



# Analyse des leviers de rénovation énergétique

Approche statistique et Machine Learning interprétable

Kiné DIOP-GUEYE



# Sommaire

- Problématique & objectifs
- Données & data engineering
- Analyse exploratoire & statistique (FAMD, ANOVA)
- Approche Machine Learning
- Évaluation des modèles
- Comparaison des variables importantes
- Interprétation SHAP
- Renovation énergétique
- Impact des variables clés sur les autres classes DPE
- Conclusion

# Problématique & objectifs

Quelles variables prioriser pour atteindre les classes DPE A et B ?

## Problématique

Sur quelles variables se concentrer en cas de rénovation pour améliorer la classe DPE d'un logement ?

## Objectifs

- Identifier les variables explicatives des classes A/B
- Distinguer variables explicatives et variables actionnables
- Mesurer l'impact réel d'une rénovation



# Données & data engineering

01

## Sources de données

Fusion de deux bases : DPE logements et consommation énergétique via join\_key

DucKDB SQL



1.119.319 lignes

02

## Nettoyage

Traitement des valeurs manquantes, détection et suppression des outliers, élimination des doublons

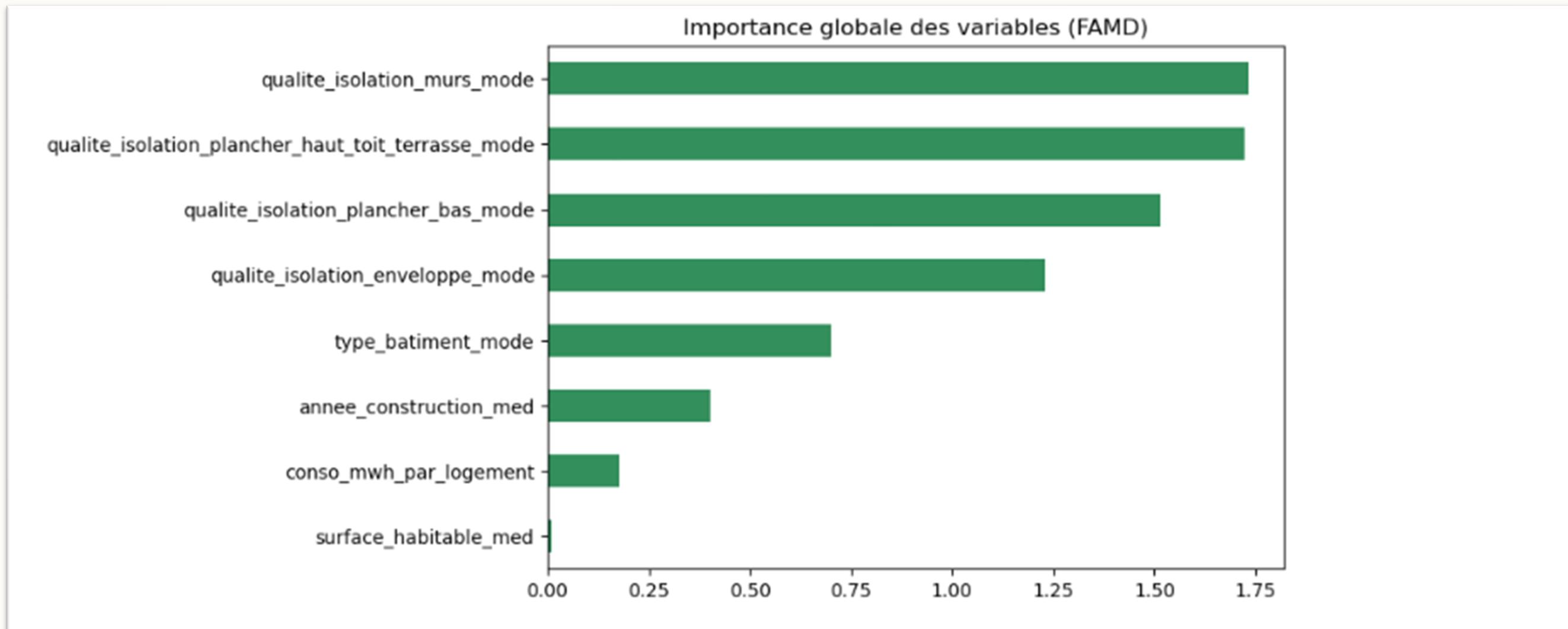
03

## Sélection

Choix des variables pertinentes pour l'analyse et encodage des variables catégorielles

