



Les Plus Beaux Logis de Paris



ESN DATA

# **ANALYSE DE L'EVOLUTION DES PRIX IMMOBILIERS AVEC PYTHON**

**Réalisé par Loïc Stéphane BAMENOU**  
**Sous supervision Jade (Dan SLAMA)**



# Plan

## 1<sup>ère</sup> Partie :

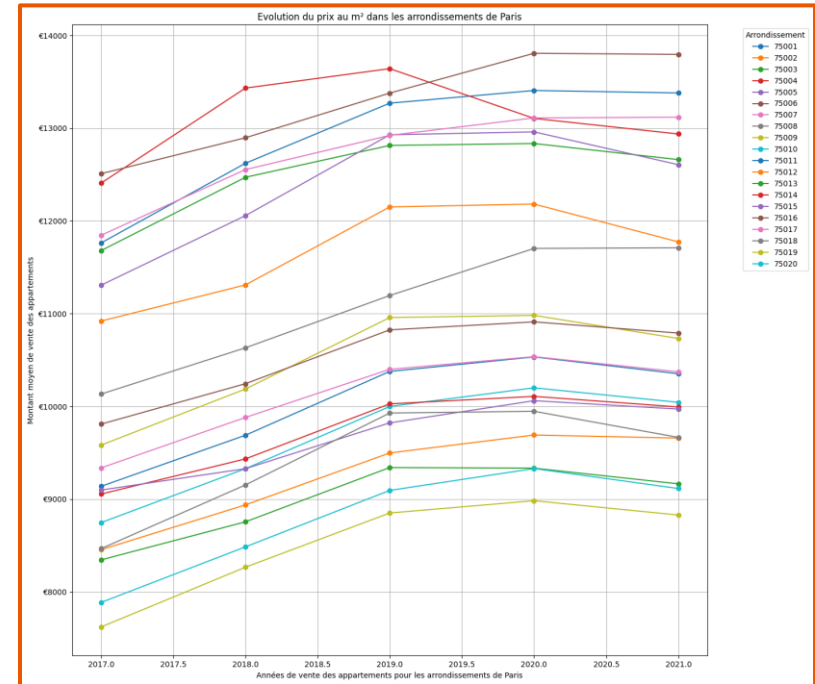
- I. Analyse du marché de l'immobilier
- II. Méthodologie suivie de l'évolution du prix (régression)
- III. Résultats des prédictions

## 2<sup>ème</sup> Partie :

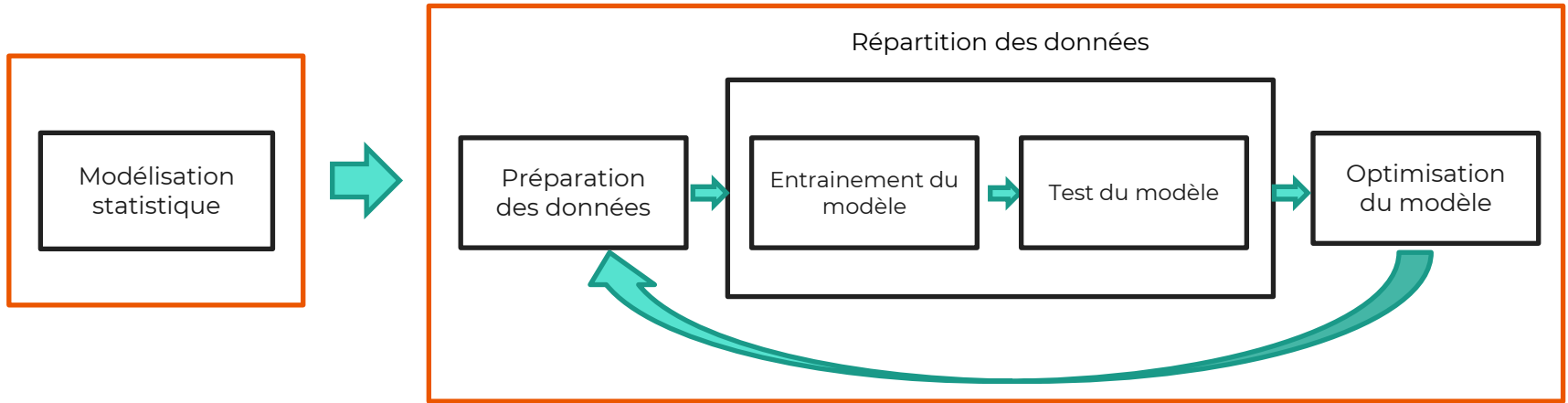
- IV. Méthodologie de classification des biens actuels
- V. Résultats de la classification

# I. Analyse du marché de l'immobilier

- ❖ *Entre 2017 et 2021, notre entreprise enregistre un background de 26 196 transactions immobiliers (soit 93 % concerne des appartements);*
- ❖ *On note une hausse de 193 €/an sur les biens immobiliers de type appartements avec une baisse de l'espace habitable ;*
- ❖ *Une baisse historique de 46 % des transactions accompagner par la hausse 257 €/an pour les appartements du sixième arrondissement ;*
- ❖ *Enregistrement d'une corrélation positive entre le prix  $m^2$  d'un bien et l'année sur tout type de bien (appartement comme locaux);*



