

# Amazon Reviews Ingestion running instructions

1. Start by cloning from GitHub:

```
git clone https://github.com/luca-mircea/amazon-reviews-ingestion
```

2. Copy the contents of the credentials file shared into a text file, save it in `src/credentials` and name it `.env`

3. To run the container without the S3 interaction (i.e.) locally:

- a. Save the clothing, shoes, and jewelry reviews dataset as `'reviews.csv'` under `data_short`; do the same for the metadata, saving it in the same location and naming it `'metadata.csv'`
- b. `docker build -t takeaway-challenge .`
- c. `docker run takeaway-challenge python entrypoint.py --task_name 'process_raw_reviews_data_without_timestamps_locally'`
- d. `docker run takeaway-challenge python entrypoint.py --task_name 'process_raw_metadata_without_timestamps_locally'`
- e. `docker run takeaway-challenge python entrypoint.py --task_name 'check_successful_completion_s3_locally'`

4. To run the container the container with the S3 interaction:

- a. `docker build -t takeaway-challenge .`
- b. `docker run takeaway-challenge python entrypoint.py --task_name 'process_raw_reviews_data_without_timestamps'`
- c. `docker run takeaway-challenge python entrypoint.py --task_name 'process_raw_metadata_without_timestamps'`
- d. `docker run takeaway-challenge python entrypoint.py --task_name 'check_successful_completion_s3'`