



Laplace Immo

PROJET DATAIMMO

Vincent ARESE
Formation Data Analyst

18/02/2022 – [OC] P03



Laplace Immo

PROJET DATAIMMO

Créer un modèle pour mieux prévoir le prix vente, analyser le marché à partir des données OpenData DVF qui ressement l'ensemble des transactions foncières en France.

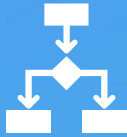
L'objectif est de se démarquer de la concurrence en permettant aux agences de mieux accompagner leurs clients pour leurs transactions immobilières.



Sommaire



PostgreSQL



- Modélisation données
 - Dictionnaire des données
 - Modèle conceptuel
 - Schéma Relationnel Normalisé



- Création des tables , normalisation & rapide des données
 - Importation des données & normalisation en SQL
 - Exploration et recherches des valeurs aberrantes : Etant donné que nous sommes au stade de POC ce jeux de données ne sera pas complètement nettoyé (trop chronophage et devra être refait sur la BDD complète pour le projet final). Bien entendus, le nettoyage sera effectivement réalisé sur le jeu de données final.
 - Recherches des Clés (Clés Primaires, Etrangères & Futurs Index pour optimiser le code SQL)



- Requêtes SQL pour les problématiques métiers



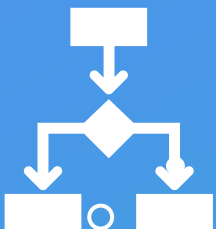
- Evaluation rapide de la fiabilité des jeux de données utilisés



- Proposition axes améliorations



DICTIONNAIRE DES DONNÉES



R	Table	Code propriété	Sinification	Exemple	Type	Observation
1,1	mutation	mutationID	identifiant unique d'une vente (N°de ligne)	2053	Entier	Clé Primaire Table
1,2	mutation	mutationyear	Année (date de mutation)	2020	numérique	Futur index table mutation
1,3	mutation	mutationquarter	trimestre (date de mutation)	2	numérique	Futur index table mutation
1,4	mutation	mutationdate	Date de signature de l'acte de vente chez le notaire format JJ/MM/AAAA)	24/06/2021	date	
1,6	mutation	cadastreidfk				Clé étrangère (référence table cadastre)
1,5	mutation	valeurfonciere	Il s'agit du montant ou de l'évaluation déclaré(e) dans le cadre d'une mutation à titre onéreux en Euros . La valeur foncière : • inclut : les frais d'agence, s'ils sont à la charge du vendeur, l'éventuelle TVA. • exclut : les frais d'agence, s'ils sont à la charge de l'acquéreur, les frais de notaires.	15300	numérique	Euros
2,1	cadastre	cadastreidtxt	Référence unique parcelle cadastrale (continué à partir des informations enregistrées au fichier du cadastre) : cadastrecodecommune_cadastreprefixesection_cadastresection_cadastreplannum_cadastrevolume_cadastreelot1	1--A-1623--40	alphanumérique	Clé primaire Table cadastre
2,2	cadastre	cadastrelot1surfacecarrez	La superficie loi Carrez en m² (La loi n°96-1107 du 18 décembre 1996, dite loi Carrez) correspond à une mesure spécifique de la superficie habitable d'un logement. Les éléments tels que murs, cloisons, marches, balcons, terrasses, embrasures de portes et fenêtres, gaines, cave et parking ne sont pas pris en compte	163	numérique	loi n°96-1107 du 18 décembre 1996, dite loi Carrez

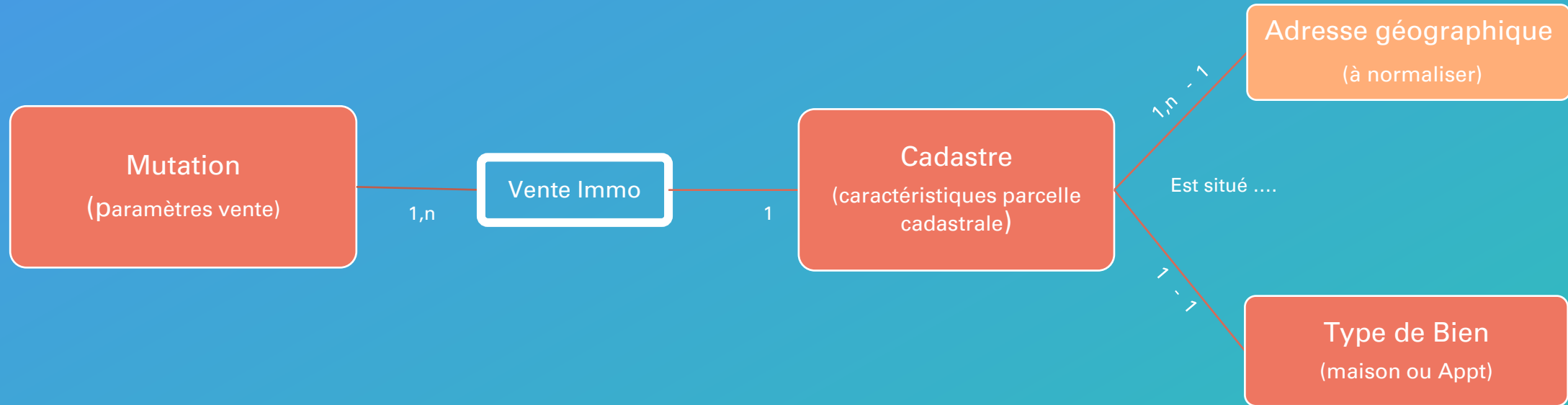
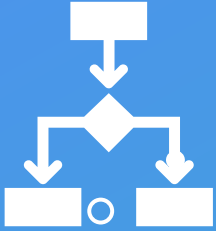