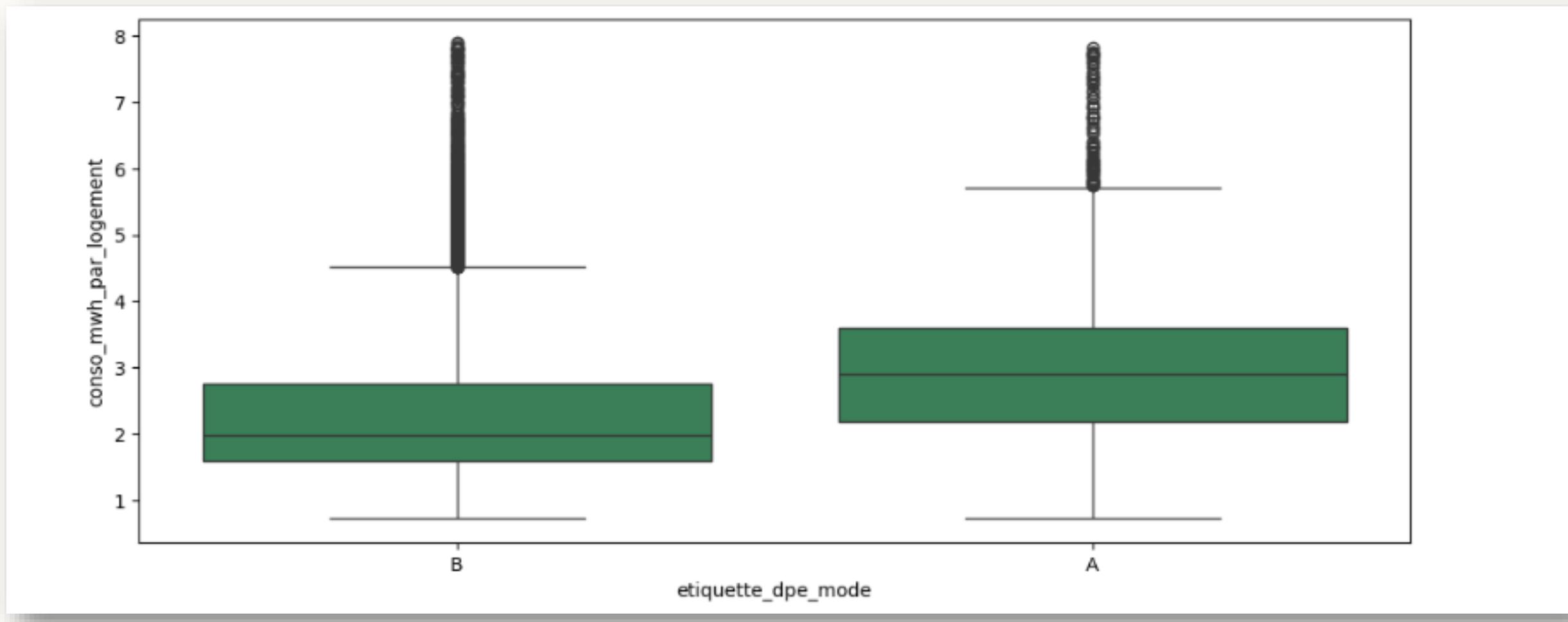
# Analyse exploratoire & statistique (FAMD, ANOVA)

FAMD Analyse factorielle des données mixtes (numériques et catégorielles)



