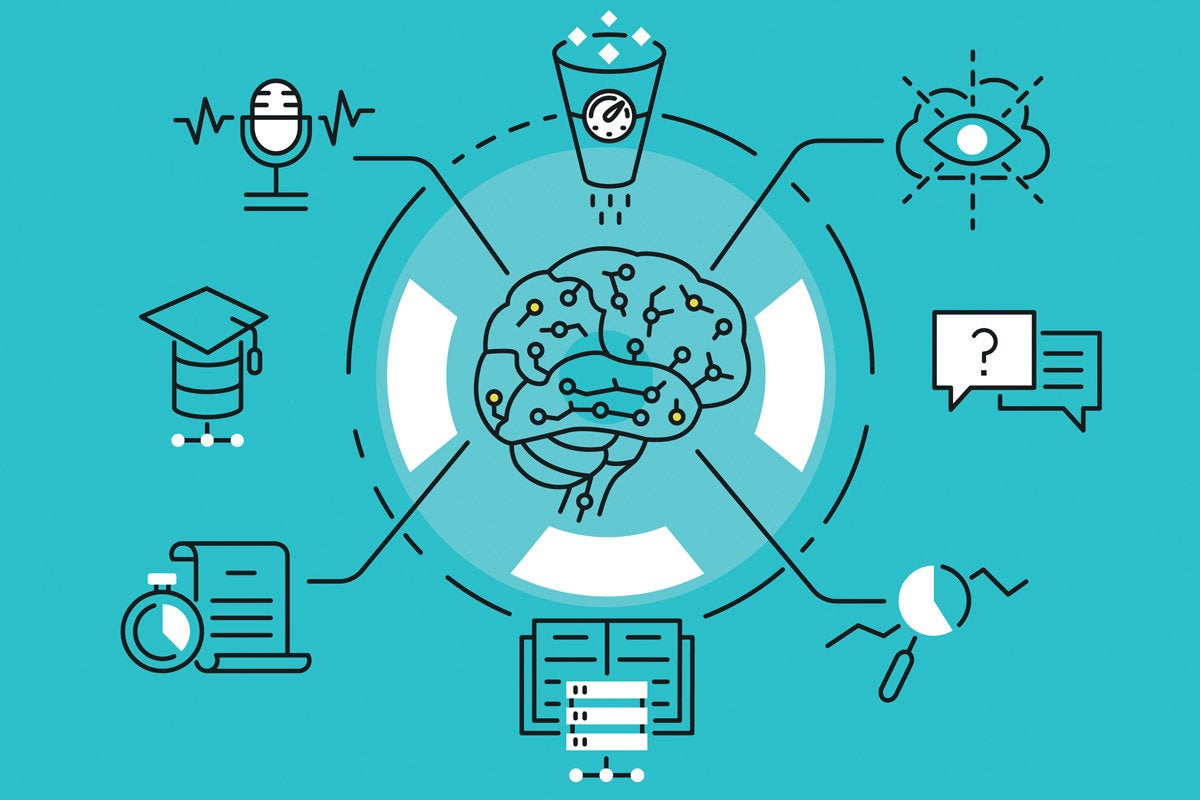
**Université Abdelmalek Essaâdi**

**Faculté des Sciences et Techniques-Tanger**

**Département Génie Informatique**

[**Machine Learning (Master SIBD)**](https://classroom.google.com/c/Njc0NzE2MzQxNjA5)

**Atelier 1 « Régression »**

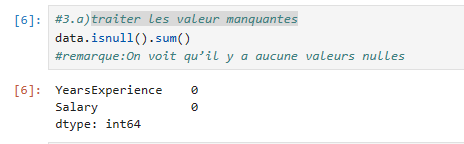
Réalisé Par : Encadre Par :

Yossra safi chetouan Pr . EL AACHAk LOTFI

**Partie 1 : Régression Simple cas Expérience Salaire**

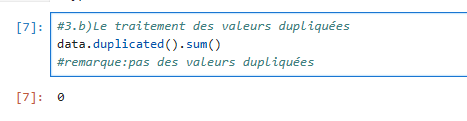
1. Préparation des données

* Traiter les valeurs manquantes



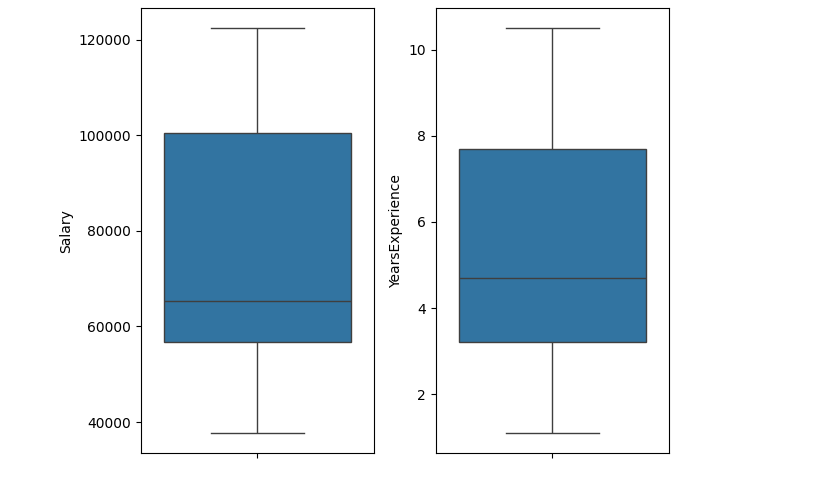
On voit qu'il n'y a pas de valeurs manquantes

