# **Your Google Data Analytics Certificate roadmap**

Use this guide to review the topics covered, tools used, and skills you will use in each course.

## Foundations:

|  |  |
| --- | --- |
| What you will learn:   * Real-life roles and responsibilities of a junior data analyst * How businesses transform data into actionable insights * Spreadsheet basics * Database and query basics * Data visualization basics |  |
| Skill sets you will build:   * Using data in everyday life * Thinking analytically * Applying tools from the data analytics toolkit * Showing trends and patterns with data visualizations * Ensuring your data analysis is fair |  |

## Ask:

|  |  |
| --- | --- |
| What you will learn:   * How data analysts solve problems with data * The use of analytics for making data-driven decisions * Spreadsheet formulas and functions * Dashboard basics, including an introduction to *Tableau*. * Data reporting basics |  |
| Skill sets you will build:   * Asking SMART and effective questions * Structuring how you think * Summarizing data * Putting things into context * Managing team and stakeholder expectations * Problem-solving and conflict-resolution |  |

## Prepare:

|  |  |
| --- | --- |
| What you will learn:   * How data is generated * Features of different data types, fields, and values * Database structures * The function of metadata in data analytics * Structured Query Language (SQL) functions |  |
| Skill sets you will build:   * Ensuring ethical data analysis practices * Addressing issues of bias and credibility * Accessing databases and importing data * Writing simple queries * Organizing and protecting data * Connecting with the data community (optional) |  |

## Process:

|  |  |
| --- | --- |
| What you will learn:   * Data integrity and the importance of clean data * The tools and processes used by data analysts to clean data. * Data-cleaning verification and reports * Statistics, hypothesis testing, and margin of error * Resume building and interpretation of job postings (optional) |  |
| Skill sets you will build:   * Connecting business objectives to data analysis * Identifying clean and dirty data * Cleaning small datasets using spreadsheet tools. * Cleaning large datasets by writing SQL queries. * Documenting data-cleaning processes |  |

## Analyze:

|  |  |
| --- | --- |
| What you will learn:   * Steps data analysts take to organize data. * How to combine data from multiple sources * Spreadsheet calculations and pivot tables * SQL calculations * Temporary tables * Data validation |  |
| Skill sets you will build:   * Sorting data in spreadsheets and by writing SQL queries. * Filtering data in spreadsheets and by writing SQL queries. * Converting data * Formatting data * Substantiating data analysis processes * Seeking feedback and support from others during data analysis |  |

## Share:

|  |  |
| --- | --- |
| What you will learn:   * Design thinking * How data analysts use visualizations to communicate about data * The benefits of Tableau for presenting data analysis findings * Data-driven storytelling * Dashboards and dashboard filters * Strategies for creating an effective data presentation |  |
| Skill sets you will build:   * Creating visualizations and dashboards in Tableau * Addressing accessibility issues when communicating about data * Understanding the purpose of different business communication tools * Telling a data-driven story * Presenting to others about data * Answering questions about data |  |

## Act:

|  |  |
| --- | --- |
| What you will learn:   * Programming languages and environments * R packages * R functions, variables, data types, pipes, and vectors * R data frames * Bias and credibility in R * R visualization tools * R Markdown for documentation, creating structure, and emphasis |  |
| Skill sets you will build:   * Coding in R * Writing functions in R * Accessing data in R * Cleaning data in R * Generating data visualizations in R * Reporting on data analysis to stakeholders |  |

## Capstone:

|  |  |
| --- | --- |
| What you will learn:   * How a data analytics portfolio distinguishes you from other candidates * Practical, real-world problem-solving * Strategies for extracting insights from data * Clear presentation of data findings * Motivation and ability to take initiative |  |
| Skill sets you will build:   * Building a portfolio * Increasing your employability * Showcasing your data analytics knowledge, skill, and technical expertise * Sharing your work during an interview * Communicating your unique value proposition to a potential employer |  |

# **Foundations: Data, Data, Everywhere**

This is the first course in the Google Data Analytics Certificate. These courses will equip you with the skills you need to apply to introductory-level data analyst jobs. Organizations of all kinds need data analysts to help them improve their processes, identify opportunities and trends, launch new products, and make thoughtful decisions. In this course, you’ll be introduced to the world of data analytics through hands-on curriculum developed by Google. The material shared covers plenty of key data analytics topics, and it’s designed to give you an overview of what’s to come in the Google Data Analytics Certificate. Current Google data analysts will instruct and provide you with hands-on ways to accomplish common data analyst tasks with the best tools and resources.

Learners who complete this certificate program will be equipped to apply for introductory-level jobs as data analysts. No previous experience is necessary.

By the end of this course, you will:

* Gain an understanding of the practices and processes used by a junior or associate data analyst in their day-to-day job.
* Learn about key analytical skills (data cleaning, data analysis, data visualization) and tools (spreadsheets, SQL, R programming, Tableau) that you can add to your professional toolbox.
* Discover a wide variety of terms and concepts relevant to the role of a junior data analyst, such as the data life cycle and the data analysis process.
* Evaluate the role of analytics in the data ecosystem.
* Conduct an analytical thinking self-assessment.
* Explore job opportunities available to you upon program completion, and learn about best practices in the job search.

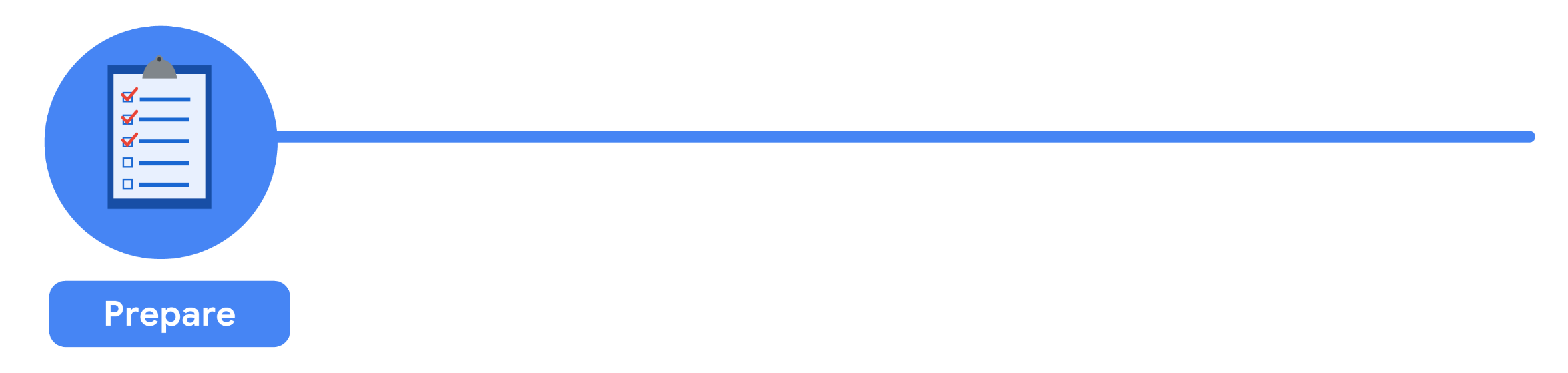
## Case Study: New Data Perspectives

|  |  |
| --- | --- |
| As you have been learning, you can find data pretty much everywhere. Any time you observe and evaluate something in the world, you’re collecting and analyzing data. Your analysis helps you find easier ways of doing things, identify patterns to save you time, and discover surprising new perspectives that can completely change the way you experience things.  Here is a real-life example of how one group of data analysts used the six steps of the data analysis process to improve their workplace and its business processes. Their story involves something called people analytics - also known as human resources analytics or workforce analytics. People analytics is the practice of collecting and analyzing data on the people who make up a company’s workforce in order to gain insights to improve how the company operates.  Being a people analyst involves using data analysis to gain insights about employees and how they experience their work lives. The insights are used to define and create a more productive and empowering workplace. This can unlock employee potential, motivate people to perform at their best, and ensure a fair and inclusive company culture.  The six steps of the data analysis process that you have been learning in this program are: **ask, prepare, process, analyze, share,** and **act**. These six steps apply to any data analysis. Continue reading to learn how a team of people analysts used these six steps to answer a business question.  An organization was experiencing a high turnover rate among new hires. Many employees left the company before the end of their first year on the job. The analysts used the data analysis process to answer the following question: **how can the organization improve the retention rate for new employees?** |  |

Here is a breakdown of what this team did, step by step.



|  |  |
| --- | --- |
| First up, the analysts needed to define what the project would look like and what would qualify as a successful result. So, to determine these things, they asked effective questions and collaborated with leaders and managers who were interested in the outcome of their people analysis. These were the kinds of questions they asked:   * What do you think new employees need to learn to be successful in their first year on the job? * Have you gathered data from new employees before? If so, may we have access to the historical data? * Do you believe managers with higher retention rates offer new employees something extra or unique? * What do you suspect is a leading cause of dissatisfaction among new employees? * By what percentage would you like employee retention to increase in the next fiscal year? |  |



|  |  |
| --- | --- |
| It all started with solid preparation. The group built a timeline of three months and decided how they wanted to relay their progress to interested parties. Also, during this step, the analysts identified what data they needed to achieve the successful result they identified in the previous step - in this case, the analysts chose to gather the data from an online survey of new employees. These were the things they did to prepare:   * They developed specific questions to ask about employee satisfaction with different business processes, such as hiring and onboarding, and their overall compensation. * They established rules for who would have access to the data collected - in this case, anyone outside the group wouldn't have access to the raw data but could view summarized or aggregated data. For example, an individual's compensation wouldn't be available, but salary ranges for groups of individuals would be viewable. * They finalized what specific information would be gathered, and how best to present the data visually. The analysts brainstormed possible project- and data-related issues and how to avoid them. |  |

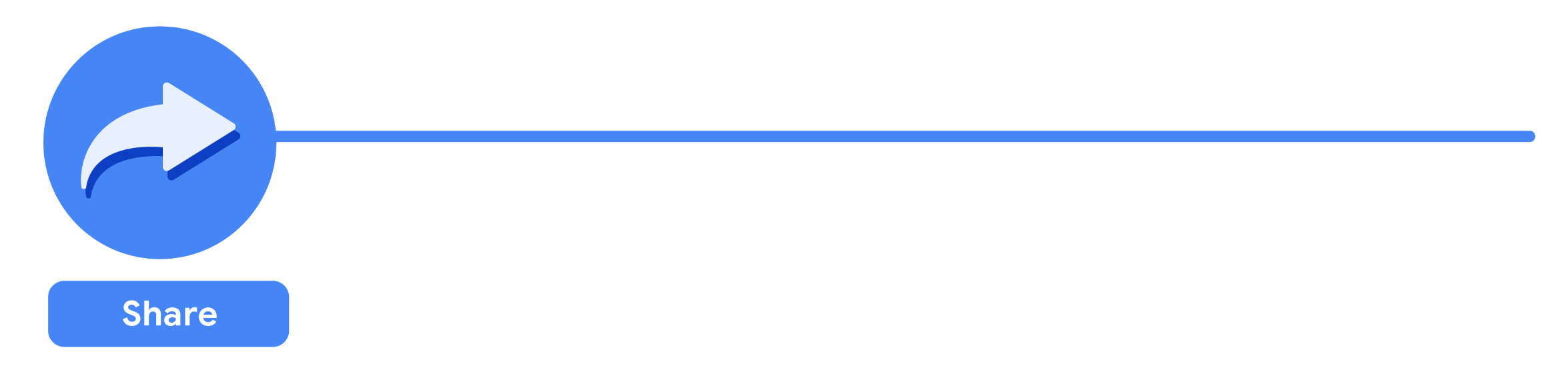


|  |  |
| --- | --- |
| The group sent the survey out. Great analysts know how to respect both their data and the people who provide it. Since employees provided the data, it was important to make sure all employees gave their consent to participate. The data analysts also made sure employees understood how their data would be collected, stored, managed, and protected. Collecting and using data ethically is one of the responsibilities of data analysts. In order to maintain confidentiality and protect and store the data effectively, these were the steps they took:   * They restricted access to the data to a limited number of analysts. * They cleaned the data to make sure it was complete, correct, and relevant. Certain data was aggregated and summarized without revealing individual responses. * They uploaded raw data to an internal data warehouse for an additional layer of security. |  |



|  |  |
| --- | --- |
| Then, the analysts did what they do best: analyze! From the completed surveys, the data analysts discovered that an employee’s experience with certain processes was a key indicator of overall job satisfaction. These were their findings:   * Employees who experienced a long and complicated hiring process were most likely to leave the company. * Employees who experienced an efficient and transparent evaluation and feedback process were most likely to remain with the company. |  |

The group knew it was important to document exactly what they found in the analysis, no matter what the results. To do otherwise would diminish trust in the survey process and reduce their ability to collect truthful data from employees in the future.



|  |  |
| --- | --- |
| Just as they made sure the data was carefully protected, the analysts were also careful sharing the report. This is how they shared their findings:   * They shared the report with managers who met or exceeded the minimum number of direct reports with submitted responses to the survey. * They presented the results to the managers to make sure they had the full picture. * They asked the managers to personally deliver the results to their teams. |  |

This process gave managers an opportunity to communicate the results with the right context. As a result, they could have productive team conversations about next steps to improve employee engagement.



|  |  |
| --- | --- |
| The last stage of the process for the team of analysts was to work with leaders within their company and decide how best to implement changes and take actions based on the findings. These were their recommendations:   * Standardize the hiring and evaluation process for employees based on the most efficient and transparent practices. * Conduct the same survey annually and compare results with those from the previous year. |  |

A year later, the same survey was distributed to employees. Analysts anticipated that a comparison between the two sets of results would indicate that the action plan worked. Turns out, the changes improved the retention rate for new employees and the actions taken by leaders were successful!

**Is people analytics right for you?**

One of the many things that makes data analytics so exciting is that the problems are always different, the solutions need creativity, and the impact on others can be great-even life-changing or lifesaving. As a data analyst, you can be part of these efforts. Maybe you’re even inspired to learn more about the field of people analytics. If so, consider learning more about this field and adding that research to your data analytics journal. You never know: One day soon, you could be helping a company create an amazing work environment for you and your colleagues!

**Additional Resource**

To learn more about some recent applications of data analytics in the business world, check out the article [“4 Examples of Business Analytics in Action”](https://online.hbs.edu/blog/post/business-analytics-examples) from Harvard Business School.  The article reveals how corporations use data insights to optimize their decision-making process. Please note that the first example in the article contains a minor error in the second paragraph, but the example is still a valid one.

## Data and Gut Instinct

*Dữ liệu và trực giác*

Detectives and data analysts have a lot in common. Both depend on facts and clues to make decisions. Both collect and look at the evidence. Both talk to people who know part of the story. And both might even follow some footprints to see where they lead. Whether you’re a detective or a data analyst, your job is all about following steps to collect and understand facts.

*Thám tử và nhà phân tích dữ liệu có rất nhiều điểm chung. Cả hai đều phụ thuộc vào sự thật và manh mối để đưa ra quyết định. Cả hai đều thu thập và xem xét các bằng chứng. Cả hai đều nói chuyện với những người biết một phần của câu chuyện. Và cả hai thậm chí có thể lần theo một số dấu chân để xem họ dẫn đến đâu. Cho dù bạn là thám tử hay nhà phân tích dữ liệu, công việc của bạn là làm theo các bước để thu thập và hiểu sự thật.*

Analysts use data-driven decision-making and follow a step-by-step process. You have learned that there are six steps to this process:

*Các nhà phân tích ra quyết định dựa trên dữ liệu và tuân theo quy trình từng bước. Có 6 bước cho quá trình này:*

* Ask questions and define the problem.

*Đặt câu hỏi và xác định vấn đề.*

* Prepare data by collecting and storing the information.

*Chuẩn bị dữ liệu bằng cách thu thập và lưu trữ thông tin.*

* Process data by cleaning and checking the information.

*Xử lý dữ liệu bằng cách làm sạch và kiểm tra thông tin.*

* Analyze data to find patterns, relationships, and trends.

*Phân tích dữ liệu để tìm các kiểu mẫu, mối quan hệ và xu hướng.*

* Share data with your audience.

*Chia sẻ dữ liệu với người xem.*

* Act on the data and use the analysis results.

*Hành động trên dữ liệu và sử dụng kết quả phân tích.*

But there are other factors that influence the decision-making process. You may have read mysteries where the detective used their gut instinct, and followed a hunch that helped them solve the case. Gut instinct is an intuitive understanding of something with little or no explanation. This isn’t always something conscious; we often pick up on signals without even realizing. You just have a “feeling” it’s right.

*Nhưng có những yếu tố khác ảnh hưởng đến quá trình ra quyết định. Có thể bạn đã đọc những câu chuyện trinh thám trong đó thám tử sử dụng trực giác của họ và làm theo linh cảm giúp họ giải quyết vụ án. Trực giác là sự hiểu biết về điều gì đó mà không cần giải thích hoặc cần ít lời giải thích. Điều này thường được coi là một loại hiểu biết không rõ nguồn gốc, đôi khi không thể giải thích bằng từ ngữ hoặc luận lý hợp lý; chúng ta thường nắm bắt được những tín hiệu mà thậm chí không nhận ra. Bạn chỉ cần có một “cảm giác” là đúng.*

### Why gut instinct can be a problem?

*Tại sao trực giác có thể là vấn đề?*

At the heart of data-driven decision making is data. Therefore, it's essential that data analysts focus on the data to ensure they make informed decisions. If you ignore data by preferring to make decisions based on your own experience, your decisions may be biased. But even worse, decisions based on gut instinct without any data to back them up can cause mistakes.

*Trọng tâm của việc ra quyết định dựa trên dữ liệu là dữ liệu. Do đó, điều cần thiết là các nhà phân tích dữ liệu tập trung vào dữ liệu để đảm bảo họ đưa ra quyết định sáng suốt. Nếu bạn bỏ qua dữ liệu bằng cách thích đưa ra quyết định dựa trên kinh nghiệm của riêng bạn, quyết định của bạn có thể bị sai lệch. Nhưng thậm chí tệ hơn, các quyết định dựa trên bản năng mà không có bất kỳ dữ liệu nào để sao lưu chúng có thể gây ra sai lầm.*

Consider an example of a restaurant entrepreneur, partnering with a well-known chef to develop a new restaurant in a bustling part of the city’s central shopping district. The well-known chef has several restaurants across the city. Banking on their reputation, the restaurant entrepreneur and chef followed gut instinct and created another uniquely themed restaurant. However, fundraising efforts fell short to fund the opening of the restaurant after months of planning and preparation. The property will go back on the market to be sold at a loss. Had the entrepreneur done more research, they would've found data showing prospective customers in this new restaurant location were very different from the chef's other restaurants.

*Hãy xem một ví dụ về một doanh nhân nhà hàng, hợp tác với một đầu bếp nổi tiếng để phát triển một nhà hàng mới tại một khu vực sầm uất của trung tâm khu mua sắm của thành phố. Đầu bếp nổi tiếng đã có một số nhà hàng trên khắp thành phố. Dựa trên danh tiếng của mình, doanh nhân nhà hàng và đầu bếp đã tuân theo trực giác và tạo ra một nhà hàng mới với chủ đề độc đáo. Tuy nhiên, công cuộc gây quỹ không đạt được mục tiêu để mở cửa nhà hàng sau nhiều tháng lập kế hoạch và chuẩn bị. Tài sản mà họ bỏ ra để tạo ra nhà hàng này sẽ được bán lại với giá thấp hơn. Nếu doanh nhân đã nghiên cứu kỹ hơn, họ sẽ tìm thấy dữ liệu cho thấy khách hàng tiềm năng tại vị trí mới của nhà hàng này rất khác biệt so với các nhà hàng khác của đầu bếp nổi tiếng.*

The more you understand the data related to a project, the easier it will be to figure out what is required. These efforts will also help you identify errors and gaps in your data so you can communicate your findings more effectively. Sometimes experience helps you make a connection that no one else would notice. For example, a detective might be able to crack open a case because they remember an old case just like the one, they’re solving today. It's not just gut instinct.

*Bạn càng hiểu nhiều dữ liệu liên quan đến một dự án, bạn càng dễ dàng tìm ra những gì được yêu cầu. Những nỗ lực này cũng sẽ giúp bạn xác định lỗi và lỗ hổng trong dữ liệu của mình để bạn có thể truyền đạt những phát hiện của mình hiệu quả hơn. Đôi khi kinh nghiệm giúp bạn tạo ra một kết nối mà không ai khác sẽ nhận thấy. Ví dụ, một thám tử có thể bẻ khóa một vụ án vì họ nhớ một vụ án cũ giống như vụ án mà họ đang giải quyết ngày hôm nay. Nó không chỉ là bản năng.*

### Data + business knowledge = mystery solved

*Dữ liệu + kiến thức kinh doanh = vấn đề được giải quyết*

Blending data with business knowledge, plus maybe a touch of gut instinct, will be a common part of your process as a junior data analyst. The key is figuring out the exact mix for each particular project. A lot of times, it will depend on the goals of your analysis. That is why analysts often ask, “How do I define success for this project?”

*Pha trộn dữ liệu với kiến thức kinh doanh, cộng với có thể là một chút bản năng, sẽ là một phần phổ biến trong quy trình của bạn với tư cách là một nhà phân tích dữ liệu đã có ít kinh nghiệm. Điều quan trọng là tìm ra sự kết hợp chính xác cho từng dự án cụ thể. Rất nhiều lần, nó sẽ phụ thuộc vào mục tiêu phân tích của bạn. Đó là lý do tại sao các nhà phân tích thường hỏi, "Làm thế nào để tôi xác định thành công cho dự án này?"*

In addition, try asking yourself these questions about a project to help find the perfect balance:

*Ngoài ra, hãy thử tự hỏi mình những câu hỏi này về một dự án để giúp tìm ra sự cân bằng hoàn hảo:*

* What kind of results are needed?

*Những loại kết quả là cần thiết?*

* Who will be informed?

*Ai sẽ được thông báo?*

* Am I answering the question being asked?

*Tôi có đang trả lời câu hỏi đang được hỏi không?*

* How quickly does a decision need to be made?

*Quyết định cần được đưa ra nhanh như thế nào?*

For instance, if you are working on a rush project, you might need to rely on your own knowledge and experience more than usual. There just isn’t enough time to thoroughly analyze all of the available data. But if you get a project that involves plenty of time and resources, then the best strategy is to be more data-driven. It’s up to you, the data analyst, to make the best possible choice. You will probably blend data and knowledge a million different ways over the course of your data analytics career. And the more you practice, the better you will get at finding that perfect blend.

*Ví dụ, nếu bạn đang làm việc trong một dự án vội vàng, bạn có thể cần phải dựa vào kiến thức và kinh nghiệm của chính mình nhiều hơn bình thường. Không có đủ thời gian để phân tích kỹ lưỡng tất cả các dữ liệu có sẵn. Nhưng nếu bạn nhận được một dự án liên quan đến nhiều thời gian và nguồn lực, thì chiến lược tốt nhất là dựa trên dữ liệu nhiều hơn. Tùy thuộc vào bạn, nhà phân tích dữ liệu, để đưa ra lựa chọn tốt nhất có thể. Bạn có thể sẽ pha trộn dữ liệu và kiến thức theo hàng triệu cách khác nhau trong suốt sự nghiệp phân tích dữ liệu của mình. Và bạn càng thực hành nhiều, bạn sẽ càng tìm thấy sự pha trộn hoàn hảo đó.*

## Origins of The Data Analysis Process

*Nguồn gốc của quá trình phân tích dữ liệu*

When you decided to join this program, you proved that you are a curious person. So, let’s tap into your curiosity and talk about the origins of data analysis. We don’t fully know when or why the first person decided to record data about people and things. But we do know it was useful because the idea is still around today!

*Khi bạn quyết định tham gia chương trình này, bạn đã chứng minh rằng bạn là một người tò mò. Vì vậy, hãy khai thác sự tò mò của bạn và nói về nguồn gốc của phân tích dữ liệu. Chúng tôi không hoàn toàn biết khi nào hoặc tại sao người đầu tiên quyết định ghi lại dữ liệu về con người và sự vật. Nhưng chúng tôi biết nó rất hữu ích vì ý tưởng này vẫn còn tồn tại cho đến ngày nay!*

We also know that 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.

*Chúng tôi cũng biết rằng phân tích dữ liệu bắt nguồn từ thống kê, có một lịch sử khá dài. Các nhà khảo cổ đánh dấu sự khởi đầu của thống kê ở Ai Cập cổ đại với việc xây dựng các kim tự tháp. Người Ai Cập cổ đại là bậc thầy trong việc tổ chức dữ liệu. Họ đã ghi lại các tính toán và lý thuyết của họ về giấy cói (vật liệu giống như giấy), hiện được xem là ví dụ sớm nhất về bảng tính và danh sách kiểm tra. Các nhà phân tích dữ liệu ngày nay nợ rất nhiều những người ghi chép xuất sắc, những người đã giúp tạo ra một quy trình kỹ thuật và hiệu quả hơn.*

It is time to enter the data analysis life cycle - the process of going from data to decision. Data goes through several phases as it gets created, consumed, tested, processed, and reused. With a life cycle model, all key team members can drive success by planning work both up front and at the end of the data analysis process. While the data analysis life cycle is well known among experts, there isn't a single defined structure of those phases. There might not be one single architecture that’s uniformly followed by every data analysis expert, but there are some shared fundamentals in every data analysis process. This reading provides an overview of several, starting with the process that forms the foundation of the Google Data Analytics Certificate.

*Đã đến lúc tham gia vào vòng đời phân tích dữ liệu, quá trình đi từ dữ liệu sang quyết định. Dữ liệu trải qua một số giai đoạn khi nó được tạo, tiêu thụ, thử nghiệm, xử lý và tái sử dụng. Với mô hình vòng đời, tất cả các thành viên chủ chốt trong nhóm đều có thể đóng góp vào sự thành công bằng cách lập kế hoạch cho công việc cả trước và cuối quy trình phân tích dữ liệu. Mặc dù vòng đời phân tích dữ liệu được các chuyên gia hiểu rõ ràng, nhưng không có một cấu trúc được xác định rõ ràng cho các giai đoạn đó. Có thể không có một kiến trúc nhất định nào được áp dụng đồng đều bởi tất cả các chuyên gia phân tích dữ liệu, nhưng có những nguyên tắc cơ bản chung trong mỗi quá trình phân tích dữ liệu. Bài đọc này cung cấp một tổng quan về một số quy trình, bắt đầu với quy trình tạo nên nền tảng cho chứng chỉ 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:

*Quy trình được trình bày như một phần của chứng chỉ Google Data Analytics, quy trình này sẽ có giá trị đối với bạn khi bạn tiếp tục phát triển sự nghiệp của mình.*

*(Vòng đời phân tích dữ liệu, quá trình đi từ dữ liệu sang quyết định)*

1. Ask: Business Challenge/Objective/Question

*Thách thức/ Mục tiêu/ Câu hỏi*

1. Prepare: Data generation, collection, storage, and data management

*Tạo, thu thập, lưu trữ và quản lý dữ liệu*

1. Process: Data cleaning/data integrity

*Làm sạch dữ liệu/toàn vẹn dữ liệu*

1. Analyze: Data exploration, visualization, and analysis

*Khám phá, trực quan hóa và phân tích dữ liệu*

1. Share: Communicating and interpreting results

*Truyền đạt và giải thích kết quả*

1. Act: Putting your insights to work to solve the problem

*Đưa những hiểu biết của bạn vào công việc để giải quyết vấn đề*

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 life cycle.

*Hiểu được quá trình này - và sự phổ biến của vòng lặp này - sẽ là một phần lớn trong việc hướng dẫn cho bạn về phân tích của riêng bạn và công việc trong chương trình này. Chúng ta hãy đi qua một vài biến thể khác của vòng đời phân tích dữ liệu.*

### EMC's data analysis life cycle

EMC Corporation's data analytics life cycle is cyclical with six steps:

*Vòng đời phân tích dữ liệu của EMC Corporation có tính chu kỳ với sáu bước:*

1. Discovery

*Khám phá*

1. Pre-processing data

*Tiền xử lý dữ liệu*

1. Model planning

*Lập kế hoạch mô hình*

1. Model building

*Xây dựng mô hình*

1. Communicate results.

*Truyền đạt kết quả*

1. Operationalize

*Vận hành*

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of real-world 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 life cycle 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.

*EMC Corporation hiện là Dell EMC. Mô hình này, được tạo ra bởi David Dietrich, phản ánh bản chất theo chu kỳ của các dự án trong thế giới thực. Các giai đoạn không cố định; mỗi bước kết nối và dẫn đến bước tiếp theo, và cuối cùng sẽ lặp lại. Các câu hỏi chính giúp các nhà phân tích kiểm tra xem họ đã hoàn thành đủ để tiếp tục và đảm bảo rằng nhóm đã dành đủ thời gian cho mỗi giai đoạn và không bắt đầu mô hình hóa trước khi dữ liệu sẵn sàng. Nó hơi khác so với vòng đời phân tích dữ liệu mà chương trình này dựa vào, nhưng có một số ý tưởng cốt lõi chung: giai đoạn đầu tiên quan tâm đến việc khám phá và đặt câu hỏi; dữ liệu phải được chuẩn bị trước khi có thể được phân tích và sử dụng; và sau đó, các kết quả nên được chia sẻ và hành động dựa trên chúng.*

For more information, refer to this e-book, [Data Science & Big Data Analytics](https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119183686).

### SAS's iterative life cycle

An iterative life cycle was created by a company called **SAS**, a leading data analytics solutions provider. It can be used to produce repeatable, reliable, and predictive results:

*Một vòng đời lặp lại được tạo ra bởi một công ty có tên là SAS, một nhà cung cấp các giải pháp phân tích dữ liệu hàng đầu. Nó có thể được sử dụng để tạo ra các kết quả lặp lại, đáng tin cậy và dự đoán được:*

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. Their life cycle has seven steps, many of which we have seen in the other models, like **Ask**, **Prepare**, **Model**, and **Act**. But this life cycle 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.

