
Les Plus Beaux Logis de Paris

Partie 1



Les Plus Beaux Logis de Paris

Analyse de l'évolution des prix de l'immobilier
Simon
Doussin
2025

I. Analyse du marché de l'immobilier

On ne constate ni de doublons, ni de valeurs aberrantes dans les données.

```
df_paris.describe()
```

	date_mutation	valeur_fonciere	adresse_numero	code_postal	code_type_local	surface_reelle
count	26196	2.619600e+04	26196.000000	26196.000000	26196.000000	26196.000000
mean	2019-06-09 14:09:14.099862784	4.916170e+05	47.449572	75012.716216	2.140709	46.859940
min	2017-01-02 00:00:00	8.519470e+04	1.000000	75001.000000	2.000000	11.000000
25%	2018-03-14 00:00:00	2.339439e+05	10.000000	75009.000000	2.000000	23.000000
50%	2019-05-23 00:00:00	3.545774e+05	27.000000	75014.000000	2.000000	35.000000
75%	2020-09-16 00:00:00	5.702197e+05	63.000000	75017.000000	2.000000	55.000000
max	2021-12-31 00:00:00	3.843359e+07	407.000000	75020.000000	4.000000	3945.000000
std	NaN	5.713292e+05	55.733432	5.269150	0.511496	52.408062

2 types de locaux figurent dans le jeu de donnée de Paris : appartement (n = 24 353), et local industriel (n = 1843).

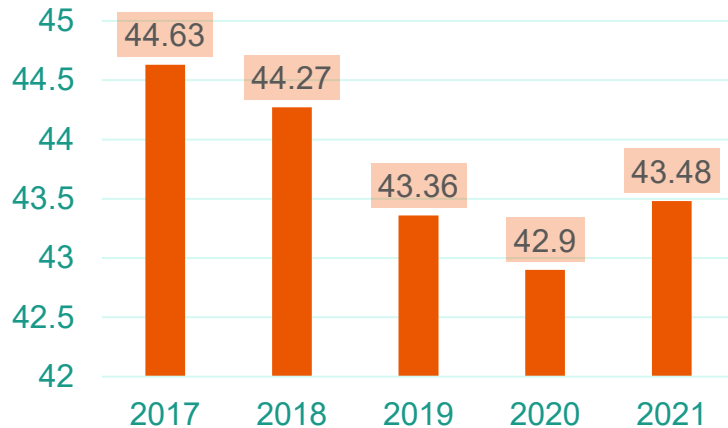
Appartement	2	24353
Local industriel, commercial ou assimilé	4	1843

On compte 26 196 transactions entre 2017 et 2021 à Paris, dont 24 353 pour les appartements et 1 843 pour les locaux commerciaux.

