Projet OPEN DATA 2021 Airbnb

HOUNKONNOU Mehdy
MENENDEZ Benjamin

université de BORDEAUX

Sommaire

université de BORDEAUX

Introduction

Airbnb → Société Américaine fondée en 2008 par Brian Chesky, Nathan Blecharczyk et Joe Gebbia.

Place de marché de la location courte durée via internet.

Objectifs :

- Croiser différentes bases de données pour obtenir de l'information nouvelle.
- Comprendre la façon dont le prix est constitué par les propriétaires
- Extraire et utiliser les sentiments des clients grâce à a leurs avis sur leurs séjour

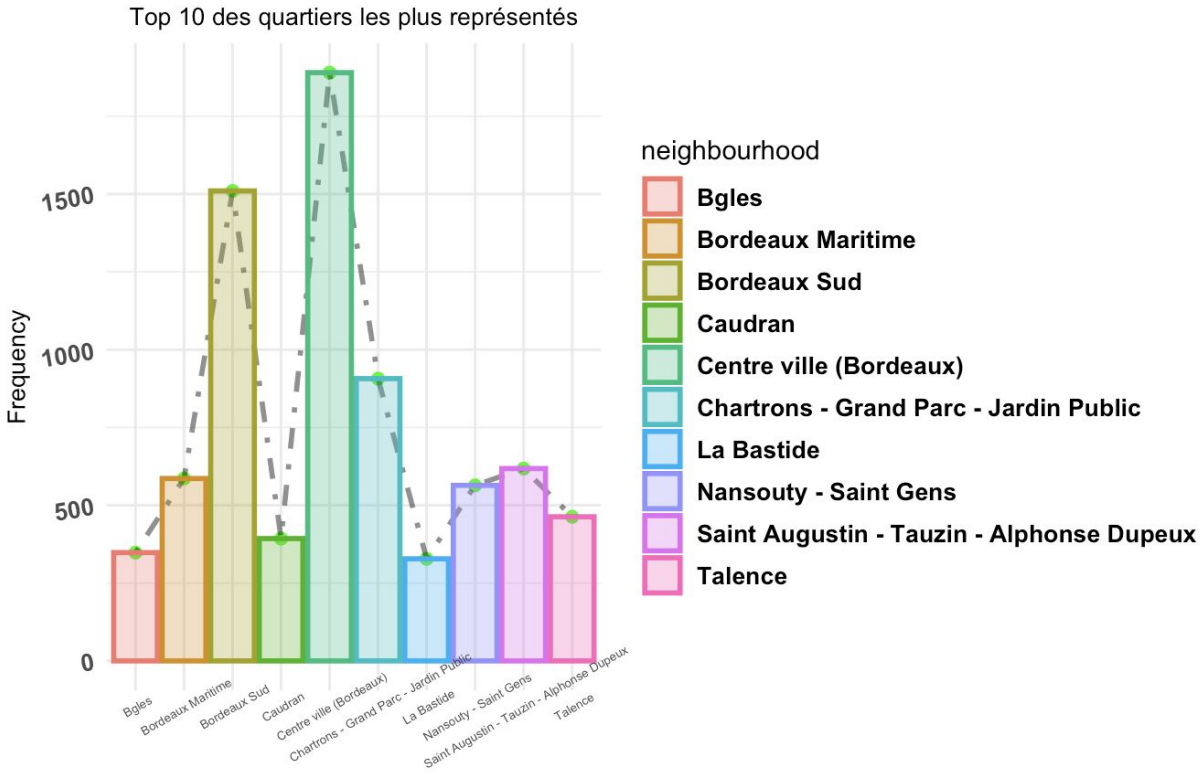
I - Quelques stats descriptives pour de l'analyse

A - Premiers pas... les données

- Données du listing brut des biens Airbnb :
  - 10562 Observations, 74 variables descriptives
  - http://insideairbnb.com/
- Données géographiques pour le marché locatif :
  - https://www.data.gouv.fr/fr/datasets/resultats-nationaux-des-observatoires-locaux-des-loyers/
- Données démographiques pour les villes impliquées :
  - https://www.data.gouv.fr/
- Données utilisateurs, commentaires sur les séjours :
  - 239 922 Observations, 6 variables descriptives
  - http://insideairbnb.com/

