

PROJECT ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

Ονοματεπώνυμο:

Ιωάννης Ξηντάρας 3150131

Δέσποινα Παπαδοπούλου 3180146

Νικόλαος Στραγγάλης 3200189

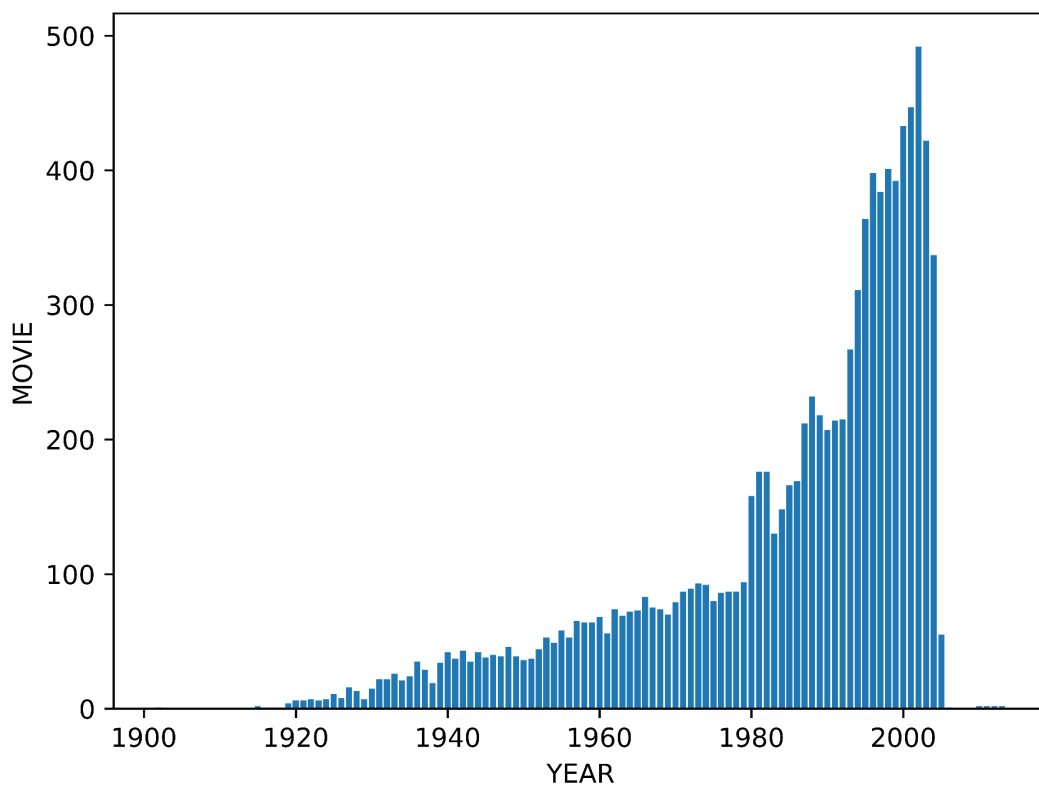
Σε όλα τα σχεδιαγράμματα ακολουθήσαμε την εξής πορεία ενεργειών:

1. Καναμέ execute τον sql κώδικα του κάθε ερωτήματος
2. Αποθηκεύουμε το fetch του σε μια μεταβλητή rows
3. Δημιουργήσαμε αντίστοιχο πλήθος πινάκων, 2 αν το σχεδιάγραμμα είναι δισδιάστατο ενώ 3 αν είναι τρισδιάστατο, ώστε να αποθηκεύσουμε την έξοδο των αποτελεσμάτων της μεταβλητής rows.
4. αποθηκεύω σε μορφή numpy array και το βάζω ως παράμετρο στο αντίστοιχο πρόγραμμα αναπαράστασης.

