

Issam Fradi  
Theo Tasset  
Edi Verbovci  
Timothée Dangleterre  
*Master 1 Economie et Finance*  
*Année 2023 - 2024*

## Projet de Data Management

---

### Sommaire

<b>Introduction</b>	Page 1
<b>I. Présentation des données</b>	Page 2
<b>II. Nettoyage et préparation des données</b>	Page 3
<b>III. Analyse géographique du marché immobilier résidentiel</b>	Page 8
<b>IV. Impact des taux d'intérêt sur le marché immobilier résidentiel</b>	Page 17
<b>V. Évaluation des variables pertinentes pour modéliser l'évolution des prix</b>	Page 20
<b>VI. Prédictions des prix par un modèle de Machine Learning</b>	Page 21
<b>Conclusions</b>	Page 22
<b>Annexes</b>	Page 23
<i>Nombre de transactions par département 2018 – 2023</i>	
<i>Résultats du Modèle de Machine Learning :</i>	

## Introduction

---

Le marché immobilier a connu une croissance importante depuis le début du XXI<sup>e</sup> siècle. Les prix ont ainsi été multipliés par près de 2,5 sur l'ensemble du territoire entre 2000 et 2023<sup>1</sup>. Par ailleurs, le nombre de ventes de logements anciens est passé d'environ 800 000 en novembre 2003 à près de 1,2 million en août 2021, avant de diminuer pour se stabiliser à 885 000 en novembre 2023.<sup>2</sup> Cette croissance du secteur a bénéficié de taux d'intérêt bas qui ont soutenu la demande de crédits immobiliers.

Pourtant, malgré cette tendance haussière, l'évolution du marché de l'immobilier est loin d'être linéaire. Les ventes de logements anciens sont ainsi passées de plus de 800 000 par an à 564 000 au moment de la crise des subprimes. Cette décroissance des ventes s'est par ailleurs accompagnée d'une légère inflexion des prix.

Le marché immobilier a ainsi connu plusieurs périodes de contraction caractérisées par un recul des transactions de logements anciens et des constructions de logements neufs. Cela s'explique par le caractère pro-cyclique du secteur : lorsque la conjoncture se retourne, les banques durcissent généralement leurs conditions de prêts, rendant plus difficile l'accès à la propriété.

Néanmoins, on peut se demander si cette dynamique affecte tous les biens immobiliers de la même manière. La crise du COVID s'est ainsi traduite par un recul des transactions immobilières lors des mesures de confinement, mais la montée du télétravail et la volonté de changer de mode de vie ont pu favoriser la demande de logements en milieu rural. La remontée des taux en 2022 a davantage affecté l'immobilier commercial, dont les valorisations ont beaucoup diminué, que l'immobilier résidentiel.

On peut ainsi se demander *si des événements conjoncturels comme le COVID ou la remontée des taux se répercutent sur le marché immobilier ? Et de quelle manière ?*

Notre étude s'appuie sur les fichiers de données *Demandes de valeurs foncières* (DVF) produits par la direction générale des Finances publiques qui recensent les transactions immobilières survenues au cours des cinq dernières années (de mi-2018 à mi-2023).

Dans un premier temps, nous allons présenter la structure du jeu de données. Nous expliquerons ensuite nos choix en matière de nettoyage et de préparation des données, préalables à la réalisation d'une analyse géographique des effets du COVID et de la hausse des taux sur l'immobilier résidentiel. Enfin, nous évaluerons les variables pertinentes pour modéliser l'évolution des prix immobiliers et présenterons un modèle de prévision.

---

<sup>1</sup> <https://france-inflation.com/evolution-immobilier-ancien.php>

<sup>2</sup> <https://www.notaires.fr/fr/article/marche-de-limmobilier-tendances-et-evolutions-des-prix-de-limmobilier-janvier-2024>

## I. Présentation des données

Les fichiers *Demandes de valeurs foncières* sont présents sur le site data.gouv (<https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/#/resources>).

Il y a 6 fichiers de données, un pour chaque année entre 2018 et 2023. Ces fichiers sont présents au format « txt » et utilisent le caractère | comme séparateur. Ils se présentent de la manière suivante (extrait du fichier de données de 2019) :

```
*Identifiant de document|Reference document|1 Articles CGI|2 Articles CGI|3 Articles CGI|4 Articles CGI|5 Articles CGI|No disposition|Date mutation|Nature mutation|Valeur foncière|No voie|B/T/Q|Type de
voie|Code voie|Voie|Code postal|Commune|Code département|Code commune|Prefixe de section|Section|No plan|No Volume|1er lot|Surface Carrez du 1er lot|2eme lot|Surface Carrez du 2eme lot|3eme lot|Surface
Carrez du 3eme lot|4eme lot|Surface Carrez du 4eme lot|5eme lot|Surface Carrez du 5eme lot|Nombre de lots|Code type local|Type local|Identifiant local|Surface réelle bâti|Nombre pieces principales|Nature
culture|Nature culture speciale|Surface terrain
|||||000001|04/01/2019|Vente|37220,00|26||RUE|2730|DE MONTMOLON|1000|BOURG-EN-BRESSE|01|53||AI|298||8|19,27|||||||1|2|Appartement||20|1||| | |
|||||000001|04/01/2019|Vente|185100,00|22||B|RUE|1650|GEN DELESTRAINT|1000|BOURG-EN-BRESSE|01|53||AM|95||17|||||||1|3|Dépendance||0|0|||
|||||000001|04/01/2019|Vente|185100,00|22||RUE|1650|GEN DELESTRAINT|1000|BOURG-EN-BRESSE|01|53||AM|95||137||154|61,51|||||||2|2|Appartement||62|3|||
|||||000001|08/01/2019|Vente|209000,00|3||RUE|0043|DES CHAMPAGNES|1160|PRIAY|01|314||E|1676|||||||0|1|Maison||90|4|S||940
|||||000001|07/01/2019|Vente|134900,00|5||LOT|A003|LE BIOLAY|1370|SAINT-ETIENNE-DU-BOIS|01|350||AA|11|||||||0|1|Maison||101|5|S||490
|||||000001|03/01/2019|Vente|192000,00|165||ALL|0445|DES LIBELLULES|1340|ATTIGNAT|01|24||AI|94|||||||0|1|Maison||88|4|S||708
|||||000001|08/01/2019|Vente|45000,00|9||RTE|0001|DU VIADUC|1250|CIZE|01|106||A|06|||||||0|1|Maison||39|2|S||631
|||||000001|08/01/2019|Vente|45000,00|||B017|SUR LA LATIE|1250|CIZE|01|106||A|975|||||||0|||||120
|||||000001|04/01/2019|Vente|65000,00|50||RUE|1240|DOC NODET|1000|BOURG-EN-BRESSE|01|53||AL|3||317|||||||1|3|Dépendance||0|0||| | | |
|||||000001|04/01/2019|Vente|65000,00|50||RUE|1240|DOC NODET|1000|BOURG-EN-BRESSE|01|53||AL|3||17||33|67,78|||||||2|2|Appartement||69|3|||
|||||000001|04/01/2019|Vente|37000,00|68||AV|2400|DE MACON|1000|BOURG-EN-BRESSE|01|53||AB|199||107|39,52|16|||||||2|2|Appartement||39|1|||
|||||000001|08/01/2019|Vente|70,00|||B037|EN CHEVRET|1250|TOSSAT|01|422||D|1312|||||||0|||||AB|32
|||||000001|08/01/2019|Vente|70,00|||B037|EN CHEVRET|1250|TOSSAT|01|422||D|1316|||||||0|||||AB|31
|||||000001|07/01/2019|Vente|116000,00|1724||RTE|0105|DE MONTLEGER|1560|MANTENAY-MONTLIN|01|230||ZH|60|||||||0|1|Maison||100|1|S||2103
|||||000001|08/01/2019|Vente|266000,00|||B013|CUET|1340|MONTREVEL-EN-BRESSE|01|266||C|714|||||||0|||||S|184
|||||000001|08/01/2019|Vente|266000,00|||B013|CUET|1340|MONTREVEL-EN-BRESSE|01|266||C|715|||||||0|||||S|8
|||||000001|08/01/2019|Vente|266000,00|1250||RTE|0037|DE CUET|1340|MONTREVEL-EN-BRESSE|01|266||C|716|||||||0|3|Dépendance||0|0|S||1067
|||||000001|08/01/2019|Vente|266000,00|1250||RTE|0037|DE CUET|1340|MONTREVEL-EN-BRESSE|01|266||C|716|||||||0|1|Maison||115|4|S||1067
|||||000001|08/01/2019|Vente|266000,00|||B013|CUET|1340|MONTREVEL-EN-BRESSE|01|266||C|718|||||||0|||||S|39
|||||000001|12/01/2019|Vente|138100,00|181||RTE|0150|DE VERNOUX|1560|COURTES|01|128||WA|34|||||||0|1|Maison||99|4|S||1500
|||||000001|12/01/2019|Vente|138100,00|181||RTE|0150|DE VERNOUX|1560|COURTES|01|128||WA|34|||||||0|3|Dépendance||0|0|S||1500
|||||000001|12/01/2019|Vente|7570,00|||B042|CROCU|1560|SAINT-TRIVIER-DE-COURTES|01|388||A|327|||||||0|||||T|8370
|||||000001|12/01/2019|Vente|7570,00|||B102|LE PIOCHAY|1560|SAINT-TRIVIER-DE-COURTES|01|388||A|333|||||||0|||||T|7100
|||||000001|12/01/2019|Vente|7570,00|||B102|LE PIOCHAY|1560|SAINT-TRIVIER-DE-COURTES|01|388||A|335|||||||0|||||J|275
|||||000001|12/01/2019|Vente|7570,00|||B075|GRANDVAL|1560|SAINT-TRIVIER-DE-COURTES|01|388||A|351|||||||0|||||P|4310
```

En moyenne, chaque fichier comprend environ 3 millions d'enregistrements pour 43 champs de données. Parmi ces champs, certains ne sont pas restitués (ex : Code service CH, article CGI, ...).<sup>3</sup>

**Numéro de disposition** : lorsqu'un acte comprend plusieurs mutations, chacune d'entre elles se voit attribuer un numéro de disposition pour pouvoir les distinguer.

