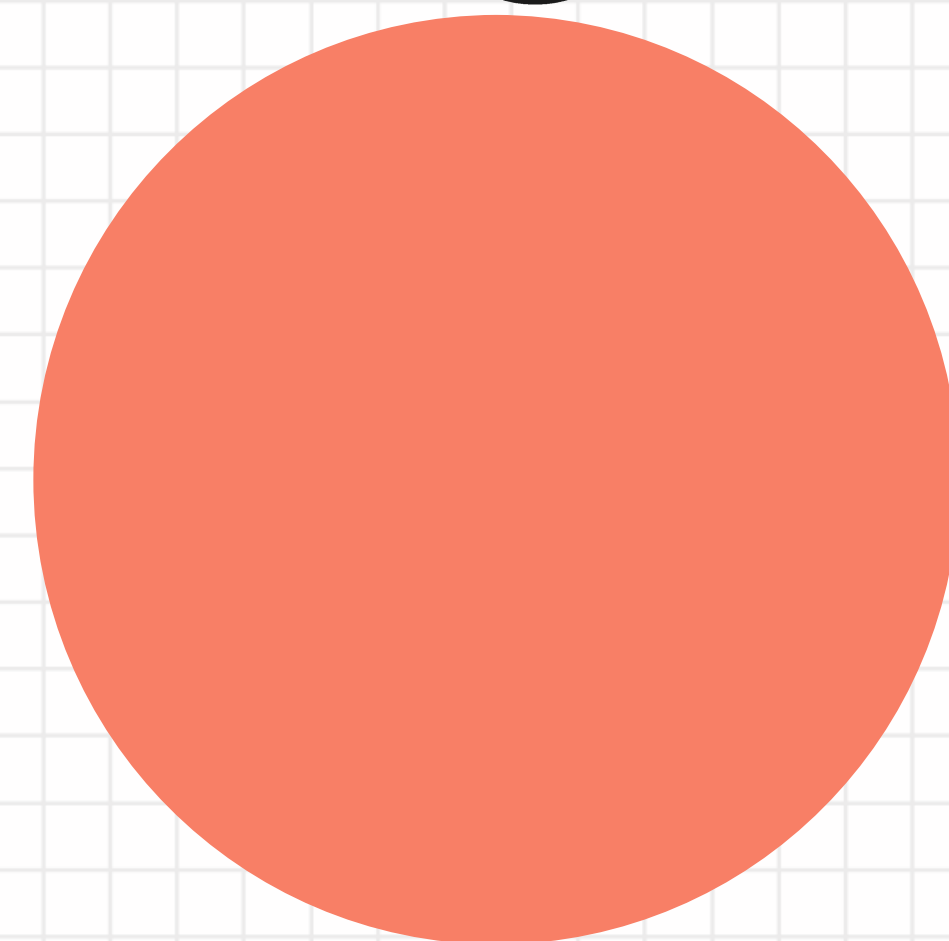
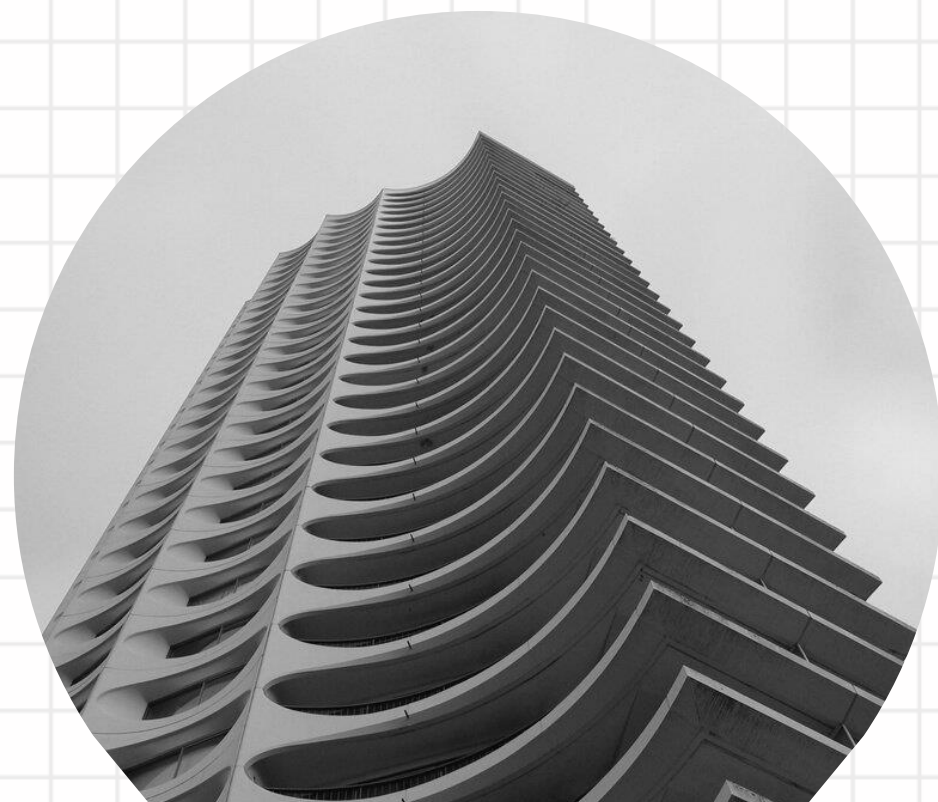




Laplace Immo

Projet 3 : DATAImmo

Azra HAMIDOVIC



Sommaire

- Démarche de création de la base de données (BDD)
- Présentation de la BDD et données importées
- Réponses aux besoins en analyse de données

Démarche de création de la base de données (BDD)

Description du jeu de données

34169

transactions

Données extraites du site open Data des Demandes de valeurs foncières (DVF) pour le premier semestre de 2020

45

colonnes au total

Informations sur les transactions (ex : date de mutation, valeur foncière, commune, type de local...)

