



Les Plus Beaux Logis de Paris

Alexandre Duvernay

19 août 2024

PROJET 8

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



Objectif de la mission



L'entreprise doit **vendre des actifs** pour se garantir un niveau suffisant de trésorerie.

Elle souhaite que soit utilisées des techniques de data science pour prédire les valorisations d'actifs en vue de statuer sur le segment le plus porteur pour les prochaines années.

Elle pourra ainsi vendre des actifs **sur le segment le moins porteur** afin de se couvrir des risques sur le marché de l'immobilier.



Introduction

--

Sur la base des données historiques des prix immobiliers de Paris entre 2017 et 2021

✓ **analyser l'évolution des prix immobiliers à Paris**

Puis, dans la liste des biens immobilier de l'entreprise

✓ **Identifier le segment le plus porteur (appartements ou locaux commerciaux)**

DONNÉES TRANSMISES



- historique_immobilier_paris_2017_2021.xlsx
- portefeuille_actifs.xlsx

ENVIRONNEMENT DE TRAVAIL



python™



1 – Analyse du marché de l'immobilier

Type de données et type de biens

```
[ ] df_historique.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26196 entries, 0 to 26195
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   date_mutation          26196 non-null  datetime64[ns]
1   valeur_fonciere        26196 non-null  float64
2   adresse_numero         26196 non-null  int64
3   adresse_nom_voie       26196 non-null  object
4   code_postal            26196 non-null  int64
5   nom_commune            26196 non-null  object
6   code_type_local        26196 non-null  int64
7   type_local             26196 non-null  object
8   surface_reelle         26196 non-null  int64
dtypes: datetime64[ns](1), float64(1), int64(4), object(3)
memory usage: 1.8+ MB
```



Dans les données historiques :
Nombre de types de biens : 2

	type_local	code_type_local
Local industriel, commercial ou assimilé	Appartement	2
		4



1 – Analyse du marché de l'immobilier

Nettoyage des données

16
lignes
dupliquées



Duplicate rows

Most frequently occurring

	date_mutation	valeur_fonciere	adresse_numero	adresse_nom_voie	code_postal	nom_commune	code_type_local	type_local
0	2017-01-04	885116.628219	8	RUE PAUL HERVIEU	75015	Paris 15e Arrondissement	2	Appartement
1	2017-05-16	108677.313288	7	RUE DOCTEUR GERMAIN SEE	75016	Paris 16e Arrondissement	2	Appartement
2	2017-05-18	246703.319425	400	RUE SAINT HONORE	75001	Paris 1er Arrondissement	4	Local industriel. c
3	2017-10-17	855144.863014	71	RUE D AMSTERDAM	75008	Paris 8e Arrondissement	2	Appartement
4	2017-12-20	132855.435616	18	RUE SIMART	75018	Paris 18e Arrondissement	2	Appartement
5	2017-12-28	148485.376712	6	RUE BLAINVILLE	75005	Paris 5e Arrondissement	2	Appartement
6	2019-03-14	109796.610000	14	RUE DE LANCRY	75010	Paris 10e Arrondissement	2	Appartement
7	2019-06-27	146977.652055	26	BD DE CLICHY	75018	Paris 18e Arrondissement	2	Appartement
8	2019-07-12	225856.479452	17	BD SAINT MARTIN	75003	Paris 3e Arrondissement	2	Appartement
9	2019-09-25	931045.625342	15	RUE RAYNOUARD	75016	Paris 16e Arrondissement	2	Appartement

