



# Full TMDb Movies Dataset 2024 (1M Movies)

*Mihailo Vukorep IN 40/2021*

*Marko Kolarski IN 60/2021*

---

# O dataset-u

- Naziv: Full TMDB Movies Dataset 2024 (1M Movies)
- Kontekst: TMDB (The Movie Database) je skup podataka o filmovima koji uključuje detalje poput naslova, ocene, datuma objavljivanja, prihoda, žanrova...
- Podaci prikupljeni sa TMDB-a (Kaggle)
- Broj zapisa: 1,281,555 filmova
- Broj atributa: 24
- Veličina: 550 MB
- Jedan dokument: “TMDB\_movie\_dataset\_v11.csv”

# Semantika obeležja

- **id** - Jedinstveni identifikator za svaki film.,
- **title** - Naslov filma.,
- **vote\_average** - Prosečan glas ili ocena koju su dali gledaoci.,
- **vote\_count** - Ukupan broj glasova dobijenih za film.,
- **status** - Status filma (npr. Objavljen, Glasine, Postprodukcija, itd.),
- **release\_date** - Datum objavljivanja filma.,
- **revenue** - Ukupan prihod koji je film generisao.,
- **runtime** - Trajanje filma u minutima.,
- **adult** - Označava da li je film pogodan samo za odraslu publiku.,
- **backdrop\_path** - URL slike pozadine za film.,
- **budget** - Budžet dodeljen za film.,
- **homepage** - Zvanični URL početne stranice filma.,
- **imdb\_id** - IMDb ID filma.,
- **original\_language** - Originalni jezik na kojem je film snimljen.,
- **original\_title** - Originalni naslov filma.,
- **overview** - Kratak opis ili rezime filma.,
- **popularity** - Ocena popularnosti filma.,
- **poster\_path** - URL slike postera filma.,
- **tagline** - Krilatica ili nezaboravna rečenica povezana sa filmom.,
- **genres** - Lista žanrova kojima film pripada.,
- **production\_companies** - Lista produkcijskih kuća uključenih u film.,
- **production\_countries** - Lista zemalja uključenih u produkciju filma.,
- **spoken\_languages** - Lista jezika koji se govore u filmu.,
- **keywords** - Ključne reči povezane sa filmom.

# Primer prostiranja dokumenata

```
{ } output.json > { } media
1  {
2    "_id": "tt1375666",
3    "movie_id": 27205,
4    "title": "Inception",
5    "original_title": "Inception",
6    "overview": "Cobb, a skilled thief who commits corporate espionage by
    infiltrating the subconscious of his targets is offered a chance to regain
    his old life as payment for a task considered to be impossible:
    'inception', the implantation of another person's idea into a target's
    subconscious.",
7    "tagline": "Your mind is the scene of the crime.",
8
9    "ratings": {
10     "vote_average": 8.364,
11     "vote_count": 34495,
12     "popularity": 83.952
13   },
14
15   "release_info": {
16     "status": "Released",
17     "release_date": {
18       "year": 2010,
19       "month": 7,
20       "day": 15,
21       "full_date": "2010-07-15"
22     },
23     "original_language": "en",
24     "spoken_languages": ["English", "French", "Japanese", "Swahili"]
25   },
26
27   "content_info": {
28     "adult": false,
29     "runtime": 148,
30     "genres": ["Action", "Science Fiction", "Adventure"]
31   },
32 }
```

```
33   "financial": {
34     "budget": 160000000,
35     "revenue": 825532764
36   },
37
38   "production": {
39     "companies": ["Legendary Pictures", "Syncopy", "Warner Bros. Pictures"],
40     "countries": ["United Kingdom", "United States of America"]
41   },
42
43   "media": {
44     "poster_path": "/oYuLEt3zVCKq57qu2F8dT7NIa6f.jpg",
45     "backdrop_path": "/8ZTVqvKDQ8emSGUEMjsS4yHAwrp.jpg",
46     "homepage": "https://www.warnerbros.com/movies/inception",
47     "imdb_id": "tt1375666"
48   },
49
50   "keywords": [
51     "rescue", "mission", "dream", "airplane", "paris", "france",
52     "virtual reality", "kidnapping", "philosophy", "spy", "allegory",
53     "manipulation", "car crash", "heist", "memory", "architecture",
54     "los angeles", "california", "dream world", "subconscious"
55   ]
56 }
```

# Primer agregacije podataka

1. Koliki je prosečan prihod po filmu produkcijskih kuća čiji su filmovi imali budžet veći od 50 miliona dolara?
2. Kako se prosečna ocena filmova u različitim žanrovima menjala tokom decenija?
3. Koji meseci u godini imaju najveću stopu objavljivanja blockbuster (hit) filmova (budžet > 100M)?
4. Koje su najprofitabilnije kombinacije žanrova?
5. Prosečno trajanje filma po zemlji produkcije sa ocenom iznad 7.0, uzimajući u obzir samo zemlje koje su proizvele više od 100 filmova?
6. Koje ključne reči su se najviše pojavljivale u poslednjih 15 godina među prvih 10% filmova sa najvećom zaradom?
7. Koji dani u nedelji imaju najviše premijera filmova sa najvećim budžetom?
8. Koje su produkcijske kuće sa najvećim profitom za horor filmove proizvedene u vrhuncu sezone (oktobar/novembar) sa budžetima većim od 5 miliona?
9. Koje produkcijske kuće prave najduže filmove za različite žanrove?
10. Koji filmovi u poslednjih 10 godina u žanru naučne fantastike imaju najveći budžet po minutu trajanja?