



Création et utilisation de la base de données

Gabriel Gwynn



Laplace Immo

Contexte du projet

Nom : DATAImmo

- Réseau national d'agences
- Anticiper les prix des biens immobiliers
- Se démarquer de la concurrence
- Base de donnée, un outil pour aider les agences



La stratégie de sauvegarde et la conformité RGPD

- RGPD : les adresses sans les noms de personnes associées respectent le RGPD. Attention pour les appartements isolés ou maison individuelles qui pourrait être considéré comme données personnelles. (association indirecte)
- Stratégie de sauvegarde (conseil) :
 - sauvegarde complète hebdomadaire
 - sauvegardes différentielles quotidiennes (sauvegarde les données modifiées depuis la dernière complète)
 - test de restauration mensuel
 - stockage au choix sur différents supports (disque local, cloud sécurisé, serveur distant)

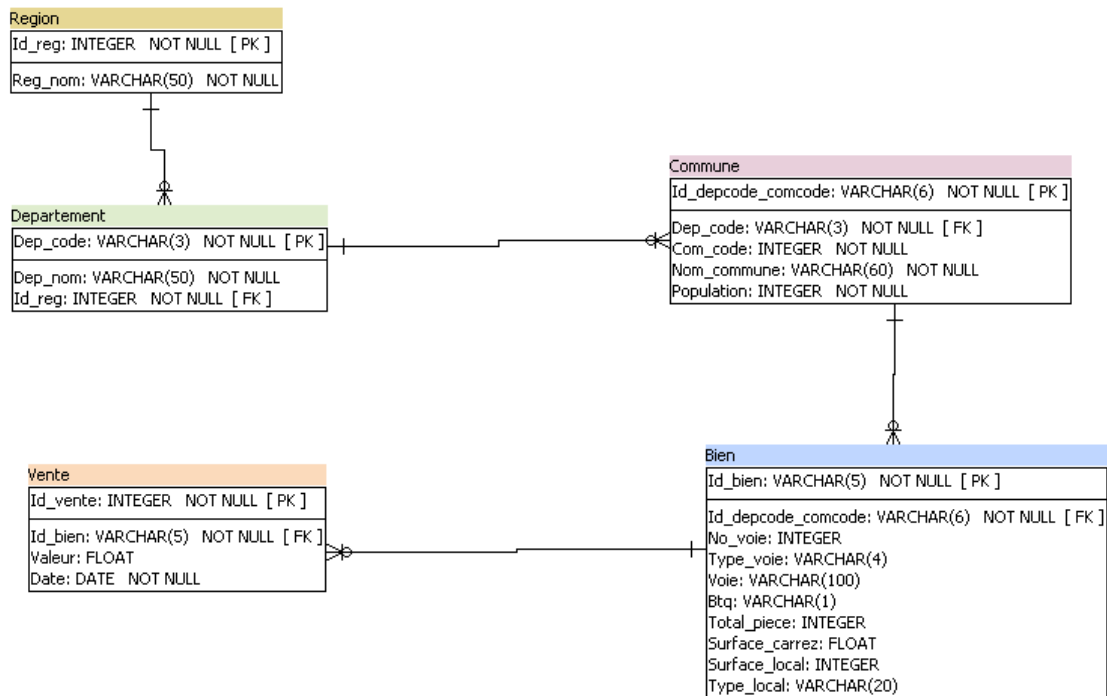
Les données initiales

- Demande des valeurs foncières : open data avec des données personnelles
- Données INSEE avec recensements de la population
- Données data.gouv référentiel géographique

