v2.3 Phase 3 Project Checklist

Presentation Content

This element assesses the content of the non-technical presentation of a student's project. Presentations are evaluated based on how clearly they demonstrate the value of the project to the business stakeholders.

- The intended audience is the project's business stakeholders, not the class or teacher
- Presentations should aim to persuade stakeholders to value the project, not just describe it
- This element assesses **the content of** a presentation, not the delivery of it

    - **Content** refers to the words and pictures describing the project
    - Delivery refers to the manner in which those words and pictures are conveyed to an audience

- This element assesses **the demonstration of** a project's value, not the value itself

Complete	**Presentation describes the project goals, data, methods, and results.** Presentation may or may 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.

- ☐ **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
    - ☐ Having a distinct introduction and conclusion

Slide Style

This element assesses the style of a student's slides. Slides are evaluated based on how well the style enables the presenter to communicate with an audience and the audience to access the content.

- Style includes all visual elements, such as layout, colors, and fonts
- Style should match the expectations of a professional business presentation

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

☐ **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 and Answers to Questions

This element assesses the delivery of the non-technical presentation of a student's project, including how well students respond to questions. Presentations are evaluated based on how well the delivery of the presentation engaged the project's business stakeholders in understanding the content, and on the clarity and appropriateness of answers to questions.

- Presentations should engage stakeholders by talking about the project's value to them
- The intended audience is the project's business stakeholders, not the class or teacher
- This element assesses **the delivery of** a presentation, not the content of it
    - Content refers to the words and pictures describing the project
    - **Delivery** refers to the manner in which those words and pictures are conveyed to an audience
- An answer is **clear** if it addresses the question directly, concisely, and in plain language
- An answer is **appropriate** if it accurately represents the project, resolves the question, and is sensitive to the audience.

Complete	**Presentation conveys the intended content, and most answers to questions are at least somewhat clear and appropriate**. Presentation may or may not be consistently clear and engaging to stakeholders. This includes presentations that are delivered too fast, explained in a confusing manner, too long, or recited monotonously. This also includes answers to questions that are long, jargon-heavy, only answer part of the question, or are not sensitive to the asker's knowledge, so long as they represent the project accurately and address the question being asked
Incomplete	**Presentation does not convey the intended content, or most answers to questions are unclear or inappropriate**. This includes presentations that are not comprehensible or that skip intended content, as well as non-responses, incorrect responses, or responses that don't resolve the question

☐ **Presentation Delivery and Answers to Questions:** Deliver your presentation clearly and engagingly by...

    ☐ Describing your project simply and succinctly in about 5 minutes
    ☐ Using pauses and emphasis while speaking at a moderate volume and pace
    ☐ Being sensitive to the knowledge level of your audience, using plain language where possible
    ☐ Directly addressing all aspects of any questions that were asked, including explaining any reasons why you cannot fully answer a question
    ☐ Answering questions accurately and succinctly

Business Understanding

This element assesses how well students explain the value of their projects. We frame value in terms of the problems and stakeholders that are the focus of the project.

- Students must explain how their projects address a ***real-world problem***
  - In Phase 1, the problem basics are provided in the project description
- A ***real-world problem*** is an actual problem faced by an actual ***stakeholder***
  - A ***stakeholder*** is a specific individual, group of people, or organization
  - A ***problem*** is a challenge faced by a stakeholder in pursuit of some goal
  - A ***real-world problem*** is one that exists in reality, i.e. students could present the project to an actual human being and they could actually find value in it.
- This element assesses ***the explanation of*** a project's value, not the value itself

Complete	**The notebook's explanation of a project's value is present, accurate, and understandable.** Notebook may or may not clearly identify how the project can help a specific stakeholder solve a real-world problem. This includes notebooks with a vague or general explanation.
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.

- ☐ **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

This element assesses how well students demonstrate the utility of their data for helping solve a business problem. We frame utility in terms of the properties, source, and business relevance of the data.

