



# L'IA au service des agents immobiliers

---







# En quoi un modèle de machine Learning peut aider

---

Difficile de faire des estimations de bien personnalisées suite à une forte demande



modèle prédictif pour prédire la valeur des logements

# Base de donnée Utilisé

Prix médians des logements pour les districts de Californie

Localisation

Population

Logement

Pas de duplication

Quelques valeurs manquantes ( 1 % colonne total\_bedroom)

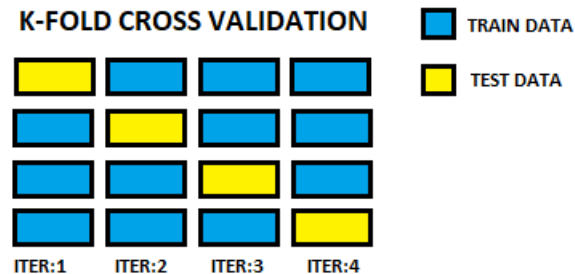
Présence d'Outliers



# Modèle choisis et les résultats obtenues

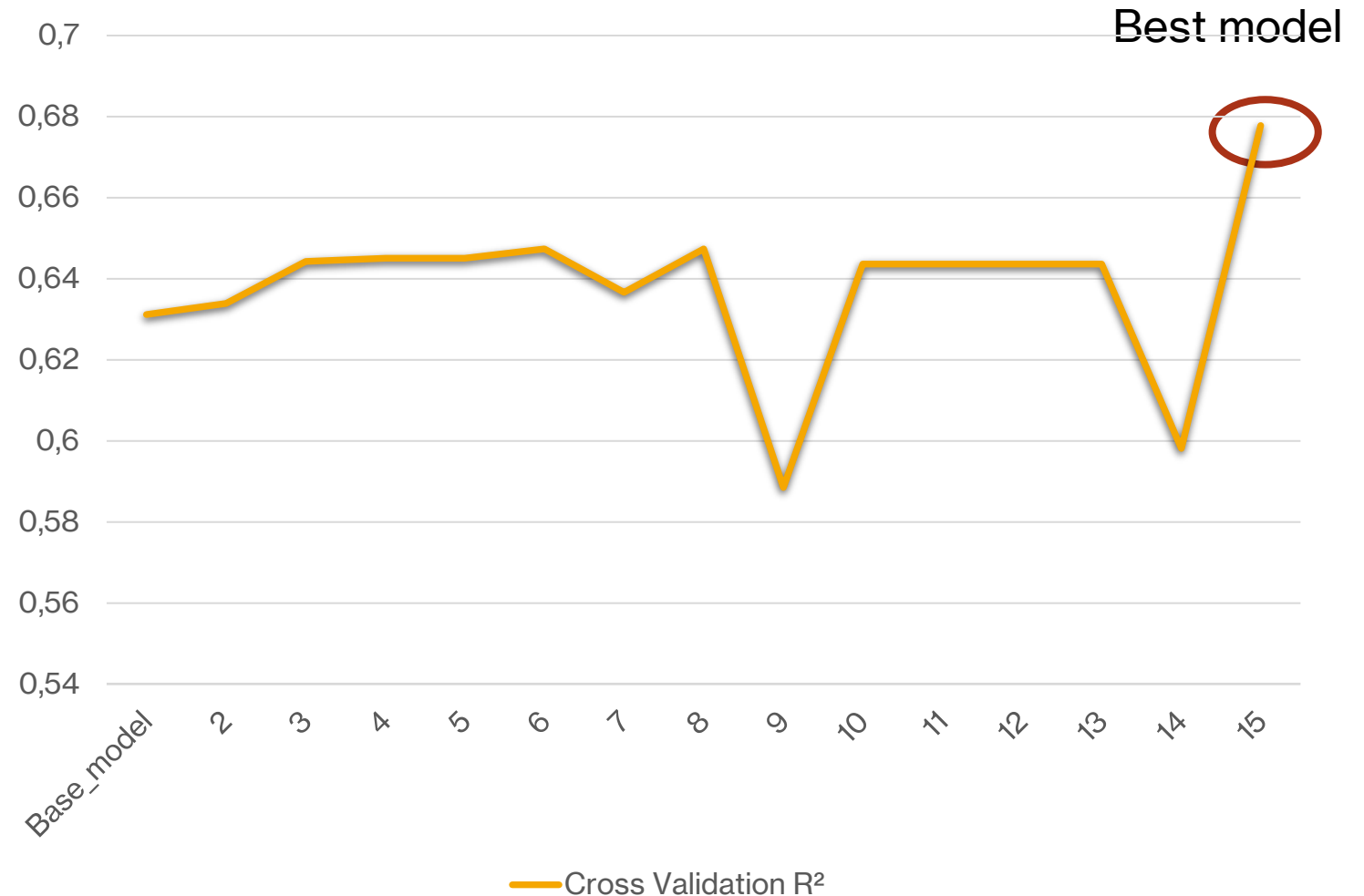
Model testé : la régression linéaire

Utilisation d'une cross validation



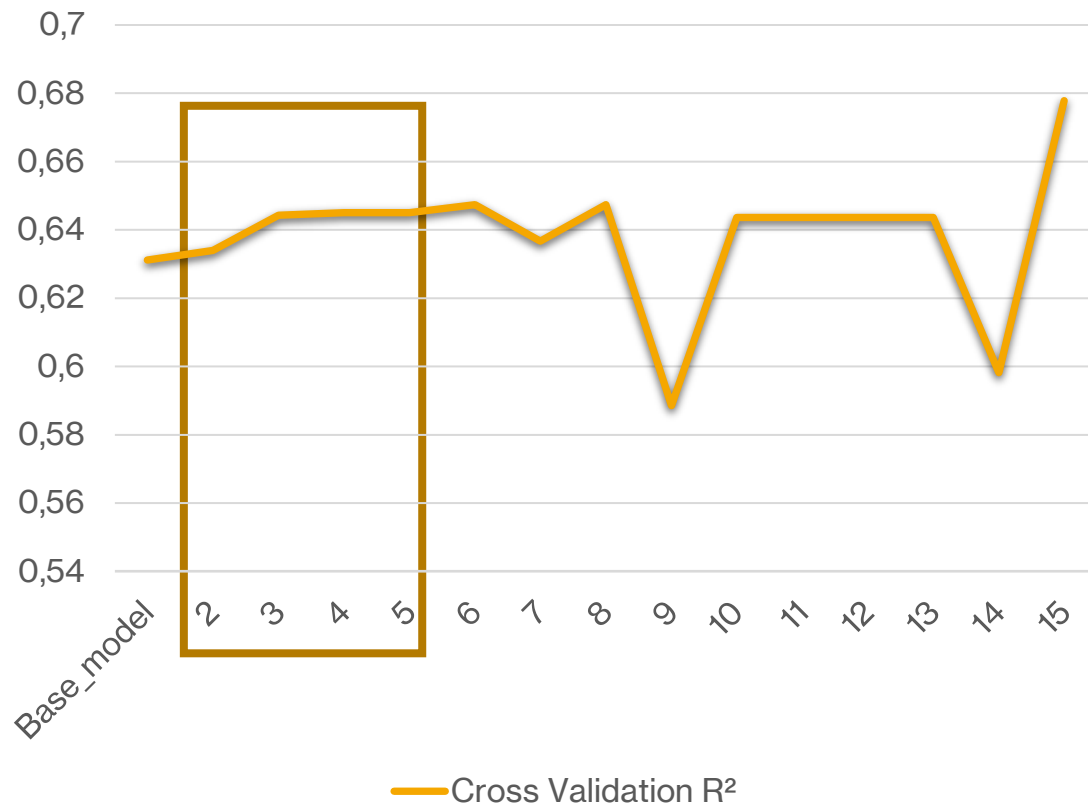
Observation de trois paramètres pour choisir la meilleurs optimisation