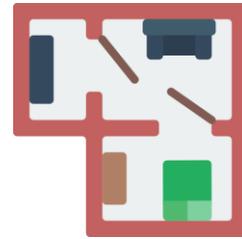


1 – Analyse du marché de l'immobilier

Nombre de transactions et intervalle temporel



26'180
transactions



24'338
appartements



1'842
locaux
commerciaux

02/01/2017

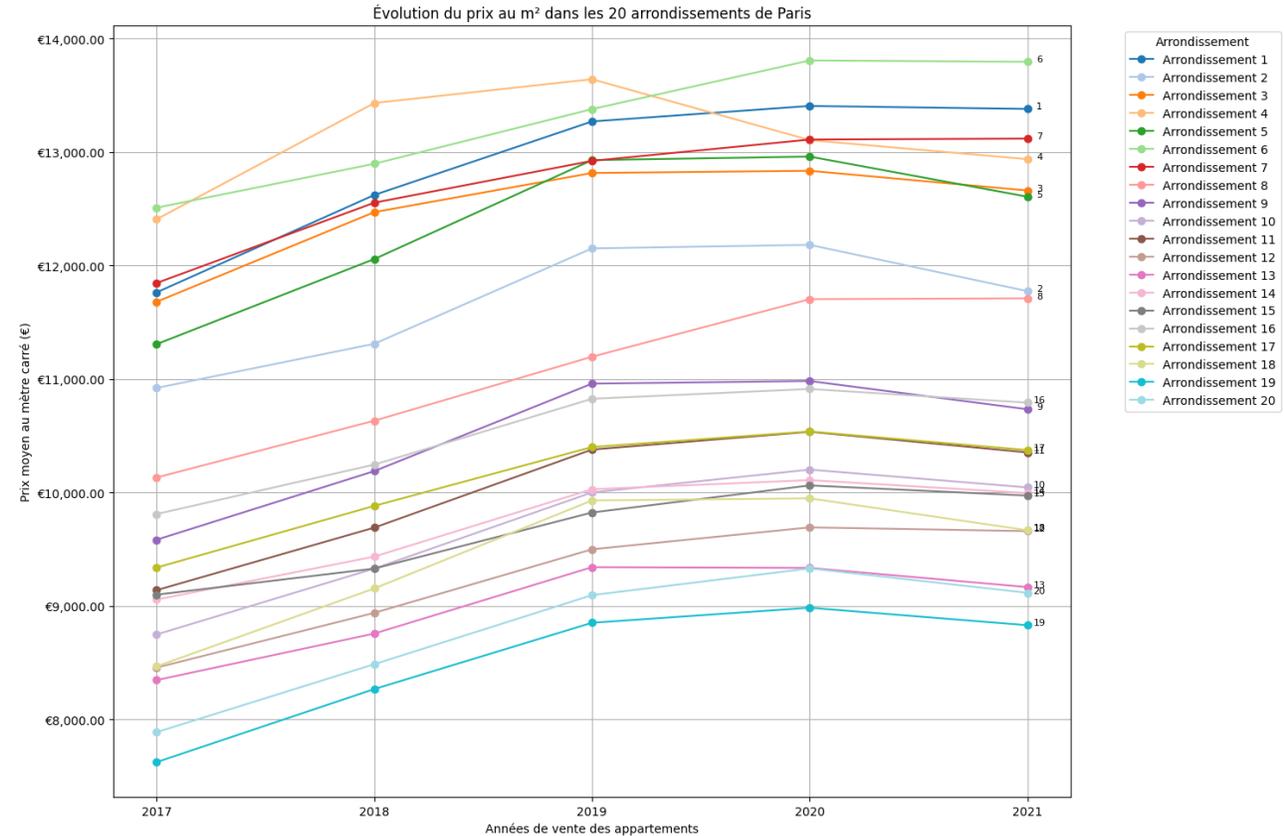
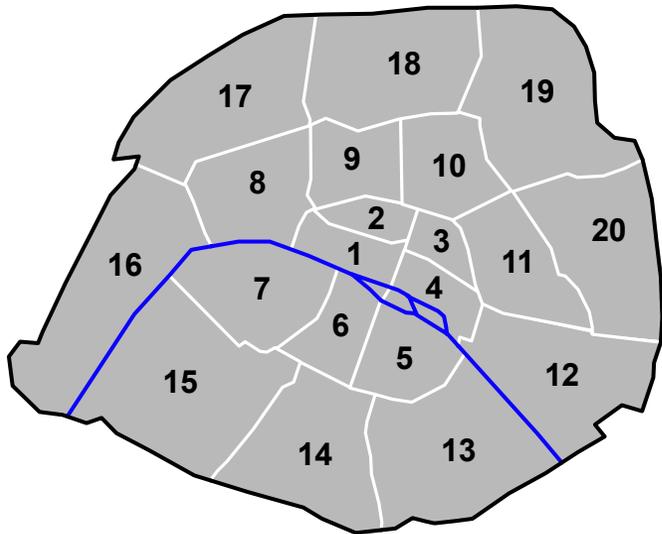


