Data Analyst Google Certification 2024

Course 1.



**MODULE 1.**

**Data Analysis:**

The collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision making.

Also is the science of data.

**Data ecosystems:**

The various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data.

**Cloud:**

A place to keep data online, rather than a computer hard drive.

**Data science:**

Creating new ways of modeling and understanding the unknown by using raw data.

**Data-driven decision-making:**

Using facts to guide business strategy.

**Test your knowledge on the data ecosystem**

Question 1

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

Answer: Data analysis

Question 2

Fill in the blank: A \_\_\_\_\_ encompasses the various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data.

Answer: Data ecosystem

Question 3

When building data into business strategy, why is it important to include insights from subject-matter experts? Select all that apply.

Answer(s): (Not Final)

-They can review analysis results and help identify inconsistencies

-Their experience and human intuition are valuable to data-driven decision-making.

Question 4

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

Answer: Data-driven decision-making

**Analytical skills:**

Qualities and characteristics associated with solving problems using facts.

1. Curiosity

2. Understanding context

3. Having a technical mindset

4. Data design

5. Data strategy

**Context:**

The condition in which something exists or happens.

**A technical mindset:**

The ability to break things down into smaller steps or pieces and work with them in an orderly and logical way.

**Data design:**

How you organize information.

**Data strategy:**

The management of the people, processes and tools used in data analysis.

Thinking about the way you organize your contacts in a new phone is an example of which analytical skill?

Answer: Data Design

**Test your knowledge on data analyst skills**

Question 1.

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

Answer: facts

Question 2.

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

Answer: Understanding context.

Question 3.

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

Answer: Technical mindset.

Question 4.

Fill in the blank: The analytical skill \_\_\_\_\_ encompasses how someone organizes information.

Answer: data design

**Analytical thinking:**

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

1. Visualization

2. Strategy

3. Problem-orientation

4. Correlation

5. Big-picture and detail-oriented thinking

**Visualization:**

The graphical representation of information

Correlation does not equal causation

To execute a plan using detail-oriented thinking, what does a data analyst consider?

Answer: The specifics

What is the root cause of the problem?

**Root cause:**

The reason why a problem occurs.

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

Where are the gaps in our process?

**Gap analysis:**

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

What did we not consider before?

1. Curiosity

2. Understanding context

3. Having a technical mindset

4. Data design

5. Data strategy

**Data-driven decision making:**

Using facts to guide business strategy.

A quartile divides data points into four equal parts.

Nonprofits are organizations dedicated to advancing a social cause or advocating for a particular effort.

* Which of your skills and ways of thinking will help you become a successful data analyst?
* Why will these skills help you become a successful data analyst?

My answer(s):

-Technical and soft skills.

-These skills can enable you to develop your critical-thinking skills to create decision decisions to drive the overall organization moving forward based on your data findings

**Test your knowledge on analytical thinking and outcomes**

Question 1

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

Answer: Identifying and defining

Question 2

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

Answer(s):

-Graph

-Chart

-Map

Question 3

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

Answer: Detail-oriented thinking

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

Answer: Gap analysis

**Glossary terms from module 1**

**Terms and definitions for Course 1, Module 1**

**Analytical skills:** Qualities and characteristics associated with using facts to solve problems

**Analytical thinking:** The process of identifying and defining a problem, then solving it by using data in an organized, step-by-step manner

**Context:** The condition in which something exists or happens

**Data:** A collection of facts

**Data analysis:** The collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making

**Data analyst:** Someone who collects, transforms, and organizes data in order to draw conclusions, make predictions, and drive informed decision-making

**Data analytics:** The science of data

**Data design:** How information is organized

**Data-driven decision-making:** Using facts to guide business strategy

**Data ecosystem:** The various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data

**Data science:** A field of study that uses raw data to createnew ways of modeling and understanding the unknown

**Data strategy:** The management of the people, processes, and tools used in data analysis

**Data visualization:** The graphical representation of data

**Dataset:** A collection of data that can be manipulated or analyzed as one unit

**Gap analysis:** A method for examining and evaluating the current state of a process in order to identify opportunities for improvement in the future

**Root cause:** The reason why a problem occurs

**Technical mindset:** The ability to break things down into smaller steps or pieces and work with them in an orderly and logical way

**Visualization:** (Refer to data visualization)

**Module 1 challenge**

QUESTIONS MAY VARY:

Question 1.

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

- Data is a collection of facts.

- One goal of data analysis is to make predictions.

- Collecting data is part of the data analysis process.

Question 2.

Fill in the blank: Data science involves using \_\_\_\_\_ data to create new ways of modeling and understanding the unknown.

raw

Question 3.

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

- Request observations from subject-matter experts

- Find relevant data

- Define the business need

Question 4

A pet shelter wants to receive 15% more donations at this year’s holiday fundraiser than were given in previous years. Data team members look into past donations compared to desired donations for this holiday season. They then develop strategies to bridge the difference between the two. What does this scenario describe?

Gap analysis

Question 5

Fill in the blank: Data analysts use a problem-oriented approach in order to identify, \_\_\_\_\_, and solve problems.

describe

Question 6

Which of the following is NOT a key aspect of analytical thinking?

Intuition

Question 7 (NOT FINAL)

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

Answer: Strategic thinking

Question 8

A data professional asks, “Why?” five times. What are they trying to figure out?

The root cause of a problem

Question 9

A data professional is always interested in learning new skills and gaining knowledge. They often seek out challenging assignments at work and professional development experiences. Which analytical skill does this scenario describe?

Curiosity

Question 10.

- A financial advisor interviews clients to assess their risk tolerance and investment goals in order to create personalized plans.

