Week 1

1. A data analyst at a book publisher is working on an urgent report for executives. They are using only historical data. What is the most likely reason for choosing to analyze only historical data?

   - ○ The data is unknown
   - ○ The data is constantly changing
   - ◉ The project has a very short time frame
   - ○ There is plenty of time to research historical data

   Correct
   The most likely reason for choosing to analyze only historical data is that a project has a very short time frame.

2. Which of the following are examples of discrete data? Select all that apply.

   - ☑ Number of actors in movie

   Correct
   The number of actors in a movie, box office returns, and the movie budget are examples of discrete data.

   - ☑ Box office returns

   Correct
   The number of actors in a movie, box office returns, and the movie budget are examples of discrete data.

   - ☐ Movie running time
   - ☑ Movie budget

   Correct
   The number of actors in a movie, box office returns, and the movie budget are examples of discrete data.

3. Nominal qualitative data has a set order or scale.

- ○ True
- ◉ False

Correct
Nominal qualitative data does not have a set order or scale.

4. Why is internal data considered more reliable and easier to collect than external data?

- ○ Internal data circumvents privacy restrictions.
- ○ Internal data has much larger sample sizes.
- ◉ Internal data lives within a company's own systems.
- ○ Internal data comes from people you know.

Correct
Internal data is considered more reliable and easier to collect than external data because it lives within a company's own systems.

5. Which of the following is an example of structured data?

- ○ Video file
- ○ Digital photo
- ○ Audio file
- ◉ Relational database

Correct
A relational database is an example of structured data.

6. Which of the following values are examples of a Boolean data type? Select all that apply.

- ☐ Yes, no, or unsure
- ☑ Yes or no

Correct
True or false and yes or no are examples of a Boolean data type.

- ☐ One, two, or three
- ☑ True or false

Correct
True or false and yes or no are examples of a Boolean data type.

7. The following is a selection from a spreadsheet:

Name	Age	Occupation
Agnes Shipton	44	Entrepreneur
Ronaldo Vincent	23	Accountant
Henry Sing	36	Editor
Krishna Bowling	62	Graphic designer

What kind of data format does it contain?

- ○ Long
- ○ Short
- ○ Narrow
- ◉ Wide

Correct
The selection from the spreadsheet contains wide data.

8. Fill in the blank: Data transformation enables data analysts to change the _____ of the data.

- ◉ structure
- ○ meaning
- ○ value
- ○ accuracy

Correct
Data transformation enables data analysts to change the structure of the data.

Week 2

1. A clinic surveys a group of male and female patients about their experience with physical therapy. The survey does not include people with disabilities. Is the survey data biased?

- (●) Yes
- ( ) No

Correct
The survey data is biased because the sample group lacks inclusivity.

2. Which type of bias is the tendency to always construe ambiguous situations in a positive or negative way?

- (●) Interpretation
- ( ) Sampling
- ( ) Confirmation
- ( ) Observer

Correct
Interpretation bias is the tendency to always construe ambiguous situations in a positive or negative way.

3. Which of the following are qualities of unreliable data? Select all that apply.

- [ ] Vetted
- [x] Inaccurate

Correct
Unreliable data is inaccurate, incomplete, and biased.

- [x] Biased

Correct
Unreliable data is inaccurate, incomplete, and biased.

- [x] Incomplete

Correct
Unreliable data is inaccurate, incomplete, and biased.

4. In data ethics, consent gives an individual the right to know the answers to which of the following questions? Select all that apply.

- [x] How long will my data be stored?

Correct
In data ethics, consent gives individuals the right to know why their data is being collected, how it will be used, and how long it will be stored.

- [x] How will my data be used?

Correct
In data ethics, consent gives individuals the right to know why their data is being collected, how it will be used, and how long it will be stored.

- [ ] Why am I being forced to share my data?
- [ ] Why is my data being collected?

You didn't select all the correct answers

5. An individual who provides their data has the right to know and understand all of the data-processing activities and algorithms used on that data. This concept refers to which aspect of data ethics?

- ( ) Ownership
- (•) Transaction transparency
- ( ) Currency
- ( ) Consent

Correct
This refers to transaction transparency, which is the idea that an individual who provides their data has the right to know and understand all of the data-processing activities and algorithms used on that data.

6. The right to inspect, update, or correct your own data is part of which aspect of data ethics?

- ( ) Data consent
- ( ) Data openness
- ( ) Data ownership
- (•) Data privacy

Correct
The right to inspect, update, or correct your own data is part of data privacy.