Cliquer pour voir
dictionnaire des données
complet sur Github



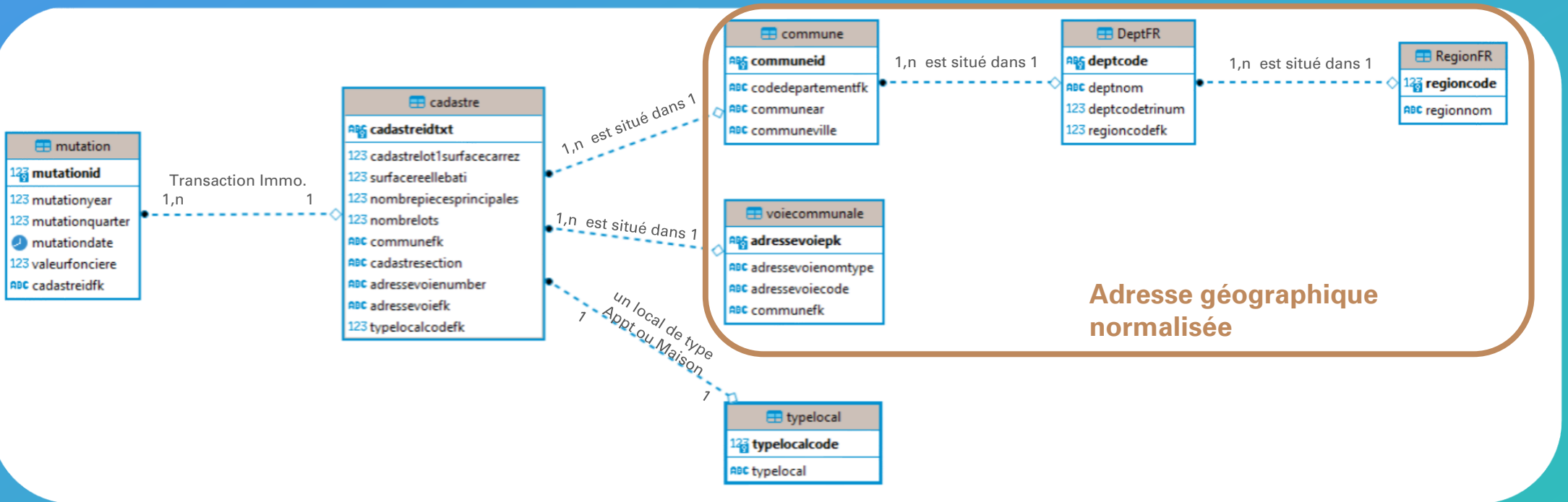
MODÉLISATION DES DONNÉES

- Modèle conceptuel des données



MODÉLISATION DES DONNÉES

- Schéma relationnel normalisé en 3NF



CRÉATION DE LA BDD POSTGRES SQL



- Création de la BDD



avec Dbeaver

PostgreSQL

- Importation des données

- RAW.csv dans PostgreSQL
- Import CSV Département FR et Région FR (réalisés à partir Excel)

RegionFR	
ABC regionnom	varchar(32)
123 regioncode	int4

DeptFR	
ABC deptnom	varchar
ABC deptcode	varchar
123 deptcodetrinum	int4
123 regioncodefk	int4

RAW		
123	mutationdispositionnum	int4
123	mutationdate	date
123	valeurfonciere	decimal
123	adressevoienum	int4
ABC	adressevoienumdetail	varchar
ABC	adressevoietype	varchar
ABC	adressevoiecode	varchar
ABC	adressevoienom	varchar
ABC	codepostal	varchar
ABC	commune	varchar
ABC	cadastrecode departement	varchar
ABC	cadastrecode commune	varchar
123	cadastre prefixe section	int4
ABC	cadastre section	varchar
123	cadastre plannul	int4
ABC	cadastre volume	varchar
ABC	cadastre lot1	varchar
123	cadastre lot1 surface carrez	decimal
123	nombre lots	int4
123	type local code	int4
123	surface reelle bati	decimal
123	nombre pieces principales	int4
123	rawid	serial4 NOT NULL



Cliquer Pour visualiser le Code
sur Github



NORMALISATION DE LA TABLE RAW

RAW		
123	mutationdispositionnum	int4
	mutationdate	date
123	valeurfonciere	decimal
123	adressevoienum	int4
ABC	adressevoienumdetail	varchar
ABC	adressevoietype	varchar
ABC	adressevoiecode	varchar
ABC	adressevoienom	varchar
ABC	codepostal	varchar
ABC	commune	varchar
ABC	cadastrecodedepartement	varchar
ABC	cadastrecodecommune	varchar
123	cadastreprefixesection	int4
ABC	cadastresection	varchar
123	cadastreplannul	int4
ABC	cadastrevolume	varchar
ABC	cadastrelot1	varchar
123	cadastrelot1surfacecarrez	decimal
123	nombrelots	int4
123	typelocalcode	int4
123	surfacereellebati	decimal
123	nombrepiecesprincipales	int4
123	rawid	serial4 NOT NULL

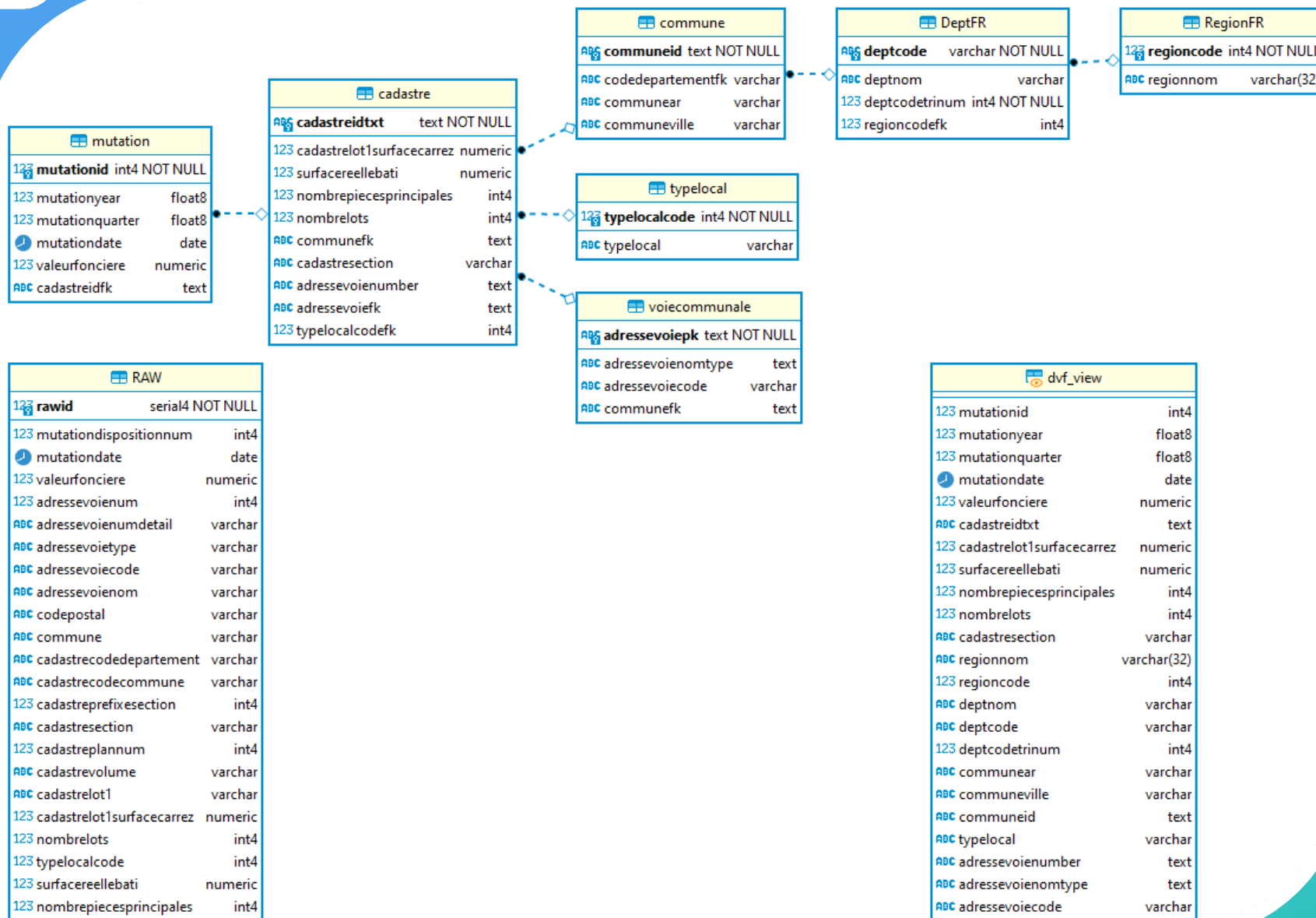
- Importation des données avec Dbeaver : Ajout Attribut rawID future PK numérique auto-incrémentée
- Recherche des clefs candidates , nettoyage rapide RAW & création des tables
 - **Table Type local** (à partir du TypeLocalCode 1: Maison – 2: Appartement)
 - **Table Commune** création Attribut CommuneVille : nettoyage attribut Commune (Paris, Lyon, Marseille)
 - **Table voiecommunale** Nettoyage raw correction erreurs saisie (tiret manquants dans les noms de rues, nom de rues partiels,)
 - **Table Cadastre** ajout attributs cadastresection(pour faire des recherche au niveau du quartier), et adresseVoieFK et adresseVoieNumber (pour pouvoir faire apparaître l'adresse et le numéro de rue dans les futures estimations)
 - **Table Mutation** : Création des attributs mutationdateY & mutationdateQ (trimestre)à partir de Mutation Date qui serviront ultérieurement d'index pour optimiser la rapidité des requêtes dans la BDD finale
 - **Définition** des PK (clef primaire) et FK (clef étrangère) pour toutes les tables