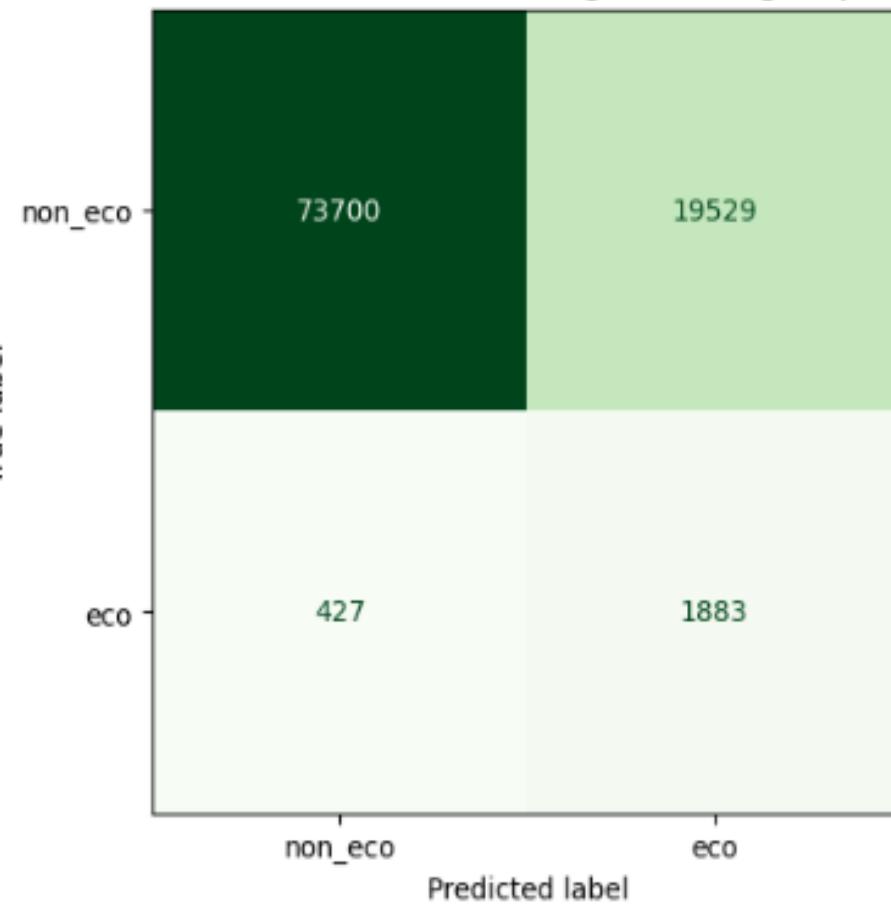
# Analyse exploratoire & statistique

ANOVA Test de significativité des variables explicatives



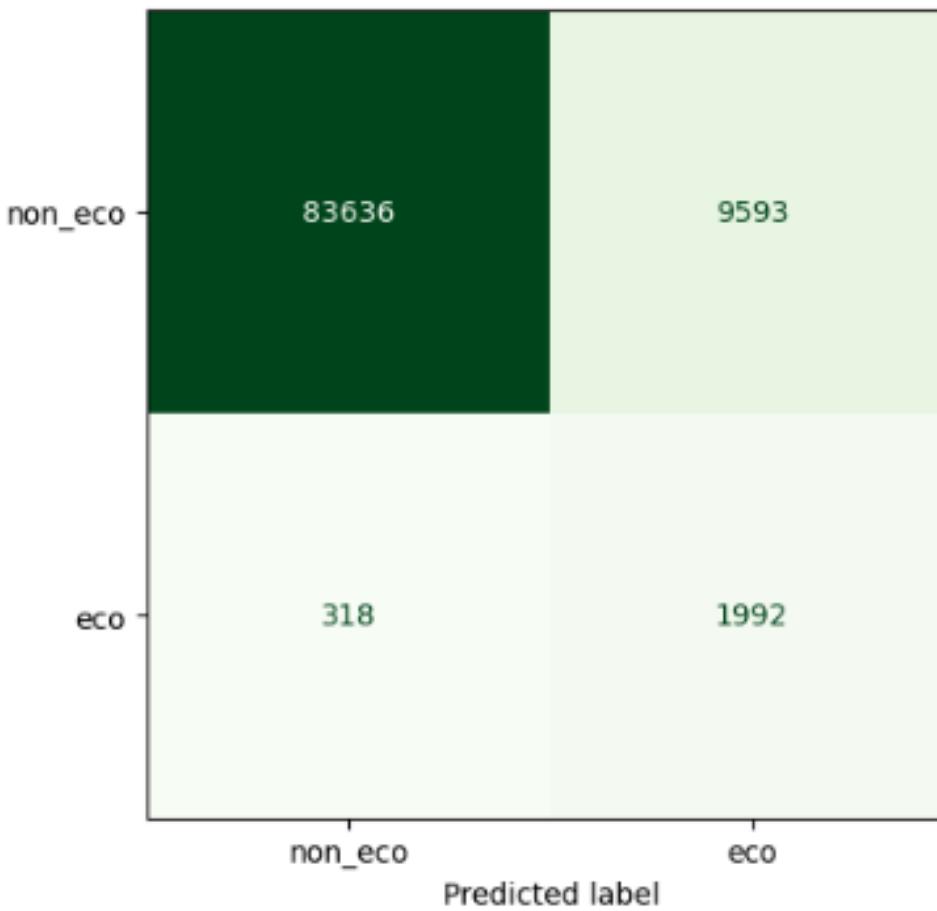