7. What is the process of protecting people's private or sensitive data by eliminating identifying information?

- ○ Data governance
- ○ Data ethics
- ○ Data design
- ◉ Data anonymization

Correct
Data anonymization removes personally identifiable information, such as home addresses, telephone numbers, credit card numbers, and medical records.

8. A key aspect of open data is free access to people's personal information.

- ○ True
- ◉ False

Correct
Open data does not involve granting free access to people's personal information.

Week 3

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

- ○ Primary
- ◉ Relational
- ○ Metadata
- ○ Normalized

Correct
Primary and foreign keys are two connected identifiers within separate tables in a relational database.

2. Data analysts use metadata for what tasks? Select all that apply.

- [x] To evaluate the quality of data

  Correct
  Data analysts use metadata to combine data, evaluate data, and interpret a database.

- [x] To combine data from more than one source

  Correct
  Data analysts use metadata to combine data, evaluate data, and interpret a database.

- [x] To perform data analyses

  This should not be selected
  Check out the video on metadata to review the material.

- [ ] To interpret the contents of a database

3. Think about data as a student at a high school. In this metaphor, which of the following are examples of metadata? Select all that apply.

- [x] Grades the student earns

  This should not be selected
  Check out the video on metadata to review the material.

- [x] Classes the student is enrolled in

  Correct
  The student ID number, enrollment date, and classes the student is enrolled in represent structural metadata.

- [x] Student's enrollment date

  Correct
  The student ID number, enrollment date, and classes the student is enrolled in represent structural metadata.

- [x] Student's ID number

  Correct
  The student ID number, enrollment date, and classes the student is enrolled in represent structural metadata.

4. What is the process that data analysts use to ensure the formal management of their company's data assets?

- ○ Data mapping
- ◉ Data governance
- ○ Data aggregation
- ○ Data integrity

Correct
Data governance is the process of ensuring the formal management of a company's data assets.

5. In what circumstance might a data analyst choose not to use external data in their analysis?

- ○ The data is too thorough
- ○ The data is free for anyone to access
- ○ The data represents diverse perspectives
- ◉ The data cannot be confirmed to be reliable

Correct
A data analyst might choose not to use external data in their analysis if the data cannot be confirmed to be reliable.

6. A data analyst reviews a database of Wisconsin car sales to find the last car models sold in Milwaukee in 2019. How can they sort and filter the data to return the last five cars sold at the top of their list? Select all that apply.

- ☑ Filter out sales not in 2019

Correct
The analyst can filter out sales outside of Milwaukee in 2019 and sort by date in descending order.

- ☐ Sort by sale date in ascending order
- ☑ Filter out sales outside of Milwaukee

Correct
The analyst can filter out sales outside of Milwaukee in 2019 and sort by date in descending order.

- ☑ Sort by sale date in descending order

Correct
The analyst can filter out sales outside of Milwaukee in 2019 and sort by date in descending order.

7. When writing a query, the name of the dataset can either be inside two backticks, or not, and the query will still run properly.

- (●) True
- ( ) False

Correct
When writing a query, the name of the dataset can either be inside two backticks, or not, and the query will still run properly.

8. You are working with a database table that contains customer data. The *first_name* column lists the first name of each customer. You are only interested in customers with the first name Mark.

You write the SQL query below. Add a WHERE clause that will return only customers named Mark.

```
1  SELECT
2  count(*)
3  FROM
4  customer
5  WHERE first_name='Mark'
```

Run
Reset

How many customers are named Mark?

- ( ) 3
- ( ) 5
- (●) 2
- ( ) 1

Correct
The clause `WHERE first_name = 'Mark'` will return only customers named Mark. The complete query is `SELECT * FROM customer WHERE first_name = 'Mark'`. The WHERE clause filters results that meet certain conditions. The WHERE clause includes the name of the column, an equals sign, and the value(s) in the column to include. Place quotes around text values. There are two customers named Mark.

Week 4

1. A data analytics team labels its files to indicate their content, creation date, and version number. The team is using what data organization tool?

- ○ File-naming attributes
- ○ File-naming references
- ○ File-naming verifications
- ◉ File-naming conventions

Correct
The team is using file-naming conventions, which are consistent guidelines that describe the content, creation date, or version of a file.

2. A data analytics team uses data about data to indicate consistent naming conventions for a project. What type of data is involved in this scenario?

- ◉ Metadata
- ○ Big data
- ○ Aggregated data
- ○ Long data

Correct
Metadata is data about data. Metadata practices can help analytics teams create consistent naming conventions and storage practices for their files.

