Final Data Science Project Report Assignment

Due: Friday, August 15th at Midnight

This final report is the culmination of your Module 3 project, building upon the exploratory analyses and modeling milestones you've already completed. Your report should clearly communicate your findings, analysis approach, and conclusions to a technical audience. The following structure and guidelines, informed by best practices, will help you prepare a professional and comprehensive document.

This document is worth 25 points towards your final project total of 100 points.

Your report must be **submitted as a PDF file** and must include the following sections:

1. Executive Summary (Abstract) [1 pt]

- Brief overview of the entire project (150–200 words)
- Clearly state the objective, approach, and key findings

2. Introduction [1 pt]

- Clearly introduce the topic and context of your project
- Describe the problem you are addressing (the problem statement)
- Clearly state the objectives and goals of your analysis

Note: You may imaginatively consider this project as taking place in a real estate company with a small data science group in-house, and write your introduction from this point of view (don't worry about verisimilitude to an actual company!).

3. Data Description [1 pt]

- Describe the source of your dataset (described in Milestone 1)
- Clearly state the characteristics of your data (size, types of features, missing values, target, etc.)

4. Methodology (What you did, and why) [10 pts]

Focus this section entirely on the steps you took and your reasoning behind them. Emphasize the process and decision-making, not the results themselves.

- Describe your analytical framework

        - Use of validation curves to see the effect of various hyperparameter choices, and
        - Choice of MAE as primary error metric
    - Clearly outline your data cleaning and preprocessing steps
        - Describe what issues you encountered in the raw data and how you addressed them.
        - Mention any key decisions (e.g., removing samples with too many missing values).
        - What worked and what didn't work?
    - Describe your feature engineering approach
        - Explain any transformations, combinations, or derived features.
        - Discuss why certain features were chosen or created, even if they were later discarded.
        - What worked and what didn't work?
    - Detail your model selection process
        - Outline the models you experimented with and why.
        - Discuss how you evaluated generalization (e.g., cross-validation, shape and relationships of plots).
        - Mention how you tuned hyperparameters or selected the final model.

5. Results and Evaluation (What you found, and how well it worked) [10 pts]

Focus purely on outcomes, with metrics, visuals, and insights. This is where you present evidence to support your conclusions.

- Provide a clear and detailed narrative of your analysis and reasoning using the analytical approach described in (4).
- Discuss model performance metrics and results (MAE, R^2, etc.)
- Include relevant visualizations (graphs, charts, tables) with appropriate labels and captions
- Error analysis
- Highlight specific patterns of error, outliers, or questionable features.
- Note anything surprising or worth improving in future iterations.

6. Conclusion [2 pts]

- Clearly state your main findings and how they address your original objectives
- Highlight the business or practical implications of your findings
- Discuss the limitations and constraints of your analysis clearly and transparently
- Suggest potential improvements or future directions