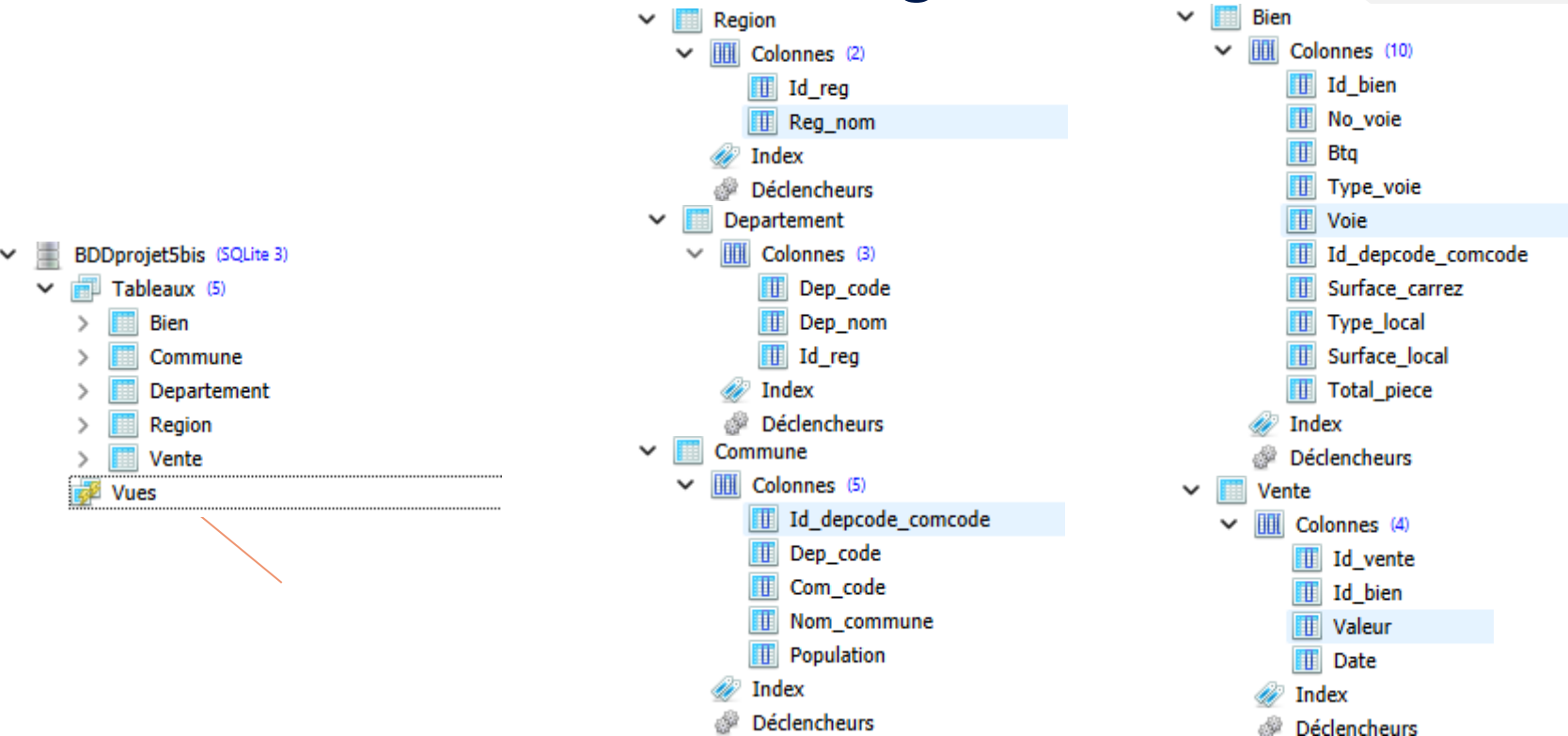
L'extrait du dictionnaire des données

	CODE	SIGNIFICATION	TYPE	LONGUEUR	CLE	NATURE	REGLE DE GESTION	REGLE DE CALCUL
Vente.cvs	Id_vente	clé vente	Integer		Clé primaire	calculé	not null	généré unique
	Id_bien	clé des biens unique	Varchar	5	Clé secondaire	calculé	not null	généré unique avec doublons renommé
	Valeur	valeur	Float					
	Date	date	date				date	
Bien.cvs	Id_bien	clé des biens unique	Varchar	5	Clé primaire	calculé	not null	généré unique + suppression doublon
	Id_depcode_comcode	clé des communes	Varchar	6	Clé secondaire	Concaténé	not null	code_département + code commune
	No_voie	numéro de la voie	Integer			Elémentaire		
	Btq	indication numéro bâtiment	Varchar	1		Elémentaire		
	Type_voie	type de la voie	Varchar	4		Elémentaire		
	Voie	nom de la voie	Varchar	100		Elémentaire		
	Total_piece	nombre de pièces	Integer			Elémentaire		
	Surface_carrez	surface loi carrez	Float			Elémentaire		
	Surface_local	surface reel	Integer			Elémentaire		
Commune.cvs	Id_depcode_comcode	clé des communes	Varchar	6	Clé primaire	Concaténé	not null	Dep_code+Com_code
	Dep_code	code du département	Varchar	3	Clé secondaire	Elémentaire	not null	
	Com_code	code de la commune	Integer			Elémentaire	not null	
	Nom_commune	nom de la commune	Varchar	60		Elémentaire	not null	
	Population	habitant commune	Integer			Elémentaire	not null	
Departement.cvs	Dep_code	code du département	Varchar	3	Clé primaire	Elémentaire	not null	
	Dep_nom	nom du département	Varchar	50		Elémentaire	not null	
	Id_reg	clé des régions	Integer		Clé secondaire	calculé	not null	généré unique
Region.cvs	Id_reg	clé des régions	Integer		Clé primaire	calculé	not null	généré unique
	Reg_nom	nom du département	Varchar	50		Elémentaire	not null	

