

OPENCLASSROOMS




Parcours Data Analyst



PROJET 3

**CRÉEZ ET UTILISEZ UNE BASE DE
DONNÉES IMMOBILIÈRE AVEC SQL**

Mor Niang



SUMMARY

-
- 1 Introduction
 - 2 Modélisation de la BDD
 - 3 Implémentation de la BDD
 - 4 Requêtes et résultats
 - 5 Conclusion



Laplace Immo : un réseau national d'agence immobilière.

Poste occupé : Data analyst

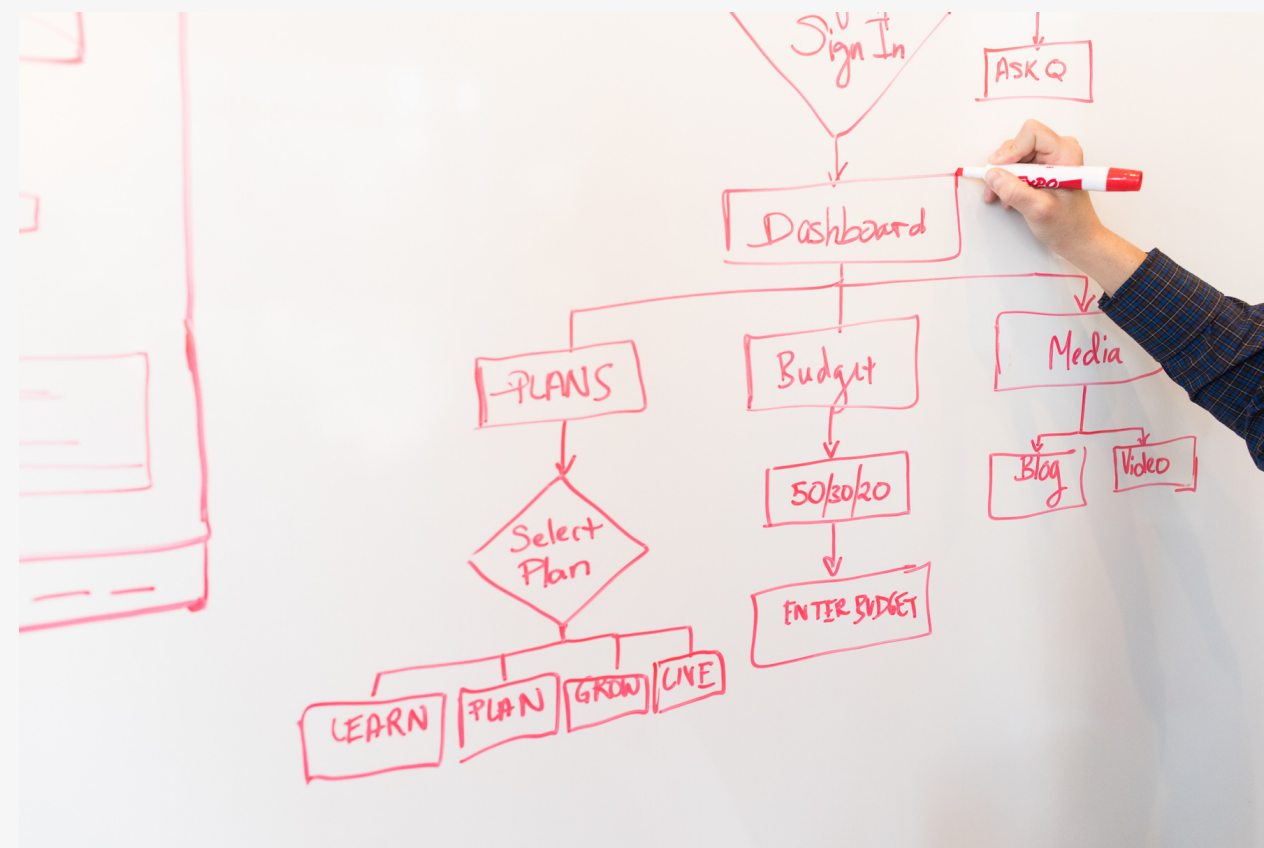
Projet DATAImmo : consiste a créer une base de données pour collecter les transactions immobilière et foncière en France.

Objectif du projet : analyser le marché de l'immobilier afin de mieux prévoir le prix de vente de l'immobilier.