* Le traitement des valeurs dupliquées



On voit qu'il n'y a pas de valeurs dupliquées

* Traitement des valeurs aberrantes (Outliers)



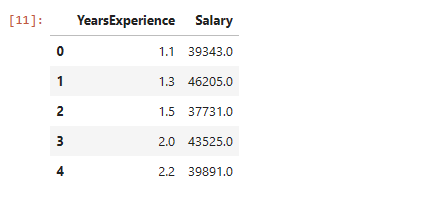
On voit qu'il n'y a pas de valeurs aberrantes

1. Visualisation des données

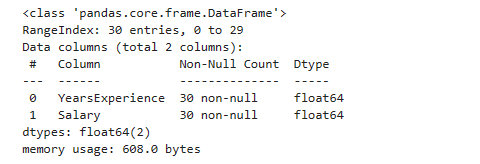
* Afficher les dernières lignes de data set



* Afficher les premières lignes



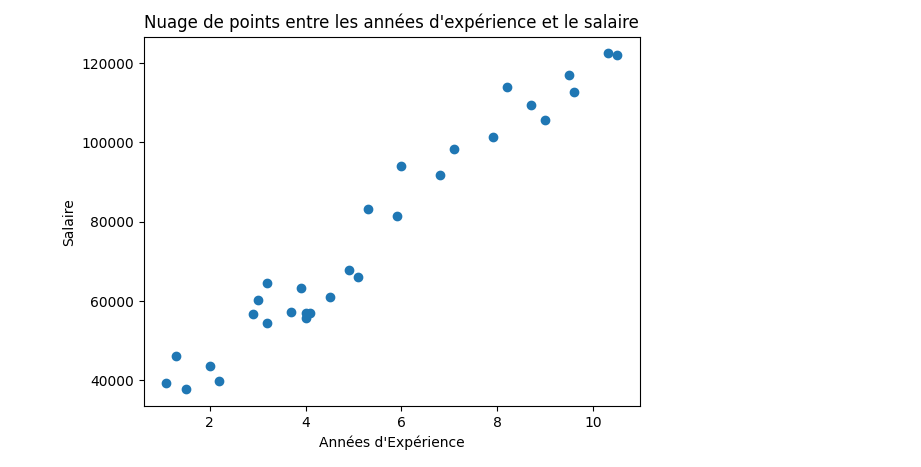
* Afficher les Informations sur les types de données et les valeurs manquantes



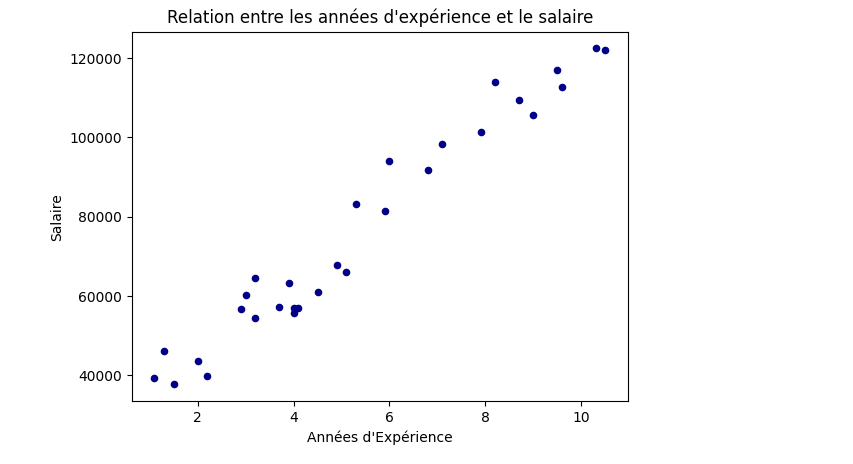
* Statistiques descriptives



* Le nombre d'observations non nulles pour les deux colonnes est de 30.
* La moyenne des salaires est de 76 003.
* La moyenne des années d'expérience est de 5,313.
* La valeur maximale dans la colonne des salaires est de 122 391.
* La valeur minimale dans la colonne des salaires est de 37 731.
* La valeur maximale dans la colonne des années d'expérience est de 10,50.
* La valeur minimale dans la colonne des années d'expérience est de 1,10.
* 25% (premier quartile): Le quartile inférieur, sous lequel se trouvent 25% des données.
* 50% (médiane ou deuxième quartile): La médiane divise les données en deux moitiés égales.
* 75% (troisième quartile): Le quartile supérieur, au-dessus duquel se trouvent 25% des données.
* Afficher la nuage des points matplotlib

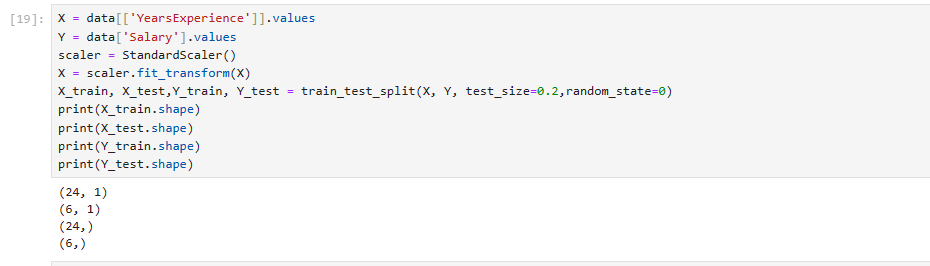


* Afficher la nuage des points Pandas

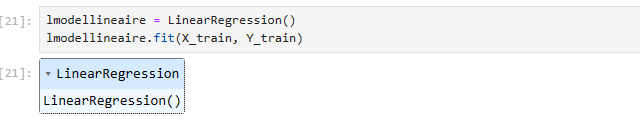


1. Entraînement de modèle

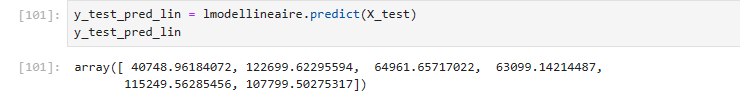
* Division data en 2 data sets : traitement \*08% et test 20%



