

# Analyse du marché immobilier à partir de données web scraping avec Python

---

## Objectifs

- **Collecter automatiquement** des données immobilières sur plusieurs sites (ex : Leboncoin, SeLoger, Logic-Immo).
  - **Nettoyer et structurer** les données (prix, surface, localisation, type de bien, nombre de pièces, etc.).
  - **Analyser** les tendances : prix moyens, variation selon les régions, rapport prix/m<sup>2</sup>, évolution dans le temps.
  - **Visualiser** les résultats à l'aide d'un **tableau de bord interactif** (Streamlit ou Dash).
- 

## Problématique

Comment le prix au mètre carré varie-t-il en fonction de la localisation, de la surface et du type de bien immobilier en France ?

---

## Sources de données possibles

Tu peux scraper (ou utiliser une API si disponible) :

- [Leboncoin Immobilier](#)
- [SeLoger](#)
- [Logic-Immo](#)
- [Bien'ici](#)

💡 **Astuce** : pour éviter les blocages, commence par scraper **une seule ville** (par ex. Paris, Lyon, Toulouse).

---

## Stack technique

Étape	Objectif	Librairies / outils
Scraping	Récupérer les données des annonces	requests, BeautifulSoup, re
Nettoyage	Nettoyer et structurer les données	pandas, numpy

Étape	Objectif	Librairies / outils
Stockage	Sauvegarder les données	CSV, SQLite, <code>pandas.DataFrame.to_csv()</code>
Analyse	Calculs statistiques, regroupements	pandas, matplotlib, seaborn
Géolocalisation	Transformer adresse → latitude/longitude	geopy, API <i>OpenStreetMap</i> ( <i>Nominatim</i> )
Visualisation	Cartes interactives	folium, plotly, streamlit
Dashboard	Interface web interactive	Streamlit ou Dash

## Structure du projet (exemple d'arborescence)

PROJET\_IMMOBILIER/

├── DATA/

│ ├── ANNONCES\_RAW.CSV

│ └── ANNONCES\_CLEAN.CSV

├── SRC/

│ ├── SCRAPER.PY # SCRIPT POUR LE SCRAPING

│ ├── CLEAN\_DATA.PY # NETTOYAGE ET FORMATAGE

│ ├── ANALYSE.PY # CALCULS STATISTIQUES ET GRAPHIQUES

│ └── DASHBOARD.PY # APPLICATION STREAMLIT

├── NOTEBOOKS/

│ └── EXPLORATION.IPYNB # ANALYSES ET TESTS

└── RAPPORT/

└── RAPPORT\_PROJET.PDF # TON RAPPORT FINAL DE M1

## Pipeline du projet

### 1. Scraping des données

- Utiliser `requests` pour récupérer le HTML des pages.
- Parser les éléments utiles avec `BeautifulSoup` :
  - Titre / Description
  - Prix
  - Surface (m<sup>2</sup>)
  - Nombre de pièces
  - Adresse / ville / code postal
  - Type (maison, appartement, studio...)

### 2. Nettoyage et structuration





- Supprimer les doublons, normaliser les unités (prix/m<sup>2</sup>).
- Extraire les nombres avec des regex (`re.findall()`).
- Compléter les localisations manquantes (géolocalisation automatique avec *Nominatim*).

### 3. Analyse statistique

- Moyenne et médiane du prix au m<sup>2</sup>.

- Répartition des prix selon les régions.
  - Corrélation entre surface et prix.
  - Histogrammes et boxplots (`matplotlib`, `seaborn`).
  - 4. **Visualisation cartographique**
    - Créer une **carte interactive** avec `folium`, où chaque bien est un point avec prix affiché au survol.
  - 5. **Tableau de bord interactif**
    - Avec **Streamlit**, permettre à l'utilisateur de :
      - Choisir une ville ou une région,
      - Filtrer par surface, prix, nombre de pièces,
      - Voir des graphiques et cartes dynamiques.
- 

## Exemples de visualisations

-  Histogramme du prix au m<sup>2</sup> par ville
  -  Carte des annonces géolocalisées (avec `folium`)
  -  Évolution temporelle du prix moyen
  -  Diagramme de corrélation entre surface et prix
- 

## Option avancée (si tu veux aller plus loin)

- Utiliser un **scheduling automatique** (ex. `schedule` ou `cron`) pour actualiser les données chaque semaine.
  - Ajouter un **module de prédiction** (ex. régression linéaire) pour estimer le prix d'un bien selon ses caractéristiques.
  - Mettre le dashboard en ligne sur **Streamlit Cloud** ou **Render**.
- 

## Structure du rapport (proposition)

1. **Introduction**
  - Contexte et enjeux du marché immobilier
  - Objectifs du projet
2. **État de l'art**
  - Présentation du scraping et des outils Python
  - Travaux similaires / études de marché existantes
3. **Méthodologie**
  - Description des sites cibles
  - Architecture technique
  - Stratégie de scraping (pagination, gestion anti-bot)
4. **Résultats et analyses**
  - Statistiques clés et visualisations
  - Tendances observées

## 5. **Discussion et limites**

- Fiabilité des données, biais potentiels
- Améliorations possibles

## 6. **Conclusion**

- Bilan du projet et perspectives