6 data analysis phrases:

Step	**Questions to ask yourself**
1. Ask • ***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	1. What are my stakeholders saying their problems are? 2. Now that I've identified the issues, how can I help the stakeholders resolve their questions?
2. Prepare What data you need to collect? • What metrics to measure • Locate data in your database • Create security measures to protect that data	1. What do I need to figure out how to solve this problem? 2. What research do I need to do?
3. Process ***Clean data*** => 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	1. What data errors or inaccuracies might get in my way of getting the best possible answer to the problem I am trying to solve? 2. How can I clean my data so the information I have is more consistent?
4. Analyze ***Sort and format data*** to make it easier to: Perform calculations Combine data from multiple sources Create tables with your results	1. What story is my data telling me? 2. How will my data help me solve this problem? 3. Who needs my company's product or service? What type of person is most likely to use it?

5. Share ***Summarize your results*** with clear and enticing visuals of your analysis using data via tools like graphs or dashboards.	1. How can I make what I present to the stakeholders engaging and easy to understand? 2. What would help me understand this if I were the listener?
6. Act 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.	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:

1. Recognizing the current problem or situation
2. Organizing available information
3. Revealing gaps and opportunities
4. Identifying your options

Commented [2HL1]: Gap analysis: A method for examining and evaluating the current state of a process in order to identify opportunities for improvement in the future

Solving problem with data: 6 common problem types

Making predictions

Using data to make informed decisions about how things may be in the future.

Eg: A company that wants to know the best advertising method to bring in new customers is an example of a problem requiring analysts to make predictions. Analysts with data on location, type of media, and number of new customers acquired as a result of past ads can't guarantee future results, but they can help predict the best placement of advertising to reach the target audience.

Categorizing things

Grouping data based on common features.

An example of a problem requiring analysts to categorize things is a company's goal to improve customer satisfaction. Analysts might classify customer service calls based on certain keywords

or scores. This could help identify top-performing customer service representatives or help correlate certain actions taken with higher customer satisfaction scores.

Spotting something unusual

Identifying data that is different from the norm.

Eg: A company that sells smart watches that help people monitor their health would be interested in designing their software to spot something unusual. Analysts who have analyzed aggregated health data can help product developers determine the right algorithms to spot and set off alarms when certain data doesn't trend normally.

Identifying themes

Recognizing broader concepts and trends from categorized data.

User experience (UX) designers might rely on analysts to analyze user interaction data. Similar to problems that require analysts to categorize things, usability improvement projects might require analysts to identify themes to help prioritize the right product features for improvement. Themes are most often used to help researchers explore certain aspects of data. In a user study, user beliefs, practices, and needs are examples of themes.

Discovering connections:

Identifying similar challenges across different entities—and using data and insights to find common solutions.

Eg: A third-party logistics company working with another company to get shipments delivered to customers on time is a problem requiring analysts to discover connections. By analyzing the wait times at shipping hubs, analysts can determine the appropriate schedule changes to increase the number of on-time deliveries.

➔ When a problem of a company can affect the other company.

Finding patterns

Using historical data about what happened in the past to understand how likely it is to happen again.

Eg: Minimizing downtime caused by machine failure is an example of a problem requiring analysts to find patterns in data. For example, by analyzing maintenance data, they might discover that most failures happen if regular maintenance is delayed by more than a 15-day window.

Craft effective questions

Avoid:

Leading question (This cake is good, isnt it?) -> need to create fair question that anyone can answer

Close-ended question (Yes-no question)

Vague (too general) question

- ⇨ Use SMART question (Specific, Measurable, Action-oriented, Relevant, Time-bound)

Eg: Start a project on weekend sales data. What is the question you should ask before starting the project?

- When is the project due?
- Are there any specific challenges to keep in mind?
- Who are the major stakeholders for this project, and what do they expect this project to do for them?
- Who am I presenting the results to?

Here are some examples of questions you might ask based on the suggested topics:

Objectives: What are the goals of the deep dive? What, if any, questions are expected to be answered by this deep dive?

Audience: Who are the stakeholders? Who is interested or concerned about the results of this deep dive? Who is the audience for the presentation?

Time: What is the time frame for completion? By what date does this need to be done?

Resources: What resources are available to accomplish the deep dive's goals?

Security: Who should have access to the information?

