

HEC MONTRÉAL

Estimation de l'impact de Airbnb sur le marché immobilier

par

Michael Pereira

**Sciences de la gestion
(Option Économie Appliquée)**

*Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)*

Décembre 2017
© Michael Pereira, 2017

Résumé

En rendant possible et en facilitant le changement entre le marché locatif long-terme et celui du court-terme, Airbnb offre un choix aux propriétaires pour augmenter la rentabilité de leurs biens immobiliers. Notre étude porte d'abord sur la corrélation entre la pénétration d'Airbnb et le prix des logements pour des villes de deux comtés de Californie. Les résultats montrent une corrélation entre l'augmentation du nombre d'annonces Airbnb et le prix des logements. Deux éléments supplémentaires semblent avoir leur importance : l'écart entre les distributions de logements et des annonces d'Airbnb et le prix du logement. Il semblerait qu'une sur-représentativité des annonces par rapport aux logements entraînerait une augmentation plus importante du prix de ces derniers. De même lorsque les prix des logements sont plus faibles, la hausse du prix corrélée aux annonces semblerait plus forte.

Le prix d'une nuitée est déterminant pour un propriétaire. Prédire ce prix lui permet d'obtenir une estimation de sa rentabilité. Donc il pourrait faire son choix entre les deux types de marchés locatifs. Pour tenter de prédire ce prix nous utilisons d'abord la régression de LASSO et les arbres de décision, puis nous effectuons des méthodes de *Machine Learning* avec ces deux techniques et avec celle de forêt d'arbres. Les résultats varient suivant la technique utilisée, certaines méthodes transforment des caractéristiques bénignes des annonces en des facteurs impactant énormément le prix. Néanmoins on retrouve parmi les facteurs impactant positivement le prix des nuitées les mêmes facteurs qui sont mis en avant par le site d'Airbnb.

Mots clés : Marché Immobilier, Prix des logements, Airbnb, Prédiction, *Machine Learning*.

Abstract

With its role as short-term rental platform, Airbnb is making easier to switch between the long-term and short-term rental markets. Airbnb offers the opportunity for owners to make a choice between those two markets in order to increase their profit. Our study deals with the correlation between Airbnb's market penetration and house prices for two counties in California. Results show a positive correlation between the number of Airbnb listings and house prices. Two additional elements seem to be important. The first one is the discrepancies between the housing and the Airbnb ads distributions. The second one is the price of housing. It appears that the over representativeness of ads relative to the number of houses, leads to a greater increase in housing prices. Similarly, when housing is more affordable, price increases are more correlated with the number of listings.

The price per night at an Airbnb location is key for an owner. Predicting this price allows her to obtain an estimate of her profitability so that she can make the choice between the two types of rental markets. To try to predict this price, we first use the LASSO regression and a decision tree method, then we use Machine Learning with those two techniques and a Random Forest. Results vary according to the technique used, some methods transform benign characteristics of the listings into factors having a huge impact on the price. Nevertheless, among the factors which positively impact the one-night price, we find factors which are highlighted by the Airbnb website.

Keywords: Housing Market, Housing Prices, Airbnb, Statistical Prediction, Machine Learning.

Table des matières

Table des figures	6
Table des abréviations	7
Avant-Propos.....	8
Remerciements	9
Introduction	10
Revue de la Littérature	13
Chapitre 1 : L'économie collaborative et la régulation.	16
1.2 Les règlementations face à l'émergence de Airbnb.....	18
1.3 La Californie, San Luis Obispo et Santa Barbara.....	21
Chapitre 2 : Fonctionnement de Airbnb	23
2.1 Fonctionnement de la plateforme Airbnb.....	23
2.2 Une annonce Airbnb.....	24
2.3 Le processus de recherche et réservation.	26
Chapitre 3 : Les données	27
3.1 Les données de Airbnb	27
3.2 Les données de Zillow.....	29
3.3 Les données Macro-économiques	31
Chapitre 4 : Méthodologie.....	33
4.1 Méthodologie pour l'estimation d'impact	33
4.1.1 Préambule.....	33
4.1.2 Panel annuel.....	35
4.1.3 Panel mensuel.....	38
4.2 Méthodologie pour prédiction prix.....	40
4.2.1 Méthodes utilisées	40
4.2.2 Les jeux de données.....	41
Chapitre 5 : Résultats et Interprétations	43
5.1 Résultats du panel annuel	43
5.2 Résultats du panel mensuel	44
5.2.1 Les différences entre les modèles sur les deux périodes	44
5.2.2 L'ordre de grandeur de l'impact et des variables macro-économiques	50
5.2.3 Les distributions des logements et des annonces.....	52

5.2.4	L'importance de Santa-Barbara.....	55
5.3	Prédiction des prix d'une nuitée.....	62
5.3.1	Estimations directes.....	62
5.3.1.1	La méthode de LASSO.....	62
5.3.1.2	Méthode d'arbre de décision.....	65
5.3.2	Méthodes d'apprentissage statistique.....	67
Chapitre 6 :	Discussions.....	70
6.1	Prix de l'immobilier.....	70
6.1.1	Les approximations.....	70
6.1.2	Le retrait de Santa-Barbara.....	70
6.1.3	Les ordres de grandeurs.....	71
6.2	Prédiction du prix.....	72
6.3	Avenues de recherche.....	73
Apport décisionnel	75
Conclusion.....	76
Bibliographie.....	78
ANNEXE A :	Résultats annuels pour les deux périodes et équivalence de noms dans les résultats.....	81
ANNEXE B :	Résultats Panel Mensuel.....	82
Période 2010-2015 – Segment 1 Chambres	82
Période 2010-2016 – Segment 4 Chambres	83
ANNEXE C :	Valeurs des coefficients estimées pour les <i>Listings</i> d'Airbnb par segment sans tenir compte de leur significativité.....	84
ANNEXE D :	Distributions des logements par ville et par type en % et distribution des annonces Airbnb par ville et par type en %.....	86
ANNEXE E :	Calcul du Lambda pour les méthodes de LASSO.....	88

Table des figures

Figure 5.1 : Cumulative de toutes les annonces d'Airbnb par mois et par ville.	45
Figure 5.2 : Cumulative de toutes les annonces d'Airbnb pour chaque ville par mois sans Santa-Barbara.	45
Figure 5.3 : Distribution totale des annonces et de l'offre de logement par type.	53
Figure 5.4 : Distribution de l'offre et des annonces par type en prenant en compte que les villes du groupe étudié.	54
Figure 5.5 : Distributions par groupe avec résultats des coefficients estimés du modèle 1 par segment pour les deux périodes.	61
Figure 5.6 : Distributions par groupe sans Santa-Barbara avec résultats des coefficients estimés du modèle 1 par segment pour les deux périodes.	61
Figure 5.7 : Prix moyen en Millions USD des logements dans les groupes de villes étudiées par segment.	62
Figure 6.1 : Exemple de l'en-tête d'une annonce d'Airbnb.	73
Tableau 1.1 : Tableau reprenant les données fournies par Airbnb pour quelques villes aux États-Unis.	18
Tableau 3.1:Description des variables explicatives utilisées (autres que les annonces d'Airbnb).	32
Tableau 4.1 : Description de la nomenclature utilisées pour les différentes segmentations des variables cumulatives des annonces d'Airbnb.	35
Tableau 4.2 : Résultats des tests pour les deux différents modèles sur les deux périodes étudiées annuellement.	37
Tableau 4.3 : Résultats des tests pour les quatre différents modèles sur les deux périodes étudiées mensuellement.	39
Tableau 5.1 : Variable dépendante : Indice de Zillow. Valeurs des coefficients estimés pour les annonces d'Airbnb par segment, leurs écarts types, le nombre d'observations et les R-carrés ajustés (option cluster par ville).	48
Tableau 5.2 : Pourcentage de chacune des distributions.	54
Tableau 5.3 : Variable dépendante : Indice de Zillow. Coefficients estimés, leurs écarts types, le nombre d'observations et les R-carrés ajustés pour les listings Airbnb par segment avec et sans Santa-Barbara pour la période 2010-2015 (option cluster par ville).	57
Tableau 5.4 : Variable dépendante : Indice de Zillow. Coefficients estimés, leurs écarts types, le nombre d'observations et les R-carrés ajustés pour les listings Airbnb par segment avec et sans Santa-Barbara pour la période 2010-2016 (option cluster par ville).	59
Tableau 5.5 : Résultats de la régression de LASSO par jeu de données.	64
Tableau 5.6 : Résultats de la technique d'arbre de décision et d'émondage de l'arbre pour le jeu de données 1.	66
Tableau 5.7 : Résultats de la technique d'arbre de décision et d'émondage de l'arbre pour le jeu de données 2.	67
Tableau 5.8 : Résultats de la technique d'arbre de décision et d'émondage de l'arbre pour le jeu de données 3.	67
Tableau 5.9 : Caractéristiques et résultats des différentes méthodes prédictives sur les trois jeux de données.	68
Tableau 5.10 : Machine Learning avec LASSO par jeu de données.	69

Table des abréviations

Abréviations	Noms
LASSO :	Least Absolute Shrinkage and Selection Operator
RMSE :	Root Mean Square Error
TRI :	Taux de Rendement Interne
CSV :	Comma Separated Values
ZHVI :	Zillow Home Value Index
FHFA :	Federal Housing Finance Agency
Log :	Logarithme
SB :	Santa Barbara
TV :	Télévision
c.p. :	Ceteris Paribus

Avant-Propos

Actuellement la location court-terme entre particuliers fait les gros titres de la presse à travers le monde pour son présumé rôle de substitution à la location longue durée dans le cadre de propriétaires investisseurs. Pointant du doigt Airbnb qui se trouve être l'acteur majeur de ce nouveau marché, les autorités publiques tentent de prendre des décisions pour réguler ce phénomène qui pourrait entraîner une augmentation des prix de l'immobilier. Notre étude trouve donc ses racines ici.

Cette étude rédigée dans ce mémoire de maitrise est le résultat d'une année de travail comprenant la récolte des données, le traitement de ces données et leurs analyses. Ce mémoire a pu être effectué grâce à l'étroite collaboration de M. Mario Samano.

Remerciements

J'avais envie d'adresser mes sincères remerciements à ceux qui ont contribué à l'élaboration de mon mémoire de maîtrise.

En premier lieu, je tiens tout particulièrement à remercier Monsieur Mario Samano, Professeur Adjoint d'Économie à HEC Montréal. En tant que directeur de mémoire il a su me soutenir, me questionner, et nous avons pu établir une relation de travail et de confiance qui a permis la réalisation de ce mémoire.

Je tiens également à remercier l'équipe administrative d'HEC Montréal qui a su être à l'écoute et s'ajuster aux différentes étapes de cette réalisation.

Pour ce qui est du soutien technique je remercie Lucas Verginer qui malgré la distance m'a donné de bons conseils et m'a aidé.

Plus personnellement, sans explicitement les citer mais ils se reconnaîtront, je remercie mes proches pour leur soutien indéfectible.

Introduction

Depuis quelques années avec le développement d'internet et de la mobilité de nombreuses *start-ups* se sont créées et ont agi comme disrupteurs de marché. Cette disruption de marché a été directement liée à l'essor de ce qu'on appelle l'économie de partage ou l'économie collaborative. Cet essor s'est fait non sans mal et a vu et voit encore s'affronter les pro-économies de partage et leurs détracteurs. Ces détracteurs souvent appuyés par les *lobbies* des industries déjà en place ont réussi à créer une fronde contre l'économie collaborative.

Dans le secteur de l'immobilier Airbnb est le plus grand disrupteur de marché. En effet Airbnb est le leader dans le domaine de la location temporaire sur internet en servant d'intermédiaire sur un marché biface (*two-sided market*). La start-up a été directement pris pour cible pour viser tout ce nouveau pan de l'économie collaborative. Actuellement on semble attribuer aux locations temporaires de type Airbnb une augmentation des prix de l'immobilier. L'argument phare de cet effet d'augmentation des prix porte sur l'idée d'un choix délibéré des propriétaires qui retirent leurs biens du marché locatif long-terme pour celui du court-terme. Le but de ces propriétaires est d'accroître leur rentabilité.

Nous nous sommes rendu compte que la plupart des études existantes portées sur les sites de publication d'annonce de type e-commerce. Une seule étude empirique avait portée sur le conflit entre les hôteliers et Airbnb. Dans ce contexte notre étude semblait d'une part être au cœur des enjeux politiques et socio-économiques actuels et d'autre part s'intégrer parfaitement dans la continuité des études effectuées.

Au vu de ces premières constatations notre étude est composée de deux grands axes distincts. Le premier est centré sur l'analyse de la corrélation entre Airbnb et les prix du marché immobilier. Ainsi nous tenterons d'abord d'estimer l'impact de l'augmentation d'annonces publiées sur la plateforme sur le prix des logements.

Le second axe va porter sur la prédiction des prix grâce à la collecte de données qui nous a permis d'obtenir de nombreuses caractéristiques pour chaque annonce Airbnb. Ici nous tenterons de prédire le prix d'une nuit Airbnb pour un propriétaire qui se

retrouve devoir choisir entre les deux marchés. Cette tentative de prédiction de prix se fera par une approche de type *Machine Learning*. Cette approche est également présentée dans le but de montrer les possibilités que peuvent engendrer les méthodes de collectes par *web-scraping* et de ne pas s'arrêter à la seule utilisation du nombre d'annonces présentes sur le site.

Pour mener à bien notre démarche, notre étude reprendra dans un premier temps la littérature existante autour des thèmes de l'entrée sur le marché des compétiteurs de l'internet, d'Airbnb et du marché immobilier. Notre première partie consistera en une mise en contexte des enjeux autour d'Airbnb et des régulations choisies à travers le monde. Dans un deuxième temps nous présenterons de manière détaillée le fonctionnement de la plateforme Airbnb car les variables explicatives d'intérêt de nos modèles proviennent directement des données du site. Puis nous présenterons les données que nous avons utilisées ainsi que la méthodologie. Suivra une partie de présentation des résultats et d'interprétations concernant le potentiel impact des annonces d'Airbnb sur les prix de l'immobilier mais aussi celles de l'approche de la prédiction du prix d'une nuitée par différentes techniques de *Machine Learning*.

Les résultats du premier axe de l'étude sur la corrélation entre le nombre d'annonces et le prix des logements semble montrer deux éléments à prendre en compte. Ces éléments sont la différence de représentativité entre le nombre d'annonces et le nombre de logements présents ainsi que le niveau de prix de ces logements. Ici une sur-représentativité des annonces par rapport aux logements entrainerait une augmentation plus importante du prix de ces derniers. De même lorsque les prix des logements sont plus faibles, la hausse du prix corrélée aux annonces semblerait plus forte.

Pour le second axe de l'étude concernant la prédiction des prix de la nuitée il s'avère que les facteurs déterminants le prix sont ceux qui viennent traditionnellement à l'esprit comme par exemple le nombre de chambres, la capacité ou la réputation de l'hôte, ce sont également les caractéristiques mises en avant par Airbnb. De plus nous avons montré que le choix des méthodes de *Machine Learning* dépend des données utilisées et que ce choix modifie les résultats obtenus.

Ces interprétations et ces résultats seront discutés en fin de rapport, au vu des différentes problématiques autour des données et des méthodes utilisées ainsi qu'au vu des futures publications sur le point d'être publiées. Enfin nous tenterons d'apporter une conclusion synthétisant l'ensemble de notre travail en essayant de répondre au mieux à la problématique soulevée plus haut.

Revue de la Littérature

La littérature actuelle s'est déjà penchée sur les nouvelles habitudes de consommation engendrée par l'arrivée d'internet et des nouveaux moyens de commercer. C'est le cas de Brynjolfsson et al. (Brynjolfsson, 2009) qui ont montré que la substitution par les achats en ligne est moins forte pour des produits de niche. Dans la même idée Forman (Forman, 2009) a montré que la proximité d'un magasin pour un client le rendait moins sensible aux promotions en ligne. Pour ce qui est des mesures de l'impact sur l'industrie de nouveaux entrants sur un marché déjà existant, certaines études ont été faites sur Craigslist. Craigslist est un site permettant la publication de petites annonces sur internet, ce type d'annonces était généralement affiché dans la presse. Seamans et Zhu (Seamans, 2013) ont montré d'une part que les journaux ont augmenté leurs tarifs d'abonnement et que le tarif de publication d'annonce avait diminué. D'autre part ils ont estimé que les afficheurs d'annonces ont économisé approximativement cinq milliards de dollars grâce à cette concurrence. Au sujet de Craigslist, les journaux auraient perdu 7% d'affichage d'annonce entre Janvier 2005 et Avril 2007 selon Kroft et Pope (Kroft, 2013).

Dans le secteur de l'économie de collaborative, des études se sont portées sur le covoiturage et son impact réel. Par exemple Cervero (Cervero, 2007) a trouvé que le covoiturage avait fait diminuer le nombre de kilomètres parcourus, la consommation d'essence et le nombre de propriétaires de véhicules. Martin (Martin, 2010) a trouvé que dans un groupe de propriétaires participants à un programme de covoiturage, le nombre de véhicules par propriétaire avait diminué de 50%.

Plus récemment, Airbnb étant le plus connu et le plus important site de location de logements entre particuliers, a été au centre de plusieurs études. Zervas et Georgios (Georgios Zervas, 2013) ont vérifié l'impact de l'arrivée de ce nouveau moyen de consommation sur les joueurs traditionnels du secteur hôtelier. Ainsi ils ont créé une cumulative pour avoir l'impact de l'entrée d'Airbnb sur le marché de l'hôtellerie. Nous utiliserons ce même concept pour notre étude. De plus par une segmentation du secteur

hôtelier ils ont pu mettre en avant que les hôtels de moins bonne qualité et ceux qui n'avaient pas d'équipement pour accueillir les gens d'affaire sont les plus affectés.

Au niveau du secteur immobilier, Dayne Lee (Lee, 2016) a essayé de montrer comment l'arrivée de Airbnb a exacerbé le marché de la location à Los Angeles. Plus particulièrement comment Airbnb a des effets sur les logements abordables. La réduction de l'offre de logements abordables se fait selon lui par deux mécanismes interconnectés. D'abord par le simple fait de retirer un bien du marché locatif longue durée et de le transformer en logement dit touristique, ce qui provoque une augmentation des loyers. Le deuxième mécanisme auquel il fait référence est le mécanisme d'« hotélisation », celui-ci est rendu possible si l'hôte d'Airbnb peut louer son bien moins cher qu'une chambre d'hôtel. Ainsi le gain pour l'hôte est garanti. Le premier mécanisme et le deuxième sont liés, le second fait accentuer le premier. En s'appuyant sur les données de Zillow et de Airbnb, qui seront également nos sources de données dans cette étude, il constate que dans les quartiers les plus densément desservis par Airbnb (sept dans tout Los Angeles), les loyers en 2014 étaient 20% plus élevés et augmentaient plus rapidement de 33% en comparaison à ceux de la ville toute entière. De plus dans ces sept quartiers les logements listés sur Airbnb représentent 3% du stock des biens à la location. Alors que le pourcentage de logements vacants est de 3,5%. Le quartier de Venice lui se trouve dans une situation bien pire avec 12,5% de logements sur Airbnb. Avec un coefficient d'élasticité du prix de 0,2 sur les loyers, un logement avec une chambre dont le loyer est 2680\$ verrait son loyer augmenter de 67\$ par mois. Le problème va être la croissance d'Airbnb et l'augmentation de la conversion des logements en biens touristiques. Globalement, en 2014, Airbnb a retiré 1% de l'offre des logements de Los Angeles et les loyers ont augmenté de 7,3% mais Airbnb n'est que partiellement responsable.

Le marché de l'immobilier est quant à lui structuré d'une demande de logement provenant des ménages. Cette demande est supposée positivement corrélée à la démographie et aux revenus. Elle est supposée être négativement corrélée au taux d'intérêt et au prix de l'immobilier. L'offre est le fait de promoteurs qui cherchent à maximiser leur profit (Levasseur, 2013). Ces investisseurs souhaitant maximiser leurs profits vont prendre en compte les éléments suivants : loyer, prix d'achat, charges

d'exploitation pour calculer le rendement locatif net et ils vont y ajouter le taux d'intérêt du financement pour calculer le Taux de Rendement Interne (Hoesli, 2008). C'est avec cette notion d'offre et de demande ainsi qu'en connaissance des enjeux pour un investisseur de prendre en compte certains paramètres que nous avons sélectionné nos données macro-économiques. De plus lors du choix du propriétaire pour passer d'un marché locatif à l'autre le calcul d'un TRI peut s'avérer utile et pour cela il lui faudra estimer le loyer qu'il pourra retirer avec la location court-terme. D'où l'intérêt de prédire le prix de sa nuitée sur Airbnb.

Chapitre 1

L'économie collaborative et la régulation.

1.1 L'Économie collaborative

L'économie collaborative touche de nombreux secteurs, les transports, le logement, les services et la production. De plus elle est souvent le prolongement de pratiques préexistantes. Elle semble se poser comme une alternative au modèle de propriété, de développement et de croissance (Simon Borel, 2015).

On retrouve l'économie collaborative dans quatre grands domaines : la production-réparation, l'éducation, le financement et la consommation. Malgré les différentes pratiques mises en place dans ces différents domaines, des principes communs peuvent être mis en avant. Tout d'abord celui de l'horizontalité qui s'appuie sur la coordination des individus au lieu de rester dans une notion de verticalité hiérarchique. De ce principe en découle celui de la démocratisation des aptitudes et des compétences. Enfin le principe d'abandon de la logique propriétaire vers la logique d'usage semble être l'essence même de cette économie (Simon Borel, 2015).

Pour caractériser les acteurs de l'économie collaborative les trois grandes tensions décrites par Simon Borel semblent être judicieuses. La première tension est entre « Désintermédiation vs ré-intermédiation », la seconde est entre « Marchand vs Non marchand » et la troisième est entre « Local vs Global ». À la lumière de ces trois tensions nous pouvons aisément caractériser la place d'Airbnb dans l'économie collaborative comme étant une plateforme de ré-intermédiation marchande globale. En effet Airbnb agit comme un nouvel intermédiaire qui monétarise son service avec un positionnement global.

L'apparition de ces nouveaux acteurs entraînent une mutation socio-économique. D'un point de vue micro-économique de nombreuses promesses sont faites sur l'augmentation du pouvoir d'achat et de revenu des individus. Au niveau macro-économique, certains pensent que globalement la création d'emploi sera supérieure à

ceux détruits. Mais ces nouveaux emplois semblent également plus précaires, on voit apparaître le terme d' « Ubérisation » pour caractériser ce phénomène. Pour les états, le revenu de la fiscalité découlant à la fois des taxes sur les entreprises et les échanges est un enjeu de taille car bien souvent il conditionne la protection sociale des individus.

