



Les Plus Beaux Logis de Paris

---

**Analysez l'évolution des prix de l'immobilier avec Python**

*Djibril*

*Soumare*

*Septembre 2023*

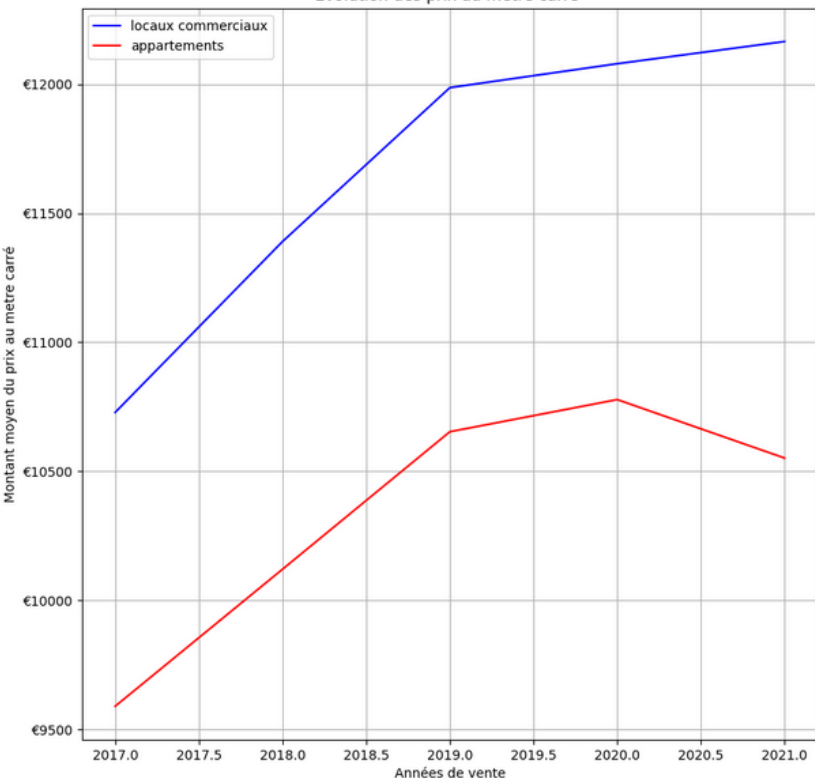
---

# Les Plus Beaux Logis de Paris

## Partie 1

# I. Analyse du marché de l'immobilier

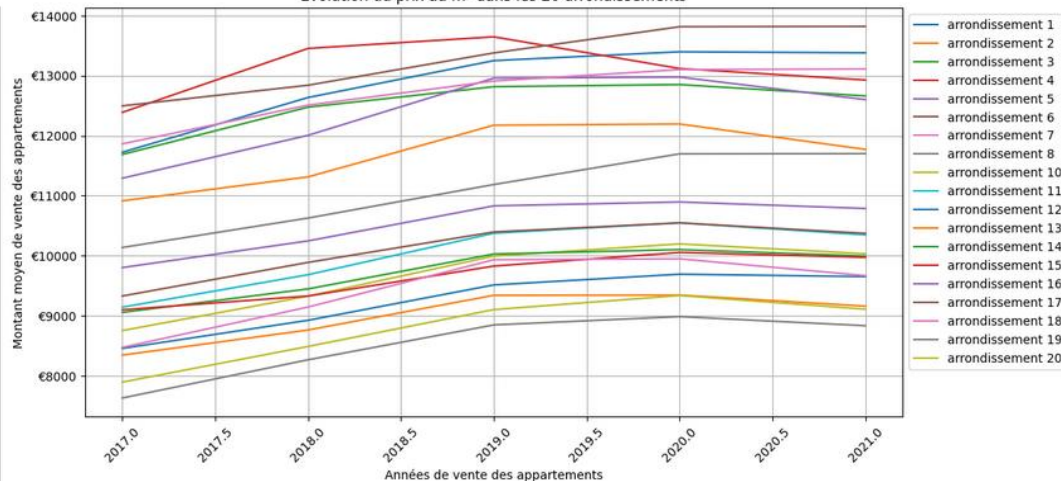
Evolution des prix au metre carré



☑ Transactions pour les appartements : 24353

☑ Transactions pour les locaux commerciaux : 1843

Evolution du prix au m<sup>2</sup> dans les 20 arrondissements



☑ Relation entre le prix au m<sup>2</sup> et la date au 6 arrondissement

Corrélation de Spearman :

R= 0,81 soit 81%

P-value = 1,05

☑ Relation entre la valeur foncière et la surface

R= 0,99

P-value = 0,0

## II. Méthodologie suivie

- ☑ **Les dimensions à utiliser pour prédire le prix au m<sup>2</sup>**
  - ✓ la surface du bien immobilier,
  - ✓ la date considérée,
  - ✓ la localisation (code postal),
  - ✓ le type de bien.
- ☑ **Prédiction de la valeur foncière d'un bien immobilier**
  - ✓ transformant les colonnes catégoriques du code postal et du type de local avec **get\_dummies**
  - ✓ un tiers de nos données (**33%**) et les garder de côté, avec `train_test_split`
  - ✓ algorithme de **régression linéaire**

Erreur moyen en pourcentage = **7,68%**    **R<sup>2</sup> = 98,50 %**

- ☑ Prédiction de la valeur foncière avec fichier de portefeuille des actifs de la société avec le modèle de prédiction.

# III. Résultat des prédictions

*La valorisation au 31 décembre 2022 du portefeuille des actifs immobiliers de l'entreprise.*

Somme/ 1.000.000

- ✓ Valorisation du portefeuille sur le segment corporate : 91,93 millions d'euros
- ✓ Valorisation du portefeuille sur le segment des particuliers : 63,16 millions d'euros

Au regard de la somme de valeurs prédites, c'est le Local industriel. commercial ou assimilé qui la plus grande valeur dont 91.93 millions d'euros soit un écart de 28,77 millions d'euros.

La limite de cette estimation est qu'elle basée sur un seul facteur prix/m2. Il y'a une marge d'erreur environ 7% environ dans le model de prédiction

---

# Les Plus Beaux Logis de Paris

## Partie 2

Classification des données issues du jeu de test

# I. Méthodologie suivie

## ☑ **Labelliser automatiquement les biens immobiliers**

- ✓ soit des Appartements

- ✓ soit des Local industriel. commercial ou assimilé

## ☑ Estimation de nombre de K avec méthode coude en utilisant **KElbowVisualizer**

## ☑ algorithme du KMeans sur le jeu de données partagé par l'entreprise.

## II. Résultat de la classification

- ☑ Données de labellisations correctes des biens immobiliers

type_local	Prix/m <sup>2</sup>		
	mean	min	max
Appartement	7408.775030	7207.21763	7666.071700
Local industriel, commercial ou assimilé	9806.924674	9194.69779	10113.195822

Au regard de la similarité entre l'analyse du résultat de la classification par le modèle KMeans et celle de labellisations correctes des biens immobiliers, on peut dire que les résultats sont satisfaisants.

- ☑ Résultat issu de la classification

cluster	Prix/m <sup>2</sup>		
	mean	min	max
Appartement	7408.775030	7207.21763	7666.071700
Local industriel commercial ou assimilé	9806.924674	9194.69779	10113.195822

Cependant, nous pouvons évaluer la qualité d'une solution de clustering K-Means en utilisant différentes métriques, telles que le score de Silhouette ou l'index Davies-Bouldin, pour évaluer la façon dont les points de données sont groupés en clusters.