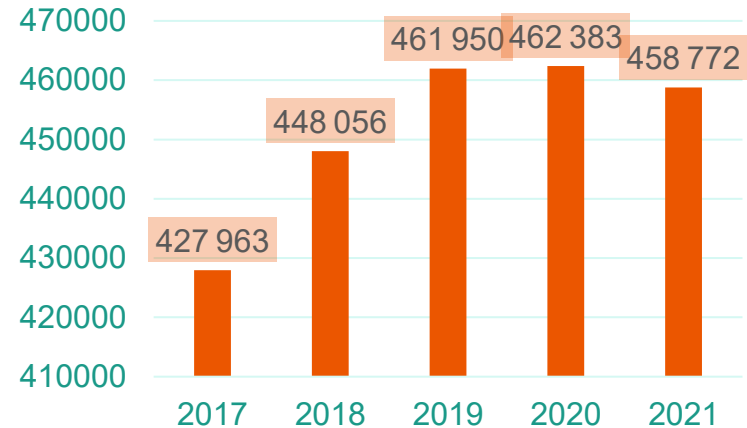
I. Analyse du marché de l'immobilier

Appartement

Surface réelle moyenne (en m²) des appartements selon les années



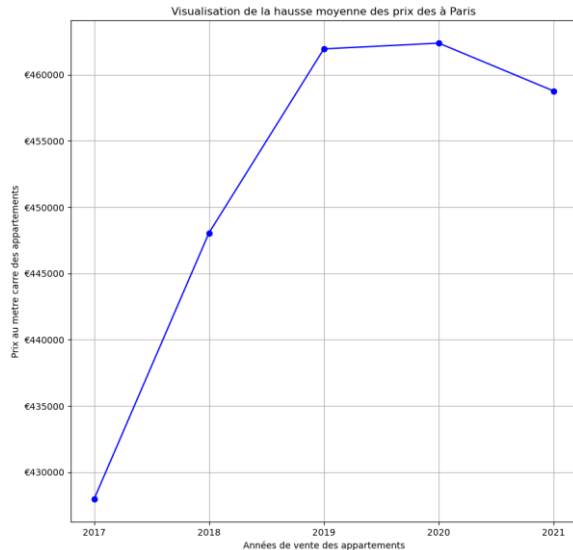
Valeur foncière moyenne (en €) des appartements selon les années



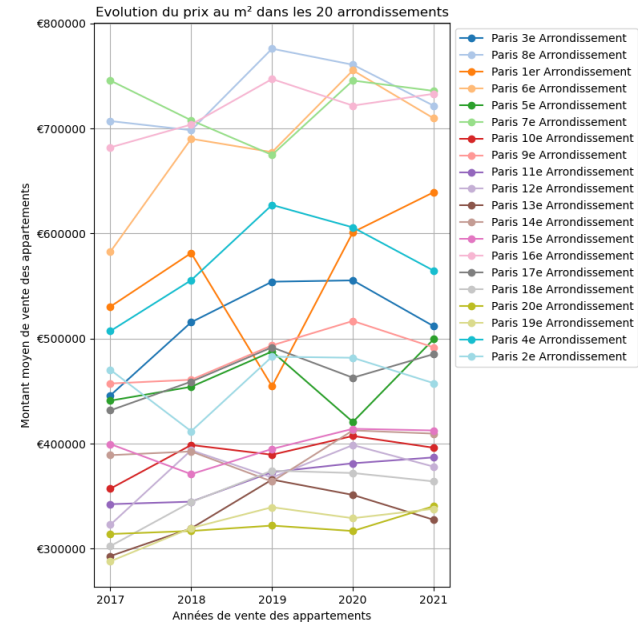
I. Analyse du marché de l'immobilier

Appartement

On constate que le moyenne des prix de ventes des appartements à Paris à tendance à augmenter depuis 2017.



On observe globalement que les prix augmentent quelque soit l'arrondissement. Le 7e arrondissement à la valeur foncière la plus élevée. Le 6e arrondissement semble être l'arrondissement qui connaît la plus grande hausse.

