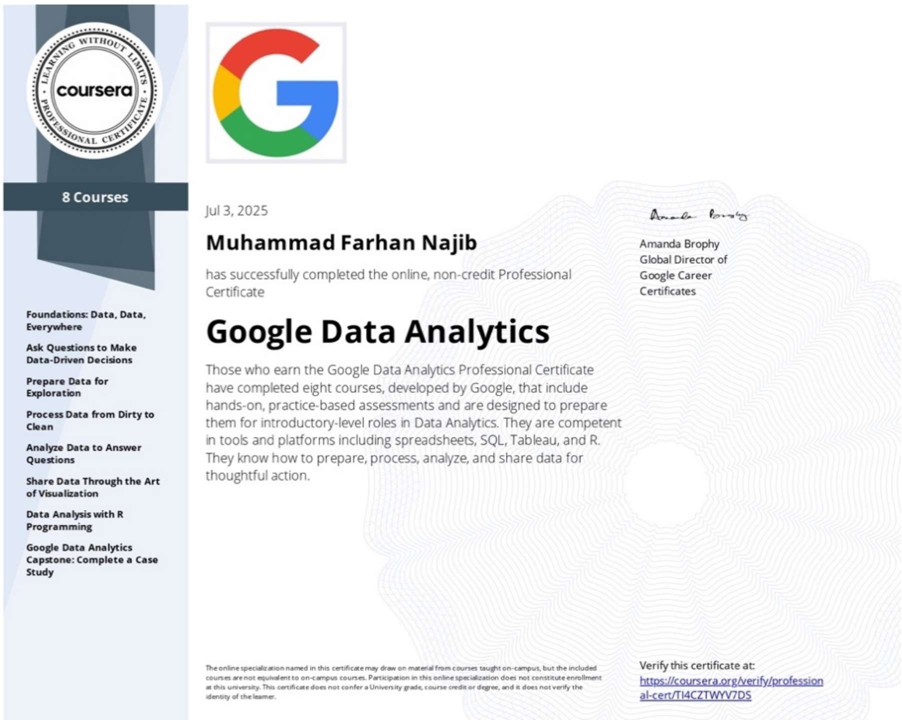
# LAMPIRAN

## Sertifikat Profesional



8 Courses

**Foundations: Data, Data, Everywhere**
**Ask Questions to Make Data-Driven Decisions**
**Prepare Data for Exploration**
**Process Data from Dirty to Clean**
**Analyze Data to Answer Questions**
**Share Data Through the Art of Visualization**
**Data Analysis with R Programming**
**Google Data Analytics Capstone: Complete a Case Study**

Jul 3, 2025

**Muhammad Farhan Najib**

has successfully completed the online, non-credit Professional Certificate

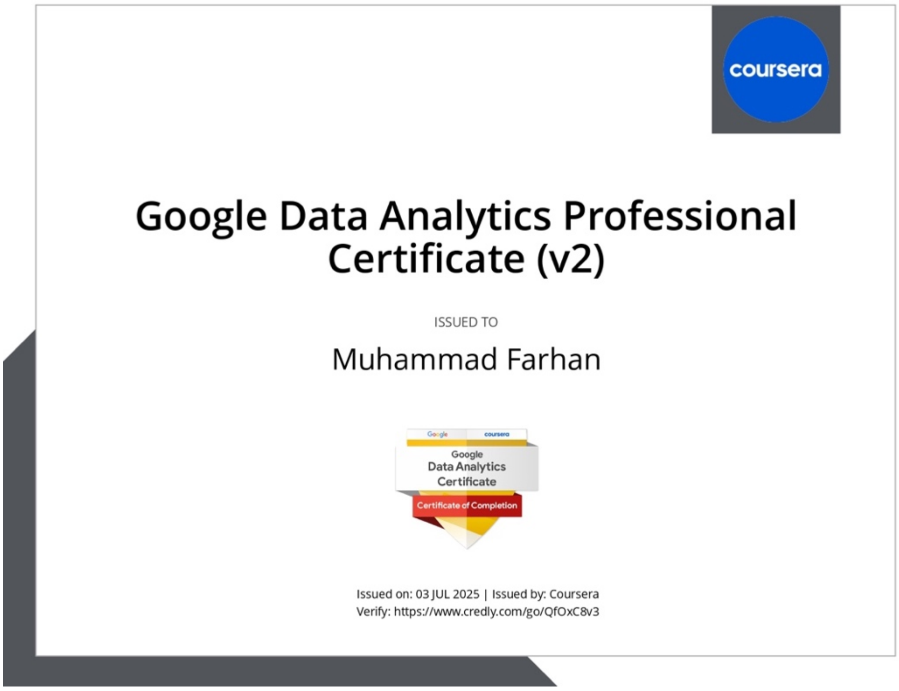
**Google Data Analytics**

Those who earn the Google Data Analytics Professional Certificate have completed eight courses, developed by Google, that include hands-on, practice-based assessments and are designed to prepare them for introductory-level roles in Data Analytics. They are competent in tools and platforms including spreadsheets, SQL, Tableau, and R. They know how to prepare, process, analyze, and share data for thoughtful action.

Amanda Brophy
Global Director of
Google Career
Certificates

The online specialization named in this certificate may draw on material from courses taught on-campus, but the included courses are not equivalent to on-campus courses. Participation in this online specialization does not constitute enrollment at this university. This certificate does not confer a University grade, course credit or degree, and it does not verify the identity of the learner.

Verify this certificate at:
https://coursera.org/verify/professional-cert/TI4CZTWYV7DS



coursera

Google Data Analytics Professional Certificate (v2)

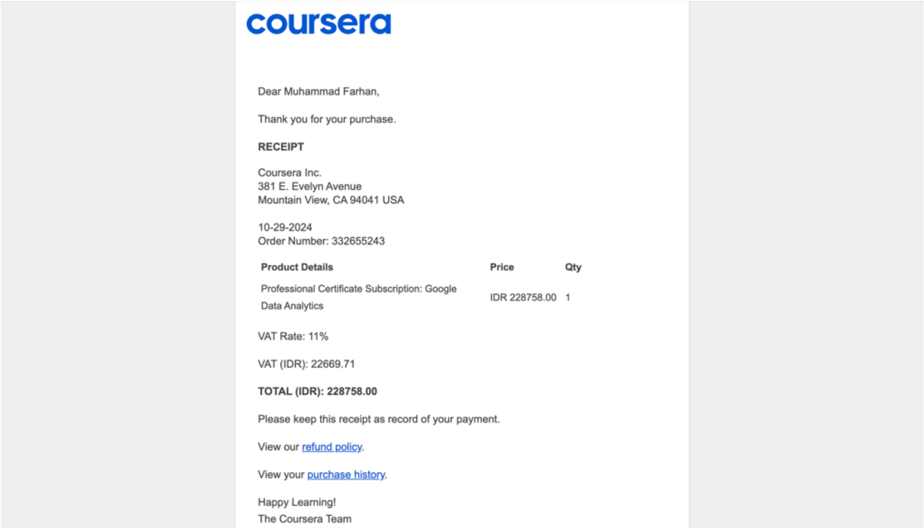
ISSUED TO

Muhammad Farhan

Google coursera
Google
Data Analytics
Certificate
Certificate of Completion

Issued on: 03 JUL 2025 | Issued by: Coursera
Verify: https://www.credly.com/go/QfOxC8v3

## Bukti Pendaftaran



Dear Muhammad Farhan,

Thank you for your purchase.

**RECEIPT**

Coursera Inc.
381 E. Evelyn Avenue
Mountain View, CA 94041 USA

10-29-2024
Order Number: 332655243

| Product Details | Price | Qty |
|---|---|---|
| Professional Certificate Subscription: Google Data Analytics | IDR 228758.00 | 1 |

VAT Rate: 11%

VAT (IDR): 22669.71

**TOTAL (IDR): 228758.00**

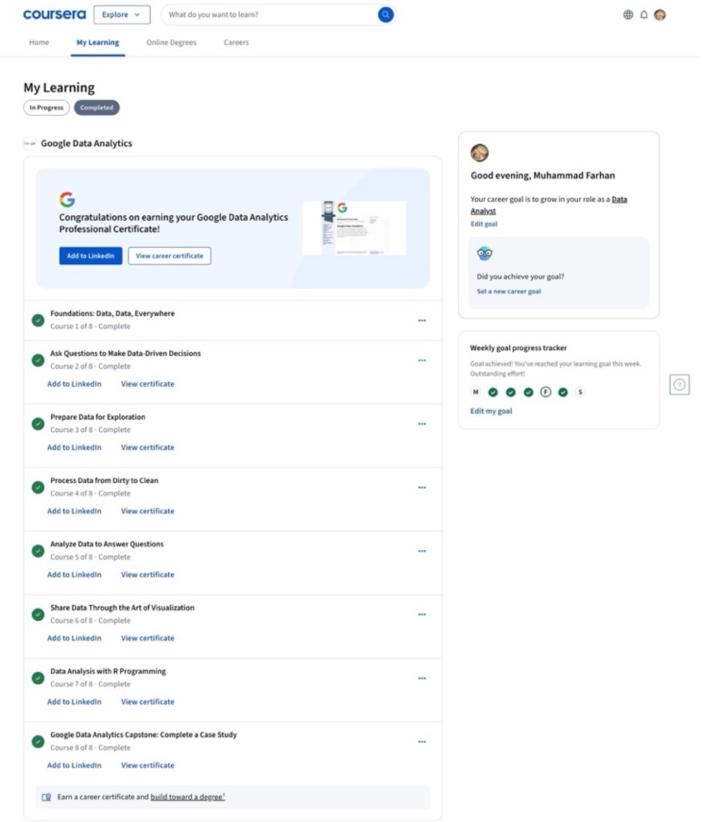
Please keep this receipt as record of your payment.

View our refund policy.

View your purchase history.

Happy Learning!
The Coursera Team

## Bukti Telah Menyelesaikan Program Google Data Analytics Professional Certificate



My Learning

Google Data Analytics

Congratulations on earning your Google Data Analytics Professional Certificate!

Foundations: Data, Data, Everywhere
Course 1 of 8 · Complete

Ask Questions to Make Data-Driven Decisions
Course 2 of 8 · Complete
Add to LinkedIn View certificate

Prepare Data for Exploration
Course 3 of 8 · Complete
Add to LinkedIn View certificate

Process Data from Dirty to Clean
Course 4 of 8 · Complete
Add to LinkedIn View certificate

Analyze Data to Answer Questions
Course 5 of 8 · Complete
Add to LinkedIn View certificate

Share Data Through the Art of Visualization
Course 6 of 8 · Complete
Add to LinkedIn View certificate

Data Analysis with R Programming
Course 7 of 8 · Complete
Add to LinkedIn View certificate

Google Data Analytics Capstone: Complete a Case Study
Course 8 of 8 · Complete
Add to LinkedIn View certificate

Earn a career certificate and build toward a degree[1]

Good evening, Muhammad Farhan

Your career goal is to grow in your role as a Data Analyst

Did you achieve your goal?

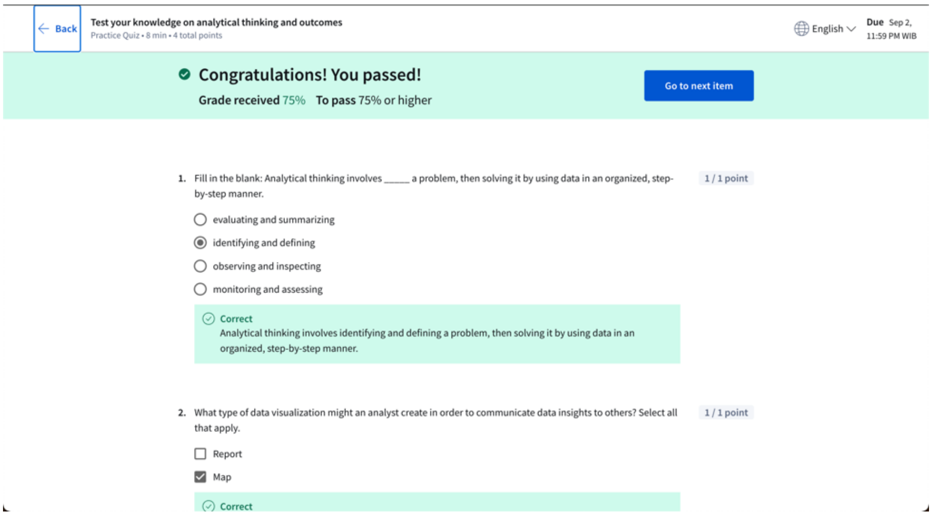
Weekly goal progress tracker

Goal achieved! You've reached your learning goal this week. Outstanding effort!

# Bukti Activity Course 1 Hingga Course 8

## Course 1: Foundation: Data, Data, Everywhere

### Modul 1



**Test your knowledge on analytical thinking and outcomes**
Practice Quiz • 8 min • 4 total points

**Due** Sep 2, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 75% **To pass** 75% or higher

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

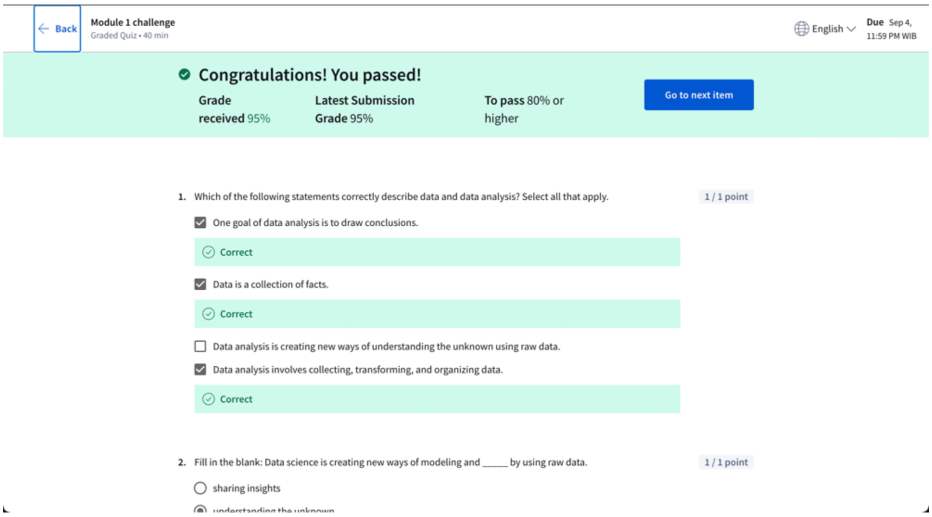
○ evaluating and summarizing
● identifying and defining
○ observing and inspecting
○ monitoring and assessing

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

2. What type of data visualization might an analyst create in order to communicate data insights to others? Select all that apply. 1 / 1 point

☐ Report
☑ Map

✓ Correct



**Module 1 challenge**
Graded Quiz • 40 min

**Due** Sep 4, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 95% **Latest Submission Grade** 95% **To pass** 80% or higher

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

☑ One goal of data analysis is to draw conclusions.

✓ Correct

☑ Data is a collection of facts.

✓ Correct

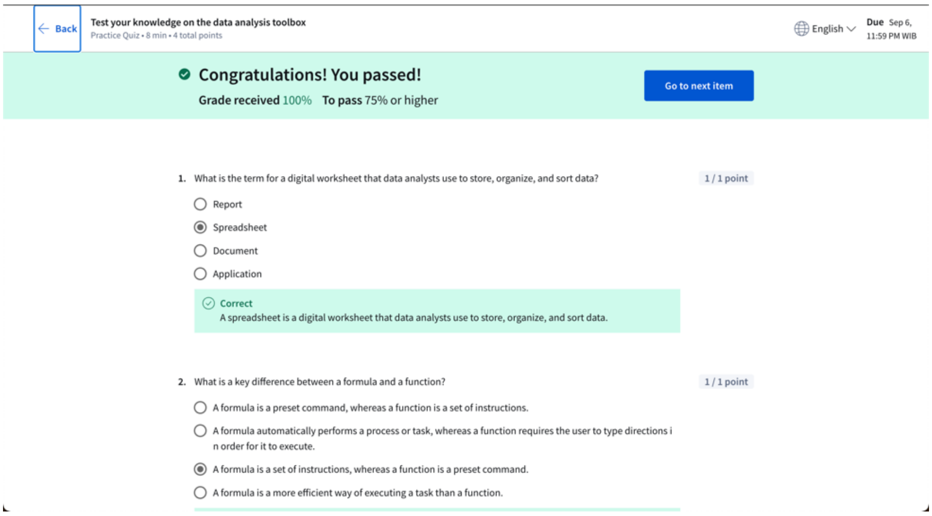
☐ Data analysis is creating new ways of understanding the unknown using raw data.
☑ Data analysis involves collecting, transforming, and organizing data.

✓ Correct

2. Fill in the blank: Data science is creating new ways of modeling and _____ by using raw data. 1 / 1 point

○ sharing insights
● understanding the unknown

# Modul 2



**Test your knowledge on the data analysis toolbox**
Practice Quiz • 8 min • 4 total points

**Due** Sep 6, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

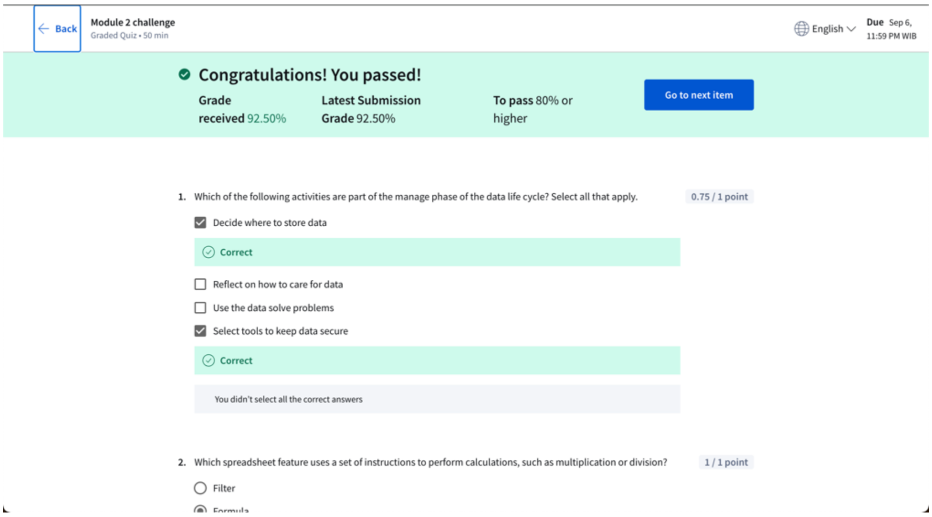
1. What is the term for a digital worksheet that data analysts use to store, organize, and sort data? 1 / 1 point

○ Report
● Spreadsheet
○ Document
○ Application

Correct
A spreadsheet is a digital worksheet that data analysts use to store, organize, and sort data.

2. What is a key difference between a formula and a function? 1 / 1 point

○ A formula is a preset command, whereas a function is a set of instructions.
○ A formula automatically performs a process or task, whereas a function requires the user to type directions i n order for it to execute.
● A formula is a set of instructions, whereas a function is a preset command.
○ A formula is a more efficient way of executing a task than a function.



**Module 2 challenge**
Graded Quiz • 50 min

**Due** Sep 6, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 92.50% **Latest Submission Grade** 92.50% **To pass** 80% or higher

1. Which of the following activities are part of the manage phase of the data life cycle? Select all that apply. 0.75 / 1 point

☑ Decide where to store data

Correct

☐ Reflect on how to care for data
☐ Use the data solve problems
☑ Select tools to keep data secure

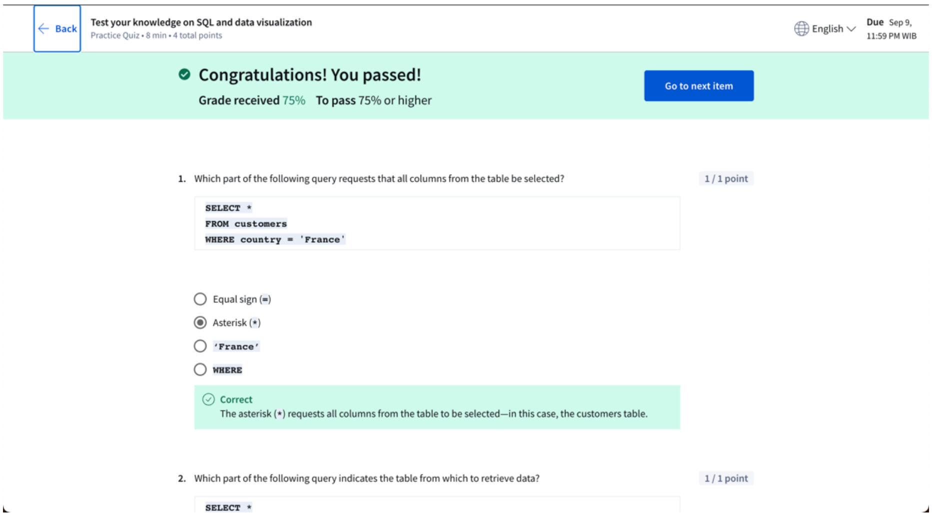
Correct

You didn't select all the correct answers

2. Which spreadsheet feature uses a set of instructions to perform calculations, such as multiplication or division? 1 / 1 point

○ Filter
● Formula

# Modul 3



**Test your knowledge on SQL and data visualization**
Practice Quiz • 8 min • 4 total points

**Due** Sep 9, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 75% **To pass** 75% or higher

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

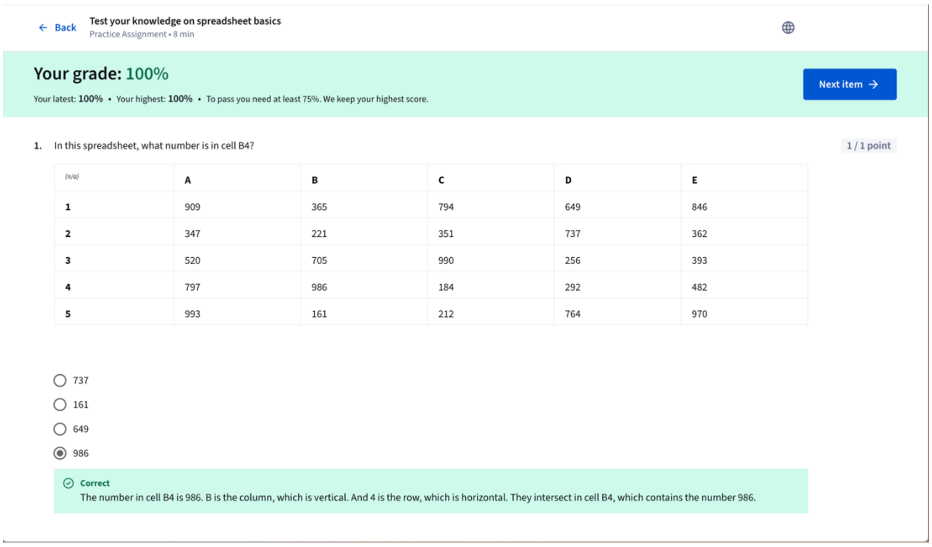
```
SELECT *
FROM customers
WHERE country = 'France'
```

○ Equal sign (=)
● Asterisk (*)
○ `'France'`
○ `WHERE`

Correct
The asterisk (*) requests all columns from the table to be selected—in this case, the customers table.

2. Which part of the following query indicates the table from which to retrieve data? 1 / 1 point

```
SELECT *
```



Test your knowledge on spreadsheet basics
Practice Assignment • 8 min

Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

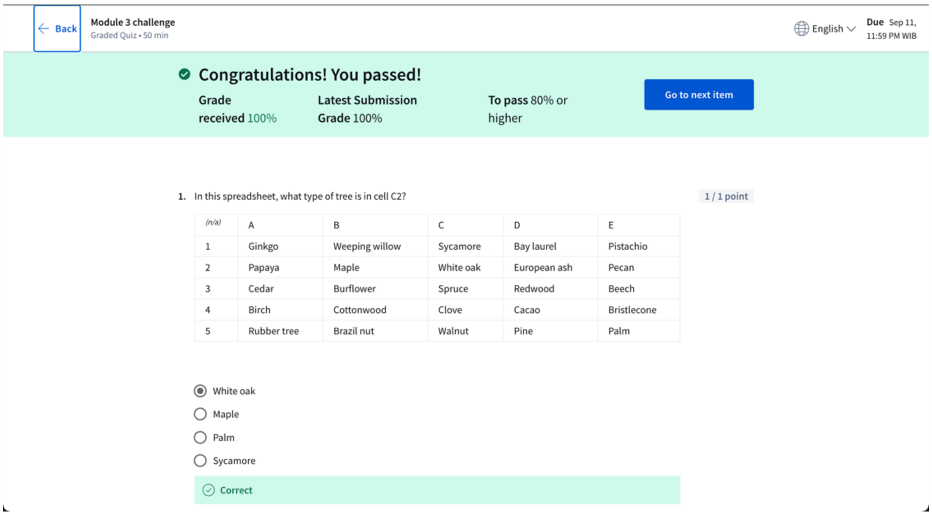
1. In this spreadsheet, what number is in cell B4?

1 / 1 point

| (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 |

○ 737
○ 161
○ 649
● 986

Correct
The number in cell B4 is 986. B is the column, which is vertical. And 4 is the row, which is horizontal. They intersect in cell B4, which contains the number 986.



Module 3 challenge
Graded Quiz • 50 min

**Due** Sep 11, 11:59 PM WIB

Congratulations! You passed!

Grade received 100% | Latest Submission Grade 100% | To pass 80% or higher

1. In this spreadsheet, what type of tree is in cell C2?

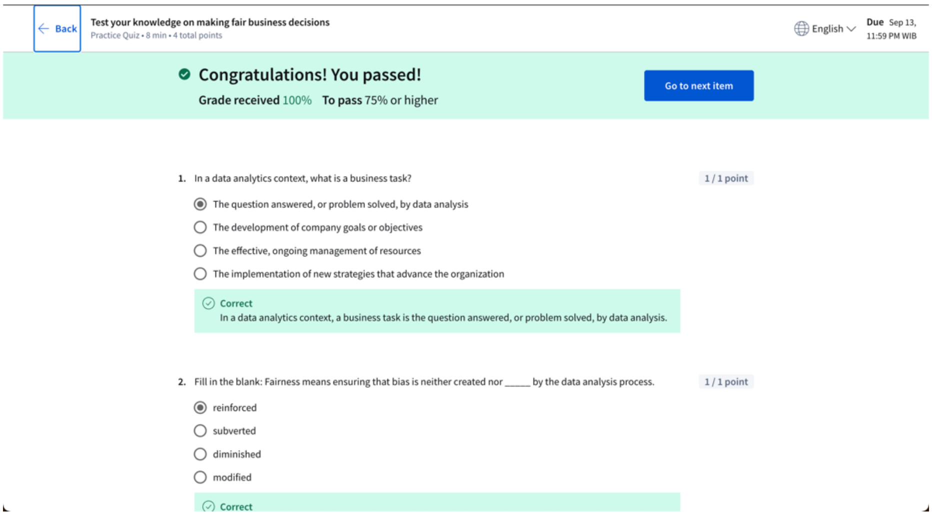
1 / 1 point

| (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 |

● White oak
○ Maple
○ Palm
○ Sycamore

Correct

# Modul 4



Test your knowledge on making fair business decisions
Practice Quiz • 8 min • 4 total points

**Due** Sep 13, 11:59 PM WIB

Congratulations! You passed!

Grade received 100% To pass 75% or higher

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

1 / 1 point

● The question answered, or problem solved, by data analysis
○ The development of company goals or objectives
○ The effective, ongoing management of resources
○ The implementation of new strategies that advance the organization

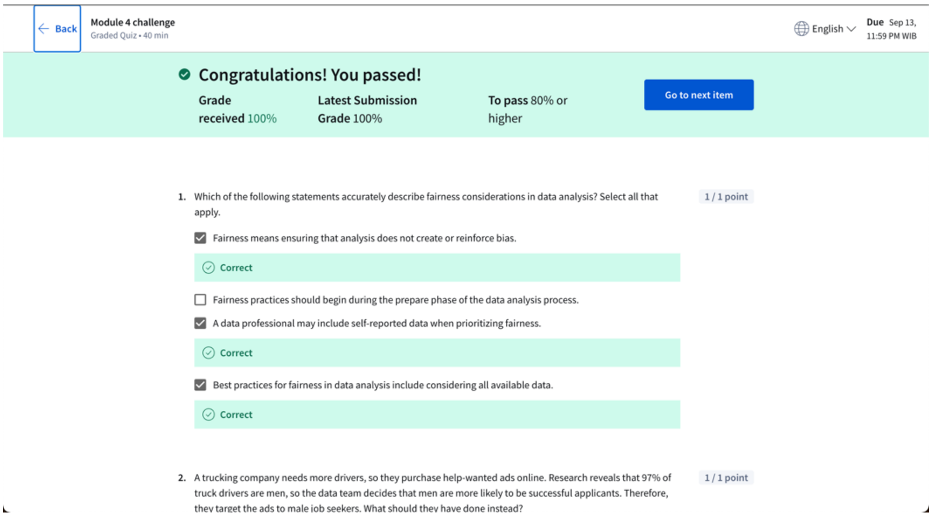
Correct
In a data analytics context, a business task is the question answered, or problem solved, by data analysis.

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

1 / 1 point

● reinforced
○ subverted
○ diminished
○ modified

Correct



**Module 4 challenge**
Graded Quiz • 40 min

**Due** Sep 13, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100%

**Latest Submission Grade** 100%

**To pass** 80% or higher

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

1 / 1 point

☑ Fairness means ensuring that analysis does not create or reinforce bias.

✓ Correct

☐ Fairness practices should begin during the prepare phase of the data analysis process.

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

✓ Correct

☑ Best practices for fairness in data analysis include considering all available data.

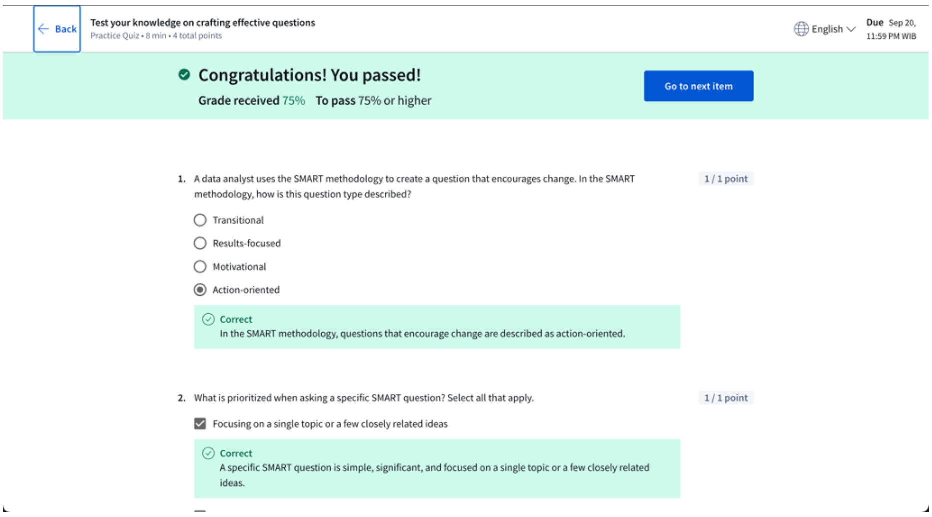
✓ Correct

2. A trucking company needs more drivers, so they purchase help-wanted ads online. Research reveals that 97% of truck drivers are men, so the data team decides that men are more likely to be successful applicants. Therefore, they target the ads to male job seekers. What should they have done instead?

1 / 1 point

# Course 2: Ask Questions to Make Data-Driven Decisions

## Modul 1



**Test your knowledge on crafting effective questions**
Practice Quiz • 8 min • 4 total points

**Congratulations! You passed!**

**Grade received** 75% **To pass** 75% or higher

1. A data analyst uses the SMART methodology to create a question that encourages change. In the SMART methodology, how is this question type described?

1 / 1 point

○ Transitional
○ Results-focused
○ Motivational
● Action-oriented

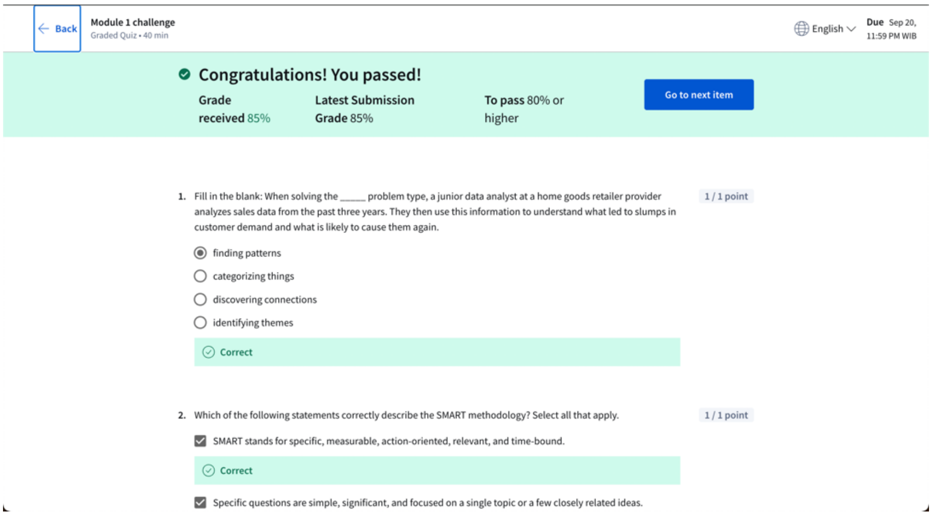
Correct
In the SMART methodology, questions that encourage change are described as action-oriented.

2. What is prioritized when asking a specific SMART question? Select all that apply.

1 / 1 point

☑ Focusing on a single topic or a few closely related ideas

Correct
A specific SMART question is simple, significant, and focused on a single topic or a few closely related ideas.



**Module 1 challenge**
Graded Quiz • 40 min

**Congratulations! You passed!**

**Grade received** 85% **Latest Submission Grade** 85% **To pass** 80% or higher

1. Fill in the blank: When solving the _____ problem type, a junior data analyst at a home goods retailer provider analyzes sales data from the past three years. They then use this information to understand what led to slumps in customer demand and what is likely to cause them again.

1 / 1 point

● finding patterns
○ categorizing things
○ discovering connections
○ identifying themes

Correct

2. Which of the following statements correctly describe the SMART methodology? Select all that apply.

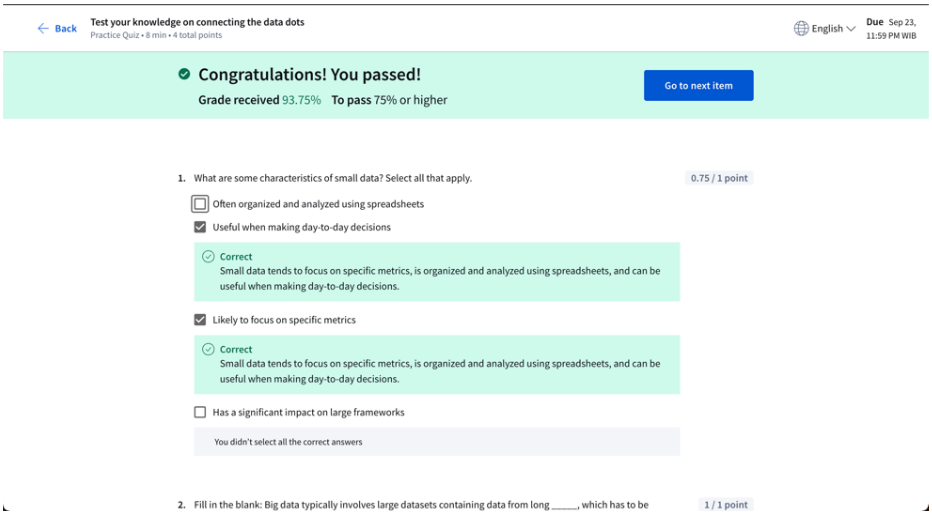
1 / 1 point

☑ SMART stands for specific, measurable, action-oriented, relevant, and time-bound.

Correct

☑ Specific questions are simple, significant, and focused on a single topic or a few closely related ideas.

# Modul 2



**Test your knowledge on connecting the data dots**
Practice Quiz • 8 min • 4 total points

**Due** Sep 23, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 75% or higher

1. What are some characteristics of small data? Select all that apply. 0.75 / 1 point

☐ Often organized and analyzed using spreadsheets

☑ Useful when making day-to-day decisions

Correct
Small data tends to focus on specific metrics, is organized and analyzed using spreadsheets, and can be useful when making day-to-day decisions.

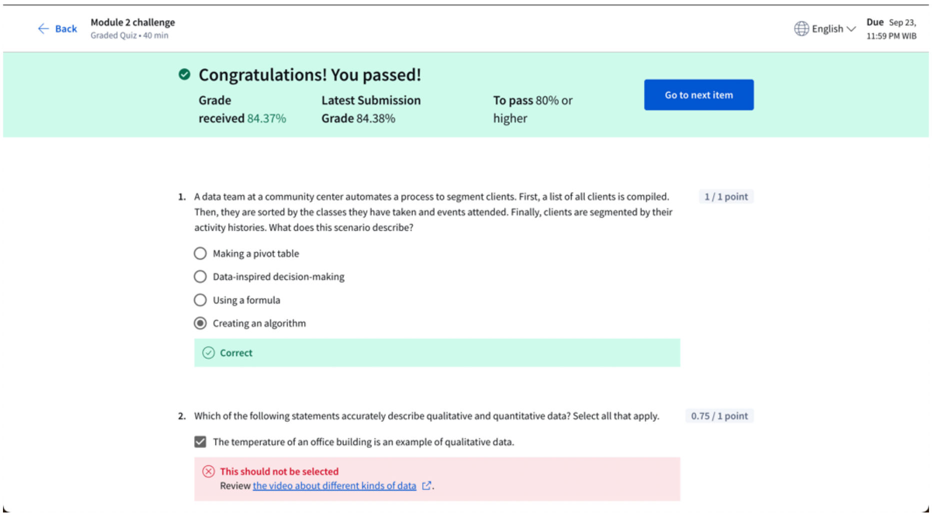
☑ Likely to focus on specific metrics

Correct
Small data tends to focus on specific metrics, is organized and analyzed using spreadsheets, and can be useful when making day-to-day decisions.

☐ Has a significant impact on large frameworks

You didn't select all the correct answers

2. Fill in the blank: Big data typically involves large datasets containing data from long _____, which has to be 1 / 1 point



**Module 2 challenge**
Graded Quiz • 40 min

**Due** Sep 23, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 84.37% **Latest Submission Grade** 84.38% **To pass** 80% or higher

1. A data team at a community center automates a process to segment clients. First, a list of all clients is compiled. Then, they are sorted by the classes they have taken and events attended. Finally, clients are segmented by their activity histories. What does this scenario describe? 1 / 1 point

○ Making a pivot table

○ Data-inspired decision-making

○ Using a formula

● Creating an algorithm

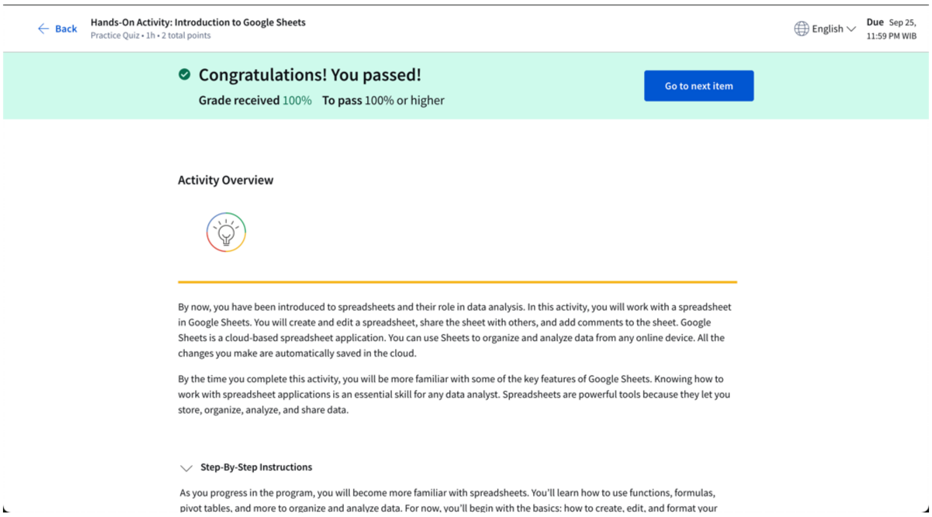
Correct

2. Which of the following statements accurately describe qualitative and quantitative data? Select all that apply. 0.75 / 1 point

☑ The temperature of an office building is an example of qualitative data.

This should not be selected
Review the video about different kinds of data.

# Modul 3



**Hands-On Activity: Introduction to Google Sheets**
Practice Quiz • 1h • 2 total points

**Due** Sep 25, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

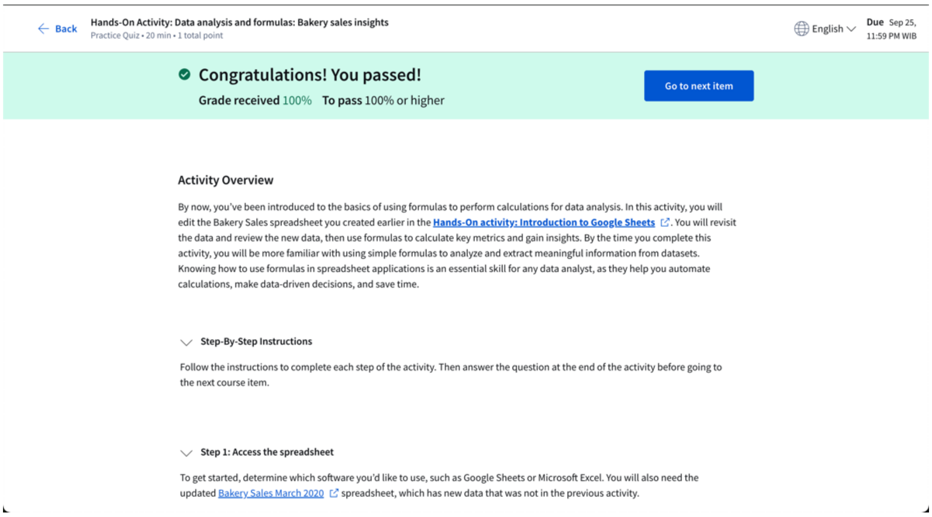
**Activity Overview**

By now, you have been introduced to spreadsheets and their role in data analysis. In this activity, you will work with a spreadsheet in Google Sheets. You will create and edit a spreadsheet, share the sheet with others, and add comments to the sheet. Google Sheets is a cloud-based spreadsheet application. You can use Sheets to organize and analyze data from any online device. All the changes you make are automatically saved in the cloud.

By the time you complete this activity, you will be more familiar with some of the key features of Google Sheets. Knowing how to work with spreadsheet applications is an essential skill for any data analyst. Spreadsheets are powerful tools because they let you store, organize, analyze, and share data.

**Step-By-Step Instructions**

As you progress in the program, you will become more familiar with spreadsheets. You'll learn how to use functions, formulas, pivot tables, and more to organize and analyze data. For now, you'll begin with the basics: how to create, edit, and format your



**Hands-On Activity: Data analysis and formulas: Bakery sales insights**

Practice Quiz • 20 min • 1 total point

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

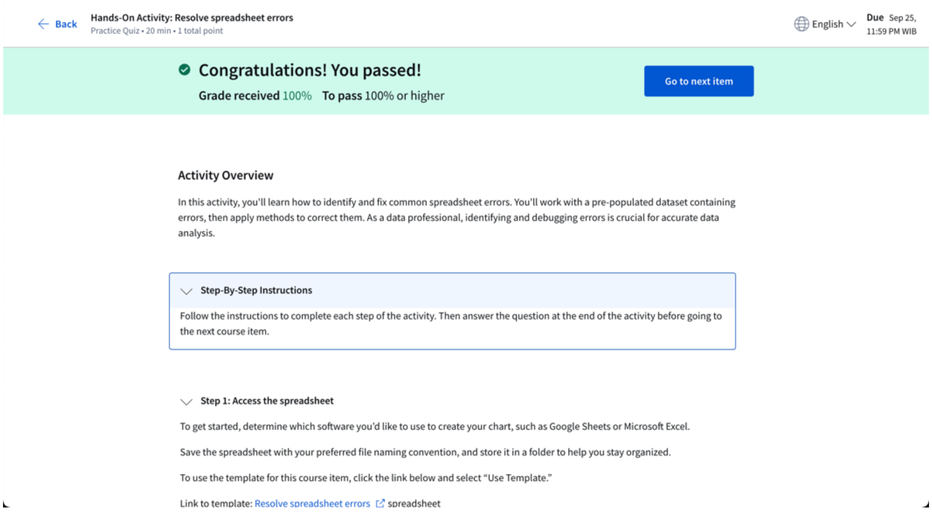
By now, you've been introduced to the basics of using formulas to perform calculations for data analysis. In this activity, you will edit the Bakery Sales spreadsheet you created earlier in the **Hands-On activity: Introduction to Google Sheets**. You will revisit the data and review the new data, then use formulas to calculate key metrics and gain insights. By the time you complete this activity, you will be more familiar with using simple formulas to analyze and extract meaningful information from datasets. Knowing how to use formulas in spreadsheet applications is an essential skill for any data analyst, as they help you automate calculations, make data-driven decisions, and save time.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to the next course item.

**Step 1: Access the spreadsheet**

To get started, determine which software you'd like to use, such as Google Sheets or Microsoft Excel. You will also need the updated Bakery Sales March 2020 spreadsheet, which has new data that was not in the previous activity.



**Hands-On Activity: Resolve spreadsheet errors**

Practice Quiz • 20 min • 1 total point

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

In this activity, you'll learn how to identify and fix common spreadsheet errors. You'll work with a pre-populated dataset containing errors, then apply methods to correct them. As a data professional, identifying and debugging errors is crucial for accurate data analysis.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to the next course item.

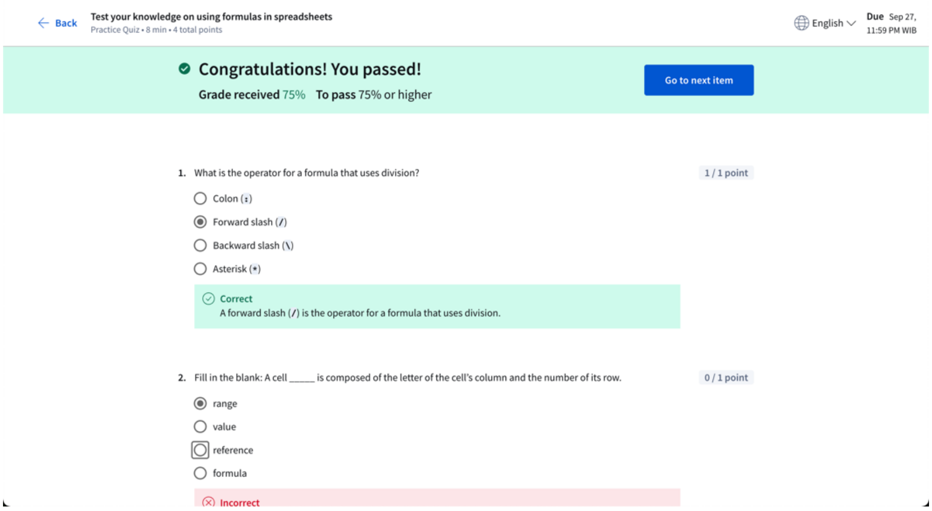
**Step 1: Access the spreadsheet**

To get started, determine which software you'd like to use to create your chart, such as Google Sheets or Microsoft Excel.

Save the spreadsheet with your preferred file naming convention, and store it in a folder to help you stay organized.

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

Link to template: Resolve spreadsheet errors spreadsheet



**Test your knowledge on using formulas in spreadsheets**

Practice Quiz • 8 min • 4 total points

**Congratulations! You passed!**

**Grade received** 75% **To pass** 75% or higher

1. What is the operator for a formula that uses division? 1 / 1 point

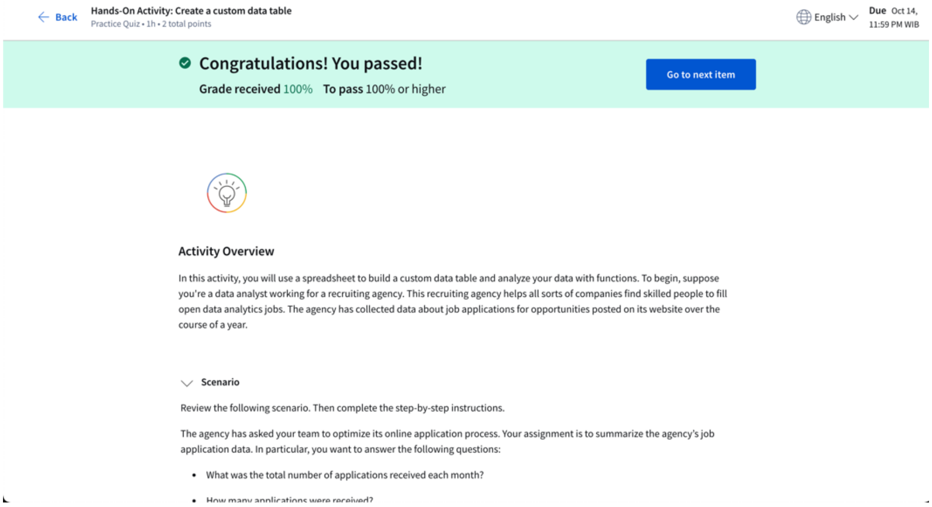
○ Colon (:)
● Forward slash (/)
○ Backward slash (\)
○ Asterisk (*)

✓ Correct
A forward slash (/) is the operator for a formula that uses division.

2. Fill in the blank: A cell _____ is composed of the letter of the cell's column and the number of its row. 0 / 1 point

● range
○ value
○ reference
○ formula

⊗ Incorrect



**Hands-On Activity: Create a custom data table**
Practice Quiz • 1h • 2 total points

**Due** Oct 14, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

In this activity, you will use a spreadsheet to build a custom data table and analyze your data with functions. To begin, suppose you're a data analyst working for a recruiting agency. This recruiting agency helps all sorts of companies find skilled people to fill open data analytics jobs. The agency has collected data about job applications for opportunities posted on its website over the course of a year.

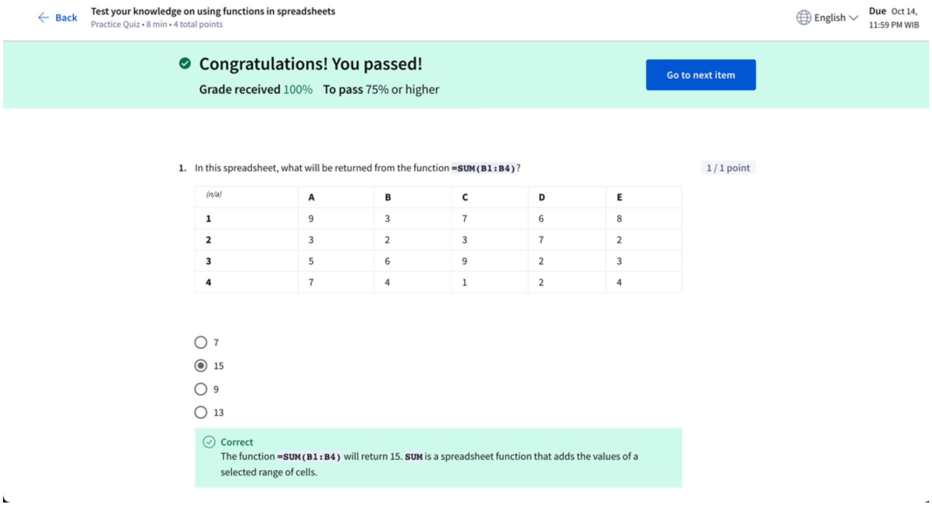
**Scenario**

Review the following scenario. Then complete the step-by-step instructions.

The agency has asked your team to optimize its online application process. Your assignment is to summarize the agency's job application data. In particular, you want to answer the following questions:

- What was the total number of applications received each month?
- How many applications were received?

---



**Test your knowledge on using functions in spreadsheets**
Practice Quiz • 8 min • 4 total points

**Due** Oct 14, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

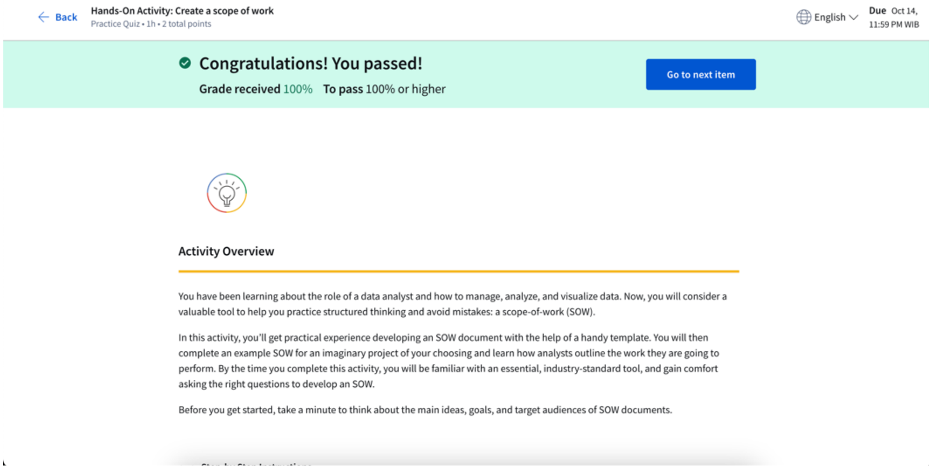
1. In this spreadsheet, what will be returned from the function `=SUM(B1:B4)`? 1 / 1 point

| *(n/a)* | **A** | **B** | **C** | **D** | **E** |
|---|---|---|---|---|---|
| **1** | 9 | 3 | 7 | 6 | 8 |
| **2** | 3 | 2 | 3 | 7 | 2 |
| **3** | 5 | 6 | 9 | 2 | 3 |
| **4** | 7 | 4 | 1 | 2 | 4 |

○ 7
● 15
○ 9
○ 13

✓ Correct
The function `=SUM(B1:B4)` will return 15. `SUM` is a spreadsheet function that adds the values of a selected range of cells.

---



**Hands-On Activity: Create a scope of work**
Practice Quiz • 1h • 2 total points

**Due** Oct 14, 11:59 PM WIB

## Congratulations! You passed!

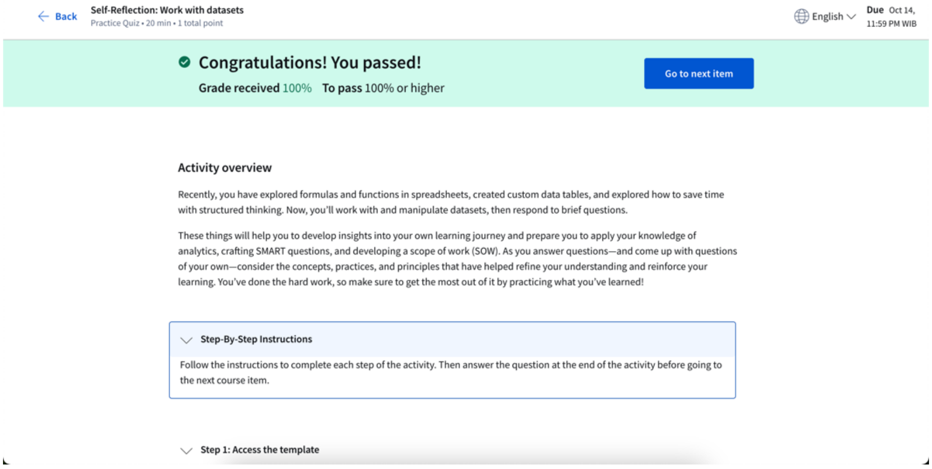
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

You have been learning about the role of a data analyst and how to manage, analyze, and visualize data. Now, you will consider a valuable tool to help you practice structured thinking and avoid mistakes: a scope-of-work (SOW).

In this activity, you'll get practical experience developing an SOW document with the help of a handy template. You will then complete an example SOW for an imaginary project of your choosing and learn how analysts outline the work they are going to perform. By the time you complete this activity, you will be familiar with an essential, industry-standard tool, and gain comfort asking the right questions to develop an SOW.

Before you get started, take a minute to think about the main ideas, goals, and target audiences of SOW documents.



**Self-Reflection: Work with datasets**
Practice Quiz • 20 min • 1 total point

**Due** Oct 14, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

**Activity overview**

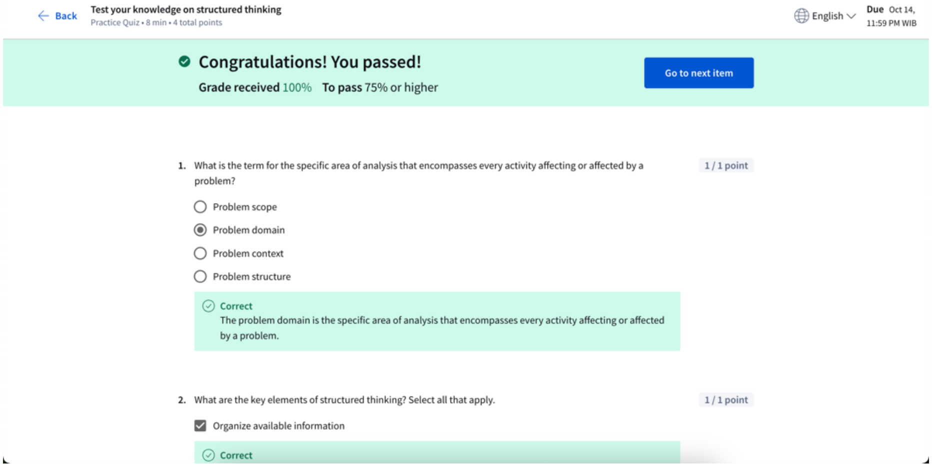
Recently, you have explored formulas and functions in spreadsheets, created custom data tables, and explored how to save time with structured thinking. Now, you'll work with and manipulate datasets, then respond to brief questions.