- A dentist’s office uses specifics from patient records to track treatment outcomes and improve quality of care.

- A school district uses report cards to track performance and identify areas where students need additional support.

**MODULE 2.**

**Planning**

**Capture**

**Manage**

**Archive**

**Destroy**

**Database:**

A collection of data stored in a computer system.

In the data life cycle, which phase involves using data to solve problems, make good decisions, and support business goals?

Answer: Analyze

**Variations of the data life cycle**

You have learned that there are six stages to the data life cycle. Here's 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.

**Note:** Be careful not to confuse the six stages of the data life cycle (plan, capture, manage, analyze, archive, and destroy) with the six phases of the data analysis process (ask, prepare, process, analyze, share, and act). They are not interchangeable.

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/program/data-management/data-management-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 takeaways**

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.

**Test your knowledge on the data life cycle**

Question 1

There are six phases in the data life cycle, which culminate with analyze, archive, and destroy. What is the correct order of the first three phases?

Answer: Plan, capture, manage

Question 2

Fill in the blank: During the \_\_\_\_\_ phase, data teams decide what kind of data is needed and who will be responsible for it.

Answer: Planning

Question 3

A data team considers how and where data will be stored, what tools are required to safeguard it, and how to maintain it properly. What phase of the data life cycle does this scenario describe?

Answer: Manage

Question 4

What is the term for the computer system in which organizations store their data?

Answer: Database

**Recall:**

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”.

Prepare: “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.”

Process: “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.”

Analyze:

“This is the point where we have to take a step back and let the data speak for itself.”

Share:

“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 action on what we’ve just learned.”

Act:

“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.”

Our job as a data analyst; 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.

**Stakeholders:**

People who have invested time and resources into a project and are interested in the outcome.

Look at the current state and identify how it’s different from the ideal state.

Key note: Determine who are the stakeholders are.

**More on the phases of data analysis and this program**

Each step in the data analysis process—ask, prepare, process, analyze, share, and act—plays a crucial role in extracting meaningful insights from data. As you navigate through each phase, from asking the right questions to taking informed actions, you harness the true power of data. In this reading, you’ll explore how the data analysis process guides this program.

**The ask phase**

At the start of any successful data analysis, the data analyst:

* Takes the time to fully understand stakeholder expectations
* Defines the problem to be solved
* Decides which questions to answer in order to solve the problem

Qualifying stakeholder expectations means determining who the stakeholders are, what they want, when they want it, why they want it, and how best to communicate with them. Defining the problem means looking at the current state and identifying the ways in which it’s different from the ideal state. With expectations qualified and the problem defined, you can derive questions that will help achieve these goals.

In an upcoming course, 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.

**The prepare phase**

In the prepare phase, the emphasis is on identifying and locating data you can use to answer your questions. In an upcoming course, 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 data and results are objective and unbiased. In other words, any decisions made from an analysis should always be based on facts and be fair and impartial.

**The process phase**

In this phase, the aim is to refine the data. Data analysts find and eliminate any errors and inaccuracies that can get in the way of results. This usually means:

* Cleaning data
* Transforming data into a more useful format
* Combining two or more datasets to make information more complete
* Removing outliers (data points that could skew the information)

After data analysts process data, they check the data they prepared to make sure it's complete and correct. This phase is all about getting the details right. Accordingly, the data analyst will refine strategies for verifying and sharing their data cleaning with stakeholders. In an upcoming course, you’ll use spreadsheets and structured query language, or SQL, to clean data.

**The analyze phase**

With a solid foundation of well-defined questions and clean data, you’ll delve into the analyze phase. This is when you turn the data you’ve gathered, prepared, and processed into actionable information. Data analysts use many powerful tools in their work. In one upcoming course you'll continue using two of them: spreadsheets and SQL. In another upcoming course you’ll explore using the programming language R to work with and analyze data.

**The share phase**

This phase is exactly what it sounds like: It’s time to share what you’ve learned with your stakeholders! In this part of the program, 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, an upcoming course will highlight why visualization is essential to getting others to understand what your data is telling you. In another upcoming course, you’ll learn how to visualize data with R.

**The act phase**

The data analysis journey culminates in the act phase, when data insights are put to work. For you, this action involves preparing for your job search and having 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 other candidates!

**Learn about the process through the program:**

1. Learn more about the **ask** phase of the process in the [Ask Questions to Make Data-Driven Decisions](https://www.coursera.org/learn/ask-questions-make-decisions/home/welcome) course.
2. Learn more about the **prepare** phase of the process in the [Prepare Data for Exploration](https://www.coursera.org/learn/data-preparation/home/welcome) course.
3. Learn more about the **process** phase of the process in the [Process Data from Dirty to Clean](https://www.coursera.org/learn/process-data/home/welcome) course.
4. Learn more about the **analyze** phase of the process in the [Analyze Data to Answer Questions](https://www.coursera.org/learn/analyze-data/home/welcome) and [Data Analysis with R Programming](https://www.coursera.org/learn/data-analysis-r/home/welcome) courses.
5. Learn more about the **share** phase of the process in the [Share Data Through the Art of Visualization](https://www.coursera.org/learn/visualize-data/home/welcome) and [Data Analysis with R Programming](https://www.coursera.org/learn/data-analysis-r/home/welcome) courses.
6. Learn more about the **act** phase of the process in the [Google Data Analytics Capstone: Complete a Case Study](https://www.coursera.org/learn/google-data-analytics-capstone/home/welcome) course.

**Note:** The course links are for you to preview. Please do not complete the courses at this time. You may mark this activity as complete after you understand how the courses align to the data analysis process.