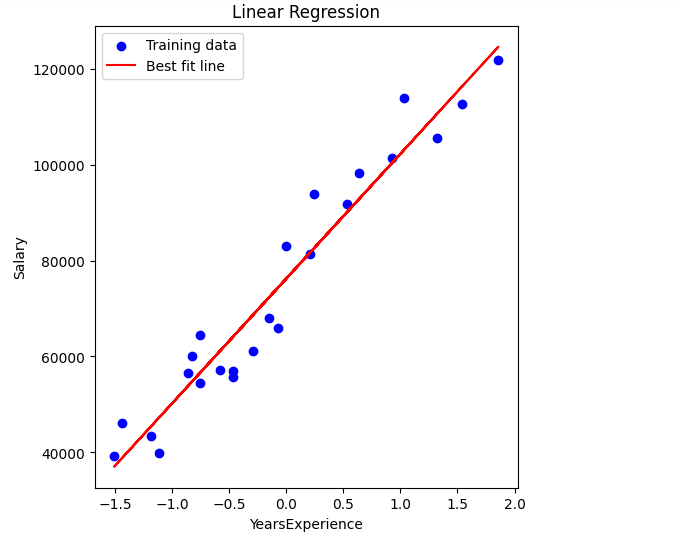
* Entraîner le modèle



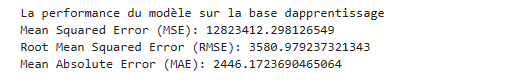
1. Prédiction sur les données de test



1. Visualiser le résultat de la régression sous forme d’un graphe



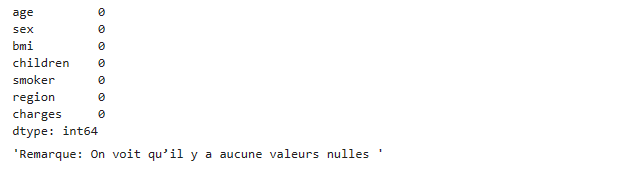
1. Évaluation de modèle



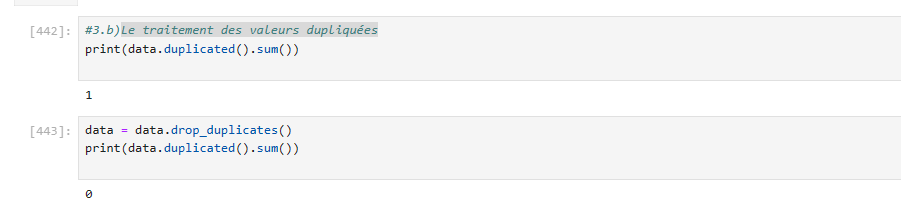
**Partie 2 : Régression multiple cas d’assurance**