**Date de mutation** : date de signature de l'acte notarié au format JJ/MM/AAAA

**Nature de la mutation** : nature de la transaction qui a eu lieu. Il peut s'agir d'une vente, d'une vente de terrain à bâtir, d'une vente en l'état futur d'achèvement, d'une adjudication, d'une expropriation ou d'un échange.

**Valeur foncière** : Montant de la transaction dans le cadre d'une mutation à titre onéreux. Cette valeur tient compte des frais d'agence, s'ils sont à la charge du vendeur, et de la TVA. Les frais de notaires, les frais d'agences à la charge de l'acquéreur et la valeur des biens meubles stipulés dans l'acte de vente ne sont en revanche pas inclus.

<sup>3</sup> Source de la description des champs : <https://static.data.gouv.fr/resources/demandes-de-valeurs-foncieres/20221017-153319/notice-descriptive-du-fichier-dvf-20221017.pdf>

Plusieurs données cadastrales : Code de département / Code commune, le libellé de la commune et le code postal, le type de voie sur lequel se situe la parcelle (rue, avenue, ...) ainsi que le numéro de voie et son Code Rivoli.

Lot : présentation des cinq premiers lots de copropriété rattachés à l'immeuble au moment de la mutation, ainsi que du nombre total de lots pour cet immeuble. Un lot de copropriété comprend une partie privative et une quote-part de partie commune.

Type local : Nature du bien vendu. Il peut s'agir d'une maison, d'un appartement, d'un local industriel et commercial ou assimilés ainsi que d'une dépendance. La notion de dépendance renvoie à des constructions accessoires au bâtiment principal (ex : piscine, jardin d'hiver, garages, caves, ...).

Surface réelle bâtie : surface mesurée au sol entre les murs ou séparations et arrondie au mètre carré inférieur.

Nombre de pièces principales : nombre de pièces principales du bien immobilier.

Code nature culture et Nature culture spéciale : Description des propriétés non bâties rattachées à la mutation et incluant les terres, les prés, les vignes, ...

Surface terrain : surface du terrain cédé.

## II. Nettoyage et préparation des données

---

### 1) Importation des données

La première étape consiste à importer les données des six fichiers et à les agréger afin de conserver un unique tableau regroupant les informations pour toutes les années de l'étude. Les fichiers « txt » ont donc été stockés dans un même répertoire. Le code demande à l'utilisateur de sélectionner le dossier des fichiers via la fonction `filedialog.askdirectory()` de la librairie `tkinter`. Le répertoire de travail est ensuite redéfini grâce à la fonction `chdir` de la librairie `os`. Une boucle permet ensuite d'importer et d'agréger les données, via les fonctions `read_csv` et `concat` du module `pandas`. L'utilisation de l'argument `Low_memory = False` permet de garantir un typage unique pour chaque champ.

L'importation permet d'obtenir le tableau suivant :

	Identifiant de document	Reference document	1 Articles CGI	2 Articles CGI	3 Articles CGI	4 Articles CGI	5 Articles CGI	No disposition	Date mutation	Nature mutation	...	Surface Carrez du 5eme lot
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	07/01/2020	Vente	...	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	02/01/2020	Vente	...	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	02/01/2020	Vente	...	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	02/01/2020	Vente	...	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	02/01/2020	Vente	...	NaN

rows x 43 columns

Ce tableau comprend près de 20 millions de lignes pour 43 colonnes. L'étude du typage des colonnes et la visualisation des données font apparaître plusieurs problèmes :

- Certaines colonnes n'ont pas le bon type de données. C'est notamment le cas des dates enregistrées au format « object » (= chaîne de caractères) ;
- Il y a de nombreuses données manquantes.

## 2) Gestion des données manquantes

Le premier retraitement effectué concerne les données manquantes. Il est nécessaire de déterminer la quantité de cellules vides pour chaque colonne, via la fonction `isna().sum()`. Cette fonction révèle que certaines colonnes sont laissées entièrement vides. C'est cohérent puisqu'il est précisé que certaines données ne sont pas reportées dans le fichier, à l'instar des articles CGI (cf : *data.gouv*). En dehors des colonnes vides, il peut y avoir beaucoup de données manquantes pour certains champs (ex : colonne *B/T/Q* avec 18 878 772 cellules vides).

Nous avons donc construit la fonction `supp_vid` afin de ne conserver que les colonnes au moins à moitié remplies. Ce choix est motivé par la taille de l'échantillon : conserver une colonne avec plus de 50% de données manquantes nécessiterait de retraiter partiellement près de 9 millions de données. Cela pourrait biaiser les résultats obtenus.

Ce premier traitement réduit le jeu de données à un tableau d'environ 20 millions de lignes et 21 colonnes contre 43 initialement. Les colonnes conservées pour la suite de l'étude sont :

```
Index(['No disposition', 'Date mutation', 'Nature mutation', 'Valeur fonciere',
      'No voie', 'Type de voie', 'Code voie', 'Voie', 'Code postal',
      'Commune', 'Code departement', 'Code commune', 'Section', 'No plan',
      'Nombre de lots', 'Code type local', 'Type local',
      'Surface reelle bati', 'Nombre pieces principales', 'Nature culture',
      'Surface terrain'],
      dtype='object')
```

## 3) Sélection des colonnes pertinentes

Parmi ces colonnes, certaines ne semblent pas apporter d'informations supplémentaires pertinentes pour évaluer les effets de la crise sanitaire et de la remontée des taux sur le

marché immobilier à la maille départementale et régionale. C'est le cas du *nombre de lots* ou du *numéro de plan*. Les références géographiques aux voies auraient présenté un intérêt dans le cadre d'une étude à une maille plus fine (ex : évaluer les effets de la remontée des taux sur les ventes de logements entre deux rues d'une même ville, ...) mais ce n'est pas l'objectif poursuivi ici. Les colonnes conservées sont ainsi :

```
['Date mutation',
 'Valeur fonciere',
 'Code departement',
 'Type local',
 'Surface reelle bati',
 'Nombre pieces principales',
 'Nature mutation',
 'Surface terrain']
```

Ce premier retraitement a permis d'obtenir un tableau comprenant les colonnes au moins à moitié remplies jugées pertinentes pour l'analyse. Avant de retraiter les valeurs plus en détail, il convient de modifier le typage de certains champs qui sont, pour la majorité d'entre eux, renseignés sous forme de chaîne de caractères (type « object ») :

Date mutation	object
Valeur fonciere	object
Code departement	object
Type local	object
Surface reelle bati	float64
Nombre pieces principales	float64
Nature mutation	object
Surface terrain	float64
dtype:	object

La colonne *Date mutation* est ainsi convertie en date à l'aide de la fonction *to\_datetime()* de la librairie *pandas*.

Les données de la colonne *Valeur fonciere* sont également transformées en données numériques. Un retraitement est par ailleurs opéré pour convertir les valeurs au format numérique français.

Maintenant que les colonnes ont le type de données adéquat, il convient de s'intéresser au cas des valeurs manquantes restantes. En effet, parmi les colonnes conservées il reste des données non renseignées :

Date mutation	0
Valeur fonciere	218532
Code departement	0
Type local	9269558
Surface reelle bati	9284102
Nombre pieces principales	9284102
Nature mutation	0
Surface terrain	7024509
dtype:	int64

Nous constatons que certaines données ne sont pas renseignées pour la valeur foncière (environ 200 000). Compte tenu de l'importance de ce champ dans notre étude, les lignes pour lesquelles il n'y a pas de valeur foncière sont supprimées.

Dans les autres cas, les valeurs manquantes sont remplacées par la valeur médiane, dans le cas des données numériques, et le mode dans le cas des données textuelles.

#### 4) Gestion des incohérences

Un contrôle est effectué sur les données obtenues afin de vérifier l'absence d'incohérences. En particulier, la première et la dernière date de l'échantillon sont bien contenues dans l'intervalle de temps S2 2018 – S1 2023.

Par ailleurs, il n'y a aucune ligne pour laquelle la surface réelle bâtie est supérieure à la surface du terrain.

#### 5) Gestion des doublons

Il peut arriver que certaines lignes fassent doublon dans le dataframe. Les lignes identiques sont donc supprimées via la méthode *drop\_duplicates()*. Le tableau de données contient désormais environ 17 millions de lignes.