# Approche Machine Learning

Matrice de confusion - Régression logistique



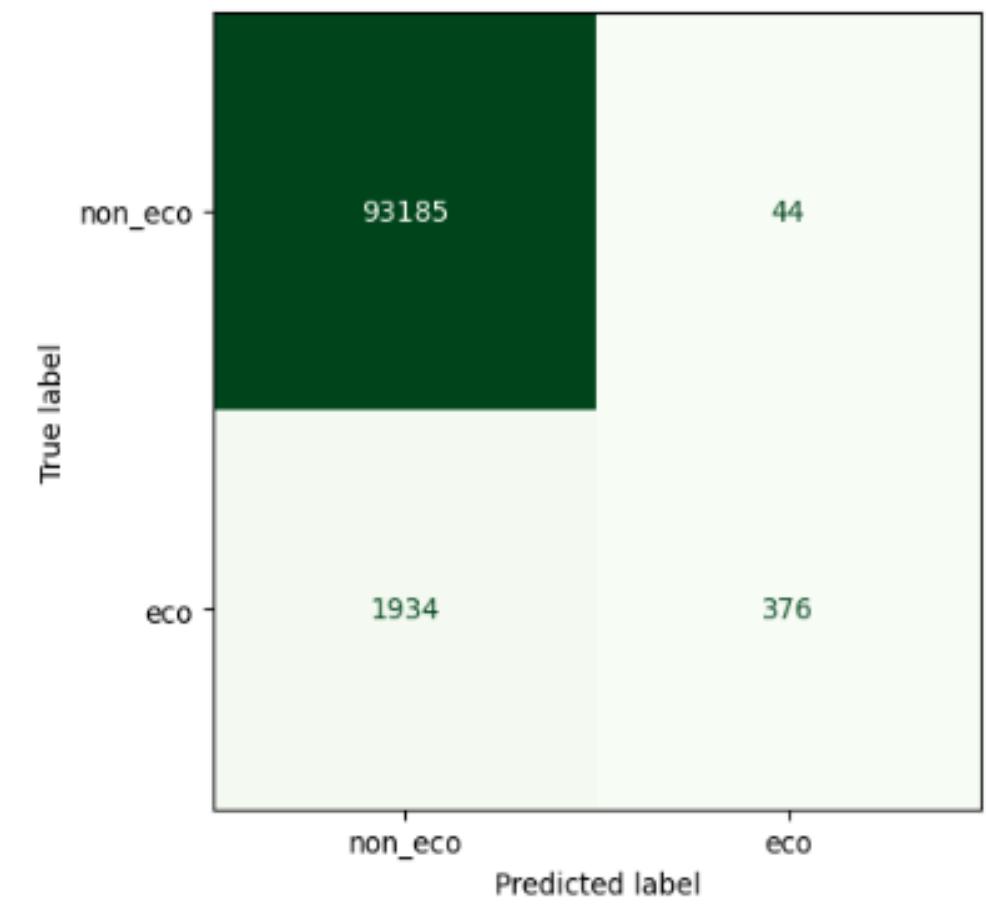
Régression logistique

Matrice de confusion - Random Forest



Random Forest