## Résultat du modèle pour les différentes itérations



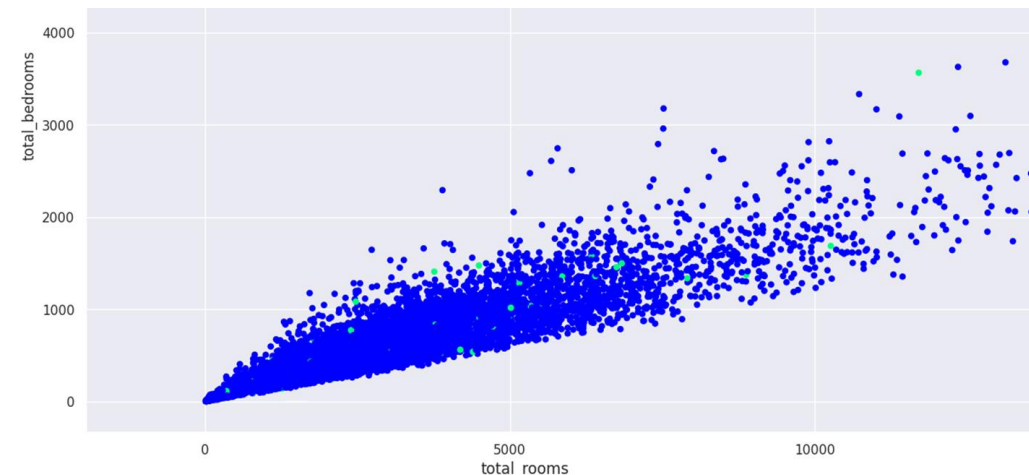
# Les pistes Étudiées

## Résultat du modèle pour les différentes itérations



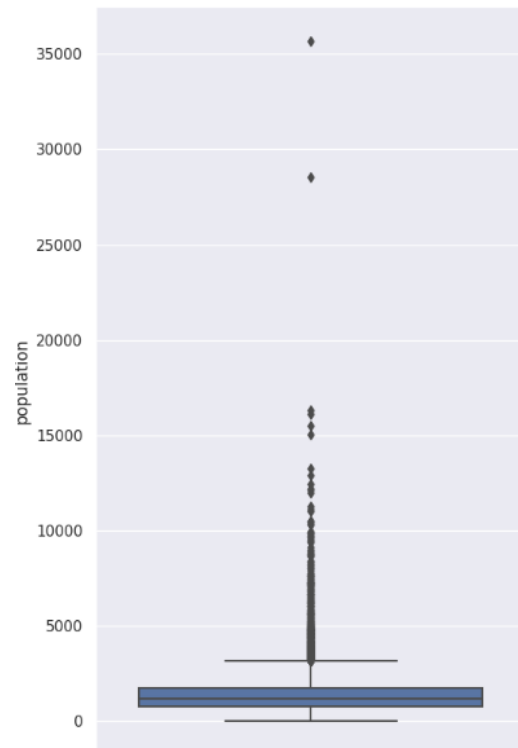
- Transformation de la colonne ocean proximity (2-3)
- imputation des valeurs manquantes (4-5)