Le tableau de données contient plusieurs types de biens immobiliers dans la colonne *Type local* :

Date mutation	
Type local	
Appartement	2757573
Dépendance	10026920
Local industriel, commercial ou assimilé	588971
Maison	3551917

#### 6) Retraitement selon le type de logement

L'essentiel des transactions concerne des *dépendances*. Or, ces dernières sont peu représentatives du marché immobilier puisqu'il s'agit de constructions accessoires au bâtiment principal tels qu'une piscine, un garage, ... Il y a par ailleurs moins d'un million de lignes associées à des transactions portant sur des locaux industriels. Afin de faciliter une

meilleure comparabilité des données, seules les transactions associées aux appartements et aux maisons sont conservées.

Ce retraitement fait apparaître un problème dans la manière dont les doublons ont été retraités. En effet, seules les lignes exactement identiques ont été supprimées du dataframe. Pourtant, il semble y avoir plusieurs lignes associées à une même date de mutation et une même valeur foncière :

Date mutation	Valeur foncière	Code postal	Commune	Code département	Code commune	Type local	Surface réelle bati	Nombre pièces principales	Nature mutation	Surface terrain
2020-01-09	72000.0	01270	COLIGNY	01	108	Maison	35.0	2.0	Vente	381.0
2020-01-06	180300.0	01000	BOURG-EN-BRESSE	01	53	Maison	75.0	4.0	Vente	525.0
2020-01-06	54800.0	01000	BOURG-EN-BRESSE	01	53	Appartement	32.0	1.0	Vente	612.0
2020-01-03	350750.0	01000	SAINT-DENIS-LES-BOURG	01	344	Maison	201.0	7.0	Vente	1497.0
2020-01-03	350750.0	01000	SAINT-DENIS-LES-BOURG	01	344	Maison	201.0	7.0	Vente	1267.0

Cette situation se produit lorsque plusieurs lignes se réfèrent à la même transaction. Une solution serait donc de ne conserver qu'une ligne en supprimant les doublons à partir des arguments *Date mutation*, et *Valeur foncière*. Cette approche ne semblait néanmoins pas satisfaisante car conserver arbitrairement la première ligne pourrait aboutir à un cas où un bien avec une valeur foncière élevée est associée à une surface terrain, une surface réelle bâtie et un nombre de pièces principales très faible (la situation s'est présentée).

Pour éviter cet écueil, les données ont été agrégées sur la base de l'ensemble des colonnes à l'exception de *Surface terrain*, *Surface réelle bati* et *Nombre pièces principales*. Ainsi, lorsqu'une transaction est associée à une unique ligne, les valeurs de ces champs seront reportées et dans le cas où plusieurs lignes sont associées à une même transaction, c'est la valeur médiane, calculée sur la base de l'ensemble des lignes associées à cette transaction, qui sera reportée pour ces trois champs.

Une fois le cas des quasi-doublons traité, un filtre est opéré à partir de la typologie des transactions présentes dans le jeu de données :

	Date mutation
Nature mutation	
Adjudication	16045
Echange	7718
Expropriation	158
Vente	5038379
Vente en l'état futur d'achèvement	139883
Vente terrain à bâtir	1394



## **7) Retraitement selon le type de transaction**

Seules les ventes et les ventes futures en l'état d'achèvement sont conservées. En effet, ces transactions sont largement majoritaires dans le jeu de données et correspondent aux transactions « classiques » qui sont à notre sens les plus représentatives du marché immobilier. Par ailleurs, seules les lignes ayant une surface réelle bâtie positive sont conservées afin d'éviter qu'un terrain puisse être mal répertorié.

## **8) Gestion des données extrêmes et standardisation**

Pour terminer la préparation des données, les valeurs extrêmes (premier et dernier quantile) associées à la *Valeur foncière* sont supprimées et les données sont standardisées. Enfin, certaines colonnes sont ajoutées : une colonne mois, année, une clé d'identification (nombre) afin de pouvoir plus facilement distinguer les lignes lors de l'entraînement du modèle et une colonne répertoriant le prix rapporté à la surface (Prix au m<sup>2</sup>). Le calcul des prix au m<sup>2</sup> par département, avec les données non standardisées, ont permis de vérifier la cohérence des prix obtenus.

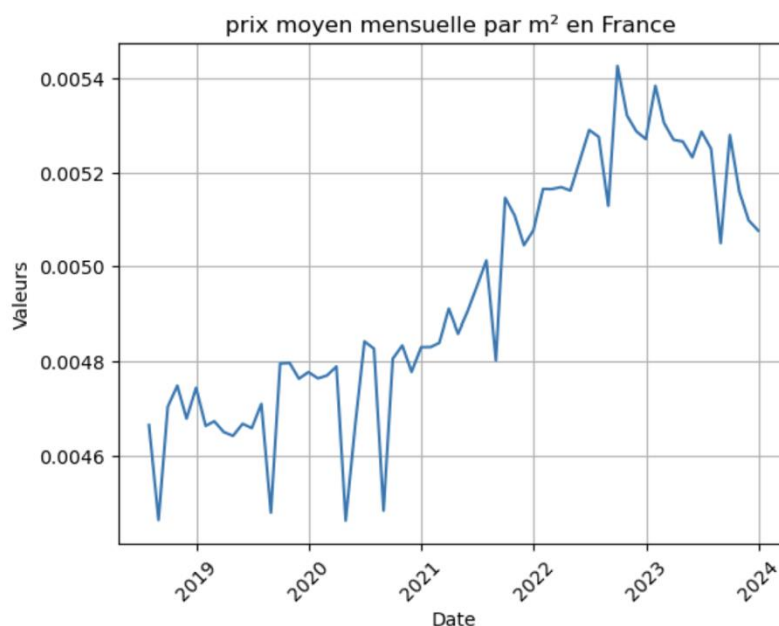
## **III. Analyse géographique du marché immobilier résidentiel français**

---

Dans un premier temps, l'analyse se concentre sur les tendances du marché immobilier résidentiel français dans son ensemble. Pour cela, nous étudions certaines métriques telles que l'évolution du prix et du nombre de transactions sur l'ensemble du territoire. Au-delà des dynamiques du territoire, nous étudions ces tendances à des mailles plus fines, régionales et départementales, afin de voir s'ils existent d'éventuelles spécificités. Enfin, nous ferons un point sur les transformations du marché provoquées par le COVID.

### **1) Évolution du prix au mètre carré moyen**

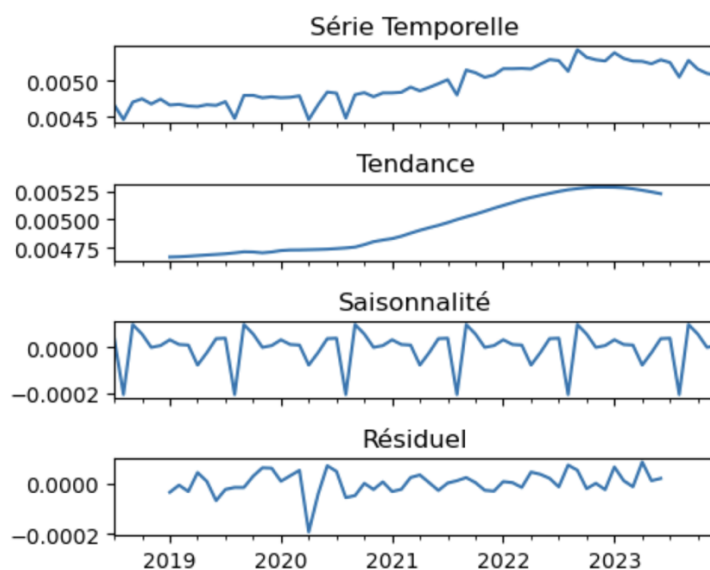
D'après le premier graphique représentant la série temporelle des prix au mètre carré de l'immobilier en France, nous observons une hausse du prix du mètre carré entre 2018 et 2023 (environ 500 € avec des données non normalisées) :



La relance économique post COVID a entraîné une forte augmentation du prix moyen au mètre carré. Néanmoins, cette tendance s'est stabilisée avec le retour de l'inflation en 2022. Depuis un an, les prix baissent légèrement pour revenir au niveau de début 2022. Au global, la tendance des prix reste largement positive sur 5 ans.

Au-delà de la tendance globale des prix, nous vérifions si les prix suivent un schéma saisonnier. On peut en effet penser que les ménages achètent plus particulièrement à certaines périodes de l'année. Pour le vérifier, nous décomposons la série temporelle des prix afin d'isoler une éventuelle composante saisonnière.

## 2) Étude de la saisonnalité des prix immobiliers

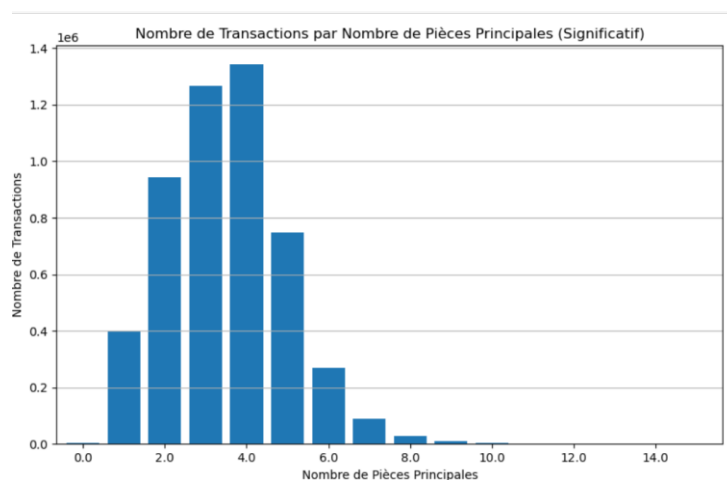


On observe des variations régulières et répétitives des prix immobiliers chaque année. Il y a ainsi une forte demande au début du printemps qui peut s'expliquer par le fait que les ménages préfèrent emménager avant la nouvelle année scolaire. En effet, l'acte de vente est généralement signé dans un horizon de 3 mois après la signature du compromis de vente. Par ailleurs, le retour des beaux jours influence le moral des ménages et renforce ainsi cette observation.

