



Laplace Immo

# Contexte

- Data Analyst
- Projet « DATAImmo » : Analyse du marché de l'immobilier

# Ressources

- Données extraites du site open data des demandes de valeur foncière
  - Template d'un dictionnaire
  - Compte-rendu de réunion

# Besoins

- Conception la BDD
- Création Base d'une données
- Elaboration requêtes SQL

# Organisation

1 - Analyse des donnée

2 - Création du dictionnaire

3 - Création d'un MCD, MLD et MPD

4 - Organisation des données par tables

5 - Création de la BDD et insertion de données

6 – Analyse de données via requêtes SQL

# 1 - Analyse des données

- Analyse et compréhension des données
- Choix des données en fonction du besoin

	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	No
1	No disposition	Date mutation	Nature mutation	Valeur foncière	No voie	B/T/Q	Code type de voie	Type de voie	Code voie	Voie	Code ID commune	Code postal	Commune	Code département	Code commune	Prefixe de section	Section	No
2	1	03/02/2020	Vente	56000	190	A	0	RUE	5	CENTRALE	0	1370.0	SAINT-ETIENNE-DU-BOIS	1	350		B	
3	1	02/01/2020	Vente	165000	347		0	RUE	20	DU CHATEAU	1	1170.0	CHEVRY	1	103		A	
4	1	08/01/2020	Vente	720000	58		1	AV	527	DU MONT BLANC	2	1220.0	DIVONNE-LES-BAINS	1	143		AK	
5	1	06/01/2020	Vente	429250	140		0	RUE	2	DE L'ABBE JOLIVET	3	1630.0	PERON	1	288		C	
6	1	07/01/2020	Vente	220900	39		0	RUE	110	BUFFON	4	1200.0	VALSERHONE	1	33		AE	
7	1	21/01/2020	Vente	42000	28		1	AV	179	JEAN FALCONNIER	5	1350.0	CULOZ	1	138		AP	
8	1	07/01/2020	Vente	262000	8		0	RUE	120	DE GENEVE	6	1630.0	ST-GENIS-POUILLY	1	354		BH	
9	1	08/01/2020	Vente	190000	2		0	RUE	210	DU RECULET	6	1630.0	ST-GENIS-POUILLY	1	354		AX	
10	1	16/01/2020	Vente	563130	1403		0	RUE	465	JEAN DE GINGINS	2	1220.0	DIVONNE-LES-BAINS	1	143		H	
11	1	17/01/2020	Vente	535000	226		2	ALL	36	DES CAPUCINES	6	1630.0	ST-GENIS-POUILLY	1	354		AB	
12	1	16/01/2020	Vente	330000	276		3	RTE	210	DE POUIGNY	3	1630.0	PERON	1	288		F	
13	1	27/01/2020	Vente	110600	79		4	CRS	1490	DE VERDUN	7	1100.0	OYONNAX	1	283		AN	
14	1	30/01/2020	Vente	50000	77	B	0	RUE	20	DU COMMERCE	8	1130.0	ST-GERMAIN-DE-JOUX	1	357		AB	
