



Université de Paris

UFR GHES - Département Lettres et Sciences Humaines

**M2 GAED - Géographie, aménagement, environnement et développement,
parcours Géoprisme - Année universitaire 2020/2021**

**Mise en oeuvre d'une chaîne de traitement de données
AirDNA : Étudier l'activité de la plateforme Airbnb
dans 3 communes franciliennes**

Par Louis Laurian

**Stage réalisé à l'UMS2414 RIATE dans le cadre du projet WisDHoM du
15/02/2021 au 30/07/2021**

Sous la direction de :

- **Marianne GUÉROIS, Maître de conférence à l'Université Paris Diderot**
- **Malika MADELIN, Maître de conférence à l'Université Paris Diderot**

Tuteur :

Ronan YSEBAERT, Ingénieur d'étude à l'UMS2414 RIATE

Session Septembre 2021

Résumé

Ce rapport de stage s'inscrit dans le cadre du projet de l'ANR WISDHOM (coord. Renaud Le Goix), et a été effectué au sein de l'UMS2414 RIATE du 15/02/2021 au 30/07/2021. L'objectif est d'analyser les inégalités liées au marché du logement *via* l'angle des locations saisonnières, et plus particulièrement Airbnb, dans des communes situées en espace périurbain des grandes métropoles. Cet angle a été abordé car il s'agit d'un phénomène récent, qui a pris de l'ampleur au cours des années 2010, à tel point que les politiques publiques ont émis des restrictions pour réguler le phénomène. En effet, un des processus accentuant la crise du logement au sein des grandes métropoles est la reconversion d'un logement vers le locatif à court terme, réduisant le parc de résidences principales d'une ville et augmentant les prix du logement. Ce travail a pour vocation de construire un socle méthodologique afin dans un premier temps d'étudier l'activité d'Airbnb dans 3 communes franciliennes, puis de transposer ces analyses à d'autres communes. En effet, un deuxième volet de ce stage est la recherche de reproductibilité, avec une méthode qui se veut extensible à d'autres terrains d'étude. Pour cela, le travail a été réalisé entièrement *via* R, langage de programmation libre et axé sur des analyses statistiques, et le code ainsi que les résultats ont été diffusés sur un site internet, disponible à cette adresse : https://riatestage.github.io/airdna_laurian/index.html.

Abstract

This internship report is part of the ANR WISDHOM project (coord. Renaud Le Goix), and was carried out within the UMS2414 RIATE from 15/02/2021 to 30/07/2021. The aim is to analyze the inequalities linked to the housing market through the scope of short-term rentals (STR), and more particularly Airbnb, in metropolitan suburban areas. This point of view was chosen because it is a recent phenomenon, which grew during the 2010s, until public policies issued restrictions to balance it. Indeed, one of the processes that emphasises the housing crisis in large metropolitan areas is the conversion of long-term to short-term rental, reducing the stock of primary residences in a city and increasing housing prices. The purpose of this work is to build a methodological basis in order first to study Airbnb's activity in three Parisian municipalities, and then to transfer these analyses to other municipalities. Finally, a second aspect of this internship is the search for reproducibility, with a method that can be extended to other study areas. In order to achieve this aim, the work was entirely done *via* R, a free programming language focused on statistical analysis, and the code and results were published on a website, available at this address : https://riatestage.github.io/airdna_laurian/index.html.

Remerciements

Je souhaite tout d'abord remercier mon tuteur Ronan Ysebaert ainsi que mes encadrantes Marianne Guérois et Malika Madelin qui m'ont accompagné jusque dans les derniers instants de ce stage, pour m'avoir permis de venir à bout de ce rapport. Je n'aurai jamais imaginé atteindre un tel niveau de maîtrise de R sans Ronan, et je remercie Marianne et Malika de m'avoir fait confiance pour une seconde année consécutive malgré mon acharnement pour les multiloueurs. Je tiens aussi à remercier Renaud Le Goix, qui m'a gentiment proposé d'effectuer mon stage au RIATE et d'intégrer le projet qu'il dirige. Je profite de ces remerciements pour m'excuser d'avoir été absent aux 3/4 des réunions hebdomadaires du RIATE, du fait de ma mémoire défaillante.

Je souhaite par la suite remercier mes collègues du RIATE, et féliciter plus particulièrement Timothée Giraud et Chloé Didelon pour l'organisation de l'école thématique du CNRS de cet été. En parlant d'école thématique, comment ne pas oublier Hugues Pécourt, avec qui j'ai eu le plaisir de partager mon premier déplacement professionnel (satanée ligne 13...). Toujours au RIATE, je remercie Nicolas et Arthur pour leur accompagnement et leur esprit de compétition tout au long du stage.

Mes pensées vont aux collègues extérieurs au RIATE, Isa, Étienne, Romain, Violaine, pour leur bonne humeur au cours de ces six mois de dur labeur, ainsi que l'irremplaçable Anne-Gaëlle, partie vers de nouvelles aventures.

En vous souhaitant une bonne continuation à tous

Table des matières :

Introduction	6
<u>Le sujet de recherche : un projet de l'ANR WISDHOM</u>	<u>6</u>
<u>Pourquoi l'angle des locations saisonnières à l'échelle locale ?</u>	<u>7</u>
<u>Objectifs du stage et hypothèses</u>	<u>9</u>
Partie 1 : Locations saisonnières et crise du logement dans les grandes métropoles	10
1.1. Une contribution à l'ANR WISDHOM	10
<u>1.1.1. Un projet axé sur le marché du logement</u>	<u>10</u>
<u>1.1.2. Intérêt d'étudier les locations saisonnières dans le cadre de ce projet</u>	<u>11</u>
1.2. Évolution de la présence de locations Airbnb à différentes échelles	12
<u>1.2.1. En France : une offre calquée sur les densités de population et de la fréquentation touristique</u>	<u>12</u>
<u>1.2.2. En IDF : Un marché parisien saturé, se diffusant vers la périphérie ?</u>	<u>13</u>
<u>1.2.3. Notre zone d'étude : Les communes d'Issy-les-Moulineaux, Pantin et Bagneux</u>	<u>14</u>
1.3. Facteurs explicatifs de la croissance de la location saisonnière	16
<u>1.3.1. Une croissance récente</u>	<u>16</u>
<u>1.3.2. Impact des mesures de régulation sur l'évolution de l'offre</u>	<u>17</u>
<u>1.3.3. Impact des locations saisonnières (et de leur expansion) sur le marché du logement</u>	<u>18</u>
Partie 2 : Une méthodologie pour une recherche reproductible	20
2.1. Vers des nouveaux standards de reproductibilité dans la recherche	20
<u>2.1.1. La reproductibilité au coeur de la démarche scientifique de la structure d'accueil : le RIATE</u>	<u>20</u>
<u>2.1.2. Une base de données particulière : AirDNA</u>	<u>21</u>
2.2. Mise en œuvre d'une chaîne de traitement permettant d'exploiter le contenu de la base de données AirDNA	22
<u>2.2.1. Pré-traitement et nettoyage des données</u>	<u>22</u>
<u>2.2.2. Traitement des valeurs NA</u>	<u>23</u>
<u>2.2.3. Création de variables</u>	<u>25</u>
<u>2.2.4. Critique des données</u>	<u>26</u>
2.3. Analyses et reproductibilité de la méthode	27
<u>2.3.1. Indicateurs utilisés pour comparer les performances d'Airbnb entre les communes</u>	<u>27</u>
<u>2.3.2. Rendre le code générique</u>	<u>29</u>
<u>2.3.3. Diffusion des résultats : site web</u>	<u>29</u>
Partie 3 : Résultats	31

3.1. Interprétation des résultats : Issy-les-Moulineaux	31
<u>3.1.1. État des lieux sur l'importance, la structure et la performance de l'offre Airbnb à Issy-les-Moulineaux (2019)</u>	<u>31</u>
<u>3.1.2. Analyse temporelle 2015-2020</u>	<u>33</u>
<u>3.1.3. Analyse infra-communale</u>	<u>35</u>
3.2. Comparaisons entre communes : Issy-les-Moulineaux et Pantin	36
<u>3.2.1. Un marché plus actif à Pantin, plus lucratif à Issy-les-Moulineaux</u>	<u>36</u>
<u>3.2.2. Un développement plus récent à Pantin</u>	<u>38</u>
<u>3.2.3. Mais une implantation spatiale qui reste la même</u>	<u>39</u>
3.3. Bagneux : Des hôtes aux profils divers dans une commune plus éloignée	40
Conclusion	43
Bibliographie	45
Annexes	48

Introduction

Le sujet de recherche : un projet de l'ANR WIsDHoM

Ce travail de recherche a pour objectif d'analyser les dynamiques spatio-temporelles du marché de la location saisonnière à une échelle locale. Il s'agit d'une contribution au projet WIsDHoM (coord. Renaud Le Goix), visant à étudier le lien entre dynamiques spatio-temporelles du marché du logement et inégalités de patrimoine. Les locations saisonnières, et en particulier la plateforme Airbnb, apparaissent comme une des clés de lecture de l'analyse du marché immobilier dans les grandes métropoles, de par leur présence importante en centre-ville qui contribue au retrait de nombreux logements du parc locatif privé, dans un contexte déjà très tendu d'accès au logement, où la hausse du prix des logements croît plus rapidement que les revenus depuis les années 90.

Ici, nous nous intéressons à des zones encore peu étudiées que sont la banlieue des grandes métropoles, et plus précisément la Petite Couronne parisienne, avec comme terrain d'étude les communes d'Issy-les-Moulineaux, de Pantin et de Bagneux. Si de nombreux rapports et études ont été publiés traitant le sujet à des échelles régionales (Shabrina et al., 2021 ; Institut Paris Région, 2021), ici le choix de cette zone d'étude est motivé par les contextes locaux : d'une part la ville d'Issy, qui compte une croissance du nombre de petites annonces mais aussi du parc locatif privé, où les premières régulations liées à l'expansion des locations saisonnières sont apparues assez tôt, et d'autre part Pantin, en pleine mutation depuis environ une décennie, avec notamment des projets d'amélioration et de rénovation du parc immobilier. Enfin, Bagneux correspond à un marché différent de celui des communes limitrophes (Le Corre, 2019)¹, où les logements sont « destinés à être vendus à un propriétaire occupant, à une administration publique ou à des investisseurs privés du secteur locatif ». On retrouve donc cette notion d'investissement mais à un degré moindre qu'à Issy-les-Moulineaux ou Pantin.

Pour analyser le marché de la location saisonnière dans ces communes, nous explorons une base de données encore peu utilisée à ses fins, AirDNA. Il s'agit d'un outil métrique destiné d'abord aux hôtes et permettant d'améliorer leurs performances commerciales à travers une veille concurrentielle et l'analyse de données statistiques (Ebelle, 2018). Concrètement, il s'agit d'une entreprise qui vend les données relatives aux logements Airbnb dans une commune donnée. Si de nombreux articles sur les plateformes saisonnières utilisent les chiffres issus d'une autre base de données, insideairbnb.com (Gutiérrez et al., 2017), ou bien réalisent directement leur propre *scraping* (Domènech et al., 2019 ; Dolnicar, 2017), peu se basent uniquement sur les données AirDNA². L'intérêt de la plateforme airdna.co se trouve parmi la diversité de l'offre en termes de données, ainsi que la multitude des territoires pour lesquels ces données sont disponibles, là où

¹ D'après la thèse de Thibault Le Corre (2019), Bagneux correspond à un « Marché de la promotion immobilière », tandis qu'Issy et Pantin font partie d'un « Marché d'investissement et d'occupation résidentielle dans les espaces centraux ».

² Certains articles citent les chiffres AirDNA dans une optique de comparaison entre plusieurs bases de données, mais n'utilisent pas les données à proprement parler (Adamiak et al., 2018 ; Boros et al., 2019)

insideairbnb.com ne propose que les données des municipalités centrales de quelques métropoles, dont la plupart aux États-Unis et en Europe. La différence entre les deux bases de données se trouve dans le prix : la base de données AirDNA est payante, à hauteur de 1.000\$ par commune pour tout l'historique, tandis que la base insideairbnb.com est gratuite, en libre accès. La profondeur temporelle des données est variable selon les communes, et correspond à l'année à partir de laquelle AirDNA a commencé à récolter les données. Par exemple, pour la commune d'Issy-les-Moulineaux, les données remontent jusqu'en 2015, tandis que pour Pantin et Bagneux, elles sont disponibles à partir de 2016.

Il est alors essentiel de documenter l'usage qui peut être fait de cette base de données peu connue ainsi que les traitements effectués, dans une optique de transparence et de reproductibilité des travaux. En effet, une des finalités de ce travail est la diffusion d'une méthode reproductible qui entre dans la démarche de la structure d'accueil du stage, le RIATE. La recherche de reproductibilité ainsi que le développement d'outils destinés à des travaux relevant des données géographiques et de l'analyse spatiale font partie intégrante des missions du RIATE, et ce travail s'intègre à la fois dans sa démarche scientifique, ainsi que dans les études réalisées par ses membres (Le Goix et al., 2021).

Concrètement, mon travail consiste en la production d'une chaîne de traitement issue de bases de données nouvelles, ainsi que des analyses spatio-temporelles sur l'activité des logements inscrits sur la plateforme Airbnb à Issy-les-Moulineaux dans un premier temps. Cette chaîne de traitements va de l'analyse critique des données et de leur nettoyage à l'analyse spatio-temporelle du marché de la location à court terme à l'échelle infra-communale. Dans un second temps, ces résultats seront comparés entre les différentes communes étudiées. Ils permettront d'alimenter le débat sur un potentiel débordement du marché locatif à court terme parisien vers les périphéries. La méthode employée a pour vocation d'être répliquable et extensible à d'autres terrains d'étude, afin de répondre à des standards élevés de reproductibilité (Peng, 2011).

Pourquoi l'angle des locations saisonnières à l'échelle locale ?

Il est nécessaire pour la compréhension de notre sujet de définir le terme de location saisonnière : elle apparaît en France avec la loi Hoguet de 1970 et désigne « la location d'un immeuble conclue pour une durée maximale et non renouvelable de quatre-vingt-dix jours consécutifs »³. Cette limite des 90 jours dépassée, il s'agit ensuite de location classique de longue durée. La notion de « meublé de tourisme » est introduite par le Code du Tourisme en 2006 et désigne « des villas, appartements, ou studios meublés, à l'usage exclusif du locataire, offerts en location à une clientèle de passage qui y effectue un séjour caractérisé par une location à la journée, à la semaine ou au mois, et qui n'y élit pas domicile » (article L324-1-1)⁴. La légère différence vient de l'usage exclusif du locataire, c'est-

³ Article 1.1. de la loi n° 70-9 du 2 janvier 1970 réglementant les conditions d'exercice des activités relatives à certaines opérations portant sur les immeubles et les fonds de commerce. URL : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000512228/>

⁴ https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000042070525/

à-dire que le propriétaire ne peut pas résider dans le logement qu'il loue. Une chambre chez l'habitant ne peut donc pas être considérée comme meublée de tourisme, comme l'indique la loi Elan de 2018. Plus récemment, et faisant suite à cette même loi Elan, le bail mobilité a été mis en place⁵. Il s'agit d'un contrat de location de moyenne durée (1 à 10 mois) non renouvelable d'un logement meublé, à un locataire considéré comme temporaire (étudiant, salarié en mission temporaire ou en formation professionnelle, etc). Le bail mobilité offre donc plus de flexibilité et de sécurité au bailleur (loyer fixé librement sauf en zone tendue, risque plus faible de non-paiement du loyer). Ces locations à moyen terme sont la cible d'une nouvelle stratégie de Airbnb⁶, leur permettant de contourner les restrictions en terme de nombre de nuitées réservées.

L'angle des locations saisonnières est abordé car il s'agit d'un jeune marché grandissant, en particulier au cours des années 2010. La France se classe en deuxième position derrière les États-Unis en nombre de locations saisonnières avec 600 000 logements en 2021 (Aulnay et Piganiol, 2021), dont 20% se trouvent en Île-de-France (chiffres AirDNA, 2019⁷). Un des facteurs de cette croissance se trouve dans l'ubiquité du concept : les logements mis en location sur ces plateformes font initialement partie du parc résidentiel « banal », c'est-à-dire non spécifiquement destinés à l'hébergement touristique, et par conséquent peuvent être omniprésents. C'est bien la source du problème : l'expansion de la plateforme entraîne une diminution du parc locatif privé qui exerce une pression sur les prix des logements restants (Barron et al., 2018). Le lien entre marché locatif à court terme et inégalités de patrimoine dans le logement a été étudié à l'échelle de grandes métropoles telles que le Greater London (Shabrina et al., 2021), en précisant qu'il s'agit bien d'un contributeur additionnel à la crise du logement, ou bien au niveau des centres-villes et aux problématiques associées de gentrification (Cocola-Gant et Gago, 2019). À l'origine de ce problème, le comportement des hôtes et les différentes stratégies immobilières mises en place : il existe différentes logiques de rentes selon le profil des hôtes : la rente de substitution du revenu, la rente de complément du revenu et la rente d'investissement (Semi et Tonetta, 2019). À Lisbonne, par exemple, les politiques publiques ont encouragé l'accession à la propriété à travers des encouragements fiscaux, encourageant une forme de professionnalisation et amenant des investisseurs étrangers et des sociétés privées spécialisées dans la locations saisonnière (Cocola-Gant et Gago, 2019).

L'originalité de cette étude repose donc sur l'échelle utilisée : si les études menées sur le sujet l'ont été à des échelles régionales ou bien en centre-ville, on s'intéresse ici à l'échelle infra-communale au sein d'un contexte régional d'une grande métropole en proie à un marché immobilier tendu. L'enjeu est de documenter le développement d'Airbnb dans les communes de banlieue, tout en les recontextualisant dans un cadre régional. Ce travail fait suite à mon mémoire de recherche de M1, portant sur la distribution spatiale des logements Airbnb dans les quartiers prioritaires de la ville de Saint-Denis, et visant à comparer les densités de logements dans ces quartiers prioritaires par rapport au nord-est parisien, et à chercher une forme de professionnalisation à travers les multiloueurs (Laurian, 2020).