S'en suivent les mois d'été, où l'activité a tendance à significativement chuter et se répercuter défavorablement sur les niveaux de prix du marché, ce qui est cohérent avec les résultats obtenus dans notre étude.

À la fin du deuxième trimestre, une forte reprise de l'activité est observée. En effet, les visites de propriétés sont généralement plus agréables en raison d'une meilleure luminosité naturelle. Cela peut augmenter leur attractivité, et par conséquent, leur prix de vente. Enfin, une légère baisse des prix est observée entre la fin du troisième trimestre et les mois de janvier/février alors que les ménages et les professionnels du secteur sont moins disponibles pour assister à des visites de biens. De plus, les températures plus basses et les vacances de Noël rendent les acquéreurs potentiels moins aptes à l'achat de biens.

### 3) Relation entre le nombre de transactions et le nombre de pièces principales

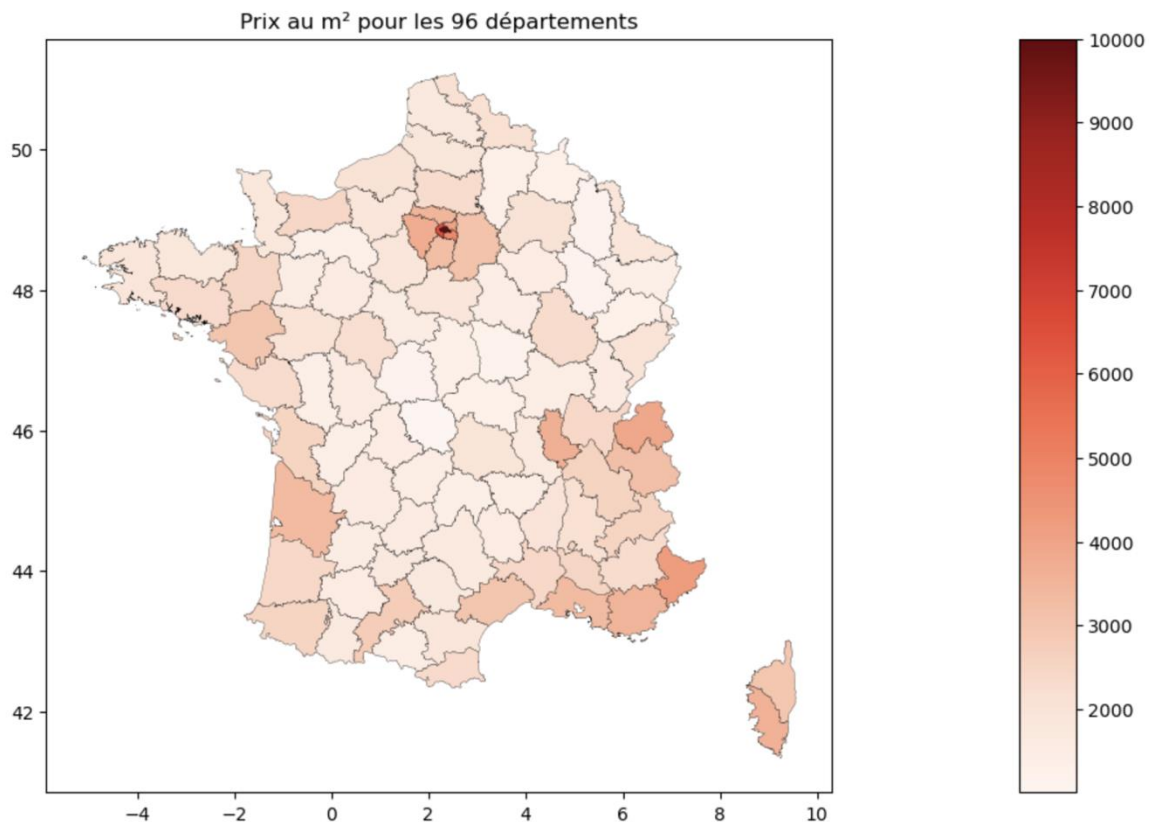


L'histogramme ci-dessus décrit le nombre de transactions immobilières en France, regroupées en fonction du nombre de pièces principales des biens vendus. Les biens immobiliers avec 3 et 4 pièces principales représentent la majorité des ventes sur le marché immobilier.

Plusieurs raisons peuvent expliquer cette observation. Les appartements ou maisons avec 3 ou 4 pièces sont souvent adaptés à la taille moyenne des familles. En effet, ce type de bien comprend généralement une chambre parentale, une ou deux chambres pour enfants ainsi qu'une pièce à vivre. Deuxièmement, les biens avec 3 ou 4 pièces sont souvent relativement moins chers au mètre carré que les plus petits logements de type studio ou une 2 pièces. Ils permettent de conserver un équilibre entre habitable et coût. De ce fait, ces biens sont considérés comme de bon investissement par les investisseurs immobiliers du fait qu'ils

constituent des actifs plus liquides sur ce marché. Nous pouvons également ajouter que les appartements de type F3 ou F4 représentent le format le plus répandu dans les zones urbaines où l'espace est plus restreint. Ce type de logement répond donc aux contraintes d'espace et au style de vie des citadins, qui représentent une grande majorité des habitants en France.

#### 4) Visualisation du niveau des prix sur la carte de France



4

Nous entreprenons une analyse des niveaux de prix à l'échelle départementale en France afin de discerner d'éventuelles spécificités géographiques. Notre objectif est d'observer les prix sur la carte de France et de déceler des tendances selon les régions et leurs caractéristiques. Cette observation nous permettra de contextualiser le prix d'une région en fonction de ses particularités. Par exemple, nous notons que Paris est le département le plus onéreux, tout comme ses départements voisins. De même, les départements frontaliers de la Suisse affichent des prix élevés. En revanche, la "diagonale du vide" présente les prix les plus bas.

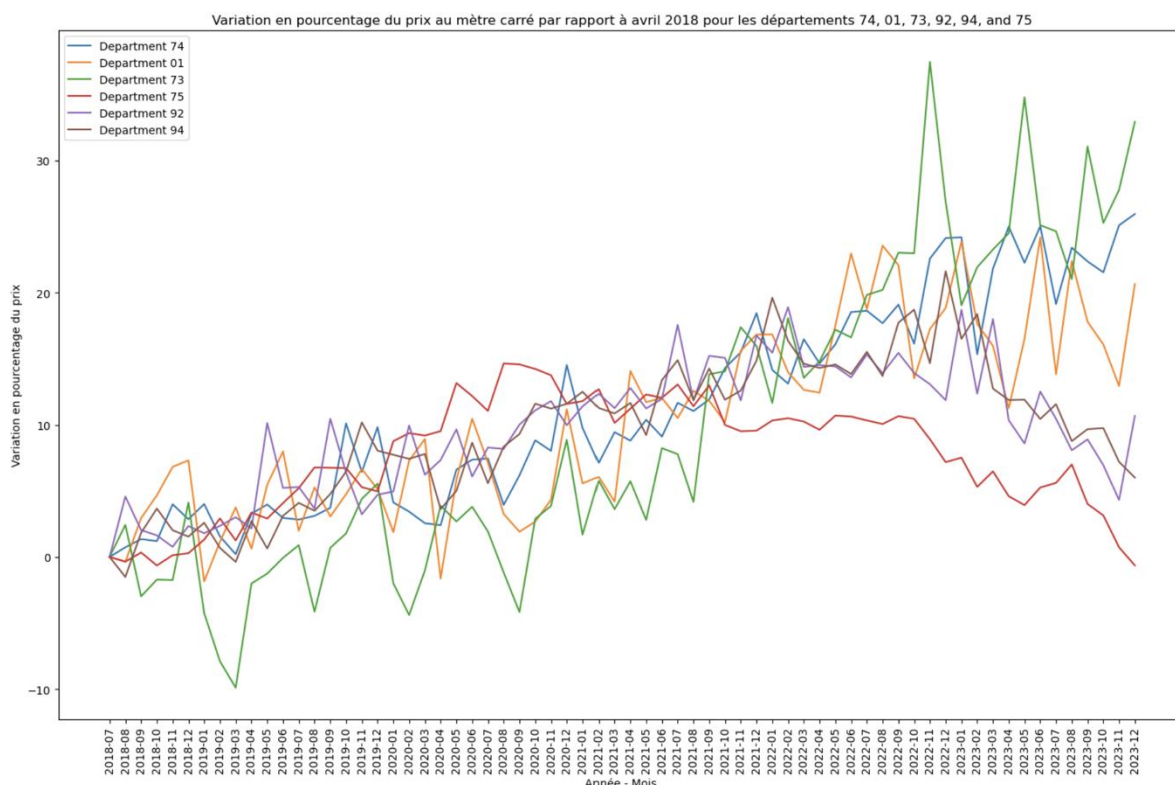
Cette première observation constitue une étape cruciale pour orienter nos futures analyses. De ce fait, elle nous conduit à entreprendre une étude plus approfondie où nous distinguons

---

<sup>4</sup> Trois départements ne sont pas affichés : il s'agit de la Moselle, du Bas-Rhin et du Haut-Rhin pour lesquels aucune donnée n'est reportée dans la base gouvernementale.

les taux de croissance en base « 0 » entre l'Île-de-France et les départements frontaliers par rapport aux prix, en raison de leur niveau élevé.