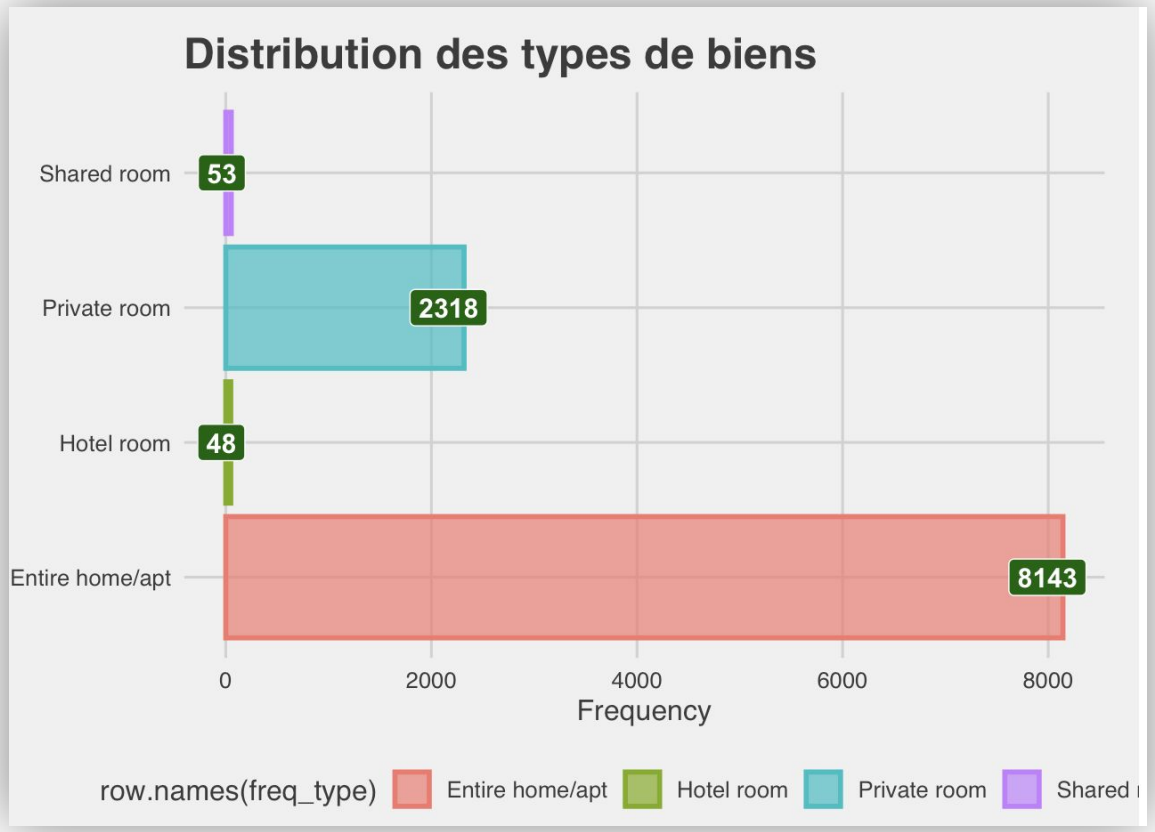
I - Quelques stats descriptives pour de l'analyse

B - Fréquence et distribution



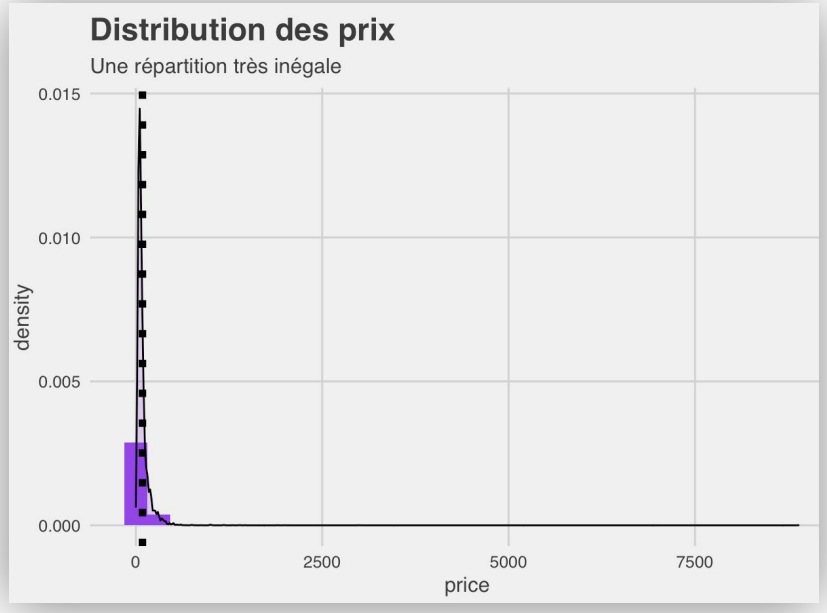
I - Quelques stats descriptives pour de l'analyse

