



Les Plus Beaux Logis de Paris

Analyse de l'évolution des prix de l'immobilier pour Les Plus Beaux Logis de Paris

Nathalie Currid

Novembre 2024

Les Plus Beaux Logis de Paris

I. Contexte du projet



Les Plus Beaux Logis de Paris

I. Contexte du projet

Les plus Beaux Logis de Paris, société de gestion immobilière gérée par Maurice Delattre et ses deux enfants Maxence et Louise, responsables respectivement des segments 'Particuliers' (154 appartements) et 'Corporate' (121 locaux industriels, commerciaux ou assimilés) souhaite vendre une partie de ses actifs pour générer de la trésorerie, et optimiser son activité en identifiant rapidement le type de biens présents sur le marché.

L'objet de ce projet est :

- de prédire l'évolution des prix de l'immobilier parisien,
- d'estimer la valeur du portefeuille d'actifs de la société et de faire de recommandations,,
- de proposer un outil de classification des données.



Sommaire

- I. Contexte du projet
- II. Analyse des données historiques du marché de l'immobilier parisien (2017-2021)
- III. Entraînement d'un modèle de régression linéaire
- IV. Prédiction de la valeur des segments 'particulier' et 'corporate' au 31 décembre 2022 (valorisation des actifs)
- V. Classification des données issues du jeu de test
- VI. Méthodologie suivie et conclusion

Les Plus Beaux Logis de Paris

II. Analyse du marché de l'immobilier



Les Plus Beaux Logis de Paris

II. Analyse du marché de l'immobilier

— Comparaison de l'évolution des prix au m2 par segment

Prix au m2 plus élevé pour les **locaux commerciaux** avec une croissance ralentie depuis 2019.

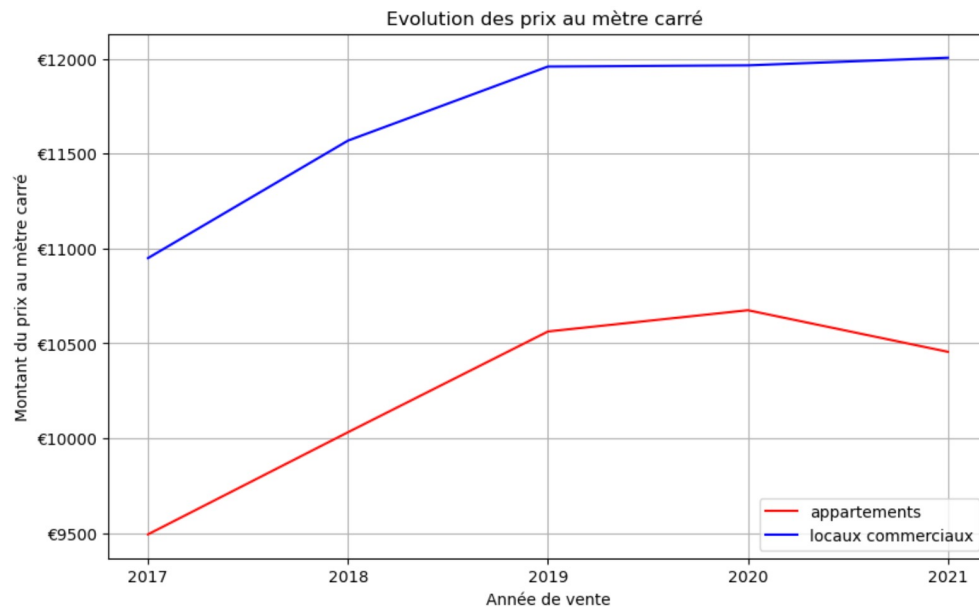
Baisse du prix au m2 pour les **appartements** depuis 2020.

Basé sur 26196 transactions dont :

- 24353 appartements
- 1843 locaux commerciaux

->10 fois plus de transactions immobilières pour les appartements.