15	1	09/01/2020	Vente	212000	240		0	RUE	650	DE PRE BAILLY	9	1170.0	GEX	1	173		AM	
16	1	15/01/2020	Vente	160000	3		0	RUE	541	TURENNE	4	1200.0	VALSERHONE	1	33		AI	
17	1	20/01/2020	Vente	561550	44		2	ALL	695	DU SQUARE DE LAUSANNE	2	1220.0	DIVONNE-LES-BAINS	1	143		AE	
18	1	28/01/2020	Vente	358000	59		0	RUE	40	ALEXANDRE BERARD	10	1500.0	AMBERIEU-EN-BUGEY	1	4		AO	
19	1	22/01/2020	Vente	220000	282		5	CHE	130	DES LONGES RAYES	11	1170.0	CESSY	1	71		AT	
20	1	27/01/2020	Vente	79300	54		6	GR	45	GRANDE RUE	12	1150.0	SAULT-BRENAZ	1	396		AD	
21	1	06/01/2020	Vente	285000	1		0	RUE	160	DE GEX	13	1210.0	FERNEY-VOLTAIRE	1	160		AE	
22	1	27/01/2020	Vente	123100	17		7	LOT	A036	LES JARDINS DE CHEVRY	1	1170.0	CHEVRY	1	103		B	
23	1	16/01/2020	Vente	107500	329		0	RUE	25	DES CARPIERES	14	1170.0	SENGY	1	399		AA	
24	1	30/01/2020	Vente	128810	369		1	AV	55	CHARLES DE GAULLE	15	1150.0	LAGNIEU	1	202		F	
25	1	05/02/2020	Vente	31000	154		0	RUE	33	SUR LA VIGNETTE	16	1130.0	LE POIZAT-LALLEYRIAT	1	204	300	AI	
26	1	21/01/2020	Vente	125000	6		0	RUE	745	DES USINIERS	9	1170.0	GEX	1	173		AC	
27	1	31/01/2020	Vente	285000	2		0	RUE	135	DU JURA	6	1630.0	ST-GENIS-POUILLY	1	354		BC	
28	1	03/02/2020	Vente	140500	54		0	RUE	455	MARCEL DEMIA	10	1500.0	AMBERIEU-EN-BUGEY	1	4		BP	
29	1	27/01/2020	Vente	101050	269		5	CHE	191	DE CHARIGNIN	17	1300.0	BELLEY	1	34		BK	
30	1	31/01/2020	Vente	530000	322		0	RUE	37	DE BRETAGNY	18	1210.0	ORNEX	1	281		AH	
31	1	29/01/2020	Vente	486415	132		2	ALL	124	DES HETRES	6	1630.0	ST-GENIS-POUILLY	1	354		BD	
32	1	28/01/2020	Vente	185000	80		0	RUE	170	DE MEYRIN	13	1210.0	FERNEY-VOLTAIRE	1	160		AE	
33	1	16/01/2020	Vente	131376	4		0	RUE	190	DE LA MAIRIE	5	1350.0	CULOZ	1	138		AO	
34	1	23/01/2020	Vente	155000	79		6	GR	182	GRANDE RUE DE BULLIEZ	15	1150.0	LAGNIEU	1	202		F	
35	1	31/01/2020	Vente	150000	8		0	RUE	15	DE LA BIENNE	19	1590.0	DORTAN	1	148		AC	
36	1	03/02/2020	Vente	67000	29		0	RUE	540	REINE CLOTILDE	10	1500.0	AMBERIEU-EN-BUGEY	1	4		BM	
37	1	13/01/2020	Vente	114000	3		0	RUE	730	STE MARIE	17	1300.0	BELLEY	1	34		AH	
38	1	07/02/2020	Vente	143850	956		3	RTE	25	DE BOURG	15	1150.0	LAGNIEU	1	202		A	