⁵ <https://www.service-public.fr/particuliers/vosdroits/F34759>

⁶ Journal du dimanche, 06/12/2020 : « Airbnb se risque en bourse »

⁷ Ces chiffres prennent aussi en compte les logements Aritel-HomeAway, qui représentent une faible part (autour de 5%)

Objectifs du stage et hypothèses

Les objectifs sont donc :

- d'analyser la contribution des locations saisonnières à la stratification du marché du logement menant à des inégalités de patrimoine, problématique qui s'inscrit à une échelle européenne. Cette question, développée dans la partie 1, prendra pour zone d'étude l'Île-de-France, et plus spécifiquement les 3 communes citées précédemment, dans une période où l'expansion des plateformes de locations saisonnières accentue la situation d'un marché de l'immobilier déjà très tendu. Nous suivons l'hypothèse d'une boucle d'auto-renforcement du projet WISDHoM (cf partie 1), alimentée par le rapport de stage de Pierre Le Brun réalisé en 2020 au RIATE.
- de documenter une nouvelle base de données encore peu utilisée ainsi que son exploration. Il s'agit là d'un travail d'analyse critique sur une source qui répond à notre problématique de recherche car il s'agit d'une des seules bases de données correspondant à notre zone d'étude (communes de banlieue) et offrant des données qu'on ne trouve pas ailleurs, notamment celles relevant de la fréquentation des logements, ainsi que des estimations des revenus engendrés. La chaîne de traitements, les outils utilisés pour l'exploration ainsi que les limites inhérentes à cette base de données seront analysées dans la partie 2 de ce rapport.
- ensuite, plus concrètement - et à l'aide de pré-traitements effectués sur cette même base de données AirDNA -, d'étudier le comportement du marché Airbnb dans la commune d'Issy-les-Moulineaux : quelle est sa dynamique depuis 2015 ? Au sein de la commune, quelle forme prend-il (est-il concentré ? où ?) ? Il sera nécessaire de définir et justifier le choix des indicateurs permettant de quantifier l'activité d'Airbnb dans la commune. L'hypothèse est la suivante : le marché parisien est débordant, et l'activité au sein des communes limitrophes suit la même dynamique qu'à Paris avec un certain temps de latence. Dans un second temps, nous comparerons les résultats d'Issy-les-Moulineaux avec les communes de Pantin et de Bagneux. Si on connaît bien la situation de Paris, celle en Petite Couronne est certainement différente. On émet l'hypothèse que les contextes locaux pèsent sur l'activité de la location saisonnière dans la commune (notamment pour les communes d'Issy et Pantin qui sont limitrophes).
- tout en s'inscrivant dans une démarche de recherche reproductible : qu'en est-il du degré de reproductibilité des analyses menées ? Si les outils utilisés permettent une excellente documentation et répliquabilité des méthodes, la question des données pose problème. La situation de la reproductibilité dans la recherche est un sujet d'actualité (Schloss, 2018), et nous expliquerons la nécessité d'une méthode reproductible, tant pour la vérification des résultats que pour d'éventuelles futures analyses sur le sujet.

Partie 1 : Locations saisonnières et crise du logement dans les grandes métropoles

1.1. Une contribution à l'ANR WISDHOM

1.1.1. Un projet axé sur le marché du logement

Ce travail de recherche s'inscrit dans le projet WISDHOM de l'Agence Nationale de Recherche (coord. Renaud Le Goix) qui vise à comprendre le rôle de l'immobilier dans le renforcement des inégalités spatiales et patrimoniales. L'hypothèse principale du projet WISDHOM est celle d'un effet d'auto-renforcement à partir de deux processus qui s'auto-entretiennent : d'un côté, les prix de vente des logements croissent plus vite que les revenus depuis les années 90, ce qui tend à exclure une part toujours croissante des catégories les plus modestes de l'accès au logement dans les zones les plus tendues. D'un autre côté, cette augmentation des prix de l'immobilier renforce l'intérêt d'investir et les perspectives de plus-value de la part des ménages les plus aisés, des ambitions plutôt à moyen terme. Tout ceci dans un contexte de transition du régime d'État-providence et de volatilité des marchés qui encouragent de plus en plus les placements immobiliers des ménages. On assiste donc à une boucle de rétroaction qui alimente sans cesse l'augmentation des prix, encouragée par des dispositifs d'aide à l'accession à la propriété. L'objectif du projet WISDHOM est donc de comprendre les interactions entre la dynamique locale des prix, les inégalités et les politiques menées à une échelle multiscalaire en s'appuyant sur trois zones d'étude que sont les *Functional Urban Areas* (FUA) de Paris (1781 communes), Lyon (316 communes) et Avignon (45 communes). Ces aires urbaines, notamment celle de Paris, n'échappent pas à cette situation de crise du marché immobilier.

Une première partie du projet WISDHOM vise à analyser des dispositifs publics de soutien à l'accession à la propriété immobilière tels que les prêts à taux zéro ou bien la loi Pinel⁸, mis en place dans les années 2010. Cette dernière encourage l'investissement locatif sur le long terme (minimum 6 ans) en tant que résidence principale non meublée. Un rapport de Pierre Le Brun (2020), stagiaire à l'UMS RIATE entre février et juin 2020, indique que « les aides fiscales à l'accumulation de patrimoine augmentent et se concentrent dans les centres urbains ». Ces aides sont destinées à des buts lucratifs telles que la possibilité d'engendrer une rente à partir du patrimoine immobilier. À l'inverse, « les aides à la construction d'un patrimoine non lucratif se limitent aux territoires ruraux et urbains périphériques ». Ce focus de l'accession à la propriété sur un marché urbain déjà très tendu ne fait qu'alimenter la boucle d'auto-renforcement.

Le deuxième volet du projet a pour objectif de cartographier les inégalités spatiales des trajectoires des prix du logement et de l'accès à la propriété immobilière. À l'aide de données relatives aux

⁸ <https://www.economie.gouv.fr/cedef/pinel-investissement-locatif>

transactions administrées par la Chambre des Notaires (bases de données BIEN et PERVAL), les ingénieurs du RIATE ont agrégé à l'échelle de la commune ou du carreau les données de 2,3 millions de transactions immobilières. Les premiers résultats indiquent que la volatilité relève de dynamiques locales et que la vulnérabilité des ménages les plus pauvres est façonnée par une combinaison de facteurs liés à l'endettement des ménages par rapport à la valeur élevée des logements, de risque de baisse de la valeur des propriétés, de discrimination sociale ou raciale des ménages conduisant à une évaluation des risques élevés (coût élevé des assurances et des prêts immobiliers), ainsi que par une précarité accrue de la situation professionnelle d'un ménage.

1.1.2. Intérêt d'étudier les locations saisonnières dans le cadre de ce projet

Mon travail s'inscrit dans le troisième axe de ce projet de recherche, qui se focalise sur les stratégies des ménages qui ont fait l'acquisition d'un logement et sur les usages associés. Parmi ces usages, la mise en location saisonnière en tant que stratégie émergente de valorisation du patrimoine immobilier. Cette stratégie est interprétée par l'Institut Paris Région (2021, p.13) comme la traduction d'une volonté des particuliers de se comporter comme gestionnaires d'actifs sur le court terme plutôt qu'en tant que gestionnaire de patrimoine sur le long terme. Cette incitation à la location à court terme a été encouragée par les dispositifs en faveur de l'investissement locatif des ménages, dont la maîtrise est progressivement passée aux mains d'acteurs privés (Pollard, 2010). Cette divergence d'interprétation de la temporalité de l'investissement provient des profils divers des hôtes et de leurs intérêts d'investir dans ce type de location, mais aussi du contexte économique : le cas d'étude du quartier d'Aurora, situé à proximité du centre-ville de Turin, met en lumière la production d'opportunités professionnelles à travers les rentes issues du secteur du logement, dans un contexte d'austérité : Semi et Tonetta (2019) distinguent 3 types de rentes (de substitution, de complément et d'investissement). Elles correspondent respectivement à une forme de protection pour les propriétaires les plus pauvres ayant perdu leur source de revenus à la suite de la perte d'un travail, à un moyen de compléter des revenus devenus insuffisants, et pour le dernier cas de figure à une stratégie qui s'inscrit dans un système italien et plus généralement sud-européen de culture de la propriété (Allen et al., 2004), où l'achat d'une seconde propriété devient un objectif d'investissement et offre une possibilité d'accumulation des richesses. Ces rentes constituent des stratégies pour résoudre partiellement ou entièrement une situation de précarité, les propriétaires voyant en la location à court terme une minimisation des risques de défaut de paiement. À Lisbonne, on observe une forme de professionnalisation parmi les hôtes, qui résulte d'un processus de gentrification au niveau du centre-ville. Les hôtes sont pour la plupart des investisseurs-promoteurs et des sociétés privatisées spécialisées dans la location saisonnière, tandis que les locaux sont exclus de ces quartiers touristiques par la prolifération des locations saisonnières. Cette part d'hôtes-investisseurs a été quantifiée à un niveau de 78% de la totalité des hôtes lisboètes (Cocola-Gant et Gago, 2019) et fait suite à des décisions politiques qui facilitent l'accession à la propriété à l'aide d'encouragements fiscaux. Là aussi, cette dynamique d'investissement dans les location à court terme est motivée par un contexte de crise du logement couplée à des orientations de la part des politiques publiques vers l'accession à la propriété, et ne fait que renforcer cette crise. Plus récemment en France, des entretiens menés par Piganiol (2020) montrent que les stratégies diffèrent

selon que l'hôte soit multiloueur ou non : si les multiloueurs sont ceux qui engendrent le plus de revenus de la part des plateformes de locations saisonnières, leur objectif étant de se constituer un capital économique, ils en sont par conséquent les plus dépendants, tandis que les monoloueurs louant occasionnellement leur logement principal voient la sous-location comme un revenu annexe. On peut recouper cette typologie avec les différents types de rentes cités précédemment.

Dans tous les cas de figure, le marché de la location saisonnière serait une des clés d'interprétation de l'analyse du marché immobilier dans les grandes métropoles. L'évolution fulgurante du parc locatif de courte durée depuis une décennie est au cœur des enjeux dans les grandes métropoles, et nous la détaillons ci-dessous à différentes échelles.

1.2. Évolution de la présence de locations Airbnb à différentes échelles

Les informations sur la présence des locations saisonnières en Île-de-France ainsi que les restrictions mises en place évoquées dans cette partie sont issues pour une grande partie du rapport de l'Institut Paris Région (2021). Ce rapport très récent, publié en mai 2021, propose un compte-rendu sur le cadre juridique en constante évolution des locations saisonnières à plusieurs échelles, ainsi qu'une analyse de l'impact qu'ont ces logements sur le parc de résidences principales en Île-de-France.

1.2.1. En France : une offre calquée sur les densités de population et de la fréquentation touristique

Du point de vue des hôtes, un des points forts de la location saisonnière, par rapport à la location classique de longue durée se trouve d'une part dans la rentabilité mais aussi dans la flexibilité. En effet, d'après une étude de l'APUR en 2011 sur les locations meublées de courte durée à Paris, louer un appartement à la semaine pendant 4 semaines revient 1,6 fois plus cher qu'une location au mois. On parle de flexibilité car ces plateformes offrent une possibilité de profits *via* la location à court terme, mais les propriétaires peuvent à tout moment revendre leur logement. À Lisbonne, on retrouve un processus de *rental gentrification*, à savoir que les propriétaires Airbnb ciblent des locataires plus aisés. De plus, les prix pratiqués ne sont pas soumis à un encadrement des loyers comme pour la location privée classique, et les charges sont moindres. Airbnb n'échappe pas à la règle et couvre désormais la quasi intégralité du territoire français. Les travaux de Vincent Aulnay⁹ (collectif ParisvsBnB) rendent compte de la distribution des quelques 600.000 annonces réparties sur l'ensemble de la France métropolitaine, enregistrées sur la plateforme entre décembre 2020 et mars 2021. La densité de l'offre est évidemment plus forte autour des grandes villes et des pôles urbains que dans les zones rurales. La littoralisation d'Airbnb est elle aussi très marquée, et le tracé

⁹ carte interactive des logements Airbnb en France : <https://parisvsbnb2.carto.com/builder/9ea3868d-2571-4984-8010-30ac096bd2b1/embed>

de l'implantation de la plateforme suit une géographie du tourisme, autour des stations de montagne ainsi que dans des zones ayant une forte identité territoriale liée à des politiques de patrimonialisation (tracé de la Loire, plages normandes...). Finalement, la carte des locations Airbnb est calquée sur celle des densités de population, de la fréquentation touristique et de l'urbanisation, car les logements loués sont pour la plupart d'entre eux habités par les propriétaires (Aulnay & Piganiol, 2021).

1.2.2. En IDF : Un marché parisien saturé, se diffusant vers la périphérie ?

C'est en Île-de-France que le marché de la location saisonnière est le plus présent avec pas moins de 122 000 annonces en 2019, dont les 2/3 seulement pour la capitale (chiffres AirDNA, 2019¹⁰). L'Institut Paris Région estime que 80% de ces annonces proviennent du site Airbnb. À la différence des autres métropoles françaises, celle du Grand Paris enregistre plus de nuitées hôtelières de la part de touristes étrangers que de touristes français en 2018¹¹. La cohabitation entre locations saisonnières et chambres d'hôtels est plutôt homogène, mais reste en proie à des dissymétries entre Petite et Grande Couronne : à Paris et en Petite Couronne, le poids en lits de la location saisonnière est semblable au poids en lits de l'offre hôtelière, tandis que pour les départements de Grande Couronne, les hôtels, campings et résidences de tourisme pèsent plus que les locations saisonnières en termes de capacité d'hébergement (Institut Paris Région, 2021). Mais si les hôtels ont des capacités d'accueil plus élevées, leur couverture géographique est bien inférieure à celle des locations saisonnières : Paris exclu, le marché hôtelier était présent dans un peu plus de 300 communes franciliennes en 2019, tandis que dans le même temps, les plateformes de locations saisonnières étaient implantées dans plus de 1100 communes (base de données Trackeet citée dans le rapport de l'Institut Paris Région, 2021). Ainsi, pour la majorité des communes de Grande Couronne, l'offre d'hébergement touristique repose uniquement sur la location saisonnière. Les enjeux de la présence des locations saisonnières varient donc selon la localisation dans la région.

En observant la carte de la distribution des locations saisonnières dans le parc de logement des communes franciliennes en 2019 (cf annexe 1), on observe sans surprise des chiffres élevés à Paris, mais aussi autour de grands pôles touristiques (Versailles, Disneyland), et à proximité des Parcs Naturels Régionaux (Vexin, Gâtinais, Haute Vallée de Chevreuse). Ces densités traduisent des formes de tourisme multiples, du tourisme de loisir à Disneyland au tourisme vert (qui correspond à une demande nationale) à Fontainebleau, en passant par le tourisme d'affaire dans des communes comme Guyancourt ou bien des étudiants venus passer une nuit près du centre d'examen de Cergy. Autant de formes de tourisme qui traduisent autant de profils de clientèle. Ces formes de tourisme peuvent aussi être hybrides, se mélanger au sein d'un même séjour, d'un même territoire, et traduisent une hétérogénéité du profil des visiteurs.

¹⁰ Ces chiffres prennent en compte les logements Airbnb et Aritel-HomeAway

¹¹ <http://www.atout-france.fr/actualites/tourisme-urbain-l-internationalisation-des-nuitées-hotelières-progresse>

1.2.3. Notre zone d'étude : Les communes d'Issy-les-Moulineaux, Pantin et Bagneux

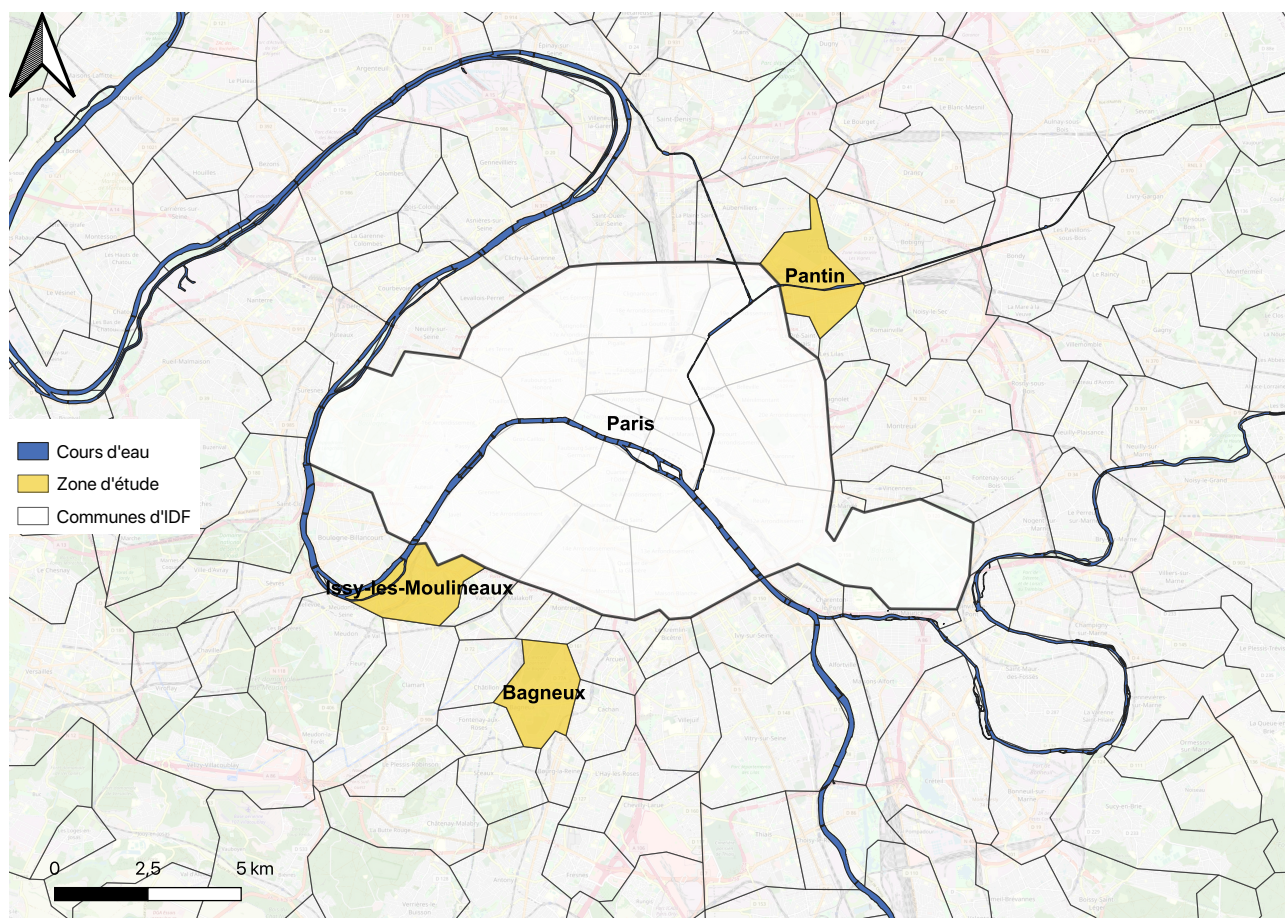
Le département Hauts-de-Seine (92) est, après Paris, celui où les locations saisonnières représentent la plus grande partie de l'offre de meublés de tourisme avec près de 15 000 offres en 2019 (IPR, 2021), soit 10% des offres totales en Île-de-France. À titre de comparaison, Paris regroupait à elle seule près de 100.000 annonces à la même période. La proximité à la capitale et à Versailles ainsi que le quartier d'affaire de La Défense peuvent expliquer ces fortes concentrations d'offres, qu'elles soient de location saisonnière ou hôtelière. En Seine-Saint-Denis (93), Airbnb a été popularisée notamment en 2016 à la suite de la coupe d'Europe de football (et la proximité au stade de France), et on y trouvera en 2024 le village olympique. Le département subit depuis peu de nombreuses opérations d'aménagement tournées sur l'immobilier et des communes telles que Saint-Denis constituent des communes à fort potentiel touristique (Laurian, 2020).