*Mô hình SAS nhấn mạnh tính vòng tròn của mô hình của họ bằng cách hình dung nó dưới dạng biểu tượng vô tận. Vòng đời của họ có bảy bước, nhiều trong số đó chúng ta đã thấy trong các mô hình khác, như* ***Ask****,* ***Prepare****,* ***Model*** *và* ***Act****. Tuy nhiên, vòng đời này cũng có một bước sau giai đoạn* ***Act*** *được thiết kế để giúp các nhà phân tích đánh giá các giải pháp của họ và có thể quay trở lại giai đoạn* ***Ask*** *một lần nữa.*

For more information, refer to [Managing the Analytics Life Cycle for Decisions at Scale](https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/manage-analytical-life-cycle-continuous-innovation-106179.pdf).

### Project-based data analytics life cycle

A project-based data analytics life cycle has five simple steps:

*Một vòng đời của dự án phân tích dữ liệu nền tảng có năm bước đơn giản:*

1. Identifying the problem

*Xác định vấn đề*

1. Designing data requirements

*Thiết kế yêu cầu dữ liệu*

1. Pre-processing data

*Tiền xử lý dữ liệu*

1. Performing data analysis

*Thực hiện phân tích dữ liệu*

1. Visualizing data

*Trực quan hóa dữ liệu*

This data analytics project life cycle was developed by Vignesh Prajapati. It doesn’t include the sixth phase, or what we have been referring to as the Act phase. However, it still covers a lot of the same steps as the life cycles we have already described. It begins with identifying the problem, preparing and processing data before analysis, and ends with data visualization.

*Vòng đời dự án phân tích dữ liệu này được phát triển bởi Vignesh Prajapati. Nó không bao gồm giai đoạn thứ sáu, hoặc những gì chúng ta đã đề cập đến là giai đoạn* ***Act****. Tuy nhiên, nó vẫn bao phủ nhiều bước giống như các vòng đời chúng ta đã mô tả trước đó. Nó bắt đầu bằng việc xác định vấn đề, chuẩn bị và xử lý dữ liệu trước khi phân tích và kết thúc bằng việc trực quan hóa dữ liệu.*

For more information, refer to [Understanding the data analytics project life cycle](http://pingax.com/understanding-data-analytics-project-life-cycle/).

### Big data analytics life cycle

Authors Thomas Erl, Wajid Khattak, and Paul Buhler proposed a big data analytics life cycle in their book, **Big Data Fundamentals: Concepts, Drivers & Techniques**. Their life cycle suggests phases divided into nine steps:

*Các tác giả Thomas Erl, Wajid Khattak và Paul Buhler đã đề xuất một vòng đời phân tích dữ liệu lớn trong cuốn sách của họ,* ***Big Data Fundamentals: Concepts, Drivers & Techniques****. Vòng đời của họ đề xuất các giai đoạn được chia thành chín bước.*

1. Business case evaluation

*Đánh giá các business case*

1. Data identification

*Nhận dạng dữ liệu*

1. Data acquisition and filtering

*Thu thập và lọc dữ liệu*

1. Data extraction

*Khai thác dữ liệu*

1. Data validation and cleaning

*Xác thực và làm sạch dữ liệu*

1. Data aggregation and representation

*Tổng hợp và biểu diễn dữ liệu*

1. Data analysis

*Phân tích dữ liệu*

1. Data visualization

*Trực quan hóa dữ liệu*

1. Utilization of analysis results

*Sử dụng kết quả phân tích*

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

*Vòng đời này có vẻ có ba hoặc bốn bước nhiều hơn so với các mô hình vòng đời phân tích dữ liệu trước đó. Nhưng thực tế, họ chỉ chia nhỏ những gì chúng ta đã đề cập là Prepare và Process thành các bước nhỏ hơn. Nó nhấn mạnh các nhiệm vụ cụ thể cần thiết để thu thập, chuẩn bị và làm sạch dữ liệu trước khi vào giai đoạn phân tích.*

For more information, refer to [Big Data Adoption and Planning Considerations](https://www.informit.com/articles/article.aspx?p=2473128&seqNum=11&ranMID=24808).

### Key takeaway

From our journey to the pyramids and data in ancient Egypt to now, the way we 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, we are here to help guide you as you continue on your data journey.

*Từ chuyến hành trình đến những kim tự tháp và dữ liệu của Ai Cập cổ đến hiện nay, cách chúng ta phân tích dữ liệu đã tiến hóa (và vẫn tiếp tục). Quá trình phân tích dữ liệu giống như kiến trúc cuộc sống thực, có nhiều cách để làm nhưng những ý tưởng cốt lõi vẫn xuất hiện trong mỗi mô hình của quá trình đó. Cho dù bạn sử dụng cấu trúc của Chứng chỉ Google Data Analytics này hoặc một trong các biến thể khác mà bạn đã học được, chúng tôi sẵn sàng giúp đỡ bạn khi bạn tiếp tục hành trình của mình với dữ liệu.*

## Key Data Analyst Skills

Analytical skills are qualities and characteristics associated with solving problems using facts. There are a lot of aspects to analytical skills, but, we'll focus on five essential points:

1. Curiosity
2. Understanding context
3. Having technical mindset
4. Data design
5. Data strategy

### Curiosity

Curiosity is all about wanting to learn something. Curious people usually seek out new challenges and experiences. This leads to knowledge.

### Understanding context

Context is the condition in which something exists or happens. This can be a structure or an environment.

*Ngữ cảnh là điều kiện trong đó một thứ gì đó tồn tại hoặc xảy ra. Điều này có thể là một cấu trúc hoặc môi trường.*

A simple way of understanding context is by counting to 5. One, two, three, four, five. All of those numbers exist in the context of one through five. But what if a friend of yours said to you, one, two, four, five, three? Well, the three will be out of context. Simple, right? But it can be a little tricky. There's a good chance that you might not even notice the three being out of context if you aren't paying close attention. That's why listening and trying to understand the full picture is critical.

*Một cách đơn giản để hiểu ngữ cảnh là đếm đến 5. Một, hai, ba, bốn, năm. Tất cả các số đó tồn tại trong ngữ cảnh từ một đến năm. Nhưng nếu một người bạn của bạn nói với bạn là một, hai, bốn, năm, ba? Vậy, số ba sẽ bị lạc khỏi ngữ cảnh. Đơn giản, phải không? Nhưng nó có thể hơi phức tạp một chút. Có khả năng cao là bạn sẽ không thể nhận ra số ba bị lạc khỏi ngữ cảnh nếu bạn không chú ý kỹ. Đó là lý do tại sao việc lắng nghe và cố gắng hiểu toàn bộ tình hình là rất quan trọng.*

In your own life, you put things into context all the time. For example, let's think about your grocery list. If you group together items like flour, sugar, and yeast, that's you adding context to your groceries. This saves you time when you're at the baking aisle at the grocery store.

*Trong cuộc sống của bạn, bạn đưa mọi thứ vào ngữ cảnh hàng ngày. Ví dụ, hãy nghĩ về danh sách đồ tạp hóa của bạn. Nếu bạn nhóm các mặt hàng như bột mì, đường và men nướng lại với nhau, đó là bạn đang thêm ngữ cảnh vào đồ tạp hóa của mình. Điều này tiết kiệm thời gian cho bạn khi bạn đến khu vực bán đồ nướng tại cửa hàng tạp hóa.*

Let's look at another example. Have you ever shuffled a deck of cards and noticed the joker? If you're playing a game that doesn't include jokers, identifying that card means you understand it's out of context. Remove it, and you're much more likely to play a successful game.

*Hãy xem một ví dụ khác. Bạn đã bao giờ xáo trộn một bộ bài và để ý đến quân Joker chưa? Nếu bạn đang chơi một trò chơi không bao gồm quân Joker, việc xác định được thẻ này có nghĩa là bạn hiểu rằng nó không thuộc ngữ cảnh. Hãy loại bỏ nó, và bạn sẽ có nhiều khả năng chơi một trò chơi thành công hơn.*

### Technical mindset

A technical mindset involves the ability to break things down into smaller steps or pieces and work with them in an orderly and logical way.

*Tư duy kỹ thuật bao gồm khả năng phân tích một vấn đề thành các bước nhỏ hơn hoặc các phần và làm việc với chúng theo cách có trật tự và logic.*

For instance, when paying your bills, you probably already break down the process into smaller steps. Maybe you start by sorting them by the date they're due. Next, you might add them up and compare that amount to the balance in your bank account. This would help you see if you can pay your bills now, or if you should wait until the next paycheck. Finally, you'd pay them. When you take something that seems like a single task, like paying your bills, and break it into smaller steps with an orderly process, that's using a technical mindset.

*Ví dụ, khi thanh toán hóa đơn, bạn có thể đã chia quá trình đó thành các bước nhỏ hơn. Có thể bạn bắt đầu bằng cách sắp xếp chúng theo ngày đáo hạn. Tiếp theo, bạn có thể cộng chúng lại và so sánh số tiền đó với số dư trong tài khoản ngân hàng của bạn. Điều này giúp bạn xem liệu bạn có thể thanh toán hóa đơn ngay bây giờ hay nên đợi đến lần nhận lương tiếp theo. Cuối cùng, bạn sẽ thanh toán chúng. Khi bạn lấy một nhiệm vụ dường như là một công việc đơn lẻ, như thanh toán hóa đơn, và chia thành các bước nhỏ hơn với quy trình có trật tự, đó là việc sử dụng tư duy kỹ thuật.*

### Data design

Data design is how you organize information. As a data analyst, design typically has to do with an actual database. But, again, the same skills can easily be applied to everyday life.

*Thiết kế dữ liệu là cách bạn tổ chức thông tin. Với vai trò là một nhà phân tích dữ liệu, thiết kế thường liên quan đến một cơ sở dữ liệu thực tế. Nhưng lại với các kỹ năng tương tự có thể dễ dàng được áp dụng vào cuộc sống hàng ngày.*

For example, think about the way you organize the contacts in your phone. That's actually a type of data design. Maybe you list them by first name instead of last, or maybe you use email addresses instead of their names. What you're really doing is designing a clear, logical list that lets you call or text a contact in a quick and simple way.

*Ví dụ, hãy nghĩ về cách bạn tổ chức danh bạ trong điện thoại của mình. Đó thực sự là một loại thiết kế dữ liệu. Có thể bạn liệt kê theo tên họ thay vì họ tên, hoặc có thể bạn sử dụng địa chỉ email thay vì tên họ. Những gì bạn thực sự đang làm là thiết kế một danh sách rõ ràng, có logic cho phép bạn gọi hoặc nhắn tin cho một liên lạc một cách nhanh chóng và đơn giản.*

### Data strategy

Data strategy is the management of the people, processes, and tools used in data analysis. Let's break that down. You manage people by making sure they know how to use the right data to find solutions to the problem you're working on. For processes, it's about making sure the path to that solution is clear and accessible. For tools, you make sure the right technology is being used for the job.

*Chiến lược dữ liệu là việc quản lý những người, quy trình và công cụ được sử dụng trong phân tích dữ liệu. Hãy phân tích điều đó. Bạn quản lý con người bằng cách đảm bảo họ biết cách sử dụng dữ liệu đúng để tìm giải pháp cho vấn đề mà bạn đang làm việc. Đối với quy trình, điều đó liên quan đến việc đảm bảo con đường đến giải pháp đó là rõ ràng và dễ tiếp cận. Đối với công cụ, bạn đảm bảo rằng công nghệ đúng đang được sử dụng cho công việc.*

Imagine mowing a lawn.

*Hãy tưởng tượng bạn đang cắt cỏ.*

* Step 1 would be reading the owner's manual for the mower. That's making sure the people involved, or you, in this example, know how to use the data available. The manual would instruct you to put on protective eyewear and closed-toe shoes.

*Bước 1 là đọc hướng dẫn sử dụng máy cắt cỏ. Điều này đảm bảo rằng những người tham gia, hoặc bạn trong ví dụ này, biết cách sử dụng dữ liệu có sẵn. Hướng dẫn sẽ chỉ dẫn bạn đeo kính bảo hộ và giày đóng để.*

* Then, it's on to step 2: making the process, the path, clear and accessible. This will involve you walking around the lawn, picking up large sticks or rocks that might get in your way.

*Sau đó, bạn tiến hành bước 2: đảm bảo quy trình, con đường, rõ ràng và dễ tiếp cận. Điều này sẽ bao gồm bạn đi quanh cỏ, nhặt các cây gậy lớn hoặc đá có thể gây trở ngại cho bạn.*

* Finally, for step 3, you check the lawn mower, your tool, to make sure it has enough gas and oil, and is in working condition, so the lawn can be mowed safely.

*Cuối cùng, cho bước 3, bạn kiểm tra máy cắt cỏ, công cụ của bạn, để đảm bảo nó đủ nhiên liệu và dầu nhớt, và đang trong tình trạng hoạt động, để cỏ có thể được cắt an toàn.*

There you have it. Now you know the five essential skills of a data analyst. Curiosity, understanding context, having a technical mindset, data design, and data strategy.

*Đó là những gì bạn cần biết về năm kỹ năng cần thiết của một nhà phân tích dữ liệu. Sự tò mò, hiểu bối cảnh, có tư duy kỹ thuật, thiết kế dữ liệu và chiến lược dữ liệu.*

## All About Thinking Analytically

People don't often think about thinking. Thinking is second nature to us. It just happens automatically, but there are actually many different ways to think. Some people think creatively, some think critically, and some people think in abstract ways. Let's talk about **analytical thinking**. Analytical thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner. As data analysts, how do we think analytically? Well, to answer that question, we will now talk about a second set of five. The five key aspects to analytical thinking:

*Mọi người không thường nghĩ về sự tư duy. Tư duy là bản chất thứ hai đối với chúng ta. Nó xảy ra một cách tự nhiên, nhưng thực tế là có nhiều cách khác nhau để tư duy. Một số người tư duy sáng tạo, một số người tư duy phê phán và một số người tư duy một cách trừu tượng. Hãy nói về* ***tư duy phân tích****. Tư duy phân tích liên quan đến việc xác định và định nghĩa vấn đề và sau đó giải quyết nó bằng cách sử dụng dữ liệu một cách có tổ chức, từng bước. Làm thế nào chúng ta (nhà phân tích dữ liệu) suy nghĩ phân tích? Để trả lời câu hỏi đó, chúng ta sẽ nói về một bộ thứ hai gồm năm yếu tố quan trọng của tư duy phân tích:*

1. Visualization

*Trực quan hóa*

1. Strategy

*Chiến lược*

1. Problem-orientation

*Định hướng vấn đề*

1. Correlation

*Sự tương quan*

1. Big-picture and detail-oriented thinking

*Suy nghĩ toàn diện và chú ý đến chi tiết*

### Visualization

In data analytics, visualization is the graphical representation of information. Some examples include graphs, maps, or other design elements. Visualization is important because visuals can help data analysts understand and explain information more effectively.

*Trong phân tích dữ liệu, trực quan hóa là biểu diễn đồ họa của thông tin. Một số ví dụ bao gồm đồ thị, bản đồ hoặc các yếu tố thiết kế khác. Trực quan hóa quan trọng vì hình ảnh có thể giúp các nhà phân tích dữ liệu hiểu và giải thích thông tin một cách hiệu quả hơn.*

Think about it like this. If you are trying to explain the Grand Canyon to someone, using words would be much more challenging than showing them a picture. A visualization of the Grand Canyon would help you make your point much quicker.

*Hãy nghĩ về nó như thế này. Nếu bạn đang cố gắng giải thích Vực Grand Canyon cho ai đó, sử dụng từ ngữ sẽ khó khăn hơn so với việc cho họ xem một bức tranh. Một biểu đồ trực quan của Vực Grand Canyon sẽ giúp bạn truyền tải ý của mình nhanh hơn.*

### Strategic

With so much data available, having a strategic mindset is key to staying focused and on track. Strategizing helps data analysts see what they want to achieve with the data and how they can get there. Strategy also helps improve the quality and usefulness of the data we collect. By strategizing, we know all our data is valuable and can help us accomplish our goals.

*Với nhiều dữ liệu có sẵn, việc có tư duy chiến lược là chìa khóa để luôn tập trung và đi đúng hướng. Chiến lược giúp cho các chuyên viên phân tích dữ liệu nhìn thấy mục tiêu mà họ muốn đạt được với dữ liệu và cách thức để đạt được điều đó. Chiến lược cũng giúp cải thiện chất lượng và tính hữu ích của dữ liệu mà chúng ta thu thập. Bằng cách lập chiến lược, chúng ta biết rằng tất cả dữ liệu của chúng ta đều có giá trị và có thể giúp chúng ta đạt được mục tiêu của mình.*

### Problem-oriented

Data analysts use a problem-oriented approach in order to identify, describe, and solve problems. It's all about keeping the problem top of mind throughout the entire project. For example, say a data analyst is told about the problem of a warehouse constantly running out of supplies. They would move forward with different strategies and processes. But the number one goal would always be solving the problem of keeping inventory on the shelves. Data analysts also ask a lot of questions. This helps improve communication and saves time while working on a solution. An example of that would be surveying customers about their experiences using a product and building insights from those questions to improve their product.

*Các chuyên gia phân tích dữ liệu sử dụng cách tiếp cận theo định hướng vấn đề để xác định, mô tả và giải quyết vấn đề. Điều quan trọng là giữ vấn đề luôn được đặt lên hàng đầu trong suốt dự án. Ví dụ, nếu một chuyên gia phân tích dữ liệu được thông báo về vấn đề của một kho hàng liên tục thiếu hàng hoá. Họ sẽ tiến lên với các chiến lược và quy trình khác nhau. Nhưng mục tiêu số một sẽ luôn là giải quyết vấn đề giữ hàng tồn kho trên kệ. Các chuyên gia phân tích dữ liệu cũng đặt nhiều câu hỏi. Điều này giúp cải thiện thông tin và tiết kiệm thời gian trong khi làm việc trên một giải pháp. Một ví dụ về điều đó là thăm dò khách hàng về trải nghiệm của họ khi sử dụng sản phẩm và xây dựng sự thấu hiểu từ những câu hỏi đó để cải thiện sản phẩm của họ.*

### Correlation

A correlation is like a relationship. You can find all kinds of correlations in data. Maybe it's the relationship between the length of your hair and the amount of shampoo you need. Or maybe you notice a correlation between a rainier season leading to a high number of umbrellas being sold. But as you start identifying correlations in data, there's one thing you always want to keep in mind: Correlation does not equal causation. In other words, just because two pieces of data are both trending in the same direction, that doesn't necessarily mean they are all related.

*Mối tương quan giống như một mối quan hệ. Bạn có thể tìm thấy tất cả các loại mối tương quan trong dữ liệu. Có thể đó là mối quan hệ giữa độ dài tóc của bạn và lượng dầu gội bạn cần. Hoặc có thể bạn nhận thấy một mối tương quan giữa mùa mưa dẫn đến số lượng ô dù bán ra cao. Tuy nhiên, khi bạn bắt đầu xác định các mối tương quan trong dữ liệu, luôn có một điều bạn muốn ghi nhớ: Tương quan không có nghĩa là nhân quả. Nói cách khác, chỉ vì hai phần dữ liệu đang có xu hướng tăng cùng một hướng, điều đó không nhất thiết có nghĩa là chúng có liên quan với nhau.*

### Big-picture and Detail-oriented thinking

**Big-picture thinking** means being able to see the big picture as well as the details. A jigsaw puzzle is a great way to think about this. Big-picture thinking is like looking at a complete puzzle. You can enjoy the whole picture without getting stuck on every tiny piece that went into making it. If you only focus on individual pieces, you wouldn't be able to see past that, which is why big-picture thinking is so important. It helps you zoom out and see possibilities and opportunities. This leads to exciting new ideas or innovations.

*Tư duy hình ảnh lớn có nghĩa là có thể nhìn thấy bức tranh lớn cũng như các chi tiết. Trò chơi ghép hình là một cách tuyệt vời để suy nghĩ về điều này. Suy nghĩ về bức tranh lớn giống như nhìn vào một câu đố hoàn chỉnh. Bạn có thể thưởng thức toàn bộ bức tranh mà không bị mắc kẹt vào từng chi tiết nhỏ đã tạo nên nó. Nếu bạn chỉ tập trung vào từng phần riêng lẻ, bạn sẽ không thể nhìn xa hơn, đó là lý do tại sao tư duy tổng thể lại quan trọng đến vậy. Nó giúp bạn thu nhỏ và nhìn thấy các khả năng và cơ hội. Điều này dẫn đến những ý tưởng hoặc đổi mới thú vị.*

On the flip side, **detail-oriented thinking** is all about figuring out all of the aspects that will help you execute a plan. In other words, the pieces that make up your puzzle. There are all kinds of problems in the business world that can benefit from employees who have both a big-picture and a detail-oriented way of thinking.

*Mặt khác, tư duy định hướng chi tiết là tìm ra tất cả các khía cạnh sẽ giúp bạn thực hiện một kế hoạch. Nói cách khác, các mảnh tạo nên câu đố của bạn. Có tất cả các loại vấn đề trong thế giới business có thể được hưởng lợi từ những nhân viên vừa có tầm nhìn bao quát vừa có lối suy nghĩ chi tiết.*

Most of us are naturally better at one or the other. But you can always develop the skills to fit both pieces together.

*Hầu hết chúng ta có thể giỏi hơn ở cái này hay cái khác. Nhưng bạn luôn có thể phát triển các kỹ năng để khớp cả hai mảnh ghép lại với nhau.*

## Exploring Core Analytical Skills

There are 5 key aspects are **Visualization**, **Strategy**, **Problem-Orientation**, **Correlation** and **Using Big-Picture And Detail-Oriented Thinking**. Different people naturally use certain types of thinking, but that you can absolutely grow and develop the skills that might not come as easily to you. This means you can become a versatile thinker, which is a very important part of data analysis. You might naturally be an analytical thinker, but you can learn to think creatively and critically, and be great at all three. The more ways you can think, the easier it is to think outside the box and come up with fresh ideas.

*Có 5 khía cạnh chính là* ***Trực quan hóa****,* ***Chiến lược****,* ***Định hướng vấn đề****,* ***Tương quan*** *và* ***Sử dụng tư duy tổng thể và định hướng chi tiết****. Những người khác nhau thì có những kiểu tư duy nhất định, nhưng bạn hoàn toàn có thể trau dồi và phát triển những kỹ năng không phải là thế mạnh của bạn. Điều này có nghĩa là bạn có thể trở thành một nhà tư tưởng linh hoạt, đây là một phần rất quan trọng trong phân tích dữ liệu. Bạn có thể bẩm sinh là một người có tư duy phân tích, nhưng bạn có thể học cách tư duy sáng tạo và phê phán, đồng thời trở nên xuất sắc ở cả ba lĩnh vực này. Bạn càng có nhiều cách để suy nghĩ, bạn càng dễ dàng suy nghĩ sáng tạo và đưa ra những ý tưởng mới.*

But why is it important to think in different ways? Because in data analysis, solutions are almost never right in front of you. You need to think critically to find out the right questions to ask. But you also need to think creatively to get new and unexpected answers.

*Nhưng tại sao điều quan trọng là phải suy nghĩ theo những cách khác nhau? Bởi vì trong phân tích dữ liệu, các giải pháp hầu như không bao giờ ở ngay trước mặt bạn. Bạn cần suy nghĩ chín chắn để tìm ra những câu hỏi phù hợp để hỏi. Nhưng bạn cũng cần suy nghĩ sáng tạo để có những câu trả lời mới và bất ngờ.*

There are some of the questions data analysts ask when they're on the hunt for a solution:

*Có một số câu hỏi mà các nhà phân tích dữ liệu đặt ra khi họ đang ở trong công cuộc săn lùng tìm kiếm một giải pháp:*

* What is the root cause of a problem?

*Nguyên nhân cốt lõi của vấn đề là gì?*

* Where are the gaps in our process?

*Đâu là lỗ hổng trong quy trình của chúng ta?*

* What did we not consider before?

*Điều gì chúng ta đã không xem xét trước đây?*

### What did we not consider before?

A root cause is the reason why a problem occurs. If we can identify and get rid of a root cause, we can prevent that problem from happening again.

*Nguyên nhân gốc rễ là lý do tại sao một vấn đề xảy ra. Nếu chúng ta có thể xác định và loại bỏ nguyên nhân gốc rễ, chúng ta có thể ngăn vấn đề đó tái diễn.*

A simple way to wrap your head around root causes is with the process called the Five Whys. In the Five Whys you ask "why" five times to reveal the root cause. The fifth and final answer should give you some useful and sometimes surprising insights.

*Một cách đơn giản để giải quyết các nguyên nhân gốc rễ là sử dụng quy trình có tên là* ***5 Whys****. Trong* ***5 Whys****, bạn hỏi* ***"Tại sao"*** *năm lần để tiết lộ nguyên nhân gốc rễ. Câu trả lời thứ năm và cũng là câu trả lời cuối cùng sẽ cung cấp cho bạn một số hiểu biết hữu ích và đôi khi gây ngạc nhiên.*

Here's an example of the Five Whys in action. Let's say you wanted to make a blueberry pie but couldn't find any blueberries. You've been trying to solve a problem by asking

* Why can't I make a blueberry pie? The answer will be, there are no blueberries at the store. There's Why Number 1.
* Then ask, why were there no blueberries at the store? Then you discover that the blueberry bushes don't have enough fruit this season. That's Why Number 2.
* Next, you'd ask, why was there not enough fruit? This would lead to the fact that birds were eating all the berries. Why Number 3, asked and answered.
* Now we get to Why Number 4. Ask why a fourth time and the answer would be that, although the birds normally prefer mulberries and don't eat blueberries, the mulberry bush didn't produce fruit this season, so the birds are eating blueberries instead.
* Finally, we get to Why Number 5, which should reveal the root cause. A late frost damaged the mulberry bushes, so it didn't produce any fruit. You can't make a blueberry pie because of the late frost months ago.

See how the Five Whys can reveal some very surprising root causes. This is a great trick to know, and it can be a very helpful process in data analysis.

***5 Whys*** *có thể tiết lộ một số nguyên nhân gốc rễ rất đáng ngạc nhiên. Đây là một mẹo rất hay và nó có thể là một quá trình rất hữu ích trong phân tích dữ liệu.*

### Where are the gaps in our process?

For this, many people will use something called gap analysis. Gap analysis lets you examine and evaluate how a process works currently in order to get where you want to be in the future. Businesses conduct gap analysis to do all kinds of things, such as improve a product or become more efficient. The general approach to gap analysis is understanding where you are now compared to where you want to be. Then you can identify the gaps that exist between the current and future state and determine how to bridge them.

*Để làm được điều này, nhiều người sẽ sử dụng một công cụ gọi là phân tích khoảng cách* ***(gap analysis)****.**Phân tích khoảng cách cho phép bạn khảo sát và đánh giá cách một quy trình hoạt động hiện tại để đạt được mục tiêu trong tương lai. Các doanh nghiệp tiến hành phân tích lỗ hổng để làm tất cả mọi thứ, chẳng hạn như cải thiện sản phẩm hoặc trở nên hiệu quả hơn. Cách tiếp cận chung để phân tích khoảng cách là hiểu bạn đang ở đâu so với nơi bạn muốn. Sau đó, bạn có thể xác định khoảng cách tồn tại giữa trạng thái hiện tại và tương lai, đồng thời xác định cách khắc phục chúng.*

### What did we not consider before?

This is a great way to think about what information or procedure might be missing from a process, so you can identify ways to make better decisions and strategies moving forward.

*Đây là một cách tuyệt vời để suy nghĩ về thông tin hoặc quy trình nào có thể bị thiếu trong một quy trình, nhờ đó bạn có thể xác định các cách để đưa ra quyết định và chiến lược tốt hơn trong tương lai.*

These are just a few examples of the kinds of questions data analysts use at their jobs every day. As you begin your career, I'm sure you'll think of a whole lot more. The way data analysts think and ask questions plays a big part in how businesses make decisions. That's why analytical thinking and understanding how to ask the right questions can have such a huge impact on the overall success of a business.

*Đây chỉ là một vài ví dụ về các loại câu hỏi mà các nhà phân tích dữ liệu sử dụng trong công việc của họ hàng ngày. Khi bạn bắt đầu sự nghiệp của mình, tôi chắc rằng bạn sẽ nghĩ đến nhiều thứ hơn nữa. Cách các nhà phân tích dữ liệu tư duy và đặt câu hỏi đóng một vai trò quan trọng trong cách các doanh nghiệp đưa ra quyết định. Đó là lý do tại sao* ***tư duy phân tích*** *và hiểu* ***cách đặt câu hỏi đúng*** *có thể có tác động rất lớn đến thành công chung của một doanh nghiệp.*

## Using Data to Drive Successful Outcomes

Data-driven decision-making involved using facts to guide business strategy. Data analysts can tap into the power of data to do all kinds of amazing things. With data, they can gain valuable insights, verify their theories or assumptions, better understand opportunities and challenges, support an objective, help make a plan, and much more. In business, data-driven decision-making can improve the results in a lot of different ways.

*Việc ra quyết định dựa trên dữ liệu liên quan đến việc sử dụng các dữ kiện để định hướng chiến lược kinh doanh. Các nhà phân tích dữ liệu có thể khai thác sức mạnh của dữ liệu để làm tất cả những điều tuyệt vời. Với dữ liệu, họ có thể thu được những hiểu biết có giá trị, xác minh lý thuyết hoặc giả định của mình, hiểu rõ hơn về cơ hội và thách thức, hỗ trợ mục tiêu, giúp lập kế hoạch, v.v. Trong kinh doanh, việc ra quyết định dựa trên dữ liệu có thể cải thiện kết quả theo nhiều cách khác nhau.*

For example, say a dairy farmer wants to start making and selling ice cream. They could guess what flavors customers would like, but there's a better way to get the information. The farmer could survey people and ask them what flavors they prefer. This gives the farmer the data they need to pick ice cream flavors people will enjoy.