## 5) Comparaison des prix immobiliers entre les départements frontaliers et l'Île de France



Le graphique illustre comment le prix de l'immobilier a évolué depuis avril 2018 dans certains départements. En Haute-Savoie, dans l'Ain et en Savoie, on note une nette montée des prix. Cette tendance pourrait s'expliquer par la proximité avec la Suisse, reconnue pour son dynamisme économique, attirant ceux qui, travaillant en Suisse, préfèrent résider dans ces départements français où la vie est souvent moins onéreuse. La qualité de vie offerte par ces régions n'est pas à négliger non plus.

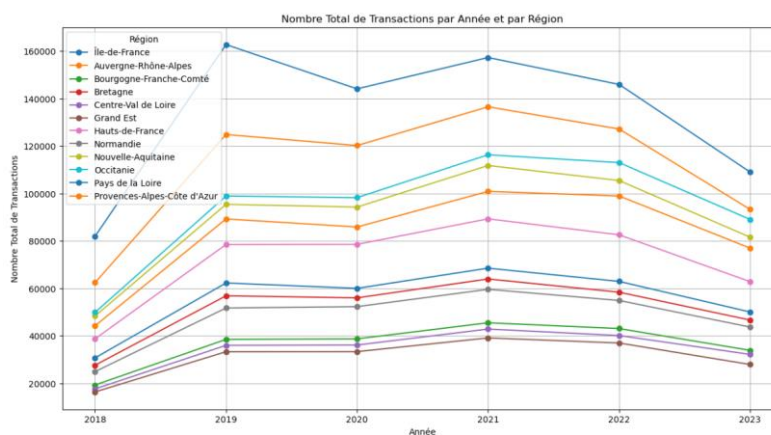
À Paris, la situation est plus nuancée : les prix grimpent moins vite, avec des moments où ils fléchissent même un peu. Cela pourrait témoigner de la tension sur le marché immobilier parisien, où l'accessibilité financière devient un casse-tête pour beaucoup. Les limitations en termes d'espace, et un marché à saturation, pourraient également expliquer une certaine stabilisation des prix, renforcée par les changements de mode de vie comme l'essor du télétravail qui incite à chercher plus d'espace ailleurs.

Quant aux Hauts-de-Seine et au Val-de-Marne, leur courbe de prix suit une ascension modérée. Ces départements, bien que proches de Paris et tout aussi recherchés, sont peut-

être perçus comme des alternatives plus abordables que la capitale, offrant un compromis intéressant entre proximité de la ville lumière et un coût de la vie légèrement moins élevé.

Nous pouvons alors vérifier si la région Auvergne-Rhône-Alpes, englobant les départements frontaliers à la Suisse, a connu une hausse du nombre de transactions immobilières sur les 5 dernières années. Analyser l'évolution du nombre de transactions par région nous permettra alors de vérifier si cette région a été plus primée que les autres malgré des prix se rapprochant de ceux de région parisienne. Additionnellement, nous analyserons l'impact de la pandémie et la hausse des taux d'intérêts sur la capacité des ménages à investir dans l'immobilier.

## 6) Évolution du nombre de transactions par région



Le graphique dépeint l'activité immobilière en France, soulignant les variations régionales selon le nombre de transactions annuelles. Globalement, nous pouvons d'ores-et-déjà observer une contraction du nombre des transactions depuis fin 2021. Cela peut s'expliquer par une montée en flèche de l'inflation qui a obligé la banque centrale européenne à relever ses taux d'intérêts directeurs à plusieurs reprises jusqu'au seuil de 4.5%. Le coût du crédit immobilier a ainsi drastiquement augmenté, réduisant l'enthousiasme des agents économiques et limitant leur capacité d'emprunt.

C'est le département de l'Ile-de-France, notamment Paris, qui connaît le fléchissement le plus important du nombre de transactions immobilières annuelles. C'est probablement la cause d'un coût de la vie qui n'a fait qu'augmenter ces dernières années et qui a encore battu des records avec le retour de l'inflation.

Les régions comme l'Auvergne-Rhône-Alpes et la Nouvelle-Aquitaine, après un pic d'intérêt lié à leur attrait touristique, économique et à un meilleur coût de vie, témoignent d'une certaine stabilisation. Elles ont cependant aussi connu une chute du nombre de transactions immobilières depuis 2022, témoignant d'un sentiment global de perte de pouvoir d'achat et d'une incapacité d'emprunter des ménages.

Avant le début de la pandémie, la Bretagne enregistrait également une hausse significative du nombre de transactions immobilières. Cependant, le retour de l'inflation et donc la hausse du coût de la vie ont freiné cette tendance. La Bretagne connaît aujourd'hui les mêmes difficultés et peine à retrouver sa croissance d'avant COVID.

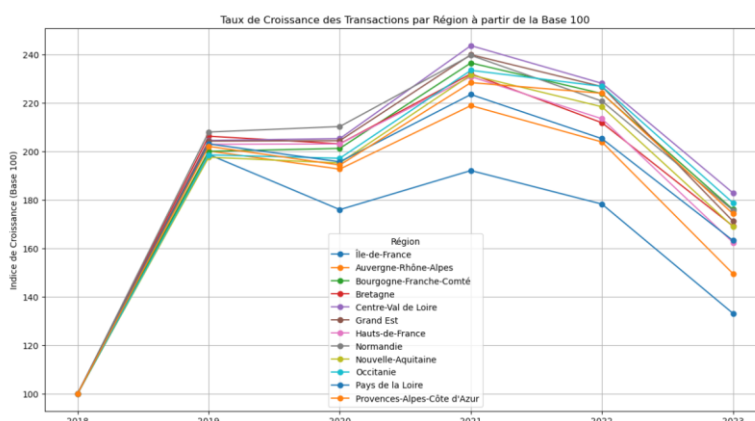
Ces tendances régionales reflètent la volatilité des cycles de crédit, où la hausse des taux d'intérêts poussée par la hausse de l'inflation perturbe les volontés d'emprunt des ménages français.

Les 3 régions avec la moindre augmentation du nombre de transactions :

	Taux de Croissance
Île-de-France	33.083129
Auvergne-Rhône-Alpes	49.466380
Hauts-de-France	62.350752

Les 3 régions avec la plus forte augmentation du nombre de transactions :

	Taux de Croissance
Bourgogne-Franche-Comté	75.963188
Occitanie	78.685499
Centre-Val de Loire	82.780779



L'examen approfondi de l'évolution du nombre de transactions immobilières par région, en se fondant sur un indice de base 100, révèle des dynamiques de marché différenciées à travers le territoire français. En particulier, les régions telles que l'Île-de-France et l'Auvergne-Rhône-Alpes, reconnues pour la vigueur de leur marché immobilier et le niveau élevé de leurs prix, ont enregistré des croissances comparativement modestes. L'Île-de-France, en particulier, avec un accroissement d'environ 33 %, affiche la progression la plus contenue, ce qui pourrait indiquer une certaine maturité, voire une saturation du marché immobilier dans cette région. Cette hypothèse est renforcée par la performance de l'Auvergne-Rhône-Alpes qui, bien que plus dynamique, reste limitée à une augmentation de 50 %. Ces observations pourraient témoigner d'une activité de transaction qui atteint un plateau, potentiellement entravée par le coût élevé de l'accès à la propriété.

À l'inverse, des régions affichant des niveaux de prix immobiliers plus modérés, notamment la Bourgogne-Franche-Comté et le Centre-Val de Loire, émergent comme des marchés en forte expansion, avec des taux de croissance de 76 % et 83 % respectivement. Ce constat pourrait s'expliquer par une attractivité accrue de ces zones qui offrent un rapport qualité-prix

avantageux, susceptible de séduire de nouveaux acquéreurs. Ces chiffres reflètent possiblement un transfert d'intérêt vers des marchés proposant un immobilier accessible, dans un contexte où la hausse générale des prix immobiliers pourrait contraindre les budgets des ménages.

L'analyse met en lumière un contraste remarquable entre les marchés régionaux : les régions traditionnellement onéreuses connaissent une croissance transactionnelle modeste tandis que les zones au coût immobilier inférieur s'avèrent être des marchés en pleine effervescence. Ces tendances pourraient illustrer une redistribution des dynamiques d'achat et d'investissement immobilier, à l'aune desquelles les régions plus abordables gagnent en popularité.

## **7) Analyse du rapport prix (/m<sup>2</sup>) /densité par département**

Le marché immobilier est un écosystème complexe où l'offre et la demande se rencontrent, l'interaction entre le prix du mètre carré (m<sup>2</sup>) et la densité de population dans chaque département permet de rendre compte des principes microéconomiques à l'œuvre.

