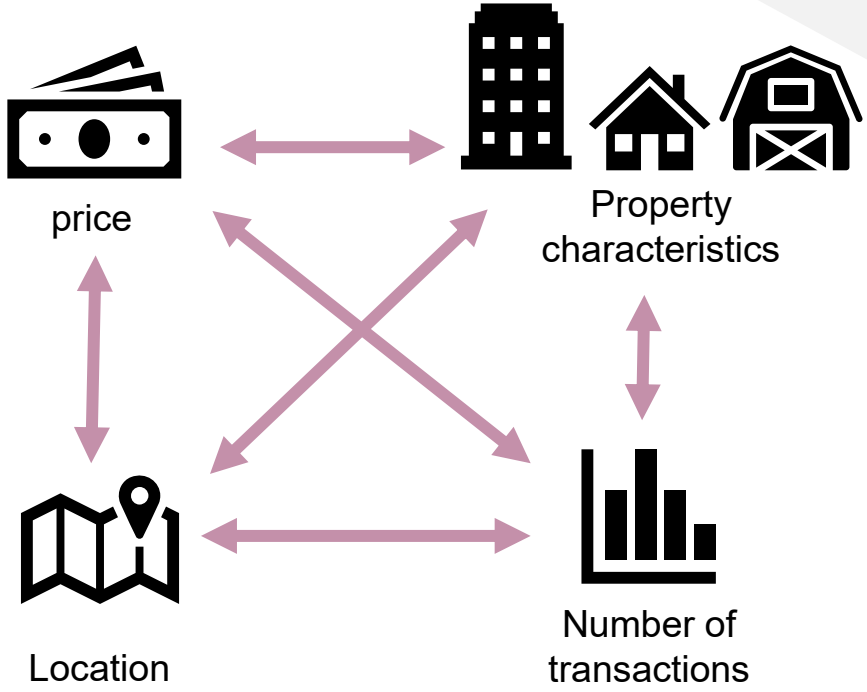
# Creation and use of the DATAImmo database

Vivien Pichon

Laplace Immo

# DATAImmo

Proof of concept: Database to follow the evolution of real estate prices




# Formatting data

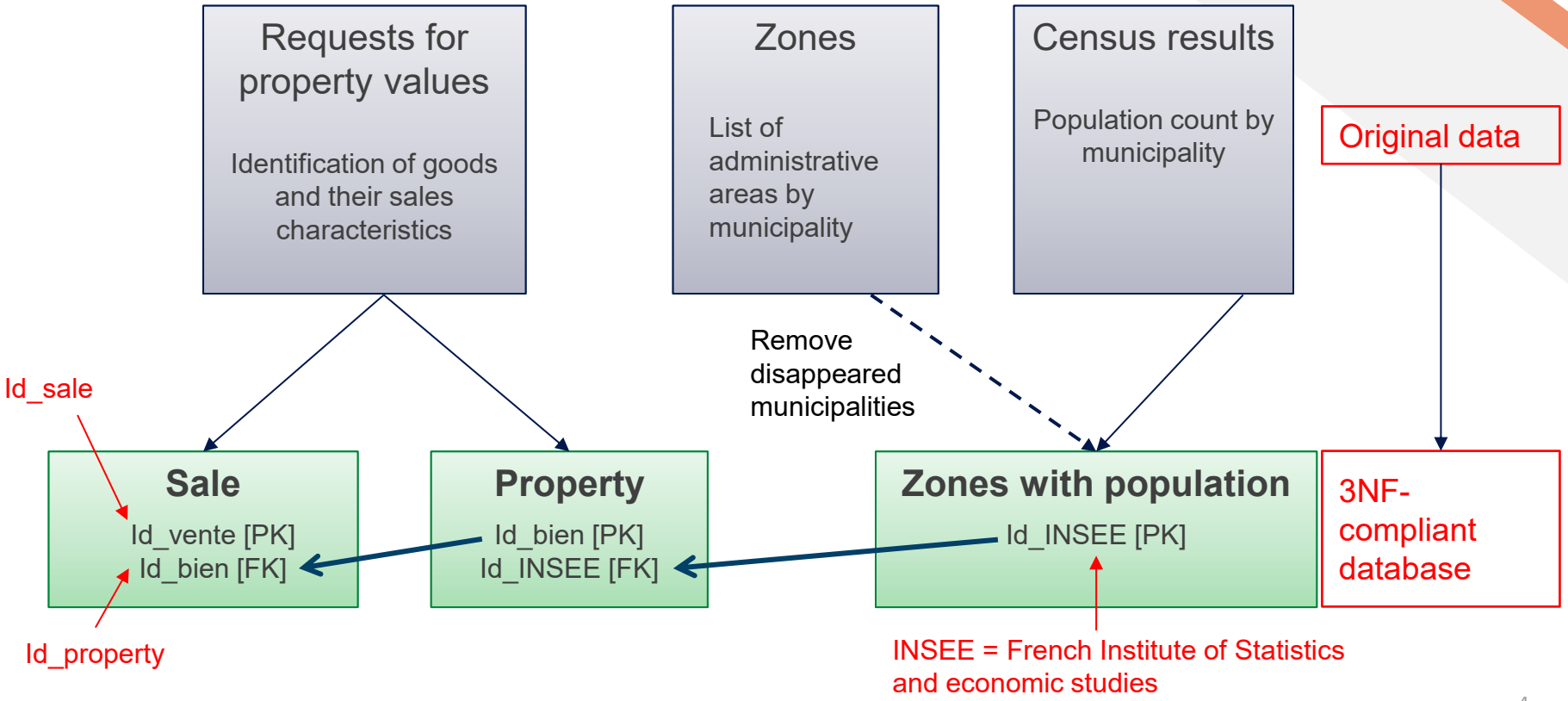
The property value table does not correspond to the 3NF standard