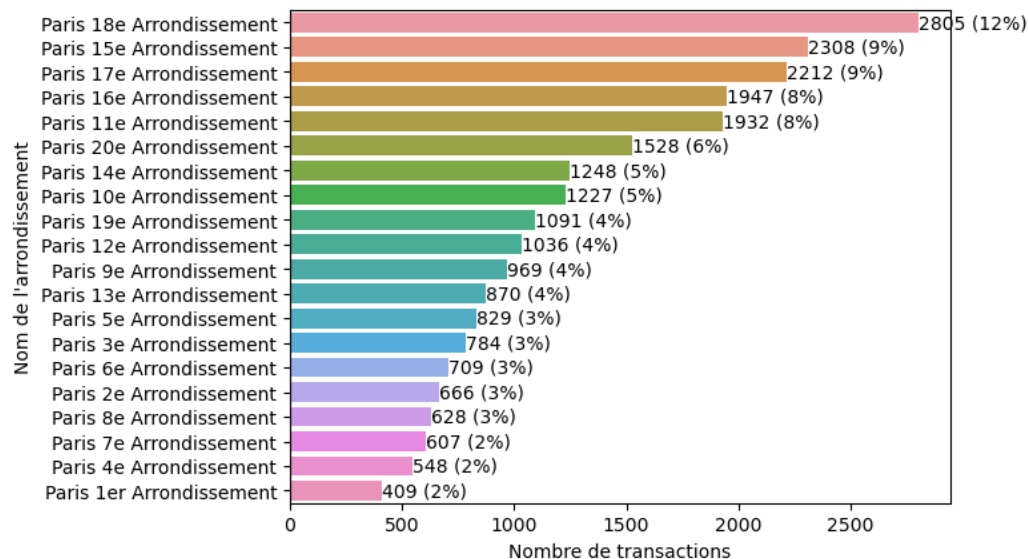


I. Analyse du marché de l'immobilier

Appartement

Il semble que le 6e arrondissement n'est pas l'arrondissement où il y a le plus de transaction étant donné la hausse des prix depuis plusieurs années.

C'est le 18e arrondissement où il y en a le plus avec 2805 transactions (dans cet arrondissement le prix au m² est plus faible mais à tendance à augmenter également depuis 2017).



I. Analyse du marché de l'immobilier

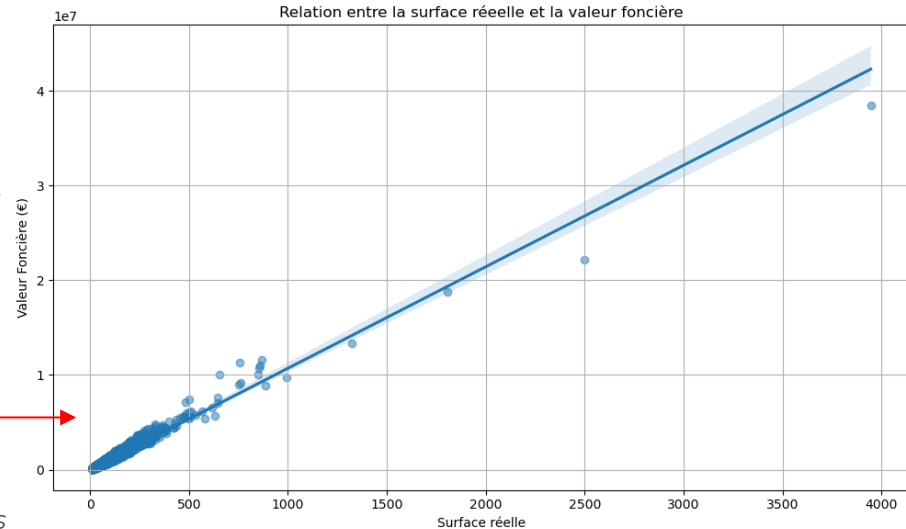
Appartement

Effectivement la moyenne des prix au mètre carré a tendance à augmenter depuis 2017.

Si on étudie la relation entre la valeur foncière et la date de transaction on s'aperçoit que le coefficient de corrélation est de 0,08 avec une pvalue de 0,0038 donc *il semble que la relation entre l'année et la valeur foncière soit une relation légèrement croissante (+0.08) et significative (car p-value < 0.05)*

La relation entre la valeur foncière et la surface réelle de l'appartement est confirmée (coeff = 0.98, pvalue = 0.0) car *il semble que la relation entre la surface réelle et la valeur foncière soit une relation croissante et significative.*

A noter que la relation entre le type de local et le prix au m² n'est pas vérifiée (coeff = 0.25 ; pvalue = 0.0).