Source de données : data.gouv.fr

Données traitées : premier semestre 2020

MODÉLISATION DE LA BDD



1. Le dictionnaire des données

2. Modèle conceptuel des données

3. Schéma relationnel Normalisé 3FN

1. Le dictionnaire de données

| Numéro | Attributs | Signification | Type | Observation |
|--------|-----------|-------------------------------|-----------------------------|--------------------------|
| 1 | IdLoc | Identifiant local | Numérique ou alphanumérique | Identifiant (artificiel) |
| 2 | TypLoc | Type local | Texte | Longueur : 20 |
| 3 | SurfLoc | Surface réelle battie | Numérique | Decimal |
| 4 | Surfter | Surface du terrain | Numérique | Decimal |
| 5 | Surfcar | Surface carrez | Numérique | Decimal |
| 6 | NbPieces | Nombre pièces local | Numérique | Entier |
| 7 | Nomrue | Nom de la rue | Texte | Longueur : 50 |
| 8 | IdCom | identifiant commune | Numérique ou alphanumérique | Identifiant |
| 9 | NomCom | Nom de la commune | Texte | Longueur : 45 |
| 10 | CodPost | Code Postal | Numérique | Longueur : 10 |
| 11 | Coddep | Code departement | Numérique ou alphanumérique | Longueur : 5 |
| 12 | IdVal | Identifiant de la Transaction | Numérique ou alphanumérique | Identifiant (artificiel) |
| 13 | DatMut | Date de la transaction | Date | |
| 14 | Valfonc | Valeur fonciere | Monetaire | Decimal |

2. Modèle conceptuel de base de données

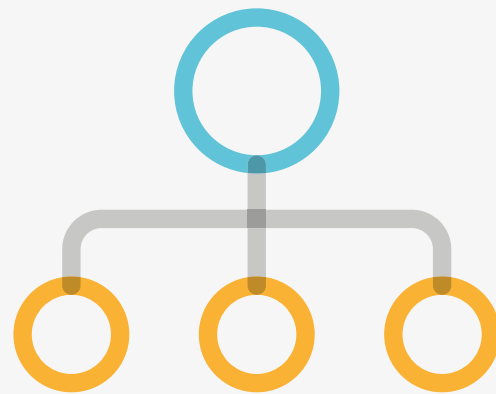
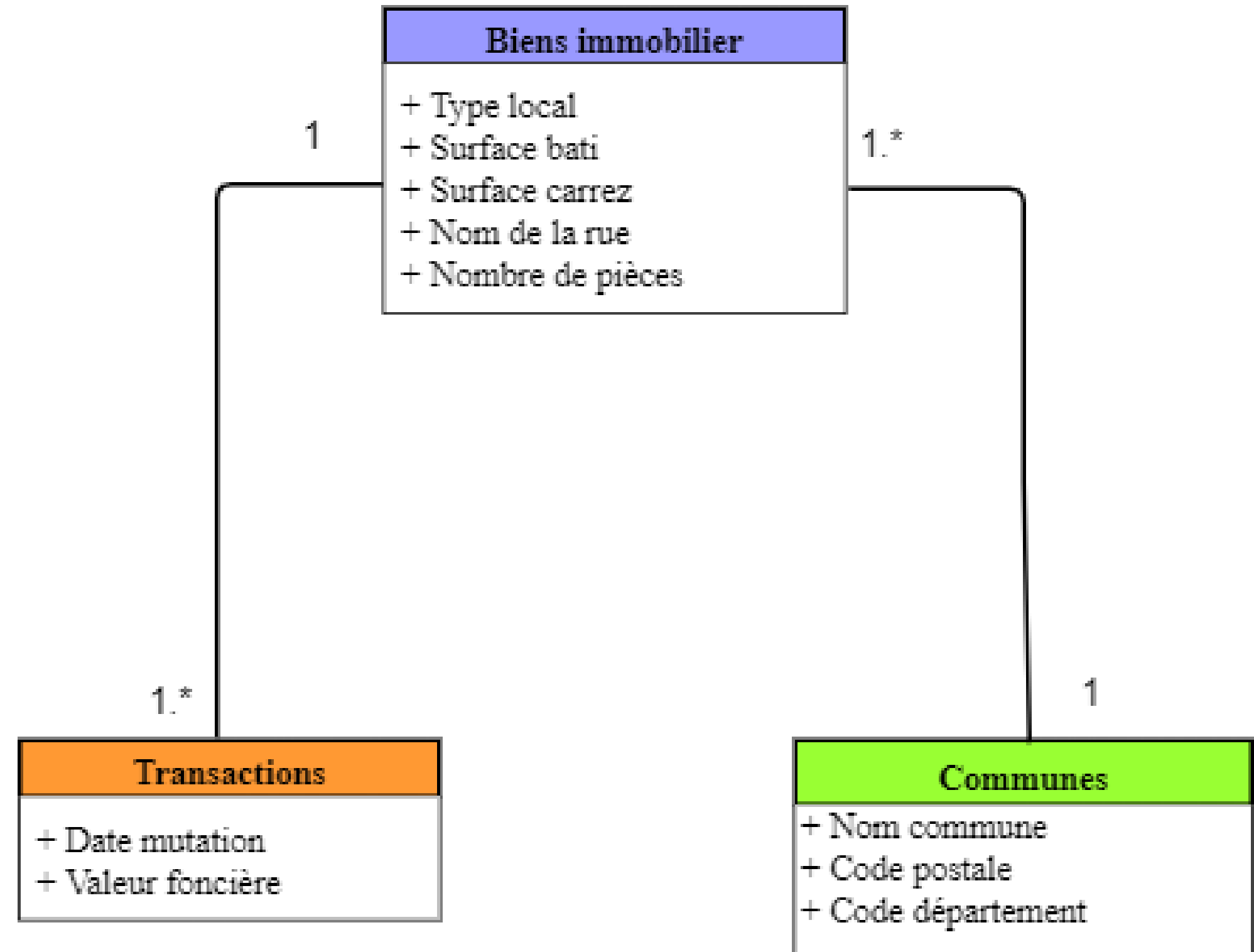
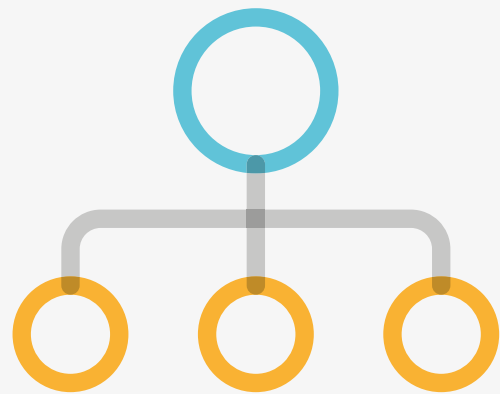


