

Projet introduction au Machine Learning en Python

Modélisation des prix des logements sur Airbnb à Lyon

Jeudi 02 Juin 2022

Projet réalisé par : Gaoussou Diakité, Abir Khan, Anis Louissi, Camil Zahi
Sous la direction de : Guillaume Clément & Paul Thavenot

|



Plan



Objectif



Dataset



Moteur de recherche
avec vue géographique



Modélisation



Prédiction



Objectif et Stratégie

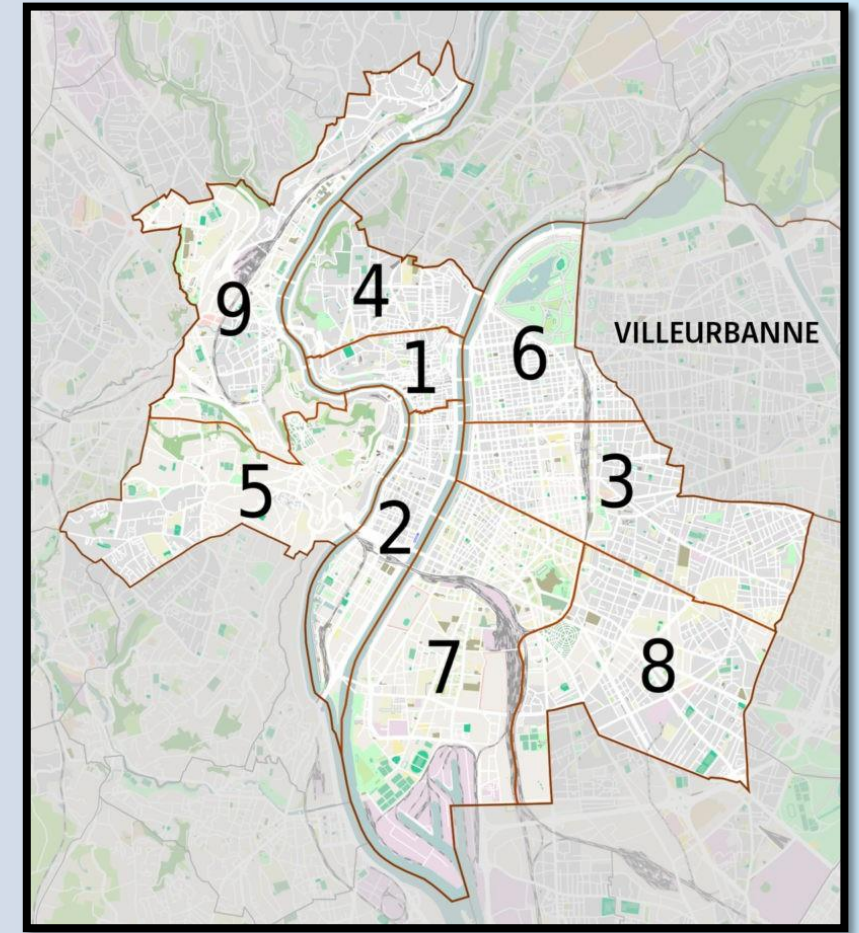
AirBnB est une place de marché pour des locations de courte durée qui permet de lister tous les logements et les mettre en location.

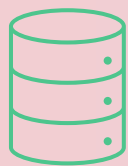
Objectif : Aider les propriétaires qui cherchent à louer leur logement, à déterminer le prix optimal du loyer pour une nuit dans la ville de Lyon sur l'année 2022.

Stratégie :

- trouver quelques annonces similaires aux nôtres ;
- faire la moyenne du prix pour les annonces les plus similaires au nôtre ;
- fixer notre prix de location à ce prix moyen calculé.

Dans notre cas nous souhaitons utiliser les données des locations locales pour prévoir le prix optimal à définir.