31/12/2021



1 – Analyse du marché de l'immobilier

Evolution du prix au m2 dans Paris

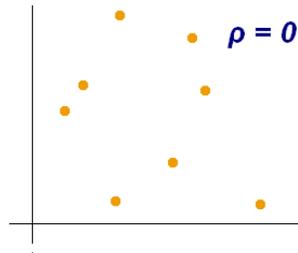
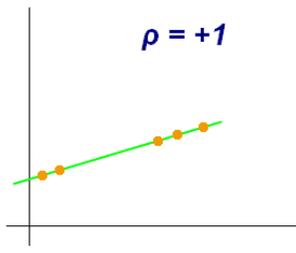
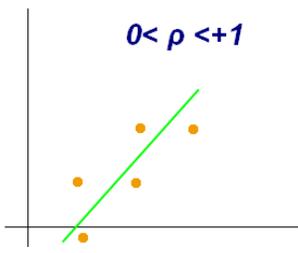
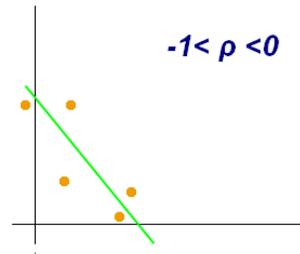
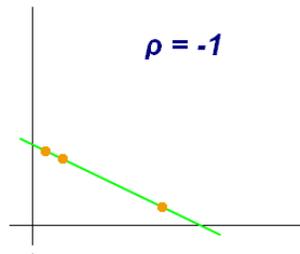




1 – Analyse du marché de l'immobilier

Corrélation entre prix m2 et date dans le 6^e arrondissement

Corrélation de Pearson



0,9

Forte corrélation positive
(qd une variable augmente l'autre aussi)

Le coefficient de corrélation de Pearson est : 0.9038

La valeur p associée est : 7.1084e-263

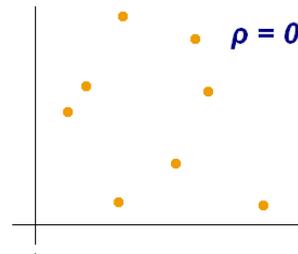
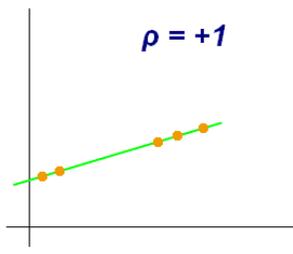
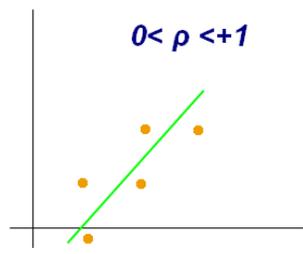
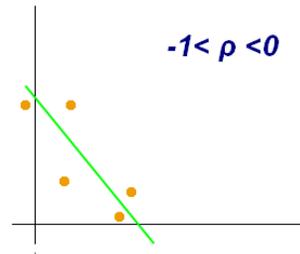
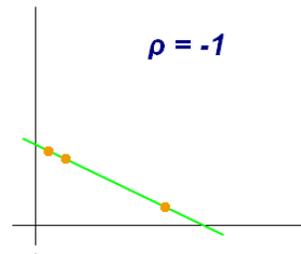
La corrélation est statistiquement significative ($p < 0.05$).