Real world case:
Asking phrase:
1. Plan for the conversation
- Prioritize your question: make most important question and interesting questions first.
- Make your time count by staying on the subject during conversation
- Clarify your understanding
2. Create questions

For instance, if you have a conversation with someone who works in retail, you might lead with questions like:

- **Specific:** Do you currently use data to drive decisions in your business? If so, what kind(s) of data do you collect, and how do you use it?
- **Measurable:** Do you know what percentage of sales is from your top-selling products?
- **Action-oriented:** Are there business decisions or changes that you would make if you had the right information? For example, if you had information about how umbrella sales change with the weather, how would you use it?
- **Relevant:** How often do you review data from your business?
- **Time-bound:** Can you describe how data helped you make good decisions for your store(s) this past year?

3. Take good notes

Exercise: Ask a Marketing team of a vehicle company
Specific: Do you currently use data as a tool to decide in the business? What marketing metrics do you use? Do you get good insight from it? And how?
Measurable: what is the percentage of sales increase from last year? What factors do you think that leads to this success/ failure?
Action-oriented: How customer insights from the data help you to make a good marketing campaign?
Revelant: How often do you review the data from the customer?
Time-bound: What is the time-bound for collecting data from customer before launching a new marketing campaign/ product?

How to utilize the power of data

Having data is not enough.

Data -> make it into information by interpreting and comparing data -> put it into knowledge

Commented [2HL2]: Need to be interpret in a complete and unite way.
Note: small amount of data # incomplete data

Quantitative data and qualitative data

Quantitative data: Chart, graph with numerical data. Answer what, how many, and how often question

➔ Specificc and objective

Qualitative data: Deals with characteristics and interpretion (insights) (from quantitative data.)

➔ Descriptive, subjective and explanatory

How to share data

	Report (Pivot table, etc.)	Dashboard (may cai bieu do nhu trong VietStock)
Def	Static collection of data given to stakeholders periodically -> past data	Monitors live, incoming data
Pros	High-level historical data Easy to design and use Pre-cleaned and sorted data	Dynamic, automatic and interactive More stakeholder access Low maintainance
Cons	Continual maintainance Less visually appealing Static	Labor-intensive design If break -> needs lots of maintainance Can be confusing Potentially uncleaned data

Dashboard

How to create dashboard:

1. Identify the stakeholders who need to see the data and how they will use it

Requirements Gathering Worksheet

2. Design the dashboard (what should be displayed)

- Use a ***clear header*** to label the information
- Add ***short text descriptions*** to each visualization
- Show the ***most important information at the top***

3. Create mock-ups if desired

This is optional, but a lot of data analysts like to sketch out their dashboards before creating them.

4. Select the visualizations you will use on the dashboard

To learn more about choosing the right visualizations, check out Tableau's galleries:

- For more samples of area charts, column charts, and other visualizations, visit Tableau's Viz Gallery
- Explore Tableau's Viz of the Day to see visualizations curated by the community.

5. Create filters as needed

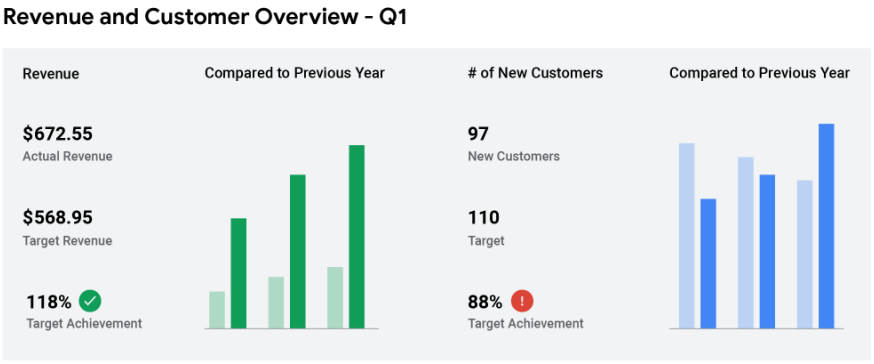
Filters show certain data while hiding the rest of the data in a dashboard. This can be a big help to identify patterns while keeping the original data intact. It is common for data analysts to use and share the same dashboard, but manage their part of it with a filter. To dig deeper into filters and find an example of filters in action, you can visit Tableau's page on Filter Actions. This is a

useful resource to save and come back to when you start practicing using filters in Tableau on your own.

