



Création et utilisation de la base de données

LEONARD Damien



Laplace Immo

Contexte du projet

- Projet “DATAIMMO”
- Modification base de données
- Analyse du marché
- Aider les différentes agences.



La stratégie de sauvegarde et la conformité RGPD

- Obligatoire pour protéger les données personnelles
- Réduction des Risques
- Continuité des Activités

Les données initiales

- Open Data : Demandes de valeurs foncières
- INSEE : les résultats des recensements de la population
- Data.gouv : Données des régions

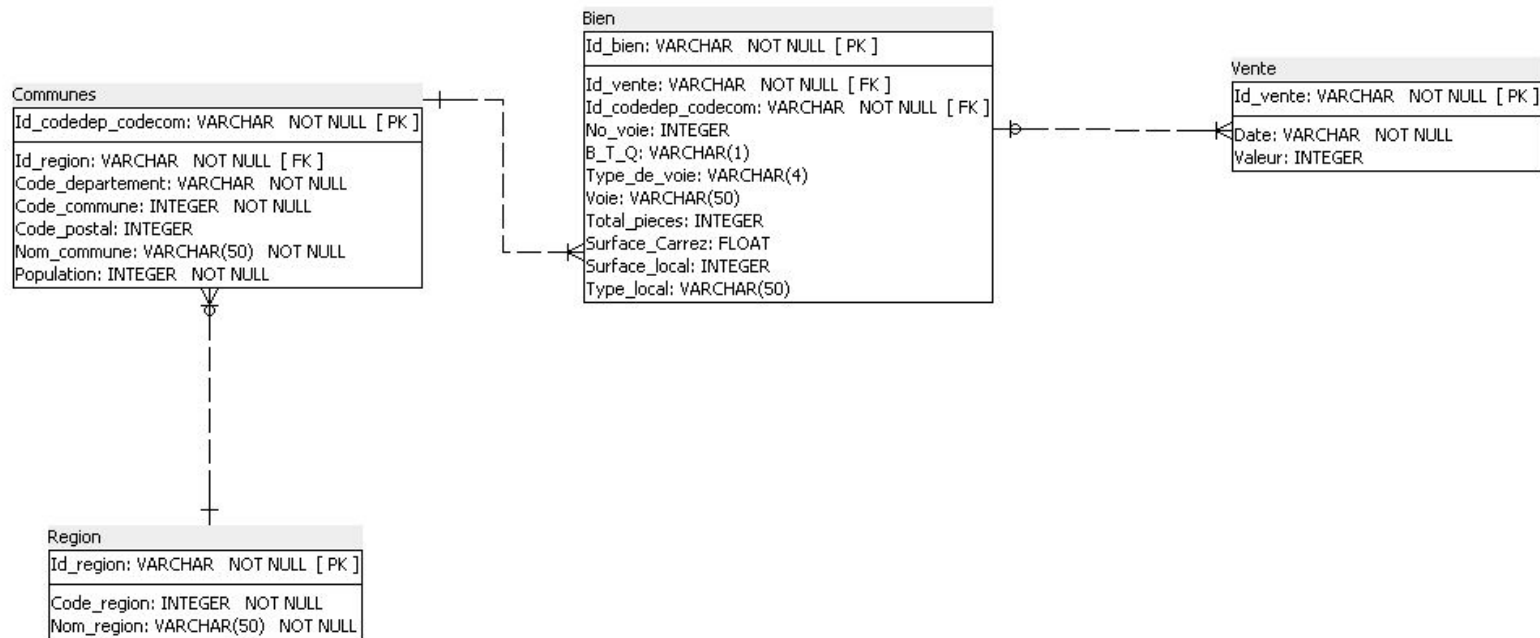


data.gouv.fr

L'extrait du dictionnaire des données

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
Id_bien	ID dans la base de données, clef primaire	Varchar	50	Elémentaire	Ne doit pas être nul	
Id_codedep_codecc	Numero concater departement et commune(clef etrangere)	Varchar	10	concaténer	Ne doit pas être nul	Concatener (Code departement , Code commune)
Id_vente	Identifiant vente	Varchar	50	Elémentaire	Ne doit pas être nul	
No_voie	Numéro des rues	Integer	NC	Elémentaire		
B/T/Q	Indicateur éventuel de répétition pour l'adresse du logement a	Varchar	1	Elémentaire		
Type_de_voie	Plusieurs valeurs (rue, avenue, chemin, etc.)	Varchar	4	Elémentaire		
Voie	Nom de la rue	Varchar	50	Elémentaire		
Total pièces	Nombres de pieces a vivre	Integer	2	Elémentaire	Ne doit pas être nul	
Surface Carrez	Surface m² du 1er lot CARREZ	Float	NC	Elémentaire		
Surface local	Nombre m² local	Integer	NC	Elémentaire		
Type local	Indication maison ou appartement	Varchar	12	Elémentaire		
Id_codedep_codecc	Numero concater departement et commune(clef etrangere)	Varchar	10	concaténer	Ne doit pas être nul	Concatener (Code departement , Code commune)
Id_bien	ID dans la base de données, clef primaire	Integer	50	Elémentaire	Ne doit pas être nul	
Code departement	Numéro du departement	Integer	3	Elémentaire	Ne doit pas être nul	
Code commune	Numero de commune	Integer	3	Elémentaire	Ne doit pas être nul	
Code postal	Code departement + code commune	Integer	6	concatener	Ne doit pas être nul	
Nom Commune	Nom de la ville	Varchar	50	Elémentaire	Ne doit pas être nul	
Population	Nombre d'habitant	Integer	50	Elémentaire	Ne doit pas être nul	
Id_vente	Identifiant vente	Varchar	50	Elémentaire	Ne doit pas être nul	