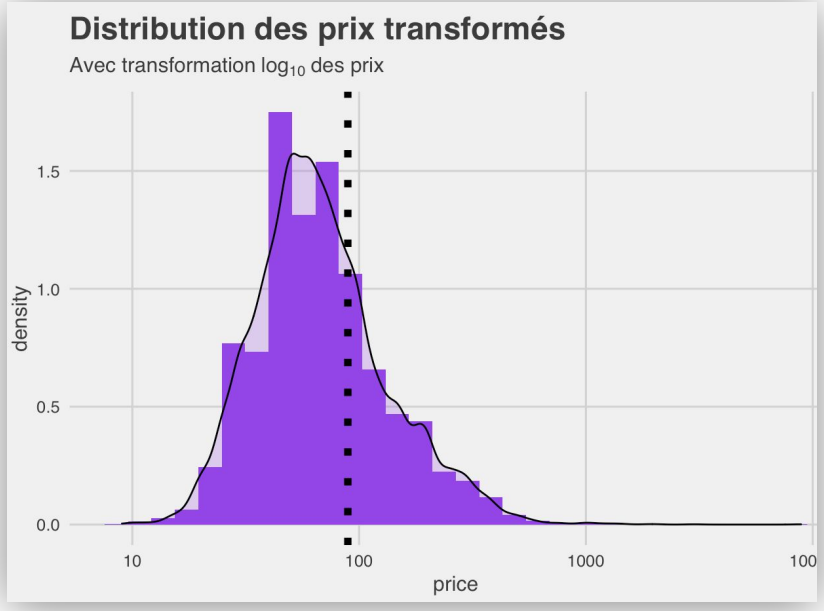
B - Fréquence et distribution



I - Quelques stats descriptives pour de l'analyse

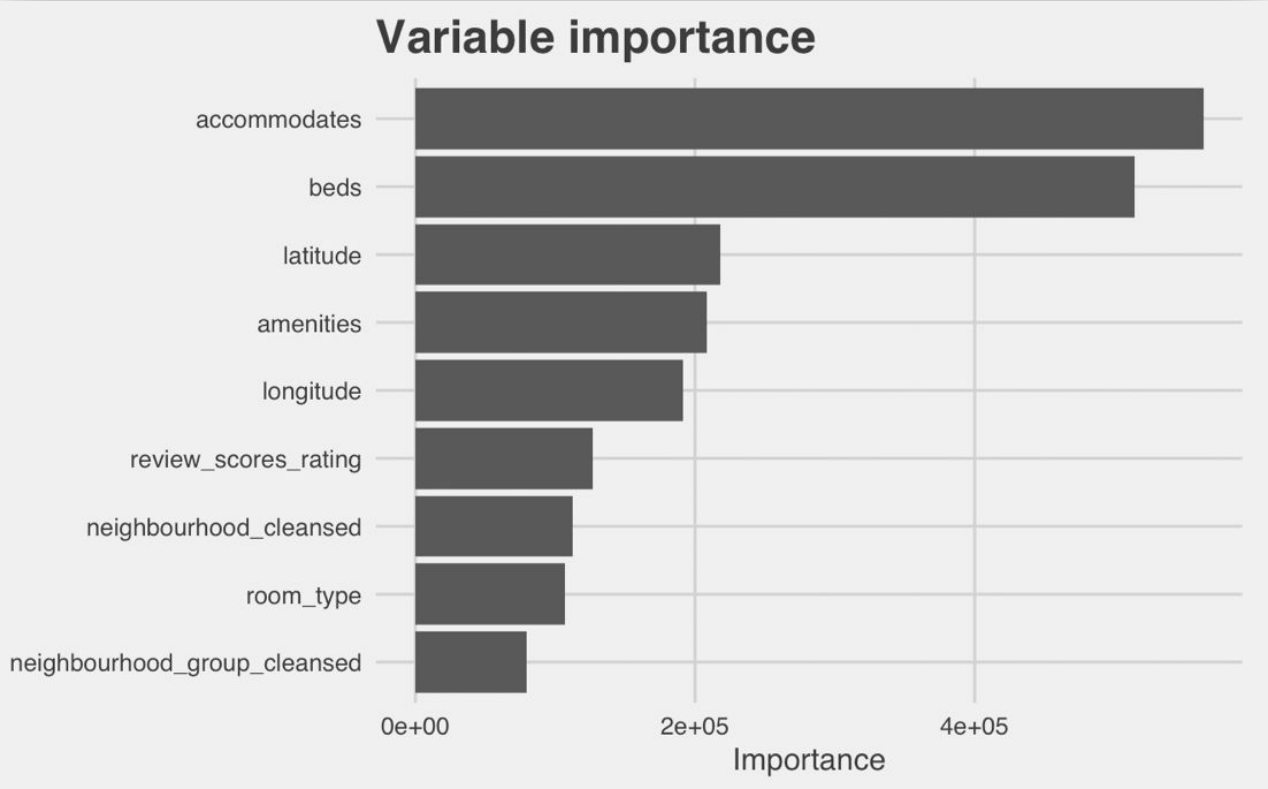
B - Fréquence et distribution





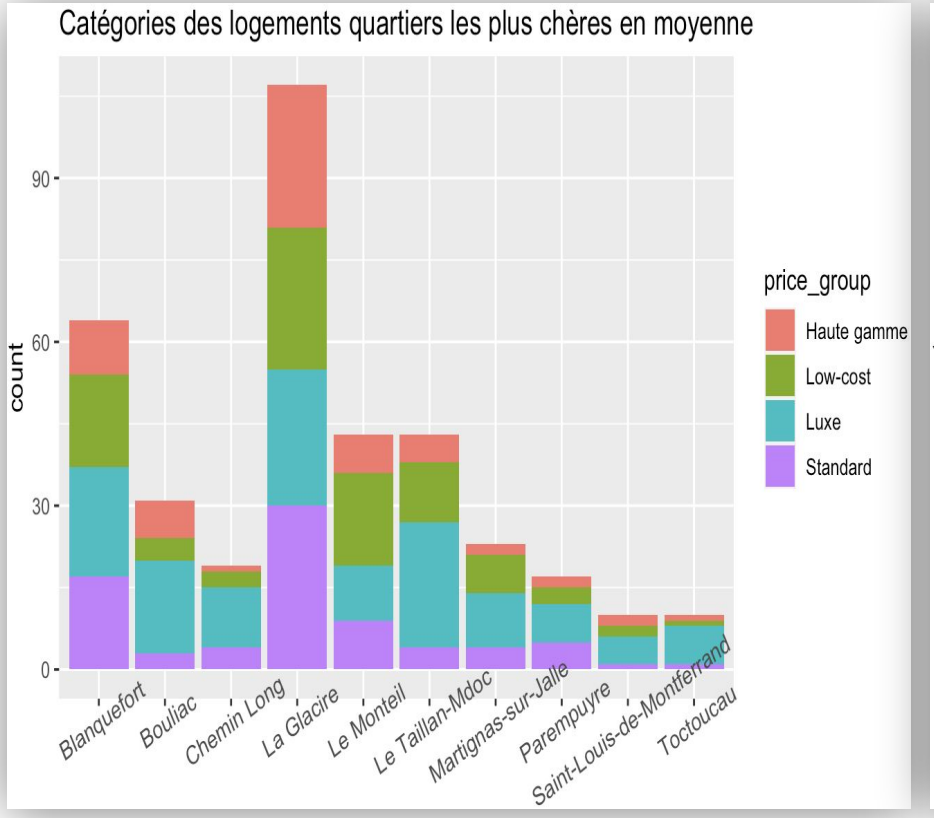
II - Étude complète sur la constitution d'un prix

