

**ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ**

**ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2022-2023**

**ΕΞΑΜΗΝΟ: Δ**

**ΕΡΓΑΣΙΑ 2**

**ΔΙΔΑΣΚΩΝΤΕΣ: ΒΑΣΙΛΗΣ ΒΑΣΣΑΛΟΣ-ΘΕΟΦΙΛΟΣ ΜΑΙΛΗΣ**

**ΟΝΟΜΑΤΕΠΩΝΥΜΟ:** ΑΙΚΑΤΕΡΙΝΗ ΚΟΡΡΕ

**ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ:** 3210084

**ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΕΥΘΥΝΣΗ:** [p3210084@aueb.gr](mailto:p3210084@aueb.gr)

**ΟΝΟΜΑΤΕΠΩΝΥΜΟ:** ΣΤΕΦΑΝΟΣ ΦΡΑΓΚΟΥΛΗΣ

**ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ:** 3210212

**ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΕΥΘΥΝΣΗ:** [p3210212@aueb.gr](mailto:p3210212@aueb.gr)

**Συνοπτική Αναφορά Εργασίας**

Σκοπός της εργασίας ήταν να δημιουργήσουμε την βάση δεδομένων ταινιών Movielens. Αρχικά, κατεβάσαμε τα δεδομένα της άσκησης από τον εξωτερικό σύνδεσμο που δινόταν. Ύστερα, επεξεργαστήκαμε το αρχείο keywords.csv που περιείχε εμφωλευμένη πληροφορία σε μορφή JSon συμβολοσειρών, δηλαδή το χωρίσαμε σε δύο csv αρχεία το Keyword.csv με πεδία το id και name και το hasKeyword.csv με πεδία το movie\_id και το keyword\_id. Η επεξεργασία του αρχείου και η επιπεδοποίησή του έγινε με κώδικα python (αρχείο my\_code.py), ο οποίος διαβάζει τα δεδομένα του αρχείου keywords ώστε να τα χωρίσει σε δύο πίνακες και να δημιουργήσει τα δύο αρχεία keyword και hasKeyword. Φυσικά, με την βοήθεια της συνάρτησης set της python αφαιρέσαμε τα διπλότυπα από το πίνακα hasKeyword. Στη συνέχεια, περάσαμε τα δεδομένα στην βάση μας με το Import wizard και φτιάξαμε τους τύπους των πεδίων κατάλληλα για να δέχονται όλα τα δεδομένα. Μετά, ορίσαμε με την βοήθεια της γλώσσας sql πρωτεύοντα κλειδιά για τους πίνακες Movie, Genre, Productioncompany, Collection, Movie\_cast, Movie\_crew, keyword και ξένα κλειδιά για τους πίνακες belongsTocollection, hasGenre, HasProductioncompany, ratings, movie\_cast, movie\_crew και hasKeyword.