

PARTIE 1

Analyse de l'évolution des prix du marché immobilier
et prédiction sur la valorisation du portefeuille au
31 Décembre 2022

I. Analyse du marché de l'immobilier

A. Choix du segment :

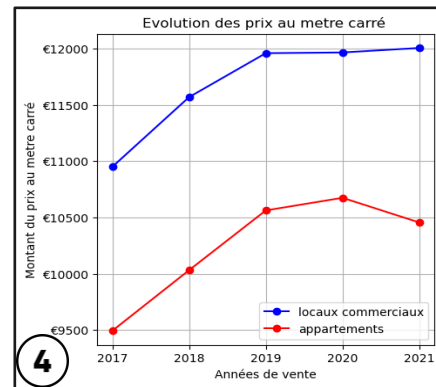
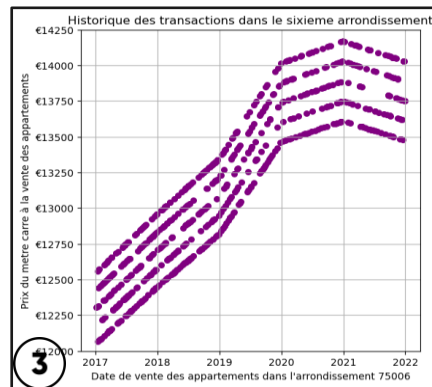
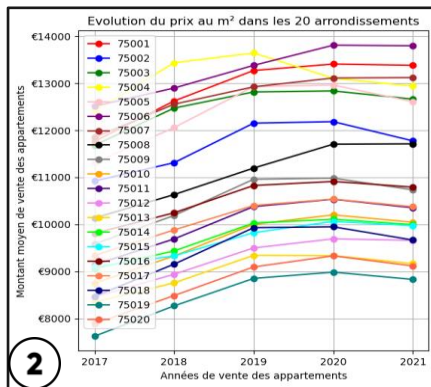
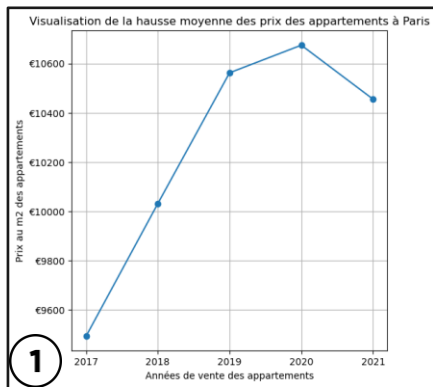
Le segment particulier (appartement) représente 92,96% du jeu de données (dataset) contenant l'historique des transactions immobilières, nous avons donc choisi de commencer à analyser celui-ci, puis de le comparer au segment corporate (locaux).

B. Résultats d'analyse : ① ② ③

- Prix moyen du m2 en augmentation quasi linéaire de 2017 à 2019
- Ralentissement généralisé de l'augmentation entre 2019 et 2020
- Légère diminution entre 2020 et 2021 (pour revenir sensiblement au prix moyen de 2019)

C. Comparaison des segments : ④

- Prix moyen du m2 plus élevé sur le segment corporate
- Ralentissement plus net entre 2019 et 2020 (que sur le segment particulier)
- Augmentation ininterrompue de 2017 à 2021

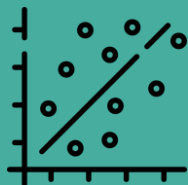


II. Méthodologie suivie



Nettoyage dataset « historique immo »

- Vérification des données contenues (types, nombre, intervalle).
- Réalisation de calculs et création de nouvelles variables (prix au m2).
- Suppression des données non pertinentes pour l'analyse.



Analyse et tests de corrélation

- Analyse segment particuliers (92,96%).
- Tests de corrélation (Pearson) pour évaluer relation entre variables.
- Résultats «statistiquement significatifs» ($p\text{-value} < 5\%$) et fortes corrélations positives (coeff corrélation $> 90\%$).



Comparaison des segments

- Prix moyen du m2 du segment corporate est plus élevé sur l'ensemble de la période.
- Segment corporate a augmenté sans interruption sur l'intervalle 2017/2021 (contrairement au segment particuliers)