| Requests for property values | Zones | Census results |
| --- | --- | --- |
| Identification of goods and their sales characteristics | List of administrative areas by municipality | Population count by municipality |

Original data

# Formatting data

The property value table does not correspond to the 3NF standard



# Data dictionary and RGPD

## Explanation of the table of property values

Only part of the data dictionary is shown here. The full document is available in the repository

**RAW DATA - Property values**

Concatenated columns to create the Property table PK.
Sales (Vente) Table
Property (Bien) Table

| CODE | MEANING | TYPE | LENGTH | NATURE | MANAGEMENT RULE | CALCULATION RULE |
|---|---|---|---|---|---|---|
| Code service CH | NA | | | | | |
| Reference document | NA | | | | | |
| 1 Articles CGI | NA | | | | | |
| 2 Articles CGI | NA | | | | | |
| 3 Articles CGI | NA | | | | | |
| 4 Articles CGI | NA | | | | | |
| 5 Articles CGI | NA | | | | | |
| No disposition | Number of mutation per deed of sale | Integer | 2 | Elementary | Not null | |
| mutation date | Date of signature of the deed of sale | Date | 10 | Elementary | Not null ; Format AAAA/MM/JJ | |
| Nature mutation | Nature of the mutation | Varchar | 35 | Elementary | Not null ; choose from : Sale, sale pending completion, sale of building land, auction, expropriation, exchange.. | |
| Valeur fonciere | Net selling price | Float | NC | Elementary | | |
| No voie | Street number | Integer | NC | Elementary | | |
| B/T/Q | Repetition index | Varchar | 1 | Elementary | | |
| Code type de voie | Number corresponding to a track type | Integer | 2 | Elementary | | |
| Type de voie | Track type (street, avenue, path, etc.) | Varchar | 4 | Elementary | | |
| Code voie | Rivoli code | Varchar | 4 | Elementary | | |

Not available (with respect to the RGPD)

Sales table

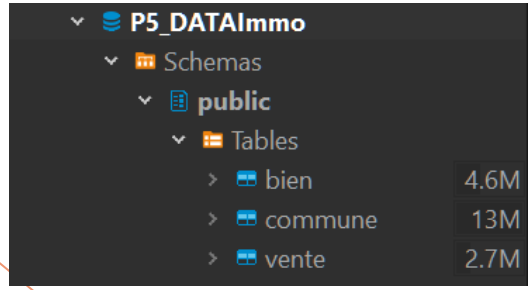
Property table

Primary key of the property table

Primary key of the sales table to be generated (incremented integer)