Diagramme de classe UML

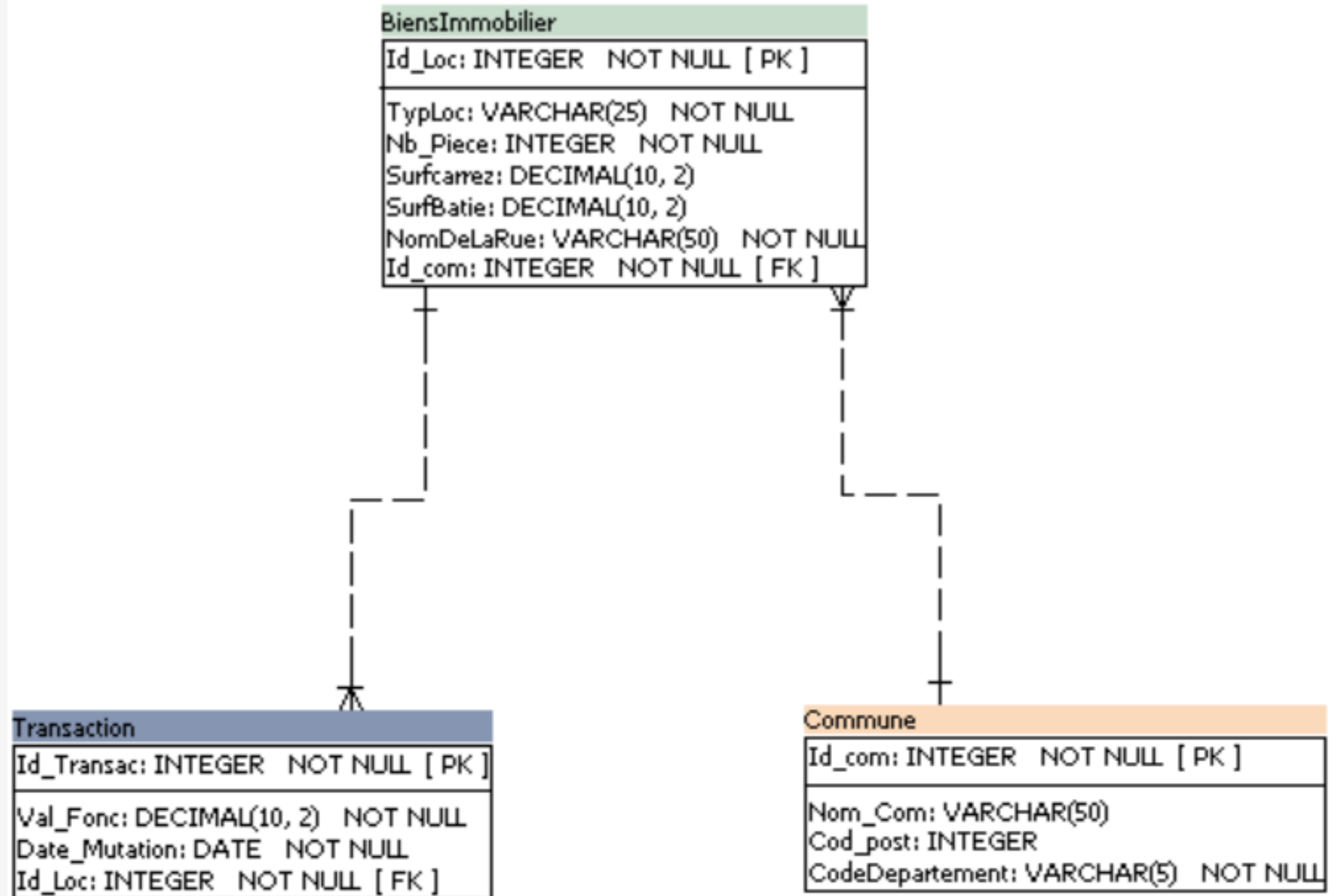
Outil utilisé: GitMind



3. Schéma relationnel Normalisé 3FN



SQL Power Architect





IMPLEMENTATION DE LA BDD

1. Création de la base de données
2. Nettoyage des données
3. Transfert des données

1. Création de la BDD



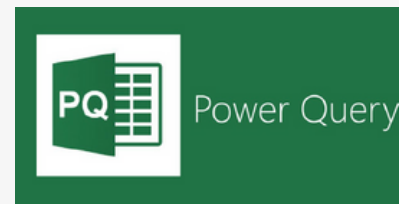
MySQL est un SGBD relationnelle.
Facile à configurer et gratuit.

Aperçu du Code de création de la BDD généré sur SQL Power Architect

```
2 • CREATE TABLE Commune (  
3     Id_com INT AUTO_INCREMENT NOT NULL,  
4     Nom_Com VARCHAR(50),  
5     Cod_post INT,  
6     CodeDepartement VARCHAR(5) NOT NULL,  
7     PRIMARY KEY (Id_com)  
8 );  
9  
10  
11 • CREATE TABLE BiensImmobilier (  
12     Id_Loc INT NOT NULL,  
13     TypLoc VARCHAR(25) NOT NULL,  
14     Nb_Piece INT NOT NULL,  
15     Surfcarrez DECIMAL(10,2),  
16     SurfBatie DECIMAL(10,2),  
17     NomDeLaRue VARCHAR(50) NOT NULL,  
18     Id_com INT NOT NULL,  
19     PRIMARY KEY (Id_Loc)  
20 );  
21
```

