**Hearful Python Developer Assignment**

Dear Candidate,

The purpose of this exercise is to test your Python knowledge and programming skills. This exercise will serve as an example of the type of work that will be done in the Python Developer role. Python is used almost entirely on Hearful’s backend. Depending on your skill level, there may be different ways to complete the assignment.

You will see a few real examples of the data Hearful works with on a daily basis. The test is meant to be broad with the ability to be approached in many different ways. It will test different skills in Python like data transformation, file I/O, use of the standard library, accessing data structures, and more.

Be as creative but efficient as possible when completing this assignment. Other than the requirements listed below, you have creative freedom to design and write your code. Create any project/directory structure you want. Feel free to showcase your knowledge and creativity. However, be able to explain why you made the decisions you did when implementing code a certain way, designing a process, using a certain Python module, etc.

Please let me know if you have any questions as you work through the assignment.

Kind regards,

Justin

**Task 1**: Create a data processing script/pipeline to transform data

*Instructions:*

Transform the data in raw\_reviews.json into the data seen clean\_reviews.json

There are two attached files **raw\_reviews.json** and **clean\_reviews.json**. Your task will be to create an ETL or data processing pipeline to transform the data in **raw\_reviews.json** into the data seen in **clean\_reviews.json**. The file **clean\_reviews.json** will serve only as an example to what the end data should look like. You will only need to access/ingest **raw\_reviews.json** into your code.

While you may not get the data to look exactly like the clean reviews data, get as close as possible to the output. If you get stuck, complete as much of the task as possible. Be prepared to discuss any difficulties you could not tackle.

Pay close attention to new fields added in the clean versus raw data. Data types may change. Certain fields may be removed. Other fields have been changed or cleaned. You are free to create external files if necessary.

Task 1 Requirements:

* Use Python 3.6+
* Create a .py file that can be run on the command line
* Directly ingest **raw\_reviews.json** into your script or a settings file
* Use OOP (Object Oriented Programming) somewhere in your code
* Output a JSON file
* Upload to a repo like Gitlab

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Task 2**: Convert JSON to CSV

*Instructions:*

Take your output from Task 1 and convert it to a flat file or CSV. The headers should be the fields from each data entry. There should be 5 rows in total. The file should contain 4 rows of data plus 1 row serving as the header. There will be some data entries that have fields which are not present in others. In this case you may leave the cell/area blank or fill in with a null value.

For Task 2, try to use the Python standard library, and modules like **csv** and **json**. While nice and easy to use, try to avoid modules like **pandas** unless you have no other option.

Task 2 Requirements:

* Use Python 3.6+
* Create a .py file that can be run on the command line
* Directly ingest **the output from task 1** into your script or a settings file
* Output a CSV file
* Upload to a repo like Gitlab