%
6	204	31 378	0.7%
7	54	31 378	0.2%
8	17	31 378	0.1%
9	8	31 378	0.0%
10	2	31 378	0.0%
11	1	31 378	0.0%

# Requête 4

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

```
/* ===== */
/* Requête 4 : Liste des 10 départements où le prix du mètre carré est le plus élevé */
/*     Prix au mètre carré d'un bien = Prix de vente du bien / Surface */
/*     Prix moyen par département = moyenne (prix au mètre carré de chaque bien du département) */
/* ===== */

SELECT DEP.dep_nom AS 'Département',
       REPLACE(LPAD(CONCAT(ROUND(AVG(VTE.valeur_fonciere_euros / BIM.surface_reelle_bati), 0), ' €'), 30, '_'), '_', '') 'Prix moyen au M2'
    , COUNT(VTE.vente_id_pk) 'Nombre de biens'
  FROM Departements DEP
  LEFT OUTER JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
  LEFT OUTER JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
  LEFT OUTER JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
  LEFT OUTER JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
 WHERE (VTE.vente_id_pk IS NULL OR (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')))
 GROUP BY DEP.dep_nom
 ORDER BY AVG(ROUND(VTE.valeur_fonciere_euros / BIM.surface_reelle_bati, 0)) DESC, DEP.dep_nom ASC
LIMIT 10
;
```

Département	Prix moyen au M2
Paris	12129 €
Hauts-de-Seine	7415 €
Val-de-Marne	5399 €
Alpes-Maritimes	4685 €
Seine-Saint-Denis	4371 €
Haute-Savoie	4150 €
Yvelines	4126 €
Rhône	4064 €
Corse-du-Sud	3922 €
Gironde	3839 €

10 row(s) returned

# Requête 5

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

```
/*
 * Requête 5 : Prix moyen du mètre carré d'une maison en Île-de-France
 */
SELECT REG.reg_nom AS 'Region',
       CONCAT(REPLACE(FORMAT(ROUND(AVG(VTE.valeur_fonciere_euros / BIM.surface_reelle_bati), 0), 0), ',', ','), ' €') AS 'Prix moyen d''une maison au M2'
FROM Regions REG
     LEFT OUTER JOIN Departements DEP ON DEP.reg_code_fk = REG.reg_code_pk
     LEFT OUTER JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
     LEFT OUTER JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
     LEFT OUTER JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
     LEFT OUTER JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
WHERE (VTE.vente_id_pk IS NULL OR (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')))
      AND BIM.type_local = 'Maison'
      AND REG.reg_nom = 'Île-de-France'
GROUP BY REG.reg_nom
;
```

	Region	Prix moyen d'une maison au M2
▶	Île-de-France	3 998 €

1 row(s) returned

# Requête 6

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

```
/*
=====
* Requête 6 : Liste des 10 appartements les plus chers avec la région et le nombre de mètres carrés
=====
SELECT REG.reg_nom AS 'Région',
       CONCAT(CONCAT(CONCAT(CONCAT(CONCAT(CONCAT(CONCAT(CONCAT(
           CONCAT(ADR.no_voie, ' '), ADR.b_t_q), ' '), ADR.voie_type), ' '), ADR.voie_nom), ' '), ADR.postal_code), ' '), COM.comm_nom), ' ('), DEP.dep_nom), ')') AS 'Appartement',
       CONCAT(REPLACE(FORMAT(VTE.valeur_fonciere_euros, 0), ',', ','), '€') AS 'Valeur de vente',
       CONCAT(LPAD(BIM.surface_reelle_bati, 10, ' '), ' m2') AS 'Surface réelle bati'
FROM Regions REG
     INNER JOIN Departements DEP ON DEP.reg_code_fk = REG.reg_code_pk
     INNER JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
     INNER JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
     INNER JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
     INNER JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
WHERE (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d'))
      AND BIM.type_local = 'Appartement'
ORDER BY VTE.valeur_fonciere_euros DESC
LIMIT 10
;
```

