

Question 1

You have been provided with information about a certain imaginary party, John Doe, contained within [several different file types and data/presentation formats](#).

For EACH of the following file types, you are to create one python 2.7 script that reads in the corresponding data file and returns all the metadata as a single level python dictionary (no nested dictionaries)

- a. **JSON** (john.json)
- b. **Plain Text** (john.txt)
- c. **HTML** (john.html)
- d. **PDF** (john.pdf)
- e. **DOCX** (john.docx) - BONUS, not mandatory to attempt

Important:

- 1. Parse the dates into a datetime object
- 2. Age should be an integer
- 3. Use pypdf to read the PDF file, beautiful soup to parse HTML & any package of choice for the word document.
- 4. Submission for this question should be 4 (or 5 python scripts if you tackle bonus question) that on execution will ingest the relevant input file, and prints the user metadata as a python dictionary.
- 5. Download the zip folder and decompress it to access the individual files inside it.

Question 2

This [html file](#) contains a list of 20 cases with some of their related metadata. Parse it in beautiful soup and capture, for each case, all the relevant labeled metadata as well as the categories, title and summary according to [this pictographic guide](#).

Important:

- a. Download the html file locally to read it, no need to download it in code.
- b. Submission for this question is a single python file that prints to the terminal an array of 20 dictionaries, each encapsulating metadata from an individual case.