Course – 2 Ask Questions to Make Data-Driven Decisions

Week1-Effective questions

The six data analysis phases

Step1: Ask
It's impossible to solve a problem if you don't know what it is. These are some things to consider:
- Define the problem you're trying to solve
- Make sure you fully understand the stakeholder's expectations
- Focus on the actual problem and avoid any distractions
- Collaborate with stakeholders and keep an open line of communication
- Take a step back and see the whole situation in context

Questions to ask yourself in this step:
- What are my stakeholders saying their problems are?
- Now that I've identified the issues, how can I help the stakeholders resolve their questions?

Step 2: Prepare
You will decide what data you need to collect in order to answer your questions and how to organize it so that it is useful.
You might use your business task to decide:
- What metrics to measure
- Locate data in your database
- Create security measures to protect that data

Questions to ask yourself in this step:
- What do I need to figure out how to solve this problem?
- What research do I need to do?

Step 3: Process
Clean data is the best data and you will need to clean up your data to get rid of any possible errors, inaccuracies, or
inconsistencies. This might mean:
- Using spreadsheet functions to find incorrectly entered data
- Using SQL functions to check for extra spaces
- Removing repeated entries
- Checking as much as possible for bias in the data

Questions to ask yourself in this step:
- What data errors or inaccuracies might get in my way of getting the best possible answer to the problem I am trying to solve?
- How can I clean my data so the information I have is more consistent?

Step 4: Analyze
You will want to think analytically about your data. At this stage, you might sort and format your data to make it easier to:
- Perform calculations
- Combine data from multiple sources
- Create tables with your results

Questions to ask yourself in this step:
- What story is my data telling me?

- How will my data help me solve this problem?
- Who needs my company's product or service? What type of person is most likely to use it?

Step 5: Share

Everyone shares their results differently so be sure to summarize your results with clear and enticing visuals of your analysis using data viz tools like graphs or dashboards.
This is your chance to show the stakeholders you have solved their problem and how you got there.
Sharing will certainly help your team:
- Make better decisions
- Make more informed decisions
- Lead to stronger outcomes
- Successfully communicate your findings

Questions to ask yourself in this step:
- How can I make what I present to the stakeholders engaging and easy to understand?
- What would help me understand this if I were the listener?

Step 6: Act

Now it's time to act on your data. You will take everything you have learned from your data analysis and put it to use.
This could mean providing your stakeholders with recommendations based on your findings so they can make
data-driven decisions.

Questions to ask yourself in this step:
- How can I use the feedback I received during the share phase (step 5) to actually meet the stakeholder's needs and expectations?

These six steps can help you to break the data analysis process into smaller, manageable parts, which is called **structured thinking**. This process involves four basic activities:

- Recognizing the current problem or situation
- Organizing available information
- Revealing gaps and opportunities
- Identifying your options

Six common types of problems that data analysts work with:

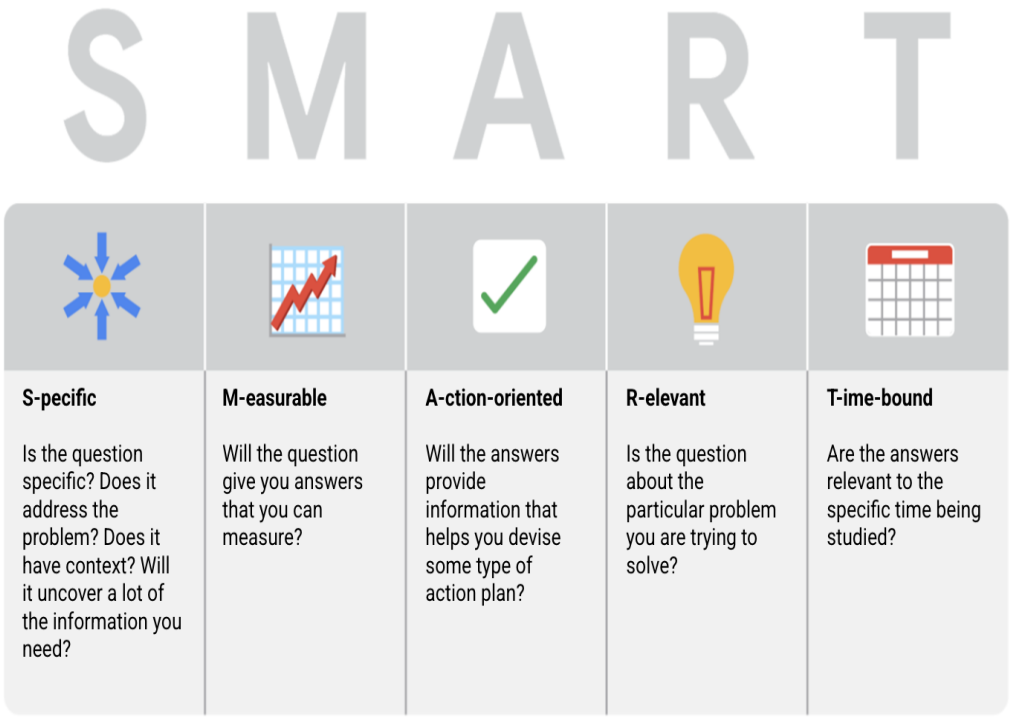
- **Making predictions**: Using data to make an informed decision about how things may be in the future.
- **Categorizing things**: Assigning information to different groups or clusters based on common features.
- **Spotting something unusual**: Identifying data that is different from the norm.
- **Identifying themes**: It takes categorization as a step further by grouping information into broader concepts.
- **Discovering connections**: Finding similar challenges faced by different entities, and then combining data and insights to address them.
- **Finding patterns**: Using historical data to understand what happened in the past and is therefore likely to happen again.

Note: Categorizing things involves assigning items to categories; identifying themes takes those categories a step further by grouping them into broader themes.

Avoid asking:

- **Leading questions**: questions that only have a particular response. Leading questions direct the respondent to a particular answer, often because they suggest the answer within the question. Example: "These are the best sandwiches ever, aren't they?" This question doesn't really give you the opportunity to share your own opinion, especially if you happen to disagree and didn't enjoy the sandwich very much. This is called a leading question because it's leading you to answer in a certain way.
- **Closed-ended questions**: questions that ask for a one-word or brief response only. Example: "Were you satisfied with the customer trial?" This question is closed-ended. That means it can be answered with a yes or no.
- **Vague questions**: questions that aren't specific or don't provide context. Example: "Does the tool work for you?" This question is too vague because there is no context.

The more questions you ask, the more you learn about your data, and the more powerful your insights will be. Asking thorough and specific questions means clarifying details until you get to concrete requirements. With clear requirements and goals, it's much easier to plan and execute a successful data analysis project and avoid time-consuming problems down the road. Effective questions follow the SMART methodology:



- **Specific**: Specific questions are simple, significant and focused on a single topic or a few closely related ideas. This helps us collect information that's relevant to what we're investigating. If a question is too general, try to narrow it down by focusing on just one element.
- **Measurable**: Measurable questions can be quantified and assessed.
- **Action-oriented**: Action-oriented questions encourage change. For example, rather than asking, "how can we get customers to recycle our product packaging?" You could ask, "what design features will make our packaging easier to recycle?" This brings you answers you can act on. The questions that are action-oriented are more likely to result in specific answers that can be acted on to lead to change.
- **Relevant**: Relevant questions matter, are important and have significance to the problem you're trying to solve.
- **Time-bound**: Time-bound questions specify the time to be studied. This limits the range of possibilities and enables the data analyst to focus on relevant data.

There's something else that's very important to keep in mind when crafting questions, fairness. Fairness means ensuring that your questions don't create or reinforce bias. Fairness also means crafting questions that make sense to everyone. It's important for questions to be clear and have a straightforward wording that anyone can easily understand. Unfair questions also can make your job as a data analyst more difficult. They lead to unreliable feedback and missed opportunities to gain some truly valuable insights. A common example of an unfair question is one that makes assumptions. These are questions that assume the answer to the question being asked.