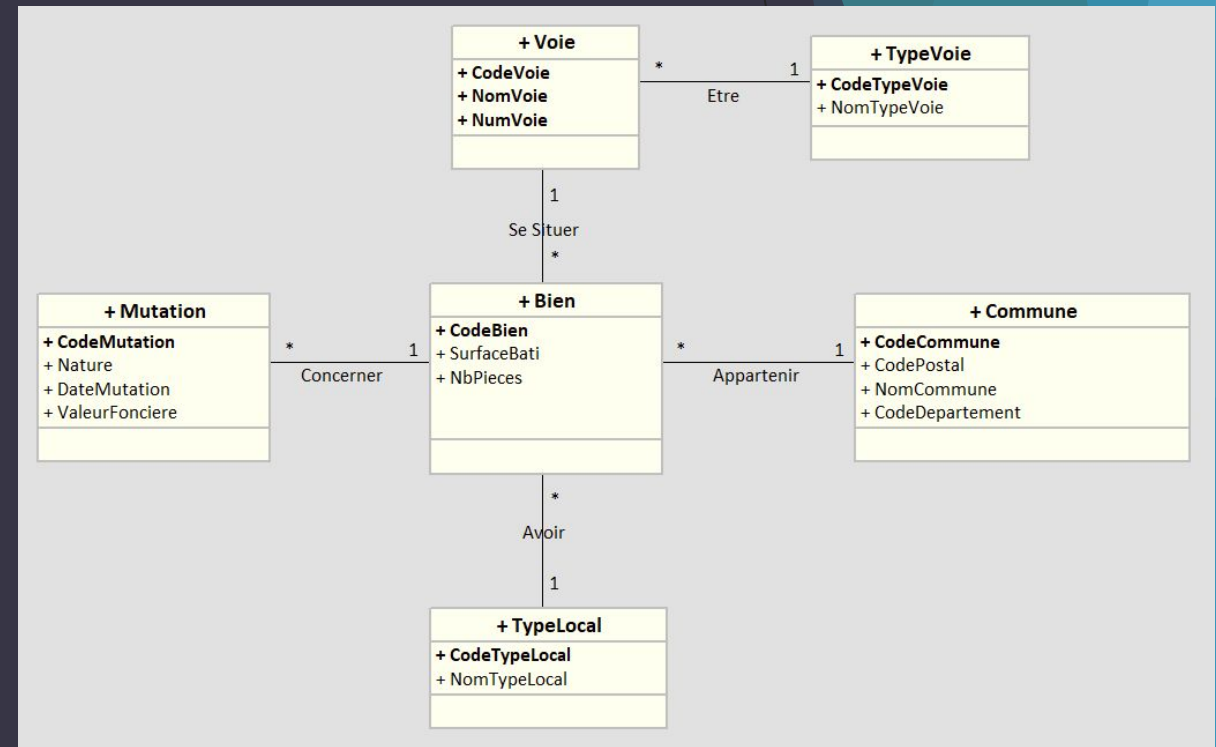
## 2 - Création du dictionnaire

- Report des données importantes dans le dictionnaire
  - Regroupement des données par thèmes

	A	B	C	D	E	F
1	Numéro	Code propriété	Signification	Type	Observation	Theme
2	1	<b>CodeMutation</b>	Identifiant de la mutation	Numérique ou alphanumérique	Longueur : 5	Mutation
3	2	DateMutation	Date mutation	Date	aaaa/mm/jj	Mutation
4	3	Nature	Nature mutation	Texte	Longueur : 20	Mutation
5	4	ValeurFonciere	Valeur foncière	Monétaire		Mutation
6	5	<b>CodeTypeVoie</b>	Code Type de Voie	Numérique ou alphanumérique	Longueur : 3	TypeVoie
7	6	NomTypeVoie	Type voie	Texte	Longueur : 10	TypeVoie
8	7	<b>CodeVoie</b>	Identifiant de la voie	Numérique ou alphanumérique	Longueur : 5	Voie
9	8	NumVoie	Numéro de voie	Numérique ou alphanumérique	Longueur : 3	Voie
10	9	NomVoie	Nom de la voie	Texte	Longueur : 100	Voie
11	10	<b>CodeCommune</b>	Code de la Commune	Numérique ou alphanumérique	Longueur : 3	Commune
12	11	CodePostal	Code Postal	Numérique ou alphanumérique	Longueur : 5	Commune
13	12	NomCommune	Nom de la commune	Texte	Longueur : 100	Commune
14	13	CodeDepartement	Code du département	Numérique ou alphanumérique	Longueur : 5	Commune
15	14	<b>CodeBien</b>	Identifiant du Bien	Numérique ou alphanumérique	Longueur : 5	Bien
16	15	SurfaceBati	Surface Batti	Numérique ou alphanumérique	Longueur : 5	Bien
17	16	NbPieces	Nombre de pieces	Numérique ou alphanumérique	Longueur : 5	Bien
18	17	<b>CodeTypeLocal</b>	Code Type Local	Numérique ou alphanumérique	Longueur : 3	TypeLocal
19	18	NomTypeLocal	Type local	Texte	Longueur : 20	TypeLocal
20						