20

colonnes sélectionnées

(certaines n'ont pas été sélectionnées pour la BDD car vides ou n'apportant pas d'informations intéressantes)

Création d'un dictionnaire des données à partir de ces colonnes + ajout d'une colonne "id"

Démarche de création de la base de données (BDD)

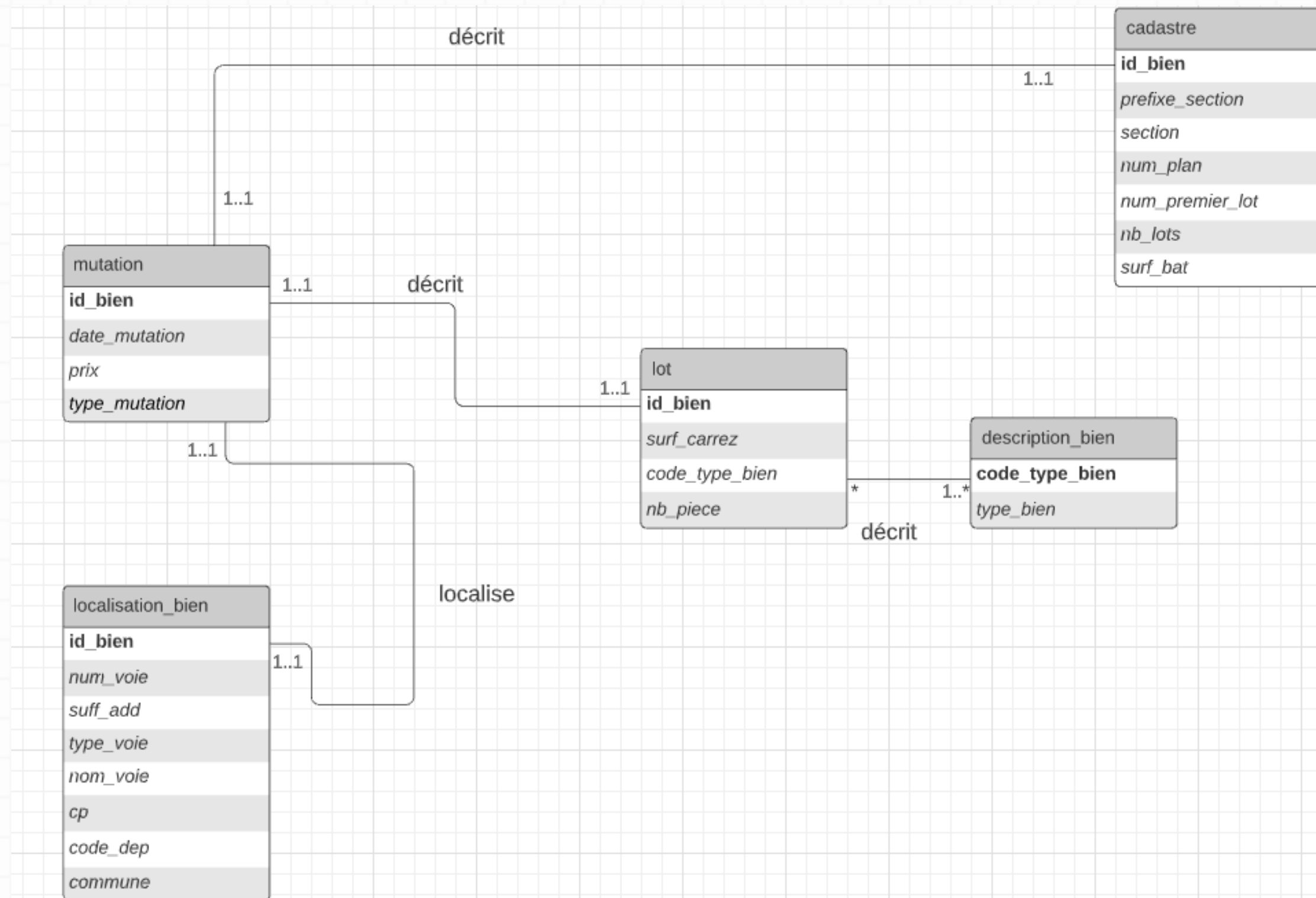
Dictionnaire des données

- Sélection des données importantes à stocker dans la BDD

Numéro ▼	Code propriété ▼	Signification ▼	Type ▼	Observation ▼
1	id	Identifiant	Numérique	INTEGER
2	date_mutation	Date de transaction	Date	AAAA-MM-JJ
3	prix	Valeur foncière	Numérique	Décimal(10,2)
4	nature_mutation	Type de mutation	Texte	Longueur : 20
5	num_voie	Numéro de voie	Numérique	Double
6	suff_add	Suffixe B/T/Q de l'adresse	Texte	Longueur : 3
7	type_voie	Type de voie	Texte	Longueur : 10
8	nom_voie	Nom de la voie	Texte	Longueur : 50
9	cp	Code postal	Numérique	Double
10	commune	Nom de la commune	Texte	Longueur : 50
11	code_dep	Code département	Numérique	Double
12	prefixe_section	Préfixe de section	Numérique	Double
13	section	Section	Texte	Varchar(10)
14	num_plan	Numéro de plan	Numérique	Double
15	num_premier_lot	1er lot	Alphanumérique	Varchar(10)
16	surf_carrez	Surface Carrez du 1er lot	Numérique	Décimal (10,2)
17	nb_lots	Nombre de lots	Numérique	Entier
18	code_type_local	Code type local	Numérique	INTEGER
19	type_bien	Type de bien	Texte	Longueur : 30
20	surf_bat	Surface réelle du bâtiment	Numérique	Décimal (10,2)
21	nb_piece	Nombre de pièces principales	Numérique	Double