Questions should be open-ended. This is the best way to get responses that will help you accurately qualify or disqualify potential solutions to your specific problem.

Week2-Data-driven decisions

Data-inspired decision making

Explores different data sources to find out what they have in common.

Algorithm

An algorithm is a process or set of rules to be followed for a specific task.

The goal of all data analysts is to use data to draw accurate conclusions and make good recommendations. That all starts with having complete, correct, and relevant data. It is possible to have solid data and still make the wrong choices. It is up to data analysts to interpret the data accurately. When data is interpreted incorrectly, it can lead to huge losses. When data is used strategically, businesses can transform and grow their revenue. There is a difference between making a decision with incomplete

data and making a decision with a small amount of data. Making a decision with incomplete data is dangerous. But sometimes accurate data from a small test can help you make a good decision.

There are a lot of different kinds of questions that data might help us answer, and these different questions make different kinds of data.

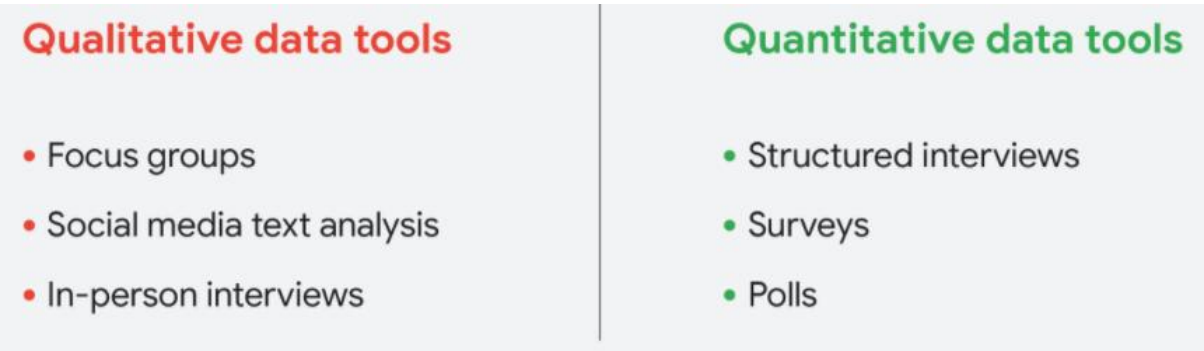
Quantitative data

Quantitative data is all about the specific and objective measures of numerical facts. This can often be the *what*, *how many*, and *how often* about a problem. In other words, things you can measure, such as a number, quantity or range.

Qualitative data

Qualitative data is a subjective and explanatory measure of a quality or characteristic. Basically, the things that can't be measured with numerical data, like your hair color. Qualitative data is great for helping us answer *why* questions.

With quantitative data, we can see numbers visualized as charts or graphs. Qualitative data can then give us a more high-level understanding of why the numbers are the way they are. It helps us add context to a problem.



Data analysts will generally use both types of data in their work. Usually, qualitative data can help analysts better understand their quantitative data by providing a reason or more thorough explanation. In other words, quantitative data generally gives you the what, and qualitative data generally gives you the why.

Two data presentation tools:

Reports

A report is a static collection of data given to stakeholders periodically.

Pros:

- Reports are great for giving snapshots of **high level historical data** for an organization.
- They can be designed and sent out periodically, often on a weekly or monthly basis, as organized and easy to reference information. They're **quick to design and easy to use** as long as you continually maintain them.
- Since reports use static data or data that doesn't change once it's been recorded, they reflect data that's already been **cleaned and sorted**.

Cons:

- Reports need **regular maintenance**.
- They are **less visually appealing**.
- Because they aren't automatic or dynamic (**static**), reports don't show live, evolving data.

One way spreadsheet data could be visualized in a report:

Pivot Table

A data summarization tool that is used in data processing. Pivot tables are used to summarize, sort, re-organize, group, count, total, or average data stored in a database. It allows its users to transform columns into rows and rows into columns.

