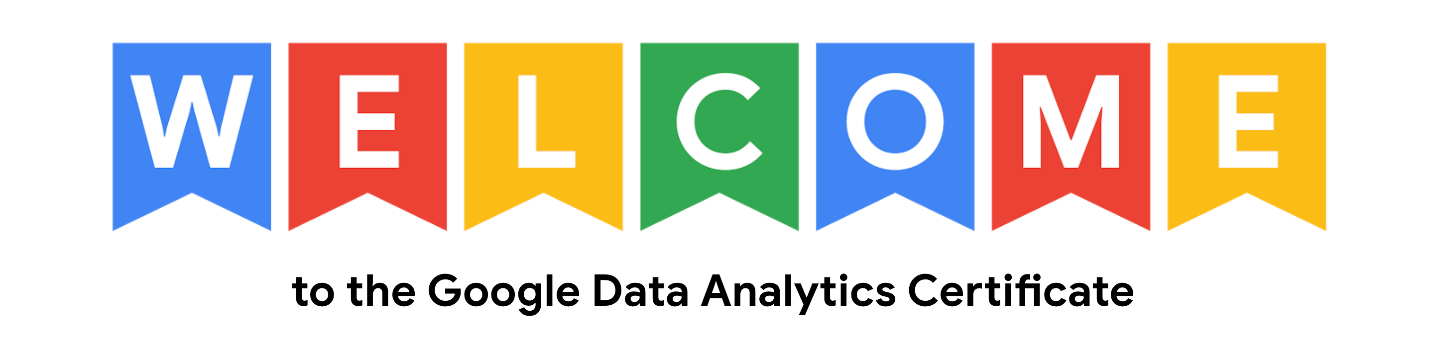
# Program description and course syllabus



Hello and welcome! The program you are about to explore is specifically designed to help every type of learner successfully finish the certificate and become an entry-level junior or associate data analyst. No previous data analytics, mathematics, or statistical experience is required. To succeed, you just need to be open to learning how data influences the world.

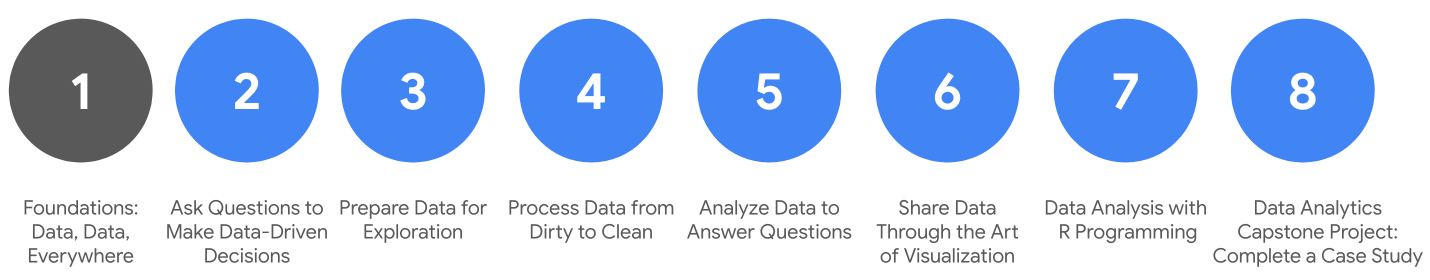
## Become job-ready

Every day, the amount of data out there gets bigger and bigger. So the ability to interpret it effectively is more important than ever before. Data analytics is becoming one of the fastest-growing and most rewarding career choices in the world. In the next decade, the demand for business analytics skills will probably be higher than the demand for any other career (10.9% vs. 5.2%) (Source: Bureau of Labor Statistics). All kinds of companies all over the world need qualified data analysts to solve problems and help them make the best possible business decisions. And right now, fifty-nine percent of companies have plans to add even more positions requiring data analysis skills (Source: SHRM). By the time you are done with this program, you will be well-prepared to make smart, strategic, data-driven recommendations for organizations in all kinds of industries.

During each course of the program, you will complete lots of hands-on assignments and projects based on both day-to-day life and the practical activities of a data analyst. Along the way, you will learn how to ask the right questions and understand objectives. You will also learn how to effectively clean and organize large amounts of data to make it ready for high-quality analysis. On top of that, you will get hands-on experience using all kinds of tools and techniques that will help you recognize patterns and uncover relationships between data points. And to help you communicate the results of your analysis, you will learn how to design visuals and dashboards. There is even an opportunity to create a case study, which you can highlight in your resume to show what you have learned to potential employers.

## Course overview

The entire program has eight courses. This is the first course and it covers about five weeks of material.



1. **Foundations: Data, Data, Everywhere** (this course)
2. [Ask Questions to Make Data-Driven Decisions](https://www.coursera.org/learn/ask-questions-make-decisions/home/welcome)
3. [Prepare Data for Exploration](https://www.coursera.org/learn/data-preparation/home/welcome)
4. [Process Data from Dirty to Clean](https://www.coursera.org/learn/process-data/home/welcome)
5. [Analyze Data to Answer Questions](https://www.coursera.org/learn/analyze-data/home/welcome)
6. [S​hare Data Through the Art of Visualization](https://www.coursera.org/learn/visualize-data/home/welcome)
7. [Data Analysis with R Programming](https://coursera.org/learn/data-analysis-r/home/welcome)
8. [Google D​ata Analytics Capstone: Complete a Case Study](https://coursera.org/learn/google-data-analytics-capstone/home/welcome)

## C​ourse content

C​ourse 1– Foundations: Data, Data, Everywhere

1. **I​ntroducing data analytics:** Data helps us make decisions, in everyday life and in business. In this first part of the course, you will learn how data analysts use tools of their trade to inform those decisions. You will also get to know more about this course and the overall program expectations.
2. **T​hinking analytically:** Data analysts balance many different roles in their work. In this part of the course, you will learn about some of these roles and the key skills that are required. You will also explore analytical thinking and how it relates to data-driven decision making.
3. **E​xploring the wonderful world of data:** Data has its own life cycle, and data analysts use an analysis process that cuts across and leverages this life cycle. In this part of the course, you will learn about the data life cycle and data analysis process. They are both relevant to your work in this program and on the job as a future data analyst. You will be introduced to applications that help guide data through the data analysis process.
4. **S​etting up a data toolbox:** Spreadsheets, query languages, and data visualization tools are all a big part of a data analyst’s job. In this part of the course, you will learn the basic concepts to use them for data analysis. You will understand how they work through examples provided.
5. **D​iscovering data career possibilities:** All kinds of businesses value the work that data analysts do. In this part of the course, you will examine different types of businesses and the jobs and tasks that analysts do for them. You will also learn how a Google Data Analytics Certificate will help you meet many of the requirements for a position with these organizations.
6. **C​ompleting the Course Challenge:** At the end of this course, you will be able to put everything you have learned into perspective with the Course Challenge. The Course Challenge will ask you questions about the main concepts you have learned and then give you an opportunity to apply those concepts in two scenarios.

## W​hat to expect

Each week of the course includes a series of lessons with many types of learning opportunities. These include:

* **V​ideos** for instructors toteach new concepts and demonstrate the use of tools
* **Readings** to introduce new ideas and build on the concepts from the videos
* [**Discussion forums**](https://www.coursera.org/learn/foundations-data/discussions) to share, explore, and reinforce lesson topics for better understanding
* **D​iscussion prompts** to promote thinking and engagement in the discussion forums
* **Practice quizzes** to prepare you for graded quizzes
* **Graded quizzes** to measure your progress and give you valuable feedback
* Also, be sure to pay attention to the **in-video questions** that will pop up from time to time. They are designed for you to check your learning.

Everyone learns differently, so this program has been designed to let you work at your own pace. Although your personalized deadlines start when you enroll, they are just a guide. Feel free to move through the program at the speed that works best for you. There is no penalty for late assignments; to earn your certificate, all you have to do is complete all of the work. If you prefer, you can extend your deadlines by returning to **Overview** in the navigation panel and clicking **Switch Sessions**. Assessments are based on the approach taken by the course to offer a wide variety of learning materials and activities that reinforce important skills. Graded and ungraded quizzes will help the content sink in. Ungraded practice quizzes are a chance for you to prepare for the graded quizzes. Both types of quizzes can be taken more than one time.

