



Projet Fil Rouge

ANALYSE DES DEMANDES DE VALEURS FONCIERES EN FRANCE (2017 – 2020)

Le projet présenté dans ce rapport est disponible sur GitHub :

github.com/laurentgagliardi/dvf_2017-20

Laurent Gagliardi, M2I Formation – *Data Analyste*, Décembre 2021

TABLE DES MATIERES

PRESENTATION DES DONNEES	3
NETTOYAGE DES DONNEES	5
STOCKAGE DES DONNEES – SQL SERVER	5
DOCKER	5
SQL SERVER MANAGEMENT STUDIO	6
VISUALISATION DES DONNEES – POWER BI	8
CONNEXION A SQL SERVER	8
PREPARATION DES DONNEES	9
CREATION DU DASHBOARD	9
ANALYSE DES DONNEES	10
EXPLORATORY DATA ANALYSIS – EDA	10
FOCUS : METROPOLE LYONNAISE	13
TIME SERIES ANALYSIS - TSA	14
CONCLUSION GENERALE	21

PRESENTATION DES DONNEES

Les données ont été récupérées sur le site data-gouv.fr : [Demandes de Valeurs Foncières](#).

Les données « *Demandes de valeurs foncières* » permettent de connaître les transactions immobilières (mutations) intervenues en France Métropolitaine et DOM-TOM (exception faite de l'Alsace, de la Moselle et de Mayotte) au cours des dernières années. Les données contenues sont issues des actes notariés et des informations cadastrales et sont publiées par la direction générale des finances publiques.

Les données obtenues sont celles des années 2017 à 2020 et sont structurées sous la forme de 4 fichiers texte (un pour chaque année). L'image ci-dessous présente à titre d'exemple le fichier **valeursfoncieres-2017** :

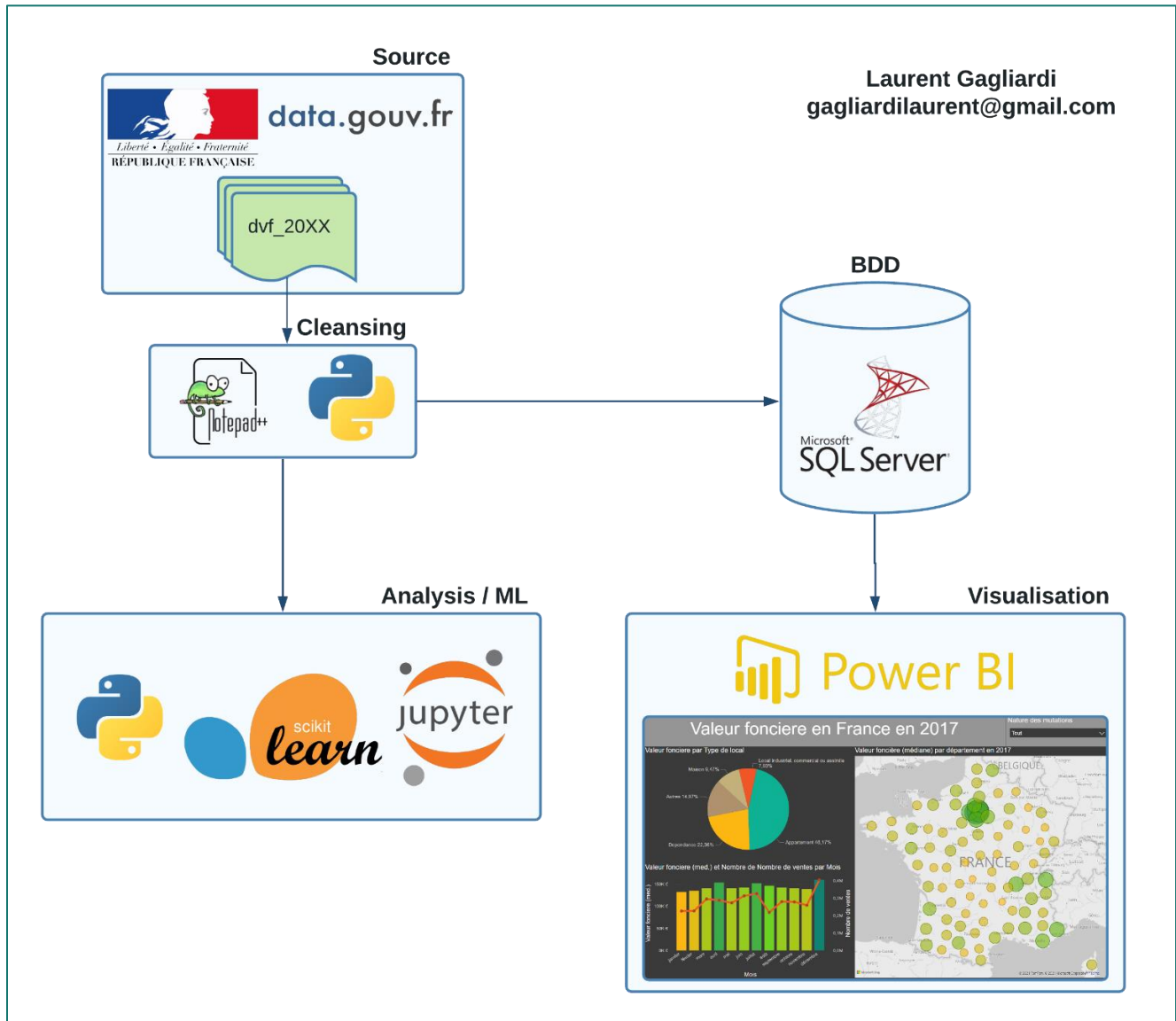
Code service	CH	Reference document	1 Articles CGI	2 Articles CGI	3 Articles CGI	4 Articles CGI	5 Articles CGI	No disposition	Date mutation	Nature mutation	Va													
000001	02/01/2017	Vente	27000,00	83	RUE 0820	CHARLES ROBIN	1000	BOURG-EN-BRESSE	01	53	BK 39	13	8	34,24	2	2	Appartement	37	2					
000001	05/01/2017	Vente	115000,00		B032	LES VAVRES	1960	PERONNAS	01	289	AR	388						AB	788					
000001	06/01/2017	Vente	1,00		B080	LA POIPE	1380	SAINT-CYR-SUR-MENTHON	01	343	ZM	197						P	42					
000001	06/01/2017	Vente	1,00		B080	LA POIPE	1380	SAINT-CYR-SUR-MENTHON	01	343	ZM	198						P	77					
000001	06/01/2017	Vente	1,00		B080	LA POIPE	1380	SAINT-CYR-SUR-MENTHON	01	343	ZM	201						P	94					
000001	09/01/2017	Vente	1,00		B205	MONTGRIMOUX CENTRE	1570	FEILLENS	01	159	AH	996						P	50					
000001	03/01/2017	Vente	258000,00	11	IMP 0384	DES PINSONS	1000	SAINT-DENIS-LES-BOURG	01	344	AK	42						0	2	Appartement	22	1	S	655
000001	03/01/2017	Vente	258000,00	11	IMP 0384	DES PINSONS	1000	SAINT-DENIS-LES-BOURG	01	344	AK	42						0	2	Appartement	22	1	S	655
000001	03/01/2017	Vente	258000,00	11	IMP 0384	DES PINSONS	1000	SAINT-DENIS-LES-BOURG	01	344	AK	42						0	3	Dependance	0	0	S	655
000001	03/01/2017	Vente	258000,00	11	IMP 0384	DES PINSONS	1000	SAINT-DENIS-LES-BOURG	01	344	AK	42						0	2	Appartement	120	5	S	655
000001	05/01/2017	Vente	175050,00	11	B244	SAINT MICHEL	1370	VAL-REVERMONT	01	426	ZI	195						0	1	Maison	99	5	S	471
000001	06/01/2017	Vente	162000,00	6	RUE 2330	LOUIS BLERIOT	1000	BOURG-EN-BRESSE	01	53	BM	425						0	3	Dependance	0	0	S	158
000001	06/01/2017	Vente	162000,00	6	RUE 2330	LOUIS BLERIOT	1000	BOURG-EN-BRESSE	01	53	BM	425	8	69,63				1	2	Appartement	69	3		
000001	06/01/2017	Vente	162000,00	6	RUE 2330	LOUIS BLERIOT	1000	BOURG-EN-BRESSE	01	53	BM	425	21					1	3	Dependance	0	0		
000001	06/01/2017	Vente	162000,00	2	RUE 2330	LOUIS BLERIOT	1000	BOURG-EN-BRESSE	01	53	BM	428						0					S	968
000001	11/01/2017	Vente	165900,00	5187	B077	LES METRAS	1250	CEYZERIAT	01	72	AC	448						0	1	Maison	90	4	S	296
000001	13/01/2017	Vente	181800,00	277	RUE 1895	DES HIRONDELLES	1000	BOURG-EN-BRESSE	01	53	CV	32						0					S	27
000001	13/01/2017	Vente	181800,00	277	RUE 1895	DES HIRONDELLES	1000	BOURG-EN-BRESSE	01	53	CV	319						0	1	Maison	80	4	S	1067
000001	04/01/2017	Vente	177000,00	26	RUE 2520	MAL FOCH	1000	BOURG-EN-BRESSE	01	53	AD	198						0	4	Local industriel, commercial ou assimile				80
000001	04/01/2017	Vente	177000,00	26	RUE 2520	MAL FOCH	1000	BOURG-EN-BRESSE	01	53	AD	198						0	2	Appartement	38	1	S	70
000001	04/01/2017	Vente	177000,00	26	RUE 2520	MAL FOCH	1000	BOURG-EN-BRESSE	01	53	AD	198						0	2	Appartement	38	1	S	70
000001	02/01/2017	Vente	130000,00	5294	B020	LHOMONT	1340	MALAFRETAT	01	229	AA	20						0	4	Local industriel, commercial ou assimile	160	0	S	
000001	03/01/2017	Vente	174500,00	121	IMP 0035	DE FEIGNOUX	1310	CONFRANCON	01	115	AA	79						0	1	Maison	80	4	S	1173
000001	06/01/2017	Vente	205788,75	5462	B011	CERTINES	1240	CERTINES	01	69	C	549						0	1	Maison	101	5	S	809
000001	06/01/2017	Vente	179000,00	190	CHEM 0330	DE MAS GRUZIN	1250	TOSSIAT	01	422	B	841						0	1	Maison	80	3	S	1283
000001	09/01/2017	Vente	47000,00	924	RUE 0170	PRINCIPALE	1270	BEAUPONT	01	29	C	1073						0	1	Maison	84	3	S	1340
000001	09/01/2017	Vente	47000,00		B002	BEAUPONT VILLAGE	1270	BEAUPONT	01	29	C	1075						0					T	630
000001	09/01/2017	Vente	47000,00		B002	BEAUPONT VILLAGE	1270	BEAUPONT	01	29	C	1076						0					J	450
000001	03/01/2017	Vente	7500,00		B003	LES ASSARDS	1851	MARBOZ	01	232	WN	111						0		BS			12290	
000001	03/01/2017	Vente	1500,00		B182	EN VERNOSSET	1270	COLIGNY	01	108	AD	88						0		P			1262	
000001	05/01/2017	Vente	138000,00	6	LOT A003	LE BIOLAY	1370	SAINT-ETIENNE-DU-BOIS	01	350	AA	12						0	1	Maison	101	5	S	419
000001	13/01/2017	Vente	161340,00	5706	B013	BROTTEAUX DU BLANCHON	1160	PONT-D AIN	01	304	AK	84						0	1	Maison	85	5	S	972
000001	12/01/2017	Vente	186000,00		B018	CUETANT	1290	SAINT-JEAN-SUR-VEYLE	01	365	C	634						0					J	375
000001	12/01/2017	Vente	186000,00		B018	CUETANT	1290	SAINT-JEAN-SUR-VEYLE	01	365	C	635						0		VI			567	
000001	12/01/2017	Vente	186000,00	5151	B018	CUETANT	1290	SAINT-JEAN-SUR-VEYLE	01	365	C	1152						0	1	Maison	160	6	S	716
000001	11/01/2017	Vente	10000,00		B160	SOUS LE BOIS	1270	COLIGNY	01	108	ZK	1						0		P			3200	
000001	11/01/2017	Vente	10000,00		B160	SOUS LE BOIS	1270	COLIGNY	01	108	ZL	29						0		BT			6920	
000001	11/01/2017	Vente	10000,00		B160	SOUS LE BOIS	1270	COLIGNY	01	108	ZL	40						0		BT			7970	
000001	12/01/2017	Vente	84000,00	64	RUE 0135	BERNARD	1340	CRAS-SUR-REYSSOUZE	01	130	B	568						0	1	Maison	154	5	S	630
000001	09/01/2017	Vente	137000,00	412	0300	GRANDE ROUTE	1190	OZAN	01	284	A	93						0	1	Maison	87	4	S	868
000001	08/01/2017	Vente	137000,00		B064	OZAN	1190	OZAN	01	284	A	93						0						1390

