

# Module 5 Summary: APIs and Data Collection

Congratulations! You have completed this module. At this point, you know that:

- Simple APIs in Python are application programming interfaces that provide straightforward and easy-to-use methods for interacting with services, libraries, or data, often with minimal configuration or complexity.
  - An API lets two pieces of software talk to each other.
  - Using an API library in Python entails importing the library, calling its functions or methods to make HTTP requests, and parsing the responses to access data or services provided by the API.
  - Pandas API processes the data by communicating with the other software components.
  - An Instance forms when you create a dictionary and then use the DataFrames constructor to create a Pandas object.
  - Method “head()” will display the mentioned number of rows from the top (default 5) of DataFrames, while method “mean()” will calculate the mean and return the values
- Rest APIs allow you to communicate through the internet, taking advantage of resources like storage, access more data, AI algorithms, and so on.
  - HTTP methods transmit data over the internet.
  - An HTTP message typically includes a JSON file with instructions for operations.
  - HTTP messages containing JSON files are returned to the client as a response from web services.
  - Dealing with time series data involves using the Pandas time series function.
  - You can get data for daily candlesticks and plot the chart using Plotly with the candlestick plot.
- The HTTP (HyperText Transfer Protocol) transfers data, including web pages and resources, between a client (a web browser) and a server on the World Wide Web.
  - The HTTP protocol is commonly used for implementing various types of REST APIs.
  - An HTTP response includes information like the type of resource, length of resource, and so on
  - Uniform resource locator (URL) is the most popular way to find resources on the web.
  - URL is divided into three parts: scheme, internet address or base URL, and route
  - The GET method is one of the popular methods of requesting information. Some other methods may also include the body.
  - Response method contains the version and body of the response.
  - POST submits data to the server, PUT updates data already on the server, DELETE deletes data from the server
- Requests is a Python library that allows you to send HTTP/1.1 requests easily
  - You can modify the results of your query with the GET method.
  - You can obtain multiple requests from a URL like name, ID, and so on with a Query string.
- Web scraping in Python involves extracting and parsing data from websites to gather information for various applications, using libraries like BeautifulSoup and requests.
  - HTML comprises text surrounded by blue text elements enclosed in angular brackets called tags.
  - You can select an HTML element on a web page to inspect the webpage.
  - Web pages may also contain CSS and JavaScript along with HTML elements.
  - Each HTML document is like an HTML Tree, which may contain strings and other tags.
  - Each HTML table is comprised of table tags and is structured with elements such as rows, headers, body and so on.
- Tabular data can also be extracted from web pages using the `read\_html` method in Pandas.
- BeautifulSoup in Python is a library for parsing and navigating HTML and XML documents, making extracting, and manipulating data from web pages more accessible.
- To parse a document, pass it through the BeautifulSoup constructor to get a beautiful soup object representing the document as a nested data structure.
- BeautifulSoup represents HTML as a set of tree-like objects with methods to parse the HTML.
- Navigable string is like a Python string that supports BeautifulSoup functionality.
- find\_all is a method used to extract content based on the tag’s name, its attributes, the text of a string, or some combination of these.
- The find\_all method looks through a tag’s descendants and retrieves all descendants that match your filters.
- The result is a Python iterable like a list.
- File formats refer to the specific structure and encoding rules used to store and represent data in files, such as .txt for plain text or .csv for comma-separated values.
- Python works with different file formats such as CSV, XML, JSON, xlsx, and so on
- The extension of a file name will let you know what type of file it is and what it needs to open with.
- To access data from CSV files, we can use Python libraries such as Pandas.
- Similarly, different methods help parse JSON, XML, and other files.

Go to next item

✓ Completed