1. Préparation des données

* Le traitement les valeurs manquantes

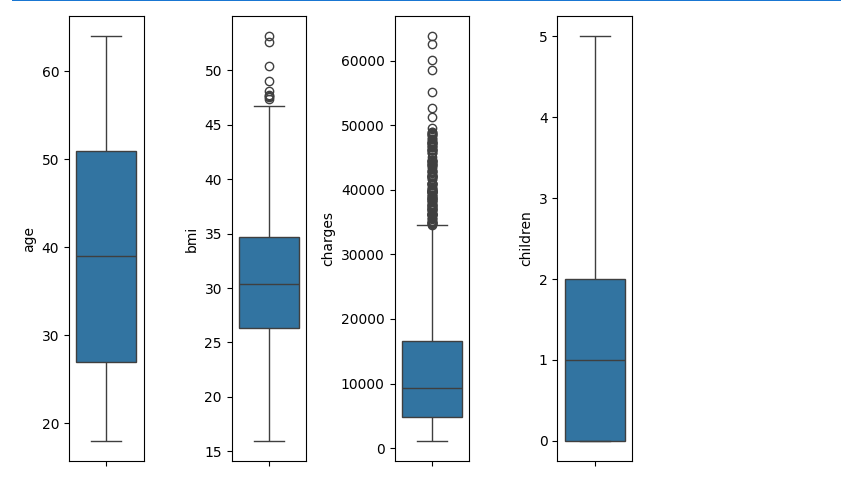


* Le traitement des valeurs dupliquées



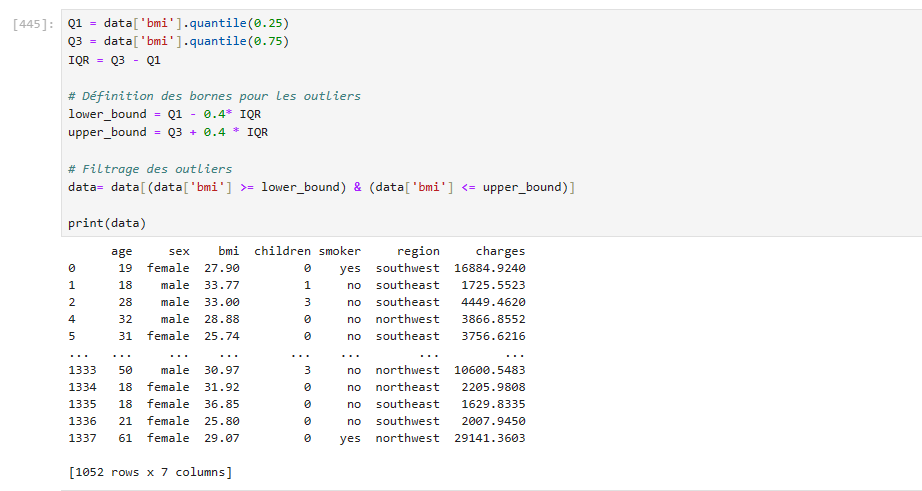
Il y a une seule valeur dupliquée

* Traitement des valeurs aberrantes (Outliers)

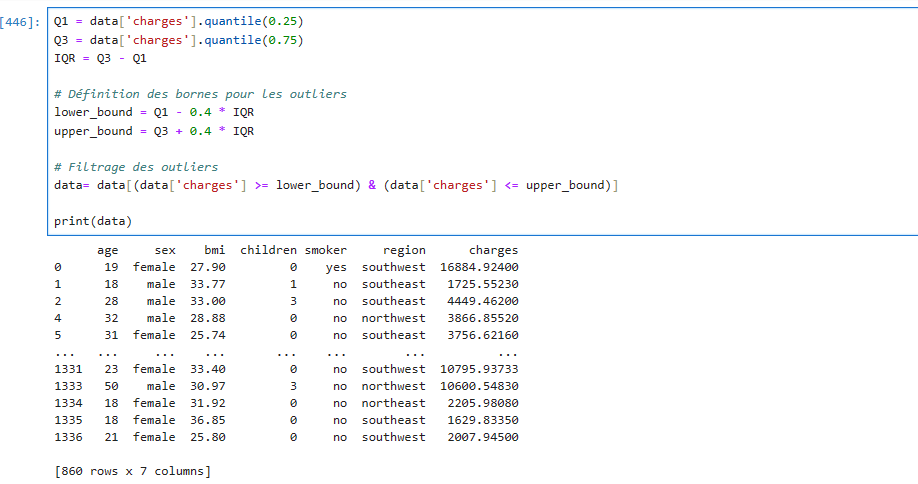


Il y a des valeurs aberrantes (outliers) pour les colonnes BMI et charges

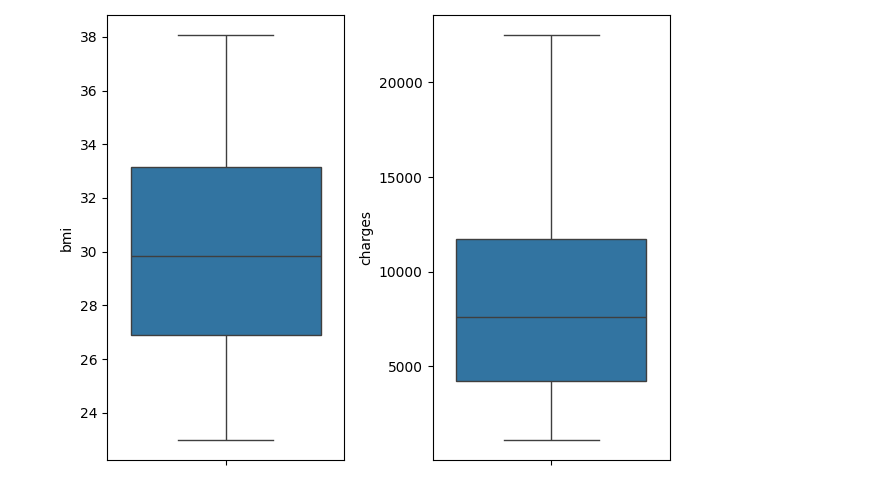
* Traiter valeurs aberrantes pour colonne BMI



* Traiter valeurs aberrantes pour colonne Charges

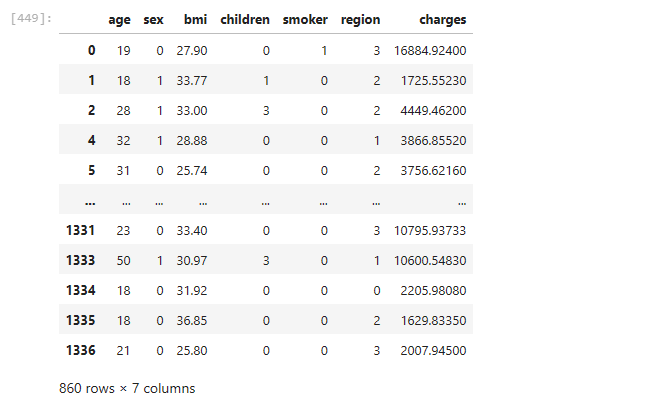


Résultat :



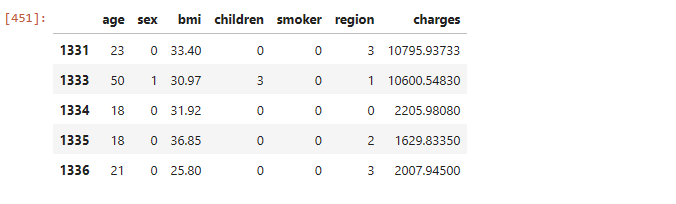
Nous ne voyons maintenant plus de valeurs aberrantes dans les colonnes BMI et charges après traitement