D'un point de vue sociétal, Rachel Bostman¹ met en avant le lien perdu que recrée ce type d'économie dans la société actuelle (Bostman, 2010). Le partage rapproche les gens entre eux et la nouvelle monnaie est la confiance entre les individus d'où l'importance de la notation entre les membres au sein d'une plateforme d'économie collaborative (Bostman, 2012). Néanmoins pour tous ceux qui ont déjà utilisé un Airbnb, il n'est pas rare de ne jamais voir le loueur du logement dans lequel nous séjournons. Ce lien reste alors un lien virtuel et une confiance virtuelle.

Ces acteurs dont Airbnb fait partie ont très bien compris les enjeux socio-économiques devant leur essor perçu comme une force destructive schumpétérienne. Le débat sociétal relayé par la presse, est un affrontement entre les acteurs traditionnels du milieu économique (ex : les hôteliers) et les nouveaux entrants disruptifs (ex : Airbnb) avec au milieu les gouvernements qui doivent arbitrer en tenant compte de l'opinion publique.

Airbnb en tant qu'acteur sait qu'il doit convaincre les gouvernements de son apport économique et des bienfaits que l'entreprise fournit à la société. C'est pour cela qu'Airbnb publie des études économiques qu'elle met à disposition de la presse et du grand public sur son site internet. Certaines études sont plus détaillées, comme par exemple celle sur la France et Paris qui est la première ville Airbnb au monde en nombre d'offre d'hébergement. Ainsi cette étude a été menée par le cabinet Astère² et se base sur une année pleine de septembre 2014 à août 2015. Au cours de cette année 176800 hébergements ont été réservés ce qui a créé 481 millions d'euros de revenus complémentaires pour les hôtes. L'étude montre que 140000 hôtes ont accueilli des visiteurs et que plus de la moitié de ces hôtes ont des revenus inférieurs au revenu médian français. Paris, à elle seule, a accueilli 1,3 million de voyageurs Airbnb, pour un revenu

¹ Co-Auteur du livre « What's mine is yours ». Elle défend l'économie collaborative.

² Astère est un cabinet spécialisé en conseil économique basé à Paris.

pour les hôtes de 218 millions d'euros. Les dépenses directes de ces voyageurs à Paris hors hébergement est de 980 millions d'euros. De l'autre côté de l'atlantique à Montréal, une étude interne d'Airbnb a été menée entre avril 2013 et mars 2014. Airbnb affirme que les retombées économiques totales sont de 54,6 Million de dollars et qu'il y a 821 emplois induits grâce à la présence de la compagnie. Les équipes d'Airbnb décrivent l'hôte montréalais typique comme percevant 280\$ par mois et louant son logement 52 nuits par an. Ils annoncent même que ce revenu complémentaire a permis à 54% des hôtes de conserver leur logement. Le tableau 1.1 reprend les informations principales de ce type d'études faites par Airbnb sur un an pour quelques grandes villes aux États-Unis (Airbnb, 2015).

Ville	Impact économique Total (USD)	Total d'emplois induits	Gain annuel moyen par hôte	Nb de nuits par an	Part des personnes ayant pu rester dans le logement grâce à Airbnb
Seattle	\$178M	1700	\$8000	79	60%
Chicago	\$209M	2100	\$8300	77	56%
Boston	\$51M	490	\$8020	68	53%
Los Angeles	\$312M	26000	\$7920	59	71%
New-York	\$632M	n.c.	\$7530	n.c.	62%

Tableau 1.1 : Tableau reprenant les données fournies par Airbnb pour quelques villes aux États-Unis.

1.2 Les réglementations face à l'émergence de Airbnb

La création et le développement rapide de Airbnb est l'un des nombreux exemples de l'effet déstabilisateur des marchés occasionné par l'économie collaborative. En l'espace de quelques années depuis sa création et son lancement en 2008, la *startup* est devenue un acteur incontournable du secteur du voyage. Basé sur deux idées simples dont la première est un concept gagnant-gagnant entre un hôte locataire de son logement souhaitant accueillir un voyageur pour pouvoir payer son loyer et un voyageur qui souhaite voyager à moindre coup. La seconde est la communauté dans laquelle chacun peut commenter et noter son expérience d'échange. Cette possibilité de commenter et que les commentaires ne soient pas supprimés a créé un espace de confiance. C'est avec cette

promesse simple et de partage qu'Airbnb est devenu une licorne³ rentable pour la première fois au deuxième semestre de 2016 soit 8 ans après son lancement. En 2017 la plateforme compte plus de deux millions d'annonces dans 34000 villes (Le Figaro, 2017). Son succès est entre-autre dû à un design permettant de réserver en moins de trois clics et à un esprit communautaire où les utilisateurs doivent mettre leur photo et se noter mutuellement (Businessinsider, 2016). Le succès étant au rendez-vous l'entreprise a connu une croissance exponentielle qui lui permet d'atteindre une valorisation de plus de 30 Milliards (Le Figaro, 2017).

Ce succès perturbe grandement les acteurs traditionnels du secteur touristique qui se sentent en danger face au type de plateforme comme Airbnb. Les sommes d'argent en jeu attirent également les pouvoirs publics locaux à demander leur part du gâteau mettant la compagnie face à de nombreux conflits à travers le monde (Boursorama, 2014).

Et pour cause, les enjeux autour de la croissance fulgurante des plateformes collaboratives comme Airbnb sont multiples pour les pouvoirs publics. Ils se situent entre la volonté de ne pas freiner le développement économique de l'économie collaborative mais aussi de faire respecter les lois et de garantir une équité et une cohésion sociale.

Dans de nombreux pays la sous-location du logement par un locataire est tout simplement illégale et le locataire n'a pas le droit de faire de bénéfice sur la période louée. C'est-à-dire que le montant gagné sur la période doit être au maximum équivalent au prorata du prix du loyer. Néanmoins au vu des statistiques fournies par Airbnb on s'aperçoit que bon nombre d'hôtes de la plateforme de location temporaire sont des locataires qui mettent à disposition leur logement lorsqu'ils s'absentent. D'un point de vue de l'équité sociale apparaît également le souci pour les autorités de récupérer les taxes, qu'elles soient touristiques ou sous forme d'impôt sur le revenu. Cette volonté d'équité vient également du fait que les pouvoirs publics ne souhaitent pas que l'arrivée sur le marché de la location temporaire de plateforme tel qu'Airbnb ne crée une concurrence déloyale vis-à-vis des hôtels ou des résidences touristiques officiellement déclarer.

³ Une licorne est une start-up dont la valorisation dépasse le milliard de dollars.

Plus récemment un nouvel argument allant contre le développement de Airbnb est apparu, celui de l'impact sur le marché immobilier. Pour un investisseur, la valeur d'un logement dépend de son rendement locatif or la location courte durée est plus fructueuse qu'une location de durée normale. Ainsi louer à la semaine devient plus rentable que de louer au mois. De plus ce rendement s'accroît si l'on se trouve sur un bien meublé et situé dans une zone touristique. Ce qui peut entraîner au moyen terme, dans les zones touristiques un risque de ne plus trouver de logements nus (Courrier Parlementaire, 2017). La raréfaction de l'offre entraîne donc une augmentation des prix des biens immobiliers.

Pour réguler ce phénomène il semble y avoir deux approches à travers le monde. La première approche consiste d'une part à imposer des contraintes dans la durée de la location des hôtes des plateformes et d'autre part à récupérer les taxes dues en partenariat avec la plateforme Airbnb. La seconde approche vise à sanctionner directement les plateformes de types Airbnb et de considérer leur pratique comme illégale. Voici quelques exemples de régulations des autorités dans des villes touristiques et leurs buts.

En Europe où Airbnb s'est développé rapidement dans les capitales et les villes touristiques l'approche de la régulation n'est pas homogène. Par exemple, Londres et Amsterdam, ont respectivement des durées de locations maximum de 90 et 60 jours. Le but avoué de ces villes est d'une part de réduire les nuisances occasionnées par un afflux de touristes dans des logements et quartiers qui ne sont pas originellement prévus à cet effet mais aussi de limiter l'impact sur les prix des logements. À Londres, la mairie a clairement dénoncé les investisseurs qui refusent de louer à longue durée et qui utilisent les plateformes de types Airbnb. À Paris qui est la première ville au monde en termes d'offre d'hébergements Airbnb, la législation est différente. La loi est nationale, elle donne le droit aux villes de plus de 200000 habitants de récupérer la taxe de séjours (taxe touristique) et limite la location à un maximum de 120 jours (Nouvel Observateur, 2016). L'ensemble de ces dispositions ont été choisies pour tenter de trouver un équilibre entre le développement économique de la plateforme et une équité avec l'économie existante. D'autres villes en Europe et à travers le monde ont pris la direction opposée en rendant les plateformes de type Airbnb soit directement illégales soit en rendant leurs démarches

marketing illégales. Par exemple dans l'état de New-York, les personnes louant leurs logements moins de 30 jours consécutifs peuvent se voir infliger une amende. De plus il est également devenu illégal de promouvoir par la publicité tout type de site ou d'annonce permettant d'accéder à des logements enfreignant la réglementation des 30 jours consécutifs. En Europe, la ville de Barcelone applique le même type de sanction contre les plateformes de location qu'elle tient pour responsable de la prolifération des locations de logements temporaires illégaux (Nouvel Observateur, 2016). À Berlin, la location via Airbnb est devenue totalement illégale peu importe la durée de celle-ci.

En ce qui concerne le Québec, le gouvernement pour tenter d'encadrer au mieux l'utilisation des plateformes a élaboré la loi 67. La loi impose à toute personne souhaitant louer son logement de manière régulière pour des périodes de moins de 31 jours consécutifs l'obtention d'une attestation auprès de Tourisme Québec (Ici Radio Canada, 2017). Plusieurs constats d'inefficacité de cette loi ont été avérés comme le témoigne l'avocate Chloé Fauchon dans *Le Devoir* en mettant en avant les zones d'ombre de la loi 67 sur la notion d'affichage qui permet de la contourner aisément (*Le Devoir*, 2016). Devant cette inefficacité flagrante et avouée des autorités, plusieurs approches sont présentées par des députés pour s'approcher d'un modèle semblable à celui d'Amsterdam (Ici Radio Canada, 2017). De plus il semblerait également que des discussions soient déjà avancées entre Airbnb et les autorités locales des villes canadiennes représentant les plus gros marchés : Montréal, Vancouver et Toronto, pour faciliter la régularisation et la collecte des taxes (*Journal de Montréal*, 2017).

1.3 La Californie, San Luis Obispo et Santa Barbara

En Californie point de départ de l'histoire de Airbnb la régulation se fait également non sans conflit. L'état de Californie tarde à statuer sur les locations de court-terme et laisse aux instances locales la possibilité de légiférer. Ainsi les villes et les comtés choisissent eux-mêmes de s'adapter face à l'expansion de l'utilisation des plateformes de type Airbnb (*LA Times*, 2017).

Quant aux comtés qui seront ceux de notre étude, la réglementation est souvent identique pour limiter le nombre d'hôtes utilisant Airbnb ou tout du moins avoir une expansion qui se conforme aux différentes règles. L'obligation est d'obtenir une attestation permettant de louer temporairement son logement. Or celle-ci est délivrée suivant des règles de concurrence géospatiale visant à limiter l'offre par zone géographique.

Dans le comté de Santa-Barbara, tout logement ou toute partie de logement qui est mis à la location moins de 30 jours consécutifs est considéré comme une location de vacances. Selon le Planning Division du comté il y a changement d'utilisation du logement et pour ce faire les règles du planificateur doivent être suivies et une autorisation doit être délivrée. Avant tout changement il est d'ailleurs fortement recommandé par le comté de Santa Barbara de remplir une pré-autorisation (Government of Santa-Barbara, 2017).

Dans le comté de San-Luis-Obispo, les dispositions légales sont identiques avec des exceptions dans certaines ville comme Cambria, Cayucos, Avila Beach et Adelaide où la distance entre deux locations de vacances est strictement réglementée. Dans le cas où la distance est jugée trop faible, le loueur se voit soit accorder une dérogation ou se voit soit dans l'obligation d'arrêter sa location. Pour faire suivre la loi, la démarche des autorités est simple. Tout d'abord identifier les adresses qui semblent louer de manière illégale. Puis leur envoyer une lettre pour qu'ils effectuent une demande de régularisation. L'une des utilisations de cette méthode a permis de faire régulariser 12% des locations jugées illégales. Par la suite si les personnes ne souhaitant pas faire de demande de régularisation continuent leurs activités, elles s'exposent à des amendes pouvant atteindre 500\$ par jour (The Tribune, 2017). Au final cette démarche effectuée dans les comtés que nous allons étudier est similaire à celle appliquée au Québec.

Chapitre 2

Fonctionnement de Airbnb

Pour mieux comprendre la collecte de données et l'estimation de l'impact d'Airbnb sur le marché immobilier il est important de connaître le fonctionnement de la plateforme et la présentation des informations présentes sur une annonce de logement. L'ensemble des informations décrites ci-dessous pour le fonctionnement, le détail de l'annonce et le processus de recherche et réservation proviennent directement de l'utilisation et de la navigation sur le site internet d'Airbnb⁴.

2.1 Fonctionnement de la plateforme Airbnb

Airbnb est une plateforme permettant l'échange entre un hôte qui met à la location un lit (chambre partagée), une chambre (chambre privée) ou son logement de manière temporaire avec un voyageur qui souhaite y séjourner. Le fait de passer par la plateforme de Airbnb permet de sécuriser la transaction entre l'hôte et le voyageur. En effet contrairement à un site d'annonce standard la transaction du paiement de la location se fait via la plateforme soit par le site internet soit par l'application mobile. L'inscription des deux types d'utilisateurs (hôte et voyageur) est totalement gratuite. La *start-up* tire ses revenus d'un pourcentage prélevé à chaque transaction. Pour le voyageur le pourcentage est inclus dans le prix. La *start-up* agit comme un intermédiaire en donnant à l'hôte la possibilité d'afficher son bien à la location avec une certaine visibilité compte tenu du rayonnement et de la crédibilité d'Airbnb et lui garantit une réception des fonds sécurisée. À l'inverse elle offre au voyageur un choix important d'annonces pour ses déplacements et lui garantit une location après avoir effectué son paiement.

Hormis l'aspect de sécurisation des transactions entre un prestataire de service (l'hôte) et son client (le voyageur), Airbnb permet d'afficher et d'avoir accès à des annonces standardisées. Chaque bien est détaillé dans une page qui lui est propre et

⁴ L'adresse de la plateforme Airbnb est : www.airbnb.com.

unique. Ainsi la qualité de l'information est clairement augmentée par rapport au site d'annonces classiques. Pour le voyageur le coût de l'accès à l'information est nul. Cela lui permet de faire aisément des comparaisons entre les différents biens proposés sur la plateforme en fonction de ses propres besoins.

Airbnb est central à l'étude et de ce fait il semble important de détailler une annonce type que l'on retrouve sur le site internet ainsi que le processus que suit un utilisateur (voyageur). En connaissant précisément les informations se trouvant sur l'annonce et le moyen d'y accéder nous allons pouvoir les collecter de manière automatisée par la méthode décrite dans la section « donnée ».

2.2 Une annonce Airbnb

Une page d'annonce d'un bien sur Airbnb comporte quatre parties avec de nombreux détails. Les parties sont la Présentation, les Commentaires, l'Hôte et l'Emplacement.

Tout d'abord la partie Présentation comporte une partie de résumé dans laquelle on retrouve une photo du bien avec le prix de la nuitée ainsi que directement la possibilité d'obtenir plus de photos du bien. Suit alors le titre de l'annonce, une photo de l'hôte et un récapitulatif des informations importantes : notation du logement par les autres voyageurs, capacité d'accueil, nombre de lits, type de logement et le nombre de salle de bain. La partie « résumé » de l'annonce se termine par une description écrite par l'hôte à l'intention des voyageurs avec la possibilité de contacter l'hôte.

Après cette partie de résumé de la partie de Présentation s'amorce le corps de l'annonce. Dans cette partie de la section Présentation apparaissent de nombreux détails. Les détails commencent par une liste des équipements fournis et absents du logement. De base une cinquantaine d'équipements possibles sont listés et peuvent s'y ajouter des équipements spécifiques à l'accès aux personnes à mobilité réduite ou souffrant d'un handicap. Ainsi le bien peut être détaillé avec plus de 70 équipements différents.

À la fin de cette section Présentation apparaît les tarifs spéciaux, un détail des couchages, le règlement intérieur, les conditions d'annulations et la durée minimale du

séjour. Les tarifs spéciaux se composent du tarif pour un voyageur supplémentaire, des tarifs semaine et week-end ainsi que des tarifs réduits. Le règlement intérieur correspond à des règles choisies par l'hôte. Les conditions d'annulations correspondent à celle choisies par l'hôtes, il y en a six possibles. Le site différencie les locations de type courte et longue durée et y attribue un choix de trois conditions d'annulations. Pour la location courte durée les conditions peuvent être flexibles, modérées ou strictes. Les deux premières permettent des remboursements en totalité avant une date limite. L'annulation de type flexibles permet d'annuler jusqu'à 24 heures avant l'arrivée avec remboursement. Ce type de condition d'annulation s'apparente à celle que l'on retrouve dans la plupart des hôtels.

La durée minimale de séjour correspond à un choix personnel de l'hôte qui peut offrir des séjours plus flexibles avec par exemple une nuit minimum demandée ou moins flexible selon sa convenance.

La seconde partie de l'annonce est la partie Commentaires. Airbnb étant une plateforme d'échange qui a pour valeur initiale la convivialité, la *start-up* souhaite que les différents utilisateurs communiquent entre eux et communiquent leurs expériences au travers d'un système de notation après chaque voyage effectué. La notation se fait avec un système allant de zéro à cinq étoiles pour six critères de notations différents. Ces critères sont la précision, la communication, la propreté, l'emplacement, l'arrivée, et le rapport qualité-prix. Enfin par la suite chacun des hôtes ayant séjourné dans la location peuvent laisser un commentaire personnalisé.

La troisième et la dernière parties sont les plus courtes. La troisième partie de l'annonce porte sur les caractéristiques de l'hôte. On y voit apparaître son nom, sa date d'inscription, sa langue, son taux de réponse et son délai de réponse. Une photo avec la présence ou non d'une médaille le désignant comme *superhost*, la possibilité d'aller sur une page de commentaires portant sur lui et un lien pour le contacter. La dernière partie porte sur l'emplacement du bien avec le nom de la ville et une zone colorée sur un carte Google Map. L'adresse exacte est toujours communiquée après que la réservation soit effectuée.

2.3 Le processus de recherche et réservation.

Lorsqu'un utilisateur (voyageur) arrive sur le site ou sur l'application de Airbnb, il commence toujours par une recherche. La plupart du temps il sélectionne ses dates de voyages, le type de logement, le nombre de personnes et il entre la destination de son choix. Il peut également y ajouter des filtres qui sont proposés par le site. Les premiers filtres proposés porte sur le prix et le type de réservation. Enfin de manière plus avancée il peut filtrer par tous les éléments précédemment cités qui se trouvent sur l'annonce.

Le résultat d'une recherche ne peut excéder 306 annonces. Ces annonces sont présentées au nombre de 18 par page avec un maximum de 18 pages. Chaque annonce apparaît avec une photo de présentation et un résumé descriptif comportant le nom de l'annonce, le type de logement, le nombre de lits, le prix par nuit et la notation du logement. L'utilisateur peut ainsi naviguer et découvrir les annonces détaillées en un simple clic. Lorsque son choix est fait il peut réserver. Deux types de réservations sont possibles, la réservation instantanée pour les hôtes qui ne veulent pas approuver chaque réservation. Avec ce type de réservation le voyageur peut aussitôt confirmer sa réservation pour le bien choisi. L'autre type de réservation est la demande de réservation. La plupart des hôtes préfèrent utiliser ce second type de réservation. Il leur permet d'approuver la réservation avant qu'elle ne soit définitive. L'hôte dispose d'un délai de 24 heures pour accepter la demande de réservation du voyageur. Lorsqu'il confirme la réservation se fait de manière automatique et le paiement est directement effectué sans que le voyageur n'ait d'autres démarches à effectuer. Lorsque le paiement est effectué le voyageur obtient l'adresse exacte et les coordonnées pour contacter l'hôte.

Chapitre 3

Les données

Pour cette étude nous avons combiné des données provenant de différentes sources. La première est le site internet de Airbnb pour les informations sur l'apparition des annonces publiées. La seconde partie des données provient du site internet de Zillow⁵ qui nous permet d'obtenir des données sur la valeur des biens immobiliers. Enfin la dernière partie des données provient de sites gouvernementaux. Ces données sont des données macroéconomiques.

3.1 Les données de Airbnb

Les données de Airbnb ont été récupérées par technique de *web-scraping*. Le *web-scraping* est une technique d'automatisation de l'exploration de page internet et de collecte des informations s'y trouvant. Le programme informatique va agir comme un voyageur qui effectue une recherche sur le site de Airbnb pour un lieu donné sans imposer de dates de voyage ni de filtre. Par la suite il se comporte comme-ci le voyageur allait cliquer sur chaque annonce et copier-coller toutes les informations s'y trouvant dans un tableur au format CSV. Le programme va directement chercher les informations dans le script des différentes pages internet analysées.

Le programme utilisé ici pour effectuer le *scraping* a été adapté d'un programme existant mis à disposition sur Github⁶. C'est un programme codé sous Python⁷ qui utilise le module scrapy⁸. Ce programme et cette méthode sont régulièrement utilisés par les *Data Scientists*. Un projet utilisant ce programme est présenté par le blog de *NYC Data Science Academy*. Dans ce projet ils ont « scrapé » le site de Airbnb pour la ville

⁵ Zillow fourni des données open-source sur sa page dédiée : www.zillow.com/research/.

⁶ Github est site d'entraîne et de publication de code en open-source. Le programme qui a servi de base est celui mis à disposition par Luca Verginer. (https://github.com/verginer/bnb_scrapy_tutorial).

⁷ Python est un langage informatique permettant de coder.

⁸ Scrapy est un module qu'il est possible d'ajouter à Python pour effectuer du « web-scraping ».

d'Orlando et comparé les résultats avec des données « scrapées » du site de Trip-Advisor (NYC DataScience, 2016).