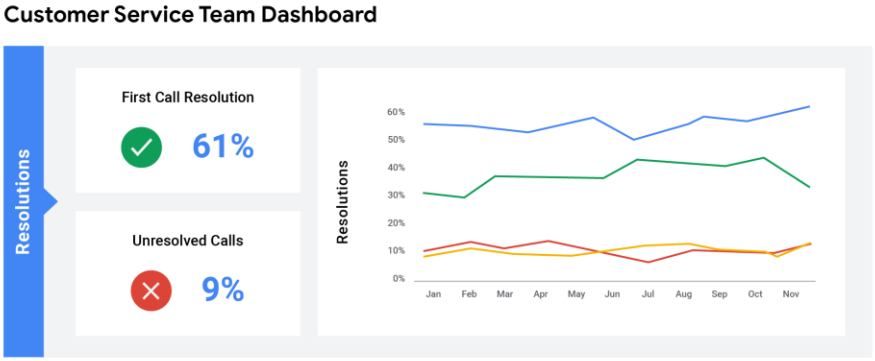
Note: Dashboard helps stakeholder navigate the path of the project inside the data. If you add **clear markers and highlight important points** on your dashboard, users will understand where your data story is headed.

Types of Dashboard

Strategic: focuses on long term goals and strategies at the highest level of metrics

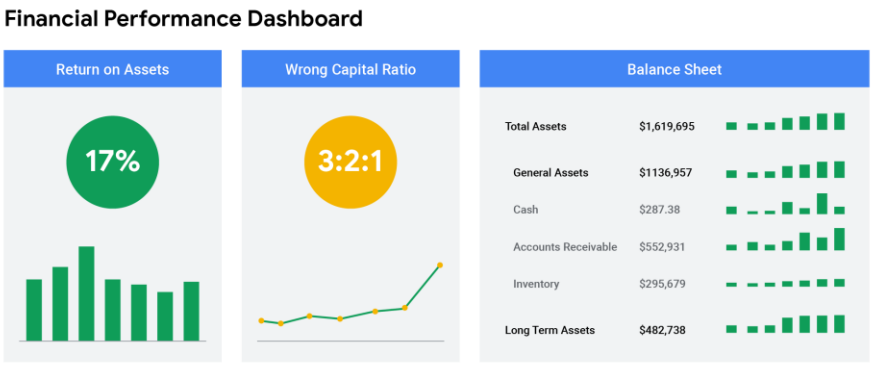


Operational: short-term performance tracking and intermediate goals -> Tracking operation process in light of stragic goals



Analytical: consists of the datasets and the mathematics used in these sets. -> Generally shared inside the team (not outside people (like upper manager) friendly to look)

➔ The analytic dashboard skips a specific timeframe. Instead, it identifies and tracks the various KPIs that may be used to assess strategic and operational goals.



Data and metrics

Data → metrics to reveal insights

Big and small data

	Small data	Big data
Period	Short	Long
Tool	Spreadsheets	Database and queried
Used by	Small – Mid businesses	Large org.
Size	Managable	Needs to be broken down for better analysis

4 V-words for big data: Volume – Variety (kinds of data) – Velocity (How fast data can be processed) – Veracity (quality and reliability of the data)

Spreadsheet

Spreadsheet task:

- Organize data
    - Pivot table

    ➔ Sort and filter data
- Calculate data

Spreadsheets and Data life cycle

- **Plan:** by developing organizational standards
- **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.
- **Manage** different kinds of data with a spreadsheet. This can involve storing, organizing, filtering, and updating information, 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.

Commented [2HL3]: •(formatting your cells, the headings you choose to highlight, the color scheme, and the way you order your data points, etc.)

Spreadsheets error and fixes

Error	Description	Example
#DIV/0!	A formula is trying to divide a value in a cell by 0 (or an empty cell with no value)	=B2/B3, when the cell B3 contains the value 0
#ERROR!	(Google Sheets only) Something can't be interpreted as it has been input. This is also known as a parsing error.	=COUNT(B1:D1 C1:C10) is invalid because the cell ranges aren't separated by a comma
#N/A	A formula can't find the data	The cell being referenced can't be found
#NAME?	The name of a formula or function used isn't recognized	The name of a function is misspelled
#NUM!	The spreadsheet can't perform a formula calculation because a cell has an invalid numeric value	=DATEDIF(A4, B4, "M") is unable to calculate the number of months between two dates because the date in cell A4 falls after the date in cell B4
#REF!	A formula is referencing a cell that isn't valid	A cell used in a formula was in a column that was deleted
#VALUE!	A general error indicating a problem with a formula or with referenced cells	There could be problems with spaces or text, or with referenced cells in a formula; you may have additional work to find the source of the problem.

