# Dataset

Το σύνολο δεδομένων για την παρούσα ανάλυση προήλθε από τον επίσημο ιστότοπο IMDb μέσω του IMDb Datasets. Δύο αρχεία μεταφορτώθηκαν: title.basics.tsv.gz (923 MB) και title.ratings.tsv.gz (25 MB)

Το αρχείο title.basics.tsv.gz περιέχει μεταδεδομένα για διάφορους τίτλους, όπως:

* tconst (string): Ένα μοναδικό αλφαριθμητικό αναγνωριστικό για κάθε τίτλο.
* titleType (συμβολοσειρά): Η μορφή του τίτλου, όπως ταινία, μικρού μήκους, τηλεοπτική σειρά κ.λπ.
* primaryTitle (συμβολοσειρά): Ο κοινώς γνωστός τίτλος, που χρησιμοποιείται συχνά σε διαφημιστικό υλικό.
* originalTitle (συμβολοσειρά): Ο τίτλος στην πρωτότυπη γλώσσα του.
* isAdult (boolean): Δείχνει αν ο τίτλος ταξινομείται ως ενήλικος (1) ή μη ενήλικος (0).
* startYear (ΧΧΧΧΧ): Το έτος κυκλοφορίας ή το έτος έναρξης της σειράς για τις τηλεοπτικές εκπομπές.
* endYear (YYYY): Το έτος λήξης μιας τηλεοπτικής σειράς ή «\N» για άλλους τύπους τίτλων.
* runtimeMinutes: Η αρχική διάρκεια του τίτλου σε λεπτά.
* genres (string array): Μέχρι τρία είδη που σχετίζονται με τον τίτλο.

Το αρχείο title.ratings.tsv.gz συμπληρώνει το αρχείο title.basics.tsv.gz παρέχοντας δεδομένα βαθμολογίας που δημιουργούνται από τους χρήστες:

* tconst (string): Το ίδιο μοναδικό αναγνωριστικό, που επιτρέπει τη συγχώνευση των δύο αρχείων.
* averageRating: Ο σταθμισμένος μέσος όρος όλων των αξιολογήσεων των χρηστών για τον τίτλο.
* numVotes: Ο συνολικός αριθμός των ψήφων που έχει λάβει ο τίτλος.

Τα αρχεία αυτά επιλέχθηκαν για το εύρος και το βάθος τους, επιτρέποντας τη λεπτομερή ανάλυση των τάσεων στα χαρακτηριστικά των τίτλων και στις προτιμήσεις των χρηστών με την πάροδο του χρόνου. Ο συνδυασμός μεταδεδομένων και αξιολογήσεων χρηστών προσφέρει μια ισχυρή βάση για τη διερεύνηση προτύπων σε διάφορα είδη, χρόνο εκτέλεσης, δημοτικότητα και άλλα.

# Ανάλυση τάσεων και μοτίβων σε δεδομένα τηλεοπτικών σειρών

## Εισαγωγή

Η μελέτη αξιοποιεί εκτεταμένα σύνολα δεδομένων τηλεοπτικών σειρών για να εξετάσει τις τάσεις στην τηλεοπτική παραγωγή, τα δημοφιλή είδη και τη διάρκεια προβολής κατά τη διάρκεια των δεκαετιών. Χρησιμοποιώντας τεχνικές οπτικοποίησης δεδομένων και στατιστικής μοντελοποίησης, η ανάλυση προσφέρει πληροφορίες για την εξέλιξη του τηλεοπτικού περιεχομένου και των προτιμήσεων του κοινού. Τα ευρήματα αναδεικνύουν κομβικούς παράγοντες που διαμορφώνουν την τηλεοπτική βιομηχανία και χρησιμεύουν ως βάση για περαιτέρω ακαδημαϊκή έρευνα.

## Περιγραφή και προεπεξεργασία δεδομένων

Τα σύνολα δεδομένων που αναλύθηκαν περιλαμβάνουν πληροφορίες σχετικά με μεταδεδομένα τηλεοπτικών σειρών (π.χ. τίτλοι, έτη έναρξης και λήξης, είδη και διάρκεια) και βαθμολογίες κοινού. Μετά την ανάγνωση των συνόλων δεδομένων, πραγματοποιήθηκαν βήματα προεπεξεργασίας για τον καθαρισμό και το φιλτράρισμα των δεδομένων. Αυτό περιελάμβανε:

1. Φιλτράρισμα των συνόλων δεδομένων ώστε να περιλαμβάνουν μόνο tvSeries και tvMiniSeries.
2. Αφαίρεση καταχωρίσεων με ελλιπή ή άκυρα δεδομένα (π.χ. τιμές \N).
3. Μετασχηματισμός στηλών, όπως διαχωρισμός των ειδών σε λίστες και εξαγωγή δεκαετιών από τα έτη έναρξης.