Cliquer Pour
visualiser le Code
sur Github



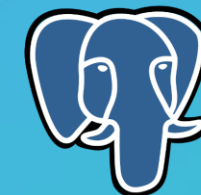
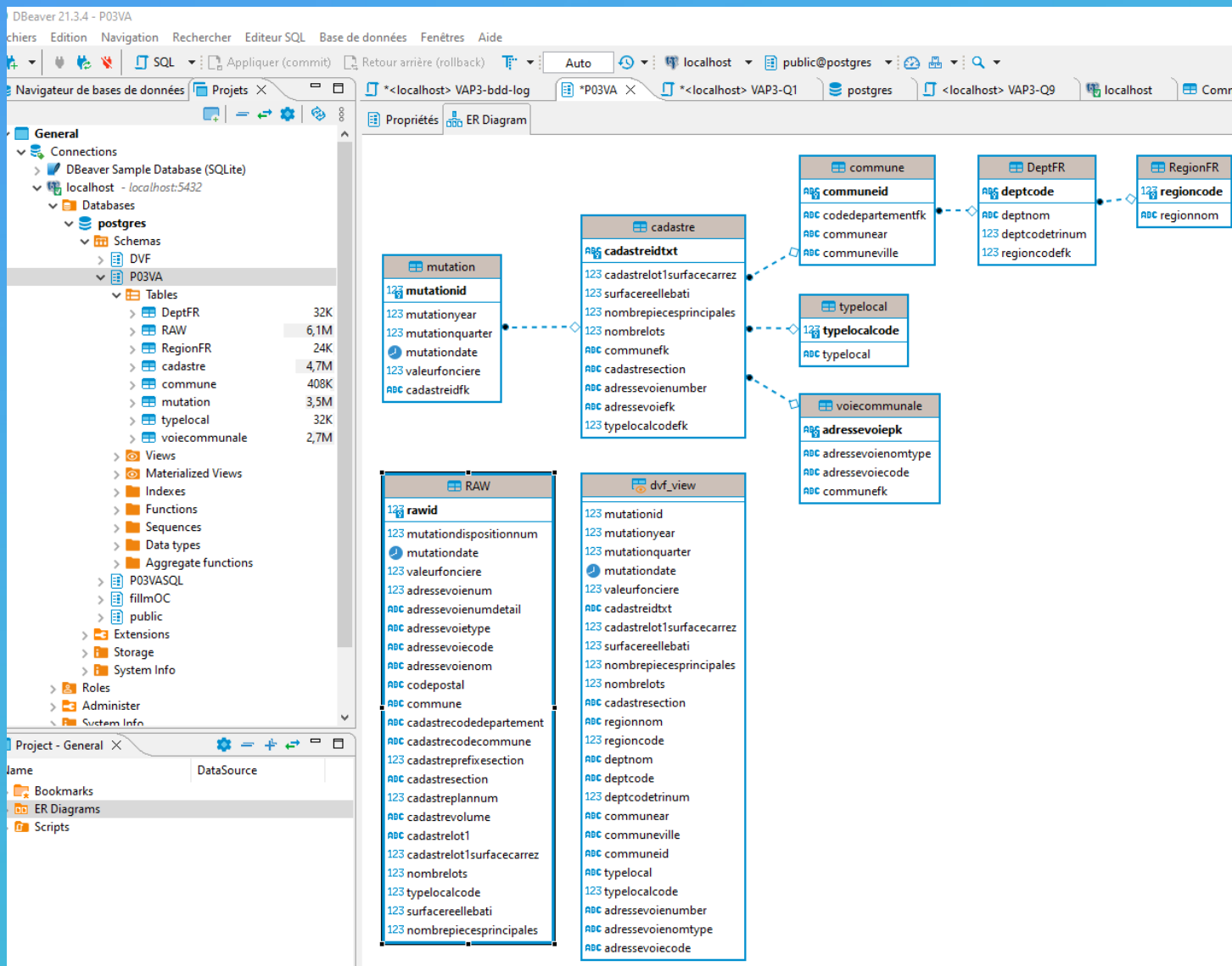
NORMALISATION DE LA TABLE RAW