A l'origine, les fichiers texte contenaient 43 colonnes mais pour l'étude (et parce qu'il y avait des données manquantes), ceux-ci ont été épurés de sorte à ce que les informations principales suivantes soient conservées :

- Date de la mutation au format (DD/MM/YYYY).
- Nature de la mutation : vente, vente en l'état futur d'achèvement, vente de terrain à bâtir, échange, adjudication, expropriation.
- Valeur foncière (séparateur décimal : virgule).
- Données géographiques :
 - Code postal (5 caractères).

- Code départemental INSEE.
- Commune.
- Type de local : appartement, dépendance, maison, local industriel/commercial ou assimilé...
- Surface réelle bâtie, surface de terrain.

Le projet Fil Rouge 🧵 a pour objectif de mobiliser les connaissances acquises lors de la formation « *Data Analyst* » et les connaissances personnelles. L'idée a donc été d'exploiter les données « DVF » pour répondre à cette problématique. Le schéma ci-dessous présente le projet dans son ensemble (et les technologies utilisées) :



Pour résumer :

- Récupération des données « DVF » via data-gouv.fr (Sources) : 4 fichiers plats.
- Nettoyage des données (Cleansing) et génération des fichiers .csv à l'aide de **Notepad++** et de scripts python (disponibles sur GitHub).

- Stockage des données sous SQL Server (en local, via Docker) et grâce à l'outil *SQL Server Management Studio*.
 - Visualisation avec Power BI (connecté à SQL Server).
- Analyse de données avec Jupyter notebook.

NETTOYAGE DES DONNEES

Les fichiers texte de chaque année (4 fichiers) contenant plus de 3M de lignes chacun, et l'ordinateur utilisé ne disposant que de 16G de RAM (résultant en des lenteurs), il a été nécessaire de spliter les fichiers à l'aide *Notepad++* pour ne pas avoir d'erreur de type « *out of memory* » durant l'exécution des traitements de nettoyage des données.

Pour chaque année, il y avait donc 7 (ou 8) fichiers texte. Ces fichiers de maximum 500 000 lignes chacun ont été utilisés comme sources par les scripts python de nettoyage et de génération des .csv (disponibles sur GitHub).

Voici les principales étapes des scripts :

- Import des fichiers textes et concaténation en un dataframe (*dvf_20XX*) à l'aide de Pandas.
- Nettoyage du dataframe :
 - Suppression des colonnes inutiles.
 - Ajout des 0 de début des champs « Code postal » et « Code départemental » pour les cas concernés.
 - Remplacement des séparateurs décimaux (virgules en points) pour le champ « Valeur foncière ».
- Chargement du dataframe nettoyé dans un .csv (*dvf_20XX.csv*).

STOCKAGE DES DONNEES – SQL SERVER

Les données étaient donc prêtes à être stockées dans une base de données. **SQL Server** a été retenu pour répondre à la problématique de stockage en raison de la simplicité de son installation et son utilisation.

DOCKER

Docker a été utilisé pour monter une image de SQL Server. Il a pour cela suffi d'ouvrir un PowerShell et d'exécuter les deux commandes suivantes :

```
docker pull mcr.microsoft.com/mssql/server:2019-CU14-ubuntu-20.04
```

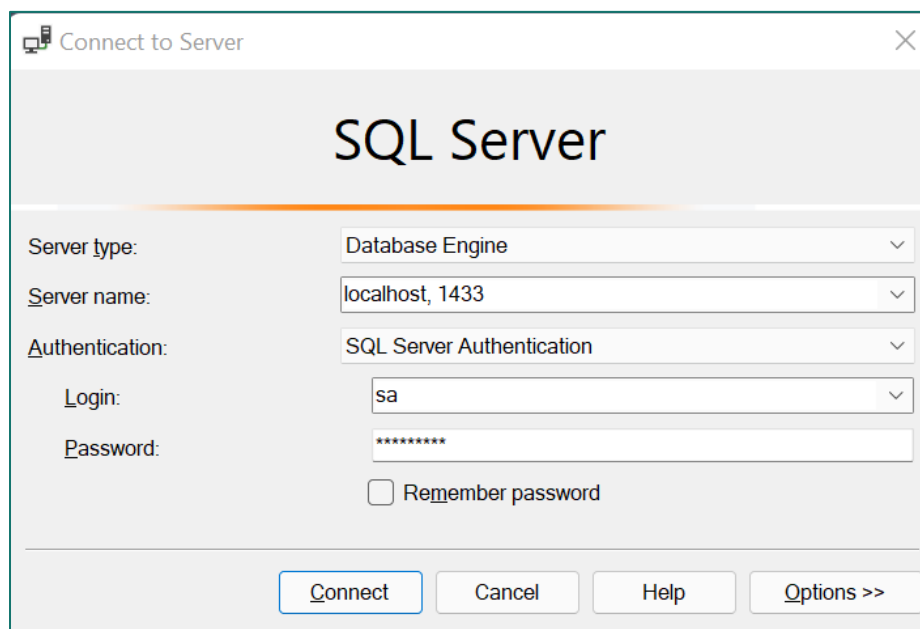
```
docker run -e "ACCEPT_EULA=Y" -e "SA_PASSWORD=<my_password>" -e "MSSQL_PID=Express" -p 1433:1433  
-d mcr.microsoft.com/mssql/server:2019-CU14-ubuntu-20.04
```

La première commande télécharge une image de SQL Server compatible avec la version 20.04 de Ubuntu. La seconde commande crée un SQL Server en *localhost:1433* avec un user « *sa* » dont vous choisissez le password.

