


# Évaluation vs Vulnérabilité

Le marché immobilier face à l'aveuglement climatique *Projet Data Visualisation Machine Learning*

Tom LAFAYE, Balthazar GOUSSET et Nathan GAUTRON

ISUP - Sorbonne Université M2 Actuariat

16 Décembre 2025

 Stack Technique : R (Shiny, sf, leaflet) · Flexdashboard · Random Forest

# Plan de la présentation

- 1 Problématique & Contexte
- 2 Données & Démarche Méthodologique
- 3 Le Dashboard & Résultats Clés
- 4 Conclusion

# L'Immobilier est-il "Aveugle" au Risque ?

## Le Constat Climatique

Le changement climatique intensifie les périls "secs" et "humides" en France. Nous nous concentrons sur deux risques majeurs gérés par le régime Cat'Nat' :

- ☀️ **Retrait-Gonflement des Argiles (RGA)**
- 🌊 **Remontée de Nappes**

## La Question Actuarielle

Ces risques sont cartographiés et connus des experts. Mais l'information redescend-elle jusqu'au prix de marché ?

*Existe-t-il une "décote de risque" intégrée dans les prix au m<sup>2</sup> ?*

**Objectif du Dashboard** : Fournir un outil interactif pour visualiser la confrontation entre l'**Aléa** (Géorisques) et l'**Enjeu Financier** (Valeur Foncière DVF).

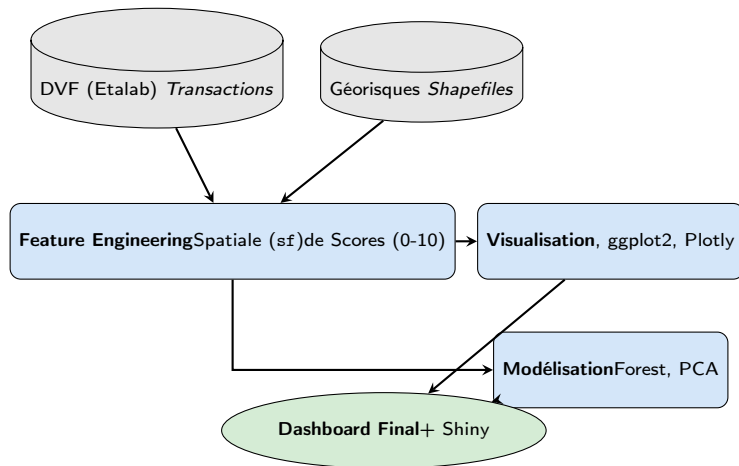
# Les Données : Big Data Immobilier & Climatique

Un croisement de sources hétérogènes volumineuses :

Source	Contenu	Usage
<b>DVF (Etalab)</b> <i>Open Data</i>	Transactions réelles Millions de lignes	Prix/m <sup>2</sup> , Volumes <i>Variable cible</i>
<b>Géorisques</b> <i>Argiles &amp; Nappes</i>	Cartes d'aléas (SHP) Polygones complexes	Exposition physique <i>Variables explicatives</i>
<b>API Géo</b>	Contours administratifs	Jointures spatiales

**Défi technique** : Intersection spatiale massive (Communes x Polygones de risque) gérée via sf et optimisation mémoire.

# Architecture des Données (Pipeline ETL)



*Défis techniques : Gestion de la mémoire pour les intersections spatiales massives Optimisation réactive Shiny.*

# Construction des Indicateurs

Pour répondre aux exigences du projet (Comparaison, Distribution, Cartographie), nous avons structuré l'analyse en 3 temps :

- ❶ **Exploration Statistique (Distribution Comparaison) :**
  - Nettoyage des outliers de prix (méthode IQR).
  - Visualisation *Raincloud Ridgeline* pour comparer les prix par Région.
- ❷ **Scoring Spatial (Cartographie) :**
  - Normalisation des surfaces à risque par commune.
  - Création d'un indicateur synthétique **Score Argile** et **Score Nappes** (0 à 10).
- ❸ **Analyse des Relations (Actuariat) :**
  - Analyse de dépendance par **Copules** (mise en évidence de la dépendance de queue).
  - Matrice de corrélation (Kendall/Spearman).

# Visualiser les Distributions

Utilisation de packages avancés (`ggdist`, `ggridges`) pour révéler la structure des prix.

## Raincloud Plot :

- Combine densité, boxplot et points bruts.
- Permet de voir les outliers et la multimodalité d'un coup d'œil.

## Ridgeline Plot :

- Comparaison des distributions de prix par Région.
- "Carte thermique" des densités.

# Visualiser la Dépendance

L'indépendance des risques est une hypothèse dangereuse. Nous l'avons testée visuellement.

## Analyse par Copules (3D)

- Utilisation de `plotly` pour visualiser la densité conjointe des rangs de risques.
- Mise en évidence de la **dépendance de queue** (Gumbel) : les extrêmes s'attirent.
- *Conclusion visuelle* : Un point rouge en haut à droite du cube 3D confirme le risque systémique.



# Le Dashboard Interactif (D mo)

L'outil final (.Rmd compil  avec Shiny) offre une exp rience utilisateur compl te :

## Fonctionnalit s Cl s :

- **Cartographie Multi-Couches :**  
Bascule dynamique entre Argiles, Nappes et Clusters.
- **🔍 Inspection :** Zoom sur les top-communes expos es.
- **🏠 ML Live :** L'utilisateur lance le Random Forest en temps r el.

### Interactivit 

Utilisation de crosstalk et shiny pour lier les filtres (D partement, Taille d' chantillon)   tous les graphiques simultan ment.

# Résultat 1 : Analyse Supervisée (Prix vs Risque)

Nous avons entraîné un modèle **Random Forest** pour prédire le prix au m<sup>2</sup>. *Variable Cible : Prix/m<sup>2</sup> | Variables : Surface, Lat, Lon, **Score Risque**.*

## Feature Importance (VIP) :

- 1 **Localisation (Lat/Lon)** : Impact critique (L'emplacement fait le prix).
- 2 **Surface** : Impact fort (Effet taille).
- 3 **Type de bien** : Impact moyen.
- 4 ...
- 5 **Score Risque** : Impact Quasi-Nul.

## Conclusion Majeure

Le modèle n'utilise pas le risque pour améliorer sa prédiction. **Le marché est donc aveugle : le risque n'est pas pricé.**

## Résultat 2 : Analyse Non-Supervisée

Conformément aux "Bonus" du projet, nous avons implémenté une réduction de dimension.

### Clustering (K-Means) PCA

Nous avons segmenté les communes en 3 clusters basés sur le Prix et le Risque.

- **Cluster 1** : Prix élevés, Risque faible (Zones refuges).
- **Cluster 2** : Prix faibles, Risque élevé (Zones vulnérables).
- **Cluster 3** : Zones intermédiaires.

**Visualisation** : Projection sur les deux premières composantes principales (PCA) intégrée interactivement dans le dashboard (plotly).

## Bilan Technique :

- ✓ Maîtrise du pipeline Rmd/Shiny.
- ✓ Traitement de données spatiales lourdes.
- ✓ 6 types de visualisations intégrés (Bar, Map, Scatter, Box, Density, PCA).

## Bilan Métier :

- ⚠ Confirmation de la "Myopie" du marché.
- 💡 **Ouverture** : Vers une "Prime de Résilience" ?  
Visualiser ce que *devrait* être le prix si le risque était intégré.

**Merci de votre attention.**

*Livrables disponibles : Code, Rapport HTML, Dashboard Interactif, Slide.*