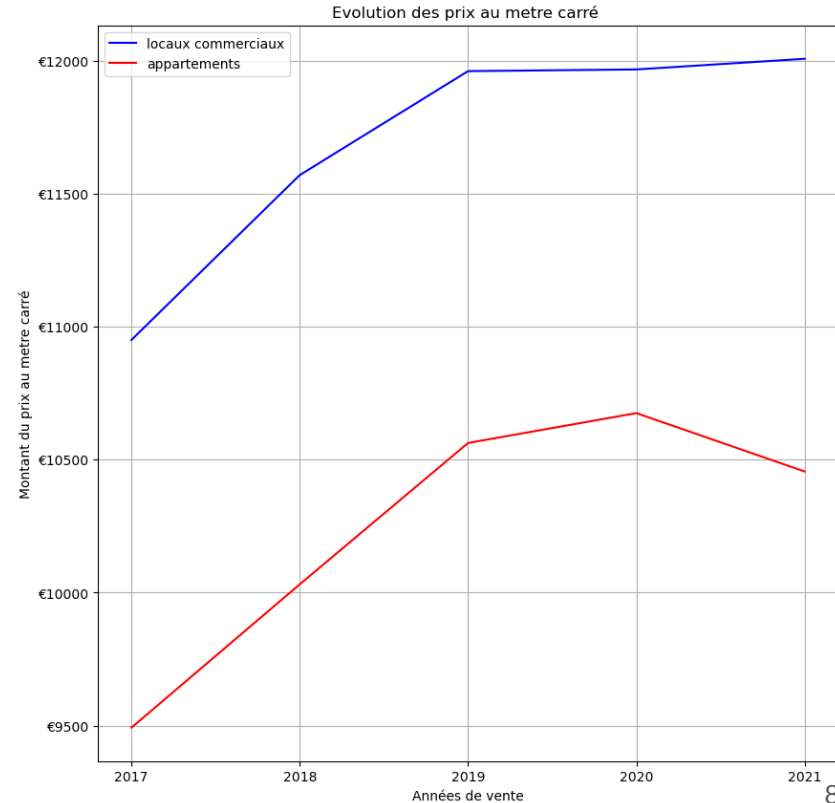
I. Analyse du marché de l'immobilier

Comparaison des prix au m² entre les appartements et les locaux commerciaux

Le prix au mètre carré des locaux commerciaux est croissant entre 2017 et 2021. Cette croissance atteint son pic en 2021 où le prix au m² est de plus de 12 006,49€.

Après ces analyses, nous pouvons conclure que les dimensions à utiliser pour prédire le prix au m² sont :

- ❖ *la surface du bien immobilier,*
- ❖ *la date considérée,*
- ❖ *la localisation (code_postal),*
- ❖ *le type de bien.*



II. Méthodologie suivie

- 1) Objectif de la prédiction : prédire la valeur foncière d'un bien immobilier
- 2) Création et préparation des données pour la prédiction (mettre le code postal en colonne pour avoir des variables binaires = `get_dummies()`, séparé le jeu de donnée en jeu de test et d'entraînement)
- 3) Entraîner le modèle de prédiction
- 4) Analyser les scores obtenus :
 - ❖ le score R^2 est bon car ils sont proches de 1 donc le modèle prédit plutôt bien les valeurs foncières).

```
train score LR = 0.9848714873845213  
test score LR = 0.9673985917135132
```
 - ❖ En moyenne le modèle se trompe d'environ 55 353€ sur les données de test dans ses prédictions (mean absolut error = ~ 55 353€, MAPE = 14,6%) et 31 847€ sur les données d'entraînement (MAPE = 8,7%).

	code_postal	code_type_local	surface_reelle	year	month	day	prediction_prix	prix_m2
0	75001	2	25	2022	12	31	466632.417866	18665.30
1	75003	2	22	2022	12	31	409507.092863	18613.96
2	75003	2	65	2022	12	31	853172.700700	13125.73
3	75008	2	74	2022	12	31	902589.822951	12197.16
4	75008	2	22	2022	12	31	366063.971613	16639.27

III. Résultat des prédictions

Expliquez ici:

- Le résultat de la prédiction sur la valorisation au 31 décembre 2022 du portefeuille des actifs immobiliers de l'entreprise.
 - ❖ le score R^2 est bon car ils sont proches de 1 donc le modèle prédit plutôt bien les valeurs foncières. Il prédit mieux sur les données d'entraînement (car R^2 plus proche de 1).

