

Rapport – Analyse Airbnb Seattle

Amine Halhali, Ismail Taaissat, Mohammed-Ibrahim Benahmida, Anas Imarrighen

2025-11-02

Contents

1	Introduction	2
1.1	Contexte de l'étude	2
1.2	Objectifs	2
2	Description du jeu de données	2
2.1	Variables disponibles	2
2.2	Description des données	3
3	Analyse descriptive	3
3.1	Statistiques de base	3
3.2	Distribution des variables clés	5
4	Analyse exploratoire bivariée	7
4.1	Corrélations entre variables numériques	7
4.2	Comparaison des prix selon des catégories (t-test, ANOVA)	7
4.3	Visualisations : scatterplots, boxplots, heatmaps	8
5	Modélisation	8
5.1	Régression linéaire multiple	8
5.2	Régression logarithmique	8
6	Analyse géographique	9
6.1	Cartographie des prix par quartier ou coordonnées	9
6.2	Cartes de densité	10
6.3	Comparaison géographique des prix	11

7	Interprétation et conclusions	11
7.1	Facteurs influençant le prix	11
7.2	Quartiers ou caractéristiques les plus chers/abordables	12
7.3	Limites de l'analyse et recommandations	12

1 Introduction

1.1 Contexte de l'étude

Le marché des locations de courte durée via des plateformes comme Airbnb connaît un développement rapide dans les grandes villes. Comprendre les facteurs qui influencent le prix d'une location est essentiel pour les propriétaires souhaitant optimiser leurs revenus, mais aussi pour les voyageurs souhaitant comparer les offres.

1.2 Objectifs

L'objectif de cette analyse est donc de pouvoir observer le lien de corrélation et/ou de causalité entre les paramètres mis à disposition dans le jeu de données. Nous nous attarderons sur les points importants comme le prix et les avis comme base pour notre étude afin de bien répondre à la problématique: Quels facteurs influencent le prix d'une location Airbnb ?

2 Description du jeu de données

2.1 Variables disponibles

On compte pas moins de 17 paramètres dans le jeu de données:

- room_id
- host_id
- room_type
- adress
- reviews
- overall_satisfaction
- accommodates
- bedrooms
- bathrooms
- price
- last_modified
- latitude
- longitude
- location
- name
- currency
- rate_type

2.2 Description des données

room_id : cet élément représente l'identifiant du rbnb à louer

host_id : cet élément représente l'identifiant de la personne qui loue le rbnb

room_type : cet élément représente le type de rbnb

adress : cet élément représente l'adresse du rbnb

reviews : cet élément représente le nombre d'avis sur le rbnb

overall_satisfaction : cet élément représente la note moyenne du rbnb

accommodates : cet élément représente la capacité de personne pouvant être dans le rbnb

bedrooms : cet élément représente le nombre de chambre du rbnb

bathrooms : cet élément représente le nombre de salle de bain du rbnb

price : cet élément représente le prix en fonction du rate_type et de currency

last_modified : cet élément représente la dernière modification du rbnb

latitude : cet élément représente la latitude du rbnb

longitude : cet élément représente la longitude du rbnb

location :