Source : *historique_immobilier_paris_2017_2021*





Les Plus Beaux Logis de Paris

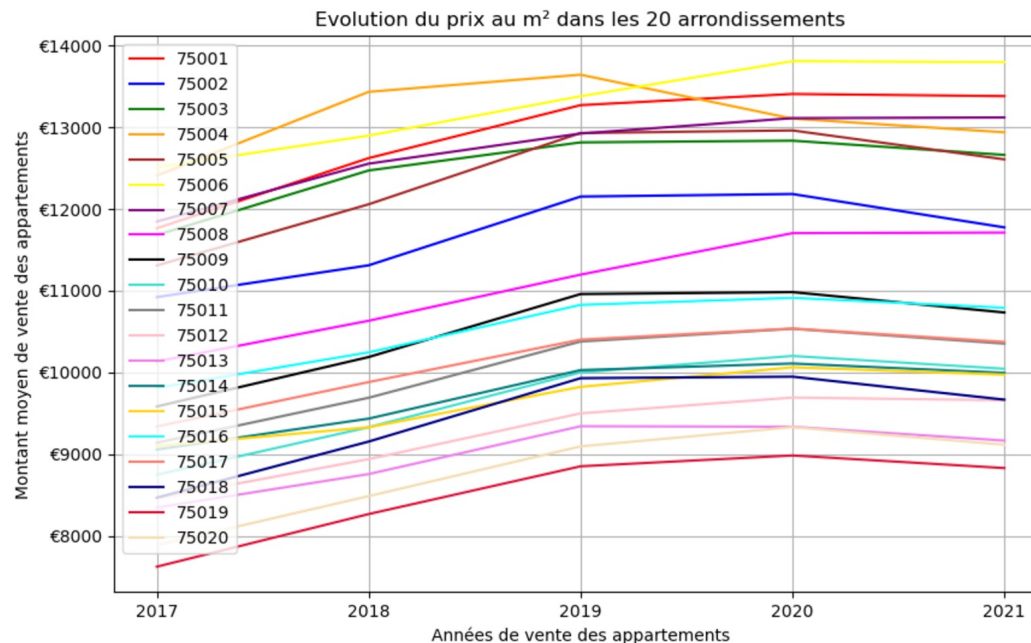
II. Analyse du marché de l'immobilier

Évolution prix m² par arrondissement pour les appartements

La tendance est à la baisse mais cela varie selon les arrondissements.

- Faire une étude par arrondissement
- Vérifier dans quels arrondissements sont les actifs de l'entreprise.
- Effectuer les mêmes analyses pour les locaux commerciaux.

Source : *historique_immobilier_paris_2017_2021*





Les Plus Beaux Logis de Paris

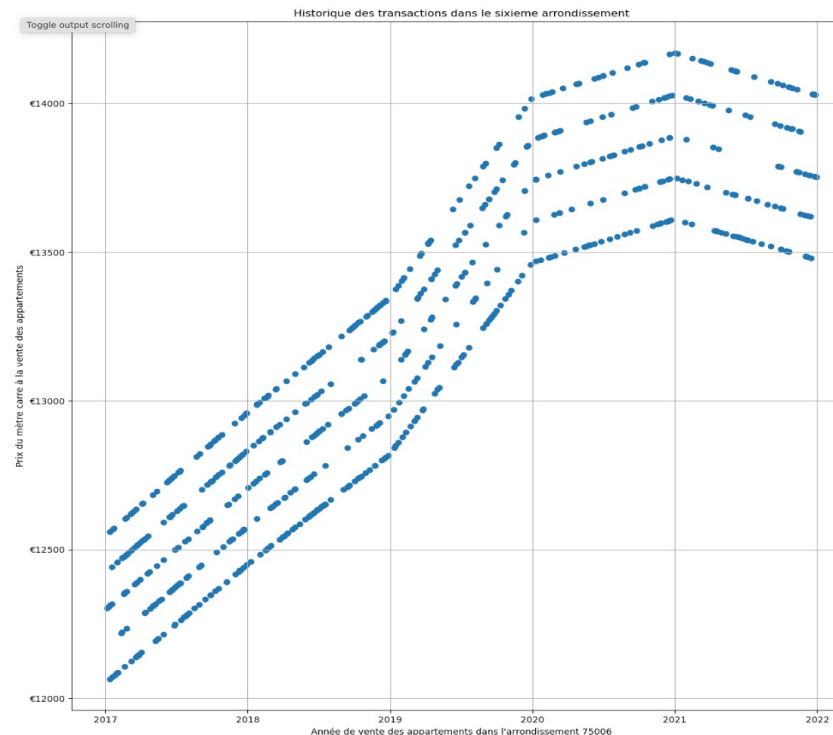
II. Analyse du marché de l'immobilier

Prix des appartements au m2 par transaction dans le 6^{ème} arrondissement de Paris

Visualisation de la dispersion des données sur 709 transactions (3% des transactions totales)

- Le nombre de transactions a diminué au fur et à mesure que le prix au m2 a augmenté.
- L'écart s'est resserré en 2019 puis creusé à partir de 2020 entre les valeurs maximum et minimum du prix au m2
- Le coefficient de corrélation de Pearson indique une corrélation très forte et positive entre les variables :
 - date et prix au m2
 - valeur foncière et surface

Source : *historique_immobilier_paris_2017_2021*





II. Analyse du marché de l'immobilier

Relation valeur foncière et surface tous segments confondus

Coefficient de corrélation de Pearson entre la valeur foncière et la surface des deux segments (appartements et locaux industriels, commerciaux ou assimilés) :

PearsonRResult(statistic=
0.98, pvalue=**0.0**)

Interprétation des valeurs de r :

La force de la corrélation peut être évaluée ainsi :

$0 \leq |r| < 0.3$: corrélation faible ou négligeable

$0.3 \leq |r| < 0.5$: corrélation modérée

$0.5 \leq |r| < 0.7$: corrélation forte

$0.7 \leq |r| < 1.0$: corrélation très forte

$|r| = 1$: Corrélation parfaite : relation linéaire positive

$|r| = -1$: Corrélation parfaite : relation linéaire négative

$r = 0.98$ indique une corrélation très forte et positive entre les deux variables valeur foncière et surface

Interprétation des valeurs de la p-value :