# Screenshots of the created database

The 3 tables




Overview of **commune** table

| | id_insee | code_region | code_dep | code_arron | code_canton | code_c |
|---|---|---|---|---|---|---|
| 1 | 01001 | 84 | 01 | 2 | 8 | 001 |
| 2 | 01002 | 84 | 01 | 1 | 1 | 002 |
| 3 | 01004 | 84 | 01 | 1 | 1 | 004 |
| 4 | 01005 | 84 | 01 | 2 | 22 | 005 |
| 5 | 01006 | 84 | 01 | 1 | 4 | 006 |
| 6 | 01007 | 84 | 01 | 1 | 1 | 007 |
| 7 | 01008 | 84 | 01 | 1 | 1 | 008 |
| 8 | 01009 | 84 | 01 | 1 | 4 | 009 |
| 9 | 01010 | 84 | 01 | 1 | 10 | 010 |
| 10 | 01011 | 84 | 01 | 4 | 14 | 011 |

```
SELECT count(*)
FROM commune c
```

| | count |
|---|---|
| 1 | 34,991 |

Expected number of rows

# Screenshots of the created database

Overview of the **vente** table

| | id_vente | id_bien | date | valeur |
|---|---|---|---|---|
| 1 | 1 | 347020117012 | 2020-01-02 | 165,000 |
| 2 | 2 | 41510006160132 | 2020-01-02 | 355,680 |
| 3 | 3 | 20B03975600099 | 2020-01-02 | 229,500 |
| 4 | 4 | 550310116700242 | 2020-01-02 | 125,000 |
| 5 | 5 | 930018A08413400218 | 2020-01-02 | 90,000 |
| 6 | 6 | 270791136002 | 2020-01-02 | 93,000 |
| 7 | 7 | 3601751513008112 | 2020-01-02 | 298,100 |
| 8 | 8 | 5076F25A312130121080 | 2020-01-02 | 163,500 |

Overview of the **bien** table

| | id_bien | id_insee | no_voie | btq | type_voie | voie |
|---|---|---|---|---|---|---|
| 1 | 347020117012 | 01103 | 347 | | RUE | DU CHATE |
| 2 | 415100061601 | 06004 | 4 | | BD | EDOUARD |
| 3 | 20B0397560009 | 06088 | 20 | B | RUE | MARCEAU |
| 4 | 5503101167002 | 06123 | 550 | | RTE | DES VESPI |
| 5 | 930018A08413 | 13005 | 9,300 | | RES | LES ARPEC |
| 6 | 270791136002 | 13028 | 27 | | RUE | DU GRANI |
| 7 | 3601751513008 | 13208 | 360 | | AV | DU PRADO |
| 8 | 5076F25A3121 | 13212 | 5,076 | F | PARC | DESSUARI |
| 9 | 1194010901451 | 14338 | 1,194 | | RUE | DE NORM |
| 10 | 302103714100 | 14366 | 30 | | ALL | DES NOIS |

```
SELECT count(*)
FROM vente v
```