**Test your knowledge on the data analysis process**

Question 1

A junior data analyst is assigned a new project. They define the problem that they need to help solve and confirm they understand expectations. Which step of the data analysis process are they working in?

Ask

Question 2

What is the term for people who invest time and resources into a project and are interested in the outcome?

Stakeholders

Question 3

Fill in the blank: In the process step, data analysts may remove \_\_\_\_\_, which are data points that lie significantly outside the overall pattern of the data and could potentially skew the information.

outliers

Question 4

Which step of the data analysis process is most likely to involve creating visuals that help people understand complex facts and figures?

Share

-Spreadsheets

-Query languages for databases

-Visualization tools

-Microsoft Excel

-Google Sheets

**Formula**

A set of instructions that performs a specific calculation using the data in a spreadsheet.

**Function**

A preset command that automatically performs a specific process or task using the data in a spreadsheet.

**Query language**

A computer programming language that allows you to retrieve and manipulate data from a database.

**S**tructured

**Q**uery

**L**anguage

**Data visualization**

The graphical representation of information

Prepare

Process

Analyse

Shared

Some better visualization tools

-Tableau, Looker

**Self-Reflection: Review past concepts**

Question 1

**Reflection**



Consider what you reviewed about the phases of data analysis and the stages of the data life cycle:

* What is the relationship between the data life cycle and the data analysis process? How are the two processes similar? How are they different?
* What is the relationship between the Ask phase of the data analysis process and the Plan phase of the data life cycle? How are they similar? How are they different?

Reflect on your learning and think about how you can apply the phases of data to future projects.

Now, write 2-3 sentences (40-60 words) in response to each of these questions. Enter your response in the text box below.

My Answer: According to LinkedIn: the data life cycle manages data from creation to end, while the data analysis process works from asking questions to getting useful insights The Ask phase in data analysis and the Plan phase in the data life cycle both kickstart a structured approach to handling data. However, they differ slightly: the Ask phase questions data for analysis, while the Plan phase maps out how to manage and use data.

**Test your knowledge on the data analysis toolbox**

**.**

Question 1

What is the term for a digital worksheet that data analysts use to store, organize, and sort data?

Answer: Spreadsheet

Question 2

What is a key difference between a formula and a function?

Answer: A formula is a set of instructions, whereas a function is a preset command.

Question 3

What does SQL stand for?

Answer: Structured query language

Question 4

Which tool in the data analysts toolbox is most likely to involve the creation of graphs, maps, or charts?

Answer: Data visualization

**Glossary terms from module 2**

**Terms and definitions for Course 1, Module 2**

**Database:** A collection of data stored in a computer system

**Formula:** A set of instructions used to perform a calculation using the data in a spreadsheet

**Function:** A preset command that automatically performs a specified process or task using the data in a spreadsheet

**Query:** A request for data or information from a database

**Query language:** A computer programming language used to communicate with a database

**Stakeholders:** People who invest time and resources into a project and are interested in its outcome

**Structured Query Language:** A computer programming language used to communicate with a database

**Spreadsheet:** A digital worksheet

**SQL:** (Refer to Structured Query Language)

Module 2 challenge

Question 1

Which of the following activities are part of the manage phase of the data life cycle? Select all that apply.

-Determine which tools will be most effective at safeguarding data

-Choose where the data will be stored

-Consider how best to care for data

Question 2

Which spreadsheet feature uses a set of instructions to perform calculations, such as addition or subtraction?

Formula

Question 3

Which of the following statements correctly describe the archive and the destroy phases of the data life cycle? Select all that apply.

- Shredders may be used to destroy data on paper.

- The archive phase involves storing data.

- A key reason for destroying data is to protect private company information.

Question 4

What tasks may occur during the ask step of the data analysis process? Select all that apply.

- Consider the current state compared to the ideal state

- Define the problem to be solved

- Stay engaged by having a dialogue with stakeholders

Question 5

Fill in the blank: During the prepare step of the data analysis process, data is collected and \_\_\_\_\_ for analysis.

stored

Question 6 (NOT FINAL)

A data team at a timber company works on a project aimed at planting more trees and eliminating deforestation. They discuss what type of data they need and how it will be managed throughout its life cycle. What phase of the data life cycle does this scenario describe?

Answer: Plan

Question 7

A car rental company is interested in improving the customer experience. A data professional fixes typos and inaccuracies from a dataset containing feedback and ratings. They also verify and share their data-cleaning procedures with stakeholders. What step of the data analysis process does this scenario describe?

Process

Question 8

Fill in the blank: During the \_\_\_\_\_ step of the data analysis process, data analytics professionals use tools such as query languages to drive informed decision-making.

analyze

Question 9

What are some key benefits of data visualizations? Select all that apply.

- Maps help people identify patterns in data.

- Data visualizations help people gain insights from data more quickly.

- Complex data can be shared more effectively with nontechnical audiences in visualizations.

Question 10

Fill in the blank: An agriculture company collects data from an outside commodity market database and internal files containing information about livestock. This work occurs during the \_\_\_\_\_ phase of the data life cycle.

Capture

# Step-by-Step: Make spreadsheets your friend (Assuming its Microsoft Excel)

This reading outlines the steps the instructor performs in the next video, [Make spreadsheets your friend.](https://www.coursera.org/learn/foundations-data/lecture/xUZNB/columns-and-rows-and-cells-oh-my) In the video, the instructor demonstrates some spreadsheet basics and discusses how to use them to organize and analyze data.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

**What you’ll need**

Use this guide as a reference if you need additional context or clarification as you watch the next video. If you’d like to follow along with the video, choose a spreadsheet tool and open a blank sheet.



