**MinMax Scaling on Boston Housing Dataset**

# **Abstract**

This project focuses on applying MinMax Scaling to the Boston Housing Dataset, a widely used dataset for regression tasks. The goal is to normalize numerical features, ensuring comparability and improving model performance. The project implements both Scikit-Learn’s MinMaxScaler and a custom scaling function. Exploratory Data Analysis (EDA) and various visualizations are conducted to analyze the data before and after scaling.

### **Introduction**

Data preprocessing is a crucial step in machine learning workflows. Feature scaling, particularly MinMax Scaling, helps in bringing numerical features to a common scale, preventing bias due to varying feature ranges. This documentation outlines the implementation of MinMax Scaling on the Boston Housing Dataset, including data preprocessing, scaling techniques, and visualization.

## 1. Dataset Overview

The dataset used in this project is the **Boston Housing Dataset**, which contains 506 rows and 15 columns. It includes various features related to housing prices in different Boston neighborhoods.

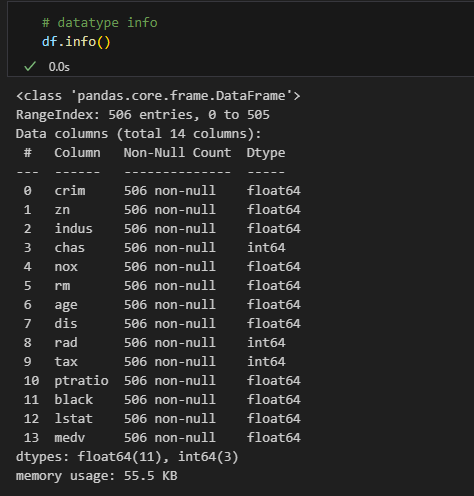
### **Dataset Column Descriptions:**

1. **crim** - Per capita crime rate by town.
2. **zn** - Proportion of residential land zoned for large lots.
3. **indus** - Proportion of non-retail business acres per town.
4. **chas** - Charles River dummy variable (1 if tract bounds river, 0 otherwise).
5. **nox** - Nitrogen oxide concentration (parts per 10 million).
6. **rm** - Average number of rooms per dwelling.
7. **age** - Proportion of owner-occupied units built before 1940.
8. **dis** - Weighted distance to employment centers in Boston.
9. **rad** - Accessibility index to radial highways.
10. **tax** - Property tax rate per $10,000.
11. **ptratio** - Pupil-teacher ratio by town.
12. **black** - Proportion of Black population per town.
13. **lstat** - Percentage of lower status population.
14. **medv** - Median value of owner-occupied homes in $1000s.

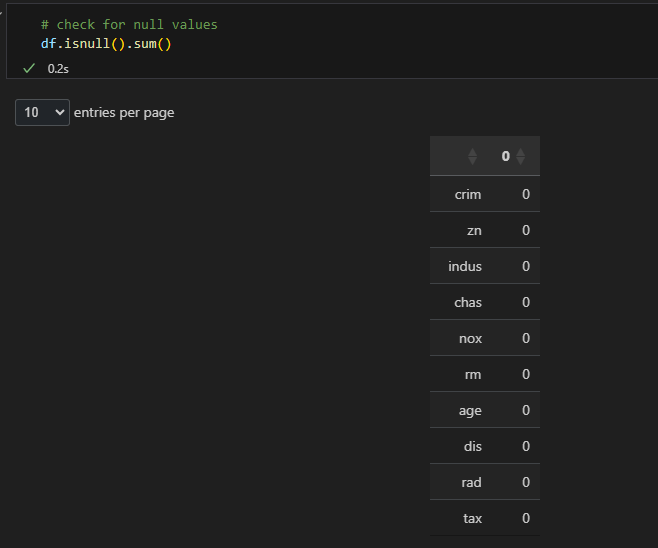
## 2. Exploratory Data Analysis (EDA)

### **Data Pre-processing:**

* **Handling Null Values:** Checked for missing values and handled them using appropriate imputation methods.
* **Correlation Analysis:** Used Pearson correlation to find relationships between numerical features.



“ Information about the dataset”

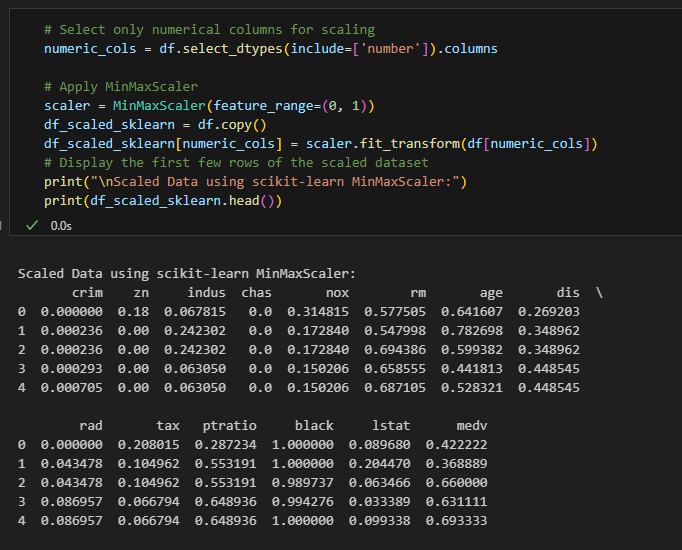


“ The dataset has no missing values”

## 3. MinMax Scaling Implementation

### **Scikit-Learn MinMaxScaler:**

* Applied MinMax scaling to transform the dataset to a range of [0, 1].
* Used MinMaxScaler from sklearn.preprocessing to normalize the numerical columns.