Multivariate Imputation by Chained Equation – MICE

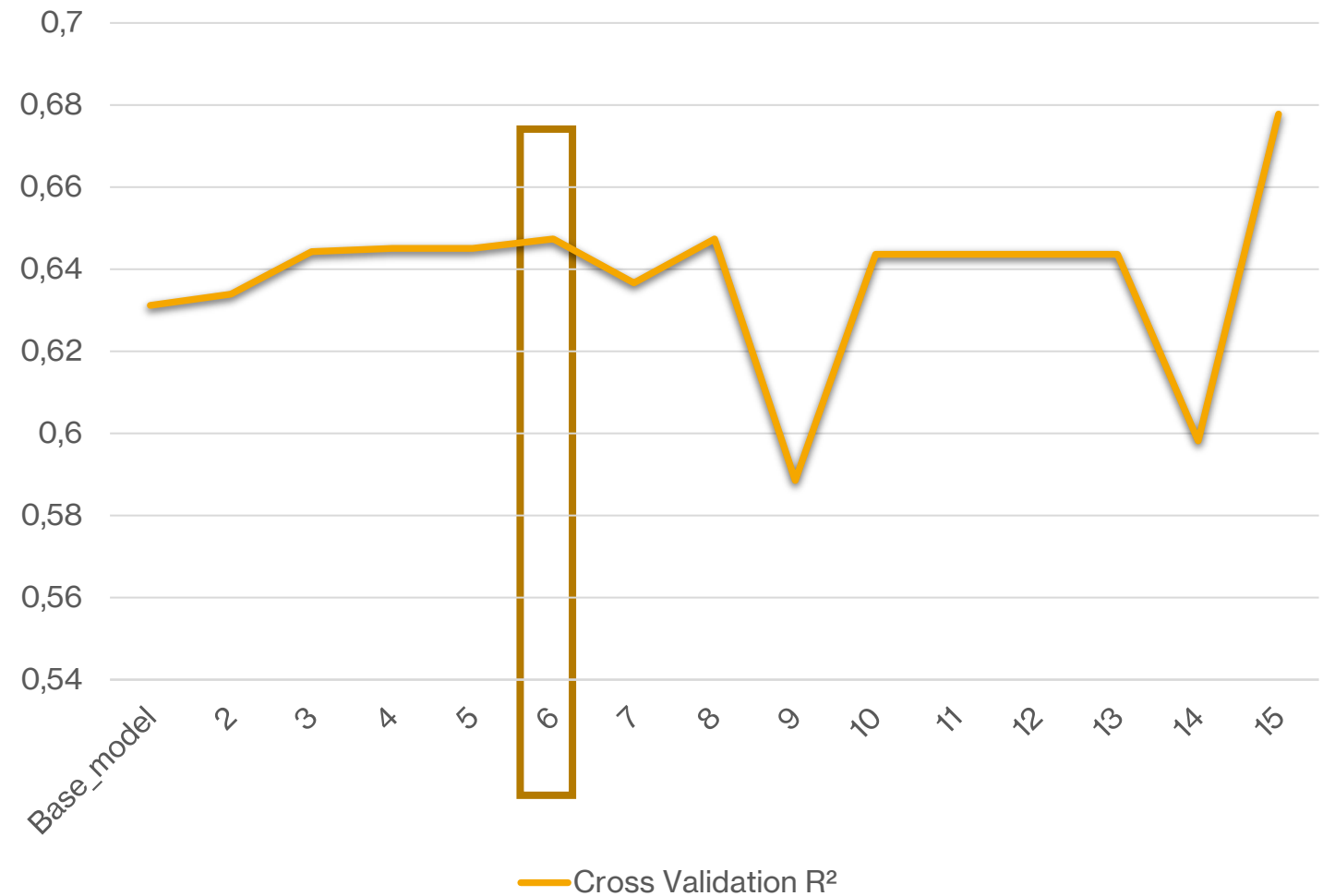


# Les pistes Etudiées

- Etude des outliers (Population )