Nos zones d'étude sont les communes d'Issy-les-Moulineaux, Pantin et Bagneux (carte 1). Dans un premier temps, notre travail s'est concentré sur la commune d'Issy-les-Moulineaux (92). Commune limitrophe de 70 000 habitants située au sud-ouest de Paris et traversée par la Seine, elle présente certains atouts touristiques, comme le Parc des Expositions de la Porte de Versailles, ainsi que l'île Saint-Germain, mais aussi de nombreux immeubles de bureaux. Depuis le début des années 90, la densité de population a augmenté de près de 50% (INSEE, 2018). Le nombre de logements suit la même tendance avec une augmentation de 54% entre 1990 et 2015. Au niveau de la composition du marché du logement, on compte 4% de résidences secondaires et 5,4% de logements vacants sur les 35182 unités de la commune. 38,5% des ménages sont propriétaires occupants de leur logement, soit environ 7 points de pourcentage de moins que la moyenne du Grand Paris Seine-Ouest (T3), qui réunit 8 communes des Hauts-de-Seine (base de données APUR, recensement commune logement 2015¹²). Tandis que le marché parisien semble montrer des signes d'accalmie, comme en témoigne la baisse des loyers dans la totalité des arrondissements parisiens du 01/04/2020 au 01/04/2021 (Le Parisien, 2021), offrant une fenêtre d'entrée aux primo-accédants, Issy-les-Moulineaux est devenue une des villes les plus chères d'Île-de-France avec un prix moyen au m² d'environ 8800€ au 1er mai 2021 selon le site Meilleurs Agents¹³.

¹² URL : <https://opendata.apur.org/datasets/recensement-commune-logement>

¹³ <https://www.meilleursagents.com/prix-immobilier/issy-les-moulineaux-92130/>

Carte 1 : Localisation de la zone d'étude (Ile-de-France)



Réalisation : Laurian (2021). Fond de carte OSM Standard

Issy-les-Moulineaux et Boulogne-Billancourt sont les premières communes à avoir mis en place une autorisation de location sous condition de compensation au premier logement, ainsi qu'une télédéclaration préalable obligatoire, et ce dès 2018¹⁴. Il s'agit aussi de communes avec un cas de figure particulier : on y observe une croissance du parc de résidences principales (RP) mais aussi hors-RP entre 2011 et 2017 (Dossier complet INSEE, commune d'Issy-les-Moulineaux, 2018). Ceci peut s'expliquer par la dynamique de construction de logements neufs dans la commune. Dans sa thèse, Thibault Le Corre classe les communes franciliennes à partir de la base des données notariales sur les transactions immobilières (BIEN) entre 2008 et 2012, et on remarque que les arrondissements de l'Est parisien ainsi que la quasi-intégralité des communes limitrophes sont considérés comme faisant partie d'un « Marché de l'investissement et d'occupation résidentielle dans les espaces centraux » (Le Corre, 2019). Issy-les-Moulineaux, où les prix au m² sont devenus parmi les plus élevés en Petite Couronne, représente donc un cas d'étude intéressant, dans la mesure où la location saisonnière s'y développait de façon prospère jusque début 2020, malgré les régulations précoces imposées dans la commune. Issy-les-Moulineaux fait partie des terrains de l'ANR WiSDHoM en tant que commune favorisée où les prix des logements sont élevés et où on trouve d'importantes stratégies d'investissements immobiliers. En effet, le développement de la

¹⁴ Suite à une délibération intercommunale de l'EPT Grand Paris Seine Ouest (T3) le 26/06/2018 : https://www.seineouest.fr/app/uploads/2020/07/Rapport_activite_GPSO_2018.pdf

location saisonnière en banlieue parisienne a pris de l'ampleur depuis plusieurs années, même s'il est loin du niveau de Paris intra-muros, et Issy-les-Moulineaux figure parmi les communes où leur nombre a le plus augmenté, dans le département de petite couronne le plus concerné.

Thibault Le Corre (2019) distingue onze types de régimes locaux du marché francilien. Issy-les-Moulineaux et Pantin appartiennent selon ses résultats à un même type de régime, à savoir un « Marché d'investissement et d'occupation résidentielle dans les espaces centraux », où l'accès au marché est concurrentiel en fonction de la rentabilité. Bagneux, quant à elle, correspond à un « Marché de la promotion immobilière » où les fonctions principales des propriétés relèvent de l'occupation résidentielle, et où l'on observe une stratification de l'accès au marché par les mécanismes de filtrage social et d'accès au crédit. Le profil socio-économique de la commune de Pantin est très différent de celui d'Issy-les-Moulineaux, tandis que celui de Bagneux est plus diversifié.

Dès 2008, l'APUR prévenait des futures opérations d'aménagement dans la ville de Pantin, notamment autour des quais du canal de l'Ourcq (APUR, 2008). Territoire anciennement marqué par les infrastructures ferroviaires, l'industrie et les abattoirs, la ville a subi de profondes mutations structurelles depuis les années 80. Plus récemment, c'est l'apparition du tramway T3 en 2012 qui a initié les grandes opérations d'aménagement des années 2010. Construction de ponts et de voies cyclables le long du canal de l'Ourcq, requalification des liaisons Paris-Pantin, relocalisation de la déchetterie, autant de projets structurants qui ont contribué à remodeler le visage de la commune, la rendant plus accueillante et plus apte à répondre à une demande touristique et à établir une recomposition sociale des résidents. Aujourd'hui, ce sont l'ensemble des canaux de l'Ourcq qui font l'objet d'une ZAC (zone d'aménagement concertée), dont un des objectifs consiste à l'amélioration du parc immobilier. En décembre 2018, la préfecture a validé par un arrêté la création de la zone touristique de la Plaine de l'Ourcq afin de créer une nouvelle destination à l'échelle du Grand Paris¹⁵.

1.3. Facteurs explicatifs de la croissance de la location saisonnière

1.3.1. Une croissance récente

Bien que le concept de locations saisonnières date de la fin du XXe siècle, elles ont connu une explosion dans les années 2010, dans un contexte de digitalisation du marché où les plateformes en ligne ont pris le rôle d'intermédiaire entre hôtes et visiteurs. Le principe se distingue par son ubiquité : les logements mis en location sur ces plateformes font initialement partie du parc résidentiel « banal », c'est-à-dire non spécifiquement destiné à l'hébergement touristique, et par

¹⁵ Est Ensemble, 2019 : <https://www.est-ensemble.fr/la-creation-de-la-zone-touristique-plaine-de-lourcq-une-etape-importante-dans-la-construction-du>

conséquent peuvent être omniprésents. Comme dit précédemment, ces plateformes couvrent un territoire bien plus large que le marché hôtelier, limité aux plus grandes communes. Ainsi, elles offrent une plus grande diversité de choix de localisation pour les visiteurs. L'intérêt de ces plateformes se trouve à la fois chez les visiteurs, avec des prix initialement plus attractifs que le marché hôtelier (ce qui n'est plus le cas dans des villes telles qu'Amsterdam ou Barcelone), mais aussi chez les hôtes, la location saisonnière peut représenter un complément de revenus occasionnel comme cité précédemment, en étant plus lucratif que la location dite classique. De plus, la mise en location d'un logement est gratuite sur certaines plateformes de location saisonnières telles que Airbnb. Cette plateforme touche tous les propriétaires, peu importe leur catégorie sociale : d'après leur site, plus de la moitié des hôtes Airbnb français ont des revenus inférieurs au revenu médian français (30 300€ par an)¹⁶. Elle n'est donc pas exclusivement réservée aux propriétaires les plus aisés.

Maintenant, du point de vue spatial, ces logements servent initialement à répondre à une demande touristique. On en retrouve donc d'abord au niveau des centres-villes. Une diffusion vers les périphéries peut être due à un marché saturé au niveau du centre, comme c'est le cas à Paris, où l'offre de logements Airbnb diminue depuis le troisième trimestre 2017, passant d'environ 70 000 à moins de 60 000 deux ans plus tard (chiffres AirDNA). Après la commune centre, ce sont les communes limitrophes qui captent les arrivées touristiques, car elles représentent une solution à moindre coût en offrant pour certaines d'entre elles une forte accessibilité à la commune centre.

1.3.2. Impact des mesures de régulation sur l'évolution de l'offre

Les années 2010, qui marquent pour l'instant l'âge d'or de la location saisonnière, sont aussi celles de l'apparition des régulations sur le marché du locatif à court terme. Les différentes réglementations mises en place ont des effets plus ou moins marqués selon les communes : à Paris, le nombre de locations saisonnières ne cesse de baisser depuis 2017, année précédant la mise en place de la majorité des restrictions. Des obligations telles que la compensation sont efficaces dans la capitale car la place est limitée et il coûte cher de construire une superficie égale. Mais avec cette réglementation s'est développé un « marché de la compensation », qui consiste en acquérir des locaux commerciaux pour les transformer en locations saisonnières. En ce qui concerne la Petite Couronne, on observe une stabilisation, voire une diminution du nombre de locations saisonnières, mais il est encore difficile de déterminer si cette dynamique est due à un effet de saturation ou aux régulations mises en place. Dans les deux cas de figure cependant (Paris et Petite Couronne), le nombre de locations saisonnières « à l'année », qui dépassent le seuil des 120 jours, ne cesse d'augmenter (Institut Paris Région, 2021).

¹⁶ source : l'impact économique d'Airbnb en France. URL : <https://blog.airbnb.com/limpact-economique-dairbnb-en-france/>

Mais les restrictions sont limitées : au 1/1/2021, seules 48 communes ont mis en place une limitation à 120 nuitées à l'année¹⁷, dont 19 se situent en Île-de-France. La mise en place de ces outils est très récente, ainsi le nombre de communes les appliquant est amené à augmenter durant les prochaines années, à moins que l'effet Covid eût été assez important pour freiner suffisamment l'expansion de la plateforme, qui sera probablement amenée à repartir. La question est de savoir à quelle vitesse ce redémarrage aura lieu, compte tenu de la flexibilité de ces plateformes. De plus, un développement entravé par des régulations mises en place au niveau des métropoles pourrait provoquer une expansion vers les communes moins régulées. C'est en tout cas la question posée par le rapport de l'IPR (2021).

Il est donc trop tôt pour mesurer l'impact des mesures de régulation. Les procédures judiciaires sont longues et souvent gelées à cause des longs délais d'application des outils. Trop peu de moyens financiers et humains sont déployés pour contrôler plusieurs dizaines de milliers d'annonces : d'après une note de l'APUR (2018), on ne compte que 29 personnes affrétées à la mise en oeuvre de la réglementation à Paris, contre 60 à Berlin et 76 à Barcelone.

1.3.3. Impact des locations saisonnières (et de leur expansion) sur le marché du logement

Le retrait d'un nombre conséquent de logements sur un marché privé déjà en proie à de fortes augmentations du prix des logements depuis les années 2000 ne fait qu'empirer avec le développement des plateformes de location saisonnières, et diminue l'offre de logements disponibles, appliquant une pression sur les prix immobiliers. Un indicateur intéressant pour mesurer l'importance que prend la location saisonnière dans une commune est l'évolution du nombre de logements au sein et en dehors du parc de résidences principales (RP). En effet, un logement devant être retiré du marché du logement classique afin d'être reconverti en location saisonnière sera classé comme hors-RP. À noter qu'un logement hors-RP ne sera pas forcément destiné à ce type de séjour. En outre, une recrudescence du nombre de résidences secondaires peut indiquer une réorientation vers le locatif à court terme. Si le phénomène est bien plus marqué à Paris (-23 000 logements RP, + 50 600 logements hors-RP entre 2011 et 2017 selon les données du recensement) que dans le reste de l'Île-de-France, c'est bien dans le département des Hauts-de-Seine qu'on observe les plus grosses pertes de résidences principales : -445 à Rueil-Malmaison, -785 à Neuilly-sur-Seine, et -1940 à Courbevoie. Dans le même temps, le parc hors-RP de ces communes a fortement augmenté (respectivement +1053, +751 et +3389). L'infographie proposée par l'Institut Paris Région (annexe 2) permet d'estimer la contribution des locations saisonnières à l'année (c'est-à-dire louée plus de 120 nuitées par an) dans l'évolution du parc hors-RP¹⁸. Elle se situe entre 15 et 17% pour les communes de Boulogne-Billancourt, Issy-les-Moulineaux, Levallois-Perret et Neuilly-sur-Seine. Il est nécessaire de préciser néanmoins qu'on ne constate pas de

¹⁷ liste : <https://www.airbnb.fr/help/article/2108/limitation-du-nombre-de-nuit%C3%A9es-en-france%C2%A0-foire-aux-questions>

¹⁸ Cette méthode se base sur le calcul du ratio (nombre de locations saisonnières louées plus de 120 jours par an) / (croissance de logements hors-RP) sur la période 2011-2017 (données du recensement). Cette méthode a ses limites car le recensement reste déclaratif.

corrélation entre le nombre de locations saisonnières à l'année d'une commune et la contribution estimée de ce parc spécifique à la hausse de son parc hors-RP.

Une conséquence de la diminution de l'offre de logement est la pression au niveau des prix : on remarque une accélération de l'augmentation des prix des logements : depuis 2010, les logements anciens ont vu leurs prix augmenter de 63% à Paris, et 39% dans le département Hauts-de-Seine (INSEE, 2020), rendant encore plus difficile l'accès à la propriété pour les primo-accédants et les ménages de catégorie modestes. Les classes moyennes sont donc exclues du centre-ville où le logement est de moins en moins accessible et on assiste donc à un processus de gentrification comme c'est le cas à Reykjavik (Mermet, 2019). De plus, toujours sur le marché du logement, Airbnb est accusée de provoquer une hausse des loyers dans les villes où les prix sont déjà très élevés (Office of the New York City Comptroller, 2018). Les plateformes de location saisonnière ne portent pas à elles seules le poids du mouvement inflationniste du marché, mais elles participent, à leur échelle, à accentuer la crise du logement qui frappe les grandes métropoles.

Partie 2 : Une méthodologie pour une recherche reproductible

2.1. Vers des nouveaux standards de reproductibilité dans la recherche

Un des objectifs du stage était d'obtenir une méthode reproductible : R concentre toutes les étapes de traitement, de l'importation des données à la diffusion des résultats en passant par les méthodes de manipulation des données et des analyses statistiques et cartographiques plus poussées. Le rendu final prend la forme d'un Markdown, généré lui aussi à l'aide de R. Le format Markdown permet de générer des rapports automatisés en mélangeant du code (que l'on peut masquer ou afficher à sa guise) et du texte (que l'on peut modéliser à partir des résultats du code). Cet outil est très utilisé dans la communication de résultats d'analyse et la documentation des chaînes de traitement.

2.1.1. La reproductibilité au coeur de la démarche scientifique de la structure d'accueil : le RIATE

Mon travail ne consiste pas à proprement parler d'étudier les stratégies des hôtes, mais se base sur des éléments de cadrage : observation des dynamiques spatio-temporelles des locations saisonnières à l'échelle communale, réflexion sur les indicateurs permettant de suivre localement la diffusion de ce phénomène et les transformations de l'offre au cours des dernières années. Il s'inscrit donc dans le prisme de la géographie quantitative, fer de lance du RIATE. Un autre volet du stage est celui de la reproductibilité. Le RIATE est un précurseur en la matière et la majorité de ses travaux sont réalisés dans une optique de transparence et de répliquabilité de la méthode. En effet, une des missions du RIATE est la diffusion et le partage des savoir-faire et des connaissances dans une démarche de science ouverte, à travers des outils et méthodes relevant de la science de l'information géographique, facilement répliquables. Les données utilisées pour les travaux du RIATE sont principalement issues de l'open data, et les méthodes sont partagées via de sites de partage et de versionnement dédiés (Github, logiciel de *version control system*). Le RIATE met l'accent sur la gestion et la diffusion des données de la recherche, ainsi que la propriété des données. Pour cela, un travail de documentation efficace doit être réalisé. Prenons l'exemple du travail sur l'étude des centralités transfrontalières¹⁹ : les ingénieurs du RIATE ont développé un site internet sous la forme de plusieurs Markdowns où chaque rubrique correspond à une étape de travail. Pour la documentation, il s'agit de la partie Données du site internet du projet en question, avec pour chaque source utilisée une description des variables disponibles ainsi que les conditions d'accès et d'utilisation de ces données. La partie Préparation détaille les opérations de pré-traitement des données, et la rubrique Analyse de Complétude OSM-BPE permet de comparer les points d'intérêt issus de la BPE avec ceux issus de la base de données OpenStreetMap. La rubrique Cas d'étude indique les traitements et analyses menées sur les 5 zones d'étude sélectionnées, et les résultats sont

¹⁹ <https://rcarto.gitpages.huma-num.fr/centralite/index.html>

indiqués dans la rubrique Analyse transfrontalière. Tout le code utilisé est visible et documenté, de sorte qu'il soit possible de reproduire le travail à l'identique depuis sa machine, à condition d'avoir les données. Ce stage a pour vocation de suivre le même modèle, à savoir proposer un site web qui concentre toutes les opérations de pré-traitement et d'analyse, ainsi que les deux fichiers contenant les résultats, à la différence que les données utilisées ici sont payantes, et non diffusables.

2.1.2. Une base de données particulière : AirDNA

Les données à notre disposition sont issues de la plateforme airdna.co. Il s'agit d'un outil métrique destiné aux hôtes Airbnb, leur permettant d'améliorer leurs performances commerciales à travers de la veille concurrentielle et l'analyse de données statistiques. Il existe une version gratuite et une autre payante, plus détaillée. Nous avons utilisé cette dernière car elle apporte une profondeur temporelle aux données, ainsi que des données relevant de la fréquentation effective des logements et des revenus engendrés par les hôtes, données disponibles nulle part ailleurs. L'extraction effectuée dans le cadre de l'ANR WISDHOM porte sur l'ensemble des logements Airbnb et HomeAway d'Issy-les-Moulineaux, Bagneux et Pantin (respectivement 1878, 505 et 1945 logements). Les données collectées par AirDNA l'ont été sur la période août 2015 - novembre 2020.

La méthode de collecte de données utilisée par la plateforme est la suivante : les données ont été récoltées à 90% par *webscraping*, c'est-à-dire des programmes informatiques chargés d'extraire de l'information dans le code source d'une page web, ici des pages correspondantes aux annonces présentes sur le site airbnb.com de chaque logement Airbnb de la ville d'Issy-les-Moulineaux. Les données récoltées dépendent donc de l'encodage du site internet. De plus, le site internet AirDNA.co propose des informations qui ne sont pas disponibles directement sur le site internet airbnb.com car elles sont fournies par les hôtes (les 10% restantes sont issues des partenaires et des hôtes eux-mêmes²⁰). On peut citer par exemple l'estimation des revenus engendrés par un hôte ou bien le taux de fréquentation.