Date	Date de changement propriétaire	Date	10	Elémentaire	Ne doit pas être nul	
Valeur	Valeur propriété	Integer	50	Elémentaire	Ne doit pas être nul	
Id_region	Numero identification region	Varchar	4	Elémentaire	Ne doit pas être nul	
reg_code	Numero region	Integer	3	Elémentaire	Ne doit pas être nul	
reg_nom	Nom de la region	Varchar	50	Elémentaire	Ne doit pas être nul	

Le schéma relationnel normalisé



screenshot qui permet de démontrer le bon chargement des données

lom	Type	Schéma
Tables (4)		
Bien		CREATE TABLE Bien (Id_bien VARCHAR NOT NULL, Id_vente VARCHAR NOT NULL, Id_codedep_codecom VARCHAR NOT NULL, No_voie INTEGER, B_T_Q VARCHAR(1), Type_de_voie VA
Id_bien	VARCHAR	"Id_bien" VARCHAR NOT NULL
Id_vente	VARCHAR	"Id_vente" VARCHAR NOT NULL
Id_codedep_codecom	VARCHAR	"Id_codedep_codecom" VARCHAR NOT NULL
No_voie	INTEGER	"No_voie" INTEGER
B_T_Q	VARCHAR(1)	"B_T_Q" VARCHAR(1)
Type_de_voie	VARCHAR(4)	"Type_de_voie" VARCHAR(4)
Voie	VARCHAR(50)	"Voie" VARCHAR(50)
Total_pieces	INTEGER	"Total_pieces" INTEGER
Surface_Carrez	FLOAT	"Surface_Carrez" FLOAT
Surface_local	INTEGER	"Surface_local" INTEGER
Type_local	VARCHAR(50)	"Type_local" VARCHAR(50)
Communes		CREATE TABLE Communes (Id_codedep_codecom VARCHAR NOT NULL, Id_region VARCHAR NOT NULL, Code_departement VARCHAR NOT NULL, Code_commune INTEGER NOT NULL, C
Id_codedep_codecom	VARCHAR	"Id_codedep_codecom" VARCHAR NOT NULL
Id_region	VARCHAR	"Id_region" VARCHAR NOT NULL
Code_departement	VARCHAR	"Code_departement" VARCHAR NOT NULL
Code_commune	INTEGER	"Code_commune" INTEGER NOT NULL
Code_postal	INTEGER	"Code_postal" INTEGER
Nom_commune	VARCHAR(50)	"Nom_commune" VARCHAR(50) NOT NULL
Population	INTEGER	"Population" INTEGER NOT NULL
Region		CREATE TABLE Region (Id_region VARCHAR NOT NULL, Code_region INTEGER NOT NULL, Nom_region VARCHAR(50) NOT NULL, CONSTRAINT Region_pk PRIMARY KEY (Id_region))
Id_region	VARCHAR	"Id_region" VARCHAR NOT NULL
Code_region	INTEGER	"Code_region" INTEGER NOT NULL
Nom_region	VARCHAR(50)	"Nom_region" VARCHAR(50) NOT NULL
Vente		CREATE TABLE Vente (Id_vente VARCHAR NOT NULL, Date VARCHAR NOT NULL, Valeur INTEGER, CONSTRAINT Vente_pk PRIMARY KEY (Id_vente))
Id_vente	VARCHAR	"Id_vente" VARCHAR NOT NULL
Date	VARCHAR	"Date" VARCHAR NOT NULL
Valeur	INTEGER	"Valeur" INTEGER

