ALGORITHM COURSEWORK 2023 TASK 2A

- Requirement:

Write a Python program **'ticketit_prep.py'** that reads the data files and creates **one flat file, 'ticketit_prep.dat' with following columns:**

id	cust_id	bkg_date	cust_seg	film_id	film_rating	genre
(from bookings)			(from customer)	(from film)		(from film film genre)

Note that this will generate more than one line of data for many bookings, as a single film can relate to many genres.

- Observation:

This requirement is based on the Bookings file and pulls data from other linked files. It obviously pulls data from the Customer file. It also uses the Film file and the Film_Film_Genres file, but it can only access these via the Film_Showing file, which provides a link between Booking and Film.

- Method Overview

We must restrict file reads to a minimum, for the sake of efficiency.

We will read the Customer, Film_Showing, Film and Film_Film_Genres files once only and create Python dictionaries with the required data only. The dictionaries will later be used as 'lookup tables' to build our line of data to be written to the output file. (Dictionaries provide direct access, via a key to the required data, whereas a list might have to be accessed via sequential reads.)

- Method Details

(Note: the steps below can be coded and tested sequentially).

1. Read customer.csv and build a dictionary with key being customer_id and value being SEG.
2. Read films.csv and build the films dictionary. This is a little complex. The key of the films dictionary is film_id. The 'value' held against the key is itself a dictionary, holding the rating value against a key of "rating" and a list of genres values under a key of "genres". This means that you will have a some code like below as you update the film rating into the film dictionary and initialise the list of genres..

   ```
   lst_genres = []
   fmdict = {"rating" : film_rating, "genres": lst_genres}
   film_dict[str(film_id)]=fmdict
   ```

3. Read the film_film_genres.csv and update the films dictionary as appropriate.
4. Read the film_showings.csv and build showings dictionary with showing_id as the key. You have a choice as to how we record the necessary link to film in the value part of the dictionary but one neat idea is to copy across the ENTIRE value part of the film dictionary entry for that film_id. This means that we won't have to reference the films dictionary again.

   ```
   fsdict = {}
   lst_genres = []
   flm_rating = ''
   ```

```
if film_id in film_dict:
    lst_genres = film_dict[film_id]["genres"]
    flm_rating = film_dict[film_id]["rating"]
fsdict = {"film" : film_id, "rating": flm_rating, "genres": lst_genres}
showing_dict[str(showing_id)]= fsdict
```

5. Open the output file for write
6. Read the booking.csv file.

   For each line

   we have booking_id, customerid and booking date.

   Get the customer SEG from the customer dictionary

   Get the film_id and film rating from the showing dictionary.

   For each film_film_genre in the list

   Write a line to the output file.

7. Close booking.csv
8. Close output file.