

# Evaluation Report for Group 4 Assignment

---

## Introduction

The following document provides a constructive and diplomatically worded evaluation of the work submitted by Group 4 for the Assignment on Multiple Linear Regression. The goal of this report is to offer balanced feedback, highlighting strengths while also identifying areas where the analysis could be improved in future iterations.

Overall, the group presents a solid starting point, with a clear structure and a well-written document. At the same time, some methodological aspects could be developed more fully to meet all requirements of the assignment.

## 1 Hypotheses and Literature Support

The group formulated several hypotheses concerning factors that may influence real estate prices. These hypotheses are clearly stated and each is accompanied by at least one reference to existing literature. This is a positive aspect of the work, as it demonstrates awareness of previous research in the field.

That said, the literature review would benefit from a slightly deeper discussion. For example, comparing or synthesizing findings from multiple studies could strengthen the theoretical foundation of each hypothesis. In some cases, the link between the hypothesis and the variables available in the dataset could be made more explicit.

## 2 Dataset Description

The dataset is described in an organized, reader-friendly manner, dividing variables into logical macro-categories. This helps provide an intuitive understanding of the structure of the data.

However, several variables (such as crime rate, walkability scores, air quality, and school ratings) are not part of the original AmesHousing dataset. While the inclusion of environmental variables is conceptually very interesting, the methodology by which these data were obtained, merged, and cleaned is not fully explained. Providing more detail on the data collection and integration process would significantly improve transparency and reproducibility.

In addition, descriptive statistics or visual summaries of the dataset would strengthen the section by offering a clearer overview of its characteristics.

## 3 Regression Model

The group estimated an OLS regression model with a comprehensive set of predictors. The model is clearly written, and the output is reported in full detail, including coefficients, standard

errors, significance levels, and model fit indicators.

There are, however, several methodological aspects that could be expanded:

- The assignment specifically requests checking the assumptions of OLS (normality, homoscedasticity, independence, absence of outliers, and multicollinearity). These diagnostics are essential for validating the model, but they are not included in the report.
- Some variables included in the model are likely to be highly correlated with each other (e.g., school ratings or size-related house characteristics). A Variance Inflation Factor (VIF) analysis or correlation matrix would help identify potential issues of multicollinearity.
- The interpretation of the regression coefficients could be expanded. For instance, discussing which variables have the most meaningful economic impact would further enrich the analysis.

Overall, the regression model represents a good starting point, but additional diagnostics and interpretation would make the results more robust and informative.

## 4 Diagnostics and Model Refinement

Diagnostic checks are a crucial part of validating a regression model. Although the assumptions are listed in the methodology section, the assignment requires applying them explicitly through statistical tests or visual inspection (e.g., residual plots, Q-Q plots, tests for heteroskedasticity, etc.).

Including such diagnostic analysis would help assess whether the estimated coefficients can be interpreted reliably. If issues were detected, refining the model (e.g., transforming variables, removing or combining predictors, or testing alternative specifications) would be the natural next step.

## 5 Conclusions

The document concludes with a complete presentation of the regression output, but a short reflective summary discussing the implications of the results, the strengths or limitations of the model, and suggestions for further work would help close the analysis more effectively.

## 6 Overall Evaluation

The work submitted by Group 4 is clearly structured and written with care. The inclusion of environmental and distance-based variables shows creativity and an intention to capture multiple dimensions of real estate valuation.

To further strengthen the analysis, the following improvements are suggested:

- Provide more details on how additional environmental variables were obtained and merged.
- Include descriptive statistics or visualizations of the dataset.

- Perform and report OLS diagnostic checks as required by the assignment.
- Assess multicollinearity and, if necessary, refine the model.
- Add a brief concluding section summarizing insights and limitations.

With these additions, the project would become much more complete from both a methodological and analytical standpoint.

## Final Remarks

The report demonstrates a solid foundation and clear effort from the group. With some further refinement in the statistical methodology and more emphasis on diagnostics, the analysis would become significantly more robust. Nonetheless, the work represents a promising attempt at applying multiple linear regression in a real-world context.