Cliquer Pour
visualiser le Code
sur Github



SCHEMAT RELATIONNEL DE BDD DATAIMMO OPERATIONNELLE



PostgreSQL



Cliquer sur le logo
pour télécharger la
version source
community

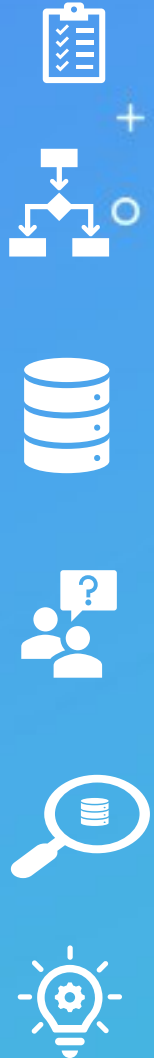


ANALYSE DES DONNÉES : REQUÊTES SQL



PostgreSQL

- 1. Nombre total d'appartements vendus au 1er semestre 2020.
- 2. Proportion des ventes d'appartements par le nombre de pièces.
- 3. Liste des 10 départements où le prix du mètre carré est le plus élevé.
- 4. Prix moyen du mètre carré d'une maison en Île-de-France.
- 5. Liste des 10 appartements les plus chers avec le département et le nombre de mètres carrés.
- 6. Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020.
- 7. Liste des communes où le nombre de ventes a augmenté d'au moins 20% entre le premier et le second trimestre de 2020
- 8. Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces.
- 9. Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69



REQUÊTES SQL 1

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

```
1 --Q1. Nombre total d'appartements vendus au 1er semestre 2020.  
2 -----  
3 select  
4 typelocal, count(typelocalcode) as nbr_appt  
5 FROM "P03VA".dvf_view  
6 where mutationyear = 2020 and mutationquarter <=2 and typelocalcode =2  
7 group by typelocal  
8 ;  
9
```

ABC typelocal	123 nbr_appt
Appartement	31 378



PostgreSQL



Cliquer Pour visualiser le Code
sur Github



REQUÊTES SQL 2



PostgreSQL

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

```
-- 2. Proportion des ventes d'appartements par le nombre de pièces.
-----
select
  typelocal as "Type Local",
  nombrepiecesprincipales as "Nombre Pieces Principales",
  count(nombrepiecesprincipales) as nbr,
  (Select count(typelocalcode) as totalappt FROM "P03VA".dvf_view as totalAppt where mutationyear = 2020 and mutationquarter <=2 and typelocalcode =2) as "Total Appt",
  ((count(nombrepiecesprincipales)::float/(Select count(typelocalcode) as totalappt FROM "P03VA".dvf_view as totalAppt where mutationyear = 2020 and mutationquarter <=2 and typelocalcode =2))*100)::DECIMAL(4,2) as "%"
FROM "P03VA".dvf_view
where mutationyear = 2020 and mutationquarter<=2 and typelocalcode=2
group by nombrepiecesprincipales , typelocal
order by nombrepiecesprincipales ;
```

	ABC Type Local	123 Nombre Pieces Principales	123 nbr	123 Total Appt	123 %
1	Appartement	0	30	31 378	0,1
2	Appartement	1	6 739	31 378	21,48
3	Appartement	2	9 783	31 378	31,18
4	Appartement	3	8 966	31 378	28,57
5	Appartement	4	4 460	31 378	14,21
6	Appartement	5	1 114	31 378	3,55
7	Appartement	6	204	31 378	0,65
8	Appartement	7	54	31 378	0,17
9	Appartement	8	17	31 378	0,05
10	Appartement	9	8	31 378	0,03
11	Appartement	10	2	31 378	0,01
12	Appartement	11	1	31 378	0



Cliquer Pour
visualiser le Code
sur Github



REQUÊTES SQL 3



PostgreSQL

- 3. Liste des 10 départements où le prix du mètre carré est le plus élevé.
- Les 10 dept ou le prix moyen au m² est le plus élevé / Le département des 10 biens ou le prix au m² est le plus élevé

```
-- Requete sur les 10 dept ou prix moyen du département au mètre carré est le plus élevé.
8 SELECT
9 deptcode, deptnom ,
10 round((AVG(valeurfonciere/surfacereellebati)),0) as Prix_m²_Moy_Dept
11 FROM "P03VA".dvf_view
12 where (valeurfonciere/surfacereellebati)>0
13 group by deptcode, deptnom
14 order by Prix_m²_Moy_Dept desc
15 limit 10;
```

	ABC deptcode	ABC deptnom	123 prix_m²_moy_dept
1	75	PARIS	12 129
2	92	HAUTS-DE-SEINE	7 415
3	94	VAL-DE-MARNE	5 399
4	06	ALPES-MARITIMES	4 685
5	93	SEINE-SAINT-DENIS	4 371
6	74	HAUTE-SAVOIE	4 150
7	78	YVELINES	4 126
8	69	RHONE	4 064
9	2A	CORSE-DU-SUD	3 922
10	33	GIRONDE	3 839

```
-- Requete sur les 10 dept ou le prix/m² d'un bien est le plus élevé.
18
19 SELECT
20 deptcode, deptnom ,
21 MAX(ROUND((valeurfonciere/surfacereellebati),0)) as Prix_m²_MAX_Dept
22 FROM "P03VA".dvf_view
23 where (valeurfonciere/surfacereellebati) is not NULL
24 group by deptcode, deptnom
25 order by Prix_m²_MAX_Dept desc
26 limit 10;
```

	ABC deptcode	ABC deptnom	123 prix_m²_max_dept
1	75	PARIS	900 000
2	94	VAL-DE-MARNE	175 000
3	91	ESSONNE	138 710
4	77	SEINE-ET-MARNE	91 076
5	59	NORD	80 833
6	92	HAUTS-DE-SEINE	78 261
7	06	ALPES-MARITIMES	70 200
8	01	AIN	58 250
9	44	LOIRE-ATLANTIQUE	57 143
10	30	GARD	42 075



Cliquer Pour
visualiser le Code
sur Github



REQUÊTES SQL 4



PostgreSQL

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

```
1 --4. Prix moyen du mètre carré d'une maison en Île-de-France.  
2 -----  
3  
4 --- Ok avec arrondi  
5 SELECT  
6 round( (AVG((valeurfonciere/surfacereellebati))),0) as Prix_m²_Bati_Maison_IDF  
7 FROM "P03VA".dvf_view  
8 where regioncode=11 and typelocalcode=1  
9 ;
```



Cliquer Pour
visualiser le Code
sur Github

123 prix_m²_bati_maison_idf	🔍 ⬆
	3 998



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

```
--5. Liste des 10 appartements les plus chers avec le département et le nombre de mètres carrés.
1
2
3 SELECT
4 typelocal,
5 valeurfonciere,
6 cadastrelot1surfacecarrez,
7 surfacereellebati,
8 deptnom,
9 (SELECT ROUND((valeurfonciere/surfacereellebati),2)) as Prix_m²_Bati,
10 (SELECT ROUND((valeurfonciere/cadastrelot1surfacecarrez),2))as Prix_m²_Carrez,
11 communeid
12 FROM "P03VA".dvf_view
13 where typelocalcode=2 and valeurfonciere>0
14 order by valeurfonciere desc ,deptcodetrinum
15 limit 10
16 ;
```



