Data Science Report Requirements

Overall Format and Navigation

Your report should be submitted in PDF document format.

Include the following:

- Table of Contents
- Five main sections:

    1. Objectives
    2. Data
    3. Modeling
    4. Conclusions
    5. Appendix

Focus on analysis and results. Do not include large blocks of code in the body (sections 1-4). Place any of your code included for reference in the Appendix. Do not include any code provided to you.

Details on what to include in each main section are outlined below.

Objectives

Include a clear description of the modeling scenario(s) and objective(s)

Data

Include the following:

- Brief description of the initial data
- Summary of
    1. data cleaning
    2. data restructuring
    3. feature engineering
- Relevant exploration results
- Description of test data split

Use visualizations (figures) where helpful. All visualizations should include axis labels and titles.

Include any code you wrote for cleaning, feature engineering, and the test data split in the Appendix.

Modeling

Include a subsection for each modeling scenario addressed.

Model Subsections

Include the following:

- Statement of the objective(s)
- Final model description:
    1. Model type
    2. Hyperparameters
    3. Required predictor features
- Training description:
    1. Hyperparameter optimization (if any)
    2. Customizations to the cost (if any) with explanation
    3. Other actions (if any, e.g. data over/under sampling) with explanation
    4. Validation method and metrics
- Test metrics
- Summary of models and/or features determined to be suboptimal

All metrics relevant to the modeling objective(s) should be provided.

Use visualizations (figures) where helpful. All visualizations should include axis labels and titles.

Include code for training (e.g. exported from the app) and testing in the Appendix.

Conclusions

Include the following:

- Overall analysis
- Any relevant comparison of the final models
- Overall recommendation(s)

Appendix

Include any code you wrote for the following purposes:

- Cleaning, feature engineering, and the test data split
- Model training, validation, and testing

Subsections here should be linked to the corresponding sections in the main body.