### 3 - Création du MCD

- Utilisation de l'outil Looping
- Création des table et liens grâce au dictionnaire de données

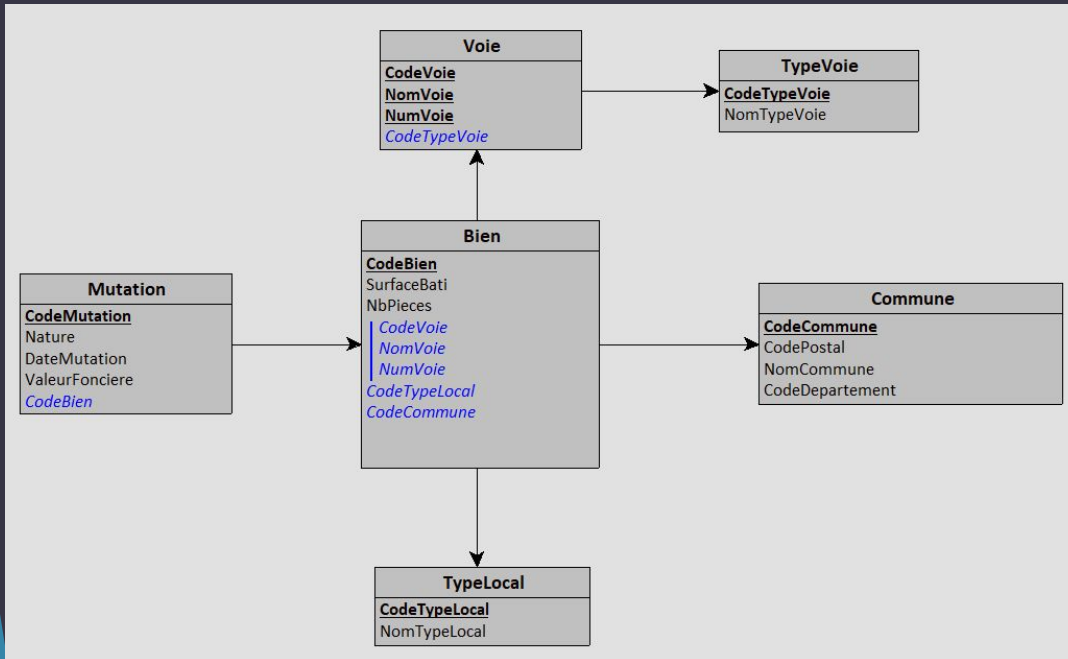




# 3 - Création du MLD et MPD

- Utilisation de l'outil Looping pour transformé le MCD en MLD puis en MPD

## MLD



## MPD

```
1 CREATE TABLE Commune(  
2     CodeCommune INT,  
3     CodePostal INT,  
4     NomCommune VARCHAR(100),  
5     CodeDepartement VARCHAR(50),  
6     PRIMARY KEY(CodeCommune)  
7 );  
8  
9 CREATE TABLE TypeVoie(  
10     CodeTypeVoie INT,  
11     NomTypeVoie VARCHAR(50),  
12     PRIMARY KEY(CodeTypeVoie)  
13 );  
14  
15 CREATE TABLE TypeLocal(  
16     CodeTypeLocal INT,  
17     NomTypeLocal VARCHAR(20),  
18     PRIMARY KEY(CodeTypeLocal)  
19 );  
20  
21 CREATE TABLE Voie(  
22     CodeVoie VARCHAR(50),  
23     NomVoie VARCHAR(100),  
24     NumVoie INT,  
25     CodeTypeVoie INT NOT NULL,  
26     PRIMARY KEY(CodeVoie, NomVoie, NumVoie, CodeTypeVoie),  
27     FOREIGN KEY(CodeTypeVoie) REFERENCES TypeVoie(CodeTypeVoie)  
28 );
```

