



Ονόματα Φοιτητών:

Πέτρος Λεχαρέας
Βασίλης Βασιλούδης
Τριαντάφυλλος Φώτογλου

Μάθημα: Εξόρυξη Δεδομένων και Γνώση

Θέμα: Αξιοποίηση δεδομένων με τη χρήση του
Recommendation System

Εισαγωγή

Πολλές εταιρείες σήμερα δεν αξιοποιούν σωστά τα δεδομένα τα οποία καταγράφουν από τους χρήστες τους, λόγω έλλειψης τεχνογνωσίας και εξειδίκευσης στην εφαρμογή μεθόδων εξόρυξης δεδομένων και γνώσης. Αυτό έχει ως κύριο αποτέλεσμα οι εταιρείες αυτές να χάνουν πιθανά κέρδη και το ενδιαφέρον των χρηστών τους. Στην προκειμένη περίπτωση, η ομάδα μας έχει παρατηρήσει ότι στο δικό σας movie site θα μπορούσαμε να αξιοποιήσουμε τα δεδομένα που συλλέγονται από τους επισκέπτες σας, με στόχο την ανάλυση των προτιμήσεων τους, έχοντας ως άμεσο στόχο τη προβολή ανάλογων ταινιών. Κάτι τέτοιο θα βοηθήσει στην καλύτερη εμπειρία του επισκέπτη αλλά και στην μεγιστοποίηση των κερδών σας.

Η παρακολούθηση της συμπεριφοράς των χρηστών σας θα βοηθήσει να εντοπίσετε πιθανές προκλήσεις που μπορεί να αντιμετωπίζουν οι χρήστες όταν περιηγούνται στην ιστοσελίδα σας. Σας βοηθά επίσης να κατανοήσετε κάθε κίνηση του χρήστη, να εξαλείψετε τυχόν εικασίες και να εστιάσετε στη λεπτομερή τμηματοποίηση χρηστών για να βελτιστοποιήσετε τις υπηρεσίες σας (το e-movie theater) για ανάπτυξη.

Με λίγα λόγια η χρήση των δεδομένων των χρηστών που έχουν συλλεχθεί είναι αυτά που θα οδηγήσουν, με τη σωστή επεξεργασία τους, τον κάθε χρήστη να έχει μία καλύτερη εμπειρία μέσα στην ιστοσελίδα σας. Μερικά από τα είδη των δεδομένων που περισυλλέγονται είναι τα παρακάτω:

```
## Classes 'data.table' and 'data.frame':  9000055 obs. of  6 variables:
## $ userId   : int  1 1 1 1 1 1 1 1 1 1 ...
## $ movieId  : num  122 185 292 316 329 355 356 362 364 370 ...
## $ rating   : num  5 5 5 5 5 5 5 5 5 5 ...
## $ timestamp: int  838985046 838983525 838983421 838983392 838983392 838984474 838983653 838984885 838983707 8389845
96 ...
## $ title    : chr  "Boomerang (1992)" "Net, The (1995)" "Outbreak (1995)" "Stargate (1994)" ...
## $ genres   : chr  "Comedy|Romance" "Action|Crime|Thriller" "Action|Drama|Sci-Fi|Thriller" "Action|Adventure|Sci-Fi"
...
## - attr(*, ".internal.selfref")=<externalptr>
```

Μηχανισμοί καταγραφής αυτών των δεδομένων θα μπορούσε κανείς να πει πως θεωρείται η βάση δεδομένων στην οποία αποθηκεύονται όλα τα δεδομένα περιήγησης του χρήστη με στόχο την αξιοποίηση μεταγενέστερα.

Παρακάτω θα σας αναφέρουμε τα είδη των βασικών δεδομένων που είναι απαραίτητα για την σωστή ανάλυση για την επίτευξη του επιθυμητού αποτελέσματος.

1. Δεδομένα συμπεριφοράς χρήστη (**Historical data**)

- Καταγραφή δραστηριότητας στον ιστότοπο: κλικ, αναζητήσεις, προβολές σελίδων και στοιχείων
- Δραστηριότητες εκτός τοποθεσίας: παρακολούθηση κλικ σε email, σε εφαρμογές για κινητά και στις ειδοποιήσεις push τους

2. Στοιχεία συγκεκριμένων αντικειμένων (**Particular item details**)

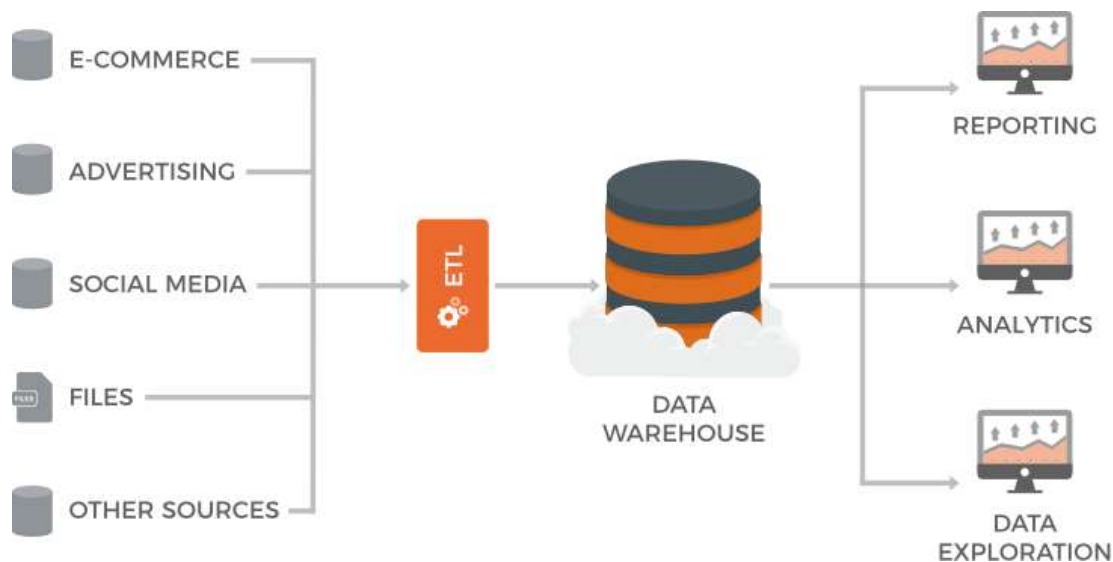
- Τίτλος
- Κατηγορία
- Είδος
- Περιγραφή
- Στυλ

3. Πληροφορίες για τα συμφραζόμενα (**Contextual information**)