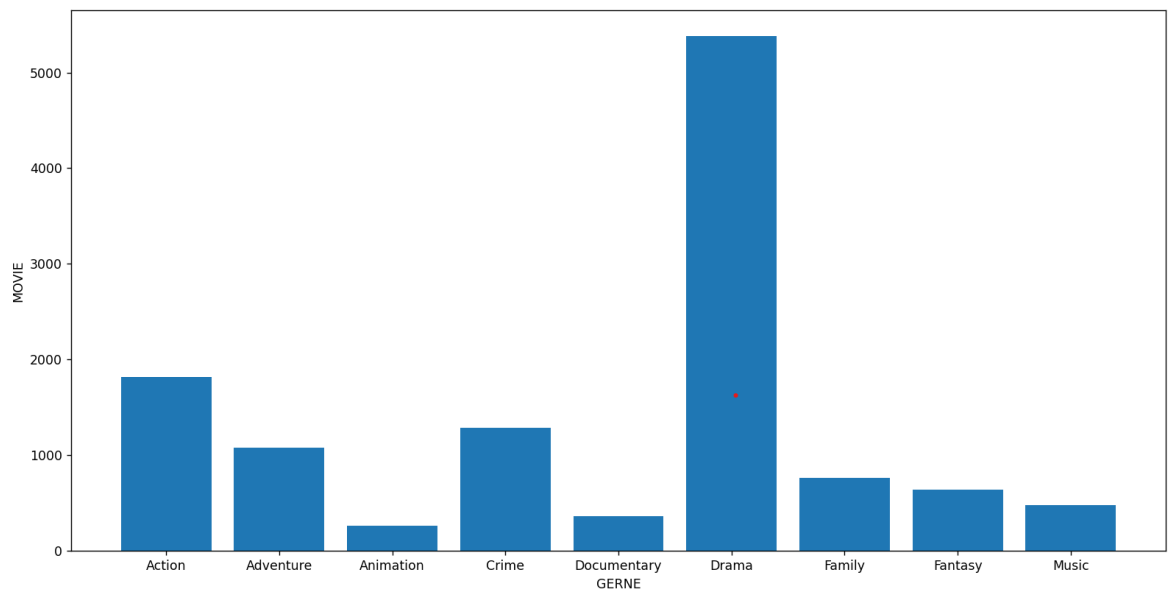
Θα ακολουθήσουν τα αντίστοιχα διαγράμματα για κάθε ερώτημα και σχόλια όπου θεωρούμε ότι απαιτείται.