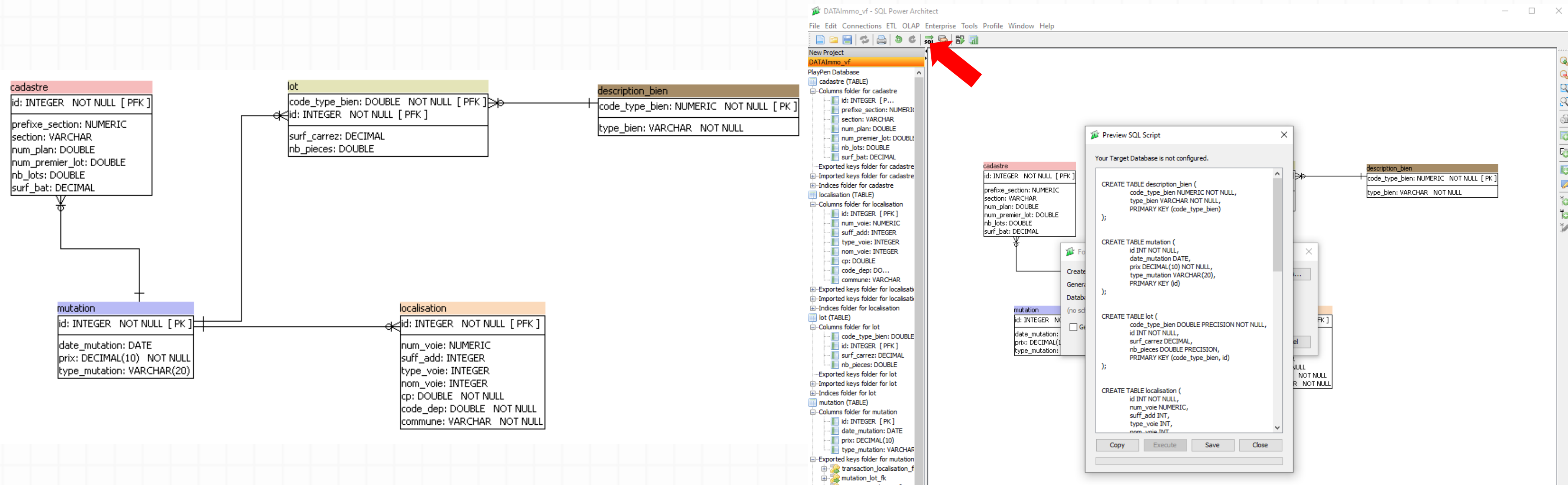
Démarche de création de la base de données (BDD)

Modèle conceptuel des données - Diagramme UML (Lucid)



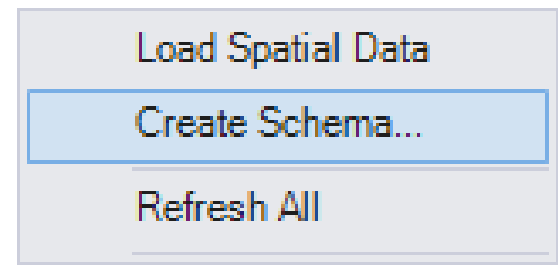
Démarche de création de la base de données (BDD)

Schéma relationnel normalisé en 3NF - Création des tables (SQL Power Architect)

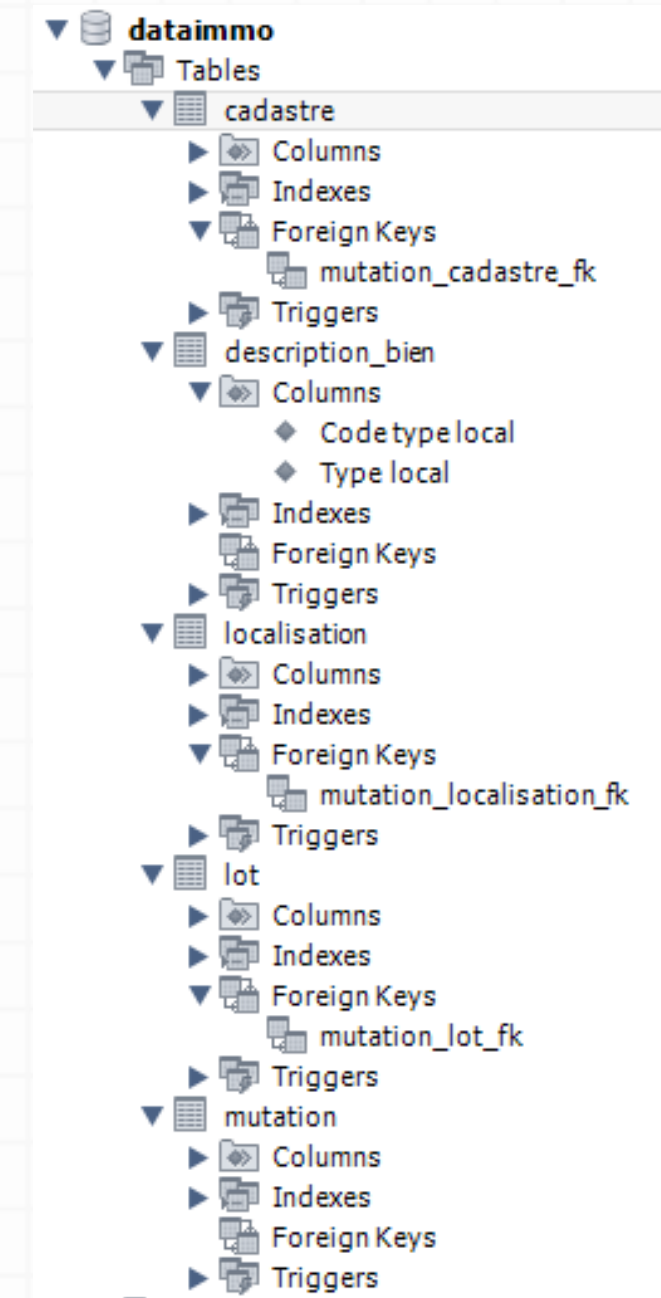


Démarche de création de la base de données (BDD)