- Συσκευή που χρησιμοποιείται
- Τωρινή τοποθεσία
- URL παραπομπής

Η καταγραφή των δεδομένων του κάθε χρήστη θα πραγματοποιείται με την αποθήκευσή τους μέσα σε Data Warehouses. Τα Data Warehouses, είναι με λίγα λόγια αποθήκες που αποθηκεύουν ξεχωριστά, δεδομένα από τις βάσεις δεδομένων, με κύρια διαφορά ότι η συλλογή αυτών των δεδομένων είναι προσανατολισμένη σε ένα συγκεκριμένο θέμα, στη δική μας περίπτωση η ανάλυση τους με στόχο την εκτίμηση ανάλογα των προσωπικών προτιμήσεων κάθε χρήστη. Ο χώρος στον οποίο δεσμεύονται τα ακατέργαστα δεδομένα κατα προτίμηση θα πρέπει να είναι ψηφιακός-εικονικός (Cloud) για να μην υπάρχει απώλεια δεδομένων σε περίπτωση που τα αποθηκεύαμε σε κάποιο φυσικό μέσο. Παρακάτω αναφέρουμε ορισμένα από τα οφέλη των Αποθηκών δεδομένων :

- Ενημερωμένη λήψη αποφάσεων
- Συγκεντρωτικά δεδομένα από πολλές πηγές
- Ανάλυση ιστορικών δεδομένων
- Ποιότητα, συνέπεια και ακρίβεια δεδομένων
- Διαχωρισμός της επεξεργασίας αναλυτικών στοιχείων από τις βάσεις δεδομένων συναλλαγών, που βελτιώνει την απόδοση και των δύο συστημάτων



Τα βασικά πρόσωπα που θα έχουν πρόσβαση στα δεδομένα προς επεξεργασία με σκοπό την ανάλυση τους και την εξαγωγή χρήσιμων πληροφοριών είναι τα εξής:

- **Marketers**(Διαφημιστές) , οι οποίοι χρησιμοποιούν δεδομένα πελατών, τάσεις του κλάδου και δεδομένα απόδοσης από προηγούμενες καμπάνιες για να σχεδιάσουν στρατηγικές μάρκετινγκ
- **Product managers**(Διευθυντές προϊόντων) , οι οποίοι αναλύουν δεδομένα αγοράς, βιομηχανίας και χρηστών για να βελτιώσουν τα προϊόντα των εταιρειών τους
- **Finance professionals**(Επαγγελματίες οικονομικών) , οι οποίοι χρησιμοποιούν ιστορικά δεδομένα απόδοσης και τάσεις του κλάδου για να προβλέψουν τις οικονομικές τροχιές των εταιρειών τους
- **Human resources and diversity, equity, and inclusion professionals**(Επαγγελματίες ανθρώπινου δυναμικού και διαφορετικότητας, δικαιοσύνης και ένταξης) , οι οποίοι αποκτούν γνώσεις για τις απόψεις, τα κίνητρα και τις συμπεριφορές των εργαζομένων και τα συνδυάζουν με δεδομένα τάσεων του κλάδου για να κάνουν σημαντικές αλλαγές στους οργανισμούς τους

Προεπεξεργασία(preprocess)

Έχοντας πάρει το dataset δεδομένων που υπάρχει στο site σας θα το χρησιμοποιήσουμε για να πάρουμε πληροφορίες σχετικά με το τι αρέσει στο κάθε επισκέπτη και πως θα μπορέσουμε να αξιοποιήσουμε αυτήν την πληροφορία ώστε να μεγιστοποιήσουμε τα κέρδη της επιχείρησης.

Το πρώτο βήμα είναι να κάνουμε Data Exploration:

userId	movieId	rating	timestamp	title	genres
<int>	<dbl>	<dbl>	<int>	<chr>	<chr>
1	122	5	838985046	Boomerang (1992)	Comedy Romance
1	185	5	838983525	Net, The (1995)	Action Crime Thriller
1	231	5	838983392	Dumb & Dumber (1994)	Comedy
1	292	5	838983421	Outbreak (1995)	Action Drama Sci-Fi Thriller
1	316	5	838983392	Stargate (1994)	Action Adventure Sci-Fi
1	329	5	838983392	Star Trek: Generations (1994)	Action Adventure Drama Sci-Fi

6 rows

Αυτό είναι ένα μικρό preview των ταινιών, με τα αντίστοιχα timestamps, τα rating, τον τίτλο και το είδος της κάθε ταινίας. Έπειτα θα κάνουμε train το data set αυτό και θα τεστάρουμε το data structure του ώστε να μας δώσει πληροφορίες σχετικά με το τι έχουμε να αντιμετωπίσουμε.

userId	movieId	rating	timestamp	title	genres
<int>	<dbl>	<dbl>	<int>	<chr>	<chr>
1	122	5	838985046	Boomerang (1992)	Comedy Romance
1	185	5	838983525	Net, The (1995)	Action Crime Thriller
1	292	5	838983421	Outbreak (1995)	Action Drama Sci-Fi Thriller
1	316	5	838983392	Stargate (1994)	Action Adventure Sci-Fi
1	329	5	838983392	Star Trek: Generations (1994)	Action Adventure Drama Sci-Fi
1	355	5	838984474	Flintstones, The (1994)	Children Comedy Fantasy

Κύρια χαρακτηριστικά του dataset είναι τα εξής:

1. **Data characteristics:** number of users, movies, and genres
2. **Timestamp:** the duration of the data collection
3. **Genre:** the labels associated with the movie data
4. **Rating:** distribution of user ratings for the movies
5. **UserId:** which users rated movies more often and their ratings

