4.2.2

A Review of Your Data Analysis Tools

In this module, you'll use two main tools for your data analyses: Jupyter Notebook and Pandas. You've already gotten an overview of these tools. Now, here comes a refresher and an explanation of why they're ideal for data analyses.

Jupyter Notebook

Jupyter Notebook is an application that allows us to run small segments of Python code while we write. This is incredibly useful for checking our results to make sure our code is operating exactly how we expect it to and for exploring datasets a bit at a time.

All of the code we write in Jupyter Notebook could just as easily be written in a plain old `.py` file, but jumping back and forth between the editor and Terminal or Anaconda Prompt is a nuisance and showing charts and graphs is more difficult without the convenience of a web browser.

Pandas

Pandas is a Python library that's specifically designed for data analysis. Pandas offers a streamlined way of reviewing datasets and includes various functions that simplify importing, updating, and analyzing data.

Pandas is designed to work with multiple file formats, especially the CSV format, which you can import and read with just a few lines of code. Additionally, Pandas has various functions for reviewing and preparing data for analysis. And once the analysis begins, Pandas is optimized to perform quickly and efficiently.

Most importantly, Pandas has a human-friendly coding syntax that's straightforward enough for beginners to understand and use. Pandas is one of Python's most powerful tools, and understanding how to work within its framework will take your data wrangling skills to the next level.

Now that you've received a review of your data analysis tools, you'll next learn what data containers are.