Spotting errors in spreadsheets with conditional formatting

1. Click the gray triangle above row number 1 and to the left of Column A to select all cells in the spreadsheet.
2. From the main menu, click **Home**, and then click **Conditional Formatting** to select **Highlight Cell Rules > More Rules**.
3. For Select a Rule Type, choose **Use a formula to determine which cells to format**.
4. For Format values where this formula is true, enter **=ISERROR(A1)**.
5. Click the **Format** button, select the Fill tab, select yellow (or any other color), and then click **OK**.
6. Click **OK** to close the format rule window.
   To remove conditional formatting, click Home and select Conditional Formatting, and then click Manage Rules. Locate the format rule in the list, click Delete Rule, and then click OK.

Structured thinking

To avoid the possibility of resolving the problem again because of incomplete data or identifying the wrong problem => important to understand the problem domain

- ⇨ Use structured thinking
- ⇨ Identify the Scope of work (SOW) : an agreed-upon outline of the work you are going to perform on a project

Scope of work (SOW)

- **Deliverables:** Items/ tasks you will complete before finishing a 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:** Time due of tasks to be done
- **Milestones:** are **significant tasks** you will **confirm** along your timeline to help everyone know the project is on track.
- **Reports:** notify everyone as you finalize deliverables and meet milestones.
  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?
  - ⇨ Think about the project requirements

Check the Scope of work template here.

The important of context

Context is the condition in which something exists or happens. Context can turn raw data into meaningful information by defining it.

Commented [2HL4]: It is different to a statement of work. A statement of work is a document that clearly identifies the products and services a vendor or contractor will provide to an organization. It includes objectives, guidelines, deliverables, schedule, and costs.
A scope of work is project-based and sets the expectations and boundaries of a project. A scope of work may be included in a statement of work to help define project outcomes.
As a junior data analyst, It's more typical to be asked to create a scope of work than a statement of work.

Commented [2HL5]: Avoid vague statements

Communication

Working with stakeholder

Stakeholder: the people who invested time, interest and resources into the project you working on.

3 common types: the executive team, the customer-facing team, and the data science team.

- **Executive team**: the person provide strategic and operational leadership to the company (CEO, chief, etc.)
  They don't have much time, so keep the presentation mostly on the headline and straight to the answer. Other details can be included in appendix/ documentation so that they can read if they have time.
  Person to keep updated and ask for guidance: your project manager
- **Customer-facing team**: anyone who interacts with customer/ potential customer
  They might have specific asks, but you should focus on the story the data actually tell, not what these stakeholders hope to find
- **Data science team**: A big part of your job will be collaborating with other data team members to find new angles of the data to explore.

How to work effectively with stakeholders:

- **Discuss goals**
- **Feel empowered to say "no."** Stakeholders don't always realize the time and effort that goes into collecting and analyzing data, or even what they actually need. If you can't, have the confidence to say "no," and provide a respectful explanation. If there's an option that would be more helpful, point the stakeholder toward those resources.
- **Plan for the unexpected**. Before you start a project, make a list of potential roadblocks. Give yourself some extra time for problem-solving at each stage of the process.
- **Know your project**. Keep track of your discussions about the project over email or reports, and be ready to answer questions about how certain aspects are important for your organization.
- **Start with words and visuals**. It is common for data analysts and stakeholders to interpret things in different ways while assuming the other is on the same page. To help avoid this, start with a description and a quick visual of what you are trying to convey.
- **Communicate often**. Share notes about project milestones, setbacks, and changes. Then use your notes to create a shareable report or a change-log.

Balancing expectations and realistic goals. From the start, share a high-level schedule with stakeholders so they can plan accordingly. If there are problems, clear communication is vital so that the stakeholders can change the schedule accordingly. State at the start that the data might not 100% cover all the story (or your theory), and also can lead to another story.

Eg: “I have a theory that people are not finding the information that they need in order to take action. Here are the proven points that I have that support that theory.”

A data analyst reframes a question. Then, they outline the problem, challenges, potential solutions, and timeframe to the stakeholders to make sure the problem is clearly understood, and they can solve the comprehensive problem.