## Optional speed track for those experienced in data analytics

The Google Data Analytics Certificate provides instruction and feedback for learners hoping to earn a position as an entry-level data analyst. While many learners will be brand new to the world of data analytics, others may be familiar with the field and simply wanting to brush up on certain skills.

If you believe this course will be primarily a refresher for you, we recommend taking the practice diagnostic quiz (you can find it in this week's content). It will enable you to determine if you should follow the speed track, which is an opportunity to proceed to Course 2 after having taken each of the Course 1 Weekly Challenges and the overall Course Challenge. Learners who score 100% on the diagnostic quiz can treat Course 1 videos, readings, and activities as optional. Learners following the speed track are still able to earn the certificate.

## Tips

* It is strongly recommended to take these courses—and go through the items in each lesson—in the order they appear because new information and concepts build on previous knowledge.
* Use the additional resources that are linked throughout the program. They are designed to support your learning.
* W​hen you encounter useful links in the course, remember to bookmark them so you can refer to the information for study or review.
* Additional resources are free, but some sites place limits on how many articles can be accessed for free each month. Sometimes you can register on the site for full access, but you can always bookmark a resource and come back to view it later.
* If something is confusing, don’t hesitate to re-watch a video, go through a reading again, and so on.
* Take part in all learning opportunities to gain as much knowledge and experience possible.

Congratulations on choosing to take this first step toward becoming part of the wonderful world of data analytics. Enjoy the journey!

# Learning Log: Think about data in daily life



## Overview



By now, you've started to discover how powerful data can be. Throughout this course, you’ll be asked to make entries in a learning log. Your log will be a personal space where you can keep track of your thinking and reflections about the experiences you will have collecting and analyzing data. Reflections may include what you liked, what you would change, and questions that were raised. By the time you complete the entry for this activity, you will have a stronger understanding of data analytics.

## Everyday data



Before you write an entry in your learning log, think about where and how you use data to make decisions. You will create a list of at least **five questions** that you might use data to answer. Here are a few examples to inspire you:

* What’s the best time to go to the gym?
* How does the length of your commute to work vary by day of the week?
* How many cups of coffee do you drink each day?
* What flavor of ice cream do customers buy?
* How many hours of sleep do you get each day?

Then, you will select **one** of the five questions from your list to explore further and write down the types of data you might collect in order to make a decision. That’s data analysis in action!



### Access your learning log

To use the learning log for this course item, click the link below and select Use Template.

Link to learning log template: [Think about data in daily life](https://docs.google.com/document/d/1j6kyM_KuLttOR8bNXbFZ9kj_WKVOdeJnChTiTxvvsv8/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the attachment below.

**Learning Log Template\_ Think about data in daily life**DOCX File

[Download file](https://d3c33hcgiwev3.cloudfront.net/DNykxIj4TLicpMSI-Dy4Dg_678c43f1933349da98708aab5e6feb7f_Learning-Log-Template_-Think-about-data-in-daily-life.docx?Expires=1648080000&Signature=fMa-cF9qld23ml3-gXwn3dA9fcPfHzFtpt~-yPCoNGeLznH9q0WGiSDyFlYFYYhe~ZQV90IVYXxYt0zUPzyZspm3sp2LVRtoQY9As2TmS~UneKIDmXPyPqxeGyZc75P6NhC-SuCNmMHB3Id0t-QfpTn7YsHr1vPn5xZO5kHOaNc_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)



## Reflection



After you consider how you use data analysis in your own life, take a moment to reflect on what you discovered. Reflections may include what you liked, what you would change, and questions that were raised. In your new learning log entry, you will write 2-3 sentences (40-60 words) in response to each question below:

* What are some considerations or preferences you want to keep in mind when making a decision?
* What kind of information or data do you have access to that will influence your decision?
* Are there any other things you might want to track associated with this decision?

When you’ve finished your entry in the learning log template, make sure to save the document so your response is somewhere accessible. This will help you continue applying data analysis to your everyday life. You will also be able to track your progress and growth as a data analyst.

# Helpful resources to get started

The Google Data Analytics Certificate is designed to provide you with new lessons every week. As you’ve learned, each one includes a series of videos, readings, peer discussions, in-video questions, practice quizzes, and graded quizzes. In this reading, you’ll learn about providing feedback on course content, obtaining the Google Data Analytics Certificate, and helpful habits for successfully completing the certificate.

## Providing feedback or getting help on course content

Please remember to give feedback on videos, readings, and materials. Just open the resource, and look for the thumbs-up and thumbs-down symbols.

* Click thumbs-up for materials that are helpful.
* Click thumbs-down for materials that are not helpful.

That feedback goes to the course developers, not other learners, and helps improve this course.

For technical help on Coursera, visit the [Learner Help Center](https://www.coursera.support/s/learner-help-center). For help accessing course materials, click the Contact us link at the bottom of the page.

## Obtaining the Google Data Analytics Certificate

After you complete all eight courses, you qualify for the Google Data Analytics Certificate.

To receive your certificate, you must:

* Pass all required assignments in the course or meet the course-passing threshold. Each graded assignment is part of a cumulative graded score, and the passing grade for the Google Data Analytics Certificate is 80%.

AND

* Pay the Course Certificate fee ($39/month, with most learners completing the material in 6 months or less), or apply and be approved for a [Coursera scholarship](https://learner.coursera.help/hc/en-us/articles/209819033-Apply-for-Financial-Aid-or-a-Scholarship).

You can review videos, readings, discussion forums, in-video questions, and practice quizzes in the program for free. However, you won’t have access to graded assignments. If you choose to go ahead and earn your certificate, you’ll need to upgrade to the certificate program, unlock the graded assessments, and finish those steps.

Helpful habits for successfully completing the certificate

As a learner, you’re bringing all of your past experiences and best learning practices to this program. The designers of this course have also put together a list of helpful habits that they believe will help you to be the most successful:

1. **Plan your time:** Setting regular study times and sticking with them each week can help you make learning a part of your routine. Use a calendar or timetable to create a schedule. Listing what you plan to do each day will break your work down into achievable goals. And creating a quiet place to watch the videos, review the readings, and complete the activities is important, so you can really focus on the material.
2. **Learn in order:** We recommend taking these courses — and the items in each lesson — in the order they appear, as new information and concepts build on previous ones. By following the order, you’ll be able to get comfortable with ideas, then practice and build on them.
3. **Be curious:** If you find an idea that gets you excited, please act on it! Ask questions, search for more details online, check out the links that interest you, and take notes on your discoveries. The little things you do to support your learning along the way will take your knowledge even further; open more doors in this new, high-growth field; and help you qualify for all kinds of new jobs.
4. **Take notes:** Notes are useful when researching something you’re curious about. This is especially helpful when a task seems important and you think it might be useful later. Or, sometimes you might come across a subject that you want to explore in more detail. Keeping notes can help you keep track of what you learn. Finally, taking notes is an effective way to help make connections between topics and gain a better understanding of them. You can use your notes to build your very own data analytics journal — a place where you can capture ideas, information, and any questions you might have. You’ll probably want to keep your notes together in one place-- whether that’s a physical journal or a document on your computer. This will make it easier to stay organized. Feel free to revisit your journal as you progress through the program, during your job hunt, and even as you settle into your new role as a data analyst.
5. **Chat (responsibly) with other learners**: If you have a question, chances are, you’re not alone. Feel free to reach out in the discussion forum to ask for help from other learners taking this program. You can also visit Coursera’s [Global Online Community](https://coursera.community/). Other important things to know while you’re making friends can be found in the [Coursera Honor Code](https://learner.coursera.help/hc/en-us/articles/209818863-Coursera-Honor-Code) and the [Code of Conduct](https://learner.coursera.help/hc/en-us/articles/208280036-Coursera-Code-of-Conduct).

