

Ξεκινάμε διασχίζοντας τα αρχεία csv και διαγράφουμε τα διπλότυπα. Έπειτα διατρέχουμε το `movies_metadata` και διορθώνουμε το πεδίο `imdb_id` όπου έχει δύο επιπλέον γράμματα (μία καταγραφή για παράδειγμα ήταν `tt1238` και μετά την διόρθωση έγινε `1238` ώστε να μπορεί να γίνει αργότερα η σύνδεση με τους άλλους πίνακες ).

Μετά δημιουργούμε set για να αποθηκεύουμε τις ταινίες όπου υπάρχουν στο πίνακα `movies_metadata` ώστε να αφαιρέσουμε από τους υπόλοιπους πίνακες τις ταινίες που δεν υπάρχουν στον `movies_metadata`. Έπειτα δημιουργούμε ένα set όπου περιέχει τα `id` του `credits` και ένα set που περιέχει τα `imdb_id` από τον `Links`. Έπειτα διατρέχουμε τα αρχικά set που φτιάξαμε από τον `movies_metadata` και σβήνουμε τα κοινά στοιχεία με τα set που φτιάξαμε από τους άλλους πίνακες με αποτέλεσμα να μένουν τα `id` όπου πρέπει να διαγραφούν.

Στη συνέχεια, ξαναδιατρέχουμε τα αρχεία, και όσες γραμμές έχουν `id` που βρίσκονται στο set δεν τα προσθέτουμε στο νέο αρχείο. Επιπλέον στο `links.csv` δημιουργούμε άλλο ένα set, στο οποίο κρατάμε τα `movieid`, τα οποία θα μας χρειαστούν στο `ratings.csv` και στο `ratings_small.csv`.

Τέλος, σβήνουμε τα προσωρινά αρχεία που δημιουργήσαμε και μετονομάζουμε τα τελικά αρχεία με τα αρχικά τους ονόματα.