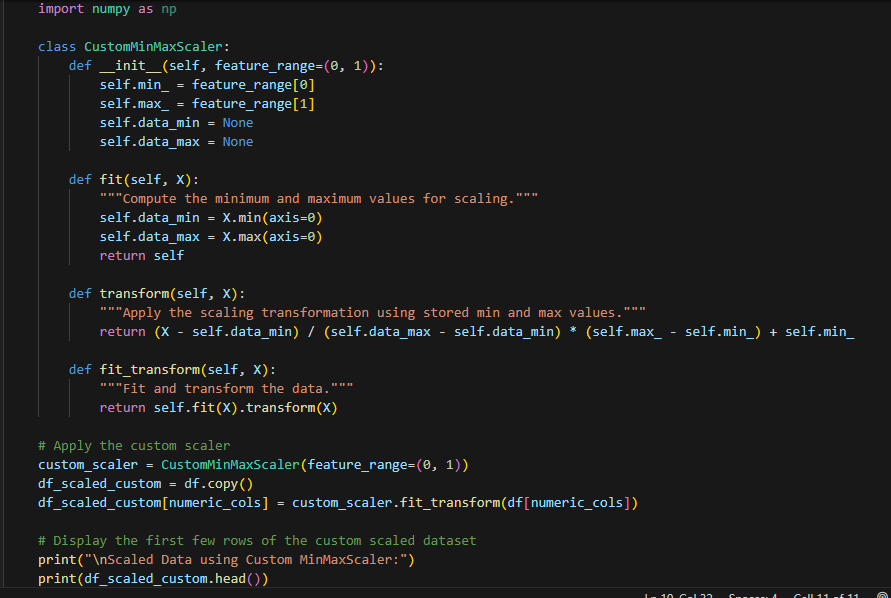


### **Custom MinMaxScaler:**

* Implemented a custom MinMax scaling function using:

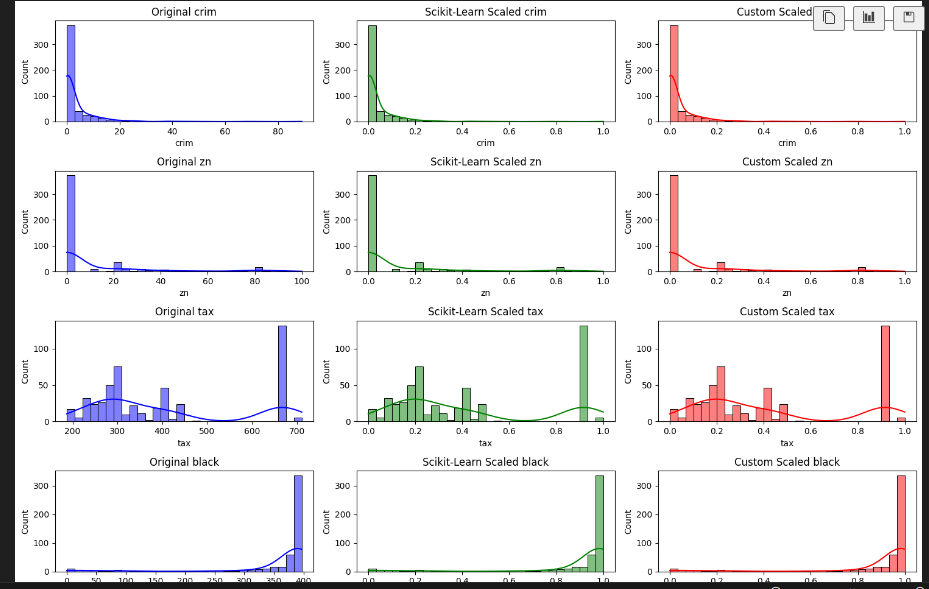


* Compared results with the scikit-learn implementation.



## 4. Data Visualization

* **Histogram**: Displayed distributions of original and scaled data.
* **Scatter Plot**: Visualized relationships between features.



## 5. Overall Summary

* Conducted data pre-processing by handling missing values and outliers.
* Applied MinMax scaling using both Scikit-Learn and a custom implementation.
* Visualized data distributions before and after scaling.
* Compared original and scaled data to understand the impact of normalization.

## 6. References

## Dataset: [Boston Dataset](https://github.com/anuselva1905/handle_missing_values/blob/main/MinMax_Scaler/Boston%20Dataset.csv)

## Github Repository: [Repository link](https://github.com/anuselva1905/handle_missing_values/blob/main/MinMax_Scaler/minmax_scaler.ipynb)