1 – Analyse du marché de l'immobilier

Corrélation entre prix au m2 et surface dans le 6^e arrondissement

Corrélation de Pearson



1

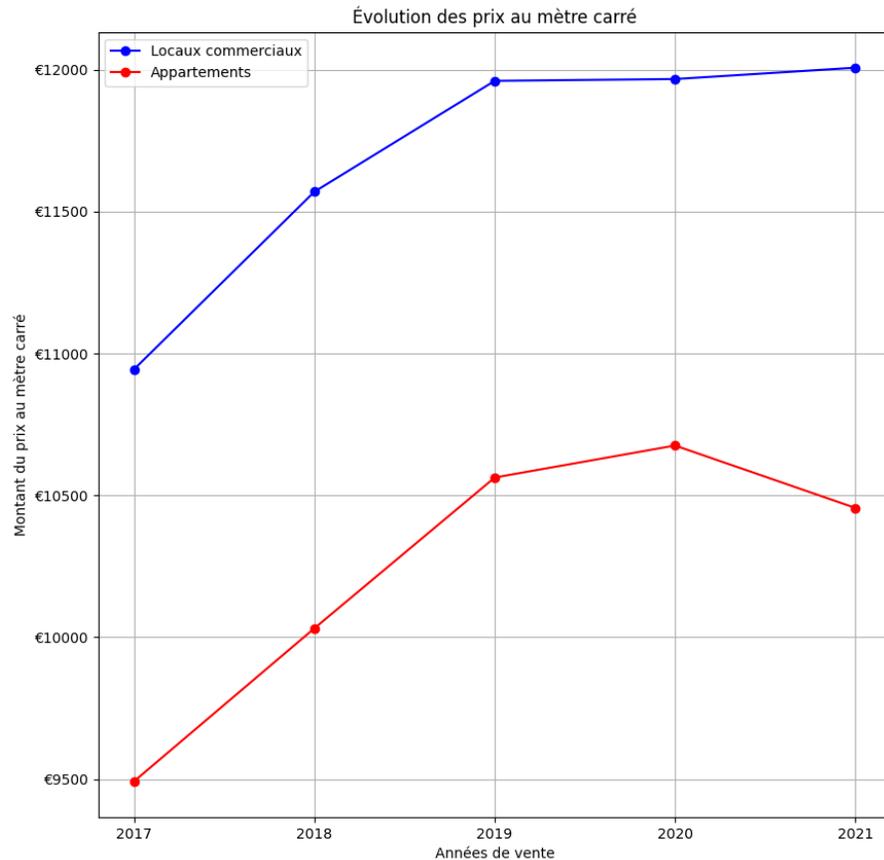
Corrélation positive parfaite
(qd une variable augmente l'autre aussi)

Le coefficient de corrélation de Pearson entre la valeur foncière et la surface est : 0.9978
La valeur p associée est : 0.0000e+00
La corrélation est statistiquement significative ($p < 0.05$).



1 – Analyse du marché de l'immobilier

Différence de prix au m2 entre locaux commerciaux et appartements



Locaux commerciaux

annee	prix_metre_carre
2017	10944.06
2018	11569.50
2019	11960.13
2020	11966.47
2021	12006.49

Appartements

annee	prix_moyen_metre_carre
2017	9492.44
2018	10031.40
2019	10562.46
2020	10675.34
2021	10455.60



1500€
d'écart moyen env.