These things will help you to develop insights into your own learning journey and prepare you to apply your knowledge of analytics, crafting SMART questions, and developing a scope of work (SOW). As you answer questions—and come up with questions of your own—consider the concepts, practices, and principles that have helped refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it by practicing what you've learned!

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to the next course item.

**Step 1: Access the template**



**Test your knowledge on structured thinking**
Practice Quiz • 8 min • 4 total points

**Due** Oct 14, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

1. What is the term for the specific area of analysis that encompasses every activity affecting or affected by a problem?

1 / 1 point

○ Problem scope
● Problem domain
○ Problem context
○ Problem structure

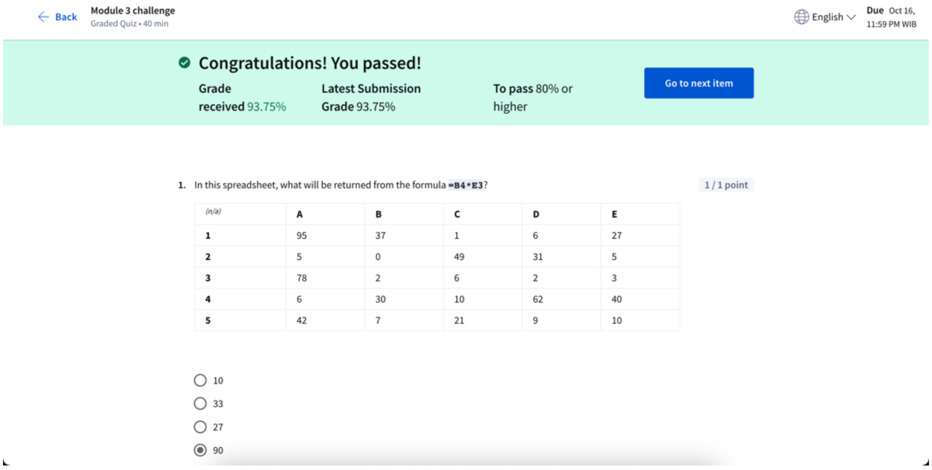
Correct
The problem domain is the specific area of analysis that encompasses every activity affecting or affected by a problem.

2. What are the key elements of structured thinking? Select all that apply.

1 / 1 point

☑ Organize available information

Correct



**Module 3 challenge**
Graded Quiz • 40 min

**Due** Oct 16, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 93.75% **Latest Submission Grade** 93.75% **To pass** 80% or higher

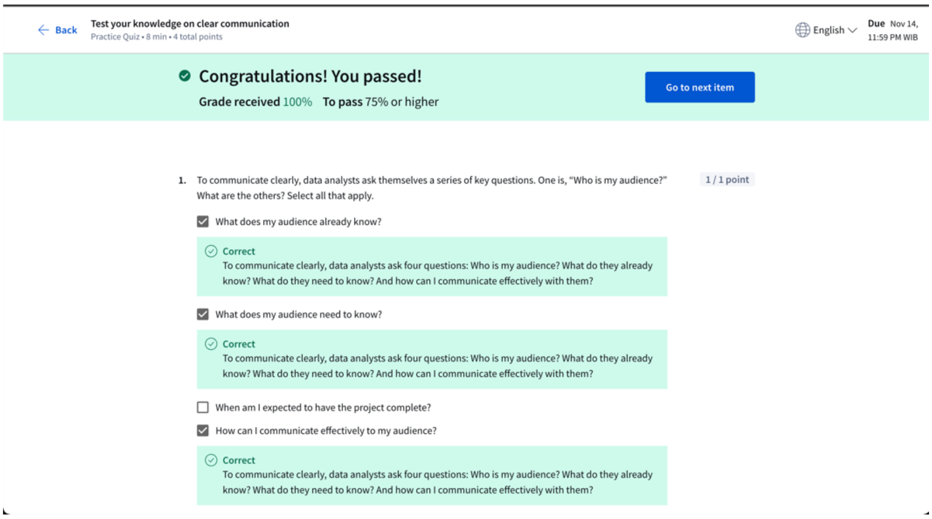
1. In this spreadsheet, what will be returned from the formula `=B4*E3`?

1 / 1 point

| *(n/a)* | A | B | C | D | E |
|---|---|---|---|---|---|
| **1** | 95 | 37 | 1 | 6 | 27 |
| **2** | 5 | 0 | 49 | 31 | 5 |
| **3** | 78 | 2 | 6 | 2 | 3 |
| **4** | 6 | 30 | 10 | 62 | 40 |
| **5** | 42 | 7 | 21 | 9 | 10 |

○ 10
○ 33
○ 27
● 90

# Modul 4



← Back **Test your knowledge on clear communication**
Practice Quiz • 8 min • 4 total points

English

**Due** Nov 14, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. To communicate clearly, data analysts ask themselves a series of key questions. One is, "Who is my audience?" What are the others? Select all that apply.

1 / 1 point

☑ What does my audience already know?

Correct
To communicate clearly, data analysts ask four questions: Who is my audience? What do they already know? What do they need to know? And how can I communicate effectively with them?

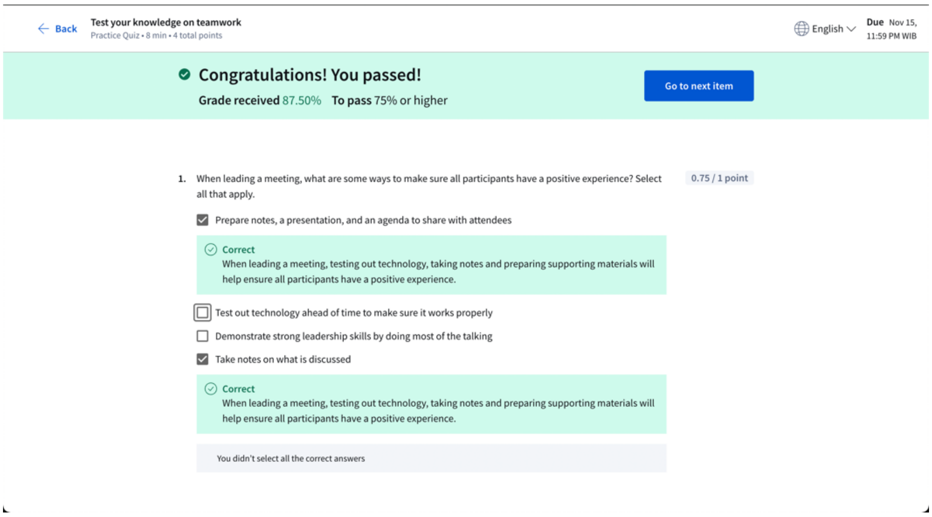
☑ What does my audience need to know?

Correct
To communicate clearly, data analysts ask four questions: Who is my audience? What do they already know? What do they need to know? And how can I communicate effectively with them?

☐ When am I expected to have the project complete?

☑ How can I communicate effectively to my audience?

Correct
To communicate clearly, data analysts ask four questions: Who is my audience? What do they already know? What do they need to know? And how can I communicate effectively with them?



← Back **Test your knowledge on teamwork**
Practice Quiz • 8 min • 4 total points

English

**Due** Nov 15, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 87.50% **To pass** 75% or higher

1. When leading a meeting, what are some ways to make sure all participants have a positive experience? Select all that apply.

0.75 / 1 point

☑ Prepare notes, a presentation, and an agenda to share with attendees

Correct
When leading a meeting, testing out technology, taking notes and preparing supporting materials will help ensure all participants have a positive experience.

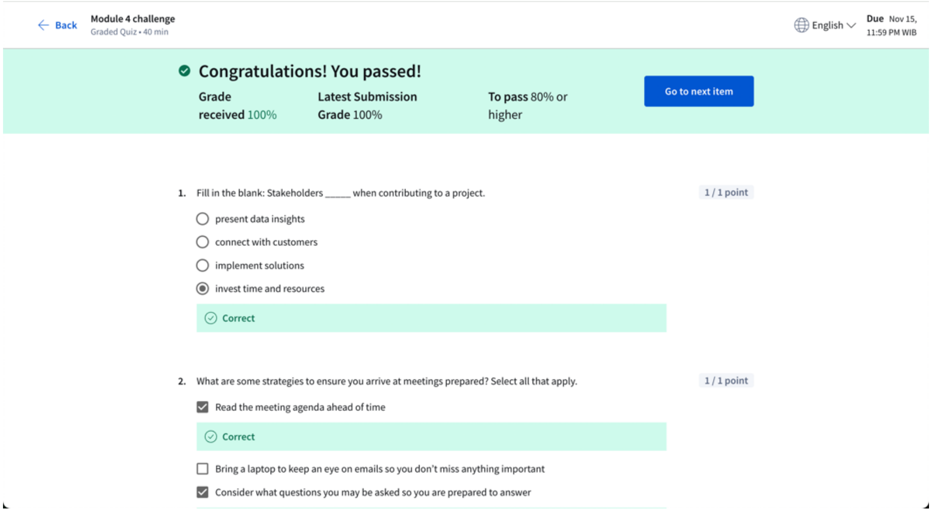
☐ Test out technology ahead of time to make sure it works properly

☐ Demonstrate strong leadership skills by doing most of the talking

☑ Take notes on what is discussed

Correct
When leading a meeting, testing out technology, taking notes and preparing supporting materials will help ensure all participants have a positive experience.

You didn't select all the correct answers



← Back **Module 4 challenge**
Graded Quiz • 40 min

English

**Due** Nov 15, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **Latest Submission Grade** 100% **To pass** 80% or higher

1. Fill in the blank: Stakeholders _____ when contributing to a project.

1 / 1 point

○ present data insights
○ connect with customers
○ implement solutions
● invest time and resources

Correct

2. What are some strategies to ensure you arrive at meetings prepared? Select all that apply.

1 / 1 point

☑ Read the meeting agenda ahead of time

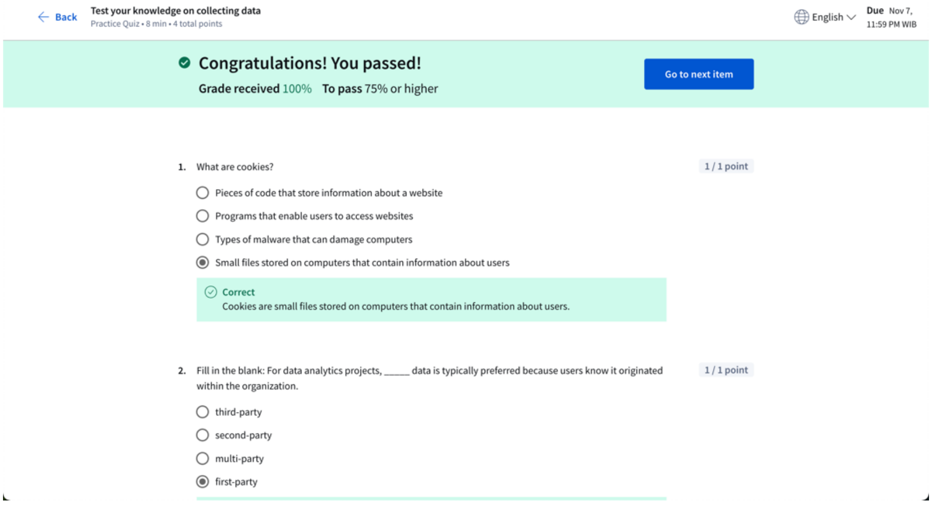
Correct

☐ Bring a laptop to keep an eye on emails so you don't miss anything important

☑ Consider what questions you may be asked so you are prepared to answer

# Course 3: Prepare Data for Exploration

## Modul 1



**Test your knowledge on collecting data**
Practice Quiz • 8 min • 4 total points

**Congratulations! You passed!**
**Grade received** 100% **To pass** 75% or higher

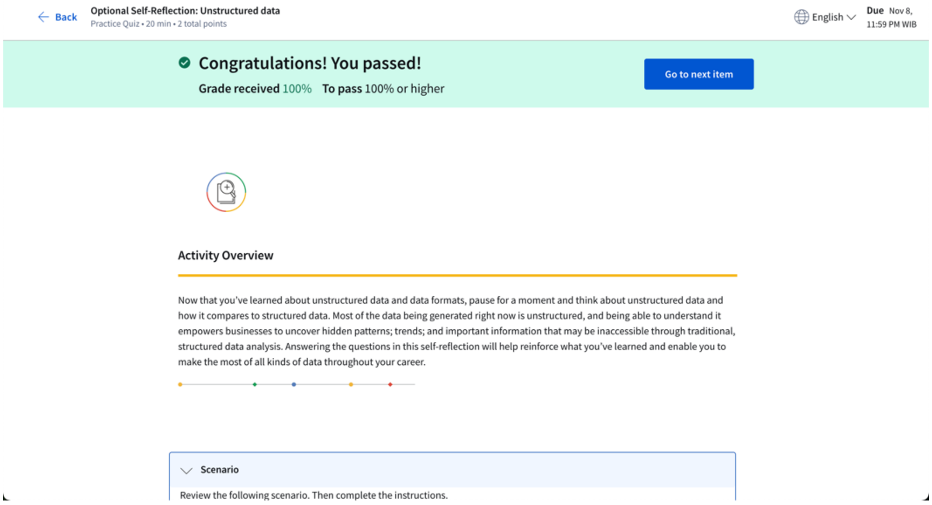
1. What are cookies? 1 / 1 point

○ Pieces of code that store information about a website
○ Programs that enable users to access websites
○ Types of malware that can damage computers
● Small files stored on computers that contain information about users

Correct
Cookies are small files stored on computers that contain information about users.

2. Fill in the blank: For data analytics projects, _____ data is typically preferred because users know it originated within the organization. 1 / 1 point

○ third-party
○ second-party
○ multi-party
● first-party



**Optional Self-Reflection: Unstructured data**
Practice Quiz • 20 min • 2 total points

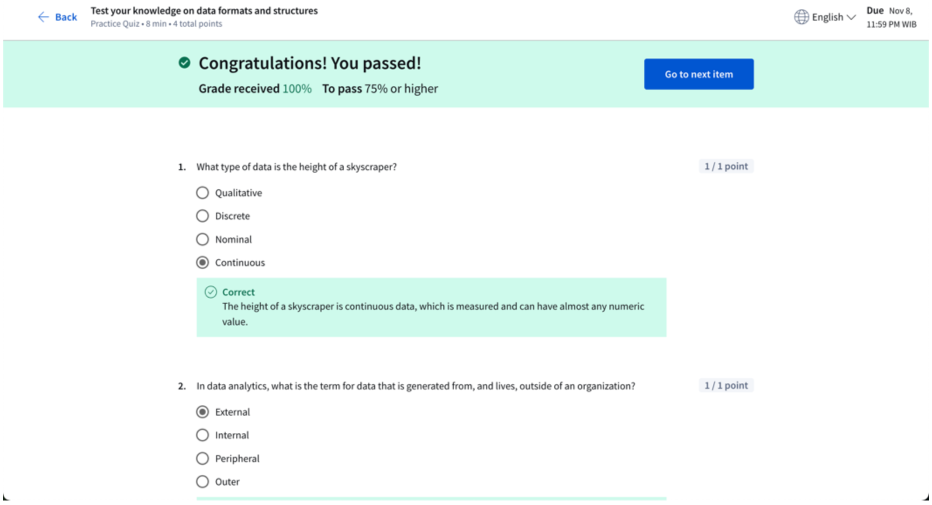
**Congratulations! You passed!**
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

Now that you've learned about unstructured data and data formats, pause for a moment and think about unstructured data and how it compares to structured data. Most of the data being generated right now is unstructured, and being able to understand it empowers businesses to uncover hidden patterns; trends; and important information that may be inaccessible through traditional, structured data analysis. Answering the questions in this self-reflection will help reinforce what you've learned and enable you to make the most of all kinds of data throughout your career.

**Scenario**

Review the following scenario. Then complete the instructions.



**Test your knowledge on data formats and structures**

Practice Quiz • 8 min • 4 total points

**Due** Nov 8, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. What type of data is the height of a skyscraper?

1 / 1 point

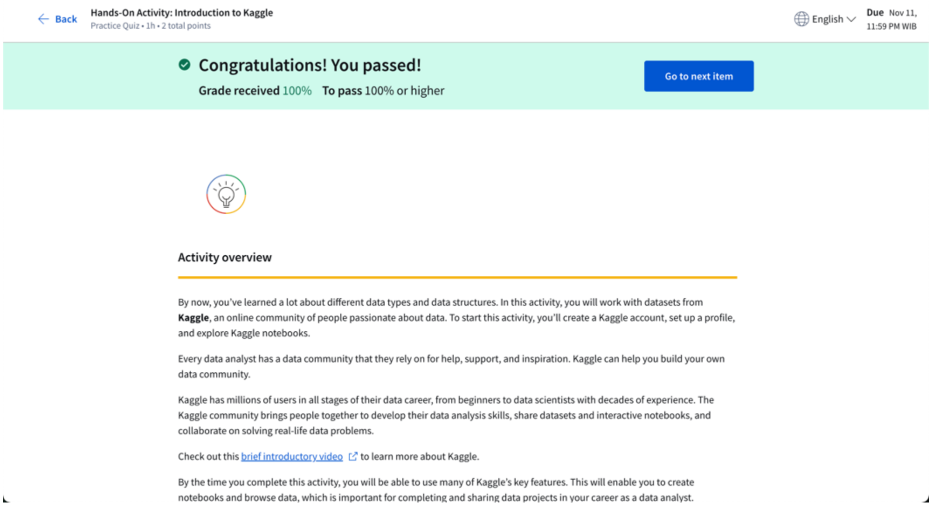
○ Qualitative
○ Discrete
○ Nominal
● Continuous

Correct
The height of a skyscraper is continuous data, which is measured and can have almost any numeric value.

2. In data analytics, what is the term for data that is generated from, and lives, outside of an organization?

1 / 1 point

● External
○ Internal
○ Peripheral
○ Outer



**Hands-On Activity: Introduction to Kaggle**

Practice Quiz • 1h • 2 total points

**Due** Nov 11, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity overview**

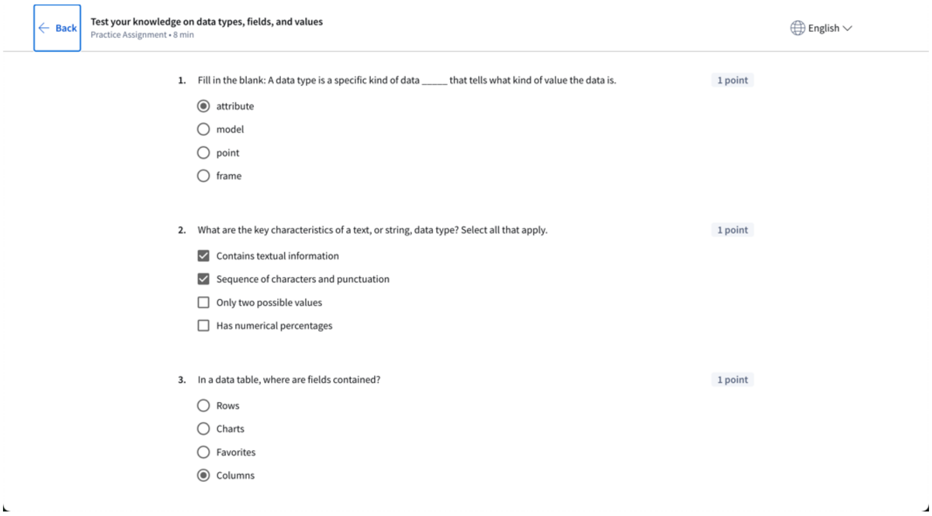
By now, you've learned a lot about different data types and data structures. In this activity, you will work with datasets from **Kaggle**, an online community of people passionate about data. To start this activity, you'll create a Kaggle account, set up a profile, and explore Kaggle notebooks.

Every data analyst has a data community that they rely on for help, support, and inspiration. Kaggle can help you build your own data community.

Kaggle has millions of users in all stages of their data career, from beginners to data scientists with decades of experience. The Kaggle community brings people together to develop their data analysis skills, share datasets and interactive notebooks, and collaborate on solving real-life data problems.

Check out this brief introductory video to learn more about Kaggle.

By the time you complete this activity, you will be able to use many of Kaggle's key features. This will enable you to create notebooks and browse data, which is important for completing and sharing data projects in your career as a data analyst.



**Test your knowledge on data types, fields, and values**

Practice Assignment • 8 min

1. Fill in the blank: A data type is a specific kind of data _____ that tells what kind of value the data is.

1 point

● attribute
○ model
○ point
○ frame

2. What are the key characteristics of a text, or string, data type? Select all that apply.

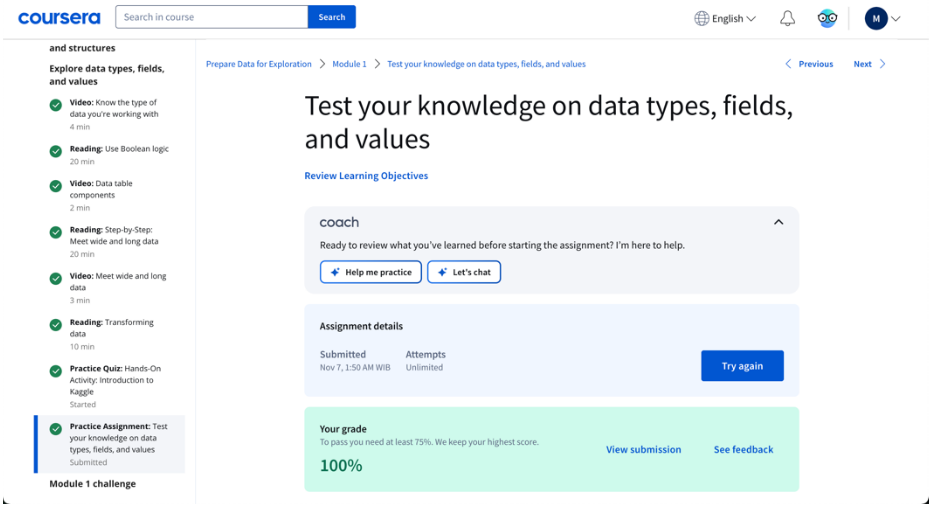
1 point

☑ Contains textual information
☑ Sequence of characters and punctuation
☐ Only two possible values
☐ Has numerical percentages

3. In a data table, where are fields contained?

1 point

○ Rows
○ Charts
○ Favorites
● Columns



and structures

**Explore data types, fields, and values**

- **Video:** Know the type of data you're working with
  4 min
- **Reading:** Use Boolean logic
  20 min
- **Video:** Data table components
  2 min
- **Reading:** Step-by-Step: Meet wide and long data
  20 min
- **Video:** Meet wide and long data
  3 min
- **Reading:** Transforming data
  10 min
- **Practice Quiz:** Hands-On Activity: Introduction to Kaggle
  Started
- **Practice Assignment:** Test your knowledge on data types, fields, and values
  Submitted

**Module 1 challenge**

Prepare Data for Exploration > Module 1 > Test your knowledge on data types, fields, and values

Test your knowledge on data types, fields, and values

Review Learning Objectives

coach

Ready to review what you've learned before starting the assignment? I'm here to help.

**Assignment details**

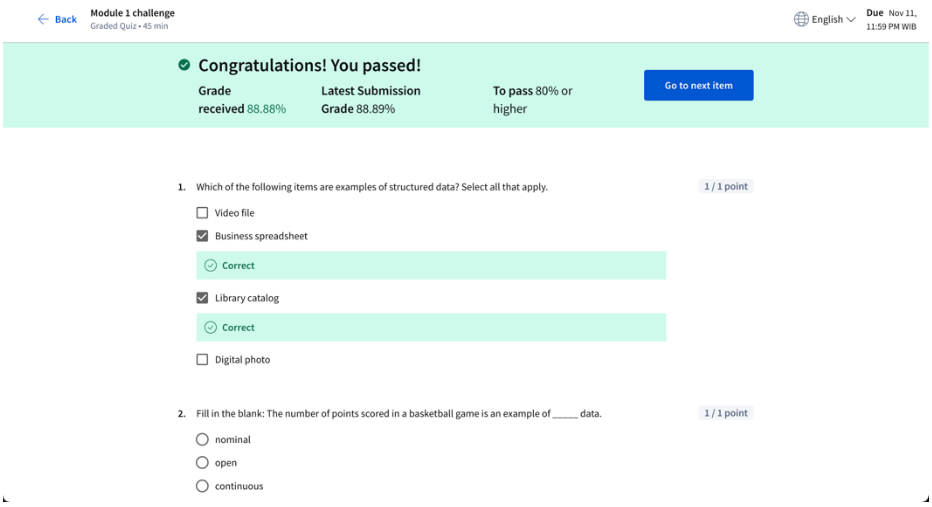
Submitted
Nov 7, 1:50 AM WIB

Attempts
Unlimited

**Your grade**

To pass you need at least 75%. We keep your highest score.

100%



**Module 1 challenge**
Graded Quiz • 45 min

**Due** Nov 11, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 88.88%

**Latest Submission Grade** 88.89%

**To pass** 80% or higher

1. Which of the following items are examples of structured data? Select all that apply. 1 / 1 point

☐ Video file

☑ Business spreadsheet

Correct

☑ Library catalog

Correct

☐ Digital photo

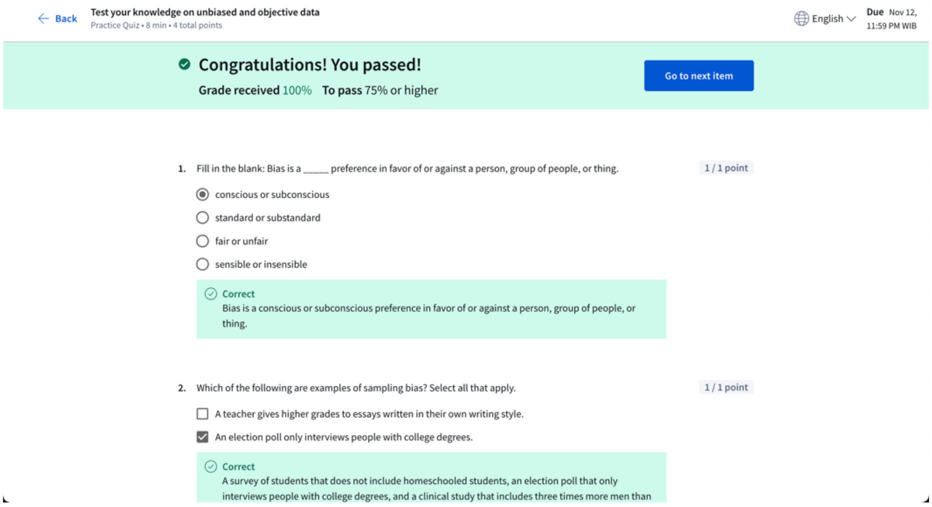
2. Fill in the blank: The number of points scored in a basketball game is an example of _____ data. 1 / 1 point

○ nominal

○ open

○ continuous

# Modul 2



**Test your knowledge on unbiased and objective data**
Practice Quiz • 8 min • 4 total points

**Due** Nov 12, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. Fill in the blank: Bias is a _____ preference in favor of or against a person, group of people, or thing. 1 / 1 point

● conscious or subconscious

○ standard or substandard

○ fair or unfair

○ sensible or insensible

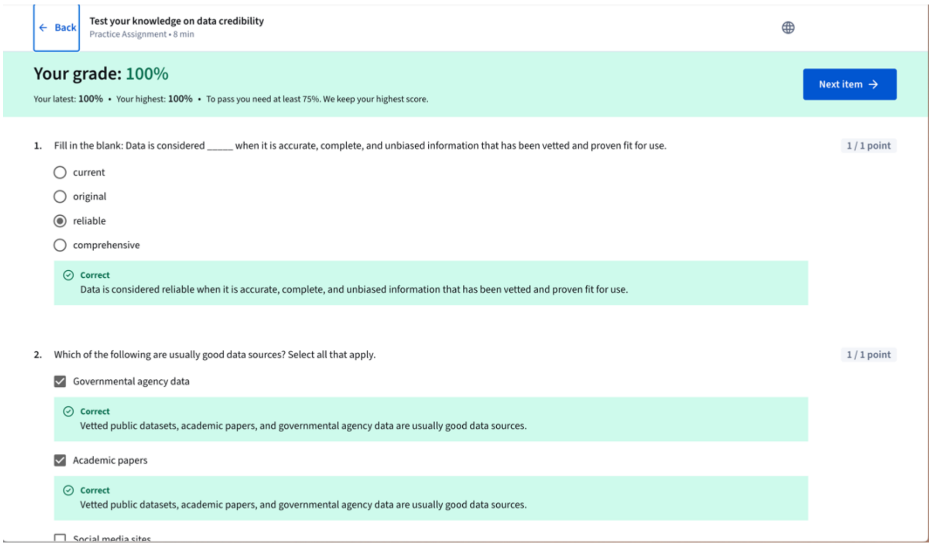
Correct
Bias is a conscious or subconscious preference in favor of or against a person, group of people, or thing.

2. Which of the following are examples of sampling bias? Select all that apply. 1 / 1 point

☐ A teacher gives higher grades to essays written in their own writing style.

☑ An election poll only interviews people with college degrees.

Correct
A survey of students that does not include homeschooled students, an election poll that only interviews people with college degrees, and a clinical study that includes three times more men than



**Test your knowledge on data credibility**

Practice Assignment • 8 min

## Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

1. Fill in the blank: Data is considered _____ when it is accurate, complete, and unbiased information that has been vetted and proven fit for use.

1 / 1 point

○ current

○ original

● reliable

○ comprehensive

**Correct**

Data is considered reliable when it is accurate, complete, and unbiased information that has been vetted and proven fit for use.

2. Which of the following are usually good data sources? Select all that apply.

1 / 1 point

☑ Governmental agency data

**Correct**

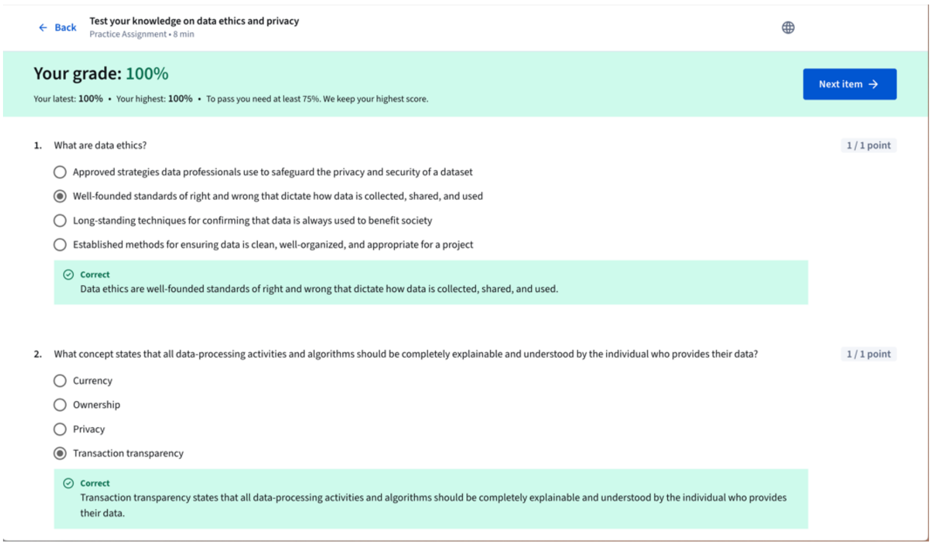
Vetted public datasets, academic papers, and governmental agency data are usually good data sources.

☑ Academic papers

**Correct**

Vetted public datasets, academic papers, and governmental agency data are usually good data sources.

☐ Social media sites



**Test your knowledge on data ethics and privacy**

Practice Assignment • 8 min

## Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

1. What are data ethics?

1 / 1 point

○ Approved strategies data professionals use to safeguard the privacy and security of a dataset

● Well-founded standards of right and wrong that dictate how data is collected, shared, and used

○ Long-standing techniques for confirming that data is always used to benefit society

○ Established methods for ensuring data is clean, well-organized, and appropriate for a project

**Correct**

Data ethics are well-founded standards of right and wrong that dictate how data is collected, shared, and used.

2. What concept states that all data-processing activities and algorithms should be completely explainable and understood by the individual who provides their data?

1 / 1 point

○ Currency

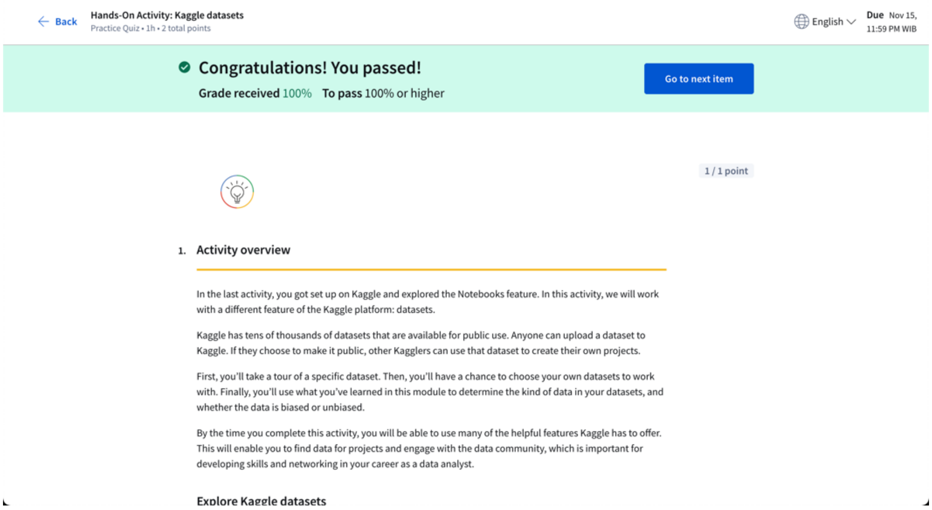
○ Ownership

○ Privacy

● Transaction transparency

**Correct**

Transaction transparency states that all data-processing activities and algorithms should be completely explainable and understood by the individual who provides their data.



**Hands-On Activity: Kaggle datasets**

Practice Quiz • 1h • 2 total points

**Due** Nov 15, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

1. **Activity overview**

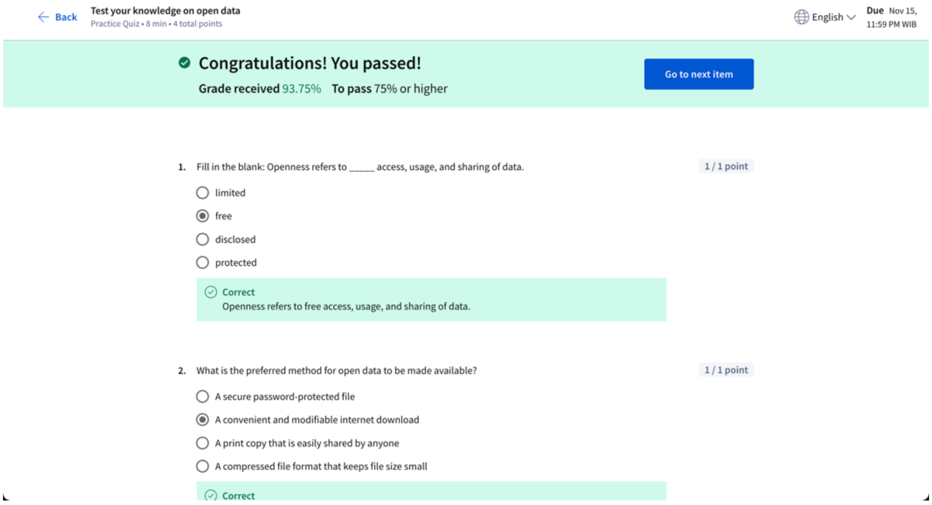
In the last activity, you got set up on Kaggle and explored the Notebooks feature. In this activity, we will work with a different feature of the Kaggle platform: datasets.

Kaggle has tens of thousands of datasets that are available for public use. Anyone can upload a dataset to Kaggle. If they choose to make it public, other Kagglers can use that dataset to create their own projects.

First, you'll take a tour of a specific dataset. Then, you'll have a chance to choose your own datasets to work with. Finally, you'll use what you've learned in this module to determine the kind of data in your datasets, and whether the data is biased or unbiased.

By the time you complete this activity, you will be able to use many of the helpful features Kaggle has to offer. This will enable you to find data for projects and engage with the data community, which is important for developing skills and networking in your career as a data analyst.

**Explore Kaggle datasets**



**Test your knowledge on open data**

Practice Quiz • 8 min • 4 total points

**Due** Nov 15, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 75% or higher

1. Fill in the blank: Openness refers to _____ access, usage, and sharing of data. 1 / 1 point

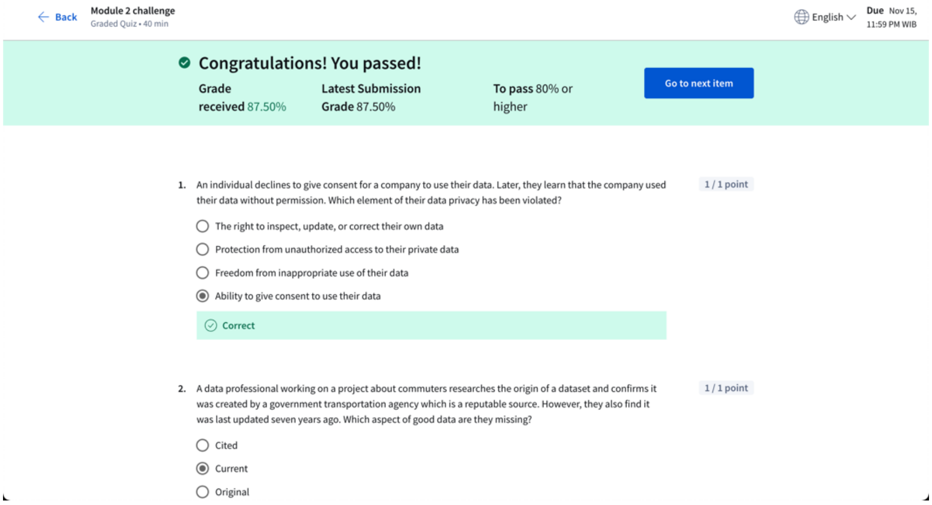
○ limited
● free
○ disclosed
○ protected

Correct
Openness refers to free access, usage, and sharing of data.

2. What is the preferred method for open data to be made available? 1 / 1 point

○ A secure password-protected file
● A convenient and modifiable internet download
○ A print copy that is easily shared by anyone
○ A compressed file format that keeps file size small

Correct



**Module 2 challenge**

Graded Quiz • 40 min

**Due** Nov 15, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 87.50% **Latest Submission Grade** 87.50% **To pass** 80% or higher

1. An individual declines to give consent for a company to use their data. Later, they learn that the company used their data without permission. Which element of their data privacy has been violated? 1 / 1 point

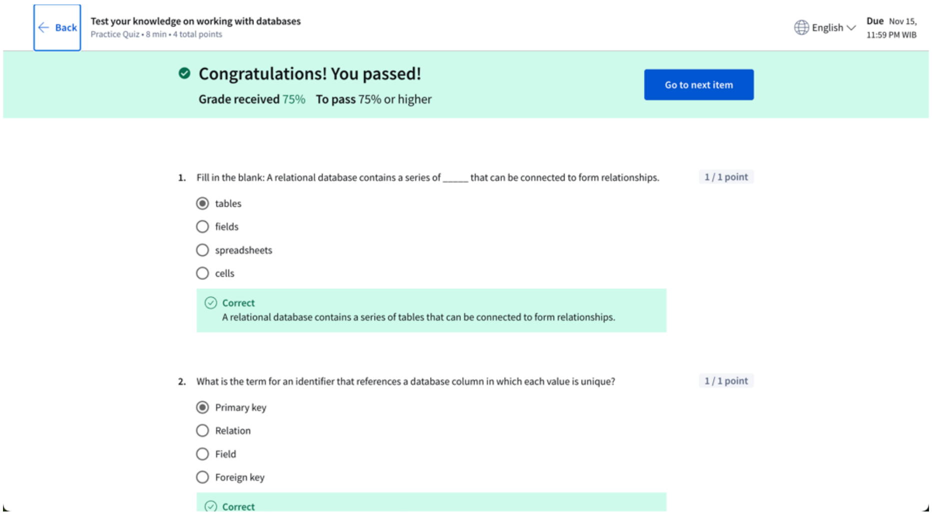
○ The right to inspect, update, or correct their own data
○ Protection from unauthorized access to their private data
○ Freedom from inappropriate use of their data
● Ability to give consent to use their data

Correct

2. A data professional working on a project about commuters researches the origin of a dataset and confirms it was created by a government transportation agency which is a reputable source. However, they also find it was last updated seven years ago. Which aspect of good data are they missing? 1 / 1 point

○ Cited
● Current
○ Original

# Modul 3



**Test your knowledge on working with databases**

Practice Quiz • 8 min • 4 total points

**Due** Nov 15, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 75% **To pass** 75% or higher

1. Fill in the blank: A relational database contains a series of _____ that can be connected to form relationships. 1 / 1 point

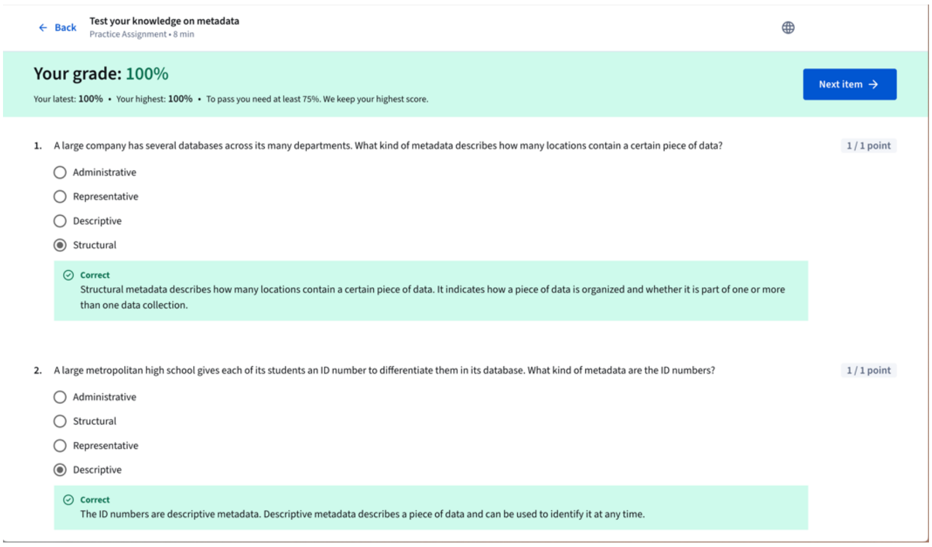
● tables
○ fields
○ spreadsheets
○ cells

Correct
A relational database contains a series of tables that can be connected to form relationships.

2. What is the term for an identifier that references a database column in which each value is unique? 1 / 1 point

● Primary key
○ Relation
○ Field
○ Foreign key

Correct



**Test your knowledge on metadata**
Practice Assignment • 8 min

## Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

1. A large company has several databases across its many departments. What kind of metadata describes how many locations contain a certain piece of data? — 1 / 1 point

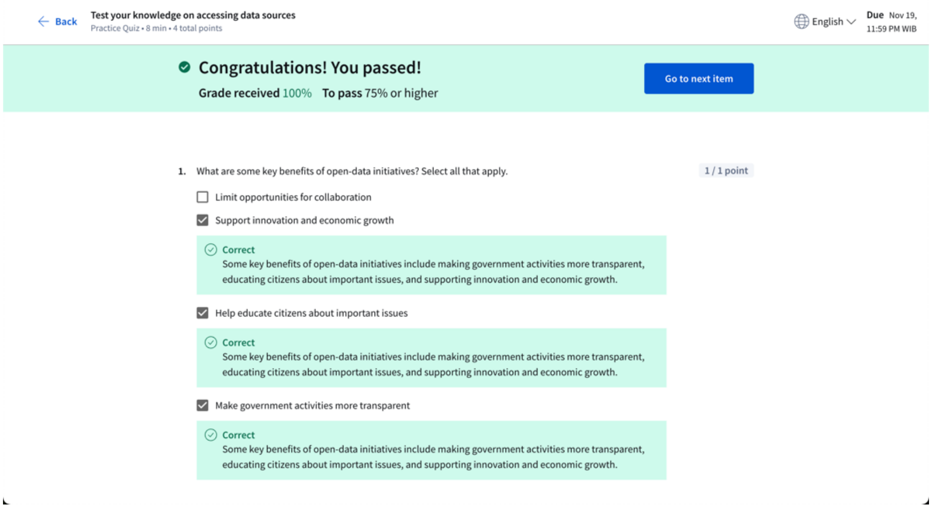
   ○ Administrative
   ○ Representative
   ○ Descriptive
   ● Structural

   ✓ **Correct**
   Structural metadata describes how many locations contain a certain piece of data. It indicates how a piece of data is organized and whether it is part of one or more than one data collection.

2. A large metropolitan high school gives each of its students an ID number to differentiate them in its database. What kind of metadata are the ID numbers? — 1 / 1 point

   ○ Administrative
   ○ Structural
   ○ Representative
   ● Descriptive

   ✓ **Correct**
   The ID numbers are descriptive metadata. Descriptive metadata describes a piece of data and can be used to identify it at any time.



**Test your knowledge on accessing data sources**
Practice Quiz • 8 min • 4 total points

**Due** Nov 19, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

1. What are some key benefits of open-data initiatives? Select all that apply. — 1 / 1 point

   ☐ Limit opportunities for collaboration
   ☑ Support innovation and economic growth

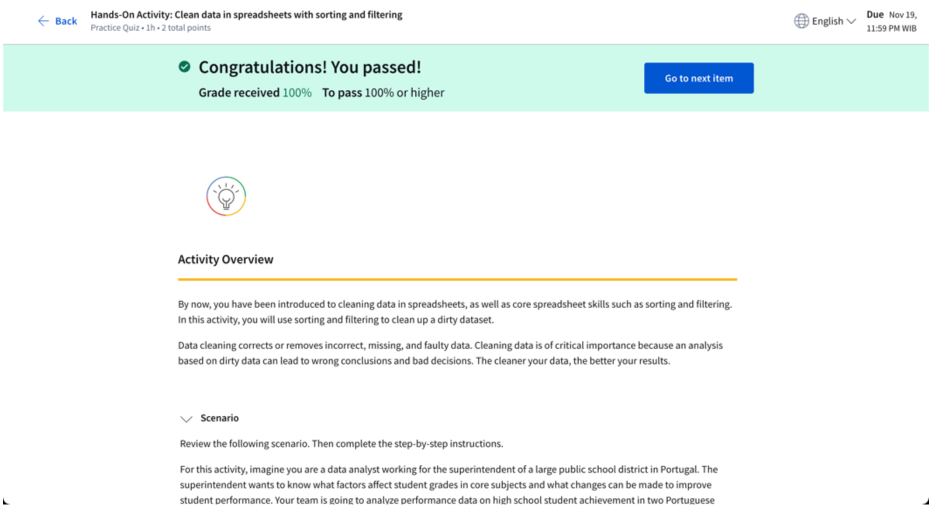
   ✓ Correct
   Some key benefits of open-data initiatives include making government activities more transparent, educating citizens about important issues, and supporting innovation and economic growth.

   ☑ Help educate citizens about important issues

   ✓ Correct
   Some key benefits of open-data initiatives include making government activities more transparent, educating citizens about important issues, and supporting innovation and economic growth.

   ☑ Make government activities more transparent

   ✓ Correct
   Some key benefits of open-data initiatives include making government activities more transparent, educating citizens about important issues, and supporting innovation and economic growth.



**Hands-On Activity: Clean data in spreadsheets with sorting and filtering**
Practice Quiz • 1h • 2 total points

**Due** Nov 19, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

### Activity Overview

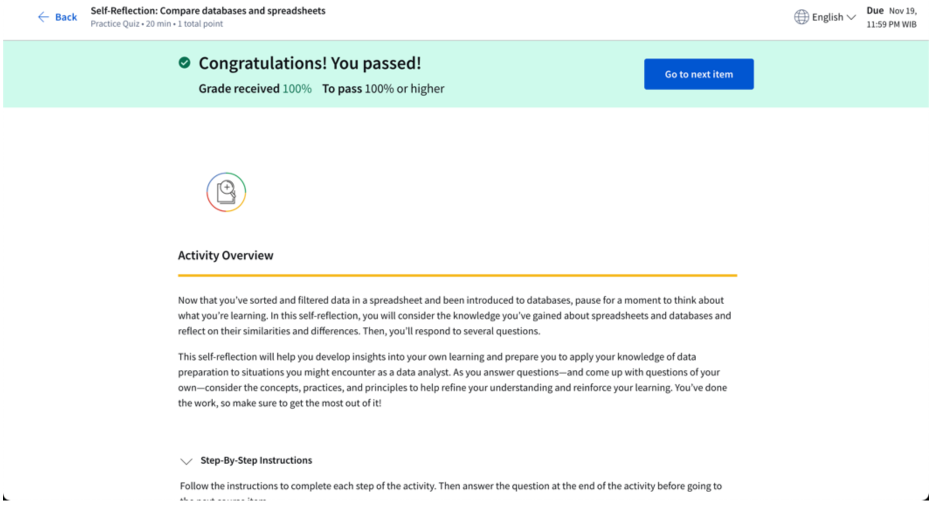
By now, you have been introduced to cleaning data in spreadsheets, as well as core spreadsheet skills such as sorting and filtering. In this activity, you will use sorting and filtering to clean up a dirty dataset.

Data cleaning corrects or removes incorrect, missing, and faulty data. Cleaning data is of critical importance because an analysis based on dirty data can lead to wrong conclusions and bad decisions. The cleaner your data, the better your results.

**Scenario**

Review the following scenario. Then complete the step-by-step instructions.

For this activity, imagine you are a data analyst working for the superintendent of a large public school district in Portugal. The superintendent wants to know what factors affect student grades in core subjects and what changes can be made to improve student performance. Your team is going to analyze performance data on high school student achievement in two Portuguese



**Self-Reflection: Compare databases and spreadsheets**
Practice Quiz • 20 min • 1 total point

**Due** Nov 19, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

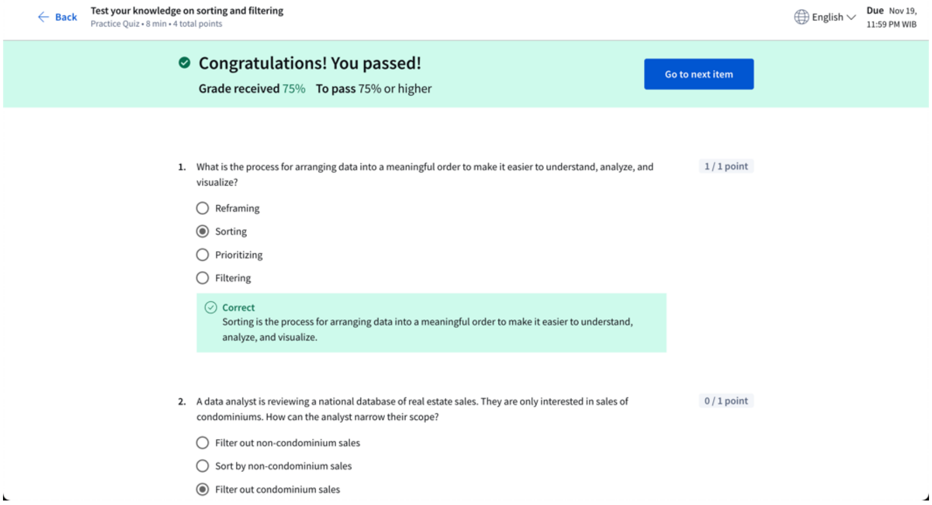
### Activity Overview

Now that you've sorted and filtered data in a spreadsheet and been introduced to databases, pause for a moment to think about what you're learning. In this self-reflection, you will consider the knowledge you've gained about spreadsheets and databases and reflect on their similarities and differences. Then, you'll respond to several questions.

This self-reflection will help you develop insights into your own learning and prepare you to apply your knowledge of data preparation to situations you might encounter as a data analyst. As you answer questions—and come up with questions of your own—consider the concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the work, so make sure to get the most out of it!

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to



**Test your knowledge on sorting and filtering**
Practice Quiz • 8 min • 4 total points

**Due** Nov 19, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 75% **To pass** 75% or higher

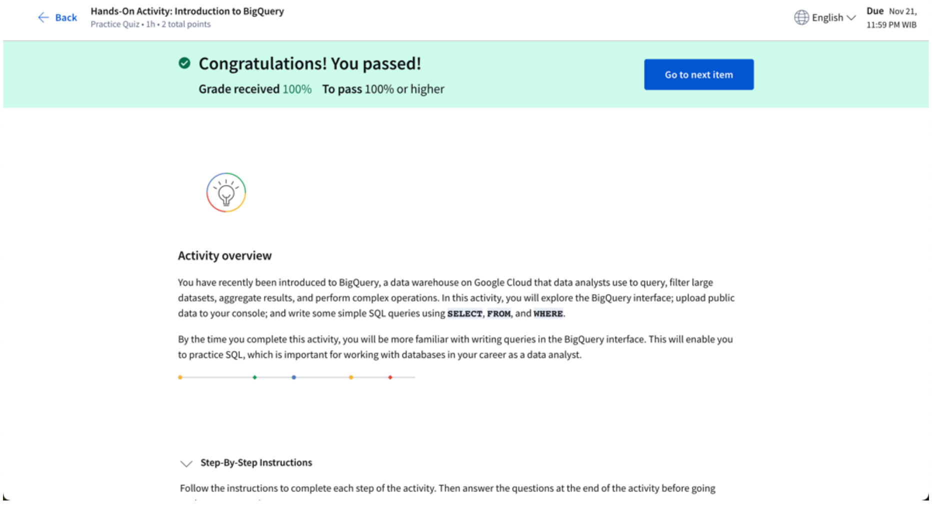
1. What is the process for arranging data into a meaningful order to make it easier to understand, analyze, and visualize? 1 / 1 point

   ○ Reframing
   ● Sorting
   ○ Prioritizing
   ○ Filtering

   Correct
   Sorting is the process for arranging data into a meaningful order to make it easier to understand, analyze, and visualize.

2. A data analyst is reviewing a national database of real estate sales. They are only interested in sales of condominiums. How can the analyst narrow their scope? 0 / 1 point

   ○ Filter out non-condominium sales
   ○ Sort by non-condominium sales
   ● Filter out condominium sales



**Hands-On Activity: Introduction to BigQuery**
Practice Quiz • 1h • 2 total points

**Due** Nov 21, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

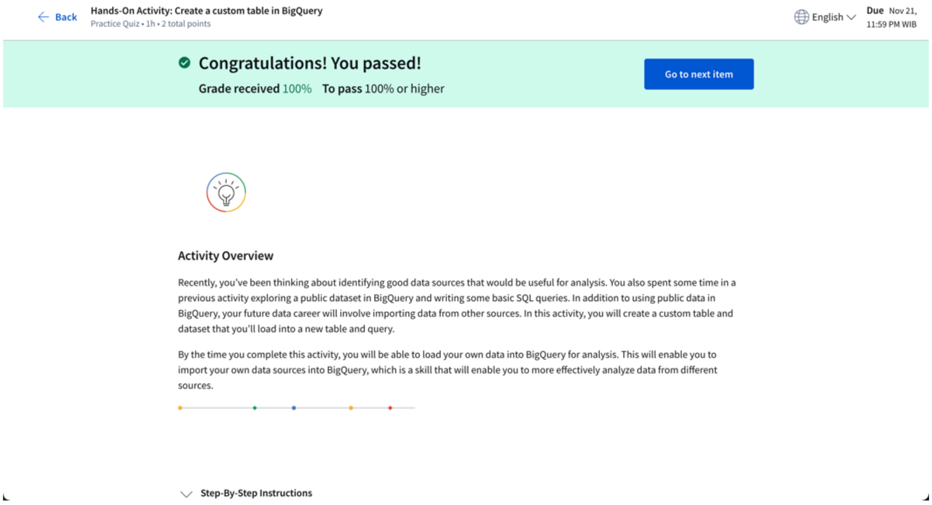
### Activity overview

You have recently been introduced to BigQuery, a data warehouse on Google Cloud that data analysts use to query, filter large datasets, aggregate results, and perform complex operations. In this activity, you will explore the BigQuery interface; upload public data to your console; and write some simple SQL queries using `SELECT`, `FROM`, and `WHERE`.

By the time you complete this activity, you will be more familiar with writing queries in the BigQuery interface. This will enable you to practice SQL, which is important for working with databases in your career as a data analyst.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going



**Hands-On Activity: Create a custom table in BigQuery**
Practice Quiz • 1h • 2 total points

**Due** Nov 21, 11:59 PM WIB

**Congratulations! You passed!**

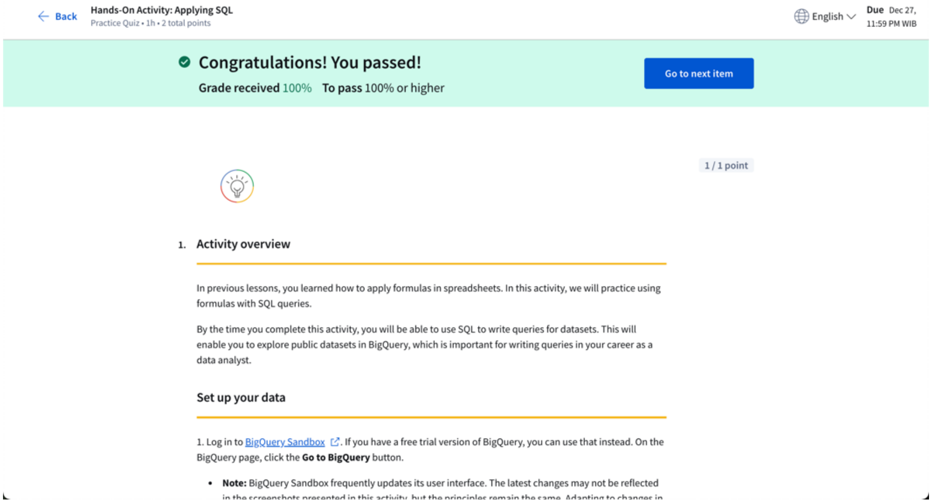
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

Recently, you've been thinking about identifying good data sources that would be useful for analysis. You also spent some time in a previous activity exploring a public dataset in BigQuery and writing some basic SQL queries. In addition to using public data in BigQuery, your future data career will involve importing data from other sources. In this activity, you will create a custom table and dataset that you'll load into a new table and query.

