Origins of the data analysis process

When you decided to join this program, you proved that you are a curious person. So now, tap into your curiosity and explore the origins of data analysis. No one knows when or why the first person decided to record data about people and things. But it was certainly a smart idea!

Data analysis is rooted in statistics, which has a pretty long history itself. Archaeologists mark the start of statistics in ancient Egypt with the building of the pyramids. The ancient Egyptians were masters of organizing data. They documented their calculations and theories on papyri (paper-like materials), which are now viewed as the earliest examples of spreadsheets and checklists. Today's data analysts owe a lot to those brilliant scribes, who helped create a more technical and efficient process.

It is time to enter the **data analysis process**—how industry professionals move from data to decision. All team members can drive success by planning work both upfront and at the end of the data analysis process. While the data analysis process is well known among experts, there isn't a single defined structure or single architecture that's uniformly followed by every data analyst. But there are some shared fundamentals. This reading provides an overview of several processes, starting with the one that forms the foundation of the Google Data Analytics Certificate.

The process presented as part of the Google Data Analytics Certificate is one that will be valuable to you as you keep moving forward in your career:

1. **Ask**: business challenge, objective, or question
2. **Prepare**: data generation, collection, storage, and data management
3. **Process**: data cleaning and data integrity
4. **Analyze**: data exploration, visualization, and analysis
5. **Share**: communicating and interpreting results
6. **Act**: putting insights to work to solve the problem

Understanding this process—and all of the iterations that helped make it popular—will be a big part of guiding your own analysis and your work in this program. Let's go over a few other variations of the data analysis process.

EMC's data analysis process

EMC Corporation's data analytics process is cyclical with six steps:
1. Discovery
2. Pre-processing data
3. Model planning
4. Model building
5. Communicate results
6. Operationalize

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of typical business projects. The phases aren't static milestones; each step connects and leads to the next, and eventually repeats. Key questions help analysts test whether they have accomplished enough to move forward and ensure that teams have spent enough time on each of the phases and don't start modeling before the data is ready. It is a little different from the data analysis process on which this program is based on, but it has some core ideas in common: the first phase is interested in discovering and asking questions; data has to be prepared before it can be analyzed and used; and then findings should be shared and acted on.

For more information, refer to this e-book, [Data Science & Big Data Analytics](#). Select **PDF** under the section that reads **Free Access**.

🔒 **Free Access**

Front Matter (Pages: i-xviii)

Summary | **PDF** | **Request permissions**

SAS's iterative process

An iterative data analysis process was created by a company called **SAS**, a leading data analytics solutions provider. It can be used to produce repeatable, reliable, and predictive results:
1. Ask
2. Prepare
3. Explore
4. Model
5. Implement
6. Act
7. Evaluate

The SAS model emphasizes the cyclical nature of their model by visualizing it as an infinity symbol. Its process has seven steps, many of which mirror the other models, like ask, prepare,

model, and act. But this process is also a little different; it includes a step after the act phase designed to help analysts evaluate their solutions and potentially return to the ask phase again.

Project-based data analytics process

A project-based data analytics process has five simple steps:
1. Identifying the problem
2. Designing data requirements
3. Pre-processing data
4. Performing data analysis
5. Visualizing data

This data analytics project process was developed by Vignesh Prajapati. It doesn't include the sixth phase, or the act phase. However, it still covers a lot of the same steps described. It begins with identifying the problem, preparing and processing data before analysis, and ends with data visualization.

For more information, refer to [Understanding the data analytics project life cycle](#).

Big data analytics process

Authors Thomas Erl, Wajid Khattak, and Paul Buhler proposed a big data analytics process in their book, **Big Data Fundamentals: Concepts, Drivers & Techniques**. Their process suggests phases divided into nine steps:
1. Business case evaluation
2. Data identification
3. Data acquisition and filtering
4. Data extraction
5. Data validation and cleaning
6. Data aggregation and representation
7. Data analysis
8. Data visualization
9. Utilization of analysis results

This process appears to have three or four more steps than the previous models. But in reality, they have just broken down what has been referred to as prepare and process into smaller steps. It emphasizes the individual tasks required for gathering, preparing, and cleaning data before the analysis phase.