Lorsque l'on observe la relation entre le prix du m<sup>2</sup> et la densité de population, une corrélation significative de 85% se dessine. Cette forte corrélation met en lumière un lien étroit entre ces deux variables : à mesure que la densité de population dans un département augmente, le prix du m<sup>2</sup> tend à suivre une tendance similaire à la hausse.

D'un côté, une densité de population élevée crée une demande accrue pour le logement dans un département donné. Plus il y a de personnes cherchant à s'installer dans une zone spécifique, plus la demande de biens immobiliers résidentiels, comme les maisons ou les appartements, augmente. Cette demande soutenue exerce une pression à la hausse sur le prix du m<sup>2</sup>, car les acheteurs potentiels sont prêts à payer une prime pour accéder à des biens immobiliers dans des zones densément peuplées.

D'un autre côté, l'offre de logements disponibles dans un département peut être limitée, en particulier dans les zones urbaines où l'espace constructible est restreint. La demande croissante associée à une offre relativement stable crée un déséquilibre sur le marché immobilier, favorisant ainsi une augmentation des prix du m<sup>2</sup>.

Le coefficient de détermination de 65% souligne la capacité de ces facteurs à expliquer et à prédire la variation des prix du m<sup>2</sup> en fonction de la densité de population dans les départements étudiés. Bien que d'autres variables puissent également influencer le marché immobilier, telles que les politiques gouvernementales, les tendances économiques ou d'autres caractéristiques géographiques, la corrélation observée entre le prix du m<sup>2</sup> et la densité de la population met en évidence l'importance de la densité d'un département.

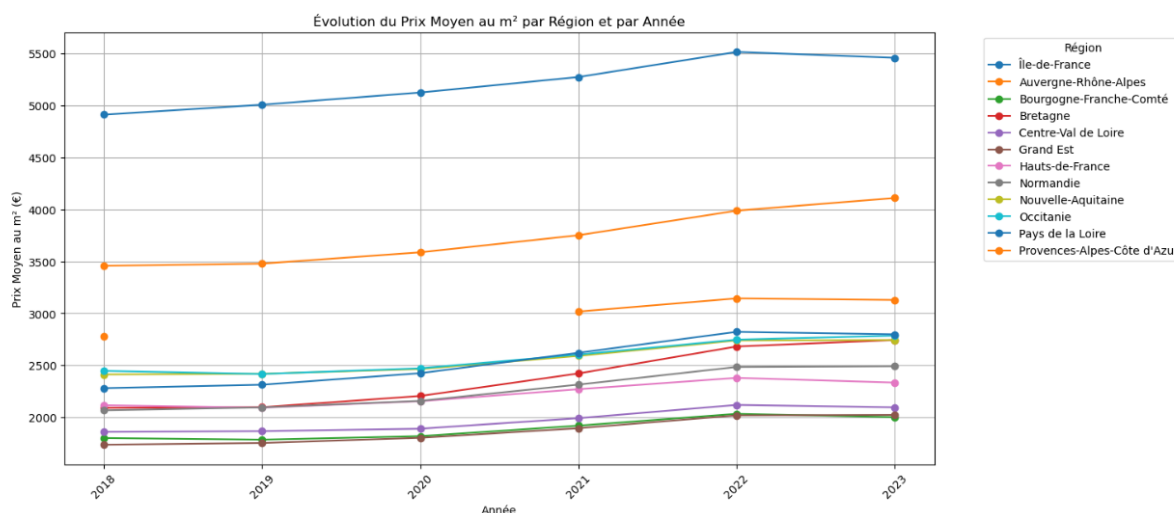


## 8) Les conséquences de la pandémie du COVID-19 sur le marché immobilier

A la suite de la mise en place d'un confinement national en mars 2020, le marché de l'immobilier a été profondément impacté. Au cours du premier confinement, et ce jusqu'à la fin de l'année 2020, nous avons observé que la demande pour les biens immobiliers poursuivait sa hausse malgré l'arrêt quasi complet de l'activité économique en France. Ce décalage a permis d'enregistrer un pic de transactions dans toutes les régions de France, sans exception. A partir du début de l'année 2021, la demande pour les biens immobiliers a cependant connu une baisse notable, attribuable en grande partie aux incertitudes économiques et aux restrictions de mobilité. Cette baisse de la demande a entraîné une chute des prix du mètre carré principalement ressentie dans les grandes métropoles.

De l'autre côté, un phénomène intéressant, bien qu'attendu, a pu se vérifier. En effet, la baisse de la demande pour les biens situés dans les régions les plus chères de France a été partiellement contrebalancée par une hausse de la demande dans les régions auparavant très peu sollicitées. La région Bourgogne-Franche-Comté ainsi que le Centre-Val de Loire, connues pour leurs zones rurales ou semi-urbaines, ont vu leur attractivité augmenter.

La pandémie du COVID-19 a mis en exergue une recherche de qualité de vie supérieure chez les Français, des espaces plus grands pour télétravailler et une volonté d'échapper à la densité urbaine.



De manière générale, nous observons que la pandémie n'a pas permis une baisse des prix à l'échelle régionale. En effet, la tendance des prix post-confinement est restée positive dans toutes les régions de France, confirmant un marché solide qui a rapidement su s'adapter à un environnement volatil. Nous retrouvons également une forte hausse des prix au mètre carré des biens situés en région Auvergne-Rhône-Alpes confirmant un attrait croissant pour cette région dynamique en forte croissance économique profitant de son rapprochement avec la Suisse et Lyon, sa capitale.

## Conclusion Partielle

---

Cette première étude du marché immobilier en France nous a permis de déterminer plusieurs tendances significatives. On remarque une tendance générale à la hausse des prix au mètre carré sur une période de cinq ans, avec des variations saisonnières annuelles. Les régions telles que l'Île de France et les départements frontaliers avec la Suisse ont affiché des prix particulièrement élevés, tandis que les régions moins densément peuplées ont vu une croissance transactionnelle plus dynamique. Cette étude nous permet de mettre en lumière l'impact de la pandémie du COVID-19 sur le marché immobilier, notamment avec des fluctuations importantes du nombre de transactions dans différentes régions.

Une étude des taux d'intérêt serait essentielle pour éclairer davantage les variations des prix immobiliers, étant donné leur influence significative sur la demande et les coûts d'emprunt.

## IV. Impact des taux d'intérêts sur le marché immobilier résidentiel

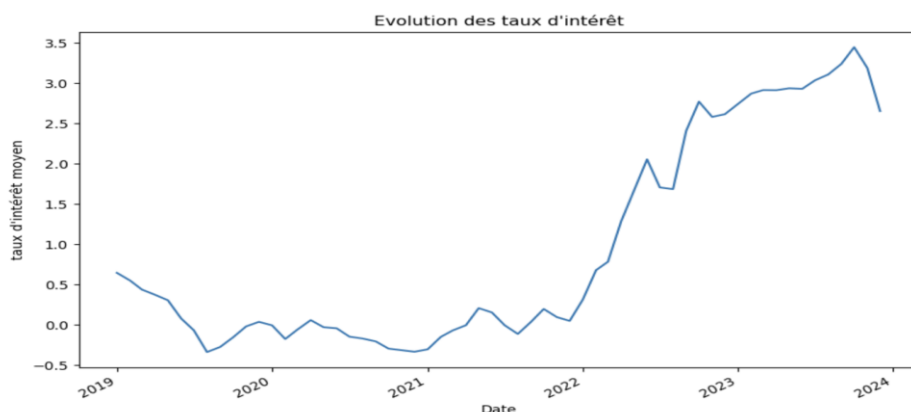
---

L'analyse des dynamiques du marché de l'immobilier par région et par département révèle des disparités importantes par rapport à la dynamique nationale. Ces différences s'illustrent particulièrement à travers les effets de la pandémie de COVID-19, très différents d'une région à l'autre. Au-delà du COVID, nous allons essayer d'évaluer les effets de la remontée des taux sur le marché immobilier français.

Pour évaluer cette dynamique, il a d'abord fallu trouver des données de taux d'intérêt. Pour cela, nous avons sélectionné la série mensuelle du taux des obligations d'État français à 10 ans, récupérée sur le site de la FRED (<https://fred.stlouisfed.org/series/IRLTLT01FRM156N>).

Le taux d'intérêt moyen sur notre échantillon est de 0.899. Il varie fortement sur la période, ce qui témoigne de l'évolution des conditions de financement des ménages. En effet, la BCE a commencé un cycle de politique monétaire restrictif début 2022, mettant fin à près de 10 ans d'une politique monétaire accommodante qui a commencé au moment de la crise des Subprimes.

