

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

26 Décembre 2025



Présentation, HTML, Dashboard

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 m² ?

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

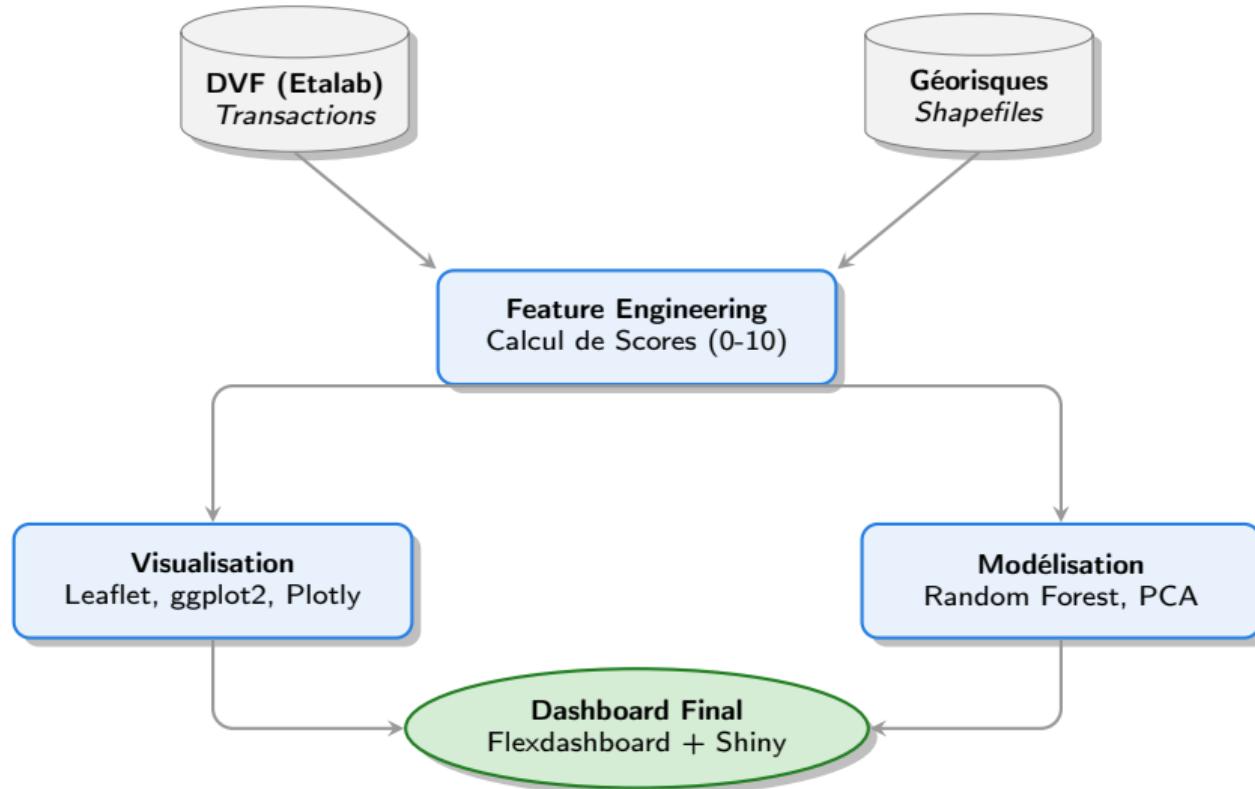
Les Données : Big Data Immobilier & Climatique

Un croisement de sources hétérogènes volumineuses :

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



Construction des Indicateurs

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

① Exploration Statistique (Distribution Comparaison) :

- Nettoyage des outliers de prix (méthode IQR).
- Visualisation pour comparer les prix par Région.