* Encodage des valeurs catégoriques vers numérique

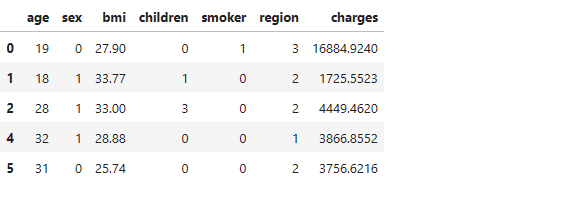


1. Visualisation des données

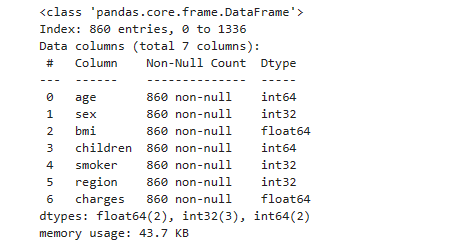
* Afficher les dernières lignes de data set



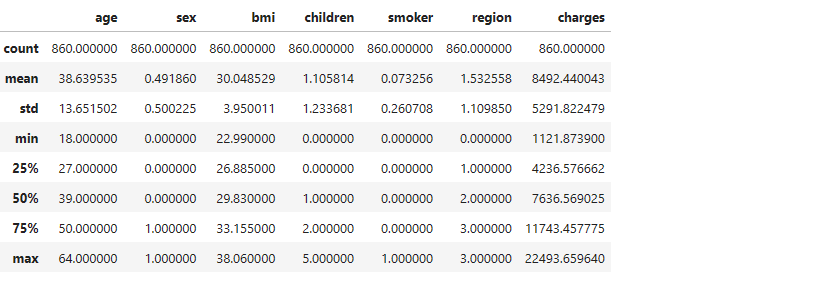
* Afficher les premières lignes



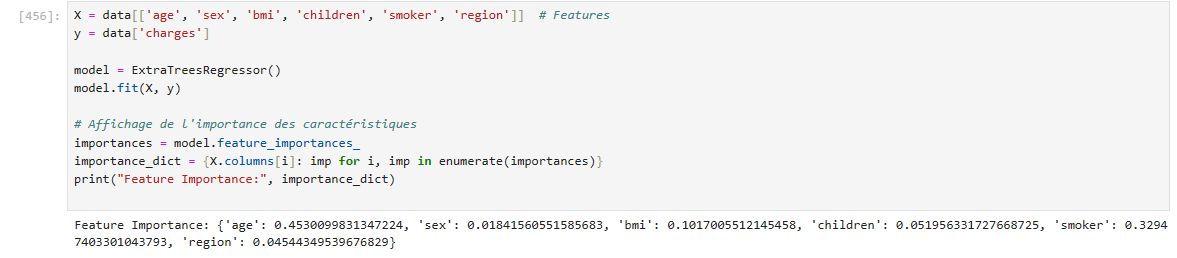
* Afficher les Informations sur les types de données et les valeurs manquantes



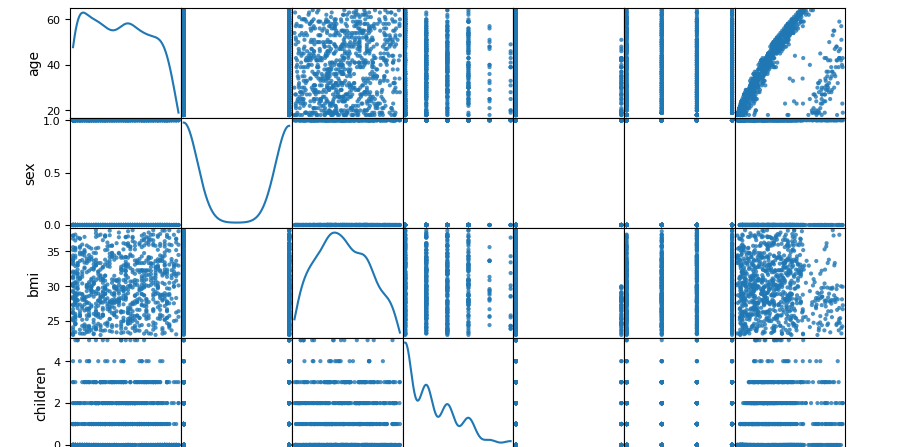
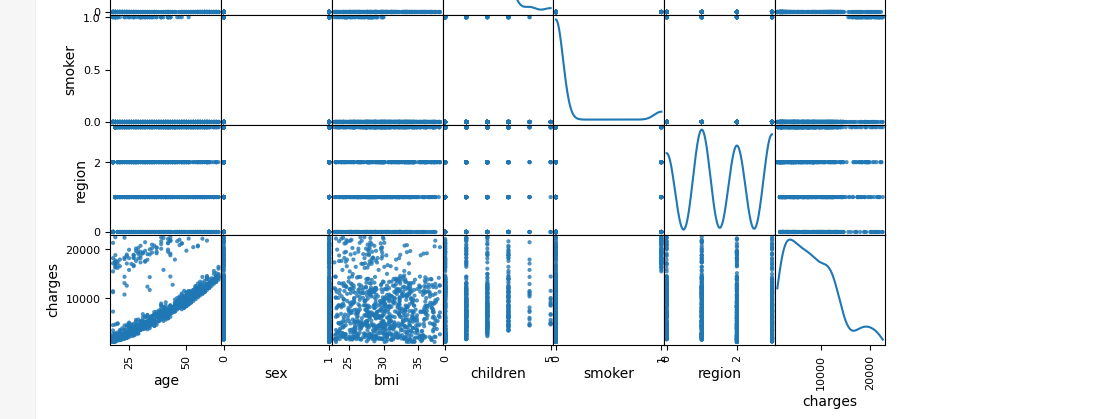
* Statistiques descriptives