*Ví dụ, giả sử một nông dân chăn nuôi bò sữa muốn bắt đầu sản xuất và bán kem. Họ có thể đoán được hương vị mà khách hàng thích, nhưng có một cách tốt hơn để lấy thông tin. Người nông dân có thể khảo sát mọi người và hỏi họ thích hương vị nào. Điều này cung cấp cho người nông dân dữ liệu họ cần để chọn hương vị kem mà mọi người sẽ thích.*

Another example. Let's say the president of an organization is curious about what perks employees value most. She asked the human resources director who says people value casual dress code. It's a gut feeling, but the HR director backs it up with the fact that he sees a lot of people wearing jeans and t-shirts. But what if this company were to use a more structured employee feedback process, such as a survey? It might reveal that employees actually enjoy free public transportation cards the most. The human resources director just didn't realize that because he drives to work.

*Một ví dụ khác. Giả sử chủ tịch của một tổ chức tò mò về những đặc quyền mà nhân viên đánh giá cao nhất. Cô hỏi giám đốc nhân sự cho rằng mọi người coi trọng quy tắc ăn mặc giản dị. Đó là một linh cảm, nhưng giám đốc nhân sự chứng minh điều đó với thực tế là anh ta nhìn thấy rất nhiều người mặc quần jean và áo phông. Nhưng điều gì sẽ xảy ra nếu công ty này sử dụng quy trình phản hồi có cấu trúc hơn, chẳng hạn như một cuộc khảo sát? Nó có thể tiết lộ rằng nhân viên thực sự thích nhất là miễn phí vé giao thông công cộng. Giám đốc nhân sự không nhận ra điều đó vì anh ta lái xe đi làm.*

These are just some of the benefits of data-driven decision-making. It gives you greater confidence about your choice and your abilities to address business challenges. It helps you become more proactive when an opportunity presents itself, and it saves you time and effort when working towards a goal.

*Đây chỉ là một số lợi ích của việc ra quyết định dựa trên dữ liệu. Nó mang lại cho bạn sự tự tin hơn về sự lựa chọn của bạn và khả năng của bạn để giải quyết những thách thức kinh doanh. Nó giúp bạn trở nên chủ động hơn khi có cơ hội, đồng thời giúp bạn tiết kiệm thời gian và công sức khi hướng tới mục tiêu.*

Now let's learn more about how these five skills help you tap into all the potential of data-driven decision-making.

### Curiosity and context

The more you learn about the power of data, the more curious you're likely to become. You'll start to see patterns and relationships in everyday life, whether you're reading the news, watching a movie, or going to an appointment across town. The analysts take their thinking a step further by using context to make predictions, research answers, and eventually draw conclusions about what they've discovered. This natural process is a great first step in becoming more data-driven.

*Bạn càng tìm hiểu nhiều về sức mạnh của dữ liệu, bạn càng có khả năng trở nên tò mò hơn. Bạn sẽ bắt đầu thấy các khuôn mẫu và mối quan hệ trong cuộc sống hàng ngày, cho dù bạn đang đọc tin tức, xem phim hay đến một cuộc hẹn trong thành phố. Các nhà phân tích đưa suy nghĩ của họ tiến thêm một bước bằng cách sử dụng bối cảnh để đưa ra dự đoán, nghiên cứu câu trả lời và cuối cùng đưa ra kết luận về những gì họ đã khám phá. Quá trình tự nhiên này là bước đầu tiên tuyệt vời để trở nên hướng dữ liệu nhiều hơn.*

### Having a technical mindset

Everyone has instincts, or as in the case of our human resources director example, gut feelings. Data analysts are no different. They have gut feelings too. But they've trained themselves to build on those feelings and use a more technical approach to explore them. They do this by always seeking out the facts, putting them to work through analysis, and using the insights they gain to make informed decisions.

*Mọi người đều có bản năng, hoặc như trong trường hợp ví dụ về giám đốc nhân sự của chúng tôi, trực giác. Các nhà phân tích dữ liệu cũng không khác. Họ cũng có trực giác. Nhưng họ đã rèn luyện bản thân để xây dựng những cảm xúc đó và sử dụng một cách tiếp cận kỹ thuật hơn để khám phá chúng. Họ làm điều này bằng cách luôn tìm kiếm sự thật, đưa chúng vào hoạt động thông qua phân tích và sử dụng những hiểu biết sâu sắc mà họ có được để đưa ra quyết định sáng suốt.*

### Data design

Data design has a strong connection to data-driven decision-making. To put it simply, designing your data so that it is organized in a logical way makes it easy for data analysts to access, understand, and make the most of available information. And it's important to keep in mind that data design doesn't just apply to databases. This kind of thinking can work with all sorts of real-life situations too. The basic idea is this. If you make decisions that are informed by data, you are more likely to make more informed and effective decisions.

*Thiết kế dữ liệu có mối liên hệ chặt chẽ với việc ra quyết định dựa trên dữ liệu. Nói một cách đơn giản, việc thiết kế dữ liệu của bạn sao cho dữ liệu được tổ chức theo cách hợp lý giúp các nhà phân tích dữ liệu dễ dàng truy cập, hiểu và tận dụng tối đa thông tin có sẵn. Và điều quan trọng cần lưu ý là thiết kế dữ liệu không chỉ áp dụng cho cơ sở dữ liệu. Kiểu suy nghĩ này cũng có thể phù hợp với tất cả các loại tình huống thực tế. Ý tưởng cơ bản là thế này. Nếu bạn đưa ra quyết định dựa trên dữ liệu, bạn có nhiều khả năng đưa ra quyết định sáng suốt và hiệu quả hơn.*

### Data strategy

Incorporates the people, processes, and tools used to solve a problem. This is a big one to remember because data strategy gives you a high-level view of the path you need to take to achieve your goals. Also, data-driven decision-making isn't a one-person job. It's much more likely to be successful if everyone is on board and on the same page, so it's important to make sure specific procedures are in place and that your technology being used is aligned with your data-driven strategy.

*Kết hợp con người, quy trình và công cụ được sử dụng để giải quyết vấn đề. Đây là một điều quan trọng cần nhớ vì chiến lược dữ liệu cung cấp cho bạn cái nhìn tổng thể về con đường bạn cần thực hiện để đạt được mục tiêu của mình. Ngoài ra, việc ra quyết định dựa trên dữ liệu không phải là công việc của một người. Sẽ có nhiều khả năng thành công hơn nếu mọi người tham gia và trên cùng một page, vì vậy điều quan trọng là phải đảm bảo các quy trình cụ thể được áp dụng và công nghệ bạn đang sử dụng phù hợp với chiến lược dựa trên dữ liệu của bạn.*

## Real-world Data Magic

There are some case studies that highlight the incredible work data analysts do. Each of these scenarios shows off the power of data-driven decision-making in unexpected ways.

*Có một số nghiên cứu điển hình làm nổi bật công việc đáng kinh ngạc của các nhà phân tích dữ liệu. Mỗi tình huống trong số này thể hiện sức mạnh của việc ra quyết định dựa trên dữ liệu theo những cách bất ngờ.*

There's really no end to what you can do as a data analyst. You'll discover even more possibilities.

*Thực sự không có giới hạn cho những gì bạn có thể làm với tư cách là một nhà phân tích dữ liệu. Bạn sẽ khám phá ra nhiều khả năng hơn nữa.*

### The first story is about Google.

Here at Google, our mission is to organize the world's information and make it universally accessible and useful. All of our products, from idea to development to launch, are built on data and data-driven decision-making.

*Tại Google, sứ mệnh của chúng tôi là tổ chức thông tin của thế giới và làm cho thông tin đó trở nên hữu ích và có thể truy cập được trên toàn cầu. Tất cả các sản phẩm của chúng tôi, từ ý tưởng đến phát triển đến ra mắt, đều được xây dựng dựa trên dữ liệu và quá trình ra quyết định dựa trên dữ liệu.*

There are tons of examples here at Google of people using facts to create business strategy. But one of the most famous ones has to do with Google's human resources. Here's how it went. The HR department wanted to know if there was value in having managers. Were their contributions worthwhile? Or should everyone just be an individual contributor?

*Có rất nhiều ví dụ ở đây tại Google về những người sử dụng dữ kiện để tạo chiến lược kinh doanh. Nhưng một trong những điều nổi tiếng nhất liên quan đến nguồn nhân lực của Google. Đây là cách nó đã đi. Bộ phận nhân sự muốn biết liệu có giá trị gì khi có các nhà quản lý hay không. Những đóng góp của họ có đáng giá không? Hay mọi người chỉ nên là một người đóng góp cá nhân?*

To answer that question, Google's people analytics team looked at past performance reviews and employee surveys. The data they found was plotted on a graph because as you've learned, visuals are extremely helpful when trying to understand a problem or concept. The graph revealed that Googlers had positive feelings about their managers, but the data was pretty general and the team wanted to learn more. So they dug deeper and split the data into quartiles. A quartile divides data points into four equal parts or quarters. Here's where the really cool stuff started happening. The data analysts discovered that there was a big difference between the very top and the very bottom quartiles. As it turned out, the teams with the best managers were significantly happier, more productive, and more likely to want to keep working at Google. This confirmed that managers were valued and make a big difference.

*Để trả lời câu hỏi đó, nhóm phân tích con người của Google đã xem xét các bài đánh giá hiệu suất trong quá khứ và các cuộc khảo sát nhân viên. Dữ liệu họ tìm thấy được vẽ trên biểu đồ vì như bạn đã biết, hình ảnh cực kỳ hữu ích khi cố gắng hiểu một vấn đề hoặc khái niệm. Biểu đồ tiết lộ rằng nhân viên Google có cảm xúc tích cực về người quản lý của họ, nhưng dữ liệu khá chung chung và nhóm muốn tìm hiểu thêm. Vì vậy, họ đào sâu hơn và chia dữ liệu thành 4 phần. Một phần tư chia các điểm dữ liệu thành bốn phần bằng nhau. Đây là nơi những thứ thực sự thú vị bắt đầu xảy ra. Các nhà phân tích dữ liệu đã phát hiện ra rằng có một sự khác biệt lớn giữa các phần tư trên cùng và dưới cùng. Hóa ra, các nhóm có người quản lý tốt nhất vui vẻ hơn, làm việc hiệu quả hơn và có nhiều khả năng muốn tiếp tục làm việc tại Google hơn. Điều này khẳng định rằng các nhà quản lý được đánh giá cao và tạo ra sự khác biệt lớn.*

But there was still more work to do. Just knowing that great managers create great results doesn't lead to actionable insights. You have to identify what exactly makes a great manager, so the team took two additional steps to collect more data. First, they launched an awards program where employees could nominate their favorite managers. For every submission you had to provide examples or data about what makes that manager great. The second step involved interviewing managers who were graphed on the top and bottom quartiles. This helped the analytics team see the differences between successful and less successful management behaviors. The best behaviors were identified as were the most common reasons for a manager needing improvement. The final step was sharing these insights and putting a procedure in place for evaluating managers with these qualities in mind. This data-driven decision continues to create an exceptional company culture for my colleagues and me. Thanks, data.

*Nhưng vẫn còn nhiều việc phải làm. Chỉ biết rằng những người quản lý tuyệt vời tạo ra kết quả tuyệt vời không dẫn đến những hiểu biết có thể hành động. Bạn phải xác định chính xác điều gì tạo nên một người quản lý tuyệt vời, vì vậy nhóm đã thực hiện thêm hai bước để thu thập thêm dữ liệu. Đầu tiên, họ đưa ra một chương trình giải thưởng để nhân viên có thể đề cử người quản lý yêu thích của họ. Đối với mỗi lần gửi, bạn phải cung cấp các ví dụ hoặc dữ liệu về điều gì làm cho người quản lý đó trở nên tuyệt vời. Bước thứ hai liên quan đến việc phỏng vấn các nhà quản lý, những người được xếp vào các nhóm trên cùng và dưới cùng. Điều này đã giúp nhóm phân tích thấy được sự khác biệt giữa các hành vi quản lý thành công và kém thành công. Các hành vi tốt nhất được xác định là những lý do phổ biến nhất khiến người quản lý cần cải thiện. Bước cuối cùng là chia sẻ những hiểu biết sâu sắc này và đưa ra một quy trình để đánh giá các nhà quản lý dựa trên những phẩm chất này. Quyết định dựa trên dữ liệu này tiếp tục tạo ra văn hóa công ty đặc biệt cho tôi và đồng nghiệp. Cảm ơn, dữ liệu.*

### Another interesting example comes from the nonprofit sector

Nonprofits are organizations dedicated to advancing a social cause or advocating for a particular effort, such as food security, education or the arts. In this case, data analysts researched how journalists can make a more meaningful impact for the nonprofits they would write about. Because journalists write for newspapers, magazines, and other news outlets, they can help nonprofits reach readers like you and me, who then take action to help nonprofits reach their goals. For instance, say you read about the problem of climate change in an online magazine. If the article is effective, you'll learn more about the cause and might even be compelled to make greener choices in your day-to-day life, volunteer for a nonprofit, or make a donation. That's an example of the journalist's work bringing about awareness, understanding, and engagement.

*Tổ chức phi lợi nhuận là các tổ chức dành riêng cho việc thúc đẩy sự nghiệp xã hội hoặc ủng hộ một nỗ lực cụ thể, chẳng hạn như an ninh lương thực, giáo dục hoặc nghệ thuật. Trong trường hợp này, các nhà phân tích dữ liệu đã nghiên cứu cách các nhà báo có thể tạo ra tác động có ý nghĩa hơn cho các tổ chức phi lợi nhuận mà họ sẽ viết về. Bởi vì các nhà báo viết báo, tạp chí và các hãng tin tức khác, họ có thể giúp các tổ chức phi lợi nhuận tiếp cận những độc giả như bạn và tôi, những người sau đó sẽ hành động để giúp các tổ chức phi lợi nhuận đạt được mục tiêu của họ. Chẳng hạn, giả sử bạn đọc về vấn đề biến đổi khí hậu trên một tạp chí trực tuyến. Nếu bài báo có hiệu quả, bạn sẽ tìm hiểu thêm về nguyên nhân và thậm chí có thể buộc phải đưa ra những lựa chọn xanh hơn trong cuộc sống hàng ngày của mình, tình nguyện cho một tổ chức phi lợi nhuận hoặc quyên góp. Đó là một ví dụ về công việc của nhà báo mang lại nhận thức, sự hiểu biết và sự tham gia.*

The data analysts used a tracker to monitor story topics, clicks, web traffic, comments, shares and more. Then they evaluated the information to make recommendations for how the journalists could do their jobs even better. In the end, they came up with some great ideas for how nonprofits and journalists can motivate people everywhere to work together and make the world a better place.

*Các nhà phân tích dữ liệu đã sử dụng trình theo dõi để theo dõi các chủ đề câu chuyện, số lần nhấp, lưu lượng truy cập web, nhận xét, lượt chia sẻ, v.v. Sau đó, họ đánh giá thông tin để đưa ra khuyến nghị về cách các nhà báo có thể làm công việc của họ tốt hơn nữa. Cuối cùng, họ đã đưa ra một số ý tưởng tuyệt vời về cách các tổ chức phi lợi nhuận và nhà báo có thể thúc đẩy mọi người ở khắp mọi nơi cùng nhau làm việc và biến thế giới thành một nơi tốt đẹp hơn.*

## Stages of the data life cycle

When you think about a life cycle, what's the first thing that comes to mind? One of the most well known examples of a life cycle is a butterfly. Butterflies begin as eggs, hatch into caterpillars and then become a chrysalis. That's where the real magic happens. Data has a life cycle of its own too. The life cycle of data is **plan**, **capture**, **manage**, **analyze**, **archive** and **destroy**.

*Khi bạn nghĩ về một vòng đời, điều đầu tiên bạn nghĩ đến là gì? Một trong những ví dụ nổi tiếng nhất về vòng đời là một con bướm. Bướm bắt đầu từ trứng, nở thành sâu bướm và sau đó trở thành nhộng. Đó là nơi phép màu thực sự xảy ra. Dữ liệu cũng có vòng đời của riêng nó. Vòng đời của dữ liệu là* ***lập kế hoạch****,* ***thu thập****,* ***quản lý****,* ***phân tích****,* ***lưu trữ*** *và* ***tiêu hủy****.*

### Planning

Let's start with the first phase, planning. This actually happens well before starting an analysis project. During planning, a business decides what kind of data it needs, how it will be managed throughout its life cycle, who will be responsible for it, and the optimal outcomes.

*Hãy bắt đầu với giai đoạn đầu tiên, lập kế hoạch. Điều này thực sự xảy ra tốt trước khi bắt đầu một dự án phân tích. Trong quá trình lập kế hoạch, doanh nghiệp quyết định loại dữ liệu nào họ cần, cách thức quản lý dữ liệu trong suốt vòng đời của nó, ai sẽ chịu trách nhiệm về dữ liệu đó và kết quả tối ưu.*

For example, let's say an electricity provider wanted to gain insights into how to save people energy. In the planning phase, they might decide to capture information on how much electricity its customers use each year, what types of buildings are being powered, and what types of devices are being powered inside of them. The electricity company would also decide which team members will be responsible for collecting, storing, and sharing that data. All of this happens during planning, and it helps set up the rest of the project.

*Ví dụ: giả sử một nhà cung cấp điện muốn hiểu rõ hơn về cách tiết kiệm năng lượng cho mọi người. Trong giai đoạn lập kế hoạch, họ có thể quyết định nắm bắt thông tin về lượng điện mà khách hàng sử dụng mỗi năm, loại tòa nhà nào đang được cấp điện và loại thiết bị nào đang được cấp điện bên trong chúng. Công ty điện lực cũng sẽ quyết định thành viên nào trong nhóm sẽ chịu trách nhiệm thu thập, lưu trữ và chia sẻ dữ liệu đó. Tất cả điều này xảy ra trong quá trình lập kế hoạch và nó giúp thiết lập phần còn lại của dự án.*

### Capture

The next phase is when you capture data. This is where data is collected from a variety of different sources and brought into the organization. With so much data being created everyday, the ways to collect it are truly endless. One common method is getting data from outside resources.

*Giai đoạn tiếp theo là khi bạn thu thập dữ liệu. Đây là nơi dữ liệu được thu thập từ nhiều nguồn khác nhau và được đưa vào tổ chức. Với rất nhiều dữ liệu được tạo ra hàng ngày, các cách để thu thập nó thực sự là vô tận. Một phương pháp phổ biến là lấy dữ liệu từ các nguồn bên ngoài.*

For example, if you were doing data analysis on weather patterns, you'd probably get data from a publicly available dataset like the National Climatic Data Center.

*Ví dụ: nếu bạn đang thực hiện phân tích dữ liệu về các kiểu thời tiết, có thể bạn sẽ lấy dữ liệu từ bộ dữ liệu có sẵn công khai như Trung tâm Dữ liệu Khí hậu Quốc gia.*

Another way to get data is from a company's own documents and files, which are usually stored inside a database. While we've mentioned databases before, we haven't gone into too much detail about what they are. A database is a collection of data stored in a computer system. In the case of our electricity provider, the business would probably measure data usage among its customers within a database that it owns. As a quick note, when you maintain a database of customer information, ensuring data integrity, credibility, and privacy are all important concerns.

*Một cách khác để lấy dữ liệu là từ các tài liệu và tệp của chính công ty, thường được lưu trữ bên trong cơ sở dữ liệu. Mặc dù chúng tôi đã đề cập đến cơ sở dữ liệu trước đây, nhưng chúng tôi chưa đi sâu vào chi tiết về chúng là gì. Cơ sở dữ liệu là một tập hợp dữ liệu được lưu trữ trong một hệ thống máy tính. Trong trường hợp nhà cung cấp điện của chúng tôi, doanh nghiệp có thể sẽ đo lường việc sử dụng dữ liệu giữa các khách hàng của mình trong cơ sở dữ liệu mà họ sở hữu. Lưu ý nhanh, khi bạn duy trì cơ sở dữ liệu thông tin khách hàng, việc đảm bảo tính toàn vẹn, độ tin cậy và quyền riêng tư của dữ liệu đều là những mối quan tâm quan trọng.*

### Manage

Now that we've captured our data, we'll move on to the next phase of the data life cycle, manage. Here we're talking about how we care for our data, how and where it's stored, the tools used to keep it safe and secure, and the actions taken to make sure that it's maintained properly. This phase is very important to data cleansing, which we'll cover later on.

*Bây giờ chúng tôi đã nắm bắt được dữ liệu của mình, chúng tôi sẽ chuyển sang giai đoạn tiếp theo của vòng đời dữ liệu, quản lý. Ở đây, chúng tôi đang nói về cách chúng tôi chăm sóc dữ liệu của mình, cách thức và vị trí dữ liệu được lưu trữ, các công cụ được sử dụng để giữ dữ liệu an toàn và bảo mật cũng như các hành động được thực hiện để đảm bảo dữ liệu được duy trì đúng cách. Giai đoạn này rất quan trọng đối với việc làm sạch dữ liệu mà chúng tôi sẽ đề cập sau.*

### Analyze

Next it's time to analyze your data. This is where data analysts really shine. In this phase, the data is used to solve problems, make great decisions, and support business goals. For example, one of our electricity company's goals might be to find ways to help customers save energy.

*Tiếp theo là thời gian để phân tích dữ liệu của bạn. Đây là nơi các nhà phân tích dữ liệu thực sự tỏa sáng. Trong giai đoạn này, dữ liệu được sử dụng để giải quyết vấn đề, đưa ra quyết định tuyệt vời và hỗ trợ các mục tiêu kinh doanh. Ví dụ, một trong những mục tiêu của công ty điện lực của chúng ta có thể là tìm cách giúp khách hàng tiết kiệm năng lượng.*

### Archive

Moving along the data life cycle now evolves to the archive phase. Archiving means storing data in a place where it's still available, but may not be used again. During analysis, analysts handle huge amounts of data. Can you imagine if we had to sort through all of the available data that's out there, even if it was no longer useful and relevant to our work? It makes way more sense to archive it than to keep it around.

*Tiếp theo trong vòng đời của dữ liệu là giai đoạn lưu trữ. Lưu trữ có nghĩa là lưu trữ dữ liệu tại một nơi nào đó, nơi nó vẫn có sẵn, nhưng có thể không được sử dụng lại. Trong giai đoạn phân tích, các nhà phân tích xử lý một lượng lớn dữ liệu. Bạn có thể tưởng tượng nếu chúng ta phải sắp xếp tất cả các dữ liệu có sẵn mà đã không còn hữu ích và liên quan đến công việc của chúng ta? Nó hiển nhiên rằng lưu trữ chúng là cách tốt hơn để giữ chúng.*

### Destroy

And finally, the last step of the data life cycle, the destroy phase. Yes, it sounds sad, but when you destroy data, it won't hurt a bit. So let's get back to our electricity provider example. They would have data stored on multiple hard drives. To destroy it, the company would use a secure data erasure software. If there were any paper files, they would be shredded too. This is important for protecting a company's private information, as well as private data about its customers. And there you have it, the data life cycle.

*Và cuối cùng, bước cuối cùng của vòng đời dữ liệu, giai đoạn phá hủy. Vâng, nghe có vẻ buồn, nhưng khi bạn hủy dữ liệu, nó sẽ không gây hại một chút nào. Vì vậy, hãy quay lại ví dụ về nhà cung cấp điện của chúng ta. Họ sẽ có dữ liệu được lưu trữ trên nhiều ổ cứng. Để phá hủy nó, công ty sẽ sử dụng một phần mềm xóa dữ liệu an toàn. Nếu có bất kỳ tệp giấy nào, chúng cũng sẽ bị cắt nhỏ. Điều này rất quan trọng để bảo vệ thông tin riêng tư của công ty, cũng như dữ liệu riêng tư về khách hàng của công ty. Và bạn đã có nó, vòng đời dữ liệu.*

## Variations of the data life cycle

You learned that there are six stages to the data life cycle. Here is a recap:

*Bạn đã học được rằng có sáu giai đoạn trong vòng đời dữ liệu. Đây là một bản tóm tắt:*

1. **Plan**: Decide what kind of data is needed, how it will be managed, and who will be responsible for it.

***Lập kế hoạch****: Quyết định loại dữ liệu nào là cần thiết, dữ liệu đó sẽ được quản lý như thế nào và ai sẽ chịu trách nhiệm về dữ liệu đó.*

1. **Capture**: Collect or bring in data from a variety of different sources.

***Thu thập****: Thu thập hoặc đưa vào dữ liệu từ nhiều nguồn khác nhau.*

1. **Manage**: Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.

***Quản lý****: Chăm sóc và duy trì dữ liệu. Điều này bao gồm việc xác định cách thức và vị trí lưu trữ cũng như các công cụ được sử dụng để làm như vậy.*

1. **Analyze**: Use the data to solve problems, make decisions, and support business goals.

***Phân tích****: Sử dụng dữ liệu để giải quyết vấn đề, đưa ra quyết định và hỗ trợ các mục tiêu kinh doanh.*

1. **Archive**: Keep relevant data stored for long-term and future reference.

***Lưu trữ****: Lưu giữ dữ liệu liên quan để tham khảo lâu dài và sau này.*

1. **Destroy**: Remove data from storage and delete any shared copies of the data.

***Hủy****: Xóa dữ liệu khỏi bộ lưu trữ và xóa mọi bản sao dữ liệu được chia sẻ.*

Warning: Be careful not to mix up or confuse **the six stages of the data life cycle (Plan, Capture, Manage, Analyze, Archive, and Destroy)** with **the six phases of the data analysis life cycle (Ask, Prepare, Process, Analyze, Share, and Act)**. They shouldn't be used or referred to interchangeably.

*Hãy cẩn thận để không trộn lẫn hoặc nhầm lẫn* ***sáu giai đoạn của vòng đời dữ liệu (Plan, Capture, Manage, Analyze, Archive và Destroy)*** *với* ***sáu giai đoạn của vòng đời phân tích dữ liệu (Ask, Prepare, Process, Analyze, Share và Act)****. Chúng không nên được sử dụng hoặc gọi thay thế cho nhau.*

The data life cycle provides a generic or common framework for how data is managed. You may recall that variations of the data analysis life cycle were described in [**Origins of the data analysis process**](#_Origins_of_The). The same can be done for the data life cycle. The rest of this reading provides a glimpse of how government, finance, and education institutions can view data life cycles a little differently.

*Vòng đời dữ liệu cung cấp một khuôn khổ chung nhất về cách quản lý dữ liệu. Bạn có thể nhớ lại rằng các biến thể của vòng đời phân tích dữ liệu đã được mô tả trong* [***Origins of the data analysis process***](#_Origins_of_The)*. Điều tương tự cũng có thể được thực hiện đối với vòng đời dữ liệu. Phần còn lại của bài đọc này cung cấp một cái nhìn thoáng qua về cách các tổ chức chính phủ, tài chính và giáo dục có thể xem các vòng đời của dữ liệu theo một cách hơi khác.*

### U.S. Fish and Wildlife Service

The U.S. Fish and Wildlife Service uses the following data life cycle:

1. Plan
2. Acquire
3. Maintain
4. Access
5. Evaluate
6. Archive

For more information, refer to [U.S. Fish and Wildlife's Data Management Life Cycle](https://www.fws.gov/data/life-cycle) page.

### The U.S. Geological Survey (USGS)

The USGS uses the data life cycle below:

1. Plan
2. Acquire
3. Process
4. Analyze
5. Preserve
6. Publish/Share

Several cross-cutting or overarching activities are also performed during each stage of their life cycle:

* Describe (metadata and documentation)
* Manage Quality
* Backup and Secure

For more information, refer to the [USGS Data Lifecycle](https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle) page.

### Financial institutions

Financial institutions may take a slightly different approach to the data life cycle as described in [The Data Life Cycle](https://sfmagazine.com/post-entry/july-2018-the-data-life-cycle/), an article in Strategic Finance magazine:

1. Capture
2. Qualify
3. Transform
4. Utilize
5. Report
6. Archive
7. Purge

### Harvard Business School (HBS)

One final data life cycle informed by Harvard University research has eight stages:

1. Generation
2. Collection
3. Processing
4. Storage
5. Management
6. Analysis
7. Visualization
8. Interpretation

For more information, refer to [8 Steps in the Data Life Cycle](https://online.hbs.edu/blog/post/data-life-cycle).