Deciding if you should take the speed track

**This reading provides an overview of a speed track we offer to those already familiar with data analytics.**

**If you are brand new to data analytics, you can skip the diagnostic quiz after this reading, and move directly to the next activity:** [**Data analytics in everyday life**](https://www.coursera.org/learn/foundations-data/lecture/N5lvQ/data-analytics-in-everyday-life).

The Google Data Analytics Certificate is a program for anyone. A background in data analysis isn’t required. But you might be someone who has some experience already. If you are this type of learner, we have designed a speed trackfor this course. Learners who opt for the speed track can refresh on the basic topics and take each of the weekly challenges and the Course Challenge at a faster pace.

To help you decide if you’re a good match for the speed trackfor this course:

1. Take the optional [diagnostic quiz](https://www.coursera.org/learn/foundations-data/quiz/4BXk3/optional-familiar-with-data-analytics-take-our-diagnostic-quiz).
2. Refer to the [scoring guide](https://www.coursera.org/learn/foundations-data/supplement/V13Ie/your-diagnostic-quiz-score-and-what-it-means) to determine if you’re a good fit for the speed track. A score of 90% or higher is the target goal for someone on the speed track.
3. Based on your individual score, follow the recommendations in the scoring guide for your next steps.

**Important reminder:** If you’re eligible for the speed track, you’re still responsible to complete all graded activities. In order to earn your certificate, you will need an overall score of 80% or higher on all graded materials in the program.

# 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.

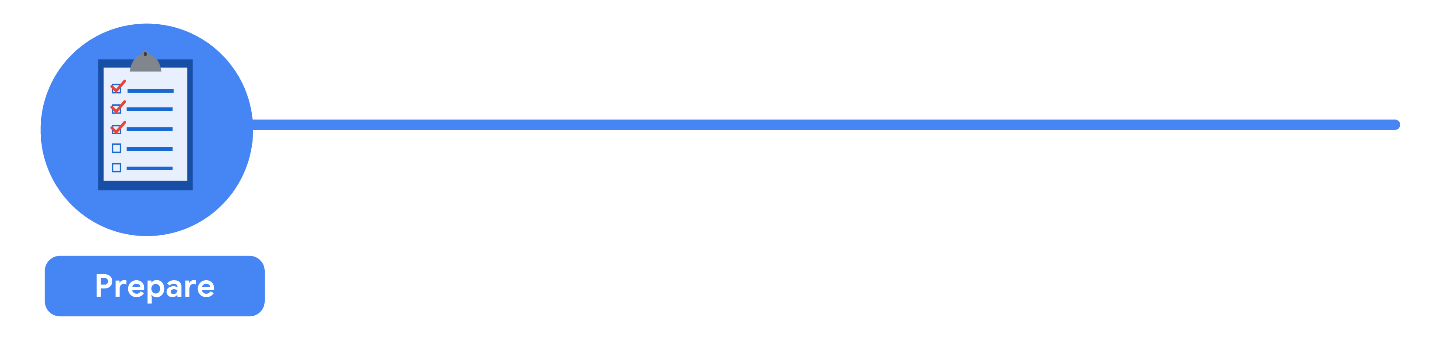
An organization was experiencing a high turnover rate among new hires. Many employees left the company before the end of their first year on the job. The analysts used the data analysis process to answer the following question: **how can the organization improve the retention rate for new employees?**

Here is a break down 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:

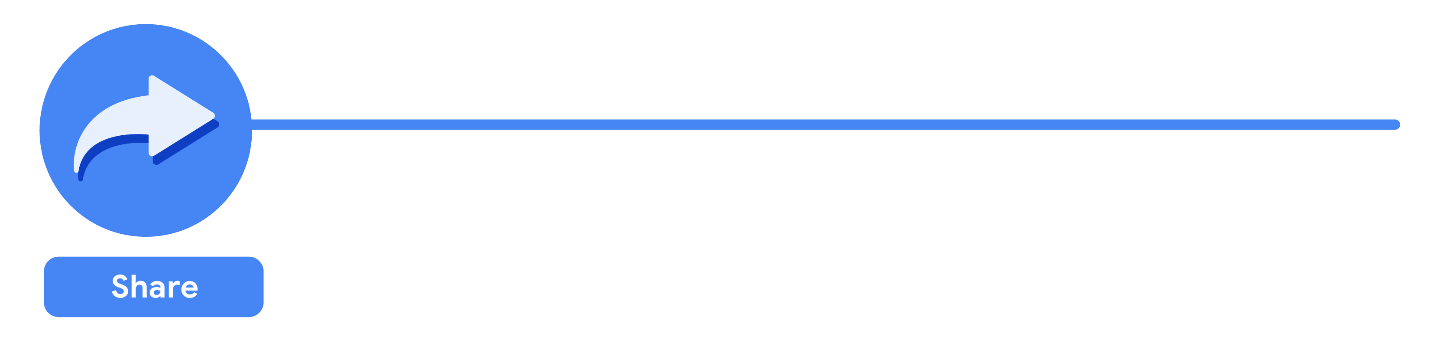
* They restricted access to the data to a limited number of analysts.
* They cleaned the data to make sure it was complete, correct, and relevant. Certain data was aggregated and summarized without revealing individual responses.
* They uploaded raw data to an internal data warehouse for an additional 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 life-saving. 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.

Correction to article in bold below: Microsoft’s Workplace Analytics team hypothesized that moving the 1,200-person group from five buildings to four could improve collaboration by **increasing** the number of employees per building and by reducing the distance that staff needed to travel for meetings.

# Learning Log: Consider how data analysts approach tasks



## Overview



Earlier you learned about how data analysts at one organization used data to improve employee retention. Now, you’ll complete an entry in your learning log to track your thinking and reflections about those data analysts' process and how they approached this problem. By the time you complete this activity, you will have a stronger understanding of how the six phases of the data analysis process can be used to break down tasks and tackle big questions. This will help you apply these steps to future analysis tasks and start tackling big questions yourself.

## Review the six phases of data analysis



Before you write your entry in your learning log, reflect on [the case study](https://www.coursera.org/learn/foundations-data/supplement/nhC19/case-study-new-data-perspectives) from earlier. The data analysts wanted to use data to improve employee retention. In order to do that, they had to break this larger project into manageable tasks. The analysts organized those tasks and activities around the six phases of the data analysis process:

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

The analysts **asked** questions to define both the issue to be solved and what would equal a successful result. Next, they **prepared** by building a timeline and collecting data with employee surveys that were designed to be inclusive. They **processed** the data by cleaning it to make sure it was complete, correct, relevant, and free of errors and outliers. They **analyzed** the clean employee survey data. Then the analysts **shared** their findings and recommendations with team leaders. Afterward, leadership **acted** on the results and focused on improving key areas.



### Access your learning log

To use the template for this course item, click the link below and select “Use Template.”

Link to learning log template: [Consider how data analysts approach tasks](https://docs.google.com/document/d/1J9N9efx3IJZxlxXBProOld0Ha22pnsKP3ZPSJnIn-UA/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the attachment below.

**Learning Log Template\_ Consider how data analysts approach tasks**DOCX File

[Download file](https://d3c33hcgiwev3.cloudfront.net/aOk-I4lyQcGpPiOJcqHBqw_63a2e80233374af58a75ca11ff93c7f5_Learning-Log-Template_-Consider-how-data-analysts-approach-tasks.docx?Expires=1648080000&Signature=bjJ0~AT3dhxrUrqCCFk5jGxmZ051YgEAltSGjfanwqW57Jj~PDFmbvHhplgTkE1~b9axMTwG~Ys7lnKIi1Y4sXICnR0LWvc0kqJOEe50ZjGza-i8D21aCCoccbR~d6H6GHBYoY~am1NG2awSikCpdogYwPcqg-uHb9~hJlurKhE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)



## Reflection



In your learning log template, write 2-3 sentences (40-60 words) reflecting on what you’ve learned from the case study by answering each of the questions below:

* Did the details of the case study help to change the way you think about data analysis? Why or why not?
* Did you find anything surprising about the way the data analysts approached their task?
* What else would you like to learn about data analysis?