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

	Id_bien	Id_vente	Id_codedep_codecom	No_voie	B_T_Q	Type_de_voie	Voie	Total_pieces	Surface_Carrez	Surface_local	Type_local ▼ ¹
	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre
1	866	V66	78640	2	NULL	RUE	DIV...	2	49.16	48	Appartement
2	8465	V465	9212	2	NULL	RUE	DE ...	2	51.34	50	Appartement
3	8661	V661	1354	2	NULL	RUE	DU ...	2	58.71	60	Appartement
4	8752	V752	74264	2	NULL	RUE	DE ...	2	45.83	45	Appartement
5	81026	V1026	77186	2	NULL	RUE	DE...	2	33.15	34	Appartement
6	81530	V1530	1479	2	NULL	RUE	CH...	2	48.35	47	Appartement
7	81582	V1582	13201	2	NULL	RUE	JEA...	2	47.9	44	Appartement
8	81583	V1583	13201	2	NULL	RUE	DE ...	2	88.3	70	Appartement
9	81694	V1694	75115	2	NULL	RUE	MIZ...	2	35.5	33	Appartement
10	81727	V1727	8021	2	NULL	RUE	D ...	2	39.24	44	Appartement

	Id_vente	Date ▲ ¹	Valeur
	Filtre	Filtre	Filtre
1	V33681	2020-06-30	29500
2	V33682	2020-06-30	36000
3	V33685	2020-06-30	64600
4	V33686	2020-06-30	425000
5	V33687	2020-06-30	970000
6	V33688	2020-06-30	328300
7	V33689	2020-06-30	445100
8	V33690	2020-06-30	54000
9	V33691	2020-06-30	750000
10	V33692	2020-06-30	34000

	Id_codedep_codecom	Id_region	Code_departement	Code_commune	Code_postal	Nom_commune	Population
	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre	Filtre
1	771	R11	77		1	77760 ACHERES-LA-FORET	1171
2	774	R11	77		4	77390 ANDREZEL	291
3	775	R11	77		5	77410 ANNET-SUR-MARNE	3326
4	7714	R11	77		14	77210 AVON	13643
5	7718	R11	77		18	77700 BAILLY-ROMAINVILLIERS	7477
6	7722	R11	77		22	77630 BARBIZON	1246
7	7724	R11	77		24	77750 BASSEVELLE	384
8	7737	R11	77		37	77590 BOIS LE ROI	6061
9	7744	R11	77		44	77720 BOMBON	968
10	7753	R11	77		53	77170 BRIE COMTE ROBERT	19129

	Id_region	Code_region	Nom_region
	Filtre	Filtre	Filtre
1	R84	84	Auvergne-Rhône-Alpes
2	R32	32	Hauts-de-France
3	R93	93	Provence-Alpes-Côte d'Azur
4	R44	44	Grand Est
5	R76	76	Occitanie
6	R28	28	Normandie
7	R75	75	Nouvelle-Aquitaine
8	R24	24	Centre-Val de Loire
9	R27	27	Bourgogne-Franche-Comté
10	R53	53	Bretagne



Requêtes SQL et résultats

Requête 1 a 12

Requête 1

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

```
1 SELECT COUNT(*) as Nombres_Appartements
2 FROM Bien
3 WHERE Type_local = 'Appartement'
4
```