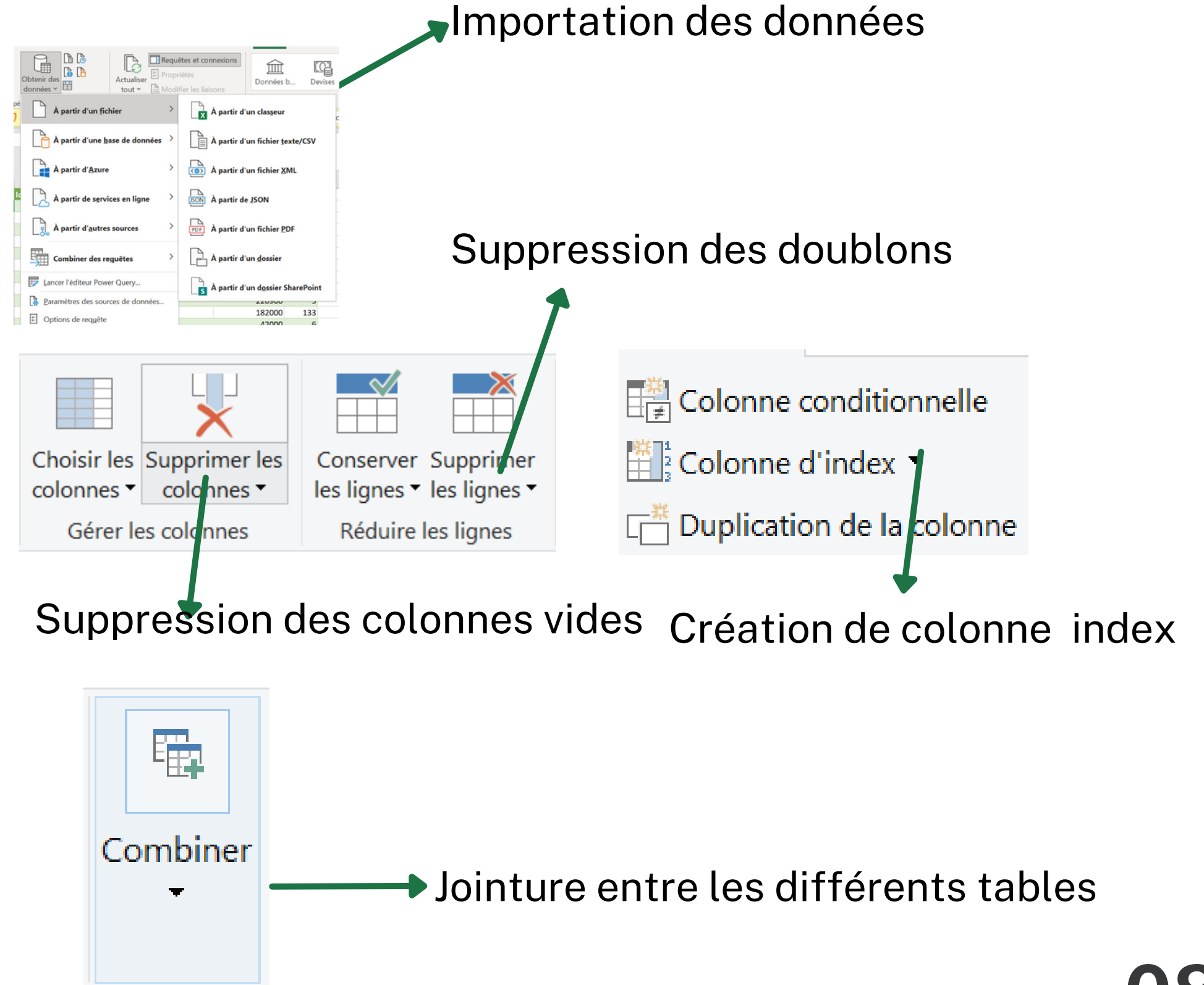
2. Nettoyage des données

Outil utilisé



Les différentes étapes :

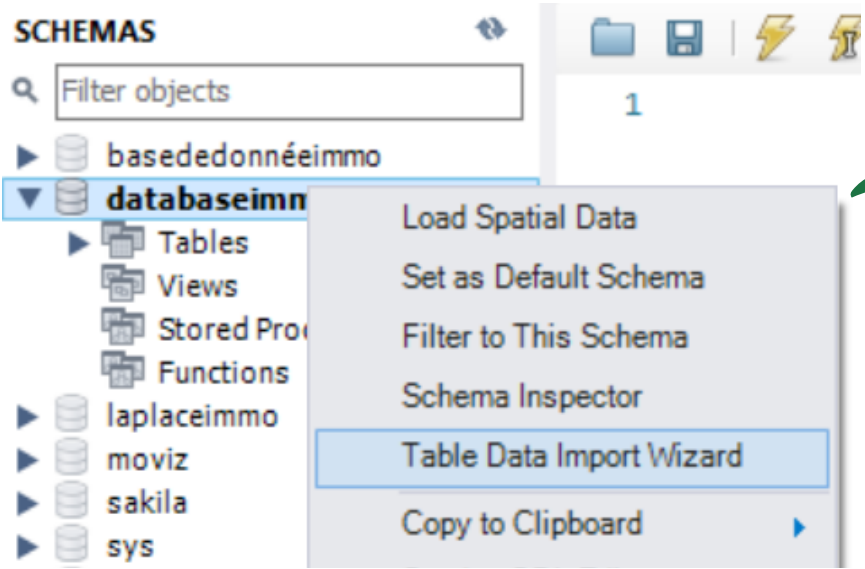
- Importation des données
- Suppression des doublons
- Supprimer les lignes vides
- Jointure des différentes tables
- Création des colonnes index