By the time you complete this activity, you will be able to load your own data into BigQuery for analysis. This will enable you to import your own data sources into BigQuery, which is a skill that will enable you to more effectively analyze data from different sources.

**Step-By-Step Instructions**



**Hands-On Activity: Applying SQL**
Practice Quiz • 1h • 2 total points

**Due** Dec 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

1. **Activity overview**

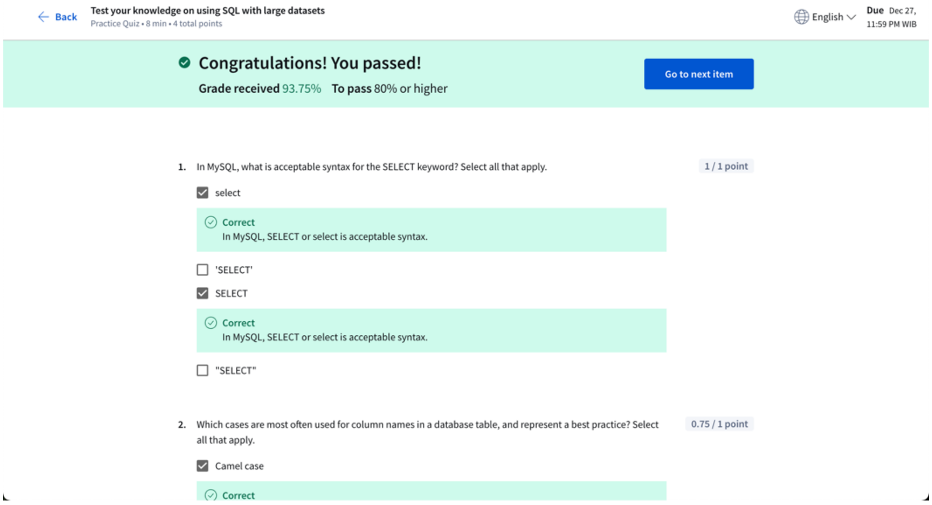
In previous lessons, you learned how to apply formulas in spreadsheets. In this activity, we will practice using formulas with SQL queries.

By the time you complete this activity, you will be able to use SQL to write queries for datasets. This will enable you to explore public datasets in BigQuery, which is important for writing queries in your career as a data analyst.

**Set up your data**

1. Log in to BigQuery Sandbox. If you have a free trial version of BigQuery, you can use that instead. On the BigQuery page, click the **Go to BigQuery** button.

- **Note:** BigQuery Sandbox frequently updates its user interface. The latest changes may not be reflected



**Test your knowledge on using SQL with large datasets**
Practice Quiz • 8 min • 4 total points

**Due** Dec 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 80% or higher

1. In MySQL, what is acceptable syntax for the SELECT keyword? Select all that apply. 1 / 1 point

☑ select

Correct
In MySQL, SELECT or select is acceptable syntax.

☐ 'SELECT'

☑ SELECT

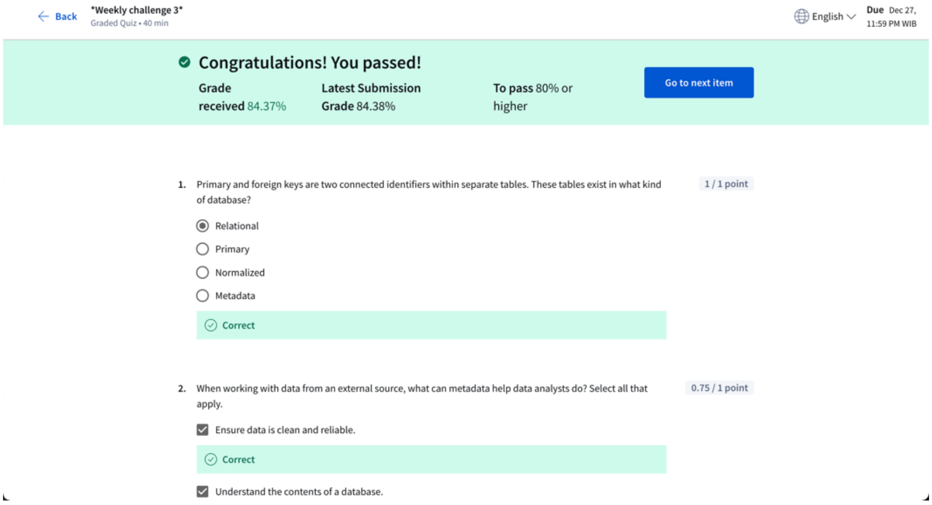
Correct
In MySQL, SELECT or select is acceptable syntax.

☐ "SELECT"

2. Which cases are most often used for column names in a database table, and represent a best practice? Select all that apply. 0.75 / 1 point

☑ Camel case

Correct



← Back **"Weekly challenge 3"**
Graded Quiz • 40 min

English ⌵ **Due** Dec 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 84.37% **Latest Submission Grade** 84.38% **To pass** 80% or higher

1. Primary and foreign keys are two connected identifiers within separate tables. These tables exist in what kind of database? 1 / 1 point

● Relational
○ Primary
○ Normalized
○ Metadata

✓ Correct

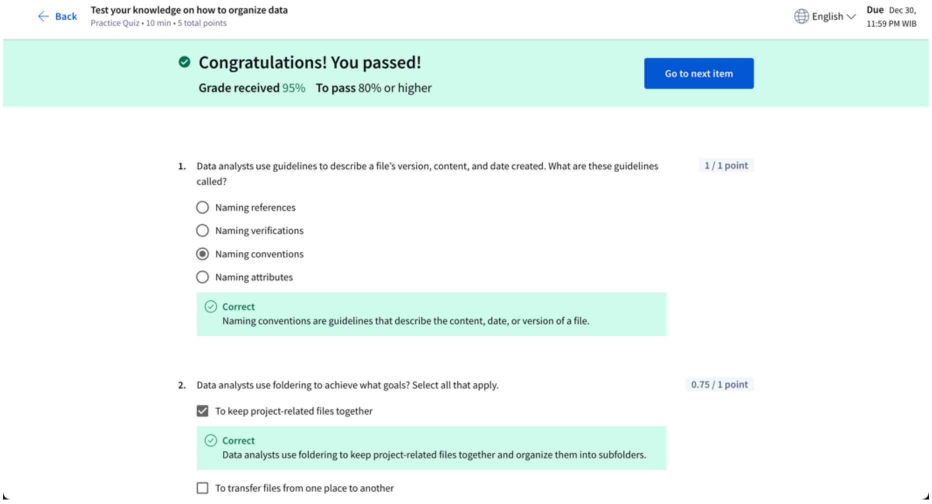
2. When working with data from an external source, what can metadata help data analysts do? Select all that apply. 0.75 / 1 point

☑ Ensure data is clean and reliable.

✓ Correct

☑ Understand the contents of a database.

# Modul 4



← Back **Test your knowledge on how to organize data**
Practice Quiz • 10 min • 5 total points

English ⌵ **Due** Dec 30, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 95% **To pass** 80% or higher

1. Data analysts use guidelines to describe a file's version, content, and date created. What are these guidelines called? 1 / 1 point

○ Naming references
○ Naming verifications
● Naming conventions
○ Naming attributes

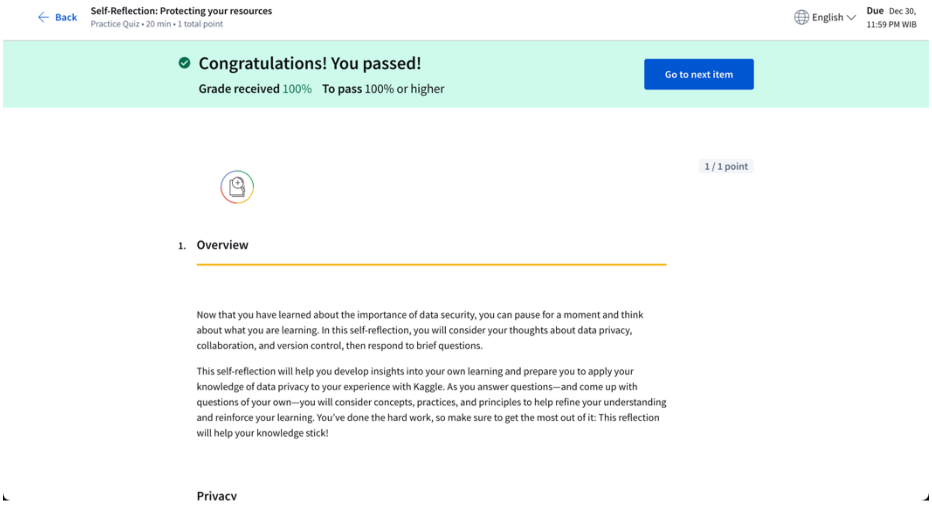
✓ Correct
Naming conventions are guidelines that describe the content, date, or version of a file.

2. Data analysts use foldering to achieve what goals? Select all that apply. 0.75 / 1 point

☑ To keep project-related files together

✓ Correct
Data analysts use foldering to keep project-related files together and organize them into subfolders.

☐ To transfer files from one place to another



← Back **Self-Reflection: Protecting your resources**
Practice Quiz • 20 min • 1 total point

English ⌵ **Due** Dec 30, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

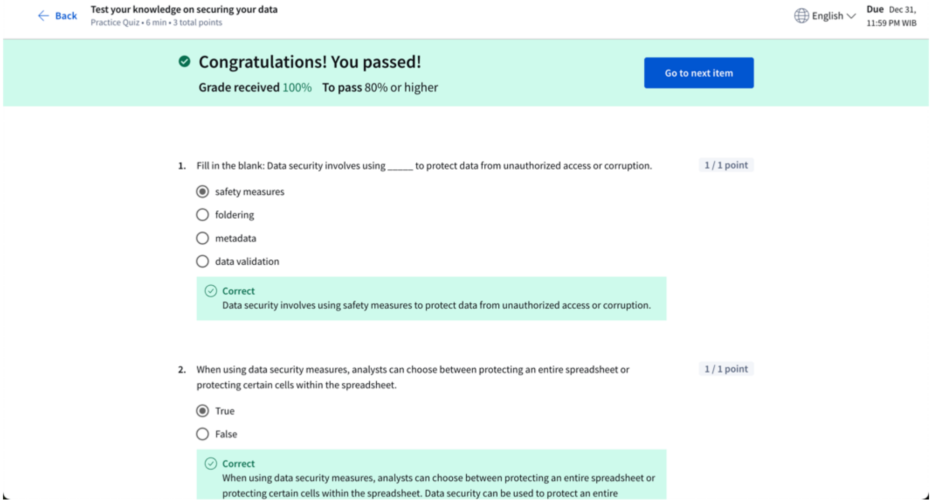
1 / 1 point

1. **Overview**

Now that you have learned about the importance of data security, you can pause for a moment and think about what you are learning. In this self-reflection, you will consider your thoughts about data privacy, collaboration, and version control, then respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to apply your knowledge of data privacy to your experience with Kaggle. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

**Privacy**



**Test your knowledge on securing your data**
Practice Quiz • 6 min • 3 total points

Due Dec 31, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100% **To pass** 80% or higher

1. Fill in the blank: Data security involves using _____ to protect data from unauthorized access or corruption. — 1 / 1 point

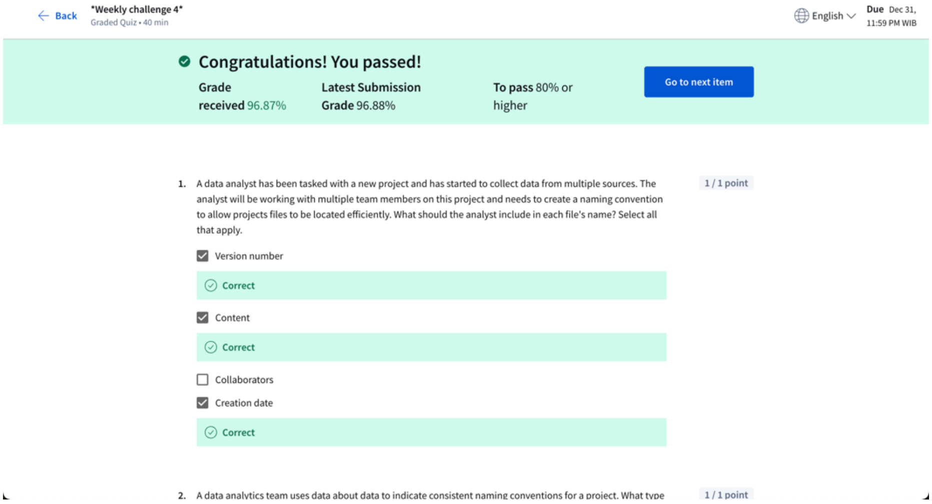
● safety measures
○ foldering
○ metadata
○ data validation

✓ Correct
Data security involves using safety measures to protect data from unauthorized access or corruption.

2. When using data security measures, analysts can choose between protecting an entire spreadsheet or protecting certain cells within the spreadsheet. — 1 / 1 point

● True
○ False

✓ Correct
When using data security measures, analysts can choose between protecting an entire spreadsheet or protecting certain cells within the spreadsheet. Data security can be used to protect an entire



***Weekly challenge 4***
Graded Quiz • 40 min

Due Dec 31, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 96.87% **Latest Submission Grade** 96.88% **To pass** 80% or higher

1. A data analyst has been tasked with a new project and has started to collect data from multiple sources. The analyst will be working with multiple team members on this project and needs to create a naming convention to allow projects files to be located efficiently. What should the analyst include in each file's name? Select all that apply. — 1 / 1 point

☑ Version number

✓ Correct

☑ Content

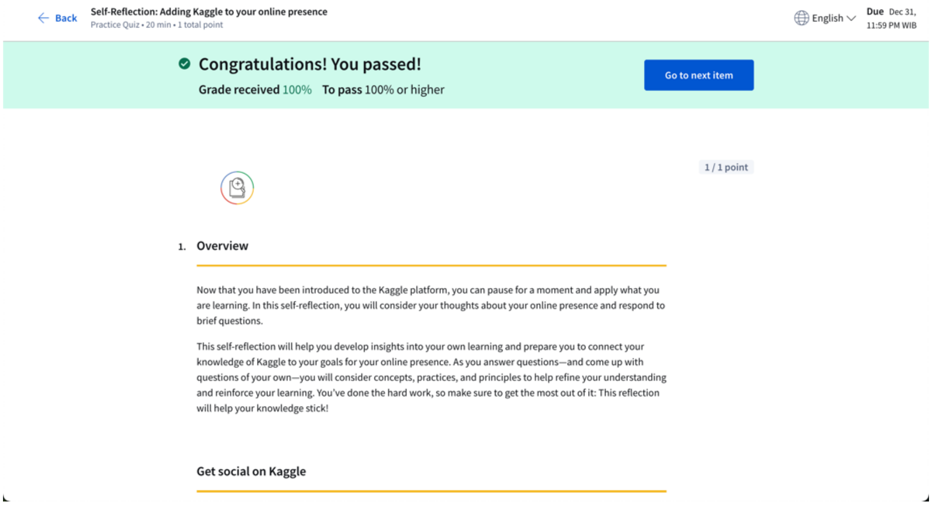
✓ Correct

☐ Collaborators
☑ Creation date

✓ Correct

2. A data analytics team uses data about data to indicate consistent naming conventions for a project. What type — 1 / 1 point

# Modul 5



**Self-Reflection: Adding Kaggle to your online presence**
Practice Quiz • 20 min • 1 total point

Due Dec 31, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100% **To pass** 100% or higher

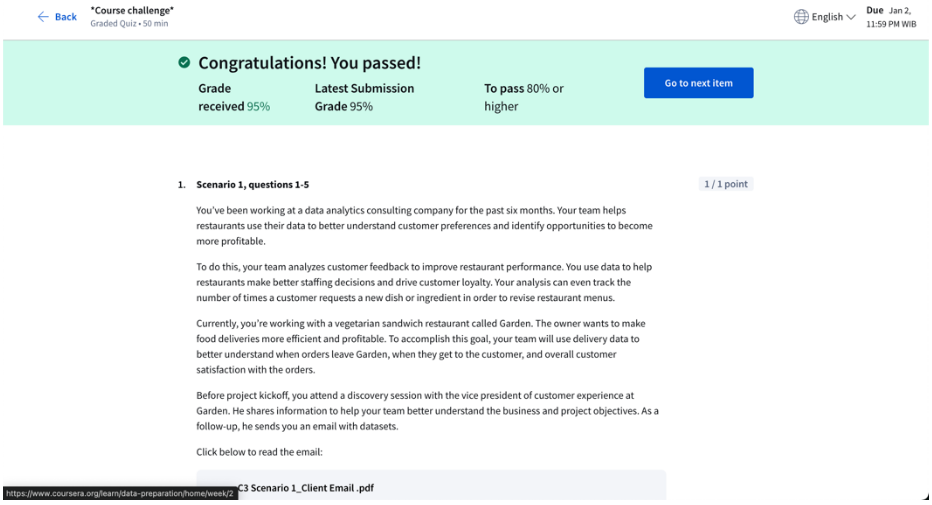
1 / 1 point

1. **Overview**

Now that you have been introduced to the Kaggle platform, you can pause for a moment and apply what you are learning. In this self-reflection, you will consider your thoughts about your online presence and respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to connect your knowledge of Kaggle to your goals for your online presence. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

**Get social on Kaggle**



***Course challenge***

Graded Quiz • 50 min

**Due** Jan 2, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 95%

**Latest Submission Grade** 95%

**To pass** 80% or higher

1. **Scenario 1, questions 1-5**

1 / 1 point

You've been working at a data analytics consulting company for the past six months. Your team helps restaurants use their data to better understand customer preferences and identify opportunities to become more profitable.

To do this, your team analyzes customer feedback to improve restaurant performance. You use data to help restaurants make better staffing decisions and drive customer loyalty. Your analysis can even track the number of times a customer requests a new dish or ingredient in order to revise restaurant menus.

Currently, you're working with a vegetarian sandwich restaurant called Garden. The owner wants to make food deliveries more efficient and profitable. To accomplish this goal, your team will use delivery data to better understand when orders leave Garden, when they get to the customer, and overall customer satisfaction with the orders.

Before project kickoff, you attend a discovery session with the vice president of customer experience at Garden. He shares information to help your team better understand the business and project objectives. As a follow-up, he sends you an email with datasets.

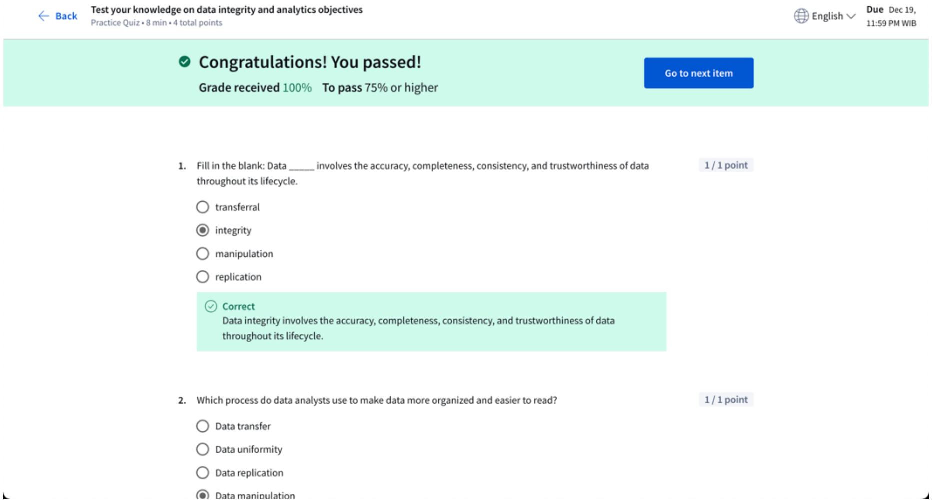
Click below to read the email:

C3 Scenario 1_Client Email .pdf

https://www.coursera.org/learn/data-preparation/home/week/2

# Course 4: *Process Data from Dirty to Clean*

## Modul 1



**Test your knowledge on data integrity and analytics objectives**
Practice Quiz • 8 min • 4 total points

**Due** Dec 19, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100% **To pass** 75% or higher

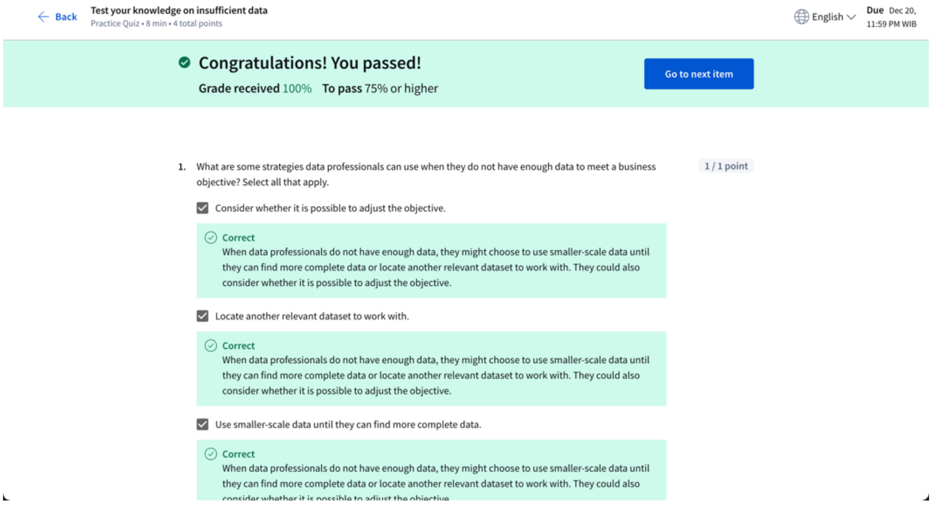
1. Fill in the blank: Data _____ involves the accuracy, completeness, consistency, and trustworthiness of data throughout its lifecycle. 1 / 1 point

○ transferral
● integrity
○ manipulation
○ replication

Correct
Data integrity involves the accuracy, completeness, consistency, and trustworthiness of data throughout its lifecycle.

2. Which process do data analysts use to make data more organized and easier to read? 1 / 1 point

○ Data transfer
○ Data uniformity
○ Data replication
● Data manipulation



**Test your knowledge on insufficient data**
Practice Quiz • 8 min • 4 total points

**Due** Dec 20, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100% **To pass** 75% or higher

1. What are some strategies data professionals can use when they do not have enough data to meet a business objective? Select all that apply. 1 / 1 point

☑ Consider whether it is possible to adjust the objective.

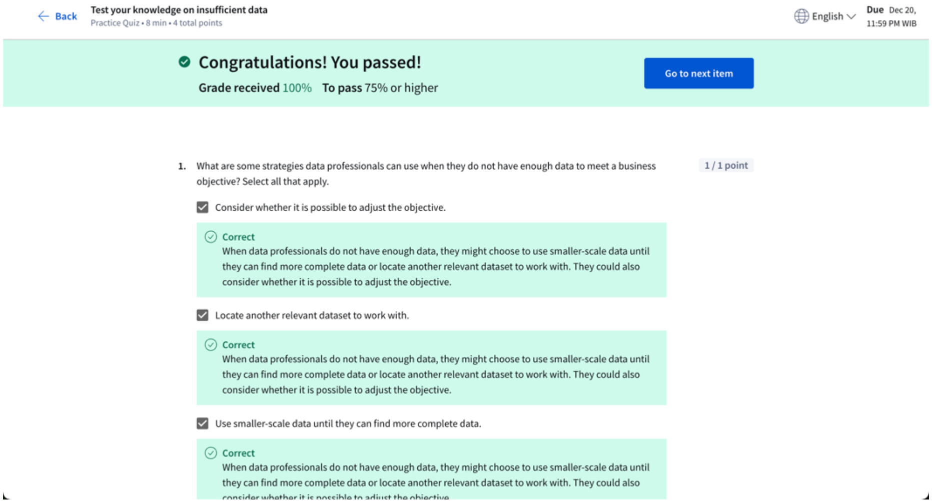
Correct
When data professionals do not have enough data, they might choose to use smaller-scale data until they can find more complete data or locate another relevant dataset to work with. They could also consider whether it is possible to adjust the objective.

☑ Locate another relevant dataset to work with.

Correct
When data professionals do not have enough data, they might choose to use smaller-scale data until they can find more complete data or locate another relevant dataset to work with. They could also consider whether it is possible to adjust the objective.

☑ Use smaller-scale data until they can find more complete data.

Correct
When data professionals do not have enough data, they might choose to use smaller-scale data until they can find more complete data or locate another relevant dataset to work with. They could also consider whether it is possible to adjust the objective.


← Back
English
Go to next item

**Test your knowledge on insufficient data**
Practice Quiz • 8 min • 4 total points

**Due** Dec 20, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100% **To pass** 75% or higher

1. What are some strategies data professionals can use when they do not have enough data to meet a business objective? Select all that apply. 1 / 1 point

☑ Consider whether it is possible to adjust the objective.

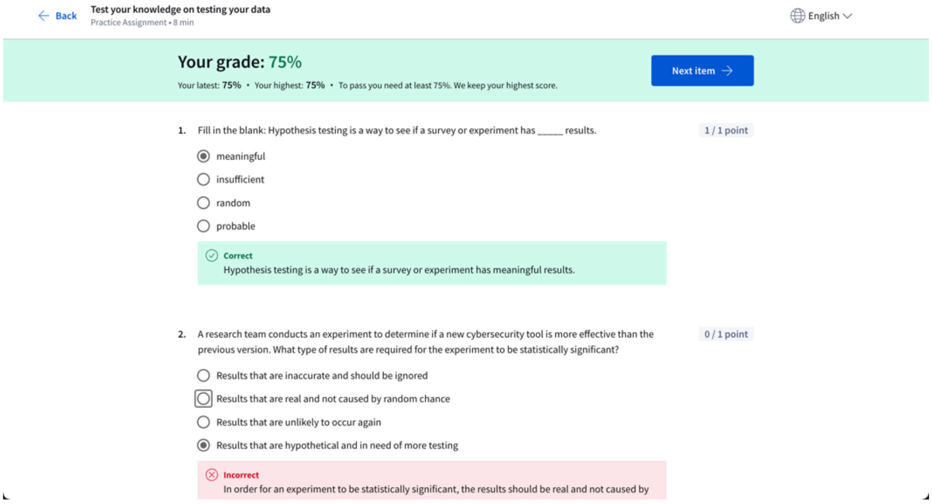
Correct
When data professionals do not have enough data, they might choose to use smaller-scale data until they can find more complete data or locate another relevant dataset to work with. They could also consider whether it is possible to adjust the objective.

☑ Locate another relevant dataset to work with.

Correct
When data professionals do not have enough data, they might choose to use smaller-scale data until they can find more complete data or locate another relevant dataset to work with. They could also consider whether it is possible to adjust the objective.

☑ Use smaller-scale data until they can find more complete data.

Correct
When data professionals do not have enough data, they might choose to use smaller-scale data until they can find more complete data or locate another relevant dataset to work with. They could also consider whether it is possible to adjust the objective.



**Test your knowledge on testing your data**
Practice Assignment • 8 min

**Your grade: 75%**

Your latest: **75%** • Your highest: **75%** • To pass you need at least 75%. We keep your highest score.

1. Fill in the blank: Hypothesis testing is a way to see if a survey or experiment has _____ results. — 1 / 1 point

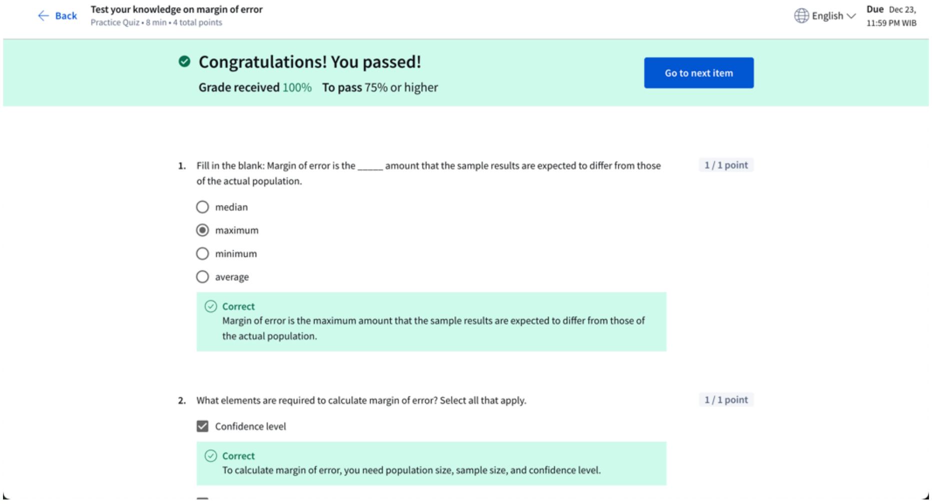
● meaningful
○ insufficient
○ random
○ probable

✓ Correct
Hypothesis testing is a way to see if a survey or experiment has meaningful results.

2. A research team conducts an experiment to determine if a new cybersecurity tool is more effective than the previous version. What type of results are required for the experiment to be statistically significant? — 0 / 1 point

○ Results that are inaccurate and should be ignored
○ Results that are real and not caused by random chance
○ Results that are unlikely to occur again
● Results that are hypothetical and in need of more testing

⊗ Incorrect
In order for an experiment to be statistically significant, the results should be real and not caused by



**Test your knowledge on margin of error**
Practice Quiz • 8 min • 4 total points

**Due** Dec 23, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. Fill in the blank: Margin of error is the _____ amount that the sample results are expected to differ from those of the actual population. — 1 / 1 point

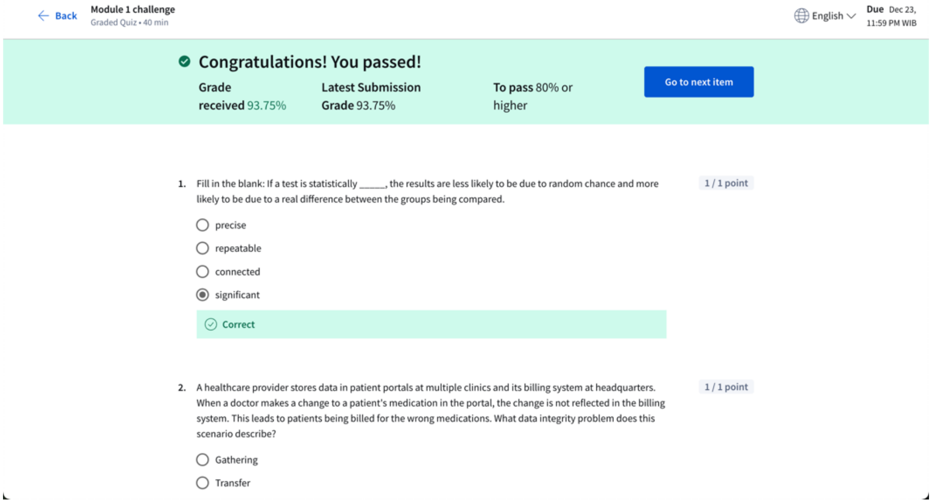
○ median
● maximum
○ minimum
○ average

✓ Correct
Margin of error is the maximum amount that the sample results are expected to differ from those of the actual population.

2. What elements are required to calculate margin of error? Select all that apply. — 1 / 1 point

☑ Confidence level

✓ Correct
To calculate margin of error, you need population size, sample size, and confidence level.



**Module 1 challenge**
Graded Quiz • 40 min

**Due** Dec 23, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **Latest Submission Grade** 93.75% **To pass** 80% or higher

1. Fill in the blank: If a test is statistically _____, the results are less likely to be due to random chance and more likely to be due to a real difference between the groups being compared. — 1 / 1 point

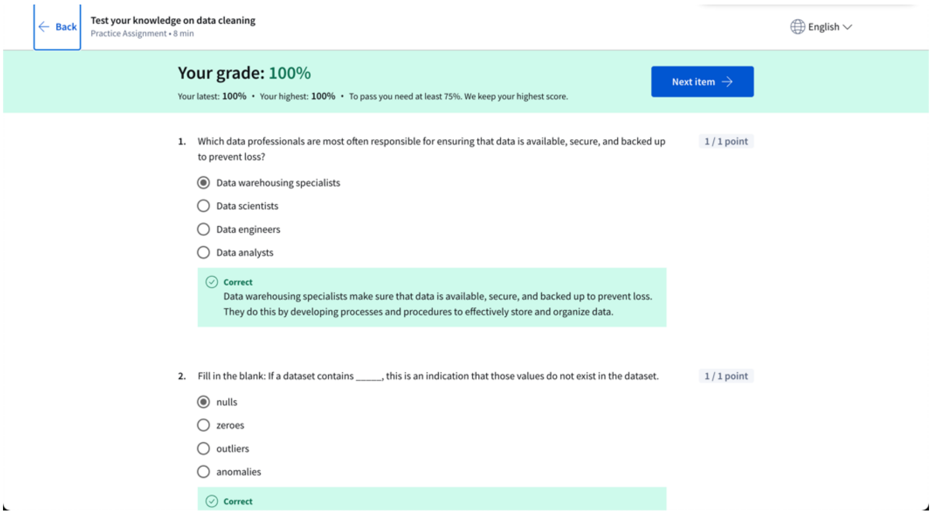
○ precise
○ repeatable
○ connected
● significant

✓ Correct

2. A healthcare provider stores data in patient portals at multiple clinics and its billing system at headquarters. When a doctor makes a change to a patient's medication in the portal, the change is not reflected in the billing system. This leads to patients being billed for the wrong medications. What data integrity problem does this scenario describe? — 1 / 1 point

○ Gathering
○ Transfer

# Modul 2



**Test your knowledge on data cleaning**
Practice Assignment • 8 min

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

1. Which data professionals are most often responsible for ensuring that data is available, secure, and backed up to prevent loss? — 1 / 1 point

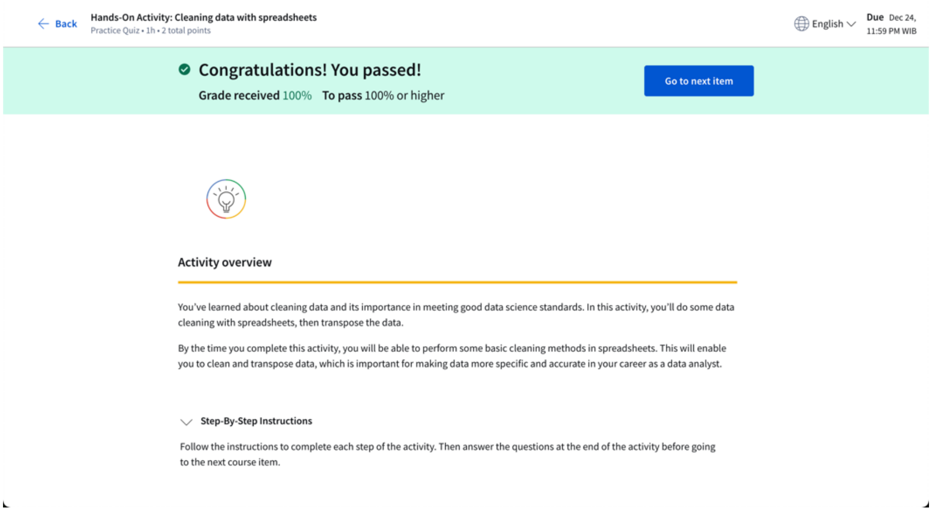
● Data warehousing specialists
○ Data scientists
○ Data engineers
○ Data analysts

✓ **Correct**
Data warehousing specialists make sure that data is available, secure, and backed up to prevent loss. They do this by developing processes and procedures to effectively store and organize data.

2. Fill in the blank: If a dataset contains _____, this is an indication that those values do not exist in the dataset. — 1 / 1 point

● nulls
○ zeroes
○ outliers
○ anomalies

✓ **Correct**



**Hands-On Activity: Cleaning data with spreadsheets**
Practice Quiz • 1h • 2 total points

**Due** Dec 24, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

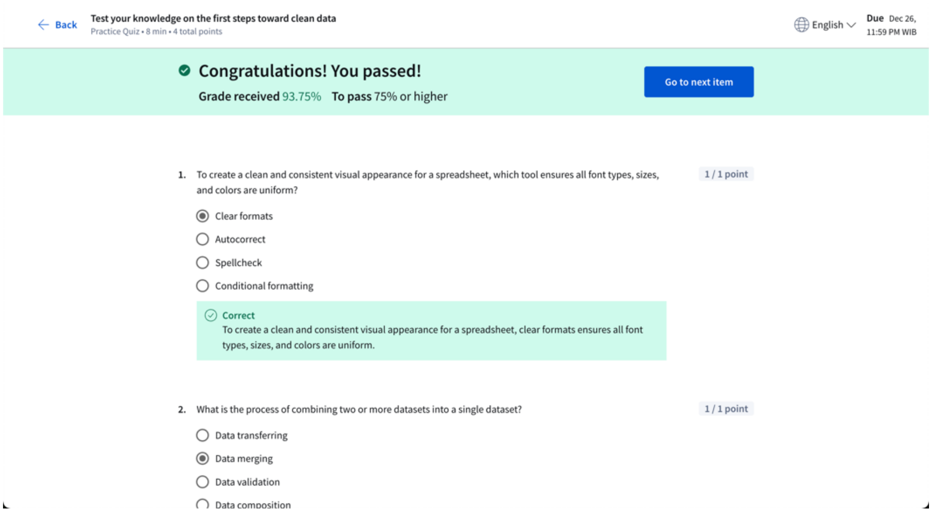
**Activity overview**

You've learned about cleaning data and its importance in meeting good data science standards. In this activity, you'll do some data cleaning with spreadsheets, then transpose the data.

By the time you complete this activity, you will be able to perform some basic cleaning methods in spreadsheets. This will enable you to clean and transpose data, which is important for making data more specific and accurate in your career as a data analyst.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.



**Test your knowledge on the first steps toward clean data**
Practice Quiz • 8 min • 4 total points

**Due** Dec 26, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 75% or higher

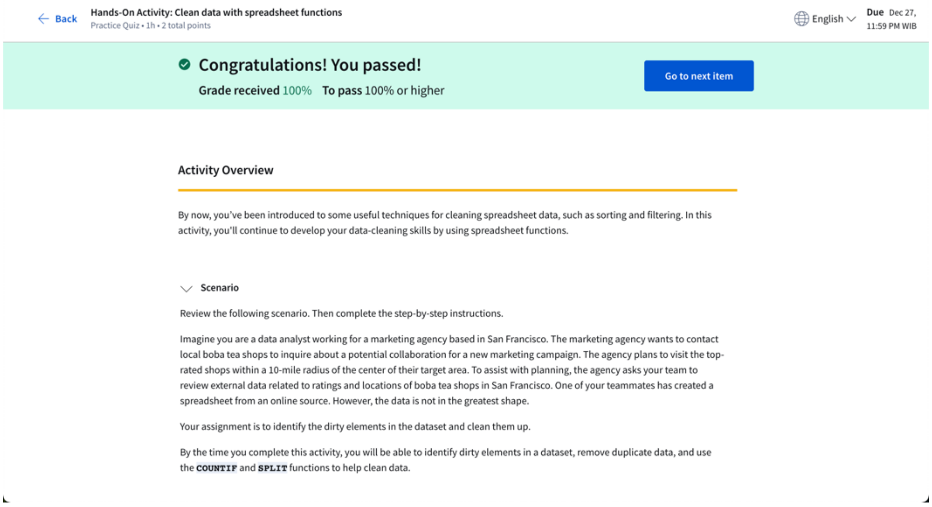
1. To create a clean and consistent visual appearance for a spreadsheet, which tool ensures all font types, sizes, and colors are uniform? — 1 / 1 point

● Clear formats
○ Autocorrect
○ Spellcheck
○ Conditional formatting

✓ **Correct**
To create a clean and consistent visual appearance for a spreadsheet, clear formats ensures all font types, sizes, and colors are uniform.

2. What is the process of combining two or more datasets into a single dataset? — 1 / 1 point

○ Data transferring
● Data merging
○ Data validation
○ Data composition



**Hands-On Activity: Clean data with spreadsheet functions**
Practice Quiz • 1h • 2 total points

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

By now, you've been introduced to some useful techniques for cleaning spreadsheet data, such as sorting and filtering. In this activity, you'll continue to develop your data-cleaning skills by using spreadsheet functions.

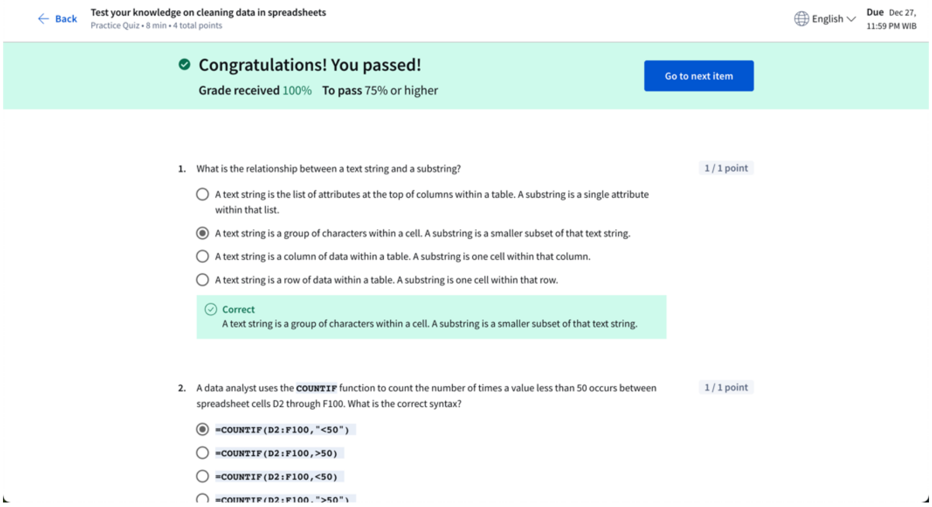
**Scenario**

Review the following scenario. Then complete the step-by-step instructions.

Imagine you are a data analyst working for a marketing agency based in San Francisco. The marketing agency wants to contact local boba tea shops to inquire about a potential collaboration for a new marketing campaign. The agency plans to visit the top-rated shops within a 10-mile radius of the center of their target area. To assist with planning, the agency asks your team to review external data related to ratings and locations of boba tea shops in San Francisco. One of your teammates has created a spreadsheet from an online source. However, the data is not in the greatest shape.

Your assignment is to identify the dirty elements in the dataset and clean them up.

By the time you complete this activity, you will be able to identify dirty elements in a dataset, remove duplicate data, and use the `COUNTIF` and `SPLIT` functions to help clean data.



**Test your knowledge on cleaning data in spreadsheets**
Practice Quiz • 8 min • 4 total points

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

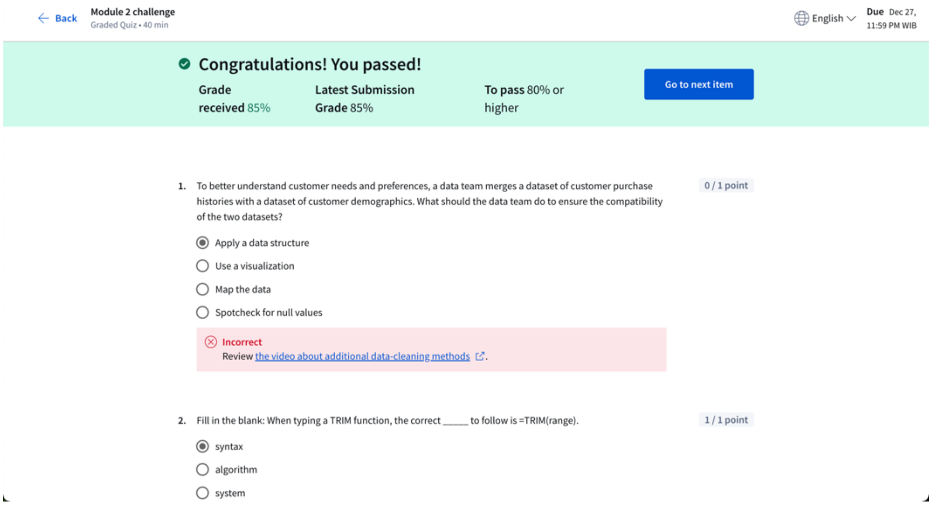
1. What is the relationship between a text string and a substring? 1 / 1 point

○ A text string is the list of attributes at the top of columns within a table. A substring is a single attribute within that list.
● A text string is a group of characters within a cell. A substring is a smaller subset of that text string.
○ A text string is a column of data within a table. A substring is one cell within that column.
○ A text string is a row of data within a table. A substring is one cell within that row.

✓ Correct
A text string is a group of characters within a cell. A substring is a smaller subset of that text string.

2. A data analyst uses the `COUNTIF` function to count the number of times a value less than 50 occurs between spreadsheet cells D2 through F100. What is the correct syntax? 1 / 1 point

● `=COUNTIF(D2:F100,"<50")`
○ `=COUNTIF(D2:F100,>50)`
○ `=COUNTIF(D2:F100,<50)`
○ `=COUNTIF(D2:F100,">50")`



**Module 2 challenge**
Graded Quiz • 40 min

**Congratulations! You passed!**

**Grade received** 85% **Latest Submission Grade** 85% **To pass** 80% or higher

1. To better understand customer needs and preferences, a data team merges a dataset of customer purchase histories with a dataset of customer demographics. What should the data team do to ensure the compatibility of the two datasets? 0 / 1 point

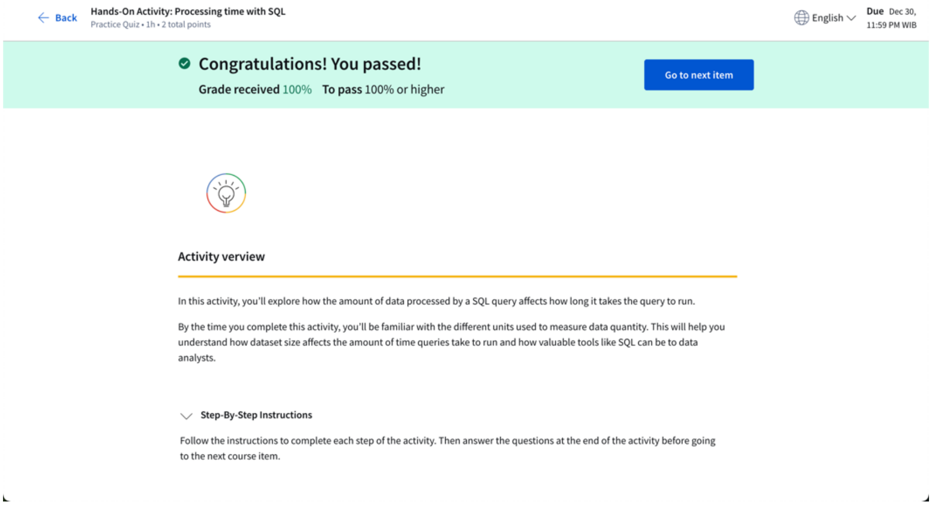
● Apply a data structure
○ Use a visualization
○ Map the data
○ Spotcheck for null values

⊗ Incorrect
Review the video about additional data-cleaning methods.

2. Fill in the blank: When typing a TRIM function, the correct _____ to follow is =TRIM(range). 1 / 1 point

● syntax
○ algorithm
○ system

# Modul 3



**Hands-On Activity: Processing time with SQL**
Practice Quiz • 1h • 2 total points

**Due** Dec 30, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

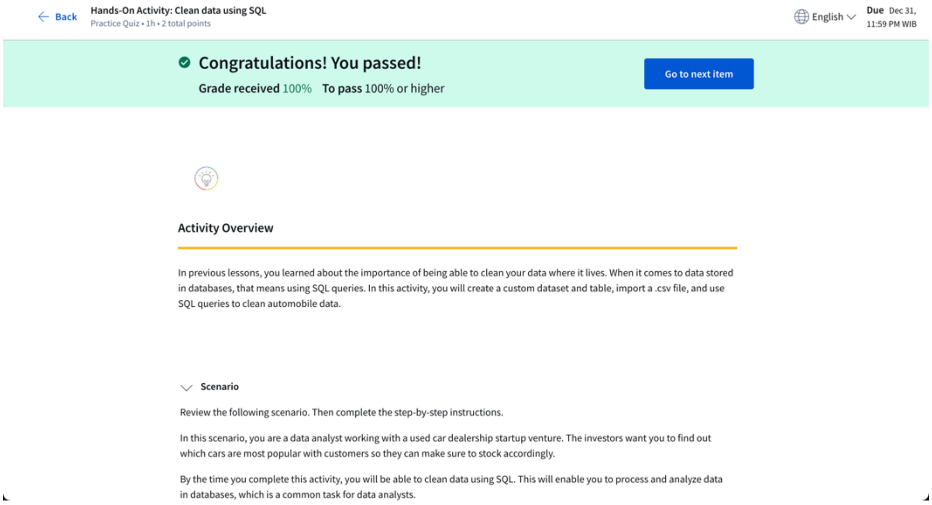
**Activity verview**

In this activity, you'll explore how the amount of data processed by a SQL query affects how long it takes the query to run.

By the time you complete this activity, you'll be familiar with the different units used to measure data quantity. This will help you understand how dataset size affects the amount of time queries take to run and how valuable tools like SQL can be to data analysts.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.



**Hands-On Activity: Clean data using SQL**
Practice Quiz • 1h • 2 total points

**Due** Dec 31, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

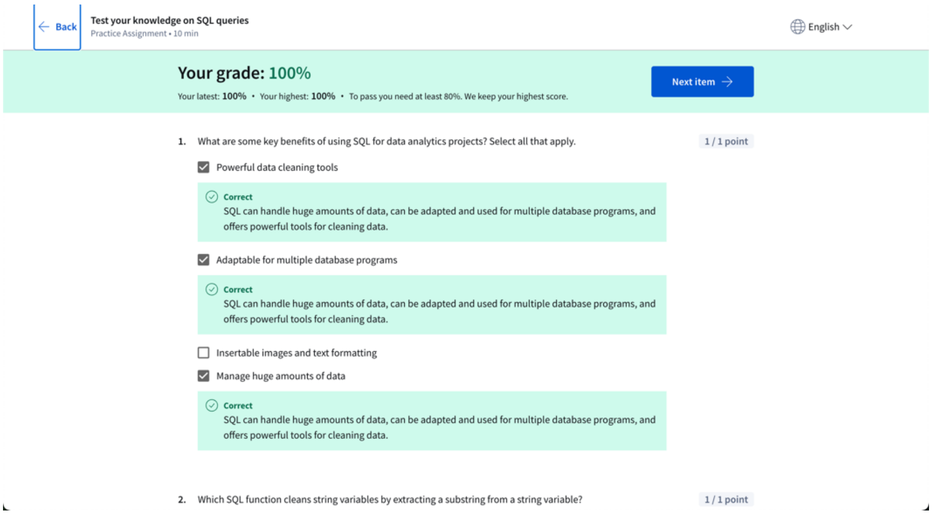
In previous lessons, you learned about the importance of being able to clean your data where it lives. When it comes to data stored in databases, that means using SQL queries. In this activity, you will create a custom dataset and table, import a .csv file, and use SQL queries to clean automobile data.

**Scenario**

Review the following scenario. Then complete the step-by-step instructions.

In this scenario, you are a data analyst working with a used car dealership startup venture. The investors want you to find out which cars are most popular with customers so they can make sure to stock accordingly.

By the time you complete this activity, you will be able to clean data using SQL. This will enable you to process and analyze data in databases, which is a common task for data analysts.



**Test your knowledge on SQL queries**
Practice Assignment • 10 min

**Your grade:** 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

1. What are some key benefits of using SQL for data analytics projects? Select all that apply. 1 / 1 point

☑ Powerful data cleaning tools

Correct
SQL can handle huge amounts of data, can be adapted and used for multiple database programs, and offers powerful tools for cleaning data.

☑ Adaptable for multiple database programs

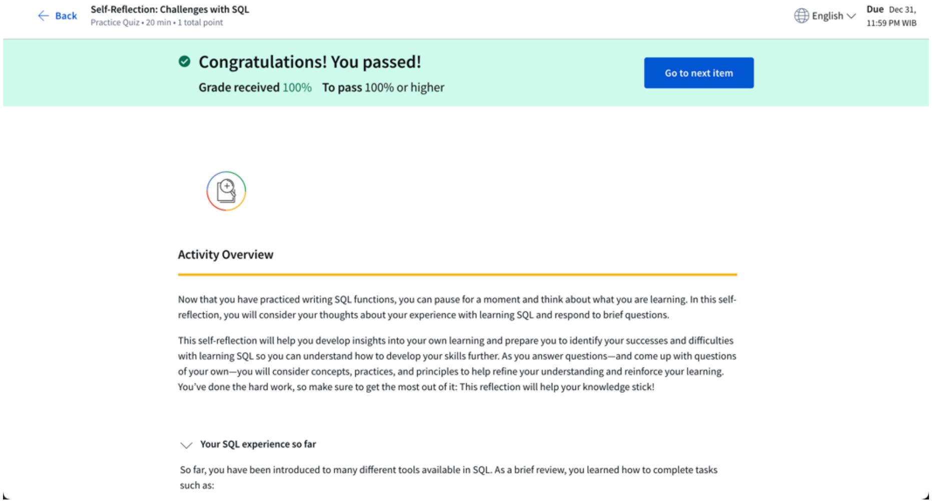
Correct
SQL can handle huge amounts of data, can be adapted and used for multiple database programs, and offers powerful tools for cleaning data.

☐ Insertable images and text formatting

☑ Manage huge amounts of data

Correct
SQL can handle huge amounts of data, can be adapted and used for multiple database programs, and offers powerful tools for cleaning data.

2. Which SQL function cleans string variables by extracting a substring from a string variable? 1 / 1 point



**Self-Reflection: Challenges with SQL**
Practice Quiz • 20 min • 1 total point

Due Dec 31, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

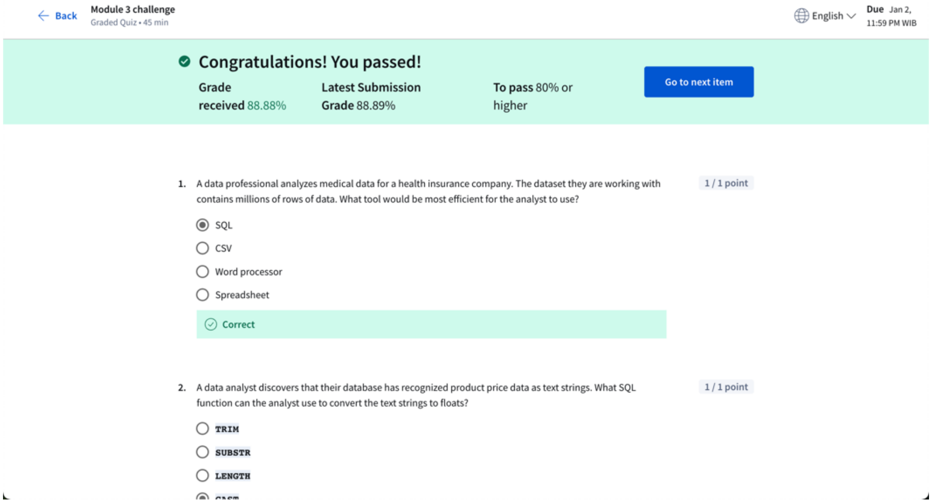
**Activity Overview**

Now that you have practiced writing SQL functions, you can pause for a moment and think about what you are learning. In this self-reflection, you will consider your thoughts about your experience with learning SQL and respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to identify your successes and difficulties with learning SQL so you can understand how to develop your skills further. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

**Your SQL experience so far**

So far, you have been introduced to many different tools available in SQL. As a brief review, you learned how to complete tasks such as:



**Module 3 challenge**
Graded Quiz • 45 min

Due Jan 2, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 88.88% **Latest Submission Grade** 88.89% **To pass** 80% or higher

1. A data professional analyzes medical data for a health insurance company. The dataset they are working with contains millions of rows of data. What tool would be most efficient for the analyst to use? 1 / 1 point

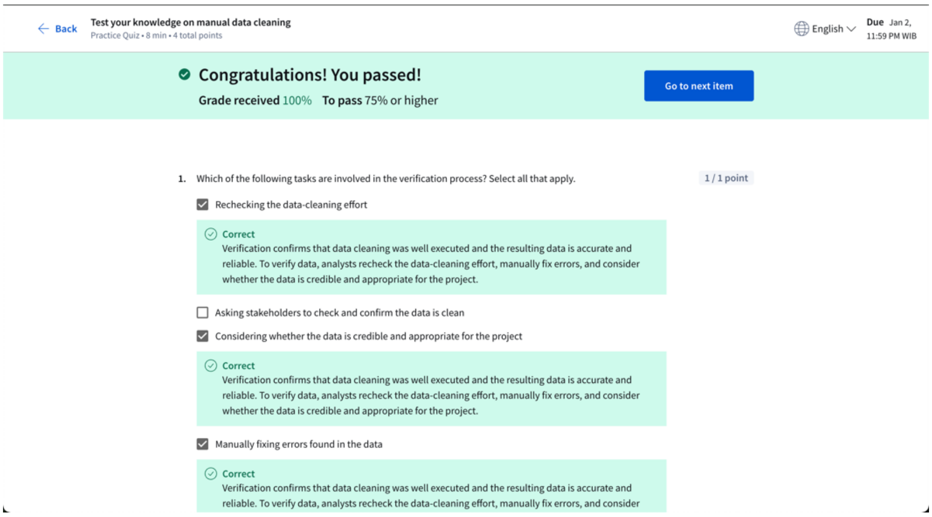
● SQL
○ CSV
○ Word processor
○ Spreadsheet

Correct

2. A data analyst discovers that their database has recognized product price data as text strings. What SQL function can the analyst use to convert the text strings to floats? 1 / 1 point

○ `TRIM`
○ `SUBSTR`
○ `LENGTH`
●

# Modul 4



**Test your knowledge on manual data cleaning**
Practice Quiz • 8 min • 4 total points

Due Jan 2, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. Which of the following tasks are involved in the verification process? Select all that apply. 1 / 1 point

☑ Rechecking the data-cleaning effort

Correct
Verification confirms that data cleaning was well executed and the resulting data is accurate and reliable. To verify data, analysts recheck the data-cleaning effort, manually fix errors, and consider whether the data is credible and appropriate for the project.