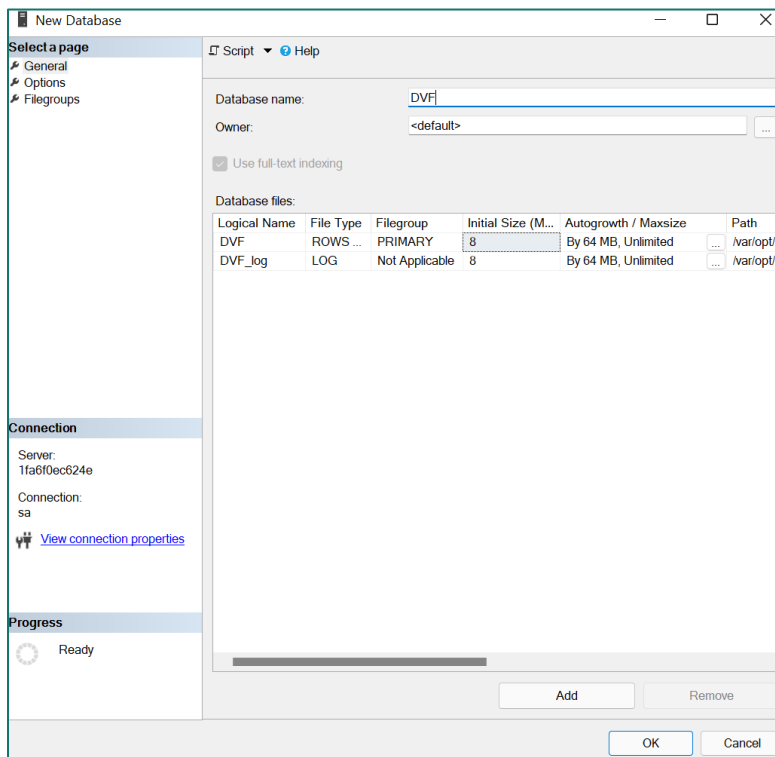
La documentation suivie pour cette installation est disponible ici : [MySQL Server via Docker](#).

SQL SERVER MANAGEMENT STUDIO

Une fois l'image de SQL Server en place, le logiciel [Microsoft SQL Server Management Studio 18](#) a été choisi pour manager le stockage des données. Pour se connecter, les informations fournies lors de la création via Docker ont été utilisées (voir l'image ci-dessous).



Une fois connecté, il a ensuite été question de créer une nouvelle database : *DVF* (voir l'image ci-dessous).



La database *DVF* désormais existante, il ne restait plus qu'à créer les tables correspondantes à chacun des fichiers *dvf_20XX.csv*. Il existe une option « Import Flat file » pour réaliser cela. Les tables ainsi créées ont un nom identiques aux *.csv* pour plus de clarté (*dbo.dvf_20XX*).

Voici, une requête *SELECT* simple permettant de voir à quoi ressemble l'une de nos tables :

SQLQuery1.sql - loc...1433.DVF (sa (54))*

```
SELECT * FROM dbo.dvf_2017;
```

100 %

Results Messages

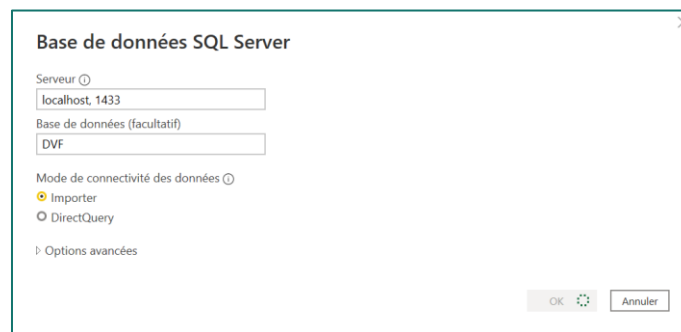
	column1	No_disposition	Date_mutation	Nature_mutation	Valeur_fonciere	No_voie	Type_de_voie	Code_voie	Voie
1	0	000001	2017-01-02	Vente	27000.00	83	RUE	0820	CHARLES ROBIN
2	1	000001	2017-01-05	Vente	115000.00	NULL	NULL	B032	LES VAVRES
3	2	000001	2017-01-06	Vente	1.00	NULL	NULL	B080	LA POIPE
4	3	000001	2017-01-06	Vente	1.00	NULL	NULL	B080	LA POIPE
5	4	000001	2017-01-06	Vente	1.00	NULL	NULL	B080	LA POIPE
6	5	000001	2017-01-09	Vente	1.00	NULL	NULL	B205	MONTGRIMOUX CENTF
7	6	000001	2017-01-03	Vente	258000.00	11	IMP	0384	DES PINSONS
8	7	000001	2017-01-03	Vente	258000.00	11	IMP	0384	DES PINSONS

VISUALISATION DES DONNEES – POWER BI

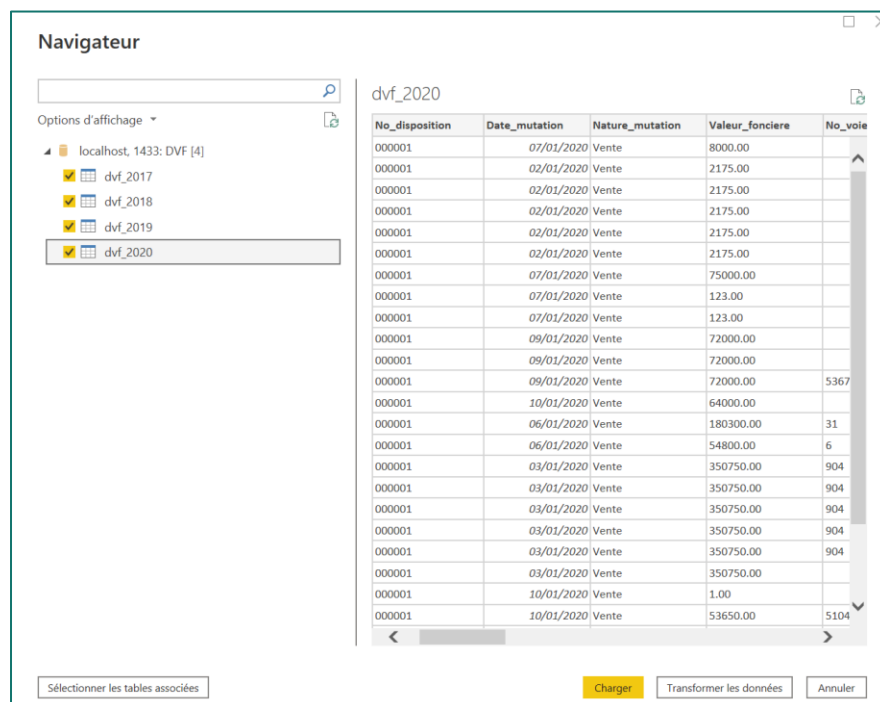
Pour la visualisation des données, l'outil Power BI a été choisi. Ce choix s'explique à travers deux principales raisons : la puissance de l'outil d'une part, et sa simplicité de connexion à SQL Server (les deux étant des outils Microsoft).

CONNEXION A SQL SERVER

Une connexion à SQL Server a été nécessaire pour récupérer les données des tables et pouvoir fournir une visualisation. Les deux outils s'associant aisément, il a suffit de renseigner les informations du SQL Server.



On obtient alors les tables dans Power BI (voir l'image ci-dessous).



PREPARATION DES DONNEES

Les données récupérées ont dû être préparées/transformées pour permettre une bonne visualisation.

Voici les principales transformations opérées :

- Transformation des données « Valeur fonciere », « Surface réelle batie » et « Surface terrain » : changement des « . » en « , » et changement du type de données : devise pour la valeur foncière et décimal pour les surfaces.
- Définition des données géographiques (*hiérarchie Géo*) :
 - Pays : ajout d'une colonne « Pays » contenant la valeur « France ».
 - Département : code département.
 - Code postal : code postal.
 - Localité : commune.

Gestion des données vides pour les types de locaux : remplacement des « *null* » par « Autres ».