3. Transfert des données

Les tables importées:

- Commune 3215 lignes :
- BiensImmobilier 34166 lignes :
- Transaction 34169 lignes :



Importation des données à partir de Table Data Import Wizard.

| | Id_Loc | TypLoc | Nb_Piece | Surfcarrez | SurfBatie | SurfTerrain | Adresse | Id_com |
|---|--------|-------------|----------|------------|-----------|-------------|-------------------------------------|--------|
| ▶ | 1 | Appartement | 2 | 50.42 | 52.00 | 0.00 | 190 RUE CENTRALE | 1 |
| | 2 | Appartement | 3 | 48.22 | 48.00 | 0.00 | 347 RUE DU CHATEAU | 2 |
| | 3 | Appartement | 6 | 130.80 | 130.00 | 0.00 | 58 AV DU MONT BLANC | 3 |
| | 4 | Maison | 5 | 109.22 | 109.00 | 0.00 | 140 RUE DE L'ABBE JOLIVET | 4 |
| | 5 | Appartement | 4 | 108.65 | 91.00 | 0.00 | 39 RUE BUFFON | 5 |
| | 6 | Appartement | 2 | 31.65 | 32.00 | 0.00 | 28 AV JEAN FALCONNIER | 6 |
| | 7 | Appartement | 2 | 52.58 | 52.00 | 0.00 | 8 RUE DE GENEVE | 7 |
| | 8 | Appartement | 2 | 58.71 | 60.00 | 0.00 | 2 RUE DU RECULET | 7 |
| | 9 | Maison | 4 | 93.23 | 96.00 | 0.00 | 1403 RUE JEAN DE GINGINS | 3 |
| | 10 | Maison | 5 | 117.00 | 117.00 | 0.00 | 226 ALL DE 1403 RUE JEAN DE GINGINS | |
| | 11 | Appartement | 2 | 35.60 | 36.00 | 0.00 | 276 RTE DE POUIGNY | 4 |
| | 12 | Appartement | 5 | 138.03 | 137.00 | 0.00 | 79 CRS DE VERDUN | 8 |
| | 13 | Appartement | 2 | 42.00 | 43.00 | 0.00 | 77 RUE DU COMMERCE | 9 |
| | 14 | Appartement | 2 | 45.36 | 45.00 | 0.00 | 240 RUE DE PRE BAILLY | 10 |



Requêtes et résultats

1. Nombre total d'appartements vendus au 1er semestre 2020

```
1 • SELECT count(Id_Transac) as "Nombre total d'appartement vendu au premier semestre 2020"  
2 FROM transaction  
3 inner JOIN biensimmobilier using(Id_Loc)  
4 WHERE TypLoc = 'Appartement'  
5 AND Date_Mutation between '2020-01-01' and '2020-06-30';  
6
```

| | |
|---|---|
| | Nombre total d'appartement vendu au premier semestre 2020 |
| ▶ | 31375 |

2. Proportion des ventes d'appartements par le nombre de pièces

```
1 • SELECT Nb_Piece,  
2   (round((count(Id_Transac)/  
3     (SELECT count(Id_Transac) as "Nombre total d'appartement vendu au premier semestre 2020"  
4     FROM transaction  
5     inner join biensimmobilier using(Id_Loc)  
6     WHERE TypLoc = 'Appartement')  
7   )*100, 2)) as "Preportion des ventes d'appartements par le nombre de piec  
8 FROM transaction  
9 inner join biensimmobilier using(Id_Loc)  
10 WHERE TypLoc = 'Appartement'  
11 GROUP BY Nb_Piece  
12 order by Nb_Piece desc;
```

| Nb_Piece | Preportion des ventes d'appartements par le nombre de pieces |
|----------|--|
| 0 | 0.10 |
| 1 | 21.48 |
| 2 | 31.18 |
| 3 | 28.57 |
| 4 | 14.21 |
| 5 | 3.55 |
| 6 | 0.65 |
| 7 | 0.17 |
| 8 | 0.05 |
| 9 | 0.03 |
| 10 | 0.01 |
| 11 | 0.00 |

