**Exercise content**

In this exercise, you'll be walking through the first three steps of the FactSheet methodology. These first three steps are designed to help you develop a Template; the list of Facts relevant to your use case. Ideally, you'd be interacting with your FactSheet consumers and Fact producers, but for the purposes of this exercise, it's more of a thought exercise.

As you walk through the first three steps of the methodology, you'll be filling out questions to help you think about the needs of your users and the use case. You're free to use your own use case, but if you don't have one, we are providing an example one for you to use. The goal of this exercise is to familiarize yourself with the methodology and end up with a draft template.

At the end of this exercise please refer to “***Methodology Exercise – Example Responses.doc***” In this document we'll show an example answer for the below steps that you can use to better understand the intent of each step and compare it with your own response.

Keep in mind that responses should be written in such a way that they can be consumed by someone who is not only a data scientist and if possible, contextualized in the business use case of selling widgets.

**Describe your use case.**

Be specific and include the role and needs of your intended FactSheet consumer. If you don't have a use case in mind, you can copy and paste the following example use case that you will use for the rest of this exercise:

* + *"Carmen, a business owner is looking to increase sales of widgets. Her data science team has developed an AI model to predict how various advertising campaigns will impact sales across demographics. Carmen has limited knowledge of data science, but still wants to understand the model's quality, how it helps her drive sales of widgets, and identify any potential bias the model may have."*

**Step 1: Know your FactSheet consumers.**

In this step, try to put yourself in the shoes of the person or people who will be reading the final FactSheet. Use the following questions to guide your thinking. (The questions below are written from the perspective of the example. If you're using your own example, adjust role names accordingly). Keep in mind that not all the questions may be relevant. If you're using our example, you may fill in the details of Carmen's role as you see fit. These questions are not comprehensive but should start your thinking along the right track.

Select and answer two of the questions below. We've bolded some suggestions, but use whichever ones are most appropriate to your use case.

* + What does Carmen do now when she performs her role?
  + What is Carmen going to be asking for when looking at a FactSheet?
  + What decisions will she be making based on the information presented?
  + How is the FactSheet going to help her do her job more effectively?
  + **What are the most important pieces of information that Carmen needs to know?**
  + What is Carmen’s level of expertise in general data science?
  + **How is Carmen’s expertise going to affect the information presented?**
  + Will there need to be additional definitions for terms that Carmen is unfamiliar with?
  + What is Carmen’s level of expertise with respect to the model algorithms being used?
  + What explanations about the model’s algorithm or results is Carmen going to need?
  + What is Carmen’s level of expertise in the problem domain?
  + How is that going to affect the information presented?
  + Will Carmen need help in mapping general knowledge of the problem domain to the particular inputs, outputs, or performance indicators associated with this model?
  + Is Carmen aware of issues related to model risk, potential harm, and regulatory compliance?
  + What information is needed to assess these issues?

**Step 2: Know your FactSheet producers.**

In this step, try to put yourself in the shoes of the person or people who will provide the Facts for you. Use the following questions to guide your thinking. If you're using our example, Priya will be your producer.

* + *"Priya was the lead data scientist on Carmen's model and she was responsible for much of its development. She was the one responsible for talking to the sales team to understand how widgets are sold, collecting the relevant historical advertising data for the model and leading the data cleaning and development efforts of the model."*

(The questions below are written from the perspective of the example. If you're using your own example, adjust role names accordingly). Keep in mind that not all the questions may be relevant. If you're using our example, you may fill in the details of Priya's role as you see fit. These questions are not comprehensive, but should start your thinking along the right track.

Select and answer two of the questions below. We've bolded some suggestions, but use whichever ones are most appropriate to your use case.

* + What facts does Priya wish she could conveniently record about the models she develops? It is often helpful to ask about the most recent model, or a model that was particularly important, or a model that was exceptionally difficult to produce, rather than discussing models in general.
  + What did Priya do during the creation of this model that is otherwise unknown to others?
  + Are there general facts about the data, the features, the model algorithm, or the training and testing Priya performs that are important to note? Why?
  + **What model-specific knowledge does she have that may not be obvious to others?**
  + What domain-specific knowledge does Priya have that may not be obvious to others?
  + **Does Priya know who will be consuming the facts she produces? We will assume it is Carmen in this particular case. Does Priya know Carmen? Have they talked about what Carmen needs to know?**
  + Is Priya aware of issues related to model risk, potential harm, and regulatory compliance?
  + What information will be needed by others to assess these issues?

**Step 3: Create a FactSheet template.**

**3.a**

In this step, you'll list the topics and questions to be included in the FactSheet, guided by your thinking in steps 1 and 2. You can think of a template as a questions part of a question/answer form or as the keys part of a key/value pairs. We've provided some questions to help guide your thinking below. Use the following questions to guide your thinking. (The questions below are written from the perspective of the example. If you're using your own example, adjust role names accordingly). Keep in mind that not all the questions may be relevant.

Some of these responses will be directly relevant to question **3.b**

Select and answer two of the questions below. We've bolded some suggestions, but use whichever ones are most appropriate to your use case.

* + - **What are the topics or categories of information needed?**
    - Do some of these categories have subcategories?
    - **What is a meaningful name for each category or subcategory?**
    - **What kinds of information should be included in each category? For example, Carmen may want to group all the model performance metrics within a category called "Model Performance". Information about the representativeness of the training data might be grouped with information on the sensitivity of the model to drift in a category called, ’Potential Sources of Error’.**
    - How should each question in a category be worded so as to be both understandable and evocative for Priya? The goal here is to encourage fact producers to answer in ways that are concise, germane, and understandable.
    - Where will the answer to a question come from? Will it be generated automatically by a tool or entered by a knowledgeable human? If the former, will Priya have some control over the frequency of fact generation or the granularity of recorded facts? If the latter, will Priya be given hints or examples of the kind of answer that would be satisfactory?
    - Are there any regulatory, legal, or business concerns that need to be considered when answering the questions in this template?
    - Are there different presentation formats needed for this information (for example, a short tabular summary of just key facts, or a slide format for presentations to review boards)? AI FactSheets 360 [[link](https://aifs360.mybluemix.net/)] shows three different formats that might be useful.
    - In addition to the human-readable content, is there a need for machine-readable content that Priya might generate?

**3.b** Finally, write out your list of Fact categories or questions to form your template in the space below.