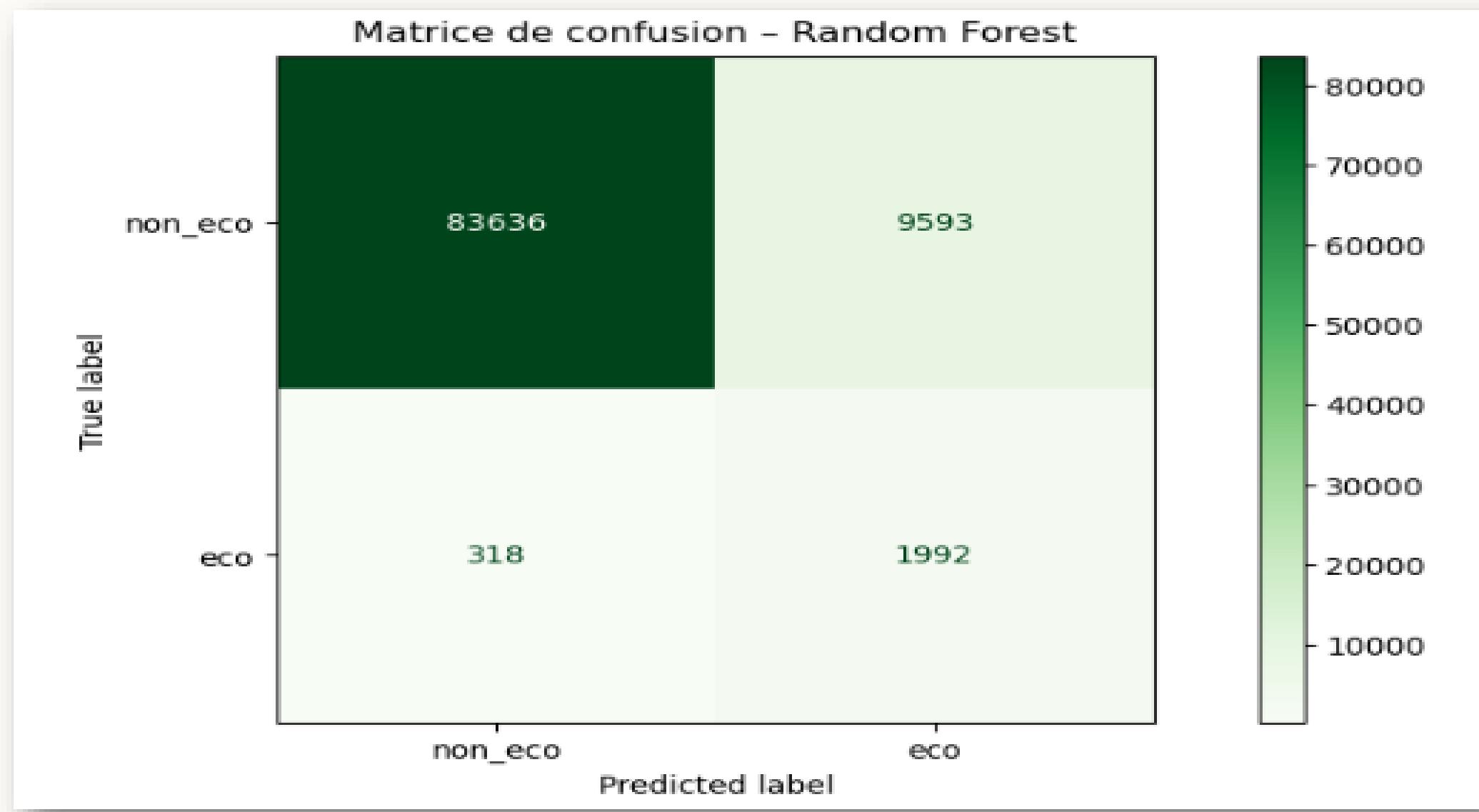
Matrice de confusion - XGBoost



XGBoost

# Approche Machine Learning

## Random Forest



# Approche Machine Learning

## Évaluation des modèles