	Région	Appartement	Valeur de vente	Surface réelle bati
►	Ile-de-France	6 , BD SUCHET 75016 Paris 16e Arrondissement (Paris)	9 000 000€	10 m <sup>2</sup>
	Ile-de-France	16 , CHE DE LA CAVIGNON 91100 Corbeil-Essonnes (Essonne)	8 600 000€	62 m <sup>2</sup>
	Ile-de-France	104 , RUE DU BAC 75007 Paris 7e Arrondissement (Paris)	8 577 713€	289 m <sup>2</sup>
	Ile-de-France	33 , RUE LEMERCIER 75017 Paris 17e Arrondissement (Paris)	7 620 000€	42 m <sup>2</sup>
	Ile-de-France	72 , RUE D ASSAS 75006 Paris 6e Arrondissement (Paris)	7 600 000€	200 m <sup>2</sup>
	Ile-de-France	8 , RUE SAINT HYACINTHE 75001 Paris 1er Arrondissement (Paris)	7 535 000€	143 m <sup>2</sup>
	Ile-de-France	36 , AV GEORGES MANDEL 75016 Paris 16e Arrondissement (Paris)	7 420 000€	357 m <sup>2</sup>
	Ile-de-France	23 , BD DE BEAUSEJOUR 75016 Paris 16e Arrondissement (Paris)	7 200 000€	241 m <sup>2</sup>
	Ile-de-France	26 , RUE CAMBON 75001 Paris 1er Arrondissement (Paris)	7 050 000€	310 m <sup>2</sup>
	Ile-de-France	108 , RUE SAINT HONORE 75001 Paris 1er Arrondissement (Paris)	6 600 000€	76 m <sup>2</sup>

10 row(s) returned

# Requête 7

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

```
/* ===== */
/* Requête 7 : Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020 */
/* ===== */
WITH VTE_EVOLUTION_T1_T2 AS
(
    SELECT COUNT(VTE.vente_id_pk) AS Ventes_2020_T1,
           0 AS Ventes_2020_T2
      FROM Ventes VTE
     WHERE VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d') AND STR_TO_DATE('2020-03-31', '%Y-%m-%d')
UNION
    SELECT 0 AS Ventes_2020_T1,
           COUNT(VTE.vente_id_pk) AS Ventes_2020_T2
      FROM Ventes VTE
     WHERE VTE.date_mutation BETWEEN STR_TO_DATE('2020-04-01', '%Y-%m-%d') AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')
)
SELECT REPLACE(FORMAT(SUM(VTE_EVOLUTION_T1_T2.Ventes_2020_T1), 0), ',', '') AS 'Ventes du premier trimestre',
REPLACE(FORMAT(SUM(VTE_EVOLUTION_T1_T2.Ventes_2020_T2), 0), ',', '') AS 'Ventes du second trimestre',
CONCAT(ROUND(( (SUM(VTE_EVOLUTION_T1_T2.Ventes_2020_T2) - SUM(VTE_EVOLUTION_T1_T2.Ventes_2020_T1)) / SUM(VTE_EVOLUTION_T1_T2.Ventes_2020_T1)) * 100, 1),
        '%'
      ) AS 'Taux d''évolution'
   FROM VTE_EVOLUTION_T1_T2
;
```

	Ventes du premier trimestre	Ventes du second trimestre	Taux d'évolution
▶	16 776	17 393	3.7%

1 row(s) returned

# Requête 8

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

Region	Prix moyen au M2
Île-de-France	8 015 €
La Réunion	3 660 €
Corse	3 046 €
Provence-Alpes-Côte d'Azur	3 005 €
Auvergne-Rhône-Alpes	2 769 €
Nouvelle-Aquitaine	2 510 €
Bretagne	2 272 €
Hauts-de-France	2 204 €
Pays de la Loire	2 187 €
Occitanie	2 096 €
Normandie	1 994 €
Centre-Val de Loire	1 429 €
Grand Est	1 313 €
Bourgogne-Franche-Comté	1 069 €
Martinique	564 €
Guadeloupe	Aucune vente enregistrée
Guyane	Aucune vente enregistrée
Collectivités d'outre-mer	Aucune vente enregistrée
Mayotte	Aucune vente enregistrée

