Data Cleaning & Transformation

Exercise

Worthy of Note

The purpose of this task is to test your thinking and technical skills.

Apply the approach you clearly understand.

Provide your solution using any Data Exploration/Processing tools (already used in class) with which you are most familiar.

Take a step further to try a different approach/tool to completing the task

Exercise Brief

The Data Engineers of a particular company often keep logs of the details of previously executed pipelines triggered by their orchestration tool (or sometimes, manually).

A new Engineering Manager has been hired and will need to know and understand the functions and processes carried out as documented in the log file.

You are to help this individual clean and transform the data found in the log file (with any methods or tools of your choice), and perform some analyses you think will be helpful to the Manager.

Files Provided:

log.txt - contains all of the tasks logged by the Data Engineering team

columns.txt - the corresponding headers of the dataset existing in the log file (in the same order) including some other necessary details

Instructions

1. Take your time to read and understand dataset.
2. Clean & Transform the dataset provided with any data cleaning/processing tool of your choice. They have to be tools already used in class.
3. Think about and execute analysis that you know could be helpful to any one (including you) trying to understand the dataset. If you were the engineering manager, what analytical questions would you ask?
4. Store and Manage your brilliant solution in GitHub. Submit the link to your repository.

HAPPY THINKING