## Example 1: Get started

Enter basic data:

1. Begin with a new spreadsheet.
2. Select cell **A2**.
3. Enter your **first name**.
4. Select cell **B2**.
5. Enter your **last name**.

Adjust the size of rows and columns:

To make the text fit in the rows and columns, adjust their sizes. Use either of the following methods:

1. If your name is longer than the width of the column, **select and drag** the right edge of the corresponding column until it fits.
2. To **wrap text**, select the cells, columns, or rows with text that you want to reformat.
3. Select the **Format** menu.
4. Under **Wrapping**, select **Wrap**.

## Example 2: Add labels

Add labels, or attributes, to help you keep track of the data:

1. Select cell **A1**.
2. Enter **First Name**.
3. Select cell **B1**.
4. Enter **Last Name**.
5. Select cells **A1** and **B1**. To do this, select a single cell and drag your cursor over to the other cell to include it in the selection.
6. From the toolbar, select the **bold icon**.

## Example 3: Add more attributes and data

Add more attributes and data to your spreadsheet:

1. Select cell **C1** and enter **Siblings**.
2. Select cell **D1** and enter **Favorite Color**.
3. Select cell **E1** and enter **Favorite Dessert**.
4. Select **all three cells** and make them bold by selecting the **bold icon** from the toolbar.
5. Adjust the columns to fit the new text.
6. Enter the corresponding data in cells **C2**, **D2**, and **E2** (your number of siblings, favorite color, and favorite dessert).
7. Add data about two more people in **rows 3 and 4**. These can be people you know or people you’ve just made up.

## Example 4: Organize your data

One way to organize your data is by sorting it.

1. Select all columns that contain data. There are a few ways to select multiple cells:
   1. To select nonadjacent cells and/or cell ranges, hold the **Command** (Mac) or **Ctrl** (PC) **key** and select the cells.
   2. To select a range of cells, hold the **Shift** key and either drag your cursor over which cells you want to include or use the arrow keys to select a range.
   3. Select a single cell and drag your cursor over the cells you want to include in your selection.
2. Select the **Data** menu.
3. Select **Sort range**, then select **Advanced range sorting options**.
4. In the **Advanced range sorting options** window, select the checkbox for **Data has header row**. Make sure that **A to Z** is selected.
5. Select the **Sort by** drop-down menu, then select **Siblings**.
6. Select **Sort**. This will organize the spreadsheet by the number of siblings, from lowest to highest.

## Example 5: Use a formula

Spreadsheets enable data professionals to analyze data. In this example, the instructor uses a formula to calculate a sum.

1. Select the next empty cell in the **Siblings** column (**C5**).
2. Enter the formula **=C2+C3+C4**.
3. Press **Enter** on your keyboard to complete the formula.
4. The formula calculates the total number of siblings.

**Attribute**

A characteristic or quality of data used to label a column in a table.

**Observation**

All of the attributes for something contained in a row of a data table

In a table, an attribute is a characteristic or quality of data used for what purpose?

To label a column

# Hands-On Activity: Generate a chart from a spreadsheet

Question 1

## Reflection



Based on the chart you created, when is the busiest time of year for your company?

Answer: December-February

Question 2

Examine your completed data visualization. In the space provided below, write 2-3 sentences (40-60 words) in response to each of the following questions:

* How does the line chart help you visualize the sales data?
* Based on your analysis, when will your company need to increase staff and inventory?

My Answer: To pinpoint the monitoring behavior in a set of sales data. Based on my analysis the company should increase around December to February to achieve this.

## Question

In a table, an attribute is a characteristic or quality of data used for what purpose?

To label a column

# Test your knowledge on spreadsheet basics

* 1. In this spreadsheet, what number is in cell B4?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *(n/a)* | **A** | **B** | **C** | **D** | **E** |
| **1** | 909 | 365 | 794 | 649 | 846 |
| **2** | 347 | 221 | 351 | 737 | 362 |
| **3** | 520 | 705 | 990 | 256 | 393 |
| **4** | 797 | 986 | 184 | 292 | 482 |
| **5** | 993 | 161 | 212 | 764 | 970 |

Answer: 986

Question 2

What spreadsheet feature is used to change the height of a cell to allow the text inside to fit?

Text wrap

Question 3

In this spreadsheet, which of the following are attributes? Select all that apply.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *(n/a)* | **A** | **B** | **C** | **D** |
| *(n/a)* | *(n/a)* | **Roses** | **Irises** | **Lilies** |
| **1** | **Revenue** | $15,837 | $10,224 | $13,857 |
| **2** | **Expenses** | $6,409 | $3,592 | $4,002 |
| **3** | **Net profit** | $9,428 | $6,632 | $9,855 |

Irises, Lilies

Question 4

What symbol is used to start a formula?

Equal sign (=)

SQL

-Store

-Organise

-Analyze

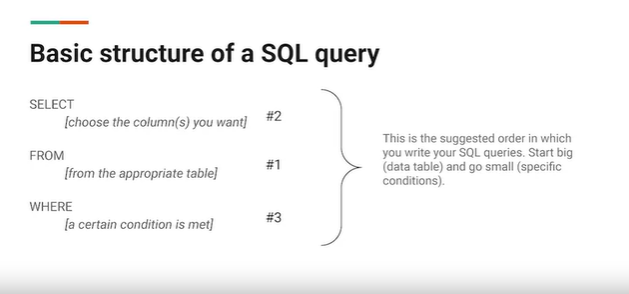
Examples include:

Oracle

MySQL

PostgreSQL

Microsoft SQL Server



Fill in the blank: A data analyst uses a SQL query to retrieve information from a database. They add a WHERE statement to \_\_\_\_\_ the data based on certain conditions

Answer: filter

Speakers for this course regarding data representation.