When you’ve finished your entry in the learning log template, make sure to save the document so your response is somewhere accessible. This will help you continue applying data analysis to your everyday life. You will also be able to track your progress and growth as a data analyst.

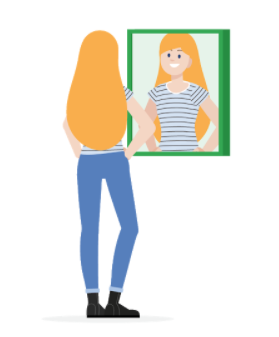
# Data and gut instinct

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.

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

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

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.



## Why gut instinct can be a problem

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.

Consider an example of a real estate developer bidding to redevelop a part of a city's central district. They were well-known for preservation of historical buildings. Banking on their reputation, the agency's planners followed gut instinct and included the preservation of several buildings to gain support and win approval for the project. However, private donations fell short and a partnership failed to materialize and save the day. The buildings eventually had to be torn down after much delay and an expensive dispute with the city.

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 past 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.

## Data + business knowledge = mystery solved

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?”

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

* What kind of results are needed?
* Who will be informed?
* Am I answering the question being asked?
* How quickly does a decision need to be made?

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.



# Origins of the data analysis process

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!



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.

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.

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

1. **Ask**: Business Challenge/Objective/Question
2. **Prepare**: Data generation, collection, storage, and data management
3. **Process**: Data cleaning/data integrity
4. **Analyze**: Data exploration, visualization, and analysis
5. **Share**: Communicating and interpreting results
6. **Act**:  Putting your insights to work to solve the problem

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

## EMC's data analysis life cycle

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

1. Discovery
2. Pre-processing data
3. Model planning
4. Model building
5. Communicate results
6. Operationalize

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of 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.

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:

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.

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:

1. Identifying the problem
2. Designing data requirements
3. Pre-processing data
4. Performing data analysis
5. Visualizing data

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.

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:

1. Business case evaluation
2. Data identification
3. Data acquisition and filtering
4. Data extraction
5. Data validation and cleaning
6. Data aggregation and representation
7. Data analysis
8. Data visualization
9. Utilization of analysis results

This 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.

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.

Learning Log: Explore data from your daily life

Background pattern

Description automatically generated with low confidence

**Overview**



In a previous learning log, you reflected on how you use data analysis in your own life to make everyday decisions. Now, you’ll complete an entry in your learning log exploring data from an area of your life. By the time you complete this activity, you will have a stronger understanding of how you can apply your data analysis skills to more specific activities and situations in your life--starting with your own everyday decisions! Later, you are going to use the data you generate for this entry to practice organizing data to draw insights from it.

**Create a list**



Before you start, pick one area of your everyday life you would like to explore further. Think about how many times in the past few weeks you made decisions about anything related to this area. Then, create a list and include details, such as the date, time, cost, quantity, size, etc. Try to focus on things that can be represented by a number or category.

Here are a few thought-starters:

* Number of cups of coffee you drink daily
* Popular workout times at the gym
* Nightly bedtime

For example, you could create a list exploring your daily coffee intake like this:

**Daily coffee intake**

* Jan. 8th 8 am - bought coffee - one 10 oz. cup
* Jan. 8th 10 am - made coffee at home - one 12 oz. cup
* Jan. 9th 8 am - bought coffee - mug
* Jan 10th 11 am - bought large coffee - 20 oz.
* Jan 11th 8 am - made coffee at home - mug

This example includes a few different details like date and time, whether the coffee was purchased or homemade, and the quantity. You can choose to focus on any area of your life you want and track the details you are interested in exploring. Try to record a week or two of data. Then, you will compile this list in a learning log template, linked below.



**Access your learning log**

To use the template for this course item, click the link below and select “Use Template.”

Link to learning log template: [Explore data from your daily life](https://docs.google.com/document/d/1V7trFOoWavtEm1OV2sEFO-7K-6p9hbnhaczvhb49Hqg/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the attachment below:

**Learning Log Template\_ Explore data from your daily life**DOCX File

[Download file](https://d3c33hcgiwev3.cloudfront.net/W6e-F-TCS_ynvhfkwnv8dw_dd82e83aa8634b449e65431a2a390b5d_Learning-Log-Template_-Explore-data-from-your-daily-life.docx?Expires=1648080000&Signature=Uq3fzrBssaM50a4iw0Rx5hz09wnVqlLBTqaOwjSRbEuc8L8xYvGYL~pRkukIRDJewGEzplreLGciXJBkglkhWQEb1mGXlntyIT2XGELWCPzlItb6Y6mRlbKdkmeUjDbHn66aKlKdEtMg78PrkwwtcJ~cA8wOWtf6xXx13M6cjEc_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)



**Reflection**



After you have finished creating your detailed list exploring data from your own life, take a moment to reflect on that data. In your learning log entry, write 2-3 sentences (40-60 words) in response to each question below:

* Are there any trends you noticed in your behavior?
* Are there factors that influence your decision-making?
* Is there anything you identified that might influence your future behavior?

When you’ve finished your entry in the learning log template, make sure to save the document so your response is somewhere accessible. This will help you continue applying data analysis to your everyday life. You will also be able to track your progress and growth as a data analyst.

Learning Log: Reflect on your skills and expectations

Background pattern

Description automatically generated with low confidence

**Overview**



You have already learned about the five essential aspects of analytical skills: **curiosity, understanding context, having a technical mindset, data design, and data strategy.** You have also discovered that you’re already practicing these skills. Now, you’ll complete an entry in your learning log exploring your own analytical strengths and weaknesses and your goals for the future. By the time you complete this activity, you will have a stronger understanding of your analytical skill set and how you can practice and improve them. These analytical skills are key to helping you solve problems and create insights using data analysis. Thinking about them now will help you grow as a data analyst!

**The analytical skills table**



First, you’ll fill out an Analytical Skills Table in your learning log entry. The table will appear like this in the template:

Analytical skill column: -Curiosity -Context -Technical mindset -Data design -Data strategy

The table has a row for each essential aspect of analytical skills:

* **Curiosity:** a desire to know more about something, asking the right questions
* **Understanding context:** understanding where information fits into the “big picture”
* **Having a technical mindset:** breaking big things into smaller steps
* **Data design:** thinking about how to organize data and information
* **Data strategy:** thinking about the people, processes, and tools used in data analysis

You will put an X in the column that you think best describes your current level with each aspect. The three ratings are:

* **Strength:** This is an area you feel is one of your strengths
* **Developing:** You have some experience with this area, but there’s still significant room for growth
* **Emerging:** This is new to you, and will gain experience in this area from this course

Then update the Comments/Plans/Goals column with a quick note to yourself about why you chose those ratings.



**Access your learning log**

To use the template for this course item, click the link below and select “Use Template.”

Link to learning log template: [Reflect on your skills and expectations](https://docs.google.com/document/d/1Wf8-Cf8bEsS_I2ckXGZCoGLSBc3F38KiUn7vCI7M1ZI/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the attachment below.

**Learning Log Template\_ Reflect on your skills and expectations**DOCX File

[Download file](https://d3c33hcgiwev3.cloudfront.net/2XJxxwH8T2OycccB_H9jWg_5d0ab96fce0b4f1f81a322aebf1bd9f1_Learning-Log-Template_-Reflect-on-your-skills-and-expectations.docx?Expires=1648080000&Signature=HpYn~bLtbWYus6vXxGmgfi-gDDegbrwSBtTF8zbImbXw97NZReVJrj7FWlaCaDalAp3BFdRuLkgGP2vTDHuFGeAwlrg~RuN4CM6q8VK4uH3dxNCdkk7Fi-9Z3JVv6sUHXZAlW4E3ZPDLMBQKyZiFgtwajvkzRAk4J9ErZ7Vh9Wk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)



**Reflection**



After you have completed the Analytical Skills Table, take a moment to reflect on your evaluations. In your learning log entry, write 2-3 sentences (40-60 words) in response to each question below:

* What do you notice about the ratings you gave yourself in each area? How did you rate yourself in the areas that appeal to you most?
* If you are asked to rate your experience level in these areas again in a week, what do you think the ratings will be, and why do you think that?
* How do you plan on developing these skills from now on?