② Scoring Spatial (Cartographie) :

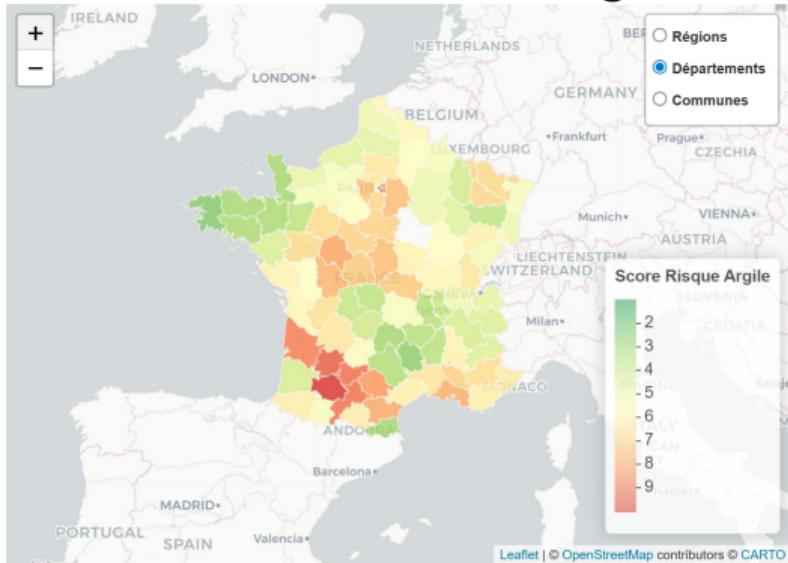
- Normalisation des surfaces à risque par commune.
- Création d'un indicateur synthétique **Score Argile** et **Score Nappes** (0 à 10).

③ Analyse des Relations (Actuariat) :

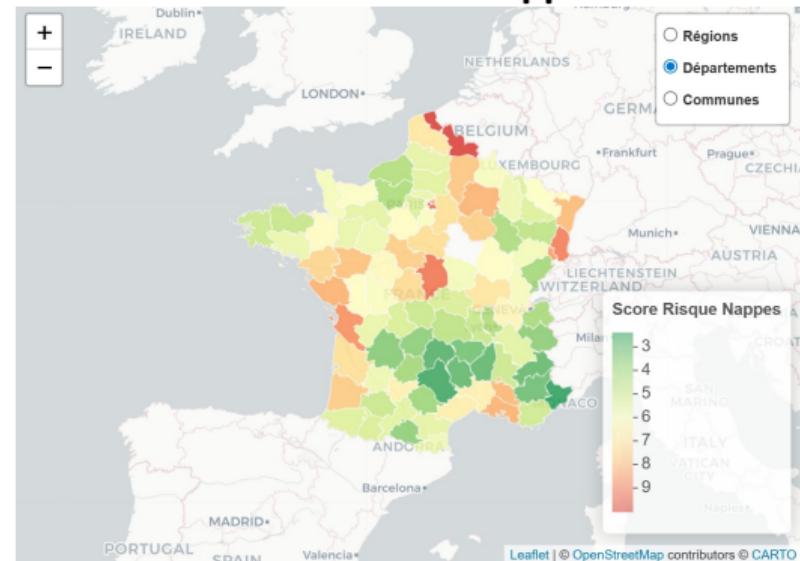
- Analyse de dépendance par **Copules** (mise en évidence de la dépendance de queue).
- Proposition de Prime de risque climatique

Cartographie des Risques (Géorisques)