* Sélection les caractéristiques importantes en utilisant la méthode d'importance des caractéristiques (FI)

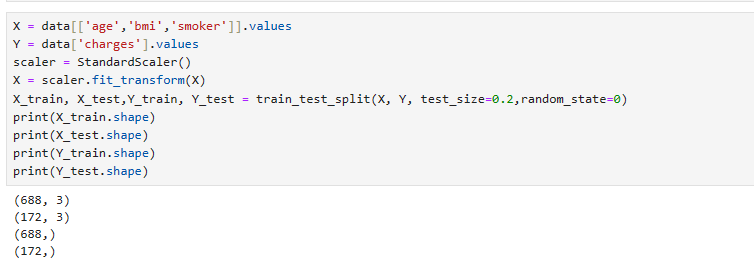


* Afficher la nuage des points

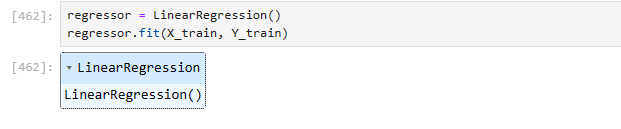


1. Entraînement de modèle

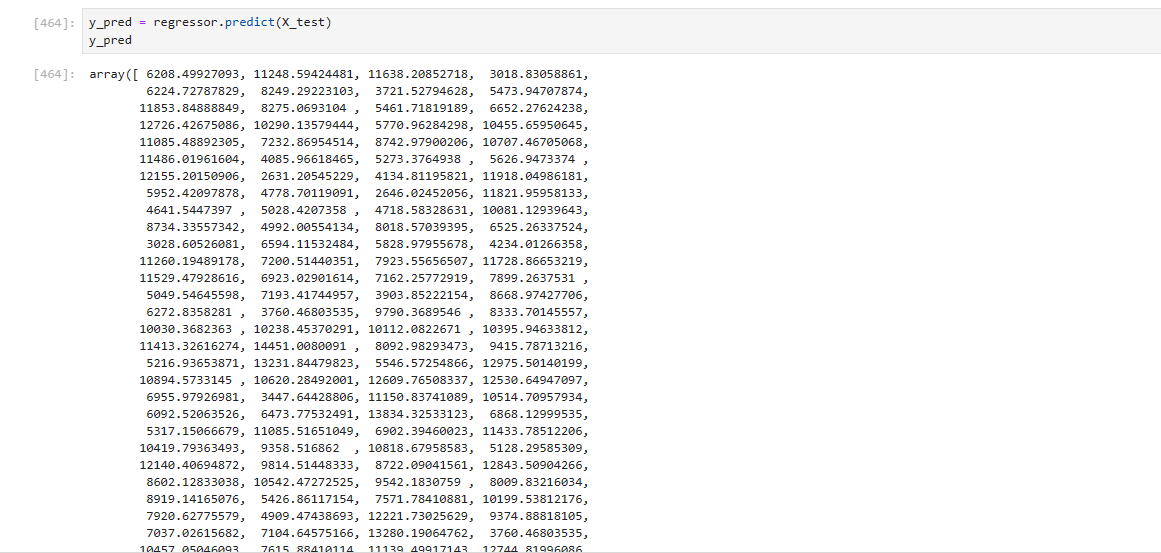
* Division data en 2 data sets : traitement \*08% et test 20%



