

CE QUE LA DATA FAIT AU MARCHÉ

Données immobilières, transparence des marchés
et fixation des prix des logements en France

Par **Alexandre COULONDRE** | Chercheur associé au LATTIS (Ecole des Ponts Paris Tech)
Claire JUILLARD | Chercheuse indépendante, Présidente d'OGGI Conseil

Recherche réalisée
avec le soutien de



Table des matières

Remerciements	3
Introduction	4
Questionner la transparence des marchés résidentiels	4
Une recherche originale sur les pratiques économiques des ménages	5
Plan du rapport et principales conclusions	7
Un vaste paysage informationnel, une diversité d'usages	9
Le prix des logements : une information multi-dimensionnelle	9
Quatre types d'informations et une préférence pour les références individuelles de prix	10
Quatre canaux de consultation et la règle du panier garni	11
Des supports mobilisés en quantité, un effet de l'offre digitale	13
S'informer sur les prix des logements : une pratique à intensité variable	16
La place de la « donnée » dans le paysage informationnel : centrale mais variable	19
Qu'est-ce que la donnée immobilière du point de vue des particuliers ?	19
Quelle place la donnée tient-elle dans le panier garni d'informations des particuliers?	20
2 vendeurs sur 5 ont recours à des données sur les prix	22
Une consommation inégale de données sur les prix	23
L'effet de l'âge et de la position sociale	23
Une consommation de données qui augmente avec la tension du marché local	25
Des inégalités d'usages cumulatives	26
Consommation de données et sentiment de transparence des marchés : un lien sous conditions	28
Un impact conditionné sur le sentiment de transparence	28
Un flou sur les prix entretenu par le foisonnement des supports numériques	29
Usages des données et pratiques marchandes des vendeurs de logements : la fixation des prix en question	33
Des consommateurs de données moins souvent accompagnés par un professionnel	33
Des consommateurs de données tentés par les valeurs supérieures	35
Les consommateurs de données : des vendeurs plus satisfaits	36
En somme : des effets contrastés de la donnée sur les marchés	38
Conclusion	40
Annexe : Méthodologie de l'enquête par questionnaires	41
Bibliographie	44



Remerciements

Cette recherche est la seconde d'une série, que nous espérons longue, sur les données immobilières en France¹. Principalement réalisée entre septembre 2020 et avril 2021, elle a reçu le soutien du PUCA, du LIFTI, du Groupe BPCE et de Crédit Logement.

Nous remercions François Ménard et Eva Simon (PUCA), Marc Kaszynski (LIFTI), Alain Tourdjman, Isabelle Friquet-Lepage et Bertrand Cartier (Groupe BPCE), Jean-Marc Vilon, Catherine Lanvario et Marie-Line Milheiro (Crédit Logement), pour leur soutien financier, leur suivi des travaux au sein du comité de pilotage de la recherche et leurs suggestions toujours pertinentes.

Nous remercions également Zoé Chaloin (FNAU), Emmanuel Perray (FNAIM) et Emmanuel Bouriau (AUDIAR) pour leur participation au comité de pilotage de la recherche, en qualité de membres et représentants du LIFTI.

Notre recherche a aussi profité d'échanges fructueux avec plusieurs personnalités que nous remercions vivement : Fabrice Bardet (Université de Lyon, ENTPE), Guilhem Boulay (Université d'Avignon), Pierre Lamarche (INSEE).

Nous remercions également Basile Pfeifer (DHUP) pour le partage de données de cadrage, par ailleurs indisponibles, sur les vendeurs de logements en France.

Nous remercions Stéphane Lefebvre (OpinionWay) pour le pilotage de la passation du questionnaire sur lequel repose cette recherche.

Merci aussi à Nathan Benamouzig, étudiant (Sciences Po), pour son aide à l'actualisation de la documentation sur les producteurs de données et les prix immobiliers.

Enfin, nous remercions l'ensemble des vendeurs de logements et des agents immobiliers qui nous ont accordé un entretien dans le cadre de notre enquête exploratoire.

Alexandre Coulondre et Claire Juillard

Contacts :

alexandre.coulondre@gmail.com

claire.juillardm@gmail.com

¹ Juillard C. (2019a), *Le tournant numérique des données immobilières : permanences et recompositions*, Etude publiée par Iread avec le soutien d'Iread, du LIFTI, du PUCA et d'Urbanics.



Introduction

Questionner la transparence des marchés résidentiels

Il paraît acquis que le libre accès à des informations économiques concourt à la transparence des marchés et que celle-ci a des effets vertueux, notamment sur les prix.


Cette idée est largement répandue dans le monde de l'immobilier. Elle stimule le développement de nouveaux services innovants notamment de la part de jeunes entreprises réunies sous le label « Proptech ». Elle sous-tend aussi la politique d'*open data* de l'Etat qui ouvre progressivement les sources d'informations dont il dispose sur le stock, la construction et la vente de logements. L'ouverture des données DVF (Demande de Valeurs Foncières) en 2019 en est un symbole marquant.

L'idée de transparence fait référence de façon explicite ou implicite à la théorie des marchés élaborée par les économistes néo-classiques. Cette théorie établit que la transparence, en instaurant la bonne information de tous (offreurs, demandeurs et intermédiaires) sur les caractéristiques et les prix des biens et des services, contribue à une concurrence pure et parfaite et ce faisant, à la rencontre de l'offre et de la demande en un point d'équilibre notamment défini en termes de prix.

Or, la difficulté d'appliquer les concepts de l'économie standard à l'immobilier a de nombreuses fois été rappelée par la communauté scientifique. Les économistes eux-mêmes ont fait le constat d'une asymétrie d'information structurelle en immobilier (Garmaise, Moskowitz 2004). Plus encore, ils se sont confrontés à une des grandes spécificités des marchés du logement : l'hétérogénéité des biens. Celle-ci rend inopérant le modèle néo-classique dans sa forme canonique. Les économistes ont alors privilégié une approche alternative de la formation des prix à partir de modèles hédoniques capables de décomposer les biens immobiliers et de mesurer l'impact de chacune de leur composante sur leur prix (Sheppard 1999, Cavailhès 2005, Simon 2008).

Sans développer vraiment le cas spécifique de l'immobilier, la sociologie a, quant à elle, critiqué les fondements mêmes du raisonnement néoclassique (comme l'a fait également la géographie sur d'autres aspects : Pumain *et al.* 2007). Elle a montré qu'une telle approche ne rendait pas vraiment compte des pratiques concrètes des acteurs économiques faites en réalité de rationalité limitée, de tâtonnements et de représentations sociales et territoriales (François, 2008). Les sociologues ont alors suggéré de ne plus appréhender le marché comme un ajustement mécanique entre l'offre et la demande autour d'un prix mais comme un ensemble de sous-espaces dans lesquels des compromis sur les prix s'opèrent entre des acteurs économiques pluriels qui ont la capacité de peser sur la fixation des valeurs et sur la définition même des biens (Vatin 2009, Desrozières 2011, Callon 2017).

Si le modèle de marché néoclassique fait tant débat en général et en particulier dans l'immobilier (Cusin 2012), il convient peut-être de ne pas prendre pour acquis son corollaire de la transparence et de ses effets vertueux sur la fixation des prix.



L'idée selon laquelle l'existence de données de marché engendre la transparence et le nivellement des prix s'appuie en fait sur trois postulats implicites :

- Le premier postulat consiste à penser que la mise à disposition d'une information implique nécessairement son usage de la part des acteurs économiques.
- Le deuxième consiste à penser que l'usage de la donnée débouche assurément sur une meilleure connaissance des marchés.
- Le troisième est que cet usage a un effet modérateur sur les prix. Or, lorsqu'un acteur économique fait effectivement usage de l'information immobilière et qu'il en retire une meilleure connaissance des marchés, doit-on supposer qu'il se conforme forcément à un prix « de marché » ?

Rien n'indique dans les faits qu'un lien aussi systématique s'établisse entre ces éléments.

Notre recherche ambitionne précisément de soumettre la théorie de la transparence à l'épreuve des faits dans le cas de l'immobilier résidentiel. Une telle démarche ne consiste pas à rejeter en bloc l'idée de transparence et de ses effets en termes de prix. Elle consiste à tester empiriquement l'ensemble des maillons du raisonnement qui relie la disponibilité des données, la transparence des marchés et la fixation des prix.


Une recherche originale sur les pratiques économiques des ménages

Les portes d'entrée dans une telle question de recherche sont nombreuses. Le point de vue défendu ici se positionne au plus près des marchés. Il consiste à questionner les pratiques concrètes des acteurs économiques dans les moments où ils établissent les prix des logements, et à interroger la place de l'information dans ce processus.

La recherche se limite au logement ancien dans la mesure où les modalités de définition des prix des logements neufs sont spécifiques et nécessitent une analyse dédiée. La recherche porte par ailleurs uniquement sur les particuliers (ménages), considérant que le comportement des personnes morales (investisseurs, bailleurs sociaux, etc.) requiert également une analyse *ad hoc*.

Plus précisément, la recherche se concentre sur les particuliers vendeurs de logements anciens. Pourquoi privilégier le prisme des vendeurs ? Il ne s'agit pas de postuler que les acheteurs ou les intermédiaires professionnels ne jouent aucun rôle dans la fixation des prix mais plutôt de penser que le vendeur est au centre du processus. C'est lui qui esquisse un premier prix pour son bien, qui, le cas échéant, en discute avec le professionnel chargé de la vente, et qui tranche les négociations avec l'acheteur.

A travers une étude approfondie des pratiques économiques des vendeurs et des étapes qu'ils empruntent pour établir le prix des logements qu'ils vendent, nous pensons que nous sommes en mesure d'identifier les modalités concrètes de fixation de la valeur et, par la même occasion, de savoir si l'information immobilière joue un rôle dans le processus.



Pour ce faire, la recherche s'appuie sur de nombreux matériaux d'enquête : une série d'entretiens exploratoires menés auprès de vendeurs de logements et d'agents immobiliers en France, une étude comparative des principaux sites Internet d'informations immobilières, et une enquête par questionnaires.

Le questionnaire constitue l'élément central du dispositif de recherche (voir annexe). Il a été administré entre décembre 2020 et janvier 2021 à un échantillon de 993 répondants représentatifs de la population des Français ayant vendu un logement dans les 24 derniers mois (2019-2020). Il couvre l'ensemble des profils sociodémographiques et des territoires et appréhende ainsi la diversité des marchés immobiliers locaux, entre zones tendues et zones détendues, métropoles, villes moyennes et communes rurales.

Le questionnaire déploie une approche détaillée des conditions de vente des logements en France. Plus précisément, il permet de décrire finement le paysage informationnel dans lequel les vendeurs évoluent, ainsi que de cerner les ressources qu'ils en tirent et les stratégies qu'ils adoptent pour établir leur prix. L'enquête prend place dans le contexte du tournant numérique de l'immobilier et se concentre donc plus particulièrement sur la catégorie spécifique d'informations immobilières qui lui est associée : la donnée de prix des logements. Elle apporte des réponses à un ensemble de questions fondamentales :

- Les vendeurs puisent-ils effectivement dans les statistiques ou les références de prix disponibles sur les marchés pour approcher le prix possible de leur logement ?
- Quelle forme l'information qu'ils mobilisent prend-elle ? S'agit-il d'indicateurs de prix moyens au mètre carré ? D'estimateurs automatiques qui s'offrent sur la toile ? Etc.
- Comment cette information sur les marchés résonne-t-elle avec les indications que les vendeurs peuvent collecter auprès de professionnels (agents immobiliers notamment) ou de leur entourage ?
- Quelle stratégie adoptent-ils et quel pari font-ils au moment d'arrêter le prix de leur bien : suivent-ils une tendance modératrice ou au contraire inflationniste ?
- Le rôle accordé à l'information immobilière dans la fixation des prix diffère-t-il selon le profil sociodémographique ou la localisation des vendeurs ?

En bref, cette recherche se penche de manière inédite sur le rôle effectif de l'information en général, et de la donnée en particulier, dans l'évolution des pratiques des vendeurs en matière de fixation des prix immobiliers. De fait, elle compte parmi les rares travaux statistiques disponibles sur les vendeurs de logements en France.



Plan du rapport et principales conclusions

Le présent rapport développe une description fine des usages effectifs que les vendeurs, le cas échéant, font des informations et parmi elles, des données disponibles dans la perspective de fixer le prix de mise en vente de leur logement. Il passe en revue les usages de l'ensemble des informations et données sur les prix des logements à disposition gratuite des particuliers, qu'elles soient publiques et offertes en *open data*, à l'instar de DVF depuis son ouverture en avril 2019, ou privées et accessibles *via* des interfaces de consultation en ligne. L'analyse a vocation à évaluer la place et le rôle spécifiques de la donnée dans la fixation du prix et son effet sur la transparence et le fonctionnement des marchés résidentiels.

Le rapport se déploie en cinq parties dont les principales conclusions sont les suivantes :

1) **Un vaste paysage informationnel, une diversité d'usages**

86% des vendeurs de logements consultent des informations sur les prix pendant leur parcours de vente. La première partie du rapport s'attache à reconstituer le paysage informationnel dans lequel ces vendeurs évoluent. Aussi vaste que divers, ce paysage est composé d'informations sur les prix de types très différents et aux canaux d'accès et supports multiples. L'analyse donne à voir la diversité du « panier garni » de références de prix que les particuliers puisent dans ce paysage. Elle révèle aussi la complémentarité du canal physique et du canal digital dans l'accès à l'information, autour du rôle central que les professionnels de l'immobilier continuent de jouer malgré les transformations du marché opérées par le tournant numérique.

2) **La place de la « donnée » dans le paysage informationnel**

La seconde partie établit que près de 40% des vendeurs s'informent sur les prix des logements en consommant des informations sous forme de données. Elle distingue les vendeurs selon l'intensité de leur usage de la donnée, à partir d'un effort original de définition et de mesure qui pallie le flou qui, dans les faits, entoure la notion même de donnée malgré son omniprésence dans les discours.

3) **Une consommation inégale de données sur les prix**

La troisième partie montre une absence de lien mécanique entre la disponibilité des données et leur utilisation. La consommation de données de prix est une pratique socialement et géographiquement située. L'enquête par questionnaires met en exergue l'impact combiné de l'âge et de la position sociale sur le recours à la donnée : les plus « grands » consommateurs ayant tendance à être plus âgés et à occuper un rang plus élevé dans la hiérarchie socioprofessionnelle. Autre constat marquant : les zones tendues concentrent les consommateurs de données et les zones détendues, les vendeurs qui s'en passent. On peut notamment y voir l'effet des profils des vendeurs, situés géographiquement.



4) **Consommation de données et sentiment de transparence des marchés**

La quatrième partie marque un temps fort de la discussion avec la théorie économique. Elle établit que la donnée a un effet conditionné sur la transparence du marché. En l'occurrence, la donnée procure un sentiment de transparence aux vendeurs qui en consomment avec modération, mais floute la vision du marché des particuliers qui en consomment beaucoup en multipliant les supports de consultation.

5) **Usage des données et fixation des prix des logements**

La cinquième et dernière partie du rapport fait un pas supplémentaire dans le débat théorique et établit que la consommation de données a tendance à augmenter la performance économique des vendeurs. En l'occurrence, les plus grands consommateurs de données sont particulièrement nombreux parmi les vendeurs qui alignent leur prix sur les valeurs hautes du marché local et qui parviennent à faire aboutir leur vente dans ces conditions. De quoi nuancer l'effet vertueux de la donnée sur la dynamique des prix.

Bref, l'enquête par questionnaires établit combien l'usage des informations disponibles sur les marchés est situé sociologiquement et géographiquement. Elle montre aussi que la donnée n'a pas d'effet mécanique mais conditionné sur la transparence des marchés et qu'elle a un impact pas toujours modérateur sur les prix des logements.



Un vaste paysage informationnel, une diversité d'usages

Pour se forger un avis sur la valeur de leur bien, 86% des vendeurs de logements consultent des informations sur les prix immobiliers locaux pendant leur parcours de vente. Ceux-là peuvent avoir accès à une multitude d'informations. Ils peuvent par exemple consulter un magazine qui leur donnera le prix moyen au mètre carré dans leur quartier. Ils peuvent faire appel à un agent immobilier qui mobilisera son expérience, voire des références statistiques et/ou un outil d'estimation pour leur fournir un avis de valeur. Ils peuvent aussi consulter un site Internet et accéder par ce biais aux prix des biens en vente ou déjà vendus dans leur secteur. Ils peuvent même mobiliser simultanément tout ou partie de ces différents types et sources d'informations et les confronter les uns aux autres.

Suivre les vendeurs de logements dans leur expérience concrète aboutit à un premier constat : celui de la diversité du paysage informationnel dans lequel ils évoluent. Un second constat tient à la cohabitation, dans le même panier, d'informations de nature différentes, à commencer par le prix des biens déjà vendus, une norme en matière de statistiques immobilières, et le prix de mise en vente, souvent utilisé par les particuliers mais moins légitime auprès des spécialistes, car plus ou moins proche du prix finalement conclu.