Retrait-Gonflement des Argiles

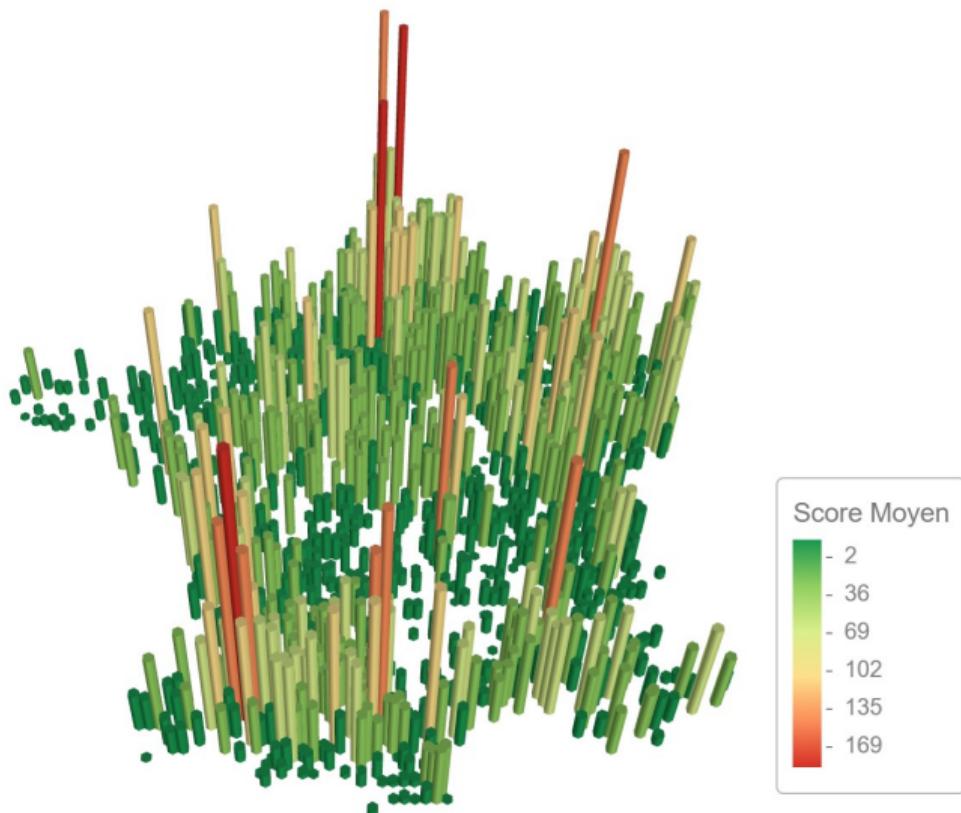


Remontée de Nappes



Score normalisé (0-10) calculé par intersection spatiale pondérée.

Risque Combiné & Points Chauds



Approche "Hotspot"

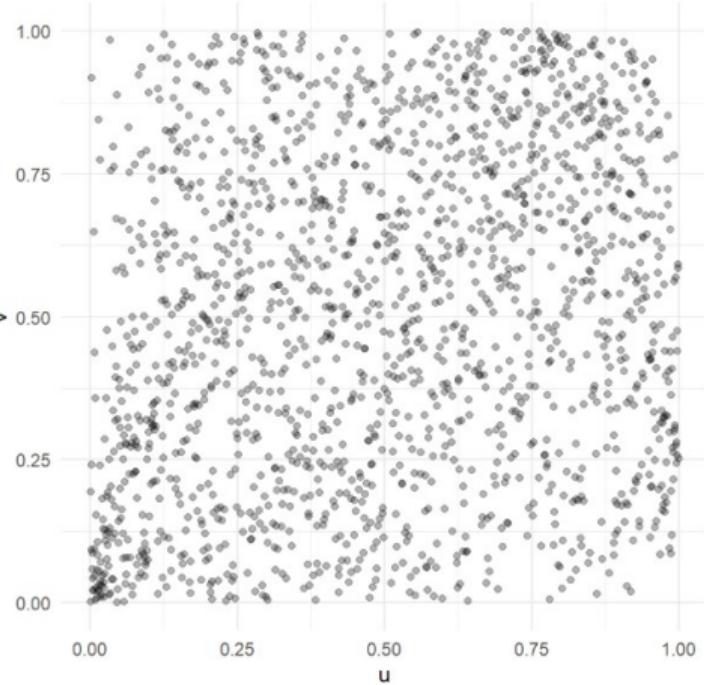
- Agrégation par hexagones.
- **Score Combiné =**
 $Score_{Argile} \times Score_{Nappe}$.
- Identification visuelle des cumuls de périls.

Visualiser la Dépendance : Vue 2D (Rangs)

L'indépendance des risques est une hypothèse dangereuse. Comparaison des rangs empiriques et simulés.