Création de la BDD (MySQL Workbench)



```
1 • CREATE SCHEMA `dataimmo` ;
2 • USE dataimmo;
3 • CREATE TABLE description_bien (
4     code_type_local NUMERIC NOT NULL,
5     type_local TEXT,
6     PRIMARY KEY (code_type_local)
7 );
8
9
10 • CREATE TABLE mutation (
11     id INT NOT NULL,
12     date_mutation DATE,
13     prix DECIMAL(10),
14     type_mutation TEXT,
15     PRIMARY KEY (id)
16 );
17
18
19 • CREATE TABLE lot (
20     code_type_bien DOUBLE PRECISION NOT NULL,
21     id INT NOT NULL,
22     surf_carrez DECIMAL(10,2),
23     nb_pieces DOUBLE PRECISION,
24     PRIMARY KEY (code_type_local, id)
25 );
26
27
28 • CREATE TABLE localisation (
29     id INT NOT NULL,
```



Présentation de la BDD et données importées

Préparation des données

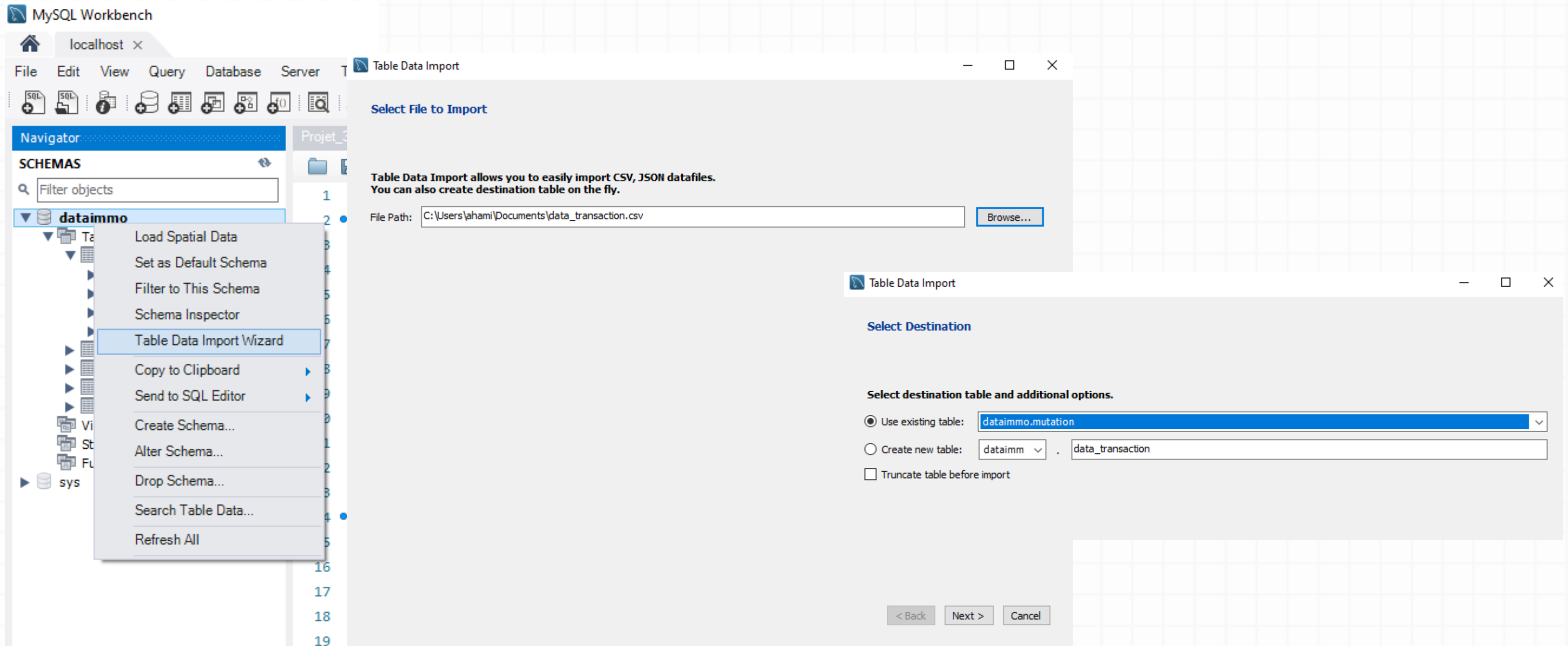
	A	B	C	D	E
1	id	Date mutation	Nature mutation	Valeur fonciere	N°
2	1431	2020-02-20	Vente	391000	
3	2403	2020-03-30	Vente	82000	
4	15186	2020-06-26	Vente	321600	
5	8325	2020-03-05	Vente	117900	
6	20189	2020-06-03	Vente	115700	
7	20200	2020-05-26	Vente	109000	
8	20075	2020-03-30	Vente	194901	
9	19969	2020-01-28	Vente	147300	
10	19993	2020-02-10	Vente	133500	
11	19988	2020-02-17	Vente	146000	
12	18660	2020-06-17	Vente	235000	
13	2035	2020-01-17	Vente	28000	
14	11813	2020-06-12	Vente	97800	
15	20711	2020-06-03	Vente	92380	
16	20701	2020-06-04	Vente	299000	
17	20821	2020-06-25	Vente	206000	
18	20167	2020-05-20	Vente	171150	
19	29325	2020-01-23	Vente	489770	
20	20569	2020-03-11	Vente	225000	
21	19835	2020-06-19	Vente	195000	
22	21636	2020-03-31	Vente	34500	
23	21720	2020-05-20	Vente	32000	
24	15197	2020-06-30	Vente	210000	
25	7697	2020-03-09	Vente	61960	
26	8934	2020-01-30	Vente	83509,5	
27	8901	2020-01-17	Vente	145000	
28	30673	2020-02-12	Vente	309000	
29	4459	2020-04-22	Vente	78000	
30	19615	2020-06-11	Vente	54500	
31	12741	2020-06-24	Vente	123500	
32	27370	2020-06-26	Vente	399000	
33	12212	2020-01-31	Vente	142500	
34	612	2020-02-14	Vente	291230	
35	21116	2020-02-26	Vente	71000	
36	21112	2020-01-24	Vente	62000	
37	13308	2020-01-20	Vente	220500	
38	2901	2020-01-13	Vente	116590	
39	11401	2020-06-23	Vente	104000	
40	8638	2020-05-28	Vente	50800	
41	8562	2020-03-13	Vente	153400	
42	8592	2020-05-06	Vente	247500	
43	12031	2020-05-28	Vente	57000	
44	12495	2020-06-19	Vente	88000	
45	12471	2020-06-25	Vente	228000	
46	14389	2020-05-26	Vente	223500	
47	21108	2020-02-08	Vente	104900	
48	740	2020-05-19	Vente	82000	
49	6182	2020-05-19	Vente	100000	

