



# Igre na Steam

Predlog projekta u  
MongoDB

# Opis skupa podataka



Skup podataka preuzet je sa sajta kaggle (link ka skupu u poslednjem slajdu). Podaci su skupljani kao vežba u obradi velikih podataka od strane autora.

Podaci se bave igrama na platformi Steam. Steam svojim korisnicima nudi elektronsku prodaju i distribuciju softvera, pretežno igara za računar.

# Opis skupa podataka



Podaci su podeljeni u tri skupa, skladištena u .csv (comma separated value) formatu.

app\_list.csv sadrži spisak svih sadržanih igara i njihov jedinstveni ID svih 29158 obrađenih igara.

steam\_app\_data.csv sadrži podatke skupljene sa same Steam platforme o različitim igrama. U pitanju su podaci koje potencijalni kupci vide o igri, na osnovu kojih može da se odluči da li želi da je kupi ili ne. Među obuhvaćenim podacima su ime igre, datum kada je ušla u prodaju i koji razvojni tim ju je napravio.

steamspy\_data.csv su podaci prikupljeni sa SteamSpy platforme, koja nudi dodatne podatke o igrama na Steam. Neki podaci u tom skupu su redundantno ponovljeni iz steam\_app\_data.csv, ali ima i novih podataka, poput skupa tagova koje su korisnici dodelili igrama, I broj pozitivnih I negativnih ocena koje je igra dobila.

# Opis semantike kolona



Sledi opis semantike dela kolona iz skupa `steam_app_data.csv`:

- Name: naziv igre
- Required\_age: minimalna preporučena starost igrača. Ukoliko ima vrednost null, smatra se da je igra prilagođena svim uzrastima.
- is\_free: da da li je potrebno kupiti igru pre instalacije.
- Controller\_support: da li igra nudi softversku podršku za periferne uređaje osim miša i tastature, poput gejmpeda, džojstika ili volana.
- Dlc: ukoliko igra ima ekspanzije, ili dlc (Down-Loadable Content), u ovom polju se nalazi spisak id-ova svih ekspanzija. Ukoliko nema vrednosti, igra nema ekspanzije na Steam platformi.
- Developers: spisak razvojnih timova koji je dizajnirao igru.
- Publishers: spisak izdavačkih kuća koje izdaju igru.
- Categories: spisak kategorija kojima igra pripada.
- Generas: spisak žanrova kojima igra pripada.
- Recommendations: Koliko korisnika je istaklo da bi preporučili igru drugima.



# Opis semantike kolona



Sledi opis semantike dela kolona iz skupa steamspy\_data.csv:

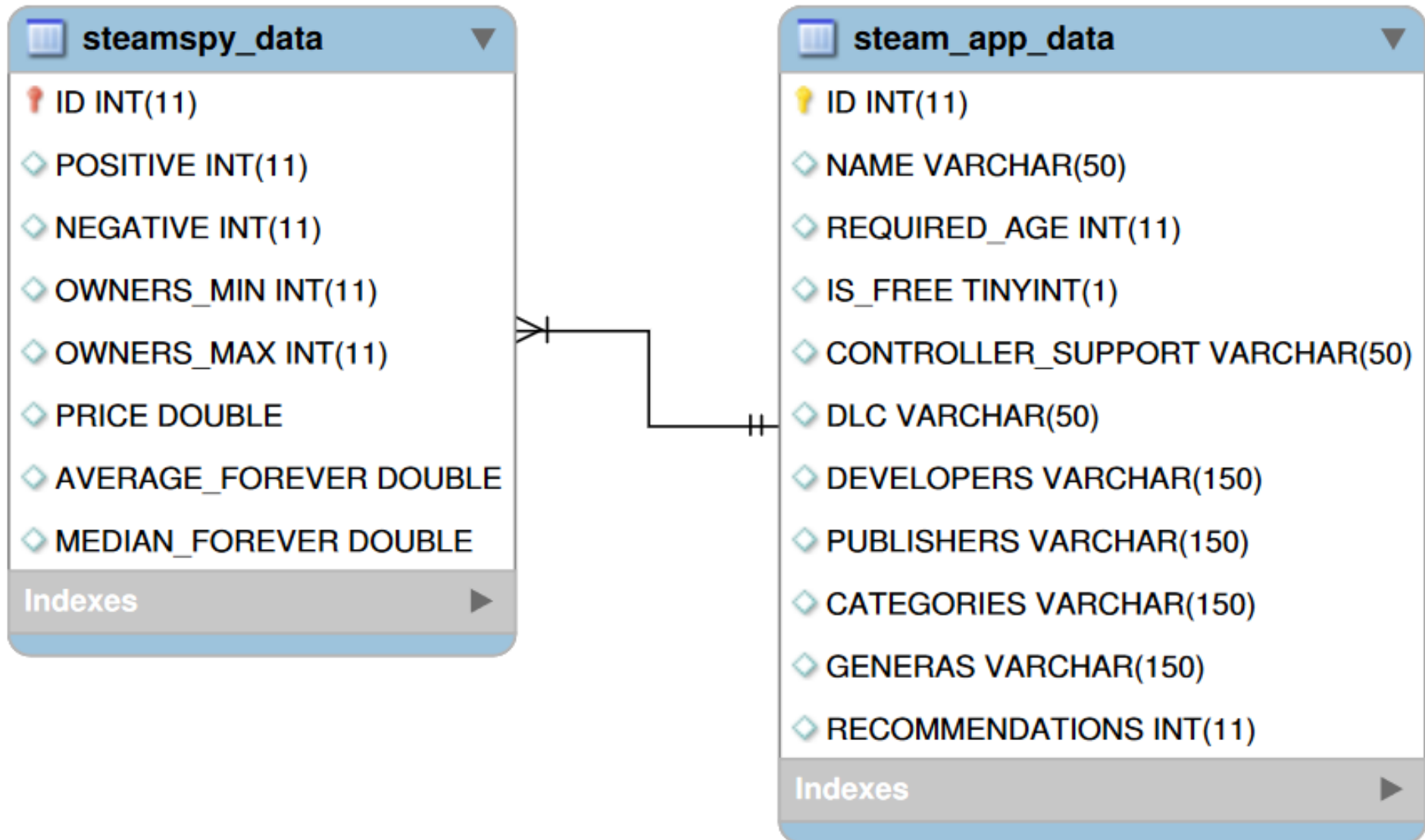
- Positive: broj korisnika koji je igru pozitivno ocenio
- Negative: broj korisnika koji je igru negativno ocenio
- Owners: procenena gornje i donje granice korisnika koji je posedovao igru. Napomena: ako je igra besplatna, smatra se da ju je posedovao svaki korisnik koji ju je bar jednom igrao.
- Price: cena igre
- Average\_forever: prosečno vreme koje igrači provedu u igri
- Median\_forever: medijana vremena koje korisnici provedu u igri

# Opis prostiranja podataka Kroz datoteke



Prostiranje podataka kroz datoteke vrši se pomoću id-polja igre, koje je za određenu igru isto u sva tri skupa podataka. Sva tri skupa podataka opisuju iste igre, ali ne sadrže iste podatke o njima. Skupljeni su sa različitih sajtova i pratili različite attribute.

# Predlog scheme baze



# Predlog agregacije podataka



- Da li su Valve igre bolje ocenjene od proseka?
- Koji razvojni studio ima najviše sati u svojim igrama?
- Koji izdavači imaju najveći broj dlc (Down-Loadable Content)?
- Koje igre su najisplativije? (broj sati igre/cena)
- Da li su besplatne igre bolje ocenjene od onih koje se naplaćuju?
- Da li više korisnika igra besplatne igre?
- Koji žanrovi su najpopularniji?
- Za 10 najpopularnijih žanrova, kojih 5 igara po žanru imaju najviše sati u igri?
- Za 10 najpopularnijih žanrova, kolika je prosečna cena igre?
- Sa koliko razvojnih studija prosečno sarađuje jedna izdavačka kuća?
- Da li postoji veza između meseca u godini i broja igara koje su u tom mesecu objavljene (postoji li nešto poput sezone za izdavanje igrica)?





# Hvala na pažnji! :)

Link ka skupu podataka:

[https://www.kaggle.com/nikdavis/steam-store-raw?select=app\\_list.csv](https://www.kaggle.com/nikdavis/steam-store-raw?select=app_list.csv)

Nikola Arsenijević

ra34 – 2016