## II. Méthodologie suivie (Régression)

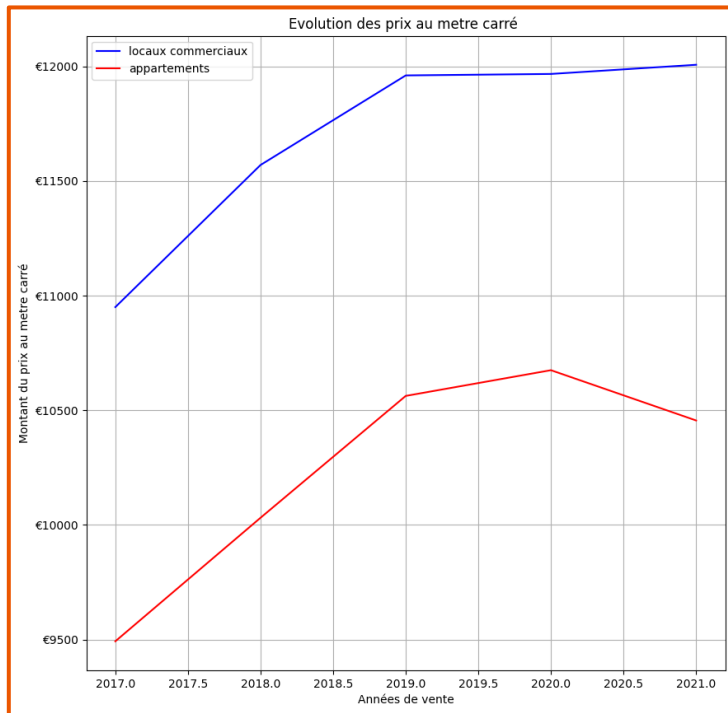


*Une modélisation statistique à travers l'analyse de l'évolution des prix des biens immobiliers par le biais du prix au mètre carré*

*Une approche supervisée pour la prédiction du secteur d'activité bénéfique :*

- ❖ 1<sup>ère</sup> étape : Data cleaning (nettoyage des données : manquantes, outliers ou autres ; numérisation des variables catégorielles) ;
- ❖ 2<sup>ème</sup> étape : Répartition des données (entraînement et test du modèle) ;
- ❖ 3<sup>ème</sup> étape : Optimisation du modèle ;

# III. Résultat des prédictions



Confrontation des résultats de la modélisation statistique à celle prédictive :

- ❖ Hypothèse : Les locaux commerciaux sont plus rentables que les appartements (**validé**)

Développement du modèle de prédiction :

- ❖ La marge d'erreur de 10 %
- ❖ Le  $R^2 = 1$ , la variable cible « valeur foncière est parfaitement expliquée par les autres variables prédictrices ;
- ❖ En présence de notre portefeuille d'actifs, le modèle issu de la régression linéaire estime que les locaux commerciaux (91 933 865 €) sont plus lucratifs que les appartements (63 159 515 €) ;
- ❖ Limites de notre prédiction :
  - ✓ Perte d'informations avec plusieurs modalités postales

# IV. Méthodologie suivie (Classification)



*Construction des Clusters (sous-ensembles) par le biais de l'algorithme de K-Means (MCO en français) sur le prix du  $M^2$*

*Phase d'initialisation :*

- ❖ Choisir le nombre de sous ensemble (Cluster  $K = 2$ );
- ❖ Déterminer le centre des clusters (centroïdes);

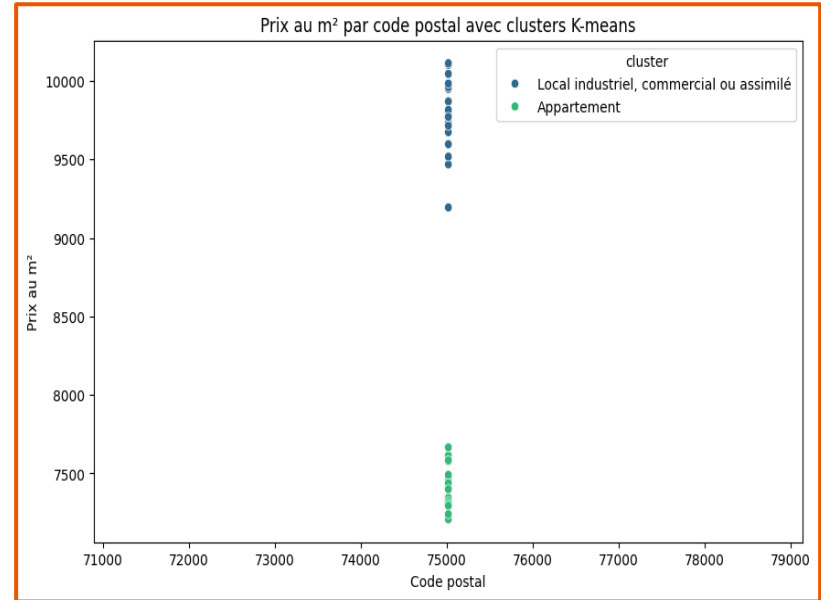
*Phase d'initialisation (suite) :*

- ❖ Attribution des données
- ❖ Mise à jour des centroïdes après disposition des données autour des centroïdes

# V. Résultat de la classification

Résultats du clustering :

- ❖ La méthode du coefficient la silhouette notre choix de faire deux groupes ( $s = 0,91 \sim 1$ );
- ❖ La valeur centrale des appartements est de 7 408 € contre 9 806 € pour les locaux commerciaux ;
- ❖ L'intervalle entre les deux biens immobiliers s'élève à 1 529 €
  - ❖ Appartement : [7 207 € ; 7 666 €]
  - ❖ Locaux commerciaux : [9 195 € ; 10 113 €]
- ❖ La performance du modèle est assez bonne à 95,8 % ;



# Recommandations

- ❖ Maintien de l'offre des biens de type appartement ;
- ❖ Diversifier le catalogue de bien de type commercial ;
- ❖ Identifier les zones commerciales à fort potentiel ;
- ❖ Adapter les valeurs locatives et les rendements en fonction des conditions du marché ;
- ❖ Mettre en place un outil de suivi d'activité (storytelling et de modélisation pour prédire les tendances du marché ;