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

```
1 • select CodeDepartement as 'Département',  
2 round(AVG(Val_Fonc/Surfcarrez),2) as 'Prix au metre carré'  
3 from commune  
4 inner join biensimmobilier using(Id_com)  
5 inner join transaction using(Id_loc)  
6 group by 1  
7 order by 2 desc  
8 limit 10
```

| | Département | Prix au metre carré |
|---|-------------|---------------------|
| ▶ | 75 | 12214.11 |
| | 92 | 7259.11 |
| | 94 | 5495.98 |
| | 74 | 4894.57 |
| | 6 | 4740.37 |
| | 93 | 4467.18 |
| | 78 | 4393.84 |
| | 73 | 4247.75 |
| | 69 | 4208.26 |
| | 2A | 4014.23 |

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

```
1 • select round(avg(Val_Fonc/Surfcarrez),2) as "Prix moyen du mètre carré d'une maison en Île-de-France"
2 from transaction
3 inner join biensimmobilier using (Id_Loc)
4 inner join commune using (Id_com)
5 where TypLoc = 'maison'
6 and CodeDepartement in (75, 77, 78, 91, 92, 93, 94, 95)
7
```

| | |
|---|---|
| | Prix moyen du mètre carré d'une maison en Île-de-France |
| ▶ | 3746.77 |

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

```
1 • select biensimmobilier.Id_Loc, TypLoc, Val_Fonc, CodeDepartement, Surfcarrez
2 from commune
3 inner join biensimmobilier using (Id_com)
4 inner join transaction using (Id_Loc)
5 where TypLoc = 'Appartement'
6 order by TypLoc, Val_Fonc desc
7 limit 10
```

| | Id_Loc | TypLoc | Val_Fonc | CodeDepartement | Surfcarrez |
|---|--------|-------------|------------|-----------------|------------|
| ► | 32275 | Appartement | 9000000.00 | 75 | 9.10 |
| | 21835 | Appartement | 8600000.00 | 91 | 64.00 |
| | 29799 | Appartement | 8577713.00 | 75 | 20.55 |
| | 7774 | Appartement | 7620000.00 | 34 | 42.77 |
| | 29850 | Appartement | 7600000.00 | 75 | 253.30 |
| | 29522 | Appartement | 7535000.00 | 75 | 139.90 |
| | 31973 | Appartement | 7420000.00 | 75 | 360.95 |
| | 32135 | Appartement | 7200000.00 | 75 | 595.00 |
| | 29353 | Appartement | 7050000.00 | 75 | 122.56 |
| | 29513 | Appartement | 6600000.00 | 75 | 79.38 |

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

```
1 • with
2
3 ⊖ table1 as (select count(id_transac) as Trimestre1
4   from transaction
5   where Date_Mutation between '2020-01-01' and '2020-03-31'),
6
7 ⊖ table2 as (select count(id_transac) as Trimestre2
8   from transaction
9   where Date_Mutation between '2020-04-01' and '2020-06-30')
10
11   select Trimestre1 as 'Nombre de vents au premier trimestre',
12   Trimestre2 as 'Nombre de vents au second trimestre',
13   round(((Trimestre2-Trimestre1)/Trimestre1)*100,2) as "taux d'evolution"
14   from table1, table2
```

| | Nombre de vents au premier trimestre | Nombre de vents au second trimestre | taux d'evolution |
|---|--------------------------------------|-------------------------------------|------------------|
| ► | 16774 | 17391 | 3.68 |

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

```
7 • with
8   Table1 as (select Nom_Com, count(id_transac) as nbventestrimestre1
9   from transaction
10  inner join biensimmobilier using (Id_Loc)
11  inner join commune using (Id_com)
12  where Date_Mutation between '2020-01-01' and '2020-03-31'
13  group by Nom_Com),
14
15  Table2 as (select Nom_Com, count(id_transac) as nbventestrimestre2
16  from transaction
17  inner join biensimmobilier using (Id_Loc)
18  inner join commune using (Id_com)
19  where Date_Mutation between '2020-03-31' and '2020-06-30'
20  group by Nom_Com)
21
22  select Nom_Com,
23  nbventestrimestre1 as 'Nombre de vents au premier trimestre',
24  nbventestrimestre2 as 'Nombre de vents au second trimestre',
25  round(((nbventestrimestre2-nbventestrimestre1)/nbventestrimestre1)*100,2) as "taux d'evolution"
26  from table1
27  inner join table2 using (Nom_Com)
28  where ((nbventestrimestre2-nbventestrimestre1)/nbventestrimestre1)*100 >= 20
```