Cette première partie étaye ces constats généraux, en rendant compte du chemin parfois sinueux que les vendeurs suivent, d'information en information, pour fixer le prix de mise en vente de leur logement. La suivante s'attachera à identifier plus particulièrement la place et le rôle de la donnée dans ce cheminement.

Le prix des logements : une information multi-dimensionnelle

L'information immobilière est multi-dimensionnelle. Rendre compte de son foisonnement et de la diversité des usages qui en est faite nécessite de la décomposer. Nous en explorerons les trois principales dimensions, et leurs combinaisons, à savoir :

- Le « **type** » d'information, entre estimation, prix d'un bien vendu, prix d'un bien en vente et prix moyen ou médian dans un secteur géographique;
- Le « **canal** » d'accès à l'information, entre canal physique, digital et papier ;
- Le « **support** », qui renvoie à l'identité du passeur d'informations, que ce soit un professionnel de l'immobilier, un site Internet, un membre de l'entourage, un journal ou magazine.

Finalement, chaque information sur les prix des logements mobilisée par les vendeurs sera définie comme un agencement spécifique de ces trois dimensions. De quoi qualifier et quantifier les informations mobilisées par les vendeurs pour mieux en appréhender les usages et les effets. De quoi également objectiver et cerner la place et le rôle spécifiques de la donnée dans le paysage informationnel, dans un contexte marqué par la démultiplication des informations qui s'offrent aux vendeurs, en particulier sur Internet.

Quatre types d'informations et une préférence pour les références individuelles de prix

Quatre « types » d'informations sur les prix immobiliers cohabitent au sein du paysage dans lequel évoluent les vendeurs de logements :

- **l'estimation de prix**, réalisée soit « à la main » par un professionnel de l'immobilier à partir de son expérience et de sa connaissance du terrain, soit par un outil automatique utilisé par le professionnel sollicité ou par le vendeur lui-même sur Internet ;
- **le prix des biens vendus**, disponible soit auprès des professionnels de l'immobilier, soit au moyen de cartes ou de listes de ventes interactives sur Internet (DVF en constitue la source principale depuis son ouverture en avril 2019), soit encore sous la forme d'un affichage de biens vendus (MeilleursAgents en fait un indicateur de performance et y encourage ses agences partenaires) ;
- **le prix des biens en vente**, disponible dans les annonces ;
- **le prix moyen ou médian au mètre carré**, proposé à la commune ou au quartier dans la plupart des cas.

De fait, pour approcher le prix possible de leur logement, les vendeurs ont recours, dans des proportions proches, à des estimations (36% d'entre eux s'y réfèrent), à des références de prix des biens vendus (35% des vendeurs en utilisent) et à des références de prix de biens en vente (32% des vendeurs) (Figure 1). Ils sont en revanche près de deux fois moins nombreux (18%) à utiliser des indicateurs synthétiques de prix moyens ou médians par mètre carré dans leur secteur.

Figure 1. Usage des types d'informations

	Vendeurs ayant recours aux types d'informations
Estimations	36%
Prix des biens vendus	35%
Prix des biens en vente (annonces)	32%
Prix au m² (moyens/médians)	18%

Ensemble des vendeurs de logements (n=993)

Remarque : la somme est supérieure à 100% car les vendeurs peuvent consulter plusieurs types d'informations à la fois.

Lecture : 32% des vendeurs s'informent sur les valeurs de marché (au moins en partie) en consultant le prix des biens en vente (annonces).

Parmi ces quatre types d'informations, les trois premiers sont certes issus de lignes de production différentes (Juillard 2019a) mais ont en commun de correspondre à des prix individuels et de renvoyer ainsi à des biens spécifiques. Ce sont largement les plus utilisés. Le quatrième type quant à lui absorbe la diversité des logements dans un périmètre donné pour en approcher une valeur moyenne ou médiane et établir un prix de comparaison. Il est significativement moins utilisé.

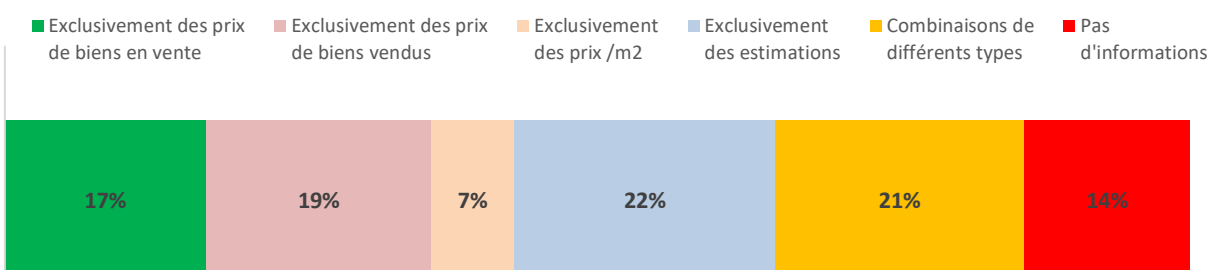
L'appétence des vendeurs pour les valeurs individuelles traduit d'abord le caractère important de l'estimation dans la perspective de vendre. Elle révèle ensuite le principal exercice auquel se prêtent les vendeurs pour fixer le prix de leur logement : la comparaison.

De fait, celle-là est établie à partir de deux types de valeurs étalons possibles : les valeurs déjà validées par le marché, à l'instar des prix des biens vendus et/ou les valeurs proposées à la mise en vente telles que les annonces des biens à vendre les indiquent. Concernant les annonces, il faut dire que leur place

privé dans le paysage informationnel n'est pas nouvelle, tel qu'en témoigne depuis longtemps le succès des journaux d'annonces. L'attrait pour les annonces rappelle également l'intérêt des particuliers pour les descriptions fines des logements disponibles sur le marché (photos, etc.).

Par ailleurs, 22% des vendeurs ont recours exclusivement à une ou plusieurs estimations pour approcher la valeur de leur logement, 19% se bornent à consulter des références de prix de biens déjà vendus, et 17% des prix de biens en vente. Cependant, 21% combinent ces types d'informations entre eux (Figure 2). Ce mode de consommation de l'information révèle une autre pratique courante : la combinaison des ressources informationnelles au sein de « paniers garnis » particulièrement divers selon les vendeurs.

Figure 2. Usages exclusifs et combinaisons des types d'informations



Ensemble des vendeurs de logements (n=993).

Lecture : 17% des vendeurs s'informent exclusivement en consultant le prix des biens en vente (annonces).

Quatre canaux de consultation et la règle du panier garni

Les vendeurs disposent par ailleurs de quatre canaux d'accès possibles à l'information sur les prix des logements :

- le **canal des professionnels** de l'immobilier (agents, mandataires, notaires) ;
- le **canal de l'entourage**, dont l'enquête permet de saisir le rôle persistant ;
 - ces deux premiers canaux formant le **canal « physique »** ;
- le **canal digital**, qui regroupe de très nombreux sites Internet et applications mobiles ;
- le **canal papier**, fait de journaux et de magazines.

Lorsqu'ils se renseignent sur la valeur possible de leur logement, 65% des vendeurs s'adressent à des professionnels, que ce soit des agents immobiliers, des mandataires ou des notaires. Le canal physique des professionnels est de loin le plus fréquemment emprunté pour accéder à l'information sur les prix, devant le canal digital, utilisé par 52% des vendeurs (Figure 3).

Figure 3. Usage des canaux d'accès à l'information sur les prix des logements

		Vendeurs ayant recours aux canaux d'informations
Canal physique	Professionnels	65%
	Entourage	39%
Canal digital	Internet	52%
Canal papier	Presse	20%

Ensemble des vendeurs de logements (n=993)

Remarque : La somme est supérieure à 100% car les vendeurs peuvent consulter plusieurs canaux à la fois.

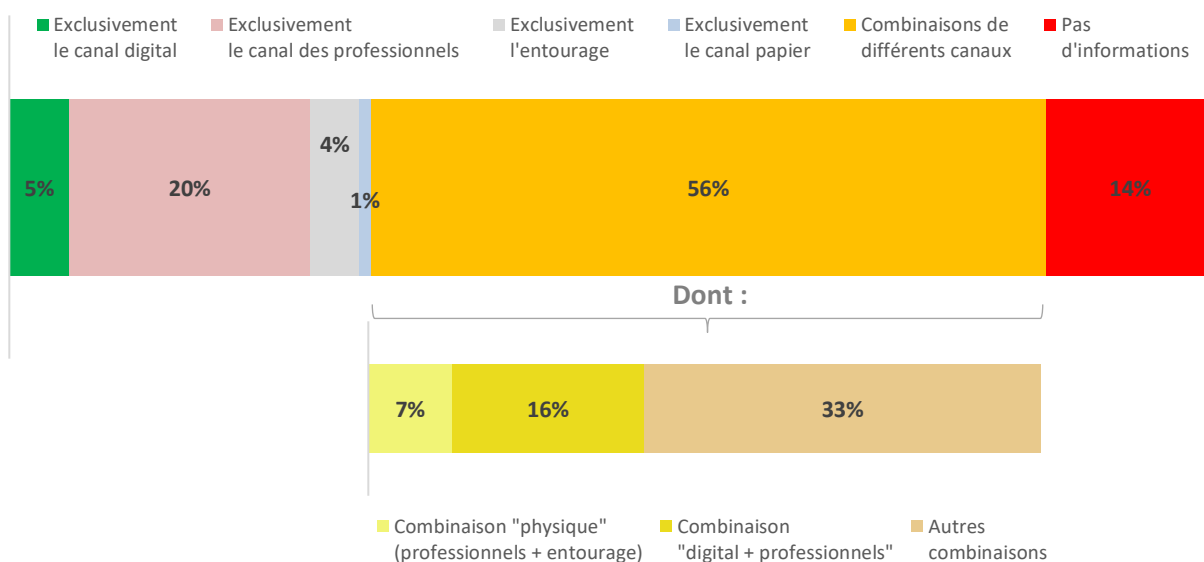
Lecture : 65% des vendeurs s'informent (au moins en partie) via des professionnels (agents immobiliers, mandataires, notaires).

Ici, plus encore que pour les types d'informations, la combinaison des canaux d'accès est la règle. En effet, 56% des vendeurs empruntent en même temps différents canaux pour accéder à l'information (Figure 4). En résulte une forte diversification de leur panier garni.

Ces résultats révèlent le rôle central des professionnels de l'immobilier dans l'accès à l'information. Ils rappellent que la valeur fonctionnelle des intermédiaires immobiliers s'apprécie notamment en regard de l'avantage informationnel obtenu auprès d'eux (S. Levitt et C. Syverson 2008). D'ailleurs, la première attente des vendeurs à l'encontre de ces

professionnels, à commencer par les agents immobiliers, est la détermination du prix de leur logement, devant la recherche d'acquéreurs et la négociation, le soutien logistique et juridique (Larceneux, Lefebvre, Simon 2014).

Figure 4. Usages exclusifs et combinaisons des canaux d'accès à l'information

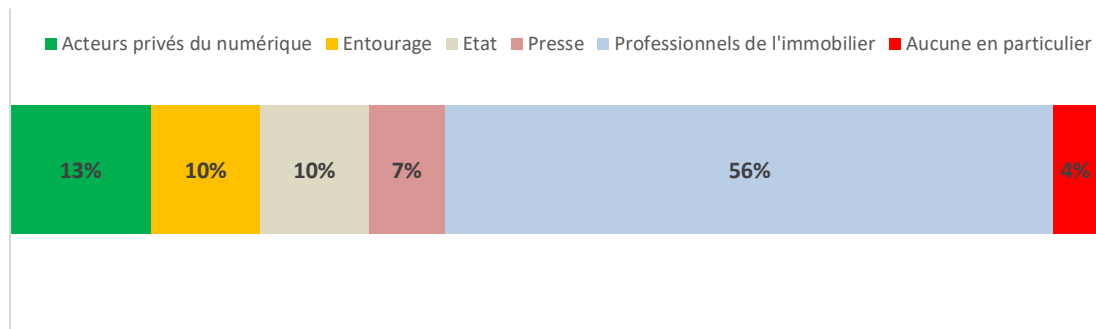


Ensemble des vendeurs de logements (n=993)

Lecture : 20% des vendeurs mobilisent uniquement des professionnels pour s'informer sur les valeurs de marché

Notre enquête établit en outre que les professionnels constituent la première source de confiance des vendeurs en matière d'informations sur les prix (56% des vendeurs donne priorité aux informations qu'ils délivrent), loin devant les acteurs du numérique (13%), l'Etat (10%) et l'entourage (10%), au coude à coude dans la hiérarchie (Figure 5).

Figure 5. Source prioritaire d'informations citée par les vendeurs



Vendeurs de logements ayant recours à l'information (n=653, question non-obligatoire)

Question : Parmi les différentes sources d'informations, laquelle vous a le plus influencée ?

Lecture : 10% des vendeurs informés considèrent l'entourage comme la source d'informations la plus décisive.

De son côté, le canal digital apparaît central dans l'accès à l'information mais il est rarement mobilisé seul. En effet, le recours à Internet a lieu dans seulement 5% des cas à titre exclusif (Figure 4). En revanche, le digital est combiné à d'autres canaux d'accès à l'information dans 47% des cas. En particulier, 16% des vendeurs recueillent des informations sur les prix des logements à la fois sur Internet et auprès de professionnels (Figure 4). Pour l'heure, l'usage de l'information disponible sur Internet et l'érosion potentielle de la rente informationnelle des professionnels qui en découle (Tucillo 1997) n'entame pas leur rôle pivot dans l'accès à l'information. Le canal digital ne remplace pas le canal physique mais s'offre en complément. Cette complémentarité s'illustre aussi dans la stabilité de la part de marché des professionnels en tant qu'intermédiaires de vente dans le contexte pourtant changeant de la « révolution numérique » (Vorms 2016).

Des supports mobilisés en quantité, un effet de l'offre digitale

En ce qu'ils articulent à la fois plusieurs types d'informations et plusieurs canaux d'accès, les comportements des vendeurs laissent une impression de dispersion. Cette impression est plus forte encore lorsque l'on tient compte de la quantité et de la diversité des supports (ou fournisseurs) d'informations que les particuliers utilisent.

Pour chaque canal d'abord, les vendeurs puisent souvent à différentes sources. Ainsi, 23% d'entre eux consultent plusieurs professionnels à la fois pour recueillir des références de prix, 16% s'adressent à plusieurs membres de leur entourage et 6% puisent des informations dans plusieurs journaux et magazines. Cette tendance s'accroît avec le digital puisque 37% des vendeurs surfent sur plusieurs sites Internet ou applications à la fois pour apprécier les prix de l'immobilier (Figure 6).

Figure 6. Propension des vendeurs à multiplier les supports selon le canal d'informations

		Vendeurs ayant recours à plusieurs supports selon les canaux d'informations
Canal physique	Professionnels	23%
	Entourage	16%
Canal digital	Internet	37%
Canal papier	Presse	6%

Ensemble des vendeurs de logements (n=993)

Remarque : La somme est différente de 100% car les vendeurs peuvent consulter plusieurs supports et plusieurs canaux à la fois.

Lecture : 23% des vendeurs s'informent en consultant plusieurs professionnels à la fois (agents immobiliers, mandataires, notaires).

Les canaux d'accès à l'information recoupent en partie les canaux d'accès au marché et les vendeurs dispersent leurs recherches d'informations comme ils dispersent leurs recherches d'acquéreurs potentiels. En effet, 47% des vendeurs sollicitent plusieurs professionnels pour mettre en vente leur logement et bien plus encore (57%) décuplent le potentiel du canal de mise vente entre particuliers en multipliant les points de contact à partir de supports dédiés (Figure 7).

L'ensemble rappelle l'opportunisme des vendeurs (Larceneux, Lefebvre, Simon 2014) et le caractère extensif de parcours de vente tendus vers le meilleur compromis possible (Bonneval 2012). Contrairement aux tendances identifiées dans certains travaux (Boulay, Blanke, Casanova Enault, Granié 2020 ; Kitchin, Lauriault 2014), on semble encore loin ici du rôle de réintermédiation du marché immobilier que l'on prête parfois aux fournisseurs de données sur Internet.

Y parviendra-t-on à l'avenir ? Rien ne permet de l'affirmer à ce stade, de même que rien ne permet (encore ?) de présager le retrait significatif des agences physiques face au développement des agences en ligne.

Cependant, « le marché du logement a bien été reconfiguré comme un flux continu de données par le déploiement des plateformes numériques » (Migozzi 2020) et le canal digital offre une diversité croissante de supports d'informations que certains vendeurs ne manquent pas d'exploiter.