3. A data analyst creates a file that lists people who donated to their organization's fund drive. An effective name f the file is FundDriveDonors_20210216_V01.

- ◉ True
- ○ False

Correct
FundDriveDonors_20210216_V03 is an effective file name because it is an appropriate length and references the project name, creation date, version.

4. What process do data analysts use to keep project-related files together and organize them into subfolders?

- ○ Editing
- ○ Naming
- ○ Encrypting
- ◉ Foldering

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

5. Data analysts use archiving to separate current from past work. It also cuts down on clutter.

- (●) True
- ( ) False

Correct
Data analysts use archiving to separate current from past work. This involves moving files from completed projects to a separate location.

6. Fill in the blank: Data analysts create _____ to structure their folders.

- ( ) ladders
- (●) hierarchies
- ( ) sequences
- ( ) scales

Correct
Data analysts create hierarchies to structure their folders.

7. A data analyst adds sharing permissions to limit who can edit the data contained within a file. This is an example of what?

- ( ) Data ethics
- ( ) Data integrity
- ( ) Data validation
- (●) Data security

Correct
A data analyst adds sharing permissions to limit who can edit the data contained within a file. This is an example of data security.

8. A data analyst creates a spreadsheet with five tabs. They want to share the data in tabs 1-4 with a client. Tab 5 contains private information about other clients. Which of the following tactics will enable them to keep tab 5 private? Select all that apply.

- [ ] Hide tab 5, then share the spreadsheet with the client.
- [ ] Rename tab 5 to include the word "private" then share the spreadsheet with the client.
- [x] Copy tabs 1-4 into a separate spreadsheet, then share the new file with the client.

Correct
Copying tabs 1-4 into a separate spreadsheet, then sharing the new file with the client will keep tab 5 private. In addition, making a copy of the spreadsheet, deleting tab 5, then sharing the new file with the client will keep tab 5 private.

- [x] Make a copy of the spreadsheet, delete tab 5, then share the new file with the client.

Correct
Copying tabs 1-4 into a separate spreadsheet, then sharing the new file with the client will keep tab 5 private. In addition, making a copy of the spreadsheet, deleting tab 5, then sharing the new file with the client will keep tab 5 private.

Course Challenge

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

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
PDF File

And click below to access the datasets:

Course 3 Final Challenge Data Sets - Customer survey data (1)
CSV File

Course 3 Final Challenge Data Sets - Delivery times_distance (1)
CSV File

Reviewing the data enables you to describe how you will use it to achieve your client's goals. First, you notice that all of the data is first-party data, which means that it was collected from outside sources.

- ○ True
- ◉ False

Correct
First-party data is data collected by an individual or group using their own resources.

2. Scenario 1 continued

Next, you review the customer satisfaction survey data. To use the template for the customer satisfaction survey data, click the link below and select "Use Template."

Link to template: Customer Satisfaction Survey data

OR

If you don't have a Google account, download the CSV file directly from the attachment below.

CustomerSurveyData - Customer survey data
CSV File

The question in column E asks, "Was your order accurate? Please respond yes or no." The responses listed in column E are an example of Boolean data.

- ◉ True
- ○ False

Correct
A Boolean data type has only two possible values, such as yes or no.

3. **Scenario 1 continued**

Now, you review the data on delivery times and the distance of customers from the restaurant.

To use the template for the dataset, click the link below and select "Use Template."

Link to template: Delivery Times/Distance

OR

If you don't have a Google account, download the CSV file directly from the attachment below.

DeliveryTimes_DistanceData - Delivery times_distance
CSV File

Fill in the blank: The data in column E is an example of _____ data. Select all that apply.

- [x] discrete

Correct
This is an example of discrete data, which is counted and has a limited number of values. It is also quantitative data, which is specific and measures numerical facts.

- [ ] continuous
- [x] quantitative

Correct
This is an example of discrete data, which is counted and has a limited number of values. It is also quantitative data, which is specific and measures numerical facts.

- [ ] qualitative

4. **Scenario 1 continued**

The next thing you review is the file containing pictures of sandwich deliveries over a period of 30 days. **This is unstructured data, which means what?**

- ( ) It's objective and measures facts.
- ( ) It's organized in a certain format.
- (•) It's not organized in an easily identifiable manner.
- ( ) It's collected by a group directly from its audience and then sold.

Correct
Unstructured data is not organized in an easily identifiable manner.

5. **Scenario 1 continued**

Now that you're familiar with the data, you want to build trust with the team at Garden.

What data-security measures do you employ? Select all that apply.

