

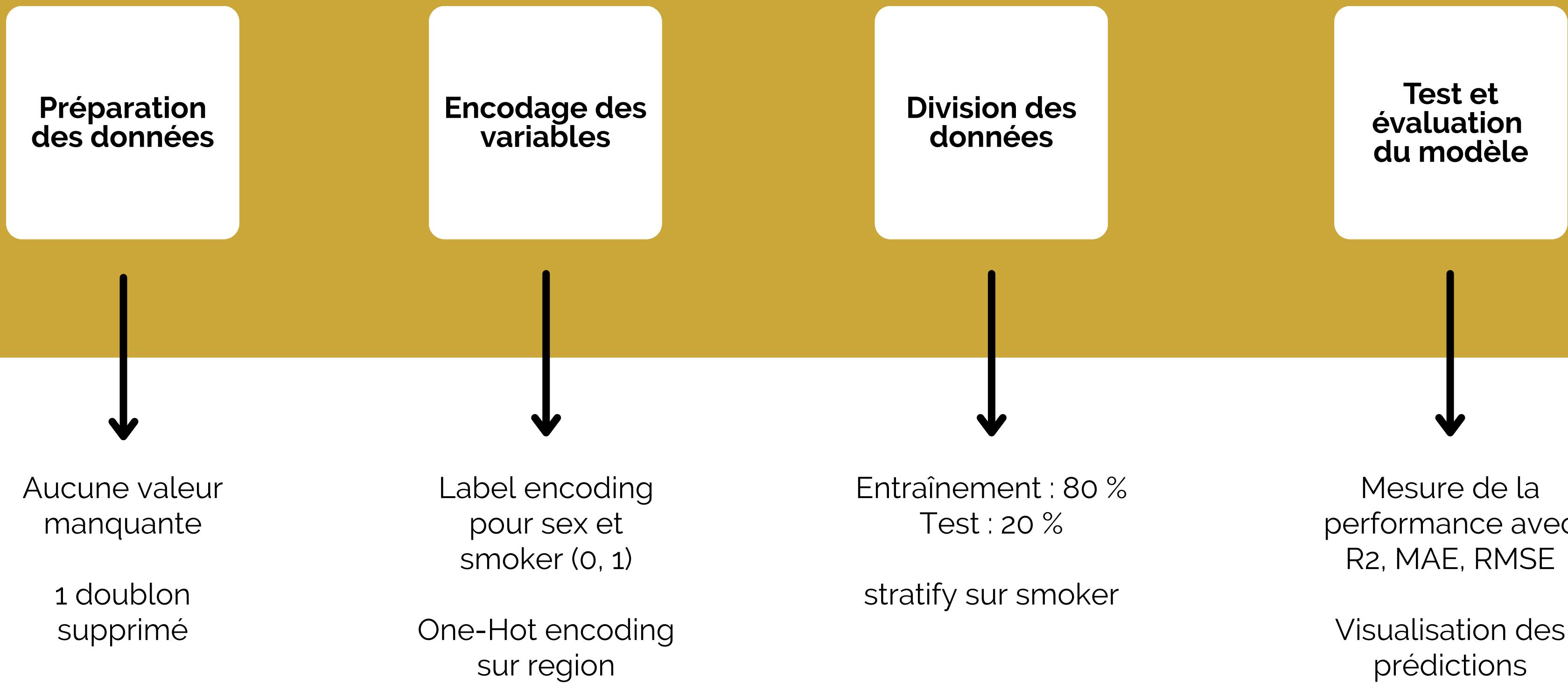
# PRÉDICTION DES CHARGES D'ASSURANCE : MODÈLE