| | count |
|---|---|
| 1 | 34,169 |

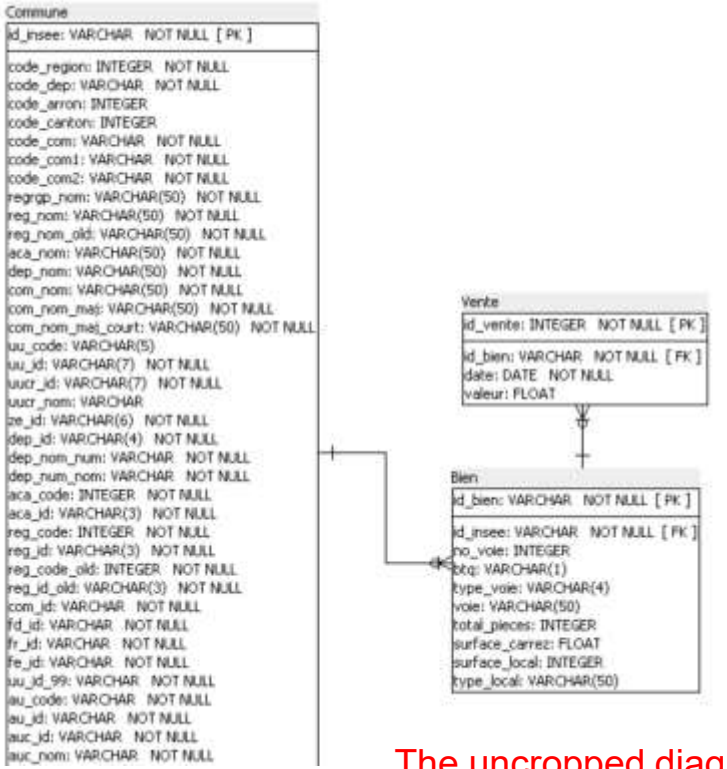
```
SELECT count(*)
FROM bien b
```

| | count |
|---|---|
| 1 | 34,169 |

Expected number of rows

# Standardised relational diagram

And other changes




**Removal of the postcode**
Risk of several many-to-many connections when we implement all the data

**The INSEE code is not an integer**
Corsican codes contain letters

**The property value is not an integer**

The uncropped diagram is in the appendix 1

# Functional database, SQL queries and results

Screenshots of the requests and results in DBeaver.

## 1. Total number of flats sold in the first half of 2020.

```
SELECT COUNT(*)
FROM vente v, bien b
WHERE v.id_bien = b.id_bien
AND
b.type_local = 'Appartement';
```

| | count |
|---|---|
| 1 | 31,378 |

## 2. Number of flat sales by region for the 1st half of 2020.

```
SELECT c.reg_nom, COUNT(v.*) AS
vente_count
FROM commune c ,vente v ,bien b
WHERE c.id_insee = b.id_insee AND
v.id_bien = b.id_bien AND
b.type_local = 'Appartement'
GROUP BY c.reg_nom
ORDER BY vente_count DESC;
```

region name

| | reg_nom | count |
|---|---|---|
| 1 | Ile-de-France | 13,995 |
| 2 | Provence-Alpes-C♦te d'Azur | 3,649 |
| 3 | Auvergne-Rh♦ne-Alpes | 3,253 |
| 4 | Nouvelle-Aquitaine | 1,932 |
| 5 | Occitanie | 1,640 |
| 6 | Pays de la Loire | 1,357 |
| 7 | Hauts-de-France | 1,254 |
| 8 | Grand Est | 984 |
| 9 | Bretagne | 983 |
| 10 | Normandie | 862 |
| 11 | Centre-Val de Loire | 696 |
| 12 | Bourgogne-Franche-Comt♦ | 376 |
| 13 | Corse | 223 |
| 14 | Martinique | 94 |
| 15 | La R♦union | 44 |
| 16 | Guyane | 34 |
| 17 | Guadeloupe | 2 |

## 3. Proportion of flat sales by number of rooms.

```
SELECT b.total_pieces ,
count(v.id_vente) ,
round(count(v.id_vente)*100.0 /
sum(count(v.id_vente)) OVER(), 2 ) AS
fraction
FROM bien b , vente v
WHERE v.id_bien = b.id_bien AND
b.type_local = 'Appartement'
GROUP BY b.total_pieces
ORDER BY b.total_pieces ;
```

Number of rooms

| | total_pieces | count | fraction |
|---|---|---|---|
| 1 | 0 | 30 | 0.1 |
| 2 | 1 | 6,739 | 21.48 |
| 3 | 2 | 9,783 | 31.18 |
| 4 | 3 | 8,966 | 28.57 |
| 5 | 4 | 4,460 | 14.21 |
| 6 | 5 | 1,114 | 3.55 |
| 7 | 6 | 204 | 0.65 |
| 8 | 7 | 54 | 0.17 |
| 9 | 8 | 17 | 0.05 |
| 10 | 9 | 8 | 0.03 |
| 11 | 10 | 2 | 0.01 |
| 12 | 11 | 1 | 0 |

