Replace with a specific project title

STAT 8561 | Fall 2026

Replace with all author names

Replace with about 150-200 words stating the problem, data or implementation, evaluation, main quantitative result, and conclusion. Write this after completing the analysis.

Keywords: Replace with 3-5 terms relevant to your application and methods.

Template note (remove before submission). Write **8-10 pages of main text**, counting the abstract, Sections 1-6, and their figures, tables, captions, equations, and code examples. Exclude a separate title-only cover page, References, Author contributions, and appendices or supplementary material. The first page here includes the abstract and counts as a main-text page. References starts on a new page to mark the end of the counted text. This short template illustrates structure, not the required report length.

1 Introduction

State the substantive question and why it matters. Distinguish the question about adjusted association from the question about predicting new outcomes. Cite relevant background and preview the report's contribution.

2 Data and study design

Describe the original data source, observational unit, response units, candidate predictors, collection process, sample size, exclusions, and missing-value handling. Explain dependence or grouping if present.

Include a motivating plot based on the training data, with units, legend, and a caption. Explain the train/test split and how it respects the study design. Keep the test outcomes out of exploration and model selection.

3 Linear regression analysis

Write and explain the model, assumptions, coding, and any transformations or interactions. Interpret at least two meaningful coefficients or contrasts with estimates and 95% confidence intervals.

Compare one scientifically motivated reduced/full model pair using the same training observations and response scale. State the restriction, test, and limitations. Include residual-versus-fitted, normal Q-Q, and influence assessments; explain any action taken. Acknowledge exploratory model selection and distinguish association from causation.

4 Machine learning method

Describe one regression algorithm and why it fits the prediction task. State the available predictors, training-based preprocessing, tuning settings, and validation design. Keep learned preprocessing inside each training fold and finalize both models before examining test performance.

5 Results and predictive comparison

Compare the training-mean baseline, final linear regression, and chosen machine learning method on the same test observations. Report test RMSE and MAE in original response units and explain any inverse response transformation.

Model	Test RMSE	Test MAE
Training-mean baseline	Replace	Replace
Multiple linear regression	Replace	Replace
Chosen machine learning method	Replace	Replace

Add observed-versus-predicted plots for both fitted models, interpret errors, and explain whether the difference matters in this application. Generate all reported numbers and figures from the submitted source.

6 Discussion and conclusion

Answer the research question. Explain substantive findings, uncertainty, practical model choice, and limits to generalization. Discuss at least one limitation and one useful next step.

References

Replace this prompt with complete author-year references for data, methods, software, and adapted code. If using a BibTeX file, add `bibliography: references.bib` to the YAML and cite entries in the text. Include the bibliography file with your submission.

Author contributions

List each author's contribution, or state that this is an individual project.

Supplementary material

Include supporting code and additional output as needed. Record seed 8561, split identifiers, software versions, data-access instructions, and fresh-session rendering steps. All main-text figures and tables should be reproducible from the submitted files.