Dataset

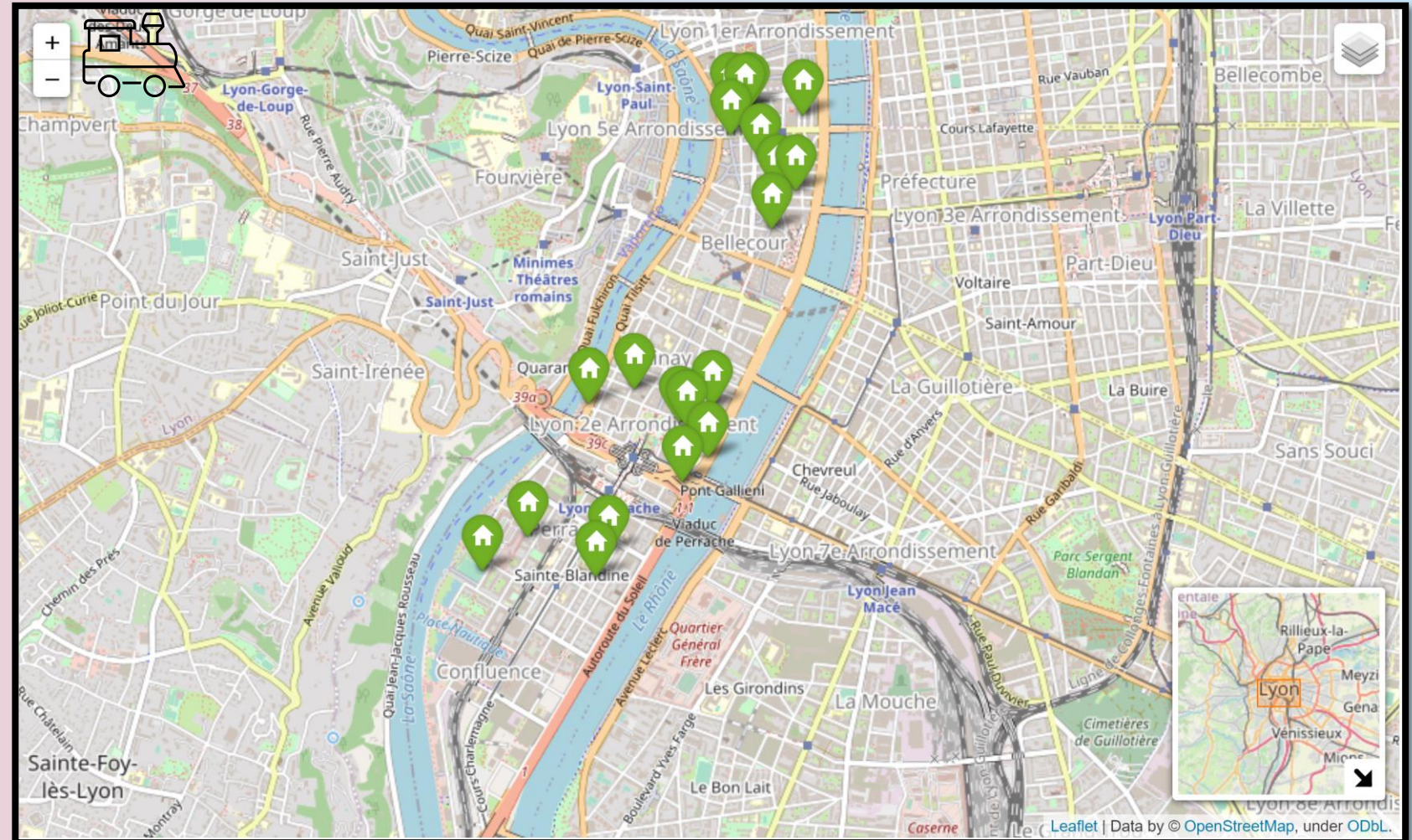
- Lien vers le site de récupération des données (Lyon - Mars 2022) : <http://insideairbnb.com/get-the-data/>
- Annonces des logements présents à **Lyon** entre le **12/03/2022** et **30/03/2023**
- Exemple de données :

	Prix par nuit	Localisation	Latitude	Longitude	Type	Nombre d'invités	Nombre de chambres	Nombre de lits	Badge hôte de qualité	Note moyenne sur 5	...	Chauffage	Climatisation	Ménage	Jardin	Terrasse	Parking	Ascenseur	Cuisine	Micro-ondes	Four
366	270.96	7	45.75366	4.84271	Entire home/apt	8	3	5	Non	-1.00	...	Oui	Non	Non	Non	Non	Non	Oui	Oui	Non	Non
3476	135.48	9	45.80087	4.83374	Entire home/apt	4	1	2	Non	-1.00	...	Oui	Non	Non	Non	Non	Non	Oui	Oui	Non	Non
3744	37.03	6	45.76454	4.86883	Private room	2	1	1	Oui	4.96	...	Oui	Non	Non	Non	Non	Non	Non	Non	Non	Oui
279	76.77	7	45.74913	4.83858	Private room	2	1	1	Non	-1.00	...	Oui	Non	Non	Non	Non	Non	Oui	Oui	Non	Non
529	47.87	2	45.75676	4.83009	Entire home/apt	2	1	1	Non	4.53	...	Oui	Non	Non	Non	Non	Non	Oui	Oui	Non	Non
4237	43.35	8	45.73717	4.86098	Entire home/apt	2	1	1	Non	4.40	...	Oui	Non	Oui	Non	Non	Oui	Non	Oui	Non	Non
1355	316.12	3	45.75266	4.88610	Entire home/apt	4	2	2	Non	-1.00	...	Oui	Non	Non	Non	Non	Oui	Oui	Oui	Non	Non
2930	40.64	8	45.73960	4.86043	Entire home/apt	4	-1	2	Non	4.56	...	Oui	Non	Non	Non	Non	Non	Non	Oui	Oui	Oui
385	45.16	7	45.75308	4.84332	Private room	2	1	1	Non	5.00	...	Oui	Non	Non	Non	Non	Non	Non	Oui	Non	Non
2496	54.19	1	45.77119	4.82631	Private room	1	1	1	Non	5.00	...	Oui	Non	Non	Non	Non	Non	Oui	Non	Non	Non