Dashboard

Monitors live, incoming data. A dashboard is a single point of access for managing a business's information. It allows analysts to pull key information from data in a quick review by visualizing the data in a way that makes findings easy to understand.

Pros:

- **Dynamic, automatic and interactive**: They give your team more access to information being recorded, you can interact through data by playing with filters, and because they're dynamic, they have long-term value.

- **More stakeholder access** and **Low maintenance**: If stakeholders need to continually access information, a dashboard can be more efficient than having to pull reports over and over.

Cons:

- **Labor-intensive design**: They take a lot of time to design and can actually be less efficient than reports, if they're not used very often.
- **Can be confusing**: If the base table breaks at any point, they need a lot of maintenance to get back up and running again.
- **Potentially uncleaned data**: Dashboards can sometimes overwhelm people with information too. If you aren't used to looking through data on a dashboard, you might get lost in it.

A dashboard would be useful for monitoring data as it becomes available. You might create a Tableau dashboard with interactive graphs that showcase multiple views of the data.

3 most common types of dashboards:

- **Strategic**: focuses on long term goals and strategies at the highest level of metrics
- **Operational**: short-term performance tracking and intermediate goals
- **Analytical**: consists of the datasets and the mathematics used in these sets. hese dashboards contain the details involved in the usage, analysis, and predictions made by data scientists.

Dashboards identify metrics: Relevant metrics may help analysts assess company performance.

Dashboards can help companies perform many helpful tasks, such as:

- Track historical and current performance.
- Establish both long-term and/or short-term goals.
- Define key performance indicators or metrics.
- Identify potential issues or points of inefficiency.

Metric

A metric is a single, quantifiable type of data that can be used for measurement. Data starts as a collection of raw facts, until we organize them into individual metrics that represent a single type of data. Metrics can also be combined into formulas that you can plug your numerical data into. Data contains a lot of raw details about the problem we're exploring. But we need the right metrics to get the answers we're looking for.

Different industries use all kinds of different metrics. But there's one thing they all have in common: they're all trying to meet a specific goal by measuring data.

Using key performance indicators to measure revenue and using annual profit targets to set and evaluate goals are examples of using metrics.

Metric Goal

A measurable goal set by a company and evaluated using metrics.

Mathematical thinking: It means looking at a problem and logically breaking it down step-by-step, so you can see the relationship of patterns in your data, and use that to analyze your problem. This kind of thinking can also help you figure out the best tools for analysis because it lets us see the different aspects of a problem and choose the best logical approach.

Small data

These kinds of data tend to be made up of datasets concerned with *specific metrics over a short, well defined period of time*. Small data can be useful for making day-to-day decisions, like deciding to drink more water. But it doesn't have a huge impact on bigger frameworks like business operations.

Example: The amount of exercise time it takes for a single person to burn a minimum of 400 calories is a problem that requires small data. It contains a specific metric (400 calories) and a short, defined period of time (amount of exercise time).

Big data

Big data has *larger, less specific datasets covering a longer period of time*. They usually have to be broken down to be analyzed. Big data is useful for looking at large- scale questions and problems, and they help companies make *big decisions*.

Small data	Big data
Describes a data set made up of specific metrics over a short, well-defined time period	Describes large, less-specific data sets that cover a long time period
Usually organized and analyzed in spreadsheets	Usually kept in a database and queried
Likely to be used by small and midsize businesses	Likely to be used by large organizations
Simple to collect, store, manage, sort, and visually represent	Takes a lot of effort to collect, store, manage, sort, and visually represent
Usually already a manageable size for analysis	Usually needs to be broken into smaller pieces in order to be organized and analyzed effectively for decision-making

Some challenges you might face when working with big data:

- A lot of organizations deal with data overload and way too much unimportant or irrelevant information.
- Important data can be hidden deep down with all of the non-important data, which makes it harder to find and use. This can lead to slower and more inefficient decision-making time frames.
- The data you need isn't always easily accessible.
- Current technology tools and solutions still struggle to provide measurable and reportable data. This can lead to unfair algorithmic bias.
- There are gaps in many big data business solutions.