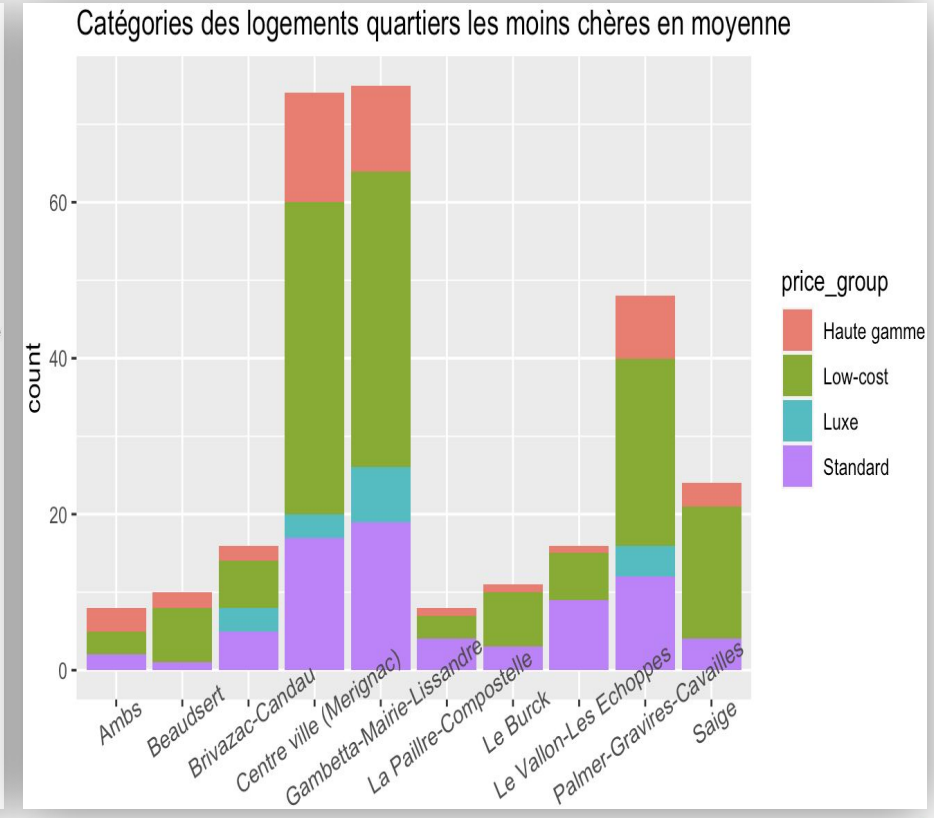
A - Prix en fonction de la localisation, du quartier



