



DATAImmo



CONTEXTE



DONNÉES



MODÉLISATION



RELATIONS ET CLÉS



**CRÉATION DE LA
BDD**



**PRÉPARATION DES
DONNÉES**



**IMPORT DES
DONNÉES**



CONCLUSION

CONTEXTE



Laplace Immo

- Réseau national d'agences immobilières

Projet DATAImmo

- Objectif : création d'un modèle pour mieux prévoir le prix de vente des biens immobiliers
- Proof of Concept :
 - collecte des transactions immobilières de France sur le premier semestre 2020
→ contexte de crise sanitaire
 - analyse du marché : par localisation, en fonction des caractéristiques du bien

DONNÉES



Tableur issu du site open data *Demandes valeurs foncières (DVF)*

- 34169 ventes réalisées en France en le 1^{er} semestre 2020
- 45 colonnes (dont certaines colonnes vides, redondantes ou non utiles pour les analyses)

Dictionnaire de données

- Définition du domaine fonctionnel
 - ➔ épuration et spécification des données
 - ➔ 11 variables retenues

Numéro	Code propriété	Signification	Type	Observation
1	ValeurFonciere	Valeur foncière du bien	Numérique ou alphanumérique	Decimal(15,2)
2	DateTransaction	Date de la transaction	Date	
3	SurfaceCarrez	Surface loi Carrez	Numérique ou alphanumérique	Decimal(6,2)
4	TypeLocal	Type de local	Texte	Varchar(20)
5	NbrPieces	Nombre de pièces	Numérique ou alphanumérique	Small Integer
6	CodeDepartement	Code du département	Texte	Varchar(5)
7	CodeCommune	Code de la commune	Texte	Varchar(5)
8	Commune	Nom de la commune	Texte	Varchar(50)
9	CodePostal	Code postal	Texte	Varchar(5)
10	TypeVoie	Type de voie	Texte	Varchar(10)
11	NomVoie	Nom de la voie	Texte	Varchar(50)

MODÉLISATION

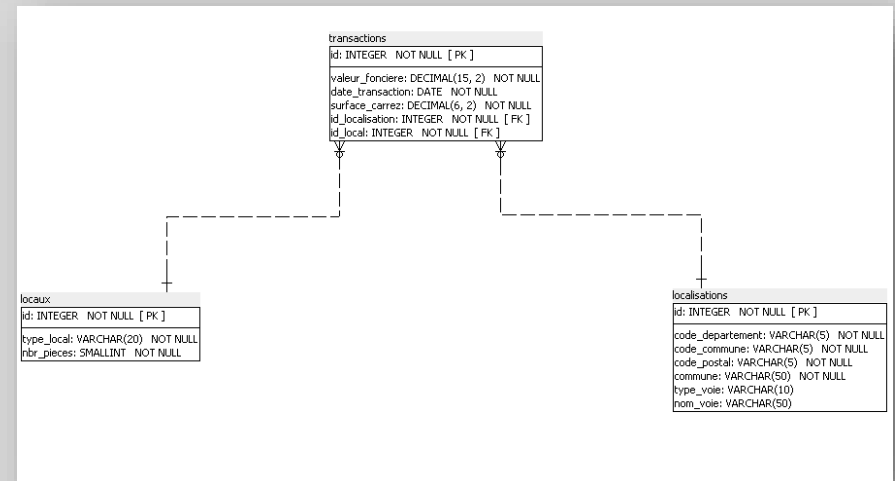
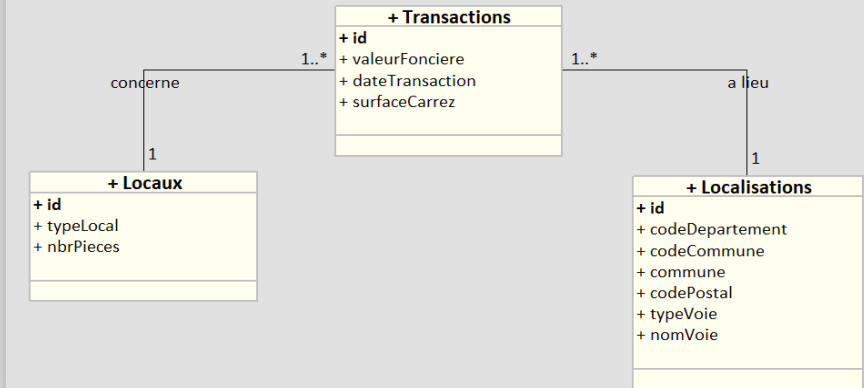
Modèle conceptuel de données