- Students must show how the data are useful for addressing a *real-world problem*
  - In Phase 1, the problem basics are provided in the project description
- A *real-world problem* is an actual problem faced by an actual *stakeholder*
  - A *stakeholder* is a specific individual, group of people, or organization
  - A *problem* is a challenge faced by a stakeholder in pursuit of some goal
  - A *real-world problem* is one that exists in reality, i.e. students could present the project to an actual human being and they could actually find value in it
- This element assesses *the demonstration of* the data's utility, not the utility itself

Complete	**The data's source and properties are described in the notebook and are potentially relevant to the real-world problem of interest.** 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. It also includes notebooks that clearly address how well-suited the data are 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 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

This element assesses how well students prepare their data for analysis. Good data preparation is reproducible, well-documented, and justified.

- Data preparation is *reproducible* if a third party with the same data access could create the same analytic dataset just by using the files and instructions provided in the repository
- A data preparation step is *justifiable* if there could be a valid rationale for how that step makes the data better suited for analyses that will address the real-world problem
- Data preparation is *well-documented* if it includes enough code comments and explanatory text for a third party to easily understand the steps that were taken and their *justifications*

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. The steps taken or rationales may or may not be clearly documented using code comments and explanatory text.
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.

☐ **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

This element assesses how well the student demonstrates an iterative approach to modeling in their notebook. Model building is iterative when it proceeds from a baseline model to more complex models based on rationales that are refined over each iteration.

- New models should be **justifiable**, progressing from a **baseline model** and yielding **improvements**

  - A **baseline model** is an extremely simple model used for comparison, such as a simple average or decision rule
  - Model iterations are **justifiable** if there could be a valid rationale for how the new model could better address the real-world problem. Justifications can include factors such as the results of prior models, properties of the data, computational power available, or the problem context
  - An **improvement** can be any reason that a new model is better than an old one, such as reduced error, computational time, or complexity

Complete	**Model iterations are developed by methodically improving from prior models**. Rationales may or may not be documented.
Incomplete	**Multiple models are not developed, or the differences between models could not have valid rationales** for why they will help solve the real-world problem.

☐ **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

This element assesses how well students interpret and evaluate the results of their models for projects in Phase 3 and later.

- Students should choose a ***final model*** to evaluate, using ***appropriate*** metrics to show its ***utility***
    - A ***final model*** is chosen, based on performance, for further evaluation and recommended use
    - A metric is ***appropriate*** if it is sensitive to the real-world consequences of errors
    - A model has ***utility*** if it could justifiably be used to solve a real-world problem
- The final model should be evaluated ***properly*** to show how well it solves the real-world problem
    - Correct use of training, validation and test data, without leakage
    - May consider factors besides metrics, such as runtime, explainability, parsimony, or ease of use
- For Phase 3, students must create a classification model
    - For Phase 4 and 5, students have the choice of a classification or regression model
- This element assesses the ***choice and evaluation of a final model***, not the performance or utility of the model itself
    - In other words, there is no benchmark requirement for model performance (e.g. "90% accuracy") that must be achieved to fulfill this requirement

Complete	**A final model is chosen based on performance and is correctly evaluated using holdout test data and an appropriate metric.** Ideally the model will be evaluated holistically to assess its utility for solving the business problem, including factors besides metrics, but this is not required for a Complete rating.
Incomplete	**No final 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.**

- ☐ **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

Code Quality

This element assesses the quality of code that students use in their project. Quality is assessed based on readability, repetitiveness, and citation.

- **_Readability_** is the ability for a newcomer to look at code and quickly understand what it does
  - Use professional coding conventions to evaluate this (e.g. PEP 8)
- **_Repetitiveness_** is the unnecessary use of similar code multiple times in one project
  - Can be avoided in many ways, such as with loops, functions, or classes
  - In other words, code should be DRY
- All code adapted from others must have proper **_citation_**
  - At a minimum, citations should include the authors' names and location of the cited material
  - Refer to the "Plagiarism Policy" section of the Flatiron School Catalog for definitions, policies, and procedures related to plagiarism and unauthorized assistance

Complete	**Code is mostly runnable, easy to read, non-repetitive, and properly cited.** Code may or may not have some substantial room for improvement. For Phase 1, 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.**

☐ **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

This element assesses the understandability of the GitHub repository that houses a student's project. Understandability is evaluated based on the README, folder structure, and commit history of the repository.

- The README should provide an overview of the project and repository structure
- Files and folders should be organized to make it easy to find any necessary files
- Commit history should reflect the project's development history

Complete	**Repository includes most items from the checklist** (4 or more out of 7)
Incomplete	**Repository does not include most items from the checklist** (0-3 out of 7)

☐ **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