II - Étude complète sur la constitution d'un prix

A - Prix en fonction de la localisation, du quartier

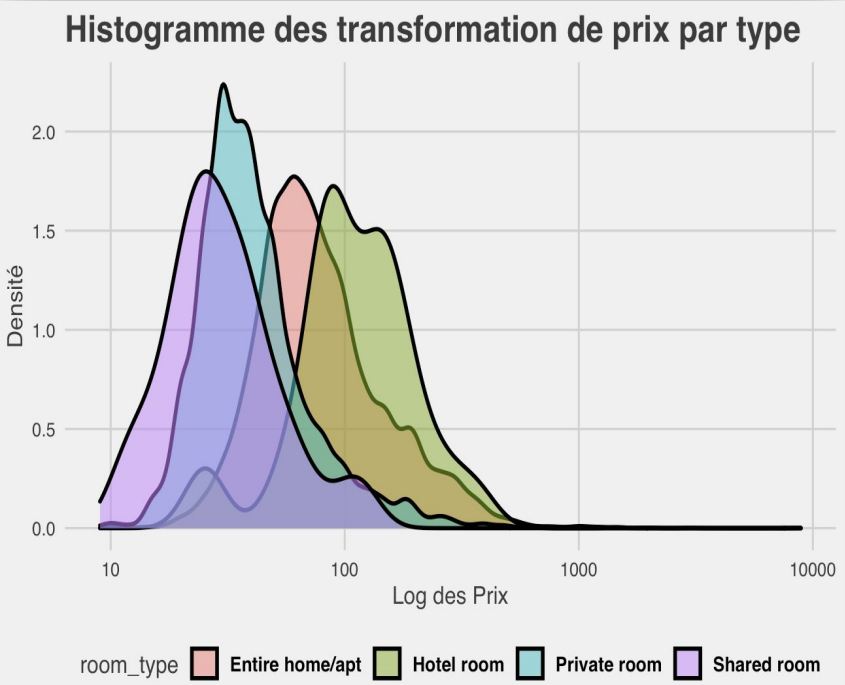


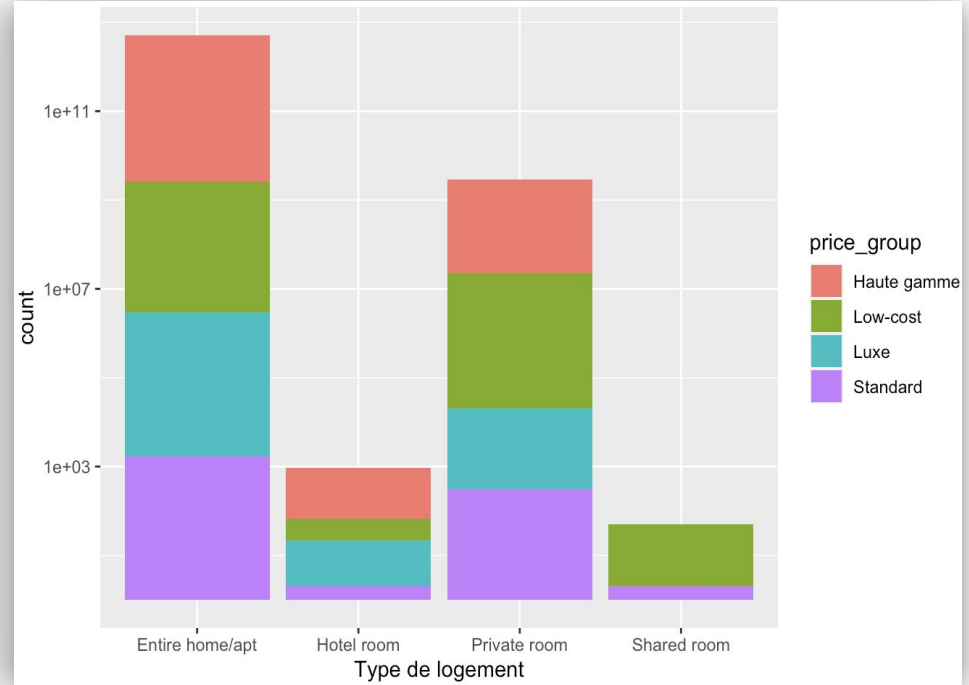


université de BORDEAUX

II - Étude complète sur la constitution d'un prix

B - Prix en fonction du type de biens et capacité d'accueille





université de BORDEAUX

II - Étude complète sur la constitution d'un prix

C - Modélisation statistique pour la variable prix

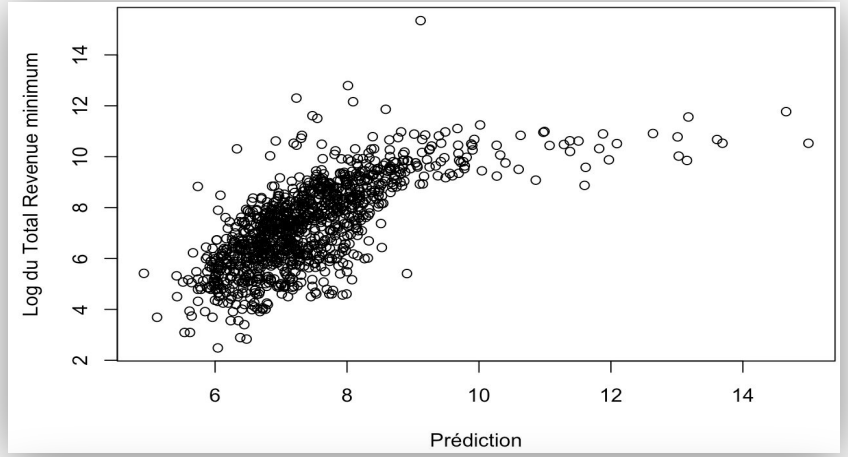
- Régression logistique avec la catégorie des prix comme variable réponse
  - Score de prédiction ⟶ 55%
  - AIC backward
  - Ajout des variables ⟶ Pas d'amélioration de la prédiction :
    - quartier nombre de lits
    - score commentaire client
    - type d'appartement
    - nombre de pièces
- Régression logistique avec les catégories extrêmes "Luxe" et "Low-cost"
  - Score de prédiction ⟶ 92%

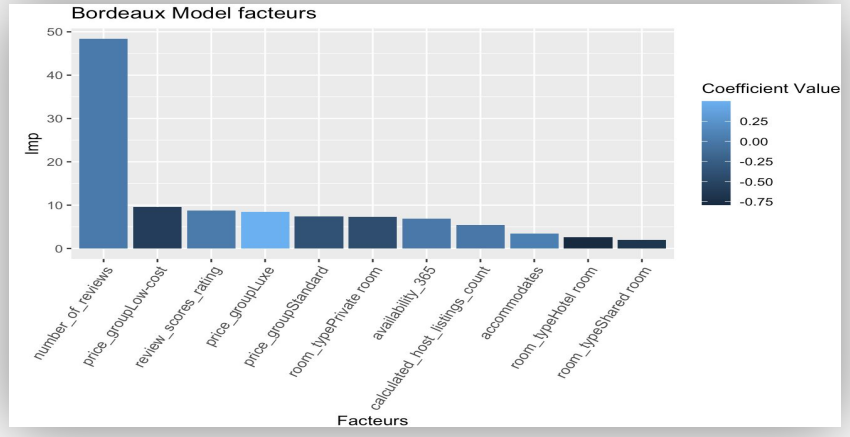
II - Étude complète sur la constitution d'un prix