“Every time I learn a new skill, I feel like I’m learning how to speak all over again.”

“Data visualizations are pictures. They are a wonderful way to take very basic ideas around data and data points and make them come alive.”

# Test your knowledge on SQL and data visualization

1. Which part of the following query requests that all columns from the table be selected?

**SELECT \***

**FROM customers**

**WHERE country = 'France'**

Asterisk (**\***)

2.

Which part of the following query indicates the table from which to retrieve data?

**FROM**

3.

In the following query, data will be retrieved from which table?

**SELECT \***

**FROM inventory**

**WHERE product = 'vacuum'**

Inventory

4.

In which of the following scenarios would a line chart be the most appropriate visualization?

When showing the change in someone's salary over time

# Glossary terms from module 3

## **Terms and definitions for Course 1, Module 3**

**Attribute**: A characteristic or quality of data used to label a column in a table

**Observation:** The attributes that describe a piece of data contained in a row of a tabl

# Module 3 challenge

Question 1

In the following spreadsheet, what type of tree is in cell D3?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (n/a) | A | B | C | D | E |
| 1 | Ginkgo | Weeping willow | Sycamore | Bay laurel | Pistachio |
| 2 | Papaya | Maple | White oak | European ash | Pecan |
| 3 | Cedar | Burflower | Spruce | Redwood | Beech |
| 4 | Birch | Cottonwood | Clove | Cacao | Bristlecone |
| 5 | Rubber tree | Brazil nut | Walnut | Pine | Palm |

Redwood

Question 2

In the following query, what will be retrieved from the database?

|  |
| --- |
| **SELECT \***  **FROM Storeroom 5**  **WHERE glasses = 'bifocals'**  All glasses in Storeroom 5 that are bifocals  Question 3 (All answers in the options are not finalised)  Which of the following statements accurately describe spreadsheet attributes and observations? Select all that apply.  -An observation includes all of the attributes contained in its row  -Attributes are also referred to as column headers |

Question 4

A data team at a public university wants to communicate to stakeholders about which foreign language courses are most popular with students. They create a data visualization that identifies the 38 different courses, then shows the number of students enrolled in each one. What type of data visualization should they create?

Bar chart

Question 5

When working in a spreadsheet, what is the correct syntax for a formula that adds the value in cell C8 to the value in cell D8?

**=C8+D8**

Question 6.

Fill in the blank: The SQL clause **SELECT \*** is used to retrieve all data from a particular \_\_\_\_\_.

Table

Question 7

In the following query, which clause indicates the table from which to retrieve data?

|  |
| --- |
| **SELECT \***  **FROM Garden**  **WHERE herb = 'basil'** |

From Garden

Question 8 (NOT FINAL)

Which text wrapping feature allows the text within a spreadsheet cell to continue into the next cell, if it is empty

Answer:Clip or Overflow

Question 9

Which of the following statements accurately describe data visualizations and visualization tools? Select all that apply.

-Data professionals can create bar graphs, pie charts, and maps in a spreadsheet.

-When working with the programming language R, data professionals can create dashboard-style data visualizations.

- Tableau is a data visualization tool that enables simple and quick data-sharing.

Question 10

A data professional wants the data in column A to be in alphabetical order from A to Z. What spreadsheet feature would enable them to accomplish this task?

|  |  |  |  |
| --- | --- | --- | --- |
| *(n/a)* | A | B | C |
| 1 | Deciani | Lin | 33 |
| 2 | Roberts | Jeremy | 28 |
| 3 | Storelli | Hyun | 64 |
| 4 | Piotrowski | Sook | 84 |
| 5 | Mondek | Priscilla | 42 |
| 6 | Lam | April | 96 |

Sort range

-The role of a Data Analyst

-Business tasks for Data Analysts

-Fairness in analysis

-Opportunities for you

-And you future success!

“In analytics, I feel like the key to success is being able to blend the personal side with the technical side.”

# Self-Reflection: Business use of data

Question 1

## Reflection



Consider the company, service, or product you chose in this reflection:

* How could it use data to improve customer experience?
* What kinds of data would it need to collect?
* How could insights from that data solve a problem?

Now, write 2-3 sentences (40-60 words) in response to each of these questions. Enter your response in the text box below.

Your team could collect many types of data, such as statistical information and qualitative data, like preferences and motivations.

Two primary types of data can be collected: quantitative data and qualitative data

An insight helps us to solve complex problems because it occurs when we put diffuse information across the brain together in a new way, rearranging it and understanding it in a way that we hadn't before.

“There’s so many different paths that you can take from the starting point that you really can’t predict your end.”

**Issue**

A topic or subject to investigate

**Question**

Designed to discover information

**Problem**

An obstacle or complication that needs to be worked out

Analyze weather data from the last decade to identify predictable patterns.

**Data-driven decision-making**

Using facts to guide business strategy

“I think one of the most important things to remember about data analytics is that data is data.”

“I found that data acts like a living and breathing thing.”

**Fairness**

Ensuring that your analysis doesn’t create or reinforce bias.

# Consider fairness

Previously, you learned that part of a data professional’s responsibility is to make certain that their analysis is fair. **Fairness** means ensuring your analysis doesn't create or reinforce bias. This can be challenging, but if the analysis is not objective, the conclusions can be misleading and even harmful. In this reading, you’re going to explore some best practices you can use to guide your work toward a more fair analysis!

