

✔ **Congratulations! You passed!**

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

[Go to next item](#)

## Self-Reflection: Why pre-cleaning activities are important

Total points 1

1.



1 / 1 point

### Overview

---

Now that you have learned about how to prepare for data cleaning, you can pause for a moment and reflect on these steps. In this self-reflection, you will consider your thoughts about the importance of pre-cleaning activities 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 pre-cleaning activities and insufficient data to your own data cleaning work. 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 data integrity

---

Before data analysts can analyze data, they first need to think about and understand the data they're working with. Assessing data integrity is a key step in this process. As you've learned in previous lessons, you should complete the following tasks before analyzing data:

1. Determine data integrity by assessing the overall accuracy, consistency, and completeness of the data.
2. Connect objectives to data by understanding how your business objectives can be served by an investigation into the data.
3. Know when to stop collecting data.

Data analysts perform pre-cleaning activities to complete these steps. Pre-cleaning activities help you determine and maintain **data integrity**, which is essential to the role of a junior data analyst.

### What makes data insufficient

---

One of the objectives of pre-cleaning activities is to address insufficient data. Recall from previous lessons that data can be insufficient for a number of reasons. Insufficient data has one or more of the following problems:

- Comes from only one source
- Continuously updates and is incomplete
- Is outdated

- Is geographically limited

To deal with insufficient data, you can:

- Identify trends within the available data
- Wait for more data if time allows
- Discuss with stakeholders and adjust your objective
- Search for a new dataset

## Reflection

---

Consider what you have learned about data insufficiency and the steps for how to avoid it:

- Why are pre-cleaning steps important to complete prior to data cleaning?
- What problems might occur if you don't follow these steps?

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

Why are pre-cleaning steps important to complete prior to data cleaning?

Pre-cleaning activities help to address insufficient data that can lead to bias. Also, pre-cleaning flag outdated, geographically limited, Continuously updates, and is incomplete datasets

What problems might occur if you don't follow these steps?

The team may analyze a biased dataset, with untrusted data.

### Correct

Great work reinforcing your learning with a thoughtful self-reflection! A good reflection on this topic would include how pre-cleaning activities are important because they increase the efficiency and success of your data analysis tasks.

If you know that your data is accurate, consistent, and complete, you can be confident that your results will be valid. Stakeholders will be pleased if you connect the data to business objectives. And, knowing when to stop collecting data will allow you to finish your tasks in a timely manner without sacrificing data integrity.

Suppose that you didn't determine data integrity. You may find that you are working with inaccurate or missing data, which can cause misleading results in your analysis. If you don't connect objectives with the data, your analysis may not be relevant to the stakeholders. Finally, not understanding when to stop collecting data can lead to unnecessary delays in completing tasks. By completing pre-cleaning activities, you avoid these problems.