C - Modélisation statistique pour la variable prix

Création d'une nouvelle variable
"total minimum revenue"

- En utilisant 3 KPI
- Transformation en log
- Adjusted R-squared ⟶ 0,47
- RMSE ⟶ 187703
- Importance des variables





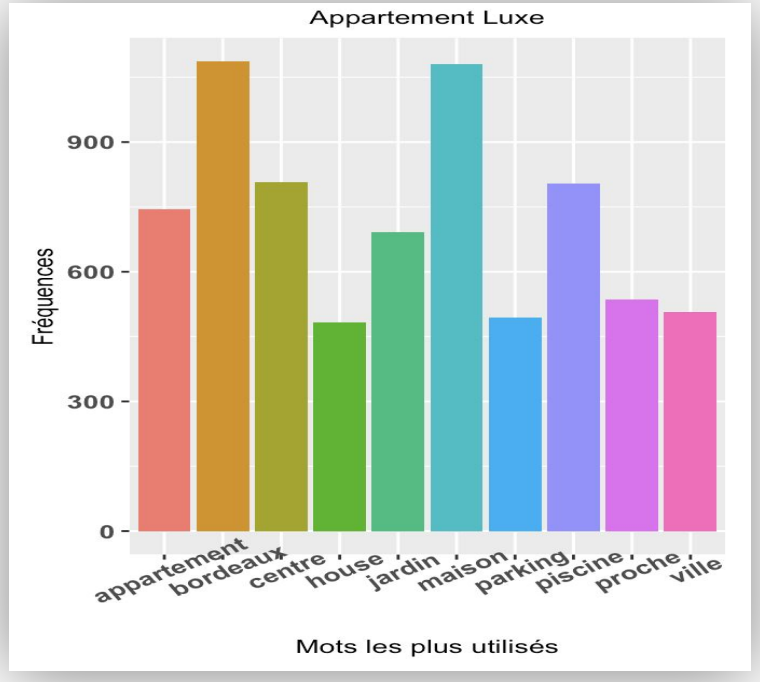
III - Étude des liens avec le marché locatif

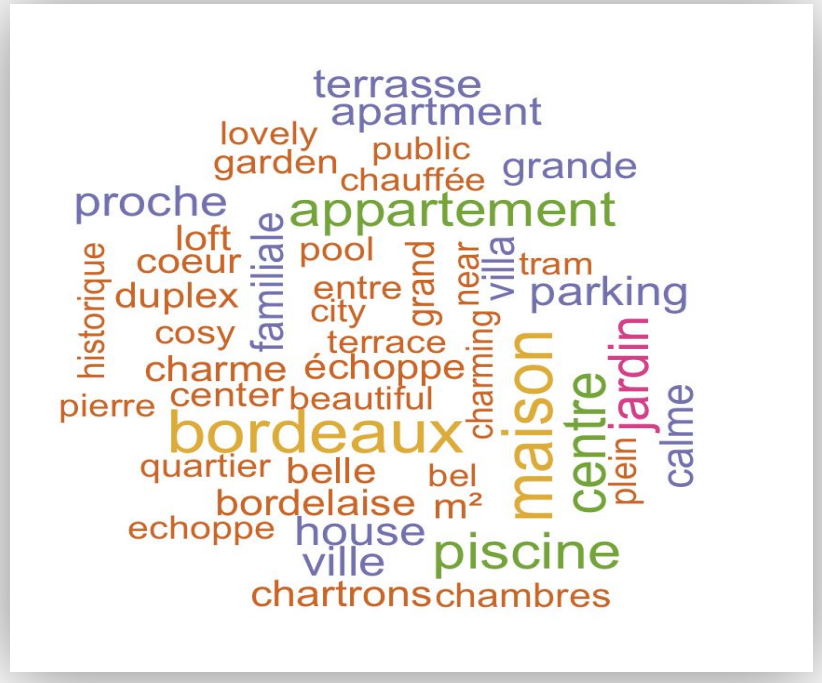
Le prix du marché locatif impact t-il le prix des biens Airbnb ?

- Carte intéractive, superposition du prix au m2 par zone + biens Airbnb par prix de la nuitée
- Toutes les “qualitées” de biens sont représentées dans toutes les zones
- Pour un même prix, la qualitée biens est meilleur dans les zones à faible prix au m2