Cliquer Pour
visualiser le Code
sur Github

	ABC typelocal	123 valeurfonciere	123 cadastrelot1surfacecarrez	123 surfacereellebati	ABC deptnom	123 prix_m²_bati	123 prix_m²_carrez	ABC communeid
1	Appartement	9 000 000	9,1	10	PARIS	900 000	989 010,99	75-PARIS 16
2	Appartement	8 600 000	64	62	ESSONNE	138 709,68	134 375	91-CORBEIL ESSONNES
3	Appartement	8 577 713	20,55	289	PARIS	29 680,67	417 406,96	75-PARIS 07
4	Appartement	7 620 000	42,77	42	PARIS	181 428,57	178 162,26	75-PARIS 17
5	Appartement	7 600 000	253,3	200	PARIS	38 000	30 003,95	75-PARIS 06
6	Appartement	7 535 000	139,9	143	PARIS	52 692,31	53 859,9	75-PARIS 01
7	Appartement	7 420 000	360,95	357	PARIS	20 784,31	20 556,86	75-PARIS 16
8	Appartement	7 200 000	595	241	PARIS	29 875,52	12 100,84	75-PARIS 16
9	Appartement	7 050 000	122,56	310	PARIS	22 741,94	57 522,85	75-PARIS 01
10	Appartement	6 600 000	79,38	76	PARIS	86 842,11	83 144,37	75-PARIS 01

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

```
1 -- 6. Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020.
2 -----
3 with
4 Q1 as (select count(*) as VQ1 FROM "P03VA".dvf_view where mutationyear=2020 and mutationquarter=1),
5 Q2 as (select count(*) as VQ2 FROM "P03VA".dvf_view where mutationyear=2020 and mutationquarter=2)
6 select VQ1,VQ2, VQ2-VQ1 as delta , (((VQ2-VQ1)::float/VQ1)*100)::DECIMAL(4,2) as "Taux variation %"
7 from Q1,Q2 ;
```



Cliquer Pour
visualiser le Code
sur Github

	123 vq1	123 vq2	123 delta	123 Taux variation %
1	16 776	17 393	617	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. Liste des communes où le nombre de ventes a augmenté d'au moins 20% entre le premier et le second trimestre de 2020
1
2
3 select TotalVentesComQ1_view.communeville ,
4 totalventescomq1 ,totalventescomq2,
5 totalventescomq2-totalventescomq1 as delta,
6 (((totalventescomq2-totalventescomq1)::float/totalventescomq1)*100)::DECIMAL (6,2) as "%Var nb ventes"
7 from TotalVentesComQ2_view,TotalVentesComQ1_view
8 where TotalVentesComQ2_view.communeville=TotalVentesComQ1_view.communeville
9 and (((totalventescomq2-totalventescomq1)::float/totalventescomq1)*100)::DECIMAL (6,2)>20
10 order by "%Var nb ventes" desc ;
11
```

	abc communeville	123 totalventescomq1	123 totalventescomq2	123 delta	123 %Var nb ventes
1	PAU	3	78	75	2 500
2	CAVAILLON	1	17	16	1 600
3	L ISLE SUR LA SORGUE	1	13	12	1 200
4	MORLAIX	1	11	10	1 000
5	LE DEVOLUY	1	11	10	1 000
6	RONCHIN	1	9	8	800
7	CHALON-SUR-SAONE	3	26	23	766,67
8	ROUEN	12	98	86	716,67
9	LUCE	1	8	7	700
10	MACON	1	7	6	600
11	ORANGE	2	14	12	600
12	LONS	1	7	6	600
13	RETHEL	1	7	6	600
14	LAMORLAYE	1	7	6	600
15	LAMBALLE-ARMOR	1	7	6	600
16	PERRAY-EN-YVELINES (LE)	1	7	6	600
17	VETRAZ-MONTHOUX	1	6	5	500
18	BELLEVILLE-EN-BEAUJOLAIS	1	6	5	500
19	NOGENT-LE-ROI	1	5	4	400
20	LE PLESSIS BELLEVILLE	1	5	4	400

123 count

546

```
--ventes Q1 2020 detail
21 create OR REPLACE VIEW TotalVentesComQ1_view as
22 select communeville , count(communeid) as TotalVentesComQ1
23 FROM "P03VA".dvf_view as VentesQ1
24 where mutationyear = 2020 and mutationquarter=1 and typelocalcode <=2
25 group by communeville
26 order by communeville asc
27
```

```
--ventes Q2 2020 detail
12 create OR REPLACE VIEW TotalVentesComQ2_view as
13 select communeville, count(communeid) as TotalVentesComQ2
14 FROM "P03VA".dvf_view as VentesQ2
15 where mutationyear = 2020 and mutationquarter=2 and typelocalcode <=2
16 group by communeville
17 order by communeville ASC;
18
```

	abc communeville	123 totalventescomq1	123 totalventescomq2	123 delta	123 %Var nb ventes
529	NOGENT SUR OISE	8	10	2	25
530	MERY-SUR-OISE	4	5	1	25
531	BONNEVILLE	4	5	1	25
532	VALENCE	28	35	7	25
533	PONTOISE	24	30	6	25
534	L'ILE-SAINT-DENIS	4	5	1	25
535	BATZ-SUR-MER	4	5	1	25
536	SAINT-GENIS-LAVAL	8	10	2	25
537	LONGPONT-SUR-ORGE	4	5	1	25
538	CADAUJAC	4	5	1	25
539	VAL DE BRIEY	4	5	1	25
540	SOUSTONS	8	10	2	25
541	RENNES	61	76	15	24,59
542	BOISSY-SAINT-LEGER	9	11	2	22,22
543	ORLY	9	11	2	22,22
544	LE CHESNAY-ROCQUENCOU	14	17	3	21,43
545	ETAMPES	24	29	5	20,83
546	LEUCATE	29	35	6	20,69



Cliquer Pour
visualiser le Cod
sur Github



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

```
1 --8. Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces.
2 -----
3
4 --OK code optimiséV1
5 with
6 P2P as (SELECT ROUND(AVG(valeurfonciere/cadastrelot1surfacecarrez),2) as PM2P FROM "P03VA".dvf_view where typelocalcode=2 and nombrepiecesprincipales=2 and ROUND((valeurfonciere/cadastrelot1surfacecarrez),2)>0),
7 P3P as (SELECT ROUND(AVG(valeurfonciere/cadastrelot1surfacecarrez),2) as PM3P FROM "P03VA".dvf_view where typelocalcode=2 and nombrepiecesprincipales=3 and ROUND((valeurfonciere/cadastrelot1surfacecarrez),2)>0)
8 select PM2P as Prix_m²_Carrez_2P, PM3P as Prix_m²_Carrez_3P, round( (((PM3P-PM2P)/PM2P)*100),2) as "Delta_3P_vs_2P_%
9 from P2P,P3P;
10
```

	123 prix_m²_carrez_2p	123 prix_m²_carrez_3p	123 Delta_3P_vs_2P_%
1	4 908,58	4 299,9	-12,4



Cliquer Pour
visualiser le Code
sur Github



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