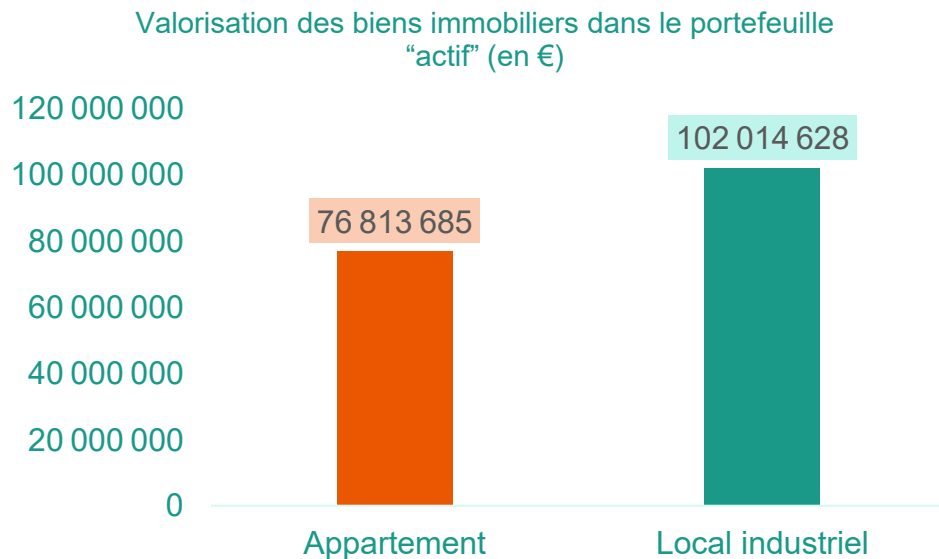
```
train score LR = 0.9848714873845213
test score LR = 0.9673985917135132
```

	adresse_numero	adresse_nom_voie	code_postal	code_commune	nom_commune	surface_carrez	code_type_local	type_local	surface_reelle_bati	nombre_pieces_principales	longitude	latitude	prediction.prix
0	127	RUE SAINT-DENIS	75001	75101	Paris 1er Arrondissement	16.41	2	Appartement	25	1	2.349593	48.863227	466632.417866
1	62	RUE NOTRE-DAME DE NAZARETH	75003	75103	Paris 3e Arrondissement	20.47	2	Appartement	22	1	2.356977	48.867907	409507.092863
2	62	RUE DE TURENNE	75003	75103	Paris 3e Arrondissement	64.48	2	Appartement	65	3	2.364855	48.859318	853172.700700
3	32	AV GEORGE V	75008	75108	Paris 8e Arrondissement	63.29	2	Appartement	74	2	2.301586	48.868893	902589.822951
4	152	BD HAUSSMANN	75008	75108	Paris 8e Arrondissement	20.58	2	Appartement	22	1	2.311643	48.875490	366063.971613
5	152	BD HAUSSMANN	75008	75108	Paris 8e Arrondissement	33.40	2	Appartement	35	1	2.311643	48.875490	500196.434448
6	52	RUE DES GRAVILLIERS	75003	75103	Paris 3e Arrondissement	21.35	2	Appartement	24	2	2.355637	48.864366	430142.702530
7	208	RUE SAINT MAUR	75010	75110	Paris 10e Arrondissement	54.69	2	Appartement	60	3	2.371885	48.873209	672172.268629
8	142	RUE DE COURCELLES	75017	75117	Paris 17e Arrondissement	9.30	2	Appartement	11	1	2.300373	48.882505	190387.469708
9	52	RUE LAMARCK	75018	75118	Paris 18e Arrondissement	12.88	2	Appartement	13	1	2.342134	48.889431	190082.154581

- Les limites et précautions à prendre avec les résultats.
 - ❖ En moyenne le modèle se trompe d'environ 33 233€ sur les données de test dans ses prédictions (mean absolut error = ~33 233€) et 34 384€ sur les données d'entraînement
 - ❖ On pourrait aller plus loin en essayant d'autre modèle, ou encore tester le modèle avec une autre source de donnée.