	Nombres_Appartements
1	31378

Requête 2

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

```
1 SELECT Region.Nom_region, COUNT(*) AS Nombre_vente
2 FROM Bien
3 JOIN Communes ON Communes.Id_codedep_codecom = Bien.Id_codedep_codecom
4 JOIN Region ON Region.Id_region = Communes.Id_region
5 WHERE Type_local = 'Appartement'
6 GROUP BY Nom_region
7
```

	Nom_region	Nombre_vente
1	Auvergne-Rhône-Alpes	3284
2	Bourgogne-Franche-Comté	376
3	Bretagne	982
4	Centre-Val de Loire	696
5	Corse	223
6	Grand Est	975
7	Guadeloupe	2
8	Guyane	34
9	Hauts-de-France	1254
10	Île-de-France	13995
11	La Réunion	44
12	Martinique	94
13	Normandie	859
14	Nouvelle-Aquitaine	1932
15	Occitanie	1639
16	Pays de la Loire	1356
17	Provence-Alpes-Côte d'Azur	3633

Requête 3

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

```
1 SELECT Total_pieces, count(*) AS Nombre_vente,  
2 round(count(*) *1.0 / (SELECT count(*) FROM Bien WHERE Type_local = 'Appartement') *100 , 2)  
3 AS Proportion  
4 FROM Bien  
5 WHERE Type_local = 'Appartement'  
6 GROUP BY Total_pieces
```

	Total_pieces	Nombre_vente	Proportion
1	0	30	0.1
2	1	6739	21.48
3	2	9783	31.18
4	3	8966	28.57
5	4	4460	14.21
6	5	1114	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	0.0

Requête 4

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

```
1 SELECT Nom_departement, round(AVG(Valeur / Surface_Carrez),2) AS Prix_metre_carré
2 FROM Bien
3 JOIN Communes ON Communes.Id_codedep_codecom = Bien.Id_codedep_codecom
4 JOIN Vente ON Vente.Id_vente = Bien.Id_vente
5 GROUP BY Nom_departement
6 ORDER BY round(AVG(Valeur / Surface_Carrez),2) DESC
7 LIMIT 10
```

	Nom_departement	Prix_metre_carré
1	Paris	12052.82
2	Hauts-de-Seine	7216.13
3	Val-de-Marne	5343.28
4	Alpes-Maritimes	4717.6
5	Haute-Savoie	4667.13
6	Seine-Saint-Denis	4344.78
7	Yvelines	4225.25
8	Rhône	4045.19
9	Corse-du-Sud	4026.97
10	Gironde	3764.14

Requête 5

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

```
1 SELECT ROUND(AVG(Valeur / Surface_Carrez),2) AS Ratio_IDF
2 FROM Bien
3 JOIN Vente ON Vente.Id_vente = Bien.Id_vente
4 JOIN Communes ON Communes.Id_codedep_codecom = Bien.Id_codedep_codecom
5 JOIN Region ON Region.Id_region = Communes.Id_region
6 WHERE Region.Code_region = '11' AND Type_local = 'Maison'
```

	Ratio_IDF
1	3742.78

Requête 6

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

```
1 SELECT Bien.Id_bien , Nom_region , Nom_commune , Valeur , Surface_Carrez
2 FROM Bien
3 JOIN Vente ON Vente.Id_vente = Bien.Id_vente
4 JOIN Communes ON Communes.Id_codedep_codecom = Bien.Id_codedep_codecom
5 JOIN Region ON Region.Id_region = Communes.Id_region
6 ORDER BY Valeur DESC
7 LIMIT 10
```

	Id_bien	Nom_region	Nom_commune	Valeur	Surface_Carrez
1	B30602	Ile-de-France	PARIS 16	9000000	9.1
2	B5260	Ile-de-France	CORBEIL ESSONNES	8600000	64.0
3	B3624	Ile-de-France	PARIS 07	8577713	20.55
4	B7601	Ile-de-France	PARIS 17	7620000	42.77
5	B9987	Ile-de-France	PARIS 06	7600000	253.3
6	B17822	Ile-de-France	PARIS 01	7535000	139.9
7	B409	Ile-de-France	PARIS 16	7420000	360.95
8	B16356	Ile-de-France	PARIS 16	7200000	595.0
9	B1923	Ile-de-France	PARIS 01	7050000	122.56
10	B19160	Ile-de-France	PARIS 01	6600000	79.38