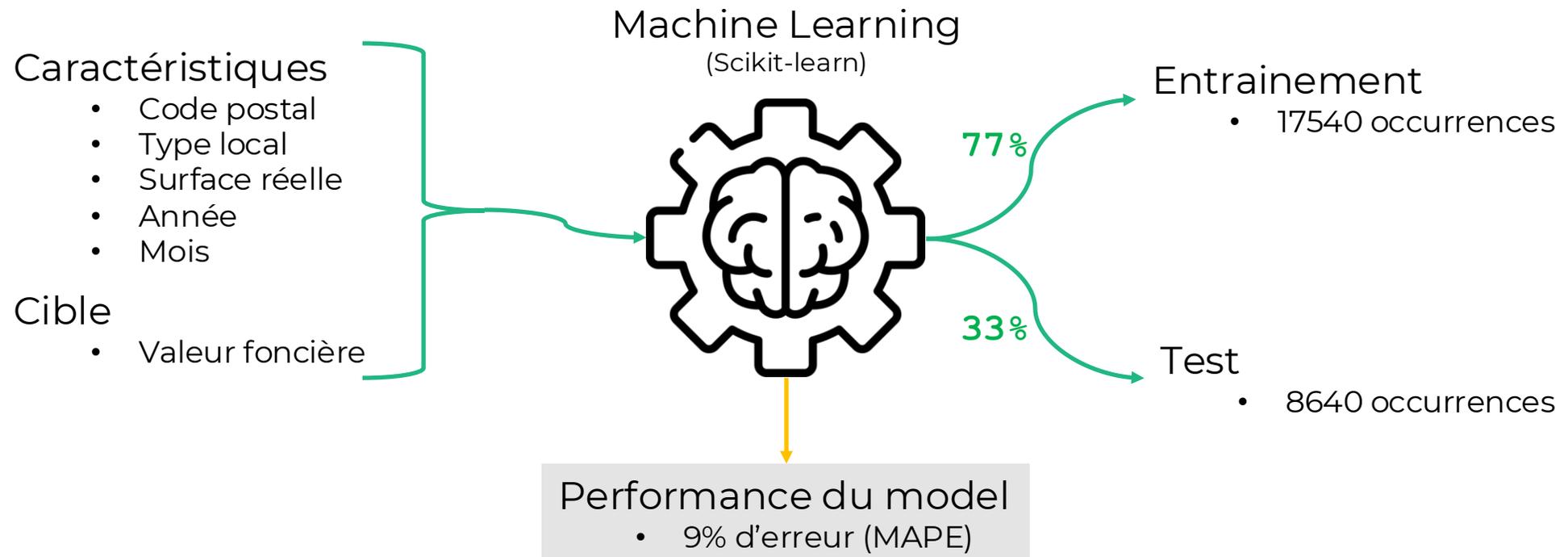


2 – ML / Traitement algorithmique

Régression linéaire > entraînement de l'algorithme



historique_immobilier_paris.xlsx





2 – ML / Traitement algorithmique

Régression linéaire > préparation des données client

2
lignes
dupliquées



Duplicate rows

Most frequently occurring

	adresse_numero	adresse_nom_voie	code_postal	code_commune	nom_commune	surface_carrez	code_type_local	type_local
0	3	RUE DAMREMONT	75018	75118	Paris 18e Arrondissement	26.86	4	Local industriel. commercial
1	104	BD VOLTAIRE	75011	75111	Paris 11e Arrondissement	38.50	4	Local industriel. commercial

Report generated by YData.



```
# Renommer la colonne 'surface_reelle_bati' en 'surface_reelle' pour correspondre au modèle  
df_portefeuille = df_portefeuille.copy()  
df_portefeuille.rename(columns={'surface_reelle_bati': 'surface_reelle'}, inplace=True)
```