CREATION DU DASHBOARD

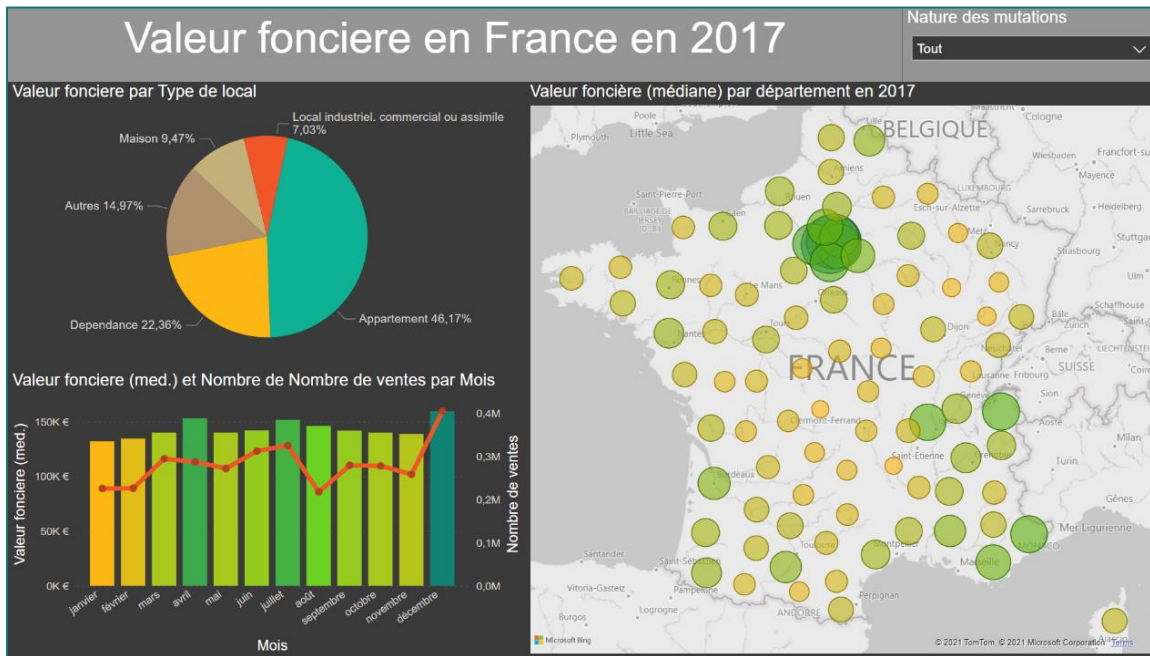
La visualisation est constituée des 4 dashboards – un pour chaque année – construits selon le même schéma détaillé ci-après :

- Titre : Valeur foncière de l'année 20XX.
- Nature de la mutation : filtre sur le type de mutation.
- Valeur foncière (médiane) par département : 3D Map.
- Valeur foncière (médiane) par type de local : pie chart sur le pourcentage de valeur foncière totale associé à chaque type de local.
- Valeur foncière (médiane) et nombre de mutations par mois : histogramme des valeurs foncières par mois groupé à une courbe du nombre de mutations par mois au cours de l'année 20XX.

Les aperçus des dashboards des autres années sont disponibles sur GitHub.

Le .pbix est, quant à lui, consultable ici : [dahsboard annuel](#).

A titre d'exemple, voici l'aspect du dashboard pour l'année 2017 (voir l'image ci-dessous) :

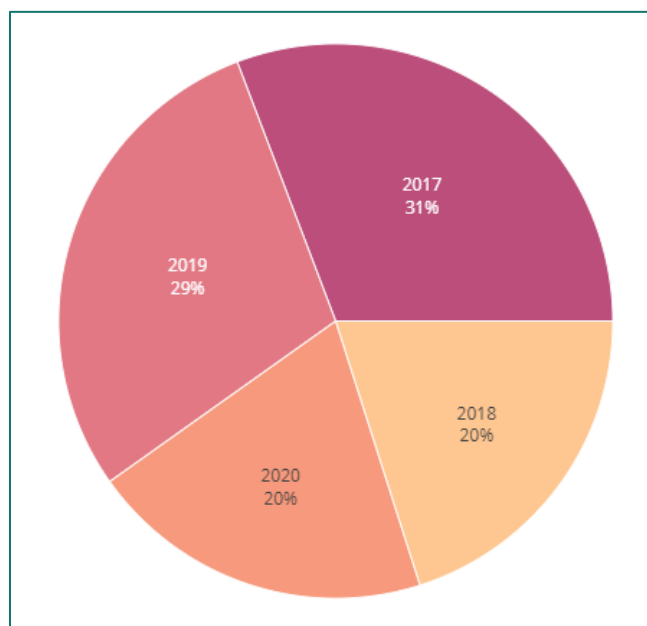


ANALYSE DES DONNEES

EXPLORATORY DATA ANALYSIS – EDA

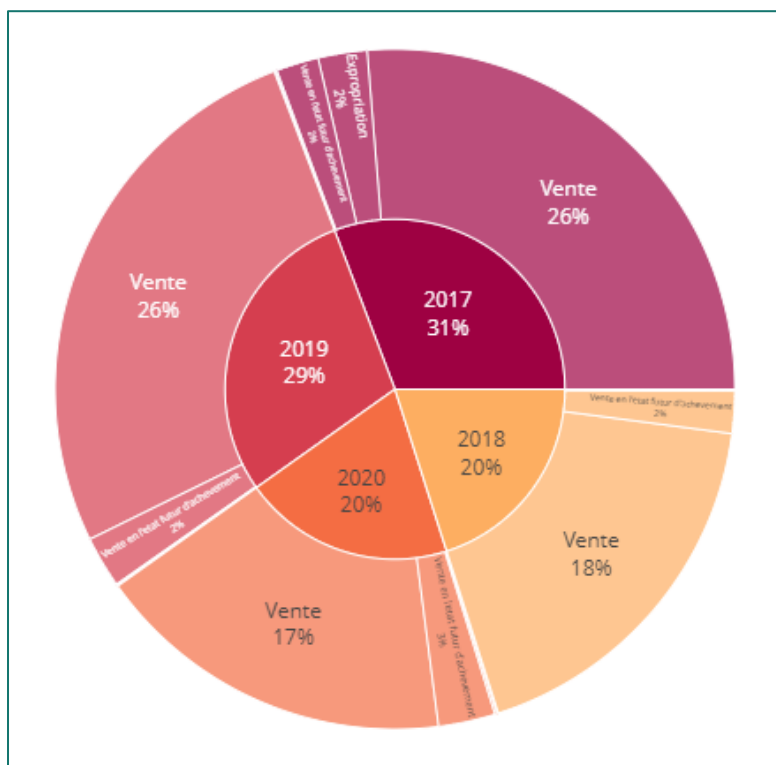
Note : l'ensemble du code ayant permis de réaliser cette analyse est disponible sur [GitHub](#)

Tout d'abord, il a été question de vérifier l'évolution de la valeur foncière sur la période 2017 – 2020. Le piechart ci-dessous présente la contribution relative des années à la valeur foncière totale.



On remarque que les années 2018 et 2020 sont à 20%, tandis que les années 2019 et 2017 sont d'environ 30% (contre 25% par année en cas d'équirépartition). Cela indique que les années 2019 et 2017 ont été plus intéressantes en terme de valeur foncière.

Par la suite, la représentation des différents types de mutations en fonction des années a été analysée. Le sunburst suivant illustre cette partie de l'analyse.



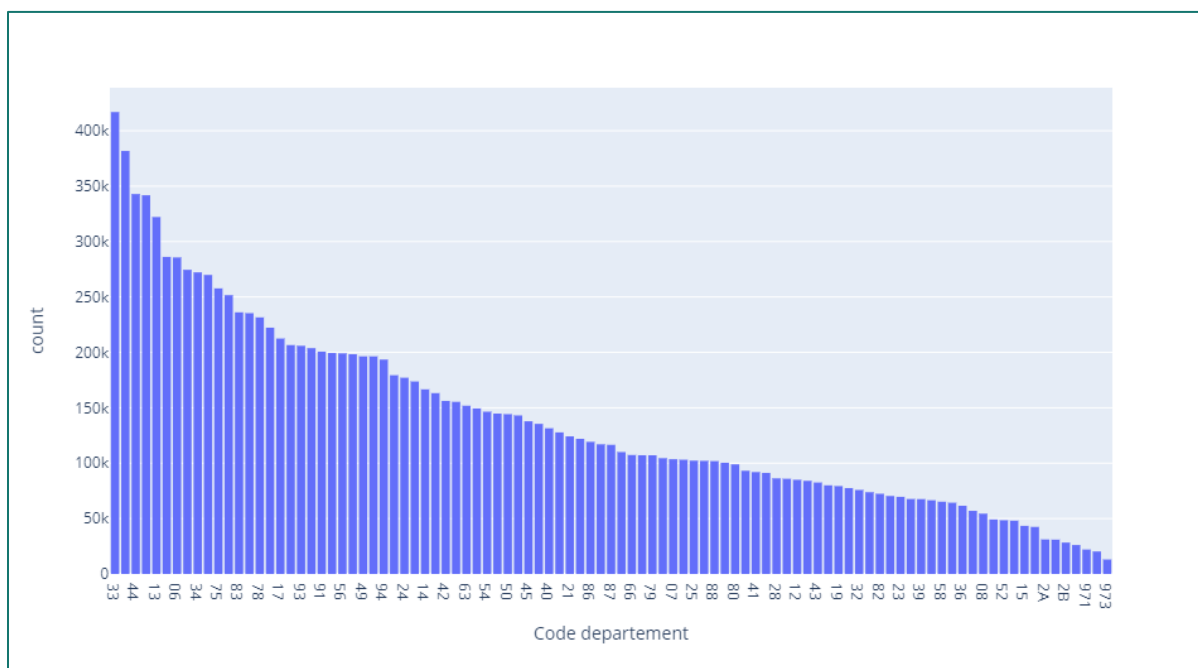
On observe le même schéma au sujet des années 2018 et 2020, à savoir une importance moindre quant au total de valeur foncière. Au sujet du second niveau d'analyse (les types de mutations par année), on constate que la majorité des transactions s'avèrent être des **ventes** et cette dominance est conservée sur les 4 années. De plus, les ventes relatives aux années 2017 et 2019 occupent une part relative plus importante (26% contre 17/18% les autres années).