19 row(s) returned

# Requête 9

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

```
/*
/* Requête 9 : Liste des communes ayant eu au moins 50 ventes au 1er trimestre */
/*
SELECT COM.comm_nom AS 'Commune',
       DEP.dep_nom AS 'Département',
       REG.reg_nom AS 'Région',
       LPAD(COUNT(VTE.vente_id_pk),10,'_') AS 'Nombre de ventes'
FROM Regions REG
     RIGHT OUTER JOIN Departements DEP ON DEP.reg_code_fk = REG.reg_code_pk
     RIGHT OUTER JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
     INNER JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
     INNER JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
     INNER JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
WHERE (VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d')
          AND STR_TO_DATE('2020-03-31', '%Y-%m-%d'))
GROUP BY REG.reg_code_pk, DEP.dep_code_pk, COM.comm_id_pk
HAVING COUNT(VTE.vente_id_pk) >= 50
ORDER BY COUNT(VTE.vente_id_pk) DESC, COM.comm_id_pk
;
```

Commune	Département	Région	Nombre de ventes
Paris 17e Arrondissement	Paris	Ile-de-France	228
Paris 15e Arrondissement	Paris	Ile-de-France	215
Paris 18e Arrondissement	Paris	Ile-de-France	209
Nice	Alpes-Maritimes	Provence-Alpes-Côte d'Azur	173
Paris 11e Arrondissement	Paris	Ile-de-France	169
Paris 16e Arrondissement	Paris	Ile-de-France	165
Bordeaux	Gironde	Nouvelle-Aquitaine	157
Paris 14e Arrondissement	Paris	Ile-de-France	146
Paris 20e Arrondissement	Paris	Ile-de-France	127
Nantes	Loire-Atlantique	Pays de la Loire	119
Paris 19e Arrondissement	Paris	Ile-de-France	116
Paris 12e Arrondissement	Paris	Ile-de-France	110
Paris 10e Arrondissement	Paris	Ile-de-France	109
Grenoble	Isère	Auvergne-Rhône-Alpes	106
Paris 9e Arrondissement	Paris	Ile-de-France	106
Boulogne-Billancourt	Hauts-de-Seine	Ile-de-France	99
Paris 13e Arrondissement	Paris	Ile-de-France	94
Paris 7e Arrondissement	Paris	Ile-de-France	87
Paris 6e Arrondissement	Paris	Ile-de-France	86
Marseille 8e Arrondissem...	Bouches-du-Rhône	Provence-Alpes-Côte d'Azur	81
Asnières-sur-Seine	Hauts-de-Seine	Ile-de-France	81
Courbevoie	Hauts-de-Seine	Ile-de-France	80
Paris 3e Arrondissement	Paris	Ile-de-France	79
Paris 5e Arrondissement	Paris	Ile-de-France	79
Toulouse	Haute-Garonne	Occitanie	78
Antibes	Alpes-Maritimes	Provence-Alpes-Côte d'Azur	77

48 row(s) returned

# Requête 10

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

**Autrement formulé : Différence en % du prix moyen au mètre carré d'un appartement de 3 pièces par rapport à un appartement de 2 pièces**

**La formule appliquée pour obtenir le % est la suivante :**

### **Différence 3 pièces – 2 pièces**

1

Prix moyen M2 2pièces

Soit -13,04%

Prix moyen au M2 appartements 3 pièces	Prix moyen au M2 appartements 2 pièces	Définition 3 pièces / 2 pièces	%définition 3 pièces / 2 pièces
► 4 285 €	4 928 €	-642 €	-13.04%

1 row(s) returned

**Le prix au mètre carré d'un appartement de 3 pièces est en moyenne inférieur de 13.04% à celui d'un appartement de 2 pièces**