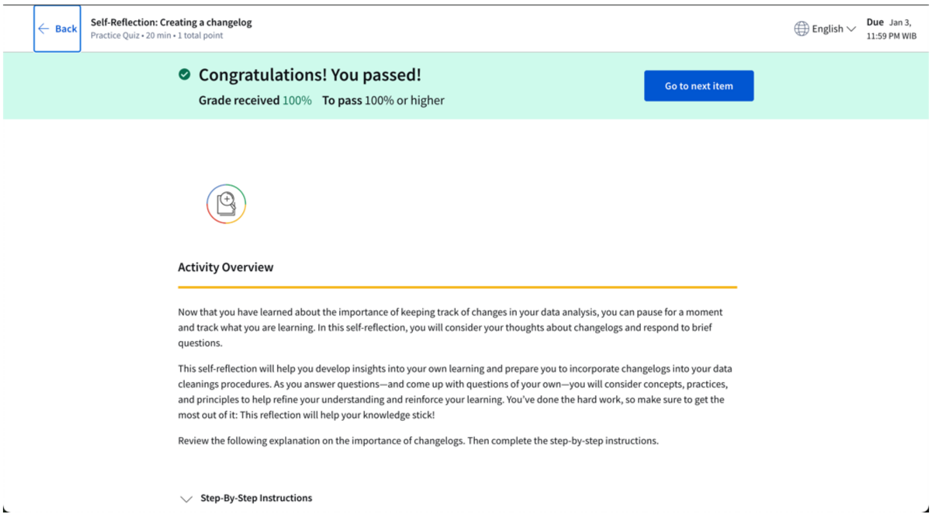
☐ Asking stakeholders to check and confirm the data is clean

☑ Considering whether the data is credible and appropriate for the project

Correct
Verification confirms that data cleaning was well executed and the resulting data is accurate and reliable. To verify data, analysts recheck the data-cleaning effort, manually fix errors, and consider whether the data is credible and appropriate for the project.

☑ Manually fixing errors found in the data

Correct
Verification confirms that data cleaning was well executed and the resulting data is accurate and reliable. To verify data, analysts recheck the data-cleaning effort, manually fix errors, and consider



**Self-Reflection: Creating a changelog**
Practice Quiz • 20 min • 1 total point

**Due** Jan 3, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

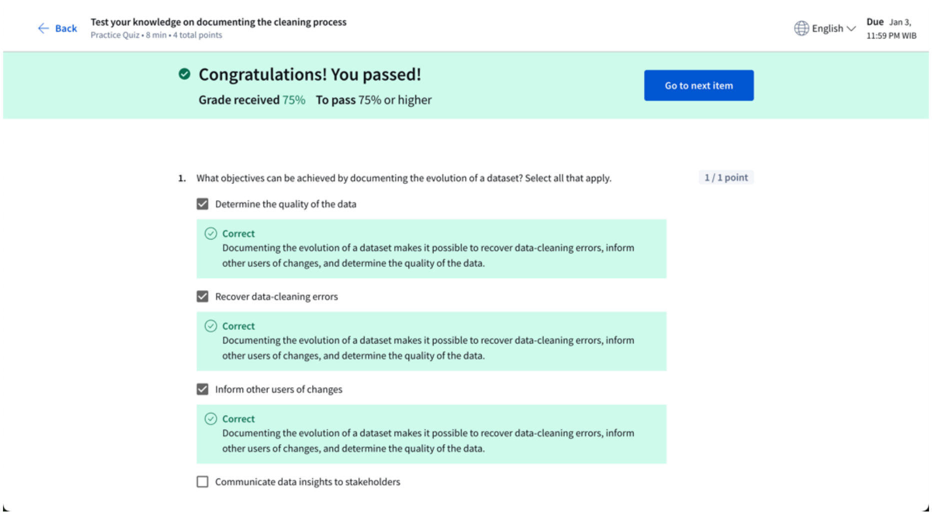
### Activity Overview

Now that you have learned about the importance of keeping track of changes in your data analysis, you can pause for a moment and track what you are learning. In this self-reflection, you will consider your thoughts about changelogs and respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to incorporate changelogs into your data cleanings procedures. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

Review the following explanation on the importance of changelogs. Then complete the step-by-step instructions.

**Step-By-Step Instructions**



**Test your knowledge on documenting the cleaning process**
Practice Quiz • 8 min • 4 total points

**Due** Jan 3, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 75% **To pass** 75% or higher

1. What objectives can be achieved by documenting the evolution of a dataset? Select all that apply. 1 / 1 point

☑ Determine the quality of the data

✓ Correct
Documenting the evolution of a dataset makes it possible to recover data-cleaning errors, inform other users of changes, and determine the quality of the data.

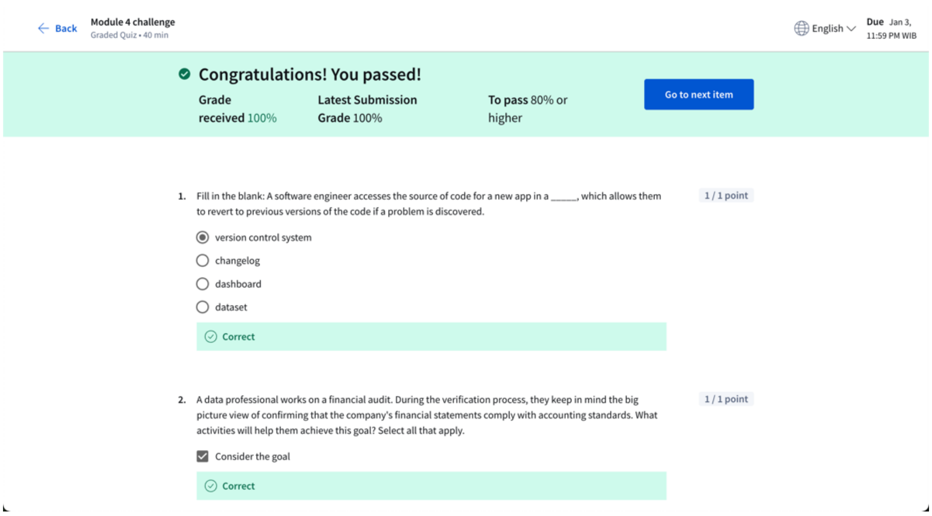
☑ Recover data-cleaning errors

✓ Correct
Documenting the evolution of a dataset makes it possible to recover data-cleaning errors, inform other users of changes, and determine the quality of the data.

☑ Inform other users of changes

✓ Correct
Documenting the evolution of a dataset makes it possible to recover data-cleaning errors, inform other users of changes, and determine the quality of the data.

☐ Communicate data insights to stakeholders



**Module 4 challenge**
Graded Quiz • 40 min

**Due** Jan 3, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **Latest Submission Grade** 100% **To pass** 80% or higher

1. Fill in the blank: A software engineer accesses the source of code for a new app in a _____, which allows them to revert to previous versions of the code if a problem is discovered. 1 / 1 point

● version control system
○ changelog
○ dashboard
○ dataset

✓ Correct

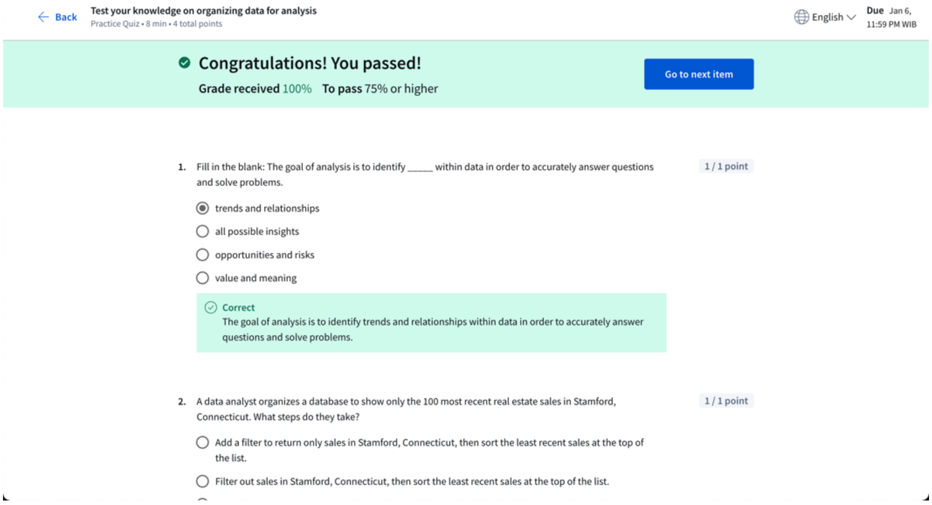
2. A data professional works on a financial audit. During the verification process, they keep in mind the big picture view of confirming that the company's financial statements comply with accounting standards. What activities will help them achieve this goal? Select all that apply. 1 / 1 point

☑ Consider the goal

✓ Correct

# Course 5: *Analyze Data to Answer Questions*

## Modul 1



**Test your knowledge on organizing data for analysis**
Practice Quiz • 8 min • 4 total points

**Due** Jan 6, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

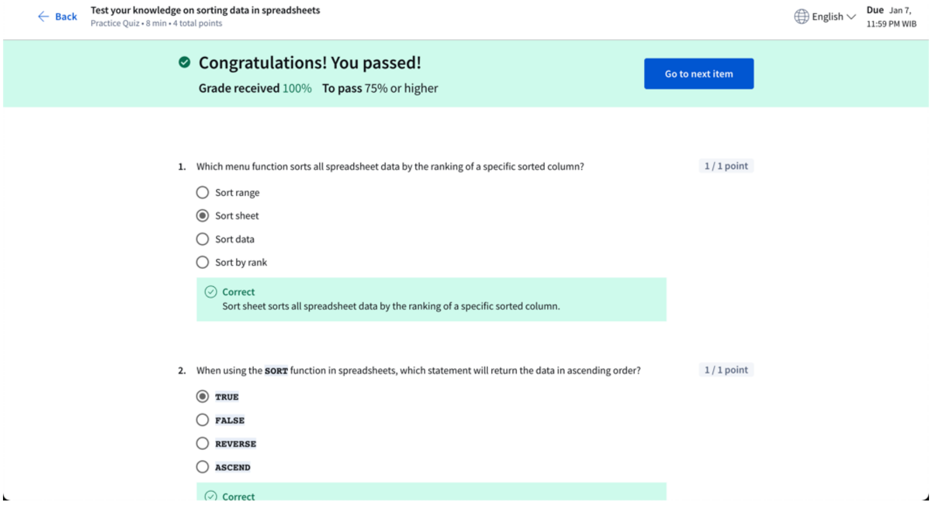
1. Fill in the blank: The goal of analysis is to identify _____ within data in order to accurately answer questions and solve problems.

- ● trends and relationships
- ○ all possible insights
- ○ opportunities and risks
- ○ value and meaning

Correct
The goal of analysis is to identify trends and relationships within data in order to accurately answer questions and solve problems.

2. A data analyst organizes a database to show only the 100 most recent real estate sales in Stamford, Connecticut. What steps do they take?

- ○ Add a filter to return only sales in Stamford, Connecticut, then sort the least recent sales at the top of the list.
- ○ Filter out sales in Stamford, Connecticut, then sort the least recent sales at the top of the list.



**Test your knowledge on sorting data in spreadsheets**
Practice Quiz • 8 min • 4 total points

**Due** Jan 7, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

1. Which menu function sorts all spreadsheet data by the ranking of a specific sorted column?

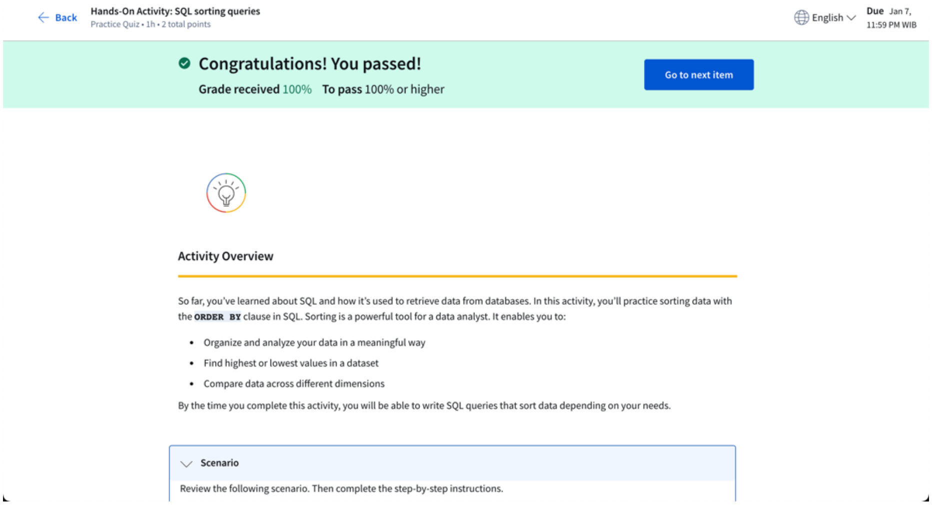
- ○ Sort range
- ● Sort sheet
- ○ Sort data
- ○ Sort by rank

Correct
Sort sheet sorts all spreadsheet data by the ranking of a specific sorted column.

2. When using the `SORT` function in spreadsheets, which statement will return the data in ascending order?

- ● `TRUE`
- ○ `FALSE`
- ○ `REVERSE`
- ○ `ASCEND`

Correct



**Hands-On Activity: SQL sorting queries**
Practice Quiz • 1h • 2 total points

**Due** Jan 7, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

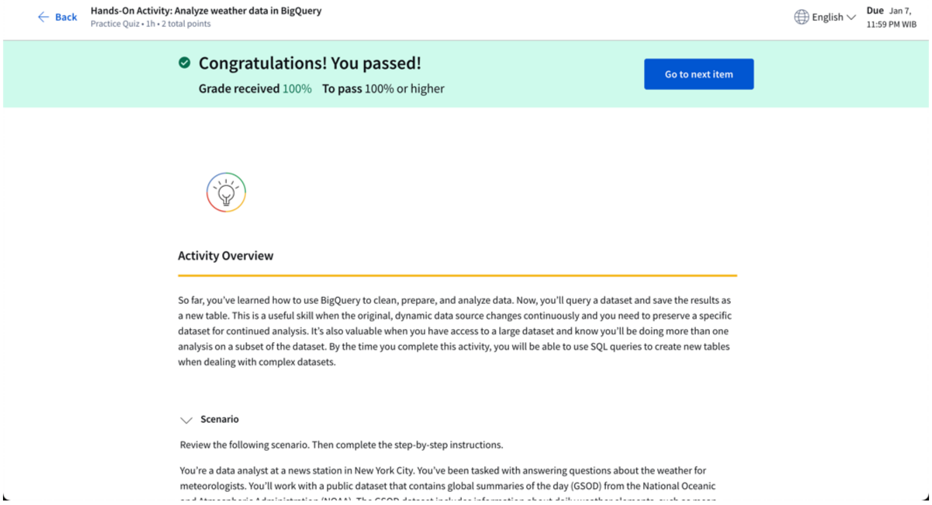
So far, you've learned about SQL and how it's used to retrieve data from databases. In this activity, you'll practice sorting data with the `ORDER BY` clause in SQL. Sorting is a powerful tool for a data analyst. It enables you to:

- Organize and analyze your data in a meaningful way
- Find highest or lowest values in a dataset
- Compare data across different dimensions

By the time you complete this activity, you will be able to write SQL queries that sort data depending on your needs.

Scenario

Review the following scenario. Then complete the step-by-step instructions.



**Hands-On Activity: Analyze weather data in BigQuery**
Practice Quiz • 1h • 2 total points

**Due** Jan 7, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

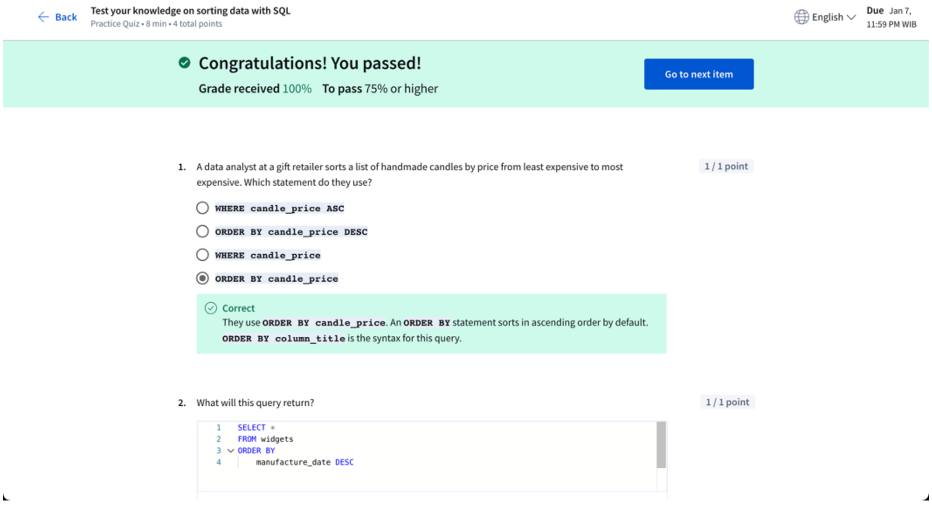
### Activity Overview

So far, you've learned how to use BigQuery to clean, prepare, and analyze data. Now, you'll query a dataset and save the results as a new table. This is a useful skill when the original, dynamic data source changes continuously and you need to preserve a specific dataset for continued analysis. It's also valuable when you have access to a large dataset and know you'll be doing more than one analysis on a subset of the dataset. By the time you complete this activity, you will be able to use SQL queries to create new tables when dealing with complex datasets.

**Scenario**

Review the following scenario. Then complete the step-by-step instructions.

You're a data analyst at a news station in New York City. You've been tasked with answering questions about the weather for meteorologists. You'll work with a public dataset that contains global summaries of the day (GSOD) from the National Oceanic



**Test your knowledge on sorting data with SQL**
Practice Quiz • 8 min • 4 total points

**Due** Jan 7, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

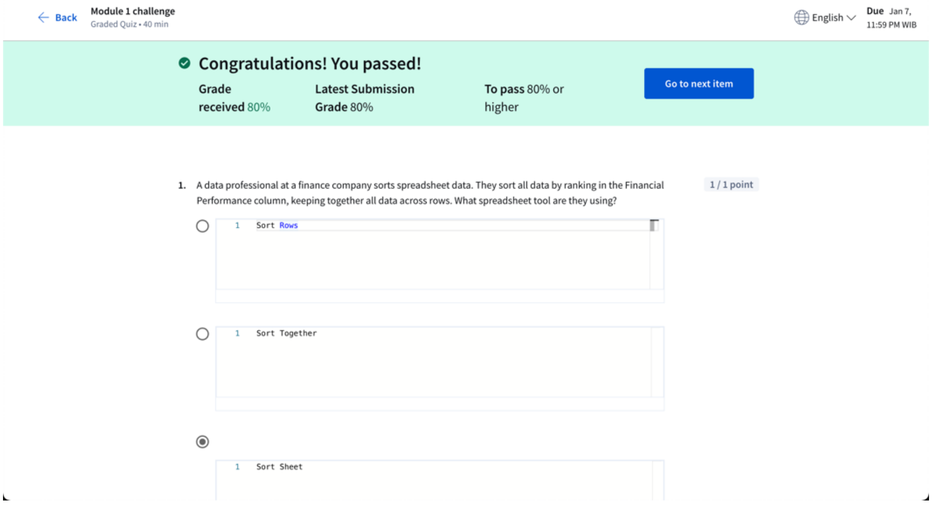
1. A data analyst at a gift retailer sorts a list of handmade candles by price from least expensive to most expensive. Which statement do they use? 1 / 1 point

○ `WHERE candle_price ASC`
○ `ORDER BY candle_price DESC`
○ `WHERE candle_price`
● `ORDER BY candle_price`

Correct
They use `ORDER BY candle_price`. An `ORDER BY` statement sorts in ascending order by default. `ORDER BY column_title` is the syntax for this query.

2. What will this query return? 1 / 1 point

```
SELECT *
FROM widgets
ORDER BY
    manufacture_date DESC
```



**Module 1 challenge**
Graded Quiz • 40 min

**Due** Jan 7, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 80% **Latest Submission Grade** 80% **To pass** 80% or higher

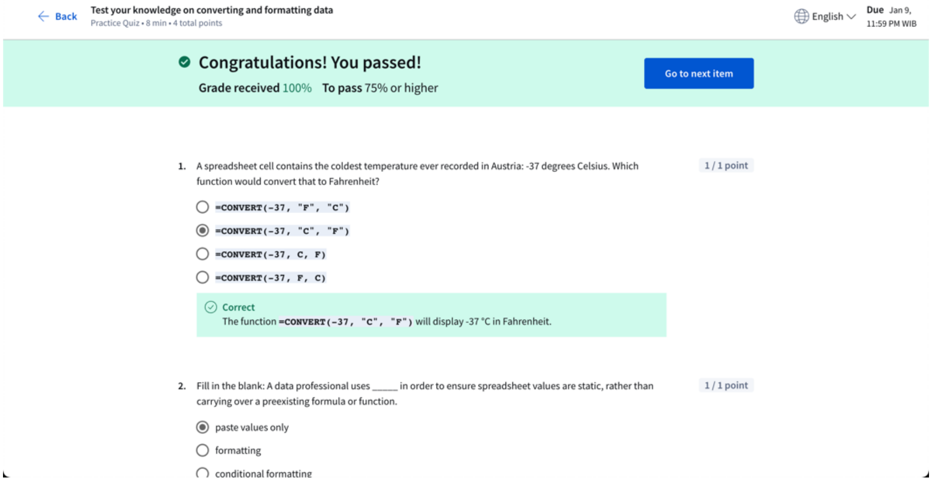
1. A data professional at a finance company sorts spreadsheet data. They sort all data by ranking in the Financial Performance column, keeping together all data across rows. What spreadsheet tool are they using? 1 / 1 point

○
```
Sort Rows
```

○
```
Sort Together
```

●
```
Sort Sheet
```

# Modul 2



**Test your knowledge on converting and formatting data**
Practice Quiz • 8 min • 4 total points

**Due** Jan 9, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

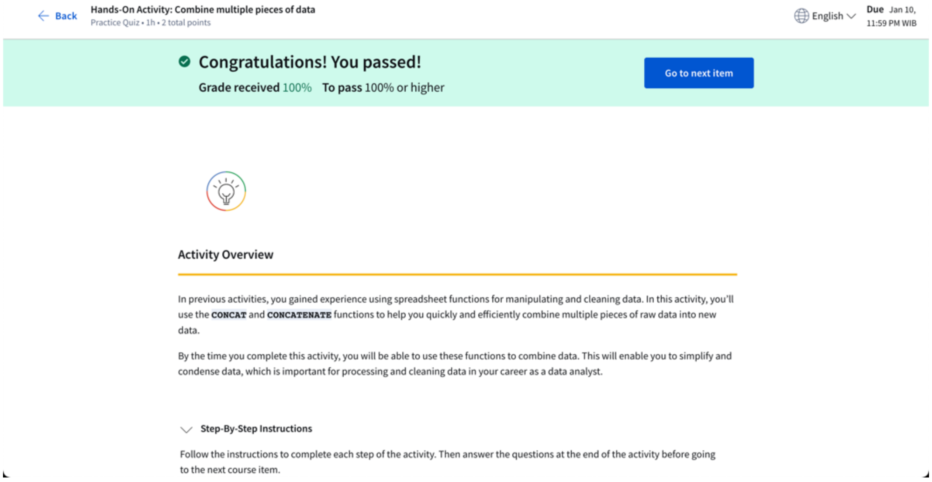
1. A spreadsheet cell contains the coldest temperature ever recorded in Austria: -37 degrees Celsius. Which function would convert that to Fahrenheit? 1 / 1 point

○ `=CONVERT(-37, "F", "C")`
● `=CONVERT(-37, "C", "F")`
○ `=CONVERT(-37, C, F)`
○ `=CONVERT(-37, F, C)`

Correct
The function `=CONVERT(-37, "C", "F")` will display -37 °C in Fahrenheit.

2. Fill in the blank: A data professional uses _____ in order to ensure spreadsheet values are static, rather than carrying over a preexisting formula or function. 1 / 1 point

● paste values only
○ formatting
○ conditional formatting



**Hands-On Activity: Combine multiple pieces of data**
Practice Quiz • 1h • 2 total points

**Due** Jan 10, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

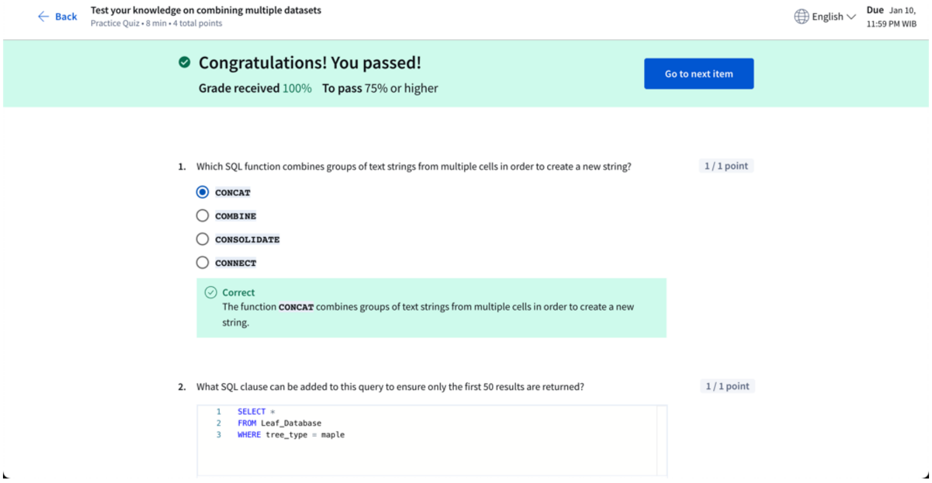
**Activity Overview**

In previous activities, you gained experience using spreadsheet functions for manipulating and cleaning data. In this activity, you'll use the `CONCAT` and `CONCATENATE` functions to help you quickly and efficiently combine multiple pieces of raw data into new data.

By the time you complete this activity, you will be able to use these functions to combine data. This will enable you to simplify and condense data, which is important for processing and cleaning data in your career as a data analyst.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.



**Test your knowledge on combining multiple datasets**
Practice Quiz • 8 min • 4 total points

**Due** Jan 10, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

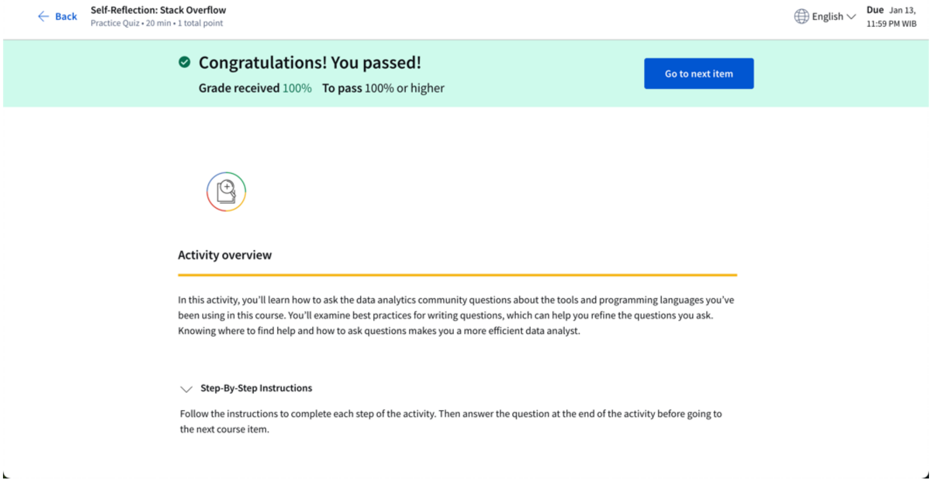
1. Which SQL function combines groups of text strings from multiple cells in order to create a new string? 1 / 1 point

● `CONCAT`
○ `COMBINE`
○ `CONSOLIDATE`
○ `CONNECT`

Correct
The function `CONCAT` combines groups of text strings from multiple cells in order to create a new string.

2. What SQL clause can be added to this query to ensure only the first 50 results are returned? 1 / 1 point

```
SELECT *
FROM Leaf_Database
WHERE tree_type = maple
```



**Self-Reflection: Stack Overflow**
Practice Quiz • 20 min • 1 total point

**Due** Jan 13, 11:59 PM WIB

**Congratulations! You passed!**

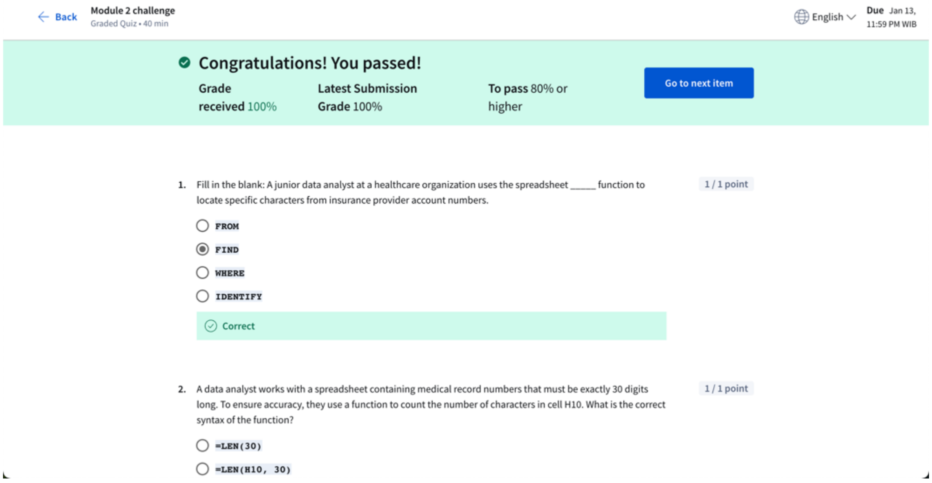
**Grade received** 100% **To pass** 100% or higher

**Activity overview**

In this activity, you'll learn how to ask the data analytics community questions about the tools and programming languages you've been using in this course. You'll examine best practices for writing questions, which can help you refine the questions you ask. Knowing where to find help and how to ask questions makes you a more efficient data analyst.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to the next course item.



**Module 2 challenge**
Graded Quiz • 40 min

**Due** Jan 13, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **Latest Submission Grade** 100% **To pass** 80% or higher

1. Fill in the blank: A junior data analyst at a healthcare organization uses the spreadsheet _____ function to locate specific characters from insurance provider account numbers. 1 / 1 point

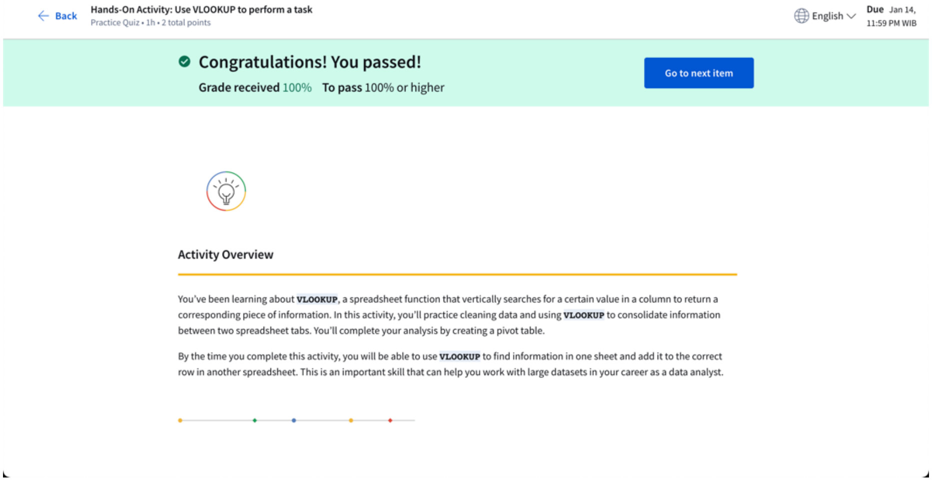
○ `FROM`
● `FIND`
○ `WHERE`
○ `IDENTIFY`

✓ Correct

2. A data analyst works with a spreadsheet containing medical record numbers that must be exactly 30 digits long. To ensure accuracy, they use a function to count the number of characters in cell H10. What is the correct syntax of the function? 1 / 1 point

○ `=LEN(30)`
○ `=LEN(H10, 30)`

# Modul 3



**Hands-On Activity: Use VLOOKUP to perform a task**
Practice Quiz • 1h • 2 total points

**Due** Jan 14, 11:59 PM WIB

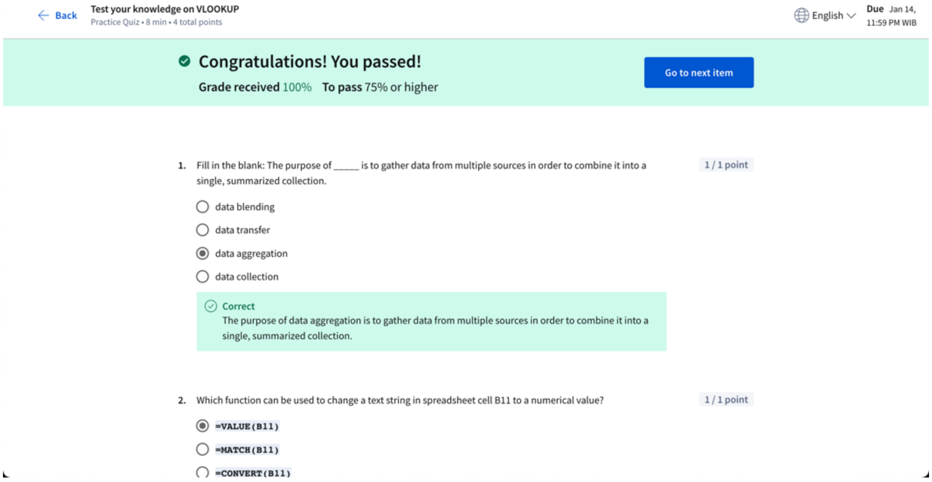
**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

You've been learning about `VLOOKUP`, a spreadsheet function that vertically searches for a certain value in a column to return a corresponding piece of information. In this activity, you'll practice cleaning data and using `VLOOKUP` to consolidate information between two spreadsheet tabs. You'll complete your analysis by creating a pivot table.

By the time you complete this activity, you will be able to use `VLOOKUP` to find information in one sheet and add it to the correct row in another spreadsheet. This is an important skill that can help you work with large datasets in your career as a data analyst.



**Test your knowledge on VLOOKUP**
Practice Quiz • 8 min • 4 total points

**Due** Jan 14, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. Fill in the blank: The purpose of _____ is to gather data from multiple sources in order to combine it into a single, summarized collection.

1 / 1 point

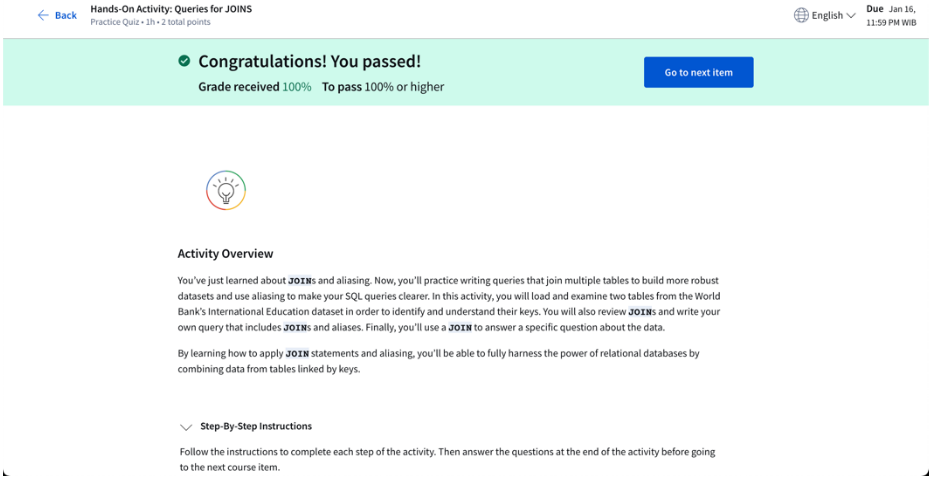
○ data blending
○ data transfer
● data aggregation
○ data collection

Correct
The purpose of data aggregation is to gather data from multiple sources in order to combine it into a single, summarized collection.

2. Which function can be used to change a text string in spreadsheet cell B11 to a numerical value?

1 / 1 point

● `=VALUE(B11)`
○ `=MATCH(B11)`
○ `=CONVERT(B11)`



**Hands-On Activity: Queries for JOINS**
Practice Quiz • 1h • 2 total points

**Due** Jan 16, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

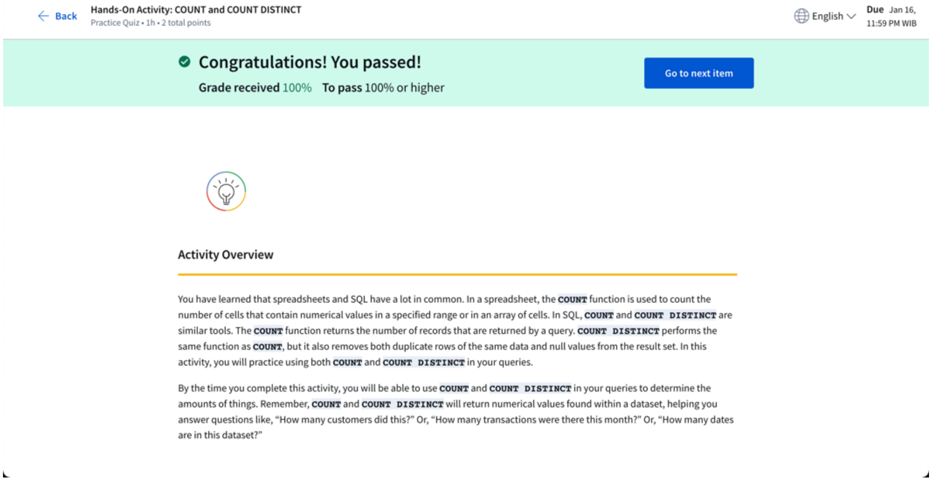
**Activity Overview**

You've just learned about `JOIN`s and aliasing. Now, you'll practice writing queries that join multiple tables to build more robust datasets and use aliasing to make your SQL queries clearer. In this activity, you will load and examine two tables from the World Bank's International Education dataset in order to identify and understand their keys. You will also review `JOIN`s and write your own query that includes `JOIN`s and aliases. Finally, you'll use a `JOIN` to answer a specific question about the data.

By learning how to apply `JOIN` statements and aliasing, you'll be able to fully harness the power of relational databases by combining data from tables linked by keys.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.



**Hands-On Activity: COUNT and COUNT DISTINCT**
Practice Quiz • 1h • 2 total points

**Due** Jan 16, 11:59 PM WIB

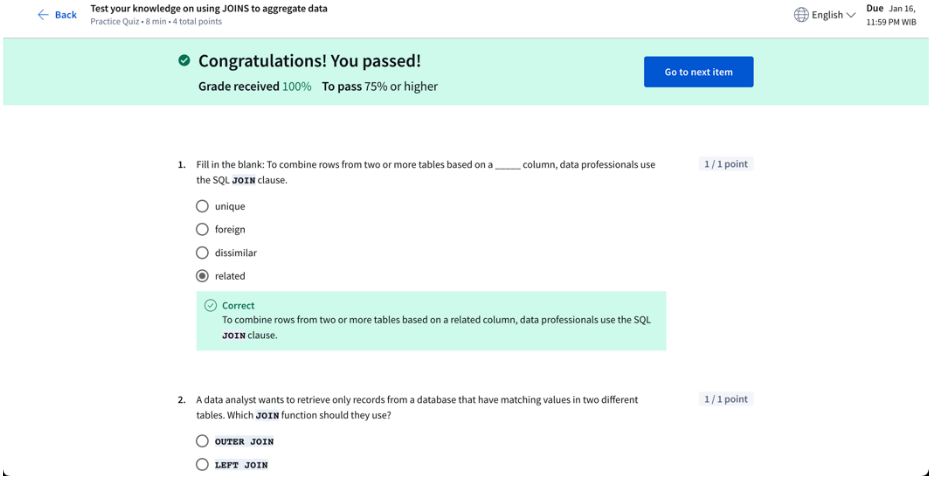
**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

You have learned that spreadsheets and SQL have a lot in common. In a spreadsheet, the `COUNT` function is used to count the number of cells that contain numerical values in a specified range or in an array of cells. In SQL, `COUNT` and `COUNT DISTINCT` are similar tools. The `COUNT` function returns the number of records that are returned by a query. `COUNT DISTINCT` performs the same function as `COUNT`, but it also removes both duplicate rows of the same data and null values from the result set. In this activity, you will practice using both `COUNT` and `COUNT DISTINCT` in your queries.

By the time you complete this activity, you will be able to use `COUNT` and `COUNT DISTINCT` in your queries to determine the amounts of things. Remember, `COUNT` and `COUNT DISTINCT` will return numerical values found within a dataset, helping you answer questions like, "How many customers did this?" Or, "How many transactions were there this month?" Or, "How many dates are in this dataset?"



**Test your knowledge on using JOINS to aggregate data**
Practice Quiz • 8 min • 4 total points

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

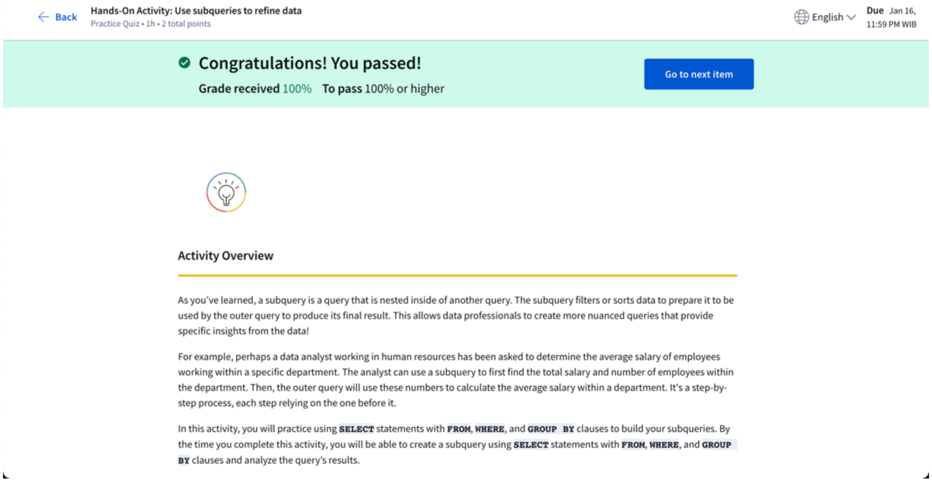
1. Fill in the blank: To combine rows from two or more tables based on a _____ column, data professionals use the SQL `JOIN` clause. 1 / 1 point

○ unique
○ foreign
○ dissimilar
● related

Correct
To combine rows from two or more tables based on a related column, data professionals use the SQL `JOIN` clause.

2. A data analyst wants to retrieve only records from a database that have matching values in two different tables. Which `JOIN` function should they use? 1 / 1 point

○ `OUTER JOIN`
○ `LEFT JOIN`



**Hands-On Activity: Use subqueries to refine data**
Practice Quiz • 1h • 2 total points

**Congratulations! You passed!**

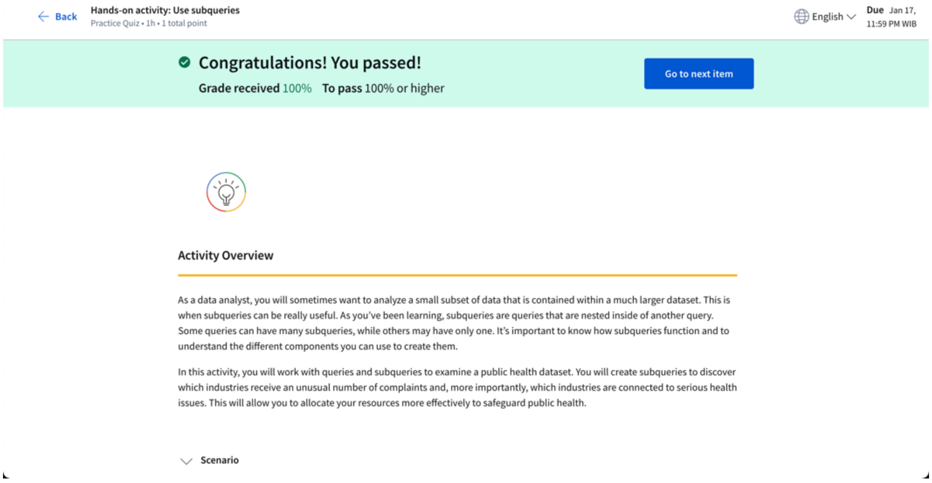
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

As you've learned, a subquery is a query that is nested inside of another query. The subquery filters or sorts data to prepare it to be used by the outer query to produce its final result. This allows data professionals to create more nuanced queries that provide specific insights from the data!

For example, perhaps a data analyst working in human resources has been asked to determine the average salary of employees working within a specific department. The analyst can use a subquery to first find the total salary and number of employees within the department. Then, the outer query will use these numbers to calculate the average salary within a department. It's a step-by-step process, each step relying on the one before it.

In this activity, you will practice using `SELECT` statements with `FROM`, `WHERE`, and `GROUP BY` clauses to build your subqueries. By the time you complete this activity, you will be able to create a subquery using `SELECT` statements with `FROM`, `WHERE`, and `GROUP BY` clauses and analyze the query's results.



**Hands-on activity: Use subqueries**
Practice Quiz • 1h • 1 total point

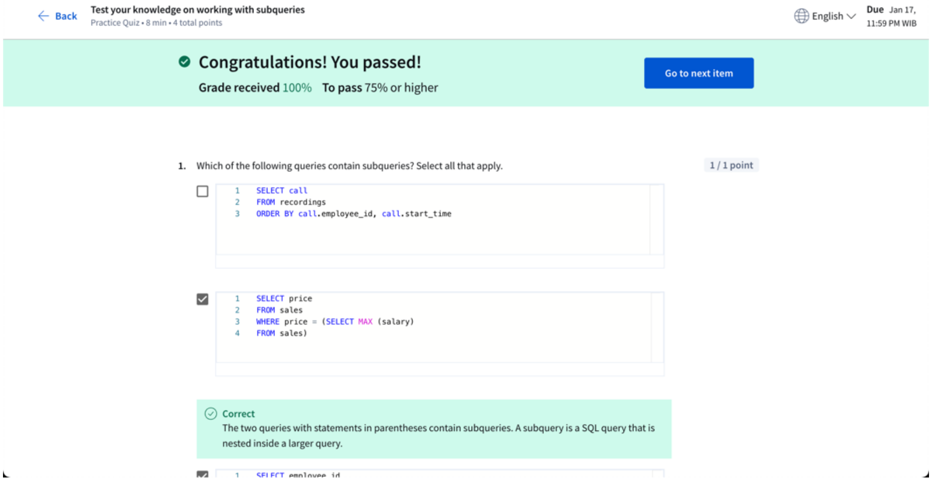
**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

As a data analyst, you will sometimes want to analyze a small subset of data that is contained within a much larger dataset. This is when subqueries can be really useful. As you've been learning, subqueries are queries that are nested inside of another query. Some queries can have many subqueries, while others may have only one. It's important to know how subqueries function and to understand the different components you can use to create them.

In this activity, you will work with queries and subqueries to examine a public health dataset. You will create subqueries to discover which industries receive an unusual number of complaints and, more importantly, which industries are connected to serious health issues. This will allow you to allocate your resources more effectively to safeguard public health.



**Test your knowledge on working with subqueries**
Practice Quiz • 8 min • 4 total points

Due Jan 17, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

1. Which of the following queries contain subqueries? Select all that apply.

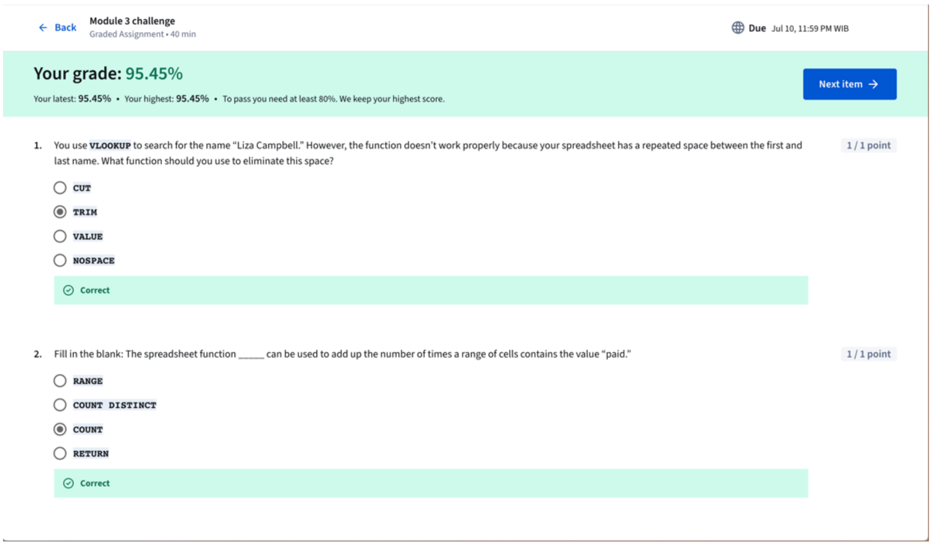
1 / 1 point

☐
```
SELECT call
FROM recordings
ORDER BY call.employee_id, call.start_time
```

☑
```
SELECT price
FROM sales
WHERE price = (SELECT MAX (salary)
FROM sales)
```

Correct
The two queries with statements in parentheses contain subqueries. A subquery is a SQL query that is nested inside a larger query.

☑
```
SELECT employee_id
```



**Module 3 challenge**
Graded Assignment • 40 min

Due Jul 10, 11:59 PM WIB

**Your grade: 95.45%**

Your latest: **95.45%** • Your highest: **95.45%** • To pass you need at least 80%. We keep your highest score.

1. You use `VLOOKUP` to search for the name "Liza Campbell." However, the function doesn't work properly because your spreadsheet has a repeated space between the first and last name. What function should you use to eliminate this space?

1 / 1 point

○ `CUT`
● `TRIM`
○ `VALUE`
○ `NOSPACE`

Correct

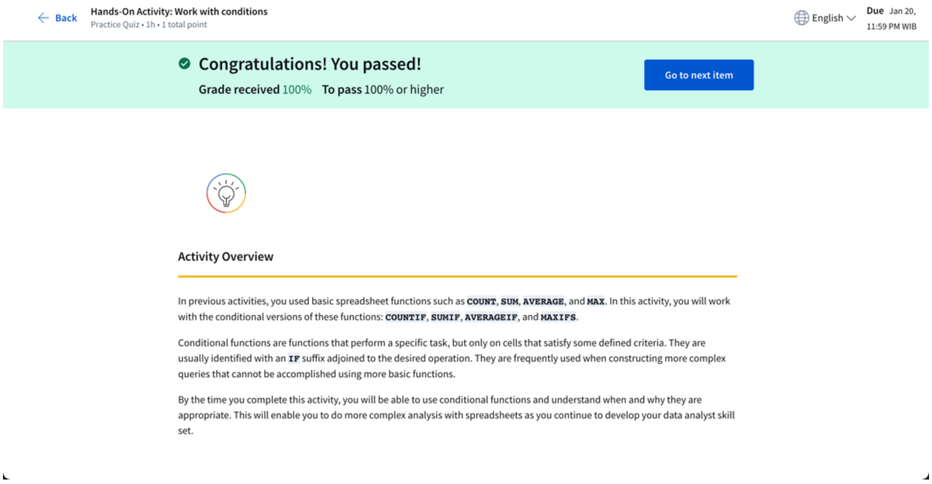
2. Fill in the blank: The spreadsheet function _____ can be used to add up the number of times a range of cells contains the value "paid."

1 / 1 point

○ `RANGE`
○ `COUNT DISTINCT`
● `COUNT`
○ `RETURN`

Correct

# Modul 4



**Hands-On Activity: Work with conditions**
Practice Quiz • 1h • 1 total point

Due Jan 20, 11:59 PM WIB

**Congratulations! You passed!**

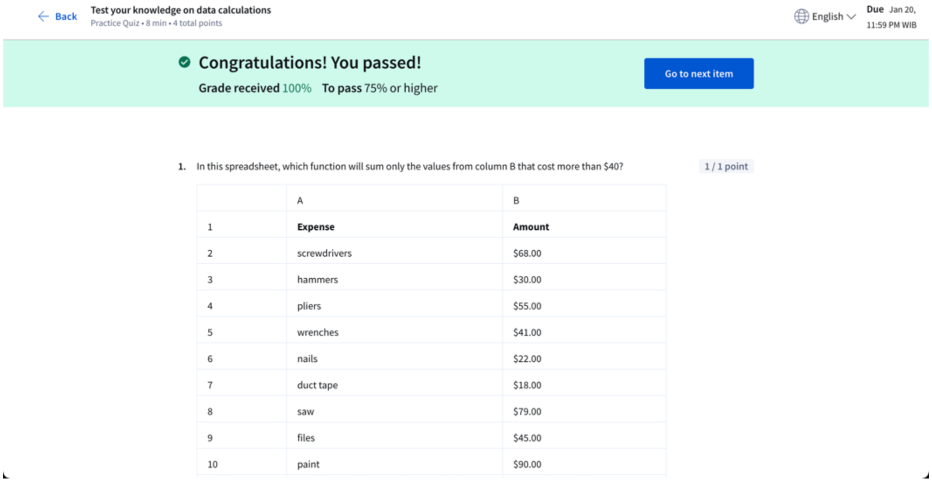
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

In previous activities, you used basic spreadsheet functions such as `COUNT`, `SUM`, `AVERAGE`, and `MAX`. In this activity, you will work with the conditional versions of these functions: `COUNTIF`, `SUMIF`, `AVERAGEIF`, and `MAXIFS`.

Conditional functions are functions that perform a specific task, but only on cells that satisfy some defined criteria. They are usually identified with an `IF` suffix adjoined to the desired operation. They are frequently used when constructing more complex queries that cannot be accomplished using more basic functions.

By the time you complete this activity, you will be able to use conditional functions and understand when and why they are appropriate. This will enable you to do more complex analysis with spreadsheets as you continue to develop your data analyst skill set.



**Test your knowledge on data calculations**
Practice Quiz • 8 min • 4 total points

**Due** Jan 20, 11:59 PM WIB

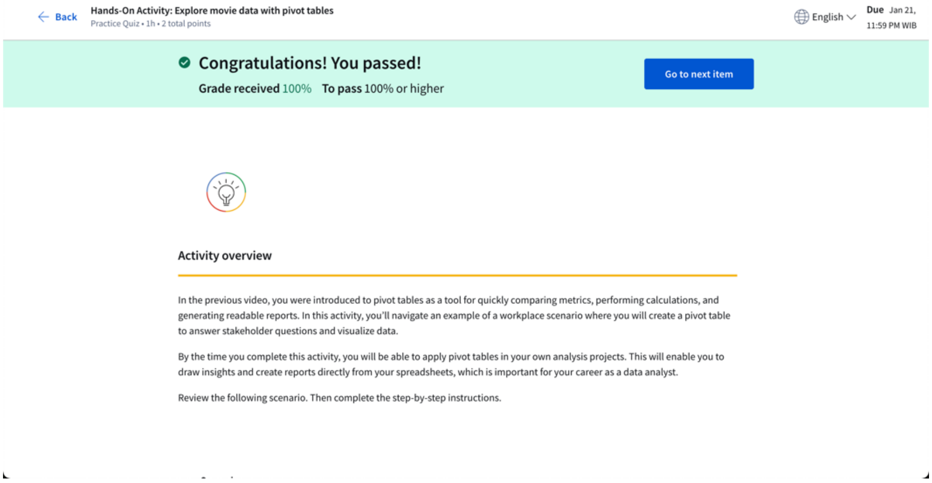
Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

**1.** In this spreadsheet, which function will sum only the values from column B that cost more than $40?

1 / 1 point

| | A | B |
|---|---|---|
| 1 | **Expense** | **Amount** |
| 2 | screwdrivers | $68.00 |
| 3 | hammers | $30.00 |
| 4 | pliers | $55.00 |
| 5 | wrenches | $41.00 |
| 6 | nails | $22.00 |
| 7 | duct tape | $18.00 |
| 8 | saw | $79.00 |
| 9 | files | $45.00 |
| 10 | paint | $90.00 |



**Hands-On Activity: Explore movie data with pivot tables**
Practice Quiz • 1h • 2 total points

**Due** Jan 21, 11:59 PM WIB

Congratulations! You passed!

