Linear Regression Analysis Report

Student Name: Daniel Muthama

Program: **M**achine Learning Regression

Course: DA1-2025

Date: July 1, 2025

1. Introduction

This report details the implementation of two simple linear regression models to predict:

Exam Scores based on study hours

House Prices based on property area

The analysis follows a complete machine learning workflow from data exploration to model evaluation.

2. Dataset Exploration

Dataset Overview

Records: 1,000

Variables:

Study_Hours (float: 2.0-10.0 hours)

Exam_Score (int: 50-100 points)

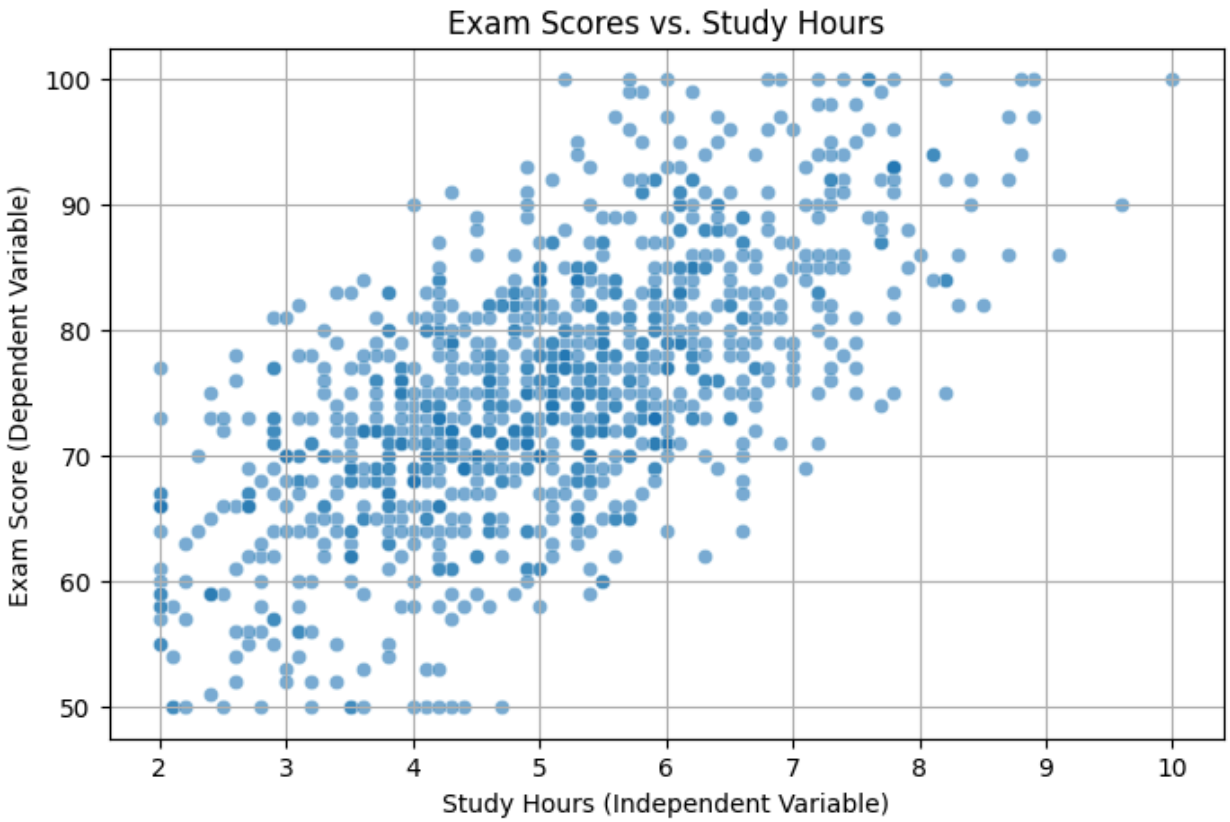
House_Area_sqft (int: 800-2,500 sq ft)

House_Price_USD (int: $172,866-$600,000)