- Modélisation graphique UML
- Utilisation du logiciel 

Schéma relationnel normalisé

- Utilisation du logiciel  SQL POWER ARCHITECT

Modèle conceptuel de données immobilières

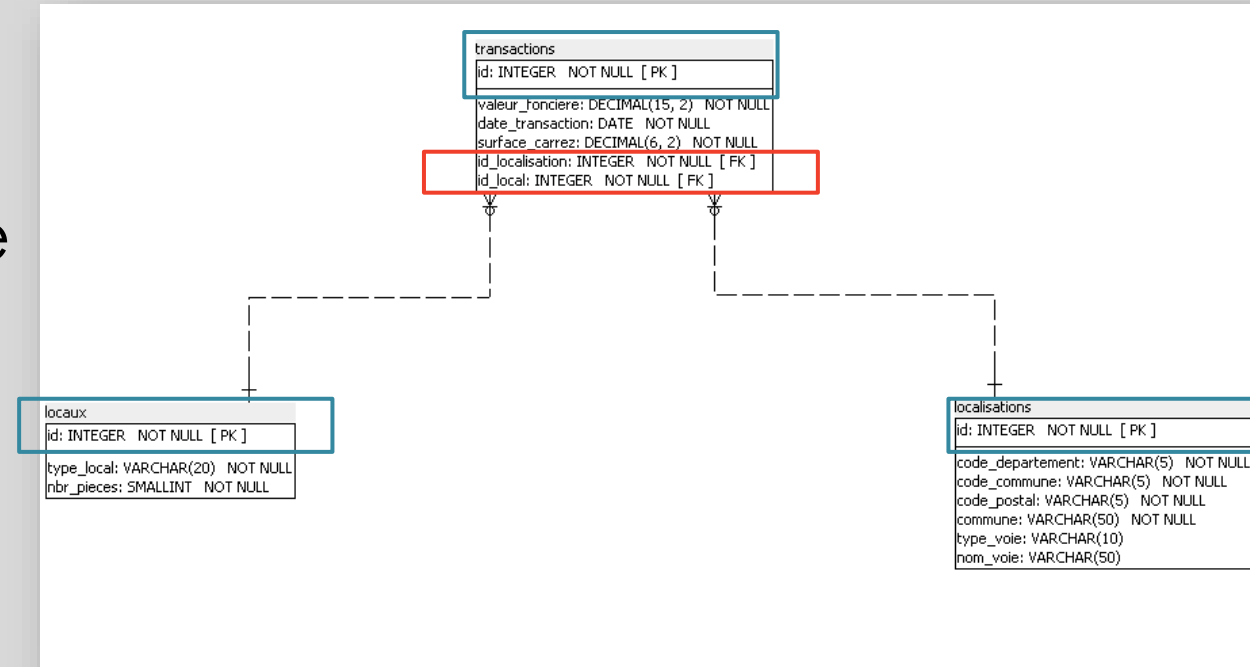


RELATIONS ET CLÉS

3 clés primaires PK : id, créés spécifiquement pour la BDD

2 relations non identifiantes de type un à plusieurs

2 clés étrangères FK :
id_localisation, id_local,
permettant de faire le lien
entre les tables



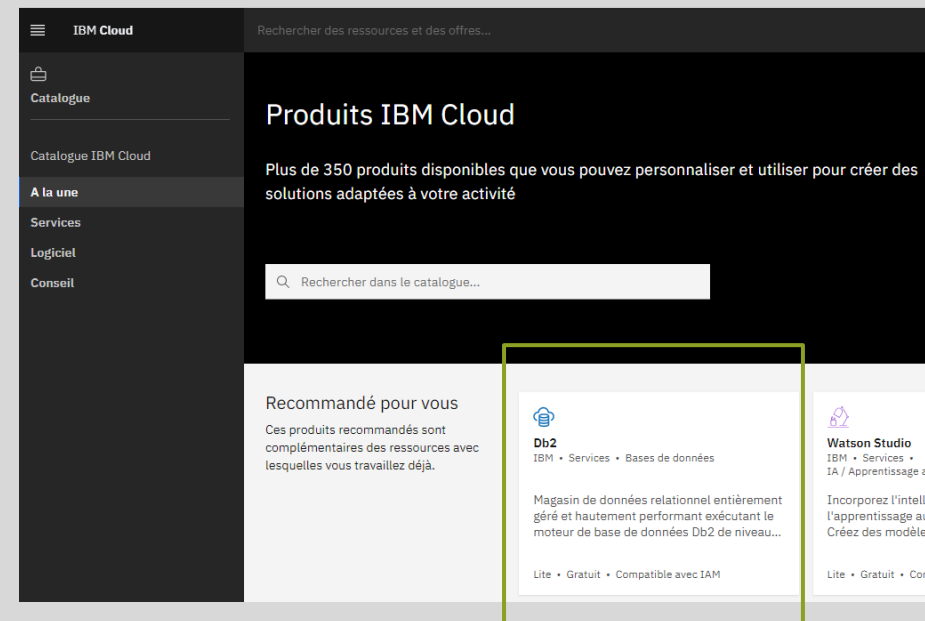
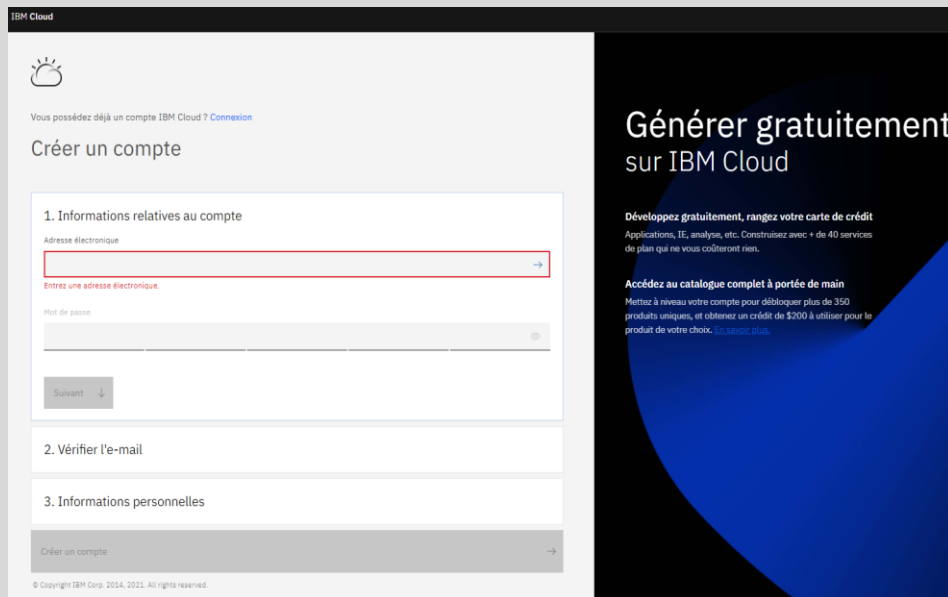
CRÉATION DE LA BDD



SGBD: IBM DB2 On Cloud, version Lite (*Création de la BDD également sur PostGres SQL*)