Plusieurs difficultés sont apparues lors de la récupération des données. Il a donc fallu contourner ces difficultés tout au long de la récupération des données. La première difficulté est venue du changement régulier de la structure du script du site internet de la part de Airbnb. En effet lors de la période de récupération des données les éléments du script des pages choisis lors des premières exécutions du programme ont été modifiés rendant ainsi inutilisable ou partiellement utilisable le programme. Il a donc fallu réécrire et réadapter le programme en fonction des modifications pour obtenir des données homogènes à chaque ronde de *scraping*. La deuxième difficulté est de pouvoir contourner la limitation du nombre d'annonces affichées pour chaque recherche qui est de 306 résultats. La technique mis en place ici, est une technique de réduction d'aire de recherche. Les premières rondes de récupération de données se sont faites au niveau de la ville sélectionnée, puis pour les différents codes postaux. Si au niveau d'un code postal le nombre d'annonces était supérieur à 300 alors un choix d'adresse pour séparer la zone était effectué. Une dernière ronde de *scraping* était alors faite avec ces différentes adresses. La dernière difficulté rencontrée est l'obtention d'un maximum de données homogènes entre elles. Les différents hôtes remplissent eux-mêmes, de manière plus ou moins assidue les informations présentées sur la fiche descriptive de l'annonce. Si un utilisateur ne remplit pas ces informations variées (précédemment détaillées) on ne peut tout simplement pas y accéder. Dans un souci d'homogénéisation maximale nous avons concentré nos efforts sur les données principales suivantes : le type de logement, le nombre de chambres, la date d'inscription et le lieu du logement. Ces données étant celles de bases caractérisant une annonce, nous avons pu nous affranchir du problème précédent. Le processus de *web-scraping*⁹ s'est effectué d'Avril à Août 2017.

Ces données ainsi récupérées nous permettent de segmenter par type de biens loués, l'entrée sur le marché de Airbnb. Cette segmentation est la même que celle utilisée par Zillow (données décrites plus bas). La segmentation ainsi possible est celui d'une zone

⁹ Un exemple du code utilisé pour le web-scraping du code postal 93101 (Santa-Barbara) est disponible sur : <https://www.dropbox.com/sh/c2kuzqd74cnxk8q/AAD3j4oFtWv57QOdE4nC9kJTa?dl=0>.

géographique par ville, et une caractérisation du bien immobilier par le nombre de chambres.

Les données fournies par le site de Airbnb nous permettent d'obtenir de l'information sur les caractéristiques du logement mis sur le marché de la location temporaire ainsi que son prix. Le but de la collecte de données sur le site de Airbnb est d'obtenir de l'information sur les biens immobiliers et leurs dates de mise en location.

Les caractéristiques spatiales et physiques des biens permettent de les comparer aux biens immobiliers mis en vente. Pour les différentes annonces présentes sur le site nous avons pu collecter le nom de la ville dans laquelle se trouve le logement ainsi que le nombre de chambres que comporte le logement.

La date de mise sur le marché nous permet quant à elle d'avoir une information sur l'extension de Airbnb dans certaines villes. Pour cette date d'entrée sur le marché de la location temporaire, la date d'inscription du loueur (hôte Airbnb) sera utilisée. Il est évident que cela revient à faire une approximation car en utilisant cette date nous ne pouvons pas prendre en compte par exemple le fait que l'hôte déménage et reste inscrit, que l'hôte retire son bien temporairement ou qu'un hôte ce soit désinscrit avant la date de collecte des données. Nous reviendrons sur ces éléments au cours de la discussion à la fin de notre étude.

3.2 Les données de Zillow

Zillow est une *start-up* qui offre via son site internet et son application mobile la possibilité d'obtenir de l'information sur le marché immobilier de manière géographique. L'intérêt de ce site pour notre étude est la partie *open-source* du site qui offre différentes données sur le marché immobilier à travers le temps. Nous avons été obligés de nous tourner vers cette ressource et donc vers une étude se trouvant sur le territoire des États-Unis car l'accès aux données immobilières dans bon nombre de pays occidentaux est soit confidentiel ou soit trop onéreux. Parfois les seules données du marché immobilier qui sont accessibles gratuitement sont d'un niveau trop agrégé.

Le site de Zillow enregistre les différentes ventes immobilières ainsi que les prix des loyers suivant les caractéristiques de bases du logement et sa localisation. Ces données sont en *open-source*. Le site crée à l'aide de sa base de données un estimateur, le *Zillow Index*.

Les équipes de Zillow ont mis en place le calcul d'un indice permettant d'avoir une estimation du prix des biens immobiliers de manière sectorielle. Cette volonté de créer un nouvel indice vient du constat que l'indice de prix de vente médian est vulnérable à des biais. Comme par exemple un nombre disproportionné de ventes de biens immobiliers dispendieux par rapport à un nombre de biens immobilier moins cher pendant une année va entraîner une appréciation du prix de l'immobilier alors que le prix des maisons n'aura pas forcément varié d'une année à l'autre.

C'est pour éviter ce biais qu'un indice idéal doit caractériser le prix de ventes des biens immobiliers en fonction de leurs caractéristiques. D'où l'utilisation d'un indice de prix permettant de prendre en compte la distinction entre les différentes gammes de maison pour chaque période temporelle. Cela revient en quelque sorte à utiliser une méthode de panier de biens comme pour l'indice des prix à la consommation. Une chose aisée qu'il est possible de faire avec des biens de consommation qui sont utilisés à chaque période de temps devient plus compliquée avec des types de biens immobiliers qui ne sont pas vendus sur toutes les périodes.

Depuis 2005 Zillow a donc développé l'indice *Zestimates* pour tenter d'approximer cet indice de prix de l'immobilier idéal. Le *Zestimates* estime le prix de vente non seulement des biens qui ont été vendus mais il estime les prix des biens non vendus sur cette période. Chaque *Zillow Home Value Index (ZHVI)* est une série temporelle mensuelle donnant une évaluation médiane des logements pour une zone géographique particulière. Dans la plupart des cas les ZHVI débutent en Avril 1996 et se trouvent à 7 niveaux géographiques différents : quartier, code postal, ville, district, comté, agglomération, état, et pays.

La construction des *Zestimates* se fait par technique d'apprentissage machine basée sur des propriétés statistiques. Cette technique prend en compte l'emplacement du bien,

les différentes transactions immobilières effectuées et les caractéristiques du bien pour ensuite pouvoir par apprentissage estimer les biens qui ne sont pas encore vendus.

Le but de l'estimateur est de donner aux consommateurs une estimation de leur domicile ainsi qu'une tendance du marché. À l'intérieur de chaque zone géographique l'indice ZHVI est calculé pour différents types de biens et permet d'avoir une vision segmentée du marché. Seules les propriétés résidentielles sont prises en compte (Zillow, 2014).

3.3 Les données Macro-économiques

Les données macro-économiques que nous avons utilisées proviennent de différentes sources gouvernementales américaines. Au niveau des institutions fédérales nos deux sources sont l'*US-Census* et la *Federal Housing Finance Agency* (FHFA). Au niveau de l'état notre source est l'office du tourisme Californien (*VisitCalifornia*) et la base de données fournie par le rapport de Dean Runyan¹⁰.

Les données sélectionnées sont le taux de chômage, la démographie, le salaire médian des ménages, les dépenses touristiques faites dans les hébergements et dans les hôtels, l'inflation des dépenses pour l'immobilier, le taux d'occupation des logements et le taux d'intérêt effectif pour un prêt immobilier. Ces différents indicateurs nous permettent d'avoir des informations macros pouvant affecter le marché immobilier et pouvant être utile pour les décisions des investisseurs comme cela a été identifié dans la revue de littérature. Ainsi on retrouve le taux de chômage, le revenu médian et la démographie pour caractériser la demande. L'inflation des dépenses pour l'immobilier donne l'évolution des dépenses courantes d'un propriétaire, le taux d'intérêt permet le calcul du TRI pour l'investisseur et le taux d'occupation résume la part de logement déjà occupé. Ces deux indicateurs caractérisent l'offre. Étant dans une zone touristique, nous utilisons les dépenses dans les hôtels et les hébergements touristiques pour caractériser l'afflux d'une population supplémentaire dans les villes sélectionnées. Nos variables

¹⁰ Dean Runyan Associates est un cabinet d'études et de conseil qui a travaillé sur l'industrie du tourisme en Californie : <http://www.deanrunyan.com/CATravelImpacts/CATravelImpacts.html>.

macroéconomiques interviennent sur trois axes. Le premier porte sur l'économie et la démographie de la zone géographique étudiée. Le second porte sur le marché immobilier et le dernier sur la fréquentation touristique.

Le détail de chacun des indicateurs sont repris dans le Tableau 3.1. Ce tableau permet de voir la fréquence de l'indicateur et son niveau géographique.

Indicateurs	Taux de chômage	Démographie	Salaire médian des propriétaires	Dépenses touristiques d'hébergements
Fréquence	Mensuelle	Annuelle	Annuelle	Annuelle
Niveau Géographique	Ville	Ville	Ville	Comté
Source	US-Census	US-Census	US-Census	VisitCalifornia
Nom	UNrate	Population	Median_Household_Income	Spending_Accommodation
Indicateurs	Dépenses touristiques d'hôtel	Inflation des dépenses immobilières	Taux d'occupation des logements	Taux d'intérêt effectif
Fréquence	Annuelle	Mensuelle	Annuelle	Mensuelle
Niveau Géographique	Comté	Pays	Ville	Pays
Source	VisitCalifornia	US-Census	US-Census	FHFA
Nom	Spending_Hotel	SAH1	Occupied_Housing_Units	Effective_Interest_Rate

Tableau 3.1 : Description des variables explicatives utilisées (autres que les annonces d'Airbnb).

Chapitre 4

Méthodologie

Deux différentes méthodes ont été utilisées. La première a pour but d'obtenir une estimation de l'impact de Airbnb sur le marché immobilier. La seconde a pour but d'utiliser des méthodes prédictives d'apprentissage statistique pour prédire le prix d'une nuit et montrer l'étendue des données récoltées et les différentes perspectives que peuvent apporter l'utilisation du *web-scraping*.

4.1 Méthodologie pour l'estimation d'impact

4.1.1 Préambule

Le préambule à la construction de nos modèles est la compilation des données pour obtenir un jeu de données homogènes et en forme de données de panel. Le choix s'est porté sur des villes proches de la côte. Villes pour lesquels nous avons pu avoir à la fois les données provenant de Airbnb et des différents sites gouvernementaux.

Ainsi la sélection que nous avons faite comporte douze villes se trouvant sur deux comtés différents, celui de Santa-Barbara et San Luis Obispo. Les villes sélectionnées sont les suivantes : Arroyo-Grande, Cambria, Carpinteria, Cayucos, Goleta, Grover Beach, Lompoc, Montecito, Morro-Bay, Oceano, Pismo Beach et Santa-Barbara. Malgré nos efforts pour sélectionner des villes ayant le maximum de données macro-économiques, nous avons des différences d'homogénéités. Certaines villes n'ayant pas les informations nécessaires pour couvrir toute la période souhaitée s'étendant jusqu'à fin 2016. Aller jusqu'à fin 2016 nous semblait important car l'essor de Airbnb dans ces villes a débuté fin 2015 et s'est poursuivi en 2016. Cet essor est sans doute lié à une démocratisation de l'utilisation de la plateforme et également à des réglementations inefficaces pour empêcher son utilisation. Pour pouvoir utiliser tous les *Listings* que nous avons récupérés, nous avons décidé de dupliquer les données macro-économiques de 2015 pour 2016 pour combler le manque d'information de l'année 2016. Cette approximation étant forte au point de vue macro-économique nous avons choisi d'utiliser

nos modèles sur deux périodes. La première période est celle allant de Janvier 2010 à Décembre 2015. La seconde période débute en Janvier 2010 et s'étant jusqu'à Décembre 2016.

À partir des données provenant de Airbnb une variable de nombre d'annonces publiées a été créée. Cette variable correspond à une valeur cumulative pour chaque mois. Elle est basée sur la date d'inscription de l'hôte. Nous considérons ici que la date d'inscription de l'hôte est identique à la date de mise à disposition du logement sur le site. Cette approximation sur la date est liée au « *scraping* », c'est la seule date que nous pouvons récupérer.

Par exemple si pour la ville de Cambria, deux annonces ont leurs hôtes qui se sont inscrit au mois de Janvier 2010 alors la valeur de la variable cumulative sera deux. Puis si pour le mois de Février 2010, trois annonces correspondent à une date d'inscription d'hôtes alors la cumulative pour ce mois sera cinq. Ces données appelées *Listings* ont été discriminées suivant le type de logement et le nombre de chambres pour obtenir la même segmentation que celle de l'indice de Zillow. Un résumé des différentes données cumulatives se trouve dans le Tableau 4.1. Néanmoins la segmentation ne peut pas se faire sur toutes les villes. L'indice de Zillow est calculé s'il y a eu assez de transactions par villes et par type de biens. Or pour les segmentations se situant aux extrémités, c'est-à-dire les logements d'une chambre et de cinq chambres et plus, les groupes d'études ne comportent pas la totalité des villes. Ainsi pour les logements ayant une seule les villes de Carpinteria, Cayucos, Grover Beach, Montecito, Lompoc, Oceano et Santa-Barbara ne sont pas prises en compte. Pour les logements à deux chambres, seule la ville de Cayucos n'a pas de données provenant de Zillow. Enfin pour les logements comportant cinq chambres et plus les villes de Carpinteria, Cayucos, Grover-Beach et Oceano ne présentent pas d'indice Zillow pour ce segment. Pour résumer les différents groupements de ville, le segment d'une chambre porte sur cinq villes, le segment de deux chambres sur onze villes, les segments de trois et quatre chambres sur douze villes et le segment de cinq chambres et plus sur huit villes.

BnB_All	Reprend toutes les annonces sans discrimination
BnB_Entire	Les annonces sont celles des logements entiers
BnB_1bdr	Annonces des logements entiers avec 1 chambre
BnB_2bdr	Annonces des logements entiers avec 2 chambres
BnB_3bdr	Annonces des logements entiers avec 3 chambres
BnB_4bdr	Annonces des logements entiers avec 4 chambres
BnB_5bdr	Annonces des logements entiers avec 5 chambres et plus

Tableau 4.1 : Description de la nomenclature utilisées pour les différentes segmentations des variables cumulatives des annonces d’Airbnb.

4.1.2 Panel annuel

La première partie de la modélisation porte sur les données annuelles. Le modèle de régression choisie est une régression Log-Linéaire avec effets fixes. On entend par Log-Linéaire que la variable d’indice des prix de Zillow utilisée sera en logarithme et que la variable explicative dont on cherche à déterminer l’impact, le « listing Airbnb » est linéaire. La motivation d’utiliser une régression Log-Linéaire vient premièrement du fait qu’en phase exploratoire des données, une régression Log-Log au niveau annuel donnait des résultats moins significatifs que le Log-Linéaire. Deuxièmement, l’indice de Zillow est une variable continue alors que notre variable d’annonces Airbnb est une cumulative de variables discrètes. Le choix de l’utilisation des effets fixes provient du fait que nous ne pouvons pas garantir l’indépendance entre les variables explicatives et le paramètre invariant temporellement qui caractérise chaque ville. En effet de nombreux paramètres fixes à la ville peuvent être corrélés avec les variables explicatives. Par exemple la réputation ou le charme de la ville, les politiques établies, les infrastructures sont des variables qui peuvent être préétablies et fixes durant notre étude mais qui peuvent avoir une corrélation avec la démographie, le nombre de touristes (donc les dépenses touristiques et le nombres d’annonces Airbnb), le taux de chômage et les salaires. À cause de cette intuition nous avons écarté le choix des effets aléatoires car l’hypothèse de non corrélation serait trop forte.

Pour les données annuelles nous utiliserons les modèles 1 et 2 qui sont respectivement décrits par les équations 1 et 2 ci-dessous :

$$\text{Équation 1: } \log(ZHIV)_{it} = \beta_0 + \beta_1.BnB_All_{it} + \beta_2.UN_rate_{it} + C_i + \varepsilon_{it}$$

$$\text{Équation 2: } \log(ZHIV)_{it} = \beta_0 + \beta_1.BnB_All_{it} + \beta_2.UN_rate_{it} + \beta_3.X_{it} + C_i + \varepsilon_{it}$$

Avec

$$X_{it} = (\log(population)_{it}, \log(Median_Household_Income)_{it}, \\ \log(Spending_Accommodations)_{it}, \\ \log(Spending_Hotel)_{it}, SAH1_{it}, Occupied_Housing_Units_{it}, \\ Effective_Interest_rate_t)$$

Le terme X_{it} correspond aux variables explicatives macro-économiques décrites dans la partie sur les données. Ainsi la différence entre le Modèle 1 et le Modèle 2 réside dans l'ajout de ces variables. Pour les résultats les noms de ces variables se trouvent dans l'annexe A avec les résultats. Le terme C_i représente toutes les caractéristiques de la ville i (observées ou non) qui ont un impact sur le prix de l'immobilier.

Compte tenu de l'approximation pour la dernière année pour certaines données macro-économiques dont nous avons discuté précédemment, nous allons faire l'étude sur deux périodes. La première allant de 2010 à 2015 et l'autre de 2010 à 2016.

Pour vérifier statistiquement le choix des effets fixes nous avons d'abord vérifié la significativité conjointe des coefficients, la présence d'effets individuels puis effectué un Test d'Hausman pour les modèles choisis. Le premier F-test pour la significativité conjointe des coefficients a pour hypothèse nulle que tous les coefficients sont nuls, pour chacun des modèles on peut rejeter cette hypothèse comme le montrent les résultats du tableau 4.2. Le F-test effectué pour vérifier la présence d'effets individuels a pour hypothèse nulle qu'il n'y a aucun effet individuel. Les résultats de ce test sont présentés dans le tableau 4.2. Ces résultats nous permettent de rejeter l'hypothèse nulle et d'effectuer le test d'Hausman. L'hypothèse nulle du Test d'Hausman est qu'il n'y a pas de différence systématique entre les coefficients des deux modèles. L'hypothèse alternative est qu'il y a une différence systématique entre les coefficients. Les résultats sont fournis dans le tableau 4.2.

Nom du Test	Période 2010 – 2015		Période 2010 – 2016	
	Modèle 1 P-value	Modèle 2 P-value	Modèle 1 P-value	Modèle 2 P-value
Coefficients nuls	0.0000	0.0000	0.0000	0.0000
Effets Individuels	0.0000	0.0000	0.0000	0.0000
Hausman	0.6004	0.0000	0.6644	0.0000
Hétéroscédasticité	0.0000	0.0000	0.0000	0.0000

Tableau 4.2 : Résultats des tests pour les deux différents modèles sur les deux périodes étudiées annuellement.

Ainsi lorsque l'on ajoute les variables macro-économiques, le modèle 2 à effets fixes convient. Lorsque le modèle est simplifié avec seulement le taux de chômage et la cumulative des annonces Airbnb (modèle 1), il faut justifier le choix des effets fixes comme nous l'avons fait au paragraphe précédent. Un test de présence d'hétéroscédasticité a également été effectué pour savoir si les écarts-types devaient être corrigés. Les résultats du test se trouvent également dans le tableau 4.2. L'hypothèse nulle du test de Breusch-Pagan est l'homoscédasticité. Les résultats nous montrent qu'il y a présence d'hétéroscédasticité pour les deux modèles et pour les deux périodes ; nous en prendrons compte lors des régressions. Il est également possible d'avoir la présence d'endogénéité. En effet le nombre d'inscriptions sur Airbnb peut être également corrélé au résidu. Néanmoins par construction les effets fixes prennent en partie compte de cette endogénéité car ils sont corrélés avec les termes d'erreurs, ce qui n'est pas le cas dans un modèle des effets aléatoires. Nous aurions pu tenter de trouver une variable instrumentale pour pallier cet effet. L'une des possibilités est l'utilisation d'une variable géographique comme la distance par rapport au centre-ville ou à un lieu touristique. Cette approche est décrite par Cavailhès pour les attributs des prix des logements (Cavailhès, 2005) . Mais nous avons souhaité nous focaliser sur l'aspect de corrélation de la relation établie par les modèles étudiés.

Le but des modèles annuels est de confirmer l'utilisation des différentes variables macroéconomiques. Celles-ci nous étant pour la plupart pas accessibles à une fréquence mensuelle.

4.1.3 Panel mensuel

La seconde partie de la modélisation porte sur les données mensuelles. Compte tenu du nombre d'observations plus grand à ce niveau, les données de panel mensuelles seront privilégiées pour nos interprétations. Nous reprenons le même type de régression Log-Linéaire avec effets fixes que celui fait sur les données annuelles. Pour les données mensuelles nous reprenons les deux modèles précédemment décrits (Équations 1 et 2) auxquels nous ajouterons des variables dichotomiques annuelles ainsi que des variables dichotomiques mensuelles (*Time_trend*). Ces deux modèles supplémentaires sont les modèles 3 et 4 et ils sont décrits par les équations 3 et 4.

Équation 3:

$$\log(ZHIV)_{it} = \beta_0 + \beta_1.BnB_All_{it} + \beta_2.UN_rate_{it} + \sum_{j=Year} \beta_j.d_j + \beta_4.Time_trend_{it} + C_i + \varepsilon_{it}$$

Équation 4:

$$\log(ZHIV)_{it} = \beta_0 + \beta_1.BnB_All_{it} + \beta_2.UN_rate_{it} + \sum_{j=Year} \beta_j.d_j + \beta_4.Time_trend_{it} + \beta_5.X_{it} + C_i + \varepsilon_{it}$$

Les variables macro-économiques du modèle annuelle sont utilisées ici par simple duplication. La valeur que nous avons au niveau annuel est la même pour chaque mois de la même année hormis pour la démographie qui a été itérée linéairement.

De la même manière que pour les données annuelles nous avons vérifié la non nullité de tous les coefficients, la présence d'effets individuels puis effectué un Test d'Hausman pour les modèles choisis. Les résultats de ces tests sont présentés dans le tableau 4.3. Ces résultats nous permettent de rejeter l'hypothèse nulle pour les deux F-Tests, il y a donc présence d'effets individuels et les coefficients ne sont pas tous nuls. Les résultats du test d'Hausman sont présentés dans le tableau 4.3. D'après ces résultats pour les modèles 2 et 4 pour les deux périodes, le choix des effets fixes convient. Pour les modèles 1 et 3 il faut justifier ce choix, la justification est la même que précédemment. Nous avons également

effectué un test de Breusch-Pagan pour déterminer la présence d'hétéroscédasticité. Les résultats du tableau 4.3 montrent clairement la présence d'hétéroscédasticité pour les quatre modèles et pour les deux périodes ; nous en prendrons également compte lors des régressions. De la même manière que précédemment pour ces modèles la présence d'endogénéité n'est pas à écarter mais nous faisons le même raisonnement que pour les données annuelles.

