



DataScientest

# Compagnon Immobilier

*Aide à la décision pour l'achat immobilier basée sur la data*

Rapport de Projet - AVR 25 CDS



- Christine Jaloux
- Walid Chaabouni
- Abdesamad Anzaghi
- Clément Visseaux

# COMPAGNON IMMOBILIER

## Fiche Projet



### CONTEXTE

- Le marché immobilier est complexe hétérogène et fortement dépendant du territoire
- Les acheteurs disposent de beaucoup d'informations mais peu d'outils pour les interpréter objectivement

#### Problématique

Comment **aider un acheteur à estimer** un bien et à **comparer** des territoires de manière **fiable** et objective ?

### OBJECTIFS

- Analyser les dynamiques territoriales immobilières,
- Estimer le prix au mètre carré d'un bien,
- Comprendre les facteurs qui influencent les prix,
- Proposer un outil simple et accessible à l'utilisateur.

L'objectif n'est pas de remplacer un expert immobilier, mais de fournir un outil **d'aide à la décision** basé sur la data.

### PERIMETRE

Pour cela, nous avons utilisé plusieurs types de données en France métropolitaine (96 fichiers CSV (~1 Go)) :

- Des données issues d'annonces immobilières,
- Des données géographiques comme les IRIS, départements et régions,
- Ainsi que des données socio-économiques.

Les données couvrent plusieurs années, incluant la période du COVID, ce qui a un impact important sur les prix et leur volatilité.

# MÉTHODOLOGIE



# COMPAGNON IMMOBILIER

## Méthodologie Projet



### Analyse exploratoire & Préparation des données

Une étape clé du projet a été la préparation des données. Nous avons travaillé sur le nettoyage des valeurs manquantes, le traitement des valeurs aberrantes et la cohérence globale des variables.

Nous avons également séparé les biens en deux catégories : maisons et appartements, car leurs logiques de prix sont différentes.

Cette étape était essentielle pour garantir des modèles fiables.

### Modélisation du prix au m<sup>2</sup>

Le problème est formulé comme un problème de **régression**. Plusieurs modèles ont été testés, notamment des modèles basés sur les arbres.

Le modèle retenu est XGBoost, car il offre le meilleur compromis entre performance, robustesse et capacité de généralisation.

Les performances obtenues sont cohérentes compte tenu de la complexité du marché immobilier.

### Interprétation du modèle

Formulé comme un problème de régression.

Plusieurs modèles ont été testés, notamment des modèles basés sur les arbres.

Le modèle retenu est XGBoost, car il offre le meilleur compromis entre performance, robustesse et capacité de généralisation.

Les performances obtenues sont cohérentes compte tenu de la complexité du marché immobilier.



# Streamlit

**Bascule sur Streamlit pour présenter:**

- L'Analyse territoriale (ACP),
- La Prédiction temporelle,
- Mise en perspective des résultats.