```
--- TOP 3 Communes de l'ensemble des departements listés !
SELECT
deptnom,communeur ,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy
FROM "P03VA".dvf_view
where deptcodetrinum in (6,13,33,59,69) and valeurfonciere IS NOT null
group by communeur,deptnom
order by Prix_Immo_Moy DESC
limit 3
;
```

```
-- TOP 3 communes de chaque departements de la liste
--Version With .... AS & Union
with
Top06 as (SELECT deptnom,communeur,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy
FROM "P03VA".dvf_view
where deptcodetrinum in (6) and valeurfonciere IS NOT null group by communeur,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top13 as (select deptnom,communeur,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy
FROM "P03VA".dvf_view
where deptcodetrinum in (13) and valeurfonciere IS NOT null group by communeur,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top33 as (SELECT deptnom,communeur,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy
FROM "P03VA".dvf_view
where deptcodetrinum in (33) and valeurfonciere IS NOT null group by communeur,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top59 as (SELECT deptnom,communeur,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy
FROM "P03VA".dvf_view
where deptcodetrinum in (59) and valeurfonciere IS NOT null group by communeur,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top69 as (SELECT deptnom,communeur,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy
FROM "P03VA".dvf_view
where deptcodetrinum in (69) and valeurfonciere IS NOT null group by communeur,deptnom order by deptnom,Prix_Immo_Moy desc limit 3)
select *from Top06 union select *from Top13 union select *from Top33 union select *from Top59 union select *from Top69
order by deptnom,Prix_Immo_Moy desc;
```

	ABC deptnom	ABC communeur	123 prix_immo_moy
1	ALPES-MARITIMES	SAINT-JEAN-CAP-FERRAT	968 750
2	ALPES-MARITIMES	EZE	655 000
3	GIRONDE	LEGE-CAP-FERRET	549 501

	ABC deptnom	ABC communeville	123 prix_immo_moy
1	ALPES-MARITIMES	SAINT-JEAN-CAP-FERRAT	968 750
2	ALPES-MARITIMES	EZE	655 000
3	ALPES-MARITIMES	MOUANS-SARTOUX	476 898
4	BOUCHES-DU-RHONE	GIGNAC-LA-NERTHE	330 000
5	BOUCHES-DU-RHONE	SAINT SAVOURNIN	314 425
6	BOUCHES-DU-RHONE	CASSIS	313 417
7	GIRONDE	LEGE-CAP-FERRET	549 501
8	GIRONDE	VAYRES	335 000
9	GIRONDE	ARCACHON	307 436
10	NORD	BERSEE	433 202
11	NORD	CYSOING	408 550
12	NORD	HALLUIN	322 250
13	RHONE	VILLE SUR JARNIOUX	485 300
14	RHONE	COLLONGES-AU-MONT- D OR	414 817
15	RHONE	ROCHETAILLEE-SUR-SAONE	393 850

Top par commune

	ABC deptnom	ABC communeur	123 prix_immo_moy
1	ALPES-MARITIMES	SAINT-JEAN-CAP-FERRAT	968 750
2	ALPES-MARITIMES	EZE	655 000
3	ALPES-MARITIMES	MOUANS-SARTOUX	476 898
4	BOUCHES-DU-RHONE	GIGNAC-LA-NERTHE	330 000
5	BOUCHES-DU-RHONE	SAINT SAVOURNIN	314 425
6	BOUCHES-DU-RHONE	CASSIS	313 417
7	GIRONDE	LEGE-CAP-FERRET	549 501
8	GIRONDE	VAYRES	335 000
9	GIRONDE	ARCACHON	307 436
10	NORD	BERSEE	433 202
11	NORD	CYSOING	408 550
12	NORD	HALLUIN	322 250
13	RHONE	VILLE SUR JARNIOUX	485 300
14	RHONE	LYON 2EME	455 217
15	RHONE	LYON 6EME	426 968

Top par commune/arrondissement

Cliquer Pour
visualiser le Code
sur Github



- 9. Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69 ainsi que le prix médian (pour limiter impact valeurs extrêmes)