IV - Traitement naturel du langage pour une étude textuel

A - Étude sur le nom titre des annonces Airbnb



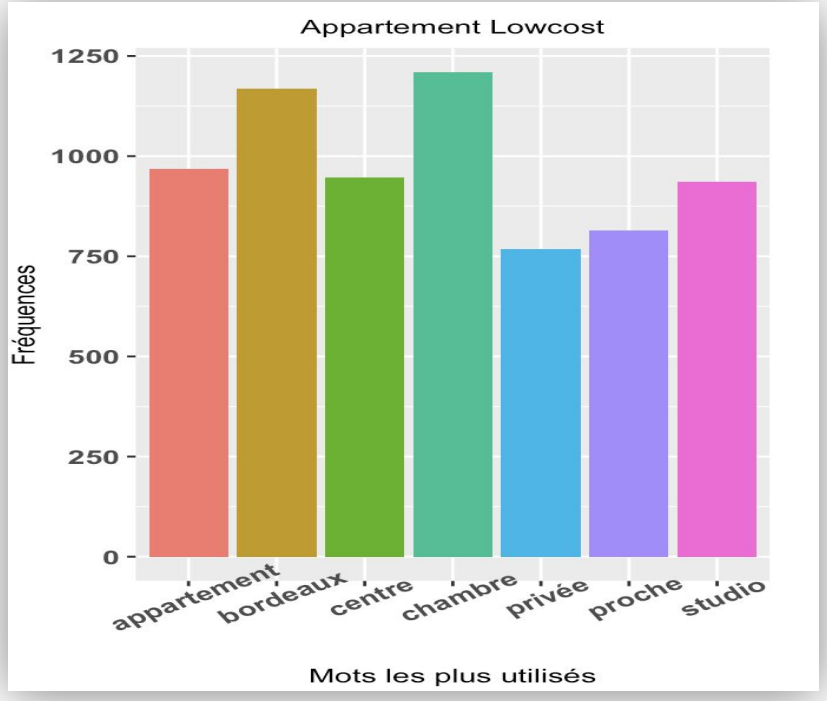


"maison", "piscine", "bordeaux", "villa", "loft", "jardin", "terrasse"

université de BORDEAUX

IV - Traitement naturel du langage pour une étude textuel

A - Étude sur le nom titre des annonces Airbnb

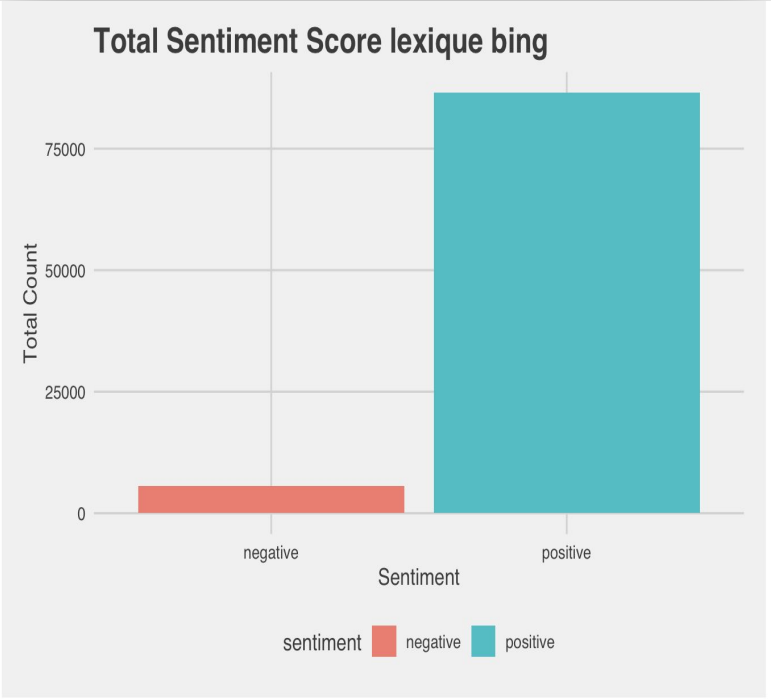


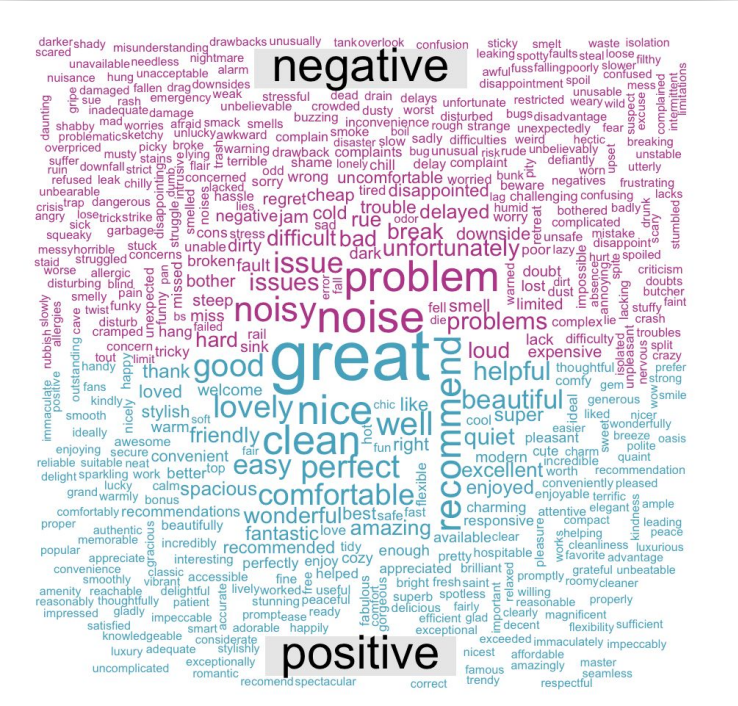
"chambre", "bordeaux", "appartement", "petit", "studio"