Sélection des
colonnes qui
constitueront chaque
table et creation d'un
fichier .csv par table

data_cadastre.csv
data_description_bien.csv
data_localisation.csv
data_lot.csv
data_transaction.csv

Présentation de la BDD et données importées

BDD sur MySQL



Présentation de la BDD et données importées

BDD sur MySQL

Table Data Import

Configure Import Settings

Detected file format: csv

Encoding: utf-8

Columns:

<input checked="" type="checkbox"/>	Source Column	Dest Column
<input checked="" type="checkbox"/>	id	id
<input checked="" type="checkbox"/>	Date mutation	date_mutation
<input checked="" type="checkbox"/>	Nature mutation	nature_mutation
<input checked="" type="checkbox"/>	Valeur foncière	prix

id	Date muta...	Nature mu...	Valeur fon...
1	2020-02-03	Vente	56000
2	2020-01-02	Vente	165000
3	2020-01-08	Vente	720000
4	2020-01-06	Vente	429250
5	2020-01-07	Vente	220900

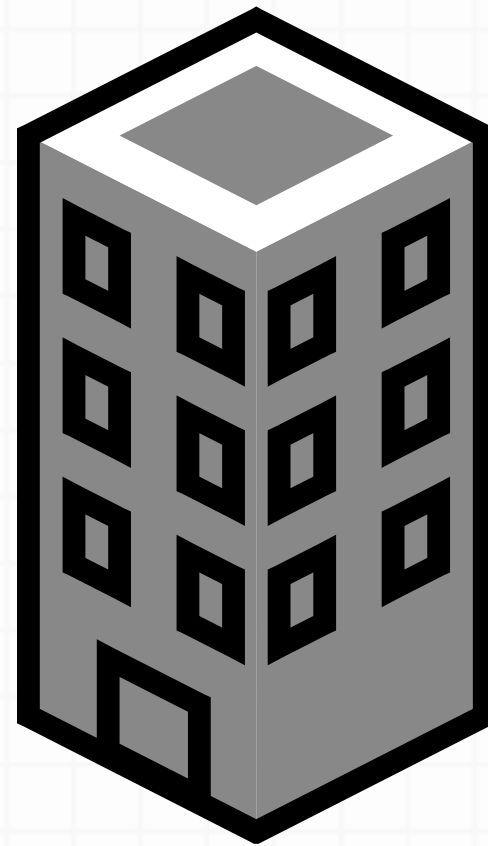
< Back Next > Cancel

Besoins en analyse des données (Proof of Concept)

Réponses aux besoins en analyse de données

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

```
SELECT  
COUNT(a.id) AS Nb_vente_Appt_Semestre1  
FROM  
    mutation AS a  
LEFT JOIN lot AS b ON  
a.id = b.id  
WHERE  
    (date_mutation < '2020-07-01')  
    AND b.code_type_local = '2';
```