When you’ve finished your entry in the learning log template, make sure to save the document so your response is somewhere accessible. This will help you continue applying data analysis to your everyday life. You will also be able to track your progress and growth as a data analyst.

# Variations of the data life cycle

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

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

**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.

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](https://www.coursera.org/learn/foundations-data/supplement/WWlrt/origins-of-the-data-analysis-process). 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.

## 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.

Learning Log: Organize your data in a table

Background pattern

Description automatically generated with low confidence

**Overview**



By now, you have started to think about data in your daily life and how you use this data to make decisions. Earlier in this course, you completed a [learning log](https://www.coursera.org/learn/foundations-data/supplement/RNx3Z/learning-log-explore-data-from-your-daily-life) where you recorded some data from your daily life. Next, you will consider how to organize this data. In this activity, you’ll write an entry in your learning log to track your thinking and reflections about how to organize data. By the time you complete your entry, you will understand how to create and format a table to store the data that you collect. Tables are one of the most common ways data is organized for analysis. This foundational skill will help you more easily analyze data, and will serve as a go-to tool in your data analyst’s toolkit.

**Structuring your data**



To get started, consider the data you have collected in your learning log entries so far in this course. Now, take a moment and prepare to organize this data. One of the simplest ways to add structure to your data is to put it in a table.

To record your data in a table, you need to understand how a table is structured:

* A table consists of rows and columns
* Each row is a different observation
* Each column is a different attribute of that observation

For example, here is a collection of observations in a learning log about how many cups of coffee are consumed each day:

1. 10/19, 2.5 cups of coffee
2. 10/20, 2 cups of coffee
3. 10/21, 1 cup of coffee
4. 10/22, 1.5 cups of coffee
5. 10/23, 1.5 cups of coffee

There are five data points. Each piece of data consists of a date and the number of cups of coffee consumed that day. You can structure this as a table with six rows and two columns. This includes five rows of data and one header row with titles:

| **Date** | **Cups of Coffee / Day** |
| --- | --- |
| 10/19 | 2.5 |
| 10/20 | 2 |
| 10/21 | 1 |
| 10/22 | 1.5 |
| 10/23 | 1.5 |

You can also create a table with more detailed data. For instance, if your data also contained information about whether there was cream and sugar in the coffee, it might appear like this:

1. 10/19, 2.5 cups, cream, sugar
2. 10/20, 2 cups, no cream, no sugar
3. 10/21, 1 cup, cream, sugar
4. 10/22, 1.5 cups, cream, no sugar
5. 10/23 1.5 cups, cream, sugar

You can represent this by adding two more columns to your table, one titled “Cream” and one titled “Sugar.”

| **Date** | **Cups Coffee/Day** | **Cream** | **Sugar** |
| --- | --- | --- | --- |
| 10/19 | 2.5 | yes | yes |
| 10/20 | 2 | no | no |
| 10/21 | 1 | yes | yes |
| 10/22 | 1.5 | yes | no |
| 10/23 | 1.5 | yes | yes |

**Now it’s your turn!**



You have been collecting data from the beginning of the course. Take a moment to consider the data you have gathered in your learning log. Now, determine how you could organize your data in a table.

Before you begin, you should decide what software you’d like to use to create your table. We suggest using Google Docs or Microsoft Word for this example; you will have a chance to use tables in spreadsheets later on. You will find detailed instructions on how to create tables when you access your learning log, below.



**Access your learning log**

To use the template for this course item, click the link below and select “Use Template.”

Link to learning log template: [Organize your data in a table](https://docs.google.com/document/d/1q7M9ZhQMaQGVMTW7iZxzLylE1Ap5DjGQHrZ1-aUcOYk/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the attachment below.

**Learning Log Template\_ Organize your data in a table**DOCX File

[Download file](https://d3c33hcgiwev3.cloudfront.net/J03wxGq-QCGN8MRqvgAhkQ_45b30b42a4594805807733582706a4da_Learning-Log-Template_-Organize-your-data-in-a-table.docx?Expires=1648080000&Signature=ayVPFTmP81DVN95QcvIc~yIAf-urAG5lZ5phSe6VKO3qaHsOGfJmggMlhqgFFFfzd9UC6lisT1XG7xPZaL9DtqKbGgddJQFsq7YKQRA3ZNke6U0bYQimGB8qZePULrlx9ziUZiOMQYovZJ7AHmmUABDvdZ8i3utgoDo2WRTgCdE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)



**Reflection**



In a new learning log entry, follow the instructions in the template, and add a table to organize your data. Then, write 3-5 sentences (60-100 words) on opportunities in your personal life or current job to organize data into tables.

When you’ve finished your entry in the learning log template, make sure to save the document so your response is somewhere accessible. This will help you continue applying data analysis to your everyday life. You will also be able to track your progress and growth as a data analyst.

Key data analyst tools

As you are learning, the most common programs and solutions used by data analysts include spreadsheets, query languages, and visualization tools. In this reading, you will learn more about each one. You will cover when to use them, and why they are so important in data analytics.



**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.

Query languages

* 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.

**Key takeaway**

You have a lot of tools as a data analyst. This is a first glance at the possibilities, and you will explore many of these tools in-depth throughout this program.

# 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.

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:

| **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. And, later in this course, you will learn about programming languages like R that give you even greater control of your data, its analysis, and the visualizations you create.

As you continue learning about these important tools, you will gain the knowledge to choose the right tool for any data job.

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:

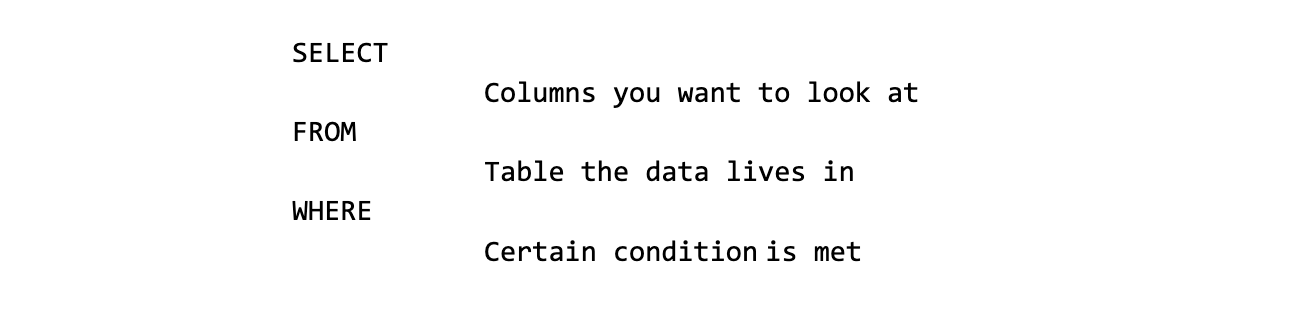
* Use **SELECT** to choose the columns you want to return.
* Use **FROM** to choose the tables where the columns you want are located.
* Use **WHERE** to filter for certain information.