Préparation données pour entraînement

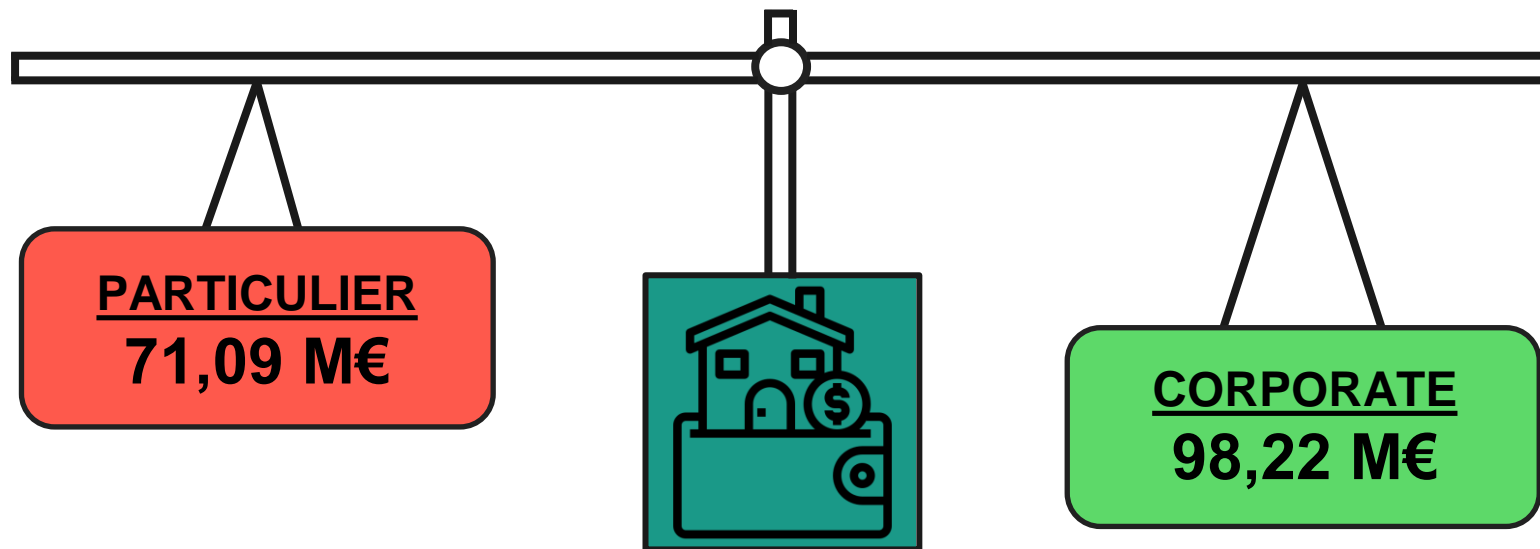
- Encodage préalable et nécessaire pour transformer variables catégoriques en variables numériques (= one hot encoding).
- Objectif : les rendre adaptées à former / entraîner notre algorithme de machine learning (apprentissage automatique)



Entraînement d'un algorithme «supervisé»

- « train_test_split » = séparation aléatoire du dataset pour entraîner et tester l'algorithme (train = 67% / test = 33%)
- « LinearRegression » = permet d'obtenir une prédiction + pourcentage d'erreur moyenne de cette prédiction (résultat obtenu = 8,78% d'erreur moyenne)

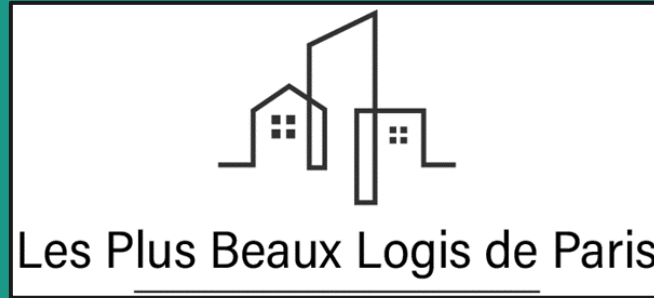
III. Résultats des prédictions



Valorisation du portefeuille d'actifs au 31.12.2022

Limites et précautions :

- Le pourcentage d'erreur moyenne sur la prédiction reste important (8,78%)
- La valorisation est réalisée à un instant T (= le portefeuille transmis n'a pas d'historique nous permettant de visualiser l'évolution du prix au m2 des segments dans le passé)
- Pas de visibilité sur l'évolution à venir (événements pouvant modifier la prédiction : guerres, krachs, épidémies....)



PARTIE 2

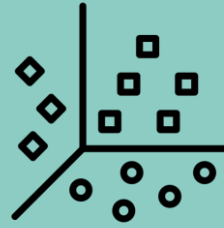
Classification automatique des nouvelles opportunités
d'acquisition de biens immobiliers grâce à un
algorithme de prédiction

I. Méthodologie suivie



Nettoyage dataset « échantillon à classer »

- Vérification des données contenues (types, nombre).
- Réalisation de calculs et création de nouvelles variables (prix au m²).
- Suppression des données non pertinentes pour l'analyse.



Utilisation d'un algorithme « non-supervisé »

- Utilisation d'un algorithme de clustering « K-means » permettant de partitionner / regrouper les données.
- 2 clusters ont été définis pour répartir l'échantillon autour de 2 centroïdes (données prototypes) et ainsi déterminer si le bien est un local ou un appartement.



Visualisation et classification

- Visualisation des valeurs attribuées aux labels (0 ou 1) et aux centroïdes (= moyenne de chaque clusters).
- Classification des données selon l'attribution faites par l'algorithme et intégration des résultats au dataset « échantillon à classer ».

II. Résultat de la classification

Résultats :

- La classification a permis d'obtenir 2 clusters (= groupes) de tailles équivalentes (20 biens par clusters).
- Le centroïd (= moyenne) du cluster corporate indique un prix moyen du m2 à 9806,92 €
- Le centroïd du cluster particulier indique un prix moyen du m2 à 7408,78 €

Limites et précautions :

- L'algorithme ne permet pas de déterminer le nombre optimal de clusters, il faut le définir en amont en fonction des besoins.
- Les résultats obtenus peuvent être différents en fonction du nombre de clusters et d'itérations définis (il s'agit d'un algorithme « non-déterministe » qui trouvera une approximation de solution).
- Les données catégoriques doivent préalablement être encodées pour être traitées car il n'est pas adapté aux données non numériques.