D'après le recensement réalisé dans le cadre de cette étude, près de soixante sites Internet et applications mettent à disposition des références de prix en France (Figure 8). Ce vaste ensemble compte des sites d'annonces immobilières, des sites spécialistes de l'information immobilière, des

Figure 7. Propension des vendeurs à multiplier les points de contact selon le canal de mise en vente

		Vendeurs ayant recours à plusieurs points de contact par canal de mise en vente
Professionnels	Agents immobiliers, mandataires, notaires	47%
Entre particuliers	Supports dédiés (sites Internet, journaux)	57%
	Bouche-à-oreille	39%

Ensemble des vendeurs de logements (n=993)

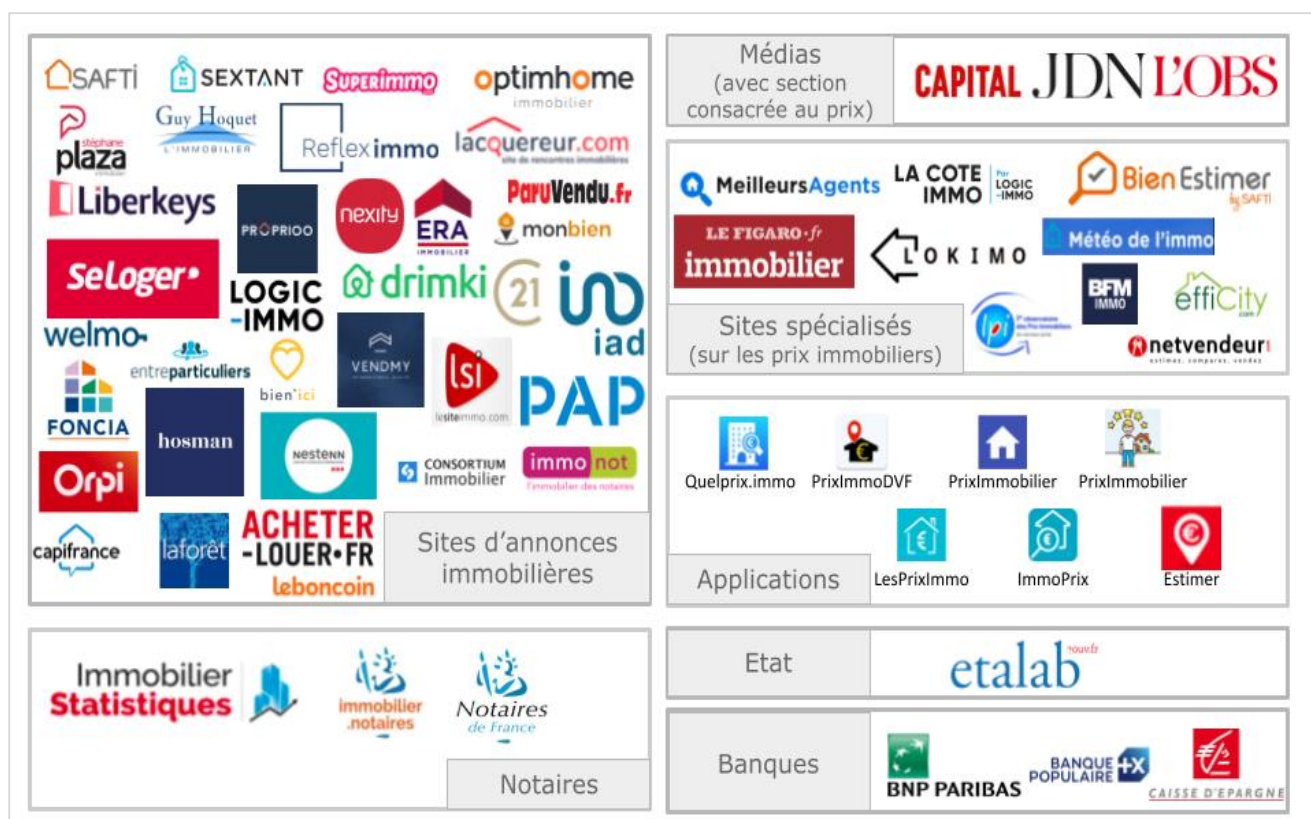
Remarque : La somme est différente de 100% car les vendeurs peuvent mobiliser plusieurs canaux et plusieurs points de contact à la fois.

Lecture : 47% des vendeurs mettent en vente leur logement en mobilisant plusieurs professionnels à la fois (agents immobiliers, mandataires, notaires).

applications dédiées, des médias, les portails du notariat (source historique de références de prix et de statistiques immobilières), des banques et le site d'Etalab (structure publique en charge de la mise en œuvre de l'ouverture de DVF, données de l'administration fiscale issues des actes authentiques de vente).

Figure 8. Sites Internet et applications fournissant des informations à destination des particuliers sur les prix des logements en France (marché de l'ancien, juin 2021)


Sont retenus ici les sites Internet et applications à destination des particuliers qui proposent des informations individuelles ou à l'échelle locale (au moins la commune) sur les prix des logements anciens pour toute la France Métropolitaine. Les sites d'annonces immobilières sont pris en compte seulement lorsqu'ils proposent une carte de prix ou un outil d'estimation automatique. De même, les sites de médias sont intégrés si l'information immobilière dépasse le contenu rédactionnel des articles (exemple : ajout de tableaux ou cartes de prix ayant vocation à être actualisées).



L'ensemble de ces sites et applications d'immobilier peuvent mêler les quatre types d'informations possibles (prix des biens en vente, prix des biens vendus, estimation de prix et prix moyen ou médian), certains se limitant à un seul type, d'autres à tout ou partie.

Une évolution récente tient à l'entrée de banques dans le paysage informationnel des vendeurs. Certains y voient le signal faible d'une nouvelle tendance à venir. Nous n'en présageons pas ici, faute d'éléments tangibles, mais y resterons attentifs à l'avenir, comme à une extension possible du marché en lien avec le marché du crédit immobilier et donc avec les enjeux du financement des achats immobiliers.

Bien plus déterminante à ce stade, une seconde évolution récente tient à l'ouverture de DVF. En effet, la mise à disposition de tous de cette manne d'informations dans sa version brute a suscité le développement d'au moins une dizaine de nouvelles interfaces de consultations à destination du



public. Le site AppDVF développé par Etalab est le premier. Il a marqué un tournant dans l'information aux particuliers, très nombreux à le consulter. La carte interactive développée et mise en service par MeilleursAgents dans la foulée de l'ouverture des données a offert la seconde interface de consultation, suivie par celle de SeLoger (Juillard 2020). Par ailleurs, une grosse poignée d'applications dédiées ont vu le jour, bien plus confidentielles mais non moins significatives de la démultiplication toujours croissante de l'information immobilière sur le canal digital. Précisons aussi que d'autres sites et applications puisent dans DVF sans le mettre directement en avant et plutôt en mélangeant cette source à d'autres données pour fournir par exemple des estimations automatiques en ligne.

De manière générale, les sites et applications d'immobilier enregistrent une vingtaine de millions de visiteurs uniques par mois. Selon les dernières données disponibles, leboncoin enregistre à lui seul plus de 12 millions de visiteurs uniques par mois, loin devant le second de la hiérarchie, SeLoger avec environ 6,5 millions de visiteurs uniques². Ce score d'audience de loin inégalé y compris lorsque la plateforme ne proposait pas encore d'outils d'estimation, ni d'autres références de prix que celles contenues dans les annonces immobilières, révèle un point important : le trafic sur les sites et applications d'immobilier est avant tout un trafic de consultation d'annonces. Premier site historiquement dédié à l'information immobilière, MeilleursAgents a lui-même accru son audience (et consolidé son modèle d'affaires) en publiant des annonces. A l'inverse, leboncoin a fait un pas dans le sens de la mise à disposition d'informations immobilières en proposant un outil d'estimation automatique depuis fin 2020.

Quoi qu'il en soit, l'information attire à elle et constitue même un appât pour capter les internautes et les redistribuer aux agences partenaires sous forme de prospects qualifiés. On l'a vu, si seulement 5% des vendeurs empruntent exclusivement le canal digital pour s'informer (Figure 4), 47% l'utilisent concomitamment à d'autres. Au total donc, 52% des vendeurs consultent des informations sur les prix à travers des sites Internet et applications d'immobilier.

S'informer sur les prix des logements : une pratique à intensité variable

Comme nous l'avons établi plus haut, 86% des vendeurs mobilisent des informations sur les prix des logements pendant leur parcours de vente (voir Encadré 1 pour une description des 14% restants).

En s'appuyant sur un score développé spécialement pour les besoins de cette recherche (Encadré 2), il est possible de mesurer pour chaque vendeur l'intensité du recours à l'information immobilière. En l'occurrence, 46% des vendeurs consomment « peu » d'informations, c'est-à-dire ont recours à une information peu diverse en termes de types d'informations et/ou de canaux d'accès. A l'inverse, 40% consomment « beaucoup » d'informations, c'est-à-dire diversifient à la fois le type d'informations consultées et le canal pour y accéder. Par ailleurs, 46% (9% + 37% : Figure 9) des vendeurs empruntent un nombre réduit de supports (de 1 à 3) et 40% (31% + 9% : Figure 9) en utilisent de plus nombreux (4 et plus).

En mobilisant à leur guise un ou plusieurs types d'informations sur un ou plusieurs canaux et à travers un ou plusieurs supports, les vendeurs de logements affichent une très grande hétérogénéité dans leurs pratiques de documentation sur les marchés du logement. Cette hétérogénéité profite de la

² Internet Global, Médiamétrie-Netratings, octobre 2019.

diversification à l'œuvre aujourd'hui dans le secteur de l'intermédiation de la donnée immobilière (Figure 8). Plus que jamais, la norme est celle du panier garni d'informations.

Figure 9. Degré d'usage de l'information sur les prix par les vendeurs de logements

		Pourcentage	Dont :	Pourcentage
Score d'usage de l'information	Beaucoup d'informations	40%	Beaucoup de supports	31%
			Peu de supports	9%
	Peu d'informations	46%	Beaucoup de supports	9%
			Peu de supports	37%
	Pas d'informations	14%		
	Total		100%	

Ensemble des vendeurs de logements (n=993)

Encadré 1. Que dire des 14% de vendeurs qui ne consomment pas d'information sur les prix ?

14% des vendeurs déclarent ne pas avoir eu recours à des informations sur le marché résidentiel local, y compris celles fournies éventuellement par les intermédiaires ou leur entourage, pour établir le prix de mise en vente de leur logement. Comment le comprendre ?

D'abord, il faut remarquer que la non-utilisation d'informations sur les prix du marché local ne signifie pas une non-utilisation absolue d'informations économiques. Les indicateurs de marché constituent une connaissance orientée « présent ». Mais les vendeurs peuvent aussi fonder l'estimation de la valeur de leur logement sur des informations issues du « passé » en s'appuyant par exemple sur le prix auquel ils ont acheté leur bien ou encore le montant des travaux réalisés pour le modifier. Ils peuvent par ailleurs être orientés « futur » en élaborant un prix de mise en vente en fonction du projet d'achat qu'ils envisagent par la suite le cas échéant.

Dans les faits, parmi les 14% qui se n'appuient pas sur des références de prix du marché local pour estimer la valeur de leur bien, la moitié (7%) s'appuie en effet sur ces informations économiques d'un autre type, à savoir le prix auquel ils ont acheté le logement dans le passé et/ou le coût des travaux réalisés et/ou le prix auquel ils projettent d'acheter un nouveau logement.

Que dire finalement des 7% restants ? De ces vendeurs de logements qui déclarent n'avoir eu recours à aucune information économique pour estimer la valeur de leur logement, que cette information soit orientée « marché local », « passé » ou « futur » ? Comment cerner cette catégorie très minoritaire mais néanmoins étonnante ? Notre enquête permet d'en approcher le portrait. Apparaît alors l'image d'une vente en zone peu dense effectuée plutôt à travers le bouche-à-oreille et portant sur des biens de qualité plutôt inférieure aux standards locaux (à refaire ou à rénover et proposés à des prix relativement bas).

Encadré 2. La méthode de construction du score d'usage de l'information sur les prix

Mesurer la consommation d'informations que mobilisent les vendeurs de logements nécessite d'objectiver à la fois la « diversité » et la « quantité » mobilisées. En reprenant la tripartition établie dans ce rapport, on peut dire que la diversité renvoie aux « types » et aux « canaux » d'accès. La quantité renvoie quant à elle au nombre de « supports » consultés par les vendeurs.

Sur cette base, un « score d'usage de l'information » a été bâti en deux étapes. Dans un premier temps, une mesure de la « diversité » a été établie en repérant le nombre de couples « types d'informations / canal d'accès » auxquels les vendeurs ont recours (Figure 10). Par construction, cette première mesure varie théoriquement entre 0 et 16 selon que le vendeur mobilise aucun ou l'ensemble de ces couples.

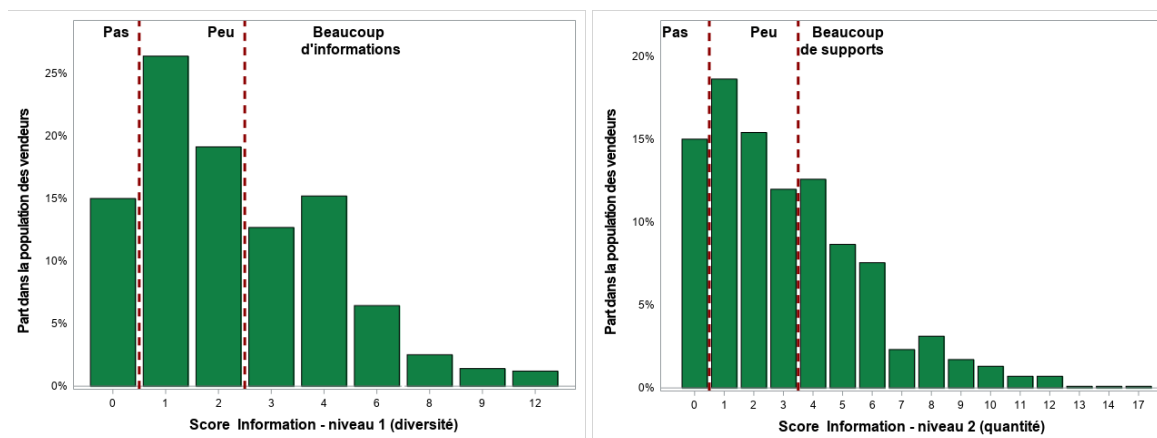
Figure 10. Les coefficients pour la mesure du versant « diversité » du score information

Canal ▼	Type ►	Prix des biens en vente (annonces)	Prix des biens vendus	Estimations	Prix au m ² (moyens ou médians)
Canal physique	Professionnels	1	1	1	1
	Entourage	1	1	1	1
Canal digital	Internet	1	1	1	1
Canal papier	Presse	1	1	1	1

Dans les faits, la mesure varie entre 0 et 12 couples d'informations (Figure 11). Elle permet de distinguer la population des vendeurs en trois groupes selon qu'ils mobilisent « pas », « peu » ou « beaucoup » d'informations (Figure 11 – gauche).

Dans un deuxième temps, une mesure de la quantité peut être établie. Elle fait la somme du nombre de supports mobilisés (nombre de professionnels, de membres de l'entourage, de journaux et sites Internet différents, etc.). Dans les faits, ce score varie entre 0 et 17 (Figure 11 - droite). Cette distribution permet de distinguer trois groupes : ceux qui mobilisent « pas », « peu » ou « beaucoup » de supports.

Figure 11. La distribution des versants « diversité » et « quantité » du score d'usage de l'information



La confrontation de ces deux catégories permet d'obtenir une variable score qui mesure l'intensité du recours à de l'information sur les prix en distinguant 5 classes : « pas d'informations » (donc pas de supports), « peu d'informations et peu de supports », « peu d'informations et beaucoup de supports », « beaucoup d'informations et peu de supports », « beaucoup d'informations et beaucoup de supports » (Figure 9).

La place de la « donnée » dans le paysage informationnel : centrale mais variable

Si le recours à l'information immobilière en général est un sujet majeur du point de vue de la théorie économique, cette recherche se focalise plus précisément sur les usages du type spécifique d'informations associé au contexte actuel de la transition numérique : la « donnée ». Fait surprenant, cet objet omniprésent dans les discours n'a en réalité qu'une définition implicite, y compris dans le champ scientifique qui consacre encore peu de réflexions aux usages concrets de la donnée en situation de marché. Il fallait donc ici poser une définition et apporter une réponse à la question suivante : parmi les informations sur les prix dont disposent les vendeurs, lesquelles renvoient à ce que l'on appelle une donnée de prix ?


Qu'est-ce que la donnée immobilière du point de vue des particuliers ?

Un vendeur qui consulte un journal de petites annonces immobilières consomme de l'information sur les prix. Mais consomme-t-il de la donnée ? Si la réponse était affirmative, tout ou presque serait « donnée ». Il serait alors impossible d'isoler une population spécifique de vendeurs caractérisée par sa consommation de données. A l'inverse, si l'on posait une définition de la donnée en se référant aux politiques publiques d'ouverture des données (Open Data), alors le terme renverrait à l'usage de fichiers informatiques assez lourds dont l'interprétation nécessite une certaine dose de compétences statistiques. Dans ce cas, la donnée accessible aux particuliers serait nulle part car, dans les faits, les particuliers ne vont pas directement puiser dans des fichiers informatiques mais recourent à des intermédiaires de l'information qui réalisent le travail de mise en forme statistique à leur place.