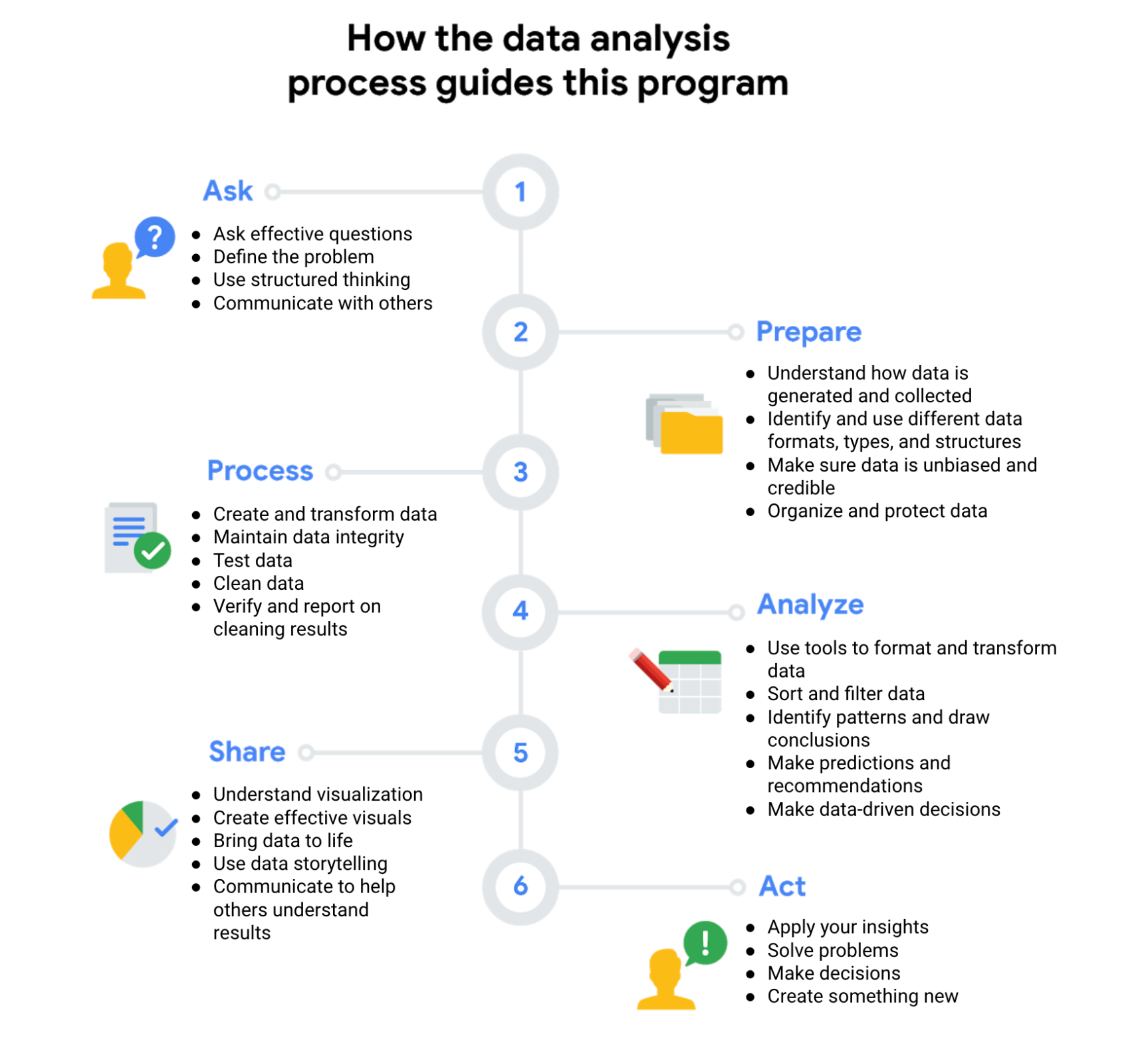
### Key takeaway

Understanding the importance of the data life cycle will set you up for success as a data analyst. Individual stages in the data life cycle will vary from company to company or by industry or sector. Historical data is important to both the U.S. Fish and Wildlife Service and the USGS, so their data life cycle focuses on archiving and backing up data. Harvard's interests are in research and teaching, so its data life cycle includes visualization and interpretation even though these are more often associated with a data analysis life cycle. The HBS data life cycle also doesn't call out a stage for purging or destroying data. In contrast, the data life cycle for finance clearly identifies archive and purge stages. To sum it up, although data life cycles vary, one data management principle is universal. Govern how data is handled so that it is accurate, secure, and available to meet your organization's needs.

## Six phases of data analysis

Now that you understand all the phases of the **Data Life Cycle**, it's time to move on to the phases of **Data Analysis**. They sound similar, but are two different things. Data analysis isn't a life cycle. It's the process of analyzing data. Coming up, we'll look at each step of the data analysis process and how it will relate to your work as a data analyst. Even this program is designed to follow these steps. Understanding these connections will help guide your own analysis and your work in this program. You've already learned that this program is modeled after the stages of the data analysis process. This program is split into courses, six of which are based upon the steps of data analysis: **Ask**, **Prepare**, **Process**, **Analyze**, **Share** and **Act**.

*Bây giờ bạn đã hiểu tất cả các giai đoạn của* ***Data Life Cycle (Vòng Đời Dữ Liệu),*** *đã đến lúc chuyển sang các giai đoạn* ***Data Analysis (PhânTích Dữ Liệu)****. Chúng nghe có vẻ giống nhau, nhưng là hai thứ khác nhau.* ***Data Analysis (PhânTích Dữ Liệu)*** *không phải là một* ***Data Life Cycle (Vòng Đời Dữ Liệu)****. Đó là quá trình phân tích dữ liệu. Sắp tới, chúng ta sẽ xem xét từng bước của quy trình phân tích dữ liệu và nó sẽ liên quan như thế nào đến công việc của bạn với tư cách là nhà phân tích dữ liệu. Ngay cả chương trình này cũng được thiết kế để thực hiện theo các bước này. Hiểu được các liên kết này sẽ giúp bạn định hướng phân tích của riêng mình và công việc của bạn trong chương trình này. Bạn đã biết rằng chương trình này được mô phỏng theo các giai đoạn của quy trình phân tích dữ liệu. Chương trình này được chia thành các khóa học, sáu trong số đó dựa trên các bước phân tích dữ liệu:* ***hỏi****,* ***chuẩn bị****,* ***xử lý****,* ***phân tích****,* ***chia sẻ*** *và* ***hành động****.*



### Ask

Let's start with the first step in data analysis, the ask phase. In this phase, we do two things. We define the problem to be solved and we make sure that we fully understand stakeholder expectations. Stakeholders hold a stake in the project. They are people who have invested time and resources into a project and are interested in the outcome.

*Hãy bắt đầu với bước đầu tiên trong phân tích dữ liệu, giai đoạn* ***Ask****. Trong giai đoạn này, chúng ta làm hai việc: Xác định vấn đề cần giải quyết và đảm bảo rằng chúng ta hiểu đầy đủ về mong muốn, kỳ vọng của các bên liên quan…*

Defining a problem means you look at the current state and identify how it's different from the ideal state. Usually there's an obstacle we need to get rid of or something wrong that needs to be fixed. For instance, a sports arena might want to reduce the time fans spend waiting in the ticket line. The obstacle is figuring out how to get the customers to their seats more quickly.

*Xác định một vấn đề có nghĩa là bạn nhìn vào trạng thái hiện tại và xác định nó khác với trạng thái lý tưởng như thế nào. Thông thường có một chướng ngại vật chúng ta cần loại bỏ hoặc một vài điều sai sót cần được sửa chữa. Ví dụ, một sân vận động thể thao có thể muốn giảm thời gian mà người hâm mộ phải đợi ở hàng mua vé. Chướng ngại vật là tìm cách đưa khách hàng đến chỗ ngồi của họ nhanh hơn.*

Another important part of the ask phase is understanding stakeholder expectations. The first step here is to determine who the stakeholders are. That may include your manager, an executive sponsor, or your sales partners. There can be lots of stakeholders. But what they all have in common is that they help make decisions, influence actions and strategies, and have specific goals they want to meet. They also care about the project and that's why it's so important to understand their expectations. For instance, if your manager assigns you a data analysis project related to business risk, it would be smart to confirm whether they want to include all types of risks that could affect the company, or just risks related to weather such as hurricanes and tornadoes.

*Một phần quan trọng khác của giai đoạn* ***Ask*** *là hiểu mong đợi, kỳ vọng của các bên liên quan. Bước đầu tiên ở đây là xác định được các bên liên quan là ai. Họ có thể bao gồm người quản lý của bạn, một nhà tài trợ điều hành hoặc các đối tác bán hàng của bạn. Và có thể có rất nhiều bên liên quan. Nhưng điểm chung của tất cả bọn họ là họ giúp đưa ra quyết định, tác động đến các hành động và chiến lược, đồng thời có những mục tiêu cụ thể mà họ muốn đạt được. Họ cũng quan tâm đến dự án và đó là lý do tại sao việc hiểu được kỳ vọng của họ lại quan trọng đến vậy. Ví dụ: nếu người quản lý của bạn giao cho bạn một dự án phân tích dữ liệu liên quan đến rủi ro kinh doanh, bạn nên xác nhận xem họ có muốn bao gồm tất cả các loại rủi ro có thể ảnh hưởng đến công ty hay chỉ những rủi ro liên quan đến thời tiết như bão và lốc xoáy.*

Communicating with your stakeholders is key in making sure you stay engaged and on track throughout the project. So as a data analyst, developing strong communication strategies is very important. This part of the ask phase helps you keep focused on the problem itself, not just its symptoms. The **5 Whys** are extremely helpful here. You'll learn how to ask effective questions and define the problem by working with stakeholders. You'll also cover strategies that can help you share what you discover in a way that keeps people interested.

*Giao tiếp với các bên liên quan là chìa khóa để đảm bảo bạn luôn tham gia và đi đúng hướng trong suốt dự án. Vì vậy, với tư cách là một nhà phân tích dữ liệu, việc phát triển các chiến lược truyền thông hiệu quả là rất quan trọng. Phần này của giai đoạn Ask giúp bạn tiếp tục tập trung vào chính vấn đề chứ không chỉ các triệu chứng của nó. 5 Whys cực kỳ hữu ích ở đây. Bạn sẽ học cách đặt câu hỏi hiệu quả và xác định vấn đề bằng cách làm việc với các bên liên quan. Bạn cũng sẽ đề cập đến các chiến lược có thể giúp bạn chia sẻ những gì bạn khám phá theo cách khiến mọi người quan tâm.*

### Prepare

After that, we'll move on to the prepare step of the data analysis process. This is where data analysts collect and store data they'll use for the upcoming analysis process. You'll learn more about the different types of data and how to identify which kinds of data are most useful for solving a particular problem. You'll also discover why it's so important that your data and results are objective and unbiased. In other words, any decisions made from your analysis should always be based on facts and be fair and impartial.

*Sau đó, chúng ta sẽ chuyển sang bước* ***Prepare*** *của quy trình phân tích dữ liệu. Đây là nơi các nhà phân tích dữ liệu thu thập và lưu trữ dữ liệu mà họ sẽ sử dụng cho quá trình phân tích sắp tới. Bạn sẽ tìm hiểu thêm về các loại dữ liệu khác nhau và cách xác định loại dữ liệu nào hữu ích nhất để giải quyết một vấn đề cụ thể. Bạn cũng sẽ khám phá ra lý do tại sao dữ liệu và kết quả của bạn phải khách quan và không thiên vị lại quan trọng đến vậy. Nói cách khác, bất kỳ quyết định nào được đưa ra từ phân tích của bạn phải luôn dựa trên thực tế và phải công bằng, không thiên vị.*

### Process

Next is the process step. Here, data analysts find and eliminate any errors and inaccuracies that can get in the way of results. This usually means cleaning data, transforming it into a more useful format, combining two or more datasets to make information more complete and removing outliers, which are any data points that could skew the information. After that, you'll learn how to check the data you prepare to make sure it's complete and correct. This phase is all about getting the details right. So you'll also fix typos, inconsistencies, or missing and inaccurate data. To top it off, you'll gain strategies for verifying and sharing your data cleansing with stakeholders.

*Tiếp theo là bước* ***Process****. Tại đây, các nhà phân tích dữ liệu tìm và loại bỏ mọi lỗi và điểm không chính xác có thể cản trở kết quả. Điều này thường có nghĩa là làm sạch dữ liệu, chuyển đổi dữ liệu thành định dạng hữu ích hơn, kết hợp hai hoặc nhiều bộ dữ liệu để làm cho thông tin đầy đủ hơn và loại bỏ các giá trị ngoại lai, tức là bất kỳ điểm dữ liệu nào có thể làm sai lệch thông tin. Sau đó, bạn sẽ học cách kiểm tra dữ liệu bạn chuẩn bị để đảm bảo dữ liệu đó hoàn chỉnh và chính xác. Giai đoạn này là tất cả về việc làm đúng các chi tiết. Vì vậy, bạn cũng sẽ sửa lỗi chính tả, sự không nhất quán hoặc dữ liệu bị thiếu và không chính xác. Trên hết, bạn sẽ có được các chiến lược để xác minh và chia sẻ quá trình làm sạch dữ liệu của mình với các bên liên quan.*

### Analyze

Then it's time to analyze. Analyzing the data you've collected involves using tools to transform and organize that information so that you can draw useful conclusions, make predictions, and drive informed decision-making. There are lots of powerful tools data analysts use in their work and in this course you'll learn about two of them, spreadsheets and structured query language, or SQL, which is often pronounced "sequel".

*Sau đó là thời gian để* ***Analyze****. Phân tích dữ liệu bạn đã thu thập liên quan đến việc sử dụng các công cụ để chuyển đổi và sắp xếp thông tin đó để bạn có thể rút ra kết luận hữu ích, đưa ra dự đoán và thúc đẩy quá trình ra quyết định sáng suốt. Có rất nhiều công cụ mạnh mẽ mà các nhà phân tích dữ liệu sử dụng trong công việc của họ và trong khóa học này, bạn sẽ tìm hiểu về hai trong số đó,* ***spreadsheets (bảng tính)*** *và* ***structured query language (ngôn ngữ truy vấn có cấu trúc)*** *hoặc* ***SQL****, thường được phát âm là* ***"sequel"****.*

### Share

The next course is based on the share phase. Here you'll learn how data analysts interpret results and share them with others to help stakeholders make effective data-driven decisions. In the share phase, visualization is a data analyst's best friend. So this course will highlight why visualization is essential to getting others to understand what your data is telling you. With the right visuals, facts and figures become so much easier to see and complex concepts become easier to understand. We'll explore different kinds of visuals and some great data visualization tools. You'll also practice your own presentation skills by creating compelling slideshows and learning how to be fully prepared to answer questions. Then we'll take a break from the data analysis process to show you all of the really cool things you can do with the programming language R. You don't need to be familiar with R or programming languages in general. Just know that R is a popular tool for data manipulation, calculation, and visualization.

*Khóa học tiếp theo dựa trên giai đoạn* ***Share****. Tại đây, bạn sẽ tìm hiểu cách các nhà phân tích dữ liệu giải thích kết quả và chia sẻ chúng với những người khác để giúp các bên liên quan đưa ra quyết định dựa trên dữ liệu hiệu quả. Trong giai đoạn chia sẻ, trực quan hóa là người bạn tốt nhất của nhà phân tích dữ liệu. Vì vậy, khóa học này sẽ nêu bật lý do tại sao trực quan hóa lại cần thiết để khiến người khác hiểu dữ liệu của bạn đang nói gì với bạn. Với hình ảnh phù hợp, các dữ kiện và số liệu trở nên dễ nhìn hơn rất nhiều và các khái niệm phức tạp trở nên dễ hiểu hơn. Chúng ta sẽ khám phá các loại hình ảnh khác nhau và một số công cụ trực quan hóa dữ liệu tuyệt vời. Bạn cũng sẽ thực hành các kỹ năng thuyết trình của riêng mình bằng cách tạo các trình chiếu hấp dẫn và học cách chuẩn bị đầy đủ để trả lời các câu hỏi. Sau đó, chúng tôi sẽ tạm dừng quá trình phân tích dữ liệu để cho bạn thấy tất cả những điều thực sự thú vị mà bạn có thể làm với ngôn ngữ lập trình R. Bạn không cần phải quen thuộc với R hoặc các ngôn ngữ lập trình nói chung. Chỉ cần biết rằng R là một công cụ phổ biến để thao tác, tính toán và trực quan hóa dữ liệu.*

### Act

For our final data analysis phase, we have Act. This is the exciting moment when the business takes all of the insights you, the data analyst, have provided and puts them to work in order to solve the original business problem and will be acting on what you've learned throughout this program. This is when you prepare for your job search and have the chance to complete a case study project. It's a great opportunity for you to bring together everything you've worked on throughout this course. Plus adding a case study to your portfolio helps you stand out from the other candidates when you interview for your first data analyst job.

*Đối với giai đoạn phân tích dữ liệu cuối cùng của chúng ta,* ***Act****. Đây là thời điểm thú vị khi doanh nghiệp sử dụng tất cả thông tin chi tiết mà bạn, nhà phân tích dữ liệu, đã cung cấp và vận dụng chúng để giải quyết vấn đề kinh doanh ban đầu và sẽ hành động dựa trên những gì bạn đã học được trong suốt chương trình này. Đây là lúc bạn chuẩn bị cho việc tìm kiếm việc làm của mình và có cơ hội hoàn thành một dự án nghiên cứu điển hình. Đó là một cơ hội tuyệt vời để bạn tập hợp mọi thứ bạn đã làm trong suốt khóa học này. Ngoài ra, việc thêm một nghiên cứu điển hình vào danh mục đầu tư của bạn sẽ giúp bạn nổi bật so với các ứng viên khác khi bạn phỏng vấn cho công việc phân tích dữ liệu đầu tiên của mình.*

### Example of the data process

The example that we are going to walk through is that of our employee engagement survey, but you could imagine that this process applies to just about any data analysis that you're going to conduct as an analyst.

*Ví dụ mà chúng tôi sẽ đi qua là cuộc khảo sát về mức độ gắn kết của nhân viên của chúng tôi, nhưng bạn có thể tưởng tượng rằng quy trình này chỉ áp dụng cho bất kỳ phân tích dữ liệu nào mà bạn sẽ thực hiện với tư cách là nhà phân tích.*

The first thing you want to do is **ask**. You want to ask all of the right questions at the beginning of the engagement so that you better understand what your leaders and stakeholders need from this analysis. The types of questions that I generally ask are around, what is the problem that we're trying to solve? What is the purpose of this analysis? What are we hoping to learn from it?

*Điều đầu tiên bạn muốn làm là* ***Ask (hỏi)****. Bạn muốn hỏi tất cả các câu hỏi phù hợp khi bắt đầu tham gia để bạn hiểu rõ hơn những gì các nhà lãnh đạo và các bên liên quan của bạn cần từ phân tích này. Các loại câu hỏi mà tôi thường hỏi xoay quanh vấn đề mà chúng tôi đang cố gắng giải quyết là gì? Mục đích của phân tích này là gì? Chúng tôi hy vọng học được gì từ nó?*

After you've asked all the right questions and you've wrapped your arms around the scope of the analysis you need to conduct, the next step is to **prepare**. We need to be thinking about what type of data we need to answer those key questions. This could be anything from quantitative data or qualitative data. It could be cross-sectional or points in time versus longitudinal over a long period of time. We need to be thinking about the type of data we need in order to answer the questions that we've set out to answer based on what we learned when we asked the right questions. We also need to be thinking about how we're going to collect that data or if we need to collect that data. It may be the case that we need to collect this data brand-new. So we need to think about what type of data we're going to be collecting and how. For our employee engagement survey, we do that via survey of both quantitative and qualitative questions. But it may actually be the case that for many analyses, the data that you're looking for already exist. Then it's a question of working with those data owners to make sure that you are able to leverage that data and use it responsibly.

*Sau khi bạn đã hỏi tất cả các câu hỏi phù hợp và bạn đã nắm được phạm vi phân tích mà bạn cần tiến hành, bước tiếp theo là* ***Prepare (chuẩn bị)****. Chúng tôi cần suy nghĩ về loại dữ liệu nào chúng tôi cần để trả lời những câu hỏi chính đó. Đây có thể là bất cứ thứ gì từ dữ liệu định lượng hoặc dữ liệu định tính. Nó có thể là mặt cắt ngang hoặc thời điểm so với chiều dọc trong một khoảng thời gian dài. Chúng tôi cần suy nghĩ về loại dữ liệu chúng tôi cần để trả lời các câu hỏi mà chúng tôi đã đặt ra để trả lời dựa trên những gì chúng tôi học được khi đặt câu hỏi đúng. Chúng tôi cũng cần suy nghĩ về cách chúng tôi sẽ thu thập dữ liệu đó hoặc liệu chúng tôi có cần thu thập dữ liệu đó hay không. Có thể xảy ra trường hợp chúng tôi cần thu thập dữ liệu hoàn toàn mới. Vì vậy, chúng tôi cần tư duy về loại dữ liệu nào chúng tôi sẽ thu thập và cách thức thu thập. Đối với khảo sát mức độ gắn kết của nhân viên, chúng tôi thực hiện điều đó thông qua khảo sát cả câu hỏi định lượng và định tính. Nhưng thực tế có thể xảy ra trường hợp đối với nhiều phân tích, dữ liệu mà bạn đang tìm kiếm đã tồn tại. Sau đó, vấn đề là làm việc với những chủ sở hữu dữ liệu đó để đảm bảo rằng bạn có thể tận dụng dữ liệu đó và sử dụng dữ liệu đó một cách có trách nhiệm.*

After you've done all the hard work to collect your data, now you need to **process** that data. It begins with cleaning. This to me is the most fun part of the data analytics process. We can think of it as the initial introduction or the handshake, hello, to your data. This is where you get a chance to understand its structure, its quirks, its nuances, and you really get a chance to understand deeply what type of data you're going to be working with and understanding what potential that data has to answer all of your questions. This is such an important part, too, where we're running through all of our quality assurance checks. For example, do we have all of the data that we anticipated we would have? Are we missing data at random or is it missing in a systematic way such that maybe something went wrong with our data collection effort? If needed, did we code all of our data the right way? Are there any outliers that we need to treat differently? This is the part where we spend a lot of time really digging deeply into the structure and nuance of the data to make sure that you're able to analyze it appropriately and responsibly.

*Sau khi bạn đã thực hiện tất cả các công việc khó khăn để thu thập dữ liệu của mình, bây giờ bạn cần* ***Process (xử lý)*** *dữ liệu đó. Nó bắt đầu với việc làm sạch. Đối với tôi, đây là phần thú vị nhất của quá trình phân tích dữ liệu. Chúng tôi có thể coi đó là mở đầu hoặc cái bắt tay, xin chào, với dữ liệu của bạn. Đây là nơi bạn có cơ hội hiểu cấu trúc, đặc điểm, sắc thái của dữ liệu và bạn thực sự có cơ hội hiểu sâu về loại dữ liệu bạn sẽ làm việc và hiểu tiềm năng của dữ liệu đó để trả lời tất cả các những câu hỏi mà bạn đã đặt ra trước đó. Đây cũng là một phần quan trọng, nơi chúng tôi đang thực hiện tất cả các kiểm tra đảm bảo chất lượng. Ví dụ: chúng tôi có tất cả dữ liệu mà chúng tôi dự đoán sẽ có không? Chúng tôi đang thiếu dữ liệu một cách ngẫu nhiên hay dữ liệu bị thiếu một cách có hệ thống đến mức có thể đã xảy ra sự cố với nỗ lực thu thập dữ liệu của chúng tôi? Nếu cần, chúng tôi đã mã hóa tất cả dữ liệu của mình đúng cách chưa? Có bất kỳ ngoại lệ nào mà chúng ta cần đối xử khác đi không? Đây là phần mà chúng tôi dành nhiều thời gian thực sự đào sâu vào cấu trúc và sắc thái của dữ liệu để đảm bảo rằng bạn có thể phân tích dữ liệu một cách thích hợp và có trách nhiệm.*

After cleaning our data and running all of our quality assurance checks, now is the point where we **analyze** our data, making sure to do so in as objective and unbiased a way as possible. To do this, the first thing we do is run through a series of analyses that we've already planned ahead of time based on the questions that we know we want to answer from the very, very beginning of the process. One thing that's probably the hardest about this particular process, the hardest thing about analyzing data, is that we as analysts are trained to look for patterns. Over time as we become better and better at our jobs, what we'll often find is that we can start to intuit what we might see in the data. We might have a sneaking suspicion as to what the data are going to tell us. This is the point where we have to take a step back and let the data speak for itself. As data analysts, we are storytellers, but we also have to keep in mind that it is not our story to tell. That story belongs to the data, and it is our job as analysts to amplify and tell that story in as unbiased and objective a way as possible.

*Sau khi làm sạch dữ liệu của chúng tôi và chạy tất cả các kiểm tra đảm bảo chất lượng, giờ là lúc chúng tôi* ***Analyze (phân tích)*** *dữ liệu của mình, đảm bảo làm như vậy theo cách khách quan và không thiên vị nhất có thể. Để làm điều này, điều đầu tiên chúng tôi làm là thực hiện một loạt phân tích mà chúng tôi đã lên kế hoạch trước dựa trên những câu hỏi mà chúng tôi biết rằng mình muốn trả lời ngay từ đầu của quá trình. Một điều có lẽ khó nhất trong quy trình cụ thể này, điều khó nhất khi phân tích dữ liệu, là chúng tôi với tư cách là nhà phân tích được đào tạo để tìm kiếm các mẫu. Theo thời gian khi chúng ta ngày càng trở nên giỏi hơn trong công việc của mình, điều mà chúng ta thường nhận thấy là chúng ta có thể bắt đầu trực cảm những gì chúng ta có thể thấy trong dữ liệu. Chúng tôi có thể có một sự nghi ngờ lén lút về những gì dữ liệu sẽ cho chúng tôi biết. Đây là điểm mà chúng ta phải lùi lại một bước và để dữ liệu tự nói lên điều đó. Là nhà phân tích dữ liệu, chúng tôi là những người kể chuyện, nhưng chúng tôi cũng phải nhớ rằng đó không phải là câu chuyện của chúng tôi để kể. Câu chuyện đó thuộc về dữ liệu và công việc của chúng tôi với tư cách là nhà phân tích là khuếch đại và kể câu chuyện đó theo cách khách quan và khách quan nhất có thể.*

The next step is to **share** all of the data and insights that you've generated from your analyses. Now typically for employee engagement survey, we start by sharing the high-level findings with our executive team. We want them to have a landscape view of how the organization is feeling, and we want to make sure that there aren't any surprises as they dig deeper and deeper into the data to understand how teams are feeling and how individual employees are feeling.

*Bước tiếp theo là* ***Share (chia sẻ)*** *tất cả dữ liệu và thông tin chi tiết mà bạn đã tạo từ các phân tích của mình. Hiện nay, thông thường đối với khảo sát về mức độ gắn kết của nhân viên, chúng tôi bắt đầu bằng cách chia sẻ những phát hiện cấp cao với nhóm điều hành của mình. Chúng tôi muốn họ có cái nhìn bao quát về cảm giác của tổ chức và chúng tôi muốn đảm bảo rằng không có bất kỳ điều gì bất ngờ khi họ đào sâu hơn nữa vào dữ liệu để hiểu cảm giác của các nhóm và cảm giác của từng nhân viên.*

All of this work from asking the right questions to collecting your data, to analyzing and sharing, doesn't mean much of anything if we aren't taking **act**ion on what we've just learned. This to me is the most critical part, especially of our employee engagement survey. I like to say that the survey is actually the easy part, and acting on the results is really where the real work begins. This is where we use all of those data-driven insights to decide what types of interventions we want to introduce, not only at the organizational level, but also at the team level as well. We might find, for example, that the organization is working on a series of interventions to help improve part of the employee experience, whereas individual teams have additional roles, responsibilities to play, to either bolster some of those efforts or to introduce new ones to better meet their team where their strengths and opportunity areas are.

*Tất cả những công việc này từ việc đặt câu hỏi phù hợp đến thu thập dữ liệu của bạn, đến phân tích và chia sẻ, sẽ không có ý nghĩa gì nhiều nếu chúng ta không thực hiện act (hành động) đối với những gì chúng ta vừa học được. Đối với tôi, đây là phần quan trọng nhất, đặc biệt là trong cuộc khảo sát về mức độ gắn kết của nhân viên của chúng tôi. Tôi muốn nói rằng cuộc khảo sát thực sự là một phần dễ dàng và hành động dựa trên kết quả thực sự là nơi công việc thực sự bắt đầu. Đây là nơi chúng tôi sử dụng tất cả những hiểu biết dựa trên dữ liệu đó để quyết định loại can thiệp nào chúng tôi muốn giới thiệu, không chỉ ở cấp độ tổ chức mà còn ở cấp độ nhóm. Ví dụ, chúng ta có thể thấy rằng tổ chức đang thực hiện một loạt các biện pháp can thiệp để giúp cải thiện một phần trải nghiệm của nhân viên, trong khi các nhóm riêng lẻ có thêm vai trò, trách nhiệm để thực hiện, để củng cố một số nỗ lực đó hoặc giới thiệu những nỗ lực mới cho gặp gỡ nhóm của họ tốt hơn ở những điểm mạnh và lĩnh vực cơ hội của họ.*

The data analysis process is rigorous, but it is lengthy. I can completely appreciate that we as data analysts, get so excited about just diving right into the data and doing what we do best. The challenge is that if we don't work through the process in its entirety, if we try to skip steps, we're not going to be able to elicit the insights that we're looking for. I absolutely love my job. I have such a deep appreciation for data and what it can do and what type of insight we can derive from it.

*Quá trình phân tích dữ liệu là nghiêm ngặt, nhưng nó kéo dài. Tôi hoàn toàn có thể đánh giá cao rằng chúng tôi với tư cách là nhà phân tích dữ liệu, rất hào hứng với việc đi sâu vào dữ liệu và làm những gì chúng tôi làm tốt nhất. Thách thức là nếu chúng ta không thực hiện toàn bộ quy trình, nếu chúng ta cố bỏ qua các bước, thì chúng ta sẽ không thể gợi ra những hiểu biết sâu sắc mà chúng ta đang tìm kiếm. Tôi rất thích công việc của mình. Tôi đánh giá rất cao dữ liệu và những gì nó có thể làm cũng như loại hiểu biết sâu sắc nào mà chúng ta có thể rút ra từ nó.*

## Exploring Data Analyst Tools

There are tons of options out there. But the most common ones you'll see analyst use are **spreadsheets**, **query languages** and **visualization tools**. And this is going to give you a quick look at how these tools are being used by data analysts everyday. Believe it or not, I was several years into my accounting and finance career before I saw all of these tools working together. At that point I was very experienced with spreadsheets, and had worked in large data sets with some of the traditional database programs. I had the foundational skill set to use query languages, and I had dabbled in visualizations, but I had never brought them all together. Then I got hired here at Google. And it was so eye-opening to come into a place like this with an abundance of information everywhere you look. As an analyst at Google, the true power of these tools became so much clearer to me. I became more focused on really maximizing everything these tools could do, streamlining my reporting and just making my work simpler. All of the sudden, I had a lot more time and space to dedicate to identifying new problems to solve and driving decision-making. Without a doubt, once you've learned the power of these tools, you will be well on your way to becoming the best data analyst you can possibly be. All right, I hope that story has you even more motivated for this course.