On considère également un pourcentage élevé d'expropriation au cours de l'année 2017. Malgré des recherches, il a été impossible de trouver une explication politique/sociale à cette différence.

Après cela, le focus a été fait sur le nombre de transactions par département afin d'analyser la composante spatiale des données.

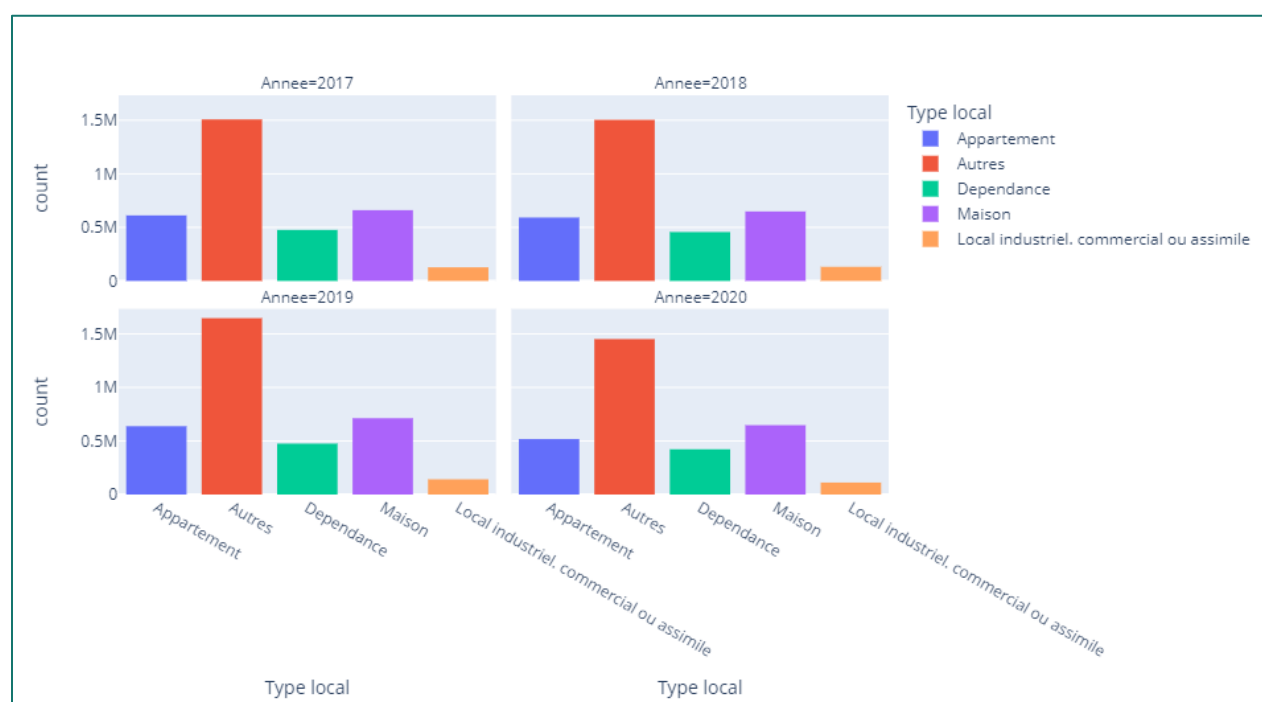
Il est apparu que les départements ayant le plus de transactions sont les départements contenant des métropoles : l'histogramme suivant présente les départements 33, 59, 44, 69 et 13 en top 5 concernant le nombre de mutations. On peut aisément associer ces départements aux **métropoles** qu'ils abritent : 33 – Bordeaux, 59 – Lille, 44 – Nantes, 69 – Lyon et 13 – Marseille.

Les métropoles étant des zones tendues du point de vue immobilier, il n'est pas étonnant de s'apercevoir qu'elles concentrent la majorité des transactions.



L’histogramme montre également, une fois les métropoles passées, que les villes à forte croissance représentent une fraction non négligeable des mutations. Cela a d’ailleurs été renforcé par l’arrivée de l’épidémie de Covid-19, et s’est donc répercuté dans les transactions en dehors des grandes villes.

Prudence néanmoins, il est important de mentionner que les départements adjacents à la métropole parisienne sont bien classés sur cette représentation. Mais cela est dû au fait que Paris est une zone tellement tendue, que la demande immobilière la concernant a créé un effet d’entraînement sur les communes limitrophes.



Un autre aspect intéressant du dataset est la variété des types de locaux. Le graphe ci-dessus est constitué de 4 histogrammes représentant chacun le nombre de mutations par type de local et ce pour chaque année.

Ces graphes sont très similaires, on note qu'il ne semble pas avoir de réelle différence de proportions entre les types de locaux au fil des années.

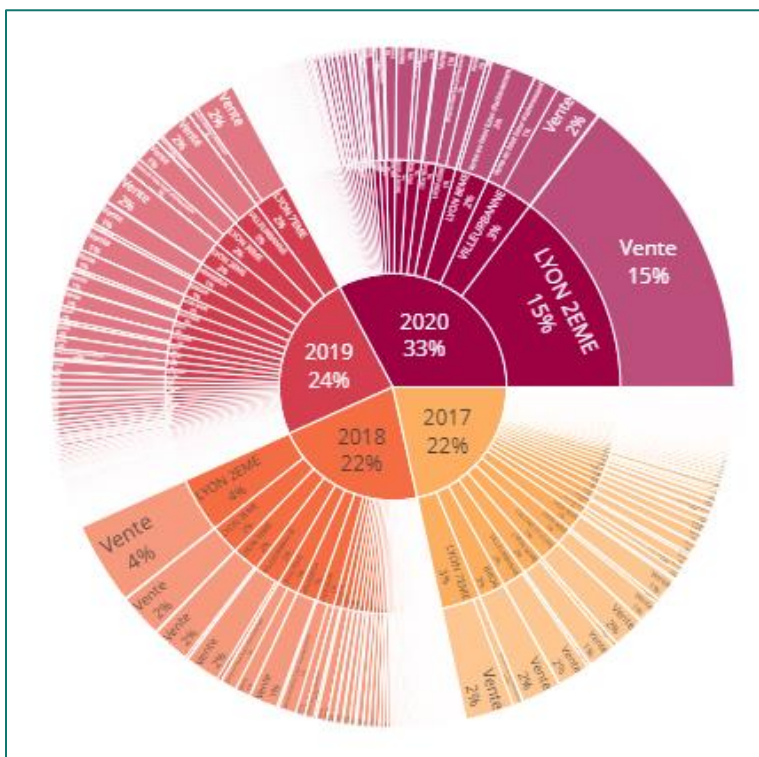
Voilà qui conclut l'analyse exploratoire sur l'ensemble des données. La suite de l'analyse constitue un focus sur les données de la métropole lyonnaise.

FOCUS : METROPOLE LYONNAISE

Ce focus démarre avec un sunburst (figure suivante) précisant la part relative des villes et des mutations en leur sein pour le département du Rhône.