Dès lors, qu'est-ce qu'une donnée ? Et qu'en est-il précisément au regard des pratiques effectives des ménages ? Pour fixer une définition, reprenons la grille de lecture proposée précédemment. Cette grille distingue deux premières dimensions : le type d'informations d'une part, le canal d'accès d'autre part. Sur cette base, nous considérons comme données les informations sur les prix des biens vendus, les estimations de prix, et les prix moyens au mètre carré recueillis sur Internet spécifiquement. Il s'agit en somme de toute l'information sur les prix présente sur la toile, à l'exception notable des références issues des annonces de biens en cours de vente (Figure 12).

Figure 12. La donnée dans le paysage informationnel sur les prix des logements

Canal ▼	Type ►	Prix des biens en vente (annonces)	Prix des biens vendus	Estimations	Prix au m ² (moyens ou médians)
Canal physique	Professionnels	-	-	-	-
	Entourage	-	-	-	-
Canal digital	Internet	-	x	x	x
Canal papier	Presse	-	-	-	-



Pourquoi distinguer précisément ces trois situations de toutes les autres ? La différence tient à trois conditions. D'abord, c'est une question de production : les prix de biens vendus, les estimations et les prix moyens ou médians au mètre carré accessibles *via* le canal digital se basent sur une exploitation systématique de fichiers informatiques structurés en vue d'une connaissance des marchés. A l'inverse, les prix des biens en vente tel que les particuliers y ont accès dans les annonces immobilières affichées sur la toile correspondent à autant d'informations éloignées d'une logique d'ingénierie statistico-informatique des marchés, sinon *ex post*, c'est-à-dire dans la perspective de produire des données à proprement parler comme de nombreux fournisseurs d'informations immobilières s'y emploient.

Ensuite, les prix des biens vendus, les estimations et les prix moyens (ou médians) au mètre carré renvoient à une logique de marché. Ils ont en commun de se référer aux transactions déjà conclues et, ce faisant, à des prix déjà validés par le marché. Inversement, si les prix des biens en vente offrent une actualisation possible des prix conclus, sorte de signal sur la tendance du marché, ils ne constituent qu'une proposition de valeur. Leur caractère provisoire les rend d'ailleurs précisément moins légitimes auprès non seulement des producteurs de données (Juillard 2019 b et c) mais des acteurs du marché, à commencer par les agents immobiliers qui déplorent les biais de compréhension qu'ils instaurent chez les particuliers qui se limitent à ce type d'information.

Enfin, les prix de biens vendus, les estimations et les prix moyens ou médians au mètre carré disponibles sur Internet spécifiquement obéissent à une condition particulière de transmission : le lien direct entre l'information et l'utilisateur, sans autre forme d'intermédiation, par le canal physique par exemple.

Quelle place la donnée tient-elle dans le panier garni d'informations des particuliers?

Une fois cette définition posée, une autre difficulté méthodologique apparaît. Comment isoler les vendeurs qui consomment de la donnée sachant que la norme concernant la collecte d'informations est celle du panier garni ? Que dire d'un vendeur qui, pour estimer son logement, demande à la fois une estimation à un agent immobilier et consulte le prix moyen au mètre carré dans son secteur sur un site Internet spécialisé ? Peut-il être catégorisé comme consommateur de données ?

Si nous pensions au commencement de ce travail de recherche qu'une approche binaire était valable, les résultats empiriques nous ont conduit à complexifier l'approche. Dans les faits, chaque vendeur de logement peut avoir, dans son panier d'informations, une part plus ou moins grande de données. En termes statistiques, cela signifie que le phénomène ne doit pas être appréhendé à travers une variable catégorielle mais à travers une variable numérique.

De façon comparable au score d'usage de l'information, nous avons donc construit un « score d'usage de la donnée » en fonction à la fois de la diversité et de la quantité de données mobilisée par chaque répondant au questionnaire (Encadré 3). Cette variable numérique a ensuite été découpée en classes pour distinguer ceux qui ne consommaient aucune donnée de ceux qui en avaient une consommation limitée ou au contraire importante.

Encadré 3. La méthode de construction du score d'usage de la donnée sur les prix

En suivant la même méthode que pour le score d'usage de l'information (Encadré 2), un « score d'usage de la donnée » a été bâti en deux étapes. D'abord en repérant la diversité des données mobilisées (Figure 13). Conformément à notre définition de la donnée, trois possibilités, toutes renvoyant au canal digital, s'offrent aux vendeurs. Par construction, la première mesure varie ainsi entre 0 et 3.

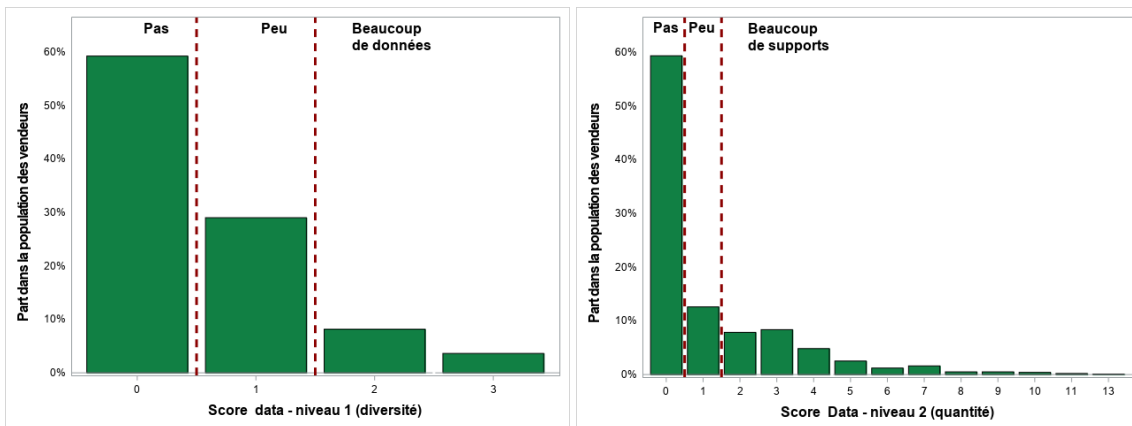
Figure 13. Les coefficients pour la mesure du versant « diversité » du score d'usage de la donnée

Canal ▼	Type ►	Prix des biens en vente (annonces)	Prix des biens vendus	Estimations	Prix au m ² (moyens ou médians)
Canal physique	Professionnels	0	0	0	0
	Entourage	0	0	0	0
Canal digital	Internet	0	1	1	1
Canal papier	Presse	0	0	0	0

Cette mesure permet de distinguer la population des vendeurs en trois groupes selon qu'ils mobilisent « pas », « peu » ou « beaucoup » de données. Dans les faits, près de 60% des répondants de notre échantillon (n=993) obtiennent un score nul. Les près de 40% restants peuvent dès lors être considérés comme des vendeurs consommant au moins en partie de la donnée sur les prix (Figure 14 - gauche).

Ensuite, une mesure de la quantité a été établie. Elle fait la somme des sites Internet et/ou applications fréquentés par un vendeur dans les moments où il se renseigne sur les prix. En pratique, ce score varie entre 0 et 13 (Figure 14 - droite). Cette distribution permet à son tour de distinguer trois groupes : ceux qui mobilisent « pas », « peu » ou « beaucoup » de supports.

Figure 14. La distribution des versants « diversité » et « quantité » du score d'usage de la donnée



La confrontation de ces deux catégories permet d'obtenir une variable score qui caractérise l'intensité de l'usage de la donnée en distinguant 5 classes : « pas de données » (donc pas de supports), « peu de données et peu de supports », « peu de données et beaucoup de supports », « beaucoup de données et peu de supports », « beaucoup de données et beaucoup de supports ».

2 vendeurs sur 5 ont recours à des données sur les prix

38% des vendeurs de logements (11% + 27% : Figure 15) consomment des données pendant leur parcours de vente. Parmi eux, 27% en mobilisent « peu » et 11% en mobilisent « beaucoup ».

Comme de manière générale en matière d'information, les vendeurs ont généralement recours à une multiplicité de supports (sites Internet et applications). Ils sont 25% dans ce cas (16% + 9% : Figure 15), contre 13% à en utiliser un seul (11% + 2% : Figure 15). Ainsi, non seulement la donnée prend place dans un vaste paysage informationnel, tel que nous l'avons déjà établi, mais la donnée appelle la donnée et ne se consomme pas sans la consultation de multiples supports digitaux.

Figure 15. Degré d'usage de la donnée parmi les vendeurs de logements

		Pourcentage	Dont :	Pourcentage
Score d'usage de la donnée	Beaucoup de données	11%	Beaucoup de supports	9%
			Peu de supports	2%
	Peu de données	27%	Beaucoup de supports	16%
			Peu de supports	11%
	Pas de données	62%		
	Total		100%	

Ensemble des vendeurs de logements (n=993)

Remarquons pour finir que 62% des vendeurs n'ont pas recours à la donnée (Figure 15) et que parmi ceux-là, 14% ne consultent aucune information sur les prix (Figure 9). Ce faisant, environ 48% des vendeurs de logements ne mobilisent pas de données alors qu'ils mobilisent de l'information pour approcher le prix de leur logement (Figure 16).

Figure 16. Les modalités d'usage de l'information immobilière parmi les vendeurs de logements

Population des vendeurs de logements	86%	38%
	des vendeurs utilisent des informations sur les prix	des vendeurs utilisent des informations sur les prix dont tout ou partie sous forme de données
	48%	des vendeurs utilisent des informations sur les prix sans utiliser de données
	14% n'en n'utilisent pas	7% utilisent d'autres types d'informations
		7% n'utilisent aucune information économique

Ensemble des vendeurs de logements (n=993)



Une consommation inégale de données sur les prix

Revenons désormais à la problématique centrale de ce rapport, à savoir le lien entre la mise à disposition de données et la transparence des marchés. Il existe un présupposé important lorsque ce lien de causalité est établi. Celui-ci renvoie à l'idée que, lorsque des données existent, les acteurs économiques s'en saisissent forcément.

Une étude des pratiques concrètes permet de tester empiriquement ce présupposé. En l'espèce, il ressort de la partie précédente que la donnée est bien utilisée par un nombre significatif de vendeurs mais non par la totalité. Ce que nous allons voir maintenant c'est que, par ailleurs, le profil des usagers est spécifique et que la consommation de données sur le prix des logements dans le cadre d'un parcours de vente immobilière est une pratique socialement et géographiquement située. Ainsi, l'idée d'un lien systématique entre ouverture des données et transparence des marchés doit être nuancée, en tout cas à ce stade de l'histoire du fonctionnement des marchés résidentiels en France et de leur tournant numérique.

L'effet de l'âge et de la position sociale

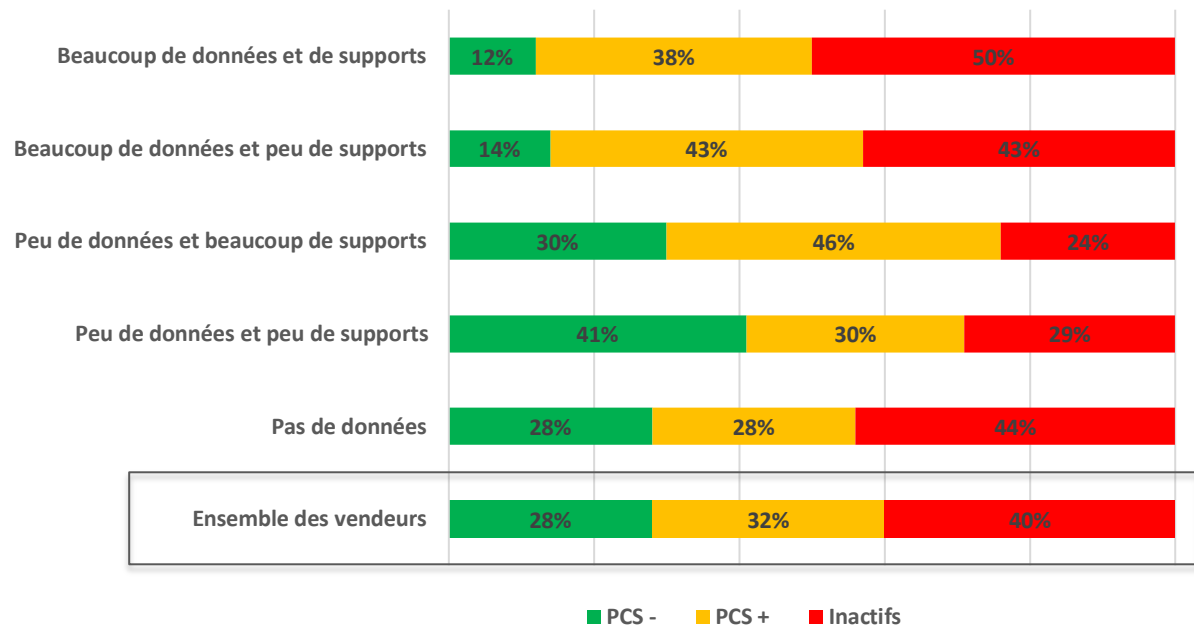
La consommation de données est une pratique liée à la position sociale du vendeur. En l'occurrence, la consommation de données renvoie plutôt aux catégories socioprofessionnelles supérieures (cadres, chefs d'entreprises, etc.), ces « PCS + » ayant tendance à être sur-représentés parmi les vendeurs qui consomment beaucoup de données (Figure 17).

De leur côté, les catégories sociales inférieures (ouvriers, employés) sont plutôt sur-représentées parmi les vendeurs qui consomment peu ou pas de données. Ce constat a déjà été établi par plusieurs travaux de recherches qui tendent à conclure que l'évolution du paysage informationnel en faveur de la donnée contribue à renforcer la ségrégation informationnelle (Boeing 2019, *idem*), de la même manière qu'elle peut exacerber la ségrégation résidentielle (Corbett, Loukissas 2019 ; Smith, Munro, Christie 2016 cités par Boulay, Blanke, Casanova Enault, Granié 2020). Il faut dire que comme l'établit l'Insee, les personnes les moins diplômées et aux revenus modestes font partie des populations les moins bien équipées et les moins compétentes au plan numérique (Leglaye, Rolland 2019).

Cependant, il ressort de notre enquête que les catégories socioprofessionnelles inférieures ne se détournent pas complètement de la donnée. Elles en font plutôt un usage limité en consultant un seul type, sur un seul canal ou un seul support (site Internet ou application).

Un des résultats originaux de l'enquête est la surreprésentation parmi les plus grands consommateurs de données de la catégorie des « inactifs ». Parmi ces derniers, ce ne sont pas tant les étudiants et les chômeurs que les retraités qui consultent des données de prix lorsqu'ils veulent évaluer le logement qu'ils souhaitent vendre. Ainsi, les plus de 65 ans ont tendance à consommer beaucoup de données, tandis que les moins de 35 ans tendent à en consulter peu (Figure 18). Finalement, la propension à recourir aux données dépend non seulement de la position sociale mais aussi de l'âge.

Figure 17. Consommation de données et catégorie socioprofessionnelle



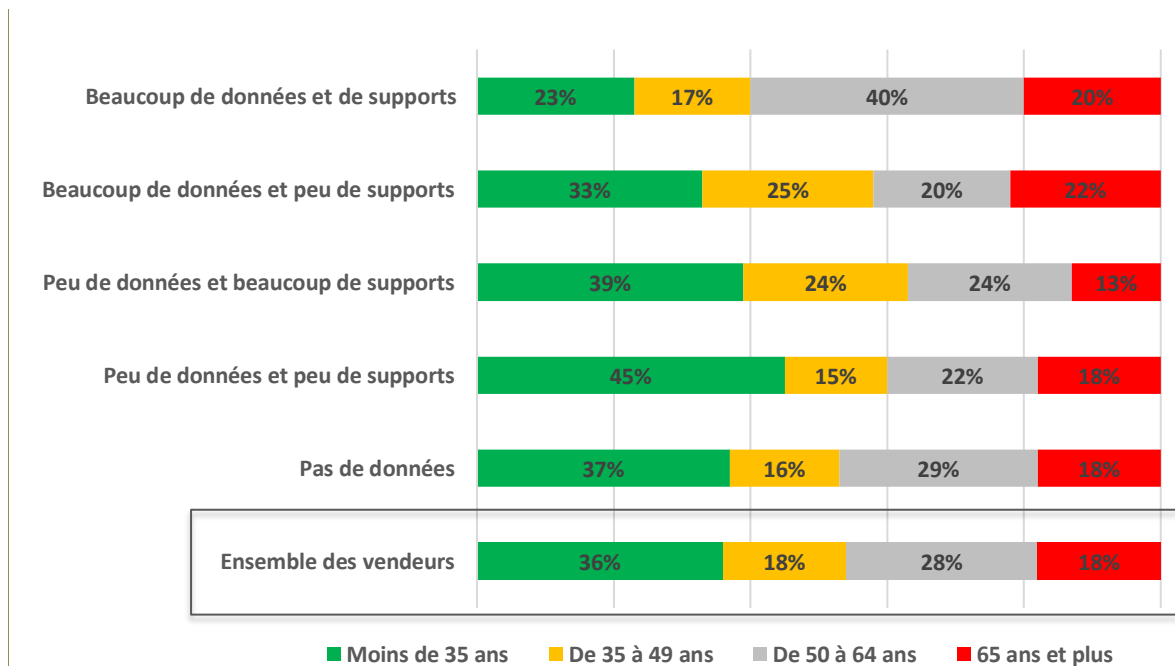
Ensemble des vendeurs de logements (n=993)

Pourcentages en ligne - Statistique du Khi-deux : DDL=8 / Valeur = 50,4 / Probabilité <0,0001

Catégories socioprofessionnelles : voir méthode de construction en annexe.

