# KV1 - Definiranje projektnog zadatka

## Projektni zadatak

[Opisati glavni cilj projekta i svrhu vizualizacije podataka.  
*Ovaj zadatak zahtijeva jasno definiranje problema koji se pokušava riješiti. Potrebno je postaviti jasne ciljeve i granice projekta, kao i odrediti koji će se podaci prikupljati i analizirati kako bi se riješio problem.*]

**Naziv zadatka:**

Vizualizacija podataka Yu-Gi-Oh! karata

**Opis problema:**

Dataset od preko 13 000 karata s desecima atributa (ATK, DEF, Level, views, ban-status, itd.) preopterećen je informacijama koje je teško brzo i intuitivno pregledati te usporediti, što onemogućuje brzu identifikaciju najmoćnijih, najpopularnijih ili zabranjenih karata.

**Opis** **zadatka**:

Razviti web-aplikaciju u D3.js koja čita podatke o kartama, obrađuje ih i prikazuje interaktivne grafikone za snagu karata (ATK vs DEF vs Level), popularnost (views, upvotes/downvotes, staple) i status zabrana (TCG/OCG/Goat) kroz vrijeme, uz filtere, sortiranje i mogućnost usporedbe odabranih karata. Podaci će biti prikazani samo za OCG format.

**Cilj projekta:**

Omogućiti igračima i analitičarima Yu-Gi-Oh! brz, jasan i dinamičan uvid u ključne karakteristike karata te promjene u popularnosti i restrikcijama, eliminirajući potrebu za ručnim pretraživanjem ili složenim upitima nad velikim datasetom.

**Poveznica na git repozitorij projekta:**

<https://github.com/durcevicjuraj/ygo-cards-data-visualization>

## Podatci

[Pronaći izvore podataka i opisati podatke koji će se koristiti za vizualizaciju.]

* + 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.*
    2. *Potrebno je opisati odabrane podatke kao i navesti pripadajuće izvore.*

Dataset je preuzet sa stranice kaggle (<https://www.kaggle.com/datasets/ioexception/yugioh-cards>).

Informacije koje dataset ima preuzete s <https://ygoprodeck.com/> putem njihovog API-ja. YGOProDeck je trenutno najkorišteniji izvor informacija o Yu-Gi-Oh! kartama i stoga sam u potpunosti siguran u pouzdanost informacija.

* + **Opis:** sadrži osnovne podatke o ~13 000 karata (Monster, Spell i Trap).
* **Polja:**
* id, name – jedinstveni identifikator i naziv karte
* type, desc – tip i opis efekta
* atk, def, level – statistike čudovišta
* race, attribute, scale – kategorijske karakteristike
* archetype, linkval, linkmarkers – napredniji atributi
* image\_url[\_small] – veze na slike karata
* ban\_tcg, ban\_ocg, ban\_goat – status zabrane u različitim formatima
* staple – oznaka često korištenih karata
* views, viewsweek, upvotes, downvotes – mjere popularnosti
* formats, treated\_as – meta-informacije o formatu igre
* tcg\_date, ocg\_date – datum prvog izdanja u TCG/OCG
* konami\_id, has\_effect – dodatni tehnički podaci

## Obrada podataka

[Opisati obavljeni postupak obrade i povezanja 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.*

Koristi se samo cards.csv iz dataseta. Ostali .csv koji se dobiju u datasetu nisu potrebni za ovaj projekt.

Brisanje stupaca:

* **ban\_tcg**, **ban\_goat** → brišem jer ne trebaju statusi zabrana za TCG i Goat formate
* **formats**, **treated\_as** → brišem jer su vezani uz meta-informacije i nisu mi potrebne
* **tcg\_date** → brišem jer je važan samo datum izdavanja u OCG (tj. **ocg\_date**)
* **konami\_id** → brišem jer ima već stupac **id**
* **viewsweek** → brišem jer mi je bitniji stupac **views**

Čišćenje i obradu podatak sam odradio pomoću Data Wrangler Visual Studio Code ekstenzije.

<https://marketplace.visualstudio.com/items?itemName=ms-toolsai.datawrangler>

Python kod:



## Relevantne vrste prikaza za korištene podatke

[Predložiti nekoliko različitih načina prikaza podataka koji bi bili prikladni za ovaj projekt]

* + 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.*
* Radar (spider) graf za usporedbu dvije karte

– Svaka os predstavlja jednu metriku: atk, def, level i has\_effect

– Možda upvotes/downvotes (omjer pozitivnih i negativnih glasova) kao još jedna os

– Dvije overlaid poligonalne linije – svaka za jednu odabranu kartu

– Sa strane će biti prikazane slike karata i njihove ostale informacije koje nisu unutar spider grafa

* Timeline (linijski ili area graf)

– X-os = godina (ocg\_date), Y-os = broj novih karata izdanih te godine.

– Dodatna linija ili shaded area za kumulativni broj karata u zabrani.

– Pomaže razumjeti kada se pojavljivalo najviše novih izdanja i kako se popunjavala ban-lista kroz vrijeme.

* Bar chart popularnosti

– Stupčasti graf top N karata po jednoj od metrika (views, upvotes, downvotes ili omjer upvotes/downvotes).

– Mogućnost odabira metrike padajućim izbornikom i tranzicija poretka (enter-update-exit).

– Jasan pregled najgledanijih ili najbolje/najlošije ocijenjenih karata.

* Bar chart duljine deskripcije karte

– Stupčasti graf koji prikazuje prosječnu duljinu deskripcije karte u određenoj godini.

– Uvid u povećanje složenosti igre kroz godine.

* Scatter plot ATK vs DEF

– Pojedina točka = karta

– X-os = atk, Y-os = def

– Veličina točke = level

– Boja = attribute ili race (ili staple status).

– Pomaže da se odmah uoče klasteri (slabe, srednje i jake karte) i outlieri.

* Stacked ili obični bar chart ban-statusa

– Četiri grupe stupaca za OCG: Banned / Limited / Semi-limited / Unlimited

– Svaki stupac prikazuje relativan udio karata u toj kategoriji zabrane.

– Brza usporedba omjera karata po statusu zabrane.