		Résultats des P-value des différents tests							
		Période 2010 – 2015				Période 2010 – 2016			
TEST	Modèle	1	3	2	4	1	3	2	4
		Coefficients nuls	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Effets Individuels	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Hausman	0.468	0.274	0.000	0.000	0.538	0.308	0.000	0.000
	Hétéroscédasticité	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Tableau 4.3 : Résultats des tests pour les quatre différents modèles sur les deux périodes étudiées mensuellement.

De la même manière et pour les mêmes raisons que les données annuelles, nous utiliserons deux périodes dans notre étude, la première allant de Janvier 2010 à Décembre 2015 et l'autre de Janvier 2010 à Décembre 2016.

Dans notre étude nous segmenterons par type de biens en fonction du nombre de chambres disponibles dans le logement. Nos modèles seront initialement utilisés avec tous les types d'annonces Airbnb et tous les types de logements, puis nous discriminerons par segment. La discrimination se fait d'abord au niveau des logements entiers puis par nombre de chambres se trouvant dans ces logements. On aura ainsi une segmentation par une, deux, trois, quatre et plus de cinq chambres. Cette segmentation permettra d'obtenir une idée de quels sont les types de logements les plus impactés par l'entrée d'Airbnb.

4.2 Méthodologie pour prédiction prix

4.2.1 Méthodes utilisées

Un certain nombre de données récoltées par *web-scraping* nous permet d'avoir un grand nombre d'informations sur les annonces et nous amène à penser qu'il serait possible de déterminer les facteurs influençant le prix d'une nuit sur Airbnb. Nous proposons d'utiliser trois méthodes d'apprentissage statistique (*Machine Learning*).

La première approche est celle par la méthode de *Least Absolute Shrinkage and Selection Operator* (LASSO). La méthode de LASSO consiste en une régression linéaire que l'on pourrait qualifier de fourre-tout. En effet cette technique permet de ne pas faire de sélection préalable des variables explicatives. À l'aide d'un critère limite, le modèle choisit de lui-même les variables explicatives utiles. Les variables considérées non-utiles se voient attribuer un coefficient estimé de zéro. Ainsi la méthode de LASSO réduit le nombre de variables explicatives, la méthode rétrécit le modèle (*Shrinkable*). Comme pour une régression linéaire la méthode de LASSO minimise la somme des carrés des résidus, la différence est que cette méthode le fait avec une contrainte (Équation 5). Dans cette équation λ est le paramètre de régularisation, les y les variables expliquées, les x les variables explicatives et les β sont les coefficients de ces variables. Nous allons commencer par faire une régression de LASSO sur l'ensemble des données puis effectuer une séparation de l'échantillon afin d'effectuer une technique de *Machine Learning*.

Équation 5: Écriture sous forme Lagrangienne de la régression LASSO.

$$\min_{\beta_0, \beta_1, \dots, \beta_k} \frac{1}{2} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^k |\beta_j|$$

La seconde méthode de *Machine Learning* mise en place est la méthode d'arbres de décision (*Tree*). Cette méthode permet d'estimer de manière automatique les variables qui impactent le plus sur le prix de la nuit d'une annonce d'Airbnb. Le rendu se présente sous la forme d'un schéma ressemblant à un arbre généalogique. Au sommet de l'arbre se trouve le facteur le plus déterminant pour la variable à expliquer puis en descendant le long des branches se succèdent les autres variables explicatives par ordre d'importance.

L'arbre ainsi trouvé peut être émondé, c'est-à-dire simplifié en moins de branches. La technique par minimum de déviance sera utilisée dans notre étude pour émonder l'arbre.

La troisième méthode utilisée est celle de la forêt d'arbres (*Random Forest*). Cette technique consiste à faire des tirages aléatoires de N échantillons et de recommencer la méthode des arbres à chaque fois. Ainsi au lieu de n'avoir qu'un seul arbre nous obtenons N arbres, autrement dit une forêt. Cette méthode se fait en boucle et est automatisée. Elle permet de réduire la variance en comparaison à celle obtenue avec un seul arbre.

4.2.2 Les jeux de données

Les données utilisées pour effectuer cette prédiction du prix de la nuit Airbnb proviennent du *web-scraping* décrit précédemment. Cette méthode de collecte s'heurte à la non homogénéité du remplissage de certains équipements compris dans les logements mis à la location. Ainsi on a une forte présence de données vides pour bon nombre d'annonces. Or pour effectuer ces techniques avec le logiciel R nous devons supprimer les annonces ayant des valeurs omises. Au vu de la diminution du nombre d'annonces qu'occasionne un retrait systématique de ces valeurs, nous avons donc décidé de faire trois jeux de données. Le premier jeu de données reprend les données avec les valeurs omises et il sera traité tel quel par le programme R, il sera nommé jeu de données 1. Le deuxième jeu de données comporte pour modification le remplacement d'une valeur omise par une valeur nulle dans les variables dichotomiques des différents équipements, ce sera le jeu de données 2. Ainsi si nous n'avons pas pu récupérer l'information sur la présence ou non de l'équipement, nous considérons qu'il n'est pas présent dans le logement. Enfin le dernier jeu de données est celui duquel nous avons retiré les caractéristiques TV et TV par câble. Ces deux variables dichotomiques sont composées de la majorité des valeurs omises, nous l'appellerons par la suite jeu de données 3. L'ensemble des techniques décrites plus haut seront effectuées sur les trois jeux de données.

Le fichier initial commun à l'ensemble des trois jeux de données décrits précédemment est composé de 924 observations et 70 variables. Nous utilisons que 70 variables car une partie des équipements possiblement offert dans les logements n'étaient pas présent pour nos annonces. La majorité de ces équipements que nous avons

directement exclus par leur non-présence systématique concernent en grand nombre les différents aménagements pour personnes handicapées ou à mobilité réduite. Cette non-présence se justifie certainement par le fait que ces caractéristiques sont nouvelles et extrêmement récentes. Elles s'incorporent dans une nouvelle stratégie d'Airbnb de diversifier encore plus l'accès aux logements présents sur la plateforme aux personnes à mobilité réduite. Cette stratégie s'est accélérée avec le rachat de la *start-up* Accomable (Techcrunch, 2017)).

Chapitre 5

Résultats et Interprétations

5.1 Résultats du panel annuel

Les résultats sur les données annuelles pour les deux périodes (jusqu'à fin 2015 et fin 2016) nous permettent de nous conforter dans le choix du Log-linéaire. En effet lorsqu'on prend les résultats du modèle 1 pour toutes les annonces disponibles, les coefficients estimés pour les annonces d'Airbnb et pour le taux de chômage sont significatifs. Ces résultats sont présentés dans l'annexe A.

Lorsque les variables explicatives macro-économiques sont ajoutées nous avons perte de la significativité des coefficients estimés des annonces d'Airbnb et du taux chômage pour les deux périodes. Seules parmi ces variables semblent être significatifs la démographie, les dépenses touristiques dans les hébergements, l'inflation spécifique aux dépenses dans l'immobilier et le taux d'intérêt effectif de l'emprunt pour la période allant jusqu'à fin 2015. Pour celle allant jusqu'à fin 2016, le coefficient estimé pour la démographie n'est pas significatif.

Ces résultats nous surprennent d'une part, par le fait que les coefficients des variables explicatives du 1^{er} modèle perdent totalement leur significativité avec le second et d'autre part par la non significativité des coefficients de la plupart des variables explicatives ajoutées dans le deuxième modèle. Or le choix de l'ajout de ces variables s'était appuyé sur la théorie et l'intuition autour des paramètres pouvant impacter le prix de l'immobilier. Néanmoins l'effet de ces variables a rendu non significatif notre variable d'intérêt qui est le *listing* d'annonce Airbnb, nous décidons de les garder pour les études mensuelles avec les modèles décrits plus haut.

5.2 Résultats du panel mensuel

Effectivement le choix d'avoir gardé les variables explicatives du modèle 2 au niveau mensuel semble judicieux. Les résultats de l'étude du panel mensuel nous montrent que ces variables voient leurs coefficients estimés devenir significatifs suivant les modèles utilisés et les segmentations choisies. Ces résultats sont présentés en partie dans l'annexe B et au complet sur ce lien¹¹.

5.2.1 Les différences entre les modèles sur les deux périodes

Pour comprendre les différences de résultats entre les deux périodes étudiées, intéressons-nous à la pénétration du marché par Airbnb que l'on peut visualiser sur les figures 5.1 et 5.2. La première nous indique clairement que la moitié des annonces se trouvant sur Airbnb sont dans la ville de Santa Barbara. En analysant la cumulative des annonces d'Airbnb pour les autres villes (sans Santa-Barbara) on s'aperçoit que l'augmentation du nombre d'annonces n'est pas linéaire et que pour certaines villes il y a clairement une rupture de pente correspondant à une forte apparition d'annonces à la fin de 2015 et au début de 2016. Cette forte augmentation se situant entre les deux périodes étudiées pourrait expliquer la différence en termes d'estimation des coefficients de la cumulative d'annonces Airbnb.

De plus la ville de Santa Barbara possédant un grand poids dans le compte total des annonces (figure 5.1) et nos modèles de régressions se faisant en niveau pour le nombre de listings, il se pourrait que des problèmes d'échelle apparaissent dans nos régressions. La possibilité d'avoir un tel effet nous a fait créer deux panels. Le premier panel sera avec la ville de Santa-Barbara (la nomenclature « avec SB ») et le second panel sera celui où l'on aura retiré la ville de Santa-Barbara (la nomenclature « sans SB »). Nous allons donc par la suite utiliser les différents modèles pour les différentes segmentations sur ces deux panels. De cette méthodologie sur les données mensuelles, nous tirerons l'essentiel des résultats et des interprétations de notre étude.

¹¹ Lien dropbox pour obtenir l'ensemble des résultats mensuels:
<https://www.dropbox.com/s/ijrfinrwnrsxygk/annexe%20A%20et%20B%20memoire.pdf?dl=0>

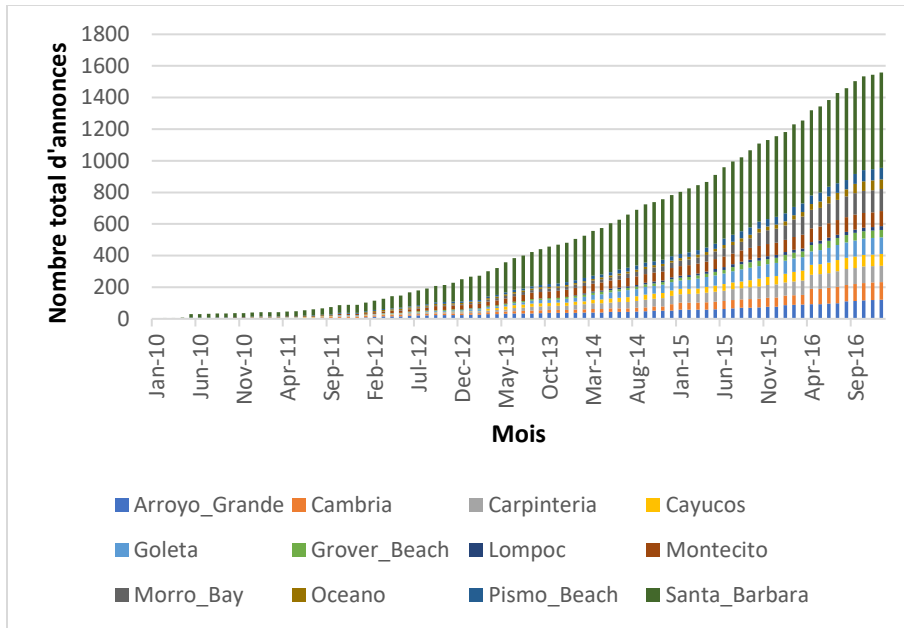


Figure 5.1 : Total de toutes les annonces d'Airbnb par mois et par ville.

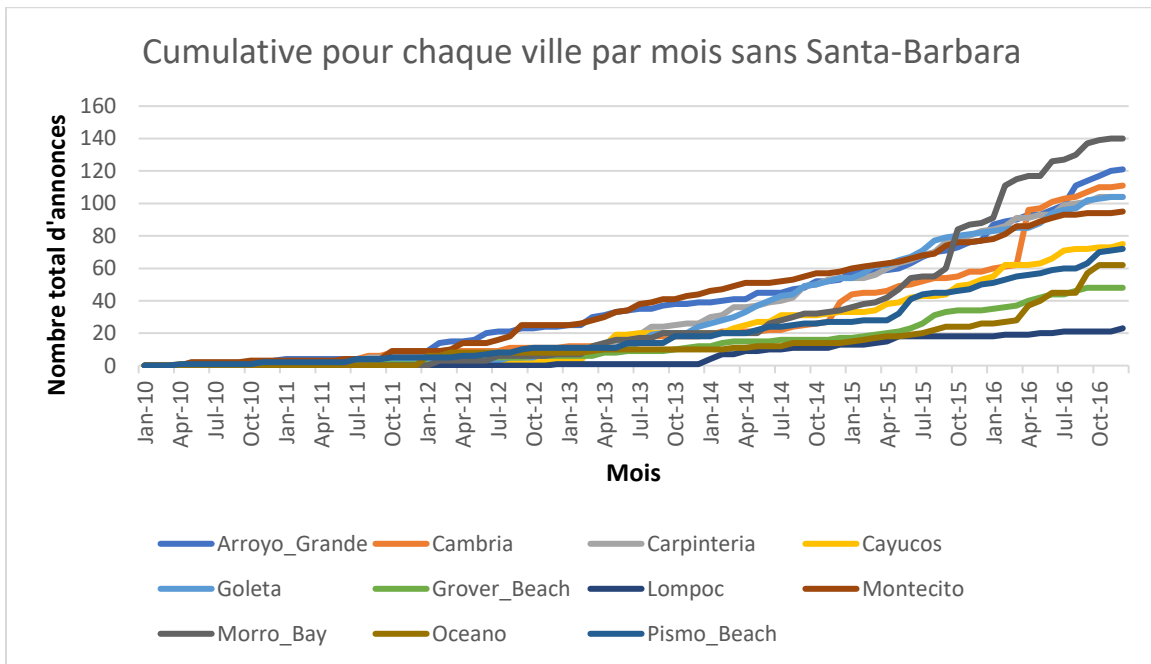


Figure 5.2 : Cumulative de toutes les annonces d'Airbnb pour chaque ville par mois sans Santa-Barbara.

Les estimations des coefficients des *Listings* d'Airbnb faites avec un *cluster* sur les villes et pour les deux périodes Janvier 2010 à Décembre 2015 et de Janvier 2010 à Décembre 2016 sont récapitulées dans le Tableau 5.1. Nous allons procéder à une reprise des résultats par segmentation pour identifier les similitudes et les différences entre les deux périodes.

- Sans segmentation et avec la segmentation logement entier

Tout d'abord sans effectuer aucune différenciation dans le type d'annonces d'Airbnb on s'aperçoit que pour les deux périodes (allant jusqu'à fin 2015 et fin 2016) seuls les modèles 1 et 2 ont des estimations significatives pour ces coefficients. Autrement dit l'ajout d'une tendance temporelle et de variables dichotomiques annuelles rend ces coefficients non significatifs sans la différenciation du type d'annonces.

On retrouve les mêmes résultats en utilisant une différenciation qui permet de garder seulement les annonces de logements entiers. Garder les logements entiers pourrait correspondre en partie à des propriétaires qui ont sorti leur bien immobilier du marché de la location long-terme.

On remarque également que les valeurs significatives des modèles 1 et 2 entre les deux périodes sont du même ordre et quasi-identiques. De plus l'ajout des variables macroéconomiques fait environ diminuer par deux les coefficients estimés des *Listings* Airbnb.

Modèle		1	2	3	4		
All	2015	0.000455** (0.000194)	0.000250** (0.000105)	0.0000926 (0.0000617)	0.0000760 (0.0000468)		
		864 0.611	864 0.818	864 0.871	864 0.917		
		0.000428** (0.000179)	0.000216** (0.0000915)	0.0000266 (0.0000403)	0.0000318 (0.0000374)		
	2016	1008 0.703	936 0.842	1008 0.906	936 0.927		
		Entire Home	2015	0.000761* (0.000348)	0.000435** (0.000196)	0.000188 (0.000133)	0.000160 (0.0000983)
				864 0.615	864 0.820	864 0.872	864 0.917
2016	0.000714** (0.000306)			0.000382** (0.000169)	0.0000732 (0.0000754)	0.0000798 (0.0000662)	
	1008 0.708	936 0.843	1008 0.906	936 0.927			
	1 Bdr	2015	0.0203** (0.00714)	0.0123 (0.00586)	0.00882 (0.00496)	-0.00014 (0.00271)	
360 0.536			360 0.805	360 0.779	360 0.881		
2016			0.0178*** (0.00249)	0.0134** (0.00352)	0.0109* (0.00483)	0.00264 (0.00419)	
		420 0.706	390 0.854	420 0.854	390 0.898		
		2 Bdrs	2015	0.00222 (0.00155)	0.000669 (0.000777)	0.00000802 (0.000724)	-0.000145 (0.000566)
789 0.505				789 0.801	789 0.851	789 0.900	
2016	0.00227 (0.00138)			0.000681 (0.000811)	-0.000428 (0.000334)	-0.000486 (0.000297)	
	921 0.630		855 0.821	921 0.896	855 0.919		

Modèle		1	2	3	4
3 Bdrs	2015	0.00569** (0.00200)	0.00359*** (0.00105)	0.00222** (0.000899)	0.00200** (0.000647)
		864	864	864	864
		0.622	0.824	0.872	0.922
	2016	0.00417*** (0.00110)	0.00272*** (0.000679)	0.000821 (0.000526)	0.000937 (0.000548)
		1008	936	1008	936
		0.705	0.843	0.897	0.925
4 Bdrs	2015	0.0122*** (0.00256)	0.00889*** (0.00163)	0.00678*** (0.00148)	0.00654*** (0.00100)
		864	864	864	864
		0.667	0.830	0.845	0.894
	2016	0.0111*** (0.00144)	0.00764*** (0.00114)	0.00489** (0.00179)	0.00521*** (0.00146)
		1008	936	1008	936
		0.727	0.844	0.865	0.897
5 + Bdrs	2015	0.00707 (0.00516)	0.00374 (0.00319)	-0.00121 (0.00329)	-0.00168 (0.00190)
		545	545	545	545
		0.553	0.688	0.708	0.799
	2016	0.0101** (0.00380)	0.00579* (0.00266)	0.000917 (0.00176)	0.000036 (0.00201)
		641	593	641	593
		0.650	0.730	0.778	0.821

Tableau 5.1 : Variable dépendante : Indice de Zillow. Valeurs des coefficients estimés pour les annonces d'Airbnb par segment, leurs écarts types, le nombre d'observations et les R-carrés ajustés (option *cluster* par ville).

- Segmentation des logements entiers à une chambre

Lorsque l'on porte attention aux résultats de la segmentation du bien immobilier comportant une chambre, les résultats sont similaires pour les deux périodes seulement pour le modèle 1. Pour ce modèle les coefficients des *Listings* d'Airbnb d'un logement entier comportant une chambre sont significatifs, positifs et semblables pour les deux périodes. Pour la période allant jusqu'à fin 2015, ils ne sont pas significatifs dès lors que

nous introduisons les variables dichotomiques annuelles et la tendance temporelle ou les variables macro-économiques. Ces cas correspondent aux modèles 2, 3 et 4. Alors que pour la période allant jusqu'à fin 2016 les estimations de ces coefficients sont significatifs pour les modèles 2 et 3 mais ne sont pas significatifs lorsque l'on introduit simultanément les variables macros, la tendance temporelle et les variables dichotomiques annuelles. De plus l'ajout de ces variables fait diminuer la valeur des estimations mais de manière moins forte que pour le cas de tous les logements entiers et celui de toutes les annonces.

- Segmentation des logements entiers à deux chambres

Pour la segmentation des biens immobiliers composés de deux chambres les résultats sont semblables sur les deux périodes pour tous les modèles. Les *Listings* d'Airbnb n'ont aucun effet significatif. Ce segment est le seul segment pour lequel l'arrivée d'Airbnb dans les villes étudiées ne semble avoir aucun effet significatif sur le prix de l'immobilier. C'est également le segment pour lequel nous avons le plus d'annonces d'Airbnb néanmoins ce segment possède seulement une dizaine d'annonces de plus que le segment des trois chambres.

- Segmentations des logements entiers à trois chambres

Les segmentations pour les logements avec trois chambres, nous donnent des résultats similaires entre les deux périodes étudiées à l'exclusion des modèles 3 et 4 pour la période comprenant 2016. Les estimations des coefficients des listings d'Airbnb sont significatifs, positifs et du même ordre de grandeur. On perd la significativité pour les modèles 3 et 4 pour la période allant jusqu'à fin 2016.

De plus l'ajout des variables dichotomiques annuelles et de la tendance temporelle font diminuer les valeurs des coefficients estimés plus fortement que les variables macros.

- Segmentation des logements entiers à quatre chambres

Les segmentations pour les logements avec quatre chambres, nous donnent des résultats similaires pour chaque modèle entre les deux périodes étudiées. Les estimations des coefficients des *listings* d'Airbnb sont très significatives, positifs et similaires. On retrouve le même phénomène qu'avec les *Listings* des trois chambres, l'ajout des variables dichotomiques annuelles et la tendance temporelle font fortement diminuer la valeur des coefficients estimés.

- Segmentations des logements entiers à cinq chambres et plus

Lorsque l'on regarde les résultats pour les logements de cinq chambres et plus, pour la période allant jusqu'à 2015, aucun modèle ne donne d'estimations significatives pour les *listings* Airbnb. Alors que pour la même segmentation pour la période allant jusqu'à décembre 2016, les modèles 1 et 2 voient les coefficients estimés des cumulatives d'annonces Airbnb être significatifs et positifs.

De manière générale les coefficients estimés des *Listings* d'Airbnb pour les deux périodes et pour les différents modèles gardent les mêmes tendances comme le montre les résultats en annexe C. Les valeurs des coefficients estimés, lorsqu'ils sont significatifs pour les deux périodes et pour le même modèle, nous montrent une légère diminution avec l'ajout de l'année 2016.

5.2.2 L'ordre de grandeur de l'impact et des variables macro-économiques

En reprenant l'ensemble des résultats des régressions (pour tous les modèles, toutes les segmentations et les deux périodes), on s'aperçoit que les plus grandes valeurs estimées des coefficients des annonces d'Airbnb se trouvent être dans le modèle 1. Ces estimations sont particulièrement plus fortes pour les segments d'une chambre, quatre chambres et plus de cinq chambres. Les valeurs respectives de ces coefficients sont 0,0203 ; 0,0122 et

0,0101. Les deux premiers résultats proviennent de la période allant jusqu'à fin 2015 et le dernier jusqu'à fin 2016 (Tableau 5.1).