1- Créer un compte, aucune installation en local requise

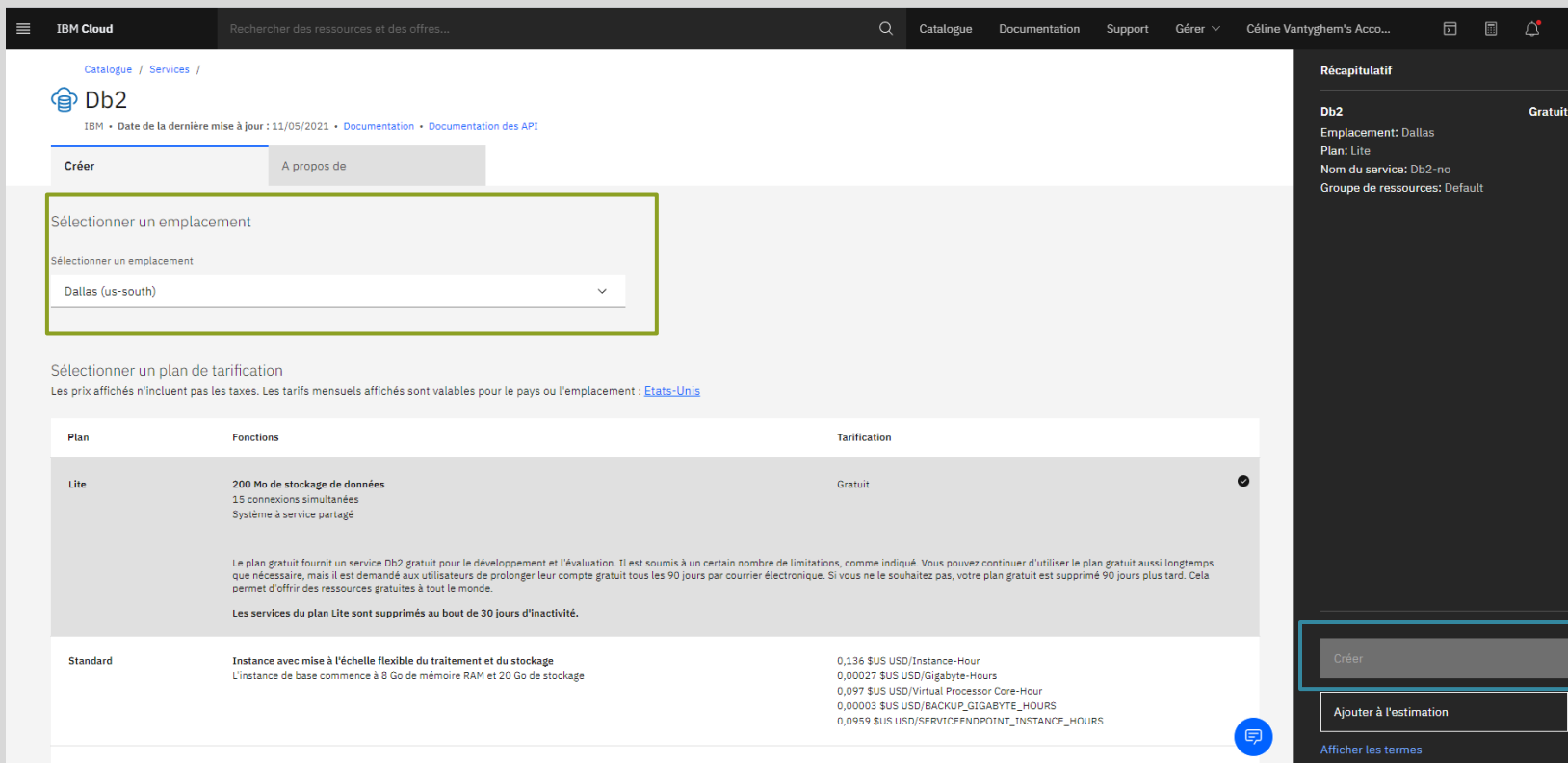
2- Dans le catalogue de services proposés, sélectionner **IBM DB2 Lite**



CRÉATION DE LA BDD



3- Sélectionner la localisation **Dallas**, créer la ressource



IBM Cloud

Rechercher des ressources et des offres...

Catalogue Documentation Support Gérer Céline Vantighem's Acco...

Catalogue / Services /

Db2

IBM • Date de la dernière mise à jour : 11/05/2021 • Documentation • Documentation des API

Créer A propos de

Sélectionner un emplacement

Sélectionner un emplacement

Dallas (us-south)

Sélectionner un plan de tarification

Les prix affichés n'incluent pas les taxes. Les tarifs mensuels affichés sont valables pour le pays ou l'emplacement : [Etats-Unis](#)

Plan	Fonctions	Tarification
Lite	200 Mo de stockage de données 15 connexions simultanées Système à service partagé	Gratuit
Standard	Instance avec mise à l'échelle flexible du traitement et du stockage L'instance de base commence à 8 Go de mémoire RAM et 20 Go de stockage	0,136 \$US USD/Instance-Hour 0,00027 \$US USD/Gigabyte-Hours 0,097 \$US USD/Virtual Processor Core-Hour 0,00003 \$US USD/BACKUP_GIGABYTE_HOURS 0,0959 \$US USD/SERVICEENDPOINT_INSTANCE_HOURS