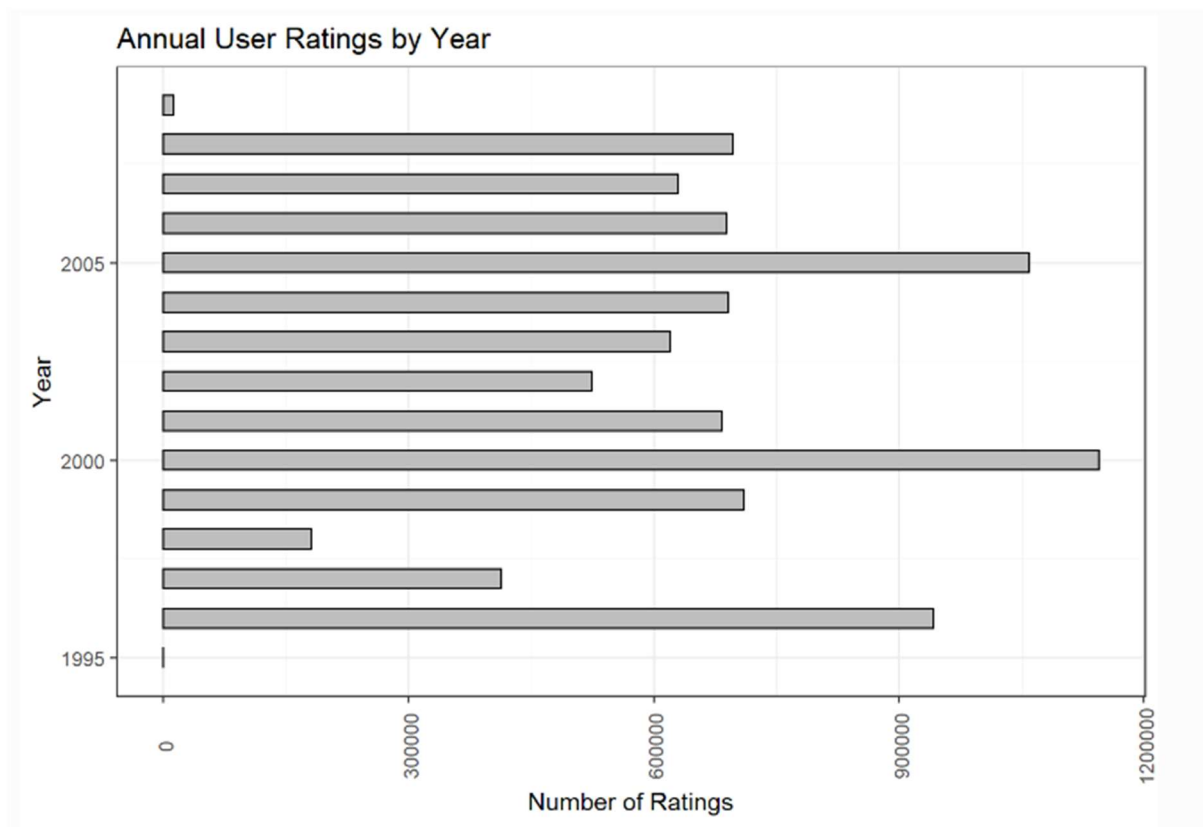
Number of Users <int>	Number of Movies <int>	Number of Genres (unique and combinations) <int>
69878	10677	797

1 row

Date of First Movie in Dataset <date>	Date of Last Movie in Dataset <date>	Duration <Duration>
1995-01-09	2009-01-05	441479727s (-13.99 years)

1 row

Από αυτά τα διαγράμματα μπορούμε εύκολα να δούμε πως υπάρχει πληροφορία για ratings σχετικά με το site σας 14 χρονων.



Υπάρχουν 19 είδη ταινιών

count	genres	Quantity of Movies by Genre
<int>	<chr>	<int>
0	(no genres listed)	7
0	Action	24482
0	Adventure	2276
0	Animation	329
0	Children	745
0	Comedy	700889
0	Crime	3197
0	Documentary	70041
0	Drama	733296
0	Fantasy	86

1-10 of 20 rows

Previous **1** **2** Next

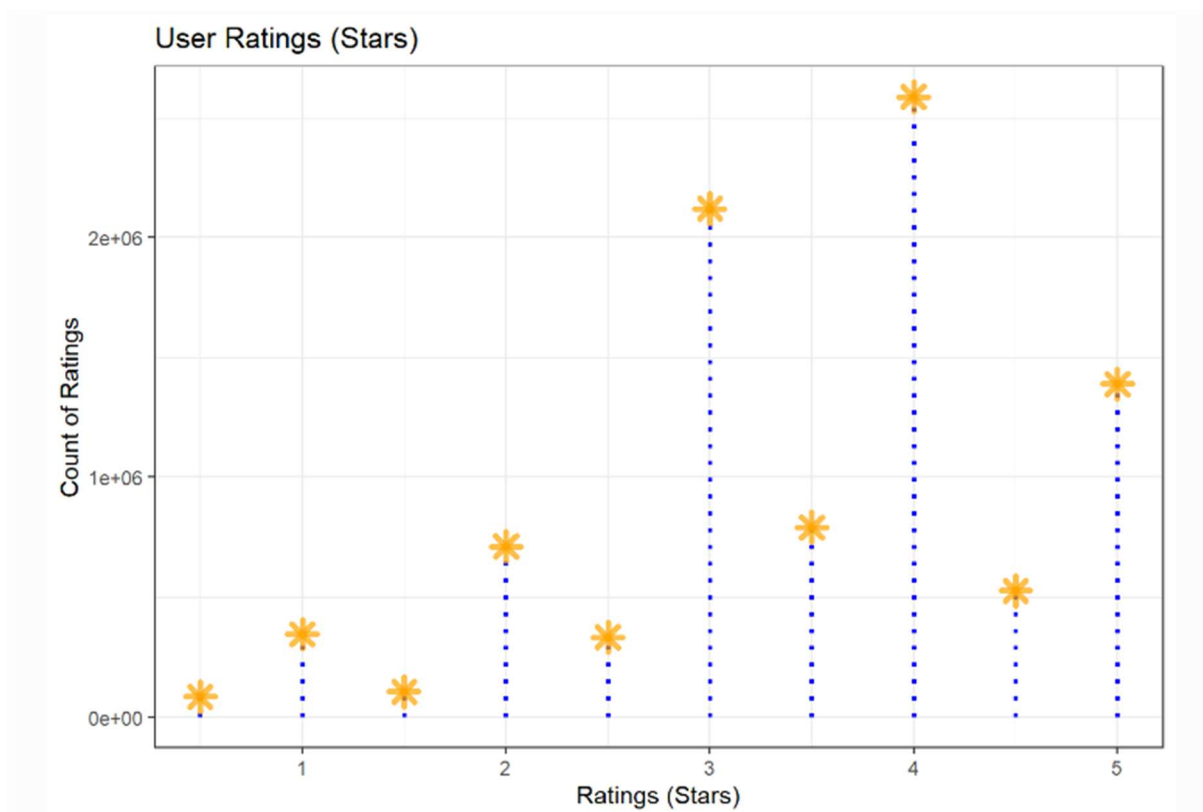
count	genres	Quantity of Movies by Genre
<int>	<chr>	<int>
0	Film-Noir	1575
0	Horror	68738
0	IMAX	14
0	Musical	3851
0	Mystery	246
0	Romance	8410
0	Sci-Fi	10125
0	Thriller	94662
0	War	2300
0	Western	15300