Ainsi le plus grand coefficient estimé tous modèles confondus et toutes périodes étudiées est 0,0203. Ce coefficient peut s'interpréter de la manière suivante : avec une annonce de plus, le prix d'un logement à une chambre augmenterait de 2,03% toutes choses étant égales par ailleurs. Ce résultat est à prendre avec précaution car dans le cas des logements à une chambre, le groupe de villes sélectionnées ne contient que cinq villes. Ces villes ne sont pas parmi les plus importantes en termes d'offre de logements et d'annonces. En effet la moyenne du nombre d'annonces pour les villes de ce groupe pour cette segmentation est de 15,2 annonces fin 2015 avec un maximum de 18 annonces et un minimum de 13 annonces. Si pour l'une de ces villes une nouvelle annonce apparaît, on aurait une augmentation d'environ 6,6% du nombre d'annonces pour cette ville. Cela nous permet donc de calculer un coefficient d'élasticité de 0,31. Or l'augmentation maximum du nombre d'annonces annuelle dans ces villes est de 9 annonces supplémentaires et l'augmentation moyenne est de 5,6 annonces pour ce segment sur l'année 2015. Ce qui veut dire l'apparition d'environ 1 nouvelle annonce tous les deux mois sur le site de Airbnb pour ce groupe. Avec ce rythme, la valeur du marché immobilier dans ce groupement aurait dû croître fortement.

En effet une augmentation de 2,3% du prix du logement lors de l'ajout d'une annonce, compte tenu du prix moyen de ce type de logement qui est d'environ 355 000 USD (figure 5.7), équivaut à une augmentation du prix moyen de ces logements de 8 165 USD. Or on voit plus de cinq annonces apparaître chaque année en moyenne. Ce qui fait une augmentation potentielle sur chaque bien de plus de 45 000 USD. On devrait donc voir une potentielle flambée des prix. De plus, si 22 transactions se font annuellement sur le marché immobilier, celui-ci devrait avoir une croissance annuelle d'environ un million de dollars.

Si l'on s'attarde aux résultats qui prennent Santa-Barbara en compte et la période allant jusqu'à fin 2016, la plus forte estimation est pour le segment des quatre chambres avec le modèle 1 soit 0,0111. Cela voudrait dire que pour une annonce supplémentaire apparaissant sur Airbnb le prix du logement à quatre chambres augmenterait de 1,11%

ceteris paribus (c.p.), de la même manière que précédemment, au vu du nombre moyen d'annonces et de ce type (6,25), nous aurons un coefficient d'élasticité de 0,07.

En comparaison avec les variables macro-économiques pour la segmentation des logements à une chambre, la démographie pourrait faire varier le prix d'un logement de plus de 2% lorsque celle-ci augmente de 1% (c.p.) dans le modèle 2. Cette augmentation de 1% de la démographie est souvent atteinte sur une année pour les plus grandes villes de notre échantillon et pour les villes se trouvant en leur périphérie. Par exemple dans nos données pour la ville de Cambria l'augmentation annuelle entre 2014 et 2015 est de 0,99% (US-Census). Il est également intéressant de voir que les dépenses touristiques dans les hôtels lorsqu'elles augmentent de 1% feraient augmenter le prix des logements à une chambre de 1,194% (c.p.) dans le modèle 4. Dans le même modèle les dépenses touristiques dans tous les hébergements en augmentant de 1% augmenteraient le prix de ces logements de 3,089%. L'augmentation de cette dépense entre 2015 et 2016 dans les comtés de Santa-Barbara et de San-Luis-Obispo sont respectivement de 5% et 1,2% (Dean Runyan Associates, 2017). Le revenu quant à lui, s'il augmentait de 1%, les prix des habitations à une chambre augmenteraient de 0,445% (c.p.) dans le modèle 4. Enfin le dernier paramètre macro-économique qui nous a attirer ici est le taux d'intérêt effectif pour les emprunts immobiliers. Si celui-ci variait de 0,1% alors le prix des logements varierait de 1,57% (c.p.) dans le modèle 2. (Annexe B)

5.2.3 Les distributions des logements et des annonces

Il est intéressant de mettre nos constatations en relief avec la distribution des logements présents dans les villes et la distribution du type de biens mis en location sur Airbnb. La distribution des logements correspond à la proportion de logements présents dans les groupes de villes sélectionnées par type de bien par rapport à l'ensemble des logements. Cette distribution donne une idée de comment est répartie l'offre de logements, elle est représentée sur la figure 5.3. La distribution des annonces Airbnb montre les proportions des annonces par type de biens par rapport au nombre d'annonces de logements entier dans les groupes de villes sélectionnées. La figure 5.3 nous montre comment sont répartis en moyenne les annonces et le type de logement dans l'ensemble

des douze villes, l'annexe D reprend le détail par ville. On s'aperçoit que les logements majoritairement présents dans les villes sélectionnées sont des biens avec deux ou trois chambres. À eux seul ce segment représente plus de 66% des logements présents. De plus les pourcentages des trois chambres et deux chambres sont très proches, moins de 1% d'écart. D'autre part lorsque l'on compare les nombres d'annonces totales sur la période Janvier 2010 à Décembre 2016, on constate que les annonces sont majoritairement pour des biens à une seule chambre. Ces annonces représentent environ un tiers de l'offre à elles seules. Par contre l'on constate également que l'offre d'Airbnb pour des locations avec deux et trois chambres sont présentes dans des proportions similaires avec pour pourcentages respectifs 25% et 23%. Ainsi pour ces deux types de biens, l'écart entre la proportion de l'offre sur le marché immobilier et les offres d'annonces est quasi-identique.

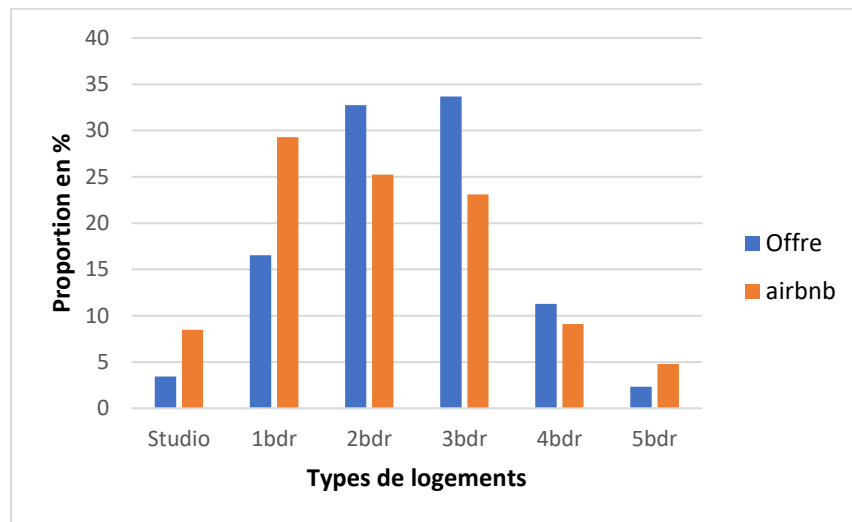


Figure 5.3 : Distribution totale des annonces et de l'offre de logement par type.

Nos coefficients ne sont pas tous estimés sur les douze villes suivant la segmentation choisie (comme vu dans le paragraphe décrivant les données). Nous avons refait les distributions en prenant en compte seulement les villes intervenant dans chaque segmentation. La figure 5.4 montre ces distributions, le tableau 5.2 reprend les parts des annonces et des logements du marché pour les groupes et le global de chaque type de biens.

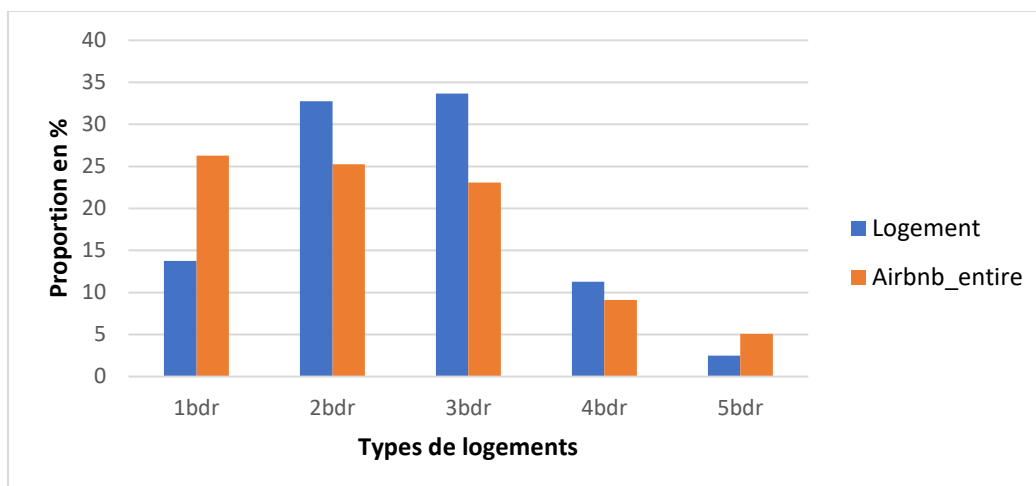


Figure 5.4 : Distribution de l'offre et des annonces par type en prenant en compte que les villes du groupe étudié.

Segment	1 bdr		2 bdr		3 bdr		4 bdr		5 et + bdr	
Distribution	Lognt	B&B	Lognt	B&B	Lognt	B&B	Lognt	B&B	Lognt	B&B
Totale	16,5	29,3	32,8	25,2	33,7	23,1	11,3	9,1	2,3	4,8
Groupe	13,7	26,3	32,8	25,2	33,7	23,1	11,3	9,1	2,5	5,1

Tableau 5.2 : Pourcentage de chacune des distributions.

Compte tenu du fait que les segments des deux et des trois chambres sont faits avec la totalité des villes les distributions demeurent inchangées. Pour le segment des quatre chambres une seule ville est retirée, cette ville n'impacte pas sur les distributions. Par contre pour les segments d'une chambre et de plus de cinq chambres les distributions varient mais l'écart entre le pourcentage de logements et le pourcentage d'annonces reste le même.

Si l'on compare maintenant les résultats obtenus pour les logements à une chambre et ceux à trois chambres. D'après les distributions de la figure 5.4, on voit clairement une sur-représentativité des annonces de logement à une chambre sur Airbnb par rapport à l'offre de ces logements dans l'ensemble des villes. Alors que pour les logements à trois chambres nous avons l'inverse. Les estimations significatives des coefficients des *listings* d'Airbnb pour les biens à une chambre sont tous d'ordre supérieur par rapport à ceux des logements composés de trois chambres pour tous les modèles étant significatifs. Malgré la différence de groupe de villes pour ces types de biens, il semblerait qu'une sur-

représentativité des annonces par rapport au nombre de logements pour un segment fasse augmenter la valeur immobilière de ce segment.

Cela semble également être confirmé lorsque l'on compare les segments des trois chambres à celui des quatre chambres. Pour les quatre chambres l'écart entre la proportion d'annonce et d'offre de logement est très faible comparé à celui des trois chambres. Les résultats montrent que les coefficients estimés des *Listings* Airbnb pour les quatre chambres sont soit autant ou plus significatifs et nettement plus élevés.

5.2.4 L'importance de Santa-Barbara

Au vu de l'importance de la contribution de Santa-Barbara en termes d'annonces d'Airbnb nous avons pensé qu'il était intéressant de refaire le même exercice que précédemment sans inclure cette ville. En effet jusqu'en Décembre 2014, les annonces pour la ville de Santa-Barbara représentent plus de 50% des annonces de tout type disponible pour les villes étudiées (Figure 5.1). Par la suite il y a un rattrapage des autres villes (Figure 5.2). Néanmoins le poids des annonces de la ville de Santa-Barbara reste supérieur à 30% par la suite. Finalement Santa-Barbara représente également plus du tiers des annonces de logements entiers

Le fait de retirer Santa-Barbara du panel nous permet de comparer des villes dont l'entrée d'Airbnb est semblable. Les résultats ainsi obtenus pour la période allant jusqu'à fin 2015 ont des coefficients estimés pour les *Listings* d'annonce d'Airbnb qui sont autant ou plus significatifs qu'avec Santa-Barbara pour la plupart des modèles et des segments. Nous constatons seulement une diminution de la significativité pour les segments trois et quatre chambres pour les modèles 3 et 4. Néanmoins toutes les estimations ont des valeurs supérieures sans Santa-Barbara pour cette période (Tableau 5.3). L'estimation de l'impact d'Airbnb sur le prix des biens immobiliers semble donc plus grand sur cette période lorsque Santa-Barbara est retirée du panel.

Pour la période comprenant l'année 2016, les résultats sont plus contrastés (Tableau 5.4). Nous retrouvons des résultats se comportant comme sur la période allant jusqu'à fin 2015 pour les segments de non différenciation des annonces, des logements entiers, des

deux chambres et cinq chambres et plus. Pour le segment à trois chambres lorsqu'il y a significativité des estimations celle-ci diminue mais les valeurs estimées se voient quasiment doubler (modèles 1 et 2). Pour le segment à quatre chambres seul le modèle 1 voit ses coefficients estimés augmenter légèrement et rester autant significatif. Pour les autres modèles il y a une diminution de la significativité voire une perte pour les modèles 3 et 4 et une diminution des valeurs estimées. Autrement dit, on a une différence lorsque l'on introduit les variables dichotomiques annuelles et la tendance temporelle.

Certains résultats attirent tout particulièrement l'attention, d'abord celui des segments des logements comportant deux chambres. Pour ce segment des deux chambres, nous avons des distributions identiques des annonces et des logements avec et sans la ville de Santa-Barbara, soit 32,8% des logements et 25,2% des annonces avec Santa-Barbara et respectivement 32,7% et 25,5% sans Santa-Barbara (figures 5.5 et 5.6). Or les coefficients estimés deviennent tous significatifs pour la période allant jusqu'à fin 2015 et ceux des modèles 1 et 2 le deviennent pour la période allant jusqu'à fin 2016. Donc dans ce cas on a augmentation de la significativité alors que les distributions n'ont pas changé. Cette constatation semble contredire celle faite précédemment au sujet des différences de distributions pour les logements à une chambre. Néanmoins en retirant Santa-Barbara du panel, certes les distributions des annonces et des logements n'ont pas changé mais la distribution des prix oui. Comme nous pouvons le constater sur la figure 5.7. Le prix de ce type de logements est globalement plus faible lorsque l'on exclue Santa-Barbara. Autrement dit, il semblerait que l'impact estimé (c.p.) est plus significatif lorsque les prix des logements sont plus faibles. Il y aurait en quelque sorte un effet prix.

Lorsque l'on regarde les résultats pour les trois chambres également pour les modèles 1 et 2, on constate qu'il y a une baisse de la significativité des estimations lorsque l'on retire Santa-Barbara du panel mais que les coefficients ont des valeurs qui se voient quasiment doubler. Or la distribution sans Santa-Barbara voit l'écart entre la proportion de logement sur le marché et d'annonce augmenter (figures 5.5 et 5.6). D'après les premières interprétations faites sur le segment d'une chambre les coefficients aurait dû devenir moins significatifs et moins élevés. Mais une fois de plus pour ce segment le prix des logements a diminué lorsque l'on retire Santa-Barbara (figure 5.7).

Période 2010-2015					
Modèle		1	2	3	4
All	avec SB	0.000455**	0.000250**	0.0000926	0.0000760
		(0.000194)	(0.000105)	(0.0000617)	(0.0000468)
		864	864	864	864
	sans SB	0.611	0.818	0.871	0.917
		0.00336***	0.00262***	0.00161**	0.00129**
		(0.000377)	(0.000293)	(0.000572)	(0.000463)
792	792	792	792		
0.688	0.848	0.876	0.920		
Entire Home	avec SB	0.000761*	0.000435**	0.000188	0.000160
		(0.000348)	(0.000196)	(0.000133)	(0.0000983)
		864	864	864	864
	sans SB	0.615	0.820	0.872	0.917
		0.00449***	0.00335***	0.00220***	0.00185***
		(0.000641)	(0.000464)	(0.000612)	(0.000449)
792	792	792	792		
0.691	0.850	0.880	0.923		
1 Bdr	avec SB	0.0203**	0.0123	0.00882	-0.00014
		(0.00714)	(0.00586)	(0.00496)	(0.00271)
		360	360	360	360
	sans SB	0.536	0.805	0.779	0.881
		0.0203**	0.0123	0.00882	-0.00014
		(0.00714)	(0.00586)	(0.00496)	(0.00271)
360	360	360	360		
0,536	0,805	0,779	0,881		
2 Bdrs	avec SB	0.00222	0.000669	0.00000802	-0.000145
		(0.00155)	(0.000777)	(0.000724)	(0.000566)
		789	789	789	789
	sans SB	0.505	0.801	0.851	0.900
		0.0140***	0.00708***	0.00617***	0.00439**
		(0.00181)	(0.00165)	(0.00145)	(0.00169)
717	717	717	717		
0.578	0.814	0.864	0.904		

Modèle		1	2	3	4
3 Bdrs	avec SB	0.00569** (0.00200)	0.00359*** (0.00105)	0.00222** (0.000899)	0.00200** (0.000647)
		864 0.622	864 0.824	864 0.872	864 0.922
	sans SB	0.0191*** (0.00440)	0.0121*** (0.00297)	0.00755* (0.00415)	0.00566* (0.00264)
		792 0.665	792 0.833	792 0.871	792 0.921
4 Bdrs	avec SB	0.0122*** (0.00256)	0.00889*** (0.00163)	0.00678*** (0.00148)	0.00654*** (0.00100)
		864 0.667	864 0.830	864 0.845	864 0.894
	sans SB	0.0193*** (0.00291)	0.0126*** (0.00310)	0.00863* (0.00399)	0.00680* (0.00305)
		792 0.660	792 0.817	792 0.831	792 0.884
5 + Bdrs	avec SB	0.00707 (0.00516)	0.00374 (0.00319)	-0.00121 (0.00329)	-0.00168 (0.00190)
		545 0.553	545 0.688	545 0.708	545 0.799
	sans SB	0.0219 (0.0153)	0.0168** (0.00658)	0.00759 (0.0109)	0.0130 (0.00821)
		473 0.562	473 0.690	473 0.700	473 0.792

Tableau 5.3 : Variable dépendante : Indice de Zillow. Coefficients estimés, leurs écarts types, le nombre d'observations et les R-carrés ajustés pour les listings Airbnb par segment avec et sans Santa-Barbara pour la période 2010-2015 (option *cluster* par ville).

Période 2010-2016					
Modèle		1	2	3	4
All	avec SB	0.000428** (0.000179)	0.000216** (0.0000915)	0.0000266 (0.0000403)	0.0000318 (0.0000374)
		1008 0.703	936 0.842	1008 0.906	936 0.927
	sans SB	0.00205*** (0.000322)	0.00144*** (0.000237)	0.0000196 (0.000406)	0.0000414 (0.000460)
		924 0.749	858 0.857	924 0.902	858 0.924
Entire Home	avec SB	0.000714** (0.000306)	0.000382** (0.000169)	0.0000732 (0.0000754)	0.0000798 (0.0000662)
		1008 0.708	936 0.843	1008 0.906	936 0.927
	sans SB	0.00275*** (0.000471)	0.00194*** (0.000349)	0.000426 (0.000482)	0.000454 (0.000519)
		924 0.751	858 0.859	924 0.903	858 0.925
1 Bdr	avec SB	0.0178*** (0.00249)	0.0134** (0.00352)	0.0109* (0.00483)	0.00264 (0.00419)
		420 0.706	390 0.854	420 0.854	390 0.898
	sans SB	0.0178*** (0.00249)	0.0134** (0.00352)	0.0109* (0.00483)	0.00264 (0.00419)
		420 0.706	390 0.854	420 0.854	390 0.898
2 Bdrs	avec SB	0.00227 (0.00138)	0.000681 (0.000811)	-0.000428 (0.000334)	-0.000486 (0.000297)
		921 0.630	855 0.821	921 0.896	855 0.919
	sans SB	0.00964*** (0.00132)	0.00549*** (0.00128)	0.00136 (0.00161)	0.000772 (0.00175)
		837 0.677	777 0.834	837 0.895	777 0.916

Modèle		1	2	3	4
3 Bdrs	avec SB	0.00417*** (0.00110)	0.00272*** (0.000679)	0.000821 (0.000526)	0.000937 (0.000548)
		1008	936	1008	936
		0.705	0.843	0.897	0.925
	sans SB	0.00772** (0.00301)	0.00536** (0.00203)	0.0000874 (0.00128)	0.000418 (0.00173)
		924	858	924	858
		0.710	0.843	0.893	0.922
4 Bdrs	avec SB	0.0111*** (0.00144)	0.00764*** (0.00114)	0.00489** (0.00179)	0.00521*** (0.00146)
		864	864	864	864
		0.667	0.830	0.845	0.894
	sans SB	0.0132*** (0.00335)	0.00761** (0.00248)	0.00235 (0.00338)	0.00225 (0.00239)
		924	858	924	858
		0.710	0.830	0.853	0.888
5 + Bdrs	avec SB	0.0101** (0.00380)	0.00579* (0.00266)	0.000917 (0.00176)	0.000036 (0.00201)
		545	545	545	545
		0.553	0.688	0.708	0.799
	sans SB	0.0187** (0.00656)	0.0131** (0.00507)	0.00510 (0.00515)	0.00191 (0.00530)
		557	515	557	515
		0.651	0.728	0.765	0.811

Tableau 5.4 : Variable dépendante : Indice de Zillow. Coefficients estimés, leurs écarts types le nombre d'observations et les R-carrés ajustés pour les listings Airbnb par segment avec et sans Santa-Barbara pour la période 2010-2016 (option *cluster* par ville).