Créer

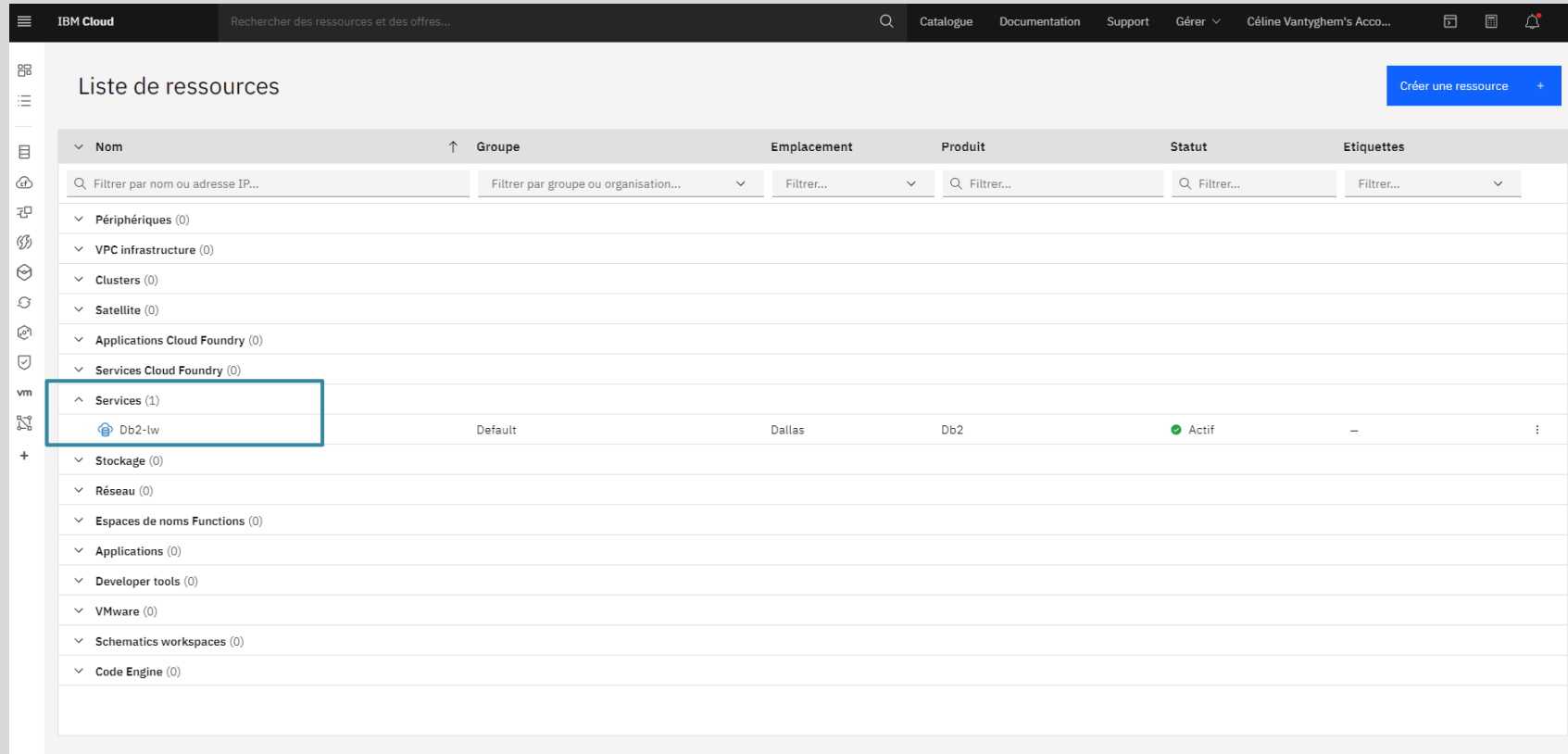
Ajouter à l'estimation

Afficher les termes

CRÉATION DE LA BDD



4- Pour accéder à la BDD, sélectionner la ressource créée (**Db2-lw**) dans la liste des ressources du compte



IBM Cloud

Rechercher des ressources et des offres...

Catalogue Documentation Support Gérer Céline Vantghem's Acco...

Liste de ressources

Créer une ressource +

Nom	Groupe	Emplacement	Produit	Statut	Etiquettes
Filtrer par nom ou adresse IP...					
Filtrer par groupe ou organisation...		Filtrer...	Filtrer...	Filtrer...	Filtrer...
Périphériques (0)					
VPC infrastructure (0)					
Clusters (0)					
Satellite (0)					
Applications Cloud Foundry (0)					
Services Cloud Foundry (0)					
Services (1)					
Db2-lw	Default	Dallas	Db2	Actif	—
Stockage (0)					
Réseau (0)					
Espaces de noms Functions (0)					
Applications (0)					
Developer tools (0)					
VMware (0)					
Schematics workspaces (0)					
Code Engine (0)					

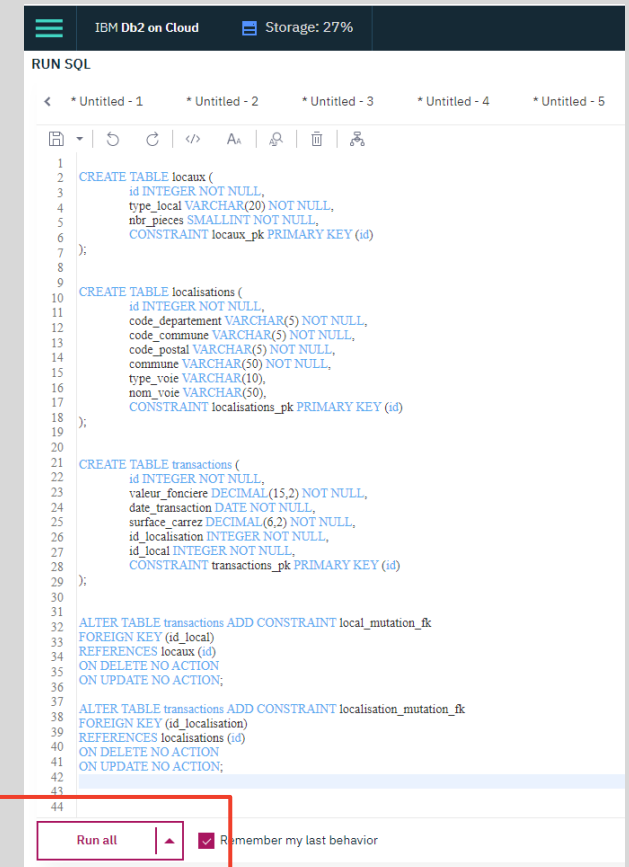
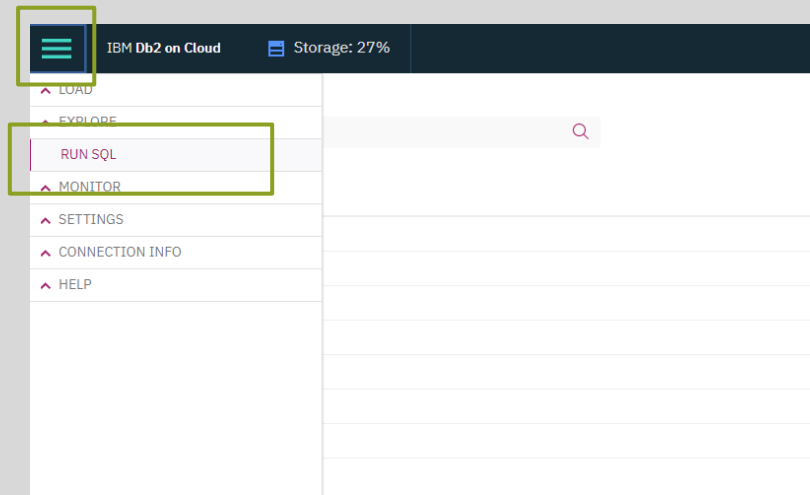
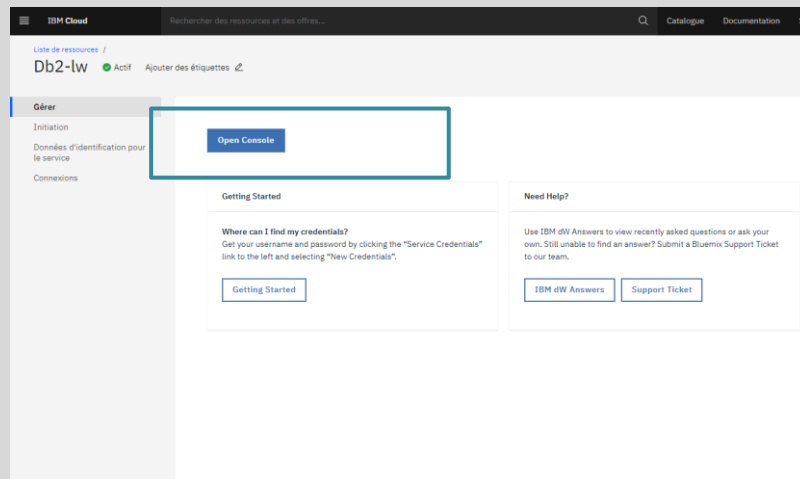
CRÉATION DE LA BDD