Lecture : 28% des vendeurs de logements appartiennent aux catégories sociales inférieures (PCS-). Ce chiffre passe à 12% lorsque l'on considère les vendeurs qui utilisent beaucoup de données sur les prix à travers de nombreux supports.

Figure 18. Consommation de données et classe d'âge



Ensemble des vendeurs de logements (n=993)

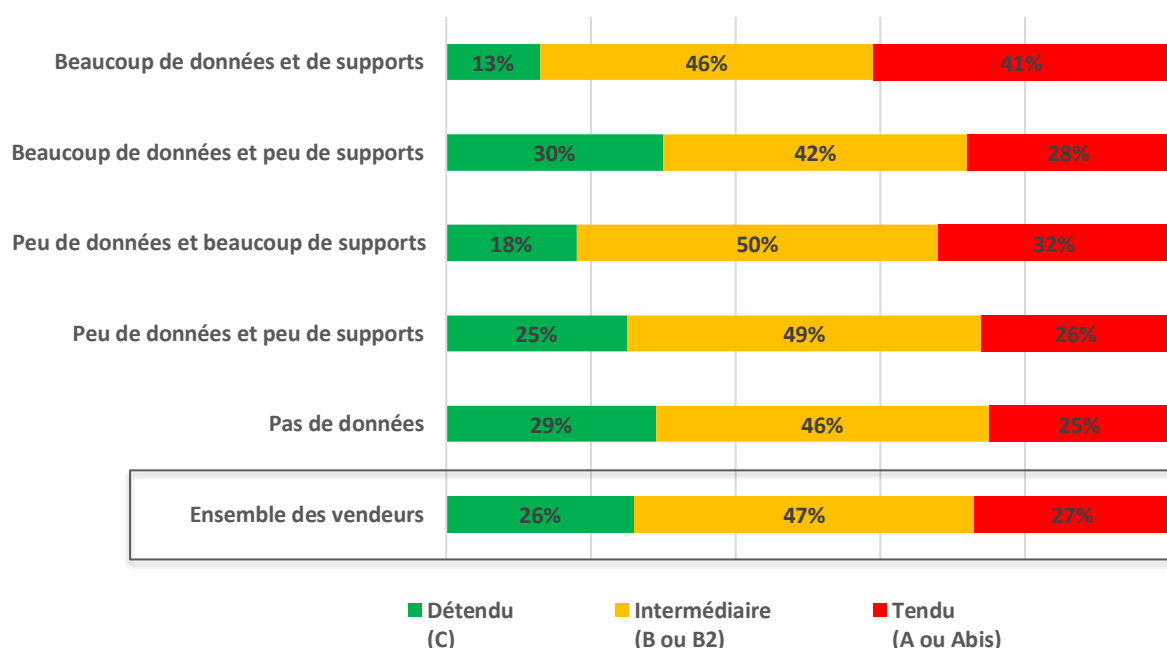
Pourcentages en ligne - Statistique du Khi-deux : DDL=8 / Valeur = 50,4 / Probabilité <0,0292

Une consommation de données qui augmente avec la tension du marché local

Si l'on se penche désormais moins sur les aspects sociodémographiques et plus sur les variables géographiques, on constate un lien statistique significatif entre la localisation du bien vendu et le recours à la donnée de prix. Plus précisément, il existe une relation croissante entre le niveau de tension du marché local (ici appréhendé au sens du zonage officiel « ABC »³) et le score d'usage de la donnée.

Dans les zones A ou Abis, les vendeurs qui recourent à beaucoup de données sur beaucoup de supports sont surreprésentés (Figure 19). Dans ces zones les plus tendues, les vendeurs qui consomment peu de données mais sur un nombre important de supports sont également surreprésentés. A l'inverse, les répondants ayant recours à aucune donnée sont davantage représentés dans les zones détendues, tout comme d'ailleurs ceux qui se passent d'informations sur les prix des logements en général.

Figure 19. Consommation de données et tension du marché local



Ensemble des vendeurs de logements (n=993)

Pourcentages en ligne - Statistique du Khi-deux : DDL=8 / Valeur = 19,87 / Probabilité = 0,0108

Remarque : zonage « ABC » reconstitué par les auteurs à partir de la localisation du bien déclarée dans le questionnaire par le vendeur.

³ <https://www.ecologie.gouv.fr/zonage-b-c>

Des inégalités d'usages cumulatives

Les effets des critères sociodémographiques et géographiques sur la consommation de données n'agissent pas indépendamment. Ils se cumulent.

Pour le comprendre, il est nécessaire de délaisser les analyses bivariées et d'opter pour des méthodes statistiques multivariées. En particulier, l'analyse des correspondances multiples (ACM) permet de détecter les liens entre les modalités d'une série de variables catégorielles et de représenter leurs proximités dans un plan. Il est ainsi possible de visualiser des grappes de modalités empiriquement liées et de savoir comment se positionne une variable d'étude (ici le score d'usage de la donnée) parmi l'ensemble des variables testées et leurs modalités.

En l'occurrence, l'ACM répartit les variables sociodémographiques et géographiques de notre questionnaire autour de deux axes (Figure 20).

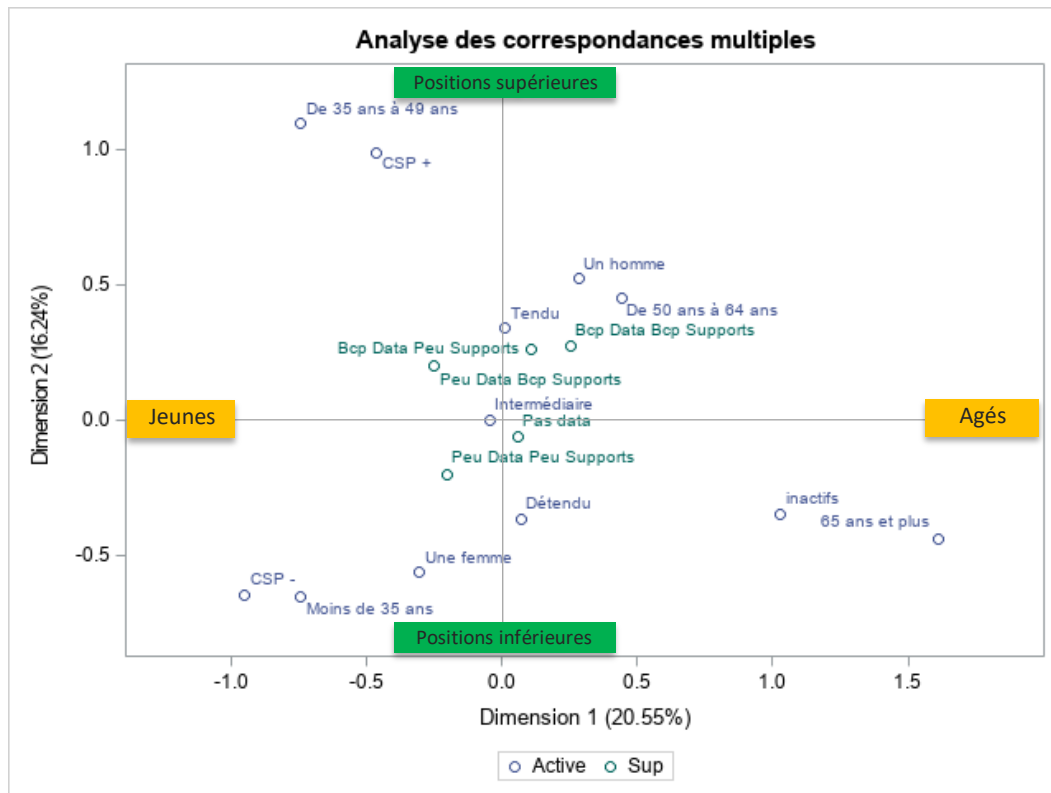
- Le premier (horizontal) peut être interprété comme une représentation de la maturité du vendeur (âge). Aux extrêmes, on trouve la modalité « moins de 35 ans » à gauche et la modalité « 65 ans et plus » à droite.
- Le deuxième axe (vertical) combine des variables sociologiques (PCS) et géographiques (tension du marché local). On repère ici que les vendeurs les plus élevés socialement ont aussi un patrimoine localisé dans des espaces valorisés et tendus sur le plan résidentiel. Cumulées, ces deux variables indiquent finalement la place des vendeurs de logements dans la hiérarchie « socio-spatiale ». En bas, les modalités « PCS – » et « Zone C » renvoient à des positions inférieures. En haut, les modalités « PCS + » et « Zone A » font référence à des profils supérieurs. Entre ces positions extrêmes s'organise un ensemble de gradients intermédiaires.

Repérons désormais la place qu'occupent les modalités de la variable score (ici présentée en tant que variable supplémentaire). En l'occurrence, celles-ci se distribuent assez bien dans l'axe vertical. Ainsi, plus un vendeur est positionné haut dans la hiérarchie socio-spatiale, plus sa propension à consommer des données augmente. Les modalités affichent aussi une certaine variation sur l'axe horizontal. Le pôle « âgés » attire davantage les modalités « beaucoup de données ». On retrouve ici l'effet de l'âge sur l'usage que nous pointons plus haut.

Finalement, ce que montre l'analyse factorielle, c'est un ensemble d'effets cumulatifs. La catégorie socioprofessionnelle se recoupe avec la localisation géographique et ces deux éléments se cumulent à l'âge dans la structuration des comportements d'utilisation de la donnée de prix.

La consommation de données de prix croît globalement avec l'âge, mais cet effet est atténué si dans le même temps la position sociale baisse. De même, l'usage de données augmente avec la tension du marché local mais, à tension équivalente, elle peut décroître avec l'âge du vendeur.

Figure 20. Liens entre consommation de données et profils des vendeurs (analyse factorielle)



Ensemble des vendeurs de logements (n=993)

Pour revenir à la question du rôle de la donnée dans la transparence des marchés, on peut conclure que quand bien même la donnée serait un gage de transparence, son usage reste dans les faits limité à certains profils de vendeurs et de territoires. La transparence des marchés trouve donc une première limite dans l'état actuel des pratiques des ménages. C'est une question d'usage des données, plus encore que d'offre de données.

Consommation de données et sentiment de transparence des marchés : un lien sous conditions

La deuxième postulat implicite du raisonnement reliant donnée et transparence des marchés renvoie à l'idée que l'usage d'informations sur les prix (notamment sous la forme de données) se traduirait mécaniquement par une bonne connaissance des marchés. Pour le tester, notre enquête s'est penchée sur le sentiment de concordance ou de discordance que les vendeurs tiraient de leur consommation de données. Elle montre que l'usage de la donnée tend bel et bien à renforcer le sentiment de concordance informationnelle mais jusqu'à un certain point seulement et que le recours à une quantité trop importante de données peut au contraire générer un sentiment de flou chez les vendeurs de logements.

Un impact conditionné sur le sentiment de transparence

A 84%, les vendeurs jugent concordantes les informations auxquelles ils ont eu accès sur les prix des logements (Figure 21). Recourir à un professionnel de l'immobilier pour assurer la transaction n'y change rien : les particuliers qui confient la vente de leur bien à un professionnel entretiennent ni un sentiment plus fort ni un sentiment moins fort de concordance informationnelle⁴. Du point de vue des vendeurs, l'intermédiation professionnelle n'a donc d'effet ni sur la transparence des marchés ni sur leur opacité.

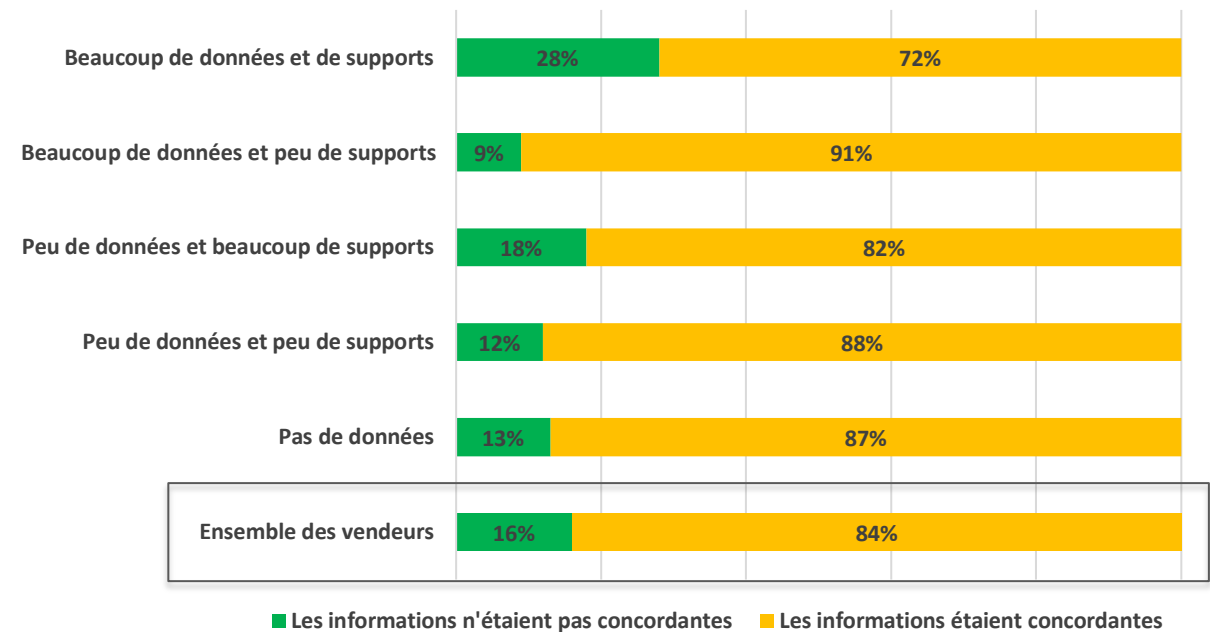
La proportion de vendeurs entretenant un sentiment de concordance des prix ne varie pas non plus avec la consommation d'informations en général⁵. En revanche, elle change précisément avec la consommation de données. Globalement en effet, consommer des données procure un sentiment de concordance (Figure 21). Cependant, ce sentiment de concordance est conditionné à la quantité de données recueillies : il domine lorsque les données sont consultées sur un nombre limité de sites Internet et applications et s'estompe, jusqu'à laisser place à une impression de flou, avec la multiplication des différents types de données mobilisés mais surtout avec la multiplication des sites Internet et applications utilisés.

Ainsi, les particuliers qui consomment beaucoup de données (de types différents) sur beaucoup de supports sont à la fois sous-représentés parmi les vendeurs qui entretiennent un sentiment de concordance informationnelle (ils sont 72% contre 84% pour l'ensemble des vendeurs), et plus fortement surreprésentés parmi les vendeurs qui considèrent que les informations qu'ils ont recueillies sont discordantes (ils sont 28% contre 16%) (Figure 21).

⁴ Le croisement entre la variable de concordance et la variable relative à l'intermédiaire de la vente affiche un test du khi-deux non significatif, ce qui indique une absence de lien statistique.

⁵ Là aussi, le croisement entre la variable de concordance et le score d'usage de l'information affiche un test du khi-deux non significatif.

Figure 21. Consommation de données et sentiment de concordance de l'information sur les prix



*Vendeurs de logements ayant recours à l'information (n=665, question non-obligatoire)
 Pourcentages en ligne - Statistique du Khi-deux : DDL=4 / Valeur = 13,7 / Probabilité = 0,008
 Question : Les informations sur les prix étaient-elles concordantes entre les différentes sources ?*

On retrouve ici l'effet ambivalent de l'extension de l'offre digitale et de la multiplication des informateurs sur les marchés. Cet effet s'avère plus fort encore que celui de la cohabitation au sein du même panier de types de données pourtant difficilement comparables car se rapportant à des biens particulièrement hétérogènes et/ou renvoyant à des modes de production très différents.

Comme on va le voir à présent, le biais induit tient à ce que la très grande hétérogénéité et l'« opacité méthodologique » (Boulay, Blanke, Casanova Enault, Ganié 2020) des dispositifs qui sous-tendent la production des données disponibles sur Internet dispersent les références de prix au sein d'un éventail très large de valeurs.

Un flou sur les prix entretenu par le foisonnement des supports numériques

Pour comprendre le sentiment de flou qui s'empare des vendeurs lorsqu'ils consultent des données sur de multiples supports, plongeons-nous dans une étude de cas et mettons-nous à la place d'un vendeur qui souhaite céder un logement dans le centre-ville de Montpellier (rue de l'Amandier).