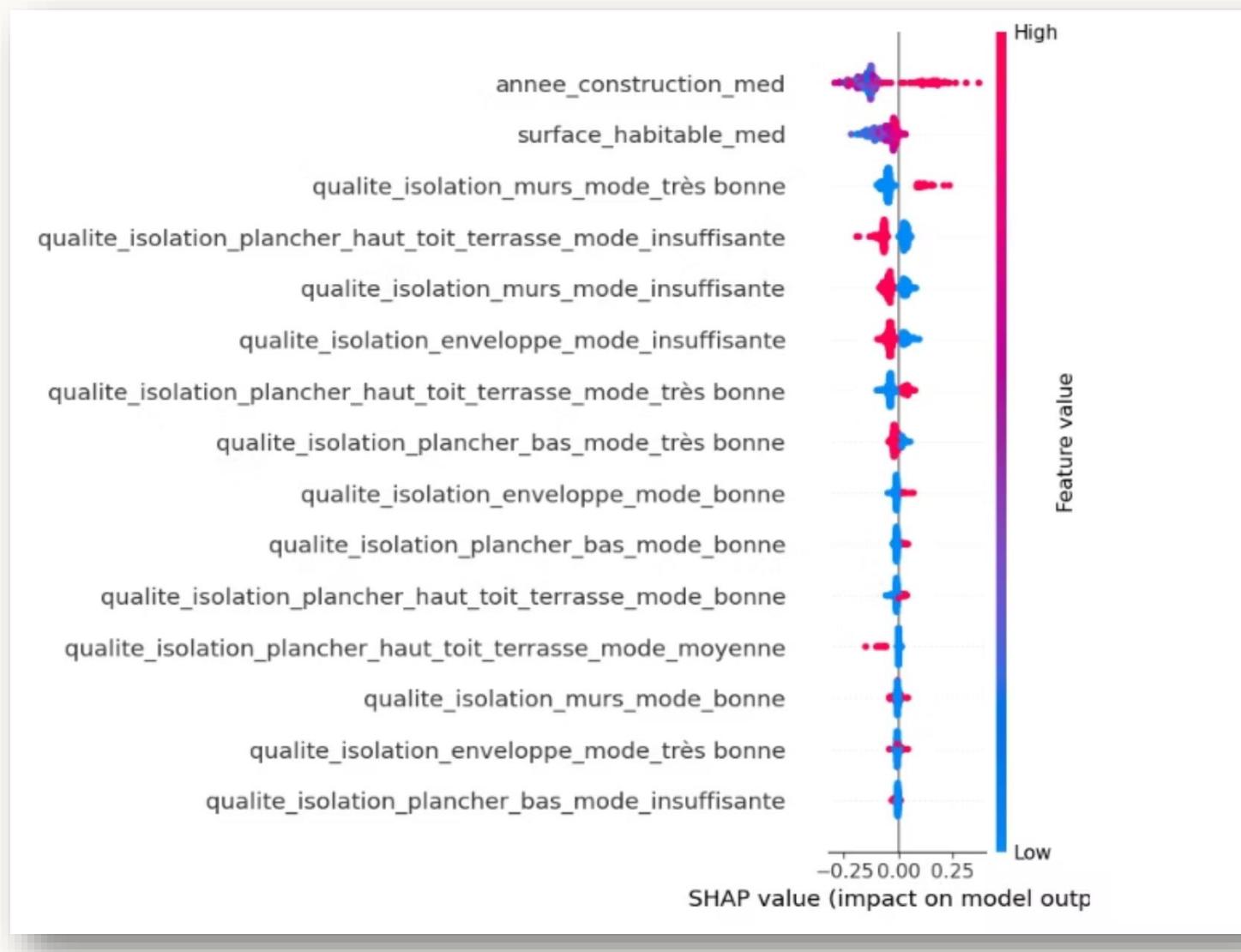
Modèle	Recall eco	F1 eco	ROC-AUC
Régression logistique	0.803	0.158	0.866
<b>Random Forest</b>	<b>0.854</b>	<b>0.283</b>	<b>0.933</b>
XGBoost	0.156	0.267	0.965

