

Τα δεδομένα περάσαν απο 3 σταδια επεξεργασιας.Πρωτα περασαν απο ενα σκριπτακι σε python όπου με χρηση του module pandas ανιχνευθηκαν τα διπλοτυπα με χρηση για την αναζητηση ενος column του καθε πινακα(όπου μελλοντικα εγινε και primary key) εκτός από τους πίνακες ratings και ratings summary.Σε δευτερο στάδιο τα δεδομένα(εκτος του Movies_metadata) πέρασαν από το δεύτερο σκριπτάκι της python ξανά με το module pandas , όπου αφαιρέθηκαν οι εγγραφές στις ταινίες στις οποίες ανφέρονταν , δεν βρίσκονταν στον πίνακα Movies_metadata.Τέλος όταν πια τα δεδομένα ανέβηκαν στη βάση , τα στοιχεία του πίνακα ratings_small ενσωματώθηκαν στον πίνακα ratings και η στήλη genres του πίνακα Movies_metadata μετά από μια μικρή επεξεργασία όπου τα “ \’ “ μετατράπηκαν με χρήση της συνάρτησης regepr_replace σε “ \ “ ώστε να μπορέσει να γίνει η στήλη cast σε τύπο json , μεταφέρθηκαν σε έναν ξεχωριστό πίνακα με όνομα Genres.