11-20 of 20 rows

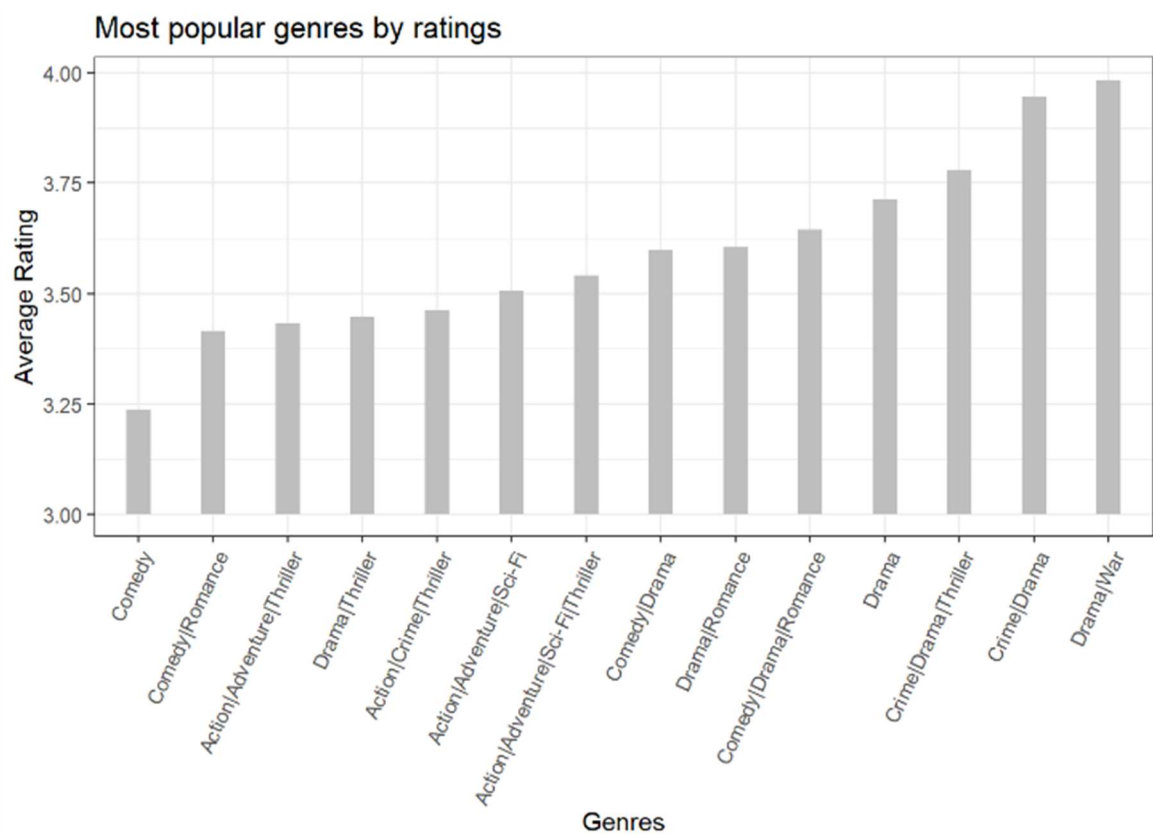
Previous **1** **2** Next

Οι ταινίες βαθμολογούνται σε κλίμακα από το 1 μέχρι το 5

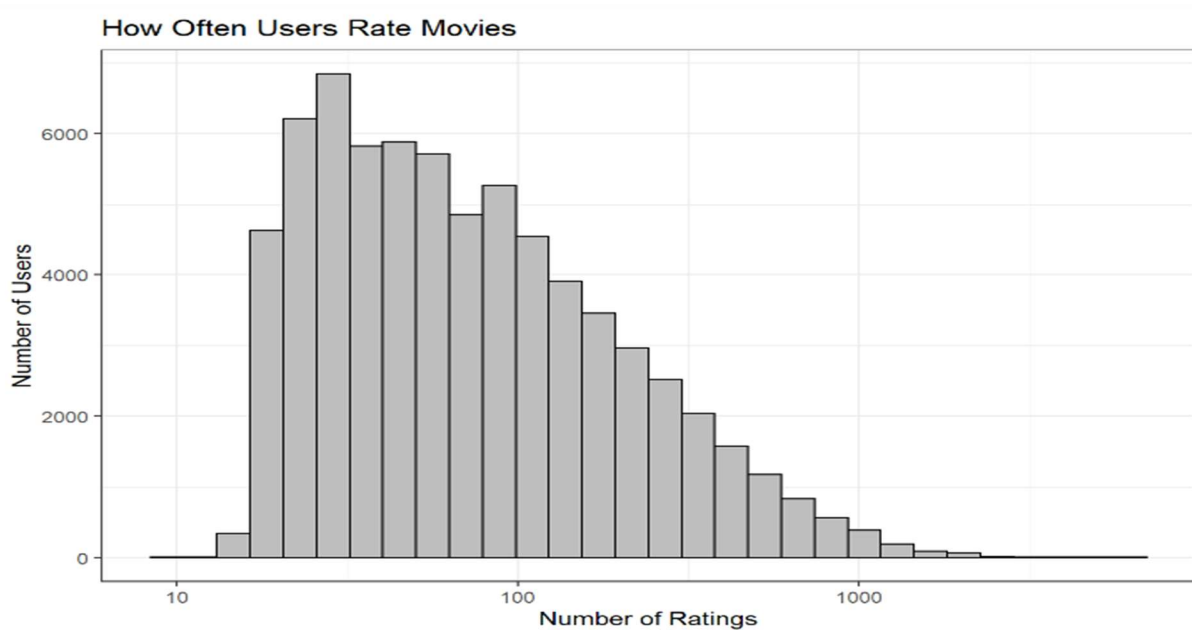
User Ratings - Stars <dbl>	Number of Ratings <int>
0.5	85374
1.0	345679
1.5	106426
2.0	711422
2.5	333010
3.0	2121240
3.5	791624
4.0	2588430
4.5	526736
5.0	1390114