Στατιστικά ταινιών

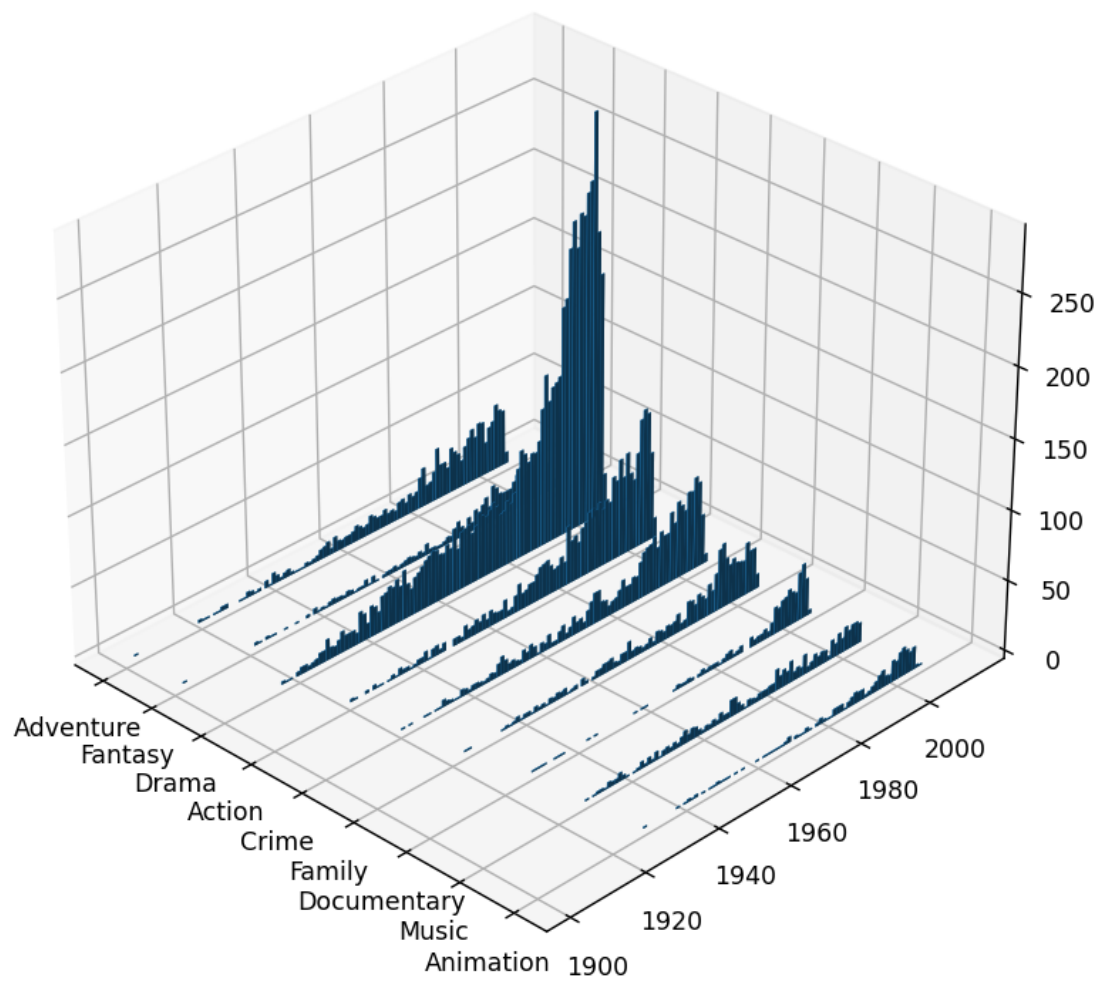
1. Αριθμός ταινιών ανά έτος



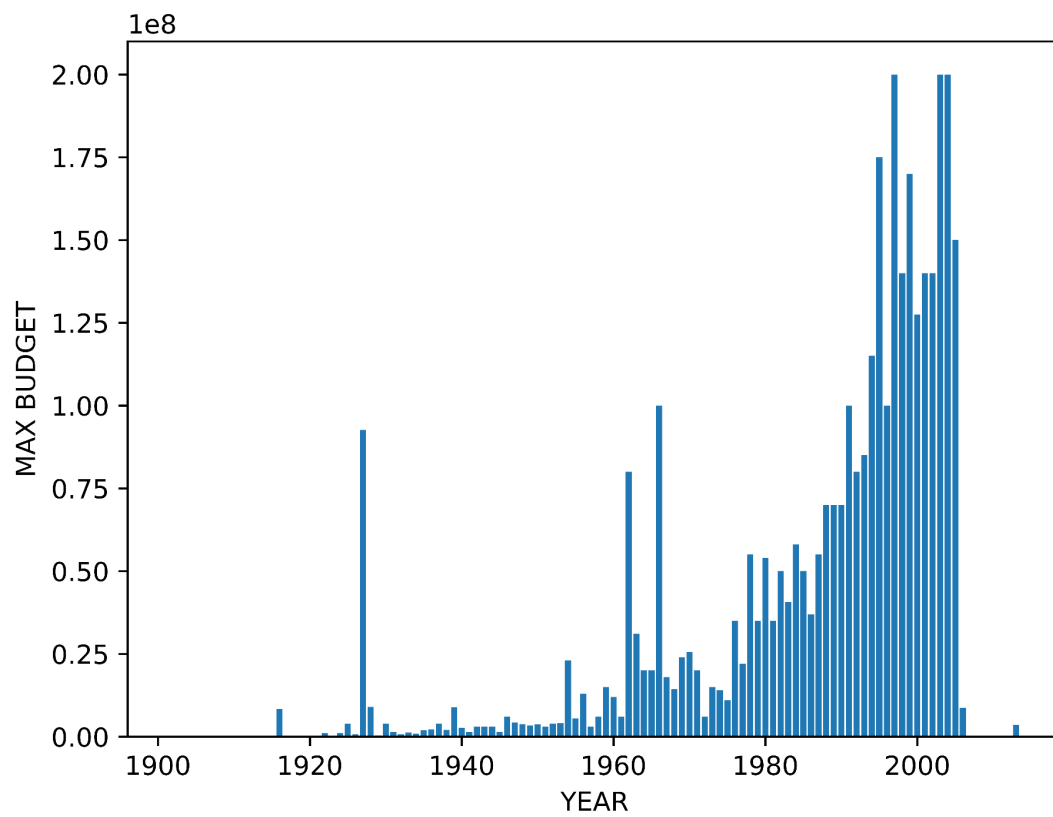
2. Αριθμός ταινιών ανά είδος (genre)



3. Αριθμός ταινιών ανά είδος (genre) και ανά έτος.