# 4 - Organisation des données par tables

Exemple avec la Table Voie

Voie	
<u>CodeVoie</u>	VARCHAR(50)
<u>NomVoie</u>	VARCHAR(100)
<u>NumVoie</u>	INT
<u>CodeTypeVoie</u>	INT



	A	B	C	D
1	CodeVoie	NomVoie	NumVoie	Code type de voie
2	5	CENTRALE	190	0
3	20	DU CHATEAU	347	0
4	527	DU MONT BLANC	58	1
5	2	DE L'ABBE JOLIVET	140	0
6	110	BUFFON	39	0
7	179	JEAN FALCONNIER	28	1
8	120	DE GENEVE	8	0
9	210	DU RECULET	2	0
10	465	JEAN DE GINGINS	1403	0
11	36	DES CAPUCINES	226	2
12	210	DE POUIGNY	276	3
13	1490	DE VERDUN	79	4
14	20	DU COMMERCE	77	0
15	650	DE PRE BAILLY	240	0
16	541	TURENNE	3	0
17	695	DU SQUARE DE LAUSANNE	44	2
18	40	ALEXANDRE BERARD	59	0
19	130	DES LONGES RAYES	282	5
20	45	GRANDE RUE	54	6
21	160	DE GEX	1	0
22	A036	LES JARDINS DE CHEVRY	17	7
23	25	DES CARPIERES	329	0
24	55	CHARLES DE GAULLE	369	1
25	33	SUR LA VIGNETTE	154	0
26	745	DES USINIERS	6	0
27	135	DU JURA	2	0
28	455	MARCEL DEMIA	54	0
29	191	DE CHARIGNIN	269	5
30	37	DE BRETIGNY	322	0
31	124	DES HETRES	132	2
32	170	DE MEYRIN	80	0

- Création d'un fichier Excel pour chaque Table
- Copie des données les fichiers Excel crée
- Ordonnées les données en respectant le MCD

# 5 - Création de la BDD et insertion de données

- Création de la base de données sur MySQL Workbench

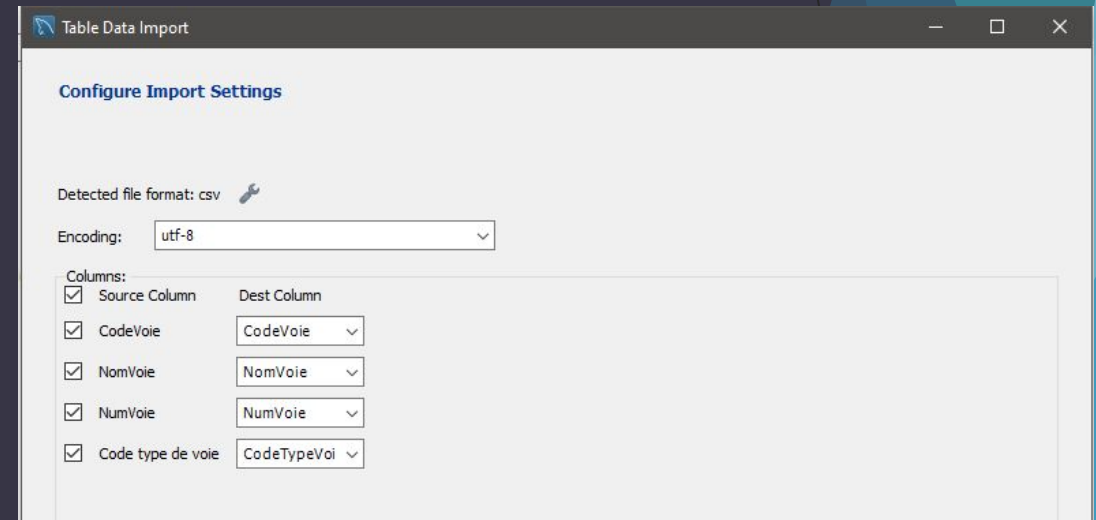
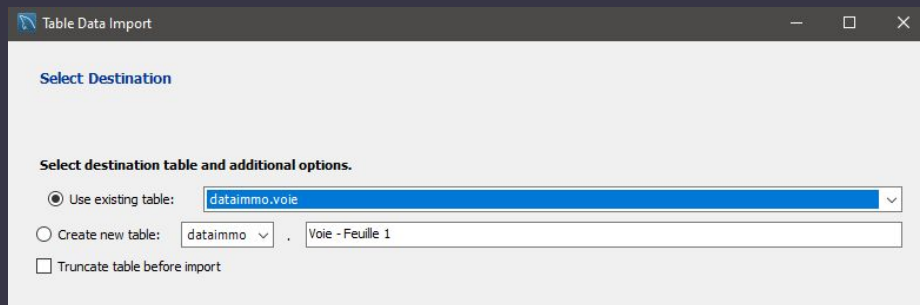
```
1 ● CREATE DATABASE dataimmo;
```