Fatima, Umberto, Flora

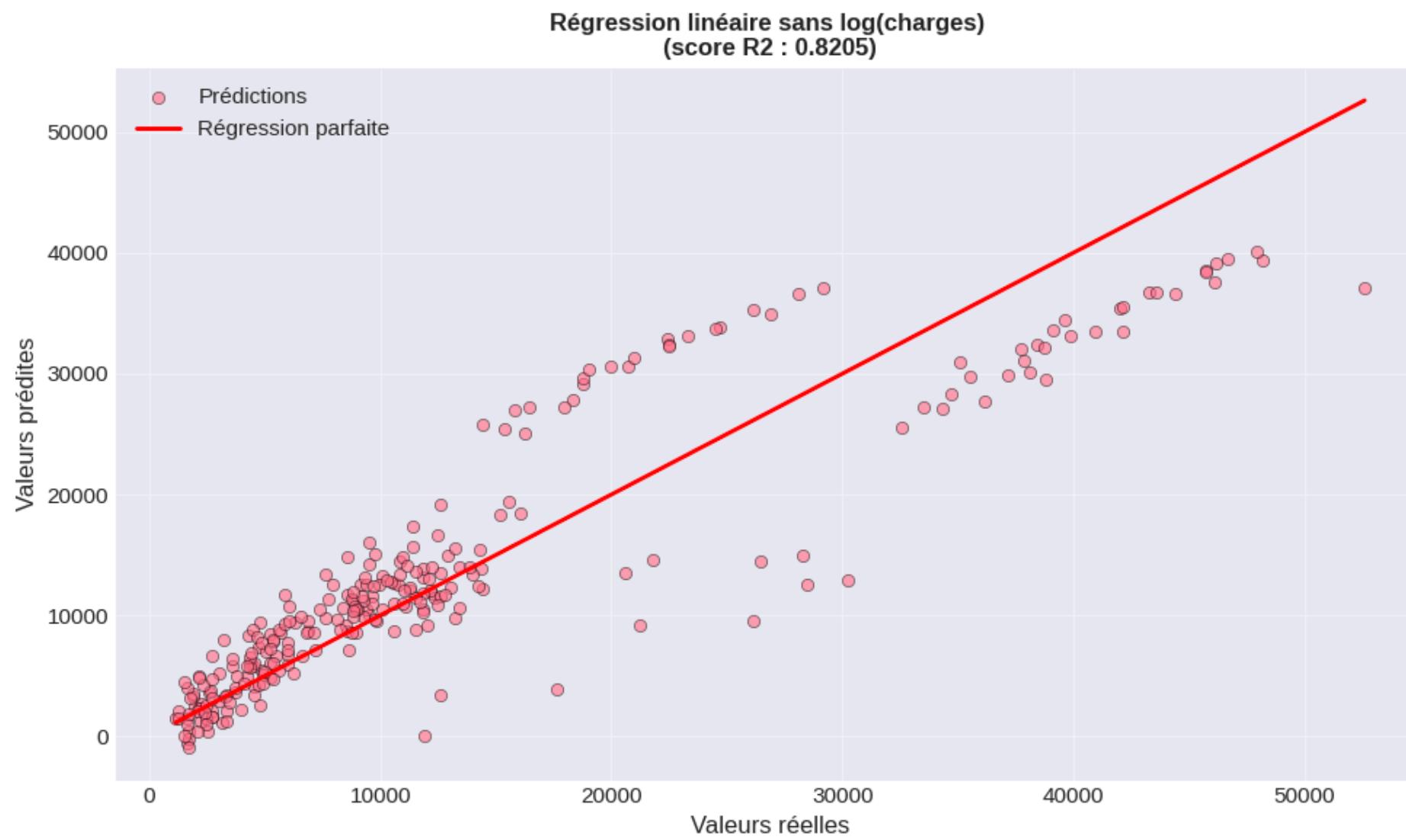
# Objectif

- **Contexte** : élaboration d'un modèle pour prédire les charges d'assurance de nouveaux clients
- **Ce qui a été fait** : analyse de données de clients existants
- **Observations principales** :
  - Fort impact du tabagisme sur les charges
  - Impact de l'obésité chez les fumeurs
  - Impact linéaire de l'âge
- **Objectif actuel** : produire un modèle de régression linéaire performant