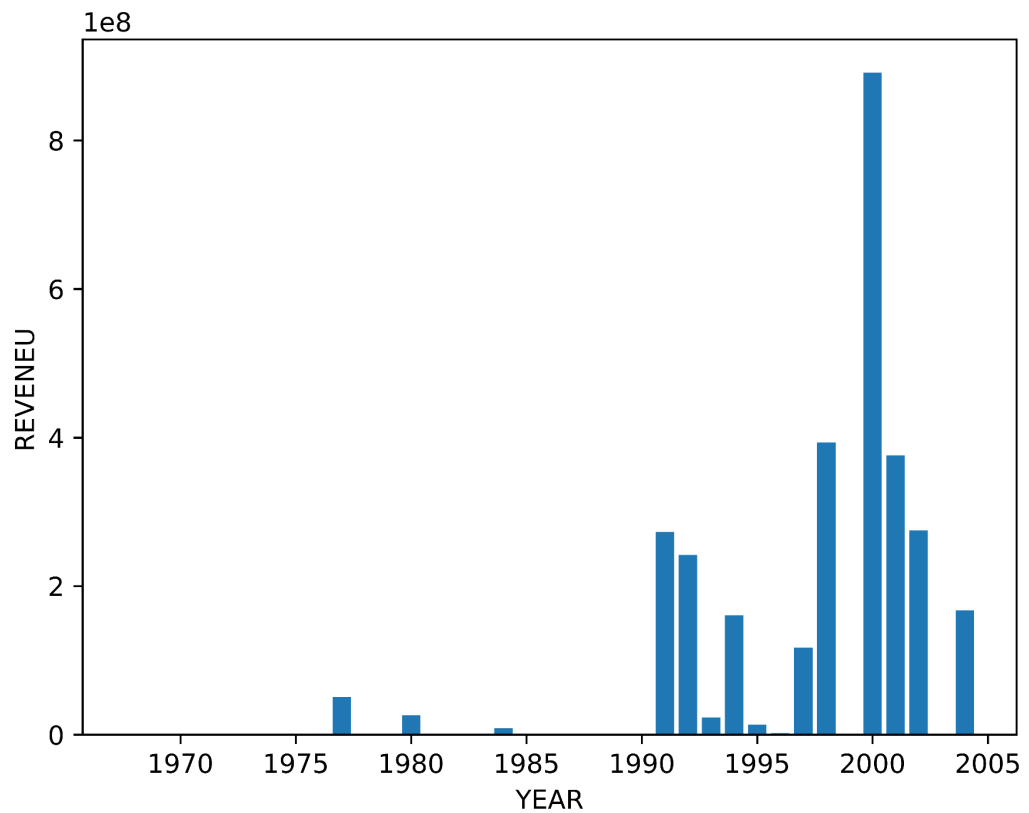


4. Το υψηλότερο budget ταινίας ανά έτος (δεν μας ενδιαφέρει για ποια ταινία)



