

## Estimation de la valeur foncière des logements dans les Deux-Sèvres (79)

### Introduction

Dans le cadre d'un projet de première année de BUT Sciences des Données, nous avons réalisé une étude visant à **estimer la valeur foncière de logements situés dans le département des Deux-Sèvres (79)**. L'objectif était de proposer une **modélisation simple mais pertinente** à partir de données brutes, en combinant l'exploitation d'un fichier Excel avec une transposition de notre méthode sur RStudio.

Le jeu de données fourni contenait plusieurs variables explicatives essentielles à la compréhension du marché immobilier local, telles que :

- La **ville** dans laquelle se situe le bien,
- Le **type de logement** (maison ou appartement),
- La **surface réelle** du logement (en  $m^2$ ),
- La **surface du terrain** (en  $m^2$ ),
- Le **nombre de pièces**,
- Et enfin, la **valeur foncière** (en euros), qui constitue la **variable cible** de notre étude.

---

### Démarche de recherche

#### 1. Création d'une variable prédictive pertinente

Dans un premier temps, nous avons construit une **variable synthétique** permettant de prédire approximativement la valeur d'un bien immobilier à partir de données disponibles et facilement interprétables.

L'idée a été la suivante :

- Pour tous **les appartements et les maisons situées à Niort**, nous avons estimé le **prix au  $m^2$**  moyen ainsi que la valeur moyenne d'une pièce. Nous avons ainsi obtenu deux coefficients :
  - Le **prix moyen au  $m^2$  d'un appartement à Niort**,
  - Le **prix moyen au  $m^2$  d'une maison à Niort**.
  - Le prix moyen d'une pièce d'une Maison à Niort
  - Le prix moyen d'une pièce d'un appartement à Niort

- Pour **les maisons situées en dehors de Niort**, nous avons élargi la logique : En plus de la surface réelle et du nombre de pièces, nous avons également tenu compte de la **surface du terrain**, car en zone rurale ou périurbaine, cette surface influe notablement sur le prix final.
- Ce choix de construction d'une **variable prédictive** repose sur une logique économique : la valeur d'un bien immobilier est généralement fonction de sa taille, de son usage (maison ou appartement) et de son emplacement (urbain ou non). Intégrer le terrain pour les maisons hors Niort ajoute une dimension réaliste au modèle.

## 2. Découpage du jeu de données par segments

Afin d'améliorer la précision des prédictions, nous avons **segmenté le jeu de données** en plusieurs sous-populations homogènes. Cette méthode permet d'ajuster plus finement le modèle aux spécificités de chaque type de bien.

Nous avons ainsi créé **huit feuilles de calcul distinctes dans Excel**, en combinant :

- **Le type de logement** (maison ou appartement),
- **La localisation** (à Niort ou hors Niort),
- **Les quartiles de la surface du terrain** (uniquement pour les maisons).

Les segments sont donc les suivants :

1. Appartements à Niort
2. Appartements hors Niort
3. Maisons à Niort, terrain  $< 283 \text{ m}^2$
4. Maisons à Niort, terrain  $\in [283 ; 578[$
5. Maisons à Niort, terrain  $\in [578 ; 810[$
6. Maisons hors Niort, terrain  $< 283 \text{ m}^2$
7. Maisons hors Niort, terrain  $\in [283 ; 578[$
8. Maisons hors Niort, terrain  $\in [578 ; 810[$

Ce découpage par quartiles nous permet de **contrôler la variabilité** dans les groupes, et de capter des tendances spécifiques à chaque configuration de bien.

## 3. Modélisation linéaire dans Excel

Dans chaque feuille de calcul, nous avons modélisé la valeur foncière en utilisant une **régression linéaire simple** basée sur notre variable prédictive.

Nous avons appliqué la formule **DROITEREG** sur Excel pour obtenir les coefficients a et b d'une droite

Cette étape nous a permis de déterminer pour chaque segment une équation propre, ajustée aux caractéristiques de la population étudiée.

#### 4. Évaluation de la précision des prédictions

Pour mesurer l'efficacité de notre modélisation, nous avons calculé, pour chaque feuille, **l'écart moyen absolu** entre les valeurs foncières réelles et les valeurs prédites.

Ce choix de métrique repose sur sa lisibilité et sa robustesse : il permet de quantifier l'écart moyen sans être influencé par d'éventuelles valeurs extrêmes (contrairement à la variance).

Cette évaluation nous a aidés à **ajuster les coefficients au m<sup>2</sup>** utilisés dans les formules, dans le but d'**optimiser l'écart moyen** et de **rapprocher au mieux les prédictions des valeurs réelles**.