Key Statistics

Variable	Mean	Std Dev	Min	Max
Study_Hours	5.03	1.45	2.0	10.0
Exam_Score	75.23	10.36	50	100
House_Area_sqft	1,149	501	800	2,500
House_Price_USD	$329,456	$101,433	$172,866	$600,000

Visualization



Positive correlation visible between study hours and exam scores

3. Model Implementation

Data Splitting

Training Set: 800 records (80%)

Test Set: 200 records (20%)

Random State: 42 (for reproducibility)

Model 1: Exam Score Prediction

Regression Equation:

Exam_Score = 4.61 × Study_Hours + 52.14

Interpretation:

Baseline score (0 hours): 52 points

Each additional study hour → +4.61 points

Model 2: House Price Prediction

Regression Equation:

Price = $194.50 × Area + $106,751.79

Interpretation:

Base land value: $106,752

Construction cost: $194.50/sq ft

4. Model Evaluation

Exam Score Model Metrics

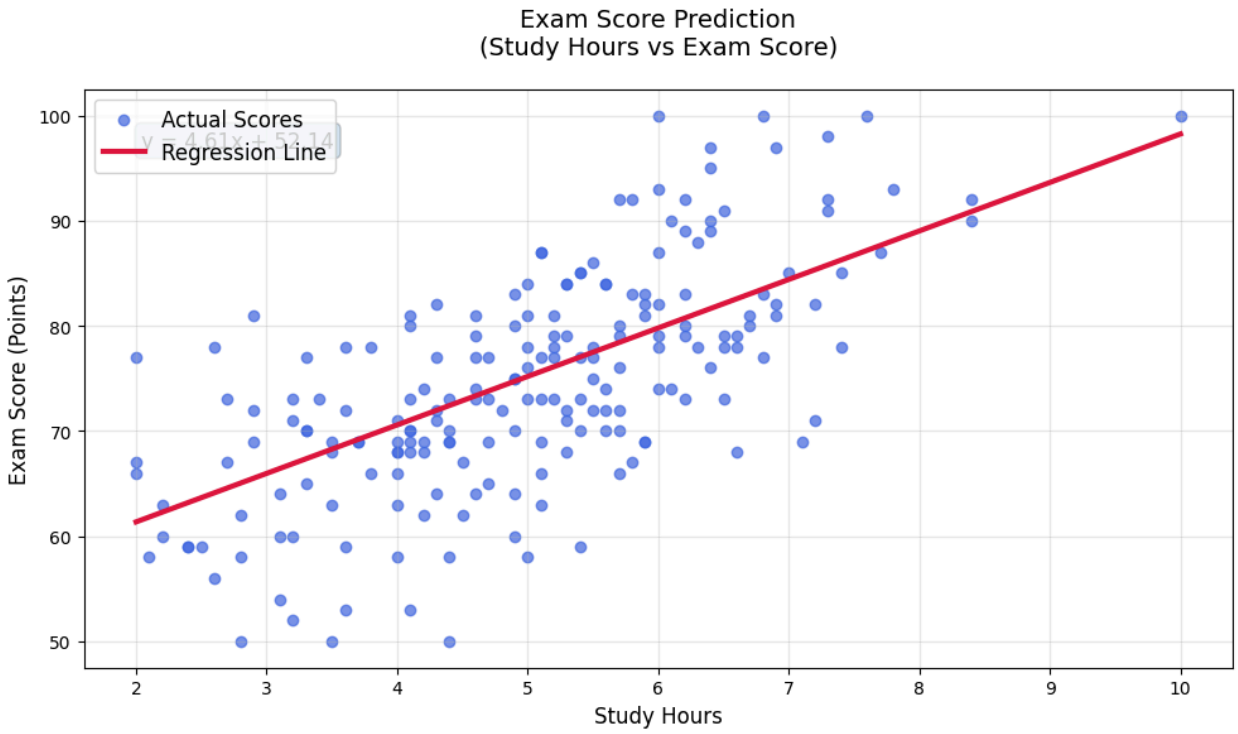
Metric	Value	Interpretation
MAE	6.18 points	Average prediction error
RMSE	7.84 points	Larger errors penalized
R^2	0.45	45% variance explained

