PCPA Post-Project Summary

The PCPA Post-Project Summary is designed to provide candidates with insightful observations on candidates' project performance, coupled with expert recommendations for improvement. This resource consists of a summary section for the PCPA Project. We will continue to provide updates and enhancements to the summary in the future.

General Comment:

Please refer to the Success Criteria for the Project and related Domains.

PCPA Project Specific Comments:

Domain A: Dealing with Data

- Common mistakes included:
  - Not splitting the data into training and testing datasets or not using cross-validation.
  - Splitting the data into training and testing but not explaining how the data was split (e.g., 70%/30% random sample).
- The key to this domain is that there is some data subsetting procedure that needs to be done in order to validate the model on data that the model has not seen yet.

Domain B: Model Diagnostics & Selection

- Common mistakes included:
  - Not validating the model using data the model has not seen yet. (Acceptable methods of splitting the data are: (a) a train/test split; (b) cross-validation.)

Domain C: Model Interpretation & Presentation

- Many candidates omitted discussion of potential flaws of modeling approach.

Released: March 18, 2026

- Common mistakes included:
    - Not providing visualizations.
    - Choosing visualizations to include in the report that do not support the predictive power of the model. While these may be helpful as intermediate review plots in a work context, the purpose of including visualizations in a report is to justify the performance of the model. Examples: histograms of model scores, QQ plots, residual plots at the policy level. In the context of insurance predictive modeling, exhibits that sort and aggregate data to show overall segmentation provide the clearest conclusions: lift plots, double-lift plots (depending on the candidate models), and Lorenz/Gini curves.
    - Not stating which dataset(s) are used in the validation exhibits. Validation exhibits should only be provided on data the model has not seen (e.g., validation dataset or out-of-fold data).
    - Providing no interpretation of model results at all.
    - Limited or omitted discussion of model coefficients and reasonableness of these coefficients.
    - Variables are listed but no coefficients are listed.
    - No discussion of model validation results.