

---

[← Pre-Course Work](#)

16 / 23

## Step 2: Preparing Our Data (Cont.)

Data scientists frequently joke that they spend about 80 percent of their time cleaning data and only 20 percent actually creating predictive models. In other words, *data cleaning is a fact of life*.

We'll learn more about the many different ways you can clean data using Python in our later lessons, including methods to:

- Format column values.
- Differentiate between qualitative and quantitative data.
- Remove unnecessary features.
- Integrate multiple data sources.
- Address missing values.
- Choose a data sampling methodology.
- Sample your data.
- Create columns derived from our data (e.g., feature engineering).

