1. Linear Regression
Linear Regression is a supervised machine learning algorithm used to predict a continuous target variable based on one or more input features. It assumes a linear relationship between the independent (input) variables and the dependent (output) variable. In this project, we use linear regression to predict house prices (`medv`) using selected features.

---

2. Data Exploration and Visualization
Before modeling, understanding the data through exploration and visualizations helps identify patterns, relationships, and anomalies. Histograms and scatter plots are used to analyze the distribution of the target variable and its relationships with input features. Heatmaps show correlation between all features.

---

3. Feature Selection
Feature selection is the process of choosing relevant input variables that contribute the most to the prediction task. Using highly correlated features helps in building better models. In this project, `rm` (number of rooms) and `lstat` (lower status population percentage) were selected for prediction based on their high correlation with the target.

---

4. Train-Test Split
This step involves splitting the dataset into training and testing sets. Typically, 80% of the data is used for training and 20% for testing. It helps evaluate the performance of the model on unseen data, which is critical to avoid overfitting.

---

5. Model Evaluation – Mean Squared Error (MSE)
MSE measures the average squared difference between the predicted and actual values. Lower values indicate a better fit of the model. It's a common metric for regression tasks, where the goal is to minimize the prediction error.

🔁 Algorithm in Short

1. **Import libraries and read CSV file**
   → **Load dataset and check the first few records.**

2. **Check for missing values**
   → Identify if any feature has missing entries.

3. **Plot histogram for `medv`**
   → Visualize the distribution of the target variable.

4. **Generate correlation heatmap**
   → Identify which features are strongly related to `medv`.

5. **Visualize relationships**
   → Scatter plots of `rm` and `lstat` against `medv`.

6. **Split dataset**
   → Create training and testing datasets.

7. **Train linear regression model**
   → Fit the model using selected features.

8. **Predict and evaluate**
   → Predict values on test set and compute MSE.

📌 Conclusion

This project demonstrates a basic yet effective machine learning pipeline for predicting housing prices using linear regression. We began with data exploration, selected meaningful features based on correlation, and trained a regression model. The model performance was evaluated using Mean Squared Error, and visualizations helped understand both the data and model predictions. The choice of features (`rm` and `lstat`) aligns well with domain knowledge, confirming their strong predictive power in the Boston housing dataset.