name : cet élément représente le nom du rbnb

currency : cet élément représente la devise utilisé pour payer

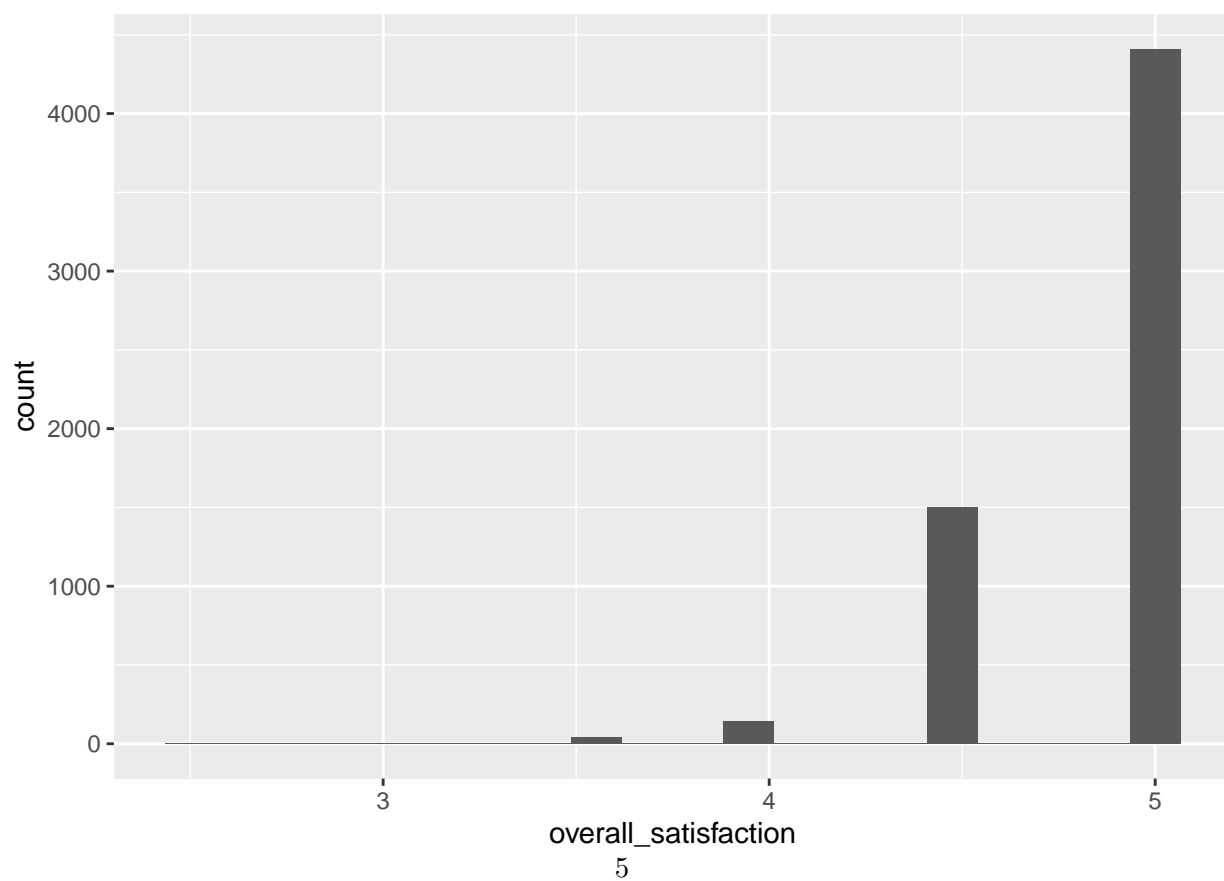
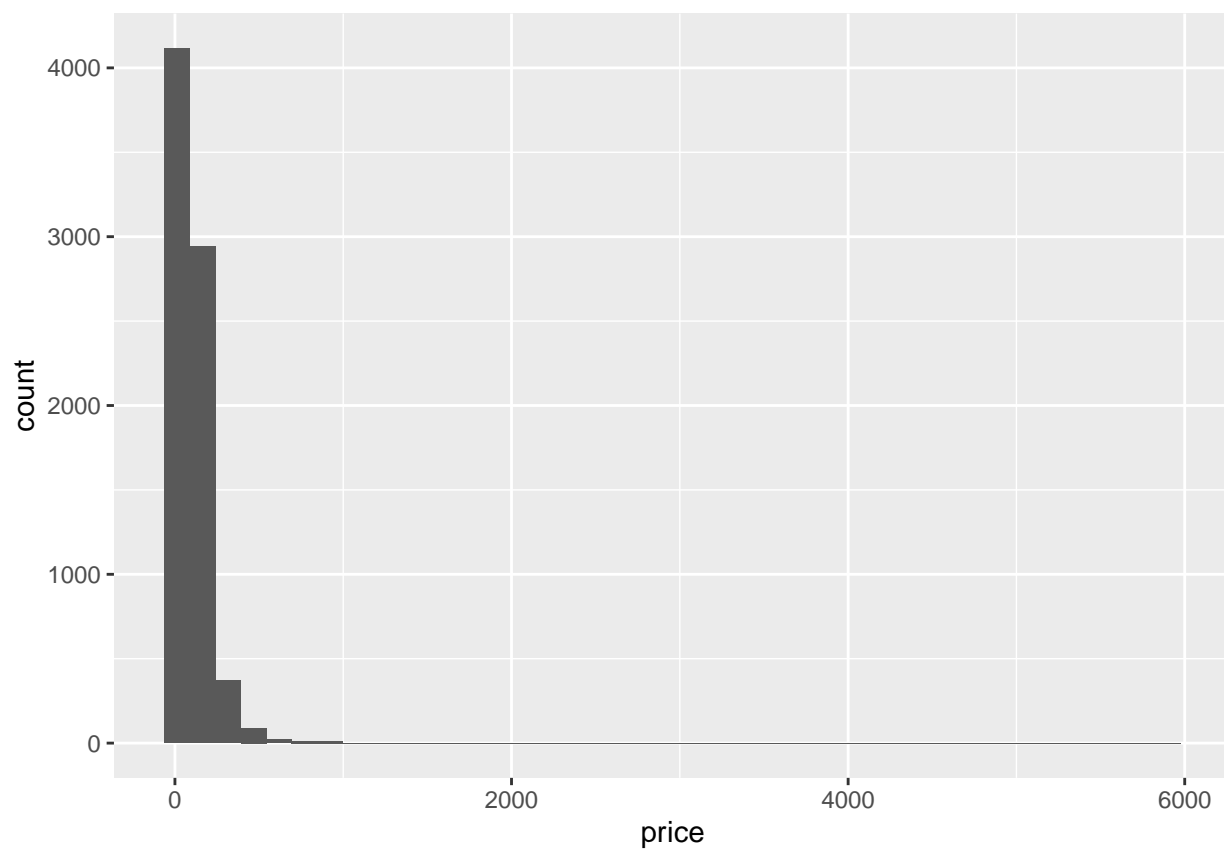
rate_type : cet élément représente le mode de facturation du rbnb