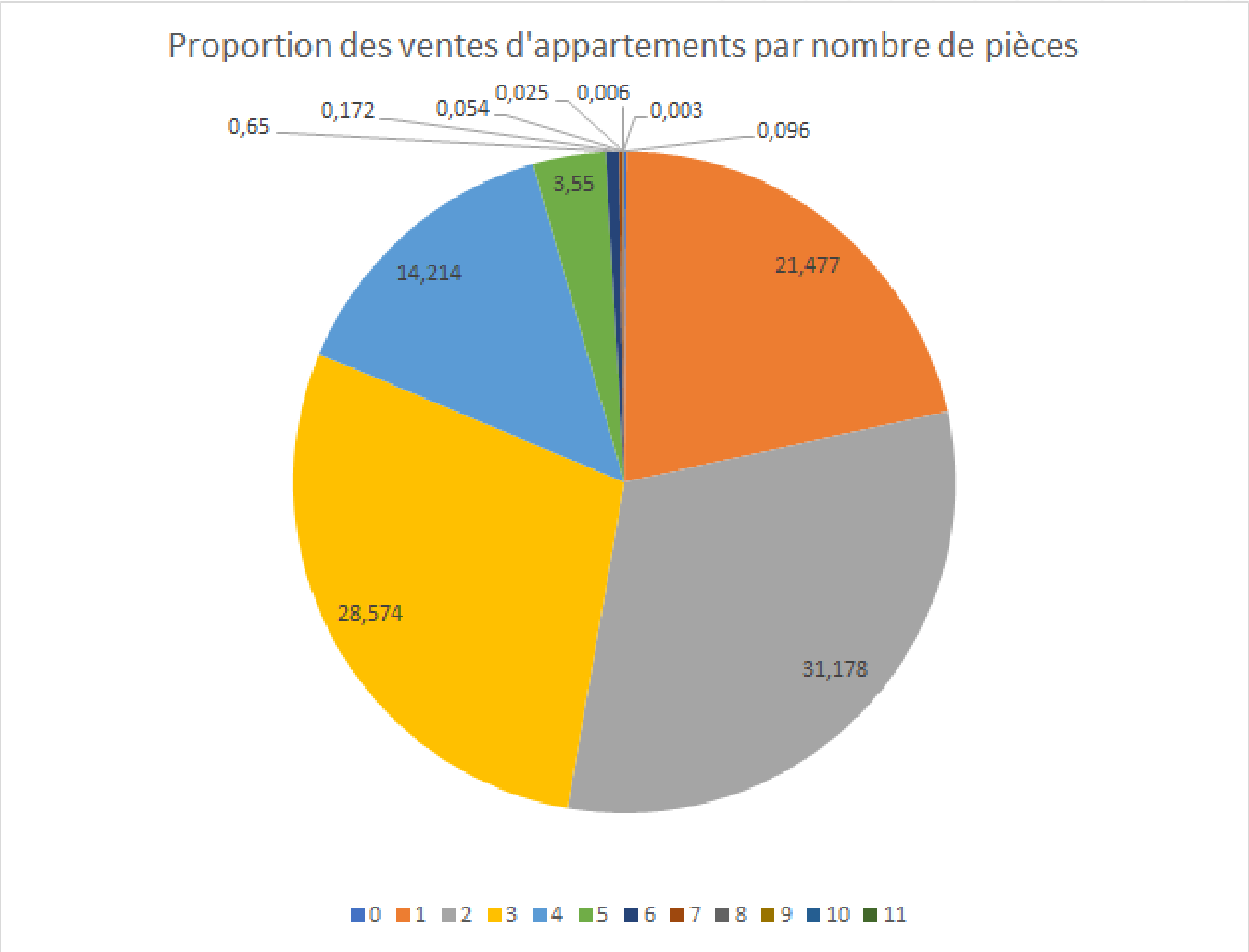
31378 appartements
vendus

Réponses aux besoins en analyse de données

Proportion des ventes d'appartements par nombre de pièces

```
SELECT nb_pieces, ROUND(COUNT(*) * 100.0 / SUM(COUNT(*)) over(),3) AS proportion_ventes
FROM lot
WHERE code_type_local = '2'
GROUP BY nb_pieces
ORDER BY nb_pieces;
```

Nombre de pièces	Proportion des ventes (%)
0	0,096
1	21,477
2	31,178
3	28,574
4	14,214
5	3,550
6	0,650
7	0,172
8	0,054
9	0,025
10	0,006
11	0,003



Réponses aux besoins en analyse de données

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

```
SELECT code_dep, ROUND(AVG(prix/surf_carrez),2) AS prix_m2_my
FROM mutation AS a
JOIN lot AS b
ON a.id = b.id
JOIN localisation c
ON b.id = c.id
GROUP BY code_dep
ORDER BY prix_m2_my DESC LIMIT 10;
```