Some benefits that come with big data:

- When large amounts of data can be stored and analyzed, it can help companies identify more efficient ways of doing business and save a lot of time and money.
- Big data helps organizations spot the trends of customer buying patterns and satisfaction levels, which can help them create new products and solutions that will make customers happy.
- By analyzing big data, businesses get a much better understanding of current market conditions, which can help them stay ahead of the competition.
- Big data helps companies keep track of their online presence—especially feedback, both good and bad, from customers. This gives them the information they need to improve and protect their brand.

V words for big data

- **Volume** describes the amount of data.
- **Variety** describes the different kinds of data.
- **Velocity** describes how fast the data can be processed.
- **Veracity** refers to the quality and reliability of the data.

These are all important considerations related to processing huge, complex data sets.

Week3-More spreadsheet basics

Spreadsheet tasks

- Organize your data
  - Pivot tables
    - Sort and filter
- Calculate your data
  - Formulas
  - Functions

▼ Spreadsheets and the data life cycle

- **Plan** for the users who will work within a spreadsheet by developing organizational standards. This can mean formatting your cells, the headings you choose to highlight, the color scheme, and the way you order your data points. When you take the time to set these standards, you will improve communication, ensure consistency, and help people be more efficient with their time.
- **Capture** data by the source by connecting spreadsheets to other data sources, such as an online survey application or a database. This data will automatically be updated in the spreadsheet. That way, the information is always as current and accurate as possible.
- **Manage** different kinds of data with a spreadsheet. This can involve storing, organizing, filtering, and updating information. Spreadsheets also let you decide who can access the data, how the information is shared, and how to keep your data safe and secure.
- **Analyze** data in a spreadsheet to help make better decisions. Some of the most common spreadsheet analysis tools include formulas to aggregate data or create reports, and pivot tables for clear, easy-to-understand visuals.
- **Archive** any spreadsheet that you don't use often, but might need to reference later with built-in tools. This is especially useful if you want to store historical data before it gets updated.
- **Destroy** your spreadsheet when you are certain that you will never need it again, if you have better backup copies, or for legal or security reasons. Keep in mind, lots of businesses are required to follow certain rules or have measures in place to make sure data is destroyed properly.

Formulas

A formula is a set of instructions that perform a specific calculation. Formulas are built on operators which are symbols that name that type of operation or calculation to be performed.

Cell reference

A cell reference is a single cell or range of cells in a worksheet that can be used in a formula. Cell references contain the letter of the column and the number of the row where the data is. A range of cells is a collection of two or more cells. A range can include cells from the same row or column, or from different columns and rows collected together. The great thing about using cell references is that they also automatically update when a formula is copied to a new cell.

Auto-fillingAbsolute referencingData rangeCombining with functions

Spreadsheet errors and fixes

Functions

A function is a preset command that automatically performs a specific process or task using the data.

Difference between formulas and functions

- A formula is a set of instructions used to perform a calculation using the data in a spreadsheet.
- A function is a preset command that automatically performs a specific process or task using the data in a spreadsheet.

Relative, absolute, and mixed references

- Relative references (cells referenced without a dollar sign, like A2) will change when you copy and paste the function into a different cell. With relative references, the location of the cell that contains the function determines the cells used by the function.
- Absolute references (cells fully referenced with a dollar sign, like A2) will not change when you copy and paste the function into a different cell. With absolute references, the cells referenced always remain the same.
- Mixed references (cells partially referenced with a dollar sign, like $A2 or A$2) will change when you copy and paste the function into a different cell. With mixed references, the location of the cell that contains the function determines the cells used by the function, but only the row or column is relative (not both).
- In spreadsheets, you can press the F4 key to toggle between relative, absolute, and mixed references in a function. Click the cell containing the function, highlight the referenced cells in the formula bar, and then press F4 to toggle between and select relative, absolute, or mixed referencing.

DAC2 Keyboard functions 1, DAC2 Keyboard functions 2.