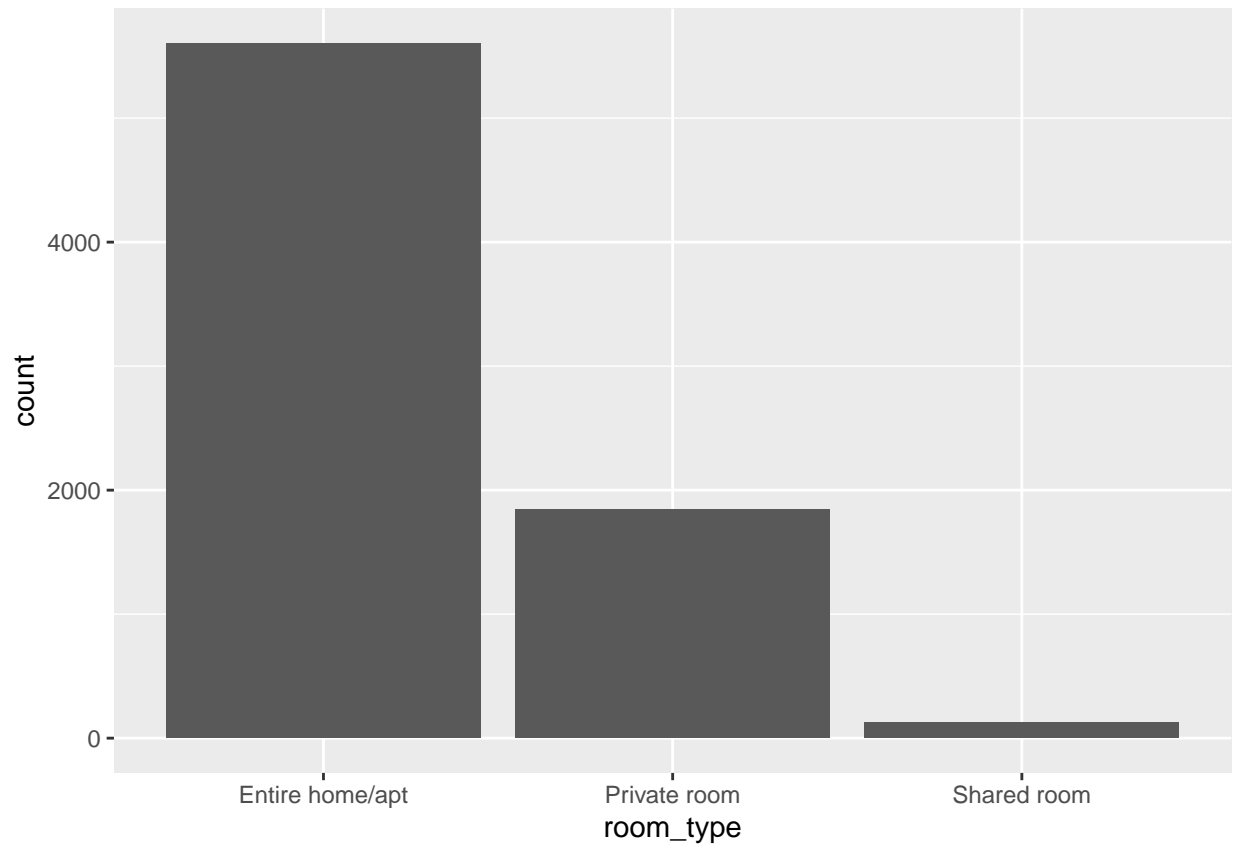
3 Analyse descriptive

3.1 Statistiques de base

min	q1	median	mean	q3	price	api	pop	city	view	new	isv	review	use	virus	sal	can	max	sat	sta	mab	dub	ma	sab	tha	wom	com
15	65	88	113.02	145	900	0	4	21	47.65	8	687	0	2.5	5	4.84	1226	473	3.85	4281	3073	267	3.68	2946			