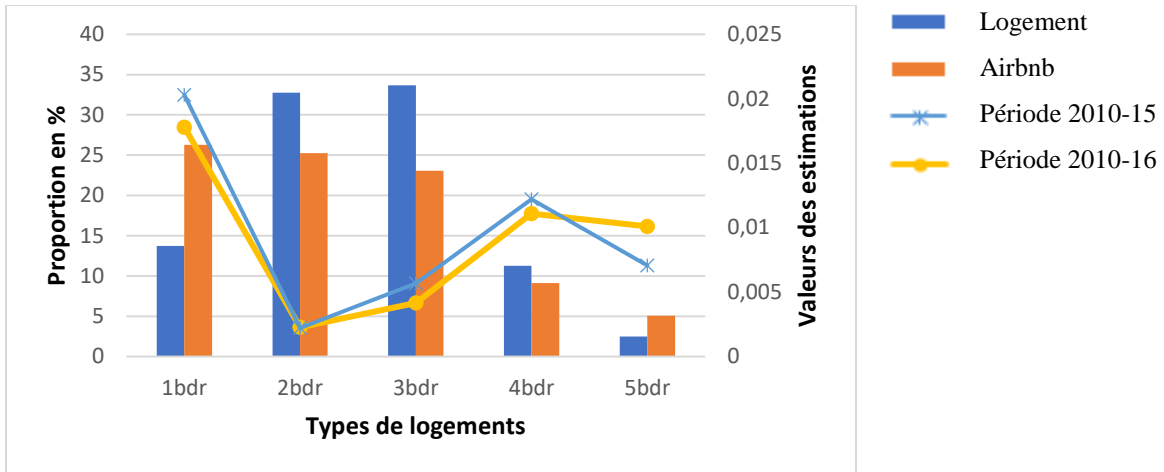


Figure 5.5 : Distributions par groupe avec résultats des coefficients estimés du modèle 1 par segment pour les deux périodes.

Au vu des résultats obtenus il semblerait que les estimations de l'impact des *Listings* d'annonces d'Airbnb varie suivant deux éléments. Le premier serait la différence entre les proportions du type de logements présent sur le marché et la proportion d'annonces Airbnb pour ce même type de bien. Lorsqu'il y aurait sur-représentativité des annonces Airbnb par rapport à l'offre du bien, les valeurs estimées seraient plus grandes et plus significatives et inversement (c.p.). L'autre élément serait le prix du bien, si celui-ci est moins cher alors l'impact des annonces sur le prix du bien serait plus grand (c.p.).

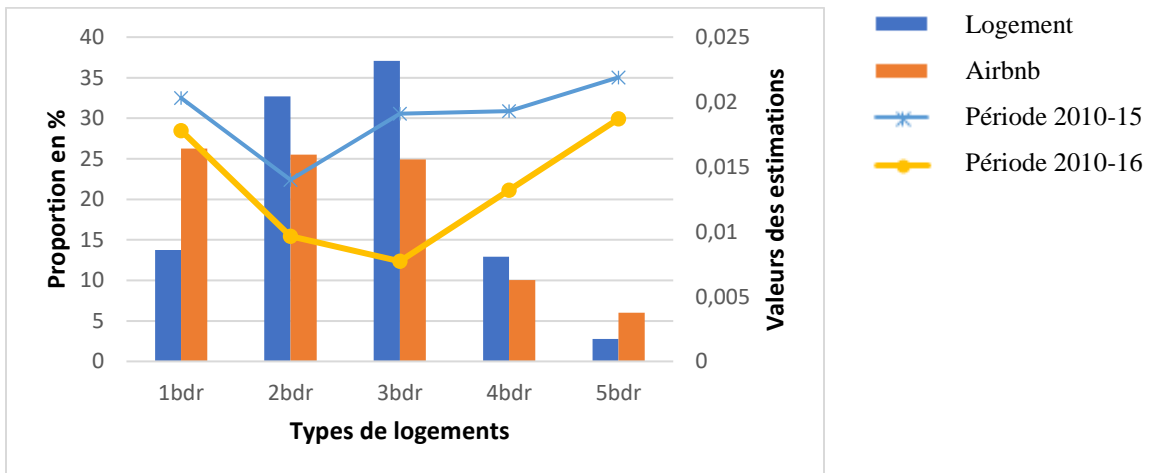


Figure 5.6 : Distributions par groupe sans Santa-Barbara avec résultats des coefficients estimés du modèle 1 par segment pour les deux périodes.

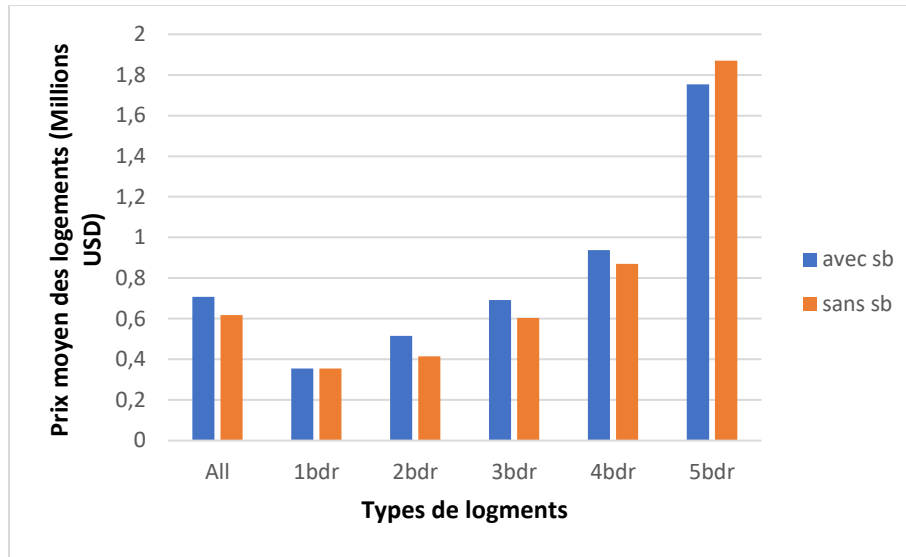


Figure 5.7 : Prix moyen en Millions USD des logements dans les groupes de villes étudiées par segment.

5.3 Prédiction des prix d'une nuitée

5.3.1 Estimations directes

5.3.1.1 La méthode de LASSO

Analysons d'abord les techniques utilisées sans apprentissage statistique qui sont la régression de LASSO (Tableau 5.5) et l'arbre de décisions (Tableaux 5.6, 5.7 et 5.8). La régression de LASSO faite sur le jeu de donnée 1 prend en compte douze paramètres alors qu'avec les deux autres jeux de données seulement sept paramètres identiques sont gardés. On retrouve d'ailleurs six de ces sept paramètres dans la régression de LASSO sur le jeu de données 1. Le seul paramètre de ces sept qui n'est pas gardé avec le jeu de donnée 1 est la ville de Montecito. Pour ce jeu de données, la régression de LASSO a exclu le fait que si le logement est situé dans la ville de Montecito cela a un impact sur le prix. Pour les deux jeux de données modifiés (2 et 3) les valeurs des coefficients estimés sont semblables.

La première chose qui nous surprend est la non prise en compte pour les trois jeux de données de variables portant sur la qualité du logement ou de l'hôte comme les commentaires ou les différentes notations. Pourtant Airbnb insiste sur le fait de faire se noter les utilisateurs entre eux.

Le deuxième élément qui nous a attiré est que la régression de LASSO effectuée sur le jeu de données 1 est la seule qui nous donne des coefficients négatifs pour de paramètres qui sont l'accueil de groupe de personnes et d'animaux. D'ailleurs c'est la possibilité d'accueillir des animaux qui selon ces résultats fait diminuer le plus le prix de la nuit. Malgré le fait que pour ce jeu de données nous avons la valeur de la *Root Mean Square Error* (RMSE) la plus élevée, ce qui ferait de ce résultat le moins bons des trois, avoir des valeurs négatives ne nous semble pas aberrant. En effet les gens ne possédant pas d'animaux, ne souhaitent très certainement pas louer un logement dans lesquels des animaux ont pu séjourner ou séjournent régulièrement. Par exemple, les voyageurs n'ont pas envie de repartir avec des poils de chat sur leurs affaires. Il en va de même pour le paramètre groupe de personnes, ce paramètre permet un accueil d'évènement pour des groupes à l'intérieur du logement. Les voyageurs peuvent être négativement sensible à ce paramètre car si le logement permet ce genre d'évènement, les autres logements autour pourraient le permettre aussi et ainsi une nuisance lors du séjour peut être présente.

Une autre chose, dont on s'aperçoit pour les trois jeux de données, est la prise en compte de certaines variables explicatives et les valeurs de leurs coefficients estimés semblent être aberrantes. C'est le cas de la variable « lit de bébé » qui est gardée dans les trois régressions et qui à la valeur de coefficient estimé la plus forte. Ainsi la présence d'un lit pour bébé dans le logement ferait augmenter le prix du logement de plus de 400\$ pour les jeux de données modifiés (2 et 3) et de plus de 1500\$ pour le jeu de donné 1 (toutes choses étant égales par ailleurs). La régression de LASSO semble complètement inadaptée à nos données.

	Jeu de données 1	Jeu de données 2	Jeu de données 3
Log (Lambda)	3,13	Log (Lambda) 3,31	Log (Lambda) 4,19
RMSE	339,91	RMSE 285,99	RMSE 320,48
Lit_bebe	1544,69	Lit_bebe 431,40	Lit_bebe 402,61
Cache_prises	46,95	Cache_prises 219,59	Cache_prises 192,67
Jacuzzi	109,23	Jacuzzi 83,94	Jacuzzi 84,45
Bdr	71,25	Bdr 64,61	Bdr 68,93
Climatisation	56,05	Climatisation 11,81	Climatisation 12,42
Person_capacity	38,10	Person_capacity 37,02	Person_capacity 35,43
Barrieres_escaliers	465,00	TownMontecito 47,00	TownMontecito 31,56
Handicap	46,67		
Sport	44,68		
Piscine	16,03		
Group	-3,67		
Animaux_accept	-24,20		

Tableau 5.5 : Résultats de la régression de LASSO par jeu de données.

Au vu de ces résultats pour la présence d'un lit bébé et de cache-prises¹² nous avons refait la méthode de LASSO en incorporant des interactions. Nous souhaitons au départ générer toutes les interactions possibles entre les 70 variables, soit nous retrouver avec 2415 interactions, mais nous atteignons les limites d'affichage du logiciel. De plus, beaucoup d'équipements sont peu présents au travers de nos *Listings*. Nous avons pris une trentaine d'équipements les plus présents dans les *Listings*. Les interactions se font entre les équipements entre eux et les différentes variables incluant le lit pour bébé et les caches prises. Ainsi nous obtenons plus de 500 interactions en plus de nos données initiales. L'idée est de vérifier si les valeurs pour ces deux équipements ne sont pas corrélées avec d'autres équipements qui n'apparaissent pas dans les résultats. Malgré le fait que le nombre de variables devient supérieur ou proche du nombre d'observations suivant les différents jeux de données, la méthode de LASSO nous donne des résultats. Ces résultats¹³ nous montrent que pour les trois jeux de données les variables lit pour bébé et caches prises ne sont plus prises en compte séparément mais en interactions.

¹² Élément permettant de cacher la prise électrique pour éviter aux jeunes enfants de s'électrocuter.

¹³ Les résultats prenant en compte un grand nombre de variables à cause des différentes interactions, il est impossible de les présenter en annexe. Ils sont donc accessibles avec ce lien : <https://www.dropbox.com/s/0pj68km42o3o80l/Lasso%20avec%20interactions.xlsx?dl=0>

Ainsi lit pour bébé est considéré via ses interactions avec les variables piscines et jacuzzi alors que la variable caches prises est prise compte via ses interactions avec les variables piscine, jacuzzi et climatisation. Néanmoins l'interaction entre caches prises et lit pour bébé est présente et elle correspond au plus grand impact pour les trois jeux de données avec à chaque fois une valeur de coefficient estimé à plus de 1300. Ainsi même en intégrant les interactions à notre régression de LASSO, il semblerait qu'il y ait toujours une surévaluation de l'impact de ces deux variables. De plus il apparaît que de nombreuses interactions sont gardées et semblent avoir des valeurs étranges. C'est par exemple le cas de des interactions shampoing avec détecteur de fumée ou avec extincteur qui diminueraient le prix de la nuit, tout comme l'interaction Climatisation et TV par câble selon les résultats du jeu de données 2 (RMSE la plus petite).

5.3.1.2 Méthode d'arbre de décision

Pour la technique d'arbre de décision, les arbres des jeux de données modifiés comportent une branche de plus que celui du jeu de données 1. La branche supplémentaire est celle du choix entre un logement entier ou non. Ainsi pour les trois jeux de données la variable explicative qui affecte le plus le prix est la capacité du logement loué. Si celle-ci est supérieure à huit personnes, la variable explicative qui impactera le prix du logement est le nombre de commentaires. Pour des logements dont la capacité est inférieure à huit personnes, le nombre de chambre est la variable qui va affecter le prix du logement. Le nombre de chambre est différencié entre un nombre de deux chambres et trois chambres et plus. La différence pour les modèles 2 et 3 survient après la différenciation du nombre de chambres. Sur la branche des deux chambres et moins, une différenciation qui influence le prix de la nuit est le choix entre un logement entier ou non. Les arbres de décision semblent cohérents. En effet si la capacité recherchée est supérieure à huit personnes, le logement sera forcément un logement entier, de plus le nombre de logements disponibles sera moins grand et ces logements auront des caractéristiques proches. Donc le moyen de différenciation sera la réputation de l'hôte et du logement, d'où l'importance des commentaires dans le choix. En dessous d'une capacité de huit personnes le nombre de chambres semblent être le facteur

déterminant lorsque plusieurs personnes voyagent ensemble. De plus le fait que la discrimination sur l'arbre se fait entre moins de deux chambres et plus de trois chambres renseigne sur la taille du groupe. Si les chambres sont de deux personnes la différenciation sur cette branche se fait donc entre moins de quatre personnes et plus de quatre personnes. Alors ajouter une branche laissant le choix entre un logement entier et un logement partagé est tout à fait cohérent lorsque les voyageurs cherchent moins de deux chambres.

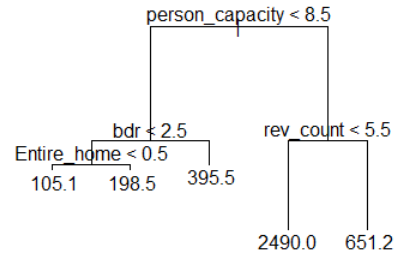
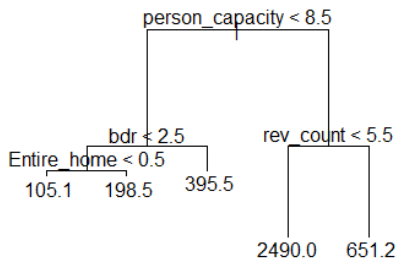
Les émondages des arbres par minimum de déviation n'ont aucun impact sur l'arbre pour les jeux de données 1 et 2 (Tableaux 5.6 et 5.7). Par contre l'arbre émondé pour le jeu de données 3 est simplifié (Tableau 5.8). Cette simplification consiste dans le retrait de la branche supplémentaire des logements entiers ou non. La méthode d'arbre de décision nous semble appropriée et cohérente pour l'ensemble des jeux de données.



Arbre de décision

Arbre de décision émondé

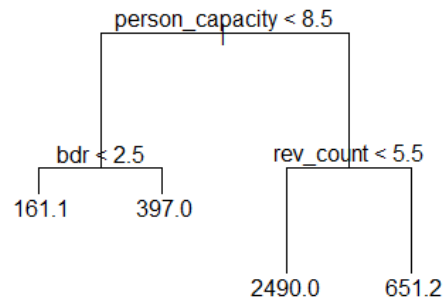
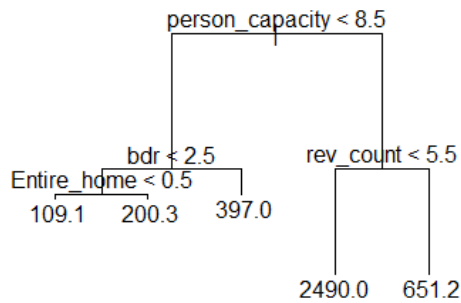
Tableau 5.6 : Résultats de la technique d'arbre de décision et d'émondage de l'arbre pour le jeu de données 1.



Arbre de décision

Arbre de décision émondé

Tableau 5.7 : Résultats de la technique d'arbre de décision et d'émondage de l'arbre pour le jeu de données 2.



Arbre de décision

Arbre de décision émondé

Tableau 5.8 : Résultats de la technique d'arbre de décision et d'émondage de l'arbre pour le jeu de données 3.

5.3.2 Méthodes d'apprentissage statistique

Les méthodes effectuées sur les différents jeux de données avec la méthode de LASSO nous donne des résultats extrêmement variés d'un jeu de données à l'autre. Le tableau 5.10 reprend les résultats de cette méthode. Le résultat le plus surprenant est le nombre de variables explicatives qui ont été gardées avec l'utilisation du jeu de données 2. En effet

celui-ci garde 34 variables explicatives alors qu'avec les jeux de données 1 et 3, les prédictions en gardent respectivement trois et deux. On remarque également des valeurs de coefficients estimés aberrantes comme celles relatives aux variables explicatives « lit bébé » et « baignoire bébé ». Pour le jeu de données 1 hormis la variables explicative « lit bébé » et pour le jeu de données 3 on retrouve des variables de l'arbre de décision.

Pour comparer les résultats des méthodes de prédiction nous utilisons la RMSE. Les résultats apparaissent dans le Tableau 5.9. Plus la RMSE est petite, meilleure est la technique utilisée. Ainsi sur les jeux de données 1 et 2 les prédictions utilisant une méthode de forêt d'arbres donne de meilleurs résultats pour les prédictions. Alors que pour le jeu de données 3 c'est la technique de prédiction avec la régression de LASSO qui est préférée. Pour rappel, cette régression de LASSO utilise seulement deux variables explicatives : la capacité d'accueil et le nombre de chambres. Le choix du paramètre Lambda est présenté en annexe E.

Jeu de données	1	2	3
Nombre d'observations	484	648	575
Nombre de variables	70	70	68
Nombre d'obs dans le <i>Training Set</i>	242	324	287
Log(best_lasso)	4,95	1,90	4,95
RMSE (LASSO)	185,26	361,63	162,21
RMSE (Tree)	390,75	357,87	414,32
RMSE (Random Forest)	180,78	343,82	198,52
Meilleure RMSE	Random Forest	Random Forest	LASSO

Tableau 5.9 : Caractéristiques et résultats des différentes méthodes prédictives sur les trois jeux de données.

Dans cette partie sur la prédiction du prix de la nuitée nous nous rendons compte que le traitement des valeurs omises et le choix de la méthode sont liés. Ce lien est présent que l'on utilise des techniques avec ou sans *Machine Learning*. Ainsi la prédiction du prix de la nuitée qui intéresse le propriétaire peut varier en valeur et en certitude.

Jeu de données 1		Jeu de données 2		Jeu de données 3	
(Intercept)	82,16	(Intercept)	-63,75	(Intercept)	48,40
Bdr	62,20	Bdr	20,60	Bdr	83,11
Person_capacity	20,69	Person_capacity	50,02	Person_capacity	20,69
Lit_bebe	890,58	Baignoire_bebe	766,99		
		TownMontecito	117,80		
		Interphone	69,99		
		Protections_fenetres	64,06		
		Climatisation	61,37		
		Stores	52,51		
		Piscine	43,22		
		chambre_verrou	42,35		
		Info_Obabysitters	36,88		
		TownCarpinteria	35,64		
		Cheminee	32,09		
		Jacuzzi	24,99		
		Seche_linge	14,57		
		Espace_PC	8,99		
		Chauffage_electrique	6,25		
		Rating_cleanliness	5,85		
		Satisfaction_guest	2,30		
		Table_langer	1,64		
		Cancel_policy	0,60		
		Baignoire	-2,28		
		Premiers_secours	-2,84		
		Internet	-3,26		
		Fumeur	-12,69		
		Animaux_accept	-14,39		
		Instant_book	-18,51		
		Produits_base	-19,25		
		TownGoleta	-20,72		
		Rating_communication	-23,61		
		Fer_repasser	-24,53		
		TownSantaMaria	-96,60		
		TownLompoc	-98,57		

Tableau 5.10 : *Machine Learning* avec LASSO par jeu de données.

Chapitre 6

Discussions

6.1 Prix de l'immobilier

6.1.1 Les approximations

La première partie de l'étude porte sur les estimations de l'impact d'Airbnb sur le prix des logements. Le fait de coupler des données provenant de différentes sources et de différents niveaux agrégés ne fait qu'accentuer les approximations. D'une part celles provenant du *web-scraping* d'Airbnb posent certaines questions. En récoltant les données nous ne pouvons pas être certains du nombre d'annonces présentes sur le site à une date précise. Les annonces qui ont été retirées ne sont pas accessibles or de nombreuses raisons peuvent pousser des utilisateurs à retirer leur annonce. L'hôte a pu déménager, il a pu également essayer temporairement Airbnb, il a également pu se forger une clientèle qui le contact directement ou des politiques de régulation mises en place par les autorités l'ont dissuadées de continuer. D'autre part les données macro-économiques ont dû pour certaines être approximées pour compléter un manque d'information au niveau locale. L'ensemble de ces approximations et possibles erreurs cumulées rendent difficile d'établir une interprétation franche. La possibilité future de l'accès à des données d'*open-source* au niveau municipal rendra sans-doute ce type d'étude plus efficace. Dans le contexte d'une politique volontariste de partage de données et d'une nouvelle vision « *smart-city* », ce genre d'études sera sans aucun doute facilité.

6.1.2 Le retrait de Santa-Barbara

Le retrait de Santa-Barbara dans une partie de notre étude peut également être discutable car la ville fait partie d'une agglomération. Dans une agglomération se pose alors le problème des interconnexions entre les villes. En étudiant le phénomène par villes nous ne pouvons pas exclure l'impact qu'ont les autres villes de la même agglomération et donc l'impact qu'à le prix de l'immobilier d'une ville sur une autre mais aussi l'impact des annonces Airbnb présentes dans une ville sur la ville voisine. Pour Santa-Barbara par exemple, les données des prix sont agrégées pour toute la ville alors

que celle-ci se retrouve coupée en deux. Une partie avec l'université proche de Goleta et une autre entourée de plusieurs villes côtières. La possibilité de passer à une étude par codes postaux serait envisageable pour les données d'Airbnb et de Zillow mais beaucoup moins envisageable dès lors que l'on souhaite utiliser des données macro-économiques. Pour notre part nous pensons que l'impact d'Airbnb sur le prix de l'immobilier peut être très locale. Un quartier extrêmement touristique attire un plus grand nombre de voyageurs aptes à utiliser les locations temporaires d'Airbnb. Si cela fait varier seulement le prix de l'immobilier dans ce quartier, l'effet sera dilué au niveau d'une ville ainsi l'impact des annonces d'Airbnb peut être sous-estimé.