## Consider fairness

Following are some strategies that support fair analysis:

| **Best practice** | **Explanation** | **Example** |
| --- | --- | --- |
| Consider all of the available data | Part of your job as a data analyst is to determine what data is going to be useful for your analysis. Often there will be data that isn’t relevant to what you’re focusing on or doesn’t seem to align with your expectations. But you can’t just ignore it; it’s critical to consider all of the available data so that your analysis reflects the truth and not just your own expectations. | A state’s Department of Transportation is interested in measuring traffic patterns on holidays. At first, they only include metrics related to traffic volumes and the fact that the days are holidays. But the data team realizes they failed to consider how weather on these holidays might also affect traffic volumes. Considering this additional data helps them gain more complete insights. |
| Identify surrounding factors | As you’ll learn throughout these courses, context is key for you and your stakeholders to understand the final conclusions of any analysis. Similar to considering all of the data, you also must understand surrounding factors that could influence the insights you’re gaining. | A human resources department wants to better plan for employee vacation time in order to anticipate staffing needs. HR uses a list of national bank holidays as a key part of the data-gathering process. But they fail to consider important holidays that aren’t on the bank calendar, which introduces bias against employees who celebrate them. It also gives HR less useful results because bank holidays may not necessarily apply to their actual employee population. |
| Include self-reported data | **Self-reporting** is a data collection technique where participants provide information about themselves. Self-reported data can be a great way to introduce fairness in your data collection process. People bring conscious and unconscious bias to their observations about the world, including about other people. Using self-reporting methods to collect data can help avoid these observer biases. Additionally, separating self-reported data from other data you collect provides important context to your conclusions! | A data analyst is working on a project for a brick-and-mortar retailer. Their goal is to learn more about their customer base. This data analyst knows they need to consider fairness when they collect data; they decide to create a survey so that customers can self-report information about themselves. By doing that, they avoid bias that might be introduced with other demographic data collection methods. For example, if they had sales associates report their observations about customers, they might introduce any unconscious bias the employees had to the data. |
| Use oversampling effectively | When collecting data about a population, it’s important to be aware of the actual makeup of that population. Sometimes, oversampling can help you represent groups in that population that otherwise wouldn’t be represented fairly. **Oversampling** is the process of increasing the sample size of nondominant groups in a population. This can help you better represent them and address imbalanced datasets. | A fitness company is releasing new digital content for users of their equipment. They are interested in designing content that appeals to different users, knowing that different people may interact with their equipment in different ways. For example, part of their user-base is age 70 or older. In order to represent these users, they oversample them in their data. That way, decisions they make about their fitness content will be more inclusive. |
| Think about fairness from beginning to end | To ensure that your analysis and final conclusions are fair, be sure to consider fairness from the earliest stages of a project to when you act on the data insights. This means that data collection, cleaning, processing, and analysis are all performed with fairness in mind. | A data team kicks off a project by including fairness measures in their data-collection process. These measures include oversampling their population and using self-reported data. However, they fail to inform stakeholders about these measures during the presentation. As a result, stakeholders leave with skewed understandings of the data. Learning from this experience, they add key information about fairness considerations to future stakeholder presentations. |

# Self-Reflection: Business cases

## Overview



Now that you have explored how businesses use data in the real world, pause for a moment and think about what you are learning. In this self-reflection, you will consider fairness and data use in three example business cases and respond to brief questions with your thoughts.

### 1.

Question 1

### Case Study #1



To improve the effectiveness of its teaching staff, the administration of a high school offered the opportunity for all teachers to participate in a workshop. They were not required to attend; instead, the administration encouraged teachers to sign up. Of the 43 teachers on staff, 19 chose to take the workshop.

At the end of the academic year, the administration collected data on teacher performance for all teachers on staff. The data was collected via student survey. In the survey, students were asked to rank each teacher's effectiveness on a scale of 1 (very poor) to 6 (very good).

The administration compared data on teachers who attended the workshop to data on teachers who did not. The comparison revealed that teachers who attended the workshop had an average score of 4.95, while teachers who did not attend had an average score of 4.22. The administration concluded that the workshop was a success.

## Reflection



Consider this scenario:

* What are the examples of fair or unfair practices?
* How could a data analyst correct the unfair practices?

Now, write 2-3 sentences (40-60 words) in response to each of these questions. Enter your response in the text box below.

1 / 1 point

What are the examples of fair or unfair practices? The statistics show similar ratings for teachers who attended the workshop as opposed to ones who did not attend the workshop. They should account other key data information regarding this issue, for instance if there are unlawful practices when it comes to students' performance and methods of teaching. They should take the workshop to able to receive feedback from students concerning about the context of the subject that was given during the curriculum. How could a data analyst correct the unfair practices? The data should make the workshops mandatory and accessible for all teachers, making data integrity and fairness to reduce bias to ensure that workshops are needed to ensure the school improves the effectiveness of its teaching staff as well as increasing efficiency in their methods of teaching. May well lead to better ratings that workshops are better and needed to understand their feedback is crucial to education success. In addition, A data analyst could correct these unfair practices by implementing a matched-pair analysis to control for pre-existing differences among teachers, using statistical methods such as regression to account for potential confounding variables, and ensuring a randomized control trial design in future evaluations to better isolate the effect of the workshop.

Correct

Great work reinforcing your learning with a thoughtful self-reflection! In your response, you likely noted that this is an example of unfair practice. It is tempting to conclude—as the administration did—that the workshop was a success. However, since the workshop was voluntary and not random, it is not appropriate to infer a causal relationship between attending the workshop and the higher rating.

The workshop might have been effective, but other explanations for the differences in the ratings cannot be ruled out. For example, another explanation could be that the staff volunteering for the workshop were the better, more motivated teachers. This group of teachers would be rated higher whether or not the workshop was effective.