Problem domain

The specific area of analysis that encompasses every activity affecting or affected by the problem.

Carefully defining a business problem can ultimately save time, money, and resources. All of this is achieved through structured thinking.

Structured thinking

The process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options.

The starting place for structured thinking is the problem domain. Once you know the specific area of analysis, you can set your base and lay out all your requirements and hypotheses before you start investigating.

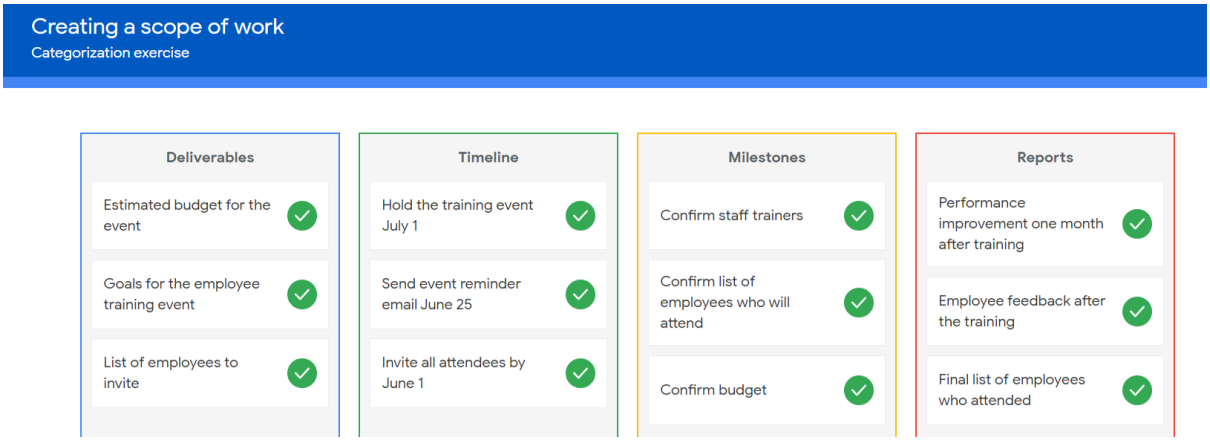
You can practice structured thinking and avoid mistakes is by using a scope of work.

Scope of work (SOW)

An agreed- upon outline of the work you're going to perform on a project. A scope of work is project-based and sets the expectations and boundaries of a project. A scope of work keeps everyone on the same page. Using structured thinking, you can define what is being delivered, when, and how you will measure success along the way.

There's no standard format for an SOW. They may differ significantly from one organization to another, or from project to project. However, they all have a few foundational pieces of content in common:

- **Deliverables** are items or tasks you will complete before you can finish the project.
  What work is being done, and what things are being created as a result of this project? When the project is complete, what are you expected to deliver to the stakeholders? Be specific here. Will you collect data for this project? How much, or for how long?
- **Timelines** include due dates for when deliverables, milestones, and/or reports are due.
  The timeline is a way of mapping expectations for how long each step of the process should take. The timeline should be specific enough to help all involved decide if a project is on schedule. When will the deliverables be completed? How long do you expect the project will take to complete? If all goes as planned, how long do you expect each component of the project will take? When can we expect to reach each milestone?
- **Milestones** are significant tasks you will confirm along your timeline to help everyone know the project is on track.
  This is closely related to your timeline. What are the major milestones for progress in your project? How do you know when a given part of the project is considered complete?
- **Reports** notify everyone as you finalize deliverables and meet milestones.
  Good SOWs also set boundaries for how and when you'll give status updates to stakeholders. How will you communicate progress with stakeholders and sponsors, and how often? Will progress be reported weekly? Monthly? When milestones are completed? What information will status reports contain?



[Data Analysis Project Scope-of-Work (SOW) Strong Example]

Usually, projects don't start until an SOW is approved with its key pieces of content: the deliverables, milestones, timeline, and reports. To collect and synthesize this information, analysts identify and formalize quantifiable project requirements. They use structured thinking to ask clarifying questions, define what to accomplish, and specify project boundaries.

