

# Summary

In this project, I implemented **K-Nearest Neighbors (KNN) Regressor** and **Multiple Linear Regression** from scratch to predict the number of **fireplaces** in a house using real estate data. The goal is to assist renovation businesses in identifying homes where adding or upgrading fireplaces could be a worthwhile investment. The dataset includes housing features like price, area, bedrooms, etc. Models were trained and tested using Python in Jupyter Notebook without machine learning libraries like scikit-learn. Evaluation was done using metrics such as **MAE** and **R<sup>2</sup> Score**, and various plots were used for analysis and visualization. While results show moderate accuracy, future improvements may include classification methods or deep learning for better performance.

## Dataset:

Real estate dataset with 3166 entries, including sold price, bedrooms, area, bathrooms, garage, and target: **fireplaces** (0–5 range).

## Use Case:

Predict the number of fireplaces in a home to **support renovation and investment planning** for fireplace companies.

## Algorithms Used:

- K-Nearest Neighbors (KNN) Regressor
- Multiple Linear Regression (Both implemented manually)

## Results:

- KNN gave better predictions with a lower MAE and higher R<sup>2</sup>.
- Linear regression had a few negative values, later rounded/clipped.

## Challenges:

- Target variable was **discrete**, but treated as regression.
- Data skewed toward 1–2 fireplaces.
- Scaling and feature selection were crucial for result

## Evaluation Metrics:

- MAE (Mean Absolute Error)
- R<sup>2</sup> Score (coefficient of determination)
- Bar plots used to visually compare both models.

## Visualization:

- Fireplace count distribution (bar graph)
- Prediction vs actual values
- Model performance comparison (MAE & R<sup>2</sup> chart)

## Limitations:

- Regression on discrete targets leads to rounding issues.
- Negative predictions from linear models require clipping.
- Not ideal for high class imbalance

## Future Work:

- Try classification instead of regression.
- Use ensemble methods or deep learning.
- Include more features like **kitchen\_features**, **HOA**, etc.