Moteur de recherche avec vue géographique

Critères :

Prix minimum : 0
Prix maximum : 200
Arrondissement (de 1 à 9) 2
Nombre d'invités minimum : 4
Nombre d'invités maximum : 8
Nombre de chambres minimum : 2
Nombre chambres maximum : 4
Nombre de lits minimum : 4
Nombre lits maximum : 8
Note minimale : 4

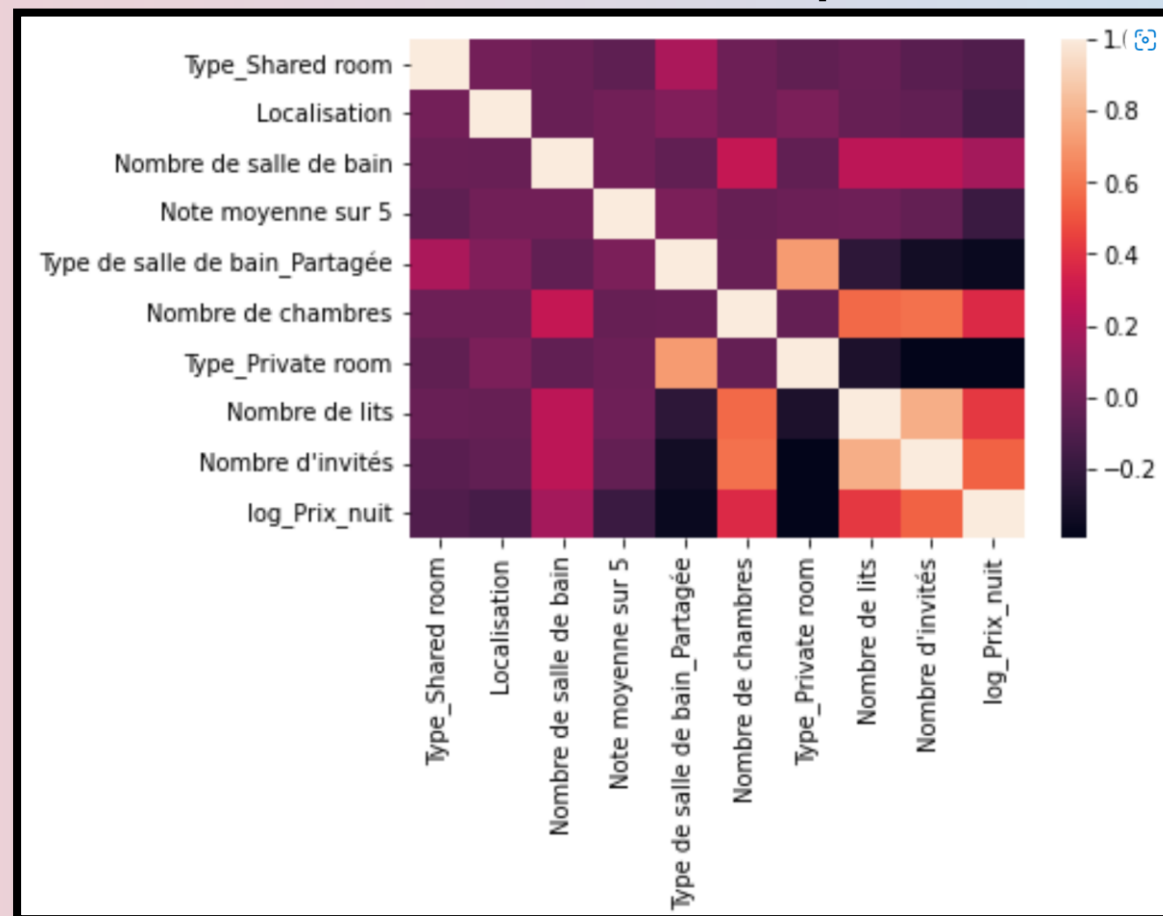


Projet réalisé par : Gaoussou Diakité, Abir Khan, Anis Louissi, Camil Zahi
Sous la direction de : Guillaume Clément & Paul Thavenot