*Có rất nhiều lựa chọn ngoài kia. Nhưng những thứ phổ biến nhất mà bạn sẽ thấy nhà phân tích sử dụng là* ***spreadsheets (bảng tính)****,* ***query languages (ngôn ngữ truy vấn)*** *và* ***visualization tools (công cụ trực quan)m****. Và điều này sẽ cung cấp cho bạn cái nhìn nhanh về cách các công cụ này đang được các nhà phân tích dữ liệu sử dụng hàng ngày. Dù bạn có tin hay không thì tùy, tôi đã có vài năm trong sự nghiệp kế toán và tài chính của mình trước khi tôi thấy tất cả những công cụ này hoạt động cùng nhau. Vào thời điểm đó, tôi đã có rất nhiều kinh nghiệm với bảng tính và đã làm việc trong các tập dữ liệu lớn với một số chương trình cơ sở dữ liệu truyền thống. Tôi đã có bộ kỹ năng cơ bản để sử dụng các ngôn ngữ truy vấn và tôi đã tìm hiểu kỹ về trực quan hóa, nhưng tôi chưa bao giờ kết hợp tất cả chúng lại với nhau. Sau đó, tôi được thuê tại Google. Và thật là mở mang tầm mắt khi đến một nơi như thế này với vô số thông tin ở mọi nơi bạn nhìn thấy. Là một nhà phân tích tại Google, sức mạnh thực sự của những công cụ này trở nên rõ ràng hơn rất nhiều đối với tôi. Tôi trở nên tập trung hơn vào việc thực sự tối đa hóa mọi thứ mà những công cụ này có thể làm, hợp lý hóa báo cáo của tôi và chỉ làm cho công việc của tôi trở nên đơn giản hơn. Đột nhiên, tôi có nhiều thời gian và không gian hơn để dành cho việc xác định các vấn đề mới cần giải quyết và thúc đẩy quá trình ra quyết định. Không còn nghi ngờ gì nữa, một khi bạn đã học được sức mạnh của những công cụ này, bạn sẽ tiếp tục trở thành nhà phân tích dữ liệu giỏi nhất có thể. Được rồi, tôi hy vọng câu chuyện đó đã giúp bạn có thêm động lực cho khóa học này.*

**Key data analyst tools**

**Spreadsheets**

Data analysts rely on spreadsheets to collect and organize data. Two popular spreadsheet applications you will probably use a lot in your future role as a data analyst are Microsoft Excel and Google Sheets. Spreadsheets structure data in a meaningful way by letting you:

* Collect, store, organize, and sort information
* Identify patterns and piece the data together in a way that works for each specific data project
* Create excellent data visualizations, like graphs and charts.

**Databases and query languages**

A database is a collection of structured data stored in a computer system. Some popular Structured Query Language (SQL) programs include MySQL, Microsoft SQL Server, and BigQuery.

* Allow analysts to isolate specific information from a database(s)
* Make it easier for you to learn and understand the requests made to databases
* Allow analysts to select, create, add, or download data from a database for analysis

**Visualization tools**

Data analysts use a number of visualization tools, like graphs, maps, tables, charts, and more. Two popular visualization tools are Tableau and Looker. These tools:

* Turn complex numbers into a story that people can understand
* Help stakeholders come up with conclusions that lead to informed decisions and effective business strategies
* Have multiple features

**Tableau**'s simple drag-and-drop feature lets users create interactive graphs in dashboards and worksheets

**Looker** communicates directly with a database, allowing you to connect your data right to the visual tool you choose

A career as a data analyst also involves using programming languages, like R and Python, which are used a lot for statistical analysis, visualization, and other data analysis.

**Choosing the right tool for the job**

As a data analyst, you will usually have to decide which program or solution is right for the particular project you are working on. In this reading, you will learn more about how to choose which tool you need and when.

*Là một nhà phân tích dữ liệu, bạn thường sẽ phải quyết định chương trình hoặc giải pháp nào phù hợp với dự án cụ thể mà bạn đang thực hiện. Trong bài đọc này, bạn sẽ tìm hiểu thêm về cách chọn công cụ nào bạn cần và khi nào.*

Depending on which phase of the data analysis process you’re in, you will need to use different tools. For example, if you are focusing on creating complex and eye-catching visualizations, then the visualization tools we discussed earlier are the best choice. But if you are focusing on organizing, cleaning, and analyzing data, then you will probably be choosing between spreadsheets and databases using queries. Spreadsheets and databases both offer ways to store, manage, and use data. The basic content for both tools are sets of values. Yet, there are some key differences, too:

*Tùy thuộc vào giai đoạn của quá trình phân tích dữ liệu mà bạn đang tham gia, bạn sẽ cần sử dụng các công cụ khác nhau. Ví dụ: nếu bạn đang tập trung vào việc tạo các hình ảnh trực quan phức tạp và bắt mắt, thì các công cụ trực quan mà chúng tôi đã thảo luận trước đó là lựa chọn tốt nhất. Nhưng nếu bạn đang tập trung vào việc sắp xếp, dọn dẹp và phân tích dữ liệu thì có thể bạn sẽ phải chọn giữa bảng tính và cơ sở dữ liệu bằng truy vấn. Cả bảng tính và cơ sở dữ liệu đều cung cấp các cách để lưu trữ, quản lý và sử dụng dữ liệu. Nội dung cơ bản của cả hai công cụ là các bộ giá trị. Tuy nhiên, cũng có một số khác biệt chính:*

| **Spreadsheets** | **Databases** |
| --- | --- |
| Software applications | Data stores - accessed using a query language (e.g. SQL) |
| Structure data in a row and column format | Structure data using rules and relationships |
| Organize information in cells | Organize information in complex collections |
| Provide access to a limited amount of data | Provide access to huge amounts of data |
| Manual data entry | Strict and consistent data entry |
| Generally one user at a time | Multiple users |
| Controlled by the user | Controlled by a database management system |

You don’t have to choose one or the other because each serves its own purpose. Generally, data analysts work with a combination of the two, as both tools are very useful in data analytics. For example, you can store data in a database, then export it to a spreadsheet for analysis. Or, if you are collecting information in a spreadsheet, and it becomes too much for that particular platform, you can import it into a database.

*Bạn không cần phải chọn cái này hay cái kia vì mỗi cái phục vụ mục đích riêng của nó. Nói chung, các nhà phân tích dữ liệu làm việc với sự kết hợp của cả hai, vì cả hai công cụ đều rất hữu ích trong phân tích dữ liệu. Ví dụ: bạn có thể lưu trữ dữ liệu trong cơ sở dữ liệu, sau đó xuất dữ liệu đó sang bảng tính để phân tích. Hoặc, nếu bạn đang thu thập thông tin trong một bảng tính và thông tin đó trở nên quá nhiều đối với nền tảng cụ thể đó, thì bạn có thể nhập thông tin đó vào cơ sở dữ liệu.*

### Spreadsheets

There are lots of different spreadsheet solutions, but two popular options are Microsoft Excel and Google Sheets. To put it simply, a spreadsheet is a digital worksheet. It stores, organizes, and sorts data. This is important because the usefulness of your data depends on how well it's structured. When you put your data into a spreadsheet, you can see patterns, group information and easily find the information you need. Spreadsheets also have some really useful features called formulas and functions. A formula is a set of instructions that performs a specific calculation using the data in a spreadsheet. Formulas can do basic things like add, subtract, multiply and divide, but they don't stop there. You can also use formulas to find the average of a number set. Look up a particular value, return the sum of a set of values that meets a particular rule, and so much more. A function is a preset command that automatically performs a specific process or task using the data in a spreadsheet. That sounds pretty technical, I know, so let's break it down. Just think of a function as a simpler, more efficient way of doing something that would normally take a lot of time. In other words, functions can help make you more efficient. Those are the spreadsheet basics for now. Later on, you'll see them in action and start working with spreadsheets yourself.

*Có rất nhiều giải pháp bảng tính khác nhau nhưng hai tùy chọn phổ biến là Microsoft Excel và Google Trang tính. Nói một cách đơn giản, bảng tính là một bảng tính kỹ thuật số. Nó lưu trữ, sắp xếp và sắp xếp dữ liệu. Điều này rất quan trọng vì tính hữu ích của dữ liệu của bạn phụ thuộc vào mức độ cấu trúc của dữ liệu đó. Khi đưa dữ liệu của mình vào bảng tính, bạn có thể xem các mẫu, nhóm thông tin và dễ dàng tìm thấy thông tin mình cần. Bảng tính cũng có một số tính năng thực sự hữu ích được gọi là formulas (công thức) và functions (hàm). Công thức là một tập hợp các hướng dẫn thực hiện một phép tính cụ thể bằng cách sử dụng dữ liệu trong bảng tính. Các công thức có thể làm những việc cơ bản như cộng, trừ, nhân và chia, nhưng chúng không dừng lại ở đó. Bạn cũng có thể sử dụng công thức để tìm giá trị trung bình của một tập hợp số. Tra cứu một giá trị cụ thể, trả về tổng của một tập hợp các giá trị đáp ứng một quy tắc cụ thể, v.v. Hàm là một lệnh đặt trước tự động thực hiện một quy trình hoặc tác vụ cụ thể bằng cách sử dụng dữ liệu trong bảng tính. Điều đó nghe có vẻ khá kỹ thuật, tôi biết, vì vậy hãy chia nhỏ nó ra. Chỉ cần nghĩ về một chức năng như một cách đơn giản hơn, hiệu quả hơn để làm một việc gì đó mà thông thường sẽ mất rất nhiều thời gian. Nói cách khác, các chức năng có thể giúp bạn làm việc hiệu quả hơn. Đó là những điều cơ bản về bảng tính cho bây giờ. Sau này, bạn sẽ thấy chúng hoạt động và tự mình bắt đầu làm việc với bảng tính.*

### Query Language

A query language is a computer programming language that allows you to retrieve and manipulate data from a database. You'll learn something called structured query language, more commonly known as SQL. SQL is a language that lets data analysts communicate with a database. A database is a collection of data stored in a computer system. SQL is the most widely used structured query language for a couple of reasons. It's easy to understand and works very well with all kinds of databases. With SQL, data analysts can access the data they need by making a query. Although query means question, I like to think of it as more of a request. So you're requesting that the database do something for you. You can ask it to do a lot of different things such as insert, delete, select or update data. Okay, that's a top level look at SQL.

*Ngôn ngữ truy vấn là ngôn ngữ lập trình máy tính cho phép bạn truy xuất và thao tác dữ liệu từ cơ sở dữ liệu. Bạn sẽ học một thứ gọi là ngôn ngữ truy vấn có cấu trúc, thường được gọi là SQL. SQL là ngôn ngữ cho phép các nhà phân tích dữ liệu giao tiếp với cơ sở dữ liệu. Cơ sở dữ liệu là một tập hợp dữ liệu được lưu trữ trong một hệ thống máy tính. SQL là ngôn ngữ truy vấn có cấu trúc được sử dụng rộng rãi nhất vì một vài lý do. Thật dễ hiểu và hoạt động rất tốt với tất cả các loại cơ sở dữ liệu. Với SQL, các nhà phân tích dữ liệu có thể truy cập dữ liệu họ cần bằng cách thực hiện truy vấn. Mặc dù truy vấn có nghĩa là câu hỏi, nhưng tôi thích coi nó giống như một yêu cầu hơn. Vì vậy, bạn đang yêu cầu cơ sở dữ liệu làm điều gì đó cho bạn. Bạn có thể yêu cầu nó thực hiện rất nhiều việc khác nhau như chèn, xóa, chọn hoặc cập nhật dữ liệu. Được rồi, đó là cái nhìn cấp cao nhất về SQL.*

### Data Visualization

You've learned that data visualization is the graphical representation of information. Some examples include graphs, maps, and tables. Most people process visuals more easily than words alone. That's why visualizations are so important. They help data analysts communicate their insights to others, in an effective and compelling way. When you think about the data analysis process, after data is prepared, processed and analyzed, the insights are visualized so it can be understood and shared. This makes it easier for stakeholders to draw conclusions, make decisions, and come up with strategies. Some popular visualization tools are Tableau and Looker. Data analysts like using Tableau because it helps them create visuals that are very easy to understand. This means that even non-technical users can get the information they need. Looker is also popular with data analysts because it gives them an easy way to create visuals based on the results of a query. With Looker, you can give stakeholders a complete picture of your work by showing them visualization data and the actual data related to it. All visualization tools have great features that are useful in different situations.

*Bạn đã học được rằng trực quan hóa dữ liệu là biểu diễn đồ họa của thông tin. Một số ví dụ bao gồm đồ thị, bản đồ và bảng. Hầu hết mọi người xử lý hình ảnh dễ dàng hơn so với chỉ từ ngữ. Đó là lý do tại sao hình dung rất quan trọng. Chúng giúp các nhà phân tích dữ liệu truyền đạt những hiểu biết của họ cho người khác theo cách hiệu quả và hấp dẫn. Khi bạn nghĩ về quy trình phân tích dữ liệu, sau khi dữ liệu được chuẩn bị, xử lý và phân tích, thông tin chi tiết sẽ được trực quan hóa để có thể hiểu và chia sẻ. Điều này giúp các bên liên quan dễ dàng đưa ra kết luận, đưa ra quyết định và đưa ra các chiến lược. Một số công cụ trực quan phổ biến là Tableau và Looker. Các nhà phân tích dữ liệu thích sử dụng Tableau vì nó giúp họ tạo ra các hình ảnh rất dễ hiểu. Điều này có nghĩa là ngay cả những người dùng không có kỹ thuật cũng có thể nhận được thông tin họ cần. Looker cũng phổ biến với các nhà phân tích dữ liệu vì nó giúp họ dễ dàng tạo hình ảnh trực quan dựa trên kết quả của truy vấn. Với Looker, bạn có thể cung cấp cho các bên liên quan một bức tranh hoàn chỉnh về công việc của mình bằng cách hiển thị cho họ dữ liệu trực quan hóa và dữ liệu thực tế liên quan đến nó. Tất cả các công cụ trực quan đều có các tính năng tuyệt vời hữu ích trong các tình huống khác nhau.*

## More Spreadsheet Resources

In the spirit of lifelong learning, it is good to have resources to turn to when you want to know more about using spreadsheets. Two of the most well known and used spreadsheet platforms are Google Sheets and Microsoft Excel. Both provide free online training resources that you can access anytime you need them. Bookmark these links if you want to access them later.

### [Google Sheets Training and Help](https://support.google.com/a/users/answer/9282959?visit_id=637361702049227170-1815413770&rd=1)

Learn even more ways to move, store, and analyze your data with the Google Sheets Training and Help page, located in the Google Workspace Learning Center. This hub offers an expanded list of tips, from beginner to advanced, along with cheat sheets, templates, guides, and tutorials.

### [Google Sheets Cheat Sheet](https://support.google.com/a/users/answer/9300022)

Want to learn more about Google Sheets? This online help article features a short list of the most important functions you will use, including rows, columns, cells, and functions.

### [Microsoft Excel for Windows Training](https://support.microsoft.com/en-us/office/excel-for-windows-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb)

Get to know Excel spreadsheets a little better by visiting this free online training center. Offering everything from a quick-start guide and introduction to tutorials and templates, you will find everything you need to know, all in one place.

## SQL Guide: Getting started

Just as humans use different languages to communicate with others, so do computers. **Structured Query Language** (or **SQL**, often pronounced “sequel”) enables data analysts to talk to their databases. SQL is one of the most useful data analyst tools, especially when working with large datasets in tables. It can help you investigate huge databases, track down text (referred to as strings) and numbers, and filter for the exact kind of data you need—much faster than a spreadsheet can.

If you haven’t used SQL before, this reading will help you learn the basics so you can appreciate how useful SQL is and how useful SQL queries are in particular. You will be writing SQL queries in no time at all.

### What is a query?

A **query** is a request for data or information from a database. When you query databases, you use SQL to communicate your question or request. You and the database can always exchange information as long as you speak the same language.

Every programming language, including SQL, follows a unique set of guidelines known as **syntax**. **Syntax** is the predetermined structure of a language that includes all required words, symbols, and punctuation, as well as their proper placement. As soon as you enter your search criteria using the correct syntax, the query starts working to pull the data you’ve requested from the target database.

The syntax of every SQL query is the same:

* Use **SELECT** to choose the columns you want to return.
* Use **FROM** to choose the tables where the columns you want are located.
* Use **WHERE** to filter for certain information.

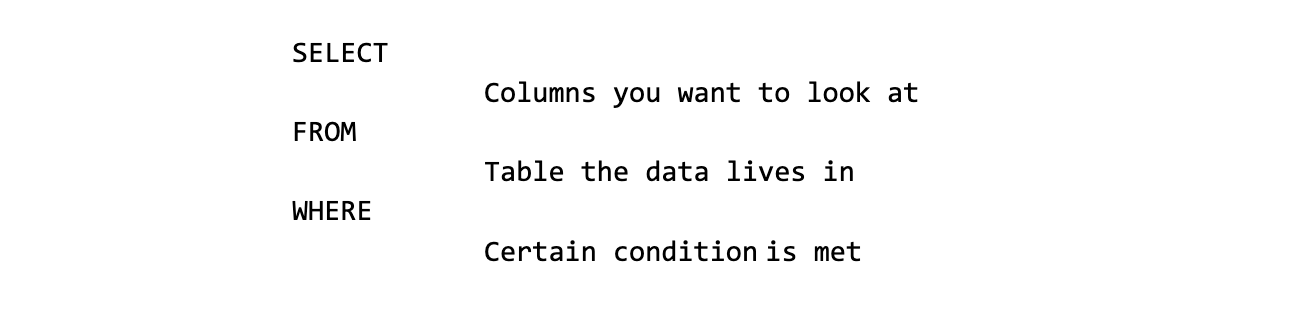
A SQL query is like filling in a template. You will find that if you are writing a SQL query from scratch, it is helpful to start a query by writing the SELECT, FROM, and WHERE keywords in the following format:

**SELECT**

**FROM**

**WHERE**

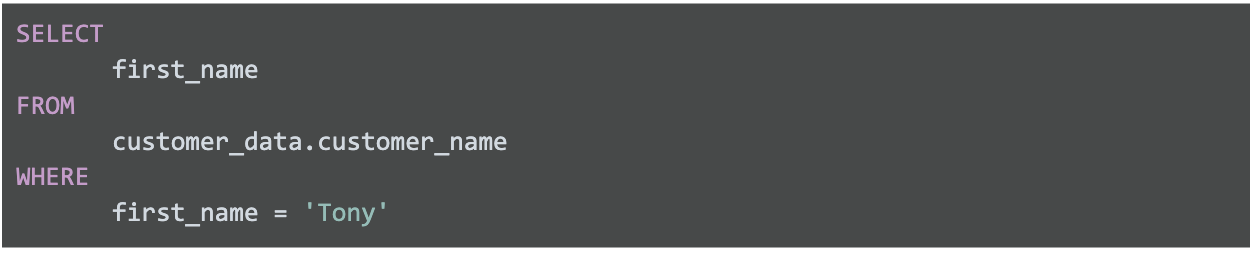
Next, enter the table name after the **FROM**; the table columns you want after the **SELECT**; and, finally, the conditions you want to place on your query after the **WHERE**. Make sure to add a new line and indent when adding these, as shown below:



Following this method each time makes it easier to write SQL queries. It can also help you make fewer syntax errors.

### Example of a query

Here is how a simple query would appear in BigQuery, a data warehouse on the Google Cloud Platform.



The above query uses three commands to locate customers with the first name Tony:

1. **SELECT** the column named **first\_name**
2. **FROM** a table named **customer\_name** (in a dataset named **customer\_data**) (The dataset name is always followed by a dot, and then the table name.)
3. But only return the data **WHERE** the first\_name is **Tony**

The results from the query might be similar to the following:

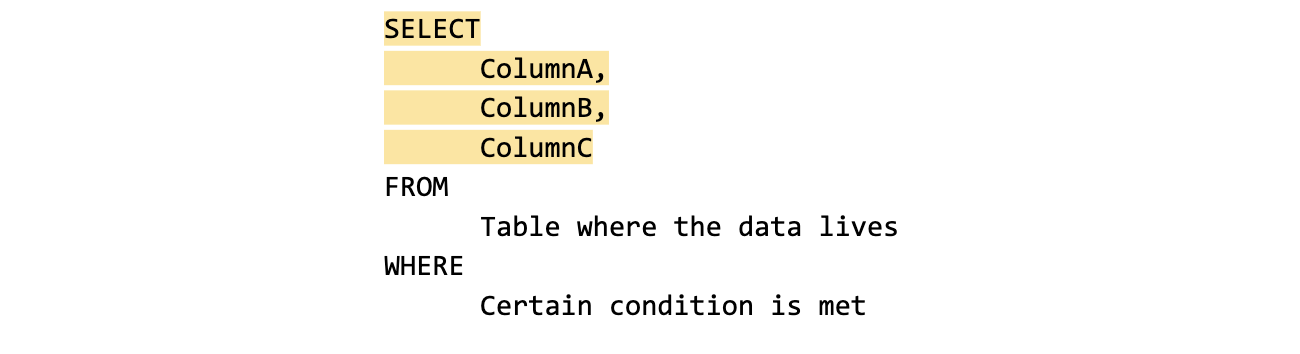
| **first\_name** |
| --- |
| Tony |
| Tony |
| Tony |

As you can conclude, this query had the correct syntax, but wasn't very useful after the data was returned.

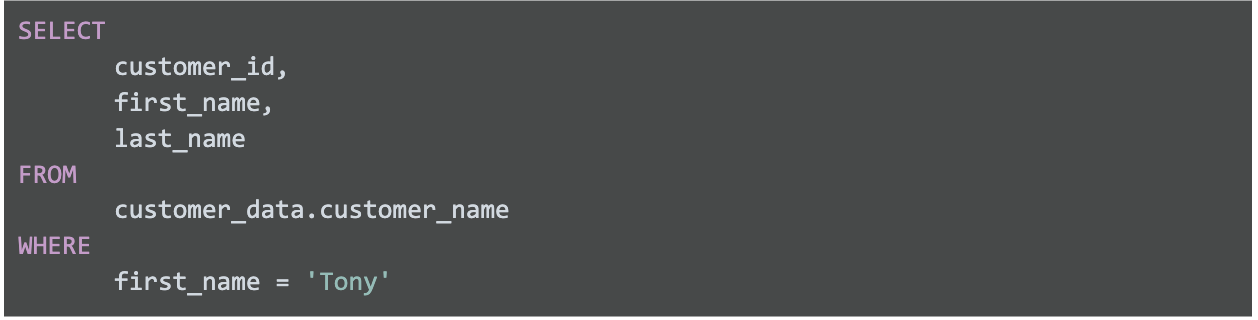
### Multiple columns in a query

In real life, you will need to work with more data beyond customers named Tony. Multiple columns that are chosen by the same SELECT command can be indented and grouped together.

If you are requesting multiple data fields from a table, you need to include these columns in your SELECT command. Each column is separated by a comma as shown below:



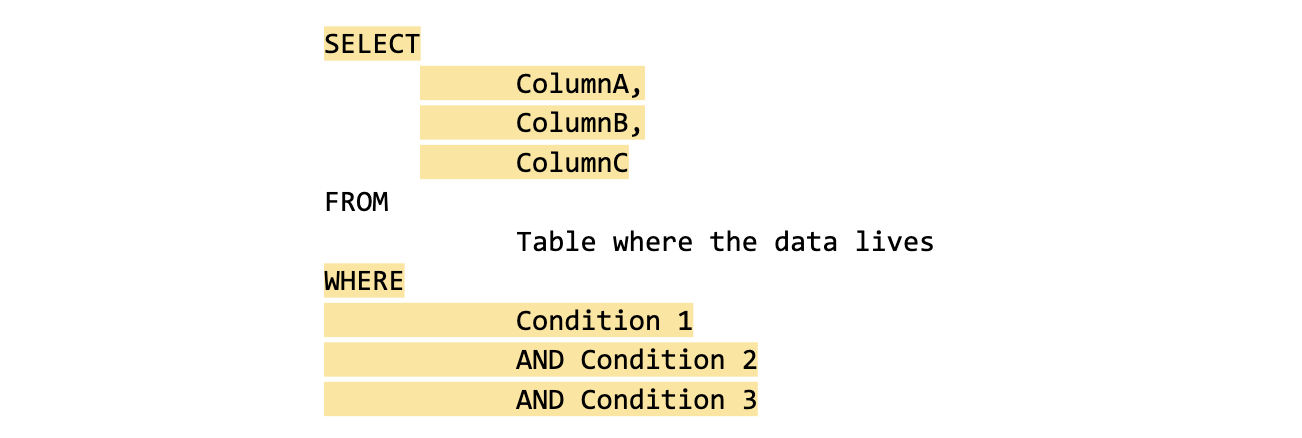
Here is an example of how it would appear in BigQuery:



The above query uses three commands to locate customers with the first name Tony.

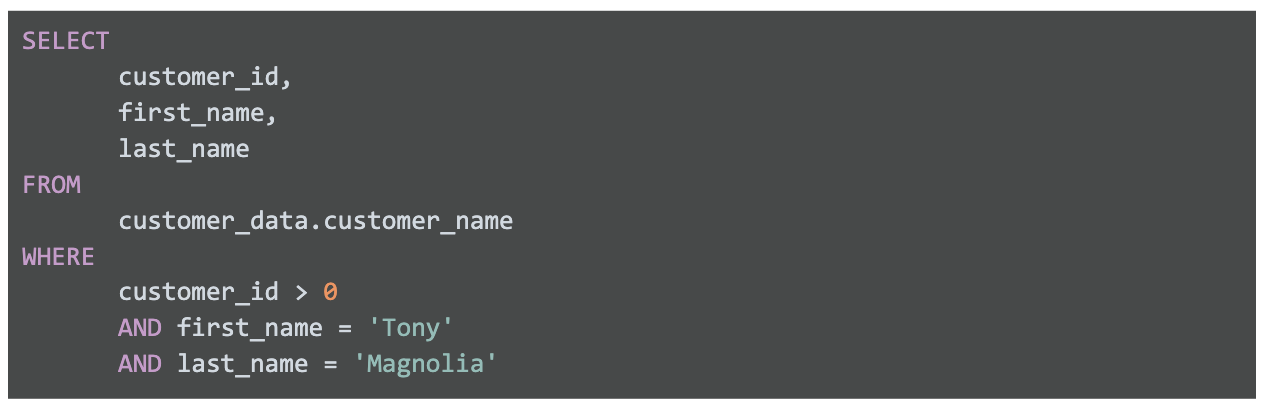
1. **SELECT** the columns named **customer\_id**, **first\_name**, and **last\_name**
2. **FROM** a table named **customer\_name** (in a dataset named **customer\_data**) (The dataset name is always followed by a dot, and then the table name.)
3. But only return the data **WHERE** the first\_name is **Tony**

The only difference between this query and the previous one is that more data columns are selected. The previous query selected first\_name only while this query selects customer\_id and last\_name in addition to first\_name. In general, it is a more efficient use of resources to select only the columns that you need. For example, it makes sense to select more columns if you will actually use the additional fields in your WHERE clause. If you have multiple conditions in your WHERE clause, they may be written like this:



Notice that unlike the SELECT command that uses a comma to separate fields/variables/parameters, the WHERE command uses the AND statement to connect conditions. As you become a more advanced writer of queries, you will make use of other connectors/operators such as OR and NOT.

Here is a BigQuery example with multiple fields used in a WHERE clause:



The above query uses three commands to locate customers with a valid (greater than 0) customer ID whose first name is Tony and last name is Magnolia.

1. **SELECT** the columns named **customer\_id**, **first\_name**, and **last\_name**
2. **FROM** a table named **customer\_name** (in a dataset named **customer\_data**) (The dataset name is always followed by a dot, and then the table name.)
3. But only return the data **WHERE** customer\_id is greater than **0**, first\_name is **Tony**, and last\_name is **Magnolia**.

Note that one of the conditions is a logical condition that checks to see if customer\_id is greater than zero.

If only one customer is named Tony Magnolia, the results from the query could be:

| **customer\_id** | **first\_name** | **last\_name** |
| --- | --- | --- |
| 1967 | Tony | Magnolia |

If more than one customer has the same name, the results from the query could be:

| **customer\_id** | **first\_name** | **last\_name** |
| --- | --- | --- |
| 1967 | Tony | Magnolia |
| 7689 | Tony | Magnolia |

### Key takeaway

The most important thing to remember is how to use SELECT, FROM, and WHERE in a query. Queries with multiple fields will become simpler after you practice writing your own SQL queries later in the program.

## Endless SQL possibilities

You have learned that a SQL query uses **SELECT**, **FROM**, and **WHERE** to specify the data to be returned from the query. This reading provides more detailed information about formatting queries, using WHERE conditions, selecting all columns in a table, adding comments, and using aliases. All of these make it easier for you to understand (and write) queries to put SQL in action. The last section of this reading provides an example of what a data analyst would do to pull employee data for a project.

### Capitalization, indentation, and semicolons

You can write your SQL queries in all lowercase and don’t have to worry about extra spaces between words. However, using capitalization and indentation can help you read the information more easily. Keep your queries neat, and they will be easier to review or troubleshoot if you need to check them later on.