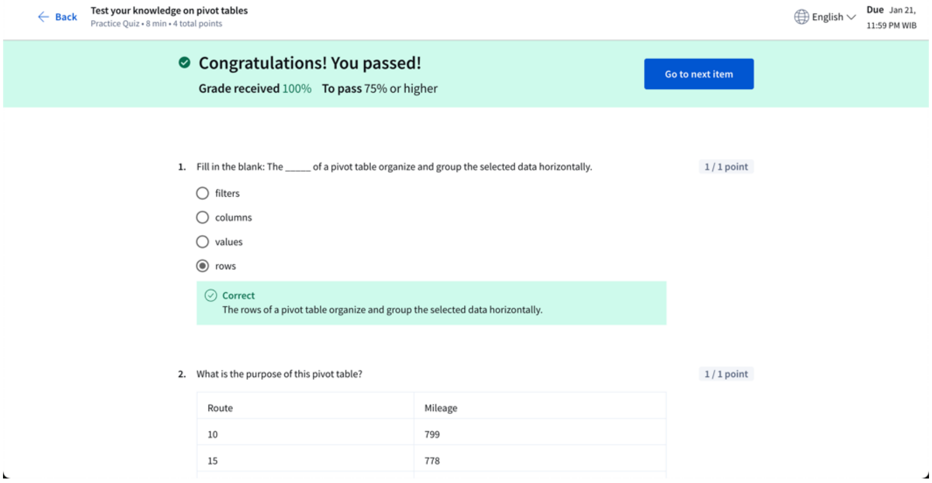
**Grade received** 100% **To pass** 100% or higher

**Activity overview**

In the previous video, you were introduced to pivot tables as a tool for quickly comparing metrics, performing calculations, and generating readable reports. In this activity, you'll navigate an example of a workplace scenario where you will create a pivot table to answer stakeholder questions and visualize data.

By the time you complete this activity, you will be able to apply pivot tables in your own analysis projects. This will enable you to draw insights and create reports directly from your spreadsheets, which is important for your career as a data analyst.

Review the following scenario. Then complete the step-by-step instructions.



**Test your knowledge on pivot tables**
Practice Quiz • 8 min • 4 total points

**Due** Jan 21, 11:59 PM WIB

Congratulations! You passed!

**Grade received** 100% **To pass** 75% or higher

**1.** Fill in the blank: The _____ of a pivot table organize and group the selected data horizontally.

1 / 1 point

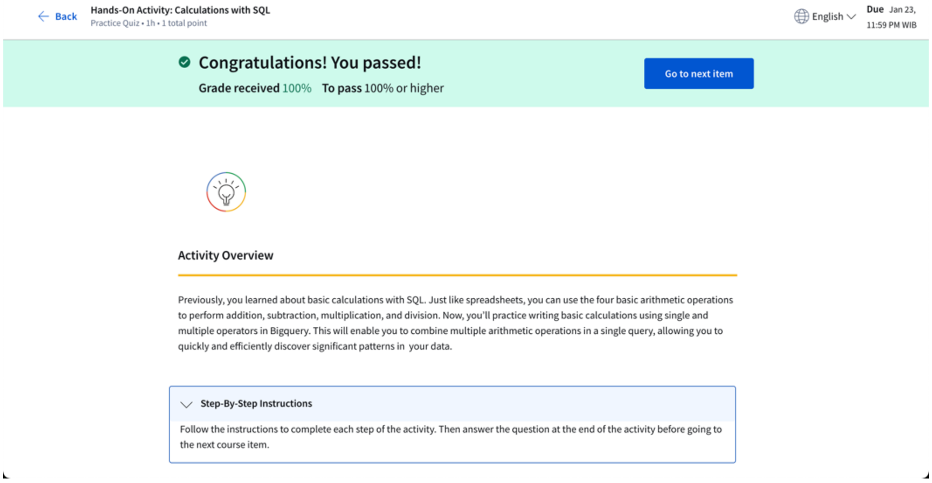
○ filters
○ columns
○ values
● rows

Correct
The rows of a pivot table organize and group the selected data horizontally.

**2.** What is the purpose of this pivot table?

1 / 1 point

| Route | Mileage |
|---|---|
| 10 | 799 |
| 15 | 778 |



**Hands-On Activity: Calculations with SQL**
Practice Quiz • 1h • 1 total point

**Due** Jan 23, 11:59 PM WIB

**Congratulations! You passed!**

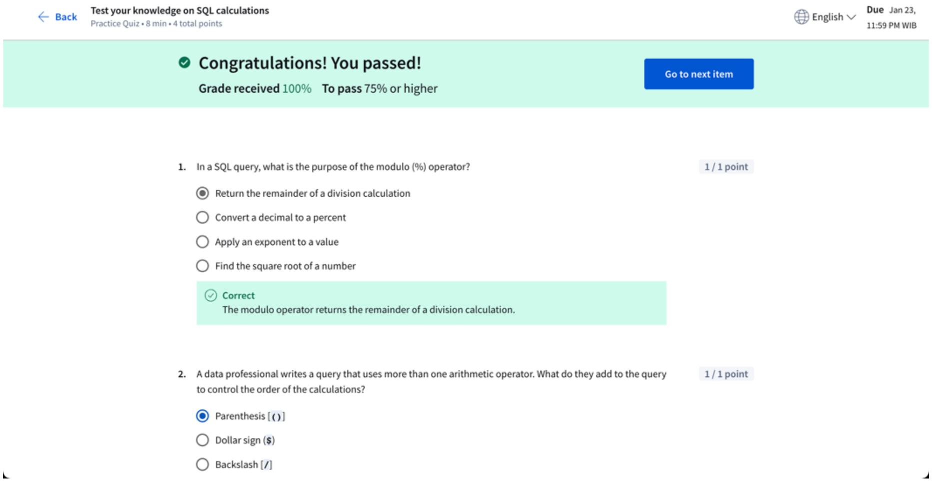
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

Previously, you learned about basic calculations with SQL. Just like spreadsheets, you can use the four basic arithmetic operations to perform addition, subtraction, multiplication, and division. Now, you'll practice writing basic calculations using single and multiple operators in Bigquery. This will enable you to combine multiple arithmetic operations in a single query, allowing you to quickly and efficiently discover significant patterns in your data.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to the next course item.



**Test your knowledge on SQL calculations**
Practice Quiz • 8 min • 4 total points

**Due** Jan 23, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

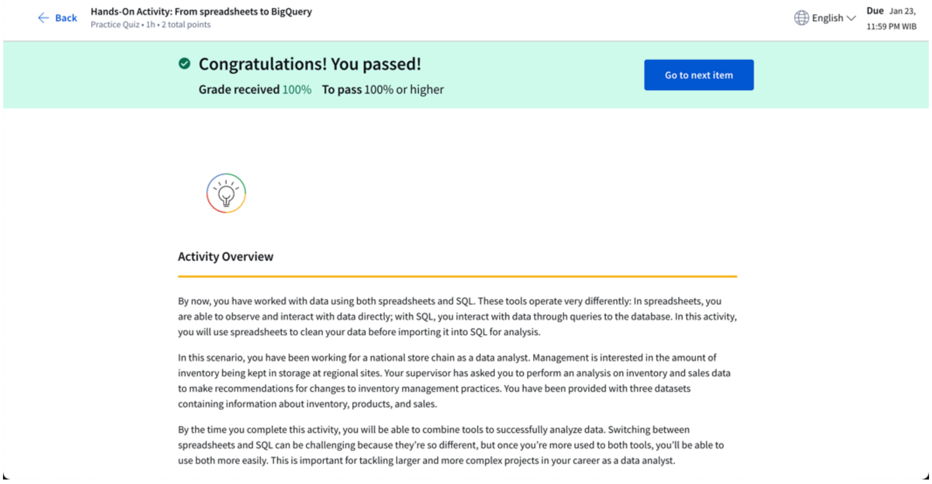
1. In a SQL query, what is the purpose of the modulo (%) operator? 1 / 1 point

● Return the remainder of a division calculation
○ Convert a decimal to a percent
○ Apply an exponent to a value
○ Find the square root of a number

Correct
The modulo operator returns the remainder of a division calculation.

2. A data professional writes a query that uses more than one arithmetic operator. What do they add to the query to control the order of the calculations? 1 / 1 point

● Parenthesis [ ( ) ]
○ Dollar sign ($)
○ Backslash [/]



**Hands-On Activity: From spreadsheets to BigQuery**
Practice Quiz • 1h • 2 total points

**Due** Jan 23, 11:59 PM WIB

**Congratulations! You passed!**

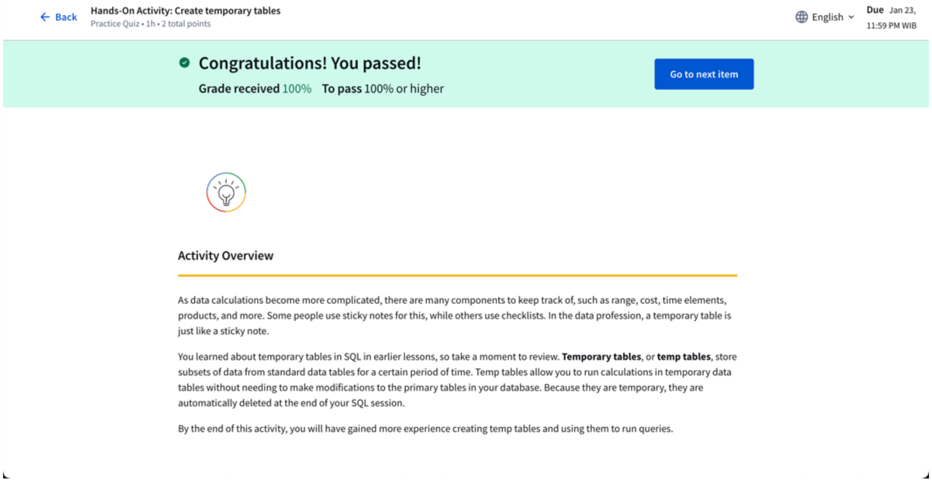
**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

By now, you have worked with data using both spreadsheets and SQL. These tools operate very differently: In spreadsheets, you are able to observe and interact with data directly; with SQL, you interact with data through queries to the database. In this activity, you will use spreadsheets to clean your data before importing it into SQL for analysis.

In this scenario, you have been working for a national store chain as a data analyst. Management is interested in the amount of inventory being kept in storage at regional sites. Your supervisor has asked you to perform an analysis on inventory and sales data to make recommendations for changes to inventory management practices. You have been provided with three datasets containing information about inventory, products, and sales.

By the time you complete this activity, you will be able to combine tools to successfully analyze data. Switching between spreadsheets and SQL can be challenging because they're so different, but once you're more used to both tools, you'll be able to use both more easily. This is important for tackling larger and more complex projects in your career as a data analyst.



**Hands-On Activity: Create temporary tables**
Practice Quiz • 1h • 2 total points

**Due** Jan 23, 11:59 PM WIB

**Congratulations! You passed!**

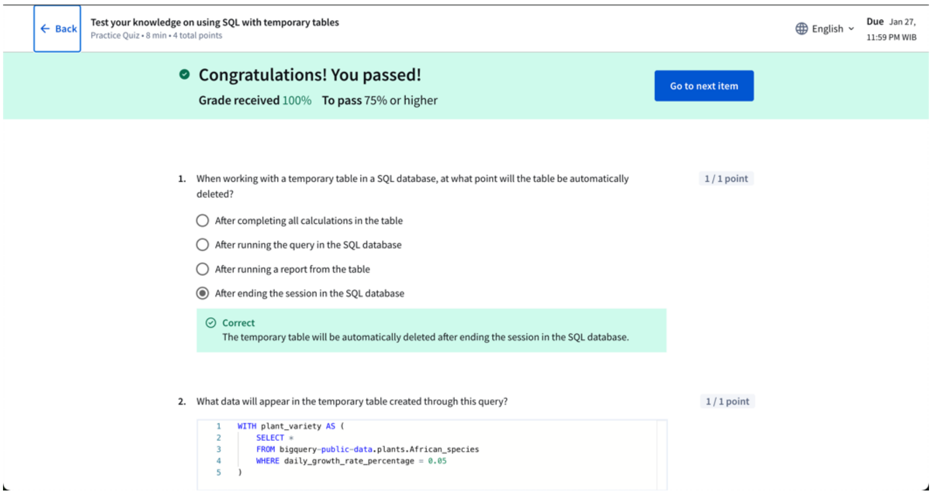
**Grade received** 100% **To pass** 100% or higher

### Activity Overview

As data calculations become more complicated, there are many components to keep track of, such as range, cost, time elements, products, and more. Some people use sticky notes for this, while others use checklists. In the data profession, a temporary table is just like a sticky note.

You learned about temporary tables in SQL in earlier lessons, so take a moment to review. **Temporary tables**, or **temp tables**, store subsets of data from standard data tables for a certain period of time. Temp tables allow you to run calculations in temporary data tables without needing to make modifications to the primary tables in your database. Because they are temporary, they are automatically deleted at the end of your SQL session.

By the end of this activity, you will have gained more experience creating temp tables and using them to run queries.



**Test your knowledge on using SQL with temporary tables**
Practice Quiz • 8 min • 4 total points

**Due** Jan 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 75% or higher

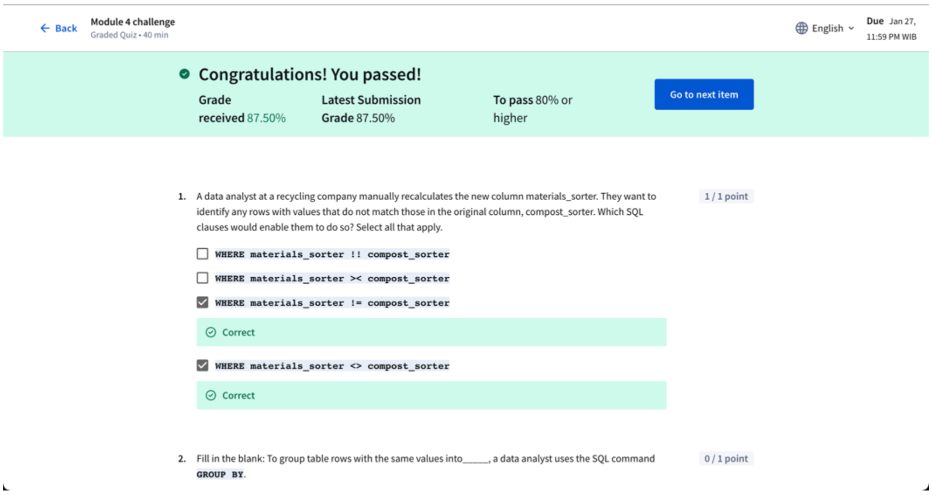
1. When working with a temporary table in a SQL database, at what point will the table be automatically deleted? 1 / 1 point

○ After completing all calculations in the table
○ After running the query in the SQL database
○ After running a report from the table
● After ending the session in the SQL database

Correct
The temporary table will be automatically deleted after ending the session in the SQL database.

2. What data will appear in the temporary table created through this query? 1 / 1 point

```
WITH plant_variety AS (
    SELECT *
    FROM bigquery-public-data.plants.African_species
    WHERE daily_growth_rate_percentage = 0.05
)
```



**Module 4 challenge**
Graded Quiz • 40 min

**Due** Jan 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 87.50% **Latest Submission Grade** 87.50% **To pass** 80% or higher

1. A data analyst at a recycling company manually recalculates the new column materials_sorter. They want to identify any rows with values that do not match those in the original column, compost_sorter. Which SQL clauses would enable them to do so? Select all that apply. 1 / 1 point

☐ `WHERE materials_sorter !! compost_sorter`
☐ `WHERE materials_sorter >< compost_sorter`
☑ `WHERE materials_sorter != compost_sorter`

Correct

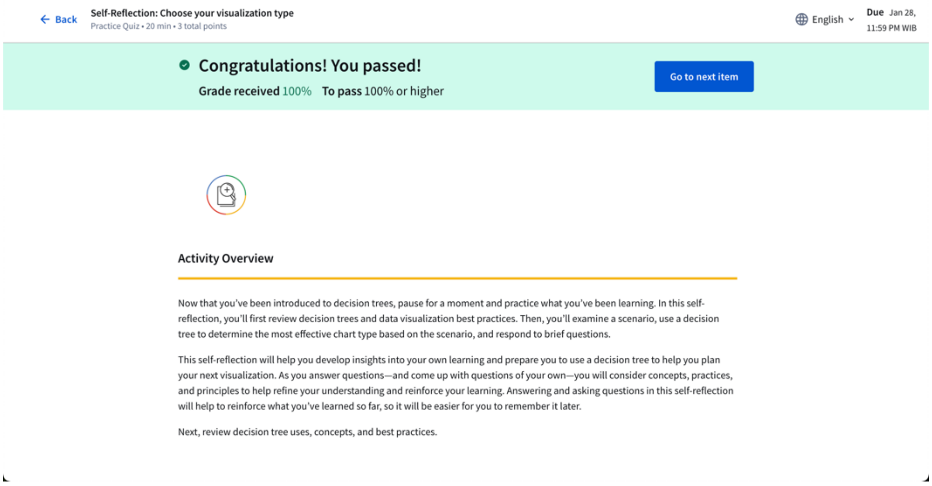
☑ `WHERE materials_sorter <> compost_sorter`

Correct

2. Fill in the blank: To group table rows with the same values into_____, a data analyst uses the SQL command `GROUP BY`. 0 / 1 point

# Course 6: Share Data Through the Art of Visualization

## Modul 1



**Self-Reflection: Choose your visualization type**
Practice Quiz • 20 min • 3 total points

**Due** Jan 28, 11:59 PM WIB

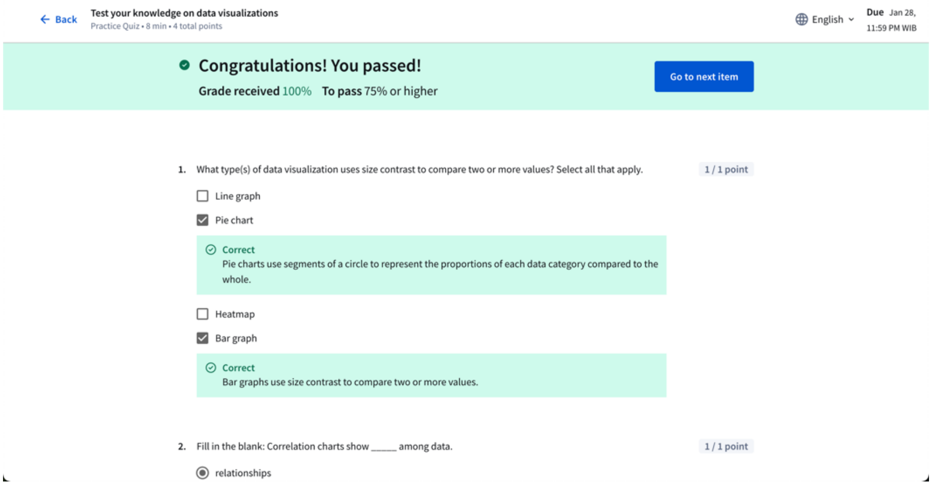
**Congratulations! You passed!**
**Grade received** 100%   **To pass** 100% or higher

**Activity Overview**

Now that you've been introduced to decision trees, pause for a moment and practice what you've been learning. In this self-reflection, you'll first review decision trees and data visualization best practices. Then, you'll examine a scenario, use a decision tree to determine the most effective chart type based on the scenario, and respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to use a decision tree to help you plan your next visualization. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. Answering and asking questions in this self-reflection will help to reinforce what you've learned so far, so it will be easier for you to remember it later.

Next, review decision tree uses, concepts, and best practices.



**Test your knowledge on data visualizations**
Practice Quiz • 8 min • 4 total points

**Due** Jan 28, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100%   **To pass** 75% or higher

1. What type(s) of data visualization uses size contrast to compare two or more values? Select all that apply. 1 / 1 point

☐ Line graph

☑ Pie chart

Correct
Pie charts use segments of a circle to represent the proportions of each data category compared to the whole.

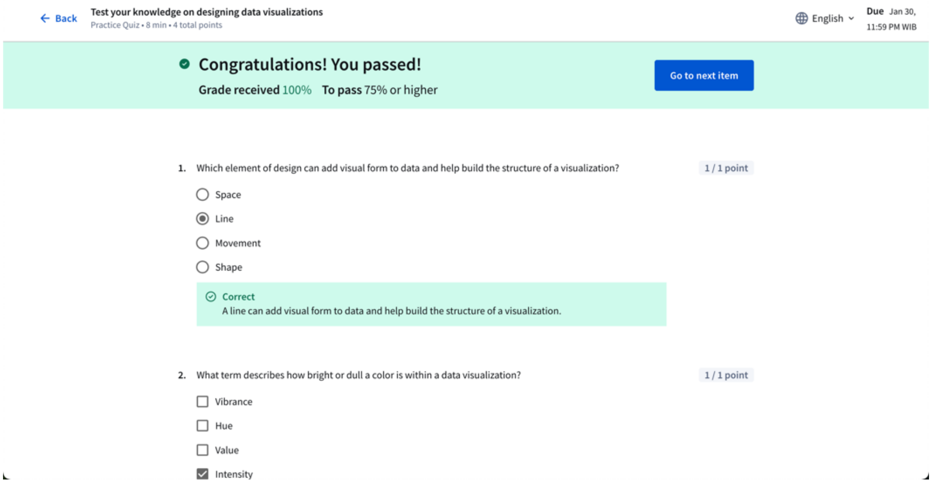
☐ Heatmap

☑ Bar graph

Correct
Bar graphs use size contrast to compare two or more values.

2. Fill in the blank: Correlation charts show _____ among data. 1 / 1 point

● relationships



**Test your knowledge on designing data visualizations**
Practice Quiz • 8 min • 4 total points

**Due** Jan 30, 11:59 PM WIB

**Congratulations! You passed!**
**Grade received** 100%   **To pass** 75% or higher

1. Which element of design can add visual form to data and help build the structure of a visualization? 1 / 1 point

○ Space

● Line

○ Movement

○ Shape

Correct
A line can add visual form to data and help build the structure of a visualization.

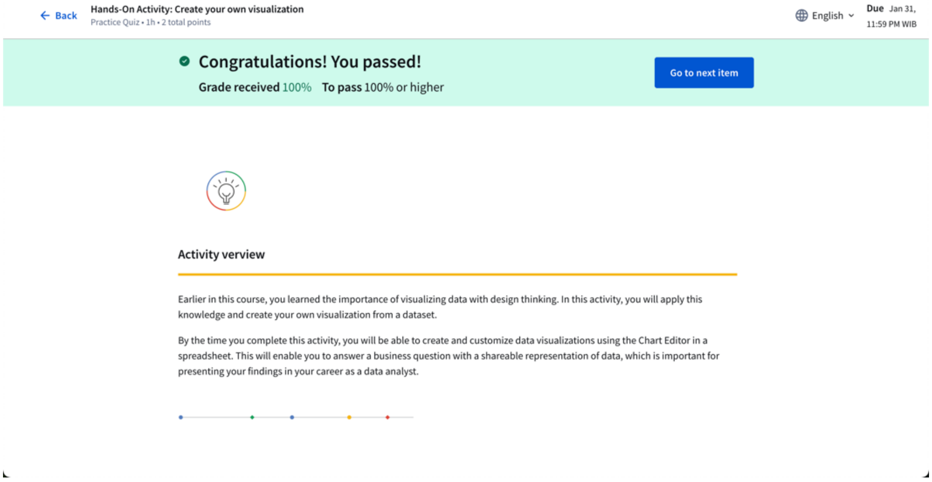
2. What term describes how bright or dull a color is within a data visualization? 1 / 1 point

☐ Vibrance

☐ Hue

☐ Value

☑ Intensity



**Hands-On Activity: Create your own visualization**
Practice Quiz • 1h • 2 total points

**Due** Jan 31, 11:59 PM WIB

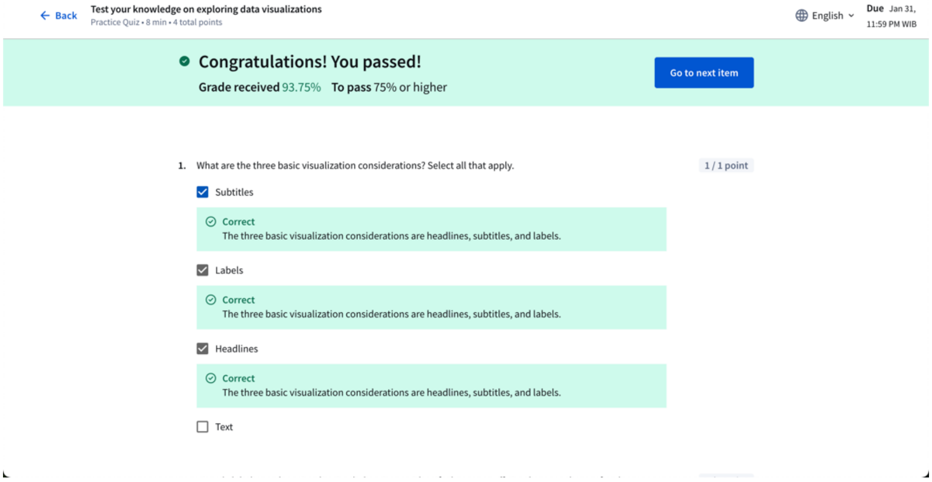
**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity verview**

Earlier in this course, you learned the importance of visualizing data with design thinking. In this activity, you will apply this knowledge and create your own visualization from a dataset.

By the time you complete this activity, you will be able to create and customize data visualizations using the Chart Editor in a spreadsheet. This will enable you to answer a business question with a shareable representation of data, which is important for presenting your findings in your career as a data analyst.



**Test your knowledge on exploring data visualizations**
Practice Quiz • 8 min • 4 total points

**Due** Jan 31, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 75% or higher

1. What are the three basic visualization considerations? Select all that apply. — 1 / 1 point

☑ Subtitles

✓ Correct
The three basic visualization considerations are headlines, subtitles, and labels.

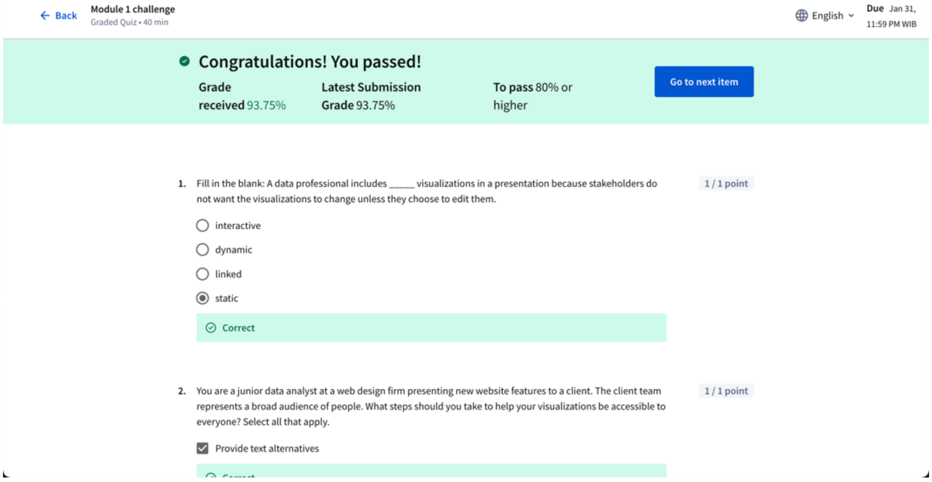
☑ Labels

✓ Correct
The three basic visualization considerations are headlines, subtitles, and labels.

☑ Headlines

✓ Correct
The three basic visualization considerations are headlines, subtitles, and labels.

☐ Text



**Module 1 challenge**
Graded Quiz • 40 min

**Due** Jan 31, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **Latest Submission Grade** 93.75% **To pass** 80% or higher

1. Fill in the blank: A data professional includes _____ visualizations in a presentation because stakeholders do not want the visualizations to change unless they choose to edit them. — 1 / 1 point

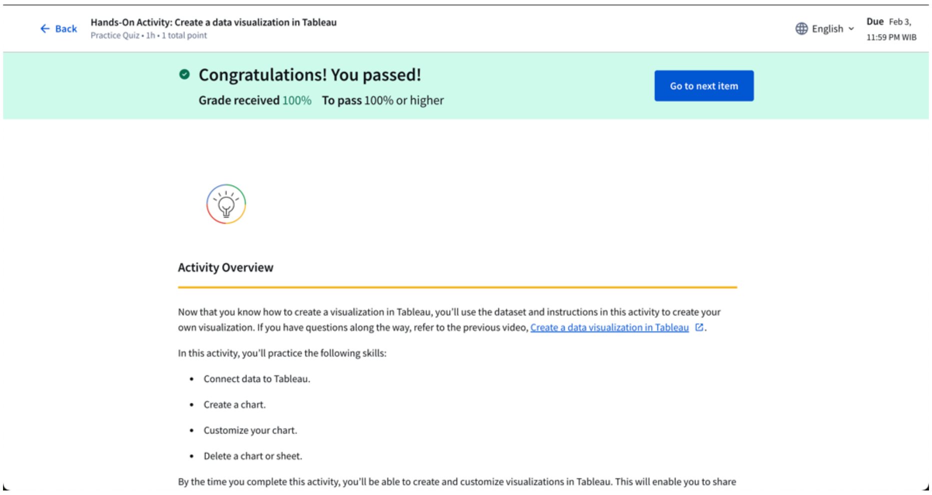
○ interactive
○ dynamic
○ linked
● static

✓ Correct

2. You are a junior data analyst at a web design firm presenting new website features to a client. The client team represents a broad audience of people. What steps should you take to help your visualizations be accessible to everyone? Select all that apply. — 1 / 1 point

☑ Provide text alternatives

# Modul 2



**Hands-On Activity: Create a data visualization in Tableau**
Practice Quiz • 1h • 1 total point

**Due** Feb 3, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

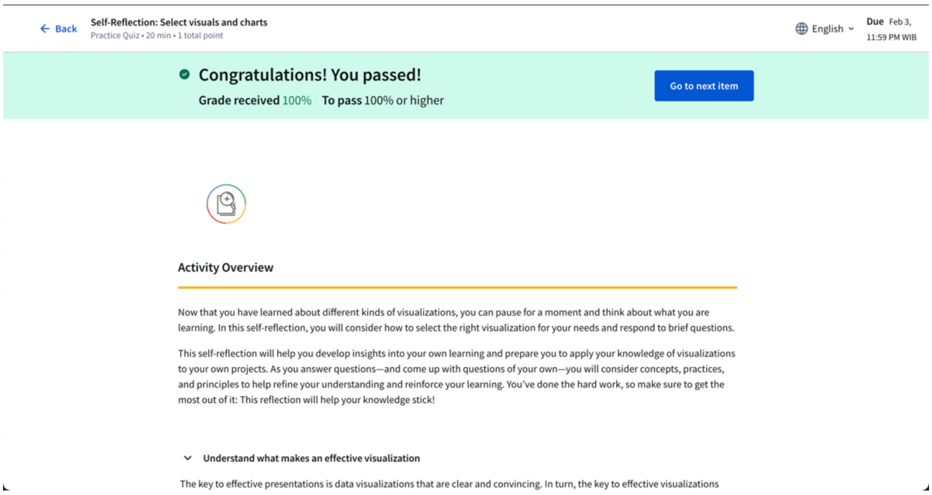
**Activity Overview**

Now that you know how to create a visualization in Tableau, you'll use the dataset and instructions in this activity to create your own visualization. If you have questions along the way, refer to the previous video, Create a data visualization in Tableau.

In this activity, you'll practice the following skills:

- Connect data to Tableau.
- Create a chart.
- Customize your chart.
- Delete a chart or sheet.

By the time you complete this activity, you'll be able to create and customize visualizations in Tableau. This will enable you to share



**Self-Reflection: Select visuals and charts**
Practice Quiz • 20 min • 1 total point

**Due** Feb 3, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

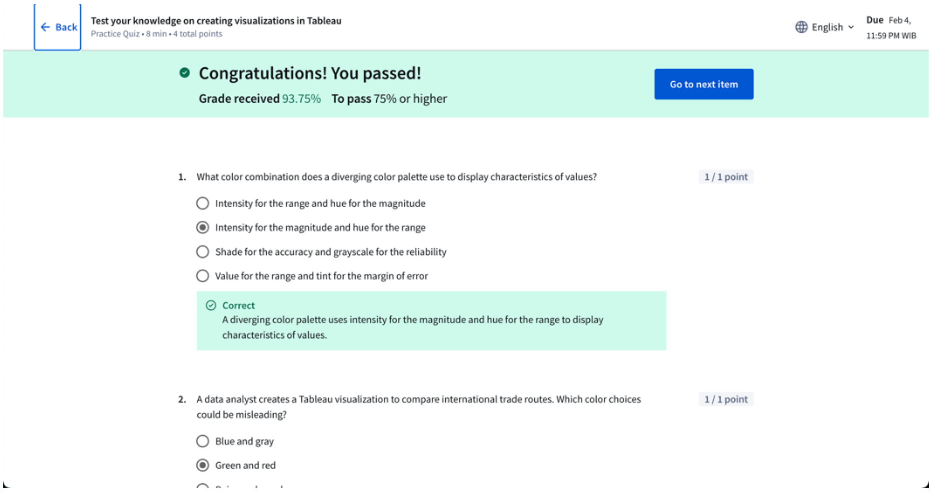
**Activity Overview**

Now that you have learned about different kinds of visualizations, you can pause for a moment and think about what you are learning. In this self-reflection, you will consider how to select the right visualization for your needs and respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to apply your knowledge of visualizations to your own projects. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

**Understand what makes an effective visualization**

The key to effective presentations is data visualizations that are clear and convincing. In turn, the key to effective visualizations



**Test your knowledge on creating visualizations in Tableau**
Practice Quiz • 8 min • 4 total points

**Due** Feb 4, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 75% or higher

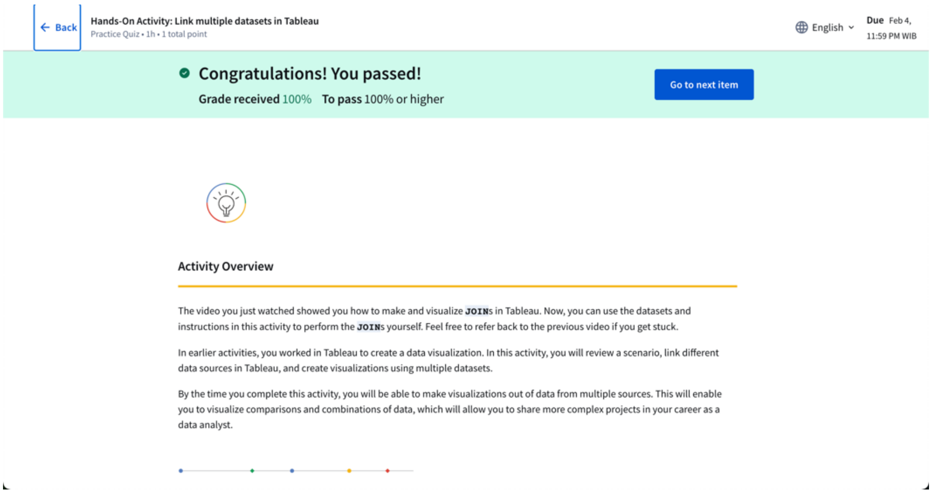
1. What color combination does a diverging color palette use to display characteristics of values? 1 / 1 point

○ Intensity for the range and hue for the magnitude
● Intensity for the magnitude and hue for the range
○ Shade for the accuracy and grayscale for the reliability
○ Value for the range and tint for the margin of error

✓ Correct
A diverging color palette uses intensity for the magnitude and hue for the range to display characteristics of values.

2. A data analyst creates a Tableau visualization to compare international trade routes. Which color choices could be misleading? 1 / 1 point

○ Blue and gray
● Green and red



Hands-On Activity: Link multiple datasets in Tableau
Practice Quiz • 1h • 1 total point

Due Feb 4, 11:59 PM WIB

Congratulations! You passed!

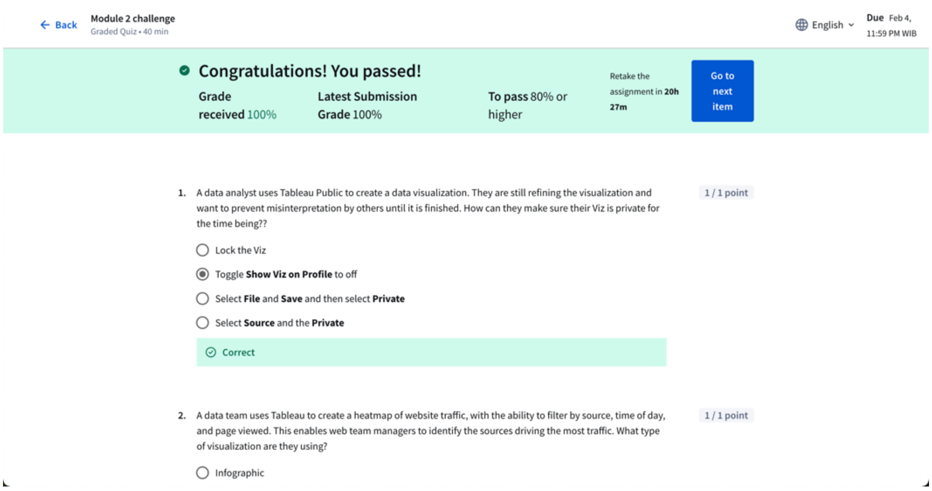
Grade received 100% To pass 100% or higher

**Activity Overview**

The video you just watched showed you how to make and visualize `JOIN`s in Tableau. Now, you can use the datasets and instructions in this activity to perform the `JOIN`s yourself. Feel free to refer back to the previous video if you get stuck.

In earlier activities, you worked in Tableau to create a data visualization. In this activity, you will review a scenario, link different data sources in Tableau, and create visualizations using multiple datasets.

By the time you complete this activity, you will be able to make visualizations out of data from multiple sources. This will enable you to visualize comparisons and combinations of data, which will allow you to share more complex projects in your career as a data analyst.



Module 2 challenge
Graded Quiz • 40 min

Due Feb 4, 11:59 PM WIB

Congratulations! You passed!

Grade received 100% Latest Submission Grade 100% To pass 80% or higher

Retake the assignment in **20h 27m**

1. A data analyst uses Tableau Public to create a data visualization. They are still refining the visualization and want to prevent misinterpretation by others until it is finished. How can they make sure their Viz is private for the time being?? 1 / 1 point

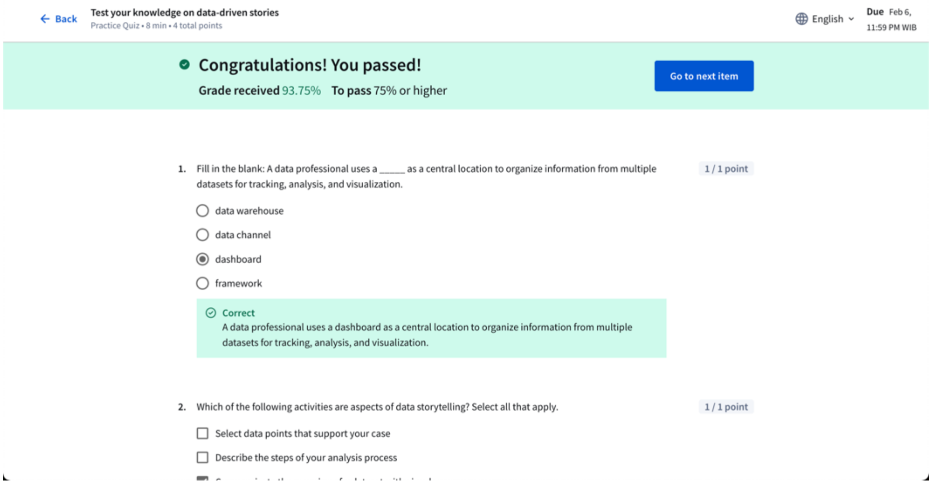
○ Lock the Viz
● Toggle **Show Viz on Profile** to off
○ Select **File** and **Save** and then select **Private**
○ Select **Source** and the **Private**

Correct

2. A data team uses Tableau to create a heatmap of website traffic, with the ability to filter by source, time of day, and page viewed. This enables web team managers to identify the sources driving the most traffic. What type of visualization are they using? 1 / 1 point

○ Infographic

Modul 3



Test your knowledge on data-driven stories
Practice Quiz • 8 min • 4 total points

Due Feb 6, 11:59 PM WIB

Congratulations! You passed!

Grade received 93.75% To pass 75% or higher

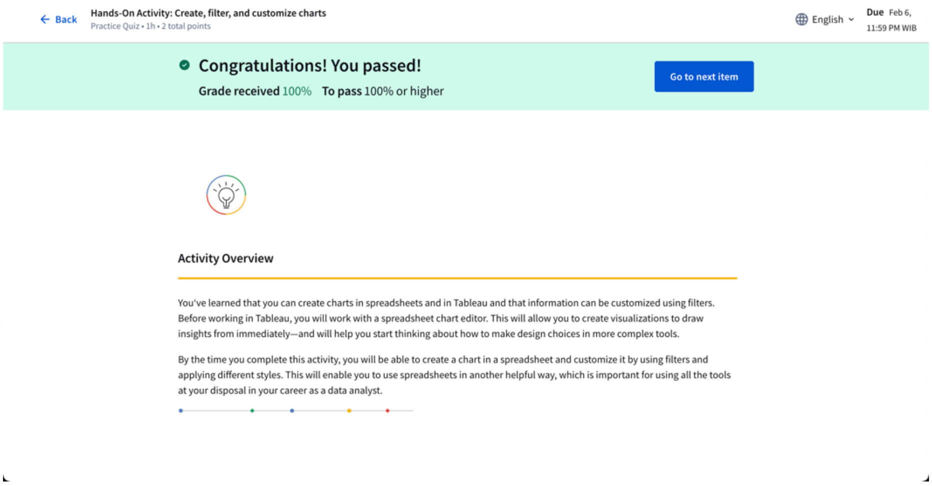
1. Fill in the blank: A data professional uses a _____ as a central location to organize information from multiple datasets for tracking, analysis, and visualization. 1 / 1 point

○ data warehouse
○ data channel
● dashboard
○ framework

Correct
A data professional uses a dashboard as a central location to organize information from multiple datasets for tracking, analysis, and visualization.

2. Which of the following activities are aspects of data storytelling? Select all that apply. 1 / 1 point

☐ Select data points that support your case
☐ Describe the steps of your analysis process



**Hands-On Activity: Create, filter, and customize charts**
Practice Quiz • 1h • 2 total points

**Due** Feb 6, 11:59 PM WIB

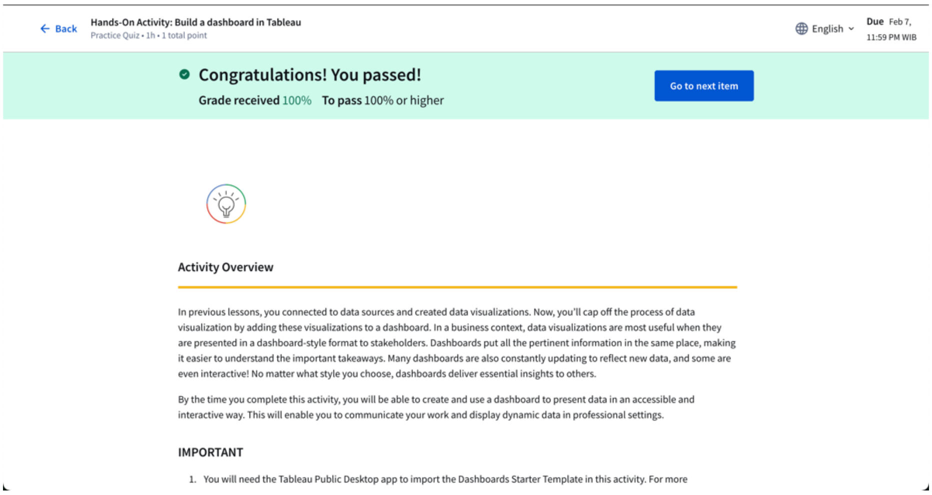
**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

**Activity Overview**

You've learned that you can create charts in spreadsheets and in Tableau and that information can be customized using filters. Before working in Tableau, you will work with a spreadsheet chart editor. This will allow you to create visualizations to draw insights from immediately—and will help you start thinking about how to make design choices in more complex tools.

By the time you complete this activity, you will be able to create a chart in a spreadsheet and customize it by using filters and applying different styles. This will enable you to use spreadsheets in another helpful way, which is important for using all the tools at your disposal in your career as a data analyst.



**Hands-On Activity: Build a dashboard in Tableau**
Practice Quiz • 1h • 1 total point

**Due** Feb 7, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

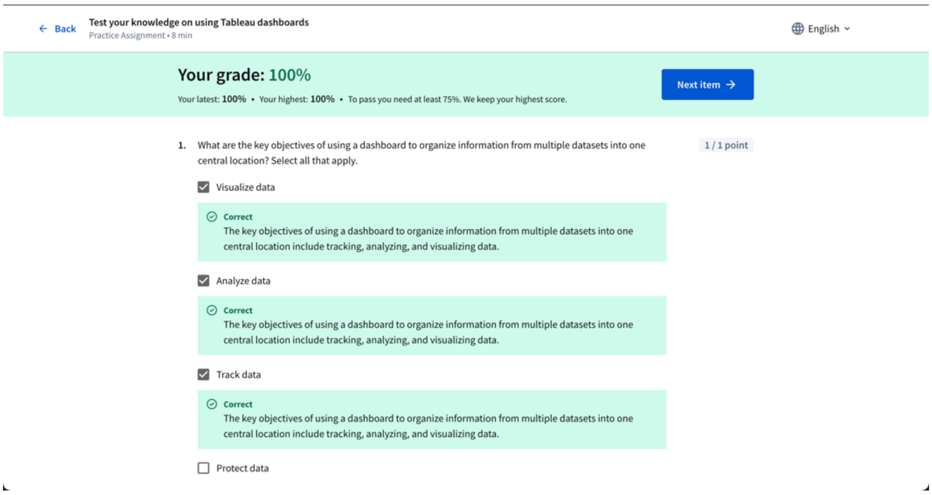
**Activity Overview**

In previous lessons, you connected to data sources and created data visualizations. Now, you'll cap off the process of data visualization by adding these visualizations to a dashboard. In a business context, data visualizations are most useful when they are presented in a dashboard-style format to stakeholders. Dashboards put all the pertinent information in the same place, making it easier to understand the important takeaways. Many dashboards are also constantly updating to reflect new data, and some are even interactive! No matter what style you choose, dashboards deliver essential insights to others.

By the time you complete this activity, you will be able to create and use a dashboard to present data in an accessible and interactive way. This will enable you to communicate your work and display dynamic data in professional settings.

**IMPORTANT**

1. You will need the Tableau Public Desktop app to import the Dashboards Starter Template in this activity. For more



**Test your knowledge on using Tableau dashboards**
Practice Assignment • 8 min

**Your grade:** 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

1. What are the key objectives of using a dashboard to organize information from multiple datasets into one central location? Select all that apply.

1 / 1 point

☑ Visualize data

✓ **Correct**
The key objectives of using a dashboard to organize information from multiple datasets into one central location include tracking, analyzing, and visualizing data.

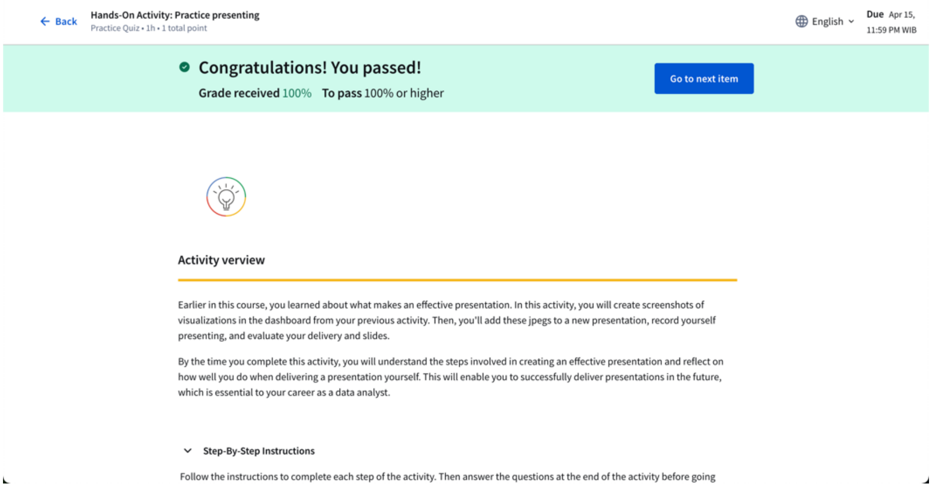
☑ Analyze data

✓ **Correct**
The key objectives of using a dashboard to organize information from multiple datasets into one central location include tracking, analyzing, and visualizing data.

☑ Track data

✓ **Correct**
The key objectives of using a dashboard to organize information from multiple datasets into one central location include tracking, analyzing, and visualizing data.

☐ Protect data



**Hands-On Activity: Practice presenting**
Practice Quiz • 1h • 1 total point

**Due** Apr 15, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

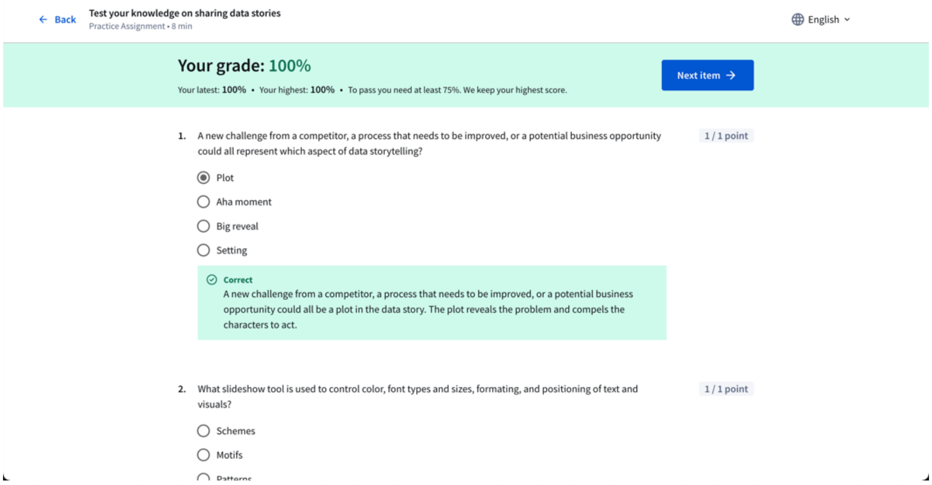
**Activity verview**

Earlier in this course, you learned about what makes an effective presentation. In this activity, you will create screenshots of visualizations in the dashboard from your previous activity. Then, you'll add these jpegs to a new presentation, record yourself presenting, and evaluate your delivery and slides.

By the time you complete this activity, you will understand the steps involved in creating an effective presentation and reflect on how well you do when delivering a presentation yourself. This will enable you to successfully deliver presentations in the future, which is essential to your career as a data analyst.

**Step-By-Step Instructions**

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going



**Test your knowledge on sharing data stories**
Practice Assignment • 8 min

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

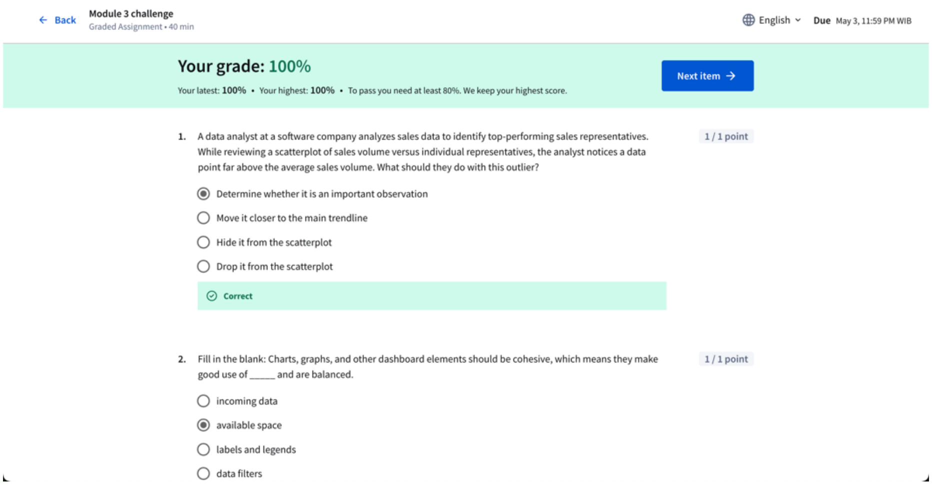
1. A new challenge from a competitor, a process that needs to be improved, or a potential business opportunity could all represent which aspect of data storytelling? 1 / 1 point

● Plot
○ Aha moment
○ Big reveal
○ Setting

Correct
A new challenge from a competitor, a process that needs to be improved, or a potential business opportunity could all be a plot in the data story. The plot reveals the problem and compels the characters to act.

2. What slideshow tool is used to control color, font types and sizes, formating, and positioning of text and visuals? 1 / 1 point

○ Schemes
○ Motifs
○ Patterns



**Module 3 challenge**
Graded Assignment • 40 min

**Due** May 3, 11:59 PM WIB

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

1. A data analyst at a software company analyzes sales data to identify top-performing sales representatives. While reviewing a scatterplot of sales volume versus individual representatives, the analyst notices a data point far above the average sales volume. What should they do with this outlier? 1 / 1 point

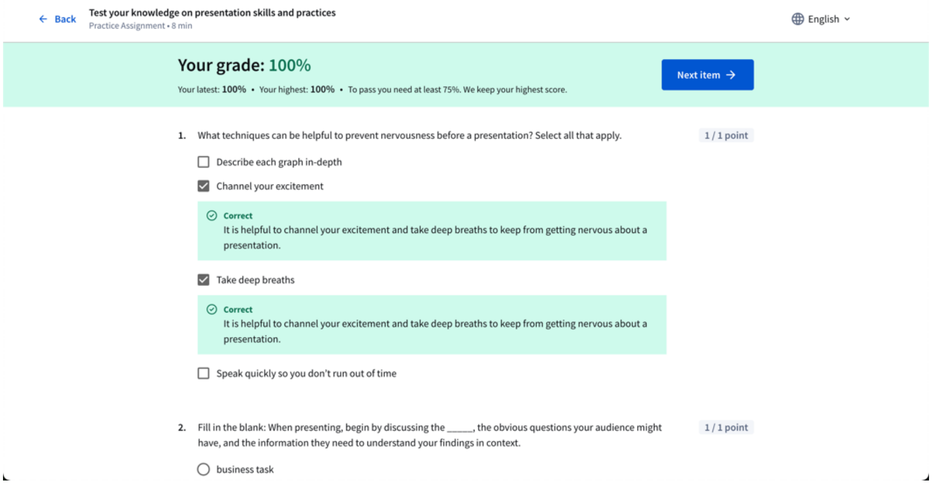
● Determine whether it is an important observation
○ Move it closer to the main trendline
○ Hide it from the scatterplot
○ Drop it from the scatterplot

Correct

2. Fill in the blank: Charts, graphs, and other dashboard elements should be cohesive, which means they make good use of _____ and are balanced. 1 / 1 point

○ incoming data
● available space
○ labels and legends
○ data filters

# Modul 4



**Test your knowledge on presentation skills and practices**
Practice Assignment • 8 min

**Your grade: 100%**
Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

1. What techniques can be helpful to prevent nervousness before a presentation? Select all that apply. — 1 / 1 point

☐ Describe each graph in-depth

☑ Channel your excitement

**Correct**
It is helpful to channel your excitement and take deep breaths to keep from getting nervous about a presentation.

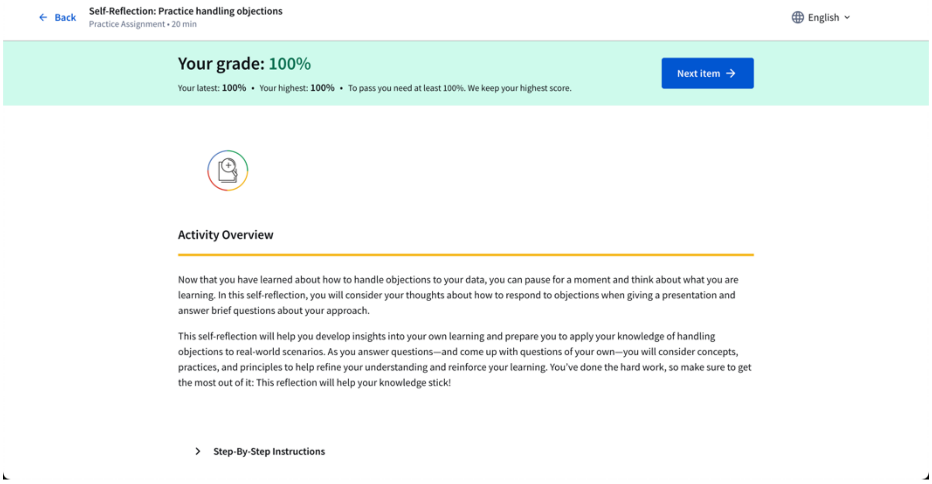
☑ Take deep breaths

**Correct**
It is helpful to channel your excitement and take deep breaths to keep from getting nervous about a presentation.

☐ Speak quickly so you don't run out of time

2. Fill in the blank: When presenting, begin by discussing the _____, the obvious questions your audience might have, and the information they need to understand your findings in context. — 1 / 1 point

○ business task



**Self-Reflection: Practice handling objections**
Practice Assignment • 20 min

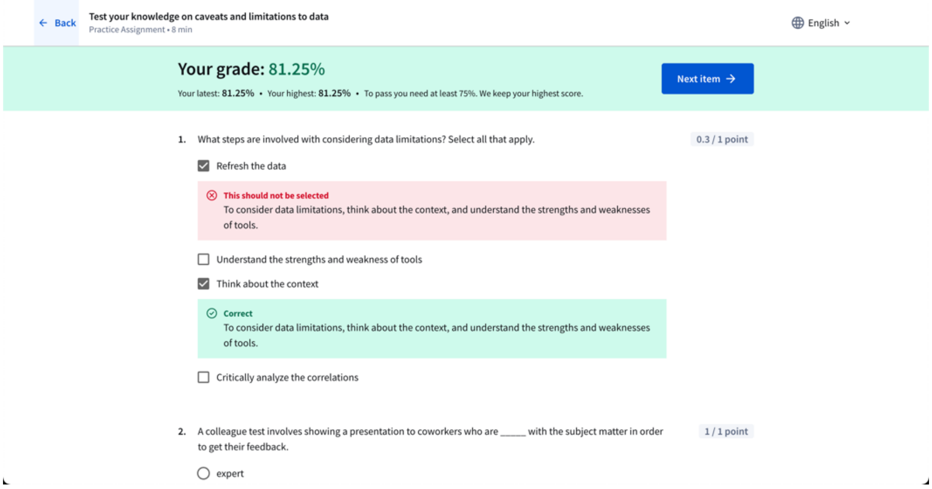
**Your grade: 100%**
Your latest: **100%** • Your highest: **100%** • To pass you need at least 100%. We keep your highest score.