Le schéma relationnel normalisé



La base de données avec les tables créées et les données chargées



Les requêtes ou screenshot qui permettent de démontrer le bon chargement des données

[10:39:15] Données importées dans le tableau 'Region' avec succès. Nombre de lignes importées : 19

[10:39:52] Données importées dans le tableau 'Departement' avec succès. Nombre de lignes importées : 109

[10:40:39] Données importées dans le tableau 'Commune' avec succès. Nombre de lignes importées : 34991

[10:41:14] Données importées dans le tableau 'Bien' avec succès. Nombre de lignes importées : 34154

[10:42:02] Données importées dans le tableau 'Vente' avec succès. Nombre de lignes importées : 34169



Requêtes SQL et résultats

Requête 1

Nombre total d'appartements vendus au 1er semestre 2020

```
SELECT COUNT(Vente.Id_bien)
FROM Vente
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
WHERE Bien.Type_local = 'Appartement'
AND Date BETWEEN '2020/01/01' AND '2020/06/30'
```

COUNT(Vente.Id bien)

31378

```
SELECT COUNT(DISTINCT Vente.Id_bien)
FROM Vente
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
WHERE Bien.Type_local = 'Appartement'
AND Date BETWEEN '2020/01/01' AND '2020/06/30'
```

COUNT(DISTINCT Vente.Id bien)

31364

Requête 2

Le nombre de ventes d'appartements par région pour le 1er semestre 2020

```
SELECT Reg_nom,COUNT(Id_vente) AS Nb_appartement
FROM Vente
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
JOIN Commune ON Bien.Id_depcode_comcode = Commune.Id_depcode_comcode
JOIN Departement ON Commune.Dep_code = Departement.Dep_code
JOIN Region ON Region.Id_reg = Departement.Id_reg
WHERE Bien.Type_local = 'Appartement'
AND Date BETWEEN '2020/01/01' AND '2020/06/30'
GROUP BY Reg_nom
ORDER BY Nb_appartement DESC;
```

	Reg_nom	Nb_appartement
1	Ile-de-France	13995
2	Provence-Alpes-Côte d'Azur	3649
3	Auvergne-Rhône-Alpes	3253
4	Nouvelle-Aquitaine	1932
5	Occitanie	1640
6	Pays de la Loire	1357
7	Hauts-de-France	1254
8	Grand Est	984
9	Bretagne	983
10	Normandie	862
11	Centre-Val de Loire	696
12	Bourgogne-Franche-Comté	376
13	Corse	223
14	Martinique	94
15	La Réunion	44
16	Guyane	34
17	Guadeloupe	2

Requête 3

Proportion des ventes d'appartements par le nombre de pièces

```
SELECT
    Total_piece,
    COUNT(Vente.Id_vente) AS 'Nombre appartements',
    ROUND(COUNT(Vente.Id_vente) * 100.0 / (
        SELECT COUNT(Vente.Id_vente)
        FROM Vente
        JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
        WHERE Bien.Type_local = 'Appartement'
    ), 2) AS Pourcentage
FROM Vente
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
WHERE Bien.Type_local = 'Appartement'
GROUP BY Total_piece;
```