# Comparaison par variables "mères"

Variable mère	Logistique	Random Forest	XGBoost
Qualité de l'isolation	✓ Dominante	✓ Dominante	✓ Dominante
Année de construction	✓ Secondaire	✓ Importante	✓ Secondaire
Surface	✓ Modérée	✓ Modérée	✗ Faible
Type de bâtiment	✗ Faible	✗ Faible	✗ Faible

Conclusion principale : La qualité de l'isolation = levier principal de rénovation énergétique

# SHAP (Random Forest)



## Résultats SHAP

### Explicative

Année de construction (impact fort mais non modifiable)

### Actionnable

Isolation murs, toiture, enveloppe

Variable explicative ≠ variable actionable : l'année explique beaucoup mais l'isolation est le levier de rénovation

# Interprétation SHAP (Random Forest)

## Pourquoi SHAP ?

- Comprendre les décisions du modèle
- Identifier le sens de l'impact (positif/négatif)
- Différencier explicabilité et actionnabilité

## Résultats SHAP

### Explicative

Année de construction (impact fort mais non modifiable)

### Actionnable

Isolation murs, toiture, enveloppe

Variable explicative ≠ variable actionnable : l'année explique beaucoup mais l'isolation est le levier de rénovation

# Lecture "rénovation énergétique"



Année ≠ action

L'année explique beaucoup mais n'est pas modifiable



Isolation = levier

Principal levier de rénovation actionnable

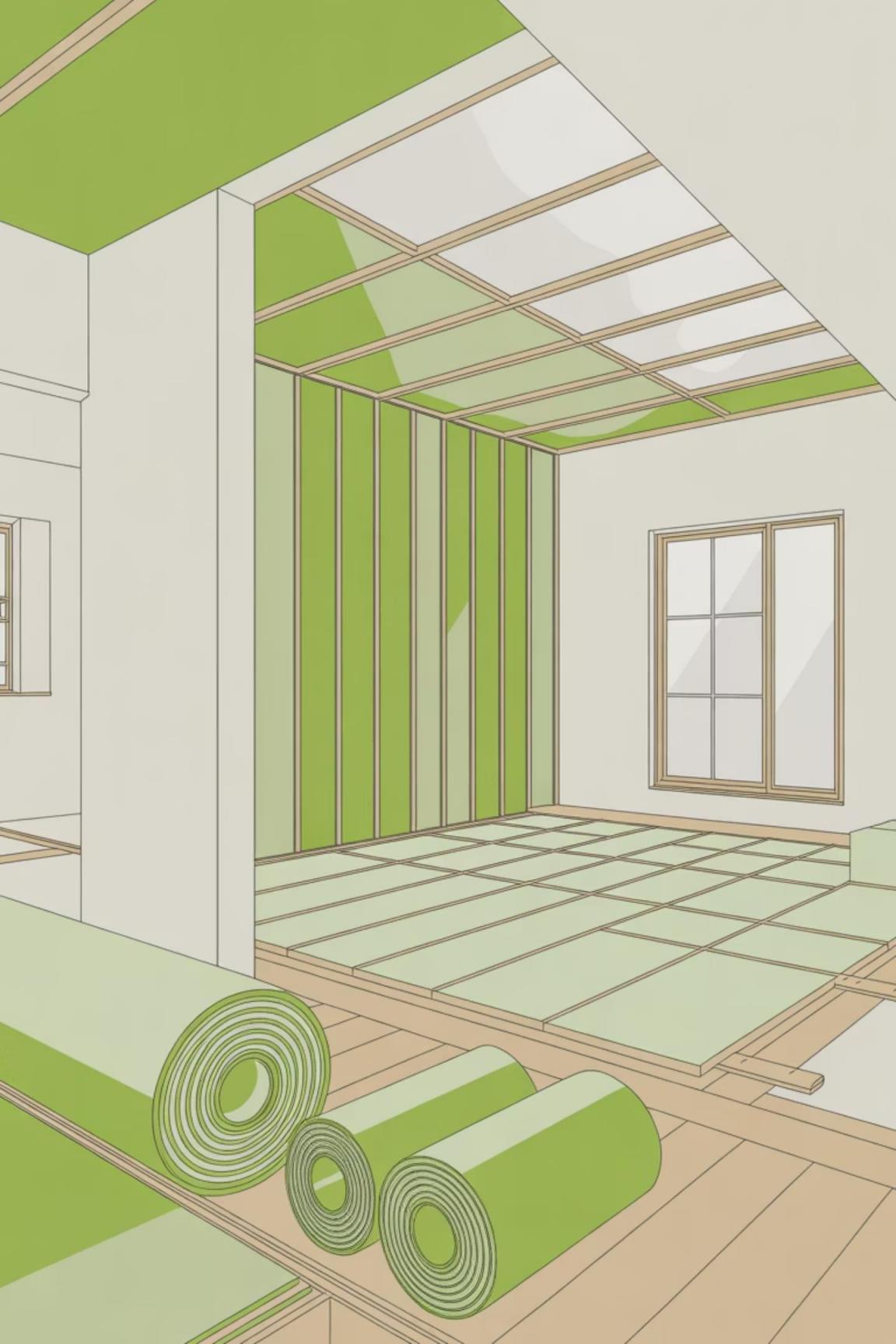


Confirmation

Statistiques, ML et SHAP convergent

## Recommandation

L'isolation est la variable retenue pour l'analyse d'impact et doit être priorisée dans les programmes de rénovation énergétique