*Bạn có thể viết các truy vấn SQL của mình bằng chữ thường và không cần phải lo lắng về khoảng trắng thừa giữa các từ. Tuy nhiên, sử dụng chữ in hoa và thụt đầu dòng có thể giúp bạn đọc thông tin dễ dàng hơn. Giữ cho các truy vấn của bạn gọn gàng và chúng sẽ dễ dàng hơn để xem xét hoặc sửa lỗi nếu bạn cần kiểm tra lại chúng sau này.*



Notice that the SQL statement shown above has a semicolon at the end. The semicolon is a statement terminator and is part of the American National Standards Institute (ANSI) SQL-92 standard, which is a recommended common syntax for adoption by all SQL databases. However, not all SQL databases have adopted or enforce the semicolon, so it’s possible you may come across some SQL statements that aren’t terminated with a semicolon. If a statement works without a semicolon, it’s fine.

*Chú ý rằng câu lệnh SQL được hiển thị ở trên có một dấu chấm phẩy ở cuối. Dấu chấm phẩy là một trình kết thúc câu lệnh và là một phần của tiêu chuẩn (ANSI) SQL-92 của American National Standards Institute Viện Tiêu chuẩn Quốc gia Hoa Kỳ (ANSI), đó là cú pháp thông thường được khuyến khích áp dụng bởi tất cả các cơ sở dữ liệu SQL. Tuy nhiên, không phải tất cả các cơ sở dữ liệu SQL đều áp dụng hoặc bắt buộc phải có dấu chấm phẩy, vì vậy có thể bạn sẽ gặp một số câu lệnh SQL không kết thúc bằng dấu chấm phẩy. Nếu một câu lệnh hoạt động mà không có dấu chấm phẩy, thì không vấn đề gì.*

### **WHERE** conditions



In the query shown above, the **SELECT** clause identifies the column you want to pull data from by name, **field1**, and the **FROM** clause identifies the table where the column is located by name, **table**. Finally, the **WHERE** clause narrows your query so that the database returns only the data with an exact value match or the data that matches a certain condition that you want to satisfy.

*Trong truy vấn được hiển thị ở trên, mệnh đề* ***SELECT*** *xác định cột mà bạn muốn trích xuất dữ liệu bằng tên của nó,* ***field1****, và mệnh đề* ***FROM*** *xác định bảng mà cột đó được đặt tên,* ***table****. Cuối cùng, mệnh đề* ***WHERE*** *giới hạn truy vấn của bạn sao cho cơ sở dữ liệu chỉ trả về dữ liệu với một giá trị chính xác hoặc dữ liệu phù hợp với một điều kiện nhất định mà bạn muốn thỏa mãn.*

For example, if you are looking for a specific customer with the last name Chavez, the WHERE clause would be:

*Ví dụ: nếu bạn đang tìm kiếm một khách hàng cụ thể có họ là Chavez, thì mệnh đề WHERE sẽ là:*

**WHERE field1 = 'Chavez'**

However, if you are looking for all customers with a last name that begins with the letters “Ch," the WHERE clause would be:

*Tuy nhiên, nếu bạn đang tìm kiếm tất cả khách hàng có họ bắt đầu bằng chữ cái “Ch", thì mệnh đề WHERE sẽ là:*

**WHERE field1 LIKE 'Ch%'**

You can conclude that the **LIKE** clause is very powerful because it allows you to tell the database to look for a certain pattern! The percent sign (%) is used as a wildcard to match one or more characters. In the example above, both **Chavez** and **Chen** would be returned. Note that in some databases an asterisk (\*) is used as the wildcard instead of a percent sign (%).

*Bạn có thể kết luận rằng mệnh đề* ***LIKE*** *rất mạnh vì nó cho phép bạn yêu cầu cơ sở dữ liệu tìm kiếm một mẫu nhất định! Dấu phần trăm (%) được sử dụng làm ký tự đại diện để khớp với một hoặc nhiều ký tự. Trong ví dụ trên, cả* ***Chavez*** *và* ***Chen*** *sẽ được trả lại. Lưu ý rằng trong một số cơ sở dữ liệu, dấu hoa thị (\*) được sử dụng làm ký tự đại diện thay vì ký hiệu phần trăm (%).*

### SELECT all columns

Can you use  **SELECT \*** ?



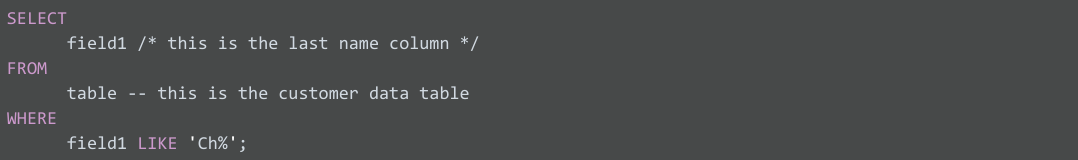
In the example, if you replace **SELECT field1** with **SELECT \*** , you would be selecting all of the columns in the table instead of the field1 column only. From a syntax point of view, it is a correct SQL statement, but you should use the asterisk (\*) sparingly and with caution. Depending on how many columns a table has, you could be selecting a tremendous amount of data. Selecting too much data can cause a query to run slowly.

*Trong ví dụ này, nếu bạn thay thế* ***SELECT field1*** *bằng* ***SELECT \**** *, bạn sẽ chọn tất cả các cột trong bảng thay vì chỉ chọn cột* ***field1****. Từ quan điểm cú pháp, đó là một câu lệnh SQL chính xác, nhưng bạn nên sử dụng dấu hoa thị* ***(\*)*** *một cách hạn chế và thận trọng. Tùy thuộc vào số lượng cột mà một bảng có, bạn có thể chọn một lượng dữ liệu khổng lồ. Chọn quá nhiều dữ liệu có thể khiến truy vấn chạy chậm.*

### Comments

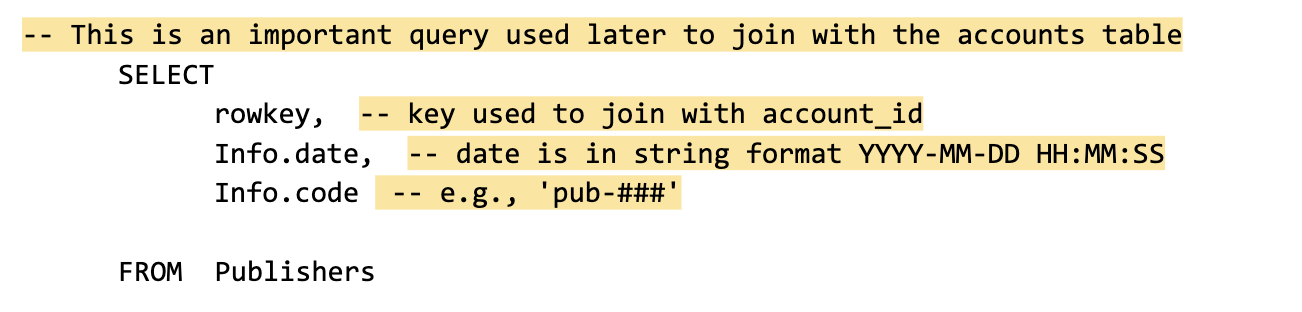
Some tables aren’t designed with descriptive enough naming conventions. In the example, **field1** was the column for a customer’s last name, but you wouldn’t know it by the name. A better name would have been something such as **last\_name**. In these cases, you can place comments alongside your SQL to help you remember what the name represents. Comments are text placed between certain characters, **/\*** and **\*/**, or after two dashes (**--**) as shown below.

*Một số bảng không được thiết kế với các quy ước đặt tên có đầy đủ mô tả. Trong ví dụ,* ***field1*** *là cột dành cho tên họ của khách hàng, nhưng bạn sẽ không biết tên đó. Một cái tên tốt hơn sẽ là một cái gì đó chẳng hạn như* ***last\_name****. Trong những trường hợp này, bạn có thể đặt* ***comment*** *bên cạnh SQL của mình để giúp bạn nhớ tên đại diện cho điều gì.* ***Comments*** *là văn bản được đặt giữa các ký tự nhất định,* ***/\**** *và* ***\*/****, hoặc sau hai dấu gạch ngang* ***(--)*** *như minh họa bên dưới.*



Comments can also be added outside of a statement as well as within a statement. You can use this flexibility to provide an overall description of what you are going to do, step-by-step notes about how you achieve it, and why you set different parameters/conditions.

***Comments*** *cũng có thể được thêm vào bên ngoài câu lệnh cũng như bên trong câu lệnh. Bạn có thể sử dụng tính linh hoạt này để cung cấp mô tả hoặc giải thích tổng thể về những gì bạn sẽ làm, comment từng bước về cách bạn đạt được mục tiêu đó và lý do bạn đặt các tham số/điều kiện khác nhau.*

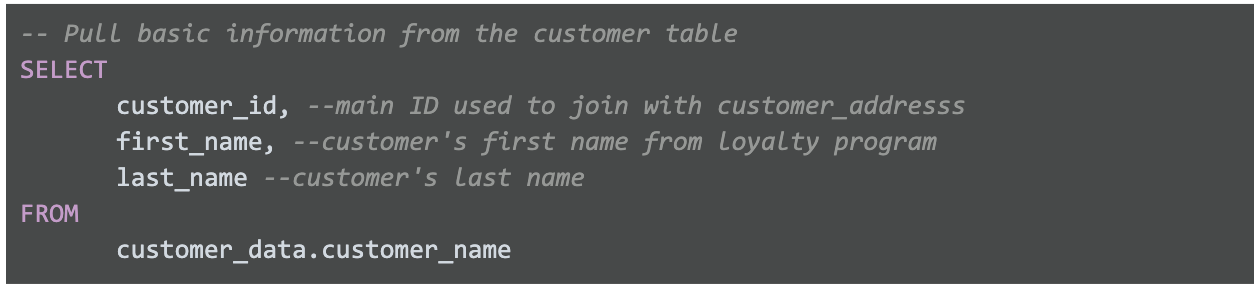


The more comfortable you get with SQL, the easier it will be to read and understand queries at a glance. Still, it never hurts to have comments in a query to remind yourself of what you’re trying to do. This also makes it easier for others to understand your query if your query is shared. As your queries become more and more complex, this practice will save you a lot of time and energy to understand complex queries you wrote months or years ago.

*Bạn càng cảm thấy thoải mái với SQL, thì càng dễ dàng đọc và hiểu các truy vấn trong nháy mắt. Tuy nhiên, sẽ không hại gì khi có nhận xét trong truy vấn để nhắc nhở bản thân về những gì bạn đang cố gắng thực hiện. Điều này cũng giúp những người khác hiểu truy vấn của bạn dễ dàng hơn nếu truy vấn của bạn được chia sẻ. Khi các truy vấn của bạn ngày càng trở nên phức tạp hơn, phương pháp này sẽ giúp bạn tiết kiệm rất nhiều thời gian và năng lượng để hiểu các truy vấn phức tạp mà bạn đã viết từ nhiều tháng hoặc nhiều năm trước.*

Here is an example of how comments could be written in BigQuery:

*Dưới đây là ví dụ về cách viết comments trong BigQuery:*



In the above example, a comment has been added before the SQL statement to explain what the query does. Additionally, a comment has been added next to each of the column names to describe the column and its use. Two dashes (--) are generally supported. So it is best to use -- and be consistent with it. You can use # in place of -- in the above query, but # is not recognized in all SQL versions; for example, MySQL doesn’t recognize #.  You can also place comments between /\* and \*/ if the database you are using supports it.

*Trong ví dụ trên, một comment đã được thêm vào trước câu lệnh SQL để giải thích truy vấn làm gì. Ngoài ra, một comment đã được thêm vào bên cạnh mỗi tên cột để mô tả cột và cách sử dụng cột đó. Hai dấu gạch ngang (--) thường được hỗ trợ. Vì vậy, tốt nhất là sử dụng -- và nhất quán với nó. Bạn có thể sử dụng # thay cho -- trong truy vấn trên, nhưng # không được nhận dạng trong tất cả các phiên bản SQL; ví dụ: MySQL không nhận ra #. Bạn cũng có thể đặt comments giữa /\* và \*/ nếu cơ sở dữ liệu bạn đang sử dụng hỗ trợ nó.*

As you develop your skills professionally, depending on the SQL database you use, you can pick the appropriate comment delimiting symbols you prefer and stick with those as a consistent style. As your queries become more and more complex, the practice of adding helpful comments will save you a lot of time and energy to understand queries that you may have written months or years prior.

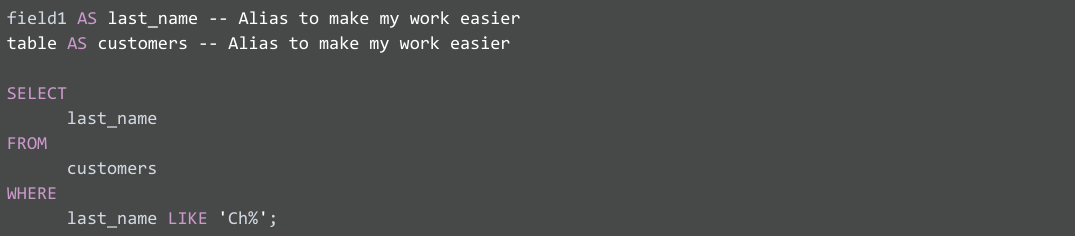
*Khi bạn phát triển các kỹ năng của mình một cách chuyên nghiệp, tùy thuộc vào cơ sở dữ liệu SQL mà bạn sử dụng, bạn có thể chọn các ký hiệu phân định comment thích hợp mà bạn thích và gắn bó với các ký hiệu đó như một phong cách nhất quán. Khi các truy vấn của bạn ngày càng trở nên phức tạp hơn, việc thêm các comments hữu ích sẽ giúp bạn tiết kiệm rất nhiều thời gian và năng lượng để hiểu các truy vấn mà bạn có thể đã viết hàng tháng hoặc hàng năm trước.*

### Aliases

You can also make it easier on yourself by assigning a new name or **alias** to the column or table names to make them easier to work with (and avoid the need for comments). This is done with a SQL AS clause. In the example below, the alias **last\_name** has been assigned to **field1** and the alias **customers** assigned to **table.** These aliases are good for the duration of the query only. An alias doesn’t change the actual name of a column or table in the database.

*Thay vì sử dụng comment để chú thích cho 1 một thành phần bạn cũng có thể tự làm việc này dễ dàng hơn bằng cách gán tên hoặc bí danh mới hoặc* ***alias*** *cho tên cột hoặc tên bảng để làm việc với chúng dễ dàng hơn (và tránh phải nhận xét). Điều này được thực hiện với mệnh đề SQL AS. Trong ví dụ bên dưới, bí danh* ***last\_name*** *đã được gán cho* ***field1*** *và bí danh* ***customers*** *được gán cho* ***table****. Những bí danh này chỉ tốt cho thời lượng của truy vấn. Bí danh không thay đổi tên thực của cột hoặc bảng trong cơ sở dữ liệu.*

**Example of a query with aliases:**



### Putting SQL to work as a data analyst

Imagine you are a data analyst for a small business and your manager asks you for some employee data. You decide to write a query with SQL to get what you need from the database.

*Hãy tưởng tượng bạn là nhà phân tích dữ liệu cho một doanh nghiệp nhỏ và người quản lý của bạn yêu cầu bạn cung cấp một số dữ liệu nhân viên. Bạn quyết định viết một truy vấn bằng SQL để lấy những gì bạn cần từ cơ sở dữ liệu.*

You want to pull all the columns: **empID**, **firstName**, **lastName**, **jobCode**, and **salary**. Because you know the database isn’t that big, instead of entering each column name in the **SELECT** clause, you use **SELECT \***.  This will select all the columns from the Employee table in the **FROM** clause.

*Bạn muốn kéo tất cả các cột:* ***empID****,* ***FirstName****,* ***lastName****,* ***jobCode*** *và salary. Vì bạn biết cơ sở dữ liệu không quá lớn nên thay vì nhập từng tên cột trong mệnh đề* ***SELECT****, bạn sử dụng* ***SELECT \*.*** *Thao tác này sẽ chọn tất cả các cột từ bảng* ***Employee*** *trong mệnh đề* ***FROM****.*



Now, you can get more specific about the data you want from the Employee table. If you want all the data about employees working in the **SFI** job code, you can use a **WHERE** clause to filter out the data based on this additional requirement. Here, you use:

*Bây giờ, bạn có thể nhận được cụ thể hơn về dữ liệu bạn muốn từ bảng* ***Employee****. Nếu bạn muốn tất cả dữ liệu về nhân viên làm việc trong mã công việc* ***SFI****, bạn có thể sử dụng mệnh đề* ***WHERE*** *để lọc ra dữ liệu dựa trên yêu cầu bổ sung này. Ở đây, bạn sử dụng:*



A portion of the resulting data returned from the SQL query might look like this:

*Một phần dữ liệu kết quả được trả về từ truy vấn SQL có thể giống như sau:*

| **empID** | **firstName** | **lastName** | **jobCode** | **salary** |
| --- | --- | --- | --- | --- |
| 0002 | Homer | Simpson | SFI | 15000 |
| 0003 | Marge | Simpson | SFI | 30000 |
| 0034 | Bart | Simpson | SFI | 25000 |
| 0067 | Lisa | Simpson | SFI | 38000 |
| 0088 | Ned | Flanders | SFI | 42000 |
| 0076 | Barney | Gumble | SFI | 32000 |

Suppose you notice a large salary range for the **SFI** job code. You might like to flag all employees in all departments with lower salaries for your manager. Because interns are also included in the table and they have salaries less than $30,000, you want to make sure your results give you only the full time employees with salaries that are $30,000 or less. In other words, you want to exclude interns with the **INT** job code who also earn less than $30,000. The **AND** clause enables you to test for both conditions. You create a SQL query similar to below, where <> means "does not equal":

*Giả sử bạn nhận thấy một phạm vi lương lớn cho mã công việc* ***SFI****. Bạn có thể muốn gắn cờ tất cả nhân viên trong tất cả các bộ phận có mức lương thấp hơn cho người quản lý của mình. Bởi vì thực tập sinh cũng được bao gồm trong bảng và họ có mức lương dưới 30.000 đô la, nên bạn muốn đảm bảo rằng kết quả của mình chỉ cung cấp cho bạn những nhân viên toàn thời gian có mức lương từ 30.000 đô la trở xuống. Nói cách khác, bạn muốn loại trừ những thực tập sinh có mã công việc* ***INT****, những người cũng kiếm được ít hơn 30.000 đô la. Mệnh đề* ***AND*** *cho phép bạn kiểm tra cả hai điều kiện. Bạn tạo một truy vấn SQL tương tự như bên dưới, trong đó <> có nghĩa là "không bằng":*

A picture containing text, screenshot, font

Description automatically generated

The resulting data from the SQL query might look like the following (interns with the job code **INT** aren't returned):

*Dữ liệu kết quả từ truy vấn SQL có thể giống như sau (thực tập sinh với mã công việc* ***INT*** *không được trả lại):*

| **empID** | **firstName** | **lastName** | **jobCode** | **salary** |
| --- | --- | --- | --- | --- |
| 0002 | Homer | Simpson | SFI | 15000 |
| 0003 | Marge | Simpson | SFI | 30000 |
| 0034 | Bart | Simpson | SFI | 25000 |
| 0108 | Edna | Krabappel | TUL | 18000 |
| 0099 | Moe | Szyslak | ANA | 28000 |

With quick access to this kind of data using SQL, you can provide your manager with tons of different insights about employee data, including whether employee salaries across the business are equitable. Fortunately, the query shows only an additional two employees might need a salary adjustment and you share the results with your manager.

*Với khả năng truy cập nhanh vào loại dữ liệu này bằng SQL, bạn có thể cung cấp cho người quản lý của mình rất nhiều thông tin chi tiết khác nhau về dữ liệu nhân viên, bao gồm cả việc lương của nhân viên trong toàn doanh nghiệp có công bằng hay không. May mắn thay, truy vấn chỉ hiển thị thêm hai nhân viên có thể cần điều chỉnh lương và bạn chia sẻ kết quả với người quản lý của mình.*

Pulling the data, analyzing it, and implementing a solution might ultimately help improve employee satisfaction and loyalty. That makes SQL a pretty powerful tool.

*Kéo dữ liệu, phân tích nó và thực hiện một giải pháp cuối cùng có thể giúp cải thiện sự hài lòng và lòng trung thành của nhân viên. Điều đó làm cho SQL trở thành một công cụ khá mạnh mẽ.*

### Resources to learn more

Nonsubscribers may access these resources for free, but if a site limits the number of free articles per month and you already reached your limit, bookmark the resource and come back to it later.

* [W3Schools SQL Tutorial](https://www.w3schools.com/sql/default.asp): If you would like to explore a detailed tutorial of SQL, this is the perfect place to start. This tutorial includes interactive examples you can edit, test, and recreate. Use it as a reference or complete the whole tutorial to practice using SQL. Click the green **Start learning SQL now** button or the **Next** buttonto begin the tutorial.
* [SQL Cheat Sheet](https://towardsdatascience.com/sql-cheat-sheet-776f8e3189fa): For more advanced learners, go through this article for standard SQL syntax used in PostgreSQL. By the time you are finished, you will know a lot more about SQL and will be prepared to use it for business analysis and other tasks.

## Planning a data visualization

Earlier, you learned that **data visualization** is the graphical representation of information. As a data analyst, you will want to create visualizations that make your data easy to understand and interesting to look at. Because of the importance of data visualization, most data analytics tools (such as spreadsheets and databases) have a built-in visualization component while others (such as Tableau) specialize in visualization as their primary value-add. In this reading, you will explore the steps involved in the data visualization process and a few of the most common data visualization tools available.

*Trước đó, bạn đã biết rằng trực quan hóa dữ liệu là biểu diễn đồ họa của thông tin. Là một nhà phân tích dữ liệu, bạn sẽ muốn tạo các hình ảnh trực quan giúp dữ liệu của bạn dễ hiểu và thú vị khi xem xét. Do tầm quan trọng của trực quan hóa dữ liệu, hầu hết các công cụ phân tích dữ liệu (chẳng hạn như spreadsheets và databases) đều có thành phần trực quan hóa tích hợp trong khi các công cụ khác (chẳng hạn như Tableau) chuyên về trực quan hóa làm giá trị gia tăng chính của chúng. Trong bài đọc này, bạn sẽ khám phá các bước liên quan đến quy trình trực quan hóa dữ liệu và một số công cụ trực quan hóa dữ liệu phổ biến nhất hiện có.*



### Steps to plan a data visualization

Let’s go through an example of a real-life situation where a data analyst might need to create a data visualization to share with stakeholders. Imagine you’re a data analyst for a clothing distributor. The company helps small clothing stores manage their inventory, and sales are booming. One day, you learn that your company is getting ready to make a major update to its website. To guide decisions for the website update, you’re asked to analyze data from the existing website and sales records. Let’s go through the steps you might follow.

*Hãy xem qua một ví dụ về tình huống thực tế mà nhà phân tích dữ liệu có thể cần tạo trực quan hóa dữ liệu để chia sẻ với các bên liên quan. Hãy tưởng tượng bạn là nhà phân tích dữ liệu cho một nhà phân phối quần áo. Công ty giúp các cửa hàng quần áo nhỏ quản lý hàng tồn kho của họ và doanh số bán hàng đang bùng nổ. Một ngày nọ, bạn biết rằng công ty của bạn đang chuẩn bị thực hiện một bản cập nhật lớn cho trang web của mình. Để hướng dẫn các quyết định cho việc cập nhật trang web, bạn được yêu cầu phân tích dữ liệu từ trang web hiện có và hồ sơ bán hàng. Hãy đi qua các bước bạn có thể làm theo.*

**Step 1: Explore the data for patterns**

First, you ask your manager or the data owner for access to the current sales records and website analytics reports. This includes information about how customers behave on the company’s existing website, basic information about who visited, who bought from the company, and how much they bought.

*Trước tiên, bạn yêu cầu người quản lý của mình hoặc chủ sở hữu dữ liệu cấp quyền truy cập vào hồ sơ bán hàng hiện tại và báo cáo phân tích trang web. Điều này bao gồm thông tin về hành vi của khách hàng trên trang web hiện có của công ty, thông tin cơ bản về những người đã truy cập, những người đã mua hàng từ công ty và họ đã mua bao nhiêu.*

While reviewing the data you notice a pattern among those who visit the company’s website most frequently: geography and larger amounts spent on purchases. With further analysis, this information might explain why sales are so strong right now in the northeast—and help your company find ways to make them even stronger through the new website.

*Trong khi xem xét dữ liệu, bạn nhận thấy một khuôn mẫu trong số những người truy cập trang web của công ty thường xuyên nhất: địa lý và số tiền lớn hơn được chi cho việc mua hàng. Với sự phân tích sâu hơn, thông tin này có thể giải thích lý do tại sao doanh số bán hàng ở phía đông bắc hiện tại rất mạnh—và giúp công ty của bạn tìm cách làm cho họ thậm chí còn mạnh mẽ hơn thông qua trang web mới.*

**Step 2: Plan your visuals**

Next it is time to refine the data and present the results of your analysis. Right now, you have a lot of data spread across several different tables, which isn’t an ideal way to share your results with management and the marketing team. You will want to create a data visualization that explains your findings quickly and effectively to your target audience. Since you know your audience is sales oriented, you already know that the data visualization you use should:

*Tiếp theo, đã đến lúc tinh chỉnh dữ liệu và trình bày kết quả phân tích của bạn. Ngay bây giờ, bạn có rất nhiều dữ liệu trải rộng trên một số bảng khác nhau, đây không phải là cách lý tưởng để chia sẻ kết quả của bạn với ban quản lý và nhóm tiếp thị. Bạn sẽ muốn tạo một hình ảnh trực quan hóa dữ liệu để giải thích những phát hiện của mình một cách nhanh chóng và hiệu quả cho đối tượng mục tiêu. Vì bạn biết đối tượng của mình hướng đến doanh số bán hàng, nên bạn đã biết rằng hình ảnh hóa dữ liệu bạn sử dụng nên:*

* Show sales numbers over time

*Hiển thị số lượng bán hàng theo thời gian*

* Connect sales to location

*Kết nối bán hàng với địa điểm*

* Show the relationship between sales and website use

*Thể hiện mối quan hệ giữa bán hàng và sử dụng website*

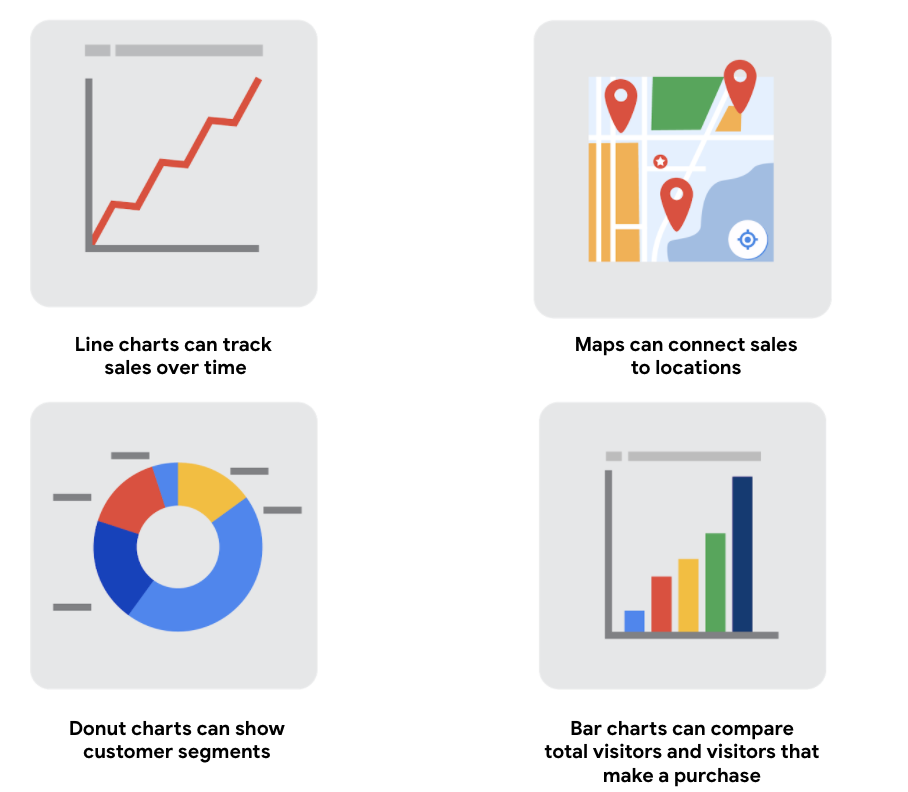
* Show which customers fuel growth

*Cho biết khách hàng nào thúc đẩy tăng trưởng*

**Step 3: Create your visuals**

Now that you have decided what kind of information and insights you want to display, it is time to start creating the actual visualizations. Keep in mind that creating the right visualization for a presentation or to share with stakeholders is a process. It involves trying different visualization formats and making adjustments until you get what you are looking for. In this case, a mix of different visuals will best communicate your findings and turn your analysis into the most compelling story for stakeholders. So, you can use the built-in chart capabilities in your spreadsheets to organize the data and create your visuals.