For more information, refer to [Big Data Adoption and Planning Considerations](#).

Key takeaway

From a journey to the pyramids and data in ancient Egypt to now, the way people analyze data has evolved (and continues to do so). The data analysis process is like real life architecture: There are different ways to do things but the same core ideas still appear in each model of the

process. Whether you use the structure of this Google Data Analytics Certificate or one of the many other iterations you have learned about, your approach will be effective.

4. An e-commerce website collects, observes, and analyzes its customers' online behaviors. Then, it uses the insights gained to choose when to put certain products on sale. What business practice does this describe?

○ Analytical thinking

○ Organizing data

○ Performance measurement

◉ Data-driven decision-making

✓ **Correct**
An e-commerce website collecting, observing, and analyzing its customers' online behaviors, then using the insights gained to choose when to put certain products on sale, describes data-driven decision making. Data-driven decision-making is using facts to guide business strategy.

1. In your role as a data professional, you examine a dataset of taxi rides to determine which hour of the day typically has the highest demand. Under which discipline does this example *best* fall?

○ Data wrangling

◉ Data analysis

○ Data analytics

○ Data science

✓ **Correct**

1. Fill in the blank: Analytical skills are the qualities and characteristics associated with using _____ to solve problems.

- ○ predictions
- ● facts
- ○ gut instinct
- ○ feedback

> ⊘ **Correct**
> Analytical skills are the qualities and characteristics associated with using facts to solve problems.

2. An acquaintance tells you that they spend many hours each day "playing." To learn more, you ask them whether they play sports, a musical instrument, or something else. Their answer helps you clarify the meaning of their statement. What does this scenario describe?

- ○ Analyzing data
- ● Understanding context
- ○ Making assumptions
- ○ Testing a hypothesis

> ⊘ **Correct**
> This scenario describes understanding context. Context is the condition in which something exists or happens.

3. Which analytical skill involves the ability to break things down into smaller steps and work with them in an orderly and logical way?

- ○ Curiosity
- ● Technical mindset
- ○ Problem-solving
- ○ Simplification

> ⊘ **Correct**
> A technical mindset involves breaking things down into smaller steps and working with them in an orderly and logical way.

4. Fill in the blank: The analytical skill _____ encompasses how someone organizes information.

- ○ data mindset
- ● data design
- ○ data strategy
- ○ data analysis

> ⊘ **Correct**
> The analytical skill data design encompasses how someone organizes information.

Use the five whys for root cause analysis

Recently, you've been learning why business solutions almost always require some data detective work. This is one way critical thinking helps data professionals determine the right questions to ask in order to arrive at those solutions. One very common question is, "What is the root cause of the problem?" A **root cause** is the reason why a problem occurs. So, by identifying and eliminating the root cause, data professionals can help stop that problem from occurring again.

The **five whys** is a simple but effective technique for identifying a root cause. It involves asking "Why?" repeatedly until the answer reveals itself. This often happens at the fifth "why," but sometimes you'll need to continue asking more times, sometimes fewer.

You recently explored a case involving lacking the necessary ingredients to bake pies; now, you'll go more in-depth with some business applications of the five whys technique to do root cause analysis.

Boost customer service

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!

Key takeaways

The five whys is a powerful tool for root cause analysis. It's simple, effective, and a great way to collaborate with colleagues and learn about other areas of the business. Plus, the five whys can be used to analyze problems in any industry, helping organizations of all kinds identify and fix business problems. As a data professional, you can turn to the five whys whenever you feel stumped by a problem and need to approach it from a different perspective.

Mark as completed

Like

Dislike

Report an issue

1. Fill in the blank: Analytical thinking involves _____ a problem, then solving it by using data in an organized, step-by-step manner.

   ○ evaluating and summarizing

   ○ monitoring and assessing

   ○ observing and inspecting

   ● identifying and defining

   ⊘ **Correct**
   Analytical thinking involves identifying and defining a problem, then solving it by using data in an organized, step-by-step manner.

