

Working with HTML, XML, JSON and Web APIs

Part I (18 points): Working with HTML, XML, and JSON

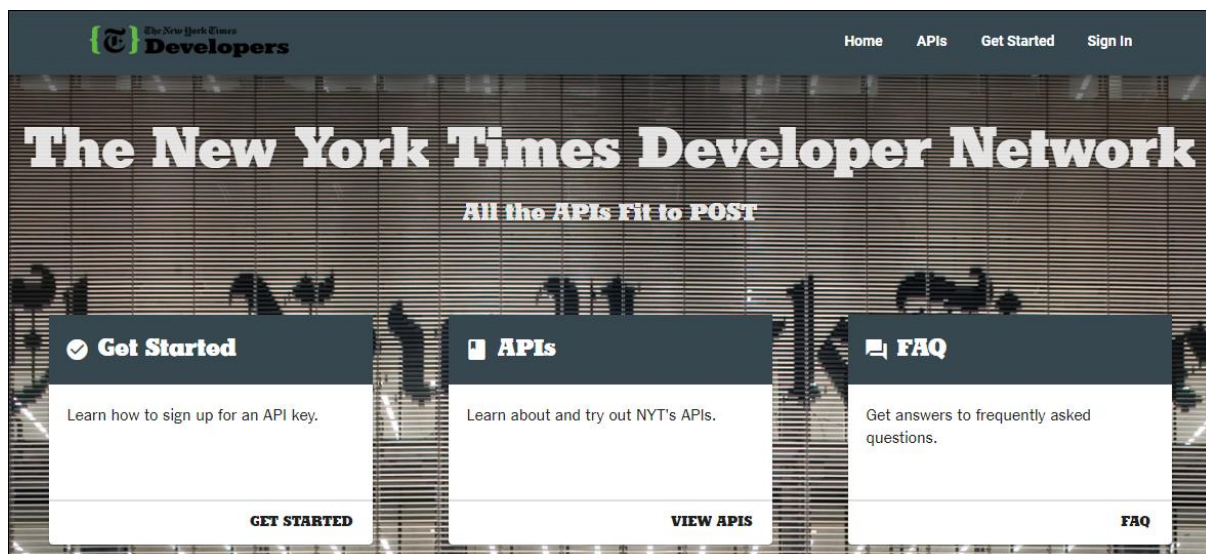
Pick three of your favorite books on one of your favorite subjects. At least one of the books should have more than one author. For each book, include the title, authors, and two or three other attributes that you find interesting. Take the information that you've selected about these three books, and separately create three files which store the book's information in HTML (using an html table), XML, and JSON formats (e.g. "books.html", "books.xml", and "books.json"). To help you better understand the different file structures, I'd prefer that you create each of these files "by hand" unless you're already very comfortable with the file formats. Write Python code, using your packages of choice, to load the information from each of the three sources into separate PANDAS data frames. Are the three data frames identical? Your deliverable is the three source files and the Python code. Post the three source files to GitHub and package your Python code within a Jupyter notebook (along with your code for Part II below) and post it to GitHub as well.

Part II (12 points): Working with Web API's

The New York Times web site provides a rich set of APIs, as described here: <https://developer.nytimes.com>

You'll need to start by signing up for an API key.

Your task is to then choose one of the New York Times APIs and construct an interface in Python to read JSON data accessible via the API and transform that data into a Pandas data frame that is suitable for use in data analysis work.



Save all of your work for this assignment within **a single Jupyter Notebook** and upload it to your online DAV5400 GitHub directory. Be sure to save your Notebook using the following nomenclature : **first initial_last name_w12_assn**" (e.g., J_Smith_w12_assn).

As a reminder, this assignment is due no later than 11.59pm on Dec 1.