Στατιστικά ηθοποιών

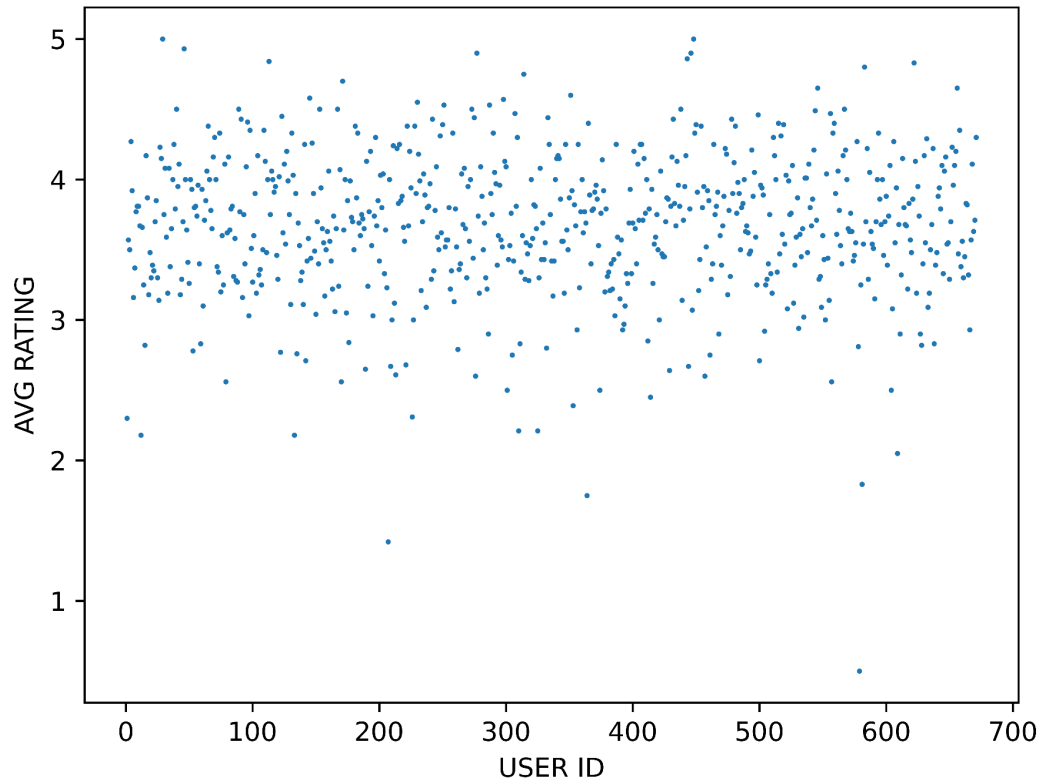
5. Για τον αγαπημένο σας ηθοποιό, το σύνολο των εσόδων (*revenue*) για τις ταινίες στις οποίες έχει συμμετάσχει ανά έτος.



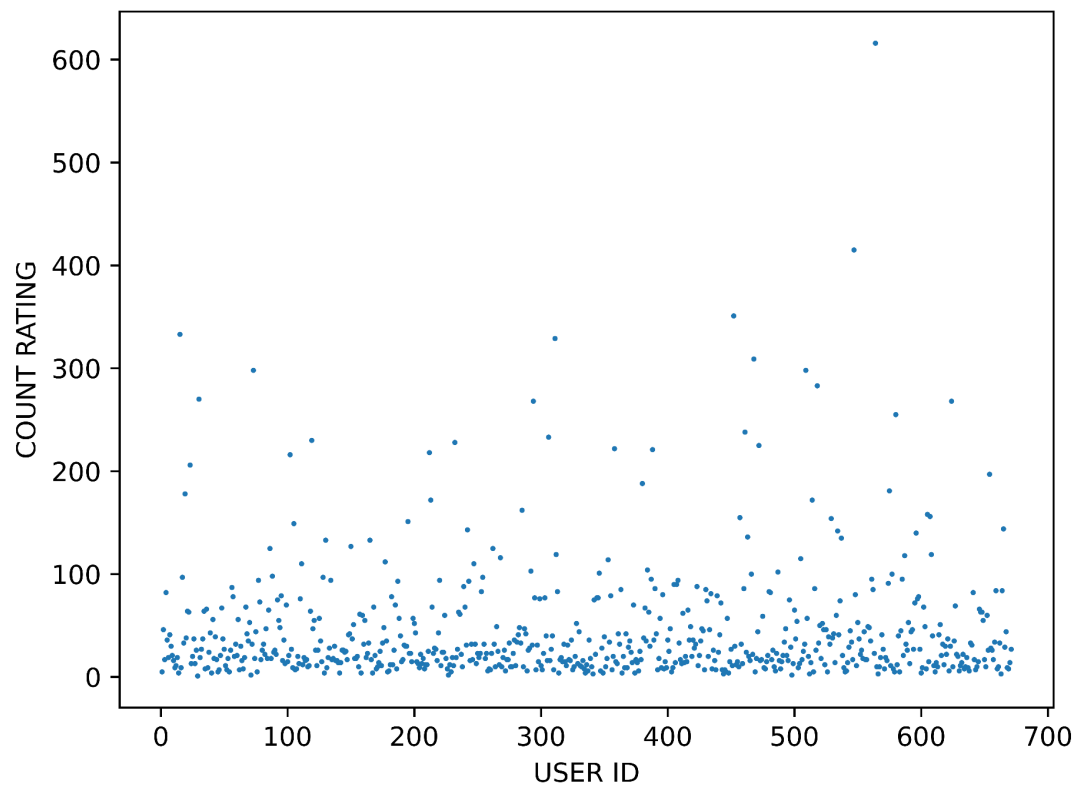
Σε αυτό το σημείο θα θέλαμε να τονίσουμε πως ξέρουμε πως υπάρχουν χρονιές που μπορεί να μην είχε συμμετάσχει σε κάποια ταινία αλλά δεν ξέρουμε με σιγουριά αν θέλατε αυτό να μην φανεί. Αν δεν έπρεπε να φανεί θα μπορούσαμε να το αποφύγουμε προσθέτοντας στον κώδικα SQL στο σημείο του `HAVING` την εντολή `AND SUM(m.revenue) > 0`.

