MODULE ANSWER

Plan and Evaluate your Template

Activity

Level: **Moderate**
Duration: ⏳⏳

Instructions

This activity requires some familiarity with the intended audience of your Data Card, and the dataset being documented.

Use this table to decide information relevant to your readers, and how much nuance they require. Use this to decide what to include, how to summarize, and when you're getting too deep in the weeds. As you fill out your Data Card, regularly evaluate it against this table, treating each row as a heuristic. Give your Data Card a severity rating using the scale provided and plan your next steps.

	Reader	Decisions	Goals	Relevance	Nuance
Priority	Who is the primary audience?	What decisions regarding the dataset will they make?	What takeaway do they want from the Data Card?	What specific content do they need from the Data Card to meet their goals?	Given what you know about the reader, how detailed or nuance must your content be?
Priority	*Production Software Engineer*	*If or not to use the dataset to test a model*	*Give me an overview of the dataset. Tell me how it is implemented.*	*Intended and Unsuitable uses, past use - results on past models*	*Highly nuanced. Focus on technical use and usability.*
1					
2					
3					
4					
5					

Takeaways: Planning Your Data Card

- **Priority and Reader:** As a group, write your primary audience groups next to their priority. Based on what success means in the context of your dataset, your dataset may have multiple audiences. For a focused Data Card, it is helpful to identify a primary audience to help you make decisions on what relevant information to include. However, if there are multiple audiences and it is important to point this out (e.g., for Data Cards for data sets used for large models), acknowledge this within the content of Data Card itself.

- **Decisions & Goals:** For each audience group, consider what decisions they might make with respect to your dataset. For example, a software engineer (SWE) who builds machine learning (ML) models for deployment in products may need to decide if or not to use a dataset to test the model they are building. Then, write down what they want from a Data Card to make those decision. This is the goal of the Data Card. For example, the SWE might need an overview of the dataset and the implementation of the dataset.

- **Relevance:** What specific content does each audience group need to easily and efficiently find in your Data Card? Keeping with our previous example, the SWE will need to quickly understand the intended uses, unsuitable uses, and past uses cases to assess if or not they want to use the dataset to evaluate their ML models. Once they've decided to use the dataset, they will also need to know the behavior of other ML models on your dataset.

- **Nuance:** Some readers will require simpler explanations, while more advanced readers will want to consume technical information with additional context. Not all readers will be able to consume information at the same level of detail. In our example, to satisfy the needs of a typically advanced SWE role, your Data Card will need to focus on performance results in past uses, and the technology stack and implementation details for your dataset.

Planning your Data Card	By focusing on the decisions that your Data Card Readers need to make, you can decide how to frame the content in your Data Card and how much detail you need to get into.Use this table as a reference each time you are unsure about what or how much content to include in your Data Card.

Heuristics: Evaluate your Data Card

Treating each row in this table as a heuristic, regularly evaluate your Data Card against this table to ensure that your high priority readers find your Data Card useful. Give your Data Card a severity rating for violations of heuristics, and plan your next steps.

A heuristic, or heuristic technique, is any approach to problem solving or self-discovery that employs a practical method that is not guaranteed to be optimal, perfect, or rational, but is nevertheless sufficient for reaching an immediate, short-term goal or approximation.

The severity scale below provides a rating of how broken something is and the impact of the issue, without regard to prioritization. These can have an impact on the trust placed by a reader in the Data Card and the usefulness of the Data Card. While precise definitions can vary by team, a more general description of severity in the context of usability is available at https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

Violation: What answers are not useful for the reader?
Severity: How badly does this need to be fixed on a scale of 1 to 5?   ☐ 1 = Catastrophic, Fix this before Data Card can be released ☐ 2 = Major problem, Important to fix, given high priority ☐ 3 = Minor problem, Fixing this should be given low priority ☐ 4 = Cosmetic problem only, Fix if time is available ☐ 5 = I don't agree that this is a problem at all
Fix: What is the solution?

Evaluating your Data Card	Many times, it simply isn't possible to get feedback from readers about the usefulness and usability of your Data Card. This can slow down the pace at which you refine your content, or can lead to providing information that readers don't find useful. Repurposing your key takeaways into heuristics can help you anticipate and avoid common pitfalls in any documentation.