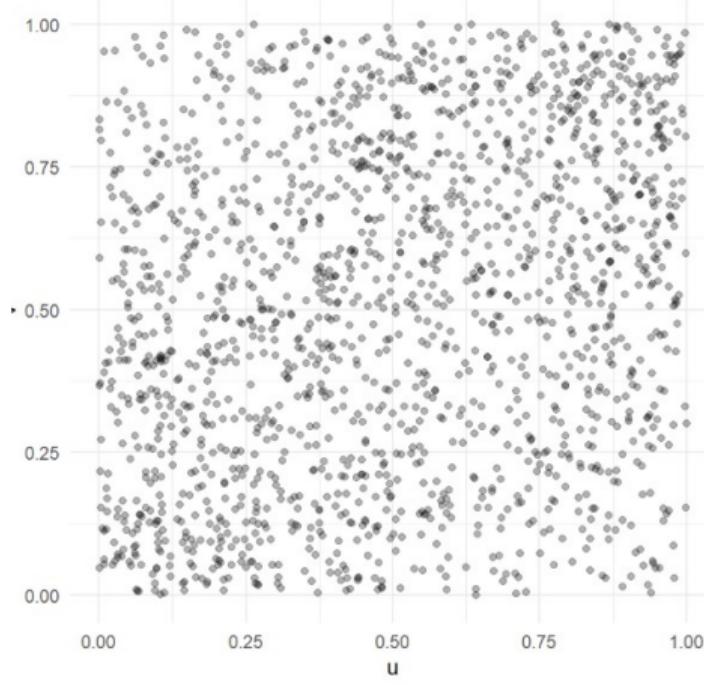
Données Empiriques

Données Originales



Simulation Théorique

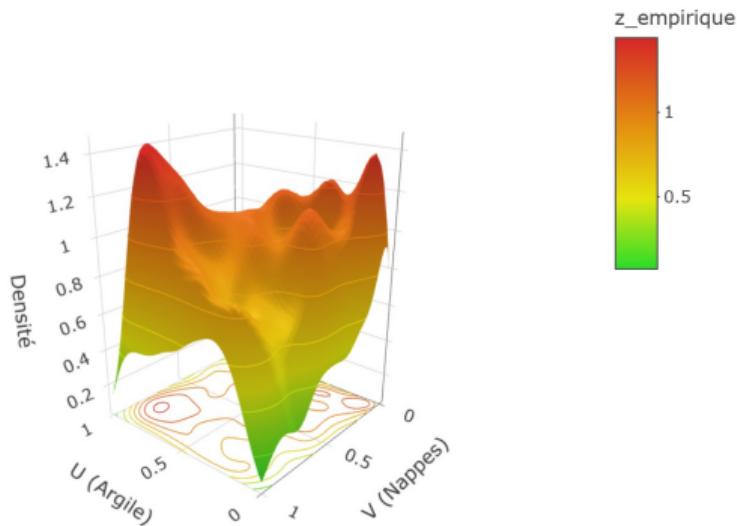
Simulation Frank (theta = 1.38)



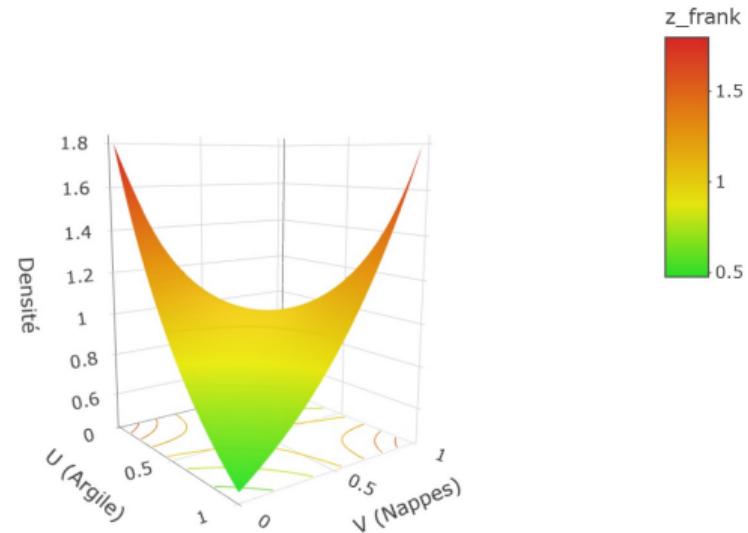
Visualiser la Dépendance : Vue 3D (Densité)

Visualisation de la densité de probabilité conjointe.