### 1) Analyse temporelle des taux d'intérêts



Le premier constat est que les taux d'intérêt ont globalement évolué à la hausse sur la période. Nous pouvons distinguer deux sous-périodes dans notre série :

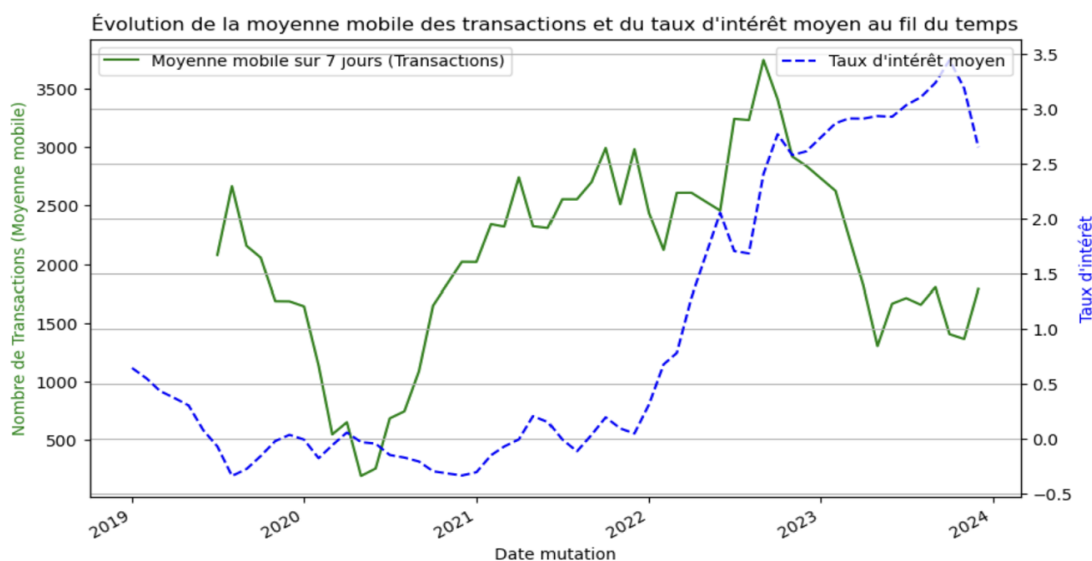
- La première, de 2019 à 2022, se caractérise par des niveaux de taux historiquement bas, voire négatifs ;
- La seconde, à partir de 2022, se caractérise par une remontée très rapide des taux d'intérêt.

Le retour de l'inflation, à la suite de la période de confinement, a nécessité une réponse de la BCE qui met en œuvre une politique de resserrement quantitatif afin de contrôler l'inflation. Cette politique se traduit par un relèvement du taux directeur, et donc par extension du taux d'intérêt français. Lorsque la banque centrale réduit la quantité de monnaie centrale à disposition des banques, ces dernières ont tendance à réduire la quantité de crédits octroyés aux ménages.

Intuitivement, on peut donc penser que les prix de l'immobilier ont diminué à partir de 2022. En effet, avec des conditions de crédit restreintes, le nombre de ménages pouvant accéder à la propriété se réduit. Les vendeurs de biens immobiliers vont donc devoir diminuer leur prix s'ils veulent réussir à trouver un acquéreur. Si ces derniers ne sont pas prêts à accepter une décote sur le prix de leurs biens, cela devrait mécaniquement entraîner une diminution du nombre de transactions.

## 2) Comparaison entre l'évolution des taux et la dynamique des transactions immobilières

En juxtaposant le nombre de transactions immobilières et l'évolution des taux d'intérêt, nous observons une corrélation négative : lorsque les taux augmentent, le volume des transactions a tendance à diminuer.



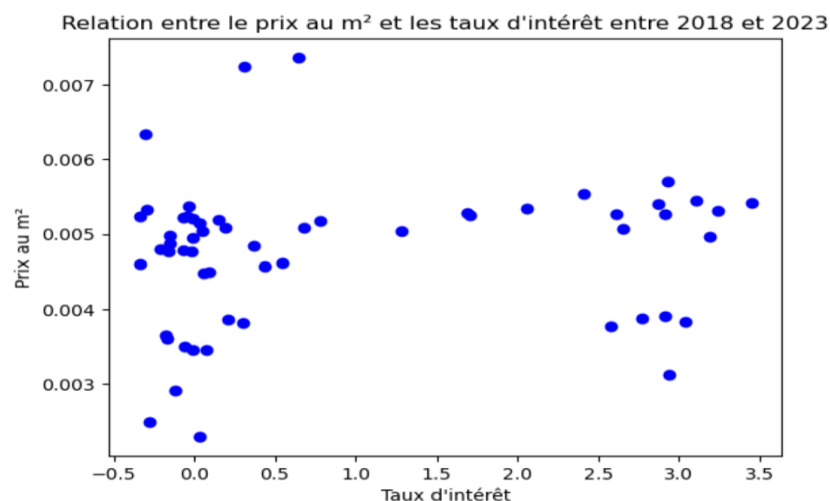
Le nombre de transactions diminue très fortement en 2020. Cela s'explique par l'arrêt du marché immobilier au moment des mesures de confinement. Ces transactions reviennent à des niveaux élevés, témoignant d'une reprise économique post COVID forte, portée par des niveaux de taux qui restent bas.

La croissance du nombre de transactions s'infléchit à mesure que les taux commencent à remonter. Lorsqu'ils dépassent le niveau de 2,5%, le nombre de transactions s'effondre.

Nous pouvons ainsi supposer qu'au-delà d'un certain taux d'intérêt, la dégradation des conditions de crédit pèsent sur le nombre de transactions.

### 3) Comparaison du taux d'intérêt avec la dynamique des prix immobiliers

Nous modélisons l'évolution des prix immobiliers en fonction des taux d'intérêts.



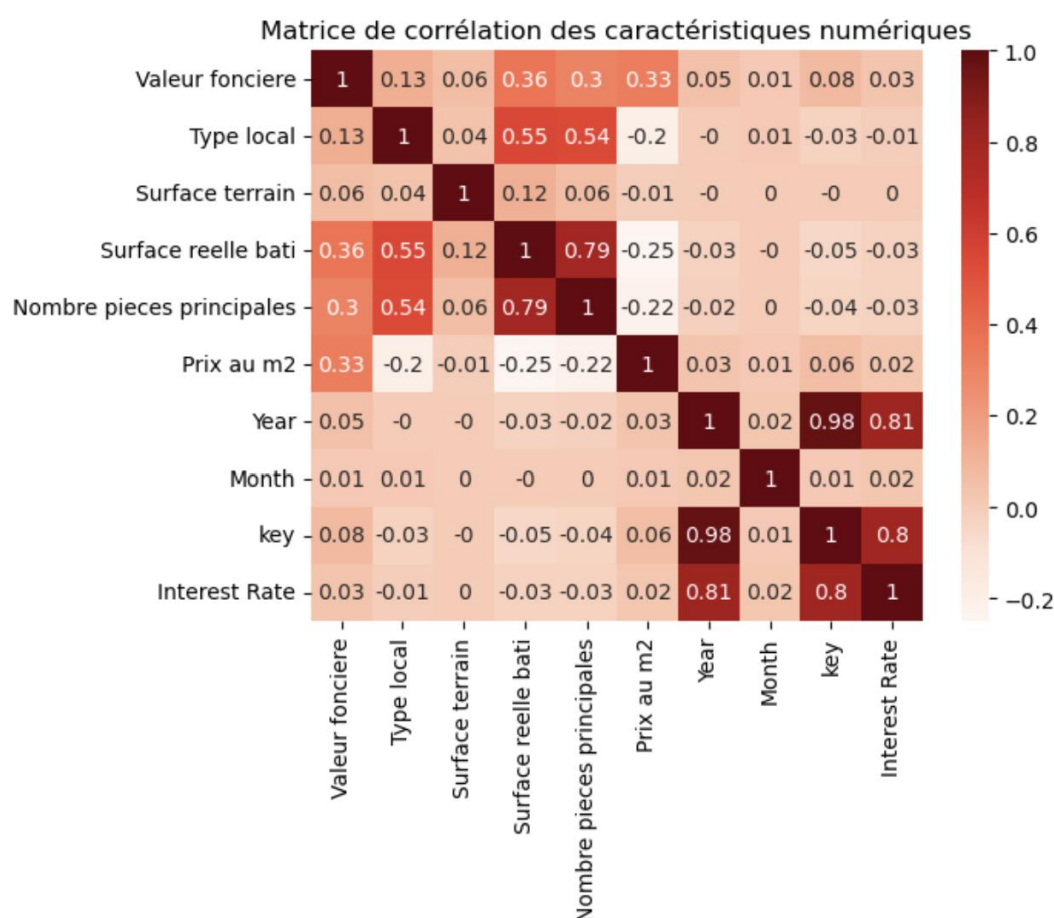
Nous nous attendions à observer une relation décroissante entre les prix et le taux d'intérêt. En effet, lorsque les taux d'intérêts augmentent, le coût du crédit augmente également ce qui renchérit un investissement immobilier. Dans le même temps, les banques ont tendance à réduire le nombre de crédits accordés (effet trend-cycle) ce qui diminue le nombre de ménages pouvant acheter des biens immobiliers. Les ménages souhaitant vendre devraient donc mécaniquement diminuer le prix de vente s'ils souhaitent trouver un acheteur (offre constante et baisse de la demande).

Malgré ces considérations théoriques, nous n'observons pas de telle relation dans les données. Cela peut s'expliquer par la nature des taux d'intérêts que nous utilisons (taux d'emprunt à 10 ans de l'Etat français). Nous observons par exemple que le taux maximal est de 3,5% sur toute la période alors que les taux immobiliers sont généralement légèrement supérieurs au taux des obligations d'Etat...

## V. Évaluation des variables pertinentes pour modéliser l'évolution des prix

Jusqu'à présent, nous avons étudié le comportement des prix immobiliers sur le territoire français au moment du COVID et comparé l'évolution des prix avec le taux d'intérêt en vigueur. L'objectif est maintenant de construire un modèle de prévision des prix immobiliers. Pour cela, il convient d'identifier les variables les plus pertinentes à utiliser dans le modèle.

