4.3.1

Introduction to Preparing Data

Maria's next set of training activities focuses on preparing data—the second phase of our three-phase data analysis process. Before handing over the activities, she stresses that nothing is infallible—especially, data collection. Remember that flawed data leads to a flawed analysis!

Have you ever worked on a spreadsheet that's full of errors, such as typos, missing data, or duplicated data? **Dirty data** like this is the scourge of analysts—and it commonly appears. Preparing any dataset by cleaning it is essential before you analyze it. And, you can do so with the built-in functions that Pandas supplies.

In this lesson, you'll learn how to prepare data for analysis. You'll use tools and techniques to handle missing data, duplicated data, messy text data, and incorrect data types. With these skills, you'll be able to tackle even the dirtiest dataset.

CHALLENGE CONNECTION

In this week's Challenge, you'll use data preparation techniques on a dataset that contains both school budget data and student data.

Now that you've received an introduction to preparing data, you'll next learn how to handle missing data.