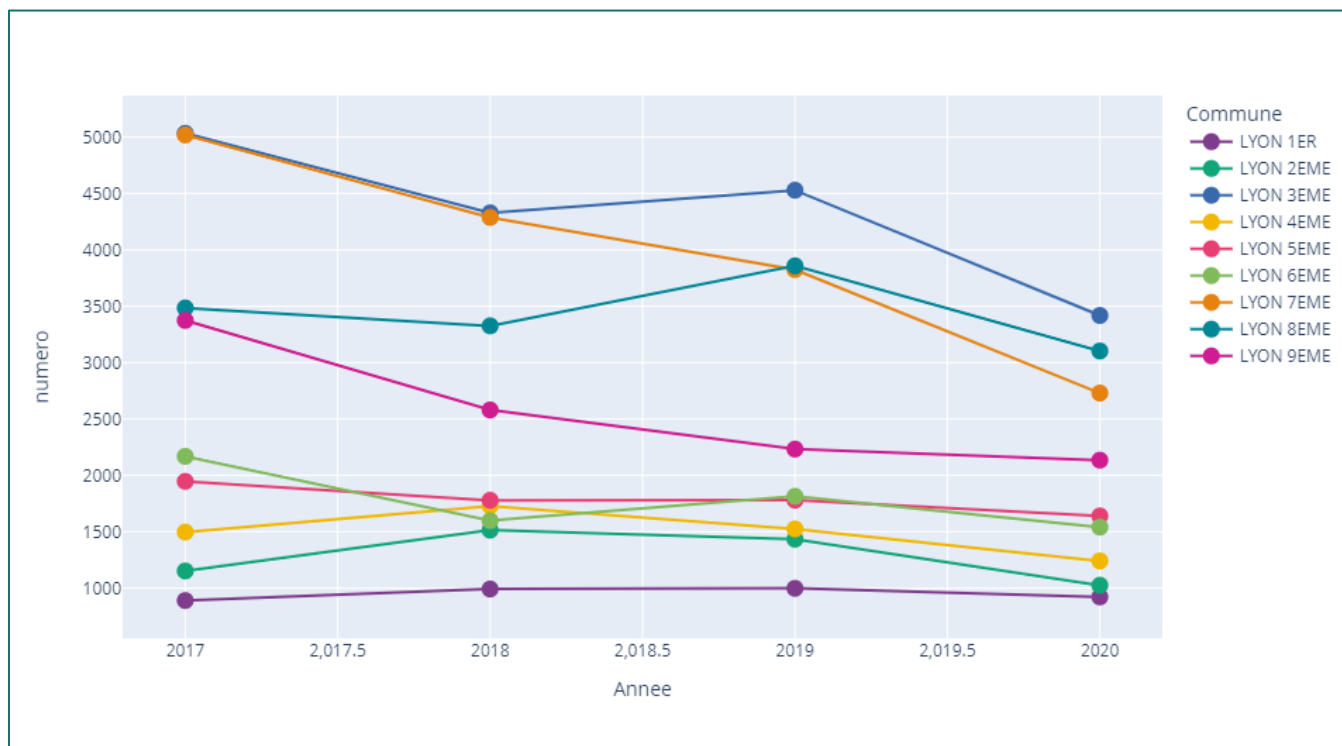
- On retrouve dans un premier temps le même constat qu'en global sur la France. A savoir que l'année 2020 a été plus importante en terme de valeur foncière (33% du chiffre global).
- Avec un deuxième degré d'analyse, on constate que l'essentiel des transactions du Rhône ont eu lieu (quelque soit l'année) sur l'ensemble de la métropole lyonnaise (Lyon et sa banlieue).
- Le dernier niveau d'analyse permet de conclure que l'essentiel des transactions est de type « Vente », et ce quelque soit la commune étudiée.

La concentration des transactions au sein de la métropole lyonnaise, a conduit à une simplification de l'analyse – à savoir le fait de limiter la suite de l'étude à l'ensemble des arrondissements de la ville de Lyon.



Le lineplot suivant présente l'évolution du nombre de transactions par année pour les différents arrondissements lyonnais. On peut considérer deux groupes :

- Le 1^{er} groupe présente un nombre élevé de transactions (2500 – 5000 par an) malgré une tendance à la baisse (surtout sur 2020).
- Le second groupe présente un nombre plus faible de transactions (1000 – 2500 par an) et une stabilité certaine sur la période de 4 ans.



Bien qu'il aurait été intéressant de corréler cela avec l'évolution du prix du m² ou avec la population (comme il a déjà été fait dans la littérature). L'étude n'a pas pris cette direction, par faute de temps.

En revanche, des travaux ont été initiés sur l'analyse temporelle des mutations lyonnaises entre 2017 et 2020. Cela constitue la suite de ce compte-rendu.

TIME SERIES ANALYSIS - TSA

L'analyse temporelle débute avec l'évolution du nombre de transactions par arrondissement de Lyon entre 2017 et 2020 (figure ci-dessous).

Il existe une **hétérogénéité** frappante entre les arrondissements quant au nombre de transactions sur la période considérée. Par exemple, Lyon 3^e, 7^e et 8^e dépassent fréquemment la barre des 200 transactions, alors que les autres que très rarement (ci ce n'est jamais). Cela témoigne de quartiers dont le paysage urbain est en pleine métamorphose.



Tout lyonnais avisé est au courant du grand plan d'urbanisme dont un certain nombre d'actions se focalisent sur ces 3 arrondissements. Les mutations sont donc naturellement plus importantes dans ces zones.

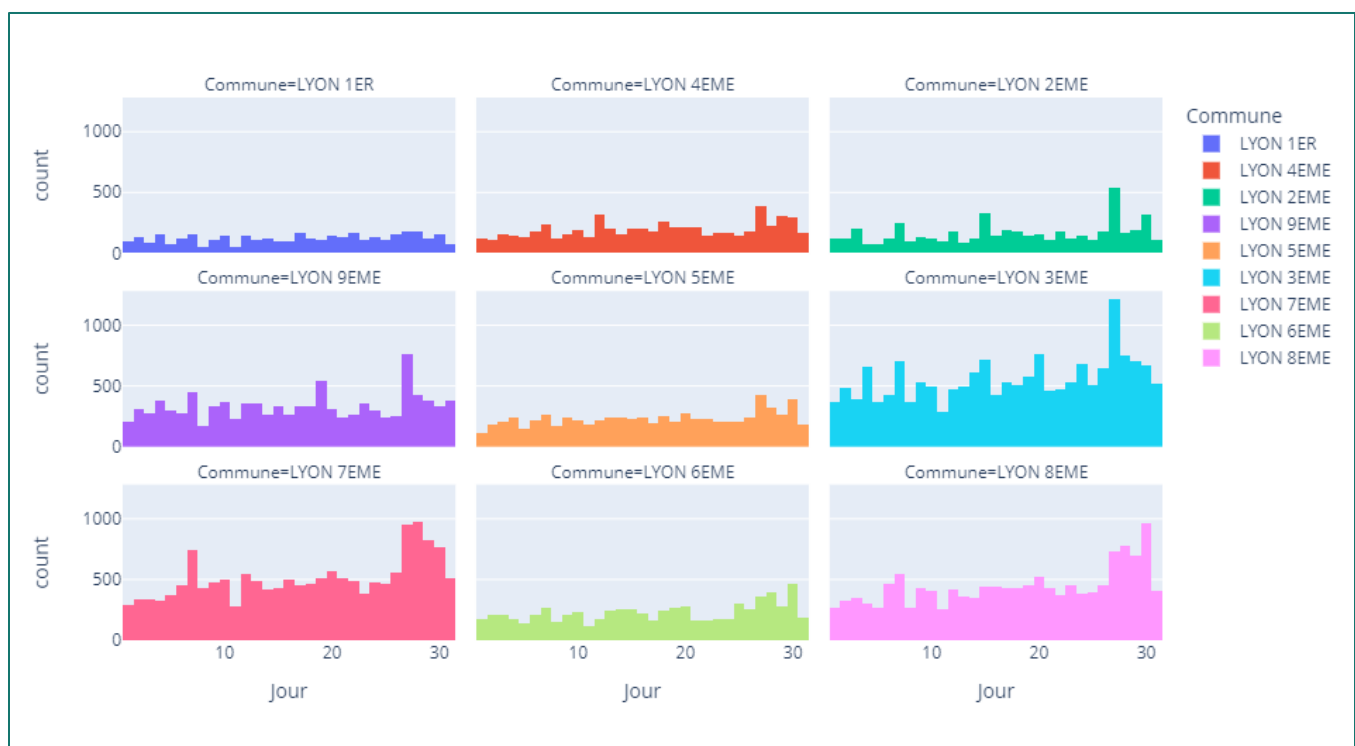
- Lyon 7^e et l'**urbanisation** du quartier « Jean Jaurès » qui a débuté avant même 2017. De nombreux locaux d'entreprises sont en construction, attirant par la même occasion de nouveaux commerces et entraînant une gentrification de la population locale.
- Lyon 3^e et la rénovation de la gare Part Dieu avec en projet connexe la construction de nouveaux locaux d'entreprise. Et cela sans compter, l'apparition massive d'immeubles dans le quartier Montchat (dont les maisons sont donc rachetées).
- Lyon 8^e et la disparition accélérée des pavillons au profit de nouveaux immeubles résidentiels.

Par la suite, il s'agit d'identifier une **saisonnalité** dans les données. Le graphique suivant est très similaire au précédent, à l'exception que les données sont regroupées par mois.

On retrouve logiquement les différences d'échelles entre les arrondissements ; mais on observe également qu'il n'y a pas de saisonnalité au niveau des mois dans les transactions. On peut toutefois apprécier une période creuse lors des vacances d'été (août principalement).