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:

Background pattern

Description automatically generated with low confidence

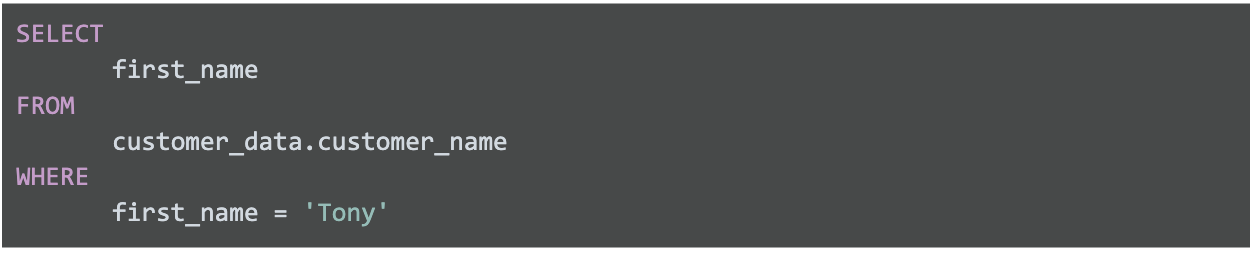
Next, enter the table name after the FROM; the table columns you want after the SELECT; and, finally, the conditions you want to place on your query after the WHERE. Make sure to add a new line and indent when adding these, as shown below:



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:

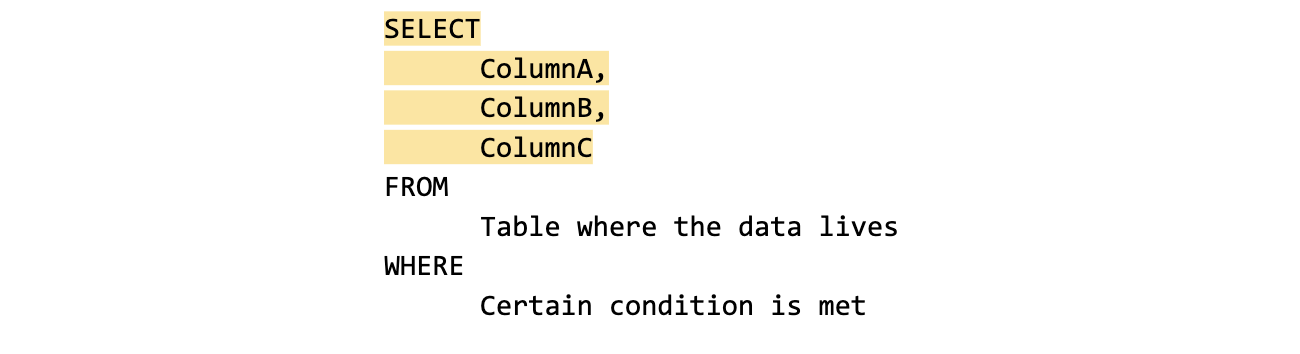
| **first\_name** |
| --- |
| Tony |
| Tony |
| Tony |

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

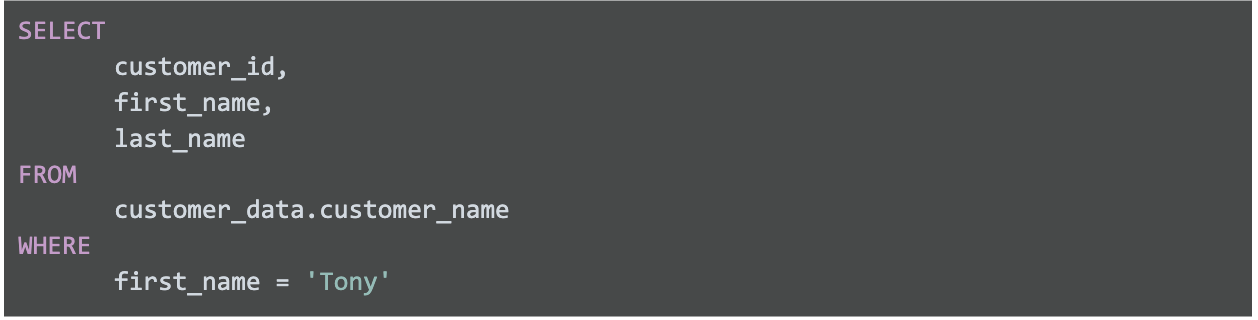
**Multiple columns in a query**

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:



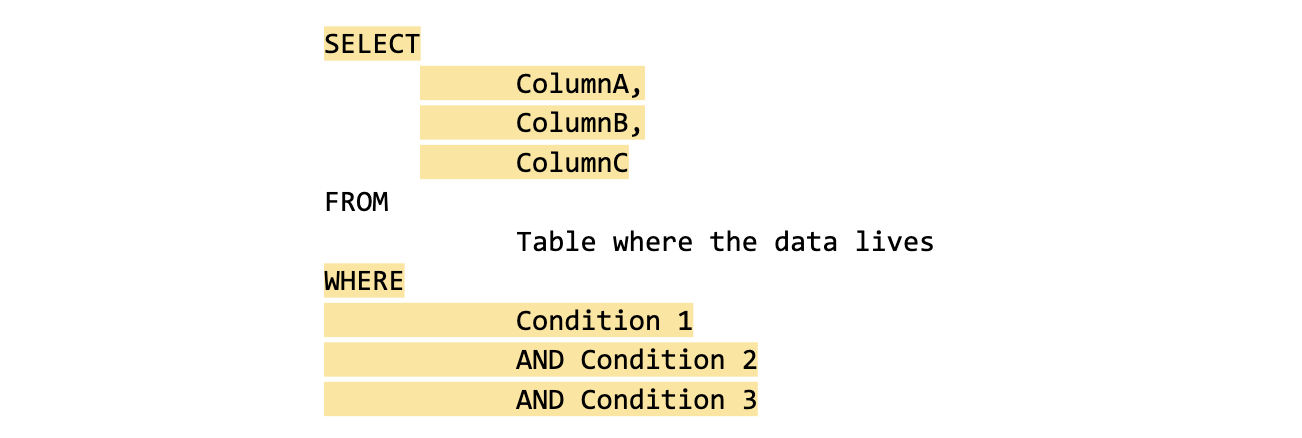
Here is an example of how it would appear in BigQuery:



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

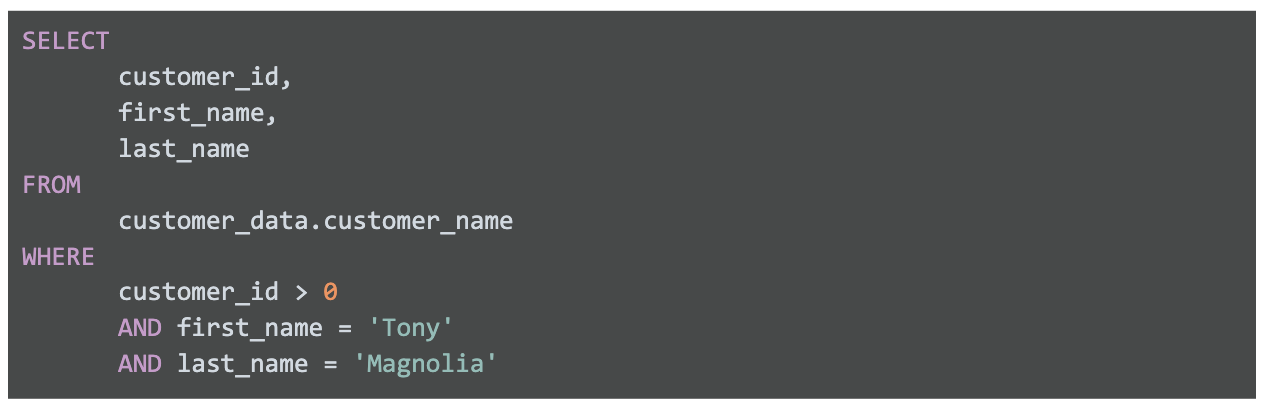
1. **SELECT** the columns named **customer\_id**, **first\_name**, and **last\_name**
2. **FROM** a table named **customer\_name** (in a dataset named **customer\_data**) (The dataset name is always followed by a dot, and then the table name.)
3. But only return the data **WHERE** the first\_name is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



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.

**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.

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

**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:

**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 (%).

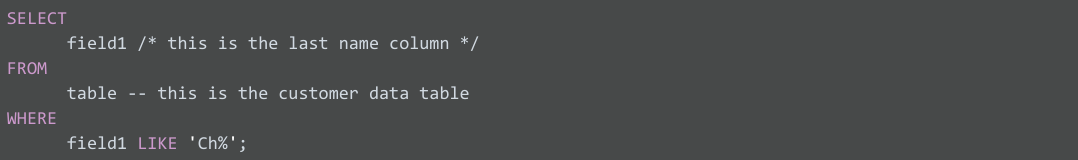
**SELECT all columns**

Can you use  **SELECT \*** ?