Notre étude été basée sur le possible phénomène de changement de marché du bien immobilier par le propriétaire. Le propriétaire en souhaitant rentabiliser le plus possible son bien se tournait vers le marché locatif temporaire. Les biens facilement louables et instinctivement qui verraient leur rentabilité augmenter rapidement sont les logements de types studio ou à une chambre. Ces biens sont des biens abordables sur le marché locatif traditionnel et il semblerait que pour les logements à une chambre les annonces d'Airbnb ont plus impacté le prix de l'immobilier. Certes nous avons vu que la représentativité des annonces est nettement supérieure à l'offre de ce type de logement. Mais nous avons également vu qu'en retirant la ville de Santa-Barbara le prix des biens diminuait et que les impacts estimés eux augmentaient. Il serait intéressant dès lors d'étudier l'impact sur les logements les plus abordables à l'achat, comme le tiers le plus bas de la distribution des valeurs de l'immobilier. Pour vérifier si effectivement les biens les plus abordables sont ceux qui voient leur valeur augmenter. Il semblerait qu'une étude de Kung allant dans ce sens est sur le point d'être publiée (Wall Street Journal, 2017). Cette étude mettrait également en avant la possibilité que le taux d'occupation par le propriétaire soit un facteur important. Il est vrai qu'un logement est possiblement transférable du marché long terme au marché court terme que si celui-ci n'est pas occupé par le propriétaire.

6.1.3 Les ordres de grandeurs

La comparaison de l'impact estimé des annonces d'Airbnb avec l'impact estimé des autres variables explicatives choisies est un enjeu. Dans notre partie sur les résultats des données mensuelles nous avons essayé de mettre ces valeurs en perspective. En faisant

abstraction des problématiques autour des approximations précédemment citées, vient la question de comment comparer des annonces avec un revenu médian ou un taux de chômage. Nous avons vu que la croissance annuelle du nombre d'annonces sur Airbnb était en train de s'accélérer lors de la dernière année (figure 5.2). Cette forte augmentation fait en sorte que les estimations des coefficients des annonces Airbnb sont à prendre en considération face aux variables macro-économiques. Néanmoins dans notre panel l'impact d'Airbnb sur le prix des logements n'est pas toujours significatif. Si la volonté des politiques publiques est de limiter l'augmentation des prix de l'immobilier alors peut-être avoir un indicateur du nombre d'annonces publiées sur Airbnb dans les agglomérations serait un indicateur utile. En effet cet impact peut être lourd socialement pour les ménages ayant les plus faibles revenus.

6.2 Prédiction du prix

Notre deuxième partie de l'étude portait sur des techniques d'apprentissage statistique. Avant de nous lancer dans les prédictions nous avons testé les méthodes de régression LASSO et d'arbre de décision sur les différents jeux de données. Si la technique par arbre de décision a semblé cohérente et répliquable sur l'ensemble des jeux de données, la méthode par régression de LASSO a, quant à elle, eu des résultats surprenants. Cette méthode a favorisé un choix de variables explicatives aberrantes en leur attribuant des coefficients estimés n'ayant pas de sens. Ce résultat est resté inchangé même avec l'ajout d'interactions entre les variables. La structure même des données a dû impacter ces choix. D'ailleurs certains problèmes liés à cette technique sont décrits dans l'article de Sendhil Mullainathan et Jann Spiess (Sendhil Mullainathan, 2017).

Pour les prédictions à l'aide de l'apprentissage statistique, nous avons utilisé trois méthodes différentes : LASSO, arbre de décision et forêt d'arbres. La constatation est que les résultats peuvent grandement varier en fonction du jeu de données utilisé. Pourtant les modifications faites sur le jeu de données original semblaient minimales. Il semblerait donc que le choix de la meilleure technique dépende de la qualité des données collectées. Le manque d'homogénéité dans les données collectées par *web-scraping* ne nous permet pas de réellement choisir la meilleure méthode pour les prix des annonces d'Airbnb.

Néanmoins le résultat qui ressort est la minimisation du RMSE tous jeux de données confondus par la régression de LASSO sur le jeu de données 3. Les variables explicatives choisies dans cette régression sont également cohérentes avec celles présentes dans l'arbre de décision. Ces variables explicatives font partis des caractéristiques qui se retrouvent dans l'entête de l'annonce sur le site d'Airbnb comme le montre la Figure 6.1.

[Présentation](#) · [Commentaires](#) · [L'hôte](#) · [Emplacement](#)

Loft ouvert sur le ciel dans le Vieux-Montréal

Loft entier · Montréal

 2 voyageurs  1 chambre  1 lit  1 salle de bain

Figure 6.1 : Exemple de l'en-tête d'une annonce d'Airbnb.

Cette correspondance entre les variables et l'entête de l'annonce intrigue. Cette mise en avant dans l'entête vient-elle du fait qu'Airbnb a ciblé ce qui prédit le mieux le prix ou est-ce parce que les utilisateurs de types voyageurs ne s'intéressent juste qu'à l'entête pour choisir leur logement ? Les commentaires n'apparaissant pas dans l'entête et n'étant pas le facteur déterminant le plus sélectionné par nos modèles prédictifs, il semblerait que les valeurs premières défendues par Airbnb de confiance et de partage entre utilisateurs aient définitivement disparu. L'intérêt se porterait donc plus sur le rapport qualité-prix et la confiance dans la transaction sécurisée de la plateforme. On apporterait ici un argument allant contre l'idée de Rachel Botsman qui voit dans la confiance une nouvelle monnaie.

6.3 Avenues de recherche

Le fait d'avoir mis en exergue un probable lien entre la représentativité du nombre d'annonces et la représentativité de l'offre de logement, nous amène à penser que le

même type d'étude pourrait être fait avec un ratio du genre annonces sur logements. Cette méthode pourrait d'ailleurs offrir l'opportunité de faire des modèles élastiques (régressions Log-Log).

Un autre point qui semble des plus importants est la collecte de données tant pour les données macro-économiques que celles « scrapées ». Un suivi du nombre de mises en chantier dans la construction d'habitation pourrait être également intéressant à incorporer pour chaque ville. La difficulté reste d'avoir suffisamment de données sur une échelle de temps mensuelle pour avoir une homogénéité. Pour ce qui est des données d'Airbnb, récupérer et utiliser les données géospatiales pourrait permettre d'intégrer des distances par rapport au centre-ville ou à d'autres points d'intérêts. Ce dernier élément pourrait être utile à la fois pour les études sur la corrélation entre l'essor d'Airbnb et le prix de l'immobilier et pour les tentatives de prédiction du prix d'une location.

Apport décisionnel

Notre étude semble établir la corrélation entre le nombre d'annonces présentes sur Airbnb et la hausse des prix de l'immobilier. Cet élément semblait déjà acquis par les autorités publiques qui souhaitaient arrêter ou réduire la propagation des annonces. Mais notre étude met avant la possibilité que la différence de proportions de logements et d'annonces joue un rôle important. Ainsi les autorités qui ne veulent pas strictement interdire la location via Airbnb pourraient surveiller la représentativité des annonces pour chaque segment de biens par rapport aux logements présents. De même notre étude pointe l'importance du prix du bien immobilier, ceux qui sont plus abordables semblent avoir leurs prix augmenter plus fortement avec le nombre d'annonces Airbnb. Ainsi les logements les plus abordables verraient leur prix augmenter plus vite avec l'arrivée d'Airbnb. Les pouvoirs publics souhaitant garder un accès abordable à la propriété pourraient considérer cet élément en adoptant des législations différenciées pour les biens. Les zones en manque de biens abordables pourraient ainsi avoir une législation différente.

Pour les propriétaires ou investisseurs notre étude leur montrent d'abord la difficulté de prédire le prix de la nuit seulement avec les caractéristiques présentes sur l'annonce. Même si la différenciation du prix semble se faire sur des critères sensés comme la capacité et le nombre de chambre. Il se peut que d'autres paramètres soient à prendre en considération. Ainsi même si les résultats proviennent des techniques les moins efficaces il semblerait qu'accueillir des animaux ou des groupes ne permette pas d'augmenter le prix de la nuit, contrairement à tous les équipements qui ont trait aux bébés. En effet avoir des équipements pour des bébés semblerait faire augmenter le prix de la nuit mais pas dans les proportions que notre étude a révélées. Certains sites internet proposent aux propriétaires des conseils pour augmenter leurs profits, ces propriétaires devraient donc porter attention aux méthodes utilisées par ces sites qui les conseillent sur le prix à pratiquer.

Conclusion

Les bruits dans la presse d'une flambée des prix de l'immobilier causée par Airbnb nous paraissent plausibles compte tenu des coefficients estimés trouvés dans notre étude et de la récente forte croissance d'annonces publiées sur Airbnb. Les valeurs annuelles nous donnent des résultats significatifs seulement pour le modèle de régression le plus simple, celui qui ne prend en compte que les *Listings* d'annonces et le taux de chômage. Dès lors que des variables explicatives supplémentaires sont ajoutées, les *Listings* d'annonces Airbnb ne semblent plus avoir d'impact au niveau annuel. Au niveau mensuel, les coefficients estimés demeurent généralement significatifs pour le modèle simple et pour celui avec les variables macroéconomiques (respectivement modèles 1 et 2) pour les différentes segmentations. Dès lors que nous introduisons les variables dichotomiques annuelles et la tendance mensuelle les résultats deviennent plus contrastés. Une fois encore le modèle le plus simpliste reste celui qui correspond aux coefficients estimés les plus élevés. Celui pour le segment d'une chambre nous donne la plus forte augmentation du prix de l'immobilier avec une augmentation de 2,03% pour une annonce supplémentaire (c.p.), ce qui correspondrait à environ une augmentation de 6,6% des annonces. Cette augmentation nous paraît considérable compte tenu de l'augmentation des annonces et au vu des autres coefficients estimés pour les autres variables explicatives comme la démographie et le revenu médian pour ne citer qu'eux.

Nous avons possiblement identifié deux phénomènes en s'attardant sur les distributions des offres de logements et du nombre d'annonces par segment (figures 5.5 et 5.6) ainsi qu'en examinant les distributions des prix avec et sans Santa-Barbara (figure 5.7). D'abord il semblerait que lors de la présence d'une sur-représentativité des annonces Airbnb par rapport à l'offre du bien, les valeurs estimées seraient plus grandes et plus significatives ; et inversement (c.p.). Puis l'autre élément serait lié au prix du bien, si celui-ci est moins cher alors l'impact des annonces sur le prix du bien serait plus grand (c.p.). Cela tendrait à nous faire penser que les logements les plus accessibles se voient plus impacter par l'entrée d'Airbnb sur le marché. Même si ces valeurs de coefficient restent discutables, le fait d'augmenter les prix des logements les plus abordables est la

base du principe de gentrification. Alors si le but des régulations mis en place par les autorités est d'empêcher ou de réduire ce phénomène, s'orienter vers des études ciblant les logements les plus abordables au niveau de la location long terme ou au niveau des ventes immobilières semble être primordial. Cela fait partie des avenues de recherches et d'études qui pourraient faciliter les décisions publiques.

Pour ce qui est de prédire le prix de la nuitée, nous avons mis de l'avant trois méthodes : la régression LASSO, les arbres de décision et la forêt d'arbres. Notre constatation est que la prédiction est fortement dépendante de la méthode utilisée et du jeu de données utilisées. Mais les résultats ont mis en avant les facteurs de capacité d'accueil, du nombre de chambres, si la location est bien pour un logement entier et dans une moindre mesure la réputation de l'hôte. Ce qui paraît être des critères tout à fait logiques et déterminants dans la prise de décision du voyageur. En affinant ces prédictions avec une plus grande collecte de données cela permettrait à un propriétaire souhaitant passer du marché locatif long-terme au marché locatif temporaire d'estimer son revenu potentiel. De plus ces techniques de *web-scraping* et d'apprentissage statistique pourraient également servir à déterminer si une saturation du marché de la location temporaire via Airbnb est envisageable.

Nos résultats et les différentes avenues de recherches proposées tant au niveau de l'impact d'Airbnb qu'au niveau de l'utilisation de la prédiction des prix de nuitée nous montre à quel point l'arrivée disruptive sur un marché existant soulève de nombreux questionnements. Questionnements auxquels les pouvoirs publics répondent parfois de manières rapides et non-nuancées sans avoir toutes les cartes en main pour décider. La question n'est pas de lutter comme le voudraient les *lobbies* traditionnels contre l'émergence de l'économie collaborative dont Airbnb est l'un des porte-étendards. La question est comment intégrer au mieux ce nouveau pan de l'économie pour qu'il soit socialement acceptable ?

Bibliographie

- Airbnb. (2015). *economic-impact-airbnb*. Consulté le Août 28, 2017, sur [blog.atairbnb.com](http://blog.atairbnb.com/economic-impact-airbnb/):
- Botsman, R. (2010, Mai). *Rachel Botsman The case for collaborative consumption*. Consulté le 09 10, 2017, sur [www.ted.com](https://www.ted.com/talks/rachel_botsman_the_case_for_collaborative_consumption?language=fr):
- Botsman, R. (2012, Juin). *Rachel Botsman The currency of the new economy is trust*. Consulté le 09 10, 2017, sur [www.ted.com](https://www.ted.com/talks/rachel_botsman_the_currency_of_the_new_economy_is_trust?language=fr):
- Boursorama. (2014, Octobre 27). *L'histoire incroyable derriere le succes d'Airbnb*. Consulté le Août 28, 2017, sur [lifestyle.boursorama.com](http://lifestyle.boursorama.com/evasion/lhistoire-incroyable-derriere-le-succes-dairbnb/):
- Brynjolfsson, E. Y. (2009). *Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition*. Management Science.
- Businessinsider. (2016, Février 23). *How 3 guys turned renting an air mattress in their apartment into a \$25 billion company*. Consulté le Août 28, 2017, sur [Businessinsider.com](http://www.businessinsider.com/how-airbnb-was-founded-a-visual-history-2016-2/#even-during-y-combinator-they-still-got-rejected-famously-by-investors-fred-wilson-of-union-square-ventures-admitted-in-2011-that-he-had-failed-to-look-past-the-air-bed-and-bre):
- Cavailhès, J. (2005). Le prix des attributs du logement. *Economie et Statistique (numéro: 381-382)*, 91-123.
- Cervero, R. A. (2007). *City CarShare: Longer-Term Travel Demand and Car Ownership Impacts*. . Journal of the Transportation.
- Courrier Parlementaire. (2017, Mai 03). *Airbnb et réglementation : ce que dit la loi*. Consulté le Août 28, 2017, sur [lecourrierduparlement.fr](http://www.lecourrierduparlement.fr/airbnb-et-reglementation-ce-que-dit-la-loi/):
- Dean Runyan Associates. (2017). *CATravelImpacts*. Récupéré sur [www.deanrunyan.com](http://www.deanrunyan.com/CATravelImpacts/CATravelImpacts.html):
- Forman, C. A. (2009). *Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live*. Management Science.
- Georgios Zervas, D. P. (2013). The Rise of the Sharing Economy : Estimating the Impact of Airbnb on the Hotel Industry. *Boston University*.

- Government of Santa-Barbara. (2017, Septembre 20). *Short-Term Rental*. Consulté le Septembre 23, 2017, sur Santabarbaraca.gov:
<http://www.santabarbaraca.gov/services/planning/mpe/stvr/default.asp>
- Hoesli, M. (2008). Chapitre 3. Dans M. Hoesli, *Investissement Immobilier, Décision et gestion du risque* (pp. 90-93). Economica.
- Ici Radio Canada. (2017, Avril 27). *La loi anti-Airbnb a peu d'effets*. Consulté le Août 28, 2017, sur Ici.radio-canada.ca: <http://ici.radio-canada.ca/nouvelle/1030589/la-loi-anti-airbnb-a-peu-effet-logements-location-quebec-montreal>
- Journal de Montréal. (2017, Avril 21). *Airbnb négocie avec Québec pour une taxe*. Consulté le Août 28, 2017, sur journaldemontreal.com:
<http://www.journaldemontreal.com/2017/04/21/airbnb-negocie-avec-quebec-pour-une-taxe>
- Kroft, K. D. (2013). *Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist*. Journal of Labor Economics.
- LA Times. (2017, Février 03). *California lawmakers can't figure out what to do with Airbnb. Here's why*. Consulté le Août 28, 2017, sur LATimes.com:
<http://www.latimes.com/politics/la-pol-sac-airbnb-laws-california-legislature-20170203-story.html>
- Le Devoir. (2016, Août 17). *La loi anti-Airbnb est aisément contournée*. Consulté le Août 28, 2017, sur LeDevoir.com: <http://www.ledevoir.com/societe/actualites-en-societe/477891/la-loi-anti-airbnb-est-aisement-contournee>
- Le Figaro. (2017, Mars 30). *Neuf ans après son lancement, Airbnb est rentable*. Consulté le Août 28, 2018, sur LeFigaro.fr: <http://www.lefigaro.fr/secteur/high-tech/2017/01/30/32001-20170130ARTFIG00072-neuf-ans-apres-son-lancement-airbnb-est-rentable.php>
- Lee, D. (2016, février). *How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations*. *Harvard Law & Policy Review*, 10. Consulté le Août 28, 2017, sur http://harvardlpr.com/wp-content/uploads/2016/02/10.1_10_Lee.pdf
- Levasseur, S. (2013). *Éléments de réflexion sur le foncier et sa contribution au prix de l'immobilier*. *Revue de l'OFCE*, 365-394.
- Martin, E. S. (2010). *Impact of carsharing on household vehicle holdings*. Journal of the Transportation Research Board.
- Nouvel Observateur. (2016, Décembre 06). *New-York, Barcelone, Amsterdam, Londres: Ces villes qui réglementent la location par airbnb*. Consulté le Août 28, 2017, sur tempsreel.nouvelobs.com:
<http://tempsreel.nouvelobs.com/immobilier/monde/20161206.OBS2272/new-york-barcelone-amsterdam-londres-ces-villes-qui-reglementent-la-location-par-airbnb.html>

- NYC DataScience. (2016, Novembre 21). *Web Scraping Project: Experience Difference between Airbnb and TripAdvisor*. Consulté le Février 6, 2017, sur blog.nycdatascience.com:
<https://blog.nycdatascience.com/student-works/help-airbnb-hosts-set-renting-price-hint-hotel-price-tripadvisor-web-scraping-project/>
- Seamans, R. F. (2013). *Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers*. Management Science.
- Sendhil Mullainathan, J. S. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 87-106. Récupéré sur <https://scholar.harvard.edu/>:
<https://scholar.harvard.edu/files/sendhil/files/jep.31.2.87.pdf>
- Simon Borel, D. M. (2015). L'Économie collaborative, entre utopie et Big business. *Editions Esprit*, 9-18.
- Techcrunch. (2017, Novembre 16). *Airbnb buys 'Airbnb for disabled people' startup Accomable in accessibility upgrade*. Consulté le Novembre 20, 2017, sur techcrunch.com:
<https://techcrunch.com/2017/11/16/airbnb-buys-accomable-a-specialist-in-travel-listings-for-disabled-people/>
- The Tribune. (2017, Mai 16). *Want to rent a vacation home in SLO County ? You may have fewer options*. Consulté le Août 28, 2017, sur [sanluisobispo.com](http://www.sanluisobispo.com):
<http://www.sanluisobispo.com/news/local/community/cambrian/article150805502.html>
- Wall Street Journal. (2017, Octobre 22). *how-airbnb-affects-home-prices-and-rents*. Consulté le Novembre 15, 2017, sur www.wsj.com: <https://www.wsj.com/articles/how-airbnb-affects-home-prices-and-rents-1508724361>
- Zillow. (2014, Janvier 3). *zhvi-methodology*. Récupéré sur [Zillow.com](http://www.zillow.com):
<https://www.zillow.com/research/zhvi-methodology-6032/>

ANNEXES

ANNEXE A : Résultats annuels pour les deux périodes et équivalence de noms dans les résultats.