Τα πιο διάσημα είδη ταινιών είναι τα παρακάτω



Το πόσο συχνά οι χρήστες κάνουν rate μία ταινία



Υλοποίηση αλγορίθμων για την επίλυση του προβλήματος

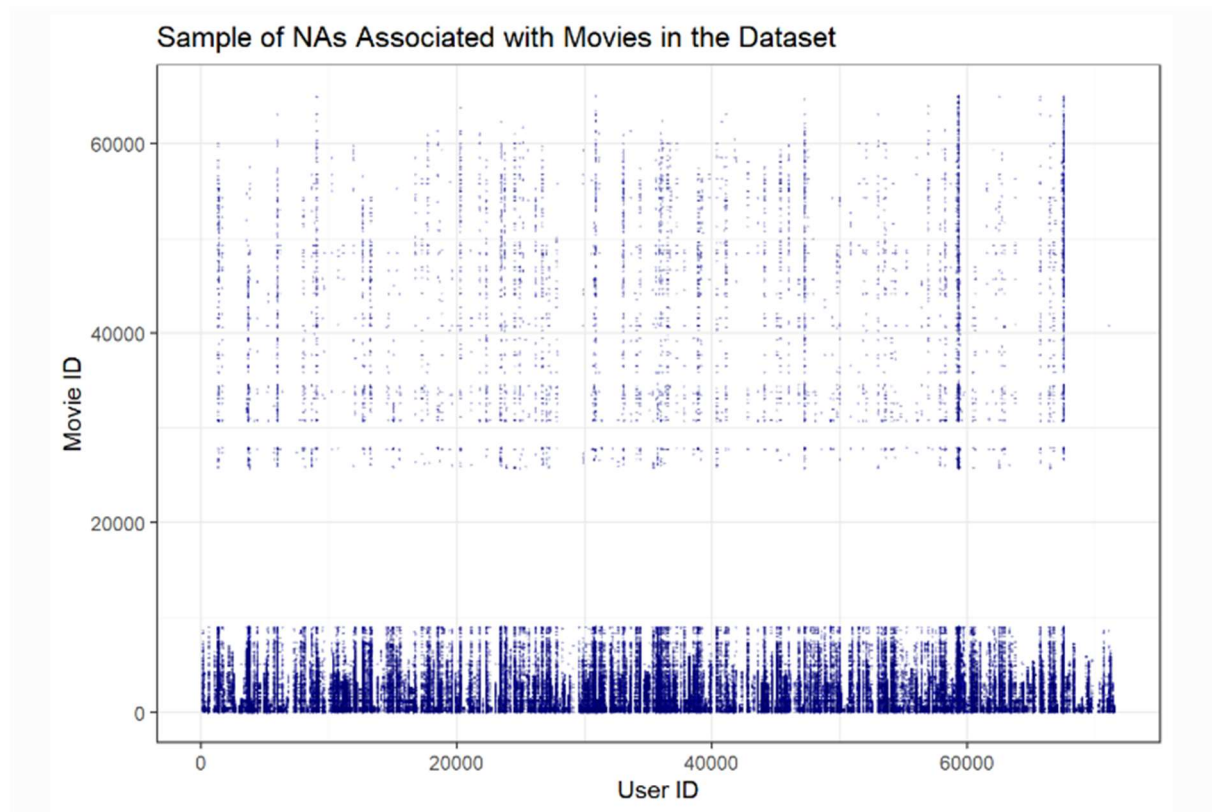
Τώρα που κατανοούμε καλύτερα το σύνολο δεδομένων, πρέπει να αξιολογήσουμε τον τύπο των δεδομένων που επιδιώκουμε να αναλύσουμε για να αυξήσουμε την αποτελεσματικότητά μας. Αυτά τα μέτρα θα επιταχύνουν τον χρόνο επεξεργασίας δεδομένων μειώνοντας την πολυπλοκότητα του μοντέλου. Για αυτό το έργο, μας ενδιαφέρει η δυνατότητα επιλογής 20 από τις κορυφαίες ταινίες από τα είδη με βάση τις αξιολογήσεις των χρηστών. Αυτές οι επιλογές θα παρέχουν τη δυνατότητα επιλογής “ταινίας” από τη λίστα. Εάν έχουμε παρακολουθήσει όλες τις ταινίες της λίστας, μπορούμε να τροποποιήσουμε τη λίστα με τις κορυφαίες ταινίες για να συμπεριλάβουμε πρόσθετες ταινίες (δηλ. κορυφαίες 50, 100 και ούτω καθεξής).

Ας ξεκινήσουμε με τη δημιουργία ενός απλού μοντέλου για το σύστημα συστάσεων, ακολουθούμενο από μεγαλύτερη πολυπλοκότητα. Ο στόχος του συστήματος συστάσεων μας είναι να δημιουργήσει ένα μοντέλο που θα κάνει τις καλύτερες δυνατές προτάσεις (με την υψηλότερη βαθμολογία). Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας regression loss algorithms και ρυθμίζοντας διάφορες παραμέτρους του μοντέλου για να γίνουν προβλέψεις. Αρχικά, θα πάρουμε ένα δείγμα ταινιών που εμφανίζονται σε μια μήτρα και θα αναγνωρίσουμε ότι ορισμένοι χρήστες δεν αξιολόγησαν τις ταινίες (εμφανίζονται ως "NA"):

userId <int>	Forrest Gump (1994) <dbl>	Jurassic Park (1993) <dbl>	Pulp Fiction (1994) <dbl>
13	NA	NA	4
16	NA	3	NA
17	NA	NA	NA
18	NA	3	5
19	4	1	NA

5 rows | 1-4 of 6 columns

Φαίνεται ότι ορισμένοι χρήστες παρείχαν περισσότερα NA από άλλους και είναι δύσκολο να εκτιμηθεί γιατί υπάρχουν. Επιπλέον, ορισμένα αναγνωριστικά ταινιών ήταν πιο επιρρεπή στο να έχουν NA από άλλα. Ο στόχος μας με τη ρύθμιση του συστήματος συστάσεων είναι να επιδιώξουμε να συμπληρώσουμε όσο το δυνατόν περισσότερα NA για να βελτιστοποιήσουμε το σύστημα συστάσεων.



Σε αυτό το σημείο θα δημιουργηθεί το Recommendation System. Το απλούστερο μοντέλο που υποθέτει την ίδια βαθμολογία για όλες τις ταινίες και τους χρήστες που εξηγούνται με τυχαία παραλλαγή χρησιμοποιεί τον τύπο:

$$Y_{u,i} = \mu + \epsilon_{u,i}$$

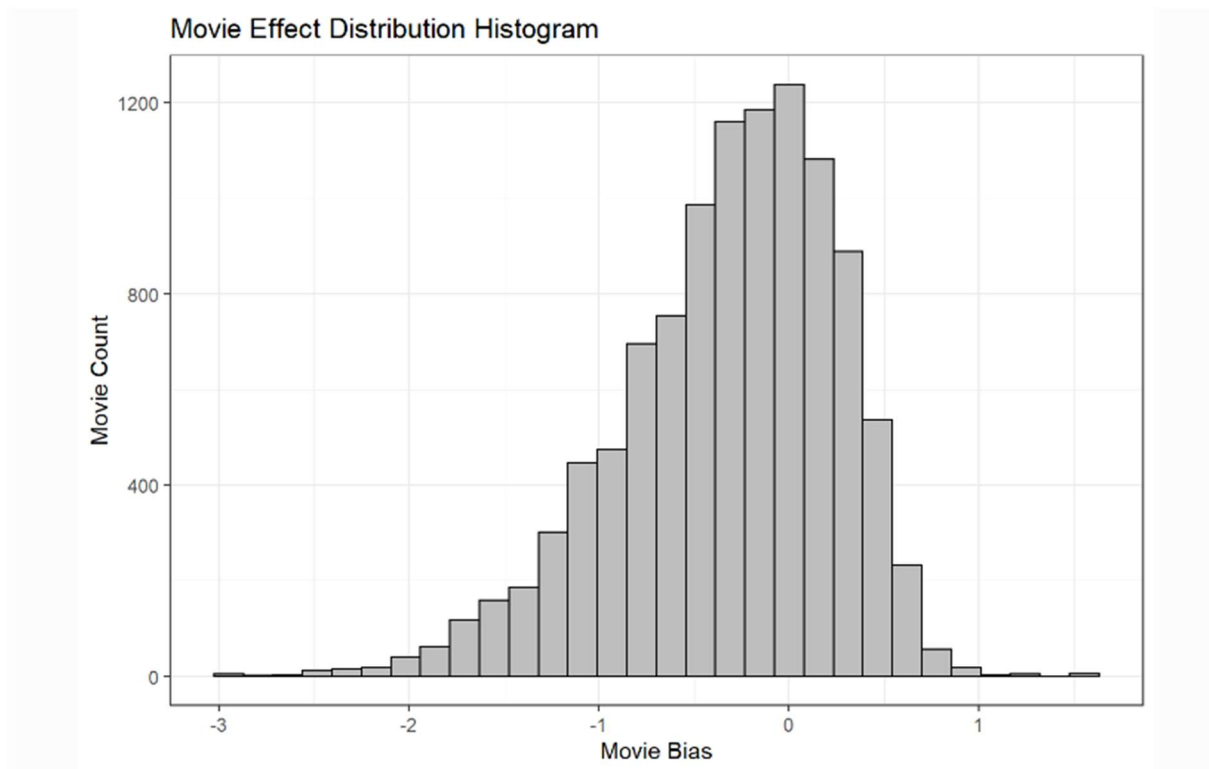
όπου $\epsilon_{u,i}$ ανεξάρτητα σφάλματα δειγματοληπτικά από τη διανομή με κέντρο το 0, και μ την «αληθινή» βαθμολογία για τις ταινίες. Το RMSE αυξάνεται με τιμές διαφορετικές από το μέσο όρο όπως γνωρίζουμε, ορισμένες ταινίες ευνοούνται έναντι άλλων και έχουν υψηλότερη βαθμολογία. Για να αντισταθμίσουμε αυτό, εισάγουμε τη μεταβλητή b_i που αναφέρεται ως προκατάληψη (bias). Αυτή η φόρμουλα έχει ως εξής:

$$Y_{u,i} = \mu + b_i + \epsilon_{u,i}$$

Μπορεί να χρησιμοποιηθεί ο ακόλουθος κώδικας, ο οποίος θα λάβει την εκτίμηση ελαχίστων τετραγώνων b_i ως μέσο όρο Y_{ui} για κάθε ταινία $\{i\}$ (τιμωρώντας τα ελάχιστα τετράγωνα):

movieId <dbl>	b_i <dbl>
1	0.4151725
2	-0.3070658
3	-0.3654817
4	-0.6481659
5	-0.4437933
6	0.3028191

6 rows



Ας δούμε πώς η προκατάληψη ταινίας βελτιώνει το RMSE για τις προβλέψεις του συστήματος συστάσεων:

method <chr>	RMSE <dbl>
Average	1.0612018
Movie Bias Effect	0.9439087

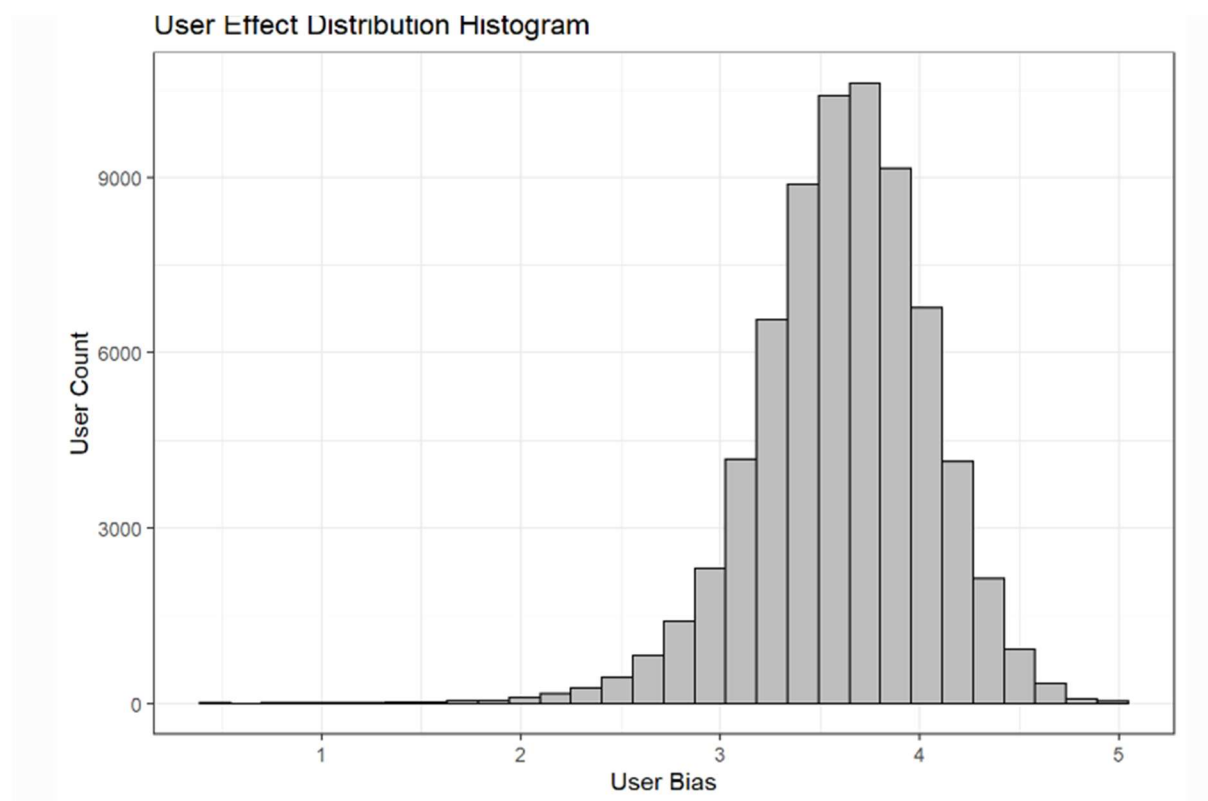
2 rows

Το RMSE μας βελτιώνεται, ωστόσο ας δούμε αν το User Effects Bias θα βελτιώσει τις προτάσεις. Αυτός είναι ο τύπος που θα χρησιμοποιήσουμε, προσθέτοντας τη μεταβλητή εφέ χρήστη στον προηγούμενο τύπο:

$$Y_{u,i} = \mu + b_i + b_u + \epsilon_{u,i}$$

method <chr>	RMSE <dbl>
Average	1.0612018
Movie Bias Effect	0.9439087
User Bias Effect	0.8653488

3 rows

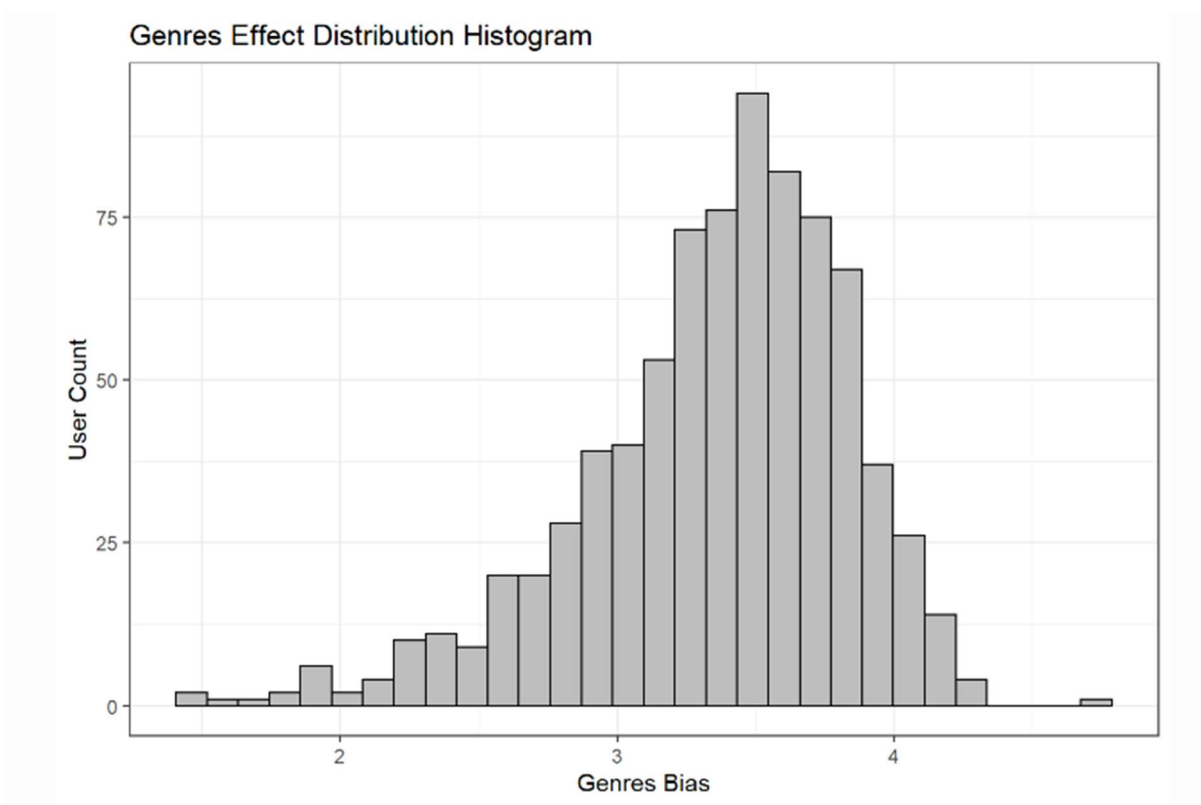


Ας δούμε πώς τα είδη επηρεάζουν τις προβλέψεις μας. Συγκριτικά με την προκατάληψη της προηγούμενης ταινίας και των εφέ χρήστη, θα προσθέσουμε τα είδη χρησιμοποιώντας τον ακόλουθο τύπο:

$$Y_{u,i} = \mu + b_i + b_u + b_g + \epsilon_{u,i}$$

method	RMSE
<chr>	<dbl>
Average	1.0612018
Movie Bias Effect	0.9439087
User Bias Effect	0.8653488
Genres Bias Effect	0.8649469

4 rows



Τώρα που συντονίσαμε αυτό το μοντέλο τόσο με την προκατάληψη ταινίας, την προκατάληψη χρήστη και την προκατάληψη του είδους, παρατηρούμε ότι το RMSE μας έχει βελτιωθεί. Ας διερευνήσουμε αρχικά τα λάθη που δημιουργήθηκαν προσθέτοντας την προκατάληψη και, στη συνέχεια, ας προσδιορίσουμε πώς να το διορθώσουμε με τακτοποίηση.

Τα 10 μεγαλύτερα λάθη (Regulation):

title
<chr>
Pok��mon Heroes (2003)
Shawshank Redemption, The (1994)
Godfather, The (1972)
Usual Suspects, The (1995)
Schindler's List (1993)
Pokemon 4 Ever (a.k.a. Pok��mon 4: The Movie) (2002)
Casablanca (1942)
Rear Window (1954)
Third Man, The (1949)
Seven Samurai (Shichinin no samurai) (1954)
1-10 of 10 rows

Οι 10 καλύτερες και χειρότερες ταινίες με βάση την προκατάληψη (Bias)

```
## [1] "Hellhounds on My Trail (1999)"
## [2] "Satan's Tango (S  t  tang  ) (1994)"
## [3] "Shadows of Forgotten Ancestors (1964)"
## [4] "Fighting Elegy (Kenka erejii) (1966)"
## [5] "Sun Alley (Sonnenallee) (1999)"
## [6] "Blue Light, The (Das Blaue Licht) (1932)"
## [7] "Who's Singin' Over There? (a.k.a. Who Sings Over There) (Ko to tamo peva) (1980)"
## [8] "Human Condition II, The (Ningen no joken II) (1959)"
## [9] "Human Condition III, The (Ningen no joken III) (1961)"
## [10] "Constantine's Sword (2007)"

## [1] "Besotted (2001)"
## [2] "Hi-Line, The (1999)"
## [3] "Accused (Anklaget) (2005)"
## [4] "Confessions of a Superhero (2007)"
## [5] "War of the Worlds 2: The Next Wave (2008)"
## [6] "SuperBabies: Baby Geniuses 2 (2004)"
## [7] "Hip Hop Witch, Da (2000)"
## [8] "Disaster Movie (2008)"
## [9] "From Justin to Kelly (2003)"
## [10] "Criminals (1996)"
```

