Amazon Reviews Ingestion running instructions

1. Start by cloning from GitHub:

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

1. Copy the contents of the credentials file shared into a text file, save it in src/credentials and name it .env
2. To run the container without the S3 interaction (i.e.) locally:
   1. 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’
   2. docker build -t takeaway-challenge .
   3. docker run takeaway-challenge python entrypoint.py --task\_name ‘process\_raw\_reviews\_data\_without\_timestamps\_locally’
   4. docker run takeaway-challenge python entrypoint.py --task\_name ‘process\_raw\_metadata\_without\_timestamps\_locally’
   5. docker run takeaway-challenge python entrypoint.py --task\_name ‘check\_successful\_completion\_s3\_locally’
3. To run the container the container with the S3 interaction:
   1. docker build -t takeaway-challenge .
   2. docker run takeaway-challenge python entrypoint.py --task\_name ‘process\_raw\_reviews\_data\_without\_timestamps’
   3. docker run takeaway-challenge python entrypoint.py --task\_name ‘process\_raw\_metadata\_without\_timestamps’
   4. docker run takeaway-challenge python entrypoint.py --task\_name ‘check\_successful\_completion\_s3’