**PREDICTION PRIX DE VENTE MAISONS**

Le but de ce projet est de prédire les prix de vente des maisons à Ames

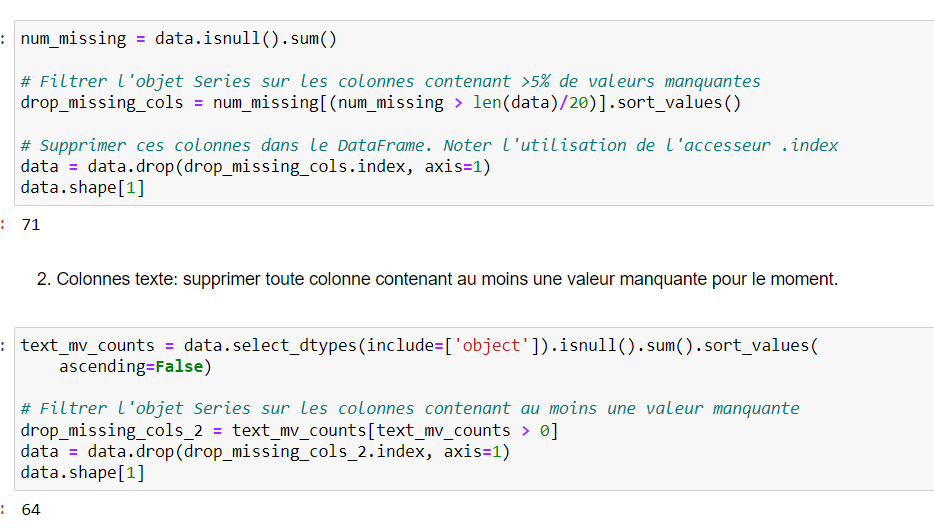
**Le workflow :**

1. Prise en main du jeu de données (Inspection / Mise en place d’hypothèses)



1. Prétraitement (Nettoyage + Normalisation)

Ex : suppression colonne



1. Choix du modèle d’apprentissage / clustering en fonction du problématique

Ex : fonction d’entrainement de modèle de régression linéaire



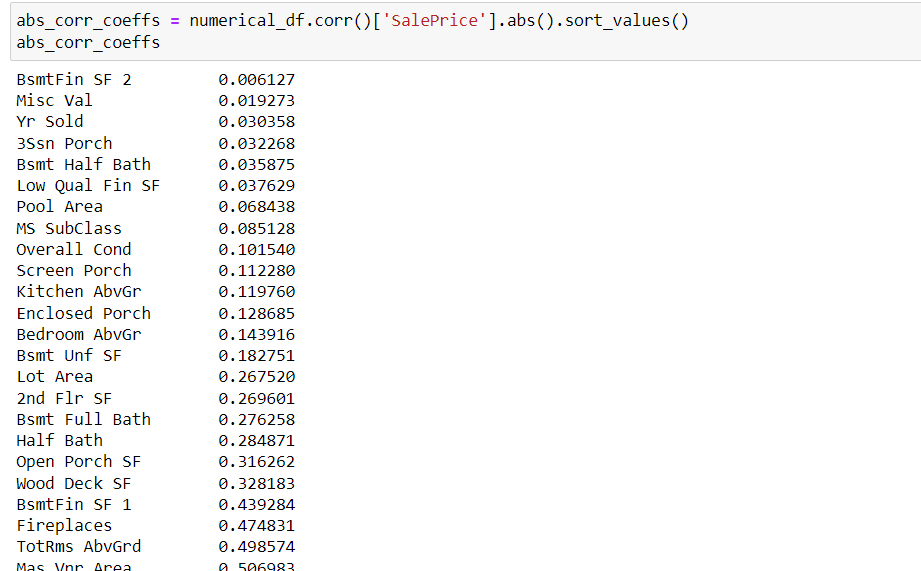
1. Evaluation des résultats

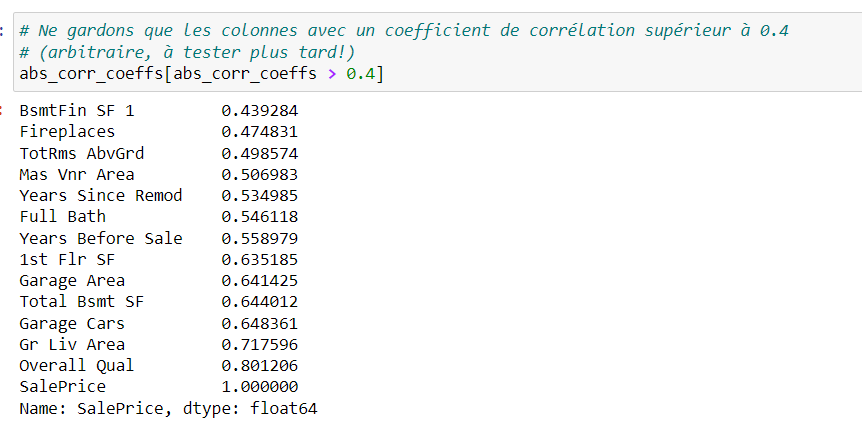
Ex : visualisation de la rmse



1. Optimisation d'hyper paramètres du modèle

Ex : sélection nouveau attributs colonne selon corrélation

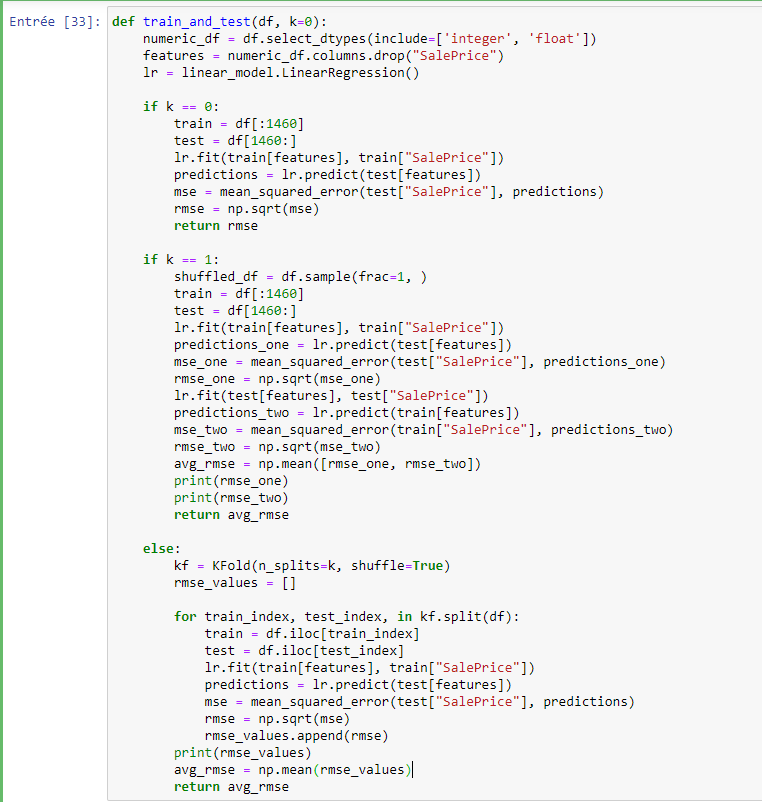




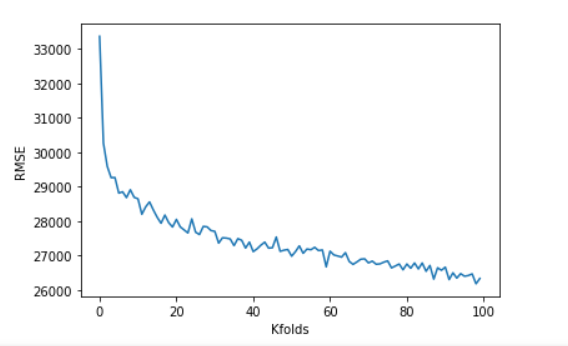
1. Evaluation finale + Interprétation

La courbe de l’erreur RMSE c’est-à-dire l’erreur au carré par rapport au prédiction et les valeurs réelles est en baisse plus le nombre de folds est elevé et stagne à partir de 90, nous pouvons donc nous arrêter à 90 folds.

Ex : ajouts kfolds à la fonction



Ex : graphique rmse et kfolds



**Les outils utilisés**

J’ai utilisé les librairies Python : pandas, numpy , sklearn, matplotlib

**Le dataset**

Le dataset liste les maisons vendu de 2006 à 2010 à Ames.

On y retrouve 82 colonnes (données : ordinal, nominal, discrète et continue) qui contiennent les caractéristiques du bien immobilier telles que : adresse, année de construction, superficie maisons/terrains/garages, chauffage, toiture ….

Nom du dataset : AmesHousing.txt

**L’outil utilisé pour la visualisation**

J’ai utilisé : matplotlib, panda