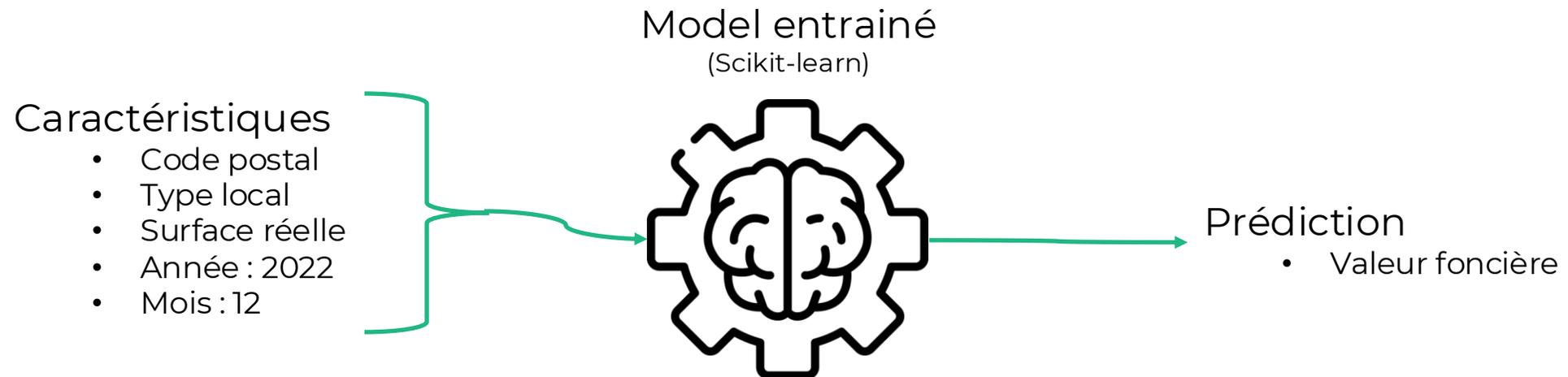


2 – ML / Traitement algorithmique

Régression linéaire > prédiction pour le client



Portefeuille_actifs.xlsx





2 – ML / Traitement algorithmique

Régression linéaire > résultat de la prédiction au 31/12/2022

Locaux commerciaux

97 M €



Appartements

71 M €



Cette estimation est fondée sur des données historiques. Si les conditions économiques changent, les prédictions peuvent ne pas refléter correctement les nouvelles réalités. De même, si les données utilisées pour l'entraînement contiennent des erreurs, cela peut affecter la précision des estimations.



3 – Etiquetage automatique des données

Préparation des données

 echantillon_a_classer.xlsx



```
# Calcul du prix au mètre carré  
df_echantillon['prix_m2'] = df_echantillon['valeur_fonciere'] / df_echantillon['surface_reelle']
```



```
# Suppression des colonnes inutiles  
df_echantillon = df_echantillon.drop(columns=['valeur_fonciere', 'surface_reelle'])
```

Appartement ou Local industriel ?



3 – Etiquetage automatique des données

Observation



```
df_echantillon.head()
```

	code_postal	nom_commune	prix_m2
0	75019	Paris 19e Arrondissement	9871.444128
1	75019	Paris 19e Arrondissement	10045.572493
2	75019	Paris 19e Arrondissement	9194.697790
3	75019	Paris 19e Arrondissement	9469.142168
4	75019	Paris 19e Arrondissement	7463.610005

Nous observons dans les données que nous avons des **valeurs différentes de prix au mètre carré pour un même arrondissement** (ici le 19ème arrondissement).

C'est sur cette variable (prix_m2) que nous allons travailler pour **scinder les données en 2 groupes**. Car nous le savons, l'analyse précédente à révéler que le prix au m2 des appartements est plus faible que celui des locaux commerciaux.

Appartement ou Local industriel ?



3 – Etiquetage automatique des données

Méthode de classification – partitionnement (ou clustering)

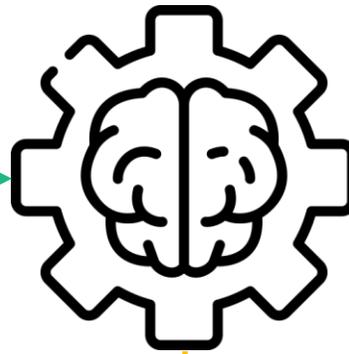


echantillon_a_classer.xlsx

Appartement ou Local industriel ?

Machine Learning
(K-means)

Prix au M2



2 groupes

label	count	mean	std	min	25%	50%
0	20.0	7408.775030	141.329995	7207.21763	7304.332833	7374.091721
1	20.0	9806.924674	232.955000	9194.69779	9705.214437	9842.717300
		75%	max			
0		7512.112911	7666.071700			
1		9979.592847	10113.195822			

```
# Préparation des données pour KMeans (on n'utilise que la colonne 'prix_m2')
X = df_echantillon[['prix_m2']].values
```

```
# Appliquer l'algorithme KMeans pour classifier en 2 clusters
kmeans = KMeans(n_clusters=2, random_state=0, n_init=10).fit(X)
```