5- Pour accéder à la console, cliquer sur [Open Console](#)

6- Cliquer sur [Run SQL](#)

7- Créer les tables en [exécutant le script SQL](#)



CRÉATION DE LA BDD



Table Transactions

IBM Db2 on Cloud Storage: 27% Préférences de cookies

TABLES

transaction x 🔍

Tables

NAME	SCHEMA	PROPERTIES
TRANSACTIONS	PKS93518	...

Table Definition

TRANSACTIONS

COLUMN NAME	DATA TYPE	NULLABLE	LENGTH	SCALE
ID	INTEGER	N		0
VALEUR_FONCIERE	DECIMAL	N	15	2
DATE_TRANSACTION	DATE	N	4	0
SURFACE_CARREZ	DECIMAL	N	6	2
ID_LOCALISATION	INTEGER	N		0
ID_LOCAL	INTEGER	N		0

PRÉPARATION DES DONNÉES



1- Nettoyage des données

- Recherche des valeurs manquantes
 - 18 valeurs manquantes pour la colonne valeur foncière (suppression des 18 instances)
 - 1 valeur manquante pour la colonne code postal (CP indiqué manuellement)
- Correction des erreurs de formatage (date, nombres décimaux)
- Recherche des doublons (aucun)
- Recherche d'outlier (toutes les instances ont été conservées)

PRÉPARATION DES DONNÉES



2- Création des ID Locaux et ID Localisations

- Concaténation des colonnes « localisation », concaténation des colonnes « locaux » dans le tableur
- Copie des colonnes « locaux » et « localisations » dans deux nouvelles feuilles de calcul
- Suppression des doublons pour conserver les valeurs uniques
 - ➔ 23 instances locaux, 22683 instances localisations
- Création des ID (=clés primaires « locaux » et « localisations »)
- Copie des ID dans le tableur avec la fonction rechercheV (création des clés étrangères id_localisation, id_local)

3- Création des 3 tableaux locaux, localisations et transactions

4- Enregistrement au format CSV pour import dans le SGBD

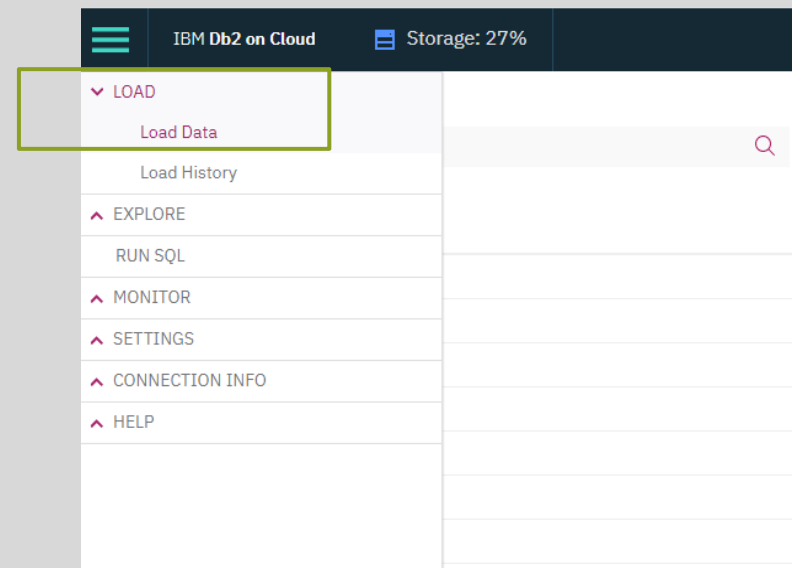
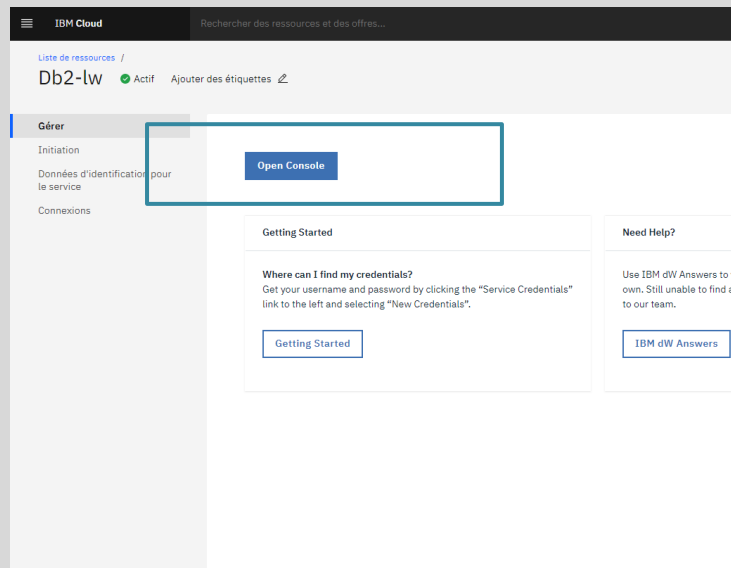
IMPORT DES DONNÉES



Utilisation de la fonction LOAD DATA d'IBM DB2 On Cloud

1- Pour accéder à la console, cliquer sur [Open Console](#)