Modélisation

Corrélations entre les variables retenues pour la modélisation :



Projet réalisé par : Gaoussou Diakité, Abir Khan, Anis Louissi, Camil Zahi
Sous la direction de : Guillaume Clément & Paul Thavenot



Régression

Résultats de nos régressions :

	Régression linéaire	Régression Ridge	XGBOOST	KNN
R2	✓ 0,4132	✓ 0,4132	✗ 0,3614	! 0,3930
MSE	✓ 0,2495	✓ 0,2495	✗ 0,2715	! 0,2581
Graphique				

Classification

Résultats de nos modèles :

**Régression logistique :
Matrice de confusion**

1026	6
80	236

**KNN :
Matrice de confusion**

988	44
78	238

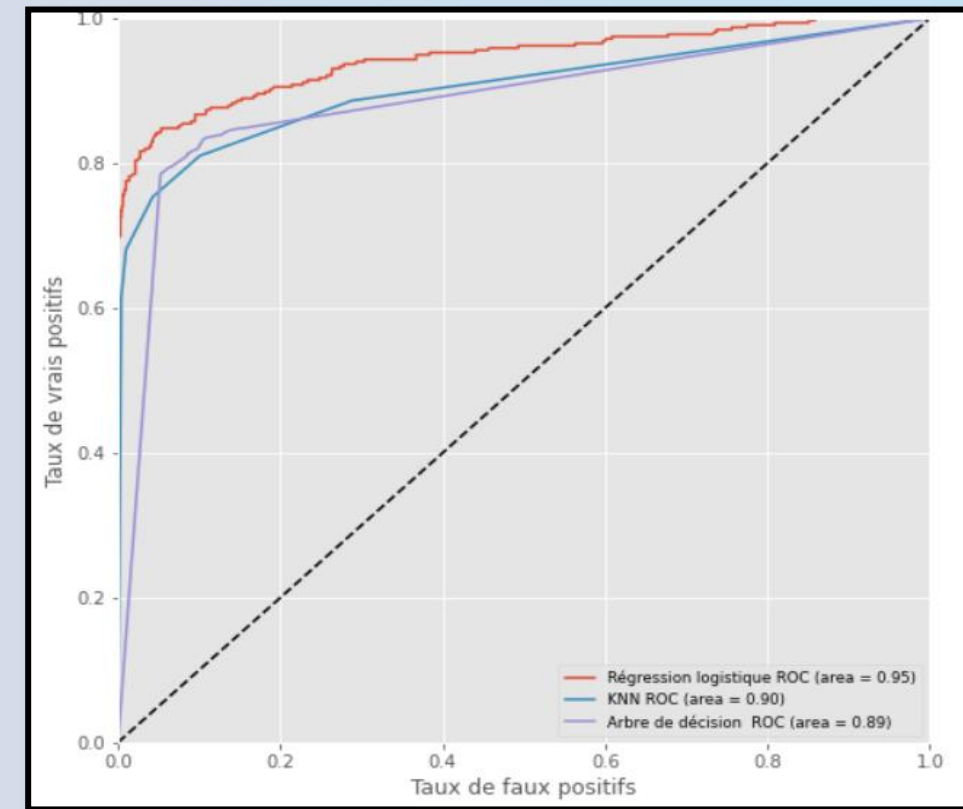
**Arbre de décision :
Matrice de confusion**

967	65
66	250

Scores de la régression logistique :

```
Accuracy 0.9362017804154302
      precision    recall
0      0.93      0.99
1      0.98      0.75
```

Sélection du meilleur
modèle en fonction de l'AUC :





Conclusion

Caractéristiques de notre logement :

```
Arrondissement (de 1 à 9) : 2
Nombre d'invités : 4
Nombre de chambres : 2
Nombre de lits : 3
Nombre de salle de bain : 1
Type_Private room (1 si Oui, 0 si Non) : 1
Type_Shared room (1 si Oui, 0 si Non) : 0
Type de salle de bain_Partagée (1 si Oui, 0 si Non) : 0
Climatisation_Oui (1 si Oui, 0 si Non) : 1
```

Prédiction du prix :

```
En fonction de ses caractéristiques, nous vous conseillons de fixer le prix du logement à : [[144.32]] €
```

Merci pour votre attention !