

Analiza filmova sa IMDb-a

Fakultet tehničkih nauka, Novi Sad
Sistemi baza podataka

Stefan Matković
Stefan Kaličanin

Opis skupa podataka

- <https://www.kaggle.com/datasets/komalkhetlani/imdb-dataset>

Skup podataka o IMDb-u u periodu od 1888 do 2023.

IMDb je platforma koja sadrži podatke o filmovima i serijama sa njihovim prosečnim ocenama, kao i podaci o glumcima, režiserima, direktorima i ostalim.

Semantika kolona - podaci o naslovima

- tconst – jedinstveni identifikator naslova
- titleType – tip naslov
- primaryTitle – naslov (popularniji)
- originalTitle – naslov (originalni)
- isAdult – da li je naslov namenjen odraslima
- startYear – godina izdanja
- endYear – godina završetka snimanja serije
- runtimeMinues – minuti trajanja
- genres – žanr
- region – mesto snimanja
- directors – jedinstveni identifikator reditelja
- writers – jedinstveni identifikator pisca
- averageRating – prosečna ocena
- numVotes – ukupan broj glasova
- parentTconst – jedinstveni identifikator serije kojoj epizoda pripada
- seasonNumber – redni broj sezone
- episodeNumber – redni broj epizode

Semantika kolona - podaci o osobama

- nconst – jedinstveni identifikator osobe
- primaryName – ime i prezime osobe
- birthYear – godina rođenja
- deathYear – godina smrti
- primaryProfession – top 3 zanimanja
- knownForTitles – naslovi po kojima je poznata osoba
- category – uloga
- characters – naziv lika

Logička šema – podaci o osobama

```
{  
  "nconst": "nm0000001",  
  "primaryName": "Fred Astaire",  
  "birthYear": 1899,  
  "deathYear": 1987,  
  "primaryProfession": "soundtrack,actor,miscellaneous",  
  "knownForTitles": "tt0053137,tt0031983,tt0050419,tt0072308"  
}
```

```
{  
  "tconst": "tt0000001",  
  "ordering": 1,  
  "nconst": "nm1588970",  
  "category": "self",  
  "job": "\\N",  
  "characters": [  
    "Self"  
  ]  
}
```

Logička šema – podaci o naslovima

```
{  
  "tconst": "tt0000001",  
  "titleType": "short",  
  "primaryTitle": "Carmencita",  
  "originalTitle": "Carmencita",  
  "isAdult": 0,  
  "startYear": 1894,  
  "endYear": "\\N",  
  "runtimeMinutes": 1,  
  "genres": "Documentary,Short"  
}
```

```
{  
  "tconst": "tt0000001",  
  "averageRating": 5.6,  
  "numVotes": 1694  
}
```

```
{  
  "tconst": "tt0041951",  
  "parentTconst": "tt0041038",  
  "seasonNumber": 1,  
  "episodeNumber": 9  
}
```

Predlozi agregacija

- Koji žanr sadrži najviše filmova?
- Koliko epizoda ima serija koja ima najveću prosečnu ocenu?
- Koja sezona je najbolje ocenjena u okviru serije?
- U kojim regionima se najčešće snimalo?
- Koliko je odstupanje prosečne ocene korisnika za neku seriju u odnosu na prosek serije dobijen na osnovu prosečnih ocena epizoda te serije?
- Koji reditelj je snimio najviše filmova?
- Top 5 glumaca u odnosu na broj snimljenih filmova?
- Pronaći osobe koji istovremeno obavljaju funkciju reditelja i glumca.
- Koja kombinacija žanra je najpopularnija?
- Koliki je broj umrlih osoba koje su učestvovala u snimanju filmova objavljenih tokom 21. veka?