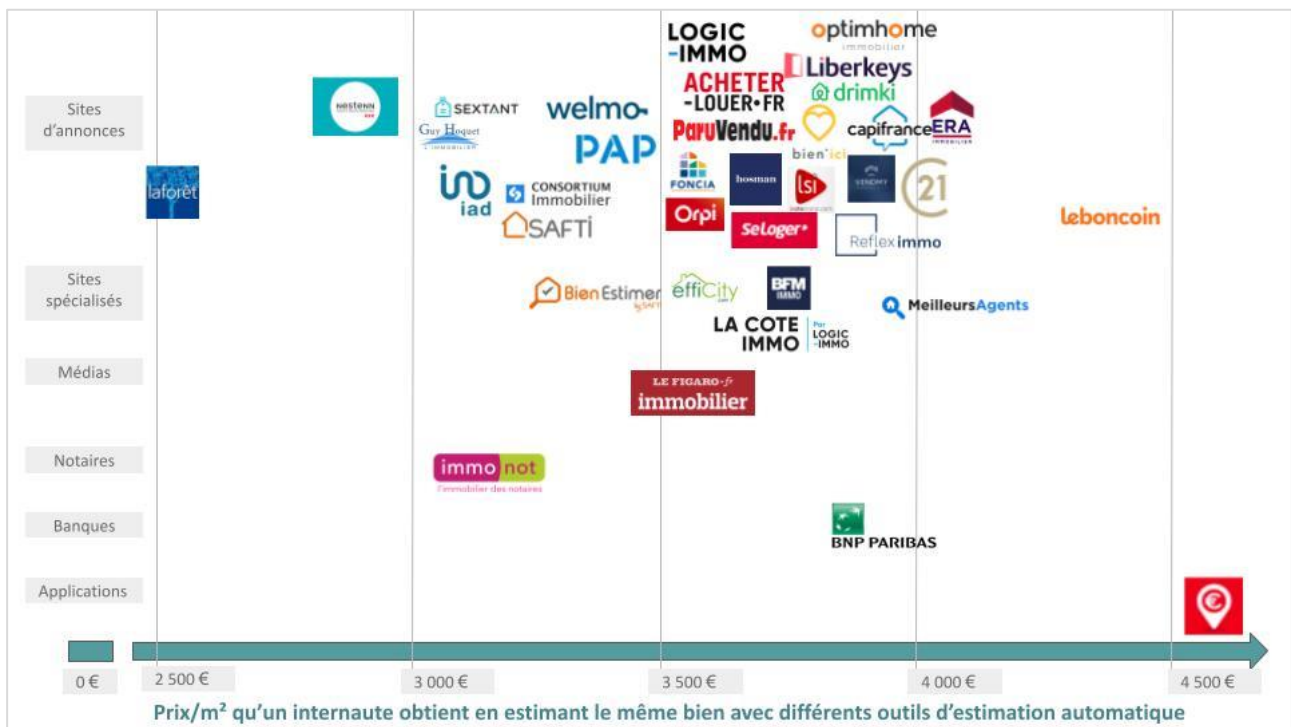
Ce logement dispose de caractéristiques standards pour le secteur. Il s'agit d'un appartement ancien de 3 pièces ayant une taille moyenne (60 m²) avec des prestations comparables à celles des autres biens (Figure 22).

Allons maintenant recueillir des informations sur les prix immobiliers sur les différents supports (sites Internet et applications) recensés dans cette recherche comme faisant partie de l'écosystème des fournisseurs de données de prix (Figure 8).

Consultons dans un premier temps, les supports qui proposent des outils d'estimation automatique. Ces outils nécessitent que l'internaute renseigne un ensemble de champs pour décrire son bien (dont l'adresse, la surface, etc.). Sur la base de ces éléments, ils estiment chacun un prix conseillé, sous forme d'une fourchette généralement. Nous avons ici retenu le prix du milieu de la fourchette suggéré par les différents sites Internet et applications. On constate alors que, pour un même bien, le prix par mètre carré que suggèrent les outils d'estimation automatique varie entre 2 600 € et 4 900 €/m², soit quasiment du simple au double (Figure 22).

Figure 22. Variation des estimations de prix selon les supports pour un même bien

*Cas d'étude : appartement ancien (1920), 3 pièces (2 chambres), 60 m², 3^{ème} étage, immeuble de 5 étages situé au 5 rue de l'amandier 34000 Montpellier (cave de 10 m², sans ascenseur, ni terrasse, ni parking).
Relevé effectué par les auteurs en juin 2021*



Consultons à présent l'indicateur de prix (moyens ou médians) par mètre carré dans le secteur géographique de ce bien de référence. Cet indicateur est proposé par un certain nombre de sites (dont certains proposant aussi un outil d'estimation automatique). Pour le cas Montpellier, ce prix est fourni la plupart du temps à l'échelle de la commune et quelques fois seulement à l'échelle du quartier. Là-encore, les variations d'un support à l'autre sont fortes puisque, selon les sites Internet et applications, les prix affichés oscillent entre 2 450 € et 3 900 €/m² (Figure 23).

De fait, il existe donc une grande hétérogénéité dans l'ensemble des niveaux de prix que les sites Internet et applications affichent respectivement sur la toile pour un même bien ou un même secteur⁶.

⁶ Une autre étude de cas a aussi été menée sur une petite commune plus « détendue » sur le plan immobilier. Les écarts constatés étaient tout aussi significatifs.

Figure 23. Variation de prix moyens ou médians au m² selon les sources pour un même secteur géographique

Cas d'étude : quartier central (incluant la rue de l'Amandier), commune de Montpellier.
 Relevé effectué par les auteurs en juin 2021




L'objectif de cette étude de cas n'est pas de démontrer l'inconsistance des outils statistiques qui sous-tendent les informations disponibles. Il ne s'agit pas non plus de les évaluer ni de distribuer de bons (et de mauvais) points aux dispositifs qui produiraient les meilleures (et les moins bonnes) informations. De fait, il est possible de rendre compte en grande partie des différences de valeurs relevées à travers des explications méthodologiques légitimes.

D'abord, tous les fournisseurs ne s'appuient pas sur le même échantillon de transactions pour consolider leurs références de prix. Certains recourent à des données publiques qui ont l'avantage d'être exhaustives mais qui sont disponibles avec un certain décalage temporel. D'autres privilégient des données plus réactives en faisant remonter des informations issues par exemple de réseaux d'agences immobilières. Elles sont en revanche moins exhaustives. D'autres encore combinent ces deux types de données.

Ensuite, si l'on considère le cas des estimateurs automatiques, la liste même des critères collectés auprès de l'internaute pour évaluer son bien varie fortement d'un outil à l'autre. Tandis que certains font varier la valeur estimée en fonction d'une série limitée de critères fondamentaux comme la surface et le nombre de pièces, d'autres multiplient les critères et s'attachent à des caractéristiques avancées telle que l'orientation (est/ouest, etc.) ou le matériau principal de construction.

Enfin, concernant les prix par mètre carré, certains supports proposent un indicateur à la commune quand d'autres le proposent au quartier. Certains affichent des prix moyens, d'autres des prix médians. On comprend alors aisément que, cumulées, ces divergences méthodologiques peuvent générer des divergences de niveaux de prix.



L'intérêt de cette étude de cas est ailleurs. Il s'agit de se mettre à la place d'un vendeur qui n'aurait pas de compétences statistiques particulières et qui consulterait différents supports pour se faire une idée du prix de son bien. Du point de vue de ce vendeur, on comprend aisément que la multiplication des sources laisse une impression de flou et que cette impression a tendance à supplanter, sinon à atténuer l'effet de transparence attendu de l'extension du paysage informationnel sur les marchés immobiliers.

Le flou s'accroît encore lorsque différents types de données cohabitent au sein du même panier garni. Il tient dans ce cas à la difficulté à embrasser des informations par nature incomparables. On en revient à l'hétérogénéité des biens auxquels se rapportent les indications individuelles de prix. Le biais qui en ressort est accentué par le fait qu'au jeu de la comparaison auquel le vendeur joue, son propre bien finit souvent bon gagnant.

Mais aussi, les vendeurs peuvent mêler des données marquées par un décalage temporel significatif, entre passé (les prix des biens vendus renvoient aux prix conclus dans un passé plus ou moins proche selon les sources) et présent (les estimations, de même que les prix des biens à vendre en dehors du champ strict de la donnée, s'ancrent dans l'actualité des marchés locaux). La distorsion qui en ressort est d'autant plus importante dans les marchés tendus aux prix qui croissent rapidement, à l'instar de Montpellier justement, et de Paris bien sûr.

Les vendeurs peuvent aussi convoquer des données fondées sur des approches et des modes de production différents, allant de l'affichage direct et sans autre forme de traitement, jusqu'à la *data science*, en passant par la constitution et l'exploitation de bases de références de prix en bonne et due forme. En la matière, le flou qui entoure le paysage informationnel dans lequel les vendeurs évoluent est renforcé par l'« opacité méthodologique » (Boulay, Blanke, Casanova Enault, Granié 2020) que la concurrence sur le marché même de la donnée instaure par nécessité. Sur ce marché en effet, les données se font capital et les « flux de données » (« flows of data ») sont des « flux de pouvoirs » (« flows of power » : Sadowski 2019). Capter des flux d'utilisateurs (et d'informations les concernant) *via* des flux de données, c'est générer des *leads* (prospects) qualifiés, en faire commerce aux intermédiaires et développer son activité.

Finalement, après avoir noté dans la partie précédente que la transparence des marchés résidentiels était restreinte par le fait que les vendeurs avaient un accès inégal à la donnée, on comprend ici que la mobilisation effective de données sur les prix ne se traduit pas non plus systématiquement par plus de transparence. L'interprétation de la donnée par les particuliers dans un contexte de prolifération des supports de consultation numérique est une activité complexe qui peut déboucher sur un sentiment d'opacité, d'autant que la toile peut afficher des niveaux de prix particulièrement variables pour un même bien ou une même localité.

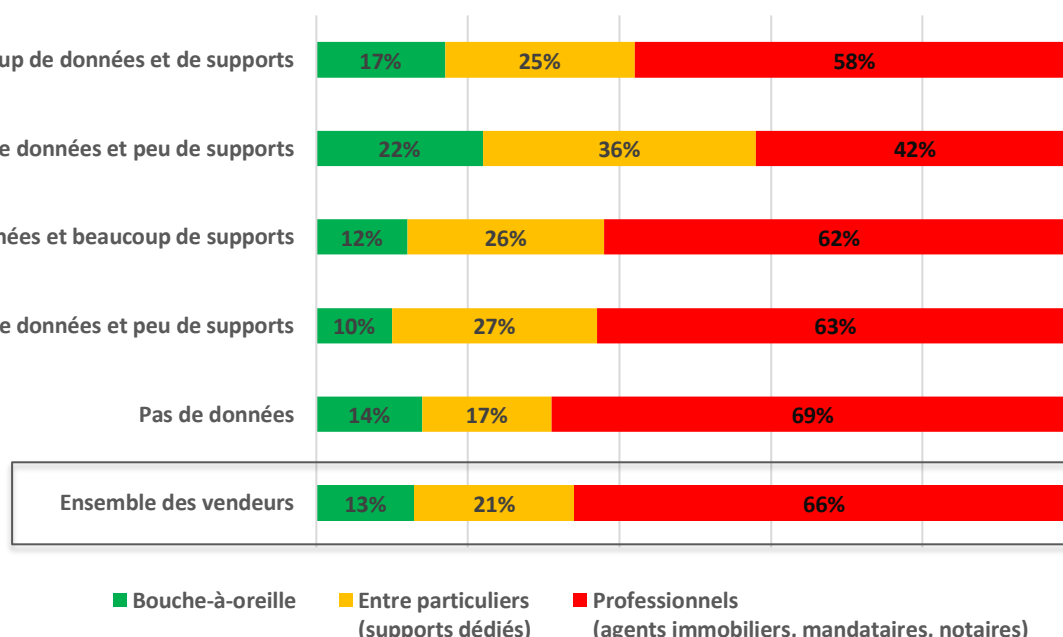
Usages des données et pratiques marchandes des vendeurs de logements : la fixation des prix en question

En dernier lieu se pose la question des pratiques marchandes des vendeurs consommateurs de données. Le choix d'un canal de vente s'accorde-t-il à un profil spécifique de consommation de données ? Qu'en est-il surtout du prix fixé par les consommateurs de données : suit-il une tendance modératrice ou au contraire inflationniste ?

Des consommateurs de données moins souvent accompagnés par un professionnel

66% des particuliers ont recours à un professionnel de l'immobilier pour réaliser la vente de leur logement (Figure 24). 21% concluent entre particuliers *via* des supports dédiés (sites Internet et journaux d'annonces entre particuliers) et 13% trouvent plus spécifiquement preneur par le bouche-à-oreille.

Figure 24. Consommation de données et canal de vente



Ensemble des vendeurs de logements (n=993)

Pourcentages en ligne - Statistique du Khi-deux : DDL=8 / Valeur = 26,34 / Probabilité = 0,009

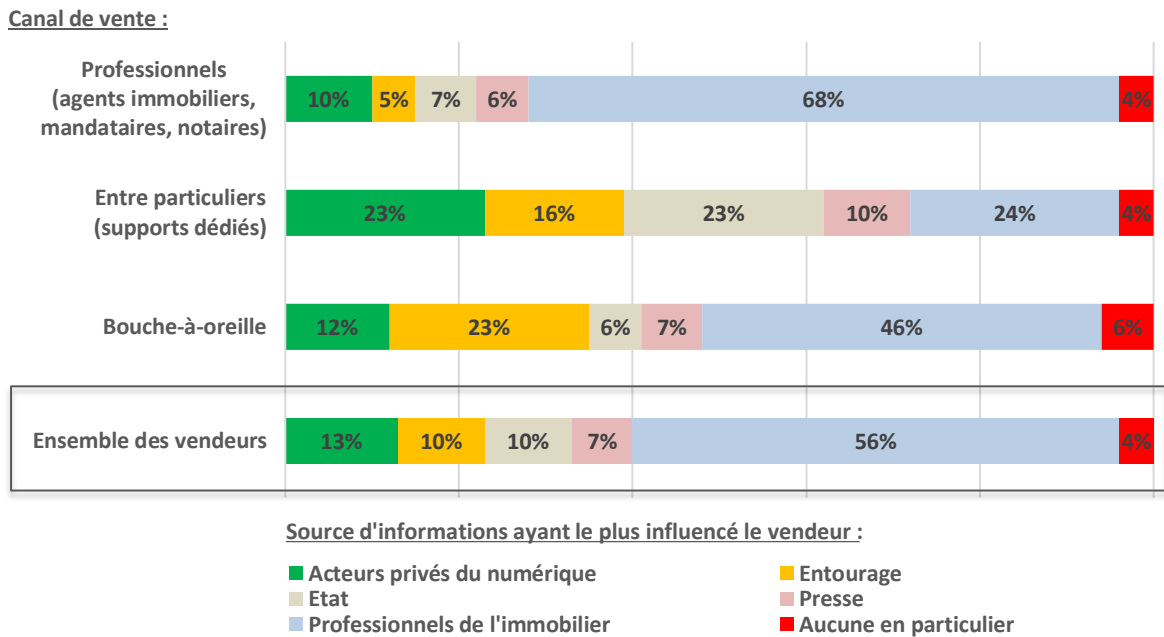
Question : Finalement, par quel intermédiaire avez-vous trouvé l'acheteur définitif ?

Le canal emprunté pour réaliser la vente est statistiquement lié à la propension à consommer des données. D'un côté, les vendeurs qui se passent de données ont tendance à avoir recours à un professionnel de l'immobilier plus que les autres. Ils sont 69% dans ce cas, contre 66% dans l'ensemble

des vendeurs. A l'inverse, les plus grands consommateurs de données sont nettement sous-représentés parmi les vendeurs qui empruntent la voie de l'intermédiation professionnelle. Ils sont de 42% à 58% (Figure 24).

En l'occurrence, les vendeurs qui concluent la vente de leur bien avec un professionnel lui accorde le crédit nécessaire pour constituer un relai d'informations privilégié sur les prix des logements. Ils sont 68%, contre 56% parmi l'ensemble des vendeurs (Figure 25).

Figure 25. Canal de vente et source d'informations prioritaire pour les vendeurs



Vendeurs de logements ayant recours à l'information (n=653, questions non-obligatoires)

Pourcentages en ligne - Statistique du Khi-deux : DDL=10 / Valeur = 112,4 / Probabilité = <0,0001

Questions : - Finalement, par quel intermédiaire avez-vous trouvé l'acheteur définitif ? - Parmi les différentes sources d'informations, laquelle vous a le plus influencé ?

Lecture : 13% des vendeurs de logements informés considèrent que les informations délivrées par les acteurs privés du numérique ont été les plus influentes dans leur expérience de vente. Ce chiffre passe à 23% parmi les vendeurs ayant vendu leur logement via des supports de mise en relation entre particuliers (journaux d'annonces, sites Internet, etc.).

