Teke the 8 movie files and turn them into silver tables(delta files)

A few requirements and things to consider:

* In the end, you should have your bronze and silver tables in Delta format all persisted to your DBFS.
* You should have at least 3 silver tables, add more if you feel it’s needed.
  + Movie
  + Genres (lookup table)
  + OriginalLanguages (lookup table)
* Harden your logic into functions and your end result should be .py files.
  + For example, to run the bronze to silver transformation, you should only need to run the bronze\_to\_silver() function with the correct parameters.
* Some movies have a negative runtime
  + First gather all such records, put them in “quarantine” area
    - Mark the record as “quarantined” in your bronze table
    - Separate these records from the “clean” records, and work on fixing them after you have loaded all the “clean” records from the 8 movie files into silver tables.
  + To fix the runtime, simply take the absolute value of the original negative value.
* Certain movies have valid ID for the genre, but the name of the genre is missing

Yes, we need, and it can be easily fixed by simply calling the dropDuplicates([“Genre\_Id”]) when we are exploding the Genre column.

* Let’s assume all the movies should have a minimum budget of 1 million

I think we should fix this when we are in the processing of “bronze to silver,” so that we can easily fix that column when we are generating the dataframe before we write it.

* We have some movies that are showing up in more than one movie files.
  + How do we ensure only one record shows up in our silver table?
    - Recall in the moovio notebooks we had a column called “status” in our bronze table to specify “new” and “loaded”

We have to use dropDuplicates() function to eliminate the repeated data when we generate the dataframe, and also we need to change the status after we uploaded the “new” status to “loaded”, also we need to change that status after we repaired the quarantined data. By doing this we can ensure that one movie has only one record.

* Commit your source code to Github.