Requête 7

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

```
1 WITH Ventes_Trimestres AS (  
2     SELECT  
3         SUM(CASE WHEN Date BETWEEN '2020-01-01' AND '2020-03-31' THEN 1 ELSE 0 END) AS Ventes_T1_2020,  
4         SUM(CASE WHEN Date BETWEEN '2020-04-01' AND '2020-06-30' THEN 1 ELSE 0 END) AS Ventes_T2_2020  
5     FROM Vente  
6 )  
7 SELECT  
8     Ventes_T1_2020,  
9     Ventes_T2_2020,  
10    round((Ventes_T2_2020 - Ventes_T1_2020) * 100 * 1.0 / Ventes_T1_2020 ,2) AS Taux_evolution  
11 FROM Ventes_Trimestres;
```

	Ventes_T1_2020	Ventes_T2_2020	Taux_evolution
1	16776	17393	3.68

Requête 8

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

```
1 SELECT Nom_region , round(avg(Valeur/Surface_Carrez),2) AS Prix_metre_carre
2 FROM Bien
3 JOIN Vente ON Vente.Id_vente = Bien.Id_vente
4 JOIN Communes ON Communes.Id_codedep_codecom = Bien.Id_codedep_codecom
5 JOIN Region ON Region.Id_region = Communes.Id_region
6 WHERE Type_local = 'Appartement'
7 AND Total_pieces > 4
8 GROUP BY Nom_region
9 ORDER BY Prix_metre_carre DESC
~
```

	Nom_region	Prix_metre_carre
1	Ile-de-France	8770.44
2	La Réunion	3641.81
3	Provence-Alpes-Côte d'Azur	3587.65
4	Corse	3104.88
5	Auvergne-Rhône-Alpes	2891.38
6	Nouvelle-Aquitaine	2465.48
7	Bretagne	2412.05
8	Pays de la Loire	2315.76
9	Hauts-de-France	2189.93
10	Occitanie	2097.23
11	Normandie	2015.77
12	Grand Est	1540.89
13	Centre-Val de Loire	1453.11
14	Bourgogne-Franche-Comté	1251.19
15	Martinique	573.48

Requête 9

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

```
1 SELECT Communes.Nom_commune,  
2        COUNT(Vente.Id_vente) AS Nombre_vente  
3 FROM Communes  
4 JOIN Bien ON Bien.Id_codedep_codecom = Communes.Id_codedep_codecom  
5 JOIN Vente ON Vente.Id_vente = Bien.Id_vente  
6 WHERE Vente.Date BETWEEN '2020-01-01' AND '2020-03-31'  
7 GROUP BY Communes.Nom_commune  
8 HAVING COUNT(Vente.Id_vente) > 50;  
9
```

	Nom_commune	Nombre_vente
1	AJACCIO	54
2	ANGERS	64
3	ANTIBES	77
4	ASNIERES-SUR-SEINE	81
5	BORDEAUX	157
6	BOULOGNE-BILLANCOURT	99
7	COURBEVOIE	80
8	GRENOBLE	106
9	LA CIOTAT	62
10	LEVALLOIS-PERRET	59
11	LILLE	67
12	MARSEILLE 1ER	71
13	MARSEILLE 4EME	72
14	MARSEILLE 8EME	81
15	MARSEILLE 9EME	66
16	MONTREUIL	65
17	NANTES	119
18	NICE	173
19	NIMES	63
20	PARIS 02	61
21	PARIS 03	79
22	PARIS 04	60
23	PARIS 05	79

	Nom_commune	Nombre_vente
23	PARIS 05	79
24	PARIS 06	86
25	PARIS 07	87
26	PARIS 08	62
27	PARIS 09	106
28	PARIS 10	109
29	PARIS 11	169
30	PARIS 12	110
31	PARIS 13	94
32	PARIS 14	146
33	PARIS 15	215
34	PARIS 16	165
35	PARIS 17	228
36	PARIS 18	209
37	PARIS 19	116
38	PARIS 20	127
39	PUTEAUX	53
40	RENNES	61
41	RUEIL-MALMAISON	68
42	SAINT-MAUR-DES-FOSSES	56
43	SETE	62
44	TOULON	59
45	TOULOUSE	78
46	VERSAILLES	54
47	VINCENNES	68

