**Gentrack Technical Test:**

**Problem Statement**

We need to process data in CSV format, however, the source files are delivered within XML. To handle this, create a set of functions that will take the XML input and generate the required CSVs:

* Read in an XML file
* Extract out the data found in the CSVIntervalData element with the following conditions,
  + Create a separate CSV file for each block of data within the CSVIntervalData element that starts with "200"
  + A single block of data is the "200" row of data, followed by the repeating rows after, until the next "200", or you hit the "900" trailing line
  + Each CSV file should have the "100" row as a header, and the "900" row as the trailer
  + Each CSV file should be named from the second field in the "200" row
  + Valid rows within the CSVIntervalData element can only start with "100", "200", "300","900"
  + The CSVIntervalData element should contain at least 1 row for each of "100", "200", "300","900"
  + "100", "900" rows should only appear once inside the CSVIntervalData element
  + "200" and "300" can repeat and will be within the header and trailer rows
  + "200" row must be followed by at least 1 "300" row
  + Remove leading and trailing white spaces, tabs, and additional newlines.

Input

The attached ZIP contains an XML: testfile.xml

The input file will always contain:

* A Header element
* A Transactions element, containing 1 Transaction element
* A Transaction element will contain transactionDate and transactionID attributes
* Required data will be in Transactions->Transaction->MeterDataNotification->CSVIntervalData

**Function Requirements break down**

* Open an xml file
* Read in the CSVIntervalData
* Break up the CSVIntervalData
* Format the CSVIntervalData for writing to csv file
* Clean up data RE leading/trailing tabs, spaces, newlines
* Make the CSV file header
* Write the CSV file

**Initial thoughts/ideas:**

Use Python programming language

* Familiar
* Excellent support and community

Upload final work to Github

Make it a single script

Clarify file location to be implemented?

Look at how to read in XML file as starting point - convert CSV data to lists.

**Initial Research:**

How to load and read an xml file: <https://www.youtube.com/watch?v=GxY0VJXACsk>

Raises the question: what is parsing and why does ElementTree do it?

What is parsing: <https://www.youtube.com/watch?v=T0BO415l3N0>

How to write to a csv file: <https://www.youtube.com/watch?v=s1XiCh-mGCA> AND <https://www.pythontutorial.net/python-basics/python-write-csv-file/>

For cleaning up non-data/non-essential string elements: <https://stackoverflow.com/questions/15558392/how-can-i-check-if-character-in-a-string-is-a-letter-python>

For conversion from string to list for mapping nodes:

<https://blog.finxter.com/how-to-convert-a-string-to-a-list-in-python/>