La méthode du *webscraping* a aussi été utilisée par Murray Cox sur son site insideairbnb.com, qui recense les informations des logements Airbnb de plus de 80 grandes villes autour du monde. De nombreuses études sur le sujet utilisent cette base de données, comme Domènech et al. (2019) ou bien Adamiak et al. (2019), qui étudient respectivement l'implantation de la plateforme en Suisse et en Espagne. La différence avec la base de données en libre accès de insideairbnb.com est que AirDNA.co propose les données des communes de Petite Couronne, tandis que les données insideairbnb.com se limitent aux villes centre des grandes métropoles. Ainsi, des études qui se basent sur une région entière, telle que le rapport de l'IPR, utilisent des données issues de la plateforme airdna.co.

Une limite majeure à la reproductibilité des analyses apparaît, inhérente aux données utilisées : de par leur caractère commercial, elles ne sont pas diffusables. Whitaker (2017, cité dans Schloss,

²⁰ <https://www.AirDNA.co/AirDNA-data-how-it-works>

2018) établit une typologie des concepts utilisés pour décrire une méthodologie (figure 1) en se basant sur la méthode et les données utilisées : le terme « répliquabilité » correspond donc partiellement à notre travail, car la méthode a été généralisée au point de correspondre avec n'importe quel jeu de données, à condition qu'il ait le même format et qu'il vienne du même fournisseur que le nôtre, à savoir AirDNA.

Figure 1 : typologie et différents degrés de reproductibilité

Methods	Same experimental system	Different experimental system
Same methods	Reproducibility	Replicability
Different methods	Robustness	Generalizability

source : Schloss, 2018

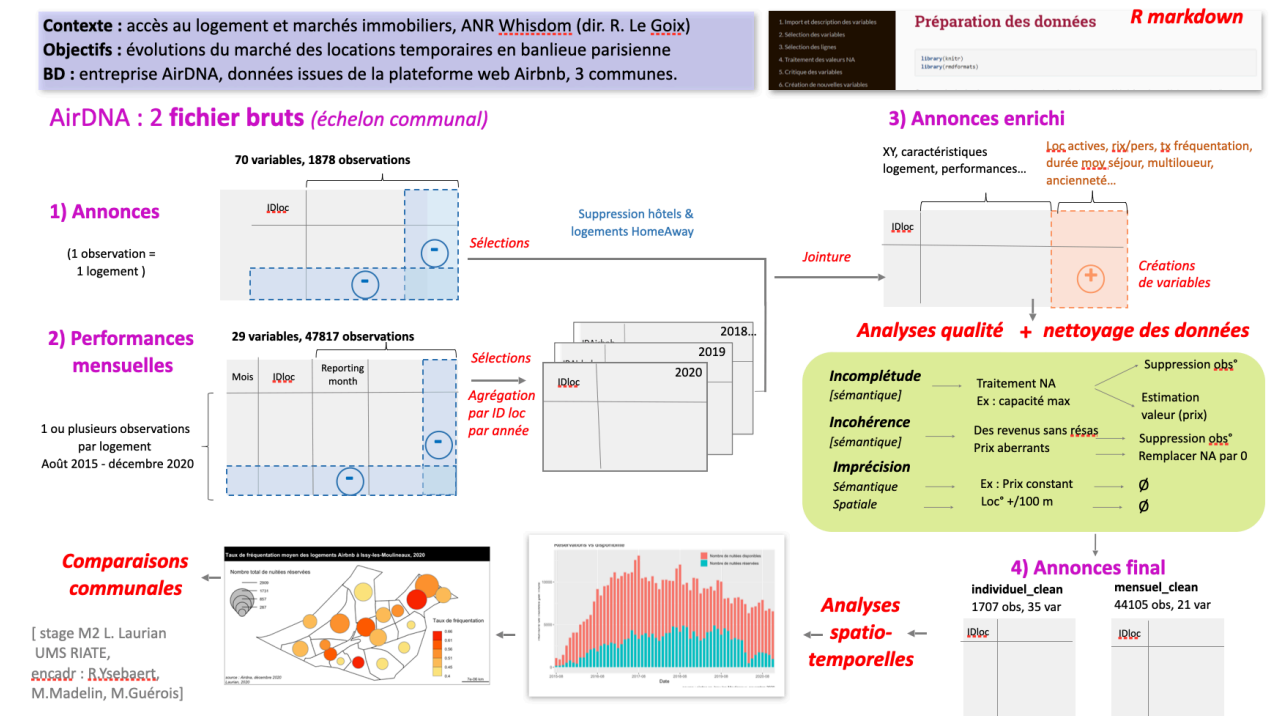
2.2. Mise en œuvre d'une chaîne de traitement permettant d'exploiter le contenu de la base de données AirDNA

On a à notre disposition deux tableaux de données (cf annexe 3). Le premier tableau, qu'on appellera par la suite tableau mensuel, contient les bilans mensuels des scraping. Cette base de données indique les mois durant lesquels chaque logement a été scrapé ainsi que les bilans en termes de fréquentation et de revenus engendrés durant le mois scrapé. Pour Issy-les-Moulineaux, il s'agit d'un tableau de 47817 observations de 29 variables. La seconde base de données, appelée tableau individuel, est plus détaillée mais est individuelle pour chaque logement Airbnb. En effet, on a 70 variables décrivant les 1878 logements depuis le début de leur mise en ligne sur le marché : chaque ligne représente un logement. On a donc pour certains d'entre eux des logements qui ne sont plus actifs voire qui ont été retirés du marché. Certaines variables sont communes aux deux bases de données. La différence entre les deux tableaux provient de la profondeur temporelle: le premier tableau résume les caractéristiques en termes de revenus et de fréquentation mois par mois depuis 2015 tandis que le second les agrège sur l'année passée (LTM = Last Twelve Months, novembre 2019-novembre 2020).

2.2.1. Pré-traitement et nettoyage des données

L'idée de cette étape est de partir des fichiers de données bruts pour arriver à des tableaux nettoyés et prêts à l'emploi. Pour cela, différents traitements ont été réalisés sous R, représentés par la chaîne de traitement ci-dessous (figure 2) :

Figure 2 : Chaîne de traitement : des fichiers bruts aux fichiers consolidés



Réalisation : Guérois, M., école thématique du CNRS, 27/06/2021 - 02/07/2021

Dans un premier temps, il convient de filtrer le tableau en ne conservant que les variables utiles à nos analyses. Il s'agit d'un premier processus de sélection où les variables sans correspondant à l'activité, la localisation et aux profils des logements seront conservées. On supprime alors 13 variables dans le tableau mensuel, et 47 dans le tableau individuel. Par la suite, seuls les logements Airbnb seront conservés : il s'agit d'un processus de filtrage destiné à supprimer les chambres d'hôtel ainsi que les logements HomeAway, eux aussi pris en compte dans les opérations de scraping de la plateforme. À la suite de ces opérations, ce sont 31 chambres d'hôtel et 82 logements HomeAway qui sont retirés de notre tableau de données individuelles, représentant 3048 observations dans notre tableau mensuel.

2.2.2. Traitement des valeurs NA

L'étape suivante réside en le traitement des valeurs NA (Not Available). Il s'agit de variables non attribuées, apparaissant sous cette forme soit par absence de données, soit par erreur dans l'encodage ou l'entrée de données. Il existe plusieurs cas de figure de valeurs NA, et différentes façons de les remplacer spécifiques à ces causes :

- Il s'agit de données manquantes et indépendantes, auquel cas il est difficile de les estimer. Une solution serait de supprimer les lignes correspondant à ces valeurs manquantes, si elles représentent une part assez faible des observations. La première phase de sélection des colonnes

a permis de se débarrasser de beaucoup de valeurs non attribuées car décrivant des indicateurs non utiles à l'analyse.

- Il s'agit de données manquantes liées à une ou plusieurs autres variables. Dans ce cas de figure, il est possible d'estimer cette valeur en fonction d'une autre variable (par exemple estimer le prix en fonction du revenu généré par la location).

Il est nécessaire de bien faire la distinction entre les mécanismes qui induisent la production des données manquantes afin de faire le juste choix de traitement car cela pourrait influencer les résultats des analyses statistiques en y implémentant de possibles biais. En effet, remplacer des valeurs non attribuées par des 0 pourrait avoir un impact plus ou moins considérable selon le nombre d'observations. La détection des valeurs manquantes ainsi que la détermination de la cause de leur présence représente donc un enjeu car la méthode d'imputation (de remplacement) dépendra de cette analyse. Par exemple, dans notre cas de figure, remplacer les valeurs non attribuées correspondant aux revenus d'un logement par des 0 n'aurait pas de sens si le logement n'a jamais été loué. On préférera par conséquent laisser les valeurs non attribuées pour ne pas fausser les analyses statistiques.

Ainsi, le remplacement des valeurs non attribuées se fait en deux étapes : il faut d'abord chercher le mécanisme d'apparition de ces valeurs (si la valeur manquante dépend ou non d'une autre variable ou s'il existe d'autres causes de l'ordre du temporel par exemple), pour ensuite choisir la méthode d'imputation, sans oublier de sélectionner seulement les observations que l'on souhaite transformer.

Certaines variables seront nécessaires à nos analyses et on ne peut pas se permettre d'avoir des valeurs non attribuées dans ces colonnes. Il s'agit d'un choix à effectuer pour chaque variable. Après vérification individuelle, on distingue plusieurs variables pour lesquelles l'apparition de valeurs NA peut induire des biais dans les analyses :

- La variable correspondant à la capacité n'est pas recensée pour certains logements. On décide de supprimer ces observations car elles ne sont pas nombreuses (6 sur 1765) et elles nous empêchent de calculer le prix par personne.
- Au niveau de la variable License, on remplace les observations par des valeurs logical qui prendraient les valeurs TRUE si l'hôte a une licence ou FALSE s'il n'en possède pas. On part du principe qu'une valeur NA dans la colonne License équivaut à un hôte qui n'a pas de licence.
- Le traitement des valeurs NA permet de détecter des anomalies au niveau de données aberrantes : c'est le cas dans nos fichiers de données de certains logements présentant des observations non attribuées dans les colonnes relevant de la fréquentation et des revenus, mais correspondant à un nombre de réservations positif. Il ne s'agit donc pas de valeurs nulles. Après comptage, nous décidons de supprimer ces 20 observations, qui représentent 1,1% de l'effectif total.
- Enfin, le prix n'est pas indiqué pour 106 logements. On décide de l'estimer lorsque c'est possible à l'aide des revenus engendrés (Average Daily Rate), à l'aide d'une régression linéaire ($r^2 = 0,93$). Pour les 30 observations pour lesquelles ni le prix ni les revenus ne sont recensés, nous n'avons pas d'autre choix que de les supprimer.

Enfin, une fois le tableau individuel nettoyé, on supprime les lignes correspondantes dans le tableau mensuel à l'aide de la colonne `id_logement`. Pour la commune d'Issy-les-Moulineaux, cela correspond à 664 observations, soit 1,5% de l'effectif total.

Après les opérations de pré-traitement des tableaux, on se trouve avec 44105 observations dans le tableau mensuel, et 1707 dans le tableau individuel, soit une différence respective de l'ordre de 7,7 et 9,1% du nombre d'observations. Bien que les mêmes logements aient été supprimés, ce chiffre diffère entre les deux tableaux de données car un même logement apparaît un nombre différent de fois dans le tableau mensuel.

2.2.3. Création de variables

Afin d'apporter des réponses aux hypothèses soulevées, et d'enrichir la base de données initiale, nous avons choisi de créer de nouvelles variables nécessaires à la création d'indicateurs. Le tableau récapitulatif des variables présentes dans les tableaux de données nettoyés est disponible en annexe (cf annexe 4).

- Le **prix par personne** : s'obtient en divisant le prix (*Average Daily Price*) par la capacité maximale d'accueil (*Max Guest*)
- La **durée des séjours** : Afin de différencier les séjours courts des séjours longs, nous avons divisé le nombre de nuitées réservées par le nombre de réservations.
- **L'ancienneté** sur le marché Airbnb : En cherchant la date la plus ancienne de récolte des données pour chaque logement, nous pouvons estimer précisément le jour où le logement a été enregistré sur Airbnb
- Le **taux de fréquentation** : il s'agit d'une réestimation de la variable *Occupancy Rate* et consiste en le nombre de nuitées réservées par rapport au nombre de nuitées disponibles pour chaque logement. Il est donc calculé sur les deux tableaux au mois et depuis la mise en ligne du logement à l'aide de :
- La **somme des nuitées disponibles, réservées, bloquées** et du nombre de réservations depuis la mise en ligne d'un logement. Ces variables ont été calculées à partir du tableau mensuel en les additionnant mois par mois.
- En cherchant le nombre d'occurrence de la variable `id_hote`, il a été possible de distinguer les logements appartenant à des multiloueurs, c'est-à-dire les logements pour lesquels cette variable apparaît plusieurs fois, des logements détenus par des monoloueurs. Cela a permis de créer la variable **multi**, une variable booléenne où les observations correspondent à TRUE si le logement est détenu par un multiloueur et FALSE sinon
- Toujours à l'aide de cette variable `id_logement`, la variable **Freq** a été créée, comptant pour chaque logement le nombre de logements totaux détenus par le propriétaire dudit logement.
- **Activité** : À l'aide des variables créées précédemment correspondant au nombre total de nuitées réservées et disponibles depuis la mise en ligne du logement, on distingue 3 types de logements selon leur activité : ceux ayant enregistré au moins une réservation depuis leur mise en ligne,

ceux ayant été disponibles mais jamais réservés, et ceux n'ayant jamais été ouverts à la réservation.

- **Jour / Mois / Année** : afin de faciliter la création de graphiques, j'ai séparé le mois correspondant au *scraping* en 3 variables jour, mois et année à l'aide du package *lubridate*
- Enfin, le tableau de données correspondant aux observations individuelles de chaque logement a été reprojété en format *sf*, afin d'obtenir une couche de points correspondant aux logements Airbnb ainsi que leurs caractéristiques

Enfin, des données externes utiles à la présentation des résultats (cartographiques et statistiques) ont été importées : c'est le cas des contours des communes et IRIS d'Île-de-France, ainsi que du carroyage INSEE de 200m et des stations RATP d'Île-de-France. Les systèmes de coordonnées de ces fichiers ont été harmonisés en WGS84 afin de faciliter les opérations sur les couches géographiques et de permettre la réalisation de cartes *leaflet*, avant d'être filtrées pour correspondre à la commune de référence. Les données de contexte tels que le nombre de logements et de chambres d'hôtel par commune ont aussi été importées. La dernière étape avant l'exportation des tableaux nettoyés et des couches géographiques fut la jointure spatiale entre couches de points et polygones : pour permettre les diverses cartographies, il a fallu faire correspondre les logements avec leurs IRIS et carreau d'appartenance.

2.2.4. Critique des données

On peut émettre quelques critiques par rapport aux bases de données utilisées. D'une part, deux limites se trouvent dans les bases de données issues de AirDNA : les prix indiqués sur le tableau individuel le sont à une période arrêtée (décembre 2020), or on sait qu'ils sont soumis à des fluctuations au fil du temps. Ils ne sont donc pas entièrement représentatifs des prix réels pratiqués, en particulier quand on creuse dans le temps. De plus, la méthodologie du site a changé en juin 2017 avec la prise en compte des logements Aritel-HomeAway, avec un travail de dédoublonnage effectué pour ne pas compter en double une même annonce postée sur les deux plateformes. Cela se ressent dans la notation des données, avec une méthodologie différente avant et après le mois de juin 2017 : si des revenus nuls sont notés comme *Not Available* avant cette période, ils seront remplacés par des 0 par la suite. On décide d'harmoniser en remplaçant toutes les valeurs *NA* par des 0.

Enfin, le premier tableau offrant des résumés mensuels, une réservation chevauchant deux mois (du 30 juin au 2 juillet par exemple) sera comptabilisée à la fois sur le mois de juin et sur celui de juillet. Cela induit un biais dans le calcul de la durée des séjours, mais le nombre de nuitées réservées reste le même.

À la source, au niveau des données Airbnb, la localisation des logements n'est pas précise car les données sont anonymisées : un point représentant une annonce se trouve en réalité dans un rayon de 150m autour de la vraie adresse, ce qui limite la précision géographique des analyses infra-communales.

2.3. Analyses et reproductibilité de la méthode

Une fois les tableaux nettoyés obtenus, une première analyse à effectuer a été d'établir un profil des communes à l'aide d'indicateurs permettant de décrire l'activité des locations saisonnières. L'année de référence choisie a été 2019 car elle correspond à la dernière année « normale » enregistrée. Tout d'abord, le nombre total de logements, ainsi que le nombre de logements actifs (c'est-à-dire ouverts à la réservation) et réservés ont été calculés pour avoir une première idée de l'activité d'Airbnb dans la commune. Par la suite, les indicateurs utilisés ont été les suivants :

2.3.1. Indicateurs utilisés pour comparer les performances d'Airbnb entre les communes

- Le taux d'activité : Il s'agit du nombre de logements réservés par rapport au nombre de logements actifs (c'est-à-dire classifiés comme ouverts à la réservation). Cet indicateur désigne pour chaque commune la probabilité qu'un propriétaire voie son logement loué au moins une nuitée s'il le propose sur la plateforme. Il prend donc en compte d'une part l'activité du propriétaire, qui décide ou non de rendre son logement disponible, mais aussi des visiteurs, qui décident ou non de réserver un logement
- Le prix moyen par personne : Afin de pouvoir comparer le prix entre les différents types de logement qui correspondent à des capacités d'accueil différentes, nous avons préféré le prix par personne par nuitée au prix par nuitée.
- Taux de fréquentation moyen : Il s'agit pour chaque logement du nombre de nuitées réservées par rapport au nombre de nuitées disponibles. Un taux de fréquentation faible n'est pas forcément synonyme de faible arrivée touristique, mais peut être dû à une plus forte disponibilité des logements.
- Le nombre moyen de nuitées réservées / disponibles : Pour avoir une idée de l'activité des locations saisonnières dans une commune.
- Part de logements louées occasionnellement / fréquemment : Cet indicateur sert à distinguer le poids des locations saisonnières « à l'année » dans une commune. Le seuil des 120 nuitées a été choisi pour qualifier un logement loué fréquemment car il s'agit de la limite utilisée pour la plupart des restrictions.
- Part des séjours de moins de 7 jours / plus de 15 jours. Ici, on s'intéresse aux types de séjours : relèvent-ils du court terme, plutôt orientés vers le tourisme, ou bien du long terme, correspondant à d'autres formes de tourisme ou bien à différents profils d'hôtes.
- Part des logements détenus par des multiloueurs : Cet indicateur révèle le poids des multiloueurs (qui ont des profils correspondant à des investisseurs) dans une commune. Il est nécessaire de préciser que ces chiffres ne sont pas exhaustifs car nous n'avons que les données pour une seule commune. Il est probable que certains propriétaires classés comme monoloueurs aient d'autres logements dans d'autres communes.