Αυτά τα βήματα προεπεξεργασίας εξασφάλισαν ότι τα σύνολα δεδομένων ήταν κατάλληλα για ακριβή ανάλυση

## Χρονική κατανομή των τηλεοπτικών σειρών

Εξετάστηκε η χρονική κατανομή της παραγωγής τηλεοπτικών σειρών από το 1927 έως το 2024. Παρατηρήθηκε αξιοσημείωτη αύξηση του αριθμού των σειρών που παράγονται μετά το 1960, με σημαντική αύξηση κατά τις δεκαετίες του 2000 και 2010. Η αύξηση αυτή ευθυγραμμίζεται με την πρόοδο της τεχνολογίας και τον πολλαπλασιασμό των πλατφορμών streaming. Η ανάλυση υπογραμμίζει μια τάση προς υψηλότερους όγκους παραγωγής τηλεοπτικών σειρών τις τελευταίες δεκαετίες.

## Δημοτικότητα του είδους

Η διερεύνηση των δεδομένων για τα είδη αποκάλυψε ότι ορισμένα είδη κυριαρχούν σταθερά, με την κωμωδία, το δράμα και την πραγματικότητα να είναι τα πιο δημοφιλή. Χρησιμοποιήθηκαν ραβδογράμματα και κυκλικά διαγράμματα για την απεικόνιση της κατανομής και των σχετικών αναλογιών αυτών των ειδών. Η ανάλυση εντόπισε επίσης κοινές δυάδες ειδών (π.χ. Δράμα-Κωμωδία), αναδεικνύοντας την επικράτηση σειρών πολλαπλών ειδών.

## Ανάλυση αξιολόγησης

Με τη συγχώνευση μεταδεδομένων τηλεοπτικών σειρών με δεδομένα τηλεθέασης, η μελέτη διερεύνησε:

1. Μέση τηλεθέαση ανά είδος: Ακολουθούν οι σειρές ντοκιμαντέρ και οι κωμωδίες. Τα ευρήματα αυτά υποδηλώνουν τις προτιμήσεις του κοινού για πνευματικά διεγερτικό και συναισθηματικά ελκυστικό περιεχόμενο.
2. Συσχέτιση μεταξύ χρόνου προβολής και βαθμολογίας: Τα διαγράμματα διασποράς έδειξαν μια μέτρια συσχέτιση μεταξύ διάρκειας προβολής και τηλεθέασης, με τις σειρές ενδιάμεσης διάρκειας (30-90 λεπτά) να λαμβάνουν γενικά υψηλότερες βαθμολογίες.
3. Δημοτικότητα σε σχέση με τις βαθμολογίες: Παρατηρήθηκε ασθενής συσχέτιση μεταξύ του αριθμού των ψήφων (δημοτικότητα) και των βαθμολογιών, γεγονός που υποδηλώνει ότι οι σειρές με υψηλή βαθμολογία δεν είναι απαραίτητα και οι πιο δημοφιλείς.

## Διαχρονικές τάσεις

Για να εκτιμηθούν οι διαχρονικές τάσεις στις μετρήσεις τηλεθέασης, χρησιμοποιήθηκε ένα μοντέλο γραμμικής παλινδρόμησης. Τα αποτελέσματα έδειξαν θετική συσχέτιση μεταξύ των ετών κυκλοφορίας και της μέσης τηλεθέασης, γεγονός που υποδηλώνει ότι οι νεότερες σειρές τείνουν να λαμβάνουν υψηλότερη τηλεθέαση. Η τάση αυτή μπορεί να αντανακλά βελτιώσεις στην ποιότητα της παραγωγής και της αφήγησης.