Focus on what matters

Working with different teams of stakeholders can be chaotic. Therefore, it is important to focus on the objective and not be distracted by answering questions

- Who are the primary and secondary stakeholders
- Who’s managing the data: to find the right data instead of finding the data needed yourself (in case there are available)
- Where you can go for help

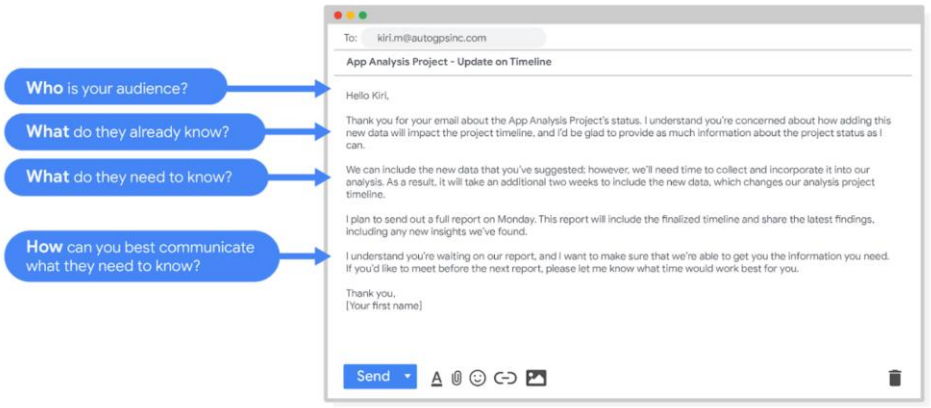
Commented [2HL6]: Normally the vice head of the department team

Commented [2HL7]: 2nd stakeholder are also typically responsible for managing the data

Blance team and stakeholder needs

Clear communication by answer these questions before communicating:

- Who your target audience are?
- What they already know?
- What they need to know?
- How you can communicate effectively about what they need to know?



Tips:

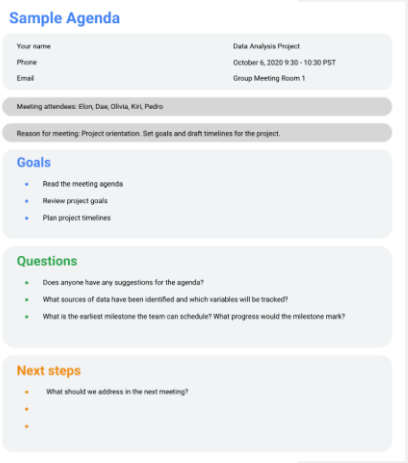
- Learn as you go and don’t afraid to ask

- Email tips: Clear, straight to the point and polite. Use headlines and bolder if necessary. If there are too many information -> meetings
- When first join a team knowing exactly where to start and what you need to do. Make sure the goal is clear and objective

Leading a meeting

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. This might 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

- 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)

A final word about meetings

Even with the most careful planning and detailed agendas, meetings can sometimes go off track. An emergency situation might steal people's attention. A recent decision might unexpectedly

change requirements that were previously discussed and agreed on. Action items might not apply to the current situation. If this happens, you might be forced to shorten or cancel your meeting. That's all right; just be sure to discuss anything that impacts your project with your manager or stakeholders and reschedule your meeting after you have more information.

Collaborate and conflict

To shift a situation from problematic to productive, data analysts can reframe a problem and start a constructive conversation with a cool head. Try to think about the group's goals and start the conversation with a friendly invitation to help so that you can understand better the stakeholders' needs.

Limitations of data

Data is an extremely powerful tool for decision-making, but if it is incomplete, misaligned, or hasn't been cleaned, then it can be misleading. Clean the data before you begin your analysis.

Imcomplete (or non exist) data. You can still use the data, but you will need to make the limits of your analysis clear.

Misaligned data. One team might define and measure things in a completely different way than another. -> Establishing how to measure things early on standardizes the data across the board for greater reliability and accuracy.

Dirty data. (data that contain errors) -> Clean the data

Tips:

- Compare the same types of data: Be sure to compare the same types of data and double-check that any segments in your chart definitely display different metrics.
- Visualize with care: A 0.01% drop in a score can look huge if you zoom in close enough. To make sure your audience sees the full story clearly, it is a good idea to set your Y-axis to 0.
- Leave out needless graphs: If a table can show your story at a glance, stick with the table instead of a pie chart or a graph.
- Test for statistical significance: Sometimes two datasets will look different, but you will need a way to test whether the difference is real and important. So remember to run statistical tests to see how much confidence you can place in that difference.
- Pay attention to sample size: Gather lots of data. If a sample size is small, a few unusual responses can skew the results. If you find that you have too little data, be careful about using it to form judgments. Look for opportunities to collect more data, then chart those trends over longer periods.
- After done, ask: Does my analysis answer the original questions? Are there any angles I haven't considered?