Densité Empirique



Densité Théorique



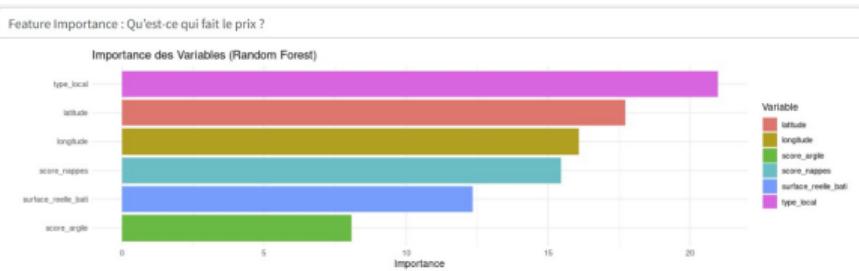
Confirmation de la dépendance de queue (les extrêmes s'attirent).

Résultat 1 : Analyse Supervisée (Prix vs Risque)

Nous avons entraîné un modèle **Random Forest** pour prédire le prix au m².

Feature Importance (VIP) :

- ① **Localisation (Lat/Lon)** : Impact critique.
- ② **Surface** : Impact fort.
- ③ **Type de bien** : Impact moyen.
- ④ ...
- ⑤ **Score Risque : Impact Quasi-Nul.**



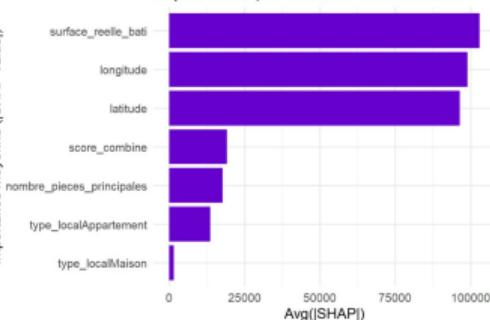
Conclusion Majeure

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

Interprétabilité du Modèle (SHAP & PDP)

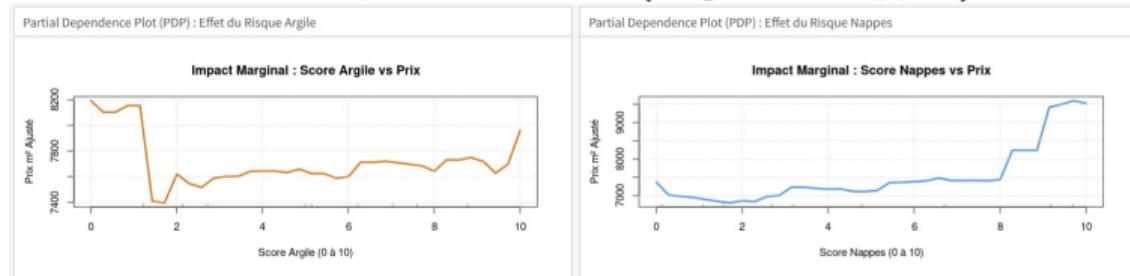
SHAP Summary Plot

Moyenne de l'impact absolu sur la valeur foncière



Contribution de chaque variable à la prédiction.

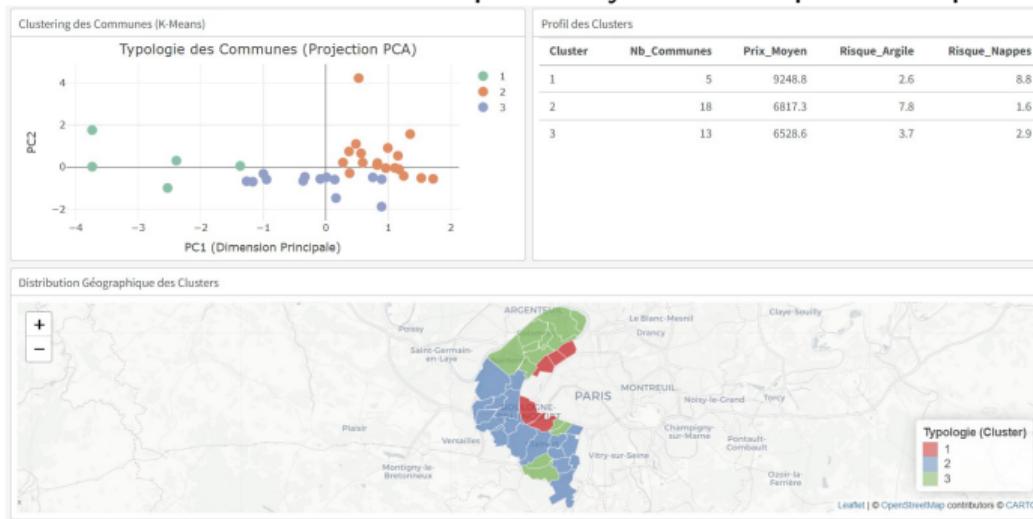
Partial Dependence Plot (Argile & Nappes)



Effet marginal du risque sur le prix (plat = aveuglement).