Context

The condition in which something exists or happens. To avoid bias when collecting data, a data analyst should keep context in mind. Context can turn raw data into meaningful information. It is very important for data analysts to contextualize their data. This means giving the data perspective by defining it. To do this, you need to identify:

- **Who**: The person or organization that created, collected, and/or funded the data collection
- **What**: The things in the world that data could have an impact on
- **Where**: The origin of the data
- **When**: The time when the data was created or collected
- **Why**: The motivation behind the creation or collection
- **How**: The method used to create or collect it

To ensure your data is accurate and fair, make sure you start with an accurate representation of the population in the sample; collect the data in an objective way; and ask questions about the data.

Week4-Always remember the stakeholder

Focusing on stakeholder expectations will help you understand the goal of a project, communicate more effectively across your team, and build trust in your work.

Working with stakeholders

By asking yourself a few simple questions at the beginning of each task, you can ensure that you're able to stay focused on your objective while still balancing stakeholder needs. You could be working on multiple projects with lots of different people but no matter what project you're working on, there are three things you can focus on that will help you stay on task.

1. Who are the primary and secondary stakeholders?
2. Who is managing the data?
3. Where can you go for help?

- There are four key questions data analysts ask themselves to communicate clearly with stakeholders and team members

    i. Who your audience is
    ii. What they already know
    iii. What they need to know
    iv. How you can communicate that effectively to them

    You'll want your emails to be just as professional as your in-person communications.

- It's important to set realistic expectations at every stage of the project. Setting expectations for a realistic timeline might involve sharing a high-level schedule with stakeholders, creating a schedule, and communicating clearly with team members.

- In the data world, speed can sometimes be the enemy of accuracy, especially when collaboration is required.

- A data analyst reframes a question. Then, they outline the problem, challenges, potential solutions, and timeframe in order to put data into context, balance speed with accuracy, and keep stakeholders informed. (To ensure their work answers the right questions and delivers useful results, the data analyst should set clear expectations, outline the problem, and reframe the question)

- Focusing on stakeholder expectations enables data analysts to understand project goals, improve communication, and build trust.
- Asking questions including, “Does my analysis answer the original question?” and “Are there other angles I haven’t considered?” enable data analysts to consider the best ways to share data with others, help their team make informed decisions, and use data to get to a solid conclusion.
- Data analysts pay attention to sample size in order to represent a diverse set of perspectives and avoid skewed results or inaccurate judgements.

- When leading a meeting, testing out technology, taking notes and preparing supporting materials will help you ensure all participants have a positive experience.

Before the meeting

- If you are organizing the meeting, you will probably talk about the data. Before the meeting:
- Identify your objective. Establish the purpose, goals, and desired outcomes of the meeting, including any questions or requests that need to be addressed.
- Acknowledge participants and keep them involved with different points of view and experiences with the data, the project, or the business.
- Organize the data to be presented. You might need to turn raw data into accessible formats or create data visualizations.
- Prepare and distribute an agenda. We will go over this next.

Crafting a compelling agenda

A solid meeting agenda sets your meeting up for success. Here are the basic parts your agenda should include:

- Meeting start and end time
- Meeting location (including information to participate remotely, if that option is available)
- Objectives
- Background material or data the participants should review beforehand

Sharing your agenda ahead of time

During the meeting

As the leader of the meeting, it's your job to guide the data discussion. With everyone well informed of the meeting plan and goals, you can follow these steps to avoid any distractions:

- Make introductions (if necessary) and review key messages
- Present the data
- Discuss observations, interpretations, and implications of the data
- Take notes during the meeting
- Determine and summarize next steps for the group

After the meeting

- Distribute any notes or data
- Confirm next steps and timeline for additional actions
- Ask for feedback (this is an effective way to figure out if you missed anything in your recap)

To shift a situation from problematic to productive, data analysts can reframe a problem and start a constructive conversation. This will give everyone the chance to share their viewpoints in a productive manner which leads to a more successful project.