2. What type of data visualization might an analyst create in order to communicate data insights to others? Select all that apply.

   ☑ Map

   ⊘ **Correct**
   An analyst might create a graph, map, or chart in order to communicate data insights to others.

   ☑ Chart

   ⊘ **Correct**
   An analyst might create a graph, map, or chart in order to communicate data insights to others.

   ☐ Report

   ☑ Graph

   ⊘ **Correct**
   An analyst might create a graph, map, or chart in order to communicate data insights to others.

3. While planning a roadtrip, you figure out all of the specific stops you need to take along the way. You also consider how often you'll stop for gas, meals, and sleep. Having this information enables you to execute your plan. What does this scenario describe?

○ Big-picture thinking

○ Problem-orientation

○ Finding a correlation

◉ Detail-oriented thinking

⊘ **Correct**
This scenario describes detail-oriented thinking, which is about figuring out all of the specifics that will help you execute a plan.

4. What is a method for examining and evaluating how a process works currently in order to get to an improved future state?

◉ Data-driven decision-making

○ Gap analysis

○ Root cause analysis

○ Five whys

⊗ **Incorrect**
Gap analysis is a method for examining and evaluating how a process works currently in order to get to an improved future state. Data-driven decision-making is using facts to guide business strategy.

1. Which of the following statements correctly describe data and data analysis? Select all that apply.

☑ Data is a collection of facts.

☑ Data analysis involves collecting, transforming, and organizing data.

☐ Data analysis is creating new ways of understanding the unknown using raw data.

☑ One goal of data analysis is to draw conclusions.

2. Fill in the blank: Data science is creating new ways of _____ and understanding the unknown by using raw data.

◉ identifying

○ strategizing

○ modeling

○ arranging

3. Which of the following activities are elements of data-driven decision-making? Select all that apply.

☐ Remove data insights that stem from human intuition

☑ Find and analyze relevant data

☑ Ask subject-matter experts to review the results

☐ Figure out the business need or problem to be solved

4. A business wants to grow its sales by 10% in the next 12 months. The data team helps achieve this goal by collecting data about current sales, their target market, and their competitors. This helps them identify the difference between where they are today and where they want to be in a year. What does this scenario describe?

○ Guiding business decisions

◉ Gap analysis

○ Future analysis

○ Planning

5. Fill in the blank: Data analysts use a problem-oriented approach in order to _____, describe, and solve problems.

○ design

○ revise

◉ identify

○ customize

6. Which of the following statements correctly describe analytical thinking? Select all that apply.

☐ Just because two pieces of data trend in the same direction does not necessarily mean they are related.

☑ Analytical thinking involves using data in an organized manner in order to modify or make adjustments to a problem.

☐ Visualization is the graphical representation of information.

☑ Figuring out all of the specifics that will help execute a plan is detail-oriented thinking.

7. A junior data analyst at a construction company develops a plan about a home-remodeling project. They employ analytical thinking to stay focused and on track. They also consider how to improve the quality and usefulness of the data they collect. Which aspect of analytical thinking does this scenario describe?

   ○ Problem-orientation

   ○ Correlation

   ● Strategic thinking

   ○ VIsualization

8. What is the purpose of the five whys?

   ○ To understand the impact of a problem

   ○ To figure out a problem's consequences

   ○ To identify key symptoms of a problem

   ● To reveal the root cause of a problem

9. A data professional is assigned a challenging data analysis task. In order to be able to work in an orderly and logical manner, they break down larger, more complex elements into smaller pieces. Which analytical skill does this scenario describe?

   ○ Data design

   ○ Curiosity

   ● Technical mindset

   ○ Understanding context

10. Which of the following examples demonstrate data-driven decision-making? Select all that apply.

   ☑ An online retailer surveys customers to develop new products that are more likely to be successful.

   ☑ A weather forecaster refers to information about past extreme weather in order to better predict future events.

   ☑ A government agency uses facts documented in police reports to help develop crime-prevention strategies.

   ☐ A transportation company prioritizes the preferences of local politicians when outlining routes and schedules.