Έπειτα χρησιμοποιούμε το Regularization Model. Με τη χρήση της τακτοποίησης μπορούμε να βελτιώσουμε περαιτέρω το μοντέλο τιμωρώντας μεγάλες εκτιμήσεις που επικεντρώνονται σε μικρά μεγέθη δειγμάτων. Είναι παρόμοια με την Bayesian προσέγγιση που συρρικνώνει τις προβλέψεις. Η δυνατότητα τιμωρίας των εκτιμήσεων παρέχει τη δυνατότητα ελέγχου της μεταβλητότητας των εφέ της ταινίας,

αντί να ελαχιστοποιείται η εξίσωση ελαχίστων τετραγώνων. Χρησιμοποιούμε το σετ εκπαίδευσης για να συντονίσουμε το μοντέλο.

Οι 10 καλύτερες/χειρότερες ταινίες μετά το Tuning

title
<chr>
Shawshank Redemption, The (1994)
Usual Suspects, The (1995)
Schindler's List (1993)
Seven Samurai (Shichinin no samurai) (1954)
Godfather, The (1972)
Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)
Star Wars: Episode IV - A New Hope (a.k.a. Star Wars) (1977)
Star Wars: Episode V - The Empire Strikes Back (1980)
Memento (2000)
Matrix, The (1999)
1-10 of 10 rows

title
<chr>
Superman IV: The Quest for Peace (1987)
Aces: Iron Eagle III (1992)
Dumb and Dumberer: When Harry Met Lloyd (2003)
Police Academy 6: City Under Siege (1989)
Amityville 1992: It's About Time (1992)
Meatballs III (1987)
Turbo: A Power Rangers Movie (1997)
Spice World (1997)
Street Fighter (1994)
Speed 2: Cruise Control (1997)
1-10 of 10 rows

Στη συνέχεια φτάνουμε στο τελευταίο και σημαντικότερο στάδιο το οποίο είναι όταν ένας χρήστης θα ψάχνει να βρει τι ταινία θέλει να δει να μην χάνει χρόνο αλλά να υπάρχουν recommended προς αυτόν.

Αν για παράδειγμα θέλει να δει Action Genre ταινία να του εμφανίζεται κάτι σαν αυτό:

title
<chr>
Seven Samurai (Shichinin no samurai) (1954)
Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)
Star Wars: Episode IV - A New Hope (a.k.a. Star Wars) (1977)
Star Wars: Episode V - The Empire Strikes Back (1980)
Matrix, The (1999)
Fight Club (1999)
Saving Private Ryan (1998)
Professional, The (Le Professionnel) (1981)
Butch Cassidy and the Sundance Kid (1969)
Braveheart (1995)
1-10 of 20 rows
Previous 1 2 Next

Αν θέλει να δει Comedie το αντίστοιχο :

title
<chr>
Wallace & Gromit: The Wrong Trousers (1993)
Pulp Fiction (1994)
Fargo (1996)
Wallace & Gromit: A Close Shave (1995)
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1964)
Eternal Sunshine of the Spotless Mind (2004)
Butch Cassidy and the Sundance Kid (1969)
Postman, The (Postino, Il) (1994)
Forrest Gump (1994)
Snatch (2000)
1-10 of 20 rows
Previous 1 2 Next

Αν θέλει Documentary:

title
<chr>
When We Were Kings (1996)
Hoop Dreams (1994)
Paradise Lost: The Child Murders at Robin Hood Hills (1996)
Celluloid Closet, The (1995)
Microcosmos (Microcosmos: Le peuple de l'herbe) (1996)
Baraka (1992)
Looking for Richard (1996)
Anne Frank Remembered (1995)
Line King: The Al Hirschfeld Story, The (1996)
Thin Blue Line, The (1988)
1-10 of 20 rows
Previous 1 2 Next

Συνοψίζοντας, Τα αποτελέσματα από αυτό το έργο συστήματος συστάσεων ταινιών ήταν εξαιρετικά επιτυχημένα. Χρησιμοποιώντας πολλές επαναλήψεις της ανάλυσης RMSE, η τιμή μειώθηκε σημαντικά για να παρέχει πιο ακριβείς προβλέψεις από τα δεδομένα της ταινίας. Οι προβλέψεις ταινιών ήταν εμφανείς μέσω της επίδειξης των επιλογών ταινιών τόσο πριν όσο και μετά τη βελτιστοποίηση του RMSE. Το τελικό σύστημα σύστασης ταινιών φάνηκε να έχει δύο ακραίες τιμές, οι οποίες δεν βαθμολογήθηκαν με 5 αστέρια.

method	RMSE
<chr>	<dbl>
Average	1.0612018
Movie Bias Effect	0.9439087
User Bias Effect	0.8653488
Genres Bias Effect	0.8649469
Regularization	0.8648165
5 rows	

Συμπέρασμα(Conclusion)

Υπολογίζεται ότι ο παγκόσμιος κινηματογράφος έχει κυκλοφορήσει περισσότερες από 500.000 ταινίες αριθμός που ξεπερνά τον έλεγχο ενός ατόμου. Με έναν τόσο τεράστιο αριθμό κινηματογραφικών ταινιών για να διαλέξετε, η ανάπτυξη και η χρήση recommendation systems είναι ένα κρίσιμο βήμα για να γίνει αυτή η διαδικασία ευκολότερη και πιο εφικτή. Συμπερασματικά λοιπόν και μετά την ανάλυση και αξιοποίηση όλων αυτών των δεδομένων, μπορούμε να πάρουμε αποφάσεις για το ποιά είδη ταινιών αρέσουν στο ευρύτερο κοινό και στοχευμένα πλέον και με βάση τις βαθμολογίες τους να προβάλλονται και να προτείνονται σε αυτούς οι ανάλογες ταινίες με βάση την προτίμηση τους. Επομένως τα κέρδη της εταιρείας σιγά σιγά θα γνωρίζουν ραγδαία αύξηση αφού πλέον όλο και περισσότεροι χρήστες θα μένουν ικανοποιημένοι από την πλατφόρμα και έτσι θα υπάρχουν και περισσότερες συστάσεις για αυτήν.