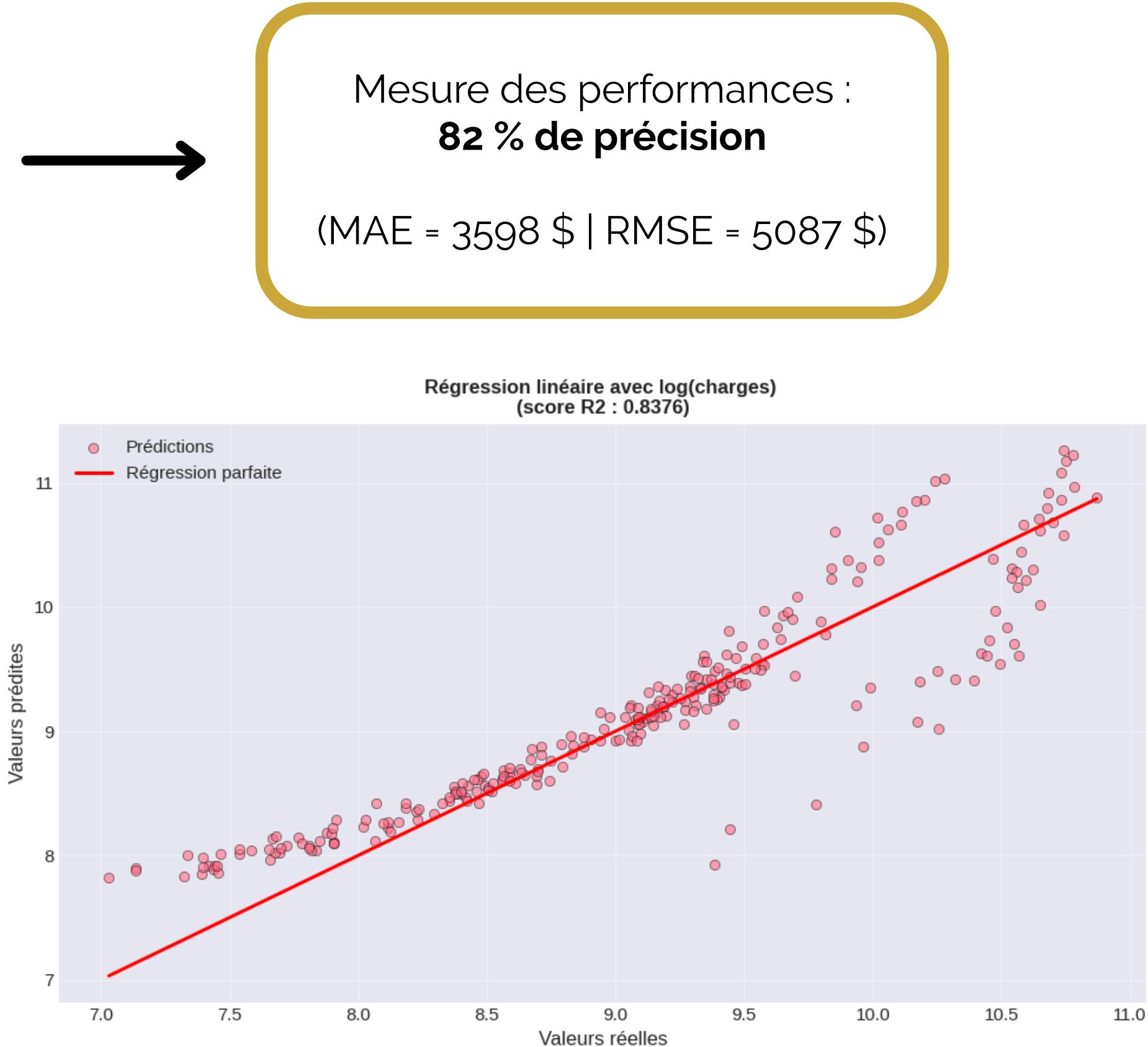
# Pipeline pour la modélisation



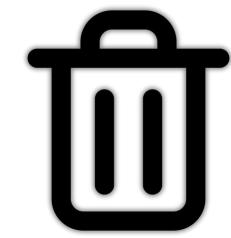
# Résultats du premier modèle



Performance avec log(charges) :  
**84 % de précision**



# Améliorations : choix techniques



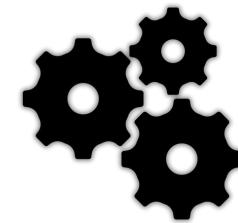
Suppression des régions



Ajout d'interactions :  
 $\text{smoker} * \text{bmi}$   
 $\text{smoker} * \text{is\_obese}$   
Suppression du bmi



Comparaison avec et  
sans log sur les charges  
→ Pas de log(charges)