	Période 2010 à 2015		Période 2010 à 2016		
	(1) L_ZHVI_All	(2) L_ZHVI_All	(1) L_ZHVI_All	(2) L_ZHVI_All	
BnB_All	0.000352* (2.00)	0.0000415 (0.89)	BnB_All	0.000340** (2.34)	0.0000666 (0.10)
UN_rate	-0.0669*** (-14.88)	-0.0172 (-1.54)	UN_rate	-0.0725*** (-15.91)	-0.00547 (-0.47)
Lpopulation		-0.420** (-2.36)	Lpopulation		0.144 (0.59)
LMed_Inc		0.0774 (0.72)	LMed_Inc		0.0753 (0.56)
L_SpendA		0.571** (2.58)	L_SpendA		0.717*** (3.12)
L_SpendH		0.358 (1.63)	L_SpendH		0.289 (1.39)
SAH1		-0.303*** (-4.04)	SAH1		-0.151* (-1.85)
Occup_House		-0.00195 (-0.96)	Occup_House		-0.00253 (-0.96)
Int_rate		0.123*** (9.67)	Int_rate		0.116*** (8.72)
_cons	13.66*** (432.12)	11.44*** (6.44)	_cons	13.71*** (444.14)	5.498** (2.28)
N	72	72	N	84	84

t statistics in parentheses

* p<.1, ** p<.05, *** p<.01

Équivalence des noms entre les noms des variables explicatives et des résultats	
Noms de la variables X	Noms équivalents dans les résultats
Log(<i>population</i>)	Lpopulation
Log(<i>Median_Household_Income</i>)	LMed_Inc
Log(<i>Spending_Accomodations</i>)	L_SpendA
Log(<i>Spending_Hotel</i>)	L_SpendH
<i>Occupied_Housing_Units</i>	Occup_House
<i>Effective_Interest_rate</i>	Int_rate

ANNEXE B¹⁴ : Résultats Panel Mensuel
Période 2010-2015 – Segment 1 Chambres

	(1)	(2)	(3)	(4)
	L_ZHVI_lbdr	L_ZHVI_lbdr	L_ZHVI_lbdr	L_ZHVI_lbdr
BnB_lbdr	0.0203** [0.00714]	0.0123 [0.00586]	0.00882 [0.00496]	-0.000140 [0.00271]
UN_rate	-0.00171 [0.0195]	-0.0229 [0.0212]	-0.00408 [0.0176]	-0.00239 [0.0176]
Lpopulation		2.488** [0.695]		0.741 [0.832]
LMed_Inc		-0.0856 [0.142]		0.445** [0.121]
L_SpendA		0.327 [0.729]		3.089*** [0.593]
L_SpendH		-0.245 [1.095]		1.194* [0.538]
SAH1		-0.116* [0.0440]		-0.0921*** [0.0131]
Occup_House		-0.0142 [0.00885]		-0.0178* [0.00678]
Int_rate		0.157** [0.0387]		0.122*** [0.0182]
_IYear_2011			-0.0938** [0.0318]	-0.317*** [0.0294]
_IYear_2012			-0.164*** [0.0350]	-0.698*** [0.101]
_IYear_2013			-0.0779 [0.0408]	-0.858*** [0.153]
_IYear_2014			0.0318 [0.0693]	-1.130*** [0.208]
_IYear_2015			0.0640 [0.0970]	-1.365*** [0.262]
time_trend			0.000914 [0.00182]	0.00542* [0.00213]
_cons	12.67*** [0.144]	-9.734 [7.188]	12.77*** [0.156]	-21.89 [11.06]
N	360	360	360	360

Standard errors in brackets
 * p<.1, ** p<.05, *** p<.01

¹⁴ Lien dropbox pour obtenir l'ensemble des résultats mensuels:
<https://www.dropbox.com/s/ijrfinwnrsxygk/annexe%20A%20et%20B%20memoire.pdf?dl=0>

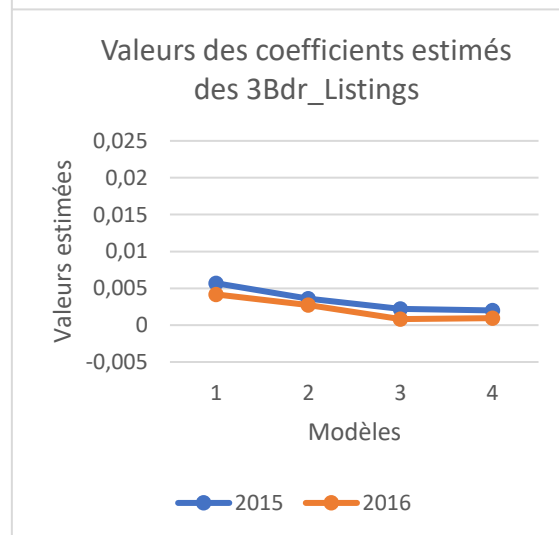
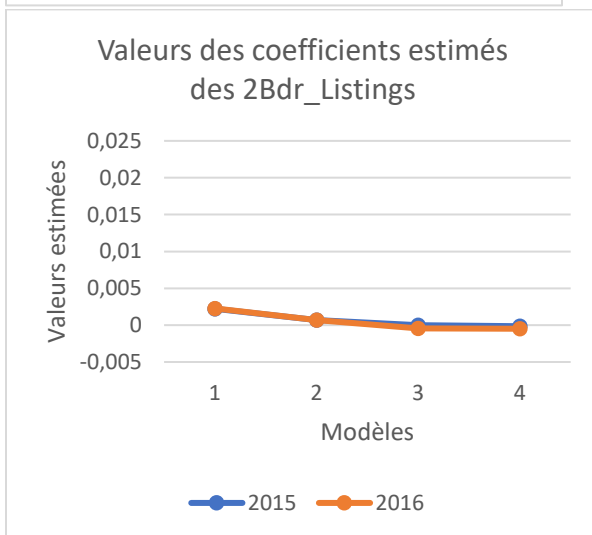
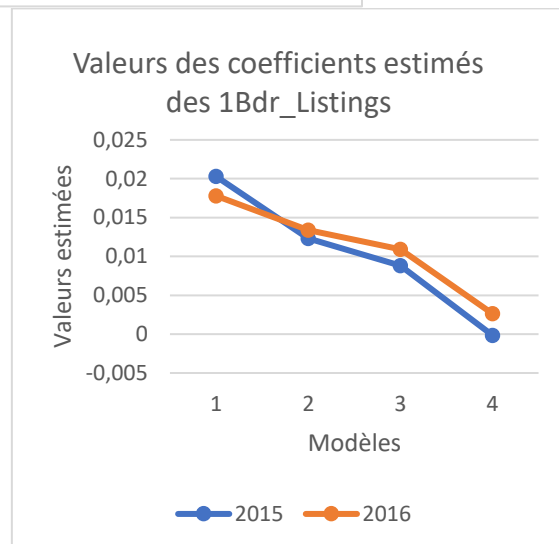
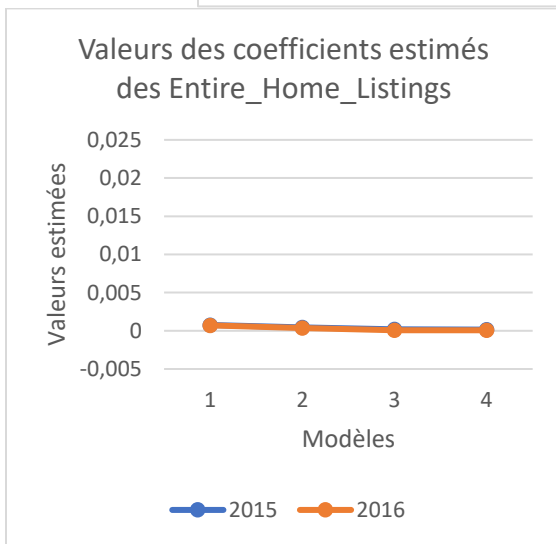
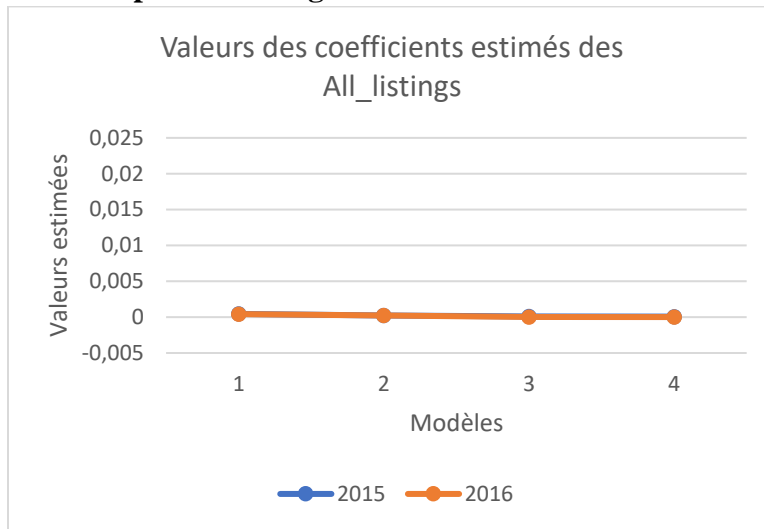
Période 2010-2016 – Segment 4 Chambres

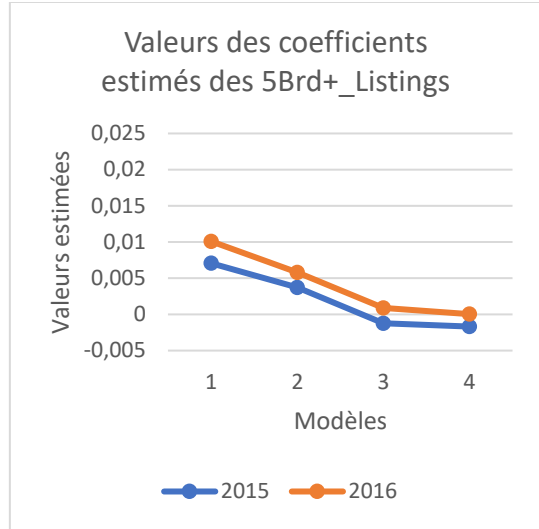
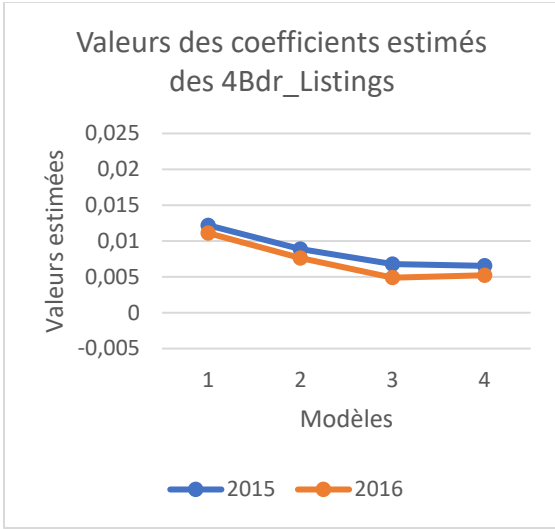
	(1)	(2)	(3)	(4)
	L_ZHVI_4bdr	L_ZHVI_4bdr	L_ZHVI_4bdr	L_ZHVI_4bdr
BnB_4bdr	0.0111*** [0.00144]	0.00764*** [0.00114]	0.00489** [0.00179]	0.00521*** [0.00146]
UN_rate	-0.0455*** [0.00380]	-0.0238*** [0.00717]	-0.00690 [0.00913]	-0.00609 [0.00793]
Lpopulation		0.216 [0.219]		0.0264 [0.255]
LMed_Inc		0.207* [0.0964]		0.259** [0.0900]
L_SpendA		0.698** [0.245]		0.933** [0.391]
L_SpendH		-0.276 [0.361]		-0.116 [0.416]
SAH1		-0.174*** [0.0176]		-0.0856*** [0.00866]
Occup_House		0.000825 [0.00295]		-0.000835 [0.00242]
Int_rate		0.0924*** [0.0154]		0.0766*** [0.00801]
_IYear_2011			-0.0730*** [0.0138]	-0.108** [0.0410]
_IYear_2012			-0.0803*** [0.0238]	-0.127 [0.106]
_IYear_2013			0.0162 [0.0339]	-0.106 [0.146]
_IYear_2014			0.0978** [0.0396]	-0.131 [0.200]
_IYear_2015			0.151** [0.0515]	-0.128 [0.255]
_IYear_2016			0.192** [0.0682]	-0.122 [0.287]
time_trend			0.00166 [0.00111]	0.00351*** [0.000940]
_cons	13.75*** [0.0245]	5.824*** [1.609]	13.48*** [0.0900]	5.192 [5.365]
N	1008	936	1008	936

Standard errors in brackets

* p<.1, ** p<.05, *** p<.01

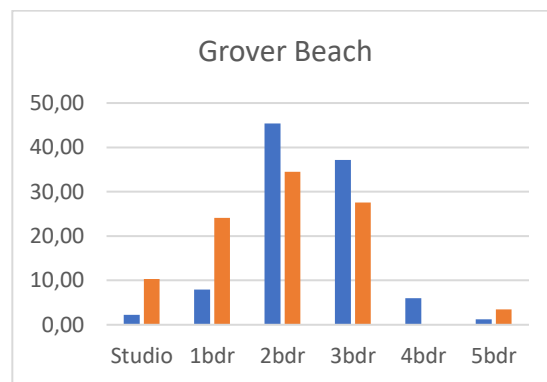
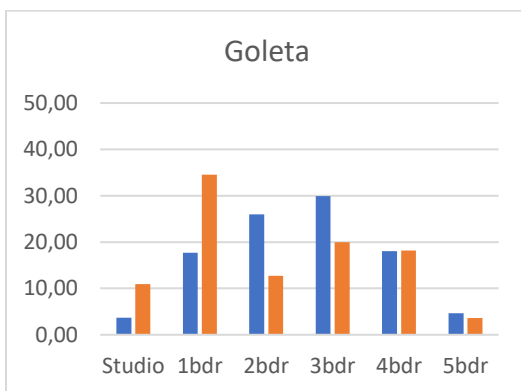
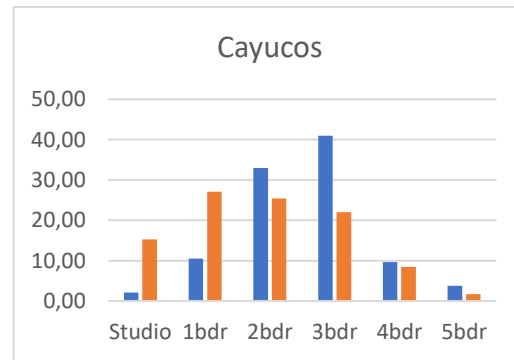
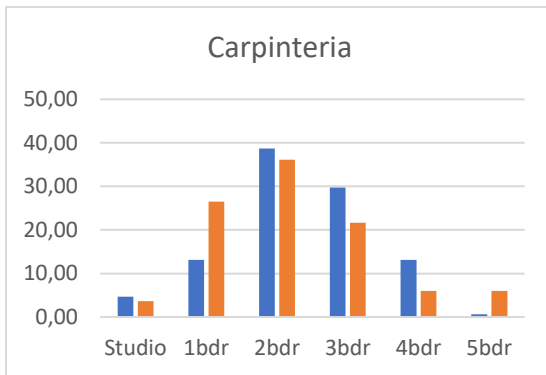
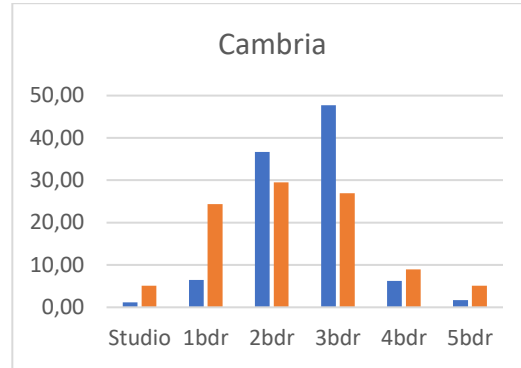
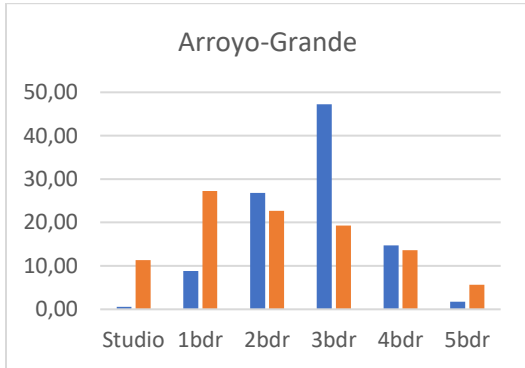
ANNEXE C : Valeurs des coefficients estimés pour les *Listings* d'Airbnb par segment sans tenir compte de leur significativité.

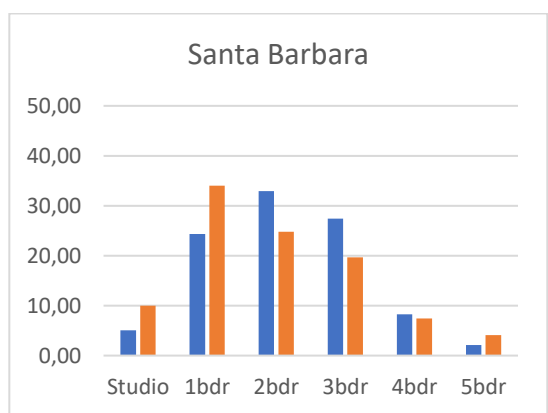
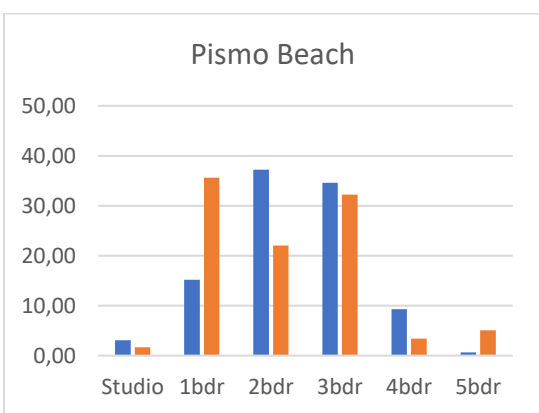
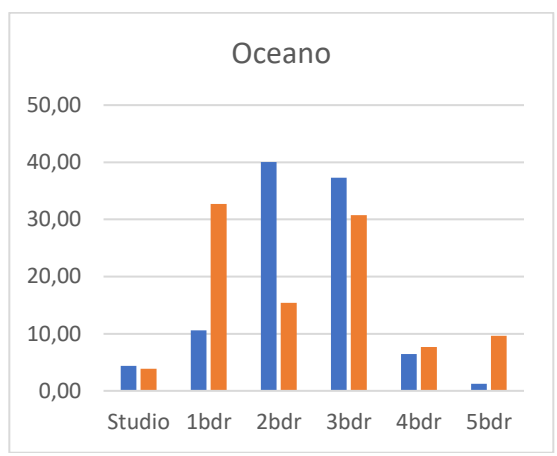
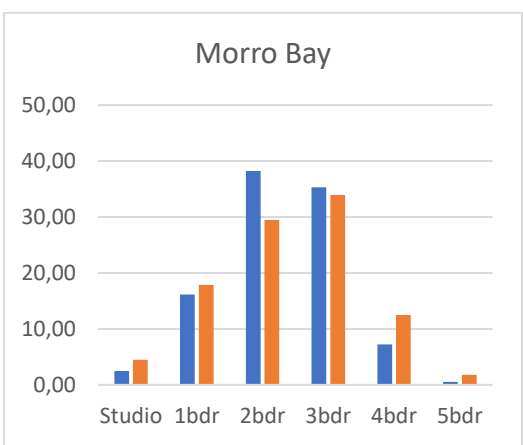
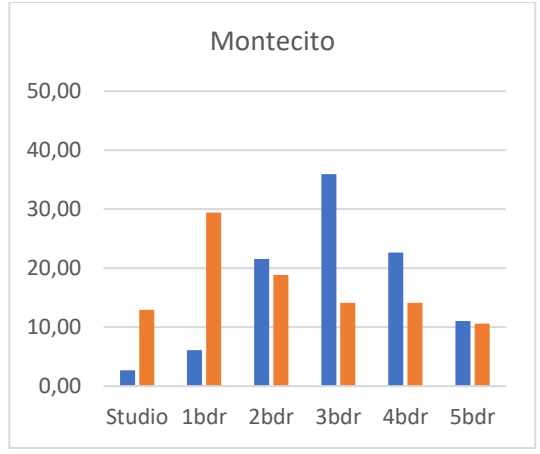
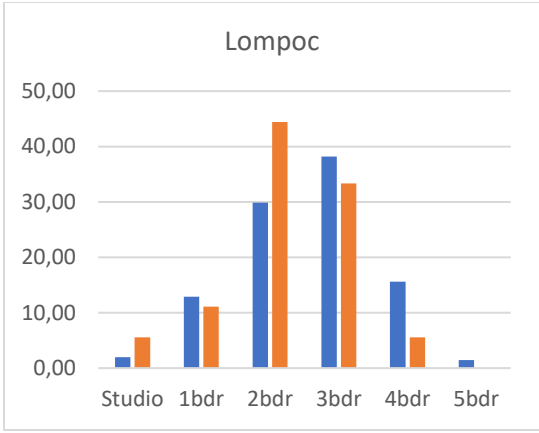




ANNEXE D : Distributions des logements par ville et par type en % et distribution des annonces Airbnb par ville et par type en %.

Les axes des abscisses représentent le type de logements, les axes des ordonnées les proportions en %

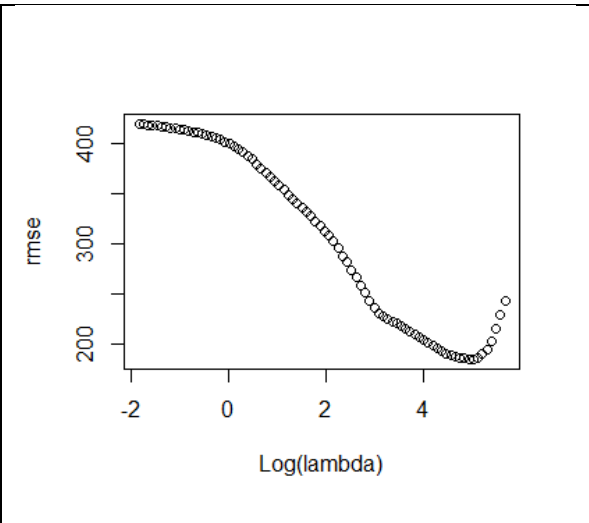
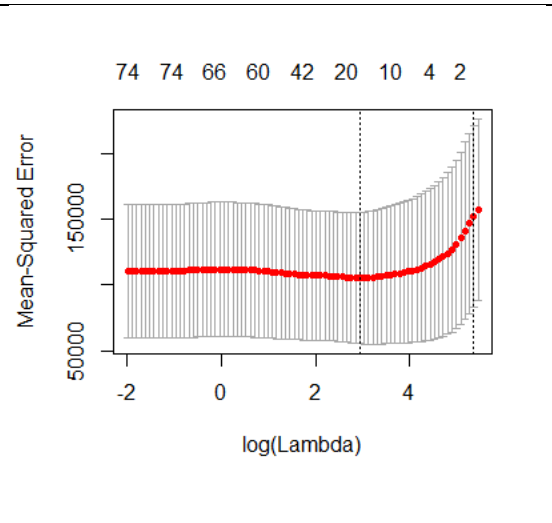




■ Offre
 ■ airbnb

ANNEXE E : Calcul du Lambda pour les méthodes de LASSO.

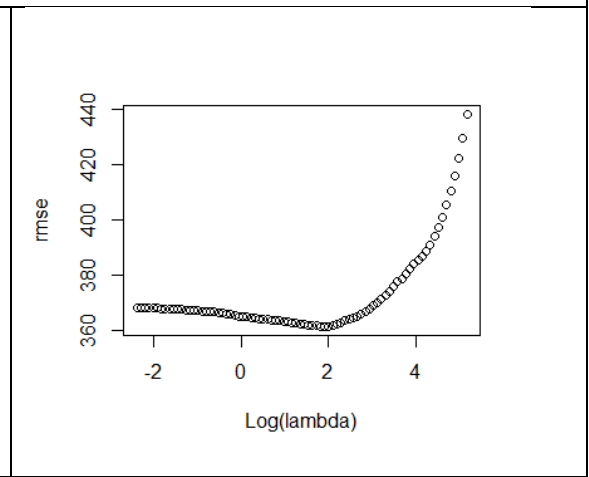
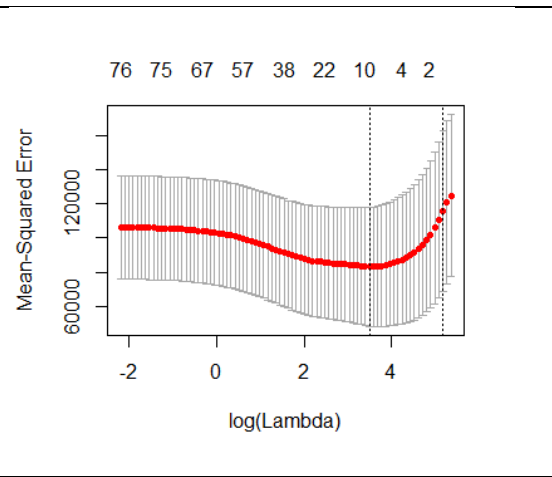
Pour le jeu de données 1.



Calcul du Lambda pour la régression de LASSO

Calcul du Lambda pour la régression de LASSO Machine Learning

Pour le jeu de données 2



Calcul du Lambda pour la régression de LASSO

Calcul du Lambda pour la régression de LASSO Machine Learning

Pour le jeu de données 3