- Revenus médians engendrés par logements : Enfin, le revenu médian a été choisi au lieu du revenu moyen car on retrouve des valeurs extrêmes notamment via des grandes maisons louées très cher et très fréquemment.

Ces indicateurs ont été calculés pour l'ensemble des logements, parmi les logements entiers et parmi les chambres privées, afin de distinguer des spécificités à ces types de logements.

Par la suite, les analyses ont porté sur la concentration des revenus et des nuitées dans la commune de référence, à l'aide de courbes de fréquences cumulées. Le code est inspiré de la chaîne de traitements²¹ produite par T. Giraud, H. Pecout et R. Ysebaert à l'occasion d'un atelier du Collège International des Sciences du Territoire sur l'accessibilité aux services de santé²². De plus, toujours dans une optique d'analyse de la concentration du marché, les hôtes engendrant le plus de revenus ont été classés, ainsi que leurs bénéficiaires estimés. Leur identifiant n'a pas été dévoilé ici par souci d'anonymisation.

Une seconde partie de l'analyse porte sur l'évolution du marché depuis le début de la récolte de données (qui varie selon les communes, 2015 pour Issy-les-Moulineaux, 2016 pour Pantin et Bagneux). Une première analyse porte sur le nombre total de nuitées réservées et disponibles mois par mois, afin de déceler des périodes correspondant à des pics de fréquentation dans la commune de référence. Pour mettre en exergue une forme de saisonnalité, les taux de fréquentation par mois ont été calculés et les courbes superposées année par année sur un graphique.

Enfin, des analyses infra-communales ont porté sur la densité de logements à l'hectare, ainsi que le prix médian au carroyage INSEE de 200m. Le package *leaflet* a permis d'une part de superposer ces deux indicateurs sur une même carte, mais aussi de proposer une cartographie interactive, où il est possible de voir les chiffres en passant le curseur sur le carreau, de changer le niveau de zoom, etc. Enfin, au niveau des IRIS, le taux de fréquentation ainsi que le nombre de nuitées réservées ont été calculés afin de distinguer les quartiers les plus fréquentés.

Au niveau des analyses comparatives, quelques indicateurs supplémentaires ont été calculés : d'une part, l'évolution du nombre de logements réservés ainsi que celle du taux de fréquentation ont été superposées sur un même graphique pour les deux communes de référence, afin de détecter des potentielles similitudes dans la dynamique de l'activité d'Airbnb dans ces communes. De plus, les indicateurs précédents ont été calculés pour les années 2019 et 2020, afin de quantifier la perte d'activité lors de la période Covid. On a donc comparé le nombre de nuitées disponibles et réservées par mois, ainsi que le revenu médian et le taux de fréquentation mensuel en 2019 et en 2020 pour les deux communes d'entrée.

²¹ <https://rcarto.gitpages.huma-num.fr/santelocal/>

²² <https://cist.cnrs.fr/accessibilite-spatiale-aux-soins-eclairages-theoriques-et-methodologiques/>

2.3.2. Rendre le code générique

Après avoir effectué les analyses pour une commune, à savoir Issy-les-Moulineaux, il a fallu généraliser le code le plus possible pour pouvoir le reproduire à d'autres communes. En partant de ce principe, le postulat de base était de construire un code visant à en modifier la plus petite partie afin de le répliquer à d'autres communes. Pour cela, une technique a été de mettre en entrée le code INSEE ainsi que le nom de la commune de référence, à modifier en début de code, permettant les opérations de sélection et de filtrage afin de distinguer les différents résultats. En somme, les parties Analyse et Interprétation ont été rendues génériques au maximum, et seuls les fichiers d'entrée (tableaux de données et couches géographiques) ainsi que le code INSEE et le nom de la commune sont à modifier en début de code.

Au niveau des interprétations, R propose des balises permettant de moduler le texte selon les résultats obtenus à l'aide d'un morceau de code précédent. Par exemple, ce morceau de code :

```
25
26 ◀ ````{r}
27 library(readr)
28 individuel_clean <- read_csv("pantin_individuel.csv")
29 ◀ ````
30
31 On dénombre `r nrow(individuel_clean)` logements Airbnb à Pantin.
```

permet d'avoir en sortie ceci :

On dénombre 1829 logements Airbnb à Pantin.

Il est donc plus ergonomique d'effectuer des comptages directement sur RMarkdown correspondant à un même morceau de code mais avec des fichiers d'entrée différents.

Des limites ont été cependant rencontrées au niveau de l'interprétation cartographique : il est difficile de rendre les analyses génériques lorsque l'on a affaire à des semis de points ou bien à des cartes de densité. En effet, une interprétation individuelle serait plus judicieuse pour pouvoir localiser des densités élevées par exemple. Un opérateur humain est nécessaire pour analyser des résultats cartographiques.

2.3.3. Diffusion des résultats : site web

Les résultats sont disponibles sur un site internet dédié spécifiquement aux travaux de stages effectués au RIATE *via* la plateforme GitHub à cette adresse : <https://riatestage.github.io/>

airdna_laurian/index.html. Il s'agit d'une page web regroupant tous les fichiers Markdown produits en format html, et structurée en 5 parties, à l'image des travaux précédents effectués par le RIATE :

- **Accueil** : Il s'agit d'une présentation des travaux effectués, du contexte du stage ainsi que de l'encadrement. Un rapide résumé du stage permet de l'ancrer dans le projet de recherche correspondant, en rappelant les objectifs et les hypothèses.
- **Préparation des données** : Il s'agit des pré-traitements effectués pour passer d'un fichier brut (*raw*) à un fichier consolidé (qui sera ensuite importé pour les analyses des parties suivantes). En somme, cette partie décrit les étapes de la chaîne de traitement illustrée ci-dessus. On trouve dans cette partie les données externes à AirDNA, qui seront utilisées à des fins d'analyse cartographiques : contours des communes, des IRIS, grille INSEE de 200m, stations RATP d'Île-de-France, nombre de logements et de chambres d'hôtel par commune.
- **Synthèse à la commune** : Il s'agit des résultats de l'analyse menée sur les données Airbnb pour chacune des 3 commune, structurée en 3 parties : État des lieux sur l'importance, la structure et la performance de l'offre Airbnb dans la commune en 2019, Analyse temporelle 2015-2020, et Analyse infra-communale
- **Comparaison** : Cette partie permet de mettre en vis-à-vis les performances d'Airbnb entre deux communes. L'activité des logements en terme de fréquentation et de revenus engendrés y est étudiée, afin de distinguer différents profils des communes.
- **À propos** : Cette partie porte sur les informations générales des travaux effectués, à savoir les droits d'utilisation et description des données mobilisées (brutes et consolidées), l'accès à ces données, mais aussi les métadonnées ainsi qu'une partie SessionInfo(), regroupant les packages utilisés pour réaliser le code.

Les données ne pouvant pas être diffusables car soumises à une convention établie entre AirDNA et le RIATE, nous avons fait le choix de ne pas les divulguer dans le code. Les résultats ainsi que le code sont cependant disponibles en ligne.

Partie 3 : Résultats

Différents indicateurs ont été retenus pour la lecture des résultats. Parmi eux, certains relèvent de l'activité de la part des hôtes (nombre de nuitées disponibles, prix par personne moyen, part des logements détenus par des multiloueurs), des visiteurs (nombre moyen de nuitées réservées, part des séjours de moins de 7 jours, de plus de 15 jours), ou bien de ces différents acteurs (taux d'activité et de fréquentation, part des logements loués fréquemment / occasionnellement, revenu médian engendré par logement). Pour la suite de nos analyses, la définition de logements actifs et logements réservés nécessite d'être expliquée. On distingue alors :

- les logements actifs : il s'agit de logements classés comme ouverts à la réservation par l'hôte au moins une fois sur l'année;
- les logements réservés : les logements ayant reçu au moins une réservation sur l'année. Il s'agit du stock de référence pour les indicateurs calculés.

L'année 2020 ayant été brusquement ralentie dès le mois de mars, et le scraping s'arrêtant en novembre, ces indicateurs ont été calculés pour une année de référence (en l'occurrence 2019) correspondant à la dernière année « normale » enregistrée. Comme dit précédemment, il existe une limite aux données disponibles : le prix des locations est fixé au mois de décembre 2020, il n'est pas forcément représentatif des prix de 2019. Dans un premier temps, les analyses ont été menées sur la commune d'Issy-les-Moulineaux, avant d'être généralisées pour potentiellement être appliquées à l'ensemble des communes d'Île-de-France.

3.1. Interprétation des résultats : Issy-les-Moulineaux

Durant la première partie du stage, seule la commune d'Issy-les-Moulineaux était étudiée. Le rendu correspond à un format court (3-4 pages), structuré en 3 parties :

3.1.1. État des lieux sur l'importance, la structure et la performance de l'offre Airbnb à Issy-les-Moulineaux (2019)

Les résultats principaux sont les suivants : en 2019, on compte 1153 logements inscrits sur la plateforme Airbnb à Issy-les-Moulineaux, dont 779 proposés au moins une nuitée à la location et 598 comptant au moins une réservation. Le nombre de logements Airbnb représente 3,2% du parc

locatif de la commune²³. En comparaison, le nombre de chambres d'hôtel à Issy-les-Moulineaux s'élève à 848 la même année²⁴.

Le taux d'activité (nombre de logements actifs sur nombre de logements réservés) s'élève à 77%, et le prix par personne et par nuitée est de l'ordre de 35\$, et ne varie que très peu selon le type de logement. Un indicateur intéressant est la part de logements réservés durant plus de 120 nuitées cumulées sur l'année. Il permet de différencier les loueurs fréquents des loueurs occasionnels. Près d'un quart des logements réservés en 2019 l'ont été plus de 4 mois en cumulé sur l'année. Ce chiffre doit être très certainement revu à la baisse aujourd'hui suite : 1. à la pandémie de Covid-19 qui a ralenti l'activité touristique et 2. aux restrictions imposées par l'UNPLV (Union Nationale pour la Promotion de la Location de Vacances) qui limitent le nombre de nuitées réservées par année civile à 120²⁵. Enfin, un quart des logements sont détenus par des multiloueurs. Cette part grimpe à 50% parmi les chambres partagées. Cet indicateur est biaisé car ces chiffres ne sont pas exhaustifs : il est probable que certains propriétaires classés comme monoloueurs aient d'autres logements dans d'autres communes.

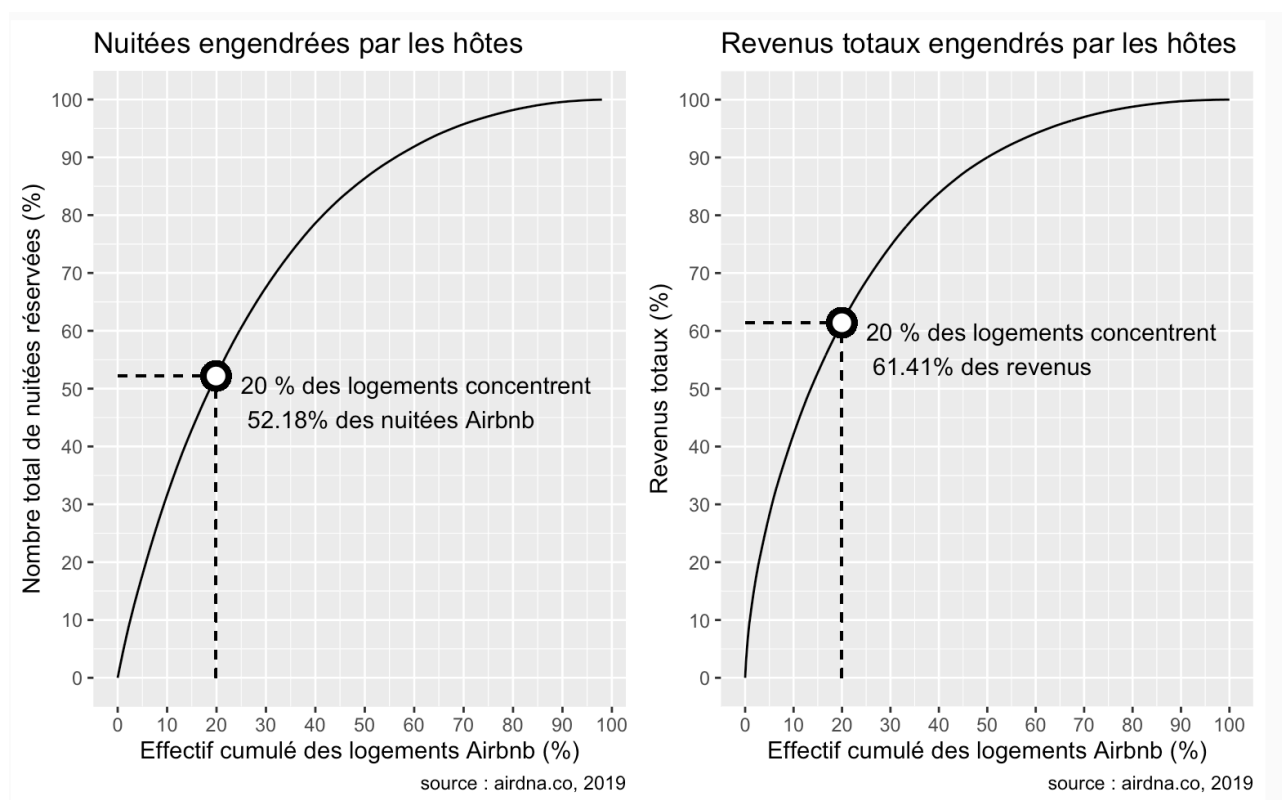
En construisant des courbes de fréquences cumulées, on se rend compte de la forte concentration du marché à Issy : en effet, plus de la moitié des 46482 nuitées réservées en 2019 se partagent entre les 20% de logements les plus fréquemment réservés. Cette concentration s'observe d'autant plus au niveau des revenus : 20% des logements engendrent 61,4% des revenus engrangés à l'aide de la plateforme Airbnb, soit près de 3 millions de dollars en 2019 (figure 3). Lorsque l'on regarde de plus près les propriétaires ayant engrangé le plus de revenus, on observe que 3 des 5 plus fortunés sont monoloueurs. De plus, l'hôte ayant généré le plus de revenus, soit près de 285 000 dollars, représente à lui seul presque 6% des revenus totaux générés sur la plateforme Airbnb à Issy-les-Moulineaux en 2019. En y regardant de plus près, il s'agit d'un propriétaire mettant en location plusieurs grands logements luxueux (4, à un prix minimum de 200\$ par nuitée), et surtout plus de 120 nuitées à l'année.

²³ Source : Nombre de Logements (totaux) des communes d'Île-de-France (Donnée INSEE), 2017. <https://data-iau-idf.opendata.arcgis.com/datasets/nombre-de-logements-totaux-des-communes-dile-de-france-donnee-insee/explore>

²⁴ Source : capacité des communes en hébergements touristiques, INSEE base 2019 <https://www.insee.fr/fr/statistiques/2021703>

²⁵ <https://www.airbnb.fr/help/article/2108/limitation-du-nombre-de-nuitées-en-france%C2%A0-foire-aux-questions>

Figure 3 : Concentration du marché à Issy-les-Moulineaux



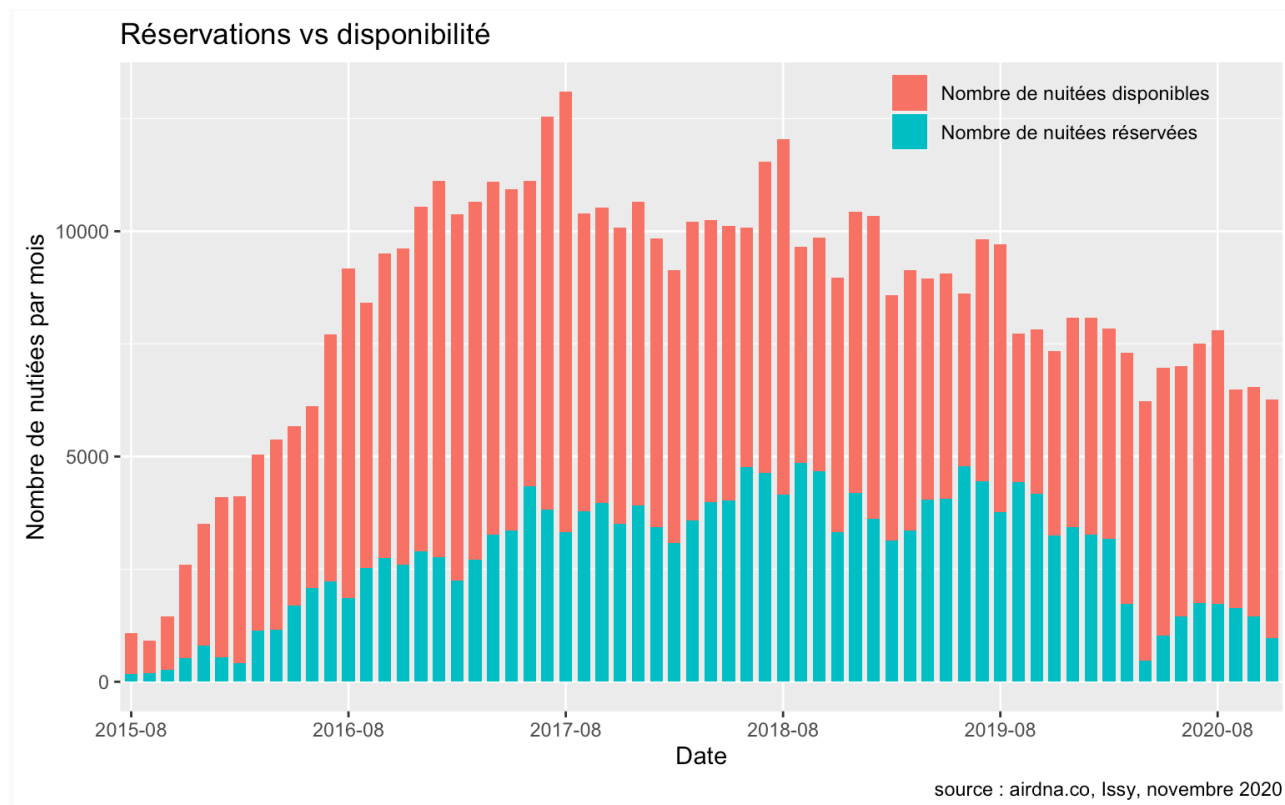
Réalisation : Laurian (2021)

3.1.2. Analyse temporelle 2015-2020

En ce qui concerne l'évolution du marché depuis la première période de récolte des données (2015), on observe une augmentation constante de l'activité jusqu'en 2017, où l'on atteint un pic de nuitées disponibles lors du seul mois d'août. Du côté des visiteurs, c'est un an plus tard en septembre 2018 que le nombre de réservations bat son plein. S'en suivent des chiffres similaires l'année suivante (figure 4). La disponibilité des logements ne cesse de baisser depuis l'été 2017, tandis que le nombre de nuitées réservées reste sensiblement le même, jusqu'à la fin de l'année 2019, où le marché présente des premiers signes de déclin, avant même l'arrivée de la Covid en 2020. En effet, après la période creuse du mois de novembre, le marché a tendance à repartir de l'avant en décembre, à l'occasion des fêtes, comme c'est le cas lors des années précédentes. Cependant, les grèves et manifestations ayant touché la capitale au mois de décembre 2019 ont probablement découragé l'arrivée de visiteurs.

