# Final Project Grading Rubric

|  |  |  |  |
| --- | --- | --- | --- |
| **Section** | **Criteria** | **Detailed Elements** | **Points** |
| A. Introduction | Overview and insights from analysis | - Clearly states problem type (regression or classification), including any conversion from regression to classification. - Provides key findings and identifies which inputs/features most influence the response/outcome. - Discusses conclusions, indicating support from predictive models, EDA, and clustering. - Reflects on specific skills learned and general skills applicable to future applications. | 5 |
| B. EDA | Data understanding and visualization quality | - Basic data info provided: rows, columns, variable names, types, missing/unique values. - Uses visualizations (counts, distributions, relationships) effectively. - Draws key insights from figures and avoids redundancy. | 10 |
| C. Clustering | Clustering execution and interpretation | - Applies clustering techniques accurately, using continuous inputs and necessary transformations (e.g., standardization, handling non-Gaussian variables).  - Justifies the optimal number of clusters using appropriate methods and explains the reasoning.  - Visualizes clusters effectively and compares cluster assignments to categorical variables or outcomes.  - Provide key characteristics of each cluster and explain how these relate to the predictive models and insights from EDA. | 15 |
| D. Models: Fitting and Interpretation | Model selection, fitting, and feature analysis | - Fits all required models, including specified formulations and two additional models based on EDA findings. - Analyzes and interprets coefficients for the top two performing models and any model of your choice, identifying statistically significant features and their positive or negative influence.  - Evaluates model performance using appropriate metrics, including Accuracy, Precision, Recall, F1 Score, and AUC. Provides visualizations such as predicted vs. observed plots, confusion matrices, and ROC curves to support a comprehensive assessment.  - Compares model performances, discussing which model is best according to different metrics and whether it outperforms the intercept-only model. | 30 |
| E. Models: Predictions | Prediction process and visualization for models | - Creates and utilizes visualization grids as per instructions, selecting the most important inputs based on model coefficients, with approaches tailored to regression or classification tasks.  - For Regression Problems, accurately visualizes predictions for both models, including average outputs, confidence intervals, and prediction intervals. For Classification Problems, accurately visualizes predicted event probabilities for both models.  - Discuss the confidence intervals, prediction intervals, or trends in uncertainty across model predictions. Explains what these indicate about the reliability of the model’s predictions and how this uncertainty might impact decision-making or further analysis. | 15 |
| F. Models: Performance and Validation | Cross-validation and best model identification | - Selects appropriate models for cross-validation and uses suitable methods and performance metrics based on the problem type.  - Applies preprocessing steps correctly within each fold (e.g., standardization), and justifies any preprocessing done before cross-validation.  - Visualizes cross-validation results with average performance metrics and 95% confidence intervals for each model.  - Compares cross-validation results with training set performance, identifies the best model, discusses any differences, and reports the number of coefficients in the best model. | 25 |
|  |  |  | 100 |