III. Résultat des prédictions

Projection des gains potentiels :

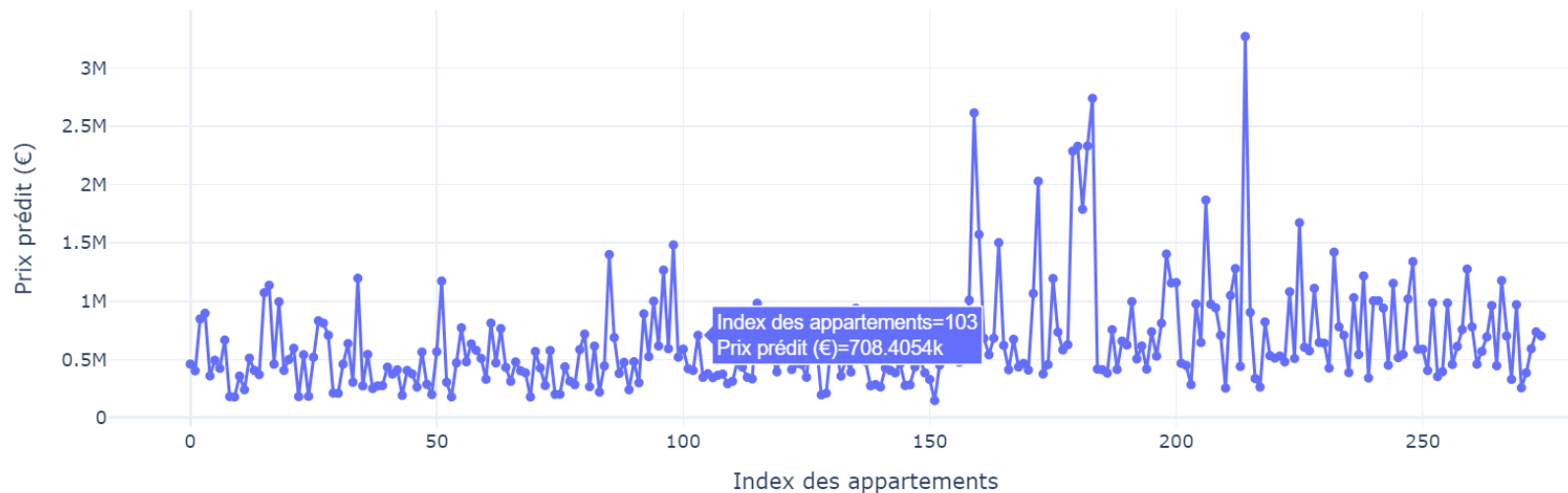


III. Résultat des prédictions

Prédictions des prix selon l'index des appartements :



Prédiction des prix selon l'index des appartements dans les données



Les Plus Beaux Logis de Paris

Partie 2



Les Plus Beaux Logis de Paris

2^e prédiction (Kmeans)

