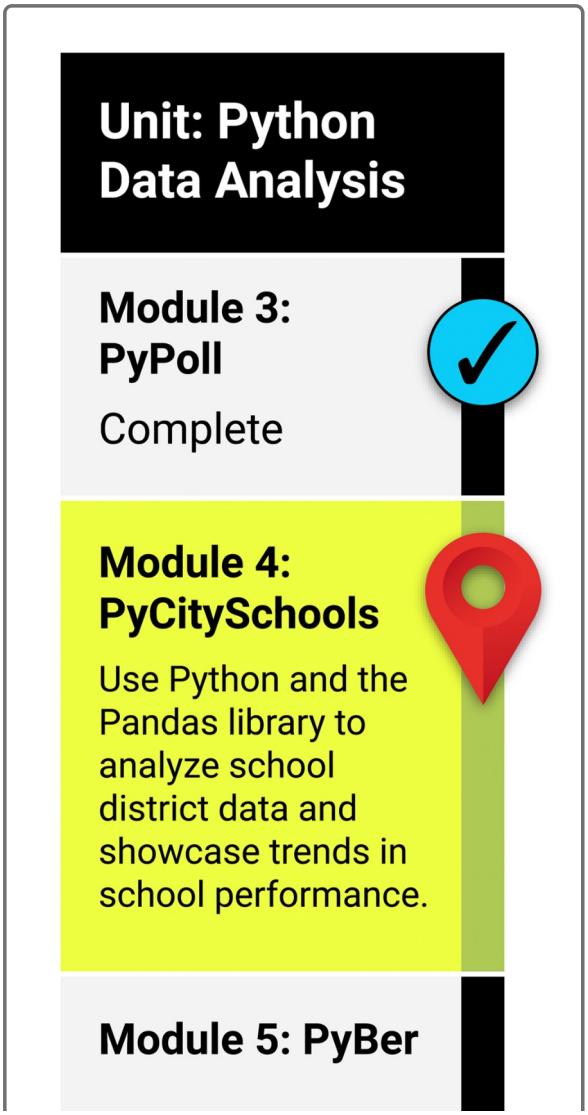
4.0.2 Module 4 Roadmap

Looking Ahead

This week you'll be introduced to the open-source distribution software called Anaconda and one of its products, Jupyter Notebook. This software allows you to create documents that contain live code using Python. Jupyter Notebook supports over 40 different programming languages, but in this module, you'll use Python as you learn the basics of the Pandas library. Pandas is an open-source library that provides high-performance data analysis tools. Using Jupyter Notebook and the Pandas library, you'll read raw data from CSV files, inspect and clean data, merge datasets, perform mathematical calculations, and visualize the data with charts and



graphs to tell a story.

What You Will Learn

By the end of this module, you will be able to:

- Open Jupyter Notebook files from local directories using a development environment.
- Read an external CSV file into a DataFrame.
- Format a DataFrame column.
- Determine data types of row values in a DataFrame.
- Retrieve data from specific columns of a DataFrame.
- Merge, filter, slice, and sort a DataFrame.
- Apply the `groupby()` function to a DataFrame.
- Use multiple methods to perform a function on a DataFrame.
- Perform mathematical calculations on columns of a DataFrame or Series.

Planning Your Schedule

Here's a quick look at the lessons and assignments you'll cover in this module. You can use the time estimates to help pace your learning and plan your schedule.

- Introduction to Module 4 (30 minutes)
- Anaconda Installation and Jupyter Notebook (1 hour)
- Creating and Activating A Development Environment (30 minutes)
- Working with Jupyter Notebook and Pandas (1 hour 30 minutes)
- Convert CSV Files to a Pandas DataFrame (1 hour)

- Exploring the Data (1 hour 30 minutes)
- Verify the Clean Student Data (15 minutes)
- Generate the School District Summary (1 hour 30 minutes)
- Generate the School Summary (1 hour 30 minutes)
- High and Low Performing Schools (1 hour)
- Average Math and Reading Scores by Grade (1 hour)
- Group Scores by School Spending per Student(1 hour)
- Group Scores by School Size (1 hour)
- Group Scores by School Type (30 minutes)
- Application (5 hours)