On retrouve le rôle central des professionnels de l'immobilier dans l'accès à l'information. La défiance n'est pourtant pas rare en la matière, ni les comportements opportunistes favorisant le choix du professionnel le mieux disant en termes d'estimation, voire la vente entre particuliers pour diminuer les coûts de transaction après avoir pourtant profité des services gratuits de professionnels (Larceneux, Lefebvre, Simon 2014). Mais lorsque la confiance est au rendez-vous, elle implique à la fois la capacité à conclure la transaction dans les meilleures conditions possibles (optimisation de la recherche d'acquéreurs, gestion de la publicité et de la commercialisation, conseil et assistance juridique, négociation et assistance pour la clôture de la transaction, etc. : Gardes 2019) et la capacité à informer et à délivrer de bons conseils en matière de prix, notamment en termes d'estimation, première attente on l'a vu, des vendeurs envers les agents immobiliers en particulier.

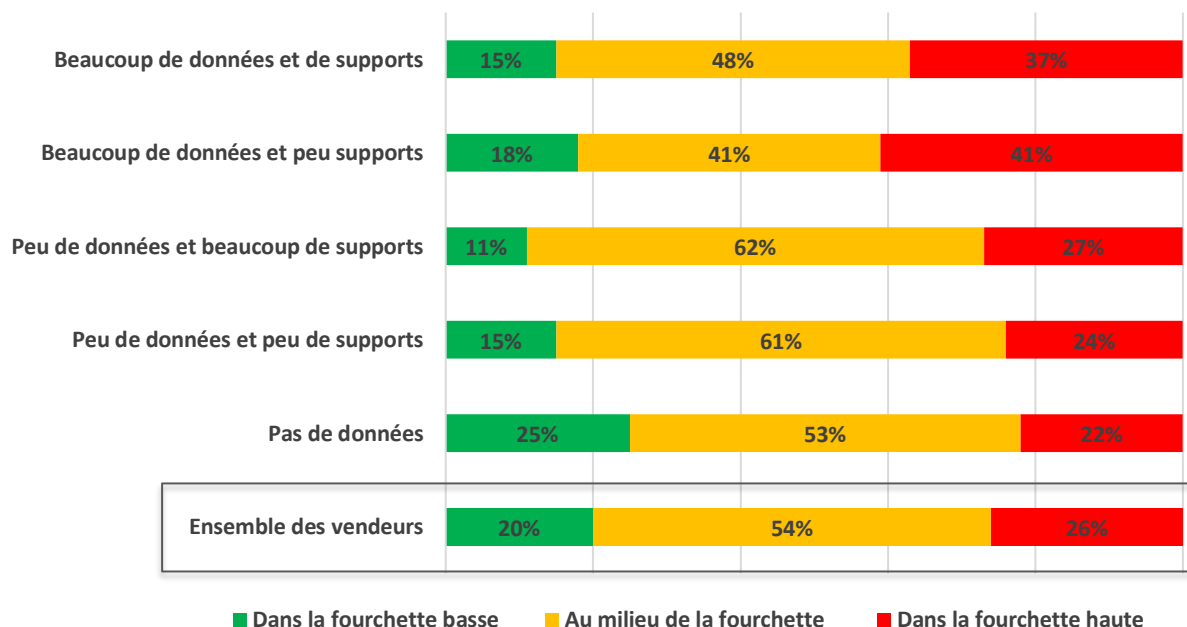
De l'autre côté, les consommateurs de données sont plus nombreux parmi les vendeurs qui contractent entre particuliers. Les plus gourmands d'entre eux sont à trouver parmi les vendeurs qui s'en remettent à des supports dédiés de désintermédiation immobilière pour conclure leur vente, davantage encore qu'au bouche-à-oreille. Ces vendeurs privilégiant la vente de particuliers à particuliers citent les acteurs du numérique comme principale source de confiance, à 23% contre 13% dans l'ensemble (Figure 25). Ils sont également la seule catégorie de vendeurs à s'en remettre de manière privilégiée aux sources étatiques d'informations, à 23% contre 10% pour l'ensemble (Figure 25). Enfin, ils assument leur choix de désintermédiation de la vente en s'adressant rarement aux professionnels pour se tenir informés des prix sur le marché, à 24% contre 56% dans l'ensemble. En matière d'informations, ils paraissent donc assumer une plus grande cohérence et un moindre opportunisme qu'on ne le prête généralement aux vendeurs de logements.

Des consommateurs de données tentés par les valeurs supérieures

54% des vendeurs de logements considèrent avoir positionné leur prix de mise en vente au milieu de la fourchette de prix, 20% dans la fourchette basse et 26% dans la fourchette haute (Figure 26).

Y a-t-il un effet de la consommation de données sur ces positionnements de prix ? En effet, la consommation de données semble favoriser un alignement sur les prix hauts et l'absence de données un alignement sur les prix bas. Ainsi, 37% des grands consommateurs de données déclarent s'être placés dans la fourchette haute de prix, contre 22% de ceux qui ne consomment pas de données et 26% dans l'ensemble des vendeurs (Figure 26).


Figure 26. Consommation de données et positionnement du prix de vente (opinion des vendeurs)



Ensemble des vendeurs de logements (n=989, question non-obligatoire)

Pourcentages en ligne - Statistique du Khi-deux : DDL=8 / Valeur = 27,6 / Probabilité = 0,0006

Question : Par rapport aux informations dont vous disposez, diriez-vous que le prix auquel vous avez mis en vente votre logement était ...



La tendance à viser les prix hauts des grands consommateurs de données peut être rapprochée de plusieurs facteurs. D'abord, on l'a vu, ces vendeurs sont plus souvent localisés en zones tendues (Figure 19), zones qui affichent de fortes augmentations de prix et qui sont propices à des tentatives d'enchères. Ensuite, ils ont une plus forte propension à entretenir un sentiment de discordance informationnelle (Figure 21). Comme si le sentiment de flou (ou le défaut de transparence) entretenait les tendances haussières de l'immobilier. Comme s'il portait à la surenchère, du moins aux paris les plus risqués. Le pari tend d'ailleurs à être gagnant : on le verra, les consommateurs de données sont surreprésentés parmi les vendeurs qui ont vendu au prix qu'ils avaient proposé initialement.

La relation entre la consommation de données et le positionnement dans la fourchette de prix semble propre à ce type spécifique d'information. En effet, le positionnement du prix de mise en vente apparaît sans lien statistique avec l'usage d'informations en général⁷. En l'occurrence, les vendeurs qui consomment beaucoup d'informations ne sont pas surreprésentés parmi ceux qui considèrent avoir fixé un prix haut.

Le recours à un professionnel de l'immobilier pour assurer la vente apparaît également sans lien avec un éventuel comportement inflationniste⁸. D'après la littérature économique, une raison en est qu'étant donné que les agents immobiliers ne reçoivent qu'une petite part du bénéfice supplémentaire lorsqu'un bien se vend à un prix plus élevé, ils sont incités, en exploitant leur avantage informationnel, à convaincre leurs clients de vendre leur bien rapidement et donc à un prix relativement bas (Levitt et Syverson 2008).

Les consommateurs de données : des vendeurs plus satisfaits

75% des vendeurs se disent satisfaits du prix auquel ils ont vendu leur logement. En particulier, la moitié s'avèrent plutôt satisfaits et un quart tout à fait satisfaits (Figure 27).

Ce sentiment a tendance à varier avec la consommation de données. De fait, les plus grands consommateurs de données ont une légère propension⁹ à être aussi les plus satisfaits du prix obtenu de la vente de leur logement (Figure 27). On touche de nouveau à la spécificité de l'effet « donnée », le sentiment de satisfaction retiré de la vente étant de son côté sans lien significatif avec la consommation d'informations en général¹⁰.

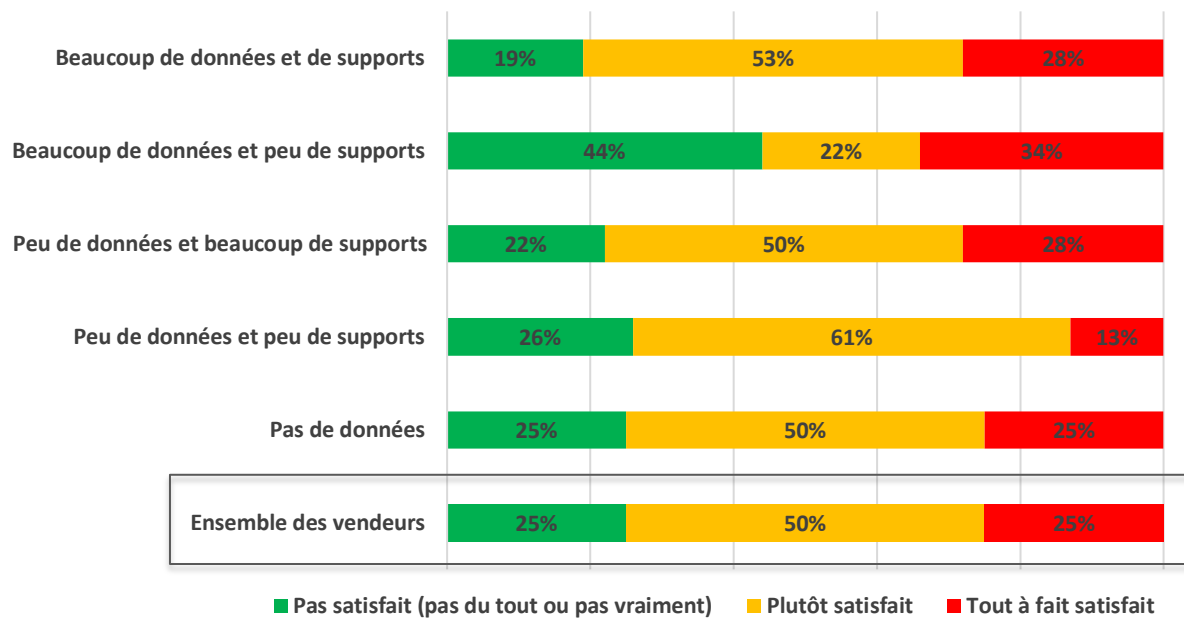
⁷ L'absence de lien statistique est établie par le fait que le croisement entre la variable de positionnement dans la fourchette de prix et le score d'usage de l'information, contrairement au score d'usage de la donnée, affiche un test du khi-deux non significatif.

⁸ Le croisement entre la variable de positionnement dans la fourchette de prix et la variable du canal de vente affiche également un test du khi-deux non significatif, ce qui dénote une absence de lien statistique entre les deux variables.

⁹ Le croisement entre la variable de satisfaction relative au prix de vente et le score d'usage de la donnée affiche un test du khi-deux significatif mais moins significatif que dans le cas des autres croisements mis en avant dans ce rapport.

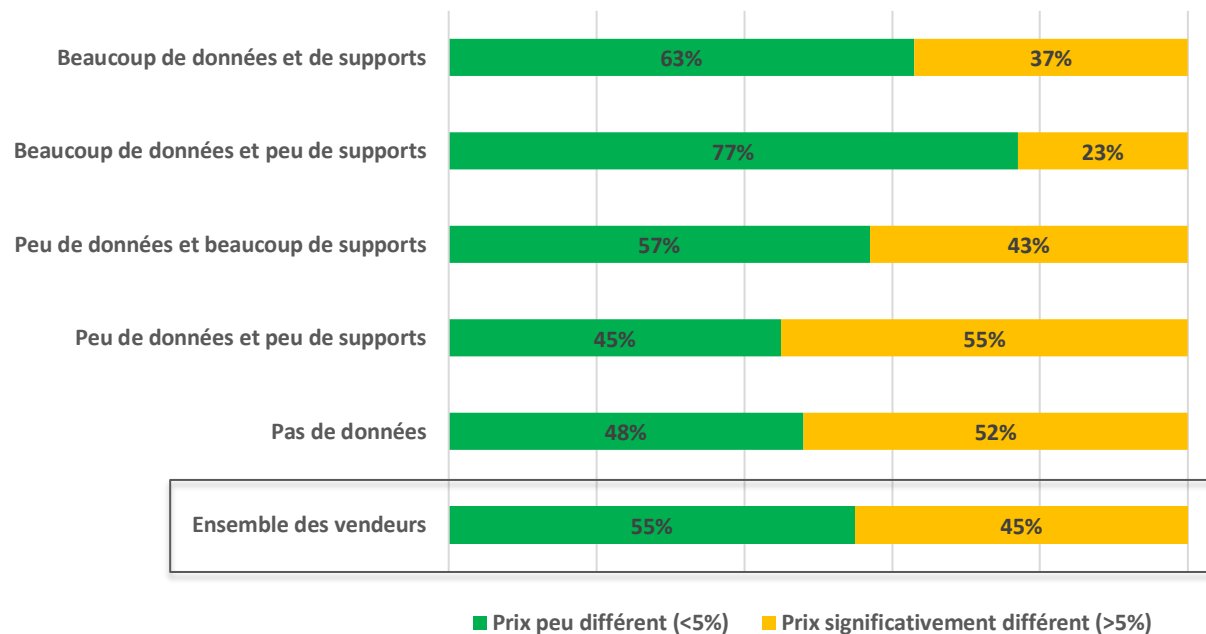
¹⁰ Dans ce cas en effet, le test du khi-deux n'est pas significatif.

Figure 27. Consommation de données et satisfaction concernant le prix de vente




Ensemble des vendeurs de logements (n=993)
 Pourcentages en ligne - Statistique du Khi-deux : DDL=8 / Valeur = 17,9 / Probabilité = 0,021
 Question : Etes-vous satisfait du prix auquel vous avez finalement vendu le logement ?

Figure 28. Consommation de données et évolution du prix au cours de la vente



Ensemble des vendeurs de logements (n=853, questions non-obligatoires)
 Pourcentages en ligne - Statistique du Khi-deux : DDL=4 / Valeur = 8,4 / Probabilité = 0,073
 Questions : -A quel prix avez-vous vendu le logement ? - Et à quel prix aviez-vous mis en vente le logement au départ ?
 Catégories : « Prix peu différent » : le prix de vente a été différent du prix de mise en vente dans des proportions inférieures à 5%. « Prix significativement différent » : le prix de vente a été différent du prix de mise en vente dans des proportions supérieures à 5%.



La satisfaction plus grande des plus grands consommateurs de données concernant le prix de vente peut tenir au fait qu'ils parviennent plus souvent à vendre leur logement à un prix final proche de celui qu'ils en avaient proposé au départ. En effet, les vendeurs qui consomment beaucoup de données sur beaucoup de supports affichent à plus de 63% des ventes à un prix faiblement négocié, contre 55% pour l'ensemble des vendeurs (Figure 28). Quant aux particuliers qui consomment beaucoup de données sur peu de supports, ils sont 77% dans ce cas, contre 45% pour les plus petits consommateurs de données.

Les plus grands consommateurs de données ont donc tendance à obtenir le prix qu'ils avaient initialement demandé pour leur bien. Plus étonnant, comme on l'a vu, ils sont aussi plus nombreux à viser la fourchette haute de prix au départ.

En somme : des effets contrastés de la donnée sur les marchés

Pour finir, tentons d'articuler dans une même analyse les différents liens entre consommation de données et pratiques marchandes des ménages. L'analyse des correspondances multiples (ACM) permet cette approche synthétique.

En l'espèce, l'ACM résume la diversité des pratiques et perceptions des vendeurs en deux axes (Figure 29) :

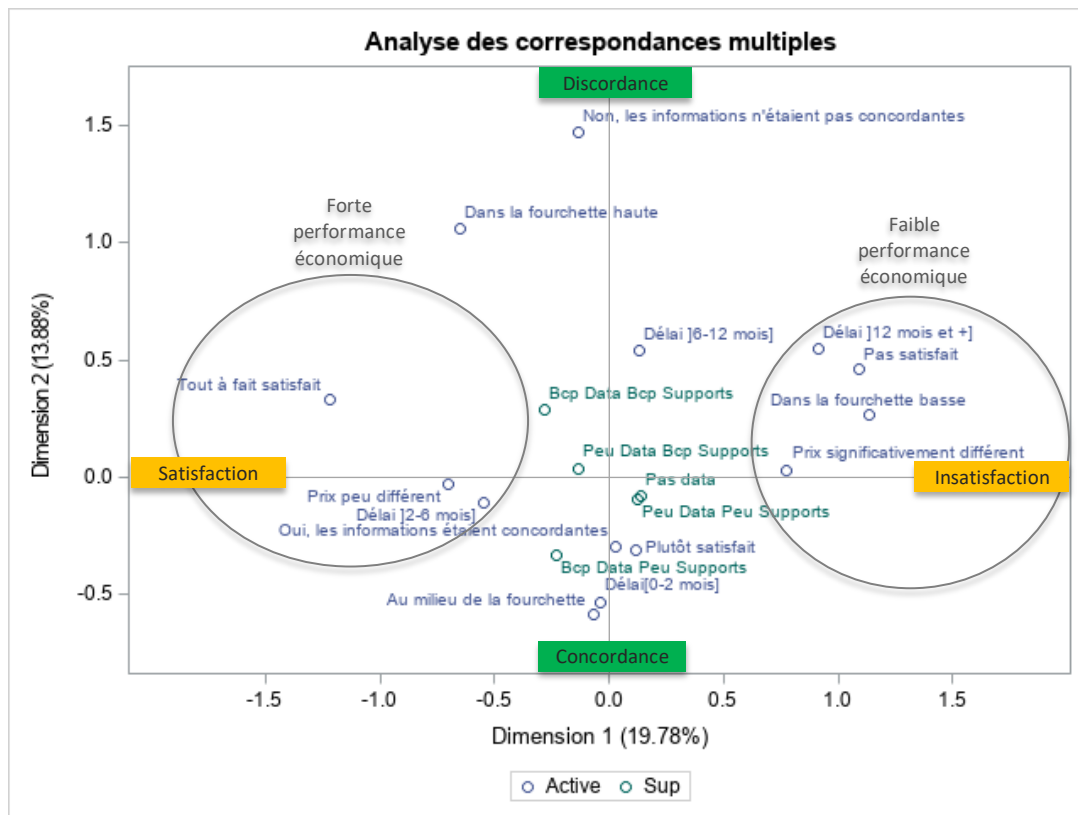
- Le premier (horizontal) peut être interprété comme une échelle de satisfaction concernant la vente immobilière. A gauche, la modalité « tout à fait satisfait ». A droite, la modalité « pas satisfait ». Dans l'intervalle, des degrés intermédiaires de satisfaction. Il est alors intéressant de constater que le niveau extrême de satisfaction est statistiquement lié au fait d'avoir vendu le logement dans des délais relativement courts (de 2 à 6 mois) et à un prix très proche du prix de mise en vente. En d'autres termes, la satisfaction dépend aux yeux des vendeurs de la performance économique dont ils ont fait preuve. A l'inverse, l'insatisfaction dérive d'une faible performance économique. A savoir : vendre dans des délais longs (plus de 12 mois), dans une fourchette jugée basse parmi les valeurs locales, et à un prix significativement différent du prix de mise en vente.
- Le deuxième axe (vertical) renvoie quant à lui à la transparence perçue, c'est-à-dire au sentiment de discordance ou de concordance que le vendeur peut retirer des informations qu'il a pu collecter sur les prix des logements.