Total_piece	Nombre appartements	Pourcentage
0	30	0.1
1	6739	21.48
2	9783	31.18
3	8966	28.57
4	4460	14.21
5	1114	3.55
6	204	0.65
7	54	0.17
8	17	0.05
9	8	0.03
10	2	0.01
11	1	0

Requête 4

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

```
SELECT Dep_code,Dep_nom,ROUND(SUM(Valeur)/SUM(Surface_local)) AS 'Prix du M^2'  
FROM Vente  
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien  
JOIN Commune ON Bien.Id_decode_comcode = Commune.Id_decode_comcode  
JOIN Departement ON Commune.Dep_code = Departement.Dep_code  
GROUP BY Dep_nom  
ORDER BY ROUND(SUM(Valeur)/SUM(Surface_local)) DESC  
LIMIT 10
```

	Dep_code	Dep_nom	Prix du M^2
1	75	Paris	11970
2	92	Hauts-de-Seine	7359
3	94	Val-de-Marne	5037
4	06	Alpes-Maritimes	4608
5	93	Seine-Saint-Denis	4121
6	74	Haute-Savoie	4046
7	78	Yvelines	4043
8	69	Rhône	3901
9	2A	Corse-du-Sud	3751
10	33	Gironde	3635

Requête 5

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

```
SELECT Reg_nom, ROUND(SUM(CAST(Valeur AS REAL))/SUM(Surface_local)) AS 'Prix du M^2  
Maison'  
FROM Vente  
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien  
JOIN Commune ON Bien.Id_decode_comcode = Commune.Id_decode_comcode  
JOIN Departement ON Commune.Dep_code = Departement.Dep_code  
JOIN Region ON Region.Id_reg = Departement.Id_reg  
WHERE Bien.Type_local = 'Maison'  
AND Reg_nom = 'Ile-de-France'  
AND Valeur IS NOT NULL  
AND Surface_local != 0
```

Reg_nom	Prix du M^2 Maison
Ile-de-France	3904

Requête 6

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

```
SELECT Id_bien,CAST(Valeur AS REAL) AS Valeur_num, Surface_local, Reg_nom  
FROM Vente  
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien  
JOIN Commune ON Bien.Id_decode_comcode = Commune.Id_decode_comcode  
JOIN Departement ON Commune.Dep_code = Departement.Dep_code  
JOIN Region ON Region.Id_reg = Departement.Id_reg  
WHERE Bien.Type_local = 'Appartement'  
AND Valeur != 0  
ORDER BY Valeur_num DESC  
LIMIT 10
```

Id bien	Valeur num	Surface local	Req nom
30603	9000000.0	10	Ile-de-France
5261	8600000	62	Ile-de-France
3625	8577713	289	Ile-de-France
7602	7620000	42	Ile-de-France
9988	7600000	200	Ile-de-France
17823	7535000	143	Ile-de-France
410	7420000	357	Ile-de-France
16357	7200000	241	Ile-de-France
1924	7050000	310	Ile-de-France
19161	6600000	76	Ile-de-France

Requête 7

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

```
SELECT  
    ROUND((((Nbr_ventes_2s2020 * 1.0)- Nbr_ventes_1s2020) / Nbr_ventes_1s2020,3) AS  
    Taux_evolution_ventes  
FROM (  
    SELECT  
        (SELECT COUNT(*)  
        FROM Vente  
        WHERE Date BETWEEN '2020/01/01' AND '2020/03/31') AS Nbr_ventes_1s2020,  
        (SELECT COUNT(*)  
        FROM Vente  
        WHERE Date BETWEEN '2020/04/01' AND '2020/06/30') AS Nbr_ventes_2s2020  
    );
```

Taux_evolution_ventes

0.037

Requête 8

Le classement des régions par rapport au prix au mètre carré des appartements de plus de 4 pièces

```
SELECT Reg_nom, ROUND(SUM(Valeur)/SUM(Surface_local)) AS 'Prix du M^2 Appartement 4+'  
FROM Vente  
JOIN Bien ON Vente.Id_Bien = Bien.Id_bien  
JOIN Commune ON Bien.Id_decode_comcode = Commune.Id_decode_comcode  
JOIN Departement ON Commune.Dep_code = Departement.Dep_code  
JOIN Region ON Region.Id_reg = Departement.Id_reg  
WHERE Bien.Type_local = 'Appartement'  
AND Total_piece >4  
AND Valeur IS NOT NULL  
AND Surface_local != 0  
GROUP BY Reg_nom  
ORDER BY ROUND(SUM(Valeur)/SUM(Surface_local)) DESC
```