université de BORDEAUX

IV - Traitement naturel du langage pour une étude textuel

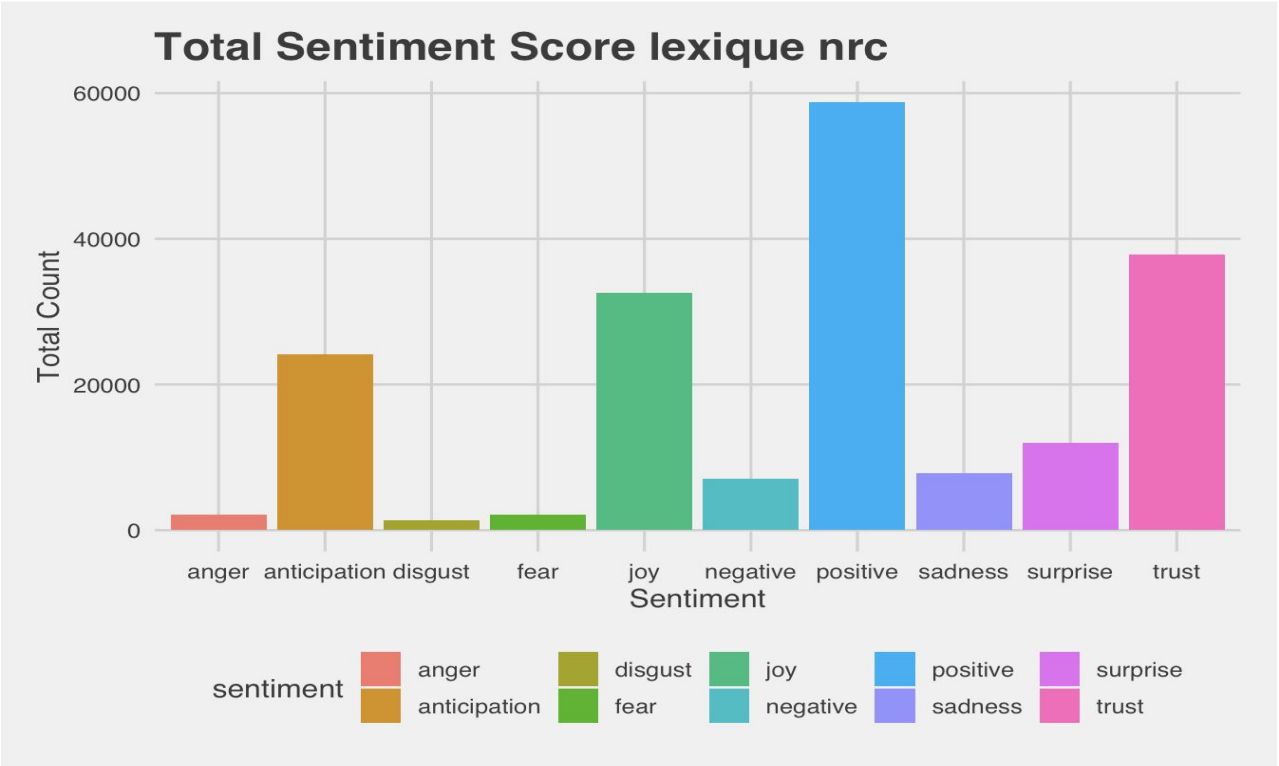
B - Étude et analyse des sentiments des commentaires des clients





IV - Traitement naturel du langage pour une étude textuel

B - Étude et analyse des sentiments des commentaires des clients



IV - Traitement naturel du langage pour une étude textuel

B - Étude et analyse des sentiments des commentaires des clients

Score	Titre de l'annonce	Prix
28.0	Experience a beautiful Bordeaux experience in the heart of Bacalan	75
26.0	Bordeaux, idéal famille !	50
25.0	Belle chambre cosy indépendante à Blanquefort	30
21.0	Maison 6 personnes/piscine collective	100
18.0	suite parentale , 2 chambres	90
18.0	Chartreuse - 75m² av terrasse, ascenseur & parking	182
17.0	Bordeaux centre spacieux T2 49 m2 avec balcon	50
17.0	STUDIO d'architecte : GRAND THEATRE	30
17.0	Maison aux portes de Bordeaux avec piscine	270
16.5	Bordeaux - Appartement calme à vue dégagée	45

Top 10 des meilleurs biens selon les sentiments

IV - Traitement naturel du langage pour une étude textuel

C - Modélisation du score avec les commentaires des clients

Variable synthétique catégorielle à partir des notes de score Airbnb ⟶ "Médiocre", "Moyen", "Haute", "Excellent"

La notes des biens correspond t-elle aux commentaires des usagers ?

Objectif : Prédire la catégorie des notes de score Airbnb en fonction des commentaires des usagers.

Méthodes : Régression logistique (pour la variable catégorielle) // Régression linéaire (pour le score brut de 0-100)

IV - Traitement naturel du langage pour une étude textuel

C - Modélisation du score avec les commentaires des clients

	precision	recall	f1-score	support
Excellent	0.83	0.82	0.83	68108
Haute	0.84	0.73	0.78	56640
Moyen	0.79	0.88	0.84	73808
Médiocre	0.00	0.00	0.00	2
accuracy			0.82	198558
macro avg	0.62	0.61	0.61	198558
weighted avg	0.82	0.82	0.82	198558

```
La performance du modèle sur la base dapprentissage
------------------------------------
Lerreur quadratique moyenne est 1.9670657643739022
le score R2 est 0.6358493633833624


La performance du modèle sur la base de test
------------------------------------
Lerreur quadratique moyenne est 1.9991469536015554
le score R2 est 0.6229673256216983
```

Conclusion

- Croiser les jeu de données pour comprendre ce qui constitue le prix d'une nuité
- Beaucoup de locations dans le centre ville de Bordeaux de façon très diversifié
- Airbnb à une politique de satisfaction complète de sa clientèle (peu d'avis négatif)
- Modélisation de la note d'un bien grâces aux sentiments des commentaires