



Legendary Performance! 🏆

Congratulations on successfully completing this assignment! Your grade has been recorded. Feel free to close this tab and return to the main course page.



Required passing grade: 50%

Status: **Passed**

Final Score: 9 / 10 (90%)

[Provide Assignment Feedback](#)



Question 1

Score: **9/10**

Upload your duly completed Jupyter Notebook for the final assignment on *Supervised Machine Learning: Regression*.

- final_project_supervisedML.ipynb [View Content](#)

Well-structured notebook with clear data exploration, multiple regression variants (SLR, MLR, Ridge, Lasso), evaluation metrics and diagnostic plots. Minor improvements recommended around encoding choices, validation, and a more detailed plan for limitations/next steps.

Detailed Breakdown:

Analysis: The submission loads and inspects the housing dataset, shows `head()/info()/describe()`, checks missing values, and visualizes distributions (histogram, boxplot) and relationships (scatter, heatmap). Categorical features are label-encoded and the train/test split, scaling, and model training (Simple LR, Multiple LR, Ridge, Lasso) are implemented. Evaluation uses R2, adjusted R2, RMSE, and MAE and includes residual diagnostics and predicted vs actual plots. Code is runnable and outputs are shown. **Weaknesses:** use of `LabelEncoder` for nominal multi-class features (`furnishingsstatus`) without one-hot encoding can impose ordinal structure; no cross-validation or hyperparameter tuning for regularized models; Lasso/Ridge alpha chosen arbitrarily; no pipeline to combine encoding and scaling (risk of leakage if reuse); minor reproducibility notes (`random_state` provided only for split).



Evaluation: Rubric decisions: data description and visualizations are complete; main objective and planning are present; multiple regression variations including regularization are implemented; clear findings and diagnostic plots are provided; limitation discussion exists but is brief. Overall earns 9/10 due to a concise but not fully detailed plan for revisiting/fixing remediation.

Explanation: Score breakdown evidence: (1) Data description (2/2): `head()`, `info()`, `describe()`, correlation heatmap, histograms and scatter plots plus a summary markdown give good coverage. (2) Objective & plan (2/2): Primary Goal stated and exploratory summary identifies target/features and preprocessing needs. (3) Regression variations (2/2): Implemented Simple LR, Multiple LR, Ridge, Lasso with metrics and comparison; regularized models included. (4) Key findings & next steps (2/2): Model comparison, residual analysis, and next steps are present. (5) Flaws & plan (1/2): Limitations and next steps are noted (limited data, missing variables, feature engineering, advanced models) but the plan is high-level and lacks concrete actionable steps (e.g., CV strategy, hyperparameter search, feature selection process), so partial credit.

Guidance: Short actionable improvements: 1) Use pipelines (`ColumnTransformer + Pipeline`) to safely encode categorical variables (one-hot for nominal features) and scale numeric features to avoid leakage. 2) Replace `LabelEncoder` on multi-class nominal columns with `OneHotEncoder` (or target encoding where appropriate). 3) Add cross-validation (e.g., `cross_val_score` or `GridSearchCV`) to tune alpha for Ridge/Lasso and to obtain robust performance estimates. 4) Consider feature engineering (area per bedroom, interaction terms) and polynomial features if nonlinearity suspected. 5) Report coefficient interpretation and check multicollinearity (VIF) before dropping/combining features. 6) Add reproducibility notes and save the final model pipeline.

[Report an issue](#)

[Retake Assignment](#)