# Analyse de l'impact des variables clés sur les autres classes DPE

Analyse par probabilité prédictive

Impact par classe DPE (C, D, E, F,G)

Probabilité de changement de classe (seuil)

# Analyse de l'impact des variables clés sur les autres classes DPE

Analyse par probabilité prédictive

	proba_avant	proba_apres
count	93229.000000	93229.000000
mean	<b>0.156651</b>	<b>0.254786</b>
50%	<b>0.064967</b>	<b>0.148134</b>



+ de 9,8 points de probabilité grâce à une action ciblée



+ de la moitié des logements gagnent au moins 3 points.

# Analyse de l'impact des variables clés sur les autres classes DPE

Impact par classe DPE (C, D, E, F, G)

Classe DPE	Impact ( $\Delta$ Probabilité)
C	0.073
D	<b>0.119</b>
E	<b>0.111</b>
F	0.084
G	0.046

La rénovation est plus efficace sur les logements intermédiaires D, E.

# Analyse de l'impact des variables clés sur les autres classes DPE

Probabilité de changement de classe (seuil)

10%

Environ 1 logement sur 10 passe réellement de non A/B → A/B

# Conclusion

## Isolation

Principal levier actionnable (fortement influente et directement actionnable)

## Année de construction

Variable explicative mais NON modifiable

## Impact mesurable

~1 logement sur 10 change de classe énergétique

## Cibles prioritaires

Classes D et E (meilleur compromis gain/effort)

- ❑ Les modèles interprétables guident efficacement les décisions de rénovation en identifiant les leviers les plus impactants

# Merci