**Activity Overview**

Now that you have learned about how to handle objections to your data, you can pause for a moment and think about what you are learning. In this self-reflection, you will consider your thoughts about how to respond to objections when giving a presentation and answer brief questions about your approach.

This self-reflection will help you develop insights into your own learning and prepare you to apply your knowledge of handling objections to real-world scenarios. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

> **Step-By-Step Instructions**



**Test your knowledge on caveats and limitations to data**
Practice Assignment • 8 min

**Your grade: 81.25%**
Your latest: **81.25%** • Your highest: **81.25%** • To pass you need at least 75%. We keep your highest score.

1. What steps are involved with considering data limitations? Select all that apply. — 0.3 / 1 point

☑ Refresh the data

**This should not be selected**
To consider data limitations, think about the context, and understand the strengths and weaknesses of tools.

☐ Understand the strengths and weakness of tools

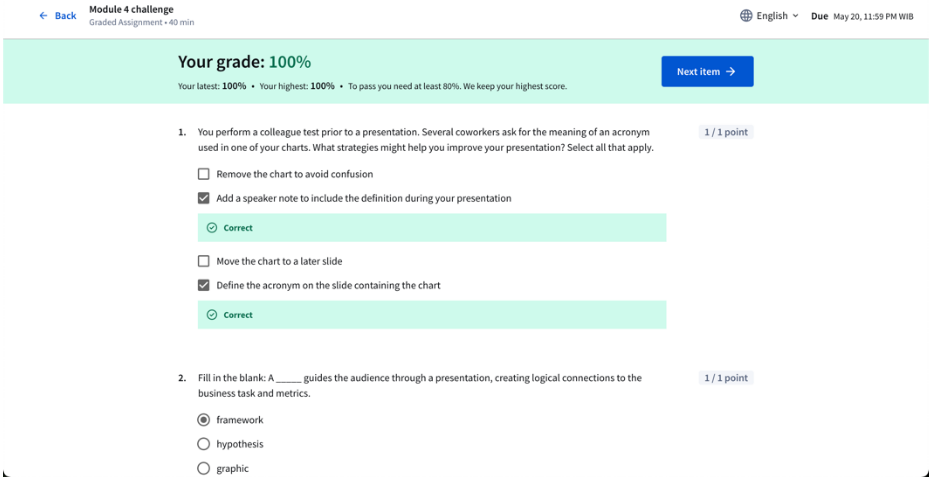
☑ Think about the context

**Correct**
To consider data limitations, think about the context, and understand the strengths and weaknesses of tools.

☐ Critically analyze the correlations

2. A colleague test involves showing a presentation to coworkers who are _____ with the subject matter in order to get their feedback. — 1 / 1 point

○ expert



**Module 4 challenge**
Graded Assignment • 40 min

**Due** May 20, 11:59 PM WIB

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

**1.** You perform a colleague test prior to a presentation. Several coworkers ask for the meaning of an acronym used in one of your charts. What strategies might help you improve your presentation? Select all that apply.

1 / 1 point

☐ Remove the chart to avoid confusion

☑ Add a speaker note to include the definition during your presentation

✓ **Correct**

☐ Move the chart to a later slide

☑ Define the acronym on the slide containing the chart

✓ **Correct**

**2.** Fill in the blank: A _____ guides the audience through a presentation, creating logical connections to the business task and metrics.

1 / 1 point

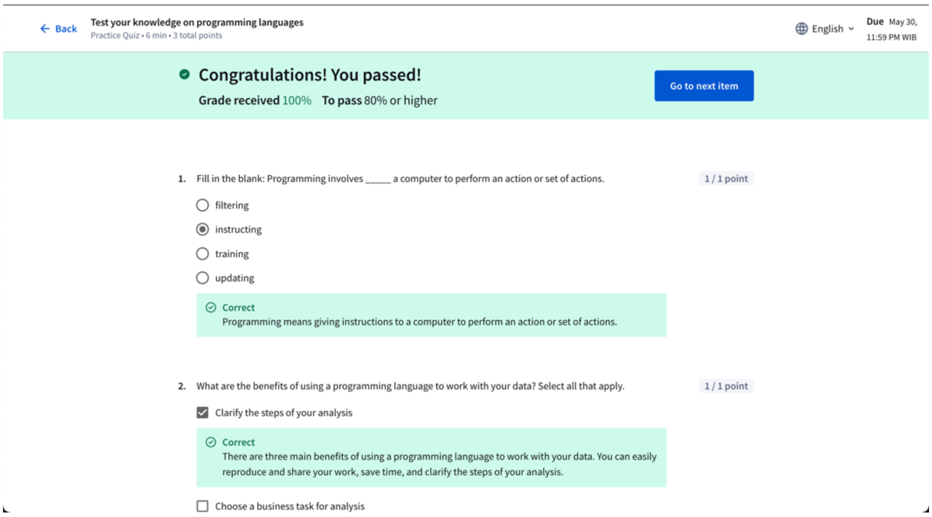
● framework

○ hypothesis

○ graphic

# Course 7: Data Analysis with R Programming

## Modul 1



**Test your knowledge on programming languages**
Practice Quiz • 6 min • 3 total points

**Due** May 30, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. Fill in the blank: Programming involves _____ a computer to perform an action or set of actions. 1 / 1 point

○ filtering
● instructing
○ training
○ updating

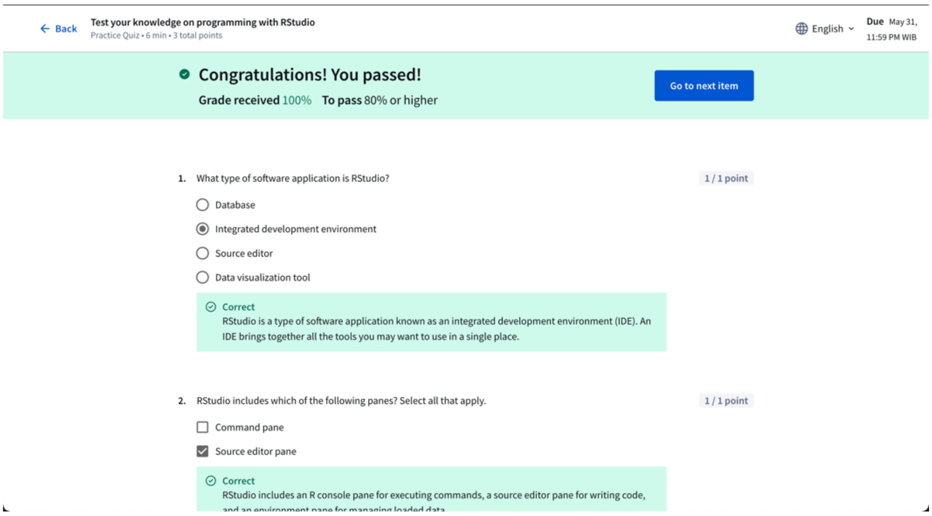
Correct
Programming means giving instructions to a computer to perform an action or set of actions.

2. What are the benefits of using a programming language to work with your data? Select all that apply. 1 / 1 point

☑ Clarify the steps of your analysis

Correct
There are three main benefits of using a programming language to work with your data. You can easily reproduce and share your work, save time, and clarify the steps of your analysis.

☐ Choose a business task for analysis



**Test your knowledge on programming with RStudio**
Practice Quiz • 6 min • 3 total points

**Due** May 31, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. What type of software application is RStudio? 1 / 1 point

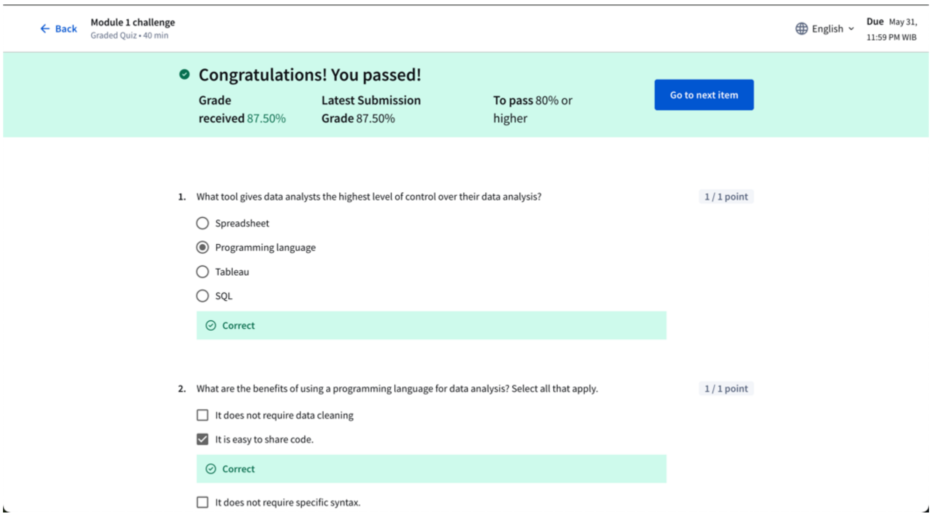
○ Database
● Integrated development environment
○ Source editor
○ Data visualization tool

Correct
RStudio is a type of software application known as an integrated development environment (IDE). An IDE brings together all the tools you may want to use in a single place.

2. RStudio includes which of the following panes? Select all that apply. 1 / 1 point

☐ Command pane
☑ Source editor pane

Correct
RStudio includes an R console pane for executing commands, a source editor pane for writing code, and an environment pane for managing loaded data.



**Module 1 challenge**
Graded Quiz • 40 min

**Due** May 31, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 87.50% **Latest Submission Grade** 87.50% **To pass** 80% or higher

1. What tool gives data analysts the highest level of control over their data analysis? 1 / 1 point

○ Spreadsheet
● Programming language
○ Tableau
○ SQL

Correct

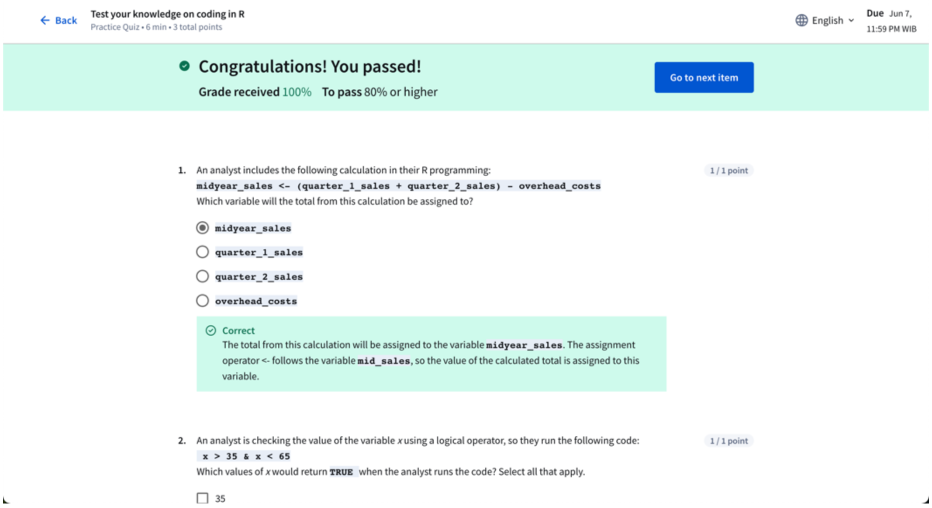
2. What are the benefits of using a programming language for data analysis? Select all that apply. 1 / 1 point

☐ It does not require data cleaning
☑ It is easy to share code.

Correct

☐ It does not require specific syntax.

# Modul 2



**Test your knowledge on coding in R**
Practice Quiz • 6 min • 3 total points

**Due** Jun 7, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. An analyst includes the following calculation in their R programming: 1 / 1 point

`midyear_sales <- (quarter_1_sales + quarter_2_sales) - overhead_costs`

Which variable will the total from this calculation be assigned to?

● `midyear_sales`
○ `quarter_1_sales`
○ `quarter_2_sales`
○ `overhead_costs`

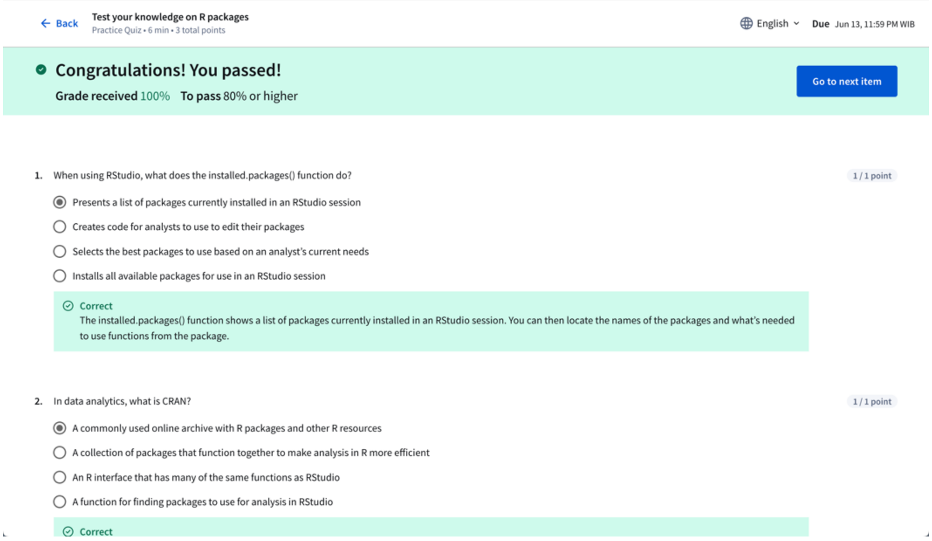
Correct
The total from this calculation will be assigned to the variable `midyear_sales`. The assignment operator <- follows the variable `mid_sales`, so the value of the calculated total is assigned to this variable.

2. An analyst is checking the value of the variable *x* using a logical operator, so they run the following code: 1 / 1 point

`x > 35 & x < 65`

Which values of *x* would return `TRUE` when the analyst runs the code? Select all that apply.

☐ 35



**Test your knowledge on R packages**
Practice Quiz • 6 min • 3 total points

**Due** Jun 13, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. When using RStudio, what does the installed.packages() function do? 1 / 1 point

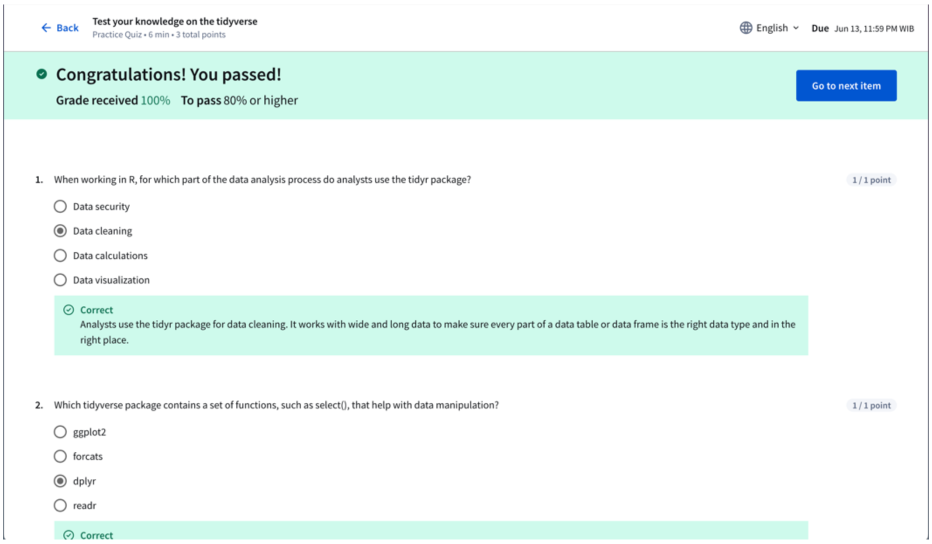
● Presents a list of packages currently installed in an RStudio session
○ Creates code for analysts to use to edit their packages
○ Selects the best packages to use based on an analyst's current needs
○ Installs all available packages for use in an RStudio session

Correct
The installed.packages() function shows a list of packages currently installed in an RStudio session. You can then locate the names of the packages and what's needed to use functions from the package.

2. In data analytics, what is CRAN? 1 / 1 point

● A commonly used online archive with R packages and other R resources
○ A collection of packages that function together to make analysis in R more efficient
○ An R interface that has many of the same functions as RStudio
○ A function for finding packages to use for analysis in RStudio

Correct



**Test your knowledge on the tidyverse**
Practice Quiz • 6 min • 3 total points

**Due** Jun 13, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. When working in R, for which part of the data analysis process do analysts use the tidyr package? 1 / 1 point

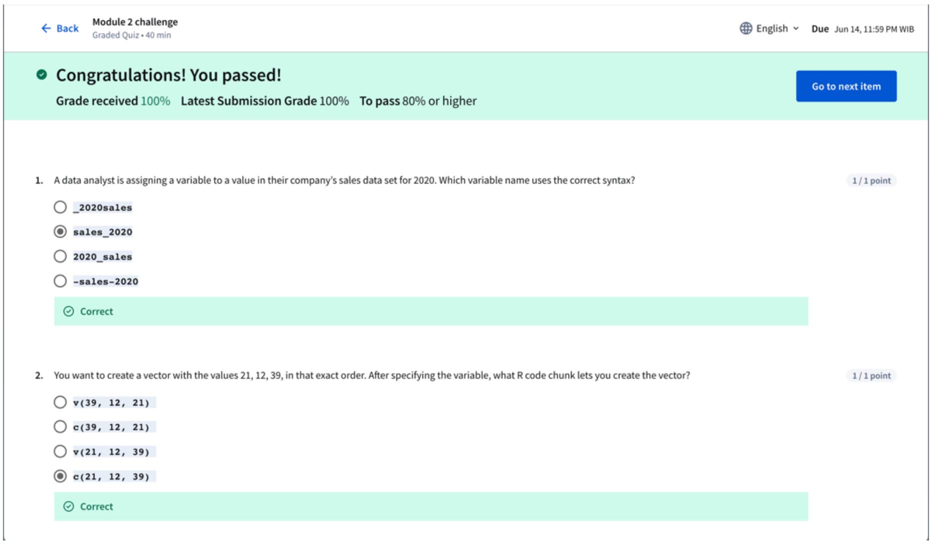
○ Data security
● Data cleaning
○ Data calculations
○ Data visualization

Correct
Analysts use the tidyr package for data cleaning. It works with wide and long data to make sure every part of a data table or data frame is the right data type and in the right place.

2. Which tidyverse package contains a set of functions, such as select(), that help with data manipulation? 1 / 1 point

○ ggplot2
○ forcats
● dplyr
○ readr

Correct



**Module 2 challenge**
Graded Quiz • 40 min

Due Jun 14, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **Latest Submission Grade** 100% **To pass** 80% or higher

1. A data analyst is assigning a variable to a value in their company's sales data set for 2020. Which variable name uses the correct syntax?

1 / 1 point

○ `_2020sales`
● `sales_2020`
○ `2020_sales`
○ `-sales-2020`

✓ Correct

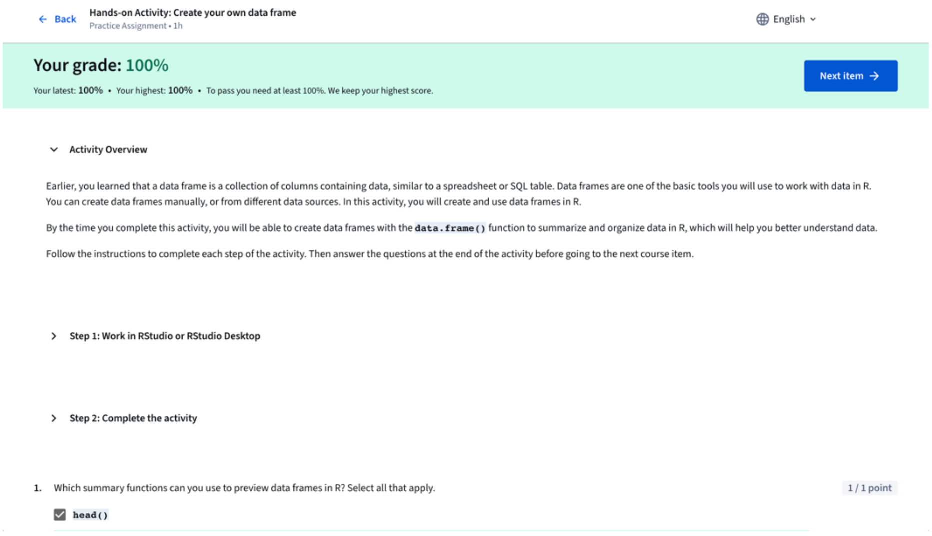
2. You want to create a vector with the values 21, 12, 39, in that exact order. After specifying the variable, what R code chunk lets you create the vector?

1 / 1 point

○ `v(39, 12, 21)`
○ `c(39, 12, 21)`
○ `v(21, 12, 39)`
● `c(21, 12, 39)`

✓ Correct

# Modul 3



**Hands-on Activity: Create your own data frame**
Practice Assignment • 1h

**Your grade:** 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 100%. We keep your highest score.

**Activity Overview**

Earlier, you learned that a data frame is a collection of columns containing data, similar to a spreadsheet or SQL table. Data frames are one of the basic tools you will use to work with data in R. You can create data frames manually, or from different data sources. In this activity, you will create and use data frames in R.

By the time you complete this activity, you will be able to create data frames with the `data.frame()` function to summarize and organize data in R, which will help you better understand data.

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.

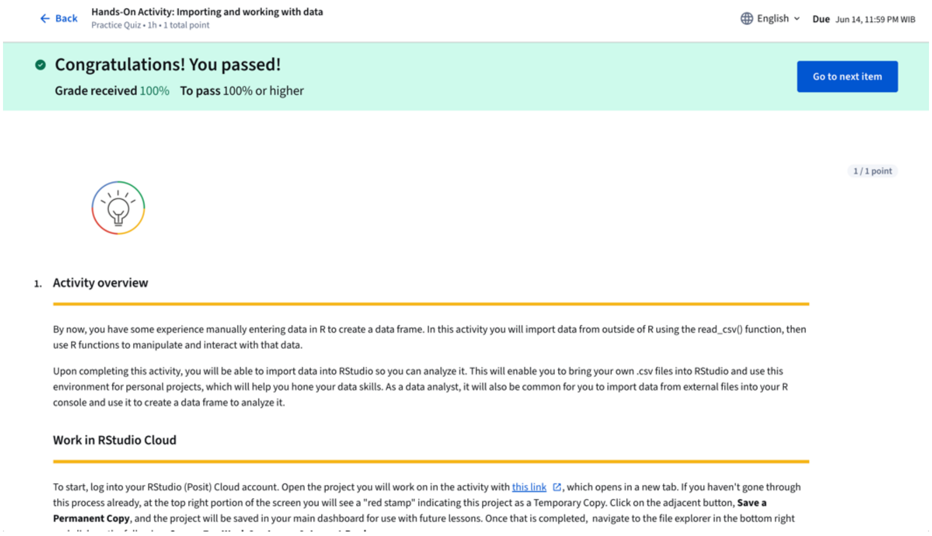
**Step 1: Work in RStudio or RStudio Desktop**

**Step 2: Complete the activity**

1. Which summary functions can you use to preview data frames in R? Select all that apply.

1 / 1 point

☑ `head()`



**Hands-On Activity: Importing and working with data**
Practice Quiz • 1h • 1 total point

Due Jun 14, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

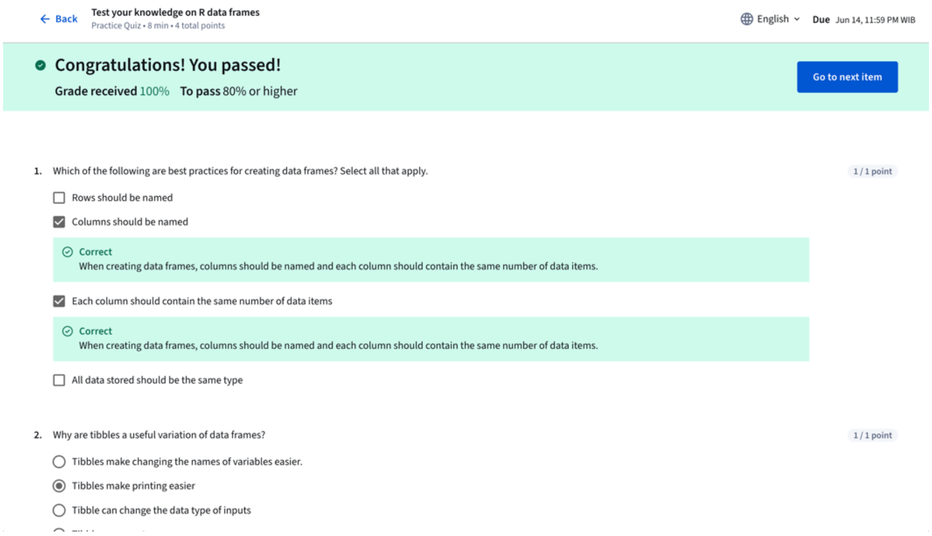
1. **Activity overview**

By now, you have some experience manually entering data in R to create a data frame. In this activity you will import data from outside of R using the read_csv() function, then use R functions to manipulate and interact with that data.

Upon completing this activity, you will be able to import data into RStudio so you can analyze it. This will enable you to bring your own .csv files into RStudio and use this environment for personal projects, which will help you hone your data skills. As a data analyst, it will also be common for you to import data from external files into your R console and use it to create a data frame to analyze it.

**Work in RStudio Cloud**

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right



**Test your knowledge on R data frames**
Practice Quiz • 8 min • 4 total points

**Due** Jun 14, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 80% or higher

1. Which of the following are best practices for creating data frames? Select all that apply. — 1 / 1 point

☐ Rows should be named

☑ Columns should be named

Correct
When creating data frames, columns should be named and each column should contain the same number of data items.

☑ Each column should contain the same number of data items

Correct
When creating data frames, columns should be named and each column should contain the same number of data items.

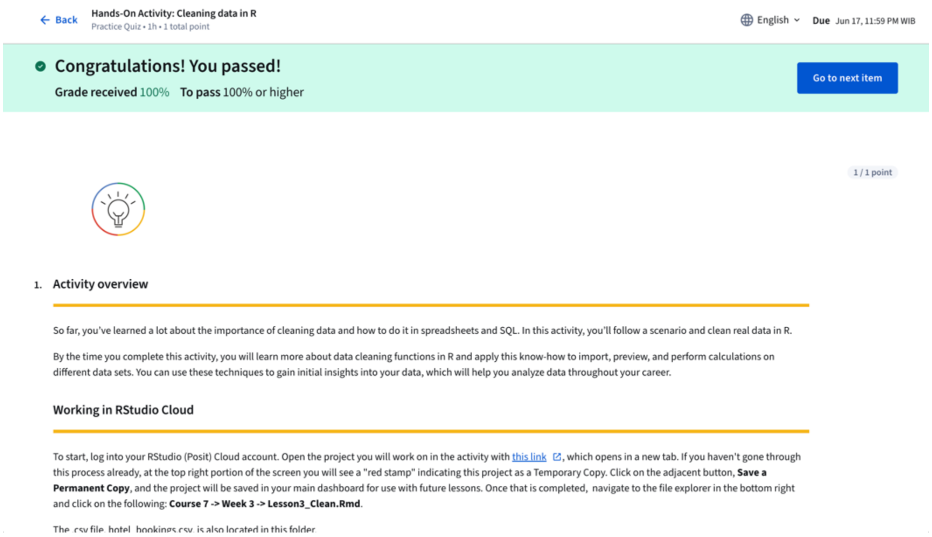
☐ All data stored should be the same type

2. Why are tibbles a useful variation of data frames? — 1 / 1 point

○ Tibbles make changing the names of variables easier.

● Tibbles make printing easier

○ Tibble can change the data type of inputs



**Hands-On Activity: Cleaning data in R**
Practice Quiz • 1h • 1 total point

**Due** Jun 17, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

1. **Activity overview**

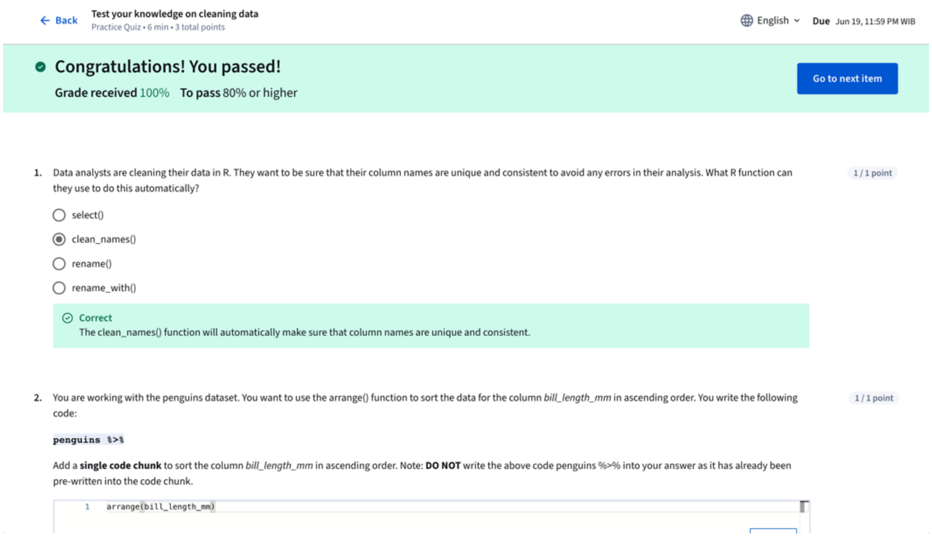
So far, you've learned a lot about the importance of cleaning data and how to do it in spreadsheets and SQL. In this activity, you'll follow a scenario and clean real data in R.

By the time you complete this activity, you will learn more about data cleaning functions in R and apply this know-how to import, preview, and perform calculations on different data sets. You can use these techniques to gain initial insights into your data, which will help you analyze data throughout your career.

**Working in RStudio Cloud**

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click on the following: **Course 7 -> Week 3 -> Lesson3_Clean.Rmd**.

The .csv file, hotel_bookings.csv, is also located in this folder.



**Test your knowledge on cleaning data**
Practice Quiz • 6 min • 3 total points

**Due** Jun 19, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 80% or higher

1. Data analysts are cleaning their data in R. They want to be sure that their column names are unique and consistent to avoid any errors in their analysis. What R function can they use to do this automatically? — 1 / 1 point

○ select()

● clean_names()

○ rename()

○ rename_with()

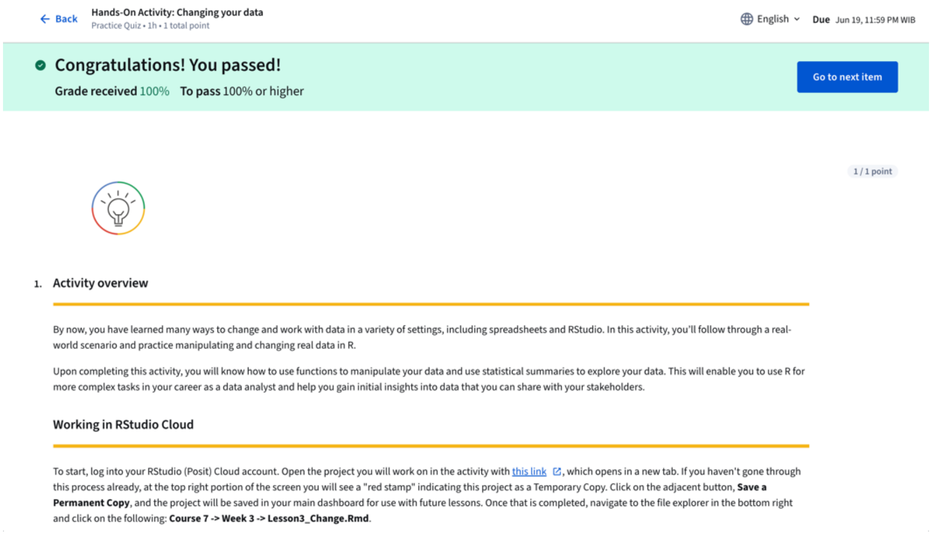
Correct
The clean_names() function will automatically make sure that column names are unique and consistent.

2. You are working with the penguins dataset. You want to use the arrange() function to sort the data for the column *bill_length_mm* in ascending order. You write the following code: — 1 / 1 point

```
penguins %>%
```

Add a **single code chunk** to sort the column *bill_length_mm* in ascending order. Note: **DO NOT** write the above code penguins %>% into your answer as it has already been pre-written into the code chunk.

```
arrange(bill_length_mm)
```



**Hands-On Activity: Changing your data**
Practice Quiz • 1h • 1 total point

Due Jun 19, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

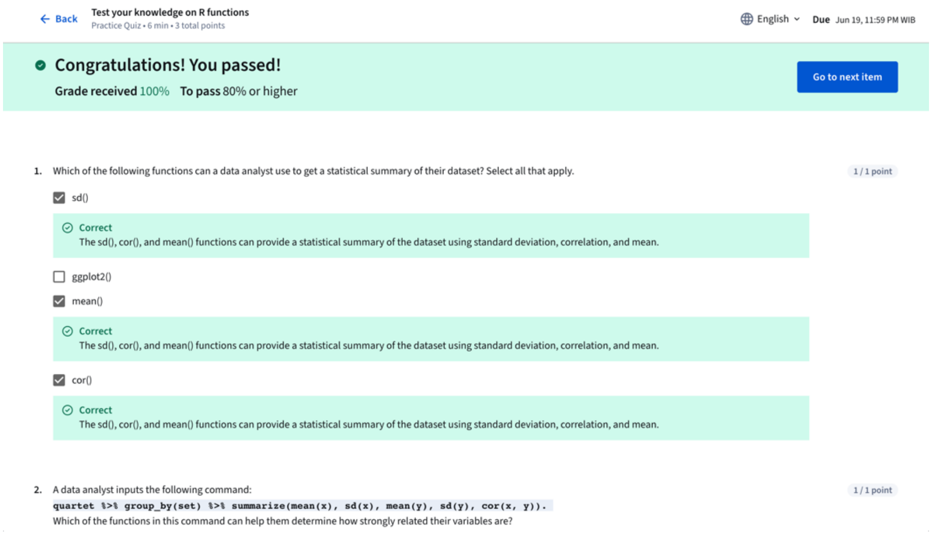
1. **Activity overview**

By now, you have learned many ways to change and work with data in a variety of settings, including spreadsheets and RStudio. In this activity, you'll follow through a real-world scenario and practice manipulating and changing real data in R.

Upon completing this activity, you will know how to use functions to manipulate your data and use statistical summaries to explore your data. This will enable you to use R for more complex tasks in your career as a data analyst and help you gain initial insights into data that you can share with your stakeholders.

**Working in RStudio Cloud**

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click on the following: **Course 7 -> Week 3 -> Lesson3_Change.Rmd**.



**Test your knowledge on R functions**
Practice Quiz • 6 min • 3 total points

Due Jun 19, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. Which of the following functions can a data analyst use to get a statistical summary of their dataset? Select all that apply. 1 / 1 point

☑ sd()

Correct
The sd(), cor(), and mean() functions can provide a statistical summary of the dataset using standard deviation, correlation, and mean.

☐ ggplot2()

☑ mean()

Correct
The sd(), cor(), and mean() functions can provide a statistical summary of the dataset using standard deviation, correlation, and mean.

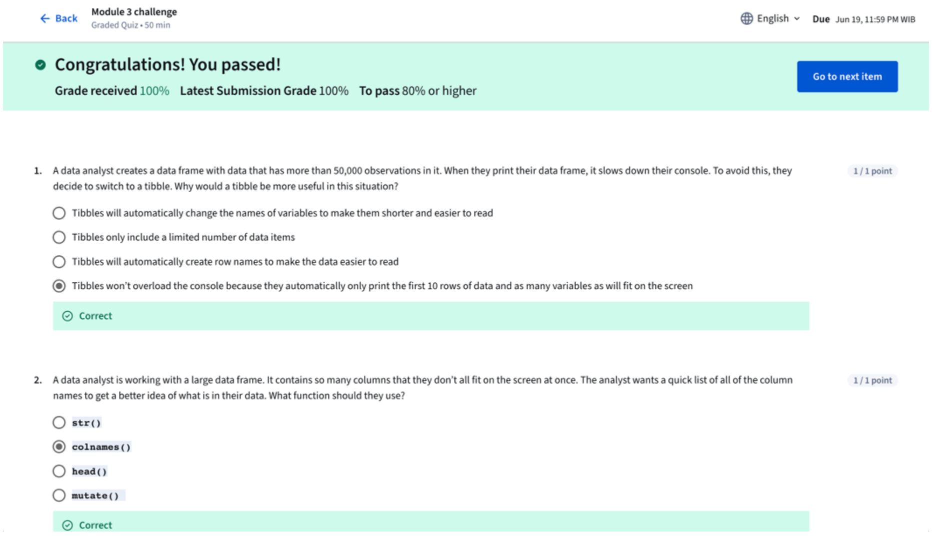
☑ cor()

Correct
The sd(), cor(), and mean() functions can provide a statistical summary of the dataset using standard deviation, correlation, and mean.

2. A data analyst inputs the following command: 1 / 1 point

```
quartet %>% group_by(set) %>% summarize(mean(x), sd(x), mean(y), sd(y), cor(x, y)).
```

Which of the functions in this command can help them determine how strongly related their variables are?



**Module 3 challenge**
Graded Quiz • 50 min

Due Jun 19, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **Latest Submission Grade** 100% **To pass** 80% or higher

1. A data analyst creates a data frame with data that has more than 50,000 observations in it. When they print their data frame, it slows down their console. To avoid this, they decide to switch to a tibble. Why would a tibble be more useful in this situation? 1 / 1 point

○ Tibbles will automatically change the names of variables to make them shorter and easier to read
○ Tibbles only include a limited number of data items
○ Tibbles will automatically create row names to make the data easier to read
● Tibbles won't overload the console because they automatically only print the first 10 rows of data and as many variables as will fit on the screen

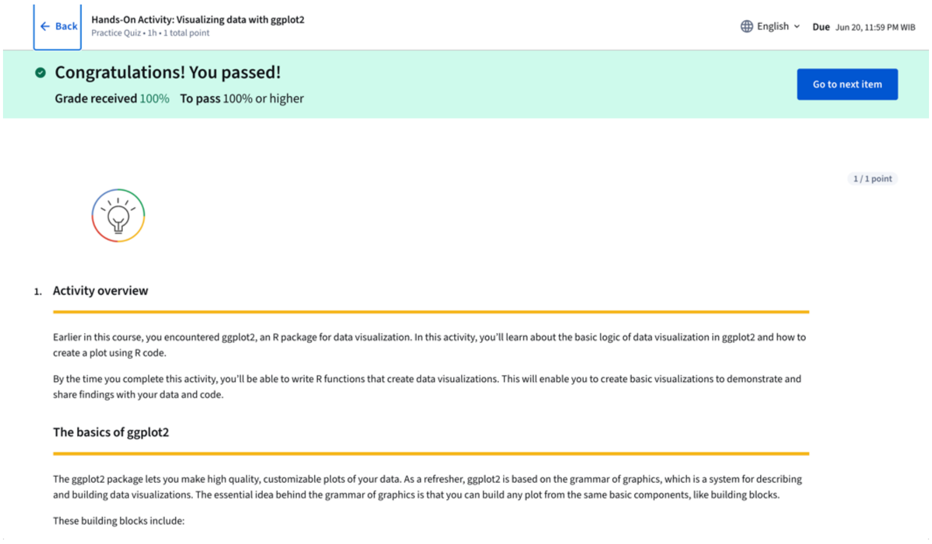
Correct

2. A data analyst is working with a large data frame. It contains so many columns that they don't all fit on the screen at once. The analyst wants a quick list of all of the column names to get a better idea of what is in their data. What function should they use? 1 / 1 point

○ `str()`
● `colnames()`
○ `head()`
○ `mutate()`

Correct

# Modul 4



**Hands-On Activity: Visualizing data with ggplot2**
Practice Quiz • 1h • 1 total point

**Due** Jun 20, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

1. **Activity overview**

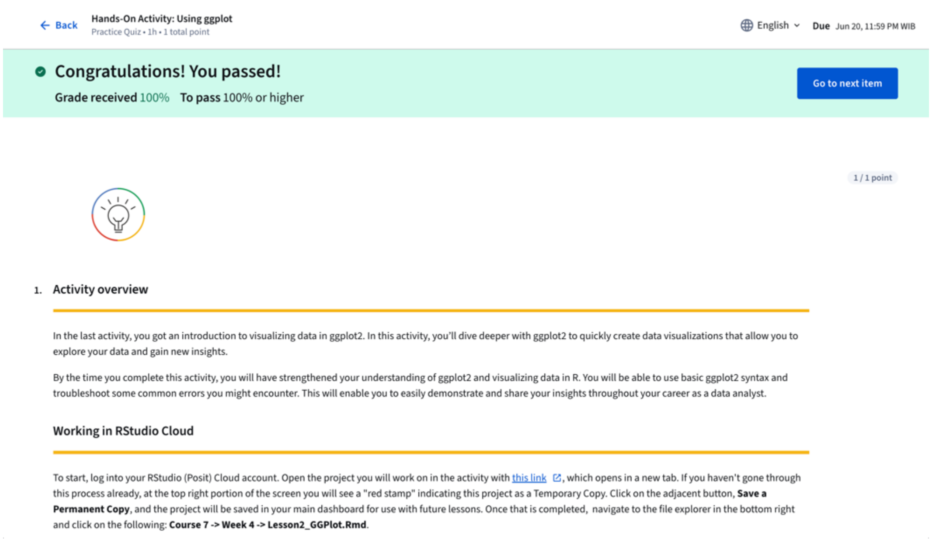
Earlier in this course, you encountered ggplot2, an R package for data visualization. In this activity, you'll learn about the basic logic of data visualization in ggplot2 and how to create a plot using R code.

By the time you complete this activity, you'll be able to write R functions that create data visualizations. This will enable you to create basic visualizations to demonstrate and share findings with your data and code.

**The basics of ggplot2**

The ggplot2 package lets you make high quality, customizable plots of your data. As a refresher, ggplot2 is based on the grammar of graphics, which is a system for describing and building data visualizations. The essential idea behind the grammar of graphics is that you can build any plot from the same basic components, like building blocks.

These building blocks include:



**Hands-On Activity: Using ggplot**
Practice Quiz • 1h • 1 total point

**Due** Jun 20, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

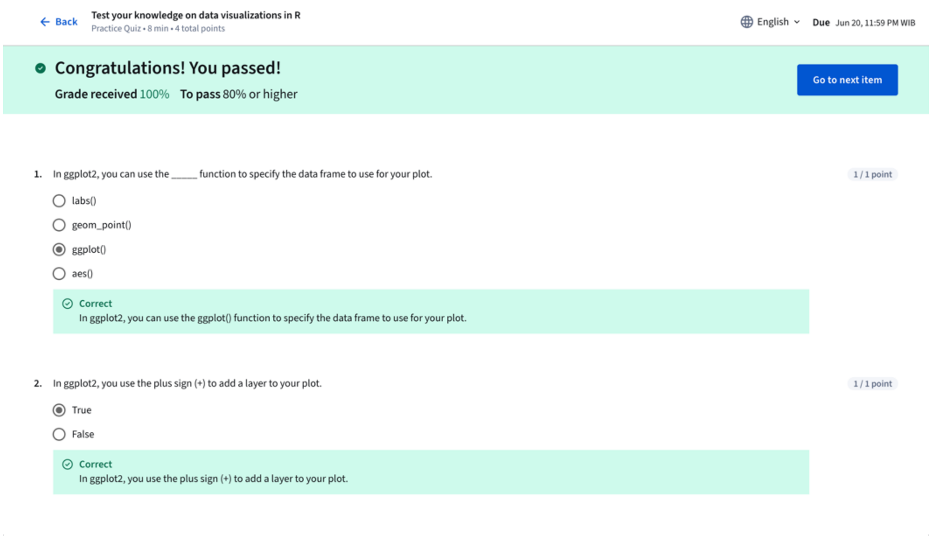
1. **Activity overview**

In the last activity, you got an introduction to visualizing data in ggplot2. In this activity, you'll dive deeper with ggplot2 to quickly create data visualizations that allow you to explore your data and gain new insights.

By the time you complete this activity, you will have strengthened your understanding of ggplot2 and visualizing data in R. You will be able to use basic ggplot2 syntax and troubleshoot some common errors you might encounter. This will enable you to easily demonstrate and share your insights throughout your career as a data analyst.

**Working in RStudio Cloud**

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click on the following: **Course 7 -> Week 4 -> Lesson2_GGPlot.Rmd**.



**Test your knowledge on data visualizations in R**
Practice Quiz • 8 min • 4 total points

**Due** Jun 20, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 80% or higher

1. In ggplot2, you can use the _____ function to specify the data frame to use for your plot.

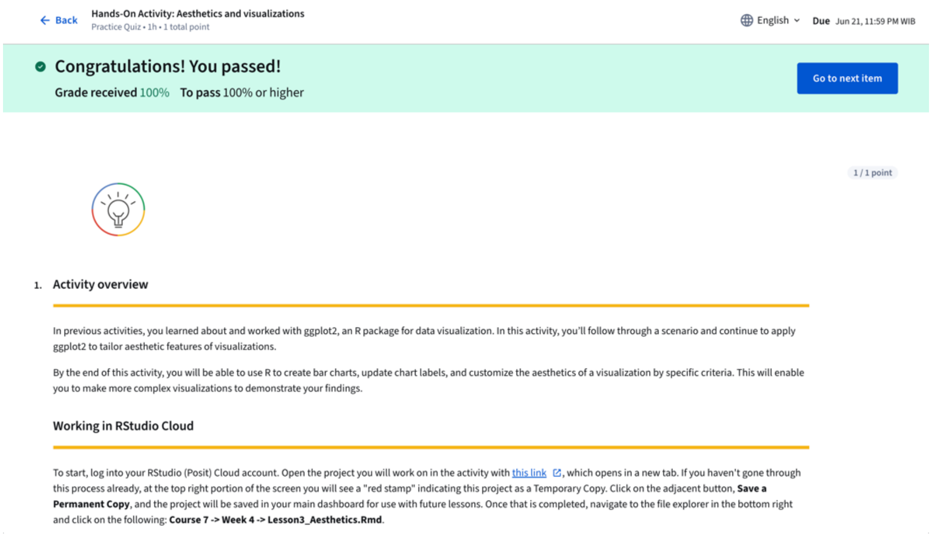
○ labs()
○ geom_point()
● ggplot()
○ aes()

Correct
In ggplot2, you can use the ggplot() function to specify the data frame to use for your plot.

2. In ggplot2, you use the plus sign (+) to add a layer to your plot.

● True
○ False

Correct
In ggplot2, you use the plus sign (+) to add a layer to your plot.



**Hands-On Activity: Aesthetics and visualizations**
Practice Quiz • 1h • 1 total point

Due Jun 21, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

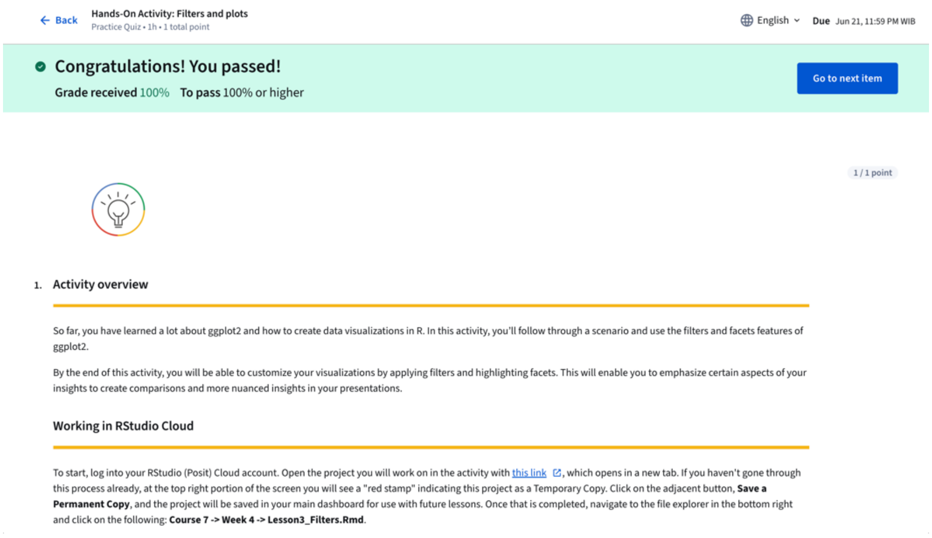
1. **Activity overview**

In previous activities, you learned about and worked with ggplot2, an R package for data visualization. In this activity, you'll follow through a scenario and continue to apply ggplot2 to tailor aesthetic features of visualizations.

By the end of this activity, you will be able to use R to create bar charts, update chart labels, and customize the aesthetics of a visualization by specific criteria. This will enable you to make more complex visualizations to demonstrate your findings.

**Working in RStudio Cloud**

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click on the following: **Course 7 -> Week 4 -> Lesson3_Aesthetics.Rmd**.



**Hands-On Activity: Filters and plots**
Practice Quiz • 1h • 1 total point

Due Jun 21, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

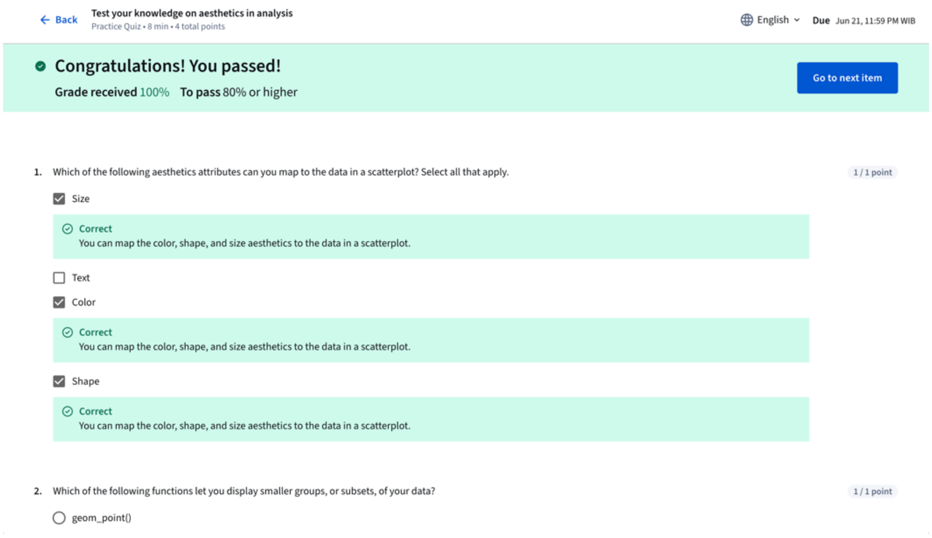
1. **Activity overview**

So far, you have learned a lot about ggplot2 and how to create data visualizations in R. In this activity, you'll follow through a scenario and use the filters and facets features of ggplot2.

By the end of this activity, you will be able to customize your visualizations by applying filters and highlighting facets. This will enable you to emphasize certain aspects of your insights to create comparisons and more nuanced insights in your presentations.

**Working in RStudio Cloud**

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click on the following: **Course 7 -> Week 4 -> Lesson3_Filters.Rmd**.



**Test your knowledge on aesthetics in analysis**
Practice Quiz • 8 min • 4 total points

Due Jun 21, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. Which of the following aesthetics attributes can you map to the data in a scatterplot? Select all that apply. 1 / 1 point

☑ Size

Correct
You can map the color, shape, and size aesthetics to the data in a scatterplot.

☐ Text

☑ Color

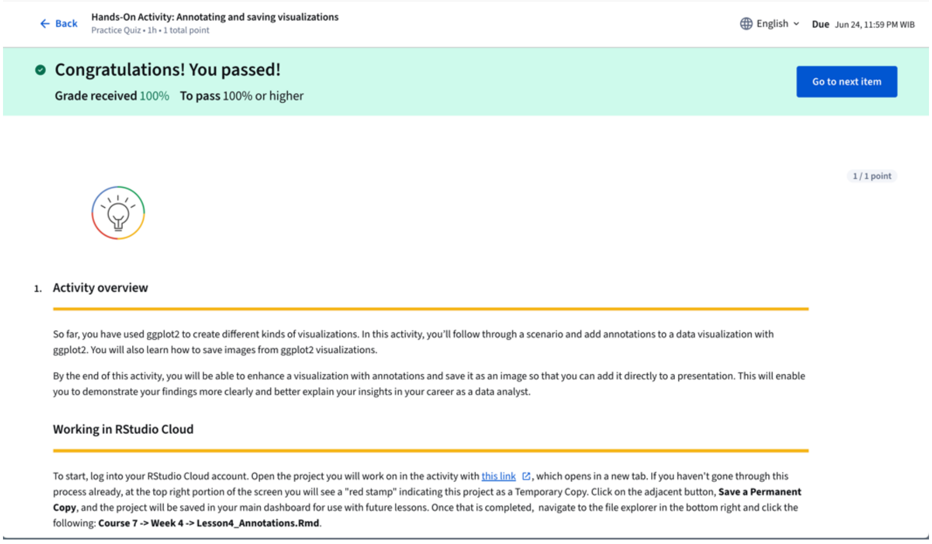
Correct
You can map the color, shape, and size aesthetics to the data in a scatterplot.

☑ Shape

Correct
You can map the color, shape, and size aesthetics to the data in a scatterplot.

2. Which of the following functions let you display smaller groups, or subsets, of your data? 1 / 1 point

○ geom_point()



**Hands-On Activity: Annotating and saving visualizations**

Practice Quiz • 1h • 1 total point

**Due** Jun 24, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

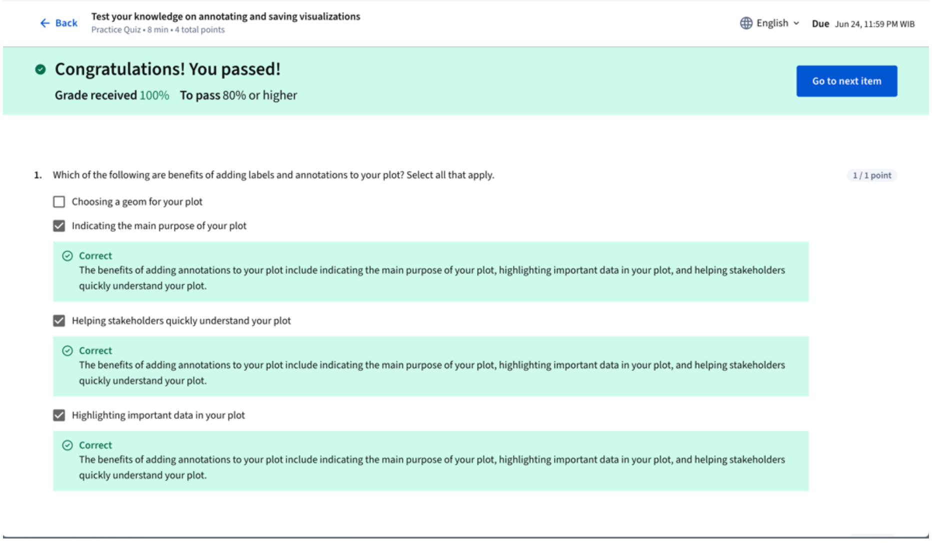
1. **Activity overview**

So far, you have used ggplot2 to create different kinds of visualizations. In this activity, you'll follow through a scenario and add annotations to a data visualization with ggplot2. You will also learn how to save images from ggplot2 visualizations.

By the end of this activity, you will be able to enhance a visualization with annotations and save it as an image so that you can add it directly to a presentation. This will enable you to demonstrate your findings more clearly and better explain your insights in your career as a data analyst.

**Working in RStudio Cloud**

To start, log into your RStudio Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, **Save a Permanent Copy**, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click the following: **Course 7 -> Week 4 -> Lesson4_Annotations.Rmd**.



**Test your knowledge on annotating and saving visualizations**

Practice Quiz • 8 min • 4 total points

**Due** Jun 24, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 100% **To pass** 80% or higher

1. Which of the following are benefits of adding labels and annotations to your plot? Select all that apply.

1 / 1 point

☐ Choosing a geom for your plot

☑ Indicating the main purpose of your plot

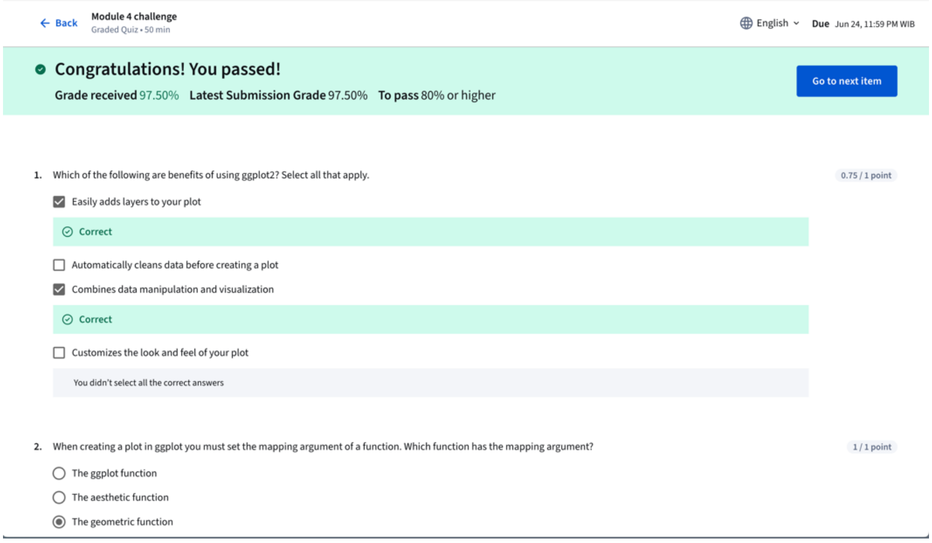
Correct
The benefits of adding annotations to your plot include indicating the main purpose of your plot, highlighting important data in your plot, and helping stakeholders quickly understand your plot.

