

**Lahore College for Women University, Lahore**

**Department: Software Engineering**

**Submitted By: Iram Ahmad**

**Roll no: 2225165111**

**Submitted to: Sir Mohsin Nazir**

**Semester: VII**

**Session: 2022-2026**

**Assignment no 5**

**Course: Applied Data Science with AI**

**Semester:** BSSE 7th  
**Week #:** 5  
**Student Name:** Iram Ahmad  
**Roll Number:** 2225165111  
**Project Title:** House Price Prediction  
**GitHub Link:** https://github.com/Iram-Ahmad/Data-Science-AI-Course

### ****1. Reading Summary (½–1 page)****

**Reading Material for this Week:**

* Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow – Regression Concepts
* Scikit-Learn Documentation – Linear Regression
* Khan Academy / Medium Articles on Regression Metrics (MAE, RMSE)

**Key Learnings:**

* **Point 1:** Learned the concept of **Supervised Learning** and how regression models predict continuous values.
* **Point 2:** Understood how to implement **Linear Regression** using Scikit-Learn and interpret coefficients.
* **Point 3:** Learned evaluation metrics such as **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** to measure model accuracy.

**Reflection:**  
These readings helped me understand how to build a **baseline predictive model** for the House Price dataset.  
By applying regression techniques, I can now relate important numerical features (like OverallQual, GrLivArea, and GarageCars) to the target variable SalePrice.  
This connects directly to our project milestone of building the **first regression model**.

### ****2. Classroom Task Documentation****

**Task Performed in Class:**  
Implemented a **Linear Regression model** using Scikit-Learn.  
Split the dataset into **training and testing sets**, trained the model, and evaluated it using **MAE and RMSE**.  
Visualized the actual vs. predicted house prices using a scatter plot.

**Screenshots / Code Snippets:**

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error

import numpy as np

X = df[['OverallQual', 'GrLivArea', 'GarageCars']]

y = df['SalePrice']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

model = LinearRegression()

model.fit(X\_train, y\_train)

y\_pred = model.predict(X\_test)

mae = mean\_absolute\_error(y\_test, y\_pred)

rmse = np.sqrt(mean\_squared\_error(y\_test, y\_pred))

**Sample Output (example):**

MAE: 23000.45

RMSE: 34000.89

### ****3. Weekly Assignment Submission****

**Assignment Title:**  
“Supervised Learning – Regression”

**Steps Taken:**

1. Loaded the cleaned dataset from previous weeks.
2. Selected top correlated features from Week 4 (OverallQual, GrLivArea, GarageCars).
3. Split data into train/test sets and trained a **Linear Regression** model.
4. Evaluated the model using **MAE** and **RMSE** metrics.
5. Visualized actual vs. predicted prices with a scatter plot.

**Output:**

* The scatter plot shows a strong positive linear trend between predicted and actual prices.
* The baseline model performed well with reasonable error values.

**Challenges Faced:**

* Initially faced file path error while loading the dataset (resolved by using absolute path).
* Needed clarification on interpreting correlation and coefficients.

**GitHub Link:**  
https://github.com/Iram-Ahmad/Data-Science-AI-Course

### ****4. Project Progress Milestone****

**This Week’s Milestone:**  
✅ Built and tested the **baseline Linear Regression model** on selected features.

**Next Week’s Goal:**  
➡ Improve model accuracy using **advanced regression techniques** (e.g., Ridge, Lasso, or Decision Trees).

### ****5. Self-Evaluation****

☑ **I completed all tasks on time.**  
⬜ I partially completed the tasks.  
⬜ I struggled with this week’s tasks and need help.

### ****6. Questions for Instructor (Optional)****

* How can we decide whether to use polynomial regression or regularization methods for better accuracy?
* What are the best practices for feature scaling before applying regression models?