Simon

Doussin

2025

I. Méthodologie suivie

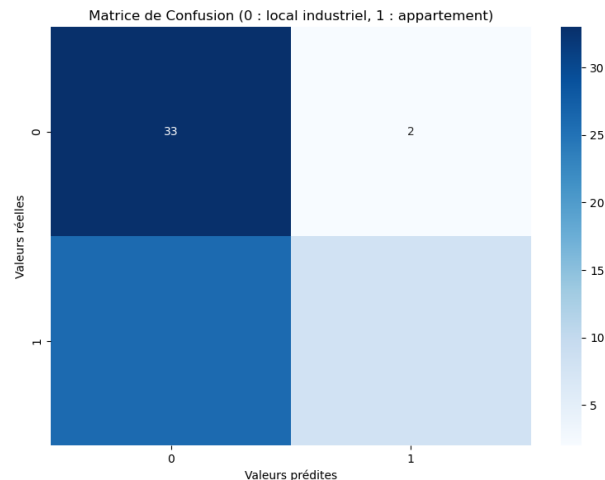
- 1) *Préparation du jeu de donnée pour l'algorithme du Kmeans :*
 - ❖ *Création d'une colonne dans le df_actifs pour prédire le prix des appartements*
 - ❖ *Création d'une colonne pour calculer le prix au m² ([prediction_prix] / [surface_réelle])*
 - ❖ *On garde uniquement les colonnes utiles : [code_postal], [code_type_local], [prix_m2]*
- 2) *Nous observons dans les données que nous avons des valeurs différentes de prix au mètre carré pour un même arrondissement (ici le 19ème arrondissement). Il semble que cela soit notre dimension à utiliser pour attribuer les prix au mètre carré les plus élevés aux locaux commerciaux, et les prix les plus bas aux appartements.*
 - ❖ *Pour vérifier ça, nous avons utilisé l'algorithme du Kmeans qui va créer 2 groupes (appartement vs locaux commerciaux) de manière autonome*

II. Résultat de la classification

Les 2 groupes ont été créés et voici les résultats :

- *1^{er} groupe (local industriel) :*
 - ✓ *Code postal moyen : 75013*
 - ✓ *Prix au m² moyen : 12 561€*
- *2^e groupe (appartement) :*
 - ✓ *Code postal moyen : 75007*
 - ✓ *Prix au m² moyen : 19 619€*
- *L'inertie est grande donc les clusters (groupes) sont bien séparé les uns des autres.*
- *L'Adjusted Rand Index (ARI) est de 0,048 donc les clusters sont peu corrélés avec les catégories de bien.*

II. Résultat de la classification



ero	adresse_nom_voie	code_postal	code_commune	nom_commune	surface_carrez	code_type_local	type_local	surface_reelle_bati	nombre_pieces_principales	longitude	latitude	prediction_prix	prix_m2	cluster	cluster_label
127	RUE SAINT-DENIS	75001	75101	Paris 1er Arrondissement	16.41	2	Appartement	25	1	2.349593	48.863227	466632.417866	18665.30	1	Appartement
62	RUE NOTRE-DAME DE NAZARETH	75003	75103	Paris 3e Arrondissement	20.47	2	Appartement	22	1	2.356977	48.867907	409507.092863	18613.96	1	Appartement
62	RUE DE TURENNE	75003	75103	Paris 3e Arrondissement	64.48	2	Appartement	65	3	2.364855	48.859318	853172.700700	13125.73	0	Local industriel, commercial ou assimilé
32	AV GEORGE V	75008	75108	Paris 8e Arrondissement	63.29	2	Appartement	74	2	2.301586	48.868893	902589.822951	12197.16	0	Local industriel, commercial ou assimilé
152	BD HAUSSMANN	75008	75108	Paris 8e Arrondissement	20.58	2	Appartement	22	1	2.311643	48.875490	366063.971613	16639.27	1	Appartement
...

II. Résultat de la classification

Les limites et précautions à prendre avec les résultats.

- *Les 2 groupes ont bien été formés.*
- *Le centre des clusters montre une différence dans les prix au m² : appartement = environ 19 619€/m², local industriel = environ 12 561€/m²*
- *Il y a peu de variables explicatives (uniquement code postal et prix au m²)*
- *On utilise le code postal comme variable numérique alors que c'est en réalité une variable catégorielle*

