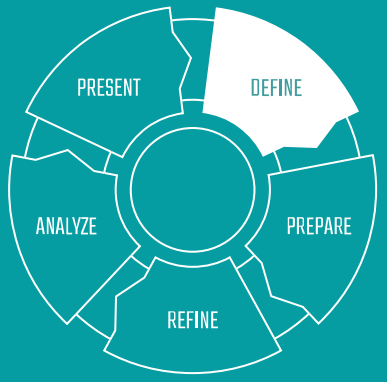
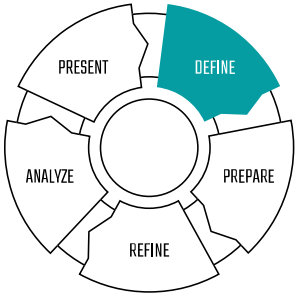


DEFINE

Laying the Foundation for Successful Data Analysis

PRAGMATIC INSTITUTE

The Data Incubator
A Pragmatic Institute Company



DEFINE

Laying the Foundation for Successful Data Analysis

YOUR STAKEHOLDER'S TIME—AND YOURS—IS PRECIOUS.

If you're a data analyst fielding a new request, it can be tempting to jump right into analysis or begin exploring solutions. But beware of that trap. While getting to work right away might seem quicker than gathering context in preparation for a project, it will lead to more work and frustration in the end. As Albert Einstein is said to have quipped at a Yale faculty club, "If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and five minutes thinking about solutions."

You can avoid disconnects and make the most of your first meetings by establishing clear goals from the outset in partnership with your stakeholder. Get to the crux of the problem by starting with the five W's (and an H). Finding answers to the basics—who, what, when, where, why and how—will head off any misunderstandings down the line and help build consensus with stakeholders.

On the following pages, find the most important questions to ask at the beginning of a data analysis project. Begin with the key questions in the chart to lay the foundation. The answers to these questions may seem obvious, but you will often be surprised.

> "If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and five minutes thinking about solutions."

After getting answers to these questions, move on to filling in the details with the subsequent list of enquiries, identifying and fleshing out the business problem to solve. Armed with the information you need, you can perform a more robust analysis that exceeds your stakeholder's expectations.

The "Define" Stage in Action

THE REQUEST

A fashion retailer asks you what products are likely to be most successful next spring based on the last few years of sales and marketing research.

1 ASK KEY QUESTIONS

Before you start setting up a regression analysis or random forest, can you answer the five W's (and an H)?

2 ENSURE IT'S SOLVABLE

If the retailer seeks to measure success by "timeless design quality," that may not be a tractable question for a data analyst. However, if the retailer releases new fashions each spring, it may be reasonable to use previous data to predict future sales.

3 CONFIRM YOUR UNDERSTANDING

After you've collected more information and considered the five Ws (and an H), you can confirm your understanding of the problem with the stakeholder. It might look something like this: "You are launching five new styles in addition to your current 20 in the spring, and you would like an estimate of how many products to order from suppliers and in what sizes and colors. Is that right?"

4 PROVIDE OPTIONS

If the retailer confirms your understanding of the problem, you can move forward with your analysis to answer this supply question. However, there may still be more value to offer. For example, additional questions could explore marketing and pricing questions to provide richer insights. Be sure to present options and anticipate the related needs of your stakeholder before finalizing your plans.

Start by Finding Answers To These Questions

WHO ...

are the stakeholders?

are the subject matter experts you can consult?

will use the analysis?

WHAT ...

are the stakeholder's goals?

are the stimulus and response variables?

is the desired outcome (e.g., description, prescription, prediction)?

WHEN ...

do decisions need to be made with this data?

was data collected last?

will the data be released, archived or destroyed?

WHERE ...

is your data coming from?

will you store data as you work with it?

will your data go afterward?

WHY ...

are these questions being posed to you now?

is this particular stakeholder involved?

would you choose one method over another?

HOW ...

do these questions relate back to the stakeholder's goals?

can you confirm or negate assumptions?

can you justify your recommendations?

Then, Build a Deeper Understanding

WHO

- **Who** will you report to directly? Who do they report to?
- **Who** gathered the data?
- **Who** is in charge of the data you'll need to access?
- **Who** is represented in the data?
- **Who** is vulnerable in this process? (Is private or personal information involved? Is proprietary company information used?)

WHAT

- **What** problems are stakeholders trying to solve with this analysis?
- **What** SMART questions stem from these goals (SMART = Specific, Measurable, Attainable, Relevant, Time-based)?
- **What** challenges might prevent you from answering the questions posed?
- **What** data is already available?
- **What** can be generalized from this data?
- **What** are the limitations in the budget?
- **What** further data is needed?
- **What** is a successful outcome of the project?

WHEN

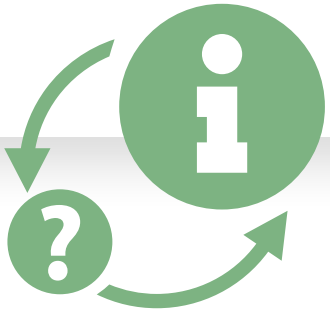
- **When** does the data need to be compiled?
- **When** will data need to be collected again?
- **When** will the analysis need to be revisited or revised?

WHERE

- **Where** will you analyze the data (e.g., cloud storage, local, copies, originals)?
- **Where** will you store your analysis?
- **What** places in the real world does your data represent?
- **Where** is geographic information applicable?
- **Where** are mapping tools required for location-based analytics?
- **Where** are regions or geographical divisions most applicable?

WHY

- **Why** would you use or exclude certain data points?
- **Why** would you frame the problem in this way and not another?
- **Why** might you choose the results of one model over another (e.g., accuracy vs. explainability, specificity vs. generalizability)?



HOW

- **How** can you confirm your understanding of the problem?
- **How** do you expect your analysis to be used?
- **How** could things go wrong in your analysis?
- **How** can you explain your findings?
- **How** can you account for uncertainty in your analysis?
- **How** do you address ethical, legal and social issues?

Want to exceed your stakeholders' expectations with critical insights?

Register for Pragmatic Institute's new course, *Business-Driven Data Analysis*. Learn a proven, repeatable approach that you can leverage across data projects and toolsets to deliver timely data analysis with actionable insights. Identify the right question and the right data, optimize results, communicate them effectively and ensure stakeholder alignment. **Learn More**

PRAGMATIC INSTITUTE

The Data Incubator
A Pragmatic Institute Company