	Req nom	Prix du M^2 Appartement 4+
1	Ile-de-France	8997
2	La Réunion	3485
3	Corse	3369
4	Provence-Alpes-Côte d'Azur	3072
5	Auvergne-Rhône-Alpes	2863
6	Nouvelle-Aquitaine	2607
7	Pays de la Loire	2400
8	Bretagne	2305
9	Hauts-de-France	2297
10	Occitanie	2214
11	Normandie	2108
12	Centre-Val de Loire	1452
13	Grand Est	1410
14	Bourgogne-Franche-Comté	1064
15	Martinique	564

Requête 9

Liste des communes ayant eu au moins 50 ventes au 1^{er} trimestre

```
SELECT Commune.Nom_commune, COUNT(Vente.Id_vente) AS Nombre_ventes
FROM Vente
JOIN Bien ON Vente.Id_bien = Bien.Id_bien
JOIN Commune ON Bien.Id_depcode_comcode = Commune.Id_depcode_comcode
WHERE Vente.Date BETWEEN '2020/01/01' AND '2020/03/31'
GROUP BY Commune.Nom_commune
HAVING COUNT(Vente.Id_vente) >= 50
ORDER BY Nombre_ventes DESC;
```

Requête 9

résultat

	Nom commune	Nombre ventes		Nom commune	Nombre ventes
1	Paris 17e Arrondissement	228	26	Antibes	77
2	Paris 15e Arrondissement	215	27	Marseille 4e Arrondissement	72
3	Paris 18e Arrondissement	209	28	Marseille 1er Arrondissement	71
4	Nice	173	29	Vincennes	68
5	Paris 11e Arrondissement	169	30	Rueil-Malmaison	68
6	Paris 16e Arrondissement	165	31	Lille	67
7	Bordeaux	157	32	Marseille 9e Arrondissement	66
8	Paris 14e Arrondissement	146	33	Montreuil	65
9	Paris 20e Arrondissement	127	34	Angers	64
10	Nantes	119	35	Nîmes	63
11	Paris 19e Arrondissement	116	36	Sète	62
12	Paris 12e Arrondissement	110	37	Paris 8e Arrondissement	62
13	Paris 10e Arrondissement	109	38	La Ciotat	62
14	Paris 9e Arrondissement	106	39	Rennes	61
15	Grenoble	106	40	Paris 2e Arrondissement	61
16	Boulogne-Billancourt	99	41	Paris 4e Arrondissement	60
17	Paris 13e Arrondissement	94	42	Toulon	59
18	Paris 7e Arrondissement	87	43	Levallois-Perret	59
19	Paris 6e Arrondissement	86	44	Saint-Maur-des-Fossés	56
20	Marseille 8e Arrondissement	81	45	Versailles	54
21	Asnières-sur-Seine	81	46	Ajaccio	54
22	Courbevoie	80	47	Puteaux	53
23	Paris 5e Arrondissement	79	48	Issy-les-Moulineaux	50
24	Paris 3e Arrondissement	79			
25	Toulouse	78			

Requête 10

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

```
SELECT
    ROUND((((Prix_M2_3pieces - Prix_M2_2pieces) * 100.0 / Prix_M2_2pieces),2) AS
    'Différence prix du M^2 entre un 2 pièces et un 3 pièces'
FROM (
    SELECT
        (SELECT SUM(Valeur)/ SUM(Surface_local)
         FROM Vente
         JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
         WHERE Bien.Type_local = 'Appartement'
         AND Valeur IS NOT NULL
         AND Total_piece = 2) AS Prix_M2_2pieces,
        (SELECT SUM(Valeur) / SUM(Surface_local)
         FROM Vente
         JOIN Bien ON Vente.Id_Bien = Bien.Id_bien
         WHERE Bien.Type_local = 'Appartement'
         AND Valeur IS NOT NULL
         AND Total_piece = 3) AS Prix_M2_3pieces
);
```