3 – Etiquetage automatique des données

Résultat de la classification

	code_postal	nom_commune	prix_m2	label	type_bien
0	75019	Paris 19e Arrondissement	9871.444128	1	Local industriel/commercial
1	75019	Paris 19e Arrondissement	10045.572493	1	Local industriel/commercial
2	75019	Paris 19e Arrondissement	9194.697790	1	Local industriel/commercial
3	75019	Paris 19e Arrondissement	9469.142168	1	Local industriel/commercial
4	75019	Paris 19e Arrondissement	7463.610005	0	Appartement
5	75019	Paris 19e Arrondissement	9982.016919	1	Local industriel/commercial
6	75019	Paris 19e Arrondissement	9674.262777	1	Local industriel/commercial
7	75019	Paris 19e Arrondissement	7348.917415	0	Appartement
8	75019	Paris 19e Arrondissement	9789.051438	1	Local industriel/commercial
9	75019	Paris 19e Arrondissement	7616.420959	0	Appartement
10	75019	Paris 19e Arrondissement	7490.469041	0	Appartement
11	75019	Paris 19e Arrondissement	9518.076103	1	Local industriel/commercial
12	75019	Paris 19e Arrondissement	7238.422973	0	Appartement
13	75019	Paris 19e Arrondissement	9869.462342	1	Local industriel/commercial
14	75019	Paris 19e Arrondissement	7343.539315	0	Appartement
15	75019	Paris 19e Arrondissement	7313.114384	0	Appartement
16	75019	Paris 19e Arrondissement	7415.564351	0	Appartement
17	75019	Paris 19e Arrondissement	9949.544452	1	Local industriel/commercial
18	75019	Paris 19e Arrondissement	9978.784822	1	Local industriel/commercial
19	75019	Paris 19e Arrondissement	7296.900510	0	Appartement

20	75019	Paris 19e Arrondissement	7322.570548	0	Appartement
21	75019	Paris 19e Arrondissement	7207.217630	0	Appartement
22	75019	Paris 19e Arrondissement	9815.972258	1	Local industriel/commercial
23	75019	Paris 19e Arrondissement	7437.880186	0	Appartement
24	75019	Paris 19e Arrondissement	7399.266027	0	Appartement
25	75019	Paris 19e Arrondissement	7306.810274	0	Appartement
26	75019	Paris 19e Arrondissement	9961.146860	1	Local industriel/commercial
27	75019	Paris 19e Arrondissement	10103.161577	1	Local industriel/commercial
28	75019	Paris 19e Arrondissement	9597.631488	1	Local industriel/commercial
29	75019	Paris 19e Arrondissement	7613.514621	0	Appartement
30	75019	Paris 19e Arrondissement	9734.720495	1	Local industriel/commercial
31	75019	Paris 19e Arrondissement	7666.071700	0	Appartement
32	75019	Paris 19e Arrondissement	10113.195822	1	Local industriel/commercial
33	75019	Paris 19e Arrondissement	7577.044521	0	Appartement
34	75019	Paris 19e Arrondissement	7294.202055	0	Appartement
35	75019	Paris 19e Arrondissement	7239.983240	0	Appartement
36	75019	Paris 19e Arrondissement	9715.531658	1	Local industriel/commercial
37	75019	Paris 19e Arrondissement	7583.980837	0	Appartement
38	75019	Paris 19e Arrondissement	9983.591800	1	Local industriel/commercial
39	75019	Paris 19e Arrondissement	9771.486096	1	Local industriel/commercial



Limites de l'analyse

Mise en garde

Le modèle est basé sur des **données historiques**. Si le marché ou les conditions économiques changent, les prédictions peuvent ne pas refléter correctement les nouvelles réalités.

Certaines données importantes et pouvant impacter significativement le prix des biens **ne sont pas présentes** et/ou n'ont pas été utilisées dans le modèle de prédiction (nombre de pièces du logement, diagnostic énergétique, étage, état du bien...) pouvant conduire à des estimations inexactes.

Les prédictions reposent sur des données passées, et **les conditions futures peuvent différer** significativement, surtout dans un marché immobilier où les tendances peuvent rapidement évoluer. Une évolution des taux d'emprunt par exemple.

Certains biens peuvent avoir des **caractéristiques uniques** non prises en compte par le modèle, ce qui pourrait conduire à des estimations erronées.

La valeur réelle peut être influencée par des **conditions spécifiques** au marché local, comme la demande actuelle ou des projets de développement urbain à proximité, qui ne sont pas pris en compte par la régression linéaire.



END

Merci pour votre écoute 😊