Résultat 2 : Analyse Non-Supervisée

Réduction de la dimension par analyse en composantes principales .



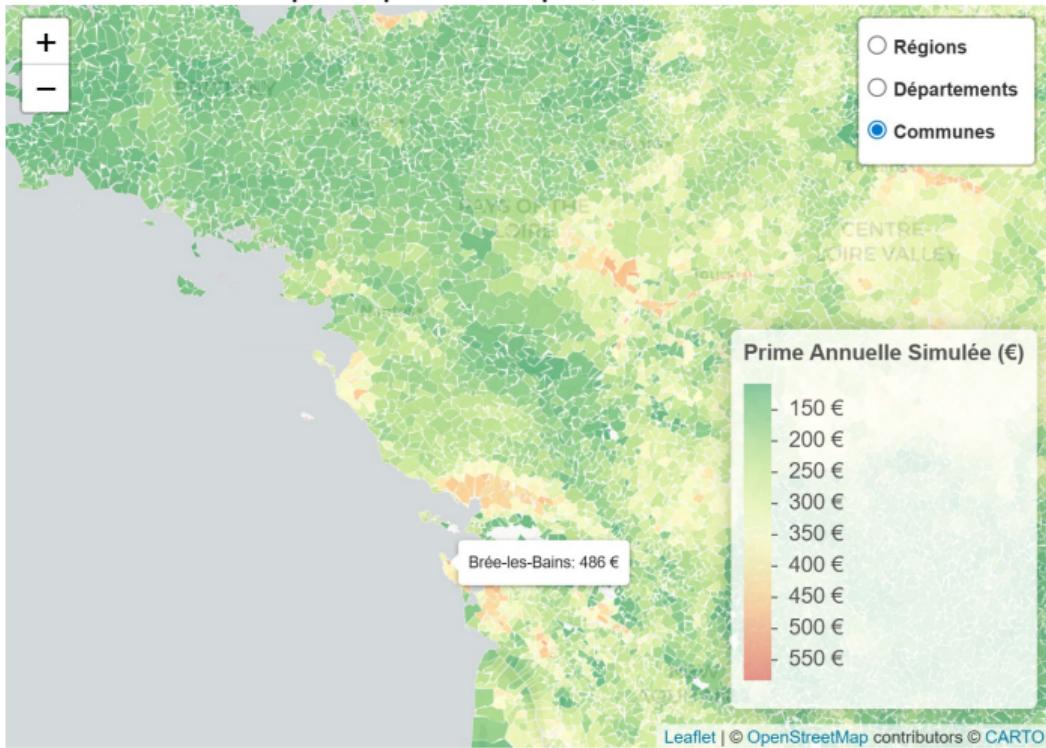
Clustering (K-Means) sur PCA

Segmentation des communes en 3 clusters :

- **Cluster 1** : Refuges (Prix ↑, Risque ↓).
- **Cluster 2** : Vulnérables (Prix ↓, Risque ↑).
- **Cluster 3** : Intermédiaires.

Vers une Tarification du Risque ?

Si le marché ne price pas le risque, l'assureur le doit.



Simulation de Prime Pure

- Projection du coût moyen des sinistres par commune.
- Basée sur l'exposition Argile + Nappes.
- Révèle les zones de sous-tarification actuelle.

Bilan Technique :

- ✓ Traitement de données spatiales lourdes.
- ✓ Différents types de visualisations intégrés (Leaflet, Hexbin, Copules, Feature Importance, SHAP, PDP, PCA).
- ✓ Maîtrise du pipeline Rmd/Shiny.

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, Slides.