Différence prix du M^2 entre un 2 pièces et un 3 pièces

-11.23

Requête 11

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

```
WITH Moyenne_Par_Commune AS (  
    SELECT  
        Commune.Nom_commune,  
        Commune.Dep_code,  
        AVG(Vente.Valeur) AS Moyenne_valeurs  
    FROM Vente  
    JOIN Bien ON Vente.Id_Bien = Bien.Id_bien  
    JOIN Commune ON Bien.Id_depcode_comcode = Commune.Id_depcode_comcode  
    WHERE Commune.Dep_code IN ('06', 13, 33, 59, 69)  
    AND Vente.Valeur IS NOT NULL  
    GROUP BY Commune.Nom_commune, Commune.Dep_code  
),  
Top3_Communes AS (  
    SELECT Nom_commune, Dep_code, Moyenne_valeurs,  
        ROW_NUMBER() OVER (PARTITION BY Dep_code ORDER BY Moyenne_valeurs  
        DESC) AS Rang  
    FROM Moyenne_Par_Commune  
)  
SELECT Dep_code, Nom_commune, ROUND(Moyenne_valeurs) AS Moyenne_valeur_fonciere  
FROM Top3_Communes  
WHERE Rang <= 3  
ORDER BY Dep_code, Moyenne_valeurs DESC
```

Requête 11

résultat

	Dep_code	Nom_commune	Moyenne_valeur_fonciere
1	06	Saint-Jean-Cap-Ferrat	968750
2	06	Eze	655000
3	06	Mouans-Sartoux	476898
4	13	Gignac-la-Nerthe	330000
5	13	Saint-Savournin	314425
6	13	Cassis	313417
7	33	Lège-Cap-Ferret	549501
8	33	Vayres	335000
9	33	Arcachon	307436
10	59	Bersée	433202
11	59	Cysoing	408550
12	59	Halluin	322250
13	69	Ville-sur-Jarnioux	485300
14	69	Lyon 2e Arrondissement	455217
15	69	Lyon 6e Arrondissement	426968

Requête 12

Les 20 communes avec le plus de transactions pour 1000 habitants pour les communes qui dépassent les 10000 habitants

```
WITH Transactions_Par_Commune AS (  
    SELECT  
        Commune.Nom_commune,  
        Commune.Population,  
        COUNT(Vente.Id_vente) AS Nombre_transactions,  
        ROUND((COUNT(Vente.Id_vente) * 1000.0 / Commune.Population),2) AS  
        Transactions_par_1000_habitants  
    FROM Vente  
    JOIN Bien ON Vente.Id_bien = Bien.Id_bien  
    JOIN Commune ON Bien.Id_decode_comcode = Commune.Id_decode_comcode  
    WHERE Commune.Population > 10000  
    GROUP BY Commune.Nom_commune, Commune.Population  
)  
SELECT Nom_commune, Population, Nombre_transactions, Transactions_par_1000_habitants  
FROM Transactions_Par_Commune  
ORDER BY Transactions_par_1000_habitants DESC  
LIMIT 20;
```

Requête 12

résultat

	Nom_commune	Population	Nombre_transactions	Transactions_par_1000_habitants
1	Paris 2e Arrondissement	21735	127	5.84
2	Paris 1er Arrondissement	16055	79	4.92
3	Paris 3e Arrondissement	34306	161	4.69
4	Arcachon	11898	55	4.62
5	La Baule-Escoublac	16797	77	4.58
6	Paris 4e Arrondissement	29390	120	4.08
7	Roquebrune-Cap-Martin	13041	52	3.99
8	Paris 8e Arrondissement	36250	139	3.83
9	Sanary-sur-Mer	17160	60	3.5
10	La Londe-les-Maures	10776	37	3.43
11	Paris 9e Arrondissement	60563	208	3.43
12	Paris 6e Arrondissement	41171	139	3.38
13	Saint-Cyr-sur-Mer	11725	38	3.24
14	Chantilly	11178	35	3.13
15	Pornichet	11440	35	3.06
16	Saint-Mandé	22576	69	3.06
17	Paris 10e Arrondissement	86863	264	3.04
18	Menton	30981	91	2.94
19	Saint-Hilaire-de-Riez	11501	33	2.87
20	Vincennes	50230	141	2.81



Merci !