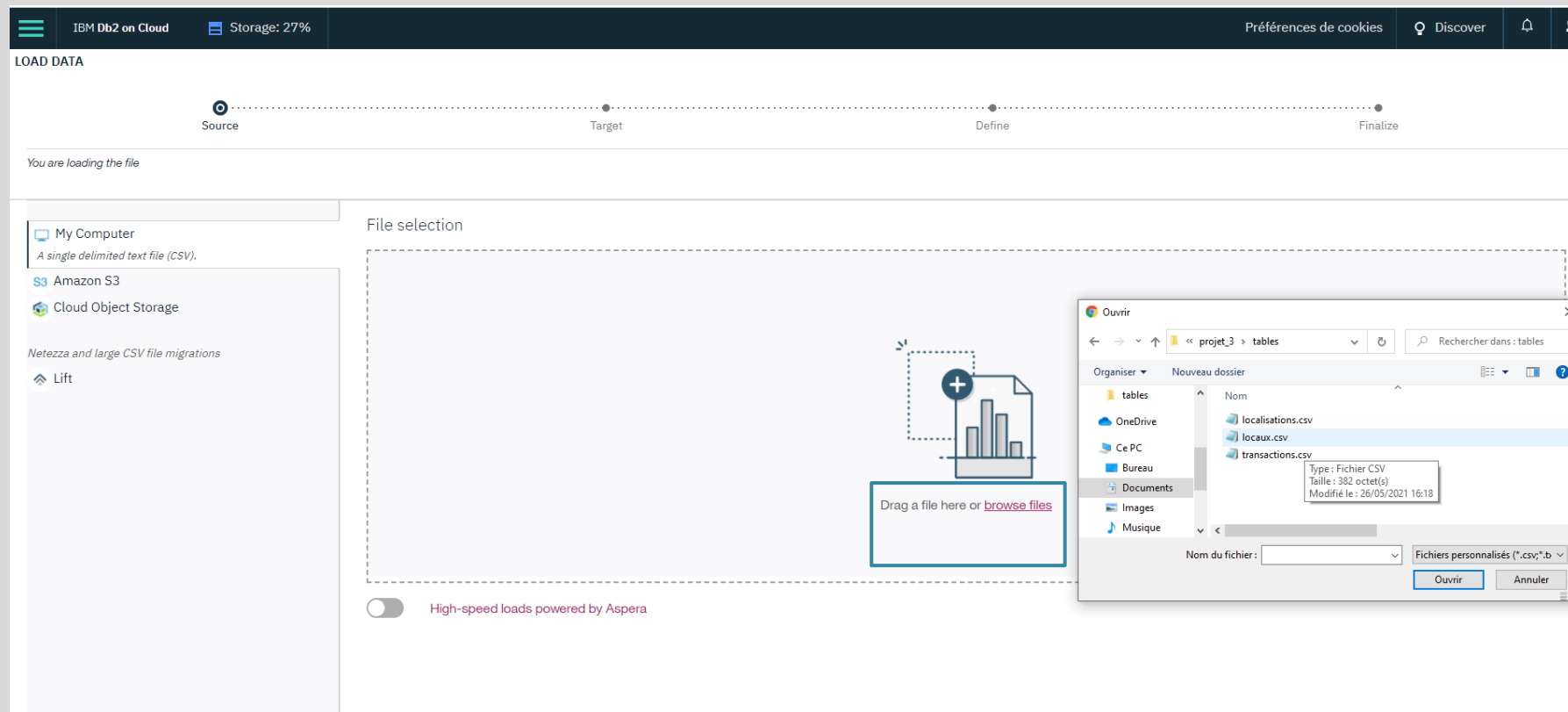
2- Cliquer sur **LOAD DATA**



IMPORT DES DONNÉES



3- Glisser-déposer ou sélectionner le fichier CSV concerné



The screenshot displays the 'LOAD DATA' interface in IBM Db2 on Cloud. The top navigation bar includes 'IBM Db2 on Cloud', 'Storage: 27%', 'Préférences de cookies', 'Discover', and user icons. The main workflow consists of four steps: Source, Target, Define, and Finalize, with 'Source' being the current step.

Under the 'Source' step, the message 'You are loading the file' is shown. On the left, a sidebar lists available sources: 'My Computer' (described as 'A single delimited text file (CSV)'), 'Amazon S3', 'Cloud Object Storage', and 'Lift' (noted for 'Netezza and large CSV file migrations').

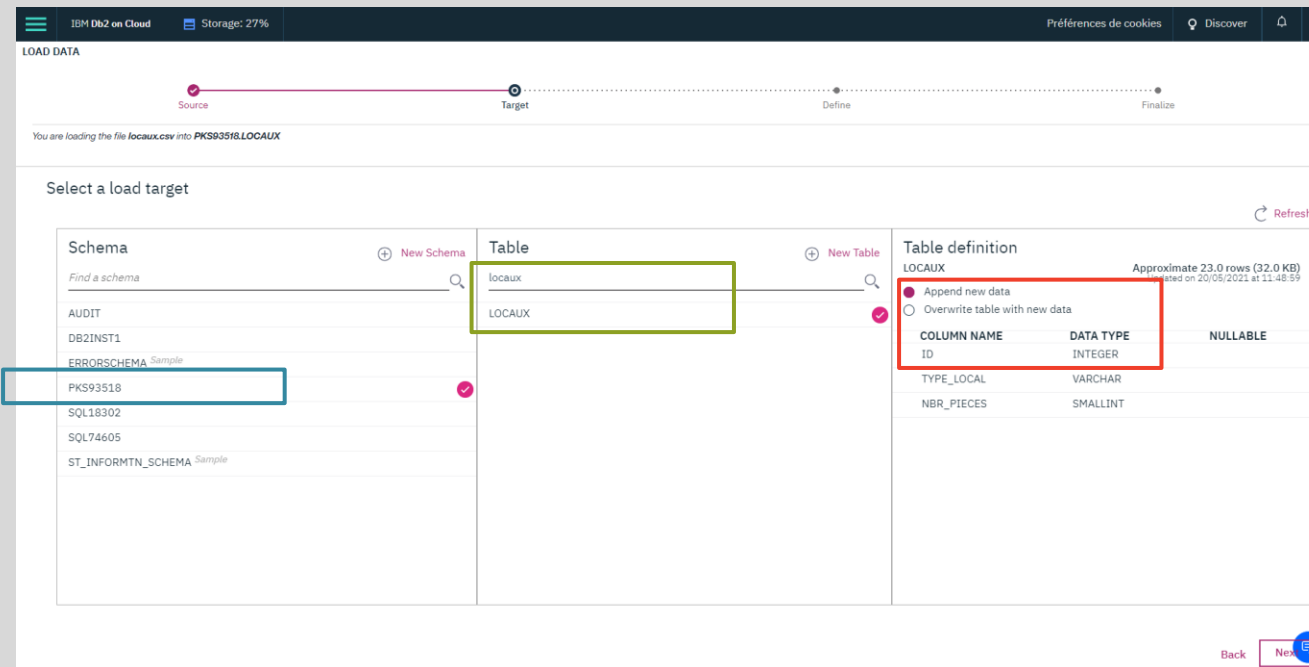
The central 'File selection' area features a large dashed box with a plus icon and a bar chart, accompanied by the text 'Drag a file here or [browse files](#)'. Below this, a toggle switch for 'High-speed loads powered by Aspera' is visible.