- [ ] Make personal copies of client files
- [x] Assign user permissions for files

Correct
Building trust can be achieved by showing a client that you will keep client data safe by using passwords and user permissions.

- [ ] Change their file naming conventions
- [x] Add passwords to files

Correct
Building trust can be achieved by showing a client that you will keep client data safe by using passwords and user permissions.

6. **Scenario 2, questions 6-10**

You've completed this program and are interviewing for a junior data scientist position at a company called Sewati Financial Services.

Click below to review the job description:

C3 Course Challenge Junior Data Scientist Job Description .pdf
PDF File

So far, you've successfully completed the first interview with a recruiter. They arrange your second interview with the team at Sewati Financial Services.

Click below to read the email from the human resources director:

Course 3 Scenario 2_Second Interview Email.pdf
PDF File

You arrive 15 minutes early for your interview. Soon, you are escorted into a conference room, where you meet Kai Harvey, the senior manager of strategy. After welcoming you, he begins the behavioral interview.

Consider and respond to the following question. Select all that apply.

Our data analytics team often surveys clients to get their feedback. If you were on the team, how would you ensure the results do not favor a particular person, group of people, or thing?

- [x] Ensure the survey sample represents the population as a whole.

  Correct
  There are several factors that influence bias in survey results. To minimize bias, consider the way questions are written, the amount of time given to answer each question, and inclusivity of participants.

- [x] Make sure the wording of the survey question does not encourage a specific response from participants.

  Correct
  The way questions are written, the amount of time given to answer each question, and the inclusivity of the participants can help ensure survey results are unbiased.

- [x] Give participants enough time to answer each survey question.

  Correct
  The way questions are written, the amount of time given to answer each question, and the inclusivity of the participants can help ensure survey results are unbiased.

- [ ] Instruct participants to share their name and contact information.

7. **Scenario 2 continued**

   Consider and respond to the following question. Select all that apply.

   Our data analytics team often uses both internal and external data. Describe the difference between the two.

   - [x] Internal data is often generated from within the company. External data is generated outside the organization.

     Correct
     Internal data lives within a company's own systems and is typically generated from within the company. External data lives in and is generated outside the organization.

   - [ ] External data came from a company's own systems. Internal data came from the organization.
   - [x] Internal data came from a company's own systems. External data comes from outside the organization.

     Correct
     Internal data lives within a company's own systems and is typically generated from within the company. External data lives in and is generated outside the organization.

   - [ ] External data is often generated from within the company. Internal data is generated outside the organization.

8. **Scenario 2 continued**

Consider and respond to the following question. Select all that apply.

Our analysts often work with the same spreadsheet, but for different purposes. How would you use sorting to help in this situation?

- [ ] Sort data to highlight the header row.
- [x] Sort the data to arrange data in a meaningful order

Correct
Sorting data enables data analysts on the same team to use the same dataset for different purposes.

- [ ] Sort data to show only the data that meets a specific criteria while hiding the rest
- [x] Sort data to make it easier to understand, analyze and visualize

Correct
Sorting data enables data analysts on the same team to use the same dataset for different purposes.

9. **Scenario 2 continued**

Next, your interviewer wants to better understand your knowledge of basic SQL commands. **He asks: How would you write a query that retrieves only data about people with the last name Hassan from the Clients table in our database?**

- ○
```
SELECT Clients
WHERE last_name = 'Hassan'
```
- ○
```
SELECT *
WHERE last_name = 'Hassan'
```

```
SELECT *
WHERE Clients = 'Hassan'
```

```
SELECT *
FROM Clients
WHERE last_name = 'Hassan'
```

Correct
To write a query that retrieves only data about people with the last name Hassan from the Clients table, type SELECT * FROM Clients WHERE last_name='Hassan'.

10. Scenario 2 continued

For your final question, your interviewer explains that Sewati Financial Services cares about data privacy. The company needs its clients' trust, and this is an important responsibility for the data analytics team.

He asks: What does data privacy involve? Select all that apply.

- [x] Encryption and sharing permissions

This should not be selected
Review the section on privacy for a refresher.

- [x] A person's legal right to their data

Correct
Data privacy deals with preserving a data subject's information and activity any time a data transaction occurs, a person's legal right to their data, and establishing privacy measures to protect people's data.

- [x] Putting privacy measures in place to protect people's data

Correct
Data privacy deals with preserving a data subject's information and activity any time a data transaction occurs, a person's legal right to their data, and establishing privacy measures to protect people's data.

- [ ] Preserving a data subject's information and activity any time a data transaction occurs