## Ανάλυση χρόνου εκτέλεσης και διάρκειας ζωής

Η ανάλυση του χρόνου προβολής και της διάρκειας ζωής των σειρών σε όλες τις δεκαετίες αποκάλυψε:

1. Τάσεις χρόνου εκτέλεσης: Ο μέσος χρόνος διάρκειας των τηλεοπτικών σειρών αυξήθηκε με την πάροδο των δεκαετιών, ιδίως από τη δεκαετία του 1980 και μετά. Αυτή η μετατόπιση αντανακλά τις μεταβαλλόμενες καταναλωτικές συνήθειες του κοινού και τις προτιμήσεις της πλατφόρμας.
2. Διάρκεια ζωής τίτλων: Η μέση διάρκεια ζωής των τηλεοπτικών σειρών κορυφώθηκε στα μέσα του 20ού αιώνα και έκτοτε έχει μειωθεί, ενδεχομένως λόγω του αυξημένου ανταγωνισμού και των μικρότερων σεζόν στις σύγχρονες παραγωγές.

## Δυναμική των ειδών σε όλες τις δεκαετίες

Η ανάλυση των τάσεων των ειδών ανέδειξε την εξελισσόμενη δημοτικότητα των ειδών με την πάροδο του χρόνου. Για παράδειγμα, είδη όπως η επιστημονική φαντασία και η φαντασία κέρδισαν έδαφος τη δεκαετία του 1980 και του 1990, πιθανότατα επηρεασμένα από τις τεχνολογικές εξελίξεις και τις πολιτιστικές αλλαγές.

## Ανάλυση παλινδρόμησης και συσχέτισης

Η μελέτη χρησιμοποίησε ανάλυση παλινδρόμησης και συσχέτισης για να εξετάσει:

1. Έτος σε σχέση με τις βαθμολογίες: Η γραμμή παλινδρόμησης έδειξε ότι οι νεότερες σειρές λαμβάνουν γενικά υψηλότερες βαθμολογίες, γεγονός που υποστηρίζεται από μια θετική κλίση στην εξίσωση παλινδρόμησης.
2. Είδος και τηλεθέαση: Διαπιστώθηκε αμελητέα συσχέτιση μεταξύ συγκεκριμένων ειδών και βαθμολογιών, γεγονός που υποδηλώνει ότι το είδος από μόνο του δεν αποτελεί καθοριστικό παράγοντα πρόβλεψης της υποδοχής από το κοινό.

## Τεχνικές οπτικοποίησης

Για την αποτελεσματική παρουσίαση των ευρημάτων χρησιμοποιήθηκε μια σειρά εργαλείων οπτικοποίησης, όπως ραβδογράμματα, κυκλικά διαγράμματα, διαγράμματα διασποράς και γραμμικά γραφήματα. Παραδείγματος χάριν:

* Τα ραβδογράμματα απεικόνιζαν τη δημοτικότητα του είδους και τη μέση βαθμολογία.
* Τα διαγράμματα διασποράς απεικόνιζαν τις συσχετίσεις μεταξύ χρόνου εκτέλεσης, δημοτικότητας και βαθμολογίας.
* Τα γραμμικά γραφήματα παρουσίασαν τις τάσεις στην παραγωγή σειρών και τις βαθμολογίες κατά τη διάρκεια δεκαετιών.

Αυτές οι απεικονίσεις βελτίωσαν την ερμηνευσιμότητα και την επικοινωνία πολύπλοκων μοτίβων μέσα στα δεδομένα.

## Συμπέρασμα και επιπτώσεις

Η ανάλυση παρέχει μια ολοκληρωμένη επισκόπηση των τάσεων στην παραγωγή τηλεοπτικών σειρών, των προτιμήσεων του κοινού και της δυναμικής του κλάδου. Τα βασικά ευρήματα περιλαμβάνουν την αυξανόμενη παραγωγή τηλεοπτικών σειρών, τη διαρκή δημοτικότητα του δράματος και της κωμωδίας, καθώς και τη θετική τάση στις μετρήσεις τηλεθέασης με την πάροδο του χρόνου. Αυτές οι γνώσεις έχουν σημαντικές επιπτώσεις για ακαδημαϊκούς, επαγγελματίες του κλάδου και δημιουργούς περιεχομένου, προσφέροντας καθοδήγηση για μελλοντική έρευνα και λήψη στρατηγικών αποφάσεων στον τηλεοπτικό κλάδο.

## Μελλοντικές ερευνητικές κατευθύνσεις

Περαιτέρω έρευνα θα μπορούσε να διερευνήσει:

* Τον αντίκτυπο των πλατφορμών streaming στην ποικιλομορφία των ειδών και στον όγκο παραγωγής.
* Μια βαθύτερη εξέταση των δημογραφικών χαρακτηριστικών του κοινού και της επιρροής τους στην τηλεθέαση.
* Συγκριτική ανάλυση σε διάφορες χώρες και πολιτισμικά πλαίσια.