Phase 3 Project Checklist

- ❏ **Business Understanding:** Notebook clearly explains the project's value for helping a specific stakeholder solve a real-world problem.
  - ❏ Introduction explains the real-world problem the project aims to solve
  - ❏ Introduction identifies stakeholders who could use the project and how they would use it
  - ❏ Conclusion summarizes implications of the project for the real-world problem and stakeholders

- ❏ **Data Understanding:** Notebook clearly describes the source and properties of the data to show how useful the data are for solving the problem of interest.
  - ❏ Describe the data sources and explain why the data are suitable for the project
  - ❏ Present the size of the dataset and descriptive statistics for all features used in the analysis
  - ❏ Justify the inclusion of features based on their properties and relevance for the project
  - ❏ Identify any limitations of the data that have implications for the project

- ❏ **Data Preparation:** Notebook shows how you prepare your data and explains why by including…
  - ❏ Instructions or code needed to get and prepare the raw data for analysis
  - ❏ Code comments and text to explain what your data preparation code does
  - ❏ Valid justifications for why the steps you took are appropriate for the problem you are solving

- ❏ **Modeling:** Notebook demonstrates an iterative approach to model building.
  - ❏ Runs and interprets a simple, baseline model for comparison
  - ❏ Introduces new models that improve on prior models and interprets their results
  - ❏ Explicitly justifies model changes based on the results of prior models and the problem context
  - ❏ Explicitly describes any improvements found from running new models

- ❏ **Evaluation:** Notebook shows how well a final model solves the real-world problem
  - ❏ Justifies choice of metrics using context of the real-world problem and consequences of errors
  - ❏ Identifies one final model based on performance on the chosen metrics with validation data
  - ❏ Evaluates the performance of the final model using holdout test data
  - ❏ Discusses implications of the final model evaluation for solving the real-world problem

- ❏ **Visualization:** Notebook includes three relevant and polished visualizations of findings that…
  - ❏ Help the project stakeholder understand the value or success of the project
  - ❏ Have text and marks to aid reader interpretation, such as graph and axis titles, axis ticks and labels, or legend (varies by visualization type)
  - ❏ Use color, size, and/or location to appropriately facilitate comparisons
  - ❏ Are not cluttered, dense, or illegible

- ❏ **Code Quality:** Code in notebook and related files meets professional standards (e.g. PEP 8)
  - ❏ Code is easy to read, using comments, spacing, variable names, and function docstrings
  - ❏ All code runs and no code or comments are included that are not needed for the project
  - ❏ Code minimizes repetition, using loops, functions, and classes
  - ❏ Code adapted from others is properly cited with author names and location of the cited material

- ❏ **GitHub Repository:** Project repository demonstrates professional “best practices”:
  - ❏ README.md includes concise summary of project with all data science steps
  - ❏ README.md links to presentation and sources
  - ❏ README.md includes instructions for navigating the repository
  - ❏ Files and folders are named briefly and descriptively, with consistent naming conventions
  - ❏ Files and folders are organized logically and consistently
  - ❏ Commit history includes regular commits with informative commit messages
  - ❏ Large or sensitive files are listed in .gitignore and not pushed to GitHub

- ❏ **Presentation Content:** Presentation clearly demonstrates the value of the project to stakeholders by…
  - ❏ Using plain language and clear visuals accessible to non-technical stakeholders
  - ❏ Describing the project goals, data, methods, and results
  - ❏ Explicitly connecting the descriptions of the project to stakeholder needs

- ❏ **Slide Style:** Slides have a professional style, such that...
  - ❏ Slides use a professional template
  - ❏ Slides are not cluttered
  - ❏ Slides are light on text
  - ❏ Slide text is easily readable
  - ❏ Visuals are easy to understand

- ❏ **Presentation Delivery:** Deliver your presentation clearly and engagingly by...
  - ❏ Speaking at a moderate volume and pace
  - ❏ Describing your project simply and succinctly in about 5 minutes
  - ❏ Using pauses, emphasis, or other variation in your speaking throughout the presentation
  - ❏ Having a distinct introduction and conclusion

- ❏ **Answers to Questions:** Answer questions clearly and appropriately by...
  - ❏ Directly addressing all aspects of the question that was asked
  - ❏ Responding accurately, succinctly, and in plain language
  - ❏ Being sensitive to the knowledge level and interests of the asker
  - ❏ Explaining any reasons why you cannot fully answer a question

Phase 3 Project Scoring Guidance

Business Understanding

Exemplary	The notebook's explanation of a project's value clearly identifies how the project can help a specific stakeholder solve a real-world problem.
Complete	The notebook's explanation of a project's value is in general terms, but does not clearly specify the problem and stakeholder the project aims to support. This includes projects that do not make it clear how the project can support specific stakeholder activities in pursuit of a specific goal.
Incomplete	The notebook's explanation of a project's value is missing, inaccurate, or difficult to understand. This includes projects that do not actually have value for any real-world problem/stakeholder.