Nous représentons donc la matrice de corrélation des variables :



Cette matrice fournit un aperçu visuel des relations entre les variables du jeu de données. Plus la teinte est foncée, plus la corrélation est forte. Ce choix de teinte s'explique par le fait que l'essentiel des corrélations sont positives, il n'est donc pas nécessaire de distinguer les corrélations positives et négatives puisqu'il y a peu de chances que les corrélations négatives soient significatives.

Compte tenu des résultats, les principales variables explicatives des prix de l'immobilier semblent être la surface réelle bâtie et le nombre de pièces principales. Les corrélations entre ces variables et les prix ne sont pourtant pas très élevées, ce qui implique que d'autres facteurs sont impliqués dans la modélisation de cette dynamique des prix. On peut en particulier

penser à la densité que nous avons utilisée pour mesurer les dynamiques de prix par zone géographique.

## **VI. Prédiction des prix par un modèle de Machine Learning**

---

Dans cette étude, nous avons entrepris une exploration exhaustive des modèles et des algorithmes pour prédire les résultats d'un ensemble de données donné. L'objectif était de déterminer le modèle le plus performant pour la prédiction des prix. Pour ce faire, nous avons utilisé une variété d'algorithmes et de techniques d'entraînement, puis nous avons évalué leurs performances pour sélectionner le modèle le plus optimal.

### **Méthodologie**

Nous avons commencé par rassembler et prétraiter les données nécessaires à notre analyse. Ensuite, nous avons divisé l'ensemble de données en ensembles d'entraînement et de test pour évaluer la performance des modèles.

Nous avons ensuite procédé à l'entraînement de plusieurs algorithmes différents.

#### **1) Les Fondements des Forêts Aléatoires**

Les Forêts Aléatoires sont une technique d'apprentissage ensembliste basée sur des arbres de décision. Elles ont été introduites par Leo Breiman et Adèle Cutler dans les années 2000. L'idée principale derrière les Forêts Aléatoires est de construire plusieurs arbres de décision en utilisant des échantillons bootstrap de l'ensemble de données d'origine et en sélectionnant aléatoirement un sous-ensemble des caractéristiques à chaque division de l'arbre.

#### **2) Construction des Forêts Aléatoires**

La construction d'une Forêt Aléatoire implique plusieurs étapes :

- a. Échantillonnage Bootstrap : Un échantillon bootstrap est créé en tirant aléatoirement avec remplacement des observations à partir de l'ensemble de données d'entraînement. Cela permet d'introduire de la variabilité dans chaque arbre.
- b. Sélection aléatoire des caractéristiques : À chaque nœud de l'arbre, un sous-ensemble aléatoire des caractéristiques est sélectionné pour la division. Cela garantit la diversité entre les arbres de la forêt.
- c. Construction de l'arbre : Chaque arbre est construit en utilisant l'algorithme de construction d'arbre de décision, tel que CART (Classification and Regression Trees) ou

C4.5. La croissance de l'arbre est limitée pour éviter le surajustement grâce à des techniques telles que la validation croisée.

### **3) Prédiction avec les Forêts Aléatoires**

Une fois que la Forêt Aléatoire est construite, elle peut être utilisée pour faire des prédictions sur de nouvelles données. Pour la classification, la prédiction finale est déterminée par un vote majoritaire des arbres, tandis que pour la régression, les prédictions sont généralement moyennées sur tous les arbres.

### **4) Avantages et Applications**

Les Forêts Aléatoires offrent plusieurs avantages, notamment :

- ✓ Robustesse aux Données de Grande Dimension : Elles peuvent gérer efficacement des ensembles de données avec un grand nombre de caractéristiques.
- ✓ Résistance au Surajustement : En combinant les prédictions de plusieurs arbres, les forêts aléatoires sont moins susceptibles de réajuster les données d'entraînement.
- ✓ Interprétabilité : Bien qu'elles ne soient pas aussi facilement interprétables qu'un seul arbre de décision, les forêts aléatoires permettent une certaine compréhension des facteurs influençant les prédictions.

Les applications des Forêts Aléatoires sont vastes et incluent la classification, la régression, la détection d'anomalies, et plus encore. Elles sont largement utilisées dans des domaines tels que la finance, la médecine, la biologie, et l'analyse de données.

## **Conclusions**

---

Au cours de cette étude, nous avons retraité un ensemble de données immobilières françaises afin d'évaluer l'impact de deux chocs conjoncturels - le COVID et la remontée des taux, sur les prix. La gestion du volume de données nous a posé des difficultés, notamment en ce qui concerne la gestion de la mémoire afin de produire un code qui s'exécute assez rapidement. Nous avons observé que la crise du COVID a eu un impact différencié sur la demande de biens immobiliers selon les régions. Certaines régions rurales ont ainsi vu le nombre de transactions augmenter alors que de grandes métropoles comme Paris ont connu une diminution des transactions. Cela témoigne d'une inflexion de la demande des ménages qui peut s'expliquer par le développement du télétravail et la volonté de certains ménages de partir s'installer à la campagne.

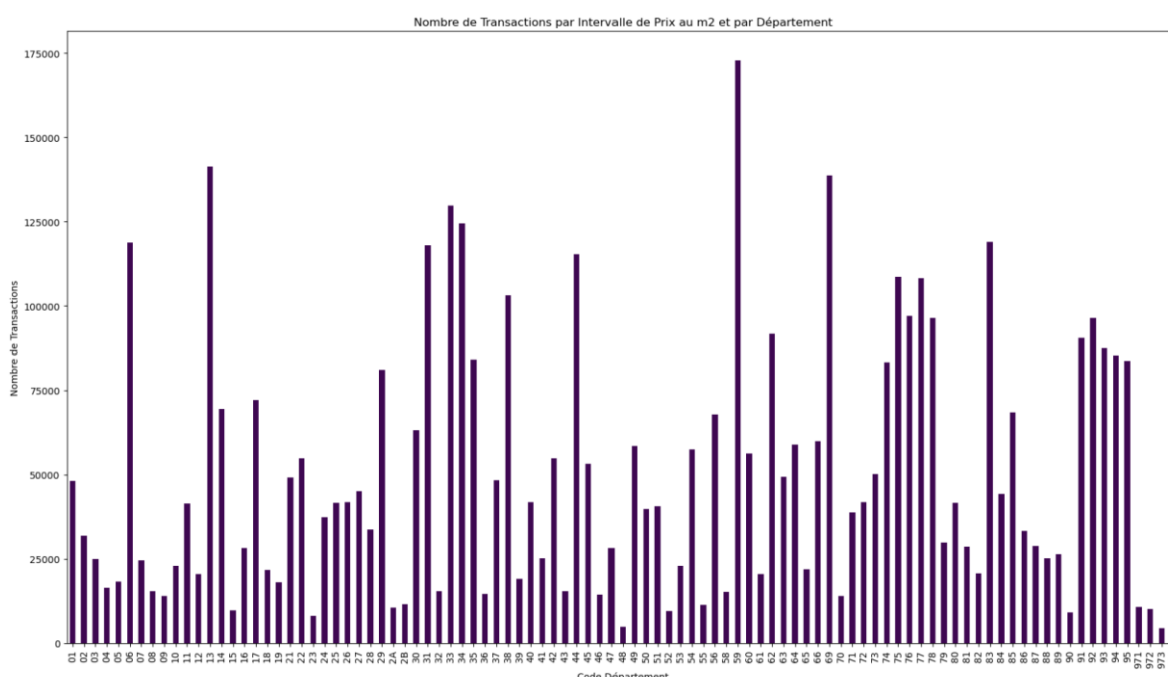
Nous avons également observé que la remontée des taux d'intérêt a fortement réduit le nombre de transactions de biens immobiliers. Il aurait été intéressant de comparer les dynamiques entre immobilier résidentiel et immobilier commercial si nous avions disposé de données sur les taux de vacances par exemple.

Enfin, nous avons sélectionné des variables pertinentes pour expliquer l'évolution des prix. Les corrélations entre les prix et les variables de notre échantillon restent assez faibles, ce qui implique que d'autres données auraient pu être utilisées pour modéliser la dynamique des prix, à l'instar de la densité que nous avons utilisé dans notre analyse géographique. Enfin, nous avons essayé de modéliser l'évolution des prix immobiliers par un modèle de machine learning. Cette discipline étant nouvelle pour nous, nous avons dû effectuer des recherches concernant les étapes à mettre en œuvre pour entraîner correctement le modèle. Les résultats du modèle ne sont cependant pas concluants. Cela peut s'expliquer par un mauvais choix de variables explicatives, un mauvais entraînement du modèle, ...

## Annexes

---

- Nombre de transactions par département 2018 - 2023 :



Nous avons également réaliser des graphiques permettant d'étudier l'évolution du nombre de transactions à la maille départementales pour chaque région.



- Résultats du Modèle de Machine Learning :

```
Training Accuracy = 0.16494820329632598  
Test Accuracy     = 0.1249857751440887
```

```
evaluate(dt, X_train, y_train)
```

```
Average Error  = 0.1507 degrees  
Model Accuracy = 0.1649 %
```

```
evaluate(dt, X_test, y_test)
```

```
Average Error  = 0.1540 degrees  
Model Accuracy = 0.1250 %
```