It’s also notable that there is no direct connection between student survey responses and workshop attendance. The data analyst could correct this by asking for the teachers to be selected randomly to participate in the workshop. They could also collect data that measures something more directly related to workshop attendance, such as the success of a technique the teachers learned in that workshop.

### 2.

Question 2

### Case Study #2



An automotive company tests the driving capabilities of its self-driving car prototype. They carry out the tests on various types of roadways—specifically, a race track, trail track, and dirt road.

The researchers only test the prototype during the daytime. They collect two types of data: sensor data from the car during the drives and video data of the drives from cameras on the car.

They review the data after the initial tests. The results illustrate that the new self-driving car meets the performance standards across each of the roadways. As a result, the car can progress to the next phase of testing, which will include driving in various weather conditions.

## Reflection



Consider this scenario:

* What are the examples of fair or unfair practices?
* How could a data analyst correct the unfair practices?

Now, write 2-3 sentences (40-60 words) in response to each of these questions. Enter your response in the text box below.

1 / 1 point

Examples of Fair or Unfair Practices Fair practices include testing the self-driving car on various types of roadways and using both sensor and video data to assess performance. Unfair practices involve only testing the car during daytime, which does not account for performance variability under different lighting conditions. How a Data Analyst Could Correct Unfair Practices A data analyst could correct these unfair practices by ensuring tests are conducted at various times of the day to include nighttime conditions, thereby providing a more comprehensive assessment of the car's capabilities across different lighting scenarios. They should also consider expanding tests to include more diverse roadway types and traffic conditions.

Correct

Once again, this case study shows an unfair practice. Your response probably mentioned that, although the researchers test the prototype on three different tracks, they only conduct tests during the day.

Conditions on each track may be very different during the day and night and this could change the results significantly. The data analyst should correct this by asking the test team to add in nighttime testing to get a full perspective of how the prototype performs at any time of the day on the tracks.

### 3.

Question 3

### Case Study #3



An amusement park plans to add new rides to their property. First, they need to determine what kinds of new rides visitors want the park to build. In order to understand their visitors’ interests, the park develops a survey.

They decide to distribute the survey near the roller coasters because the lines are long enough that visitors will have time to answer all of the questions. After collecting this survey data, they find that most of the respondents want more roller coasters at the park. They conclude that they should add more roller coasters, as most of their visitors prefer them.

### Reflection



Consider this scenario:

* What are the examples of fair or unfair practices?
* How could a data analyst correct the unfair practices?

Now, write 2-3 sentences (40-60 words) in response to each of these questions. Enter your response in the text box below.

1 / 1 point

Examples of Fair or Unfair Practices Fair practices include developing a survey to understand visitor preferences and collecting feedback directly from visitors. Unfair practices involve distributing the survey only near roller coasters, leading to a biased sample that likely overrepresents the preferences of roller coaster enthusiasts. How a Data Analyst Could Correct Unfair Practices A data analyst could correct these unfair practices by distributing the survey across various areas of the park, ensuring a more diverse and representative sample of visitors. This would include locations near different types of rides and attractions, allowing for a more accurate assessment of overall visitor preferences.

Correct

A thoughtful reflection would have included information about the decision to distribute surveys in places where visitors would have time to respond. Otherwise, it accidentally introduces sampling bias.

The only respondents to the survey are people waiting in line for the roller coasters. This may unfairly bias survey results, because respondents might prefer roller coasters. A data analyst could reduce sampling bias by distributing the survey at the entrance and exit of the amusement park. This would avoid targeting roller coaster fans and provide results from the park’s general audience.

# Case opinion

Recently, you were presented with cases about data analytics in the real world. One case involved an unfair conclusion about the performance of women who worked at a business. It demonstrated that data can sometimes be true, yet unfair. In addition, it highlighted the importance of asking, "Why?" when reviewing the results of data analysis.

Another example involved data analysts prioritizing fairness and going out of their way to ensure their data was as fair as possible. Because they were working with sensitive and potentially biased health data, they chose to collaborate with social scientists in order to better understand the social context behind that data.

If you need to, return to the video to refresh your understanding of the examples before you continue. Then, discuss the first case and how the analysts at that company could improve their process:

* What could they have done differently to be fairer in their analysis?
* What could have made their conclusion less biased?

Improving Fairness in Analysis

In the first case involving an unfair conclusion about the performance of women at a business, the analysts could have taken several steps to ensure a fairer analysis. Firstly, they should have examined potential biases in the data collection process, such as whether performance metrics were applied consistently across genders and if there were any underlying biases in the evaluation criteria. They could have implemented blind evaluations to reduce the influence of gender stereotypes. Additionally, conducting intersectional analysis to understand how various factors such as race, age, and tenure intersect with gender could provide a more nuanced understanding of performance differences.

Reducing Bias in Conclusions

To make their conclusions less biased, the analysts should have included contextual factors that might explain performance disparities. For instance, if women were disproportionately assigned to less profitable departments or given fewer opportunities for advancement, these contextual elements should be accounted for in the analysis. Collaborating with social scientists or gender studies experts could provide insights into structural inequalities and help interpret the data more comprehensively. By adopting a holistic approach and considering external factors that influence performance, the analysts would be better positioned to draw fair and unbiased conclusions, ultimately leading to more equitable business practices.

‘How do we actually improve the lives of people by using data?”

“At the end of the day, data are people.”

# Test your knowledge on making fair business decisions

Question 1

In a data analytics context, what is a business task?

The question answered, or problem solved, by data analysis

In a data analytics context, a business task is the question answered, or problem solved, by data analysis

Question 2

Fill in the blank: Fairness means ensuring that bias is neither created nor \_\_\_\_\_ by the data analysis process.