## 4. List of the 10 departements with the highest price per square meter.

```
SELECT c.dep_nom_num, c.code_dep ,
avg(v.valeur/b.surface_carrez)::INTEGER AS prix_m2
FROM commune c ,vente v ,bien b
WHERE c.id_insee = b.id_insee AND
          v.id_bien = b.id_bien
GROUP BY c.dep_nom_num,
          c.code_dep
ORDER BY prix_m2 DESC
LIMIT 10 ;
```

Department name and code

| | dep_nom_num | prix_m2 |
|---|---|---|
| 1 | Paris (75) | 12,053 |
| 2 | Hauts-de-Seine (92) | 7,219 |
| 3 | Val-de-Marne (94) | 5,343 |
| 4 | Alpes-Maritimes (06) | 4,700 |
| 5 | Haute-Savoie (74) | 4,667 |
| 6 | Seine-Saint-Denis (93) | 4,345 |
| 7 | Yvelines (78) | 4,225 |
| 8 | Rh&#65533;ne (69) | 4,059 |
| 9 | Corse-du-Sud (2A) | 4,027 |
| 10 | Gironde (33) | 3,764 |

## 5. Average price per square meter of a house in Île-de-France.

```
SELECT avg(v.valeur/
b.surface_carrez)::INTEGER
FROM commune c , bien b , vente v
WHERE c.id_insee = b.id_insee AND
        v.id_bien = b.id_bien AND
        b.type_local = 'Maison' AND
        c.reg_code = 11
```

| | avg |
|---|---|
| 1 | 3,745 |

## 6. List of the 10 most expensive flats with the region and the number of square meters.

```
SELECT b.id_bien , v.valeur,
b.surface_carrez , c.reg_nom
FROM commune c , bien b , vente v
WHERE c.id_insee = b.id_insee AND
        v.id_bien = b.id_bien AND
        b.type_local = 'Appartement' AND
        v.valeur IS NOT NULL
ORDER BY v.valeur DESC
LIMIT 10
```

| | id_bien | valeur | surface_carrez | reg_nom |
|---|---|---|---|---|
| 1 | 615910775016853 | 9,000,000 | 9.1 | Ile-de-France |
| 2 | 16585091100322 | 8,600,000 | 64 | Ile-de-France |
| 3 | 104062075007100 | 8,577,713 | 20.55 | Ile-de-France |
| 4 | 33055037501715 | 7,620,000 | 42.77 | Ile-de-France |
| 5 | 72049975006104 | 7,600,000 | 253.3 | Ile-de-France |
| 6 | 8086417500179 | 7,535,000 | 139.9 | Ile-de-France |
| 7 | 3614141750165 | 7,420,000 | 360.95 | Ile-de-France |
| 8 | 23157947501611 | 7,200,000 | 595 | Ile-de-France |
| 9 | 2601449750012 | 7,050,000 | 122.56 | Ile-de-France |
| 10 | 1080863575001181 | 6,600,000 | 79.38 | Ile-de-France |

## 7. Rate of change in the number of sales between the first and second quarters of 2020.

```sql
WITH tri1 AS
(SELECT count(*)
FROM vente v
WHERE v."date" BETWEEN
'2020-01-01' AND '2020-03-31'),
tri2 AS
(SELECT count(*)
FROM vente v
WHERE v."date" BETWEEN
'2020-04-01' AND '2020-06-30')
SELECT ((tri2.count::NUMERIC
/tri1.count)-1)*100 AS
evol_ventes
FROM tri2, tri1;
```

| | evol_ventes |
|---|---|
| 1 | 3.6778731521 |

3.7% increase in the sales

## 8. Ranking of regions by price per square metre of flats with more than 4 rooms.

```sql
SELECT c.reg_nom, round(
avg(v.valeur/b.surface_carrez) ::
NUMERIC, 2) AS prix_m2
FROM commune c , bien b , vente v
WHERE c.id_insee = b.id_insee AND
        v.id_bien = b.id_bien AND
        b.type_local = 'Appartement' AND
        b.total_pieces > 4 AND
        v.valeur IS NOT NULL
GROUP BY c.reg_nom
ORDER BY prix_m2 DESC
```