House Price Model Metrics

Metric	Value	Interpretation
MAE	$20,526	Typical prediction error
RMSE	$26,429	Sensitive to large errors
R^2	0.94	94% variance explained

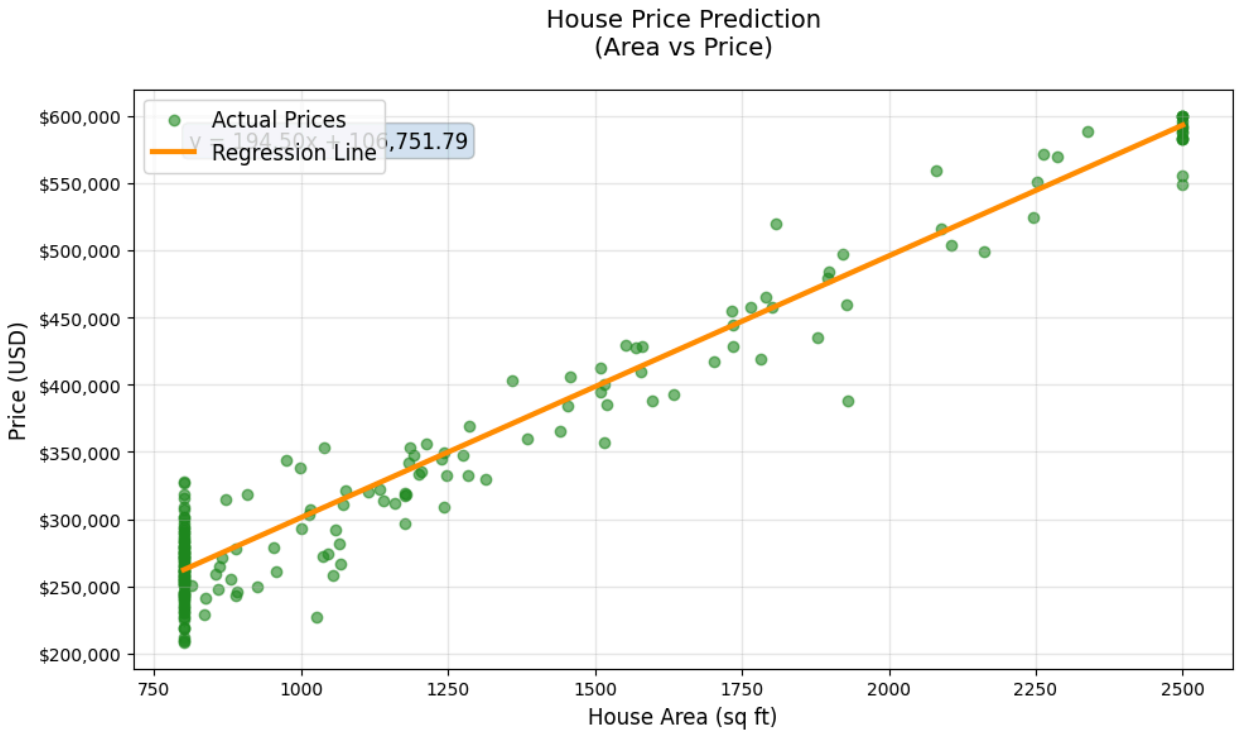
5. Visualization Results

Exam Score Prediction



Regression line shows moderate predictive power

House Price Prediction



Tight fit indicates excellent predictive accuracy

6. Conclusion

Key Findings

Study hours moderately predict exam scores (R^2=0.45)

House area strongly predicts prices (R^2=0.94)

House price model shows professional-grade accuracy (MAE ±$20.5K)

Learnings

Simple linear regression works exceptionally well for clear linear relationships

Domain context matters (e.g., minimum house value reflects land costs)

Evaluation metrics should align with business needs (dollar errors vs points)

Recommendations

For exam predictions: Incorporate additional features (e.g., prior GPA)

For housing model: Explore location-based segmentation

7. Project Artifacts

Code Repository: Link

Notebook: Link

Dataset: exam_house_dataset.csv

All plots and metrics generated from reproducible code (see Appendix for full implementation)