Data Understanding

Exemplary	The notebook's descriptions of the data's source or properties clearly address how well-suited the data are for addressing the real-word problem.
Complete	The data's source and properties are described in the notebook and are potentially relevant to the real-world problem of interest, but the connection is not made clearly. This includes exploratory analyses that reveal properties of data but that does not go further to explain what those properties imply about the potential value of the data for addressing the real-world problem.
Incomplete	The notebook's description of the data's source or properties is missing, inaccurate, difficult to understand, or irrelevant to the real-world problem of interest. This includes projects using data that are not appropriate for solving the chosen real-world problem.

Data Preparation

Exemplary	Data preparation is fully reproducible using instructions and code that is contained in, imported by, or referenced by the demonstration notebook. AND Preparation steps have code comments, explanatory text, and valid justifications that clearly explain how the data is being manipulated and why those manipulations make the data more useful for solving the real-world problem addressed by the project.
Complete	Data preparation is fully reproducible using instructions and code that is contained in, imported by, or referenced by the demonstration notebook. AND Preparation steps could have valid rationales for why they will help with solving the real-world problem addressed by the project, but the steps taken or rationales are not clearly documented.
Incomplete	Data preparation is not fully reproducible using instructions and code that is contained in, imported by, or referenced by the demonstration notebook.

	OR Preparation steps could not have valid rationales for why they will help with solving the real-world problem addressed by the project.

Modeling

Exemplary	Model iterations are developed by methodically improving from prior models, have clear rationales for why they will help solve the real-world problem, and describe any improvements found.
Complete	Model iterations are developed by methodically improving from prior models and could have valid rationales for why they will help solve the real-world problem, but the rationales are not clearly documented or the model improvements are not clearly discussed.
Incomplete	Model changes are not developed by methodically improving from prior models or could not have valid rationales for why they will help solve the real-world problem.

Classification Results

Exemplary	A final classification model is chosen based on performance and evaluated holistically to assess its utility for solving the business problem. This includes not only calculating performance on an appropriate metric using holdout test data, but also evaluation of other relevant properties of the model and a detailed discussion of how well the model solves the real-world problem informed by those properties and metrics.
Complete	A final classification model is chosen based on performance and is correctly evaluated using holdout test data and an appropriate metric, but the evaluation does not go beyond calculating performance on the metric.
Incomplete	No final classification model is explicitly chosen, the final model is not correctly evaluated using holdout test data, or the chosen metric is not appropriate for the real-world problem.

Visualization

Exemplary	Three visualizations in the notebook are relevant and polished, including all elements from the checklist
Complete	Three visualizations in the notebook are relevant and interpretable, but missing elements from the checklist
Incomplete	There are not three relevant and interpretable visualizations in the notebook

Code Quality

Exemplary	Code is all runnable, easy to read, non-repetitive, and properly cited, with no substantial room for improvement
Complete	Code is mostly runnable, easy to read, non-repetitive, and properly cited, but has some

	substantial room for improvement. This includes projects that have lots of unnecessary comments or code that is not used or needed.
Incomplete	Code is mostly not runnable, difficult to read, repetitive, or improperly cited.

GitHub Repository

Exemplary	Repository includes all items from the checklist (7 out of 7)
Complete	Repository includes most items from the checklist (4-6 out of 7)
Incomplete	Repository does not include most items from the checklist (0-3 out of 7)

Presentation Content

Exemplary	Presentation clearly describes the project goals, data, methods, and results, using this information to help demonstrate how the project's value to stakeholder
Complete	Presentation describes the project goals, data, methods, and results, but does not clearly convey the value of the project to the business stakeholders. This includes presentations that are somewhat confusing, include irrelevant information, or omit evidence-based recommendations
Incomplete	Presentation does not describe the project goals, data, methods, or results. This includes presentations that omit critical information, have substantial errors, or that are too confusing for stakeholders to follow.

Slide Style

Exemplary	Most slides include all items from the checklist (5 out of 5)
Complete	Most slides include most items from the checklist (3-4 out of 5)
Incomplete	Most slides do not include most items from the checklist (0-2 out of 5)

Presentation Delivery

Exemplary	Presentation engages the audience throughout, clearly and compellingly conveying the content and connecting it to stakeholder needs
Complete	Presentation conveys the intended content, but is not consistently clear and engaging to stakeholders. This includes presentations that are delivered too fast, explained in a confusing manner, too long, or recited monotonously
Incomplete	Presentation does not convey the intended content. This includes presentations that are not comprehensible or that skip intended content

Answers to Questions

Exemplary	Most answers are fully clear and appropriate
Complete	Most answers are somewhat clear and appropriate. This includes responses that are long,

	jargon heavy, only answer part of the question, or are not sensitive to the asker's knowledge
Incomplete	Most answers are unclear or inappropriate. This includes non-responses, incorrect responses, or responses that don't resolve the question