| | Nom_Com | Nombre de vents au premier trimestre | Nombre de vents au second trimestre | taux d'evolution ▼ |
|---|-------------------------------|--------------------------------------|-------------------------------------|--------------------|
| ► | BILLERE | 1 | 18 | 1700.00 |
| | PAU | 5 | 72 | 1340.00 |
| | CAVAILLON | 1 | 12 | 1100.00 |
| | BELLEVILLE-EN-BEAUJOLAIS | 1 | 8 | 700.00 |
| | ROUEN | 13 | 97 | 646.15 |
| | LE CREUSOT | 1 | 7 | 600.00 |
| | PETIT QUEVILLY (LE) | 1 | 7 | 600.00 |
| | MORLAIX | 2 | 13 | 550.00 |
| | AUNEAU-BLEURY-SAINT-SYMPHO... | 1 | 6 | 500.00 |
| | VITRE | 2 | 12 | 500.00 |
| | SAINT-MARTIN-LE-VINOUX | 1 | 6 | 500.00 |
| | LE PALAIS | 1 | 6 | 500.00 |
| | LAMORLAYE | 1 | 6 | 500.00 |
| | RAMONVILLE ST AGNE | 1 | 6 | 500.00 |
| | LYON 7EME | 12 | 65 | 441.67 |
| | ORCIERES | 1 | 5 | 400.00 |

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

```
7 • with
8
9 ⊖ Table1 as (select round(avg(Val_Fonc/Surfcarrez), 2) as prix2pieces
10 from biensimmobilier
11 inner join transaction using (Id_Loc)
12 where Nb_Piece = 2 and TypLoc = 'appartement'),
13
14 ⊖ Table2 as (select round(avg(Val_Fonc/Surfcarrez),2) as prix3pieces
15 from biensimmobilier
16 inner join transaction using (Id_Loc)
17 where Nb_Piece = 3 and TypLoc = 'appartement')
18
19 select prix2pieces as "Prix au metre carre appartement 2 piece",
20 prix3pieces as "Prix au metre carre appartement 3 piece",
21 round(((prix3pieces-prix2pieces)/prix2pieces)*100,2) as 'Difference en poucentage'
22 from Table1, Table2
```

| | Prix au metre carre appartement 2 piece | Prix au metre carre appartement 3 piece | Difference en poucentage |
|---|---|---|--------------------------|
| ► | 4903.56 | 4300.36 | -12.30 |

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

```
with
table1 as (select Nom_Com, CodeDepartement,
round(avg(Val_Fonc), 2) as vm ,
rank()
over (partition by CodeDepartement order by avg(Val_Fonc) desc) as rang
from commune
inner join biensimmobilier using (Id_Com)
inner join transaction using (Id_Loc)
where CodeDepartement in ('6', '13', '33', '59', '69')
group by 1, 2)

select Nom_Com, CodeDepartement, vm, rang
from table1
where rang < 4
```

| | Nom_Com | CodeDepartement | vm | rang |
|---|---------------------|-----------------|-----------|------|
| ► | GIGNAC-LA-NERTHE | 13 | 330000.00 | 1 |
| | SAINT SAVOURNIN | 13 | 314425.00 | 2 |
| | CASSIS | 13 | 310292.63 | 3 |
| | LEGE-CAP-FERRET | 33 | 549500.64 | 1 |
| | VAYRES | 33 | 335000.00 | 2 |
| | LE TAILLAN MEDOC | 33 | 303700.00 | 3 |
| | BERSEE | 59 | 433202.00 | 1 |
| | CYSOING | 59 | 408550.00 | 2 |
| | MOUVAUX | 59 | 331727.14 | 3 |
| | EZE | 6 | 655000.00 | 1 |
| | SAINT-JEAN-CAP-F... | 6 | 582088.89 | 2 |
| | MOUANS-SARTOUX | 6 | 476898.10 | 3 |
| | VILLE SUR JARNIOUX | 69 | 485300.00 | 1 |
| | LYON 2EME | 69 | 458667.22 | 2 |
| | LYON 6EME | 69 | 439409.97 | 3 |

Conclusion



Compétences développées :

1. Conception d'un modèle relationnel
2. Les formes normales
3. Création de base de données
4. MySQL, SQL Power Architect
5. Requêtes avancées
6. Excel, Power Query