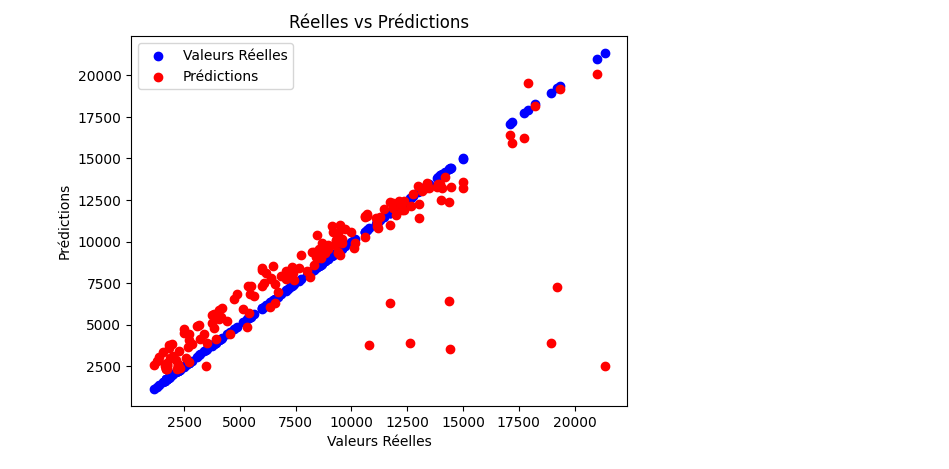
* Entraîner le modèle



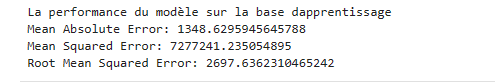
1. Prédiction sur les données de test



1. Visualiser le résultat de la régression sous forme d’un graphe



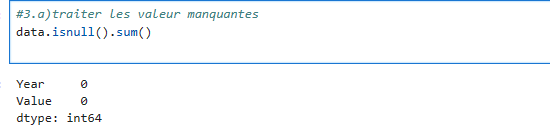
1. Évaluation de modèle

****

**Partie 4 : Régression linière polynomial multiple cas de china GDP**

1. Préparation des données

* Traitement les valeurs manquantes



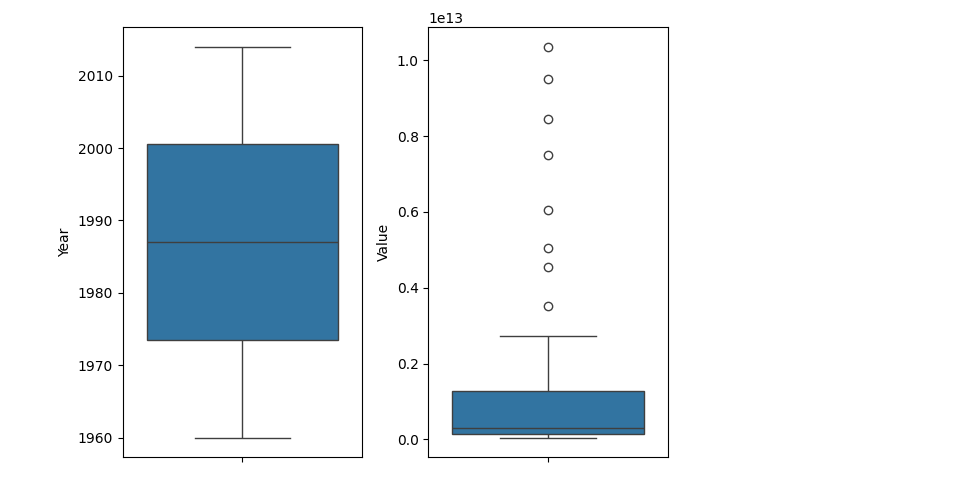
On voit qu'il n'y a pas de valeurs manquantes

* Traitement des valeurs dupliquées



On voit qu'il n'y a pas de valeurs dupliquées

* Traitement des valeurs aberrantes (Outliers)

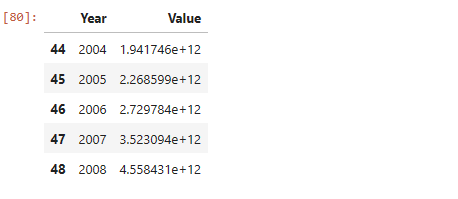


Il y a des valeurs aberrantes dans la colonne Value

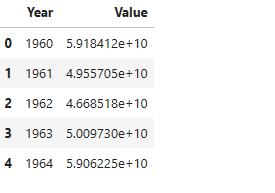


1. Visualisation des données

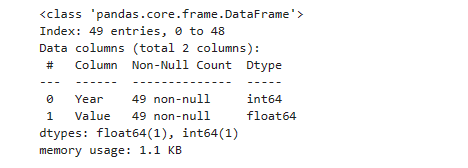
* Afficher les dernières lignes de data set



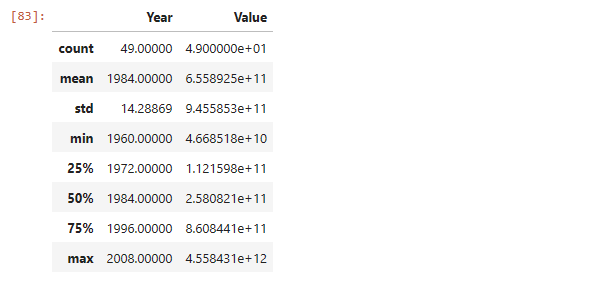
* Afficher les premières lignes



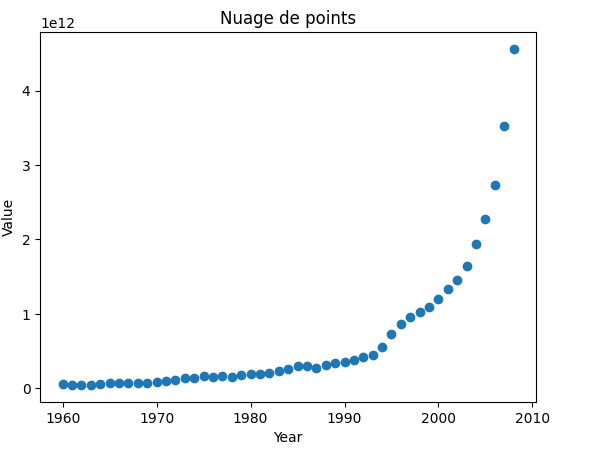
* Afficher les Informations sur les types de données et les valeurs manquantes