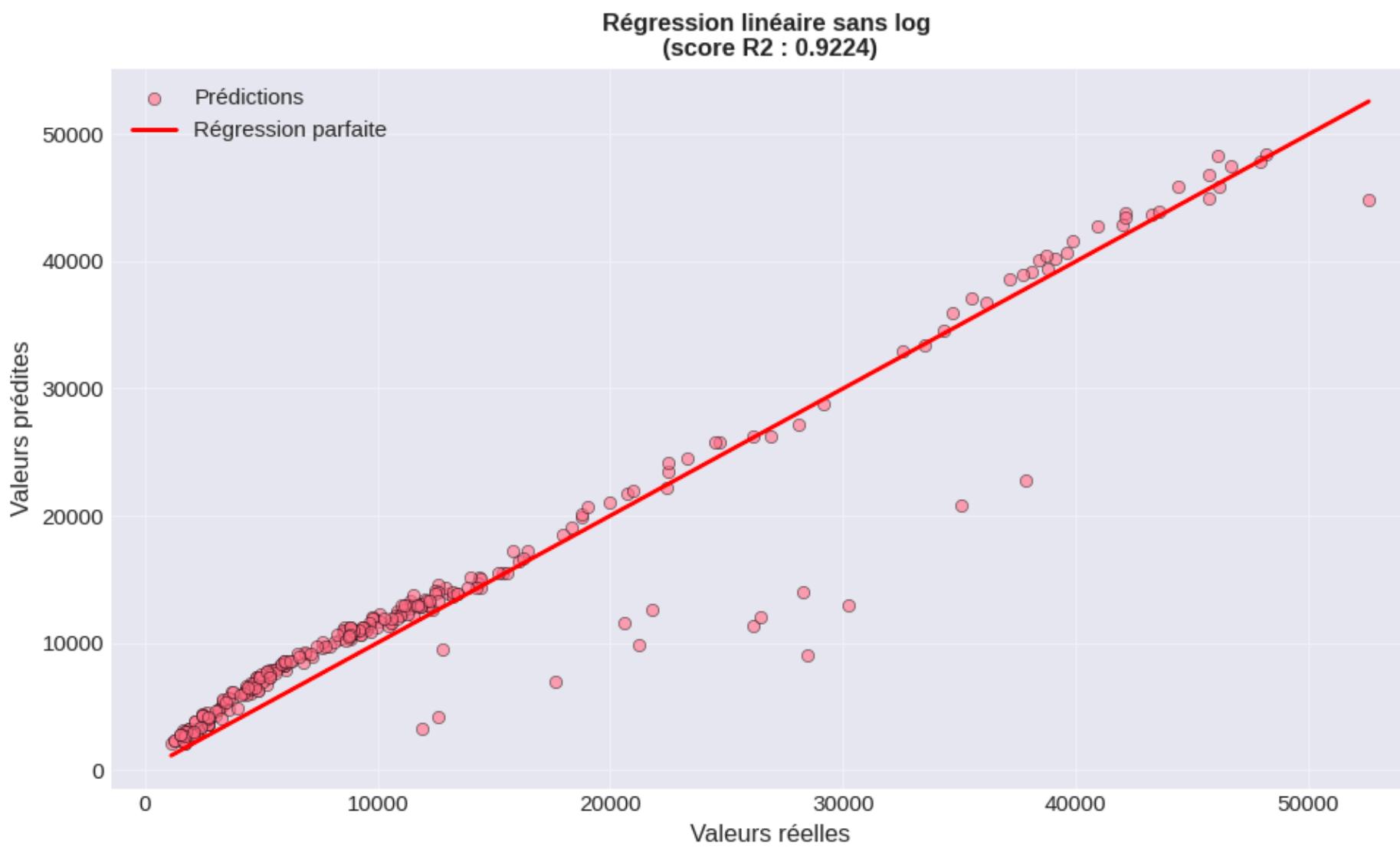
Standardisation de  
age, children  
après le split pour  
éviter la fuite de  
données

# Résultats après amélioration

Ancien score :  
**R2 = 84 %**



Modèle amélioré :  
**R2 = 92%**  
MAE = 2027 \$  
RMSE = 3345 \$



## Performance du modèle

Analyse  
des résidus

Ridge  
& Lasso

Polynomial  
features

Cross-  
validation

## Automatisation et déploiement

Pipeline  
automatisée

Application  
Streamlit

**Axes  
d'amélioration**

# Merci pour votre attention !

---

**Des questions ?**