An 'Ouvrir' (Open) file explorer window is overlaid on the right, showing the directory path 'proj_3 > tables'. It lists three files: 'localisations.csv', 'locaux.csv', and 'transactions.csv'. A tooltip for 'localisations.csv' indicates it is a 'Fichier CSV' (382 octet(s), modified 26/05/2021 16:18). The window includes search, navigation, and file management controls.

IMPORT DES DONNÉES



- 4- Sélectionner le **schéma** dans lequel se trouve la table cible
- 5- Sélectionner la **table**
- 6- Choisir la **modalité d'import** (ajouter des données ou remplacer les données existantes)



LOAD DATA

Source Target Define Finalize

You are loading the file locaux.csv into PKS93518.LOCAUX

Select a load target

Schema

Find a schema

AUDIT

DB2INST1

ERRORSCHEMA Sample

PKS93518

SQL18302

SQL74605

ST_INFORMTN_SCHEMA Sample

Table

locaux

LOCAUX

Table definition

LOCAUX

Approximate 23.0 rows (32.0 KB)

Created on 20/05/2021 at 11:48:59

☒ Append new data

☐ Overwrite table with new data

COLUMN NAME	DATA TYPE	NULLABLE
ID	INTEGER	
TYPE_LOCAL	VARCHAR	
NBR_PIECES	SMALLINT	

Back Next

IMPORT DES DONNÉES

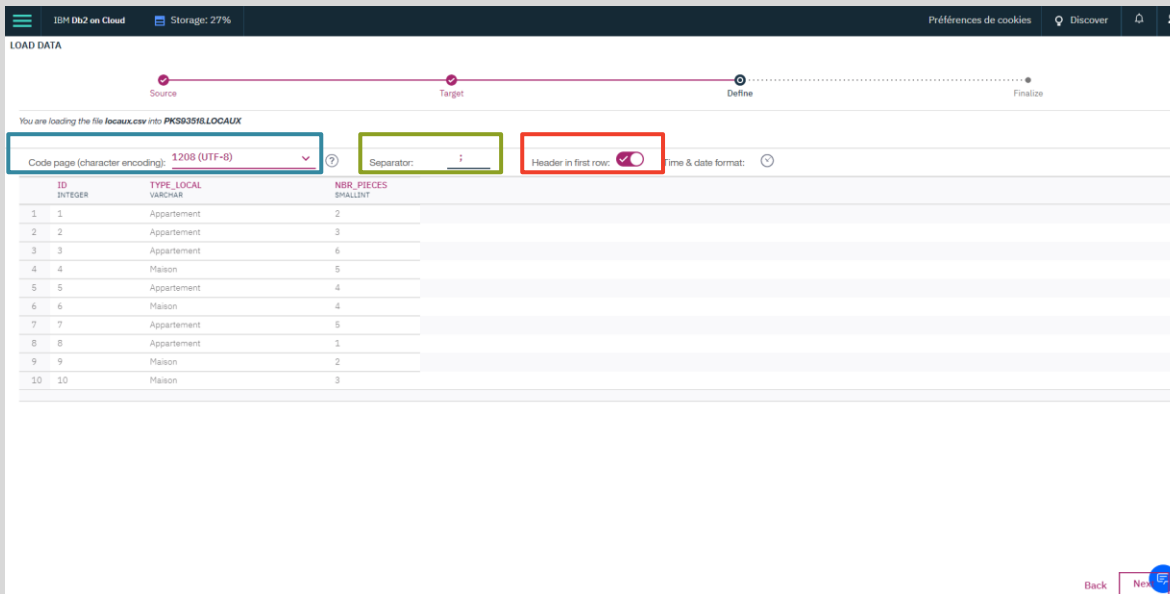


7- Sélectionner l'**encodage** du fichier CVS

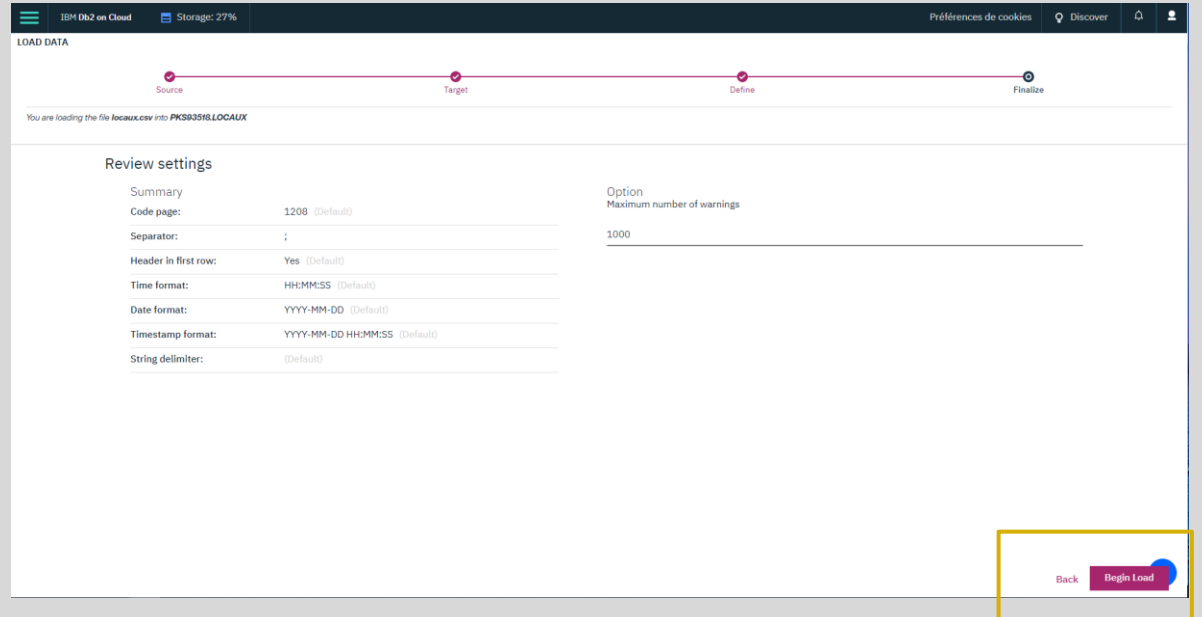
8- Indiquer le **séparateur**

9- Indiquer si la première ligne du fichier est une **en-tête**

10- **Charger** les données



ID	TYPE_LOCAL	NBR_PIECES
1	Appartement	2
2	Appartement	3
3	Appartement	6
4	Maison	5
5	Appartement	4
6	Maison	4
7	Appartement	5
8	Appartement	1
9	Maison	2
10	Maison	3



Review settings

Summary

Code page: 1208 (Default)

Separator: ;

Header in first row: Yes (Default)

Time format: HH:MM:SS (Default)

Date format: YYYY-MM-DD (Default)

Timestamp format: YYYY-MM-DD HH:MM:SS (Default)

String delimiter: (Default)

Option Maximum number of warnings: 1000

Back Begin Load

IMPORT DES DONNÉES



IBM Db2 on Cloud Storage: 27%

Préférences de cookies Discover

TABLES

Back

PKS93518.LOCAUX

ID	TYPE_LOCAL	NBR_PIECES
1	Appartement	2
2	Appartement	3
3	Appartement	6
4	Maison	5
5	Appartement	4
6	Maison	4
7	Appartement	5
8	Appartement	1
9	Maison	2
10	Maison	3
11	Maison	6
12	Appartement	0
13	Maison	1
14	Maison	8
15	Appartement	7
16	Appartement	8
17	Maison	7
18	Maison	9
19	Appartement	9
20	Maison	10
21	Maison	0
22	Appartement	11
23	Appartement	10

Table
Locaux
avec
données

CONCLUSION



Intérêt majeur de la création d'une BDD immobilière pour Laplace Immo afin :

- de connaître les indicateurs clés du marché : par localisation, en fonction des caractéristiques du bien
- d'accompagner les clients plus précisément dans l'estimation et la vente de leur bien

Limites actuelles de l'outil :

- mise à jour semestrielle sur le site DVF (avril et octobre) : pas de vue sur l'état du marché en temps réel