Βαθμολογίες Χρήστη

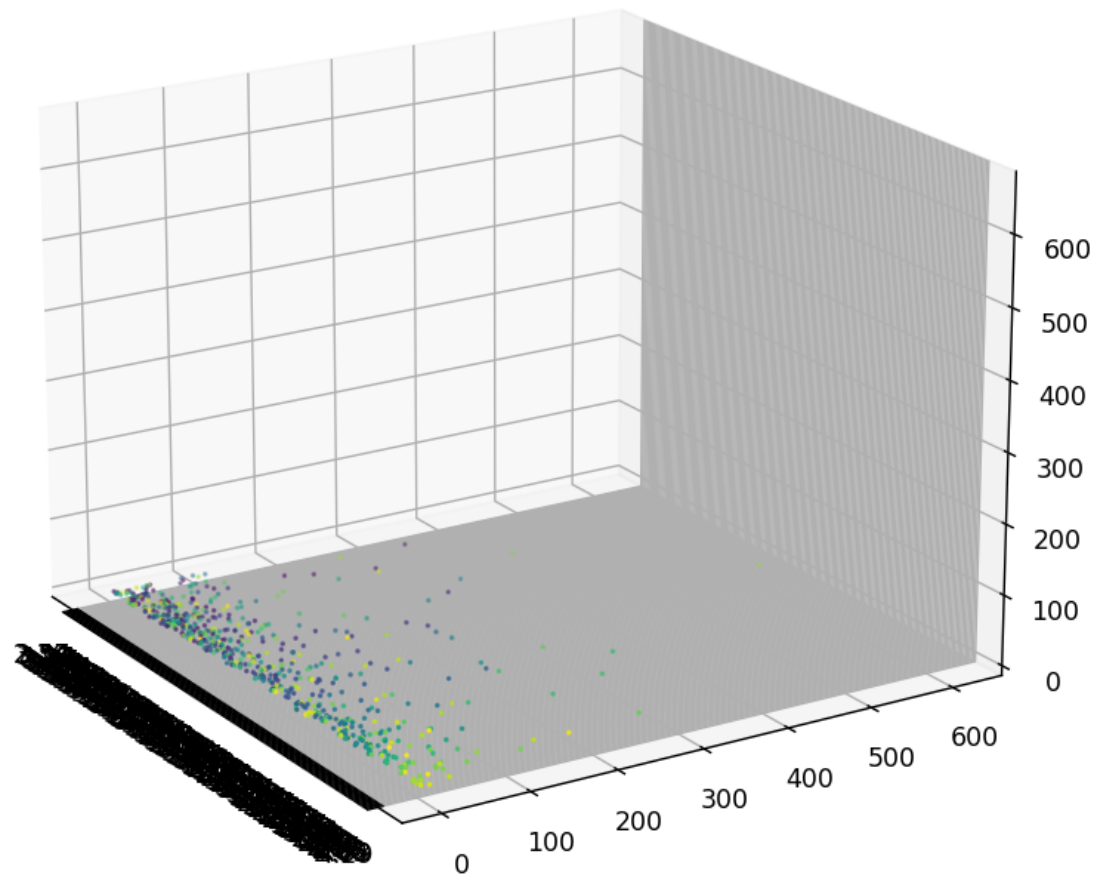
6. Μέση βαθμολογία (*rating*) ανά χρήστη (*scatter plot*).



7. Αριθμός από βαθμολογίες ανά χρήστη (scatter plot).



8. Scatter plot το οποίο θα έχει ένα σημείο για κάθε χρήστη που στον x άξονα φαίνεται ο αριθμός των αξιολογήσεων του χρήστη και στον y άξονα η μέση βαθμολογία του.



9. Μέση βαθμολογία (rating) ανά είδος ταινίας.