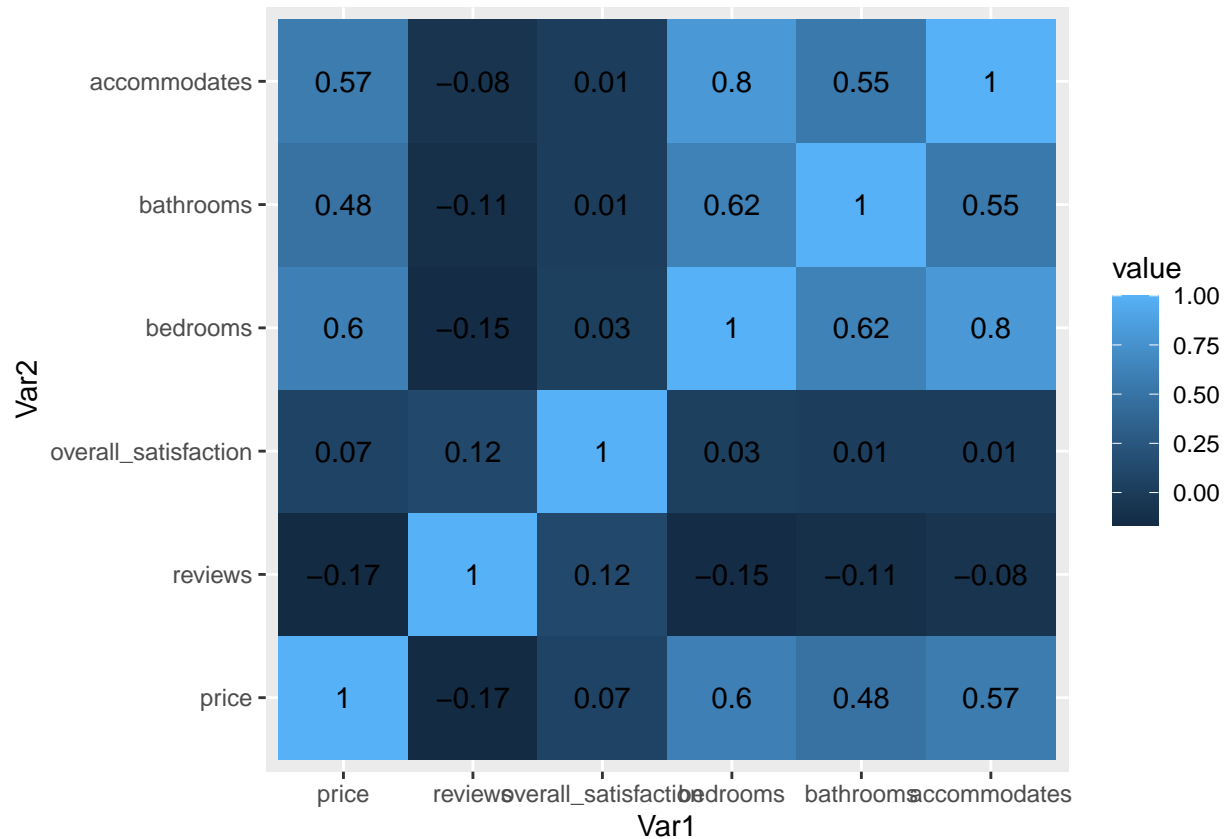
3.2 Distribution des variables clés





4 Analyse exploratoire bivariée

4.1 Corrélations entre variables numériques

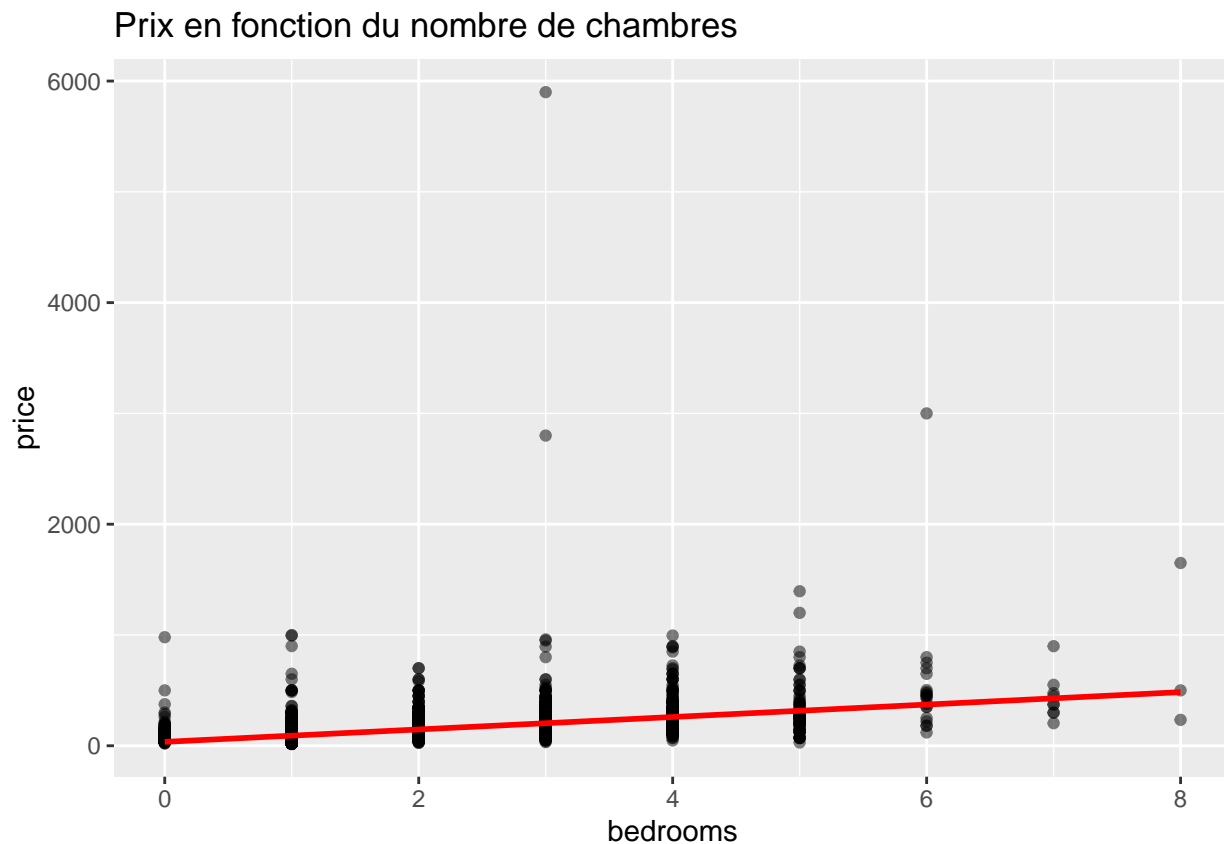


4.2 Comparaison des prix selon des catégories (t-test, ANOVA)

Table 2: Statistiques descriptives des variables

price	reviews	overall_satisfaction	bedrooms	bathrooms	accommodates
Min. : 15	Min. : 0.00	Min. :2.500	Min. :0.000	Min. :0.000	Min. : 1.000
1st Qu.: 65	1st Qu.: 4.00	1st Qu.:4.500	1st Qu.:1.000	1st Qu.:1.000	1st Qu.: 2.000
Median : 88	Median : 21.00	Median :5.000	Median :1.000	Median :1.000	Median : 3.000
Mean : 113	Mean : 47.66	Mean :4.841	Mean :1.385	Mean :1.307	Mean : 3.683
3rd Qu.: 125	3rd Qu.: 64.00	3rd Qu.:5.000	3rd Qu.:2.000	3rd Qu.:1.500	3rd Qu.: 4.000
Max. :5900	Max. :687.00	Max. :5.000	Max. :8.000	Max. :8.000	Max. :28.000
NA	NA	NA's :1473	NA	NA's :2	NA

4.3 Visualisations : scatterplots, boxplots, heatmaps



5 Modélisation

5.1 Régression linéaire multiple

Table 3: Résumé du modèle linéaire

term	estimate	std.error	statistic	p.value
(Intercept)	-72.596	14.182	-5.119	0
reviews	-0.111	0.012	-9.118	0
overall_satisfaction	19.337	2.920	6.623	0
bedrooms	26.307	1.491	17.644	0
bathrooms	21.717	1.718	12.640	0
accommodates	7.964	0.596	13.357	0