```
-- TOP 3 communes de chaque departements de la liste Commune Arrondissements + Prix MEDIAN !!!!!!!
with
Top06 as (SELECT deptnom,commune,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy,
percentile_cont(0.5) within group (order by "P03VA".dvf_view.valeurfonciere) as "Prix median"
FROM "P03VA".dvf_view where deptcodetrinum in (6) and valeurfonciere IS NOT null group by commune,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top13 as (select deptnom,commune,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy,
percentile_cont(0.5) within group (order by "P03VA".dvf_view.valeurfonciere) as "Prix median"
FROM "P03VA".dvf_view where deptcodetrinum in (13) and valeurfonciere IS NOT null group by commune,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top33 as (SELECT deptnom,commune,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy,
percentile_cont(0.5) within group (order by "P03VA".dvf_view.valeurfonciere) as "Prix median"
FROM "P03VA".dvf_view where deptcodetrinum in (33) and valeurfonciere IS NOT null group by commune,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top59 as (SELECT deptnom,commune,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy ,
percentile_cont(0.5) within group (order by "P03VA".dvf_view.valeurfonciere) as "Prix median"
FROM "P03VA".dvf_view where deptcodetrinum in (59) and valeurfonciere IS NOT null group by commune,deptnom order by deptnom,Prix_Immo_Moy desc limit 3),
Top69 as (SELECT deptnom,commune,round(( AVG(valeurfonciere)),0) as Prix_Immo_Moy ,
percentile_cont(0.5) within group (order by "P03VA".dvf_view.valeurfonciere) as "Prix median"
FROM "P03VA".dvf_view where deptcodetrinum in (69) and valeurfonciere IS NOT null group by commune,deptnom order by deptnom,Prix_Immo_Moy desc limit 3)
select *from Top06 union select *from Top13 union select *from Top33 union select *from Top59 union select *from Top69
order by deptnom,Prix_Immo_Moy desc;
```

	ABC deptnom	ABC commune	123 prix_immo_moy	123 Prix median
1	ALPES-MARITIMES	SAINT-JEAN-CAP-FERRAT	968 750	1 075 000
2	ALPES-MARITIMES	EZE	655 000	655 000
3	ALPES-MARITIMES	MOUANS-SARTOUX	476 898	365 000
4	BOUCHES-DU-RHONE	GIGNAC-LA-NERTHE	330 000	330 000
5	BOUCHES-DU-RHONE	SAINT SAVOURNIN	314 425	314 425
6	BOUCHES-DU-RHONE	CASSIS	313 417	285 750
7	GIRONDE	LEGE-CAP-FERRET	549 501	498 500
8	GIRONDE	VAYRES	335 000	335 000
9	GIRONDE	ARCACHON	307 436	260 000
10	NORD	BERSEE	433 202	433 202
11	NORD	CYSOING	408 550	408 550
12	NORD	HALLUIN	322 250	127 750
13	RHONE	VILLE SUR JARNIOUX	485 300	485 300
14	RHONE	LYON 2EME	455 217	370 000
15	RHONE	LYON 6EME	426 968	348 250

Cliquer Pour
visualiser le Code
sur Github



ANALYSE DES DONNÉES : REQUÊTES SQL

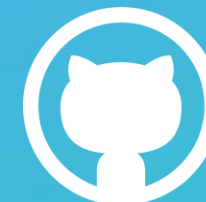


PostgreSQL

- 1. Nombre total d'appartements vendus au 1er semestre 2020.
- 2. Proportion des ventes d'appartements par le nombre de pièces.
- 3. Liste des 10 départements où le prix du mètre carré est le plus élevé.
- 4. Prix moyen du mètre carré d'une maison en Île-de-France.
- 5. Liste des 10 appartements les plus chers avec le département et le nombre de mètres carrés.
- 6. Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020.
- 7. Liste des communes où le nombre de ventes a augmenté d'au moins 20% entre le premier et le second trimestre de 2020
- 8. Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces.
- 9. Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69



FIABILITÉ DONNÉES DU DATA SET TEST ?



Cliquer Pour
visualiser le Code
sur Github

- Quelques valeurs Foncières manquantes (18 sur 34169 transactions immo)
- Quelques transactions sur des biens dont le nombre de pièces principales =0 (33 sur 34 169 transactions immo)
- Prix au m² douteux (souvent la surface, le nombre de pièces ne semblent pas corrélés)

	ABC typelocal	123 valeurfonciere	123 cadastrelot1surfacecarrez	123 surfacereellebati	123 nombrepiecesprincipales	123 prix_m²_bati	123 prix_m²_carrez	ABC deptnom	ABC communeid
1	Appartement	9 000 000	9,1	10	1	900 000	989 011	PARIS	75-PARIS 16
2	Appartement	1 290 000	115,37	5	1	258 000	11 181	PARIS	75-PARIS 16
3	Appartement	2 703 385	148,36	12	1	225 282	18 222	PARIS	75-PARIS 07
4	Appartement	2 634 000	173,01	13	1	202 615	15 225	PARIS	75-PARIS 12
5	Appartement	7 620 000	42,77	42	2	181 429	178 162	PARIS	75-PARIS 17
6	Appartement	175 000	9,62	1	1	175 000	18 191	VAL-DE-MARNE	94-MAISONS-ALFORT
7	Appartement	1 298 500	7,79	8	1	162 313	166 688	PARIS	75-PARIS 06
8	Appartement	8 600 000	64	62	3	138 710	134 375	ESSONNE	91-CORBEIL ESSONNES
9	Appartement	2 200 000	18,64	19	1	115 789	118 026	PARIS	75-PARIS 16
10	Appartement	1 900 000	116,33	17	1	111 765	16 333	PARIS	75-PARIS 18
11	Appartement	1 750 000	129,09	16	1	109 375	13 556	PARIS	75-PARIS 01
12	Appartement	1 724 500	96,86	17	1	101 441	17 804	PARIS	75-PARIS 09
13	Appartement	4 098 416,2	1 483,78	45	2	91 076	2 762	SEINE-ET-MARNE	77-BRIE COMTE ROBERT
14	Appartement	1 303 000	15,07	15	1	86 867	86 463	PARIS	75-PARIS 16
15	Appartement	6 600 000	79,38	76	3	86 842	83 144	PARIS	75-PARIS 01
16	Appartement	1 940 000	24,35	24	2	80 833	79 671	NORD	59-LILLE
17	Appartement	3 600 000	46,59	46	3	78 261	77 270	HAUTS-DE-SEINE	92-ASNIERES-SUR-SEINE
18	Appartement	1 321 200	18,66	18	1	73 400	70 804	PARIS	75-PARIS 10
19	Appartement	2 500 000	2 910,92	35	1	71 429	859	VAL-DE-MARNE	94-BOISSY-SAINT-LEGER
20	Appartement	2 350 000	73,77	13	1	71 154	13 530	PARIS	75-PARIS 02

FIABILITÉ DONNÉES DU DATA SET TEST ?

- Des prix et/ou surfaces à vérifier car données surprenantes : 166 mutations dont la surface moyennes des pièces est inférieure à 10m^2 ($347 \leq 10\text{m}^2$)

	ABC typelocal	123 valeurfonciere	123 carrez	123 surfacereellebati	123 Nb piece p.	123 Estim surf. moy/piece	123 prix_m²_bati	123 prix_m²_carrez	ABC deptnom	ABC communeid
1	Appartement	175 000	9,62	1	1	1	175 000	18 191	VAL-DE-MARNE	94-MAISONS-ALFORT
2	Appartement	242 000	15,35	11	10	1	22 000	15 765	PARIS	75-PARIS 04
3	Appartement	184 762	20,25	16	9	2	11 548	9 124	HAUTS-DE-SEINE	92-LEVALLOIS-PERRET
4	Appartement	1 290 000	115,37	5	1	5	258 000	11 181	PARIS	75-PARIS 16
5	Appartement	105 000	8,5	5	1	5	21 000	12 353	PARIS	75-PARIS 12
6	Appartement	296 500	63,67	25	5	5	11 860	4 657	VAL-DE-MARNE	94-SAINT-MAUR-DES-FOSSES
7	Appartement	15 000	3,64	5	1	5	3 000	4 121	PARIS	75-PARIS 08
8	Appartement	139 000	50,91	50	11	5	2 780	2 730	SEINE-MARITIME	76-ROUEN
9	Appartement	376 200	27	6	1	6	62 700	13 933	PARIS	75-PARIS 06
10	Appartement	151 000	11,72	6	1	6	25 167	12 884	PARIS	75-PARIS 06
11	Appartement	128 000	26,75	6	1	6	21 333	4 785	CALVADOS	14-TOURGEVILLE
12	Appartement	91 000	6,75	6	1	6	15 167	13 481	PARIS	75-PARIS 05
13	Appartement	89 000	6,26	6	1	6	14 833	14 217	PARIS	75-PARIS 02
14	Appartement	87 000	7,71	6	1	6	14 500	11 284	PARIS	75-PARIS 14
15	Appartement	80 000	6,25	6	1	6	13 333	12 800	PARIS	75-PARIS 14
16	Appartement	210 000	20,12	17	3	6	12 353	10 437	PARIS	75-PARIS 10
17	Appartement	70 000	7,16	6	1	6	11 667	9 777	PARIS	75-PARIS 12
18	Appartement	15 050	5,27	6	1	6	2 508	2 856	PARIS	75-PARIS 05
19	Appartement	23 000	17,44	11	2	6	2 091	1 319	EURE	27-EVREUX
20	Appartement	208 000	16,97	7	1	7	29 714	12 257	PARIS	75-PARIS 11



Cliquer Pour
visualiser le Code
sur Github

123 < 10m²	166
------------	-----

123 <= 10m²	347
-------------	-----



- 53,75% des Appartements ont une surface Réelle Bati < Surface Carrez !
(ce qui n'est pas possible et indique donc des erreurs de saisies ou un problème de fiabilité des données sources)

123 Nb Appt Pb Surface Bati < Carrez	123 Total Appt	123 % erreur
18 367	34 169	53,75

	123 delta Surface (Bati-Carrez)	123 surfacereellebati	123 Surface Carrez	123 Nb Pieces princip	mutationdate	123 valeurfonciere	adresse	123 rawid
1	-5 100	53	5 153	3	2020-02-05	265 000	90 DE LA JUSTICE 91-MONTGERON	21 858
2	-4 884	52	4 936	3	2020-02-20	223 645	2 DES 6 FR RUELLAN 35-SAINT-MALO	8 136
3	-2 875,92	35	2 910,92	1	2020-02-21	2 500 000	3 DES COQUELICOTS 94-BOISSY-SAINT-LEGER	27 145
4	-1 438,78	45	1 483,78	2	2020-02-07	4 098 416,2	12 DU PARC DES SPORTS 77-BRIE COMTE ROBERT	16 693
5	-1 191,44	100	1 291,44	3	2020-01-16	700 000	23 GAY LUSSAC 13-MARSEILLE 14EME	2 622
6	-734	81	815	3	2020-03-11	154 000	34 B VICTOR HUGO 17-SAUJON	3 978
7	-702,17	40	742,17	2	2020-02-27	1 683 000	5229 DE LA PRIVADIERE 30-GARRIGUES SAINTE EULALIE	5 422
8	-490,85	19	509,85	1	2020-02-28	285 000	37 DU COLLEGE 39-POLIGNY	9 079
9	-460	110	570	6	2020-04-23	95 000	5 A FERNAND BAZIN 02-COEUVRES ET VALSERY	250
10	-407,5	15	422,5	1	2020-03-06	116 000	68 DU PRESBYTERE 70-VILLERSEXEL	15 483
11	-403,3	47	450,3	2	2020-03-13	1 600 000	298 LAFAYETTE 69-LYON 3EME	14 587
12	-354	241	595	8	2020-03-26	7 200 000	23 DE BEAUSEJOUR 75-PARIS 16	32 135
13	-310	40	350	2	2020-05-19	153 000	1 SAINT ROMME 38-ROYBON	9 051
14	-303,07	48	351,07	3	2020-05-27	2 410 000	74 HENRI BARBUSSE 92-NANTERRE	23 553
15	-256,46	34	290,46	2	2020-03-05	147 000	26 DE LA LIBERTE 83-Draguignan	19 750
16	-246,58	25	271,58	1	2020-03-03	1 755 000	8 DES CRETES 06-NICE	644
17	-238	36	274	2	2020-05-13	450 000	1 B DE FETAN BERR 56-Languidic	12 339
18	-226,11	190	416,11	6	2020-06-19	1 643 570	26 THOMAS LEMAITRE 92-NANTERRE	23 589
19	-214	61	275	3	2020-02-20	215 000	9002 A DU CHAMPS DE MARS 26-DONZERE	4 509

- Par exemple :

Ligne1 : 3 pièces pour 5153m² Carrez (53m² Surface Réelle Bati)

Ligne2 : 2 pièces pour 4936m² Carrez (53m² Surface Réelle Bati)



Cliquer Pour
visualiser le Code
sur Github



- Pour valider si il s'agit d'un problème sur le jeu de données test, une recherche sur les mutations dont la surface réelle du bâti est < à la surface Carrez du lot 1 a été réalisée sur les données data.gouv DVF 2020.