- réalisation de statistiques descriptives a posteriori uniquement

Perspectives :

- import des données du second semestre 2020
- import des données disponibles depuis 2016

→ création d'un modèle afin de prévoir le prix de vente des biens immobiliers

Précautions concernant les analyses effectuées :

- *Difficulté à différencier les valeurs atypiques des valeurs aberrantes dans le fichier source (possibles erreurs manuelles lors de la saisie des données sur DVF ?)*

Ex : rapport surface carrez / surface réelle bâtie , prix/m2 par rapport à la moyenne du département

- *Pas de pondération ou filtre par le nombre de ventes réalisées par commune*

Ex : moyenne valeur foncière de la commune peu représentative si peu de ventes sur la période

→ impact sur les analyses, notamment lorsqu'elles sont réalisées à l'échelle des communes

ANNEXES – ANALYSE R3

Top 10 des départements où le prix/m² est le plus chers :

- **Paris 75** se distingue par un prix au m² moyen de **12052,88 €**
- **5 départements sur 10** sont des départements d'Île-de-France
- **Autres départements :**
 - Alpes-Maritimes
 - Haute-Savoie
 - Rhône
 - Corse
 - Gironde

MOYENNE_PRIX_PAR_M2	DEPARTEMENT
12052.88	75
7219.38	92
5343.27	94
4700.32	6
4667.12	74
4344.77	93
4225.25	78
4059.31	69
4026.97	2A
3764.14	33

ANNEXES – ANALYSE R5

Top 10 des appartements les plus chers :

- Possible valeur aberrante concernant l'appartement n°1, qui présente un prix d'**1 million d'euros du m²**
- **9 appartements sur 10** se situent à **Paris**
- **L'ensemble des appartements** se situent en **Île-de-France**

VALEUR_APPARTEMENT	SURFACE	DEPARTEMENT
9000000.00	9.10	75
8600000.00	64.00	91
8577713.00	20.55	75
7620000.00	42.77	75
7600000.00	253.30	75
7535000.00	139.90	75
7420000.00	360.95	75
7200000.00	595.00	75
7050000.00	122.56	75
6600000.00	79.38	75

ANNEXES – ANALYSE R9

Top 20 des communes ayant les valeurs foncières moyennes les plus élevées :

- **Une seule vente** a été réalisée sur la période d'étude pour les communes encadrées en rouge: la valeur moyenne est donc **peu représentative**
 - possibilité de réaliser la requête en incluant une condition sur le nombre minimal de ventes réalisées
- **12 communes/arrondissements sur 20** se situent en **Île-de-France**

COMMUNE	VALEUR_FONCIERE_MOYENNE	CODE_DEPARTEMENT
GARRIGUES SAINTE EULALIE	1683000.00	30
LES PORTES-EN-RE	1400000.00	17
VERNEUIL EN HALATTE	1280000.00	60
PARIS 07	1251288.40	75
PARIS 16	1028243.69	75
PARIS 01	1013503.91	75
PARIS 08	1010536.07	75
PARIS 06	1005955.56	75
SAINT-JEAN-CAP-FERRAT	968750.00	6
VAL-D ISERE	880000.00	73
NEUILLY-SUR-SEINE	813661.07	92
PARIS 17	720622.89	75
BARBIZON	713500.00	77
PARIS 03	655953.78	75
EZE	655000.00	6
PARIS 04	650908.76	75
RAMATUELLE	633000.00	83
CRESPIERES	632065.20	78
GARCHES	615348.57	92
DUINGT	610800.00	74