## 1. What did you implement?

1. i implemented Linear Regression (LR) and Random Forest Regressor (RF) to predict house prices
2. I loaded a cleaned dataset
3. The dataset was split into 80% training and 20% testing
4. I trained both models on the training data
5. I evaluated the models using standard metrics: $R^2$, MAE, MSE, RMSE

## 2. Comparison of Models (Sanity Check)

waxan ku tijaabiyeye modelka 4 sample rows si aan u arko kan ugu dow saadaasha/sadaalinta natiijaduna waa tan:

### example1: size_sqft single-row sanity check:

Actual size sqft: 812,100

LR Pred: 656,755

RF Pred: 789,031

### example2: price single-row sanity check:

Actual price: $806,000

LR Pred: $609,615

RF Pred: $538,756

### example3: HouseAge single-row sanity check

actual HouseAge: 654200.0

LR prded : 718486.321659815

RF pred : 777862.0

example4: size_per_bedroom single-row sanity check

actual size_per_bedroom:  366000.0

LR prded : 299971.0381155011

RF pred : 311571.0

example1: Random Forest u dhow

example2: Linear Regression u dhow

example3: Both models overshot, lakiin Random Forest u dhow

example4: Random Forest u dhow

## 3: Understanding Random Forest:

Random Forest is an ensemble machine learning technique that combines multiple decision trees to produce a more accurate and robust prediction for both classification and regression tasks

How does it work?

Each tree makes a prediction, and the final result is the average (for regression) of all tree outputs, this reduces overfitting and increases accuracy compared to a single decision tree

Random Forest had slightly better $R^2$ (explained more variance in house prices)

Random Forest also had lower error metrics (MAE, RMSE), meaning its predictions were more accurate and stable

Linear Regression was simpler but less flexible, struggling with non-linear relationships

## 4: Metrics Discussion

$R^2$ (Coefficient of Determination):

RF got a higher $R^2$ (0.859 vs 0.848)

This means that RF explained the changes in house prices better than LR

Errors (MAE & RMSE):

RF has a smaller error (both MAE and RMSE).

This means that the average prediction error is less than LR

## Strengths & Weaknesses

Linear Regression (LR):

1. Strengths: Simple, fast, works well on data with linear relationships
2. Weaknesses: Struggles with data with complex or non-linear relationships

Random Forest (RF):

1. Strengths: Works well on complex, non-linear data, and provides accurate and stable predictions
2. Weaknesses: Can be computationally expensive when the dataset is very large, and the results cannot be easily explained like LR

# Your Findings:

## Summarize in 1–2 paragraphs which model you prefer for price prediction models and why?

After training and testing both models, I found that Random Forest performed better than Linear Regression for house price prediction

Random Forest achieved a slightly higher $R^2$ score (0.859 vs 0.848) and lower error values (MAE and RMSE), which means it explained the variation in housing prices more effectively and made more accurate predictions the sanity check examples also showed that Random Forest's predictions were often closer to the actual values

I would prefer to use Random Forest for house price prediction While Linear Regression is simpler and easier to interpret, it struggles when the relationships between features and prices are non-linear, Random Forest, on the other hand, is more flexible, handles complex patterns well, and consistently provided stronger results with this dataset.

By: Mohamed Mohamud