Sur la période 2019-2020, le nombre de locations réservées a subi une chute de 27%, passant de 598 à 438 unités. Seuls 18,6% des logements ont connu une hausse du nombre de leurs visites sur cette période. Enfin, l'arrêt du tourisme se ressent au niveau des revenus : ils sont passés de près de 4,9 à 1,8 millions de dollars entre 2019 et 2020.

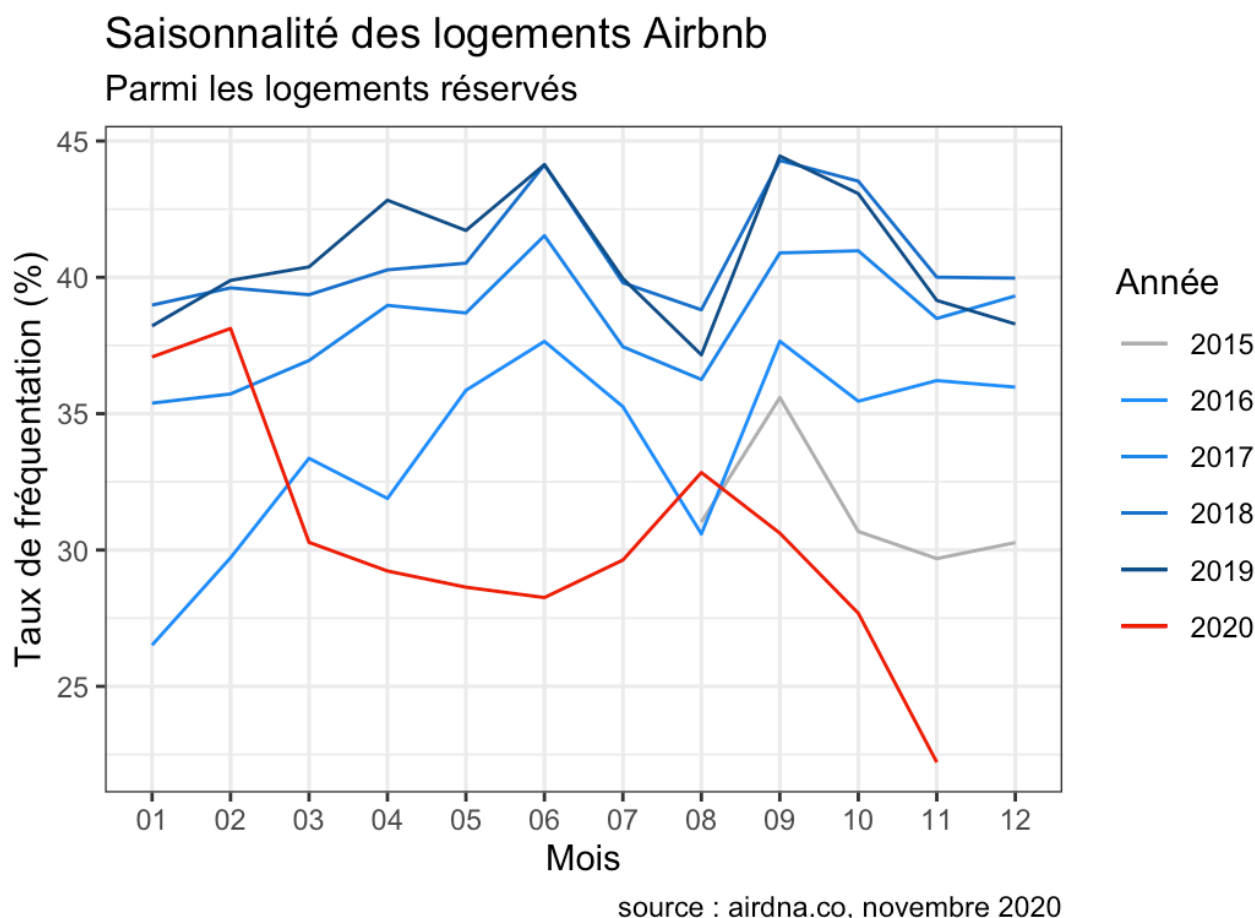
Figure 4 : Un déclin amorcé chez les hôtes dès 2017, chez les visiteurs fin 2019



Réalisation : Laurian (2021)

Lorsque l'on s'intéresse au taux de fréquentation (nombre de nuitées réservées par rapport au nombre de nuitées disponibles), on observe une certaine saisonnalité : les mois de juin et septembre sont ceux où le taux de fréquentation est le plus élevé, tandis que les mois d'août et de novembre correspondent à des saisons creuses (figure 5). Pour le mois d'août, une interprétation serait la suivante : beaucoup de propriétaires laissent leur logement disponible au moment de partir en vacances; ainsi le nombre de nuitées réservées étant constant, le taux de fréquentation diminue. Enfin, on constate une chute moyenne de 47% du taux de fréquentation par logement sur l'année de 2019 à 2020. On observe sur la figure suivante que la chute du taux de fréquentation par rapport à l'année précédente est entamée dès le mois d'octobre 2019. S'agit-il des premiers signes de saturation du marché à Issy-les-Moulineaux ?

Figure 5 : Une forme de saisonnalité à Issy-les-Moulineaux

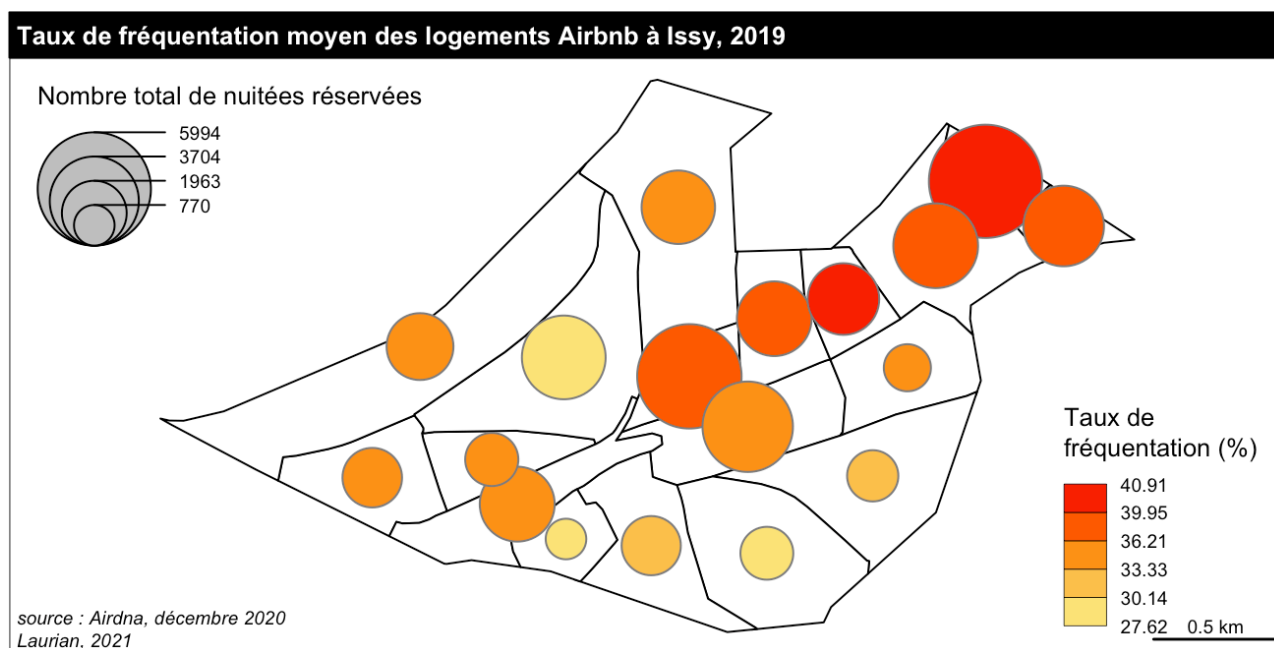


Réalisation : Laurian, 2021

3.1.3. Analyse infra-communale

Enfin, une analyse infra-communale des prix par personne au carroyage INSEE de 200m fait apparaître des quartiers très fournis en offre Airbnb près de la capitale avec des densités atteignant 8 logements réservés par hectare, et dans une moindre mesure au sein du centre-ville et le long de la ligne 12 du métro parisien (cf annexe 5). Cette distribution spatiale s'observe dans la carte ci-dessous (carte 2), où les IRIS jouxtant Paris sont ceux avec le plus de nuitées réservées, mais aussi ceux où le taux de fréquentation est le plus élevé. On retrouve des taux de fréquentation similaires en centre-ville. Le sud de la ville correspond à une activité plus faible de la plateforme Airbnb, en termes de nombre de logements, nombre de nuitées réservées et taux de fréquentation. Une accessibilité moindre à la capitale pourrait en être la cause. La proximité à Paris semble donc augmenter l'offre en terme de logements, ainsi que la demande de la part de visiteurs.

Carte 2 : Une concentration en centre-ville et à proximité de la capitale



3.2. Comparaisons entre communes : Issy-les-Moulineaux et Pantin

Dans un second temps, un fichier a été produit pour comparer l'activité de la plateforme Airbnb dans deux communes. Ici, ce sont les indicateurs pour Issy-les-Moulineaux et Pantin qui ont été sélectionnés, mais nous avons aussi les résultats pour Bagneux, où le locatif à court terme tient une place plus faible dans le marché immobilier que dans les communes limitrophes.

3.2.1. Un marché plus actif à Pantin, plus lucratif à Issy-les-Moulineaux

Les indicateurs vont dans le sens d'une plus forte présence à Pantin qu'à Issy-les-Moulineaux (figure 6). En effet, avec 650 logements réservés pour 898 logements actifs en 2019, on dénombre pas moins de 1238 logements enregistrés sur le site Airbnb dans la commune. Parmi les logements réservés, on note une part plus importante de chambres privées à Pantin qu'à Issy. De plus, Airbnb y représente 4,7% du parc de logements de la commune. On note une sensible différence au niveau du prix par personne : Pantin est 40% moins chère que Issy (25\$ par personne et par nuitée en moyenne contre 35\$ à Issy). Cet écart s'accroît entre les chambres privées des deux communes. Cette différence de prix peut s'expliquer dans la diversité des profils socio-économiques et du marché du logement des deux communes. En effet, Pantin est une ville à profil socio-résidentiel très mixte, en pleine mutation vers un processus de moyennisation qui n'exclut pas les catégories les

plus modestes, tandis qu'Issy-les-Moulineaux est composée d'un « Marché de catégories supérieures dans les espaces privilégiés » (Le Corre, 2019), avec une sur-représentation des cadres et professions intellectuelles supérieures qui représentent la majorité des acheteurs.

Au niveau de l'activité des logements, les deux communes présentent des profils assez similaires, en terme de taux de réservation ou bien de nuitées réservées. À Issy cependant, 15% des chambres privées sont louées occasionnellement, contre 9% à Pantin. On observe ici des premiers signes structurants qui permettent de différencier le profil des deux communes : les chambres privées prennent une place plus importante dans l'activité de la plateforme Airbnb à Pantin qu'à Issy-les-Moulineaux, en captant une plus grande partie des nuitées réservées (91 à Pantin en 2019 contre 73 à Issy-les-Moulineaux). Cette divergence dans les types de logement pourrait être due à des formes différentes de tourisme ou de clientèle au sein de ces deux communes. Si les chambres privées sont plutôt recherchées par une clientèle de passage, ou moins aisée (jeunes), les logements entiers permettent une expérience différente de la ville, correspondant à une expérience de tourisme plus consommatrice et adaptées aux visiteurs plus aisés.

Enfin, une différence notable se remarque au niveau du revenu médian engendré. Les prix pratiqués plus faibles à Pantin expliquent ces écarts, qui se creusent entre logements entiers (35% plus lucratifs à Issy). Cependant, la tendance s'inverse parmi les chambres privées : elles engrangent en moyenne 15% de revenus de plus à Pantin qu'à Issy (respectivement 2867\$ contre 2052\$). Ces chiffres corroborent les résultats précédents, indiquant une activité plus forte du côté des chambres privées à Pantin. Une dernière différence s'observe au niveau des courbes de concentration des revenus dans les communes : 20% des logements Airbnb engendrent 60.3% des revenus totaux à Issy contre 55.6% à Pantin, tandis que la concentration des nuitées est relativement similaire (20% des logements concentrent environ 52% des nuitées dans les deux communes).

Figure 6 : Comparaison des indicateurs entre Issy-les-Moulineaux et Pantin

Indicateur (2019)	Tous les logements		Logements entiers		Chambres privées	
	Issy	Pantin	Issy	Pantin	Issy	Pantin
Taux d'activité (%)	74	72	75	73	68	72
Prix par personne moyen	35	25	34	25	37	25
Taux de fréquentation moyen (%)	55	57	58	61	44	50
Nombre moyen de nuitées disponibles	146	146	143	133	164	182
Nombre moyen de nuitées réservées	80	83	83	81	73	91
Part de logements loués fréquemment (>= 120 nuitées)	25	24	26	24	26	27
Part de logements loués occasionnellement (<= 7 nuitées)	10	9	9	9	15	9
Part des séjours de moins de 7 jours	79	75	79	74	84	78
Part des séjours de 15 jours ou plus	4	4	3	5	3	1
Part de logements détenus par des multiloueurs	27	30	21	24	52	47
Revenu médian engendré par logement (\$)	4393	3714	5474	4050	2502	2867

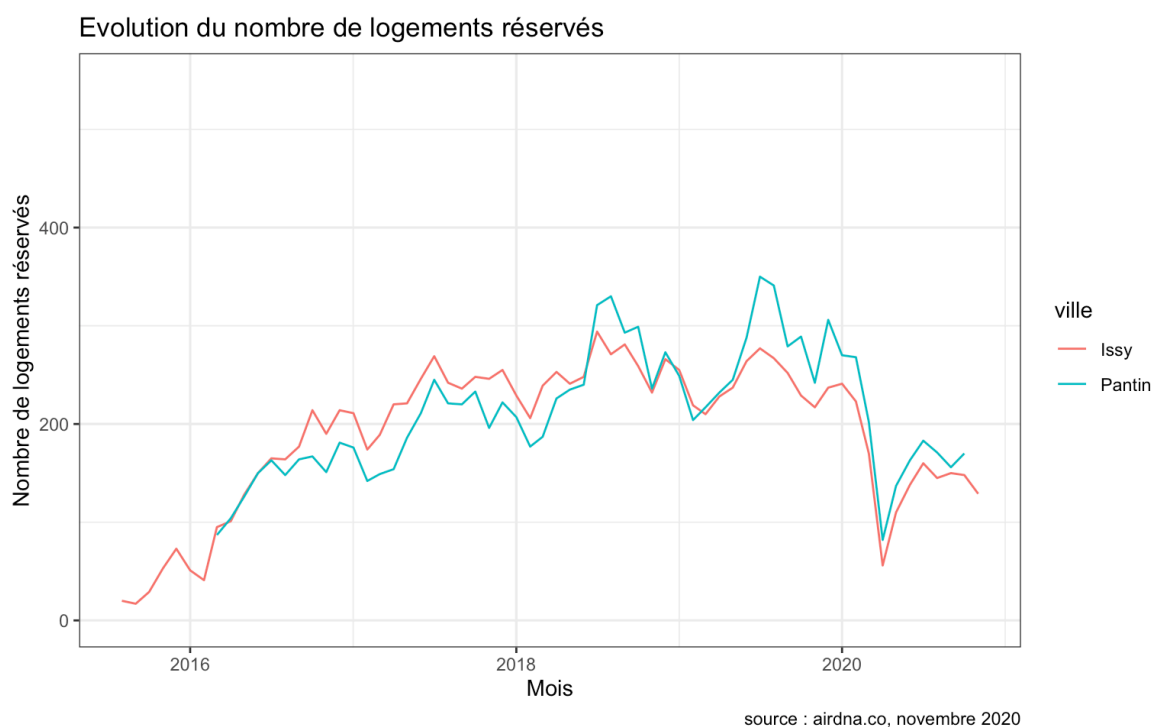
Réalisation : Laurian (2021)

3.2.2. Un développement plus récent à Pantin

En regardant l'évolution du marché dans les deux communes, on observe que le pic d'activité (nombre de logements réservés) est atteint à Pantin en juillet 2019 avec 350 logements réservés. À Issy, le pic apparaît plus tôt en juillet 2018 avec 294 logements réservés au cours de ce mois. Ce chiffre est difficilement comparable avec les quelque 70 000 annonces enregistrées à Paris au troisième trimestre 2017 (chiffres AirDNA). C'est durant ce même mois de juillet 2018 que le nombre de logements réservés à Pantin va dépasser celui d'Issy pour la première fois et le restera jusqu'au mois de mars 2020 où les premières restrictions de déplacement ont été mises en place (figure 7). Globalement, les trajectoires sont les mêmes entre les communes, avec des chiffres plus élevés à Pantin. Les deux communes suivent une même saisonnalité, à savoir des pics de réservation en juin-juillet et novembre, contre des saisons creuses en février et en octobre.