# Requête 11

## Moyennes des valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69

```
/*
 * Requête 11 : moyenne des valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69 */
WITH VTE_VIEW AS
(
    SELECT RANK() OVER (
        PARTITION BY DEP.dep_code_pk ORDER BY AVG(VTE.valeur_fonciere_euros) DESC
    ) AS RANG_PAR_DEP,
    DEP.dep_code_pk AS CODE_DEPARTEMENT,
    DEP.dep_nom AS NOM_DEPARTEMENT,
    COM.comm_nom AS NOM_COMMUNE,
    AVG(VTE.valeur_fonciere_euros) AS VALEUR_FONCIERE_MOYENNE
    FROM Departements DEP
    RIGHT OUTER JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
    RIGHT OUTER JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
    RIGHT OUTER JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
    RIGHT OUTER JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
    WHERE VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d')
        AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')
        AND DEP.dep_code_pk IN ('06', '13', '33', '59', '69')
    GROUP BY DEP.dep_code_pk, COM.comm_id_pk
)
SELECT CONCAT(VTE_VIEW.CODE_DEPARTEMENT, CONCAT(' - ', VTE_VIEW.NOM_DEPARTEMENT)) AS 'Département',
    VTE_VIEW.NOM_COMMUNE AS 'Commune',
    LPAD ( CONCAT(REPLACE(FORMAT(ROUND(VTE_VIEW.VALEUR_FONCIERE_MOYENNE, 0), 0), ',', '.'), ' €'), 30,
        ) AS 'Valeur foncière moyenne'
    FROM VTE_VIEW
    WHERE VTE_VIEW.RANG_PAR_DEP <= 3
    ORDER BY VTE_VIEW.CODE_DEPARTEMENT ASC, VTE_VIEW.VALEUR_FONCIERE_MOYENNE DESC, VTE_VIEW.NOM_COMMUNE ASC
```

	Département	Commune	Valeur foncière moyenne
▶	06 - Alpes-Maritimes	Saint-Jean-Cap-Ferrat	968 750 €
	06 - Alpes-Maritimes	Eze	655 000 €
	06 - Alpes-Maritimes	Mouans-Sartoux	476 898 €
	13 - Bouches-du-Rhône	Gignac-la-Nerthe	330 000 €
	13 - Bouches-du-Rhône	Saint-Savournin	314 425 €
	13 - Bouches-du-Rhône	Cassis	313 417 €
	33 - Gironde	Lège-Cap-Ferret	549 501 €
	33 - Gironde	Vayres	335 000 €
	33 - Gironde	Arcachon	307 436 €
	59 - Nord	Bersée	433 202 €
	59 - Nord	Cysoing	408 550 €
	59 - Nord	Halluin	322 250 €
	69 - Rhône	Ville-sur-Jarnioux	485 300 €
	69 - Rhône	Lyon 2e Arrondissement	455 217 €
	69 - Rhône	Lyon 6e Arrondissement	426 968 €

15 row(s) returned

# Requête 12

## Liste des 20 communes de plus de 10 000 habitants qui ont la plus de transactions pour 1000 habitants

```
/*
/* Requête 12 : Les 20 communes avec le plus de transactions pour 1000 habitants pour les communes qui dépassent les 10 000 habitants */
SELECT    CONCAT(DEP.dep_code_pk, CONCAT(' - ', DEP.dep_nom)) AS 'Département',
          COM.comm_nom AS 'Commune',
          LPAD(ROUND(COM.population_totale / 1000, 0), 4, ' ') AS 'Population (en milliers d''habitants)',
          LPAD(COUNT(VTE.vente_id_pk), 3, ' ') 'Nombre de ventes',
          LPAD(ROUND(COUNT(VTE.vente_id_pk) / (COM.population_totale / 1000), 2), 5, ' ') AS 'Nb transactions pour mille habitants'
FROM      Departements DEP
          RIGHT OUTER JOIN Communes COM ON COM.dep_code_fk = DEP.dep_code_pk
          RIGHT OUTER JOIN Adresses ADR ON ADR.comm_id_fk = COM.comm_id_pk
          RIGHT OUTER JOIN Biens_immobiliers BIM ON BIM.adr_id_fk = ADR.adr_id_pk
          RIGHT OUTER JOIN Ventes VTE ON VTE.bien_immo_id_fk = BIM.bien_immo_id_pk
WHERE     VTE.date_mutation BETWEEN STR_TO_DATE('2020-01-01', '%Y-%m-%d')
          AND STR_TO_DATE('2020-06-30', '%Y-%m-%d')
          AND COM.population_totale > 100000
GROUP BY  DEP.dep_code_pk, COM.comm_id_pk
ORDER BY  (COUNT(VTE.vente_id_pk) / (COM.population_totale / 1000)) DESC
LIMIT 20
```