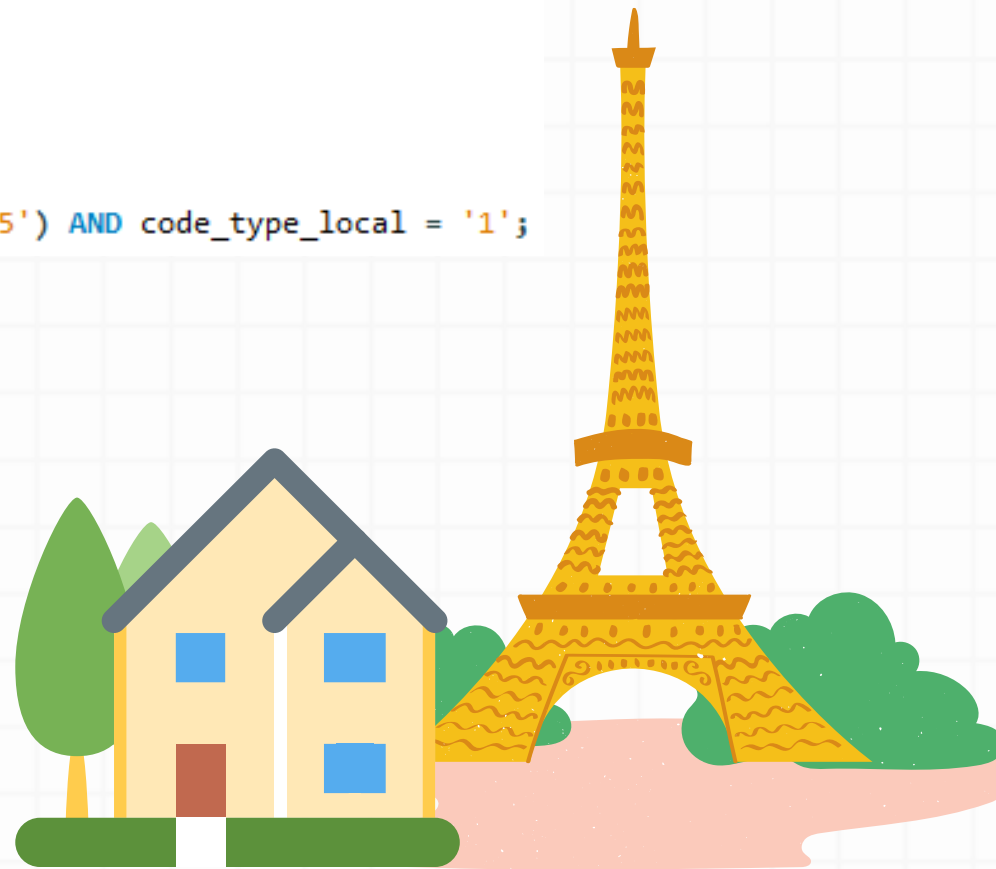


Département	Prix moyen du m ² (€)
75 - Paris	12091,87
92 - Hauts-de-Seine	7300,66
94 - Val-de-Marne	5430,68
74 - Haute-Savoie	4781,39
6 - Alpes Maritimes	4759,17
93 - Seine-Saint-Denis	4393,91
78 - Yvelines	4275,58
69 - Rhône	4100,33
33 - Gironde	3807,24
83 - Var	3766,35

Réponses aux besoins en analyse de données

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

```
SELECT code_type_local, ROUND(AVG(prix/surf_carrez),2) AS prix_m2_my
FROM mutation AS a
JOIN lot AS b
ON a.id = b.id
JOIN localisation c
ON b.id = c.id
WHERE code_dep IN ('75', '77', '78', '91', '92', '93', '94', '95') AND code_type_local = '1';
```



**3764,82€/m2 en moyenne
pour une maison en IDF**

Réponses aux besoins en analyse de données

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

```
SELECT code_type_local, cp, prix, surf_carrez
FROM mutation AS a
JOIN lot AS b
ON a.id = b.id
JOIN localisation c
ON b.id = c.id
WHERE b.code_type_local = '2'
ORDER BY prix DESC LIMIT 10;
```

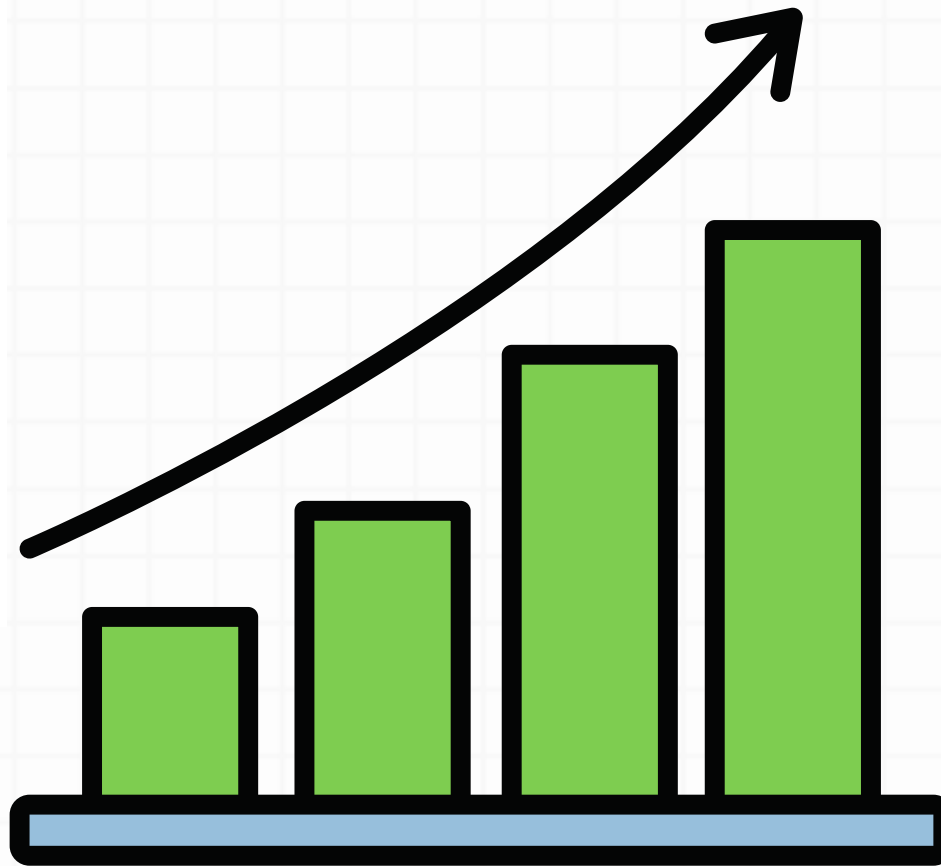
Type de local	CP	Prix (€)	Surface carrez (m²)
Appartement	75016	9000000,00	9,00
Appartement	91100	8600000,00	64,00
Appartement	75007	8577713,00	20,00
Appartement	75017	7620000,00	42,00
Appartement	75006	7600000,00	253,00
Appartement	75001	7535000,00	139,00
Appartement	75016	7420000,00	360,00
Appartement	75016	7200000,00	595,00
Appartement	75001	7050000,00	122,00
Appartement	75001	6600000,00	79,00