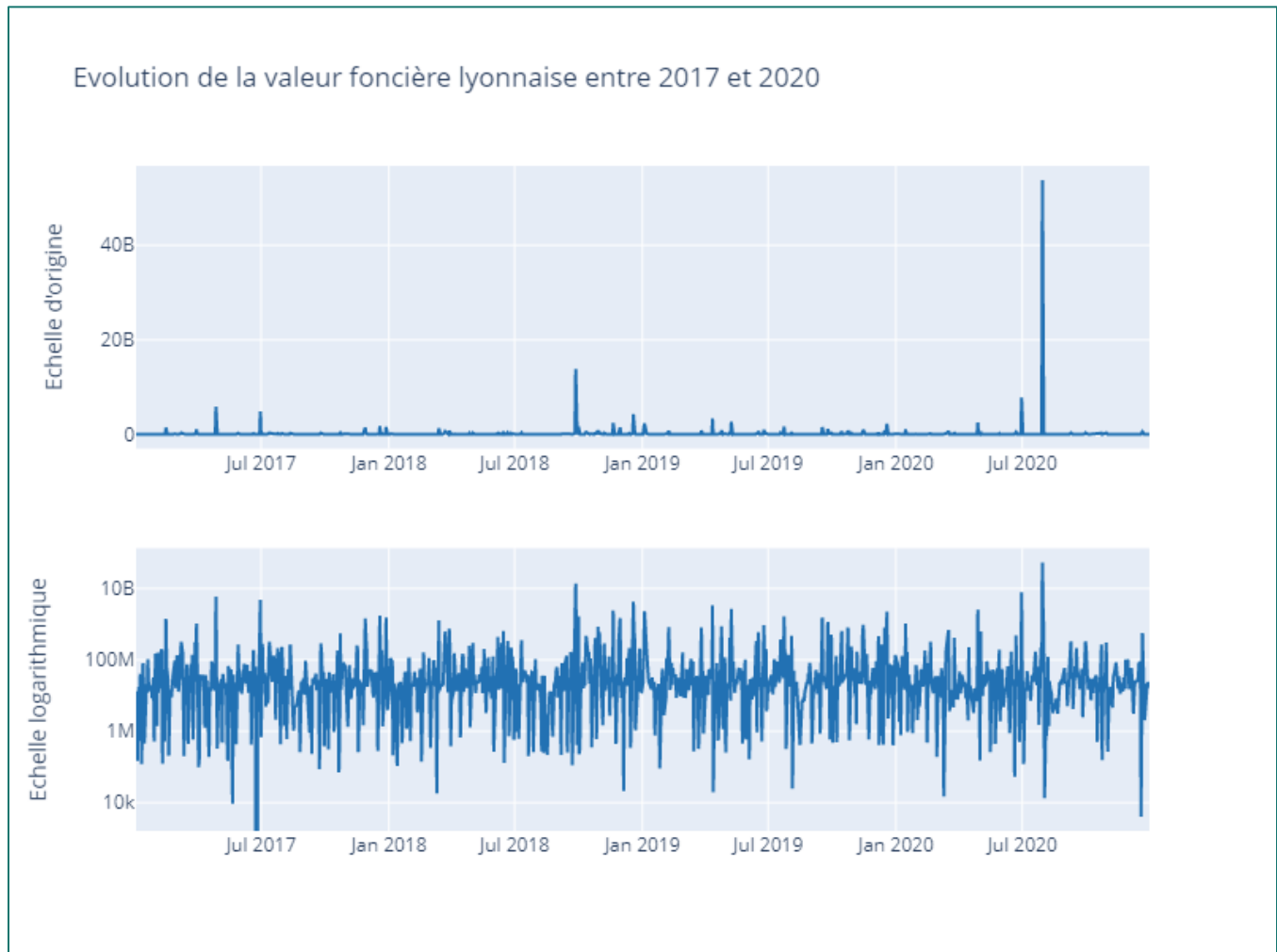


La recherche de saisonnalité a toutefois démontré une augmentation du nombre de transactions sur la fin du mois (voir la figure suivante). Il y a un motif repeté de façon mensuelle.



Pour la suite de l'analyse, il a été décidé d'analyser les données sur Lyon (au global, sans entrer dans le détail des arrondissements dans un premier temps).

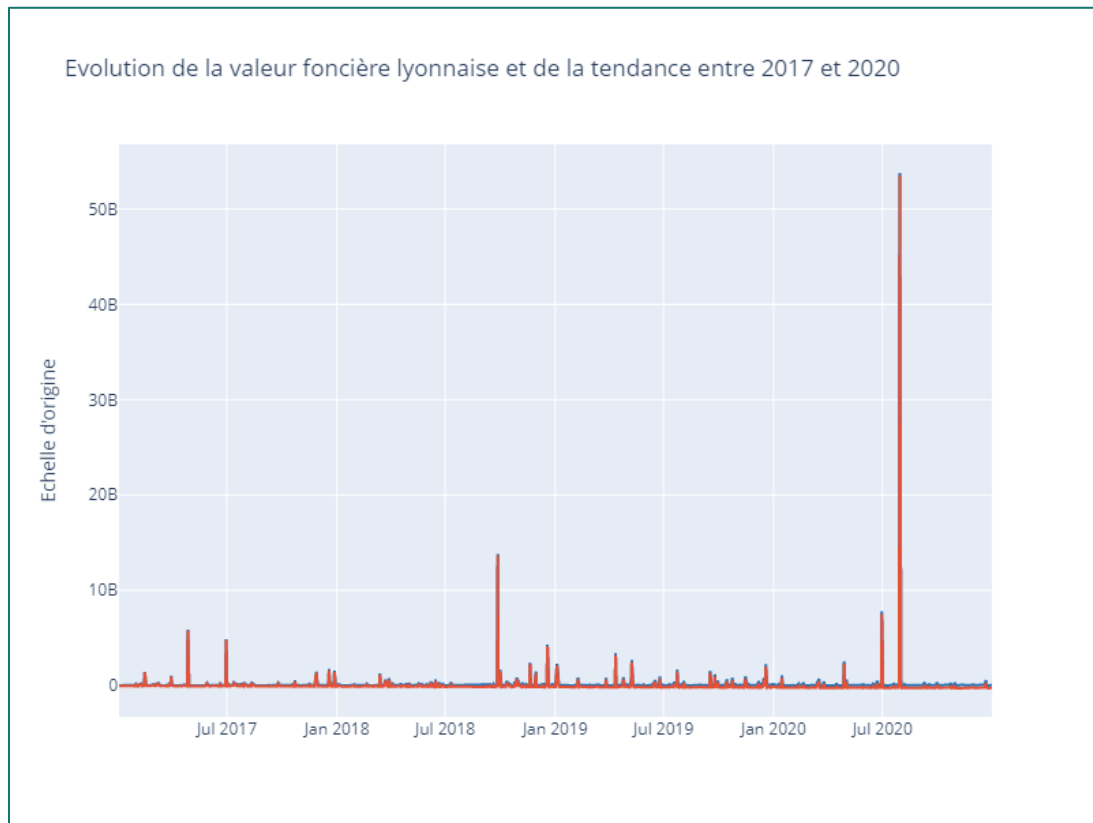
La figure suivante illustre l'évolution de la valeur foncière cumulée de vente entre 2017 et 2020.



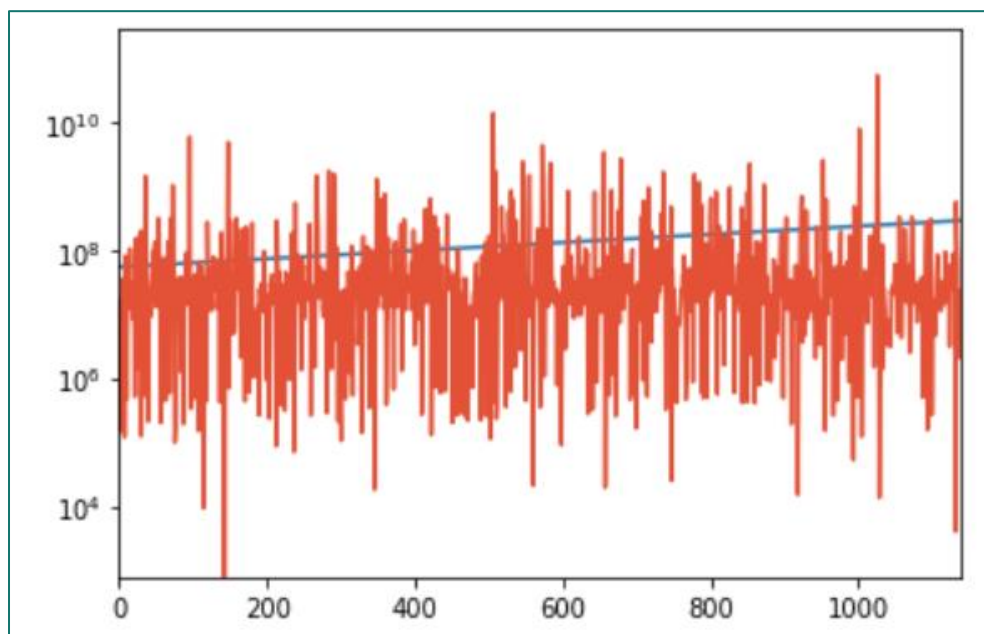
Il apparaît clairement que le premier graphique (échelle d'origine) ne permet pas d'observer de façon correcte les variations dans la valeur foncière lyonnaise. Le deuxième graphique (échelle logarithmique) répond à cette problématique. On remarque tout d'abord que les variations sont majoritairement de l'ordre du million / centaine de millions d'euros. A noter que cette échelle montre que pour le sujet d'étude (les variations de valeurs foncières), une journée avec un delta de plusieurs milliards d'euros n'est pas une abberation (valeur abberante ou *outlier*).

Pour creuser ces variations observées, il a fallu s'intéresser à la tendance. C'est ce qu'illustre le graphique suivant. La courbe en bleu représente toujours nos données ; et la courbe en rouge représente les données après élimination des tendances . Il est frappant dans notre cas que la courbe rouge se superpose

parfaitement à la bleu. Si on regarde dans le détail, on note que la courbe bleue possède des pics plus élevés. On peut penser qu'il y a une tendance à la hausse à la vue de nos données.