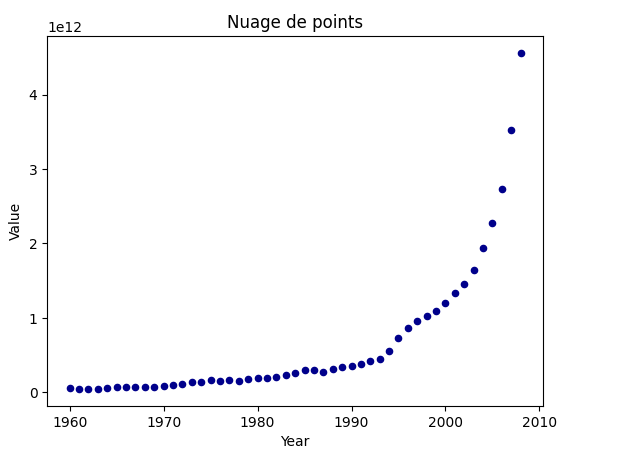
* Statistiques descriptives



* Afficher la nuage des points matplotlib

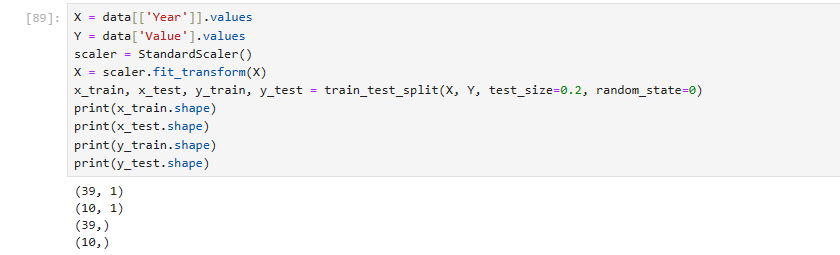


* Afficher la nuage des points Pandas

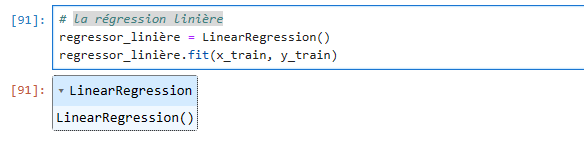


1. Entraînement de modèle

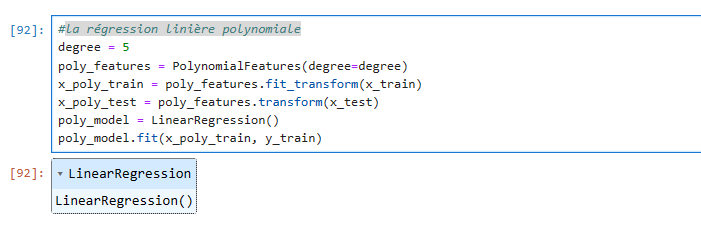
* Division data en 2 data sets : traitement \*08% et test 20%



* Entraîner le modèle
* La régression linière

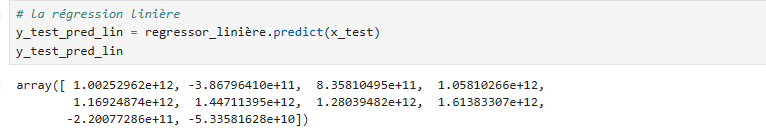


* La régression linière polynomiale

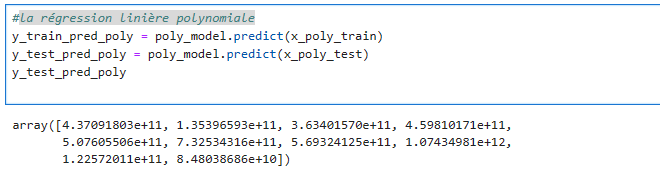


1. Prédiction sur les données de test

* La régression linière

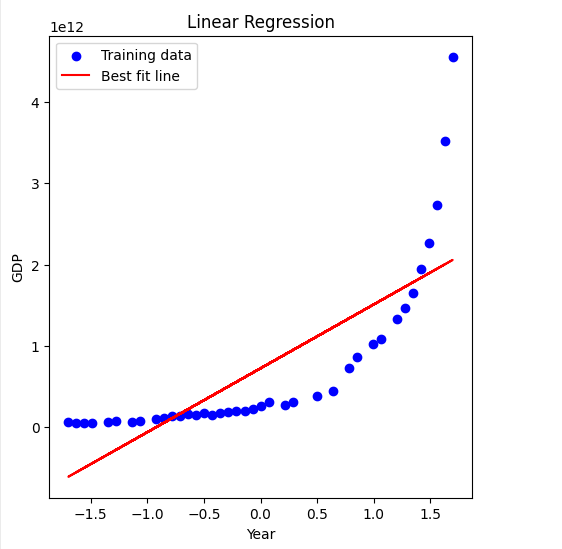


* La régression linière polynomiale

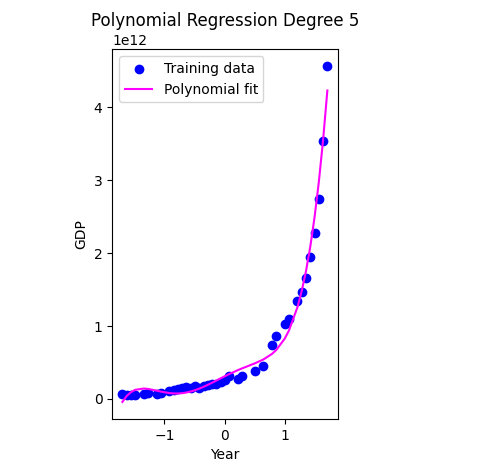


1. Visualiser le résultat de la régression sous forme d’un graphe

* La régression linière

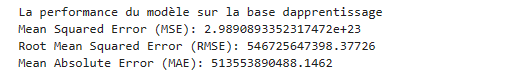


* La régression linière polynomiale

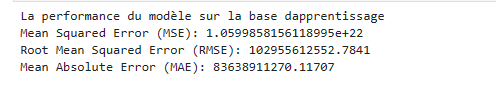


1. Évaluation de modèle

* La régression linière

****

* La régression linière polynomiale

****