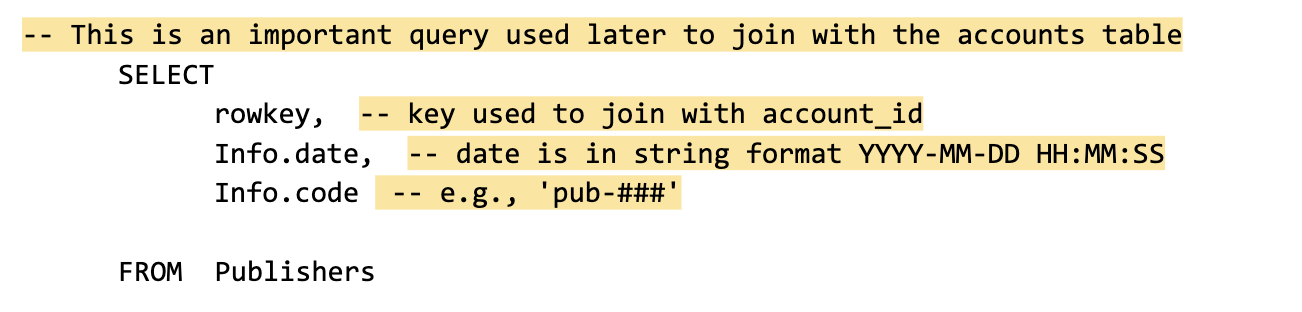
In the example, if you replace **SELECT field1** with **SELECT \*** , you would be selecting all of the columns in the table instead of the field1 column only. From a syntax point of view, it is a correct SQL statement, but you should use the asterisk (\*) sparingly and with caution. Depending on how 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.

**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.



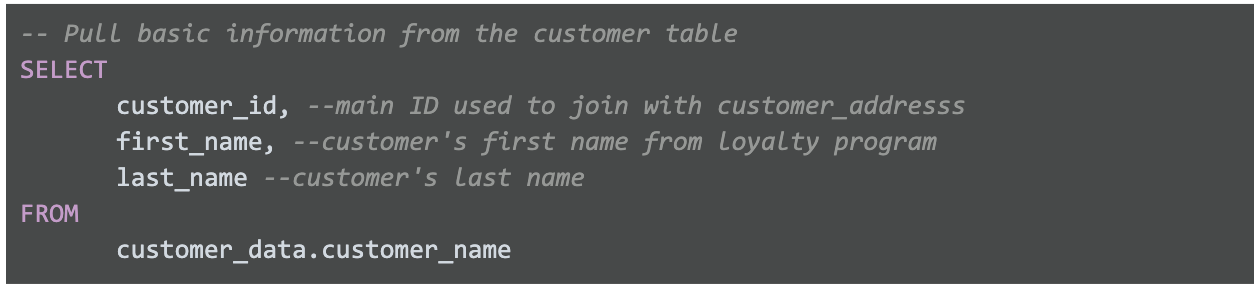
Comments can also be added outside of a statement as well as within a statement. You can use this flexibility to provide an overall description of what you are going to do, step-by-step notes about how you achieve it, and why you set different parameters/conditions.



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.

**Example of a query with comments**

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



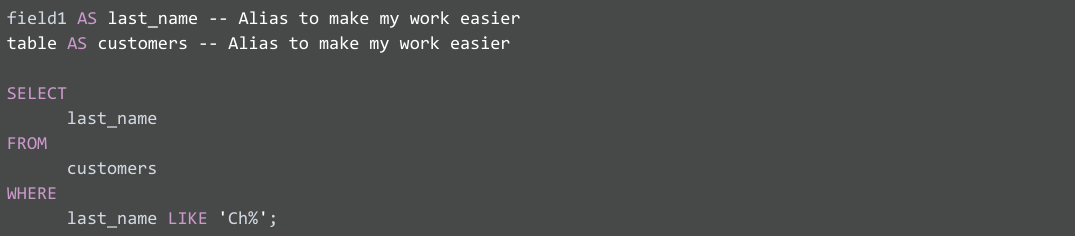
In the above example, a comment has been added before the SQL statement to explain what the query does. Additionally, a comment has been added next to each of the column names to describe the column and its use. Two dashes (--) are generally supported. So it is best to use -- and be consistent with it. 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.

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.

**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.

**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.

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.



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:



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

| **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":

Shape

Description automatically generated

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

| **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.

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

**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.

Learning Log: Reflect on the data analysis process

Background pattern

Description automatically generated with low confidence

**Overview**



By now, you have started getting familiar with the data analysis process. Now, you’ll complete an entry in your learning log reflecting on your experience with the data analysis process and your progress in this course. By the time you complete this activity, you will have a stronger understanding of how to use the steps of this process to organize data analysis tasks and solve big problems with data. This framework will continue to help guide you through your own work in this course--and as a junior data analyst!

**The data analysis process so far**



Take a moment to appreciate all the work you have done in this course. You identified a question to answer, and systematically worked your way through the data analysis process to answer that question—just like professional data analysts do every day!

In reviewing the data analysis process so far, you have already performed a lot of these steps. Here are some examples to think about before you begin writing your learning log entry:

* You **asked** an interesting question and defined a problem to solve through data analysis to answer that question.
* You thought deeply about what data you would need and how you would collect it in order to **prepare** for analysis.
* You **processed** your data by organizing and structuring it in a table and then moving it to a spreadsheet.
* You **analyzed** your data by inspecting and scanning it for patterns.
* You **shared** your first data visualization: a bar chart.
* Finally, after completing all the other steps, you **acted**: You reflected on your results, made decisions, and gained insight into your problem--even if that insight was that you didn't have enough data, or that there were no obvious patterns in your data.

As you progress through the rest of the program, you will continue using and revisiting these steps to help guide you through your own analysis tasks. You will also learn more about different tools that can help you along the way!



**Access your learning log**

To use the template for this course item, click the link below and select Use Template.

Link to learning log template: [Reflect on the data analysis process](https://docs.google.com/document/d/1MCgIOQk2eW77ZBpJ6ygRcrkdJ1aLuQd3IrkqsMpys5c/template/preview)

OR

If you don’t have a Google account, you can download the template directly from the attachment below.

**Learning Log Template\_ Reflect on the data analysis process**DOCX File

[Download file](https://d3c33hcgiwev3.cloudfront.net/bRFBueRvSTeRQbnkb1k3Qw_ac819fa396dd47e89701e527dfdd97e7_Learning-Log-Template_-Reflect-on-the-data-analysis-process.docx?Expires=1648080000&Signature=No~2NMgVJByxd8ghR3RDM1aHyO-e1rOJVEhv6vblFmB2Q5Olc8OMfelYHZZdNHEWDVyvLFOtv7QknGjOBDawdwOyoDn3HADaw6Pu3aQzoacH2Mh9NDCfYOB~tFZjB-uudIeA9-xGZdlpdkrxHayekeuM6aRzQOWkkANkwi1r0QE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)



**Reflection**



In your learning log, write 2-3 sentences (40-60 words) reflecting on the data analysis process and your experiences so far by answering each of the questions below:

* Which part(s) of the data analysis process did you enjoy the most? What did you enjoy about it?
* What were some of the key ideas you learned in this course?
* Are there concepts or portions of the content that you would like to learn more about? If so, what are they? Which upcoming course do you think would teach you the most about this area?
* Now that you’ve gained experience doing data analysis, how do you feel about becoming a data analyst? Have your feelings changed since you began this course? If so, how?

When you’ve finished your entry in the learning log template, make sure to save the document so your response is somewhere accessible. This will help you continue applying data analysis to your everyday life. You will also be able to track your progress and growth as a data analyst.

