

Python Interview Project

Brief:

This technical challenge requires the creation of a file loading and processing python job. There is no restrictions on what libraries or version of Python should be used. Instead, the candidate is free to use those they feel best enables them to perform the task at hand.

For each file that is to be processed, there is also a corresponding 'request' instructions that inform what steps need to be performed on the data. For the sake of this demo, the candidate will need to create 3 transformation steps, but not all three will need to be applied to each data set.

Objectives:

- 1.) Load the specified file into a data frame
- 2.) Execute the required rules against the specified data frame
- 3.) Save the required output as a json formatted file

Besides the 3 primary objectives above, the candidate should also factor in adequate error handling of code, loading and transforming of files. Logging should also be provided, giving an audit trail of each file that is processed and how the data is modified along the way.

Transformation Rules:

Rule one – Age Band Group:

When provided with an age, date of birth or year of birth, convert the value to ten year age bands.

0 – 10, 11-20, 21-30, etc. to 90+

Rule Two – Post Code Trim – Precision 3

For all records, trim the postcode to the first 3 characters

Rule Three – Post Code Trim – Precision 2

This rule should only be applied after Rule Two. Group the records by the trimmed postcode and where the number of records per postcode is under 10, apply further trimming of the post code to only the first 2 characters

Logging Requirements:

A log should be created for each file that is being processed and log the following attributes:

- Process request id
- File to process
- Date and time processing started
- Rules to process
- Each rule processed and the time to complete
- Job completion and status (success or fail)

Request 1:

Request ID: 123

File to Process: customers.csv

Rules to Apply: [Rule 1]

Request 2:

Request ID: 7282

File to process: PatientCohorts.json

Rules to Apply: ["Rule 1", "Rule 2"]

Request 3:

Request ID: 92421

File to process: ResearchList.xls

Rules to Apply: ["Rule 2", "Rule 3"]