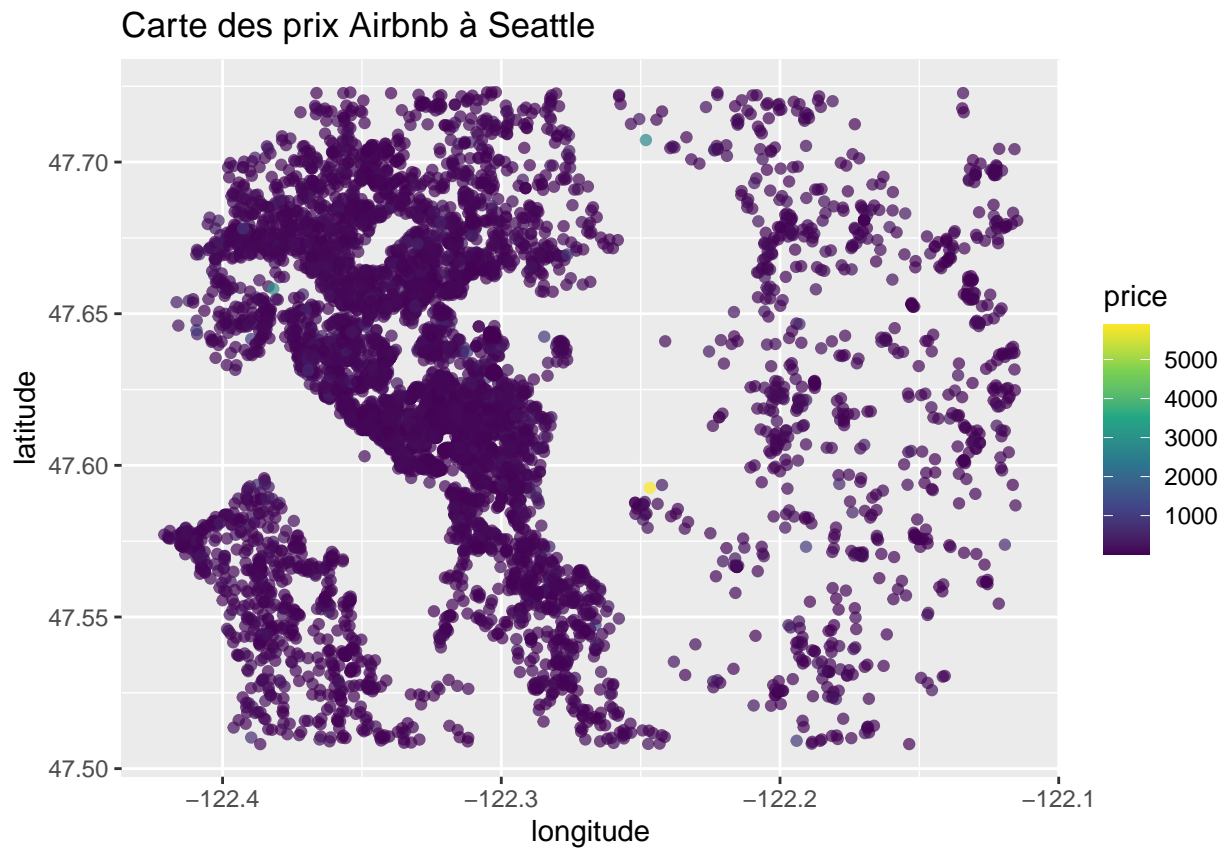
5.2 Régression logarithmique

Table 4: Résumé du modèle linéaire

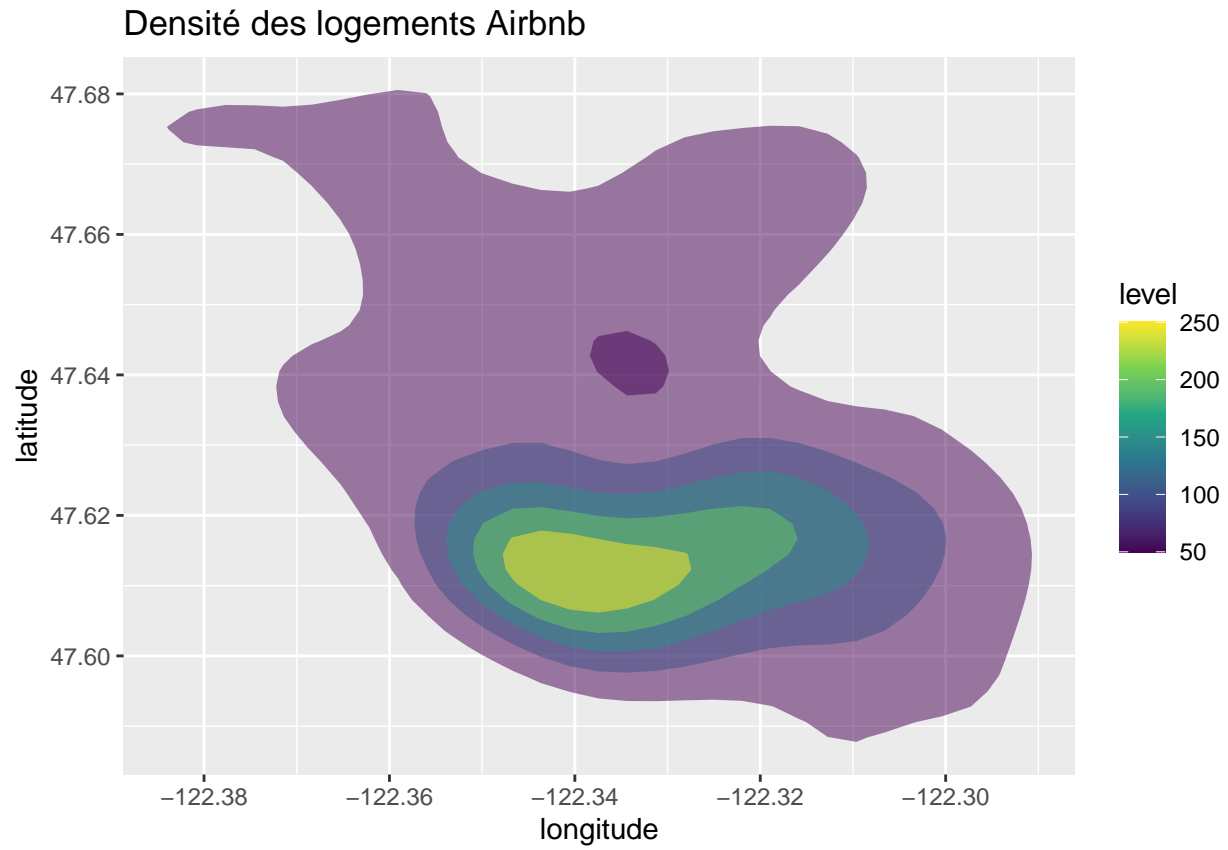
term	estimate	std.error	statistic	p.value
(Intercept)	2.915	0.100	29.052	0.000
reviews	-0.001	0.000	-11.805	0.000
overall_satisfaction	0.220	0.021	10.651	0.000
bedrooms	0.140	0.011	13.296	0.000
bathrooms	0.032	0.012	2.632	0.009
accommodates	0.092	0.004	21.906	0.000