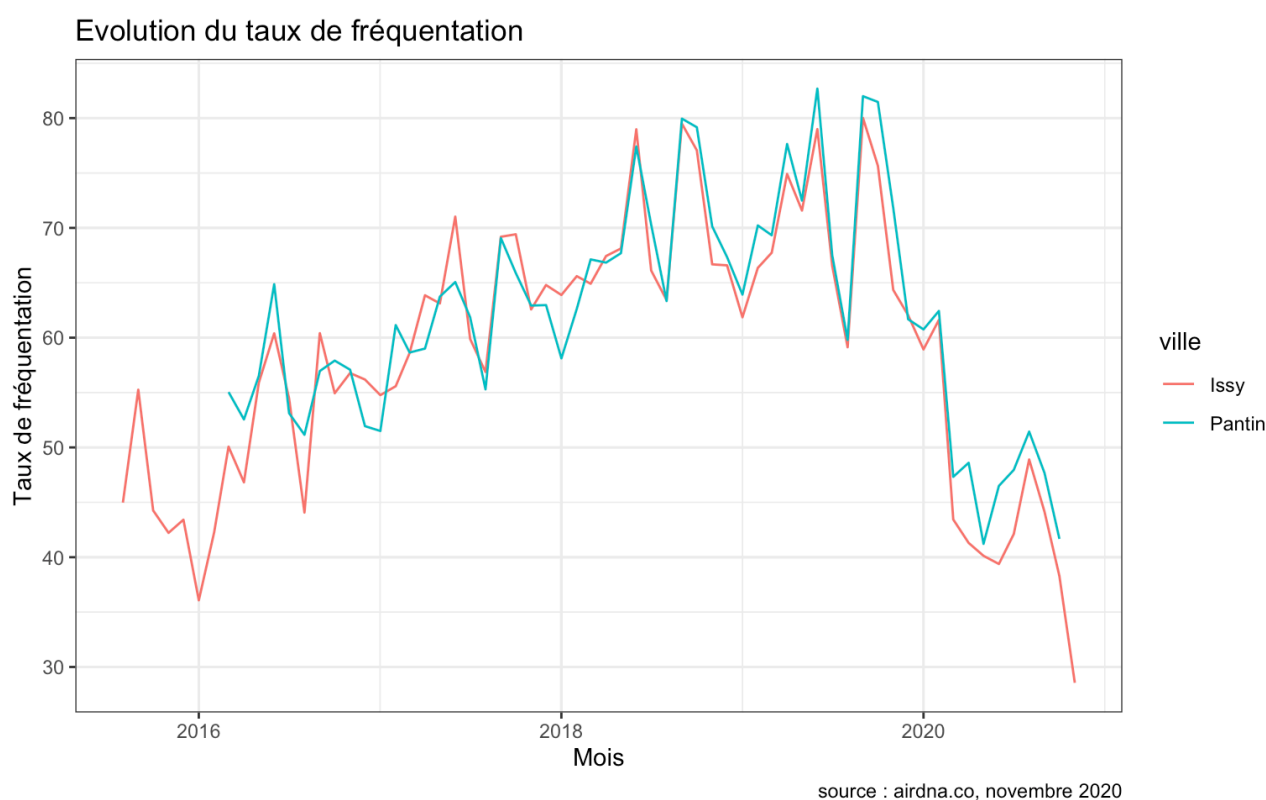
Nous émettons l'hypothèse que l'année 2018 a été un tournant dans la fréquentation touristique - et notamment en ce qui concerne les locations saisonnières - de Pantin et d'Issy-les-Moulineaux, dont la période correspond à l'apparition des premières régulations liées à Airbnb (Pantin n'est pas touchée par la régulation des 120 nuitées). Les trajectoires décisionnelles de ces deux communes (régulations à Issy, opérations d'aménagement axées sur le logement à Pantin cf partie 1) peuvent y expliquer les dynamiques de l'activité d'Airbnb.

Figure 7 : Un marché en proie aux restrictions (Issy-les-Moulineaux) contre un en expansion (Pantin)



Au niveau du taux de fréquentation, les similitudes entre les deux communes sont plus évidentes : elles suivent quasiment le même tracé au chiffre près à partir de 2018 (figure 8). Deux acteurs sont à prendre en compte lorsque l'on analyse le taux de fréquentation : les hôtes, qui décident de rendre leur logement plus ou moins disponible, et les visiteurs, qui contribuent au nombre de nuitées réservées. Lorsque l'on s'intéresse à la période Covid plus en détail, on observe que cette baisse du taux de fréquentation n'est pas due aux hôtes : en effet, le nombre de nuitées disponibles stagne entre 2019 et 2020, voire augmente à Issy-les-Moulineaux (+4,5%). C'est au niveau des nuitées réservées que le marché s'effondre : -28,3% à Pantin, -30,5% à Issy, où les hôtes ont subi de plus grosses pertes financières avec un revenu médian diminué de 41,3%, contre 35% à Pantin.

Figure 8 : Un taux de fréquentation quasiment identique entre les deux communes

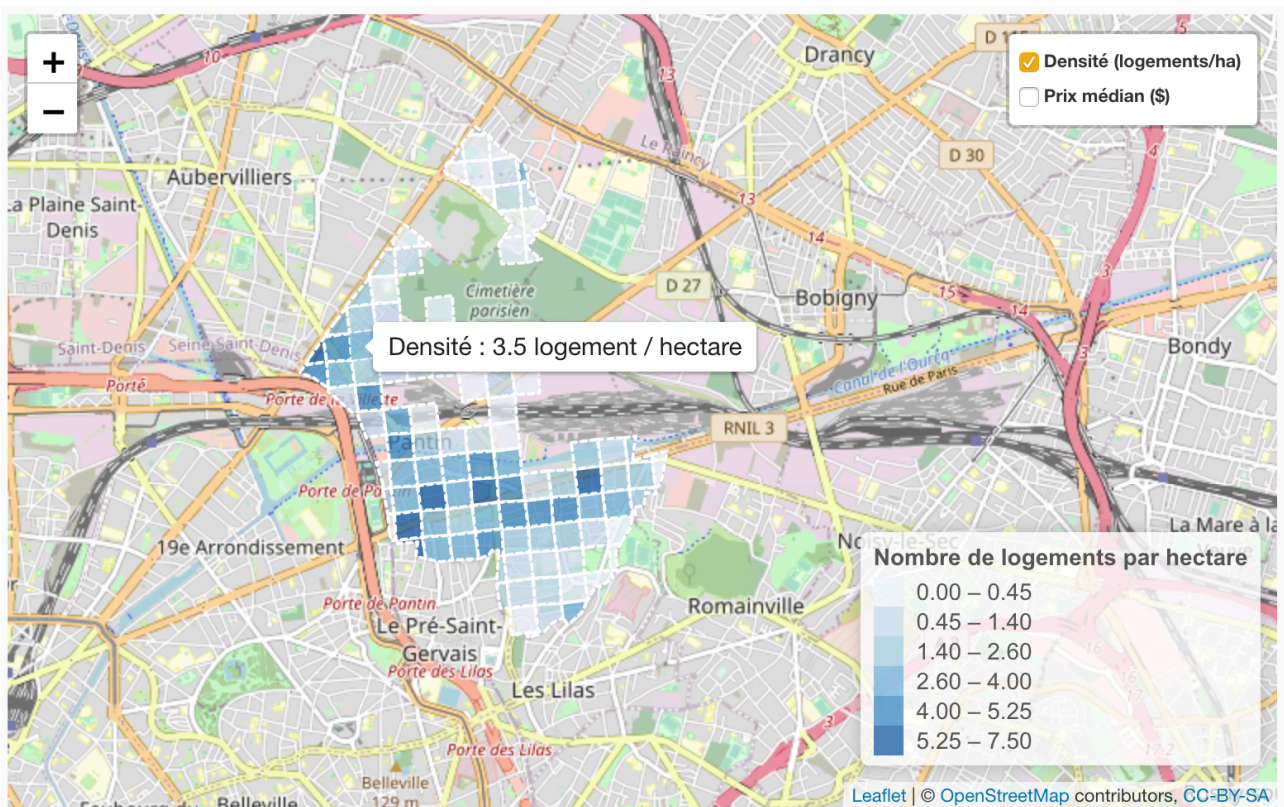


Réalisation : Laurian (2021)

3.2.3. Mais une implantation spatiale qui reste la même

Dans les deux communes, un schéma similaire apparaît : comme on pouvait s'y attendre, les densités des logements Airbnb sont beaucoup plus fortes près de la capitale et autour du centre-ville. (carte 3) À Pantin, on observe cependant deux zones polarisantes des logements Airbnb : une première le long du périphérique, et une autre à la frontière avec Aubervilliers, au niveau de la station de métro Quatre Chemins de la ligne 7. Le tracé du canal de l'Ourcq, qui se confond avec la ligne 5 du métro parisien, semble lui aussi concentrer des fortes densités de logement, ce qui est moins le cas au niveau du Bras de la Seine à Issy-les-Moulineaux.

Carte 3 : Concentration du marché à Pantin



Réalisation : Laurian (2021)

Au niveau des taux de fréquentation, le gradient centre-périphérie est plus évident à Pantin qu'à Issy, où les logements situés près des stations de métro, qui ne sont pas forcément les plus réservés, enregistrent des taux de fréquentation record. Si à Issy les logements les plus réservés sont aussi ceux où le taux de fréquentation est le plus élevé, ce n'est pas forcément le cas à Pantin.

Une limite se pose aux méthodes et outils employés : il est difficile de comparer des analyses cartographiques entre deux communes de manière générique. En effet, des contextes locaux sont propres à chaque commune, et peuvent expliquer individuellement la localisation ou la concentration de logements à certains endroits (par exemple le cimetière de Pantin explique les faibles densités, ou bien l'écoquartier des Épinettes à Issy-les-Moulineaux peut expliquer des prix plus élevés). Cela nécessite donc un opérateur humain, qui constitue une limite à la reproductibilité.

3.3. Bagneux : Des hôtes aux profils divers dans une commune plus éloignée

Le phénomène Airbnb est beaucoup moins marqué à Bagneux que dans les deux autres communes étudiées, avec seulement 1,9% du marché du logement occupé. L'hypothèse principale est qu'il s'agit d'un gradient centre-périphérie naissant, où le marché parisien se diffuse, mais peine à

dépasser les communes limitrophes. Cela s'observe au niveau du nombre d'annonces sur la plateforme : seulement 335 pour l'année de référence 2019. Le taux d'activité cependant reste le même que pour Issy et Pantin, à savoir 77% : sur les 232 logements actifs, 179 ont été réservés durant l'année. La commune de Bagneux ne compte qu'un seul hôtel, pour 47 chambres.

On trouve des prix légèrement plus élevés qu'à Pantin, de l'ordre de 27\$ par personne par nuitée. L'activité est globalement plus faible que dans les deux autres communes étudiées, avec un taux de fréquentation inférieur à 50%, et environ 70 nuitées réservées sur l'année en moyenne par logement. Comme pour Pantin, les chambres privées concentrent la majorité des nuitées réservées (29% de plus que les logements entiers), mais il s'agit surtout du type de logement qui concentre les locations saisonnières « à l'année » (> 120 nuitées réservées) : si elles représentent un quart des locations à Issy et à Pantin peu importe le type de logement, ce chiffre monte à 30% parmi les chambres privées de Bagneux, soit deux fois plus qu'entre les logements entiers. Au niveau de la durée des séjours, 12% des séjours s'effectuent sur une durée supérieure ou égale à 15 jours, soit 3 fois plus qu'à Issy et Pantin. Enfin, plus d'un tiers des logements appartiennent à des multiloueurs, dont 60% des chambres privées. Les multiloueurs représentent 17% des propriétaires de la commune, contre respectivement 12 et 13% pour Issy et Pantin. Si les revenus sont inférieurs aux deux autres communes, ils ne diffèrent que très peu entre chambres privées et logements entiers.

Au niveau de la concentration du marché, on note des chiffres similaires à ceux des autres communes : 20% des logements concentrent 56% des nuitées, et 59% des revenus. À des niveaux moindres, l'hôte le plus rentable a engendré 59 000\$, soit 6,6% de la totalité des revenus engendrés par le marché Airbnb à Bagneux en 2019. On remarque que sur les 5 hôtes engrangeant le plus de revenus, 4 sont monoloueurs. Bagneux pourrait donc se différencier par le profil de ses propriétaires : on y trouve plus de multiloueurs qu'à Issy et Pantin, mais il ne s'agit pas des plus gros hôtes en terme de revenus.

Lorsque l'on s'intéresse à la dynamique de l'activité d'Airbnb dans la commune, on observe que le nombre de logements réservés stagne depuis la mi-2017, où on compte 71 logements réservés. Les variations sont ensuite très faibles, et le pic est atteint en juillet 2019 avec 82 logements réservés. Il est plus difficile de détecter une saisonnalité dans les taux de fréquentation recensés. Les mois de juin et septembre apparaissent comme ceux étant les plus actifs, mais le comportement du marché diffère d'année en année. Le taux de fréquentation en août 2020 est plus élevé qu'à la même période lors des années précédentes. Si l'activité des logements semble suivre un parcours en dent de scie après la dégringolade de mars 2020 à Issy et Pantin, il se maintient à un niveau constant à Bagneux dès le mois de mai. L'effet de la crise sanitaire a été moins violent à Bagneux mais reste notable : -25,3% de nuitées réservées par mois, -29,6% des revenus mensuels médians. La situation s'inverse : un hôte balnéolais engrange plus de revenus que son homologue pantinois en 2020 (673\$ contre 637\$, contre respectivement 956 et 977\$ en 2019).

Enfin, au niveau infra-communal, il est difficile de distinguer un gradient centre-périphérie depuis la capitale. Les densités sont très faibles (3,7 logements par hectare au maximum, sur carroyage non complet), mais semblent être parmi les plus élevées autour de la station de métro Châtillon-

Montrouge de la ligne 4 du métro parisien, ainsi qu'autour du fort de Montrouge. Encore une fois, c'est l'accessibilité à la capitale qui semble être facteur de densités plus élevées. Autour de la mairie, les densités oscillent entre 0,5 et 1,5 logement par hectare. Quant au prix, trop peu de données sont disponibles, mais la station de RER Arcueil-Cachan, pôle universitaire, semble faire hausser les tarifs. Si le nord de la commune ainsi que le centre-ville semblent concentrer le plus de nuitées réservées, c'est bien au sud que l'on observe des taux de fréquentation plus élevés. De nombreuses nuitées réservées ne riment donc pas forcément avec un fort taux de fréquentation.

Peut-on classer ces communes selon l'activité qui s'y produit ? Pantin se distingue par de gros propriétaires proposant jusqu'à 12 logements, tandis que les portefeuilles des plus gros multiloueurs à Bagneux et à Issy montent respectivement à 4 et 5 logements. Mais ces multiloueurs ont des profils différents selon les communes : à Bagneux, 60% des logements loués par des multiloueurs sont des chambres privées ou partagées. Ce chiffre descend à 38% à Pantin et 35% à Issy. De plus, l'arrivée précoce des restrictions dans la commune d'Issy-les-Moulineaux entraîne une stabilisation de l'activité d'Airbnb dans la commune, à l'image de la capitale quelques années plus tôt. On peut émettre l'hypothèse que le marché présente des premiers signes de saturation dans cette commune, tandis que le phénomène monte à Pantin, encouragé par les opérations de rénovations immobilières. On remarque que le marché redémarre mieux à Pantin qu'à Issy à la suite des restrictions de déplacement. Enfin, Bagneux semble trop éloignée pour capter le marché parisien. On émet l'hypothèse qu'il puisse s'agir d'un marché différent, plus local.

Conclusion

Ce travail de recherche a permis d'explorer l'activité de la location saisonnières dans des territoires encore peu documentés, les communes situées en périphérie des grandes métropoles. Les résultats principaux sont les suivants : dans un marché saturé au niveau de la ville-centre, le phénomène se diffuse d'abord vers les communes limitrophes, puis dans une moindre mesure dans les aires périurbaines, avec une certaine latence. La dynamique n'est cependant pas linéaire dans le temps, des restrictions apparaissant à la fin des années 2010 dans les communes ayant connu une explosion précoce du marché de la location saisonnière, telles qu'Issy-les-Moulineaux. La période 2017-2019 représente (pour l'instant) l'apogée de ces plateformes, avant les effets des premières régulations et surtout l'écroulement du marché début 2020. Les premiers résultats tendent donc à montrer que l'implantation d'Airbnb est plus forte au sein des communes limitrophes, mais la densité des logements ainsi que leur dynamique restent spécifique à des contextes locaux, notamment en ce qui concerne les politiques communales de logement et de tourisme. Au sein des communes, le marché est plutôt concentré et une poignée de propriétaires se distinguent notamment par une plus forte activité en ce qui concerne les nuitées et revenus engendrés, qui se chiffrent en dizaines voire centaines de milliers de dollars par an.

Les données à notre disposition nous permettent de creuser de nouvelles pistes quant à la fréquentation ou bien les revenus engendrés de chaque logement, et plus précisément de chaque hôte. Il est donc possible *via* ces bases de données de distinguer les locations saisonnières « à l'année » de celles plus occasionnelles, ainsi qu'avoir une estimation des revenus engendrés par Airbnb au sein d'une commune. Le plus surprenant est la faible proportion d'hôtes enregistrés : seulement 5%, tandis qu'on compte plus de 10% de multiloueurs à Issy. Il est donc relativement simple à l'aide de ces données de détecter les hôtes ne respectant pas la réglementation. Du point de vue des pouvoirs publics (type mairies), le point fort de ces données se trouve dans les détails de chaque annonce, permettant de retrouver les potentiels propriétaires fraudeurs. Pour les hôtes, AirDNA offre plutôt une analyse du marché local afin de les aider à maximiser leurs profits. Enfin, du côté de la recherche, AirDNA offre une source de données supplémentaires, complétant les chiffres mis à disposition par d'autres plateformes telles que insideairbnb ou bien Airbnb elle-même, notamment en dédoublonnant les annonces (permet de détecter si une même annonce est présente sur deux plateformes différentes). Un des points forts de AirDNA par rapport aux autres bases de données se trouve dans la plus grande couverture géographique de ses données. Il ne faut pas oublier cependant que ces données sont des estimations générées par un algorithme et ne sont donc pas parfaitement exactes. De plus, leurs méthodes de calcul sont incompatibles avec les standards de l'industrie (Agarwal et al., 2019), mais restent néanmoins utilisables. C'est pourquoi une phase de pré-traitement des données est nécessaire afin de minimiser les possibles biais inhérents à cette base de données. Une solution pour communiquer les résultats sans diffuser les données de la base AirDNA serait de constituer une base de données avec des chiffres aléatoires, comme ce qui a été réalisé par Thibault Le Corre pour le compte de l'ANR WISDHOM avec la base BIEN. Pour des raisons de confidentialité, les données de la Chambre des Notaires n'ont pas pu être diffusées, et des échantillons tests ont été générés afin de permettre à l'utilisateur de prendre

connaissance des données sources, et dans un souci de reproductibilité du travail (base de données CASSMIR²⁶). Cela nécessiterait de plus amples travaux qui dépassent le cadre de ce stage.

Les méthodes et outils mis en oeuvre permettent un haut degré de reproductibilité de ce travail. Elles permettent en effet une forte vérification des résultats par le lecteur, mais aussi pour de recherches futures, un socle de méthodes réutilisables pour par exemple réaliser une étude similaire en comparant avec d'autres bases de données, ou bien modifier l'année ainsi que la commune de référence. Une limite à la méthode employée est l'accessibilité aux données, qui restent payantes car issues d'une entreprise. Ce travail s'inscrit dans la démarche scientifique du RIATE, qui exige une forte transparence et documentation de la méthode utilisée. Ainsi, le code réalisé et les fichiers de sortie se trouvent en libre accès sur la plateforme GitHub riateStage à cette adresse : https://riatestage.github.io/airdna_laurian/about.html. Cette disponibilité des sources et des méthodes permet à tous de reproduire ou d'approfondir ces analyses, via la plateforme GitHub.

