# KV1 - Definiranje projektnog zadatka

## Projektni zadatak

Naziv zadatka: Vizualizacija top 250 najbolje ocijenjenih filmova sa IMBD-a 2023. godine

Opis problema: Podatci sadrže informacije o 250 najpopularnijih/najboljih filmova 2023. godine. Potrebno je proučiti podatke radi stvaranja veza I zaključaka vezano uz razne aspekte ovih filmova kao što su: godinja izdanja, trajanje, ocjena , sveukupni rank, žanra itd.

Opis zadatka: Vizualizirati podatke i sazanja o najbolje ocijenjenim filmovima u 2023. godini.

Cilj projekta: Steći uvid u odnose među podatcima koji opisiju filmove iz ovog podatkovnog skupa.

Poveznica na git repozitorij projekta: https://github.com/adrijanmalinovic/Vizualizacija-Podataka-KV

## Podatci

* + 1. *U ovom zadatku potrebno je pronaći odgovarajuće izvore podataka koji će se koristiti za rješavanje problema definiranog u prvom zadatku. Važno je osigurati da su podaci kvalitetni i relevantni za problem koji se rješava te da su dostupni za upotrebu.*

[*https://www.kaggle.com/datasets/ashishjangra27/imdb-top-250-movies*](https://www.kaggle.com/datasets/ashishjangra27/imdb-top-250-movies)

* + 1. *Potrebno je opisati odabrane podatke*

Podatkovni skup sadrži podatke o 250 filmova prikupljenih 2023. g. Podatci su u csv formatu, a sadrže razne stupce kao što su ime filma, rang po ocjenjenosti (1 –“najbolji”, 250 – “najlošij”), naslovna slika, godina izdanja, glasovi na IMBD-u, ocjenu sa IMBD-a, trajanj, listu direktora, glumaca, pisaca te druge podatke.

## Obrada podataka

* + 1. *Obraditi prikupljene podatke i povezati ih kako bi se stvorio cjelovit skup podataka. Ovo uključuje čišćenje i obradu podataka, kao i provjeru njihove konzistentnosti, aktualnosti, cjelovitosti, tj. kvalitete i ispravnosti.*

Podatke treba restrukturirati u JSON format radi lakšeg korištenja u D3 biblioteci. Također je potrebno obrisati razne stupce sa metapodatcima vezanim uz dataset kao što su ID filma, glumaca I direktora. Oni se neće koristiti u vizualizaciji. Neke stupci kao što su direktori i glumci, koji su originalno obliku jednog stringa, je poželjno prebaciti u polja stringova radi lakšeg pristupa određenim elementima ako za to bude potrebe. Također je poželjno preimenovati imena JSON ključeva radi konzistentnosti i jasnoće. Budući da u podatkovnom skupu nema nedostajućih ili pogrešnih vrijednosti, u tom pogledo nije potrebno ništa mijenjati.

## Relevantne vrste prikaza za korištene podatke

* + 1. *Predložiti moguće načine prikaza podataka koji će pomoći u razumijevanju podataka i rješavanju problema koji je postavljen u prvom zadatku. Ovo može uključivati odabir najprikladnijeg načina vizualizacije podataka, ali to je zadatak iduće vježbe.*

Koristit će se ponajviše stupčasti dijagrami, histogrami, pita dijagrami, liste i text, te *grid* prikazi sa slikama. Ideja je vidjeti, primjerice, kako na rangiranje/ocjenu filma ovisi trajanje filma, koji žanrovi su najbolje rangirani, utječe li zanimljivija naslovna slika na ocjenu filma itd.