< 0.05 : petite p-value : résultats peu susceptibles d'être dus au hasard

≥ 0.05 : grande p-value : les données ne sont pas suffisamment probantes

0.0 indique que la probabilité que les résultats observés soient dus au hasard est extrêmement faible

Les Plus Beaux Logis de Paris

III. Entraînement d'un modèle de régression linéaire

III. Entraînement d'un modèle de régression linéaire pour prédire le prix au m2 d'un bien immobilier



Les Plus Beaux Logis de Paris

Justification du choix :

Algorithme de régression linéaire approprié car :

- existence d'une relation linéaire entre les variables 'valeur foncière' et 'surface'
- données nettoyées

Étapes

- Préparation des données : transformation des colonnes 'code_postal' et 'type_local' grâce au *one hot encoder* de la bibliothèque *scikit-learn*, importée via *sklearn* (solution prête à l'emploi). Chaque catégorie devient une colonne : un **1** indique la présence de la catégorie et un **0** son absence. Création d'une colonne 'timestamp' à partir des données de 'date_mutation'.
- Utilisation du `train_test_split` pour prélever 1/3 des données.
- Entraînement de l'algorithme sur les 2/3 avec `model.fit(X_train, y_train)`.
- Prédictions effectuées avec `model.predict(X_test)`.
- Comparaison entre les valeurs réelles et les prédictions.

III. Entraînement d'un modèle de régression linéaire pour prédire le prix au m2 d'un bien immobilier



Les Plus Beaux Logis de Paris

Résultat

Notre algorithme a fait une erreur moyenne de **8.93%** sur la prédiction de la valeur foncière, ce qui est acceptable du fait que les prédictions concernent un domaine où les variations de prix sont élevées et difficiles à prévoir (l'immobilier).

Limites

Les données utilisées couvrent une période durant laquelle le prix au m2 a beaucoup évolué. Le prix au m2 diffère selon le type de bien et l'arrondissement où il se trouve, ce qui n'est pas reflété dans cette prédiction.

Recommandations

Entraîner l'algorithme sur les données :

- par type de local
- par arrondissement de Paris
- sur les années 2019 à 2021

Les Plus Beaux Logis de Paris

IV. Prédiction de la valeur des segments 'particulier' et 'corporate' au 31 décembre 2022

IV. Prédiction de la valeur des segments 'particulier' et 'corporate' au 31 décembre 2022



Les Plus Beaux Logis de Paris

Étapes

- Préparation des données du portefeuille d'actifs au format des données utilisées pour l'entraînement de l'algorithme.
- Prédictions effectuées

Résultat

- La valorisation du **segment particulier** (154 appartements) est : **71.24 M€**
- La valorisation du **segment corporate** (121 locaux) est : **98.31 M€**

Recommandations

D'après cette prédiction de la valorisation des actifs de l'entreprise, le segment du portefeuille qui aura la plus grande valorisation est le segment corporate. Cependant, les deux portefeuilles n'ont pas le même nombre de biens par arrondissement et il se peut que les biens du segment corporate soient situés dans des arrondissements où le prix au m2 est moins élevé. Il serait donc judicieux d'entraîner l'algorithme par type de bien pour avoir une estimation plus juste.

Les Plus Beaux Logis de Paris

V. Classification des données issues du jeu de test

V. Classification des données issues du jeu de test



Les Plus Beaux Logis de Paris

Labellisation automatique des biens immobiliers comme étant :

- soit 'Appartement'
- soit 'Local industriel. commercial ou assimilé'

- Préparation des données de l'échantillon à classer (ajout prix/m2)
- Utilisation de *k-means* pour regrouper les données en groupes de caractéristiques similaires (clustering)
- Prédiction effectuée
- Remplacement des valeurs des labels (0 = Appartement, 1 = local commercial)

Conclusions sur les limites de l'exercice

Cette analyse permet d'effectuer une présélection pour trier les types de bien mais elle n'est pas fiable à 100% car, par exemple, le prix du m2 diffère d'un arrondissement à l'autre, il se pourrait donc qu'un appartement dans un arrondissement où le coût au m2 est élevé se trouve classifié comme 'Local industriel. commercial ou assimilé'.

Les Plus Beaux Logis de Paris

VI. Méthodologie suivie et conclusion

VI. Méthodologie suivie et conclusion



Les Plus Beaux Logis de Paris

J'ai utilisé les techniques de machine learning suivantes pour faire des prédictions à partir des données préparées :

- Apprentissage supervisé : **régression linéaire** pour prédire une variable cible continue en fonction de variables indépendantes ;
- Apprentissage non supervisé : **clustering** pour partitionner les données en groupes basés sur leurs caractéristiques (k-means)

Conclusion et recommandations

Bien que cette étude estime une plus grande valorisation du portefeuille 'corporate', nous vous recommandons d'effectuer les analyses complémentaires suggérées lors de cette présentation avant de prendre des décisions stratégiques. Une fois celles-ci effectuées, il faudrait également considérer le point suivant : avec la volatilité du marché, est-il pertinent de se limiter à un portefeuille d'actifs ou plus judicieux de diversifier son portefeuille de biens, afin de garder une certaine agilité ? (appartements plus faciles à la revente)



Merci