Une question que nous pouvons nous poser (et qui pourra être répondue à l'aide de la chaîne de traitement réalisée) est la manière dont le marché réagira lors de l'après Covid. D'un côté, la flexibilité du fonctionnement de la plateforme permettrait une rapide reprise de l'activité Airbnb. En même temps, et en particulier dans les villes les plus marquées par le phénomène Airbnb telles que Paris ou Bordeaux, il est de plus en plus difficile d'obtenir une licence auprès de la commune. Quelles en seront les conséquences sur l'évolution du marché au sein des communes périphériques ? Quels types de logements sauront le mieux résister à cette pandémie ? Si le tourisme infra-national est celui qui a le mieux résisté à la crise, à quel type de territoire correspond-il ? Les résultats du rapport de l'IPR indiquent que l'offre de location saisonnière en espace rural répond à une demande nationale, voire régionale. Avec la démocratisation du télétravail, peut-on s'attendre à une recrudescence du marché de la location à court terme en espace peu denses, au détriment des grandes villes ? Nous pouvons émettre l'hypothèse que dans un futur plus ou moins proche, les espaces urbains correspondront à une clientèle internationale, tandis que les territoires moins denses répondront à une demande de la part d'une clientèle de proximité.

Deuxièmement, et plus en relation avec le projet Wisdrom, il serait possible de retourner le problème : la quasi-absence de tourisme en période de pandémie a-t-elle permis d'inverser la tendance en ce qui concerne la disponibilité de ces logements sur le marché ? En d'autres termes, assiste-t-on à une re-reconversion du logement du locatif à court terme vers un retour au marché classique ? Nous n'avons pas encore assez de recul mais cette période Covid pourrait correspondre à un freinage de la crise du logement, les ménages français ayant réalisé des économies durant plus d'un an et en parallèle les propriétaires remettant leur logements sur le marché de la location classique, voire le revendent. À Paris, l'offre locative a augmenté de 64% entre le 01/06/2018 et le 30/07/2019, la location rentable étant moins rentable mais aussi moins risquée que le court terme²⁷. On constate des dynamiques similaires à Barcelone et à Lisbonne. Les prix des loyers quant à eux restent stable sur cette période. Aucune étude n'a encore été réalisée pour prouver le lien entre retrait des logements du marché à court terme et stagnation voire baisse des loyers dans certaines villes.

²⁶ <https://tlecorre.gitpages.huma-num.fr/cassmir/>

²⁷ source : <https://edito.seloger.com/actualites/france/crise-coronavirus-a-t-booste-marche-de-location-article-39175.html>

Bibliographie

- Adamiak, C., Szyda, B., Dubrownik, A., Garcia-Alvarez, D., 2019, « Airbnb offer in Spain - spatial analysis of the pattern and determinants of its distribution », *ISPRS international Journal of Geo-Information* 8 (3), 155. URL : <https://doi.org/10.3390/ijgi8030155>
- Allen, J., Barlow, L., Leal, J., Maloutas, T., Padovani, L., 2004, *Housing and Welfare in Southern Europe*, Oxford, Blackwell Publishing, 240 p.
- Andreau, S., Basini, B., 2020, « Airbnb se risque en bourse », *Journal du Dimanche*, n.3856, p. 26
- APUR (Atelier Parisien d'Urbanisme) :
 - Paris Pantin. Autour du tramway, enjeux et projets, juillet 2008. URL : <https://www.apur.org/fr/nos-travaux/paris-pantin-autour-tramway-enjeux-projets>
 - Note n. 52, novembre 2011, « Les locations meublées de courte durée à Paris ». URL : <https://www.apur.org/fr/nos-travaux/locations-meublees-courte-duree-paris-lieux-propositions>
 - Note n. 128, juin 2018, « Locations meublées de courte durée : Quelles réponses publiques ? » URL : <https://www.apur.org/sites/default/files/documents/APBROAPU554.pdf>
- Aulnay, V., Piganiol, V., 2021, « La France d'Airbnb », Carte à la une de Géoconfluences. URL : <http://geoconfluences.ens-lyon.fr/informations-scientifiques/a-la-une/carte-a-la-une/france-airbnb>
- Barron, K., Kung, E., Proserpio, D., 2018, The sharing economy and housing affordability: Evidence from Airbnb. SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3006832
- Boros, L., Dudás, G., Kovalcsik, T., Pap, S., Vida, G., 2018, « Airbnb in Budapest : Analysing spatial patterns and room rates of hotels and peer-to-peer accommodations », *in Geojournal of Tourism and Geosites*, 21(1), pp.26-38
- Cocola-Gant, A., Gago, A., 2019. Airbnb, buy-to-let investment and tourism-driven displacement: A case study in Lisbon. *Environment and Planning A: Economy and Space*. doi: [10.1177/0308518X19869012](https://doi.org/10.1177/0308518X19869012)
- Dolnicar, S., 2017, *Peer-to-Peer Accommodation Networks : Pushing the boundaries*, Oxford : Goodfellow Publishers, 296 p.
- Domènech, A., Larpin, B., Scaglione, M., Schegg, R., 2019, « Disentangling the geographical logic of Airbnb in Switzerland », *Erdkunde*, Vol. 73-4, pp.245-258, DOI : 10.3112/erdkunde.2019.04.01
- Ebelle, K., 2018, "AirDNA.co : le site qui vous dit combien louer votre Airbnb", Génération Voyage, décembre 2018. URL : <https://generationvoyage.fr/AirDNA.co-analyse-Airbnb/>
- Gutiérrez, J., García-Palomares, J.C., Romanillos, G., Salas-Olmedo, M.H., 2017, « The Eruption of Airbnb in Tourist Cities : Comparing Spatial Patterns of Hotels and Peer-to-Peer Accommodation in Barcelona », *Tourism Management*, 62, pp. 278–291.
- INSEE :
 - Dossier complet - Issy-les-Moulineaux (2018). URL : <https://www.insee.fr/fr/statistiques/>

[2011101?geo=COM-92040](https://www.insee.fr/fr/statistiques/5058675)

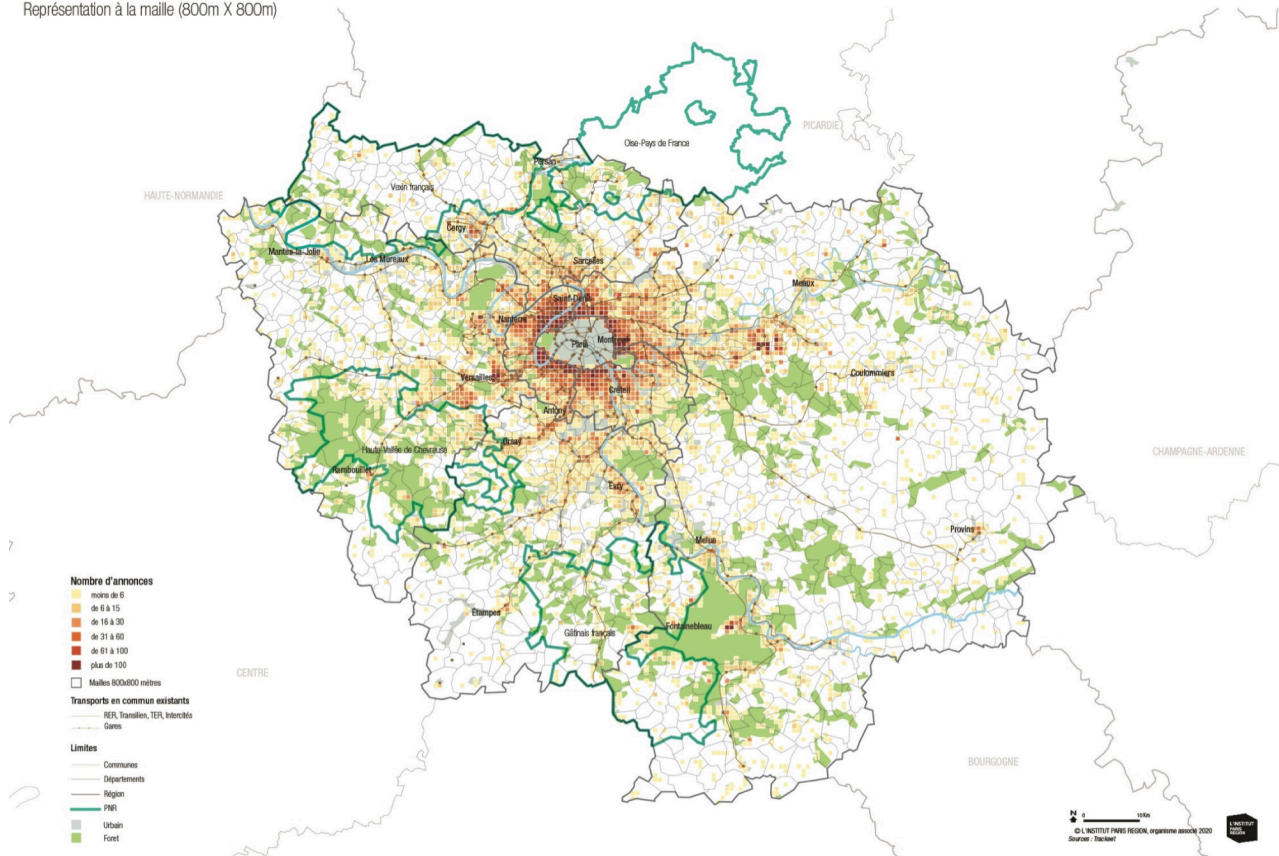
- Indices Notaires-Insee des prix des logements anciens - quatrième trimestre 2020 (2021). URL : <https://www.insee.fr/fr/statistiques/5058675>
- Base du dossier complet (2021). URL : <https://www.insee.fr/fr/statistiques/5359146>
- Institut Paris Région, 2021., Les locations saisonnières en Île-de-France, État des lieux d'avant crise. URL : https://www.institutparisregion.fr/fileadmin/NewEtudes/000pack2/Etude_2610/Rapport_Airbnb_250521complet.pdf
- Laurian, L., 2020, "Airbnb au sein des quartiers prioritaires de la ville : le cas de Saint-Denis", sous la direction de Guérois, M. et Madelin, M., Paris, Université de Paris, 92p.
- Le Brun, P., 2021. « Cartographier les aides publiques à l'investissement immobilier résidentiel en France (2003-2019) ». Working paper ANR WISDHOM. URL : <https://hal.archives-ouvertes.fr/hal-03157640>
- Le Corre, T., 2019. Paris à tous prix. Analyse des inégalités par une géographie de l'investissement sur le marché immobilier résidentiel en Île-de-France. Géographie. Université Paris 1 – Panthéon Sorbonne. Français
- Le Goix, R., Ysebaert, R., Giraud, T., Lieury, M., Boulay, G., Coulon, M., Rey-Coyrehourcq, S., Lemoy, R., J. Ramasco, J., Mazzoli, M., Colet, P., Theurillat, T., Segessemann, A., Marcińczak, S., et Bartosiewicz, B., « Unequal housing affordability across European cities. The ESPON Housing Database, Insights on Affordability in Selected Cities in Europe », *Cybergeog : European Journal of Geography* [En ligne], Data papers, document 974, mis en ligne le 16 avril 2021, consulté le 23 juillet 2021. URL : <http://journals.openedition.org/cybergeog/36478> ; DOI : <https://doi.org/10.4000/cybergeog.36478>
- Le Monde, 2021, Airbnb condamnée à 8 millions d'euros d'amende à Paris pour des annonces sans numéro d'enregistrement. URL : https://www.lemonde.fr/smart-cities/article/2021/07/01/la-plateforme-airbnb-condamnee-a-8-millions-d-euros-d-amende-a-paris_6086578_4811534.html
- Lovelace, R., Nowosad, J., & Muenchow, J., 2019. Geocomputation with R. *CRC Press*. <https://geocompr.robinlovelace.net/intro.html>
- Mermet A.-C., 2019, « Entendu / Entretien : Airbnb et la gentrification touristique des villes », *Urbanités*, mai 2019, URL <http://www.revue-urbanites.fr/entendu-entretien-mermet/>
- Peng, RD. Reproducible research in computational science. *Science*. 2011 Dec 2;334(6060):1226-7. doi: 10.1126/science.1213847. PMID: 22144613; PMCID: PMC3383002.
- Piganiol, V., « La planète Airbnb au risque de la COVID-19 », *Téoros* [Online], | 2020, Online since 29 October 2020. URL : <http://journals.openedition.org/teoros/6518>
- Pollard, J., 2010, « Soutenir le marché : les nouveaux instruments de la politique du logement », *Sociologie du travail* [En ligne], Vol. 52 - n° 3 | Juillet-Septembre 2010, mis en ligne le 18 août 2010. URL : <http://journals.openedition.org/sdt/14211> ; DOI : <https://doi.org/10.4000/sdt.14211>
- Schloss, P. D., 2018, Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. (J. Ravel, Ed.) *mBio*, 9(3), e00525–18. doi : 10.1128/mbio.00525-18

- Semi G., Tonetta M., 2019, « Plateformes locatives en ligne et rente urbaine à Turin : les classes moyennes face à l'austérité », *Annales de géographie*, 2019/3, no. 727, pp. 40-61, <https://www.cairn.info/revue-annales-de-geographie-2019-3-page-40.htm>
- Shabrina Z, Arcaute E, Batty M., 2021, Airbnb and its potential impact on the London housing market. *Urban Studies*. doi:10.1177/0042098020970865

Annexes

Annexe 1 : Carte des locations saisonnières en Île-de-France en 2019 (Institut Paris Région, 2021)

Nombres d'annonces de locations saisonnières
Représentation à la maille (800m X 800m)



Annexe 2 : Carte de l'évolution du parc RP et hors-RP (Institut Paris Région, 2021)



Annexe 3 : Extraits des tableaux de données bruts anonymisés

- Tableau mensuel :

Property ID	Property Type	Listing Type	Bedrooms	Reporting Month	Occupancy Rate	Revenue (USD)	Revenue (Native)	ADR (USD)
1	Apartment	Private room	1	2016-04-01	0.000	NA	0	NA
2	Apartment	Private room	1	2016-05-01	0.429	348	300	58.00
3	Apartment	Private room	1	2016-06-01	NA	NA	0	NA
4	Apartment	Private room	1	2016-07-01	NA	NA	0	NA
5	Apartment	Private room	1	2016-08-01	NA	NA	0	NA
6	Apartment	Private room	1	2016-09-01	NA	NA	0	NA
7	Apartment	Private room	1	2016-10-01	0.000	NA	0	NA
8	Apartment	Private room	1	2016-11-01	0.000	NA	0	NA
9	Apartment	Private room	1	2016-12-01	0.000	NA	0	NA
10	Apartment	Private room	1	2017-01-01	0.000	NA	0	NA
11	Apartment	Private room	1	2017-02-01	0.000	NA	0	NA
12	Apartment	Private room	1	2017-03-01	0.000	NA	0	NA
13	Apartment	Private room	1	2017-04-01	0.000	NA	0	NA

Showing 1 to 12 of 47,817 entries, 29 total columns

- Tableau individuel :

Property ID	Listing Title	Property Type	Listing Type	Created Date	Last Scr Date
1	Nice flat with a superbe view	Apartment	Private room	2010-07-07	20
2	STUDIO FLAT ON A BARGE (PARIS)	Boat	Entire home/apt	2011-02-02	20
3	Paris view-30m2-bright,confy,pratik	Apartment	Entire home/apt	2011-08-01	20
4	Original Modern Studio on the Seine	Apartment	Private room	2011-11-06	20
5	1 Bedroom-Upscale Paris Apartment	Apartment	Private room	2011-11-20	20
6	Terrace Bedroom in Paris	Apartment	Private room	2011-11-20	20
7	Home sweat home very near Paris	Apartment	Entire home/apt	2019-12-13	20
8	LOFT APARTMENT IN A BARGE (PARIS)	Boat	Entire home/apt	2012-04-02	20
9	Quiet sunny room Pt Versailles Expo	Apartment	Private room	2012-04-15	20
10	RENT CHARMING APARTMENT IN A BOAT !	Boat	Entire home/apt	2012-06-01	20
11	Spacious flat, close to Paris	Apartment	Entire home/apt	2012-06-12	20
12	Chambre moderne dans appartement 70 m2	Private room in apartment	Private room	2012-06-29	20

Showing 1 to 11 of 1,878 entries, 70 total columns

Annexe 4 : Variables post-traitement présentes dans les fichiers nettoyés

variable	fichier	créé ?	modifications / suppressions ?	thématique
id_logement	both	FALSE	FALSE	profil logement
id_hote	both	FALSE	FALSE	profil hote
Freq	individuel_clean	TRUE	FALSE	profil hote
titre	individuel_clean	FALSE	FALSE	profil logement
type	both	FALSE	TRUE	profil logement
mois	mensuel_clean	FALSE	FALSE	temporalité
debut	individuel_clean	FALSE	FALSE	profil logement
fin	individuel_clean	FALSE	FALSE	profil logement
revenus_mois	mensuel_clean	FALSE	TRUE	revenus
revenus_nuitee	both	FALSE	FALSE	revenus
revenus_annee	individuel_clean	FALSE	FALSE	revenus
Occupancy Rate LTM	individuel_clean	FALSE	FALSE	fréquentation
Occupancy Rate	mensuel_clean	FALSE	FALSE	fréquentation
nb_resa	mensuel_clean	FALSE	FALSE	fréquentation
nb_resa_annee	individuel_clean	FALSE	FALSE	fréquentation
capacite	individuel_clean	FALSE	FALSE	profil logement
Calendar Last Updated	individuel_clean	FALSE	FALSE	temporalité
prix	individuel_clean	FALSE	TRUE	revenus
duree_min	individuel_clean	FALSE	FALSE	temporalité
nb_nuitées	both	FALSE	FALSE	fréquentation
nb_dispo	both	FALSE	FALSE	fréquentation
nb_bloques	both	FALSE	FALSE	fréquentation
Amenities	individuel_clean	FALSE	FALSE	profil logement
Latitude	both	FALSE	FALSE	localisation
Longitude	both	FALSE	FALSE	localisation
License	individuel_clean	FALSE	FALSE	profil hote
Active	mensuel_clean	FALSE	FALSE	fréquentation
licence	individuel_clean	TRUE	FALSE	profil hote
estim	individuel_clean	TRUE	FALSE	revenus
prixpers	individuel_clean	TRUE	FALSE	revenus
duree	mensuel_clean	TRUE	FALSE	fréquentation
ancienneté	individuel_clean	TRUE	FALSE	profil logement
year	mensuel_clean	TRUE	FALSE	temporalité
tauxfreq	both	TRUE	FALSE	fréquentation
month	mensuel_clean	TRUE	FALSE	temporalité
day	mensuel_clean	TRUE	FALSE	temporalité
multi	individuel_clean	TRUE	FALSE	profil hote
nb_nuitées_tot	individuel_clean	TRUE	FALSE	fréquentation
nb_dispo_tot	individuel_clean	TRUE	FALSE	fréquentation
nb_bloques_tot	individuel_clean	TRUE	FALSE	fréquentation
nb_resa_tot	individuel_clean	TRUE	FALSE	fréquentation
activite	both	TRUE	FALSE	fréquentation
actif	mensuel_clean	TRUE	FALSE	fréquentation

Annexe 5 : carte de densité des logements Airbnb à Issy-les-Moulineaux (carroyage INSEE 200m)