☑ Helping stakeholders quickly understand your plot

Correct
The benefits of adding annotations to your plot include indicating the main purpose of your plot, highlighting important data in your plot, and helping stakeholders quickly understand your plot.

☑ Highlighting important data in your plot

Correct
The benefits of adding annotations to your plot include indicating the main purpose of your plot, highlighting important data in your plot, and helping stakeholders quickly understand your plot.



**Module 4 challenge**

Graded Quiz • 50 min

**Due** Jun 24, 11:59 PM WIB

## Congratulations! You passed!

**Grade received** 97.50% **Latest Submission Grade** 97.50% **To pass** 80% or higher

1. Which of the following are benefits of using ggplot2? Select all that apply.

0.75 / 1 point

☑ Easily adds layers to your plot

Correct

☐ Automatically cleans data before creating a plot

☑ Combines data manipulation and visualization

Correct

☐ Customizes the look and feel of your plot

You didn't select all the correct answers

2. When creating a plot in ggplot you must set the mapping argument of a function. Which function has the mapping argument?

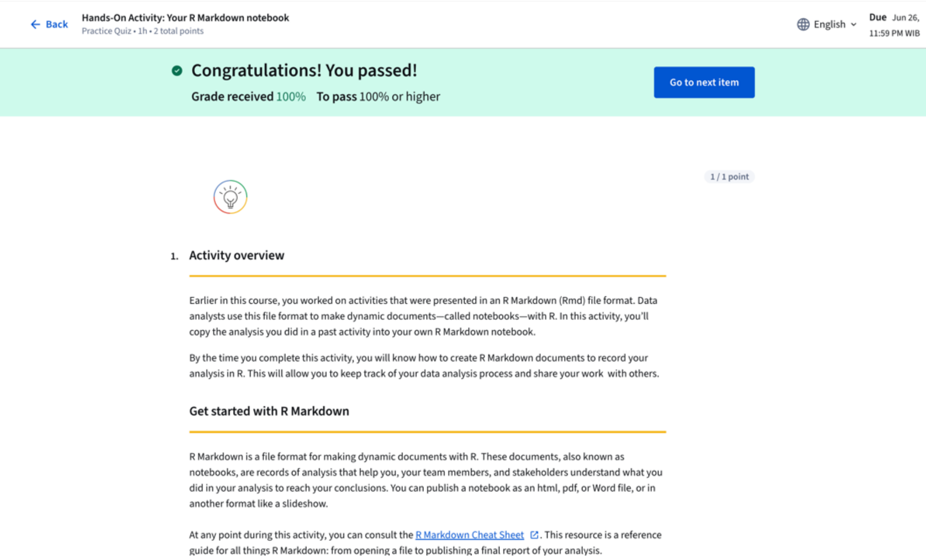
1 / 1 point

○ The ggplot function

○ The aesthetic function

● The geometric function

# Modul 5



**Hands-On Activity: Your R Markdown notebook**

Practice Quiz • 1h • 2 total points

**Due** Jun 26, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

1. **Activity overview**

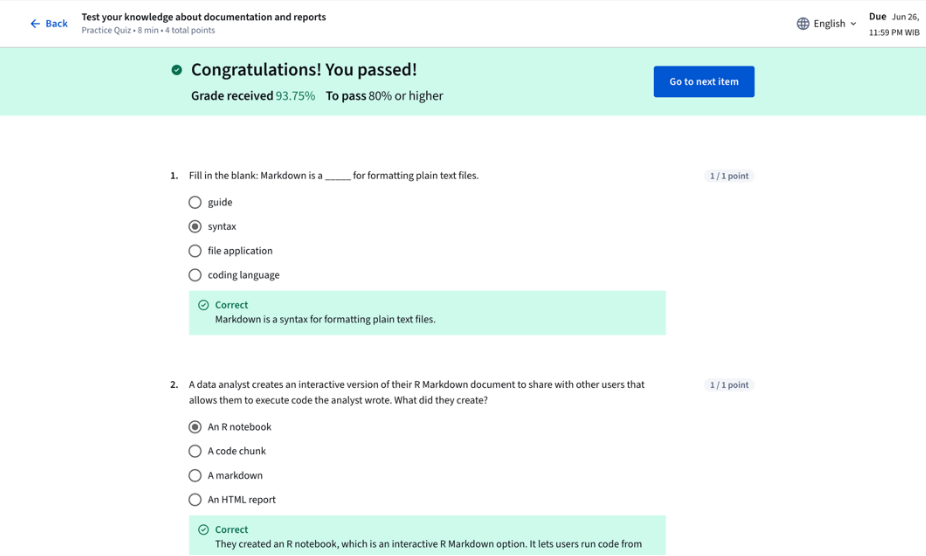
Earlier in this course, you worked on activities that were presented in an R Markdown (Rmd) file format. Data analysts use this file format to make dynamic documents—called notebooks—with R. In this activity, you'll copy the analysis you did in a past activity into your own R Markdown notebook.

By the time you complete this activity, you will know how to create R Markdown documents to record your analysis in R. This will allow you to keep track of your data analysis process and share your work with others.

**Get started with R Markdown**

R Markdown is a file format for making dynamic documents with R. These documents, also known as notebooks, are records of analysis that help you, your team members, and stakeholders understand what you did in your analysis to reach your conclusions. You can publish a notebook as an html, pdf, or Word file, or in another format like a slideshow.

At any point during this activity, you can consult the R Markdown Cheat Sheet. This resource is a reference guide for all things R Markdown: from opening a file to publishing a final report of your analysis.



**Test your knowledge about documentation and reports**

Practice Quiz • 8 min • 4 total points

**Due** Jun 26, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 93.75% **To pass** 80% or higher

1. Fill in the blank: Markdown is a _____ for formatting plain text files. 1 / 1 point

○ guide

● syntax

○ file application

○ coding language

Correct

Markdown is a syntax for formatting plain text files.

2. A data analyst creates an interactive version of their R Markdown document to share with other users that allows them to execute code the analyst wrote. What did they create? 1 / 1 point

● An R notebook

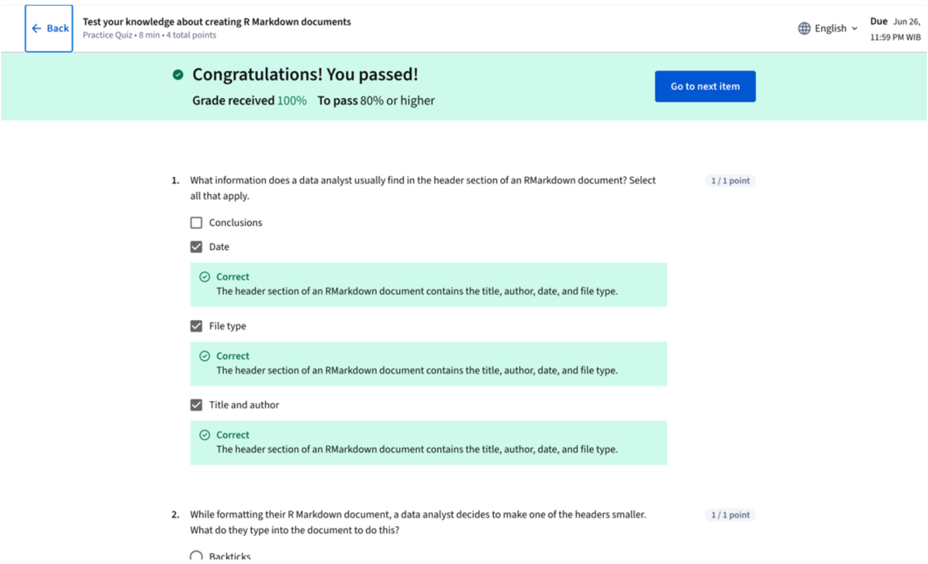
○ A code chunk

○ A markdown

○ An HTML report

Correct

They created an R notebook, which is an interactive R Markdown option. It lets users run code from



**Test your knowledge about creating R Markdown documents**

Practice Quiz • 8 min • 4 total points

**Due** Jun 26, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. What information does a data analyst usually find in the header section of an RMarkdown document? Select all that apply.

1 / 1 point

☐ Conclusions

☑ Date

Correct
The header section of an RMarkdown document contains the title, author, date, and file type.

☑ File type

Correct
The header section of an RMarkdown document contains the title, author, date, and file type.

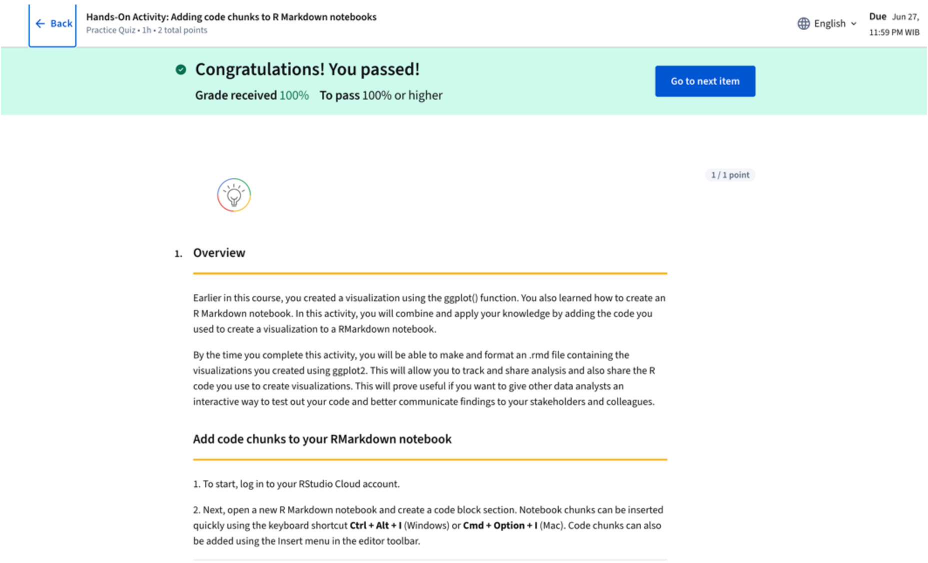
☑ Title and author

Correct
The header section of an RMarkdown document contains the title, author, date, and file type.

2. While formatting their R Markdown document, a data analyst decides to make one of the headers smaller. What do they type into the document to do this?

1 / 1 point

○ Backticks



**Hands-On Activity: Adding code chunks to R Markdown notebooks**

Practice Quiz • 1h • 2 total points

**Due** Jun 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 100% or higher

1 / 1 point

1. **Overview**

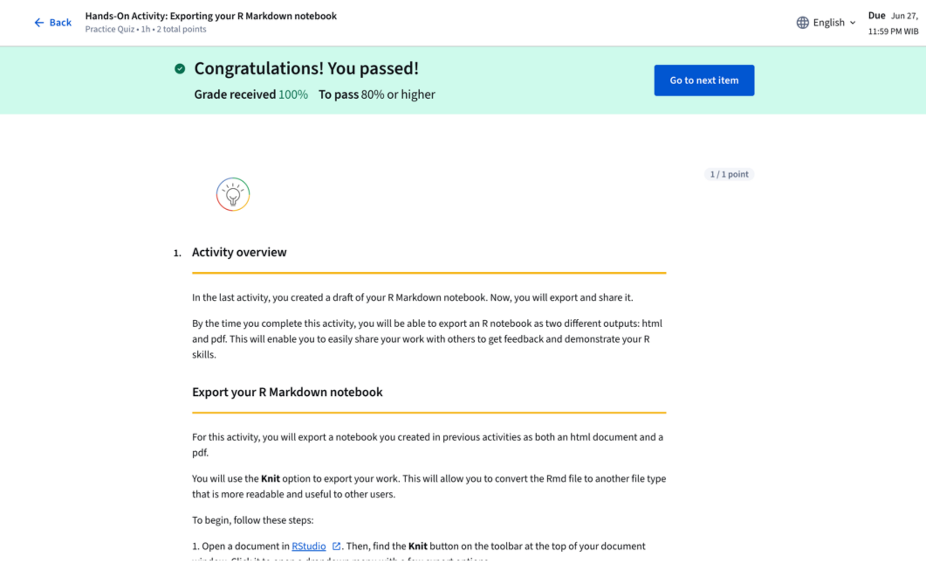
Earlier in this course, you created a visualization using the ggplot() function. You also learned how to create an R Markdown notebook. In this activity, you will combine and apply your knowledge by adding the code you used to create a visualization to a RMarkdown notebook.

By the time you complete this activity, you will be able to make and format an .rmd file containing the visualizations you created using ggplot2. This will allow you to track and share analysis and also share the R code you use to create visualizations. This will prove useful if you want to give other data analysts an interactive way to test out your code and better communicate findings to your stakeholders and colleagues.

**Add code chunks to your RMarkdown notebook**

1. To start, log in to your RStudio Cloud account.

2. Next, open a new R Markdown notebook and create a code block section. Notebook chunks can be inserted quickly using the keyboard shortcut **Ctrl + Alt + I** (Windows) or **Cmd + Option + I** (Mac). Code chunks can also be added using the Insert menu in the editor toolbar.



**Hands-On Activity: Exporting your R Markdown notebook**

Practice Quiz • 1h • 2 total points

**Due** Jun 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1 / 1 point

1. **Activity overview**

In the last activity, you created a draft of your R Markdown notebook. Now, you will export and share it.

By the time you complete this activity, you will be able to export an R notebook as two different outputs: html and pdf. This will enable you to easily share your work with others to get feedback and demonstrate your R skills.

**Export your R Markdown notebook**

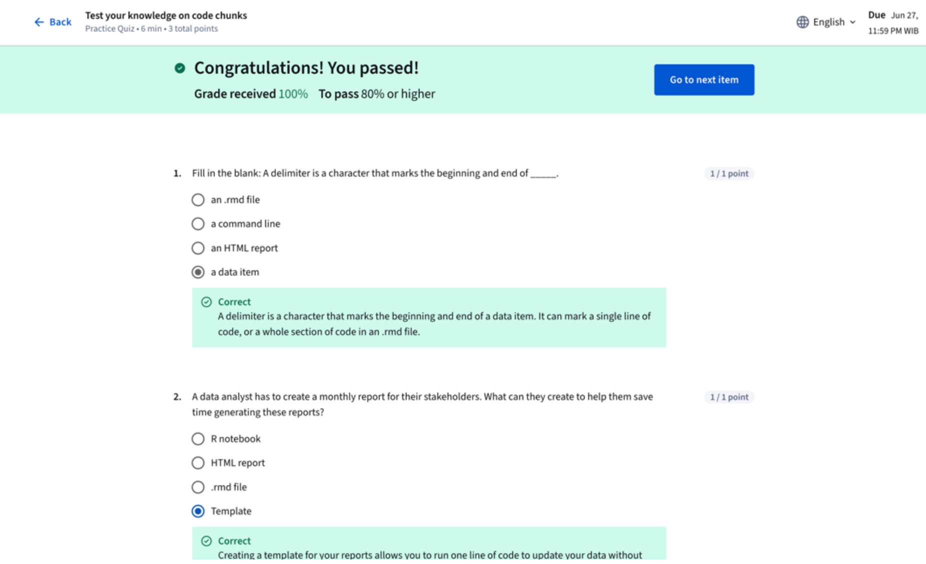
For this activity, you will export a notebook you created in previous activities as both an html document and a pdf.

You will use the **Knit** option to export your work. This will allow you to convert the Rmd file to another file type that is more readable and useful to other users.

To begin, follow these steps:

1. Open a document in RStudio. Then, find the **Knit** button on the toolbar at the top of your document

---



**Test your knowledge on code chunks**

Practice Quiz • 6 min • 3 total points

**Due** Jun 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 100% **To pass** 80% or higher

1. Fill in the blank: A delimiter is a character that marks the beginning and end of _____. 1 / 1 point

○ an .rmd file

○ a command line

○ an HTML report

● a data item

Correct
A delimiter is a character that marks the beginning and end of a data item. It can mark a single line of code, or a whole section of code in an .rmd file.

2. A data analyst has to create a monthly report for their stakeholders. What can they create to help them save time generating these reports? 1 / 1 point

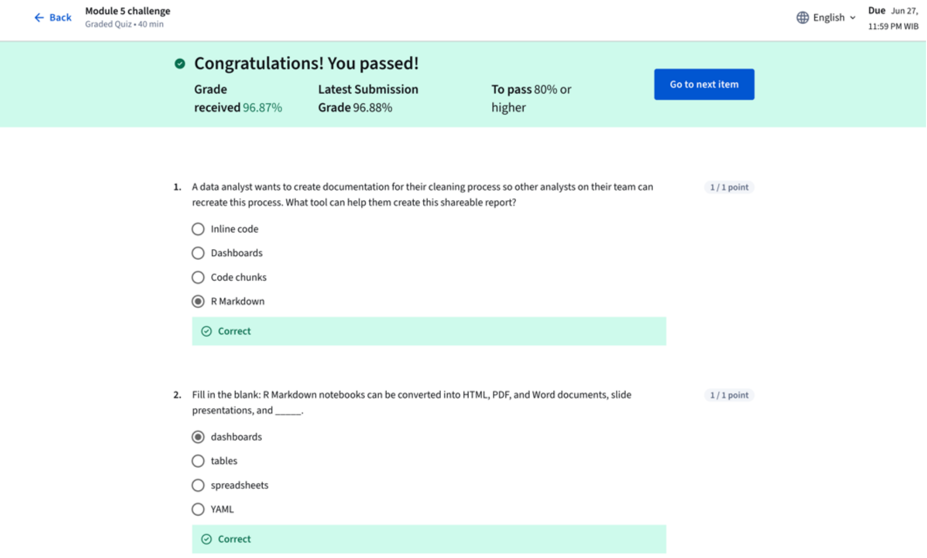
○ R notebook

○ HTML report

○ .rmd file

● Template

Correct
Creating a template for your reports allows you to run one line of code to update your data without



**Module 5 challenge**
Graded Quiz • 40 min

**Due** Jun 27, 11:59 PM WIB

**Congratulations! You passed!**

**Grade received** 96.87%

**Latest Submission Grade** 96.88%

**To pass** 80% or higher

1. A data analyst wants to create documentation for their cleaning process so other analysts on their team can recreate this process. What tool can help them create this shareable report?

1 / 1 point

○ Inline code
○ Dashboards
○ Code chunks
● R Markdown

✓ Correct

2. Fill in the blank: R Markdown notebooks can be converted into HTML, PDF, and Word documents, slide presentations, and _____.

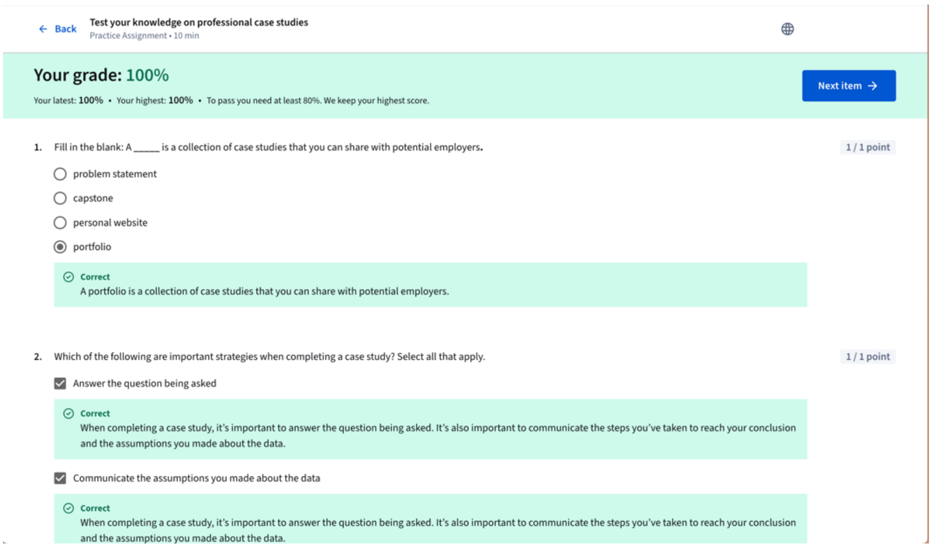
1 / 1 point

● dashboards
○ tables
○ spreadsheets
○ YAML

✓ Correct

# Course 8: Google Data Analytics Capstone: Complete a Case Study

## Modul 1



**Test your knowledge on professional case studies**

Practice Assignment • 10 min

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

1. Fill in the blank: A _____ is a collection of case studies that you can share with potential employers.

1 / 1 point

○ problem statement

○ capstone

○ personal website

● portfolio

**Correct**

A portfolio is a collection of case studies that you can share with potential employers.

2. Which of the following are important strategies when completing a case study? Select all that apply.

1 / 1 point

☑ Answer the question being asked

**Correct**

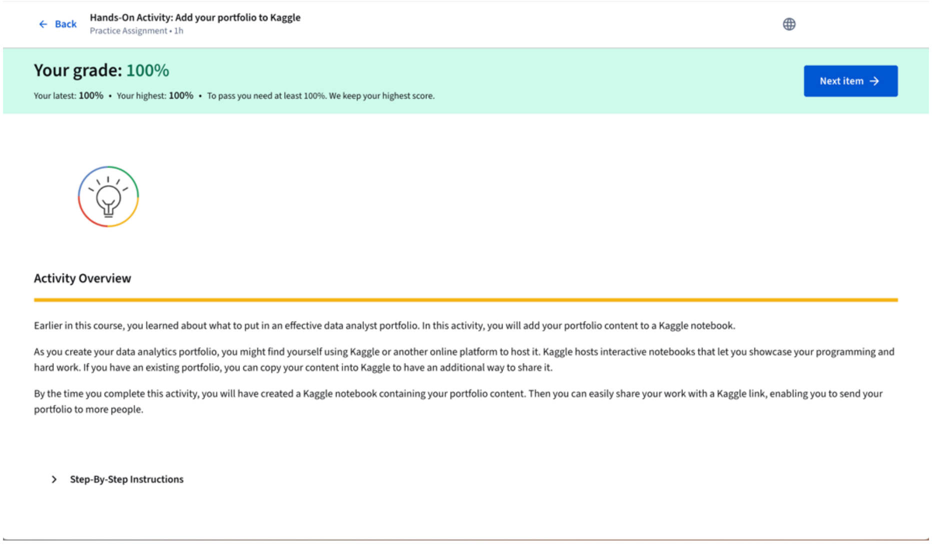
When completing a case study, it's important to answer the question being asked. It's also important to communicate the steps you've taken to reach your conclusion and the assumptions you made about the data.

☑ Communicate the assumptions you made about the data

**Correct**

When completing a case study, it's important to answer the question being asked. It's also important to communicate the steps you've taken to reach your conclusion and the assumptions you made about the data.

## Modul 2



**Hands-On Activity: Add your portfolio to Kaggle**

Practice Assignment • 1h

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 100%. We keep your highest score.

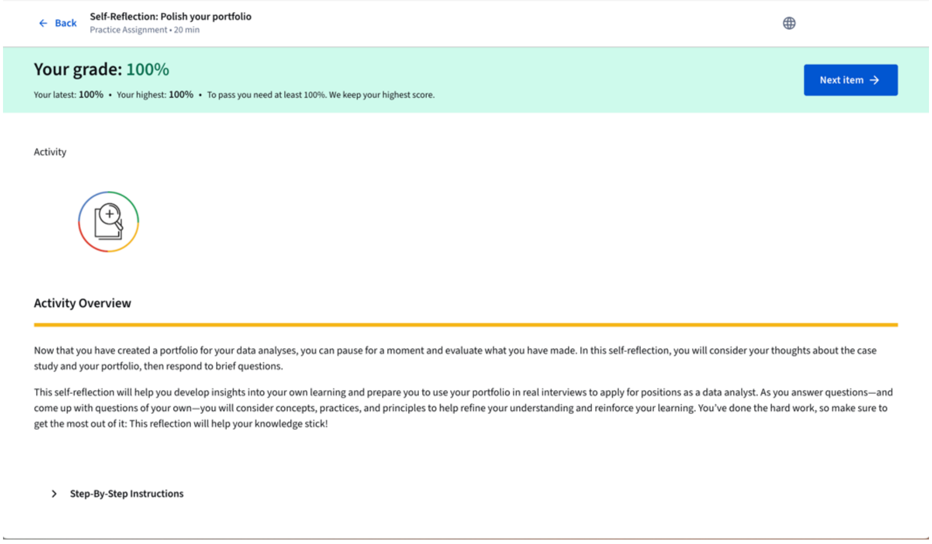
**Activity Overview**

Earlier in this course, you learned about what to put in an effective data analyst portfolio. In this activity, you will add your portfolio content to a Kaggle notebook.

As you create your data analytics portfolio, you might find yourself using Kaggle or another online platform to host it. Kaggle hosts interactive notebooks that let you showcase your programming and hard work. If you have an existing portfolio, you can copy your content into Kaggle to have an additional way to share it.

By the time you complete this activity, you will have created a Kaggle notebook containing your portfolio content. Then you can easily share your work with a Kaggle link, enabling you to send your portfolio to more people.

## Modul 3



**Self-Reflection: Polish your portfolio**
Practice Assignment • 20 min

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 100%. We keep your highest score.

Activity

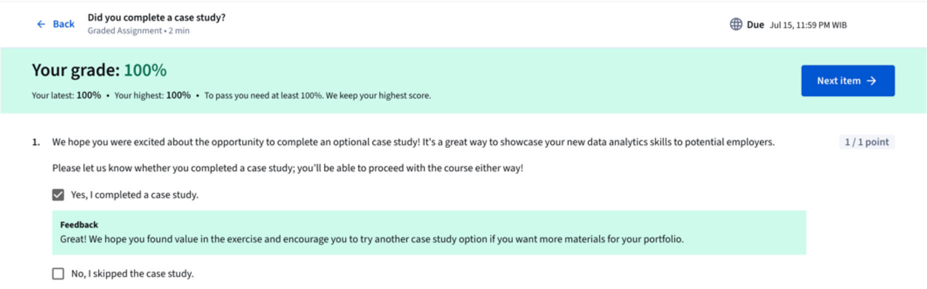
**Activity Overview**

Now that you have created a portfolio for your data analyses, you can pause for a moment and evaluate what you have made. In this self-reflection, you will consider your thoughts about the case study and your portfolio, then respond to brief questions.

This self-reflection will help you develop insights into your own learning and prepare you to use your portfolio in real interviews to apply for positions as a data analyst. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

**Step-By-Step Instructions**

## Modul 4



**Did you complete a case study?**
Graded Assignment • 2 min

**Due** Jul 15, 11:59 PM WIB

**Your grade: 100%**

Your latest: **100%** • Your highest: **100%** • To pass you need at least 100%. We keep your highest score.

1. We hope you were excited about the opportunity to complete an optional case study! It's a great way to showcase your new data analytics skills to potential employers. 1 / 1 point

Please let us know whether you completed a case study; you'll be able to proceed with the course either way!

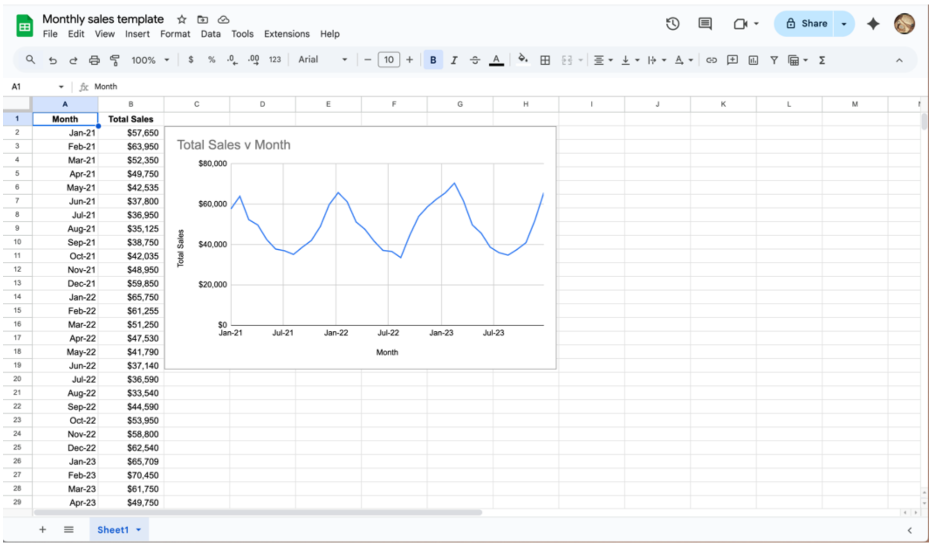
☑ Yes, I completed a case study.

**Feedback**
Great! We hope you found value in the exercise and encourage you to try another case study option if you want more materials for your portfolio.

☐ No, I skipped the case study.

# Hands-On Course 1 Hingga Course 7

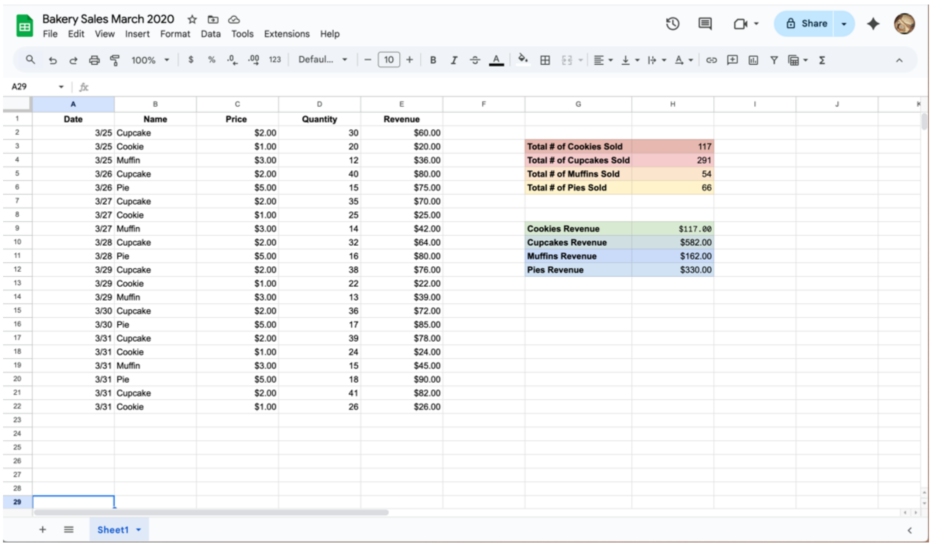
## Course 1



| Month | Total Sales |
|---|---|
| Jan-21 | $57,650 |
| Feb-21 | $63,950 |
| Mar-21 | $52,350 |
| Apr-21 | $49,750 |
| May-21 | $42,535 |
| Jun-21 | $37,800 |
| Jul-21 | $36,950 |
| Aug-21 | $35,125 |
| Sep-21 | $38,750 |
| Oct-21 | $42,035 |
| Nov-21 | $48,950 |
| Dec-21 | $59,850 |
| Jan-22 | $65,750 |
| Feb-22 | $61,255 |
| Mar-22 | $51,250 |
| Apr-22 | $47,530 |
| May-22 | $41,790 |
| Jun-22 | $37,140 |
| Jul-22 | $36,590 |
| Aug-22 | $33,540 |
| Sep-22 | $44,590 |
| Oct-22 | $53,950 |
| Nov-22 | $58,800 |
| Dec-22 | $62,540 |
| Jan-23 | $65,709 |
| Feb-23 | $70,450 |
| Mar-23 | $61,750 |
| Apr-23 | $49,750 |

Membuat chart dari *table data*

## Course 2

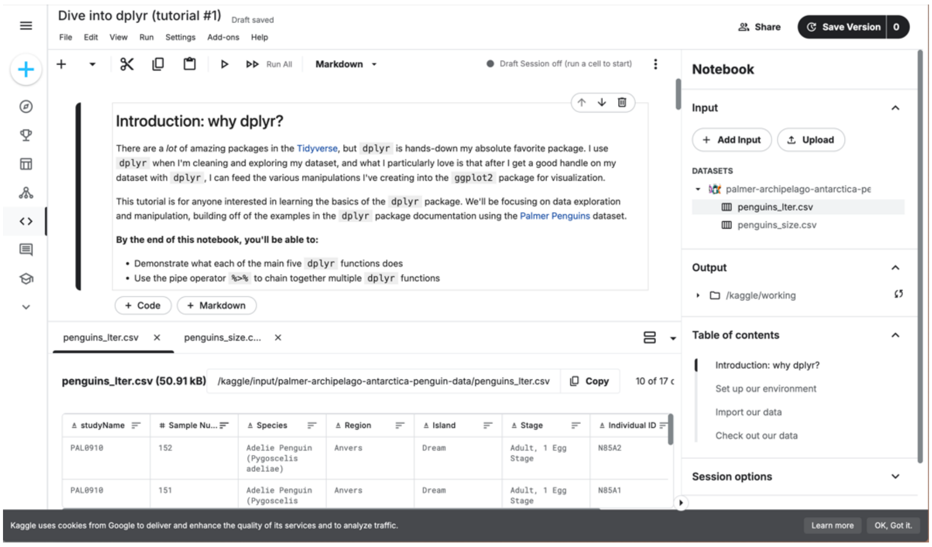


| Date | Name | Price | Quantity | Revenue |
|---|---|---|---|---|
| 3/25 | Cupcake | $2.00 | 30 | $60.00 |
| 3/25 | Cookie | $1.00 | 20 | $20.00 |
| 3/25 | Muffin | $3.00 | 12 | $36.00 |
| 3/26 | Cupcake | $2.00 | 40 | $80.00 |
| 3/26 | Pie | $5.00 | 15 | $75.00 |
| 3/27 | Cupcake | $2.00 | 35 | $70.00 |
| 3/27 | Cookie | $1.00 | 25 | $25.00 |
| 3/27 | Muffin | $3.00 | 14 | $42.00 |
| 3/28 | Cupcake | $2.00 | 32 | $64.00 |
| 3/28 | Pie | $5.00 | 16 | $80.00 |
| 3/29 | Cupcake | $2.00 | 38 | $76.00 |
| 3/29 | Cookie | $1.00 | 22 | $22.00 |
| 3/29 | Muffin | $3.00 | 13 | $39.00 |
| 3/30 | Cupcake | $2.00 | 36 | $72.00 |
| 3/30 | Pie | $5.00 | 17 | $85.00 |
| 3/31 | Cupcake | $2.00 | 39 | $78.00 |
| 3/31 | Cookie | $1.00 | 24 | $24.00 |
| 3/31 | Muffin | $3.00 | 15 | $45.00 |
| 3/31 | Pie | $5.00 | 18 | $90.00 |
| 3/31 | Cupcake | $2.00 | 41 | $82.00 |
| 3/31 | Cookie | $1.00 | 26 | $26.00 |

| | |
|---|---|
| **Total # of Cookies Sold** | 117 |
| **Total # of Cupcakes Sold** | 291 |
| **Total # of Muffins Sold** | 54 |
| **Total # of Pies Sold** | 66 |
| | |
| | |
| **Cookies Revenue** | $117.00 |
| **Cupcakes Revenue** | $582.00 |
| **Muffins Revenue** | $162.00 |
| **Pies Revenue** | $330.00 |

Menggunakan formula untuk mendapatkan *insight*

## Course 3



### Introduction: why dplyr?

There are a *lot* of amazing packages in the Tidyverse, but `dplyr` is hands-down my absolute favorite package. I use `dplyr` when I'm cleaning and exploring my dataset, and what I particularly love is that after I get a good handle on my dataset with `dplyr`, I can feed the various manipulations I've creating into the `ggplot2` package for visualization.

This tutorial is for anyone interested in learning the basics of the `dplyr` package. We'll be focusing on data exploration and manipulation, building off of the examples in the `dplyr` package documentation using the Palmer Penguins dataset.

**By the end of this notebook, you'll be able to:**

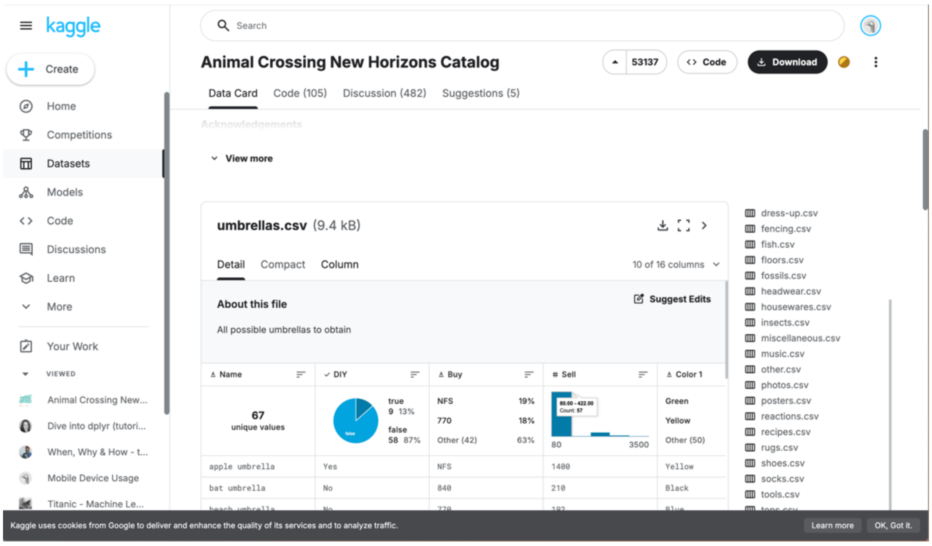
- Demonstrate what each of the main five `dplyr` functions does
- Use the pipe operator `%>%` to chain together multiple `dplyr` functions

**penguins_lter.csv (50.91 kB)** /kaggle/input/palmer-archipelago-antarctica-penguin-data/penguins_lter.csv

| studyName | Sample Nu... | Species | Region | Island | Stage | Individual ID |
|---|---|---|---|---|---|---|
| PAL0910 | 152 | Adelie Penguin (Pygoscelis adeliae) | Anvers | Dream | Adult, 1 Egg Stage | N85A2 |
| PAL0910 | 151 | Adelie Penguin (Pygoscelis | Anvers | Dream | Adult, 1 Egg Stage | N85A1 |

Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic.

Pengenalan *Kaggle*



### Animal Crossing New Horizons Catalog

**umbrellas.csv** (9.4 kB)

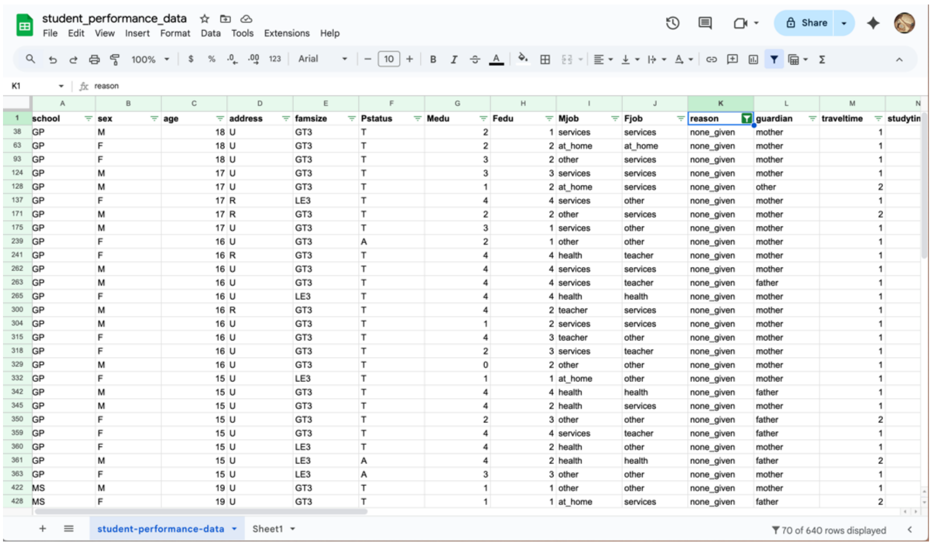
**About this file**

All possible umbrellas to obtain

| Name | DIY | Buy | Sell | Color 1 |
|---|---|---|---|---|
| 67 unique values | | NFS 19%<br>770 18%<br>Other (42) 63% | | Green<br>Yellow<br>Other (50) |
| apple umbrella | Yes | NFS | 1400 | Yellow |
| bat umbrella | No | 840 | 210 | Black |

Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic.

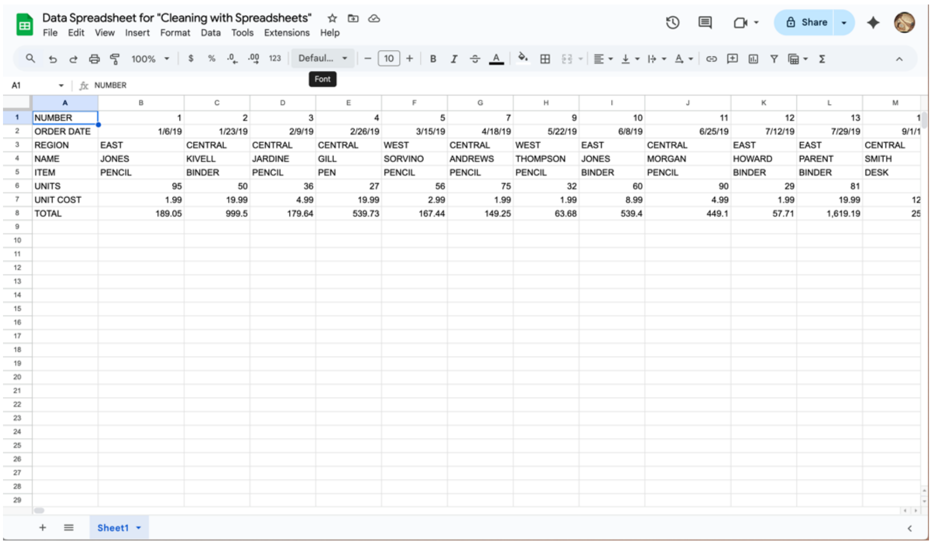
Eksplor *Dataset*



| 1 | school | sex | age | address | famsize | Pstatus | Medu | Fedu | Mjob | Fjob | reason | guardian | traveltime | studytin |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 38 | GP | M | 18 | U | GT3 | T | 2 | 1 | services | services | none_given | mother | 1 | |
| 63 | GP | F | 18 | U | GT3 | T | 2 | 2 | at_home | at_home | none_given | mother | 1 | |
| 93 | GP | F | 18 | U | GT3 | T | 3 | 2 | other | services | none_given | mother | 1 | |
| 124 | GP | M | 17 | U | GT3 | T | 3 | 3 | services | services | none_given | mother | 1 | |
| 128 | GP | M | 17 | U | GT3 | T | 1 | 2 | at_home | services | none_given | other | 2 | |
| 137 | GP | F | 17 | R | LE3 | T | 4 | 4 | services | other | none_given | mother | 1 | |
| 171 | GP | M | 17 | R | GT3 | T | 2 | 2 | other | services | none_given | mother | 2 | |
| 175 | GP | M | 17 | U | GT3 | T | 3 | 1 | services | other | none_given | mother | 1 | |
| 239 | GP | F | 16 | U | GT3 | A | 2 | 1 | other | other | none_given | mother | 1 | |
| 241 | GP | F | 16 | R | GT3 | T | 4 | 4 | health | teacher | none_given | mother | 1 | |
| 262 | GP | M | 16 | U | GT3 | T | 4 | 4 | services | services | none_given | mother | 1 | |
| 263 | GP | M | 16 | U | GT3 | T | 4 | 4 | services | teacher | none_given | father | 1 | |
| 265 | GP | F | 16 | U | LE3 | T | 4 | 4 | health | health | none_given | mother | 1 | |
| 300 | GP | M | 16 | R | GT3 | T | 4 | 2 | teacher | services | none_given | mother | 1 | |
| 304 | GP | M | 16 | U | GT3 | T | 1 | 2 | services | services | none_given | mother | 1 | |
| 315 | GP | F | 16 | U | GT3 | T | 4 | 3 | teacher | other | none_given | mother | 1 | |
| 318 | GP | F | 16 | U | GT3 | T | 2 | 3 | services | teacher | none_given | mother | 1 | |
| 329 | GP | M | 16 | U | GT3 | T | 0 | 2 | other | other | none_given | mother | 1 | |
| 332 | GP | F | 15 | U | LE3 | T | 1 | 1 | at_home | other | none_given | mother | 1 | |
| 342 | GP | M | 15 | U | GT3 | T | 4 | 4 | health | health | none_given | father | 1 | |
| 345 | GP | M | 15 | U | GT3 | T | 4 | 2 | health | services | none_given | mother | 1 | |
| 350 | GP | F | 15 | U | GT3 | T | 2 | 3 | other | other | none_given | father | 2 | |
| 359 | GP | F | 15 | U | GT3 | T | 4 | 4 | services | teacher | none_given | father | 1 | |
| 360 | GP | F | 15 | U | LE3 | T | 4 | 2 | health | other | none_given | mother | 1 | |
| 361 | GP | M | 15 | U | LE3 | A | 4 | 2 | health | health | none_given | father | 2 | |
| 363 | GP | F | 15 | U | LE3 | A | 3 | 3 | other | other | none_given | mother | 1 | |
| 422 | MS | M | 19 | U | GT3 | T | 1 | 1 | other | other | none_given | mother | 1 | |
| 428 | MS | F | 19 | U | GT3 | T | 1 | 1 | at_home | services | none_given | father | 2 | |

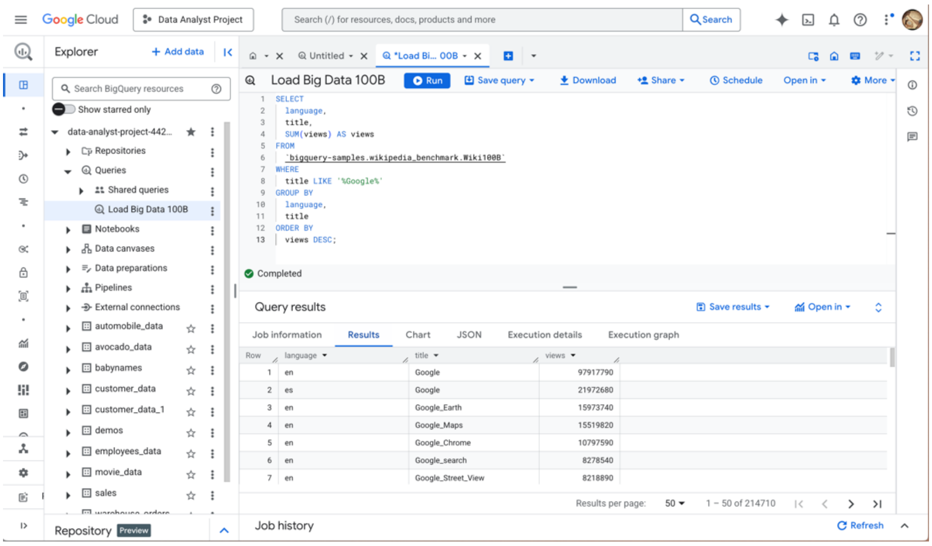
*Cleaning data* dengan *filtering* dan *sorting*

## Course 4



| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NUMBER | 1 | 2 | 3 | 4 | 5 | 7 | 9 | 10 | 11 | 12 | 13 | 1 |
| 2 | ORDER DATE | 1/6/19 | 1/23/19 | 2/9/19 | 2/26/19 | 3/15/19 | 4/18/19 | 5/22/19 | 6/8/19 | 6/25/19 | 7/12/19 | 7/29/19 | 9/1/1 |
| 3 | REGION | EAST | CENTRAL | CENTRAL | CENTRAL | WEST | CENTRAL | WEST | EAST | CENTRAL | EAST | EAST | CENTRAL |
| 4 | NAME | JONES | KIVELL | JARDINE | GILL | SORVINO | ANDREWS | THOMPSON | JONES | MORGAN | HOWARD | PARENT | SMITH |
| 5 | ITEM | PENCIL | BINDER | PENCIL | PEN | PENCIL | PENCIL | PENCIL | BINDER | PENCIL | BINDER | BINDER | DESK |
| 6 | UNITS | 95 | 50 | 36 | 27 | 56 | 75 | 32 | 60 | 90 | 29 | 81 | |
| 7 | UNIT COST | 1.99 | 19.99 | 4.99 | 19.99 | 2.99 | 1.99 | 1.99 | 8.99 | 4.99 | 1.99 | 19.99 | 12 |
| 8 | TOTAL | 189.05 | 999.5 | 179.64 | 539.73 | 167.44 | 149.25 | 63.68 | 539.4 | 449.1 | 57.71 | 1,619.19 | 25 |

Hapus data kosong, *transpose* data, *trim* data string



```sql
SELECT
  language,
  title,
  SUM(views) AS views
FROM
  `bigquery-samples.wikipedia_benchmark.Wiki100B`
WHERE
  title LIKE '%Google%'
GROUP BY
  language,
  title
ORDER BY
  views DESC;
```

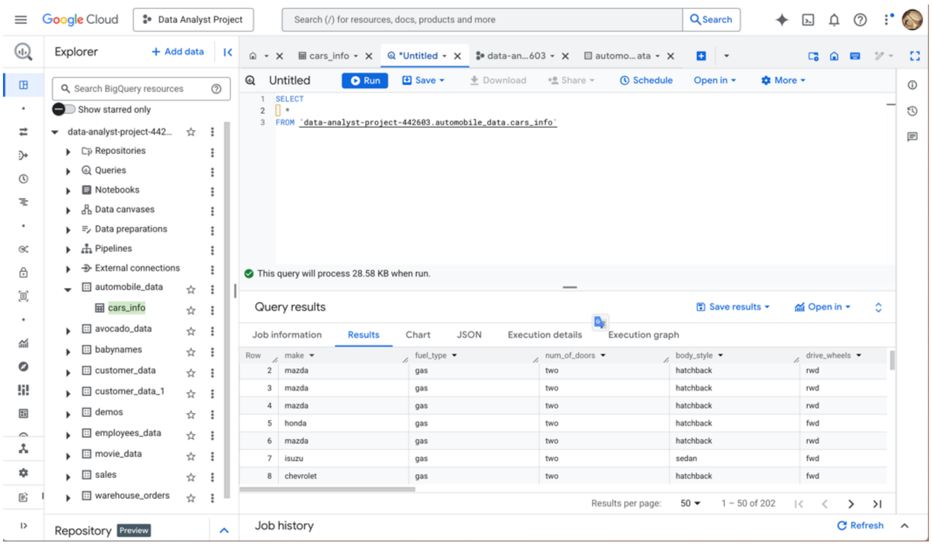
Completed

Query results

| Row | language | title | views |
|---|---|---|---|
| 1 | en | Google | 97917790 |
| 2 | es | Google | 21972680 |
| 3 | en | Google_Earth | 15973740 |
| 4 | en | Google_Maps | 15519820 |
| 5 | en | Google_Chrome | 10797590 |
| 6 | en | Google_search | 8278540 |
| 7 | en | Google_Street_View | 8218890 |

Results per page: 50 1 – 50 of 214710

Menjalankan *query* dari data sebanyak 100 miliar dengan SQL



```sql
SELECT
  *
FROM `data-analyst-project-442603.automobile_data.cars_info`
```

This query will process 28.58 KB when run.