Réponses aux besoins en analyse de données

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

```
WITH
ventes_t1 AS (
  SELECT COUNT(id) AS vt1
  FROM mutation
  WHERE date_mutation BETWEEN '2020-01-01' AND '2020-03-31'),
ventes_t2 AS (
  SELECT COUNT(id) AS vt2
  FROM mutation
  WHERE date_mutation BETWEEN '2020-04-01' AND '2020-06-30')

SELECT ROUND(((vt2 - vt1) / vt1) * 100, 2) AS Taux_evolution_pourcentage
FROM ventes_t1
FULL JOIN ventes_t2;
```



Augmentation de **3,68% du nombre de ventes
entre le 1er et le 2e trimestre de 2020**

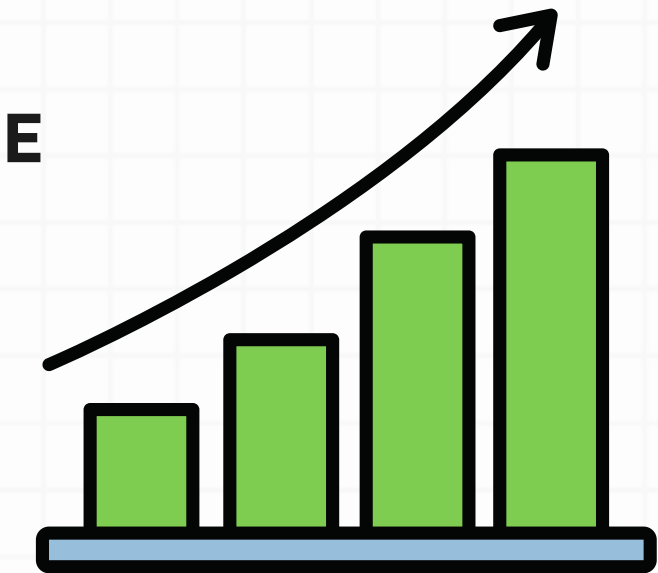
Réponses aux besoins en analyse de données

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

```
WITH
ventes_t1_com AS (
  SELECT commune, COUNT(a.id) AS vt1_com
  FROM mutation AS a
  JOIN localisation AS b
  ON a.id = b.id
  WHERE date_mutation BETWEEN '2020-01-01' AND '2020-03-31'
  GROUP BY commune),
ventes_t2_com AS (
  SELECT commune, COUNT(a.id) AS vt2_com
  FROM mutation AS a
  JOIN localisation AS b
  ON a.id = b.id
  WHERE date_mutation BETWEEN '2020-04-01' AND '2020-06-30'
  GROUP BY commune)

SELECT *, ROUND(((vt2_com - vt1_com) / vt1_com) * 100, 2) AS taux_evolution_pourcentage
FROM ventes_t1_com
FULL JOIN ventes_t2_com USING (commune)
WHERE ROUND(((vt2_com - vt1_com) / vt1_com) * 100, 2) >= '20'
ORDER BY taux_evolution_pourcentage DESC;
```

PAU
CAVAILLON
L ISLE SUR LA SORGUE
LE DEVOLUY
MORLAIX
LYON 8EME
RONCHIN
LYON 7EME
CHALON-SUR-SAONE
ROUEN
LUCE
LYON 3EME
RETHEL
LAMBALLE-ARMOR
LAMORLAYE ... (Total : 572)



Réponses aux besoins en analyse de données

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

```
WITH
prix_m2_2p AS (
  SELECT ROUND(AVG(prix/surf_carrez),2) AS prix_m2_my_2p
  FROM mutation AS a
  JOIN lot AS b
  ON a.id = b.id
  WHERE nb_pieces = '2'),
prix_m2_3p AS (
  SELECT ROUND(AVG(prix/surf_carrez),2) AS prix_m2_my_3p
  FROM mutation AS a
  JOIN lot AS b
  ON a.id = b.id
  WHERE nb_pieces = '3')

SELECT ROUND(((prix_m2_my_3p - prix_m2_my_2p) / prix_m2_my_2p) * 100, 2) AS diff_pourcentage
FROM prix_m2_2p
FULL JOIN prix_m2_3p;
```



Prix moyen du m2 pour un 2 pièces (4931.37€) plus cher de **13% par rapport au prix moyen du m2 pour un 3 pièces (4282.06€)**

Réponses aux besoins en analyse de données

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

```
WITH
prix_my_par_commune_selec AS (
  SELECT commune, code_dep, ROUND(AVG(prix),2) AS prix_my
  FROM mutation AS a
  JOIN localisation AS b
  ON a.id = b.id
  WHERE code_dep IN ('6', '13', '33', '59', '69')
  GROUP BY commune
  ORDER BY prix)

SELECT commune, code_dep, prix_my
FROM (
  SELECT ROW_NUMBER() OVER (PARTITION BY code_dep ORDER BY prix_my DESC) AS row_num, commune, code_dep, prix_my
  FROM prix_my_par_commune_selec
) AS top3
WHERE top3.row_num <= 3
ORDER BY code_dep;
```

	Commune	Département	Prix (€)
Alpes Maritimes	SAINT-JEAN-CAP-FERRAT	6	968750,00
	EZE	6	655000,00
	MOUANS-SARTOUX	6	476898,13
Bouches-du-Rhône	GIGNAC-LA-NERTHE	13	330000,00
	SAINT SAVOURNIN	13	314425,00
	CASSIS	13	313416,88
Gironde	LEGE-CAP-FERRET	33	549500,64
	VAYRES	33	335000,00
	ARCACHON	33	307435,93
Nord	BERSEE	59	433202,00
	CYSOING	59	408550,00
	HALLUIN	59	322250,00
Rhône	VILLE SUR JARNIOUX	69	485300,00
	LYON 2EME	69	455217,26
	LYON 6EME	69	426968,25

**Merci pour
votre
attention**