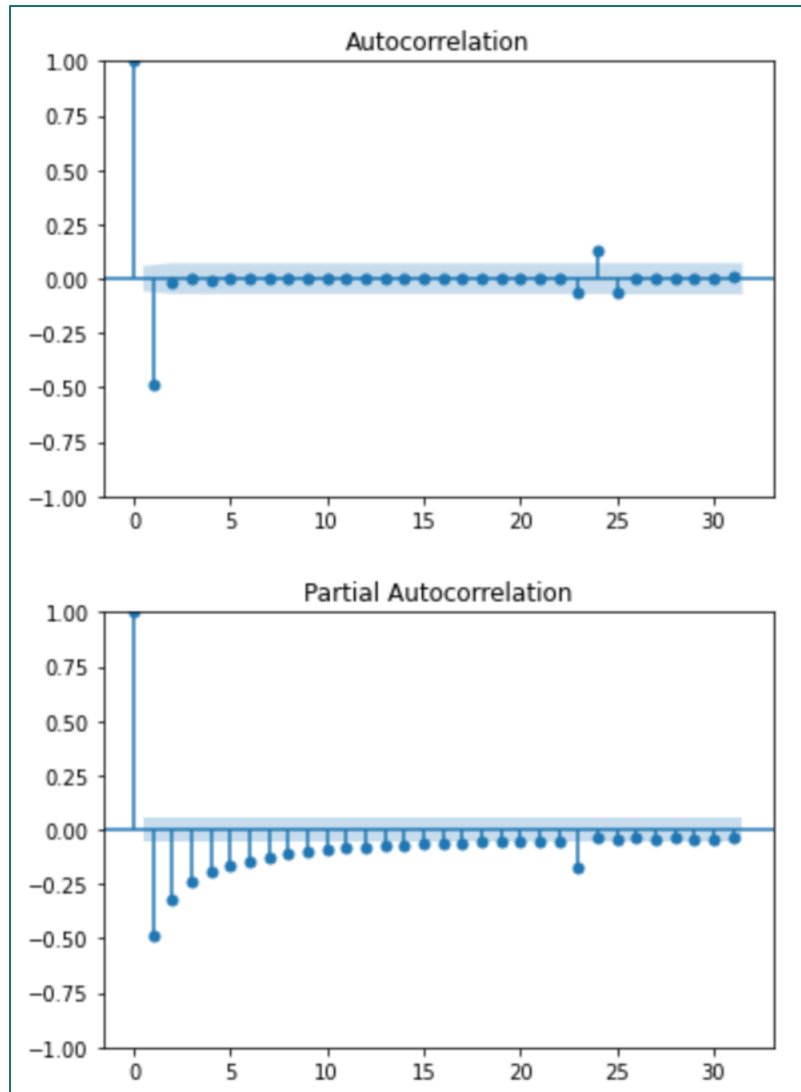


Pour aller plus loin, une tendance par regression linéaire a été opérée à l'aide d'un *abline_plot*. Ce graphique présente l'évolution de la valeur foncière (en rouge) ainsi que la droite de régression associée (en bleu). Cette droite a clairement une pente positive et indique une tendance à la hausse de la valeur foncière.



Ensuite, l'autocorrélation et l'autocorrélation partielle ont été étudiées.

Ces graphiques présentent la valeur de l'autocorrélation (partielle ou non) en fonction d'un lag. Ainsi, l'autocorrélation avec un lag de 0 représente l'autocorrélation de la donnée avec elle-même. Chaque point qui est en dehors de l'aire bleue est considéré comme statistiquement significatif.

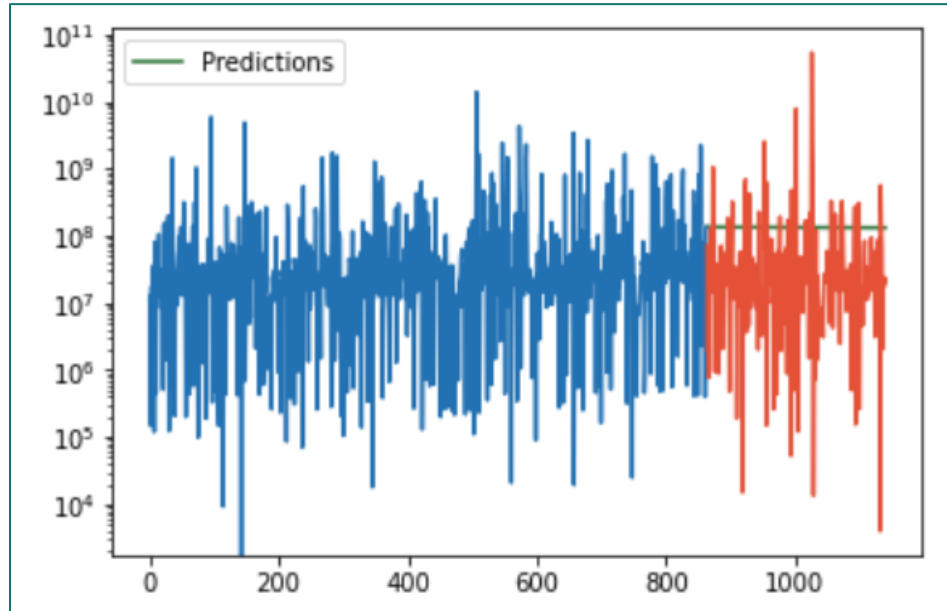


Ainsi, on peut conclure que seul un lag de un jour est pertinent. Il n'y a donc pas de réelle autocorrélation sur la valeur foncière. Quant à l'autocorrélation partielle, celle-ci semble s'étendre jusqu'à un lag d'environ 15 jours (même si les valeurs restent proches de la limite de l'aire jusqu'à lag = 30).

Les données ayant été classées par ordre chronologiques, on avait donc des lags qui correspondaient parfaitement à l'influence de jour - n (n étant la valeur du lag). Il ne semble donc pas avoir de réelle autocorrélation.

Après ces conclusions, l'analyse aurait très bien pu s'arrêter ici. Mais il a été décidé d'essayer de produire une ARIMA pour vérifier ce qu'il a déjà été établi : nos données ne permettent pas dans l'état de faire de prédictions quant aux valeurs foncières à venir.

Voici donc une représentation dans laquelle il a été question d'essayer de prédire les valeurs de 2020 grâce aux valeurs des 3 années précédentes :



Sans surprise, on constate que les prédictions (quasi droite du fait de l'échelle ; affichée en vert) ne sont pas bonnes. On aurait donc pu essayer d'établir un modèle prédictif, mais cela n'aurait pas été pertinent. C'est donc sur cela que ce termine la partie analyse.

CONCLUSION GENERALE

L'étude a permis de mettre en évidence une certaine constance de la valeur foncière à l'échelle globale qu'est la France. L'essentiel des transactions demeurent de type « Vente ».

L'analyse spatialisée des données a permis l'observation d'un effet métropole – à savoir une concentration des transactions dans les grandes villes de France. Cet effet est également plus marqué en 2020 dans les grandes villes de province (en opposition à la capitale) – on peut aisément corrélér cela avec l'effet épidémie de Covid qui a poussé de nombreux parisiens à quitter la capitale à la faveur de ces métropoles (Lyon, Bordeaux, Marseille, Nantes, etc.). Cet exode « urbain » des parisiens a entraîné une hausse du prix moyen des transactions dans les métropoles.

Le focus sur la métropole lyonnaise a confirmé ce constat : la valeur foncière est en hausse. L'essentiel des transactions ont lieu dans les 3^e, 7^e et 8^e arrondissements de Lyon. Cela concorde avec le plan d'urbanisme de la métropole pour la ville. Ces arrondissements étant en pleine métamorphose. On sait par exemple que certains quartiers de Lyon 7^e et 8^e sont en pleine gentrification.

L'analyse temporelle a décelé une saisonnalité dans les transactions : au niveau mensuel, il y a plus de transactions en fin de mois ; au niveau annuel, on observe une baisse correspondant au creux des vacances d'été – août principalement. L'analyse a également permis de mettre en évidence une tendance à la hausse de la valeur foncière. En revanche, l'analyse a échoué à identifier clairement une autocorrélation de nos données ; même si une autocorrélation partielle existe. La construction de modèle de prédiction a donc été abandonnée, faute de pertinence.

Par manque de temps, de compatibilité entre technologies utilisées et de données, l'analyse n'a pas pu être poussée jusqu'aux ambitions initiales. L'idée était aussi de faire des représentations spatialisées de nos données (un peu comme avec Power BI, des cartes choroplètes avec des colorations de surface pour la métropole lyonnaise par exemple) et d'essayer de construire un modèle prédictif sur les valeurs foncières de 2021.

Toutefois, ce projet a répondu à son objectif pédagogique – à savoir entraîner une mise en pratique des connaissances acquises au cours de la formation. Cela va du raisonnement sur l'architecture (simple dans ce cas précis) jusqu'à l'analyse souhaitée (et le code nécessaire pour sa réalisation).

