

# DataScience

## Házi feladat választás

### Adatszerkezet

4-es számú adathalmaz – [10,000 Data about movies](#)

movie_name:	A film címe	szöveg
year	A film megjelenésének éve	Egész
rating	IMDb felhasználó értékelés	Tört
metascore	Kritikusok összesített értékelése	Tört
gross_income	Bevétel	Pozitív egész szám
votes	Felhasználói értékelések száma	Pozitív egész szám
runtime	Játékidő	Pozitív egész szám
genre	Műfaj	Szöveg
certificate	Besorolás	Szöveges
description	Rövid ismertető	Szöveg
directors.	A film rendezője	Szöveges
stars	Főszereplők	Szöveges

### Adatok és téma bemutatása

Az adatok IMDb nevű interneten elérhető filmdatbázisból lettek kinyerve CSV formátumban kutatási és oktatási célokra. Az adathalmaz 10,000 rekordot tartalmaz különböző 1915 és 2023 között megjelent filmekről. A filmek neve és megjelenésük éve mellett olyan adatokat is tartalmaz, mint felhasználói és filmkritikusi értékelések, bevétel, filmhossz, műfaj, besorolás, rendező, rövid leírás, illetve a filmben szereplő színészek.

### Hipotézisek, feltárando elemzések

- Egyszerű összefüggések különböző értékek között
  - Mi kedvenc műfaj?
    - Értékelés-műfaj
  - Mi a legjövedelmezőbb műfaj?
    - Műfaj-bevétel
  - Kik a legnépszerűbb színészek?
  - Metascore és felhasználói értékelések kapcsolata
    - Mennyire korrelálnak ezek az értékek?
    - Mely filmeknél a legmarkánsabb az eltérés?
  - A felhasználói értékelések száma és az értékelések átlaga

- A népszerűbb filmeknél nagyobb a szórás?
- Mely filmeknél a legegységesebb?
- Stb.
- Trendelemzések
  - Hogyan változnak a különböző évtizedekben a filmek tulajdonságai?
    - Játékidő, besorolás, műfaj
  - Hogyan változott a bevétel?
    - Mai dollárárfolyamra átszámolva
- Komplexebb vizsgálat
  - A nemzetközi események hogyan hatnak a háborús filmekre?
    - Tükröződik az aktuális „főellenség”
    - Érezhetően megnövekedik a háborús filmek száma, valamilyen nemzetközi események hatására
    - Milyen a fogadtatásuk? Népszerűek-e egyáltalán?