- Création des tables grâce au script SQL

```
1 CREATE TABLE Commune(  
2     CodeCommune INT,  
3     CodePostal INT,  
4     NomCommune VARCHAR(100),  
5     CodeDepartement VARCHAR(50),  
6     PRIMARY KEY(CodeCommune)  
7 );  
8  
9 CREATE TABLE TypeVoie(  
10     CodeTypeVoie INT,  
11     NomTypeVoie VARCHAR(50),  
12     PRIMARY KEY(CodeTypeVoie)  
13 );  
14  
15 CREATE TABLE TypeLocal(  
16     CodeTypeLocal INT,  
17     NomTypeLocal VARCHAR(20),  
18     PRIMARY KEY(CodeTypeLocal)  
19 );  
20  
21 CREATE TABLE Voie(  
22     CodeVoie VARCHAR(50),  
23     NomVoie VARCHAR(100),  
24     NumVoie INT,  
25     CodeTypeVoie INT NOT NULL,  
26     PRIMARY KEY(CodeVoie, NomVoie, NumVoie, CodeTypeVoie),  
27     FOREIGN KEY(CodeTypeVoie) REFERENCES TypeVoie(CodeTypeVoie)  
28 );
```

# 5 - Création de la BDD et insertion de données

- Enregistrer les fichiers Excel au format CSV
- Importation des fichiers CSV correspondant aux bonnes tables dans le bon ordre



# 6 - Analyse de données via requêtes SQL

Requête n°1 : Nombre total d'appartements vendus au 1er semestre 2020

## Requête

```
SELECT count(*) AS NbAppartVendu  
FROM mutation m, bien b, typelocal tl  
WHERE b.CodeBien = m.CodeBien  
AND tl.CodeTypeLocal = b.CodeTypeLocal  
AND nature = 'vente'  
AND NomTypeLocal = 'Appartement'  
AND DateMutation BETWEEN "2020-01-01" AND  
"2020-06-30";
```

## Résultat

	NbAppartVendu
▶	31375

# 6 - Analyse de données via requêtes SQL

Requête n°2 : Proportion des ventes d'appartements par le nombre de pièces

## Requête

```
WITH  
table1 AS (  
SELECT count(*) AS NbAppartVendu  
FROM mutation m, bien b, typelocal tl  
WHERE b.CodeBien = m.CodeBien  
AND tl.CodeTypeLocal = b.CodeTypeLocal  
AND nature = 'vente'  
AND NomTypeLocal = 'Appartement')  
  
SELECT NbPieces,  
ROUND((count(*)/table1.NbAppartVendu)*100, 2) AS  
ProportionVentes  
FROM table1, bien  
GROUP BY NbPieces  
ORDER BY NbPieces ASC;
```

