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

Question 1



### Data integrity



In this activity, you will consider what you've learned about how to prepare for data analysis and deal with insufficient data. Before you get started, first reflect on the steps to take before you begin analysis of your data. In particular, consider the following:

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

To address these steps, data analysts complete pre-cleaning activities. These activities relate to Data integrity, which is where you will spend most of your time as a junior data analyst.

### Insufficient data



As you've learned in previous lessons, insufficient data can come from a number of sources:

* Data from only one source
* Data that keeps updating
* Outdated data
* Geographically limited data

For these concerns, the solutions you've reviewed include:

* 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



Take a moment to reflect on what you’ve just reviewed: the steps to address data integrity, and the solutions for data insufficiency. Now, write a 3-5 sentence (50-100 word) response addressing why these steps are important to complete prior to data cleaning.

**1 / 1 point**

Because all the rest of the data analysis depend on accurate data if we don't have accurate data our analysis will be incomplete. So to get accurate data we need to perform these steps before using them in action.

**Correct**

Thank you for your response! These steps 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 have confidence that your results will be valid. The stakeholders will be pleased if you connect the data to business objectives. Finally, knowing when to stop collecting data will allow you to finish your tasks in a timely manner without sacrificing data integrity.

2.

Question 2

Can you think of any problems that might occur if you don't follow these steps? What are they? Write a 5 - 7 sentence response (100 - 150 words) reflecting these questions.

**1 / 1 point**

We may get inappropriate results or biased conclusions which is not good for data analysts.

**Correct**

Failing to follow these steps can produce problems when you try to explore and analyze data. Suppose that you didn't determine data integrity. Then you may find that you are working with inaccurate or missing data. This can cause misleading results in your analysis. If you neglect to 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.