Requête 10

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

```
1 SELECT Total_pieces, avg(Valeur/Surface_Carrez) AS Prix_metre_carre,
2 ROUND(100.0 * ((AVG(Valeur / Surface_local)
3 -- LAG(AVG(Valeur / Surface_Carrez)) OVER (ORDER BY Total_pieces))
4 / LAG(AVG(Valeur / Surface_Carrez)) OVER (ORDER BY Total_pieces)), 2)
5 AS Difference_pourcentage
6 FROM Bien
7 JOIN Vente ON Vente.Id_vente = Bien.Id_vente
8 WHERE Total_pieces IN (2, 3)
9 GROUP BY Total_pieces
10 ORDER BY Total_pieces;
```

	Total_pieces	Prix_metre_carre	Difference_pourcentage
1	2	4870.05534927306	NULL
2	3	4243.71280748603	-12.91

Requête 11

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

```
1 SELECT Nom_commune, Code_departement, Moyenne_Valeur
2 FROM (
3     SELECT
4         Nom_commune, Code_departement, AVG(Valeur) AS Moyenne_Valeur,
5         ROW_NUMBER() OVER (PARTITION BY Code_departement ORDER BY AVG(Valeur) DESC) AS Rang
6     FROM Communes
7     JOIN Bien ON Bien.Id_codedep_codecom = Communes.Id_codedep_codecom
8     JOIN Vente ON Vente.Id_vente = Bien.Id_vente
9     WHERE Code_departement IN (6, 13, 33, 59, 69)
10    GROUP BY Nom_commune, Code_departement
11 ) AS Classement
12 WHERE Rang <= 3;
```

	Nom_commune	Code_departement	Moyenne_Valeur
1	GIGNAC-LA-NERTHE	13	330000.0
2	SAINT SAVOURNIN	13	314425.0
3	CASSIS	13	313416.875
4	LEGE-CAP-FERRET	33	549500.636363636
5	VAYRES	33	335000.0
6	ARCACHON	33	307435.927272727
7	BERSEE	59	433202.0
8	CYSOING	59	408550.0
9	HALLUIN	59	322250.0
10	SAINT-JEAN-CAP-FERRAT	6	968750.0
11	EZE	6	655000.0
12	MOUANS-SARTOUX	6	476898.1
13	VILLE SUR JARNIOUX	69	485300.0
14	LYON 2EME	69	455217.266666667
15	LYON 6EME	69	426968.25

Requête 12

Les 20 communes avec le plus de transactions pour 1000 habitants pour les communes qui dépassent les 10 000 habitants

```
1 WITH TransactionsParHabitant AS (  
2     SELECT Communes.Nom_commune, COUNT(Vente.Id_vente) AS Nombre_transactions, Communes.Population,  
3         round(COUNT(Vente.Id_vente) * 1000.0 / Communes.Population,2) AS Transactions_pour_1000_habitants  
4     FROM Communes  
5     JOIN Bien ON Bien.Id_codedep_codecom = Communes.Id_codedep_codecom  
6     JOIN Vente ON Vente.Id_vente = Bien.Id_vente  
7     WHERE Communes.Population > 10000  
8     GROUP BY Communes.Nom_commune, Communes.Population  
9 )  
10 SELECT Nom_commune, Transactions_pour_1000_habitants  
11 FROM TransactionsParHabitant  
12 ORDER BY Transactions_pour_1000_habitants DESC  
13 LIMIT 20;
```

	Nom_commune	Transactions_pour_1000_habitants
1	PARIS 02	5.84
2	PARIS 01	4.92
3	PARIS 03	4.69
4	ARCACHON	4.62
5	LA BAULE-ESCOUBLAC	4.58
6	PARIS 04	4.08
7	ROQUEBRUNE CAP MARTIN	3.99
8	PARIS 08	3.83
9	SANARY SUR MER	3.5
10	LA LONDE LES MAURES	3.43
11	PARIS 09	3.43
12	PARIS 06	3.38
13	SAINT-CYR-SUR-MER	3.24
14	CHANTILLY	3.13
15	PORNICHET	3.06
16	SAINT-MANDE	3.06
17	PARIS 10	3.04
18	MENTON	2.94
19	SAINT-HILAIRE-DE-RIEZ	2.87
20	VINCENNES	2.81



Merci !