Reinforced

Fairness means ensuring that bias is neither created nor reinforced by the data analysis process.

Question 3

What are some strategies data professionals use to ensure their data analysis is unbiased? Select all that apply.

Prioritize fairness during data collection

Consider the complicated social contexts that could create bias

Ensure anything shared with stakeholders is presented in context

To ensure their data analysis is unbiased, data professionals consider the complicated social contexts that could create bias, prioritize fairness during data collection, and ensure anything shared with stakeholders is presented in context.

To ensure their data analysis is unbiased, data professionals consider the complicated social contexts that could create bias, prioritize fairness during data collection, and ensure anything shared with stakeholders is presented in context.

To ensure their data analysis is unbiased, data professionals consider the complicated social contexts that could create bias, prioritize fairness during data collection, and ensure anything shared with stakeholders is presented in context.

Question 4

A data professional notices that their dataset is imbalanced and therefore is not representative of the entire population. They work to get more responses from

nondominant groups in order to address this issue and ensure data insights are fair. What does this scenario describe?

Oversampling

This scenario describes oversampling. Oversampling is the process of increasing the sample size of nondominant groups in order to better represent them and address imbalance datasets.

# 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 there are over 380,000 job openings in the Data Analytics field in the United States\*. 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.

\* Burning Glass data, Feb 1, 2021 - Jan 31, 2022, US

## Decoding the job description

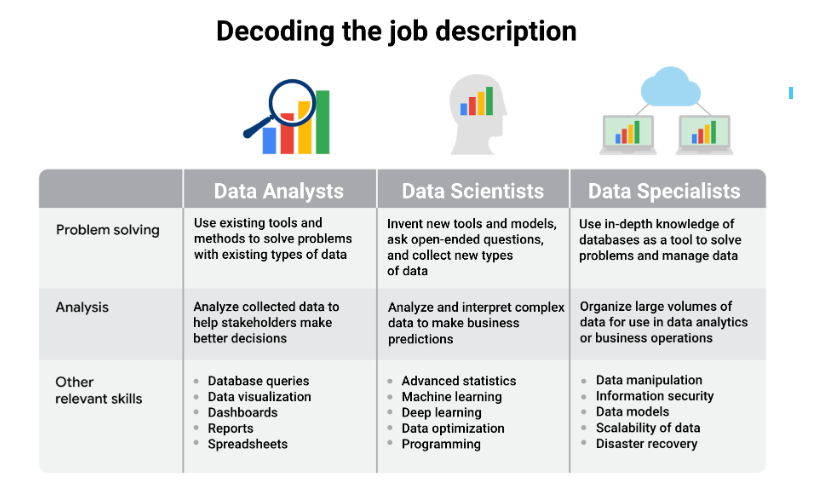
The data analyst role is one of many job titles that contain the word “analyst.”

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:



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 takeaways

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.

“Think about a time where you’ve used data to solve a problem, whether it’s in your professional or personal projects.”

“Increase your professional network.”

“It’s really important to have your LinkedIn updated along with websites, like GitHub where you can showcase a lot of the data analyst projects you’ve done.”

“Prepare questions for the interviewer.”

“Sometimes there is no right answer, and a lot of times interviewers are looking to see your thought process and the way you get to your solution.”

“Look for the recruiter. Look for the hiring manager online. See if you can reach out to them and set up a coffee chat or send them your resume directly.”

# Glossary terms from module 4

## **Terms and definitions for Course 1, Module 4**

**Business task:** The question or problem data analysis resolves for a business

**Fairness:** A quality of data analysis that does not create or reinforce bias

**Oversampling:** The process of increasing the sample size of nondominant groups in a population. This can help you better represent them and address imbalanced datasets

**Self-reporting:** Adata collection technique where participants provide information about themselves

# Module 4 challenge

1 (Some of the answers are not final)

Which of the following statements accurately describe fairness considerations in data analysis? Select all that apply.

- Best practices for fairness in data analysis include identifying surrounding factors.

- Effective data analysts help create systems that are fair and inclusive to everyone.

- A data professional may include self-reported data when prioritizing fairness.

### 2.

An HR firm is filling open positions for administrative assistants, so they purchase online help-wanted ads. Research reveals that 87% of administrative assistants are women, so the data team decides that women are more likely to be successful applicants. Therefore, they target the ads to female job seekers. What should they have done instead?

Conduct more research to understand the surrounding factors of this situation.

### 3.

### Which fairness best practice is intended to help data teams understand circumstances that could have an impact on their data insights?

Identify surrounding factors

### 4.(NOT FINAL)

While conducting a survey to learn about employee satisfaction, a data analyst notices that the majority of respondents are people who work in the operations department. The fairness of their survey could be improved by over-sampling which employees?

People in operations who have not yet responded to the survey

5.

Fill in the blank: An executive might ask data team members to work on a \_\_\_\_\_ in order to address a particular question or problem using data analysis.

business task

6.

A data professional at a grocery store considers fairness when collecting data. Rather than having store associates share observations, they create a survey that asks customers to provide information about their own shopping experiences. This helps avoid any unconscious bias that might be introduced by the sales associates. Which fairness best practice does this scenario describe?

Self-reporting

7.

A data analyst at a government agency researches ways to reduce traffic congestion. Rather than using only data about when and where traffic occurs, the analyst considers public transportation usage, road and weather conditions, and more. This enables them to achieve more insightful results. Which fairness best practice does this scenario describe?

Considering all available data

8.

Fill in the blank: A data professional ensures their data analysis is fair by considering fairness from \_\_\_\_\_ to the point when their organization acts on the data insights.

the start of a project