*Bây giờ bạn đã quyết định loại thông tin và thông tin chi tiết nào bạn muốn hiển thị, đã đến lúc bắt đầu tạo trực quan hóa thực tế. Hãy nhớ rằng việc tạo hình ảnh trực quan phù hợp cho bản trình bày hoặc để chia sẻ với các bên liên quan là cả một quá trình. Nó liên quan đến việc thử các định dạng trực quan hóa khác nhau và thực hiện các điều chỉnh cho đến khi bạn nhận được thứ mình đang tìm kiếm. Trong trường hợp này, sự kết hợp của các hình ảnh khác nhau sẽ truyền đạt tốt nhất những phát hiện của bạn và biến phân tích của bạn thành câu chuyện hấp dẫn nhất cho các bên liên quan. Vì vậy, bạn có thể sử dụng các chức năng biểu đồ tích hợp sẵn trong bảng tính của mình để sắp xếp dữ liệu và tạo hình ảnh trực quan.*



### Build your data visualization toolkit

There are many different tools you can use for data visualization.

*Có nhiều công cụ khác nhau mà bạn có thể sử dụng để trực quan hóa dữ liệu.*

* You can use the visualizations tools in your spreadsheet to create simple visualizations such as line and bar charts.

*Bạn có thể sử dụng các công cụ trực quan hóa trong bảng tính của mình để tạo trực quan hóa đơn giản, chẳng hạn như biểu đồ đường và thanh*

* You can use more advanced tools such as Tableau that allow you to integrate data into dashboard-style visualizations.

*Bạn có thể sử dụng các công cụ nâng cao hơn như* ***Tableau*** *cho phép bạn tích hợp dữ liệu vào trực quan hóa kiểu bảng điều khiển.*

* If you’re working with the programming language R you can use the visualization tools in RStudio.

*Nếu bạn đang làm việc với ngôn ngữ lập trình R, bạn có thể sử dụng các công cụ trực quan hóa trong RStudio.*

Your choice of visualization will be driven by a variety of drivers including the size of your data, the process you used for analyzing your data (spreadsheet, or databases/queries, or programming languages). For now, just consider the basics.

*Lựa chọn hình ảnh hóa của bạn sẽ được thúc đẩy bởi nhiều trình điều khiển bao gồm kích thước dữ liệu của bạn, quy trình bạn đã sử dụng để phân tích dữ liệu của mình (bảng tính hoặc cơ sở dữ liệu/truy vấn hoặc ngôn ngữ lập trình). Hiện tại, chỉ cần xem xét những điều cơ bản.*

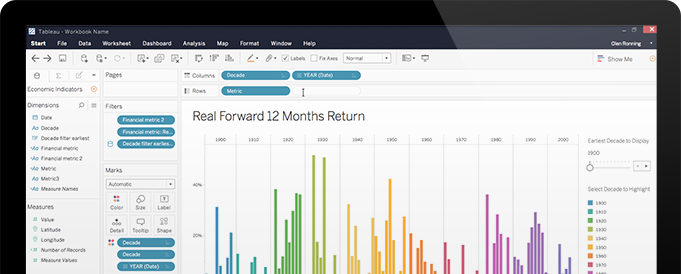
### Spreadsheets (Microsoft Excel or Google Sheets)

In our example, the built-in charts and graphs in spreadsheets made the process of creating visuals quick and easy. Spreadsheets are great for creating simple visualizations like bar graphs and pie charts, and even provide some advanced visualizations like maps, and waterfall and funnel diagrams (shown in the following figures).

But sometimes you need a more powerful tool to truly bring your data to life. Tableau and RStudio are two examples of widely used platforms that can help you plan, create, and present effective and compelling data visualizations.

### Visualization software (Tableau)

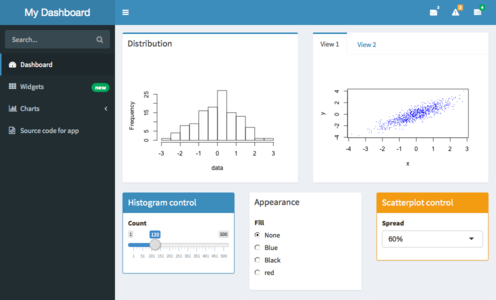
Tableau is a popular data visualization tool that lets you pull data from nearly any system and turn it into compelling visuals or actionable insights. The platform offers built-in visual best practices, which makes analyzing and sharing data fast, easy, and (most importantly) useful. Tableau works well with a wide variety of data and includes an interactive dashboard that lets you and your stakeholders click to explore the data interactively.



You can start exploring Tableau from the [How-to Video](https://public.tableau.com/en-us/s/resources) resources. Tableau Public is free, easy to use, and full of helpful information. The Resources page is a one-stop-shop for how-to videos, examples, and datasets for you to practice with. To explore what other data analysts are sharing on Tableau, visit the [Viz of the Day](https://public.tableau.com/en-us/gallery/?tab=viz-of-the-day&type=viz-of-the-day) page where you will find beautiful visuals ranging from the [Hunt for (Habitable) Planets](https://public.tableau.com/en-us/gallery/hunt-habitable-planets?tab=viz-of-the-day&type=viz-of-the-day) to [Who’s Talking in Popular Films](https://public.tableau.com/en-us/gallery/whos-talking-popular-films?tab=viz-of-the-day&type=viz-of-the-day).

### Programming language (R with RStudio)

A lot of data analysts work with a programming language called R. Most people who work with R end up also using RStudio, an integrated developer environment (IDE), for their data visualization needs. As with Tableau, you can create dashboard-style data visualizations using RStudio.



Check out their website to learn more about [RStudio](https://rstudio.com/).

You could easily spend days exploring all the resources provided at RStudio.com, but the [RStudio Cheatsheets](https://rstudio.com/resources/cheatsheets/) and the [RStudio Visualize Data Primer](https://rstudio.cloud/learn/primers/3) are great places to start. When you have more time, check out the webinars and videos which offer advice and helpful perspectives for both beginners and advanced users.

### Key takeaway

The best data analysts use lots of different tools and methods to visualize and share their data. As you continue learning more about data visualization throughout this course, be sure to stay curious, research different options, and continuously test new programs and platforms to help you make the most of your data.

# **Ask Questions to Make Data-Driven Decisions**

This is the second course in the Google Data Analytics Certificate. These courses will equip you with the skills needed to apply to introductory-level data analyst jobs. You’ll build on your understanding of the topics that were introduced in the first Google Data Analytics Certificate course. The material will help you learn how to ask effective questions to make data-driven decisions, while connecting with stakeholders’ needs. Current Google data analysts will continue to instruct and provide you with hands-on ways to accomplish common data analyst tasks with the best tools and resources.

Learners who complete this certificate program will be equipped to apply for introductory-level jobs as data analysts. No previous experience is necessary.

By the end of this course, you will:

* Learn about effective questioning techniques that can help guide analysis.
* Gain an understanding of data-driven decision-making and how data analysts present findings.
* Explore a variety of real-world business scenarios to support an understanding of questioning and decision-making.
* Discover how and why spreadsheets are an important tool for data analysts.
* Examine the key ideas associated with structured thinking and how they can help analysts better understand problems and develop solutions.
* Learn strategies for managing the expectations of stakeholders while establishing clear communication with a data analytics team to achieve business objectives.

## Data in action

In this video, I'm going to share an interesting data analytics case study, it will illustrate how problem solving relates to each phase of the data analysis process and shed some light on how these phases work in the real world.

It's about a small business that used data to solve a unique problem it was facing. The business is called Anywhere Gaming Repair. It's a service provider that comes to you to fix your broken video game systems or accessories. The owner wanted to expand his business. He knew advertising is a proven way to get more customers, but he wasn't sure where to start. There are all kinds of different advertising strategies, including print, billboards, TV commercials, public transportation, podcasts, and radio. One of the key things to think about when choosing an advertising method is your target audience, in other words, the specific people you're trying to reach. For example, if a medical equipment manufacturer wanted to reach doctors, placing an ad in a health magazine would be a smart choice. Or if a catering company wanted to find new cooks, it might advertise using a poster at a bus stop near a cooking school. Both of these are great ways to get your ad seen by your target audience. The second thing to think about is your budget and how much the different advertising methods will cost. For instance, a TV ad is likely to be more expensive than a radio ad. A large billboard will probably cost more than a small poster on the back of a city bus. The business owner asked a data analyst, Maria, to make a recommendation. She started with the first step in the data analysis process, Ask. Maria began by defining the problem that needed to be solved. To do this, she first had to zoom out and look at the whole situation in context. That way she could be sure that she was focusing on the real problem and not just its symptoms. This leads us to another important part of the problem solving process, collaborating with stakeholders and understanding their needs. For Anywhere Gaming Repair, stakeholders included the owner, the vice president of communications, and the director of marketing and finance. Working together, Maria and the stakeholders agreed on the problem, not knowing their target audience's preferred type of advertising. Next step was the prepare phase, where Maria collected data for the upcoming analysis process. But first, she needed to better understand the company's target audience, people with video game systems. After that, Maria collected data on the different advertising methods. This way, she would be able to determine which was the most popular one with the company's target audience. Then she moved on to the process step. Here Maria cleaned the data to eliminate any errors or inaccuracies that could get in the way of the result. As we've learned, when you clean data, you transform it into a more useful format, create more complete information and remove outliers. Then it was time to analyze. In this step, Maria wanted to find out two things. First, who's most likely to own a video gaming system? Second, where are these people most likely to see an advertisement? Maria, first discovered that people between the ages of 18 and 34 are most likely to make video game related purchases. She could confirm that Anywhere Gaming Repair's target audience was people 18-34 years old. This was who they should be trying to reach. With this in mind, Maria then learned that both TV commercials and podcasts are very popular with people in the target audience. Because Maria knew Anywhere Gaming Repair had a limited budget and understanding the high cost of TV commercials, her recommendation was to advertise in podcasts because they are more cost-effective. Now that she had her analysis, it was time for Maria to share her recommendation so the company could make a data driven decision. She summarized her results using clear and compelling visuals of the analysis. This helped her stakeholders understand the solution to the original problem. Finally, Anywhere Gaming Repair took action, they worked with a local podcast production agency to create a 30 second ad about their services. The ad ran on podcast for a month, and it worked. They saw an increase in customers after just the first week. By the end of week 4, they had 85 new customers. There you go. Effective problem solving using data analysis phases in action. Now, you've seen how the six phases of data analysis can be applied to problem solving and how you can use that to solve real world problems.

## From issue to action: The six data analysis phases

There are six data analysis phases that will help you make seamless decisions: ask, prepare, process, analyze, share, and act. Keep in mind, these are different from the data life cycle, which describes the changes data goes through over its lifetime. Let’s walk through the steps to see how they can help you solve problems you might face on the job.



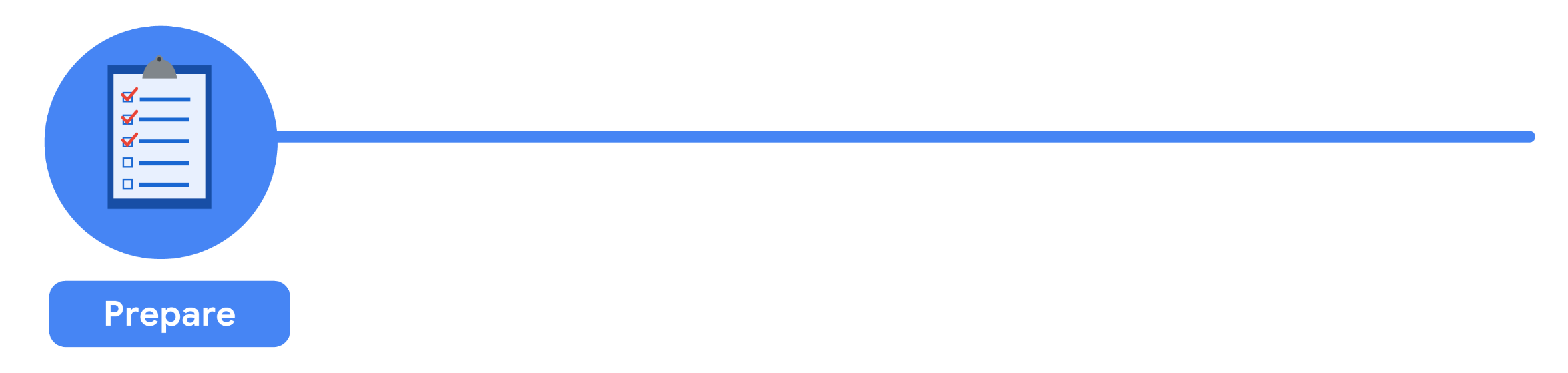
### Step 1: Ask

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:

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?



### 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:

1. What do I need to figure out how to solve this problem?
2. 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:

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?



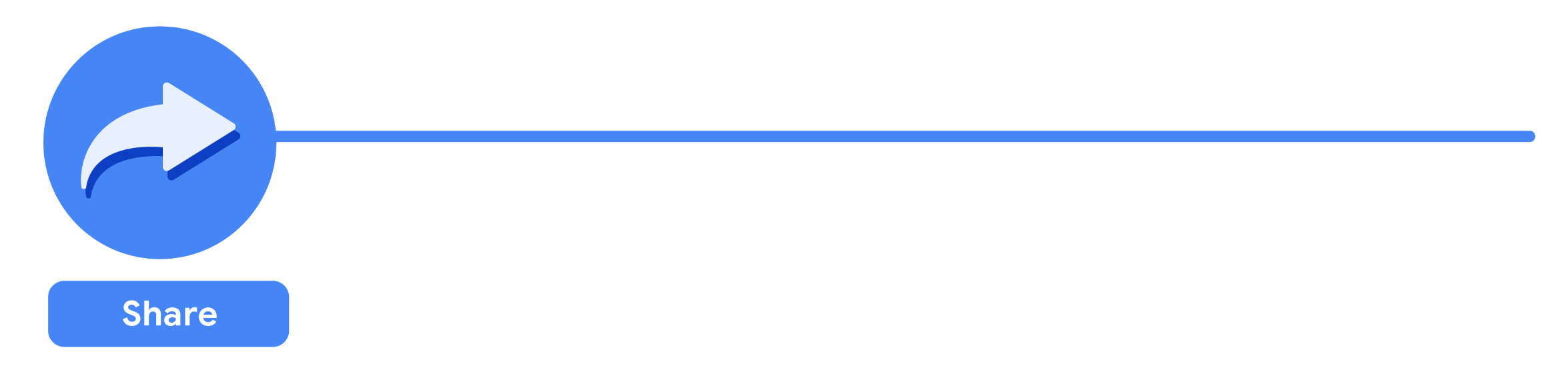
### Step 4: Analyze

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:

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?



### 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 via 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:

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?



### 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:

1. 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

When you are starting out in your career as a data analyst, it is normal to feel pulled in a few different directions with your role and expectations. Following processes like the ones outlined here and using structured thinking skills can help get you back on track, fill in any gaps and let you know exactly what you need.

## Common problem types

In the previous part, We were shared how data analysis helped a company figure out where to advertise its services. An important part of this process was strong problem-solving skills. As a data analyst, you'll find that problems are at the center of what you do every single day, but that's a good thing. Think of problems as opportunities to put your skills to work and find creative and insightful solutions. Problems can be small or large, simple or complex, no problem is like another and they all require a slightly different approach but the first step is always the same: Understanding what kind of problem you're trying to solve and that's what we're going to talk about now.

*Ở phần trước, trước ta đã được chia sẻ về một trường hợp mà công ty tìm ra được địa điểm quảng cáo dịch vụ của mình thông qua việc phân tích dữ liệu. Một phần quan trọng trong quy trình này là kỹ năng giải quyết vấn đề tốt. Là một nhà phân tích dữ liệu, bạn sẽ thấy các vấn đề là trọng tâm của mọi công việc bạn làm hàng ngày, nhưng đó là điều tốt. Hãy coi các vấn đề là cơ hội để áp dụng các kỹ năng của mình và tìm ra các giải pháp sáng tạo và sâu sắc. Vấn đề có thể lớn hoặc nhỏ, đơn giản hoặc phức tạp, không có vấn đề nào giống nhau, chúng đòi hỏi các cách tiếp cận khác nhau, nhưng bước đầu tiên thì luôn giống nhau: hiểu rõ vấn đề mà bạn đang muốn giải quyết và bây giờ chúng ta sẽ nói về việc này.*

Data analysts work with a variety of problems. In this section, we're going to focus on six common types. These include:

*Việc phân tích dữ liệu có thể áp dụng với nhiều loại vấn đề. Trong phần này, ta sẽ tập trung vào sáu loại vấn đề thường gặp. Các loại này bao gồm:*

* Making predictions

*Dự đoán*

* Categorizing things

*Phân loại*

* Spotting something unusual

*Phát hiện điểm khác biệt*

* Identifying themes

*Xác định chủ đề*

* Discovering connections

*Phát hiện các liên kết*

* Finding patterns

*Tìm ra mẫu*

As a future data analyst, this is going to be valuable knowledge for your career.

*Là một nhà phân tích dữ liệu trong tương lai, đây sẽ là những kiến thức quý giá cho sự nghiệp của bạn.*

### Making predictions

First, making predictions. This problem type involves using data to make an informed decision about how things may be in the future.

*Đầu tiên là dự đoán. Loại vấn đề này liên quan đến việc sử dụng dữ liệu để đưa ra quyết định sáng suốt về diễn biến của mọi chuyện trong tương lai.*

For example, a hospital system might use a remote patient monitoring to predict health events for chronically ill patients. The patients would take their health vitals at home every day, and that information combined with data about their age, risk factors, and other important details could enable the hospital's algorithm to predict future health problems and even reduce future hospitalizations.

*Ví dụ, hệ thống bệnh viện có thể áp dụng việc giám sát bệnh nhân từ xa để dự đoán các biến cố sức khỏe của các bệnh nhân mãn tính. Bệnh nhân sẽ được đo sinh hiệu tại nhà mỗi người, thông tin đó kết hợp với dữ liệu về tuổi tác, yếu tố rủi ro, và các thông tin quan trọng khác có thể giúp thuật toán của bệnh viện dự đoán được các vấn đề về sức khỏe trong tương lai và giảm thiểu số lần phải nhập viện.*

### Categorizing things

The next problem type is categorizing things. This means assigning information to different groups or clusters based on common features.

*Loại vấn đề tiếp theo là phân loại. Tức là chỉ định thông tin vào các nhóm hoặc các cụm dựa trên các thuộc tính giống nhau.*

An example of this problem type is a manufacturer that reviews data on shop floor employee performance. An analyst may create a group for employees who are most and least effective at engineering. A group for employees who are most and least effective at repair and maintenance, most and least effective at assembly, and many more groups or clusters.

*Một ví dụ về loại vấn đề này là khi một nhà sản xuất xem xét dữ liệu về hiệu suất của nhân viên tại khu vực cửa hàng. Nhà phân tích có thể tạo một nhóm cho các nhân viên làm việc hiệu quả nhất và kém hiệu quả nhất về kỹ thuật. Nhóm cho các nhân viên làm việc hiệu quả nhất và kém hiệu quả nhất về sửa chữa và bảo dưỡng, làm việc hiệu quả nhất và kém hiệu quả nhất về lắp ráp, cũng như là nhiều nhóm và cụm khác.*

### Spotting something unusual

Next, we have spotting something unusual. In this problem type, data analysts identify data that is different from the norm.

*Tiếp theo là phát hiện điểm khác biệt. Với loại vấn đề này, các nhà phân tích dữ liệu xác định các dữ liệu khác với tiêu chuẩn.*

An instance of spotting something unusual in the real world is a school system that has a sudden increase in the number of students registered, maybe as big as a 30 percent jump in the number of students. A data analyst might look into this upswing and discover that several new apartment complexes had been built in the school district earlier that year. They could use this analysis to make sure the school has enough resources to handle the additional students.

*Ví dụ về việc phát hiện điểm khác biệt trong thực tế là ở một hệ thống trường học, số học sinh đăng ký đột ngột tăng lên, con số cao đến mức tăng 30% số học sinh. Nhà phân tích dữ liệu có thể nhìn vào mức tăng này và phát hiện nhiều khu tổ hợp căn hộ mới đã được xây dựng trong khu vực vào đầu năm. Họ có thể áp dụng phân tích này để đảm bảo trường học có đủ nguồn lực để phục vụ các học sinh mới.*

### Identifying themes

Identifying themes is the next problem type. Identifying themes takes categorization as a step further by grouping information into broader concepts.

*Loại vấn đề tiếp theo là xác định chủ đề. Việc xác định chủ đề chính là phân loại ở mức độ cao hơn bằng các nhóm thông tin thành các khái niệm rộng hơn.*

Going back to our manufacturer that has just reviewed data on the shop floor employees. First, these people are grouped by types and tasks. But now a data analyst could take those categories and group them into the broader concept of low productivity and high productivity. This would make it possible for the business to see who is most and least productive, in order to reward top performers and provide additional support to those workers who need more training.

*Quay lại ví dụ về nhà sản xuất xem xét lại dữ liệu về nhân viên tại khu vực cửa hàng. Đầu tiên, những người này được phân nhóm theo loại và nhiệm vụ. Nhưng bây giờ, nhà phân tích dữ liệu có thể sử dụng các danh mục này và phân nhóm theo các khái niệm rộng hơn, bao gồm năng suất thấp và năng suất cao. Như vậy, doanh nghiệp có thể biết được ai có năng suất cao nhất và ai có năng suất thấp nhất, để trao thưởng cho những người có thành tích tốt nhất và hỗ trợ thêm cho những nhân viên cần đào tạo nhiều hơn.*

### Discovering connections

The problem type of discovering connections enables data analysts to find similar challenges faced by different entities, and then combine data and insights to address them.

*Bây giờ, loại vấn đề phát hiện các liên kết cho phép các nhà phân tích dữ liệu tìm thấy các thử thách tương tự mà nhiều thực thể khác nhau gặp phải, sau đó kết hợp các dữ liệu và thông tin chuyên sâu để xử lý vấn đề.*

Say a scooter company is experiencing an issue with the wheels it gets from its wheel supplier. That company would have to stop production until it could get safe, quality wheels back in stock. But meanwhile, the wheel companies encountering the problem with the rubber it uses to make wheels, turns out its rubber supplier could not find the right materials either. If all of these entities could talk about the problems they're facing and share data openly, they would find a lot of similar challenges and better yet, be able to collaborate to find a solution.

*Giả sử một công ty xe tay ga đang gặp phải vấn đề với bánh xe nhận được từ nhà cung cấp bánh xe. Công ty đó phải ngừng sản xuất đến khi có được nguồn bánh xe an toàn, chất lượng. Trong khi đó, công ty bánh xe gặp phải vấn đề với loại cao su được dùng để tạo ra bánh xe, hóa ra nhà cung cấp cao su không thể tìm được đúng loại nguyên liệu. Nếu tất cả các thực thể này có thể trao đổi về vấn đề mà họ đang gặp phải và chia sẻ dữ liệu một cách cởi mở, họ sẽ phát hiện thấy rất nhiều khó khăn giống nhau, và tốt hơn là, họ có thể hợp tác để tìm ra giải pháp.*

### Finding patterns

The final problem type is finding patterns. Data analysts use data to find patterns by using historical data to understand what happened in the past and is therefore likely to happen again.

*Loại vấn đề cuối cùng là tìm ra mẫu. Nhà phân tích dữ liệu sử dụng dữ liệu để tìm ra mẫu thông qua các dữ liệu trước đó để hiểu về những chuyện đã xảy ra trước đây và có khả năng sẽ tái diễn.*

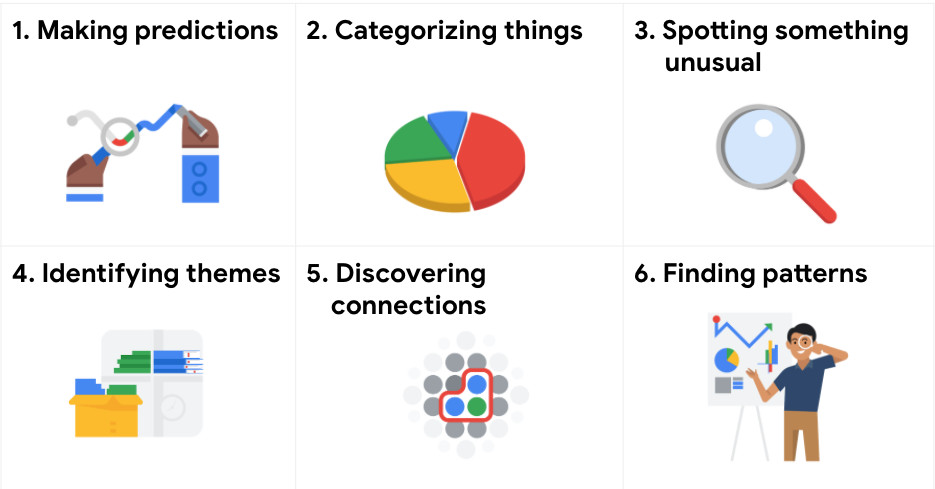
Ecommerce companies use data to find patterns all the time. Data analysts look at transaction data to understand customer buying habits at certain points in time throughout the year. They may find that customers buy more canned goods right before a hurricane, or they purchase fewer cold-weather accessories like hats and gloves during warmer months. The ecommerce companies can use these insights to make sure they stock the right amount of products at these key times.

*Các công ty thương mại điện tử thường xuyên sử dụng dữ liệu để tìm ra mẫu. Nhà phân tích xem xét dữ liệu giao dịch để hiểu được thói quen mua hàng của khách hàng ở những thời điểm nhất định trong năm. Họ có thể phát hiện rằng khách hàng mua nhiều đồ hợp hơn trước các trận bão, hoặc họ ít mua phụ kiện mùa lạnh như mũ hay găng tay vào các tháng ấm hơn. Các công ty thương mại điện tử có thể sử dụng các thông tin này để đảm bảo họ có đúng số lượng hàng trong kho vào những thời điểm quan trọng.*

## Six problem types

Data analytics is so much more than just plugging information into a platform to find insights. It is about solving problems. To get to the root of these problems and find practical solutions, there are lots of opportunities for creative thinking. No matter the problem, the first and most important step is understanding it. From there, it is good to take a problem-solver approach to your analysis to help you decide what information needs to be included, how you can transform the data, and how the data will be used.

### Data analysts typically work with six problem types



A video, [Common problem types](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/E8HxZ/common-problem-types), introduced the six problem types with an example for each. The examples are summarized below for review.

### Making predictions

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

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

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

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.

By now you might be wondering if there is a difference between categorizing things and identifying themes. The best way to think about it is: categorizing things involves assigning items to categories; identifying themes takes those categories a step further by grouping them into broader themes.

### Discovering connections

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.

### Finding patterns

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.

### Key takeaway

As you move through this program, you will develop a sharper eye for problems and you will practice thinking through the problem types when you begin your analysis. This method of problem solving will help you figure out solutions that meet the needs of all stakeholders.

## Problems in the real world

You've been learning about six common problem types of data analysts encounter:

* Making predictions
* Categorizing things
* Spotting something unusual
* Identifying themes
* Discovering connections
* Finding patterns.

Let's think back to our real world example from a previous section. In that example, anywhere gaming repair wanted to figure out how to bring in new customers. So the problem was, how to determine the best advertising method for anywhere gaming repair's target audience. To help solve this problem, the company used data to envision what would happen if it advertised in different places. Now nobody can see the future but the data helped them make an informed decision about how things would likely work out. So, their problem type was making predictions.

*Hãy nghĩ về ví dụ thực tế trong video trước. Trong ví dụ đó, Anywhere Gaming Repair muốn tìm cách thu hút khách hàng mới. Vấn đề là làm thế nào để xác định phương pháp quảng cáo hay nhất với đối tượng mục tiêu của Anywhere Gaming Repair. Để hỗ trợ giải quyết vấn đề này, công ty sử dụng dữ liệu để hình dung chuyện gì sẽ xảy ra nếu ta quảng cáo ở một địa điểm khác. Không ai có thể biết được tương lai nhưng dữ liệu sẽ giúp họ đưa ra quyết định sáng suốt về việc mọi chuyện sẽ diễn ra như thế nào. Như vậy, loại vấn đề của họ là* ***dự đoán****.*

Now let's think about the second problem type, categorizing things. Here's an example of a problem that involves categorization. Let's say a business wants to improve its customer satisfaction levels. Data analysts could review recorded calls to the company's customer service department and evaluate the satisfaction levels of each caller. They could identify certain key words or phrases that come up during the phone calls and then assign them to categories such as politeness, satisfaction, dissatisfaction, empathy, and more. Categorizing these key words gives us data that lets the company identify top performing customer service representatives, and those who might need more coaching. This leads to happier customers and higher customer service scores.

*Giả sử một doanh nghiệp muốn cải thiện mức độ hài lòng của khách hàng. Nhà phân tích dữ liệu có thể xem xét các cuộc gọi được ghi lại tại bộ phận chăm sóc khách hàng ở công ty và đánh giá mức độ hài lòng của từng người gọi. Họ có thể nhận diện một vài từ khóa hoặc cụm từ xuất hiện trong các cuộc gọi và phân loại vào các nhóm như lịch sự, hài lòng, không hài lòng, thấu cảm, và hơn thế nữa. Phân loại những từ khóa này cho chúng ta các dữ liệu để công ty xác định những người đại diện dịch vụ khách hàng có thành tích tốt nhất, và những người cần đào tạo thêm. Điều này giúp nhiều khách hàng hài lòng hơn và có được điểm số dịch vụ khách hàng cao hơn.*

Okay, now let's talk about a problem that involves spotting something unusual. Some of you may have a smart watch, my favorite app is for health tracking. These apps can help people stay healthy by collecting data such as their heart rate, sleep patterns, exercise routine, and much more. There are many stories out there about health apps actually saving people's lives. One is about a woman who was young, athletic, and had no previous medical problems. One night she heard a beep on her smartwatch, a notification said her heart rate had spiked. Now in this example think of the watch as a data analyst. The watch was collecting and analyzing health data. So when her resting heart rate was suddenly 120 beats per minute, the watch spotted something unusual because according to its data, the rate was normally around 70. Thanks to the data her smart watch gave her, the woman went to the hospital and discovered she had a condition which could have led to life threatening complications if she hadn't gotten medical help.