region name

| | reg_nom | prix_m2 |
|---|---|---|
| 1 | Ile-de-France | 8,770.44 |
| 2 | La R□union | 3,641.81 |
| 3 | Provence-Alpes-C□te d'Azur | 3,587.65 |
| 4 | Corse | 3,104.88 |
| 5 | Auvergne-Rh□ne-Alpes | 2,891.38 |
| 6 | Nouvelle-Aquitaine | 2,465.48 |
| 7 | Bretagne | 2,412.05 |
| 8 | Pays de la Loire | 2,315.76 |
| 9 | Hauts-de-France | 2,189.93 |
| 10 | Occitanie | 2,097.23 |
| 11 | Normandie | 2,015.77 |
| 12 | Grand Est | 1,540.89 |
| 13 | Centre-Val de Loire | 1,453.11 |
| 14 | Bourgogne-Franche-Comt□ | 1,251.19 |
| 15 | Martinique | 573.48 |

## 9. List of towns with at least 50 sales in the 1st quarter

```
SELECT c.id_insee ,
c.dep_nom_num ,
c.com_nom ,
COUNT(v.id_vente) AS nb_ventes
FROM commune c , bien b , vente v
WHERE c.id_insee = b.id_insee AND
v.id_bien = b.id_bien AND
v."date" <= '2020.03.31'
GROUP BY c.id_insee
HAVING COUNT(v.id_vente) >= 50
ORDER BY c.com_nom ;
```

| | id_insee | dep_nom_num | com_nom | nb_ventes |
|---|---|---|---|---|
| 1 | 2A004 | Corse-du-Sud (2A) | Ajaccio | 54 |
| 2 | 49007 | Maine-et-Loire (49) | Angers | 64 |
| 3 | 06004 | Alpes-Maritimes (06) | Antibes | 77 |
| 4 | 92004 | Hauts-de-Seine (92) | Asni◆res-sur-Seine | 81 |
| 5 | 33063 | Gironde (33) | Bordeaux | 157 |
| 6 | 92012 | Hauts-de-Seine (92) | Boulogne-Billancourt | 99 |
| 7 | 92026 | Hauts-de-Seine (92) | Courbevoie | 80 |

## 10. Percentage difference in price per square meter between a 2-room flat and a 3-room flat.

```
WITH prix3 AS
( SELECT avg(v.valeur/b.surface_carrez) AS col3
FROM vente v , bien b
WHERE v.id_bien = b.id_bien AND
b.type_local = 'Appartement' AND
b.total_pieces = 3),
prix2 AS
(SELECT avg(v.valeur/b.surface_carrez) AS col2
FROM vente v , bien b
WHERE v.id_bien = b.id_bien AND
b.type_local = 'Appartement' AND
b.total_pieces = 2)
SELECT round(((prix3.col3 / prix2.col2)-1) ::
NUMERIC , 3)*100 AS difference
FROM prix3, prix2;
```

| | difference |
|---|---|
| 1 | -12.4 |

3-rooms flats are 12.4% cheaper per square meter than 2-rooms flats.

## 11. Average property values for the top 3 municipalities in the departments of 6, 13, 33, 59 and 69.

```
WITH depval AS (
SELECT c.id_insee , c.com_nom,
c.code_dep , c.dep_nom_num ,
avg(v.valeur) :: integer AS average
FROM commune c , bien b , vente v
WHERE c.id_insee = b.id_insee AND
v.id_bien = b.id_bien AND
c.code_dep IN ('06', '13', '33',
'59','69')
GROUP BY c.id_insee
ORDER BY average DESC)
SELECT *
FROM (SELECT ROW_NUMBER() OVER
(PARTITION BY code_dep ORDER BY
average DESC) AS rownb,
depval.*
FROM depval) AS reftab
WHERE reftab.rownb <=3
```