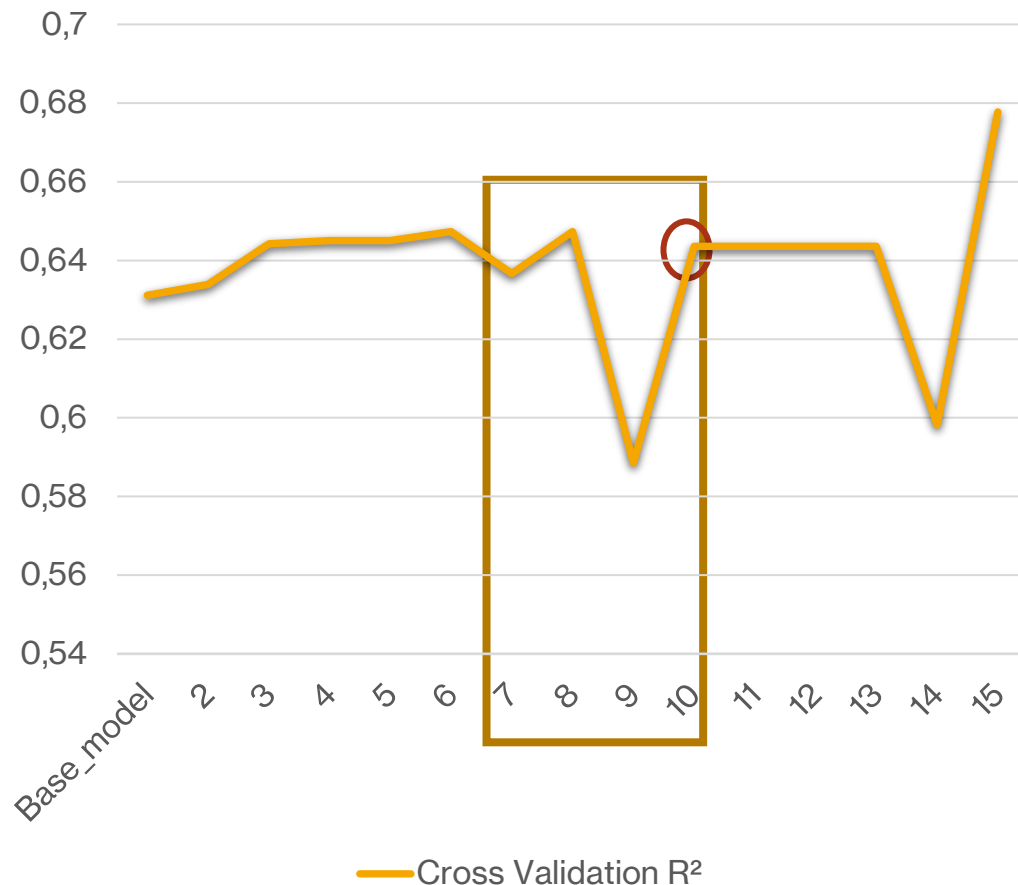
## Résultat du modèle pour les différentes itérations



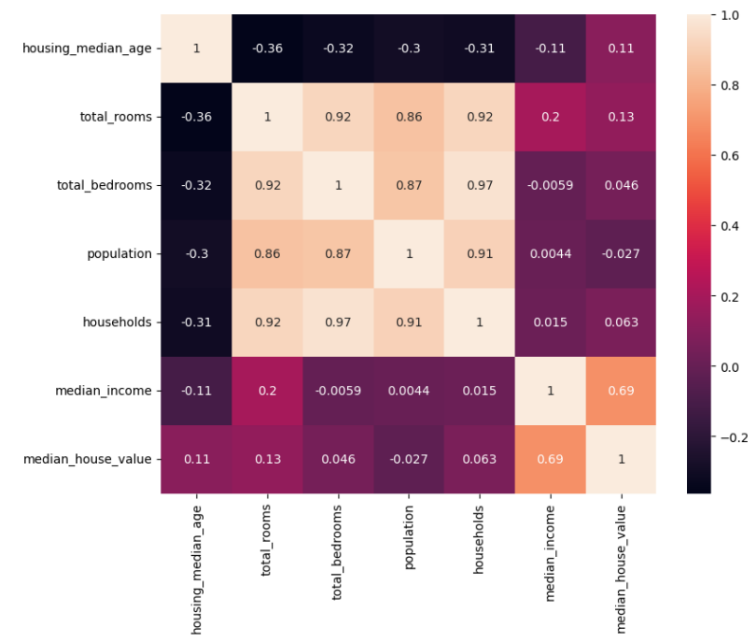


# Les pistes Etudiées

## Résultat du modèle pour les différentes itérations



- Sélection des features:



- Création d'une nouvelle colonne: le nombre de room par ménage

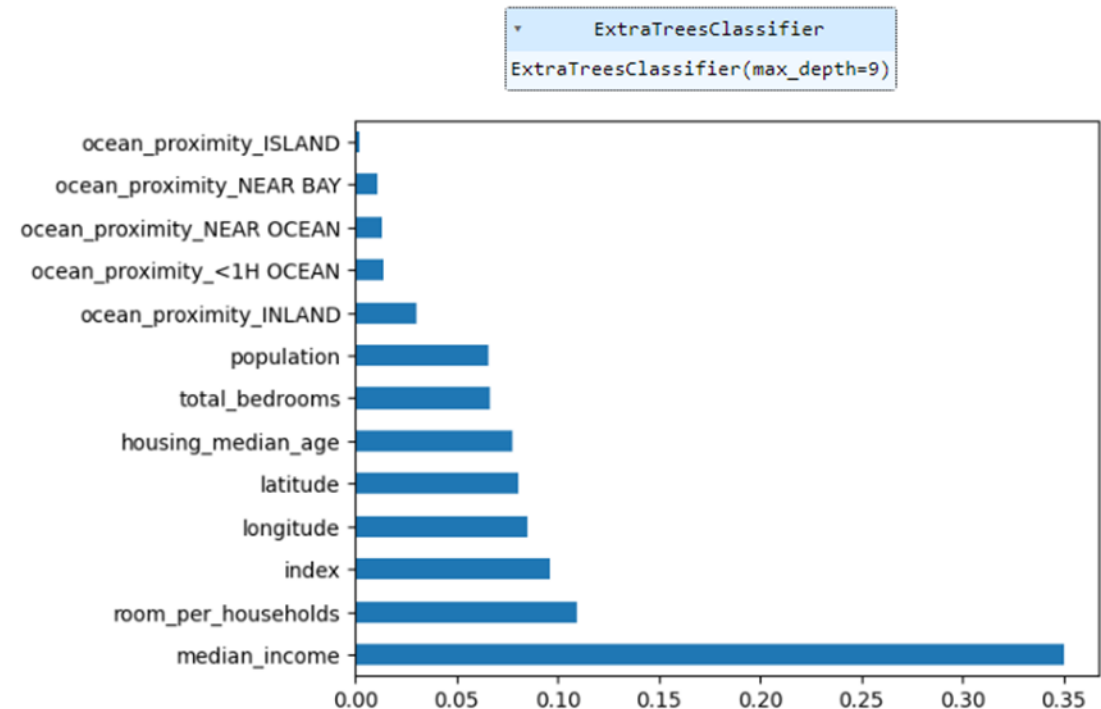
# Les pistes Estudiés

## Résultat du modèle pour les différentes itérations



- Sélection des features:

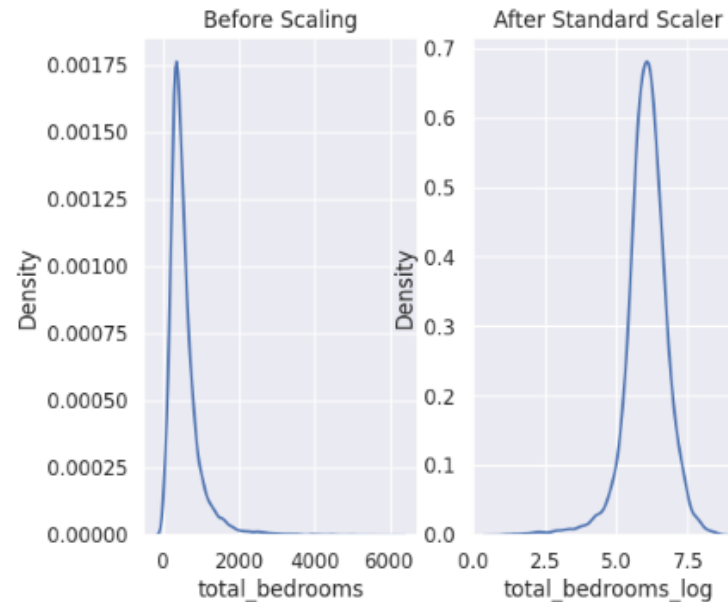
## Feature importance





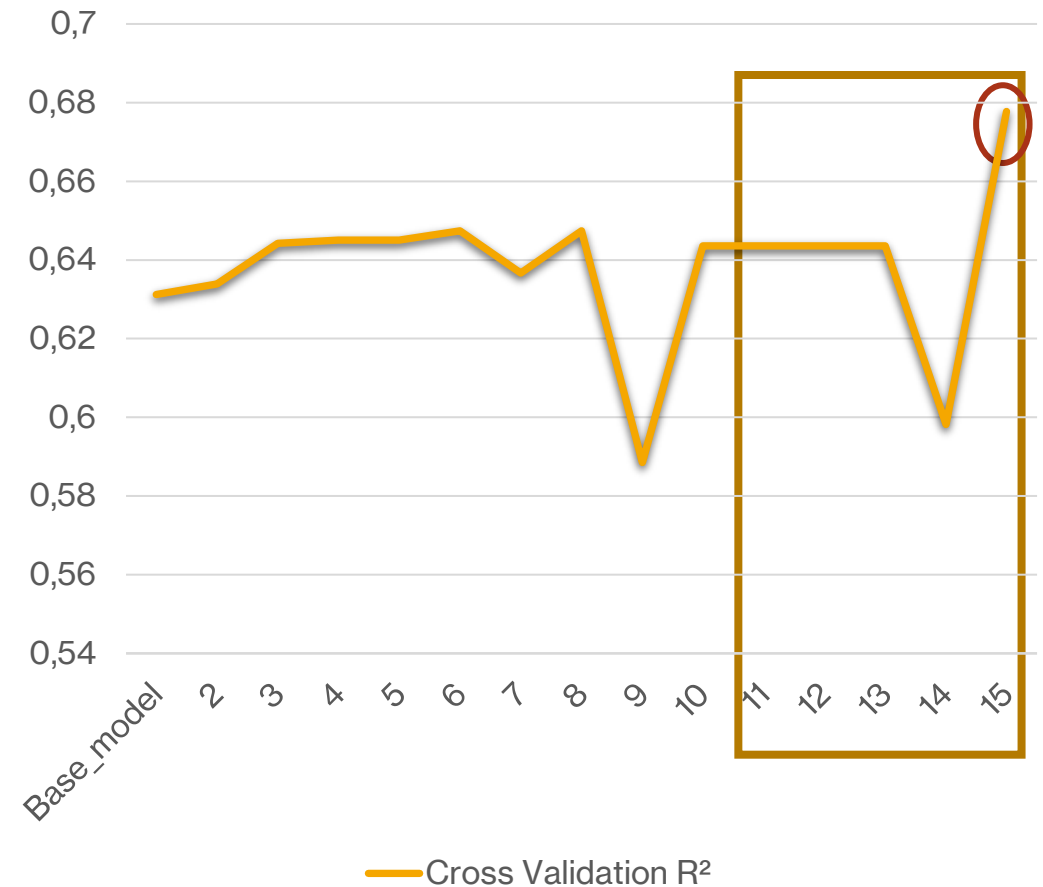
# Les pistes Etudiés

- Scaling des données (différents test):  
Plus efficace = log



A obtenue les meilleur résultats

## Résultat du modèle pour les différentes itérations





# Perspective

---

- Meilleure gestion des outliers
- Amélioration du modèle ( enrichissement du dataframe)
- Test d'autres modèles de régression linéaire (plus adaptés aux outliers...)
- Refactorisation du code
- Utilisation d'une Pipeline