*Được rồi, bây giờ hãy nói về vấn đề liên quan đến việc phát hiện các điểm khác biệt. Có thể các bạn có đồng hồ thông minh, ứng dụng yêu thích của tôi là theo dõi sức khỏe. Ứng dụng này có thể giúp mọi người giữ gìn sức khỏe bằng cách thu thập các dữ liệu như nhịp tim, chu kỳ ngủ, thói quen tập thể dục, và nhiều hơn thế. Có rất nhiều câu chuyện về việc ứng dụng sức khỏe có thể cứu sống mạng người. Câu chuyện đầu tiên là về một phụ nữ trẻ, năng hoạt động và trước đây không có vấn đề nào về sức khỏe. Một đêm, cô ấy nghe thấy một tiếng bíp từ đồng hồ thông minh của mình, có thông báo rằng nhịp tim của cô ấy tăng cao. Trong ví dụ này, hãy coi đồng hồ là nhà phân tích dữ liệu. Đồng hồ đã thu thập và phân tích các dữ liệu sức khỏe. Khi nhịp tim lúc đang nghỉ của cô ấy tăng đến 120 nhịp một phút, đồng hồ phát hiện có điều gì đó bất thường vì theo dữ liệu, nhịp tim bình thường vào khoảng 70. Nhờ có dữ liệu từ đồng hồ thông minh, cô ấy đã tới bệnh viện và phát hiện mình gặp phải một căn bệnh có thể dẫn đến biến chứng nguy hiểm tới tính mạng nếu không được hỗ trợ về y tế.*

Now let's move on to the next type of problem: identifying themes. We see a lot of examples of this in the user experience field. User experience designers study and work to improve the interactions people have with products they use every day. Let's say a user experience designer wants to see what customers think about the coffee maker his company manufactures. This business collects anonymous survey data from users, which can be used to answer this question. But first to make sense of it all, he will need to find themes that represent the most valuable data, especially information he can use to make the user experience even better. So the problem the user experience designer's company faces, is how to improve the user experience for its coffee makers. The process here is kind of like finding categories for keywords and phrases in customer service conversations. But identifying themes goes even further by grouping each insight into a broader theme. Then the designer can pinpoint the themes that are most common. In this case he learned users often couldn't tell if the coffee maker was on or off. He ended up optimizing the design with improved placement and lighting for the on/off button, leading to the product improvement and happier users.

*Tiếp theo là loại vấn đề xác định chủ đề. Ta sẽ gặp nhiều ví dụ về loại này trong lĩnh vực trải nghiệm người dùng. Nhà thiết kế trải nghiệm người dùng nghiên cứu và làm việc để cải thiện tương tác của mọi người với sản phẩm họ dùng hàng ngày. Giả sử nhà thiết kế trải nghiệm người dùng muốn biết cảm nhận của khách hàng về máy pha cà phê mà công ty đó sản xuất. Doanh nghiệp này thu thập dữ liệu khảo sát ẩn danh từ người dùng, ta có thể sử dụng dữ liệu này để trả lời câu hỏi. Nhưng để hiểu được tất cả, họ cần tìm ra chủ đề chung đại diện cho dữ liệu có giá trị nhất, nhất là những thông tin họ có thể sử dụng để cải thiện trải nghiệm người dùng. Vấn đề mà công ty của nhà thiết kế trải nghiệm người dùng gặp phải là làm thế nào để cải thiện trải nghiệm người dùng cho máy pha cà phê. Quy trình ở đây giống như tìm kiếm các mục đối với từ khóa và cụm từ trong các cuộc trao đổi dịch vụ khách hàng. Nhưng việc xác định chủ đề còn hơn thế, ta sẽ nhóm các thông tin vào một chủ đề chung. Sau đó nhà thiết kế có thể xác định các chủ đề phổ biến mất. Trong trường hợp này, họ biết rằng người dùng không thể phân biệt máy pha cà phê có đang bật hay không. Cuối cùng họ đã tối ưu hóa thiết kế bằng cách cải thiện vị trí và ánh sáng của nút bật/tắt, từ đó cải tiến sản phẩm và khiến người dùng hài lòng hơn.*

Now we come to the problem of discovering connections. This example is from the transportation industry and uses something called third party logistics. Third party logistics partners help businesses ship products when they don't have their own trucks, planes or ships. A common problem these partners face is figuring out how to reduce wait time. Wait time happens when a truck driver from the third party logistics provider arrives to pick up a shipment but it's not ready. So she has to wait. That costs both companies time and money and it stops trucks from getting back on the road to make more deliveries. So how can they solve this? Well, by sharing data the partner companies can view each other's timelines and see what's causing shipments to run late. Then they can figure out how to avoid those problems in the future. So a problem for one business doesn't cause a negative impact for the other. For example, if shipments are running late because one company only delivers Mondays, Wednesdays and Fridays, and the other company only delivers Tuesdays and Thursdays, then the companies can choose to deliver on the same day to reduce wait time for customers.

*Bây giờ là đến vấn đề phát hiện liên kết. Ví dụ này là từ ngành giao thông vận tải, ta áp dụng khái niệm dịch vụ hậu cần của bên thứ ba. Các đối tác hậu cần bên thứ ba giúp các doanh nghiệp vận chuyển sản phẩm khi không có xe tải, máy bay hoặc tàu riêng. Một vấn đề phổ biến mà các đối tác này gặp phải là việc giảm thời gian chờ đợi. Thời gian chờ đợi là khi tài xế xe tải của nhà cung cấp dịch vụ hậu cần bên thứ ba đến nơi nhận lô hàng nhưng lô hàng lại chưa sẵn sàng. Vậy là họ phải đợi. Việc này làm tiêu tốn cả thời gian và tiền bạc, khiến xe tải không thể tiếp tục hành trình vận chuyển. Làm thế nào để giải quyết vấn đề này? Bằng cách chia sẻ dữ liệu, các công ty đối tác có thể xem lịch trình của nhau và biết được nguyên nhân giao hàng chậm trễ. Từ đó họ có thể tìm cách ngăn chặn các vấn đề này trong tương lai. Vậy là vấn đề của một doanh nghiệp sẽ không gây ra tác động tiêu cực cho doanh nghiệp kia. Ví dụ, nếu lô hàng bị chậm trễ vì có một công ty chỉ giao hàng vào thứ Hai, thứ Tư và thứ Sáu, và công ty khác chỉ giao hàng vào thứ Ba và thứ Năm, thì các công ty có thể lựa chọn giao hàng vào cùng ngày để giảm thiểu thời gian chờ đợi cho khách hàng.*

All right, we've come to our final problem type, finding patterns. Oil and gas companies are constantly working to keep their machines running properly. So the problem is, how to stop machines from breaking down. One way data analysts can do this is by looking at patterns in the company's historical data. For example, they could investigate how and when a particular machine broke down in the past and then generate insights into what led to the breakage. In this case, the company saw pattern indicating that machines began breaking down at faster rates when maintenance wasn't kept up in 15 day cycles. They can then keep track of current conditions and intervene if any of these issues happen again. Pretty cool, right? I'm always amazed to hear about how data helps real people and businesses make meaningful change. I hope you are too. See you soon.

*Và bây giờ là loại vấn đề cuối cùng, tìm ra mẫu. Các công ty dầu khí phải liên tục làm việc để giữ cho máy móc của họ hoạt động đúng cách. Vì vậy, vấn đề là, làm thế nào để ngăn máy móc ngừng hoạt động. Nhà phân tích dữ liệu có thể xem xét các mẫu trong dữ liệu trước đó của công ty. Ví dụ, họ có thể tìm hiểu cách thức và thời điểm một máy cụ thể ngừng hoạt động trước đây và tìm hiểu thông tin về nguyên nhân gây ra việc đó. Trong trường hợp này, công ty phát hiện được mẫu chỉ ra rằng máy bắt đầu hỏng ở tốc độ nhanh hơn nếu không thể bảo dưỡng theo chu kỳ 15 ngày. Sau đó họ có thể theo dõi điều kiện hiện tại và can thiệp nếu xảy ra vấn đề này lần nữa.*

Pretty cool, right? I'm always amazed to hear about how data helps real people and businesses make meaningful change.

*Khá là tuyệt, phải không? Tôi luôn ngạc nhiên khi nghe về việc dữ liệu có thể giúp các cá nhân và doanh nghiệp trong thực tế tạo ra thay đổi có ý nghĩa như thế nà*

## SMART questions

Now that we've talked about six basic problem types, it's time to start solving them. To do that, data analysts start by asking the right questions. In this section, we're going to learn how to ask effective questions that lead to key insights you can use to solve all kinds of problems.

*Bây giờ chúng ta đã nói về sáu loại vấn đề cơ bản, đã đến lúc bắt đầu giải quyết chúng. Để làm được điều đó, các nhà phân tích dữ liệu bắt đầu bằng cách đặt những câu hỏi phù hợp. Trong phần này, chúng ta sẽ tìm hiểu cách đặt câu hỏi hiệu quả dẫn đến những hiểu biết chính mà bạn có thể sử dụng để giải quyết mọi loại vấn đề.*

As a data analyst, I ask questions constantly. It's a huge part of the job. If someone requests that I work on a project, I ask questions to make sure we're on the same page about the plan and the goals. And when I do get a result, I question it. Is the data showing me something superficially? Is there a conflict somewhere that needs to be resolved? The more questions you ask, the more you'll learn about your data and the more powerful your insights will be at the end of the day.

*Là một nhà phân tích dữ liệu, tôi liên tục đặt câu hỏi. Đó là một phần rất lớn của công việc. Nếu ai đó yêu cầu tôi làm việc trong một dự án, tôi sẽ đặt câu hỏi để đảm bảo rằng chúng tôi có cùng quan điểm về kế hoạch và mục tiêu. Và khi tôi nhận được một kết quả, tôi tiếp tục đặt câu hỏi về nó. Là bởi vì dữ liệu cho tôi thấy một cái gì đó hời hợt chăng? Có xung đột ở đâu đó cần được giải quyết không? Bạn càng đặt nhiều câu hỏi, bạn càng tìm hiểu kỹ hơn về dữ liệu của mình và sự hiểu biết sâu sắc của bạn sẽ càng mạnh mẽ hơn vào cuối ngày.*

Some questions are more effective than others. Let's say you're having lunch with a friend and they say, "These are the best sandwiches ever, aren't they?" Well, that 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. Or maybe you're working on a project and you decide to interview a family member. Say you ask your uncle, did you enjoy growing up in Malaysia? He may reply, "Yes." But you haven't learned much about his experiences there. Your question was closed-ended. That means it can be answered with a yes or no. These kinds of questions rarely lead to valuable insights. Now what if someone asks you, do you prefer chocolate or vanilla? Well, what are they specifically talking about? Ice cream, pudding, coffee flavoring or something else? What if you like chocolate ice cream but vanilla in your coffee? What if you don't like either flavor? That's the problem with this question. It's too vague and lacks context. Knowing the difference between effective and ineffective questions is essential for your future career as a data analyst.

*Một số câu hỏi hiệu quả hơn những câu hỏi khác. Giả sử bạn đang ăn trưa với một người bạn và họ nói, "Đây là những chiếc bánh mì kẹp ngon nhất từ trước đến nay phải không?" Chà, câu hỏi đó không thực sự cho bạn cơ hội để chia sẻ ý kiến ​​của riêng mình, đặc biệt nếu bạn không đồng ý và không thích bánh sandwich lắm. Đây được gọi là câu hỏi dẫn dắt vì nó dẫn bạn đến câu trả lời theo một cách nhất định. Hoặc có thể bạn đang thực hiện một dự án và bạn quyết định phỏng vấn một thành viên trong gia đình. Giả sử bạn hỏi chú của bạn, bạn có thích lớn lên ở Malaysia không? Anh ta có thể trả lời, "Có." Nhưng bạn chưa học được nhiều về kinh nghiệm của anh ấy ở đó. Câu hỏi của bạn đã được đóng lại. Điều đó có nghĩa là nó có thể được trả lời bằng có hoặc không. Những loại câu hỏi này hiếm khi dẫn đến những hiểu biết có giá trị. Bây giờ nếu ai đó hỏi bạn, bạn thích sô cô la hay vani hơn? Chà, cụ thể họ đang nói về cái gì vậy? Kem, bánh pudding, hương liệu cà phê hay thứ gì khác? Nếu bạn thích kem sô cô la nhưng vani trong cà phê của bạn thì sao? Nếu bạn không thích một trong hai hương vị thì sao? Đó là vấn đề với câu hỏi này. Nó quá mơ hồ và thiếu ngữ cảnh. Biết được sự khác biệt giữa các câu hỏi hiệu quả và không hiệu quả là điều cần thiết cho sự nghiệp tương lai của bạn với tư cách là một nhà phân tích dữ liệu.*

After all, the data analyst process starts with the ask phase. So it's important that we ask the right questions. Effective questions follow the **SMART** methodology. That means they're:

*Rốt cuộc, quá trình phân tích dữ liệu bắt đầu với giai đoạn hỏi. Vì vậy, điều quan trọng là chúng tôi đặt câu hỏi đúng. Các câu hỏi hiệu quả tuân theo phương pháp* ***SMART****. Điều đó có nghĩa là:*

* ***S***pecific *cụ thể*
* ***M***easurable *có thể đo lường*
* ***A***ction-oriented *định hướng hành động*
* ***R***elevant *phù hợp*
* ***T***ime-bound *có thời hạn*

### Specific questions

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. For example, instead of asking a closed-ended question, like, are kids getting enough physical activities these days? Ask what percentage of kids achieve the recommended 60 minutes of physical activity at least five days a week? That question is much more specific and can give you more useful information.

*Các câu hỏi cụ thể đơn giản, quan trọng và tập trung vào một chủ đề duy nhất hoặc một vài ý tưởng liên quan chặt chẽ. Điều này giúp chúng tôi thu thập thông tin liên quan đến những gì chúng tôi đang điều tra. Nếu một câu hỏi quá chung chung, hãy cố gắng thu hẹp nó lại bằng cách chỉ tập trung vào một yếu tố. Ví dụ, thay vì hỏi một câu hỏi đóng, chẳng hạn như ngày nay trẻ em có hoạt động thể chất đủ không? Hãy hỏi có bao nhiêu phần trăm trẻ em đạt được 60 phút hoạt động thể chất được khuyến nghị ít nhất năm ngày một tuần? Câu hỏi đó cụ thể hơn nhiều và có thể cung cấp cho bạn nhiều thông tin hữu ích hơn.*

### Measurable questions

Now, let's talk about measurable questions. Measurable questions can be quantified and assessed. An example of an unmeasurable question would be, why did a recent video go viral? Instead, you could ask how many times was our video shared on social channels the first week it was posted? That question is measurable because it lets us count the shares and arrive at a concrete number.

*Bây giờ, hãy nói về những câu hỏi có thể đo lường được. Các câu hỏi đo lường có thể được định lượng và đánh giá. Một ví dụ về câu hỏi không thể đo lường được là tại sao một video gần đây lại lan truyền như vậy? Thay vào đó, bạn có thể hỏi video của chúng tôi đã được chia sẻ bao nhiêu lần trên các kênh xã hội trong tuần đầu tiên nó được đăng? Câu hỏi đó có thể đo lường được vì nó cho phép chúng tôi đếm các lượt chia sẻ và đi đến một con số cụ thể.*

### Action-oriented questions

Okay, now we've come to action-oriented questions. Action-oriented questions encourage change. You might remember that problem solving is about seeing the current state and figuring out how to transform it into the ideal future state. Well, action-oriented questions help you get there. So 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.

*Được rồi, bây giờ chúng ta đến với các câu hỏi định hướng hành động. Các câu hỏi định hướng hành động khuyến khích sự thay đổi. Bạn có thể nhớ rằng giải quyết vấn đề là nhìn thấy trạng thái hiện tại và tìm ra cách biến nó thành trạng thái lý tưởng trong tương lai. Chà, các câu hỏi định hướng hành động sẽ giúp bạn đạt được điều đó. Vì vậy, thay vì hỏi, làm thế nào chúng ta có thể khiến khách hàng tái chế bao bì sản phẩm của mình? Bạn có thể hỏi, những tính năng thiết kế nào sẽ giúp bao bì của chúng ta dễ tái chế hơn? Điều này mang lại cho bạn câu trả lời mà bạn có thể hành động.*

### Relevant questions

All right, let's move on to relevant questions. Relevant questions matter, are important and have significance to the problem you're trying to solve. Let's say you're working on a problem related to a threatened species of frog. And you asked, why does it matter that Pine Barrens tree frogs started disappearing? This is an irrelevant question because the answer won't help us find a way to prevent these frogs from going extinct. A more relevant question would be, what environmental factors changed in Durham, North Carolina between 1983 and 2004 that could cause Pine Barrens tree frogs to disappear from the Sandhills Regions? This question would give us answers we can use to help solve our problem.

*Được rồi, hãy chuyển sang các câu hỏi liên quan. Các câu hỏi liên quan quan trọng, quan trọng và có ý nghĩa đối với vấn đề bạn đang cố gắng giải quyết. Giả sử bạn đang giải quyết một vấn đề liên quan đến một loài ếch đang bị đe dọa. Và bạn đã hỏi, tại sao ếch cây Pine Barrens bắt đầu biến mất? Đây là một câu hỏi không liên quan vì câu trả lời sẽ không giúp chúng ta tìm ra cách ngăn chặn những con ếch này tuyệt chủng. Một câu hỏi phù hợp hơn sẽ là, những yếu tố môi trường nào đã thay đổi ở Durham, Bắc Carolina từ năm 1983 đến năm 2004 có thể khiến ếch cây Pine Barrens biến mất khỏi Vùng Sandhills? Câu hỏi này sẽ cung cấp cho chúng tôi câu trả lời mà chúng tôi có thể sử dụng để giúp giải quyết vấn đề của mình.*

### Time-bound questions

The example above is also a great example for our final point, time-bound questions. Time-bound questions specify the time to be studied. The time period we want to study is 1983 to 2004. This limits the range of possibilities and enables the data analyst to focus on relevant data.

*Ví dụ ở trên cũng là một ví dụ tuyệt vời cho điểm cuối cùng của chúng tôi, các câu hỏi giới hạn thời gian. Câu hỏi giới hạn thời gian xác định thời gian được nghiên cứu. Khoảng thời gian chúng tôi muốn nghiên cứu là từ năm 1983 đến năm 2004. Điều này giới hạn phạm vi khả năng và cho phép nhà phân tích dữ liệu tập trung vào dữ liệu liên quan.*

### Fairness

Okay, now that you have a general understanding of SMART questions, there's something else that's very important to keep in mind when crafting questions, fairness. We've touched on fairness before, but as a quick reminder, fairness means ensuring that your questions don't create or reinforce bias.

*Được rồi, bây giờ bạn đã có hiểu biết chung về các câu hỏi SMART, có một điều khác rất quan trọng cần lưu ý khi đặt câu hỏi, đó là sự công bằng. Chúng tôi đã đề cập đến tính công bằng trước đây, nhưng xin nhắc lại, tính công bằng có nghĩa là đảm bảo rằng các câu hỏi của bạn không tạo ra hoặc củng cố sự thiên vị.*

To talk about this, let's go back to our sandwich example. There we had an unfair question because it was phrased to lead you toward a certain answer. This made it difficult to answer honestly if you disagreed about the sandwich quality.

*Để nói về điều này, hãy quay lại ví dụ về bánh sandwich của chúng ta. Ở đó, chúng tôi có một câu hỏi không công bằng vì nó được đặt ra để dẫn bạn đến một câu trả lời nhất định. Điều này khiến bạn khó trả lời một cách trung thực nếu bạn không đồng ý về chất lượng bánh sandwich.*

Another common example of an unfair question is one that makes assumptions. For instance, let's say a satisfaction survey is given to people who visit a science museum. If the survey asks, what do you love most about our exhibits? This assumes that the customer loves the exhibits which may or may not be true. 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.

*Để nói về điều này, hãy quay lại ví dụ về bánh sandwich của chúng ta. Ở đó, chúng tôi có một câu hỏi không công bằng vì nó được đặt ra để dẫn bạn đến một câu trả lời nhất định. Điều này khiến bạn khó trả lời một cách trung thực nếu bạn không đồng ý về chất lượng bánh sandwich. Một ví dụ phổ biến khác về câu hỏi không công bằng là câu hỏi đưa ra các giả định. Ví dụ: giả sử một cuộc khảo sát về mức độ hài lòng dành cho những người đến thăm bảo tàng khoa học. Nếu cuộc khảo sát hỏi, bạn thích điều gì nhất về triển lãm của chúng tôi? Điều này giả định rằng khách hàng yêu thích các cuộc triển lãm, điều này có thể đúng hoặc không. Công bằng cũng có nghĩa là đặt ra những câu hỏi có ý nghĩa với tất cả mọi người. Điều quan trọng là các câu hỏi phải rõ ràng và có cách diễn đạt đơn giản mà bất kỳ ai cũng có thể dễ dàng hiểu được. Những câu hỏi không công bằng cũng có thể khiến công việc của bạn với tư cách là một nhà phân tích dữ liệu trở nên khó khăn hơn. Chúng dẫn đến phản hồi không đáng tin cậy và bỏ lỡ cơ hội để đạt được một số hiểu biết thực sự có giá trị.*

## More about SMART questions

Companies in lots of industries today are dealing with rapid change and rising uncertainty. Even well-established businesses are under pressure to keep up with what is new and figure out what is next. To do that, they need to ask questions. Asking the right questions can help spark the innovative ideas that so many businesses are hungry for these days.

*Các công ty trong nhiều ngành công nghiệp ngày nay đang phải đối phó với sự thay đổi nhanh chóng và sự không chắc chắn gia tăng. Ngay cả những doanh nghiệp lâu đời cũng phải chịu áp lực phải bắt kịp những gì mới và tìm ra những gì tiếp theo. Để làm được điều đó, họ cần phải đặt câu hỏi. Đặt câu hỏi đúng có thể giúp khơi dậy những ý tưởng đổi mới mà rất nhiều doanh nghiệp đang khao khát hiện nay.*

The same goes for data analytics. No matter how much information you have or how advanced your tools are, your data won’t tell you much if you don’t start with the right questions. Think of it like a detective with tons of evidence who doesn’t ask a key suspect about it.  Coming up, you will learn more about how to ask highly effective questions, along with certain practices you want to avoid.

*Điều tương tự cũng xảy ra với phân tích dữ liệu. Cho dù bạn có bao nhiêu thông tin hay công cụ của bạn tiên tiến đến đâu, dữ liệu của bạn sẽ không cho bạn biết nhiều nếu bạn không bắt đầu với những câu hỏi phù hợp. Hãy tưởng tượng như một thám tử có hàng đống bằng chứng mà không hỏi một nghi phạm chủ chốt về chúng. Tiếp theo, bạn sẽ tìm hiểu thêm về cách đặt câu hỏi hiệu quả cao, cùng với một số thực hành nhất định mà bạn muốn tránh.*

### Highly effective questions are SMART questions:

A picture containing text, screenshot, font, number

Description automatically generated

Các câu trả lời có liên quan đến thời gian cụ thể đang được nghiên cứu không?

Câu hỏi này có liên quan đến vấn đề cụ thể mà bạn đang cố gắng giải quyết không?

Các câu trả lời có thể cung cấp thông tin giúp bạn lập kế hoạch thực hiện một loại kế hoạch hành động không?

Câu hỏi này có thể cung cấp cho bạn những câu trả lời có thể đo lường được không?

Câu hỏi này có rõ ràng không? Nó có liên quan đến vấn đề được đề cập? Nó có ngữ cảnh không? Nó có thể khám phá được nhiều thông tin mà bạn cần không?

### Examples of SMART questions

Here's an example that breaks down the thought process of turning a problem question into one or more SMART questions using the SMART method:

**What features do people look for when buying a new car?**

*Mọi người tìm kiếm những tính năng nào khi mua một chiếc ô tô mới?*

* **Specific**: Does the question focus on a particular car feature?

*Câu hỏi có tập trung vào một tính năng cụ thể của ô tô không?*

* **Measurable**: Does the question include a feature rating system?

*Câu hỏi có bao gồm hệ thống đánh giá tính năng không?*

* **Action-oriented**: Does the question influence creation of different or new feature packages?

*Câu hỏi có ảnh hưởng đến việc tạo ra các gói tính năng mới hoặc khác biệt không?*

* **Relevant**: Does the question identify which features make or break a potential car purchase?

*Câu hỏi có xác định những tính năng nào tạo ra hoặc phá vỡ khả năng mua xe hơi không?*

* **Time-bound**: Does the question validate data on the most popular features from the last three years?

*Câu hỏi có xác thực dữ liệu về các tính năng phổ biến nhất trong ba năm qua không?*

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. So, based on the thought process, possible SMART questions might be:

*Các câu hỏi nên có kết thúc mở. Đây là cách tốt nhất để nhận phản hồi giúp bạn xác định chính xác hoặc loại bỏ các giải pháp tiềm năng cho vấn đề cụ thể của mình. Vì vậy, dựa trên quá trình suy nghĩ, các câu hỏi SMART có thể là:*

* On a scale of 1-10 (with 10 being the most important) how important is your car having four-wheel drive?

*Trên thang điểm từ 1-10 (với 10 là quan trọng nhất), xe của bạn có hệ dẫn động bốn bánh quan trọng như thế nào?*

* What are the top five features you would like to see in a car package?

*5 tính năng hàng đầu bạn muốn thấy trong một gói xe hơi là gì?*

* What features, if included with four-wheel drive, would make you more inclined to buy the car?

*Những tính năng nào, nếu đi kèm với hệ dẫn động bốn bánh, sẽ khiến bạn có xu hướng mua xe hơn?*

* How much more would you pay for a car with four-wheel drive?

*Những tính năng nào, nếu đi kèm với hệ dẫn động bốn bánh, sẽ khiến bạn có xu hướng mua xe hơn?*

* Has four-wheel drive become more or less popular in the last three years?

*Xe bốn bánh đã trở nên phổ biến hơn hay ít hơn trong ba năm qua?*

### Things to avoid when asking questions

Những điều nên tránh khi đặt câu hỏi

**Leading questions**: Questions that only have a particular response. Example: **This product is too expensive, isn’t it?**

*Câu hỏi chỉ có một câu trả lời cụ thể. Ví dụ: Sản phẩm này đắt quá phải không?*

This is a leading question because it suggests an answer as part of the question. A better question might be, “What is your opinion of this product?” There are tons of answers to that question, and they could include information about usability, features, accessories, color, reliability, and popularity, on top of price. Now, if your problem is actually focused on pricing, you could ask a question like “What price (or price range) would make you consider purchasing this product?” This question would provide a lot of different measurable responses.

*Đây là một câu hỏi dẫn dắt bởi vì nó gợi ý một câu trả lời như một phần của câu hỏi. Một câu hỏi hay hơn có thể là, “Ý kiến của bạn về sản phẩm này như thế nào?” Có rất nhiều câu trả lời cho câu hỏi đó và chúng có thể bao gồm thông tin về khả năng sử dụng, tính năng, phụ kiện, màu sắc, độ tin cậy và mức độ phổ biến cùng với giá cả. Bây giờ, nếu vấn đề của bạn thực sự tập trung vào giá cả, bạn có thể đặt câu hỏi như “Mức giá (hoặc phạm vi giá) nào sẽ khiến bạn cân nhắc mua sản phẩm này?” Câu hỏi này sẽ cung cấp rất nhiều câu trả lời có thể đo lường khác nhau.*

**Closed-ended questions**: questions that ask for a one-word or brief response only. Example: **Were you satisfied with the customer trial?**

*Câu hỏi chỉ yêu cầu một từ hoặc câu trả lời ngắn gọn. Ví dụ: Bạn có hài lòng với bản dùng thử của khách hàng không?*

This is a closed-ended question because it doesn’t encourage people to expand on their answer. It is really easy for them to give one-word responses that aren’t very informative. A better question might be, “What did you learn about customer experience from the trial.” This encourages people to provide more detail besides “It went well.”

*Đây là một câu hỏi đóng vì nó không khuyến khích mọi người mở rộng câu trả lời của họ. Họ thực sự dễ dàng đưa ra những câu trả lời một từ không có nhiều thông tin. Một câu hỏi hay hơn có thể là, "Bạn đã học được gì về trải nghiệm của khách hàng từ bản dùng thử." Điều này khuyến khích mọi người cung cấp thêm thông tin chi tiết bên cạnh câu “Mọi việc diễn ra tốt đẹp”.*

**Vague questions:** questions that aren’t specific or don’t provide context. Example: **Does the tool work for you?**

*Câu hỏi không cụ thể hoặc không cung cấp ngữ cảnh. Ví dụ: Công cụ này có phù hợp với bạn không?*

This question is too vague because there is no context. Is it about comparing the new tool to the one it replaces? You just don’t know. A better inquiry might be, “When it comes to data entry, is the new tool faster, slower, or about the same as the old tool? If faster, how much time is saved? If slower, how much time is lost?” These questions give context (data entry) and help frame responses that are measurable (time).

*Câu hỏi này quá mơ hồ vì không có ngữ cảnh. Có phải là so sánh công cụ mới với công cụ mà nó thay thế không? Bạn chỉ không biết. Một câu hỏi hay hơn có thể là, “Khi nói đến mục nhập dữ liệu, công cụ mới nhanh hơn, chậm hơn hay gần giống như công cụ cũ? Nếu nhanh hơn thì tiết kiệm được bao nhiêu thời gian? Nếu chậm hơn thì mất bao nhiêu thời gian?” Những câu hỏi này đưa ra ngữ cảnh (nhập dữ liệu) và khung trợ giúp có thể đo lường được (thời gian).*