## Résultat

	NbPieces	ProportionVentes
▶	0	0.11
	1	20.97
	2	31.40
	3	29.94
	4	17.27
	5	5.27
	6	1.17
	7	0.31
	8	0.09
	9	0.04
	10	0.01
	11	0.00

# 6 - Analyse de données via requêtes SQL

Requête n°3 : Liste des 10 départements où le prix du mètre carré est le plus élevé

## Requête

```
SELECT CodeDepartement,  
ROUND(AVG(ValeurFonciere/ SurfaceBati), 2) AS  
PrixM²  
FROM mutation m, bien b, commune c  
WHERE b.CodeBien = m.CodeBien  
AND c.CodeCommune = b.CodeCommune  
GROUP BY CodeDepartement  
ORDER BY PrixM² DESC  
LIMIT 10;
```

## Résultat

	CodeDepartement	PrixM²
►	75	12121.88
	92	7415.28
	94	5395.98
	6	4681.76
	93	4363.01
	74	4149.56
	78	4126.19
	69	4064
	2A	3910.52
	33	3838.74

# 6 - Analyse de données via requêtes SQL

Requête n°4 : Prix moyen du mètre carré d'une maison en Île-de-France

## Requête

```
SELECT ROUND(AVG(ValeurFonciere/SurfaceBati), 2)  
AS PrixMoyenM²  
FROM mutation m, bien b, commune c, typelocal tl  
WHERE b.CodeBien = m.CodeBien  
AND c.CodeCommune = b.CodeCommune  
AND tl.CodeTypeLocal = b.CodeTypeLocal  
AND NomTypeLocal = 'Maison'  
AND CodeDepartement IN( 75 , 77 , 78 , 91 , 92 , 93  
, 94 , 95);
```

## Résultat

	PrixMoyenM²
▶	3997.71



# 6 - Analyse de données via requêtes SQL

Requête n°5 : Liste des 10 appartements les plus chers avec le département et le nombre de mètres carrés

## Requête

```
SELECT CodeDepartement, SurfaceBati,  
b.CodeBien, ValeurFonciere  
FROM bien b , commune c, mutation m, typelocal tl  
WHERE c.CodeCommune = b.CodeCommune  
AND b.CodeBien = m.CodeBien  
AND tl.CodeTypeLocal = b.CodeTypeLocal  
AND tl.NomTypeLocal = 'appartement'  
ORDER BY m.ValeurFonciere DESC  
LIMIT 10;
```

## Résultat

	CodeDepartement	SurfaceBati	CodeBien	ValeurFonciere
►	75	10	31572	9000000
	91	62	21325	8600000
	75	289	29109	8577710
	75	42	31729	7620000
	75	200	29160	7600000
	75	143	28837	7535000
	75	357	31270	7420000
	75	241	31432	7200000
	75	310	28668	7050000
	75	76	28828	6600000

# 6 - Analyse de données via requêtes SQL

Requête n°6 : Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020

## Requête

```
WITH  
table1 as (  
SELECT count(CodeMutation) AS VentesPremierTrimestre  
FROM mutation m  
WHERE DateMutation BETWEEN "2020-01-01" AND "2020-03-31"),  
  
table2 as (  
SELECT count(CodeMutation) AS VentesDeuxiemeTrimestre  
FROM mutation m  
WHERE DateMutation BETWEEN "2020-04-01" AND "2020-06-30")  
  
SELECT VentesPremierTrimestre, VentesDeuxiemeTrimestre,  
ROUND((VentesDeuxiemeTrimestre-VentesPremierTrimestre)/Vent  
esPremierTrimestre*100, 2) AS TauxEvolution  
FROM table1, table2;
```

## Résultat

	VentesPremierTrimestre	VentesDeuxiemeTrimestre	TauxEvolution
►	16775	17391	3.67

# 6 - Analyse de données via requêtes SQL

Requête n°7 : Liste des communes où le nombre de ventes a augmenté d'au moins 20% entre le premier et le second trimestre de 2020

