

Foundations: Data, Data Everywhere

★ Data Analysts Data-Driven Decision-Making Process:

1. **Ask** questions and define the problem.
2. **Prepare** data by collecting and storing the information.
3. **Process** data by cleaning and checking the information.
4. **Analyze** data to find patterns, relationships, and trends.
5. **Share** data with your audience.
6. **Act** on the data and use the analysis results.

Analytical thinking

Identifying and defining a problem and then solving it by using data in an organized, step-by-step manner

5 aspects of analytical thinking

1. Visualization
2. Strategy
3. Problem-orientation
4. Correlation
5. Big-picture and detail-oriented thinking

★ Some Common Questions Analysts Ask:

What is the root cause of the problem?

Root cause

The reason why a problem occurs

Ask, “**why?**” five times to reveal the root cause

Where are the gaps in our process?

Gap analysis

A method for examining and evaluating how a process works currently in order to get where you want to be in the future

What did we not consider before?

★ Examples of 5 Whys in Action

An online grocery store was receiving numerous customer service complaints about poor deliveries. To address this problem, a data analyst at the company asked their first “why?”

Why #1. “Customers are complaining about poor grocery deliveries. Why?”

The data analyst began by reviewing the customer feedback more closely. They noted the vast majority of complaints dealt with products arriving damaged. So, they asked “why?” again.

Why #2. “Products are arriving damaged. Why?”

To answer this question, the data analyst continued exploring the customer feedback. It turned out that many customers said products were not packaged properly.

Why #3. “Products are not packaged properly. Why?”

After asking their third “why,” the data analyst did some further detective work. They ultimately learned that their company’s grocery packers were not adequately trained on packing procedures.

Why #4. “Grocery packers are not adequately trained. Why?”

This “why” enabled the data analyst to uncover that nearly 35% of all packers were new to the company. They had not yet had the chance to complete all required training, yet they were already being asked to pack groceries for customer orders.

Why #5. “Packers have not completed required training. Why?”

This final “why?” led the data analyst to find out that the human resources department had not provided necessary training to any newly hired packers. This was because HR was in the middle of reworking the training program. Rather than training new hires using the old system, they had provided them with a quick one-page guide, which was insufficient.

So, in this example, the root cause of the problem was that HR had not completed the training program updates and was using a less-thorough guide to train new packers. Fortunately, this was a problem that the grocer could control. And thanks to the data analyst’s work, they provided more support to the HR department to complete the training and retrain all newly hired grocery packers!

Advance quality control

An irrigation company was experiencing an increase in the number of defects in their water pumps. The company's data team used the five whys to analyze the situation:

Why #1. "There has been an increase in the number of defects in water pumps. Why?"

To answer this question, the data team set up a meeting with shop floor engineers. They asked for some insights into machine performance and manufacturing processes. After some exploration, it was discovered that the machines used to produce the pumps were not properly calibrated.

Why #2. "The machines are not properly calibrated. Why?"

After more brainstorming with the engineering team, it was determined that the machines were miscalibrated during the last maintenance cycle.

Why #3. "The machines were miscalibrated during maintenance. Why?"

Next, the data team investigated the procedures involved with machine calibration. They found out that the current method was inappropriate for the machines.

Why #4. "The calibration method is inappropriate for the machines. Why?"

This "why" led them to discover that the company had recently installed new software in their machines. Because it was a minor software upgrade, the engineers didn't realize it would affect calibration. They didn't have the information they needed to properly calibrate the upgraded machines.

Why #5. "The engineers don't have the information they need to calibrate the upgraded machines. Why?"

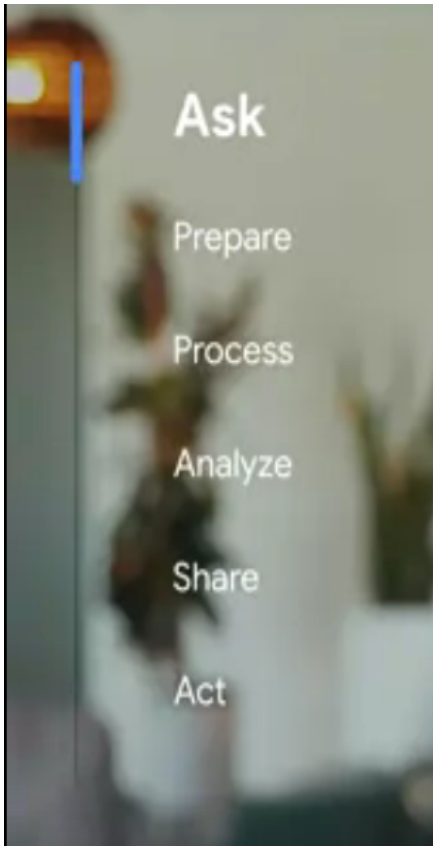
The fifth and final "why" turned up even more evidence: The installation team had upgraded machine software, but had failed to share the corresponding calibration procedures with the engineers.

So, in this example, the root cause of the problem was that the engineers lacked important information about how to calibrate the machines using the new software system. The solution was found, and the irrigation company was able to implement it right away. Soon, the engineers had the necessary calibration instructions, and the pump defects were eliminated!

★ Data Life Cycle:

1. **Plan:** Decide what kind of data is needed, how it will be managed, and who will be responsible for it.
2. **Capture:** Collect or bring in data from a variety of different sources.
3. **Manage:** Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.
4. **Analyze:** Use the data to solve problems, make decisions, and support business goals.
5. **Archive:** Keep relevant data stored for long-term and future reference.
6. **Destroy:** Remove data from storage and delete any shared copies of the data.

★ Phases of Data Analysis:



★ Some Common Data Analysts Tools:

- Spreadsheets
- Query languages for databases
- Visualization tools

Decoding the job description



	Data Analysts	Data Scientists	Data Specialists
Problem solving	Use existing tools and methods to solve problems with existing types of data	Invent new tools and models, ask open-ended questions, and collect new types of data	Use in-depth knowledge of databases as a tool to solve problems and manage data
Analysis	Analyze collected data to help stakeholders make better decisions	Analyze and interpret complex data to make business predictions	Organize large volumes of data for use in data analytics or business operations
Other relevant skills	<ul style="list-style-type: none">• Database queries• Data visualization• Dashboards• Reports• Spreadsheets	<ul style="list-style-type: none">• Advanced statistics• Machine learning• Deep learning• Data optimization• Programming	<ul style="list-style-type: none">• Data manipulation• Information security• Data models• Scalability of data• Disaster recovery