Il est alors particulièrement intéressant de voir comment se positionne la consommation de données dans cette galaxie de pratiques et de ressentis.

En l'occurrence, les modalités du score d'usage de la donnée (intégré dans l'ACM en variable supplémentaire) se distribuent assez bien sur l'axe vertical. Ce qui est important à noter est l'effet du nombre de supports. Lorsque la donnée est consommée, même en quantité, sur un nombre limité de sites Internet ou applications, elle favorise une lecture concordante des marchés et donc une forme de transparence. En revanche, quand la consommation s'établit sur un nombre important de supports, les vendeurs tendent à juger discordantes les informations recueillies sur les prix. Ceci confirme l'analyse que nous proposons plus haut à savoir que la consommation de données améliore la

transparence mais que la multiplication des supports la restreint. Dans ce cas, le sentiment de flou reprend le dessus.

Figure 29. Liens entre consommation de données et pratiques marchandes des vendeurs (analyse factorielle)



Ensemble des vendeurs de logements (n=993)

Les modalités du score d'usage de la donnée se répartissent moins nettement sur l'axe horizontal. Néanmoins, toutes les modalités renvoyant à une consommation importante de données se situent plutôt à gauche de l'axe vertical, c'est-à-dire vers le pôle incarnant une plus grande efficacité et une plus grande satisfaction des vendeurs.

En somme, l'usage de la donnée améliore la transparence mais seulement dans certains cas. En revanche, elle peut inciter à tirer les valeurs vers le haut et tend à renforcer la performance économique de ceux qui la consomment.

Conclusion

Les résultats de cette enquête empirique originale sur les pratiques économiques des particuliers montrent que les consommateurs de données accumulent davantage de critères de réussite dans la vente immobilière. L'écart entre le prix auquel ils mettent leur logement en vente initialement et son prix final est plus faible que pour les autres vendeurs. Les consommateurs de données vendent aussi plus souvent dans des délais courts. Finalement, ils se disent davantage satisfaits de leur expérience de vente immobilière. A une échelle individuelle, l'usage de données semble ainsi s'associer à une plus grande performance sur les marchés.

A une échelle collective, les résultats de l'enquête montrent que les effets de la diffusion de la donnée sont plus contrastés. En règle générale, l'usage de données de prix renforce le sentiment de transparence des marchés résidentiels. Mais il est socialement et géographiquement situé. Tous les ménages ne consomment pas (pas encore en tout cas) des données immobilières. Ceux qui le font ont plutôt un âge avancé et une position sociale élevée. Ils sont par ailleurs situés plus souvent dans des marchés locaux dits tendus. Si la diffusion des données apporte globalement de la transparence, elle ne la procure pas à tous à un même degré.

Par ailleurs, il faut insister sur l'existence d'une barrière technique à l'usage de la donnée par les particuliers. Les ménages qui veulent vendre un logement ne puisent pas directement dans les fichiers sources pour obtenir des références de prix. Le cas échéant, ils utilisent la connaissance que produisent pour eux des intermédiaires qui prennent en charge l'exploitation de ces *big data*. On assiste depuis quelques années à un développement très rapide du nombre d'intermédiaires de données (*data brokers*) qui, sur Internet, proposent de la connaissance sur les prix immobiliers. Dans cet écosystème foisonnant, il n'existe pas de référentiel méthodologique partagé et les méthodes restent souvent opaques. Ainsi, des informations très différentes peuvent être collectées par les ménages lorsqu'ils s'informent sur les prix.

Une autre barrière technique tient précisément à la difficulté des vendeurs à interpréter ces différences. Finalement, plus un vendeur recourt à un nombre important d'intermédiaires de données sur Internet, plus il collecte des informations qui sont difficiles à comparer et peuvent donc apparaître peu cohérentes. L'enquête met en exergue un effet de seuil dans l'usage des données de prix à partir duquel la transparence laisse place au flou !

Pour revenir au débat théorique initié en introduction, on peut dire que l'existence d'informations économiques, notamment sous forme de données, ne se traduit pas systématiquement par plus de transparence sur les marchés ni par des effets vertueux sur les prix. Les liens qu'entretiennent ces variables sont complexes. Elles dépendent du contexte social et territorial dans lequel les données sont diffusées. En l'occurrence, la France de 2021 oppose à la libération des données un contexte hétérogène tant sur le plan des populations et des territoires que des acteurs numériques de la connaissance immobilière.

Les marchés du logement et de la donnée évoluent vite. Cette recherche les saisit dans leur actualité mais repose sur un cadre d'interprétation valable pour décrypter leurs évolutions à venir. Histoire à suivre...

Annexe : Méthodologie de l'enquête par questionnaires

Les résultats de cette recherche s'appuient essentiellement sur une enquête par questionnaires menée entre décembre 2020 et janvier 2021. Le questionnaire a été diffusé à 993 répondants parmi la population des personnes ayant vendu un logement en France au cours des 24 derniers mois (en 2019 ou 2020) qu'il s'agisse ou non d'une résidence principale.

La rédaction du questionnaire a été assurée par l'équipe de recherche sur la base d'un premier volet qualitatif ayant permis de réaliser des entretiens semi-directifs avec une dizaine de vendeurs et autant d'agents immobiliers. La passation du questionnaire a été assurée par l'institut Opinion Way. Il s'agissait d'un questionnaire auto-administré en ligne.

La population d'enquête

Aussi surprenant que cela puisse paraître, il n'existe pas dans la statistique publique d'enquêtes portant sur la population des vendeurs de logements en France. A notre connaissance, il n'y a pas non plus d'enquête d'envergure menée sur les vendeurs par des universitaires ou des acteurs de la ville et de l'immobilier. Ce manque de référence implique une difficulté méthodologique majeure : cette population est assez mal connue. Or, sans une description précise de la population d'enquête, difficile de façonner un échantillon représentatif.

La première étape a donc consisté à utiliser le sondage *omnibus* qu'Opinion Way diffuse de façon hebdomadaire à un échantillon de 1000 répondants représentatifs de la population française. La question posée était : « Avez-vous vendu un logement au cours des 24 derniers mois ? ». Les modalités de réponses possibles étaient les suivantes : « Oui, une résidence principale » ; « Oui, une résidence secondaire » ; « Oui, un bien mis en location » ; « Oui, un autre bien (héritage, etc.) » ; « Non ». En croisant ces réponses avec les caractéristiques démographiques des répondants de ce sondage *omnibus*, une description statistique de la population a pu être proposée.

Construction de l'échantillon

Pour construire l'échantillon, la méthode des quotas a d'abord été envisagée. Plusieurs critères sont apparus pertinents pour les construire. D'abord trois critères sociodémographiques que sont : le sexe, la classe d'âge et la catégorie socioprofessionnelle. Il s'agissait en effet d'être représentatif de la diversité des profils de vendeurs de logements en France. Ensuite, il s'agissait aussi d'être représentatif de la diversité des territoires à l'échelle du pays. Deux autres critères ont donc été mobilisés : la taille de l'unité urbaine et la région de résidence du répondant (Figure 30).

La combinaison de 5 critères de quotas rend plus difficile l'obtention de réponses en nombre suffisant pour les profils de répondants peu fréquents. Ainsi, le sondage n'est pas parvenu à remplir complètement ses objectifs de quotas. Le profil des 993 répondants n'étaient pas parfaitement conforme à la structure de la population. Des poids de redressement ont donc été appliqués sur l'échantillon lors des analyses.

Représentativité de l'échantillon

L'échantillon redressé obtenu finalement reprend assez bien la structure de la population des vendeurs de logements récents (voir Figure 30).

Figure 30. Représentativité de l'échantillon sur 5 variables sociodémographiques et territoriales

	Population (Sondage Omnibus)	Echantillon sans redressement (n=993)	Echantillon redressé (n=993)
Sexe			
Masculin	52,1%	53,5%	52,0%
Féminin	47,9%	46,5%	48,0%
Âge			
18-34 ans	36,1%	21,4%	36,9%
35-49 ans	17,8%	25,1%	17,6%
50-64 ans	26,2%	32,3%	28,1%
65 ans et +	19,9%	21,2%	17,4%
Région (secteurs)			
IDF	18,4%	22,2%	18,5%
Nord-Ouest	27,1%	23,9%	27,1%
Nord Est	16,9%	18,8%	16,9%
Sud-Ouest	10,6%	10,2%	10,6%
Sud Est	27,0%	24,9%	26,9%
PCS			
PCS +	32,2%	39,0%	32,1%
PCS -	27,7%	27,2%	27,8%
Inactifs	40,2%	33,8%	40,1%
Tranche d'unité urbaine			
Commune rurale	21,1%	15,8%	16,9%
Unité urbaine de 2000 à moins de 20 000 habitants	23,5%	18,5%	20,6%
Unité urbaine de 20 000 à moins de 200 000 habitants	10,3%	22,6%	23,1%
Unité urbaine de 200 000 habitants ou plus (sauf agglomération parisienne)	26,7%	22,8%	22,5%
Agglomération parisienne	18,4%	20,3%	16,9%

Pour réaliser les redressements, la nomenclature des PCS (professions et catégories socioprofessionnelles) de l'INSEE a été utilisée en procédant à des regroupements de catégories (voir Figure 31).

Figure 31. Modalités de regroupement des PCS

PCS (INSEE)	Regroupement
1 - Agriculteurs	PCS +
2- Artisans, commerçants, chefs d'entreprises	
3 - Cadres et professions intellectuelles supérieures	
4 - Professions intermédiaires	
5 - Employés	PCS -
6 - Ouvriers	
7 - Retraités	Inactifs
8 - Autres sans activités professionnelles	

Bibliographie

Boeing G. (2019), « Online rental housing market representation and the digital reproduction of urban inequality », *Environment and Planning A - Economy and space* : https://www.researchgate.net/publication/335427185_Online_rental_housing_market_representation_and_the_digital_reproduction_of_urban_inequality

Bonneval L. (2012), « Évolutions du métier d'agent immobilier et marchés du logement », in Bergé M., Rougé L. (dir.), *Être logés, se loger, habiter*, Paris, L'Harmattan (Habitat et Société), pp. 17-24

Boulay G., Blanke D., Casanova Enault L., Granié A. (2020), « Moving from Market Opacity to Methodological Opacity: Are Web Data Good Enough for French Property Market Monitoring? », *The Professional Geographer*, pp.1-16

Callon M. (2017), *L'emprise des marchés*, Paris, La Découverte, 501 p.

Cavailhès J. (2005), « Le prix des attributs du logement », *Economie et statistique*, n°381-382 : https://www.persee.fr/doc/estat_0336-1454_2005_num_381_1_7210

Cusin F. (2012), « La ville à l'épreuve du marché. Contribution à une sociologie économique de la ville et des mobilités résidentielles », HDR en Sociologie, Université Paris Dauphine, 401 p.

Desrozières A. (2011), « Les origines statisticiennes de l'économie des conventions : réflexivité et expertise », *Oeconomia*, 1-2: <https://journals.openedition.org/oeconomia/1770>

François P. (2008), *Sociologie des marchés*, Armand Colin, 331 p.

Gardes N. (2019), « Digitalisation du secteur immobilier : la proposition de valeur phytigale au cœur de la performance », *La Revue des sciences de gestion*, n°299-300


Garmaise M.J. et Moskowitz T.J. (2004), « Confronting Information Asymmetries: Evidence from Real Estate Markets », *The Review of Financial Studies*, Vol.17, n°2 : <https://www.nber.org/papers/w8877>

Juillard C. (2019a), *Le tournant numérique des données immobilières : permanences et recompositions*, Etude publiée par Iread avec le soutien d'Iread, du LIFTI, du PUCA et d'Urbanics: https://uploads.strikinglycdn.com/files/37cade11-94f9-41d6-b935-dcc186fbebfe/Le%20tournant%20num%C3%A9rique%20des%20donn%C3%A9es%20immobili%C3%A8res_CJuillard_mai%202019.pdf

Juillard C. (2019b), « Faire des références de loyers à partir de données d'annonces. Le pari disruptif du Ministère de la Cohésion des Territoires », *Le Journal de l'Agence*, juillet 2019: <https://www.journaldelagence.com/1167305-la-puissance-publique-producteur-disruptif-de-donnees-immobilieres-claire-juillard-sociologue-recherche-et-conseil-en-immobilier>

Juillard C. (2019c), « Prix immobiliers : de quoi parle-t-on? », *My Expat*, octobre 2019: <https://www.myexpat.fr/prix-immobiliers-de-quoi-parle-t-on/>

Juillard C. (2020), « Produire des données de prix et de loyers à l'heure de la PropTech : quel rôle pour l'Etat ? », *Working paper pour la Cities and Digital Technology Chair*, Sciences Po, Paris, janvier 2020: https://www.sciencespo.fr/ecole-urbaine/sites/sciencespo.fr.ecole-urbaine/files/2020_01%20-%20Juillard%20copie.pdf



Larceneux F., Lefebvre T., Simon A. (2014), « La perspective des coûts de transaction perçus : une explication de l'intermédiation immobilière », *Revue d'économie régionale et urbaine*, 2014/3, octobre: <https://www.cairn.info/revue-d-economie-regionale-et-urbaine-2014-3-page-499.html>

Leglaye S., Rolland A. (2019), « Une personne sur six n'utilise pas Internet, plus d'un usager sur trois manque de compétences numériques de base », *Insee Première*, n°1780 : <https://www.insee.fr/fr/statistiques/4241397>

Levitt S. et Syverson C. (2008), « Market distortions when agents are better informed : The value of information in real estate transactions », *Review of Economics and Statistics*, Vol. 90, n°4 : https://www.nber.org/system/files/working_papers/w11053/w11053.pdf

Migozzi J. (2020), *Une ville à vendre : numérisation et financiarisation du marché du logement au Cap: stratification et ségrégation de la métropole émergente*, Thèse de doctorat en géographie, Université Grenoble Alpes

Pumain D., Thisse J-F., Walliser B., Thomas I., Chamussy H., Taillard C., Durand-Dastès F., Marie-Claire Robic M-C., et Saint-Julien T. (2007), « Nouvelle économie géographique et géographie : quel dialogue ? », *L'Espace Géographique*, Vol. 3, n° 3, p.193-214 : <https://www.cairn.info/revue-espace-geographique-2007-3-page-193.htm>

Sadowski J. (2019), « When data is capital : datafication, accumulation and extraction », *Big data & society*, Vol.6, n°1 : <https://journals.sagepub.com/doi/full/10.1177/2053951718820549>

Sheppard S. (1999), « Hedonic Analysis of Housing Markets », in Mills E.S. and Cheshire P. (dir.), *Handbook of Regional and Urban Economics*, vol. 3, *Applied Urban Economics*.

Simon A. (2008), « Les déterminants de la valeur des appartements : une approche hédoniste », *Etudes Foncières*, Compagnie d'édition foncière

Vatin F. (dir). (2009), *Evaluer et valoriser. Une sociologie économique de la mesure*, Toulouse, Presses universitaires du Mirail

Vorms B., avec Cusset P-Y (2016), *La Révolution numérique et le marché du logement. Nouveaux usages, nouveaux acteurs, nouveaux enjeux*, Paris, France Stratégie : https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/rapport-logement-vorms-11-2016_0.pdf