| | rownb | id_insee | com_nom | code_dep | dep_nom_num | average |
|---|---|---|---|---|---|---|
| 1 | 1 | 06121 | Saint-Jean-Cap-Ferrat | 06 | Alpes-Maritimes (06) | 968,750 |
| 2 | 2 | 06059 | Eze | 06 | Alpes-Maritimes (06) | 655,000 |
| 3 | 3 | 06084 | Mouans-Sartoux | 06 | Alpes-Maritimes (06) | 476,898 |
| 4 | 1 | 13043 | Gignac-la-Nerthe | 13 | Bouches-du-Rh\ufffdne (1 | 330,000 |
| 5 | 2 | 13101 | Saint-Savournin | 13 | Bouches-du-Rh\ufffdne (1 | 314,425 |
| 6 | 3 | 13022 | Cassis | 13 | Bouches-du-Rh\ufffdne (1 | 313,417 |
| 7 | 1 | 33236 | L\ufffdge-Cap-Ferret | 33 | Gironde (33) | 549,501 |
| 8 | 2 | 33539 | Vayres | 33 | Gironde (33) | 335,000 |
| 9 | 3 | 33009 | Arcachon | 33 | Gironde (33) | 307,436 |
| 10 | 1 | 59071 | Bers\ufffde | 59 | Nord (59) | 433,202 |
| 11 | 2 | 59168 | Cysoing | 59 | Nord (59) | 408,550 |
| 12 | 3 | 59279 | Halluin | 59 | Nord (59) | 322,250 |
| 13 | 1 | 69265 | Ville-sur-Jarnioux | 69 | Rh\ufffdne (69) | 485,300 |
| 14 | 2 | 69382 | Lyon 2e | 69 | Rh\ufffdne (69) | 455,217 |
| 15 | 3 | 69386 | Lyon 6e | 69 | Rh\ufffdne (69) | 426,968 |

12. The 20 towns with the most transactions per 1,000 inhabitants for towns with more than 10,000 inhabitants.

```
SELECT c.id_insee , c.com_nom,
c.dep_nom_num ,
round ( count(v.id_vente)*1000 ::
NUMERIC /c.pop_total, 2) AS
vente_parmille
FROM commune c , bien b , vente v
WHERE c.id_insee = b.id_insee AND
          v.id_bien = b.id_bien AND
          c.pop_total > 10000
GROUP BY c.id_insee
ORDER BY vente_parmille DESC
LIMIT 20
```

| | id_insee | com_nom | dep_nom_num | vente_parmille |
|---|---|---|---|---|
| 1 | 75102 | Paris2e | Paris (75) | 5.84 |
| 2 | 75101 | Paris1er | Paris (75) | 4.92 |
| 3 | 75103 | Paris3e | Paris (75) | 4.69 |
| 4 | 33009 | Arcachon | Gironde (33) | 4.62 |
| 5 | 44055 | La Baule-Escoublac | Loire-Atlantique (44) | 4.58 |
| 6 | 75104 | Paris4e | Paris (75) | 4.08 |
| 7 | 06104 | Roquebrune-Cap-Martin | Alpes-Maritimes (06) | 3.99 |
| 8 | 75108 | Paris8e | Paris (75) | 3.83 |
| 9 | 83123 | Sanary-sur-Mer | Var (83) | 3.5 |
| 10 | 75109 | Paris9e | Paris (75) | 3.43 |
| 11 | 83071 | La Londe-les-Maures | Var (83) | 3.43 |
| 12 | 75106 | Paris6e | Paris (75) | 3.38 |
| 13 | 83112 | Saint-Cyr-sur-Mer | Var (83) | 3.24 |
| 14 | 60141 | Chantilly | Oise (60) | 3.13 |
| 15 | 44132 | Pornichet | Loire-Atlantique (44) | 3.06 |
| 16 | 94067 | Saint-Mand  | Val-de-Marne (94) | 3.06 |
| 17 | 75110 | Paris 10e | Paris (75) | 3.04 |
| 18 | 06083 | Menton | Alpes-Maritimes (06) | 2.94 |
| 19 | 85226 | Saint-Hilaire-de-Riez | Vend  e (85) | 2.87 |
| 20 | 94080 | Vincennes | Val-de-Marne (94) | 2.81 |

# Thank you for your attention

# Appendix 1: Full standardized relational diagram