	Département	Commune	Population (en milliers d'habitants)	Nombre de ventes	Nb transactions pour mille habitants
▶	75 - Paris	Paris 17e Arrondissement	168 Kh	470	2.80 %
	75 - Paris	Paris 18e Arrondissement	193 Kh	516	2.67 %
	75 - Paris	Paris 11e Arrondissement	146 Kh	383	2.62 %
	75 - Paris	Paris 16e Arrondissement	168 Kh	394	2.35 %
	75 - Paris	Paris 12e Arrondissement	140 Kh	254	1.81 %
	75 - Paris	Paris 15e Arrondissement	233 Kh	407	1.75 %
	75 - Paris	Paris 14e Arrondissement	137 Kh	222	1.63 %
	75 - Paris	Paris 20e Arrondissement	196 Kh	303	1.55 %
	75 - Paris	Paris 19e Arrondissement	186 Kh	267	1.44 %
	38 - Isère	Grenoble	160 Kh	226	1.41 %
	93 - Seine-Saint-Denis	Montreuil	112 Kh	138	1.23 %
	33 - Gironde	Bordeaux	264 Kh	302	1.14 %
	06 - Alpes-Maritimes	Nice	346 Kh	393	1.14 %
	83 - Var	Toulon	181 Kh	182	1.01 %
	30 - Gard	Nîmes	151 Kh	147	0.97 %
	76 - Seine-Maritime	Rouen	114 Kh	110	0.97 %
	54 - Meurthe-et-Moselle	Nancy	107 Kh	102	0.96 %
	44 - Loire-Atlantique	Nantes	324 Kh	299	0.92 %
	69 - Rhône	Lyon 3e Arrondissement	104 Kh	89	0.86 %
	45 - Loiret	Orléans	119 Kh	97	0.82 %

20 row(s) returned

# Synthèse de l'étude (1/2)

## Tendances observées :

- ✓ Des ventes essentiellement concentrées sur Paris (13 995 biens vendus pour un total national de 31 378, soit (44%) . Vers une diversification du marché ?
- ✓ Un prix moyen au M2 très largement supérieur à Paris (12 129 €), avec pour second les Hauts de Seine (7 415 €). Vers une concentration sur le marché parisien ?
- ✓ Des ventes sous-représentées dans les DOM-TOM (<100 ventes par département). Les investir ?
- ✓ Des ventes d'appartements concentrées sur les 2 (31.2%) et 3 pièces (28.6%)
- ✓ Une évolution substantielle du nombre de ventes de 2020/T1 à 2020/T2 (+3.7%)
- ✓ Des incohérences observées sur les données :
  - Vente de 30 appartements avec 0 pièces
  - Des biens vendus entre 6M€ et 9M€ avec une surface comprise entre 10m2 et 76m2 (dont 10m2 vendus pour 9M€ à Paris 16<sup>e</sup>)

# Synthèse de l'étude (2/2)

## Limites de l'étude :

### ✓ Fonctionnelles :

- Exploitation des données de géolocalisation (cartographie du nombre de ventes et des prix moyens au m<sup>2</sup>)
- Détection et analyse exhaustive des inconsistances de données
- Elargissement de l'échantillon sur 2 ans (évolution trimestrielle, évolution annuelle)
- Modélisation des cas d'usage suivants : plusieurs ventes d'un même bien, ventes englobant plusieurs biens

### ✓ Techniques :

- Tests de sauvegarde
- Typage des données numériques (int, short, long, float, long, signed, unsigned), cryptage des données sensibles (nom de l'acquéreur)
- Restitution des données numériques alignées à droite avec des espaces à gauche (au lieu du caractère '\_')



Merci !