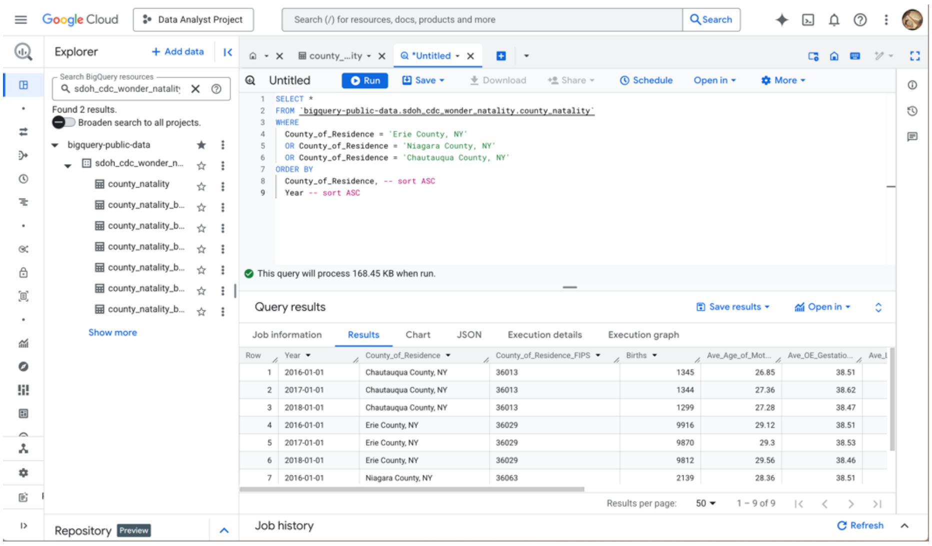
Query results

| Row | make | fuel_type | num_of_doors | body_style | drive_wheels |
|---|---|---|---|---|---|
| 2 | mazda | gas | two | hatchback | rwd |
| 3 | mazda | gas | two | hatchback | rwd |
| 4 | mazda | gas | two | hatchback | rwd |
| 5 | honda | gas | two | hatchback | fwd |
| 6 | mazda | gas | two | hatchback | rwd |
| 7 | isuzu | gas | two | sedan | fwd |
| 8 | chevrolet | gas | two | hatchback | fwd |

Results per page: 50 1 – 50 of 202

*Cleaning* data dengan SQL menggunakan *SELECT, DISTINCT, UPDATE, DELETE*

## Course 5



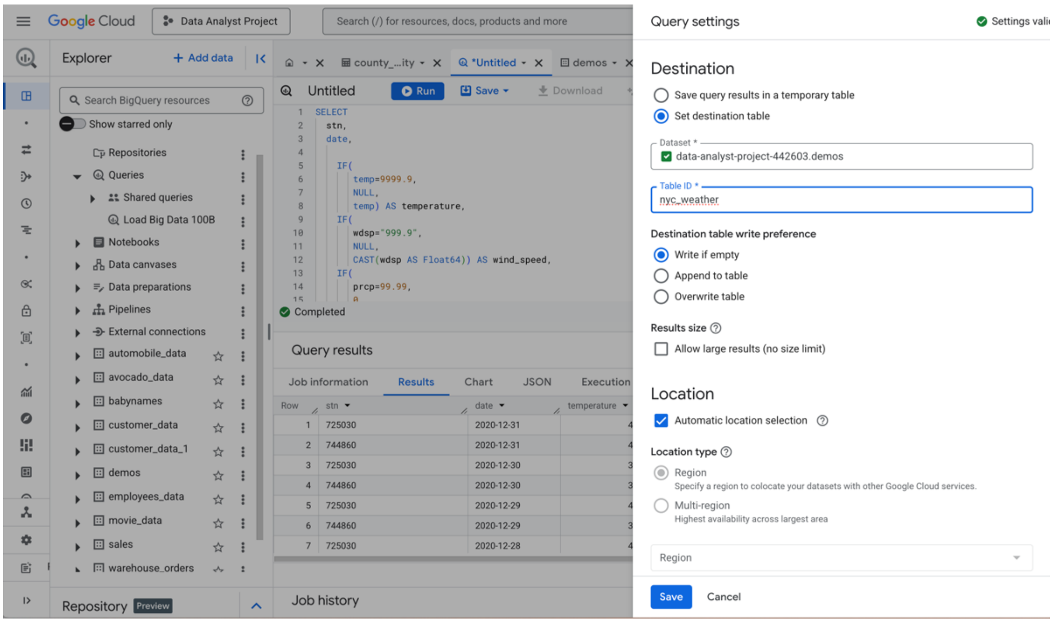
```
SELECT *
FROM `bigquery-public-data.sdoh_cdc_wonder_natality.county_natality`
WHERE
  County_of_Residence = 'Erie County, NY'
  OR County_of_Residence = 'Niagara County, NY'
  OR County_of_Residence = 'Chautauqua County, NY'
ORDER BY
  County_of_Residence, -- sort ASC
  Year -- sort ASC
```

This query will process 168.45 KB when run.

| Row | Year | County_of_Residence | County_of_Residence_FIPS | Births | Ave_Age_of_Mot... | Ave_OE_Gestatio... |
|---|---|---|---|---|---|---|
| 1 | 2016-01-01 | Chautauqua County, NY | 36013 | 1345 | 26.85 | 38.51 |
| 2 | 2017-01-01 | Chautauqua County, NY | 36013 | 1344 | 27.36 | 38.62 |
| 3 | 2018-01-01 | Chautauqua County, NY | 36013 | 1299 | 27.28 | 38.47 |
| 4 | 2016-01-01 | Erie County, NY | 36029 | 9916 | 29.12 | 38.51 |
| 5 | 2017-01-01 | Erie County, NY | 36029 | 9870 | 29.3 | 38.53 |
| 6 | 2018-01-01 | Erie County, NY | 36029 | 9812 | 29.56 | 38.46 |
| 7 | 2016-01-01 | Niagara County, NY | 36063 | 2139 | 28.36 | 38.51 |

Results per page: 50 1 – 9 of 9

*Filtering* dan *sorting query SQL*



```
SELECT
  stn,
  date,

    IF(
      temp=9999.9,
      NULL,
      temp) AS temperature,
    IF(
      wdsp="999.9",
      NULL,
      CAST(wdsp AS Float64)) AS wind_speed,
    IF(
      prcp=99.99,
      0
```

Completed

| Row | stn | date |
|---|---|---|
| 1 | 725030 | 2020-12-31 |
| 2 | 744860 | 2020-12-31 |
| 3 | 725030 | 2020-12-30 |
| 4 | 744860 | 2020-12-30 |
| 5 | 725030 | 2020-12-29 |
| 6 | 744860 | 2020-12-29 |
| 7 | 725030 | 2020-12-28 |

Query settings

Settings vali

Destination

○ Save query results in a temporary table
● Set destination table

Dataset *: data-analyst-project-442603.demos

Table ID *: nyc_weather

Destination table write preference

● Write if empty
○ Append to table
○ Overwrite table

Results size

☐ Allow large results (no size limit)

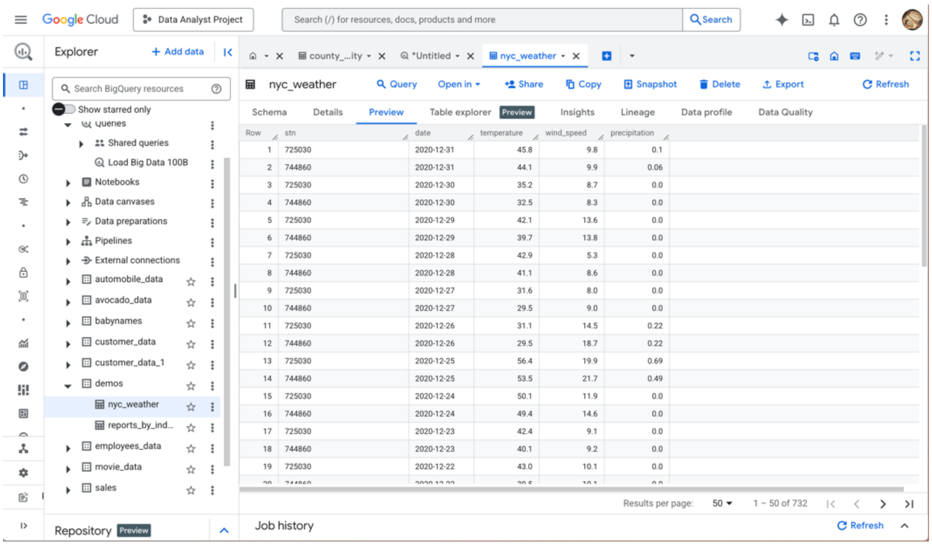
Location

☑ Automatic location selection

Location type

● Region
Specify a region to colocate your datasets with other Google Cloud services.
○ Multi-region
Highest availability across largest area

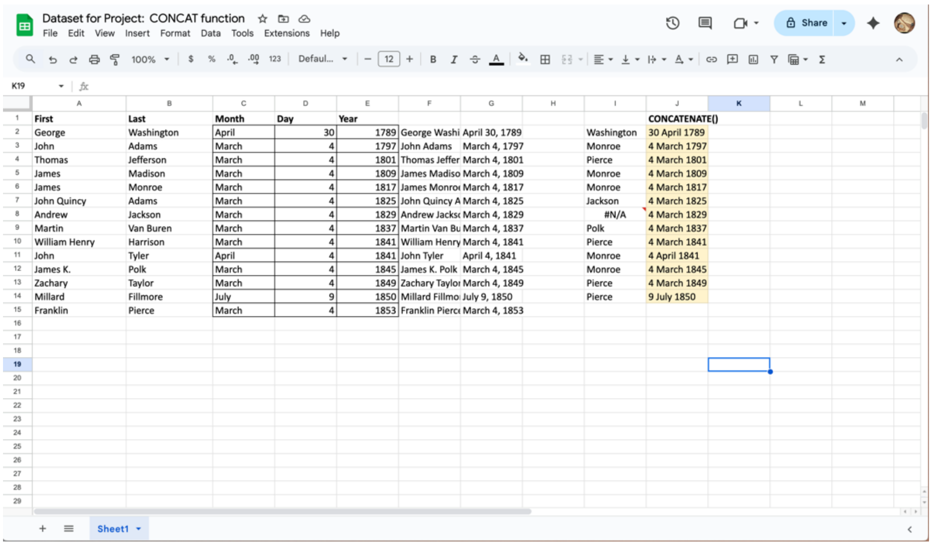
Region



nyc_weather

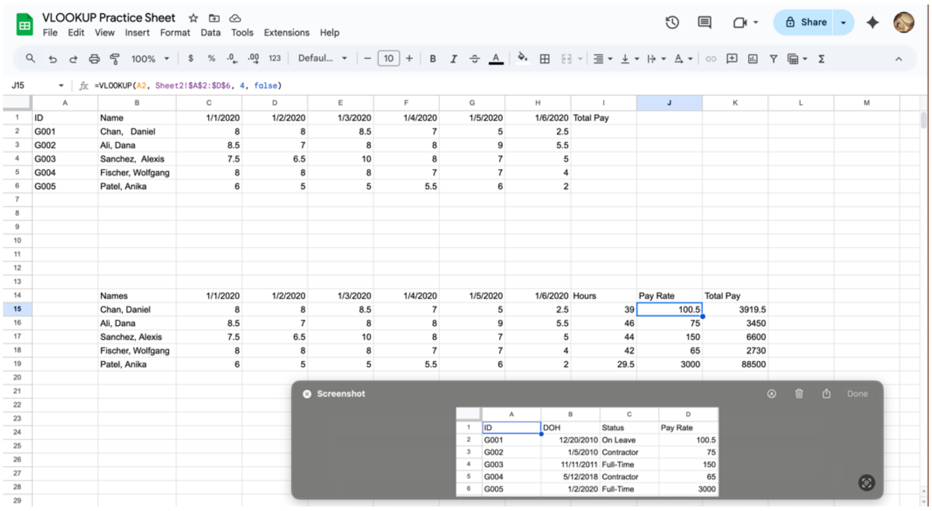
| Row | stn | date | temperature | wind_speed | precipitation |
|---|---|---|---|---|---|
| 1 | 725030 | 2020-12-31 | 45.8 | 9.8 | 0.1 |
| 2 | 744860 | 2020-12-31 | 44.1 | 9.9 | 0.06 |
| 3 | 725030 | 2020-12-30 | 35.2 | 8.7 | 0.0 |
| 4 | 744860 | 2020-12-30 | 32.5 | 8.3 | 0.0 |
| 5 | 725030 | 2020-12-29 | 42.1 | 13.6 | 0.0 |
| 6 | 744860 | 2020-12-29 | 39.7 | 13.8 | 0.0 |
| 7 | 725030 | 2020-12-28 | 42.9 | 5.3 | 0.0 |
| 8 | 744860 | 2020-12-28 | 41.1 | 8.6 | 0.0 |
| 9 | 725030 | 2020-12-27 | 31.6 | 8.0 | 0.0 |
| 10 | 744860 | 2020-12-27 | 29.5 | 9.0 | 0.0 |
| 11 | 725030 | 2020-12-26 | 31.1 | 14.5 | 0.22 |
| 12 | 744860 | 2020-12-26 | 29.5 | 18.7 | 0.22 |
| 13 | 725030 | 2020-12-25 | 56.4 | 19.9 | 0.69 |
| 14 | 744860 | 2020-12-25 | 53.5 | 21.7 | 0.49 |
| 15 | 725030 | 2020-12-24 | 50.1 | 11.9 | 0.0 |
| 16 | 744860 | 2020-12-24 | 49.4 | 14.6 | 0.0 |
| 17 | 725030 | 2020-12-23 | 42.4 | 9.1 | 0.0 |
| 18 | 744860 | 2020-12-23 | 40.1 | 9.2 | 0.0 |
| 19 | 725030 | 2020-12-22 | 43.0 | 10.1 | 0.0 |

Menyimpan hasil *query* ke *table*



|  | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | First | Last | Month | Day | Year |  |  |  |  | CONCATENATE() |
| 2 | George | Washington | April | 30 | 1789 | George Washi | April 30, 1789 |  | Washington | 30 April 1789 |
| 3 | John | Adams | March | 4 | 1797 | John Adams | March 4, 1797 |  | Monroe | 4 March 1797 |
| 4 | Thomas | Jefferson | March | 4 | 1801 | Thomas Jeffer | March 4, 1801 |  | Pierce | 4 March 1801 |
| 5 | James | Madison | March | 4 | 1809 | James Madiso | March 4, 1809 |  | Monroe | 4 March 1809 |
| 6 | James | Monroe | March | 4 | 1817 | James Monro | March 4, 1817 |  | Monroe | 4 March 1817 |
| 7 | John Quincy | Adams | March | 4 | 1825 | John Quincy A | March 4, 1825 |  | Jackson | 4 March 1825 |
| 8 | Andrew | Jackson | March | 4 | 1829 | Andrew Jacks | March 4, 1829 |  | #N/A | 4 March 1829 |
| 9 | Martin | Van Buren | March | 4 | 1837 | Martin Van Bu | March 4, 1837 |  | Polk | 4 March 1837 |
| 10 | William Henry | Harrison | March | 4 | 1841 | William Henry | March 4, 1841 |  | Pierce | 4 March 1841 |
| 11 | John | Tyler | April | 4 | 1841 | John Tyler | April 4, 1841 |  | Monroe | 4 April 1841 |
| 12 | James K. | Polk | March | 4 | 1845 | James K. Polk | March 4, 1845 |  | Monroe | 4 March 1845 |
| 13 | Zachary | Taylor | March | 4 | 1849 | Zachary Taylo | March 4, 1849 |  | Pierce | 4 March 1849 |
| 14 | Millard | Fillmore | July | 9 | 1850 | Millard Fillmo | July 9, 1850 |  | Pierce | 9 July 1850 |
| 15 | Franklin | Pierce | March | 4 | 1853 | Franklin Pierc | March 4, 1853 |  |  |  |

Menggabungkan 2 atau lebih kolom dengan *CONCATENATE*



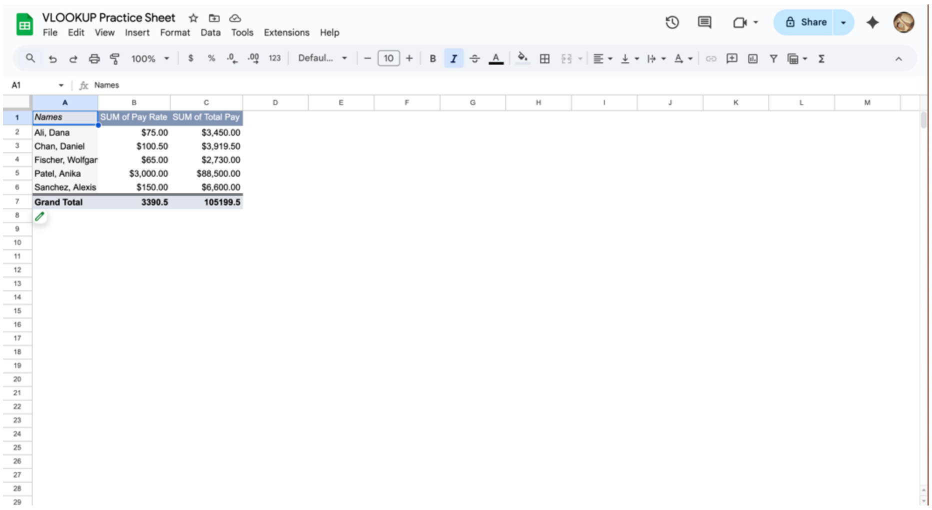
J15 `=VLOOKUP(A2, Sheet2!$A$2:$D$6, 4, false)`

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ID | Name | 1/1/2020 | 1/2/2020 | 1/3/2020 | 1/4/2020 | 1/5/2020 | 1/6/2020 | Total Pay |
| 2 | G001 | Chan, Daniel | 8 | 8 | 8.5 | 7 | 5 | 2.5 | |
| 3 | G002 | Ali, Dana | 8.5 | 7 | 8 | 8 | 9 | 5.5 | |
| 4 | G003 | Sanchez, Alexis | 7.5 | 6.5 | 10 | 8 | 7 | 5 | |
| 5 | G004 | Fischer, Wolfgang | 8 | 8 | 8 | 7 | 7 | 4 | |
| 6 | G005 | Patel, Anika | 6 | 5 | 5 | 5.5 | 6 | 2 | |

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | | Names | 1/1/2020 | 1/2/2020 | 1/3/2020 | 1/4/2020 | 1/5/2020 | 1/6/2020 | Hours | Pay Rate | Total Pay |
| 15 | | Chan, Daniel | 8 | 8 | 8.5 | 7 | 5 | 2.5 | 39 | 100.5 | 3919.5 |
| 16 | | Ali, Dana | 8.5 | 7 | 8 | 8 | 9 | 5.5 | 46 | 75 | 3450 |
| 17 | | Sanchez, Alexis | 7.5 | 6.5 | 10 | 8 | 7 | 5 | 44 | 150 | 6600 |
| 18 | | Fischer, Wolfgang | 8 | 8 | 8 | 7 | 7 | 4 | 42 | 65 | 2730 |
| 19 | | Patel, Anika | 6 | 5 | 5 | 5.5 | 6 | 2 | 29.5 | 3000 | 88500 |

| | A | B | C | D |
|---|---|---|---|---|
| 1 | ID | DOH | Status | Pay Rate |
| 2 | G001 | 12/20/2010 | On Leave | 100.5 |
| 3 | G002 | 1/5/2010 | Contractor | 75 |
| 4 | G003 | 11/11/2011 | Full-Time | 150 |
| 5 | G004 | 5/12/2018 | Contractor | 65 |
| 6 | G005 | 1/2/2020 | Full-Time | 3000 |

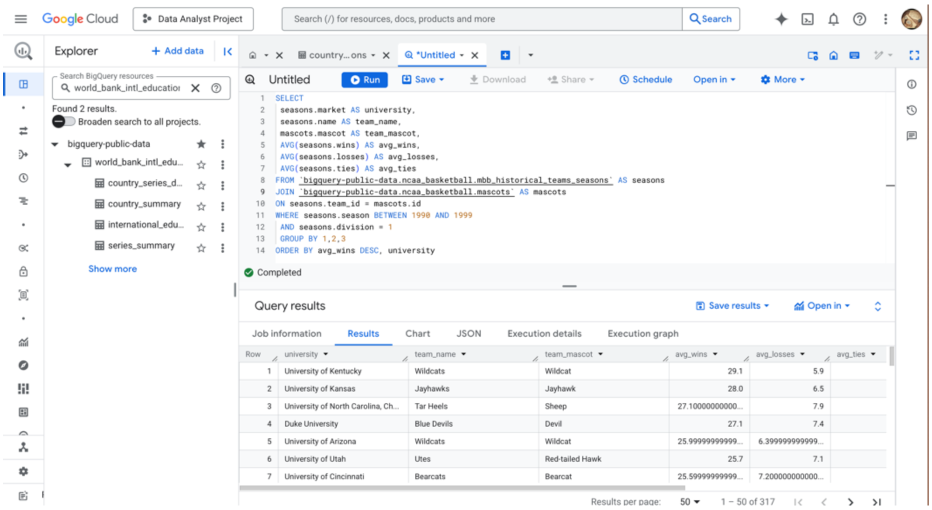
*VLOOKUP* dengan *spreadsheet*



A1 Names

| | A | B | C |
|---|---|---|---|
| 1 | *Names* | SUM of Pay Rate | SUM of Total Pay |
| 2 | Ali, Dana | $75.00 | $3,450.00 |
| 3 | Chan, Daniel | $100.50 | $3,919.50 |
| 4 | Fischer, Wolfgar | $65.00 | $2,730.00 |
| 5 | Patel, Anika | $3,000.00 | $88,500.00 |
| 6 | Sanchez, Alexis | $150.00 | $6,600.00 |
| 7 | **Grand Total** | **3390.5** | **105199.5** |

*Pivot table*

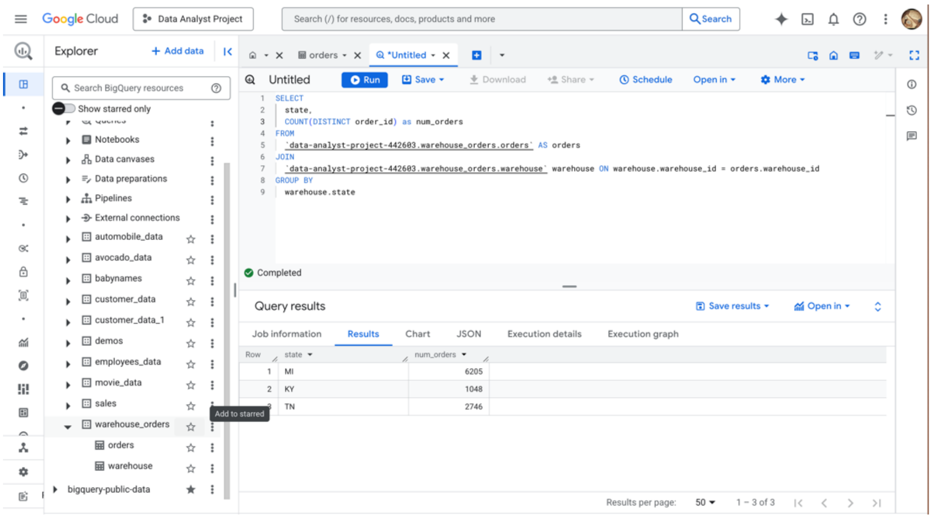


```sql
SELECT
 seasons.market AS university,
 seasons.name AS team_name,
 mascots.mascot AS team_mascot,
 AVG(seasons.wins) AS avg_wins,
 AVG(seasons.losses) AS avg_losses,
 AVG(seasons.ties) AS avg_ties
FROM `bigquery-public-data.ncaa_basketball.mbb_historical_teams_seasons` AS seasons
JOIN `bigquery-public-data.ncaa_basketball.mascots` AS mascots
ON seasons.team_id = mascots.id
WHERE seasons.season BETWEEN 1990 AND 1999
 AND seasons.division = 1
 GROUP BY 1,2,3
ORDER BY avg_wins DESC, university
```

| Row | university | team_name | team_mascot | avg_wins | avg_losses | avg_ties |
|---|---|---|---|---|---|---|
| 1 | University of Kentucky | Wildcats | Wildcat | 29.1 | 5.9 | |
| 2 | University of Kansas | Jayhawks | Jayhawk | 28.0 | 6.5 | |
| 3 | University of North Carolina, Ch... | Tar Heels | Sheep | 27.10000000000... | 7.9 | |
| 4 | Duke University | Blue Devils | Devil | 27.1 | 7.4 | |
| 5 | University of Arizona | Wildcats | Wildcat | 25.99999999999... | 6.399999999999... | |
| 6 | University of Utah | Utes | Red-tailed Hawk | 25.7 | 7.1 | |
| 7 | University of Cincinnati | Bearcats | Bearcat | 25.59999999999... | 7.200000000000... | |

Results per page: 50 1 – 50 of 317

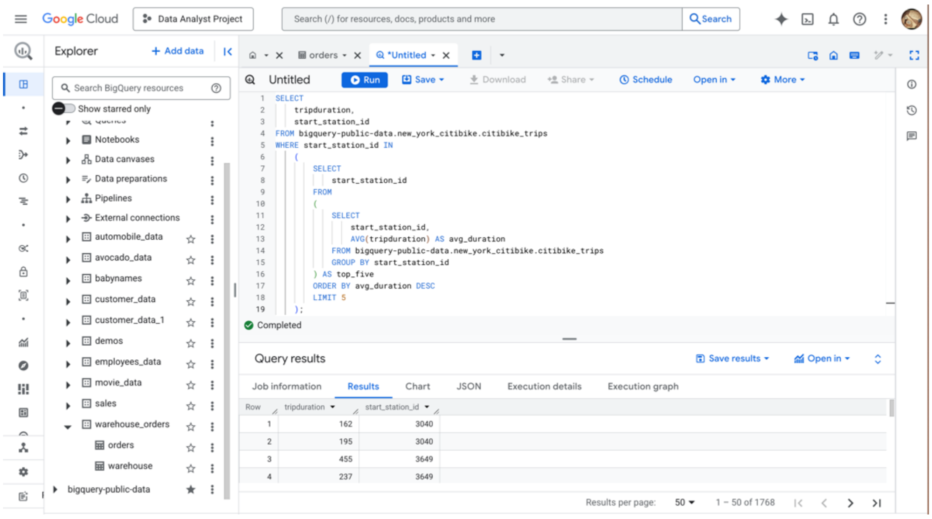
*INNER JOIN* pada *query SQL*



```sql
SELECT
  state,
  COUNT(DISTINCT order_id) as num_orders
FROM
  `data-analyst-project-442603.warehouse_orders.orders` AS orders
JOIN
  `data-analyst-project-442603.warehouse_orders.warehouse` warehouse ON warehouse.warehouse_id = orders.warehouse_id
GROUP BY
  warehouse.state
```

| Row | state | num_orders |
|---|---|---|
| 1 | MI | 6205 |
| 2 | KY | 1048 |
| 3 | TN | 2746 |

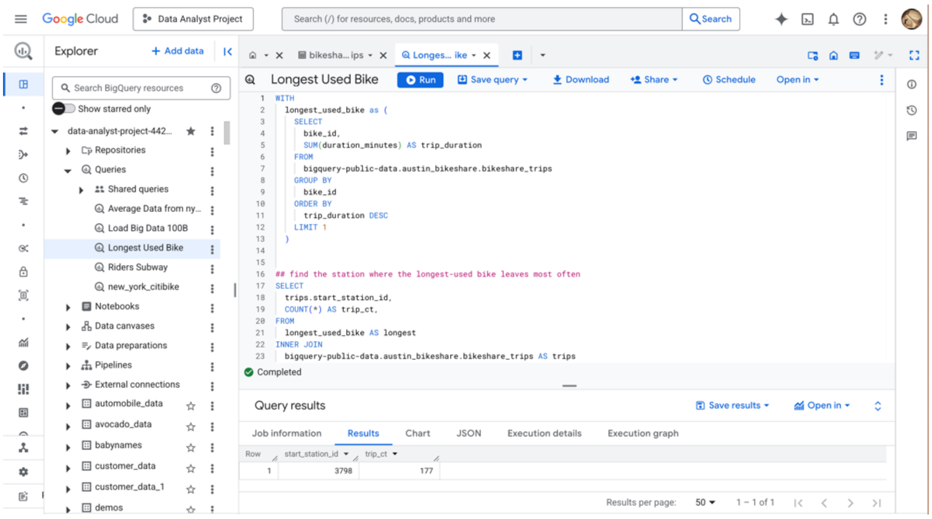
*Count* pada *query SQL*



```sql
SELECT
    tripduration,
    start_station_id
FROM bigquery-public-data.new_york_citibike.citibike_trips
WHERE start_station_id IN
    (
        SELECT
            start_station_id
        FROM
        (
            SELECT
                start_station_id,
                AVG(tripduration) AS avg_duration
            FROM bigquery-public-data.new_york_citibike.citibike_trips
            GROUP BY start_station_id
        ) AS top_five
        ORDER BY avg_duration DESC
        LIMIT 5
    );
```

| Row | tripduration | start_station_id |
|---|---|---|
| 1 | 162 | 3040 |
| 2 | 195 | 3040 |
| 3 | 455 | 3649 |
| 4 | 237 | 3649 |

*Subquery* pada *SQL*

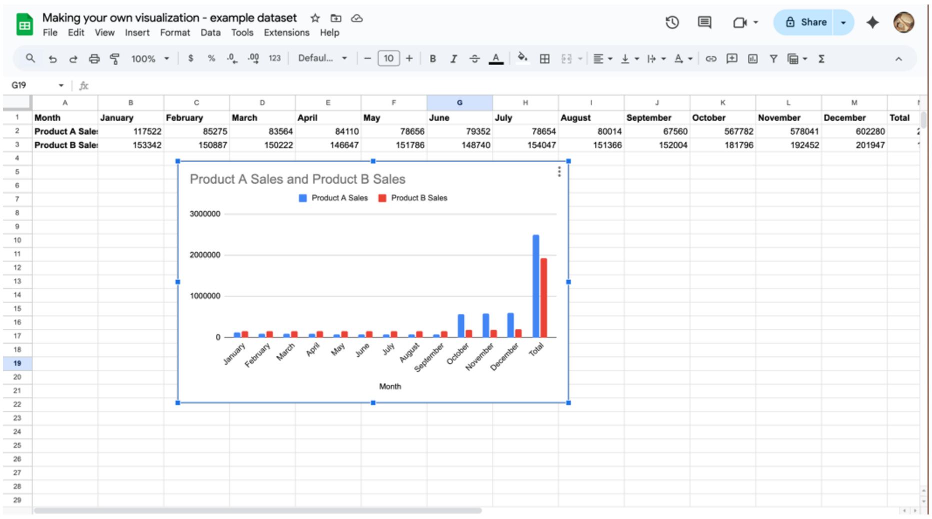


```sql
WITH
  longest_used_bike as (
    SELECT
      bike_id,
      SUM(duration_minutes) AS trip_duration
    FROM
      bigquery-public-data.austin_bikeshare.bikeshare_trips
    GROUP BY
      bike_id
    ORDER BY
      trip_duration DESC
    LIMIT 1
  )


## find the station where the longest-used bike leaves most often
SELECT
  trips.start_station_id,
  COUNT(*) AS trip_ct,
FROM
  longest_used_bike AS longest
INNER JOIN
  bigquery-public-data.austin_bikeshare.bikeshare_trips AS trips
```

| Row | start_station_id | trip_ct |
|---|---|---|
| 1 | 3798 | 177 |

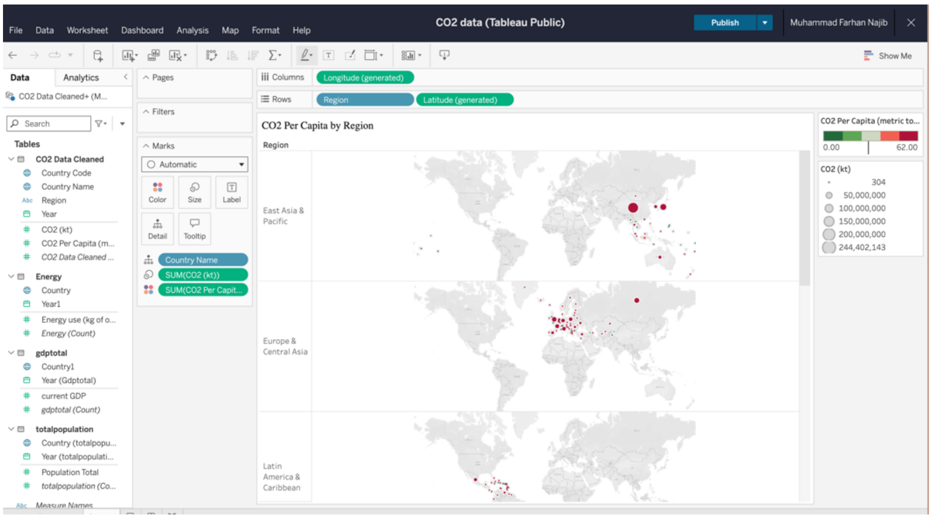
*Temporary Table*

# Course 6

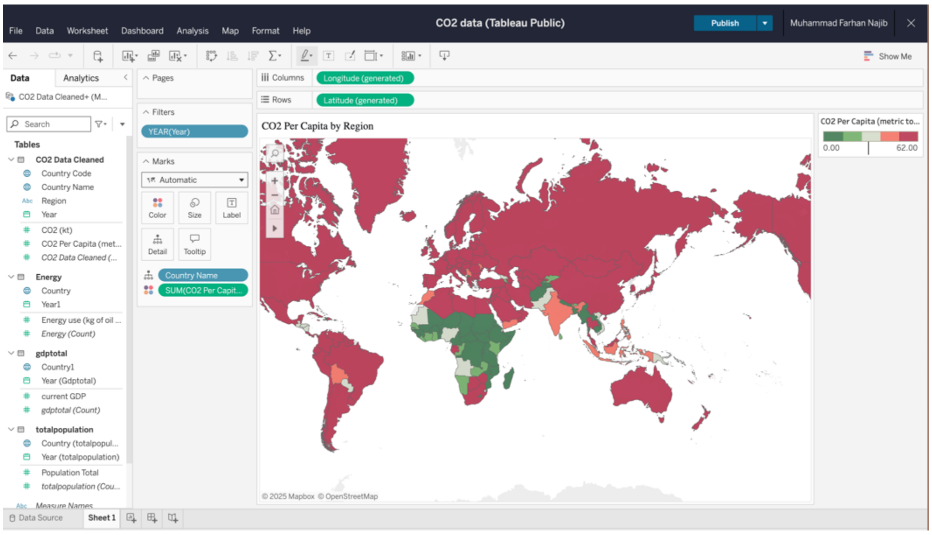


| Month | January | February | March | April | May | June | July | August | September | October | November | December | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Product A Sale | 117522 | 85275 | 83564 | 84110 | 78656 | 79352 | 78654 | 80014 | 67560 | 567782 | 578041 | 602280 | |
| Product B Sale | 153342 | 150887 | 150222 | 146647 | 151786 | 148740 | 154047 | 151366 | 152004 | 181796 | 192452 | 201947 | |

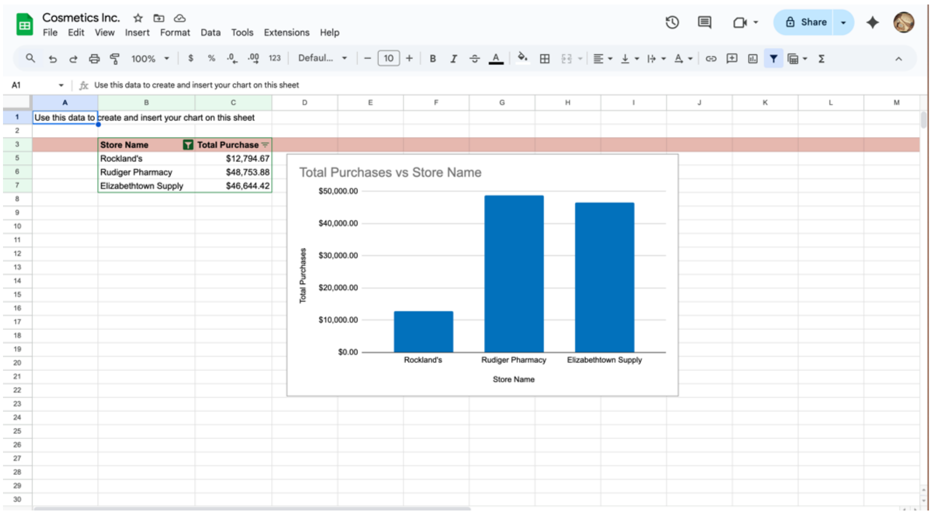
Visualisasi dengan *Excel*



Visualisasi dengan *Tableau*



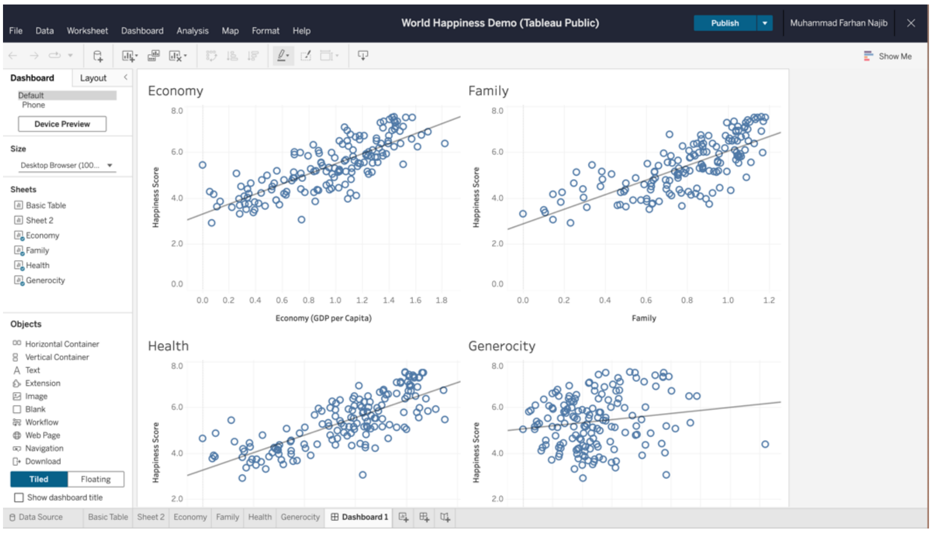
Visualisasi dengan lebih dari 1 sumber data



Use this data to create and insert your chart on this sheet

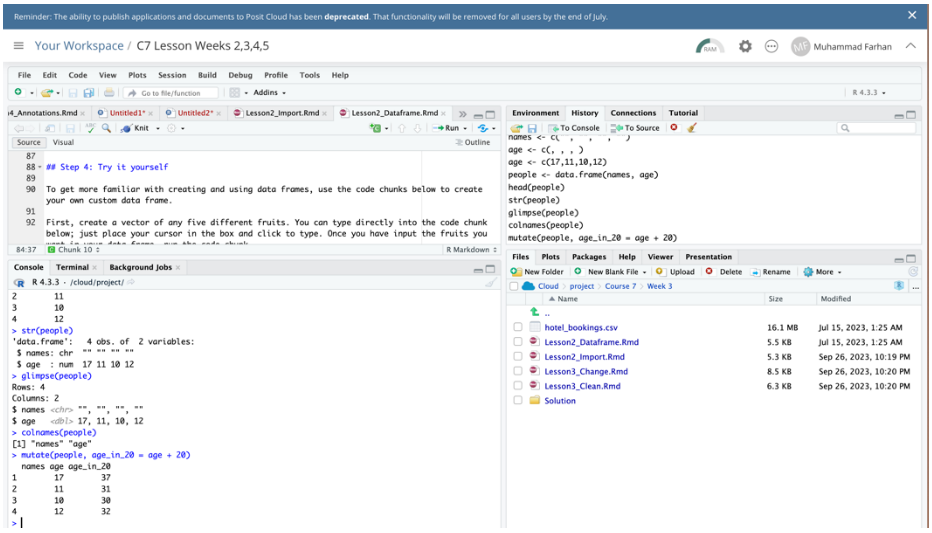
| Store Name | Total Purchase |
|---|---|
| Rockland's | $12,794.67 |
| Rudiger Pharmacy | $48,753.88 |
| Elizabethtown Supply | $46,644.42 |

Membuat diagram, filter dan kustomisasi



Membuat *dashboard* pada *Tableau*

# Course 7



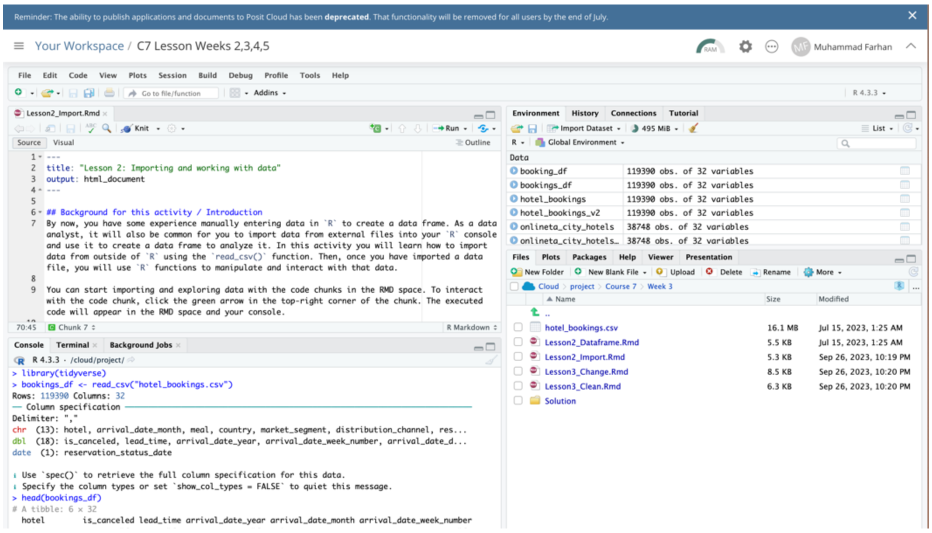
```
## Step 4: Try it yourself

To get more familiar with creating and using data frames, use the code chunks below to create
your own custom data frame.

First, create a vector of any five different fruits. You can type directly into the code chunk
below; just place your cursor in the box and click to type. Once you have input the fruits you
```

```
2        11
3        10
4        12
> str(people)
'data.frame':	4 obs. of  2 variables:
 $ names: chr  "" "" "" ""
 $ age  : num  17 11 10 12
> glimpse(people)
Rows: 4
Columns: 2
$ names <chr> "", "", "", ""
$ age   <dbl> 17, 11, 10, 12
> colnames(people)
[1] "names" "age"
> mutate(people, age_in_20 = age + 20)
  names age age_in_20
1        17        37
2        11        31
3        10        30
4        12        32
> |
```

```
age <- c(, , , )
age <- c(17,11,10,12)
people <- data.frame(names, age)
head(people)
str(people)
glimpse(people)
colnames(people)
mutate(people, age_in_20 = age + 20)
```

| Name | Size | Modified |
|---|---|---|
| .. | | |
| hotel_bookings.csv | 16.1 MB | Jul 15, 2023, 1:25 AM |
| Lesson2_Dataframe.Rmd | 5.5 KB | Jul 15, 2023, 1:25 AM |
| Lesson2_Import.Rmd | 5.3 KB | Sep 26, 2023, 10:19 PM |
| Lesson3_Change.Rmd | 8.5 KB | Sep 26, 2023, 10:20 PM |
| Lesson3_Clean.Rmd | 6.3 KB | Sep 26, 2023, 10:20 PM |
| Solution | | |

Membuat *dataframe* pada R


Reminder: The ability to publish applications and documents to Posit Cloud has been deprecated. That functionality will be removed for all users by the end of July.
Your Workspace / C7 Lesson Weeks 2,3,4,5
Muhammad Farhan
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.3
Lesson2_Import.Rmd
Knit
Run
Source
Visual
Outline
70:45
Chunk 7
R Markdown
Console
Terminal
Background Jobs
R 4.3.3 · /cloud/project/
Environment
History
Connections
Tutorial
Import Dataset
495 MiB
List
Global Environment
Files
Plots
Packages
Help
Viewer
Presentation
New Folder
New Blank File
Upload
Delete
Rename
More
Cloud > project > Course 7 > Week 3

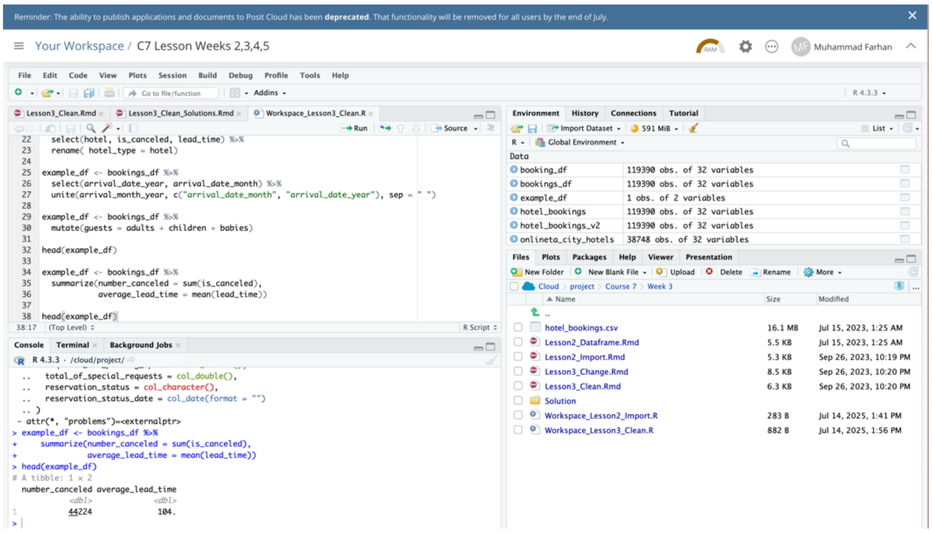
```
---
title: "Lesson 2: Importing and working with data"
output: html_document
---

## Background for this activity / Introduction
By now, you have some experience manually entering data in `R` to create a data frame. As a data
analyst, it will also be common for you to import data from external files into your `R` console
and use it to create a data frame to analyze it. In this activity you will learn how to import
data from outside of `R` using the `read_csv()` function. Then, once you have imported a data
file, you will use `R` functions to manipulate and interact with that data.

You can start importing and exploring data with the code chunks in the RMD space. To interact
with the code chunk, click the green arrow in the top-right corner of the chunk. The executed
code will appear in the RMD space and your console.
```

```
> library(tidyverse)
> bookings_df <- read_csv("hotel_bookings.csv")
Rows: 119390 Columns: 32
── Column specification ────────────────────────────────────────────────────
Delimiter: ","
chr  (13): hotel, arrival_date_month, meal, country, market_segment, distribution_channel, res...
dbl  (18): is_canceled, lead_time, arrival_date_year, arrival_date_week_number, arrival_date_d...
date  (1): reservation_status_date

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
> head(bookings_df)
# A tibble: 6 × 32
  hotel        is_canceled lead_time arrival_date_year arrival_date_month arrival_date_week_number
```

| Data | |
|---|---|
| booking_df | 119390 obs. of 32 variables |
| bookings_df | 119390 obs. of 32 variables |
| hotel_bookings | 119390 obs. of 32 variables |
| hotel_bookings_v2 | 119390 obs. of 32 variables |
| onlineta_city_hotels | 38748 obs. of 32 variables |
| onlineta_city_hotels… | 38748 obs. of 32 variables |

| Name | Size | Modified |
|---|---|---|
| .. | | |
| hotel_bookings.csv | 16.1 MB | Jul 15, 2023, 1:25 AM |
| Lesson2_Dataframe.Rmd | 5.5 KB | Jul 15, 2023, 1:25 AM |
| Lesson2_Import.Rmd | 5.3 KB | Sep 26, 2023, 10:19 PM |
| Lesson3_Change.Rmd | 8.5 KB | Sep 26, 2023, 10:20 PM |
| Lesson3_Clean.Rmd | 6.3 KB | Sep 26, 2023, 10:20 PM |
| Solution | | |

*Import* data pada R



Reminder: The ability to publish applications and documents to Posit Cloud has been **deprecated**. That functionality will be removed for all users by the end of July.

Your Workspace / C7 Lesson Weeks 2,3,4,5
Muhammad Farhan
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.3
Lesson3_Clean.Rmd
Lesson3_Clean_Solutions.Rmd
Workspace_Lesson3_Clean.R
Run
Source
38:17
(Top Level)
R Script
Console Terminal Background Jobs
R 4.3.3 · /cloud/project/
Environment History Connections Tutorial
Import Dataset
591 MiB
List
Global Environment
Files Plots Packages Help Viewer Presentation
New Folder
New Blank File
Upload
Delete
Rename
More
Cloud > project > Course 7 > Week 3

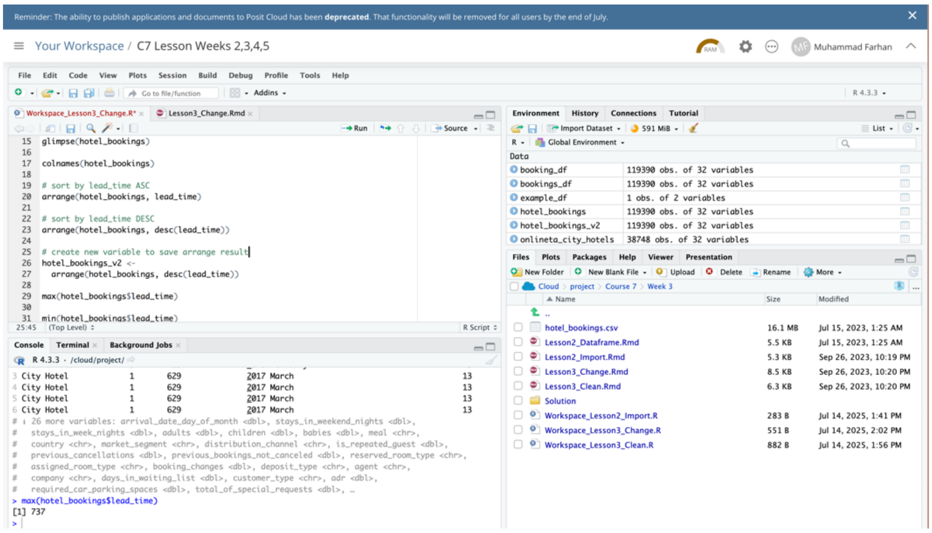
```r
  select(hotel, is_canceled, lead_time) %>%
  rename( hotel_type = hotel)

example_df <- bookings_df %>%
  select(arrival_date_year, arrival_date_month) %>%
  unite(arrival_month_year, c("arrival_date_month", "arrival_date_year"), sep = " ")

example_df <- bookings_df %>%
  mutate(guests = adults + children + babies)

head(example_df)

example_df <- bookings_df %>%
  summarize(number_canceled = sum(is_canceled),
            average_lead_time = mean(lead_time))

head(example_df)
```

```
 ..   total_of_special_requests = col_double(),
 ..   reservation_status = col_character(),
 ..   reservation_status_date = col_date(format = "")
 .. )
 - attr(*, "problems")=<externalptr>
> example_df <- bookings_df %>%
+     summarize(number_canceled = sum(is_canceled),
+               average_lead_time = mean(lead_time))
> head(example_df)
# A tibble: 1 × 2
  number_canceled average_lead_time
            <dbl>             <dbl>
1           44224              104.
> 
```

| Data | |
|---|---|
| booking_df | 119390 obs. of 32 variables |
| bookings_df | 119390 obs. of 32 variables |
| example_df | 1 obs. of 2 variables |
| hotel_bookings | 119390 obs. of 32 variables |
| hotel_bookings_v2 | 119390 obs. of 32 variables |
| onlineta_city_hotels | 38748 obs. of 32 variables |

| Name | Size | Modified |
|---|---|---|
| .. | | |
| hotel_bookings.csv | 16.1 MB | Jul 15, 2023, 1:25 AM |
| Lesson2_Dataframe.Rmd | 5.5 KB | Jul 15, 2023, 1:25 AM |
| Lesson2_Import.Rmd | 5.3 KB | Sep 26, 2023, 10:19 PM |
| Lesson3_Change.Rmd | 8.5 KB | Sep 26, 2023, 10:20 PM |
| Lesson3_Clean.Rmd | 6.3 KB | Sep 26, 2023, 10:20 PM |
| Solution | | |
| Workspace_Lesson2_Import.R | 283 B | Jul 14, 2025, 1:41 PM |
| Workspace_Lesson3_Clean.R | 882 B | Jul 14, 2025, 1:56 PM |

*Clean* data pada R



Reminder: The ability to publish applications and documents to Posit Cloud has been **deprecated**. That functionality will be removed for all users by the end of July.

Your Workspace / C7 Lesson Weeks 2,3,4,5
Muhammad Farhan
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.3
Workspace_Lesson3_Change.R*
Lesson3_Change.Rmd
Run
Source
25:45
(Top Level)
R Script
Console Terminal Background Jobs
R 4.3.3 · /cloud/project/
Environment History Connections Tutorial
Import Dataset
591 MiB
List
Global Environment
Files Plots Packages Help Viewer Presentation
New Folder
New Blank File
Upload
Delete
Rename
More
Cloud > project > Course 7 > Week 3

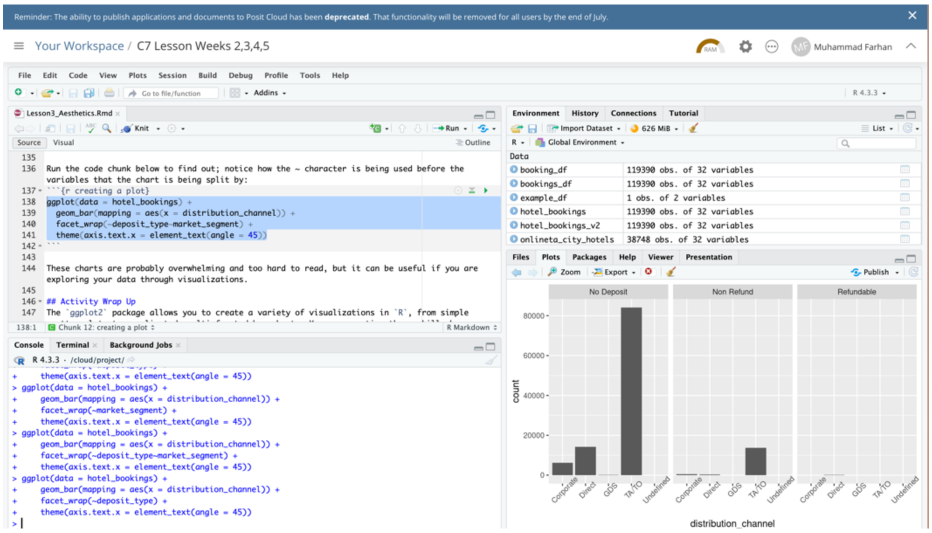
```r
glimpse(hotel_bookings)

colnames(hotel_bookings)

# sort by lead_time ASC
arrange(hotel_bookings, lead_time)

# sort by lead_time DESC
arrange(hotel_bookings, desc(lead_time))

# create new variable to save arrange result
hotel_bookings_v2 <-
  arrange(hotel_bookings, desc(lead_time))

max(hotel_bookings$lead_time)

min(hotel_bookings$lead_time)
```

```
3 City Hotel            1        629             2017 March                              13
4 City Hotel            1        629             2017 March                              13
5 City Hotel            1        629             2017 March                              13
6 City Hotel            1        629             2017 March                              13
# ℹ 26 more variables: arrival_date_day_of_month <dbl>, stays_in_weekend_nights <dbl>,
#   stays_in_week_nights <dbl>, adults <dbl>, children <dbl>, babies <dbl>, meal <chr>,
#   country <chr>, market_segment <chr>, distribution_channel <chr>, is_repeated_guest <dbl>,
#   previous_cancellations <dbl>, previous_bookings_not_canceled <dbl>, reserved_room_type <chr>,
#   assigned_room_type <chr>, booking_changes <dbl>, deposit_type <chr>, agent <chr>,
#   company <chr>, days_in_waiting_list <dbl>, customer_type <chr>, adr <dbl>,
#   required_car_parking_spaces <dbl>, total_of_special_requests <dbl>, …
> max(hotel_bookings$lead_time)
[1] 737
> 
```

| Data | |
|---|---|
| booking_df | 119390 obs. of 32 variables |
| bookings_df | 119390 obs. of 32 variables |
| example_df | 1 obs. of 2 variables |
| hotel_bookings | 119390 obs. of 32 variables |
| hotel_bookings_v2 | 119390 obs. of 32 variables |
| onlineta_city_hotels | 38748 obs. of 32 variables |

| Name | Size | Modified |
|---|---|---|
| .. | | |
| hotel_bookings.csv | 16.1 MB | Jul 15, 2023, 1:25 AM |
| Lesson2_Dataframe.Rmd | 5.5 KB | Jul 15, 2023, 1:25 AM |
| Lesson2_Import.Rmd | 5.3 KB | Sep 26, 2023, 10:19 PM |
| Lesson3_Change.Rmd | 8.5 KB | Sep 26, 2023, 10:20 PM |
| Lesson3_Clean.Rmd | 6.3 KB | Sep 26, 2023, 10:20 PM |
| Solution | | |
| Workspace_Lesson2_Import.R | 283 B | Jul 14, 2025, 1:41 PM |
| Workspace_Lesson3_Change.R | 551 B | Jul 14, 2025, 2:02 PM |
| Workspace_Lesson3_Clean.R | 882 B | Jul 14, 2025, 1:56 PM |

Manipulasi data pada R


Reminder: The ability to publish applications and documents to Posit Cloud has been **deprecated**. That functionality will be removed for all users by the end of July.
Your Workspace / C7 Lesson Weeks 2,3,4,5
Muhammad Farhan
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.3
Lesson3_Aesthetics.Rmd
Knit
Run
Source Visual
Outline
138:1
Chunk 12: creating a plot
R Markdown
Environment History Connections Tutorial
Import Dataset
626 MiB
List
Global Environment
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
No Deposit
Non Refund
Refundable
count
0
20000
40000
60000
80000
Corporate
Direct
GDS
TA/TO
Undefined
distribution_channel

~~~
Run the code chunk below to find out; notice how the ~ character is being used before the
variables that the chart is being split by:
```{r creating a plot}
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel)) +
  facet_wrap(~deposit_type~market_segment) +
  theme(axis.text.x = element_text(angle = 45))
```

These charts are probably overwhelming and too hard to read, but it can be useful if you are
exploring your data through visualizations.

## Activity Wrap Up
The `ggplot2` package allows you to create a variety of visualizations in `R`, from simple
~~~

Console Terminal Background Jobs

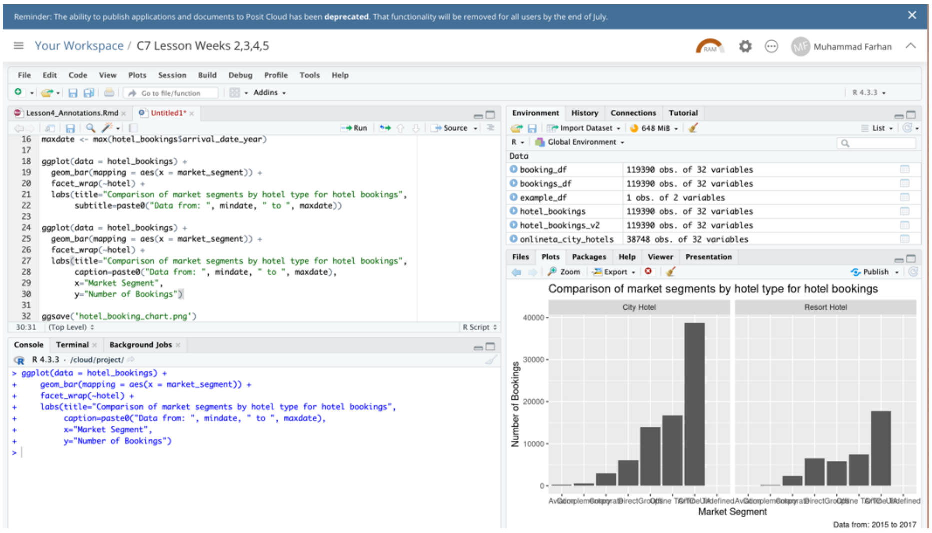
R 4.3.3 · /cloud/project/

```
+     theme(axis.text.x = element_text(angle = 45))
> ggplot(data = hotel_bookings) +
+     geom_bar(mapping = aes(x = distribution_channel)) +
+     facet_wrap(~market_segment) +
+     theme(axis.text.x = element_text(angle = 45))
> ggplot(data = hotel_bookings) +
+     geom_bar(mapping = aes(x = distribution_channel)) +
+     facet_wrap(~deposit_type~market_segment) +
+     theme(axis.text.x = element_text(angle = 45))
> ggplot(data = hotel_bookings) +
+     geom_bar(mapping = aes(x = distribution_channel)) +
+     facet_wrap(~deposit_type) +
+     theme(axis.text.x = element_text(angle = 45))
> |
```

Data

```
booking_df            119390 obs. of 32 variables
bookings_df           119390 obs. of 32 variables
example_df            1 obs. of 2 variables
hotel_bookings        119390 obs. of 32 variables
hotel_bookings_v2     119390 obs. of 32 variables
onlineta_city_hotels  38748 obs. of 32 variables
```

Visualisasi data dengan *ggplot2*



```
maxdate <- max(hotel_bookings$arrival_date_year)

ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = market_segment)) +
  facet_wrap(~hotel) +
  labs(title="Comparison of market segments by hotel type for hotel bookings",
       subtitle=paste0("Data from: ", mindate, " to ", maxdate))

ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = market_segment)) +
  facet_wrap(~hotel) +
  labs(title="Comparison of market segments by hotel type for hotel bookings",
       caption=paste0("Data from: ", mindate, " to ", maxdate),
       x="Market Segment",
       y="Number of Bookings")

ggsave('hotel_booking_chart.png')
```

Console Terminal Background Jobs

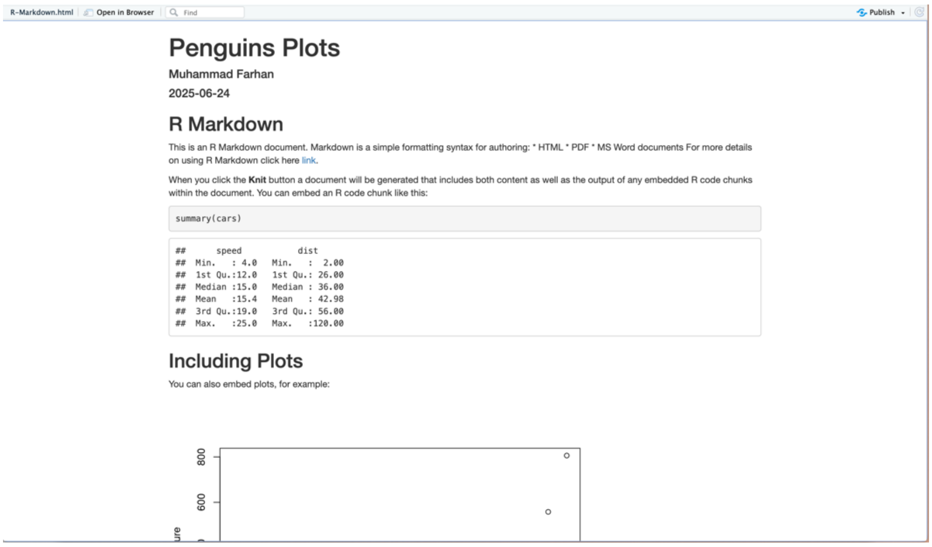
R 4.3.3 · /cloud/project/

```
> ggplot(data = hotel_bookings) +
+     geom_bar(mapping = aes(x = market_segment)) +
+     facet_wrap(~hotel) +
+     labs(title="Comparison of market segments by hotel type for hotel bookings",
+          caption=paste0("Data from: ", mindate, " to ", maxdate),
+          x="Market Segment",
+          y="Number of Bookings")
> |
```

Data

```
booking_df            119390 obs. of 32 variables
bookings_df           119390 obs. of 32 variables
example_df            1 obs. of 2 variables
hotel_bookings        119390 obs. of 32 variables
hotel_bookings_v2     119390 obs. of 32 variables
onlineta_city_hotels  38748 obs. of 32 variables
```

Memberikan anotasi dan menyimpan hasil visualisasi



# Penguins Plots

Muhammad Farhan

2025-06-24

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring: * HTML * PDF * MS Word documents For more details on using R Markdown click here link.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

## Including Plots

You can also embed plots, for example:

Membuat dokumentasi dengan *R Markdown*