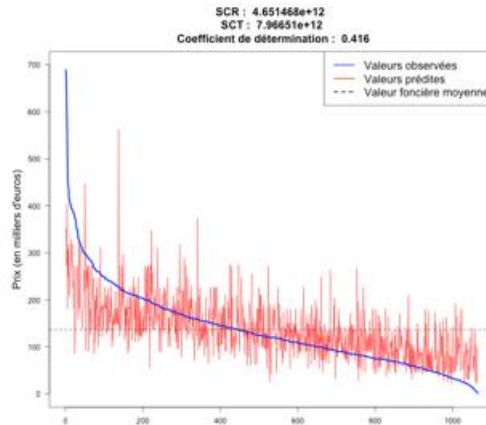
Commune	Code.depart	Type.local	type local	Surface.reelle	Nombre.pieci	surface terrain	Valeur.fonciere	prix moyen au m2 vs nb pieces						
ADILLY	79	Maison	1	75	3	1381	119900	179753	0,50301393	36151,0737	126569,336	62758,611		
AIFFRES	79	Maison	1	117	6	866	265000	239377			156561,039			
AIFFRES	79	Maison	1	103	4	1182	2,00E+05	218987			146304,585			
AIFFRES	79	Maison	1	85	3	835	273000	173575			123461,716			
AIFFRES	79	Maison	1	143	4	909	310000	266893			170401,97			
AIFFRES	79	Maison	1	96	4	820	63000	195032			134254,886			
AIFFRES	79	Maison	1	95	4	823	137000	193689			133579,338			
AIFFRES	79	Maison	1	168	5	1510	330000	332156			203230,168			
AIFFRES	79	Maison	1	113	4	915	307130	223411			148529,918			
AIFFRES	79	Maison	1	121	5	834	281550	237989			155862,855			
AIFFRES	79	Maison	1	169	4	1609	454000	331375			202837,314			
AIFFRES	79	Maison	1	108	5	948	280000	223380			148514,325			
AIFFRES	79	Maison	1	112	4	948	260000	223208			148427,806			
AIFFRES	79	Maison	1	104	4	1347	230000	226714			150191,373			
AIGONDIGNE	79	Maison	1	83	5	839	165000	182813			128108,559			
AIGONDIGNE	79	Maison	1	170	7	2019	305000	366412			220461,413			
AIGONDIGNE	79	Maison	1	108	4	1782	272000	249072			161437,759			
AIGONDIGNE	79	Maison	1	110	5	1060	130000	230550			152120,935			
AIGONDIGNE	79	Maison	1	190	5	1461	175000	362348			218417,164			
AIGONDIGNE	79	Maison	1	102	3	1390	3,00E+05	219434			146529,432			
AIGONDIGNE	79	Maison	1	111	5	1200	166384	237327			155529,86			
AIGONDIGNE	79	Maison	1	95	4	1010	182650	200795			137153,755			
AIGONDIGNE	79	Maison	1	80	4	1048	167000	180384			126886,738			
AIGONDIGNE	79	Maison	1	185	5	881	350000	333023			203666,281			
AIGONDIGNE	79	Maison	1	90	4	1100	155000	196930			135209,606			
AIGONDIGNE	79	Maison	1	90	4	1035	135100	194460			133967,162			
AIGONDIGNE	79	Maison	1	73	4	1605	237000	234230			133966,530			

#### Modèle retenu

Le modèle final retenu repose sur une **formule de type linéaire** propre à chaque segment de bien, avec des coefficients calculés de manière empirique à partir des prix au m<sup>2</sup>. Ce choix est **pertinent pour plusieurs raisons** :

- Il est **simple à interpréter** et à mettre en œuvre sur Excel.
- Il repose sur une **logique économique concrète** (surface x prix unitaire).
- Il **s'adapte** à chaque type de bien grâce au découpage segmenté.

Nous avons ensuite **reproduit l'ensemble de notre travail sur RStudio**, en utilisant des fonctions comme `lm()` pour ajuster les modèles linéaires, et `subset()` pour isoler les différents groupes. Cette étape nous a permis de **valider la cohérence de notre méthode** et d'utiliser un environnement statistique plus rigoureux.



## Conclusion

Ce projet nous a permis de combiner l'analyse de données, la logique économique et la modélisation statistique dans un cadre réaliste. Grâce à des outils simples (Excel et RStudio), nous avons appris à :

- Créer des **variables pertinentes** à partir de données brutes,
- Segmenter un jeu de données de manière intelligente,
- Modéliser les relations linéaires et interpréter les coefficients,
- Évaluer la qualité de nos prédictions.

La transition d'Excel vers R nous a permis de mieux comprendre les forces et limites de chaque outil, et d'anticiper les perspectives d'amélioration (modèles plus complexes, analyse multivariée, etc.).

Ce travail fut particulièrement formateur, car il nous a demandé de **réfléchir aux hypothèses sous-jacentes**, de **justifier nos choix méthodologiques** et de **traduire une situation réelle en équations statistiques**.

---