6 Analyse géographique

6.1 Cartographie des prix par quartier ou coordonnées



6.2 Cartes de densité



6.3 Comparaison géographique des prix



7 Interprétation et conclusions

7.1 Facteurs influençant le prix

```
##  
## Call:  
## lm(formula = price ~ reviews + overall_satisfaction + bedrooms +  
##     bathrooms + accommodates, data = airbnb)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -299.40  -32.86   -6.42   20.52 1158.06   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   -72.59598    14.18227   -5.119 3.17e-07 ***  
## reviews       -0.11082     0.01215  -9.118 < 2e-16 ***  
## overall_satisfaction 19.33744     2.91959   6.623 3.81e-11 ***  
## bedrooms       26.30651     1.49096  17.644 < 2e-16 ***
```

```
## bathrooms          21.71726      1.71817  12.640 < 2e-16 ***
## accommodates       7.96399      0.59623  13.357 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.65 on 6095 degrees of freedom
## (1475 observations effacées parce que manquantes)
## Multiple R-squared:  0.411, Adjusted R-squared:  0.4105
## F-statistic: 850.5 on 5 and 6095 DF, p-value: < 2.2e-16
```

7.2 Quartiers ou caractéristiques les plus chers/abordables

location	prix_moyen	nb_annonces
0101000020E6100000CE3461FBC98F5EC0E4F736FDD9CB4740	5900	1
0101000020E6100000976F7D586F985EC04A5E9D6340D44740	3000	1
0101000020E6100000C8B1F50CE18F5EC00EDAAB8F87DA4740	2800	1
0101000020E6100000B0592E1B9D975EC015C440D7BED04740	1650	1
0101000020E61000001F84807C09915EC0F8C3CF7F0FC64740	1395	1
0101000020E61000003B35971B0C975EC035EECD6F98D04740	1200	1
0101000020E6100000D7187442E8965EC086E28E37F9D54740	999	1
0101000020E61000004B5B5CE3338C5EC0DC2BF3565DC94740	995	1
0101000020E61000004E7E8B4E96965EC0E38A8BA372D34740	995	1
0101000020E6100000C1E09A3BFA935EC0ABD0402C9BD14740	980	1

7.3 Limites de l'analyse et recommandations