## Requête

```
WITH
table1 as (
SELECT NomCommune, count(*) AS VentesPremierTrimestre
FROM mutation m, bien b, commune c
WHERE b.CodeBien = m.CodeBien
AND c.CodeCommune = b.CodeCommune
AND DateMutation BETWEEN "2020-01-01" AND "2020-03-31"
GROUP BY NomCommune),

table2 as (
SELECT NomCommune, count(*) AS VentesDeuxiemeTrimestre
FROM mutation m, bien b, commune c
WHERE b.CodeBien = m.CodeBien
AND c.CodeCommune = b.CodeCommune
AND DateMutation BETWEEN "2020-04-01" AND "2020-06-30"
GROUP BY NomCommune)

SELECT table1.NomCommune, VentesPremierTrimestre, VentesDeuxiemeTrimestre,
ROUND(((VentesDeuxiemeTrimestre-VentesPremierTrimestre)/VentesPremierTrimestre*100, 2) AS TauxEvolution
FROM table1, table2
WHERE
((VentesDeuxiemeTrimestre-VentesPremierTrimestre)/VentesPremierTrimestre*100) > 20
AND table1.NomCommune = table2.NomCommune
ORDER BY TauxEvolution;
```

## Résultat

	NomCommune	VentesPremierTrimestre	VentesDeuxiemeTrimestre	TauxEvolution
►	LEUCATE	29	35	20.69
	ETAMPES	24	29	20.83
	LE CHESNAY-ROCQUENCOURT	14	17	21.43
	ORLY	9	11	22.22
	BOISSY-SAINT-LEGER	9	11	22.22
	PARIS 08	62	77	24.19
	RENNES	61	76	24.59
	VALENCE	28	35	25.00
	CADAUJAC	4	5	25.00
	SOUSTONS	8	10	25.00
	BATZ-SUR-MER	4	5	25.00
	VAL DE BRIEY	4	5	25.00
	NOGENT SUR OISE	8	10	25.00

# 6 - Analyse de données via requêtes SQL

Requête n° 8 : Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces

## Requête

```
WITH
table1 as (
SELECT AVG(ValeurFonciere/ SurfaceBati) AS PrixM²Appart2Pieces
FROM mutation m, bien b, typelocal tl
WHERE b.CodeBien = m.CodeBien
AND tl.CodeTypeLocal = b.CodeTypeLocal
AND NomTypeLocal = 'Appartement'
AND NbPieces = 2),

table2 as (
SELECT AVG(ValeurFonciere/ SurfaceBati) AS PrixM²Appart3Pieces
FROM mutation m, bien b, typelocal tl
WHERE b.CodeBien = m.CodeBien
AND tl.CodeTypeLocal = b.CodeTypeLocal
AND NomTypeLocal = 'Appartement'
AND NbPieces = 3)

SELECT
ROUND((PrixM²Appart3Pieces-PrixM²Appart2Pieces)/PrixM²Appart2Pieces*100,
2) AS DifférencePrixM²
FROM table1, table2
```

## Résultat

	DifférencePrixM²
▶	-13.03

# 6 - Analyse de données via requêtes SQL

Requête n°9 : Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69

## Requête

```
SELECT CodeDepartement, NomCommune,  
AVG(ValeurFonciere) AS Moyen_ValeurFonciere  
FROM mutation m, bien b, commune c  
WHERE b.CodeBien = m.CodeBien  
AND c.CodeCommune = b.CodeCommune  
AND CodeDepartement IN (6, 13, 33, 59, 69)  
GROUP BY NomCommune  
ORDER BY Moyen_ValeurFonciere DESC  
LIMIT 3;
```

## Résultat

	CodeDepartement	NomCommune	Moyen_ValeurFonciere
►	6	SAINT-JEAN-CAP-FERRAT	968750
	6	EZE	655000
	33	LEGE-CAP-FERRET	549500.6363636364