Data analyst roles and job descriptions

As technology continues to advance, being able to collect and analyze the data from that new technology has become a huge competitive advantage for a lot of businesses. Everything from websites to social media feeds are filled with fascinating data that, when analyzed and used correctly, can help inform business decisions. A company’s ability to thrive now often depends on how well it can leverage data, apply analytics, and implement new technologies.

This is why skilled data analysts are some of the most sought-after professionals in the world. A study conducted by IBM estimates that companies in the United States will fill 2,720,000 Data Science and Analytics jobs by 2020\*. Because the demand is so strong, you’ll be able to find job opportunities in virtually any industry. Do a quick search on any major job site and you’ll notice that every type of business from zoos, to health clinics, to banks are seeking talented data professionals. Even if the job title doesn’t use the exact term “data analyst,” the job description for most roles involving data analysis will likely include a lot of the skills and qualifications you’ll gain by the end of this program. In this reading, we’ll explore some of the data analyst-related roles you might find in different companies and industries.

\**“The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market,” by Will Markow, Soumya Braganza, and Bledi Taska, with Steven M. Miller and Debbie Hughes.* [*https://www.ibm.com/downloads/cas/3RL3VXGA*](https://www.ibm.com/downloads/cas/3RL3VXGA)

**Decoding the job description**

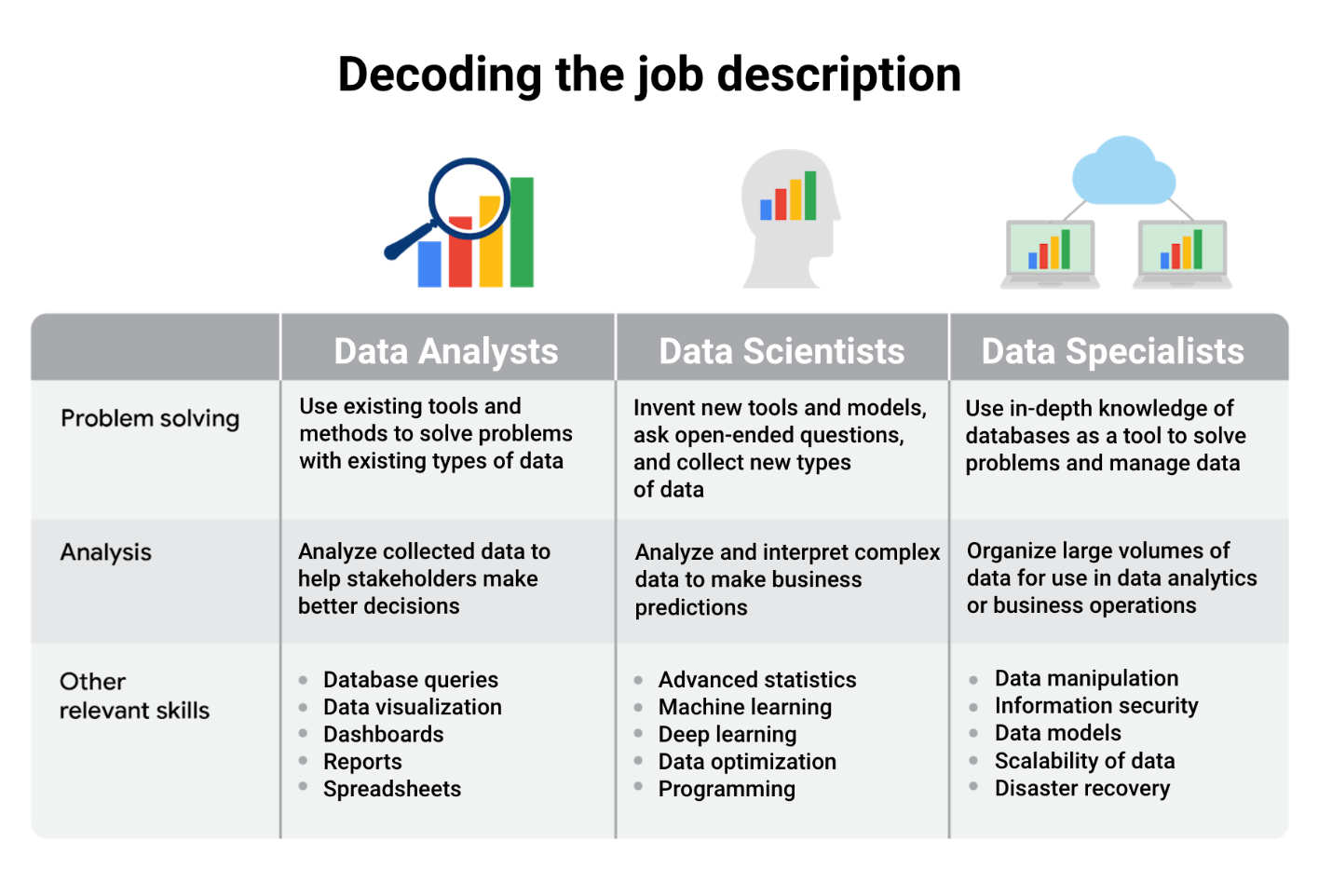
The data analyst role is one of many job titles that contain the word “analyst.”

To name a few others that sound similar but may not be the same role:

* Business analyst — analyzes data to help businesses improve processes, products, or services
* Data analytics consultant — analyzes the systems and models for using data
* Data engineer — prepares and integrates data from different sources for analytical use
* Data scientist — uses expert skills in technology and social science to find trends through data analysis
* Data specialist — organizes or converts data for use in databases or software systems
* Operations analyst — analyzes data to assess the performance of business operations and workflows

Data analysts, data scientists, and data specialists sound very similar but focus on different tasks. As you start to browse job listings online, you might notice that companies’ job descriptions seem to combine these roles or look for candidates who may have overlapping skills. The fact that companies often blur the lines between them means that you should take special care when reading the job descriptions and the skills required.

The table below illustrates some of the overlap and distinctions between them:

Title: Decoding the job description data analysts: -problem solving: Use existing tools and methods to solve problems with existing types of data -analysis: Analyze collected data to help stakeholders make better decisions -other relevant skills: database queries, data visualization, dashboards, reports and spreadsheets data scientists: -problem solving: Invent new tools and models, ask open-ended questions, and collect new types of data -analysis: Analyze and interpret complex data to make business predictions -other relevant skills: advanced statistics, machine learning, deep learning, data optimization, and programming data specialists: -problem solving: Use in-depth knowledge of databases as a tool to solve problems and manage data -analysis: Organize large volumes of data for use in data analytics or business operations -other relevant skills: data manipulation, information security, data models, scalability of data, and disaster recovery

We used the role of data specialist as one example of many specializations within data analytics, but you don’t have to become a data specialist! Specializations can take a number of different turns. For example, you could specialize in developing data visualizations and likewise go very deep into that area.

**Job specializations by industry**

We learned that the data specialist role concentrates on in-depth knowledge of databases. In similar fashion, other specialist roles for data analysts can focus on in-depth knowledge of specific industries. For example, in a job as a business analyst you might wear some different hats than in a more general position as a data analyst. As a business analyst, you would likely collaborate with managers, share your data findings, and maybe explain how a small change in the company’s project management system could save the company 3% each quarter. Although you would still be working with data all the time, you would focus on using the data to improve business operations, efficiencies, or the bottom line.

Other industry-specific specialist positions that you might come across in your data analyst job search include:

* Marketing analyst — analyzes market conditions to assess the potential sales of products and services
* HR/payroll analyst — analyzes payroll data for inefficiencies and errors
* Financial analyst — analyzes financial status by collecting, monitoring, and reviewing data
* Risk analyst — analyzes financial documents, economic conditions, and client data to help companies determine the level of risk involved in making a particular business decision
* Healthcare analyst — analyzes medical data to improve the business aspect of hospitals and medical facilities

**Key takeaway**

Explore data analyst job descriptions and industry-specific analyst roles. You will start to get a better sense of the different data analyst jobs out there and which types of roles you’re most interested to go after.