```
--estimation avec %
select
(select count (*) FROM "DVF2020"."valeursfoncieres-2020.txt" where "Surface reelle bati"<"Surface Carrez du 1er lot") as "Nb Pb Surface Bati < Carrez",
(select count (*) FROM "DVF2020"."valeursfoncieres-2020.txt") as "Total mutations",
(((select count (*) FROM "DVF2020"."valeursfoncieres-2020.txt" where "Surface reelle bati"<"Surface Carrez du 1er lot")/
(select count (*) FROM "DVF2020"."valeursfoncieres-2020.txt")::float))*100)::decimal(4,2)
as "% erreurs"
FROM "DVF2020"."valeursfoncieres-2020.txt" limit 1.
```

	123 Nb Pb Surface Bati < Carrez	123 Total mutations	123 % erreurs
1	143 994	3 149 482	4,57

Pour le fichier de 2020 il y a 143 994 résultat douteux (surface bâti < surface carrez lot1) sur 3 149 492 mutation soit 4,57% de résultats douteux. Un nettoyage en profondeur de la base sera donc nécessaire pour exclure les mutations douteuses mais globalement il semble convenable.

=> C'est donc très probablement la data set de test qui a été corrompu au moment de la simplification des données !



Cliquer Pour
visualiser le Code
sur Github



SUGGESTIONS D'AXES D'AMÉLIORATIONS



- Intégrer les zones cadastrales pour filtrer des biens/transactions au niveau d'un quartier et les rues (pour les références dans les estimations)



- Enrichir la base avec d'autres données INSEE pour faire des recherches sur des zones géographiques ciblées pour les acheteurs (en fonctions de la localisation de leur travail / de leurs activités)
 - Bassins de vie : Le bassin de vie constitue le plus petit territoire sur lequel les habitants ont accès aux équipements et services les plus courants. (source [Insee](#))
 - Zone d'emploi : Une zone d'emploi est un espace géographique à l'intérieur duquel la plupart des actifs résident et travaillent, et dans lequel les établissements peuvent trouver l'essentiel de la main d'œuvre nécessaire pour occuper les emplois offerts. (source : [Insee](#))



- Enrichissement de la base de données avec d'autres sources (en fonction des retour /requêtes des agences lors de prochaines mises à jour de l'application)



- Création d'un module web (SAAS) pour faire les estimations (avec les références des biens vendus à proximité issues de la BDD DataImmo)





Laplace Immo

Synergy

SVA SA

PROJET DATAIMMO

Des questions avant de valider le POC ?

Planifier la mise en production d'une version démo DATAimmo ?





MERCI POUR VOTRE
ATTENTION



