

TRIDGE ENGINEERING

The following assignment has a goal to evaluate your ability to analyze a dataset, find the value in the dataset and find an appropriate way to extract this value from it.

There is a preference of Python for the language used in its resolution, but you are free to submit your assignment in any language of your choice. The ability to justify another language chosen is also valuable (technical limitations, personal proficiency, language design fit).

The first part of this assignment will ask you to understand the dataset and extract certain information.

The second part of this assignment will ask you to suggest improvements or additional value you could extract from it.

A pseudo-implementation can be accepted in case of implementations requiring external processing like machine learning.

You can find the dataset attached with the assignment communication.

Allotted Time:

The schedule and allotted time to submit this assignment should be given in the communication (email, phone) related to this assignment. If you did not receive them please confirm them with the contact point for your application.

Submission:

The following format are accepted:

- Public repository (Github, etc..)
- Zipped folder

Your submission should explain any steps required to start your program.

Dataset:

The attached dataset represents a list of prices across three years for certain products. Those products are identified by their specifications such as product name, variety, grade, label, region.

Part 1:

Please implement the following logic:

A) Extract the Weekly/Monthly Trends (% change) for each Row for the entire duration of the dataset range

B) There are multiple aggregation levels possible for this dataset. One example could be aggregating the price and trends at the Product-Region level. Please implement a different aggregation level for this dataset.

You will also note that some rows do not have data on specific dates, how to handle this situation is also part of the exercise.

Expected output:

- A: CSV, the column/row structure is up to your judgement
- B: CSV, the column/row structure is up to your judgement

Part 2:

This part is an open question. You are free to make any suggestions and do not require implementation. A non-code answer is acceptable.

- What additional information, not asked in the previous exercise, do you think can be extracted from this dataset ?
- How would you collect or extrapolate such information ? Please describe a vague algorithm/implementation you would use.
- Assume this same dataset structure now contains 200,000,000 rows, would your logic scale ? If not, how would you make it scale? If yes, how does it handle the scale?

If you need any clarifications for this assignment regarding the exercises, submission or allotted time, please communicate them with your Tridge contact point for your application.

Good luck :)